Molecular mechanisms underlying pancreatic identity and plasticity in mammalian species

INAUGURAL-DISSERTATION

to obtain the academic degree

Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry and Pharmacy

of Freie Universität Berlin

by

Silvia Ruzittu

Berlin, 2020

Die vorliegende Arbeit wurde unter der Anleitung von Dr. Francesca M. Spagnoli am Max-Delbrück-Centrum für Molekulare Medizin (MDC) in Berlin-Buch von Juni 2015 bis Dezember 2019 angefertigt.

1. Gutachter: Dr. Francesca M. Spagnoli

2. Gutachter: Prof. Dr. Simone Spuler

Disputation am: 01.07.2020

Selbstständigkeitserklärung

Ich erkläre gegenüber der Freien Universität Berlin, dass ich die vorliegende Dissertation selbstständig und ohne Benutzung anderer als der angegebenen Quellen und Hilfsmittel angefertigt habe.

Die vorliegende Arbeit ist frei von Plagiaten. Alle Ausführungen, die wörtlich oder inhaltlich aus anderen Schriften entnommen sind, habe ich als solche kenntlich gemacht.

Ich bestätige, dass ich meine Dissertation im Einvernehmen mit meiner Betreuerin in Teilen veröffentlicht habe.

Diese Arbeit wurde in gleicher oder ähnlicher Form noch bei keiner anderen Universität als Prüfungsleistung eingereicht.

09.01.2020

Datum

Unterschrift

TABLE OF CONTENTS

Acknowledgements

First, I would like to express my deepest gratitude to my supervisor Dr. Francesca M. Spagnoli, for giving me the possibility to work in her lab on this exciting project. Her teaching and guidance have been fundamental for my professional and personal growth.

I am grateful to Prof. Dr. Simone Spuler and Prof. Dr. Kai Schmidt-Ott for the critical evaluation of my research project and the precious advice during the Ph.D. committee meetings.

I would like to thank collaboration partners at the MDC and BIH in Berlin and KCL in London, who offered their help in multiple occasions.

I wish to thank the TransCard Ph.D. program for giving me the chance to present and discuss my research studies in seminars and attend useful scientific meetings.

I wish to thank the German Stem Cell Network for providing funding for me to attend international conferences.

A warm thank to past and current members of the Spagnoli team, Heike, Sophie, Andrea, Juan, Heather and Jean- François. I am very lucky to have lived with them this exciting journey. You have been a family to me, giving important scientific and emotional support and unforgettable joyful moments. I am particularly grateful to Laura and David, who constantly supported me in any aspect of my work and life.

I would like to thank the friends of the Italian community at the MDC: they have made my experience in Berlin incredibly full of fun. In particular, an immense thank to Silvia, who has been a perfect mate and invaluable supporting presence. I owe her a lot.

I am profoundly grateful to my mum Adriana, my dad Salvatore and my sister Laura, who constantly supported and motivated me throughout my studies and were always close to me despite the distance. I dedicate all my important achievements to them.

Last but not least, a special thank to Nino, who is a special piece of my life, a source of strength and love.

ABSTRACT

Lineage reprogramming of somatic cells to generate pancreatic β-like cells represents a promising strategy for developing a cell-based therapy to treat diabetes. The close developmental origin with the pancreas and its regenerative ability make the liver an ideal tissue source for generating new β-cells. We recently identified the transcription factor TG-interacting factor 2 (TGIF2) as a developmental regulator of the cell fate decision between liver and pancreas in the mouse. Consistently, stable lentiviral expression of TGIF2 is sufficient to promote a pancreatic progenitor state in adult mouse liver cells, by repressing the hepatic transcriptional program and initiating the pancreatic progenitor one.

The studies of my Ph.D. thesis have focused on further investigating the biological function of TGIF2 in the mouse and translating these findings in human cells. Specifically, I undertook an *in vivo* loss-of-function approach based on the Cre/LoxP recombination system in the mouse to define the requirements of *Tgif2* during pancreas embryonic development. Whole transcriptome analysis showed that TGIF2 acts as regulator of binary choices: first, it establishes and maintains pancreatic identity *instead of* a liver fate; secondly, it controls pancreatic endocrine lineage differentiation at expenses of the acinar one. These findings were further supported by *in vitro* reprogramming experiments, whereby enforced expression of human *TGIF2* in human primary hepatocytes repressed the liver features and promoted the induction of a pancreatic state. Moreover, I expanded the study of TGIF2 reprogramming potentials to different cellular contexts, including fibroblast cells. Overall, the results presented in this work suggest a conserved function of TGIF2 in initiating a pancreatic program in different cellular contexts. Further comprehension of TGIF2 and the molecular mechanisms regulating pancreatic identity and cellular plasticity will ultimately lead to the development of innovative cell-based therapies to cure diabetes.

Keywords: Cell identity, Cell plasticity, Direct lineage reprogramming, Liver, Pancreas

ZUSAMMENFASSUNG

Die Umprogrammierung von somatischen Zellen zur Herstellung pankreatischer β-ähnlicher Zellen stellt eine vielversprechende Strategie für die Entwicklung einer zellulären Therapie zur Behandlung von Diabetes dar. Auf Grund der engen entwicklungsbiologischen Verwandtschaft mit dem Pankreas und ihres regenerativen Potentials, stellt die Leber ein ideales Gewebe dar, welches als zelluläre Quelle für die Herstellung von β-Zellen dienen kann. Vor Kurzem identifizierten wir den Transkriptionsfaktor TG-interacting factor 2 (TGIF2) als einen entwicklungsbiologischen Regulator der Entscheidung von Vorläuferzellen zur pankreatischen und nicht hepatischen Differenzierung in der Maus. Übereinstimmenderweise ist die stabile lentivirale Expression von TGIF2 ausreichend, um in Leberzellen adulter Mäuse eine Zellidentität ähnlich der pankreatischer Vorläuferzellen zu induzieren, indem hepatische Transkriptionsprogramme unterdrückt und pankreatische eingeleitet werden.

In der vorliegenden Arbeit habe ich die biologische Funktion von TGFI2 in der Maus weiter aufgeklärt und meine Befunde auf humane Zellen übertragen. Im Einzelnen habe ich TGIF2 *in vivo* mit Hilfe des Cre/loxP Rekombinationssystem genetisch inaktiviert, um die Erfordernis von TGIF2 während der embryonalen Entwicklung des Pankreas in der Maus zu definieren. Analyse des gesamten Transkriptoms von pankreatischen Vorläuferzellen zeigte, dass TGIF2 als Regulator von zwei binären Entscheidungen während der Zelldifferenzierung agiert. Zum Ersten etabliert und erhält TGIF2 pankreatische Zellidentität während es hepatische Zellidentität unterdrückt. Zum Zweiten, fördert TGIF2 endocrine pankreatische Zelldifferenzierung auf Kosten der Differenzierung von azinären Drüsenzellen. Diese Befunde wurden im Weiteren von *in vitro* Umprogrammierungsexperimenten unterstützt, bei denen die erzwungene Expression von humanen TGIF2 Proteinen in primären humanen Hepatozyten zur Unterdrückung hepatischer Charakteristika und gleichzeitig zur Induktion einer pankreatischen Zellidentität führte. Außerdem habe ich das Potential von TGIF2 zur zellulären Umprogrammierung in weiteren zellulären Kontexten, einschließlich Fibroblasten, untersucht. Zusammengenommen zeigen die Ergebnisse dieser Arbeit eine konservierte Funktion von TGIF2 in der Initiation pankreatischer Zellidentität in verschiedenen zellulären Kontexten. Weitere Untersuchungen zur Funktion von TGIF2 und den molekularen Mechanismen die pankreatische Zellidentität und zelluläre Plastizität regulieren, werden schlussendlich zur Entwicklung von innovativen zellulären Therapien zur Heilung von Diabetes führen.

Stichworte: zelluläre Identität, zelluläre Plastizität, direkte zelluläre Umprogrammierung, Leber, Pankreas

1. INTRODUCTION

1.1 Pancreas and liver, two close metabolic organs

1.1.1 The human adult pancreas

Structure and function of the organ

The pancreas is a vital glandular organ of the digestive system in vertebrates, exerting essential metabolic functions. It is composed of an exocrine domain, which contributes to the body digestive function, and an endocrine compartment, which maintains glucose homeostasis. The exocrine tissue accounts for the majority of the pancreatic mass and comprises acinar and ductal cells. Acinar cells secrete the pancreatic juice, containing digestive enzymes, such as carboxypeptidases, amylase, chymotrypsin, trypsin and lipases, that mediate the breakdown of proteins, carbohydrates and lipids of the food entering the gastro-intestinal tract. The pancreatic juice is transported to the duodenum through the ductal network formed by ductal cells secreting mucin and bicarbonate that neutralizes the gastric acidity in the duodenum[1,2].

The pancreatic endocrine function is performed by clusters of cells, so-called “islets of Langherans”, which are embedded in the exocrine tissue, innerved and surrounded by a vascular network. Pancreatic islets represent approximately the 2% of the total organ mass and comprise 5 endocrine cell types (α, β, PP, δ, ε) releasing different hormones (insulin, glucagon, pancreatic polypepetide, somatostatin, and ghrelin, respectively)[3] (Fig.1).

Insulin and glucagon are the most fundamental peptide hormones released by the pancreatic cells, which act in an antagonistic way to maintain normoglycaemia, aka normal blood glucose levels. Glucagon is a catabolic hormone, which triggers endogenous production of glucose via gluconeogenesis or glycogenolysis in the liver, resulting in increased blood glucose levels in fasting state. By contrast, insulin mediates an anabolic function: it is released in hyperglycaemia and inhibits gluconeogenesis while promoting glucose uptake by insulin-responsive tissues, such as the muscle and fat, in order to lower high blood glucose levels[4].

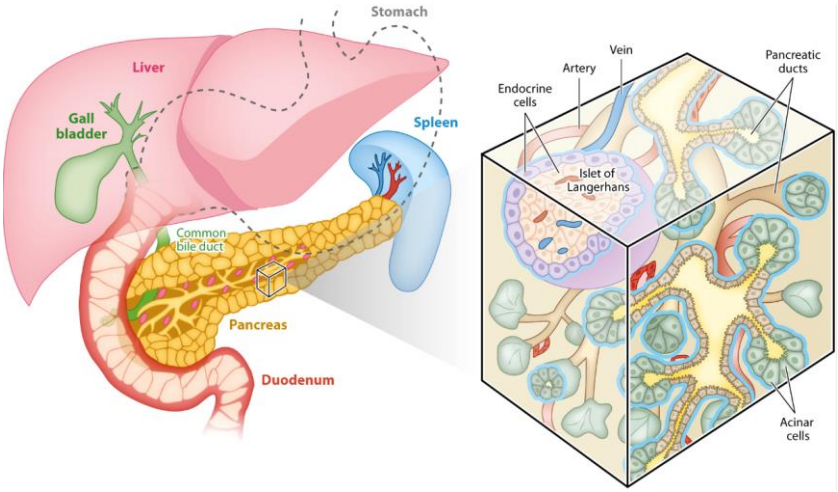

Figure 1. Anatomy of the adult pancreas.
The pancreas is a compound gland of the digestive system and is located in the abdominal cavity, behind the stomach and adjacent to the liver. Morphologically, it is divided into a head, which is connected to the duodenum through the pancreatic duct, and the body and tail, which extend across the midline and contact the spleen. A histological view of the pancreas (right) illustrates the exocrine and the endocrine compartments. Hormone-releasing cells are clustered together in the well-vascularized Islets of Langerhans, while cells secreting digestive enzymes are organized in acini at the termini of the pancreatic ducts, which form a complex intra-organ tree-like network. Taken from Shih et al., 2013.

Physiologic function and pathology of the islet β-cells

The most represented and important cell type in the pancreatic islets is the insulin-secreting β-cell[5]. In physiologic conditions, β-cells are sensitive to the hematic concentration of glucose, which enters the cell through the β-cell specific glucose transporter GLUT-2 and is converted via glycolysis first to pyruvate and subsequently to Acetil-CoA[4,6]. As a result of the metabolic oxidation, released energy is used to form ATP molecules, which induce the closure of the ATP-sensitive K^+ channels and polarization of the plasma membrane. This causes the voltage-dependent Ca^{2+} channels to open and allow Ca^{2+} to enter the cytosol and trigger the fusion of the insulin-containing secretory granules with the cell membrane. The hormone is therefore released in the blood stream and can reach insulin-responsive tissues exerting its biologic function, resulting in a decrease of the blood glucose levels[4,6].

The pancreatic function is compromised in pathological conditions such as pancreatitis, pancreatic adenocarcinoma and diabetes, for which a definitive cure is not available[7–9].

Diabetes is a chronic, metabolic disease that affects over 420 million people worldwide (http://www.who.int/diabetes/en/) and its incidence is dramatically increasing. It exists in different forms and has diverse aetiologies. Type 2 diabetes (T2D) (90% of the cases), is the most common form of the disease, it develops during adulthood and is characterized by β-cell dysfunction, insulin resistance and metabolic stress, which eventually lead to β-cell mass reduction[4,5]. Type 1 diabetes (T1D) (10% of the cases) is an autoimmune disorder characterized by specific autoimmune destruction of pancreatic β-cells, which leads to

insufficient production of insulin and consequent hyperglycaemia[5,10]. Moreover, the so-called maturity onset diabetes of the young (MODY) represents a rare form of hereditary diabetes that is caused by monogenic mutations in regulators of the pancreas embryonic development, which affect the formation of β-cells[11].

β-cell regenerative ability in the adult pancreas is very limited and does not compensate for cell loss in diabetic patients[7]. Lack (T1D) or dysfunction (T2D) of insulin-producing cells result in impaired control of glycaemia, which leads to long-term complications, such as retinopathies, kidney failure, heart attacks, that ultimately reduce life expectancy[5,12]. Patients suffering from diabetes require sustained administration of exogenous insulin throughout life, however this treatment is not a definitive cure and the risk of long-term complications persists. Current efforts in regenerative medicine aim at restoring glycaemic control in diabetic patients either by replacing beta-cells or by preserving or enhancing the remnant endogenous beta-cell mass[13–16] and will be further discussed in Chapter 1.3.

1.1.2 The human adult liver

The liver is the most voluminous organ of our body, which is situated in the abdomen, and plays fundamental roles in numerous physiological processes. It performs a wide range of metabolic functions, such as nutrient processing, maintenance of blood metabolites and protein concentrations, secreting plasma proteins and bile, drug detoxification, cholesterol synthesis transport as well as glycogen storage[17]. During embryonic development the liver is also responsible for haematopoiesis[17].

Various cell types compose the liver. Hepatocytes, which constitute the majority of the parenchymal compartment, and biliary epithelial cells (also called cholangiocytes) are the most abundant populations. Other less represented cell types are sinusoidal endothelial cells, which form the sinusoidal plexus to facilitate blood circulation, Kupffer cells (resident liver macrophages), pit cells (natural killer cells) and hepatic stellate cells, which maintain the extracellular matrix and control the microvascular tone[17] (Fig. 2).

Anatomically, the liver is formed by lobes (four in the human), which are structured into hexagonal functional units called lobules[18,19]. The organ is highly vascularized, receiving blood supply from the hepatic portal vein and the hepatic arteries, which together with a bile duct form the portal triad at each corner of the lobules (Fig. 2). Hepatocytes are arranged in chords and secrete endocrine products directly in the blood stream, which flows through sinusoidal capillaries radially oriented toward the central efferent vein. Bile acid is secreted by the hepatocytes and collected by the surrounding bile duct canaliculi to be transported to the portal triad[17,19]. The bile is subsequently stored in the gall bladder and discharged in the duodenum via the common bile duct to mediate lipid digestion.

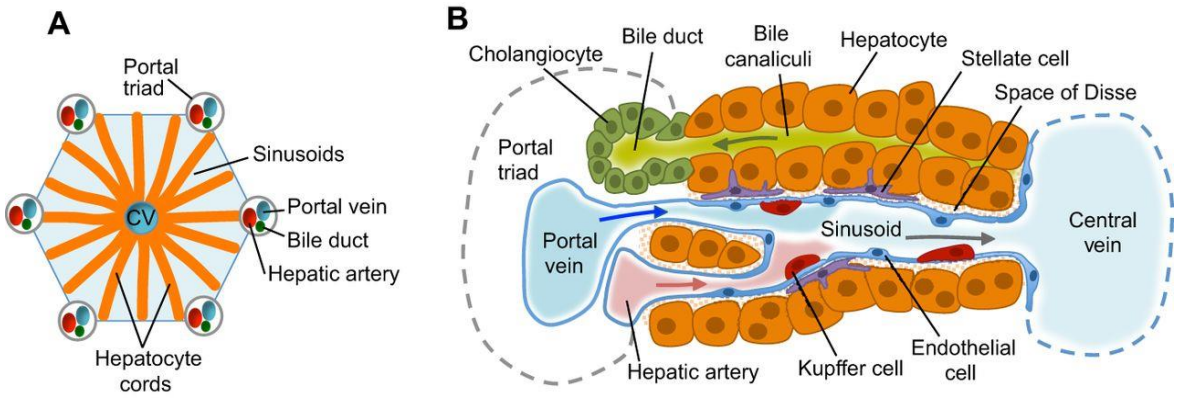

Figure 2. Anatomy and cell types of the adult hepatic lobules.
(A) The functional unit of the liver is the lobule, which displays a hexagonal shape, with a portal triad at each corner and radial cords of hepatocytes converging toward the central vein (CV). The portal triad consists of a portal vein, a hepatic artery and a biliary duct. **(B)** Within each lobule, the single-cell sheets of hepatocytes are lined by sinusoids that carry blood from the portal triad to the central vein. These discontinuous vessels, formed by fenestrated endothelial cells, allow direct release of metabolite products into the blood stream. Other cell types present in the hepatic lobule are stellate cells, located in the space of Disse between the hepatocyte cords and the sinusoids, the Kupffer cells, which are specialized cells of the immune system, and the cholangiocytes, which form the bile ducts. Hepatocytes secrete bile acid into the bile canaliculi that lead to the bile duct. Taken from Gordillo et al., 2015.

The hepatic microenvironment is characterized by a spatial variability pattern defined as “liver zonation”, according to which hepatocytes possess a distinct functional specialization based on their position along the porto-central axis of the liver lobule[20]. “Periportal” hepatocytes (close to the portal veins) are metabolically active and involved in cholesterol synthesis, fatty acid oxidation, and bile acid production. “Pericentral” hepatocytes (adjacent to the central vein) are responsible for glycolysis and xenobiotic metabolism. This sub-specialization occurs in response to a gradient of oxygen, nutrients and metabolites, generated by the blood stream, and morphogens, including Wingless-type MMTV integration site family (WNT) signal[20].

The adult liver is endowed with an exceptional regenerative ability, which can compensate for surgical removal of up to 75% of the total liver mass[14]. Upon injury, terminally differentiated hepatocytes can re-enter cell cycle and proliferate[21,22]. However, in case of more severe insults, hepatocytes are unable to respond to the growth stimuli and other different mechanisms have been proposed to ensure organ mass restoration. Specifically, biliary-like progenitor cells (aka “oval cells”) arising in the ductal region[14,23,24] or diploid cells located around the central vein[25] have been suggested to constitute a stem cell pool in the adult liver. Nevertheless, the presence of such stem cell compartment in the human liver remains an open question. Alternatively, hepatocytes or cholangiocytes have been reported to de-differentiate into a hybrid bi-potent progenitor[26–29] acting as facultative stem cells to maintain tissue turnover in chronically injured liver, underlying intra-hepatic cell plasticity[21].

1.1.3 Embryonic development of pancreas and liver in the mouse

Every single cell of our body acquires a specific cell identity and function by a process called cell fate specification, resulting from a series of cell fate choices during the embryonic development.

The first cell specification event during embryonic life leads to the separation between the embryonic (inner cell mass) and the extra-embryonic (trophoblast) fate within the blastocyst[30]. The embryonic domain, also called epiblast, contains pluripotent cells that will form the three germ layers, ectoderm, mesoderm and endoderm, undergoing a process called gastrulation, that starts around embryonic day (E) 6.5 of mouse development[30] (Fig. 3).

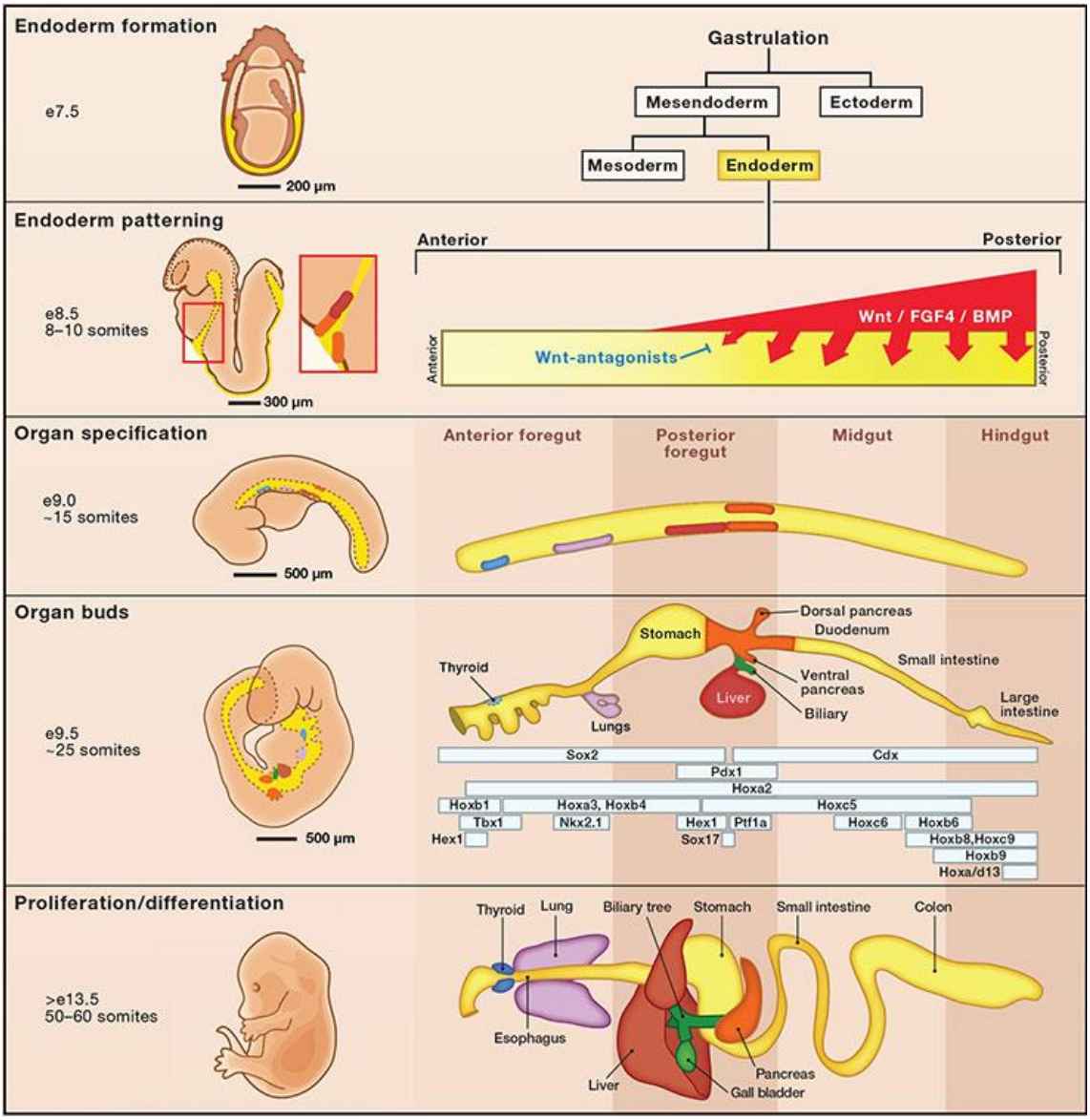

Figure 3. Patterning of the endoderm germ layer.
Overview illustrating the process of establishment of the different organ domains in the endoderm germ layer, from gastrulation to endoderm patterning, organ specification, and subsequent organogenesis. Wnt inhibition imparts an anterior character in the endoderm, while FGF and BMP promote posterior fates. The endoderm ultimately gives rise to thyroid, thymus, lungs, liver, biliary system, pancreas, and intestines. Taken from McGrath and Wells, 2015.

Subsequently, all the different cell lineages are generated by multi-step transition mechanisms. In particular, the definitive endodermal (DE) fate is elicited by high Nodal, a signalling factor of the Transforming Growth Factor β (TGF-β) superfamily[31–33], and is marked by the expression of *Eomesodermin* (*Eomes*)*, Forkhead box protein A2* (*Foxa2*)*, Sex-determining region Y box 17* (*Sox17*)*, C-X-C chemokine receptor type 4* (*CXCR-4*) and members of the *GATA* and *Mix-like* (*Mixl*) family[19,32,34,35] (Fig. 3). The endoderm is patterned along anterior-posterior axis (A/P-axis) into foregut, midgut and hindgut[33,36]. The most anterior part gives rise to the lung, the thymus, the thyroid and the oesophagus, while the liver, the pancreas, the biliary system and the epithelial cells lining the digestive systems. become specified posteriorly[33,35] (Fig. 3).

Patterning of the foregut endoderm

The earliest step in the development of the digestive system is marked by the patterning of the endoderm along the A/P-axis into foregut, midgut and hindgut[33,36] (Fig. 3). This process mainly occurs in response to gradients of soluble factors secreted by the adjacent lateral plate mesoderm and the endoderm itself[1,37]. In particular, Fibroblast growth factor 4 (FGF4), Bone morphogenetic proteins (BMPs), Activin and Retinoic acid (RA) ligands have posteriorizing effects[38,39], while inhibition of WNT signalling promotes the anterior endoderm fates[35] (Fig. 3). Distinct segments of the endoderm layer exhibit different competences to respond to the inductive signals from the surrounding mesoderm and, as a consequence, will be programmed to become the various organs[37]. The anterior-most region of the developing gut tube is named foregut, which maintains the expression of *Hematopoietically-expressed homeobox protein* (*Hhex*)*, Sox2* and *Foxa2* and exclude factors of the *Caudal type homeobox* (*Cdx*) family to define the boundary with the adjacent hindgut domain fates[35] (Fig.3). The foregut comprises a dorsal region, that will give rise to pancreatic tissue only, and a ventral region, that harbours progenitors of the liver, pancreas, gallbladder, and bile ducts[35,40]. At the embryonic stage E8.5, pancreatic and hepatic organ fate specification occurs in the foregut endoderm[36] (Fig. 3; 4; 5; 6).

Signalling cues and factors controlling the specification of the hepato-pancreatic domains

Upon primitive gut tube closure and endoderm patterning, specific tissue interactions with mesodermal tissues instruct the regions of the foregut to acquire distinct positional identities and, consequently, different cell fates[36,37]. The pancreatic and hepatic territories are concomitantly specified in the posterior ventral foregut in close proximity to each other (Fig.4). All the main signalling pathways, such as BMP, FGF, RA, Sonic hedgehog (SHH), Wnt and

Notch, have been reported to play a role at different stages of the liver and pancreas development[1,2,36].

The pancreas emerges as two distinct buds, one dorsal and one ventral, expressing the homeodomain transcription factor *Pancreatic and duodenal homeobox 1* (*Pdx1*)[2,41,42]. Remarkably, pancreatic specification occurs with distinct molecular mechanisms in the ventral and the dorsal foregut, being the two rudiments exposed to different signalling environments and establishing independent cell interactions with the surrounding tissues. Nevertheless, they give rise to the same pancreatic genetic program[36].

Interaction with the notochord, the aorta and the mesenchyme control the specification, growth, morphogenesis and differentiation of the pancreas.

First, the dorsal pancreatic endoderm is initially associated with the lateral plate mesoderm, which secretes RA, and the notochord, that releases Activin-βB and FGF2 ligands, which in turn represses the *Shh* expression in the pre-pancreatic territory[43,44] (Fig. 4). Both RA signalling and inhibition of SHH are necessary for initiation of *Pdx1* expression in the dorsal pancreatic endoderm[45,46]. However, SHH is required later on in embryonic life for the expansion of the pancreatic epithelium and the regulation of the expression of insulin in the mature β-cells[47,48]. At E8.5 the paired dorsal aortae fuse in the midline, between the dorsal pancreatic endoderm and the notochord[49,50] and provide direct inductive influences to the nascent pancreatic bud (Fig. 4; 6). Several *in vitro* tissue recombination studies have shown that the vascular endothelium provides specific cues that are fundamental for proper pancreas development, arguing for a central role of the close physical association of blood vessels and pancreatic epithelium over developmental time[49,50]. In addition, aortic endothelial cells regulate pancreatic induction by promoting survival of mesenchymal *LIM homeobox transcription factor Islet1* *Islet1* (*Isl1*)$^{+}$ cells[51].

The ventral pancreatic region is specified adjacent to the prospective hepato-biliary domain in the posterior ventral foregut[19,52] (Fig. 4). The onset of the two fates is marked by a common set of transcription factors (TFs), such as *Prospero homeobox 1* (*Prox1*), *Hhex* and *Foxa* and *GATA* family[35,53–57]. Fate mapping and lineage tracing experiments in different vertebrate models suggested that both lineages arise from a bipotent progenitor domain[35,40,58,59] and gradually segregate into lineage restricted progenitor populations[58–60]. However, the precise nature of this population as well as the temporal dynamics of the segregation are still undefined. The hepato-pancreatic domain is in contact with the cardiac mesoderm and the septum transversum mesenchyme (STM), which derives from the lateral plate mesoderm[58,61–63].

Experiments performed using explant cultures of E8.0-8.5 mouse endoderm identified FGF signalling released from the cardiac mesoderm and BMPs from the STM as determinant of hepatic fate specification and inhibitors of the ventral *Pdx1* expression domain[55,58,59,64]. In the

absence of these pro-hepatic signals, the “default fate” of this foregut region results to be pancreatic tissue[54,58]. Inhibition of BMP and FGF signalling is therefore essential for the establishment of the pancreatic cell fate in the ventral foregut[63]. However, it is not completely understood how exactly FGF and BMP signalling are controlled *in vivo*. The dosage and the duration of their activity dictate the initiation of the hepatic or pancreatic program, marked by albumin and Pdx1 expression respectively[58,61] (Fig. 4). Moreover, few hours later in development, BMP signalling seems to be necessary to maintain *Pdx1* expression, indicating dynamic changes in BMP and FGF signalling requirements during pancreas organogenesis[65].

FGF10 signalling has been well characterized for its roles in growth and differentiation of pancreatic progenitors in the mouse[66] and also reported or being essential in lineage segregation between hepatic, pancreatic, and biliary tissues in zebrafish[67]. Notch pathway also participates in this fate decision; specifically, its downstream effector Hes-1 is required for gall bladder formation and its loss promotes conversion of biliary tissue into pancreatic fate in the mouse embryo[68,69].

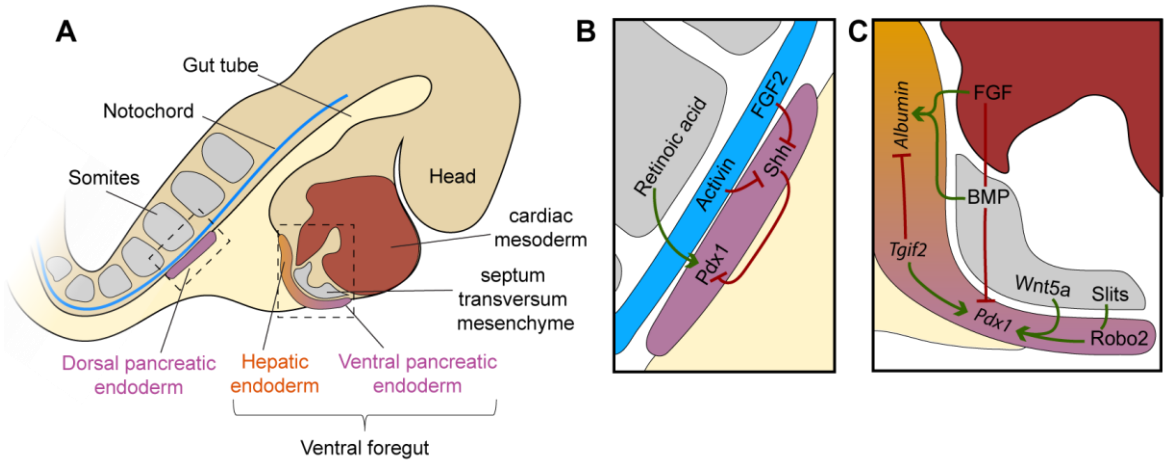

Figure 4. Specification of the hepatic and pancreatic organ domains.
(A) Sagittal view of a mouse embryo at E8.5. The foregut endoderm (light beige) receives inductive signals from neighbouring mesodermal tissues (septum transversum mesenchyme, cardiac mesoderm, somites, notochord) and is consequently specified in hepatic (orange) and pancreatic (purple) organ domains, ventrally and dorsally. **(B)** Schematics of the factors controlling fate specification of the dorsal pancreatic endoderm. Activin and FGF2 secreted by the notochord repress expression of Shh in the presumptive dorsal pancreatic endoderm, which in turn allows Pdx1 expression. Pdx1 expression is further promoted by retinoic acid from the somites. **(C)** Schematic overview of the factors controlling the fate specification of hepatic and ventral pancreatic endoderm. FGFs secreted from the cardiac mesoderm and BMPs produced by the septum transversum mesenchyme promote hepatic fate in the anterior ventral foregut, while suppressing ventral pancreatic identity. In the posterior ventral foregut, extrinsic signalling cues, such as Wnt5a, and cell-intrinsic transcriptional regulators, such as Tgif2, establish ventral pancreatic identity and repress the hepatic fate. Subsequently, specified organ domains turn on hepatic (Albumin) or pancreatic marker genes (Pdx1). Slit ligands from the overlying mesenchyme bind to Robo2 receptors expressed in the ventral pancreatic endoderm and facilitate the maintenance of pancreatic identity. Adapted from Ruzittu et al., 2019.

Recently, an RNA-Seq analysis of hepatic and pancreatic progenitors isolated from mouse embryos at the time of their lineage divergence have provided key insights into intrinsic and extrinsic developmental regulators of the fate decision[70]. This study unveiled unique signalling signatures correlating with the pancreatic progenitor state, including non-canonical Wnts and Roundabout (Robo) signaling pathway[70]. Specifically, non-canonical WNT ligands and receptors have been found enriched in foregut and pancreatic progenitors, while absent in hepatoblasts. They have been proposed to control the acquisition of the pancreatic identity both *in vivo* and in *in vitro* liver-to-pancreas transition experiments[70,71]. Similar findings have been obtained in a study performed by coupling laser capture technology with deep sequencing analysis in human embryonic hepatic and pancreatic tissues[72]. In this work, gene ontology (GO) analyses highlighted enrichment for components of Notch, BMP, and Wnt signalling in human pancreatic progenitors, suggesting that the genetic programs directing lineage differentiation are conserved between mouse and human[70,72]. More recent evidence showed that the Slit/Robo guidance pathway is critical for establishing a pro-pancreatic *niche* to preserve pancreatic identity in cells of the ventral foregut, immediately after fate specification[73]. *Slit3* ligand is expressed in the mesenchyme surrounding the ventral pancreatic bud at E10.5, while its receptors, ROBO1/2, are present at the membranes of epithelial cells in the ventral pancreas. In the absence of the *Robo1/2* receptors hepatic features are induced in progenitor cells within the ventral pancreas[73]. In addition to extrinsic factors, it has been recently uncovered the role of the TF TG-Interacting Factor 2 (TGIF2), member of the Three-Amino-Acid Loop Extension (TALE) homeoprotein family, as a developmental regulator of the liver *versus* pancreas cell fate decision at early stage during mouse embryogenesis (Fig. 4C). *In vivo* studies demonstrated that its absence in the epiblast impairs the specification of the ventral pancreatic bud and, simultaneously, promotes an expansion of the liver bud volume[71].

Overall experimental perturbation of the identified signalling molecules leads to a shift in the balance between hepatic and pancreatic progenitor domains underlining the developmental plasticity of ventral foregut cells to contribute to either tissue[16].

Formation of the liver bud

The liver is specified as a single domain in the posterior ventral foregut[19,74]. Liver progenitor cells, also called hepatoblasts, initially acquire a columnar shape, proliferate and subsequently undergo a transition to form a pseudostratified epithelium[52,75]. For hepatic organogenesis to proceed, further delamination of the hepatoblasts into the septum transversum mesenchyme is required[76] (Fig. 5) This process is promoted by the TF Hex and its downstream effector, *Prox1*, which mediates the downregulation of *E-cadherin* to allow hepatoblasts migration[53]. In mice deficient for these factors, a liver bud fails to develop and liver progenitors are packed

close to the gut tube[75,77,78]. The earliest stages of liver organogenesis implicate the synergistic activity of several other TFs, including *One-cut-domain 1* (*Onecut1*)[79], *T-box transcription factor 3* (*Tbx3*), which stimulates organ expansion by extracellular matrix (ECM) remodelling[80], *Hepatocyte nuclear factor 4-alpha* (*Hnf4a*), that is essential for the hepatic lineage specification and differentiation, being later needed for bile acid biosynthesis and lipid homeostasis[81], and *Foxa2*, that opens the chromatin at the Albumin genetic site, enabling its transcription[74]. Moreover, shortly after specification, the hepatic bud is invaded by hematopoietic progenitors and endothelial cells and rapidly grows as a vascularized organ, with sinusoids being the first vessels to form by angiogenesis[82] (Fig.5).

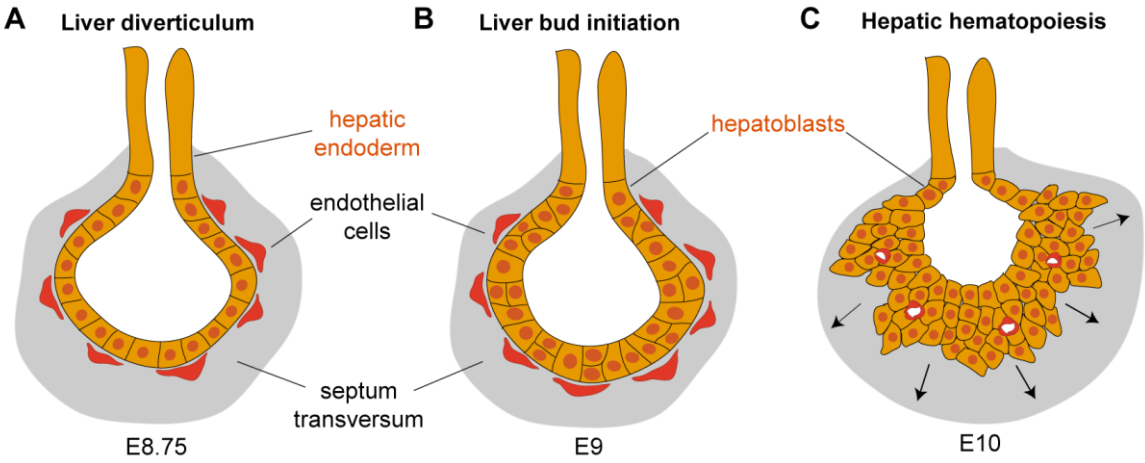

Figure 5. Early morphogenetic events during development of the hepatic domain
(A) At E8.75, endothelial cells (red) surround the thickened hepatic endoderm (orange), which initiates a budding process into the septum transversum (grey). **(B)** At E9, the hepatic endoderm transitions from a columnar to a pseudostratified epithelium. **(C)** At E10, hepatoblasts proliferate and migrate into the septum transversum to form the liver bud. Hematopoietic progenitor cells invade the hepatic tissue and establish liver foetal haematopoiesis. Adapted from Gordillo et al. 2015.

