Decoding Heterogeneity: Exploring Cellular Landscapes through Single Nuclei RNA Sequencing at the Pre-eclamptic Maternal-Fetal Interface and in Pancreatic Neuroendocrine Carcinoma for Precision Medicine Insights

Inaugural-Dissertation
to obtain the academic degree
Doctor rerum naturalium (Dr. rer. nat.)
submitted to the Department of Biology, Chemistry, Pharmacy
of Freie Universität Berlin

Submitted by
Olivia Debnath
2024

Doctorate Duration: August 2019 – December 2023

Doctorate Supervisor: Prof. Dr. Roland Eils

Doctorate Hosting Institutes: BIH Center of Digital Health, Berlin Institute of Health (BIH) at Charité – Universitätsmedizin Berlin 10117 Berlin, Germany

First Reviewer: Prof. Dr. Roland Eils
BIH Center of Digital Health, Berlin Institute of Health (BIH) at Charité-Universitätsmedizin Berlin

&

Department of Mathematics and Computer Science, Institute of Mathematics, Freie Universität Berlin

&

Health Data Science Unit, Bioquant, Medical Faculty, University of Heidelberg, Germany

Second Reviewer: Prof. Dr. Sigmar Stricker
Institute of Chemistry and Biochemistry, Department of Biology, Chemistry and Pharmacy,
Freie Universität Berlin

Date of Thesis Defense: 24.06.2024

Acknowledgment

I am filled with gratitude as I take this opportunity to acknowledge the extraordinary people who have been instrumental in shaping both my journey and the realization of this doctoral thesis. This has truly been an incredible experience, and its realization wouldn't have been possible, nor half as exhilarating, without the unwavering support, encouragement, guidance, and expertise generously offered by those around me. Working, collaborating, and sharing moments with such a wonderful group of people has been a privilege, and I am exceptionally fortunate to have had this opportunity.

As I conclude another chapter of my life, reflecting on this journey brings me joy — a journey that began as a curious teenager in a small town near Kolkata. It involved sleepless nights in the Molecular Signaling laboratory at IISER, conducting molecular cloning, biochemistry assays, exploring protein structures through MD simulations, Biophysics, and network analysis. My passion for medical research burgeoned from witnessing grappling poverty in India, and how the lack of equitable treatment plans exacerbates the struggle for basic access to healthcare. Siddhartha Mukherjee's 'Emperor of Maladies' deepened my interest in cancer research, highlighting the unshakable spirit of scientists, patients, and caregivers in the ongoing struggle against this ancient and formidable disease. Discovering the Human Cell Atlas project, popularly known as the 'Periodic Table of Cells,' during my undergraduate years, which promised to unravel the complexities of human biology at the cellular level and offer a pathway for personalized and targeted medical interventions, inspired my pursuit of doctoral training within the Computational Oncology group at the Center of Digital Health, BIH/Charité.

My heartfelt gratitude extends to my supervisors, Dr. Naveed Ishaque, and Prof. Dr. Roland Eils. Your profound expertise, guidance, and encouragement, coupled with the exceptional work environment you have cultivated, played a pivotal role in shaping my growth as a researcher. The independence you granted, particularly in participating in critical translational projects and collaborations, has fostered resourcefulness, integrity, and determination. I genuinely thank you for the opportunities provided, for sparking scientific curiosity in me, and for being incredibly supportive mentors. Naveed, my heartfelt thanks for closely supervising throughout the two major studies of my doctoral project. Your scientific insights, guidance, and constructive criticisms have been instrumental in developing an analytical mindset, adaptability, and maintaining a collaborative spirit. Beyond these, I am thankful for your emphasis on instilling good scientific practices, promoting transparency, and nurturing a

commitment to moral values within the group. I firmly believe that pursuing a doctorate in a different lab would not have been as enriching, and I am humbled that our paths crossed.

I express sincere gratitude to Prof. Dr. Sigmar Stricker, who graciously agreed to serve as my second reviewer for my doctoral thesis. I am thankful for his kindness, insightful scientific exchanges, and valuable feedback during the thesis writing process.

My deepest gratitude extends to Prof. Dr. Bertram Widenmann, Prof. Aurel Perren, Dr. Katharina Detjen, and Dr. Hilmar Berger for their invaluable collaboration on pancreatic neuroendocrine carcinoma during my doctoral project. Each of you, as physician scientists, brought expertise and insights that added immense depth to our work. Special appreciation to Katharina and Hilmar for the enriching scientific discussions that expanded my knowledge on pancreas, neuroendocrine biology, and bioinformatics, respectively. I would sincerely miss working with you. Bertram, I am indebted for your exceptional mentorship, both within and beyond the workplace, for inspiring me toward translational research and offering generous assistance during my illness.

A heartfelt thanks to Dr. Luca Tosti and Dr. Sören Lukassen, individuals I deeply admire. Luca, your guidance on the HCA pancreas project not only sparked my interest in pancreas biology but also instilled in me the importance of being inquisitive and rigorous in scientific pursuits. Sören, our collaboration on the COVID-19 project accelerated my learning curve in Bioinformatics, and I appreciate the valuable scientific exchanges and your support throughout my doctorate.

During this time, I had the privilege of collaborating with two distinguished people- Prof. Dr. Christian Conrad and Prof. Dr. Irina Lehmann on major side projects that significantly enriched my research journey. I am thankful that working with you enhanced the overall depth and impact of our joint projects.

Sebastian Tiesmeyer (“t-SNE”), my companion, has been a beacon of brilliance and kindness throughout our academic journey together. As we approach the end, I appreciate our shared memories, scientific exchanges, and your unyielding support. Special thanks for your extensive spatial transcriptomics analysis on the pre-eclampsia project, serving as a vital validation of my computational findings, and opening doors to new projects. Let's not be strangers, and I look forward to our future collaborations, perhaps over bread-butter and Indian curries!

To our collaborators on the maternal-fetal atlas and pre-eclampsia project in Berlin and Graz, I thank you for your trust and enriching exchanges. Witnessing the evolution of our work from comprehensive single-cell omics analytics to gaining mechanistic insights into pre-eclampsia biology fills me with joy. I am grateful to my namesake, Dr. Olivia Nonn and Dr. Florian Herse for their invaluable assistance in sample recruitment, project management, scientific discussions and organizing validations of my key data-driven results in independent validation cohorts. I am equally thankful to Prof. Dominik Muller, and Dr. Ralf Denchend for allowing me to make significant contributions to this project. I thoroughly enjoyed intriguing discussions with Prof. Berthold Huppertz on placenta biology, particularly in gaining a deeper understanding of the function of a novel trophoblast state that I deciphered through single-nuclei RNA-seq analysis. I also extend my gratitude to other project members- Daniela, Katja & Cornelius for their support and helpful discussions. Overall, this collaboration has fueled my enthusiasm for pregnancy research and reproductive medicine. This prompted me to start a new project with Dr. Andreas Ian Lackner and Prof. Jürgen Pollheimer, delving into the investigation of the maternal immune system's role in supporting pregnancy through multi-omics approaches. I also thank Andreas and Jürgen for insightful discussions on decidua and placenta biology. Many thanks to Kerim for our consistent scientific discussions and collaboration on the pre-eclampsia project. Your steadfast mental support during challenging periods, including family bereavement and burnout phases, helped me navigate through tough times during doctorate.

To the members of our Computational Oncology group- Niklas, Shashwat, Johanna, Alexander & Sebastian M, working alongside professionals of your caliber has been a privilege. I appreciate the moments we've shared, your support in both good and bad times, and engaging scientific discussions. Special thanks to Niklas and Johanna for proofreading and assisting with the German translation of my abstract, an important step given my non-German-speaking background. I would sincerely thank the cluster managers, especially Stefan, for all his help and support while working on Eils-HPC and navigating technical issues. I also thank Franzi, Alexandra and Wiebke for their cooperation throughout my doctorate.

I had the privilege of working with Prof. Dr. Pascal Falter Braun at HMGU Munich on protein interaction networks, an intellectually enriching experience during my academic journey. His guidance significantly contributed to my understanding of complex biological systems, intricacies of biological networks, and inspired me to pursue doctoral research. I am also thankful to Dr. Rahul Das for the great scientific exchanges, advice, and mentorship during my Master’s thesis dissertation at his lab.

To the members of Global Shapers Berlin (including Felix, Caroline & Marcela), I feel obliged to have been part of your projects addressing pressing social issues that extend beyond conventional volunteering. I appreciate that our collaborative efforts not only generated meaningful impact but also fostered lasting friendships. I am truly thankful to a special group of individuals at Entrepreneur First, Excellerator'22 and my collaborators during Leaders of Tomorrow 2020 Cambridge (especially Matthew & Arinze) for inspiring me to aim for an impact beyond academia and offered support to venture into the realm of entrepreneurship. Manuel, Gundars & Agnes- I am certainly glad that our paths crossed, and I look forward to many more years of friendship.

I want to whole-heartedly thank my cousins and wonderful friends outside of work, including the meaningful connections forged at IISER, Munich, Berlin, and beyond. To Sambhasan, Subhodeep & Shashank, I am grateful for your timely advice, feedback on figure panels, and your guidance in navigating Adobe Illustrator. Subhodeep, a special thanks for introducing me to digital painting & InDesign; it really helped to sketch the introductory figures for my thesis. Sambhashan & Abhijeeth, thank you for being such generous friends and offering help whenever one needs. To Abdelwahed, Saroj & Angelo – thank you for providing me with a home away from home, cooking warm meals after my long working hours, and sharing memorable travels together. My life would not be the same without you! Arnab & Saurav, thanks for being my constants for the last 14 years, for your inspiration and celebrating small achievements together. Debasri, I will always admire you for being my comfort place over the last decade. Archismita, I am always grateful for your kindness and support throughout the last couple of years; I've found a sister in you.

To my best friend, I cordially thank you for your constant support throughout my doctoral journey, engaging in scientific discussions and for helping me prioritize my well-being. Your encouragement to maintain composure and resilience amidst uncertain times is truly uplifting. I do cherish our friendship, and I look forward to seeing you soon across the Atlantic.

Lastly, but certainly not least, I want to express my deepest gratitude to my family, particularly my mother, who has been a consistent pillar of support and encouragement throughout my life. My late grandfather and elder uncle, though physically absent, continue to inspire me eternally. Their wisdom, kindness, and steadfast support have profoundly shaped my character, and I embrace their enduring influence as a guiding light in my life.

Declaration of Authorship

I hereby declare that I alone am responsible for the content of my doctoral dissertation and that I have only used the sources or references cited in the dissertation.

Olivia Debnath
Doctoral Candidate

Scientific contributions

I deeply appreciate the interdisciplinary and collaborative nature of my scientific work, which fostered efforts from multiple institutes, and resulted in the successful completion of two major projects addressing two distinct biological systems— human maternal-fetal interface, and pancreatic neuroendocrine carcinoma in the context of development and disease. Separate ‘scientific acknowledgement’ sections are included within each results chapter to express gratitude to individuals involved in particular aspects of the project and to uphold transparency. In cases where there is involvement in generating figures, specific acknowledgments were provided within the figure legends. Also, the rationale for providing selected collaborative data figures were for validation purposes of my work and aiding in storytelling.

If not stated otherwise, all data, results, figures, and discussion presented in this thesis were performed by me under the supervision of Dr. Naveed Ishaque and Prof. Dr. Roland Eils at Digital Health Center, Berlin Institute of Health and Charité – Universitätsmedizin Berlin.

Olivia Debnath
Doctoral Candidate

This thesis is dedicated to the cherished memory of my grandfather
whose guidance shaped my journey,

and

my mother

Abstract

Over the past decade, technological progress has facilitated profiling single cell and single nuclei transcriptomes through innovative next-generation sequencing approaches that empowered the exploration of cellular phenotypes in healthy and diseased tissues with unprecedented precision. This dissertation aimed to unravel a deep understanding of human development and disease biology by studying the underlying cellular architecture and molecular mechanisms using single nuclei RNA sequencing (snRNA-seq). Specifically, two distinct systems were investigated, namely, a) maternal-fetal interface in healthy and early-onset pre-eclamptic pregnancies that comprised of human decidua and placenta and b) pancreatic neuroendocrine carcinoma and its comparison with healthy adult pancreatic cell types or states.

Pre-eclampsia (PE) stands out as one of the most severe pregnancy disorders, characterized by hypertension, proteinuria, cardio-metabolic dysfunctions, and various multi-organ complications, ultimately leading to preterm delivery, maternal mortality, and associated morbidities. Early-onset pre-eclampsia (eoPE) is particularly formidable, typically necessitating delivery before the 34th week of gestation and leading to approximately 80,000 maternal and 500,000 fetal deaths annually. Importantly, eoPE currently lacks adequate biomarkers for early screening and clinical management. Diagnosis relies solely on clinical and biomarker signs in late pregnancy, where maternal and fetal morbidity is often irreversible. This dissertation presents a comprehensive cell atlas of the human maternal-fetal interface, comprising of around 225,000 nuclei from the maternal decidua and fetal placenta obtained during the first trimester, healthy term and eoPE pregnancies, utilizing complex snRNA-seq data harmonization. A novel nuclear state termed juvenile syncytiotrophoblast (STBjuv) was identified, demonstrating previously unexplored transcriptomic diversity and a division of labor within the placental syncytiotrophoblast. Notably, the comparative analysis of decidua and placenta samples from term controls and eoPE revealed differences in composition, differential gene expression and transcriptional regulation patterns, along with changes in signaling pathways. These findings provide mechanistic insights into cell type or state-specific dysregulations occurring in eoPE. Moreover, employing advanced spatial transcriptomics techniques such as In-situ RNA sequencing (ISS) and 10X Visium sequencing facilitated the spatial delineation of cell types and states within the human placenta and established a connection between spatial and transcriptomic heterogeneity, presenting a novel discovery in this evolving field of pregnancy research.

The computational analysis identified a dysregulated syncytiotrophoblast development to serve as an initiation point for eoPE that is characterized by perturbations in transcriptional factors/co-activators, specifically the master-regulator *EP300, FOXO1, SCRT2, FOXO4, FOS,* and *PAX5.* Significantly, the enriched transcriptional regulators exhibited a noteworthy overlap in their downstream targets, imparting functional implications across various signaling pathways, including HIF1, AP1, TGFb, Wnt, PI3K-Akt signaling, and vesicle-mediated transport. Of note, a significant proportion of the perturbed syncytiotrophoblast differentiation drivers included *EP300* (or, p300) regulated fusogenic targets that suggest impaired trophoblast syncytialization as a significant contributing factor in the development of eoPE.

Significantly, the discoveries from this work indicate that eoPE possibly originates in the outer syncytiotrophoblast sub-states in the fetal placenta and is marked by an augmented senescence-associated secretory phenotype (SASP) profile. The heightened senescence resulted from elevated ligand pressure, facilitated by the secretion of *GDF15, INHBA, HSPG2, MIF, TGM2, ADAM9,* and *ADAM12* that could potentially traverse the maternal-fetal interface, and translate the disease from the fetal to the maternal side. Of note, In-situ RNA sequencing (ISS) analysis unveiled a statistically significant proximity between the senescent marker (*INHBA*) and markers of fetal vessels in eoPE. This association is not detected in term controls. Consequently, the findings of this thesis emphasize that a disrupted communication between syncytiotrophoblast sub-states and maternal decidua is the key to a dysregulated maternal-fetal barrier and potentially compromised maternal uterine vessel remodeling in eoPE. Remarkably, the presented data suggest a potential strategy for the prevention, intervention, and clinical management of eoPE through the pharmacological inhibition of these ligands associated with the senescence-associated secretory phenotype (SASP), including *GDF15, HSPG2* and *INHBA.*

The subsequent chapter of this dissertation delved into the molecular intricacies guiding the development of healthy pancreatic islets— specifically, the beta cell type, and next assessed the role of these differentiation drivers in the context of pancreatic cancer. Specifically, this thesis focused on high-grade pancreatic neuroendocrine carcinoma (panNEC) with large cell morphology, a subtype presenting challenges in classification and treatment. While advancements in molecular genetics have progressively uncovered significant inter-tumor heterogeneity, there remains an unexplored realm regarding the extent of intra-tumoral heterogeneity and lineage plasticity. Like the pre-eclampsia study, a snRNA-seq approach was utilized to deconstruct the cellular landscape of panNEC that delineated both shared and

unique malignant sub-states associated with specific signaling pathways and transcriptomic regulatory programs driving tumor pathophysiology and heterogeneous clinical behavior.

Notably, this work identified a shared neuroendocrine sub-state characterized by robust induction of heat shock protein encoding mRNA (HSP+), exhibiting signatures indicative of activation of the unfolded protein response, hypoxia, mTORC/PI3K-AKT signaling, and glycolytic shift. In one patient sample, a unique stromal sub-state depicted enriched YAP/TAZ-associated Hippo signaling alongside mesenchymal and basaloid programs expression, reflecting transcriptomic similarities with pancreatic ductal-adenocarcinoma (PDAC). Furthermore, one of the shared neuroendocrine sub-states was highly proliferative and was characterized by overexpressed *E2F* targets, including Enhancer of Zeste homolog 2 (*EZH2*). Notably, this sub-state demonstrated significant enrichment for PTF1A-regulated targets specific to the brain while repressing the pancreas-specific targets. This observation suggests a shift from a pancreatic lineage program in the cell of origin towards a more generic neuronal phenotype and a concomitant enrichment of signatures related to the DNA damage response, vulnerabilities in cancer stem cells, and chemotherapeutic resistance. Of note, this de-differentiated neuronal program could be exploited as the achilles heel of panNEC for devising therapeutic strategies.

Hence, the presented data unmasked considerable heterogeneity and therapeutic vulnerabilities in high-grade panNEC, emphasizing the importance of tumor profiling for personalized treatment approaches. Of note, it highlights the prospect of clinical intervention to target two shared neuroendocrine sub-states suggesting the feasibility of personalized combination therapies in clinical settings.

In a nutshell, this dissertation extensively explored two distinct but critical systems at a cellular resolution in the context of disease development and progression. The snRNA-seq approach presented an unbiased and deep understanding of disease biology that demonstrated significant translational potential. Notably, in both systems, specific cell (or, nuclear) sub-states were ascribed to the disease origin that added to our understanding on how a disease evolves on a molecular level and suggested potential avenues for therapeutic development. This transition from concentrating on a single molecular target to addressing the underlying cellular dysfunction presents novel opportunities for the clinical targeting of cellular signatures for complex diseases and for developing a new generation of therapeutics.

Zusammenfassung

In den letzten zehn Jahren hat der technologische Fortschritt die Erstellung von Profilen der Transkriptome einzelner Zellen und Zellkerne durch innovative Sequenzierungsansätze der nächsten Generation erleichtert, die die Erforschung zellulärer Phänotypen in gesunden und kranken Geweben mit bisher unerreichter Präzision ermöglichen. Ziel dieser Dissertation war es, durch die Untersuchung der zugrundeliegenden zellulären Architektur und molekularen Mechanismen mit Hilfe der Einzelkern-RNA-Sequenzierung (snRNA-seq) ein tieferes Verständnis der menschlichen Entwicklung und Krankheitsbiologie zu erlangen. Konkret wurden zwei verschiedene Systeme untersucht, nämlich a) die mütterlich-fötale Schnittstelle in gesunden und früh einsetzenden präeklampsischen Schwangerschaften, die aus menschlicher Dezidua und Plazenta besteht, und b) das neuroendokrine Pankreaskarzinom und sein Vergleich mit gesunden erwachsenen Pankreaszelltypen oder -zuständen.

Präeklampsie (PE) ist eine der schwersten Schwangerschaftserkrankungen und ist durch Bluthochdruck, Proteinurie, kardio-metabolische Störungen und verschiedene Multiorgan-Komplikationen gekennzeichnet.Letztlich führt sie zu Frühgeburten, Müttersterblichkeit und damit verbundenen Krankheiten. Besonders bedrohlich ist die früh einsetzende Präeklampsie (eoPE), die in der Regel eine Entbindung vor der 34. Schwangerschaftswoche erforderlich macht und jährlich zu etwa 80 000 mütterlichen und 500 000 fötalen Todesfällen führt. Wichtig ist, dass es bei eoPE derzeit keine geeigneten Biomarker für ein frühzeitiges Screening und klinisches Management gibt. Die Diagnose stützt sich ausschließlich auf klinische Anzeichen und Biomarker in der Spätschwangerschaft, wo die mütterliche und fetale Morbidität oft irreversibel ist. In dieser Dissertation wird zum ersten Mal ein umfassender Zellatlas der menschlichen mütterlich-fetalen Schnittstelle vorgestellt, der 225.000 Zellkerne aus der mütterlichen Dezidua und der fetalen Plazenta umfasst, die während des ersten Trimesters, gesunder Terminschwangerschaften und eoPE-Schwangerschaften gewonnen wurden, wobei eine komplexe snRNA-seq-Datenharmonisierung eingesetzt wurde. Es wurde ein neuartiger Kernzustand identifiziert, der als juveniler Synzytiotrophoblast (STBjuv) bezeichnet wird und eine bisher unerforschte transkriptomische Vielfalt sowie eine Arbeitsteilung innerhalb des plazentaren Synzytiotrophoblasten aufzeigt. Die vergleichende Analyse von Dezidua- und Plazentaproben von Terminkontrollen und eoPE ergab Unterschiede in der Zusammensetzung, der differentiellen Genexpression und den Mustern der Transkriptionsregulierung sowie Veränderungen in den Signalübertragungswegen. Diese Ergebnisse liefern mechanistische Einblicke in zelltyp- oder zustandsspezifische

Dysregulationen, die bei eoPE auftreten. Darüber hinaus erleichterte der Einsatz fortschrittlicher räumlicher Transkriptomik-Techniken wie In-situ-RNA-Sequenzierung (ISS) und 10-fache Visium-Sequenzierung die räumliche Abgrenzung von Zelltypen und -zuständen innerhalb der menschlichen Plazenta und stellte eine Verbindung zwischen räumlicher und transkriptomischer Heterogenität her, was eine neue Entdeckung in diesem sich entwickelnden Bereich der Schwangerschaftsforschung darstellt.

Die computergestützte Analyse identifizierte eine dysregulierte Synzytiotrophoblasten-Entwicklung, die als Initiationspunkt für eoPE dient und durch Störungen bei Transkriptionsfaktoren/Co-Aktivatoren gekennzeichnet ist, insbesondere dem Master-Regulator *EP300, FOXO1, SCRT2, FOXO4, FOS* und *PAX5*. Es ist bezeichnend, dass die angereicherten Transkriptionsregulatoren eine bemerkenswerte Überschneidung bei ihren nachgeschalteten Zielen aufweisen, was funktionelle Auswirkungen auf verschiedene Signalwege hat, darunter HIF1, AP1, TGFb, Wnt, PI3K-Akt-Signalisierung und Vesikel-vermittelter Transport. Bemerkenswert ist, dass ein signifikanter Anteil der gestörten Synzytiotrophoblasten-Differenzierungstreiber *EP300/p300*-regulierte fusogene Ziele umfasste, was darauf hindeutet, dass eine gestörte Trophoblasten-Synzytialisierung ein wesentlicher Faktor für die Entwicklung von eoPE ist.

Die Entdeckungen dieser Arbeit deuten stark darauf hin, dass eoPE in den äußeren Synzytiotrophoblasten-Substadien der fetalen Plazenta entsteht und durch ein erhöhtes Profil des Seneszenz-assoziierten sekretorischen Phänotyps (SASP) gekennzeichnet ist. Die erhöhte Seneszenz resultiert aus einem erhöhten Ligandendruck, der durch die Sekretion von *GDF15, INHBA, HSPG2, MIF, TGM2, ADAM9,* und *ADAM12* begünstigt wird, die potenziell die mütterlich-fetale Schnittstelle passieren und die Krankheit von der fetalen auf die mütterliche Seite übertragen könnten. Bemerkenswert ist, dass die In-situ-RNA-Sequenzierungsanalyse (ISS) eine statistisch signifikante Ähnlichkeit zwischen dem Seneszenzmarker (INHBA) und Markern der fetalen Gefäße bei eoPE aufdeckte. Diese Assoziation wurde bei Terminkontrollen nicht festgestellt. Folglich unterstreichen die Ergebnisse dieser Arbeit, dass eine gestörte Kommunikation zwischen Synzytiotrophoblasten-Subzuständen und der mütterlichen Dezidua der Schlüssel zu einer dysregulierten mütterlich-fetalen Barriere und einem potenziell beeinträchtigten mütterlichen Uterusgefäßumbau bei eoPE ist. Bemerkenswerterweise legen die vorgestellten Daten eine potenzielle Strategie zur Prävention, Intervention und klinischen Behandlung von eoPE durch die pharmakologische Hemmung dieser Liganden nahe, die mit dem seneszenzassoziierten sekretorischen Phänotyp (SASP) assoziiert sind, einschließlich *GDF15, HSPG2* und *INHBA*.

Das folgende Kapitel dieser Dissertation befasste sich mit den molekularen Feinheiten der Entwicklung gesunder Pankreasinseln, einschließlich der Beta- und Alphazelltypen, und bewertete anschließend die Rolle dieser Differenzierungsfaktoren im Zusammenhang mit Pankreaskrebs. Im Mittelpunkt dieser Arbeit stand das hochgradige neuroendokrine Karzinom der Bauchspeicheldrüse (panNEC) mit großzelliger Morphologie, ein Subtyp, der eine Herausforderung bei der Klassifizierung und Behandlung darstellt. Während Fortschritte in der Molekulargenetik nach und nach eine erhebliche Heterogenität zwischen den Tumoren aufgedeckt haben, bleibt das Ausmaß der intra-tumoralen Heterogenität und die Plastizität der Abstammung unerforscht. Wie bei der Präeklampsie-Studie wurde ein snRNA-seq-Ansatz verwendet, um die zelluläre Landschaft von panNEC zu dekonstruieren. Dabei wurden sowohl gemeinsame als auch einzigartige maligne Unterzustände beschrieben, die mit spezifischen Signalwegen und transkriptomischen Regulationsprogrammen verbunden sind, die die Pathophysiologie und das heterogene klinische Verhalten des Tumors bestimmen.

In dieser Arbeit wurde ein gemeinsamer neuroendokriner Substatus identifiziert, der durch eine starke Induktion von Hitzeschockprotein-mRNA (HSP+) gekennzeichnet ist und Signaturen aufweist, die auf eine Aktivierung der "Unfolded Protein Response", Hypoxie, mTORC/PI3K-AKT-Signalisierung und eine glykolytische Verschiebung hinweisen. In einer Patientenprobe zeigte sich ein einzigartiger stromalähnlicher Sub-Status mit angereicherter YAP/TAZ-assoziierter Hippo-Signalisierung neben der Expression mesenchymaler und basaloider Programme, was transkriptomische Ähnlichkeiten mit dem duktalen Pankreas-Adenokarzinom (PDAC) widerspiegelt. Darüber hinaus war einer der gemeinsamen neuroendokrinen Substadien hochgradig proliferativ und zeichnete sich durch überexprimierte *E2F*-Ziele aus, darunter Enhancer of Zeste homolog 2 (*EZH2*). Bemerkenswerterweise zeigte dieser Subtype eine signifikante Anreicherung von *PTF1A*-Zielgenen, die gewebespezifisch im Gehirn exprimiert werden, während die Pankreas-spezifische Zielgene unterdrückt wurden. Diese Beobachtung deutet auf eine Verschiebung der zellulären Identität von einem pankreatischen Abstammungsprogramm in der Ursprungszelle hin zu einem allgemeineren neuronalen Phänotyp hin, und einer damit einhergehenden Anreicherung von Signaturen mit Bezug zu DNA-Reparaturprogrammen, Krebsstammzellen und Chemotherapieresistenz. Möglicherweise könnte dieses gegenüber dem Organgewebe dedifferenzierte neuronale Programm als Achillesferse von panNEC für die Entwicklung therapeutischer Strategien genutzt werden.

Die vorgestellten Daten haben also eine beträchtliche Heterogenität und therapeutische Schwachstellen bei hochgradigen panNEC aufgedeckt, was die Bedeutung der

Tumorprofilierung für personalisierte Behandlungsansätze unterstreicht. Bemerkenswert ist auch die Aussicht auf klinische Interventionen, die auf zwei gemeinsame neuroendokrine Substadien abzielen und die Machbarkeit personalisierter Kombinationstherapien im klinischen Umfeld nahelegen.

Zusammenfassend lässt sich sagen, dass in dieser Dissertation zwei unterschiedliche, aber kritische Systeme auf zellulärer Ebene im Zusammenhang mit der Entstehung und dem Fortschreiten von Krankheiten eingehend untersucht wurden. Der snRNA-seq-Ansatz ermöglichte ein unvoreingenommenes und tiefes Verständnis der Krankheitsbiologie, welches ein erhebliches translationales Potenzial aufweist. Insbesondere wurden in beiden Systemen spezifische Zell- (oder Kern-) Unterzustände dem Krankheitsursprung zugeschrieben, die zu unserem Verständnis der Krankheitsentwicklung auf molekularer Ebene beitrugen und mögliche Wege für die therapeutische Entwicklung aufzeigten. Dieser Übergang von der Konzentration auf ein einzelnes molekulares Ziel zur Behandlung der zugrundeliegenden zellulären Dysfunktion bietet neue Möglichkeiten für die klinische Ausrichtung auf zelluläre Signaturen für komplexe Krankheiten und für die Entwicklung einer neuen Generation von Therapeutika.

Table of contents

1 Introduction

Cells are the most fundamental unit of life on earth, and every multicellular organism critically depends on the intricate interplay of a multitude of cells that contribute to the formation of intricate tissues. The genesis of most multicellular organisms commences from a single cell, involving a series of events encompassing cell division, differentiation, and specialization. Through these processes, cells organize into distinct tissues, and organs assume specific functions.

Despite sharing the same genetic code, these cells manifest remarkably diverse phenotypes. The transcriptome serves as a pivotal mediator orchestrating distinct cell types and states within functional tissues [1]. The foundational concept of cells serving as the fundamental building blocks of organisms traces its roots to seminal works like the 'Cell Theory,' propounded by Rudolf Virchow and his contemporaries during the 19th century [2]. These insights were forged through a culmination of two centuries of scientific exploration, beginning with Robert Hooke's inaugural depiction of a cell in 1665, made possible by Antoni van Leeuwenhoek's pioneering microscope [3]. After these pivotal revelations, investigators have endeavored to categorize cells into distinct types characterized by their unique responses and interactions with their surroundings, thereby influencing formation of tissues and ultimately, entire organisms. The initial taxonomy was grounded in morphological distinctions among cells. However, as our comprehension of the molecular mechanisms governing cellular phenotypes has deepened, the classifications have evolved to encompass the diverse molecular strata comprising DNA, RNA, and proteins.

Cellular organelles, including the nucleus, mitochondria, endoplasmic reticulum, and others, play specialized roles in maintaining the structural and functional integrity of cells. Among these organelles, the nucleus holds particular significance within the context of this dissertation. It houses genetic material in the form of DNA, enclosed by a nuclear membrane with pores that regulate molecular movement between the nucleus and cytoplasm. A standard cell typically contains about 6 pg of DNA, 50,000-300,000 mRNA molecules (5-30 pg), and millions of proteins (20-200 pg) [4-6]. The nucleus assumes a crucial role in governing gene expression and safeguarding the integrity of DNA.

1.1 Central Dogma of life – from DNA to protein

Biological systems inherently exhibit a natural flow of information. Within cells, this information flow entails the transfer of genetic information from DNA to RNA and, ultimately, to proteins **(Figure 1.1).** In the initial step, transcription occurs within the nucleus, where a segment of DNA is transcribed into RNA [7-9]. The resulting messenger RNA molecule (mRNA) is then transported out of the nucleus to a ribosome, where the mRNA is decoded in the translation process, giving rise to proteins that perform specific functions within or outside the cell. The regulation of this information flow is a sophisticated and tightly controlled process crucial for cellular functionality. Understanding the underlying mechanisms is essential for advancing our comprehension of molecular cell biology and developing new therapeutic strategies.

1.1.1 Decoding life's blueprint – DNA and the genome

The genome serves as the intricate blueprint that choreographs the complexity of life within organisms, encapsulating the complete set of genetic instructions essential for their growth, development, and functioning. At the heart of this genetic blueprint lies the remarkable interplay of genes, DNA, and the fundamental process of replication.

Human genetic information is comprised of DNA molecules organized into condensed structures known as chromosomes. Somatic cells, found in most tissues, possess two sets of chromosomes totalling 46 (44 autosomal and two sex chromosomes), while cells in the human reproductive system carry only half, or 23 chromosomes (22 autosomal and one sex chromosome), facilitating their fusion during fertilization to generate a new organism with a distinct combination of genetic material from both parents [10-12]. The DNA molecule exhibits a double helix structure, consisting of pairs of four bases- cytosine (C), guanine (G), adenine (A), and thymine (T). Each base is linked to a 2-deoxyribose sugar molecule and three phosphate groups, forming the basic building block of DNA, a nucleotide. Adhering to specific base-pairing rules, A pairs with T, and C pairs with G, establishing hydrogen bonds that maintain the structural integrity of the DNA. The sequence of these bases along the DNA strands forms the genetic code, akin to the language that directs the synthesis of proteins.

Replication, a fundamental biological process, ensures the faithful transmission of genetic information from one generation of cells to the next. Replication involves the unwinding of the DNA double helix, followed by the synthesis of complementary strands facilitated by the DNA polymerase enzyme [13, 14]. This meticulous process guarantees that each daughter cell

inherits a complete and accurate copy of the genetic material, maintaining the continuity and integrity of the genomic information.

1.1.2 RNA Symphony– gene expression dynamics and the transcriptome

Ribonucleic Acid (RNA) stands as a versatile and dynamic macromolecule crucial for various cellular functions. Derived from DNA through the process of transcription, RNA comes in different forms, each serving specific roles within the cell. Contrasting with DNA, the nucleotides found in RNA molecules contain ribose sugar (as opposed to 2-deoxyribose), three phosphate groups, and feature uracil (U) in lieu of thymine (T). RNA can be categorized into two major groups: coding and non-coding [15, 16]. Transfer RNA (tRNA) assists in assembling amino acids during protein synthesis, and ribosomal RNA (rRNA) forms an integral part of the cellular machinery responsible for protein production. Both tRNA and rRNA represent non-coding RNA. Given that the research presented in this thesis centres on analysing expression of protein-coding genes, the subsequent section will exclusively concentrate on coding RNA.

Messenger RNA (mRNA) [15] carries the genetic instructions from DNA to the ribosomes, where proteins are synthesized. mRNA synthesis is accomplished through the process of transcription that begins with an initiation phase, where an enzyme called DNA-dependent RNA polymerase binds to a specific region of DNA known as the promoter. The promoter region acts as a signal for the start of transcription and provides a binding site for RNA polymerase. Once bound to the promoter, RNA polymerase unwinds the DNA double helix, exposing a segment of the DNA template. This exposed DNA strand serves as the template for the synthesis of a complementary RNA molecule. As RNA polymerase moves along the template, it catalyzes the addition of nucleotides to the growing RNA chain, following the base-pairing rules: adenine (A) pairs with uracil (U), guanine (G) pairs with cytosine (C). The elongation phase continues until RNA polymerase reaches a termination signal in the DNA sequence. At this point, the newly synthesized RNA molecule is released, and RNA polymerase detaches from the DNA template. The transcribed RNA molecule, often referred to as pre-mRNA, undergoes further post-transcriptional processing steps, including splicing, 5’ end capping and 3’ end polyadenylation. Of note, splicing indicates the removal of non-coding regions (introns) and the joining together of coding regions (exons). On the other hand, a modified guanine nucleotide is added at the 5' end of the pre-mature mRNA. This cap structure not only protects the RNA from degradation but also facilitates its transport out of the nucleus and aids in translation initiation during protein synthesis. At the 3' end of the pre-mature mRNA, a poly-A tail composed of adenine nucleotides is added. This

polyadenylation contributes to mRNA stability, enhances mRNA export from the nucleus, and influences translation efficiency. Hence, these post-transcriptional modifications contribute to the complexity and diversity of the transcriptome, allowing cells to fine-tune gene expression in response to internal and external signals. Subsequently, the final mature mRNA is then transported out of the nucleus and into the cytoplasm, where it serves as a template for protein synthesis during the process of translation [17].

The transcriptome, in essence, represents the complete set of RNA molecules present in a cell at any given moment. It provides a dynamic snapshot of the cellular activity, reflecting not only the genes expressed but also the complex regulatory networks underlying cellular functions. Understanding the transcriptome is akin to deciphering the symphony of genetic expression that underlies the diverse cellular phenotypes observed in different physiological states or environmental conditions.

However, in 1970, the concept of reverse transcription, a reversal of the flow of genetic information, was first elucidated in viruses [18]. This phenomenon involves the conversion of RNA into a DNA molecule and is facilitated by the enzyme RNA-dependent DNA polymerase, commonly known as reverse transcriptase. This enzymatic process enables the virus to integrate its genetic material into the host cell's genome, effectively co-opting the cellular machinery to synthesize viral proteins and assemble new viral particles. The revelation of reverse transcription marked a groundbreaking moment in molecular biology, giving rise to the ability to generate complementary DNA (cDNA) from RNA molecules [18]. This technique has since become a pivotal tool in research, particularly for quantifying mRNA levels in cells and tissue samples to investigate gene expression. Additionally, it has found widespread utility in various applications such as cloning and vaccine development.

1.1.3 Probing the proteome – decoding protein function and diversity

Proteins are complex macromolecules essential for the structure, function, and regulation of cells. Composed of amino acid chains folded into unique three-dimensional structures, proteins carry out diverse biological functions. They serve as enzymes catalyzing biochemical reactions, structural components providing cellular support, transporters facilitating the movement of molecules, and signalling molecules regulating various cellular processes.

Translation [17] is a central process in molecular biology where the genetic information encoded in mRNA (messenger RNA) is used to synthesize proteins. It is achieved with the

help of ribosomes, which read the mRNA codons and recruit corresponding amino acids. A complex interplay of transfer RNA (tRNA) molecules ensures the accurate alignment of amino acids, forming a polypeptide chain. As the ribosome moves along the mRNA, the amino acids are sequentially added, ultimately resulting in the synthesis of a functional protein. In the translation process, proteins are constructed using a set of up to 20 amino acids. The specific order of these amino acids is dictated by the sequence of nucleotides in mRNA molecules. Each group of three consecutive nucleotides forms a codon, which serves as either the starting or stopping point for translation, or codes for a single amino acid.

The proteome [19] encompasses the entire set of proteins expressed by an organism, tissue, or cell at a specific time under defined conditions. The proteome is influenced by factors such as genetics, environment, and cell state, making it a dynamic and informative reflection of the cellular landscape [20, 21].

1.2 Cellular phenotypes – What is a cell type/state?

A cellular phenotype refers to a cell's observable characteristics or traits resulting from the interactions between its genetic makeup and environment [1, 22]. These characteristics include a cell's shape, size, behavior, and function. The phenotype of a cell is determined by its genotype (the genetic information encoded in its DNA) and its environment (which can include factors such as nutrients, chemical signals, temperature, and physical forces). Changes in either the genotype or the environment can alter the phenotype of a cell [24]. Hence, the cell phenotype can provide important information about its health, developmental stage, and function [23-25]. For example, abnormal cellular phenotypes are often associated with several metabolic, hematology, oncology, and developmental diseases [26-32]. Investigating the changes in how cells behave can provide valuable insights into an organism's underlying biology and disease mechanisms, and hence can help identify markers for diseases and develop treatment strategies.

Our cells can be classified into different types based on structure, function, and molecular markers. For example, humans have various types of cells, such as neurons, muscle cells, skin cells, blood cells, and many more, each with unique structures and functions. In contrast, cell states refer to the different states or stages that a cell can be in, which can be influenced by various factors such as environmental conditions, developmental stage, and disease state. For instance, differentiation during embryonic development starts with pluripotent stem cells (PSCs), which exhibit the capability to differentiate into all cell types within an organism. These cells possess the dual capacity to self-renew and generate daughter cells that retain

their original state, and under specific conditions, can undergo differentiation into specialized cell types. A mature cell may be in a more specialized and differentiated state with limited potential change. Understanding cell types and states is critical to understanding how the body works and how diseases develop, as well as developing therapies and treatments for various diseases [33-35].

Traditionally, the classification of phenotypes at the individual cell level has predominantly concentrated on protein identification through microscopy techniques, such as immunofluorescence light microscopy. Over time, with continuous advancements in these methodologies, the classification of cells based on imaging has evolved to include the quantitative characterization of molecular features. This expanded scope encompasses the assessment of RNA molecules through techniques like fluorescence in-situ hybridization (RNA-FISH) and the examination of structural attributes of DNA using diverse fluorescent dyes [36-38]. Nevertheless, it is important to note that these approaches necessitate a prerequisite understanding of the target, such as knowledge of protein structure for epitope detection via antibodies or awareness of RNA molecules for hybridization to a consensus sequence.

1.2.1 Analysis of gene expression with next-generation sequencing

The initial stride toward an impartial approach for exploring the molecular makeup of cells began with microarray technology. Microarrays operate on the principle of hybridization between labeled target nucleic acids (e.g., DNA or RNA) from a sample and complementary probe immobilized on a solid support [39]. The detection of hybridized targets allows the simultaneous analysis of gene expression levels, genetic variations, or other genomic features across thousands of genes. While this enabled profiling of many genes, the first unbiased technique materialized with the capability to reconstruct the sequence of an unidentified DNA molecule [40].

Presently, a diverse array of technologies, collectively termed next-generation sequencing (NGS), empowers the precise determination of the sequence of any given DNA or RNA molecule [41-43]. The Human Genome Project (HGP) played a pivotal role in revolutionizing NGS technology by setting the stage for large-scale sequencing endeavors [44]. The HGP, completed in 2003, provided a foundational reference genome, stimulating advancements in sequencing techniques, reducing costs, and enhancing the accuracy and throughput of NGS. This synergy between the Human Genome Project and NGS has facilitated groundbreaking discoveries in genomics and propelled the field of personalized medicine. While the HGP

incurred a staggering cost of up to $3 billion, the advent of new sequencing technologies has revolutionized the landscape, allowing for the complete sequencing of a human genome within a few days at a mere cost of a few thousand US dollars. This dramatic reduction in both time and expense renders it practical to address scientific inquiries swiftly and comprehensively on a large scale [44].

NGS is a high-throughput DNA sequencing technology that allows for the rapid and parallel sequencing of millions of DNA fragments [41, 45]. Illumina, being a marker leader revolutionized NGS, particularly in terms of throughput and cost-effectiveness, making it highly accessible and one of the most widely utilized platforms [46]. Their method employs sequencing-by-synthesis on a solid surface, involving repetitive cycling of fluorescently labelled nucleotides that integrate into growing DNA strands. At the end of each cycle, a high-resolution image is captured to determine the sequences of immobilized molecules. This iterative process is replicated numerous times, producing millions to billions reads per run (depending on the system used) that are subsequently assembled into a complete sequence [47]. This transformative approach has significantly increased the speed and efficiency of DNA sequencing, alongside a simultaneous reduction in the cost per sequenced base. This shift facilitated comprehensive genomic analysis and development of bench-top technologies for both research and diagnostic applications [48, 49].

1.2.2 RNA-sequencing – from bulk to single cell

While proteins traditionally serve as the ultimate effectors of cellular functions, RNA has emerged as a valuable tool for studying cellular phenotypes. The choice of RNA over proteins is grounded in the dynamic nature of the transcriptome. Since RNA levels are more dynamic and responsive to cellular changes than proteins, RNA analyses provide a real-time assessment of cellular responses, offering a more nuanced and temporally precise perspective compared to protein-based approaches. The higher abundance of RNA relative to proteins allows for the detection of a broader range of genes and their expression levels. Additionally, studying RNA allows researchers to gain insights into post-transcriptional modifications, alternative splicing events, and non-coding RNA molecules, offering a more comprehensive understanding of cellular dynamics. Furthermore, the use of RNA in cellular studies often overcomes challenges associated with the stability and turnover rates of proteins.

The advent of reverse transcription and NGS has broadened the spectrum of possibilities for advancing novel approaches in studying gene expression. Since then, bulk RNA sequencing

(bulk RNA-seq) has emerged and expanded its use in both basic research and clinical applications [42, 43, 50-52]. In this methodology, total RNA is extracted from a cell population or tissue, converted into cDNA, and subsequently sequenced on NGS platforms. Consequently, this yields aggregated data from a large number of cells, encompassing various distinct cell types. Bulk RNA-seq boasts several advantages, such as the detection of low-abundant transcripts, identification of novel transcripts, and are relatively high throughput.

However, it presents a limitation by offering only an averaged snapshot of gene expression for the entire sample, thereby masking variations in gene expression levels among individual cells [53]. Even when individual cell types were isolated through sorting techniques, both bulk RNA-seq and microarray analyses provide averaged measurements across a population of cells. As a result, these techniques most likely overlook cellular heterogeneity, missing variations in gene expression that exist among individual cells within a sample [54, 55].

The transcriptomic signature of individual cells mirrors their respective states and is influenced by both systematic and local stimuli. The growing fascination with exploring cell diversity and heterogeneity has propelled advancements in technology, particularly in the realm of single-cell RNA sequencing (scRNA-seq) [56-59]. Over the last couple of years, scRNA-seq technologies have emerged to address the limitations of bulk RNA-seq by providing a more granular and precise understanding of gene expression at the individual cell level [60, 61]. Even cells that belong to the same cell type and reside in the same tissue can exhibit variable gene expression owing to cellular microenvironment or stochastic fluctuations in gene regulation [60-64]. By identifying the sources of variability within a cell population, researchers can better understand how cells respond to their environment and maintain their identity and function.

While scRNA-seq emerged as the current state-of-the-art technique for transcriptome profiling, it presents several drawbacks when compared to bulk RNA-seq. The higher cost of scRNA-seq, stemming from specialized reagents and the requirement for deeper sequencing, can be a limiting factor for large-scale studies. Technical variability is amplified in scRNA-seq due to the need for single-cell isolation protocols and the associated challenges of low RNA input, leading to increased technical noise [65], dropout events and high data sparsity [66, 67] that are detailed in subsequent sections. The limited range and transcript coverage in scRNA-seq may affect the accuracy of quantitative measurements, especially for low-abundance transcripts. Despite its revolutionary impact, careful consideration of these downsides is crucial for researchers selecting the most suitable sequencing approach for their specific biological questions.

1.3 Single cell or nucleus RNA sequencing

Single-cell RNA-seq [56-59] is a technique that allows the measurement of gene expression at a cellular resolution. It involves isolating and capturing the RNA content of individual cells, followed by cDNA synthesis, library preparation, and high-throughput sequencing. Presently, widely adopted scRNA-seq techniques typically involve hybridizing the polyA tails of mRNA molecules to barcoded polyT primers, prior to library preparation and sequencing [68].

On the other hand, single nucleus RNA sequencing (snRNA-seq) is a technique that allows the measurement of gene expression from the nuclei of individual cells [69]. snRNA-seq involves isolating and capturing individual nuclei, amplifying, and sequencing their transcriptome. This technique is particularly useful for analyzing samples that are difficult to isolate, or cells that have undergone significant degradation, exhibit high ribonuclease activity such as pancreatic acinar cells and for fresh-frozen biobanked samples [70-73]. In addition, the size of nuclei varies less than cells, therefore snRNAseq cell proportions are more representative than that of scRNAseq. Although snRNA-seq can function as an alternative to scRNA-seq, it is important to recognize that the former selectively captures mRNA within the nucleus. Exclusion of cytoplasmic mRNA can result in a limited ability to capture certain biological processes [74].

Both techniques provide notable insights into cellular heterogeneity and diversity, allowing the identification of known and rare cell types/states and characterization of gene expression patterns at a cellular level. This can also help classify cells into groups based on their functional properties and developmental origins, providing insights into how cells differentiate and specialize over time [75, 76]. Overall, both techniques are powerful for studying cellular variability and can help uncover the underlying mechanisms that drive cellular phenotypes in health and disease. The choice between these techniques depends on the biological question being addressed, the type of tissue or cells being analyzed, and the experimental constraints.

1.3.1 Brief history of single cell/nuclei RNA-seq technology

In recent years, single cell transcriptomics has played a pivotal role in advancing precision medicine, offering unprecedented insights into the molecular and genetic heterogeneity at the individual cell level. The previous limitations of NGS for single-cell sequencing primarily revolved around the challenges associated with the amount of material required and the absence of technologies for capturing and amplifying nucleic acids at single-cell resolution.

That former means the material requirements of NGS technologies far surpassed the typical quantities of DNA or RNA present in a single cell.

The first scRNA-seq technology was developed in 2009 by Tang et al. [57], who described a single-cell digital gene expression profiling assay that enabled the amplification and sequencing of full-length mRNA transcripts from individual cells. However, it was limited in throughput and could only sequence several cells simultaneously. Among the initial techniques for single-cell analysis, Single-Cell Tagged Reverse Transcription Sequencing (STRT-seq) developed by Islam et. al. [77] stands out as a method that relies on 5' tag counting of transcripts. When conducted on the microfluidics Fluidigm C1 platform, STRT-seq [78] emerged as a versatile scRNA-seq, offering precise, sensitive, and crucially, molecular counting of transcripts at the individual cell level. In 2015, Macosko et al. [79] developed the Drop-seq method that leveraged microfluidic droplets to isolate and barcode individual cells for high-throughput sequencing that allowed the analysis of thousands of cells in a single experiment and enabled the identification of known and rare cell types and subpopulations. Since then, various other scRNA-seq technologies have been developed, including SMART-Seq [58], SMART-Seq2 [80, 81], CEL-Seq [82], CEL-Seq2 [83], inDrop [84] and MARS-Seq [85]. Importantly, SMART-Seq (Switching Mechanism At 5' end of RNA Template) and its successor SMART-Seq2 protocols were developed to capture and amplify full-length transcripts from individual cells. SMART-Seq was initially introduced by Ramskold et. al. 2012 [58]. This method utilized template-switching reverse transcription to capture the entire transcript, providing a comprehensive view of the transcriptome from single cells. SMART-Seq2 incorporated improvements such as reduced amplification bias, increased sensitivity, and lower input requirements, making it a powerful tool for profiling the transcriptomes of individual cells with high resolution. Taken together, these technologies have improved sensitivity, accuracy, and throughput and have enabled the analysis of large numbers of cells at a lower cost [86-88].

Of note, 10X Genomics, a biotechnology company, introduced its GemCode technology [89] in 2015, marking a milestone in single-cell genomics. GemCode utilized a microfluidic system to encapsulate single cells with unique barcoded gel beads, each containing a unique molecular identifier (UMI). Subsequently, 10X Genomics released the Chromium scRNA-seq Solution, an upgraded version of GemCode, providing a robust platform for high-throughput single-cell transcriptomics that enabled the simultaneous analysis of thousands to millions of cells. The Chromium system utilizes barcoded gel beads to uniquely label RNA molecules within each cell. This barcoding strategy enables the distinction of transcripts originating from the same cell, overcoming challenges related to amplification biases. Importantly, Cell

hashing [90] is a technique that involves labeling cells with unique oligonucleotide barcodes during sample preparation. This allows for the multiplexing of samples, enabling the simultaneous analysis of multiple samples in a single experiment. To increase the scalability and cost-effectiveness, 10X Genomics introduced Cell Hashing, allowing multiplexing of samples, by using oligo-tagged antibodies to uniquely label cells before pooling them, facilitating the analysis of multiple samples in a single experiment.

The 10X single cell or nuclei RNA-seq platform found widespread applications in various biological fields, including immunology, hematology, oncology, and neuroscience [91-93].

1.3.2 10X single nuclei isolation, library construction and sequencing

The sequential steps of a typical 10X snRNA-seq experiment include the nuclei isolation, GEM generation and barcoding, cDNA QC and quantification, 3’ gene expression library construction and sequencing followed by CellRanger mapping and downstream analysis [72-74].

Initially, nuclei isolation is a critical step, achieved through specialized buffers and enzymatic or mechanical dissociation techniques. The isolated nuclei are then individually encapsulated into droplets or wells using microfluidic devices, ensuring the preservation of cellular heterogeneity within complex tissues. During encapsulation, each poly-A tailed RNA molecule is captured and labelled with unique barcodes, and subsequently are reverse transcribed to yield cDNA. The cDNA is then processed and amplified using the 10X Genomics Single Cell 3' Solution protocol, which includes fragmentation, tagging, and amplification of the cDNA. Barcoding is a crucial step in 10X Genomics-based protocols. Here, unique barcode sequences are introduced during PCR amplification to enable the pooling of samples for parallel sequencing while still distinguishing between individual nuclei. For optimal experimental results, it is crucial to utilize high-quality tissue samples.

Quality control steps may include checking the size distribution of the library fragments [94, 95]. In this context RNA integrity is typically assessed by the RNA Integrity Number (RIN) [96]. Typically, an RIN score exceeding 8 is deemed satisfactory for advancing to subsequent downstream experiments.

For single nuclei experiments, it is imperative to generate pre-mRNA reference files [97] in order to distinguish between reads that align to exonic (mature mRNA) and intronic regions, aiding in the quantification of transcriptional activity within the cell.

1.3.3 Technical differences between scRNA and snRNA-seq

Single-nucleus RNA-seq is specifically designed to analyze gene expression at a cellular resolution while circumventing the challenges associated with traditional scRNA-seq methodologies. For instance, cells in certain tissues like skeletal muscle, adipose tissue or pancreas are difficult to dissociate or isolate without causing major damage to their integrity [70-74]. In another case, some cell types are challenging to isolate intact, and the isolation process may lead to biased representation or loss of specific cell populations in a scRNA-seq [73, 74]. Also, snRNA-seq is useful for analyzing both fresh and frozen (or archived) tissues-so, this eliminates the requirement of immediate sample processing and helps reduce any changes caused by the stress response to enzymatic treatment [70-74]. Moreover, earlier investigations indicate that snRNA-seq consistently demonstrated satisfactory sensitivity – a measure of the identified UMIs or genes counts per cell in datasets sampled with an equal number of reads per cell [74]

Given the loss of cytoplasm, nuclei typically exhibit lower mRNA levels compared to whole cells, resulting in a reduced total number of detected genes and unique molecular identifiers (UMIs) per cell [74]. Ultimately, the variations in sensitivity between single-nucleus and single-cell RNA sequencing depend largely on the sample specific characteristics. Generally, the quality control (QC) cut-off(s) can be decided by investigating the relationship among number of UMI, genes and %MT-transcripts per nuclei [98, 99].

In a snRNA-seq experiment, the cytoplasmic proteins released during the extraction process includes RNAses that could potentially destroy the transcripts after encapsulation and lysis of the nuclei. All procedural steps are required to be conducted in a cold environment to prevent unwanted artefactual transcriptional changes, taking advantage of the stabilizing effect of the cool milieu.

1.4 snRNA-seq data analysis

The computational analysis of both single cell or nuclei RNA-seq data involves several steps: read alignment, gene expression quantification, QC, normalization, dimensionality reduction, and identification of differentially expressed genes to identify cell type or states [98, 99]. There are several challenges in the downstream single cell or nuclei RNA-seq analysis, starting from quality control and data harmonization to ensure appropriate data interpretation. A few major points are summarized below:

1.4.1 Handling technical noise and quality control

snRNA-seq data can be noisy due to various technical factors such as cellular barcoding and indexing errors, amplification bias, sequencing depth, dropout events, and other batch effects. These technical factors [65-67] can introduce unavoidable variability in downstream data analyses, such as cell type identification and differential gene expression analysis. Quality control measures, such as assessing the percentage of reads with valid barcodes, can help identify and filter out cells affected by barcoding errors. On the other hand, using unique molecular identifiers (UMIs) to tag individual RNA molecules, and optimizing PCR conditions can help reduce amplification-related noise.

Also, contamination from ambient RNA or from adjacent cells during tissue dissociation can introduce extraneous RNA- often, leading to misinterpretation of cell type composition and gene expression profiles [100, 101]. To computationally mitigate these artifacts, state-of-the-art tools like Cellbender employ a statistical approach to distinguish ambient RNA from true cellular RNA and correct for this contamination [101]. Also, single cell/nuclei profiling is prone to dropout events where certain transcripts fail to be detected due to technical limitations, leading to false negatives in gene expression analysis [66, 67, 102].

Technical doublets in snRNA-seq data refer to instances where two or more nuclei are mistakenly captured and sequenced as a single unit [103-105]. These artifacts can arise during the nuclei isolation process and lead to misinterpretation of the true biological heterogeneity. Tools like Scrublet [103] specifically for the identification and removal of technical doublets. Scrublet operates by creating a nearest-neighbor classifier for each nucleus based on its transcriptomic similarity to other nuclei in the dataset. It then assesses the observed and expected rates of doublets, assigning a doublet score to each nucleus. Nuclei with higher doublet scores are considered more likely to be technical doublets. Researchers can use these scores to confidently flag and if required, exclude potential doublets from downstream analyses [103-105]. Nonetheless, there is a lack of gold standard tools for flagging biological doublets that can be optimally labelled by investigating expression of conflicting cell type/state markers and is detailed in sub-section 1.4.3.

In addition, single nuclei isolation can be challenging, and some nuclei could be damaged or incomplete, resulting in incomplete transcripts and lower sequencing quality [106]. Often, such nuclei exhibit relatively low UMI counts but a high percentage mitochondrial transcripts per nucleus- possibly, derived from the ‘soup’ of the single-nuclear isolation preparation. Hence, both the number of UMI and %mitochondrial transcripts per nucleus should be

modelled as technical covariates in downstream analysis. Also, some nuclei have a high number of UMI(s) but only a few numbers of genes that could be explained by nuclei undergoing apoptosis or necrosis. Such populations can represent low complexity cell types. Hence, the number of UMI or genes per nuclei alone might not be a reliable proxy of a good sample quality in isolation. In this regard, it is recommended to investigate sample complexity that is defined as the number of genes detected per UMI [107]. For example, a higher number of genes detected per UMI indicates higher complexity or high sample quality. Usually, a score > 0.8 suggests a good sample quality [107].

snRNA-seq data normalization is a critical step in data analysis to correct differences in sequencing depth and technical factors that can bias gene expression measurements between cells or nuclei [98, 99, 108]. However, selecting the appropriate normalization method can be challenging as different methods may perform better in different scenarios.

1.4.2 Challenges in snRNA-seq data harmonization

Harmonizing snRNA-seq data from different experiments or platforms is challenging due to several technical variations underlying the data. For instance, processing samples in different batches or at different sequencing runs can induce sample specific effects that potentially introduce technical variation unrelated to biological differences. Additional complexities might arise from variability in tissue dissociation protocols, nuclei isolation methods, and library preparation techniques- hence, modelling appropriate covariates is required depending on the experimental set-up. Hence, to ensure robust downstream analyses, appropriate batch effect removal and data harmonization methodologies are required to optimally align data stemming from different sources [99, 109-116]. Moreover, to achieve effective data harmonization, one has to struggle with the trade-off between two scenarios — (1) properly mixing datasets to prevent separation by biological or technical replicates and (2) retaining the original data structure, as illustrated in **Figure 1.1.**

Several algorithms require dimensional reduction methods before data integration which is described below.

- **Linear and non-linear dimension reduction techniques**

Linear and non-linear dimension reduction techniques are commonly used in snRNA-seq analysis for cellular phenotyping. These techniques reduce the high-dimensional gene expression data into lower-dimensional representations that can be analyzed and visualized.

Linear dimension reduction techniques, such as principal component analysis (PCA), independent component analysis (ICA), and factor analysis (FA) are based on linear transformations of the data. PCA [117, 118] is a general-purpose linear dimensionality reduction technique that aims to capture the maximum variance in the data by projecting it onto a new set of orthogonal axes, called principal components. PCA, in its traditional form, is not explicitly designed for batch correction and generally used for capturing variance within a single dataset. In contrast to many other methods, principal components (PC) are easier to select for and interpret as they are ranked by variance i.e., the first PC accounts for the most variance, followed by the second PC (orthogonal to the first) and so on [117, 118]. In contrast, Independent Components (IC) [119, 120] and Factors [121] are not naturally ordered like PC(s) that complicates their interpretation without further analysis.

Several software tools leverage linear dimension reduction strategies for the snRNA-seq data harmonization. A popular tool, Harmony [110] is grounded on adjusting PC(s) to effectively reduce batch effect during data harmonization. In the transformed PC space, Harmony employs k-means clustering to delineate clusters having maximum diversity. The diversity of each cluster is evaluated based on whether it includes balanced proportions of cells from each batch (such as donor or condition) intended for integration. Once diverse clusters are defined, Harmony assesses the impact of a cell’s batch identity on its PC coordinates and applies a correction to 'shift' the cell toward the centroid of its respective cluster. Subsequently, cells are projected again using these corrected PCs, and the iterative repetition of this process continues until convergence is achieved [110].

On the other hand, BBKNN (Batch Balanced k-Nearest Neighbors) [111] aims to adjust the shared neighbor graph of the cells. Firstly, the algorithm identifies k-nearest neighbors for each cell (or, nucleus), and iteratively correcting for batch effects by applying correction vectors based on the differences in expression profiles between cells and their neighbors across different batches. This approach ensures a balanced representation of cells from various batches, ultimately harmonizing the data and facilitating effective integration for downstream analyses such as clustering and visualization. Moreover, it is technically feasible to employ a sequential integration approach, firstly adjusting the principal component(s) with Harmony, and subsequently using the corrected principal component(s) to build an adjusted k-nearest neighbors (KNN) graph through BBKNN.

Furthermore, Seurat utilizes another linear dimension reduction method, called Canonical Correlation Analysis (CCA) [116] that is designed to uncover linear relationships between two or more sets of variables. It aims to find linear combinations of variables (canonical variables)

in each set such that their correlation is maximized. CCA is commonly used in multivariate statistics and data integration tasks where there are two sets of related variables [116]. In biology, this means integrating datasets having common anchors i.e., common cell types/states. Also, the Reciprocal PCA [122] approach implemented in Seurat involves computing the PCs separately for each batch and then adjusting the PCs to minimize differences between batches, providing a corrected representation of the data that is more amenable to integration and downstream analyses. The reciprocal nature of the approach ensures mutual correction between batches, enhancing the harmonization of diverse datasets in a reciprocal manner.

In contrast, non-linear dimension reduction techniques, such as t-distributed stochastic neighbor embedding (t-SNE) [123, 124] and uniform manifold approximation and projection (UMAP) [125, 126], are based on non-linear transformations of the data. These techniques aim to preserve the local structure of the data while reducing the dimensionality. t-SNE models the high-dimensional data as a probability distribution of pairwise similarities and maps it to a low-dimensional space. UMAP is a relatively newer technique developed by McInnes et. al 2018 [125] that also preserves the local structure of the data but uses a different optimization approach compared to t-SNE. However, t-SNE and UMAP are widely used for data visualization rather than dimension reduction for sc/sn RNA-seq datasets [124, 126].

Hence, various tools for data harmonization have been developed to address batch effects, each with its specific strengths and limitations. However, recent methodological progress has propelled the adoption of deep-learning frameworks because of their previously demonstrated efficacy in handling complex integration tasks.

- **Deep learning techniques**

Deep learning algorithms based on artificial neural networks (ANNs) can be leveraged for various tasks in single-cell analysis, such as cell type classification, gene expression imputation, and data integration. These algorithms can learn patterns and relationships from the data without explicit feature engineering or manual annotation, making them highly adaptable and flexible [127-129]. For example, convolutional neural networks (CNNs) have been used for image-based analyses of spatial transcriptomics data [130]. In contrast, recurrent neural networks (RNNs) have been used for time-series analyses [131, 132].

Deep learning-based data integration tools have gained prominence in the analysis of single-cell or nuclei RNA sequencing data due to their ability to capture complex non-linear relationships and extract high-level features from large-scale datasets [112-115]. They often provide a more nuanced representation of biological variability compared to linear methods like PCA. Two popular methods for deep learning-based data integration are autoencoder-based approaches and Deep Generative Models, including Variational autoencoders (VAE) [133, 134].

Autoencoders can be employed to integrate sc/sn RNA-seq data from multiple conditions or technologies. By training on diverse datasets, the autoencoder learns a low-dimensional representation capturing the shared features and variability across datasets. VAE is a type of deep generative model that incorporates a probabilistic framework, wherein they map input data to a probability distribution within the latent space [112, 113, 135-138]. VAE(s) are trained to generate a latent space that follows a specific probability distribution and allows more nuanced uncertainty modelling. Several recent studies [139-141] have delved into the challenge of incorporating bias and uncertainty into the modeling of single-cell data. A prevalent approach in these studies involves treating each data point (representing cell-gene pairs) as a random variable and fitting a probabilistic model accordingly.

Deep generative techniques, such as scVI (single-cell Variational Inference) [112], extend the VAE framework to sc/sn RNA-seq data. scVI utilizes variational inference to estimate gene expression variability across cells, reducing noise in sc/sn RNA-seq data, and captures both the global structure of the data and the variability specific to individual cells, facilitating data integration and batch correction. scVI adopts a probabilistic approach and models the underlying gene expression distribution using a Zero-inflated negative binomial model (ZINB) that enhances its ability to capture the variability present in sc/sn RNA-seq data compared to traditional methods. It is known that nuclei have lower RNA content compared to whole cells, leading to increased dropout rates and decreased sensitivity in detecting low-abundance transcripts, potentially influencing downstream analyses in snRNA-seq. Hence, scVI's probabilistic modelling coupled with its ability to impute missing data seem to reliably account for the uncertainty associated with low expression levels and high dropouts in individual nuclei in snRNA-seq data [112].

On a similar note, single-cell ANnotation using Variational Inference (scANVI) [113] can be applied on semi-annotated data or label transfer from reference to perturbed/disease datasets. Both scVI/scANVI frameworks [112, 113] allow encoding different categorical and continuous covariates besides the key batch to remove technical artifacts from downstream

analysis effectively. Notably, scVI/scANVI were top performing tools for complex data harmonization tasks that balanced both batch effect removal and preserving biological heterogeneity in an extensive benchmarking study by Luecken et. al 2021 [114]. Furthermore, in addition to scVI/scANVI, numerous other studies have shown the effectiveness of ANN(s) and VAE(s) in integrating scRNA-seq datasets in a scalable fashion [135-138].

Hence, by applying generative models like scVI/scANVI, researchers can gain new insights into the complexity of cellular heterogeneity and the underlying biological processes. These techniques provide relatively more efficient ways to analyze large-scale scRNA-seq datasets, enabling researchers to identify new cell types, understand cell states and transitions, and unravel the mechanisms driving disease and development.

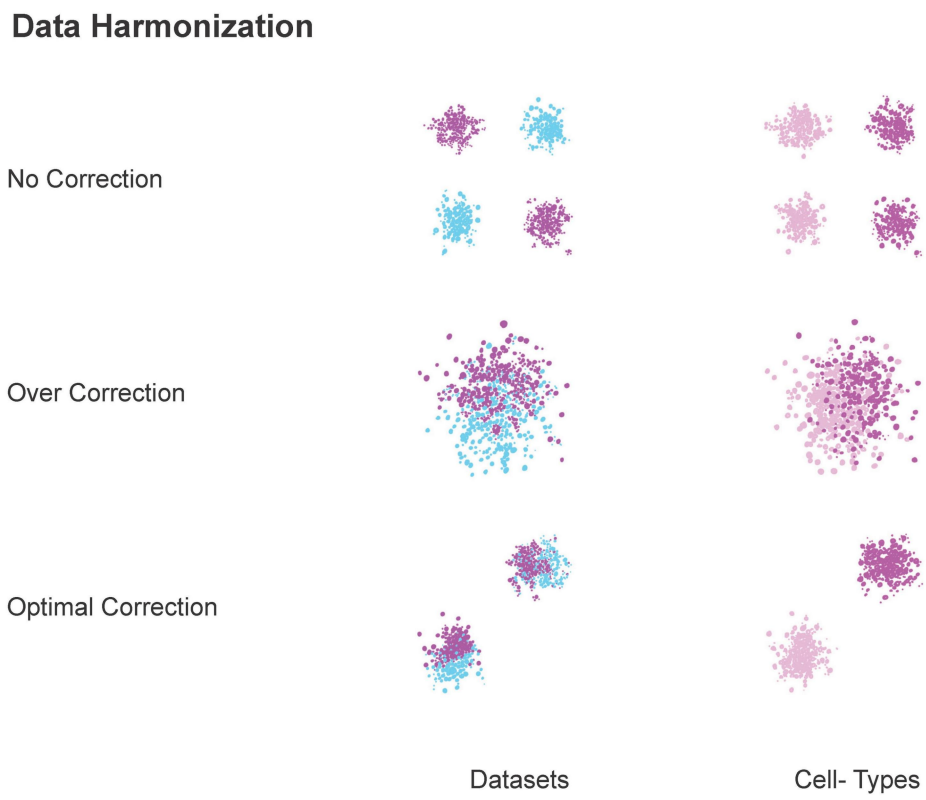


Figure 1.1: Challenges of single cell or single nuclei RNA-seq data harmonization.
Schematic illustrating uncorrected, over-corrected and optimally batch corrected data harmonization. A challenge in performing optimal data harmonization comes with achieving two goals- (1) mixing datasets properly to avoid separation by biological/technical replicates and (2) retaining the original structure of datasets. In the first scenario (No Correction), the datasets are not mixed and hence, strongly separated into batches. In the second scenario (Over Correction), the first goal is improved, but the second goal becomes worse. The third scenario (Optimal Correction) reflects balancing two goals to achieve adequate data harmonization needed for downstream analysis.

1.4.3 Cell type/state identification

After adequate batch effect removal and data harmonization, the task is to identify cell type or cell states, which is typically done through clustering gene expression profiles obtained from single cells/nuclei. Louvain [142] and Leiden [143] clustering are community detection algorithms commonly applied in snRNA-seq data analysis to identify distinct cell populations or clusters. The Louvain algorithm, based on modularity optimization, iteratively refines partitions of cells into communities to maximize the density of connections within communities compared to between them [142]. Leiden clustering, an extension of the Louvain method, introduces a resolution parameter that modulates the size and granularity of identified clusters, allowing for a more flexible adjustment of cluster boundaries [143]. Both algorithms are widely used for their efficiency and effectiveness in uncovering meaningful cell populations in heterogeneous datasets. On a technical note, they both are relatively robust to noise and outlier than traditional methods like K-means clustering that can improve cell-type or -state identification.

It is also important to evaluate cluster quality using metrics like adjusted silhouette score or adjusted Rand index [114, 115] before finalizing annotation. Additionally, visually inspect clustering results to ensure biological relevance and coherence. If large clusters contain multiple cell types, sub-clustering is considered to further dissect cell type heterogeneity within major groups [144, 145].

Apart from filtering out technically low-quality cells in QC step, it is critical to investigate biological doublets i.e., nuclei confounded by looking into the expression of conflicting lineage markers after initial annotation of clusters [146-148]. For example, if a cluster expresses marker for both astrocytes and vascular cells- it indicates potentially confounded cells, and thus should be removed from downstream analysis [147].

1.4.4 Data visualization

Non-linear techniques like t-SNE [123, 124] and UMAP [125, 126] were developed for effective visualization of sc/snRNA-seq data. Although linear methods such as PCA [117, 118] can capture and visualize the main sources of variations in the data, it is usually unable to capture subtle and complex relationships among the clusters.

While t-SNE is still a useful tool for visualizing and analyzing single-cell data, UMAP offers several advantages that make it a more attractive option for many researchers in the field [125, 126].

- **Scalability and speed:** UMAP is generally faster and more scalable than t-SNE, making it advantageous for larger datasets. Also, UMAP can handle larger datasets than t-SNE. While t-SNE can struggle with datasets containing more than 50,000 cells, UMAP can efficiently process datasets with hundreds of thousands of cells.

- **Better preservation of data structure:** UMAP tends to better preserve the global structure of the data compared to t-SNE. It maintains the relationships between distant points in the high-dimensional space more accurately when projecting them into lower-dimensional space. This means that the overall layout of clusters or cell populations is better maintained in UMAP, reflecting the broader relationships between different groups of cells. Also, UMAP is known for its ability to maintain more faithful representations of local neighborhoods and distances between neighboring data points. This results in improved preservation of local relationships among cells, capturing both fine-scale and coarse-scale structures within clusters.

- **Flexibility in parameter tuning:** UMAP offers greater flexibility in adjusting parameters, such as the number of neighbors and the minimum distance, allowing users to fine-tune the trade-off between global and local structure preservation based on the characteristics of their data.

- **Parameter independence:** UMAP's performance is less dependent on the choice of hyperparameters compared to t-SNE. While t-SNE's perplexity parameter requires careful tuning, UMAP is less sensitive to variations in its parameters, making it more user-friendly.

- **Interpretability:** UMAP provides more consistent embeddings across different runs or subsamples of the same dataset. This stability is crucial for obtaining reliable and reproducible results, especially when exploring the same dataset under different conditions or at different time points.

1.4.5 Trajectory analysis

It is to be noted that cell differentiation and maturation are inherently dynamic and might not be adequately captured by discrete analytical approaches like clustering. Hence, trajectory inference and estimation of pseudotime (proxy of developmental time) has surfaced that allow researchers to reconstruct developmental trajectories and capture dynamic transitions among cell populations within complex tissues [149-152]. This approach unveils the nuanced dynamics of cellular differentiation and maturation by modelling critical decision points, often referred to as branch points of a lineage trajectory. Through this analysis, crucial insights into

key regulatory genes and signalling pathways are revealed, providing a comprehensive understanding of the intricate nature of cellular development.

Several state-of-the-art tools have been developed for inferring trajectory analysis. For example, STREAM [149] grounded on inferring Elastic Principal Graph optimization [153] is a popular tool for disentangling complex trajectories, with multiple lineage differentiation paths and subsequently, visualizing them. Another well-known method is Diffusion Pseudotime (DPT) [150] that infers the progression of cells along developmental trajectories by modelling the diffusion of cellular states according to global transcriptomic similarities.

1.4.6 Receptor-ligand interaction analysis

Receptor-ligand interaction analysis using snRNA-seq data provides a system-level understanding of cell communication within tissues [154-157]. It helps identify potential signalling networks that contribute to cellular function, development, and disease. This approach is particularly powerful in capturing the nuanced and cell type/state-specific nature of intercellular communication within complex tissues. Differential expression analysis such as, Wilcoxon rank sum or multivariate Logistic Regression test may be performed to identify significant changes in receptors or ligands expression between different conditions or cell types [156].

1.4.7 Transcription regulation and network modelling

PySCENIC (Python Single-Cell rEgulatory Network Inference and Clustering) [158, 159] is a powerful tool used for transcription factor (TF) inference analysis in snRNA-seq data. This computational framework aims to uncover regulatory networks by identifying potential transcriptional regulators and their target genes within individual nuclei. PySCENIC employs the GENIE3 algorithm to predict regulatory interactions based on gene expression patterns, capturing both direct and indirect relationships between TFs and their targets. By integrating these predictions with known cis-regulatory motifs, PySCENIC infers TF activity in single nuclei, providing insights into the transcriptional landscape of individual nuclei. The resulting information aids in understanding the regulatory dynamics of cellular states, offering a valuable resource for deciphering the intricacies of gene regulatory networks in complex tissues at a single-cell resolution.

In general, sc/snRNA sequencing exhibit high drop-out events and low library sizes, posing challenges in detecting differentially expressed TFs and co-regulators [160]. Previously,

methods like scCapture-Seq [160] were developed for the targeted sequencing of TFs. The authors observed a 36-fold enrichment for TF reads compared to pre-capture scRNA-seq data, leading to an increased detection of total TFs in the sample. Moreover, each cell was shown to express a broader range of TFs based on log-fold change thresholds. Hence, it is possible to utilize a list of differentially expressed genes (such as cell or disease markers) and reverse-engineer the enriched TF or co-regulators acting on distinct target sets. In this regard, iRegulon [161] combines motif enrichment analysis, motif scanning, and a genome-wide ranking-and-recovery method to assign an enrichment score to each TF or co-regulator, reflecting the likelihood that the TF regulates an optimal set of direct targets. In this way, tools such as PySCENIC and iRegulon can unveil the transcriptional regulatory network associated with a cell type/state or in disease vs healthy conditions [158, 159, 161].

On other hand, network analysis of differentially expressed genes (DEGs) inferred from snRNA-seq data involves the construction and exploration of biological networks to uncover interactions and functional relationships among genes. Nodes in the network represent genes, and edges represent interactions or correlations between them [162, 163]. Various types of networks can be constructed, such as co-expression networks (based on expression patterns), protein-protein interaction networks, or regulatory networks. Importantly, clustering algorithms, such as Glay [164] or MCODE [165] are often applied to identify modules or groups of genes that are tightly interconnected within the network. These modules may represent functional units or pathways, and hence are required for organization and coordination of cellular activities. Of note, most of the biological networks are scale free i.e., where a few nodes (hubs) have a disproportionately high number of connections (degree) compared to most nodes. This topology follows a power-law distribution and contributes to the robustness and resilience of biological networks [163, 166-168]. Identifying the hub nodes using tools like NetworkAnalyzer [169] aids in uncovering central regulators of the network and is crucial for understanding the hierarchical organization and control points within biological systems.

1.5 Spatial transcriptomics

Recent advancements in single cell and nuclei RNA sequencing techniques and Human Cell Atlas (HCA) initiative [170], enabled researchers to highlight the diversity within cell populations. Complementing this insight by showing where specific cell types or states are located within tissues contributes to a more detailed understanding of cellular heterogeneity and functions. The recognition of the need to incorporate information into a spatial context has driven the development of novel techniques collectively known as spatially resolved transcriptomics (SRT). Spatially resolved transcriptomics [171, 172] serves as an umbrella term encompassing various methods employed to associate gene expression information with its spatial location. In recent years, the efficacy of these approaches has been demonstrated across diverse tissues and systems, ranging from oncology research to studies on organ development [173-179]. In context cancer, understanding the spatial distribution of gene expression can shed light into tumor heterogeneity and tumor microenvironment [173, 177]. This knowledge could be translated into developing targeted therapies and precision medicine approaches. In case of embryonic development [178] or tissue regeneration [179], spatial transcriptomics helps map the precise gene expression patterns in different regions. This adds to the understanding of how cells differentiate and organize themselves spatially to form functional tissues and organs.

1.5.1 In-situ sequencing

Efforts to gauge gene expressions in a spatial context have been established through techniques such as multiplexed fluorescence in situ hybridization (FISH) [180-182] and in situ/intact tissue sequencing [183-188]. These methods enable the simultaneous measurement of mRNA localization in a spatial context. Like FISH, in-situ sequencing (ISS) [183-185] employs fluorescence microscopy to visualize individual transcripts with subcellular precision. In contrast to sequential imaging of barcoded FISH probes, ISS is a technique that allows for the direct imaging and sequencing of RNA molecules within intact cells or tissues. To be precise, this method involves the use of sequence (or, gene) specific padlock probes, which hybridize to target RNA sequences, and rolling-circle amplification (RCA) to generate localized signals [185]. These signals are then read using sequential imaging and sequencing steps, providing spatial information about the distribution of specific RNA molecules within their native cellular context. ISS offers a powerful approach to studying gene expression patterns at the single-cell level, enabling researchers to analyze the spatial organization of RNA in complex biological samples.

1.5.2 10X Visium

Sequencing-based technologies, constituting the second group of SRT methods, involve the utilization of a solid surface with spatially indexed locations for the capture of RNA molecules. A technique introduced by Ståhl et al. in 2016 [187], referred to as spatial transcriptomics (ST), utilizes glass slides featuring microarrays comprised of capture regions arranged in a spot pattern. In this configuration, each spot incorporates millions of oligonucleotides that bear a spatial barcode designating the spot's position, a unique molecular identifier (UMI), and a polyT sequence designed for capturing polyadenylated transcripts. Subsequent refinements by 10x Genomics enhanced the method's resolution by reducing the spot diameter from 100 µm to 55 µm. Additionally, the total number of spots per capture area increased from ~1,000 to ~5,000 and ~14,000 spots [189]. This modified technique was subsequently commercialized as the Visium spatial gene expression assay. 10X Visium technique was originally designed to work on fresh frozen samples but was later expanded to FFPE samples [190]. Upon placing fresh-frozen tissue sections onto the Visium capture areas, the samples are fixed using methanol and then stained, commonly with immunostaining or Hematoxylin and Eosin, to acquire histological images. The Visium assay offers a significant advantage by combining histological images with transcriptomics data, enabling the integration of gene expression and histological information. This integration provides a spatial perspective to transcriptome profiling. Following the imaging step, tissue sections undergo enzymatic permeabilization to open the plasma membrane, facilitating the diffusion of RNA out of the cells toward the array surface. Given that each tissue possesses a unique cell composition with distinct properties, it is crucial to optimize the permeabilization time for each tissue type to achieve uniform permeabilization across the entire section. After RNA hybridization to the array surface, a reverse transcription reaction extends the surface probes. This merging of the RNA sequence with its spatial position barcode enables the simultaneous readout of both transcript and location information. The subsequent stage involves synthesizing complementary strands from cDNA, which are then collected for the final library preparation and subsequent sequencing.

1.5.3 Advantages and limitations of ISS and 10X Visium

Identifying the project goals accurately is crucial for informed decision-making in choosing between targeted ISS and untargeted 10X Visium approaches and determining the requisite resolution to address raised questions and hypotheses. While imaging-based ISS offers subcellular resolution, it struggles with the optical crowding effect, limiting the number of imaged molecules. Partial mitigation involves serial staining rounds or molecule separation

via sample expansion, but both methods increase imaging time, a notable bottleneck. Tissue expansion diminishes the area captured per field of view, restricting imaging throughput. ISS is known to have a higher precision for the identification of individual cells, given targeted sequencing of RNA molecules [185]. In contrast, sequencing based 10X Visium ensures faster sample processing and higher throughput while targeting the entire transcriptome at a lower resolution. Hence, it is difficult to capture extremely fine details of cellular interactions due to the size of tissue spots [53, 187-191]. In this regard, emerging methodologies such as the Xenium in-situ technique [192] have demonstrated significant enhancements in throughput and sensitivity compared to other in ISS methods. It has garnered attention for its capability to map hundreds of transcripts in situ with subcellular resolution.

Generally, for analyzing a selected gene set, targeted approaches with subcellular resolution are preferable, while untargeted methods are common in exploratory research. Untargeted methods serve as a starting point to identify genes of interest for subsequent targeted analysis, offering higher resolution measurements and further validation of findings.

1.5.4 Building an atlas- integrating snRNA-seq with spatial transcriptomics

A pivotal technology employed in constructing cell atlases is single cell and single nuclei RNA seq. As described before, this approach facilitates the exploration of cellular heterogeneity at the transcriptomic level, enabling the categorization of cells based on their gene expression profiles and the discovery of previously unidentified cell types/states [192-194]. Despite the merits of single cell or nuclei RNA-seq, the overall tissue organization is not investigated that creates a knowledge gap in linking transcriptomic heterogeneity with the spatial organization of the tissue. While imaging can offer insights into the histological-level spatial organization of cells within each tissue, the advent of SRT methods has been instrumental in studying in situ gene expression, cell types/states, their spatial relationships.

- **Integration of snRNA-seq data with 10X Visium**

snRNA-seq provides gene expression data at the individual nuclei level but does not profile the spatial organization of cells. 10X Visium [187] gene expression retains spatial information, but the resolution of each spot is constrained, as each spot encompasses multiple cells, typically ranging from 1 to 10 cells. Hence, this integration allows researchers to leverage the complementary strengths of each technology. For instance, deconvolution methods strive to discern the cell types/states and their respective proportions contributing to a spot, while mapping methods aim to assign the most probable dominant cell type to a spot [195-197].

Deconvolution approach-based tools such as RCTD [195] and Spotlight [196] employ probabilistic models based on references to identify cell types within a single spot, which may contain a mixture of cell types. It utilizes a maximum-likelihood estimation to infer the proportions of different cell types and subsequently projects this information onto a spatial map of cell types.

- **Integration of snRNA-seq data with ISS**

Aligning the spatial information from ISS data with snRNA-seq inferred gene expression profiles can be challenging. This involves mapping the spatial coordinates of in situ data to the corresponding cell identities in snRNA-seq data [198]. Mostly, snRNA-seq and ISS are performed on separate samples of a tissue- so, highly specific, and robust cell type or condition specific markers can be used as a proxy to integrate two data modalities and investigate spatial relationships between clusters. Unlike 10X Visium but like FISH, ISS utilizes pre-designed probes for labelling target transcripts. Hence, differential gene expression analysis using snRNA-seq can be used to pre-determine marker genes for the targeted ISS probe panels. This predefined repertoire of target genes results in a biased detection of transcriptomes. Often, a correlation analysis is used between the snRNA-seq derived signatures, and the observed mRNA counts within the cells of segmented spatial data.

A meticulously constructed cell atlas functions as a valuable resource for the scientific community, bearing substantial implications for comprehending physiological processes. Furthermore, it establishes a benchmark for investigating the underlying mechanisms of diseases. Cell atlases play a pivotal role in the identification of novel therapeutic targets or biomarkers. By offering a comprehensive understanding of cell identity and function, they contribute significantly to the development of innovative medical approaches.

1.6 Single nuclei profiling in advancing Precision Medicine

The development of snRNA-seq revolutionized the field of Precision Medicine. It allowed researchers to examine the transcriptomes of individual cells, providing insights into the diversity of cell types, gene expression patterns, and regulatory networks in both healthy and diseased tissues. In summary, the history of snRNA-seq in precision medicine has evolved from uncovering genetic variations to providing detailed insights into cellular functions and interactions [199-202]. Its integration into clinical practice holds great promise for tailoring medical treatments to individual patients based on the specific characteristics of their cells.

Like single cell profiling, snRNA-seq has made significant contributions to precision medicine across various fields, providing high-resolution insights into gene expression profiles at the cellular resolution. Here are some notable advancements of snRNA-seq in precision medicine:

- **Heterogeneity Analysis:** snRNA-seq allows for the profiling of individual cell nuclei, providing a detailed understanding of cellular heterogeneity within tissues. This level of resolution is crucial for identifying rare cell populations, understanding cellular diversity, and characterizing subtypes within complex tissues or tumors.

- **Clinical Biomarker Discovery:** In diseases such as cancer, neurodegenerative disorders, and autoimmune conditions, snRNA-seq has been instrumental in discovering molecular subtypes based on gene expression profiles. These subtypes may have distinct clinical outcomes and responses to treatments, allowing for more precise disease stratification and personalized therapeutic strategies.

- **Identification of rare cell types:** In tissues where certain cell types are rare but functionally critical, such as stem cells or specific immune cell subsets, snRNA-seq facilitates their identification and characterization. This is particularly relevant in the context of diseases where these rare populations may play a crucial role.

- **Inferring developmental trajectories:** In the context of Precision Medicine, trajectory analysis offers valuable insights into the heterogeneity of cellular responses in various disease states or during therapeutic interventions. In diseases such as cancer, understanding the trajectories of malignant cells can aid in identifying specific driver genes and potential therapeutic targets. Moreover, trajectory analysis can inform the development of targeted interventions tailored to individual patient profiles. Leveraging the information derived from snRNA-seq data, Precision Medicine strategies can be refined, allowing for the design of personalized therapeutic approaches that consider the specific developmental trajectories and molecular signatures of patient-derived cells.

- **Mapping Tissue Architecture:** Integrating spatial information with gene expression data is essential for understanding tissue architecture. Techniques like 10X Visium and spatial transcriptomics combined with snRNA-seq enable the mapping of gene expression within the spatial context of tissues, providing insights into how cells interact and organize in their native environments.

- **Drug Response Prediction:** By characterizing the transcriptomic profiles of individual cells, snRNA-seq contributes to the identification of potential drug targets and predicts responses to specific treatments. This can inform the development of personalized treatment plans tailored to the molecular characteristics of a patient's disease.

- **Longitudinal Studies:** snRNA-seq enables the study of how gene expression profiles change over time. This longitudinal approach is valuable for understanding disease progression, identifying early markers of pathology, and monitoring the effects of interventions over time.

- **Integration with Multi-Omics Data:** Integrating snRNA-seq data with other omics data, such as genomics, epigenomics, and proteomics, allows for a more comprehensive understanding of cellular function. This holistic approach is essential for unravelling the complexity of diseases and identifying multi-faceted therapeutic targets.

As technology matures, snRNA-seq is moving toward clinical translation. Its potential for diagnosing diseases, predicting disease progression, and guiding treatment decisions underscores its role in advancing precision medicine into clinical practice [199-202]. In summary, snRNA-seq has significantly advanced precision medicine by providing unparalleled insights into the cellular and molecular landscapes of tissues and diseases. The integration of differentiation trajectory analysis with Precision Medicine holds great promise for advancing our understanding of disease mechanisms, optimizing treatment strategies, and ultimately improving patient outcomes by tailoring interventions to the individual cellular and molecular characteristics of each patient.

1.7 Human pregnancy and maternal-fetal interface

Human pregnancy is a complex and dynamic process that involves the growth and development of a fertilized egg into a fully formed fetus, culminating in the birth of a new life. The pregnancy is divided into three trimesters, each lasting approximately 12-13 weeks [203]. During the first trimester, the fertilized egg lodges itself in the uterus, initiating the formation of an embryo enveloped by a structure known as the gestational sac, supplying nutrients and support for its growth [204, 205]. As the embryo grows, it forms placenta, a temporary organ which connects to the mother's uterine wall and facilitates nutrients and oxygen exchange between the mother and the developing fetus [206-208]. During the second trimester, the fetus considerably develops, and its underlying organs start to mature. The placenta keeps growing and expanding, providing increased nourishment to the developing fetus. The fetus undergoes its final development during the third trimester and prepares for birth. The mother's body also undergoes significant changes, including producing hormones that help relax the uterus muscles and prepare for labor and delivery.

The maternal-fetal interface is the bridge between the maternal and fetal tissues [146, 209-211]. This interface is formed by the placenta (fetal tissue) and decidua (maternal tissue) and

acts as a protective barrier between the two, as illustrated in **Figure 1.2**. In this regard, the placenta serves as a shield that prevents the mother's immune system from attacking the developing baby. This point of contact is a complex and dynamic system orchestrated by various cells and signaling molecules, including immune cells, cytokines, and hormones. Disruptions to this interface can lead to serious complications during pregnancy, including preterm labor, pre-eclampsia, and fetal growth restriction [212, 213]. Elucidating the pathophysiological mechanisms that regulate the maternal-fetal interface is a growing area of research that has significant implications for the mother and her child.

Previous major single-cell studies, such as those by Rosa Vento Tormo et al. in 2018 [209] and Surywavanshi et al. in 2018 [146], were focused solely on the first trimester of pregnancy. These studies comprehensively characterized the cell types and states at the maternal-fetal interface, providing insights into trophoblasts (the outer layer of the developing placenta), fetal vessels, and distinct decidual cell types, encompassing stromal, endothelial, immune, and epithelial cells **(Figure 1.2).**

The development of the human placental trophoblast is a complex and dynamic process that begins shortly after fertilization and continues throughout pregnancy. Trophoblast cells are the first cells to differentiate from the fertilized egg and are responsible for the formation and function of the placenta [146, 209-211, 214]. In this regard, major trophoblasts subtypes were identified, including invasive extra-villous trophoblasts (EVT) that help to anchor the placenta to the uterine wall and ultimately, invade the maternal decidua to drive uterine vessel remodelling [215-218]. Syncytiotrophoblasts (STB) are relatively homogenous and are in contact with the maternal blood [209-212]. Hence, STB cell type is involved with the gas and nutrient exchange alongside hormone production required to sustain healthy pregnancies. Importantly, cytotrophoblasts (CTB) play a crucial role in the early stages of pregnancy by serving as proliferative and undifferentiated cells that give rise to EVT and STB [209, 219], thereby contributing to placental development and establishing the maternal-fetal interface **(Figure 1.2)**.

The maternal decidua constitutes the specialized lining of the uterus that forms during pregnancy under the influence of progesterone**.** Former scRNA-seq studies [146, 209] profiling decidua decoded major subtypes of stromal cells (DSCs), fibroblasts (FB), pericytes, endothelial cells (EC) epithelial cells (EpC) and different subsets of immune cells **(Figure 1.2)**. Importantly, maternal decidua is an immunologically fertile site, where several types of immune cells intermingle, including natural killer (NK) cells, macrophages, and T-cells.

Overall, the placental and decidual cells cooperate to support fetal growth and development and protect it from external injuries. Hence, understanding the physiological roles of these cells in healthy and diseased pregnancies is an important area of research that has important implications for maternal and fetal health.

1.7.1 Cell types/states of human decidua

Several cell types have been identified within the decidua, which are detailed as follows:

- **Decidual stromal cells:** Decidual stromal cells (DSC) [146, 209, 220, 221] play a pivotal role in pregnancy by contributing to the formation and maintenance of the decidua, the specialized uterine lining essential for successful implantation and embryo development. These cells undergo dynamic changes, exhibiting immunomodulatory properties that regulate maternal-fetal immune tolerance, ensuring a supportive environment for embryonic growth. Additionally, decidua stromal cells participate in tissue remodeling and angiogenesis, crucial processes for establishing a functional placenta and sustaining a healthy pregnancy.

- **Immune cells:** The decidua is an immunologically active site [222] that harbors native immune cells such as macrophages (MAC), dendritic cells (DC), natural killer (NK) cells, and T cells. These cells play central roles in regulating maternal immune responses throughout gestation and protecting the developing fetus from pathogens [146, 209, 222-225].

- **Decidual endothelial cells:** Major endothelial cell types (EC) [146, 209, 226-229] include the vascular and lymphatic endothelial cells. Vascular endothelial cells (VEC) form the inner lining of blood vessels within the decidua, playing a crucial role in facilitating nutrient and oxygen exchange between the maternal blood and the developing fetus. On the other hand, lymphatic endothelial cells (LEC) are involved in the formation and maintenance of lymphatic vessels in the decidua. These vessels are essential for draining excess fluids and immune cells from the tissue, contributing to the regulation of the local immune response and tissue homeostasis during pregnancy. Both types of EC(s) are integral components of the complex microenvironment in the decidua, supporting the physiological processes associated with pregnancy.

- **Decidual epithelial cells:** This cell type or EpC line the surfaces of the maternal and fetal tissues and play an important role in maintaining the integrity of the barrier between the two tissues. A subset of EpC also forms the lining of the uterine cavity and those that make up the chorion (a membrane that surrounds the developing fetus) [146, 209, 230].

- **Extravillous trophoblasts:** These specialized cells are originally derived from the placenta and invade the maternal tissues of the decidua, where they play important

roles in establishing and maintaining the maternal-fetal interface [215-218].

1.7.2 Cell types/states of human placenta

As mentioned before, the human placenta is a pivotal organ supporting fetal growth and development during pregnancy. It comprises several different cell types that are detailed below:

- **Syncytiotrophoblasts:** These are the outermost cells of the placenta and are involved in the nutrient and gas exchange between the mother and fetus. Importantly, STB is endocrine in nature i.e., they produce placental hormones critical for maintaining pregnancy, such as human chorionic gonadotropin (hCG), placental lactogen *CSH1* and *CSH2* [231, 232]. *CGA* is a subunit of hCG and is integral to STB functions, including implantation, corpus luteum support, and immunomodulation [233, 234].

- **Cytotrophoblasts:** These cells are located beneath the STB and are involved in the development of the placenta, including the formation of new blood vessels [146, 209-212].

- **Cell column trophoblasts:** The placental part of EVT (pre-invasion) is also called as cell column trophoblasts (CCT) that play a crucial role in invasion and remodeling of maternal tissues, facilitating the establishment of the maternal-fetal interface and ensuring proper nutrient and oxygen exchange between the mother and the developing fetus. Their invasive properties are essential for processes such as implantation, uterine spiral artery modification, and the formation of a functional placental bed to support a healthy pregnancy [235, 236].

- **Hofbauer cells:** Hofbauer cells (HBC) [237-240], also known as fetal macrophages, are specialized immune cells found within the placenta. These macrophages hold multifactorial roles, including immune modulation, phagocytosis, and the regulation of inflammation, contributing to the dynamic interactions at the maternal-fetal interface. Their presence in the placenta suggests a role in both immune defense and the maintenance of a tolerogenic environment to support fetal development.

- **Placenta associated maternal macrophages:** Apart from the fetal-originating HBC, a group of maternal macrophages linked to the placenta (PAMM) [238] has recently been identified, and these can be observed attached to the surface of placental villi. In contrast to HBC, PAMM(s) are *HLA-DRA* positive and *FOLR2* negative cells [238]. PAMM is suggested to play a role in inhibiting microbe transmission at areas where the syncytium is damaged. Conversely, they could potentially serve as a pathway for infection by allowing the survival and replication of microbes within macrophages.

- **Fibroblasts:** These connective tissue cells are involved in tissue development,

wound healing, and the synthesis of extracellular matrix components (ECM), contributing to the structural and functional integrity of during embryonic development [146, 241].

- **Endothelial cells:** Placental EC(s) mainly constitute the vascular endothelial cells (VEC) that line the blood vessels and regulate blood flow to the developing fetus [195, 196].

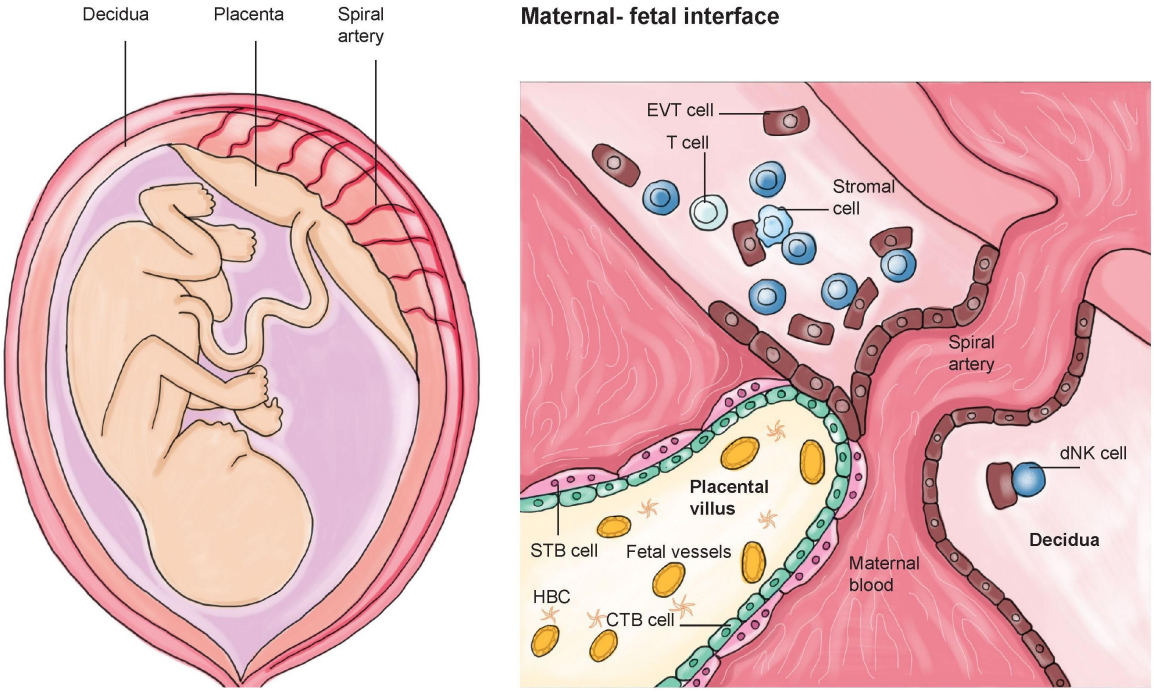


Figure 1.2: Illustration of maternal-fetal interface and underlying cell type/state atlas.
Cell types derived from the early embryo include villous cytotrophoblast (CTB) cells, which line placental structures called villi, syncytiotrophoblast (STB) cells covering the villus surface, and extravillous trophoblast (EVT) cells lining maternal blood vessels and intermingling with maternal cells in the decidua. Previous studies identified various maternal immune cell types, including T-cells and decidual natural killer (dNK) cells, along with stromal cells providing structural and functional support to the decidua throughout the pregnancy. Figure adapted from Rajagopalan et. al 2018 [242] and digitally painted from scratch using InDesign software.

1.7.3 Development of human placental trophoblast

During the first week of development, the fertilized egg undergoes several cell divisions to form a blastocyst structure. The genesis of all trophoblast lineages originates from the trophectoderm (TE) cells, the outer layer of blastocyst **(Figure 1.3).** Placental trophoblast

stem (TS) cells are a population of precursor cells that give rise to the various subtypes of placental trophoblast cells. These progenitors are believed to be present in the early stages of placental development and are responsible for initiating the process of trophoblast differentiation [219].

TS cells exhibit the ability for self-renewal, maintaining an undifferentiated state through cell division facilitated by signalling pathways, including the fibroblast growth factor (FGF) and bone morphogenetic protein (BMP) [219]. Also, placental TS cells are considered pluripotent and express markers associated with pluripotency, such as transcription factors *Oct4* (Octamer-binding transcription factor 4), *Sox2* (SRY-Box 2), and *Nanog* [243-247]. *CDX2* and *EOMES* are two key transcription factors that orchestrate the differentiation and function of TS cells [219, 247, 248]. *CDX2* plays a crucial role in the early specification of trophectoderm and the maintenance of trophoblast identity, while *EOMES* is also involved in the differentiation of mouse and human trophoblast progenitors [247, 248]. After implantation into the uterine wall, the TS cells differentiate into major trophoblast subtypes.

Biologically, the CTB is known to harbor TS progenitors that can further differentiate into two divergent lineage paths- STB and EVT [209, 219] **(Figure 1.3).** The process begins with the proliferation of mononuclear CTB cells that initially populate the chorionic villi. Further, CTB(s) undergo fusion to form multinucleated STB(s), which create a syncytial layer crucial for nutrient exchange and hormonal production. STB forms a layer on the surface of the chorionic villi, which are finger-like projections of the placenta that extend into the maternal blood supply.

Of note, a previous scRNA-seq study focussed on TS cells derived from cultured human blastocysts and were able to recapitulate emergence of two lineages, including STB and migratory trophoblasts transcriptomically same as EVT [249]. The synchronized processes of proliferation and differentiation among these lineages play a crucial role in ensuring the success of a pregnancy. Disruptions in trophoblast development and function are believed to be associated with diverse pregnancy complications, encompassing miscarriage, pre-eclampsia, and intrauterine growth restriction [212, 213, 250]. Hence, unravelling the molecular mechanisms governing early placental development have potential implications for reproductive medicine and understanding pregnancy-related disorders.

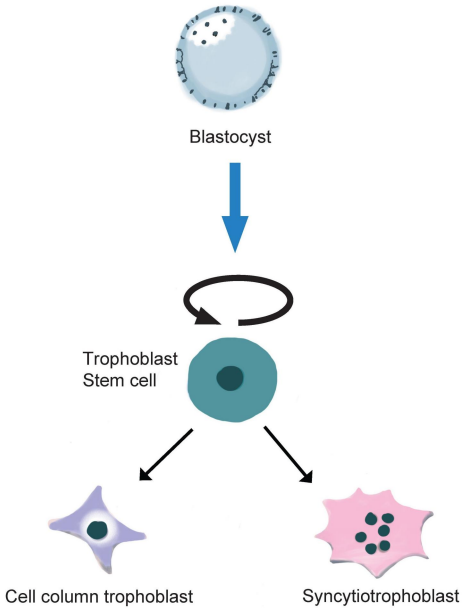


Figure 1.3: Schematic illustrating trophoblast development.
Blastocyst derived trophoblast stem cells differentiating into two divergent lineages, syncytiotrophoblast (STB) cells and cell-column trophoblasts (CCT). Biologically, cytotrophoblast (CTB) act as progenitors and are known to harbor trophoblast stem cells. Figure adapted from Okae et al 2018 [219] and created using InDesign software.

1.8 Early onset Pre-eclampsia

Pre-eclampsia (PE) is a pregnancy-specific disorder characterized by elevated blood pressure and proteinuria (protein in the urine) occurring after 20 weeks of gestation [251-254]. PE and associated hypertensive disorders are reported to contribute to ~14% of annual maternal deaths worldwide, ranking second only to hemorrhage [255, 256], and incur an approximate healthcare cost of nearly 2 billion USD within the first year following delivery [257].

The condition is classified into early-onset pre-eclampsia (eoPE) [258], which manifests before 34 weeks of gestation, and late-onset pre-eclampsia, with onset after 34 weeks. eoPE is typically considered more severe and is linked to higher risks of adverse maternal and fetal outcomes, including preterm delivery, abnormal maternal spiral arteries remodelling, intrauterine growth restriction (IUGR), and fetal distress [259-261]. Despite causing approximately 80,000 maternal and 500,000 fetal deaths annually [262], eoPE remains an under-investigated area of research. In contrast, late-onset pre-eclampsia generally presents milder symptoms, and a lower risk of adverse outcomes [263]. This thesis focuses on eoPE

due to its strong association with placental dysfunction, abnormal fetal development, and maternal factors such as chronic hypertension, autoimmune diseases, and genetic predisposition [264-267].

Generally, it is acknowledged that eoPE unfolds in two stages: irregular placentation during early pregnancy, succeeded by systemic endothelial dysfunction [268, 269]. In normal pregnancy, invading trophoblast cells (EVT) undergo a process of remodelling maternal spiral arteries to ensure proper blood flow to the developing placenta [215-218]. However, in eoPE, inadequate remodelling of these uterine vessels leads to impaired perfusion, hypoxia, and enhanced oxidative stress, contributing to the development of the condition [270, 271]. Additionally, dysregulated interactions between invading EVT and maternal immune cells (like, dNK cells) further exacerbate the inflammatory response, leading to the characteristic features of eoPE, including endothelial dysfunction and maternal vascular complications, as illustrated in **Figure 1.4**.

If untreated, eoPE can lead to lifelong cardiovascular and metabolic complications for the mother and child, that could potentially lead to stroke, liver and kidney failure, and even end-organ damage [264-267]. In terms of clinical management and treatment, eoPE often requires earlier delivery of the baby and more intensive monitoring and medical management of maternal and fetal health. Significantly, once eoPE is established, delivery becomes the sole recourse to address the maternal crisis [259, 272]. The placenta plays a pivotal role in the genesis of eoPE, which is evident in the alleviation of symptoms upon placental delivery [268-270, 273]. Nevertheless, the early pathological mechanisms within the placenta that culminate in the clinical manifestations of PE remain a mystery.

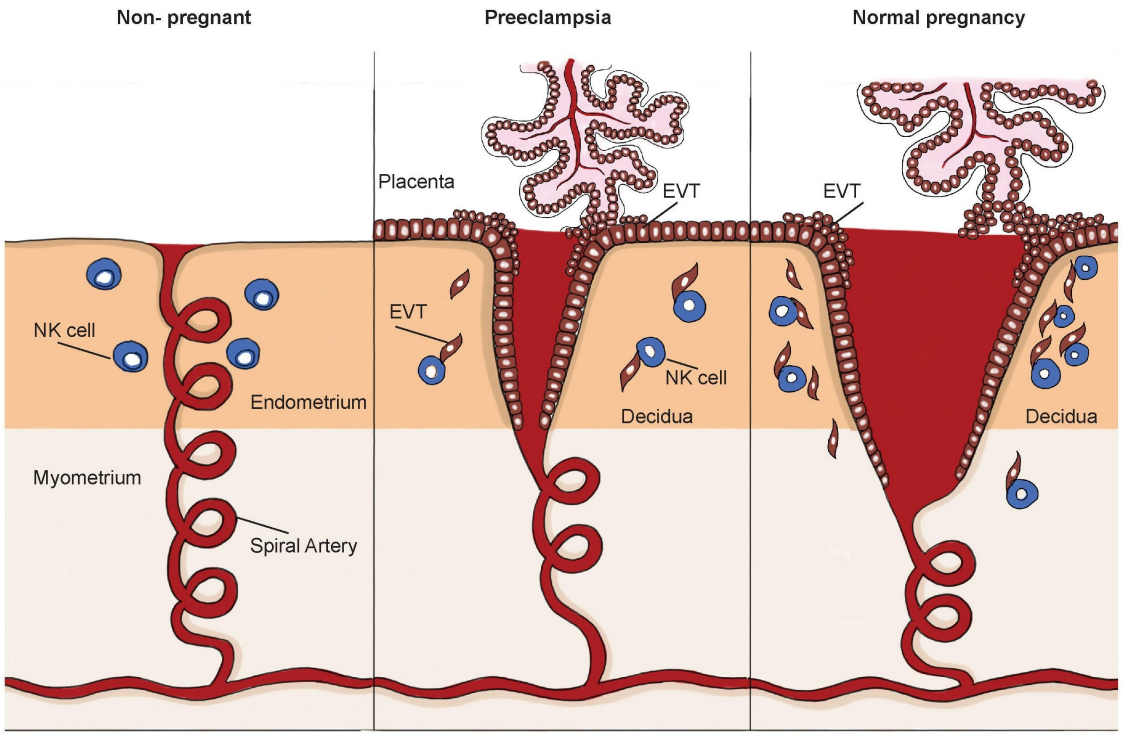


Figure 1.4: Physiological adaptations in pre-eclampsia relative to non-pregnancy and normal pregnancy conditions.
Schematic illustrating maternal spiral artery (uterine vessels) in non-pregnant individuals, and how the same is remodelled during healthy and preeclamptic pregnancies. Early-onset pre-eclampsia (eoPE) is distinguished by aberrant trophoblastic invasion, stemming from shallow and constricted transformation of spiral arteries. This condition results in placental ischemia and heightens oxidative stress. Figure created using InDesign software, and partly adapted from Dimitriadis E et al 2023 [293]

1.8.1 Genomics and bulk transcriptomics of early onset pre-eclampsia

Pre-eclampsia is known to result from intricate interactions involving multiple maternal and fetal genes, and demonstrate evident heritability, estimated to range between 31% and 54% [274, 275]. Past Whole Genome/Exome Sequencing studies examining pre-eclampsia provided valuable insights into the genetic landscape of the disorder that further illuminated the dysregulation of molecular pathways associated with trophoblast function, vascular remodelling, and immune regulation [274-278]. Former studies found specific genetic variants in or proximal to genes such as *RAS, AGT, ACE, CTLA4, ENG, F5, F2, FV, APOE, GST, LPL, NOX1* and *SERPINE1* [274, 278]. Significantly, several of these variants are also recognized as risk factors for cardiovascular disease via the renin–angiotensin system, fibrinolysis and coagulation regulation, oxidative stress, and lipid metabolism- indicating shared genetic predispositions between pre-eclampsia and cardiovascular conditions.

Prominent diagnostic biomarkers identified in the context of pre-eclampsia, include *sFlt-1* (Soluble Fms-Like Tyrosine Kinase-1) [279] and *PlGF* (Placental Growth Factor) [281] that offer valuable insights into the pathophysiology for the condition. A previous study analyzed over 7 million genetic variants in 2,658 offspring from preeclamptic individuals and 308,292 controls from the general population pinpointed a singular association with the *FLT1* (Fms-like tyrosine kinase 1) gene on chromosome 13 [279]. The *FLT1* gene encodes *sFLT1*, a splice variant of the vascular endothelial growth factor (*VEGF*) receptor renowned for its antiangiogenic properties, achieved through the inhibition of proangiogenic factor signaling [279, 280]. Importantly, pre-eclampsia is characterized by elevated levels of *sFlt-1*, which hinders *PlGF* and *VEGF*, contributing to endothelial dysfunction and hypertension [279].

Major advances over the last decade explored the diagnostic and prognostic value of RNA-based markers in maternal blood or placental tissues, using bulk RNA-seq and microarray approaches [282-284]. Of note, previous studies identified potential biomarkers, such as *sFLT1, PGF, PAPPA2,* and *INHBA*, and linked altered placental gene expression to impaired trophoblast invasion, abnormal vascular remodelling, and compromised maternal-fetal exchange. A previous study based on bulk RNA-seq, and weighted gene correlation network analysis (WGCNA) postulated eoPE±FGR samples were consistently distinguished by glycolysis/gluconeogenesis metabolic pathway [285]. This could be related to placental metabolic reprogramming under oxidative stress and hypoxia during eoPE.

Having said that, traditional bulk RNA-seq and microarray methods mask the heterogeneity of individual cell populations, limiting our understanding of the specific cell types and states involved in the pathogenesis of eoPE. Hence, single cell profiling of maternal-fetal interface can enhance our knowledge regarding the intricate biology of eoPE by unravelling key dysregulated pathways, novel biomarkers, and offers potential targets at a cellular resolution for therapeutic intervention. Although two previous studies [286, 287] performed scRNA-seq of PE-affected placenta- they were very limited in terms of sample size, number of cells profiled and lacked a comprehensive overview of molecular dysregulations across maternal-fetal interface.

1.8.2 scRNA-seq of early onset pre-eclampsia

In a recent scRNA-seq study conducted by Admati et al. in 2023 [288], significant dysregulation of major placental cell types was identified in eoPE, but not in late-onset PE. The study further highlighted heightened cellular stress and pre-apoptotic states specific to the placental vasculature in eoPE, a distinction not observed in late-onset PE [288]. Also, eoPE was found to exhibit relatively higher prevalence of inflammatory cell types, and increased expression of pro-inflammatory cytokines in various myeloid cell types as compared to late PE. Taken together, these observations indicated that eoPE severely affects the underlying maternal-fetal cell types/states relative to late PE.

However, the investigation did not uncover detailed molecular dysregulations in STB beyond cell-autonomous transcriptional regulation of FLT1/PGF [288]. Of note, it is known that STB nuclei undergo senescence as part of the differentiation process, ultimately shedding into the maternal bloodstream [236, 289, 290]. Along with their secreted factors and cellular debris, STB nuclei contribute to the release of up to 3 grams of fetal-origin protein mass daily into the maternal circulation [236]. Prior research has indicated that factors shed from the pre-eclamptic placenta are implicated in the development of cardiovascular and hypertensive symptoms in the mother, subsequently categorized as the clinical syndrome recognized as the “maternal pre-eclampsia syndrome” [291]. Hence, a detailed understanding of STB specific markers and its dysregulated secreted factors are required to be profiled to understand the disease pathogenesis. Moreover, no sc/snRNA-seq studies delved into understanding the role of abnormal trophoblast development in eoPE- mainly, how molecular disruptions in STB can potentially lead to exacerbated maternal and fetal outcomes. Importantly, no studies so far clearly identified the cell-of-origin of eoPE and how the disease is translated between placenta and decidua.

1.8.3 Spatial transcriptomics of early onset pre-eclampsia

A recent study by Arutyunyan, Roberts et. al 2023 [292] constructed a spatially detailed multi-omics single-cell atlas encompassing the first-trimester human maternal–fetal interface, inclusive of the myometrium. Leveraging this cellular map, the authors inferred potential transcription factors orchestrating EVT differentiation, invasion, demonstrating their conservation in in-vitro models derived from primary trophoblast organoids and trophoblast stem cells. However, no former studies spatially profiled the maternal-fetal interface in healthy term and eoPE affected pregnancies.

Hence, integration of spatial transcriptomics with snRNA-seq can significantly advance eoPE research by building a comprehensive disease atlas. For example, it would enable the identification and localization of dysregulated genes associated with eoPE, shedding light on the molecular mechanisms underlying the disorder. Mapping the spatial relationships between the placental cell types/states can elucidate how communication is disrupted in eoPE. Moreover, the approach will facilitate the characterization of vascular remodeling patterns associated with eoPE. Overall, integration of spatial transcriptomic data may unveil targetable biomarkers, enhance our understanding of disease progression, and open avenues for personalized therapeutic strategies in eoPE.

1.9 Human pancreatic islet development

In the early stages of development, the endoderm is responsible for generating the epithelial lining of the digestive tract. The dorsal pancreatic bud emerges from the posterior foregut, and simultaneously, the ventral pancreatic bud originates from the hepatic diverticulum. Eventually, these buds merge to constitute the pancreas.

The pancreas is a dual-function organ with endocrine and exocrine functions [72, 294, 295]. The endocrine portion, called the islets of Langerhans, produces hormones such as insulin and glucagon, regulating blood sugar levels. The exocrine part secretes digestive enzymes into the small intestine to aid in the digestion of food. The details of pancreatic cell types are discussed below:

- **Exocrine cells:** These cells are organized into small clusters called acinar cells that produce a mixture of enzymes that help to break down carbohydrates, proteins, and fats in food. The two main types of exocrine cells are:
    - **Acinar cells:** These cells produce digestive enzymes and other substances that are released into the ducts.
    - **Duct cells**: line the ducts that carry digestive enzymes from the acini to the small intestine. Additionally, multipotent progenitors can differentiate into ductal progenitors, which play a crucial role in pancreas homeostasis and regeneration. The dynamic interplay of these cell types orchestrates the complex functions of the mature pancreas.
- **Endocrine cells:** These cells are responsible for producing and releasing hormones directly into the bloodstream. The endocrine cells of the pancreas are clustered together in groups called islets of Langerhans. The islets contain several different types of endocrine cells, including glucagon secreting alpha cells, insulin producing

beta cells, somatostatin producing delta cells, gamma cells, and pancreatic polypeptide secreting PP cells which helps to regulate digestion. These cell types form the islets of Langerhans and regulate hormonal balance.

The pancreatic microenvironment plays a role in influencing cell fate decisions and differentiation. Previous studies have explored how the cellular context and microenvironment affect alpha and β cell differentiation and regeneration [296-299]. In this regard, both in vivo animal models and in vitro cell culture systems have been employed to explore islet regeneration mechanisms in the adult pancreas. Moreover, various approaches have been utilized to stimulate the regeneration of β cells, encompassing pancreatectomy, partial duct ligation, and inducing massive β cell loss through chemicals [300-303]. Comprehending the factors underlying the reduced proliferation observed in aging human pancreatic islet β cells can provide significant insights in type 1 diabetes treatment, where there is a deficiency of insulin-producing beta cells.

Previous scRNA-seq studies of human islet cells have unveiled signatures linked to both type 1 (T1D) and type 2 (T2D) diabetes [304-307], markers associated with islet maturation, islet dedifferentiation, aging, and islet transdifferentiation [308-310]. However, these former studies mostly investigated sorted human fetal pancreatic cells and transplanted insulin-secreting β cells that closely resemble adult islets. It is known that beta cells are incompletely functional at the neonatal stage; however, their roles, such as glucose-regulated insulin secretion, develop and mature over time [311]. Nevertheless, no former studies extensively investigated age-dependent transcriptomic changes in healthy islets by performing comparative studies using neonatal and adult pancreas.

1.10 Pancreatic neuroendocrine carcinoma

Neuroendocrine neoplasms (NEN) [312-314], marked by neuroendocrine differentiation, can develop in various epithelial organs throughout the body. Neuroendocrine cells [315] exhibit features of both nerve cells (neurons) and endocrine cells, which release hormones into the bloodstream. These cells play crucial roles in regulating various physiological processes, including the control of hormone secretion, response to stress, and modulation of the nervous system. In the context of diseases like cancer, understanding the neuroendocrine cell state is essential for unravelling the complexities of tumor heterogeneity and behavior.

The gastrointestinal (GI) system is the predominant site, representing two-thirds of NEN(s), with the pancreas being a major primary location [316, 317]. These neoplasms encompass several distinct entities with diverse etiologies, clinical features, morphological and genomic

characteristics. In this regard, the World Health Organization (WHO) classification of Digestive System Tumour categorized NEN into three groups: well-differentiated neuroendocrine tumors (NETs), poorly differentiated neuroendocrine carcinomas (NECs), and mixed neuroendocrine-non-neuroendocrine neoplasms (MiNEN) [316, 318]. High-grade gastroenteropancreatic NENs (GEP-NENs), including poorly differentiated GEP-NECs and high-grade well-differentiated GEP-NETs, constitute a group of highly aggressive and clinically diverse cancers. Clinical treatment currently utilized for GEP-NENs have been adapted from those employed for small-cell lung cancers (SCLC) owing to their evident clinical and histomorphological resemblances [316, 322]. Taken together, there is a pressing demand for effective therapeutic interventions to address these challenging malignancies [319-322].

Pancreatic neuroendocrine carcinoma (panNEC) is a rare yet highly aggressive cancer among GEP-NECs, comprising only 1-5% of all pancreatic neoplasms [323, 324]. Molecular studies have identified genetic alterations associated with panNEC, including mutations in *TP53, RB1, KRAS*, and alterations in chromatin remodeling genes [323, 324]. However, the defining characteristic of panNEC lies in its undifferentiated morphology that distinguishes it from well-differentiated panNET. This differentiation status is determined by assessing proliferation activity, indicated by a mitotic rate and Ki-67 proliferation index exceeding 20%. The Ki-67 index, a measure of cell proliferation, is determined based on the percentage of actively dividing tumor cells. panNETs exhibit histologically low-grade nuclear characteristics and are classified as G1, G2, or G3 based on their Ki-67 index [318, 319]. In contrast, panNECs exhibit a very high Ki-67 index and carcinoma-like nuclear features, indicating aggressive clinical behavior, frequent metastases, chemotherapeutic resistance, and worse survival outcomes. Although G1 panNETs reportedly demonstrate an overall survival of > 10 years, G2 panNETs typically exhibit the same of around 6 years [319]. Conversely, high-grade panNENs present poorer survival outcomes, with patients diagnosed with NECs surviving less than 10 months [325]. Notably, panNECs exhibit distinct morphological and genetic features compared to well-differentiated tumors [323, 338]. Additionally, panNECs carry a higher mutational burden compared to NET and share genomic traits with adenocarcinomas of the same anatomical sites, such as frequent KRAS mutations in panNEC and pancreatic ductal adenocarcinoma (PDAC) [324]. Patient-derived (PD) xenografts of GEP-NENs provide insufficient insights into the functional and mechanistic aspects of drug responses [326, 327]. Additionally, the restricted availability of NEN cell lines fails to accurately represent panNEC biology [326, 327]- thereby, posing considerable challenges in developing novel treatments and combination therapies for GEP-NEN patients, including panNEC.

Routine diagnostic markers for panNEC include Chromogranin A (CgA) and Synaptophysin (SYP) [328, 329]. CgA is a protein found in the secretory granules of neuroendocrine cells, and elevated levels in the blood or detected by immunohistochemistry in tissue samples can indicate the presence of NE tumors or carcinoma. The presence of SYP is an indicator of neuroendocrine differentiation. CD56, or NCAM1 (neural cell adhesion molecule) [330], is a cell surface protein that is often expressed in neuroendocrine tumors. In combination with CgA and SYP, it can be used as an immunohistochemical marker to confirm the neuroendocrine nature of the carcinoma [331].

panNECs can manifest as either large cell or small cell forms, though this difference lacks a clear molecular basis and clinical implications. panNEC(s) are typically diagnosed at an advanced, metastatic stage, precluding curative surgical intervention. Palliative treatments mainly involve platinum-based chemotherapy, particularly Cisplatin in combination with Etoposide. Also, Temozolomide-based chemotherapy is presently employed in clinical settings for high-grade GEP-NET [332] due to observed lower response rates with platinum-based therapies [325]. Despite these aggressive treatments, the median overall survival for patients with metastatic GEP-NECs remained less than one year [319-321]. The complications in therapeutic response and emergence of multi-drug resistance in panNEC majorly likely stem from the substantial intra-tumoral heterogeneity and plasticity within a tumor that is not properly investigated yet. One of the primary challenges in developing therapeutic strategies for panNEC is the lack of a clear understanding of the fundamental mechanisms that drive tumor biology. This is further complicated by limited access to tumor tissue and the absence of suitable in-vitro and in-vivo experimental models. Also, patients with panNEC frequently exhibit vague symptoms, posing a challenge for early diagnosis. Therefore, there is a pressing need to create a detailed cellular map of panNEC, aiming to unravel its tumor biology, identify early diagnostic and novel therapeutic targets, and broaden the horizons of precision medicine in panNEC treatment.

A recent scRNA-seq investigation [333] has offered valuable insights into the anatomical subtypes and molecular heterogeneity of well-differentiated GEP-NET, along with its interaction with the immune microenvironment. While the study extensively explored the heterogeneity of lymphoid and myeloid immune cell types or states, there is limited information about the subtyping of neuroendocrine (NE) tumor [333]. Importantly, this study did not profile poorly differentiated NEC. Hence, the cause and consequences of the significant clinical heterogeneity observed in panNEC are still in the early stages of understanding, necessitating further single-cell research. While scRNA-seq has previously been instrumental in characterizing shared and distinct cellular phenotypes in several solid

tumors, applying it to the pancreas is reported to be challenging due to its high intrinsic nuclease activity [72]. Instead, snRNA-seq emerged as a promising alternative that can be utilized with frozen samples, facilitating high-resolution profiling of the tumor and immune microenvironments.

Of note, a recent study conducted by Domenico et al. in 2020 [334] proposed that panNETs has at least two origins- either from α- or β-cells within the islets of Langerhans, based on epigenetic data. Furthermore, this study revealed that relapses and metastases most frequently occurred within an intermediate group, which was found to have closer associations with α-cells rather than β-cells. Interestingly, other studies based on gene expression and enhancer analysis have also suggested the panNET origin from the same [335, 336]. Conversely, Sadanandam et al. 2015 [337] found that a subset of aggressive panNETs, referred to as "metastasis-like primary," exhibit a stemness associated phenotype demarcated by the expression of pancreatic progenitor-specific genes, in comparison to well-differentiated tumors. This suggests a potential shared origin for high-grade pancreatic NENs, whether from immature multipotent pancreatic progenitors or specialized tip, trunk, or endocrine progenitors [308]. Another seminal study by Yachida et al 2022 [338] postulated that panNEC could be categorized into two subgroups, namely "ductal-type" and "acinar-type," distinguished by their genomic and epigenomic characteristics. In the "ductal-type panNEC," transcription factors such as *SOX2, ASCL1, NKX2-1, EZH2*, and *E2F1* were consistently overexpressed, with most exhibiting loss of RB1 protein and *TP53* mutations [338]. Conversely, the "acinar-type panNEC" displayed alterations in WNT signaling (including mutations in *APC* and *CTNNB1* genes), alongside with overexpression of transcription factors like *PTF1A, GATA4, NR5A2,* and *RBPJL* dedicated for acinar lineage. However, it was observed that phenotypically distinct cells shared a common origin only when they exhibited specific genomic abnormalities. Previously, no deep phenotyping approaches utilizing single-cell or nuclei RNA-seq were conducted to systematically examine developmental trajectories and investigate transcriptomic resemblances among panNEC samples or specific sub-states, aiming to explore potential common or multiple origins.

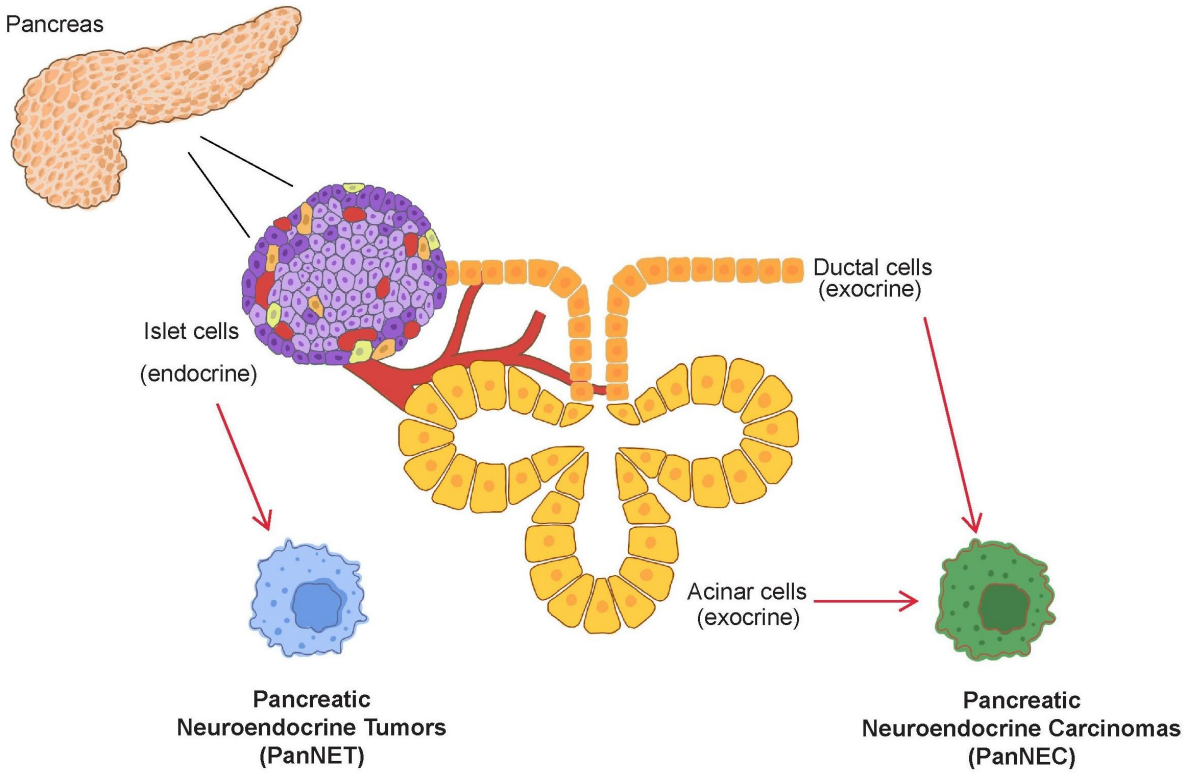


Figure 1.5: Schematic illustrating possible origins of well differentiated pancreatic neuroendocrine tumors (PanNET) and carcinoma (PanNEC).
Adapted from Domenico et. al. 2020 [334] and Yachida et. al. 2022 [338]. Illustration created using InDesign software.

2. Aims and objectives:

Within healthy tissues, a diverse array of specialized cells exists, each capable of adopting specific functional states to execute distinct functions in a coordinated fashion. The transcriptome of individual cells plays a crucial role in translating genetic and epigenetic information into observable phenotypic traits, facilitating cellular communication with the surrounding environment. Consequently, the transcriptome serves as a surrogate indicator for cellular identity in both health and perturbed situations, including diseases and upon drug treatment. The human placenta and pancreas are two strikingly different organs of endodermal origin, serving vastly different functions in the body. Nonetheless, the development of both organs involves intricate and complex processes of cellular differentiation, migration, and cellular communication that can be investigated using approaches like single cell or nuclei RNA sequencing. An explicit focus of this dissertation is on the snRNA-seq based deep phenotyping approach that was deliberately applied to different biological systems, having unconnected translational needs, and have remained unsolved in prior research works.

Recent investigations utilizing single-cell sequencing techniques have yielded valuable insights into the maternal-fetal interface, trophoblast subtypes, and endometrium within healthy female reproductive tissues, both in non-pregnant states and during the first trimester of pregnancy [146, 209]. A significant gestational time gap is unavoidable for studies comparing maternal-fetal tissues in healthy term pregnancies and those affected by eoPE. Early diagnosis of eoPE before the 12th gestational week is currently impractical, and delayed diagnosis in later stages of pregnancy results in irreversible health consequences. Therefore, a comprehensive understanding of the early patho-mechanisms underlying eoPE is crucial for timely diagnosis and improved clinical management. The challenge lies in obtaining uteroplacental tissue samples for eoPE studies, as this is associated with an elevated risk of preterm miscarriage.

The primary objective of the first segment of this thesis was to establish a reference cellular atlas of early (first-trimester) pregnancies, healthy term controls, and eoPE affected pregnancies using snRNA-seq. A key goal was to delineate cell type/state specific differential gene and transcriptomic dysregulation in eoPE in comparison to term controls to get mechanistic insights into disease pathomechanisms. For comparative analysis, term controls were selected instead of preterm controls, given that the latter group is often confounded by clinical factors such as respiratory distress syndrome, developmental delays, and long-term

cardio-metabolic complications [264-267]. Subsequently, the study aimed to investigate if a significant fraction of trophoblast development drivers were dysregulated during eoPE, by leveraging trajectory analysis. Since conducting longitudinal studies on placental tissue from the same pregnancy during pregnancy poses challenges, it is proposed that computational trajectory modelling of snRNA-seq data can partially overcome this limitation. Here, an integration of snRNA-seq data analysis with a multi-omics framework encompassing 10X Visium spatial transcriptomics and proteomics was considered to validate differentiation markers characterizing the divergent trophoblast lineages. On this note, the trophoblast syncytium was a major focus, with an emphasis on uncovering how perturbed syncytium fusion could impact the development of eoPE. Importantly, this work majorly aimed to understand how dysregulated senescence associated secretory phenotype (SASP) in placenta could possibly associate with exaggerated hallmarks of eoPE, including placental dysfunction, oxidative stress, apoptosis, and autophagy. Additionally, ISS was applied to study the localization of specific SASP mRNA(s) in eoPE relative to term controls placenta. In parallel, this thesis investigated the molecular and signalling level perturbations in the analogous immune cell types/states in decidua and villi, using snRNA-seq data. Notably, this dissertation sought to identify the potential cell(s) of origin for eoPE and specifically examined whether the disease originates autonomously in both maternal and fetal tissues or if it is translated from one to the other. Taken together, this extensive multi-omics investigation is anticipated to uncover additional biomarkers crucial to understanding the pathogenesis of eoPE and potentially pave the way for innovative treatments.

The second segment of this dissertation delved into understanding the developmental progression and maturation of healthy islets (emphasized on beta cells) by comparing neonatal and adult pancreas. To be precise, trajectory modelling of snRNA-seq data and statistical frameworks such as Generalized Additive Models (GAM) were applied to find significant gene expression changes as the neonatal cells differentiate towards adult. This approach provided insights into the previously unexplored age-dependent developmental trajectory of islets. The analyses relate to physiological processes and represent a different use-case for snRNA-seq for otherwise unresolvable knowledge gaps in developmental biology. Overall, integrating snRNA-seq data from neonatal and adult pancreases would facilitate a comprehensive understanding of age-dependent changes in human islets- mainly, the maturation of β cells and the establishment of glucose homeostasis during postnatal development. In terms of translation, the results are expected to be relevant to diabetes research and possibly to well differentiated panNET(s).

Subsequently, the primary objective of the last segment of this thesis was to comprehensively elucidate the cellular landscape of large cell pancreatic neuroendocrine carcinoma (panNEC) by addressing the gaps in understanding the fundamental mechanisms driving tumor biology. The rationale was to identify distinct malignant sub-states, and how their signatures associate with tumor growth, progression, and clinical variability. Also, this research aimed to overcome some of the challenges posed by limited access to tumor tissue and the absence of suitable experimental models, by establishing the first reference snRNA-seq based atlas for panNEC. To gain insights into lineage plasticity and possible developmental origins of panNEC sub-states, single cell and nuclei RNA-seq data from both healthy mouse and adult pancreas cell atlas were utilized [72, 339]. Specifically, adult pancreas snRNA-seq data [72] was computationally integrated. It was assumed that juxtaposing inferred panNEC sub-states with key cell markers and transcription factors specific to various healthy pancreas cell types would shed light on their molecular origins. Similarly, this thesis sought to explore the transcriptomic similarities between panNEC and NEC in the lungs [340] and prostate [341], as well as with previously decoded oncogenic programs found in PDAC [342]. A major focus was to investigate if any of the oncogenic sub-state(s) demonstrate therapeutic vulnerabilities and could be exploited to develop treatment strategies for panNEC. Through this endeavor, this work attempted to unravel novel therapeutic targets, and advance precision medicine approaches for the treatment of panNEC.

3 Results | Maternal-fetal interface in eoPE

- **Scientific acknowledgement**

For the first segment of thesis, the maternal-fetal interface and pre-eclampsia was investigated using single nuclei RNA-seq and spatial transcriptomic approaches. For this study, acquisition and 10X sequencing of human decidua and placenta samples were facilitated through collaboration with Dr. Olivia Nonn, Prof. Martin Gauster, and Dr. Florian Herse. 10X single nuclei library preparation and sequencing were performed at BIH Core Unit Genomics. Initial data pre-processing was done by Dr. Cornelius Fischer at the core facility. Daniela Sofia Valdes contributed to the biological interpretation of cell type annotation together with me and helped in curating fusiogenic gene list (related to Figure 2.23a). I also had helpful discussions about trophoblast progenitors with Daniela Sofia Valdes, and about uterine immunobiology with Andreas Ian Lackner & Jurgen Pollheimer. 10X Visium and In-situ RNA sequencing experiments were conducted by Ali Kerim Secener and Dr. Katja Sallinger, respectively. Ali Kerim Secener also contributed to the 10X Visium analysis jointly with me. In-situ RNA sequencing pre-processing, analysis and figure generation were carried out by Sebastian Tiesmeyer, under the supervision of Dr. Naveed Ishaque. I aided in cell typing in ISS analysis. The spatial proteomics experiment and analysis were undertaken by Jose Nimo and Dr. Fabian Cosica. Immunohistochemistry analysis was performed by Lena Neuper under the supervision of Prof. Berthold Huppertz. Alina Forlova performed microarray analysis. I would love to thank Dr. Olivia Nonn for productive discussions on placenta biology, and for organizing experimental validation of my key computational results in independent validation cohorts.

If not stated otherwise, I conducted all snRNA-seq data analytics, created figures, interpreted results, and wrote the content presented in this section. Collaborative data figures were selectively incorporated to reinforce my findings and validate key results.

3.1 Cohort establishment and data generation

The adopted study design involved separate sampling of 34 decidua and placenta villi samples followed by performing snRNA-seq, comprehensive computational data analytics, and multi-level validation of major data-driven findings, as illustrated in **Figure 3.1.** The study population was carefully selected (see Methods; **Supplementary Table 1**), ensuring homogeneity of recruited samples by excluding pre-eclamptic patients confounded with other maternal disease conditions. Specifically, samples associated with acute kidney injury, liver dysfunction, neurological symptoms, hemolysis, and thrombocytopenia [272] were removed. Instead, the recruitment focused exclusively on one diagnostic criterion Fetal Growth Restriction (FGR) that serves as the sole indicator for diagnosing severe fetal outcomes associated with severe eoPE. In this way, the heterogeneous confounders associated with this multi-factorial disease were reduced by solely focusing on PE placenta(s) with FGR. This aligns with the International Society for the Study of Hypertension in Pregnancy (ISSHP) [272], which considers FGR as one of the criteria for diagnosing PE in a patient who develops high blood pressure after 20 weeks of pregnancy. This is because both conditions often coexist and share underlying pathophysiological mechanisms linked to impaired placental function [343, 344]. Subsequent investigation aimed to determine whether FGR is independently linked to the PE signature or if it is characterized by an overlap with it. Analysis of placental microarray data [345] from cases of eoPE with FGR (n=18) and eoPE without FGR (n=19) did not reveal significant differences. This indicated that FGR did not introduce major confounding factors independent of eoPE (Supplementary Table 1).

The temporal progression of eoPE was examined using uteroplacental tissue samples i.e., decidua and placenta villi stemming from early controls (n=13), healthy term controls (n=10), and eoPE pregnancies (n=11) **(Figure 3.1, Supplementary Table 1).** Due to limitations in longitudinal sampling, all snRNA-seq samples were obtained after elective surgical termination in early pregnancy (at 5-10 weeks of gestation i.e., first-trimester), and during a cesarean section in term pregnancy. In cases of eoPE, cesarean delivery occurred at 27-33 weeks of gestation. Labor-related confounding factors were mitigated by recruiting cesarean section deliveries for both healthy term and eoPE groups. A concise cohort description table is shown in **Table 1**.

	First-trimester healthy		**Term control (healthy)**		**eoPE (diseased)**	
	Early villi	Early decidua	Term villi	Term decidua	eoPE villi	eoPE decidua
Gestational age at delivery (in days)	59	43	273.5	277.5	214	192
Maternal age (in years)	29.5	31	36	36	33	32
Maternal BMI (in $kg/m^{2)}$	21.007	21.8299522	24.5	24.05	26.1	23.2
Total Samples (N=34)	10	3	6	4	6	5
Male (fetal sex)	3	1	5	3	6	4
Female (fetal sex)	7	2	1	1	0	1
FGR/SGA*						
No FGR	10	3	6	4	2 (status unknown)	0
YES FGR	0	0	0	0	4	5
* Fetal growth restriction / small for gestational age						

Table 1: Cohort characteristics table

Table presenting an overview of the cohort characteristics, encompassing samples from the first trimester (early), term controls, and eoPE villi and decidua. Median values for gestational age at delivery (in days), maternal age (in years), and BMI (Body mass index; in kg/m^2) are reported. The total number of samples for early, term, and eoPE villi and decidua is summarized (N=34). Additionally, the distribution of male and female samples, based on fetal sex, is provided. Samples identified as male fetuses are categorized accordingly. Furthermore, the presence or absence of FGR/SGA status in an eoPE patient sample is indicated. FGR was determined based on pathological fetal doppler results and/or a baby percentile below 3, while the absence of FGR was defined according to the consensus set by DRT, 20190114. For detailed references, please refer to Supplementary Table 1. To address potential confounding factors related to labor, cesarean section deliveries were included for both the healthy term and eoPE groups- hence, this is not separately tabulated.

As securing placental biopsies is not feasible in ongoing healthy pregnancies, term instead of preterm controls were utilized in this study. As mentioned in previous section 1.11, healthy term was considered as a preferred control group, given that preterm deliveries are at higher risk of developing certain diseases and often confounded by super-imposed clinical conditions [272, 293, 346-352]. Hence, there exists a gap of around 6-8 gestational weeks in eoPE, as compared to term controls. To address this temporal gap and compensate for the lack of gestation age matched controls, an external scRNA-seq data from Regi et al. 2019 [353] **(n=16; Supplementary Table 1)** that profiled uteroplacental samples of non-preeclamptic, non-hypertensive preterm deliveries (before 34th week) were utilized. This integration adjusted for major cell type/state-specific signatures that could introduce bias in gene expression owing to preterm birth. **(Figure 3.1; also, section 3.7 & Methods).** Potential confounder related to labor was eliminated by only including samples from births delivered by cesarean section prior to labor.

However, the above scRNA-seq study did not profile vSTB, the largest cell type constituting the placenta. These cells have numerous nuclei but share a common cytoplasm, posing challenges for isolating individual nuclei for sequencing. Additionally, the high level of RNA degradation and the loss of nuclear integrity during cell isolation further complicate the accurate profiling of gene expression in multinucleated cells using scRNA-seq techniques. In this regard, a bulk microarray dataset [345] obtained from placental preterm and term deliveries was leveraged to account for vSTB specific preterm signatures (see Methods & Supplementary Table 1). This is because vSTB comprises of most cells in term villi, specifically, the surface area of vSTB measures approximately 12 square meters at term, compared to only 5 square meters at the 28th week of gestation [354]. More precisely, earlier studies indicated that the number of vSTB nuclei reaches approximately 58 billion at term, representing approximately 90% of the total villi [355-357]. Samples most suitable for gestational age correction were identified from that study by matching the characteristics of this study cohort: age (18 – 40), BMI (18.5 – 35, WHO normal to obese), delivery mode (caesarean section), and the absence of FGR and chronic hypertension. After excluding individuals with chronic hypertension **(Supplementary Table 1),** 10 preterm (<34 weeks) and 16 term (>37 weeks) samples were retained respectively and were considered for downstream analysis.

Subsequently, a multi-omics analytics framework encompassing spatial proteomics and 10X Visium [187] based spatial transcriptomic profiling were included for validating snRNA-seq based findings w.r.t early control placenta **(Figure 3.1).** Also, for the first time, ISS technique [185] was applied to spatially resolve and compare eoPE and term placenta samples to gain mechanistic insights about disease biology **(Figure 3.1).** In a nutshell, multi-omics with a major focus on snRNA-seq analysis was utilized to understand the longitudinal pathogenesis of eoPE.

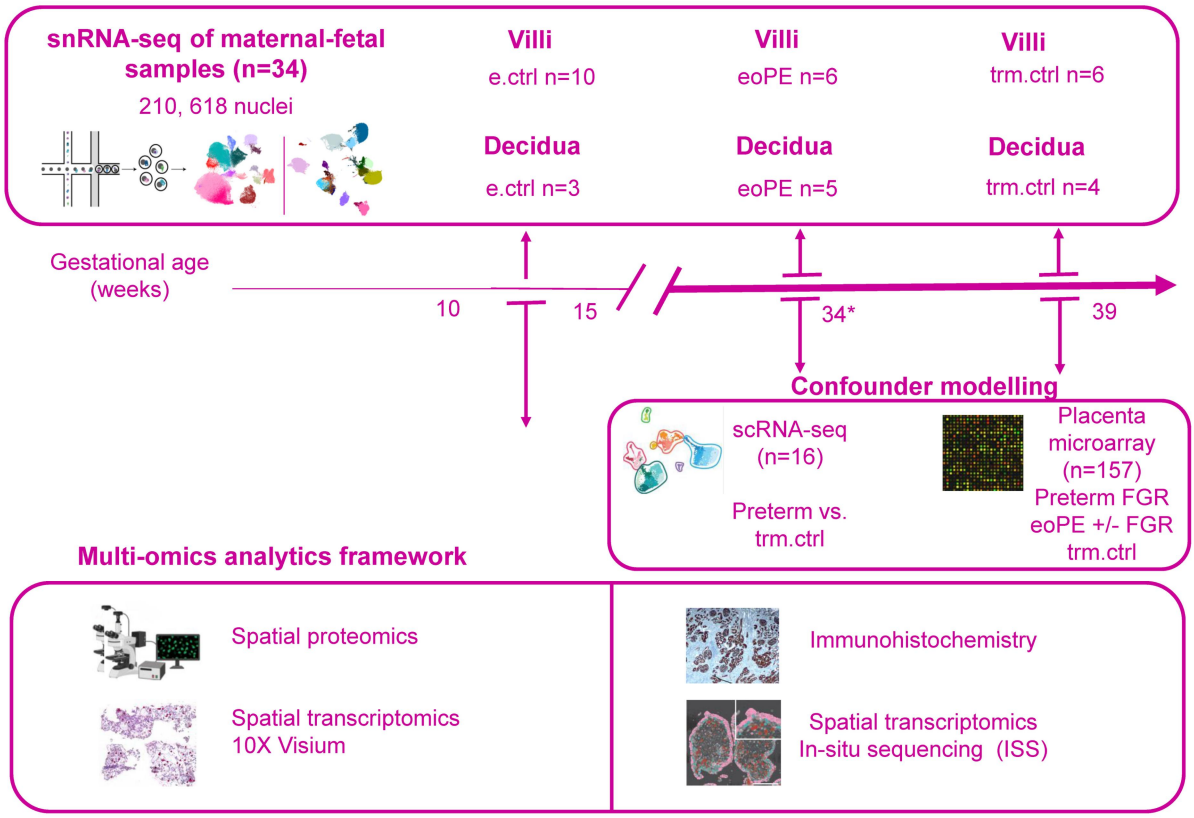


Figure 3.1: Clinical and multi-omics study design with patient cohorts.
Schematic illustration of experimental design across stages of gestation - early control or first-trimester pregnancy (e.ctrl), early onset pre-eclampsia (eoPE) versus late term control pregnancies (trm.ctrl). Uteroplacental tissues were surgically procured and separately sampled to collect villi and decidua from mostly same patients for snRNA-seq. Early controls (e.ctrl) corresponds to gestational age ranging between 5-10 weeks, early onset pre-eclampsia (eoPE) induces delivery before 34 weeks (ranged between 27-33 gestational weeks), and healthy term control (trm.ctrl) at 39 weeks (ranged between 38-40 gestational weeks). The gestational age difference between healthy term controls (trm.ctrl) and diseased pre-term eoPE were adjusted by using additional scRNA-seq data from preterm delivered non-hypertensive obstetric pathologies.

3.2 A snRNA-seq atlas of the healthy and eoPE maternal-fetal interface

In this study, a separate sampling of placental villi (fetal part) and decidua (maternal part) was performed to delineate the cellular architecture of the maternal-fetal interface, as illustrated in **Figure 3.2**. Also, the cellular origin could be distinguished through distinct sampling and processing of villous and decidual tissues **(Figure 3.2 & Figure 3.3).** It is to be noted that the multinucleated STB layer covering the surface of placental villi posed a challenge during the fluorescence-activated cell sorting (FACS) step for preparation in the scRNA-seq analysis pipeline, potentially causing fragmentation. It is known that multinucleated cells like vSTB are tedious to isolate and may exhibit functional or transcriptional heterogeneity among their nuclei [357, 358]. Thus, traditional scRNA-seq, designed for single-cell analysis, may not faithfully capture vSTB heterogeneity, as it tends to

provide an average representation of gene expression across all nuclei within a multinucleated cell. In contrast, a more comprehensive exploration of the maternal-fetal interface in contact with maternal circulation and the depiction of the diverse nuclear content of this multinucleated cell layer were facilitated by employing snRNA-seq profiling **(Figure 3.2).** Subsequently, the cellular landscape underlying the maternal-fetal interface was deciphered using 10X snRNA-seq bioinformatic analysis (conceptualized in **Figure 3.2**).

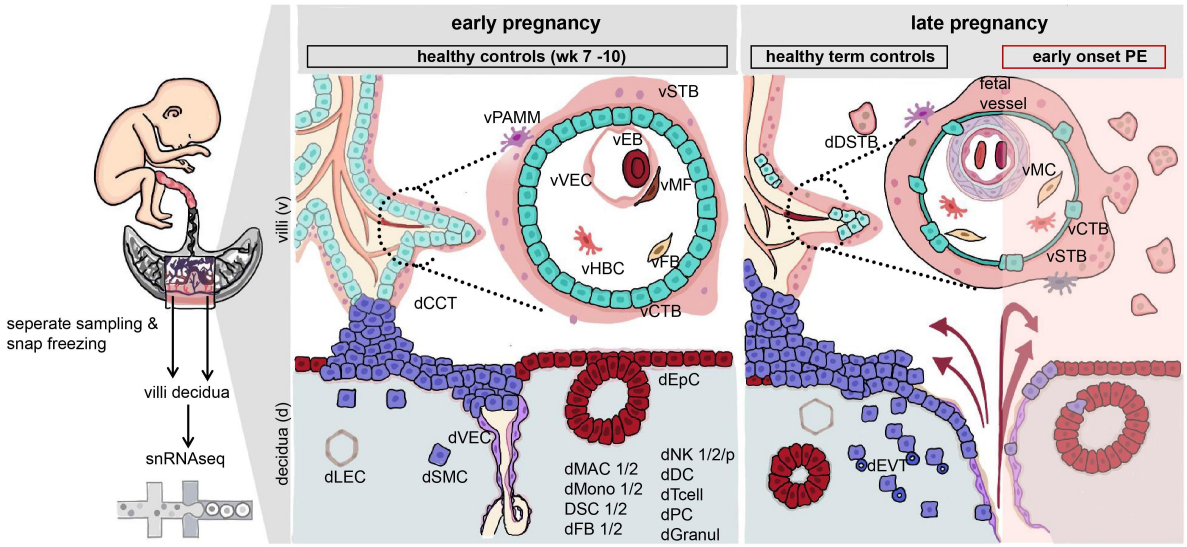


Figure 3.2:Schematic illustration of experimental design and histological changes across gestation in healthy and early onset preeclamptic (PE) pregnancies.
Placental tissue was separately sampled to collect villi and decidua from mostly the same patients for 10X single-nuclei RNA sequencing (snRNA-seq). Early tissues correspond to gestational ages between 7-10 weeks. Late pregnancy included healthy term controls between 39-40 weeks and early onset pre-eclampsia (early onset PE) to < 34 weeks. Figure adapted from Nonn, Debnath & Valdes et al. 2022 [363] (unpublished) and re-sketched using Adobe InDesign software.

In total, there were 10 early villous samples, 6 term control samples, and 5 eoPE villi samples profiled. Regarding decidua, there were 3 early samples, 4 term control samples, and 5 eoPE samples collected. Firstly, Cellranger detected a median of 50,659 sequenced reads (52,833 in villi, and 43,473 in decidua), with 91.15% median mapped reads (91.35% in villi, and 90.25% in decidua). Secondly, QC measures included the removal of cell barcodes confounded with ambient RNA(s) and random barcode swapping using CellBender [101], followed by traditional QC analysis using scanpy toolkit [98], and subsequent elimination of ambiguous clusters expressing conflicting cell type markers **(see Methods).** A concise summary tabulating quantitative QC metrics at these three levels of analysis is depicted in **Table 2.** Extensive QC per nuclei and per cell type/state were visualized and tabulated **(Extended Data Figure 1e, f; Supplementary Table 3-5)**.

	Number of nuclei recovered	Number of UMI per nuclei	Number of genes detected per nuclei
CellBender level QC statistics (before filtering cells)	6134.5 (6211 in villi, and 5534 in decidua)	1418.25 (1469.5 in villi, and 1058 in decidua)	984.5 (1026 in villi, and 834 in decidua)
Scanpy level QC statistics	5786.5 (6032.5 in villi, and 5470.5 in decidua)	1521 (1633 in villi, and 1113.5 in decidua)	1128 (1224.5 in villi, and 882.5 in decidua)
Downstream filtering	5729.5 (6026 in villi, and 5470.5 in decidua)	1516 (1669.5 in villi, and 1112.25 in decidua)	1125.25 (1254.25 in villi, and 880.75 in decidua)

Table 2: QC statistics summarizing number of nuclei recovered, number of UMI per nuclei and number of genes detected per nuclei across major levels of analysis.
QC summarized across three levels of computational analysis- i) CellBender [101] (denotes QC after adjusting for ambient RNA expression and random barcode swapping, but before downstream filtering of nuclei), ii) Scanpy [98] (filtering high quality nuclei, after filtering for %mtRNA, #UMI, #Genes) and iii) Downstream filtering (after removing donor-specific and ambiguous clusters). Only median values are tabulated.

Additionally, a detailed breakdown of QC across different gestational conditions and library chemistry (10X V3 or V2) is summarized in **Table 3.**

Gestational condition	10X V3 samples	10X V2 samples

Early villi	n=7
	TC= 7035
	#genes= 3088
	N= 67037
		n=3
		TC= 1867
		#genes= 1465
		N= 12848

Early decidua	Not added	n=3
		TC= 1661
		#genes= 1248
		N=15582

Late term control villi	Not added	n=6
		TC= 1157
		#genes= 851
		N= 29702

Late term control decidua	n=1
	TC= 3374
	#genes= 1942
	N= 6554
		n=3
		TC= 801
		#genes= 661
		N= 14649

eoPE villi	n=2
	TC= 4707
	#genes= 1892
	N= 12365
		n=3
		TC= 1123
		#genes= 790
		N= 23685

eoPE decidua	n=2
	TC= 3233
	#genes= 1887
	N= 15405
		n=3
		TC= 898
		#genes= 731
		N= 13910

Table 3: Quantitative QC analysis presenting nuclei recovery, UMI, and gene detection metrics across early, late term, and eoPE villi/decidua, stratified by 10X V3 and 10X V2 library chemistry.
Only median values are tabulated. TC= total UMI counts per nuclei; #genes= the number of genes with at least 1 count in a nucleus. and N= #nuclei per condition; n sample size. The median

values for TC and #genes are noted. 10X V3 detected a substantially higher number of UMI and genes than 10X V2 for which 'library chemistry' was considered as a covariate during data integration and downstream cell type/state markers analysis (see Methods).

Ultimately, snRNA-seq data analysis unveiled 24 and 15 cell types/states in the decidua and villi, respectively encompassing immune, vascular-endothelial, matrisome, and trophoblast compartments **(Figure 3.3 a-b).** This attributed to a total of 210,618 sequenced nuclei in both tissues, after removing low-quality and confounded cells. Precisely, 95,253 nuclei originated from the early controls (e.ctrl), 64,850 from late term controls (trm.ctrl) and 50,515 from severe eoPE pregnancies **(Figure 3.3 c-d; Supplementary Table 2-4).** For data harmonization of samples stemming from different gestational time points (i.e., early, term controls, and severe eoPE), scVI/scANVI [112, 113] deep generative model-based tool was applied. Individual donor ID served as a key batch, and additional covariates such as sampling site (procurement center of the sample) and gestational condition were encoded in the VAE model to minimize technical and batch-specific effects in the downstream analysis **(see Methods).** The aim was to achieve optimal clusters, i.e., cell types or states segregated by biological variability, as extensively discussed in **"Section 3.6 scVI data harmonization and evaluation of performance."** Of note, this complex integration and clustering was separately performed for decidua and villi to delineate tissue specific cell types or states (see Methods). Robust and specific markers characterizing decidual and villi cell types/states were computed using the Logistic Regression framework [360, 361], out of which a few key markers were visualized **(Figure 3.4, Figure 3.5, Supplementary Table 8),** which are further elaborated in **sub-sections 3.2.1 & 3.2.2**, respectively.

Variations in cell composition within the immune, vascular-endothelial, matrisome, and trophoblast compartments were observed at different gestational time points, reflecting specific physiological and functional adaptations at different stages of pregnancy **(Figure 3.3 c-d; Extended Data Figure 1c, d & g; Supplementary Table 2).** Statistically significant shifts in cell type/state composition were noted between early and term control pregnancies **(Extended Data Figure 1g).** For instance, in early placenta samples, there were statistically higher levels of vCTB and vCTBp (FDR < 0.01, two-tailed Wilcoxon rank sum test [362] corrected for multiple states testing). This agrees with the known biology of first-trimester villi [146, 209] where vCTBp represents mitotically active trophoblast progenitors whose active proliferation is required for the expansion and growth of a functional placenta [214, 219,359]. Additionally, various cell types/states in villi, such as vFB and vHBC also showed differential abundance in early samples compared to term controls (FDR < 0.01, two-tailed Wilcoxon rank sum test [362] corrected for multiple states testing) **(see Extended Data Figure 1g).**

Nevertheless, the limited number of samples in the late term and eoPE groups posed challenges for statistical analysis of cell type/state composition.

In the villi, the syncytiotrophoblast (vSTB) was found to exist in various sub-states, namely vSTB1, vSTB2 and vSTBjuv (illustrated in **Figure 3.3b**) and constituted the predominant cell type of villi. Of note, 69.4% of villous nuclei in early pregnancy, 89.2% in healthy term pregnancies, and 92.7% in severe eoPE belonged to trophoblast cell-types forming a barrier separating maternal and fetal circulation **(Figure 3.3d).** Importantly, both vSTB1 and vSTB2 were found to be differentially abundant in term controls relative to early (FDR < 0.01, two-tailed Wilcoxon rank sum test [362] corrected for multiple states testing) **(Extended Data Figure 1g).** This is to support the continuous growth and maturation of the placenta to accommodate the increasing needs of the developing fetus and to ensure efficient maternal-fetal exchange of nutrients and gases [355-357]. Importantly, nuclei of fetal origin expressing STB and extravillous trophoblasts (EVT) profiles were identified in decidua and termed as dDSTB and dEVT, respectively **(Figure 3.3a).**

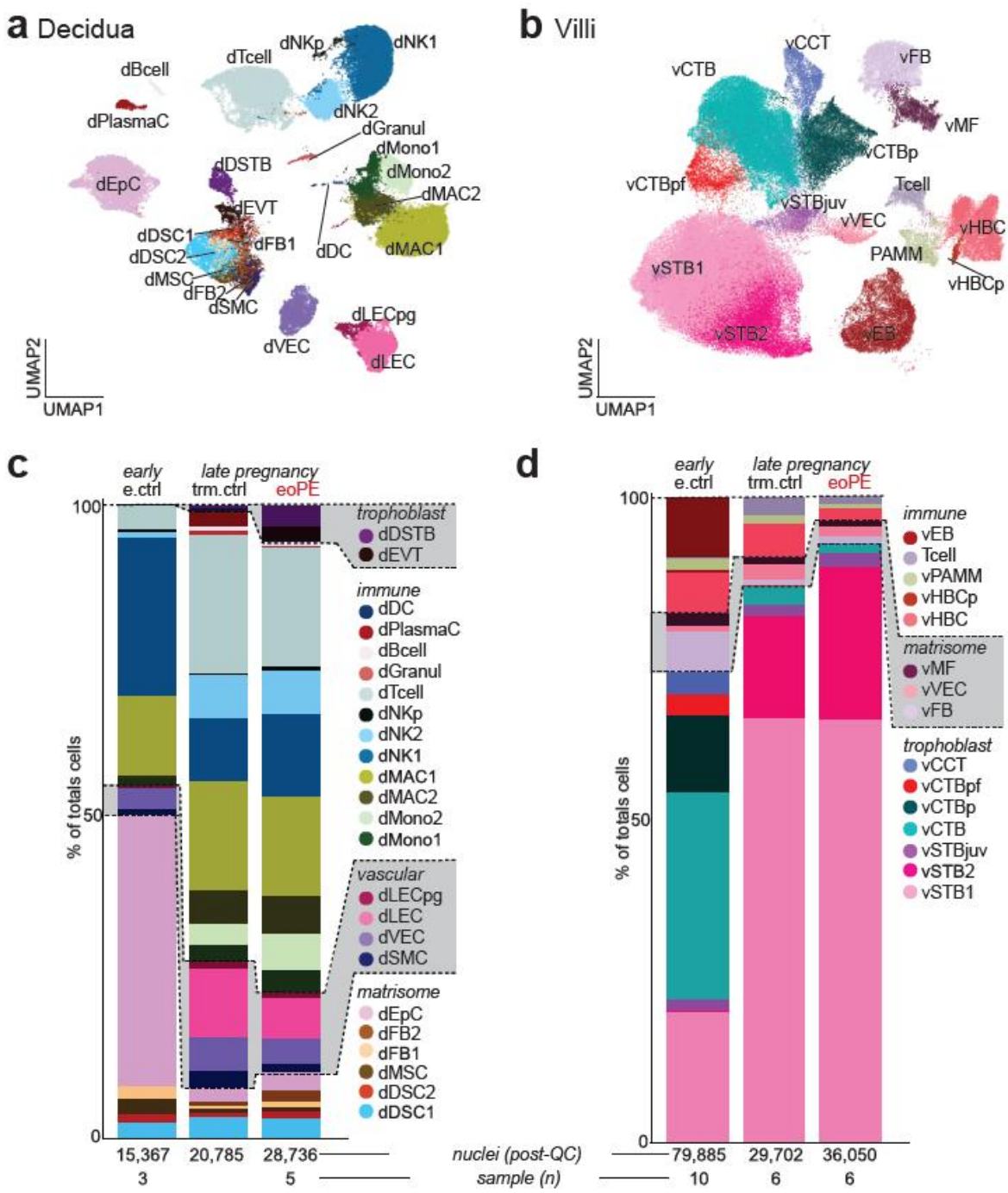


Figure 3.3: Cellular landscape of maternal-fetal interface in healthy controls and early-onset pre-eclampsia.

a) Decidual (maternal; n=12 samples) and b) placenta villi (fetal; n=22 samples) cell types/states visualized using UMAP. Each dot represents a nuclear transcriptome of decidua and villi in (a) and (b) respectively, colored by the cell type or state. The samples within decidua and placenta were separately integrated using scVI/scANVI (single-cell variational inference) across all gestational time-points to allow comparative analysis. (c,d) Cell composition (%) distribution per cell type of state, numbers under bars indicate the sample size of sequenced nuclei. Compositions presented across gestational time points (e.ctrl, trm.ctrl) and additionally for a disease state (eoPE) in late pregnancy for (c) decidua and (d) placenta villi. e). Cell name abbreviations for annotated cell types and states presented in this study. CCT, cell column trophoblast; CTB, villous cytotrophoblast; CTBpf, cytotrophoblast post-fusion; STB, syncytiotrophoblast; DSTB, deported STB; VEC, vascular

endothelial cell; LEC, lymphatic endothelial cell; LECP, LEC progenitor; SMC, smooth muscle cell; MC, myocyte; FB, fibroblast; EpC, epithelial cell; MSC, mesenchymal stem cell; DSC, decidual stromal cell; EB, erythroblast; HBC, Hofbauer cell; PAMM, placenta-associated maternal macrophage; Mono, monocyte; MAC1, M1-like macrophage; MAC2, M2-like macrophage; NK, natural killer cell; PC, plasma cell; DC, dendritic cell, Granul, granulocyte; v, villous/placental; d, decidual; p, proliferative; juv, juvenile.

3.2.1 Identification of decidual cell types and states

The decidual cell types/states were deconstructed using key markers extracted from Logistic Regression based GLM test [360, 361] **(Figure 3.4; Supplementary Table 7).** Additionally, the cell type or state associated markers were also verified using the Negative Binomial framework. In the mesenchymal compartment, progenitor-like dMSC and two sub-states of decidua stromal cell, namely, DSC1 and DSC2 were detected. dDSC1 exhibited high expression of markers such as *ABI3BP, SYT1, SCARA5, RGCC*, and *SIPA1L2*, whereas dDSC2 expressed robust *IGFBP1, IGFBP2, LUM* and classical decidual marker prolactin (*PRL*) genes **(Figure 3.4)**. A previous scRNA-seq study by Vento-Tormo et. al. 2018 [209] revealed that decidua compacta, a site of initial EVT invasion, is significantly enriched in *IGFBP1*+ stromal cells. Also, two distinct sets of decidual fibroblasts- dFB1 and dFB2 were identified that share key markers such as *COL1A1, COL1A2* and *COL3A1* **(Figure 3.4)**. Additionally, dFB2 noted specific expressions of genes such as *EDIL3, TNC, SLIT2*, and *SULF1* **(Figure 3.4)** associated with cell adhesion, migration, and ECM remodelling [364-367]. Interestingly, dFB2 was detected only in eoPE and term conditions, not in early samples **(Figure 3.3c, Extended Figure 1d & Supplementary Table 2).** Of note, dSMC showed robust expression of *GUCY1A2, ADGRL3, ACTA2*, and specific *RGS5* expression **(Figure 3.4)**. Activation of *GUCY1A2* (guanylate cyclase) and subsequent cGMP production contribute to vasodilation and relaxation of smooth muscle cells [368]. In smooth muscle, *RGS5* has been implicated in regulating vasoconstriction and blood vessel tone by modulating signaling pathways involved in the response to vasoactive agents [369, 370].

Importantly, the cell types/states underlying decidual lymphoid and myeloid immune lineage were characterized using previously demonstrated mutually exclusive key markers **(Figure 3.4)**. Decidual Natural Killer cell type (dNK) was found to exist in three sub-states, namely, dNK1, dNK2 and dNKp **(Figure 3.3a & 3.4)**. As expected, all these three states exhibited prominent expression of *CD96* - an immune checkpoint cell surface receptor contributing to NK cell activation and cytotoxicity [371, 372] **(Figure 3.4)**. Another marker, *NCAM1*, or *CD56* [373], is a neural cell adhesion molecule having an established role in NK cell activation and

cell adhesion. Moreover, *NCAM1 or CD56* expression is known to be associated with different uterine NK cell subsets, where *CD56bright* NK cells are considered more immunoregulatory and *CD56dim* NK cells are more cytotoxic [223, 374-376]. Two other marker genes included *KLRD1* (also known as *CD94*) and *KLRC1* (also known as *NKG2A*) that are known cell surface receptors playing crucial roles regulating functions of dNK, such as, immune tolerance and immune surveillance [377]. In contrast to dNK1, the dNK2 sub-state exhibited low but specific expression of *CD160* and *ITGAE* **(dNK2; Figure 3.4)**. *CD160* is a glycoprotein expressed on the surface of NK cells that can act as both co-stimulatory and co-inhibitory receptor, depending on the context and the ligands it interacts with [378, 379]. *ITGAE* (or, *CD103*) is known to be involved in promoting the retention of NK cells within epithelial tissues, where it interacts with E-cadherin and facilitates local immune surveillance and responses to infections or tumors [380-382]. On this note, dNKp is a proliferating sub-state expressing *MKI67, TOP2A, POLQ* and *CENPF,* while sharing other dNK1/2 markers **(dNKp; Figure 3.4)**. dTcells were characterized using key expressions of *THEMIS, CD3D, CD2, CD8A, IL7R, CCL5* and *CAMK4* **(dTcells; Figure 3.4)**. *THEMIS* is known for its role in T-cell development and activation [383, 384]. *CD3D* is a component of the T-cell receptor (TCR) complex and plays a central role in signal transduction during T-cell activation [385]. *CD2* is a cell adhesion molecule and co-stimulatory receptor involved in T-cell activation, adhesion to antigen-presenting cells, and the formation of immunological synapses [386]. Previous literature suggests that *CD8+* dTcells exhibit atypical cytokine profile that is possibly linked to the local effects of progesterone [387]. Also, maternal–fetal tolerance may be favored by *CD8+* dTcells, and the decidual microenvironment promotes the residency of *CD8+* T cells to balance between tolerance and defense.

Furthermore, dMono1 and dMono2 constituted decidual monocyte sub-states characterized by distinct set of markers. dMono1 exhibited specific expressions of *CTSS, CD300E, FCN1, LYZ, PRAM1*, relatively robust *LST1* while also expressing *CD14 and MS4A6A* **(dMono1; Figure 3.4)**. This might indicate the presence of classical monocytes [388]. On the other hand, expression of *S100A9, S100A8, AQP9, CSF3R, FCGR3B*, and *TNFRSF10C* in dMono2 indicate it to be neutrophil-like monocytes [389] **(dMono2; Figure 3.4).** Of note, neutrophil-like monocytes might contribute to the immune surveillance within the decidua, actively patrolling the tissue to detect and eliminate potentially harmful microorganisms or abnormal cells [390]. Also, they are expected to play a role in regulating local inflammatory responses, and tissue remodelling. Of note, AQP9 (Aquaporin 9) [391] encodes a water channel protein whose expression might indicate metabolic adaptations required to adjust with decidual immune responses. A former study by Shi et al 2022 [392] showed compromised monocyte chemotaxis in *AQP9*$^{-/-}$ mice. Importantly, dMAC1 sub-state was

characterized using key markers such as CD14, *MRC1 (or, CD206), SPP1, CD163, F13A1, RBPJ,* and *SELENOP* **(dMAC1; Figure 3.4).** A previous study by Vondra et al 2023 [393] performed flow cytometry of CD14+ dMAC cells and found two subpopulations- *CD163+CD206+* macrophage-like and *CD163-CD206-* monocyte-like. Of note, dMAC1 closely resembled their *CD14+ CD163+ CD206+* cells. Next, dDC was characterized using well-known dendritic cell markers, like *HLA-DRA, HLA-DRB1* that encode major histocompatibility complex class II (MHC-II) molecules*, IRF8* and *XCR1* and absence of other immune lineage genes **(dDC; Figure 3.4)**. A small subgroup of cells (n=203) expressing *MS4A1, BANK1, BACH2*, and *FCRL2* were found and annotated as dBcells **(dBcells; Figure 3.4).**

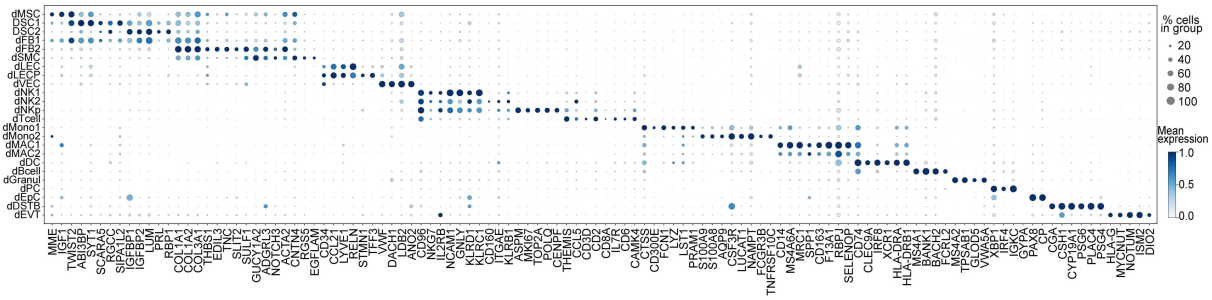


Figure 3.4: Known and novel cell types and states of maternal decidua characterised by discrete and functionally relevant markers.
Dot-plots featuring known and specific novel markers characterising each cell type or state in decidua as computed by Logistic Regression based Generalized Linear model (Bonferroni adjusted two-sided, P < 0.05 and log2FC cut-off of ±0.25). Genes are scaled across clusters. dDSTB, deported STB; EVT, extravillous trophoblast; VEC, vascular endothelial cell; LEC, lymphatic endothelial cell; LECP, LEC progenitor; SMC, smooth muscle cell; MC, myocyte; FB, fibroblast; EpC, epithelial cell; MSC, mesenchymal stem cell; DSC, decidual stromal cell; EB, erythroblast; Mono, monocyte; MAC1, M1-like macrophage; MAC2, M2-like macrophage; NK, natural killer cell; PC, plasma cell; DC, dendritic cell, Granul, granulocyte; d, decidual; p, proliferative.

3.2.2 Identification of placental cell types and states

Known and novel cell types/states underlying the human placenta were annotated using literature resources and Logistic Regression-inferred markers [360, 361] **(Figure 3.5; Supplementary Table 7).** On a computational note, the cell type or state associated markers were also verified using the Negative Binomial framework.

vCTB cell type exhibited robust expression markers, including *YAP1, LGR5, TP63, PEG10, PARP1*, and *FRAS1*, aligning with the known biology of vCTB [146, 209]. *YAP1* [394], a

transcriptional co-activator, plays a crucial role in regulating cell proliferation and apoptosis in CTB(s). *LGR5* [292, 395, 396] may be involved in regulating the balance between self-renewal and differentiation, contributing to the dynamic process of CTB differentiation. *TP63* is also indicative of vCTB stemness [395] and is a member of the p53 family of transcription factors, likely contributes to cell fate decisions, differentiation, and the establishment of epithelial integrity [397]. On the other hand, *PEG10* is a well-known CTB marker that is an imprinted gene involved in cell proliferation and differentiation [398]. *PARP1* is an enzyme involved in DNA repair and likely contributes to the maintenance of genomic integrity, ensuring proper cell function and response to stressors during trophoblast differentiation [399]. *FRAS1* is a component of the extracellular matrix [400] and might be involved in the regulation of cell adhesion, migration, and tissue architecture during trophoblast differentiation. Importantly, an actively proliferating sub-state of CTB, termed as vCTBp, which is believed to represent trophoblast progenitors, was detected. vCTBp expresses markers associated with cell-cycle regulation and G2M checkpoint kinases such as *MKI67, TOP2A, POLQ, CENPE,* and *CENPK* **(vCTBp; Figure 3.5).** Apart from this, vCTBp expresses major characteristic markers of vCTB, as mentioned above.

An intermediate trophoblast sub-state specifically expressing fusiogenic markers such as *GREM2, ERVFRD-1, ERVV-1/2, OTUB2, ABTB2*, and *DYSF,* that recapitulated the transcriptomic profile previously described by Liu et al. 2018 [401] as *ERVFRD-1* positive CTB cells was identified. This sub-state was regarded as vCTB pre-fusion (vCTBpf) **(vCTBpf; Figure 3.5).** During early human placentation, the mononuclear layer of cytotrophoblasts fuses to establish multinucleated syncytia responsible for hormonal production and nutrient exchange between the mother and the child. Furthermore, all three vSTB states express markers related to hormone secretion, such as *CSH1, CSH2*, and *CGA* [231-234], into the maternal circulation, playing a crucial role in maintaining and modulating pregnancy **(vSTBjuv/1/2; Figure 3.5).**

Also, a cell-column trophoblast (CCT) cell type was detected using well-known markers like *HLA-G, MYCNUT, NOTUM, DIO2, LAIR2,* and *ASCL2* **(vCCT; Figure 3.5).** *HLA-G* and *LAIR2* are both implicated in immune tolerance, contributing to the immunosuppressive environment at the maternal-fetal interface, facilitating successful pregnancy [402-404]. *MYCN* upstream transcript (MYCNUT) and Achaete-scute homolog 2 (*ASCL2*) are transcription factors known for regulating trophoblast differentiation and invasion. A detailed description of vCCT and the role of different markers in regulating its functional heterogeneity is provided in “Section 2.6 Subcluster analysis of invasive-phenotype cell column cytotrophoblast cell type.”

In terms of immune cell states, placental F13A1+/FGF13+ resident macrophages, known as Hofbauer cells (vHBC), which uniquely express the hyaluronan receptor *LYVE1* in the immune cell subset, were identified [237-240]. These cells are suggested to maintain arterial tone and possess pro-angiogenic functions. They are HLA-DR (-) but express markers such as *MRC1, SPP1, VSIG4, CD36, RBPJ,* and *MAF,* revealing their transcriptomic similarity to decidual macrophages [238] **(vHBC; Figure 3.5).** Interestingly, a highly proliferative sub-state of vHBC characterized by *MKI67, TOP2A, CENPE,* and potentially denoting progenitors of this immune cell type was detected **(Figure 3.5; Supplementary Table 7)**. Together with vHBC cells, vHBCp possibly contributes to maintaining immune tolerance, production of various cytokines and growth factors, angiogenesis, and interaction with trophoblasts [237-240]. Additionally, antigen-presenting *HLA-DRA+* placenta-associated maternal monocytes/macrophages (vPAMM), which are villi-associated and of extra-embryonic or maternal origin, were identified. The annotation and transcriptomic profile of this sub-state agrees with Thomas et al 2020 that characterized PAMMs using flow cytometry [238]. Villi fibroblast (vFB) was characterized using key fibroblast markers, such as *COL1A1, COL1A2, COL3A1, CDH11*, *DCN,* and *SOX5* **(vFB; Figure 3.5)**. Villi myocyte (vMC) was identified based on their *GUCY1A2, AGTR1* expression **(vMC; Figure 3.5).** Importantly, vVEC was characterized using key vascular endothelial cell markers, including *CD34, MEOX2, PECAM1, LDB2* and *DACH1* **(vVEC; Figure 3.5).**

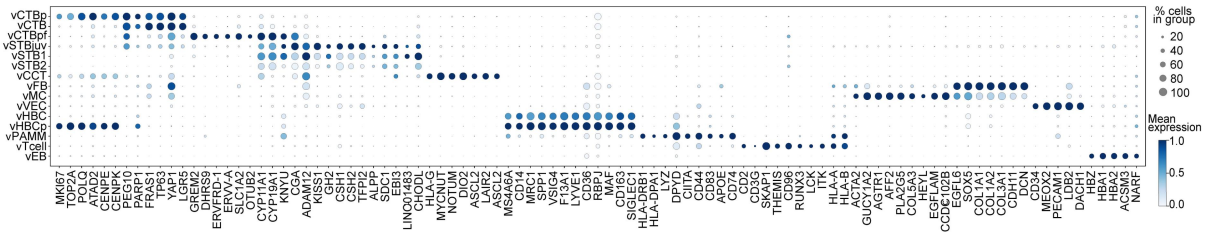


Figure 3.5: Known and novel cell types and states of placenta villi characterised by discrete and functionally relevant markers.

Dot-plots featuring known and specific novel markers characterising each cell type or state in the villi computed by Logistic Regression based Generalized Linear model (Bonferroni adjusted two-sided, P < 0.05 and log2FC cut-off of ±0.25). Genes are scaled across clusters. Placental F13A1+/FGF13+ resident macrophages (Hofbauer cells, vHBC) uniquely express hyaluronan receptor LYVE1 in the immune cell subset, suggested to maintain arterial tone and have pro-angiogenic functions [,]. Additionally, antigen presenting HLA-DRA+ placenta associated maternal monocytes/macrophages (vPAMM) were identified as villi-associated and of extra-embryonic (or maternal) origin, respectively []. Villi myocytes were identified through their expression of AGTR1. CTB, villous cytotrophoblast; STB, syncytiotrophoblast; DSTB, deported STB; CCT, cell column trophoblast; EVT, extravillous trophoblast; VEC, vascular endothelial cell; MC, myocyte; FB, fibroblast; HBC, Hofbauer cell; PAMM, placenta-associated maternal macrophage; v, villous; d, decidual; p, proliferative; juv, juvenile; pf, pre-fusion.

3.2.3 Global transcriptomic similarity among clusters

A hierarchical clustering of cell types/states and Pearson correlation analysis was conducted to infer the similarities in the global transcriptomic profile among the clusters identified in the decidua and villi **(Figure 3.6)**. As anticipated, vSTB1, vSTB2, and dDSTB were clustered together, reflecting their conserved expression of key STB signatures and overall STB phenotype. Additionally, vPAMM closely clustered with decidual macrophages, i.e., dMAC1 and dMAC2, as the former represents maternal-origin macrophages infiltrated into villi to support the growing placenta. vHBC- being functionally analogous to decidual macrophages also revealed high positive Pearson correlation scores between dMAC1 and dMAC2.

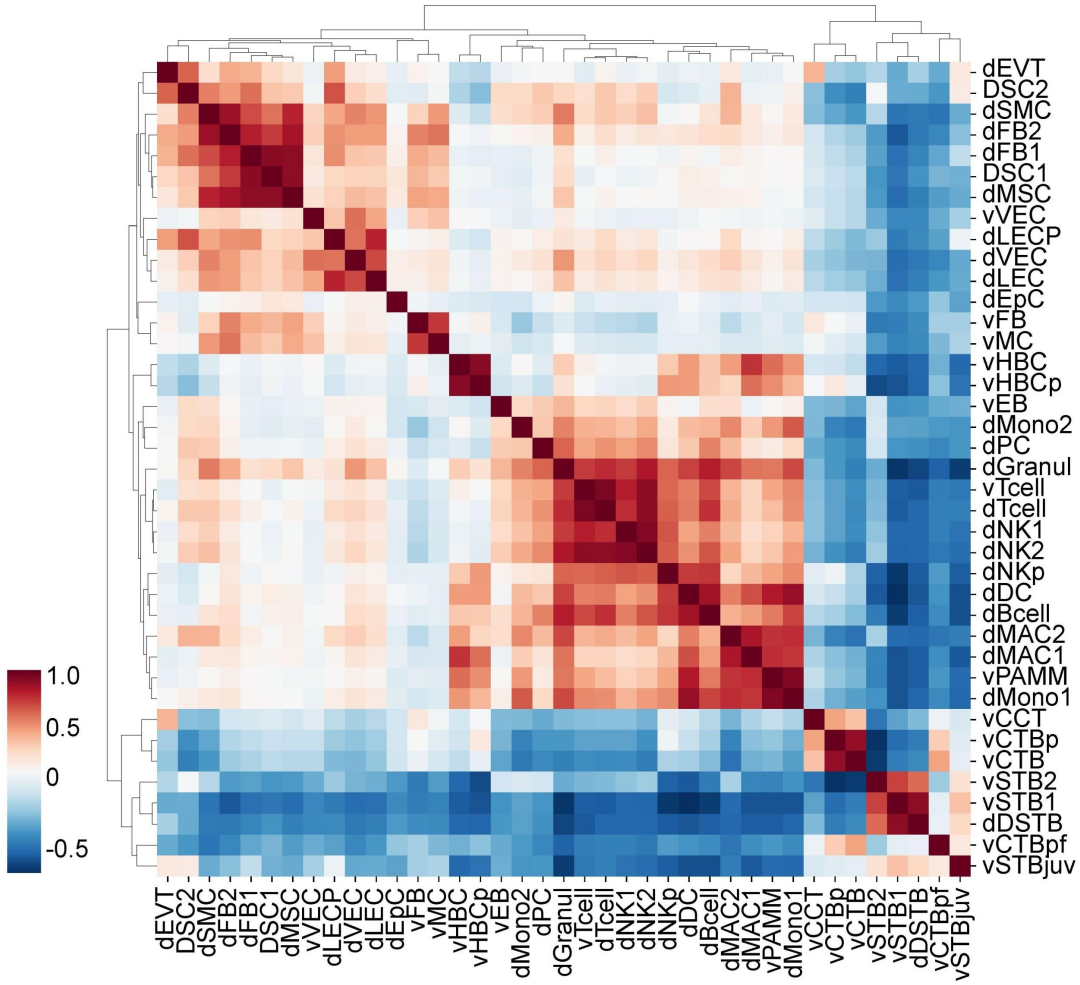


Figure 3.6: Global transcriptomic similarities among decidua and villi clusters.
Pearson correlation map of the highly variable genes (n = 6000) for identified cell-types and states. The clear uniqueness shared by STB groups also reveals clear differences in transcriptome between the novel vSTBjuv nuclei state and other STB nuclei subgroups. Because traditional characterization markers (such as CGA, CYP19A1, KISS1) were used to describe STB are fulfilled by all groups, this suggests that vSTBjuv has additional unknown functions in pregnancy. Expression data of early late control and preeclampsia samples is shown.

3.2.4 Cell types/states of maternal and fetal origin

Additionally, the examination of the maternal and fetal origin of cell types and states in both decidua and villi was conducted through the analysis of chromosome X- and Y-linked gene expression in cell clusters of samples from male fetuses and their mothers. As anticipated, the villi clusters originating from male fetuses solely exhibited expression of fetal Y-linked genes such as *TTTY14* and *USP9Y.* In contrast, X Inactive Specific Transcript (*XIST*) was specifically expressed in vPAMM and vTcell in villi of male fetuses, indicating that these cells are of maternal origin, and either infiltrated into the fetal tissue or attached to the villous surface from maternal circulation **(Figure 3.7).** On similar note, the fetal origin of dDSTB was confirmed by the expression of Y-specific genes and negligible *XIST*, as shown in **Figure 3.7.**

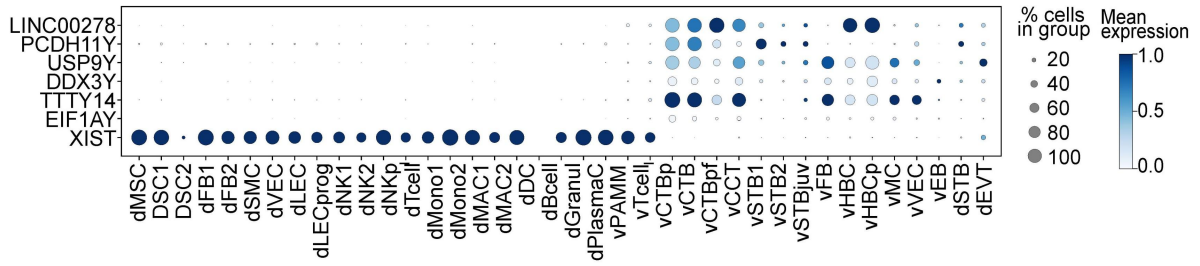


Figure 3.7: Cell types and states of maternal and fetal origin.
Dotplot featuring gene expression of key sex-associated genes on male fetus placenta samples only. Decidual cell groups are the only ones with high expression of XIST, suggesting decidual samples are of maternal orgin. XIST expression in the villous derived T cell & placenta-associated maternal macrophage (PAMM) groups suggest these are also maternal in origin and captured in the villi as invaded or adhered cells.

3.3 Functional heterogeneity of vSTB1 and vSTB2

snRNA-seq analysis demonstrated notable transcriptional heterogeneity among vSTB nuclei and supported segregating vSTB into three distinct sub-states (vSTB1, vSTB2, vSTBjuv), all characterized by a pronounced secretory capacity and different gene enrichment profiles **(Figure 3.3b & 3.5a).** To investigate functional differences between vSTB1 and vSTB2 sub-states, a multi-variate Logistic Regression test was conducted (average log2FC >=0.25, Bonferonni adjusted p-value < 0.01) **(Supplementary Table 7).** vSTB1 was found to be the most prevalent nuclear state within the syncytial nuclei and displayed a transcriptomic profile associated with typical vSTB functions, including hormone secretion (e.g., *CSH1, CSHL1, CSH2*), estrogen biosynthesis (*CYP19A1*), and protein secretion (*ADAM12, PAPPA,*

PAPPA2). Notably, robust, and specific expression of *KYNU*, a gene known for regulating placental NAD+ synthesis and the supply of fetal tryptophan and kynurenine metabolites [405] **(Figure 3.8a),** was observed in vSTB1.

In contrast, lower expression of these markers was observed in vSTB2 nuclei, implicating partial transcriptional inactivation of traditional vSTB programs. These nuclei were characterized by higher expression of *PDE4D*, an enzyme involved in cAMP and cGMP regulation **(Figure 3.8a).** Previously, *PDE4D* was described to induce intra-uterine inflammation in gestational tissues and preterm labor [406]. Of note, selective inhibitors against PDE4 are currently in clinical trials for the disease treatment related to inflammatory disorders [407]. vSTB2 also expresses *TENT5A*, which is implicated in mRNA stabilization and potential response to infection [408] **(Figure 3.8a).** Unique genes upregulated in vSTB2 are associated with pathways related to ubiquitin-mediated proteolysis and suppression of genes involved in the cell cycle. Additionally, vSTB2 exhibited repression of genes associated with DNA damage response (e.g., *DDX5, DDX17, DDX39B*) and cyclin-dependent kinases (e.g., *CDK19*), suggesting that vSTB2 might be a senescent STB sub-state **(Figure 3.8a)** and possibly represents the terminal fate in the differentiation trajectory. This could be further supported by the downregulation of fusion genes in STB2, including *LIFR, GCM1*, and *DYSF* **(Figure 3.8a; Supplementary Table 7).**

To further understand the functional relevance of vSTB1 and vSTB2, Metascape [409] analysis was performed to identify enriched pathways and signaling programs using the top 100 markers for these two nuclear states **(Figure 3.8b).** Both vSTB1 and vSTB2 exhibited enrichment in cell surface interactions at the vascular wall (R-HSA-202733, q-value<0.000001) and NABA Matrisome Associated (M5885, q-value=0.000007). Additionally, vSTB1 marker genes were uniquely enriched for processes related to steroid hormone metabolism (R-HSA-196071, q-value=0.000008), the JAK-STAT signaling pathway (hsa04630, q-value=0.006642), and the PID integrin A9B1 pathway (M118, q-value=0.022035). On the other hand, the top 100 marker genes for vSTB2 were uniquely enriched in estrogen-dependent nuclear events downstream of ESR-membrane signaling (R-HSA-9634638, q-value=0.009465) and the PID ARF6 pathway (M86, q-value=0.047032), as detailed in Supplementary Table13. Furthermore, for both vSTB1 and vSTB2, markers showed enrichment in transcription factors and regulators, including *TFAP2A* (q-value=0.00004), *ESR1* (q-value=0.006.31), *SP1* (q-value=0.015849), and *CEBPB* (q-value=0.019953). Presence of these shared transcription factors in both vSTB1 and vSTB2 possibly indicate ongoing hormonal activity in these nuclei states. Of note, *TFAP2A* is a TF known to initiate early specification of trophoblast progenitors by placental genes activation

and by supressing pluripotency regulator *OCT4* [410], and further for regulating terminal differentiation of vSTB [411]. A previous study detected *ESR1* protein in the nuclei of cultured human vSTB cells [412]. In another investigation by O'Neil and colleagues [413], estrogen was shown to activate the *LEP* promoter in choriocarcinoma JEG-3 cells through *ESR1*. This suggests that the regulation of leptin biosynthesis might rely on the presence of a functional *ESR*.

Next, vSTB1 marker genes were found to be uniquely enriched for the transcription factors *CREB1* (q-value=0.001995), *MYBL2* (q-value=0.00631), *ATF1* (q-value=0.019953), and *NR5A1* (q-value=0.031623). In contrast, the top 100 vSTB2 marker genes did not exhibit unique enrichment for any transcriptional regulator, as outlined in **Supplementary Table 7**. The enriched transcription factors *CREB1, MYBL2, ATF1,* and *NR5A1* in vSTB1 markers are associated with the EGF/EGFR-pathway. This suggests EGFR signaling, a pathway commonly linked to vCTB continues to regulate vSTB1. Also, *NR5A1* (or, *SF1*) is known to regulate steroidogenesis [414]- a process essential for maintaining a supportive hormonal environment during pregnancy.

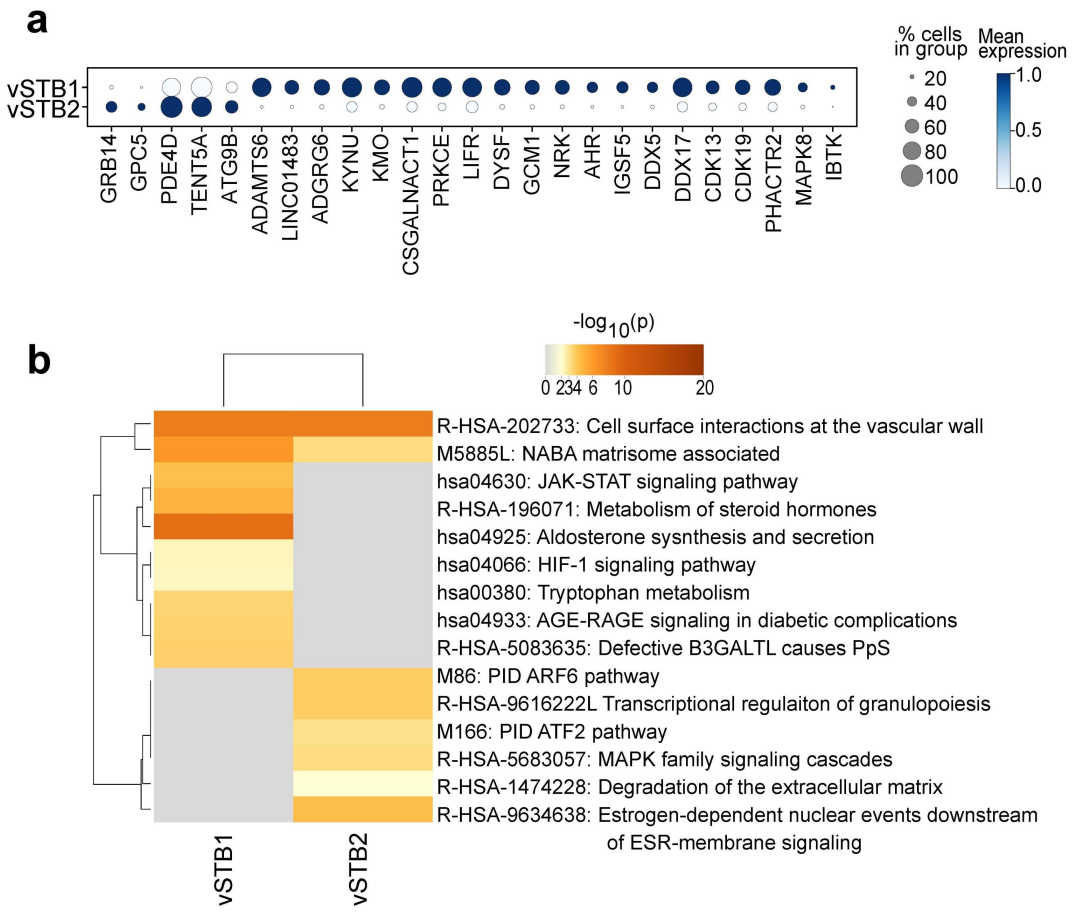


Figure 3.8: Functional heterogeneity between vSTB1 and vSTB2 nuclear sub-states.

(a) Dotplot featuring gene expression of markers characterising the two predominant nuclear states in the syncytial layer (STB). (b) Enrichment heatmap displaying top shared and unique pathways comparing vSTB1 and vSTB2 sub-states, using a discrete color-scale to represent statistical significance (-log$_{10}$(p)). Gray color indicates a lack of significance.

3.4 Validation of decidua-derived STB sub-state

Decidua-derived dDSTB nuclei of fetal origin were characterized by robust expression of traditional STB markers such as *CGA, CSH1, CYP19A1*, and *PSG4*, but not EVT lineage marker *HLA-G* (**Figure 3.4**). This finding was validated by localizing the dDSTB population in maternal blood by immunohistochemistry (IHC) in decidual sections **(Figure 3.9).** Specifically, IHC robustly stained GDF15, and β -hCG (encoding *CSH1*), but was negative for HLA-G. However, it remains to be elucidated whether dDSTB results from sampling techniques or are caused by the physiological shedding of STB nuclei, as it is known that STB sheds off its nuclei into the maternal circulation.

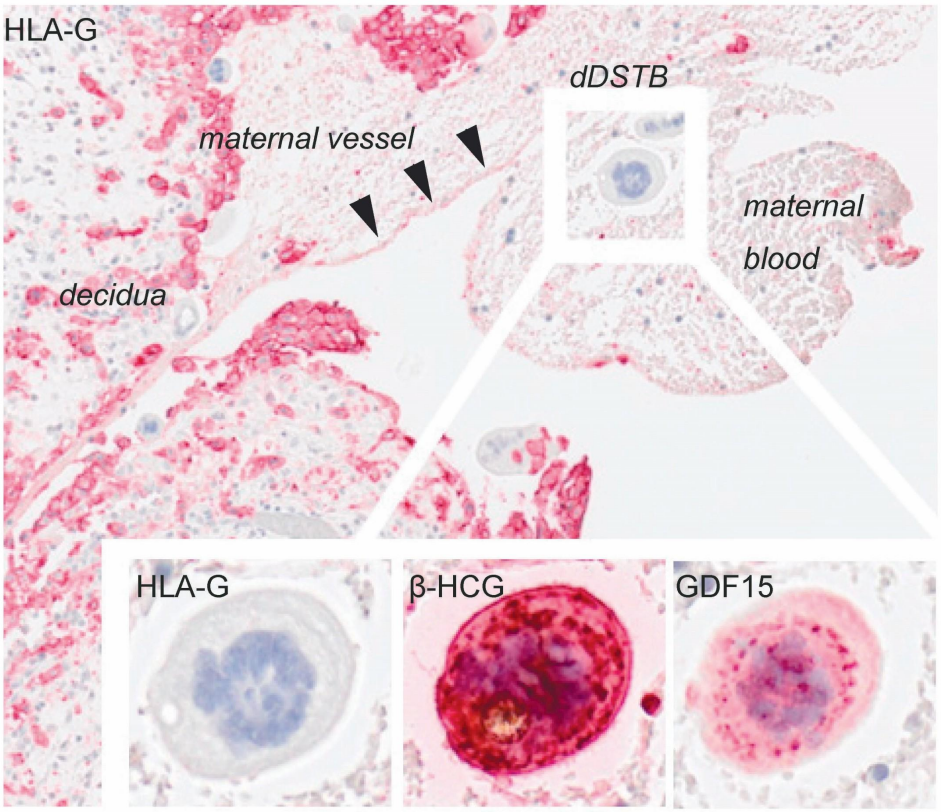


Figure 3.9: Immunohistochemical validation of dDSTB in maternal decidua.
GDF15 expressed in dDSTB is localised by immunohistochemistry in decidual tissue. dDSTB is identified among maternal erythrocytes in maternal vessels as an HLA-G negative β-hCG positive GDF15pos fragment (n=3, representative figure shown). Acknowledgment: Lena Neuper & Prof. Berthold Huppertz at the Division of Cell Biology, Histology and Embryology, Gottfried Schatz

Research Center, Medical University of Graz, Austria, and is provided as a validation of my computational finding of dDSTB sub-state in decidua snRNA-seq data (taken with permission).

3.5 Juvenile syncytiotrophoblast state

Among the three distinct vSTB nuclear states, a novel sub-state of vSTB termed as "juvenile STB" (vSTBjuv) was uncovered. Possibly, this reflects an immature nuclear state of vSTB that originates before committing to the fully differentiated STB, as discussed later in the trajectory modelling (refer to Section 2.8). vSTBjuv sub-state displayed relatively higher expression of hormone encoding genes, including placental lactogen (*CSH2*), and classical vSTB markers such as *KISS1, CGA, CYP19A1 TFPI,* and *TFPI2*, along with exocytotic expression signatures such as *HSPB1, CD63, FURIN* **(Figure 3.10a).**

A comparison of vSTBjuv with other vSTB types revealed differentially upregulated genes, including *DLK1, ACTB, TMSB10, FSTL1*, and *SPARC* (Logistic Regression, Bonferroni adjusted p-value < 0.001)- known to be associated with cytoskeletal stability and extracellular matrix organization, as detailed in **Supplementary Table 7**. Hence, it is speculated that vSTBjuv aids in the formation and maintenance of the outer placental wall before acquiring a traditional STB phenotype (i.e., endocrine, and secretory by nature). This heterogeneity might reflect a “division of labor” in the STB barrier of the placenta. No significant differences in the abundance of STBjuv nuclei between eoPE and term controls were observed, suggesting a gestation-independent role of this state. However, given the low number of samples in eoPE and term groups, it is recommended to validate this observation in a larger cohort before drawing conclusions.

To experimentally localize vSTBjuv nuclei within the multinucleated vSTB layer, markers that technically best distinguished vSTBjuv from vSTB1 and vSTB2 **(Supplementary Table 7)** were considered. This included *TENM3* (FC= 12.4, BH-corrected p-value adj. < 0.0001), which promotes homophilic adhesion [415], and *DLK1* (FC= 6.2, BH-corrected p-value adjusted < 0.0001), a paternally imprinted gene correlated with birth weight [416]. Thereafter, In-situ mRNA hybridization [183, 418] was performed with *DLK1* and *TENM3* rolling padlock-probes. Specific high-affinity probes that can distinguish single nucleotide differences in sequences were used [418]. Moreover, key marker and important pregnancy hormone, human chorionic gonadotropin β-hCG was added to capture the vSTB layer. Together, this allowed the precise localization of $TENM3^{pos}/DLK1^{pos}$ nuclei within a $\beta\text{-hCG}^{pos}$ STB cytoplasm layer in the early and late pregnancy (or, term control) placental villi **(Figure 3.10b).**

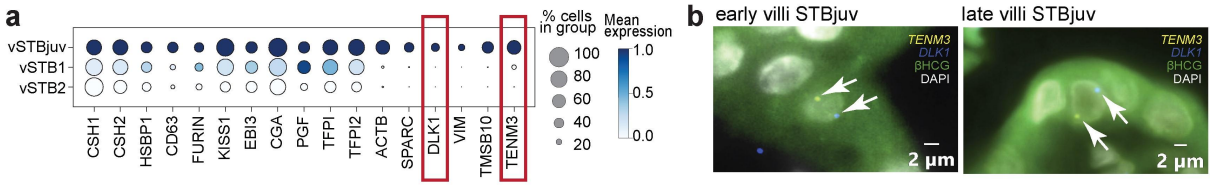


Figure 3.10: | vSTBjuv sub-state characterized by computationally and experimentally validated markers.

(a). Dot-plot with markers characterising the novel vSTBjuv nucleus state; computed by Logistic Regression based Generalized Linear model (Bonferroni adjusted two-sided, P < 0.05 and log2FC cut-off of ±0.25). Genes are scaled across clusters. (b) Representative images showing the localisation of the novel vSTB nucleus state vSTBjuv; immunofluorescence staining with vSTB protein marker β-hCG (green) combined with padlock probe based in situ mRNA hybridisation for vSTBjuv markers DLK1 and TENM3 (arrows). n=3 independent experiments with 2 biological replicates each per gestational time. CTB, cytotrophoblast; STB, syncytiotrophoblast; p, proliferating; pf, pre-fusion; juv, juvenile state. Rolling padlock probe experiment and figure generation for (b) was performed at Dr. Muller/Denchend lab at MDC Berlin, Germany, and used to support my computational findings of novel vSTBjuv sub-state as shown in (a).

3.5.1 Technical evaluation of novel STBjuv sub-state

To account for possible technical confounders in the snRNA-seq data for diligently identifying cell type/states, a mitochondrial (MT) read cut-off at 5% was employed as a quality control threshold. Additionally, the percentage of MT-transcripts per nuclei was used as a continuous covariate in the scVI data integration model and differential gene expression tests. Hence, the finding of novel STBjuv sub-state was not driven by artifacts such as %MT-transcripts per nuclei. Doublet analysis using Scrublet [103] was also performed to flag technical doublets, that later confirmed that vSTBjuv nuclei were not confounded by doublets **(Extended Data Figure 2).**

Furthermore, amortized Latent Dirichlet Allocation (LDA) [417] was performed to investigate whether any cluster-associated topics were confounded by contradictory marker expression as demonstrated by Surywanshi et al. in 2018 [146] **(see Methods).** The presence of genes from conflicting lineages (such as astrocytes and vascular cells, or STB and EVT) could indicate biological doublets and hence, introduce contamination in cell type/state annotation. Conceptually, a distinct cell type should map to a unique topic inferred by LDA as depicted in **Extended Data Figure 3**. Cell states or sub-clusters usually share the topics of the mother cluster (or “cell type”) and additionally may harbor a unique topic. However, LDA analysis showed that vSTBjuv robustly maps to ‘topic-14’ dominated by vSTB features, and it did not bear other conflicting topics, i.e., mesenchymal, or immune related topics **(Extended Data**

Figure 3). Also, the topics describing vSTBjuv was not dominated by mitochondrial genes. A full breakdown of topics is also provided in **Extended Data Figure 3.**

3.6 scVI data harmonization and evaluation of performance

For addressing the complex task of data harmonization, a state-of-the-art data integration tool, scVI [112], was employed, as demonstrated to be a consistent top performer in Luecken et al. 2022 [114]. This technique is referred to as 'data harmonization' rather than batch effect correction to emphasize that the input datasets stemmed from very different sources (e.g., library, sampling sites) and from samples from different gestational time-points having different underlying cell type/state composition.

The individual donor (Donor ID) was considered as the key batch in the variational encoder model. Furthermore, batch effect owing to sample procurement center (sampling site), biological (gestational time points), and technical covariates (library chemistry, total UMI counts, total number of genes detected, percentage of mitochondrial transcripts per nuclei, %XIST per nuclei) was considered. Performance of the integration was evaluated using well-established metrics Adjusted Rand Index (ARI), Adjusted Mutual Information (AMI), and cell-type adjusted silhouette width (ASW) per batch as used in prior integration benchmarking studies [114, 115].

For this study, a very detailed description of the strategy investigating the batch effect across four analyses is included:

- Case 1: Technical replicates to establish an upper bound of integration quality
- Case 2: Different library preparation methods
- Case 3: Different sampling sites
- Case 4: Different sampling site and library preparation methods

After an exhaustive batch effect analysis, the integration results were compared to other studies, and it was concluded that majority of the batch effects had been regressed out. Only results on those cell states for which there was high confidence that integration performed reliably were presented. Below, details of each of the four groups of analysis are provided. A multi-panel figure for the four cases is presented as **Figure 3.11.** A comprehensive table summarizing the statistics is reported as **Supplementary Table 6.**

Case 1: Batch effects evaluation between technical replicates

In the consideration of modeling batch effects in the experimental design, technical replicates were included for a sample (577_1 and 577_2) to establish an upper bound for the quality of integration. Comparable numbers of cells (6185 and 6081 cells, respectively), median genes expressed (788 and 756 genes), and similar cell state compositions characterized both samples **(Figure 3.11 a-d, technical replicates).** The relative proportion of cell-states in the two technical replicates is depicted using a kernel density plot (KDE) reflecting the density of cells in an embedded space and concordance in cell composition. As anticipated, both the ARI (0.0037) and AMI (0.003) scores were close to 0, indicating near-perfect integration. Similarly, the average ASW per batch (0.92) was close to 1, also indicating decent batch effect removal. Furthermore, the ASW at the cell-state level was investigated, indicating excellent scores for a larger number of cell states, with intermediate scores for villi myocytes and slightly lower scores for T-cells. These cell states were not the focus of our study (Supplementary Table 6, Technical replicate 557).

Case 2: Batch effects evaluation between different library preparation chemistries

Similar to establishing an upper bound for the similarity of samples through the analysis of technical replicates, a more extreme effect of using different library preparations was also investigated.

In the case of early (first-trimester pregnancy), samples were processed using 10X V3 (n=7) and 10X V2 (n=3) chemistries. The groups had 67037 and 12848 cells, with median numbers of genes expressed being 3088 and 1465 using 10X V3 and 10X V2 chemistries, respectively. The relative cell state composition was also similar **(Figure 3.11 e-h, Library chemistry)** with more vHBCp, vCTBp, vFB, vMC, vHBC, and vPAMM in 10x V3 samples (**Figure 3.11 f-h, Library chemistry**). ARI and AMI values (0.002 and 0.009) were close to zero, suggesting a negligible effect of library chemistry. The average ASW per batch was 0.88, indicating a decent removal of library effects. Furthermore, the ASW at the cell-state level was investigated, indicating excellent scores (>0.90) for a larger number of cell-states relevant to the first-trimester pregnancy, such as vCTB, vCTBp, vCTBpf, vCCT, and vSTB1 (**Supplementary Table 6, Library chemistry**). All late term controls were processed using 10X V2 chemistry. Hence, the library was not considered a confounder. Differences in eoPE samples are presented in Case 4.

Case 3: Batch effects evaluation between sampling sites (procurement) in late samples

Late samples were equally split between both Oslo and Graz sites (n=3 and n=3, respectively). Both samples had comparable numbers of cells (15733 and 13969 cells, respectively), median genes expressed (827 and 894 genes), and similar cell state compositions with some exceptions **(Figure 3.11 i-l, Sampling site term controls).** The ARI (0.016) and AMI (0.027) were close to zero, indicating that our cell state labels were not influenced by the sampling site. The average cell state ASW per batch was 0.88, indicating appropriate removal of sampling effects within a cell state identity. Cell-state ASW per batch was very good (> 0.90) for the major cell states such as vCTB, vSTB1/2 & vSTBjuv, indicating optimal batch-mixing while conserving biological information **(Figure 3.11j & Supplementary Table 6).** There were increased proportions of PAMM, vT-cells, vFB, and vMC in Graz samples **(Figure 3.11 j-k).** Hence, no key conclusions were drawn from these cell populations in the thesis. Despite observing slightly more vCTB in Graz samples, they mixed well between sites as reflected by ASW (Supplementary Table 6, Sampling site late controls).

Case 4: Batch effects evaluation between sampling site and chemistry in eoPE samples

Furthermore, eoPE samples were split between both Oslo and Graz sites (n=3 and n=2, respectively). The numbers of cells were 17604 and 12365 cells, respectively (median ~5900 cells), median genes expressed 790 and 1892 genes, respectively, and similar cell state compositions **(Figure 3.11 m-p, Sampling site eoPE).** A higher number of profiled genes were because the Graz samples were processed using 10X V3 chemistry compared to 10X V2 used for Oslo samples. Hence, the library was an overlapping confounding factor here. The ARI (0.004), and AMI (0.028) scores were close to zero, indicating that our cell-state labels were not majorly influenced by sampling sites and library chemistry. The average ASW is 0.80, indicating good integration. ASW for major cell states such as vSTB1/2/juv were very high, which rules out the influence of sampling site (Supplementary Table6). However, we observed relatively more vFB in Graz eoPE samples that might have lowered its ASW **(Figure 3.11 n).** Immune cells such as PAMM & vTcells were slightly depleted in Graz samples relative to Oslo, which apparently also have lowered their ASW score. However, no major conclusions were drawn for these cell states at any point in the manuscript. Even though vSTB2 is relatively more in Oslo, this group is well integrated (ASW= 0.88) (**Supplementary Table 6, eoPE sampling site**).

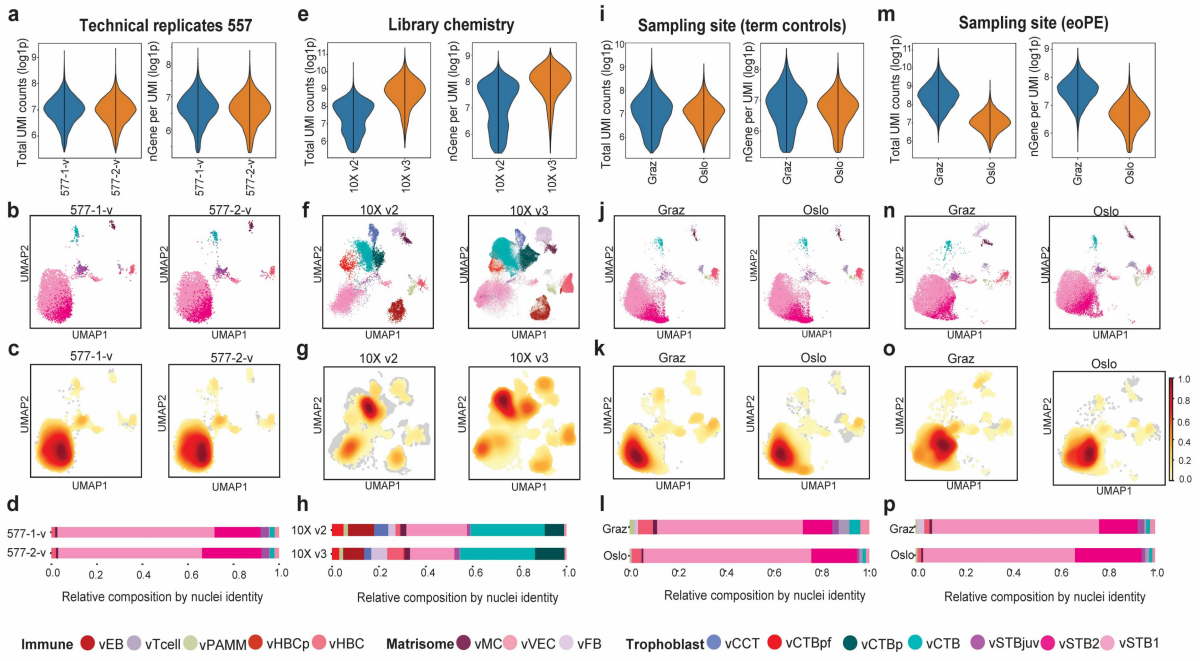


Figure 3.11: Batch effect assessment from snRNAseq placenta samples in this study.
(a) Violin plots illustrating the detected (log1p per sample) total UMI counts representative of total RNA transcripts (left) and number of total genes having at least one positive count in a cell (right) for technical replicates 557_1 and 557_2 (n= 6185 and 6081 nuclei respectively). (b) UMAP embeddings split by technical replicates 557_1 and 557_2 visualizing distribution and concordance of cell-types/states. (c) Kernel density estimation revealing similar composition in technical replicates 557_1 and 557_2. Values are scaled from 0-1 for comparison and high-density values suggest strong contribution of cells to the overall dataset. (d) Stacked bar-plot depicting similar relative composition of cell types/states in technical replicates 557_1 and 557_2. (e) Violin plots illustrating the detected (log1p per sample) total UMI counts representative of total RNA transcripts (left) and number of total genes having at least one positive count in a cell (right) for early (first trimester) pregnancy samples split by library condition- 10X V2 (n=3; 12848 nuclei) and 10X V3 chemistry (n=7; 67037 nuclei). (f) UMAP embeddings of early pregnancy samples split by 10X V2 and 10X V3 library samples visualizing distribution and concordance of cell-types/states. (g) Kernel density estimation reflecting contribution of cells to overall composition in 10X V2 and 10X V3 library samples (early). Values are scaled from 0-1 for comparison and high-density values suggest strong contribution of cells to the overall dataset. (h) Stacked bar-plot depicting similar relative composition of cell types/states in 10X V2 and 10X V3 library samples (early) with more vHBCp, vCTBp, vFB, vMC, vHBC, and vPAMM in 10x V3 samples. (i) Violin plots illustrating the detected (log1p per sample) total UMI counts representative of total RNA transcripts (left) and number of total genes having at least one positive count in a cell (right) for late term controls split by sampling site- Graz (n=3; 13969 nuclei) and Oslo (n=3, 15733 nuclei). (j) UMAP embeddings of term control samples split by sampling site- Graz and Oslo visualizing similar distribution and concordance of cell-types/states in the two categories. (k) Kernel density estimation reflecting contribution of cells to overall composition in Graz and Oslo term control samples. (l) Stacked bar-plot depicting similar relative composition of cell types/states in Graz and Oslo term controls showing increased proportions of PAMM, vT-cells, vFB and vMC in Graz samples. (m) Violin plots illustrating the detected (log1p per sample) total UMI counts representative of total RNA transcripts (left) and number of total genes having at least one positive count in a cell (right) for eoPE samples split by sampling site- Graz (n=2; 12365 nuclei) and Oslo (n=3, 17604 nuclei). (n) UMAP embeddings of eoPE samples split by sampling site- Graz and Oslo visualizing similar distribution and concordance of cell-types/states in the two categories. (o) Kernel density estimation reflecting contribution of cells to overall composition in Graz and Oslo eoPE samples. (p) Stacked bar-

plot depicting similar relative composition of cell types/states in Graz and Oslo eoPE samples showing slight depletion in immune cells (PAMM & vTcells) in Graz samples relative to Oslo.

3.7 Subcluster analysis of invasive-phenotype cell column cytotrophoblast cell type

vCCT cells were defined as the 'cell column trophoblast' that has the potential to differentiate into decidua invading extravillous trophoblast (dEVT) during early placental development **(Figure 3.3 a-b).** The formation of this trophoblast sub-state takes place at locations where placental villi attach to the decidua, and hence, referred to as anchoring villi that can give rise to structures known as cell columns. However, these cell columns are not uniform. As cells progress distally within the column, they undergo phenotypic changes and acquire EVT gene expression patterns. Hence, sub-clustering analysis of vCCT was performed to dissect this heterogeneity on a transcriptomic level.

The Leiden algorithm (scanpy implementation) with low resolution (=0.3) was applied, yielding 7 clusters. The subgroups were then merged based on the expression of key markers to distinguish proximal CCT (pCCT), distal CCT (dCCT/pEVT), and an additional transition state of CCT **(Figure 3.12a)**. Overall, pCCT is characterized by pronounced levels of *EGFR, TEAD4, TP63,* and *NOTCH2* that is regarded as the 'EGFR signature group' **(Figure 3.12 b-d).** During pCCT formation, *EGFR* activation likely contributes to acquiring migration and invasive properties of these cells [419], allowing them to penetrate the maternal decidua and remodel maternal spiral arteries. On a similar note, Hippo pathway target, *TEAD4* indicates stemness and self-renewing property [420, 421], and possibly regulate vCCT differentiation together with *YAP1* [421]. Moreover, YAP-TEAD4 complexes were found to suppress markers of vSTB differentiation [412].

The proliferative proximal trophoblasts (pCCT) progress toward the distal end of the column and transition into non-dividing trophoblasts (dCCT). This transition is marked by increased expression of EVT-specific marker genes, including *HLA-G, NOTUM,* and *HPGD* in the dCCT or, pEVT **(Figure 3.12b).** Human Leukocyte Antigen-G or HLA-G is an immune tolerance molecule. During EVT differentiation, it helps regulate the immune response by suppressing the activation of maternal immune cells, thus preventing immune rejection of the developing embryo [402, 403]. Of note, *NOTUM* encodes an enzyme modulating Wnt signaling, which is critical for trophoblast invasion and placental development [422-425]. *HPGD* (15-Hydroxyprostaglandin Dehydrogenase) is an enzyme that plays a role in prostaglandin metabolism [426]. Specifically, it helps regulate inflammation by breaking down

prostaglandins, and maybe involved in regulating the inflammatory response in the maternal-fetal interface during EVT invasion.

Moreover, a significantly increased expression of ECM remodeling genes such as *ITGA1, ITGA5, TIMP2, MMP2, ADAM12* in the dCCT **(Figure 3.12b).** Integrins are cell surface receptors that mediate cell adhesion and signaling. *ITGA1* and *ITGA5* likely facilitate EVT attachment and invasion by interacting with extracellular matrix components and signaling pathways in the uterine environment [427]. *MMP(s)* such as *MMP2* are a family of enzymes that play a crucial role in trophoblast invasion [427-429]. They degrade the extracellular matrix, allowing EVT cells to penetrate maternal tissues. Specifically, *MMP2* was found to be induced by canonical Wnt and PI3K-AKT pathway [428]. In this regard, *TIMP2* (Tissue Inhibitor of Metalloproteinases 2) is involved in regulating the activity of matrix metalloproteinases *(MMPs).* In EVT differentiation, *TIMP2* helps control the activity of *MMP(s)*, which are enzymes that facilitate tissue remodeling and invasion by degrading extracellular matrix components [429].

Notably, robust expression of *TEAD1* (TEA Domain Transcription Factor 1) and *WWTR1* (WW Domain Containing Transcription Regulator 1) was also recapitulated, known as critical players in the regulatory network that governs the invasive properties of EVT **(Figure 3.12b)** [420, 421, 430]**.** These two proteins are part of the Hippo signaling pathway and interact with each other to control various cellular processes. In conjunction with *WWTR1*, *TEAD1* acts as a transcriptional co-activator and forms a complex that binds to specific DNA sequences, known as TEAD-binding sites. By doing so, they regulate the transcription of genes that are critical for EVT differentiation and invasion. Their downstream targets include genes responsible for extracellular matrix degradation, cell adhesion such as *ITGA1, ITGA5,* and *MMP2.* Moreover, *TEAD1* and *WWTR1* are involved in controlling the balance between EVT proliferation and differentiation. Specifically, they play a role in promoting the differentiation of CTB(s) into invasive EVTs, which is essential for placental development and the implantation process. Importantly, the activity of *TEAD1* and *WWTR1* can be influenced by other signaling pathways, such as the Wnt and TGF-β pathways. This crosstalk allows for the fine-tuning of EVT behavior in response to various signals from the uterine microenvironment.

Taken together, the dCCT group robustly expresses the above-mentioned genes which we regarded as the 'ITGA1 signature group' **(Figure 3.12 b-d).** After invading the decidua, one of the primary functions of EVTs is to transform spiral arteries for which the candidate members of the ITGA1 group are required. This is to establish the uteroplacental blood circulation necessary for providing essential nutrients and oxygen to the developing baby.

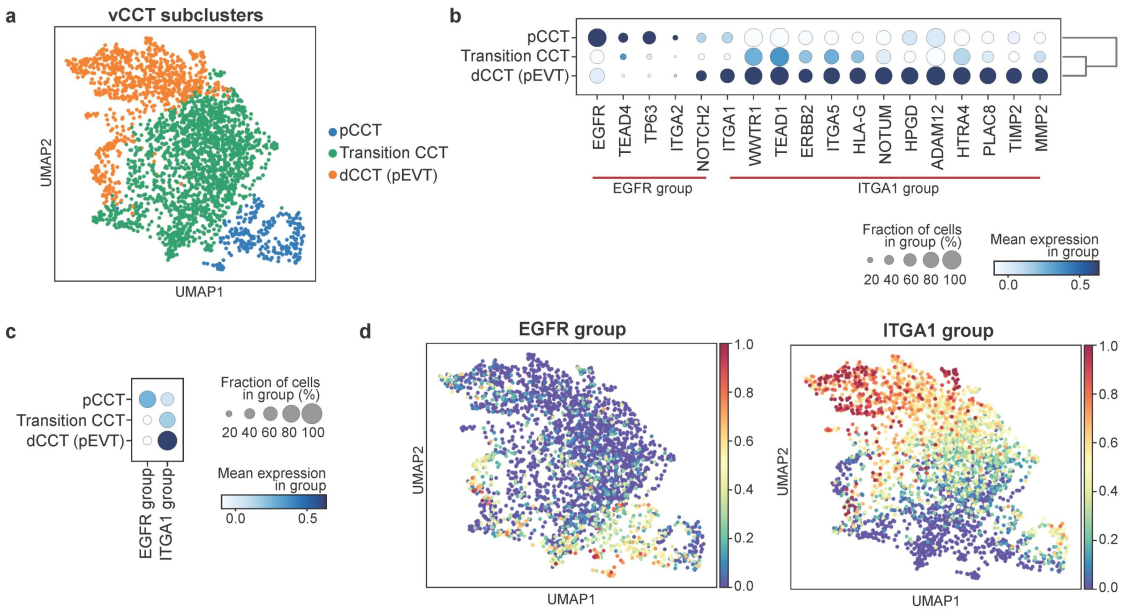


Figure 3.12: Subcluster analysis of invasive-phenotype cell column cytotrophoblast cell type.
(a) UMAP embedding visualizing vCCT subclusters based on robust and specific marker profiles. (b) Dotplot depicting expression of key genes relevant for each subgroup. The fraction of cells per group expressing a gene is size coded; normalized mean expression is color-coded. (c) Module scores of two signature sets (derived from computational and curated subcluster markers) revealing relatively robust expression of Set1 (EGFR, TEAD4, TP63, ITGA2, NOTCH2) in pCCT and that of Set2 (ITGA1-MMP2). (d) Feature plot showing the signature scores of EGFR-group and ITGA1-group.

3.8 Trophoblast trajectory modelling

In this study, trajectory analysis was conducted using STREAM (see Methods) [149] and subsequently validated with additional methods such as Diffusion Pseudotime (DPT) [150] and force-directed graphs to ensure its robustness **(Figure 3.13a & Extended Data Figure 4).** STREAM predicts cellular trajectories through the utilization of ElPiGraph [153] that is entirely revamped algorithm for previously known elastic principal graph optimization.

Leveraging existing knowledge on trophoblast development [209, 219], vCTBp was designated as the root progenitor cell state in downstream analysis **(Figure 3.13a).** Importantly, the two divergent lineages of the bipotent trophoblast, namely vSTB and vCCT were characterized by identifying markers that (i) exhibit dynamic regulation across pseudotime, referred to as "transition markers," and (ii) display robust expression at the terminal branches, defining cell-fate commitment, known as "leaf markers" (**see Methods**).

Of note, a large overlap was noticed between the leaf and transition markers **(Supplementary Table 16).**

Trophoblast differentiation is inherently linked to the spatial organization of its underlying cell types/states, which is organized in two layers. To validate the vSTB and vCCT transition markers, a spatial proteomics method adapted from deep visual proteomics [431] was utilized. Laser microdissection-based proteomics enabled the isolation of single cells or multi-cellular regions of interest (ROIs) with preserved spatial resolution after microscopic inspection. In the spatial proteomics approach, vCTB, vSTB, and vCCT cell types were identified on formalin-fixed, paraffin-embedded (FFPE) sections, laser-microdissected cell type-specific regions, and performed liquid chromatography-mass spectrometry (LC-MS) (n=4; **Figure 3.13 b-c**). The dataset included replicates of the same cell class, demonstrating high proteome consistency with excellent proteome correlations (Pearson r > 0.9 for cell types; refer to **Supplementary Table 17**). The major rationale for using this technique is to validate snRNA-seq trajectory inferred transition markers on a protein level and identify markers characterizing the vSTB and vCCT fates. The subset of overlapping markers essential for both lineages is discussed in subsequent sections.

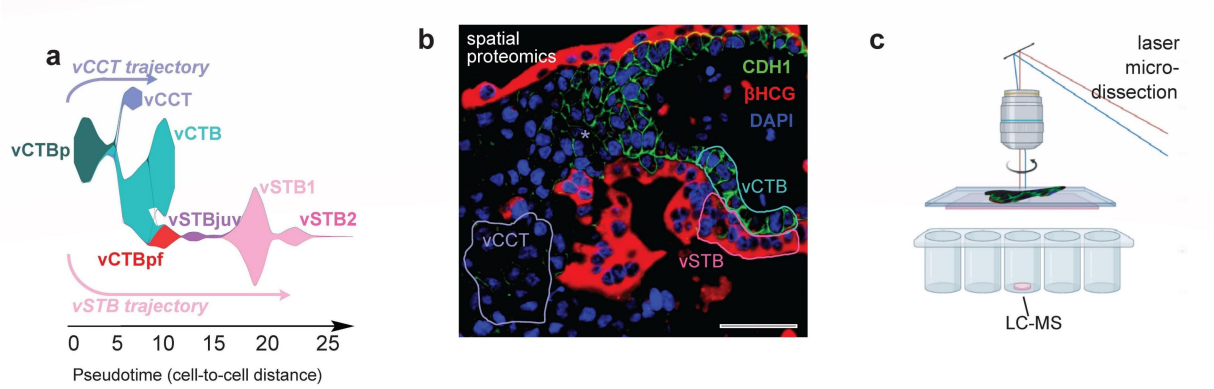


Figure 3.13: Trajectory modelling of first-trimester trophoblast recapitulates STB and CCT differentiation.

(a). Stream plot showing the developmental trajectory of early trophoblast and cell density across pseudotime. Branch length represents pseudotime progression, branch width is directly proportional to cell numbers at a given pseudotime. (b) Stained using immunofluorescence markers E-Cadherin (CDH1) and human chorionic gonadotrophin (β-hCG) to discriminate between vCTB, vSTB, and CCT; (c) set areas were laser micro-dissected, captured and processed for untargeted proteomics using LC-MS). Acknowledgement: Proteomics figures (b) and (c) were produced by our collaborators- Dr. Fabian Cosica & Jose Nimo (MDC Berlin) and are included (with permission) to support subsequent figures associated with sections 2.7.1 & 2.7.3. STB, syncytiotrophoblast; CCT, cell column trophoblast; CTB, cytotrophoblast; p, proliferative; pf, pre-fusion; juv, juveline.

3.8.1 vCCT differentiation lineage

The vCCT differentiation path was characterized by the dynamically increasing expression of key transition markers, namely *HLA-G, NOTUM, TEAD1,* and *FOS* **(Figure 3.14a),** further validated by proteomics **(Figure 3.14b, d).** Moreover, the vCCT cell fate was characterized by the elevated expression of the AP1 signaling pathway **(Figure 3.14a),** where transition markers *TEAD1* and *WWTR1* are key contributors. As already discussed in section 2.6, both are key regulators of EMT signaling required for decidual invasion, and hence sustaining a functional maternal-fetal interface during pregnancy [420, 421, 430]. Additionally, vCCT lineage path exhibited increasing expression of transition markers, including *MYCNUT, TGFB1, LAIR2, ANGPT4, GALNT2, HPGD, HTRA4, ITGA5, MMP12,* and *ISM2*, holding multi-factorial roles in metabolism, cell adhesion, and migration required for acquiring an invasive phenotype [404, 426-429] **(Figure 3.14c).** Importantly, the signatures characterizing vCCT fate were confirmed through the intersection of markers identified as transition genes, and those that were also substantially abundant at a protein level **(Figure 3.14d).**

Next, the spatial distribution of key cell types or states within the villi was investigated by integrating snRNA-seq data with 10X Visium spatial transcriptomics **(Figure 3.14e).** For the analysed sample, number of spots under tissue was 1917; mean reads per spot was 153,206; median genes per spot was 1429 and median UMI counts per spot was 2375, as reported by SpaceRanger software [453]. Subsequently, SPOTlight [196] was employed to deconvolute spatial transcriptomics capture locations, commonly referred to as spots. Each spot is expected to have ~10 cells. This approach revolves around an initialized seeded non-negative matrix factorization (NMF) regression using specified cell-type marker genes and non-negative least squares (NNLS). Subsequently, a vCCT signature score was computed using the transition gene list visualized in Figure 2.14d, and this module was found to be robustly expressed across the deconvoluted vCCT region **(Figure 3.14f).**

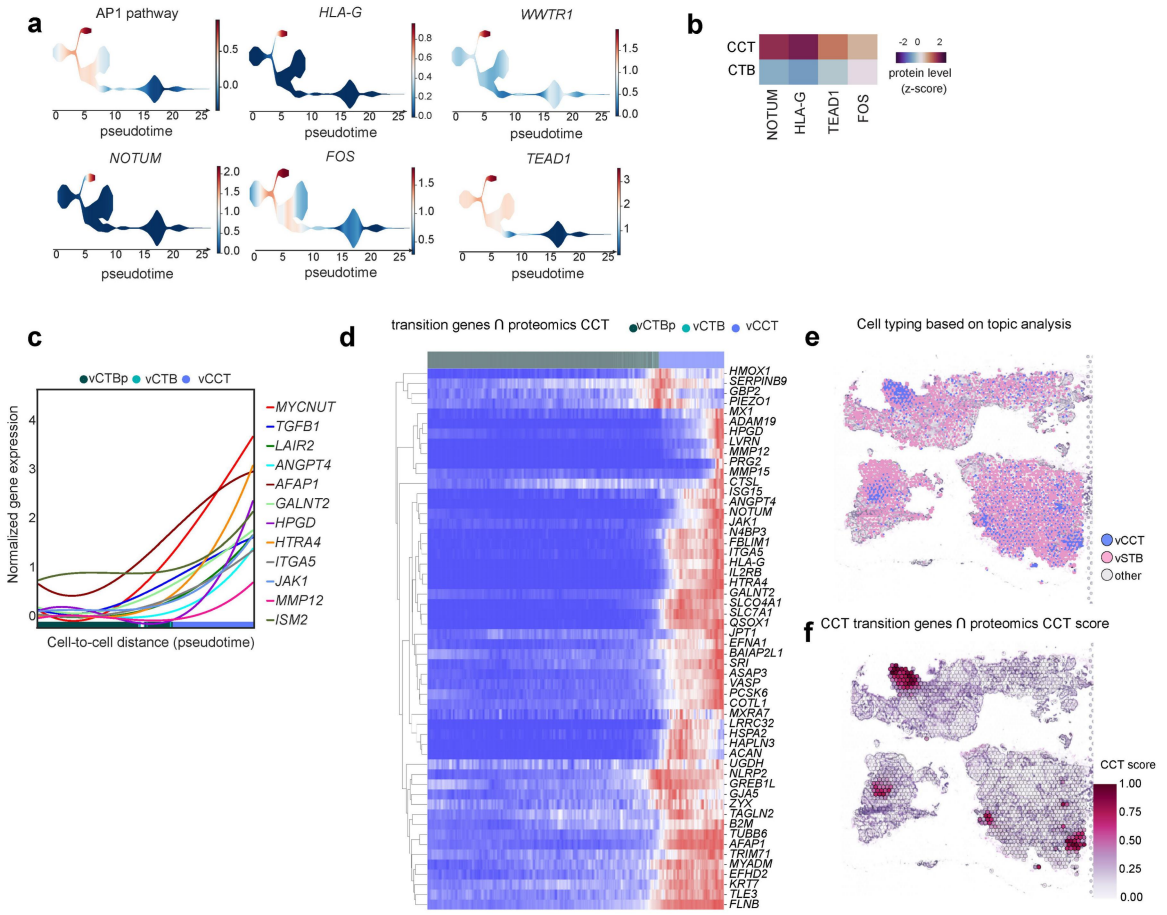


Figure 3.14: Characterization of vCCT lineage path.
(a) Stream plot displaying scaled expression of critical vCCT lineage signatures, including AP1 pathway and selected transition markers HLA-G, WWTR1, NOTUM, FOS, and TEAD1. Branch length represents pseudotime progression, branch width is directly proportional to cell numbers at a given pseudotime. (b) Heatmap depicting scaled (z-scored) protein abundance levels of key markers characterizing vCCT fate. (c) Selected lineage-specific transition markers significantly correlating to pseudotime (cell-cell distance) progression for vCCT identity. Lines are polynomial regression fits of normalized expression data. Cell-type membership is incorporated on the x-axis and ordered by pseudotime. (d) Heatmap describing mean gene expression changes along the inferred pseudotime for vCCT lineage. Genes shown are detected as transition marker genes that are also found to be differentially abundant in spatial proteomics analysis. Cells ordered by pseudotime values (top bar) and coloured according to (a). Gene expression is scaled row-wise. (e) 10X Visium-based spatial profiling revealing vSTB and vCCT enriched spots using non-negative least squares (NNLS) and non-negative matrix factorisation (NMF) based Spotlight deconvolution approach. (f) Module scores of vCCT developmental drivers derived by intersecting snRNA-seq based trajectory analysis and spatial proteomics based differentially abundant proteins. Specifically, an intersected list of transition genes (expression correlates with pseudotime) and key proteins (from proteomics) plotted in (d) were used to compute a module-score. For the vCCT path, the z-score of protein expression comparing a given lineage against others was > 1.5 & n = 6 biological replicates.

10X Visium analysis also allowed the identification of spatially variable developmental drivers crucial for determining cell fate commitment. Notably, candidates like *HLA-G, HPGD, NOTUM, ASCL2* and *ISM2* were detected as spatially variable genes that strongly

contributed to this module score **(Figure 3.15).** A previous study by Varberg et. al. 2021 [432] found *ASCL2* to be conserved and critical regulator orchestrating deep trophoblast invasion and maternal spiral arteries remodelling in human, and rat models. In the aforementioned research [432], *ASCL2* transcript expression was observed in the EVT column and junctional zone, representing tissue sources of invasive trophoblast progenitor cells within human and rat placentation sites, respectively. This observation aligns with the 10X Visium data that recapitulated strong *ASCL2* expression in vCCT dominated spots. In the same regard, *LAIR2* is part of the LAIR family of receptors, and it is known to interact with collagens II/III, suggesting a potential involvement in ECM interactions [433]. Former investigation by Founds et. al. 2013 [404] demonstrated that LAIR2 expressing EVT(s) invade the decidua, participate in maternal spiral arterioles remodelling, and modulate innate immune response at maternal–fetal interface.

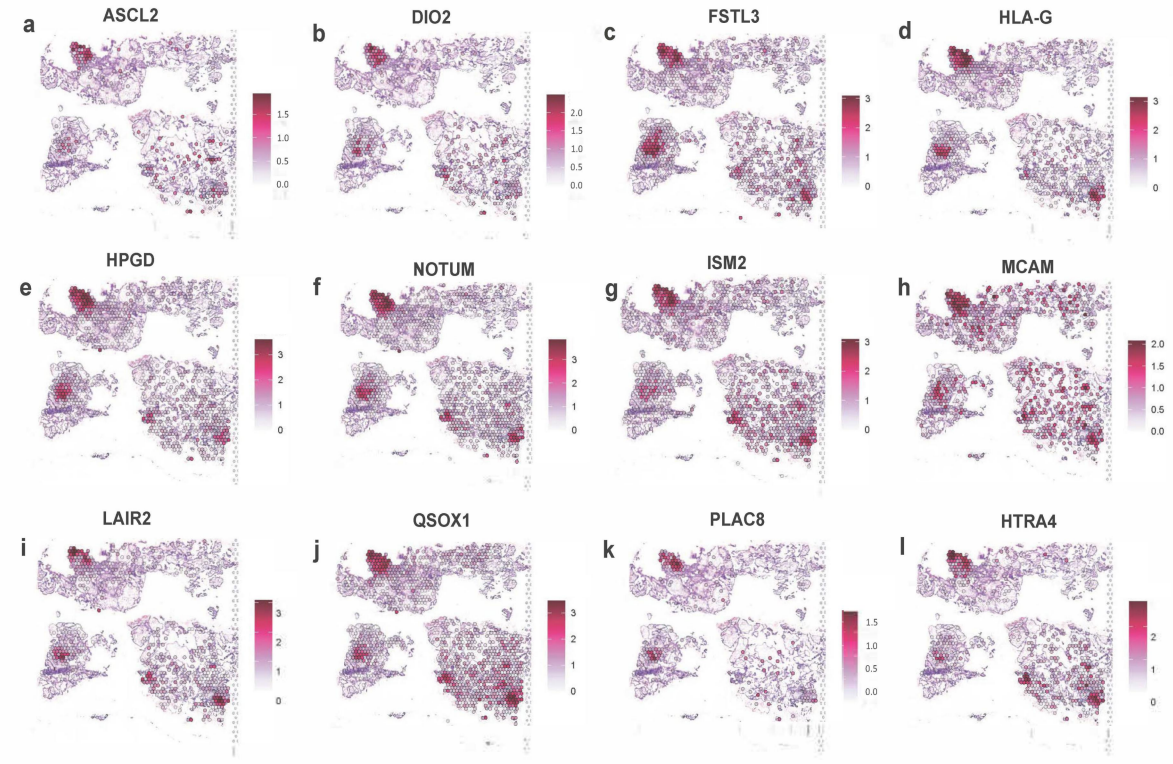


Figure 3.15: Spatially variable genes detected as transition markers for vCCT lineage.
Selected spatially variable genes inferred from 10X Visium analysis that are also vCCT transition markers are visualized (see Methods).

3.8.2 BMP signaling blocks spontaneous fusion in vCTBpf

Trajectory inference suggested that dynamic transitional state vCTBpf acts as a vSTB progenitor **(Figure 3.13a)**. During early placental development, vCTBpf regulates cell-to-cell fusion as revealed from its transcriptomic profile [401, 434] (as discussed in **section 3.1.2**)- hence, it is very likely that mononucleated vCTBpf differentiates to give rise to multinucleated vSTB.

It is worth noting that the BMP-antagonist *GREM2* [435] demonstrated robust expression in the vCTBpf. Interestingly, the expression of the BMP/Activin receptor dimer *ACVR2A/BMPR1A* and BMP7 (a BMP signaling agonist) [436, 437] was identified in the vCTB branch but was repressed in the vCTBpf section **(Figure 3.16)**. This temporary inhibition of BMP in vCTBpf could potentially serve as a prerequisite for the fusion and trophoblast differentiation from vCTB to vSTB. Collectively, these findings indicate that "pre-fusion vCTB or vCTBpf" represents a vCTB sub-state that undergoes evolution to become vSTB through a process involving transient downregulation of *BMP7* and an increase in *ERVFRD-1* (syncytin-2) expression.

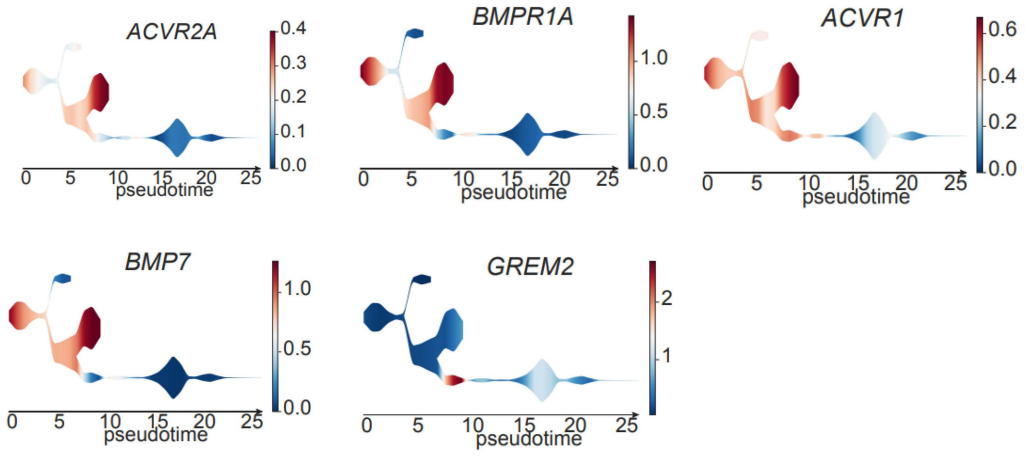


Figure 3.16: vCTBpf characterized by BMP signaling genes.
Stream plot depicting mean gene expression of vCTBpf associated developmental markers that are members of BMP signaling pathway [435-437]. BMP-inhibitor GREM2 and BMP agonizer BMP7 expression are mutually exclusive. Branch length represents pseudotime progression, branch width is directly proportional to cell numbers at a given pseudotime.

3.8.3 vSTB differentiation lineage

A total of 280 transition markers showing dynamic up or downregulation during vSTB differentiation were detected **(Supplementary Table 16).** In the vSTB lineage, the increasing expression of genes such as *CGA, KYNU, ARHGAP26, ADAM12, AFF1, PAPPA2, LIFR,* and *PSG4* was positively correlated with the pseudotime **(Figure 3.17b).** *CGA* is a subunit of human chorionic gonadotropin (hCG), a hormone produced by vSTB cells. Importantly, hCG plays a crucial role in early pregnancy, particularly in the maintenance of the corpus luteum, which continues to produce progesterone to support the uterine lining. *CGA* also possesses immunomodulatory properties that help protect the fetus from the maternal immune system. Pregnancy-specific beta-1-glycoproteins (*PSGs*) are a family of protein encoding genes produced by the placenta during pregnancy and known to play a role in immunomodulation [438-441] and may support a tolerogenic environment [442]. *ARHGAP26* is a Rho GTPase-activating protein involved in controlling cell migration, cytoskeletal dynamics, and cell adhesion [443]. *ADAM12* and *PAPPA2* are both secreted metalloproteinase and previously described as first-trimester markers of healthy trophoblast [444]. *ADAM12* is known as a regulator of cell adhesion, migration, and proteolysis of ECM [444, 445]. Importantly, *ADAM12, PAPPA2,* and *LIFR* genes are regulated by *EP300/p300* and hence, governs placental syncytialization process [434, 446] required to maintain multinucleated vSTB barrier.

Subsequently, to understand the spatial localization of vSTB developmental drivers, a module score was computed using the transition markers plotted in **Figure 3.17c**, and this signature module was robustly expressed across the deconvoluted vSTB region, as shown in **Figure 3.17e.**

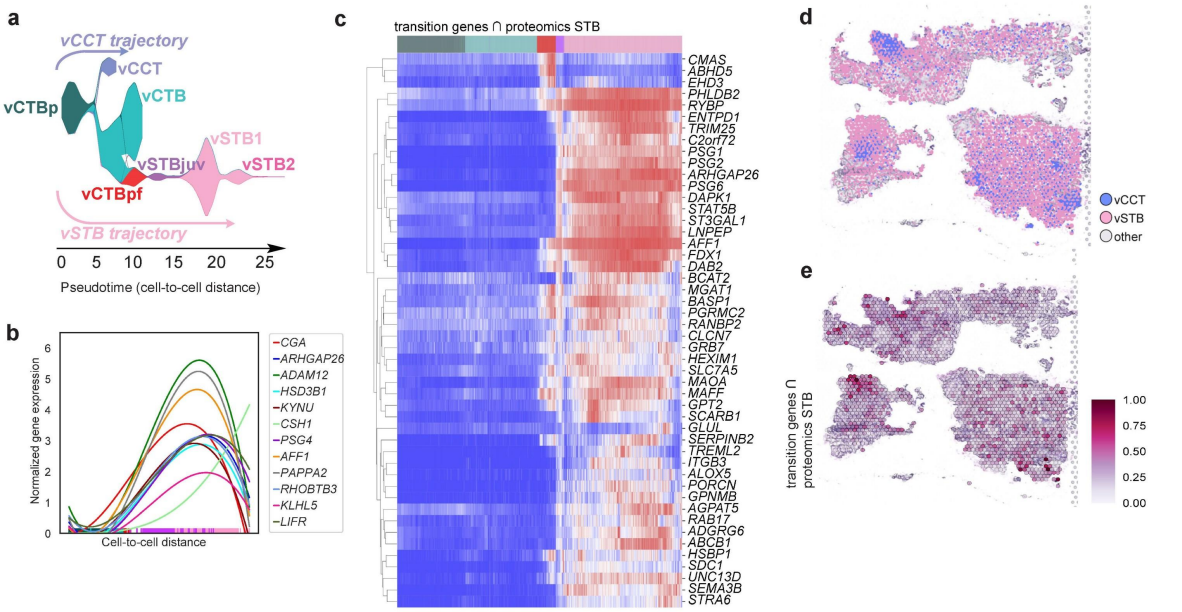


Figure 3.17: Characterization of vSTB lineage path.
(a). Stream plot showing the developmental trajectory of early trophoblast and cell density across pseudotime. Branch length represents pseudotime progression, branch width is directly proportional to cell numbers at a given pseudotime. This figure is repetitive of figure 2.13a and added for ease in understanding rest figures in this panel. b). Selected lineage-specific transition markers significantly correlating to pseudotime (cell-cell distance) progression for vSTB identity. Lines are polynomial regression fits of normalized expression data. Cell-type membership is incorporated on the x-axis and ordered by pseudotime. (d) Heatmap describing mean gene expression changes along the inferred pseudotime for vSTB lineage. Genes shown are detected as transition marker genes that are also found to be differentially abundant in spatial proteomics analysis. Cells ordered by pseudotime values (top bar) and coloured according to (a). Gene expression is scaled row-wise. (e) 10X Visium-based spatial profiling revealing vSTB and vCCT enriched spots using non-negative least squares (NNLS) and non-negative matrix factorisation (NMF) based Spotlight deconvolution approach. (f) Module scores of vSTB developmental drivers derived by intersecting snRNA-seq based trajectory analysis and spatial proteomics based differentially abundant proteins. Specifically, an intersected list of transition genes (expression correlates with pseudotime) and key proteins (from proteomics) plotted in (d) were used to compute a module-score.

Unlike the vCCT lineage, *TEAD1*, a negative transition marker for the vSTB lineage is differentially downregulated as trophoblasts transitioned from vCTB to vSTB **(**previously depicted in **Figure 3.14a, b).** This observation aligns with the spatial proteomics analysis, which validated lower levels of the YAP1 protein in vSTB compared to vCTB **(YAP1; Figure 3.18).** Taken together, this confirms the repression of YAP/TAZ pathway as vSTB differentiation progresses- in agreement with prior investigations [421, 447, 448]. Furthermore, few key vSTB lineage markers, including SDC1, CGA, and GDF15 were validated on a protein level **(Figure 3.18).** In this regard, *SDC1* is a cell surface proteoglycan that is expressed in the STB, and is known to modulate adhesion, and ECM interactions required for communication with maternal tissues [449]. On the other hand, *GDF15* expressed by vSTB cells is released into the maternal circulation, where it might contribute to

modulating maternal-fetal communication, immune tolerance, and nutrient transport [450]. *LNPEP* is an enzyme involved in the metabolism of peptides, including vasopressin and oxytocin [451]. *BMP1* is an enzyme that belongs to the bone morphogenetic protein family and drives ECM remodelling and tissue development [452]. In the placenta, *BMP1* may be involved in processes such as trophoblast invasion and tissue remodeling, which are essential for placental development and function **(Figure 3.18).**

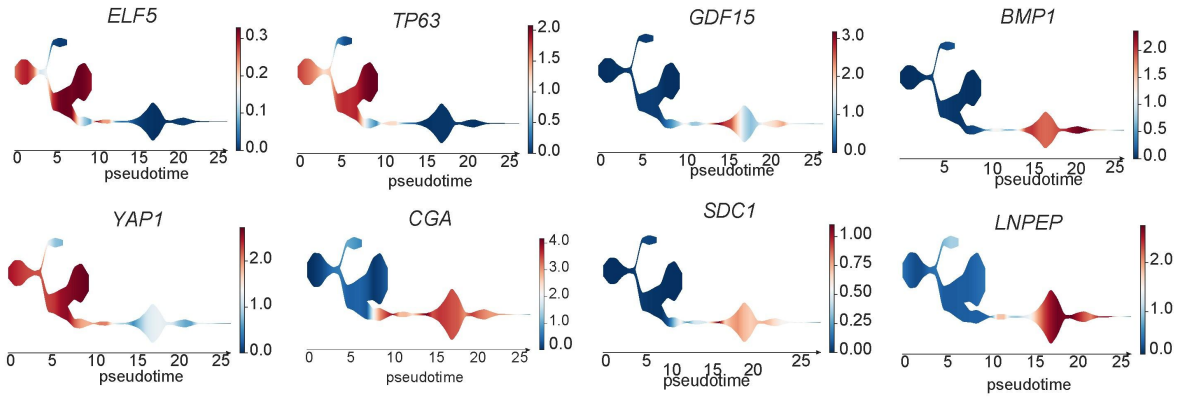


Figure 3.18: Stream plot depicting key vSTB lineage markers.
Lineage markers with decreasing (ELF5, TP63, YAP1) and increasing (CGA, SDC1, GDF15, BMP1, LNPEP) expression across pseudotime are plotted. Branch length represents pseudotime progression, branch width is directly proportional to cell numbers at a given pseudotime.

3.9 Comparative investigation of eoPE vs term controls

In this study, a significant computational challenge was encountered in addressing variability associated with 'preterm' gestational age. This challenge arose due to the collection of eoPE patient samples around the 34th week, while the term control samples were acquired between the 38th to 40th week of pregnancy. To tackle this issue, harmonization of an external scRNA-seq dataset [353] that compared non-pathological preterm and term pregnancy samples was performed using scVI/scANVI [112, 113]. The integrated preterm controls were reported to maintain normotensive pregnancies. Only eoPE and term control samples from this study were considered to understand the relationship among eoPE, preterm controls, and term controls. Of note, this facilitated the identification of cell type/state-specific signatures associated with preterm relative to term controls (refer to Methods). It is important to note that this integration was performed separately for placenta villi and decidua samples.

Upon projecting the cells into UMAP embeddings, effective batch mixing of the datasets was observed, especially for conserved cell types/states like vCTB **(Figure 3.19).** To quantify the integration, the adjusted rand index (ARI), adjusted mutual information (AMI), and cell-type-specific absolute silhouette width (ASW) were computed (see Methods). The ARI and AMI were 0.051 and 0.117, respectively, indicating satisfactory integration for the placental clusters **(Supplementary Table 9).** Additionally, the mean ASW (scaled between 0-1, where 1 signifies perfect integration) per donor was 0.807, with generally high scores per cell type. In the decidua, the ARI and AMI were 0.03 and 0.12, respectively, while the mean ASW per donor was 0.826, indicating appropriate integration **(Supplementary Table 9)**.

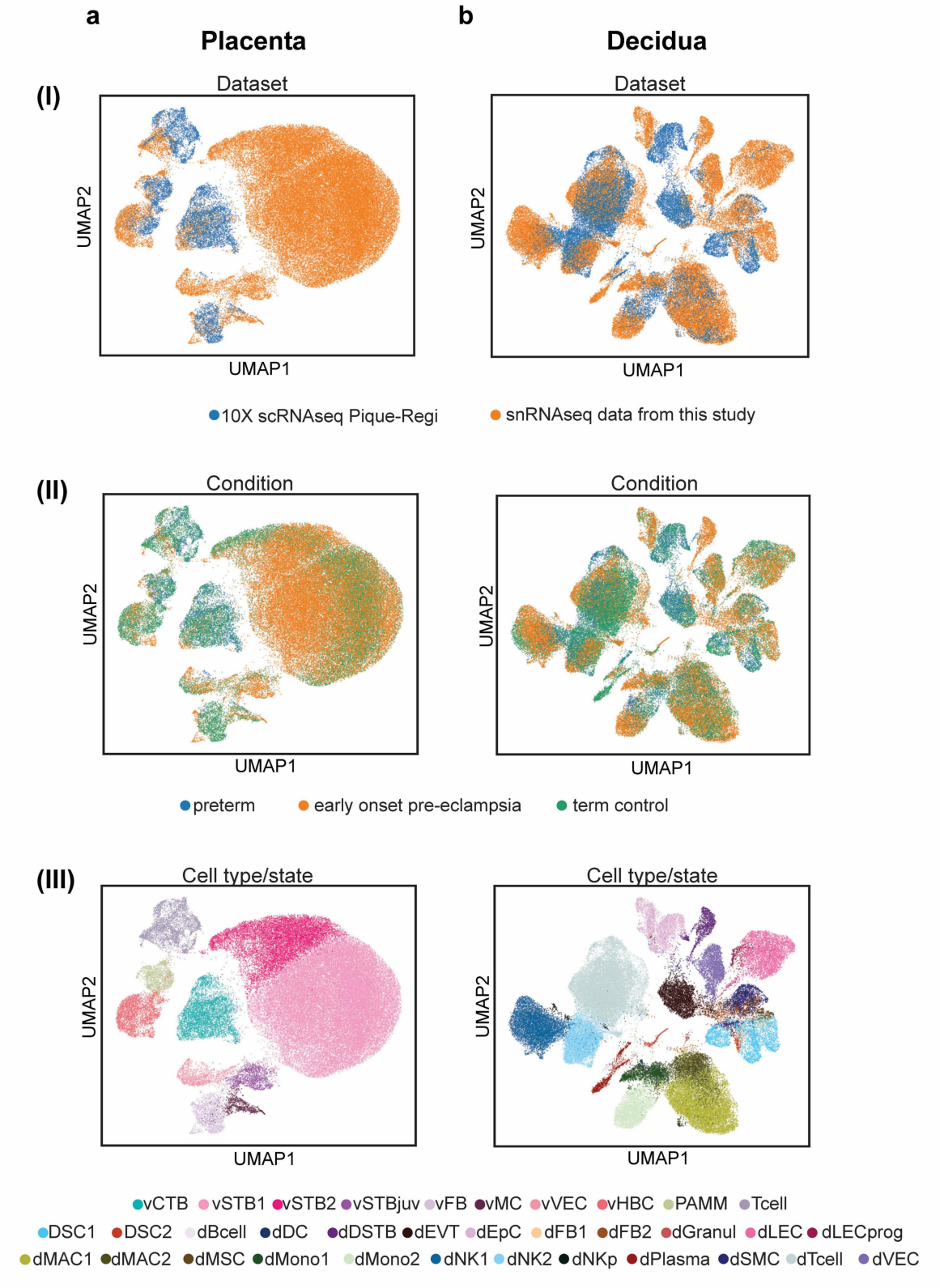


Figure 3.19: Integration visualization summary for gestational age correction in differential gene expression inference.
UMAP visualization of the 10X snRNA-seq from this study integrated with 10X scRNA-seq Pique-Regi data [353] of (a) placenta villi and (b) decidua respectively colored by (I) dataset of origin, (II) condition and (III) cell types/states. Data harmonization and label transfer was performed using scANVI. Since vSTB(s) were not profiled by Regi et al 2019, downstream analysis was restricted to conserved cell

types/states in both placenta and decidua. Batch-effect investigation using statistics are tabulated in Supplementary Table 9.

As previously mentioned, performing statistical analysis of cell type or state composition was difficult due to the limited number of samples (n=5 biological replicates for eoPE). Therefore, Kernel density estimation (KDE) was employed to visually inspect the abundance or depletion of cell type/state (see **Extended Data Figure 5**). An abundance of dNK1 and dNK2 lymphoid sub-states, as well as dMAC1 and dMAC2, was observed in eoPE. However, it is important to note that this observation should be confirmed in a larger cohort for statistical validation.

3.10 Immune specific dysregulation in eoPE vs term controls

Maternal characteristics such as BMI, age, coexisting maternal disease (e.g. diabetes mellitus), and ethnicity could incorporate potential confounders in the downstream analysis for eoPE relative to term controls. Hence, samples were carefully selected between conditions to control these effects (refer to **Supplementary Table 1**). Importantly, no significant differences in maternal age or BMI were found between late term controls and eoPE groups using unpaired t-test with Welch's correction **(Extended Data Figure 6).** Additionally, co-occurring diseases were reported for the majority (80%) of the samples, including one known case of hyperlipidemia, one case of asthma, and one case of mild allergies. Although ethnicity was not recorded at the time of sampling, samples were derived from cohorts in Graz (Austria) and Oslo (Norway) where the population is homogeneous Caucasian. Potential technical effects of sampling site and ethnicity were minimized by including the procurement center of samples as a categorical variable when performing scVI data integration (refer to Methods and section 2.5).

Molecular dysregulations per cell type/state were discerned using Logistic Regression (avg log2FC cut-off of 0.25; Bonferroni corrected p-value < 0.05) after adjusting for covariates arising from technical factors and preterm signatures **(Methods & Supplementary Table 11).** Downstream analysis and interpretations are provided for the cell types/states for which adequate preterm corrections were possible**.**

Firstly, the investigation was focused on specific perturbations in the immune populations- both lymphoid and myeloid lineages present at the maternal-fetal interface. Importantly, the

examination involved the assessment of transcriptomic changes occurring in analogous cell types/states – maternal dMAC1/2 sub-states and fetal vHBC. The comparison between dMAC1/2 and vPAMM as well as dTcell and vTcell were excluded. This is because integration and composition of vPAMM and vTcell were affected by sampling sites in late pregnancies **(refer to section 3.5).**

3.10.1 Dysregulation of dNK and dTcell types

Firstly, Logistic Regression analysis **(Supplementary Table 11)** comparing eoPE relative to term controls displayed differentially upregulated candidates such as *FKBP5* and *SMURF2* in dNK1 and dNK2 sub-states **(Figure 3.20a).** *SMURF2* (SMAD-specific E3 ubiquitin protein ligase 2) plays a regulatory role in NK cells by controlling turnover of signaling molecules and negatively regulating TGF-β signalling [454]. In turn, this could impact NK cell cytotoxicity, target cell recognition, and the balance between activating and inhibitory signals [455]. A previous study by Yang et al 2009 suggests that spatio-temporal migration and invasion of murine trophoblast cells might be regulated by *Smurf2*, potentially involving downregulation of the TGF-β type-I receptor [456]. Elevated expression of *CCL4*, a chemokine previously associated with PE [457], is also detected in dTcells **(Figure 3.20a)**. Additionally, increased levels of *XAF1* (X-linked inhibitor of apoptosis (XIAP)-associated factor 1), [458] a proapoptotic protein functioning as an alternative pathway for TNF-α induced apoptosis, was noted in dTcell.