1.1.4 Morphogenesis and cell fate allocation during pancreatic organogenesis

The pancreatic gene expression program is established in the ventral and the dorsal foregut in response to external cues and through the action of internal factors, which are spatiotemporally controlled[1,41] (Fig. 4). After E11.5, when gut rotation occurs, the two buds are brought together and eventually fuse in one single organ that form the adult pancreas[36,83]. The dorsal pancreatic bud eventually contributes to the head, neck, body and tail regions of the adult pancreas, whereas the ventral bud generates the posterior part of the head region[84].

Morphogenesis and cell fate diversification in the pancreatic epithelium are tightly coordinated during embryonic development. Nevertheless, the precise dynamics of these early events are not completely understood.

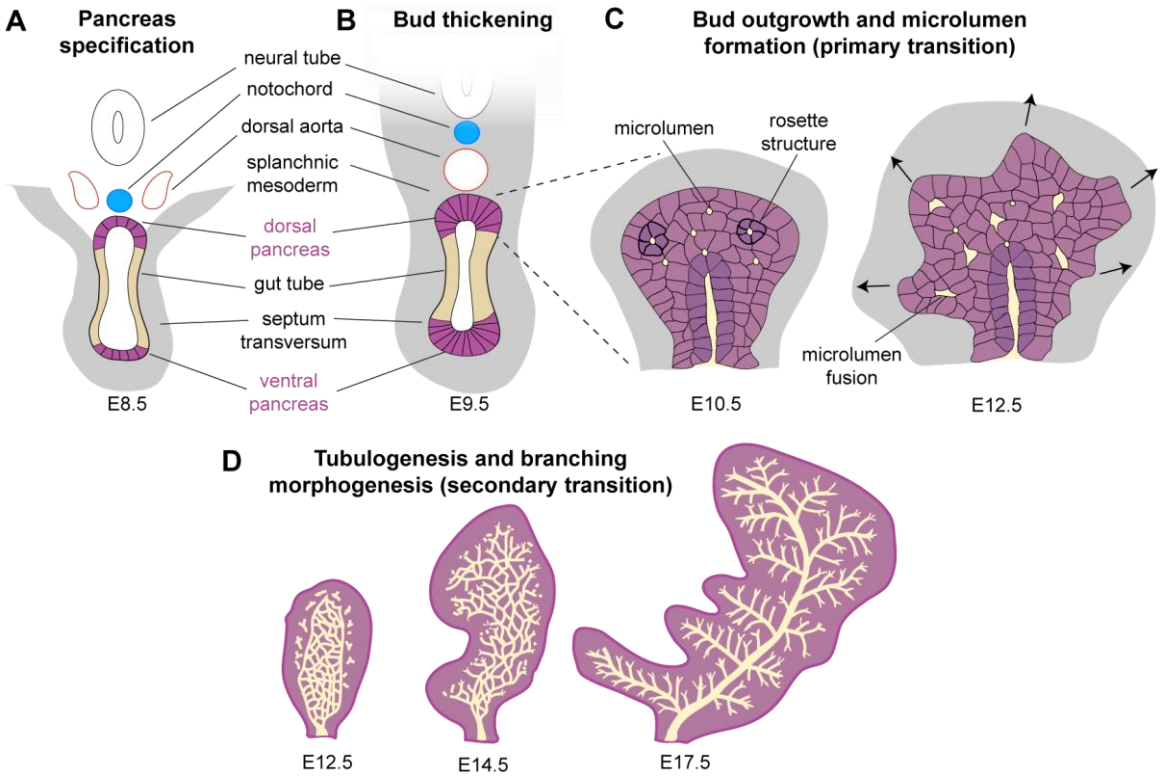

Figure 6. Schematic overview of the morphogenetic events of the nascent pancreas.
(A-B) The ventral and the dorsal pancreatic buds are specified in opposite regions of the gut tube at E8.5. The dorsal bud initially interacts with the notochord **(A)** and subsequently establishes contact with the dorsal aorta that become fused between the pancreatic endoderm and the notochord **(B).** At E9.5, cells of the pancreatic domain adopt a columnar shape, resulting in a thickening of the rudiment. Ventral and dorsal buds evaginate into the surrounding mesenchyme at E9.0 and E11.5 respectively**. (C)** At the onset of the primary transition, the pancreatic epithelium consists of a multi-layered core of cells, with the innermost cells exhibiting apical polarity and the outer cap cells having basal polarity. Rosette structures start to form within the pseudostratified epithelium, which will create central microlumens. At E12.5 the fusion of the microlumens results in the generation of the primary ducts. This early pancreatic epithelium is composed of proliferating multipotent pancreatic progenitor cells that initiate processes of branching and cell differentiation. **(D)** Starting from E12.5, the secondary transition coincides with the epithelium progressively remodelling to eventually form a well-ramified ductal system. Adapted from Bastidas-Ponce et al., 2017 and Villasenor et al., 2010.

In mice, two sequential waves frame the pancreatic organogenesis process, referred to as “primary transition” and “secondary transition”[85] (Fig. 6).

During the first wave, occurring between E9.5 and E12.5, the pancreatic epithelium comprises a pool of multipotent progenitor cells (MPC)[86–88] that actively proliferates and self-renew to form a pseudostratified epithelium. While the initial thickenings protrude out from the gut tube, epithelial cells acquire apico-basal cell polarity, undergo apical constriction and form rosette epithelial structures[89]. These structures develop microlumens at their centres, which subsequently coalesce to form the ductal system[90,91] (Fig. 6). The founder MPC are critical for determining the final cell number of the pancreatic populations as well as organ size[88] and give rise to all the pancreatic lineages (acinar, endocrine and ductal), as demonstrated by lineage tracing experiments[92,93].

Starting from E12.5, when the “secondary transition” takes place, the progenitor epithelium undergoes dramatic morphogenetic changes involving massive epithelial remodelling, tubulogenesis, branching morphogenesis and growth[42,90,94,95], to finally generate a ramifying tree-like glandular structure. During this phase all the different specialized cell types are formed[89] (Fig. 6).

This stage is marked by the segregation of the epithelial “tip” and “trunk” compartments[1] (Fig. 7). The tip domains contain MPC co-expressing Pdx1, Pancreas specific transcription factor 1a (Ptf1a), Myelocytomatosis oncogene (c-Myc), Gata4 and Carboxypetidase A1 (Cpa1), whose fate becomes restricted to the exocrine identity after E14.5[1,87,96]. By contrast, cells located in the trunk are characterized by the expression of NK6 homeobox 1 (Nkx6-1), Nkx6-2, SRY-box 9 (Sox9), HNF1 homeobox B (Hnf1b), Onecut1, Prox1 and Hairy and enhancer of split (Hes1) and constitute a bipotent progenitor pool, being able to give rise to both the ductal and the endocrine cell lineages[2,93,97–99].

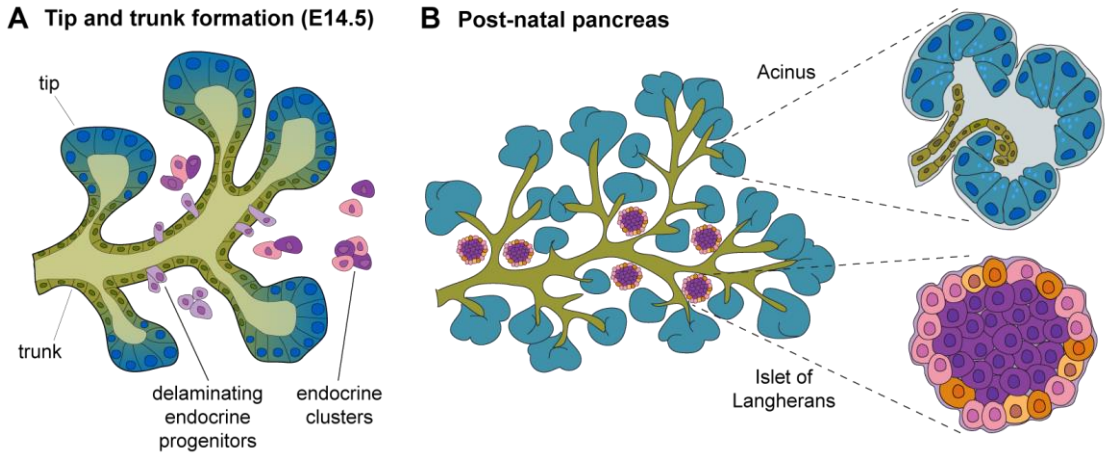

Figure 7. Branching morphogenesis and lineage diversification in the pancreas.
Illustrated overview of the developing and post-natal pancreas. **(A)** At E14.5 the pancreatic luminal epithelium, formed by fusion of the microlumens, enlarges and progressively remodels into a single cell-layered epithelium. Branching morphogenesis leads to the segregation into multipotent tip and bipotent trunk domains. A subset of trunk cells transiently activates the expression of Ngn3 and delaminates from the epithelium, being committed to become endocrine progenitor cells (lilac). **(B)** The mature pancreas comprises a complex ductal system with terminal acini and endocrine islets interspersed in the embedding mesenchyme. The insets on the right show the functional exocrine secretory unit (acinar compartment) and the Islet of Langherans (endocrine compartment), that in the mouse is composed of a core of β-cells (purple) and an outer layer constituted by other hormone-releasing cell types (α, δ, PP, ε).

Acinar lineage differentiation

The commitment of pancreatic progenitors to become acinar cells primarily requires the mutual repression between 2 TFs, *Nkx6-1/Nkx6-2* and *Ptf1a*, which controls how the pancreatic cells segregate to tip and trunk domains and drives the acinar versus ductal cell

fate choice (Fig. 8). Specifically, NKX6-1 induces trunk formation, while PTF1A, which is initially expressed broadly in the epithelium, increases in the tip cells and simultaneously opposes to *Nkx6.1*[100]. PTF1a is the major transcriptional determinant of acinar fate: as multipotent cells transition to acinar-lineage-restricted fates, PTF1a forms different protein-protein interactions to promote the activation of an acinar gene regulatory network[101]. For instance, the activity of the trimeric PTF1A-L complex, enhanced by nuclear receptor subfamily 5, group A, member 2 (NR5A2)[102], is critical for the maturation of the acinar cells, by directly activating genes encoding hydrolytic enzymes, mitochondrial components, and exocytosis machinery, all hallmark components of functionally mature acini[101,103,104]. Acinar maturation is achieved also by the activity of *Mist1*, a basic helix-loop-helix (bHLH) TF, which regulates acinar cell polarity and exocytosis[105,106]. Differentiating cells initially express digestive zymogens, such as *Cpa1* and *Elastase*, and *Amylase* later in development. By E15.5 most acinar cells have differentiated, and their expansion largely contributes to the increase of the organ size[42]. The adult acinar cells are secretory polarized cells with a pyramidal shape (Fig. 7B) and express the TFs *Ptf1a, Rbpjl, Gata4*, *bHLH family member a15 (Bhlha15, aka Mist1)* and *Nr5a2*[107,108].

Multiple extrinsic signals and cell-cell interactions have been shown to contribute to the establishment of the acinar fate[97,107,109,110]. Among these, the mesenchyme influences acinar development by secreting the Activin repressor follistatin, which inhibits the endocrine lineage program[111,112], and FGF10, which is critical for modulating pancreatic epithelial proliferation and differentiation as well as *Ptf1a* expression[51]. Inhibitions of the FGF pathway impairs branching morphogenesis, while its over-activation leads to pancreatic hyperplasia and abrogates cell differentiation[113–116]. Specifically, FGF10 operates through Notch signalling, whose effector HES1 directly interacts with PTF1A antagonizing its activity[66,115–117].

Endocrine lineage differentiation

Few endocrine cells are formed during the primary transition, including mostly glucagon-producing cell clustering in the dorsal pancreatic epithelium, but their contribution to the final endocrine compartment is controversial[85,118,119]. The major production of endocrine cells takes place during the secondary transition. At E12.5, the trunk region of the pancreatic epithelium harbours progenitors for both ductal and endocrine lineages. *Sox9*$^{+}$ cells that activate the expression of *Neurogenin-3* (*Ngn3*) are fated to become endocrine cells, exit cell cycle and go through an epithelial-to-mesenchymal-transition (EMT), underlying the delamination process[92,120,121]. Precursor endocrine cells individually leave the epithelium and form clusters in the surrounding tissue[122] (Fig. 7A). The levels of *Ngn3* are critical for triggering endocrine commitment, whereby bipotent progenitors that fail to reach a specific threshold of *Ngn3*

expression default to a ductal or acinar fate[92,123]. Of note, the activation of an alternative pancreatic fate in failed endocrine progenitors is stage dependent, with induction of acinar fate being possible only before E12.5[124]. This finding further supports the notion that during the secondary transition endocrine progenitors arise from bipotent "trunk" progenitors, whereas early endocrine progenitors arise from multipotent progenitors[87].

A recent study pointed out at the relevance of cellular position and mechano-signalling cues in ductal versus endocrine cell fate choice[125]. The interesting concept emerging from this study is that endocrinogenesis is influenced by the surrounding ECM proteins and cell shape. Specifically, bi-potent and ductal progenitor cells express *Itga5* and are exposed to fibronectin, which promotes cellular spreading and active Yes-associated protein (YAP) signalling, that in turn represses *Ngn3* expression. By contrast, endocrine precursors mainly contact laminin and have a more confined shape, with consequent reduced levels of nuclear YAP, resulting in the initiation of the endocrine program[125].

Unipotent *Ngn3*$^{+}$ cells generate all the different hormone-releasing endocrine cells[126] in a timely controlled fashion, with glucagon-producing α-cells being specified first, followed by insulin-producing β-cells, δ-cells and PP-cells[42]. At early stages bi-hormonal cells positive for glucagon and insulin are detected, but these are not mature endocrine cells[42].

A series of NGN3 downstream target genes[1,83,85], individually described in Introduction section 1.1.6, guide the process of differentiation and maturation of hormone-producing cells (Fig. 8).

After birth, the endocrine clusters coalesce and round up to form the mature Langherans islets, that in the mouse are constituted by a mantle of α, δ, ε, and PP-cells surrounding a large mass of β-cells, which undergo post-natal functional maturation (Fig. 7B)[97,109]. Islet architecture and function are also modulated by the interaction with endothelial cells and neurons, which invade the pancreatic tissue early during development[2].

Ductal lineage differentiation

Cells of the trunk epithelium that are not committed to become endocrine cells are directed to the ductal fate[2,93,97]. The molecular mechanisms involved in ductal specification and development are poorly understood[97]. During embryonic development, ductal cells express *Pdx1*, *Nkx6-1*, *Sox9, Hes1, Hnf1b, Hnf6 (*aka *Onecut1)*, *GLIS family zinc finger 3 (Glis3)* and *Prox1*[97]. In particular, HNF6 regulates maturation of the ductal epithelium and primary cilia formation, which are important for proper ductal morphogenesis[127]. Al later stages, they lose the expression of *Pdx1* and *Nkx6-1,* while maintaining the expression of all the other TFs[97].

1.1.5 Regulators of pancreatic development

Mesenchymal cues influencing pancreatic development

Several evidence, including seminal *in vitro* tissue recombination experiments[128], suggests that the crosstalk between the pancreatic epithelium and surrounding mesenchyme governs the whole process of pancreatic organogenesis[129]. Indeed, the mesenchyme secrets soluble factors and establish cell-cell and cell-matrix interactions, supporting the initial steps of pancreatic specification, branching morphogenesis and differentiation of the pancreatic lineages, eventually regulating the relative proportions between the different pancreatic lineages[111,128–131].

Multiple signals have been reported to regulate mesenchymal-epithelial interactions [83,129,132]. For instance, FGF signalling has been shown to maintain cell fate commitment and proliferation of the pancreatic epithelium at the expense of cellular differentiation both *in vitro* and *in vivo*[111,113,133,134]. Among the FGF family members, FGF1, FGF7 and FGF10 ligands are secreted by the pancreatic mesenchyme while their receptor FGFR2B is expressed by epithelial cells, a pattern that is conserved in human embryos[112]. It has been proposed that FGF10 and FGFR2B establish a feed-forward loop with the downstream pancreatic TF SOX9, which is responsible for the maintenance of the pancreatic identity[134]. Disruption of this molecular circuitry leads to pancreas-to-liver conversion in *Fgf10*- and *Fgfr2b*-null mutant embryos[134]. Moreover, FGF10 activates epithelial Notch signalling to maintain a pool of undifferentiated progenitor cells supporting epithelial growth[114–116,135,136]. Consistently, long-term culture of human pancreatic progenitors can be achieved by stimulation of cell proliferation via exposure to FGF10, Wnt and epidermal growth factor (EGF) signaling[137]. During pancreas development, FGF10 positively regulates *Ptf1a* expression in the tip cells[51].

Several components of the TGF-β signalling family are also implied in modulating pancreatic growth and differentiation. The mammalian TGF-β isoforms (TGF-β1, -β2 and -β3) are initially localized broadly in the E12.5 pancreatic epithelium and, subsequently, become restricted to the acinar compartment[138,139]. The TGFBR2 receptor expression is detected in both epithelium and mesenchyme at early stages of pancreas development and is limited to the islets and ducts by late gestation[140]. Analysis of genetic mouse models suggested a role for TGF-β signalling in controlling the ability of endocrine progenitors to migrate and coalesce in the mesenchyme[141] and to organize into mature clusters[142]. Inhibition of the TGF-β pathway results in enhanced ductal proliferation and accumulation of periductal endocrine cells, suggesting its role in ductal/endocrine lineages restriction[140,143].

Activin signalling is important for early pancreatic morphogenesis, promoting endocrine differentiation and limiting the exocrine one. Loss-of-function of *Activin receptor type IIB* (*Acvr2b*) and *type IIA (Acvr2a*) results in pancreatic hypoplasia with reduced number and size

of endocrine islets[144]. Pancreatic explants exposed to follistatin, a known inhibitor of Activin present in early pancreatic mesenchyme, undergo enhanced exocrine differentiation at the expenses of endocrine cells growth[111]. Consistently, *ex vivo* Activin treatment of pancreatic explants inhibits branching morphogenesis and increases insulin-producing cells number[145,146].

BMP molecules are expressed in the developing pancreas, however the role of BMP signalling in the morphogenesis of the organ is not clear yet, with contrasting results being reported[147–149]. For example, BMP4/7, derived from the mesenchyme, has been reported to signal through the BMP receptor, type 1A (BMPR1a, aka ALK3) receptor to regulate expansion and branching of the developing pancreas[147]. On the other hand, pancreatic Bmp4-deficient mice do not display pancreatic defects[148], whereas the deletion of Bmp6 in the pancreatic domain results in complete pancreatic agenesis[149].

1.1.6 Transcription factors involved in pancreatic development

A complex transcriptional cascade orchestrates the multistep differentiation process that leads progenitor cells to acquire a fully differentiated and functional pancreatic state[1,2] (Fig. 8).

In first place, organ territories in the gastrointestinal tract are defined by factors that act as key master developmental regulators. These have been identified by loss-of-function or gene-dosage studies (*Sox2, Sox17, Ptf1a, Cdx2, Pdx1*), whereby genetic dysregulation impairs the maintenance of inter-organ boundaries and triggers switches in cell fates[150–155].

Subsequently, the pancreatic cell lineages are determined by the action of TFs which establish specific gene regulatory networks. Of note, mutual inhibitory interactions between TFs are mechanisms commonly used to impart and reinforce cell identity, for example for both tip and trunk cell allocation (as seen between *Nkx6-1* and *Ptf1a*) and the α- versus β-cell lineage decision. Many of these genes influence the onset and progression of pancreatogenesis in multiple ways, with the same TF exerting different functions depending on the cellular context[1,42,83] (Fig. 8). Their roles have been identified mainly by genetic studies in vertebrate models. The section below reports various examples.

Pancreatic specification: Pdx1, Sox9, Ptf1a

Pdx1

Pdx1, also known as insulin promoter factor 1 (Ipf1), is a TF belonging to the ParaHox family. Its expression starts as early as E8.5 of mouse development, initially in the ventral pancreatic domain and 12 h later also in the dorsal pancreatic bud. From E9.5, its expression expands to the proximal regions of stomach, duodenum and gallbladder. At later stages of development, *Pdx1* is maintained only in the pancreatic β-cells[42], where it is needed for

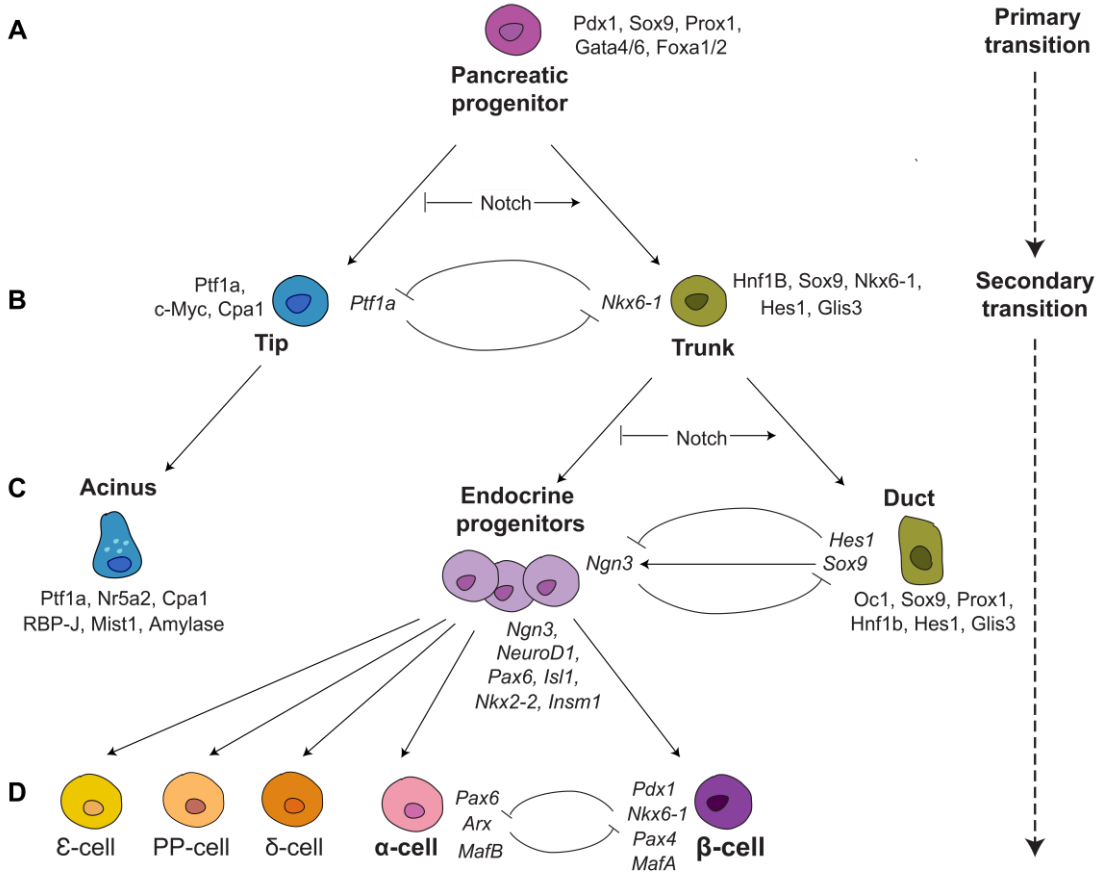

Figure 8. Cascade of TFs regulating lineage decisions during pancreas development.
Schematic overview of the pancreatic cell differentiation along the acinar, endocrine and ductal lineages. **(A)** Pancreatic expansion and cell identity in MPC is maintained by a combination of TFs, including Pdx1, Ptf1a and Sox9. This first phase of pancreatic organogenesis is named "primary transition" **(B)** At the onset of the secondary transition, Notch activity and a cross-repression between Ptf1a and Nkx6.1 mediate tip/trunk separation. **(C)** Tip cells mature into acinar cells through the action of the PTF1-L complex, which establishes a positive-regulatory loop with Nr5a2. Trunk cells commit to either an endocrine precursor state, marked by the expression of Ngn3, or a ductal identity. This segregation depends on the Notch signaling and transcriptional interactions between Ngn3, Hes1 and Sox9. **(D)** Endocrine progenitors further differentiate into the five different hormone-producing cells. Repressive interaction between the α-cell specific TF Arx and the β-cell specific TFs Pax4, Nkx6.1 and Pdx1 is required to determine and maintain an α- or β-cell identity. Adapted from Shih et al., 2013.

sustained insulin gene expression[156].

Full *Pdx1* knock-out (KO) results in severe pancreatic hypoplasia with production of few insulin- and glucagon-positive cells[157]. Mice conditionally deprived of *Pdx1* in β-cells develop diabetes, uncovering its requirement for the maintenance of β-cells functions and identity[157–159]. Cases of homozygous and heterozygous mutations of human *PDX1* gene have been reported to be characterized by pancreatic agenesis and MODY, respectively. This suggests that this TF has a decisive role in pancreas formation and function also in the human[160,161].

Mechanistically, it is still not clear how *Pdx1* is initially activated in the foregut endoderm[36]. However, much more is known about its establishment and subsequent maintenance in the pancreatic cell types. Four conserved *cis*-regulatory regions (termed Areas I-IV), located upstream *Pdx1* transcription start site (TSS) have been described, which are bound by a set of endodermal TF regulating *Pdx1* at different stages of pancreas development, from growth and differentiation to post-natal maturation of β-cells[162–166]. For instance, FOXA1 and FOXA2 bind to a distal enhancer in the *Pdx1* promoter to regulate its early gene activation[167], and, similarly, PTF1a trans-activates another enhancer in the *Pdx1* promoter, contributing to the pancreas-wide expression of *Pdx1*[168]. Area II contributes to all endocrine-specific functions of *Pdx1*, including endocrine progenitor specification, β-cell versus α-cell lineage allocation and post-natal β-cell maturation, upon binding by v-Maf musculoaponeurotic Fibrosarcoma oncogene family proteins A and B (MAFA/MAFB) and other TFs[162].

Sox9

Sox9 encodes for a TF belonging to the SOX (sex-determining region on Y box) family. It starts to be expressed in the pancreatic epithelium shortly after *Pdx1* and its expression is restricted to the ductal cells after E13.5. SOX9 is a regulator of pancreatic cell identity, contributing to the maintenance of multipotency, proliferation and survival of the pancreatic progenitors, by positively regulating the Notch effector Hes1[169]. In fact, conditional ablation of *Sox9* in the whole pancreatic epithelium using the transgenic *Pdx1*-Cre mouse transgenic strain results in pancreatic aplasia[169], while *Sox9*-null embryos have a reduction of *Pdx1* expression levels and display pancreas-to-liver identity conversion[134]. These studies underscored the requirement of *Sox9* in establishing pancreatic cell identity through a feed-forward loop with *Fgf10* and its receptor *Fgfr2*[134]. SOX9 is also considered to cooperate together with PDX1 at early stages for the establishment of a boundary with the prospective intestinal lineage, by activating pancreatic genes and repressing the intestinal ones[154]. Moreover, SOX9 is essential for regulating the endocrine *versus* ductal cell fate decision[134,169–171], being required for cell-autonomous induction of the pro-endocrine gene *Ngn3* in a dose-dependent manner[170]. Once *Ngn3* expression has been initiated, endocrine differentiation can proceed only upon subsequent *Sox9* downregulation[171]. In fact, *Ngn3* establishes a negative feedback loop to specify endocrine cells at the expenses of ductal cells, if Notch levels are high and *Sox9* expression is maintained, endocrine precursor cells are directed to a ductal fate. Finally, *Sox9* is necessary for *Ngn3*-negative ductal differentiation and maintenance. Indeed, *Sox9* mutant pancreata display disorganized polycystic ducts lacking primary cilia[171] .

Ptf1a

The bHLH factor PTF1A (aka PTF1-p48) exerts its functions in a trimeric complex together with a canonical E-protein partner (E12, E47, or HEB) and the recombination signal binding protein RBPJ[101]. Its expression starts shortly after *Pdx1* and the two TFs initially overlap in both pancreatic buds. After E10.5, PTF1A is present only in the pancreatic epithelium and becomes necessary to form a boundary with the duodenum[42]. From E12.5, *Ptf1a* expression segregates to the tips of the branching epithelium and, eventually, to the acinar cells[101,150,172].

Several evidences proved that PTF1A controls pancreatic cell identity and forms a boundary with developmentally closely related organs. Specifically, loss-of -function of *Ptf1a* gene causes pancreatic agenesis and cell conversion toward an intestinal fate[150,172]. Consistently, transient misexpression of *Ptf1a* in the endoderm causes conversion of stomach, extra-hepatic biliary system and rostral duodenum into pancreatic tissue[155]. PTF1A is not only important for preserving the pancreatic identity at early stages, but is also fundamental for establishing the acinar cell fate[101]. Pancreata from *Ptf1a*-deficient embryos have few endocrine cells and completely lack exocrine tissue[172]. When the epithelium starts to branch during the secondary transition, cross-repressive interactions are established between *Ptf1a* and *Nkx6-1/6-2* determining the endocrine/exocrine lineage segregation[100]. At the molecular level, this double role of PTF1A is due to the dynamic binding of PTF1 with interacting factors. At early stages, pancreatic multipotency is regulated by the PTF1-J complex, where PTF1 is associated with the nuclear mediator of Notch signalling RBP-J kappa (RBPJk). As multipotent cells transition to acinar lineage fates at the tips, PTF1-J activates the expression of RBPJL, that, in turn, displaces RBPJk. The newly formed protein complex, named PTF1-L, is responsible for activating genes essential for acinar cell identity, such as *Cpa1* and other digestive enzymes [101].

Endocrine differentiation: Ngn3

NGN3 is a bHLH TF with a biphasic expression dynamic in the developing mouse pancreas[173]: it is first expressed between E8.5 and E11 and then, after E12, in a second temporal wave[173]. From the moment of the tip and trunk segregation, *Ngn3* expression is observed in cells scattered throughout the primitive ducts. Mouse genetics and lineage tracing experiments provided evidence that *Ngn3* is required for endocrine differentiation. All the endocrine cell types are descendent of the $NGN3^+$ cells[92] and *Ngn3*-null mice entirely lack the endocrine compartment[120]. Moreover, transient ectopic expression of NGN3 in the pancreas is sufficient to generate α, β, δ and PP cells[92]. $Ngn3^+$ cells are also found in the enteroendocrine cells of the gastric epithelium and during neurogenesis in the neural tube[174–176]. In the

pancreas, *Ngn3* expression is restricted to the embryonic life and is not found in adult islet cells[120].

Ngn3 activation is controlled by a mechanism of lateral inhibition mediated by the Notch signalling[135]. Specifically, the Notch downstream target and effector HES1 has been shown to repress *Ngn3* transcription and accelerate NGN3 protein degradation[136,177], but how different ductal cells acquire different Notch levels is still not clear. More recently, it has been proposed that NGN3 protein stability is regulated by the phosphorylation status of NGN3, which is cell-cycle dependent[178,179]. *Ngn3* promotes cell cycle arrest and indirectly controls post-transcriptionally SNAIL2 protein, that in turn suppresses the epithelial marker *E-cadherin* and triggers cell delamination[122,180]. Cells committed to the endocrine fate increase the expression of *Ngn3*, undergo EMT and exit the epithelium to intercalate in the mesenchyme, where they progress toward a fully differentiate endocrine state. These studies suggested that *Ngn3* controls both epithelium delamination and endocrine cell differentiation[122]. Moreover, recent studies highlighted the importance of the level of expression of NGN3: low NGN3 protein level ($NGN3^{LO}$) would be associated with maintenance of a mitotic status of endocrine precursors in the epithelium, while the $NGN3^{HI}$ state dictates the exit from the epithelium and promotes the transition to endocrine lineage commitment[181].

Ngn3 is the master regulator of the endocrine lineage, by directly activating downstream endocrine lineage-specific factors[120,182]. Among these are *Isl1*[183,184], *Insulinoma-associated 1* (Insm1)[185,186], *Neurogenic differentiation 1* (*NeuroD1*)[187,188], *NK2 homeobox 2* (Nkx2-2), *Paired box 4 (Pax4*), *Pax6*, *Regulatory factor X6* (*Rfx6*)[189] and *Myelin transcription factor 1* (*Myt1*)[190]. Most of them start to be expressed in the mouse pancreatic endoderm at E9.0, in particular in glucagon-expressing cells, and become endocrine progenitor-specific later in development, often establishing feed-forward loops with *Ngn3*[190].

β-cell differentiation: *Nkx6-1 and Nkx2-2*

The NK-homeodomain genes *Nkx6-1* and *Nkx2-2* are present in pancreatic progenitors and during the major wave of endocrine differentiation their expression pattern becomes confined to the trunk cells[42]. While *Nkx2-2* transcription is maintained in α-, β- and PP-cells of the mature islets, *Nkx6-1* in uniquely present in the β-cells, ensuring their functional state[42,191–193]. The two factors are not required for early stages of pancreatic development, but β-cell differentiation is impaired when one of the two is fully ablated, indicating a major role in the acquisition of the beta-cell identity[191,192,194]. Specifically, *Nkx6-1* full KO mice display loss of β-cell precursors, whereas in *Nkx2-2* mutants β-cells survive but fail to become functional[191,192]. Furthermore, *Nkx6-1* expression is lost in endocrine cells of *Nkx2-2*-null embryos, suggesting

that *Nkx6-1* lies downstream of *Nkx2-2* in the transcriptional cascade driving β-cell differentiation[191,192].

Mechanistically, it has been shown that both these factors preserve β-cell fate at expenses of the α-cell fate, by repressing the α-cell-specific TF *Aristaless related homeobox (Arx)* in β-cell precursors, directly (NKX6-1)[195] or forming repression complexes with other TFs (NKX2-2)[196].

Regulation of α- versus β-cell fate decision: Pax4, Pax6, Arx, Mafa, Mafb

Pax4, Pax6

The expression of *Pax6* and *Pax4* starts at early stages in the developing pancreas, respectively at E9.0 and E9.5, and after the second transition they both become restricted to the endocrine lineage[197]. *Pax4* expression is confined to the β-cells perinatally[198], before completely disappearing in the adult pancreas[199]. Instead, Pax6 is still present in adult cells secreting insulin, glucagon, somatostatin and PP[197].

Consistently with their expression pattern, *Pax4* and *Pax6*, are required for the proper development of endocrine cells; in fact, *Pax4/Pax6* double mutants entirely lack endocrine hormone-producing cells[200]. α-cells are abolished in *Pax6* homozygous mutant mice, while a hypomorphic mutation in *Pax6* impairs proper islet organization and differentiation to all hormone-expressing cells[197,200]. *Pax6* ablation after birth results in reduction of all hormones except for ghrelin, which is strongly upregulated[201]. Transcriptional profiling and lineage tracing experiments in adult *Pax6*-deficient β-cells showed that PAX6 works as a direct transcriptional activator of β-cell specific genes and a repressor or alternative islet cell genes[202]. Collectively, all these studies pointed at PAX6 as a regulator of both cell identity acquisition and maintenance within the endocrine lineage, being important also in the maintenance of adult pancreatic islet function and β-cell identity. Moreover, there is evidence that its role is conserved in the human[202]. Mice lacking *Pax4* exhibit loss of β- and δ-cells, and concomitant increase of α-cells, indicating that PAX4 has a role in the cell fate decision of β- *versus* α-cells [198]. Of note, ectopic expression of PAX4 in pancreatic or endocrine progenitors causes α-to-β cell conversion, resulting in the generation of islet mainly composed of insulin-producing cells [203].

Arx

ARX is another TF required for cell fate establishment of endocrine cell types. It is expressed in the pancreatic bud at E 9.5, and later on in endocrine progenitors and, ultimately, in α- and PP-cells[204]. *Arx*-deficient mice show increased number of β- and δ-cells and

reduction of α-cells, indicating that ARX exerts an opposite role compared to PAX4 [204]. Accordingly, *Arx* and *Pax4* have been shown to repress each other[205] and the addition of *Pax4* or repression of *Arx* are equivalent mechanisms to induce an α-to-β cell conversion *in vivo*[206,207]. Remarkably, the establishment of β-cell identity also relies on mutual repression between ARX and NKX6-1[195] and a failure of the derepression of *Arx* leads to β-to-α-cell conversion[196].

Mafa, Mafb

The basic leucine-zipper transcription factors *Mafa* and *Mafb* display a differential expression between α- and β-cells[208] and both participate to their maturation process[209–211]. *Mafb* is first detected from E10.5 in early immature hormonal cells, subsequently in the glucagon and insulin secreting cells and, in the adult, becomes specific to the α-cells[212], where it activates the glucagon gene[208,212]. *Mafb* mutant embryos display defective beta-cell maturation and reduction of *Mafa* transcript, which is required for beta-cell identity, implying a regulative action of *Mafb* on *Mafa* gene[209–211]. During pancreatogenesis, *Mafa* is expressed uniquely in the insulin$^+$ cells arising during the secondary transition. The switch from *Mafb* to *Mafa* is necessary for proper differentiation of insulin-secreting cells. In fact, MAFA positively regulates insulin expression in β-cells[208,213]. Consistently*, in vivo* studies showed that *Mafa* deficiency causes dysfunctional glucose-stimulated insulin secretion and abnormal islets morphogenesis and eventually leads to diabetes in adult mice[209]. Of note, MAFA was shown to reinforce α-to-β cell conversion *in vivo* and is one of the major pancreatic TFs that is used to establish a β-cell identity in different cell types[16,214–216].

In summary, the majority of the TFs described above were shown to exert diverse roles during pancreatic development and in the adult organ, implying multi-step transcriptional states from specification to maturation of the different pancreatic lineages.

To date, a gap in knowledge exists about the precise order of transcriptional events that controls the acquisition of the pancreatic and endocrine cell fate, especially as regards the integration of the intrinsic factors with the external inputs, including morphogens and cell-matrix interactions.

1.1.7 Human pancreas development

Our understanding of the pancreas ontogeny and transcriptional regulation comes from extensive studies in multiple vertebrate animal models, such as mouse, chicken and frog [1,36,42,83,217], which highlight general inter-species conservation of morphogenesis and

differentiation processes. The study of the human pancreas organogenesis is limited by ethical restrictions and access to embryonic and foetal tissue. Only recently, molecular and morphogenetic features defining the early stages of the human pancreas development, started to be unravelled[72,218–220]. In particular, recent transcriptome analyses revealed that many gene expression signatures are common between human and rodent pancreatic progenitors[72,220].