Secondly, to delineate the transcriptomic regulatory landscape of dNK1, dNK2 and dTcell, iRegulon [161] was applied on differentially upregulated genes (see Methods). In this regard, dNK1 was found to be specifically regulated by *PRDM2, RUNX1, CBFB, RBBP* and *FOXP1* **(Figure 3.20b).** *PRDM2* is known to induce apoptosis and G2/M arrest [459], and hence might cause accelerated apoptosis of dNK cells in eoPE. *RUNX1* is a transcription factor regulating HSCs [460]. In a recent study, Kannan et al. 2023 [461] demonstrated, using a conditional knockout mouse model, that Runx1 is linked with compromised uterine angiogenesis, trophoblast differentiation, and dysregulated vascular remodeling. On the other hand, *ZNF513, RREB1, EP300, TCF7L2, NFYA* and *ACO1* constitute transcriptional regulators of dNK2 sub-state **(Figure 3.20b)**. dTcell is found to be specifically regulated by pro-apoptotic factor *TGIF1* **(Figure 3.20b)**, a co-repressor in the TGF-β1 pathway. Precisely, *TGIF1* suppresses TGF-β1 signaling by a direct competition with the transcriptional co-activator *p300/CBP* for *Smad2* interaction, resulting in the repression of genes activated by TGF-β1 [462]. Beyond its association with Smad signalling, *TGIF1* has been identified as a

crucial component in the TNF-α cytotoxic program [463]. This implies a potential involvement of *TGIF1* in initiating a pro-inflammatory response associated with eoPE. Other dTcell regulators include *CEBPD, ATF4, SRF, KAT2A* and *NFKB1* **(Figure 3.20b)**. *CEBPD* is known to demonstrate a context-dependent role in apoptosis. For instance, it can promote mitochondrial-mediated apoptosis by inducing CASP3/8 activity [464]. In context of cancer, *CEBPD* is pivotal in mediating HMDB-induced apoptosis through activation of *p38/CREB* pathway [465]. Another dTcell TF, *ATF4* is also known to regulate apoptosis associated genes, including those involved in the unfolded protein response (UPR), endoplasmic reticulum (ER) stress and oxidative stress adaptation [466]. During PE, heightened ER stress was shown to result in increased PERK and IRE1 expression that can further activate ATF genes linked with abnormal placentation [467]. Also, elevated expression of NFκB was previously observed in PE, where it stimulates the expression of both pro-inflammatory and anti-angiogenic proteins, thereby exacerbating oxidative stress, inflammation, and vascular dysfunction characteristic of PE [468].

Thirdly, Metascape [409] was utilized to perform a differential pathway enrichment analysis using significantly upregulated genes in dNK1, dNK2, and dTcell types **(Figure 3.20c)**. Here, the comparison between dTcells and vTcells was avoided as the latter was affected by sampling site. Both dNK1 and dNK2 sub-states exhibit enrichment of processes like transcriptional misregulation in cancer, transcriptional regulation by RUNX3, TGF-β and FGF pathways **(Figure 3.20c).** As discussed before, TGF-β is a key immunomodulatory cytokine that plays a crucial role in immune tolerance and vascular remodelling. In the context of dNK cells, TGF-β signaling can contribute to a tolerogenic maternal-fetal interface, promote immune tolerance and prevent excessive immune responses against fetal tissues [469]. A previous study found elevated TGF-β1 expression in the PE decidua as compared to normal pregnancies [470]. Of note, this was reported to inhibit functional activation of dNK subsets, and ultimately resulted in impaired decidualization and placentation [470]. dTcell exhibited enrichment of specific pathways, including the CCR5 pathway, ILK pathway, NFAT pathway, and cytokine signaling in eoPE condition **(Figure 3.20c)**. Notably, CCR5 signaling was mediated by candidates such as *CALM2, FOS, GNAQ, CCL4, PRKACB*, and *MAPK1* **(Supplementary Table 19)** and might be linked to inflammation secondary to endothelial dysfunction and related vascular disorders occurring during PE. Previous research also linked *CCR5* polymorphisms to PE susceptibility in Caucasians [471]. On the other hand, exposure to hypoxia elicits an excessive unfolded protein response (UPR) and ER stress, which might cause elevated levels of *HSP90AA1, ILK, RAC1, IQGAP1, IGF1R, MAPK1*, and *ARF1,* contributing to the Integrin-linked kinase (ILK) pathway. However, specific associations between ILK and eoPE require more investigation. Importantly, estrogen

signaling is exhibited by both dTcell and dNK1 **(Figure 3.20c)**. On this note, both preclinical and clinical studies show a decrease in 2-methoxyestradiol levels in PE, resulting in declined estrogen levels [472-474]. It might be possible that increased estrogen signaling in decidual immune sub-states act as a compensatory mechanism to increase 2-Methoxyestradiol levels that can potentially activate HIF1α and vascular endothelial growth factor receptors (VEGFR-2) to maintain placental perfusion by increasing angiogenesis [281, 475].

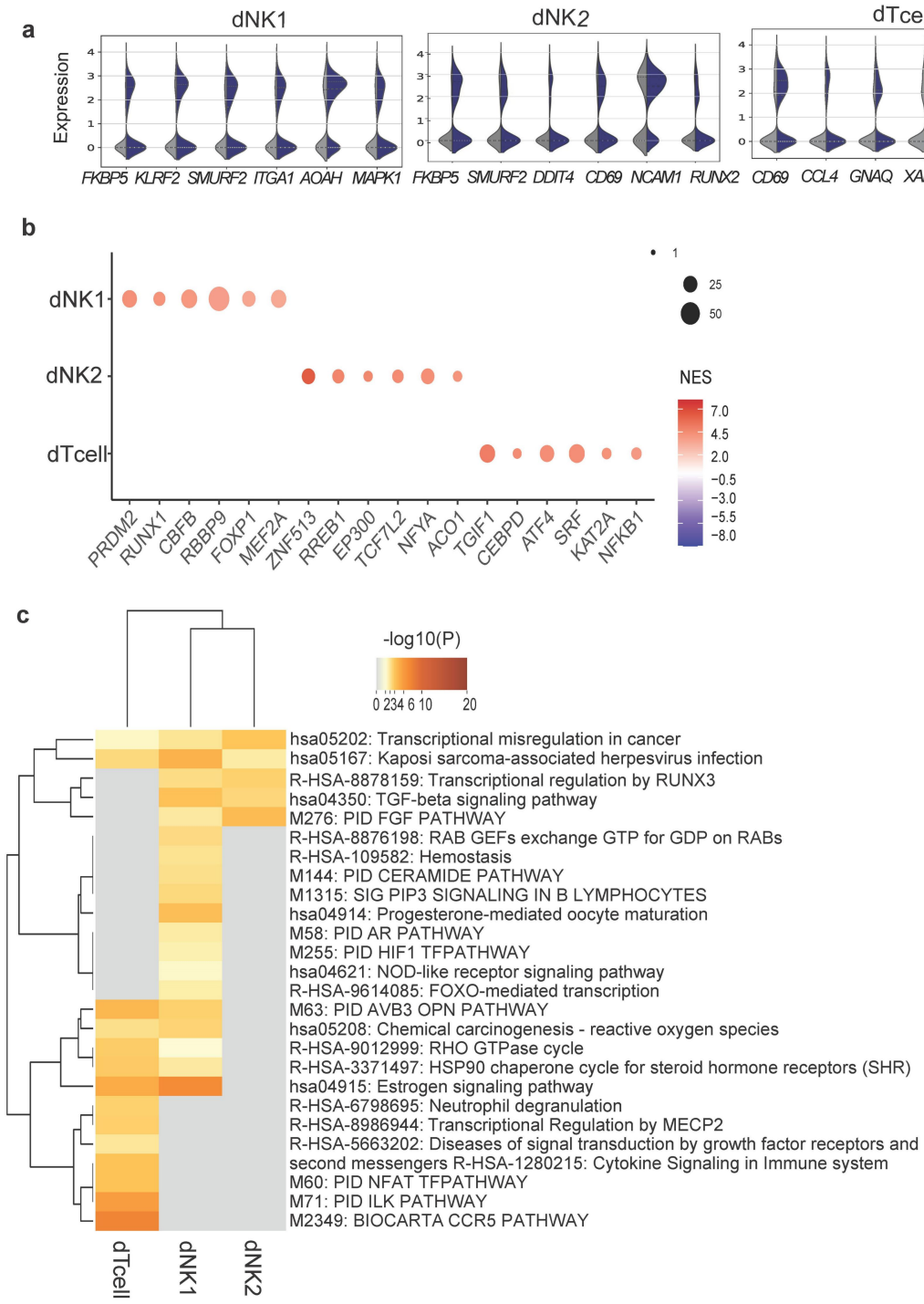


Figure 3.20: Dysregulation of dNK and dTcell in eoPE

(a*) Violin plot depicting normalized expression of selected differentially upregulated genes in dNK1, dNK2 and dTcell in eoPE vs term controls. (b) Dotplot depicting transcriptional activities based on top-100 differentially upregulated targets of dNK1, dNK2 and dTcell (Logistic Regression avg_log2FC > 0.25 & Bonferroni adjusted p-value < 0.01) using iRegulon. The number of targets regulated by a transcription factor is size-coded (see Methods), while the normalized enrichment score (NES) is color-coded. A higher number of regulated targets generally indicates a higher centrality of a transcription factor. (c) Heatmap visualizing differentially enriched pathways in dTcell, dNK1, and dNK2 cell types in eoPE relative to term controls. A discrete color-scale was used to represent statistical significance. Gray color indicates a lack of significance.*

3.10.2 Dysregulation of analogous dMAC and vHBC cell types

Firstly, Logistic Regression analysis revealed important differentially upregulated genes such as, *ANKUB1, DUSP1, IRAK3, C5AR1, MAFB* for dMAC1 and *CTSZ, DUSP1, FOS, SIGLEC1, MAF, GLUL* for dMAC2 sub-states **(Figure 3.21a).** In contrast, vHBC displayed dysregulation of genes such as *PAPPA2, MS5A7, RABGEF1, IFNAR2, CABLES1* and *SPRED1* (**Figure 3.21a**). *DUSP1* was previously found to reduce levels of *p-MAPK1* and *ERK1/2* expression to promote PE [476]. *MAFB* polymorphisms have been linked to variations in lipid levels, coronary disease, and atherosclerosis [477, 478]. Also, *MAFB* gene expression was previously detect to rise in response to stimuli that encourage the polarization of macrophages into the M2 state and facilitate cholesterol efflux [479]. On similar note, high SIGLEC1 expression in pregnant women were associated with increased risk of congenital heart block [480]. Pregnancy-Associated Plasma Protein-A2 (*PAPPA2*) is involved in regulating insulin-like growth factor (IGF) bioavailability by cleaving IGF-binding proteins. In the context of PE, abnormal levels, or dysregulation of *PAPPA2* have been associated with impaired placental development, altered trophoblast function, and disrupted angiogenesis [481]. Additionally, PAPPA2 is considered a potential biomarker for preeclampsia due to its altered expression and activity in affected pregnancies [482, 483].

Secondly, TF and co-regulators acting on differentially upregulated genes in dMono1, dMAC1, dMAC2, and, notably, vHBC analogous to decidual macrophages was decoded using iRegulon (see Methods). In this context, the major TF regulators of dMono1 were found to be constituted by *TCF7L2, PAX3, JUND,* and *BNC1* **(Figure 3.21b).** Conversely, the regulators of dMAC1 were identified as *GMEB2, TBX21, ARNTL, RORC, PRDM2*, and *FOXK1* **(Figure 3.21b).** In the case of decidual macrophages, *GMEB2* might play a role in responding to glucocorticoid signals and influencing the immune response [484]. Albeit it is known that glucocorticoids can induce immunomodulation and affect endothelial function [485], further investigation is warranted to investigate if disrupted glucocorticoid signaling cause endothelial dysfunction in eoPE. dMAC2 was found to be regulated by *POLR2A,*

FOXJ3, FOS, MAF, and *BCLAF1* **(Figure 3.21b).** *FOS* and *MAF*, both being members of the *AP-1* (Activator Protein 1) transcription factor complex, are often known to form heterodimers for the regulation of gene expression. *BCLAF1*, a paternally imprinted transcription factor interacting with anti-apoptotic members of the *BCL2* family, acts as an inducer of apoptosis [486, 487]. Another significant transcription factor regulating dMAC2 is identified as SRF, which has previously been implicated in the migration and phagocytosis of macrophages [488]. Additionally, *SRF* maintains mitochondrial dynamics, regulates fatty acid translocation, and controls the expression of electron transport chain (ETC) complex proteins [489, 490], thereby influencing cell metabolism. *SRF* has also been implicated in the apoptosis regulation by controlling expression of anti-apoptotic molecules like *BCL2*. [491]

Importantly, differential pathway enrichment analysis performed on differentially upregulated genes (Logistic Regression, Bonferroni corrected adjusted p-value < 0.05) revealed conserved and specific signalling programs per cell type (**Figure 3.21c**). dMAC1/2 and dMono1 revealed enriched cytokine signaling via upregulated genes, including *FOS, DUSP6, IRAK3, IFIT2, IFNAR2, IFNGR2, IL6ST, MX1, MX2, TNFRSF11A,* and *ISG15* (Suppl Table). Both dMAC1 and vHBC are enriched for GPCR signaling. Importantly, dMAC2 exhibited robust enrichment for vesicle mediated transport, AP1 pathway, Receptor tyrosine-kinase (RTK) signaling, and Plasma lipoprotein clearance. Of note, RTK signalling is mediated by dysregulated candidates such as *APOE, FOS, FOSB, JUND, DUSP6*, and *FN1* **(Supplementary Table 19)**. Previously, disruptions in RTK signaling pathways were implicated in the pathophysiology of PE, affecting processes such as angiogenesis and endothelial function [492, 493]. Further, *DUSP1, FOS, MAF,* and *MAFB* constitute AP1 pathway **(Figure 3.21 a,c; Supplementary Table 19)**. Noteworthy to mention, dysregulated AP-1 signalling might lead to an imbalance in cytokine production, promoting a pro-inflammatory environment that has been implicated in the endothelial dysfunction and vascular damage associated with PE [494, 495].

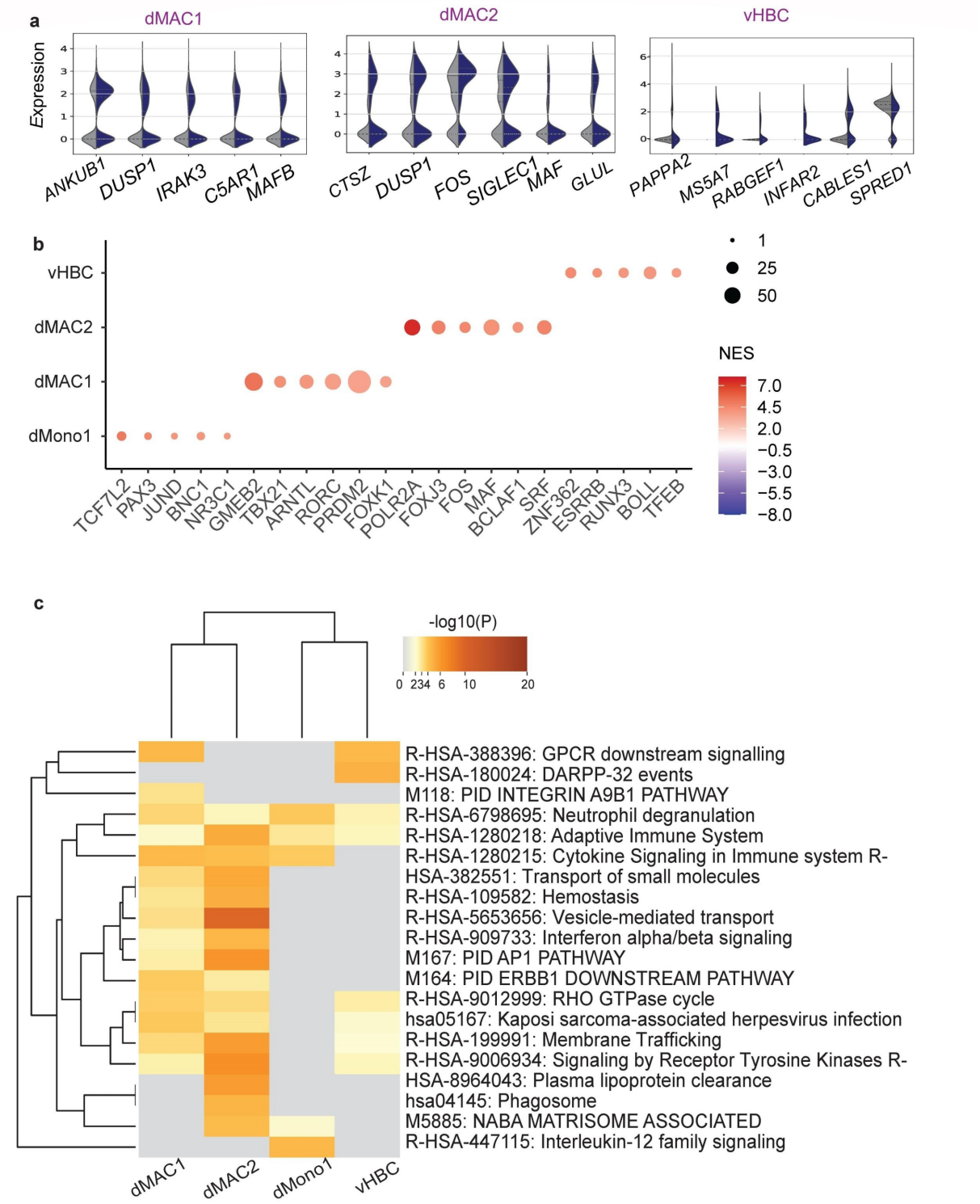


Figure 3.21: Dysregulation of analogous dMAC1/2 and vHBC.
(a) Violin plot depicting normalized expression of selected differentially upregulated genes in dMAC1, dMAC2, dMono1 and vHBC for eoPE vs term controls. (b) Dotplot demonstrating iRegulon inferred transcriptional activities based on top-100 differentially upregulated targets of vHBC, dMAC1/2, and Mono1 (Logistic Regression avg_log2FC > 0.25 & Bonferroni adjusted p-value < 0.01). The number of targets regulated by a transcription factor is size-coded (see Methods), while the normalized enrichment score (NES) is color-coded. A higher number of regulated targets generally indicates a higher centrality of a transcription factor (c) Heatmap visualizing differentially enriched pathways in dMAC1, dMAC2, dMono1 and dHBC1 cell types in eoPE relative to term controls. A discrete color-scale was used to represent statistical significance. Gray color indicates a lack of significance.

3.11 vSTB majorly dysregulated in eoPE

Subsequently, eoPE centric dysregulation within the placental vSTB layer was examined. Despite their shared secretory function and a cytoplasmic surface area in direct communication with maternal blood (approx.12 m2) [354], distinct and shared dysregulated profiles were observed among vSTB nuclei sub-states (uniquely dysregulated: 660 out of 980 STB DEGs, 67.3%). Specifically, 59, 289, and 412 unique DEG(s) were identified in STBjuv, STB1, and STB2 nuclei, respectively **(Supplementary Table 12).** Albeit shared DEG(s) represented a smaller fraction of all differentially expressed disease genes (shared: 320 of 980 DEGs, 32.3%), they displayed a predominantly consistent regulation across all nuclear sub-states (96.25% of 320 genes with concordant regulation between vSTB1, vSTB2, and vSTBjuv sub-states). Importantly, these shared DEG(s) indicated a core dysregulation within the vSTB compartment, with only a few exceptions such as *FTX, CHODL*, and *CSH2* (accounting 3.75%, 12 of 320 genes) that were shared across all vSTB sub-states but exhibited different directions of dysregulation **(Supplementary Table 12)**.

At first, vSTB1 and vSTB2 emerged as the top two most dysregulated cell types/states within the placenta in terms of the identified number of DEG(s) in eoPE vs term controls. However, since they are also the most prevalent nucleus types, downsampling was considered based on the upper quartile (Q3) or 75% of villi cell type cell numbers. Albeit reducing the total number of DEG(s) per sub-state, this resulted in the confirmation of key drivers of eoPE, including *FLT1, PAPPA2, LEP, SAT1, DUSP1, ENG*, and *SH3PXD2A.*

Transcriptomic profiles uniquely downregulated in the STB1 sub-state were associated with lipid biosynthetic processes (GO:0008610; q-value = 0.016595869) and the regulation of RNA splicing (GO:0043484; q-value = 0.046773514) **(Supplementary Table 14).** In contrast, the uniquely upregulated genes in STB2 exhibited enrichment for ubiquitin-mediated proteolysis (hsa04120; q-value = 0.005128614) **(Supplementary Table 14),** a crucial mechanism in cellular protein load regulation.

3.11.1 *EP300* central to vSTB fusion perturbation

Placental fusion is a fundamental regulator of vCTB differentiation to vSTB (as illustrated in **Figure 3.22a**) — hence, disruptions in fusion mechanism is expected to significantly contribute to the overall eoPE pathogenesis [446]. To understand the molecular mechanisms behind dysregulated fusion, the transcription factors/co-regulators acting on the differentially

regulated vSTB targets, previously inferred by Logistic Regression analysis (see Methods), were decoded. To be precise, iRegulon was applied to dissect the transcriptomic regulatory landscape of targets perturbed in at least two vSTB sub-states. Notably, the transcriptional co-activator *EP300* (or, p300) emerged as the most enriched regulator acting on n = 133 dysregulated vSTB targets and showing normalized enrichment score = 4.585, FDR < 0.001 **(Figure 3.22b; Supplementary Table 15).** p300/EP300 is known to be associated with cell fusion [512] and cell cycle arrest [513] — therefore, enhanced nuclear activity of this transcriptional co-activator likely causes disruptions in trophoblast fusion processes leading to aberrant vSTB differentiation. Apart from p300/*EP300*, a significant enrichment of transcription factors such as *FOXO1, SCRT2, FOXO4, CHD1, FOS, PAX5, UBXN1,* and *PAX2* was revealed by iRegulon analysis **(Figure 3.22b)**.

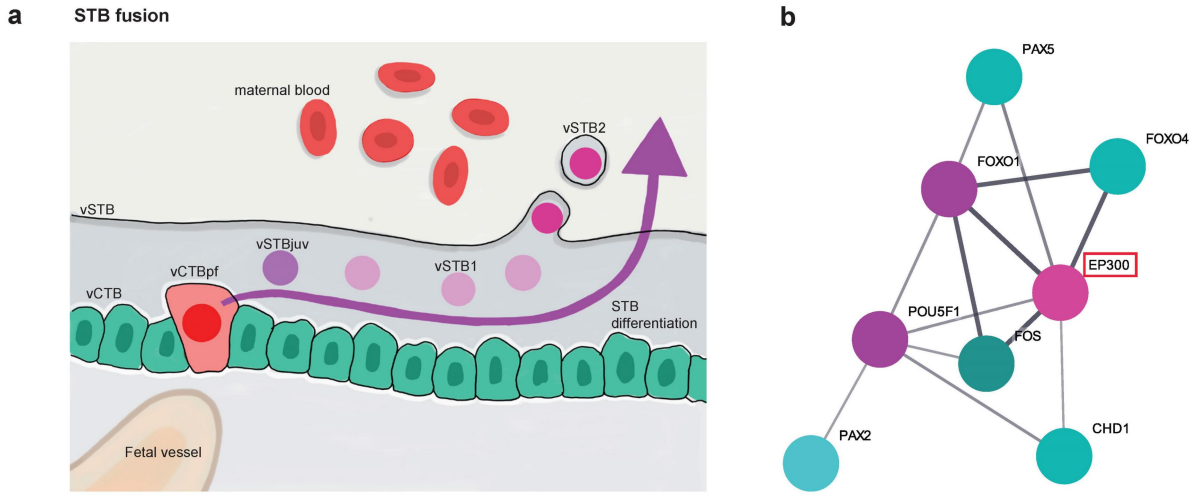


Figure 3.22: p300 is central to STB dysregulation in eoPE.
(a) Schematic representation of syncytial barrier formation in the human placenta. Cells in the cytotrophoblast monolayer fuse (in red) to generate the outer multinuclear syncytiotrophoblast (STB) with distinct nuclear transcriptomic states across the lineage differentiation trajectory (denoted by purple arrow). The STB layer provides the interface for nutrient transport and gas exchange in pregnancy and is a key endocrine cell type. Figure adapted from Nonn, Debnath & Valdes et al. 2022 [363] (unpublished) and re-sketched using Adobe InDesign software. (b) Predicted transcription factors and regulators inferred from placental barrier dysregulation, where motifs of dysregulated shared STB-genes were used to predict its upstream transcription factors (threshold normalized enrichment score i.e., NES > 3, FDR on motif similarity < 0.001); EP300 is the most dysregulated candidate with the highest normalized enrichment score (boxed in red).

To investigate the potential perturbation of fusion-associated genes in the vSTB sub-states, a list of 27 fusogenic genes relevant to syncytium formation **(Supplementary Table 15)** was curated. Remarkably, the shared differentially expressed gene(s) across the three vSTB sub-

states revealed that 18.8% of p300 targets were linked to the fusion process, encompassing genes such as *GCM1, ADAM12,* and *DYSF* (5.6-fold enrichment, hypergeometric p-value = 5.602047e-05, over-enrichment test) **(Figure 3.23a).** *GCM1*, a marker for trophoblast fusion, is crucial for vSTB differentiation [446]. *ADAM12* is implicated in cell adhesion and migration processes within the placenta [444, 445]. On the other hand, *DYSF* plays a role in membrane repair- mainly, at sites where syncytial knots are detached from vSTB surface [496].

Additionally, the p300 targets included previously known candidates for PE progression including *ENG, SAT1, FOS, SERPINE1, GADD45G, INHBA*, and *HTRA1* **(Figure 3.23a)** that are perturbed in both vSTB1 and vSTB2 sub-states. Dysregulation of *ENG* has been associated with impaired angiogenesis and vascular dysfunction, contributing to the pathogenesis of preeclampsia [497-499]. *SAT1* is involved in polyamine metabolism and may play a role in cellular growth and differentiation [500]. Of note, *FOS* is a transcription factor involved in cell proliferation, differentiation, and apoptosis [501]. As mentioned before, *FOS* appeared to be one of the major transcriptional regulators of dysregulated STB genes, as inferred by iRegulon **(Figure 3.22b).** Altered *SERPINE1* expression contributes to abnormal coagulation and fibrinolysis observed in PE [502]. Dysregulation of *GADD45G* may impact trophoblast cell cycle control and DNA repair mechanisms, given its pro-apoptotic effects [503]. Importantly, *INHBA* and *SERPINE1* constitute canonical TGF-β/BMP signaling targets and were previously implicated in the pathogenesis of PE and abnormal placental development [504-507].

Subsequently, a network analysis [162] was performed to understand the connectivity of the *p300* regulated targets and decomposed into community subgraph structures, from which the first largest community ‘C1’ is shown **(Figure 3.23b).** This is grounded on the idea that each community consists of nodes (targets) that are tightly intra-connected and loosely inter-connected with the same from other communities [163, 164]. This is influenced by the fact that members of each community participate in similar signaling pathways and programs. For instance, *ATF3, DUSP1, FOS, FOSB, JUNB, JUND* are implicated in AP1 pathway **(Supplementary Table 20)** and form a closely knit community subgraph in the network analysis of dysregulated targets **(Figure 3.23b).**

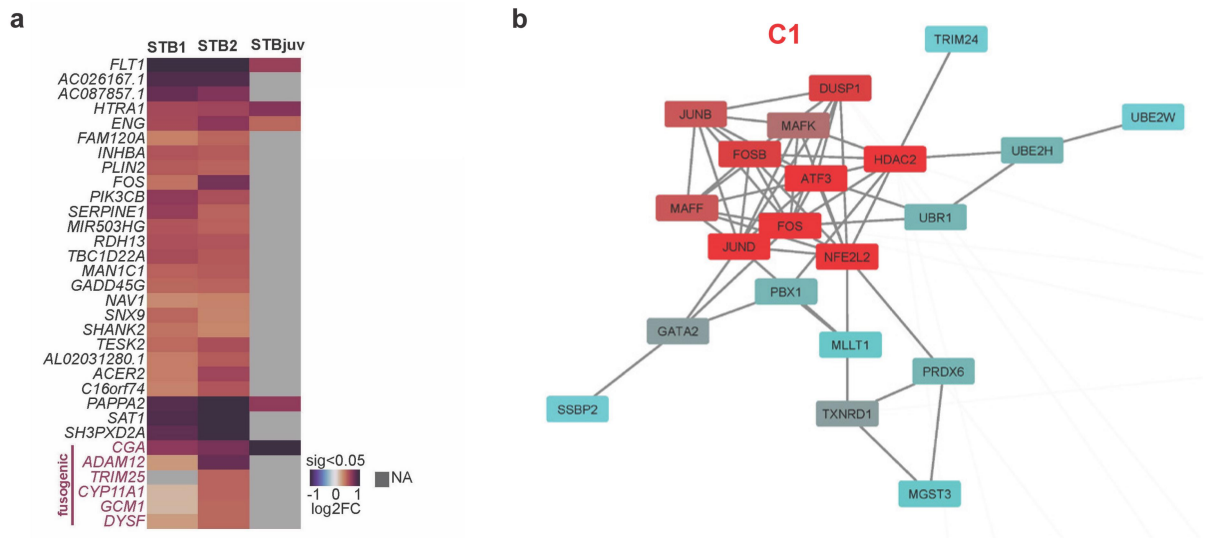


Figure 3.23: p300 regulated vSTB targets perturbed in eoPE.
(a) Heatmap depicting log2 fold changes of p300 related targets dysregulated in the placenta barrier in eoPE (≥ 2 STB nuclei states), genes involved in fusion marked in purple. Logistic regression was used for differential testing with a min. log2FC 0.4, genes expressed in at least 30% of nuclei and Bonferroni adjusted p-val < 0.05. (b) Largest community subgraph (C1) derived from network analysis of EP300 dysregulated targets. Subgraph communities are inferred by Glay algorithm based on Newmann-Girvan's edge betweenness [164].

3.11.2 Shared dysregulated targets drive hypoxia in eoPE

Of note, a substantial fraction of vSTB perturbed targets is found to be jointly regulated by multiple TF/co-regulators inferred by iRegulon. Specifically, a significant overlap of *p300/EP300* targets co-regulated by *FOXO1, FOXO4, FOS,* and *PAX5* is revealed through a pairwise comparison of TF-targets, using Jaccard's index **(Figure 3.24a).** To understand the functional implications of these common targets in eoPE pathogenesis, a pathway enrichment analysis of the same regulated by at least 5 TF/co-regulators was performed. Here, significant enrichment of several signaling cascades comprising of HIF1A, AP1, Phosphatidylinositol (PI3K-Akt), and TGFβ pathways is observed **(Figure 3.24b).** For example, dysregulated candidates like *PLIN2, ENG, FOS, LEP, FURIN* and *NDRG1* **(Supplementary Table 20 & Extended Data Figure 7)** are found to be cross-regulated by Hypoxia-inducible factor-1 alpha (HIF-1α) transcription factor. Increased levels of *HIF-1α* in PE affected placentas were reported in previous studies [508, 509], leading to the induction of soluble vascular endothelial growth factor receptor-1 (sFLT-1), a key factor in the pathogenesis of preeclampsia and intrauterine growth restriction (IUGR).

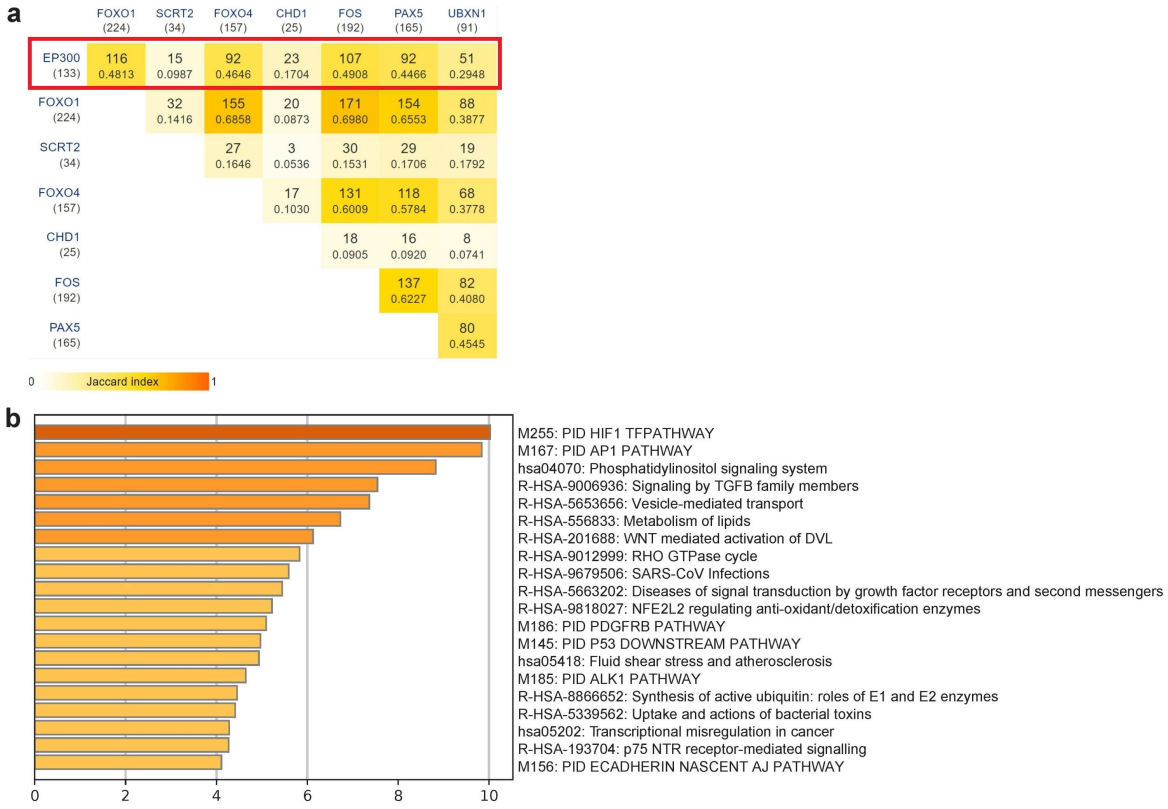


Figure 3.24: Shared dysregulated vSTB targets drive hypoxia in eoPE.
(a) Overlap analysis depicting the absolute number of shared targets, and significance of overlap using Jaccard index among iRegulon inferred transcription factors and regulators- EP300, FOXO1, SCRT2, FOXO4, CHD1, FOS, PAX5 and UBXN1. The number of targets for each TF or a regulator is mentioned in a bracket and the overlap of EP300 targets with the rest are boxed in red. (b) Enrichment heatmap illustrating the enriched signalling pathways and processes of dysregulated vSTB targets that are regulated by at least 5 TF/regulators as inferred by iRegulon.

3.11.3 vSTB developmental drivers dysregulated in eoPE

One of the important aims of this study was to investigate the impact of eoPE on the cellular differentiation of vCTB to vSTB, and to decipher the specific stage at which cells affected by eoPE deviate from their normal developmental trajectory. To accomplish this, an overlap analysis was conducted, comparing DEG(s) associated with eoPE in all vSTB sub-states and vSTB transition markers inferred from early trophoblast trajectory. The analysis revealed a significant enrichment of eoPE DEG signatures in transition genes associated with vSTB development (enrichment = 29.96-fold, p-value = 0.002, Exact hypergeometric test & **Supplementary Table 16**) as depicted in **Figure 3.25.**

However, it was acknowledged that the composition of vSTB1 and vSTB2 was significantly higher than that of vSTBjuv. Therefore, vSTB1 and vSTB2 were randomly downsampled to match the number of vSTBjuv nuclei, ensuring consistent statistical power for DEG analysis. Next, the DEG marker identification was repeated for vSTB1 and vSTB2. Of note, the significant enrichment of vSTB transition genes within eoPE DEG signatures persisted post downsampling (enrichment = 44.94-fold, p-value = 0.0008, hypergeometric test) **(Supplementary Table 16; transition_eoPE_overlap)**.

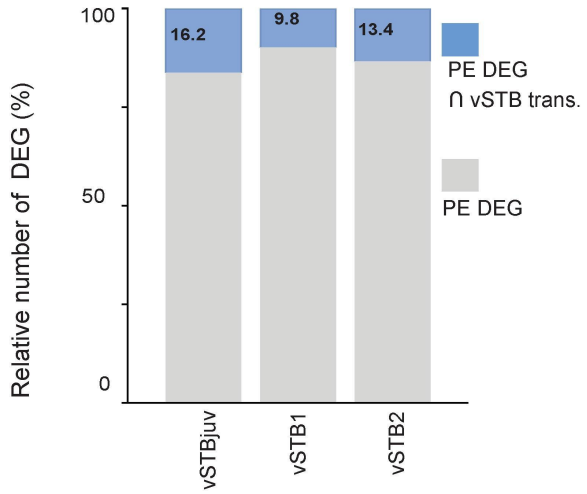


Figure 3.25: vSTB differentiation drivers significantly dysregulated in eoPE.
Relative number of differentially expressed genes (DEGs) per STB nuclei states in eoPE that overlap with variable genes across the differentiation trajectory (STB transition). Overlap highlighted in blue, with exact percentage contribution written.

Additionally, a significant overlap was observed between CTB transition genes and eoPE DEGs in CTB (hypergeometric p-value = 1.32e-12; enriched 15.94-fold compared to expectations) **(Supplementary Table 16).** Taken together, these results implicate that dysregulation arising early in trophoblast development might be involved in the eoPE pathogenesis, by affecting trophoblast fusion dynamics.

3.12 Dysregulated SASP(s) translates eoPE from fetus to mother

The next major aim of this thesis was to determine whether the localized senescence observed in eoPE within the trophoblast interface, connecting fetal circulation and maternal blood, resulted in the release of senescence-associated secretory phenotype (SASP) factors

into the maternal circulation. Among the shared DEGs in vSTB sub-states, 12.2% were identified as SASP encoding genes, revealing a significant 1.94-fold enrichment (39/320 genes; hypergeometric p-value 6.02e-05; **Figure 3.26a & Supplementary Table 18)**. Overall, SASP genes constituted 12% of all dysregulated STB genes (1.91-fold enrichment, hypergeometric p-value 5.75e-13) **(Figure 3.26a & Supplementary Table 18).**

3.12.1 p300 significantly regulates perturbed SASP targets in eoPE

iRegulon analysis previously implied that the normal fusion of vCTB is required for vSTB differentiation that is disrupted by enhanced activity of *p300/EP300* in eoPE. Correspondingly, senescence-associated p300 targets were found to be dysregulated in all three vSTB nuclear states in eoPE, as depicted in the heatmap **(Figure 3.26b).** The signatures of senescence were particularly enhanced in vSTB1 and vSTB2, representing the terminal nuclear states along the STB differentiation trajectory path. Given the dysregulation of relevant SASP factors, it is postulated that *p300/EP300* likely induces premature senescence in the vSTB nuclear states, ultimately leading to accelerated placental aging in eoPE.

Prior investigations have demonstrated that two SASP encoding genes, Activin A encoding *INHBA* [504-506], and GDF15 [514, 515], hold potential for predicting PE at 36th week. Both factors belong to the secreted TGFβ superfamily, with GDF15 serving as a recognized marker for the multinucleated vSTB barrier facing maternal blood. Activin A is a known SASP target and functions as an activin-pathway ligand.

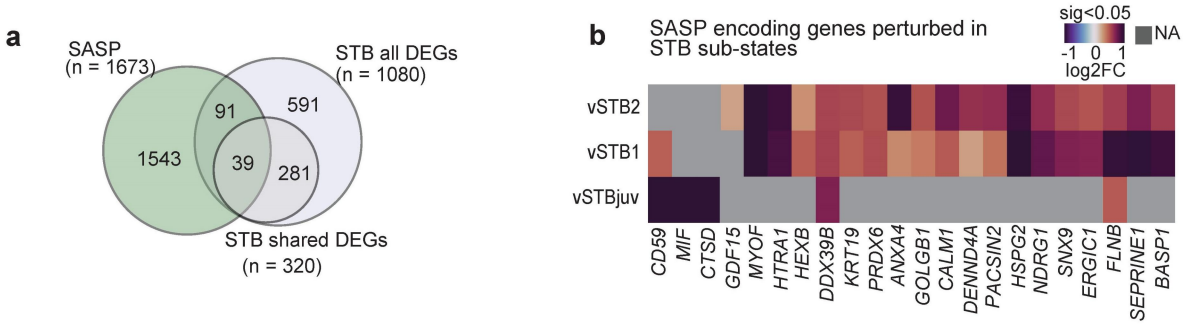


Figure 3.26: SASP(s) substantially dysregulated in eoPE vSTB sub-states.
(a) Overlap of STB dysregulation (≥2 nuclei states, logistic regression used for differential testing with a min. log2FC 0.4, genes expressed in ≥30% of nuclei and adjusted p-value <0.05.), human secreted genes from the human Secretome database (SPRomeDB) [510], senescence-associated genes from the Molecular Signatures Database (MSigDB) and senescence-associated secretory phenotype atlas (SASP, Buck Institute) [511]. (b) Heatmap illustrating log2FC of SASP associated genes dysregulated in the placenta barrier in eoPE (≥ 2 vSTB nuclear states). Logistic regression used for differential

testing with a min log2FC 0.4, expressed in ≥ 30% of nuclei and Bonferroni adjusted p-value <0.05 per gene. STB, syncytiotrophoblast; juv, juvenile.

3.12.2 SASP mediated disrupted cellular communication at maternal-fetal interface

Firstly, the cellular communication between dDSTB and maternal decidua was investigated. Specifically, ligands secreted by dDSTB that could bind to their corresponding receptors in the maternal endothelium (dVEC, dSMC) are depicted **(Figure 3.27a).** It is important to mention that dDSTB was found only in the eoPE condition **(Supplementary Table 2)**- hence, the R-L interactions profiled are very likely disease-centric **(Supplementary Table 13; dDSTB)**.

Notably, emphasis was placed on modelling cellular communication across the maternal-fetal interface using receptor-ligand interactions to simulate the transmission of signals from the fetus to the mother, considering the secretion of SASP encoding ligands **(Figure 3.27b & Supplementary Table 13; vSTB).** To accomplish this, dVEC and dSMC cell-types were utilized as a proxy of maternal vessels from the snRNA-seq decidua data. The aim was to understand how the dysregulated SASP ligands expressed in STB nuclei sub-states in eoPE could translate the placental perturbations to affect the maternal decidual endothelium and, subsequently, the systemic circulation **(Figure 3.27b).** By employing Connectome [156] and various databases implemented in LIANA framework [157] for modelling R-L network, the communication between dysregulated secreted factors originating from vSTB and decidual vascular cells (dVEC, dSMC) was investigated. This analysis highlighted specific secreted ligands like *LEP*, *EBI3* and SASPs including *TGM2*, *HSPG2, ADAM9, GDF15* and *INHBA* (log2FC > 0.25, with a cut-off of genes detected in at least 10% of diseased cells and an adjusted p-value < 0.01). Observations revealed that the interactions between vSTB-dVEC and vSTB-dSMC were predominantly associated with SASP genes (67%, 6 out of 9), which notably included INHBA and GDF15.

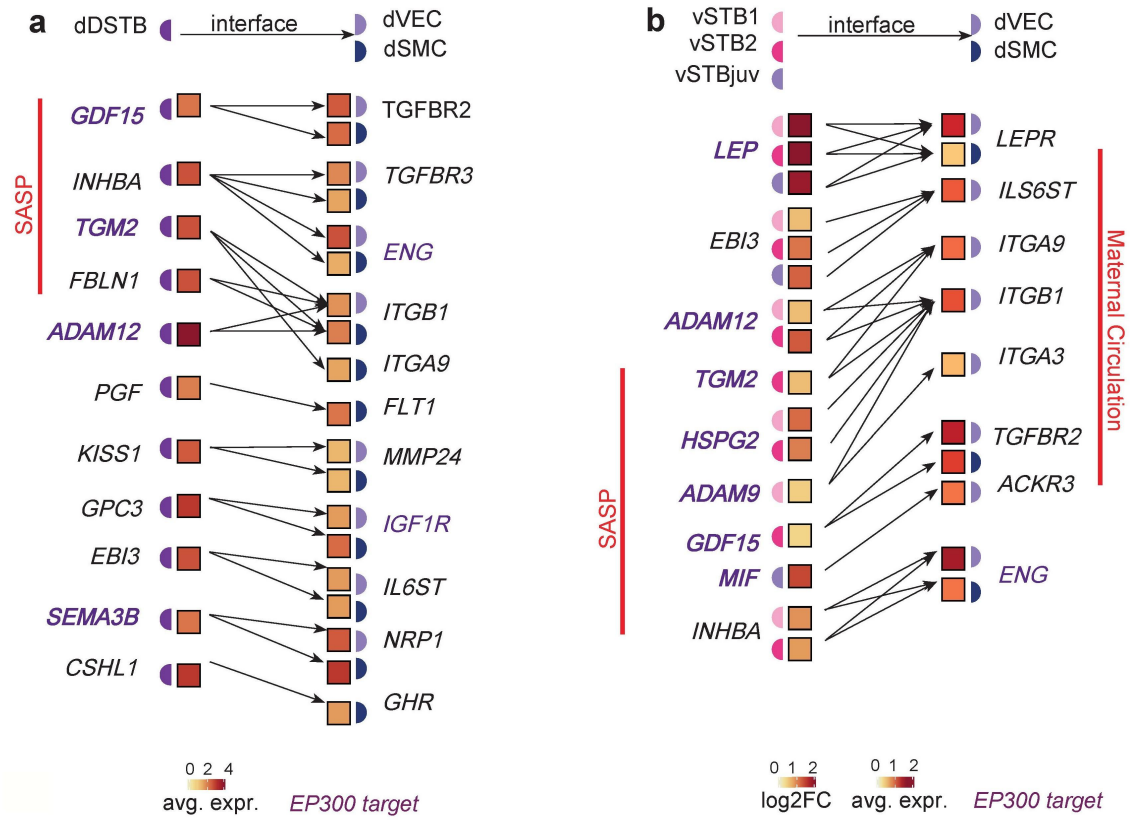


Figure 3.27: Elevated SASP mediates dysregulated cellular-crosstalk across maternal-fetal interface.
(a) Receptor-ligand interaction analysis depicting ligands expressed by dDSTB and its potential interaction partners in decidua endothelium receptors (dVEC, dSMC). Results are restricted to eoPE condition. (b) Plot demonstrating secreted ligands in vSTB altered in eoPE relative to term controls with preterm gestational age correction acting on highly expressed decidual endothelial receptors highlight ligand pressure, i.e., increased ligand expression with unaltered receptor expression, at the maternal-fetal interface. Receptor-ligand interaction pairs are shown (Wilcoxon rank-sum test, p<0.05; identified by multiple tools and databases including Connectome [156], and SingleCellSignalR, NATMI, CellphoneDB, CellChat as implemented in LIANA framework [157]) using arrows from ligands to receptors. Dot colors represent cell states/types, ligand squares illustrate average log2FC between conditions (only upregulated candidates are included from logistic regression; log2FC>0.25 and p-value<0.01 after Bonferroni correction) and receptors squares encode average expression. Figure style adapted from Nonn, Debnath & Valdes et al. 2022 [363] (unpublished).

Of note, Leptin (*LEP*) is a central regulator of metabolism and has been linked to endothelial dysfunction. In PE, elevated levels of *LEP* are associated with abnormal placental development and vascular complications, contributing to pathophysiology [516, 517]. Importantly, serum *LEP* was found to be a prominent biomarker distinguishing eoPE from late-onset PE [516]. Epstein-barr virus induced 3 *(EBI3)* is associated with immunomodulation and inflammatory responses, and a previous study by Guo et. al. 2020

[518] detected 2X fold increased circulating *EBI3* in eoPE relative to late-onset PE. Furthermore, this study observed a positive correlation between plasma *EBI3* levels with clinical parameters, like uric acid and proteinuria that suggested elevated *EBI3* as a highly sensitive biomarker for eoPE [518]. Perlecan (*HSPG2*) is a key component of the extracellular matrix and plays a role in vascular development [519]. Transglutaminase (*TGM2*) has multifactorial roles in ECM remodelling, apoptosis, oxidative stress response and inflammation [520]. Importantly, a previous study based on mouse model [521] described *TGM2* as an essential link between inflammation and hypertension. Elevated *GDF15* expression was previously detected in women very likely to develop PE or already with this disease [515]. This observation was valid in several international cohorts [515, 522] and related to endothelial dysfunction, inflammation, and overall vascular complications observed in the disorder. Of note, *ADAM9* and *ADAM12* are involved in cell adhesion and migration. A recent study by Liu et al. 2023 [523] proved that *ADAM9* upregulation considerably suppressed proliferation, migration, and EMT progression in trophoblasts via *USP22* mediated de-ubiquitination.

To compare the spatial distribution of senescence within the vSTB layer between healthy and eoPE placentas, a high-resolution spatial transcriptomics method, in-situ sequencing (ISS) [185], was employed. This analysis was performed to interrogate the implication of elevated SASP(s) like *INHBA* on placenta vascular biology **(see acknowledgement; Figure 3.28)**. The maternal-fetal interface, crucial for nutrient and oxygen exchange, involves the crossing of molecules from maternal blood through the syncytium and trophoblast wall, comprising vCTB and vSTB. Spatial regulation in this region is pivotal for effective mother-to-fetus communication. In the high-resolution approach, an ISS panel detecting 174 genes characterizing the main villous cell types based on markers identified in the single-nucleus RNA sequencing (snRNA-seq) was established (see Methods).

Utilizing plankton.py [524], signals were captured, decoded to x and y coordinates, and assigned to specific cell types or states based on RNA signals. Villi wall location was detected using protein staining signal intensity, after applying a Gaussian filter to integrate local signal intensity. Subsequently, all transcripts associated with the villus wall were identified- precisely, mRNA molecules were differentiated between within and outside villi regions based on the detected local protein density (see Methods). Thereafter, the positioning of senescence associated *INHBA* transcripts within the vSTB layer relative to markers associated with fetal vessels *(CDH5, IDO2, KDR, TEK, ZEB1)* was evaluated. Fetal vessels were localized within a 5 µm radius of the villus wall, near the trophoblast barrier, facilitating direct nutrient and oxygen exchange. To visually investigate the spatial

relationship between senescence and vascularization in both term control and eoPE tissue samples, molecules associated with the 'wall' were plotted on a DAPI stain rendering, color-coded based on their gene assignments for 'vascularization' and 'senescence,' with other molecules represented in white **(Figure 3.28a).** The study found that in the term control sample, senescence markers (like *INHBA*) were situated far away from vessel markers (*KDR*) as compared to eoPE sample. This finding was statistically analyzed by categorizing gene molecules based on their proximity to vascular markers and conducting a binomial test to evaluate gene distributions. Two categories, vessel proximal and vessel distal, were defined (refer to Methods for details). According to the null hypothesis, gene distribution should be uniform across both categories. Subsequently, for each gene, a p-value indicating deviation from the null hypothesis was determined **(Figure 3.28c & Methods).** Of note, the occurrence of senescence markers *INHBA* and *MMP11* were found to be considerably lower in the vicinity of placenta vessel markers in term control samples (both p-values <0.01, representing the lowest p-values among all genes examined). However, this pattern was lost in the eoPE sample **(Figure 3.28c).** In eoPE sample, *INHBA* in the villi wall appeared to be closer to fetal vessels compared to term controls **(Figure 3.28 a-c).** This spatial disorganization in the villous wall, characterized by increased senescence in areas of maternal-fetal crosstalk, may disrupt the exchange of nutrients, potentially contributing to fetal growth restriction (FGR) associated with eoPE.

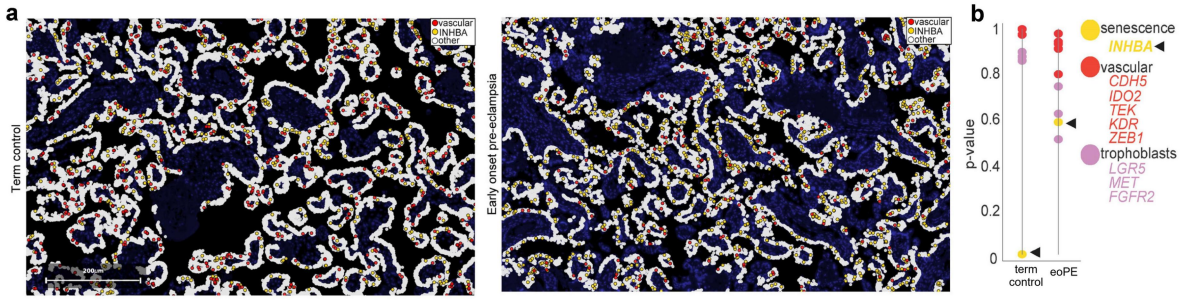


Figure 3.28: In-situ transcriptomic analysis reveals proximity of SASP associated INHBA to fetal vessels in eoPE placenta.

(a). Topography of wall molecules in the term control and early-onset pre-eclamptic (eoPE) samples over a DAPI stain background. n = 2 (1 slide per group). Yellow circles indicate INHBA molecules, red circles indicate vascular marker molecules, and all other molecules in the wall are rendered in white. Sensescence-associated INHBA expression is spatially variable in ISS data, and (b). significantly far from the placenta vessel markers in term controls relative to eoPE. Acknowledgement: Sebastian Tiesmeyer (Digital Health Center, BIH & Charite) performed the analysis under the supervision of Dr. Naveed Ishaque. I aided in the cell typing analysis of ISS. Data figures used with permission to support the significance of SASP on placental vascular biology.

3.13 Disrupted R-L interactions within decidua and villi

To explore disrupted cellular communication within the decidua and villi, R-L interaction analysis was leveraged to model dual ligand/receptor increases or decreases, aligning with edge activation or deactivation, respectively. Specifically, edge activation suggests a robust association with eoPE, as both receptors and ligands exhibit differential upregulation compared to term controls. A schematic of potential R-L interactions governing perturbed cellular crosstalk in both tissues is depicted in **Figure 3.29.**

Apolipoprotein E (*APOE*) is strongly secreted by decidual smooth muscle cells (dSMC) and dMAC2 (log2FC= 1.1656 & 1.0445, Logistic Regression Bonferroni adjusted p-value < 0.05) that could bind to receptors *LRP1* and *SORL1* in DSC1 and dMAC1, respectively, exhibiting a classic case of edge activation **(Figure 3.29a)**. Of note, a multifaceted role for *APOE* in endothelial function, inflammation, vascular remodeling, and lipid metabolism is suggested, all of which are relevant to the PE pathophysiology [525, 526]. A prior study concluded that *ApoE*-knockout mice recapitulated pathologic process of PE- possibly mediated by *TLF4* and *sFlt-1* [525]. Another study based on untargeted proteomics found increased levels of *APOE* in late-onset PE, while it was marginally increased in eoPE [526]. In this regard, it is well accepted that *APOE* is involved with oxidative stress and mitochondrial dysfunction, which are also key factors in PE pathophysiology [527]. *FN1* is moderately upregulated in eoPE that possibly interacts with several receptors in DSC1, dLEC, and dNK1. *FN1* serves as a key player in the interplay between macrophages and the extracellular matrix, influencing multiple aspects of macrophage function in immune responses, tissue repair, and homeostasis [528]. *FN1* was also reported to mediate autophagy and apoptosis of human umbilical vessels [529], and hence, might contribute to PE associated maternal systemic inflammation. On similar note, *FN1* ligand is found to be dysregulated in vCTB and vMC **(Figure 3.29b)** that could potentially interact with *C5AR1* receptor differentially upregulated in vHBC and vPAMM. *FN1* is a glycoprotein that plays a vital role in cell adhesion, migration, and wound healing, while *C5AR1* is a receptor for the complement component C5a, a key mediator of the immune response. Also, in case of dEVT R-L analysis in eoPE, FN1 emerged as a central ligand binding to multiple decidual receptors such as integrins *(ITGA3, ITGA4, ITGA5, ITGAV)* across several cell types/states **(Extended Data Figure 8).** Previous data suggested that overexpressed *FN1* promotes PE pathogenesis by contributing to apoptosis and autophagy, associated with the PI3K/AKT/mTOR signaling pathway [529]. While this observation was described in human umbilical vessles, the same might hold for other cell types like and would be a matter of investigation.

Importantly, *CD36* (log2FC=0.444) is a differentially upregulated receptor in dVEC that could potentially bind to several elevated *COL1A1* and *THSB1* ligands **(Figure 3.29a)**. Of note, CD36 is involved in the recognition and clearance of oxidized low-density lipoprotein (oxLDL) and is associated with oxidative stress and inflammation [530]. Elevated oxidative stress and inflammation are key features of PE and can contribute to endothelial dysfunction. *CD36* is further implicated in endothelial cell activation that could lead to impaired vasodilation, altered vascular permeability, and increased blood pressure in stroke [531]- which might be relevant for eoPE. Furthermore, *CD36* plays a role in fatty acid metabolism, thrombosis, and coagulation pathways [532]. In PE, there is an increased risk of abnormal blood clotting, and *CD36*-mediated processes might influence these hemostatic changes. As CD36 serves as a receptor for THBS1, numerous studies have illustrated its capability to impede the activation of the cAMP signaling pathway [533]. Notably, in conditions such as cardiovascular diseases and cancer, *THBS1* hinders the response of vascular smooth muscle cells by modulating cAMP and cGMP [534]. Additionally, increased collagen signaling via DSC1 may contribute to vascular changes observed in eoPE, influencing endothelial function and vascular remodelling [535]. Overall, these alterations in the vascular system can lead to hypertension, a hallmark of eoPE, and compromise maternal health.

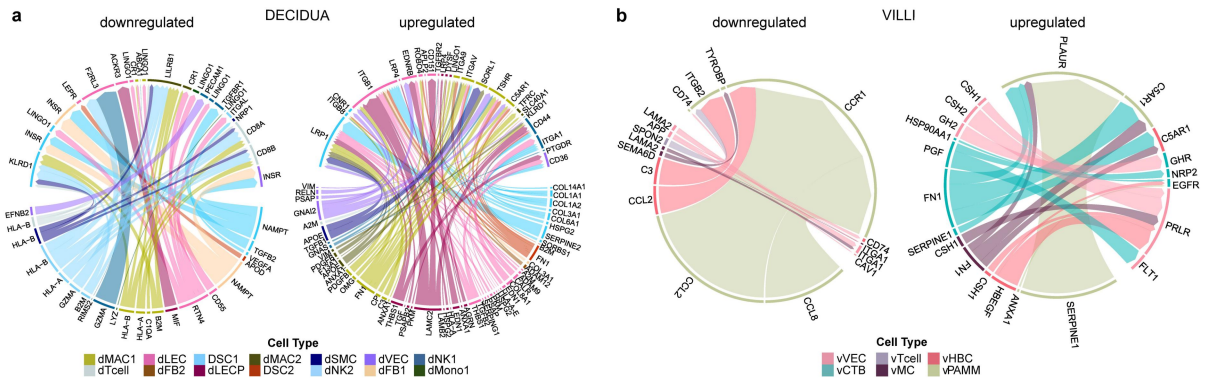
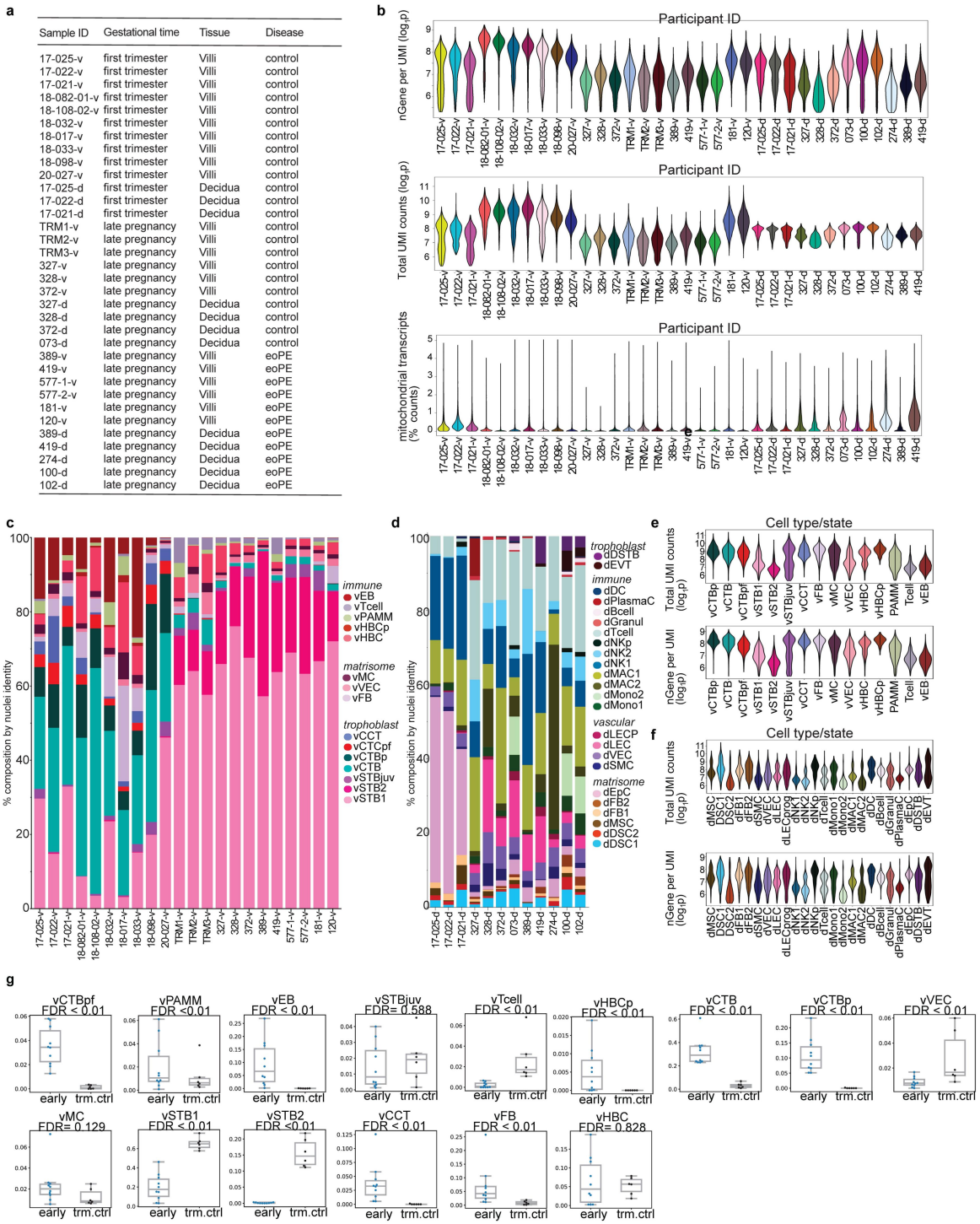


Figure 3.29: Global dysregulation in decidua and villi tissues in early onset preeclampsia modelled by differential receptor-ligand interactions.
(a,b) Overview of selected receptor-ligand interaction network revealing dysregulation during early onset pre-eclampsia (eoPE) within immune and vascular matrisome compartments in decidua (a) and villi (b). Interactions were analyzed using Connectome and differentially expressed receptors/ligands were profiled using multivariate Logistic Regression (p-values adjusted for multiple testing). Ribbon color represents the respective source (or, ligand sending) cell types/states, and is proportional to the edge-weight score (product of normalized LFC of both partners). Networks shown are limited to edges where both receptor and ligand log2FC ± 0.25 between conditions (positive indicates upregulated in eoPE and vice-versa). The minimum percentage of cell groups expressing a receptor (or, ligand) is 10%.

3.14 Extended Data Figures

a

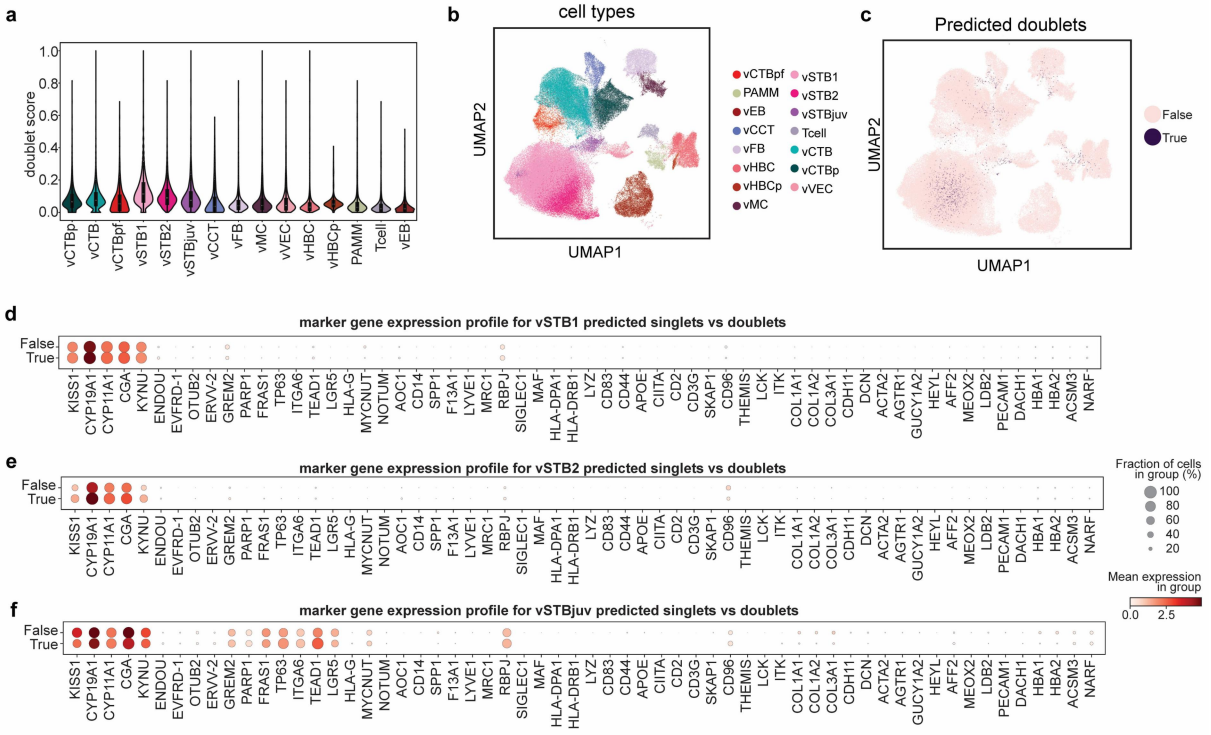
Sample ID	Gestational time	Tissue	Disease
17-025-v	first trimester	Villi	control
17-022-v	first trimester	Villi	control
17-021-v	first trimester	Villi	control
18-082-01-v	first trimester	Villi	control
18-108-02-v	first trimester	Villi	control
18-032-v	first trimester	Villi	control
18-017-v	first trimester	Villi	control
18-033-v	first trimester	Villi	control
18-098-v	first trimester	Villi	control
20-027-v	first trimester	Villi	control
17-025-d	first trimester	Decidua	control
17-022-d	first trimester	Decidua	control
17-021-d	first trimester	Decidua	control
TRM1-v	late pregnancy	Villi	control
TRM2-v	late pregnancy	Villi	control
TRM3-v	late pregnancy	Villi	control
327-v	late pregnancy	Villi	control
328-v	late pregnancy	Villi	control
372-v	late pregnancy	Villi	control
327-d	late pregnancy	Decidua	control
328-d	late pregnancy	Decidua	control
372-d	late pregnancy	Decidua	control
073-d	late pregnancy	Decidua	control
389-v	late pregnancy	Villi	eoPE
419-v	late pregnancy	Villi	eoPE
577-1-v	late pregnancy	Villi	eoPE
577-2-v	late pregnancy	Villi	eoPE
181-v	late pregnancy	Villi	eoPE
120-v	late pregnancy	Villi	eoPE
389-d	late pregnancy	Decidua	eoPE
419-d	late pregnancy	Decidua	eoPE
274-d	late pregnancy	Decidua	eoPE
100-d	late pregnancy	Decidua	eoPE
102-d	late pregnancy	Decidua	eoPE



Extended Data Figure 1: Quality control and characteristics of the single nuclei samples used for analysis.

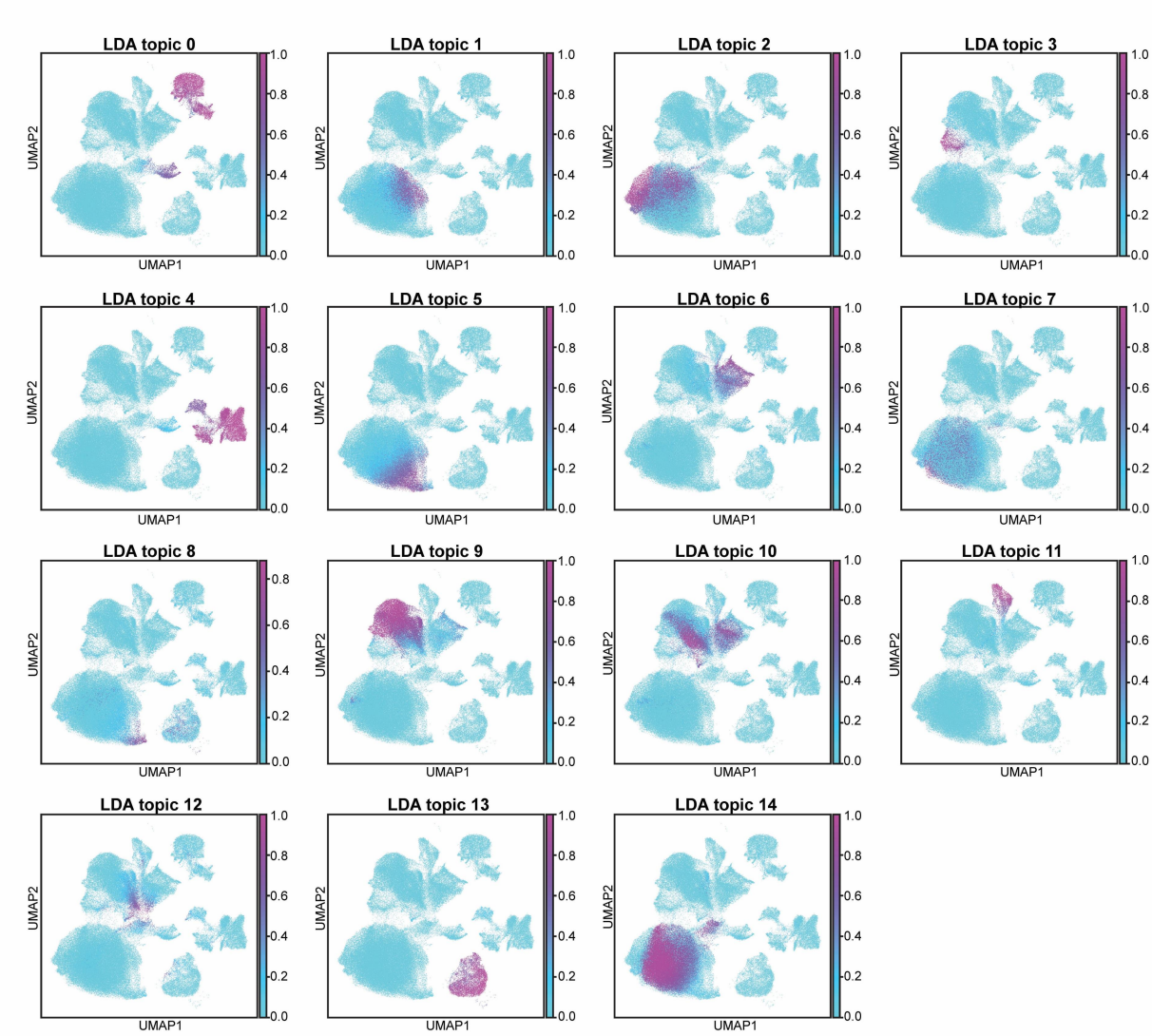
(a) Pseudonymised patient-ID of maternal(decidua)-foetal(villi) interface sequenced samples with information regarding sample gestational age and preeclampsia status. (b) Violin plots illustrating the detected (log1p per sample) number of total genes having at least one positive count (UMI) in a cell; total UMI counts representative of total RNA transcripts; and percentage mitochondrial transcript per nuclei. Note that 577-1-v and 557-2-v are technical replicates. (c) Annotated cell type or state

composition (%) per individual biological sample in villi illustrated as stacked bar plots. n = 21 villi (10 early, 6 late control, 5 eoPE) (d) Annotated cell type or state composition (%) per individual biological sample in decidua illustrated as stacked bar plots. n = 12 deciduas (3 early, 4 late control, 5 eoPE). (e) Violin plots illustrating the detected (log1p) and number of total genes having at least one positive count (UMI) per nuclei in villi annotated cell types and states. (f) Violin plots illustrating the detected (log1p) and number of total genes having at least one positive count (UMI) per nuclei in villi annotated cell types and states. (g) Box plots comparing composition per cell type or state compared between early and healthy term control (trm.ctrl) pregnancy samples. CTB, villous cytotrophoblast; STB, syncytiotrophoblast; DSTB, deported STB; CCT, cell column trophoblast; EVT, extravillous trophoblast; VEC, vascular endotheli-al cell; LEC, lymphatic endothelial cell; LECPprog, LEC progenitor; SMC, smooth muscle cell; MC, myocyte; FB, fibroblast; EpC, epithelial cell; MSC, mesenchymal stem cell; DSC, decidual stromal cell; EB, erythroblast; HBC, Hofbauer cell; PAMM, placenta-associated maternal macrophage; Mono, monocyte; MAC1, M1-like macrophage; MAC2, M2-like macrophage; NK, natural killer cell; PC, plasma cell; DC, dendritic cell, Granul, granulocyte; v, villous; d, decidual; p, proliferative; juv, juvenile; pf, pre-fusion.

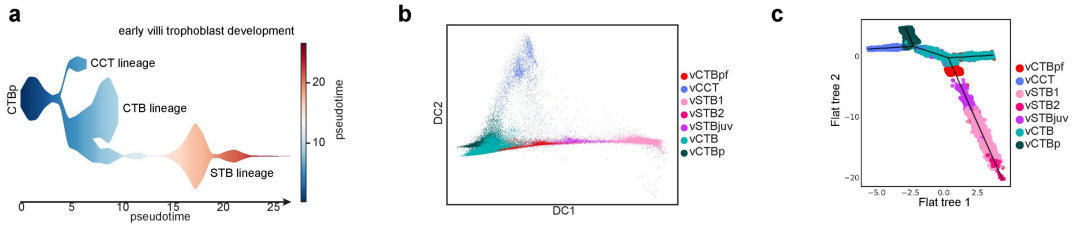


Extended Data Figure 2: Doublet nuclei inference in villous cell types and states

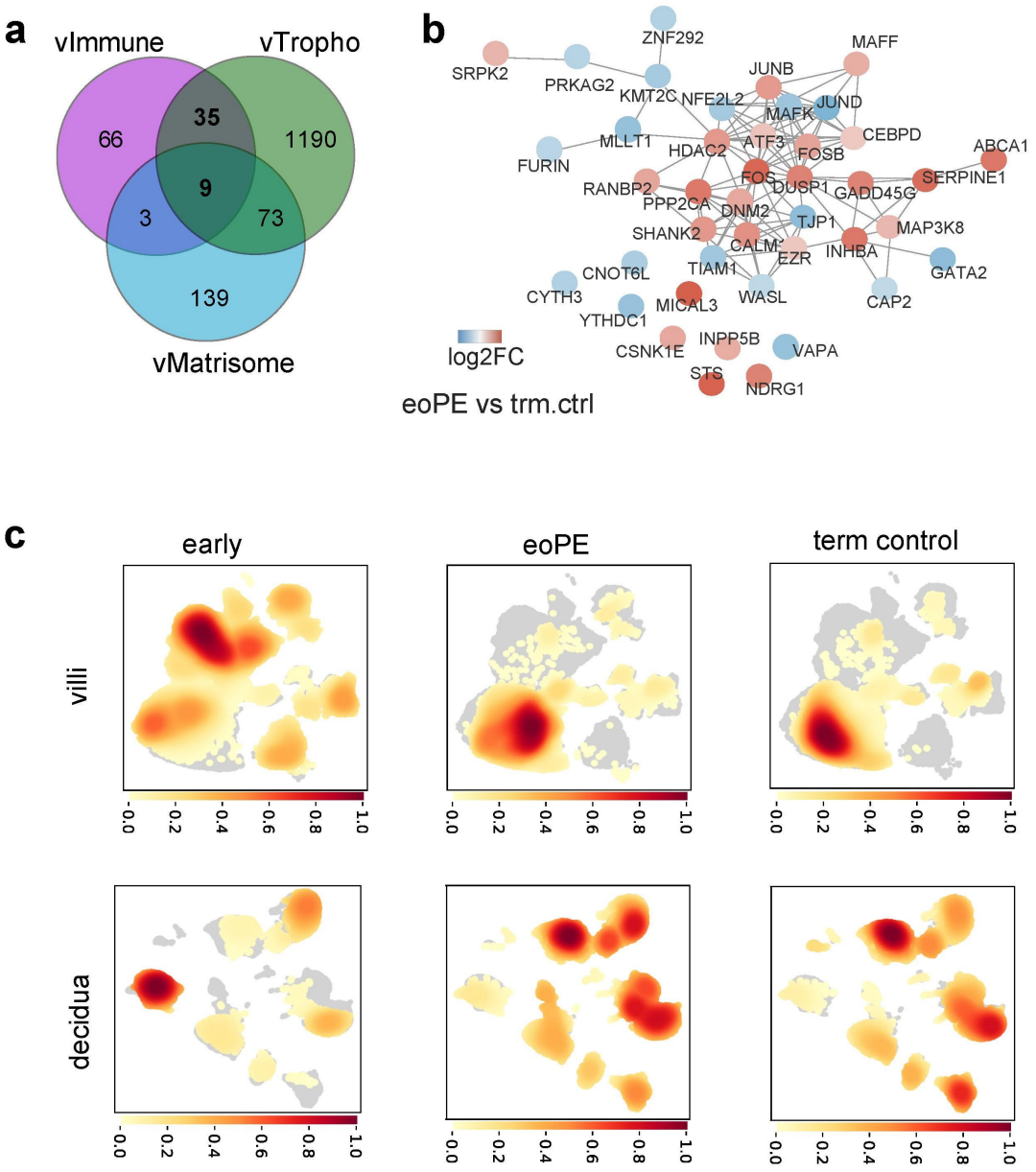
(a) Violin plot depicting doublet score distribution across villi cell types. Scores ranged from 0 to 1. (b) UMAP embeddings reflecting cell type/state distribution. (c) Predicted doublets using a threshold of 0.35 from doublet score histogram as predicted by Scrublet. (d-f) Dot plots showing expression of key marker genes for predicted doublets against singlets for (d) vSTB1, (e) vSTB2, (f) vSTBjuv. Absence of contradictory lineage specific genes reveals lack of biological doublets in our cell types or states. CTB, villous cytotrophoblast; STB, syncytiotrophoblast; DSTB, deported STB; CCT, cell column trophoblast; EVT, extravillous trophoblast; VEC, vascular endothelial cell; MC, myocyte; FB, fibroblast; EB, erythroblast; HBC, Hofbauer cell; PAMM, placenta-associated maternal macrophage; v, villous; p, proliferative; juv, juvenile; pf, pre-fusion.



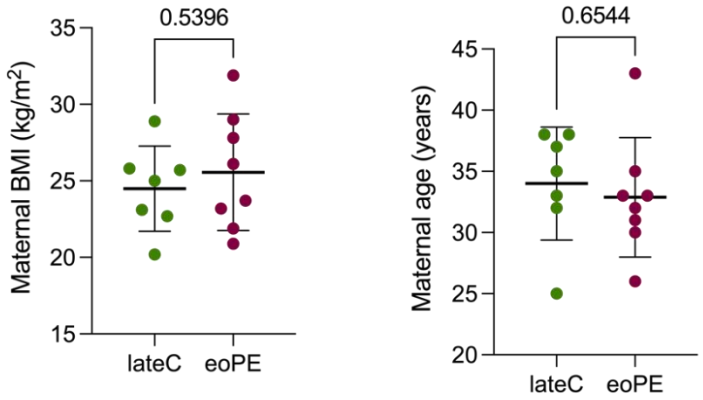
Extended Data Figure 3: UMAP colored by topic profiles identified using Latent Dirichlet Allocation (LDA) analysis.



Extended Data Figure 4: Trajectory inference on trophoblast from early samples
(a) STREAM based on elastic principal graph approximation and colored by pseudotime, (b) diffusion pseudotime based on unsupervised ordering of nuclei based on their global transcriptomic similarity, and (c) force-directed graph based on Fruchterman-Reingold algorithm.

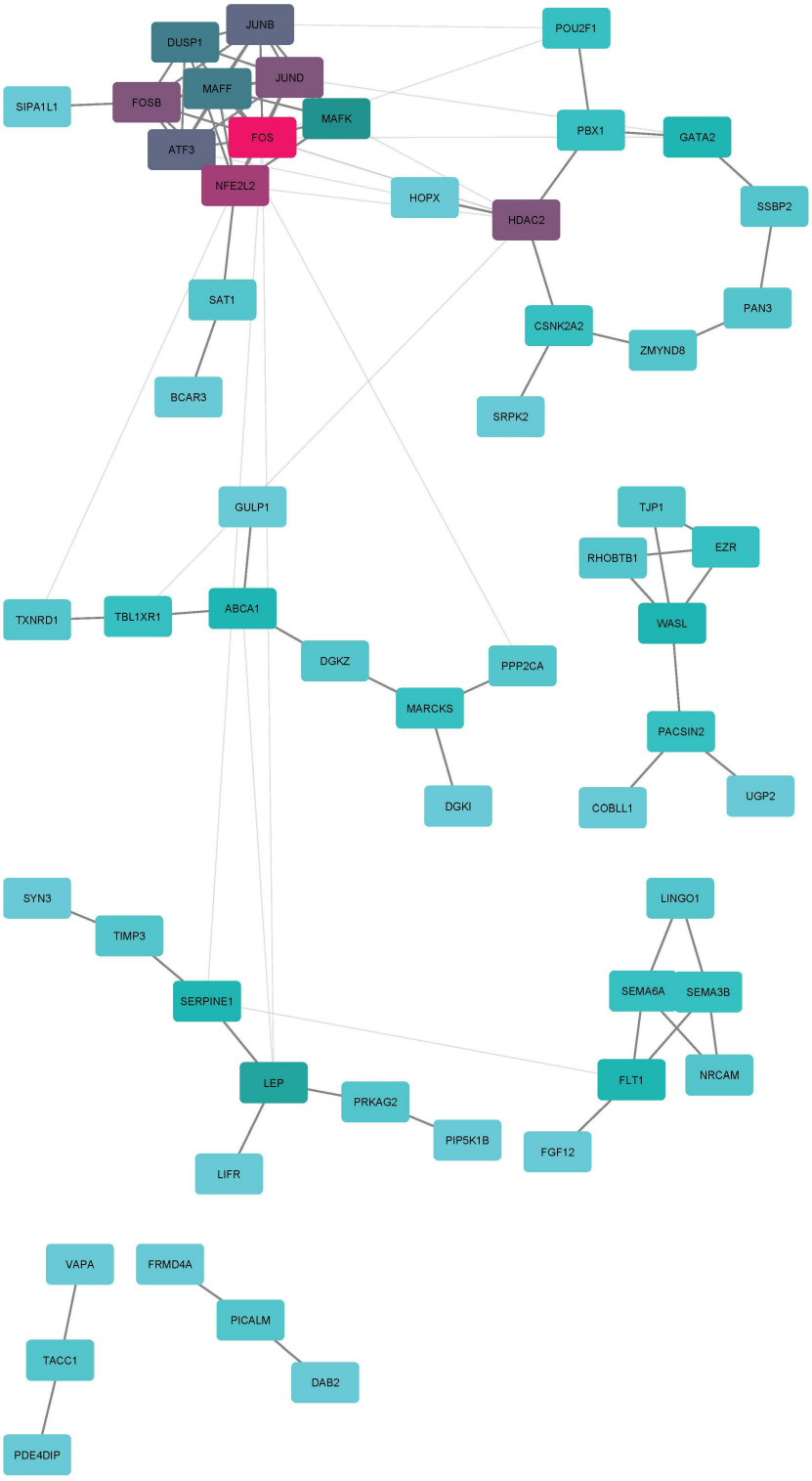


Extended Data Figure 5: eoPE mediated molecular and compositional changes.
(a). Venn Diagram showing overlap of eoPE differentially expressed genes identified in trophoblast, immune and matrisomal compartments of villi, as tabulated in Supplementary Table 12 (excluding vPAMM and vTcells in the immune group). (b). Detected community subgraph of statistically significant vSTB hubs (defined as highly connected nodes). Each hub node reflects a DEG and is colored by log2FC inferred using Logistic Regression (Supplementary Table 11). Node degree is size coded. The final visualized network was achieved by merging hubs (see Methods). Acknowledgement: Dr. Olivia Nonn with my contributions. (c). Kernel density estimation (KDE) reflecting the cell type/state composition change per condition (early, eoPE and term control). KDE values are scaled between 0-1 to enable comparison with the condition.



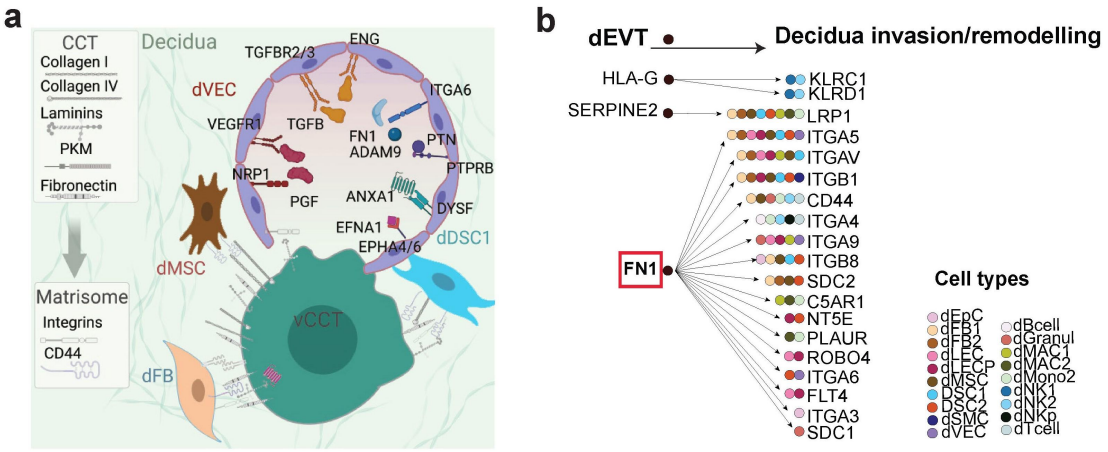
Extended Data Figure 6: Comparison of maternal body mass index (BMI) and age for samples included in the snRNA-seq experiments.

p-values shown derived from unpaired t-test with Welch's correction. For underlying data, refer to Supplementary Table-1. Done on GraphPad with assistance from Daniela S. Valdes.



Extended Data Figure 7: Community network analysis of targets regulated by a minimum of 5 iRegulon inferred transcription factors/regulators.

Network community detection was conducted using Glay algorithm [164] and depicted using Cytoscape. Each node represents a differentially upregulated vSTB gene regulated by at least 5 TFs, as inferred by iRegulon. The degree of a node (the number of incoming and outgoing connections) is visually indicated through color-coding. Isolated nodes (genes) were removed.



Extended Data Figure 8: Mechanisms of EVT invasion and eoPE driven cellular communication in decidua.

(a) Schematic summary of the cell interactions that occur during CCT from the villi onto the vascular and matrisome cell compartments of the decidua in early pregnancy (See supplementary Table 6 for a detailed summary of interactions). Schematic created with BioRender.com. n = 4 early control deciduas, n = 10 early control placentas. (b) Overview of selected receptor-ligand interaction network revealing dysregulation during early onset pre-eclampsia (eoPE) within immune and vascular matrisome compartments in decidua. Interactions were analyzed using Connectome and differentially expressed receptors/ligands were profiled using multivariate Logistic Regression (p-values adjusted for multiple testing). Ribbon color represents the respective source (or, ligand sending) cell types/states, and is proportional to the edge-weight score (product of normalized LFC of both partners). Networks shown are limited to edges where both receptor and ligand log2FC ± 0.25 between conditions (positive indicates upregulated in eoPE and vice-versa). The minimum percentage of cell groups expressing a receptor (or, ligand) is 10%. (c) Summary of selective receptor-ligand interactions for invasive extravillous trophoblasts (dEVT) onto decidual receptors at the maternal-fetal interface analyzed using Connectome. The interaction edges for whom ligand and receptor z-score > 0.25 were included. Further edges were filtered using the Diagnostic Odds Ratio of dEVT cut-off of 5, mean ligand expression > 1.5, edge score > 3, and a minimum percentage of source expressing the ligand: 50%. n = 9 deciduas (4 term control, 5 eoPE). Values are scaled from 0-1 for comparison. CCT, cell column trophoblast; EVT, extravillous trophoblast; p, proximal; d, distal. Acknowledgement: Figure (a, b) was created on Adobe Illustrator with the help of Daniela S. Valdes from my analysis results.

4 Results | Development of human pancreatic islets

- **Scientific acknowledgement:**

For the second chapter focussed on development and maturation of human beta cells, I have separately enclosed the published manuscript in Gastroenterology, according to the rule of Freie Universität Berlin BCP on monograph dissertation. My selected contributions are listed below. I shall remain grateful to Dr. Luca Tosti, Prof. Dr. Christian Conrad, and Prof. Dr. Roland Eils for allowing me to contribute to such an important project.

- **Publication related to chapter:**

Published: Luca Tosti, Yan Hang, Olivia Debnath, Sebastian Tiesmeyer, Timo Trefzer, Katja Steiger, Foo Wei Ten, Sören Lukassen, Simone Ballke, Anja A. Kühl, Simone Spieckermann, Rita Bottino, Naveed Ishaque, Wilko Weichert, Seung K. Kim, Roland Eils, Christian Conrad, Single-Nucleus and In Situ RNA–Sequencing Reveal Cell Topographies in the Human Pancreas. Gastroenterology, Volume 160, Issue 4, 2021, Pages 1330-1344.e11, ISSN 0016-5085, https://doi.org/10.1053/j.gastro.2020.11.010 .

- **Detailed contributions discussed in thesis:**

- Investigated the developmental relationships among adult cell types/states using pseudotime analysis (Figure 4F).

- Modelled the differentiation trajectory of beta cell development integrating neonatal and adult snRNA-seq data (Figure 6I & Extended Figure 6).

5 Results | Unanticipated heterogeneity in high-grade pancreatic neuroendocrine carcinoma pinpoints cell sub-state specific therapeutic targets

- **Scientific acknowledgement:**

For the pancreatic neuroendocrine carcinoma project, Prof. Dr. Bertram Weidenmann, Dr. Katharina Detjen, Dr. Ilaria Marioni, and Prof. Aurel Perren were responsible for obtaining clinical patient samples. 10X single nuclei library preparation and sequencing were performed at BIH Core Unit Genomics facility under the supervision of Dr. Thomas Conrad. Dr. Hilmar Berger performed the pre-processing of raw 10X single nuclei sequencing data with help from Dr. Philipp Kirchner. I express immense gratitude to Dr. Katherina Detjen for all the meaningful discussions that contributed to data interpretation and enhanced overall insights into pancreas and neuroendocrine tumor biology. I shall also remain indebted to Dr. Hilmar Berger for fruitful discussions on bioinformatic analysis. Dr. Aurel Perren provided immunohistochemistry-based characterization and marker validations per patient samples.

5.1 Study design and cohort characteristics

Single nuclei were extracted from surgically resected, histologically confirmed fresh frozen archival specimens of high-grade large cell pancreatic neuroendocrine carcinoma (panNEC) obtained from five patients as shown in **Figure 5.1**. Among these patients, four were treatment-naive, while one (referred to as P1) had received prior treatment with Cisplatin/Etoposide and FOLFIRINOX. Patient characteristics are tabulated in **Table 5.1**. Single nuclei were isolated and sequenced to dissect the molecular landscape of panNECs and their microenvironment. After removing ambient RNA using SoupX [100] and conducting quality control analyses [100], approximately 45,015 single nuclei were profiled. Detailed sample-level statistics and quality control metrics are presented in **Extended Data Figure 9a, b** and **Supplementary Table 22.** Following this, the aim was to elucidate the cell types and sub-states present within the diverse profiles of panNEC patient samples. The specific purpose was to decipher the overarching cell types or cellular communities underlying both tumor and immune niches, with the goal of revealing biological variability, cell-specific roles, and potential targets for future precision medicine endeavors. To achieve this, a chain integration technique was employed for data harmonization across patients to mitigate technical sources of variation (see Methods). Initially, Harmony [110] was employed to adjust the principal components- hence, accounting for batch-specific effects. Subsequently, BBKNN [111] was applied using the corrected principal components to adjust the neighborhood graph, ensuring a balanced representation of contributions from different batches, and facilitating appropriate cell connections based on biological similarities rather than batch-specific variations **(see Methods for details & parameters used).**

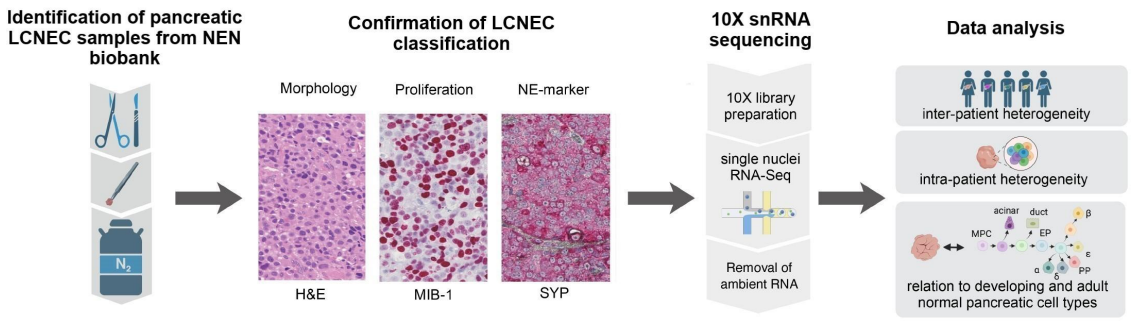


Figure 5.1: Overview of study design.
LCNEC patient samples meeting appropriate criteria were selected from the NEN biobank (see Patient recruitment and ethics approval in the Methods section). Subsequent confirmation of the large cell morphology and classification as NEC was achieved through H&E staining (for morphology), along with MIB1 staining for proliferation and synaptophysin (SYP) staining for neuroendocrine markers,

illustrating key features of NEC. Thereafter, selected patient samples (n=5) were subjected to 10X snRNA-sequencing and downstream data analysis.

Sample ID	Stage	Sample characteristics from H& E stains and immunohistochemistry								
		Ki67	Necrosis	CgA	Syp	SSTR2	SOX9	CK (pan)	p53	Rb1
		>55% ● <55% ○	none ○ punctate ● geographic ●	+ ++ +++ Mix	+ ++ +++ Mix	- + Mix	- + ++ Mix	100% ● 5% ●	Mut ● WT ○ Mix	Mut ● WT ○
P1	IV	●	●	●	●	◎	NA	NA	○	○
P2	IIa	●	●	●	●	◎	●	●	◎	●
P3	IIb	●	○	●	●	●	○	●	●	○
P4	IV	●	●	●	●	○	●	●	○	○
P5	III	○	○	◎	◎	○	◎	●	○	○

Table 4: Patient sample characteristics
Sample characteristics table illustrating the stage, Ki-67 (proliferative) index, necrosis status, and protein expression levels of neuroendocrine markers- chromogranin A (CgA), synaptophysin (SYP), SSRT2, SOX9 and pancreatic cytokeratin (CK pan). In addition, mutation status of p53 and Rb1 are shown, which were profiled using panel sequencing by Simon L. April-Monn, Philipp Kirchner, Dr. Katharina Detjen & Prof. Dr. Aurel Perren at the Bern Center for Precision Medicine, University & University Hospital of Bern. For immunohistochemistry, refer to the Methods. For panel sequencing methods using cancer mutational panel genes, refer to April-Monn et al. 2024 [822].

5.2 Cellular landscape of panNEC

Clustering of single-nucleus transcriptomic profiles revealed the tissue composition across 5 panNEC patient samples and identified five distinct cell types: neuroendocrine (NE), amphicrine, amphicrine progenitor, stromal (normal), and immune populations **(Figure 5.2a, b).** Four of the patient samples (P1-P4) primarily clustered together and significantly contributed to the shared neuroendocrine (NE) cell type **(Figure 5.2a, b)**. In contrast, patient P5 stood out owing to having two unique clusters with robust expression of key acinar markers like *GP2* and the pancreatic multipotent progenitor marker *PDX1*, contributing to the presence of amphicrine acinar and progenitor-like cell types, respectively. **(Figure 5.2 a-d)**. Composition analysis revealed a relatively higher proportion of shared NE cell type in P2 and P4 that is majorly depleted in P5 **(Figure 5.2c)**. However, given a small and heterogenous sample cohort, no statistical analysis was performed to investigate compositional shifts in cell types. The NE cell type exhibited variable expression levels of established diagnostic neuroendocrine markers like *CHGA, SYP*, and *NCAM1* at the mRNA level **(Figure 5.2d)**, which was also validated through immunohistochemical staining at a protein level **(Extended**

Data Figure 10). CHGA and SYP [328, 329] are used as routine diagnostic markers for neuroendocrine neoplasms, including those in the pancreas and particularly useful for distinguishing NEN from other types of pancreatic tumors, such as adenocarcinomas [536]. *SYP* is also considered a reliable NE marker in poorly differentiated NEC [537, 538]. NCAM1, also known as CD56 [331], is a cell adhesion molecule and is used as an immunohistochemical marker to identify neuroendocrine differentiation in tumor cells. Moreover, the shared NE cell type also revealed specific expression of differentially upregulated genes, like *RBFOX1, RIMBP2, EYA2, ADARB2* related to neuronal regulation and novel lncRNA transcripts- *LINC01876* and *LINC01811* (Wilcoxon rank sum test; average log2FC >= 0.25 & Benjamini-Hochberg adjusted p-value < 0.01) **(Figure 5.3).**

The amphicrine cell type was assigned based on the co-expression of pancreatic acinar marker *GP2* [539, 540] and simultaneous expression of neuroendocrine markers- *CHGA, SYP* and *NCAM1* (**Figure 5.2d)**. Apart from *GP2*, key markers indicative of acinar lineage and differentiation, such as *RBPJL* [72, 540, 541, 544]*, NR5A2* [541, 542, 544]*, MECOM* [543, 544] and *CEL* were found to be robustly upregulated in the amphicrine cell type relative to the rest (Wilcoxon rank sum test; average log2FC >= 0.25 & Benjamini-Hochberg adjusted p-value < 0.01) **(Figure 5.3)**. von Figura et al. 2014 [542] reported that *Nr5a2* is a pivotal regulator of acinar plasticity, essential for both maintaining acinar identity and reinstating acinar fate during regeneration. *MECOM* is an oncogenic transcription factor required for acinar cell dedifferentiation [543]. Also, digestive enzyme trypsinogen (*PRSS1*) was specifically expressed in this cell type, albeit at a lower level **(Figure 5.3).** The amphicrine progenitor-like cell type prominently expressed the early pancreatic progenitor marker *PDX1* [545-547] and displayed negligible *GP2* expression **(Figure 5.2d)**. Importantly, this cell type also demonstrated specific expression of pancreatic progenitor marker *SOX9* [545, 547-549], and robust expression of *FREM2*, *ALK* and *RASGRF1* **(Figure 5.3).** Of note, *FREM2* is an early endodermal marker [550] which is not typically expressed in the adult pancreatic cell types [72]. Moreover, *Frem2* is known to dynamically modify the ECM to create a conducive substrate for cell migration and rearrangements throughout embryogenesis [550]. Conversely, a stromal cell type was identified exhibiting strong and specific expression of *VIM* but lacked expression of key NE markers such as *SYP* **(Figure 5.2d)**. Furthermore, ECM remodelling genes such as *CALD1, COL1A1, COL1A2,* and *CASC15* were found as markers of this cell type **(Figure 5.3)**. Additionally, this cell type was observed to be more abundant in P1, which had undergone previous chemotherapy treatments (**Figure 5.2c)**. Given the absence of NE marker genes, this cell type was annotated as Stroma (normal)- this classification was later confirmed through experimental validation using IHC **(refer to section 5.4).** Immune cell type was characterized using prominent markers like *PTPRC, CD247, THEMIS, ARHGAP15* and

DOCK2- having previously established function in immune biology **(Figure 4.2d & 4.3).** Stroma (normal) and immune cell types, being non-malignant, are most likely components of the panNEC tumor microenvironment (TME). Importantly, this acinar-like distinction in P5 was validated by localizing focal trypsin immunohistochemical staining, which was absent in the other tumors **(Extended Data Figure 10).**

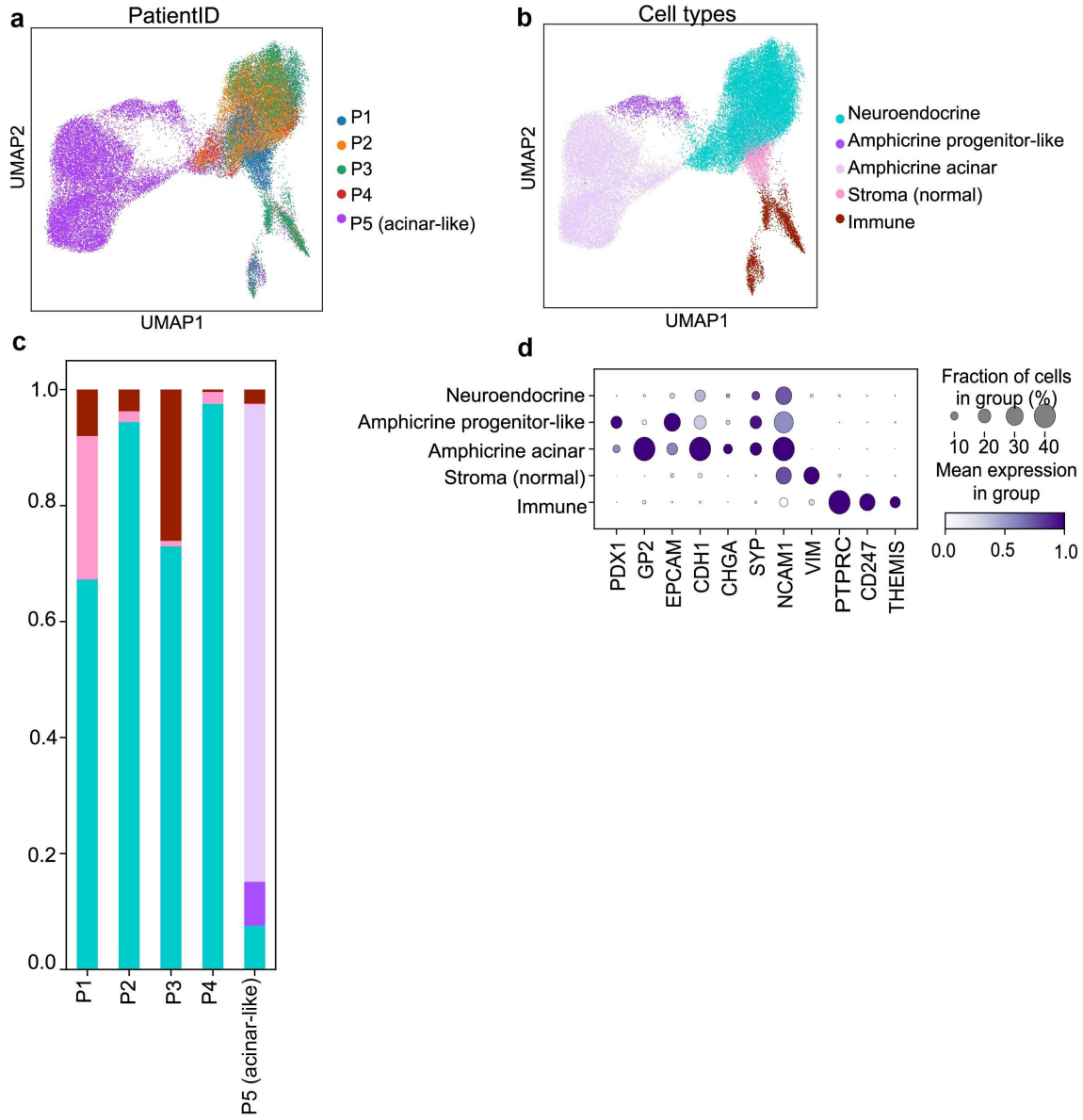


Figure 5.2: Cellular landscape demonstrating shared and unique cell types underlying panNEC.

UMAP visualization of the cell type organization in LCNEC samples colored by (a) Patient-ID, and (b) broad cell types inferred from integrated snRNA-seq data (n=5). Each dot (single nucleus profile) represents Patient-ID and cell type in (a) and (b) respectively. (c) Proportions (y-axis) of cell types (color legend, shared with panel b) per patient sample depicted using stacked bar-plot. (d) Dot-plot depicting key neuroendocrine markers CHGA, SYP and NCAM1 across cell-types. Other

distinguishing markers characterizing early pancreatic progenitor (PDX1), acinar lineage (GP2), epithelial (EPCAM, CDH1), mesenchymal (VIM) and immune (PTPRC, CD247, THEMIS) are shown. Nuclei in P5 clusters exhibited a distinct transcriptomic profile of both acinar- and neuroendocrine-specific genes. Mean expression within each cluster is color-coded and percentage of nuclei expressing a marker is size-coded.

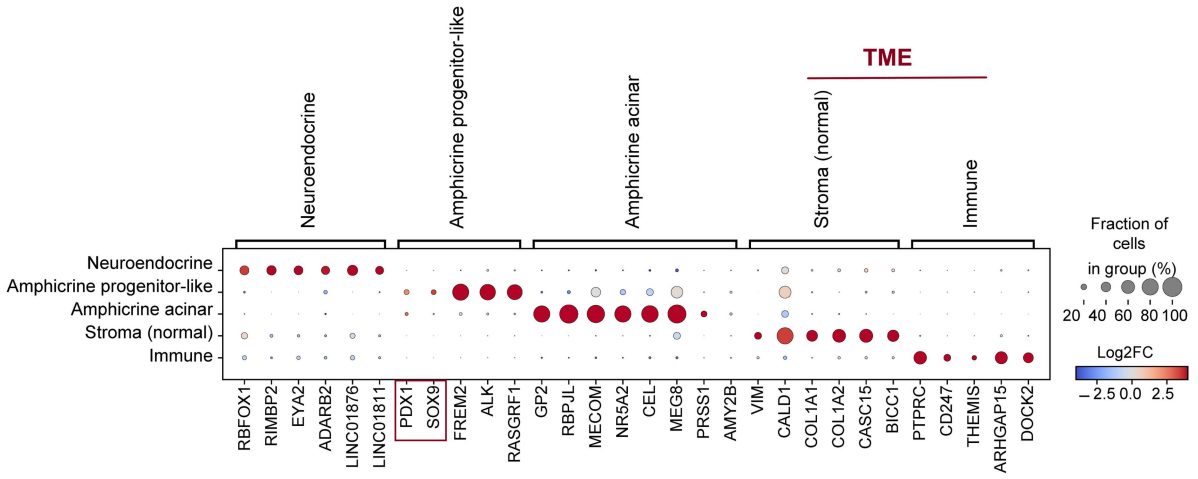


Figure 5.3: Neuroendocrine and amphicrine cell types characterized by robust and specific markers.
Dot-plot depicting robust and specific markers across each cell type inferred using Wilcoxon rank sum test (average log2FC >= 0.25 & Benjamini-Hochberg adjusted p-value < 0.01). Log2FC of a marker within a cell type is color-coded and percentage of nuclei expressing a marker is size-coded. The cell types associated with tumor microenvironment (TME) i.e., stroma (normal) and immune are marked. PDX1 and SOX9 were manually added to the figure, given their biological relevance [545-549]. Both depict positive average log2FC in amphicrine progenitor-like relative to other cell types. FC= fold change.

5.2.1 Shared panNEC sub-states

Subsequently, the cell types were subclustered into sub-states **(Figure 5.4a)** defined by robust and specific markers inferred using differential gene expression analysis (average log2FC >= 0.25 & Bonferroni adjusted p-value < 0.01, multi-variate Logistic Regression) **(Figure 5.5, Extended Data Figure 11, Supplementary Table 23).** Relative composition of sub-states per patient sample is depicted by **Figure 5.4b**. Given the small number of heterogeneous samples (n=5), no statistical test is shown to make conclusive remarks about abundance or depletion of a sub-state. Nevertheless, a kernel density estimation (KDE) was performed to obtain a qualitative understanding of the composition shift among the patient samples. For example, a noticeable trend of enrichment was observed for neuroendocrine proliferating in case of P2 **(Figure 5.4b & Extended Data Figure 9c).**

Despite salient variability among patients, common sub-states within the neuroendocrine cell type were identified, exhibiting enrichment in specific functional pathways and oncogenic processes **(Figure 5.6 & Supplementary Table 24).** The shared neuroendocrine sub-state was characterized by the presence of established NE markers like *RIMBP2* and *ADARB2*, linked to neuronal system, as well as *CACNB2*, *CACNA1A*, and *KCNJ3* involved in voltage gated $Ca2^{+}$ and K^{+} channels, respectively [551] **(Figure 5.5 & Extended Data Figure 11).** Together, these markers contribute to enriched pre-synaptic depolarization **(Figure 5.6 & Supplementary Table 24).** The calcium channel, encoded by the *CACNB2* gene, is a membrane-associated guanylate kinase (MAGUK) protein, which serves as the β2 subunit of the L-type cardiac calcium channel *CACNA1C*. Of note, L-type calcium channels facilitate the entry of Ca2+ into the cytoplasm and control both cardiac excitability and excitation-contraction coupling (EC coupling) [552].

Additionally, a closely related neuroendocrine proliferating sub-state marked by the expression of *MKI67, DIAPH3*, and *CENPP* **(Figure 5.5, Extended Data Figure 11)** was enriched for processes associated with cell-cycle regulation and chromatin remodeling pathways, including the G2M checkpoint, E2F transcriptional targets, DNA damage response (DDR), and Aurora kinase activation **(Figure 5.6 & Supplementary Table 24)**. Notably, the Fanconi anemia (FA) signaling pathway, governed by *FANCA, FANCI,* tumor suppressor genes *BRCA1/2*, and the *BRCA* interacting protein-encoding gene *BRIP1* (**Figure 5.5 & Extended Data Figure 11**) were found to be enriched in the NE proliferating sub-state **(Figure 5.6).** Mutations or deficiencies in FA genes can lead to impaired DNA repair, and lead to chromosomal abnormalities and genomic instability, making it a genetic susceptibility factor for cancer predisposition [553]. Elevated expression of FA genes was previously found to be correlated with chemotherapy resistance in cisplatin-resistant non-small cell lung cancer (NSCLC) [554] and ovarian cancer cell lines [555]. In the case of patient P5, a distinct proliferating sub-state was detected that harboured significantly high expression of acinar genes, as discussed in subsequent **sub-section 5.1.2**. This cluster was annotated as "amphicrine acinar proliferating" sub-state, which shared a clustering pattern with the amphicrine sub-states on the UMAP.

Furthermore, a shared neuroendocrine sub-state (“NE HSP+ (hypoxic)”) demarcated by high expression of heat-shock protein-encoding genes, including *HSP90AA1, HSP90AB1, HSPE1*, and *HSPH1,* along with hypoxia-induced genes such as *VEGFA, KDM3A,* and *NDRG1* **(Figure 5.5 & Extended Data Figure 11)** was unveiled. Importantly, this sub-state showed notable enrichment in terms related to the heat stress response, along with hypoxia, mTORC signaling, glycolysis, upregulated MYC targets, and MAPK signaling **(Figure 5.6 &**

Supplementary Table 24). Cells encountering environmental stressors such as hypoxia may activate pathways like mTORC/PI3-AKT and generate heat stress response to adapt and survive [556, 557]. The overactivation of mTORC signaling promotes cell proliferation, survival, and metabolism, thereby contributing to tumor growth and progression [558], a phenomenon frequently observed in pNET [559]. Prior studies have reported frequent abnormal activation of mTOR, often attributed to inactivating mutations in genes encoding negative regulators of the pathway or through indirect mechanisms [559]. Clinically, the overexpression of mTOR and its downstream targets were associated with worse prognosis in various NETs [560, 561]. In a former study by Shida et al. in 2019 [562], IHC analysis of tumor tissues revealed elevated mTOR expression in poorly differentiated NENs, with a reported expression rate of 67% that declined to 27% in well-differentiated counterparts. Similarly, Catena et al. in 2011 [563] found mTOR expression in 80% of patients with poorly differentiated NEC, irrespective of tumor origin (including pancreas, colon, lung, and small intestine), or the rate of proliferation.

Another sub-state, termed "neuroendocrine stromal-like," displays moderate expression of *ROBO2*, low expression of NE genes like *KCNB2,* and stromal genes, such as*, COL1A2* and *CALD1*. No highly specific genes were found to distinguish neuroendocrine stromal-like relative to other sub-states, and hence might represent a transitional state between NE and stromal phenotype. In contrast, "stroma (normal)" sub-state is notably more prevalent in patient P1 and is characterized by the specific expression of epithelial-mesenchymal transition (EMT) mediators like *COL1A1, COL1A2, CDH11, FN1,* and *LAMA4* **(Figure 5.5 & Extended Data Figure 11).** Interestingly, this sub-state also revealed robust expression of Hippo signaling members *YAP1* and *WWTR1 (TAZ)* **(Figure 5.5),** known as critical regulators of pancreatic tissue regeneration, and neoplastic transformation in both PDAC and pancreatitis [564, 565]. Also, *YAP1* is inactivated following normal endocrine specification and not expressed in pancreatic islets [566] and rather established as a potential oncogene in PDAC [567]. This further prompted an inquiry about whether *YAP1* originates from the tumor subpopulation or adjacent stroma in P1, using IHC **(refer to section 5.4).** *CDH11* is recognized for promoting immunosuppression, extracellular matrix deposition [568], and development of desmoplastic stroma. Previous *CDH11* knockdown (KD) studies have reported reduced cell migration in PDAC [569]. Albeit the transcriptional profile of stroma (normal) resembles cancer associated fibroblasts (CAFs), this cell type lacks key CAF markers like *FAP* and *S100A11* [570]. Analysis of differential pathways revealed biological processes such as R-mediated phagocytosis and PPI synapse in both stromal sub-states **(**refer to **Figure 5.6)**. Stroma (normal) exhibited significant enrichment in specific pathways related to Focal adhesion, NABA matrisome, ALK, and PDGF signaling **(Figure 5.6 &**

Supplementary Table 24). Activation of PDGF signaling is known to promote the recruitment of endothelial cells and stimulate angiogenesis within the tumor microenvironment, facilitating tumor growth and metastasis [571, 572]. Also, the signaling pathways can influence the immune response within the tumor microenvironment. For example, alterations in ECM composition mediated by focal adhesion and NABA matrisome signaling can impact immune cell infiltration and function. Additionally, paracrine signaling through ALK and PDGF pathways may affect the immune cell recruitment and activity within the tumor stroma [573]. Taken together, the stroma (normal) might function like CAFs and promote ECM remodelling, cell migration, invasion, and metastasis. To decipher the biological differences between neuroendocrine stromal-like and stroma (normal) sub-states, a marker analysis was performed comparing the two that revealed relatively increased levels of neural genes like *RIMS2, PTPRN2, CACNA1A* and *PLCG2* (average log2FC >= 0.25, adjusted p-value < 0.01, Logistic Regression analysis) **(Supplementary Table 23)** in the former.

The immune cell type was further subdivided into lymphocytes and macrophages **(Figure 5.4a)** and characterized by well known marker genes **(Figure 5.5 & Extended Data Figure 11)**. Of note, *PLXDC2*- a LR inferred marker of macrophage was previously reported as a TME-related signature and highly correlated to CD163 M2 macrophages [574]. A previous study by Tubau-Juni et al 2020 [575] postulated *PLXDC2* to be highly immunomodulatory, and Plxdc2 deficiency resulted in the production of pro-inflammatory macrophages in their ex-vivo experimental system. In metastatic populations, tumor-infiltrating macrophages exhibit strong expression of *CTSB*, another marker of this sub-state [576]. Lin et al. 2022 [577] discovered that *CTSB*+ macrophages suppress the anti-tumor immune response through various immune checkpoints. Another study [578] reported that DOCK8 protein regulates macrophage migration. Another marker associated with this sub-state, *DOCK2*, was identified as regulating macrophage migration, phagocytosis, and ROS production [579].

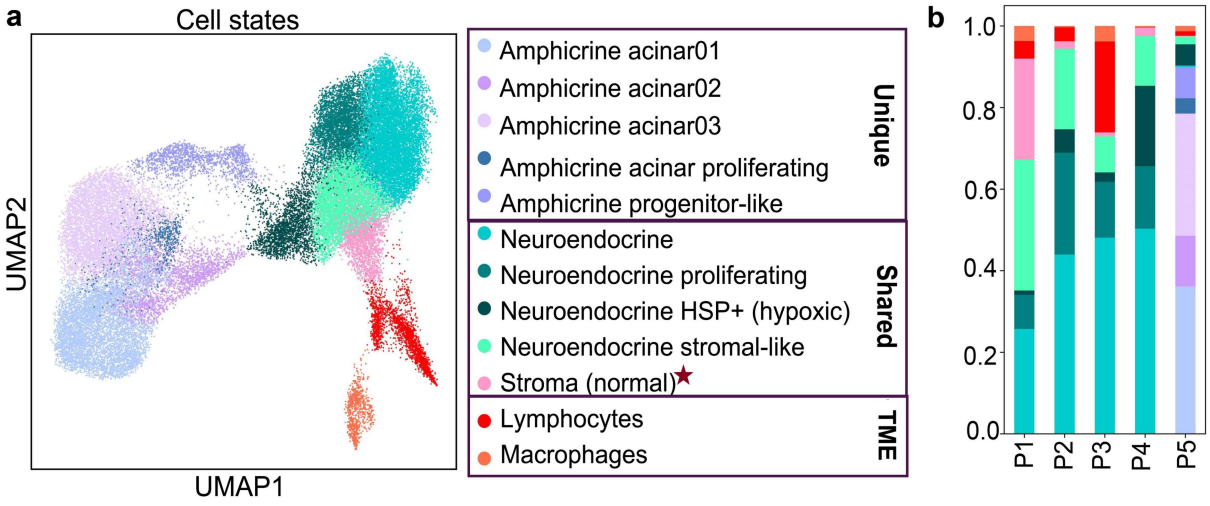


Figure 5.4: Cellular landscape underlying panNEC constitute shared and unique sub-states.

(a) UMAP embedding of single-nucleus profiles (dots) of panNEC samples (n=5) colored by cell sub-states derived from integrated snRNA-seq, and sub-clustering cell types shown in figure 4.2b. The unique, shared, and tumor microenvironment (TME) sub-states are labelled. Stroma (normal) is marked (in asterisk) as it is a shared non-malignant sub-state. (b) Stacked bar-plot revealing relative composition (y-axis) of inferred cell sub-states per patient sample. The unique sub-states are specific to P5. Color legend shared with (a).

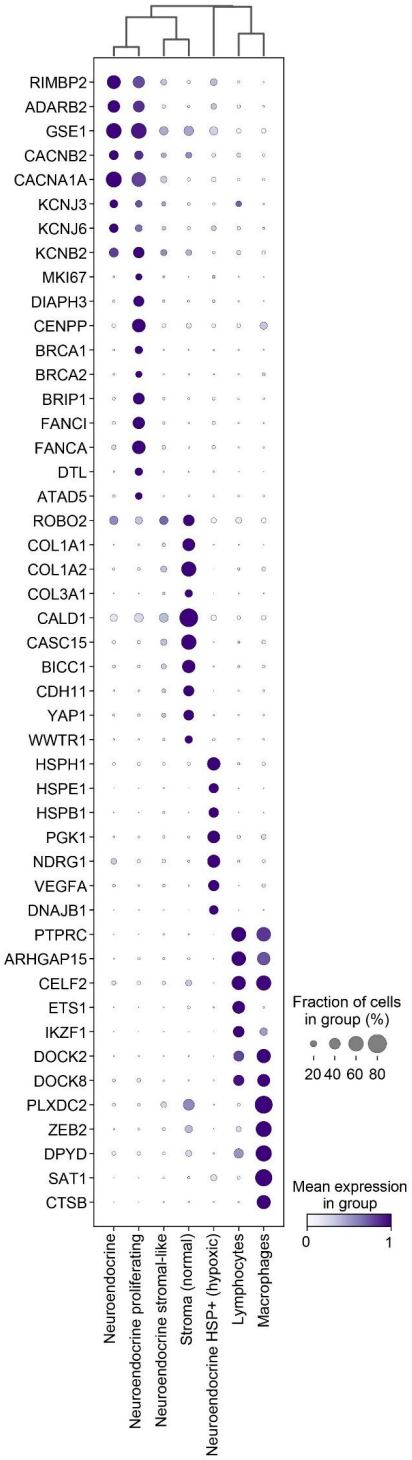


Figure 5.5: Shared sub-states characterized by robust and specific markers.
Dot-plot displaying selected key markers for shared sub-states inferred using multi-variate Logistic Regression framework (avg log2FC > 0.25 & Bonferroni adjusted p-value < 0.01). Mean marker expression within the cluster is color-coded and percentage of nuclei expressing a marker is size-coded. The dot-plot comparing all sub-states (unique and shared) is shown in Extended Data Figure 11 and tabulated in Supplementary Table 23.

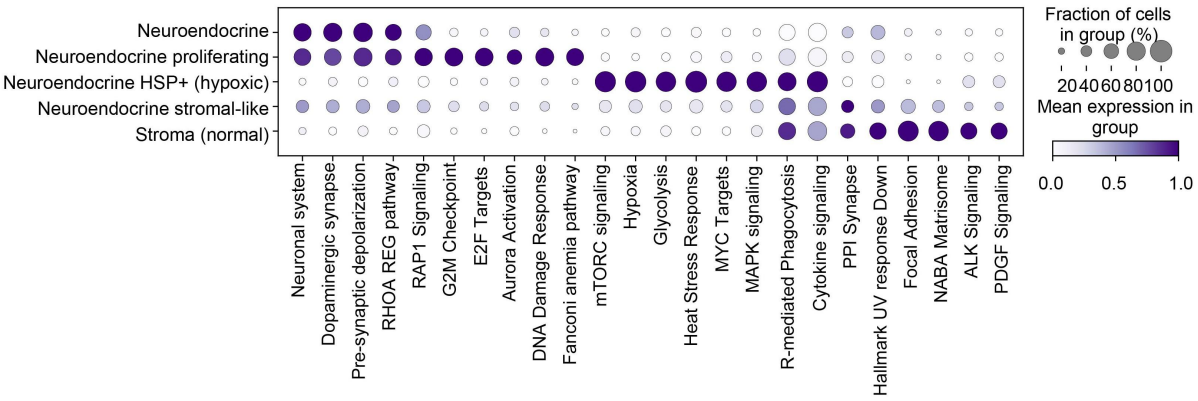


Figure 5.6: Shared sub-states characterized by differential pathway signaling programs.
Dot-plot depicting signature scores for differentially expressed signaling pathways and processes for shared sub-states. Expression is color-coded and percentage of nuclei expressing a pathway module is size-coded. Pathways were inferred from the top-100 differentially upregulated genes per sub-state and listed in Supplementary Table 24.

Notably, two genes that direct neuronal development- *RBFOX1* [580] and *NKX2-2* [581] were found to be specific to the shared sub-states and were not expressed in the amphicrine sub-states **(Figure 5.7)**. *RBFOX1* was identified as a differentially upregulated gene in both NE and NE proliferating (average log2FC= 1.4077 & 0.6638, respectively, Bonferroni adjusted p-value < 0.0001, Logistic Regression), while it was moderately expressed in NE stromal-like **(Figure 5.7)**. Interestingly, *RBFOX1* is not expressed in the embryonic pancreas or the adult endocrine pancreas under physiological conditions [582]. This was further confirmed by integrating the adult pancreas snRNA-seq dataset from Tosti et al. 2020 [72], which showed the absence of *RBFOX1* in the native pancreas cell types or states **(Figure 5.7)**. Moreover, the RNA binding protein RBFOX1 has been shown to regulate splicing and impair the stability of LSD1 (KDM1A) [583], thereby adjusting the ratio of LSD1 and its neuro-specific isoform LSD1n, resulting in de-repression of neuroplastic gene expression [584]. Specifically, RBFOX1 has been linked to calcium signaling, including the expression of calcium channels [585]. Hence high *RBFOX1* expression, in addition to serving as a marker gene, could provide a tentative mechanistic link to the prominent expression of calcium channels present

in the shared NE states [585]. Also, *NKX2-2* was found to be specific only for the NE and NE proliferating sub-states **(Figure 5.7).** Nevertheless, *NKX2-2* was not identified as an LR marker — it might be due to the limited capability of snRNA-seq in detecting transcription factors as markers when their expression is low [160]. Given that *NKX2-2* is of prime importance in pancreatic islet development [581], its expression was also found in adult endocrine cell types i.e., alpha, beta, and delta **(Figure 5.7)**. Importantly, amphicrine progenitor-like sub-state was found to robustly express *PROX1* that was considerably decreased in the other amphicrine sub-states (avg log2FC= 1.0677, Bonferroni adjusted p-value < 7.02E-179, Logistic Regression) **(Figure 5.7).** Of note, *PROX1* is a transcription factor known for driving neurogenesis in the adult brain [586]**,** is highly expressed in the undifferentiated progenitors of mouse pancreas and controls pancreatic morphogenesis [587]. Moreover, *PROX1* is known to be highly expressed in the multipotent progenitors of mouse embryonic pancreas, whereas it is absent in mature acinar cells in the postnatal murine pancreas [588, 589]. The latter fact was verified as *PROX1* expression was negligible in the Tosti's acinar sub-states **(Figure 5.7).**

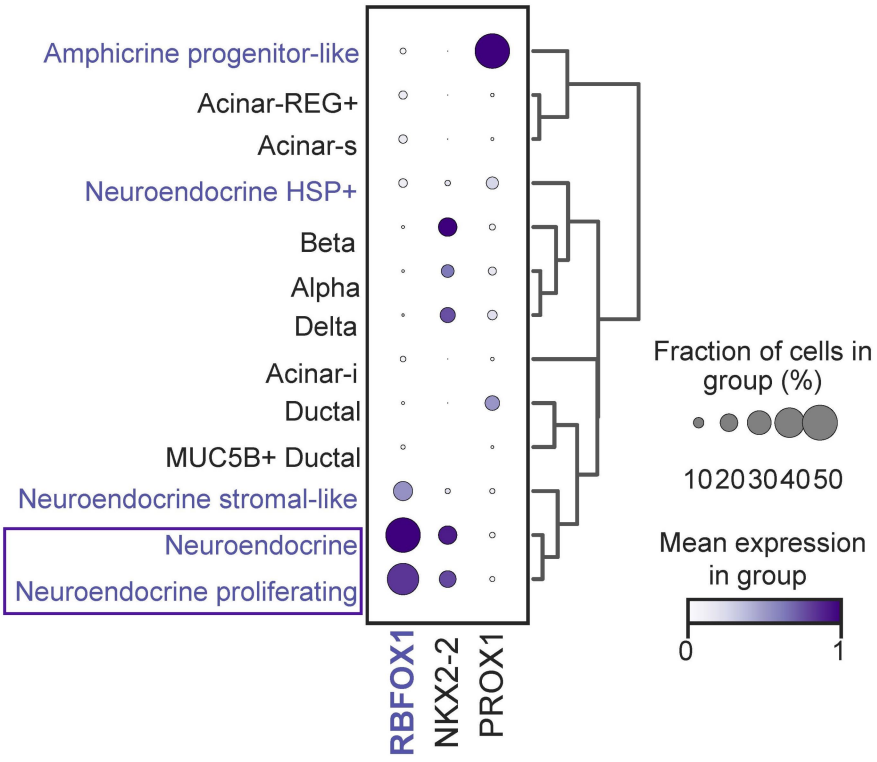


Figure 5.7: Shared NE and NE proliferating panNEC sub-states distinguished by classical neurogenesis markers- RBFOX1 and NKX2-2.

Dot-plot illustrating the normalized expression of RBFOX1, NKX2-2 and PROX1 across panNEC and adult pancreas snRNA-seq dataset from Tosti et al. 2020 [72]. panNEC sub-states are shown in purple. Average gene expression within a sub-state is color coded and and dot size reflects the percentage of nuclei expressing a gene, capped at 40% for visualization. Notably, RBFOX1 emerged as differentially upregulated in NE and NE proliferating sub-states and is highlighted in purple. PROX1 is significantly

upregulated in amphicrine progenitor-like (multivariate Logistic Regression, see Supplementary Table 23). Neuroendocrine HSP+ (hypoxic) was written as Neuroendocrine HSP+.

5.2.2 Amphicrine sub-states unique to sample P5

In the instance of patient P5, a distinct amphicrine sub-state termed 'Amphicrine progenitor-like' was identified— characterized by the expression of key markers and transcription factors that orchestrate differentiation of early pancreas progenitors, including, multipotent pancreas progenitors (MPP) and tip progenitors [547, 590, 591]. Notable MPP markers *PDX1* and *SOX9* [545-549] were jointly expressed in the amphicrine progenitor-like sub-state **(Figure 5.8).** Of note, several former lineage investigation studies associated *PDX1* and *SOX9* with MPP that could differentiate into acinar, ductal, and endocrine lineage paths [545, 547-549, 592-594].

Taken together, the amphicrine progenitor-like sub-state reflected an immature, stem cell-like phenotype. While amphicrine acinar03 exhibited *PDX1* expression at a similar level to amphicrine progenitor-like cells, the former did not express *SOX9* and instead displayed acinar markers such as *GP2* **(Figure 5.8 & 5.9a).** Furthermore, *PTF1A*- a major tip progenitor marker [547, 590, 591], known for maintaining the identity and function of mature acinar cells was lowly expressed in the amphicrine acinar sub-states, including amphicrine acinar03 (approximately 5% of nuclei per sub-state express *PTF1A*; **Figure 5.8**). It might be possible that the sparse portion of nuclei expressing *PDX1+/PTF1A+* in amphicrine acinar03 indicated the existence of tip progenitors or early acinar progenitors before fully committing to mature acinar fate. In this context, a noteworthy study by Cogger et al. 2017 [595] postulated that *GP2+/PTF1A+/NKX6-1+* subpopulation might correspond to MPP. However, *NKX6-1* was not expressed in the panNEC dataset. Furthermore, the presence of mouse multipotent progenitor markers *PDX1* and *ONECUT1* was noted in both amphicrine progenitor-like and amphicrine acinar03 sub-states [596] (**Figure 5.8).** On the other hand, other early pancreatic progenitor markers, including *HNF1B* [597-599]*, GATA6* [600-602]*, ETV4* and *MEIS1/2* [603] were found to be expressed in all the amphicrine sub-states, although *HNF1B* was relatively robust for amphicrine acinar sub-states **(Figure 5.8)**.