Similarly to other vertebrates, also the human pancreas develops from two distinct protrusions of the primitive gut epithelium, visible starting 4 weeks post-conception (wpc), upon SHH-driven patterning from the lining notochord[218]. Pancreatic identity in the human is established upon expression of PDX1, SOX9 and GATA4[221,222]. PTF1A, which is critical of the establishment mouse pancreatic identity, has also been found in embryonic human dorsal pancreatic bud[72]. Consistently, truncation of the C-terminal region of PTF1a affecting one of its enhancers was identified in cases of pancreatic agenesis[223,224].

The first distinction between central trunk cells and acinar tip cells occurs at 7 wpc, according to different expression levels of the trunk-enriched factors *SOX9/NKX6-1* and acinar-restricted factor GATA4. Complete segregation of the acinar lineage from the duct one is delayed compared to the mouse, with exclusion of *SOX9* from tip cells happening between 10 and 14 wpc[100,222].

Unlike the mouse, NKX2-2 protein expression is not observed in the human pancreatic progenitor cell population prior to endocrine commitment[222], which starts at 7 wpc. Moreover, in contrast to other models, an early endocrine differentiation wave, corresponding to the “primary transition”, does not occur in the human. This can probably be due to the lack of pro-endocrine signals deriving from the paired dorsal aortae to the pancreatic endoderm, as seen in the early mouse and chicken embryos[49,225]. NEUROG3 expression is detected transiently as a single phase from 8 weeks post conception (wpc) and is absent after 35 wpc[222,226]. The first foetal β-cells emerge at ~8 wpc, followed by the formation of glucagon-expressing cells at 9 wpc[222,227]. Remarkably, human endocrine differentiation is completely dependent on NEUROG3[228], which implies that β-cell allocation in utero is completed at least 5 weeks prior to term and that proliferation of already specified foetal β-cell is the main mechanism to control the beta-cell mass[218]. Of note, the architecture of human islets changes as development progresses: at 14 wpc, β-cells are mainly present in the core and α-cells at the periphery, as reported in the mouse, but by 21 wpc both cell types are intermingled within the islets [3,227]. This different configuration of endocrine cells might be required for proper mature functionality. Moreover, in stark contrast with the rodent model, MAFB seems to have a major role in the post-natal maturation of human β-cells, that instead express low levels of MAFA[229].

Recent advances in methods based on transcriptomic analysis such as single cell RNA-sequencing (scRNA-seq) revealed a high degree of heterogeneity in the human pancreatic islet cells[220,230,231]. Remarkably, different antigenic subtypes of β cell, which have distinct gene

expression profiles, have been identified. All these notions have important medical implications, since foetal maturation and distribution of endocrine cells are profoundly altered in T2D[232].

In vitro systems to study human pancreatic development and disorders

Genome-wide association (GWA) studies have identified a significant amount of sequence variances across the human genome that associate with T2D[219,233]. Most of the reported single nucleotide polymorphisms (SNPs) fall close to genetic loci that are involved in events during pancreas development, such as progenitor cell proliferation or β-cell differentiation, including genes with key roles in pancreas development, such as *PDX1, HNF1B, HNF4A, GLIS3, HHEX, NEUROD1, NOTCH2* and *PROX1*[218,233]. Some of these genes and others have also been found mutated in cases of MODY[234,235], whereby organ defects resembles those observed in the respective null mice[11]. Also, it has been hypothesized that suboptimal development of β-cell mass in foetal life correlates with T2D risk[236,237]. All these studies support the notion that a certain degree of conservation exists between rodent and human species and prompt to further understand how much of the concepts gained using mouse genetics and preclinical models can be translated into possible therapeutic approaches[9].

In vitro systems based on differentiation of pluripotent stem cells (PSCs) provide a valuable platform to model pancreatic organogenesis and diseases[9,13]. Due to general conservation of gene functions and developmental processes across species, protocols to differentiate human Embryonic Stem Cell (hESC) or induced Pluripotent Stem Cells (iPSC) into the β-cell lineage were initially based on the knowledge of mouse pancreas development, by mimicking those signals that drive pancreatic development *in* vivo in the mouse[31]. Over the years, these differentiation protocols have been further refined to generate more functional mono-hormonal endocrine cells[238–240]. PSC-based *in vitro* systems, coupled with gene editing tools, are considerably important to gain insight into the molecular mechanisms regulating human pancreas development, to confirm similarities or discover differences with the other vertebrate models. For example, efforts in Huangfu lab aimed at elucidating the function of various TFs by using an inducible Cas9-hESC line to create KO for *PDX1, RFX6, PTF1A, GATA6, GLIS3,* Motor neuron and pancreas homeobox 1 (*MNX1), NEUROG3, HES1, ARX* and *FOXA2*, showing how these genes are essential at different stages of human pancreatic development[241,242].

In addition, these cellular products have an extremely important potential for drug screening and regenerative medicine purposes, as will be discusses in Introduction section 1.3.

1.2 TALE homeodomain proteins

The Three-Amino-Acid Loop Extension (TALE) superfamily comprises transcription factors with three-amino-acid loop extension between helices 1 and 2 of their homeodomain[243]. The genes encoding for these transcriptional regulators are highly conserved, being present in the common ancestor of plants, fungi, and animals[244].

These homeodomain-proteins are further categorized into 5 subclasses: the PBC family, consisting of the Drosophila Extradenticle (Exd) and the vertebrate Pre-B Cell Leukemia Homeobox (PBX) proteins; the MEIS subclass, including Drosophila Homothorax (Hth) and the vertebrate Myeloid Ecotropic Viral Integration Site (MEIS) and Pbx regulating protein (PREP); the TG-Interacting Factor (TGIF) subclass; the Iroquois (IRX) proteins; the Mohawk protein family[245].

The TALE proteins are involved in many developmental processes and diseases[246–248]. They have been described as partners of the Homeobox (HOX) proteins, which are evolutionarily conserved TFs with fundamental roles in determining cell identity during embryonic development among other processes[249–251]. In vertebrates, *Hox* genes encode homeodomain-containing DNA-binding proteins that control the A/P body axis patterning during embryogenesis[252,253]. *Hox* genes are organized in clusters that are characterized by "collinearity": their location along the chromosome corresponds to their expression along the body axis[254]. The integration of the positional signalling and the identity information is determined by specific interaction of the HOX protein with the co-factors TALE homeodomains, which confer DNA-binding site selectivity upon complex formation[255–258]. TALE homeoproteins can also function through HOX-independent mechanisms[246,259] and interact with each other, forming homo and heterodimers[246,260]. For instance, PBX associates with MEIS/PREP through conserved motifs situated at the N-terminus of their respective homeodomains[261,262] and form protein aggregates that regulate the subcellular localization and stability of PBX protein. The same multimeric complex also participates in the binding of Hox complexes to the DNA[260,262–264].

The TALE homeodomain-proteins can act both as transcriptional activators or repressors and in a synergistic as well as antagonistic fashion, complicating the prediction of the functional outcome of the binding pattern[246,247,265]. PBX, MEIS and HOX cooperatively exert their function by associating in heterotrimeric complexes[247,266,267]. On the other hand, TALE proteins possess overlapping consensus binding and, therefore, compete for the same DNA target region[267]. For example, MEIS and TGIF proteins are antagonists for a common binding site on the promoter sequence of the human dopamine receptor D1A[244].

Several cellular differentiation processes are driven by the interaction between PBX and bHLH factors. This type of cooperation represents a recurrent mechanism to merge positional and cellular identity and drive specific activation of cell differentiation programs. In fact, homeodomain factors are mostly described as positional identity regulators, whereas bHLH factors impart cell identity[246].

In summary, the relative cellular abundance of different TALE proteins and the quality of the combinatorial interaction with HOX factors or other homeodomain proteins are critical for transcriptional regulation of morphogenesis and cellular differentiation processes[246,268].

1.2.1 The roles of TALE homeodomain proteins during embryonic development

The TALE homeoproteins are widely expressed in most tissues during embryogenesis and exert multiple developmental functions *in vivo*, in concert with or independently from HOX proteins[246,247]. Only a subset of the TALE loss-of-function phenotypes can be explained by their role as HOX co-factors[265]. In this section, I focus on the most characterized families, the PBX and the MEIS, which were initially identified as proto-oncogenes in various forms of leukemia and as developmental regulators in embryogenesis[247], and the TGIF.

PBX family

The PBC subclass contains TALE proteins with a PBC domain[269], which mediates the interaction with members of the MEIS/MEINOX[245,269] and the PREP[270] subfamilies. The vertebrate genome contains four *Pbx* genes: *Pbx1*, which was discovered first for its role in the human pre-B cell acute leukemia, and *Pbx2, Pbx3* and *Pbx4*, which were subsequently identified because of sequence homology to *Pbx1*[271–274]. The expression patterns of *Pbx1*, *Pbx2* and *Pbx3* mostly overlap during embryonic development, suggesting a certain degree of redundancy between the different members, whereas *Pbx4* is exclusively expressed in testes [271–274].

Pbx1-null mice are embryonically lethal, with multiple organogenesis defects, such as craniofacial and axial skeleton malformations, impaired haematopoiesis, absence of spleen and pancreas hypoplasia[246,272,275–278]. PBX factors have been shown to control the balance between cellular proliferation and terminal differentiation during processes such as ossification, patterning and morphogenesis of the limbs[268,277] and expansion of the developing spleen[268,275]. In addition, a PBX1-dependent gene regulatory network controls epithelial properties in facial morphogenesis, by directly regulating the expression of the EMT-driver *Snail1*[279].

PBX proteins have also been described to act as “pioneer factors”, working as molecular beacons at tissue-specific loci to remodel condensed chromatin and allow the binding of downstream master regulators[280,281]. Typically, during early embryogenesis, PBX proteins

directly interact with HOX or other TFs and recruit epigenetic factors on gene promoters. For instance, the process of myogenic differentiation requires PBX1 poising Myogenic Differentiation 1 (MYOD) target genes, such as the myogenin, (*Myog*) to facilitate transcriptional activation by MYOD itself at this site[282–284].

MEIS family

Meis1, Meis2 and *Meis3* and two Prep members, *Prep1* and *Prep2* (also known as *Pknox*) belong to the *Meis* subfamily of TALE homeoproteins[245,247]. The MEIS proteins are major *in vivo* DNA-binding partners of the PBX proteins, forming PBX/PREP dimers, which bind preferentially to promoters, and PBX/MEIS dimers, which recognize binding sites enriched in enhancers, intergenic regions and intragenic regions[285], and can interact with HOX proteins in ternary complexes[247,266,286,287]. *Meis1*-deficient embryos die at E14.5 displaying hematopoietic, cardiovascular and ocular defects that in part phenocopy *Pbx* gene defects, demonstrating the genetic interaction between the two groups of factors[288,289]. However, some aspects of the MEIS activity are unique, suggesting also PBX-independent functions. For instance, in the eye MEIS1, but not PBX1, directly regulates *Pax6* gene expression, which is an evolutionarily conserved master regulator of neuroepithelial cell differentiation into retinal fate[251,289,290]. Moreover, MEIS has also been suggested to control the transcriptional network driving cardiac lineage differentiation. In particular, global gene expression and chromatin organization analysis identified MEIS1 and MEIS2 as regulators of mammalian cardiac lineage development, with MEIS binding motifs enriched at cardiac-relevant enhancers[291,292].

TGIF family

The *Tgif* family comprises *Tgif1*, *Tgif2* and *TgifXL*. Structurally, TGIF1 and TGIF2 share a highly conserved TALE homeodomain and a carboxy-terminal domain, which includes two Mitogen-Activated Protein Kinase (MAPK) phosphorylation sites and confers repressive activity[293–295]. Post-transcriptional modification, mediated by the Ras/MAP kinase pathway, increases protein stability in TGIFs[296]. The central region of the proteins is less conserved, with TGIF1 exhibiting a PLDLS motif at its N-terminus, responsible for the interaction with the C-terminal-binding protein 1 (CtBP) co-repressor which is absent in the TGIF2 sequence[295] (Fig.9). Both TGIF1 and TGIF2 have been described as context-independent transcriptional repressors, either by a DNA-dependent or -independent mechanism[295]. As expected from the high degree of conservation between the homeodomains, both TGIF1 and TGIF2 recognize and bind a similar DNA consensus sequence[293,295]. Other mechanisms of transcriptional repression shared by the two factors include the recruitment of histone deacetylases (HDAC) and the SIN3 co-repressor at gene regulatory regions[295,297], with which they can interact

through the C-terminus repressive domains. TGIF1 was originally identified as a transcriptional repressor able to bind to the retinoid-responsive element on the Cellular Retinol-Binding Protein II (CRBP II) gene, thus regulating RA-mediated response[298,299]. TGIFs have been also shown to interact with TGF-β- and BMP-activated Small body size Mothers Against Decapentaplegic (SMADs), resulting in the repression of TGF-β/BMP-responsive transcription[63,295,299]. Upon TGF-β signalling stimulation, cytoplasmic SMAD2 and SMAD3 are phosphorylated by the TGF-β receptor and, subsequently, associate with SMAD4 to form a complex that translocates to the nucleus[300], where they activate target gene expression, either directly or upon interaction with general coactivators, such as p300/CBP[301,302]. Tgifs are able to bind to the activated SMAD complexes in the nucleus, resulting in displacement of coactivators and attenuation of the TGF-β-activated gene expression[63,303].

Mutations of *TGIF1* in humans are associated with holoprosencephaly[294], a genetic disorder characterized by craniofacial malformation, which is caused by incomplete cleavage of the ventral forebrain[304]. This defect suggests that *TGIF1* has a critical role during early development of the neural system[304], probably due to disruption of the SHH signalling, as demonstrated by mouse genetics experiments[305,306].

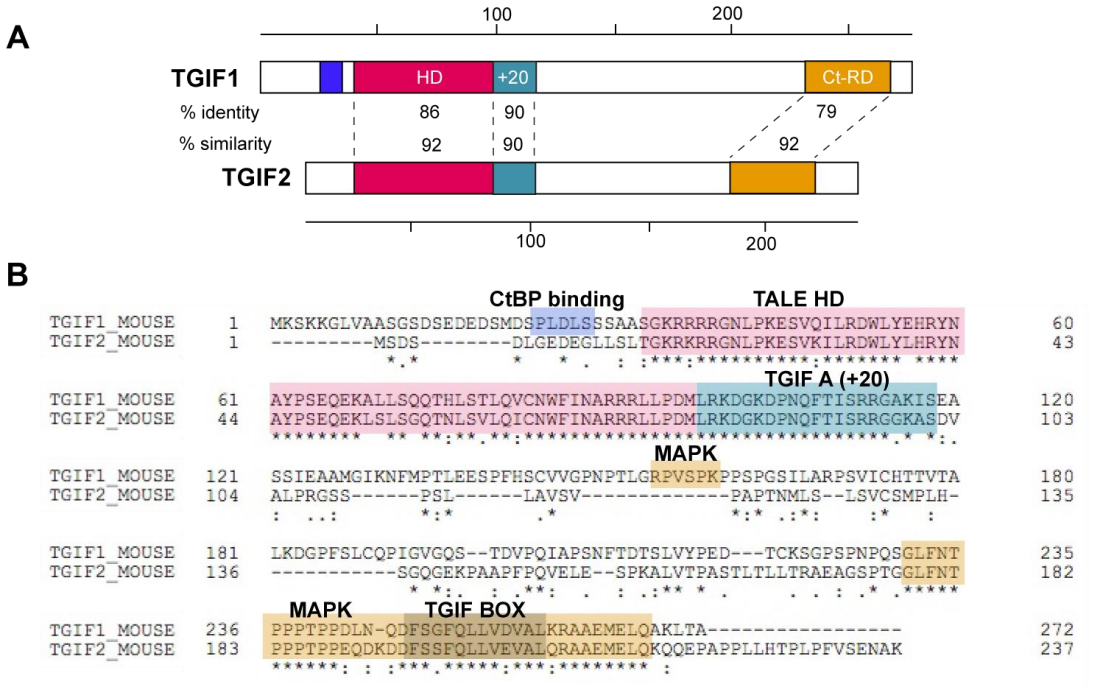

Figure 9. Conserved properties in TGIF proteins.
Schematic representation **(A)** and sequence alignment **(B)** of the mouse TGIF1 and TGIF2 proteins, displaying major features and the high percentage of identity and similarity for the conserved domains. Both TGIF proteins have the TALE homeodomain (HD) (pink), that mediates the interaction with other homeodomain proteins. A conserved 20 amino acid region is present immediately downstream of the homeodomain (TGIF A or +20) (turquoise) and another highly similar motif is at their C-terminus, termed TGIF box (gray), overlapping with the repression domain (Ct-RD) and MAPK phosphorylation site (orange). TGIF proteins interact with HDAC and SMAD proteins to exert their repressive function. The blue box represents the amino-terminus PLDLS motif, which is critical for the recruitment of the corepressor CtBP (carboxyl terminus-binding protein), that is found in TGIF1 but not TGIF2. Adapted from Wotton and Taniguchi, 2018.

Tgif1 and *Tgif2* expression is found broadly at E6.5 of mouse development[299,307,308]. *Tgif1* null mutant (*Tgif1*$^{-/-}$) mice exhibit a spectrum of mild abnormalities dependent on the genetic background, whereas *Tgif2*-deficient mutants (*Tgif2*$^{-/-}$) do not display overt phenotype[306,308–311]. Double homozygous mutants (*Tgif1*$^{-/-}$; *Tgif2*$^{-/-}$) instead die at early stages of embryonic development due to gastrulation failure[308], but the presence of at least one wild-type (WT) allele of either *Tgif1* or *Tgif2* in the epiblast is able to rescue the gastrulation defects, suggesting that TGIF activity is required for proper development of extra-embryonic endodermal structures[71,305,308].

Lastly, *in vitro* stem cell experiments showed that TGIF1 acts as an integral component of the ESC core regulatory network, balancing the expression levels of pluripotency factors[312].

1.2.2 TALE homeodomain proteins in the pancreas

Among the TALE homeoproteins, PBX transcriptional regulators are the most studied in different developmental contexts[268]. They are abundantly expressed in both the pancreatic epithelium and mesenchyme in the mouse starting from E10.5[246,278]. *Pbx1* mutant embryos phenotypes closely resembles some of the defects of *Pdx1* mutant embryos[278,313]. *Pbx1* full KO embryos do not proceed after E15.5 of development and display a broad spectrum of defects including pancreatic hypoplasia and aberrant acinar and endocrine cell differentiation. PBX1 interacts with PDX1 *in vivo*, as indicated by the analysis of double heterozygous *Pbx1*$^{+/-}$; *Pdx1*$^{+/-}$ which develop age-related diabetes in adulthood, unlike either single heterozygous mutant[278]. Indeed, the two proteins form heterodimers that engage with regions at promoters of pancreatic-specific genes, such as somatostatin, insulin, and elastase[314–316]. In pancreatic exocrine cells *in vitro*, a trimeric complex between PBX1, PDX1, and MEIS2 mediates the activation of *Elastase 1* gene expression by cooperation with PTF1A[314,316]. This research underscores that cell lineage-specific activities of pancreatic TFs, such as PDX1, depend at least in part on the availability of TALE homeoproteins[268].

Current studies in our lab. generated evidence that PBX1 has distinct requirements in the pancreatic epithelium and the surrounding mesenchyme. Specifically, genetic inactivation of this factor in a subset of cells in the pancreatic mesenchyme expressing *Nkx2-5* or *Nkx3-2* (aka *Bapx1*) factors has a non-cell-autonomous impact on the development of endocrine cells, which is mediated by ECM-integrin interactions and soluble molecules. This study underscores the requirement of Pbx1 in the pancreatic microenvironment to direct the endocrine pancreatic differentiation (Cozzitorto et al., manuscript under revision). Overall, the molecular mechanisms underlying the role of PBX proteins in mammalian pancreatic development and cell differentiation remain largely unresolved.

Detailed analysis of *Pax6* enhancers has suggested that MEIS requires interaction with PBX to regulate the *Pax6* enhancer active in the pancreas, whereas in the *Pax6* lens enhancer MEIS might bind to different co-factors[317]. Studies in zebrafish uncovered a possible role of MEIS in endodermal patterning, acting upstream of Shh signaling[318], which seems to be conserved in human according to observation from *in vitro* analysis on differentiated pancreatic progenitors[319] and transcriptomic analysis on human embryonic dorsal pancreatic bud[218]. Besides these findings, the function of MEIS proteins during pancreatic development is not well characterized. To date, defects in the developing pancreas have not been reported in the *Meis1*-deficient embryos, suggesting possible compensation by PBX proteins or redundancy with other MEIS family members.

1.2.3 TGIF2 as a developmental regulator of the liver *versus* pancreas cell fate decision

Compared to other TALE homeoproteins, much less is known about the function of TGIF proteins in development and, specifically, in endoderm-derived organs. TGIF2 was identified as target of the endodermal TF *Gata5* in the Xenopus laevis[63]. Gain-of-function experiments in the frog demonstrated that TGIF2 acts as a modifier of the endodermal fate, inducing *Pdx1* expression in posterior endoderm and limiting the expression of *Hex* in the prospective hepatic domain[63]. Silencing of *Tgif2* leads to the loss of the pancreatic region in the frog development and downregulation of the pancreatic markers *Pdx1* and *Insulin* in an insulinoma cell line model[63]. These findings underscored the requirement of TGIF2 for both the proper establishment of pancreatic fate and the maintenance of the beta-cell identity.

Gene and protein sequences of Tgif2 are well-conserved across species[63,71]. In the mouse, *Tgif2* is detected broadly in the foregut endoderm and, subsequently, its expression becomes restricted to the pancreatic progenitors and excluded from the hepatic domain[63,71]. In the adult pancreas its expression is enriched in the mature islet cells[71]. Recent studies in the Spagnoli lab. described TGIF2 as a developmental regulator of the liver *versus* pancreas cell fate decision at early stage during mouse embryogenesis[71]. *Tgif2* conditional deletion in the epiblast impairs the specification of the ventral pancreatic bud and, simultaneously, promotes an expansion of the liver bud volume[71]. This is probably due to crosstalk between TGIF2 and BMP signalling, which is known to be required for the establishment of the pancreatic identity[61,63]. Alternatively, TGIF2 can modulate the non-canonical Wnt signalling, which defines the divergence between liver and pancreatic lineages[70] (see Introduction section 1.1.3). Nevertheless, it is still unknown if TGIF2 in the endoderm also functions through other

mechanisms of action, for example by directly binding DNA through its homeodomain, or through interaction with other transcriptional factors.

In line with its developmental activity *in vivo*, TGIF2 acts as a liver-to-pancreas reprogramming factor. Specifically, stable lentiviral expression of TGIF2 in murine adult liver cells is sufficient to unlock and repress the hepatic identity and initiate the pancreatic progenitor one[71] (see Introduction section 1.3.3).

In summary, studies in the frog and in the mouse revealed that TGIF2 exerts a pro-pancreatic function at the expense of hepatic identity, which is conserved across vertebrate species[63,71]. Whether TGIF2 plays a similar biological role in the humans has not been explored yet.

1.3 Cell-based therapies to treat diabetes

Diabetes is a chronic metabolic disorder that affects over 420 million people worldwide (http://www.who.int/diabetes/). Despite the availability of insulin as treatment to temporarily restore the impaired glucose homeostasis in diabetic patients, this remedy is unable to avoid either the acute dangers of hypoglycaemia or the long-term complications of hyperglycaemia. Current efforts in regenerative medicine aim at restoring glycaemic control in diabetic patients either by replacing beta-cells or by preserving or enhancing the remnant endogenous beta-cell mass[7,8,13,16] (Fig. 10). To date, it is still matter of debate whether an adult pancreatic stem cell compartment exists and to what extent endogenous β-cells can be induced to proliferate[13,320–325]. Among all possible therapeutic approaches, whole-organ transplantation and pancreatic islets transplantation have been successful in achieving glucose control in diabetic patients[326,327]. Islet transplantation, performed according to the Edmonton protocol[327], combined together with proper immune-modulatory medication, leads to insulin independence for 5 years in at least 25% of patients in certain specialized centres[8,326,327]. Thus, the proof-of-concept for cell replacement therapy in T1D has been established.

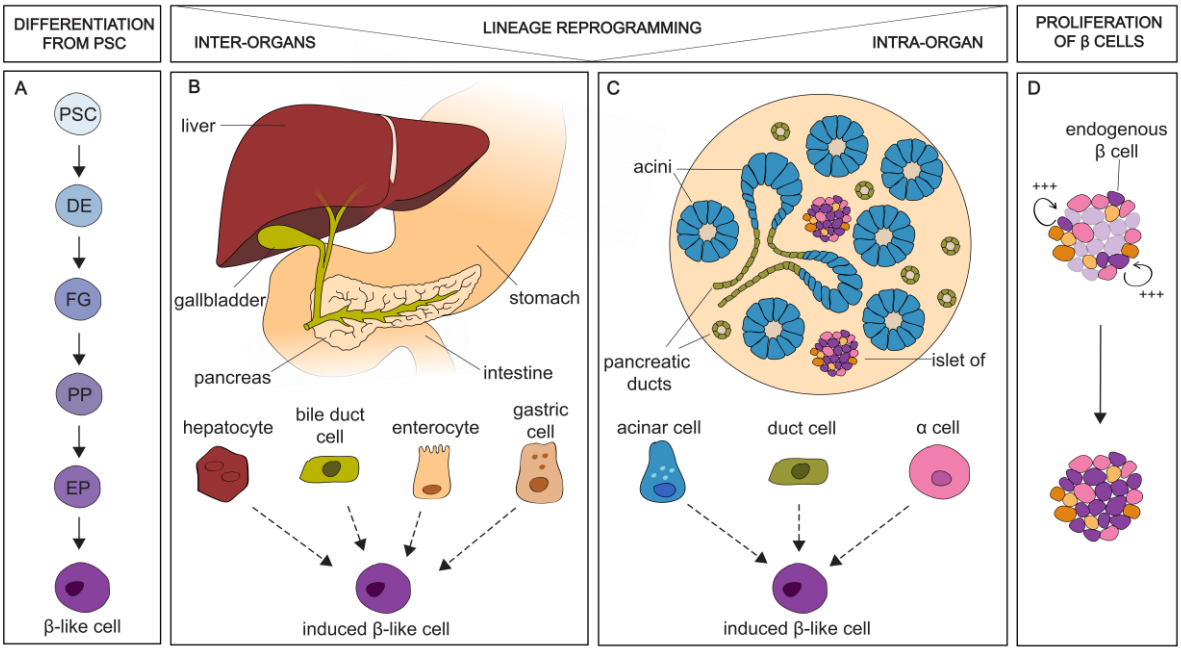

Figure 10. Strategies of cell replacement therapies to target diabetes.
Schematic overview of the different approaches to generate *de novo* β-cells to restore the glycemic function in the context of diabetes. A β-like state can be induced through a multi-step differentiation process **(A)** from pluripotent stem cells, or by direct lineage reprogramming of endoderm-derived organs **(B)** (liver, gallbladder, intestine and stomach) or other pancreatic cell types **(C)** (acinar cells, bile ducts, endocrine α-cells). Alternatively, the proliferative capacity of the endogenous β-cells could be re-awaken to replenish the pool of insulin secreting cells. PSC, pluripotent stem cells; DE, definitive endoderm; FG, foregut; PP, pancreatic progenitor; EP, endocrine progenitor. Taken from Ruzittu et al., 2019.

However, this approach is profoundly limited by the scarcity of donor islets and the risks linked to immunosuppressive regimens. Establishing a renewable source of insulin-producing cells that can be used in transplantation would eliminate the reliance on cadaveric islets and be a huge step forward for islet cell transplantation, permitting the broad application of these therapies. Moreover, if the cellular source could be autologous in origin (from the same patient), this would help to address the adverse consequences of the immune response normally triggered by the transplanted exogenous tissue[8]. Thus, recently, most efforts have focused on generating β-cell equivalents from patient-specific sources, either by differentiation of PSCs or by reprogramming of adult somatic cells of other lineages[13,16] (Fig. 10).

1.3.1 Sources and strategies to generate therapeutic pancreatic β-cells

β-cell equivalents derived from stem cells

ESCs are pluripotent stem cells that are established from the inner cell mass of the blastocyst, i.e. the pre-implantation stage embryo. They have unlimited self-renewal capacity and can give rise to all of the body cell types, with cell differentiation potential being progressively restricted during multistep developmental events[328–330]. Until recently, adult somatic cells were thought to be locked in their final differentiated state. The ground-breaking discovery that somatic cell nuclear transfer[331,332] or ectopic expression of a small set of TFs[333] can revert the somatic epigenome to a pluripotent state has opened up unprecedented opportunities for regenerative medicine, human disease modelling and drug discovery[8,13,326]. Induced pluripotent stem cells (iPSCs) can be derived from terminally differentiated cells upon ectopic expression of Oct-4, Sox2, Klf4 and c-Myc (termed OSKM factors)[333]. Induced pluripotency confers unlimited self-renewal and differentiation potential properties similarly to the ESCs. The major utility of the culture of these pluripotent cells is the generation of disease-relevant cell types.

In the context of diabetes, there is enormous interest in obtaining insulin-producing β-cell supply for disease modelling and future patient-tailored cell replacement therapies. Different groups have developed protocols to achieve *in vitro* cell differentiation of pluripotent stem cells into endoderm and pancreatic lineages by exposure to sequential combinations of compounds and growth factors, which recapitulate in a dish the signals that drive pancreatic development in the embryo[31,238–240]. iPSC and ESC directed to a pancreatic progenitor state can then further acquire the functional β-cell properties *in vivo* upon transplantation. Indeed, it has been shown that immature endocrine cells can further mature when implanted under the kidney capsule of immunocompromised mice and ameliorate the glycaemic state in chemically induced diabetic mice[334]. In the last decade, considerable advances have been achieved in the control and optimization of the differentiation process to obtain more mature glucose-responsive insulin-

producing β-cells[335]. Moreover, significant efforts are directed towards defining strategies on how to protect transplanted cells from the body's immune response either by improving immune modulation or by encapsulation devices[13]. A first clinical trial has been recently launched by the company Viacyte based on humans hESC-derived pancreatic endoderm cells (NCT #NCT03163511). In this study stem cell-derived pancreatic endoderm cells (PEC) are delivered though subcutaneous injection in an encapsulation device allowing vascularization and protection from the patient's immune system. First assessments showed that the implanted cells successfully engrafted and matured into pancreatic islet cells in the hosts.

In an alternative approach to force induction of cell identity, transient expression of OSKM factors has been used to generate an intermediate cell state, which is more amenable to be converted to the desired cell type by exposure to exogenous signals. This technique, called cell-activation and signaling-directed (CASD) lineage conversion has been applied in the reprogramming of fibroblasts to expandable pancreatic β-like cells[336,337] among other cell types[338,339].

β-cell identity induced by direct lineage reprogramming of somatic cells

Terminally differentiated cells can switch cell identity and acquire a new one without the generation of a pluripotent intermediate, in a process called direct lineage reprogramming or transdifferentiation[15,340,341]. Examples of lineage reprogramming events have been reported to occur naturally in model organisms[342,343], upon injury, in human pathological conditions[344,345], or can be experimentally induced by manipulating genetic factors or adding small molecules and growth factors[341,346,347].

The concept of direct lineage reprogramming has its fundaments in developmental biology[15]. Indeed, it is possible to rewire the genetic program of adult somatic cells by over-expressing TFs that are required during embryonic development to establish and maintain the desired cellular identity[346–348]. The first example of experimental cell conversion to be published was the induction of a myoblast state in mouse fibroblasts upon forced expression of the muscle lineage 'master regulator' MyoD1[349]. Afterwards, many examples of direct lineage reprogramming have followed in all different germ layer derivatives, including neuroectoderm, mesoderm and endoderm, across different cell lineages, *in vitro* as well as *in vivo*[15,341,347,350].

To design optimal strategies for direct lineage conversion, several milestones need to be take into account[15]. First, the minimum set of TFs must be defined[351]. This largely depends on the selection of the source cell type, which imposes a specific molecular context resulting from its gene expression profile and epigenetic signature. Cell lineages that are closely related in development, deriving from a common progenitor, can interconvert more easily than more distant ones since less transcriptional remodeling and epigenetic changes are required[351,352].

Moreover, it is suitable to choose self-renewing target cells located in a proper place of the body to avoid loss of function of the cellular source after conversion. An important requirement in a successful direct reprogramming strategy is the repression of the identity of the original cell type[15]. Ectopic expression of suitable lineage-instructive TFs should act to erase the function of terminally differentiate cells and impose a new molecular program. An incomplete conversion of cell identity leads to hybrid states which result in defective functional properties and consequent unwanted outcomes in the host. Then, the obtained desired cell types need to be thoroughly characterized for molecular profile and physiological behavior by comparing them with their *in vivo* counterparts and employing stringent *ex vivo* or i*n vivo* functional tests[71,350]. In particular, it is relevant to evaluate if the transdifferentiated cells acquired a stable new identity (e.g. do not revert to the original or other cell states) and are able to integrate into the appropriate anatomical and restore the lost function in the recipient[15]. Finally, safe non-invasive strategies to deliver the reprogramming agents must be defined in order to avoid adverse side effects[15].

Direct cell identity conversion is conceptually attractive in regenerative medicine providing an alternative therapeutic avenue with a relatively lower risk of tumorigenesis when compared to the use of pluripotent stem cells. It also provides the opportunity to directly convert cells *in situ*, which is relevant in certain regenerative strategies[15]. Lineage reprogramming potentially offers a cure to diabetes by generating β-cell equivalents from various somatic cells sources and using different strategies[15,16,353].

Intra-pancreatic reprogramming resulting in β-cells formation

Recent discovery of high degree of cellular plasticity in the adult pancreas has pointed to pancreas-resident cells as potential sources for new β-cells [354,355] (Fig. 10). Seminal work from Melton and colleagues provided the first evidence that pancreatic acinar cells can be directed to acquire a beta-cell phenotype upon *in vivo* forced expression of the TFs Pdx1, MafA and Ngn3 (hereafter referred to as PMN), which are key developmental regulators of β-cell fate[216,354]. Since then, the same triplet of TFs has been applied to convert multiple cell types into beta-like fate, including pancreatic ducts[356], as well as non-pancreatic cells types, such as enteroendocrine cells[357], liver cells[358] and bile ducts[359,360].

Within the exocrine compartment, pancreatic ductal cells have also been studied for their potential to switch to the endocrine lineage. Sancho et al. reported that *in vivo* inactivation of *Fbw7* in mouse ductal cells stabilizes NGN3 protein resulting in their conversion into hormone-secreting cells[361]. Alternatively, neogenesis of β-cells in pancreatic ducts can be triggered *in vivo* by overexpression of PMN [356] or in combination with Pax6[362].

Given the close developmental lineage and endocrine functions[1,363], fate interconversion between the different pancreatic islet cells has been extensively explored by various groups[364]. For example, Herrera and colleagues have reported that fate interconversion between endocrine cells can occur spontaneously, following extreme toxin-induced β-cell damage in adult mice[365]. More recently, the same group demonstrated that also human α-cells, obtained from cadavers, can serve as source to generate *de novo* beta-cells upon introduction of Pdx1 and MafA[366]. Notably, alpha cells are characterized by a bivalent chromatin signature at genomic sites of α-cell-specific transcription factor (Arx) and β-cell specific (Pdx1 and MafA) regulator genes[367], which is in support of the great plasticity that these cells retain. Consistently, multiple studies have shown that pancreatic α-cells can be converted into β-like cells after deletion of *Arx* alone[207,368] or simultaneously with the DNA methyltransferase (*Dnmt1*) epigenetic modifier[369], overexpression of TFs necessary for β-cell differentiation such as *Pax4*[203] or *Pdx1* and *Mafa*[214] in mice.

Inter-organ reprogramming resulting in β-cell formation

The intra-pancreatic cell fate conversion presents an undeniable limitation, which is the anatomically inaccessible location of the pancreas. Thus, an alternative approach with more immediate therapeutic applications might be to promote lineage reprogramming in other more easily accessible cell types outside of the pancreas. *In vivo* reprogramming screens using mouse transgenic for the three TFs PMN identified enteroendocrine progenitor cells of the gut as well as the stomach as sources of insulin-producing cells[357,370]. This is conceivable since the gastrointestinal tract is also derivative of the endoderm germ layer, therefore it shares a close origin with the pancreas. Zhou and colleagues showed that antral stomach endocrine cells undergo conversion into insulin$^+$ cells more efficiently than intestinal enteroendocrine cells and displayed some close molecular and functional similarity to endogenous β-cells[370]. Similarly, enteroendocrine cells in the intestinal crypts respond to *Foxo1* depletion *in vivo* in the mouse by inducing ectopic insulin expression[371]. Nevertheless, the induced insulin$^+$ cells retain intestinal properties and might not survive long-term, if definitively lacking the important protector of β-cell against cell stress, Foxo1[371].

Finally the gallbladder, which is part of the extrahepatic biliary tree and shares common development with the ventral pancreatic rudiment, retains a certain potential to reprogram to the pancreatic lineage[68]. Grompe's group recently reported that adenoviral (Ad)-mediated overexpression of the PMN transgenic cassette in the biliary epithelium, along with inhibition of hedgehog and BMP signalling, induced endocrine molecular features that can be further enhanced by the addition of the TF Pax6, both in mouse and human *ex vivo* systems[359,360].

However, the obtained cell products do not inactivate completely the gallbladder program and, moreover, display multi-hormonal properties.

1.3.2 Reprogramming of liver cells into pancreatic cell types

The pancreas and the liver are responsible for the hormonal control of energy metabolism. Because of the close embryonic origin, the anatomic close position, the transcriptional and functional similarities and the intercellular plasticity between the two organs (see Introduction section 1.1.3), the liver represents an attractive source for generating new pancreatic β-cells. Additionally, the high regenerative ability of the adult liver after injury and accessibility make it a very good candidate tissue from the clinical perspective.

To date, most of the reprogramming efforts targeting hepatic cells have focused on forcing the expression of pancreatic-specific TFs. Table 1 lists the *in vitro* and *in vivo* attempts of liver-to-pancreas reprogramming in different organisms and using different methods.

Pioneering studies in the laboratory of Sarah Ferber started hepato-pancreatic transdifferentiation experiments based on transient Ad-mediated expression of Pdx1[372]. This work was then expanded by her group and others by combining Pdx1 with other pancreatic TFs and/or by supplementing the culture medium with various small molecules and growth factors in both mouse and human cultures (Table 1). For example, adding EGF and nicotinamide to the culture improved PDX1 ability to partially convert human adult and foetal liver cells *in vitro* to an immature endocrine state, as indicated by the expression of *Ngn3* in the resulting culture[373].

Quite commonly these studies showed that Ad-Pdx1-transduced mouse adult liver cells cultured *in vitro* are able to produce insulin to some extent and ameliorate the glycaemic status of diabetic mouse model upon transplantation[372–374] (Table 1).

Despite the well-established and fundamental role of Pdx1 in pancreatic development, little is known about its downstream transcriptional cascade and how this TF might work in the context of liver-to-pancreas conversion. Pdx1 has been shown to promote liver-to-pancreas transition more efficiently in Xenopus and mouse when fused to the VP16 transcriptional activation domain[375–377]. These findings suggested that Pdx1 works together with an activator in the context of a pancreatic progenitor and perhaps the identification of such factor might help Pdx1 reprogramming activity for example in a hepatocyte context. More recent work from Dunn's group showed by ChIP-Seq and ChIP-qPCR PDX1 occupancy and binding at hepatic genes not only in differentiating hESCs but also in HepG2 hepatoma cells[378].

SPECIES	CELL SOURCE	SYSTEM	REPROGRAMMING STRATEGY	*IN VIVO* FUNCTIONAL OUTCOME	REF.
Mouse	Adult liver cells	*in vivo*	Pdx1 (Ad)	Amelioration of hyperglycemia	372
Mouse	Adult liver cells	*in vivo*	Pdx1 (Ad)	Amelioration of hyperglycemia	374
Mouse	Adult liver cells	*in vivo*	NeuroD1 (HDAd), Betacellulin	Amelioration of hyperglycemia	379
Mouse	Adult liver cells	*in vivo*	Pdx1(OE in Alb-Cre mice)	n.d.	380
Mouse	Adult liver cells	*in vivo*	Pdx1, Ngn3 (AAV8)	No amelioration of hyperglycemia	381
Mouse	Adult liver: hepatocytes and oval cells	*in vivo*	Ngn3 (HDAd), Betacellulin	Transient expression of insulin	382
Mouse	Hepatic ducts	*in vivo*	Pdx1, Ngn3, MafA (Ad)	Long-term reversal of hyperglycemia	358
Mouse	Adult liver cells	*in vivo*	Pdx1, NeuroD1, MafA (Ad)	Short-term reversal of hyperglycemia	383
Mouse	Adult liver cells	*in vivo*	Ngn3 (HDAd), Betacellulin, Socs1	Long-term reversal of hyperglycemia	384
Xenopus and human	Immature liver of transgenic tadpoles (TTR-Xlhbox8) and HepG2	*in vivo* and *in vitro*	Pdx1-VP16	n.d.	375
Mouse	Adult primary hepatocytes and BAML	*in vivo* and *in vitro*	Tgif2 (Lv and AAV8)	Amelioration of hyperglycemia	71
Mouse	Hepatic oval cell line	*In vitro*	Pdx1-VP16, exposure to high glucose	Amelioration of hyperglycemia	376
Mouse	Adult primary hepatocytes	*In vitro*	NeuroD, Ngn3, Pax4 (Ad), HGF, Betacellulin, Dexamethasone (-)	n.d.	385
Mouse	Embryonic liver cells	*In vitro*	Pdx1, Ngn3, MafA (Ad)	n.d.	386
Rat	Liver stem cells (oval)	*In vitro*	High glucose	Amelioration of hyperglycemia	387
Rat	Liver epithelial WB cells (stem-cell like)	*In vitro*	Pdx1-VP16, Pax4 (Lv)	Amelioration of hyperglycemia	388
Rat	Liver epithelial WB cells (stem-cell like)	*In vitro*	5-AZA, TSA, RA, insulin, transferrin and selenite (ITS) and nicotinammide	Amelioration of hyperglycemia	389
Human	Fetal liver cells	*In vitro*	hTERT, Pdx1 (Lv)	Amelioration of hyperglycemia	390
Human	Adult and fetal liver cells	*In vitro*	Pdx1 (Ad), Nicotinamide, EGF	Amelioration of hyperglycemia	373
Human	Adult liver cells	*In vitro*	Pdx1, Nkx6.1(Ad)	n.d.	391
Human	Adult liver cells	*In vitro*	Pdx1 (Lv)	n.d.	392
Human	Hepatocellular carcinoma cells (Huh7)	*In vitro*	Pdx1, Pdx1-VP16, Ngn3 (Lv)	n.d.	393
Pig	Neonatal liver cells	*In vitro*	Pdx1-VP16, Beta2/NeuroD1, MafA (Ad)	Amelioration of hyperglycemia	394

Table 1. Liver-to-pancreas reprogramming studies.
This table summarizes the studies performed to investigate hepatic-pancreatic direct reprogramming and plasticity. Lv, lentivirus; AAV, adeno-associated virus; Ad, adenovirus; HDAd, helper-dependant adenoviral vectors; OE, over-expression; n.d., not determined; ref, reference. Adapted from Ruzittu et al., 2019.

Consistently, Pdx1 overexpression in hepatoma lines suppresses expression of a sub-set of endogenous liver genes, such as *Albumin*[378] as well as *Hnf1a* and *Hnf4a*[393]. Thus, PDX1 might act as a context-dependent transcriptional repressor and activator within the same cell type, possibly explaining its activity when overexpressed in liver cells.

Yechoor et al. showed that *in vivo* delivery of Ngn3 transgene is sufficient to activate the expression of some pancreatic endocrine genes in periportal hepatic progenitor cells resulting in a transient secretion of insulin along with other pancreatic hormones[382]. Different

combinations of various pro-endocrine factors, such as Ngn3, NeuroD1, Nkx6.1, MafA, Pax6, have been tested to improve the maturation of reprogrammed hepatic cells[379,383,391,394]. Remarkably, the most characterized combination of reprograming factors includes Pdx1, Mafa and Ngn3, expressed in equal molarity by the transgenic cassette "PMN" (Table 1). Ad vectors have been used by various groups to induce abundant expression of the PMN transgene *in vivo* in the liver as well as *in vitro* in liver cultures (Table 1). Slack and colleagues showed that PMN overexpression in the liver of diabetic mice results in a partial reprogramming to insulin-positive cells with a mixed phenotype, displaying some typical features of duct cells and some properties of beta cells[358]. Moreover, the induced insulin$^+$ duct-like cells were shown to arise from SOX9$^+$ cell population, which are probably cells of small bile ducts[358]. The PMN transgene induces a similar partial reprogramming also in foetal liver cultures, whereby hepatoblasts are the most amenable cells to undergo reprogramming and their competence declines during development[386].

Also, in all these studies the conversion into pancreatic identity is incomplete since cells still exhibit features of the original cell type and fail to sense glucose levels, resulting in a partial correction of the glycaemic state in diabetic mice. Thus, PMN-mediated reprogramming in a liver context results in an immature or multiendocrine cell type, yet to be fully characterized.

It should be noted that all the studies aforementioned employed adenoviral vectors for mediating transgene expression. Actually, when the same TFs (*e.g.* Pdx1 and Ngn3) were delivered to the liver of diabetic mice using adeno-associated virus (AAV), which has a high transduction efficiency, but lower immunogenic potential compared to adenoviruses, the mice remained hyperglycemic[381]. This study also showed that irrelevant elements of the adenovirus capsid might be responsible together with the pancreatic TFs for the reprogramming[381], possibly by inducing an inflammatory response that is more conducive to reprogramming, as previously suggested in other contexts[395]. Hence, it seems that the most appropriate route for reprogramming hepatocytes into pancreatic β-cells has yet to be defined.

1.3.3 TGIF2 as a lineage reprogramming factor

In an effort to define additional reprogramming factors and successful strategies for converting hepatocytes into pancreatic cells, our group started thorough analyses of developmental regulators controlling the fate decision between the two lineages as well as of the lineage dynamics[70,73]. From these studies, we found that TGIF2 acts as a regulator of cell fate binary choice between liver and pancreas[71]. In the embryo, TGIF2 promotes the establishment of pancreatic identity at the expenses of the hepatic one[71] (as described in Introduction section 1.1.3), being therefore a suitable candidate to induce cell fate conversion. Consistently, when ectopically expressed by a lentiviral vector in mouse hepatocytes *in vivo*

and *in vitro*, TGIF2 mediates repression of the hepatic phenotype and triggers the activation of a pancreatic progenitor program[71]. The obtained reprogrammed cells can be further differentiated along pancreatic cell lineage and secrete some insulin upon transplantation in diabetic mouse models[71]. These findings suggest that this might be a starting point for achieving full conversion into mature β-cells upon combination of TGIF2 with additional TFs or proper culture conditions. Finally, identifying epigenetic barriers that need to be override or the existence of putative privileged hepatocytes populations that can more easily change identity would be of considerable interest to effectively produce the desired pancreatic cell type.

If TGIF2 can convert also human liver cells into pancreatic cell types is still unknown and would deserve extensive investigation for its clinical relevance.

Moreover, it is still unexplored whether the reprogramming activity of TGIF2 is lineage-restricted or it applies to multiple contexts in terms of cell origin and differentiation state. For example, the same factor could be used to reprogram more easily accessible tissues, i.e. skin or blood, or unlimited self-renewal sources, i.e. pluripotent stem cells. Answering to these questions will be fundamental to elucidate the mechanisms underlying TGIF2 biological activity and tissue plasticity and, ultimately, to develop a possible strategy in the context of a cell-based therapy to cure diabetes.

2. AIMS OF THE STUDY

The liver and the pancreas share a close developmental origin, being both organs specified at the same developmental stage during embryogenesis and from common endoderm progenitors. Moreover, the two adult organs share transcriptional and functional similarities, as they perform essential roles in the control of the body energy metabolism. However, to date, the molecular basis of hepatic and pancreatic cell lineage restriction is still poorly understood. This knowledge would be extremely important for the establishment of a liver-to-pancreas direct lineage conversion approach to generate new β-cells that might function to rescue glycaemic control in diabetic patients.

We recently uncovered the role of the Three-Amino-Acid Loop Extension (TALE) homeodomain transcriptional regulator TG-interacting factor 2 (TGIF2) in controlling liver *versus* pancreas cell fate decision[71]. In the early mouse embryo, TGIF2 expression is required for the allocation of the pancreatic cell fate. In line with this developmental activity, TGIF2 acts as a liver-to-pancreas reprogramming factor, being sufficient to repress hepatic genes expression and to establish a pancreatic cell identity in liver cells. These results in the mouse were consistent with previous observations made in other vertebrate species, such as Xenopus, suggesting conservation of its function across species[63,71]. Overall, these previous studies unveiled the reprogramming activity of the TALE homeoprotein TGIF2 and raised interesting questions regarding the mechanisms of action of this factor and the translational impact of TGIF2-mediated cell fate conversion.

My studies focused on further exploring the biological function of TGIF2 and addressed the conservation of the reprogramming function in the human cells.

The first aim of the present work was to define the *in vivo* requirements of TGIF2 in the embryonic pancreas, after the time of lineage specification. I inter-crossed the mouse *Tgif2* floxed allele line with the *Pdx1-Cre* transgenic strain to genetically inactivate*Tgif2* in the pancreatic rudiment. The analysis of *Tgif2*-deficient embryos revealed a crucial role of TGIF2 in the maintenance of the pancreatic identity, preventing the acquisition of hepatic features in the developing pancreas. Additionally, I found that ablation of both *Tgif1* and *Tgif2 genes* impairs endocrine cell fate differentiation, which is accompanied by an expansion of the acinar compartment, suggesting that TGIF TALE homeoproteins control binary cell fate choice at different stages of pancreatic development.

Secondly, I investigated whether the reprogramming activity of TGIF2 is conserved in human liver cells and how to improve the maturation of pancreatic progenitor cells obtained through the current reprogramming strategy. Stable lentiviral-mediated forced expression of

TGIF2 in human adult primary hepatocytes results in the robust repression of the original hepatic identity and induces the expression of key pancreatic genes to some extent.

Finally, the third aim of my thesis was to determine if TGIF2 reprogramming activity is lineage-restricted or it applies to multiple contexts in terms of cell origin and differentiation state. To investigate the reprogramming potentials of TGIF2 outside an endodermal cellular context and better understand its mechanisms of action, I have established different *ex vivo* lineage reprogramming strategies in mouse and human fibroblasts. Ectopic expression of TGIF2 promoted morphological changes in fibroblasts, inducing the acquisition of an epithelial shape, and activated the expression of pancreatic marker genes.

Taken together, these multiple approaches enabled me to further elucidate the molecular mechanisms regulating pancreatic identity and plasticity in mammalian species. These findings will help to ultimately develop an innovative cell-based therapy to cure diabetes.

3. MATHERIALS AND METHODS

3.1 Animal experiments

3.1.1 Mouse care and husbandry

Mouse strains used in this study are: wild-type *C57BL/6*, *B6.FVB-Tg(Ipf1-cre)^{1Tuv}* (Hingorani et al. 2003)[396]; TGIF1$^{tm1aP\ ah}$ (Mar et al. 2006)[397]**,** TGIF2$^{tm1a(EUCOMM)Wtsi}$. For lineage tracing studies the following lines was used: Gt(ROSA)26Sor$^{tm2(HIST1H2BB/EGFP)Sia}$ (Abe et al. 2011)[398]. Timed matings were set up in the evening to obtain embryos at different stages of development. The presence of a vaginal plug in the morning was considered as day 0.5 post coitum (dpc). Mice were kept in standard conditions and manipulated according to the regulation of the local animal protection authority (Landesamt für Gesundheit und Soziales, Berlin).

3.2 Cell culture methods

3.2.1 Culture of liver cell lines

Bipotential adult mouse liver (BAML) cell lines[399] were cultured in William's E medium without phenol red (Sigma), 2 mm stabile glutamine, Penicillin / Streptomycin, 10% Fetal Bovine Serum (FBS), 30 ng/mL IGF II, 50 ng/mL EGF, 10 µg/mL Insulin, 0.1 µm dexamethasone and 10 mM nicotinamide (Sigma) on collagen type I (Serva) coated dishes.

3.2.2 Culture of primary human hepatocytes

Plateable cryopreserved primary hepatocytes were purchased from Invitrogen (Cat. HMCPTS, Lot.HU8216; Cat. HMCPP5, Lot.HPP20160501) or Lonza (Cat. HUCPG, Lot. HUM4191, Cat. HUCPQ, Lot. HUM4180A, Lot. HUM4075B). They were thawed following the manufacturer's protocols on collagen type I (Serva) treated 12-well plates at a density of $7.8x10^5$ cells/well. Briefly, cells were diluted in commercial thawing medium (Invitrogen or Lonza, respectively) and centrifuged for 10 minutes at 100 g. The cell pellet was resuspended in plating medium consisting of William's E medium (Sigma) supplemented with commercial Serum-containing Hepatocyte Plating Supplement Pack (Invitrogen) or commercial Plating medium (Lonza), respectively. After 6 to 8 hours post-plating, the medium was replaced by "HEP medium" composed by William's E medium (Sigma), 2 mM stabile glutamine, Penicillin / Streptomycin, 5% FBS, 50 ng/mL EGF (PeproTech), ITS (Gibco), 0.1 µm dexamethasone (Sigma) and 10 mM nicotinamide (Sigma). Alternatively, HLIM medium was used to culture the cells, which

was composed of Advanced DMEM-F12 (Gibco), 1% FBS, Penicillin / Streptomycin, B27 supplement (minus vitamin A) (Sigma), 1mM N-acetyl-cysteine (Sigma), 10 mM nicotinamide (Sigma), 100 ng/ml FGF10 (R&D Systems), 50 ng/ml EGF (R&D Systems), 25 ng/ml Hepatocyte Growth Factor (HGF) (PeproTech), 10 µM Rho-associated protein kinase (ROCK) inhibitor Y-27632 (Sigma), 5 µM ALK5 inhibitor II (Enzo), 50 ng/ml Wnt3a (R&D Systems). This is a modified version of the medium composition described in Zhang et al. 2018[400].

3.2.3 Culture of mouse fibroblasts

Primary murine fibroblasts were isolated from the tip of the tail of adult mice. Briefly, 1 cm of mouse tail was cut into small pieces with sterile scissors and digested with an enzymatic solution made of 2U/ml Dispase I (Sigma) and 0.2% Collagenase A (Sigma) in PBS. After 1 h incubation shaking at 37°C, the cell suspension was centrifuged for 5 minutes at 100 g. The supernatant was removed and the pellet was resuspended in complete medium DMEM (Gibco), 10% FBS, Penicillin / Streptomycin. Cells were seeded on 6cm plates and fed with fresh medium the day after.

3.2.4 Culture of human fibroblasts

Human adult dermal fibroblasts were obtained from skin of healthy donors. In-house cells were MDC000122 and MDC0001223 fibroblasts, kindly provided by Dr. Diecke (BIH, Berlin) and CD90+ FACS-enriched dermal fibroblasts, provided by Dr. Pilippeos (Watt group, KCL, London), obtained as described before[401] from surgical waste skin of a 17-years old female donor. Commercial samples were Normal Adult Human Dermal Fibroblasts (NHDF-Ad) purchased from Lonza (Cat. CC-2511, lot 489572, male, 32). Cells were cultured on 0.1% Gelatine-coated plates (Sigma) in complete medium composed of High Glucose DMEM (Gibco), 10% FBS, 2mM Glutamax, NEAA and Penicillin / Streptomycin, supplemented with 8 ng/ml basic FGF (R&D Systems). Early passage cells were used for reprogramming experiments.

3.2.5 Culture of human induced pluripotent stem cells (hiPSCs)

Human iPSC lines (iXM001 and BIH004) were maintained on Geltrex-coated (Invitrogen) plates in home-made E8 media, as reported by Chen et al. (2011)[402], under hypoxic conditions. The medium was changed daily and cells were passaged every ~3 days as cell clumps or single cells using 0.5 mM EDTA (Gibco) or Accutase (Gibco), respectively. Medium was supplemented with 10 µM Rho-associated protein kinase (ROCK) inhibitor Y-27632 (Sigma) when iPSCs were thawed or passaged as single cells.

3.2.6 Differentiation of pluripotent iPSCs into pancreatic β-like cells

Differentiation was carried out following a 21-day protocol described by Russ et al. (2015). Briefly, iPSCs were dissociated using Accutase and seeded at a density of 5.5×10^6 cells per well in ultra-low attachment 6-well plates (Thermo Fisher Scientific) in 5.5 ml E8-home medium supplemented with 10 µM ROCK inhibitor, 10 ng/ml activin A (R&D Systems) and 10 ng/ml heregulin-b1 (Peprotech). Plates were placed on an orbital shaker at 100 rpm to induce sphere formation at 37°C in a humidified atmosphere containing 5% CO_2.

To induce definitive endoderm differentiation, cell clusters were collected after 36 h in a 50-ml Falcon tube, washed with PBS and re-suspended in d1 media [RPMI (Invitrogen) containing 0.2% FBS, 1:5.000 ITS (Invitrogen), 100 ng/ml activin A and 50 ng/ml WNT3a (R&D Systems)]. Clusters from two wells were combined into one well and distributed into low-attachment plates in 5.5 ml d1 media. Subsequently, cell clusters were differentiated into β-like cells by exposure to the appropriate media as previously published (Russ et al., 2015)[240]. Exactly: d2 [RPMI containing 0.2% FBS, 1:2,000 ITS, and 100 ng/ml activin A]; d3 [RPMI containing 0.2% FBS, 1:1,000 ITS, 2.5 µM TGFbi IV (CalBioChem), and 25 ng/ml KGF (R&D Systems)]; d4-5 [RPMI containing 0.4% FBS, 1:1,000 ITS, and 25 ng/ml KGF]; d6-7 [DMEM (Gibco) with 25 mM glucose containing 1:100 B27 (Gibco), 3 nM TTNBP (Sigma)]; d8 [DMEM with 25 mM glucose containing 1:100 B27, 3 nM TTNBP, and 50 ng/ml EGF (R&D Systems)]; d9 [DMEM with 25 mM glucose containing 1:100 B27, 50 ng/ml EGF, and 50 ng/ml KGF]; d10-14 [DMEM with 25 mM glucose containing 1:100 B27, 500 nM LDN-193189 (Stemgent), 30 nM TBP (Millipore), 1mM ALKi II (Axxora), and 25 ng/ml KGF]; and d15-21 [DMEM with 2.8 mM glucose containing 1:100 Glutamax (Gibco) and 1:100 NEAA (Gibco)].

3.2.7 Treatment of human primary hepatocytes with pancreatic differentiation cytokines

Based on the media composition described in section 3.2.6, hepatocytes and fibroblasts undergoing reprogramming were exposed to a combination of pro-pancreatic cytokines, corresponding to the ones promoting transition from a pancreatic progenitor (PP) to an endocrine progenitor (EP) state. Specifically, primary hepatocytes were grown in Advanced DMEM-F12 (Gibco), 1% FBS, 2mM Glutamax, Penicillin / Streptomycin containing 500 nM LDN-193189 (Stemgent), 30 nM TBP (Millipore), 1mM ALK5 inhibitor II (Axxora), and 25 ng/ml FGF10 (R&D Systems).

3.2.8 Treatment of human fibroblasts with chemical compounds

Human fibroblast cells were treated with various chemical compounds (described in Table 3 in Result section 4.3.2.) in the following way in independent experiments: LV.hTGIF2-transduced cells were treated after 2 weeks of culture with 1mM ALK5 inhibitor II (Axxora) until the day of analysis; exposure to 1µM 5-azacytidine (5-AZA) was carried out on fibroblasts for 18 h prior to lentiviral transduction, as described before[403]; treatment with pancreatic differentiation cytokines was carried out at 3 weeks of culture as described in 3.2.7.

3.2.9 3D culture of reprogrammed human fibroblasts

Fibroblasts undergoing reprogramming were cultured for 3 weeks in monolayer as described in 3.2.4. Next, they were dissociated using 0.05% Trypsin-EDTA (Gibco) and seeded in ultra-low attachment 6-well plates (Thermo Fisher Scientific) in 5.5 ml DMEM (Gibco) with 25 mM glucose 1% FBS, Penicillin / Streptomycin, supplemented with 500 nM LDN-193189 (Stemgent), 30 nM TBP (Millipore), 1,000 nM ALK5 inhibitor II (Axxora), and 25 ng/ml FGF10 (R&D Systems). Plates were placed on an orbital shaker at 100 rpm to induce sphere formation at 37°C in a humidified atmosphere containing 5% CO_2.

3.2.10 Culture of Human Embryonic Kidney cells (HEK)

Human embryonic kidney (HEK) cells were cultured in High Glucose DMEM (Invitrogen), 10% FBS and Penicillin / Streptomycin. Cells were passaged every 2-3 days, when they reach 90% confluence culture, by trypsinization.

3.2.11 Transfection of eukaryotic cells with DNA

Eukaryotic cells were transfected using Polyethylenimine "Max" (PEI) (Polysciences). Up to 10 µg of plasmid DNA was mixed with 20 µg PEI by vortexing. The mixture was incubated for 30 minutes at room temperature. The cells were kept in OptiMEM medium (Invitrogen) during the transformation procedure.

3.2.12 Production of LV

Lentiviral particles were generated in the lab by transient co-transfection of HEK 293T cells with the 2nd generation packaging plasmid, psPAX2, envelope plasmid, VSV-G, and the transgene expressing plasmid (see Methods section 3.3.6). Cells were transfected using PEI (Polysciences). Transfection efficiency was measured by detection of GFP expression in the

HEK cells 2 days post-transfection. Culture supernatants were harvested after 48-72 hours post-transfection and filtered under vacuum using 0.45 µm filter units to remove cell debris. The virus was concentrated by ultracentrifugation through a 20% sucrose cushion at a speed of 17000 rpm for 3 hours or at speed of 11000 rpm for 18 hours at 4°C. The pellet was resuspended in PBS in 1:100-1:300 of the original volume, aliquoted and stored at -80°C. Virus titer was determined by Fluorescent Activated Cell Sorting (FACS) on HEK cells transduced with a serial virus dilution, as previously described[404].

3.2.13 Production of AAV

AAV2/8 virus production was performed by the University of Pennsylvania Vector Core.

3.2.14 Eukaryotic cell transduction

Fibroblast and hepatic cell cultures were transduced with a MOI (Molteplicity Of Infection) of 10 to 20 of the lentiviral vectors. AAV transduction was carried out calculating a MOI of $5x10^5$. The efficiency of infection was monitored by detection of the fluorescence emitted by the GFP or mCherry fluorophores. Cell morphology and gene expression changes were analysed by IF analysis and quantitative RT-qPCR assay at different time points after transduction.

3.3 Molecular biology methods

3.3.1 Culture and cryopreservation of bacterial cells

DH5α, One Shot TOP10 and MAX efficiency Stbl2 (Invitrogen) competent E.coli were grown in Luria Broth Base (LB) (Sigma), containing 10 g/L tryptone, 5 g/L yeast extract and 5 g/L NaCl. For long-term storage, 500 µL of an overnight culture of E.coli was mixed with 500 µL of 50% glycerol and frozen at -80°C.

3.3.2 Transformation of bacteria with DNA

Chemically competent E.coli DH5α, TOP10 or Stbl2 (Invitrogen) were transformed with purified plasmid DNA or DNA-ligation reactions using a standard heat-shock protocol. Briefly, 100 ng of plasmid DNA was mixed with 50 µL commercial heat-shock competent bacteria cells and incubated on ice for 30 minutes. Subsequently, bacteria cells were heat-shocked at 42°C for 30 seconds. After letting the cells recover for 2 minutes on ice, they were gently mixed with 900 µL SOC medium (Invitrogen) and incubated at 37°C with shaking for 1 hour. Stbl2 cells

were shaken at 30°C for 90 min. Subsequently, transformed bacteria were plated on LB agar plates (15 g/L Bacto Agar) with the appropriate antibiotics (Kanamycin at 50 µg/mL or Ampicillin at 50 µg/mL) and placed in the incubator overnight at 37°C (DH5α ant TOP10) or 30°C (Stbl2) to let bacteria colonies grow.

3.3.3 Isolation of plasmid DNA from bacteria

5 mL to 100 mL of LB medium with the appropriate antibiotics were inoculated with a single colony of E.coli picked from an LB agar plate or a glycerol stock. The bacteria were grown overnight at 37 °C with shaking at 180 rpm. The next day, the plasmid DNA was isolated using a plasmid miniprep (Zymo), endotoxin-free maxiprep or megaprep kits (Qiagen).

3.3.4 Enzymatic digest of DNA

Restriction enzymes were bought from New England Biolabs, USA. Digests were performed with 0.5 U – 1 U/µg DNA at the adequate temperature for 2 hours to overnight in the recommended buffers.

3.3.5 Isolation of DNA from agarose gels

Polymerase chain reaction (PCR) products or digested DNA were resolved on 1.5% agarose gels and the corresponding bands were cut from the gel. DNA was extracted using the ZymoClean gel DNA recovery kit (Zymo).

3.3.6 Cloned Constructs

Generation of AAV.hTGIF2-2A-GFP. The open reading frame (ORF) coding sequence of human *TGIF2*, excluding the STOP codon, was amplified by PCR from the construct pOTB7-hTGIF2 using primers with restriction enzyme sites (referred to as *hTGIF2-EcoRI-MluI-F* and *hTGIF2-ClaI-R*, see Table 2). First, it was cloned into pBSKII vector containing the T2A-GFP sequence (enhanced green fluorescent protein linked to the 2A self-cleaving sequence[405]). The pBSKII-TGIF2-2A-GFP plasmid was therefore created by ligation of vector and insert cut with EcoRI and ClaI. The hTGIF2-2A-GFP sequence was then inserted into the pENN.AAV.TBG.PI vector (Penn Vector Core, AV-8-PV0146) using MluI and KpnI sites, leaving intact the ITR and other regulatory sequences. The final plasmid DNA is pENN.AAV.TBG.hTGIF2-2A-GFP (AAV.hTGIF2).

Generation of LV.hTGIF2-2A-GFP. The same fragment TGF2-2A-GFP (see above) was cloned into the lentiviral backbone pRRL.SIN.cPPT.PGK.WPRE (Addgene plasmid Cat. 12252)[406] to create the pPGK-hTGIF2-2A-GFP lentiviral expression vector (LV.hTGIF2). In this case both the lentiviral vector and the insert were digested with BamHI and AgeI enzymes before ligation.

Generation of LV.mTGIF2-2A-GFP. This vector was previously generated in the lab[71].

Generation of AAV.mPMN-H2B-GFP. The transgenic cassette mPMN-H2B-GFP (composed of mouse Pdx1, MafA and Ngn3 ORFs interspaced by 2A and linked to the H2B-GFP by another 2A peptide) was isolated from the pENTR/D-TOPO.mPMN-H2B-GFP vector by digesting with the enzyme PsiI. The pENN.AAV.TBG plasmid (Penn Vector Core, AV-8-PV0146) was cut with HincII. The pENN.AAV.TBG.PMN-H2B-GFP (AAV.mPMN) product resulted from the ligation of vector and insert with blunt ends. The correct orientation of the transgenic cassette was checked.

Generation of LV.hPBX1a-2A-mCherry. The *PBX1a* ORF fragment, excluding the STOP codon, was amplified by PCR from the CAGGS-PBX1a, plasmid using primers with restriction enzyme sites (referred to as *hPBX1a-EcoRV-F* and *hPBX1a-BamHI-R*, see Table 2). It was then cloned into the pCR-Blunt II-TOPO vector (Invitrogen) containing the 2A-mCherry sequence, using EcoRV and BamHI enzymes. The lentiviral vector pRRL.SIN.cPPT.PGK.WPRE was linearized with AgeI and SalI, instead TOPO-BluntII.hPBX1a-2A-mCherry was cut with EcoRV and HindIII to isolate the transgenic cassette hPBX1a-2A-mCherry. The Klenow fragment (see Methods section 3.3.9) was used to generate blunt ends in both the vector and the insert. The LV.PGK.hPBX1a-2A-mCherry construct (LV.hPBX1) was finally produced by ligation of the two DNA.

Generation of LV.hPDX1-2A-mCherry. The human *PDX1* ORF fragment, excluding the STOP codon, was amplified by PCR from the plasmid pcDNA3.1-PDX1 (GenScript, Cat. No: OHu19441D) using primers with restriction enzyme sites (referred to as *hPDX1-BamHI-F* and *hPDX1-BamHI-R,* see Table 2). The PCR product was directly cloned into the pRRL.SIN.cPPT.PGK.WPRE vector containing the 2A-mCherry sequence by using BamHI. The resulting lentiviral vector is LV.PGK.hPDX1-2A-mCherry (LV.hPDX1).

Generation of LV.hNKX6.1-2A-mCherry. The human *NKX6.1* ORF fragment, excluding the STOP codon, was amplified from pMXs-NKX6.1 (Addgene #32934) using primers with restriction enzyme sites (referred to as *hNKX6.1-XhoI-F* and *hNKX6.1-EcoRI-R*, see Table 2). First, it was cloned into pCR-Blunt II-TOPO vector upstream of the T2A-mCherry sequence (mCherry linked by the T2A self-cleaving sequence), using XhoI and EcoRI. The resulting TOPO-BluntII.hNKX6.1-2A-mCherry plasmid was cut with XbaI and HindIII, while the lentiviral

plasmid vector pRRL.SIN.cPPT.PGK.WPRE was linearized using AgeI and SalI. Next, the ends of both vector and insert DNA were rendered blunt by using the Kenow fragment and ligated afterwards to generate the lentiviral vector pCCL.PGK.hNKX6.1-2A-mCherry (LV.NKX6.1).

3.3.7 Polymerase chain reaction for cloning

For cloning, PCR reactions were performed on 500 ng of plasmid DNA with either proof-reading ExTaq DNA Polymerase (Takara) or DreamTaq Green DNA Polymerase (Thermofisher) in the presence of 250 nM primer mix and 0.1 mM dNTPs and relative PCR buffers (containing 2mM $MgCl_2$). The primers used for PCR for cloning are listed in Table 2. Cycling parameters are listed in Table 3.

Primer name	Sequence
hTGIF2-EcoRI-MluI-F	5'-CGG AAT TCC GAC GCG TCG ATA TGT CGG ACA GTG ATC-3'
hTGIF2-ClaI-R	5'- CCA TCG ATG GAC TTC TGG GGA TTT TCA GAG AC-3'
hPBX1a-EcoRV-F	5'-CGG ATA TCC GAT TAT GGA CGA GCA GCC CAG G-3'
hPBX1a-BamHI-R	5'-CGG GAT CCC GAG TTG GAG GTA TCA GAG TGA-3'
hPDX1-BamHI-F	5'-CGG GAT CCC GAA TAT GAA CGG CGA GGA GCA-3'
hPDX1-BamHI-R	5'-CGG GAT CCC GAA TAT CGT GGT TCC TGC GG -3'
hNKX6.1-XhoI-F	5'-CCC TCG AGG GAT TAT GTT AGC GGT GGG GGC AAT-3'
hNKX6.1-EcoRI-R	5'-GGA ATT CCA AGG ATG AGC TCT CCG GCT CGG ACG-3'

Table 2. Primers for cloning

Cloning PCR		
Temperature	**Time**	
94°C	2 min	
94°C	1 min	30 cycles
56 - 62°C	1 min	
72°C	1 min	
72°C	5 min	

Table 3. Protocols for cloning PCR

3.3.8 Ligation of PCR products and digested DNA fragments

Before ligation, digested vector fragments were dephosphorylated using rAPid Alkaline phosphatase (Roche). Ligation of DNA fragments was performed with T4 ligase (NEB) overnight at 16°C using a vector-to-insert molar ratio of 1:3.

3.3.9 DNA blunting of PCR products and vectors

Klenow Fragment was used to produce blunt ends in PCR products and linearized vectors by fill-in of 5'-overhangs or removal of 3'-overhangs. Klenow Fragment is a DNA polymerase exhibiting 5'→3' polymerase activity and 3'→5' exonuclease (proofreading) activity. The manufacturer's instructions were followed. Briefly 3 µg of DNA was processed with 5 U of Klenow Fragment enzyme (Thermo Fisher Scientific) in presence of 50 mM Tris-HCl (pH 8.0), 5 mM $MgCl_2$, 1 mM DTT, 50 nM dNTPs. After incubation at 37°C for 10 minutes, the reaction was stopped by heating at 75°C for 10 minutes. Afterward a DNA clean & concentrator kit (Zymogen) was used to purify the DNA.

3.3.10 DNA sequencing

Sequencing of cloned expression vectors to confirm correct cloning was performed by the companies Source BioScience and LGC Genomics using the Sanger sequencing method.

3.3.11 Isolation of genomic DNA from mouse tail biopsies

Adult mouse tail biopsies were lysed in 250 µL Tail lysis buffer (50 mM Tris pH 8, 100 mM EDTA pH 8, 100 mM NaCl, 1% SDS and 1.5 µL Proteinase K (Roche) shaking at 55°C for 3 hours to overnight. Next, the samples were cleared by centrifugation. The supernatant was transferred into a new reaction tube. The DNA was precipitated with 250 µL isopropanol, centrifuged, the supernatant discarded, and the DNA pellet air-dried for 10min, before resuspending in 75-100 µL TE buffer (10 mM Tris pH 7.5, 1 mM EDTA) for 1 hour at 37°C. Embryo mouse tail biopsies were lysed in 75 µL Embryo lysis buffer (100mM Tris pH 8, 50mM KCl, 2mM $MgCl_2$, 0.1 mg/mL Gelatin, 0.45% NP40, 0.45%Tween and 2.5 µL Proteinase K (Roche)) shaking at 55°C for 3 hours to overnight. The Proteinase K was then heat-inactivated shacking the lysate at 95°C for 10 minutes. The lysate was then ready for genotyping.

3.3.12 Genotyping

DNA from tail biopsies from step 3.3.8. was used as template for genotyping PCRs. The PCR reaction contained µL tail DNA, 300 – 400 nM of each oligonucleotide, and 1x DreamTaqT M mix, which contains Taq Polymerase, PCR buffer, 4 mM $mgCl_2$, 40 mM dNTPs and loading dye. The primer combinations are described in Table 4. The cycler programs are described in Table 5. Genotyping PCRs for adult or embryonic tissues were analysed by standard agarose gel electrophoresis and genotypes assigned according to band size of the PCR products (see Table 4).

3.3.13 Isolation of total RNA from cells and tissues samples

Total RNA from cells and tissues was isolated through the guanidinium-thyocianate phenol-cloroform extraction method or using High Pure RNA Isolation kit (Roche), following manufacturer instructions. For guanidinium-thyocianate phenol-cloroform extraction, 1 mL Trizol® Reagent (Qiagen, Hilden, Germany) was used to extract RNA. 200 µL chloroform were added and the samples were shaken vigorously for 15 seconds. After centrifugation for 15 minutes at 4°C, the aqueous phase was transferred to a fresh tube. The RNA was precipitated overnight at -20°C with 600 µL isopropanol. An RNA pellet was obtained after centrifugation at 4°C for 15 minutes. It was washed in 70% ice-cold ethanol DEPC-treated, then dried and dissolved in 20 µL DEPC-treated water. For RNA-Seq, RNA purification on column was performed using the RNeasy RNA Mini Kit (Qiagen), starting from the aqueous phase resulted from the first centrifugation step of the protocol above described (adjusted to half of all the volumes) and following the manufacturer's instructions.

Allele	Sequence	Band size
Pdx1-Cre internal control	5'-CTA GGC CAC AGA ATT GAA AGA TCT-3'	324 bp
	5'-GTA GGT GGA AAT TCT AGC ATC ATC C-3'	
Pdx1-Cre (TG)	5'-GCG GTC TGG CAG TAA AAA CTA TC-3'	100 bp
	5'-GTG AAA CAG CAT TGC TGT CAC TT 3'	
H2B-EGFP (WT)	5'-AAC CCC AGA TGA CTA CCT ATC CTC C-3'	217 bp
	5'-TCC CTC GTG ATC TGC AAC TCC AGT C-3'	
H2B-EGFP (TG)	5'-GGG GGA GGA TTG GGA AGA CAA TAG C-3'	297 bp
	5'-AAC CCC AGA TGA CTA CCT ATC CTC C-3'	
Tgif1 (WT)	5'-TTC CCT GCT GGT GAA AGC AA-3'	220 bp
	5'-TGT TCA TAC AGC CAG TCT CG-3'	
Tgif1 (KO)	5'-TTC CCT GCT GGT GAA AGC AA-3'	220 bp
	5'-GGC CTC TTC GAT ATT ACG CC-3'	
Tgif2 (WT or TG)	5'-CCA AAT AGC TCG TTT GTT CGG C-3'	201 bp (WT)
	5'-GCC AGG CTG TCC TGG AAA CTT AG-3'	341 bp (TG)
Tgif2 (KO)	5'-CCA AAT AGC TCG TTT GTT CGG C-3'	332 bp
	5'-CTC CTC CTC TTC CTT CTC AGG G-3'	

Table 4. Primers for genotyping PCR

Genotyping Pdx1-Cre

Temperature	Time	
94°C	2 min	
94°C	30 sec	35 cycles
56°C	30 sec	
72°C	30 sec	
72°C	2 min	

Genotyping 26RH2B-EGFP

Temperature	Time	
94°C	2 min	
94°C	30 sec	30 cycles
65°C	30 sec	
72°C	30 sec	
72°C	5 min	

Genotyping Tgif1

Temperature	Time	
94°C	2 min	
94°C	30 sec	30 cycles
56°C	30 sec	
72°C	30 sec	
72°C	2 min	

Genotyping Tgif2

Temperature	Time	
94°C	2 min	
94°C	30 sec	35 cycles
62°C	30 sec	
72°C	1 min	
72°C	7 min	
20°C	5 min	

Table 5. Protocols for genotyping PCR

3.3.14 DNA and RNA concentration determination

RNA concentration was measured spectrophotometrically using a NanoDrop (PEQLAB Biotechnologie, Erlangen, Germany). The Optical Density (OD) at 260 nm was used to calculate the concentration, while the ratios between the absorbances at 230, 260 and 280 nm were used to determine the quality of the sample.