Moreover, amphicrine progenitor-like and amphicrine acinar01 sub-states showed relatively higher expression of *ONECUT2* and *FOXQ1,* respectively **(Figure 5.8)**. These two genes are recognized markers for metaplastic acinar cells undergoing transdifferentiation towards ductal or other secretory epithelial cell states following injury or oncogene activation in a

process referred to as acinar to ductal metaplasia (ADM), as described in Schlesinger et al. 2020 [604]. Moreover, the amphicrine acinar sub-states strongly expressed regenerative protein-encoding gene *REG1A*, transcription factors *ONECUT1* [604, 606], *HNF4A* [605, 606] and *HNF4G*- critical for embryonic and pancreatic development; *HNF4A* and *GATA6* are known drivers of classical phenotype in PDAC [605] **(Figure 5.8)**. Collectively, this suggested a functional analogy of the amphicrine acinar sub-states with the acinar-REG+ cell-state described in Tosti et al. 2020 [72] and Muraro et al. 2016 [294] **(Extended Data Figure 12)**. In contrast, the shared sub-states showed negligible expression of *GATA6*, and extremely low levels of *HNF4A/HNF4G* **(Figure 5.8).** This finding aligned with Kawasaki et al. 2020 organoid study [607], indicating that shared panNEC sub-states might have lost of their original cell identity and were reprogrammed into the neuroendocrine lineage**.** Overall, based on the expression of genes relevant for PDAC and pancreatic neoplasia like *HNF1B, GATA6, ONECUT2, HNF4A* and *HNF4G,* the amphicrine sub-states also recapitulated similarities with MUC5B+ ductal population from Tosti et al. 2020 [72] **(Extended Data Figure 12)**.

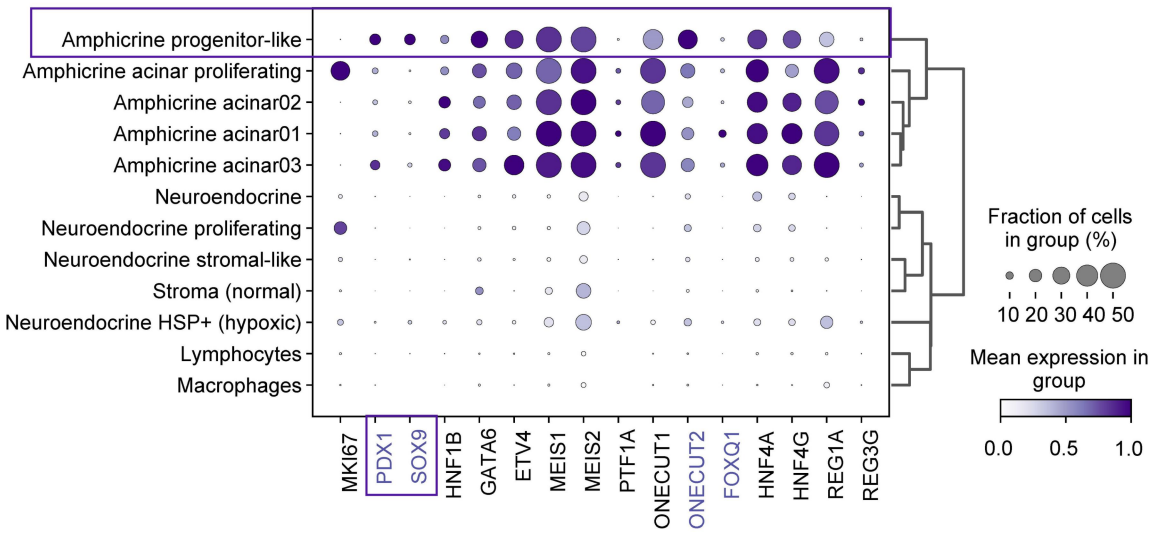


Figure 5.8: Amphicrine progenitor-like and amphicrine acinar sub-states characterized by early pancreatic progenitors and metaplastic acinar signatures.
Dot-plot illustrating key markers and transcription factors pertinent to early pancreas development and the regulation of pancreatic progenitors (including multipotent pancreatic progenitors, tip and trunk progenitors) as outlined in the text. Visualized gene-set is curated from several literature references, as described in the text. PDX1 and SOX9 are well-established regulators of multipotent pancreatic progenitors [545-549, 592-594] and highlighted for their prime importance in early pancreas development. ONECUT2 and FOXQ1 are indicative of late metaplastic stage, as reported by Schlesinger et al. 2020 [604]. HNF4A and GATA6 are known drivers of PDAC classical phenotype [605]. The presence of MKI67 in the amphicrine acinar proliferating cluster suggests a self-renewing sub-state. NEUROG3 and FEV were not detected in the dataset, indicating the absence of classical pancreatic endocrine progenitors. Average gene expression within each cluster is color-coded, while the percentage of nuclei expressing a marker is indicated by marker size. Gene expression levels are standardized across columns for comparison. For the comparative analysis with healthy pancreatic cell types/states from Tosti et al. 2020 [72], refer to Extended Data Figure 12.

As previously stated, repression of acinar lineage encoding genes like *RBPJL, GP2,* and digestive enzymes such as *PRSS1, PRSS2* was observed in amphicrine progenitor-like sub-state, emphasizing an immature and less differentiated state **(Figure 5.9a & Supplementary Table 23)**. Consistent with its progenitor characteristics, genes associated with pancreatoblastoma (PBL) [608], an immature childhood tumor with multilineage features were prominently expressed in this sub-state. These genes included *LEF1, LGR5, BMP4, ID1*, and *TCF7* **(Figure 5.9a & Supplementary Table 23)**. Of note, the positive markers for amphicrine progenitor-like sub-state were also statistically over-enriched in PBL signatures extracted from [608] (representation factor: 18.0, hypergeometric p-value < 5.507e-08) **(Figure 5.9b, c).** This observation suggested the presence of WNT-BMP-NOTCH signaling pathway, which seemed to be a notable feature of this sub-state, implying a dynamic interplay of signaling cascades within this unique cell population. Crosstalk among these pathways is known to be associated with the maintenance of stemness and progenitor cell properties, in promoting cell survival and inhibiting differentiation in PBL [608]. Importantly, pathway enrichment analysis using differentially upregulated markers (n=200, avg log2FC >=0.25 & Bonferroni adjusted p-value < 0.01, Logistic Regression) also suggested Wnt signaling to be a critical regulator of this sub-state **(Figure 5.9d),** governed by mediators such as *CTNNB1, AXIN2, LGR5, DKK4, LEF1, BMP4, RNF43 and NOTUM* **(Figure 5.9a, b & Supplementary Table 24)**. Previously inferred, neurogenesis driver *PROX1* is known to be induced by Wnt signaling [609, 610] **(Figure 5.7).** Owing to the noted resemblances with early pancreatic progenitors — including, MPP and trunk progenitors (such as, *PDX1+/SOX9+*), PBL stemness signatures, and the absence of mature acinar markers, this sub-state was labeled as "amphicrine progenitor-like." Furthermore, a significant upregulation of genes linked to invasiveness and cell migration, including *NOTUM, RASGRF1*, and *ALK* **(Figure 5.9a)** was observed in this progenitor-like sub-state. This heightened expression of genes associated with cellular mobility suggested a potentially more aggressive and invasive nature of this sub-state within the tumor. Interestingly, this sub-state also demonstrated enrichment in transport of ions **(Figure 5.9d)**, influenced by overexpressed markers like *ADCY2, ATP1A1, SLC24A1, SLC7A11,* and *SLC22A15* **(Supplementary Table 24)**. This observation implied altered cellular homeostasis and metabolic dynamics in P5. Of note, amphicrine progenitor-like and shared NE HSP+ (hypoxic) sub-states shared a statistically significant marker overlap (representation factor: 6.8 & hypergeometric p-value < 2.833e-37). Performing pathway enrichment using common markers indicated enrichment in processes such as fluid shear stress, presenilin PS1 pathway, RAC1 pathway, thyroid signaling, and focal adhesion **(Figure 5.9e & Supplementary Table 24)**.

Apart from this, three distinct amphicrine acinar sub-states, namely Amphicrine acinar01, Amphicrine acinar02, and Amphicrine acinar03 were detected in the amphicrine acinar cell

type **(Figure 5.4a).** These three sub-states were characterized using key acinar progenitor markers like *RBPJL, GP2, NR5A2, MECOM* and *CEL* **(Figure 5.9a, Extended Data Figure 11 & Supplementary Table 23).** *NR5A2* regulates various stages of pancreatic development over time, which involve regulatory mechanisms pertinent to pancreatic oncogenesis and the preservation of the exocrine phenotype [611]. During the last phase of acinar cell maturation, the lack of *Nr5a2* was found to influence *Ptf1a* and *Rbpjl* expression that resulted in incomplete differentiation of the limited acinar cells formed [611]. Amphicrine acinar-01 additionally expressed upregulated trypsinogen *(PRSS1 & PRSS2*) levels **(Figure 5.9a)**. The amphicrine acinar-01 and -03 sub-states demonstrated significant enrichment of *PDX1*-regulated targets [612], with the former exhibiting the highest enrichment. Also, module score analysis revealed that amphicrine acinar-01 sub-state is statistically enriched in one of the metaplastic acinar programs previously described in Schlesinger et al. 2020 in a K-Ras driven mouse model of acinar to ductal metaplasia [604] **(Figure 5.9b)**. Furthermore, DEG analysis of amphicrine acinar-01 relative to amphicrine acinar-02 and -03 sub-states revealed differentially upregulated genes like *CLPS, CPA1, SERPINI2, CPA2, PDGFD, CTRB1* and *GALNT17* (avg log2FC > 1.5 & Benjamini-Hochberg adjusted p-value < 0.01; Wilcoxon rank sum test) **(Figure 5.10).** The concurrent presence of metaplastic signatures and acinar secreted enzyme encoding genes (like *CPA1* and *CPA2)* suggested an intermediate phenotype bridging early and late metaplastic states, which is known to be induced during pancreatic injury and ADM [604]. Of note, Amphicrine acinar02 exhibited relatively increased expression of serotonin biosynthesis associated gene *TPH1* [613] as compared to other two amphicrine acinar subgroups **(Figure 5.9a & Extended Data Figure 11)**. In the context of small bowel neuroendocrine tumors (SBNETs), cells with *TPH1* knockdown generated smaller tumors in vivo, which exhibited reduced vascularity [614]. Additionally, the amphicrine acinar02 sub-state strongly expressed a few ECM remodeling genes such as *ABI3BP, FKBP9*, *IGFN1* and *PROM1* associated with ductal-like features **(Figure 5.9a & 4.10)**. Taken together, amphicrine substates co-expressed markers of multiple pancreatic lineages or their progenitors, which have been described in the trans-differentiation of pancreatic cells in mouse models of pancreatic injury and regeneration, where they can give rise to sub-states with enteroendocrine or gastric differentiation profiles [615]. A proliferating amphicrine sub-state called amphicrine acinar proliferating was detected, having a similar molecular profile to the shared neuroendocrine proliferating sub-state, and characterized by markers like *MKI67, DIAPH3, CENPP, BRCA1* and *FANCA* **(Figure 5.9a).**

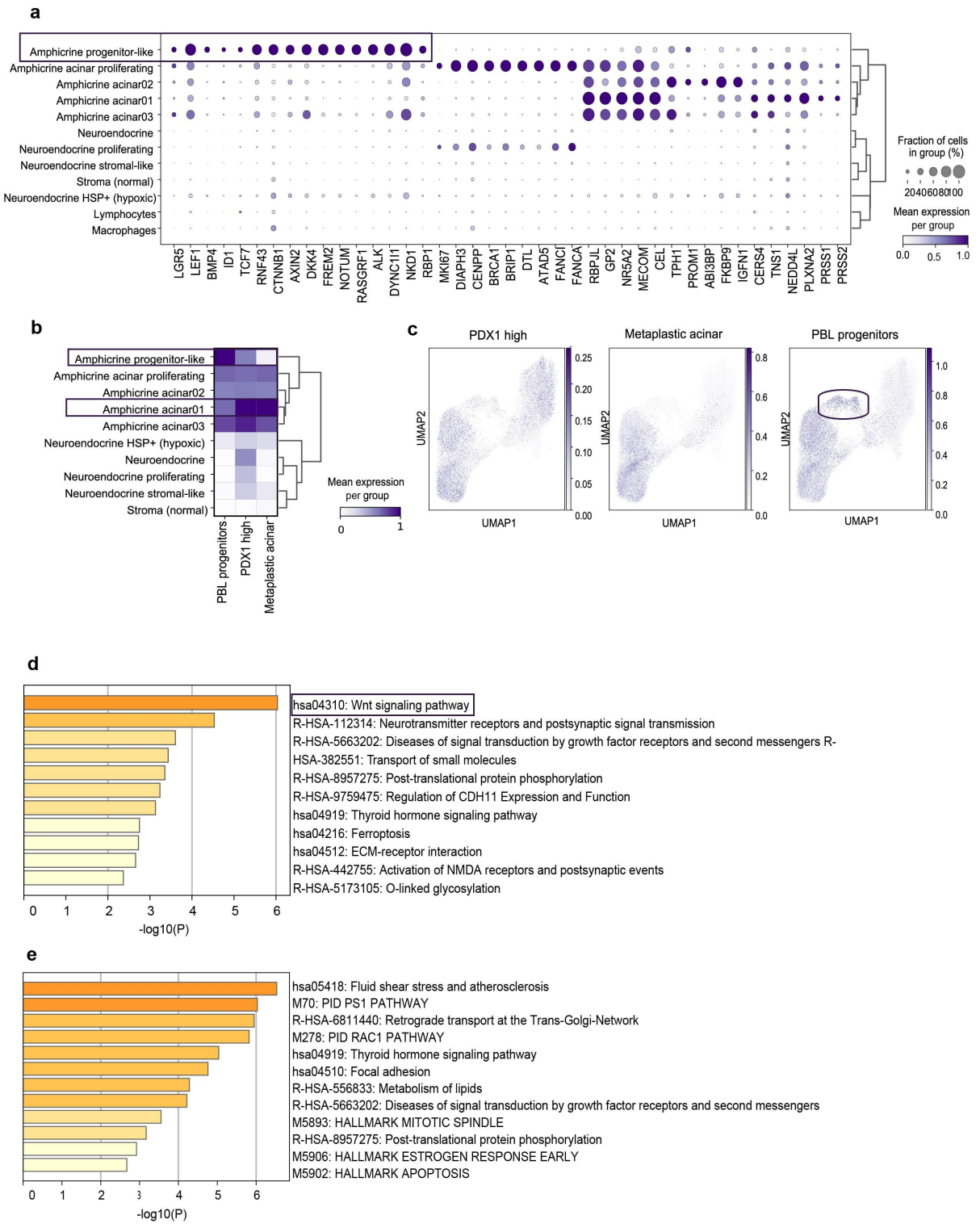


Figure 5.9: Amphicrine progenitor-like characterized by WNT-BMP-NOTCH signaling pathway.

(a) Dot-plot illustrating key positive markers for amphicrine progenitor-like and other amphicrine acinar sub-states inferred using multivariate Logistic Regression (avg log2FC >= 0.25 & Bonferroni corrected p-value < 0.01). Only genes specific to amphicrine sub-states were included for clarity in visualization; however, the entire plot encompassing all sub-state markers can be referred to in Extended Data Figure 11. The markers crucial for amphicrine progenitor-like sub-state are highlighted. Mean gene expression within a sub-state is color-coded, and the percentage of nuclei expressing a marker is size-coded. (b) Matrix plot displaying module scores computed using signatures of PBL progenitors from

[608], PDX1 high pancreatic cells from [612], and metaplastic acinar cell state from Schlesinger et al. 2020 [604]. (c) Feature plot demonstrating module scores depicted in (b) per nucleus. The amphicrine progenitor-like sub-state is circled to emphasize the enrichment of PBL progenitor score for clarity. (d) Bar graph presenting enriched pathways and processes identified using upregulated markers of the amphicrine progenitor-like sub-state. (e) Bar graph depicting pathway enrichment analysis using overlapped markers between amphicrine progenitor-like and Neuroendocrine HSP+ (hypoxic) sub-states. A discrete color scale is utilized to represent statistical significance, with gray indicating a lack of significance.

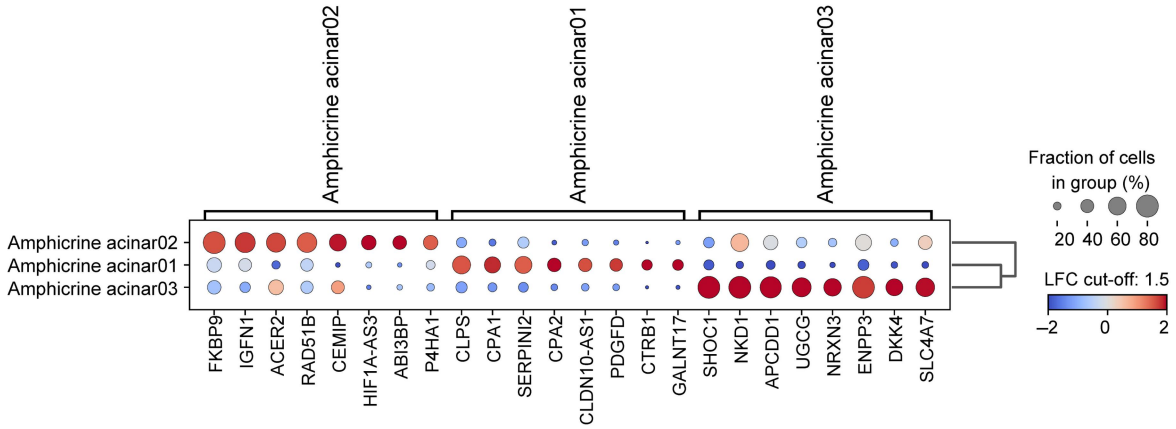


Figure 5.10: Classification of Amphicrine acinar sub-states within panNEC.
Dot-plot illustrating differentially expressed markers per amphicrine acinar sub-states. DEG(s) were recalculated using only amphicrine acinar sub-states and by using Wilcoxon rank sum test (avg log fold change or LFC >= 1.5, Benjamini-Hochberg corrected p-value < 0.01). LFC (avg log2FC) per marker within a sub-state is encoded by the dot color. Fraction of cells in a sub-state (in %) expressing a marker is size-coded.

5.3 Lineage plasticity and malignant cell state programs in panNEC

5.3.1 Pancreas development

Considering the influence of lineage plasticity on malignant behavior, a module score analysis was conducted using curated signature lists derived from various studies centered on pancreatic development, with a focus on acinar [547], ductal [616] and endocrine lineage differentiation [619, 620]. Further, an over-enrichment test based on the hypergeometric distribution (see Methods) was performed to determine if the observed enrichment for a certain signature was also statistically significant. The P5-specific amphicrine progenitor-like sub-state was found to be enriched for gene-sets for trunk and endocrine progenitors known to differentiate into ductal and endocrine cell types or states, respectively **(Figure 5.11).**

However, this sub-state revealed no enrichment for the tip progenitor- a population typically located at the tips or edges of developing pancreatic buds and is known to differentiate into mature acinar cells **(Figure 5.11)**. This can be explained by lack of *PTF1A, CEL, RBPJL, CPA1* and digestive enzymes like *PRSS1* in this sub-state as discussed before **(Figure 5.9a & Extended Data Figure 11)**. On the contrary, the amphicrine acinar sub-states exhibited variable expression of the tip progenitor module, with the most significant enrichment observed in amphicrine acinar 01 followed by amphicrine acinar 03 sub-state **(Figure 5.11).** This could be explained by the significant expression of key acinar markers including *RBPJL*, *GP2*, and *CEL* in the amphicrine acinar sub-states **(Figure 5.9a)**.

In a similar vein, another investigation was performed to find resemblance of the inferred panNEC sub-states to various types of ductal progenitor signatures extracted from Qadir et al. 2020 [616] **(Figure 5.11).** While the NE HSP+ (hypoxic) sub-state apparently showed strong expression for both trunk and migratory ductal progenitor module scores, it was later found that this observation was driven by few features and lacked statistical significance. On the other hand, amphicrine acinar01 demonstrated significant enrichment for Ductoacinar 1 and 2 signatures, implying a transition state between acinar and ductal, and plasticity owing to de-differentiated acinar cells [616]. Additionally, the amphicrine acinar sub-states were found to be enriched for signatures of centro-acinar cell type that are believed to harbor pancreatic multipotent progenitors [617] **Figure 5.11)**. Moreover, the centroacinar cells were suggested to have a direct role in oncogenic transformation through their ductal transdifferentiation [618].

Subsequently, the transcriptomic resemblances between panNEC sub-states, endocrine progenitors and cell types of endocrine lineage were explored [619, 620] **(Figure 5.11)**. Module score analysis showed a mutual exclusivity between the gene-sets indicative of alpha and beta cell fate programs, with alpha being associated with shared panNEC sub-states and beta with amphicrine ones. Especially, beta cell program was found to be most enriched in amphicrine acinar 01 **(Figure 5.11)**. Moreover, amphicrine acinar 01 displayed enrichment for gamma cell types, while amphicrine progenitor-like sub-state exhibited enrichment for delta cells **(Figure 5.11)**. It is important to mention that the hypergeometric test did not reveal any significant enrichment for the alpha cell program in the shared sub-states and hence, should be carefully interpreted. Notably, none of the key markers associated with mature endocrine cell fates, such as *GCG, INS,* and *SST* were expressed in any of the sub-states **(Figure 5.12).** Other critical genes defining the alpha lineage like *MAFB* and *IRX2* were also not expressed **(Figure 5.12)**. Albeit *ARX,* a TF known to direct alpha cell development [620, 621] was very lowly expressed in the shared sub-states— it specifically

stemmed from P1 sample. On a similar note, prominent beta markers like *IAPP* and *MAFA* were negligibly expressed in the panNEC sub-states **(Figure 5.12).** Although *DLK1, TGFBR3,* and *SMAD9* expression contributed partially to the enrichment of the beta program in amphicrine sub-states **(Figure 5.12)**, their presence does not necessarily signify a typical beta-like NEC in the absence of *INS*. For instance, *DLK1* is known to be expressed in the developing pancreas as well as in a large variety of cancers, including well differentiated NET and SCLC [622, 623]. Furthermore, *SST* and *PPY*- characteristic markers for delta and PP cell types were not expressed in the panNEC sub-states **(Figure 5.12).** The lack of characteristic lineage markers such as *GCG* or *INS* confirmed that the panNEC sub-states did not recapitulate the transcriptomic profile of mature islets, unlike their well-differentiated pNET counterparts that is known to closely resemble the transcriptomic profile of alpha, beta, gamma, or delta cell types.

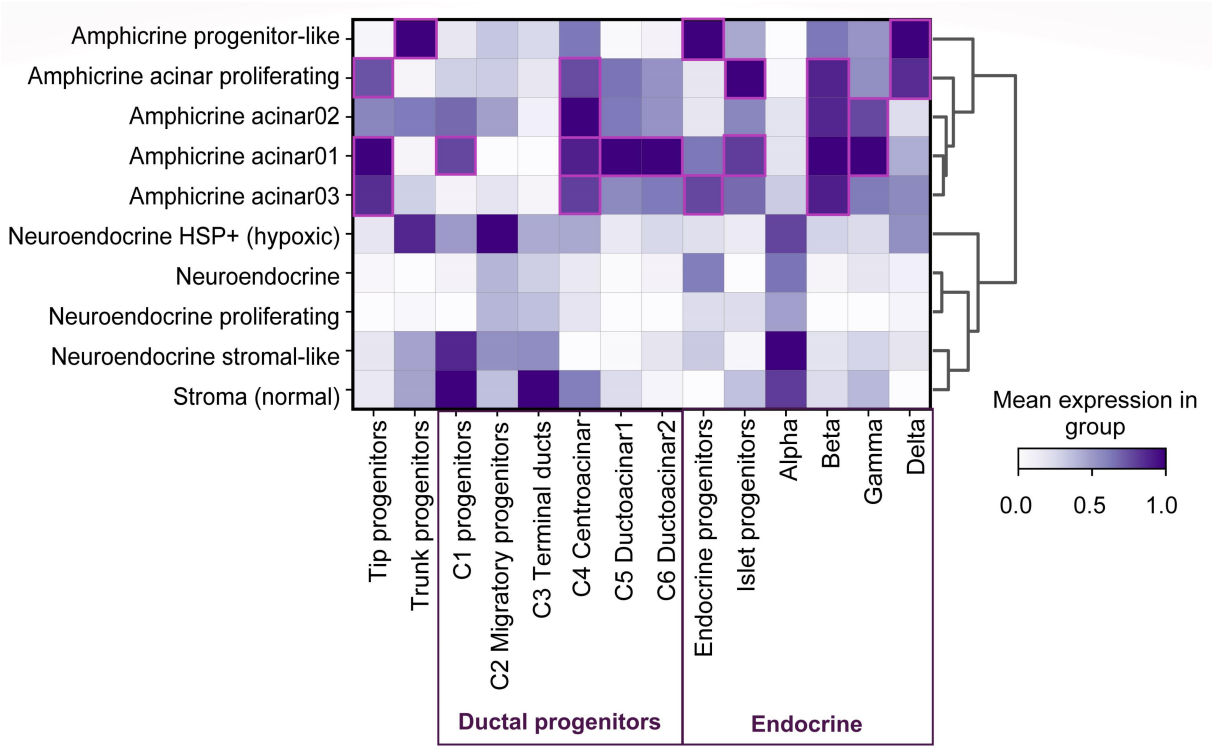


Figure 5.11: Transcriptomic similarity of panNEC sub-states with signature modules indicative of developmental (or, lineage) plasticity.

Matrix plot depicting mean normalized scores of various cell type/state signature modules associated with pancreatic development. Module scores represent the mean expression of a specified gene-set subtracted by the average expression of a reference set of genes, with the reference set randomly sampled from the data for each binned expression value using scanpy. Marker lists for ductal progenitors were curated from Qadir et al. 2020 [616], a scRNA-seq study profiling variability of major pancreatic duct progenitor cells; signatures for Endocrine & Islet progenitors were specifically extracted from Ramond et al. 2018 [619], while alpha and beta cell type signatures were sourced from

Muraro et al. 2016 [294] and Sean et al. 2022 [620], and later combined during module scoring. Gamma and Delta markers were taken from Muraro et al. 2016 [294]. Gene lists used for computing module scores were included in Supplementary Table 25. Sub-states demonstrating statistically significant over-enrichment for a signature (representation factor > 3 and hypergeometric p-value < 0.01) are boxed in purple. The p-value indicating under- or over-enrichment was calculated using the cumulative distribution function (CDF) of the hypergeometric distribution. While Neuroendocrine HSP+ (hypoxic) exhibited visibly high expression of trunk progenitors and C1 ductal progenitors, this observation was driven by two features and was not a statistically significant enrichment. Similarly, the enrichment of neuroendocrine stromal-like, and stroma (normal) for C1 progenitors and alpha module were not significant. Therefore, these instances and others not outlined in purple indicate a lack of notable enrichment.

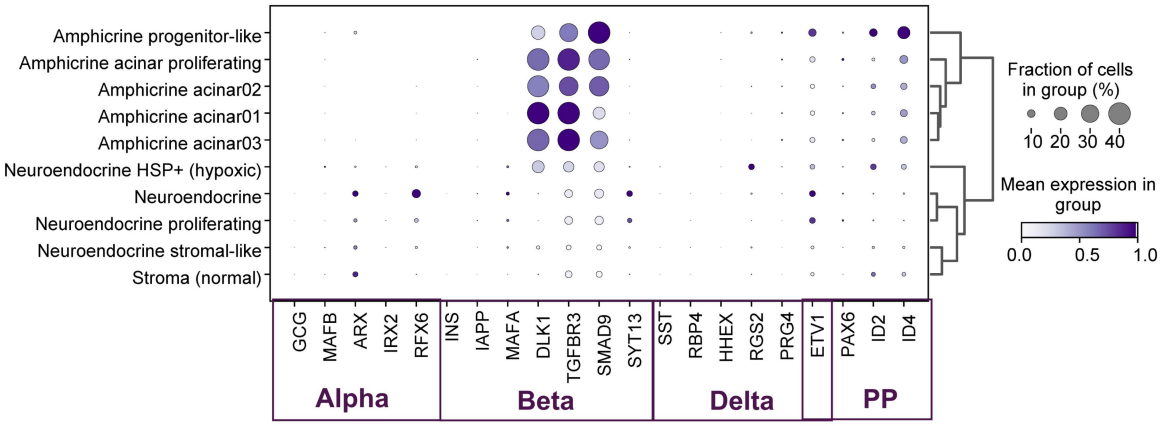


Figure 5.12: panNEC sub-states lack key endocrine lineage drivers— GCG, INS and SST

Dot-plot depicting characteristics markers defining mature endocrine cell types comprising alpha, beta, delta, and pancreatic polypeptide (PP). Markers were curated from Muraro et. al 2016 [294] and verified from other sources [619, 620]. GCG, INS and SST are established lineage markers for alpha, beta and gamma cell types respectively. PPY, characteristic marker of PP cell type was not expressed in the dataset. Average gene expression within a sub-state is color-coded and percentage of cells (or, nuclei) expressing a marker is size-coded.

5.3.2 Neuroendocrine carcinoma

NE differentiated cancer sub-states are increasingly recognized across the full spectrum of cancer entities [624, 625], with particularly high frequencies in lung and androgen resistant prostate cancer [624-627]. Indeed, the clinical management of panNEC largely follows the experience gathered from trials in NE small cell carcinomas of the lung, which occur at much higher frequency [628]. Hence, transcriptomic signatures observed in NEC of non-pancreatic origins were utilized to determine whether their malignant phenotypes were recapitulated by

the panNEC sub-states. Specifically, module score analysis and an over-enrichment test (see Methods) were employed to assess the enrichment of signatures within the panNEC sub-states using publicly available signatures of malignant sub-types and class programs from lung and prostate NEC datasets **(Figure 5.13).**

Signatures were initially extracted from the study by George et al. 2018 [340], which focused on high-grade lung NEC subtypes, including a sizable cohort of LCNEC, which was considered relevant to the pancreatic LCNEC sample set presented in this thesis. Signatures derived from exclusive clustering of lung LCNEC samples resulted in a broad segregation of two subtypes (Type I and II) that were associated with distinct mutational profiles; either with bi-allelic *TP53* and *STK11/KEAP1* alterations (Type-I) or bi-allelic inactivation of *TP53* and *RB1 (*Type-II*).* It is important to note that the transcriptional profile of Type-I was defined as *ASCL1^{high}/DLL3^{high}/NOTCH^{low}* and reported to possess high neuroendocrine expression. Based on this observation, Type-I LCNEC transcriptionally mirrors 'SCLC-like tumors' despite sharing a mutational profile like NSCLC [652, 653] **(**as marked in **Figure 5.13).** In contrast, Type-II LCNECs demonstrated reduced expression of neuroendocrine genes and *ASCL1^{low}/DLL3^{low}/NOTCH^{high}* pattern in George et. al 2018 study— hence, it resembles NSCLC tumors based on transcriptomic pattern [652] **(**as marked in **Figure 5.13).** Additionally, the authors described distinct malignant sub-classes (Class D and E), which crossed the conventional classification categories of small cell lung carcinoma (SCLC) and LCNEC. Signatures of these sub-classes were found to broadly separate the NE lung cancer types from other histotypes of lung cancer.

The Type-I LCNEC signature exhibited variable expression across most panNEC sub-states, with the highest statistical enrichment specifically observed in amphicrine acinar01 **(Figure 5.13).** Interestingly, both the proliferative and progenitor sub-states completely lacked representation of the Type-I signature. Biologically, LCNEC subtype type-I has been linked to elevated ATP synthesis, oxidative phosphorylation, and electron transport chain activity, but reduced cell cycle and DNA repair that broadly aligns with the transcriptomic profile of the amphicrine acinar and shared NE sub-states. Instead, the Type-II lung LCNEC showed robust over-enrichment for the amphicrine proliferating sub-state (representation factor: 10.1, hypergeometric p-value < 6.287e-17, over-enrichment test based on hypergeometric distribution), though lesser- for the shared NE proliferating sub-state (representation factor: 5, hypergeometric p-value < 0.003) **(Figure 5.13).** This aligns with the previously described type-II feature, which has been linked to increased cell activation and proliferation. Enrichment in NE proliferating was driven by genes like *DNMT1, POLA1, FANCC, RNGTT* and *TIAM1*. Dysregulation of *DNMT1* activity can lead to aberrant DNA methylation patterns,

contributing to cancer development and progression by silencing tumor suppressor genes or activating oncogenes [629]. Mutations in *FANCC* and other FA pathway genes can lead to chromosomal instability and increased susceptibility to cancer, particularly hematologic malignancies, and solid tumors [553]. *TIAM1* is implicated in cancer metastasis by promoting tumor cell invasion and migration through cytoskeletal rearrangements and changes in cell adhesion properties [630].

Further signatures from the George et al 2018 [340] study, which represent distinct transcriptomic sub-classes within neuroendocrine lung cancers also exhibited distinct enrichment patterns in the panNEC sub-states. Both Class I and II NE lung cancer sub-class signatures were highly represented in the amphicrine sub-states **(Figure 5.13)**. Also, the Class I signature that is representative of lung LCNEC with strong NE marker expression was moderately represented by the shared NE sub-state (representation factor: 3.5, hypergeometric p-value < 0.008) **(Figure 5.13)**. In contrast, shared panNEC sub-states generally lacked expression of the class II signature, which in lung cancer denoted predominantly LCNEC with lower-level NE marker expression. **(Figure 5.13).** Importantly, SCLC-like transcriptome features (Class III and IV) were strongly over-enriched in both shared NE and NE-proliferating sub-states (representation factor: 3.5 respectively, hypergeometric p-value < 0.008) **(Figure 5.13)**. Taken together, these analyses suggested that despite their large cell morphology, the shared NE substates in this panNEC cohort share transcriptomic features with SCLC rather than with LCNEC of the lung.

On the other hand, class D and E signatures were upregulated in both amphicrine and NE proliferating sub-states **(Figure 5.13)**. This is mainly because of class D and E being associated with chromatin modifications, DNA repair, and DNA damage response. More specifically, the class D signature that represented the majority of lung LCNEC was relatively more enriched in the proliferating amphicrine sub-state (representation factor: 9.2, hypergeometric p-value < 1e-09), than its shared counterpart (representation factor: 5, hypergeometric p-value < 0.003). However, the shared NE proliferating revealed stronger over-enrichment of the class E signature derived predominantly from SCLC samples (representation factor: 6, hypergeometric p-value < 0.003) relative to LCNEC dominant Class-D. Considering the clinical relevance of shared NE-proliferating, markers that contributed to the enrichment of Class D and E signatures were inspected. Again, *DNMT1* turned up together with genes such as *CCDC14, KNTC1*, and *RBL1* for Class-D, and *LIG1, NASP, POLA1, CACNA1A, CBFA2T2,* and *KHDRBS1* for Class-E **(Supplementary Table 23)**. Collectively, these genes are implicated in DNA repair mechanisms, maintenance of genomic integrity, and neuronal regulation. In a previous study, Naert et al. 2020 [631]

demonstrated that *RBL1 (p107)* acts as a tumor suppressor in glioblastoma and small-cell pancreatic neuroendocrine carcinoma using a Xenopus model. Lazaro et al. 2019 [632] linked *RBL1* inactivation to the differential development of LCNEC, whereas George et al. validated the involvement of *RBL1* in the progression of human SCLC.

Differentially upregulated markers for both NE and NE-proliferating sub-states also demonstrated a significant over-enrichment of a prostate cancer signature extracted from Alshalalfa et al. 2019 [341], that was found suitable to distinguish small cell NE prostate cancer from adenocarcinoma in treatment naïve, early stage specimens (representation factors of 10.7 and 11.8, respectively, hypergeometric p-value < 1e-07) **(Figure 5.13)**. To be precise, this observation was driven by markers, including *CAMK1D, CHD7, CACNA1A, KCNB2, NFASC, NRXN1* for both sub-states, and additionally by *EZH2, RFX2, LIG1, LMNB1* in NE-proliferating. Similarly, both sub-states exhibited significant enrichment in upregulated NE markers obtained from a study focused on an SCLC mouse model [633].

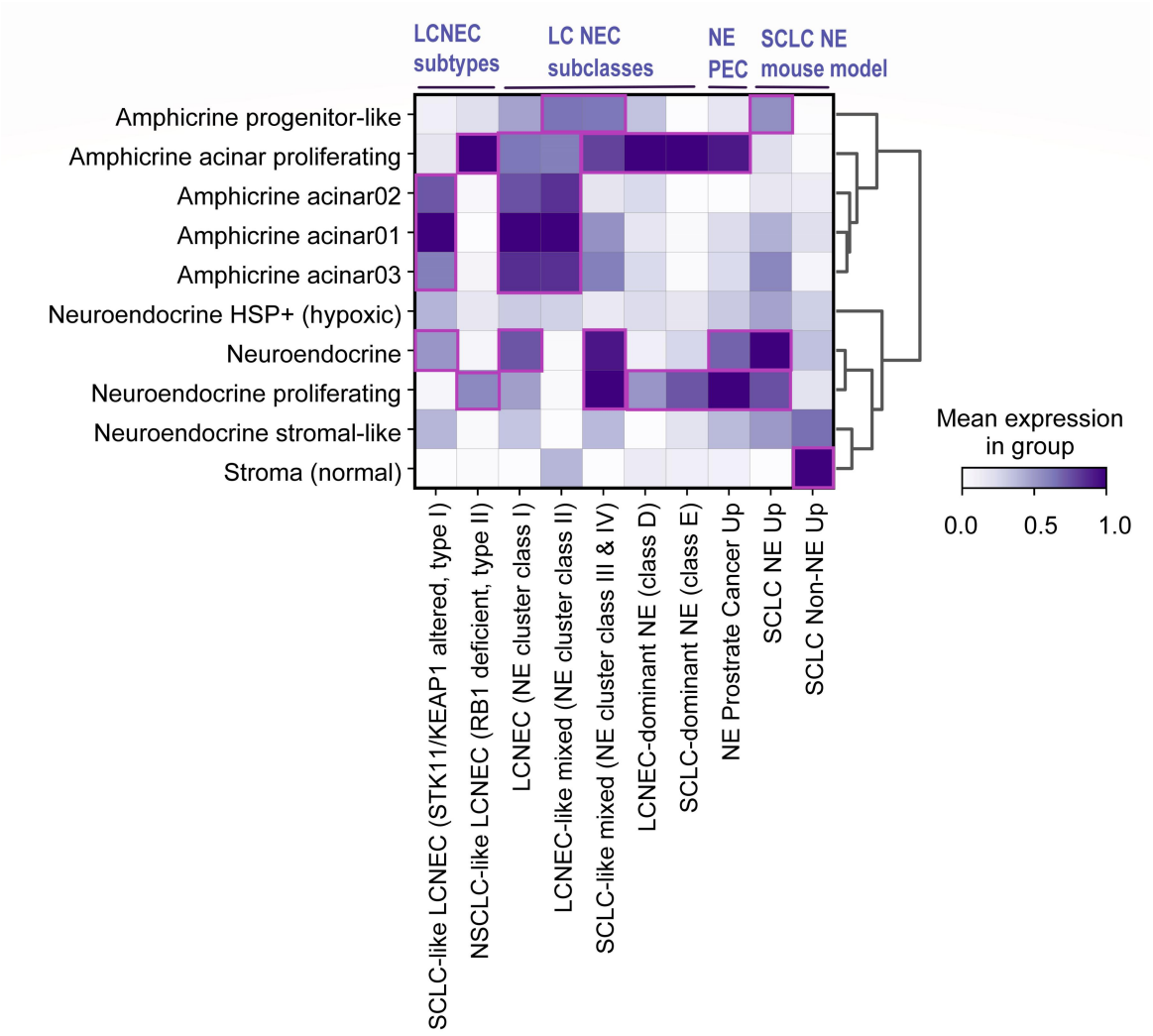


Figure 5.13: Transcriptomic similarity of panNEC sub-states with signature modules indicative of neuroendocrine lung and prostate cancer subtypes and subclasses.
Matrix plot illustrating the normalized mean score of curated signature modules extracted from lung and prostate NEC. The module score represents the mean expression of a specified gene-set subtracted from the average expression of a reference set of genes, with the reference set randomly sampled from the data for each binned expression value as implemented in scanpy. Signatures corresponding to LCNEC subtypes (type-I and II), lung NEC sub-classes (I, II, III+IV), and Class-D and E were sourced from George et al. 2018 [340]. Additionally, signatures upregulated in NE Prostate Cancer were included from Alshalalfa et al. 2019 [341]. Differentially upregulated NE and non-NE genes in an SCLC mouse model were extracted from [633]. Gene lists used for performing module score analysis was included in Supplementary Table 25. Enclosed in purple boxes are sub-states with statistically significant over-enrichment for a malignant signature (representation factor > 3 and hypergeometric p-value < 0.01). The p-value denoting under- or over-enrichment was calculated using the CDF of the hypergeometric distribution. Note: According to previous data [340, 652], type I LCNEC exhibited a mutational pattern resembling NSCLC, yet displayed transcriptomic profiles characteristic of SCLC. Conversely, type II LCNEC demonstrated a mutational pattern akin to SCLC but exhibited low expression of neuroendocrine markers, akin to NSCLC. Hence, type I and II LCNEC(s) were marked as SCLC-like and NSCLC-like respectively, based on transcriptomic patterns.

5.3.3 Pancreatic Ductal Adenocarcinoma

Previous literature suggests that panNEC shares the mutation spectrum of PDAC rather than pancreatic NET [324, 636]. Also, panNEC can co-exist with PDAC in a rare phenomenon known as mixed or combined neuroendocrine-non-neuroendocrine neoplasm (MiNEN or MANEC) [637, 638]. These tumors typically exhibit both neuroendocrine and non-neuroendocrine components within the same lesion [638, 639]. This observation inspired an investigation into whether any of the panNEC sub-states transcriptomically recapitulate malignant cell states and lineage programs observed in PDAC. To achieve this objective, a module score analysis was conducted using disease signatures compiled from Hwang et al. 2022 [342], a snRNA-seq study that characterized PDAC tumor specimens subjected to either neoadjuvant therapy or from treatment-naïve conditions **(Figure 5.14)**. Of note, the latter inferred several malignant programs representing either cell state (Cycling-S, Cycling-G2/M, MYC, Adhesive, Interferon, TNF-NFκB signaling) or lineage (Acinar-like, Classical-like, Basaloid, Squamoid, Mesenchymal, Neuroendocrine-like, Neural-like progenitor). Next, an over-enrichment test based on hypergeometric distribution (see Methods) was performed to assess the significance of enrichment. The signatures of 'Classical-like' showed strong overlap with the previously defined Moffitt's classical signature [342, 634].

As expected, both NE and amphicrine proliferating sub-states revealed significant over-enrichment for Cycling (S) and Cycling (G2/M) programs associated with cell-cycle regulation and *E2F1* targets **(Figure 5.14)**. Amphicrine acinar-like 01 demonstrated the most significant enrichment for the 'acinar-like' program. Of note, it also showed intriguing characteristics concerning the Classical-like subgroup **(Figure 5.14)**. This aligned with the earlier findings in **Figure 5.8**, where substantial expression of genes that contribute to the classical phenotype signature, e.g. *GATA6* and *HNF4A* were demonstrated in the amphicrine sub-states. *GATA6*, a transcription factor pivotal for maintaining epithelial identity and promoting acinar cell differentiation, is known to be notably expressed in the classical PDAC subtype [605]. A study by Kloesch et al. 2022 [635] revealed that *GATA6* loss, accompanied by concurrent downregulation of *HNF4A*, is necessary for transitioning to the basal phenotype in PDAC. Importantly, this loss of Gata6 was also linked to enhanced metastatic potential and hence, more aggressive phenotype.

The stroma (normal) sub-state demonstrated elevated scores for basaloid and squamoid signaling programs **(Figure 5.14).** However, this overexpression was primarily driven by specific genes such as *VIM* and *WSB1*, resulting in an insignificant enrichment (hypergeometric p-value < 0.461 & < 0.107, respectively). Both basaloid and squamoid

programs were reported to show substantial overlap with the Moffitt basal-like signature established for PDAC [634], suggesting that stroma (normal) cannot be directly associated with PDAC. Mesenchymal program was robustly over-enriched in stroma (normal) sub-state (Representation factor: 7.8 & hypergeometric p-value= 4.474e-09) **(Figure 5.14)**. Nevertheless, stromal cells or CAFs are typically anticipated to exhibit upregulated mesenchymal genes associated with stemness, cell invasion and EMT. Thus, this observation may not be relevant to specific PDAC biology. Additionally, stroma (normal) showed significant over-enrichment for adhesive signatures (Representation factor: 4.4 & hypergeometric p-value < 4.882e-04), Interferon signaling (Representation factor: 7.2, hypergeometric p-value < 3.871e-08) and TNF-NFkb signaling (Representation factor: 8.3 & hypergeometric p-value < 4.790e-10). **(Figure 5.14)** Interestingly, this recapitulated previously known PDAC biology. While the module score analysis suggested a high expression of the squamoid program in shared NE and Interferon signaling in NE-proliferating sub-states, these observations were found to be statistically insignificant when examined using the over-enrichment test.

Moreover, both the shared NE and NE proliferating sub-states displayed notable over-enrichment for the 'neuroendocrine-like' module **(Figure 5.14).** Specifically, the NE sub-state exhibited relatively higher enrichment score (representation factor: 17.6 & hypergeometric p-value < 7.215e-24) as compared to NE proliferating (representation factor: 10 & hypergeometric p-value < 1.810e-14). Of note, the 'neuroendocrine-like' module was associated with genes regulating nervous system development, neurotransmitter secretion, vesicle-mediated and cation membrane mediated transport [342]. Conversely, the shared NE HSP+ (hypoxic) sub-state showed no enrichment for the 'neuroendocrine-like' module, indicating a departure from the typical functional traits of NE cells such as secretion and transport **(Figure 5.14).** Additionally, the shared NE sub-state demonstrated significant over-enrichment for Neural-like progenitor (NRP) targets (representation factor: 9.9 & hypergeometric p-value < 2.010e-10), while NE-proliferating also exhibited enrichment for the same (representation factor: 6.5 & hypergeometric p-value < 1.320e-07) **(Figure 5.14)**. It is worth mentioning that Hwang et al. [342] characterized the NRP malignant program as 'brain tissue enhanced,' associating it with axonal guidance, tumor-nerve interactions, and a high incidence of perineural invasion. Of note, their study demonstrated enriched NRP malignant program in residual PDAC tumors and patient derived PDAC organoids following treatment that was associated with poor prognosis [342].

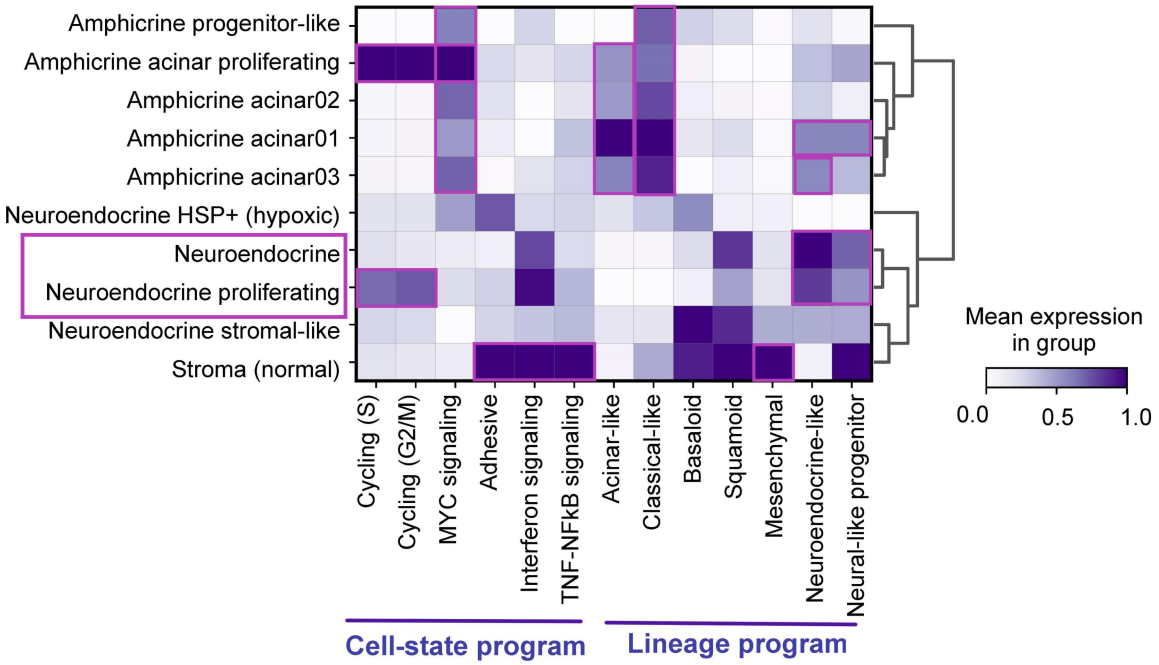


Figure 5.14: Transcriptomic similarity of panNEC sub-states with malignant cell states and lineage programs of PDAC.
Matrix plot displaying the normalized mean score of curated PDAC malignant cell-state and lineage specific signatures sourced from Hwang et al 2022 [342]. The module score represents the mean expression of a specified gene-set subtracted from the average expression of a reference set of genes, sampled randomly from the data for each binned expression value using scanpy. Signature lists used for computing module scores per sub-state were included in Supplementary Table 25. Sub-states exhibiting statistically significant over-enrichment for a signature (representation factor > 3 and p-value < 0.01) are enclosed in purple boxes. The p-value indicating under- or over-enrichment was computed using the cumulative distribution function (CDF) of the hypergeometric distribution. The cases that showed visible enrichment but lacked statistical significance according to hypergeometric test was not marked in purple as they should not be used for drawing conclusions. Shared neuroendocrine and neuroendocrine-proliferating sub-states are highlighted given their substantial enrichment for "Neuroendocrine-like" and "Neural-like progenitor" programs, also extremely relevant in context of panNEC.

5.4 Transcriptional regulation and multi-lineage profiles of panNEC sub-states

To dissect the transcriptional regulation within panNEC sub-states, SCENIC [158] and iRegulon [161] methodologies were employed (see Methods). Hierarchical clustering based on regulon activity revealed two major subgroups, encompassing the amphicrine sub-states in patient P5 and the shared neuroendocrine sub-states.

The transcription factors identified using SCENIC in the amphicrine sub-states were generally also found in healthy pancreas, given their physiological roles in acinar or

endocrine cells. Inferred tumoral transcription factor activity either surpassed physiological activity in the corresponding endocrine or exocrine cell populations (such as, *SOX13, PDX1, RUNX1, PBX1, LEF1*) or was diminished (*SOX9, HNF1B, HNF4G, MECOM, PTF1A*) **(Figure 5.15a)**. Of note, the amphicrine progenitor-like sub-state displayed differential regulation by transcription factors such as *SOX9, HNF1B, SOX13, ONECUT1, PDX1*, and *LEF1* **(Figure 5.15a).** Further, differential gene expression analysis by comparing respective panNEC sub-states with healthy adult pancreas cell types sourced from Tosti et al. [72] (see Methods), revealed that *SOX9* and *HNF1B* were differentially upregulated in the healthy pancreas. This observation was expected as *SOX9* and *HNF1B* are recognized for their roles in regulating the differentiation of pancreatic multipotent and trunk progenitors. Increased *LEF1* activity in panNEC might imply a potential role in NEC initiation, likely through maintenance stemness via transactivation of *Wnt/CTNNB1* responsive genes [640, 641]. In addition, both highly proliferative sub-states shared elevated transcription factor activity for several *E2F* family members **(Figure 5.15a)**. Amphicrine acinar 01 and 03 sub-states demonstrate the highest activity of *PTF1A* and *MECOM* **(Figure 5.15a),** with *PTF1A* primarily involved in specifying acinar cell identity, while *MECOM* is a critical regulator of acinar cell dedifferentiation and pancreatic tumors [543]. Backx et al. 2021 [543] detected *MECOM* expression during acinar development and found it to be reactivated during acute and chronic pancreatitis. Moreover, both sub-states demonstrate strong *HNF4A* activity **(Figure 5.15a)**, which is associated with normal pancreatic development and is necessary for the expression of early pancreatic progenitor genes [642, 643], particularly within the context of the PDAC classical subtype [605]. Overall, within the amphicrine cluster, the amphicrine progenitor-like sub-state showcased the most unique regulatory profile, characterized by lower inferred activity for acinar-related transcription factors, consistent with earlier marker and pathway analyses.

Shared sub-states, NE and NE-proliferating, exhibited distinctive activities of transcription factors involved in pancreatic and/or, brain organogenesis, including *PAX6, ISL1, SOX5, ETV1*, and *TCF3* **(Figure 5.15a)**. Of note, the tumoral activity of these transcription factors in the NE sub-states were elevated when compared to their physiological activity in the endocrine pancreas. More specifically, *PAX6* and *ISL1* are known for their roles in transcriptionally and epigenetically controlling pancreatic islet differentiation and the production of functional alpha and beta cells [643-648]. Furthermore, *ISL1* and *PAX6* regulate survival and differentiation of pancreatic endocrine progenitors [645, 648]. *PAX6* also serves as a master regulator of neuronal development [649-651] and was found to be upregulated in NEC relative to healthy pancreatic cell types in the presented analysis **(Figure 5.15a)**. A notable study by Axelsson et al. 2017 [654] reported that knocking down *Sox5* leads to gene expression alterations akin to type 2 diabetes (T2D) and diabetic animals.

Additionally, *Sox5* knock down has been shown to significantly impact insulin secretion, cause decreased depolarization-evoked calcium influx and β-cell exocytosis [654]. Interestingly, *ETV1*, known as a master regulator in neuronal-derived gastrointestinal stromal tumors [655], was found to be increased in the panNEC samples relative to healthy pancreatic controls. As described before, *NKX2-2*, a multipotent neuronal progenitor marker [651, 656], was specifically expressed in shared NE and NE-proliferating sub-states **(Figure 5.7)**, which is also implicated in islet differentiation, together with *PAX6* and *ISL1.*

Thereafter, iRegulon [161] was applied to unravel the transcriptional regulation landscape of the shared panNEC sub-states based on differentially upregulated targets (average log2FC >= 0.25 & Bonferroni adjusted p-value < 0.01)**.** The shared NE sub-state was specifically regulated by *REST, SOX10*, and *RELA* **(Figure 5.15b)**. Of note, *REST* is a transcriptional repressor and recognized as a master regulator of hypoxia-induced neuroendocrine differentiation in prostate cancer cells [657]. A previous study on metastatic castration-resistant prostate cancer by Labrecque et. al 2019 [658] found that loss of *REST* repressor activity encourages the emergence of a tumor phenotype marked by the expression of NE genes in the absence of androgen (AR) activity. As anticipated, the NE-proliferating sub-state was uniquely regulated by *E2F* family transcription factors and *FOXM1*, both associated with tumor differentiation, proliferation, and metastasis in GEP NENs **(Figure 5.15b)**. On the other hand, the NE HSP+ (hypoxic) sub-state demonstrated significantly high activity of *HSF1* and *HIF1A* **(Figure 5.15b)**, with *HSF1* being known for driving cancer cell growth and survival, associated with poor prognosis [659]. On a physiological level, cross-regulation between *HIF1A* and *HSF1* was previously described [660] and suggested a link between low oxygen response and the activation of heat-shock proteins as an adaptive mechanism. Other regulators of NE HSP+ (hypoxic) sub-state included *ATF5* and *NKX2-2* **(Figure 5.15b)**. *ATF5* expression was found to inversely correlate with patient survival in diverse cancer types [661]. Moreover, it has been linked to cancer cell survival and proliferation and can be targeted to induce apoptosis in cancer cells while sparing normal cells selectively [661]. Moreover, Li et al. 2011 [662] documented that ATF5 was significantly upregulated during cellular stress, and HSP70-mediated stabilization enhanced ATF5 activity, resulting in increased transcription of its downstream targets like BCL2. This highlighted a significant post-translational mechanism through which cancer cells can dramatically upregulate ATF5 activity by prolonging ATF5 half-life.

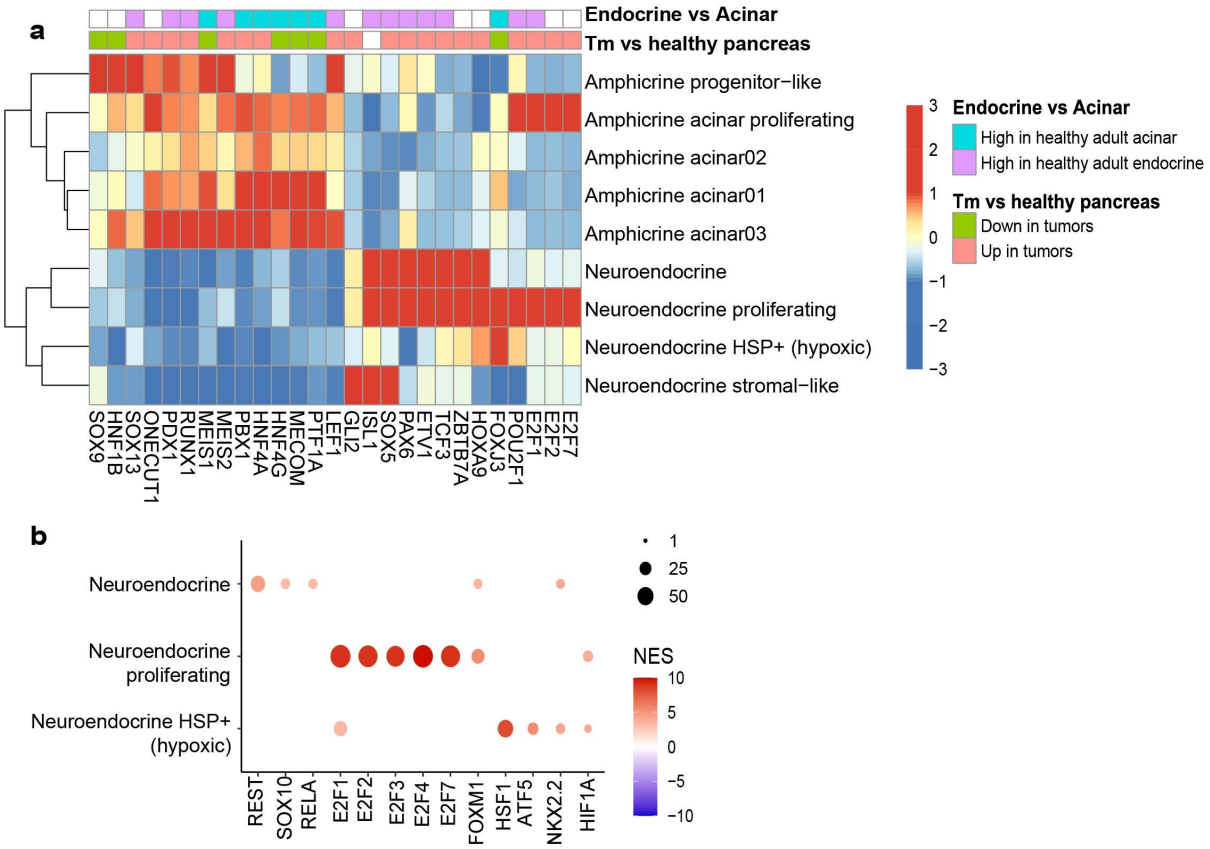


Figure 5.15: Transcriptional regulatory landscape underlying panNEC progression and pathophysiology.
(a) Differential regulation of transcription factors and regulators across NEC cell sub-states inferred using pySCENIC. The significance status of a transcription factor in panNEC tumors Tm relative to healthy pancreatic cell-types from Tosti et al. 2020 [72] is annotated (see Methods). If a transcription factor is expressed in the healthy pancreas cell types (aggregating exocrine and endocrine cell types/states), a differential gene analysis was performed to determine if it is upregulated in acinar or endocrine cell types (see Methods). Expression values are represented as z-scores. (b) Inference of transcriptional activities based on top-100 differentially expressed targets (upregulated only; Logistic Regression avg_log2FC > 0.25 & Bonferroni adjusted p-value < 0.01) using iRegulon. The number of targets regulated by a transcription factor is size-coded (see Methods), while the normalized enrichment score (NES) is color-coded. A higher number of regulated targets generally indicates a higher centrality of a transcription factor.

5.4.1 *PTF1A* regulated brain signatures associated aberrant neuronal phenotype

Previous investigations, employing DNA methylation and mutational profile analyses of panNEC, indicate an exocrine cell origin [338] and a mutational spectrum that aligns with exocrine PDAC, rather than well-differentiated NET [324, 636]. Within this framework, the neuroendocrine phenotype could develop through acquisition of characteristics resembling

either neural or neuroendocrine islet cell types. Subsequently, this study sought to determine if any of the panNEC sub-states exhibited tissue-specific target enrichment for key developmental factors such as *PTF1A, PAX6,* and *NKX2-2*. This analysis was guided by the distinct regulatory roles these transcription factors play in the pancreas and the brain, each associated with a distinct set of targets. For instance, *PTF1A* is indispensable regulating the proliferation of multipotent pancreatic progenitor cells as well as in specifying and in sustaining acinar cells [547, 590, 591, 595]. Conversely, *PTF1A* also regulates brain and neuronal development — precisely, molecular specification of inhibitory neurons and generates correctly balanced neuronal circuits [663-665]. However, the brain vs pancreas specific enrichment of *PTF1A* was not captured in the SCENIC analysis as this (and similar) tool(s) does not incorporate tissue-specific target information in its reference database.

Since *PTF1A* is not expected to be present in non-acinar cells in the fully differentiated adult state, murine tissue-specific developmental signatures from E12.5 neural tube and E17.5 pancreas could be utilized for analysis. Subsequently, specific targets regulated by *PTF1A* in the pancreas and brain (precisely, neural tube) were obtained from Meredith et al. 2013 [666], which elucidated the tissue-specific functions of *Ptf1a* by identifying bound genomic regions during the development of both tissues in vivo. Of note, *PTF1A* regulated pancreatic target signature was substantially enriched in the amphicrine acinar 01, and variably expressed in the other amphicrine sub-states **(Figure 5.16a)**, as previously revealed by SCENIC analysis **(**referring to **Figure 5.15a)**. In contrast, the shared NE and NE-proliferating sub-states did not express the pancreatic targets module. This discrepancy arose because the *PTF1A*-regulated pancreatic targets predominantly encompassed genes encoding digestive enzymes and that with secretory functions associated with the normal and inflammatory conditions of the exocrine pancreas, such as Chronic Pancreatitis. Instead, this analysis observed a notable over-enrichment of *PTF1A*-regulated brain target signatures in the NE proliferating sub-state, featuring genes like *HPIBP3, KHDRBS1*, *SSBP3, THRAP3, PRRC2B, ARID1A, LMNB1,* and *NCAM1* (representation factor of 5.33 & hypergeometric p-value = 0.0001420) as shown in **Figure 5.16a, b.** *HP1BP3* activation of *WNT7B* through *EZH2* interaction mediates drug resistance in glioblastoma cells, with correlated increases in cell proliferation and temozolomide resistance [667]. Additionally, *KHDRBS1*, upregulated by c-Myc, plays roles in various cancers, influencing proliferation, survival, and cancer stem cell vulnerability via Wnt/β-catenin signaling [668-671]. Meanwhile, *SSBP3*, associated with pancreatic β-cell function, affects neuronal morphology and synaptic vesicle biogenesis [672, 673], whereas *THRAP3* regulates the DNA Damage response, influencing sensitivity to DNA-damaging agents [674]. Finally, *PRRC2B* facilitates protein translation, crucial for cell cycle progression and proliferation [675], suggesting potential therapeutic avenues for pancreatic

neuroendocrine carcinoma. In Alshalalfa et al. 2019 [341], *LMNB1* and *NCAM1* were identified as differentially upregulated in primary prostatic tumors resembling prostate small cell neuroendocrine carcinoma. Although the shared NE and NE HSP+ subtypes exhibited a slight overrepresentation of *PTF1A* brain signatures (representation factor: 2.8 and 2.3, respectively), these findings lacked statistical significance (hypergeometric p-values > 0.05, with observed p-values of 0.1 and 0.07, respectively). Except for *NCAM1*, none of the *PTF1A* brain targets were expressed in adult pancreas cell types or states, suggesting their role in neuronal dedifferentiation and maintenance of neuronal functions **(Extended Data Figure 13).**

Amphicrine proliferating cells also showed enrichment for *PTF1A* brain signatures, although to a lesser extent compared to their shared counterpart (representation factor of 2.3 & hypergeometric p-value < 0.022). Further inspection revealed that this enrichment of *PTF1A* brain target signature mostly involves distinct targets between both proliferating substates rather than shared ones**.** *LMNB1* was identified as an LR-inferred marker of amphicrine proliferating cells, contributing to the overlap similarly to NE-proliferating cells.

Interestingly, *ARID1A* is one of the major downstream targets of *PTF1A* involved in regulating both acinar and islet cell development [676-678]. Previously, *ARID1A* loss in pancreas was found to cause islet developmental defect and metabolic disturbance [677]. Importantly, a gene list of the pancreas beta cell hallmark signature, reportedly downregulated in Arid1a-depleted islets in KPC mice (a PDAC model), was secured [678]. This list of genes was utilized to compute a score for the 'ARID1A Islet target' signature and was found to be expressed in the shared NE, NE proliferating, and amphicrine acinar 01 during initial analysis **(Figure 5.16a).** However, no statistical enrichment was noted here (hypergeometric p-value > 0.05 for both cases). This once again related to the earlier discussion regarding the absence of typical islet signatures in the panNEC sub-states.

5.4.2 *PAX6* and *NKX2.2* regulation

As previously mentioned, *PAX6* and *NKX2-2* are key regulators of endocrine cell differentiation, particularly in the development of insulin-producing beta cells. In parallel, they are also key regulators of central nervous system development [649-651, 656]. As no databases or studies have delineated their specific targets in the pancreas and brain, GTEX bulk RNA-seq data [679] was leveraged to conduct DEG analysis to identify genes differentially expressed between these tissues. Subsequently, *PAX6* and *NKX2-2* target lists

from MSigDB [727] were intersected with the DEG list to create two distinct tissue-specific lists **(see Methods).** Similar to the *PTF1A* case, the amphicrine sub-state displayed varying expression levels for pancreatic targets regulated by *PAX6* and *NKX2-2*, reaching its highest enrichment in Amphicrine acinar-like 01 **(Figure 5.16a).** Notably, it was found that NE proliferating sub-state was significantly enriched for *PAX6* regulated brain specific targets (representation factor= 4.4, hypergeometric p-value < 0.013) mediated by *NFASC, PKD1, PRKCE,* and *STXBP1* genes **(Figure 5.16a, b).** Also, the shared NE sub-state demonstrated prominent enrichment for the same (representation factor= 4.7 & hypergeometric p-value < 0.027) **(Figure 5.16a).** Despite the relatively high *PAX6* brain target score observed in the amphicrine progenitor-like cells **(Figure 5.16a)**, this finding lacked statistical significance (hypergeometric p-value < 0.145). Additional examination uncovered intriguing *PAX6* brain targets such as *SYTL2*, *DKK4* and *NRXN3* to be differentially upregulated within this sub-state. However, upon repeating the overlap analysis with the top 100 DEGs, this intersection was determined to be highly significant (representation factor= 7.3 & hypergeometric p-value < 0.008). Hence, it is suggested to repeat this analysis in a larger cohort for future investigations, where this sub-state could be recapitulated in a higher number of samples. In contrast to *PTF1A*, the *PAX6* regulated brain targets were observed to be lowly expressed in adult pancreas cell types or states **(Extended Data Figure 13).** *PRKCE* was the only overlapped target that was robustly expressed in the alpha cell type.

The significant enrichment of the amphicrine progenitor-like sub-state with the *NKX2-2* brain signature (representation factor= 4.7 & hypergeometric p-value < 7.022e-04) **(Figure 5.16a)**. indicates a regulatory function of this transcription factor in the context of P5. This could also be attributed to the WNT regulation of the amphicrine progenitor-like state, as *NKX2-2* levels have been documented to be modulated by TCF4-mediated Wnt signaling in the ventral neural tubes [680]. However, none of the shared sub-states showed any enrichment for either pancreas or brain specific *NKX2-2* targets **(Figure 5.16a)**.

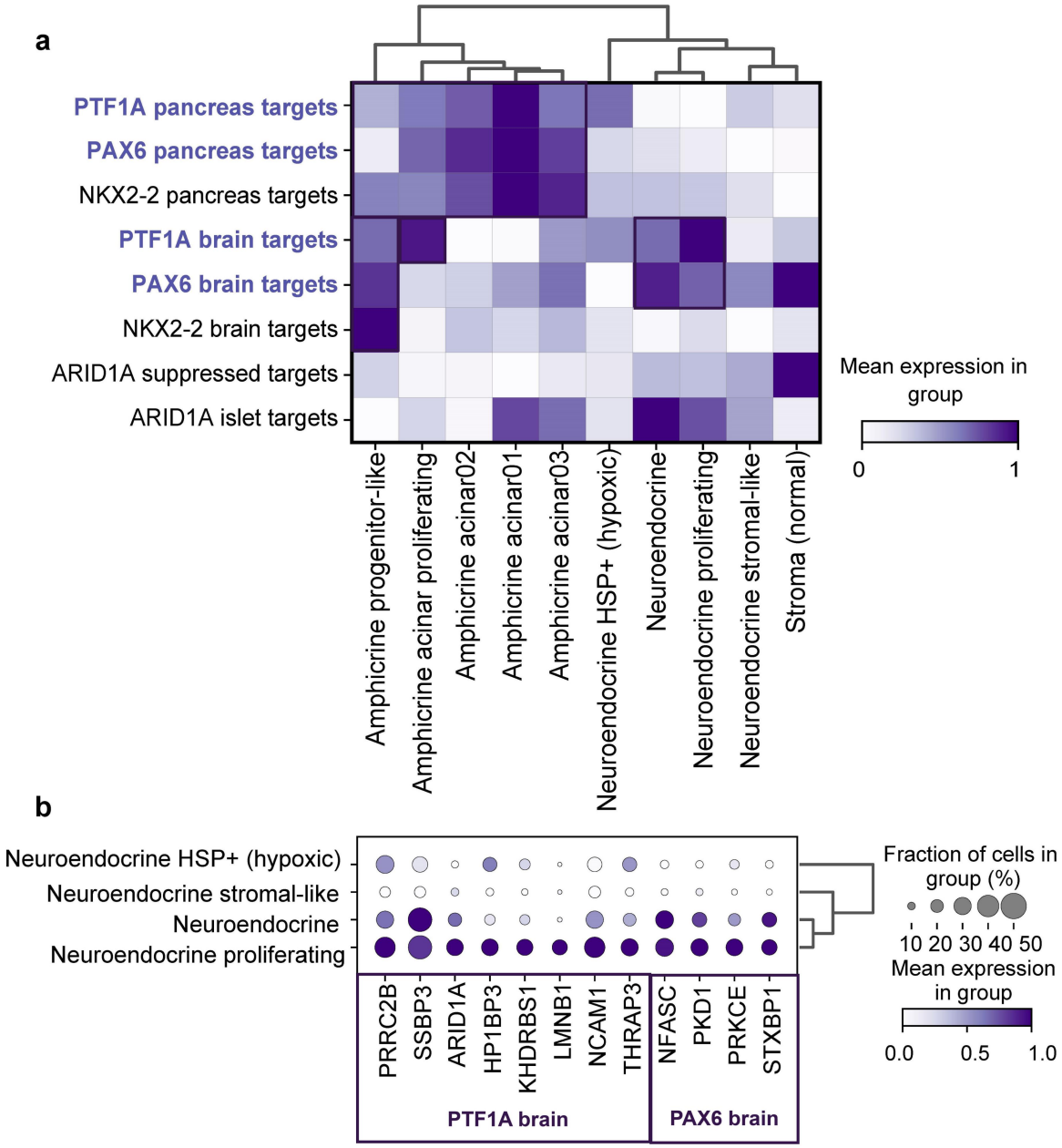


Figure 5.16: Enrichment of PTF1A and PAX6 regulated brain specific targets demonstrating aberrant de-differentiation in the shared panNEC sub-states.
(a) Matrix plot displaying the mean normalized scores of PTF1A, PAX6, and NKX2.2 target genes specific to the pancreas and brain. PTF1A targets were sourced from Meredith et al. 2013 [666], while tissue-specific targets for PAX6 and NKX2.2 were identified using bulk RNA-seq data from GTEx [679] and DESeq2 analysis [728] to identify differentially expressed targets between brain and pancreas (see Methods). ARID1A targets were obtained from a study that profiled targets following ARID1A conditional knockout in KPC mice [678]. TF target lists used for performing module score analysis was included in Supplementary Table 25. The module score represents the mean expression of a specified gene-set subtracted from the average expression of a reference set of genes, sampled randomly from the data for each binned expression value using scanpy. Statistically significant observations according to over-enrichment test were outlined in purple. (b) Dot-plot illustrating the overlapped genes contributing to PTF1A, and PAX6 brain signature over-enrichment in the NE proliferating sub-state. Average gene expression within a sub-state is color-coded, and the percentage of cells (or nuclei) expressing a marker is size-coded. Refer to Extended Data Figure 13 for marker expression across panNEC sub-states and adult pancreas cell type or states from Tosti et al. 2020 [72].

5.4.3 EZH2 regulation

EZH2 orchestrates developmental processes in several organs, including both the pancreas [681, 682] and nervous system [683-686] by modulating gene expression patterns crucial for cell fate determination and tissue morphogenesis. *EZH2* is a well-known transcriptional repressor within the Polycomb Repressive Complex 2 (PRC2). Through its methyltransferase activity, *EZH2* catalyzes the trimethylation of histone H3 lysine 27 (H3K27me3), leading to gene silencing and regulation of various cellular processes. Particularly, *EZH2* is known to regulate the proliferation and fate determination of neural stem cells by repressing various gene sets in the central nervous system [683, 684]. Ezh2 deletion in mouse neural progenitors inhibits progenitor cell proliferation and leads to premature neuronal differentiation and impaired neuronal migration [684-686]. Nevertheless, *EZH2* can also function as a transcriptional activator independently of *PRC2* and its methyltransferase activities, as previously demonstrated in prostate cancer [687]. To investigate effects of *EZH2* tissue specific regulation on panNEC sub-states, a strategy similar to *PAX6* and *NKX2-2* was employed to derive *EZH2* target lists exclusively for brain and pancreas (see Methods).

The amphicrine sub-states exhibited enrichment for *EZH2* pancreas target scores, with the amphicrine acinar01 revealing the most enrichment **(Figure 5.17a).** Notably, the shared NE and NE-proliferating demonstrated substantial over-enrichment for *EZH2* brain specific signatures **(Figure 5.17a)**. The shared NE revealed relatively higher enrichment for the same (representation factor: 2.4 & hypergeometric p-value < 2.013e-09), as compared to NE proliferating (representation factor: 1.6 & hypergeometric p-value < 0.001). In contrast, the pancreas module was not enriched in the shared sub-states **(Figure 5.17a).** Considering that *EZH2* primarily functions as a repressor, *EZH2* pancreas and brain targets expression in specific panNEC sub-states may suggest a diminished *EZH2* repressive activity in a context specific manner. Previous studies like Zhang et al. 2023 [684] have reported prominently up and downregulated gene-sets in neural progenitors and post-mitotic neurons using *EZH2* conditional knockout models. Hence, additional experimental studies are needed to determine exact *EZH2* regulatory mechanisms in panNEC models i.e., if loss of *EZH2* can up or downregulate selected candidates.

Of note, a significant proportion of brain targets mediated the over-enrichment in both shared sub-states (representation factor: 421.4 & hypergeometric p-value < 2.173e-83). The overlapped features included *ADARB2, ADGRB3*, calcium channel encoding genes *CACNA1A, CACNA1C, CACNA2D3,* prominent neurogenesis marker *RBFOX1* and other

neuronal genes *NRXN1, NPAS3, RIMBP2* and *RIMS2*. Overall, these genes function in the regulation of neuronal system, axon guidance, depolarization, synaptic functions, and neurite growth. A former study by Luen Yu et al. 2011 [688] showed that repressed *EZH2* levels stimulate the expression of *PIP5K1C*, leading to the upregulated intracellular calcium signaling. Afterward, a union of markers that contributed to this over-enrichment in both sub-states was compiled, and an iRegulon analysis was conducted to deduce additional transcription factors/regulators influencing the same. *REST* appeared as the most enriched TF (NES= 5.609, #targets= 23) **(Figure 5.17b)**, followed by *GFI1* (NES= 5.483, #targets= 17), and *TEAD4* (NES= 4.791, #targets= 8) **(Supplementary Table 26)**. Of note, *GFI1* is a transcriptional repressor protein crucial for neuronal cell development [689]. Additionally, *Gfi1* acts as an oncogene in SCLC [689, 690]. Another notable transcription factor identified was *NKX2.5* (NES= 3.921, #targets= 7), previously recognized for its role in regulating conduction and contraction in the perinatal heart through the induction of Ca2+ channels and gap junctions [691].The downstream targets regulated either uniquely by *REST* or in combination with three or at least two other transcription factors/regulators (like *GFI1, TEAD4,* and *NKX2.5*; as predicted by iRegulon) were visually presented in **Figure 5.17c.** Interestingly, a moderate portion of *REST* targets were shared with *GFI1* (Jaccard index=0.6 on a scale of 0 to 1) **(Supplementary Table 26).** Moreover, most of the regulated targets (including *DLGAP1, CALN1, GNG4, CAMTA1, NOS1AP*, and *RBFOX1*) were robustly expressed in the NE and NE-proliferating sub-states, as compared to adult pancreas cell types/states **(Extended Data Figure 14)**. These candidates are recognized for their roles in regulating neuronal development, synaptic transmission, and plasticity, suggesting their potential utility in discerning panNEC functions relative to healthy islets.

Further, Metascape enrichment analysis using PaGenBase database [692] revealed that *EZH2* brain signatures were enriched in dorsal root ganglion (DRG) cell type and cerebellar region **(Figure 5.17d).** Additionally, enrichment analysis conducted with the DisGeNET database [693] indicated associations with neurodevelopmental disorders and cerebellar ataxia **(Figure 5.17e)**. In this context, a former study by Feng et al 2016 [694] reported impaired neurogenesis and profound regulatory dysregulations upon *Ezh2* deletion in the murine embryonic cerebellum. Specifically, a conditional KO mouse model (Ezh2cKO) was utilized by Feng and colleagues to study the loss of function effects. Considering that my analysis linked shared NE and NE proliferating targets to the cerebellum, the decision was made to utilize the EZH2cKO downregulated genes from Feng et al. to compute cerebellar specific signature score. Of note, this program displayed moderate enrichment in the shared NE sub-state (representation factor = 4.3, hypergeometric test p-value < 0.033) **(Figure 5.17a),** while this observation was rendered insignificant in the other shared sub-states.

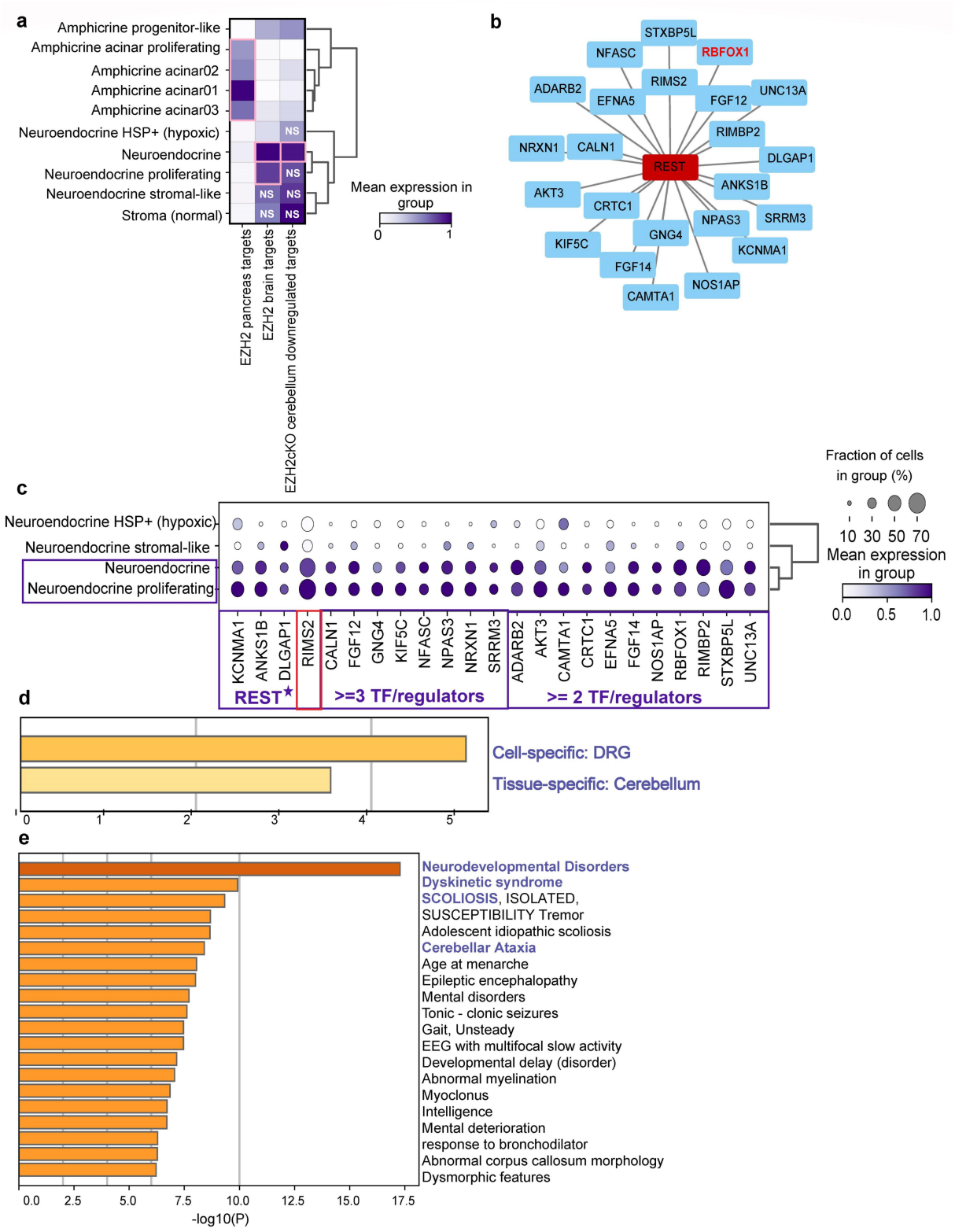


Figure 5.17 : EZH2 regulated brain specific targets demonstrating aberrant de-differentiation in the shared panNEC sub-states.

(a) Matrix plot illustrating the mean normalized scores of EZH2 targets specific to the pancreas and brain. Additionally, EZH2 targets from the cerebellum were sourced from Feng et al. 2016 [694]. The module score represents the mean expression of a specified gene-set subtracted from the average expression of a reference set of genes, sampled randomly from the data for each binned expression value using scanpy. Cases outlined in pink boxes were significant enrichment according to hypergeometric test. Cases revealing a visible (highest) enrichment for a signature score but lacking statistical significance according to over-enrichment test were marked as 'non-significant' (NS) to

prevent confusion. Cases not outlined in pink box were also NS. (b) Network visualization of the downstream targets of REST, identified as the most enriched TF regulator of EZH2 brain-specific signatures (NES= 5.609, #targets= 23). The neurogenesis marker RBFOX1 is highlighted for its relevance to neuronal regulation. (c) Dot-plot depicting the normalized expression of selected EZH2 brain targets jointly regulated by REST, either alone (labeled as REST) or in combination with at least three or two other TF(s) (see Supplementary Table 26). Expression is color-coded, and the percentage of cells (or nuclei) expressing a marker is size-coded. Enrichments of gene lists are identified across (d) PaGenBase [692], and (e) DisGeNET [693] ontology categories performed using Metascape. All genes in the genome served as the enrichment background. Terms meeting criteria of p-value < 0.01, minimum count of 3, and an enrichment factor > 1.5 are collected and clustered based on membership similarities.*

5.5 Intra-patient heterogeneity links to tumoral plasticity and lineage properties

Immunohistochemistry validations confirmed the presence (or, expression) of sub-state markers inferred through snRNA-seq analysis at protein level, while highlighting the substantial variability in individual patients. In the case of Patient P1, Ki-67 staining revealed a moderate ~40% proliferation rate of tumor cells, mainly at tumor margins. Protein levels of EZH2 positively correlated with that of Ki-67, depicting strongest staining at the periphery, but stained negative in the center in case of P1 **(Figure 5.18a-c).** In contrast, P2 demonstrated higher fractions of MKI*67* (*Ki-67)* and *EZH2* positive cells in snRNA-seq data **(Figure 5.18l, m)**, which corresponded to a more homogenous positivity of the tumor cells for both markers. Consistent with the presence of cells with lower level *EZH2* expression in the snRNA-seq data, EZH2 was moderately expressed in sample P1 **(Figure 5.18a-c).** Patient P2 exhibited a higher abundance of the shared NE proliferating sub-state, possibly because of *RB1* loss **(Extended Figure 9 & Table 4.1)**.

In patient sample P1, there was a notable *YAP1* expression confined to the surrounding stroma (non-tumor) but absent in the tumor cells, as illustrated in **Figure 5.18d.** This observation aligned with the previously identified P1 enriched Stroma (normal) sub-state that was found to significantly express *YAP1* **(Figure 5.18k).** Conversely, YAP1 expression was weaker in P2 and reflected heterogeneity of individual glands with or without nuclear and cytoplasmic staining **Figure 5.18h,** which is in-line with snRNA-seq observation as shown in **Figure 5.18n.**

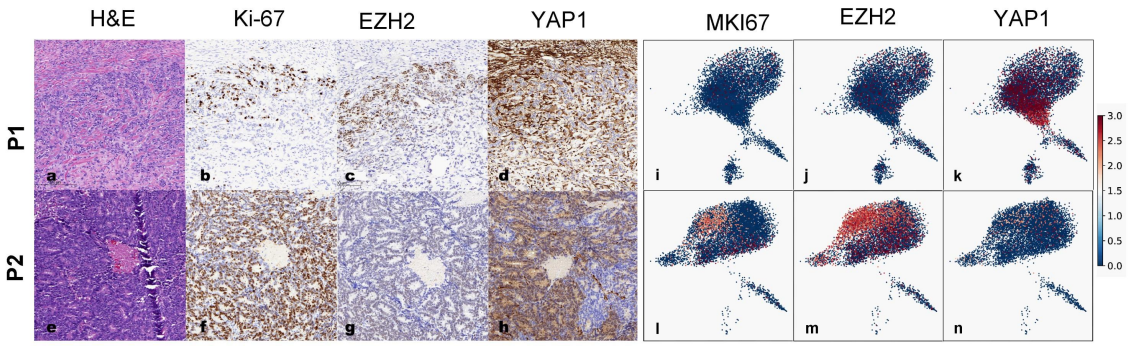


Figure 5.18: Immunohistochemistry depicting variable expression levels of EZH2 and YAP1 in two patient samples.
(a-d) Immunohistochemistry on sample P1 (NEC metastasis), shown at 15x magnification: (a) tumor cell nests at the periphery, single tumor cell infiltrates towards the center of the metastasis (bottom of panel, HE): (b) Ki-67 staining showing 40%, proliferating rate of tumor cells in the periphery; (c) EZH2 staining positively correlates with Ki-67, strongest staining at the periphery, negative staining in the center (bottom of panel); (d) YAP1 staining negative in tumor cells, but positive stromal sub-populations. (e-h) NEC immunohistochemistry for sample P2 (primary tumor), shown at 15x magnification: (e) tumor cells in pseudo glandular structures (HE); (f) homogenous Ki-67 proliferation rate of 80%; (g) homogenous weak positivity for EZH2; (h) YAP1 staining shows heterogeneity of individual glands with or without nuclear and cytoplasmic staining. Endothelial cells (lower right) are negative (control). (i-n) Feature plots showing EZH2, MKI67, and YAP1 single-nuclei normalized expression in the patient sample P1 (i-k) and P2 (l-n). Acknowledgement: Dr. Aurel Perren performed IHC staining that recapitulated my observations from snRNA-seq data.

5.6 panNEC sub-states differentiation recapitulate human small intestinal signatures

Subsequently, an attempt was made to understand if any of the panNEC sub-states are enriched for human small intestine-specific cell type/state signatures, such as the enteroendocrine (EEC), Tuft, Paneth, or Goblet cells [695]. The initial inspiration for examining these signatures originated from observations in PDAC and injury models, which indicate trajectories starting from acinar to ductal metaplasia and subsequently progressing toward endocrine or intestinal de-differentiation, featuring characteristics of intestinal secretory cell types [615]. This analysis also facilitated the investigation of whether intrinsic tumor plasticity permits the adoption of programs from non-pancreatic intestinal cell types, even though these programs still pertain to the secretory lineage. This adaptability enables tumor cells to exhibit a spectrum of differentiation features.

In this context, the expression of Intestinal Stem Cell (ISC) signatures was predominantly observed in the amphicrine sub-states **(Figure 5.19a).** The amphicrine progenitor-like sub-state displayed relatively strong expression of intestinal Paneth signatures, driven by genes

associated with focal adhesion **(Figure 5.19a; Supplementary Table 26).** As anticipated, both proliferating sub-states specifically expressed signatures characteristic of transient amplifying (TA01) cells, consistent with their high proliferative status. Notably, significant expression of small intestinal Enteroendocrine Cell (EEC) signature patterns was exhibited by the shared NE sub-state.

It is important to highlight that the NE sub-state specific to patient P3 demonstrated robust expression levels of the EEC program, along with *GIPR* and *TPH1*, distinguishing it from other patients **(Figure 5.19b & Extended Figure 13).** Subsequently, a trajectory graph learning was conducted using STREAM [149], revealing additional intrinsic variability within the P3 NE sub-state, as illustrated by its subclusters and lineage branches **(Figure 5.19c & Methods).** Interestingly, NE01 closely resembled ISC and intestinal EEC, while the NE02 subgroup transcriptomically appeared like Tuft, Goblet, and EEC cell types **(Figure 5.19d).** To gain a deeper insight into the molecular mechanisms underlying P3 NE02, an iRegulon analysis was conducted to deconstruct the transcription factor (TF) regulatory landscape of the B0B1 and B0B2 branches. To be precise, the transition genes inferred for both branches were used to predict the TF/regulators.

Notably, the B0B2 branch displayed enrichment for several TF targets, including *SRF, FOXP2, ONECUT2, HOXA7, FOXN4, STAT1,* and *CTBP2* **(Figure 5.19e)**, implicated in cancer progression, playing crucial roles in various cellular processes [696-705]. For example, *SRF* regulates cell migration and proliferation [696], whereas *FOXP2* was shown to enhance stemness and migration in lung epithelial cells upon enforced expression [701]. *ONECUT2* mediates androgen receptor-independent cell growth, driving neuroendocrine differentiation in prostate cancer and associating with worse clinical outcomes and early stage PDAC development [702-704]. Moreover, increased *HOXA7* expression correlates with heightened proliferation in liver cancer, suggesting its potential as a molecular target for diagnosis and treatment [705]. In contrast, B1B0 state revealed enriched *SOX, ELF5* and *CREB1* targets **(Figure 5.19e).** *ELF5* has been implicated in various cancers — importantly, breast cancer, where it acts as an EMT repressor [706]. A pioneering study by Kim et al. 2021 [707] found that effectors of oncogenic *KRAS* activate *CREB1* that further drives PDAC metastasis via upregulating *FOXA1* and WNT/β-catenin signaling. Kim and colleagues also suggested *CREB1* as a potentially therapeutic target for pancreatic cancers [707].

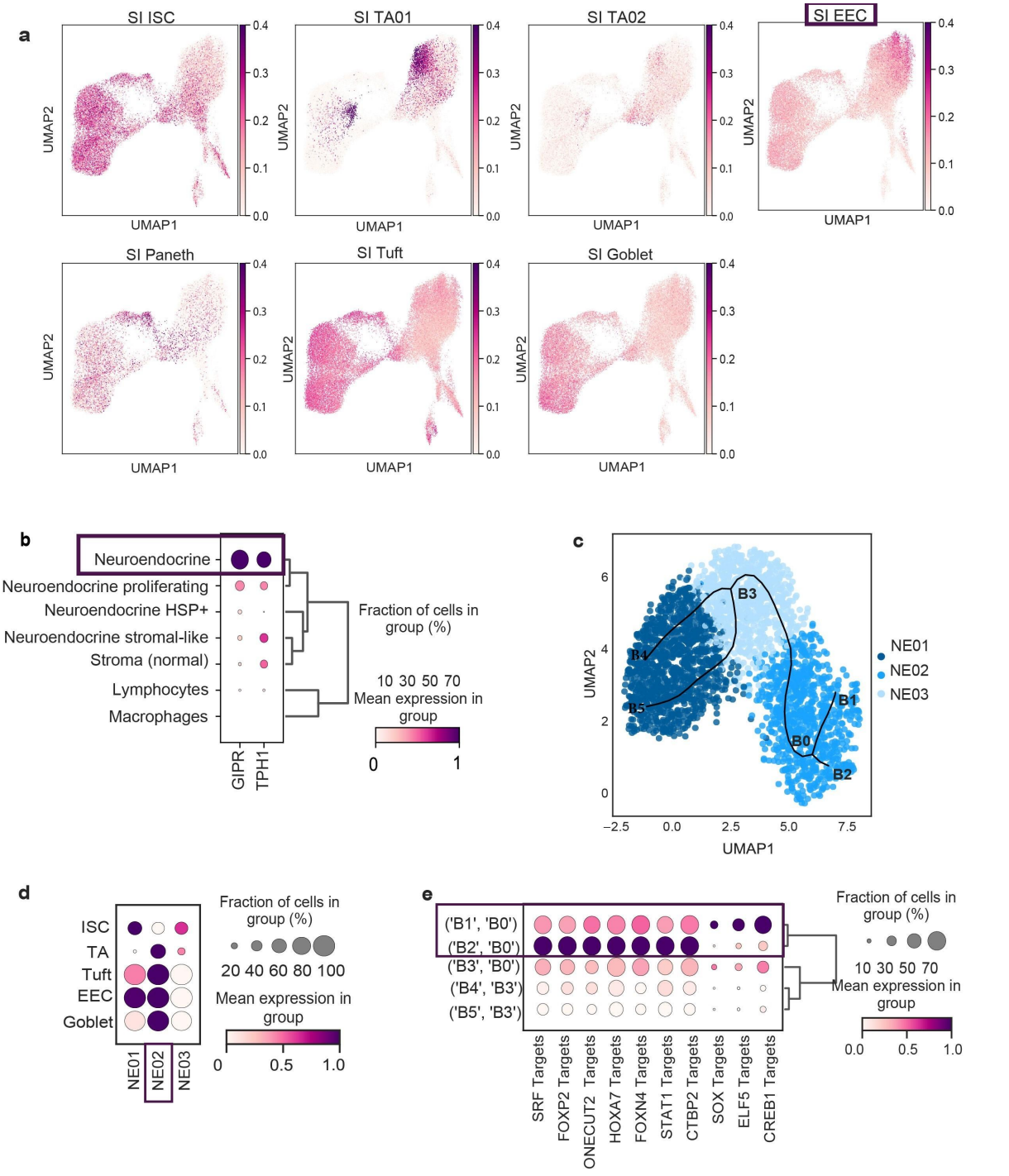


Figure 5.19: Distinct sub-states of panNEC exhibit differentiation reminiscent of human small intestinal signatures.

(a) Feature plots showing the signature score per nucleus for human small intestinal (SI) cell type/state programs, including Intestinal Stem Cell (ISC), transient amplifying (TA01 & TA02), Enteroendocrine cells (EEC), Paneth, Tuft, and Goblet, sourced from Burclaff et al. 2022 [695]. Expression values are min-max scaled for comparison. (b) Dot-plot illustrating GIPR and TPH1 expression across shared NE and immune sub-states for P3. Expression is color-coded, and the percentage of cells (or nuclei) expressing a marker is size-coded. (c) STREAM-based trajectory analysis demonstrating lineage plasticity and class switching among the NE sub-clusters. Branches are marked. (d) Dot-plot displaying selected SI programs (ISC, TA, Tuft, EEC, and Goblet) of biological interest for NE sub-clusters. (e) Dot-plot showcasing TF target signatures enriched for B0B1 and B0B2 branches. TF/regulators were inferred from the positive transition markers for the respective branches using iRegulon (see Methods). No significant enrichment was observed for B0B3, B3B4, and

B3B5 branches. Average expression of a gene (in b & d) or target score (in e) within a sub-state is color-coded, and the percentage of cells (or nuclei) expressing the target signature module is size-coded.

5.7 Therapeutic vulnerability of shared panNEC sub-states

Subsequently, the potential for sub-state directed treatments was addressed by jointly investigating the enriched pathways and drug activities. Differential pathway analysis revealed the enrichment of numerous pathways across the common sub-states, suggesting therapeutic implications. As seen before, the shared NE proliferating sub-state exhibited heightened expression of cell-cycle checkpoint kinases, *E2F* targets, and genes involved in DNA repair mechanisms, which are well targeted by the current first-line treatment with platinum-based drugs, e.g. Cisplatin **(Figure 5.20a)**. Additionally, elevated expression of *EZH2* suggests the potential application of EZH2 inhibitors [708] (such as Tazemetostat). Importantly, gene set enrichment analysis (GSEA) [709] was performed using the LR inferred differentially upregulated genes (average log2FC >= 0.25 & Bonferroni adjusted p-value < 0.01; **Supplementary Table 23**) and using DrugBank [710] and GLAD4U [711] as underlying databases. Here, the rationale was to infer possible drugs that would act on the targets specific to a sub-state. The shared NE proliferating and NE HSP+ (hypoxic) sub-states demonstrated substantial enrichment in signaling pathways and known drugs (both clinically approved and under trials). Conversely, no such enrichment was observed for the shared NE sub-state. Analyzing the NE stromal-like sub-state posed challenges due to the absence of robust and specific differential genes, which complicated statistical analysis and interpretation. As anticipated, GSEA significantly associated NE proliferating upregulated targets with Cisplatin (NES= 2.0493 & FDR= 0.005223) as a prominent pharmacological inhibitor **(Figure 5.20b).** Cisplatin, a widely recognized platinum-based chemotherapeutic agent, induces DNA damage in malignant cells and aligns with current treatment approaches for panNEC [712-715]. The computation of a signature score using Cisplatin targets reflects therapeutically vulnerable cells within this sub-state that could be strategically exploited for combination therapies **(Figure 5.20c).** Additionally, hydroxyurea (NES= 2.1714 & FDR <2.2e-16), and mitomycin (NES= 2.0102 & FDR= 0.0063844) were identified, although they were disregarded as they are not widely recognized as first-line treatments for solid tumors, including pancreatic cancers.

Furthermore, the shared NE HSP+ (hypoxic) sub-state emerged as a pathological hotspot, exhibiting enrichment in several pathways with oncogenic potential. Apart from accelerated induction of heat stress response and hypoxia, this sub-state is marked by glycolysis/PI3K-

AKT (*PFKBP4, ENO1, P4HA1, PGK1*), mTORC pathway (*ACSL3, ACTR3, PSMD14, RAB1A, TCEA1*) and MAPK signaling (*SEPTIN7, FOXO1, MAPK6*) — suggesting shared vulnerable targets across all five patient samples **(Figure 5.20b).** On a clinical note, this could be exploited by repurposing of approved drugs specific to these signaling programs, e.g. Bevacizumab or Alpelisib to specifically target VEGF or PI3K pathway, respectively. In this context, GSEA identified Geldanamycin (NES= 2.1501 & FDR= 0.0043136) **(Figure 5.20d, e)** and Tanespimycin (NES= 2.0704 & FDR= 0.003505) **(Figure 5.20f, g)** as highly significant drugs for targeting NE HSP+ (hypoxic). Furthermore, the significant dysregulation of *HSP90* levels, encoded by *HSP90AA1* and *HSP90AB1*, indicates the potential repurposing of drugs such as Pimitespib (TAS-116) [716]. Pimitespib, the first HSP90 inhibitor approved in Japan, is indicated for the treatment of gastrointestinal stromal tumors (GIST) that have progressed post-chemotherapy [717, 718]. *HSP90* is extremely well known for its distinctive role in evolution by preserving the activity of mutant proteins and acting as a buffer to mitigate phenotypic variation and drug resistance in cancer [719-721]. Of note, one HSP90 inhibitor, 17-allylamino, 17-demethoxygeldanamycin (17-AAG), an analog of Geldanamycin, has undergone phase II clinical trials in various cancers [719-723] (refer to http://www.clinicaltrials.gov for more information on clinical trials), and demonstrated its ability to enhance the cytotoxic effects of paclitaxel in lung and breast cancer cells both in vitro and in vivo [719, 724]. Additionally, Geldanamycin has been shown to inhibit androgen signaling in neuroendocrine prostate cancer, although it did not induce neuroendocrine transdifferentiation [725]. Previously iRegulon analysis also decoded transcriptional regulatory activities by *HSF1* and *HIF1A* that also underscores therapeutic potential (refer to **Figure 5.15b**). Of note, *HSF1* is a hub transcription factor regulating a multitude of targets in NE HSP+ (hypoxic), including the major heat stress encoding genes.

Consecutively, the therapeutic implication of P1 derived stroma (normal) was investigated. Apart from enriched EMT program and Hippo Signaling pathway, snRNA-seq data identified specific expression of *PDGFRA* gene in this sub-state **(Figure 5.20a)**. Of note, *PDGFRA*, is a marker associated with poor prognosis, and suggests the potential utility of targeted therapy using clinically available inhibitors such as the multi-kinase inhibitor Lenvatinib [726]. In this way, co-targeting of a tumor supportive stromal niche might complement existing therapies, in situations of emerging chemotherapy resistance.

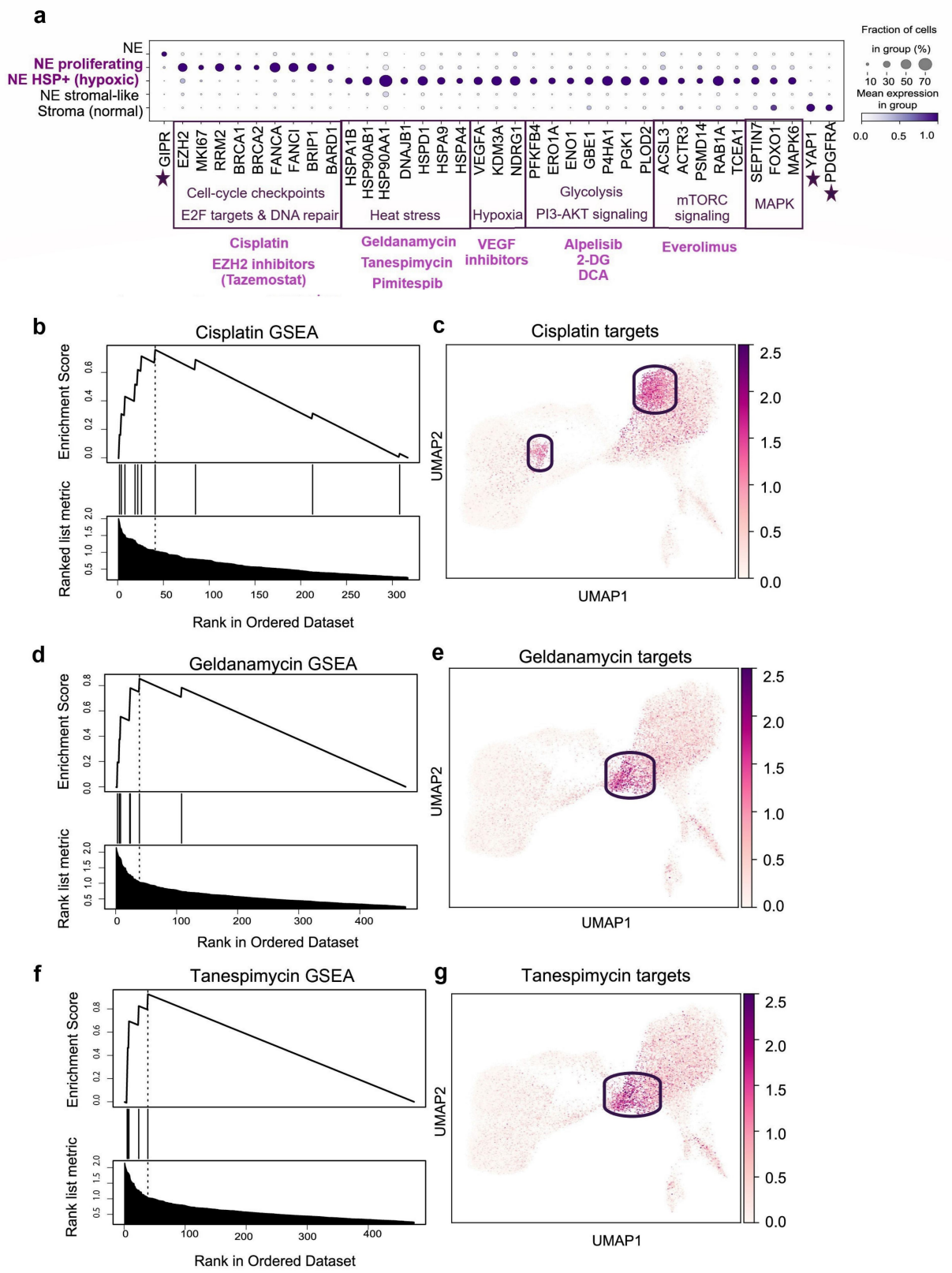


Figure 5.20: Shared NE proliferating and NE HSP+ (hypoxic) sub-states represent pathological hubs with therapeutic significance.

(a) Dot-plot showing differentially upregulated marker expression associated with druggable pathways across shared panNEC sub-states for all patients. Average gene expression within a sub-state is color-coded, and the percentage of cells (or nuclei) expressing a marker is size-coded. GIPR (for

shared NE) and YAP1/PDGFRA (for stromal normal) are included due to their therapeutic relevance in P3 and P1, respectively. Since both of these candidates stem from intra-patient heterogeneity analysis, they are marked with an asterisk (*). Drugs listed underneath are suggested based on enriched targets and pathways and require further experimental validations. Gene Set Enrichment Analysis (GSEA) revealing drug targets: (b) Cisplatin based on enrichment of differentially upregulated genes in NE proliferating, (d) Geldanamycin, and (f) Tanespimycin based on enrichment of differentially upregulated genes in NE HSP+ (hypoxic). Feature plots identifying nuclei enriched in (c) Cisplatin targets in proliferating sub-states, (e) Geldanamycin, and (g) Tanespimycin enriched targets in NE HSP+ (hypoxic).

5.8 Developmental origins of panNEC sub-states

Following that, an investigation was aimed at examining the developmental resemblances between panNEC sub-states and adult pancreatic cell types or states, as well as exploring potential cell of origins. The rationale behind this analysis was to determine whether the diverse panNEC sub-states originate from a single common pancreatic cell type/state acting as a progenitor, or if they have multiple origins. In this regard, adult pancreas snRNA-seq dataset from Tosti et al. [72] was integrated with panNEC dataset presented in this dissertation (see Methods). No additional QC based filtering was performed for the adult pancreas data apart from what was defined in the original publication. A chain integration technique comprising of Harmony [110] and BBKNN [111] was utilized to harmonize two data modalities and mitigating technical sources of variations across samples.

5.8.1 Integration of panNEC and adult pancreas snRNA-seq data

Both exocrine and endocrine cell types or states originating from a common multipotent pancreatic progenitor i.e., all acinar, ductal, and endocrine subtypes from Tosti et al. [72] were incorporated into the data integration process. Given that the combined dataset was predominantly dominated by exocrine cell types, particularly with Acinar-I being the most abundant, the initial integration was possibly biased towards cell types with a higher cell count. Hence, adult pancreatic cell types whose nuclei number exceeded the median of that of the panNEC sub-state were downsampled (see Methods). Consequently, highly variable genes (HVG) were identified separately using sample ID as a batch key in both Tosti and panNEC datasets, and the dataset was concatenated based on outer joining of HVG (n=5351). An attempt was made to merge the datasets by inner joining of HVG (n= 406); however, this approach failed to capture any inherent biological patterns or distinctions among cell types/states. Given that none of the biological truths, such as the separation of acinar, ductal, and endocrine subtypes, were preserved after data integration, the decision

was made not to proceed with inner joining and to adhere to the outer joined dataset. Thereafter, a chain integration step using Harmony and BBKNN was performed (see Methods). The adult pancreas cell types, particularly the ductal and endocrine subtypes, were predominantly distinct from the panNEC sub-states, whereas the acinar cells exhibited proximity to the shared NE cell type **(Extended Data Figure 15).** A discernible segregation existed within the shared NE sub-state, with one cluster being proximal with NE stromal-like cells and another cluster associated with P5 amphicrine sub-states **(Extended Data Figure 15)**. This pattern persisted even after adjusting the number of principal components and internal parameters of BBKNN and UMAP (refer to Methods). On this note, a Pearson correlation analysis was conducted using the highly variable genes, revealing that acinar subtypes (primarily acinar-i) were closely clustered with NE stromal-like cells **(Extended Data Figure 16)**. The shared NE and NE proliferating sub-states were positioned between endocrine and ductal subtypes in the hierarchical clustering **(Extended Data Figure 16)**. While the shared NE sub-state exhibited moderately high positive correlation scores with acinar and ductal subtypes, it did not particularly resemble any of the adult pancreatic clusters.

While the transcriptomic resemblance between acinar subtypes and shared panNEC sub-states (particularly, NE stromal-like) was evident, a re-integration was considered solely with shared NE sub-states for better comprehension of the developmental mechanisms involved. The fact that none of the P5 sub-states closely clustered with Tosti's acinar subtypes prompted an investigation into the HVG(s) and principal component(s) chosen for integration, questioning whether the data had been over-corrected. It was hypothesized that a re-computation of HVG(s) followed by principal component(s) within shared sub-states (comprising P1-P4 data; excluding P5) would lead to a more faithful representation of developmental trajectory in data integration. Therefore, two separate integrations were conducted as two case studies: i) P1-P4 data, and ii) P5.

5.8.2 Integration of shared panNEC sub-states and adult pancreas cell types

Data integration of P1-P4 panNEC patient samples with Tosti et al. snRNA-seq data demonstrated a close association of acinar subtypes with the shared panNEC sub-states — specifically, shared NE stromal-like **(Figure 5.21).** Nevertheless, the ductal and endocrine subtypes remained distinctly separated in the integrated data, indicating that panNEC sub-states exhibited closer transcriptomic similarities to acinar rather than ductal or endocrine cell types/states. This observation was confirmed through Diffusion Pseudotime analysis (DPT) [150], which arranged cells (or nuclei) in an unbiased manner according to their overall transcriptomic profile similarities, without being biased towards any specific cell type or state.

Of note, DPT also revealed a close connection of acinar subtypes with panNEC shared sub-states.

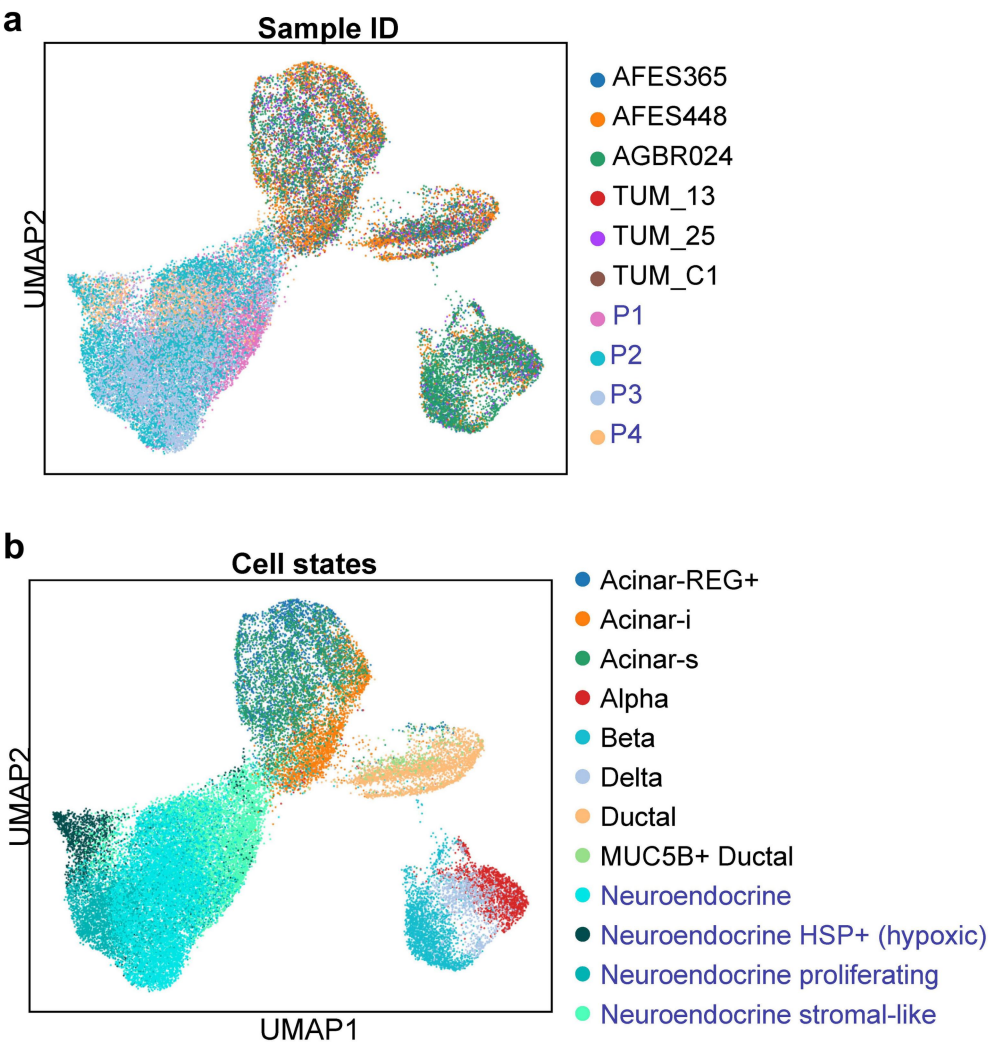


Figure 5.21: Data integration of panNEC and adult pancreas snRNA-seq datasets.
UMAP visualizing integrated datasets encompassing panNEC and adult pancreatic cell types/states from Tosti et al. 2020 [72]. Each nucleus represents either (a) Sample ID or (b) cell states (or, sub-states). Integration was performed using Harmony [110] and BBKNN [111] algorithms in a sequential manner (see Methods). panNEC patient samples and shared sub-states were marked in purple in (a) and (b) respectively, for convenience in visualization.

Following that, a trajectory analysis was performed using STREAM [149], based on the elastic principal graph learning approach [153]. Significantly, STREAM identified a branch linking acinar-I to the shared panNEC sub-states **(S0S3 branch; Figure 5.22a)**. However, ductal, and endocrine cell types/states were located on a separate branch and did not show proximity to panNEC sub-states. Of note, the fact that acinar could be a potential origin of panNEC was previously shown by Yachida et al. [338] based on genomic and epigenomic features of the carcinoma. The "acinar-type panNEC" described by Yachida and colleagues

was found to exhibit perturbed WNT signaling (like mutations in *APC* and *CTNNB1*), as well as altered *CDKN2A*. Moreover, this group exhibited elevated expressions of TF(s) such as *PTF1A, GATA4, NR5A2*, and *RBPJL*. However, these pivotal markers were not detected in the shared sub-states encompassing P1-P4, being limited to the P5 sample. To interrogate possible drivers of the acinar to NE differentiation, transition markers were computed for S0S3 branch and found upregulated elevated expression of genes such as *AUTS2, CACNA1A, RBFOX1* and *ZBTB20* **(Figure 5.22b & Methods).** However, this was not sufficient to describe molecular factors within adult acinar subtypes that could potentially initiate this transition towards NE sub-states. In this context, it can be argued that adult pancreatic snRNA-seq datasets, such as those from Tosti et al., lacked essential cell types or states, such as multipotent pancreatic progenitors (MPPs), tip and trunk cells, and endocrine progenitors — either of which could act as potential starting point ('root node') of the trajectory. Additionally, intermediate states crucial for modelling the progression to panNEC sub-states were not adequately represented. These transitional states were more comprehensively captured in developing mouse scRNA-seq datasets, such as those by Bastidas-Ponce et al. 2019 [339].

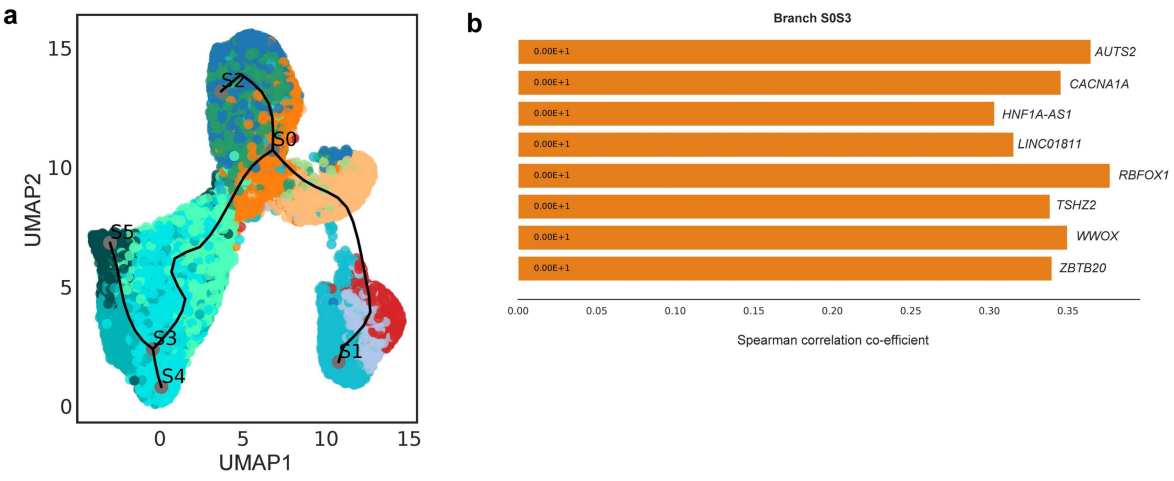


Figure 5.22: Reconstructed lineage trajectory graph showing probable route of acinar to panNEC transdifferentiation

(a) STREAM based trajectory analysis of shared panNEC sub-states integrated with adult pancreas cell types or states revealed an elastic principal graph illustrating possible lineage paths. This includes a lineage path from the acinar subtype (specifically acinar-I at S0) towards panNEC (branch S0S3), as well as a normal differentiation path leading towards ductal/endocrine (branch S0S1). (b) Spearman correlation co-efficient of identified transition markers in branch S0S3 demonstrating (moderately) positive correlation with pseudotime (proxy of developmental time) (see Methods).

5.8.3 Global transcriptomic similarities among panNEC sub-states and adult pancreas cell types

A Pearson correlation analysis was performed using a union of all genes expressed in both datasets to obtain a comprehensive understanding of the global transcriptomic similarities between the panNEC sub-states and adult pancreatic subtypes **(Figure 5.23)**. It was observed that the amphicrine acinar sub-states closely clustered with the acinar-REG+/acinar-s cell states identified by Tosti et al. [72] study. This strong positive correlation was influenced by several genes of interest shared between the two cases, including *GP2, RBPJL, NR5A2*, and *MECOM,* which were extensively discussed in sub-section 5.1.2. It is worth mentioning that Tosti and colleagues associated acinar-REG+ cells with chronic pancreatitis [72], thus, the proximity of amphicrine acinar sub-states to acinar-REG+ further supported the metaplastic nature of the latter sub-states. Interestingly, the amphicrine acinar sub-states did not align with acinar-I, an idle acinar state originally described in Tosti et al., primarily due to the relative immaturity and plasticity of acinar-I, which lacked genes encoding digestive enzymes. Utilizing all expressed genes, as opposed to solely highly variable genes, allowed for recapitulation of the relationship between P5 and acinar subtypes. However, the previously observed high correlation score between NE stromal-like and acinar-I was not retained in this analysis **(Figure 5.23)**. Of note, NE HSP+ (hypoxic) exhibited a moderately high positive correlation score with adult alpha, beta, and delta cell types (refer to **Figure 5.23**). Despite none of the panNEC sub-states displaying expression of key islet markers, this correlation analysis suggested an overall transcriptomic resemblance of the NE HSP+ (hypoxic) sub-state to the adult islets. On the other hand, shared NE, NE proliferating and NE stromal-like sub-states were broadly clustered with ductal subtypes **(Figure 5.23)**. Although they displayed some degree of transcriptomic similarity with acinar-I and ductal subtypes, indicated by low to medium positive correlation scores, these shared sub-states clustered closely together among themselves and did not particularly resemble any native pancreatic cell types/states. **(Figure 5.23).** However, it can be contended that conducting such analysis is more reliable when executed with a carefully chosen sets of key genes and transcription factors driving pancreatic development and differentiation. This approach can help mitigate the influence of non-specific genes that do not have roles in deciding developmental origins.

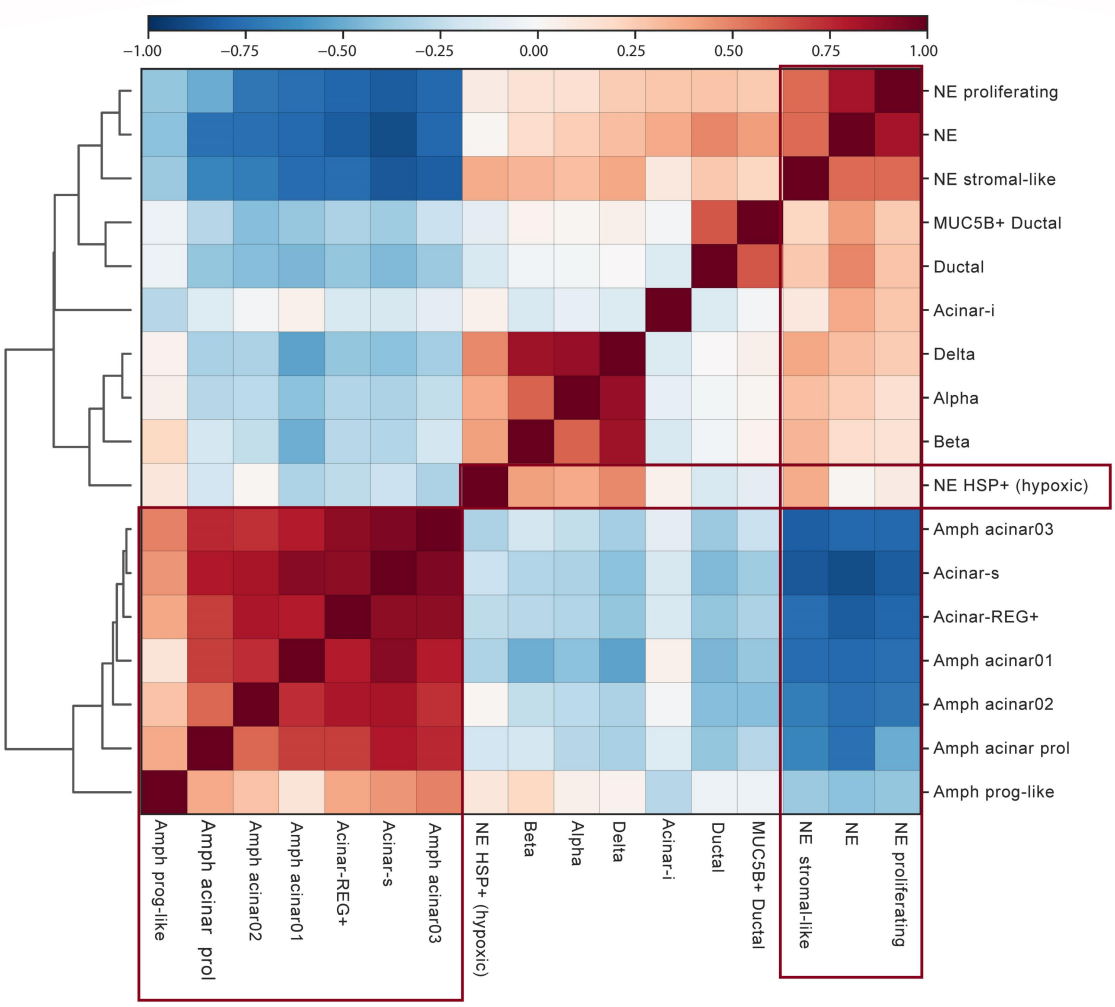


Figure 5.23: Pearson correlation analysis revealing global transcriptomic similarities among panNEC and adult pancreatic cell types or states.
Pearson correlation map depicting the overall transcriptomic similarities between the identified panNEC sub-states and adult pancreatic cell types/states from Tosti et al. [72]. This analysis involved computing correlations using the union of all expressed genes between both datasets and was not restricted to using highly variable gene set. NE= neuroendocrine sub-state; prol= proliferating and prog-like= progenitor-like.

5.8.4 Developmental origins of patient samples

As dedifferentiated NEC might recapitulate aspects of earlier pancreatic development, a selected set of key markers and transcription factors pivotal for regulating pancreatic progenitors (including multipotent pancreatic progenitors), tip, acinar, ductal, and endocrine lineages, sourced from various previous publications [72, 545-549, 592-594, 620] was leveraged. P5 was observed to closely align with the adult acinar cell subtypes **(Figure 5.24),** consistent with its transcriptomic profile of mature exocrine digestive enzymes. Additionally, P3 exhibited close clustering with the endocrine cell types, a relationship reinforced by islet differentiation markers such as *NKX2-2, RFX3, RFX6*, and *SCGN* **(Figure 5.24)**. P2 and P4

closely clustered among themselves and due to *EZH2* and *DNMT1* expression, and revealed limited similarities with endocrine cell types, rather than acinar or ductal subtypes. This was in-line with the observations in Extended Data Figure 18. Lastly, P1 revealed a stronger association with ductal subtypes, given expression of genes like *YAP1* **(Figure 5.24)**.