3.3.15 Agarose gel electrophoresis of DNA

DNA fragments were separated according to their molecular weight on 1%, 1.5% or 2.5% agarose gels containing ethidium bromide in TAE buffer.

3.3.16 Reverse transcription reaction

Extracted RNA samples were first treated with DNase to remove genomic DNA. 3 µg of RNA were treated with 1-2 µL DNase Turbo (Ambion) in presence of RNAse inhibitor (Promega) and Turbo reaction buffer (Ambion) for 30 minutes at 37°C. The RNA was purified through a phenol-chloroform extraction by adding 1 volume of phenol-chloroform-isoamyl alcohol to the sample and mixing both phases by vigorous shaking. Following centrifugation, the aqueous phase was transferred to a new tube. The RNA was precipitated at -20°C overnight with 1/10 volumes 3 M sodium acetate, 2 volumes ice-cold 100% ethanol and 20 µg glycogen. The following day, the tubes were centrifuged for 15 minutes at 4°C and the pellets washed with 70% ethanol. The RNA was resuspended in 20 µL DEPC-treated water. 10 µL of the total RNA were used for reverse transcription (RT), 10 µL for a control without reverse transcriptase (-RT control). The reverse transcription was carried out either with the Transcriptor First Strand cDNA Synthesis kit (Roche). A mixture of random hexamers and oligo-dT primers was used for priming.

3.3.17 Quantitative real-time polymerase chain reaction and transcriptional analysis

Quantitative real-time polymerase chain reactions (RT-qPCR) were carried out using the SYBR Green Master Mix (Roche) on ABI StepOne Plus system. Succinate dehydrogenase (*Sdha*) or 36B4 were used as reference genes for murine material; glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) was used as reference gene for human material. Primers used for RT-qPCR are listed in Tables 6 and 7. All the values were normalized to the reference genes and calculated using the software REST[407]. 'Undetermined' data points were assigned a Ct of '40' to enable calculation of fold change. Unless stated otherwise, data are

shown as mean ± SEM and statistical significance ($p<0.05$) was determined using the REST randomization test (*)[407]. Statistical significance between groups in the *in vivo* experiments was evaluated using two-tailed Student's t-test (#).

Primer name	Sequence
36B4-F	5'-GGC CCT GCA CTC TCG CTT TC-3'
36B4-R	5'-TGC CAG GAC GCG CTT GT-3'
AFP-F	5'-CGA GGA GTG TTG CCA AGG AAA-3'
AFP-R	5'-CAG AAG CCT AGT TGG ATC ATG-3'
Albumin-F	5'-TGC TGC TGA TTT TGT TGA GG-3'
Albumin-R	5'-GCA GCA CTT TTC CAG AGT GG-3'
Cdh1-F	5'-AAC CCA AGC ACG TAT CAG GG -3'
Cdh1-R	5'-GAG TGT TGG GGG CAT CAT CA-3'
Cela1-F	5'-GAG GAA CTC TTG GCC GTC TC-3'
Cela1-R	5'-CAG TTG CTT CGG ATG AGG GT-3'
Celsr3-F	5'-CCA GGC CAA GTC ACA CTT TTG-3'
Celsr3-R	5'-TAG GGA TGG GCC ATT GTG AGT-3'
Col1a1-F	5'-CTG ACG CAT GGC CAA GAA GA-3'
Col1a1-R	5'-CTC GGG TTT CCA CGT CTC AC-3'
Cpa1-F	5'-AGC CTT GAC ATC TCT ACA CGG-3'
Cpa1-R	5'-CTG GCT GTA GGT CCA GTC AAT-3'
Fzd2-F	5'-GCA CCA TCA TGA AGC ACG AC-3'
Fzd2-R	5'-TAC CGT GTA GAG CAC CGA GA-3'
Hnf4a-F	5'-AAC CAC GCT ACT TGC CTT TGC T-3'
Hnf4a-R	5'-TCT GAT GGG ACA CAG CCT ACT TCT-3'
Iapp-F	5'-TGA TAT TGC TGC CTC GGA CC-3'
Iapp-R	5'-CGT GTT GCA CTT CCG TTT GT-3'
Ins2-F	5'-ACC CAC AAGT GGC ACA ACT G-3'
Ins2-R	5'-TAC AAT GCC ACG CTT CTG C-3'
Insm1-F	5'-GCC CAG GTG TTC CCC TGC AA-3'
Insm1-R	5'-AGG CCC GGG GAG CTG TAG AA-3'
Isl1-F	5'-TGC AAA TGG CAG CCG AAC CCA-3'
Isl1-R	5'-AGG TCC GCA AGG TGT GCA GC-3'
Itga6-F	5'-CGG TCT CCG GAG TCG CTA AGC-3'
Itga6-R	5'-TCA AGG TTG CTG TGC CGA GGT T-3'
Lama4-F	5'-GGT TCA GGA AGC TAC GGA CC-3'
Lama4-R	5'-CTG TAC CAG CCC GTT CAT GT-3'
Lamc1-F	5'-CTG CTA AGA AGG GAC GCA GT-3'
Lamc1-R	5'-TTC CGC GGC TGT CTT GTT AT-3'
Nestin-F	5'-GCT GTG GAA GCC CTG GAG CA-3'
Nestin-R	5'-TCA GCC TCC AGC AGA GTC CTG T-3'
Neurod1-F	5'-AAG GCA AGG TGT CCC GAG GC-3'

Continued on next page

Neurod1-R	5'-CAT CAG CCC GCT CTC GCT GT-3'
Ngn3-F	5'-GAG CAC TGG TGT TCT CAG ACT-3'
Ngn3-R	5'-GCT TGA GAG CCT CCA CTA CC-3'
Nkx2.2-F	5'-TCA GGG ACG GCA AAC CGT GC-3'
Nkx2.2-R	5'-GCC GAG CTG TAC TGG GCG TT-3'
Osr1-F	5'-CTC TGG TCA CTC AAG TCC AGC-3'
Osr1-R	5'-CTG GGA ACC GCA ATG ATT TCA A-3'
Osr2-F	5'-TCC GGA CCT AGT CTG AAC CG-3'
Osr2-R	5'-GAG GGT GTG AGG GGG AAA AG-3'
Pax6-F	5'-CAA CCT GCC TAT GCA ACC CCC A-3'
Pax6-R	5'-GGG CAG CAT GCA CGA GTA CGA-3'
Pdx1-F	5'-CCA CCA AAG CTC ACG CGT GGA-3'
Pdx1-R	5'-GGC GGG GCC GGG AGA TGT ATT-3'
Ptf1a-F	5'-TTC CTG AAG CAC CTT TGA CAG A-3'
Ptf1a-R	5'-ACG GAG TTT CCT GGA CAG AGT-3'
Rfx6-F	5'-TGC AGT GGC TTG AAG ACA ATT AC-3'
Rfx6-R	5'-CTT CGA GTT GTT AGG AGG GGA-3'
Sdha-F	5'-TGT TCA GTT CCA CCC CAC A-3'
Sdha-R	5'-TCT CCA CGA CAC CCT TCT GT-3'
Sox17-F	5'-GTA AAG GTG AAA GGC GAG GTG-3'
Sox17-R	5'-GTC AAC GCC TTC CAA GAC TTG-3'
Sox9-F	5'-AGA CTC ACA TCT CTC CTA ATG CT-3'
Sox9-R	5'-ACG TCG GTT TTG GGA GTG G-3'
Tgif2-F	5-CTA TCT GCA CCG CTA CAA CG-3
Tgif2-R	5-GGG CAT TGA TGA ACC AGT TAC-3
Tgif2-UTR-F	5-GGC GAG AAT GGC TAT TTG TAA-3
Tgif2-UTR-R	5-AGC CTG CTG GGT CTC CTA C-3
Ttr-F	5'-CGT TCC ATG AAT TCG CGG ATG TGG T-3'
Ttr-R	5'-GCA GGG CTG CGA TGG TGT AGT-3'
Vimentin-F	5'-GGA TCA GCT CAC CAA CGA CA-3'
Vimentin-R	5'-AAG GTC AAG ACG TGC CAG AG-3'

Table 6. Mouse-specific primers for RT-qPCR

Primer name	Sequence
ALBUMIN-F	5'-CCT GTT GCC AAA GCT CGA TG-3'
ALBUMIN-R	5'-GAA ATC TCT GGC TCA GGC GA-3'
APOA2-F	5'-GTC AAG AGC CCA GAG CTT CA-3'
APOA2-R	5'-GCT GTG TTC CAA GTT CCA CG-3'
ARX-F	5'-GGA GGC AGA AAG GCA CAA AGA-3'
ARX-R	5'-GGT GGG GTT AGA TAG CGG GTT-3'
CDH1-F	5'-CAC CAC GGG CTT GGA TTT TG-3'
CDH1-R	5'-TGG GGG CTT CAT TCA CAT CC-3'
CELSR3-F	5'-AAG TCC AAT GTG CGT GGG AT-3'

Continued on next page

CELSR3-R	5'-GCT GAT TGT GGG TCC TGT GA-3'
FOXA2-F	5'-TGC ACT CGG CTT CCA GTA TG-3'
FOXA2-R	5'-CAT GTT GCT CAC GGA GGA GT-3'
GAPDH-F	5'-AGC TCA CTG GCA TGG CCT TC-3'
GAPDH-R	5'-CGC CTG CTT CAC CAC CTT CT-3'
G6PC-F	5'-CAC TTC CGT GCC CCT GAT AA-3'
G6PC-R	5'-TAG TAT ACA CCT GCT GTG CCC-3'
HNF4a-F	5'-ACA TGG ACA TGG CCG ACT AC-3'
HNF4a-R	5'-CGT TGA GGT TGG TGC CTT CT-3'
MAFA-F	5'-GGC TTC AGC AAG GAG GAG GT-3'
MAFA-R	5'-TGG CAC TTC TCG CTC TCC AG-3'
NEUROD1-F	5'-GAC ACG AGG AAT TCG CCC AC-3'
NEUROD1-R	5'-CCC ACT CTC GCT GTA CGA TTT-3'
NGN3-F	5'-TTT TCT CCT TTG GGG CTG GG-3'
NGN3-R	5'-CTC ACG GGT CAC TTG GAC AG-3'
NKX6-1-F	5'-TGG CCT ATT CGT TGG GGA TG-3'
NKX6-1-R	5'-TGT CTC CGA GTC CTG CTT CT-3'
NKX2-2-F	5'-TTC CAG AAC CAC CGC TAC AAG-3'
NKX2-2-R	5'-GGG CGT CAC CTC CAT ACC T-3'
PAX6-F	5'-AGG ATG TTG AAC GGG CAG AC-3'
PAX6-R	5'-TCT CCC CCT CCT TCC TGT TG-3'
PCK1-F	5'-GGC CAG GAT CGA AAG CAA GA-3'
PCK1-R	5'-GGA TGA CGT ACA TGG TGC GA-3'
PDX1-F	5'-TGG AGC TGG CTG TCA TGT TGA-3'
PDX1-R	5'-CGC TTC TTG TCC TCC TCC TTT T-3'
PDX1-F2	5'-CAC ATC CCT GCC CTC CTA C-3'
PDX1-R2	5'-GAA GAG CCG GCT TCT CTA AAC-3'
PTF1A-F	5'-GTC ATC ATC TGC CAT CGG G-3'
PTF1A-R	5'-CTA GGG GAG GGA GGC CAT AA-3'
S100a4-F	5'-TCT TGG TTT GAT CCT GAC TGC T-3'
S100a4-R	5'-GGG TCA GCA GCT CCT TTA GT-3'
SERPINA1-F	5'-TAA ATA CGG ACG AGG ACA GGG-3'
SERPINA1-R	5'-ATG CCC CAC GAG ACA GAA G-3'
SOX17-F	5'-TTC GTG TGC AAG CCT GAG AT-3'
SOX17-R	5'-TAA TAT ACC GCG GAG CTG GC-3'
SOX9-F	5'-GGA GAC TTC TGA ACG AGA GCG-3'
SOX9-R	5'-GAG ATG TGC GTC TGC TCC G-3'
TGIF1-F	5'-CCG CAT CGG TGG GAA CTT-3'
TGIF1-R	5'-CTG AGC CAG CGG ATG AAG AA-3'
TGIF2-F UTR	5'-AAG CCG CTG GGA AAA GTT-3'
TGIF2-R UTR	5'-GAT CTT CAC CGA CTC CTT GG-3'
TGIF2-F (TG)	5'-GCA CCG CTA CAA CGC CTA-3'
TGIF2-R (TG)	5'-GGG CAT TGA TGA ACC AGT TAC-3'
TTR-F	5'-TTA CTG GAA GGC ACT TGG CA-3'
TTR-R	5'-CGG AGT CGT TGG CTG TGA AT-3'
VIMENTIN-F	5'-GCT AAC CAA CGA CAA AGC CC-3'
VIMENTIN-R	5'-CGT TCA AGG TCA AGA CGT GC-3'

Table 7. Human-specific primers for RT-qPCR

3.3.18 *In vitro* transcription of digoxigenin labeled RNA

For generation of mouse *Tgif2* and *Pdx1 in situ* probes, 2 µg of the containing vectors (previously generated in the lab) were linearized with SalI and transcribed using T7 polymerase (Roche) in the presence of digoxigenin-labeled UTP (Roche), DTT and RNase inhibitor. The reaction was incubated for 2 hours at 37 °C and, subsequently, treated with 1 µL of DNase I Turbo (Ambion) for further 15 minutes at 37 °C. The RNA probe was then precipitated on ice for 30 minutes using 0.1 volumes of 4M LiCl and 2.5 volumes of ice-cold 100% ethanol. An RNA pellet was obtained by centrifugation at 4 °C for 10 minutes and then washed with 70% ethanol/DEPC-treated water. The pellet was dried and resuspended in 20 – 50 µL DEPC-treated water.

3.3.19 *In situ hybridization on cryosections*

In situ hybridization (ISH) on cryosections (10 µm) was carried out as previously described[408]. Briefly, cryosections were treated with 5 µg/mL proteinase K (Sigma) for 2 minutes and post-fixed in 4% PFA. Next, the slides were acetylated with 0.0025% acetic anhydride in 0.1M triethanolamine-HCl pH 7.5, dehydrated in 70% and 95% ethanol and washed in Tris/glycine buffer. Hybridization was performed overnight at 65 °C in Hybe Mix (see below) with 1 µg/mL RNA probe. Washes were performed in 5x and 2x SSC, 0.5x SSC / 20% formamide at 60 °C, and NTE buffer. The probe was degraded by incubation with 10 µg/mL RNase A. Blocking was carried out with 1% blocking solution in MAB buffer (Roche). Hybridized cRNA probes were detected with sheep anti-DIG AP FAB antibody (Roche). Slides were washed with TBS and Tween-20/Levamisole. The colorimetric reaction with BM Purple (Roche) to detect the signal was stopped with 1mM EDTA/PBS.

Tris/glycine buffer: 0.1 M Tris, 0.1 M glycine
Hybe Mix: 40% formamide, 5x SSC, 10x Denhardt's, 100 µg/mL salmon sperm DNA, 100µg/mL torula tRNA
NTE: 0.5 M NaCl, 10 mM Tris-HCl pH 7.0, 5 mM EDTA
MAB: 100 mM maleic acid, 150 mM NaCl
TBS: 136 mM NaCl, 20 mM Tris-HCl pH 7.6
10% Tween/Levamisol: 0.1% Tween-20, 50 mg/mL levamisole

3.4 Histology methods

3.4.1 Samples embedding and cryosectioning

Tissue was fixed in 4% paraformaldehyde at 4 °C overnight. Samples were washed three times with PBS, equilibrated in 20% sucrose/PBS solution overnight and embedded in OCT compound (Sakura). Cryosections were cut with 10 µm thickness on a CM3050s cryostat (Leica).

3.4.2 Immunofluorescence on cryosections

Cryosection slides were dried at room temperature, then rehydrated in PBS and washed in PBS with 0.1% Triton X-100. Antigen retrieval was performed when needed by boiling the slides for 20 minutes in citrate buffer pH 6.1 (Target Retrieval Solution, Dako). Blocking was performed for one hour at room temperature with TSA solution (Perkin Elmer), before incubating with primary antibodies (overnight, at 4 °C) and subsequently secondary antibodies (1 h at RT). Washes were performed by applying 0.1% Triton X-100 for 5 minutes for 3 times. Primary and secondary antibody incubations were done in 3% horse serum, 0.3% BSA and 0.1% Triton X-100 solution. The antibodies and dilutions used are listed in Table 8. Alexa-conjugated secondary antibodies (Invitrogen) were used at a dilution of 1:750 (see Table 9). Hoechst 33342 (Invitrogen) counterstaining was used at a concentration of 20 µg/mL. The slides were mounted using Dako Fluorescence Mounting Medium.

3.4.3 Immunofluorescence on fixed cells

For immunofluorescence, cells were grown on glass coverslip, plastic 24-well plates or 8-well chambers (Ibidi). Cells fixed for 20 minutes with 4% paraformaldehyde at room temperature and then permeabilized for 20 minutes with 0.2% Triton X-100 before blocking. Unspecific binding was blocked using a solution constituted of 2% donkey serum, 2% BSA, 50 mM glycine in PBS for 1 h at room temperature. Incubation with primary antibody was performed overnight at 4°C in blocking solution. Samples were then washed 3 times for 10 minutes in PBS-0.1% Tween. Incubation with secondary antibody was performed at room temperature for 1 h in blocking solution and washes as above. Coverslips were mounted on slides using Dako mounting medium. The antibodies and dilutions used are listed in table 12. Alexa-conjugated secondary antibodies (Invitrogen) were used at a dilution of 1:750 (see Tables 8 and 9). Hoechst 33342 counterstaining was used at a concentration of 20 µg/mL.

3.4.4 Image acquisition and analysis

Images were acquired on Zeiss AxioObserver and Zeiss LSM 700 laser scanning microscope. For counting, pancreatic tissue of at least three controls and three mutant embryos per genotype were cut into serial sections and stained cells were counted on every other section for E12.5 embryos every third section for E14.5 and E16,5 ones. The region-of-interest (ROI) tool of ImageJ was applied to quantify fluorescence intensities. E-cadherin (E-cad)-positive pancreatic epithelium was measured using ImageJ. The number of cells positive for selected markers were counted manually. Quantification of immunohistochemical markers is expressed as mean ± SEM and significance of differences between groups was evaluated using two-tailed Student's t-test.

Antibody	Catalog number	Host species	Dilution
anti-ALBUMIN	DakoCytomation, A0001	Rabbit	1:600 (1:200 cells)
anti-ALBUMIN	Bethyl, A80-129A	Goat	1:600
anti-CPA1	R&D, AF2765	Goat	1:1000
anti-E-CADHERIN	Sigma, U3254	Rat	1:1000
anti-α-FETOPROTEIN	DakoCytomation, A000829-2	Rabbit	1:600
anti-GFP	Aves, GFP-1020	Chicken	1:500
anti-GLUCAGON	Immunostar, 20076	Rabbit	1:500
anti-HNF4A	Abcam, ab41898	Mouse	1:500 (1:200 cells)
anti-INSULIN	Invitogen, PAI-26938	Guinea pig	1:400
anti-ISLET 1	Hybridoma Bank, 39.4D5	Mouse	1:100
anti-MAFA	Abcam, ab26405	Rabbit	1:500
anti-NKX6-1	Hybridoma Bank, F55A10	Mouse	1:500
anti-NEUROD1	Abcam, ab213725	Rabbit	1:800
anti-NGN3	BCBC, AB2774	Goat	1:1200
anti-PAX6	Covance, PRB-278P	Rabbit	1:500
anti-PDX1	Abcam, ab47308	Guinea pig	1:1000 (1:100 cells)
anti-pH3	Millipore, 06-570	Rabbit	1:500
anti-PROX1	RELIATech GmbH, 102-PA32S	Rabbit	1:500
anti-PTF1A	C. Wright Lab	Goat	1:2000
anti-SOX9	Millipore, AB5535	Rabbit	1:1000 (1:200 cells)
anti-TGIF2	Santa Cruz, sc-390870 X	Mouse	1:300 cells
anti-VIMENTIN	Cell Signaling, 5741S	Rabbit	1:100 cells

Table 8. Primary antibodies used for Immunofluorescence

Antibody	Catalog number	Host species	Dilution
Alexa Fluor 488-labeled Anti-Chicken IgG	Invitrogen, A11039	Goat	1:750
Alexa Fluor 488-labeled Anti-Goat IgG	Invitrogen, A11055	Donkey	1:750
Alexa Fluor 594-labeled Anti-Goat IgG	Invitrogen, A11058	Donkey	1:750
Alexa Fluor 488-labeled Anti-Guinea Pig IgG	Dianova, 706-545-148	Donkey	1:750
Alexa Fluor 594-labeled Anti-Guinea Pig IgG	Invitrogen, A11058	Goat	1:750
Alexa Fluor 647-labeled Anti-Guinea Pig IgG	Invitrogen, A21450	Goat	1:750
Alexa Fluor 647-labeled Anti-Guinea Pig IgG	Dianova, 706-605-148	Donkey	1:750
Alexa Fluor 594-labeled Anti-Mouse IgG	Invitrogen, A21203	Donkey	1:750
Alexa Fluor 647-labeled Anti-Mouse IgG	Invitrogen, A31571	Donkey	1:750
Alexa Fluor 488-labeled Anti-Rabbit IgG	Invitrogen, A21206	Donkey	1:750
Alexa Fluor 594-labeled Anti-Rabbit IgG	Invitrogen, A21207	Donkey	1:750
Alexa Fluor 647-labeled Anti-Rabbit IgG	Invitrogen, A31573	Donkey	1:750
Alexa Fluor 488-labeled Anti-Mouse IgG	Invitrogen, A21208	Donkey	1:750
Alexa Fluor 594-labeled Anti-Mouse IgG	Invitrogen, A21209	Donkey	1:750
Alexa Fluor 647-labeled Anti-Mouse IgG	Dianova, 712-605-153	Donkey	1:750

Table 9. Secondary antibodies used for Immunofluorescence

3.5 Bioinformatics tools

3.5.1 Total RNA Sequencing and bioinformatic analysis

Total RNA-seq was performed using 100 ng of Total RNA quantified using a Nanodrop. The quality of RNA samples prior to library preparation was determined using an Agilent Bioanalyzer and only samples with RIN (RNA integrity number) scores > 8.5 were further processed. Library preparation, sequencing and bioinformatic analysis were performed by Star Seq company, Mainz, Germany. Briefly, Illumina sequencing libraries from stranded cDNA were sequenced, pair-ended, on NextSeq 500TM. Subsequently, the data sets had been mapped against mouse mm10 mouse genome with Tophat (2.1.0) and analysed with Cufflinks (2.2.1) for expression level measurement. Heatmaps illustrating differential gene expression of selected genes were generated by using R package gplots (http://www.R-project.org/).

3.5.2 Gene Ontology analysis

Gene Ontology (GO) analysis was performed using DAVID online tool (https://david.ncifcrf.gov/home.jsp)[409].

4. RESULTS

4.1 *In vivo* study of TGIF2 function in the developing pancreas

Previous studies in our lab. identified the TALE homeoprotein TGIF2 as an endodermal modifier in Xenopus laevis[63] and showed a conserved similar role for it within the foregut endoderm of the mouse embryo at the time of pancreatic and hepatic specification[71].

Specifically, *Tgif2* expression is widespread in the mouse embryo during gastrulation (E6.5-E7.5) and, subsequently, marks the posterior foregut endoderm at E8.5, as shown by RNA-seq analysis performed in our lab. on PROX1$^+$ dissected embryonic endodermal tissue[70] and *in situ* hybridization (ISH)[71]. *Tgif2* transcript becomes then differentially expressed in endoderm-derivatives, being progressively enriched in the pancreatic domain and downregulated in the hepatic one[70,71]. After the establishment of the pancreatic fate, *Tgif2* expression is detected throughout both dorsal and ventral pancreatic buds[71] (Fig.11A), being enriched at the epithelial tips. In the adult, its expression is confined to the pancreatic islet cells[71]. To study the early role of TGIF2 in the foregut endoderm, *Tgif2* floxed mouse transgenic line was inter-crossed with the *Sox2-Cre* transgenic strain to conditionally ablate the gene in all epiblast derivatives from gastrulation onwards in a *Tgif1*-deficient genetic background[309]. Indeed, functional redundancy between *Tgif1* and *Tgif2* has been previously reported in other studies, whereby double homozygous mutants are embryonically lethal due to a requirement of both the TALE homeoproteins for gastrulation and neural development[305,410]. Compound mutant embryos with only one functional *Tgif* allele were analysed and they displayed a reduced SOX17$^+$/PDX1$^+$ ventral pancreatic bud and an expansion of the PROX1$^+$ hepatic bud volume compared to littermate controls[71]. This result indicated a role of TGIF2 in the establishment of the pancreatic identity at the expenses of hepatic fate. Besides this initial observation, the *in vivo* role of TGIF factors in the mouse pancreatic development has remained unexplored.

During my Ph.D. studies, I undertook an *in vivo* loss-of-function approach based on the Cre/LoxP recombination system, which allows for tissue- and temporal-specific deletion of loxP flanked genes[411], to investigate TGIF2 function after pancreatic fate specification. *Tgif2* floxed mice[71] were bred with a mouse transgenic strain carrying the Cre-recombinase under the constitutive control of the *Pdx1* promoter, which is active in all pancreatic lineages from E9.5 onwards[396] (Fig. 11B-C). Lineage tracing analysis employing the *Pdx1-Cre* transgenic line and the *R26-H2B-GFP* reporter line showed that the recombinase activity in my experimental setting is active in approximately 80% (78,7% ±10,1) of endogenous PDX1$^+$ cells in

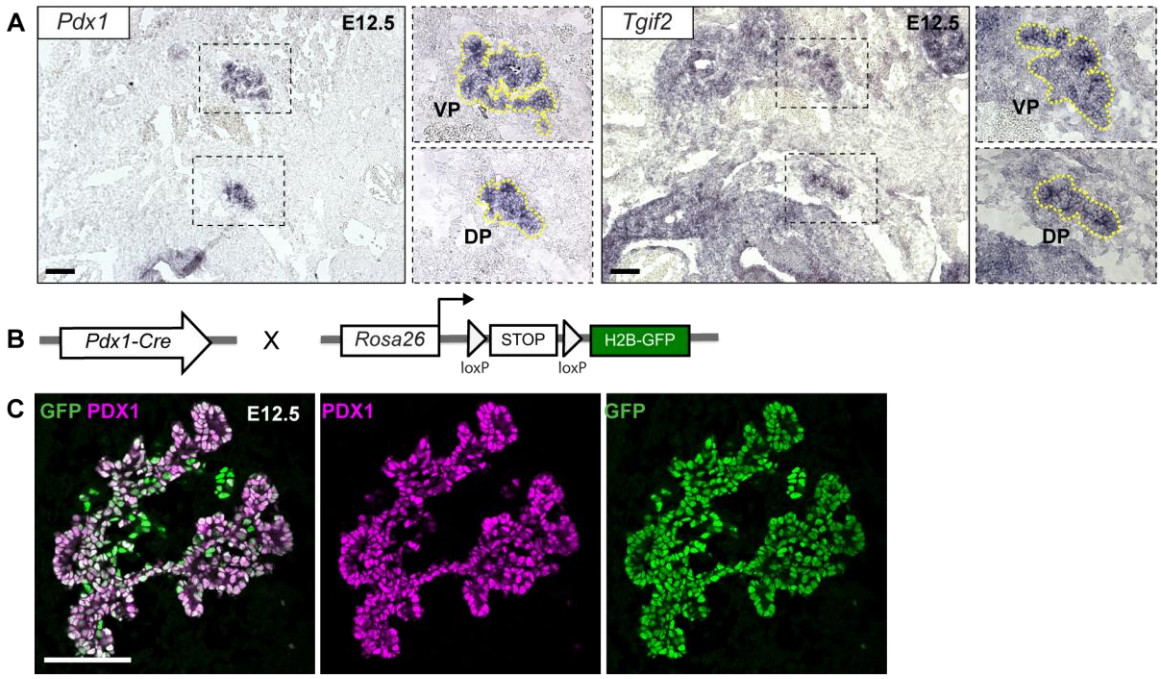

Figure 11. Expression domain of *Tgif2* and *Pdx1-Cre* recombinase activity in the embryonic pancreas. **(A)** In situ hybridization showing *Pdx1* and *Tgif2 expression* in E12.5 pancreata. The dotted boxes indicate higher magnification views of the ventral (VP) and dorsal (DP) pancreatic buds, demarcated by yellow dotted lines. **(B)** Schematic showing the lineage tracing strategy used to detect the activity of the Cre recombinase. The *Pdx1-Cre* driver line has been crossed with the *R26-H2B-GFP* (Abe et al., 2011) reporter line to excise the floxed STOP codon and allow expression of the GFP fluorescent protein. **(C)** IF staining for GFP (green) and PDX1 (purple) shows the efficiency of Cre recombinase in the pancreatic epithelium of E12.5 Pdx1-Cre; *R26-H2B-GFP* transgenic embryos. Of note, the GFP reporter overlaps with endogenous PDX1 expression domain and also marks delaminating cells, which have reduced levels of PDX1 protein at this stage. Scale bars, 100 µm.

E12.5 embryos (Fig. 11C). Exposure to the Cre recombinase results in the excision of the second exon of the *Tgif2* gene, inducing a frameshift mutation and triggering nonsense-mediated decay of the mutant transcript[71] (Fig. 12A).

To avoid functional redundancy with the close family member *Tgif1*, I examined embryos deficient for *Tgif2* specifically in the pancreatic epithelium in a *Tgif1* null background (*Tgif2$^{fl/fl}$; Tgif1$^{-/-}$; Pdx1-Cre),* referred to as T1/T2 cdKO (*Tgif1/Tgif2* compound double knock-out) (Fig. 12A-B).

In order to analyse the consequences of *Tgif2* deletion in the pancreatic epithelium and gain a deep insight into cell plasticity and differentiation during early pancreatogenesis, I performed bulk RNA-seq analysis at E12.5, when the allocation of the different pancreatic cell lineages starts (Fig. 12B-D). RNA was extracted from pancreatic buds that were manually dissected from control and T1/T2 cdKO mutant embryo littermates. Two dorsal pancreata were pooled per genotype to generate the transcriptomes.

A similar number of high-quality raw reads was obtained from each sample and used to estimate the relative abundance of transcripts. This was calculated by estimating the fragments

per kilobase of exon per million mapped fragments (FPKM). To create a list of differentially expressed genes, I considered transcripts with a log2 fold change greater than 0.5 or less than 0.5 in gene expression between mutant and controls. Based on these criteria, 418 genes were found upregulated in the mutant pancreata compared to controls, whereas 451 were found downregulated.

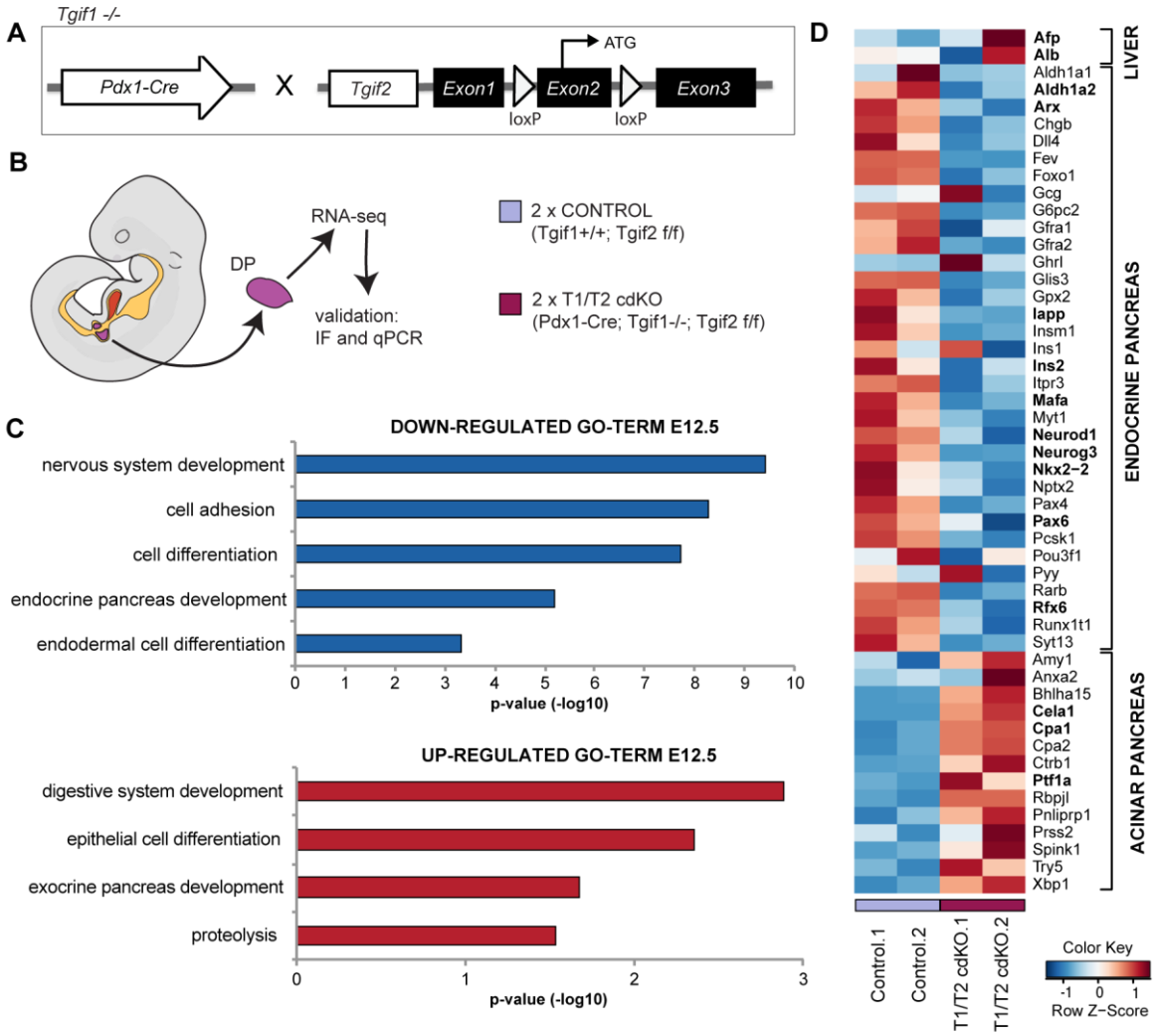

Figure 12. Study of TGIF2 roles in the mouse pancreas by a transcriptome approach.
(A) Schematic representation of the transgenes used in *Tgif2* loss-of-function approach. **(B)** Schematic showing the experimental setting used for RNA-seq analysis. Pancreatic buds were manually dissected from E12.5 control (Tgif1$^{+/+}$; Tgif2$^{f/f}$) and T1/T2 cdKO compound mutant (Pdx1-cre; Tgif1$^{-/-}$; Tgif2$^{f/f}$) embryos and processed to obtain RNA. Subsequently, RNA-seq results were validated by RT-qPCR or IF approaches. Duplicates for each genotype were analysed. **(C)** GO-term analysis of downregulated and upregulated gene categories based on the RNA-seq analysis. **(D)** Heatmap illustrating the relative expression levels of a sub-set of selected genes from the RNA-seq dataset. Colours represent high (red) or low (blue) expression values based on Z-score normalized FPKM values for each gene. White represents the average between red (high) and blue (low) expression values. Genes from liver, endocrine pancreas and acinar pancreas categories validated by either RT-qPCR or IF are in bold. Dorsal pancreas, DP.

Gene Ontology (GO) analysis of biological processes (BP) showed significant enrichment of categories related to nervous system development and endocrine pancreas development, sharing genes, such as *Neurog3, Myt1, Fev, Arx, Insm1 and NeuroD1,* among the top downregulated ones in mutant pancreata compared to controls (Fig. 12C-D). By contrast, terms associated with exocrine pancreas development (*e.g. Xbp1, Ptf1a, Cela1*) and digestive system development (e.g. *Barx1, Nkx3-2*) were induced in T1/T2 cdKO pancreata (Fig. 12C-D). Most strikingly, I found that *Tgif2* ablation results in the aberrant induction of liver hallmark genes, such as *Albumin* and *Alpha-fetoprotein* (AFP) in the pancreatic epithelium (Fig. 12D).

I then focused my analysis on all these cell fate specification and differentiation defects by using immunofluorescence and molecular approaches at different embryonic stages.

4.1.1 Genetic deletion of *Tgif2* in the pancreas results in the aberrant activation of liver genes

The ectopic presence of liver markers, such as AFP and ALBUMIN, in the pancreatic epithelium of *Tgif* mutant embryos was confirmed by immunofluorescence (IF) techniques at E12.5 and persisted at later embryonic stages, E14.5 and E16.5 (Fig. 13A).

In order to define the different contributions of the two closely related TALE homeoproteins, TGIF1 and TGIF2, to the observed phenotype, I included single mutant embryos in my analysis. Specifically, *Tgif1*$^{-/-}$*; Tgif2*$^{f/f}$ (referred to as T1 KO) and *Pdx1-Cre; Tgif1*$^{+/+}$*; Tgif2*$^{f/f}$ (referred to as T2 KO) mutant embryos were compared for defects in cell identity and cell type specification. IF analysis of liver markers showed that the loss of *Tgif2* activity alone in the pancreas is sufficient to perturb a metastable pancreatic identity and to induce the alternative hepatic identity (Fig. 13B). This effect appears to be amplified by *Tgif1* deficiency in the T1/T2 cdKO embryos (Fig. 13B).