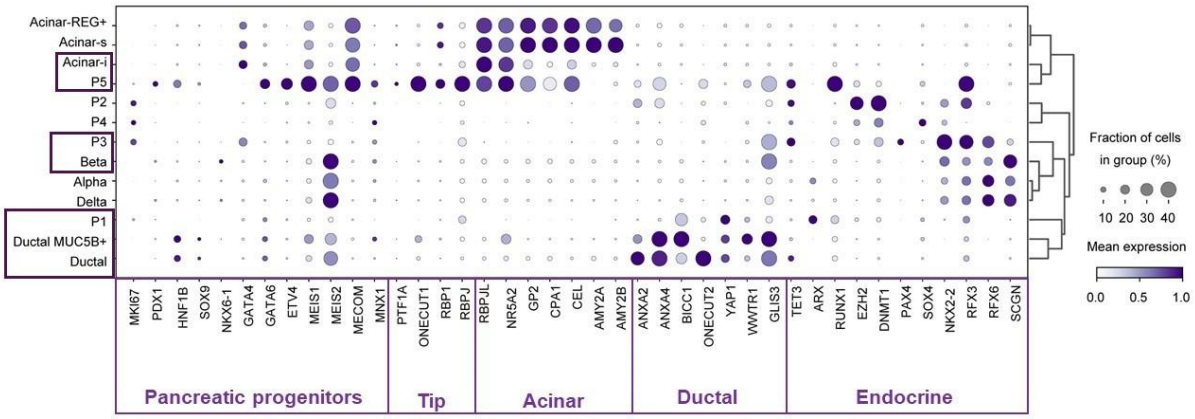
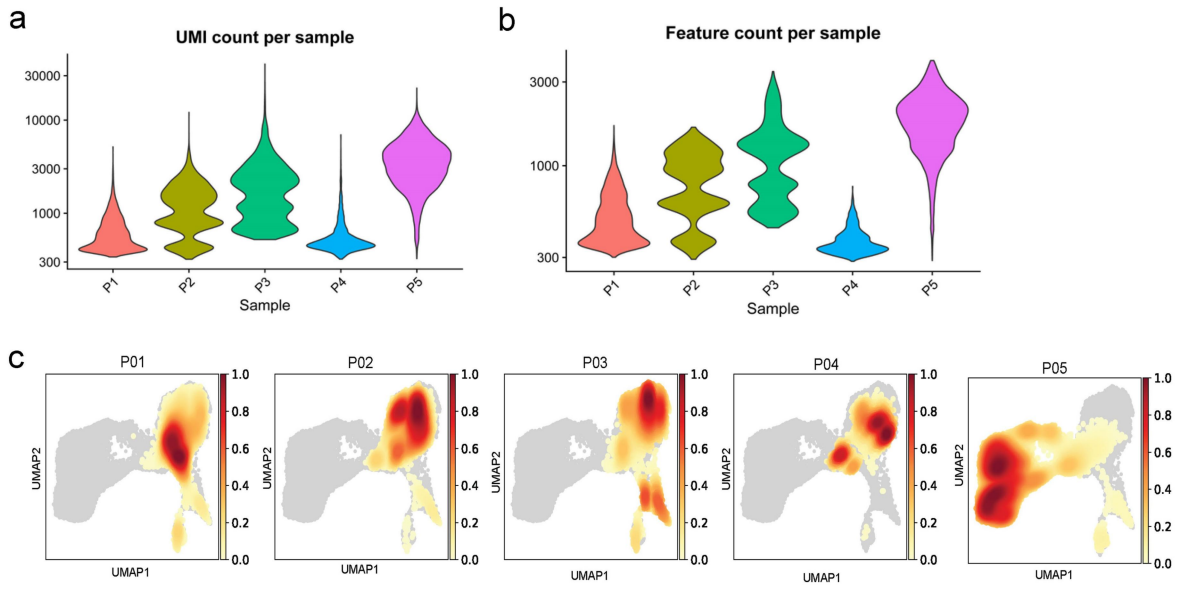
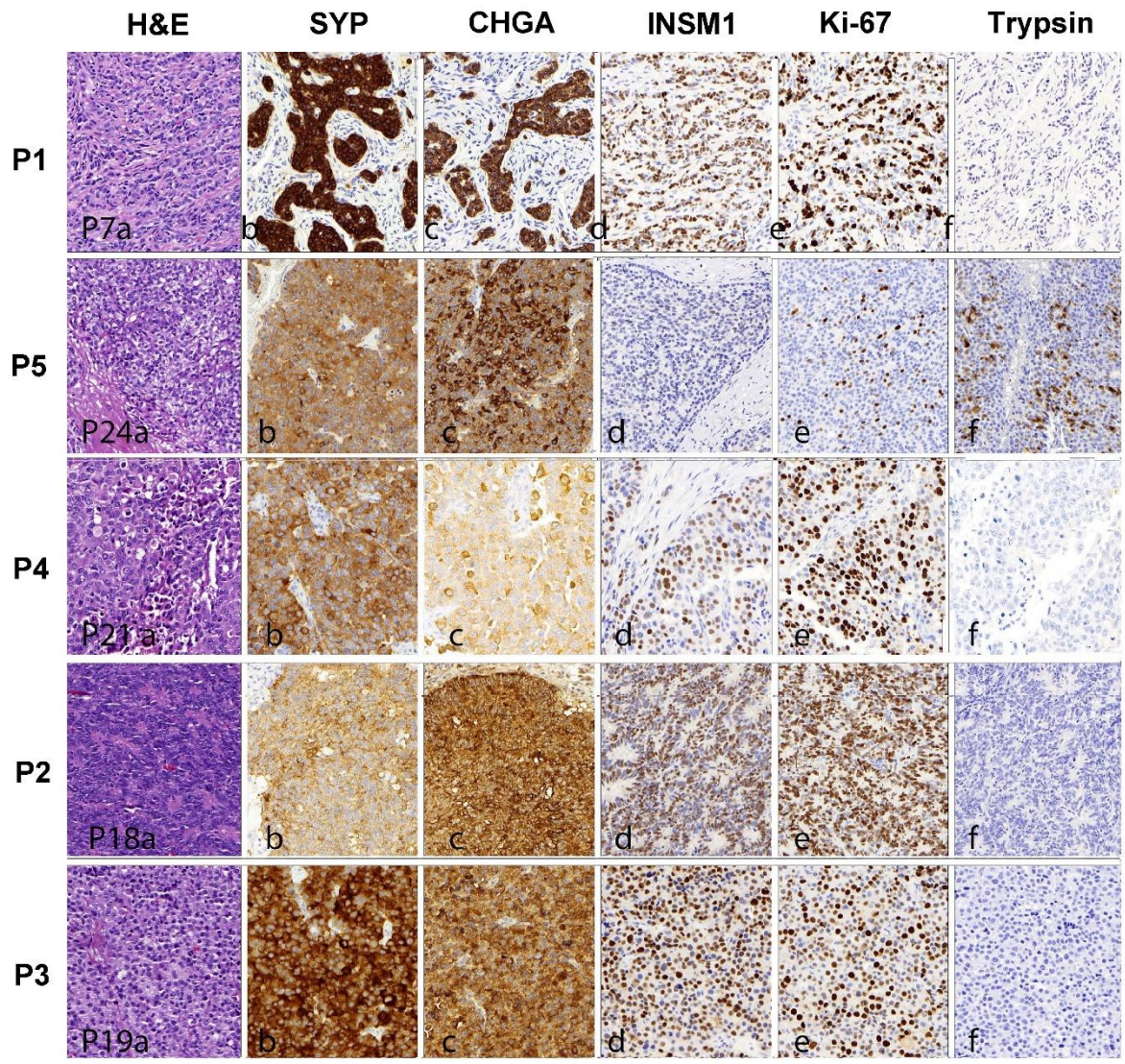


Figure 5.24: | snRNA-seq data suggests distinct developmental origins for panNEC patients.
Dot-plot illustrating markers expression of early pancreas progenitors, tip cells, acinar, ductal, and endocrine cell types. Tip cells are known to differentiate to acinar lineage [547, 590, 591]. Average gene expression within a sub-state is color-coded, and the percentage of cells (or nuclei) expressing a marker is size-coded. Sources of markers plotted were enlisted in Supplementary Table 25.

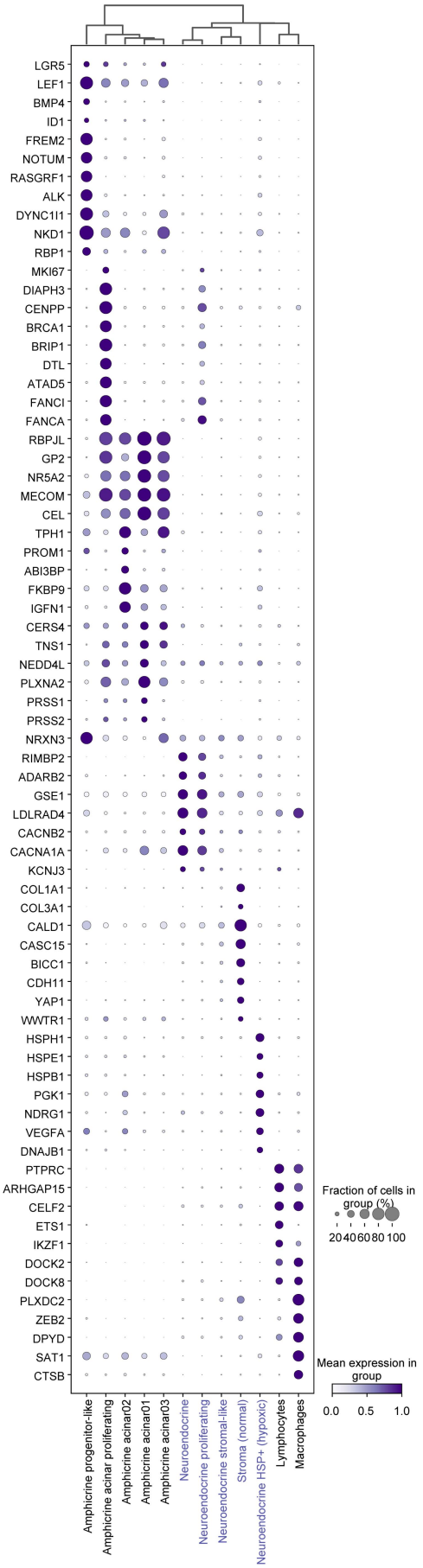
5.9 Extended Data Figures



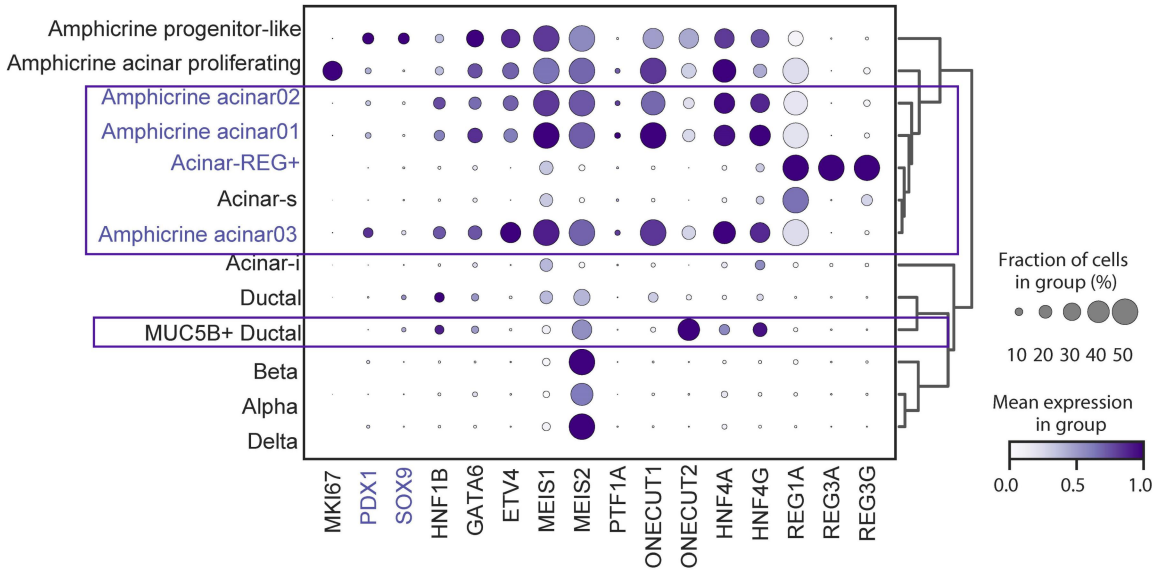
Extended Data Figure 9: Quality control analysis of panNEC samples
Illustrative violin plots depicting (a) total UMI counts and (b) the count of features (i.e., expressed genes) per panNEC sample. (c) Gaussian Kernel density estimation (KDE) analysis showcasing the distribution of nuclei density in UMAP embedding across panNEC samples. Density values are scaled between 0 to 1 for comparative analysis within each category.



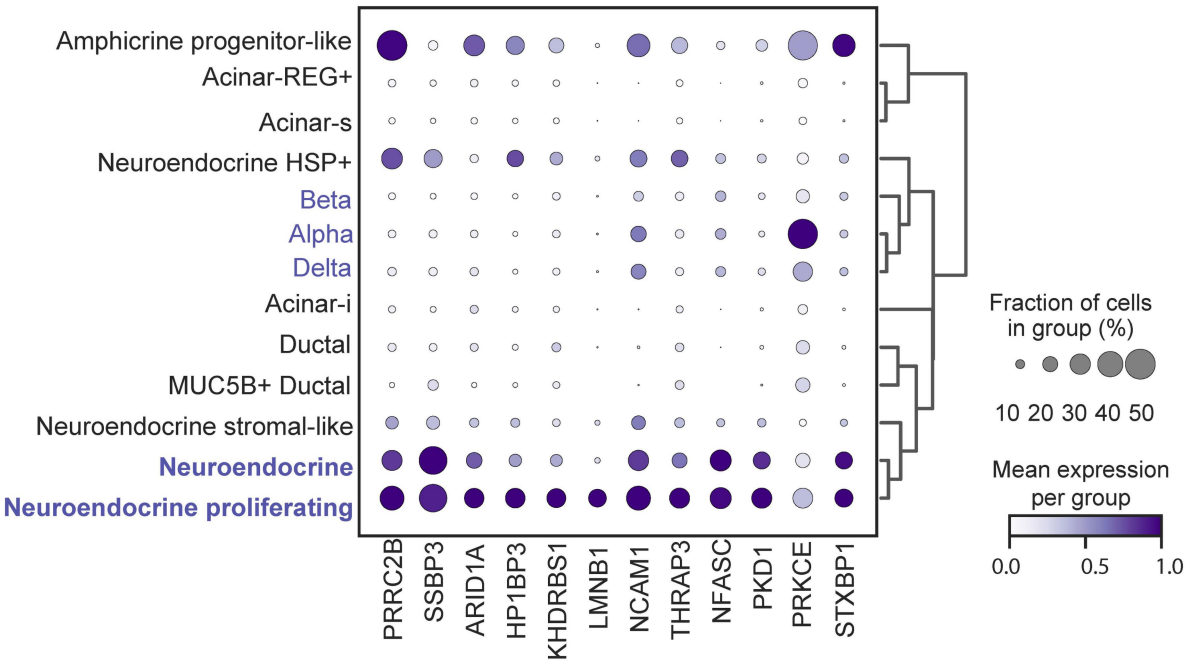
Extended Data Figure 10: Immunohistochemical confirmation of poorly differentiated large cell morphology and neuroendocrine differentiation of panNEC samples.
Representative (a) Hematoxylin and eosin (HE) staining, (b) neuroendocrine diagnostic markers Synaptophysin (SYN) (c) Chromogranin (CHGA) (d) Insulinoma-associated 1 (INSM1) (e) Ki-67 and (f) trypsin immunolabeling in the original tumor tissues. >20% proliferative fraction as determined by Ki-67 immunohistochemical detection determination of proliferating cells. Uniform NE marker expression and detection of trypsin reactivity in a subpopulation of tumor cells in sample P5 confirm the amphicrine differentiation and highlight tumor cell heterogeneity in this sample. P7a, P24a, P21a, P18a and P19a are internal clinical ID(s) and map to P1, P5, P4, P2 and P3 respectively. Only clinical ID P24a or sample ID P5 was tested trypsin positive that confirm "acinar-like" nature of P5. Original magnification was 30x. Acknowledgement: Prof. Dr. Aurel Perren performed the IHC at Institute of Tissue Medicine and Pathology, University of Bern.


LGR5
LEF1
BMP4
ID1
FREM2
NOTUM
RASGRF1
ALK
DYNC1I1
NKD1
RBP1
MKI67
DIAPH3
CENPP
BRCA1
BRIP1
DTL
ATAD5
FANCI
FANCA
RBPJL
GP2
NR5A2
MECOM
CEL
TPH1
PROM1
ABI3BP
FKBP9
IGFN1
CERS4
TNS1
NEDD4L
PLXNA2
PRSS1
PRSS2
NRXN3
RIMBP2
ADARB2
GSE1
LDLRAD4
CACNB2
CACNA1A
KCNJ3
COL1A1
COL3A1
CALD1
CASC15
BICC1
CDH11
YAP1
WWTR1
HSPH1
HSPE1
HSPB1
PGK1
NDRG1
VEGFA
DNAJB1
PTPRC
ARHGAP15
CELF2
ETS1
IKZF1
DOCK2
DOCK8
PLXDC2
ZEB2
DPYD
SAT1
CTSB
Amphicrine progenitor-like
Amphicrine acinar proliferating
Amphicrine acinar02
Amphicrine acinar01
Amphicrine acinar03
Neuroendocrine
Neuroendocrine proliferating
Neuroendocrine stromal-like
Stroma (normal)
Neuroendocrine HSP+ (hypoxic)
Lymphocytes
Macrophages
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
0.0 0.5 1.0

Extended Data Figure 11: panNEC sub-states characterized by robust and specific markers.
Dot-plot illustrating key markers across unique and shared panNEC sub-states inferred using multivariate Logistic Regression (average log2FC >= 0.25 & Bonferroni adjusted p-value < 0.01; Supplementary Table 23). Average gene expression within each cluster is color-coded, while the percentage of nuclei expressing a marker is indicated by marker size. Gene expression levels are standardized across columns for comparison. Hierarchical clustering was computed using all genes expressed in the concatenated dataset to reflect global transcriptomic similarities.

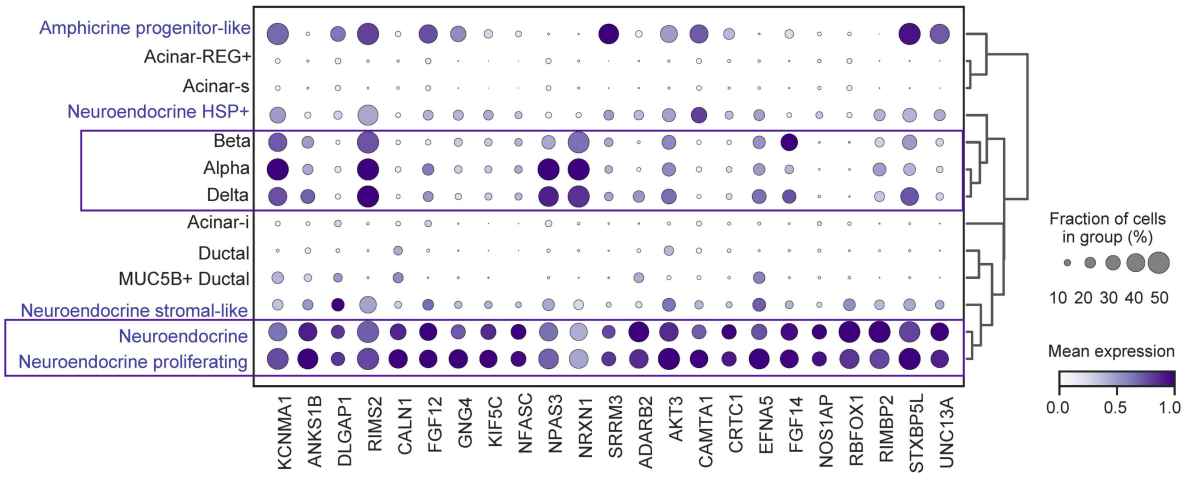


Extended Data Figure 12: Early pancreatic progenitor and metaplastic markers in panNEC and adult pancreatic cell types or states.
Dot-plot depicting prominent markers and transcription factors known to regulate early pancreas development and control differentiation of pancreatic progenitors (including multipotent pancreatic progenitors, tip, and trunk progenitors) as outlined in main figure 4.8. PDX1 and SOX9 are prime regulators of multipotent pancreatic progenitors [545-549, 592-594] and marked for their significance in early pancreas development. ONECUT2 is indicative of late metaplastic stage, as reported by Schlesinger et al. 2020 [604]. GATA6 and HNF4A are implicated in PDAC classical subtype. The presence of MKI67 in the amphicrine acinar proliferating cluster suggests a self-renewing sub-state. NEUROG3 and FEV were not detected in the dataset, indicating the absence of classical pancreatic endocrine progenitors. Average gene expression within each cluster is color-coded, while the percentage of nuclei expressing a marker is indicated by marker size. Gene expression levels are standardized across columns for comparison. Hierarchical clustering was computed using all genes expressed in the concatenated dataset to reflect global transcriptomic similarities.



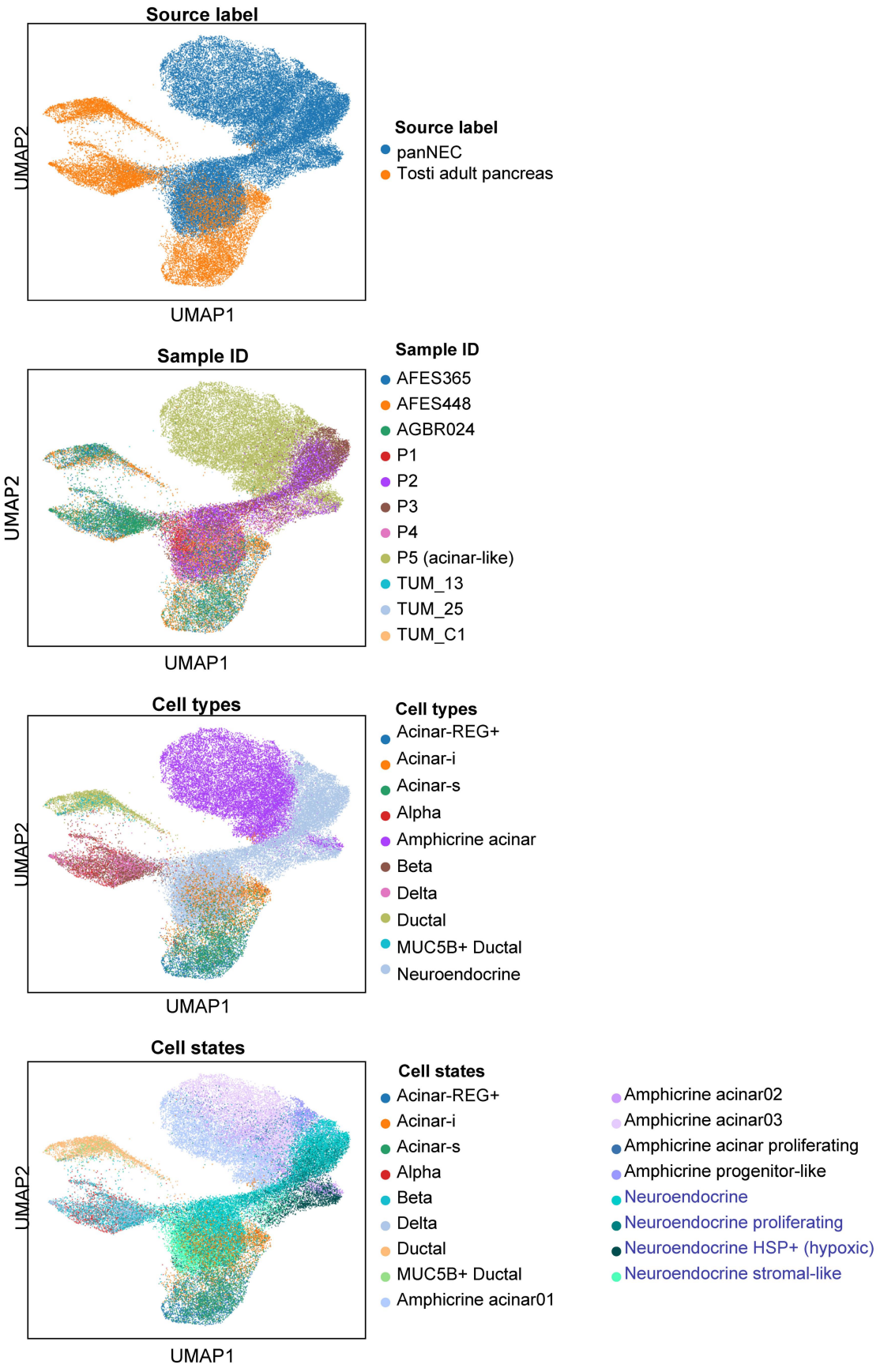
Extended Data Figure 13: PTF1A regulated brain targets mostly specific to shared NE and NE proliferating sub-states.

Dot-plot illustrating the prominent PTF1A and PAX6 regulated brain targets across panNEC sub-states and adult pancreas cell types/states integrated from Tosti et al. 2020 [72]. Adult islets (alpha, beta, and delta) exhibit negligible expression of PTF1A brain targets, while PAX6 brain targets (NFASC to STXBP1) are expressed at low levels in the islets. P5 amphicrine acinar sub-states were excluded from the plot. Average gene expression within each cluster is color-coded, and the percentage of nuclei expressing a marker is represented by marker size. Gene expression levels are standardized across columns for comparison. Hierarchical clustering was performed using all genes expressed in the concatenated dataset to illustrate global transcriptomic similarities.



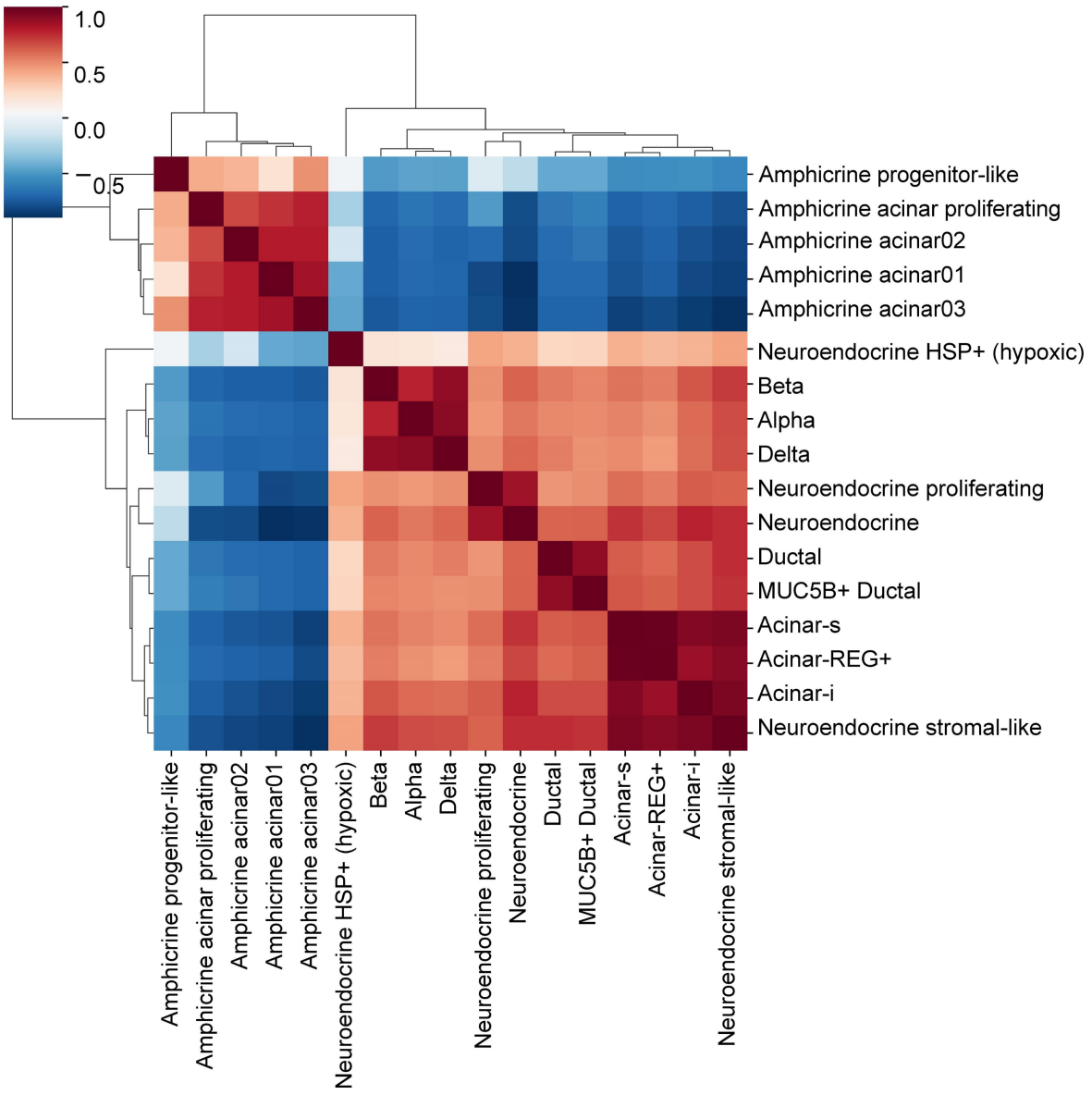
Extended Data Figure 14: EZH2 and REST targets robustly expressed in shared NE and NE proliferating sub-states.

Dot-plot demonstrating the expression of selected markers jointly regulated by EZH2 and REST across panNEC sub-states and adult pancreas cell types/states integrated from Tosti et al. 2020 [72]. EZH2 brain targets were inferred through analysis of the GTEX dataset (see Methods), followed by iRegulon prediction of REST as a central regulator of the EZH2 brain targets. P5 amphicrine acinar sub-states were excluded from the plot. Average gene expression within each cluster is color-coded, while the percentage of nuclei expressing a marker is indicated by marker size. Gene expression levels are standardized across columns for comparison. Hierarchical clustering was conducted using all genes expressed in the concatenated dataset to illustrate global transcriptomic similarities.



Extended Data Figure 15: Data integration of panNEC and adult pancreas snRNA-seq datasets.

UMAP visualizing harmonized snRNA-seq datasets encompassing panNEC (n=5) and adult pancreatic cell types/states from Tosti et al. 2020 [72]. Each nucleus denotes (a) Source label (i.e., Tosti or panNEC), (b) Sample ID, (c) Cell types and (b) Cell states (where panNEC sub-states were visualized). Chain integration was performed using Harmony [110] and BBKNN [111] algorithms in a sequential manner (see Methods).



Extended Data Figure 16: Pearson correlation analysis revealing transcriptomic similarities among panNEC and adult pancreatic cell types or states.
Pearson correlation map depicting transcriptomic similarities between the identified panNEC sub-states and adult pancreatic cell types/states based on highly variable gene expression (see Methods).

6 Discussion

6.1 Generating a comparative atlas of maternal-fetal interface in healthy and eoPE pregnancies

Understanding the complex pathological mechanisms of eoPE is challenging due to its multifactorial origins, constraints in collecting longitudinal samples during pregnancy, and the necessity for obtaining gestational age-matched healthy controls. This pilot study aimed to overcome these challenges by utilizing snRNA-seq analytics, enabling the longitudinal exploration of disease mechanisms. This involved comparing placenta and decidua samples collected during the first trimester (early), term (late) control, as well as those affected by eoPE. snRNA-seq profiling was performed instead of scRNA-seq to ensure appropriate capturing of heterogeneity underlying multinucleated syncytiotrophoblast cells, and to enable a more comprehensive exploration of the maternal-fetal interface. To maintain a uniform study population, only cases of eoPE associated with fetal growth restriction (FGR) [272, 343, 344] were investigated. It is to be noted that this choice acknowledges the incomplete representation of the syndrome, which typically involves maternal organ dysfunctions associated with PE. Importantly, analysis of placental microarray data did not reveal significant differences between eoPE with and without FGR, indicating FGR did not introduce major confounding factors in downstream analysis.

The study utilized separate sampling of placental villi and decidua, enabling the investigation of tissue-specific cellular landscape of the maternal-fetal interface. To overcome challenges in harmonizing data, the study leveraged the deep learning framework scVI/scANVI [112, 113], emphasizing its effectiveness in integrating diverse datasets from different sources and time points during gestation. This variational inference model treated individual donors as the primary batch and included additional covariates to account for batch effects associated with sampling sites, as well as biological and technical factors. Through rigorous analysis under various scenarios, the integration outcomes were compared with those of other studies, demonstrating successful regression of the majority of batch effects and ensuring reliable harmonization of the samples presented in the study. Ultimately, this approach followed by clustering and cell marker analysis successfully deconstructed cellular hierarchies, revealing various cell types or states within immune, vascular-endothelial, matrisome, and trophoblast components across both tissues. Nevertheless, the increased proportions of PAMM, vT-cells, vFB, and vMC in samples from Graz compared to those from Oslo in the term control condition indicated differences in cellular composition among sampling sites. Similarly, the slight reduction of PAMM and vT-cells in Graz samples compared to Oslo samples within the

eoPE group, coupled with the observed reduction in their ASW score, adds complexity to understanding their contributions to the pathogenesis of eoPE. Therefore, these observations necessitate further exploration to clarify the significance of these cellular alterations in the context of eoPE.

The multivariate Logistic regression-based GLM tests utilized in this study provided insight into characteristic markers of decidual and placental cell types/states. Within the decidua, distinct populations were identified, including progenitor-like decidual mesenchymal stem cells (dMSC) and two sub-states of decidua stromal cells (DSC1 and DSC2), each expressing unique marker genes. Furthermore, the characterization of decidual fibroblasts (dFB1 and dFB2) revealed their roles in ECM remodeling and structural support provision [364-367], with dFB2 exhibiting specificity to both eoPE and term conditions. Additionally, the study delineated the maternal immune landscape, uncovering sub-states of decidual natural killer cells (dNK1, dNK2, dNKp), decidual T cells (dTcells), decidual macrophages (dMAC1 and dMAC2), and decidual monocytes (dMono1 and dMono2), each exhibiting specific marker expression indicative of their functional roles in immune regulation, maternal-fetal tolerance establishment, and tissue remodeling to support fetal development [377, 380-382, 390-393]. Moreover, the identification of decidua-derived dDSTB nuclei expressing traditional STB markers prompts questions regarding their presence, whether attributed to sampling techniques or reflecting physiological shedding of STB nuclei into the maternal circulation, as suggested by IHC staining of GDF15 and β-hCG in decidual sections, while lacking the EVT lineage marker HLA-G [402, 403]. Together, these findings reveal intricate cellular dynamics within the decidua and raise intriguing inquiries regarding their implications in maternal-fetal physiology during eoPE.

Distinct trophoblast cell types/states, including villous cytotrophoblasts (vCTB) expressing markers such as *YAP1, LGR5, TP63,* and *PARP1,* were identified, along with a proliferative sub-state, vCTBp, representing trophoblast progenitors [146, 209, 394-400]. A pre-fusion state, vCTBpf, expressing fusiogenic markers, resembled the transcriptomic profile of *ERVFRD-1*+ trophoblasts, previously described by Liu et al. 2018 [401]. The study delineated two major trophoblast lineages- villous syncytiotrophoblasts (vSTB) expressing hormone secretion markers [231-234] and cell-column trophoblasts (vCCT) expressing markers for immune modulation and maternal-spiral artery remodeling [402-404, 432]. To date, there remains no standardized consensus nomenclature for the cell column, and this annotation was derived from the widely recognized book "Benirschke's Pathology of the Human Placenta" [236]. Despite that, it was recognized that the proximal and distal cell columns located in vCCT serve different functions, and the distal part potentially differentiates into

decidua-invading extravillous trophoblasts (dEVT) during placental development. This prompted to conduct subclustering analysis that detected notable heterogeneity within the vCCT group, revealing distinct subtypes: proximal (pCCT), distal (dCCT/pEVT), and an intermediate transitional state. The pCCT subtype exhibited an EGFR signature associated with migration and invasion. The transition from proliferative pCCT to non-dividing dCCT marked an increase in EVT-specific markers (*HLA-G, NOTUM, HPGD*), emphasizing their role in immune tolerance, Wnt signaling, and inflammation regulation during EVT invasion [402, 403, 422-426]. Additionally, dCCT exhibited elevated expression of extracellular matrix remodeling genes, like integrins (*ITGA1, ITGA5*), *MMP2*, Hippo signaling pathway members (*TEAD1, WWTR1*) that are essential for EVT differentiation, invasion, and establishment of uteroplacental blood circulation [420, 421, 427-429, 430]. Conversely, placental resident macrophages (vHBC) with pro-angiogenic functions and a proliferative sub-state (vHBCp) were identified, along with placenta-associated maternal monocytes/macrophages (vPAMM) [237-240]. Importantly, a comparison between analogous cell types, specifically dMAC and fetal vHBC, was conducted using differential gene analysis, examination of transcriptomic regulation, and analysis of biological signaling pathways.

Statistically significant shifts in cell-type compositional abundance were observed in early samples compared to term controls for all vCTBp, vCTB, vCTBpf, and vCCT. While no composition shift was observed for vHBC between early and term, vHBCp was uniquely identified in the early placenta, indicating its physiological importance in the first trimester. This includes involvement in tissue remodeling processes and the secretion of growth factors and cytokines essential for fetal growth and development. On this note, the recapitulation of vHBCp aligns with a previous scRNA-seq study by Thomas et. al. 2020 [238] suggesting that vHBCs are produced via primitive hematopoiesis and follow a distinct trajectory path through the cell cycle, without requiring a monocyte intermediate. The prevalence of vSTB populations was significantly higher in late villi compared to early villi, with seemingly greater abundance observed in tissues affected by eoPE. Nevertheless, the statistical analysis of composition shifts between eoPE and term controls was complicated by the small sample size. In contrast, vCTBp, vCTBpf, and vCCT were extremely sparse in both eoPE and term control groups for which no eoPE-based analysis was performed for them.

EVT invasion into the decidua is a requirement for sustaining healthy pregnancies, and it has been previously theorized that inadequate spiral artery remodeling by EVT contributes to PE [270, 730]. However, this study faced limitations as only a few dEVT(s) were obtained from the decidua samples despite a specialized sampling procedure involving vacuum suctioning, and both dEVTs and vCCTs were too sparse for comparative analyses. This aligned with

prior research, which employed pre-enrichment of HLA-G+ cells to increase the EVT/CCT yield to compensate for the low EVT abundance [209]. Additionally, a high pulsatility index of the uterine artery preceding spiral arteries in the decidua has been associated with defective EVT remodeling. While the predictive sensitivity of this index can range between 12 to 61% for early gestational PE cases, it demonstrates higher sensitivity for FGR [731, 732] but less predictive potential for term PE [731, 733]. These prior findings suggested that the development of PE is likely linked not only to impaired uterine artery flow but also to multifactorial causes [734, 735].

The study faced a computational challenge in addressing gestational age variabilities, particularly in preterm eoPE samples collected around the 34th week compared to term controls obtained between the 38th and 40th week. To compensate for this lack of gestational age-matched controls, harmonization of an external scRNA-seq dataset [353] comparing non-pathological, non-hypertensive preterm and term pregnancies was conducted. The integration demonstrated effective batch mixing, especially for conserved cell types like vCTB, with satisfactory integration metrics, such as adjusted rand index (ARI) and adjusted mutual information (AMI), supporting the reliability of the combined placental and decidua datasets. Subsequently, a preterm signature score specific to each cell type/state was calculated using differentially expressed genes associated with preterm births relative to term controls. This score was then modeled as a covariate to account for preterm effects in the subsequent analysis comparing eoPE versus term samples. Preterm effects could not be adjusted for during the analysis of eoPE markers in vFB, vMC, and vVEC clusters, as these clusters had too few cells in the scRNA-seq preterm data, resulting in the inability to identify a reliable preterm gene set for confounder modeling. Hence, no significant conclusions were drawn for these cell types during downstream analysis.

Since the aforementioned scRNA-seq data did not profile sufficient numbers of vSTB, a bulk microarray dataset [345] comparing preterm and term controls was leveraged to account for vSTB-specific preterm signatures. Given the lack of publicly available single-cell/nucleus data of pre-term STBs, the fact that STB accounts for the majority (~90% of cells) of cells in the villi [355-357] was rationalized to justify the use of bulk gene expression data to represent the transcriptome of STB cells. Subsequently, differentially expressed genes between pre-term and term controls that were also expressed to varying extents in our STB populations were identified. A vSTB-specific preterm score was then computed using this gene set and used as a confounding variable in the Logistic Regression model when identifying eoPE-specific vSTB genes. While this correction barely affected vSTB1 and vSTB2, larger differences were observed in the vSTBjuv sub-state that lost STB fusion genes post-

correction– this is in part due to *ERVFRD-1* and *ISYNA1* being identified as vSTB pre-term genes.

Despite extensive computational efforts, the absence of gestational age matching between pre-eclamptic and control subjects could still be considered a limitation of this study. This discrepancy arises because women with PE frequently undergo premature delivery as a primary intervention to mitigate risks to both mother and fetus, given that termination of pregnancy is a primary intervention for PE management [272, 736]. Premature delivery, whether by vaginal birth or cesarean section, is not typical in uncomplicated pregnancies but mainly occurs due to underlying pathology, often related to inflammation or infection [737]. However, such factors might be clinically unreported in the scRNA-seq studies investigating preterm controls. Therefore, recruiting age-matched premature control pregnancies for utero-placental tissue samples may be challenging. While adjusting for gestational age in the PE group was statistically feasible for most cell types/states, and was necessary for diligent downstream analysis, it might come with biological shortcomings, as women with PE who deliver prematurely are likely to have more severe disease compared to those who deliver at term. This highlights an unavoidable challenge inherent in eoPE research.

Ethical considerations precluded the collection of longitudinal biopsies of placenta samples. Consequently, the study traced the eoPE-centric dysregulation signatures from the first-trimester trophoblast differentiation trajectory. The hypothesis posited that eoPE has its origins in early development, characterized by disrupted trophoblast differentiation. Consequently, the comparison of differentially expressed disease genes and pseudotime transition markers was employed to trace and identify the dysregulated developmental drivers underlying this condition. Multi-omics analytics, including spatial proteomics and 10X Visium spatial transcriptomics [187], were employed for the validation of developmental drivers inferred from trajectory analysis.

In summary, this comprehensive reference cell atlas will offer valuable insights into the diverse array of cell types and states, along with their molecular features and shifts in cellular composition across first-trimester, healthy term, and eoPE pregnancies. With its detailed information, the atlas is poised to be instrumental in unraveling the molecular intricacies of eoPE through the identification of differentially expressed disease signatures, transcriptional landscapes, and changes in cellular interactions compared to term controls. Such comparative analyses are essential for understanding disease mechanisms, discovering biomarkers, and devising targeted therapeutic strategies.

6.1.1 Functional specialization within syncytiotrophoblasts

snRNA-seq analysis revealed notable transcriptional heterogeneity among villous STB nuclei, leading to the identification of three distinct sub-states, namely vSTB1, vSTB2, and vSTBjuv. This study elaborated on the observation that these represent different states of nuclei within the syncytiotrophoblast, rather than distinct cell types. It was noted that each placenta contains only one syncytiotrophoblast layer, housing nuclei ranging from recently fused with ongoing transcriptional activity (vSTBjuv) to gradually becoming transcriptionally inactive (vSTB1) and senescent (vSTB2) nuclei, which are destined to form syncytial knots and enter the maternal bloodstream. The vSTB1 subset constituted the majority of STB nuclei, while vSTB2 nuclei represented a smaller subset of aged nuclei requiring shedding, akin to a stratified epithelium. All three sub-states expressed genes related to the syncytial secretion of hormones into the maternal circulation required for the maintenance and immunomodulation of pregnancy. However, it is essential to acknowledge the limitations of the snRNA-seq data, which solely captured the transcriptomic profiles of vSTB nuclei and might exhibit reduced complexity compared to in vivo conditions. This limitation includes the potential oversight of vCTB-to-vSTB RNA transfer, a phenomenon recently unveiled in non-trophoblastic cells involving syncytin-mediated transfer through tunneling nanotubes [738].

- **Functional differences between vSTB1 and vSTB2**

The transcriptomic profiles of vSTB1 were indicative of typical syncytiotrophoblast functions, encompassing hormone secretion, estrogen biosynthesis, and protein secretion. Conversely, vSTB2 displayed partial transcriptional inactivation of traditional STB programs but exhibited differential expression of genes associated with senescence. Notably, vSTB1 demonstrated robust expression of *KYNU* [405], a gene pivotal in regulating placental NAD+ synthesis and facilitating the provision of fetal tryptophan and kynurenine metabolites, whereas this feature was relatively subdued in vSTB2. Additionally, vSTB2 exhibited relatively elevated expression of the *PDE4D* gene [406], previously implicated in inducing intra-uterine inflammation in gestational tissues and preterm labor. Downregulation of DNA damage response genes such as *DDX5* and *DDX17*, along with cyclin-dependent kinases like *CDK19*, was observed in vSTB2 compared to vSTB1, suggesting a potential inclination towards senescence in STB2 and implying a later occurrence in differentiation. Importantly, key fusion genes like *LIFR, GCM1*, and *DYSF* were significantly downregulated in vSTB2, suggesting it to be a terminal sub-state of vSTB.

Functional disparities between vSTB1 and vSTB2 were explored, revealing distinct gene enrichment profiles, and signaling pathways associated with steroid hormone metabolism, JAK-STAT signaling, and integrin pathways in vSTB1. Conversely, vSTB2 showed enrichment in estrogen-dependent nuclear events and the ARF6 pathway. Transcription factors such as *TFAP2A, ESR1, SP1*, and *CEBPB* were identified in both vSTB1 and vSTB2, indicating ongoing hormonal activity and regulation of key pathways in these sub-states. Notably, *TFAP2A*, a renowned transcription factor, is implicated in initiating the early specification of trophoblast progenitors and regulating the terminal differentiation of vSTB [410, 411]. Additionally, ESR1 protein presence has been observed in the nuclei of cultured human vSTB cells, and previous research suggests its involvement in estrogen-mediated regulation of leptin biosynthesis through the *LEP* promoter in choriocarcinoma JEG-3 cells, highlighting the potential regulatory role of *ESR1* in leptin biosynthesis [413].

However, thorough experimental validations are required to accurately describe these nucleus states biologically. As this exceeds the scope of my thesis, the two states have been conservatively labeled with numbers rather than specific phenotypes.

- **Discovery of novel vSTBjuv nuclear sub-state**

This study unveiled a novel nuclear sub-state within the vSTB barrier known as juvenile syncytiotrophoblast (vSTBjuv). The designation of this potentially immature nuclear sub-state is based on its position immediately after the fusion event (i.e., after vCTBpf transitional state), as modeled in the trophoblast pseudotime differentiation trajectory. However, the functional distinctions of STBjuv compared to other vSTB states remain elusive. Before delving deeper into the biology, additional QC investigations were conducted, including assessments of doublet scores, the distribution of mitochondrial reads, and conflicting topics using LDA analysis, to rule out potential artifacts in defining this nuclear sub-state. As seen before, the overall distribution of doublet scores did not exhibit bimodality, indicating that this dataset was not systematically affected by potential doublets. This observation aligned with known expectations for single-"cell" isolation, where a substantial portion of doublets arises from incomplete tissue dissociation. In the case of single-"nuclei" analysis, the sources of doublets are either nuclei sticking together, or the chip being overloaded, resulting in multiple nuclei per droplet. Overall, this reduction in potential causes of doublets should result in a decrease in the observed doublet rate.

Differential gene expression analysis revealed specific upregulated genes associated with cytoskeletal stability and extracellular matrix organization in vSTBjuv, hinting at a role in

forming and maintaining the outer placental wall. The abundance of vSTBjuv nuclei showed no significant difference between eoPE and term controls, suggesting a gestation-independent role, although larger cohort studies are needed to confirm this observation. Reduced expression of senescence markers in vSTBjuv was observed, and it was found that less senescent regions in the trophoblast layer occur around fetal vessels in healthy placentas. Hence, it is conceivable that these vSTBjuv nuclei might be positioned in proximity to regions of maternal-fetal syncytium interaction, potentially playing a role in transplacental transport.

6.1.2 Early trophoblast trajectory

In this study, trajectory analysis was performed using STREAM [149], and validated with additional methods like Diffusion Pseudotime [150]. Of note, Diffusion Pseudotime is an unsupervised non-linear algorithm that orders cells based on global transcriptomic similarity in a Markovian space and consistently placed vCTBpf and vSTBjuv as transition states before committing to vSTB path. By considering vCTBp as the root progenitor cell state, two divergent lineages i.e., vSTB and vCCT, were characterized by dynamic transition and leaf markers. Spatial proteomics validated vSTB and vCCT transition markers, ensuring consistency in mRNA and protein levels.

The differentiation branch of vCCT exhibited heightened expression of *HLA-G, NOTUM, TEAD1*, and *FOS*, accompanied by increased AP1 signaling, as corroborated by proteomic data. This lineage demonstrated the presence of transition markers associated with metabolic processes, cell adhesion, and migration, a finding validated at the protein level. AP-1 transcriptional activity, known to stimulate *MMP* production in human trophoblast cells, influences key cellular processes such as proliferation, survival, and migration [420, 421, 430]. FOS, notably detected in EVT cells within proximal cell columns of early human placentae, plays a pivotal role in trophoblast proliferation by regulating cyclin gene transcription and repressing cyclin-dependent kinase inhibitors [501, 739]. Integration with 10X Visium spatial transcriptomics revealed spatially variable developmental drivers, including *HLA-G, HPGD, NOTUM, ASCL2, LAIR2*, and *ISM2,* contributing to vCCT fate commitment through their involvement in immunomodulation and maternal spiral arteries remodeling [402-404, 422-426, 432]. Spatial distribution analysis confirmed robust expression of vCCT signatures, underscoring the significance of these genes in trophoblast development and the regulation of deep decidual invasion.

The differentiation of vSTB was characterized by the expression of transition markers, including genes like *CGA, KYNU, ARHGAP26, ADAM12, AFF1, PAPPA2, LIFR*, and *PSG4*, which exhibited a positive correlation with pseudotime. These genes play pivotal roles in various aspects of placental development, such as hormone production, immunomodulation, and maintenance of a tolerogenic environment [231-234, 438-442]. Notably, *ADAM12, PAPPA2*, and *LIFR* are regulated by *EP300/p300*, thus influencing the placental syncytialization [434, 446] process necessary for maintaining the multinucleated vSTB barrier. Spatial analysis confirmed the robust expression of the vSTB signature module, validating the spatial localization of vSTB developmental drivers. Additionally, the downregulation of *TEAD1* and repression of the YAP/TAZ pathway during vSTB differentiation was observed, further supporting the known biology of the vSTB lineage. Key vSTB lineage markers, such as *SDC1, CGA, GDF15, LNPEP1*, and *BMP1*, were validated at the protein level, highlighting their significance in placental function and development. Importantly, *GDF15* is a SASP ligand that likely contributes to maternal-fetal communication, and immune tolerance [450]. *LNPEP1* is an enzyme involved in peptide metabolism, including oxytocin and vasopressin, and may facilitate peptide exchange and nutrient transport across the vSTB barrier [451]. One question that emerges from these findings is: which of the transition markers identified are later dysregulated in eoPE, and what are the precise mechanisms underlying the progression of the disease?

- **vCTBpf transitional state regulated by BMP-antagonist *GREM2***

To trace the differentiation trajectory of mononuclear vCTB to vSTB lineage- two transitional states, termed as vCTBpf and vSTBjuv, were closely examined. Of note, trajectory analysis suggests that the dynamic transitional state vCTBpf serves as a progenitor for vSTB and fuses to juvenile STBs during early placental development, potentially regulating cell-to-cell fusion. Syncytin-2 (*ERVFRD-1*), a potent fusion inducer and an *EP300/p300* target, was identified as a specific marker for vCTBpf, aligning with previous data from Liu et al. [401]. In this regard, similar retroviral encoding genes such as *ERVV-1* and *ERVV-2* were found to be specific markers of the vCTBpf state and could regulate gene expression controlling trophoblast differentiation and syncytialization [401, 740], potentially affecting the formation of the vSTB layer. Other markers of the vCTBpf nuclear state include *OTUB2*, which might impact the ubiquitination of proteins [741] involved in trophoblast differentiation and syncytialization, influencing the formation and function of the syncytiotrophoblast.

Notably, the BMP-antagonist Gremlin 2 *(GREM2)* [435] was robustly expressed in vCTBpf, and the temporary inhibition of BMP in this state, as indicated by the downregulation of

BMP7, may play a crucial role in the transition from vCTBpf to vSTB. By modulating BMP signaling, *GREM2* might potentially oversee placental angiogenesis, thus influencing its vascular branching patterns crucial for transporting nutrients and oxygen to the developing fetus [742]. Also, the functional aspect of vCTBpf in regulating trophoblast fusion was subsequently confirmed by Arutyunyan et al. 2023 [292], who also generated a transcriptomic profile of this cell state in first-trimester placental villi. Taken together, these findings highlighted the significance of vCTBpf as a distinct sub-state undergoing evolution towards vSTB through specific molecular changes regulated by BMP signaling.

- **Experimental validations of vCTBpf (insights from collaborative data)**

The computational findings regarding vCTBpf solicited additional experimental validations beyond the scope of this dissertation. In collaboration with clinical partners, GREM2+ vCTBpf cells were localized in early villous tissue using immunofluorescence and were found to express the classical vCTB marker CDH1, indicating their status as mono-nucleated vCTB subtypes prior to fusion with vSTB. Flow cytometry analysis of primary CDH1+ vCTB cells isolated from first-trimester placentas confirmed the presence of a GREM2 and ABTB2-positive vCTBpf cell fraction. Moreover, our collaborators validated that BMP7 inhibited trophoblast fusion in BeWo cells- a model for trophoblast syncytialization [743, 744]. This supported the initial hypothesis that temporary inhibition of BMP signaling by GREM2 in vCTBpf cells is essential for trophoblast fusion and differentiation into vSTB, highlighting the importance of the "pre-fusion vCTB (vCTBpf)" state in syncytiotrophoblast formation.

6.1.3 Immune dysregulation in eoPE vs. term controls

- **Dysregulated lymphoid cell types/states in decidua**

The application of Logistic Regression in this study unveiled molecular dysregulations within lymphoid and myeloid immune populations at the maternal-fetal interface. Differential upregulation of genes and transcriptional regulators in dNK1 and dNK2 sub-states suggested potential impact on NK cell cytotoxicity and TGF-β signaling [454-456], while factors influencing differentially upregulated dTcell genes like *XAF1* indicated involvement in pro-inflammatory responses and apoptosis [458]. These findings shed light on potential mechanisms underlying the pathogenesis of eoPE, including the role of transcription factors like *ATF4* and *NFκB* in apoptosis regulation and inflammatory responses, respectively in dTcells. For instance, *ATF4* regulates apoptosis-associated genes, functioning through regulating unfolded protein response (UPR), endoplasmic reticulum (ER) stress, and oxidative stress adaptation [466]. During PE, heightened ER stress was shown to result in

increased PERK and IRE1α expression that could further activate *ATF* genes and lead to disrupted vasculature and abnormal placentation [467]. Also, elevated expression NFκB was previously observed in PE, where it stimulates the expression of both pro-inflammatory and anti-angiogenic proteins, thereby exacerbating oxidative stress, inflammation, and vascular dysfunction characteristic of PE [468].

Furthermore, pathway enrichment analysis revealed enrichments in processes such as transcriptional misregulation in cancer, transcriptional regulation by RUNX3 and TGF-β pathways in dNK1/2 subtypes. Increased TGF-β signaling in eoPE could contribute to impaired immune tolerance and inflammation [469]. Previously, increased decidual TGF-β1 levels were found to inhibit the activation of particular subsets of dNK cells, thereby contributing to the uteroplacental pathology associated with the onset of PE [470]. Also, Yazaki et al. identified that TGF-β1 induced both ROS generation and Nrf2 activation in specific cells [745]. Of note, *NRF2* is a pivotal regulator of inflammation, cellular oxidative response, and autophagy [746], and could be potentially induced during ROS imbalance [747]. Other studies reported that increased Nrf2 expression resulted in decreased levels of ROS under hypoxic conditions as an adaptive mechanism to oxidative stress [746, 748-750]. These collective findings suggest that TGF-β1 might contribute to the development of PE through the ROS-Nrf2 signaling pathway [750].

The enrichment of pathways such as CCR5, ILK, NFAT, cytokine signaling, and HSP90 chaperone signaling in dTcells suggested potential involvement in the inflammation linked to endothelial dysfunction and vascular disorders in eoPE. Chronic hypoxia associated with PE can lead to heightened ER stress, exacerbating unfolded protein response (UPR) activity [751] and compromising autophagy [752]. Moreover, ILK signaling, crucial for various cellular processes including cell migration and angiogenesis, may be influenced by elevated UPR and hypoxia [753] during eoPE, potentially impacting decidual function. A study by Chou et. al. 2015 deciphered a regulatory loop between HIF-1α and ILK, promoting hypoxia-induced expression of HIF-1α and EMT in prostate and breast cancer cells [753]. Further investigations are needed to elucidate the molecular mechanisms by which hypoxia and the ILK pathway contribute to the progression of eoPE in the decidua, as well as whether they induce disturbances in placental function. Notably, increased estrogen signaling observed in dTcell and dNK1 might serve as a compensatory mechanism to counteract decreased 2-methoxyestradiol levels in PE [472-474], aiming to activate HIF1α and VEGFR-2 to maintain placental perfusion through increased angiogenesis [281, 475].

- **Dysregulations in analogous cell types- maternal macrophages and fetal Hofbauer cells**

Significantly, the analysis focused on evaluating transcriptomic alterations in similar cell types or states, specifically maternal dMAC1/2 sub-states, dMono1 and fetal vHBC. dMono2 was not included in the analysis due to its presence exclusively in samples generated from the 10X v3 library. Exclusion of the comparison between dMAC1/2 and vPAMM, as well as dTcell and vTcell, was warranted due to potential sampling site effects observed in late pregnancies impacting vPAMM and vTcell integration.

Similar to lymphoid group analysis, Logistic regression identified differentially upregulated genes in decidual macrophages (dMAC1 and dMAC2) and decidual monocytes (dMono1). *DUSP1* found dysregulated in both dMAC1 and dMAC2, is implicated in reducing p-MAPK1 and *ERK1/2* expression, potentially contributing to PE [476]. *MAFB*, associated with dMAC1, is linked to lipid level variations and atherosclerosis, suggesting a role in macrophage polarization and cholesterol efflux [477, 478]. *PAPPA2,* found dysregulated in vHBC, plays a critical role in regulating insulin-like growth factor (IGF) bioavailability by cleaving IGF-binding proteins. Dysregulation of PAPPA2 in PE is associated with adverse effects, including disrupted placental development, impaired angiogenesis, and perturbed trophoblast function [481]. PAPPA2 is also previously suggested as a potential biomarker for PE, highlighting its diagnostic and therapeutic significance [482, 483].

Subsequently, iRegulon analysis inferred key transcriptional regulators for each cell type or state. Significantly, *GMEB2* acting on dysregulated targets in dMAC1 could be in response to glucocorticoid signals [484]. While it is acknowledged that glucocorticoids can induce immunomodulation and impact endothelial function [485], additional research is needed to explore whether disrupted glucocorticoid signaling contributes to endothelial dysfunction in eoPE. The transcriptional landscape of dMAC2 involves key regulators such as *FOS, MAF, BCLAF1,* and *SRF*. Dysregulation of *FOS* in maternal macrophages could potentially lead to an imbalance in cytokine production, promoting a pro-inflammatory environment [754, 755]. This inflammatory state is implicated in endothelial dysfunction and vascular damage associated with eoPE. *MAF*, another member of the AP-1 complex, often forms heterodimers with *FOS* and is an important regulator of M2-like macrophages. Additional investigation is needed to explore whether transcription factors such as *FOS* and *MAF* can disrupt the regulatory mechanisms controlling the M1/M2 balance, potentially contributing to eoPE, fetal growth restriction, and preterm delivery. *BCLAF1* is a transcription factor that interacts with anti-apoptotic members of the *BCL2* family [486, 487]. Dysregulation of *BCLAF1* could impact the apoptotic balance and cellular turnover [486, 487] in maternal macrophages,

potentially contributing to altered immune responses and cellular dynamics associated with eoPE. Another noteworthy transcription factor that regulates dMAC2 is *SRF*, known for its broader cellular functions, including the regulation of mitochondrial dynamics that can impact energy metabolism within maternal macrophages, influencing their overall activity [488-490]. Moreover, *SRF* controls fatty acid translocation, suggesting a role in lipid metabolism as well as the expression of electron transport chain (ETC) complex proteins [489, 490]. Of note, dysregulated lipid metabolism can lead to increased oxidative stress in eoPE. Oxidative stress contributes to the release of reactive oxygen species (ROS), which further exacerbates endothelial dysfunction and contributes to the overall pathogenesis of the condition [756]. Moreover, disruptions in the translocation of fatty acids may contribute to insulin resistance [757, 758] in maternal tissues, a condition frequently linked to PE [759, 760]. This contributes to both metabolic dysfunction and impairment of endothelial function and would serve as an interesting line of investigation. Additionally, *SRF* has been implicated in the regulation of apoptosis by controlling the expression of anti-apoptotic molecules such as *BCL2* [761]. The dysregulation of these cellular processes, governed by *SRF*, might contribute to the pathophysiology of eoPE, affecting maternal macrophage functions and potentially leading to complications in pregnancy.

Differential pathway enrichment analysis demonstrated enriched cytokine signaling in dMono1 and dMAC1/2, suggesting an aberrant activation of immune responses that may contribute to the development of the eoPE pathophysiology. Inflammation is a key component of PE, affecting endothelial function and vascular integrity [515, 522, 762, 763]. Dysregulated cytokine signaling can contribute to an inflammatory environment, leading to endothelial dysfunction, vascular damage, and ultimately, the clinical manifestations of PE [762, 763]. Further understanding of the intricate connections between cytokine signaling and the immunological aspects of eoPE is required to unravel the underlying mechanisms of the disorder. It may provide insights into potential therapeutic targets aimed at modulating immune responses and mitigating the inflammatory processes associated with this condition.

6.1.4 vSTB is majorly dysregulated in eoPE

The molecular pathomechanisms underlying eoPE trophoblast and the identification of at-risk mothers have remained elusive. This study specifically concentrated on the disease-centric dysregulation within the placental vSTB, given its direct contact with maternal blood and the potential to transmit eoPE signals from the fetus to the mother. Distinct and shared perturbed profiles were observed among the three vSTB nuclei sub-states. The shared profile,

reflecting a core dysregulation within the vSTB compartment, indicated consistent pattern of gene regulation across nuclear sub-states.

Generally, single-cell/nuclei RNA sequencing encounters challenges with dropout rates and low library sizes, hindering the comprehensive capture of transcription factors [160]. Methods like scCapture-Seq [160] improve TF detection but might overlook markers expressed in small cell fractions. To overcome this constraint, iRegulon [161] was applied on differentially expressed vSTB genes for identifying significant transcriptional regulators and co-factors, enhancing comprehension of the molecular regulation within vSTB.

- **Dysregulated p300 mediated trophoblast fusion in eoPE**

The disruptions of molecular processes involved in placental fusion, which regulates the differentiation of vCTB into vSTB, are expected to be a key factor in the development of eoPE. Transcriptional analysis using iRegulon revealed that the co-activator *EP300*, associated with cell cycle arrest [512, 513], emerged as the most enriched regulator acting on dysregulated vSTB targets, suggesting its involvement in trophoblast fusion disruptions [764]. To validate this computational finding, my clinical collaborators compared *EP300* expression in eoPE and gestational age-matched non-PE preterm pathologies in cohorts from four other centers. A significantly higher expression of *EP300* was observed in eoPE compared to preterm placentas across these cohorts, confirming the findings predicted from snRNA-seq data. Importantly, no difference in *EP300* expression was found in eoPE with and without FGR, suggesting that *EP300* upregulation in placental tissue is specific to the pathology of PE.

Notably, fusogenic genes linked to syncytium formation, including *GCM1* [446], *ADAM12* [444, 445], and *DYSF* [496], were perturbed in the vSTB sub-states, indicating their potential role in eoPE. Furthermore, dysregulation of p300 targets included known candidates for PE progression, such as *ENG* [497-499], *FOS* [501], *SAT1* [500], *SERPINE1* [502, 507], *GADD45G* [503], and *INHBA* [504-506], implicated in angiogenesis, vascular dysfunction, coagulation, and signaling pathways. Elevated serum ENG (sEng) level is known to impede TGF-β1 signaling, thereby decreasing activation of vasodilators like nitric oxide synthase [498, 499]. This interplay between sEng and TGF-β highlights its role in the pathogenesis of impaired placentation and compromised blood flow due to vaso-constrictions in PE. Interestingly, *FOS* happened to be another major transcriptional regulator of dysregulated STB genes, originally inferred by iRegulon analysis. In essence, *FOS,* operating as a constituent of the AP-1 transcription factor complex [501], oversees a range of cellular processes such as proliferation, differentiation, apoptosis, hypoxia response, angiogenesis

[765], and steroidogenesis [766]. Previously, transgenic overexpression of SAT1 or its pharmacological activation resulted in increased mono- or diacetylated spermidine and spermine levels that ultimately limited cellular proliferation by inducing apoptosis [767, 768]. Importantly, maternal serum levels of diacetylspermine metabolite were found to be significantly high in PE but not in FGR [769]. *SERPINE1*, a member of the serine protease inhibitor (SERPIN) gene family, is upregulated in response to hypoxia and exhibits altered expression in placentas from preeclamptic pregnancies [507]. Previous research has demonstrated changes in plasma levels of *SERPINE1* in women with PE as compared to controls [507, 770, 771]. However, *SERPINE1* levels in the placenta may not adequately reflect changes observed in maternal serum, as *SERPINE1* is secreted from sources other than placental villi — for example, in decidual cells [772]. It should be still considered that *SERPINE1* functions as a key regulator in various molecular pathways, including inflammation, coagulation, fibrinolysis, complement activation, and phagocytosis, all of which play crucial roles in placental diseases [507, 773]. In a recent study by Yang et al. 2023 [774], *SERPINE1* emerged as one of the seven major diagnostic candidates for PE, along with *FLT1*, and operates within the HIF1-signaling pathway. Consecutively, network analysis of *EP300*/*p300* targets highlighted a closely interconnected community subgraph, including *ATF3, DUSP1, FOS, FOSB, JUNB*, and *JUND,* implicated in the AP1 pathway. In PE, abnormal levels of *DUSP1* might result in impaired placental development via MAPK/ERK signaling [476].

Collectively, the findings strongly suggested that disrupted *EP300* activity impairs normal fusion and vSTB differentiation during early gestation, resulting in aberrant alterations in transcriptional signatures across all STB nuclear sub-states later in eoPE.

- **Shared dysregulated targets drive hypoxia in eoPE**

In the vSTB, a significant subset of disturbed targets was co-regulated by multiple transcription factors and co-regulators, including master-regulator *EP300, FOS, FOXO1, FOXO4,* and *PAX5*. The functional implications of these shared targets in eoPE pathogenesis involve the enrichment of signaling cascades such as HIF-1α, AP1, PI3K-Akt, and TGFβ pathways.

In PE, there is shallow and incomplete placental endovascular invasion caused by inappropriate changes in adhesion molecules, resulting in placental hypoxia [775, 776]. Increased levels of HIF-1α during hypoxia can boost secretion of various mediators, including sFlt-1, sEng, and angiotensin II converting enzyme (ACE), into the maternal bloodstream that can promote eoPE pathogenesis and abnormal placentation [497, 777]. In this regard, prior

studies had reported that upregulated HIF-1α enhanced sFLT-1 expression levels in PE-affected placentas [778, 779]. Notably, sFLT-1 is a key factor in both PE and IUGR [780]. Additionally, hypoxia-induced activation of AP-1 has been observed in various cell lines, where it collaborates with HIF-1α to enhance VEGF gene transcription under hypoxic conditions [781]. Hypoxia can also amplify TGF-β/Smad signaling through HIF-1α, which could lead to the upregulation of ECM components, including PAI-1 (encoded by *SERPINE1*) and TIMP [782, 783]. Moreover, the TGFβ-SMAD pathway regulates trophoblast syncytialization and proliferation through GCM1 and cell cycle regulators, influencing trophoblast fusion and migration [784]. Hence, dysregulated TGFβ signaling has the potential to drive aberrant trophoblast differentiation and impaired placental development through a complex interplay with hypoxia during eoPE.

A promising area for investigation involves inducing *EP300* or other inferred transcription factors combinatorially in first-trimester trophoblast organoids or cell lines to examine whether this induction triggers eoPE phenotypes such as hypoxia and senescence. Utilizing these models can aid researchers in elucidating the mechanisms underlying eoPE, screening prospective drugs, and devising personalized medicine strategies.

- **vSTB developmental drivers dysregulated in eoPE**

Furthermore, this dissertation leveraged the first trimester STB differentiation trajectory for connecting evidence from eoPE arising later in gestation to the early stages when diagnosis is not feasible. In simpler terms, the assessment of vSTB differentiation was carried out by comparing disease signatures and pseudotime markers to identify molecular programs causing eoPE-affected cells to deviate from their normal developmental trajectory. The analysis demonstrated a noteworthy enrichment of eoPE differentially expressed signatures in transition genes associated with vSTB development, indicating a substantial impact on the vSTB differentiation trajectory. Despite the initial higher composition of vSTB1 and vSTB2 compared to vSTBjuv, a careful downsampling approach was employed to ensure consistent statistical power for DEG analysis. Remarkably, even after downsampling, the significant enrichment of vSTB transition genes within eoPE DEG signatures persisted, underscoring the robust association between eoPE and altered vSTB developmental processes.

An illustrative example involves the dysregulation of pseudotime markers *LEP* and *LNPEP* observed in eoPE. Specifically, *LEP* was significantly upregulated across all three vSTB sub-states, while LNPEP1 was downregulated in vSTB1 compared to term controls. Previous research highlighted elevated maternal LEP levels in both serum and placenta as a PE

characteristic feature [516, 517, 785]. Also, LEP is well known for its involvement in endothelial dysfunction, and atherosclerosis considering cardiovascular disorders, often in conjunction with components of the renin-angiotensin system (RAS) [786]. An interesting study by Nonn et al. 2021 [787] proposed a mechanism wherein maternal Angiotensin IV (Ang IV), which increases during pregnancy, acts on trophoblastic *LNPEP*, leading to a reduction in basal mitochondrial respiration. This reduction subsequently triggers an increase in *LEP* expression, which in turn, negatively regulates *LNPEP* expression at both mRNA and protein levels through a negative feed-forward loop. While targeting *LEP* pharmacologically may pose challenges due to its central role in cellular metabolism regulation, *LNPEP* shed into maternal plasma could be a viable consideration to mitigate detrimental effects associated with increased maternal LEP levels, which are implicated in the pathogenesis of eoPE.

6.1.5 vSTB-derived elevated SASPs mediate the transition of eoPE from child to mother

This dissertation extensively investigated whether localized senescence observed in eoPE within the trophoblast interface, linking fetal circulation and maternal blood, leads to the release of senescence-associated secretory phenotype (SASP) factors into the maternal circulation. Among the DEG(s) shared across various vSTB sub-states, approximately 12.2% were identified as genes encoding SASP. Dysregulation of senescence-associated *EP300* targets was evident in all three nuclear states of vSTB in eoPE, particularly pronounced in vSTB1 and vSTB2, suggesting a potential role of the transcriptional co-activator *EP300* in inducing premature senescence in these states during eoPE. This association between PE and defects in villous trophoblast fusion leading to senescence, mediated by *EP300*, may represent a fundamental aspect across all PE subtypes, regardless of their multifactorial origins. Various triggers and clinical risk factors for PE, such as obesity, may converge in dysfunction of the maternal-fetal barrier, accelerating the turnover of villous trophoblasts and promoting premature senescence. This hypothesis finds support in similar findings of SASP profiles observed in the maternal circulation of both late-onset and early-onset PE cohorts.

- **vSTB as possible cell-of-origin in eoPE**

Notably, the discovery from this dissertation strongly indicated that eoPE majorly dysregulates vSTB sub-states lining the outer layer of the fetal placenta, characterized by an elevated SASP profile. Receptor-ligand interaction analysis that profiled perturbed cellular communication between vSTB and maternal endothelium receptors revealed increased

secretion of ligands such as *GDF15, INHBA, HSPG2, MIF, TGM2, EZR, ADAM12,* and *ADAM9.* This implied that a heightened ligand pressure originates in vSTB sub-states and causes accelerated senescence at the fetal side during eoPE (as illustrated in **Figure 6.1**). Impressively, these ligands can cross the maternal-fetal interface and can potentially translate eoPE from the fetal to the maternal side. Therefore, this also implies that eoPE may potentially arise within the outer layer of vSTB, with the SASP facilitating the transmission of the disease from the fetal to the maternal side. It would be also noteworthy to explore the impact of *EP300* agonism in an in vitro primary trophoblast model and assess whether it induces specific SASP expression in a time- or dose-dependent manner, thereby mimicking eoPE progression.

Collectively, these SASP-mediated changes in vSTB function could potentially disrupt maternal-fetal interactions, leading to impaired nutrient exchange, hypoxia, and placental oxidative stress — known hallmarks of eoPE [788, 790]. For instance, *GDF-15* responds robustly to oxidative stress and simulated ischemia/reperfusion injury in cultured cardiomyocytes [789]. Similarly, the escalation of oxidative and nitrosative stresses in the placenta is characteristic of PE, along with ischemia/reperfusion events [790]. Moreover, *GDF15* is a stress-responsive cytokine related to TGF-β and is increasingly recognized as a novel risk indicator in individuals with cardiovascular diseases [791, 792]. A prior study conducted by Sugulle et al. in 2009 [793] revealed elevated levels of GDF-15 in patients with PE and superimposed PE in those with diabetes mellitus when compared to control subjects. *INHBA,* a subunit of Activin-A and Inhibin-A cytokines belonging to the TGF-β superfamily is known to regulate cell growth and differentiation and has been implicated in mediating cellular senescence in various contexts [794, 795]. However, the diverse inputs influencing the release of activin A likely contribute to its wide-ranging serum values across various pregnancy pathologies, limiting its efficacy as a standalone biomarker [796]. Therefore, exploring its clinical utility in conjunction with other biomarkers like *GDF15* or *HSPG2* is imperative.

Elevated levels of placental perlecan (*HSPG2*) expression have been linked to hypoxic-ischemic placental injury seen in miscarriages and eoPE, regardless of HELLP syndrome presence [797]. Given its role in vascular malperfusion [797], targeting *HSPG2* activity may offer the potential to mitigate vascular complications associated with PE. Additionally, *ADAM12* and *ADAM9* play roles in cell adhesion and migration pertinent to placental function, making them potential therapeutic targets for addressing PE pathophysiology. Placental ischemia or hypoxia induces ROS production, disrupting *ADAM* gene expression and activity, thereby disturbing the balance between pro and anti-angiogenic factors, and promoting the

release of inflammatory cytokines, that contribute to the inflammatory response and endothelial dysfunction characteristic of PE [798]. Of note, a previous study found elevated *ADAM9* expression levels in PE, and its stabilization was facilitated through de-ubiquitination by USP22 [523]. This stabilization led to the inhibition of trophoblast cell proliferation, migration, invasion, and EMT progression. These findings suggest a novel pathway involving the USP10/RUNX1 axis in the pathogenesis of PE.

Simply put, the findings of this thesis proposed that a disrupted cellular communication between vSTB sub-states and maternal decidua is the key to a dysregulated maternal-fetal barrier and potentially compromised maternal uterine vessel remodeling in eoPE. Hence, a potential strategy for the prevention, intervention, and clinical management of eoPE could be accomplished through the pharmacological inhibition of the SASP ligands, including *GDF15, INHBA, ADAM9,* and *HSPG2*.

Furthermore, ISS-based spatial transcriptomic analysis revealed a significant loss of spatial exclusivity between the senescent marker *INHBA* and markers of fetal vessels in eoPE, providing a potential reason behind impaired nutrient transport and vascular dysfunction within the placenta. This association was not detected in term controls and possibly, suggests impaired nutrient transport, and vascular dysfunction within the eoPE-affected placenta. Taken together, this spatial proximity underscored the role of cellular senescence in causing tissue level perturbations and altered cellular function, offering a novel insight into eoPE pathophysiology that warrants further validation in larger cohorts.

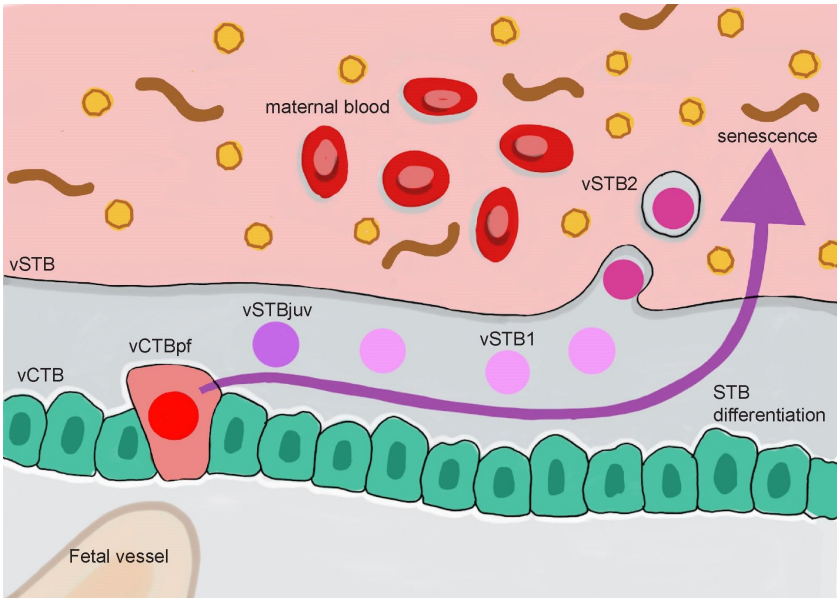


Figure 6.1: Illustration depicting elevated SASP expression in maternal blood during early-onset pre-eclampsia.

6.1.6 Future Outlook and Precision Medicine in PE

The validation studies, inspired by the computational findings of this thesis, were conducted through extensive collaborative efforts involving clinicians, scientists, and statisticians across various centers in Berlin, Graz, Vienna, and Melbourne, thereby significantly advancing research on PE.

- **Diagnostics avenues (taken insights from additional collaborative data)**

To assess the predictive potential and pathomechanistic relevance of the computationally derived SASP markers, experimental validations are planned to be conducted on two prospective longitudinal cohorts. The first cohort involves women who provided blood samples at the 36th week preceding delivery and PE diagnosis. Early research revealed increased levels of PAI-1 (*SERPINE1*) and GDF-15 in the maternal serum of samples collected at this stage that were predictive of eoPE development. Multivariable logistic regression analyses demonstrated that these associations remained significant even after adjusting for BMI and nulliparity. The role of PAI-1 in predicting eoPE prior to disease onset indicated its connection to premature vSTB senescence. However, additional markers such as *HSPG2, ADAM9* and *ADAM12* await validation.

Of note, validations are scheduled in a subsequent cohort, where maternal (first-trimester) serum was obtained from participants who enrolled during the first trimester but were later diagnosed with eoPE during late gestation. Preliminary data indicated that Activin A (INHBA) served as a notable predictor of the disease during the early stages of pregnancy for women who later developed PE. In this context, statistical modeling conducted using the Fetal Maternal Foundation (FMF) algorithm [799] indicated a significant correlation between elevated levels of circulating Activin A in the first trimester and a heightened estimated risk of PE development. Nevertheless, further research is required to identify an optimal set of early diagnostic markers before the eoPE symptoms manifest and cause irreversible harm to both mother and child.

In summary, it was revealed that placental dysregulation during vCTB to vSTB differentiation initiates eoPE and culminates in premature SASP. This dysregulation can be detected in placental tissue and maternal blood as early as the first trimester, and ultimately, this SASP is responsible for translating the placental differentiation defect into a maternal

cardiovascular syndrome. These findings also align with observations in cardiovascular disease beyond pregnancy, where SASP is known to play a role.

- **Therapeutic avenues- pinpointing towards TGF-β signaling inhibitors**

Targeting SASP components or modulating the senescence process could offer novel therapeutic strategies for eoPE management. Interventions aimed at reducing SASP-associated inflammation or promoting senescent cell clearance might mitigate the progression of eoPE and its associated complications. Importantly, therapies aimed at reducing senescence or targeting the same early in pregnancy could potentially prevent or delay the onset of eoPE in at-risk individuals.

Based on snRNA-seq data, it is suggested to consider targeting *EP300*, as it is believed to potentially contribute to the exaggerated SASP observed during eoPE. Although directly targeting *EP300* presents significant challenges given its fundamental physiological functions in the placenta, it might be more feasible to pharmacologically inhibit its downstream targets, such as *FOS*, which play significant roles in the progression of eoPE. Previous studies postulated that targeting the acetyltransferase *p300* could be a promising avenue for developing epigenetic drugs aimed at mitigating accelerated senescence and aging-related cardiovascular diseases [800]. While various small molecule inhibitors directed against p300 have demonstrated efficacy in preclinical models for alleviating diverse cardiovascular pathologies, none have progressed to clinical trials. Therefore, further research is warranted to optimize and evaluate the therapeutic potential of *p300* inhibitors in clinical settings, specifically for pre-eclampsia.

Therapeutically, targeting *GDF15* signaling pathways may have the potential in managing conditions associated with inflammation, oxidative stress, and metabolic disorders [793]. However, challenges persist in understanding the regulatory mechanisms controlling GDF15 expression in response to drugs, hindering the development of effective therapeutic interventions. Recent research implicates BET proteins, particularly BRD4, in GDF15 regulation [801]. BET proteins, which regulate inflammation-related genes like GDF15-activating factors NR5A2 and NFkB, play a crucial role in exacerbating cytokine expression in diseases such as cancer or heart failure [801]. Despite these complexities, numerous clinical studies have explored BET inhibitors [801], primarily in cancer treatment, indicating the therapeutic potential of targeting BET proteins in various pathological conditions. However, extensive research is needed to investigate if GDF15 could be practically targeted in eoPE.

On a similar note, interventions directed at HSPG2-mediated pathways hold promise for tissue repair, angiogenesis, and conditions involving extracellular matrix dysregulation. Like *GDF15* and *INHBA*, regulation of HSPG2 is also responsive to TGF-β signaling, suggesting that investigating TGF-β blockades could be a compelling avenue for exploring therapeutic strategies in eoPE. However, as of now, there are no approved drugs specifically targeting HSPG2 and other mentioned biomarkers for the treatment of PE. Further research is solicited to explore their therapeutic potential and develop targeted interventions for PE management.

- **Clinical management of eoPE**

Developing "gestational clocks" for diagnosing and managing eoPE would involve integrating data from snRNA-seq, along with other multi-omics data and clinical information sourced from databases like the UK Biobank [802]. For instance, monitoring the levels of specific SASP components (originally inferred using snRNA-seq) in maternal circulation could provide valuable insights into disease severity and progress. Additionally, combining multi-omics data, including proteomics and metabolomics, would offer a holistic understanding of molecular changes linked to eoPE compared to healthy pregnancies.

By analyzing biomarkers encoding hormones, SASP proteins, and genetic markers alongside clinical data, machine-learning algorithms can be trained to accurately estimate disease risk before symptoms arise. This approach enhances clinicians' ability to monitor pregnancy progression effectively and identify deviations from the norm, such as eoPE, enabling earlier diagnosis and intervention.

The implications of gestational clocks for early eoPE diagnosis are expected to be crucial for timely medical interventions to prevent complications for both the mother and the fetus. By identifying patterns in biomarker levels that are indicative of PE development, gestational clocks can provide warning signs earlier in pregnancy, allowing healthcare providers to closely monitor at-risk pregnancies and implement preventive measures or treatment strategies promptly. For example, if GDF-15 remains consistently elevated in maternal serum post-childbirth, it suggests a heightened risk of cardiovascular disorders and a potential predisposition to PE in subsequent pregnancies. This correlation stems from the association between elevated GDF-15 levels in circulation and an augmented risk of cardiovascular events in non-pregnant individuals [515].

6.2 Development of human pancreatic islets

At the perinatal stage, beta cells are not fully functional, but their functions, including glucose-regulated insulin secretion, mature with age [311]. To investigate transcriptomic changes across the two different age groups (neonatal and adult), the data sets were combined and diffusion pseudotime analyses were performed to explore age-dependent development. Pseudotime ordering was found to recapitulate donor age, thereby allowing gene expression modeling using a generalized additive model and the identification of highly dynamic genes. Genes depicting increasing or decreasing expression levels across pseudotime were identified, as discussed in the manuscript. Genes implicated in beta cell proliferation, such as *PDZD2* [803] and *IGFBP5* [804], exhibit lower expression levels in adult samples. *PLAG1*, a protein known to decline within a few days after birth and known to inhibit insulin secretion in neonatal murine islets [805], was also detected in neonatal samples. In contrast, *CD99*, the expression level of which was previously reported to increase in adult mouse islets [806], also had increased mRNA levels in adult human beta cells. In human adults, higher levels of *RASD1* and *SYT16*, genes previously reported to be upregulated in human islets exposed to relatively high glucose, were detected [807, 808]. In adult beta cells, there is higher expression of genes that encode members of the secretogranin-chromogranin family, including *SCG2, SCG5,* and *CHGB*, which are recognized for their crucial function within the insulin secretory granule. Therefore, the utilization of snRNA-seq on tissues obtained at distinct developmental stages unveiled markers and previously under-recognized regulators of human beta cell functional maturation, shedding light on age-related restrictions in beta cell proliferation and the progression of secretory functions.

6.3 Generating first reference atlas for pancreatic neuroendocrine carcinoma

The study delved into the intricate molecular landscape of high-grade large cell pancreatic neuroendocrine carcinoma (panNEC) through snRNA-seq based transcriptomic profiling of surgically resected fresh frozen specimens from five patients, one of whom (P1) had received prior treatment. By employing advanced computational data analytics, approximately 45,015 single nuclei were meticulously analyzed. Through unsupervised clustering, five distinct cell types were identified across the patient samples: neuroendocrine (NE), amphicrine, amphicrine progenitor, stromal (normal), and immune populations. While four patients (P1-P4) predominantly exhibited shared neuroendocrine cell types, the fifth patient (P5) demonstrated unique clusters with pronounced expression of acinar markers [539, 540]. Notably, NE cell types displayed variable expression of diagnostic markers such as *CHGA, SYP,* and *NCAM1* [328-331, 536-538], also validated through immunohistochemical staining. The amphicrine cell type showcased co-expression of pancreatic acinar [539-544] and neuroendocrine markers, while the amphicrine progenitor-like cell type expressed early pancreatic progenitor markers, such as *PDX1* and *SOX9* [545-549]. Stromal cells, characterized by ECM remodeling genes, were more abundant in P1, previously treated with chemotherapy. The immune cell type harbored markers indicative of its immune function. Hence, for the first time, snRNA-seq data analysis uncovered tissue compostion and distinct cell type classifications, emphasizing the heterogeneity within panNEC and its tumor microenvironment.

6.3.1 Shared panNEC sub-states reveals diverse functional profiles

Through meticulous subclustering of cell types into sub-states, this dissertation uncovered distinct functional diversity within high-grade panNEC patients. Despite striking patient-to-patient variability, common sub-states within the neuroendocrine cell type were discerned, each characterized by distinct molecular signatures indicative of functional pathways and oncogenic processes. The shared neuroendocrine (NE) sub-state demonstrated a distinct molecular profile characterized by the presence of well-established neuroendocrine markers such as *RIMBP2* and *ADARB2*, associated with neuronal systems, alongside *CACNB2, CACNA1A,* and *KCNJ3*, pivotal in voltage-gated Ca2+ and K+ channel regulation [551]. This amalgamation of markers contributed significantly to enriched pre-synaptic depolarization, indicating an active role in neuronal regulatory processes. Of note, the calcium channel encoded by the *CACNB2* gene, a membrane-associated guanylate kinase (MAGUK) protein, serves as the β2 subunit of the L-type cardiac calcium channel *CACNA1C*. This discovery underscores the importance of L-type calcium channels in facilitating Ca2+ influx into the

cytoplasm, thereby regulating cardiac excitability and excitation-contraction coupling (EC coupling) [552]. It is well known that ion channel-specific genes play pivotal roles in various cellular processes, including electrical excitability, cell signaling, and hormone release. In the context of panNEC, dysregulation of ion channels might influence the tumor microenvironment, cellular communication, and hormone secretion, all of which are hallmark features of neuroendocrine cells. Hence, understanding the significance of these genes could offer insights into the panNEC pathophysiology. Of note, previous studies in glioblastoma have highlighted the contribution of electrochemical signaling to tumor growth in preclinical models [803]. Therefore, the neural characteristics of LCNEC may present potential targets for NEC-specific therapies, distinguishing them from treatments for pancreatic ductal adenocarcinoma (PDAC) [804]. Notably, neuroendocrine proliferating sub-state (NE proliferating) is an actively cycling sub-set of shared NE cluster, which exhibited enrichment in cell-cycle regulation, DNA damage response pathways and BRCA/FANCA signaling programs [553], with implications for tumor aggressiveness and therapeutic response. Previously, heightened expression of Fanconi anemia genes has been associated with resistance to chemotherapy in cisplatin-resistant non-small cell lung cancer (NSCLC) [554] as well as in ovarian cancer cell lines [555]. In case of P5, another proliferating amphicrine sub-state was detected, resembling the shared neuroendocrine proliferating sub-state, but additionally characterized by acinar markers.

Moreover, a hypoxic neuroendocrine sub-state, namely NE HSP+ (hypoxic) displayed activation of mTOR signaling, glycolysis, and heat stress response pathways, suggesting adaptations to tumor microenvironmental stresses [556, 557]. The excessive activation of mTORC signaling is recognized for its role in promoting cell proliferation, survival, and metabolism, thereby fueling tumor growth and progression [558]. This phenomenon is frequently observed in pancreatic neuroendocrine tumors (panNET) [559]. Previous studies have indicated a frequent abnormal activation of mTOR, often attributed to inactivating mutations in genes encoding negative regulators of the pathway or through indirect mechanisms [559]. Clinically, overexpression of mTOR and its downstream targets has been linked to a poorer prognosis in various neuroendocrine tumors [560, 561]. For instance, Shida et al. 2019 observed elevated mTOR expression in poorly differentiated neuroendocrine neoplasms (NENs), with a reported expression rate of 67%, contrasting with a reduced rate of 27% in well-differentiated counterparts [562]. Similarly, Catena et al. 2011 found mTOR expression in 80% of patients with poorly differentiated NEC, regardless of tumor origin or proliferation rate [563]. These findings underscore the clinical significance of mTOR dysregulation in the pathogenesis and prognosis of neuroendocrine tumors. Additionally, Stroma (normal) sub-state exhibited differential expression of genes involved in

ECM remodeling, immune modulation, and angiogenesis, hinting at their role in tumor progression and immune evasion. Importantly, the identification of distinct macrophage sub-states highlighted their immunomodulatory functions and potential as therapeutic targets. This comprehensive analysis elucidates potential avenues for targeted therapies and underlined the importance of understanding cellular diversity in advanced pancreatic neuroendocrine carcinoma patients.

6.3.2 *RBFOX1* and *NKX2-2* are crucial regulators of shared neuroendocrine and neuroendocrine proliferating sub-states

The shared panNEC sub-states revealed distinctive transcriptomic signatures indicative of underlying neuroendocrine characteristics. Specifically, *RBFOX1* [580] and *NKX2-2* [581] emerged as notable markers exclusive to NE and NE proliferating sub-states, which were absent in the amphicrine counterparts. *RBFOX1* exhibited statistically significant upregulation in both NE and NE proliferating sub-states, playing a crucial role in neuroplastic gene expression regulation through splicing mechanisms [583, 584]. Intriguingly, association of *RBFOX1* with calcium signaling pathways suggests a potential mechanistic link to the observed prominence of calcium channels within shared NE states [585]. Meanwhile, *NKX2-2*, despite not being identified as a computational marker, retained importance in pancreatic islet development, reflecting its expression in adult endocrine cell types [581]. Conversely, the amphicrine progenitor-like sub-state displayed robust *PROX1* expression, a transcription factor crucial for neurogenesis and pancreatic morphogenesis [586, 587].

6.3.3 Molecular landscape of P5-specific amphicrine sub-States: Insights into Early Pancreatic Progenitor signatures

In patient P5, a distinct amphicrine sub-state labeled 'Amphicrine progenitor-like' was identified, characterized by the expression of key markers and transcription factors associated with the differentiation of early pancreas progenitors, including multipotent pancreas progenitors (MPP) and tip progenitors [547, 590, 591]. Noteworthy markers *PDX1* and *SOX9* were co-expressed in this sub-state, indicating multilineage potential, as demonstrated in previous lineage investigation studies [545-549, 592, 593]. For example, Gu et al., 2002 [592] demonstrated that Pdx1 expressing cells could generate all adult pancreatic cell types. A previous study by Trott et al., 2017 [593], revealed that cultured pancreatic progenitor cells express *PDX1* and *SOX9*, which, upon exposure to differentiation signals, differentiate into pancreatic acinar, ductal, and endocrine lineages— thereby, indicating multilineage potential. *SOX9* serves as the primary marker and maintenance factor for multipotent progenitors during pancreas organogenesis [545, 547-549, 592-594]. In the

developing pancreas, *SOX9* expression is confined to a subset of actively dividing *PDX1+* pluripotent progenitors that are responsive to Notch signaling, while it is absent from committed endocrine precursors or differentiated cells [594].

Conversely, the amphicrine acinar03 sub-state, while expressing *PDX1* similar to amphicrine progenitor-like cells, lacked *SOX9* and instead displayed acinar lineage markers such as *GP2* [539, 540], indicating a more committed acinar fate. Additionally, the co-expression of *PDX1* and *ONECUT1* in both amphicrine progenitor-like and amphicrine acinar03 sub-states suggests their potential cooperativity in pancreatic cell differentiation and maturation [596]. Kropp et al 2018 [596] hypothesized that *Pdx1* and *Onecut1* are co-expressed in mouse MPPs and their cooperativity in MPP impacts postnatal β-cell maturation and functions. Moreover, early pancreatic progenitor markers like *HNF1B, GATA6,* and *ETV4,* were expressed across all amphicrine sub-states, with *HNF1B* being relatively robust in amphicrine acinar sub-states [597-602]. In this regard, Pinho et al. 2011 [597] observed that *Pdx1/Ptf1a* acinar cells in suspension culture dedifferentiate into an embryonic progenitor-like phenotype when exposed to *Hnf1b*. Haumaitre et al. 2005 [598] postulated that *HNF1B+* cells within the pancreatic trunk epithelium serve as multipotent pancreatic progenitors involved in both endocrine and exocrine development. Early deletion of *Hnf1b* results in a diminished pool of pancreatic MPPs due to reduced proliferation and increased apoptosis [599]. De Vas et al. 2015 [599] further demonstrated that *Hnf1b* plays a pivotal role in the regulatory networks governing pancreatic MPP expansion, maintenance of acinar cell identity, duct morphogenesis, and the generation of endocrine precursors. Of note, Allen et al. 2011 [600] showed that *GATA6* haploinsufficiency is responsible for pancreatic agenesis. Other studies utilizing genome editing techniques have shown that human pluripotent cell lines carrying mutations in both alleles of *GATA6* failed to differentiate into definitive endoderm, leading to a developmental block at the primitive streak stage- a defect that was rescued by re-expressing *GATA6* [601]. Also, Carrasco et al. 2012 [602] reported a decrease in *PDX1+* multipotent pancreatic progenitors' population in the Gata6 mutant pancreatic epithelium. Similarly, earlier studies such as Swift et al. 1998 [603] showed that *MEIS* subclasses (MEIS1/2) create a multimeric complex with *PDX1*, altering its transcriptional activity. In acinar cell lines, the exocrine activity is facilitated by a complex comprising *PDX1, PBX1*, and *MEIS2*. Hence, the presence of *MEIS* encoding genes together with *PDX1* and *SOX9* indicate similarities with early pancreatic progenitors.

6.3.4 Amphicrine progenitor-like sub-state recapitulates invasive cancer stem cell like phenotype

P5-specific amphicrine progenitor-like sub-state revealed a unique cellular phenotype characterized by repression of acinar lineage genes like *RBPJL, GP2* and *NR5A2*, as well as absence of digestive enzyme encoding genes like *PRSS1/2* — indicative of an immature and less differentiated state. This sub-state prominently expressed signatures associated with pancreatoblastoma (PBL) [608], an immature childhood tumor with multilineage features, including *LEF1, LGR5, BMP4, ID1*, and *TCF7,* suggesting a resemblance to PBL oriented molecular profile. Moreover, pathway enrichment analysis indicated the activation of WNT-BMP-NOTCH signaling, which is acknowledged for its role in maintaining stemness and suppressing differentiation in pancreatoblastoma (PBL) [608]. This finding was further corroborated by the enrichment of Wnt signaling mediators such as *CTNNB1, AXIN2, LGR5, DKK4, LEF1, BMP4, RNF43,* and *NOTUM*. In this context, it is noteworthy that the amphicrine progenitor-like sub-state prominently expressed the neurogenesis driver *PROX1*, which is known to be induced by Wnt signaling [609, 610]. Additionally, an observed upregulation of genes linked to invasiveness and cell migration, along with enrichment in ion transport processes, hints at a potentially aggressive and metabolically dynamic nature of this sub-state within the tumor microenvironment. Importantly, a significant fraction of markers overlapped with shared NE HSP+ (hypoxic) sub-state that emphasized the complex interplay of signaling pathways involved in cellular adaptation and tumor progression, highlighting the multifaceted nature of panNEC biology.

6.3.5 Amphicrine acinar-like sub-states reminiscent of pancreatic injury and acinar-to-ductal metaplasia

The amphicrine acinar cell type encompassed three discernible sub-states, identified as Amphicrine acinar01, Amphicrine acinar02, and Amphicrine acinar03. These sub-states were delineated based on key acinar progenitor markers such as *RBPJL, GP2, NR5A2, MECOM*, and *CEL*. *NR5A2*, critical for pancreatic development and exocrine phenotype preservation, demonstrated varying expression levels across the sub-states, suggesting its regulatory role in pancreatic oncogenesis [611]. Notably, Amphicrine acinar-01 exhibited elevated trypsinogen levels and enrichment of *PDX1*-regulated targets, with significant resemblance to a metaplastic acinar program observed in K-Ras-driven models of acinar to ductal metaplasia [604]. Differential gene expression analysis comparing these three sub-states elucidated specific gene profiles in Amphicrine acinar-01, including *CPA1, CPA2, CLPS* and *CTRB1*, compared to the other sub-states, suggesting an intermediate phenotype bridging early and late metaplastic states, as described by Schlesinger et al. 2020 [604]. In this

regard, the amphicrine progenitor-like sub-state showed elevated expression of *ONECUT2*, a marker indicative of acinar to ductal metaplasia (ADM), suggesting a transitional state characterized by lineage dedifferentiation [604]. Similarly, the amphicrine acinar01 sub-state showed heightened expression of *FOXQ1*, associated with metaplastic changes in acinar cells undergoing transdifferentiation [604]. Moreover, both amphicrine progenitor-like and acinar sub-states displayed strong expression of regenerative genes such as *REG1A* and transcription factors including *ONECUT1, HNF4A*, and *HNF4G*, reminiscent of classical PDAC phenotypes [605]. *ONECUT1 (HNF6)* has been proposed to play a role in the development of pancreatic metaplasia lesions [604, 606]. *HNF4A* expression has previously been observed in a subgroup of pancreatic and duodenal NEN, where it was associated with poor survival outcomes [606]. Conversely, shared sub-states exhibit negligible expression of GATA6 and minimal levels of *HNF4A/HNF4G*, suggesting a neuroendocrine reprogramming reminiscent of organoid studies [607]. Amphicrine acinar02 notably expressed serotonin biosynthesis-associated gene TPH1, linked to tumor vascularity in small bowel neuroendocrine tumors [613, 614]. Moreover, ECM remodeling genes expressed in Amphicrine acinar02 hinted at ductal-like features.

This dynamic interplay between lineage plasticity and oncogenic reprogramming underscored the heterogeneous nature of panNEC, with amphicrine sub-states resembling both metaplastic and ductal-like phenotypes, offering potential insights into therapeutic targets and further avenues for investigation. Additionally, the expression profiles of key genes relevant to PDAC and pancreatic neoplasia in amphicrine sub-states mirrored those of MUC5B+ ductal populations, highlighting shared molecular signatures across distinct pancreatic tumor phenotypes [72]. These findings collectively suggest a dynamic interplay of pancreatic lineage markers and metaplastic signatures within amphicrine sub-states, reminiscent of trans-differentiation events observed in pancreatic injury and regeneration models.

6.3.6 Developmental programs and malignant cell state programs in panNEC

- **Resemblance with pancreatic development**

The examination of how lineage plasticity influences the malignant behavior of panNEC involved conducting module score analyses and over-enrichment tests using carefully curated signature lists from studies on pancreatic development, encompassing acinar, ductal, and endocrine lineage signatures [547, 616, 619, 620]. Specifically, the amphicrine progenitor-like sub-state unique to patient P5 displayed enrichment for trunk and endocrine progenitor gene sets, implying its potential to differentiate towards ductal and endocrine cell

lineages [547, 616]. Conversely, amphicrine acinar sub-states exhibited variable expression of the tip progenitor module, responsible for acinar lineage differentiation, with significant enrichment observed particularly in amphicrine acinar01 and 03 [547]. Additionally, amphicrine acinar01 demonstrated significant enrichment for ductal progenitor signatures, indicating a transitional state between acinar and ductal phenotypes, while also showing signatures of centro-acinar cells implicated in oncogenic transformation through ductal transdifferentiation [616, 617]. Furthermore, the exploration of transcriptomic similarities between panNEC sub-states, endocrine progenitors, and various endocrine lineage cell types unveiled distinctive patterns [619, 620]. Module score analysis revealed mutual exclusivity between signatures associated with alpha and beta cell fate programs, with alpha linked to shared panNEC sub-states and beta to amphicrine ones, prominently enriched in amphicrine acinar01 [619]. Notably, key markers associated with mature endocrine cell fates were absent in all sub-states, including *GCG, INS*, and *SST,* alongside critical genes defining the alpha lineage such as *MAFB* and *IRX2* [620]. Furthermore, prominent beta markers like *IAPP* and *MAFA* exhibited negligible expression in the panNEC sub-states [620]. Although *DLK1, TGFBR3*, and *SMAD9* contributed partially to the enrichment of the beta program in amphicrine sub-states, their presence did not indicate a typical beta-like NEC in the absence of INS [620]. Additionally, *SST* and *PPY,* characteristic markers for delta and PP cell types, were not expressed in the panNEC sub-states [620]. The absence of characteristic lineage markers like *GCG* or *INS* confirmed that the panNEC sub-states did not reflect the transcriptomic profile of mature islets, unlike well-differentiated pNET counterparts, which closely resemble the transcriptomic profile of various endocrine cell types [620].

- **Resemblance with lung and prostate neuroendocrine carcinoma**

Neuroendocrine (NE) differentiated cancer sub-states are increasingly acknowledged across various cancer types, notably in lung and androgen-resistant prostate cancer [624-627]. The clinical management of panNEC draws heavily from experiences with NE small cell carcinomas of the lung, which are more prevalent [628]. To evaluate if the malignant programs observed in non-pancreatic NE carcinomas were recapitulated in panNEC sub-states, transcriptomic signatures from lung and prostate NEC types and subclasses were examined using module score analysis and an over-enrichment test. Specifically, transcriptomic signatures from George et al. 2018 [340] study profiling lung neuroendocrine carcinoma (LCNEC) subtypes were examined, revealing distinct enrichment patterns across panNEC sub-states. George and colleagues segregated LCNEC into two subtypes (Type I and II) with distinct mutational profiles, either involving bi-allelic TP53 and STK11/KEAP1 alterations (Type-I) or bi-allelic inactivation of TP53 and RB1 (Type-II) [340]. Type-I LCNEC

showed high neuroendocrine expression akin to small cell lung carcinoma (SCLC) despite sharing a mutational profile with non-small cell lung carcinoma (NSCLC), while Type-II LCNEC resembled NSCLC based on transcriptomic patterns [652, 653]. Moreover, distinct malignant sub-classes (Class D and E) were identified, bridging SCLC and LCNEC categories. The Type-I LCNEC signature was notably enriched in the amphicrine acinar01 panNEC sub-state, aligning with its high neuroendocrine expression and metabolic characteristics. Conversely, Type-II LCNEC signature exhibited significant enrichment in the amphicrine proliferating and shared NE proliferating sub-states, indicative of increased cell activation and proliferation. This enrichment was driven by genes associated with DNA methylation, chromosomal instability, and cancer metastasis, highlighting the biological relevance of these sub-states in pancreatic NECs [629, 630]. Additionally, signatures from George et al.'s 2018 study delineating transcriptomic sub-classes within neuroendocrine lung cancers demonstrated distinct enrichment patterns in panNEC sub-states. Class I and II neuroendocrine lung cancer sub-class signatures were prominently represented in amphicrine sub-states, with Class I moderately represented in the shared neuroendocrine sub-state. On the other hand, shared panNEC sub-states generally lacked expression of the Class II signature, associated with LCNEC with lower-level neuroendocrine marker expression. Notably, SCLC-like transcriptome features (Class III and IV) were strongly over-enriched in both shared neuroendocrine and neuroendocrine-proliferating sub-states, suggesting a shared transcriptomic profile with SCLC rather than LCNEC of the lung. Furthermore, class D and E signatures, linked to chromatin modifications, DNA repair, and damage response, were upregulated in both amphicrine and neuroendocrine-proliferating sub-states, with neuroendocrine-proliferating showing stronger enrichment of class E, derived mainly from SCLC samples, relative to the LCNEC-dominant Class D [340].

Differentially upregulated markers in both NE and NE-proliferating sub-states demonstrated significant over-enrichment of a prostate cancer (PC) signature identified by Alshalalfa et al. 2019 [341]. This PC signature was reportedly effective in distinguishing small cell neuroendocrine prostate cancer from adenocarcinoma in early-stage, treatment-naïve specimens.

Overall, this analysis suggested that there might be common underlying molecular pathways or mechanisms driving the neuroendocrine phenotype across different cancer types. This shared molecular signature could potentially be targeted therapeutically, allowing for the development of novel treatment strategies that could benefit patients with NEC regardless of the primary tumor site. Additionally, it underscored the importance of considering the similarities in molecular profiles when designing clinical trials and treatment approaches for

NEC, as therapies effective in lung or prostate NEC may also hold promise for pancreatic NEC patients. Moreover, identifying shared molecular signatures may aid in refining diagnostic and prognostic markers, leading to more accurate classification and personalized management of NEC patients across different anatomical sites.

- **Resemblance with Pancreatic Ductal Adenocarcinoma**

Previous molecular characterization of panNEC suggests a closer mutational resemblance to pancreatic ductal adenocarcinoma (PDAC) rather than pancreatic neuroendocrine tumors (NET) [324, 636]. Additionally, panNEC can manifest as mixed neuroendocrine-non-neuroendocrine neoplasms (MiNEN or MANEC), featuring both neuroendocrine and non-neuroendocrine components within the same lesion [637, 638]. Inspired by these observations, an investigation was conducted to determine whether any of the panNEC sub-states transcriptomically mirror malignant cell states and lineage programs observed in PDAC. Employing module score analysis using disease signatures compiled from Hwang et al. 2022 [342], it was observed that amphicrine acinar-like 01 exhibited the most pronounced enrichment for the 'acinar-like' program, with intriguing characteristics observed regarding the Classical-like subgroup. This observation was driven by genes like *GATA6* and *HNF4A*. Of note, *GATA6*, a critical transcription factor for maintaining epithelial identity and facilitating acinar cell differentiation, is a major marker for classical PDAC subtype [605]. A recent study by Kloesch et al. 2022 [635] decoded that *GATA6* loss, coupled with concurrent downregulation of *HNF4A*, is essential for transitioning to the basal phenotype in PDAC. Importantly, this loss of *GATA6* has been associated with increased metastatic potential and a more aggressive phenotype.

Both shared NE and NE proliferating sub-states displayed notable enrichment for ‘Neuroendocrine-like’ program, associated with nervous system development and neurotransmitter secretion, while shared NE HSP+ (hypoxic) sub-state exhibited no such enrichment, suggesting a departure from typical neuroendocrine functional traits. Furthermore, the shared NE sub-state demonstrated significant over-enrichment for Neural-like progenitor (NRP) targets, associated with axonal guidance and tumor-nerve interactions, indicating potential implications for disease progression and prognosis [342]. In this regard, Hwang and colleagues reported significant enrichment of the NRP malignant program in residual PDAC tumors and patient derived PDAC organoids post-treatment, which correlated with an unfavorable prognosis [342]. These findings shed light on the potential similarities in the underlying molecular mechanisms driving tumorigenesis and progression between these

two types of pancreatic cancer. Moreover, this emphasized the need for tailored treatment strategies informed by molecular subtyping.

6.3.7 Transcriptional regulation and lineage plasticity of panNEC

The application of single-cell regulatory network inference and clustering (SCENIC) revealed that transcription factors identified in amphicrine sub-states are typically present in healthy pancreas, reflecting their physiological roles in acinar or endocrine cells. The inferred activity of tumoral transcription factors either exceeded or was reduced compared to their physiological activity in corresponding cell populations [72]. Notably, the amphicrine progenitor-like sub-state exhibited differential regulation by transcription factors such as *SOX9, HNF1B, SOX13, ONECUT1, PDX1,* and *LEF1,* suggesting potential roles in pancreatic cancer initiation and stemness maintenance [640, 641]. Furthermore, highly proliferative sub-states demonstrated elevated activity of *E2F* family members, while amphicrine acinar sub-states displayed notable *PTF1A* and *MECOM* activity, which are pivotal in specifying acinar cell identity and regulating pancreatic tumor progression [543]. Strong *HNF4A* activity observed in both sub-states is associated with normal pancreatic development, early pancreatic progenitor gene expression, and PDAC classical subtype, highlighting its relevance of developmental drivers in pancreatic cancer pathogenesis [642, 643]. Overall, the amphicrine progenitor-like sub-state showcased a unique regulatory profile with lower inferred activity for acinar-related transcription factors, consistent with its distinct molecular characteristics within the amphicrine cluster.

The shared sub-states, NE and NE-proliferating displayed distinct activities of transcription factors linked to pancreatic and/or brain organogenesis, including *PAX6, ISL1, SOX5, ETV1*, and *TCF3* [643-656]. Notably, the tumoral activity of these transcription factors in the NE sub-states exceeded their physiological levels in the endocrine pancreas. *PAX6* and *ISL1,* known for their roles in regulating pancreatic islet differentiation and the production of alpha and beta cells, also regulate survival and differentiation of pancreatic endocrine progenitors [643-648]. *PAX6*, a master regulator of neuronal development, was found to be upregulated in NEC relative to healthy pancreatic cell types, suggesting a potential role in pancreatic neuroendocrine tumorigenesis [649-651]. During embryonic development, *ISL1* is known to control motor neuron generation, axon growth and cranial ganglia neurons during brain development [809]. In addition, *ISL1* overexpression was observed in a variety of cancers, including brain, lung, breast, gastric and rhabdosarcoma [810]. With respect to NEC, an immunohistochemical study on large cells and small cell NEC from lung or

gastroenteropancreatic (GEP) origin had reported *ISL1* expression in half of the GEP cases with a trend to higher expression in small cell tumors [811]. The presence of *ISL1* in non-pancreatic NEC including neuroblastomas [812] would likely align with the functional context of neural specific transcriptional programs rather than a beta islet cell specific context. The specific expression of *NKX2-2,* a multipotent neuronal progenitor marker [651, 656], in shared NE and NE-proliferating sub-states, alongside *PAX6* and *ISL1,* suggested a potential endocrine origin for these sub-states. However, further investigations are warranted to elucidate whether *NKX2-2* functions analogously to neural or islet cells in the context of panNEC.

SOX5 belongs to the SOXD family of transcription factors that are strongly involved in cell fate determination during development and previously described to regulate neuronal migration, differentiation and projection [813]. Knocking down *Sox5* has been associated with gene expression changes resembling type 2 diabetes (T2D) and compromised insulin secretion [654]. This was primarily due to diminished mitochondrial activity, reduced expression of L-type Ca2+ channels, and a decrease in depolarization-evoked Ca2+ influx. As intracellular Ca2+ plays a pivotal role in regulating various enzymes, including ATP synthase, the concurrent mitochondrial defect and impaired Ca2+ influx likely synergistically exacerbated the secretory impairment. In recent studies, *SOX5* oncogene has emerged as a crucial factor in driving the metastasis and advancement of diverse tumors, including breast cancer, prostate cancer, hepatocellular carcinoma, and nasopharyngeal carcinoma [814]. Previously, it had been implicated in promoting metastasis and EMT programs in human prostate cancer [814]. Mechanistically, *SOX5* exerts its influence by activating Twist1 through binding to the Twist1 promoter, thereby facilitating the EMT process [814]. In context of human lung adenocarcinoma (LAD), a previous study delineated the role of SOX5 transactivation in promoting chemoresistance by upregulating FOX-A1 gene [815]. Similar to LAD and many other cancers, EMT-driven chemoresistance is a major roadblock for pNEC. Hence, the presented data suggests further investigation of *SOX5* in mediating panNEC malignant programs and repurposing it for therapeutics.

Similarly, the ETS-transcription factor or *ETV1* is known for its involvement in EMT during pancreatic development and was found to be upregulated in mouse pancreatic intraepithelial neoplasia (PanIN) and PDAC [816]. Moreover, *ETV1,* implicated as a master regulator in neuronal-derived gastrointestinal stromal tumors, showed increased expression in panNEC samples compared to healthy pancreatic controls [655]. Previously implicated in promoting neuroendocrine prostate cancer (NEPC), *ZBTB7A* has been associated with a pronounced dependence on RET kinase activity in NEPC cancer cells exhibiting a high small-cell

neuroendocrine (SCN) phenotype score, indicating a strong correlation between *RET* and *ZBTB7A* dependencies [817]. Silencing *ZBTB7A* in a NEPC cell line confirmed its role in supporting cell growth by inhibiting the G1/S transition in the cell cycle and inducing apoptosis, highlighting its oncogenic function [817]. Therefore, further investigation is warranted to explore the potential involvement of *ZBTB7A* in promoting cell proliferation and growth in the shared NE sub-states. *HOXA9* has been identified as capable of stimulating the Wnt/β-catenin pathway in pancreatic cancer cells, promoting stemness properties and self-renewal of cancer cells. Conversely, knockdown of *HOXA9* impedes sphere formation and leads to the downregulation of stemness genes [818]. Another inferred TF in both shared NE and NE-proliferating sub-states is *TCF3*, a member of the Wnt/β-catenin signaling pathway that was previously found to be highly expressed in endocrine progenitors along with *NEUROG3* [620, 819]. In adult human organisms, de novo expression of *TCF3* has been reported in islets from individuals with type II diabetes, suggesting its potential role in reflecting endocrine cell stress/injury [820].

6.3.8 *PTF1A/PAX6* regulated brain targets associated with aberrant neuronal phenotype

Previous studies, utilizing DNA methylation and mutational profile analyses of panNEC, have suggested an exocrine cell origin and a mutational spectrum resembling that of exocrine PDAC, rather than well-differentiated NET [338, 324, 636]. Within this context, the neuroendocrine phenotype could arise from the acquisition of characteristics resembling either neural or neuroendocrine islet cell types. Therefore, this study aimed to investigate whether any of the panNEC sub-states exhibited tissue-specific target enrichment for key developmental factors such as *PTF1A, PAX6,* and *NKX2-2.* These transcription factors play crucial regulatory roles in both pancreas and brain development, each associated with distinct sets of targets. For example, *PTF1A* is essential for regulating the proliferation of multipotent pancreatic progenitor cells and specifying and sustaining acinar cells [547, 590, 591, 595]. Additionally, *PTF1A* also governs brain and neuronal development, particularly the molecular specification of inhibitory neurons and the generation of correctly balanced neuronal circuits [663-665]. However, it's important to note that the SCENIC analysis used in this study does not incorporate tissue-specific target information in its reference database, limiting the assessment of brain versus pancreas-specific enrichment of *PTF1A* and similar factors.

This dissertation leveraged murine tissue-specific developmental signatures from E12.5 neural tube and E17.5 pancreas to explore the regulatory role of *PTF1A* in different cell types.

Target signatures governed by *PTF1A* in both pancreas and neural tube were obtained from Meredith et al. 2013 [666], revealing distinct tissue-specific functions of *PTF1A* during development. The pancreatic target signature of *PTF1A* was notably enriched in the amphicrine acinar 01 sub-state, aligning with SCENIC analysis results. Conversely, shared NE and NE-proliferating sub-states did not express the pancreatic targets module due to its association with genes related to exocrine pancreatic function. Interestingly, the NE proliferating sub-state exhibited significant enrichment of *PTF1A*-regulated brain target signatures, featuring genes implicated in various cellular processes. Notably, HPIBP3, known for its role in drug resistance and cell proliferation, showed over-enrichment in this sub-state, along with KHDRBS1, a regulator of cancer stem cell properties. Additionally, *SSBP3, PRRC2B, THRAP3, ARID1A, LMNB1*, and *NCAM1* were among the genes enriched in this sub-state, each playing crucial roles in various cellular functions [341, 666-675]. Although the NE and NE HSP+ (hypoxic) sub-states showed a slight overrepresentation of *PTF1A* brain signatures, statistical significance was not reached. Nevertheless, these findings underscore the potential role of PTF1A-regulated brain target genes in neuronal dedifferentiation and panNEC progression, emphasizing their functional significance in this context.

Of particular interest, *HP1BP3* is known to directly interact with *EZH2*, and epigenetically activate *WNT7B* that promotes drug resistance in glioblastoma cells [667]. *HP1BP3* overexpression was also reported to result in significantly increased cell proliferation, self-renewal, and promotion of temozolomide (TMZ) resistance [667]. Moreover, splicing factor *KHDRBS1* (aka Sam68) is known to be tightly controlled by c-Myc proto-oncogene, and previously found to be enhanced in several human malignancies, including prostate cancer, where it influences high proliferation and survival [668]. Importantly, *KHDRBS1* is also recognized to regulate human cancer stem cell (CSC) vulnerability by Wnt/β-catenin signaling [669, 670]. Notably, Farini and colleagues showed that *KHDRBS1 (Sam68)* regulates synaptic connections through controlling dynamic splicing [671]. *SSBP3* is a DNA-binding protein encoding gene that is known to co-regulate the function of pancreatic β-cells, often in conjunction with *ISL1* [672]. Additionally, *SSBP3* is recognized for its role in overseeing the morphology of neurons and the processes involving the biogenesis and transport of synaptic vesicles [673]. Of note, RNA processing factor *THRAP3* is known to regulate DNA Damage response (DDR), and previously, mass spectrometry-based proteomics study found that *THRAP3* depletion causes cellular hypersensitivity to DNA-damaging agents [674]. PRRC2B plays a vital role in effectively translating specific proteins necessary for cell cycle advancement and cellular proliferation [675]

The regulatory roles of *PAX6* and *NKX2-2* extend beyond endocrine cell differentiation, encompassing crucial functions in central nervous system development [649-651, 656]. Leveraging GTEX bulk RNA-seq data [679], differential expression analysis between pancreatic and brain tissues was conducted, intersecting *PAX6* and *NKX2-2* target lists from MSigDB [727]. The amphicrine sub-state exhibited varying expression levels for pancreatic targets regulated by *PAX6* and *NKX2-2*, with the NE proliferating sub-state significantly enriched for *PAX6*-regulated brain-specific targets. However, the shared sub-states did not demonstrate significant enrichment for pancreas or brain-specific NKX2-2 targets. This might suggest an autonomous function of *NKX2-2* in panNEC development that is tissue agnostic.

6.3.9 Intra-tumor plasticity and role of EZH2 regulation in panNEC

Immunohistochemistry validations confirmed the protein-level expression of sub-state markers identified through snRNA-seq analysis, underscoring significant variability among individual patients. In Patient P1, Ki-67 staining indicated a moderate proliferation rate of approximately 40% of tumor cells, primarily localized at tumor margins. Moreover, EZH2 protein levels positively correlated with Ki-67 expression, exhibiting the strongest staining at the periphery but negative staining in the center. Conversely, Patient P2 displayed higher fractions of MKI67 (Ki-67) and EZH2 positive cells, corresponding to a more uniform positivity among tumor cells. Notably, in Patient P1, YAP1 expression was confined to the surrounding stroma, while absent in tumor cells, consistent with the enriched stroma sub-state observed in this patient. Conversely, Patient P2 exhibited heterogeneous YAP1 expression within individual glands, aligning with the observed heterogeneity in the snRNA-seq data. These findings highlighted the importance of considering intra-tumoral heterogeneity in characterizing panNEC sub-states and underscore the utility of combining snRNA-seq with immunohistochemistry for comprehensive molecular profiling of tumor samples.

EZH2, a crucial orchestrator of cell fate determination and organogenesis in the pancreas [681, 682] and nervous system [683-686]. Acting predominantly as a transcriptional repressor within the Polycomb Repressive Complex 2 (PRC2), *EZH2* catalyzes histone H3 lysine 27 trimethylation (H3K27me3), thereby silencing genes and regulating diverse cellular processes. In panNEC sub-states, *EZH2* targets showed tissue-specific enrichment, with amphicrine sub-states displaying heightened EZH2 pancreas target scores, and shared NE sub-states exhibiting over-enrichment for *EZH2* brain-specific signatures. Notably, a significant proportion of brain targets mediated this over-enrichment, including genes involved in neuronal system regulation, axon guidance, and synaptic functions. Given that

EZH2 predominantly serves as a transcriptional repressor, the enriched expression of *EZH2* targets within specific panNEC sub-states might indicate a context-dependent attenuation of *EZH2's* repressive function. Previous research, exemplified by Zhang et al. in 2023 [684], has demonstrated profound alterations in gene expression profiles in neural progenitors and post-mitotic neurons following the use of *EZH2* conditional knockout models. Therefore, further experimental investigations are necessary to elucidate the precise regulatory mechanisms of EZH2 in panNEC models, particularly in discerning whether the loss of *EZH2* can lead to the upregulation or downregulation of specific target genes.

Additionally, computational analyses identified several transcription factors, such *as REST, GFI1, TEAD4,* and *NKX2.5*, as potential regulators of *EZH2*-mediated gene expression in panNEC sub-states, with downstream targets implicated in neuronal development and synaptic transmission [689-691]. Furthermore, enrichment analyses revealed associations of *EZH2* brain signatures with neurodevelopmental disorders and cerebellar ataxia, corroborating previous findings linking *EZH2* dysregulation to impaired neurogenesis [694]. These findings underscore the intricate regulatory roles of *EZH2* in panNEC sub-states and highlight its potential implications in neurodevelopmental disorders and cerebellar functions.

6.3.10 Recapitulation of human small intestinal signatures in pNEC sub-states

The exploration of human small intestine-specific cell type/state signatures in panNEC sub-states sheds light on tumor plasticity and potential differentiation trajectories [695, 615]. Notably, the amphicrine sub-states predominantly express Intestinal Stem Cell (ISC) signatures, with the amphicrine progenitor-like sub-state exhibiting strong expression of intestinal Paneth signatures [695]. Additionally, the shared NE sub-state displayed significant expression of small intestinal Enteroendocrine Cell (EEC) signatures, with patient-specific variations observed, particularly in the NE sub-state of patient P3 [695]. Further analysis using STREAM based trajectory analysis revealed intrinsic variability within the P3 NE sub-state, with distinct lineage branches showing enrichment for specific transcription factor targets, such as *SRF, FOXP2, ONECUT2, HOXA7, FOXN4*, *STAT1,* and *CTBP2* [695]. *SRF* has been associated with cell proliferation and migration and is particularly known for its involvement in the control of cytoskeletal and muscle-related genes [696]. *FOXP2* functions physiologically in organ development, e.g. brain, lung, skeletal system [697], in GI tract development as well as the enteric nervous system [698]. *FOXP2*'s role can shift between that of an oncogene or a repressor depending on the context [699, 700]. For example, *FOXP2* acts as an oncogene in multiple myeloma and diffuse large B-cell lymphoma, yet functions as a tumor suppressor in hepatocellular carcinoma (HCC) and gastric cancer [699,

700]. In another study conducted by Chen et al. in 2023 [701], it was shown that *FOXP2* is a direct target of PRC2-mediated gene silencing in malignant lung epithelial cells. Chen and colleagues also observed that the enforced expression of *FOXP2* resulted in heightened stemness and increased migration of lung epithelial cells [701]. *ONECUT2* plays a pivotal role as a mediator in androgen receptor-independent cell growth and acts as a driver for neuroendocrine differentiation during the transition from adenocarcinoma in castration-resistant prostate cancer [702-704]. A prior investigation by Schlesinger et. al. 2020 [604] identified *ONECUT2* as a key marker for advanced metaplastic cells, related it to worse clinical outcomes and implied it as a major driver of early stage PDAC development. Elevated expression of the *HOXA7* gene has been associated with heightened proliferation in HCC via cyclin E1/CDK2 regulation making it a prospective and significant molecular target for the diagnosis and treatment of liver cancer [705].

6.3.11 snRNA-seq provided insights into panNEC therapeutics

The investigation into potential treatments tailored to specific panNEC sub-states involved a comprehensive analysis of enriched pathways and drug activities. Differential pathway analysis unveiled pathways enriched across common sub-states, suggesting therapeutic avenues. Notably, the shared NE proliferating sub-state exhibited heightened expression of cell-cycle checkpoint kinases, *E2F* targets, and DNA repair genes, indicative of susceptibility to platinum-based drugs like Cisplatin [712-715]. Moreover, the elevated expression of *EZH2* suggested potential application of EZH2 inhibitors such as Tazemetostat, GSK343, and GSK926 [708]. Importantly, Gene set enrichment analysis (GSEA) also associated NE proliferating sub-state with Cisplatin, aligning with current treatment strategies.

Of note, the shared NE HSP+ (hypoxic) sub-state emerged as a pathological hotspot, showcasing enrichment in pathways with oncogenic potential, including heat stress response, hypoxia, glycolysis/PI3K-AKT, mTORC, and MAPK signaling, suggesting vulnerable targets across all patient samples. This observation presented an opportunity for repurposing approved drugs targeting these signaling pathways, such as Bevacizumab or Alpelisib, to specifically inhibit VEGF or the PI3K pathway, respectively. Geldanamycin and Tanespimycin emerged as highly significant drugs for targeting NE HSP+ (hypoxic), highlighting therapeutic potential [732]. Notably, dysregulation of *HSP90* levels, encoded by *HSP90AA1* and *HSP90AB1*, suggested repurposing of drugs like Pimitespib (TAS-116), known for its efficacy in gastrointestinal stromal tumors post-chemotherapy [716-718]. *HSP90* inhibitors like 17-AAG have shown promise in clinical trials across various cancers, including their ability to enhance cytotoxic effects of paclitaxel [719-725]. Overall, these findings illuminate potential

targeted therapeutic strategies for managing panNEC by exploiting specific molecular vulnerabilities within the NE HSP+ (hypoxic) sub-state.

6.3.12 Possible developmental origins of panNEC

Integration of panNEC patient samples (P1-P4) with Tosti et al.'s snRNA-seq data revealed a close association of acinar subtypes with shared panNEC sub-states, particularly shared NE stromal-like, indicating greater transcriptomic similarities to acinar rather than ductal or endocrine cell types/states [150, 149]. Diffusion Pseudotime analysis confirmed this observation, revealing a connection between acinar subtypes and panNEC shared sub-states [150]. Trajectory analysis using STREAM unveiled a branch linking acinar-I to shared panNEC sub-states, further supporting the potential origin of panNEC from acinar cells [153, 338]. However, markers identified by Yachida et al. for "acinar-type panNEC" were not detected in the shared sub-states, prompting investigation into potential drivers of acinar to NE differentiation [338]. Transition markers revealed elevated expressions of genes like *AUTS2, CACNA1A, RBFOX1*, and *ZBTB20*, though these markers alone did not fully explain the molecular factors driving this transition [338]. Moreover, it was noted that adult pancreatic snRNA-seq datasets may lack essential cell types or states, such as multipotent pancreatic progenitors and intermediate states crucial for modeling progression to panNEC sub-states, which are more comprehensively captured in developing mouse scRNA-seq datasets [339]. These findings underlined the complexity of panNEC development and the need for further exploration, such as integration of human embryonic or fetal pancreas single cell or nuclei RNA-seq data to elucidate the precise mechanisms underlying its pathogenesis.

7 Methods

7.1 Maternal-fetal interface and eoPE study

7.1.1 Patient recruitment

Tissue sampling was conducted within a multicenter framework. Patients were enrolled from various locations including Berlin (Germany), Graz and Vienna (Austria), Oslo (Norway), London (UK), and Melbourne (Australia). Approval for the studies was obtained from the respective regional committees, with each study being individually described and methodological details outlined. This part was performed in collaboration with Dr. Florian Herse, Dr. Olivia Nonn and Dr. Martin Gauster, and snRNA-seq experiments at BIH Genomics Core facility. Method described below is as per [363].

7.1.2 Sample preparation

- **First-trimester (or early samples) snRNA-seq**

Placental and corresponding decidual tissue were obtained from voluntarily terminated pregnancies with the informed consent of healthy individuals (gestational age 5 – 11 weeks). Exclusion criteria included maternal age below 18, maternal BMI exceeding 25, and self-reported maternal pathologies. Ethical approval was granted by the Medical University of Graz Ethics Committee (31-019 ex 18/19; 26-132 ex 13/14). Following surgical extraction, tissue was promptly stored at 4°C in DMEM/F12 culture medium (1:1 ratio, 1 g/dL glucose, Gibco®, Life Technologies (TM), Thermo Fisher Scientific, Vienna, Austria) and processed within a maximum of 4 hours. The amnion was excised, and the decidua dissected. Villous and decidual tissue were individually rinsed twice in cold (4°C) 0.9% NaCl solution to eliminate blood, then snap-frozen in liquid nitrogen and stored at -80°C until processing. Patient characteristics are elucidated in Extended Data Table 1.

- **Healthy term tissue snRNA-seq**

Samples were obtained immediately after delivery from the inpatient clinic of the Department of Obstetrics and Gynaecology, University Hospital Graz, Austria. The study received approval from the local Ethics Committee at the Medical University of Graz (approval numbers: 31-019 ex 18/19; 26-132 ex 13/14), and informed consent was obtained from all

participating women. Representative tissue samples (1x1x1 cm) were taken from the medial third of the placenta, specifically from vital cotyledons that exhibited no macroscopic signs of infarction or other apparent pathologies resulting from delivery. This approach aimed to minimize the risk of sampling degraded RNA and to ensure a high-quality yield for subsequent analysis, recognizing the potential bias towards inaccurate phenotypes in both disease and healthy samples. The amnion was dissected, and residual tissue was rinsed twice in cold (4°C) 0.9% NaCl solution to remove blood before being snap-frozen in liquid nitrogen and stored at -80°C until processing. Patient characteristics are provided in Extended Data Table 1 and Supplementary Table 1.

- **Early-onset pre-eclampsia (eoPE) and healthy term pregnancies snRNA-seq**

Pregnant women were enrolled at Oslo University Hospital before elective caesarean section following informed written consent, as previously outlined, from individuals with either eoPE or normotensive pregnancies. eoPE was defined as the onset of hypertension (blood pressure ≥140/90 mmHg) and proteinuria (≥1+ on dipstick or ≥30 protein/creatinine ratio) after 20 weeks of gestation but with delivery occurring before the 34th week. Placental villous tissue biopsies were obtained from the central, visually normal appearing cotyledons, snap-frozen in liquid nitrogen, and stored at −80°C until use. The study received approval from the regional committee for Medical and Health Research Ethics in South-Eastern Norway and was conducted in accordance with the principles of the Helsinki Declaration. Patient characteristics are detailed in Extended Data Table 1 and Supplementary Table 1.

7.1.3 From Capture to Sequence: Nuclei Isolation, Library Preparation, and Sequencing Process

Approximately 100-200 mg of frozen placental tissue and corresponding separately sampled decidual tissue were processed using an optimized nuclei isolation protocol developed by Krishnaswami et al. (ref. 6). In brief, the frozen tissue was disrupted using a pre-cooled glass Dounce in homogenization buffer (1X NIM2 [1X protease inhibitor, 1 µM DDT, 250 mM sucrose, 25 mM KCl, 5 mM MgCl2, 10 mM pH 8.0 Tris], 0.4 U/µL RNAseIn, 0.2 U/µL Superasin, 0.1% v/v Triton X-100) and filtered through a flow cytometry tube (BD Falcon) with a 35 µm cell sieve cap. The homogenate was incubated in the dark on ice for two minutes with DAPI (5 µg/µL) and then centrifuged for eight minutes at 1,000xg and 4°C. The resulting pellet was resuspended with staining buffer, transferred to a FACS tube (BD Falcon) with a 35 µm cell sieve cap, and analyzed using the BD FACS ARIA III flow cytometer with the BD FACSDiva software (BD Bioscience). Following FACS sorting with a cut-off at 90%

viable single nuclei, nuclei from the landing buffer (1% BSA, 0.2 U/µL RNAseIn) were counted using a digital counting chamber (Elvira) to achieve a concentration of 400-500 nuclei/µl and were loaded onto 10x Genomics Chromium chips. Single-index v2 and v3 libraries from 10x Genomics were prepared following the guidelines provided by the manufacturer (Chromium Single Cell 3' Kits v2 User Guide – CG00052, Chromium Single Cell 3' Kits v3.1 Dual Index User Guide – CG000315). The libraries were sequenced on an Illumina HiSeq-4000 (paired-ended), aiming for a minimum coverage of 50,000 raw reads per nucleus.