Next, I assessed the nature of the cells expressing hepatic markers by performing double-IF stainings using cell type-specific pancreatic markers, including PDX1, SOX9, CPA1, INSULIN and GLUCAGON (Fig.13, 14). Interestingly, cells that acquired the expression of hepatic markers were found in PROX1^{+} cell clusters, containing delaminating endocrine precursor cells, at E12.5 and E14.5 (Fig.13A). At later stages, ALBUMIN^{+} cells were present exclusively in the ductal (SOX9^{+}) and acinar (CPA1^{+}) compartments, mostly at the edges of the pancreatic tissue, but absent in INSULIN^{+} or GLUCAGON^{+} cellular clusters (Fig. 14). Overall, co-expression of ALBUMIN and PDX1, SOX9 or CPA1 suggests that these cells may represent hepato-pancreatic hybrid cellular states in *Tgif2*-deficient embryos.

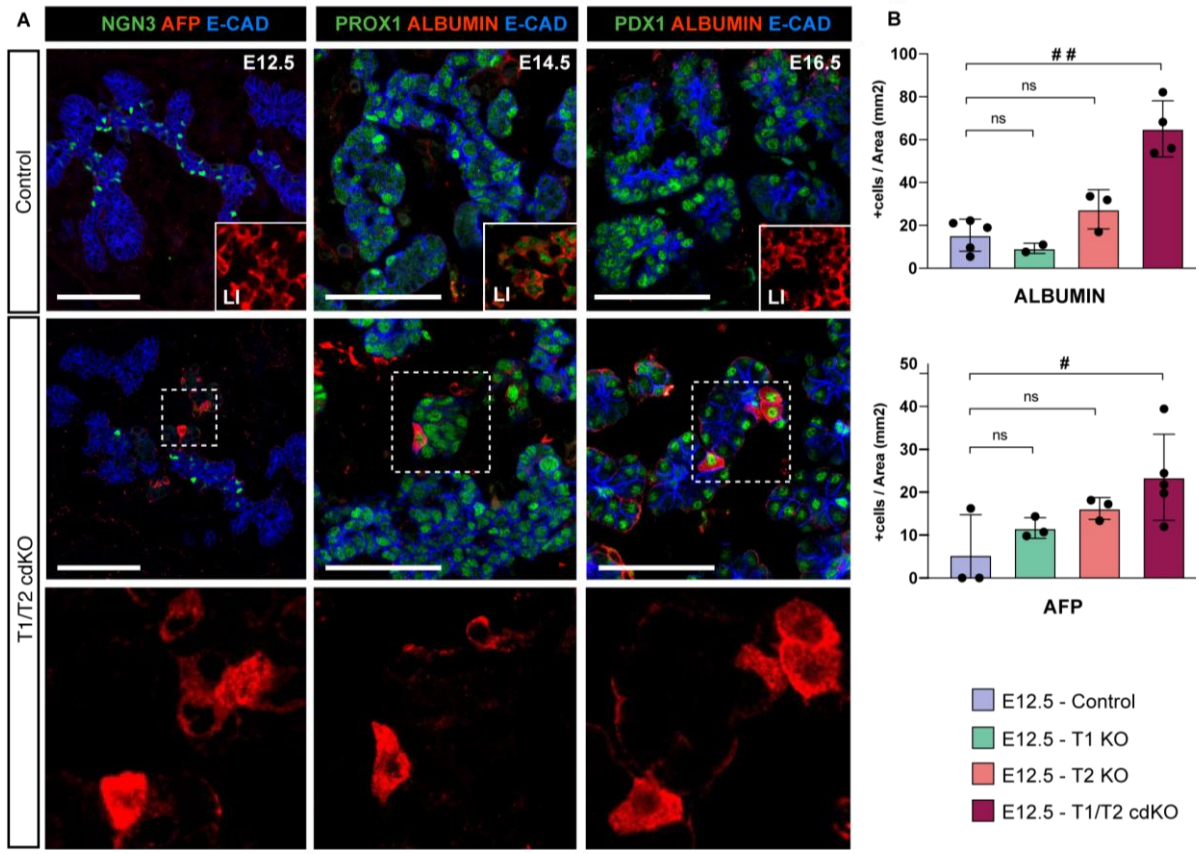

Figure 13. *Tgif2* genetic deletion results in the activation of liver genes in the developing pancreas.
(A) IF analysis on cryosections showing cells expressing hepatic markers in the pancreatic epithelium at different embryonic stages (E12.5, E14.5 and E16.5). Inset panels show expression of the same hepatic markers in the liver. Bottom panels show ALBUMIN- or AFP-single channel images of the area in the dashed box. **(B)** Quantification of ALBUMIN$^+$ and AFP$^+$ cells based on IF of E12.5 embryos. Controls, single mutants (T1 KO and T2 KO) and compound mutants (T1/T2 cdKO) were analysed. Number of cells positive for liver markers was normalised to the sum of the E-cadherin (E-CAD)$^+$ epithelium area (μm^2). Error bars represent +/- SEM. Two-tailed Student's T-test (#) P<0.05; (##) P<0.01; ns, not significant; Liver, LI. Scale bars, 100 µm.

Notably, the examined pancreatic epithelia did not express marker genes of other endodermal lineages, such as stomach-specific SOX2 or intestinal-specific CDX2 (data not shown).

Taken together, these observations are consistent with previous *in vitro* gain-of-function studies[71] in the mouse and indicate that TGIF2 not only controls pancreatic fate allocation but also preserves pancreatic identity, once established in the mouse embryo. In the absence of TGIF2 a subset of pancreatic cells undergoes a fate switch acquiring features of the closely developmentally-related hepatic lineage.

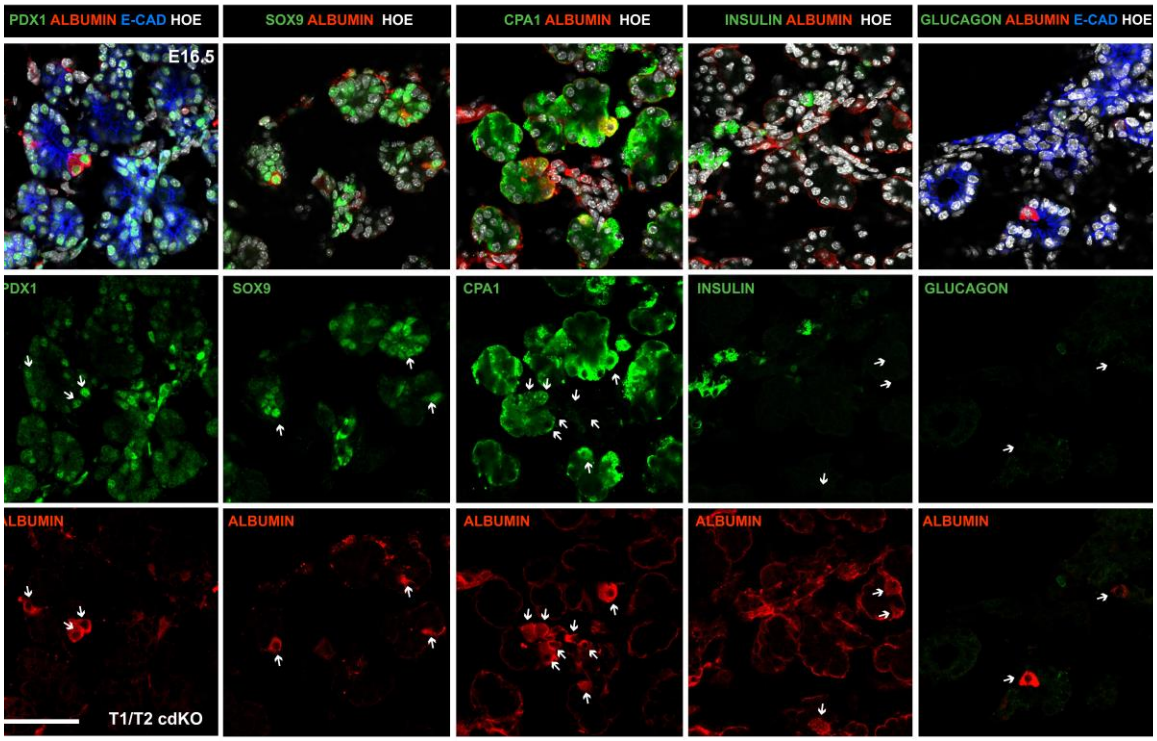

Figure 14. Characterization of ALBUMIN$^+$ cells in *Tgif2*-deficient pancreatic epithelia at E16.5. Co-immunostaining of the hepatic marker ALBUMIN and major pancreatic subtypes markers, including PDX1, SOX9, CPA1, INSULIN and GLUCAGON in E16.5 mutant (T1/T2 cdKO) pancreata. Middle and bottom panels show single channel images of indicated markers. Hoechst (HOE) nuclear counterstain in grey. Arrows indicate ALBUMIN$^+$ cells. Scale bars, 100 µm.

4.1.2 TGIF2 controls the differentiation of the pancreatic endocrine lineage

The transcriptome analysis performed on control and mutant T1/T2 cdKO pancreata revealed differences in the endocrine and exocrine cell lineages differentiation. In particular, pancreata from mutant embryos were characterized by significant downregulation of a wide set of genes regulating the transcriptional cascade driving endocrine cell differentiation (Fig. 12C-D). This includes *Ngn3*, which is required for the commitment of pancreatic multipotent progenitor cells along the endocrine lineage[120], as well as of the TFs *Insm1, Rfx6, Nkx2-2, Pax4, Pax6, NeuroD1, Glis3* (Fig. 12D). I validated this transcriptional signature by assessing changes in the expression of relevant genes involved in endocrinogenesis by RT-qPCR (Fig. 15C). Consistently with the transcriptome data, IF analysis showed robust reduction of NGN3$^+$ progenitor cells in the absence of *Tgif2* at E12.5 and E14.5 (Fig. 13A; 15B, D-E). PAX6, another key transcriptional factor marking endocrine cells, which is a direct target of NGN3, and later required for differentiation of α- or β-cells, was also downregulated in the T1/T2 cdKO mutants (Fig. 15A, C). Furthermore, the striking downregulation of MAFA indicates a more specific impairment of the endocrine β-cell lineage differentiation (Fig. 15D, E).

The analysis of the single mutants suggests that the effect on the endocrine differentiation is due to a synergistic activity of TGIF2 with TGIF1, since the simultaneous absence of both factors leads to a more pronounced endocrine phenotype (Fig. 15B, C).

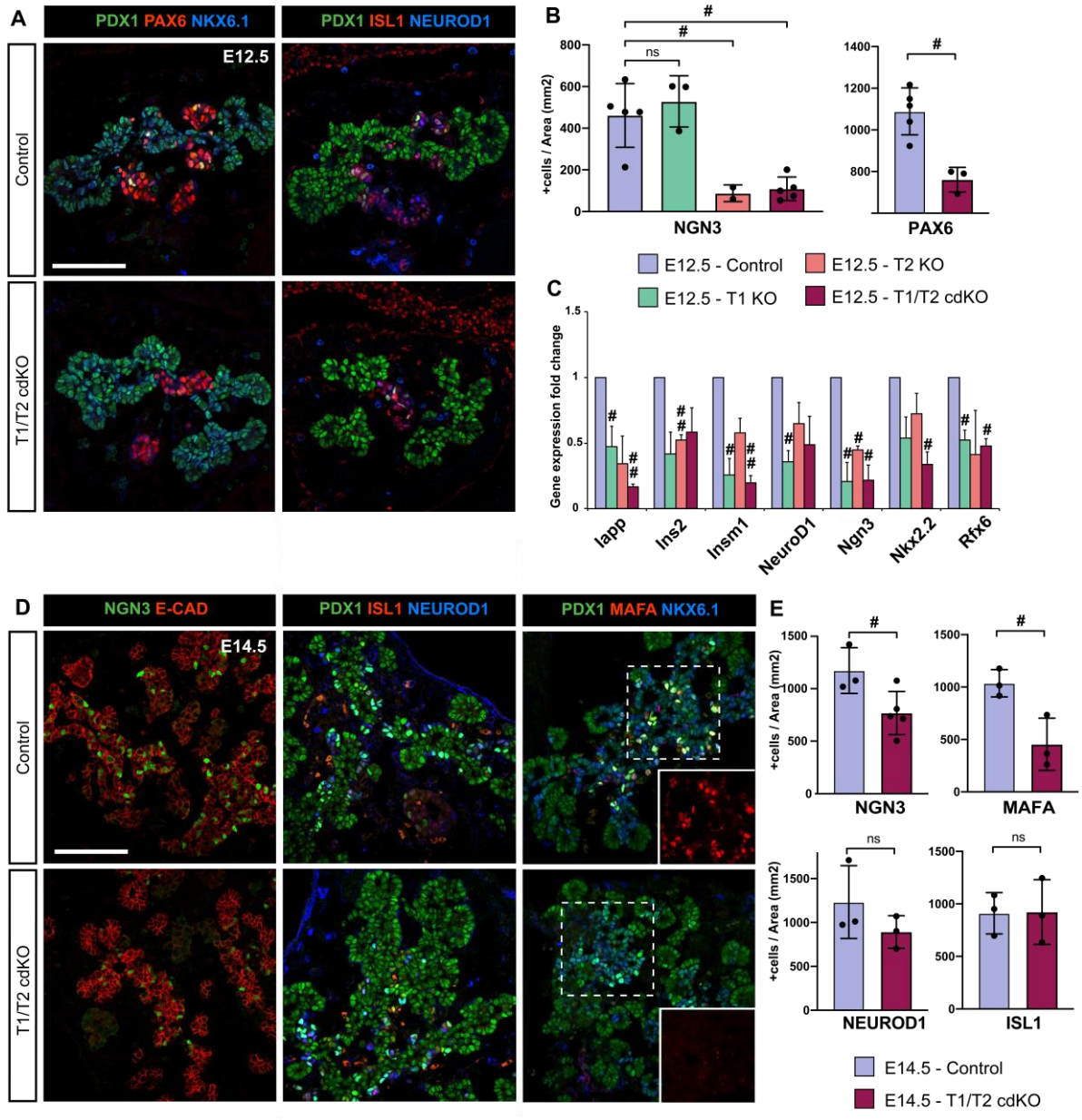

Figure 15. Deletion of Tgif(s) results in a reduced amount of pancreatic endocrine progenitors.
(A) IF pictures of dorsal pancreata and relative quantification **(B)** of cells positive for endocrine progenitor markers, including NGN3 and PAX6, of E12.5 control, single (T1 KO and T2 KO) and compound (T1/T2 cdKO) mutant embryos. **(C)** RT-qPCR analysis of endocrine markers in control and mutant E12.5 pancreata, confirming the differential gene expression observed in the RNA-seq analysis. RT-qPCR data were normalized to *36B4* and represented as fold change expression compared to control (set to 1 as calibrator) (n=3 for all genotypes). **(D)** Representative micrographs of dorsal pancreata and relative quantitative analysis **(E)** of the number of cells expressing NGN3, MAFA, NEUROD1 and ISL1 based on IF stainings of E14.5 embryos. Inset panels depict MAFA single channel images of the areas in the dashed boxes. Number of cells positive for endocrine markers was normalised to the sum of the E-cadherin (E-CAD)$^{+}$ epithelium area (µm^2). Error bars represent +/- SEM. Two-tailed Student's T-test. (#) P<0.05; (##) P<0.001. ns, not significant. Scale bars, 100 µm.

To investigate if such a reduced number of endocrine progenitor pool subsequently gives rise to a reduced number of differentiating endocrine cells, I performed IF analysis to detect INSULIN$^+$ and GLUCAGON$^+$ cells. The number of INSULIN+ β-cells was lower in T1/T2 cdKOs compared to controls at E14.5 and E16.5 stages, while the amount of GLUCAGON$^+$ α-cells were slightly decreased only at E16.5 (Fig. 16A, B). Notably, in mutant embryos the majority of the GLUCAGON-expressing cells were organized in cords instead of forming small groups of delaminating endocrine cells (Fig. 16A). This phenotype suggests a morphogenetic defect that occurs in *Tgif(s)* deficiency, whereby committed endocrine cells might fail to properly detach from the trunk epithelium and remain connected to it.

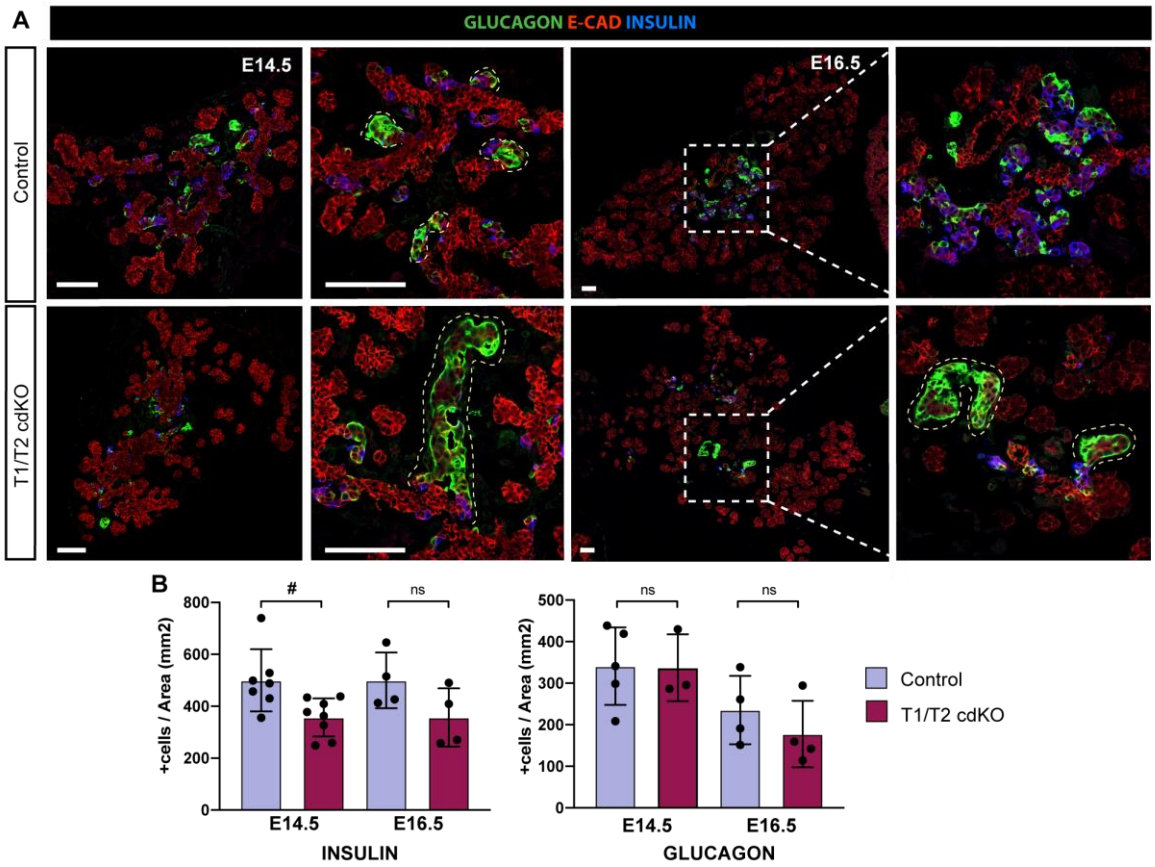

Figure 16. Analysis of INSULIN$^+$ and GLUCAGON$^+$ cells in *Tgif*-deficient embryos.
(A) IF analysis of INSULIN and GLUCAGON hormones on cryosections of E14.5 and E16.5 pancreatic tissues from controls and T1/T2 mutants. Tgif2 deletion results in morphogenetic defects of the endocrine clusters, as indicated by yellow dotted lines. Panels on the right show insets of the dashed lines boxes. **(B)** Quantification of cells expressing INSULIN and GLUCAGON based on IF stainings performed on E14.5 and E16.5 embryos, normalised to the sum of the E-cadherin (E-CAD)$^+$ epithelium area (μm^2). Error bars represent +/- SEM. Two-tailed Student's T-test. (#) P<0.05. ns, not significant. Scale bars, 100 μm

4.1.3 Genetic deletion of *Tgif(s)* results in the expansion of the acinar compartment

Oppositely to the endocrine differentiation markers, genes associated with the acinar lineage differentiation were found to be upregulated in *Tgif*-deficient embryos as compared to controls in the RNA-seq analysis (Fig. 12C-D). This observation was validated by independent approaches, including IF and RT-qPCR analyses on pancreatic tissue from control and T1/T2 cdKO embryos (Fig. 17). In particular, I found that the number of cells expressing PTF1A, which is confined at the tips of the pancreatic epithelium by E12.5 developmental stage, is higher in the compound mutants compared to littermate controls. No relevant defect was detected in the absence of the *Tgif* factors singularly, suggesting a combinatorial activity of the two TALE homeoproteins in the acinar lineage (Fig. 17B, C). Increased expression of the acinar markers CPA1 and CELA1 was observed also at E14.5 and E16.5 (Fig. 17D-F). In addition, CPA1 fluorescent intensity was markedly stronger in the examined mutant pancreatic cells, which indicate an increase also at the protein levels (Fig. 17E).

To assess whether defects in endocrine and acinar differentiation are due to changes in cell death and/or proliferation in the *Tgif* mutants, I performed IF staining for CASPASE3, a hallmark of apoptosis, and the phosphorylated form of the histone H3 (pHH3), a well-known marker of mitosis. Only very few apoptotic events were detected in both control and mutant pancreata at E14.5, with no differences between the two genotypes, ruling out the possibility of increased cell death in the endocrine compartment (Fig. 18A). Moreover, depletion of *Tgif2* in the pancreas did not alter the total number of $pHH3^+$ cells or pHH3/CPA1-double positive cells (Fig. 18A-C), ruling out the possibility of an over-proliferation of the acinar compartment in the *Tgif* mutants. Consistently, the size of the pancreata was unchanged between the controls and mutants (Fig. 18D).

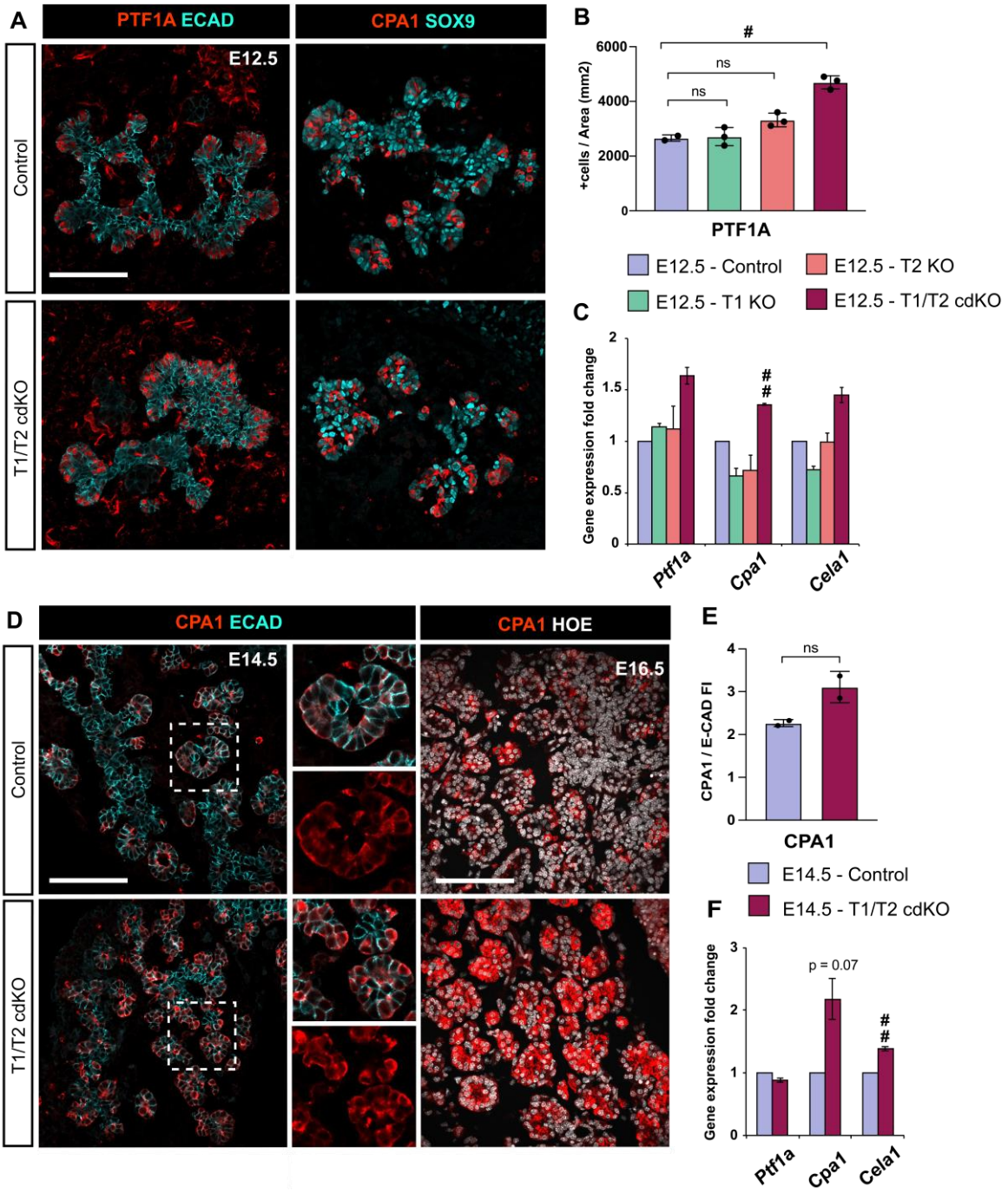

Figure 17. *Tgif(s)* deficiency results in higher levels of expression of pancreatic acinar markers. **(A-C)** IF analysis of acinar markers in E12.5 dorsal pancreata of control and mutant embryos. **(B)** Quantitative analysis of PTF1A^{+} cells based on IF stainings at E12.5. **(C)** RT-qPCR analysis of acinar markers at E12.5 **(D)** IF analysis of acinar markers at E14.5 and E16.5. Pictures of E16.5 embryos are maximum intensity projections of 7.5 µm thick z-stacks **(E)** Measurement of fluorescence intensity (FI) of CPA1 normalized to the FI of E-cadherin (E-CAD) on the stainings performed at E14.5. **(F)** RT-qPCR of acinar markers at E14.5. RT-qPCR data shown here are normalized to *36B4* and represented as fold changes compared to the control embryos (set to 1 as calibrator) (n=3). Number of cells positive for acinar markers was normalised to the sum of the E-CAD^{+} epithelium area (µm^{2}). Error bars represent +/- SEM. Two-tailed Student's T-test (##) P<0.001. ns, not significant; p, p-value. Scale bars, 100 µm.

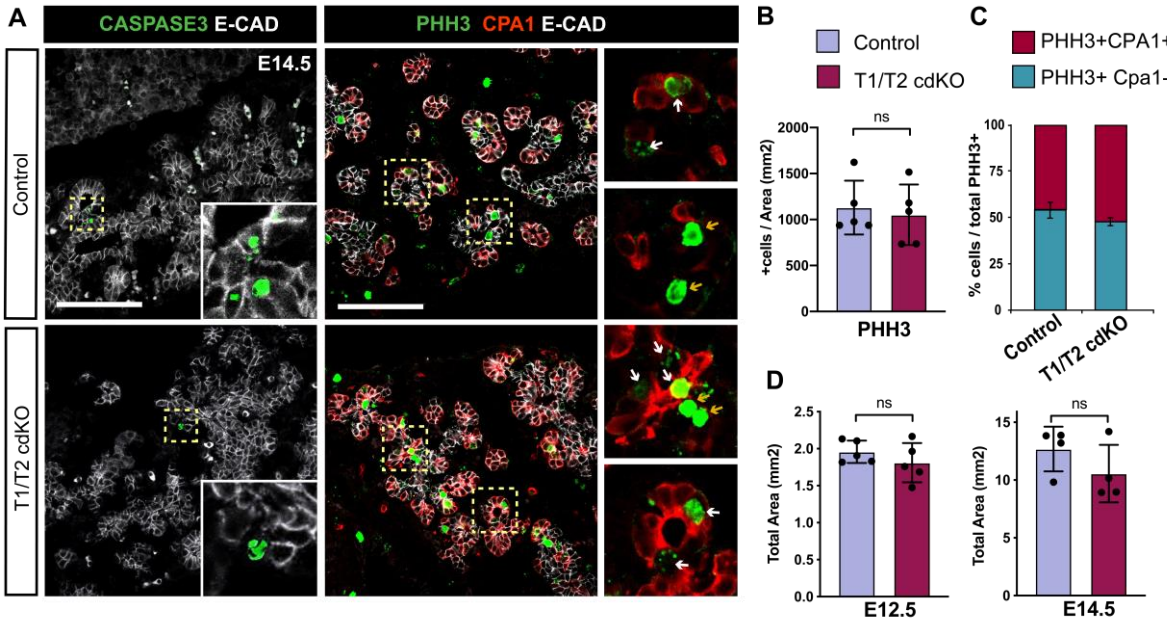

Figure 18. *Tgif2* deletion in the pancreas does not cause proliferation defects or cell death.
(A) IF analysis for the apoptosis marker CASPASE3 (green) and the proliferation marker PHH3 (green) in the pancreatic epithelium, as marked by E-CAD (white). Inset show magnified pictures of the areas in the yellow dashed boxes, with (left) or without (right) E-CAD. Arrows indicate co-expression of PHH3 and CPA1 in pancreatic cells. **(B)** Quantitative analysis of PHH3$^+$ proliferative cells based on the IF stainings of control and T1/T2 cdKO embryos. **(C)** Total area measurements of control and mutant pancreata at E12.5 and E14.5. **(D)** Quantification of the pancreatic E-CAD$^+$ epithelium area (µm^2) at E12.5 and E14.5. Number of cells positive for PHH3 was normalised to the sum of the E-CAD$^+$ epithelium area (µm^2). Error bars represent +/- SEM. Two-tailed Student's T-test; ns, not significant. Scale bar, 100 µm.

4.1.4 Signalling pathways perturbed upon pancreatic *Tgif2*-deletion *in vivo*

Late stage *Tgif*-deficient mutant embryonic pancreata displayed morphogenetic defects, such as abnormally large cystic ductal structures (Fig. 19), which are generally associated with dysregulation of cell polarity and primary cilia formation[412].

To better understand global genetic and morphogenetic alterations occurring in the absence of *Tgif2*, I focused on changes in signalling pathways, which are known to play a role in cell polarity and morphogenesis. First, I investigated the planar cell polarity (PCP), which is an evolutionarily conserved pathway that defines cellular polarization in the plane perpendicular to the apico-basal axis[413]. This pathway has an important role in vertebrate morphogenesis by controlling processes such as oriented cell division, cell migration, cell differentiation, orientation of cytoskeletal components and positioning of cell extensions, such as cilia and axons[413]. I found that in the RNA-seq dataset many genes involved in the PCP pathway (e.g. *Celsr2, Celsr3, Dvl1, Fat2, Fat4, Gpc4, Vangl2*) were less abundant in the mutant samples (Fig. 20A). By contrast, the non-canonical Wnt ligand *Wnt5a* was induced in the T1/T2 cdKO (Fig. 20A), probably reflecting an attempt of compensation by the surrounding pancreatic

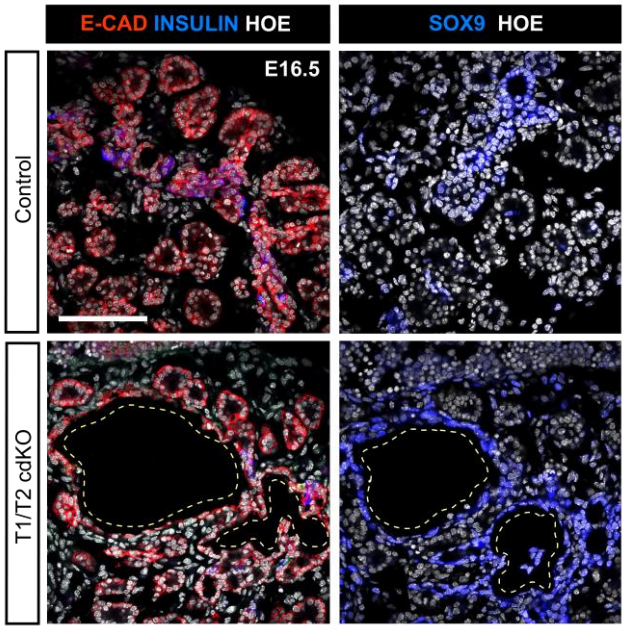

Figure 19. Genetic deletion of *Tgif2* in the pancreas leads to the formation of cystic ducts.
Large cystic ducts, marked by E-cadherin (E-CAD, red, left) and SOX9 (blue, right) are observed in the pancreatic tissue of E16.5 *Tgif2*-deficient embryos. Scale bar, 100 μm

mesenchymal cells to sustain the signalling. These data agree with previous findings pointing at PCP molecules as targets of TGIF2-mediated transcriptional regulation and acting as novel players in the liver *versus* pancreas cell fate decision[70,71]. Interestingly, independent *in vivo* studies in the mouse showed that CELSR2 and CELSR3 receptors promote endocrine differentiation during pancreas development[414].

Other biological processes and pathways were also examined (Fig. 20B). Some ECM components (e.g. *Col4a1, Col7a1, Fn, Lama4, Lamc1*) were downregulated as well as epithelial cell adhesion molecules, which are located at the cell surface and mediate ECM- and cell-cell interactions. These include Ca^{2+}-dependent CAMs (e.g. *Epcam, Ncam1, Vcam1*) and integrins (e.g. *Itga4, Itga6, Itgam*) transmembrane receptors. RT-qPCR analysis in Fig. 20B shows validation of gene expression changes of selected markers from the abovementioned categories in mutant *versus* control pancreata.

Additionally, the axon guidance Slit/Robo pathway resulted to be differentially regulated in the RNA-seq analysis. The expression of *Robo1* and *Robo2* genes, which recent findings from our lab. reported as regulators of pancreatic progenitor cell identity[73], increased in cdKO embryos, while the expression of *Slit* ligands was inconsistent among the different samples (Fig. 20B). Semaphorin genes (e.g. *Sema3a, Sema3g*), which encode molecules secreted by the peripheral mesenchyme, their Neuropilin receptors, *Nrp1* and *Nrp2*, and their Plexin co-receptor *Plxna4*, which are expressed by delaminating endocrine clusters, were all reduced in

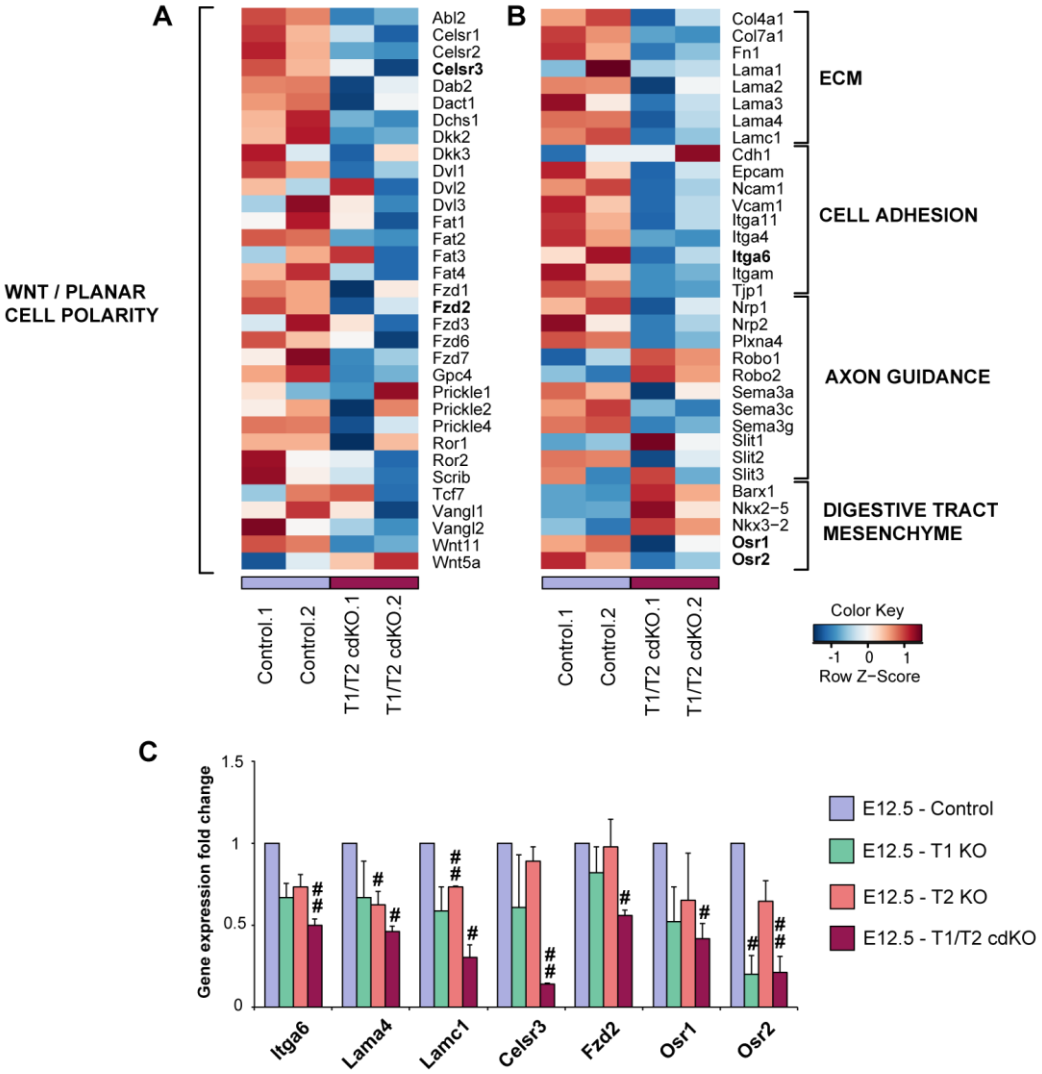

Figure 20. Modulation of signalling pathways upon genetic deletion of *Tgif2* in the pancreas.
Heatmaps illustrating differentially regulated genes in cdKO mutants *versus* control from the RNA-seq dataset. A subset of markers associated with Wnt/PCP pathway **(A),** ECM, cell adhesion, axon guidance and digestive tract mesenchyme **(B)** are shown. Colours represent high (red) or low (blue) expression values based on Z-score normalized FPKM values for each gene. White represents the average between red (high) and blue (low) expression values. Genes which were validated by RT-qPCR are in bold. **(C)** RT-qPCR analysis of markers in control and mutant E12.5 pancreata, confirming the differential gene expression observed at the RNA-seq analysis. RT-qPCR data were normalized to *36B4* and represented as fold change expression compared to control (set to 1 as calibrator) (n=3 for all genotypes). Error bars represent SEM. Two-tailed Student's T-test. (#) P<0.05; (##) P<0.001.

the mutant embryonic pancreata (Fig. 20B). This observation might explain the endocrine phenotype I observed in T1/T2 mutants, as the semaphorin-to-neuropilin signaling axis has been recently described to control pancreatic islet morphogenesis[415]. Specifically, this study showed that a disruption of this pathway leads to defects in endocrine cells migration in the surrounding mesenchyme with consequent formation of cords of GLUCAGON$^+$ cells close to the epithelial trunk[415].