7.1.4 snRNA-seq data preprocessing

- **Matrix pre-processing and quality control**

The Cell Ranger software (versions 3.0.2, 6.0.1 & 6.1.2) from 10x Genomics was utilized to process and identify Unique Molecular Identifiers (UMI) and extract nucleus barcodes from demultiplexed FASTQs of raw 3' snRNA-Seq data. Specifically, the SP014 (10X V2 library chemistry), SP082, and SP136 batches (10X V3 library chemistry) were processed with versions 3.0.2, 6.0.1, and 6.1.2, respectively. The transcripts were aligned to the pre-built human reference genome GRCh38 pre-mRNA version 3.0.0, which was derived from the precompiled GRCh38 reference and tailored for snRNA-Seq data following the protocol provided by 10x Genomics.

The unfiltered feature-barcode matrix per sample was considered for further removal of ambient RNA and technical artifacts. Systematic biases and empty droplets were subsequently addressed by modeling and filtering counts to eliminate contaminated ambient RNA reads and random barcode swapping. This process performed using the remove-background function in CellBender v0.2.0 with model= "full," maintained a default target false positive rate (FPR) of 0.01. The --expected-cells parameter was set to the estimated number of cells from CellRanger. Following evaluation of the CellRanger rank-ordered UMI plot, a threshold of 15,000 --total-droplets-included was deemed suitable for modeling ambient RNA. For three villi samples (18-033-v, 18-098-v, and 20-027-v), --total-droplets-included was set to the default 25,000, resulting in highly consistent numbers of cells after CellBender filtering (16,442, 12,922, and 12,839, respectively). The number of cells per sample after CellBender filtering and corresponding QC metrics were reported in Supplementary Table 3 and 4 for placenta and decidua, respectively.

Thereafter, matrices filtered for ambient RNA(s) were loaded into Python v3.7.9 and processed using scanpy v1.8.2. Post-quantification quality control involved excluding nuclei with fewer than 200 expressed genes or total mitochondrial transcript expression higher than 5%. Only genes expressed in more than three nuclei were included. Data quality was assessed by plotting the number of UMIs and total number of genes per sample. After quality control filtering, samples were log-normalized to 10,000 reads using scanpy. The resulting number of cells and corresponding QC per sample were reported for both placenta and decidua in Supplementary Table 3 and 4, respectively. On average, scanpy filtering removed 5.62% of nuclei after quality check for both tissue samples.

- **Doublet detection**

Detecting doublets using computational algorithms is a challenging task, and there is a lack of universally effective tools for every dataset. Initially, we employed the Scrublet scr.scrub_doublets() function to predict neotypic doublets. This function simulates doublets using observed data and employs a k-nearest-neighbor classifier to compute a continuous doublet score for each nucleus, ranging from 0 to 1. The overall distribution of doublet scores did not exhibit a clear bimodal pattern, suggesting that our dataset was not systematically influenced by potential doublets.

This observation aligned with our expectations, considering that in single "cell" isolation, a significant portion of doublets may result from incomplete dissociation of tissue. However, in single "nuclei" isolation, doublets can arise from nuclei sticking together or the chip being overloaded, leading to multiple nuclei per droplet. This reduction in potential causes of doublets should result in a decreased observed doublet rate. Nevertheless, identified doublets were flagged using a threshold of 0.35 based on the doublet score histogram and were cautiously retained for downstream analysis instead of immediate filtration.

The decision to retain these flagged doublets was motivated by the understanding that droplets appearing to harbor doublets could arise from various factors, such as cellular differentiation or insufficient sequencing. By including them in the analysis, potential doublets could be more comprehensively interpreted in the context of marker expression. Subsequent to cell typing, the doublet analysis was re-evaluated, considering contradictory marker genes that characterize ambiguous cells. The findings were further discussed in the "Evaluation of biological doublets" section.

7.1.5 snRNA-seq data harmonization and clustering

- **Harmonization of data, clustering, and cell annotations for placental analysis**

In the context of placenta samples, data harmonization, clustering, and cell annotations were conducted as follows:

Initially, 6,000 highly variable genes were computed using the scanpy's highly_variable_genes function, employing the dispersion-based method (flavor='seurat_v3') with default parameters. The donor identifier served as the key batch to minimize the selection of batch-specific genes. Subsequently, a Bayesian variational inference model, scVI v0.14.5, was utilized for sample integration. The scvi.model.SCVI and get_latent_representation functions in scVI were employed to infer a shared latent space of 15 dimensions for all placental single nuclei. The model specifications included 128 nodes per hidden layer, 2 hidden layers for both encoder and decoder in variational inference, and a 0.1 drop-out rate. Gene expression was modeled using a zero-inflated negative binomial distribution (ZINB). To account for technical variation in cell typing, additional categorical covariates (10X library chemistry, procurement center of samples, gestational week) and continuous covariates (total counts, total number of genes with at least one positive count, percentage of mitochondrial expression, percentage of XIST expression per nucleus) were incorporated.

The K-nearest neighbor graph was computed on the scVI inferred latent space using the pp.neighbors function in scanpy with $k = 15$ and default parameters. For dimensionality reduction to 2D, Uniform Manifold Approximation and Projection (UMAP) from the umap-learn v0.5.2 implementation in Python was employed, with 500 iterations for improved convergence and a random state of 0 for reproducibility.

Cell-typing (annotations) began with the control placenta samples (both early and late gestation), relying on robust and specific expression of marker genes. Unsupervised clustering using the Leiden community algorithm in scanpy (with an initial resolution limit of 2) was followed by annotation using marker genes from literature and top signatures obtained from Seurat's FindAllMarkers Logistic Regression (LR) method and Empirical Bayes method by the model.differential_expression function in scVI. Clusters that did not exhibit robust or specific biological markers were consolidated. Subsequently, a LR classifier model in Celltypist v0.2.09, optimized by the stochastic gradient descent algorithm, was trained based on control cluster labels and used to predict cell annotations in diseased (eoPE) samples.

The confidence score of the inferred annotation per cell was reported, and a confusion matrix was utilized for performance evaluation.

Misclassified nuclei were excluded based on specific criteria, including the removal of fibroblast and erythroblast subpopulations and the exclusion of nuclei tightly clustering with vCTBp but also expressing high vSTB/dEVT markers. Internal validation of cluster phenotypes was conducted using module scores computed with known marker lists. Finally, clusters assigned to a phenotype were evaluated using robust and specific marker genes, as detailed in the Differential Expression Analysis section, and illustrated in Figure 3.5 & Supplementary Table 7.

- **Harmonization of data, clustering, and cell annotations for decidua analysis**

Similar to the approach taken for placenta, the identification of the top 6,000 highly variable genes in decidua involved utilizing scanpy's highly_variable_genes function with donor_id as the batch key. Initially, cell typing was conducted on the 10X V2 samples, which were sequenced earlier, through the annotation of unsupervised Leiden clusters based on robust and specific marker expression. A shared latent space of 10 dimensions was inferred using the get_latent_representation function in scVI, maintaining the parameters previously employed for placenta. Markers were derived from literature and top signatures obtained from the Bayes-method scVI model's differential_expression function and Seurat's FindAllMarkers() Logistic Regression method. Clusters uniquely mapped to a donor were excluded for comparative cell typing. Subsequently, cell labels were transferred to the 10X V3 samples using scANVI10.

Concurrently, an LR classifier model from Celltypist was trained using annotated cluster labels to predict cell annotations in 10X V3 samples. A confusion matrix was utilized to evaluate the performance of the CellTypist classifier and scANVI (predicted labels). Final decisions for each cluster were made after inspecting them using biological markers knowledge.

One cluster, initially annotated as stromal due to its proximity to the DSC1/2 and consisting of 2911 nuclei, was later excluded due to the expression of conflicting markers such as *NOTUM, HPGD,* and *HLA-G* (indicating EVT lineage) and certain macrophage genes, making classification challenging. The CellTypist LR classifier assigned a very low confidence score (~0), suggesting likely contamination. Another cluster (initially considered as NKT cells; 1119 nuclei) was removed due to high macrophage gene expression. The cell

type/state annotations of decidua were ultimately presented using a set of robust and specific marker genes, as depicted in Figure 3.4 and Supplementary Table 7.

The resulting number of cells and corresponding QC metrics (total UMI counts and the number of genes expressing at least one positive count per nucleus) for each placenta and decidua sample were documented in Supplementary Table 10 and 11, respectively. A total of 5.31% additional nuclei were removed after filtering out donor-specific and ambiguous clusters in both tissues. The final UMAP embedding illustrating cellular hierarchies for decidua and villi is shown in Figure 3.3a & b. Importantly, cell type/state level QC was also evaluated, as presented in Extended Data Figure 1e, f and Supplementary Table 5.

7.1.6 Assessment of clustering robustness

- **Evaluation of biological doublets**

Scrublet predictions of neotypic doublets in the placenta exhibited a bias towards vSTB1. Among the total 2443 identified doublets, Scrublet predicted 1656, 136, and 50 doublets for vSTB1, vSTB2, and vSTBjuv cell states, respectively, accounting for 2.8%, 1.02%, and 1.90% of their respective populations. These predicted doublets were subjected to further investigation. Specifically, for each cluster, marker gene expression was scrutinized, distinguishing between predicted doublets and singlets. This analysis aimed to ascertain whether cross-contamination of marker genes occurred between predicted doublets and non-doublets for the vSTB states, as illustrated in Extended Data Figure 2. The examination revealed that no doublet-centric clusters were identified, and there was no evidence of conflicting marker expression leading to confusion with other cell types/states.

- **LDA analysis**

To ensure the accurate annotation of cell types/states, amortized Latent Dirichlet Allocation (LDA) implemented within scVI was employed to derive topic profiles for both placental and decidual tissues. In essence, each distinct cell type/state was expected to correspond to a unique topic. Subclusters, while sharing topics with the parent cluster, typically exhibited additional unique topics. For instance, dNK1 and dNK2 shared certain topics but also possessed distinct ones. This modeling strategy was not only effective for identifying unique cell types/states but could also be utilized to detect potential doublets in cases where cells demonstrated conflicting topics, often attributed to opposing lineage markers. This

methodology aligns with marker-based approaches utilized in other single-cell studies focusing on placenta and decidua.

The LDA was executed at various stages, initially employing the number of Leiden clusters equal to the number of topics and ultimately extending to the number of final cell type labels. This approach aimed to assess whether the learned topics predominantly influenced cells located in close proximity in the UMAP space. When a topic proved dominant in multiple clusters on the UMAP, it indicated a similarity between clusters despite their spatial separation in the embedding. This occurrence suggested that local relationships were not preserved beyond a specific threshold. By confirming that problematic clusters did not map to unique or known topics, they were subsequently excluded from all downstream analyses. Additionally, this process served as a quality check for the UMAP embedding, and a breakdown of LDA topics is visualized in Extended Data Figure 3.

7.1.7 Biological validation of cell types/states

- **Detection of syncytiotrophoblast sub-clusters**

For the first time, a substantial heterogeneity was observed within the vSTB group using snRNA-seq based approach. Notably, a distinct state, termed juvenile STB, was identified. This state is characterized by the expression of paternally imprinted *DLK1*, a regulator of cell growth and differentiation, along with *SPARC*, *TMSB10*, and *ACTB*, suggesting an association with extracellular matrix remodeling and the promotion of changes to cell shape. Intriguingly, the juvenile nuclei exhibited robust expression of the secretory phenotype, characterized by several *PSGs* and maternally imprinted *TFPI2*. Simultaneously, they displayed a classical STB-like profile, marked by the expression of *CGA, CYP19A1, KISS1, ADAM12, SDC1*, and others. Consequently, this unique state was classified under the vSTB group.

- **Other placenta cell types/states**

vCTBp was recognized as the trophoblast progenitor due to their active cycling, evidenced by the expression of genes such as *MKI67, TOP2A, STMN1*, and *CENPK/CENPE.* Notably, these cells exhibited robust expression of *YAP1, TEAD1, TP63, CCNA2*, and *ITGA6,* all well-known for their roles in the progenitor cell phenotype. On the other hand, vCTBpf primarily displayed a fusogenic profile and was characterized by specific markers including *GREM2,*

ERVFRD-1, ERVV-1/2, OTUB2, and *DYSF.* Placental F13A1+/FGF13+ resident macrophages, also known as Hofbauer cells (vHBC), uniquely expressed the hyaluronan receptor LYVE1 in this immune cell subset, suggesting a role in maintaining arterial tone and possessing pro-angiogenic functions. Additionally, we identified antigen-presenting *HLA-DRA+* placenta-associated maternal monocytes/macrophages (vPAMM), which were villi-associated and of extra-embryonic or maternal origin. A cycling population of vHBC (vHBCp) was identified, characterized by traditional HBC genes and the expression of proliferative genes such as MKI67 and TOP2A. Villi myocytes (vMC) were distinguished by their AGTR1 expression. Other cell types were mapped based on well-known marker genes (refer to Supplementary Table 7). Appropriate references for markers and associated cell types/states were provided in the text.

7.1.8 Differential expression analysis for inferring cell type/state markers

Cell-type marker analyses were conducted for both decidua and placenta using a multivariate logistic regression (LR) generalized linear model, implemented in Seurat's FindAllMarkers(). Internal validation was further performed using the empirical Bayes method within the scVI model differential expression function.

In the LR analysis, the number of unique molecular identifiers (UMI), the number of genes, and the percentage of mitochondrial transcripts per nucleus were utilized as continuous covariates. Additionally, categorical covariates such as ~disease (indicating whether a nucleus is from a control or eoPE sample) and library type (10X V2 or V3 chemistry) were incorporated to minimize the effects of eoPE and library variations. Only genes exhibiting a log-fold change cut-off of 0.25 and expressed in at least 25% of cells within each cluster were considered as significant cell markers. An adjusted p-value cut-off of 0.05 was set following Bonferroni correction for multiple testing of states.

7.1.9 Integration of publicly available placenta scRNA-seq data with our snRNA-seq data

The 10X single-cell RNA sequencing (scRNA-seq) data from Pique-Regi et al. 2019 [353] were harmonized with our own 10X single-nucleus RNA sequencing (snRNA-seq) data to pinpoint differentially expressed preterm gene sets per cell type/state. These gene sets were

subsequently employed to correct for gestational age or preterm effects in the analysis comparing early-onset preeclampsia (eoPE) vs term controls.

Data integration was carried out solely on relevant samples: eoPE and late term control samples from this study and preterm and term samples from the Pique-Regi study, separately for villi and decidua samples. Due to the absence of pre-processed count matrices from the Pique-Regi study, raw FASTQ data were obtained from dbGaP accession phs001886.v4.p1, and feature-barcode count matrices were generated using CellRanger 6.1.2. The results exhibited high concordance with a mean difference of only 43.9 cells compared to the original study, all with highly similar unique molecular identifier (UMI) and gene counts (refer to Supplementary Table 8). Subsequently, data harmonization and label transfer were performed using scVI/scANVI, based on Bayesian variational inference. The donor_id (each sample) served as the batch key, and additional categorical covariates (dataset: scRNA-seq vs snRNA-seq, 10X library chemistry: 10X V2 vs 10X V3, procurement center of samples: Oslo, Graz & Detroit, condition: preterm vs term, fetal sex) and continuous covariates (total counts, total number of genes with at least one positive count, percentage of mitochondrial expression, XIST counts per nucleus) were included to minimize the influence of batches and technical variation in cell typing. Projection of cells into UMAP embeddings revealed effective batch mixing for conserved cell types/states, such as vCTB, as depicted in Figure 2.19. Integration was evaluated using adjusted rand index (ARI), adjusted mutual information (AMI), and cell-type-specific absolute silhouette width (ASW) per batch. Specifically, ARI and AMI were 0.051 and 0.118, respectively, indicating good integration for the placental clusters (refer to Supplementary Table 09). Mean ASW (scaled between 0-1, where 1 signifies perfect integration) per donor was 0.803, with fairly high scores per cell type (Supplementary Table 9). For the decidua, the ARI and AMI were 0.03 and 0.12, respectively; mean ASW per donor was 0.826, indicating appropriate integration (Supplementary Table 9). After data integration, differential gene expression analysis was conducted while accounting for batch effects. For two cell-state/class populations of interest (preterm vs term controls), differentially expressed genes were identified using scvi.model.SCVI.differential_expression() in the 'change' DE mode and adjusted for covariates specified during integration. The list was filtered using false discovery rate (FDR) < 0.05 and Bayes factor > 3. Term controls for each cluster were downsampled to match the cell numbers in preterm controls to avoid composition-specific effects. Subsequently, unique preterm genes for each conserved cell type/state were identified and reported in Supplementary Table 10. Preterm genes were not identified in vFB, vMC, and vVEC cell groups due to extreme sparsity in preterm control groups. For the decidua, the dNK1/2 and dMAC1/2 subclusters were merged into dNK and dMAC classes for preterm gene calculation.

Preterm genes were not identified for dVEC, dLEC, and dLECp owing to sparsity of these populations in the preterm control group. The preterm genes for both placenta and decidua cell types/states were reported in Supplementary Table 10. After identifying preterm genes, a preterm score per cell type/state was computed, and ultimately, eoPE-associated genes were identified as described below.

7.1.10 Differential analysis of eoPE disease markers and gestational age correction

To identify differentially expressed genes associated with eoPE compared to late controls within specific cell types/states, the logistic regression (LR) framework was employed, utilizing Seurat's FindMarkers function. Continuous covariates, including the number of unique molecular identifiers (UMI), gene counts, percentage of mitochondrial transcripts, and percentage of sex-specific transcripts per nucleus, were incorporated into the LR model. Of particular significance, a cell type/state-specific preterm score was computed using the significant genes differentiating preterm vs term groups (determined through a microarray approach for vSTB groups and scVI-integrated Pique-Regi et al. data for other cell types/states, as detailed in the previous section). This preterm score served as a continuous covariate in the LR model, explicitly designed to mitigate pronounced preterm-specific effects in the analysis, considering the earlier onset of eoPE by 6-8 weeks prior to healthy term.

Genes with a log2 fold-change cut-off of 0.25 and expressed in at least 10% of cells within each group were deemed significant, based on an adjusted p-value < 0.05 (Bonferroni corrected). Both up- and down-regulated genes were determined. Notably, no significant compositional shifts were observed in cell types/states associated with eoPE relative to term controls, except for vHBC. Therefore, down-sampling was exclusively performed for vHBC. For certain cell types, such as vCCT, dEVT, vCTBp, vCTBpf, dDSTB, dPC, and dBcells, no analysis was conducted due to extreme sparsity in the eoPE group. It's worth mentioning that none of our samples exhibited confounding effects from major co-occurring diseases.

7.1.11 Reconstruction of developmental trajectories and pseudotime analysis

To deduce the cluster and lineage relationships among various trophoblast cell types/states, STREAM v1.1 (https://github.com/pinellolab/STREAM) and diffusion pseudotime were employed. The trajectory inference focused on early controls of trophoblast cell types,

including vCTBp (progenitor), vCTB, vCTBpf, vSTBjuv, vSTB1/2 and vCCT, considering the notable discrepancy in the late term controls' cell-type composition. Initially, scVI-harmonized control data were subsetted for the relevant cell types to learn the trajectory principal graph using STREAM 1.1. Utilizing previously computed latent variables, cells were clustered in the reduced UMAP space to recover the main and possibly finer structures of trophoblast differentiation. The principal graph was then inferred on the manifold learned from the dimension_reduction function, using the first six components. K-means clustering facilitated initial graph seeding through seed_elastic_principal_graph(). The resulting elastic principal graphs, structured data approximators, featured each cell as a vertex interconnected by edges. The graph's inference relied on a greedy optimization procedure, constructing a minimum spanning tree (MST) using Kruskal’s algorithm. No branch pruning or shifting of nodes was performed to obtain the optimal principal graph.

Subsequently, transition and leaf markers were computed for all lineage paths (vSTB, vCTB, and vCCT), considering MKI67-positive vCTBp as the root node (start of the pseudotime). Transition genes, dynamic in nature, were calculated based on the fold change in the average gene expression of the first 20% and the last 80% of cells along an individual branch, considering the inferred pseudotime. For genes with a log2 fold change cut-off of 0.20, Spearman’s rank correlation was calculated between pseudotime and gene expression along individual branches. Genes surpassing a predetermined correlation threshold (set at =0.35) were identified as transition genes. For leaf gene detection, z-scores of all leaf branches were calculated based on average normalized gene expressions. Kruskal–Wallis H-test followed by a post-hoc pairwise Conover’s test was used for multiple comparisons, with a z-score cut-off of 1 and a p-value cut-off of 0.01. Highly robust cell fate markers along the pseudotime provided validation for trajectories (refer to Figures 2.14 & 2.15 for vCCT and Figure 2.17 & 2.18 for vSTB lineages). The lists of leaf and transition genes are provided in Supplementary Table 16. To further evaluate lineage relationships and global transcriptomic similarity among different trophoblast cell types, diffusion pseudotime analysis (Haghvardi et al., 2016) was performed. This analysis orders cells based on their transcriptomic similarity in a Markovian space, considering each cell as represented by a Gaussian wave function. Diffusion distances were based on a robust connectivity measure between cells estimated over all possible paths of a certain length between the cells. The eigenfunctions of the Markovian transition probability matrix, referred to as diffusion components (DC1 and DC2), were used to create a low-dimensional representation and visualize trophoblast data (Extended Figure 4). Additionally, a force-directed graph based on the Fruchterman-Reingold algorithm was presented (Extended Figure 4).

7.1.12 Receptor-Ligand Interaction Analysis

- **Choice of Databases**

The choice of the database significantly influences the outcomes of Receptor-Ligand (R-L) interaction studies. We utilized two widely recognized databases, CellChatDB and FANTOM5, enabling the identification of established interactions, such as MIF-ACKR3/CXCR7 and INHBA-ENG/END, exclusive to CellChatDB and FANTOM5, respectively.

- **Differential Analysis of Receptor-Ligand Interactions between eoPE and term Controls**

Differential analysis of ligand-receptor interactions was conducted using Connectome v1.0.1, specifically employing the differential connectomics pipeline. In the context of the maternal-fetal interface, the approach involved using secretory ligands exclusively for vSTB groups. These ligands, practical in crossing the maternal-fetal barrier and interacting with maternal blood in the decidua, were selected. Ligands that showed significant upregulation in eoPE compared to term controls, with a log2 fold change cut-off of 0.25 and detected in at least 10% of diseased cells, were included in the analysis. Assuming that an activated ligand would bind to a receptor, Multivariate Logistic Regression was employed for differential calculation, maintaining consistency with the previously described Differentially Expressed Genes (DEG) test and accounting for covariates. The resulting figure was visualized in Figure 2.27b.

For within-tissue interaction mapping (decidua and villi interaction) involving immune and endothelial cell types, LR and Connectome were applied. For both upregulated and downregulated differential candidates, a log2 fold change cut-off of 0.25 and ligands/receptors detected in at least 10% of diseased cells were taken into account. The p-value was adjusted for covariates, as outlined in the eoPE vs late term DEG analysis. Visualization was achieved through circos plots.

- **Analysis of Receptor-Ligand Interactions in eoPE**

The analysis focused on decidual Syncytiotrophoblasts (dDSTB) and Extravillous Trophoblasts (dEVT) ligands in conjunction with maternal Vascular Endothelial Cells (VEC) and decidual Smooth Muscle Cells (dSMC) specifically within eoPE samples. Interactions

were derived using Connectome, utilizing both FANTOM5 and CellChatDB databases. A minimum cell count per identification was set at 75, and Diagnostic Odds Ratio (DOR) was calculated for each interaction pair, with a high DOR indicating high specificity and sensitivity along with a low rate of false positives and false negatives.

For dDSTB interactions, the interaction list was refined based on sender and receiver percentages, ligand expression values as well as DOR.source. Stringent criteria were applied to ensure specificity and sensitivity (Figure 2.27a). In interactions with dEVT, rigorous criteria, including DOR.source, edge strength, and minimum percentage of ligand-expressing source, were used to ensure cell-specific communication. The resulting Receptor-Ligand figure is provided in Extended Data Figure 8.

- **Validation of Key Receptor-Ligand Interactions through Computational Methods**

All Connectome results were cross-checked using SingleCellSignalR for interactions involving vSTB, dDSTB, and dEVT. Subsequently, additional tools (NATMI, logFC Mean, CellphoneDB, CellChat) within the LIANA framework were employed to recapitulate all R-L interactions across multiple databases.

7.1.13 Spatial transcriptomics using 10X Visium

Kerim Secener is duly acknowledged for performing the 10X Visium sample preparation and helping me with data analysis. Visium method is as per [363].

- **Sample preparation**

The initial trimester tissue was obtained as detailed previously, dissected under a stereomicroscope, and promptly frozen by immersion in isopentane within a liquid nitrogen bath. To mitigate potential large batch effects, numerous sections of placental tissue were incorporated into a 6.5x6.5 mm cryo-mould utilizing OTC cryo-embedding medium (TissueTek). Subsequently, the samples were stored at -80°C overnight and cryo-sectioned at -20°C. To assess the morphological integrity of the embedded tissues, standard H&E staining was conducted. The resulting 10 µm cryo-sections were then transferred onto spatial transcriptomics slides (Visium, 10x Genomics) and positioned within a single tissue optimization and gene expression slide capture area.

- **10X Visium data sequencing**

After establishing an optimal permeabilization duration of 18 minutes in accordance with the tissue optimization protocol (10x Genomics – CG000238 Rev A), the gene expression experiment was conducted following the manufacturer's user guide (10x Genomics – CG000239 Rev A). The acquired images were then scanned using the Slide Scanner Pannoramic MIDI (3DHISTECH) with the objective plan-apochromat 20×/0.8× (Zeiss).

Following this, the dual-indexed Visium library was loaded at a concentration of 200 pM and underwent sequencing on a HiSeq-4000 (Illumina) with the specified configuration: 28-10-10-90. Please refer to the sequencing requirements for Visium Spatial Gene Expression (refer to https://support.10x_genomics.com/spatial-gene-expression/sequencing/doc/specifications-sequencing-requirements-for-visium-spatial-gene-expression).

- **Spaceranger count matrix generation**

The base call (BCL) files obtained from the Illumina run to FASTQ reads were converted using bcl2fastq (Illumina). Afterwards, the reads were aligned to the human reference dataset GRCh38 (build 2020-A; refdata-gex-GRCh38-2020-A) using the spaceranger count pipeline (Space Ranger v1.1.0), incorporating automatic fiducial alignment and tissue detection. A total of 1,387 spots were identified within the tissue, yielding an average of 201,176 reads and 3,561 median genes per spot.

- **Data processing of Spaceranger count matrix**

This step involved reading the 10x output folder using the Load10X_Spatial function implemented in Seurat (v3). The object was then normalized with the SCTransform function. Principal Component Analysis (PCA) was performed using RunPCA, computing 50 principal components and selecting the first 20 for identifying the k-nearest neighbors of each spot with the FindNeighbors function. Subsequently, clustering was carried out using FindClusters with a resolution of 0.2.

- **Visium data deconvolution using NMF/NNLS**

Spotlight-based deconvolution of the Visium data was accomplished by generating a spotlight object using the spotlight_deconvolution function in SPOTlight (Version 0.1.7), with the early villi subset (from the placental single-nuclei data) serving as the reference. The marker table for the nuclei clusters, initially generated based on the Logistic Regression method implemented in Seurat as discussed earlier, was then filtered to obtain the most

representative topic profile for each cell type in the dataset. Non-negative matrix factorization (NMF - nsNMF) regression, as well as Non-negative Least Squares (NNLS) regression, were employed for deconvolution in SPOTlight. Spots contributing to less than 1% of the spot composition were excluded (min_cont = 0.01). The deconvoluted spots were assessed by examining the topic profiles of the cell type (Extended Data Figure 1f) and the characteristics of individual topics within a cell type.

7.1.14 In-situ sequencing (ISS)

Dr. Katja Sallinger and Sebastian Tiesmeyer are duly acknowledged for performing ISS experiments (library preparation, imaging, hybridization & sequencing) and extensive data analysis respectively. The method summarized here is according to [363].

- **Library preparation**

Fresh tissue samples from early villi were subjected to formalin-fixed paraffin-embedding (FFPE) processing and stored at +4°C. A customized gene panel was employed to identify specific genes associated with cell types and pathways of interest. The in-situ sequencing method was conducted following the manufacturer's instructions (Cartana, a part of 10x Genomics). Tissue sections of 5 µm thickness were baked at 60°C for one hour, deparaffinized in xylene, rehydrated in 100% and 70% ethanol, and permeabilized using citrate buffer (pH 6) for 45 minutes at >95°C in a steamer. The sections were then dehydrated in an ethanol series from 70% to 100% and air-dried using Secure Seal (Grace Biolabs, Bend, United States). Gene-specific chimeric padlock probes were introduced, directly hybridized to the RNA at 37°C in an RNAse-free humid chamber overnight and ligated at 37°C for 2 hours. The circular oligonucleotide structures (padlocks) derived from ligation were amplified overnight at 30°C. To minimize RNA degradation during tissue processing, 0.1% v/v diethyl pyrocarbonate (DEPC) was added to all buffers and reagents not provided by the manufacturer.

- **Imaging**

Imaging was conducted using a digital slide scanner (Olympus SLIDEVIEW VS200) connected to an external LED source (Excelitas Technologies, X-Cite Xylis). Fluorescence filter cubes and wheels were outfitted with a pentafilter (AHF) featuring excitation ranges of 352-404 nm, 460-488 nm, 542-566 nm, 626-644 nm, and 721-749 nm, and emission ranges of 416-452 nm, 500-530 nm, 579-611 nm, 665-705 nm, and 767-849 nm. Additionally, single

cube filters (Kromnigon) including SpectraSplit 440, SpectraSplit 488, SpectraSplit Cy3, and SpectraSplit 594 were utilized. Images were captured using an sCMOS camera (2304 × 2304, ORCA-Fusion C14440-20UP, 16 bit, Hamamatsu) and an Olympus Universal-Plansuperapochromat 40× objective (0.95 NA/air, UPLXAPO40X). To prevent signal crosstalk, the pentafilter was utilized to capture DAPI, Cy5, and AF750 signals, while single cubes were used for AF488 and Cy3 signals. The imaged regions were recorded to facilitate repetitive cycle imaging. After imaging, the labeling mix was removed from each slide by adding three rounds of 100% formamide for 1 minute, followed by a washing step.

- **ISS Hybridization and Sequencing processes**

The ISS steps were iterated six times using six distinct adapter probe pools, each subjected to imaging in five channels (DAPI, FITC, Cy3, Cy5, AF750). Following the removal of adapter probes, hybridization of adapter probes was performed at 37°C for 1 hour in an RNAse-free humid chamber. The sections were then washed, and sequencing probes were allowed to hybridize at 37°C for 30 minutes in an RNAse-free humid chamber. After washing, the sections underwent dehydration in an ethanol series, air-drying, and were finally mounted with SlowFade Gold Antifade Mountant (Thermo Fisher Scientific). Library preparation protocols were fine-tuned for placental tissue using high (MALAT1) and low (RPLP0) control probes before the implementation of the final probe panels. To account for autofluorescence, a background without any adapter probe pool was imaged in six channels for subtraction.

- **Spot calling analyses**

The imaging data were analyzed using a custom pipeline provided by CARTANA, which manages image processing and gene calling. All code was developed in MATLAB, and an additional CellProfiler pipeline (v.2.1.1) was employed, incorporating the ImageJ plugins MultiStackReg, StackReg, and TurboReg as previously detailed. In summary, TIFF images from all sequencing cycles were aligned to the general stain of library preparation and segmented into smaller images. The median intensity of all RCP signals for each channel was computed using another CellProfiler pipeline (v.4.0.7). This value was utilized to normalize RCP signal intensities for each channel to a pixel intensity of 10,000. The multiplication factor value acquired for each channel was incorporated into the CellProfiler pipeline, and background subtraction was performed for each sequencing cycle to minimize tissue autofluorescence. A pseudo-anchor was created for each cycle by combining the four readout detection probe channels into a merged image. This pseudo-anchor aided in a more accurate secondary alignment. Reusable counting particles (RCPs) from the labeling mix

were identified, and their x and y coordinates were recorded along with the measured fluorescence intensities. The maximum intensity value in each sequencing cycle was designated as a positive event and utilized for decoding in MATLAB. To visualize the signals, the selected transcripts were plotted onto an image stained with DAPI.

- **ISS data handling**

The management of in situ sequencing data was executed using the plankton.py v0.1.0 package (https://github.com/HiDiHlabs/planktonpy; acknowledgement: Sebastian Tiesmeyer) using Python 3.10.4. In the comprehensive analysis procedure, the ISS data was presented as decoded spots with x and y coordinates of all identified mRNA molecules, each paired with an associated gene label. A total of three in situ sequencing slide scans were examined (106KS, 107KS, 156KS). The 156KS (early control) slide included genes from the custom placenta/cell typing panel designed to identify cell and tissue types. Both the 106KS (late control) and 107KS (eoPE) slides comprised genes from the custom/pathway panel, which was tailored to analyze cell state and metabolic activity. The cell typing sample was collected during the early stage of pregnancy. To visualize the detected mRNA molecules in their histological context, matching DAPI stains of each sample slide were pre-processed. This involved transforming them to grayscale, normalizing the color values between 0 and 1, and enhancing the low-exposure areas by raising all values to the power of 0.4.

- **Cell type specific markers in the placenta panel**

The analysis of placenta cell typing data aimed to provide context to the major cell types identified through snRNA-Seq analysis. Genes from the cell typing panel were considered as markers for specific cell types, particularly vCTB, vSTB, and vHBC. These cell types were chosen for further spatial analysis and plotting due to their robust marker coverage and their significance as spatial landmarks in villi anatomy, where vSTB(s) and vCTB(s) form the layered walls, and vHBC(s) constitute distinct, compact cells in the intra-villous matrix.

To quantify the affinity of each gene to a specific cell type, a gene-cell-type affinity measure was derived from the molecule counts in the snRNA-Seq dataset for vCTB, vSTB, and vHBC. This was achieved by contrasting molecule counts in cells belonging to a specific cell type against an opposing set of cells, utilizing the score_affinity() implemented in plankton.py framework. For each analysis, three distinct sets of cell types were delineated as follows: (i) vCTB(s) vs. vSTB(s) for vCTB(s); (ii) vSTB(s) vs. vCTB(s) for vSTB(s); and (iii) vHBC(s) vs. all other clusters for vHBC(s). To ascertain definitive cell type markers, vCTB(s) and vSTB(s)

were compared against each other to eliminate potential shared trophoblast marker genes. The mean molecule counts of each gene in the two contrastive cell type sets were calculated, and the logarithm of the ratio of these mean count indicators served as a score for a gene's affinity to a specific cell type. A threshold of 0.5 was applied to assign color labels to each gene in both analyses for visual clarity during plotting. Genes exceeding this affinity score threshold were identified as markers for CTBs (*ASPM, ATAD2, BRIP1, CD24, CDH1, CENPE, DIAPH3, FBN2, KANK1, SEMA6D, TIMP3*), STBs (*ADAMTS20, CGA, CYP19A1, ENTPD1, KISS1, KLRD1, LEP, LINC00474, PAPPA2, PLAC4, PLXDC2*), and HBCs (*CD163L1, CD36, F13A1, FGF13, LYVE1, MEF2C, SPP1*), while the remaining genes were categorized as 'other.'

- **Spatial analysis of mRNA molecules**

Utilizing the plankton.py's run_umap function, we constructed a weighted neighborhood graph based on the 800 nearest neighbors of each molecule. The neighbors were weighted using a Gaussian probability density function (PDF) with a bandwidth of 9 µm, approximately covering the area of a single cell and its immediate environment. Subsequently, we generated a model of the local mRNA distribution for all genes by summing the weights of each gene's molecules. To introduce regularization, each distribution's value for the gene of its molecule of origin was increased by 1.15. A 2D embedding that reflects recurring spatial context was obtained by applying the umap-learn v0.5.3 (python implementation of umap algorithm) to the local distributions. The UMAP algorithm was configured with 24 neighbors, a minimal distance of 0.2 between points in the embedded representation, and a Euclidean distance metric. The algorithm was initiated with a random state of 42. The resulting gene-cell type associations were utilized to create a cell type visualization plot for the early placenta sample 156KS. All molecules were depicted as a scatter plot overlaid on the greyscale renderings of the DAPI stain. Molecules with a cell type association affinity score exceeding 0.5 were colored accordingly, while the remaining molecules were rendered in grey.

- **Detection of villi wall**

After establishing the credibility of spatial information in our in-situ sequencing data through cell-typing analysis, our experimental strategy necessitated a subsequent comparative examination of pathways between a late control and an eoPE sample. The pathway categories chosen for this subsequent experiment included (genes *IDO2, TEK, CDH5, KDR, ZEB1*), senescence (genes *INHBA, MMP11*), and trophoblasts (genes *LGR5, MET, FGFR2*). Spatial analysis was confined to the densely populated and well-structured villi walls in both

samples, representing the most organized part of the tissue. Identification of villi walls involved employing a basic edge detection algorithm on the corresponding DAPI signal. Villi walls were distinctly marked by dense nucleation. A greyscale rendering of the DAPI stain underwent smoothing with an optical Gaussian filter with a 2 µm bandwidth. The villi walls were extracted using Scikit-image's (v0.19.2) feature.canny() implementation of the canny edge detection algorithm (sigma value of 3.7) on the smoothed image. Molecules within a 5 µm radius of any point on the detected edges were considered part of the villi walls, while others were excluded from further analysis. The wall filter algorithm was visualized by plotting the underlying DAPI stain in matplotlib's violet-blue 'magma' color scheme. The identified edges from the second step of the wall filter algorithm were superimposed on the stain as orange lines. In the bottom-right half of the plot, the existing mRNA molecules were plotted, with colors indicating their classification as wall or not-wall members.

- **Spatial localization of vascular and senescence markers**

To visualize the spatial relationship between senescence and vascularization, all 'wall' molecules were plotted on a black-and-white rendering of the DAPI stain, color-coded based on their gene assignments for 'vascularization' (red) and 'senescence' (yellow), while other molecules were represented in white. A visual examination of the dispersed senescence and vascularization markers suggested that the topography of senescence markers exhibited more structure in the control sample compared to the eoPE sample, with reduced senescence marker expression around vascularization clusters in the tissue.

To statistically assess this observation, villi wall molecules were categorized into two groups based on their expression locations: (i) a vessel-proximal category containing molecules within regions of 5 µm of another vessel marker; and (ii) a vessel-distal category containing the remaining molecules. A null hypothesis was formulated, positing that the gene distributions should be equal within the two categories. The p-value indicating the deviation from this null hypothesis was determined for each gene using a binomial test. Scipy's (v1.8.1) stats.binom.cdf() function was utilized, with parameters 'p' defined as the overall percentage of 'proximal' molecules, 'k' representing the gene-specific count of proximal molecules, and 'n' indicating the total count of molecules of the respective gene in the sample. The sorted p-values for all genes in the pathway sample were presented in a vertical bar graph, with bars colored according to their membership in the categories 'senescence,' 'vascularization,' 'trophoblast,' or 'other' **(Figure 3.28).** The p-values for senescence, vascularization, and a control category of 'trophoblasts' were extracted and plotted per sample as scatters on a vertical line. The scatters for both samples were juxtaposed for visual comparison **(Figure 3.28).**

7.1.15 Immunohistochemistry (IHC)

Placenta tissue sections fixed in formalin and embedded in paraffin (FFPE) underwent deparaffinization using standard procedures. Antigen retrieval (AGR) was carried out in citrate buffer at pH 6 for 40 minutes using a microwave oven. Following a wash with TBS/T, sections were treated with Hydrogen Peroxide Block (Epredia, Netherlands) to neutralize endogenous peroxidase, and subsequently blocked with UltraVision Protein Block (Epredia). Primary antibodies were appropriately diluted in antibody diluent and applied to the sections for 45 minutes at room temperature. After washing with TBS/T, the UltraVision LP HRP Polymer Detection System (Epredia) was employed as per the manufacturer's instructions. The polymer complex was visualized using AEC (AEC substrate kit, Abcam, UK), and the sections were counterstained with hematoxylin before being mounted with Kaisers glycerin gelatine (Merck, Germany). Scanning of the slides was performed using an Olympus VS200 slide scanner. Prof. Berthold Huppertz and Lena Neuper are duly acknowledged for providing IHC validations for dDSTB sub-state.

Target	Clone	Species	Dilution	Company (cat.#)
GDF15	-	rabbit	1:1000 (1st Trim) 1:250 (Term)	Sigma-Aldrich, #HPA011191
HLA-G	4H84	mouse	1:6000	BD Biosciences, #557577
βhCG	-	rabbit	1:1000	Thermo Scientific, #RB-059-A
PAI-1	PR17272-21	rabbit	1:1000	Abcam, ab182973

7.1.16 Spatial proteomics

Placenta tissue sections (5 µm) fixed in formalin and embedded in paraffin (FFPE) were affixed to PPS FrameSlides (Leica). After standard deparaffinization, antigen retrieval (AGR) was performed in the incubator using Pepsin solution for 10 minutes at 37°C. Following this, sections were washed with PBS/T and treated with Ultra V Block for 10 minutes at room temperature (RT). For double staining, primary antibodies were combined, diluted in antibody diluent, and applied to sections overnight at 4°C. Subsequently, slides were washed with PBS/T and exposed to secondary anti-mouse or anti-rabbit antibodies for 30 minutes at RT. Finally, slides were washed, and tissue sections were mounted with SlowFade Diamond

Mounting media containing DAPI (Invitrogen), followed by the acquisition of fluorescence micrographs.

Target	Clone	Species	Dilution	AGR	Company
b-CG	5H4-E2	mouse	1:300	Pepsin	Abcam #ab9582
E-Cadherin	24E10	rabbit	1:200	Pepsin	Cell Signaling, mAb #3195
anti-rabbit IgG (Alexa Fluor 488)		goat	1:300		Life Technologies
anti-mouse IgG (Alexa in donkey Cy3)			1:200		Jackson Immuno Research

Regions of interest were isolated via laser microdissection (LMD) on a Leica LMD7 microscope, utilizing a 20x objective in fluorescence mode. Approximately 50,000 µm² of each sample was collected into 384-well plates (Eppendorf #0030129547). Following LMD, tissue samples underwent processing for bottom-up LC-MS-based proteomics, following a recently described protocol with slight modifications. Specifically, 4 µl of 60 mM triethylammonium bicarbonate (TEAB, Sigma #T7408) was added to each well, briefly centrifuged (2,000xg, 1 min), and the plate was heated at 95°C for 60 min in a thermal cycler (Biorad's S1000 with 384-well reaction module) at a constant lid temperature of 110°C. Subsequently, 1 µl of ACN was added to each well (20% final concentration), and the plate was heated again at 75°C for 60 min in the thermal cycler. Samples were briefly cooled to room temperature, and 2 µl of pre-diluted LysC (Promega) in ultra-pure water to 2 ng/µl was added, followed by digestion for 4 h at 37°C in the thermal cycler. Next, 2 µl of pre-diluted trypsin (Promega Trypsin Gold) in ultra-pure water to 2 ng/µl was added, and incubation occurred overnight at 37°C in the thermal cycler. The following day, digestion was halted by adding trifluoroacetic acid (TFA, final concentration 1% v/v), and samples were vacuum-dried (approx. 60 min at 60°C). Ultimately, 4 µl of MS loading buffer (3% acetonitrile in 0.2% TFA) was added, followed by vortexing the plate for 10 seconds and centrifuging it for 5 minutes at 2,000xg. Subsequently, samples were stored at -20°C until LC-MS analysis.

- **LC-MS analysis**

Liquid chromatography mass spectrometry (LC-MS) analysis was conducted using an EASY-nLC-1200 system (Thermo Fisher Scientific) coupled to a trapped ion mobility spectrometry quadrupole time-of-flight mass spectrometer (timsTOF SCP, Bruker Daltonik GmbH, Germany) equipped with a nano-electrospray ion source (Captive spray, Bruker Daltonik GmbH). The autosampler was configured to pick up samples from 384-well plates. Peptides

were loaded onto an in-house packed HPLC column measuring 20 cm in length and having a diameter of 75 µm. The column was packed with 1.9 µm ReproSilPur C18-AQ silica beads obtained from Dr. Maisch GmbH, Germany. Peptide separation was carried out using a 32-minute gradient at a flow rate of 250 nL, during which the concentration of buffer B (0.1% formic acid, 90% ACN in LC-MS grade H2O) increased gradually up to 60%. Buffer A, comprising 3% ACN, 0.1% formic acid in LC-MS grade H2O, was used as the mobile phase. The total duration of the gradient was 44 minutes, while the column temperature was maintained at a constant 40°C.

Mass spectrometric analysis was conducted in data-independent (diaPASEF) mode using the default method for extended gradients, with a cycle time of 1.8 seconds. Ion accumulation and ramp times in the dual TIMS analyzer were set to 100 milliseconds each, and the ion mobility range was examined from 1/K0 = 1.6 Vs cm-2 to 0.6 Vs cm-2. The total m/z range was configured from 100 to 1,700 m/z. Collision energy was adjusted linearly with increasing mobility, commencing at 59 eV at 1/K0 = 1.6 Vs cm-2 and decreasing to 20 eV at 1/K0 = 0.6 Vs cm-2. Singly charged precursor ions were filtered out using a polygon filter (timsControl software, Bruker Daltonik GmbH).

- **Proteomics data analysis**

Proteomics measurements were analyzed using timsControl software (Bruker Daltonik GmbH, v. 3.1). For diaPASEF measurements, raw files were processed with DIA-NN (v. 1.8) in library-free mode, relying on a predicted human spectral library (Uniprot 2021 release). Default settings were applied with minor adjustments, such as a mass range set to 100 – 1,700 m/z, precursor charge state set at 2 - 4, and a maximum of 2 allowed miscleavages. MS1 and MS2 mass accuracies were set to 15 ppm, and the match-between-runs option was activated. The quantification strategy was set to ‘Robust LC’. For subsequent data analysis, we utilized the protein FDR filtered pg.matrix.tsv and unique.genes.matrix.tsv outputs from DIA-NN, which were further analyzed using Perseus (v. 1.6.15.0) and the Protigy package from Broad Institute (v. 1.0.2, https://github.com/broadinstitute/protigy).

Missing values underwent imputation based on a normal distribution (with a width of 0.3 and a downshift of 1.8) following rigorous data filtering, which necessitated 70% quantified values across samples. Prior to principal component analysis (PCA), batch effects were rectified using the proBatch R package (version 1.10.0) employing the ComBat method (https://doi.org/10.3929/ethz-b-000307772). Pathway enrichment analysis was conducted with the clusterProfiler R package (version 4.2.2, https://doi.org/10.1089/omi.2011.0118). Dr.

Fabian Cosica and Jose Nimo are duly acknowledged for performing the proteomics experiments and data generation. The method summarized for spatial proteomics is as per as [363].

7.1.17 Data analysis softwares

The alignment and pre-processing of single-nucleus RNA sequencing (snRNA-Seq) data were conducted using Cellranger versions 3.0.2, 6.0.1 & 6.1.2. Additionally, ambient RNA and background noise correction were carried out using CellBender 0.2.0, followed by data processing using scanpy 1.8.2 in Python 3.7.9. Data harmonization was performed using scvi-tools 0.14.5, and UMAP computation utilized umap-learn 0.5.2. Trajectory analysis was executed using stream 1.1 and scanpy 1.8.2, while marker analysis utilized Seurat 4.0. Cell-cell interaction analyses were conducted with Connectome 1.0.1 and LIANA 0.1.4. Analysis and visualization of transcription factor regulatory networks were conducted using STRING, iRegulon, and Cytoscape 3.8.2, with visualization supported by igraph 1.3.2, circlize 0.4.15, dplyr 1.0.9, ComplexHeatmap 2.10.0, seaborn 0.10.0, and python-igraph 0.7.1. General operations utilized scikit-learn 1.0.2, statsmodel 0.12.1, scipy 1.5.3, pandas 1.1.4, and numpy 1.19.4. For the analysis of 10X Visium data, Spotlight 1.0.0 was employed. In situ sequencing (ISS) analysis utilized Python 3.10.4 and Jupyter 1.0.0, with data handling facilitated by plankton 0.1.0 (utilizing pandas 1.4.3). Plot generation utilized matplotlib 3.5.2, and snRNA-Seq data integration was performed using scanpy 1.9.1. Image analysis for villi wall detection was conducted with Scikit-image 0.19.2, and spatial model building, and nearest neighbor analysis utilized scikit-learn 1.1.1. All algebraic operations on matrix representations of the data were executed with numpy 1.22.4, while statistical model building during pathway analysis utilized scipy 1.8.1. Spatial proteomic analyses were conducted utilizing timsControl software (version 3.1, Bruker Daltonik GmbH), DIA-NN 1.8, Protigy R 1.0.2, proBatch R 1.10.0, clusterProfiler 4.2.2, and Perseus 1.6.15. Pathway enrichment analysis was conducted with clusterProfiler R 4.2.2 packages. For conditional regression model analyses and visualizations, R 4.1.2, magrittr 2.0.2, Matching 4.9-11, tidyr 1.2.0, survival 3.2-13, and pROC 1.18.0 were utilized. Software versions from ISS analysis and spatial proteomics were secured from Sebastian Tiesmeyer and Dr. Fabian Cosica, respectively (with permission).

7.1.18 Data analysis collection

No software or tools were utilized for sample data collection.

7.1.19 Data and code availability

The raw single-nucleus RNA sequencing (snRNA-Seq) data from the 33 villi and decidua samples generated in this study have been deposited in the European Genome-Phenome Archive under accession number EGAS00001005681. Due to the sensitive nature of sequencing data, access to the data would be made available under controlled access. Interested parties can request access by contacting the appropriate Data Access Committee listed for each dataset in the study. Access will be granted to both commercial and non-commercial entities in accordance with patient consent forms and data transfer agreements. Images of the in-situ sequencing (ISS) data are accessible via Zenodo (doi: 10.5281/zenodo.5243240), while the snRNA-seq and Visium data can be found on Zenodo (doi: 10.5281/zenodo.8159511) with controlled access. Other data are either included within the thesis, Extended Data Figures or Tables, or Supplementary Tables. Scripts used for snRNA-seq and 10X Visium data analysis and figure generation are available at https://github.com/Olivia117/Maternal_fetal_interface_atlas2022 (with a copy at https://github.com/HiDiHlabs/preeclamspsia_Nonn_etal/tree/main for the pre-print).

7.2 panNEC snRNA-seq study

7.2.1 Patient cohort and ethics approval

Because of the rarity of the disease, a retrospective study design was employed, allowing for strict sample inclusion criteria. Inclusion criteria were histopathological confirmation of G3 NEC of large cell morphology, tumor localization in the pancreas and availability of both flash frozen and FFPE tumor tissue. A cohort comprising six patients diagnosed with high-grade GEP-NEN was curated from two ENETS Centers of Excellence: the University Hospital Charité Berlin, Germany, and the University Cancer Institute of the Inselspital and the University of Bern, Switzerland. Samples were reviewed and (in the case of older samples) reclassified according to the WHO 2019 criteria by a board-certified pathologist (Prof. Aurel Perren), as detailed in Table 4. TNM staging information is given (for inferring tumor, nodes, or metastases) according to the 8th edition UICC/AJCC guidelines. Ethical approval for the study was obtained from local authorities (Ethics approval EA1/229/17 at Charité and KEK-BE 105/2015 from Kantonale Ethikkomission Bern), in compliance with the Declaration of Helsinki. Patient samples were anonymized for the current analysesBoth formalin-fixed paraffin-embedded (FFPE) and fresh frozen tumor tissues were obtained. FFPE sections underwent H&E staining and immunohistochemical detection of chromogranin A,

Synaptophysin, Insulinoma-associated protein 1, Pancytokeratin, Trypsin, Ki67, TP53, Rb1, and Sox9, utilizing routine clinical diagnostic procedures available at the Institute of Pathology at Bern University. Sample preparations for snRNA-seq were conducted from fresh frozen tissues at three different time points: the first run involved P1 as a pilot for technical feasibility, the second run included samples P2-P5, and the third run consisted of a repetition of the P5 preparation due to poor quality observed in the initial preparation, which was performed in two batches (technical replicates).

7.2.2 panNEC sample preparation and snRNA-seq

Tumor tissue preserved by cryopreservation was disrupted using NP-40 lysis buffer (10mM Tris-HCl at pH 7.4, 10mM NaCl, 3mM MgCl2, 0.01% NP-40, 1mM DTT, 2% BSA, 1U/µl RNAse inhibitor, and Complete EDTA-free protease inhibitor) in a 1.5ml Eppendorf tube with a plastic pestle. Following a 5-minute incubation on ice, the suspension underwent filtration through a 70 µm pre-separation strainer, followed by centrifugation at 4°C and resuspension in nuclei wash buffer (PBS, 1% BSA, and 0.4U/µl RNAse inhibitor). DAPI (3.5 µl) was added, and the mixture was filtered through a 40 µm Flowmi cell strainer before sorting on an Aria Fusion instrument (BD) into an Eppendorf tube containing 200 µL of sort buffer (PBS, 2% BSA, 1U/µl RNASe inhibitor) using a 100 µm nozzle.

Single-nuclei libraries were prepared following the protocol outlined in the Chromium Next GEM Single Cell 3' Reagent Kits v3.1 (Dual Index) user guide (CG0003154) from 10x Genomics. Briefly, individual nuclei were encapsulated alongside gel beads containing barcoded primers using the Chromium controller. mRNA within the droplets was reverse transcribed into barcoded cDNA and subsequently amplified. The amplified cDNA underwent fragmentation, end-repair, A-tailing, adaptor ligation, and sample index PCR. The resulting library was sequenced using a NovaSeq 6000 system (Illumina) to achieve a depth of 40,000 mean reads per nuclei.

7.2.3 panNEC snRNA-seq data analysis

- **Data pre-processing and QC**

Single-nucleus RNA reads were aligned to the GRCh38-2020-A reference using CellRanger v6.0.1 using exonic and intronic (nuclear pre-mRNA) reads. One biological sample was removed from further downstream analysis as it did not pass basic Cellranger QC criteria. P5 was run as two technical replicates and later merged during analysis. Background reads

were eliminated using the R package SoupX [100], with parameters set to tfidfMin = 0.6, soupQuantile = 0.9, and priorRho = 0.1. The resulting expression values, after background subtraction, were rounded to integer values. Cells were filtered based on criteria such as having less than 300 or more than 4000 RNA features, exceeding 5% mitochondrial read content, or genes with fewer than two reads across all samples. MT and ribosomal genes were removed from further downstream analysis to better adjust for technical noise. Additionally, for joint analysis with the snRNA data from Tosti et al. 2020 [72], all samples were downsampled to a median read count of 500 reads for validating major results.

- **Data integration and clustering of panNEC patient samples**

The SoupX corrected anndata matrices per patient were merged using scanpy [98], followed by log-normalization for subsequent analysis. Highly variable genes (n=4000) were determined using sample ID as a batch key for lightweight batch correction and default parameters in scanpy's highly_variable_genes() function. Subsequently, principal component analysis (PCA) was performed using tl.pca(), employing svd_solver="arpack" implemented in scanpy. This function inferred PCA coordinates, loadings, and variance decomposition, which were then visualized using sc.pl.pca_overview() and heatmaps to display top features associated with each principal component. The calculation of K-nearest neighbor (KNN) graph and UMAP visualization using default parameters (by fine-tuning the number of principal components and number of neighbors) resulted in strongly separated patient samples or batches.

To attain data harmonization and mitigate batch effects across 5 panNEC samples, a sequential integration approach (also, known as chain integration) was applied. Initially, Harmony [110] (implemented within scanpy external's harmony_integrate function) was run using sample ID as the batch key. This function utilizes the Python version of Harmony, called harmonypy, to harmonize single-cell data stored within an AnnData object. Since Harmony operates by batch correcting the PC(s), it was run after PCA but before computing the neighbor graph. Harmony adjusted PC1 vs PC2 graph indicated that P5 was distinct from the other four patients, with this segregation was primarily influenced by acinar lineage-specific genes such as *GP2*, encoded by PC1. However, such information is largely biological and not a technical artifact (as P5's immunohistochemistry confirmed trypsin positive and hence, similarities with acinar). Subsequently, BBKNN was employed to construct a batch-corrected mutual neighbor graph, using sce.pp.bbknn(). The parameters used were neighbors_within_batch=5, trim=80 and use_rep='X_pca_harmony' to consider Harmony adjusted PC(s) for constructing further batch balanced mutual neighbor graph. The

Leiden algorithm was then run with a resolution parameter set to 1 for clustering and inferring cell sub-states. In cases where robust and specific markers were lacking, clusters were merged. Also, leiden-12 having n=540 nuclei was removed from further downstream analysis, given its ambiguous transcriptomic profile.

Of note, case-studies were performed by adjusting the number of PC(s) from n=10 to 50 and observed changes in data structure were noted. The choice of PC=10 was considered somewhat suboptimal and yielded a clumsy structure. Importantly, when using a higher number of PCs (40 & 50), a distinct but ambiguous cluster with contradictory marker profile (leiden-12 consisting of n= 540 nuclei, originating from P5) was clearly separated on the UMAP plot. An improvement in batch mixing, especially among patient samples P1-P4, was noticed, as evidenced by relatively higher ASW scores per cell type/state with an increase in the number of PCs to 40 or 50. Hence, BBKNN's default PC=50 was considered for final analyses. UMAP was computed using specific random_state=0 (for reproducibility) and maxiter=200 (for better optimization of the embedding layout and ensures convergence of algorithm). Initially, Leiden cluster 1 was labeled as "Stromal/Mesenchymal" containing 8133 nuclei. However, upon further sub-clustering at a medium resolution (set at 0.4 for Leiden clustering), it was divided into two distinct groups: Stromal-like NE (n= 5668 nuclei) and Stromal (normal) (n=2465 nuclei).

Initially, coarse-grained cell types were determined through Leiden clustering at a resolution of 0.3. However, this initial classification was later fine-tuned by consolidating sub-states that exhibited distinct markers for amphicrine, neuroendocrine, stromal, and immune cells, respectively. Although all sub-states specific to P5 were combined under the label "Amphicrine," the Amphicrine progenitor-like group remained distinct due to its strong and specific expression of markers linked to pancreatic progenitors, genes associated with cancer stemness, and other signatures related to Wnt-Notch-BMP signaling. This observation was verified by re-running Logistic Regression marker analysis and Wilcoxon rank sum test, implemented in scanpy's tl.rank_genes_groups() using corr_method='benjamini-hochberg.'

- **Sub-state marker analysis**

Markers analysis of panNEC sub-states was conducted using a multivariate logistic regression (LR) generalized linear model, implemented in Seurat's FindAllMarkers() function. In the LR analysis, the number of unique molecular identifiers (UMI), the number of genes, and the percentage of mitochondrial transcripts per nucleus were employed as continuous

covariates. Significant cell markers were identified among genes showing a log2 fold change cut-off of 0.25 and expression in at least 25% of cells within each cluster. An adjusted p-value cut-off of 0.01 was applied following Bonferroni correction for multiple testing of states. Internal validation was additionally conducted using the Negative binomial method implemented in Seurat.

For comparative analysis among specific sub-states of amphicrine acinar groups, Wilcoxon rank sum method implemented in scanpy's tl.rank_genes_groups() was used and verified using Logistic Regression. Statistically significant sub-state markers were identified based on log2 fold change cut-off of 1.5 and expression in at least 25% of cells within each cluster. An adjusted p-value cut-off of 0.01 was applied using Benjamini-Hochberg adjustment for multiple testing (default correction method in scanpy).

- **Inference of transcriptional regulation**

The pySCENIC tool [158] was employed to forecast transcription factor activities on the downsampled data. The analysis utilized motifs from version 9 of the motif database (moitfs-v9) and binding sites were identified within the human genome version hg19, with sequences located 500bp upstream, as well as 5kb and 10kb centered at the transcription start site (TSS). Differences in transcription factor activity between groups were assessed using the FindMarkers function from the R package Seurat. This was carried out by integrating transcription factor marker activities, estimated by pySCENIC, as a new assay, followed by the application of a Wilcoxon rank sum test.

- **GTEX bulk RNA-seq analysis for inferring brain vs pancreas DEG(s)**

Bulk RNA-seq count data matrix and sample annotations were acquired from the GTEX portal (https://gtexportal.org/home/downloads/adult-gtex/bulk_tissue_expression, file version bulk-gex_v8_rna-seq_GTEx_Analysis_2017-06-05_v8_RNASeQCv1.1.9_gene_reads.gct.gz). Briefly speaking, DESeqDataSetFromMatrix() was used to construct a DESeqDataSet object (dds) from count data matrix (expr_mat) and sample metadata (colData). A comparison of pancreas and brain tissue expression was performed using a likelihood-ratio test facilitated by the R package DESeq2, specifically using fitType = "glmGamPoi", full = ~ SMTS, reduced = ~ 1, test = "LRT." The fitType=glmGamPoi feature in DESeq2 employed a generalized linear model (GLM) with a gamma-Poisson distribution to analyze count data in RNA-Seq studies. This model is adept at handling overdispersion often encountered in RNA-Seq experiments. Genes whose

absolute value of the log2FoldChange column < 500 and where log2FoldChange was not missing (NA) was tabulated as differentially expressed gene (DEG). Furthermore, a cut-off of logFC of 2 & adjusted p-value < 0.001 were considered to derive mutually exclusive pancreas and brain specific DEG lists.

Subsequently, both DEG lists were intersected with *PAX6* and *NKX2-2* target lists secured from MsigDB (https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/PAX6_TARGET_GENES & https://www.gsea-msigdb.org/gsea/msigdb/human/geneset/NKX2_2_TARGET_GENES respectively) to derive brain and pancreas target lists. For *PTF1A* pancreas and brain targets, the signature lists were secured from Meredith et al. 2013 [666]. *EZH2* target list was sourced from ENCODE transcription factor targets (https://maayanlab.cloud/Harmonizome/gene_set/EZH2/ENCODE+Transcription+Factor+Targets).

- **Module score analysis and hypergeometric test**

All module score analysis performed as a part of this study was computed using sc.tl.score_genes(), as implemented in scanpy. Precisely, the module score was calculated by subtracting the average expression of a specific set of genes from the average expression of a reference set of genes. The reference set is randomly selected from the gene_pool for each expression value bin. All external signature lists used for module score analysis were tabulated in Supplementary Table 25 (given reviewers' access).

Furthermore, an over-enrichment test was conducted based on the cumulative distribution function of the hypergeometric distribution, as explained in https://systems.crump.ucla.edu/hypergeometric/index.php. Precisely, the overrepresentation is depicted by the representation factor, calculated as the number of overlapping genes divided by the expected number of overlapping genes from two independent groups. A representation factor greater than 1 suggests higher overlap than expected, while a factor less than 1 suggests less overlap. A representation factor of 1 denotes the expected overlap for independent gene groups. This representation factor can be calculated using the formula:

k = number of overlapping genes between two groups (also, denoted as number of successes)
s = number of genes in group 1 (for example, number of transition genes)
M = number of genes in group 2 (for example, number of DEGs)

N = total genes expressed or tested.
C(a,b) is the number of combinations of a things taken b at a time.
The representation factor = x / expected number of genes. Where, expected no of genes = (s*M) / N

7.2.4 panNEC data integration with adult pancreas snRNA-seq

The first step involved integrating the panNEC snRNA-seq data with Tosti's adult pancreas snRNA-seq data [72]. The exclusion of stromal (normal) and immune cell types from the panNEC data was conducted. Importantly, cell types derived from a common multipotent progenitor during embryonic development, namely, acinar, and ductal cells in the exocrine compartment and alpha, beta, gamma, and delta cells in the endocrine compartment, were included from the adult pancreas data. Since the adult pancreas data was dominated by the acinar-i, acinar-s and ductal cell types (n= 42968, 31570 & 20371 respectively), downsampling of the data was contemplated to avoid compositional bias during data integration. In particular, the mentioned cell types were randomly downsampled to n=2444 nuclei (the median value derived from the distribution of panNEC sub-state nuclei numbers). Next, highly variable genes were computed for each dataset individually, using sc.pp.highly_variable_genes() in scanpy, with batch_key = 'Sample ID' and default parameters. Genes deemed variable in at least 3 samples were selected for subsequent analysis to mitigate sample-specific biases. Following that, panNEC and adult pancreas anndata(s) were concatenated using join= “outer” to preserve the union of expressed features in both datasets. Thereafter, PCA was conducted using sc.tl.pca(), followed by an examination of each principal component (PC) using sc.pl.pca_overview() and the plotting of heatmaps displaying the top 10 positive and negative features associated with each PC. The data harmonization was achieved using the previously described chain integration method. Initially, Harmony was executed using sce.pp.harmony_integrate() with Sample ID serving as the key batch. Subsequently, the PC(s) adjusted by Harmony were fed into the BBKNN algorithm, which generated the batch balanced neighbor graph. Precisely, BBKNN was run using sce.pp.bbknn() and internal parameters batch_key= 'PatientID', neighbors_within_batch=5. Of note, the number of (corrected) PCs were fine tuned between 10 to 30, but the acinar and NE groups started segregating from PC=20 onwards. This splitting got severe from PC=25 onwards and was difficult to explain biologically. The result with PC=10 was shown as Extended Data Figure 15.

As P1-P4 exhibited tight clustering during panNEC integration and clustering analysis, primarily showing neuroendocrine-like characteristics, we endeavored to integrate these

shared sub-states with adult pancreas data. Firstly, P5 sample was excluded from the analysis and the chain integration steps (described above) were rerun. Importantly, the number of principal components (PCs) was adjusted from n=10 to 30 to ensure the stability of data structures. Nevertheless, beyond PC=14, both the acinar and NE groups exhibited division into multiple subparts, indicating a technical rather than biological origin. Moreover, such partitioning proved challenging to justify using biological markers. Guided by the ranked variance plot of PCs and UMAP visualization of integrated data, PC=13 was ultimately selected for the final analysis (Figure 5.21). Also, for the UMAP, min_dist=0.3 and maxiter=200 was used for better optimization of data structure and convergence of algorithm.

Following that, trajectory analysis was conducted using STREAM, wherein the UMAP coordinates of integrated data were utilized to learn the elastic principal graph. While seeding the principal graph using st.seed_elastic_principal_graph(), K-means clustering and n_cluster was set to 10. Principal graph learning was performed using st.seed_elastic_principal_graph(). The trajectory was visualized in Figure 5.22a. No overbranching or pruning of branches were performed by fine-tuning internal parameters. Transition marker analysis were performed using st.detect_transition_markers() using S0 as a root node and logFC cut-off was kept at 0.25. The positive markers for S0S3 branch (increasing as a function of pseudotime) was visualized as Figure 5.22b.

To confirm the close proximity and continuum between the neuroendocrine cell type and adult acinar states, diffusion pseudotime analysis using sc.tl.diffmap() was employed. This unsupervised algorithm revealed that the NE group aligns with acinar states, while the adult endocrine cell types mainly diverged. This behavior was verified by inspecting combinations of top 6 DC pairs, such as DC1 vs DC2, DC2 vs DC3 or DC1 vs DC5.

7.2.5 panNEC Immunohistochemistry (IHC) validations

All the IHC markers were repeated on freshly cut tissue blocks and re-evaluated by an NEN expert pathologist (Prof. Aurel Perren). For immunohistochemistry, the paraffin-embedded material was cut into 2.5-µm-thick serial sections and then deparaffinized, rehydrated, and antigens were retrieved with an automated immunostainer (Bond RX, Leica Biosystems). Antigen retrieval was performed in a Tris-EDTA buffer or Citric acid buffer for generally 30 min at 95 °C for selected candidates (please refer to the table below for antibody description and other details). The primary antibody incubation lasted 30 min at the specified dilutions. Visualization was carried out using a Bond Polymer Refine Detection Kit (Leica, #DS9800) (RRID: AB_2891238); DAB (3,3′-Diaminobenzidine) was the chromogen. Slides

were counterstained with hematoxylin. An automated slide scanner Panoramic 250 (3DHistech) at 20x magnification was utilized to capture scans, and images were acquired with QuPath software.

Antibody	Supplier	ID	Dilution	Antigen retrieval	Temperature	Time
CgA	CellMarque	138M-94 LK2H10	1:400	Tris-EDTA	100°C	30 min
Ki-67	Dako	M7240	1:200	Tris-EDTA	95°C	30 min
RB1	BD Pharmingen	554136 G3-245	1:200	Citric buffer	100°C	30 min
SOX9	Cell Signaling	82630 T D8G8H	1:100	Tris-EDTA	95°C	30 min
SSTR2A	BioTrend	SS-8000-RM UMB-1	1:50	Tris-EDTA	100°C	30 min
Synaptophysin	Novocastra	27G12	1:100	Tris-EDTA	100°C	30 min
TP53	Dako	M7001 DO-7	1:800	Citric buffer	95°C	20 min
Trypsin 1	Chemicon	MAB1482	1:20000	proteinase K solution	37°C	5 min
EZH2	Cell Signaling	5246	1:400	Citric buffer	100°C	30 min
INSM1	Santa Cruz	Sc-271408	1:200	Tris-EDTA	95°C	40min
Pan-Cytokeratin	DAKO	M3515	1:400	Tris-EDTA	95°C	30min
YAP1	Abcam	Ab52771	1:200	Tris-EDTA	95°C	30min

7.2.6 panNEC data analysis softwares

The alignment and pre-processing of single-nucleus RNA sequencing (snRNA-Seq) data were conducted using Cellranger version 6.0.1. Additionally, ambient RNA and background noise correction were executed using SoupX 1.5.2, followed by data processing using Seurat 4.1.1 on R version 4.1.3 (2022-03-10). Python based analyses were performed using scanpy 1.8.2 in Python 3.7.9. anndata 0.7.8 & annoy 1.17.0 was used. General python operations used scikit-learn 1.0.2, statsmodel 0.12.1, scipy 1.5.3, pandas 1.1.4, numpy 1.19.4 and pytables 3.6.1. Data harmonization was performed using harmonypy 0.0.5

(python portal of Harmnoy), bbknn 1.5.1 and UMAP computation utilized umap-learn 0.5.2. Trajectory analysis was executed using stream 1.1 and scanpy 1.8.2, while marker analysis utilized Seurat 4.0.0 & 4.1.1. Conversion of anndata to Seurat was performed using SeuratObject 4.1.0. pySCENIC 0.11.2 was used for transcription factor inference. Additional analysis and visualization of transcription factor regulatory networks were conducted using STRING, iRegulon, and Cytoscape 3.8.2. DESeq2 1.34.0 was used for bulk RNA-seq data analysis. R-based data visualization was performed using png 0.1-7, pheatmap 1.0.12, ggplot2 3.3.6, cowplot 1.1.1, and sp 1.4-7. Furthermore, data processing were facilitated by packages such as dplyr 1.0.9, magrittr 2.0.3, naturalsort 0.1.3, data.table 1.14.2, writexl 1.4.0, and readxl 1.4.0. iPython 7.31.1 and jupyter-client 7.1.2 were used for notebook based analysis.

7.2.7 panNEC Data and code availability

The raw single-nucleus RNA sequencing (snRNA-Seq) data from 5 patient samples will be deposited in the European Genome-Phenome Archive, after submission of manuscript (in preparation) based on this thesis results. Given the sensitive nature of sequencing data, data access will be controlled. All raw and processed data files are stored and managed internally at Eils-HPC, which will be shared with the doctoral committee upon request. Data analysis scripts will be made available at https://github.com/Olivia117/panNEC_analysis2024.

8 Extended Summary & Conclusion

The advent of single-cell and single-nuclei sequencing technologies has revolutionized our capacity to explore cellular characteristics in both healthy and diseased states across various species. Collaborative initiatives, frequently within extensive international consortia such as the Human Cell Atlas are actively engaged in mapping the multitude of cell types and states present in human tissues. Analyzing single cells or nuclei in disease contexts yields comparative insights that hold the potential to deepen our comprehension of pathophysiological mechanisms.

Both early-onset pre-eclampsia (eoPE) and pancreatic neuroendocrine carcinoma (panNEC) represent significant clinical challenges necessitating in-depth molecular investigations. In both cases, single-nuclei sequencing, and advanced informatics analytics were leveraged to dissect the disease biology at a cellular level. Both studies uncovered aberrant expression of markers, transcriptional regulators, signaling programs and cellular interactions indicative of pathogenesis. Interestingly, both diseases exhibit heterogeneity and developmental plasticity, underscoring the complexity of their biology.

8.1. Summary: Maternal-fetal interface in healthy and eoPE pregnancies

The first results chapter of this dissertation reported the creation of a comprehensive reference atlas of the maternal-fetal interface, encompassing first-trimester, healthy term, and early-onset pre-eclamptic pregnancies. Leveraging deep learning models such as scVI/scANVI, distinct cell types or states comprising the cellular landscape of placenta and decidua were identified. Also, this study decoded the transcriptional heterogeneity and functional significance of distinct sub-states within villous syncytiotrophoblast (vSTB) nuclei, including the novel vSTBjuv state, which offered valuable insights into placental development and maternal-fetal syncytium interaction. However, further experimental validations are imperative to fully grasp the biological implications of these sub-states. Despite rigorous computational efforts to address gestational age disparities between preterm eoPE samples and term controls, the absence of gestational age matched healthy controls remained a constraint, highlighting an unavoidable limitation in eoPE research. Another constraint was the small sample size of the eoPE group, which rendered statistical compositional analysis between the disease and term groups challenging.

Importantly, this dissertation and associated preprint publication [363] marks the first comprehensive investigation into the molecular dysregulations within the vSTB layer of the

placenta, addressing critical inquiries regarding how abnormalities in cell-cell fusion during vSTB differentiation may expedite vSTB senescence and its implications for maternal symptoms. By elucidating the trophoblast differentiation trajectory of healthy human pregnancy and establishing connections between findings from eoPE in later gestational stage and early pregnancy, this study sheds light on the elusive molecular mechanisms underlying eoPE and the identification of high-risk mothers. Notably, a novel connection between eoPE dysregulation and an EP300-dependent defect in villus cytotrophoblast fusion has been established, alongside the detection of increased senescence associated with eoPE. This association might represent a fundamental aspect across all pre-eclampsia subtypes, with various triggers and clinical risk factors potentially converging to accelerate the turnover of villous trophoblasts and promote premature senescence. Computational analysis revealed that a subset of p300-regulated perturbed vSTB targets was found to be co-regulated by multiple transcription factors, including *FOS, FOXO1, FOXO4*, and *PAX5*, with functional implications involving the enrichment of signaling pathways like HIF-1α, AP1, PI3K-Akt, and TGFβ. Dysregulated HIF-1α and AP-1 pathways, influenced by placental hypoxia, contribute to abnormal placentation and impaired vascular perfusion, promoting the secretion of factors like sFlt-1 and sEng into the maternal bloodstream, potentially exacerbating eoPE pathogenesis [497-499, 777-779, 781]. Investigating induction of inferred transcription factor activity in first-trimester trophoblast organoids or cell lines might offer a promising avenue for studying eoPE mechanisms and screening prospective drugs, facilitating the development of personalized medicine strategies.

Finally, we found that elevated secretion of specific senescence associated secretory phenotype (SASP) encoding ligands from vSTB sub-states may trigger senescence, potentially initiating eoPE pathogenesis at the fetal side and transmitting it to the maternal side. The data suggested that eoPE could originate within the vSTB layer, disrupting maternal-fetal interactions and lead to complications like impaired nutrient exchange and oxidative stress. Overall, we propose that eoPE syndrome stems from a defect in vCTB to vSTB differentiation, influenced by factors like oxidative stress and altered vascular perfusion, ultimately culminating in placental dysfunction and premature senescence. Notably, SASP ligands such as *GDF15, INHBA, HSPG2, ADAM9,* and *ADAM12* showed promise as potential biomarkers for eoPE, emphasizing the importance of exploring their clinical utility in conjunction with other markers to improve diagnostic accuracy and guide therapeutic interventions. The development of new pharmacological approaches for treating PE at early time points in pregnancy is expected to decrease the risk of cardiovascular disease in post-preeclamptic mothers and their children in the long term.

8.2. Summary: Pancreatic Neuroendocrine Carcinoma

Pancreatic neuroendocrine carcinoma (panNEC) presents a substantial clinical challenge, with large cell NEC (LCNEC) being more prevalent than small cell NEC in the pancreas. By deeply phenotyping the transcriptome of panNEC tissue samples at a single- nucleus level, this research elucidated shared and distinct cell sub-states associated with oncogenic programs and lineage plasticity. A novel finding is the identification of amphicrine sub-states within one patient sample, despite the prototypical morphology of LCNEC. In contrast to all other tumor samples, these amphicrine clusters demonstrated similarities to the exocrine pancreas, particularly the acinar cells, and exhibited a progenitor-like sub-state reminiscent of early pancreatic progenitors. It is worth mentioning that acinar carcinoma often presents with neuroendocrine differentiated parts, prompting further question on whether this sample provides more insight into the transitional states between acinar and neuroendocrine differentiation associated with pancreatic acinar carcinomas rather than prototypical NEC. Follow-up studies in pancreatic acinar carcinoma are warranted to ascertain the relevance of the progenitor-like substate concerning the cell of origin in these rare malignancies. Despite striking variability among patient samples, this study identified shared panNEC sub-states displaying distinct pathway expressions associated with hallmarks of malignancy. Identification of distinct cellular subpopulations within patient samples helped us stratify patients based on molecular profiles, allowing for more targeted and personalized treatment approaches. If validated on a larger sample size, this stratification of amphicrine vs neuroendocrine types could lead to more homogeneous patient cohorts in clinical trials, increasing the likelihood of detecting treatment effects.

Moreover, the shared neuroendocrine and neuroendocrine proliferating sub-states showed widespread expression of signatures linked to neuronal or islet/endocrine cells, which are hard to distinguish. Importantly, the shared sub-states lacked key markers like *INS, GCG* or *SST*, indicative of mature islet cell types, unlike its well differentiated panNET counterpart. Absence of prominent endocrine progenitor markers like *Neurog 3* implied that shared panNEC sub-states do not recapitulate the blueprint of pancreatic endocrine cell development. This supports emerging concepts that the neuroendocrine signatures defining NEC reflect an overarching plasticity state transcending organ-specific program. It is known that islet cells and brain utilize a largely overlapping set of transcription factors with tissue specific targets likely determined by biological context. On this note, this dissertation pinpointed that transcriptional regulator such as *PTF1A* and *PAX6,* having established functions in the brain and pancreas, orchestrate the regulatory landscape underlying the neuroendocrine clusters. Intriguingly, the markers of shared neuroendocrine and

neuroendocrine proliferating sub-states shared notable overlap with brain-specific targets rather than pancreas-specific ones, denoting a role of *PTF1A/PAX6* in mediating an aberrant neuronal phenotype in panNEC. This finding is consistent with the notion that the neuroendocrine phenotype does not mirror bonafide islet cell differentiation programs. Additionally, the panNEC sub-states were found to variably represent NEC signatures from non-gastrointestinal tissues, such as lung and prostate, which are distinct from islet cells. Of note, shared panNEC sub-states exhibited greater transcriptomic similarity to small cell lung cancer (SCLC) signatures as compared to lung LCNEC signatures. On the other hand, lung LCNEC signatures appeared more aligned with the amphicrine sub-states. This observation underlined the complexity of NEC biology and its potential transcriptomic similarities across different tissue types. Interestingly, the shared neuroendocrine and neuroendocrine proliferating sub-states exhibited significant over-enrichment of neuroendocrine and neural-like progenitor programs, formerly associated with worse prognosis in PDAC. Moreover, trajectory analysis of pancreatic NEC sub-states integrated with adult pancreas snRNA-seq data suggested a developmental relation between shared NE cluster and acinar-I cells, supporting the idea of an acinar origin for NEC as has been suggested based on DNA methylation profiles.

Of note, this dissertation revealed the emergence of two shared sub-states of prime therapeutic importance — neuroendocrine proliferating and neuroendocrine HSP+ (hypoxic). Neuroendocrine proliferating demonstrated overexpressed genes linked to *E2F/EZH2* targets, cell-cycle regulation, DNA Damage Response and Fanconi Anemia pathway that could be targeted by Cisplatin, possibly with EZH2-inhibitors as a combination treatment. Notably, the association of NE HSP+ (hypoxic) with accelerated induction of heat stress response, hypoxia and metabolic stress might contribute to the clinical observation of frequent necrotic areas in pancreatic NEC. Exploiting these vulnerabilities could involve repurposing approved or under-trials drugs targeting specific signaling programs. In particular, the findings put forward indicate that utilizing combination therapies involving Geldanamycin, Tanespimycin, or Pimitespib, which target the dysregulated HSP90, alongside Cisplatin/EZH2 inhibitors, might lead to enhanced clinical outcomes. Hence, this paradigm shifts from targeting individual molecules to addressing pathological cell states shows great promise in enhancing treatment outcomes, particularly for diseases that have proven resistant to conventional therapies.

8.3. Summary: Impact of single nuclei sequencing in Precision Medicine

Understanding the cell of origin is crucial for comprehending disease transformation because it provides insights into the initial cellular events that lead to pathological changes. By identifying the specific cell type or lineage from which a disease originates, researchers can elucidate the molecular mechanisms underlying disease initiation and progression. Moreover, understanding the cell of origin can help identify potential therapeutic targets that are specific to the diseased cell type, leading to more effective treatment strategies. It is worth mentioning that we endeavored to delineate the developmental origins of both eoPE and panNEC utilizing snRNA-seq data analysis methodologies. While there is a possibility that eoPE emanates from the outer syncytiotrophoblast layer of the placenta and is transmitted to the maternal endothelium through Senescence-Associated Secretory Phenotype encoding ligands capable of traversing the maternal-fetal barrier, further experiments are necessary to confirm the acinar origin of panNEC patients.

Utilizing snRNA-seq data, it becomes feasible to predict significant transcriptional factors or 'master-regulators' associated with specific diseases, a method applied in both the studies on eoPE and panNEC. Since snRNA-seq has limitations in profiling transcription factor expression comprehensively due to high dropout rates and low library sizes [160], additional complementary experiments are required to validate the findings on a larger scale, potentially at the single-cell level. As an example, the data-driven identification of p300 as a master-regulator in eoPE was confirmed by my collaborators through qPCR analysis and examination of maternal serum in additional cohorts (sourced from Graz and Vienna medical hospitals). For panNEC, we are eagerly awaiting validation to ascertain whether *HSF1* and *HIF1A* co-express within tumor tissue, elucidating the transcriptional regulation underlying NE HSP+ (hypoxic) states. Additionally, we are keen on confirming the involvement of *ATF5* and *NKX2.2* in the interplay between enhanced heat stress and hypoxia.

A prominent role of SASP(s) was decoded in eoPE study, where overexpressed SASP ligands demonstrated the ability to traverse the maternal-fetal barrier, thereby facilitating the transmission of the disease from the fetal to maternal side. However, no enrichment of SASP was observed in panNEC sub-states. Notably, a prior investigation by Raynard et al. 2022 [821] identified role of SASP(s) in promoting neuroendocrine transdifferentiation through SASP-induced Ca2+ signaling in breast cancer cells. They reported that analyses of human breast cancer datasets indicate that neuroendocrine differentiation primarily occurs in p53 wild-type tumors and in older patients, aligning with the involvement of senescent cells and

their secretome, which increase with aging. Consequently, we intend to conduct a more detailed examination of the effects of SASP on panNEC differentiation in future studies.

The collective findings presented in this dissertation, spanning disparate biological systems, underscore the importance of adopting a holistic approach to understanding disease biology and addressing unmet clinical needs. Through the utilization of single-nuclei sequencing technologies and advanced data-driven approaches, as demonstrated in the study of early-onset preeclampsia, adult pancreatic islet maturation, and pancreatic neuroendocrine carcinoma, significant insights into disease mechanisms and potential therapeutic avenues have been uncovered, thereby propelling precision medicine forward. Embracing such a comprehensive perspective from the outset is paramount for identifying novel biomarkers and crafting tailored therapies.

9. Reference

1. Wooten DJ, Quaranta V. Mathematical models of cell phenotype regulation and reprogramming: Make cancer cells sensitive again! Biochim Biophys Acta Rev Cancer. 2017 Apr;1867(2):167-175. doi: 10.1016/j.bbcan.2017.04.001. Epub 2017 Apr 7. PMID: 28396217; PMCID: PMC7572175.

2. Alberts, Bruce M. "cell theory". Encyclopedia Britannica, 4 Mar. 2024, https://www.britannica.com/science/cell-theory.

3. Harris, H., The birth of the cell. 2000: Yale University Press.

4. Marinov, G.K., et al., From single-cell to cell-pool transcriptomes: stochasticity in gene expression and RNA splicing. Genome Res, 2014. 24(3): p. 496-510.

5. Cheung, M.C., et al., Intracellular protein and nucleic acid measured in eight cell types using deep-ultraviolet mass mapping. Cytometry A, 2013. 83(6): p. 540-51.

6. Ho, B., A. Baryshnikova, and G.W. Brown, Unification of Protein Abundance Datasets Yields a Quantitative Saccharomyces cerevisiae Proteome. Cell Syst, 2018. 6(2): p. 192-205 e3.

7. Poveda AM, Le Clech M, Pasero P. Transcription and replication: breaking the rules of the road causes genomic instability. Transcription. 2010 Sep-Oct;1(2):99-102. doi: 10.4161/trns.1.2.12665. PMID: 21326900; PMCID: PMC3023637.

8. Alberts B, Johnson A, Lewis J, et al. New York: Garland Science; 2002.

9. Vijayakumar A, Park A, Steitz JA. Modulation of mRNA 3'-End Processing and Transcription Termination in Virus-Infected Cells. Front Immunol. 2022 Feb 10; 13:828665. doi: 10.3389/fimmu.2022.828665. PMID: 35222412; PMCID: PMC8866245.

10. https://www.genome.gov/25520285/online-education-kit-1955-a-46-human-chromosomes

11. Mayrose I, Lysak MA. The Evolution of Chromosome Numbers: Mechanistic Models and Experimental Approaches. Genome Biol Evol. 2021 Feb 3;13(2):evaa220. doi: 10.1093/gbe/evaa220. PMID: 33566095; PMCID: PMC7875004.

12. Nurk S, Koren S, Rhie A. The complete sequence of a human genome. Science. 2022 Apr;376(6588):44-53. doi: 10.1126/science.abj6987. Epub 2022 Mar 31. PMID: 35357919; PMCID: PMC9186530.

13. Lygerou, Z., Koutroumpas, K.K., Lygeros, J. (2013). DNA Replication. In: Dubitzky, W., Wolkenhauer, O., Cho, KH., Yokota, H. (eds) Encyclopedia of Systems Biology. Springer, New York, NY. https://doi.org/10.1007/978-1-4419-9863-7_40

14. Leman AR, Noguchi E. The replication fork: understanding the eukaryotic replication machinery and the challenges to genome duplication. Genes (Basel). 2013 Mar 1;4(1):1-32. doi: 10.3390/genes4010001. PMID: 23599899; PMCID: PMC3627427.

15. https://us.vwr.com/cms/function-of-mRNA

16. Hombach S, Kretz M. Non-coding RNAs: Classification, Biology and Functioning. Adv Exp Med Biol. 2016;937:3-17. doi: 10.1007/978-3-319-42059-2_1. PMID: 27573892.

17. Jia, X., He, X., Huang, C. et al. Protein translation: biological processes and therapeutic strategies for human diseases. Sig Transduct Target Ther 9, 44 (2024). https://doi.org/10.1038/s41392-024-01749-9

18. Coffin JM. 50th anniversary of the discovery of reverse transcriptase. Mol Biol Cell. 2021 Jan 15;32(2):91-97. doi: 10.1091/mbc. E20-09-0612. PMID: 33448895; PMCID: PMC8120695.

19. Al-Amrani S, Al-Jabri Z, Al-Zaabi A, Alshekaili J, Al-Khabori M. Proteomics: Concepts and applications in human medicine. World J Biol Chem. 2021 Sep 27;12(5):57-69. doi: 10.4331/wjbc.v12.i5.57. PMID: 34630910; PMCID: PMC8473418.

20. Harper JW, Bennett EJ. Proteome complexity and the forces that drive proteome imbalance. Nature. 2016 Sep 15;537(7620):328-38. doi: 10.1038/nature19947. PMID: 27629639; PMCID: PMC5204264.

21. Duan G, Walther D. The roles of post-translational modifications in the context of protein interaction networks. PLoS Comput Biol. 2015 Feb 18;11(2):e1004049. doi: 10.1371/journal.pcbi.1004049. PMID: 25692714; PMCID: PMC4333291.

22. Rose, F. Analysis of phenotypic and spatial cellular heterogeneity from large scale microscopy data. Cellular Biology, Université Paris sciences et lettres. 2019. https://theses.hal.science/tel-03116062/file/Rose-2019-These.pdf

23. Karr JR, Sanghvi JC, Macklin DN, Gutschow MV, Jacobs JM, Bolival B Jr, Assad-Garcia N, Glass JI, Covert MW. A whole-cell computational model predicts phenotype from genotype. Cell. 2012 Jul 20;150(2):389-401. doi: 10.1016/j.cell.2012.05.044. PMID: 22817898; PMCID: PMC3413483.

24. Park J, Brureau A, Kernan K, Starks A, Gulati S, Ogunnaike B, Schwaber J, Vadigepalli R. Inputs drive cell phenotype variability. Genome Res. 2014 Jun;24(6):930-41. doi: 10.1101/gr.161802.113. Epub 2014 Mar 26. PMID: 24671852; PMCID: PMC4032857.

25. Toth T, Balassa T, Bara N, Kovacs F, Kriston A, Molnar C, Haracska L, Sukosd F, Horvath P. Environmental properties of cells improve machine learning-based phenotype recognition accuracy. Sci Rep. 2018 Jul 4;8(1):10085. doi: 10.1038/s41598-018-28482-y. PMID: 29973621; PMCID: PMC6031649.

26. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011 Mar 4;144(5):646-74. doi: 10.1016/j.cell.2011.02.013. PMID: 21376230.

27. Fuchs E, Raghavan S. Getting under the skin of epidermal morphogenesis. Nat Rev Genet. 2002 Mar;3(3):199-209. doi: 10.1038/nrg758. PMID: 11972157.

28. Carmeliet P, Jain RK. Angiogenesis in cancer and other diseases. Nature. 2000 Sep 14;407(6801):249-57. doi: 10.1038/35025220. PMID: 11001068.

29. Wendimu MY, Hooks SB. Microglia Phenotypes in Aging and Neurodegenerative Diseases. Cells. 2022 Jun 30;11(13):2091. doi: 10.3390/cells11132091. PMID: 35805174; PMCID: PMC9266143.

30. Cords L, Tietscher S, Anzeneder T, Langwieder C, Rees M, de Souza N, Bodenmiller B. Cancer-associated fibroblast classification in single-cell and spatial proteomics data. Nat Commun. 2023 Jul 18;14(1):4294. doi: 10.1038/s41467-023-39762-1. PMID: 37463917; PMCID: PMC10354071.

31. Coppé JP, Desprez PY, Krtolica A, Campisi J. The senescence-associated secretory phenotype: the dark side of tumor suppression. Annu Rev Pathol. 2010;5:99-118. doi: 10.1146/annurev-pathol-121808-102144. PMID: 20078217; PMCID: PMC4166495.

32. Wynn TA, Vannella KM. Macrophages in Tissue Repair, Regeneration, and Fibrosis. Immunity. 2016 Mar 15;44(3):450-462. doi: 10.1016/j.immuni.2016.02.015. PMID: 26982353; PMCID: PMC4794754.

33. Arendt D, Musser JM, Baker CVH, Bergman A, Cepko C, Erwin DH, Pavlicev M, Schlosser G, Widder S, Laubichler MD, Wagner GP. The origin and evolution of cell types. Nat Rev Genet. 2016 Dec;17(12):744-757. doi: 10.1038/nrg.2016.127. Epub 2016 Nov 7. PMID: 27818507.

34. Travaglini KJ, Nabhan AN, Penland L, Sinha R, Gillich A, Sit RV, Chang S, Conley SD, Mori Y, Seita J, Berry GJ, Shrager JB, Metzger RJ, Kuo CS, Neff N, Weissman IL, Quake SR, Krasnow MA. A molecular cell atlas of the human lung from single-cell RNA sequencing. Nature. 2020 Nov;587(7835):619-625. doi: 10.1038/s41586-020-2922-4. Epub 2020 Nov 18. PMID: 33208946; PMCID: PMC7704697.

35. Trapnell C. Defining cell types and states with single-cell genomics. Genome Res. 2015 Oct;25(10):1491-8. doi: 10.1101/gr.190595.115. PMID: 26430159; PMCID: PMC4579334.

36. Mannack LV, Eising S, Rentmeister A. Current techniques for visualizing RNA in cells. F1000Res. 2016 Apr 28;5:F1000 Faculty Rev-775. doi: 10.12688/f1000research.8151.1. PMID: 27158473; PMCID: PMC4850879.

37. Weil TT, Parton RM, Davis I. Making the message clear: visualizing mRNA localization. Trends Cell Biol. 2010 Jul;20(7):380-90. doi: 10.1016/j.tcb.2010.03.006. Epub 2010 May 3. PMID: 20444605; PMCID: PMC2902723.

38. Ploeger LS, Dullens HF, Huisman A, van Diest PJ. Fluorescent stains for quantification of DNA by confocal laser scanning microscopy in 3-D. Biotech Histochem. 2008 Apr;83(2):63-9. doi: 10.1080/10520290802127586. PMID: 18568680.

39. Hoheisel JD. Microarray technology: beyond transcript profiling and genotype analysis. Nat Rev Genet. 2006 Mar;7(3):200-10. doi: 10.1038/nrg1809. PMID: 16485019.

40. Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet. 2010 Jan;11(1):31-46. doi: 10.1038/nrg2626. Epub 2009 Dec 8. PMID: 19997069.

41. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet. 2016 May 17;17(6):333-51. doi: 10.1038/nrg.2016.49. PMID: 27184599; PMCID: PMC10373632.

42. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009 Jan;10(1):57-63. doi: 10.1038/nrg2484. PMID: 19015660; PMCID: PMC2949280.

43. Hong M, Tao S, Zhang L, Diao LT, Huang X, Huang S, Xie SJ, Xiao ZD, Zhang H. RNA sequencing: new technologies and applications in cancer research. J Hematol Oncol. 2020 Dec 4;13(1):166. doi: 10.1186/s13045-020-01005-x. PMID: 33276803; PMCID: PMC7716291.

44. Collins FS, Morgan M, Patrinos A. The Human Genome Project: lessons from large-scale biology. Science. 2003 Apr 11;300(5617):286-90. doi: 10.1126/science.1084564. PMID: 12690187.

45. Satam H, Joshi K, Mangrolia U, Waghoo S, Zaidi G, Rawool S, Thakare RP, Banday S, Mishra AK, Das G, Malonia SK. Next-Generation Sequencing Technology: Current Trends and Advancements. Biology (Basel). 2023 Jul 13;12(7):997. doi: 10.3390/biology12070997. PMID: 37508427; PMCID: PMC10376292.

46. https://www.illumina.com/science/technology/next-generation-sequencing/beginners/ngs-cost.html

47. Bentley DR, Balasubramanian S, Swerdlow HP. Accurate whole human genome sequencing using reversible terminator chemistry. Nature. 2008 Nov 6;456(7218):53-9. doi: 10.1038/nature07517. PMID: 18987734; PMCID: PMC2581791.

48. Sabour L, Sabour M, Ghorbian S. Clinical Applications of Next-Generation Sequencing in Cancer Diagnosis. Pathol Oncol Res. 2017 Apr;23(2):225-234. doi: 10.1007/s12253-016-0124-z. Epub 2016 Oct 8. PMID: 27722982.

49. Hussen BM, Abdullah ST, Salihi A, Sabir DK, Sidiq KR, Rasul MF, Hidayat HJ, Ghafouri-Fard S, Taheri M, Jamali E. The emerging roles of NGS in clinical oncology and personalized medicine. Pathol Res Pract. 2022 Feb;230:153760. doi: 10.1016/j.prp.2022.153760. Epub 2022 Jan 10. PMID: 35033746.

50. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods. 2008 Jul;5(7):621-8. doi: 10.1038/nmeth.1226. Epub 2008 May 30. PMID: 18516045.

51. Kuksin M, Morel D, Aglave M, Danlos FX, Marabelle A, Zinovyev A, Gautheret D, Verlingue L. Applications of single-cell and bulk RNA sequencing in onco-immunology. Eur J Cancer. 2021 May; 149:193-210. doi: 10.1016/j.ejca.2021.03.005. Epub 2021 Apr 15. PMID: 33866228.

52. Cloonan N, Forrest AR, Kolle G. Stem cell transcriptome profiling via massive-scale mRNA sequencing. Nat Methods. 2008 Jul;5(7):613-9. doi: 10.1038/nmeth.1223. Epub 2008 May 30. PMID: 18516046.

53. Li X, Wang CY. From bulk, single-cell to spatial RNA sequencing. Int J Oral Sci. 2021 Nov 15;13(1):36. doi: 10.1038/s41368-021-00146-0. PMID: 34782601; PMCID: PMC8593179.

54. Li L, Clevers H. Coexistence of quiescent and active adult stem cells in mammals. Science. 2010 Jan 29;327(5965):542-5. doi: 10.1126/science.1180794. PMID: 20110496; PMCID: PMC4105182.

55. Huang S. Non-genetic heterogeneity of cells in development: more than just noise. Development. 2009 Dec;136(23):3853-62. doi: 10.1242/dev.035139. PMID: 19906852; PMCID: PMC2778736.

56. Eberwine J, Yeh H, Miyashiro K, Cao Y, Nair S, Finnell R, Zettel M, Coleman P. Analysis of gene expression in single live neurons. Proc Natl Acad Sci U S A. 1992 Apr 1;89(7):3010-4. doi: 10.1073/pnas.89.7.3010. PMID: 1557406; PMCID: PMC48793.

57. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, Wang X, Bodeau J, Tuch BB, Siddiqui A, Lao K, Surani MA. mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods. 2009 May;6(5):377-82. doi: 10.1038/nmeth.1315. Epub 2009 Apr 6. PMID: 19349980.

58. Ramsköld D, Luo S, Wang YC, Li R, Deng Q, Faridani OR, Daniels GA, Khrebtukova I, Loring JF, Laurent LC, Schroth GP, Sandberg R. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nat Biotechnol. 2012 Aug;30(8):777-82. doi: 10.1038/nbt.2282. Erratum in: Nat Biotechnol. 2020 Mar;38(3):374. PMID: 22820318; PMCID: PMC3467340.

59. Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB. Massively parallel digital transcriptional profiling of single cells. Nat Commun. 2017 Jan 16;8:14049. doi: 10.1038/ncomms14049. PMID: 28091601; PMCID: PMC5241818.

60. Svensson V, Natarajan KN, Ly LH, Miragaia RJ, Labalette C, Macaulay IC, Cvejic A, Teichmann SA. Power analysis of single-cell RNA-sequencing experiments. Nat Methods. 2017 Apr;14(4):381-387. doi: 10.1038/nmeth.4220. Epub 2017 Mar 6. PMID: 28263961; PMCID: PMC5376499.

61. Tanay A, Regev A. Scaling single-cell genomics from phenomenology to mechanism. Nature. 2017 Jan 18;541(7637):331-338. doi: 10.1038/nature21350. PMID: 28102262; PMCID: PMC5438464.

62. Gustafsson J, Held F, Robinson JL, Björnson E, Jörnsten R, Nielsen J. Sources of variation in cell-type RNA-Seq profiles. PLoS One. 2020 Sep 21;15(9):e0239495. doi: 10.1371/journal.pone.0239495. PMID: 32956417; PMCID: PMC7505444.

63. Battich N, Stoeger T, Pelkmans L. Control of Transcript Variability in Single Mammalian Cells. Cell. 2015 Dec 17;163(7):1596-610. doi: 10.1016/j.cell.2015.11.018. PMID: 26687353.

64. Phillips NE, Mandic A, Omidi S, Naef F, Suter DM. Memory and relatedness of transcriptional activity in mammalian cell lineages. Nat Commun. 2019 Mar 14;10(1):1208. doi: 10.1038/s41467-019-09189-8. PMID: 30872573; PMCID: PMC6418128.

65. Ding J, Adiconis X, Simmons SK, Kowalczyk MS, Hession CC, Marjanovic ND. Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. Nat Biotechnol. 2020 Jun;38(6):737-746. doi: 10.1038/s41587-020-0465-8. Epub 2020 Apr 6. Erratum in: Nat Biotechnol. 2020 Apr 27;: PMID: 32341560; PMCID: PMC7289686.

66. Qiu P. Embracing the dropouts in single-cell RNA-seq analysis. Nat Commun. 2020 Mar 3;11(1):1169. doi: 10.1038/s41467-020-14976-9. PMID: 32127540; PMCID: PMC7054558.

67. Xu C, Cai L, Gao J. An efficient scRNA-seq dropout imputation method using graph attention network. BMC Bioinformatics. 2021 Dec 7;22(1):582. doi: 10.1186/s12859-021-04493-x. PMID: 34876032; PMCID: PMC8650344.

68. Jovic D, Liang X, Zeng H, Lin L, Xu F, Luo Y. Single-cell RNA sequencing technologies and applications: A brief overview. Clin Transl Med. 2022 Mar;12(3):e694. doi: 10.1002/ctm2.694. PMID: 35352511; PMCID: PMC8964935.

69. Wu H, Kirita Y, Donnelly EL, Humphreys BD. Advantages of Single-Nucleus over Single-Cell RNA Sequencing of Adult Kidney: Rare Cell Types and Novel Cell States Revealed in Fibrosis. J Am Soc Nephrol. 2019 Jan;30(1):23-32. doi: 10.1681/ASN.2018090912. Epub 2018 Dec 3. PMID: 30510133; PMCID: PMC6317600.

70. Habib N, Avraham-Davidi I, Basu A, Burks T, Shekhar K, Hofree M, Choudhury SR, Aguet F, Gelfand E, Ardlie K, Weitz DA, Rozenblatt-Rosen O, Zhang F, Regev A. Massively parallel single-nucleus RNA-seq with DroNc-seq. Nat Methods. 2017 Oct;14(10):955-958. doi: 10.1038/nmeth.4407. Epub 2017 Aug 28. PMID: 28846088; PMCID: PMC5623139.

71. Drokhlyansky E, Smillie CS, Van Wittenberghe N, Ericsson M, Griffin GK, Eraslan G, Dionne D, Cuoco MS, Goder-Reiser MN, Sharova T, Kuksenko O, Aguirre AJ, Boland GM, Graham D, Rozenblatt-Rosen O, Xavier RJ, Regev A. The Human and Mouse Enteric Nervous System at Single-Cell Resolution. Cell. 2020 Sep 17;182(6):1606-1622.e23. doi: 10.1016/j.cell.2020.08.003. Epub 2020 Sep 3. PMID: 32888429; PMCID: PMC8358727.

72. Tosti L, Hang Y, Debnath O, Tiesmeyer S, Trefzer T, Steiger K, Ten FW, Lukassen S, Ballke S, Kühl AA, Spieckermann S, Bottino R, Ishaque N, Weichert W, Kim SK, Eils R, Conrad C. Single-Nucleus and In Situ RNA-Sequencing Reveal Cell Topographies in the Human Pancreas. Gastroenterology. 2021 Mar;160(4):1330-1344.e11. doi: 10.1053/j.gastro.2020.11.010. Epub 2020 Nov 16. PMID: 33212097.

73. Kim N, Kang H, Jo A, Yoo SA, Lee HO. Perspectives on single-nucleus RNA sequencing in different cell types and tissues. J Pathol Transl Med. 2023 Jan;57(1):52-59. doi: 10.4132/jptm.2022.12.19. Epub 2023 Jan 10. PMID: 36623812; PMCID: PMC9846005.

74. Slyper M, Porter CBM, Ashenberg O, Waldman J, Drokhlyansky E. A single-cell and single-nucleus RNA-Seq toolbox for fresh and frozen human tumors. Nat Med. 2020 May;26(5):792-802. doi: 10.1038/s41591-020-0844-1. Epub 2020 May 11. Erratum in: Nat Med. 2020 Jun 25; PMID: 32405060; PMCID: PMC7220853.

75. Fernández-García J, Franco F, Parik S, Altea-Manzano P. CD8+ T cell metabolic rewiring defined by scRNA-seq identifies a critical role of ASNS expression dynamics in T cell differentiation. Cell Rep. 2022 Nov 15;41(7):111639. doi: 10.1016/j.celrep.2022.111639. PMID: 36384124.

76. Jerber J, Seaton DD, Cuomo ASE, Kumasaka N, Haldane J, Steer J. Population-scale single-cell RNA-seq profiling across dopaminergic neuron differentiation. Nat Genet. 2021 Mar;53(3):304-312. doi: 10.1038/s41588-021-00801-6. Epub 2021 Mar 4. PMID: 33664506; PMCID: PMC7610897.

77. Islam S, Kjällquist U, Moliner A, Zajac P, Fan JB, Lönnerberg P, Linnarsson S. Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. Genome Res. 2011 Jul;21(7):1160-7. doi: 10.1101/gr.110882.110. Epub 2011 May 4. PMID: 21543516; PMCID: PMC3129258.

78. Natarajan KN. Single-Cell Tagged Reverse Transcription (STRT-Seq). Methods Mol Biol. 2019; 1979:133-153. doi: 10.1007/978-1-4939-9240-9_9. PMID: 31028636.

79. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, Trombetta JJ, Weitz DA, Sanes JR, Shalek AK, Regev A, McCarroll SA. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell. 2015 May 21;161(5):1202-1214. doi: 10.1016/j.cell.2015.05.002. PMID: 26000488; PMCID: PMC4481139.

80. Picelli S, Björklund ÅK, Faridani OR, Sagasser S, Winberg G, Sandberg R. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat Methods. 2013 Nov;10(11):1096-8. doi: 10.1038/nmeth.2639. Epub 2013 Sep 22. PMID: 24056875.

81. Picelli S, Faridani OR, Björklund AK, Winberg G, Sagasser S, Sandberg R. Full-length RNA-seq from single cells using Smart-seq2. Nat Protoc. 2014 Jan;9(1):171-81. doi: 10.1038/nprot.2014.006. Epub 2014 Jan 2. PMID: 24385147.

82. Grün D, Kester L, van Oudenaarden A. Validation of noise models for single-cell transcriptomics. Nat Methods. 2014 Jun;11(6):637-40. doi: 10.1038/nmeth.2930. Epub 2014 Apr 20. PMID: 24747814.

83. Hashimshony T, Senderovich N, Avital G, Klochendler A, de Leeuw Y, Anavy L, Gennert D, Li S, Livak KJ, Rozenblatt-Rosen O, Dor Y, Regev A, Yanai I. CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. Genome Biol. 2016 Apr 28;17:77. doi: 10.1186/s13059-016-0938-8. PMID: 27121950; PMCID: PMC4848782.

84. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, Peshkin L, Weitz DA, Kirschner MW. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell. 2015 May 21;161(5):1187-1201. doi: 10.1016/j.cell.2015.04.044. PMID: 26000487; PMCID: PMC4441768.

85. Jaitin DA, Kenigsberg E, Keren-Shaul H, Elefant N, Paul F, Zaretsky I, Mildner A, Cohen N, Jung S, Tanay A, Amit I. Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. Science. 2014 Feb 14;343(6172):776-9. doi: 10.1126/science.1247651. PMID: 24531970; PMCID: PMC4412462.

86. Ziegenhain C, Vieth B, Parekh S, Reinius B, Guillaumet-Adkins A, Smets M, Leonhardt H, Heyn H, Hellmann I, Enard W. Comparative Analysis of Single-Cell RNA Sequencing Methods. Mol Cell. 2017 Feb 16;65(4):631-643.e4. doi: 10.1016/j.molcel.2017.01.023. PMID: 28212749.

87. Tirosh I, Izar B, Prakadan SM. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science. 2016 Apr 8;352(6282):189-96. doi: 10.1126/science.aad0501. PMID: 27124452; PMCID: PMC4944528.

88. Gokce O, Stanley GM, Treutlein B, Neff NF, Camp JG, Malenka RC, Rothwell PE, Fuccillo MV, Südhof TC, Quake SR. Cellular Taxonomy of the Mouse Striatum as Revealed by Single-Cell RNA-Seq. Cell Rep. 2016 Jul 26;16(4):1126-1137. doi: 10.1016/j.celrep.2016.06.059. Epub 2016 Jul 14. PMID: 27425622; PMCID: PMC5004635.

89. https://www.decibio.com/insights/10x-genomics-launches-gemcode

90. Stoeckius M, Zheng S, Houck-Loomis B, Hao S, Yeung BZ, Mauck WM 3rd, Smibert P, Satija R. Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. Genome Biol. 2018 Dec 19;19(1):224. doi: 10.1186/s13059-018-1603-1. PMID: 30567574; PMCID: PMC6300015.

91. Patel AP, Tirosh I, Trombetta JJ, Shalek AK. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science. 2014 Jun 20;344(6190):1396-401. doi: 10.1126/science.1254257. Epub 2014 Jun 12. PMID: 24925914; PMCID: PMC4123637.

92. Puram SV, Tirosh I, Parikh AS, Patel AP. Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. Cell. 2017 Dec 14;171(7):1611-1624.e24. doi: 10.1016/j.cell.2017.10.044. Epub 2017 Nov 30. PMID: 29198524; PMCID: PMC5878932.

93. Liau ES, Jin S, Chen YC, Liu WS, Calon M, Nedelec S, Nie Q, Chen JA. Single-cell transcriptomic analysis reveals diversity within mammalian spinal motor neurons. Nat Commun. 2023 Jan 3;14(1):46. doi: 10.1038/s41467-022-35574-x. PMID: 36596814; PMCID: PMC9810664.

94. Rousselle TV, McDaniels JM, Shetty AC, Bardhi E, Maluf DG, Mas VR. An optimized protocol for single nuclei isolation from clinical biopsies for RNA-seq. Sci Rep. 2022 Jun 14;12(1):9851. doi: 10.1038/s41598-022-14099-9. PMID: 35701599; PMCID: PMC9198012.

95. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, Szcześniak MW, Gaffney DJ, Elo LL, Zhang X, Mortazavi A. A survey of best practices for RNA-seq data analysis. Genome Biol. 2016 Jan 26;17:13. doi: 10.1186/s13059-016-0881-8. Erratum in: Genome Biol. 2016;17(1):181. PMID: 26813401; PMCID: PMC4728800.

96. Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, Lightfoot S, Menzel W, Granzow M, Ragg T. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC Mol Biol. 2006 Jan 31; 7:3. doi: 10.1186/1471-2199-7-3. PMID: 16448564; PMCID: PMC1413964.

97. Building a Custom Reference for 3' Single Cell Gene Expression: https://www.10xgenomics.com/support/software/cell-ranger/latest/analysis/inputs/cr-3p-references

98. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 2018 Feb 6;19(1):15. doi: 10.1186/s13059-017-1382-0. PMID: 29409532; PMCID: PMC5802054.

99. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3rd, Hao Y, Stoeckius M, Smibert P, Satija R. Comprehensive Integration of Single-Cell Data. Cell. 2019 Jun 13;177(7):1888-1902.e21. doi: 10.1016/j.cell.2019.05.031. Epub 2019 Jun 6. PMID: 31178118; PMCID: PMC6687398.

100. Young MD, Behjati S. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. Gigascience. 2020 Dec 26;9(12):giaa151. doi: 10.1093/gigascience/giaa151. PMID: 33367645; PMCID: PMC7763177.

101. Fleming SJ, Chaffin MD, Arduini A, Akkad AD, Banks E, Marioni JC, Philippakis AA, Ellinor PT, Babadi M. Unsupervised removal of systematic background noise from droplet-based single-cell experiments using CellBender. Nat Methods. 2023 Sep;20(9):1323-1335. doi: 10.1038/s41592-023-01943-7. Epub 2023 Aug 7. PMID: 37550580.

102. Hicks SC, Townes FW, Teng M, Irizarry RA. Missing data and technical variability in single-cell RNA-sequencing experiments. Biostatistics. 2018 Oct 1;19(4):562-578. doi: 10.1093/biostatistics/kxx053. PMID: 29121214; PMCID: PMC6215955.

103. Wolock SL, Lopez R, Klein AM. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. Cell Syst. 2019 Apr 24;8(4):281-291.e9. doi: 10.1016/j.cels.2018.11.005. Epub 2019 Apr 3. PMID: 30954476; PMCID: PMC6625319.

104. McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. Cell Syst. 2019 Apr 24;8(4):329-337.e4. doi: 10.1016/j.cels.2019.03.003. Epub 2019 Apr 3. PMID: 30954475; PMCID: PMC6853612.

105. Xi NM, Li JJ. Benchmarking Computational Doublet-Detection Methods for Single-Cell RNA Sequencing Data. Cell Syst. 2021 Feb 17;12(2):176-194.e6. doi: 10.1016/j.cels.2020.11.008. Epub 2020 Dec 17. PMID: 33338399; PMCID: PMC7897250.

106. Butto T, Mungikar K, Baumann P, Winter J, Lutz B, Gerber S. Nuclei on the Rise: When Nuclei-Based Methods Meet Next-Generation Sequencing. Cells. 2023 Mar 30;12(7):1051. doi: 10.3390/cells12071051. PMID: 37048124; PMCID: PMC10093037.

107. Introduction to Single-cell RNA-seq: https://hbctraining.github.io/scRNA-seq/lessons/04_SC_quality_control.html

108. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol. 2019 Dec 23;20(1):296. doi: 10.1186/s13059-019-1874-1. PMID: 31870423; PMCID: PMC6927181.

109. Lotfollahi M, Naghipourfar M, Luecken MD, Khajavi M, Büttner M, Wagenstetter M, Avsec Ž, Gayoso A, Yosef N, Interlandi M, Rybakov S, Misharin AV, Theis FJ. Mapping single-cell data to reference atlases by transfer learning. Nat Biotechnol. 2022 Jan;40(1):121-130. doi: 10.1038/s41587-021-01001-7. Epub 2021 Aug 30. PMID: 34462589; PMCID: PMC8763644.

110. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, Baglaenko Y, Brenner M, Loh PR, Raychaudhuri S. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods. 2019 Dec;16(12):1289-1296. doi: 10.1038/s41592-019-0619-0. Epub 2019 Nov 18. PMID: 31740819; PMCID: PMC6884693.

111. Polański K, Young MD, Miao Z, Meyer KB, Teichmann SA, Park JE. BBKNN: fast batch alignment of single cell transcriptomes. Bioinformatics. 2020 Feb 1;36(3):964-965. doi: 10.1093/bioinformatics/btz625. PMID: 31400197; PMCID: PMC9883685.

112. Lopez R, Regier J, Cole MB, Jordan MI, Yosef N. Deep generative modeling for single-cell transcriptomics. Nat Methods. 2018 Dec;15(12):1053-1058. doi: 10.1038/s41592-018-0229-2. Epub 2018 Nov 30. PMID: 30504886; PMCID: PMC6289068.

113. Xu C, Lopez R, Mehlman E, Regier J, Jordan MI, Yosef N. Probabilistic harmonization and annotation of single-cell transcriptomics data with deep generative models. Mol Syst Biol. 2021 Jan;17(1): e9620. doi: 10.15252/msb.20209620. PMID: 33491336; PMCID: PMC7829634.

114. Luecken MD, Büttner M, Chaichoompu K, Danese A, Interlandi M, Mueller MF, Strobl DC, Zappia L, Dugas M, Colomé-Tatché M, Theis FJ. Benchmarking atlas-level data integration in single-cell genomics. Nat Methods. 2022 Jan;19(1):41-50. doi: 10.1038/s41592-021-01336-8. Epub 2021 Dec 23. PMID: 34949812; PMCID: PMC8748196.

115. Sikkema L, Ramírez-Suástegui C, Strobl DC, Gillett TE. An integrated cell atlas of the lung in health and disease. Nat Med. 2023 Jun;29(6):1563-1577. doi: 10.1038/s41591-023-02327-2. Epub 2023 Jun 8. PMID: 37291214; PMCID: PMC10287567.

116. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol. 2018 Jun;36(5):411-420. doi: 10.1038/nbt.4096. Epub 2018 Apr 2. PMID: 29608179; PMCID: PMC6700744.

117. Greenacre, M., Groenen, P.J.F., Hastie, T. et al. Principal component analysis. Nat Rev Methods Primers 2, 100. 2022. doi: 10.1038/s43586-022-00184-w

118. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. Philos Trans A Math Phys Eng Sci. 2016 Apr 13;374(2065):20150202. doi: 10.1098/rsta.2015.0202. PMID: 26953178; PMCID: PMC4792409.

119. Hérault J, Ans B. Réseau de neurones à synapses modifiables: décodage de messages sensoriels composites par apprentissage non supervisé et permanent [Neuronal network with modifiable synapses: decoding of composite sensory messages under unsupervised and permanent learning]. C R Acad Sci III. 1984;299(13):525-8. French. PMID: 6437617.

120. Sompairac N, Nazarov PV, Czerwinska U, Cantini L, Biton A, Molkenov A, Zhumadilov Z, Barillot E, Radvanyi F, Gorban A, Kairov U, Zinovyev A. Independent Component Analysis for Unraveling the Complexity of Cancer Omics Datasets. Int J Mol Sci. 2019 Sep 7;20(18):4414. doi: 10.3390/ijms20184414. PMID: 31500324; PMCID: PMC6771121.

121. Tavakol M, Wetzel A. Factor Analysis: a means for theory and instrument development in support of construct validity. Int J Med Educ. 2020 Nov 6; 11:245-247. doi: 10.5116/ijme.5f96.0f4a. PMID: 33170146; PMCID: PMC7883798.

122. https://satijalab.org/seurat/articles/integration_rpca.html

123. Kobak D, Berens P. The art of using t-SNE for single-cell transcriptomics. Nat Commun. 2019 Nov 28;10(1):5416. doi: 10.1038/s41467-019-13056-x. PMID: 31780648; PMCID: PMC6882829.

124. van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. J. Mach. Learning Res 2008. 9, 2579–2605.

125. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform manifold approximation and projection for dimension reduction. https://arxiv.org/abs/1802.03426 (2018).

126. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, Ginhoux F, Newell EW. Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol. 2018 Dec 3. doi: 10.1038/nbt.4314. Epub ahead of print. PMID: 30531897.

127. Geubbelmans M, Rousseau AJ, Burzykowski T, Valkenborg D. Artificial neural networks and deep learning. Am J Orthod Dentofacial Orthop. 2024 Feb;165(2):248-251. doi: 10.1016/j.ajodo.2023.11.003. PMID: 38302219.

128. Drew PJ, Monson JR. Artificial neural networks. Surgery. 2000 Jan;127(1):3-11. doi: 10.1067/msy.2000.102173. PMID: 10660751.

129. Prisciandaro E, Sedda G, Cara A, Diotti C, Spaggiari L, Bertolaccini L. Artificial Neural Networks in Lung Cancer Research: A Narrative Review. J Clin Med. 2023 Jan 22;12(3):880. doi: 10.3390/jcm12030880. PMID: 36769528; PMCID: PMC9918295.

130. Gao J, Hu K, Zhang F, Cui X. Hexagonal image segmentation on spatially resolved transcriptomics. Methods. 2023 Dec; 220:61-68. doi: 10.1016/j.ymeth.2023.11.002. Epub 2023 Nov 4. PMID: 37931852.

131. Min S, Lee B, Yoon S. Deep learning in bioinformatics. Brief Bioinform 2017; 18:851–69.

132. Mikolov T, Karafiát M, Burget L, et al.. Recurrent neural network based language model. In: INTERSPEECH, 11th Annual Conference of the International Speech Communication Association, Makuhari, Chiba, Japan, 2010;2010:1045–8.

133. Marino J. Predictive Coding, Variational Autoencoders, and Biological Connections. Neural Comput. 2021 Dec 15;34(1):1-44. doi: 10.1162/neco_a_01458. PMID: 34758480.

134. Eltager M, Abdelaal T, Charrout M, Mahfouz A, Reinders MJT, Makrodimitris S. Benchmarking variational AutoEncoders on cancer transcriptomics data. PLoS One. 2023 Oct 5;18(10): e0292126. doi: 10.1371/journal.pone.0292126. PMID: 37796856; PMCID: PMC10553230.

135. Ding J, Condon A, Shah SP. Interpretable dimensionality reduction of single cell transcriptome data with deep generative models. Nat Commun. 2018 May 21;9(1):2002. doi: 10.1038/s41467-018-04368-5. PMID: 29784946; PMCID: PMC5962608.

136. Grønbech CH, Vording MF, Timshel PN, Sønderby CK, Pers TH, Winther O. scVAE: variational auto-encoders for single-cell gene expression data. Bioinformatics. 2020 Aug 15;36(16):4415-4422. doi: 10.1093/bioinformatics/btaa293. PMID: 32415966.

137. Eraslan G, Simon LM, Mircea M, Mueller NS, Theis FJ. Single-cell RNA-seq denoising using a deep count autoencoder. Nat Commun. 2019 Jan 23;10(1):390. doi: 10.1038/s41467-018-07931-2. PMID: 30674886; PMCID: PMC6344535.

138. Wang D, Gu J. VASC: Dimension Reduction and Visualization of Single-cell RNA-seq Data by Deep Variational Autoencoder. Genomics Proteomics Bioinformatics. 2018 Oct;16(5):320-331. doi: 10.1016/j.gpb.2018.08.003. Epub 2018 Dec 18. PMID: 30576740; PMCID: PMC6364131.

139. Pierson E, Yau C. ZIFA: Dimensionality reduction for zero-inflated single-cell gene expression analysis. Genome Biol. 2015 Nov 2; 16:241. doi: 10.1186/s13059-015-0805-z. PMID: 26527291; PMCID: PMC4630968.

140. Risso D, Perraudeau F, Gribkova S, Dudoit S, Vert JP. A general and flexible method for signal extraction from single-cell RNA-seq data. Nat Commun. 2018 Jan 18;9(1):284. doi: 10.1038/s41467-017-02554-5. Erratum in: Nat Commun. 2019 Feb 4;10(1):646. PMID: 29348443; PMCID: PMC5773593.

141. Prabhakaran S, Azizi E, Carr A, Pe'er D. Dirichlet Process Mixture Model for Correcting Technical Variation in Single-Cell Gene Expression Data. JMLR Workshop Conf Proc. 2016; 48:1070-1079. PMID: 29928470; PMCID: PMC6004614.

142. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. J. Stat. Mech. Theory Exp. 2008. 10008, 6, doi: 10.1088/1742-5468/2008/10/P10008.

143. Traag VA, Waltman L, van Eck NJ. From Louvain to Leiden: guaranteeing well-connected communities. Sci Rep. 2019 Mar 26;9(1):5233. doi: 10.1038/s41598-019-41695-z. PMID: 30914743; PMCID: PMC6435756.

144. Leary JR, Xu Y, Morrison AB, Jin C, Shen EC, Kuhlers PC, Su Y, Rashid NU, Yeh JJ, Peng XL. Sub-Cluster Identification through Semi-Supervised Optimization of Rare-Cell Silhouettes (SCISSORS) in single-cell RNA-sequencing. Bioinformatics. 2023 Aug 1;39(8): btad449. doi: 10.1093/bioinformatics/btad449. PMID: 37498558; PMCID: PMC10412410.

145. https://hbctraining.github.io/scRNA-seq_online/lessons/seurat_subclustering.html

146. Suryawanshi H, Morozov P, Straus A, Sahasrabudhe N, Max KEA, Garzia A, Kustagi M, Tuschl T, Williams Z. A single-cell survey of the human first-trimester placenta and decidua. Sci Adv. 2018 Oct 31;4(10): eaau4788. doi: 10.1126/sciadv.aau4788. PMID: 30402542; PMCID: PMC6209386.

147. Zeisel A, Hochgerner H, Lönnerberg P, Johnsson A, Memic F, van der Zwan J, Häring M, Braun E, Borm LE, La Manno G, Codeluppi S, Furlan A, Lee K, Skene N, Harris KD, Hjerling-Leffler J, Arenas E, Ernfors P, Marklund U, Linnarsson S. Molecular Architecture of the Mouse Nervous System. Cell. 2018 Aug 9;174(4):999-1014.e22. doi: 10.1016/j.cell.2018.06.021. PMID: 30096314; PMCID: PMC6086934.

148. Cao J, O'Day DR, Pliner HA, Kingsley PD, Deng M, Daza RM, Zager MA, Aldinger KA, Blecher-Gonen R, Zhang F, Spielmann M, Palis J, Doherty D, Steemers FJ, Glass IA, Trapnell C, Shendure J. A human cell atlas of fetal gene expression. Science. 2020 Nov 13;370(6518): eaba7721. doi: 10.1126/science.aba7721. PMID: 33184181; PMCID: PMC7780123.

149. Chen H, Albergante L, Hsu JY, Lareau CA, Lo Bosco G, Guan J, Zhou S, Gorban AN, Bauer DE, Aryee MJ, Langenau DM, Zinovyev A, Buenrostro JD, Yuan GC, Pinello L. Single-cell trajectories reconstruction, exploration and mapping of omics data with STREAM. Nat Commun. 2019 Apr 23;10(1):1903. doi: 10.1038/s41467-019-09670-4. PMID: 31015418; PMCID: PMC6478907.

150. Haghverdi L, Büttner M, Wolf FA, Buettner F, Theis FJ. Diffusion pseudotime robustly reconstructs lineage branching. Nat Methods. 2016 Oct;13(10):845-8. doi: 10.1038/nmeth.3971. Epub 2016 Aug 29. PMID: 27571553.

151. Wolf FA, Hamey FK, Plass M, Solana J, Dahlin JS, Göttgens B, Rajewsky N, Simon L, Theis FJ. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. Genome Biol. 2019 Mar 19;20(1):59. doi: 10.1186/s13059-019-1663-x. PMID: 30890159; PMCID: PMC6425583.

152. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol. 2014 Apr;32(4):381-386. doi: 10.1038/nbt.2859. Epub 2014 Mar 23. PMID: 24658644; PMCID: PMC4122333.

153. Albergante L, Mirkes E, Bac J, Chen H, Martin A, Faure L, Barillot E, Pinello L, Gorban A, Zinovyev A. Robust and Scalable Learning of Complex Intrinsic Dataset Geometry via ElPiGraph. Entropy (Basel). 2020 Mar 4;22(3):296. doi: 10.3390/e22030296. PMID: 33286070; PMCID: PMC7516753.

154. Efremova M, Vento-Tormo M, Teichmann SA, Vento-Tormo R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. Nat Protoc. 2020 Apr;15(4):1484-1506. doi: 10.1038/s41596-020-0292-x. Epub 2020 Feb 26. PMID: 32103204.

155. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, Myung P, Plikus MV, Nie Q. Inference and analysis of cell-cell communication using CellChat. Nat Commun. 2021 Feb 17;12(1):1088. doi: 10.1038/s41467-021-21246-9. PMID: 33597522; PMCID: PMC7889871.

156. Raredon MSB, Yang J, Garritano J, Wang M, Kushnir D, Schupp JC, Adams TS, Greaney AM, Leiby KL, Kaminski N, Kluger Y, Levchenko A, Niklason LE. Computation and visualization of cell-cell signaling topologies in single-cell systems data using Connectome. Sci Rep. 2022 Mar 9;12(1):4187. doi: 10.1038/s41598-022-07959-x. PMID: 35264704; PMCID: PMC8906120.

157. Dimitrov D, Türei D, Garrido-Rodriguez M, Burmedi PL, Nagai JS, Boys C, Ramirez Flores RO, Kim H, Szalai B, Costa IG, Valdeolivas A, Dugourd A, Saez-Rodriguez J. Comparison of methods and resources for cell-cell communication inference from single-cell RNA-Seq data. Nat Commun. 2022 Jun 9;13(1):3224. doi: 10.1038/s41467-022-30755-0. PMID: 35680885; PMCID: PMC9184522.

158. Van de Sande B, Flerin C, Davie K, De Waegeneer M, Hulselmans G, Aibar S, Seurinck R, Saelens W, Cannoodt R, Rouchon Q, Verbeiren T, De Maeyer D, Reumers J, Saeys Y, Aerts S. A scalable SCENIC workflow for single-cell gene regulatory network analysis. Nat Protoc. 2020 Jul;15(7):2247-2276. doi: 10.1038/s41596-020-0336-2. Epub 2020 Jun 19. PMID: 32561888.

159. Kumar N, Mishra B, Athar M, Mukhtar S. Inference of Gene Regulatory Network from Single-Cell Transcriptomic Data Using pySCENIC. Methods Mol Biol. 2021;2328:171-182. doi: 10.1007/978-1-0716-1534-8_10. Erratum in: Methods Mol Biol. 2021;2328:C1. PMID: 34251625.

160. Pokhilko A, Handel AE, Curion F, Volpato V, Whiteley ES, Bøstrand S, Newey SE, Akerman CJ, Webber C, Clark MB, Bowden R, Cader MZ. Targeted single-cell RNA sequencing of transcription factors enhances the identification of cell types and trajectories. Genome Res. 2021 Jun;31(6):1069-1081. doi: 10.1101/gr.273961.120. Epub 2021 May 19. PMID: 34011578; PMCID: PMC8168586.

161. Janky R, Verfaillie A, Imrichová H, Van de Sande B, Standaert L, Christiaens V, Hulselmans G, Herten K, Naval Sanchez M, Potier D, Svetlichnyy D, Kalender Atak Z, Fiers M, Marine JC, Aerts S. iRegulon: from a gene list to a gene regulatory network using large motif and track collections. PLoS Comput Biol. 2014 Jul 24;10(7): e1003731. doi: 10.1371/journal.pcbi.1003731. PMID: 25058159; PMCID: PMC4109854.

162. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003 Nov;13(11):2498-504. doi: 10.1101/gr.1239303. PMID: 14597658; PMCID: PMC403769.

163. Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Berriz GF, Gibbons FD, Dreze M, Ayivi-Guedehoussou N, Klitgord N, Simon C, Boxem M, Milstein S, Rosenberg J, Goldberg DS, Zhang LV, Wong SL, Franklin G, Li S, Albala JS, Lim J, Fraughton C, Llamosas E, Cevik S, Bex C, Lamesch P, Sikorski RS, Vandenhaute J, Zoghbi HY, Smolyar A, Bosak S, Sequerra R, Doucette-Stamm L, Cusick ME, Hill DE, Roth FP, Vidal M. Towards a proteome-scale map of the human protein-protein interaction network. Nature. 2005 Oct 20;437(7062):1173-8. doi: 10.1038/nature04209. Epub 2005 Sep 28. PMID: 16189514.

164. Su G, Kuchinsky A, Morris JH, States DJ, Meng F. GLay: community structure analysis of biological networks. Bioinformatics. 2010 Dec 15;26(24):3135-7. doi: 10.1093/bioinformatics/btq596. PMID: 21123224; PMCID: PMC2995124.

165. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics. 2003 Jan 13; 4:2. doi: 10.1186/1471-2105-4-2. Epub 2003 Jan 13. PMID: 12525261; PMCID: PMC149346.

166. Albert R, Jeong H, Barabási AL. Diameter of the World-Wide Web. Nature. 1999 Sep 9;401(6749):130-131. doi: 10.1038/43601.

167. Barabasi AL, Albert R. Emergence of scaling in random networks. Science. 1999 Oct 15;286(5439):509-12. doi: 10.1126/science.286.5439.509. PMID: 10521342.

168. Carlson JM, Doyle J. Highly optimized tolerance: a mechanism for power laws in designed systems. Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics. 1999 Aug;60(2 Pt A):1412-27. doi: 10.1103/physreve.60.1412. PMID: 11969901.

169. Doncheva NT, Assenov Y, Domingues FS, Albrecht M. Topological analysis and interactive visualization of biological networks and protein structures. Nat Protoc. 2012 Mar 15;7(4):670-85. doi: 10.1038/nprot.2012.004. PMID: 22422314.

170. Rood JE, Maartens A, Hupalowska A, Teichmann SA, Regev A. Impact of the Human Cell Atlas on medicine. Nat Med. 2022 Dec;28(12):2486-2496. doi: 10.1038/s41591-022-02104-7. Epub 2022 Dec 8. PMID: 36482102.

171. Williams CG, Lee HJ, Asatsuma T, Vento-Tormo R, Haque A. An introduction to spatial transcriptomics for biomedical research. Genome Med. 2022 Jun 27;14(1):68. doi: 10.1186/s13073-022-01075-1. PMID: 35761361; PMCID: PMC9238181.

172. Marx V. Method of the Year: spatially resolved transcriptomics. Nat Methods. 2021 Jan;18(1):9-14. doi: 10.1038/s41592-020-01033-y. Erratum in: Nat Methods. 2021 Feb;18(2):219. PMID: 33408395.

173. Kalita-de Croft P, Sadeghi Rad H, Gasper H, O'Byrne K, Lakhani SR, Kulasinghe A. Spatial profiling technologies and applications for brain cancers. Expert Rev Mol Diagn. 2021 Mar;21(3):323-332. doi: 10.1080/14737159.2021.1900735. Epub 2021 Mar 19. PMID: 33685321.

174. Wirth J, Huber N, Yin K, Brood S, Chang S, Martinez-Jimenez CP, Meier M. Spatial transcriptomics using multiplexed deterministic barcoding in tissue. Nat Commun. 2023 Mar 18;14(1):1523. doi: 10.1038/s41467-023-37111-w. PMID: 36934108; PMCID: PMC10024691.

175. Park J, Choi W, Tiesmeyer S, Long B, Borm LE, Garren E, Nguyen TN, Tasic B, Codeluppi S, Graf T, Schlesner M, Stegle O, Eils R, Ishaque N. Cell segmentation-free inference of cell types from in situ transcriptomics data. Nat Commun. 2021 Jun 10;12(1):3545. doi: 10.1038/s41467-021-23807-4. Erratum in: Nat Commun. 2021 Jun 28;12(1):4103. PMID: 34112806; PMCID: PMC8192952.

176. Rao A, Barkley D, França GS, Yanai I. Exploring tissue architecture using spatial transcriptomics. Nature. 2021 Aug;596(7871):211-220. doi: 10.1038/s41586-021-03634-9. Epub 2021 Aug 11. PMID: 34381231; PMCID: PMC8475179.

177. Krausgruber T, Redl A, Barreca D, Doberer K, Romanovskaia D, Dobnikar L, Guarini M, Unterluggauer L, Kleissl L, Atzmüller D, Mayerhofer C, Kopf A, Saluzzo S, Lim CX, Rexie P, Weichhart T, Bock C, Stary G. Single-cell and spatial transcriptomics reveal aberrant lymphoid developmental programs driving granuloma formation. Immunity. 2023 Feb 14;56(2):289-306.e7. doi: 10.1016/j.immuni.2023.01.014. Epub 2023 Feb 6. PMID: 36750099; PMCID: PMC9942876.

178. Srivatsan SR, Regier MC, Barkan E, Franks JM, Packer JS, Grosjean P, Duran M, Saxton S, Ladd JJ, Spielmann M, Lois C, Lampe PD, Shendure J, Stevens KR, Trapnell C. Embryo-scale, single-cell spatial transcriptomics. Science. 2021 Jul 2;373(6550):111-117. doi: 10.1126/science.abb9536. PMID: 34210887; PMCID: PMC9118175.

179. Foster DS, Januszyk M, Yost KE, Chinta MS, Gulati GS, Nguyen AT, Burcham AR, Salhotra A, Ransom RC, Henn D, Chen K, Mascharak S, Tolentino K, Titan AL, Jones RE, da Silva O, Leavitt WT, Marshall CD, des Jardins-Park HE, Hu MS, Wan DC, Wernig G, Wagh D, Coller J, Norton JA, Gurtner GC, Newman AM, Chang HY, Longaker MT. Integrated spatial multiomics reveals fibroblast fate during tissue repair. Proc Natl Acad Sci U S A. 2021 Oct 12;118(41): e2110025118. doi: 10.1073/pnas.2110025118. PMID: 34620713; PMCID: PMC8521719.

180. Codeluppi S, Borm LE, Zeisel A, La Manno G, van Lunteren JA, Svensson CI, Linnarsson S. Spatial organization of the somatosensory cortex revealed by osmFISH. Nat Methods. 2018 Nov;15(11):932-935. doi: 10.1038/s41592-018-0175-z. Epub 2018 Oct 30. PMID: 30377364.

181. Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. Science. 2015 Apr 24;348(6233):aaa6090. doi: 10.1126/science.aaa6090. Epub 2015 Apr 9. PMID: 25858977; PMCID: PMC4662681.

182. Lubeck E, Coskun AF, Zhiyentayev T, Ahmad M, Cai L. Single-cell in situ RNA profiling by sequential hybridization. Nat Methods. 2014 Apr;11(4):360-1. doi: 10.1038/nmeth.2892. PMID: 24681720; PMCID: PMC4085791.

183. Ke R, Mignardi M, Pacureanu A, Svedlund J, Botling J, Wählby C, Nilsson M. In situ sequencing for RNA analysis in preserved tissue and cells. Nat Methods. 2013 Sep;10(9):857-60. doi: 10.1038/nmeth.2563. Epub 2013 Jul 14. PMID: 23852452.

184. Lee JH, Daugharthy ER, Scheiman J, Kalhor R, Ferrante TC, Terry R, Turczyk BM, Yang JL, Lee HS, Aach J, Zhang K, Church GM. Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. Nat Protoc. 2015 Mar;10(3):442-58. doi: 10.1038/nprot.2014.191. Epub 2015 Feb 12. PMID: 25675209; PMCID: PMC4327781.

185. Gyllborg D, Langseth CM, Qian X, Choi E, Salas SM, Hilscher MM, Lein ES, Nilsson M. Hybridization-based in situ sequencing (HybISS) for spatially resolved transcriptomics in human and mouse brain tissue. Nucleic Acids Res. 2020 Nov 4;48(19): e112. doi: 10.1093/nar/gkaa792. Erratum in: Nucleic Acids Res. 2021 Sep 7;49(15):9002. PMID: 32990747; PMCID: PMC7641728.

186. Wang X, Allen WE, Wright MA, Sylwestrak EL, Samusik N, Vesuna S, Evans K, Liu C, Ramakrishnan C, Liu J, Nolan GP, Bava FA, Deisseroth K. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. Science. 2018 Jul 27;361(6400): eaat5691. doi: 10.1126/science.aat5691. Epub 2018 Jun 21. PMID: 29930089; PMCID: PMC6339868.

187. Ståhl PL, Salmén F, Vickovic S, Lundmark A, Navarro JF, Magnusson J, Giacomello S, Asp M, Westholm JO, Huss M, Mollbrink A, Linnarsson S, Codeluppi S, Borg Å, Pontén F, Costea PI, Sahlén P, Mulder J, Bergmann O, Lundeberg J, Frisén J. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Science. 2016 Jul 1;353(6294):78-82. doi: 10.1126/science.aaf2403. PMID: 27365449.

188. Vickovic S, Eraslan G, Salmén F, Klughammer J, Stenbeck L, Schapiro D, Äijö T, Bonneau R, Bergenstråhle L, Navarro JF, Gould J, Griffin GK, Borg Å, Ronaghi M, Frisén J, Lundeberg J, Regev A, Ståhl PL. High-definition spatial transcriptomics for in situ tissue profiling. Nat Methods. 2019 Oct;16(10):987-990. doi: 10.1038/s41592-019-0548-y. Epub 2019 Sep 9. PMID: 31501547; PMCID: PMC6765407.

189. Wang T, Harvey K, Reeves J, Roden DL, Bartonicek N, Yang J, Al-Eryani G, Kaczorowski D, Chan CL, Powell J, O'Toole S, Lim E, Swarbrick A. An experimental comparison of the Digital Spatial Profiling and Visium spatial transcriptomics technologies for cancer research. bioRxiv. 2023 Apr 6. doi: 10.1101/2023.04.06.535805.

190. https://www.10xgenomics.com/support/spatial-gene-expression-ffpe/documentation/steps/library-construction/visium-spatial-gene-expression-reagent-kits-for-ffpe-user-guide

191. Li B, Zhang W, Guo C, Xu H, Li L, Fang M, Hu Y, Zhang X, Yao X, Tang M, Liu K, Zhao X, Lin J, Cheng L, Chen F, Xue T, Qu K. Benchmarking spatial and single-cell transcriptomics integration methods for transcript distribution prediction and cell type deconvolution. Nat Methods. 2022 Jun;19(6):662-670. doi: 10.1038/s41592-022-01480-9. Epub 2022 May 16. PMID: 35577954.

192. Salas SM, Czarnewski P, Kuemmerle LB, Helgadottir S, Matsson-Langseth C, Tismeyer S, Avenel C, Rehman H, Tiklova K, Andersson A, Chatzinikolaou M, Theis FJ, Luecken MD, Wählby C, Ishaque N, Nilsson M. Optimizing Xenium In Situ data utility by quality assessment and best practice analysis workflows. bioRxiv. 2023 Feb 13. doi: 10.1101/2023.02.13.528102.

193. Hildebrandt F, Andersson A, Saarenpää S et al. Spatial Transcriptomics to define transcriptional patterns of zonation and structural components in the mouse liver. Nat Commun. 2021 Dec 2;12(1):7046. doi: 10.1038/s41467-021-27354-w. PMID: 34857782; PMCID: PMC8640072.

194. Elmentaite R, Kumasaka N, Roberts K, Fleming A et al. Cells of the human intestinal tract mapped across space and time. Nature. 2021 Sep;597(7875):250-255. doi: 10.1038/s41586-021-03852-1. Epub 2021 Sep 8. PMID: 34497389; PMCID: PMC8426186.

195. Cable DM, Murray E, Zou LS, Goeva A, Macosko EZ, Chen F, Irizarry RA. Robust decomposition of cell type mixtures in spatial transcriptomics. Nat Biotechnol. 2022 Apr;40(4):517-526. doi: 10.1038/s41587-021-00830-w. Epub 2021 Feb 18. PMID: 33603203; PMCID: PMC8606190.

196. Elosua-Bayes M, Nieto P, Mereu E, Gut I, Heyn H. SPOTlight: seeded NMF regression to deconvolute spatial transcriptomics spots with single-cell transcriptomes. Nucleic Acids Res. 2021 May 21;49(9):e50. doi: 10.1093/nar/gkab043. PMID: 33544846; PMCID: PMC8136778.

197. Coleman K, Hu J, Schroeder A, Lee EB, Li M. SpaDecon: cell-type deconvolution in spatial transcriptomics with semi-supervised learning. Commun Biol. 2023 Apr 7;6(1):378. doi: 10.1038/s42003-023-04761-x. PMID: 37029267; PMCID: PMC10082183.

198. Andersson A, Andrusivová Ž, Czarnewski P, Li X, Sundström E, Lundeberg J. A Landmark-based Common Coordinate Framework for Spatial Transcriptomics Data. bioRxiv. 2021 Nov 11. doi: 10.1101/2021.11.11.468178.

199. Kim J, Park J. Single-cell transcriptomics: a novel precision medicine technique in nephrology. Korean J Intern Med. 2021 May;36(3):479-490. doi: 10.3904/kjim.2020.415. Epub 2021 Mar 18. PMID: 33076636; PMCID: PMC8137400.

200. Nath A, Bild AH. Leveraging Single-Cell Approaches in Cancer Precision Medicine. Trends Cancer. 2021 Apr;7(4):359-372. doi: 10.1016/j.trecan.2021.01.007. Epub 2021 Feb 6. PMID: 33563578; PMCID: PMC7969443.

201. Van de Sande B, Lee JS, Mutasa-Gottgens E, Naughton B, Bacon W, Manning J, Wang Y, Pollard J, Mendez M, Hill J, Kumar N, Cao X, Chen X, Khaladkar M, Wen J, Leach A, Ferran E. Applications of single-cell RNA sequencing in drug discovery and development. Nat Rev Drug Discov. 2023 Jun;22(6):496-520. doi: 10.1038/s41573-023-00688-4. Epub 2023 Apr 28. PMID: 37117846; PMCID: PMC10141847.

202. Shalek AK, Benson M. Single-cell analyses to tailor treatments. Sci Transl Med. 2017 Sep 20;9(408): eaan4730. doi: 10.1126/scitranslmed.aan4730. PMID: 28931656; PMCID: PMC5645080.

203. https://my.clevelandclinic.org/health/articles/7247-fetal-development-stages-of-growth

204. Burton GJ, Watson AL, Hempstock J, Skepper JN, Jauniaux E. Uterine glands provide histiotrophic nutrition for the human fetus during the first trimester of pregnancy. J Clin Endocrinol Metab. 2002 Jun;87(6):2954-9. doi: 10.1210/jcem.87.6.8563. PMID: 12050279.

205. Hempstock J, Cindrova-Davies T, Jauniaux E, Burton GJ. Endometrial glands as a source of nutrients, growth factors and cytokines during the first trimester of human pregnancy: a morphological and immunohistochemical study. Reprod Biol Endocrinol. 2004 Jul 20; 2:58. doi: 10.1186/1477-7827-2-58. PMID: 15265238; PMCID: PMC493283.

206. Dockery P, Bermingham J, Jenkins D. Structure-function relations in the human placenta. Biochem Soc Trans. 2000 Feb;28(2):202-8. doi: 10.1042/bst0280202. PMID: 10816128.

207. Desforges M, Sibley CP. Placental nutrient supply and fetal growth. Int J Dev Biol. 2010;54(2-3):377-90. doi: 10.1387/ijdb.082765md. PMID: 19876836.

208. Maltepe E, Fisher SJ. Placenta: the forgotten organ. Annu Rev Cell Dev Biol. 2015; 31:523-52. doi: 10.1146/annurev-cellbio-100814-125620. Epub 2015 Oct 5. PMID: 26443191.

209. Vento-Tormo R, Efremova M, Botting RA, Turco MY et al. Single-cell reconstruction of the early maternal-fetal interface in humans. Nature. 2018 Nov;563(7731):347-353. doi: 10.1038/s41586-018-0698-6. Epub 2018 Nov 14. PMID: 30429548; PMCID: PMC7612850.

210. Ander SE, Diamond MS, Coyne CB. Immune responses at the maternal-fetal interface. Sci Immunol. 2019 Jan 11;4(31): eaat6114. doi: 10.1126/sciimmunol.aat6114. PMID: 30635356; PMCID: PMC6744611.

211. Balasundaram P, Farhana A. Immunology at the Maternal-Fetal Interface. 2023 Aug 28. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan–. PMID: 34662056.

212. Aplin JD, Myers JE, Timms K, Westwood M. Tracking placental development in health and disease. Nat Rev Endocrinol. 2020 Sep;16(9):479-494. doi: 10.1038/s41574-020-0372-6. Epub 2020 Jun 29. PMID: 32601352.

213. Burton GJ, Jauniaux E. Pathophysiology of placental-derived fetal growth restriction. Am J Obstet Gynecol. 2018 Feb;218(2S): S745-S761. doi: 10.1016/j.ajog.2017.11.577. PMID: 29422210.

214. Knöfler M, Haider S, Saleh L, Pollheimer J, Gamage TKJB, James J. Human placenta and trophoblast development: key molecular mechanisms and model systems. Cell Mol Life Sci. 2019 Sep;76(18):3479-3496. doi: 10.1007/s00018-019-03104-6. Epub 2019 May 3. PMID: 31049600; PMCID: PMC6697717.

215. Albrecht ED, Pepe GJ. Regulation of Uterine Spiral Artery Remodeling: a Review. Reprod Sci. 2020 Oct;27(10):1932-1942. doi: 10.1007/s43032-020-00212-8. Epub 2020 Jun 16. PMID: 32548805; PMCID: PMC7452941.

216. Burton GJ, Woods AW, Jauniaux E, Kingdom JC. Rheological and physiological consequences of conversion of the maternal spiral arteries for uteroplacental blood flow during human pregnancy. Placenta. 2009 Jun;30(6):473-82. doi: 10.1016/j.placenta.2009.02.009. Epub 2009 Apr 17. PMID: 19375795; PMCID: PMC2697319.

217. Jauniaux E, Watson AL, Hempstock J, Bao YP, Skepper JN, Burton GJ. Onset of maternal arterial blood flow and placental oxidative stress. A possible factor in human early pregnancy failure. Am J Pathol. 2000 Dec;157(6):2111-22. doi: 10.1016/S0002-9440(10)64849-3. PMID: 11106583; PMCID: PMC1885754.

218. Greenbaum S, Averbukh I, Soon E, Rizzuto G, Baranski A et al. A spatially resolved timeline of the human maternal-fetal interface. Nature. 2023 Jul;619(7970):595-605. doi: 10.1038/s41586-023-06298-9. Epub 2023 Jul 19. PMID: 37468587; PMCID: PMC10356615.

219. Okae H, Toh H, Sato T, Hiura H, Takahashi S, Shirane K, Kabayama Y, Suyama M, Sasaki H, Arima T. Derivation of Human Trophoblast Stem Cells. Cell Stem Cell. 2018 Jan 4;22(1):50-63. e6. doi: 10.1016/j.stem.2017.11.004. Epub 2017 Dec 14. PMID: 29249463.

220. Muñoz-Fernández R, de la Mata C, Prados A, Perea A, Ruiz-Magaña MJ, Llorca T, Fernández-Rubio P, Blanco O, Abadía-Molina AC, Olivares EG. Human predecidual stromal

cells have distinctive characteristics of pericytes: Cell contractility, chemotactic activity, and expression of pericyte markers and angiogenic factors. Placenta. 2018 Jan; 61:39-47. doi: 10.1016/j.placenta.2017.11.010. Epub 2017 Nov 22. PMID: 29277270.

221. Vinketova K, Mourdjeva M, Oreshkova T. Human Decidual Stromal Cells as a Component of the Implantation Niche and a Modulator of Maternal Immunity. J Pregnancy. 2016; 2016:8689436. doi: 10.1155/2016/8689436. Epub 2016 Apr 28. PMID: 27239344; PMCID: PMC4864559.

222. Murata H, Tanaka S, Okada H. Immune Tolerance of the Human Decidua. J Clin Med. 2021 Jan 18;10(2):351. doi: 10.3390/jcm10020351. PMID: 33477602; PMCID: PMC7831321.

223. Koopman LA, Kopcow HD, Rybalov B, Boyson JE, Orange JS, Schatz F, Masch R, Lockwood CJ, Schachter AD, Park PJ, Strominger JL. Human decidual natural killer cells are a unique NK cell subset with immunomodulatory potential. J Exp Med. 2003 Oct 20;198(8):1201-12. doi: 10.1084/jem.20030305. PMID: 14568979; PMCID: PMC2194228.

224. Jiang X, Du MR, Li M, Wang H. Three macrophage subsets are identified in the uterus during early human pregnancy. Cell Mol Immunol. 2018 Dec;15(12):1027-1037. doi: 10.1038/s41423-018-0008-0. Epub 2018 Apr 4. PMID: 29618777; PMCID: PMC6269440.

225. Aluvihare VR, Kallikourdis M, Betz AG. Regulatory T cells mediate maternal tolerance to the fetus. Nat Immunol. 2004 Mar;5(3):266-71. doi: 10.1038/ni1037. Epub 2004 Feb 1. PMID: 14758358.

226. Agostinis C, Masat E, Bossi F, Ricci G, Menegazzi R, Lombardelli L, Zito G, Mangogna A, Degan M, Gattei V, Piccinni MP, Kishore U, Bulla R. Transcriptomics and Immunological Analyses Reveal a Pro-Angiogenic and Anti-Inflammatory Phenotype for Decidual Endothelial Cells. Int J Mol Sci. 2019 Mar 31;20(7):1604. doi: 10.3390/ijms20071604. PMID: 30935090; PMCID: PMC6479455.

227. Red-Horse K, Crawford Y, Shojaei F, Ferrara N. Endothelium-microenvironment interactions in the developing embryo and in the adult. Dev Cell. 2007 Feb;12(2):181-94. doi: 10.1016/j.devcel.2007.01.013. PMID: 17276337.

228. Pawlak JB, Bálint L, Lim L, Ma W, Davis RB, Benyó Z, Soares MJ, Oliver G, Kahn ML, Jakus Z, Caron KM. Lymphatic mimicry in maternal endothelial cells promotes placental spiral artery remodeling. J Clin Invest. 2019 Nov 1;129(11):4912-4921. doi: 10.1172/JCI120446. PMID: 31415243; PMCID: PMC6819089.

229. Aird WC. Endothelial cell heterogeneity. Cold Spring Harb Perspect Med. 2012 Jan;2(1): a006429. doi: 10.1101/cshperspect.a006429. PMID: 22315715; PMCID: PMC3253027.

230. Liu M, Ji M, Cheng J, Li Y, Tian Y, Zhao H, Wang Y, Zhu S, Zhang L, Xu X, Feng GS, Liang X, Bao H, Tang Y, Kong S, Lu J, Wang H, Lu Z, Deng W. Deciphering a critical role of uterine epithelial SHP2 in parturition initiation at single cell resolution. Nat Commun. 2023 Nov 14;14(1):7356. doi: 10.1038/s41467-023-43102-8. PMID: 37963860; PMCID: PMC10646072.

231. George DL, Phillips JA 3rd, Francke U, Seeburg PH. The genes for growth hormone and chorionic somatomammotropin are on the long arm of human chromosome 17 in region q21 to qter. Hum Genet. 1981;57(2):138-41. doi: 10.1007/BF00282009. PMID: 6262212.

232. Männik J, Vaas P, Rull K, Teesalu P, Rebane T, Laan M. Differential expression profile of growth hormone/chorionic somatomammotropin genes in placenta of small- and large-for-gestational-age newborns. J Clin Endocrinol Metab. 2010 May;95(5):2433-42. doi: 10.1210/jc.2010-0023. Epub 2010 Mar 16. PMID: 20233782; PMCID: PMC2869554.

233. Syversen U, Opsjøn SL, Stridsberg M, Sandvik AK, Dimaline R, Tingulstad S, Arntzen KJ, Brenna E, Waldum HL. Chromogranin A and pancreastatin-like immunoreactivity in normal pregnancies. J Clin Endocrinol Metab. 1996 Dec;81(12):4470-5. doi: 10.1210/jcem.81.12.8954061. PMID: 8954061.

234. Yabe S, Alexenko AP, Amita M, Yang Y, Schust DJ, Sadovsky Y, Ezashi T, Roberts RM. Comparison of syncytiotrophoblast generated from human embryonic stem cells and from term placentas. Proc Natl Acad Sci U S A. 2016 May 10;113(19):E2598-607. doi: 10.1073/pnas.1601630113. Epub 2016 Apr 5. PMID: 27051068; PMCID: PMC4868474.

235. Zhou Y, Fisher SJ, Janatpour M, Genbacev O, Dejana E, Wheelock M, Damsky CH. Human cytotrophoblasts adopt a vascular phenotype as they differentiate. A strategy for successful endovascular invasion? J Clin Invest. 1997 May 1;99(9):2139-51. doi: 10.1172/JCI119387. PMID: 9151786; PMCID: PMC508044.

236. Baergen, R. N., Burton, G. J. & Kaplan, C. G. Benirschke's Pathology of the Human Placenta. 7 edn, (Springer Nature, 2023).

237. Reyes L, Wolfe B, Golos T. Hofbauer Cells: Placental Macrophages of Fetal Origin. Results Probl Cell Differ. 2017; 62:45-60. doi: 10.1007/978-3-319-54090-0_3. PMID: 28455705.

238. Thomas JR, Appios A, Zhao X, Dutkiewicz R, Donde M, Lee CYC, Naidu P, Lee C, Cerveira J, Liu B, Ginhoux F, Burton G, Hamilton RS, Moffett A, Sharkey A, McGovern N. Phenotypic and functional characterization of first-trimester human placental macrophages, Hofbauer cells. J Exp Med. 2021 Jan 4;218(1): e20200891. doi: 10.1084/jem.20200891. PMID: 33075123; PMCID: PMC7579740.

239. Reyes L, Golos TG. Hofbauer Cells: Their Role in Healthy and Complicated Pregnancy. Front Immunol. 2018 Nov 15; 9:2628. doi: 10.3389/fimmu.2018.02628. PMID: 30498493; PMCID: PMC6249321.

240. Ozarslan N, Robinson JF, Buarpung S, Kim MY, Ansbro MR, Akram J, Montoya DJ, Kamya MR, Kakuru A, Dorsey G, Rosenthal PJ, Cheng G, Feeney ME, Fisher SJ, Gaw SL. Distinct transcriptional profiles of maternal and fetal placental macrophages at term are associated with gravidity. bioRxiv. 2023 Sep 25. doi: 10.1101/2023.09.25.559419.

241. Ilic D, Kapidzic M, Genbacev O. Isolation of human placental fibroblasts. Curr Protoc Stem Cell Biol. 2008 Jun;Chapter 1:Unit 1C.6. doi: 10.1002/9780470151808.sc01c06s5. PMID: 18770629.

242. Rajagopalan S, Long EO. Cell atlas reveals the landscape of early pregnancy. Nature. 2018 Nov;563(7733):337-338. doi: 10.1038/d41586-018-07317-w.

243. Adachi K, Nikaido I, Ohta H, Ohtsuka S, Ura H, Kadota M, Wakayama T, Ueda HR, Niwa H. Context-dependent wiring of Sox2 regulatory networks for self-renewal of embryonic and trophoblast stem cells. Mol Cell. 2013 Nov 7;52(3):380-92. doi: 10.1016/j.molcel.2013.09.002. Epub 2013 Oct 10. PMID: 24120664.

244. Naama M, Rahamim M, Zayat V, Sebban S, Radwan A, Orzech D, Lasry R, Ifrah A, Jaber M, Sabag O, Yassen H, Khatib A, Epsztejn-Litman S, Novoselsky-Persky M, Makedonski K, Deri N, Goldman-Wohl D, Cedar H, Yagel S, Eiges R, Buganim Y. Pluripotency-independent induction of human trophoblast stem cells from fibroblasts. Nat Commun. 2023 Jun 8;14(1):3359. doi: 10.1038/s41467-023-39104-1. PMID: 37291192; PMCID: PMC10250329.

245. van den Berg DL, Zhang W, Yates A, Engelen E, Takacs K, Bezstarosti K, Demmers J, Chambers I, Poot RA. Estrogen-related receptor beta interacts with Oct4 to positively regulate Nanog gene expression. Mol Cell Biol. 2008 Oct;28(19):5986-95. doi: 10.1128/MCB.00301-08. Epub 2008 Jul 28. PMID: 18662995; PMCID: PMC2547019.

246. Zhang X, Zhang J, Wang T, Esteban MA, Pei D. Esrrb activates Oct4 transcription and sustains self-renewal and pluripotency in embryonic stem cells. J Biol Chem. 2008 Dec 19;283(51):35825-33. doi: 10.1074/jbc.M803481200. Epub 2008 Oct 28. PMID: 18957414.

247. Bačenková D, Trebuňová M, Čížková D, Hudák R, Dosedla E, Findrik-Balogová A, Živčák J. In Vitro Model of Human Trophoblast in Early Placentation. Biomedicines. 2022 Apr 15;10(4):904. doi: 10.3390/biomedicines10040904. PMID: 35453654; PMCID: PMC9029210.

248. Latos PA, Hemberger M. From the stem of the placental tree: trophoblast stem cells and their progeny. Development. 2016 Oct 15;143(20):3650-3660. doi: 10.1242/dev.133462. PMID: 27802134.

249. Wang Y, Jiang X, Jia L, Wu X, Wu H, Wang Y, Li Q, Yu R, Wang H, Xiao Z, Liang X. A Single-Cell Characterization of Human Post-implantation Embryos Cultured In Vitro Delineates Morphogenesis in Primary Syncytialization. Front Cell Dev Biol. 2022 Jun 15; 10:835445. doi: 10.3389/fcell.2022.835445. PMID: 35784461; PMCID: PMC9240912.

250. Norwitz ER. Defective implantation and placentation: laying the blueprint for pregnancy complications. Reprod Biomed Online. 2006 Oct;13(4):591-9. doi: 10.1016/s1472-6483(10)60649-9. PMID: 17007686.

251. Wu P, Green M, Myers JE. Hypertensive disorders of pregnancy. BMJ. 2023 Jun 30;381: e071653. doi: 10.1136/bmj-2022-071653. PMID: 37391211.

252. Roberts JM, Redman CW. Pre-eclampsia: more than pregnancy-induced hypertension. Lancet. 1993 Jun 5;341(8858):1447-51. doi: 10.1016/0140-6736(93)90889-o. Erratum in: Lancet 1993 Aug 21;342(8869):504. PMID: 8099148.

253. Phipps EA, Thadhani R, Benzing T, Karumanchi SA. Pre-eclampsia: pathogenesis, novel diagnostics and therapies. Nat Rev Nephrol. 2019 May;15(5):275-289. doi: 10.1038/s41581-019-0119-6. Erratum in: Nat Rev Nephrol. 2019 Jun;15(6):386. PMID: 30792480; PMCID: PMC6472952.

254. Staff AC, Ranheim T, Khoury J, Henriksen T. Increased contents of phospholipids, cholesterol, and lipid peroxides in decidua basalis in women with preeclampsia. Am J Obstet Gynecol. 1999 Mar;180(3 Pt 1):587-92. doi: 10.1016/s0002-9378(99)70259-0. PMID: 10076133.

255. Khan KS, Wojdyla D, Say L, Gülmezoglu AM, Van Look PF. WHO analysis of causes of maternal death: a systematic review. Lancet. 2006 Apr 1;367(9516):1066-1074. doi: 10.1016/S0140-6736(06)68397-9. PMID: 16581405.

256. Say L, Chou D, Gemmill A, Tunçalp Ö, Moller AB, Daniels J, Gülmezoglu AM, Temmerman M, Alkema L. Global causes of maternal death: a WHO systematic analysis. Lancet Glob Health. 2014 Jun;2(6): e323-33. doi: 10.1016/S2214-109X(14)70227-X. Epub 2014 May 5. PMID: 25103301.

257. Stevens W, Shih T, Incerti D, Ton TGN, Lee HC, Peneva D, Macones GA, Sibai BM, Jena AB. Short-term costs of preeclampsia to the United States health care system. Am J Obstet Gynecol. 2017 Sep;217(3):237-248.e16. doi: 10.1016/j.ajog.2017.04.032. Epub 2017 Jul 11. PMID: 28708975.

258. von Dadelszen P, Magee LA, Roberts JM. Subclassification of preeclampsia. Hypertens Pregnancy. 2003;22(2):143-8. doi: 10.1081/PRG-120021060. PMID: 12908998.

259. Hauth JC, Ewell MG, Levine RJ, Esterlitz JR, Sibai B, Curet LB, Catalano PM, Morris CD. Pregnancy outcomes in healthy nulliparas who developed hypertension. Calcium for Preeclampsia Prevention Study Group. Obstet Gynecol. 2000 Jan;95(1):24-8. doi: 10.1016/s0029-7844(99)00462-7. PMID: 10636496

260. Lisonkova S, Joseph KS. Incidence of preeclampsia: risk factors and outcomes associated with early- versus late-onset disease. Am J Obstet Gynecol. 2013 Dec;209(6): 544.e1-544.e12. doi: 10.1016/j.ajog.2013.08.019. Epub 2013 Aug 22. PMID: 23973398.

261. Rocha G. Consequences of early-onset preeclampsia on neonatal morbidity and mortality. Minerva Pediatr (Torino). 2023 Feb;75(1):87-97. doi: 10.23736/S2724-5276.22.06714-3. Epub 2022 Apr 4. PMID: 35373936.

262. Ndwiga C, Odwe G, Pooja S, Ogutu O, Osoti A, E Warren C. Clinical presentation and outcomes of pre-eclampsia and eclampsia at a national hospital, Kenya: A retrospective cohort study. PLoS One. 2020 Jun 5;15(6): e0233323. doi: 10.1371/journal.pone.0233323. PMID: 32502144; PMCID: PMC7274433.

263. Wójtowicz A, Zembala-Szczerba M, Babczyk D, Kołodziejczyk-Pietruszka M, Lewaczyńska O, Huras H. Early- and Late-Onset Preeclampsia: A Comprehensive Cohort Study of Laboratory and Clinical Findings according to the New ISHHP Criteria. Int J Hypertens. 2019 Sep 17; 2019:4108271. doi: 10.1155/2019/4108271. PMID: 31637053; PMCID: PMC6766116.

264. Williams D. Long-term complications of preeclampsia. Semin Nephrol. 2011 Jan;31(1):111-22. doi: 10.1016/j.semnephrol.2010.10.010. PMID: 21266269.

265. Birukov A, Herse F, Nielsen JH, Kyhl HB, Golic M et al. Blood Pressure and Angiogenic Markers in Pregnancy: Contributors to Pregnancy-Induced Hypertension and Offspring Cardiovascular Risk. Hypertension. 2020 Sep;76(3):901-909. doi: 10.1161/HYPERTENSIONAHA.119.13966. Epub 2020 Jun 8. PMID: 32507044.

266. Brouwers L, van der Meiden-van Roest AJ, Savelkoul C, Vogelvang TE, Lely AT, Franx A, van Rijn BB. Recurrence of pre-eclampsia and the risk of future hypertension and cardiovascular disease: a systematic review and meta-analysis. BJOG. 2018 Dec;125(13):1642-1654. doi: 10.1111/1471-0528.15394. Epub 2018 Aug 10. PMID: 29978553; PMCID: PMC6283049.

267. Ray JG, Vermeulen MJ, Schull MJ, Redelmeier DA. Cardiovascular health after maternal placental syndromes (CHAMPS): population-based retrospective cohort study. Lancet. 2005 Nov 19;366(9499):1797-803. doi: 10.1016/S0140-6736(05)67726-4. PMID: 16298217.

268. Redman CW. Current topic: pre-eclampsia and the placenta. Placenta. 1991 Jul-Aug;12(4):301-8. doi: 10.1016/0143-4004(91)90339-h. PMID: 1946241.

269. Staff AC. The two-stage placental model of preeclampsia: An update. J Reprod Immunol. 2019 Sep;134-135:1-10. doi: 10.1016/j.jri.2019.07.004. Epub 2019 Jul 8. PMID: 31301487.

270. Brosens I, Pijnenborg R, Vercruysse L, Romero R. The "Great Obstetrical Syndromes" are associated with disorders of deep placentation. Am J Obstet Gynecol. 2011 Mar;204(3):193-201. doi: 10.1016/j.ajog.2010.08.009. Epub 2010 Nov 20. PMID: 21094932; PMCID: PMC3369813.

271. Staff AC, Fjeldstad HE, Fosheim IK, Moe K, Turowski G, Johnsen GM, Alnaes-Katjavivi P, Sugulle M. Failure of physiological transformation and spiral artery atherosis: their roles in preeclampsia. Am J Obstet Gynecol. 2022 Feb;226(2S): S895-S906. doi: 10.1016/j.ajog.2020.09.026. Epub 2020 Sep 21. PMID: 32971013.

272. Brown MA, Magee LA, Kenny LC, Karumanchi SA, McCarthy FP, Saito S, Hall DR, Warren CE, Adoyi G, Ishaku S; International Society for the Study of Hypertension in Pregnancy (ISSHP). Hypertensive Disorders of Pregnancy: ISSHP Classification, Diagnosis, and Management Recommendations for International Practice. Hypertension. 2018 Jul;72(1):24-43. doi: 10.1161/HYPERTENSIONAHA.117.10803. PMID: 29899139.

273. Roberts JM, Escudero C. The placenta in preeclampsia. Pregnancy Hypertens. 2012 Apr 1;2(2):72-83. doi: 10.1016/j.preghy.2012.01.001. PMID: 22745921; PMCID: PMC3381433.

274. Parada-Niño L, Castillo-León LF, Morel A. Preeclampsia, Natural History, Genes, and miRNAs Associated with the Syndrome. J Pregnancy. 2022 Feb 14; 2022:3851225. doi: 10.1155/2022/3851225. PMID: 35198246; PMCID: PMC8860533.

275. Salonen Ros H, Lichtenstein P, Lipworth L, Cnattingius S. Genetic effects on the liability of developing pre-eclampsia and gestational hypertension. Am J Med Genet. 2000 Apr 10;91(4):256-60. PMID: 10766979.

276. Williams PJ, Broughton Pipkin F. The genetics of pre-eclampsia and other hypertensive disorders of pregnancy. Best Pract Res Clin Obstet Gynaecol. 2011 Aug;25(4):405-17. doi: 10.1016/j.bpobgyn.2011.02.007. Epub 2011 Mar 22. PMID: 21429808; PMCID: PMC3145161.

277. Nilsson E, Salonen Ros H, Cnattingius S, Lichtenstein P. The importance of genetic and environmental effects for pre-eclampsia and gestational hypertension: a family study. BJOG. 2004 Mar;111(3):200-6. doi: 10.1111/j.1471-0528.2004.00042x.x. PMID: 14961879.

278. Buurma AJ, Turner RJ, Driessen JH, Mooyaart AL, Schoones JW, Bruijn JA, Bloemenkamp KW, Dekkers OM, Baelde HJ. Genetic variants in pre-eclampsia: a meta-analysis. Hum Reprod Update. 2013 May-Jun;19(3):289-303. doi: 10.1093/humupd/dms060. Epub 2013 Jan 8. PMID: 23300202.

279. Gray KJ, Saxena R, Karumanchi SA. Genetic predisposition to preeclampsia is conferred by fetal DNA variants near FLT1, a gene involved in the regulation of angiogenesis. Am J Obstet

Gynecol. 2018 Feb;218(2):211-218. doi: 10.1016/j.ajog.2017.11.562. Epub 2017 Nov 11. PMID: 29138037; PMCID: PMC5807126.

280. McGinnis R, Steinthorsdottir V, Williams NO, Thorleifsson G et al. Variants in the fetal genome near FLT1 are associated with risk of preeclampsia. Nat Genet. 2017 Aug;49(8):1255-1260. doi: 10.1038/ng.3895. Epub 2017 Jun 19. PMID: 28628106.

281. Chau K, Hennessy A, Makris A. Placental growth factor and pre-eclampsia. J Hum Hypertens. 2017 Dec;31(12):782-786. doi: 10.1038/jhh.2017.61. Epub 2017 Aug 24. PMID: 29115294; PMCID: PMC5680413.

282. Hesson AM, Langen ES, Plazyo O, Gudjonsson JE, Ganesh SK. Placental transcriptome analysis of hypertensive pregnancies identifies distinct gene expression profiles of preeclampsia superimposed on chronic hypertension. BMC Med Genomics. 2023 May 2;16(1):91. doi: 10.1186/s12920-023-01522-x. PMID: 37131171; PMCID: PMC10152005.

283. Tarca AL, Romero R, Erez O, Gudicha DW, Than NG, Benshalom-Tirosh N, Pacora P, Hsu CD, Chaiworapongsa T, Hassan SS, Gomez-Lopez N. Maternal whole blood mRNA signatures identify women at risk of early preeclampsia: a longitudinal study. J Matern Fetal Neonatal Med. 2021 Nov;34(21):3463-3474. doi: 10.1080/14767058.2019.1685964. Epub 2020 Jan 3. PMID: 31900005; PMCID: PMC10544754.

284. Kang Q, Li W, Xiao J, Yu N, Fan L, Sha M, Ma S, Wu J, Chen S. Integrated analysis of multiple microarray studies to identify novel gene signatures in preeclampsia. Placenta. 2021 Feb; 105:104-118. doi: 10.1016/j.placenta.2021.01.023. Epub 2021 Feb 2. PMID: 33571845.

285. Ackerman WE 4th, Buhimschi CS, Brown TL, Zhao G, Summerfield TL, Buhimschi IA. Transcriptomics-Based Subphenotyping of the Human Placenta Enabled by Weighted Correlation Network Analysis in Early-Onset Preeclampsia With and Without Fetal Growth Restriction. Hypertension. 2023 Jun;80(6):1363-1374. doi: 10.1161/HYPERTENSIONAHA.122.20807. Epub 2023 Mar 29. PMID: 36987911; PMCID: PMC10192030.

286. Zhou W, Wang H, Yang Y, Guo F, Yu B, Su Z. Trophoblast Cell Subtypes and Dysfunction in the Placenta of Individuals with Preeclampsia Revealed by Single-Cell RNA Sequencing. Mol Cells. 2022 May 31;45(5):317-328. doi: 10.14348/molcells.2021.0211. PMID: 35289305; PMCID: PMC9095508.

287. Zhang T, Bian Q, Chen Y, Wang X, Yu S, Liu S, Ji P, Li L, Shrestha M, Dong S, Guo R, Zhang H. Dissecting human trophoblast cell transcriptional heterogeneity in preeclampsia using single-cell RNA sequencing. Mol Genet Genomic Med. 2021 Aug;9(8): e1730. doi: 10.1002/mgg3.1730. Epub 2021 Jul 2. PMID: 34212522; PMCID: PMC8404237.

288. Admati I, Skarbianskis N, Hochgerner H, Ophir O, Weiner Z, Yagel S, Solt I, Zeisel A. Two distinct molecular faces of preeclampsia revealed by single-cell transcriptomics. Med. 2023 Oct 13;4(10):687-709.e7. doi: 10.1016/j.medj.2023.07.005. Epub 2023 Aug 11. PMID: 37572658.

289. Negre-Salvayre A, Swiader A, Salvayre R, Guerby P. Oxidative stress, lipid peroxidation and premature placental senescence in preeclampsia. Arch Biochem Biophys. 2022 Nov 15; 730:109416. doi: 10.1016/j.abb.2022.109416. Epub 2022 Sep 27. PMID: 36179910.

290. Huppertz B, Frank HG, Kingdom JC, Reister F, Kaufmann P. Villous cytotrophoblast regulation of the syncytial apoptotic cascade in the human placenta. Histochem Cell Biol. 1998 Nov;110(5):495-508. doi: 10.1007/s004180050311. PMID: 9826129.

291. Huppertz B, Sammar M, Chefetz I, Neumaier-Wagner P, Bartz C, Meiri H. Longitudinal determination of serum placental protein 13 during development of preeclampsia. Fetal Diagn Ther. 2008;24(3):230-6. doi: 10.1159/000151344. Epub 2008 Aug 28. PMID: 18753763.

292. Arutyunyan A, Roberts K, Troulé K, Wong FCK et al. Spatial multiomics map of trophoblast development in early pregnancy. Nature. 2023 Apr;616(7955):143-151. doi: 10.1038/s41586-023-05869-0. Epub 2023 Mar 29. PMID: 36991123; PMCID: PMC10076224.

293. Dimitriadis E, Rolnik DL, Zhou W, Estrada-Gutierrez G, Koga K, Francisco RPV, Whitehead C, Hyett J, da Silva Costa F, Nicolaides K, Menkhorst E. Author Correction: Pre-eclampsia. Nat Rev Dis Primers. 2023 Jul 3;9(1):35. doi: 10.1038/s41572-023-00451-4. Erratum for: Nat Rev Dis Primers. 2023 Feb 16;9(1):8. PMID: 37400556.

294. Muraro MJ, Dharmadhikari G, Grün D, Groen N, Dielen T, Jansen E, van Gurp L, Engelse MA, Carlotti F, de Koning EJ, van Oudenaarden A. A Single-Cell Transcriptome Atlas of the Human Pancreas. Cell Syst. 2016 Oct 26;3(4):385-394.e3. doi: 10.1016/j.cels.2016.09.002. Epub 2016 Sep 29. PMID: 27693023; PMCID: PMC5092539.

295. Karpińska M, Czauderna M. Pancreas-Its Functions, Disorders, and Physiological Impact on the Mammals' Organism. Front Physiol. 2022 Mar 30; 13:807632. doi: 10.3389/fphys.2022.807632. PMID: 35431983; PMCID: PMC9005876.

296. Wang KL, Tao M, Wei TJ, Wei R. Pancreatic β cell regeneration induced by clinical and preclinical agents. World J Stem Cells. 2021 Jan 26;13(1):64-77. doi: 10.4252/wjsc.v13.i1.64. PMID: 33584980; PMCID: PMC7859987.

297. Rieck S, Kaestner KH. Expansion of beta-cell mass in response to pregnancy. Trends Endocrinol Metab. 2010 Mar;21(3):151-8. doi: 10.1016/j.tem.2009.11.001. Epub 2009 Dec 16. PMID: 20015659; PMCID: PMC3627215.

298. Kahn SE, Hull RL, Utzschneider KM. Mechanisms linking obesity to insulin resistance and type 2 diabetes. Nature. 2006 Dec 14;444(7121):840-6. doi: 10.1038/nature05482. PMID: 17167471.

299. Mezza T, Muscogiuri G, Sorice GP, Clemente G, Hu J, Pontecorvi A, Holst JJ, Giaccari A, Kulkarni RN. Insulin resistance alters islet morphology in nondiabetic humans. Diabetes. 2014 Mar;63(3):994-1007. doi: 10.2337/db13-1013. Epub 2013 Nov 11. PMID: 24215793; PMCID: PMC3931397.

300. Thorel F, Népote V, Avril I, Kohno K, Desgraz R, Chera S, Herrera PL. Conversion of adult pancreatic alpha-cells to beta-cells after extreme beta-cell loss. Nature. 2010 Apr 22;464(7292):1149-54. doi: 10.1038/nature08894. Epub 2010 Apr 4. PMID: 20364121; PMCID: PMC2877635.

301. Xu X, D'Hoker J, Stangé G, Bonné S, De Leu N, Xiao X, Van de Casteele M, Mellitzer G, Ling Z, Pipeleers D, Bouwens L, Scharfmann R, Gradwohl G, Heimberg H. Beta cells can be generated from endogenous progenitors in injured adult mouse pancreas. Cell. 2008 Jan 25;132(2):197-207. doi: 10.1016/j.cell.2007.12.015. PMID: 18243096.

302. Chera S, Baronnier D, Ghila L, Cigliola V, Jensen JN, Gu G, Furuyama K, Thorel F, Gribble FM, Reimann F, Herrera PL. Diabetes recovery by age-dependent conversion of pancreatic δ-cells into insulin producers. Nature. 2014 Oct 23;514(7523):503-7. doi: 10.1038/nature13633. Epub 2014 Aug 20. PMID: 25141178; PMCID: PMC4209186.

303. Bonner-Weir S, Baxter LA, Schuppin GT, Smith FE. A second pathway for regeneration of adult exocrine and endocrine pancreas. A possible recapitulation of embryonic development. Diabetes. 1993 Dec;42(12):1715-20. doi: 10.2337/diab.42.12.1715. PMID: 8243817.

304. Segerstolpe Å, Palasantza A, Eliasson P, Andersson EM, Andréasson AC, Sun X, Picelli S, Sabirsh A, Clausen M, Bjursell MK, Smith DM, Kasper M, Ämmälä C, Sandberg R. Single-Cell Transcriptome Profiling of Human Pancreatic Islets in Health and Type 2 Diabetes. Cell Metab. 2016 Oct 11;24(4):593-607. doi: 10.1016/j.cmet.2016.08.020. Epub 2016 Sep 22. PMID: 27667667; PMCID: PMC5069352.

305. Chen K, Zhang J, Huang Y, Tian X, Yang Y, Dong A. Single-cell RNA-seq transcriptomic landscape of human and mouse islets and pathological alterations of diabetes. iScience. 2022 Oct 14;25(11):105366. doi: 10.1016/j.isci.2022.105366. PMID: 36339258; PMCID: PMC9626680.

306. Weng C, Gu A, Zhang S, Lu L, Ke L, Gao P, Liu X, Wang Y, Hu P, Plummer D, MacDonald E, Zhang S, Xi J, Lai S, Leskov K, Yuan K, Jin F, Li Y. Single cell multiomic analysis reveals diabetes-associated β-cell heterogeneity driven by HNF1A. Nat Commun. 2023 Sep 5;14(1):5400. doi: 10.1038/s41467-023-41228-3. PMID: 37669939; PMCID: PMC10480445.

307. Hrovatin K, Bastidas-Ponce A, Bakhti M, Zappia L, Büttner M, Salinno C, Sterr M, Böttcher A, Migliorini A, Lickert H, Theis FJ. Delineating mouse β-cell identity during lifetime and in diabetes with a single cell atlas. Nat Metab. 2023 Sep;5(9):1615-1637. doi: 10.1038/s42255-023-00876-x. Epub 2023 Sep 11. PMID: 37697055; PMCID: PMC10513934.

308. Ma Z, Zhang X, Zhong W, Yi H, Chen X, Zhao Y, Ma Y, Song E, Xu T. Deciphering early human pancreas development at the single-cell level. Nat Commun. 2023 Sep 2;14(1):5354. doi: 10.1038/s41467-023-40893-8. PMID: 37660175; PMCID: PMC10475098.

309. Olaniru OE, Hook P, Persaud SJ. Using single-cell multi-omics screening of human fetal pancreas to identify novel players in human beta cell development. Diabet Med. 2022 Dec;39(12): e14992. doi: 10.1111/dme.14992. Epub 2022 Nov 2. PMID: 36302085; PMCID: PMC9828353.

310. de la O S, Liu Z, Sun H, Yu SK, Wong DM, Chu E, Rao SA, Eng N, Peixoto G, Bouza J, Shen Y, Knox SM, Tward AD, Gloyn AL, Sneddon JB. Single-Cell Multi-Omic Roadmap of Human Fetal Pancreatic Development. bioRxiv. 2022 Feb 17. doi: 10.1101/2022.02.17.480942.

311. Arda HE, Li L, Tsai J, Torre EA, Rosli Y, Peiris H, Spitale RC, Dai C, Gu X, Qu K, Wang P, Wang J, Grompe M, Scharfmann R, Snyder MS, Bottino R, Powers AC, Chang HY, Kim SK. Age-Dependent Pancreatic Gene Regulation Reveals Mechanisms Governing Human β Cell Function. Cell Metab. 2016 May 10;23(5):909-20. doi: 10.1016/j.cmet.2016.04.002. Epub 2016 Apr 28. PMID: 27133132; PMCID: PMC4864151.

312. Grimley PM, Albores-Saavedra J. Neoplasms with neuroendocrine differentiation: implications of molecular pathology. J Exp Pathol. 1987 Spring;3(2):155-76. PMID: 3320292.

313. DeLellis RA. The neuroendocrine system and its tumors: an overview. Am J Clin Pathol. 2001 Jun;115 Suppl: S5-16. doi: 10.1309/7GR5-L7YW-3G78-LDJ6. PMID: 11993690.

314. Ahmed M. Gastrointestinal neuroendocrine tumors in 2020. World J Gastrointest Oncol. 2020 Aug 15;12(8):791-807. doi: 10.4251/wjgo.v12.i8.791. PMID: 32879660; PMCID: PMC7443843.

315. Tischler AS. The dispersed neuroendocrine cells: the structure, function, regulation and effects of xenobiotics on this system. Toxicol Pathol. 1989;17(2):307-16. doi: 10.1177/019262338901700207. Erratum in: Toxicol Pathol 1990;18(1 Pt 1):78. PMID: 2675281.

316. Pavel M, Öberg K, Falconi M, Krenning EP, Sundin A, Perren A, Berruti A; ESMO Guidelines Committee. Electronic address: clinicalguidelines@esmo.org. Gastroenteropancreatic neuroendocrine neoplasms: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2020 Jul;31(7):844-860. doi: 10.1016/j.annonc.2020.03.304. Epub 2020 Apr 6. PMID: 32272208.

317. Danti G, Flammia F, Matteuzzi B, Cozzi D, Berti V, Grazzini G, Pradella S, Recchia L, Brunese L, Miele V. Gastrointestinal neuroendocrine neoplasms (GI-NENs): hot topics in morphological, functional, and prognostic imaging. Radiol Med. 2021 Dec;126(12):1497-1507. doi: 10.1007/s11547-021-01408-x. Epub 2021 Aug 24. PMID: 34427861; PMCID: PMC8702509.

318. Rindi G, Mete O, Uccella S, Basturk O, La Rosa S, Brosens LAA, Ezzat S, de Herder WW, Klimstra DS, Papotti M, Asa SL. Overview of the 2022 WHO Classification of Neuroendocrine Neoplasms. Endocr Pathol. 2022 Mar;33(1):115-154. doi: 10.1007/s12022-022-09708-2. Epub 2022 Mar 16. PMID: 35294740.

319. Dasari A, Shen C, Halperin D, Zhao B, Zhou S, Xu Y, Shih T, Yao JC. Trends in the Incidence, Prevalence, and Survival Outcomes in Patients With Neuroendocrine Tumors in the United States. JAMA Oncol. 2017 Oct 1;3(10):1335-1342. doi: 10.1001/jamaoncol.2017.0589. PMID: 28448665; PMCID: PMC5824320.

320. Heetfeld M, Chougnet CN, Olsen IH, Rinke A, Borbath I, Crespo G, Barriuso J, Pavel M, O'Toole D, Walter T; other Knowledge Network members. Characteristics and treatment of patients with G3 gastroenteropancreatic neuroendocrine neoplasms. Endocr Relat Cancer. 2015 Aug;22(4):657-64. doi: 10.1530/ERC-15-0119. Epub 2015 Jun 25. PMID: 26113608.

321. Sorbye H, Welin S, Langer SW, Vestermark LW, Holt N, Osterlund P, Dueland S, Hofsli E, Guren MG, Ohrling K, Birkemeyer E, Thiis-Evensen E, Biagini M, Gronbaek H, Soveri LM, Olsen IH, Federspiel B, Assmus J, Janson ET, Knigge U. Predictive and prognostic factors for treatment and survival in 305 patients with advanced gastrointestinal neuroendocrine carcinoma (WHO G3): the NORDIC NEC study. Ann Oncol. 2013 Jan;24(1):152-60. doi: 10.1093/annonc/mds276. Epub 2012 Sep 11. PMID: 22967994.

322. Garcia-Carbonero R, Sorbye H, Baudin E, Raymond E, Wiedenmann B, Niederle B, Sedlackova E, Toumpanakis C, Anlauf M, Cwikla JB, Caplin M, O'Toole D, Perren A; Vienna Consensus Conference participants. ENETS Consensus Guidelines for High-Grade Gastroenteropancreatic Neuroendocrine Tumors and Neuroendocrine Carcinomas. Neuroendocrinology. 2016;103(2):186-94. doi: 10.1159/000443172. Epub 2016 Jan 5. PMID: 26731334.

323. Venizelos A, Elvebakken H, Perren A, Nikolaienko O, Deng W, Lothe IMB, Couvelard A, Hjortland GO, Sundlöv A, Svensson J, Garresori H, Kersten C, Hofsli E, Detlefsen S, Krogh M, Sorbye H, Knappskog S. The molecular characteristics of high-grade gastroenteropancreatic neuroendocrine neoplasms. Endocr Relat Cancer. 2021 Nov 11;29(1):1-14. doi: 10.1530/ERC-21-0152. PMID: 34647903; PMCID: PMC8630776.

324. Konukiewitz B, Jesinghaus M, Steiger K, Schlitter AM, Kasajima A, Sipos B, Zamboni G, Weichert W, Pfarr N, Klöppel G. Pancreatic neuroendocrine carcinomas reveal a closer relationship to ductal adenocarcinomas than to neuroendocrine tumors G3. Hum Pathol. 2018 Jul; 77:70-79. doi: 10.1016/j.humpath.2018.03.018. Epub 2018 Mar 26. PMID: 29596894.

325. Elvebakken H, Perren A, Scoazec JY, Tang LH, Federspiel B, Klimstra DS, Vestermark LW, Ali AS, Zlobec I, Myklebust TÅ, Hjortland GO, Langer SW, Gronbaek H, Knigge U, Tiensuu Janson E, Sorbye H. A Consensus-Developed Morphological Re-Evaluation of 196 High-Grade Gastroenteropancreatic Neuroendocrine Neoplasms and Its Clinical Correlations. Neuroendocrinology. 2021;111(9):883-894. doi: 10.1159/000511905. Epub 2020 Oct 1. PMID: 33002892.

326. Rinke A, Auernhammer CJ, Bodei L, Kidd M, Krug S, Lawlor R, Marinoni I, Perren A, Scarpa A, Sorbye H, Pavel ME, Weber MM, Modlin I, Gress TM. Treatment of advanced gastroenteropancreatic neuroendocrine neoplasia, are we on the way to personalised medicine? Gut. 2021 Sep;70(9):1768-1781. doi: 10.1136/gutjnl-2020-321300. Epub 2021 Mar 10. PMID: 33692095.

327. Detjen K, Hammerich L, Özdirik B, Demir M, Wiedenmann B, Tacke F, Jann H, Roderburg C. Models of Gastroenteropancreatic Neuroendocrine Neoplasms: Current Status and Future Directions. Neuroendocrinology. 2021;111(3):217-236. doi: 10.1159/000509864. Epub 2020 Jul 2. PMID: 32615560.

328. Tomita T. Significance of chromogranin A and synaptophysin in pancreatic neuroendocrine tumors. Bosn J Basic Med Sci. 2020 Aug 3;20(3):336-346. doi: 10.17305/bjbms.2020.4632. PMID: 32020844; PMCID: PMC7416176.

329. Wang YH, Yang QC, Lin Y, Xue L, Chen MH, Chen J. Chromogranin A as a marker for diagnosis, treatment, and survival in patients with gastroenteropancreatic neuroendocrine neoplasm. Medicine (Baltimore). 2014 Dec;93(27):e247. doi: 10.1097/MD.0000000000000247. PMID: 25501094; PMCID: PMC4602794.

330. Wick MR. Immunohistology of neuroendocrine and neuroectodermal tumors. Semin Diagn Pathol. 2000 Aug;17(3):194-203. PMID: 10968705.

331. Kanehira K, Khoury T. Neuroendocrine markers expression in pancreatic serous cystadenoma. Appl Immunohistochem Mol Morphol. 2011 Mar;19(2):141-6. doi: 10.1097/PAI.0b013e3181f5023d. PMID: 21030861.

332. Al-Toubah T, Pelle E, Valone T, Haider M, Strosberg JR. Efficacy and Toxicity Analysis of Capecitabine and Temozolomide in Neuroendocrine Neoplasms. J Natl Compr Canc Netw. 2021 Aug 24;20(1):29-36. doi: 10.6004/jnccn.2021.7017. PMID: 34433130.

333. Hoffman SE, Dowrey TW, Villacorta Martin C, Bi K et al. Intertumoral lineage diversity and immunosuppressive transcriptional programs in well-differentiated gastroenteropancreatic

neuroendocrine tumors. Sci Adv. 2023 Sep 29;9(39): eadd9668. doi: 10.1126/sciadv.add9668. Epub 2023 Sep 27. PMID: 37756410; PMCID: PMC10530100.

334. Di Domenico A, Pipinikas CP, Maire RS, Bräutigam K, Simillion C, Dettmer MS, Vassella E, Thirlwell C, Perren A, Marinoni I. Epigenetic landscape of pancreatic neuroendocrine tumours reveals distinct cells of origin and means of tumour progression. Commun Biol. 2020 Dec 7;3(1):740. doi: 10.1038/s42003-020-01479-y. PMID: 33288854; PMCID: PMC7721725.

335. Chan CS, Laddha SV, Lewis PW, Koletsky MS, Robzyk K, Da Silva E, Torres PJ, Untch BR, Li J, Bose P, Chan TA, Klimstra DS, Allis CD, Tang LH. ATRX, DAXX or MEN1 mutant pancreatic neuroendocrine tumors are a distinct alpha-cell signature subgroup. Nat Commun. 2018 Oct 12;9(1):4158. doi: 10.1038/s41467-018-06498-2. PMID: 30315258; PMCID: PMC6185985.

336. Cejas P, Drier Y, Dreijerink KMA, Brosens LAA et al. Enhancer signatures stratify and predict outcomes of non-functional pancreatic neuroendocrine tumors. Nat Med. 2019 Aug;25(8):1260-1265. doi: 10.1038/s41591-019-0493-4. Epub 2019 Jul 1. Erratum in: Nat Med. 2019 Oct;25(10):1627. PMID: 31263286; PMCID: PMC6919319.

337. Sadanandam A, Wullschleger S, Lyssiotis CA, Grötzinger C et al. A Cross-Species Analysis in Pancreatic Neuroendocrine Tumors Reveals Molecular Subtypes with Distinctive Clinical, Metastatic, Developmental, and Metabolic Characteristics. Cancer Discov. 2015 Dec;5(12):1296-313. doi: 10.1158/2159-8290.CD-15-0068. Epub 2015 Oct 7. PMID: 26446169; PMCID: PMC4946251.

338. Yachida S, Totoki Y, Noë M, Nakatani Y et al. Comprehensive Genomic Profiling of Neuroendocrine Carcinomas of the Gastrointestinal System. Cancer Discov. 2022 Mar 1;12(3):692-711. doi: 10.1158/2159-8290.CD-21-0669. PMID: 34880079; PMCID: PMC9394397.

339. Bastidas-Ponce A, Tritschler S, Dony L, Scheibner K, Tarquis-Medina M, Salinno C, Schirge S, Burtscher I, Böttcher A, Theis FJ, Lickert H, Bakhti M. Comprehensive single cell mRNA profiling reveals a detailed roadmap for pancreatic endocrinogenesis. Development. 2019 Jun 17;146(12): dev173849. doi: 10.1242/dev.173849. PMID: 31160421.

340. George J, Walter V, Peifer M, Alexandrov LB et al. Integrative genomic profiling of large-cell neuroendocrine carcinomas reveals distinct subtypes of high-grade neuroendocrine lung tumors. Nat Commun. 2018 Mar 13;9(1):1048. doi: 10.1038/s41467-018-03099-x. PMID: 29535388; PMCID: PMC5849599.

341. Alshalalfa M, Liu Y, Wyatt AW, Gibb EA, Tsai HK, Erho N, Lehrer J, Takhar M, Ramnarine VR, Collins CC, Den RB, Schaeffer EM, Davicioni E, Lotan TL, Bismar TA. Characterization of transcriptomic signature of primary prostate cancer analogous to prostatic small cell neuroendocrine carcinoma. Int J Cancer. 2019 Dec 15;145(12):3453-3461. doi: 10.1002/ijc.32430. Epub 2019 Jun 10. PMID: 31125117; PMCID: PMC6852174.

342. Hwang WL, Jagadeesh KA, Guo JA, Hoffman HI et al. Single-nucleus and spatial transcriptome profiling of pancreatic cancer identifies multicellular dynamics associated with neoadjuvant treatment. Nat Genet. 2022 Aug;54(8):1178-1191. doi: 10.1038/s41588-022-01134-8. Epub 2022 Jul 28. PMID: 35902743; PMCID: PMC10290535.

343. Cluver CA, Bergman L, Bergkvist J, Imberg H, Geerts L, Hall DR, Mol BW, Tong S, Walker SP. Impact of fetal growth restriction on pregnancy outcome in women undergoing expectant

management for preterm pre-eclampsia. Ultrasound Obstet Gynecol. 2023 Nov;62(5):660-667. doi: 10.1002/uog.26282. Epub 2023 Oct 9. PMID: 37289938.

344. Weiler J, Tong S, Palmer KR. Is fetal growth restriction associated with a more severe maternal phenotype in the setting of early onset pre-eclampsia? A retrospective study. PLoS One. 2011;6(10): e26937. doi: 10.1371/journal.pone.0026937. Epub 2011 Oct 28. PMID: 22046419; PMCID: PMC3203930.

345. Leavey K, Bainbridge SA, Cox BJ. Large scale aggregate microarray analysis reveals three distinct molecular subclasses of human preeclampsia. PLoS One. 2015 Feb 13;10(2): e0116508. doi: 10.1371/journal.pone.0116508. PMID: 25679511; PMCID: PMC4332506.

346. Wu P, Gulati M, Kwok CS, Wong CW, Narain A, O'Brien S, Chew-Graham CA, Verma G, Kadam UT, Mamas MA. Preterm Delivery and Future Risk of Maternal Cardiovascular Disease: A Systematic Review and Meta-Analysis. J Am Heart Assoc. 2018 Jan 15;7(2): e007809. doi: 10.1161/JAHA.117.007809. PMID: 29335319; PMCID: PMC5850169.

347. Crump C, Howell EA, Stroustrup A, McLaughlin MA, Sundquist J, Sundquist K. Association of Preterm Birth With Risk of Ischemic Heart Disease in Adulthood. JAMA Pediatr. 2019 Aug 1;173(8):736-743. doi: 10.1001/jamapediatrics.2019.1327. PMID: 31157896; PMCID: PMC6547251.

348. Heikkilä K, Metsälä J, Pulakka A, Nilsen SM, Kivimäki M, Risnes K, Kajantie E. Preterm birth and the risk of multimorbidity in adolescence: a multiregister-based cohort study. Lancet Public Health. 2023 Sep;8(9): e680-e690. doi: 10.1016/S2468-2667(23)00145-7. PMID: 37633677.

349. Davies EL, Bell JS, Bhattacharya S. Preeclampsia and preterm delivery: A population-based case-control study. Hypertens Pregnancy. 2016 Nov;35(4):510-519. doi: 10.1080/10641955.2016.1190846. Epub 2016 Jun 20. PMID: 27322489.

350. An H, Jin M, Li Z, Zhang L, Li H, Zhang Y, Ye R, Li N. Impact of gestational hypertension and pre-eclampsia on preterm birth in China: a large prospective cohort study. BMJ Open. 2022 Sep 27;12(9): e058068. doi: 10.1136/bmjopen-2021-058068. PMID: 36167382; PMCID: PMC9516080.

351. Vogel, Sinae, Roxanne Rajaii, Ari Kane, Yvonne Cheng, Geri Ottaviano, and Aaron Caughey. "Preeclampsia as an etiology for preterm birth: what is the burden of disease?" American Journal of Obstetrics & Gynecology. 2008 Dec doi: 10.1016/j.ajog.2009.10.817.

352. Beardmore-Gray A, Seed PT, Fleminger J, Zwertbroek E, Bernardes T, Mol BW, Battersby C, Koopmans C, Broekhuijsen K, Boers K, Owens MY, Thornton J, Green M, Shennan AH, Groen H, Chappell LC. Planned delivery or expectant management in preeclampsia: an individual participant data meta-analysis. Am J Obstet Gynecol. 2022 Aug;227(2):218-230.e8. doi: 10.1016/j.ajog.2022.04.034. Epub 2022 Apr 26. PMID: 35487323.

353. Pique-Regi R, Romero R, Tarca AL, Sendler ED, Xu Y, Garcia-Flores V, Leng Y, Luca F, Hassan SS, Gomez-Lopez N. Single cell transcriptional signatures of the human placenta in term and preterm parturition. Elife. 2019 Dec 12;8: e52004. doi: 10.7554/eLife.52004. PMID: 31829938; PMCID: PMC6949028.

354. Ellery PM, Cindrova-Davies T, Jauniaux E, Ferguson-Smith AC, Burton GJ. Evidence for transcriptional activity in the syncytiotrophoblast of the human placenta. Placenta. 2009

Apr;30(4):329-34. doi: 10.1016/j.placenta.2009.01.002. Epub 2009 Feb 11. PMID: 19215981; PMCID: PMC3712185.

355. Simpson RA, Mayhew TM, Barnes PR. From 13 weeks to term, the trophoblast of human placenta grows by the continuous recruitment of new proliferative units: a study of nuclear number using the disector. Placenta. 1992 Sep-Oct;13(5):501-12. doi: 10.1016/0143-4004(92)90055-x. PMID: 1470609.

356. Lu X, Wang R, Zhu C, Wang H, Lin HY, Gu Y, Cross JC, Wang H. Fine-Tuned and Cell-Cycle-Restricted Expression of Fusogenic Protein Syncytin-2 Maintains Functional Placental Syncytia. Cell Rep. 2018 Jun 26;23(13):3979. doi: 10.1016/j.celrep.2018.06.043. Erratum for: Cell Rep. 2017 Oct 31;21(5):1150-1159. PMID: 29949779.

357. Wang M, Liu Y, Sun R, Liu F, Li J, Yan L, Zhang J, Xie X, Li D, Wang Y, Li S, Zhu X, Li R, Lu F, Xiao Z, Wang H. Single-nucleus multi-omic profiling of human placental syncytiotrophoblasts identifies cellular trajectories during pregnancy. Nat Genet. 2024 Feb;56(2):294-305. doi: 10.1038/s41588-023-01647-w. Epub 2024 Jan 24. PMID: 38267607; PMCID: PMC10864176.

358. Jaremek A, Jeyarajah MJ, Jaju Bhattad G, Renaud SJ. Omics Approaches to Study Formation and Function of Human Placental Syncytiotrophoblast. Front Cell Dev Biol. 2021 Jun 15; 9:674162. doi: 10.3389/fcell.2021.674162. PMID: 34211975; PMCID: PMC8240757.

359. Xiao Z, Yan L, Liang X, Wang H. Progress in deciphering trophoblast cell differentiation during human placentation. Curr Opin Cell Biol. 2020 Dec; 67:86-91. doi: 10.1016/j.ceb.2020.08.010. Epub 2020 Sep 18. PMID: 32957014.

360. Ntranos V, Yi L, Melsted P, Pachter L. A discriminative learning approach to differential expression analysis for single-cell RNA-seq. Nat Methods. 2019 Feb;16(2):163-166. doi: 10.1038/s41592-018-0303-9. Epub 2019 Jan 21. PMID: 30664774.

361. Pullin JM, McCarthy DJ. A comparison of marker gene selection methods for single-cell RNA sequencing data. Genome Biol. 2024 Feb 26;25(1):56. doi: 10.1186/s13059-024-03183-0. PMID: 38409056; PMCID: PMC10895860.

362. Wilcoxon F. Individual comparisons by ranking methods. Biometrics Bulletin. 1945 Dec;1(6):80-83.

363. Nonn O, Debnath O, Valdes DS, Sallinger K, Secener AK, Haider S, Fischer C, Tiesmeyer S, Nimo J, Kuenzer T, Maxian T, Knöfler M, Karau P, Bartolomaeus H, Kroneis T, Frolova A, Neuper L, Haase N, Kräker K, Kedziora S, Forstner D, Verlohren S, Stern C, Coscia F, Sugulle M, Jones S, Tilaganathan B, Eils R, Huppertz B, El-Heliebi A, Staff AC, Müller DN, Dechend R, Gauster M, Ishaque N, Herse F. Disturbed trophoblast transition links preeclampsia progression from placenta to the maternal syndrome. bioRxiv. 2022 Oct 10:2022.10.10.511539. doi: 10.1101/2022.10.10.511539.

364. Oplawski M, Dziobek K, Zmarzły N, Grabarek B, Tomala B, Leśniak E, Adwent I, Januszyk P, Dąbruś D, Boroń D. Evaluation of Changes in the Expression Pattern of EDIL3 in Different Grades of Endometrial Cancer. Curr Pharm Biotechnol. 2019;20(6):483-488. doi: 10.2174/1389201020666190408112822. PMID: 30961491; PMCID: PMC6806535.

365. Tucić M, Stamenković V, Andjus P. The Extracellular Matrix Glycoprotein Tenascin C and Adult Neurogenesis. Front Cell Dev Biol. 2021 Apr 29; 9:674199. doi: 10.3389/fcell.2021.674199. PMID: 33996833; PMCID: PMC8117239.

366. Vicente CM, Lima MA, Yates EA, Nader HB, Toma L. Enhanced tumorigenic potential of colorectal cancer cells by extracellular sulfatases. Mol Cancer Res. 2015 Mar;13(3):510-23. doi: 10.1158/1541-7786.MCR-14-0372. Epub 2014 Dec 4. PMID: 25477293.

367. Tong M, Jun T, Nie Y, Hao J, Fan D. The Role of the Slit/Robo Signaling Pathway. J Cancer. 2019 Jun 2;10(12):2694-2705. doi: 10.7150/jca.31877. PMID: 31258778; PMCID: PMC6584916.

368. Papapetropoulos A, Marczin N, Mora G, Milici A, Murad F, Catravas JD. Regulation of vascular smooth muscle soluble guanylate cyclase activity, mRNA, and protein levels by cAMP-elevating agents. Hypertension. 1995 Oct;26(4):696-704. doi: 10.1161/01.hyp.26.4.696. PMID: 7558233.

369. Cho H, Park C, Hwang IY, Han SB, Schimel D, Despres D, Kehrl JH. Rgs5 targeting leads to chronic low blood pressure and a lean body habitus. Mol Cell Biol. 2008 Apr;28(8):2590-7. doi: 10.1128/MCB.01889-07. Epub 2008 Feb 11. PMID: 18268011; PMCID: PMC2293121.

370. Holobotovskyy V, Manzur M, Tare M, Burchell J, Bolitho E, Viola H, Hool LC, Arnolda LF, McKitrick DJ, Ganss R. Regulator of G-protein signaling 5 controls blood pressure homeostasis and vessel wall remodeling. Circ Res. 2013 Mar 1;112(5):781-91. doi: 10.1161/CIRCRESAHA.111.300142. Epub 2013 Jan 9. PMID: 23303165.

371. Wang CQ, Choy FC, Sanny A, Murakami T, Tan AH, Lam KP. An Inhibitory Role for Human CD96 Endodomain in T Cell Anti-Tumor Responses. Cells. 2023 Jan 13;12(2):309. doi: 10.3390/cells12020309. PMID: 36672244; PMCID: PMC9856660.

372. Georgiev H, Ravens I, Papadogianni G, Bernhardt G. Coming of Age: CD96 Emerges as Modulator of Immune Responses. Front Immunol. 2018 May 17; 9:1072. doi: 10.3389/fimmu.2018.01072. PMID: 29868026; PMCID: PMC5966540.

373. Van Acker HH, Capsomidis A, Smits EL, Van Tendeloo VF. CD56 in the Immune System: More Than a Marker for Cytotoxicity? Front Immunol. 2017 Jul 24; 8:892. doi: 10.3389/fimmu.2017.00892. PMID: 28791027; PMCID: PMC5522883.

374. Gaynor LM, Colucci F. Uterine Natural Killer Cells: Functional Distinctions and Influence on Pregnancy in Humans and Mice. Front Immunol. 2017 Apr 24;8:467. doi: 10.3389/fimmu.2017.00467. PMID: 28484462; PMCID: PMC5402472.

375. Poli A, Michel T, Thérésine M, Andrès E, Hentges F, Zimmer J. CD56bright natural killer (NK) cells: an important NK cell subset. Immunology. 2009 Apr;126(4):458-65. doi: 10.1111/j.1365-2567.2008.03027.x. PMID: 19278419; PMCID: PMC2673358.

376. Cooper MA, Fehniger TA, Caligiuri MA. The biology of human natural killer-cell subsets. Trends Immunol. 2001 Nov;22(11):633-40. doi: 10.1016/s1471-4906(01)02060-9. PMID: 11698225.

377. Meza Guzman LG, Keating N, Nicholson SE. Natural Killer Cells: Tumor Surveillance and Signaling. Cancers (Basel). 2020 Apr 11;12(4):952. doi: 10.3390/cancers12040952. PMID: 32290478; PMCID: PMC7226588.

378. Le Bouteiller P, Barakonyi A, Giustiniani J, Lenfant F, Marie-Cardine A, Aguerre-Girr M, Rabot M, Hilgert I, Mami-Chouaib F, Tabiasco J, Boumsell L, Bensussan A. Engagement of CD160 receptor by HLA-C is a triggering mechanism used by circulating natural killer (NK) cells to mediate cytotoxicity. Proc Natl Acad Sci U S A. 2002 Dec 24;99(26):16963-8. doi: 10.1073/pnas.012681099. Epub 2002 Dec 16. PMID: 12486241; PMCID: PMC139252.

379. He W, Wang B, Li Q, Yao Q, Jia X, Song R, Li S, Zhang JA. Aberrant Expressions of Co-stimulatory and Co-inhibitory Molecules in Autoimmune Diseases. Front Immunol. 2019 Feb 20; 10:261. doi: 10.3389/fimmu.2019.00261. PMID: 30842773; PMCID: PMC6391512.

380. Hu X, Li YQ, Li QG, Ma YL, Peng JJ, Cai SJ. ITGAE Defines CD8+ Tumor-Infiltrating Lymphocytes Predicting a better Prognostic Survival in Colorectal Cancer. EBioMedicine. 2018 Sep; 35:178-188. doi: 10.1016/j.ebiom.2018.08.003. Epub 2018 Aug 9. PMID: 30100393; PMCID: PMC6154785.

381. Corgnac S, Boutet M, Kfoury M, Naltet C, Mami-Chouaib F. The Emerging Role of CD8+ Tissue Resident Memory T (TRM) Cells in Antitumor Immunity: A Unique Functional Contribution of the CD103 Integrin. Front Immunol. 2018 Aug 15; 9:1904. doi: 10.3389/fimmu.2018.01904. PMID: 30158938; PMCID: PMC6104123.

382. Perera Molligoda Arachchige AS. Human NK cells: From development to effector functions. Innate Immunity. 2021 Mar;27(3):212-229. doi: 10.1177/17534259211001512.

383. Yang C, Blaize G, Marrocco R, Rouquié N, Bories C, Gador M, Mélique S, Joulia E, Benamar M, Dejean AS, Daniels-Treffandier H, Love PE, Fazilleau N, Saoudi A, Lesourne R. THEMIS enhances the magnitude of normal and neuroinflammatory type 1 immune responses by promoting TCR-independent signals. Sci Signal. 2022 Jul 12;15(742): eabl5343. doi: 10.1126/scisignal.abl5343. Epub 2022 Jul 12. PMID: 35857631.

384. Mehta M, Brzostek J, Chen EW, Tung DWH, Chen S, Sankaran S, Yap J, Rybakin V, Gascoigne NRJ. Themis-associated phosphatase activity controls signaling in T cell development. Proc Natl Acad Sci U S A. 2018 Nov 27;115(48):E11331-E11340. doi: 10.1073/pnas.1720209115. Epub 2018 Nov 9. PMID: 30413615; PMCID: PMC6275480.

385. Mariuzza RA, Agnihotri P, Orban J. The structural basis of T-cell receptor (TCR) activation: An enduring enigma. J Biol Chem. 2020 Jan 24;295(4):914-925. doi: 10.1074/jbc.REV119.009411. Epub 2019 Dec 17. PMID: 31848223; PMCID: PMC6983839.

386. Bachmann MF, Barner M, Kopf M. CD2 sets quantitative thresholds in T cell activation. J Exp Med. 1999 Nov 15;190(10):1383-92. doi: 10.1084/jem.190.10.1383. PMID: 10562314; PMCID: PMC2195700.

387. Lissauer D, Eldershaw SA, Inman CF, Coomarasamy A, Moss PA, Kilby MD. Progesterone promotes maternal-fetal tolerance by reducing human maternal T-cell polyfunctionality and inducing a specific cytokine profile. Eur J Immunol. 2015 Oct;45(10):2858-72. doi: 10.1002/eji.201445404. Epub 2015 Aug 28. PMID: 26249148; PMCID: PMC4833190.

388. Hu Y, Shao X, Xing L, Li X, Nonis GM, Koelwyn GJ, Zhang X, Sin DD. Single-Cell Sequencing of Lung Macrophages and Monocytes Reveals Novel Therapeutic Targets in

COPD. Cells. 2023 Dec 5;12(24):2771. doi: 10.3390/cells12242771. PMID: 38132091; PMCID: PMC10741950.

389. Sprenkeler EGG, Zandstra J, van Kleef ND, Goetschalckx I, Verstegen B, Aarts CEM, Janssen H, Tool ATJ, van Mierlo G, van Bruggen R, Jongerius I, Kuijpers TW. S100A8/A9 Is a Marker for the Release of Neutrophil Extracellular Traps and Induces Neutrophil Activation. Cells. 2022 Jan 11;11(2):236. doi: 10.3390/cells11020236. PMID: 35053354; PMCID: PMC8773660.

390. Presicce P, Senthamaraikannan P, Alvarez M, Rueda CM, Cappelletti M, Miller LA, Jobe AH, Chougnet CA, Kallapur SG. Neutrophil recruitment and activation in decidua with intra-amniotic IL-1beta in the preterm rhesus macaque. Biol Reprod. 2015 Feb;92(2):56. doi: 10.1095/biolreprod.114.124420. Epub 2014 Dec 23. PMID: 25537373; PMCID: PMC4342792.

391. da Silva IV, Garra S, Calamita G, Soveral G. The Multifaceted Role of Aquaporin-9 in Health and Its Potential as a Clinical Biomarker. Biomolecules. 2022 Jun 27;12(7):897. doi: 10.3390/biom12070897. PMID: 35883453; PMCID: PMC9313442.

392. Shi Y, Yasui M, Hara-Chikuma M. AQP9 transports lactate in tumor-associated macrophages to stimulate an M2-like polarization that promotes colon cancer progression. Biochem Biophys Rep. 2022 Aug 7;31:101317. doi: 10.1016/j.bbrep.2022.101317. PMID: 35967760; PMCID: PMC9372591.

393. Vondra S, Höbler AL, Lackner AI, Raffetseder J, Mihalic ZN, Vogel A, Saleh L, Kunihs V, Haslinger P, Wahrmann M, Husslein H, Oberle R, Kargl J, Haider S, Latos P, Schabbauer G, Knöfler M, Ernerudh J, Pollheimer J. The human placenta shapes the phenotype of decidual macrophages. Cell Rep. 2023 Mar 28;42(3):112285. doi: 10.1016/j.celrep.2023.112285. Epub 2023 Mar 13. Erratum for: Cell Rep. 2023 Jan 31;42(1):111977. PMID: 36917611.

394. Mizutani T, Orisaka M, Miyazaki Y, Morichika R, Uesaka M, Miyamoto K, Yoshida Y. Inhibition of YAP/TAZ-TEAD activity induces cytotrophoblast differentiation into syncytiotrophoblast in human trophoblast. Mol Hum Reprod. 2022 Sep 29;28(10): gaac032. doi: 10.1093/molehr/gaac032. PMID: 35993908.

395. Haider S, Meinhardt G, Saleh L, Kunihs V, Gamperl M, Kaindl U, Ellinger A, Burkard TR, Fiala C, Pollheimer J, Mendjan S, Latos PA, Knöfler M. Self-Renewing Trophoblast Organoids Recapitulate the Developmental Program of the Early Human Placenta. Stem Cell Reports. 2018 Aug 14;11(2):537-551. doi: 10.1016/j.stemcr.2018.07.004. Epub 2018 Aug 2. PMID: 30078556; PMCID: PMC6092984.

396. Barker N, van Es JH, Kuipers J, Kujala P, van den Born M, Cozijnsen M, Haegebarth A, Korving J, Begthel H, Peters PJ, Clevers H. Identification of stem cells in small intestine and colon by marker gene Lgr5. Nature. 2007 Oct 25;449(7165):1003-7. doi: 10.1038/nature06196. Epub 2007 Oct 14. PMID: 17934449.

397. Soares E, Zhou H. Master regulatory role of p63 in epidermal development and disease. Cell Mol Life Sci. 2018 Apr;75(7):1179-1190. doi: 10.1007/s00018-017-2701-z. Epub 2017 Nov 4. PMID: 29103147; PMCID: PMC5843667.

398. Hishida T, Naito K, Osada S, Nishizuka M, Imagawa M. peg10, an imprinted gene, plays a crucial role in adipocyte differentiation. FEBS Lett. 2007 Sep 4;581(22):4272-8. doi: 10.1016/j.febslet.2007.07.074. Epub 2007 Aug 10. PMID: 17707377.

399. Ray Chaudhuri A, Nussenzweig A. The multifaceted roles of PARP1 in DNA repair and chromatin remodelling. Nat Rev Mol Cell Biol. 2017 Oct;18(10):610-621. doi: 10.1038/nrm.2017.53. Epub 2017 Jul 5. PMID: 28676700; PMCID: PMC6591728.

400. Pavlakis E, Chiotaki R, Chalepakis G. The role of Fras1/Frem proteins in the structure and function of basement membrane. Int J Biochem Cell Biol. 2011 Apr;43(4):487-95. doi: 10.1016/j.biocel.2010.12.016. Epub 2010 Dec 21. PMID: 21182980.

401. Liu Y, Fan X, Wang R, Lu X, Dang YL, Wang H, Lin HY, Zhu C, Ge H, Cross JC, Wang H. Single-cell RNA-seq reveals the diversity of trophoblast subtypes and patterns of differentiation in the human placenta. Cell Res. 2018 Aug;28(8):819-832. doi: 10.1038/s41422-018-0066-y. Epub 2018 Jul 24. PMID: 30042384; PMCID: PMC6082907.

402. Ferreira LMR, Meissner TB, Tilburgs T, Strominger JL. HLA-G: At the Interface of Maternal-Fetal Tolerance. Trends Immunol. 2017 Apr;38(4):272-286. doi: 10.1016/j.it.2017.01.009. Epub 2017 Mar 6. PMID: 28279591.

403. Tilburgs T, Evans JH, Crespo ÂC, Strominger JL. The HLA-G cycle provides for both NK tolerance and immunity at the maternal-fetal interface. Proc Natl Acad Sci U S A. 2015 Oct 27;112(43):13312-7. doi: 10.1073/pnas.1517724112. Epub 2015 Oct 12. PMID: 26460007; PMCID: PMC4629323.

404. Founds SA, Fallert-Junecko B, Reinhart TA, Conley YP, Parks WT. LAIR2 localizes specifically to sites of extravillous trophoblast invasion. Placenta. 2010 Oct;31(10):880-5. doi: 10.1016/j.placenta.2010.07.005. Epub 2010 Aug 6. PMID: 20692035.

405. Broekhuizen M, Danser AHJ, Reiss IKM, Merkus D. The Function of the Kynurenine Pathway in the Placenta: A Novel Pharmacotherapeutic Target? Int J Environ Res Public Health. 2021 Nov 3;18(21):11545. doi: 10.3390/ijerph182111545. PMID: 34770059; PMCID: PMC8582682.

406. Schmitz T, Souil E, Hervé R, Nicco C, Batteux F, Germain G, Cabrol D, Evain-Brion D, Leroy MJ, Méhats C. PDE4 inhibition prevents preterm delivery induced by an intrauterine inflammation. J Immunol. 2007 Jan 15;178(2):1115-21. doi: 10.4049/jimmunol.178.2.1115. PMID: 17202375.

407. Crocetti L, Floresta G, Cilibrizzi A, Giovannoni MP. An Overview of PDE4 Inhibitors in Clinical Trials: 2010 to Early 2022. Molecules. 2022 Aug 4;27(15):4964. doi: 10.3390/molecules27154964. PMID: 35956914; PMCID: PMC9370432.

408. Liudkovska V, Krawczyk PS, Brouze A, Gumińska N, Wegierski T, Cysewski D, Mackiewicz Z, Ewbank JJ, Drabikowski K, Mroczek S, Dziembowski A. TENT5 cytoplasmic noncanonical poly(A) polymerases regulate the innate immune response in animals. Sci Adv. 2022 Nov 18;8(46): eadd9468. doi: 10.1126/sciadv.add9468. Epub 2022 Nov 16. PMID: 36383655; PMCID: PMC9668313.

409. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun. 2019 Apr 3;10(1):1523. doi: 10.1038/s41467-019-09234-6. PMID: 30944313; PMCID: PMC6447622.

410. Krendl C, Shaposhnikov D, Rishko V, Ori C et al. GATA2/3-TFAP2A/C transcription factor network couples human pluripotent stem cell differentiation to trophectoderm with repression

of pluripotency. Proc Natl Acad Sci U S A. 2017 Nov 7;114(45): E9579-E9588. doi: 10.1073/pnas.1708341114. Epub 2017 Oct 25. PMID: 29078328; PMCID: PMC5692555.

411. Hubert MA, Sherritt SL, Bachurski CJ, Handwerger S. Involvement of transcription factor NR2F2 in human trophoblast differentiation. PLoS One. 2010 Feb 25;5(2):e9417. doi: 10.1371/journal.pone.0009417. PMID: 20195529; PMCID: PMC2828470.

412. Gambino YP, Maymó JL, Pérez-Pérez A, Dueñas JL, Sánchez-Margalet V, Calvo JC, Varone CL. 17Beta-estradiol enhances leptin expression in human placental cells through genomic and nongenomic actions. Biol Reprod. 2010 Jul;83(1):42-51. doi: 10.1095/biolreprod.110.083535. Epub 2010 Mar 17. PMID: 20237333.

413. O'Neil JS, Burow ME, Green AE, McLachlan JA, Henson MC. Effects of estrogen on leptin gene promoter activation in MCF-7 breast cancer and JEG-3 choriocarcinoma cells: selective regulation via estrogen receptors alpha and beta. Mol Cell Endocrinol. 2001 May 15;176(1-2):67-75. doi: 10.1016/s0303-7207(01)00473-7. PMID: 11369444.

414. Köhler B, Lin L, Ferraz-de-Souza B, Wieacker P, Heidemann P, Schröder V, Biebermann H, Schnabel D, Grüters A, Achermann JC. Five novel mutations in steroidogenic factor 1 (SF1, NR5A1) in 46,XY patients with severe underandrogenization but without adrenal insufficiency. Hum Mutat. 2008 Jan;29(1):59-64. doi: 10.1002/humu.20588. PMID: 17694559; PMCID: PMC2359628.

415. Berns DS, DeNardo LA, Pederick DT, Luo L. Teneurin-3 controls topographic circuit assembly in the hippocampus. Nature. 2018 Feb 15;554(7692):328-333. doi: 10.1038/nature25463. Epub 2018 Feb 7. PMID: 29414938; PMCID: PMC7282895.

416. MacDonald TM, Walker SP, Hiscock R, Cannon P, Harper A, Murray E, Hui L, Dane K, Middleton A, Kyritsis V, de Alwis N, Hannan NJ, Tong S, Kaitu'u-Lino TJ. Circulating Delta-like homolog 1 (DLK1) at 36 weeks is correlated with birthweight and is of placental origin. Placenta. 2020 Feb; 91:24-30. doi: 10.1016/j.placenta.2020.01.003. Epub 2020 Jan 8. PMID: 32174303.

417. Blei DM, Ng AY, Jordan MI. Latent Dirichlet Allocation. J Mach Learn Res. 2003; 3:993-1022.

418. Nilsson M, Malmgren H, Samiotaki M, Kwiatkowski M, Chowdhary BP, Landegren U. Padlock probes: circularizing oligonucleotides for localized DNA detection. Science. 1994 Sep 30;265(5181):2085-8. doi: 10.1126/science.7522346. PMID: 7522346.

419. Huang M, Anand S, Murphy EA, Desgrosellier JS, Stupack DG, Shattil SJ, Schlaepfer DD, Cheresh DA. EGFR-dependent pancreatic carcinoma cell metastasis through Rap1 activation. Oncogene. 2012 May 31;31(22):2783-93. doi: 10.1038/onc.2011.450. Epub 2011 Oct 3. PMID: 21963850; PMCID: PMC3711644.

420. Varberg KM, Dominguez EM, Koseva B, Varberg JM et al. Extravillous trophoblast cell lineage development is associated with active remodeling of the chromatin landscape. Nat Commun. 2023 Aug 10;14(1):4826. doi: 10.1038/s41467-023-40424-5. PMID: 37563143; PMCID: PMC10415281.

421. Meinhardt G, Haider S, Kunihs V, Saleh L, Pollheimer J, Fiala C, Hetey S, Feher Z, Szilagyi A, Than NG, Knöfler M. Pivotal role of the transcriptional co-activator YAP in trophoblast stemness of the developing human placenta. Proc Natl Acad Sci U S A. 2020 Jun

16;117(24):13562-13570. doi: 10.1073/pnas.2002630117. Epub 2020 Jun 1. PMID: 32482863; PMCID: PMC7306800

422. Shukla V, Moreno-Irusta A, Varberg KM, Kuna M, Iqbal K, Galligos AM, Aplin JD, Choudhury RH, Okae H, Arima T, Soares MJ. NOTUM-MEDIATED WNT SILENCING DRIVES EXTRAVILLOUS TROPHOBLAST CELL LINEAGE DEVELOPMENT. bioRxiv [Preprint]. 2024 Feb 14:2024.02.13.579974. doi: 10.1101/2024.02.13.579974. PMID: 38405745; PMCID: PMC10888853.

423. Haider S, Lackner AI, Dietrich B, Kunihs V, Haslinger P, Meinhardt G, Maxian T, Saleh L, Fiala C, Pollheimer J, Latos PA, Knöfler M. Transforming growth factor-β signaling governs the differentiation program of extravillous trophoblasts in the developing human placenta. Proc Natl Acad Sci U S A. 2022 Jul 12;119(28):e2120667119. doi: 10.1073/pnas.2120667119. Epub 2022 Jul 6. PMID: 35867736; PMCID: PMC9282384.

424. Willert K, Nusse R. Wnt proteins. Cold Spring Harb Perspect Biol. 2012 Sep 1;4(9): a007864. doi: 10.1101/cshperspect.a007864. PMID: 22952392; PMCID: PMC3428774.

425. Dietrich B, Haider S, Meinhardt G, Pollheimer J, Knöfler M. WNT and NOTCH signaling in human trophoblast development and differentiation. Cell Mol Life Sci. 2022 May 13;79(6):292. doi: 10.1007/s00018-022-04285-3. PMID: 35562545; PMCID: PMC9106601.

426. Pang H, Lei D, Chen T, Liu Y, Fan C. The Enzyme 15-Hydroxyprostaglandin Dehydrogenase Inhibits a Shift to the Mesenchymal Pattern of Trophoblasts and Decidual Stromal Cells Accompanied by Prostaglandin Transporter in Preeclampsia. Int J Mol Sci. 2023 Mar 7;24(6):5111. doi: 10.3390/ijms24065111. PMID: 36982197; PMCID: PMC10049104.

427. Meinhardt G, Haider S, Haslinger P, Proestling K, Fiala C, Pollheimer J, Knöfler M. Wnt-dependent T-cell factor-4 controls human etravillous trophoblast motility. Endocrinology. 2014 May;155(5):1908-20. doi: 10.1210/en.2013-2042. Epub 2014 Feb 26. PMID: 24605829.

428. Sonderegger S, Haslinger P, Sabri A, Leisser C, Otten JV, Fiala C, Knöfler M. Wingless (Wnt)-3A induces trophoblast migration and matrix metalloproteinase-2 secretion through canonical Wnt signaling and protein kinase B/AKT activation. Endocrinology. 2010 Jan;151(1):211-20. doi: 10.1210/en.2009-0557. Epub 2009 Nov 3. PMID: 19887570; PMCID: PMC2974214.

429. Zhu JY, Pang ZJ, Yu YH. Regulation of trophoblast invasion: the role of matrix metalloproteinases. Rev Obstet Gynecol. 2012;5(3-4):e137-43. PMID: 23483768; PMCID: PMC3594863.

430. Ray S, Saha A, Ghosh A, Roy N, Kumar RP, Meinhardt G, Mukerjee A, Gunewardena S, Kumar R, Knöfler M, Paul S. Hippo signaling cofactor, WWTR1, at the crossroads of human trophoblast progenitor self-renewal and differentiation. Proc Natl Acad Sci U S A. 2022 Sep 6;119(36):e2204069119. doi: 10.1073/pnas.2204069119. Epub 2022 Aug 29. PMID: 36037374; PMCID: PMC9457323.

431. Mund A, Coscia F, Kriston A, Hollandi R, Kovács F, Brunner AD, Migh E, Schweizer L, Santos A, Bzorek M, Naimy S, Rahbek-Gjerdrum LM, Dyring-Andersen B, Bulkescher J, Lukas C, Eckert MA, Lengyel E, Gnann C, Lundberg E, Horvath P, Mann M. Deep Visual Proteomics defines single-cell identity and heterogeneity. Nat Biotechnol. 2022 Aug;40(8):1231-1240. doi: 10.1038/s41587-022-01302-5. Epub 2022 May 19. PMID: 35590073; PMCID: PMC9371970.

432. Varberg KM, Iqbal K, Muto M, Simon ME, Scott RL, Kozai K, Choudhury RH, Aplin JD, Biswell R, Gibson M, Okae H, Arima T, Vivian JL, Grundberg E, Soares MJ. ASCL2 reciprocally controls key trophoblast lineage decisions during hemochorial placenta development. Proc Natl Acad Sci U S A. 2021 Mar 9;118(10): e2016517118. doi: 10.1073/pnas.2016517118. PMID: 33649217; PMCID: PMC7958375.

433. Lebbink RJ, Raynal N, de Ruiter T, Bihan DG, Farndale RW, Meyaard L. Identification of multiple potent binding sites for human leukocyte associated Ig-like receptor LAIR on collagens II and III. Matrix Biol. 2009 May;28(4):202-10. doi: 10.1016/j.matbio.2009.03.005. Epub 2009 Apr 2. PMID: 19345263.

434. Gauster M, Siwetz M, Huppertz B. Fusion of villous trophoblast can be visualized by localizing active caspase 8. Placenta. 2009 Jun;30(6):547-50. doi: 10.1016/j.placenta.2009.03.007. Epub 2009 Apr 5. PMID: 19345995.

435. Sanders LN, Schoenhard JA, Saleh MA, Mukherjee A, Ryzhov S, McMaster WG Jr, Nolan K, Gumina RJ, Thompson TB, Magnuson MA, Harrison DG, Hatzopoulos AK. BMP Antagonist Gremlin 2 Limits Inflammation After Myocardial Infarction. Circ Res. 2016 Jul 22;119(3):434-49. doi: 10.1161/CIRCRESAHA.116.308700. Epub 2016 Jun 9. PMID: 27283840; PMCID: PMC4961528.

436. Salazar VS, Gamer LW, Rosen V. BMP signalling in skeletal development, disease and repair. Nat Rev Endocrinol. 2016 Apr;12(4):203-21. doi: 10.1038/nrendo.2016.12. Epub 2016 Feb 19. PMID: 26893264.

437. Bai L, Chang HM, Cheng JC, Chu G, Leung PCK, Yang G. ALK2/ALK3-BMPR2/ACVR2A Mediate BMP2-Induced Downregulation of Pentraxin 3 Expression in Human Granulosa-Lutein Cells. Endocrinology. 2017 Oct 1;158(10):3501-3511. doi: 10.1210/en.2017-00436. PMID: 28977600.

438. Snyder SK, Wessner DH, Wessells JL, Waterhouse RM, Wahl LM, Zimmermann W, Dveksler GS. Pregnancy-specific glycoproteins function as immunomodulators by inducing secretion of IL-10, IL-6 and TGF-beta1 by human monocytes. Am J Reprod Immunol. 2001 Apr;45(4):205-16. doi: 10.1111/j.8755-8920.2001.450403.x. PMID: 11327547.

439. Timganova VP, Zamorina SA, Litvinova LS, Todosenko NM, Bochkova MS, Khramtsov PV, Rayev MB. The effects of human pregnancy-specific β1-glycoprotein preparation on Th17 polarization of CD4+ cells and their cytokine profile. BMC Immunol. 2020 Oct 30;21(1):56. doi: 10.1186/s12865-020-00385-6. PMID: 33126863; PMCID: PMC7602336.

440. Blois SM, Sulkowski G, Tirado-González I, Warren J, Freitag N, Klapp BF, Rifkin D, Fuss I, Strober W, Dveksler GS. Pregnancy-specific glycoprotein 1 (PSG1) activates TGF-β and prevents dextran sodium sulfate (DSS)-induced colitis in mice. Mucosal Immunol. 2014 Mar;7(2):348-58. doi: 10.1038/mi.2013.53. Epub 2013 Aug 14. PMID: 23945545; PMCID: PMC3844031.

441. Zamorina SA, Litvinova LS, Yurova KA, Dunets NA, Khaziakhmatova OG, Timganova VP, Bochkova MS, Khramtsov PV, Rayev MB. The effect of pregnancy-specific β1-glycoprotein 1 on the transcription factor FOXP3 expression by immunocompetent cells. Dokl Biochem Biophys. 2016 Sep;470(1):361-363. doi: 10.1134/S1607672916050161. Epub 2016 Nov 6. PMID: 27817017.

442. Rayev MB, Litvinova LS, Yurova KA, Dunets NA, Khaziakhmatova OG, Timganova VP, Bochkova MS, Khramtsov PV, Zamorina SA. Role of the pregnancy-specific glycoprotein in regulation of the cytokine and chemokine profiles of intact mononuclear cells. Dokl Biol Sci. 2017 Jul;475(1):180-182. doi: 10.1134/S001249661704007X. Epub 2017 Sep 1. PMID: 28861873.

443. Zhang L, Zhou A, Zhu S, Min L, Liu S, Li P, Zhang S. The role of GTPase-activating protein ARHGAP26 in human cancers. Mol Cell Biochem. 2022 Jan;477(1):319-326. doi: 10.1007/s11010-021-04274-3. Epub 2021 Oct 30. PMID: 34716859; PMCID: PMC8755663.

444. Christians JK, Beristain AG. ADAM12 and PAPP-A: Candidate regulators of trophoblast invasion and first trimester markers of healthy trophoblasts. Cell Adh Migr. 2016 Mar 3;10(1-2):147-53. doi: 10.1080/19336918.2015.1083668. Epub 2015 Sep 29. PMID: 26417939; PMCID: PMC4853044.

445. Aghababaei M, Hogg K, Perdu S, Robinson WP, Beristain AG. ADAM12-directed ectodomain shedding of E-cadherin potentiates trophoblast fusion. Cell Death Differ. 2015 Dec;22(12):1970-84. doi: 10.1038/cdd.2015.44. Epub 2015 Apr 24. PMID: 25909890; PMCID: PMC4816105.

446. Gauster M, Moser G, Orendi K, Huppertz B. Factors involved in regulating trophoblast fusion: potential role in the development of preeclampsia. Placenta. 2009 Mar;30 Suppl A:S49-54. doi: 10.1016/j.placenta.2008.10.011. Epub 2008 Nov 21. PMID: 19027159.

447. Saha B, Ganguly A, Home P, Bhattacharya B, Ray S, Ghosh A, Rumi MAK, Marsh C, French VA, Gunewardena S, Paul S. TEAD4 ensures postimplantation development by promoting trophoblast self-renewal: An implication in early human pregnancy loss. Proc Natl Acad Sci U S A. 2020 Jul 28;117(30):17864-17875. doi: 10.1073/pnas.2002449117. Epub 2020 Jul 15. PMID: 32669432; PMCID: PMC7395512.

448. Dong C, Fu S, Karvas RM, Chew B, Fischer LA, Xing X, Harrison JK, Popli P, Kommagani R, Wang T, Zhang B, Theunissen TW. A genome-wide CRISPR-Cas9 knockout screen identifies essential and growth-restricting genes in human trophoblast stem cells. Nat Commun. 2022 May 10;13(1):2548. doi: 10.1038/s41467-022-30207-9. PMID: 35538076; PMCID: PMC9090837.

449. Yang N, Friedl A. Syndecan-1-Induced ECM Fiber Alignment Requires Integrin αvβ3 and Syndecan-1 Ectodomain and Heparan Sulfate Chains. PLoS One. 2016 Feb 24;11(2):e0150132. doi: 10.1371/journal.pone.0150132. PMID: 26909794; PMCID: PMC4766302.

450. Zeng YT, Liu WF, Zheng PS, Li S. GDF15 deficiency hinders human trophoblast invasion to mediate pregnancy loss through downregulating Smad1/5 phosphorylation. iScience. 2023 Sep 13;26(10):107902. doi: 10.1016/j.isci.2023.107902. PMID: 37766993; PMCID: PMC10520888.

451. Paladini F, Fiorillo MT, Tedeschi V, Mattorre B, Sorrentino R. The Multifaceted Nature of Aminopeptidases ERAP1, ERAP2, and LNPEP: From Evolution to Disease. Front Immunol. 2020 Jul 23; 11:1576. doi: 10.3389/fimmu.2020.01576. PMID: 32793222; PMCID: PMC7390905.

452. Vadon-Le Goff S, Hulmes DJ, Moali C. BMP-1/tolloid-like proteinases synchronize matrix assembly with growth factor activation to promote morphogenesis and tissue remodeling. Matrix Biol. 2015 May-Jul;44-46:14-23. doi: 10.1016/j.matbio.2015.02.006. Epub 2015 Feb 18. PMID: 25701650.

453. https://www.10xgenomics.com/support/software/space-ranger/latest

454. Bai Y, Ying Y. The Post-translational Modifications of Smurf2 in TGF-β Signaling. Front Mol Biosci. 2020 Jul 7;7:128. doi: 10.3389/fmolb.2020.00128. PMID: 32733916; PMCID: PMC7358609.

455. Regis S, Dondero A, Caliendo F, Bottino C, Castriconi R. NK Cell Function Regulation by TGF-β-Induced Epigenetic Mechanisms. Front Immunol. 2020 Feb 25;11:311. doi: 10.3389/fimmu.2020.00311. PMID: 32161594; PMCID: PMC7052483

456. Yang Q, Chen SP, Zhang XP, Wang H, Zhu C, Lin HY. Smurf2 participates in human trophoblast cell invasion by inhibiting TGF-beta type I receptor. J Histochem Cytochem. 2009 Jun;57(6):605-12. doi: 10.1369/jhc.2009.953166. Epub 2009 Mar 2. PMID: 19255252; PMCID: PMC2690412.

457. Holder BS, Tower CL, Jones CJ, Aplin JD, Abrahams VM. Heightened pro-inflammatory effect of preeclamptic placental microvesicles on peripheral blood immune cells in humans. Biol Reprod. 2012 Apr 5;86(4):103. doi: 10.1095/biolreprod.111.097014. PMID: 22205696.

458. Straszewski-Chavez SL, Visintin IP, Karassina N, Los G, Liston P, Halaban R, Fadiel A, Mor G. XAF1 mediates tumor necrosis factor-alpha-induced apoptosis and X-linked inhibitor of apoptosis cleavage by acting through the mitochondrial pathway. J Biol Chem. 2007 Apr 27;282(17):13059-72. doi: 10.1074/jbc.M609038200. Epub 2007 Feb 28. PMID: 17329253.

459. Wei D, Yiyuan C, Qian L, Jianhua L, Yazhuo Z, Hua G. The absence of PRDM2 involved the tumorigenesis of somatotroph adenomas through regulating c-Myc. Gene. 2020 May 5;737:144456. doi: 10.1016/j.gene.2020.144456. Epub 2020 Feb 7. PMID: 32044406.

460. Friedman AD. Cell cycle and developmental control of hematopoiesis by Runx1. J Cell Physiol. 2009 Jun;219(3):520-4. doi: 10.1002/jcp.21738. PMID: 19235904; PMCID: PMC4741264.

461. Kannan A, Beal JR, Neff AM, Bagchi MK, Bagchi IC. Runx1 regulates critical factors that control uterine angiogenesis and trophoblast differentiation during placental development. PNAS Nexus. 2023 Jun 30;2(7):pgad215. doi: 10.1093/pnasnexus/pgad215. PMID: 37416873; PMCID: PMC10321400.

462. Hneino M, François A, Buard V, Tarlet G, Abderrahmani R, Blirando K, Hoodless PA, Benderitter M, Milliat F. The TGF-β/Smad repressor TG-interacting factor 1 (TGIF1) plays a role in radiation-induced intestinal injury independently of a Smad signaling pathway. PLoS One. 2012;7(5):e35672. doi: 10.1371/journal.pone.0035672. Epub 2012 May 2. PMID: 22567107; PMCID: PMC3342305.

463. Demange C, Ferrand N, Prunier C, Bourgeade MF, Atfi A. A model of partnership co-opted by the homeodomain protein TGIF and the Itch/AIP4 ubiquitin ligase for effective execution of TNF-alpha cytotoxicity. Mol Cell. 2009 Dec 25;36(6):1073-85. doi: 10.1016/j.molcel.2009.12.009. PMID: 20064471.

464. Chuang CH, Wang WJ, Li CF, Ko CY, Chou YH, Chuu CP, Cheng TL, Wang JM. The combination of the prodrugs perforin-CEBPD and perforin-granzyme B efficiently enhances the activation of caspase signaling and kills prostate cancer. Cell Death Dis. 2014 May 8;5(5):e1220. doi: 10.1038/cddis.2014.106. PMID: 24810056; PMCID: PMC4047860.

465. Pan YC, Li CF, Ko CY, Pan MH, Chen PJ, Tseng JT, Wu WC, Chang WC, Huang AM, Sterneck E, Wang JM. CEBPD reverses RB/E2F1-mediated gene repression and participates in HMDB-induced apoptosis of cancer cells. Clin Cancer Res. 2010 Dec 1;16(23):5770-80. doi: 10.1158/1078-0432.CCR-10-1025. Epub 2010 Oct 22. PMID: 20971808; PMCID: PMC7325841.

466. Wortel IMN, van der Meer LT, Kilberg MS, van Leeuwen FN. Surviving Stress: Modulation of ATF4-Mediated Stress Responses in Normal and Malignant Cells. Trends Endocrinol Metab. 2017 Nov;28(11):794-806. doi: 10.1016/j.tem.2017.07.003. Epub 2017 Aug 7. PMID: 28797581; PMCID: PMC5951684.

467. Lian IA, Løset M, Mundal SB, Fenstad MH, Johnson MP, Eide IP, Bjørge L, Freed KA, Moses EK, Austgulen R. Increased endoplasmic reticulum stress in decidual tissue from pregnancies complicated by fetal growth restriction with and without pre-eclampsia. Placenta. 2011 Nov;32(11):823-9. doi: 10.1016/j.placenta.2011.08.005. Epub 2011 Sep 9. PMID: 21907405; PMCID: PMC3210381.

468. Armistead B, Kadam L, Drewlo S, Kohan-Ghadr HR. The Role of NFκB in Healthy and Preeclamptic Placenta: Trophoblasts in the Spotlight. Int J Mol Sci. 2020 Mar 5;21(5):1775. doi: 10.3390/ijms21051775. PMID: 32150832; PMCID: PMC7084575.

469. Yang D, Dai F, Yuan M, Zheng Y, Liu S, Deng Z, Tan W, Chen L, Zhang Q, Zhao X, Cheng Y. Role of Transforming Growth Factor-β1 in Regulating Fetal-Maternal Immune Tolerance in Normal and Pathological Pregnancy. Front Immunol. 2021 Aug 31;12:689181. doi: 10.3389/fimmu.2021.689181. PMID: 34531852; PMCID: PMC8438197.

470. Zhang J, Dunk CE, Shynlova O, Caniggia I, Lye SJ. TGFb1 suppresses the activation of distinct dNK subpopulations in preeclampsia. EBioMedicine. 2019 Jan;39:531-539. doi: 10.1016/j.ebiom.2018.12.015. Epub 2018 Dec 19. PMID: 30579870; PMCID: PMC6355656.

471. Gurdol F, Yurdum LM, Ozturk U, Isbilen E, Cakmakoglu B. Association of the CC chemokine receptor 5 (CCR5) polymorphisms with preeclampsia in Turkish women. Arch Gynecol Obstet. 2012 Jul;286(1):51-4. doi: 10.1007/s00404-012-2244-3. PMID: 22314435.

472. Wantania J, Attamimi A, Siswishanto R. A Comparison of 2-Methoxyestradiol Value in Women with Severe Preeclampsia Versus Normotensive Pregnancy. J Clin Diagn Res. 2017 Mar;11(3):QC35-QC38. doi: 10.7860/JCDR/2017/21516.9603. Epub 2017 Mar 1. PMID: 28511459; PMCID: PMC5427385.

473. Shen Z, Wu Y, Chen X, Chang X, Zhou Q, Zhou J, Ying H, Zheng J, Duan T, Wang K. Decreased maternal serum 2-methoxyestradiol levels are associated with the development of preeclampsia. Cell Physiol Biochem. 2014;34(6):2189-99. doi: 10.1159/000369662. Epub 2014 Dec 4. PMID: 25562165.

474. Zhang Y, Wang T, Shen Y, Wang X, Baker PN, Zhao A. 2-Methoxyestradiol deficiency is strongly related to hypertension in early onset severe pre-eclampsia. Pregnancy Hypertens. 2014 Jul;4(3):215-9. doi: 10.1016/j.preghy.2014.04.004. Epub 2014 May 3. PMID: 26104608.

475. Regnault TR, de Vrijer B, Galan HL, Davidsen ML et al. The relationship between transplacental O2 diffusion and placental expression of PlGF, VEGF and their receptors in a placental insufficiency model of fetal growth restriction. J Physiol. 2003 Jul 15;550(Pt 2):641-56. doi: 10.1113/jphysiol.2003.039511. Epub 2003 May 9. PMID: 12740423; PMCID: PMC2343042.

476. Wang Y, Cheng K, Zhou W, Liu H, Yang T, Hou P, Li X. miR-141-5p regulate ATF2 via effecting MAPK1/ERK2 signaling to promote preeclampsia. Biomed Pharmacother. 2019 Jul;115:108953. doi: 10.1016/j.biopha.2019.108953. Epub 2019 May 7. PMID: 31075732.

477. Yang Q, Yin RX, Zhou YJ, Cao XL, Guo T, Chen WX. Association of polymorphisms in the MAFB gene and the risk of coronary artery disease and ischemic stroke: a case-control study. Lipids Health Dis. 2015 Jul 25;14:79. doi: 10.1186/s12944-015-0078-2. PMID: 26204962; PMCID: PMC4513700.

478. Yang Q, Zhou Y, Yin R. Polymorphisms in the MAFB gene are associated with the risk of coronary artery disease and ischemic stroke. J Am Coll Cardiol. 2015 Oct;66(16 Suppl):C139-C140.

479. Kim H. The transcription factor MafB promotes anti-inflammatory M2 polarization and cholesterol efflux in macrophages. Sci Rep. 2017 Aug 8;7(1):7591. doi: 10.1038/s41598-017-07381-8. PMID: 28790455; PMCID: PMC5548719.

480. Lisney AR, Szelinski F, Reiter K, Burmester GR, Rose T, Dörner T. High maternal expression of SIGLEC1 on monocytes as a surrogate marker of a type I interferon signature is a risk factor for the development of autoimmune congenital heart block. Ann Rheum Dis. 2017 Aug;76(8):1476-1480. doi: 10.1136/annrheumdis-2016-210927. Epub 2017 May 13. PMID: 28501799.

481. Macintire K, Tuohey L, Ye L, Palmer K, Gantier M, Tong S, Kaitu'u-Lino TJ. PAPPA2 is increased in severe early onset pre-eclampsia and upregulated with hypoxia. Reprod Fertil Dev. 2014 Jan;26(2):351-7. doi: 10.1071/RD12384. PMID: 23484525.

482. Neuman RI, Alblas van der Meer MM, Nieboer D et al. PAPP-A2 and Inhibin A as Novel Predictors for Pregnancy Complications in Women With Suspected or Confirmed Preeclampsia. J Am Heart Assoc. 2020 Oct 20;9(19):e018219. doi: 10.1161/JAHA.120.018219. Epub 2020 Sep 29. PMID: 32990126; PMCID: PMC7792419.

483. Keikkala E, Forstén J, Ritvos O, Stenman UH, Kajantie E, Hämäläinen E, Räikkönen K, Villa PM, Laivuori H. Serum Inhibin-A and PAPP-A2 in the prediction of pre-eclampsia during the first and second trimesters in high-risk women. Pregnancy Hypertens. 2021 Aug; 25:116-122. doi: 10.1016/j.preghy.2021.05.024. Epub 2021 Jun 2. PMID: 34116346.

484. Ning Z, Wu Z, Zhang F, Yang M, Lu Z, Yu B, Long F, Guo Y, Yang K, Hu G, Zhang Y, Li X, Li L, Lin C. GMEB2 Promotes the Growth of Colorectal Cancer by Activating ADRM1 Transcription and NF-κB Signalling and Is Positively Regulated by the m6A Reader YTHDF1. Cancers (Basel). 2022 Dec 8;14(24):6046. doi: 10.3390/cancers14246046. PMID: 36551532; PMCID: PMC9776391.

485. Zielińska KA, Van Moortel L, Opdenakker G, De Bosscher K, Van den Steen PE. Endothelial Response to Glucocorticoids in Inflammatory Diseases. Front Immunol. 2016 Dec 14;7:592. doi: 10.3389/fimmu.2016.00592. PMID: 28018358; PMCID: PMC5155119.

486. Yu Z, Zhu J, Wang H, Li H, Jin X. Function of BCLAF1 in human disease. Oncol Lett. 2022 Feb;23(2):58. doi: 10.3892/ol.2021.13176. Epub 2021 Dec 22. PMID: 34992690; PMCID: PMC8721854.

487. Kasof GM, Goyal L, White E. Btf, a novel death-promoting transcriptional repressor that interacts with Bcl-2-related proteins. Mol Cell Biol. 1999 Jun;19(6):4390-404. doi: 10.1128/MCB.19.6.4390. PMID: 10330179; PMCID: PMC104398.

488. Sullivan AL, Benner C, Heinz S, Huang W, Xie L, Miano JM, Glass CK. Serum response factor utilizes distinct promoter- and enhancer-based mechanisms to regulate cytoskeletal gene expression in macrophages. Mol Cell Biol. 2011 Feb;31(4):861-75. doi: 10.1128/MCB.00836-10. Epub 2010 Dec 6. PMID: 21135125; PMCID: PMC3028656.

489. Patyal P, Nguyen B, Zhang X, Azhar G, Ameer FS, Verma A, Crane J, Kc G, Che Y, Wei JY. Rho/SRF Inhibitor Modulates Mitochondrial Functions. Int J Mol Sci. 2022 Sep 29;23(19):11536. doi: 10.3390/ijms231911536. PMID: 36232837; PMCID: PMC9570101.

490. Guo Y, Jardin BD, Zhou P, Sethi I, Akerberg BN et al. Hierarchical and stage-specific regulation of murine cardiomyocyte maturation by serum response factor. Nat Commun. 2018 Sep 21;9(1):3837. doi: 10.1038/s41467-018-06347-2. PMID: 30242271; PMCID: PMC6155060.

491. Rodenberg JM, Hoggatt AM, Chen M, Touw K, Jones R, Herring BP. Regulation of serum response factor activity and smooth muscle cell apoptosis by chromodomain helicase DNA-binding protein 8. Am J Physiol Cell Physiol. 2010 Nov;299(5):C1058-67. doi: 10.1152/ajpcell.00080.2010. Epub 2010 Aug 25. PMID: 20739623; PMCID: PMC2980307.

492. Shibuya M. VEGFR and type-V RTK activation and signaling. Cold Spring Harb Perspect Biol. 2013 Oct 1;5(10):a009092. doi: 10.1101/cshperspect.a009092. PMID: 24086040; PMCID: PMC3783052.

493. Fong GH, Rossant J, Gertsenstein M, Breitman ML. Role of the Flt-1 receptor tyrosine kinase in regulating the assembly of vascular endothelium. Nature. 1995 Jul 6;376(6535):66-70. doi: 10.1038/376066a0. PMID: 7596436.

494. Jena MK, Sharma NR, Petitt M, Maulik D, Nayak NR. Pathogenesis of Preeclampsia and Therapeutic Approaches Targeting the Placenta. Biomolecules. 2020 Jun 24;10(6):953. doi: 10.3390/biom10060953. PMID: 32599856; PMCID: PMC7357118.

495. Marzioni D, Todros T, Cardaropoli S, Rolfo A, Lorenzi T, Ciarmela P, Romagnoli R, Paulesu L, Castellucci M. Activating protein-1 family of transcription factors in the human placenta complicated by preeclampsia with and without fetal growth restriction. Placenta. 2010 Oct;31(10):919-27. doi: 10.1016/j.placenta.2010.08.001. Epub 2010 Aug 30. PMID: 20800894.

496. Vandré DD, Ackerman WE 4th, Kniss DA, Tewari AK, Mori M, Takizawa T, Robinson JM. Dysferlin is expressed in human placenta but does not associate with caveolin. Biol Reprod. 2007 Sep;77(3):533-42. doi: 10.1095/biolreprod.107.062190. Epub 2007 Jun 6. PMID: 17554076.

497. Moghaddas Sani H, Zununi Vahed S, Ardalan M. Preeclampsia: A close look at renal dysfunction. Biomed Pharmacother. 2019 Jan;109:408-416. doi: 10.1016/j.biopha.2018.10.082. Epub 2018 Nov 3. PMID: 30399576.

498. Venkatesha S, Toporsian M, Lam C, Hanai J et al. Soluble endoglin contributes to the pathogenesis of preeclampsia. Nat Med. 2006 Jun;12(6):642-9. doi: 10.1038/nm1429. Epub 2006 Jun 4. Erratum in: Nat Med. 2006 Jul;12(7):862. PMID: 16751767.

499. Oujo B, Perez-Barriocanal F, Bernabeu C, Lopez-Novoa JM. Membrane and soluble forms of endoglin in preeclampsia. Curr Mol Med. 2013 Sep;13(8):1345-57. doi: 10.2174/15665240113139990058. PMID: 23826920.

500. Ou Y, Wang SJ, Li D, Chu B, Gu W. Activation of SAT1 engages polyamine metabolism with p53-mediated ferroptotic responses. Proc Natl Acad Sci U S A. 2016 Nov 1;113(44):E6806-E6812. doi: 10.1073/pnas.1607152113. Epub 2016 Oct 3. PMID: 27698118; PMCID: PMC5098629.

501. Gormley M, Ona K, Kapidzic M, Garrido-Gomez T, Zdravkovic T, Fisher SJ. Preeclampsia: novel insights from global RNA profiling of trophoblast subpopulations. Am J Obstet Gynecol. 2017 Aug;217(2):200.e1-200.e17. doi: 10.1016/j.ajog.2017.03.017. Epub 2017 Mar 24. PMID: 28347715.

502. Zhao L, Bracken MB, Dewan AT, Chen S. Association between the SERPINE1 (PAI-1) 4G/5G insertion/deletion promoter polymorphism (rs1799889) and pre-eclampsia: a systematic review and meta-analysis. Mol Hum Reprod. 2013 Mar;19(3):136-43. doi: 10.1093/molehr/gas056. Epub 2012 Nov 23. PMID: 23180602.

503. Tamura RE, de Vasconcellos JF, Sarkar D, Libermann TA, Fisher PB, Zerbini LF. GADD45 proteins: central players in tumorigenesis. Curr Mol Med. 2012 Jun;12(5):634-51. doi: 10.2174/156652412800619978. PMID: 22515981; PMCID: PMC3797964.

504. Muttukrishna S, Knight PG, Groome NP, Redman CW, Ledger WL. Activin A and inhibin A as possible endocrine markers for pre-eclampsia. Lancet. 1997 May 3;349(9061):1285-8. doi: 10.1016/s0140-6736(96)09264-1. Erratum in: Lancet 1997 Aug 2;350(9074):372. PMID: 9142063.

505. Salomon LJ, Benattar C, Audibert F, Fernandez H, Duyme M, Taieb J, Frydman R. Severe preeclampsia is associated with high inhibin A levels and normal leptin levels at 7 to 13 weeks into pregnancy. Am J Obstet Gynecol. 2003 Dec;189(6):1517-22. doi: 10.1016/s0002-9378(03)00902-5. PMID: 14710054.

506. Yong HE, Melton PE, Johnson MP, Freed KA, Kalionis B, Murthi P, Brennecke SP, Keogh RJ, Moses EK. Genome-wide transcriptome directed pathway analysis of maternal pre-eclampsia susceptibility genes. PLoS One. 2015 May 26;10(5):e0128230. doi: 10.1371/journal.pone.0128230. PMID: 26010865; PMCID: PMC4444079.

507. Chelbi ST, Mondon F, Jammes H, Buffat C, Mignot TM et al. Expressional and epigenetic alterations of placental serine protease inhibitors: SERPINA3 is a potential marker of preeclampsia. Hypertension. 2007 Jan;49(1):76-83. doi: 10.1161/01.HYP.0000250831.52876.cb. Epub 2006 Nov 6. PMID: 17088445.

508. Albers RE, Kaufman MR, Natale BV, Keoni C, Kulkarni-Datar K, Min S, Williams CR, Natale DRC, Brown TL. Trophoblast-Specific Expression of Hif-1α Results in Preeclampsia-Like Symptoms and Fetal Growth Restriction. Sci Rep. 2019 Feb 26;9(1):2742. doi: 10.1038/s41598-019-39426-5. PMID: 30808910; PMCID: PMC6391498.

509. Sallais J, Park C, Alahari S, Porter T, Liu R, Kurt M, Farrell A, Post M, Caniggia I. HIF1 inhibitor acriflavine rescues early-onset preeclampsia phenotype in mice lacking placental prolyl hydroxylase domain protein 2. JCI Insight. 2022 Dec 8;7(23):e158908. doi: 10.1172/jci.insight.158908. PMID: 36227697; PMCID: PMC9746916.

510. Chen G, Chen J, Liu H, Chen S, Zhang Y, Li P, Thierry-Mieg D, Thierry-Mieg J, Mattes W, Ning B, Shi T. Comprehensive Identification and Characterization of Human Secretome Based on Integrative Proteomic and Transcriptomic Data. Front Cell Dev Biol. 2019 Nov 21;7:299. doi: 10.3389/fcell.2019.00299. PMID: 31824949; PMCID: PMC6881247. SPRomeDB freely accessible at: www.unimd.org/SPRomeDB

511. Basisty N, Kale A, Jeon OH, Kuehnemann C, Payne T, Rao C, Holtz A, Shah S, Sharma V, Ferrucci L, Campisi J, Schilling B. A proteomic atlas of senescence-associated secretomes for aging biomarker development. PLoS Biol. 2020 Jan 16;18(1):e3000599. doi: 10.1371/journal.pbio.3000599. PMID: 31945054; PMCID: PMC6964821.

512. Polesskaya A, Naguibneva I, Fritsch L, Duquet A, Ait-Si-Ali S, Robin P, Vervisch A, Pritchard LL, Cole P, Harel-Bellan A. CBP/p300 and muscle differentiation: no HAT, no muscle. EMBO J. 2001 Dec 3;20(23):6816-25. doi: 10.1093/emboj/20.23.6816. PMID: 11726517; PMCID: PMC125755.

513. Wang X, Pan L, Feng Y, Wang Y, Han Q, Han L, Han S, Guo J, Huang B, Lu J. P300 plays a role in p16(INK4a) expression and cell cycle arrest. Oncogene. 2008 Mar 20;27(13):1894-904. doi: 10.1038/sj.onc.1210821. Epub 2007 Oct 1. PMID: 17906698.

514. Wang D, Day EA, Townsend LK, Djordjevic D, Jørgensen SB, Steinberg GR. GDF15: emerging biology and therapeutic applications for obesity and cardiometabolic disease. Nat Rev Endocrinol. 2021 Oct;17(10):592-607. doi: 10.1038/s41574-021-00529-7. Epub 2021 Aug 11. PMID: 34381196.

515. Cruickshank T, MacDonald TM, Walker SP et al. Circulating Growth Differentiation Factor 15 Is Increased Preceding Preeclampsia Diagnosis: Implications as a Disease Biomarker. J Am Heart Assoc. 2021 Aug 17;10(16):e020302. doi: 10.1161/JAHA.120.020302. Epub 2021 Aug 13. PMID: 34387117; PMCID: PMC8475051.

516. Salimi S, Farajian-Mashhadi F, Naghavi A, Mokhtari M, Shahrakipour M, Saravani M, Yaghmaei M. Different profile of serum leptin between early onset and late onset preeclampsia. Dis Markers. 2014; 2014:628476. doi: 10.1155/2014/628476. Epub 2014 Jan 23. PMID: 24591763; PMCID: PMC3925616.

517. Daskalakis G, Bellos I, Nikolakea M, Pergialiotis V, Papapanagiotou A, Loutradis D. The role of serum adipokine levels in preeclampsia: A systematic review. Metabolism. 2020 May; 106:154172. doi: 10.1016/j.metabol.2020.154172. Epub 2020 Feb 4. PMID: 32027908.

518. Guo F, Zhang B, Yang H, Fu Y, Wang Y, Huang J, Cheng M, Li X, Shen Z, Li L, He P, Xiang AP, Wang S, Zhang H. Systemic transcriptome comparison between early- And late-onset pre-eclampsia shows distinct pathology and novel biomarkers. Cell Prolif. 2021 Feb;54(2):e12968. doi: 10.1111/cpr.12968. Epub 2020 Dec 17. PMID: 33332660; PMCID:

519. Ishijima M, Suzuki N, Hozumi K, Matsunobu T, Kosaki K, Kaneko H, Hassell JR, Arikawa-Hirasawa E, Yamada Y. Perlecan modulates VEGF signaling and is essential for vascularization in endochondral bone formation. Matrix Biol. 2012 May;31(4):234-45. doi: 10.1016/j.matbio.2012.02.006. Epub 2012 Mar 7. PMID: 22421594; PMCID: PMC3340529.