Finally, terms associated with digestive tract mesenchyme (e.g. *Barx1, Nkx2-5, Nkx3-2, Osr1, Osr2*) were also differentially regulated between control and mutant pancreata. The detection of these transcripts in the RNA-seq data suggests the presence of contaminant mesenchymal tissue surrounding and/or intercalating in the epithelial pancreatic buds at the moment of the dissection. These changes therefore suggest a switch in the surrounding mesenchymal identity more likely due to the lack of *Tgif1* since *Tgif2* was deleted specifically in *Pdx1*$^{+}$ cells. Alternatively, dysregulation of mesenchymal genes might reflect an epithelium-to-mesenchyme conversion. Further investigation with *in situ* approaches will help to elucidate these observations.

In summary, the *in vivo* data indicate that the pancreatic-specific ablation of *Tgif2* leads to 'unstable' pancreatic identity, defective endocrine lineage differentiation and concomitant expansion of the acinar domain, suggesting a possible role in the allocation of the two pancreatic fates within multipotent pancreatic progenitors. Moreover, *Tgif(s)* loss-of-function results in morphogenetic defects of endocrine cell clustering, cystic duct formation and perturbation of signalling pathways including Wnt/PCP.

4.2 Study of the liver-to-pancreas lineage reprogramming role of the human TGIF2

We previously showed that the homeoprotein TGIF2 promotes the conversion of hepatic cells into a pancreatic progenitor state both *in vitro* and *in vivo* in murine model systems[71]. Such lineage reprogramming occurs in a step-wise manner through an initial down-regulation of the original hepatic features followed by the activation of the expression of key pancreatic progenitor genes. Since most of the transcription factors that are crucial for pancreatic development are conserved across species[218], during my Ph.D. I investigated whether the same change in identity can be prompted in human liver cells upon overexpression of TGIF2.

First, I characterized the gene expression pattern of *TGIF*(s) in human pancreatic and hepatic lineages (Fig. 21). By RT-qPCR I detected *TGIF1* and *TGIF2* mRNAs in human adult pancreatic tissue, liver tissue and human pluripotent stem cells, which were differentiated along the pancreatic lineage. To obtain iPSC-derived β-like cells, I used a robust differentiation protocol established in our lab. and based on recent publications (Fig. 21A)[240,416]. Contrarily to the close family member *TGIF1*, *TGIF2* gene expression levels increased as pluripotent stem cells transition to a pancreatic β-cell state. In addition, in line with previous results in the mouse[70,71], *TGIF2* transcript levels were higher in the adult human pancreas than in liver (Fig. 21B).

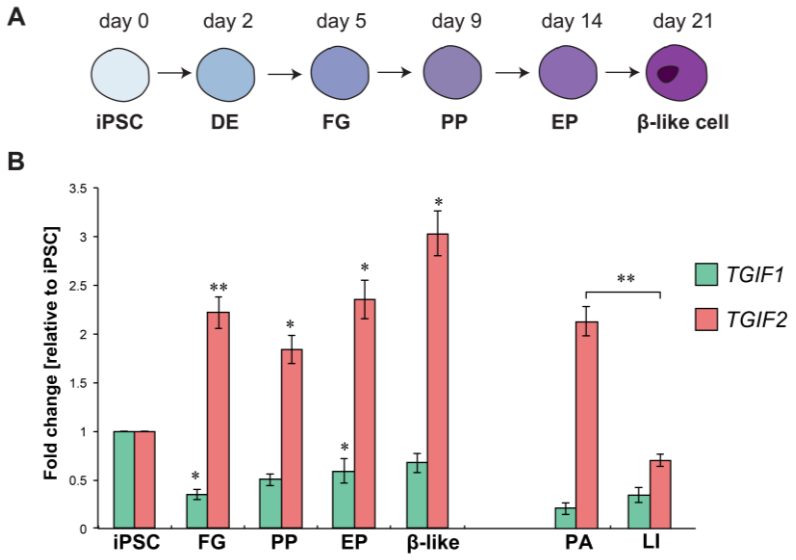

Figure 21. Expression analysis of *TGIF1 and TGIF2* in human pancreatic and hepatic cells.
(A) Schematic showing the multiple steps of the differentiation process of iPSCs along the pancreatic lineage (Helker et al. 2019). **(B)** Expression levels of *TGIF1* and *TGIF2* in iPSCs differentiated toward the β-like state and primary adult pancreatic and hepatic tissues. Data were normalized to *GAPDH* and represented as fold changes compared to iPSC (set to 1 as calibrator). Error bars represent +/- SEM. (*) P<0.05; (**) P<0.001. iPSC, induced pluripotent stem cells; DE, definitive endoderm; FG, foregut; PP, pancreatic progenitor; EP, endocrine progenitor; PA, adult pancreas; LI, adult liver.

4.2.1 Conservation of the human TGIF2 reprogramming activity in mouse liver cells

To start addressing the conservation of the human TGIF2 in a liver cellular context, I first assessed the functionality of the human *TGIF2* sequence by ectopically expressing it in bipotential adult mouse liver (BAML) cells[71,399]. This is a non-transformed hepatic cell line derived from adult mouse liver which is endowed of both bile duct and hepatocyte properties[399]. The human *TGIF2* ORF (open reading frame) was cloned in a lentivirus (LV) vector, upstream the sequence coding for the fluorescent reporter protein GFP, separated by a self-cleaving 2A peptide (Fig. 22A). Recapitulating the mouse TGIF2-induced reprogramming events, transduction of BAML cells with LV.hTGIF2 resulted in a marked reduction of the transcription levels of typical hepatic genes, such as *Albumin, Afp, Transthyretin* (*Ttr*) and *Hnf4a* after 4 weeks of culture (Fig. 22B). At the same time, the expression of pancreatic progenitor genes, such as *Pdx1, Pax6, Nestin, NeuroD1, Insm1* and *Isl1,* was turned on in the transduced liver cells (Fig. 22B).

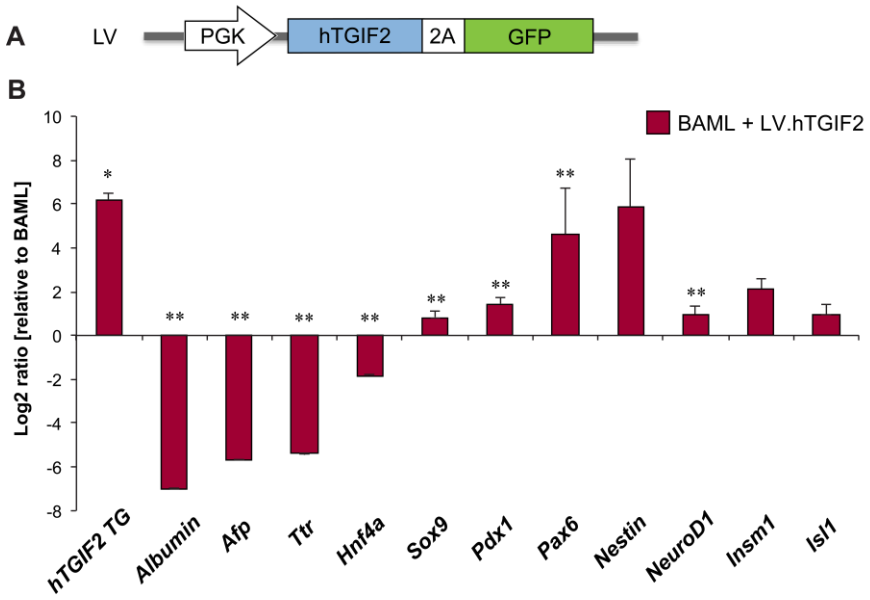

Figure 22. Ectopic expression of human TGIF2 downregulates the hepatic function and activates the expression of pancreatic genes in mouse liver cells.
(A) Schematic showing the lentiviral construct used to drive human TGIF2 (hTGIF2) expression in BAML cells. **(B)** RT-qPCR analysis of BAML cells transduced with LV.hTGIF2 showing reduction of hepatic markers and induction of pancreatic genes after 28 days of culture. Data were normalized to *Sdha* and represented as Log_2 fold change expression relative to non-transduced BAML controls. Error bars represent SEM (n=1). (*) P<0.05; (**) P<0.001.

4.2.2 Establishment of an *in vitro* culturing system of human primary hepatocytes

Adult primary hepatocytes are the golden standard tool for most of the *ex vivo* studies of liver homeostasis and diseases (e.g. to model liver disease or assess the metabolic activity of drugs)[417]. Despite their high regenerative potential *in vivo*, hepatocytes are characterized by reduced cell proliferation rate *in vitro*, which impedes any expansion after plating. Moreover, they undergo rapid dedifferentiation when cultured in a Petri dish, limiting long-term experiments[417–419]. These features largely restrict the applications of this type of cells. To overcome these limitations, I initially tested different culture conditions to establish the most suitable ones for preserving a bonafide hepatocyte phenotype for longer time in culture. Commercially-available cryopreserved human primary hepatocytes were thawed and plated on collagen-coated dishes and cultured with two different growth media, which were adapted from previously published ones. Specifically, I used one medium containing EGF, dexamethasone, Insulin-Transferrin-Selenium (ITS) and nicotinamide[71], hereafter referred to as standard HEP medium, and another medium called HLIM, which was richer in growth factors[400] (see Methods section 3.2.2.). At the morphologic analysis, cells maintained in the HLIM medium better preserved the polygonal epithelial shape typical of primary hepatocytes (Fig. 23A). Gene expression, assessed at 14 and 21 days of culture, revealed that this medium also partially limited the loss of hepatic function compared to the cells cultured in HEP medium (Fig. 23B). For instance, mature hepatocytes markers, including *ALBUMIN, APOA2, G6PC,*

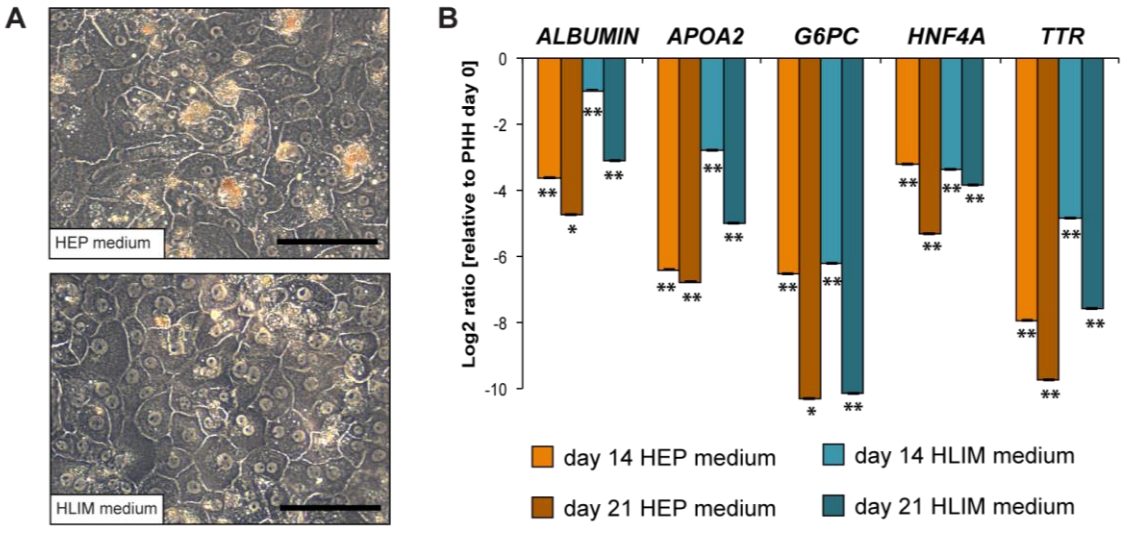

Figure 23. Establishment of the culture conditions of human adult primary hepatocytes.
(A) Transmitted light microscopy images showing the morphology of primary human hepatocytes (PHH) cultured with HEP and HLIM media 1 week post-plating. **(B)** RT-qPCR analysis showing the reduction of gene expression of typical markers of liver function, after 2 and 3 weeks of culture. Data were normalized to *GAPDH* and represented as Log_2 fold change compared to freshly isolated hepatocytes (PHH day 0). Data shown correspond to one representative experiment (n>3). Error bars represent SEM. (*) P<0.05; (**) P<0.001.

HNF4A and *TTR,* progressively decreased over time, more rapidly and in a more pronounced way when the HEP medium was applied (Fig. 23B). Nevertheless, both culture conditions were suitable for *in vitro* culture of hepatocytes for few weeks and were used in the reprogramming experiments.

4.2.3 Ectopic expression of human TGIF2 promotes a liver-to-pancreas conversion in primary human hepatocytes

Primary human hepatocytes (PHHs) from different donors were exposed to hTGIF2 delivered by a lentiviral approach, maintained in standard HEP medium (Fig. 24A) or HLIM medium for almost 2 weeks and analysed by RT-qPCR and IF techniques. I used non-transduced cells at matching time points as control. *TGIF2* overexpression promoted down-regulation of typical hepatic gene markers, as indicated by the reduced levels of *ALBUMIN, HNF4a, G6PC* and *PCK1* at both transcript and protein levels (Fig. 24B-C).

A comparable repression of the hepatic state was observed in both HEP medium and HLIM medium (Fig. 24B-C and data not shown). However, only in the former condition a panel of essential genes regulating pancreatic cell identity, including *PDX1, PTF1a*, *NKX6-1, NKX2-2*, was concomitantly induced (Fig. 24C). Interestingly, transcription factors involved in endocrine cell identity, such as *NGN3, PAX6, NEUROD1, ARX, MAFA*, were also induced in the reprogrammed hepatocytes (Fig. 24C).

Remarkably, the levels of induction of pancreatic TFs gene expression were comparable to those obtained by differentiating human pluripotent stem cells to a pancreatic progenitor (PP) state, by applying the multi-step differentiation protocol described in Chapter 4.2. (Fig. 21; 25). Notably, I detected the induction of the membrane receptor *CELSR3* upon exposure to the LV.hTGIF2 (Fig. 24C). These results are in line with previous reprogramming experiments in the mouse as well as *in vivo* studies showing an interplay between TGIF2 and the non-canonical Wnt/PCP signalling pathway in the allocation of the pancreatic fate in the mouse embryo[70,71].

Importantly, pancreatic differentiation markers, e.g. *INSULIN*, or genes typically expressed in other endodermal lineages, including the intestinal-specific *CDX2* or the stomach-specific *SOX2,* were not detected by RT-qPCR (Fig. 24E). Also, LV.GFP transduction in the hepatic cells did not elicit any change in the liver function or induction of non-hepatic genes, behaving similarly to non-transduced cells (Fig. 24D).

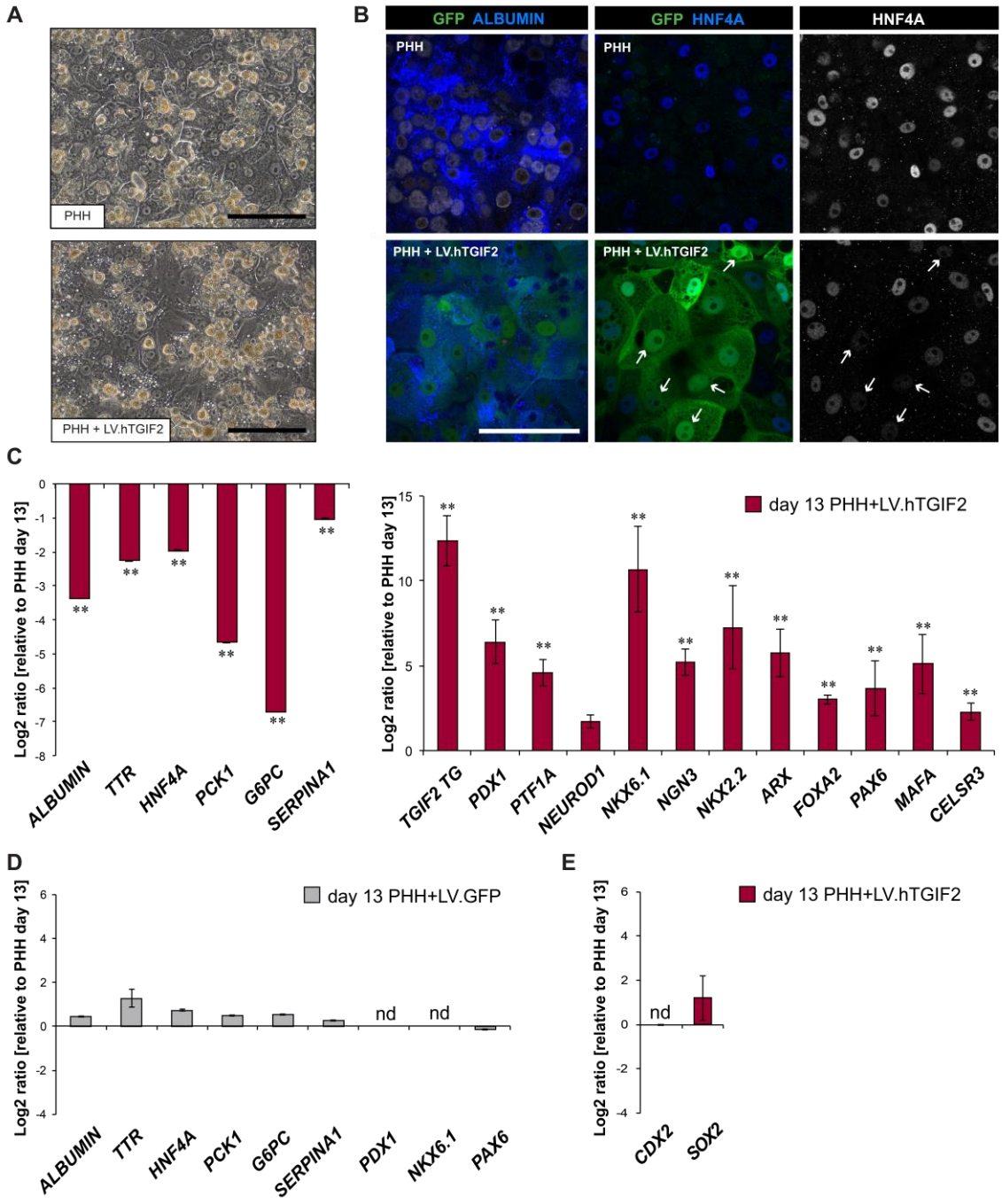

Figure 24. Human TGIF2 downregulates hepatic features and activates the expression of pancreatic genes in human primary hepatocytes.

(A) Micrograph of primary human hepatocytes control (PHH) and transduced with LV.hTGIF2 (PHH+LV.hTGIF2). **(B)** IF analysis of indicated liver markers (blue) in PHH cultured for 13 days upon transduction with LV.hTGIF2. Arrows indicate nuclei of transduced cells with reduced levels of the hepatic transcription factor HNF4A. **(C)** RT-qPCR analysis showing the repression of liver genes and induction of pancreatic marker genes in PHH transduced with LV.hTGIF2 with MOI=20. **(D)** RT-qPCR analysis showing that LV.GFP transduction does not elicit liver-to-pancreas reprogramming. **(E)** RT-qPCR analysis showing that non-pancreatic genes *CDX2* and *SOX2* are not induced upon expression of hTGIF2. RT-qPCR data were normalized to *GAPDH* and represented as Log2 expression ratio compared to relative controls at the equivalent stage. Data shown correspond to one representative experiment performed on human primary hepatocytes grown in HEP medium (n>3). Error bars represent +/- SEM. (*) P<0.05; (**) P<0.001; nd, not detected. Scale bars, 100 µm.

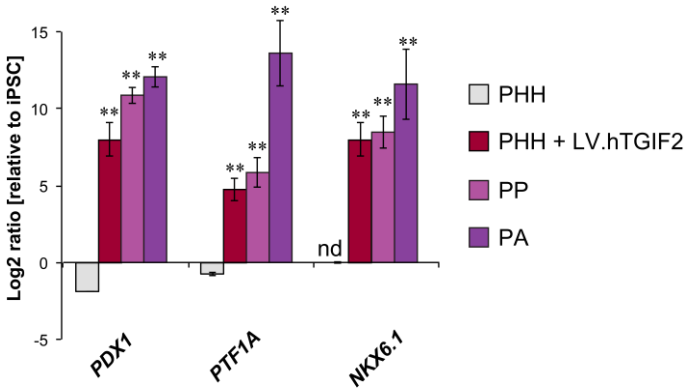

Figure 25. Induction of pancreatic genes in direct cell reprogramming and differentiation.
RT-qPCR analysis showing the levels of gene expression of three crucial pancreatic factors, *PDX1, PTF1A* and *NKX6.1,* in reprogrammed hepatocytes (PHH + LV.hTGIF2), iPSC-derived pancreatic progenitors (PP) and adult pancreatic tissue (PA), as benchmark for the analysis. Data were normalized to *GAPDH* and represented as Log_2 fold change relative to iPSCs. Error bars represent +/- SEM. (*) P<0.05; (**) P<0.001; nd, not detected.

4.2.4 Gene delivery systems to overexpress TGIF2 in human primary hepatocytes

Next, to further investigate the TGIF2 reprogramming function in the human hepatocytes, I used two different delivery systems to assess if there is any difference in transduction efficiency and functional response. The human *TGIF2* was expressed upon stable genomic integration using a lentivirus vector[406] or in a transient way through AAV viral vectors (Fig. 26A). In the latter approach, to confer hepatic cell-type specificity, I used the hepatotropic serotype AAV2/8[420] and the Thyroxine binding globulin (TBG) promoter to drive hTGIF2 expression (Fig. 26A). PHH were transduced with AAV.hTGIF2 with multiplicity of infection (MOI) of 5.10^5 or LV.hTGIF2 with MOI of 20. Contrarily to the effect of LV-mediated reprogramming, AAV-mediated expression of TGIF2 led to a less significant repression of the livers markers and moderate induction of pancreatic genes (Fig. 26B and data not shown). Such discrepancy might be due to the lower levels of transgene expression driven by the AAV compared to the LV vector. Indeed, AAV-based system has been shown to be a powerful tool of gene delivery *in vivo*, but has a poor transduction efficiency in *ex vivo* cell cultures and is an episomal vector system[421].

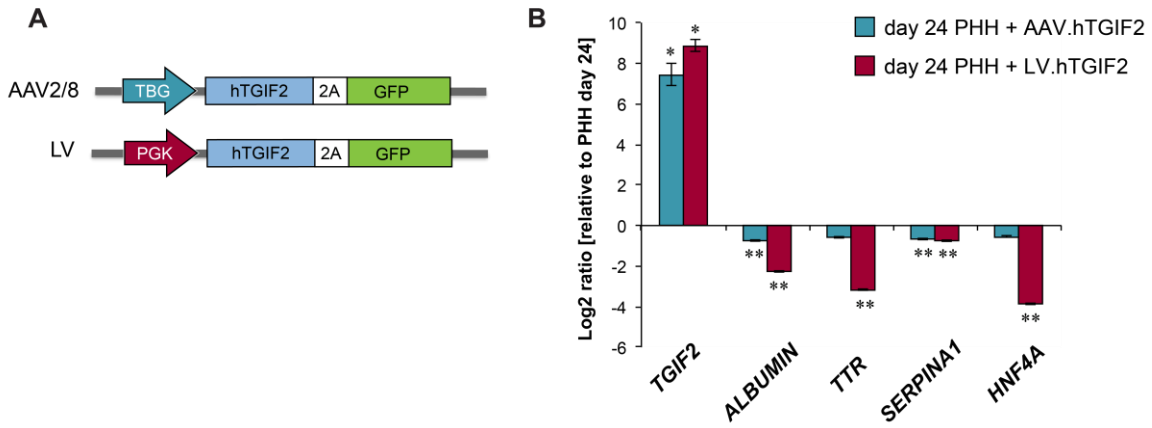

Figure 26. Comparison between the AAV and the LV-based methods to mediate reprogramming of hepatocytes by TGIF2.
(A) Schematics of the vectors employed in this study. The sequence of the human *TGIF2,* cloned upstream the reporter green fluorescence protein (GFP) is expressed by a lentiviral vector (LV), under the control of the constitutive PGK promoter, or by an hepatotropic AAV virus, under the control of the TBG promoter. **(B)** RT-qPCR analysis showing reduced expression of typical marker genes encoding liver functions in human primary hepatocytes (PHH) AAV- and LV-transduced with TGIF2 after 24 days of culture. Data were normalized to *GAPDH* and represented as Log_2 fold change relative to matching stage hepatocytes transduced with AAV.GFP and LV.GFP (PHH day 24) respectively. Data shown correspond to one representative experiment (n>3). Error bars represent +/- SEM. (*) P<0.05; (**) P<0.001.

4.2.5 Optimizing the conditions for an efficient TGIF2-induced liver-to-pancreas reprogramming

Reprogramming (a.k.a. transdifferentiation) is a phenomenon that requires heavy rewiring of the original fate into the new identity program[15]. Both genetic and epigenetic events must occur to unlock those loci that confer resistance to changes in cell state, while stable activation of genes regulating the new desired functions needs to be ensured[15,352]. Exogenous expression of the single factor TGIF2 confers pancreatic progenitor characteristics to hepatic cells[71]. Remarkably, low doses of LV.hTGIF2 result in downregulation of hepatic features, but the activation of the pancreatic program is successful primarily when high levels of transgene expression are reached (Fig. 24). In order to further extend the study of the hepato-pancreatic plasticity and increase the efficiency of the reprogramming driven by TGIF2, I tested multiple parameters, which are summarized in Table 10.

First, I used PHHs from donors of different ages and sexes and found that the downregulation of the hepatic function is consistently achieved in all cell sources upon exposure to hTGIF2, with some variability in the magnitude of the effect. Intriguingly, the expression of pancreatic markers was triggered at high levels only in PHH from male donor (Fig. 24; Table 10). This might be due to specific genetic or epigenetic restrictions in the cellular source that limit the propensity to cell fate conversion.

	Experimental conditions	Liver genes	Pancreatic genes	n.
PHH SOURCE	Lonza Lot.HUM4191 (Male, 27)	↓↓↓	↑↑↑	4 (*)(*)(*)
	Lonza Lot.HUM4180A (Female, 8)	↓↓	no change	1
	Lonza Lot.HUM4075B (Female, 38)	↓↓↓	no change	1
	Invitrogen Lot.HU8216 (Female, 58)	↓↓	no change	7
	Invitrogen Lot.HPP20160501 (5 donors pooled)	↓	no change	2
DELIVERY SYSTEM	LV.hTGIF2	↓↓↓	↑↑↑	13 (*)(*)
	AAV.hTGIF2	↓	↑	5 (*)
CELL CULTURE MEDIUM	HEP medium + ITS + Dex + EGF	↓↓↓	↑↑↑	12 (*)(*)
	HEP medium + ITS + EGF	no change	no change	1
	HEP medium + ITS + Dex	no change	no change	1
	HEP medium + FGF10	↓	no change	1
	HEP medium + panc. diff. cytokines	no change	no change	1
	HLIM medium	↓↓↓	no change	4 (*)

Table 10. Experimental conditions to enhance the hTGIF2-mediated liver-to-pancreas reprogramming in human primary hepatocytes.
Different sources of primary human hepatocytes (PHH), gene delivery methods and modified cell culture media were tested to enhance TGIF2-driven reprogramming process in hepatocytes. Arrows indicate the magnitude of repression or induction of hepatic and pancreatic marker genes, respectively, as measured by RT-qPCR. (*) indicates conditions used in the experiments shown in Fig. 24 and 26; (*) of the same colour marks the conditions used simultaneously in the same reprogramming experiments.
Abbreviations: dexamethasone, dex; Insulin-Transferrin-Selenium, ITS; pancreatic differentiation, panc. diff; number of experiments, n.

Secondly, as previously discussed, I established that LV transduction of TGIF2 leads to a more robust liver-to-pancreas lineage reprogramming, considering the changes in the transcript levels of the two programmes compared to the AAV method (Fig. 26).

Finally, after establishing that the HEP medium was more beneficial for the acquisition of the pancreatic state compared to the HLIM medium (Fig. 24), I further modified this culturing condition by removing from the medium some growth factors that are known to be relevant for preserving an hepatic phenotype. For instance, depletion of EGF or dexamethasone from the complete medium did not result into any amelioration of the reprogramming. Alternatively, I supplemented the culture medium with “pancreatic differentiation cytokines”, which have been shown to be crucial in the expansion and further differentiation of pancreatic progenitor cells, derived *in vitro* from pluripotent stem cells[240]. To this aim, I added to the standard HEP medium the small compounds LDN-193189, TBP, ALK5 inhibitor II and FGF10 after 2 weeks of culture,

when pancreatic gene expression starts to be detected in the TGIF2-transduced cells. Nevertheless, the exposure to these compounds as well as addition of FGF10 alone, known to be fundamental for expanding mouse pancreatic progenitors[66,137], did not further facilitate the reprogramming process (Table 10).

4.2.6 Addition of possible co-factors of TGIF2 does not enhance liver-to-pancreas lineage reprogramming

Direct lineage reprogramming often implies the activity of multiple transcription factors that behave as repressors of the original state or activators of the new acquired fate[422]. Members of the TALE homeoprotein superfamily can function as context-dependent gene repressor, activator or both and can cooperate with each other in complexes to exert their biological functions during development and differentiation processes[423]. Therefore, I investigated whether TGIF2 reprogramming activity could be enhanced by the concomitant overexpression of additional members of the same family (Fig. 27A). A valid candidate in this perspective is PBX1, which has been shown in our laboratory to be fundamental for the embryonic development of pancreas (Cozzitorto C et al., manuscript under revision). PBX1 might act as a pioneer factor[280] for the subsequent activation of the cell identity regulator TGIF2, as shown in other cellular contexts[280,281]. Alternatively, PBX1 could directly bind to regulatory elements of transcriptional targets (e.g. pancreatic or hepatic genes) in hetero-multimeric complexes together with TGIF2[424]. Human PBX1 ORF, fused to the red fluorophore mCherry by the 2A self-cleaving peptide, was cloned into a lentiviral vector[406] under the expression of the constitutive PGK promoter (Fig. 27B). Ectopic expression of PBX1 on PHH was carried out at the same time or after 2 weeks of stable introduction of LV.hTGIF2. However, both simultaneous or consecutive ectopic expression of the two TFs did not lead to any improvement in the liver-to-pancreas reprogramming in the tested experimental settings (Fig. 27C).

Next, I assessed whether additional pancreatic TFs might work in a synergistic manner together with TGIF2 and stabilize or promote a more differentiated pancreatic phenotype in the transduced cells. The combination Pdx1-MafA-Ngn3 (called PMN) has been shown to convert different somatic cells into pancreatic fate when overexpressed by adenoviral approach in mouse[356,358,359]. Thus, I cloned the mouse multicistronic cassette, interspaced by self-cleaving 2A peptides and followed by the GFP reporter sequence, into the same AAV2/8 vector[420] previously used for ectopic expression of TGIF2 in hepatocytes (Fig. 26A). AAV.mPMN alone or the addition of AAV.mPMN to the LV.hTGIF2-treated cells did not elicit any conversion to

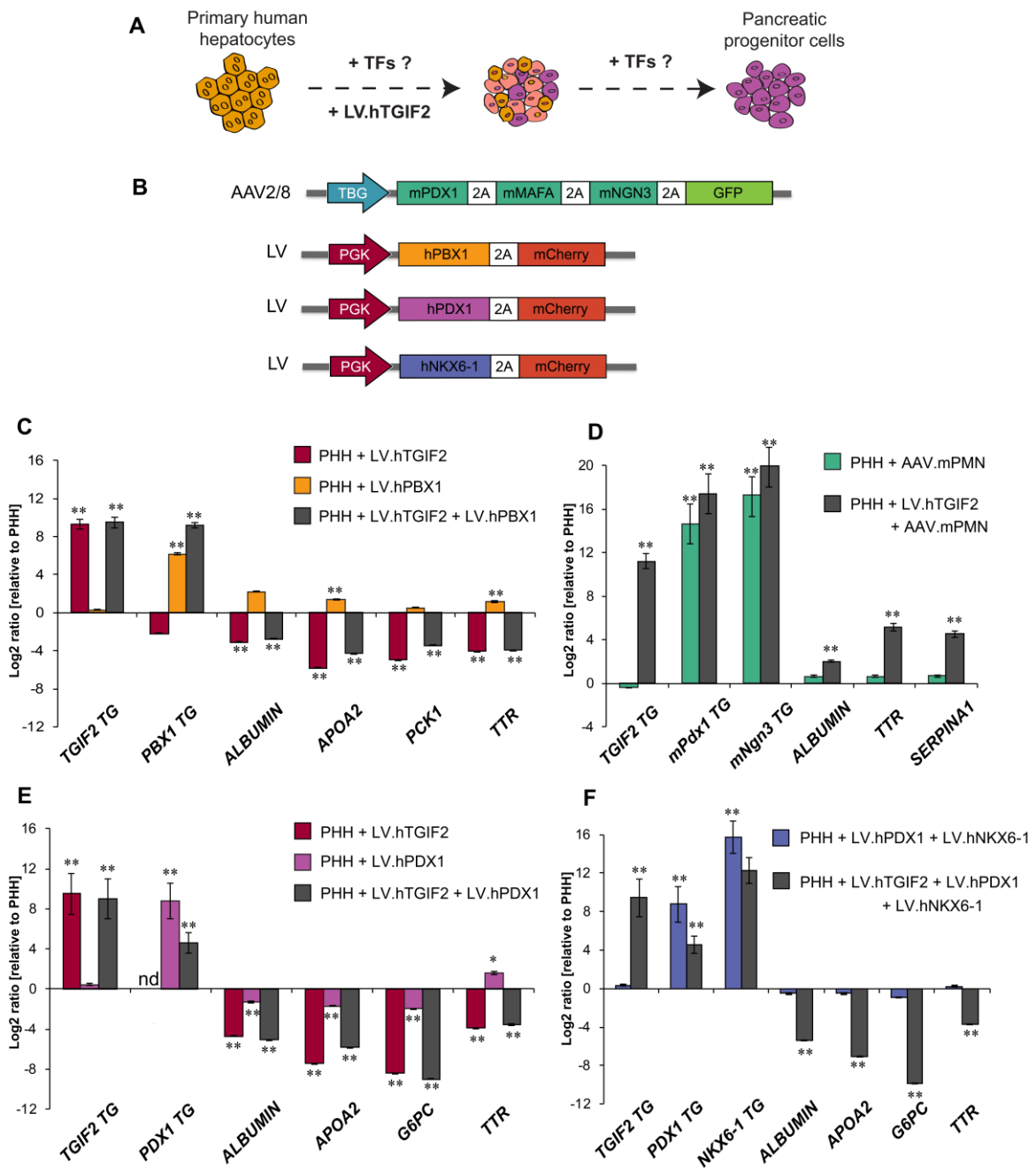

Figure 27. Addition of possible co-factors of TGIF2 does not enhance liver-to-pancreas lineage reprogramming.

(A) Schematic of the experimental strategy used to study the combinatorial activity of TGIF2 with other TFs in the reprogramming of PHHs. Relevant TFs were added simultaneously or consecutively to LV.hTGIF2 introduction in the cellular system. **(B)** Schematics of the vectors employed in this study. The sequences of the mouse Pdx1-MafA-Ngn3 (PMN), cloned upstream the GFP reporter, are expressed by a multicistronic hepatotropic AAV virus under the control of the TBG promoter. The sequences of the human TALE homeoprotein PBX1 and the pancreatic TFs PDX1 and NKX6-1 were cloned individually in lentiviral vectors (LV), under the control of the constitutive PGK promoter, upstream the red fluorophore mCherry. **(C-F)** RT-qPCR analysis of cells transduced with TGIF2 along with the different TFs ectopically expressed at the same time: hPBX1 **(C)** analysed at day 14; AAV.mPMN **(D)** analysed at day 26; hPDX1 alone **(E)** or in combination with hNKX6-1 **(F)** analysed at day 26 of cellular culture. Data were normalized to *GAPDH* and represented as Log_2 fold change relative to matching stage non transduced hepatocytes. Data shown correspond to single representative experiments (n=2 for each experiment). Error bars represent +/- SEM. (*) P<0.05; (**) P<0.001; nd, not detected.

the pancreatic fate or improve TGIF2 activity (Fig. 27D). This result is in contrast with published data in the mouse; the discrepancy might be due to the different approach used to deliver the genetic material *in vitro* compared to previous studies, which employed adenovirus vectors, or to species-specific differences.

To start addressing potential differences between human and mouse pancreatic TFs, I engineered transgenic constructs in which hPDX1 and hNKX6.1 ORF, fused to the mCherry reporter by a self-cleaving 2A peptide, were individually cloned into lentiviral vectors under the expression of the constitutive PGK promoter (Fig. 27B). Next, I exposed TGIF2-expressing human hepatic cells to LV.hPDX1 alone or in combination with LV.hNKX6.1 and analysed the phenotype of reprogrammed cells by RT-qPCR. Preliminary results showed that the addition of PDX1 and/or NKX6.1 does not exert any positive influence in the reprogramming process compared to the transduction with LV.hTGIF2 alone (Fig. 27E-F). Further investigations are required to address the right combination of TFs and/or to recapitulate the correct temporal succession of events during reprogramming.

Taken together, these results demonstrate that the ability of TGIF2 to robustly repress hepatic functions and to promote pancreatic progenitor genes, to some extent, in primary hepatocytes cultured *in vitro* is conserved in humans. Such conversion occurs with greater efficiency when mediated by an integrating viral vector, such as the lentivirus, and at a high dose. The identification of molecular agents that possibly enhance the phenomenon deserves further investigation.

4.3 Study of the reprogramming activity of TGIF2 in non-endodermal *ex vivo* cellular contexts

Cells of close embryonic lineage are more prone to undergo lineage reprogramming compared to cells arising from the different germ layers[15,351]. Nevertheless, several studies reported cell fate conversion between developmentally and epigenetically distant types of cells driven by small molecules or defined transcription factors, within or outside the endodermal cell lineages[422,425,426]. For instance, fibroblast cells have been converted to a neuronal fate[422] or hepatocyte-like state[425] upon ectopic expression of key TFs regulating distinct cell types identity. TGIF2 reprogramming potential has not yet been characterized in a non-endodermal context.

During my Ph.D. studies, I asked whether TGIF2 can induce endodermal or, more specifically, pancreatic fate in a lineage different from the hepatic one.

4.3.1 Ectopic expression of TGIF2 activates pancreatic gene expression in mouse and human fibroblasts

To expand the study of the reprogramming function of TGIF2, I ectopically expressed this factor in mouse adult fibroblasts (MAF) and human adult dermal fibroblasts (hADF), using a lentiviral vector approach (Fig. 28), which mediates high efficiency gene delivery.