520. Robinson NJ, Baker PN, Jones CJ, Aplin JD. A role for tissue transglutaminase in stabilization of membrane-cytoskeletal particles shed from the human placenta. Biol Reprod. 2007 Oct;77(4):648-57. doi: 10.1095/biolreprod.107.061747. Epub 2007 Jul 11. PMID: 17625111.

521. Luo R, Liu C, Elliott SE, Wang W, Parchim N et al. Transglutaminase is a Critical Link Between Inflammation and Hypertension. J Am Heart Assoc. 2016 Jun 30;5(7):e003730. doi: 10.1161/JAHA.116.003730. PMID: 27364991; PMCID: PMC5015405.

522. Wertaschnigg D, Rolnik DL, Nie G, Teoh SSY, Syngelaki A, da Silva Costa F, Nicolaides KH. Second- and third-trimester serum levels of growth-differentiation factor-15 in prediction of pre-eclampsia. Ultrasound Obstet Gynecol. 2020 Dec;56(6):879-884. doi: 10.1002/uog.22070. Epub 2020 Nov 9. PMID: 32388891.

523. Liu J, Wang Y, Zhang S, Sun L, Shi Y. ADAM9 deubiquitination induced by USP22 suppresses proliferation, migration, invasion, and epithelial-mesenchymal transition of trophoblast cells in preeclampsia. Placenta. 2024 Feb; 146:50-57. doi: 10.1016/j.placenta.2023.12.008. Epub 2023 Dec 11. PMID: 38176298.

524. https://github.com/HiDiHlabs/planktonpy

525. Sun W, Cui B, Hong F, Xu Y. Establishment of ApoE-knockout mouse model of preeclampsia and relevant mechanisms. Exp Ther Med. 2016 Oct;12(4):2634-2638. doi: 10.3892/etm.2016.3678. Epub 2016 Sep 6. PMID: 27698766; PMCID: PMC5038201.

526. Chen H, Aneman I, Nikolic V, Karadzov Orlic N et al. Maternal plasma proteome profiling of biomarkers and pathogenic mechanisms of early-onset and late-onset preeclampsia. Sci Rep. 2022 Nov 9;12(1):19099. doi: 10.1038/s41598-022-20658-x. PMID: 36351970; PMCID: PMC9646706.

527. Yin J, Reiman EM, Beach TG, Serrano GE, Sabbagh MN, Nielsen M, Caselli RJ, Shi J. Effect of ApoE isoforms on mitochondria in Alzheimer disease. Neurology. 2020 Jun 9;94(23): e2404-e2411. doi: 10.1212/WNL.0000000000009582. Epub 2020 May 26. PMID: 32457210; PMCID: PMC7455369.

528. Witherel CE, Sao K, Brisson BK, Han B, Volk SW, Petrie RJ, Han L, Spiller KL. Regulation of extracellular matrix assembly and structure by hybrid M1/M2 macrophages. Biomaterials. 2021 Feb; 269:120667. doi: 10.1016/j.biomaterials.2021.120667. Epub 2021 Jan 7. PMID: 33450585; PMCID: PMC7870567.

529. Wu H, Liu K, Zhang J. Excess fibronectin 1 participates in pathogenesis of pre-eclampsia by promoting apoptosis and autophagy in vascular endothelial cells. Mol Hum Reprod. 2021 May 29;27(6):gaab030. doi: 10.1093/molehr/gaab030. PMID: 33881516.

530. Silverstein RL, Febbraio M. CD36, a scavenger receptor involved in immunity, metabolism, angiogenesis, and behavior. Sci Signal. 2009 May 26;2(72):re3. doi: 10.1126/scisignal.272re3. PMID: 19471024; PMCID: PMC2811062.

531. Cho S. CD36 as a therapeutic target for endothelial dysfunction in stroke. Curr Pharm Des. 2012;18(25):3721-30. doi: 10.2174/138161212802002760. PMID: 22574985; PMCID: PMC3411860.

532. Yang M, Cooley BC, Li W, Chen Y, Vasquez-Vivar J, Scoggins NO, Cameron SJ, Morrell CN, Silverstein RL. Platelet CD36 promotes thrombosis by activating redox sensor ERK5 in hyperlipidemic conditions. Blood. 2017 May 25;129(21):2917-2927. doi: 10.1182/blood-2016-11-750133. Epub 2017 Mar 23. PMID: 28336528; PMCID: PMC5445574.

533. Duan FM, Fu LJ, Wang YH, Adu-Gyamfi EA, Ruan LL, Xu ZW, Xiao SQ, Chen XM, Wang YX, Liu TH, Ding YB. THBS1 regulates trophoblast fusion through a CD36-dependent inhibition of cAMP, and its upregulation participates in preeclampsia. Genes Dis. 2020 Jun 8;8(3):353-363. doi: 10.1016/j.gendis.2020.05.007. PMID: 33997182; PMCID: PMC8093648.

534. Yao M, Roberts DD, Isenberg JS. Thrombospondin-1 inhibition of vascular smooth muscle cell responses occurs via modulation of both cAMP and cGMP. Pharmacol Res. 2011 Jan;63(1):13-22. doi: 10.1016/j.phrs.2010.10.014. Epub 2010 Oct 29. PMID: 20971192; PMCID: PMC3026097.

535. Ma Z, Mao C, Jia Y, Fu Y, Kong W. Extracellular matrix dynamics in vascular remodeling. Am J Physiol Cell Physiol. 2020 Sep 1;319(3):C481-C499. doi: 10.1152/ajpcell.00147.2020. Epub 2020 Jun 24. PMID: 32579472; PMCID: PMC7509265.

536. Konukiewitz B, Kasajima A, Schmitt M, Schwamborn K et al. Neuroendocrine Differentiation in Conventional Colorectal Adenocarcinomas: Incidental Finding or Prognostic Biomarker? Cancers (Basel). 2021 Oct 12;13(20):5111. doi: 10.3390/cancers13205111. PMID: 34680258; PMCID: PMC8533893.

537. Rekhtman N. Lung neuroendocrine neoplasms: recent progress and persistent challenges. Mod Pathol. 2022 Jan;35(Suppl 1):36-50. doi: 10.1038/s41379-021-00943-2. Epub 2021 Oct 18. PMID: 34663914; PMCID: PMC8695375.

538. Basturk O, Tang L, Hruban RH, Adsay V, Yang Z et al. Poorly differentiated neuroendocrine carcinomas of the pancreas: a clinicopathologic analysis of 44 cases. Am J Surg Pathol. 2014 Apr;38(4):437-47. doi: 10.1097/PAS.0000000000000169. PMID: 24503751; PMCID: PMC3977000.

539. Lin Y, Nakatochi M, Sasahira N, Ueno M, Egawa N, Adachi Y, Kikuchi S. Glycoprotein 2 in health and disease: lifting the veil. Genes Environ. 2021 Dec 3;43(1):53. doi: 10.1186/s41021-021-00229-8. PMID: 34861888; PMCID: PMC8641183.

540. Baldan J, Houbracken I, Rooman I, Bouwens L. Adult human pancreatic acinar cells dedifferentiate into an embryonic progenitor-like state in 3D suspension culture. Sci Rep. 2019 Mar 11;9(1):4040. doi: 10.1038/s41598-019-40481-1. PMID: 30858455; PMCID: PMC6411888.

541. Pan L, Mulaw MA, Gout J, Guo M, Zarrin H, Schwarz P, Baumann B, Seufferlein T, Wagner M, Oswald F. RBPJ Deficiency Sensitizes Pancreatic Acinar Cells to KRAS-Mediated Pancreatic Intraepithelial Neoplasia Initiation. Cell Mol Gastroenterol Hepatol. 2023;16(5):783-807. doi: 10.1016/j.jcmgh.2023.07.013. Epub 2023 Aug 4. PMID: 37543088; PMCID: PMC10520364.

542. von Figura G, Morris JP 4th, Wright CV, Hebrok M. Nr5a2 maintains acinar cell differentiation and constrains oncogenic Kras-mediated pancreatic neoplastic initiation. Gut. 2014 Apr;63(4):656-64. doi: 10.1136/gutjnl-2012-304287. Epub 2013 May 3. PMID: 23645620; PMCID: PMC3883808.

543. Backx E, Wauters E, Baldan J, Van Bulck M, Michiels E et al. MECOM permits pancreatic acinar cell dedifferentiation avoiding cell death under stress conditions. Cell Death Differ. 2021 Sep;28(9):2601-2615. doi: 10.1038/s41418-021-00771-6. Epub 2021 Mar 24. PMID: 33762742; PMCID: PMC8408219.

544. Hoang CQ, Hale MA, Azevedo-Pouly AC, Elsässer HP et al. Transcriptional Maintenance of Pancreatic Acinar Identity, Differentiation, and Homeostasis by PTF1A. Mol Cell Biol. 2016 Nov 28;36(24):3033-3047. doi: 10.1128/MCB.00358-16. PMID: 27697859; PMCID: PMC5126291.

545. Aigha II, Memon B, Elsayed AK, Abdelalim EM. Differentiation of human pluripotent stem cells into two distinct NKX6.1 populations of pancreatic progenitors. Stem Cell Res Ther. 2018 Apr 3;9(1):83. doi: 10.1186/s13287-018-0834-0. PMID: 29615106; PMCID: PMC5883581.

546. Offield MF, Jetton TL, Labosky PA, Ray M, Stein RW, Magnuson MA, Hogan BL, Wright CV. PDX-1 is required for pancreatic outgrowth and differentiation of the rostral duodenum. Development. 1996 Mar;122(3):983-95. doi: 10.1242/dev.122.3.983. PMID: 8631275.

547. Gonçalves CA, Larsen M, Jung S, Stratmann J et al. A 3D system to model human pancreas development and its reference single-cell transcriptome atlas identify signaling pathways required for progenitor expansion. Nat Commun. 2021 May 25;12(1):3144. doi: 10.1038/s41467-021-23295-6. PMID: 34035279; PMCID: PMC8149728.

548. Van Hoof D, D'Amour KA, German MS. Derivation of insulin-producing cells from human embryonic stem cells. Stem Cell Res. 2009 Sep-Nov;3(2-3):73-87. doi: 10.1016/j.scr.2009.08.003. Epub 2009 Aug 26. PMID: 19766074.

549. Lynn FC, Smith SB, Wilson ME, Yang KY, Nekrep N, German MS. Sox9 coordinates a transcriptional network in pancreatic progenitor cells. Proc Natl Acad Sci U S A. 2007 Jun 19;104(25):10500-5. doi: 10.1073/pnas.0704054104. Epub 2007 Jun 11. PMID: 17563382; PMCID: PMC1965542.

550. Timmer JR, Mak TW, Manova K, Anderson KV, Niswander L. Tissue morphogenesis and vascular stability require the Frem2 protein, product of the mouse myelencephalic blebs gene. Proc Natl Acad Sci U S A. 2005 Aug 16;102(33):11746-50. doi: 10.1073/pnas.0505404102. Epub 2005 Aug 8. PMID: 16087869; PMCID: PMC1183448.

551. Molina-Navarro MM, Roselló-Lletí E, Ortega A et al. Differential gene expression of cardiac ion channels in human dilated cardiomyopathy. PLoS One. 2013 Dec 5;8(12): e79792. doi: 10.1371/journal.pone.0079792. PMID: 24339868; PMCID: PMC3855055.

552. Catterall WA. Structure and regulation of voltage-gated Ca2+ channels. Annu Rev Cell Dev Biol. 2000; 16:521-55. doi: 10.1146/annurev.cellbio.16.1.521. PMID: 11031246.

553. Liu W, Palovcak A, Li F, Zafar A, Yuan F, Zhang Y. Fanconi anemia pathway as a prospective target for cancer intervention. Cell Biosci. 2020 Mar 16; 10:39. doi: 10.1186/s13578-020-00401-7. PMID: 32190289; PMCID: PMC7075017.

554. Chen P, Li J, Chen YC, Qian H, Chen YJ, Su JY, Wu M, Lan T. The functional status of DNA repair pathways determines the sensitization effect to cisplatin in non-small cell lung cancer cells. Cell Oncol (Dordr). 2016 Dec;39(6):511-522. doi: 10.1007/s13402-016-0291-7. Epub 2016 Jul 29. PMID: 27473273.

555. Taniguchi T, Tischkowitz M, Ameziane N, Hodgson SV, Mathew CG, Joenje H, Mok SC, D'Andrea AD. Disruption of the Fanconi anemia-BRCA pathway in cisplatin-sensitive ovarian tumors. Nat Med. 2003 May;9(5):568-74. doi: 10.1038/nm852. Epub 2003 Apr 7. PMID: 12692539.

556. Heberle AM, Prentzell MT, van Eunen K, Bakker BM, Grellscheid SN, Thedieck K. Molecular mechanisms of mTOR regulation by stress. Mol Cell Oncol. 2014 Dec 3;2(2):e970489. doi: 10.4161/23723548.2014.970489. PMID: 27308421; PMCID: PMC4904989.

557. Wu CW, Storey KB. mTOR Signaling in Metabolic Stress Adaptation. Biomolecules. 2021 May 1;11(5):681. doi: 10.3390/biom11050681. PMID: 34062764; PMCID: PMC8147357.

558. Tian T, Li X, Zhang J. mTOR Signaling in Cancer and mTOR Inhibitors in Solid Tumor Targeting Therapy. Int J Mol Sci. 2019 Feb 11;20(3):755. doi: 10.3390/ijms20030755. PMID: 30754640; PMCID: PMC6387042.

559. Zanini S, Renzi S, Giovinazzo F, Bermano G. mTOR Pathway in Gastroenteropancreatic Neuroendocrine Tumor (GEP-NETs). Front Endocrinol (Lausanne). 2020 Nov 16; 11:562505. doi: 10.3389/fendo.2020.562505. PMID: 33304317; PMCID: PMC7701056.

560. Marinoni I, Kurrer AS, Vassella E, Dettmer M, Rudolph T, Banz V, Hunger F, Pasquinelli S, Speel EJ, Perren A. Loss of DAXX and ATRX are associated with chromosome instability and reduced survival of patients with pancreatic neuroendocrine tumors. Gastroenterology. 2014 Feb;146(2):453-60.e5. doi: 10.1053/j.gastro.2013.10.020. Epub 2013 Oct 19. PMID: 24148618.

561. Scarpa A, Chang DK, Nones K, Corbo V et al. Whole-genome landscape of pancreatic neuroendocrine tumours. Nature. 2017 Mar 2;543(7643):65-71. doi: 10.1038/nature21063. Epub 2017 Feb 15. Erratum in: Nature. 2017 Sep 27; PMID: 28199314.

562. Shida T, Kishimoto T, Furuya M, Nikaido T, Koda K et al. Expression of an activated mammalian target of rapamycin (mTOR) in gastroenteropancreatic neuroendocrine tumors. Cancer Chemother Pharmacol. 2010 Apr;65(5):889-93. doi: 10.1007/s00280-009-1094-6. Epub 2009 Aug 6. PMID: 19657638.

563. Catena L, Bajetta E, Milione M, Ducceschi M, Valente M, Dominoni F, Colonna V. Mammalian target of rapamycin expression in poorly differentiated endocrine carcinoma: clinical and therapeutic future challenges. Target Oncol. 2011 Jun;6(2):65-8. doi: 10.1007/s11523-011-0171-z. Epub 2011 Apr 6. PMID: 21468754.

564. Gruber R, Panayiotou R, Nye E, Spencer-Dene B, Stamp G, Behrens A. YAP1 and TAZ Control Pancreatic Cancer Initiation in Mice by Direct Up-regulation of JAK-STAT3 Signaling. Gastroenterology. 2016 Sep;151(3):526-39. doi: 10.1053/j.gastro.2016.05.006. Epub 2016 May 20. PMID: 27215660; PMCID: PMC5007286.

565. Zhang W, Nandakumar N, Shi Y, Manzano M, Smith A, Graham G et al. Downstream of mutant KRAS, the transcription regulator YAP is essential for neoplastic progression to pancreatic ductal adenocarcinoma. Sci Signal. 2014 May 6;7(324):ra42. doi: 10.1126/scisignal.2005049. PMID: 24803537; PMCID: PMC4175524.

566. George NM, Day CE, Boerner BP, Johnson RL, Sarvetnick NE. Hippo signaling regulates pancreas development through inactivation of Yap. Mol Cell Biol. 2012 Dec;32(24):5116-28. doi: 10.1128/MCB.01034-12. Epub 2012 Oct 15. PMID: 23071096; PMCID: PMC3510525.

567. Tu B, Yao J, Ferri-Borgogno S, Zhao J et al. YAP1 oncogene is a context-specific driver for pancreatic ductal adenocarcinoma. JCI Insight. 2019 Nov 1;4(21): e130811. doi: 10.1172/jci.insight.130811. PMID: 31557131; PMCID: PMC6948828.

568. Peran I, Dakshanamurthy S, McCoy MD, Mavropoulos A et al. Cadherin 11 Promotes Immunosuppression and Extracellular Matrix Deposition to Support Growth of Pancreatic Tumors and Resistance to Gemcitabine in Mice. Gastroenterology. 2021 Mar;160(4):1359-1372.e13. doi: 10.1053/j.gastro.2020.11.044. Epub 2020 Dec 9. PMID: 33307028; PMCID: PMC7956114.

569. Birtolo C, Pham H, Morvaridi S, Chheda C et al. Cadherin-11 Is a Cell Surface Marker Up-Regulated in Activated Pancreatic Stellate Cells and Is Involved in Pancreatic Cancer Cell Migration. Am J Pathol. 2017 Jan;187(1):146-155. doi: 10.1016/j.ajpath.2016.09.012. Epub 2016 Nov 14. PMID: 27855278; PMCID: PMC5225304.

570. Linares J, Marín-Jiménez JA, Badia-Ramentol J, Calon A. Determinants and Functions of CAFs Secretome During Cancer Progression and Therapy. Front Cell Dev Biol. 2021 Jan 22;8:621070. doi: 10.3389/fcell.2020.621070. PMID: 33553157; PMCID: PMC7862334.

571. Raica M, Cimpean AM. Platelet-Derived Growth Factor (PDGF)/PDGF Receptors (PDGFR) Axis as Target for Antitumor and Antiangiogenic Therapy. Pharmaceuticals (Basel). 2010 Mar 11;3(3):572-599. doi: 10.3390/ph3030572. PMID: 27713269; PMCID: PMC4033970.

572. Wu K, Tang H, Lin R, Carr SG, Wang Z et al. Endothelial platelet-derived growth factor-mediated activation of smooth muscle platelet-derived growth factor receptors in pulmonary arterial hypertension. Pulm Circ. 2020 Sep 10;10(3):2045894020948470. doi: 10.1177/2045894020948470. PMID: 33294172; PMCID: PMC7707860.

573. Guo Y, Guo H, Zhang Y, Cui J. Anaplastic lymphoma kinase-special immunity and immunotherapy. Front Immunol. 2022 Jul 25; 13:908894. doi: 10.3389/fimmu.2022.908894. PMID: 35958559; PMCID: PMC9359062.

574. Guan Y, Du Y, Wang G, Gou H, Xue Y, Xu J, Li E, Chan DW, Wu D, Xu P, Ni P, Xu D, Hu Y. Overexpression of PLXDC2 in Stromal Cell-Associated M2 Macrophages Is Related to EMT and the Progression of Gastric Cancer. Front Cell Dev Biol. 2021 May 28;9:673295. doi: 10.3389/fcell.2021.673295. PMID: 34124056; PMCID: PMC8194078.

575. Tubau-Juni N, Bassaganya-Riera J, Leber A, Zoccoli-Rodriguez V, Kronsteiner B, Viladomiu M, Abedi V, Philipson CW, Hontecillas R. Identification of new regulatory genes through expression pattern analysis of a global RNA-seq dataset from a Helicobacter pylori co-culture system. Sci Rep. 2020 Jul 13;10(1):11506. doi: 10.1038/s41598-020-68439-8. PMID: 32661418; PMCID: PMC7359330.

576. Vasiljeva O, Papazoglou A, Krüger A, Brodoefel H et al. Tumor cell-derived and macrophage-derived cathepsin B promotes progression and lung metastasis of mammary cancer. Cancer Res. 2006 May 15;66(10):5242-50. doi: 10.1158/0008-5472.CAN-05-4463. PMID: 16707449.

577. Lin C, Yang H, Zhao W, Wang W. CTSB+ macrophage repress memory immune hub in the liver metastasis site of colorectal cancer patient revealed by multi-omics analysis. Biochem Biophys Res Commun. 2022 Oct 20; 626:8-14. doi: 10.1016/j.bbrc.2022.06.037. Epub 2022 Jul 3. PMID: 35964553.

578. Shiraishi A, Uruno T, Sanematsu F, Ushijima M, Sakata D, Hara T, Fukui Y. DOCK8 Protein Regulates Macrophage Migration through Cdc42 Protein Activation and LRAP35a Protein Interaction. J Biol Chem. 2017 Feb 10;292(6):2191-2202. doi: 10.1074/jbc.M116.736306. Epub 2016 Dec 27. PMID: 28028174; PMCID: PMC5313093.

579. Ji L, Xu S, Luo H, Zeng F. Insights from DOCK2 in cell function and pathophysiology. Front Mol Biosci. 2022 Sep 29; 9:997659. doi: 10.3389/fmolb.2022.997659. PMID: 36250020; PMCID: PMC9559381

580. Fogel BL, Wexler E, Wahnich A, Friedrich T, Vijayendran C, Gao F, Parikshak N, Konopka G, Geschwind DH. RBFOX1 regulates both splicing and transcriptional networks in human neuronal development. Hum Mol Genet. 2012 Oct 1;21(19):4171-86. doi: 10.1093/hmg/dds240. Epub 2012 Jun 23. PMID: 22730494; PMCID: PMC3441119.

581. Seymour PA, Serup P. Differential use of the Nkx2.2 NK2 domain in developing pancreatic islets and neurons. Genes Dev. 2023 Jun 1;37(11-12):451-453. doi: 10.1101/gad.350895.123. Epub 2023 Jun 30. PMID: 37399332; PMCID: PMC10393196.

582. Moss ND, Sussel L. mRNA Processing: An Emerging Frontier in the Regulation of Pancreatic β Cell Function. Front Genet. 2020 Sep 1; 11:983. doi: 10.3389/fgene.2020.00983. PMID: 33088281; PMCID: PMC7490333.

583. Forastieri C, Italia M, Toffolo E, Romito E, Bonasoni MP, Ranzani V, Bodega B, Rusconi F, Battaglioli E. Evolution Increases Primates Brain Complexity Extending RbFOX1 Splicing Activity to LSD1 Modulation. J Neurosci. 2022 May 4;42(18):3689-3703. doi: 10.1523/JNEUROSCI.1782-21.2022. Epub 2022 Mar 29. PMID: 35351830; PMCID: PMC9087731.

584. Laurent B, Ruitu L, Murn J, Hempel K, Ferrao R, Xiang Y, Liu S, Garcia BA, Wu H, Wu F, Steen H, Shi Y. A specific LSD1/KDM1A isoform regulates neuronal differentiation through H3K9 demethylation. Mol Cell. 2015 Mar 19;57(6):957-970. doi: 10.1016/j.molcel.2015.01.010. Epub 2015 Feb 12. PMID: 25684206; PMCID: PMC4369399.

585. Perron G, Jandaghi P, Moslemi E, Nishimura T, Rajaee M, Alkallas R, Lu T, Riazalhosseini Y, Najafabadi HS. Pan-cancer analysis of mRNA stability for decoding tumour post-transcriptional programs. Commun Biol. 2022 Aug 20;5(1):851. doi: 10.1038/s42003-022-03796-w. PMID: 35987939; PMCID: PMC9392771.

586. Stergiopoulos A, Elkouris M, Politis PK. Prospero-related homeobox 1 (Prox1) at the crossroads of diverse pathways during adult neural fate specification. Front Cell Neurosci. 2015 Jan 26; 8:454. doi: 10.3389/fncel.2014.00454. PMID: 25674048; PMCID: PMC4306308.

587. Wang J, Kilic G, Aydin M, Burke Z, Oliver G, Sosa-Pineda B. Prox1 activity controls pancreas morphogenesis and participates in the production of "secondary transition" pancreatic endocrine cells. Dev Biol. 2005 Oct 1;286(1):182-94. doi: 10.1016/j.ydbio.2005.07.021. PMID: 16122728.

588. Martinez-Ramirez AS, Borders TL, Paul L, Schipma M, Wang X, Korobova F, Wright CV, Sosa-Pineda B. Specific Temporal Requirement of Prox1 Activity During Pancreatic Acinar Cell Development. Gastro Hep Adv. 2022;1(5):807-823. doi: 10.1016/j.gastha.2022.05.013. Epub 2022 May 26. PMID: 37829188; PMCID: PMC10569262.

589. Westmoreland JJ, Kilic G, Sartain C, Sirma S, Blain J, Rehg J, Harvey N, Sosa-Pineda B. Pancreas-specific deletion of Prox1 affects development and disrupts homeostasis of the exocrine pancreas. Gastroenterology. 2012 Apr;142(4):999-1009.e6. doi: 10.1053/j.gastro.2011.12.007. Epub 2011 Dec 13. PMID: 22178591; PMCID: PMC3398795.

590. Villani V, Thornton ME, Zook HN, Crook CJ, Grubbs BH, Orlando G, De Filippo R, Ku HT, Perin L. SOX9+/PTF1A+ Cells Define the Tip Progenitor Cells of the Human Fetal Pancreas of the Second Trimester. Stem Cells Transl Med. 2019 Dec;8(12):1249-1264. doi: 10.1002/sctm.19-0231. Epub 2019 Oct 21. PMID: 31631582; PMCID: PMC6877773.

591. Zhou Q, Law AC, Rajagopal J, Anderson WJ, Gray PA, Melton DA. A multipotent progenitor domain guides pancreatic organogenesis. Dev Cell. 2007 Jul;13(1):103-14. doi: 10.1016/j.devcel.2007.06.001. PMID: 17609113.

592. Gu G, Dubauskaite J, Melton DA. Direct evidence for the pancreatic lineage: NGN3+ cells are islet progenitors and are distinct from duct progenitors. Development. 2002 May;129(10):2447-57. doi: 10.1242/dev.129.10.2447. PMID: 11973276.

593. Trott J, Tan EK, Ong S, Titmarsh DM, Denil SLIJ, Giam M, Wong CK, Wang J, Shboul M, Eio M, Cooper-White J, Cool SM, Rancati G, Stanton LW, Reversade B, Dunn NR. Long-Term Culture of Self-renewing Pancreatic Progenitors Derived from Human Pluripotent Stem Cells. Stem Cell Reports. 2017 Jun 6;8(6):1675-1688. doi: 10.1016/j.stemcr.2017.05.019. PMID: 28591650; PMCID: PMC5470345.

594. Seymour PA, Freude KK, Tran MN, Mayes EE, Jensen J, Kist R, Scherer G, Sander M. SOX9 is required for maintenance of the pancreatic progenitor cell pool. Proc Natl Acad Sci U S A. 2007 Feb 6;104(6):1865-70. doi: 10.1073/pnas.0609217104. Epub 2007 Jan 31. PMID: 17267606; PMCID: PMC1794281.

595. Cogger KF, Sinha A, Sarangi F, McGaugh EC, Saunders D, Dorrell C, Mejia-Guerrero S, Aghazadeh Y, Rourke JL, Screaton RA, Grompe M, Streeter PR, Powers AC, Brissova M, Kislinger T, Nostro MC. Glycoprotein 2 is a specific cell surface marker of human pancreatic progenitors. Nat Commun. 2017 Aug 24;8(1):331. doi: 10.1038/s41467-017-00561-0. PMID: 28835709; PMCID: PMC5569081.

596. Kropp PA, Dunn JC, Carboneau BA, Stoffers DA, Gannon M. Cooperative function of Pdx1 and Oc1 in multipotent pancreatic progenitors impacts postnatal islet maturation and adaptability. Am J Physiol Endocrinol Metab. 2018 Apr 1;314(4):E308-E321. doi: 10.1152/ajpendo.00260.2017. Epub 2017 Dec 12. PMID: 29351489; PMCID: PMC5966755.

597. Pinho AV, Rooman I, Reichert M, De Medts N, Bouwens L, Rustgi AK, Real FX. Adult pancreatic acinar cells dedifferentiate to an embryonic progenitor phenotype with concomitant activation of a senescence programme that is present in chronic pancreatitis. Gut. 2011 Jul;60(7):958-66. doi: 10.1136/gut.2010.225920. Epub 2010 Dec 30. PMID: 21193456.

598. Haumaitre C, Barbacci E, Jenny M, Ott MO, Gradwohl G, Cereghini S. Lack of TCF2/vHNF1 in mice leads to pancreas agenesis. Proc Natl Acad Sci U S A. 2005 Feb 1;102(5):1490-5. doi: 10.1073/pnas.0405776102. Epub 2005 Jan 24. PMID: 15668393; PMCID: PMC547822.

599. De Vas MG, Kopp JL, Heliot C, Sander M, Cereghini S, Haumaitre C. Hnf1b controls pancreas morphogenesis and the generation of Ngn3+ endocrine progenitors. Development. 2015 Mar 1;142(5):871-82. doi: 10.1242/dev.110759. PMID: 25715395; PMCID: PMC4352981.

600. Allen HL, Flanagan SE, Shaw-Smith C, De Franco E, Akerman I, Caswell R; International Pancreatic Agenesis Consortium; Ferrer J, Hattersley AT, Ellard S. GATA6 haploinsufficiency causes pancreatic agenesis in humans. Nat Genet. 2011 Dec 11;44(1):20-22. doi: 10.1038/ng.1035. PMID: 22158542; PMCID: PMC4062962.

601. Tiyaboonchai A, Cardenas-Diaz FL, Ying L, Maguire JA et al. GATA6 Plays an Important Role in the Induction of Human Definitive Endoderm, Development of the Pancreas, and Functionality of Pancreatic β Cells. Stem Cell Reports. 2017 Mar 14;8(3):589-604. doi: 10.1016/j.stemcr.2016.12.026. Epub 2017 Feb 9. PMID: 28196690; PMCID: PMC5355564.

602. Carrasco M, Delgado I, Soria B, Martín F, Rojas A. GATA4 and GATA6 control mouse pancreas organogenesis. J Clin Invest. 2012 Oct;122(10):3504-15. doi: 10.1172/JCI63240. Epub 2012 Sep 24. PMID: 23006330; PMCID: PMC3461915.

603. Swift GH, Liu Y, Rose SD, Bischof LJ, Steelman S, Buchberg AM, Wright CV, MacDonald RJ. An endocrine-exocrine switch in the activity of the pancreatic homeodomain protein PDX1 through formation of a trimeric complex with PBX1b and MRG1 (MEIS2). Mol Cell Biol. 1998 Sep;18(9):5109-20. doi: 10.1128/MCB.18.9.5109. PMID: 9710595; PMCID: PMC109096.

604. Schlesinger Y, Yosefov-Levi O, Kolodkin-Gal D, Granit RZ, Peters L, Kalifa R, Xia L, Nasereddin A, Shiff I, Amran O, Nevo Y, Elgavish S, Atlan K, Zamir G, Parnas O. Single-cell transcriptomes of pancreatic preinvasive lesions and cancer reveal acinar metaplastic cells' heterogeneity. Nat Commun. 2020 Sep 9;11(1):4516. doi: 10.1038/s41467-020-18207-z. PMID: 32908137; PMCID: PMC7481797.

605. Brunton H, Caligiuri G, Cunningham R, Upstill-Goddard R, Bailey UM et al. HNF4A and GATA6 Loss Reveals Therapeutically Actionable Subtypes in Pancreatic Cancer. Cell Rep. 2020 May 12;31(6):107625. doi: 10.1016/j.celrep.2020.107625. PMID: 32402285; PMCID: PMC9511995.

606. Prévot PP, Simion A, Grimont A, Colletti M, Khalaileh A, Van den Steen G, Sempoux C, Xu X, Roelants V, Hald J, Bertrand L, Heimberg H, Konieczny SF, Dor Y, Lemaigre FP, Jacquemin P. Role of the ductal transcription factors HNF6 and Sox9 in pancreatic acinar-to-ductal metaplasia. Gut. 2012 Dec;61(12):1723-32. doi: 10.1136/gutjnl-2011-300266. Epub 2012 Jan 22. PMID: 22271799; PMCID: PMC3898034.

607. Kawasaki K, Toshimitsu K, Matano M, Fujita M et al. An Organoid Biobank of Neuroendocrine Neoplasms Enables Genotype-Phenotype Mapping. Cell. 2020 Nov 25;183(5):1420-1435.e21. doi: 10.1016/j.cell.2020.10.023. Epub 2020 Nov 6. PMID: 33159857.

608. Isobe T, Seki M, Yoshida K, Sekiguchi M, Shiozawa Y et al. Integrated Molecular Characterization of the Lethal Pediatric Cancer Pancreatoblastoma. Cancer Res. 2018 Feb 15;78(4):865-876. doi: 10.1158/0008-5472.CAN-17-2581. Epub 2017 Dec 12. PMID: 29233928.

609. Nakazawa N, Sohda M, Ide M, Shimoda Y, Tateno K, Watanabe T, Sano A, Sakai M, Yokobori T, Ogawa H, Oyama T, Shirabe K, Saeki H. PROX1 Was Associated with LGR5 and Wnt Signaling and Contributed to Poor Prognosis in Gastric Cancer. Oncology. 2022;100(11):569-575. doi: 10.1159/000526734. Epub 2022 Sep 14. PMID: 36103842.

610. Zhu L, Tian Q, Gao H, Wu K, Wang B, Ge G, Jiang S, Wang K, Zhou C, He J, Liu P, Ren Y, Wang B. PROX1 promotes breast cancer invasion and metastasis through WNT/β-catenin

pathway via interacting with hnRNPK. Int J Biol Sci. 2022 Feb 28;18(5):2032-2046. doi: 10.7150/ijbs.68960. PMID: 35342346; PMCID: PMC8935233.

611. Hale MA, Swift GH, Hoang CQ, Deering TG, Masui T, Lee YK, Xue J, MacDonald RJ. The nuclear hormone receptor family member NR5A2 controls aspects of multipotent progenitor cell formation and acinar differentiation during pancreatic organogenesis. Development. 2014 Aug;141(16):3123-33. doi: 10.1242/dev.109405. Epub 2014 Jul 25. PMID: 25063451; PMCID: PMC4197540.

612. Yang X, Raum JC, Kim J, Yu R, Yang J, Rice G, Li C, Won KJ, Stanescu DE, Stoffers DA. A PDX1 cistrome and single-cell transcriptome resource of the developing pancreas. Development. 2022 Jul 1;149(13): dev200432. doi: 10.1242/dev.200432. Epub 2022 Jun 29. PMID: 35708349; PMCID: PMC9340549.

613. Sinenko SA, Kuzmin AA, Skvortsova EV, Ponomartsev SV, Efimova EV, Bader M, Alenina N, Tomilin AN. Tryptophan Hydroxylase-2-Mediated Serotonin Biosynthesis Suppresses Cell Reprogramming into Pluripotent State. Int J Mol Sci. 2023 Mar 2;24(5):4862. doi: 10.3390/ijms24054862. PMID: 36902295; PMCID: PMC10003565.

614. Tow DH, Tran CG, Borbon LC, Ridder M, Li G, Kaemmer CA et al. Inhibition of serotonin biosynthesis in neuroendocrine neoplasm suppresses tumor growth in vivo. bioRxiv [Preprint]. 2023 Apr 8:2023.04.07.536013. doi: 10.1101/2023.04.07.536013. PMID: 37066322; PMCID: PMC10104152.

615. Ma Z, Lytle NK, Chen B, Jyotsana N, Novak SW et al. Single-Cell Transcriptomics Reveals a Conserved Metaplasia Program in Pancreatic Injury. Gastroenterology. 2022 Feb;162(2):604-620.e20. doi: 10.1053/j.gastro.2021.10.027. Epub 2021 Oct 23. PMID: 34695382; PMCID: PMC8792222.

616. Qadir MMF, Álvarez-Cubela S, Klein D, van Dijk J, Muñiz-Anquela R et al. Single-cell resolution analysis of the human pancreatic ductal progenitor cell niche. Proc Natl Acad Sci U S A. 2020 May 19;117(20):10876-10887. doi: 10.1073/pnas.1918314117. Epub 2020 Apr 30. Erratum in: Proc Natl Acad Sci U S A. 2021 Feb 9;118(6): PMID: 32354994; PMCID: PMC7245071.

617. Rovira M, Scott SG, Liss AS, Jensen J, Thayer SP, Leach SD. Isolation and characterization of centroacinar/terminal ductal progenitor cells in adult mouse pancreas. Proc Natl Acad Sci U S A. 2010 Jan 5;107(1):75-80. doi: 10.1073/pnas.0912589107. Epub 2009 Dec 15. PMID: 20018761; PMCID: PMC2806716.

618. Stanger BZ, Stiles B, Lauwers GY, Bardeesy N, Mendoza M, Wang Y, Greenwood A, Cheng KH, McLaughlin M, Brown D, Depinho RA, Wu H, Melton DA, Dor Y. Pten constrains centroacinar cell expansion and malignant transformation in the pancreas. Cancer Cell. 2005 Sep;8(3):185-95. doi: 10.1016/j.ccr.2005.07.015. PMID: 16169464.

619. Ramond C, Beydag-Tasöz BS, Azad A, van de Bunt M, Petersen MBK, Beer NL, Glaser N, Berthault C, Gloyn AL, Hansson M, McCarthy MI, Honoré C, Grapin-Botton A, Scharfmann R. Understanding human fetal pancreas development using subpopulation sorting, RNA sequencing and single-cell profiling. Development. 2018 Aug 15;145(16): dev165480. doi: 10.1242/dev.165480. PMID: 30042179; PMCID: PMC6124547.

620. de la O, Sean, Zhe Liu, Han Sun, Shengyang K. Yu, Daniel M. Wong, Emily Chu, Sneha A. Rao, Nicolas Eng, Gabriel Peixoto, Jacquelyn Bouza, Yin Shen, Sarah M. Knox, Aaron D.

Tward, Anna L. Gloyn, and Julie B. Sneddon. "Single-Cell Multi-Omic Roadmap of Human Fetal Pancreatic Development." bioRxiv (2022): 2022.02.17.480942.

621. Itoh M, Takizawa Y, Hanai S, Okazaki S, Miyata R, Inoue T, Akashi T, Hayashi M, Goto Y. Partial loss of pancreas endocrine and exocrine cells of human ARX-null mutation: consideration of pancreas differentiation. Differentiation. 2010 Sep-Oct;80(2-3):118-22. doi: 10.1016/j.diff.2010.05.003. Epub 2010 Jun 9. PMID: 20538404.

622. Takagi H, Zhao S, Muto S, Yokouchi H et al. Delta-like 1 homolog (DLK1) as a possible therapeutic target and its application to radioimmunotherapy using 125I-labelled anti-DLK1 antibody in lung cancer models (HOT1801 and FIGHT004). Lung Cancer. 2021 Mar; 153:134-142. doi: 10.1016/j.lungcan.2021.01.014. Epub 2021 Jan 14. PMID: 33508526.

623. Altenberger T, Bilban M, Auer M, Knosp E, Wolfsberger S, Gartner W, Mineva I, Zielinski C, Wagner L, Luger A. Identification of DLK1 variants in pituitary- and neuroendocrine tumors. Biochem Biophys Res Commun. 2006 Feb 17;340(3):995-1005. doi: 10.1016/j.bbrc.2005.12.094. Epub 2005 Dec 27. PMID: 16403460.

624. Oronsky B, Ma PC, Morgensztern D, Carter CA. Nothing But NET: A Review of Neuroendocrine Tumors and Carcinomas. Neoplasia. 2017 Dec;19(12):991-1002. doi: 10.1016/j.neo.2017.09.002. Epub 2017 Nov 5. PMID: 29091800; PMCID: PMC5678742.

625. Hofland J, Kaltsas G, de Herder WW. Advances in the Diagnosis and Management of Well-Differentiated Neuroendocrine Neoplasms. Endocr Rev. 2020 Apr 1;41(2):371–403. doi: 10.1210/endrev/bnz004. PMID: 31555796; PMCID: PMC7080342.

626. Fine SW. Neuroendocrine tumors of the prostate. Mod Pathol. 2018 Jan;31(S1):S122-132. doi: 10.1038/modpathol.2017.164. PMID: 29297494.

627. Merkens L, Sailer V, Lessel D, Janzen E, Greimeier S, Kirfel J, Perner S, Pantel K, Werner S, von Amsberg G. Aggressive variants of prostate cancer: underlying mechanisms of neuroendocrine transdifferentiation. J Exp Clin Cancer Res. 2022 Feb 2;41(1):46. doi: 10.1186/s13046-022-02255-y. PMID: 35109899; PMCID: PMC8808994.

628. Ooki A, Osumi H, Fukuda K, Yamaguchi K. Potent molecular-targeted therapies for gastro-entero-pancreatic neuroendocrine carcinoma. Cancer Metastasis Rev. 2023 Sep;42(3):1021-1054. doi: 10.1007/s10555-023-10121-2. Epub 2023 Jul 8. PMID: 37422534; PMCID: PMC10584733.

629. Zhang J, Yang C, Wu C, Cui W, Wang L. DNA Methyltransferases in Cancer: Biology, Paradox, Aberrations, and Targeted Therapy. Cancers (Basel). 2020 Jul 31;12(8):2123. doi: 10.3390/cancers12082123. PMID: 32751889; PMCID: PMC7465608.

630. Izumi D, Toden S, Ureta E, Ishimoto T, Baba H, Goel A. TIAM1 promotes chemoresistance and tumor invasiveness in colorectal cancer. Cell Death Dis. 2019 Mar 19;10(4):267. doi: 10.1038/s41419-019-1493-5. PMID: 30890693; PMCID: PMC6425043.

631. Naert T, Dimitrakopoulou D, Tulkens D, Demuynck S, Carron M, Noelanders R, Eeckhout L, Van Isterdael G, Deforce D, Vanhove C, Van Dorpe J, Creytens D, Vleminckx K. RBL1 (p107) functions as tumor suppressor in glioblastoma and small-cell pancreatic neuroendocrine carcinoma in Xenopus tropicalis. Oncogene. 2020 Mar;39(13):2692-2706. doi: 10.1038/s41388-020-1173-z. Epub 2020 Jan 30. PMID: 32001819.

632. Lázaro S, Pérez-Crespo M, Lorz C, Bernardini A, Oteo M, Enguita AB, Romero E, Hernández P, Tomás L, Morcillo MÁ, Paramio JM, Santos M. Differential development of large-cell neuroendocrine or small-cell lung carcinoma upon inactivation of 4 tumor suppressor genes. Proc Natl Acad Sci U S A. 2019 Oct 29;116(44):22300-22306. doi: 10.1073/pnas.1821745116. Epub 2019 Oct 14. PMID: 31611390; PMCID: PMC6825275.

633. Ireland AS, Micinski AM, Kastner DW, Guo B, Wait SJ et al. MYC Drives Temporal Evolution of Small Cell Lung Cancer Subtypes by Reprogramming Neuroendocrine Fate. Cancer Cell. 2020 Jul 13;38(1):60-78.e12. doi: 10.1016/j.ccell.2020.05.001. Epub 2020 May 30. PMID: 32473656; PMCID: PMC7393942.

634. Moffitt RA, Marayati R, Flate EL, Volmar KE, Loeza SG et al. Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. Nat Genet. 2015 Oct;47(10):1168-78. doi: 10.1038/ng.3398. Epub 2015 Sep 7. PMID: 26343385; PMCID: PMC4912058.

635. Kloesch B, Ionasz V, Paliwal S, Hruschka N et al. A GATA6-centred gene regulatory network involving HNFs and ΔNp63 controls plasticity and immune escape in pancreatic cancer. Gut. 2022 Apr;71(4):766-777. doi: 10.1136/gutjnl-2020-321397. Epub 2021 Apr 12. PMID: 33846140; PMCID: PMC9733634.

636. Bevere M, Gkountakos A, Martelli FM, Scarpa A, Luchini C, Simbolo M. An Insight on Functioning Pancreatic Neuroendocrine Neoplasms. Biomedicines. 2023 Jan 21;11(2):303. doi: 10.3390/biomedicines11020303. PMID: 36830839; PMCID: PMC9953748.

637. Toor D, Loree JM, Gao ZH, Wang G, Zhou C. Mixed neuroendocrine-non-neuroendocrine neoplasms of the digestive system: A mini-review. World J Gastroenterol. 2022 May 21;28(19):2076-2087. doi: 10.3748/wjg.v28.i19.2076. PMID: 35664032; PMCID: PMC9134131.

638. La Rosa S, Marando A, Sessa F, Capella C. Mixed Adenoneuroendocrine Carcinomas (MANECs) of the Gastrointestinal Tract: An Update. Cancers (Basel). 2012 Jan 16;4(1):11-30. doi: 10.3390/cancers4010011. PMID: 24213223; PMCID: PMC3712682.

639. Imaoka K, Fukuda S, Tazawa H, Kuga Y, Mochizuki T, Hirata Y, Fujisaki S, Takahashi M, Nishida T, Sakimoto H. A mixed adenoneuroendocrine carcinoma of the pancreas: a case report. Surg Case Rep. 2016 Dec;2(1):133. doi: 10.1186/s40792-016-0263-1. Epub 2016 Nov 15. PMID: 27848241; PMCID: PMC5110455.

640. Zhao B, Li Z, Yu S, Li T, Wang W, Liu R, Zhang B, Fang X, Shen Y, Han Q, Xu X, Wang K, Gong W, Li T, Li A, Zhou T, Li W, Li T. LEF1 enhances β-catenin transactivation through IDR-dependent liquid-liquid phase separation. Life Sci Alliance. 2023 Sep 1;6(11): e202302118. doi: 10.26508/lsa.202302118. PMID: 37657935; PMCID: PMC10474303.

641. Singhi AD, Lilo M, Hruban RH, Cressman KL, Fuhrer K, Seethala RR. Overexpression of lymphoid enhancer-binding factor 1 (LEF1) in solid-pseudopapillary neoplasms of the pancreas. Mod Pathol. 2014 Oct;27(10):1355-63. doi: 10.1038/modpathol.2014.40. Epub 2014 Mar 21. PMID: 24658583; PMCID: PMC4318228.

642. Ng NHJ, Jasmen JB, Lim CS, Lau HH, Krishnan VG, Kadiwala J, Kulkarni RN, Ræder H, Vallier L, Hoon S, Teo AKK. HNF4A Haploinsufficiency in MODY1 Abrogates Liver and Pancreas Differentiation from Patient-Derived Induced Pluripotent Stem Cells. iScience. 2019

Jun 28;16:192-205. doi: 10.1016/j.isci.2019.05.032. Epub 2019 May 27. PMID: 31195238; PMCID: PMC6562146.

643. Bohuslavova R, Fabriciova V, Smolik O, Lebrón-Mora L, Abaffy P, Benesova S, Zucha D, Valihrach L, Berkova Z, Saudek F, Pavlinkova G. NEUROD1 reinforces endocrine cell fate acquisition in pancreatic development. Nat Commun. 2023 Sep 9;14(1):5554. doi: 10.1038/s41467-023-41306-6. PMID: 37689751; PMCID: PMC10492842.

644. Bohuslavova R, Fabriciova V, Lebrón-Mora L, Malfatti J, Smolik O, Valihrach L, Benesova S, Zucha D, Berkova Z, Saudek F, Evans SM, Pavlinkova G. ISL1 controls pancreatic alpha cell fate and beta cell maturation. Cell Biosci. 2023 Mar 10;13(1):53. doi: 10.1186/s13578-023-01003-9. PMID: 36899442; PMCID: PMC9999528.

645. Du A, Hunter CS, Murray J, Noble D, Cai CL, Evans SM, Stein R, May CL. Islet-1 is required for the maturation, proliferation, and survival of the endocrine pancreas. Diabetes. 2009 Sep;58(9):2059-69. doi: 10.2337/db08-0987. Epub 2009 Jun 5. PMID: 19502415; PMCID: PMC2731519.

646. St-Onge L, Sosa-Pineda B, Chowdhury K, Mansouri A, Gruss P. Pax6 is required for differentiation of glucagon-producing alpha-cells in mouse pancreas. Nature. 1997 May 22;387(6631):406-9. doi: 10.1038/387406a0. PMID: 9163426.

647. Gosmain Y, Marthinet E, Cheyssac C, Guérardel A, Mamin A, Katz LS, Bouzakri K, Philippe J. Pax6 controls the expression of critical genes involved in pancreatic {alpha} cell differentiation and function. J Biol Chem. 2010 Oct 22;285(43):33381-33393. doi: 10.1074/jbc.M110.147215. Epub 2010 Jun 30. PMID: 20592023; PMCID: PMC2963382.

648. Ediger BN, Du A, Liu J, Hunter CS, Walp ER, Schug J, Kaestner KH, Stein R, Stoffers DA, May CL. Islet-1 Is essential for pancreatic β-cell function. Diabetes. 2014 Dec;63(12):4206-17. doi: 10.2337/db14-0096. Epub 2014 Jul 15. PMID: 25028525; PMCID: PMC4237994.

649. Thakurela S, Tiwari N, Schick S, Garding A, Ivanek R, Berninger B, Tiwari VK. Mapping gene regulatory circuitry of Pax6 during neurogenesis. Cell Discov. 2016 Feb 9; 2:15045. doi: 10.1038/celldisc.2015.45. PMID: 27462442; PMCID: PMC4860964.

650. Sansom SN, Griffiths DS, Faedo A, Kleinjan DJ, Ruan Y, Smith J, van Heyningen V, Rubenstein JL, Livesey FJ. The level of the transcription factor Pax6 is essential for controlling the balance between neural stem cell self-renewal and neurogenesis. PLoS Genet. 2009 Jun;5(6): e1000511. doi: 10.1371/journal.pgen.1000511. Epub 2009 Jun 12. PMID: 19521500; PMCID: PMC2686252.

651. Ericson J, Rashbass P, Schedl A, Brenner-Morton S, Kawakami A, van Heyningen V, Jessell TM, Briscoe J. Pax6 controls progenitor cell identity and neuronal fate in response to graded Shh signaling. Cell. 1997 Jul 11;90(1):169-80. doi: 10.1016/s0092-8674(00)80323-2. PMID: 9230312.

652. Andrini E, Marchese PV, De Biase D, Mosconi C, Siepe G, Panzuto F, Ardizzoni A, Campana D, Lamberti G. Large Cell Neuroendocrine Carcinoma of the Lung: Current Understanding and Challenges. J Clin Med. 2022 Mar 7;11(5):1461. doi: 10.3390/jcm11051461. PMID: 35268551; PMCID: PMC8911276.

653. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. Nature. 2014 Jul 31;511(7511):543-50. doi: 10.1038/nature13385. Epub

2014 Jul 9. Erratum in: Nature. 2014 Oct 9;514(7521):262. Rogers, K [corrected to Rodgers, K]. Erratum in: Nature. 2018 Jul;559(7715):E12. PMID: 25079552; PMCID: PMC4231481.

654. Axelsson AS, Mahdi T, Nenonen HA, Singh T et al. Sox5 regulates beta-cell phenotype and is reduced in type 2 diabetes. Nat Commun. 2017 Jun 6; 8:15652. doi: 10.1038/ncomms15652. PMID: 28585545; PMCID: PMC5467166.

655. Chi P, Chen Y, Zhang L, Guo X, Wongvipat J, Shamu T, Fletcher JA, Dewell S, Maki RG, Zheng D, Antonescu CR, Allis CD, Sawyers CL. ETV1 is a lineage survival factor that cooperates with KIT in gastrointestinal stromal tumours. Nature. 2010 Oct 14;467(7317):849-53. doi: 10.1038/nature09409. Epub 2010 Oct 3. PMID: 20927104; PMCID: PMC2955195.

656. Sun T, Dong H, Wu L, Kane M, Rowitch DH, Stiles CD. Cross-repressive interaction of the Olig2 and Nkx2.2 transcription factors in developing neural tube associated with formation of a specific physical complex. J Neurosci. 2003 Oct 22;23(29):9547-56. doi: 10.1523/JNEUROSCI.23-29-09547.2003. PMID: 14573534; PMCID: PMC6740479.

657. Lin TP, Chang YT, Lee SY, Campbell M, Wang TC, Shen SH, Chung HJ, Chang YH, Chiu AW, Pan CC, Lin CH, Chu CY, Kung HJ, Cheng CY, Chang PC. REST reduction is essential for hypoxia-induced neuroendocrine differentiation of prostate cancer cells by activating autophagy signaling. Oncotarget. 2016 May 3;7(18):26137-51. doi: 10.18632/oncotarget.8433. PMID: 27034167; PMCID: PMC5041970.

658. Labrecque MP, Coleman IM, Brown LG, True LD et al. Molecular profiling stratifies diverse phenotypes of treatment-refractory metastatic castration-resistant prostate cancer. J Clin Invest. 2019 Jul 30;129(10):4492-4505. doi: 10.1172/JCI128212. PMID: 31361600; PMCID: PMC6763249.

659. Carpenter RL, Gökmen-Polar Y. HSF1 as a Cancer Biomarker and Therapeutic Target. Curr Cancer Drug Targets. 2019;19(7):515-524. doi: 10.2174/1568009618666181018162117. PMID: 30338738; PMCID: PMC6472998.

660. Baird NA, Turnbull DW, Johnson EA. Induction of the heat shock pathway during hypoxia requires regulation of heat shock factor by hypoxia-inducible factor-1. J Biol Chem. 2006 Dec 15;281(50):38675-81. doi: 10.1074/jbc.M608013200. Epub 2006 Oct 13. PMID: 17040902.

661. Sears TK, Angelastro JM. The transcription factor ATF5: role in cellular differentiation, stress responses, and cancer. Oncotarget. 2017 Sep 20;8(48):84595-84609. doi: 10.18632/oncotarget.21102. PMID: 29137451; PMCID: PMC5663623.

662. Li G, Xu Y, Guan D, Liu Z, Liu DX. HSP70 protein promotes survival of C6 and U87 glioma cells by inhibition of ATF5 degradation. J Biol Chem. 2011 Jun 10;286(23):20251-9. doi: 10.1074/jbc.M110.211771. Epub 2011 Apr 25. PMID: 21521685; PMCID: PMC3121475.

663. Jin K, Xiang M. Transcription factor Ptf1a in development, diseases and reprogramming. Cell Mol Life Sci. 2019 Mar;76(5):921-940. doi: 10.1007/s00018-018-2972-z. Epub 2018 Nov 23. PMID: 30470852.

664. Borromeo MD, Meredith DM, Castro DS, Chang JC, Tung KC, Guillemot F, Johnson JE. A transcription factor network specifying inhibitory versus excitatory neurons in the dorsal spinal cord. Development. 2014 Jul;141(14):2803-12. doi: 10.1242/dev.105866. Epub 2014 Jun 12. Erratum in: Development. 2014 Aug;141(15):3102. Erratum in: Development. 2017 Jul 1;144(13):2539. PMID: 24924197; PMCID: PMC4197617.

665. Mona B, Villarreal J, Savage TK, Kollipara RK, Boisvert BE, Johnson JE. Positive autofeedback regulation of Ptf1a transcription generates the levels of PTF1A required to generate itch circuit neurons. Genes Dev. 2020 May 1;34(9-10):621-636. doi: 10.1101/gad.332577.119. Epub 2020 Apr 2. PMID: 32241803; PMCID: PMC7197352.

666. Meredith DM, Borromeo MD, Deering TG, Casey BH, Savage TK, Mayer PR, Hoang C, Tung KC, Kumar M, Shen C, Swift GH, Macdonald RJ, Johnson JE. Program specificity for Ptf1a in pancreas versus neural tube development correlates with distinct collaborating cofactors and chromatin accessibility. Mol Cell Biol. 2013 Aug;33(16):3166-79. doi: 10.1128/MCB.00364-13. Epub 2013 Jun 10. PMID: 23754747; PMCID: PMC3753895.

667. Yu T, Zhou F, Tian W, Xu R, Wang B, Zeng A, Zhou Z, Li M, Wang Y, Zhang J. EZH2 interacts with HP1BP3 to epigenetically activate WNT7B that promotes temozolomide resistance in glioblastoma. Oncogene. 2023 Feb;42(6):461-470. doi: 10.1038/s41388-022-02570-w. Epub 2022 Dec 14. PMID: 36517590.

668. Caggiano C, Pieraccioli M, Panzeri V, Sette C, Bielli P. c-MYC empowers transcription and productive splicing of the oncogenic splicing factor Sam68 in cancer. Nucleic Acids Res. 2019 Jul 9;47(12):6160-6171. doi: 10.1093/nar/gkz344. PMID: 31066450; PMCID: PMC6614821.

669. Benoit YD, Mitchell RR, Risueño RM, Orlando L, Tanasijevic B et al. Sam68 Allows Selective Targeting of Human Cancer Stem Cells. Cell Chem Biol. 2017 Jul 20;24(7):833-844.e9. doi: 10.1016/j.chembiol.2017.05.026. Epub 2017 Jun 22. PMID: 28648376.

670. Fu K, Wan F. Sam68 Offers Selectively Aimed Modulation of Transcription in Cancer Stem Cells. Cell Chem Biol. 2017 Jul 20;24(7):777-779. doi: 10.1016/j.chembiol.2017.07.003. PMID: 28732196; PMCID: PMC5909376.

671. Farini D, Cesari E, Weatheritt RJ, La Sala G, Naro C, Pagliarini V, Bonvissuto D, Medici V, Guerra M, Di Pietro C, Rizzo FR, Musella A, Carola V, Centonze D, Blencowe BJ, Marazziti D, Sette C. A Dynamic Splicing Program Ensures Proper Synaptic Connections in the Developing Cerebellum. Cell Rep. 2020 Jun 2;31(9):107703. doi: 10.1016/j.celrep.2020.107703. PMID: 32492419.

672. Wade AK, Liu Y, Bethea MM, Toren E, Tse HM, Hunter CS. LIM-domain transcription complexes interact with ring-finger ubiquitin ligases and thereby impact islet β-cell function. J Biol Chem. 2019 Aug 2;294(31):11728-11740. doi: 10.1074/jbc.RA118.006985. Epub 2019 Jun 11. PMID: 31186351; PMCID: PMC6682733.

673. Hashimoto Y, Muramatsu K, Kunii M, Yoshimura S, Yamada M, Sato T, Ishida Y, Harada R, Harada A. Uncovering genes required for neuronal morphology by morphology-based gene trap screening with a revertible retrovirus vector. FASEB J. 2012 Nov;26(11):4662-74. doi: 10.1096/fj.12-207530. Epub 2012 Aug 8. PMID: 22874834; PMCID: PMC3475256.

674. Beli P, Lukashchuk N, Wagner SA, Weinert BT, Olsen JV, Baskcomb L, Mann M, Jackson SP, Choudhary C. Proteomic investigations reveal a role for RNA processing factor THRAP3 in the DNA damage response. Mol Cell. 2012 Apr 27;46(2):212-25. doi: 10.1016/j.molcel.2012.01.026. Epub 2012 Mar 15. PMID: 22424773; PMCID: PMC3565437.

675. Jiang F, Hedaya OM, Khor E, Wu J, Auguste M, Yao P. RNA binding protein PRRC2B mediates translation of specific mRNAs and regulates cell cycle progression. Nucleic Acids

Res. 2023 Jun 23;51(11):5831-5846. doi: 10.1093/nar/gkad322. PMID: 37125639; PMCID: PMC10287950.

676. Livshits G, Alonso-Curbelo D, Morris JP 4th, Koche R, Saborowski M, Wilkinson JE, Lowe SW. Arid1a restrains Kras-dependent changes in acinar cell identity. Elife. 2018 Jul 17;7: e35216. doi: 10.7554/eLife.35216. PMID: 30014851; PMCID: PMC6050044.

677. Kuo TL, Cheng KH, Chen LT, Hung WC. ARID1A loss in pancreas leads to islet developmental defect and metabolic disturbance. iScience. 2022 Dec 27;26(1):105881. doi: 10.1016/j.isci.2022.105881. PMID: 36654862; PMCID: PMC9840936.

678. Wang W, Friedland SC, Guo B, O'Dell MR, Alexander WB et al. ARID1A, a SWI/SNF subunit, is critical to acinar cell homeostasis and regeneration and is a barrier to transformation and epithelial-mesenchymal transition in the pancreas. Gut. 2019 Jul;68(7):1245-1258. doi: 10.1136/gutjnl-2017-315541. Epub 2018 Sep 18. PMID: 30228219; PMCID: PMC6551318.

679. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013 Jun;45(6):580-5. doi: 10.1038/ng.2653. PMID: 23715323; PMCID: PMC4010069.

680. Lei Q, Jeong Y, Misra K, Li S, Zelman AK, Epstein DJ, Matise MP. Wnt signaling inhibitors regulate the transcriptional response to morphogenetic Shh-Gli signaling in the neural tube. Dev Cell. 2006 Sep;11(3):325-37. doi: 10.1016/j.devcel.2006.06.013. PMID: 16950124.

681. Mallen-St Clair J, Soydaner-Azeloglu R, Lee KE, Taylor L, Livanos A, Pylayeva-Gupta Y, Miller G, Margueron R, Reinberg D, Bar-Sagi D. EZH2 couples pancreatic regeneration to neoplastic progression. Genes Dev. 2012 Mar 1;26(5):439-44. doi: 10.1101/gad.181800.111. PMID: 22391448; PMCID: PMC3305982.

682. Dumasia NP, Pethe PS. Pancreas development and the Polycomb group protein complexes. Mech Dev. 2020 Dec; 164:103647. doi: 10.1016/j.mod.2020.103647. Epub 2020 Sep 28. PMID: 32991980.

683. Li D, Wang HL, Huang X, Gu X, Xue W, Xu Y. Identification and Functional Characterization of a New Splicing Variant of EZH2 in the Central Nervous System. Int J Biol Sci. 2019 Jan 6;15(1):69-80. doi: 10.7150/ijbs.28129. PMID: 30662348; PMCID: PMC6329929.

684. Zhang M, Zhang Y, Xu Q, Crawford J, Qian C, Wang GH, Qian J, Dong XZ, Pletnikov MV, Liu CM, Zhou FQ. Neuronal Histone Methyltransferase EZH2 Regulates Neuronal Morphogenesis, Synaptic Plasticity, and Cognitive Behavior in Mice. Neurosci Bull. 2023 Oct;39(10):1512-1532. doi: 10.1007/s12264-023-01074-1. Epub 2023 Jun 16. PMID: 37326884; PMCID: PMC10533778.

685. Zemke M, Draganova K, Klug A, Schöler A, Zurkirchen L, Gay MH, Cheng P, Koseki H, Valenta T, Schübeler D, Basler K, Sommer L. Loss of Ezh2 promotes a midbrain-to-forebrain identity switch by direct gene derepression and Wnt-dependent regulation. BMC Biol. 2015 Nov 30; 13:103. doi: 10.1186/s12915-015-0210-9. PMID: 26621269; PMCID: PMC4665911.

686. Pereira JD, Sansom SN, Smith J, Dobenecker MW, Tarakhovsky A, Livesey FJ. Ezh2, the histone methyltransferase of PRC2, regulates the balance between self-renewal and differentiation in the cerebral cortex. Proc Natl Acad Sci U S A. 2010 Sep 7;107(36):15957-62. doi: 10.1073/pnas.1002530107. Epub 2010 Aug 23. PMID: 20798045; PMCID: PMC2936600.

687. Kim J, Lee Y, Lu X, Song B, Fong KW, Cao Q, Licht JD, Zhao JC, Yu J. Polycomb- and Methylation-Independent Roles of EZH2 as a Transcription Activator. Cell Rep. 2018 Dec 4;25(10):2808-2820.e4. doi: 10.1016/j.celrep.2018.11.035. PMID: 30517868; PMCID: PMC6342284.

688. Yu YL, Chou RH, Chen LT, Shyu WC, Hsieh SC, Wu CS, Zeng HJ, Yeh SP, Yang DM, Hung SC, Hung MC. EZH2 regulates neuronal differentiation of mesenchymal stem cells through PIP5K1C-dependent calcium signaling. J Biol Chem. 2011 Mar 18;286(11):9657-67. doi: 10.1074/jbc.M110.185124. Epub 2011 Jan 7. PMID: 21216957; PMCID: PMC3059014.

689. Kazanjian A, Gross EA, Grimes HL. The growth factor independence-1 transcription factor: new functions and new insights. Crit Rev Oncol Hematol. 2006 Aug;59(2):85-97. doi: 10.1016/j.critrevonc.2006.02.002. Epub 2006 May 23. PMID: 16716599; PMCID: PMC2830489.

690. Kazanjian A, Wallis D, Au N, Nigam R, Venken KJ, Cagle PT, Dickey BF, Bellen HJ, Gilks CB, Grimes HL. Growth factor independence-1 is expressed in primary human neuroendocrine lung carcinomas and mediates the differentiation of murine pulmonary neuroendocrine cells. Cancer Res. 2004 Oct 1;64(19):6874-82. doi: 10.1158/0008-5472.CAN-04-0633. PMID: 15466176.

691. Briggs LE, Takeda M, Cuadra AE, Wakimoto H et al. Perinatal loss of Nkx2-5 results in rapid conduction and contraction defects. Circ Res. 2008 Sep 12;103(6):580-90. doi: 10.1161/CIRCRESAHA.108.171835. Epub 2008 Aug 8. PMID: 18689573; PMCID: PMC2590500.

692. Pan JB, Hu SC, Shi D, Cai MC, Li YB, Zou Q, Ji ZL. PaGenBase: a pattern gene database for the global and dynamic understanding of gene function. PLoS One. 2013 Dec 2;8(12): e80747. doi: 10.1371/journal.pone.0080747. PMID: 24312499; PMCID: PMC3846610.

693. Piñero J, Bravo À, Queralt-Rosinach N, Gutiérrez-Sacristán A, Deu-Pons J, Centeno E, García-García J, Sanz F, Furlong LI. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. Nucleic Acids Res. 2017 Jan 4;45(D1): D833-D839. doi: 10.1093/nar/gkw943. Epub 2016 Oct 19. PMID: 27924018; PMCID: PMC5210640.

694. Feng X, Juan AH, Wang HA, Ko KD, Zare H, Sartorelli V. Polycomb Ezh2 controls the fate of GABAergic neurons in the embryonic cerebellum. Development. 2016 Jun 1;143(11):1971-80. doi: 10.1242/dev.132902. Epub 2016 Apr 11. PMID: 27068104; PMCID: PMC4920161.

695. Burclaff J, Bliton RJ, Breau KA, Ok MT, Gomez-Martinez I, Ranek JS, Bhatt AP, Purvis JE, Woosley JT, Magness ST. A Proximal-to-Distal Survey of Healthy Adult Human Small Intestine and Colon Epithelium by Single-Cell Transcriptomics. Cell Mol Gastroenterol Hepatol. 2022;13(5):1554-1589. doi: 10.1016/j.jcmgh.2022.02.007. Epub 2022 Feb 15. PMID: 35176508; PMCID: PMC9043569.

696. Miano JM, Long X, Fujiwara K. Serum response factor: master regulator of the actin cytoskeleton and contractile apparatus. Am J Physiol Cell Physiol. 2007 Jan;292(1):C70-81. doi: 10.1152/ajpcell.00386.2006. Epub 2006 Aug 23. PMID: 16928770.

697. Herrero MJ, Gitton Y. The untold stories of the speech gene, the FOXP2 cancer gene. Genes Cancer. 2018 Jan;9(1-2):11-38. doi: 10.18632/genesandcancer.169. PMID: 29725501; PMCID: PMC5931254.

698. Fujita-Jimbo E, Nakamura S, Momoi T, et al. Abnormal development of gastrointestinal tract and enteric nervous system of homozygous Foxp2(R552H)-mutated mice. August 8, 2023. Preprint (Version 1). Available at Research Square. doi: 10.21203/rs.3.rs-3201691/v1

699. Wong KK, Gascoyne DM, Soilleux EJ, Lyne L, Spearman H, Roncador G, Pedersen LM, Møller MB, Green TM, Banham AH. FOXP2-positive diffuse large B-cell lymphomas exhibit a poor response to R-CHOP therapy and distinct biological signatures. Oncotarget. 2016 Aug 16;7(33):52940-52956. doi: 10.18632/oncotarget.9507. PMID: 27224915; PMCID: PMC5288160.

700. Kim JH, Hwang J, Jung JH, Lee HJ, Lee DY, Kim SH. Molecular networks of FOXP family: dual biologic functions, interplay with other molecules and clinical implications in cancer progression. Mol Cancer. 2019 Dec 9;18(1):180. doi: 10.1186/s12943-019-1110-3. PMID: 31815635; PMCID: PMC6900861.

701. Chen F, Byrd AL, Liu J, Flight RM, DuCote TJ, Naughton KJ, Song X, Edgin AR, Lukyanchuk A, Dixon DT, Gosser CM, Esoe DP, Jayswal RD, Orkin SH, Moseley HNB, Wang C, Brainson CF. Polycomb deficiency drives a FOXP2-high aggressive state targetable by epigenetic inhibitors. Nat Commun. 2023 Jan 20;14(1):336. doi: 10.1038/s41467-023-35784-x. PMID: 36670102; PMCID: PMC9859827.

702. Choi WW, Boland JL, Lin J. ONECUT2 as a key mediator of androgen receptor-independent cell growth and neuroendocrine differentiation in castration-resistant prostate cancer. Cancer Drug Resist. 2022 Feb 8;5(1):165-170. doi: 10.20517/cdr.2021.108. PMID: 35582526; PMCID: PMC8992592.

703. Guo H, Ci X, Ahmed M, Hua JT, Soares F et al. ONECUT2 is a driver of neuroendocrine prostate cancer. Nat Commun. 2019 Jan 17;10(1):278. doi: 10.1038/s41467-018-08133-6. PMID: 30655535; PMCID: PMC6336817.

704. Rotinen M, You S, Yang J, Coetzee SG, Reis-Sobreiro M et al. ONECUT2 is a targetable master regulator of lethal prostate cancer that suppresses the androgen axis. Nat Med. 2018 Dec;24(12):1887-1898. doi: 10.1038/s41591-018-0241-1. Epub 2018 Nov 26. PMID: 30478421; PMCID: PMC6614557.

705. Li Y, Yang XH, Fang SJ, Qin CF, Sun RL, Liu ZY, Jiang BY, Wu X, Li G. HOXA7 stimulates human hepatocellular carcinoma proliferation through cyclin E1/CDK2. Oncol Rep. 2015 Feb;33(2):990-6. doi: 10.3892/or.2014.3668. Epub 2014 Dec 11. PMID: 25501982.

706. Chakrabarti R, Hwang J, Andres Blanco M, Wei Y, Lukačišin M et al. Elf5 inhibits the epithelial-mesenchymal transition in mammary gland development and breast cancer metastasis by transcriptionally repressing Snail2. Nat Cell Biol. 2012 Nov;14(11):1212-22. doi: 10.1038/ncb2607. Epub 2012 Oct 21. PMID: 23086238; PMCID: PMC3500637.

707. Kim MP, Li X, Deng J, Zhang Y, Dai B et al. Oncogenic KRAS Recruits an Expansive Transcriptional Network through Mutant p53 to Drive Pancreatic Cancer Metastasis. Cancer Discov. 2021 Aug;11(8):2094-2111. doi: 10.1158/2159-8290.CD-20-1228. Epub 2021 Apr 10. PMID: 33839689; PMCID: PMC8338884.

708. Straining R, Eighmy W. Tazemetostat: EZH2 Inhibitor. J Adv Pract Oncol. 2022 Mar;13(2):158-163. doi: 10.6004/jadpro.2022.13.2.7. Epub 2022 Mar 25. PMID: 35369397; PMCID: PMC8955562.

709. Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Res. 2019 Jul 2;47(W1):W199-W205. doi: 10.1093/nar/gkz401. PMID: 31114916; PMCID: PMC6602449.

710. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, Chang Z, Woolsey J. DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res. 2006 Jan 1;34(Database issue): D668-72. doi: 10.1093/nar/gkj067. PMID: 16381955; PMCID: PMC1347430.

711. Jourquin J, Duncan D, Shi Z, Zhang B. GLAD4U: deriving and prioritizing gene lists from PubMed literature. BMC Genomics. 2012;13 Suppl 8(Suppl 8):S20. doi: 10.1186/1471-2164-13-S8-S20. Epub 2012 Dec 17. PMID: 23282288; PMCID: PMC3535723.

712. Kunz PL, Reidy-Lagunes D, Anthony LB, Bertino EM, Brendtro K, Chan JA, Chen H, Jensen RT, Kim MK, Klimstra DS, Kulke MH, Liu EH, Metz DC, Phan AT, Sippel RS, Strosberg JR, Yao JC; North American Neuroendocrine Tumor Society. Consensus guidelines for the management and treatment of neuroendocrine tumors. Pancreas. 2013 May;42(4):557-77. doi: 10.1097/MPA.0b013e31828e34a4. PMID: 23591432; PMCID: PMC4304762.

713. Palmieri LJ, Dermine S, Barré A, Dhooge M, Brezault C, Cottereau AS, Coriat R. Medical Treatment of Advanced Pancreatic Neuroendocrine Neoplasms. J Clin Med. 2020 Jun 15;9(6):1860. doi: 10.3390/jcm9061860. PMID: 32549203; PMCID: PMC7355438.

714. Morizane C, Machida N, Honma Y, Okusaka T, Boku N, Kato K, Nomura S, Hiraoka N, Sekine S, Taniguchi H, Okano N, Yamaguchi K, Sato T, Ikeda M, Mizuno N, Ozaka M, Kataoka T, Ueno M, Kitagawa Y, Terashima M, Furuse J; Japan Clinical Oncology Group (JCOG). Effectiveness of Etoposide and Cisplatin vs Irinotecan and Cisplatin Therapy for Patients With Advanced Neuroendocrine Carcinoma of the Digestive System: The TOPIC-NEC Phase 3 Randomized Clinical Trial. JAMA Oncol. 2022 Oct 1;8(10):1447-1455. doi: 10.1001/jamaoncol.2022.3395. PMID: 35980649; PMCID: PMC9389440.

715. Iwasa S, Morizane C, Okusaka T, Ueno H, Ikeda M, Kondo S, Tanaka T, Nakachi K, Mitsunaga S, Kojima Y, Hagihara A, Hiraoka N. Cisplatin and etoposide as first-line chemotherapy for poorly differentiated neuroendocrine carcinoma of the hepatobiliary tract and pancreas. Jpn J Clin Oncol. 2010 Apr;40(4):313-8. doi: 10.1093/jjco/hyp173. Epub 2010 Jan 4. PMID: 20047862.

716. Kawazoe A, Itahashi K, Yamamoto N, Kotani D, Kuboki Y, Taniguchi H, Harano K, Naito Y, Suzuki M, Fukutani M, Higuchi T, Ikeno T, Wakabayashi M, Sato A, Koyama S, Nishikawa H, Shitara K. TAS-116 (Pimitespib), an Oral HSP90 Inhibitor, in Combination with Nivolumab in Patients with Colorectal Cancer and Other Solid Tumors: An Open-Label, Dose-Finding, and Expansion Phase Ib Trial (EPOC1704). Clin Cancer Res. 2021 Dec 15;27(24):6709-6715. doi: 10.1158/1078-0432.CCR-21-1929. Epub 2021 Sep 30. PMID: 34593531.

717. Doi T, Yamamoto N, Ohkubo S. Pimitespib for the treatment of advanced gastrointestinal stromal tumors and other tumors. Future Oncol. 2023 Dec 5. doi: 10.2217/fon-2022-1172. Epub ahead of print. PMID: 38050698.

718. Teranishi R, Takahashi T, Obata Y, Nishida T, Ohkubo S et al. Combination of pimitespib (TAS-116) with sunitinib is an effective therapy for imatinib-resistant gastrointestinal stromal tumors. Int J Cancer. 2023 Jun 15;152(12):2580-2593. doi: 10.1002/ijc.34461. Epub 2023 Feb 21. PMID: 36752576.

719. Lu X, Xiao L, Wang L, Ruden DM. Hsp90 inhibitors and drug resistance in cancer: the potential benefits of combination therapies of Hsp90 inhibitors and other anti-cancer drugs. Biochem Pharmacol. 2012 Apr 15;83(8):995-1004. doi: 10.1016/j.bcp.2011.11.011. Epub 2011 Nov 22. PMID: 22120678; PMCID: PMC3299878

720. Rutherford SL, Lindquist S. Hsp90 as a capacitor for morphological evolution. Nature. 1998 Nov 26;396(6709):336-42. doi: 10.1038/24550. PMID: 9845070.

721. Li ZN, Luo Y. HSP90 inhibitors and cancer: Prospects for use in targeted therapies (Review). Oncol Rep. 2023 Jan;49(1):6. doi: 10.3892/or.2022.8443. Epub 2022 Nov 11. PMID: 36367182; PMCID: PMC9685368.

722. Solit DB, Osman I, Polsky D, Panageas KS, Daud A, Goydos JS, Teitcher J, Wolchok JD, Germino FJ, Krown SE, Coit D, Rosen N, Chapman PB. Phase II trial of 17-allylamino-17-demethoxygeldanamycin in patients with metastatic melanoma. Clin Cancer Res. 2008 Dec 15;14(24):8302-7. doi: 10.1158/1078-0432.CCR-08-1002. PMID: 19088048; PMCID: PMC2629404.

723. Modi S, Stopeck A, Linden H, Solit D, Chandarlapaty S, Rosen N, D'Andrea G, Dickler M, Moynahan ME, Sugarman S, Ma W, Patil S, Norton L, Hannah AL, Hudis C. HSP90 inhibition is effective in breast cancer: a phase II trial of tanespimycin (17-AAG) plus trastuzumab in patients with HER2-positive metastatic breast cancer progressing on trastuzumab. Clin Cancer Res. 2011 Aug 1;17(15):5132-9. doi: 10.1158/1078-0432.CCR-11-0072. Epub 2011 May 10. PMID: 21558407.

724. Nguyen DM, Lorang D, Chen GA, Stewart JH 4th, Tabibi E, Schrump DS. Enhancement of paclitaxel-mediated cytotoxicity in lung cancer cells by 17-allylamino geldanamycin: in vitro and in vivo analysis. Ann Thorac Surg. 2001 Aug;72(2):371-8; discussion 378-9. doi: 10.1016/s0003-4975(01)02787-4. PMID: 11515869.

725. Frigo DE, McDonnell DP. Differential effects of prostate cancer therapeutics on neuroendocrine transdifferentiation. Mol Cancer Ther. 2008 Mar;7(3):659-69. doi: 10.1158/1535-7163.MCT-07-0480. PMID: 18347151.

726. Zschäbitz S, Grüllich C. Lenvantinib: A Tyrosine Kinase Inhibitor of VEGFR 1-3, FGFR 1-4, PDGFRα, KIT and RET. Recent Results Cancer Res. 2018; 211:187-198. doi: 10.1007/978-3-319-91442-8_13. PMID: 30069768.

727. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015 Dec 23;1(6):417-425. doi: 10.1016/j.cels.2015.12.004. PMID: 26771021; PMCID: PMC4707969.

728. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550. doi: 10.1186/s13059-014-0550-8. PMID: 25516281; PMCID: PMC4302049.

729. Hoeller A, Ehrlich L, Golic M, Herse F, Perschel FH, Siwetz M, Henrich W, Dechend R, Huppertz B, Verlohren S. Placental expression of sFlt-1 and PlGF in early preeclampsia vs. early IUGR vs. age-matched healthy pregnancies. Hypertens Pregnancy. 2017 May;36(2):151-160. doi: 10.1080/10641955.2016.1273363. Epub 2017 Feb 10. PMID: 28609172.

730. Rana S, Lemoine E, Granger JP, Karumanchi SA. Preeclampsia: Pathophysiology, Challenges, and Perspectives. Circ Res. 2019 Mar 29;124(7):1094-1112. doi: 10.1161/CIRCRESAHA.118.313276. Erratum in: Circ Res. 2020 Jan 3;126(1):e8. PMID: 30920918.

731. Tian Y, Yang X. A Review of Roles of Uterine Artery Doppler in Pregnancy Complications. Front Med (Lausanne). 2022 Mar 3; 9:813343. doi: 10.3389/fmed.2022.813343. PMID: 35308523; PMCID: PMC8927888.

732. Pilalis A, Souka AP, Antsaklis P, Basayiannis K, Benardis P, Haidopoulos D, Papantoniou N, Mesogitis S, Antsaklis A. Screening for pre-eclampsia and small for gestational age fetuses at the 11-14 weeks scan by uterine artery Dopplers. Acta Obstet Gynecol Scand. 2007;86(5):530-4. doi: 10.1080/00016340601155056. PMID: 17464579.

733. Melchiorre K, Wormald B, Leslie K, Bhide A, Thilaganathan B. First-trimester uterine artery Doppler indices in term and preterm pre-eclampsia. Ultrasound Obstet Gynecol. 2008 Aug;32(2):133-7. doi: 10.1002/uog.5400. PMID: 18615872.

734. Roberts JM, Rich-Edwards JW, McElrath TF, Garmire L, Myatt L; Global Pregnancy Collaboration. Subtypes of Preeclampsia: Recognition and Determining Clinical Usefulness. Hypertension. 2021 May 5;77(5):1430-1441. doi: 10.1161/HYPERTENSIONAHA.120.14781. Epub 2021 Mar 29. PMID: 33775113; PMCID: PMC8103569.

735. Robillard PY, Dekker G, Scioscia M, Bonsante F, Iacobelli S, Boukerrou M, Hulsey TC. The blurring boundaries between placental and maternal preeclampsia: a critical appraisal of 1800 consecutive preeclamptic cases. J Matern Fetal Neonatal Med. 2022 Jul;35(13):2450-2456. doi: 10.1080/14767058.2020.1786516. Epub 2020 Jul 6. PMID: 32627713.

736. O'Gorman N, Wright D, Poon LC, Rolnik DL, Syngelaki A, de Alvarado M, Carbone IF, Dutemeyer V, Fiolna M, Frick A, Karagiotis N, Mastrodima S, de Paco Matallana C, Papaioannou G, Pazos A, Plasencia W, Nicolaides KH. Multicenter screening for pre-eclampsia by maternal factors and biomarkers at 11-13 weeks' gestation: comparison with NICE guidelines and ACOG recommendations. Ultrasound Obstet Gynecol. 2017 Jun;49(6):756-760. doi: 10.1002/uog.17455. Erratum in: Ultrasound Obstet Gynecol. 2017 Dec;50(6):807. PMID: 28295782.

737. Romero R, Espinoza J, Gonçalves LF, Kusanovic JP, Friel L, Hassan S. The role of inflammation and infection in preterm birth. Semin Reprod Med. 2007 Jan;25(1):21-39. doi: 10.1055/s-2006-956773. PMID: 17205421; PMCID: PMC8324073.

738. Zhang C, Schekman R. Syncytin-mediated open-ended membrane tubular connections facilitate the intercellular transfer of cargos including Cas9 protein. Elife. 2023 Mar 10;12:e84391. doi: 10.7554/eLife.84391. PMID: 36896791; PMCID: PMC10112890.

739. Renaud SJ, Kubota K, Rumi MA, Soares MJ. The FOS transcription factor family differentially controls trophoblast migration and invasion. J Biol Chem. 2014 Feb 21;289(8):5025-39. doi: 10.1074/jbc.M113.523746. Epub 2013 Dec 30. PMID: 24379408; PMCID: PMC3931062.

740. Roberts RM, Ezashi T, Schulz LC, Sugimoto J, Schust DJ, Khan T, Zhou J. Syncytins expressed in human placental trophoblast. Placenta. 2021 Sep 15; 113:8-14. doi: 10.1016/j.placenta.2021.01.006. Epub 2021 Jan 15. PMID: 33504453; PMCID: PMC8280254.

741. Altun M, Walter TS, Kramer HB, Herr P, Iphöfer A, Boström J, David Y, Komsany A, Ternette N, Navon A, Stuart DI, Ren J, Kessler BM. The human otubain2-ubiquitin structure provides insights into the cleavage specificity of poly-ubiquitin-linkages. PLoS One. 2015 Jan 15;10(1):e0115344. doi: 10.1371/journal.pone.0115344. PMID: 25590432; PMCID: PMC4295869.

742. Akiyama I, Yoshino O, Osuga Y, Shi J, Harada M, Koga K, Hirota Y, Hirata T, Fujii T, Saito S, Kozuma S. Bone morphogenetic protein 7 increased vascular endothelial growth factor (VEGF)-a expression in human granulosa cells and VEGF receptor expression in endothelial cells. Reprod Sci. 2014 Apr;21(4):477-82. doi: 10.1177/1933719113503411. Epub 2013 Sep 10. PMID: 24023033; PMCID: PMC3960841.

743. Liu F, Soares MJ, Audus KL. Permeability properties of monolayers of the human trophoblast cell line BeWo. Am J Physiol. 1997 Nov;273(5):C1596-604. doi: 10.1152/ajpcell.1997.273.5.C1596. PMID: 9374645.

744. Bode CJ, Jin H, Rytting E, Silverstein PS, Young AM, Audus KL. In vitro models for studying trophoblast transcellular transport. Methods Mol Med. 2006;122:225-39. doi: 10.1385/1-59259-989-3:225. PMID: 16511984; PMCID: PMC2265081.

745. Yazaki K, Matsuno Y, Yoshida K, Sherpa M, Nakajima M, Matsuyama M, Kiwamoto T, Morishima Y, Ishii Y, Hizawa N. ROS-Nrf2 pathway mediates the development of TGF-β1-induced epithelial-mesenchymal transition through the activation of Notch signaling. Eur J Cell Biol. 2021 Sep-Nov;100(7-8):151181. doi: 10.1016/j.ejcb.2021.151181. Epub 2021 Nov 3. PMID: 34763128.

746. Wang R, Liang L, Matsumoto M, Iwata K, Umemura A, He F. Reactive Oxygen Species and NRF2 Signaling, Friends or Foes in Cancer? Biomolecules. 2023 Feb 11;13(2):353. doi: 10.3390/biom13020353. PMID: 36830722; PMCID: PMC9953152.

747. Buelna-Chontal M, Zazueta C. Redox activation of Nrf2 & NF-κB: a double end sword? Cell Signal. 2013 Dec;25(12):2548-57. doi: 10.1016/j.cellsig.2013.08.007. Epub 2013 Aug 29. PMID: 23993959.

748. Bae T, Hallis SP, Kwak MK. Hypoxia, oxidative stress, and the interplay of HIFs and NRF2 signaling in cancer. Exp Mol Med. 2024 Mar 1. doi: 10.1038/s12276-024-01180-8. Epub ahead of print. PMID: 38424190.

749. Xue D, Zhou X, Qiu J. Emerging role of NRF2 in ROS-mediated tumor chemoresistance. Biomed Pharmacother. 2020 Nov;131:110676. doi: 10.1016/j.biopha.2020.110676. Epub 2020 Aug 25. PMID: 32858502.

750. Tossetta G, Fantone S, Piani F, Crescimanno C, Ciavattini A, Giannubilo SR, Marzioni D. Modulation of NRF2/KEAP1 Signaling in Preeclampsia. Cells. 2023 Jun 4;12(11):1545. doi: 10.3390/cells12111545. PMID: 37296665; PMCID: PMC10252212.

751. Bartoszewska S, Collawn JF, Bartoszewski R. The Role of the Hypoxia-Related Unfolded Protein Response (UPR) in the Tumor Microenvironment. Cancers (Basel). 2022 Oct 5;14(19):4870. doi: 10.3390/cancers14194870. PMID: 36230792; PMCID: PMC9562011.

752. Zaarour RF, Azakir B, Hajam EY, Nawafleh H, Zeinelabdin NA, Engelsen AST, Thiery J, Jamora C, Chouaib S. Role of Hypoxia-Mediated Autophagy in Tumor Cell Death and Survival.

Cancers (Basel). 2021 Jan 30;13(3):533. doi: 10.3390/cancers13030533. PMID: 33573362; PMCID: PMC7866864.

753. Chou CC, Chuang HC, Salunke SB, Kulp SK, Chen CS. A novel HIF-1α-integrin-linked kinase regulatory loop that facilitates hypoxia-induced HIF-1α expression and epithelial-mesenchymal transition in cancer cells. Oncotarget. 2015 Apr 10;6(10):8271-85. doi: 10.18632/oncotarget.3186. PMID: 25821081; PMCID: PMC4480751.

754. Zhou L, Xue C, Chen Z, Jiang W, He S, Zhang X. c-Fos is a mechanosensor that regulates inflammatory responses and lung barrier dysfunction during ventilator-induced acute lung injury. BMC Pulm Med. 2022 Jan 6;22(1):9. doi: 10.1186/s12890-021-01801-2. PMID: 34986829; PMCID: PMC8734268.

755. Tugal D, Liao X, Jain MK. Transcriptional control of macrophage polarization. Arterioscler Thromb Vasc Biol. 2013 Jun;33(6):1135-44. doi: 10.1161/ATVBAHA.113.301453. Epub 2013 May 2. PMID: 23640482.

756. Scioli MG, Storti G, D'Amico F, Rodríguez Guzmán R, Centofanti F, Doldo E, Céspedes Miranda EM, Orlandi A. Oxidative Stress and New Pathogenetic Mechanisms in Endothelial Dysfunction: Potential Diagnostic Biomarkers and Therapeutic Targets. J Clin Med. 2020 Jun 25;9(6):1995. doi: 10.3390/jcm9061995. PMID: 32630452; PMCID: PMC7355625.

757. Boden G. Effects of free fatty acids (FFA) on glucose metabolism: significance for insulin resistance and type 2 diabetes. Exp Clin Endocrinol Diabetes. 2003 May;111(3):121-4. doi: 10.1055/s-2003-39781. PMID: 12784183.

758. Sears B, Perry M. The role of fatty acids in insulin resistance. Lipids Health Dis. 2015 Sep 29; 14:121. doi: 10.1186/s12944-015-0123-1. PMID: 26415887; PMCID: PMC4587882.

759. Hauth JC, Clifton RG, Roberts JM, Myatt L, Spong CY, Leveno KJ, Varner MW, Wapner RJ, Thorp JM Jr, Mercer BM, Peaceman AM, Ramin SM, Carpenter MW, Samuels P, Sciscione A, Tolosa JE, Saade G, Sorokin Y, Anderson GD; Eunice Kennedy Shriver National Institute of Child Health and Human Development Maternal-Fetal Medicine Units Network. Maternal insulin resistance and preeclampsia. Am J Obstet Gynecol. 2011 Apr;204(4):327.e1-6. doi: 10.1016/j.ajog.2011.02.024. PMID: 21458622; PMCID: PMC3127262.

760. Scioscia M, Gumaa K, Rademacher TW. The link between insulin resistance and preeclampsia: new perspectives. J Reprod Immunol. 2009 Nov;82(2):100-5. doi: 10.1016/j.jri.2009.04.009. Epub 2009 Jul 22. PMID: 19628283.

761. Schratt G, Philippar U, Hockemeyer D, Schwarz H, Alberti S, Nordheim A. SRF regulates Bcl-2 expression and promotes cell survival during murine embryonic development. EMBO J. 2004 Apr 21;23(8):1834-44. doi: 10.1038/sj.emboj.7600188. Epub 2004 Apr 1. PMID: 15057274; PMCID: PMC394242.

762. Harmon AC, Cornelius DC, Amaral LM, Faulkner JL, Cunningham MW Jr, Wallace K, LaMarca B. The role of inflammation in the pathology of preeclampsia. Clin Sci (Lond). 2016 Mar;130(6):409-19. doi: 10.1042/CS20150702. PMID: 26846579; PMCID: PMC5484393.

763. Cornelius DC. Preeclampsia: From Inflammation to Immunoregulation. Clin Med Insights Blood Disord. 2018 Jan 10; 11:1179545X17752325. doi: 10.1177/1179545X17752325. PMID: 29371787; PMCID: PMC5772493.

764. Li X, Li ZH, Wang YX, Liu TH. A comprehensive review of human trophoblast fusion models: recent developments and challenges. Cell Death Discov. 2023 Oct 10;9(1):372. doi: 10.1038/s41420-023-01670-0. PMID: 37816723; PMCID: PMC10564767.

765. Hamik A, Wang B, Jain MK. Transcriptional regulators of angiogenesis. Arterioscler Thromb Vasc Biol. 2006 Sep;26(9):1936-47. doi: 10.1161/01.ATV.0000232542.42968.e3. Epub 2006 Jun 15. PMID: 16778118.

766. Stocco DM, Wang X, Jo Y, Manna PR. Multiple signaling pathways regulating steroidogenesis and steroidogenic acute regulatory protein expression: more complicated than we thought. Mol Endocrinol. 2005 Nov;19(11):2647-59. doi: 10.1210/me.2004-0532. Epub 2005 Apr 14. PMID: 15831519.

767. Mandal S, Mandal A, Johansson HE, Orjalo AV, Park MH. Depletion of cellular polyamines, spermidine and spermine, causes a total arrest in translation and growth in mammalian cells. Proc Natl Acad Sci U S A. 2013 Feb 5;110(6):2169-74. doi: 10.1073/pnas.1219002110. Epub 2013 Jan 23. PMID: 23345430; PMCID: PMC3568356.

768. Pietilä M, Alhonen L, Halmekytö M, Kanter P, Jänne J, Porter CW. Activation of polyamine catabolism profoundly alters tissue polyamine pools and affects hair growth and female fertility in transgenic mice overexpressing spermidine/spermine N1-acetyltransferase. J Biol Chem. 1997 Jul 25;272(30):18746-51. doi: 10.1074/jbc.272.30.18746. PMID: 9228047.

769. Gong S, Sovio U, Aye IL, Gaccioli F, Dopierala J, Johnson MD, Wood AM, Cook E, Jenkins BJ, Koulman A, Casero RA Jr, Constância M, Charnock-Jones DS, Smith GC. Placental polyamine metabolism differs by fetal sex, fetal growth restriction, and preeclampsia. JCI Insight. 2018 Jul 12;3(13):e120723. doi: 10.1172/jci.insight.120723. PMID: 29997303; PMCID: PMC6124516.

770. Estellés A, Gilabert J, Keeton M, Eguchi Y, Aznar J, Grancha S, Espña F, Loskutoff DJ, Schleef RR. Altered expression of plasminogen activator inhibitor type 1 in placentas from pregnant women with preeclampsia and/or intrauterine fetal growth retardation. Blood. 1994 Jul 1;84(1):143-50. PMID: 8018914.

771. Braekke K, Holthe MR, Harsem NK, Fagerhol MK, Staff AC. Calprotectin, a marker of inflammation, is elevated in the maternal but not in the fetal circulation in preeclampsia. Am J Obstet Gynecol. 2005 Jul;193(1):227-33. doi: 10.1016/j.ajog.2004.11.055. PMID: 16021084.

772. Uszyński M, Maciejewski K, Uszyński W, Kuczyński J. Placenta and myometrium--the two main sources of fibrinolytic components during pregnancy. Gynecol Obstet Invest. 2001;52(3):189-93. doi: 10.1159/000052971. PMID: 11598362.

773. Silverman GA, Bird PI, Carrell RW, Church FC, Coughlin PB, Gettins PG et al. The serpins are an expanding superfamily of structurally similar but functionally diverse proteins. Evolution, mechanism of inhibition, novel functions, and a revised nomenclature. J Biol Chem. 2001 Sep 7;276(36):33293-6. doi: 10.1074/jbc.R100016200. Epub 2001 Jul 2. PMID: 11435447.

774. Yang X, Yu L, Ding Y, Yang M. Diagnostic signature composed of seven genes in HIF-1 signaling pathway for preeclampsia. BMC Pregnancy Childbirth. 2023 Apr 5;23(1):233. doi: 10.1186/s12884-023-05559-9. PMID: 37020283; PMCID: PMC10074875.

775. Kingdom JC, Kaufmann P. Oxygen and placental vascular development. Adv Exp Med Biol. 1999; 474:259-75. doi: 10.1007/978-1-4615-4711-2_20. PMID: 10635006.

776. Jauniaux E, Gulbis B, Burton GJ. Physiological implications of the materno-fetal oxygen gradient in human early pregnancy. Reprod Biomed Online. 2003 Sep;7(2):250-3. doi: 10.1016/s1472-6483(10)61760-9. PMID: 14567901.

777. Hong K, Kim SH, Cha DH, Park HJ. Defective Uteroplacental Vascular Remodeling in Preeclampsia: Key Molecular Factors Leading to Long Term Cardiovascular Disease. Int J Mol Sci. 2021 Oct 18;22(20):11202. doi: 10.3390/ijms222011202. PMID: 34681861; PMCID: PMC8539609.

778. Tal R, Shaish A, Barshack I, Polak-Charcon S, Afek A, Volkov A, Feldman B, Avivi C, Harats D. Effects of hypoxia-inducible factor-1alpha overexpression in pregnant mice: possible implications for preeclampsia and intrauterine growth restriction. Am J Pathol. 2010 Dec;177(6):2950-62. doi: 10.2353/ajpath.2010.090800. Epub 2010 Oct 15. PMID: 20952590; PMCID: PMC2993274.

779. Nevo O, Soleymanlou N, Wu Y, Xu J, Kingdom J, Many A, Zamudio S, Caniggia I. Increased expression of sFlt-1 in in vivo and in vitro models of human placental hypoxia is mediated by HIF-1. Am J Physiol Regul Integr Comp Physiol. 2006 Oct;291(4): R1085-93. doi: 10.1152/ajpregu.00794.2005. Epub 2006 Apr 20. PMID: 16627691; PMCID: PMC6428068.

780. Velegrakis A, Kouvidi E, Fragkiadaki P, Sifakis S. Predictive value of the sFlt-1/PlGF ratio in women with suspected preeclampsia: An update (Review). Int J Mol Med. 2023 Oct;52(4):89. doi: 10.3892/ijmm.2023.5292. Epub 2023 Aug 18. PMID: 37594116; PMCID: PMC10500221.

781. Michiels C, Minet E, Michel G, Mottet D, Piret JP, Raes M. HIF-1 and AP-1 cooperate to increase gene expression in hypoxia: role of MAP kinases. IUBMB Life. 2001 Jul;52(1-2):49-53. doi: 10.1080/15216540252774766. PMID: 11795593.

782. Mallikarjuna P, Zhou Y, Landström M. The Synergistic Cooperation between TGF-β and Hypoxia in Cancer and Fibrosis. Biomolecules. 2022 Apr 25;12(5):635. doi: 10.3390/biom12050635. PMID: 35625561; PMCID: PMC9138354.

783. Mingyuan X, Qianqian P, Shengquan X, Chenyi Y, Rui L, Yichen S, Jinghong X. Hypoxia-inducible factor-1α activates transforming growth factor-β1/Smad signaling and increases collagen deposition in dermal fibroblasts. Oncotarget. 2017 Dec 14;9(3):3188-3197. doi: 10.18632/oncotarget.23225. PMID: 29423039; PMCID: PMC5790456.

784. Zhou H, Zhao C, Wang P, Yang W, Zhu H, Zhang S. Regulators involved in trophoblast syncytialization in the placenta of intrauterine growth restriction. Front Endocrinol (Lausanne). 2023 Jan 31;14:1107182. doi: 10.3389/fendo.2023.1107182. PMID: 36798658; PMCID: PMC9927020.

785. Pérez-Pérez A, Toro A, Vilariño-García T, Maymó J, Guadix P, Dueñas JL, Fernández-Sánchez M, Varone C, Sánchez-Margalet V. Leptin action in normal and pathological pregnancies. J Cell Mol Med. 2018 Feb;22(2):716-727. doi: 10.1111/jcmm.13369. Epub 2017 Nov 21. PMID: 29160594; PMCID: PMC5783877.

786. Ugovšek S, Šebeštjen M. Lipoprotein(a)-The Crossroads of Atherosclerosis, Atherothrombosis and Inflammation. Biomolecules. 2021 Dec 24;12(1):26. doi: 10.3390/biom12010026. PMID: 35053174; PMCID: PMC8773759.

787. Nonn O, Fischer C, Geisberger S, El-Heliebi A, Kroneis T, Forstner D et al. Maternal Angiotensin Increases Placental Leptin in Early Gestation via an Alternative Renin-Angiotensin System Pathway: Suggesting a Link to Preeclampsia. Hypertension. 2021 May 5;77(5):1723-1736. doi: 10.1161/HYPERTENSIONAHA.120.16425. Epub 2021 Mar 29. PMID: 33775117.

788. Gupta S, Agarwal A, Sharma RK. The role of placental oxidative stress and lipid peroxidation in preeclampsia. Obstet Gynecol Surv. 2005 Dec;60(12):807-16. doi: 10.1097/01.ogx.0000193879.79268.59. PMID: 16359563.

789. Kempf T, Eden M, Strelau J, Naguib M, Willenbockel C, Tongers J, Heineke J, Kotlarz D, Xu J, Molkentin JD, Niessen HW, Drexler H, Wollert KC. The transforming growth factor-beta superfamily member growth-differentiation factor-15 protects the heart from ischemia/reperfusion injury. Circ Res. 2006 Feb 17;98(3):351-60. doi: 10.1161/01.RES.0000202805.73038.48. Epub 2006 Jan 5. PMID: 16397141.

790. Aouache R, Biquard L, Vaiman D, Miralles F. Oxidative Stress in Preeclampsia and Placental Diseases. Int J Mol Sci. 2018 May 17;19(5):1496. doi: 10.3390/ijms19051496. PMID: 29772777; PMCID: PMC5983711.

791. George M, Jena A, Srivatsan V, Muthukumar R, Dhandapani VE. GDF 15--A Novel Biomarker in the Offing for Heart Failure. Curr Cardiol Rev. 2016;12(1):37-46. doi: 10.2174/1573403x12666160111125304. PMID: 26750722; PMCID: PMC4807717.

792. Li M, Duan L, Cai YL, Li HY, Hao BC, Chen JQ, Liu HB. Growth differentiation factor-15 is associated with cardiovascular outcomes in patients with coronary artery disease. Cardiovasc Diabetol. 2020 Aug 3;19(1):120. doi: 10.1186/s12933-020-01092-7. PMID: 32746821; PMCID: PMC7398317.

793. Sugulle M, Dechend R, Herse F, Weedon-Fekjaer MS, Johnsen GM, Brosnihan KB, Anton L, Luft FC, Wollert KC, Kempf T, Staff AC. Circulating and placental growth-differentiation factor 15 in preeclampsia and in pregnancy complicated by diabetes mellitus. Hypertension. 2009 Jul;54(1):106-12. doi: 10.1161/HYPERTENSIONAHA.109.130583. Epub 2009 May 26. PMID: 19470878; PMCID: PMC4167791.

794. Chen S, Gong Y, Shen Y, Liu Y, Fu Y, Dai Y, Rehman AU, Tang L, Liu H. INHBA is a novel mediator regulating cellular senescence and immune evasion in colorectal cancer. J Cancer. 2021 Aug 13;12(19):5938-5949. doi: 10.7150/jca.61556. PMID: 34476008; PMCID: PMC8408109.

795. Kawagishi-Hotta M, Hasegawa S, Hasebe Y, Inoue Y, Okuno R, Arima M, Iwata Y, Sugiura K, Akamatsu H. Increase in inhibin beta A/Activin-A expression in the human epidermis and the suppression of epidermal stem/progenitor cell proliferation with aging. J Dermatol Sci. 2022 Jun;106(3):150-158. doi: 10.1016/j.jdermsci.2022.05.001. Epub 2022 May 7. PMID: 35610160.

796. Barber CV, Yo JH, Rahman RA, Wallace EM, Palmer KR, Marshall SA. Activin A and pathologies of pregnancy: a review. Placenta. 2023 May;136:35-41. doi: 10.1016/j.placenta.2023.03.008. Epub 2023 Apr 3. PMID: 37028223.

797. Szenasi NL, Toth E, Balogh A, Juhasz K, Karaszi K, Ozohanics O, Gelencser Z et al. Proteomic identification of membrane-associated placental protein 4 (MP4) as perlecan and

characterization of its placental expression in normal and pathologic pregnancies. PeerJ. 2019 Jun 20;7:e6982. doi: 10.7717/peerj.6982. PMID: 31259093; PMCID: PMC6589330.

798. Qu H, Khalil RA. Role of ADAM and ADAMTS disintegrin and metalloproteinases in normal pregnancy and preeclampsia. Biochem Pharmacol. 2022 Dec; 206:115266. doi: 10.1016/j.bcp.2022.115266. Epub 2022 Oct 1. PMID: 36191626.

799. Ontario Health (Quality). First-Trimester Screening Program for the Risk of Pre-eclampsia Using a Multiple-Marker Algorithm: A Health Technology Assessment. Ont Health Technol Assess Ser. 2022 Dec 8;22(5):1-118. PMID: 37772268; PMCID: PMC10530459.

800. Ghosh AK. Acetyltransferase p300 Is a Putative Epidrug Target for Amelioration of Cellular Aging-Related Cardiovascular Disease. Cells. 2021 Oct 22;10(11):2839. doi: 10.3390/cells10112839. PMID: 34831061; PMCID: PMC8616404.

801. Winter LM, Reinhardt D, Schatter A, Tissen V, Wiora H, Gerlach D, Tontsch-Grunt U, Colbatzky F, Stierstorfer B, Yun SW. Molecular basis of GDF15 induction and suppression by drugs in cardiomyocytes and cancer cells toward precision medicine. Sci Rep. 2023 Jul 26;13(1):12061. doi: 10.1038/s41598-023-38450-w. PMID: 37495707; PMCID: PMC10372009.

802. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M, Liu B, Matthews P, Ong G, Pell J, Silman A, Young A, Sprosen T, Peakman T, Collins R. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. 2015 Mar 31;12(3): e1001779. doi: 10.1371/journal.pmed.1001779. PMID: 25826379; PMCID: PMC4380465.

803. Suen PM, Zou C, Zhang YA, Lau TK, Chan J, Yao KM, Leung PS. PDZ-domain containing-2 (PDZD2) is a novel factor that affects the growth and differentiation of human fetal pancreatic progenitor cells. Int J Biochem Cell Biol. 2008;40(4):789-803. doi: 10.1016/j.biocel.2007.10.020. Epub 2007 Oct 24. PMID: 18037333.

804. Gleason CE, Ning Y, Cominski TP, Gupta R, Kaestner KH, Pintar JE, Birnbaum MJ. Role of insulin-like growth factor-binding protein 5 (IGFBP5) in organismal and pancreatic beta-cell growth. Mol Endocrinol. 2010 Jan;24(1):178-92. doi: 10.1210/me.2009-0167. Epub 2009 Nov 6. PMID: 19897600; PMCID: PMC2802899.

805. Hoffmann A, Spengler D. Transient neonatal diabetes mellitus gene Zac1 impairs insulin secretion in mice through Rasgrf1. Mol Cell Biol. 2012 Jul;32(13):2549-60. doi: 10.1128/MCB.06637-11. Epub 2012 Apr 30. PMID: 22547676; PMCID: PMC3434484.

806. Aguayo-Mazzucato C, van Haaren M, Mruk M, Lee TB Jr, Crawford C, Hollister-Lock J, Sullivan BA, Johnson JW, Ebrahimi A, Dreyfuss JM, Van Deursen J, Weir GC, Bonner-Weir S. β Cell Aging Markers Have Heterogeneous Distribution and Are Induced by Insulin Resistance. Cell Metab. 2017 Apr 4;25(4):898-910.e5. doi: 10.1016/j.cmet.2017.03.015. PMID: 28380379; PMCID: PMC5471618.

807. Hall E, Dekker Nitert M, Volkov P, Malmgren S, Mulder H, Bacos K, Ling C. The effects of high glucose exposure on global gene expression and DNA methylation in human pancreatic islets. Mol Cell Endocrinol. 2018 Sep 5; 472:57-67. doi: 10.1016/j.mce.2017.11.019. Epub 2017 Nov 26. PMID: 29183809.

808. Huang C, Walker EM, Dadi PK, Hu R, Xu Y, Zhang W et al. Synaptotagmin 4 Regulates Pancreatic β Cell Maturation by Modulating the Ca2+ Sensitivity of Insulin Secretion Vesicles.

Dev Cell. 2018 May 7;45(3):347-361.e5. doi: 10.1016/j.devcel.2018.03.013. Epub 2018 Apr 12. PMID: 29656931; PMCID: PMC5962294.

809. Liang X, Song MR, Xu Z, Lanuza GM, Liu Y, Zhuang T, Chen Y, Pfaff SL, Evans SM, Sun Y. Isl1 is required for multiple aspects of motor neuron development. Mol Cell Neurosci. 2011 Jul;47(3):215-22. doi: 10.1016/j.mcn.2011.04.007. Epub 2011 May 4. PMID: 21569850; PMCID: PMC3200226.

810. Li L, Sun F, Chen X, Zhang M. ISL1 is upregulated in breast cancer and promotes cell proliferation, invasion, and angiogenesis. Onco Targets Ther. 2018 Feb 14; 11:781-789. doi: 10.2147/OTT.S144241. PMID: 29497310; PMCID: PMC5818873.

811. Lee H, Fu Z, Koo BH, Sheehan CE, Young GQ, Lin J, Patil DT, Yang Z. The expression of TTF1, CDX2 and ISL1 in 74 poorly differentiated neuroendocrine carcinomas. Ann Diagn Pathol. 2018 Dec; 37:30-34. doi: 10.1016/j.anndiagpath.2018.09.005. Epub 2018 Sep 13. PMID: 30236546.

812. Agaimy A, Erlenbach-Wünsch K, Konukiewitz B, Schmitt AM, Rieker RJ, Vieth M, Kiesewetter F, Hartmann A, Zamboni G, Perren A, Klöppel G. ISL1 expression is not restricted to pancreatic well-differentiated neuroendocrine neoplasms, but is also commonly found in well and poorly differentiated neuroendocrine neoplasms of extrapancreatic origin. Mod Pathol. 2013 Jul;26(7):995-1003. doi: 10.1038/modpathol.2013.40. Epub 2013 Mar 15. PMID: 23503646.

813. Kwan KY, Lam MM, Krsnik Z, Kawasawa YI, Lefebvre V, Sestan N. SOX5 postmitotically regulates migration, postmigratory differentiation, and projections of subplate and deep-layer neocortical neurons. Proc Natl Acad Sci U S A. 2008 Oct 14;105(41):16021-6. doi: 10.1073/pnas.0806791105. Epub 2008 Oct 7. PMID: 18840685; PMCID: PMC2572944.

814. Hu J, Tian J, Zhu S, Sun L, Yu J, Tian H, Dong Q, Luo Q, Jiang N, Niu Y, Shang Z. Sox5 contributes to prostate cancer metastasis and is a master regulator of TGF-β-induced epithelial mesenchymal transition through controlling Twist1 expression. Br J Cancer. 2018 Jan;118(1):88-97. doi: 10.1038/bjc.2017.372. Epub 2017 Nov 9. PMID: 29123266; PMCID: PMC5765224.

815. Chen D, Wang R, Yu C, Cao F, Zhang X, Yan F, Chen L, Zhu H, Yu Z, Feng J. FOX-A1 contributes to acquisition of chemoresistance in human lung adenocarcinoma via transactivation of SOX5. EBioMedicine. 2019 Jun; 44:150-161. doi: 10.1016/j.ebiom.2019.05.046. Epub 2019 May 27. PMID: 31147293; PMCID: PMC6607090.

816. Heeg S, Das KK, Reichert M, Bakir B, Takano S, Caspers J et al. ETS-Transcription Factor ETV1 Regulates Stromal Expansion and Metastasis in Pancreatic Cancer. Gastroenterology. 2016 Sep;151(3):540-553.e14. doi: 10.1053/j.gastro.2016.06.005. Epub 2016 Jun 16. PMID: 27318148; PMCID: PMC5002361.

817. Bae SY, Bergom HE, Day A, Greene JT, Sychev ZE, Larson G et al. ZBTB7A as a novel vulnerability in neuroendocrine prostate cancer. Front Endocrinol (Lausanne). 2023 Mar 29; 14:1093332. doi: 10.3389/fendo.2023.1093332. PMID: 37065756; PMCID: PMC10090553.

818. Bhatlekar S, Viswanathan V, Fields JZ, Boman BM. Overexpression of HOXA4 and HOXA9 genes promotes self-renewal and contributes to colon cancer stem cell overpopulation. J Cell Physiol. 2018 Feb;233(2):727-735. doi: 10.1002/jcp.25981. Epub 2017 Jul 11. PMID: 28464221.

819. Duvall E, Benitez CM, Tellez K, Enge M, Pauerstein PT, Li L, Baek S et al. Single-cell transcriptome and accessible chromatin dynamics during endocrine pancreas development. Proc Natl Acad Sci U S A. 2022 Jun 28;119(26): e2201267119. doi: 10.1073/pnas.2201267119. Epub 2022 Jun 22. PMID: 35733248; PMCID: PMC9245718.

820. Lee SH, Demeterco C, Geron I, Abrahamsson A, Levine F, Itkin-Ansari P. Islet specific Wnt activation in human type II diabetes. Exp Diabetes Res. 2008; 2008:728763. doi: 10.1155/2008/728763. Epub 2009 Jan 20. PMID: 19165345; PMCID: PMC2628766.

821. Raynard C, Ma X, Huna A, Tessier N, Massemin A, Zhu K, Flaman JM et al. NF-κB-dependent secretome of senescent cells can trigger neuroendocrine transdifferentiation of breast cancer cells. Aging Cell. 2022 Jul;21(7):e13632. doi: 10.1111/acel.13632. Epub 2022 Jun 2. PMID: 35653631; PMCID: PMC9282844.

822. April-Monn SL, Kirchner P, Detjen K, Bräutigam K, Trippel MA, Grob T, Statzer C, Maire RS, Kollàr A, Chouchane A, Kunze CA, Horst D, Sadowski MC, Schrader J, Marinoni I, Wiedenmann B, Perren A. Patient derived tumoroids of high grade neuroendocrine neoplasms for more personalized therapies. NPJ Precis Oncol. 2024 Mar 1;8(1):59. doi: 10.1038/s41698-024-00549-2. PMID: 38429350; PMCID: PMC10907580.

10.Supplementary notes and list of Supplementary Tables

Supplemental Note 1. (data integration and correction of batch effects)

In the presented multi-centric study design, samples were collected from various sites to incorporate a diverse set of patients. Confounding by different sampling sites may be induced, but this factor could be adjusted for. Our strategy involved the investigation of batch effect correction for our snRNA-seq data across 4 analyses, employing state-of-the-art deep learning models to regress out any batch effects and the individual, sampling site, gestational age, and technical covariates for the cell types that were analyzed in detail. Importantly, near-perfect integration was revealed by our global average silhouette width (ASW) per batch for our technical replicates (0.92), which was also comparable to our overall integration using all samples (0.863) (refer to Figure 2.11).

Supplemental Note 2. (modeling gestational age in downstream analysis)

As mentioned previously, difficulties are encountered in studying pre-eclamptic diseases due to challenges in matching controls. Placentas affected by early-onset pre-eclampsia (eoPE) typically occur prematurely. While gestational age-matched counterparts without hypertensive conditions could potentially serve as controls, they often present with other underlying obstetric pathologies that lead to preterm birth. Therefore, healthy term samples were included as controls in this thesis, with an additional integration of other preterm datasets to account for the confounding effects of gestational age. In the explorative pilot-study setting of this dissertation, efforts were concentrated on developing a model to longitudinally explore pathomechanisms, hence recruitment focused on the most severe fetal involvement in the disease, fetal growth restriction (FGR). This approach facilitated the presentation of a homogeneous study population, albeit at the expense of depicting the full heterogeneity of the syndrome, such as maternal organ dysfunctions like pulmonary edema or acute kidney insufficiency associated with preeclampsia. FGR was also considered as a potential confounder in our corrections, and clinical cohorts were utilized to validate findings from the adjusted snRNA-seq data. FGR, along with preterm birth, can occur independently of eoPE, potentially overlaying a specific transcriptional profile. For the first time, the multi-omics profile of the preeclamptic maternal-fetal interface is demonstrated with comprehensive adjustment for clinical, biological, and technical confounders.

All Supplementary Tables are enclosed in the FU Box with controlled access to the reviewers and doctoral committee. The detailed description of the files are summarized below:

Supplementary Table 1: Detailed snRNAseq and validation pregnancy cohorts' characteristics

Detailed maternal characteristics from both early and late pregnancy tissue samples of control and early onset pre-eclampsia, utilized for snRNAseq, are provided, categorized by sample ID. Additional tabs include summaries of baseline characteristics comparing groups across our multi-site and large-scale cohorts.

Supplementary Table 2: Cell type or state composition per biological sample sequenced.

The first tab presents the contribution of nuclei numbers per condition, including early and term late controls, as well as pregnancies with early onset pre-eclampsia (eoPE). Each biological replicate was annotated to a specific cell type or state in the maternal-fetal interface, with contributions provided as both absolute and relative values, categorized by gestational age, timepoint, and condition. Additionally, the total number of cells per donor, condition, and cell type/state were reported. The second and third tabs organized the data by tissue of origin.

Supplementary Table 3: Quality Control metrics for Placenta villi

A comprehensive table detailing placenta quality control (QC) measures at various stages: CellRanger metrics, including mean reads per cell, reads confidently mapped to the genome, and fraction of reads in cells; Cellbender filtering outcomes post removal of ambient RNA and random barcode swapping; Scanpy filtering results; and final downstream filtering statistics. For each step, the estimated number of cells, median number of unique molecular identifiers (UMI), and median number of genes per sample were reported at the donor level in the first tab, while the second tab presents these metrics at the cell level. Additional metrics included are total UMI counts (total_counts), number of genes with at least one count (n_genes_by_counts), and the percentage of mitochondrial transcripts per nuclei (pct_counts_MT_genes).

Supplementary Table 4: Quality Control metrics for Decidua

A detailed table summarizing the quality control (QC) assessments of decidua tissue at various stages: CellRanger metrics including mean reads per cell, reads confidently mapped to the genome, and fraction of reads in cells; outcomes of Cellbender filtering post-removal of

ambient RNA and random barcode swapping; results of Scanpy filtering; and final downstream filtering outcomes. The estimated number of cells, median number of unique molecular identifiers (UMI), and median number of genes per sample at each QC step are reported at the donor level in the first tab, while the second tab provides these metrics at the cell level. Additional metrics such as total UMI counts (total_counts), number of genes with at least one count (n_genes_by_counts), and the percentage of mitochondrial transcripts per nuclei (pct_counts_MT_genes) are also included.

Supplementary Table 5: Quality Control metrics for each cell type/state of placenta and decidua.

Total number of UMI counts (total_counts), number of genes with at least one count per nuclei(n_genes_by_counts), %MT-transcripts per nuclei (pct_counts_MT_genes) are reported for each cell type/state.

Supplementary Table 6: Investigation of batch-effects using technical parameters.

The cell-type/state metrics, including absolute silhouette width (ASW), adjusted rand index (ARI), and adjusted mutual information (AMI) scores, are organized into a table per batch, particularly focusing on technical replicates (samples 557_1 and 557_2), late term and early onset pre-eclampsia (eoPE) pregnancies in the first tab, and early pregnancy in the second tab. Additionally, comparisons between our study metrics and those from published datasets are presented in the third tab. ASW values were standardized from 0 to 1, where a score of 1 denotes perfect integration or batch removal. Scores close to zero for ARI/AMI indicate that the batch and cluster labels are independent of each other.

Supplementary Table 7: Markers identifying annotated cell types and states.

The summary presents significant differential genes identified through logistic regression analysis, following covariate correction, that characterize each annotated cell type or state as described in the thesis. These lists were organized based on the tissue of origin (placenta, decidua). The criteria for inclusion are restricted to Bonferroni-adjusted p-values less than 0.05 and log2 fold-change within ± 0.25. Additionally, significant markers relative to vSTB1/2 compared to vSTBjuv were identified using logistic regression, with the list restricted to Bonferroni-adjusted p-values less than 0.01 and log2 fold-change within ± 0.25; the top 100 markers were reported. To avoid composition bias, the number of vSTB1/2 was downsampled to match that of vSTBjuv, as detailed in the third tab.

Supplementary Table 8: Pique-Regi samples QC.

Quality control (QC) metrics are presented for the integrated Pique-Regi samples, comprising 16 samples including placental villi and basal plates. For each sample, the estimated number of cells, median mapped reads per cell, and median number of genes and unique molecular identifiers (UMI) per cell are reported. These QC metrics are then compared between the values reported in Regi et al.'s study published in eLife 2019 and those obtained from our Cellranger run.

Supplementary Table 9: Investigation of batch-effects of harmonized 10X snRNA-seq with external scRNA-seq data (Pique-Regi et. al 2019).

The table presents the absolute silhouette width (ASW), adjusted rand index (ARI), and adjusted mutual information (AMI) scores per batch for placenta and decidua. These metrics are reported for conserved cell types/states that were validated by key markers in both datasets. ASW values range from 0 to 1, where 1 indicates perfect integration or batch removal. ARI/AMI scores close to zero suggest that the batch and cluster labels are independent of each other. Additionally, the composition of conserved cell types/states between the two integrated datasets is provided in a separate tab.

Supplementary Table 10: Preterm genes per cell annotated cell type or state.

Gene sets associated with preterm birth were identified by comparing preterm versus term controls for each conserved cell type/state following data harmonization using snRNA/scRNA-seq data, as outlined in the Methods section. The analysis employed the scVI (change mode) algorithm. Metrics including proba_de (probability of a gene to be differential), bayes Factor, LFC (log-fold change), cell-type proportions (non-zeros proportion), and raw normalized mean counts are reported for both preterm (group 1) and term (group 2) samples. Consistent with scVI conventions, the mean, median, and standard deviation (std) are reported for the LFC effect size variable. Results are presented separately for placenta (first tab) and decidua (second tab).

Supplementary Table 11: eoPE markers per cell type/state.

Differential expression analysis between eoPE and late term pregnancy controls was conducted using multivariate logistic regression. Bonferroni adjustments were applied to p-values for various factors, including the number of UMIs, gene counts, % mitochondrial transcripts, % sex transcripts, and preterm gene scores per nucleus. The resulting list is limited to Bonferroni adjusted p-values < 0.05 and log2 fold-change within ± 0.25. Tabs categorize the results based on cell type/state.

Supplementary Table 12: Differentially expressed genes overlapping in villi compartments.

Significantly dysregulated genes in early onset pre-eclampsia that overlap between Immune, Meso/endothelial and Trophoblast villi compartments.

Supplementary Table 13: Receptor ligand interaction analyses

Receptor ligand differentially upregulated interactions of secreted vSTB and highly expressed dDSTB ligands in eoPE binding to maternal dVEC and dSMC. dEVT ligands interacting with decidua; vCCT ligands interacting with decidua matrisome; up/downregulated interactions within villi and within decidua.

Supplementary Table 14: Pathway enrichment analysis of syncytiotrophoblast functional states- STB1 and STB2.

Metascape enrichment analysis of eoPE genes uniquely dysregulated in vSTB1 (first tab); vSTB2 (second tab); vSTBjuv (third tab) and iRegulon identified transcription factors for enriched unique STB targets (fourth tab). eoPE genes were as per Supplementary Table 11.

Supplementary Table 15: List of syncytial fusion genes and p300 targets.

A compilation of fusion genes associated with placental morphogenesis and physiology is presented in the first tab, while the second tab contains a roster of iRegulon-predicted p300 targets (totaling 133), identified with an enrichment score cutoff of 1.5 and FDR < 0.001. These targets were selected based on a putative regulatory region centered 20 Kb around the transcription start site. The third tab features a gene-attribute matrix detailing gene-transcription factor associations derived from DNA-binding ChIP-seq data obtained from the ENCODE database.

Supplementary Table 16: Recapitulation of early villi trophoblast development.

The complete catalog of developmental regulators is provided per lineage of early villi trophoblast trajectory. Genes are categorized by trophoblast cell type and their role in transitions (dynamic expression across pseudotime) and leaf nodes (characterizing cell-fate commitment per lineage). Both shared and unique transition or leaf genes per path are included. Additionally, the hypergeometric test assessing the significance of overlap between transition markers (positively correlated with pseudotime) and early onset pre-eclampsia (eoPE) markers listed in Supplementary Table 11 is presented in the fourth tab.

Supplementary Table 17: Spatial proteomic results from early villi tissues.
The average gene Z-scores derived from the villi spatial proteomics analysis are presented in the first tab. Proteins that coincide with transition genes in the trophoblast trajectory were listed in the second and third tabs. A summary of the individual biological replicates (P) and technical replicate Z-scores, contributing to the mean cell type score, is provided in the last tab.

Supplementary Table 18: Syncytiotrophoblast dysregulated genes associated with a senescence secretory phenotype.
The intersection of dysregulated genes within the syncytiotrophoblast nuclei states is examined. Genes present in one or more of these states are linked to a senescence phenotype compartment through comparison with the human Senescence Associated Proteins (SASP) atlas (PMID: 31945054).

Supplementary Table 19: Pathway enrichment analysis of dysregulated eoPE genes encompassing lymphoid and myeloid cell types or states at maternal-fetal interface.
Enrichment analysis conducted via Metascape revealed pathways and processes associated with dysregulated eoPE genes in lymphoid cell types, including dNK1, dNK2, and dT cells (first tab), as well as myeloid cell types, such as dMono1, dMAC1, dMAC2, and vHBC (second tab). eoPE genes were as per Supplementary Table 11.

Supplementary Table 20: Pathway enrichment analysis of commonly affected vSTB eoPE genes co-regulated by at least 5 transcription factors/co-regulators.
Metascape enrichment analysis unveiled pathways and processes associated with genes implicated in early-onset pre-eclampsia (eoPE). These genes were found to be dysregulated in at least two villous syncytiotrophoblast (vSTB) sub-states and were inferred to be co-regulated by at least 5 transcription factors/co-regulators, including EP300. Detailed information regarding the eoPE genes can be found in Supplementary Table 11.

Supplementary Table 21: List of abbreviations for gene names discussed in the maternal-fetal interface study.

Supplementary Table 22: Quality Control metrics and composition analysis of 5 panNEC samples.
A comprehensive table summarizing the quality control (QC) assessments of panNEC patient samples (n=5) at various stages, including SoupX and Seurat filtered matrices. The table presents the estimated number of nuclei, median number of unique molecular

identifiers (UMI), and median number of genes per sample at each QC step at the donor level in the first tab. The second tab provides these metrics at the cell level. Additional metrics such as total UMI counts (total_counts), number of genes with at least one count (n_genes_by_counts), and the percentage of mitochondrial transcripts per nuclei (pct_counts_MT_genes) are also included. The third tab displays the distribution of nuclei numbers across panNEC sub-states for each patient sample.

Supplementary Table 23: Markers annotating panNEC sub-states.

The summary outlines significant differentially expressed genes identified via multi-variate logistic regression analysis, with covariate correction, delineating each annotated panNEC sub-state as outlined in the thesis. Inclusion criteria were limited to Bonferroni-adjusted p-values below 0.01 and log2 fold-change within ± 0.25.

Supplementary Table 24: Differentially expressed pathways associated with shared panNEC sub-states and amphicrine progenitor-like.

Metascape enrichment analysis revealed pathways and processes linked to differentially upregulated genes (log2 fold-change >= 0.25 and Bonferroni-adjusted p-values < 0.01) across shared sub-states, including NE, NE proliferating, NE HSP+ (hypoxic), NE Stromal-like, and Stromal (normal). Additionally, pathways associated with Amphicrine progenitor-like cells were documented to corroborate the findings presented in Figure 5.9.

Supplementary Table 25: Signature lists associated with module score analysis.

Gene lists used for computing module scores for analysis presented in Figure 5.11 (pancreatic developmental signatures; first tab), Figure 5.13 (neuroendocrine lung and prostate cancer subtypes and subclasses signatures; second tab), Figure 5.14 (PDAC; third tab), Figure 5.16 (*PTF1A, PAX6* and *NKX2-2* brain and pancreas targets; fourth tab).

Supplementary Table 26: iRegulon inferred transcription factors regulating *EZH2* brain targets.

List of regulated targets controlled by *REST, GFI1, TEAD4*, and *NKX2.5* (first tab) and their intersection (second tab). The TF/regulators were predicted by applying iRegulon on the list of *EZH2* regulated brain targets.

Supplementary Table 27: List of abbreviations for gene names discussed in the panNEC study.

11. Abbreviations

ARI	Adjusted rand index
AMI	Adjusted mutual information
ASW	Absolute silhouette width
Bcell	B lymphocyte
BBKNN	Batch Balanced k-Nearest Neighbors
CCT	Cell column trophoblast
CTB	Cytotrophoblast
CTBp	Cytotrophoblast proliferating
CTBpf	Cytotrophoblast pre-fusion
DC	Dendritic Cells
DSTB	Deported syncytiotrophoblast
DSC	Decidua Stromal Cell
EB	Erythroblast
EpC	Epithelial Cell
eoPE	Early onset pre-eclampsia
EVT	Extravillous trophoblast
FB	Fibroblast
GEP	Gastroentereopancreatic
Granul	Granulocyte
HBC	Hofbauer cell
HBCp	Hofbauer cell proliferating
IHC	Immunohistochemistry
ISS	In-situ sequencing
LEC	Lymphatic endothelial cell
LECp	Lymphatic endothelial cell progenitor
MAC1	M1-like macrophage
MAC2	M2-like macrophage
MC	Myocyte
Mono	Monocyte
MSC	Mesenchymal stem cell
NE	Neuroendocrine
NEC	Neuroendocrine carcinoma
NEN	Neuroendocrine neoplasm
NK	Natural killer cell
NKp	Natural killer cell proliferating
PAMM	Placenta-associated maternal macrophage
panNEC	Pancreatic NEC
PC	Plasma Cell
PE	Pre-eclampsia
pySCENIC	Python Single-Cell rEgulatory Network Inference and Clustering
scRNA-seq	Single cell RNA sequencing
scVI	single-cell Variational Inference
scANVI	single-cell ANnotation using Variational Inference
snRNA-seq	Single nuclei RNA sequencing

STB	Syncytiotrophoblast
STBjuv	Syncytiotrophoblast juvenile
SMC	Smooth muscle cell
Tcell	T lymphocyte
TF	Transcription Factor
t-SNE	t-distributed stochastic neighbor embedding
UMAP	Uniform manifold approximation and projection
Prefix	
d	Denotes decidua as tissue of origin
v	Denotes villi as tissue of origin

*Full gene names are supplied in Supplementary Table 21 (for maternal-fetal interface study) and Supplementary Table 27 (for panNEC study)

12.List of figures

13. List of extended figures

14. List of tables