Interestingly, I observed that murine fibroblasts cultured in standard fibroblast medium for 35 days after LV.mTGIF2 transduction (Fig. 28A) undergo a striking change in morphology, adopting an epithelial-like cuboidal shape (Fig. 28B) accompanied by down-regulation of fibroblast-specific markers, such as *Vimentin* and *Col1a1,* and induction of the epithelial marker *Cdh1* (Fig. 28C). Most importantly, I found significant induction of expression of relevant endodermal transcription factors, such as *Sox17*, and the pancreatic genes, *Pdx1, Ptf1a, NeuroD1,* by RT-qPCR (Fig. 28C). By performing IF on fixed cells, I visually detected the presence of PDX1 and SOX9 proteins in patches of cells transduced with LV.mTGIF2 (Fig. 28D).

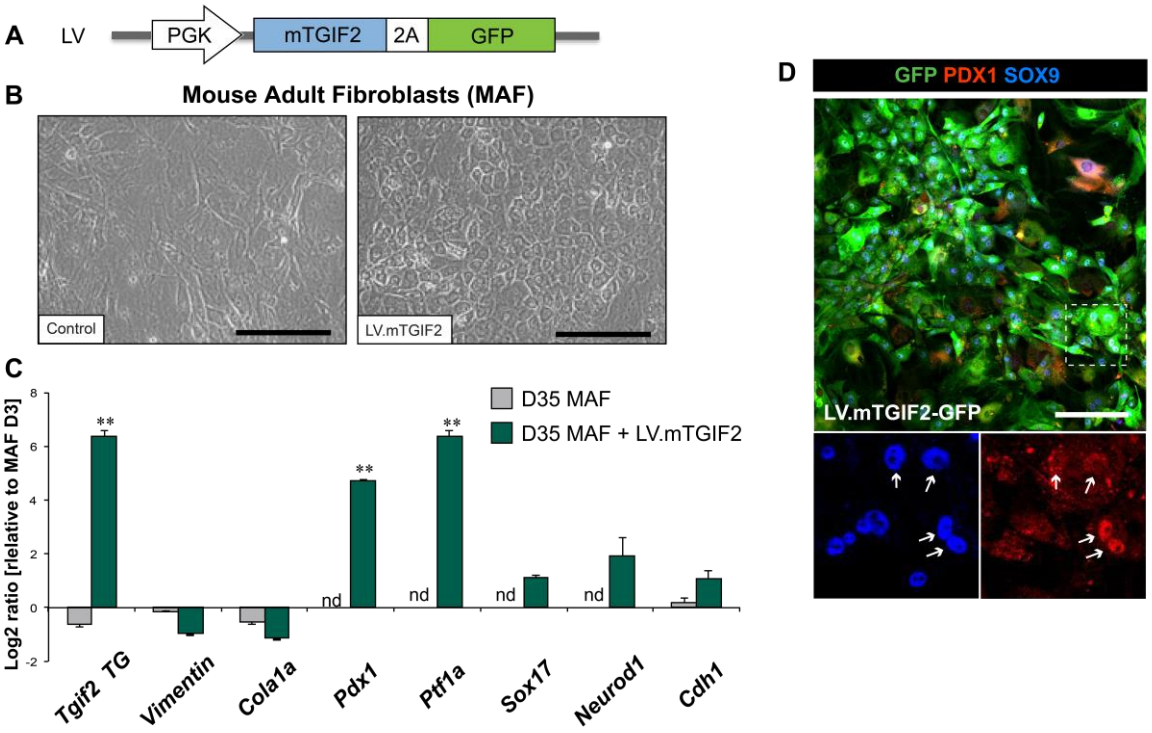

Figure 28. Ectopic expression of *Tgif2* in mouse adult fibroblasts induces changes in morphology and expression of pancreatic genes.
(A) Schematic of the LV.mTGIF2 vector. **(B)** Morphology of control and LV.mTGIF2-transduced Mouse Adult Fibroblasts (MAF). **(C)** RT-qPCR analysis of MAF cells after 35 day of transduction with LV.mTGIF2. Data were normalized to *Sdha* and represented as Log2 expression ratio compared to control at day 3. Data shown correspond to one representative experiment (n=2). Error bars represent SEM; (**) P<0.001; n.d. not detectable. **(D)** IF staining of PDX1 and SOX9 in LV.mTGIF2-transduced MAFs. Lower panels show single channels images of the area in the dashed box. Arrows indicate mTGIF2-GFP-transduced cells co-expressing SOX9 and PDX1. Scale bars, 100 µm.

hADF, which were isolated from healthy donors, displayed a similar change in morphological properties upon transduction with LV.hTGIF2, rearranging themselves in clusters of epithelial-like cells (Fig. 29A,B). Compared to non-transduced cells at day 0, the expression of fibroblast markers, such as *VIMENTIN* and *S100A4,* was downregulated (Fig. 29C, D). This was accompanied by a remarkable induction of the expression of the pancreatic genes *PDX1* and *PTF1a,* the endodermal marker *SOX17* and the epithelial marker *CDH1* in the RT-qPCR analysis (Fig. 29C). IF analysis confirmed the presence of PDX1 and the exclusion of VIMENTIN in hTGIF2-GFP^{+} cells (Fig. 29D).

Notably, transcriptional expression of the non-canonical Wnt/PCP component *CELSR3* was also found to be induced in this cellular context (Fig. 29C), supporting the idea of a direct correlation between TGIF2 activity and PCP components, which might help to drive cell fate transitions.

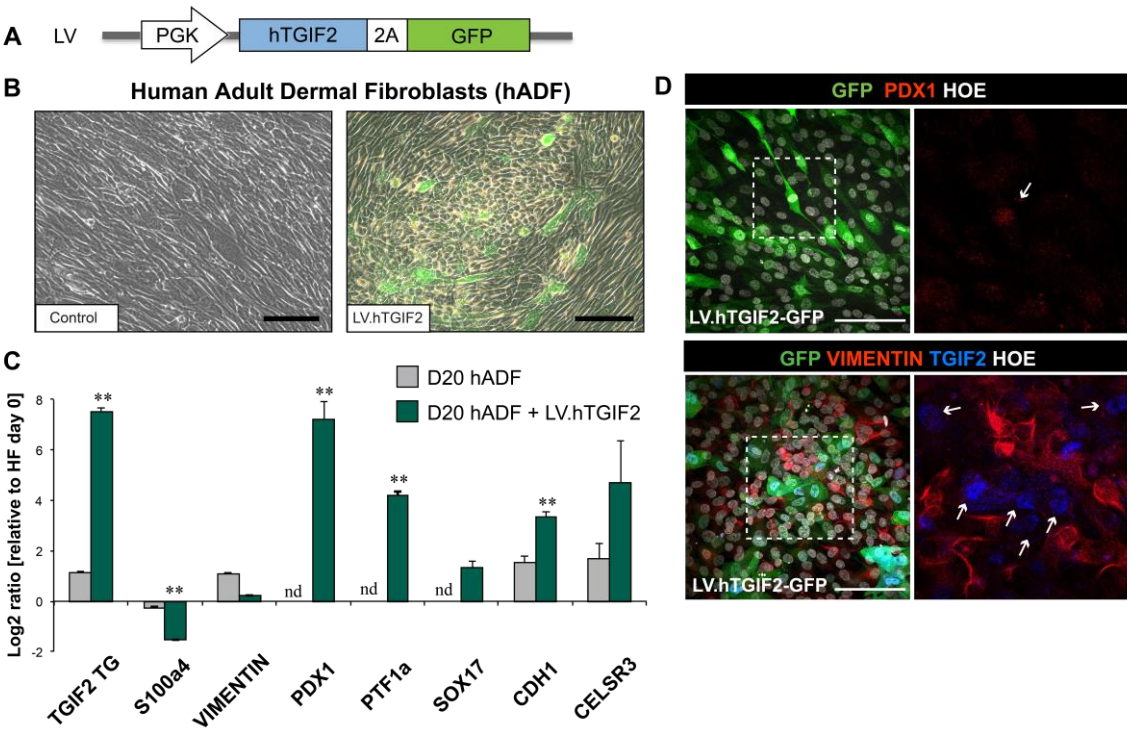

Figure 29. Ectopic expression of *TGIF2* in human adult dermal fibroblasts induces changes in morphology and expression of pancreatic genes.
(A) Schematic of the LV.hTGIF2 vector. **(B)** Morphology of control and LV.hTGIF2-transduced human Adult Dermal Fibroblasts (hADF) at day 20. **(C)** RT-qPCR analysis of hADF cells at day 20 transduced with LV.hTGIF2. Data were normalized to *GAPDH* and represented as Log_2 expression ratio compared to control at day 0. Data shown correspond to one representative experiment (n>3). Error bars represent SEM. (**) P<0.001; n.d. not detectable. **(D)** IF staining of PDX1, VIMENTIN and TGIF2 in LV.hTGIF2-transduced hADFs. On the left, magnified pictures of the area in the dashed boxes. Hoechst (HOE) nuclear counterstain in grey. Arrows indicate hTGIF2-GFP-transduced cells positive for PDX1 or negative for VIMENTIN. Scale bars, 100 µm.

4.3.2 Optimization of the TGIF2-mediated lineage reprogramming in fibroblasts

Prompted by the promising results obtained upon expression of TGIF2 alone in fibroblasts, I sought to improve the reprogramming process, including the efficiency and differentiation of the obtained reprogrammed cells, by modifying cell culture conditions and/or adding other relevant TFs as co-factors. Table 11 provides a summary of the experiments performed.

Specifically, I employed different sources of hADFs, including commercially available dermal fibroblast cells (purchased from Lonza) or in-house fibroblasts isolated from the skin biopsy tissue of a young healthy female donor. The latter, provided by Watt lab (King's College London), were FACS-purified using the pan-fibroblast marker CD90[401]. Among the different cellular sources tested, commercially-available and $CD90^+$ fibroblast cells were the ones that best responded to the reprogramming cues (Table 11).

	Experimental conditions	Epithelial morphology	Pancreatic genes	n.
hADF SOURCE	Lonza NHDF Lot. 489572 (male)	++	↑↑	4 (*)(*)
	in-house MDCH00122 (male), MDCH00123 (female)	no change	no change	1
	in-house CD90^{+} (female)	++	↑↑	3 (*)
CELL CULTURE MEDIUM	FIB medium	++	↑↑	4 (*)
	FIB medium + ALK5i	no change	no change	1
	FIB medium + 5-AZA	+	no change	1
	FIB medium + panc.diff.cytokines	no change	no change	1
	FIB medium + 3D suspension	Round clusters	↑↑	1 (*)

Table 11. Experimental conditions to enhance TGIF2-mediated reprogramming of fibroblasts.
Different sources of human Adult Dermal Fibroblasts (hADF) and modified cell culture media were used to promote TGIF2-driven reprogramming process in fibroblasts. Arrows indicate the magnitude of induction of pancreatic marker genes, as judged by RT-qPCR. (*) indicates conditions used in the same experiments used to show representative data in Fig. 28 and 29. (*) of the same colour indicates conditions used simultaneously in the same reprogramming experiments.
Abbreviations: 5-azacytidine (5-AZA); pancreatic differentiation, panc. diff; number of experiments, n.

As gene delivery method, I exclusively used lentiviral (LV) vectors for their great efficiency in transducing dividing and non-dividing cells. Different doses of LV (MOI from 5 to 40) and culture periods were tested. I found that MOI=20 is the most sustainable dosage for avoiding cell stress and 3-weeks duration is the minimum time window to allow expression of pancreatic markers.

Inhibition of the TGF-β signalling is known to regulate epithelial morphogenesis, drive mesenchymal-to-epithelial transition (MET) and enhance reprogramming of somatic cells to a pluripotent state[427]. Based on this knowledge, I added to the cell culture medium the ALK5 inhibitor, which blocks the TGF-β signalling by competing for the TGF-β RI kinase. The addition of this compound did not seem to improve the outcome of the reprogramming in the tested conditions (Table 11).

Chromatin modifiers have a primary role in the transition between different epigenetic states during reprogramming. Recent studies showed that treatment with the DNA methyl-transferase inhibitor 5-azacytidine (5-AZA) promotes the establishment of a pancreatic phenotype in human skin fibroblasts, when followed by a multi-step differentiation protocol based on a cocktail of cytokines to drive pancreatic differentiation[428,429]. I therefore included 5-AZA in the TGIF2-based reprogramming protocol by exposing the fibroblasts to the chemical compound for 18 hours prior to LV.hTGIF2 transduction. I subsequently transduced and grew the treated cells in standard fibroblast medium (Table 11). Even if the addition of 5-AZA to fibroblasts initially resulted in cell stress followed by acquisition of epithelial-like cell shape, this was not accompanied by expression of pancreatic genes as judged by RT-qPCR analysis (Table 11).

4.3.3 3D culture conditions enhance TGIF2-mediated reprogramming of fibroblasts

In an effort to simulate the differentiation protocol used to generate β-like cells from iPSCs[240], I decided to adapt this three-dimensional (3D) culture condition for human fibroblast cells undergoing lineage reprogramming. Specifically, I cultured control non-transduced and LV.hTGIF2-transduced hADFs for 3 weeks in 2D and, subsequently, transferred them in 3D suspension culture by growing the cells on an orbital rotator shaking at the speed of 100 rpm. Serum-reduced medium was supplemented with the set of cytokines (ALK5 inhibitor, LDN-193189, TBP and FGF10), which is used to induce the transition from pancreatic progenitors to endocrine progenitors in human PSC-based differentiation protocol[240] (Fig. 30A). Clusters of cells initially formed from both control fibroblasts and hTGIF2.LV-cells. While control clusters rapidly disaggregated, the hTGIF2-reprogrammed cells evolved into round and dense multicellular structures (Fig. 30B). This morphological organization was accompanied by significant induction of various pancreatic genes, including *PDX1, PTF1A, SOX9, NKX6.1, NEUROD1 and MAFA* (Fig. 30C), at higher levels when compared to the simple 2D condition (Fig. 29C). Importantly, the sole addition of pancreatic differentiation cytokines to cells undergoing reprogramming in 2D did not enhance the expression of pancreatic markers in these experimental settings (Table 11).

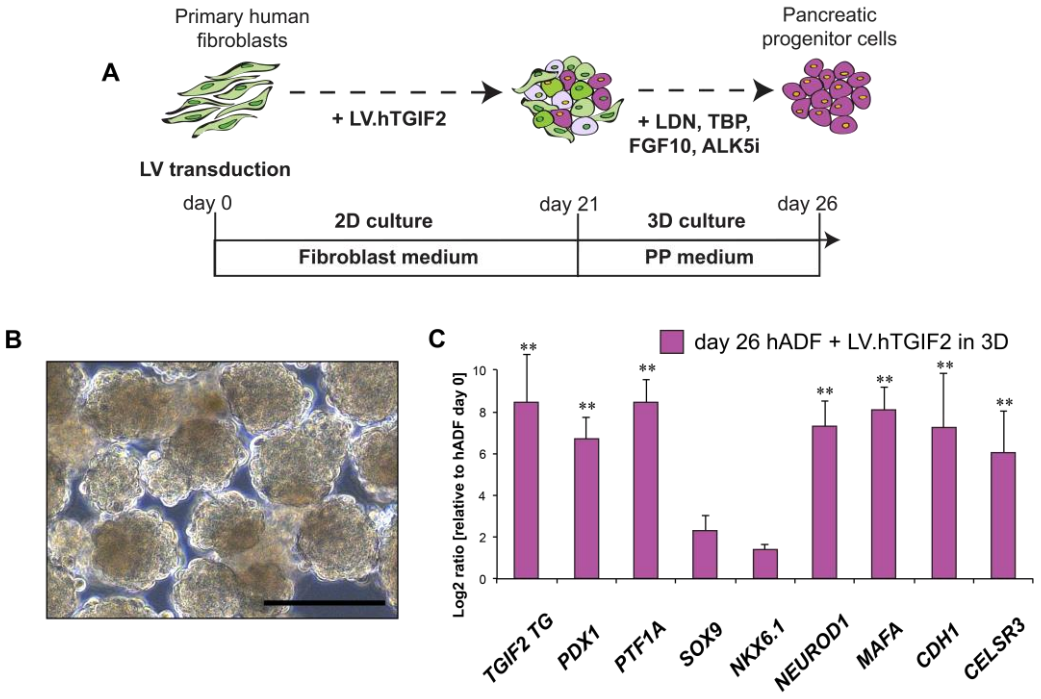

Figure 30. 3D culture of TGIF2-reprogrammed fibroblasts results in robust expression of pancreatic markers.
(A) Schematic of the experimental strategy. **(B)** Morphology of LV.hTGIF2-transduced human Adult Dermal Fibroblasts (hADF) at day 26, after 5 days of 3D culture. **(C)** RT-qPCR analysis of hADFs reprogrammed by TGIF2. Data were normalized to *GAPDH* and represented as Log_2 fold change expression ratio compared to control fibroblasts (n=1). Error bars represent SEM. (**) P<0.001. Scale bar, 100 μm.

Taken together, these results provide evidence that relevant genes of the pancreatic progenitor program can be turned on in different cellular contexts by TGIF2 ectopic expression. Remarkably, these observations support the interdependence between the two transcription factors, TGIF2 and PDX1, even outside an endodermal context and in different mammalian species, and unveil the tight connection between changes in morphology and gene expression programs.

5. DISCUSSION

5.1 TGIF2 safeguards pancreatic identity *in vivo*

To date, extrinsic signals and intrinsic factors that control and restrict lineage identity in the endoderm derivatives are only partially elucidated[2,16,129]. Recent studies in the Spagnoli lab. identified TGIF2 as a novel regulator of the pancreatic fate, sitting upstream of the transcriptional cascade that drives pancreatic specification[71]. Precisely, this TALE homeoprotein has a dual activity in the endoderm, acting as a repressor of the hepatic fate and an activator of the pancreatic one, in a conserved way between amphibians and mammalian species[63,71]. Genetic experiments in the mouse model showed that early embryos lacking *Tgif2* displayed larger liver buds and, at the same time, reduced size of pancreatic rudiments[71]. Notably, TGIF2 also emerged among the top pan-epithelium TFs expressed at early stages in the pancreas[87], further supporting the relevance of this TALE homeoprotein in the pancreatic progenitor domain. The crucial developmental role of *Tgif2* in endodermal patterning and its restricted presence in the pancreatic lineages[70,71] led us to further investigate its cell-autonomous activity during pancreatogenesis, after lineage segregation occurs.

The first aim of this thesis was to elucidate the *in vivo* requirement of *Tgif2* in the formation of the pancreas in the mouse embryo. For this purpose, I employed a loss-of-function approach based on the Cre recombination system to enable pancreatic tissue-specific deletion of loxP flanked *Tgif2* allele. To avoid functional redundancy with the close family member TGIF1, I intercrossed *Tgif1* +/- KO mice with *Tgif2* conditional KO mice. I focused my analysis on pancreatic fate specification, differentiation and proliferation of different pancreatic cell types at different embryonic stages in control and mutant littermates. The genetic analysis of *Tgif2* in the mouse embryo indicates that this homeodomain protein not only controls pancreatic fate allocation but also preserves the maintenance of the pancreatic identity in the mouse embryo and prevents the acquisition of different close cellular fates, such as the hepatic one. In fact, conditional deletion of *Tgif2* in the *Pdx1* domain (Fig. 11) results in the aberrant expression of liver hallmarks in the pancreatic epithelium (Fig. 13; 14). Simultaneous expression of ALBUMIN with PDX1 or PROX1 at early stages and with PDX1, SOX9 or CPA1 at later stages indicates that epithelial pancreatic cells do not completely switch to a hepatoblast state. The metastable identity of pancreatic cells rather develops into a hepato-pancreatic hybrid condition. It has to be addressed whether such cell plasticity occurs also in new-born and adult mice and whether it affects the normal metabolic function of the mature organ in adult *Tgif2* mutants. Consistently, the *in vivo* effects of *Tgif2* deletion are exactly the opposite of the ones

achieved in *in vitro* gain-of-function experiments using the same TF, whereby the hepatic features are erased[71].

5.2 TGIF2 controls pancreatic endocrine differentiation in the mouse embryo

Pancreatic-specific ablation of *Tgif2* results also in the perturbation of endocrine and exocrine lineage specification (Fig. 12), suggesting a possible role of this TF in the allocation of the two pancreatic fates in multipotent progenitors. Specifically, *Tgif*-mutant pancreata exhibit reduced endocrine progenitor cell formation and differentiation (Fig. 15). Consequently, the content of mature INSULIN^{+} cells is lower compared to control pancreata (Fig. 16). Glucagon-expressing cells, instead, are not significantly affected in numbers, but form cords which appear to be stuck inside the main epithelial trunk. Nevertheless, the total number of hormone releasing cells is higher than expected considering the dramatic reduction of endocrine NGN3+ precursor pool, therefore it is likely that mechanisms of compensation take place during development to restore the normal content of hormone-releasing cells. It remains to be addressed whether maturation of endocrine cells is impaired at birth and whether adult animals develop hyperglycaemia or any form of pre-diabetic state.

Concomitantly, an expansion of acinar compartment was observed in the mutant pancreata (Fig. 17). Given the fact that differentiation of the pancreatic lineages is highly dynamic at early stages of development, with MPC being able to differentiate along either the endocrine or acinar fate until E14.5, it is conceivable that TGIF2 controls the balance of such cell fate choice during the first or the secondary transition of pancreas development. The fact that *Tgif2* expression is abundant at the tips of the developing pancreas (Fig. 11), which normally contributes to the acinar lineage, suggests an inhibitory activity of the TF towards the acinar fate. Importantly, a proliferative advantage of acinar cells can be excluded as no difference in cell mitosis or cell death was detected (Fig.18). Also, organ size as well as gene expression in MPC or ductal cells (*e.g. Hnf1b, Nkx6.1, Pdx1, Prox1, Sox9*) were unchanged between controls and mutants at E12.5, supporting the hypothesis that loss of *Tgif2* substantially impairs subsequent cell differentiation.

The comparison of the phenotypes between single and compound mutants suggests that TGIF1 has a role in these developmental processes (Fig. 15; 17). This is not surprising as TALE homeoproteins are known to interact with each other and form complexes to exert their biological functions[423]. One can hypothesize that their combinatorial activity is required for proper endocrine differentiation and that the absence of one of the two TFs prevents the activation of the heterodimer complex, provoking endocrine defects (Fig. 15). On the other

hand, the acinar fate differentiation occurs normally when one of the two TFs is present (Fig. 17), suggesting possible functional redundancy between the two TFs. However, since *Tgif1* is knocked out also in the surrounding mesenchyme, a non-cell autonomous role of TGIF1 on pancreatic lineage decision cannot be excluded.

Several TFs have been described to play distinct roles during pancreas organogenesis, dependently on the embryonic stage and the cellular context[1,42,83]. For instance, PTF1A is crucial for establishing the pancreatic progenitor identity forming a boundary with other organ-specific endodermal regions, while at later stages it regulates the acinar identity and differentiation[104,150]. Likewise, TGIF2 might function as a pancreatic developmental regulator driving binary fate choices. In first place, it restricts the identity of endodermal progenitors in the foregut to establish a pancreatic identity at expenses of the hepatic one. After lineage segregation, TGIF2 becomes crucial in maintaining the pancreatic identity and plays a role in the endocrine *versus* acinar cell fate decision (Fig. 31).

Overall, the experiments described in this thesis dissect the tissue-specific requirements of TGIF2 in pancreas development, uncovering a role in maintaining the pancreatic identity and regulating the endocrine differentiation cascade. These findings need further investigation in *ex vivo* systems and might be harnessed to improve maturation toward hormone-releasing endocrine cell types in *in vitro* differentiation processes.

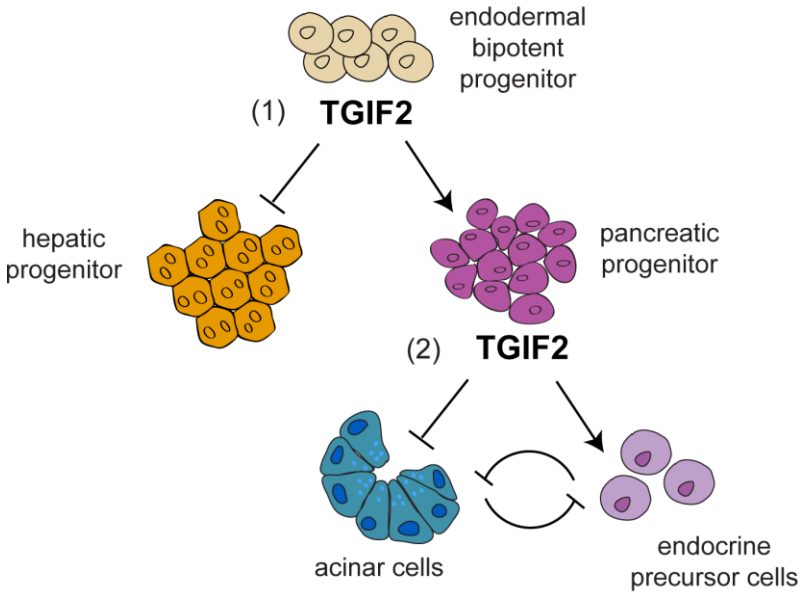

Figure 31. Model summarizing the *in vivo* functions of TGIF2 during pancreas development.
TGIF2 controls binary choices at multiple steps of pancreas organogenesis. (1) During the patterning of the endoderm, TGIF2 inhibits the hepatic fate and promotes the establishment of the pancreatic cell identity. (2) Within the pool of multipotent pancreatic cells, TGIF2 plays a role in the allocation of the pancreatic cell types by promoting endocrine differentiation and limiting the expansion of the acinar compartment.

5.3 Conservation of TGIF2-mediated liver-to-pancreas conversion in human cells

Developmental factors acting at the branchpoint between alternate lineages are particularly suitable for lineage reprogramming[71,430]. The regulatory activity of TGIF2 of pancreas *versus* liver cell decision makes it an ideal candidate for reprogramming cell identity[16,71]. Consistently with *in vivo* observations, previous gain-of-function *in vitro* studies demonstrated that TGIF2 has a liver-to-pancreas reprogramming function in the mouse. Specifically, lentiviral overexpression of TGIF2 is sufficient to erase the genetic program of mouse liver cells and redirect it to a pancreatic fate[71]. The second aim of my Ph.D. studies was to extend these findings in humans, as this knowledge would be extremely valuable to generate pancreatic cell types for therapeutic purposes.

The human and mouse TGIF2 protein sequences share a high degree (94%) of similarity[297]. Moreover, stable expression of the human TGIF2 ORF in mouse liver cells recapitulates the liver-to-pancreas conversion process observed in previous *in vitro* experiments, demonstrating a conservation of the functionality of the gene across the two species (Fig. 22).

After establishing the culture conditions to grow primary human hepatocytes (Fig. 23), I performed viral transduction experiments to ectopically express the human TGIF2 ORF in human adult liver cells (Fig. 24). TGIF2 repressed typical hepatic genes, including *ALBUMIN*, *TTR*, *HNF4A* and *APOA2*. In particular, *G6PC* (Glucose-6-phosphatase), encoding a membrane protein, which functions in gluconeogenesis and glycogenolysis, was the gene that most dramatically responded to the reprogramming factor. This might imply a major impact on the glucose metabolism during TGIF2-mediated reprogramming that needs to be studied.

Importantly, lentiviral-mediated introduction of TGIF2 in the hepatic cell culture turned on the expression of key pancreatic genes, including *PDX1, PTF1A, NKX6.1* and *SOX9* (Fig. 24). Differentiation markers such as *NGN3, NEUROD1, MAFA, PAX6* and *ARX* were also induced at high levels, indicating a mixed progenitor/differentiation state of the resulting cell culture (Fig. 24; 25). Remarkably, the expression of the same endocrine genes was perturbed in the embryonic pancreas after *Tgif2* depletion (Fig. 12; 15), possibly suggesting that they are direct targets of TGIF2.

Identity of the reprogrammed cells

A good efficiency of transduction of the primary hepatocytes was achieved by using the lentiviral method (>90%), with high levels of the TF transgene being expressed (Fig. 24; 26). However not all the transduced cells successfully undergo cell fate conversion and the overall

efficiency of reprogramming in humans remains lower than in the mouse. In addition, the outcome of the process can be influenced by variability of the cellular source, underlying a certain predisposition to change identity depending on the donor's gender and age (Table 10).

The bulk population transcriptional analysis does not resolve potential subtle changes in gene expression or different transition states occurring in culture. In order to determine the composition of the culture and possible heterogeneity, an extensive characterization of the phenotype of the human TGIF2-transduced cells is necessary. A thorough single-cell transcriptome analysis (scRNA-seq) would allow to fully appreciate cellular heterogeneity during reprogramming of hepatocytes into pancreatic cells, discover cell-state-specific genes and alternative "off-target" programs emerging during reprogramming. This would help to better understand the reasons why a full or more efficient conversion between the two cellular states fails and to control the cell identity switch. For instance, the lack of induction of pancreatic genes, as assessed by RT-qPCR, in a sub-set of the TGIF2-mediated reprogramming experiments (Table 10) might be explained by the heterogeneous cellular response to the reprogramming and the limited sensitivity of the bulk approaches used.

Finally, reprogrammed cells will need to be compared to their *in vivo* counterparts to assess how similar and interchangeable they are. Specifically, surrogate pancreatic β-cells induced by lineage reprogramming should undergo deep phenotyping at the transcriptome, epigenome, and functional levels (e.g., insulin content, secretion, and dynamic response to blood glucose)[431].

Improving and tuning the liver-to-pancreas reprogramming

Overall, the results described in this thesis support conservation of the reprogramming role of TGIF2 between mouse and human. However, transposing onto human model systems the findings obtained in the mouse have required important modifications of the experimental conditions[430] that still need to be improved. Reprogramming strategies are often based on the combination of multiple TFs to superimpose the program of the desired cell type; such combinations can either facilitate stepwise conversion from a progenitor to a mature state, for example with one TF being involved in the initial fate specification and the other one in subsequent maturation, or, alternatively, one of the TFs might act as a repressor to erase the original cellular identity[15,346,422,430]. Also pioneer TFs are particularly relevant in the context of direct reprogramming, as they act as master regulators of cell fate during embryonic development via their interaction with chromatin and in cooperation with lineage-specific TFs, including FOXA, GATA4, C/EBPα, ASCL1[347,425,432,433].

In an effort to improve the TGIF2-mediated reprogramming activity to obtain more homogeneous and mature induced pancreatic progenitor cells, I tested modified versions of

“reprogramming factors cocktail” and various culture conditions (see Result sections 4.2.5 and 4.2.6). In my experimental settings, ectopic expression of TGIF2 together with other relevant pro-pancreatic factors, such as PDX1 and NKX6.1, as well as the PMN transgenic cassette did not enhance the conversion process (Fig. 27). No improvement of the reprogramming was obtained either by the concomitant expression with another TALE homeoprotein, PBX1 (Fig. 27), which has been described as a pioneer factor[280] as well as co-factor of TGIF proteins in hetero-multimeric complexes[246] and identified as a crucial factor during embryonic development of pancreas (Cozzitorto C. et al., manuscript under revision). This might be due to species-specific differences between mouse and human, being these TFs mostly studied in the mouse. Alternatively, the experimental conditions and/or transgene dosage might still need to be improved to attain full conversion.

Finally, I assessed different culture conditions with the purpose of further improving the reprogramming process (see Result section 4.2.5). For instance, I withdrew typical hepatic growth factors from the culture medium or exposed the transduced cells to the cytokines that drive the differentiation of pluripotent stem cells into pancreatic β-like cells[240,416]. However, such manipulations of the culture conditions were also not beneficial for the reprogramming to a more stable pancreatic fate (Table 10). Future studies are required to define a pro-pancreatic environment and identify an appropriate combination of TFs to further stabilize the fate of the converted cells in humans and push it toward fully mature insulin producing β-cells, which would provide an extremely valuable source for clinical applications.

Having access to reliable and abundant sources of human hepatocytes has been an impediment for many fields of research and clinical therapies. Even though methods for hepatocyte isolation have improved, resulting into higher yields of viable cells, to culture them *in vitro* and for long-term remains problematic[434]. Indeed, adult hepatocytes are challenging to cultivate due to the lack of proliferative potential and the rapid loss of an hepatic differentiated phenotype in culture[21]. Moreover, high-quality primary hepatocytes are mainly reserved to the clinics for transplantation and less available for research. In order to overcome the limitations derived by the shortage of donor material and the difficulties in culturing, alternative human hepatocytes *ex vivo* models could be considered to further investigate the hepato-pancreatic plasticity and TGIF2-mediated reprogramming. In particular, hepatic cells obtained upon differentiation of iPSCs would represent a renewable and expandable cellular source[435] that could be used also for biochemical approaches to elucidate TGIF2 mechanism of action. In addition, 3D organoid cultures can help to improve long-term culture and expansion of adult human liver cells and also to develop human liver tissue models[436]. These models will move forward the *ex vivo* reprogramming efforts in human cells or organ-like contexts[16].

Finally, the key for the optimization of the reprogramming process might reside in the precise understanding of the developmental function of the TGIF2 during human pancreas

organogenesis. By applying a differentiation protocol to derive *in vitro* β-like cells from PSCs[240,416], I found that the endogenous expression of the human TGIF2 gene increases over time during pancreatic differentiation (Fig. 21). In the adult, its expression is more abundant in the pancreas compared to the liver, mirroring the situation in the mouse[71]. However, whether the intrinsic developmental activity exerted by TGIF2 is the same between mouse and human is not known yet and deserves further investigation. *In vitro* differentiation models, which closely recapitulated the *in vivo* organogenesis of pancreas and liver, would be useful to address this question. For instance, systems based on differentiation of pluripotent stem cells toward the two endodermal lineages could be instrumental to investigate whether TGIF2 overexpression further promotes the acquisition of the pancreatic and, more specifically, endocrine state and repress the acquisition of the hepatic fate as previously shown in the mouse[71].

5.4 TGIF2 activates a pancreatic transcriptional cascade in different *ex vivo* cellular systems

The third aim of my thesis was to determine if TGIF2 reprogramming activity is endoderm lineage-restricted or applies to multiple contexts in terms of cell origin and differentiation state. To explore the reprogramming potentials of TGIF2 outside the endoderm and better understand its mechanisms of action, I have used somatic cell types other than the hepatic ones.

Lentiviral-mediated expression of *TGIF2* induces morphologic and transcriptional changes in mouse and human fibroblasts (Fig. 28; 29). In particular, the transduced cells acquire an epithelial ‘cuboidal’ shape and form patches of tightly packed cells. This phenomenon is accompanied by induced gene expression of the epithelial marker *CDH1*, the endoderm gene *SOX17* and fundamental pancreatic identity markers, including *PDX1* and *PTF1A* (Fig. 27; 28). Experiments aiming at improving the reprogramming process uncovered the decisive positive effect of the 3D culture (Fig. 30). Indeed, culturing TGIF2-transduced cells as 3D clusters resulted in higher levels of expression of *CDH1* and pancreatic markers, including *PDX1, PTF1A, SOX9, NKX6.1, NEUROD1* and *MAFA* (Fig. 30). Interestingly, the induction of endocrine markers gene expression is in line with the reprogramming experiments performed with hepatocytes and the downregulation of the same TFs in *Tgif2*-deficient pancreata (see Result sections 4.1.2 and 4.2.3).

The TGIF2-reprogramming approach requires optimization under multiple points of view. First of all, the lentiviral transduction needs fine-tuning, to avoid cell toxicity when using high

doses of virus or excessive culture heterogeneity in case of low amount of virus. Indeed, lentiviral transduction of a small subpopulation of fibroblasts often results in a mixed cell culture, since untransduced fibroblasts have proliferative advantage and take over at late time points. Attempts of enriching in TGIF2-GFP$^+$ populations by FACS, not shown in this thesis, have not been successful in post-sorting culture, with cells undergoing stress and, subsequently, cell death. Also, disruption of cell adhesive junctions seems to be not beneficial for the overall reprogramming, therefore new methods to isolate and characterize the patches of TGIF2-induced epithelial cells should be considered. In addition, the use of a *PDX1*-transgenic reporter cell line to select and enrich the population transitioning to a pancreatic progenitor state might be an important step to undertake.

Of note, the correlation between histone modifications at target genes and reprogramming potential should be taken into account, as it was shown, for instance, to be determinant in the success rate of ASCL1-mediated reprogramming of fibroblasts into neurons[437]. Therefore, epigenome remodelling might help to increase permissiveness to reprogramming[352].

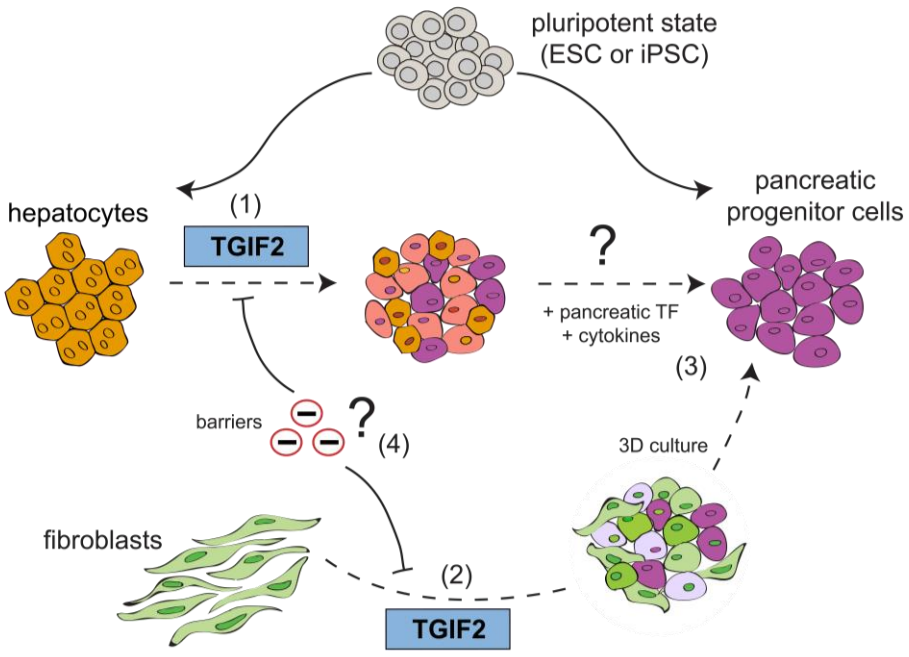

Figure 32. Reprogramming activities of TGIF2.
Ectopic expression of the single factor TGIF2 can convert liver cells **(1)** and fibroblasts **(2)** to a pancreatic progenitor state, behaving as a broad reprogramming factor. In both circumstances, the resulting cell culture is composed of mixed cell types, where only a subset of cells turns off the original program and activate the new pancreatic one**. (3)** 3D culture, additional pancreatic TFs and cytokines have been assessed for their potential to improve the process and further facilitate the maturation of the reprogrammed cell products. **(4)** The reprogramming processes are hampered by yet unidentified barriers (e.g. epigenetic), which need to be overcome to increase the efficiency of cell fate conversion.

Collectively, my studies highlight the ability of TGIF2 in destabilizing lineage commitment of multiple adult somatic cells and inducing a pancreatic identity, even across distant germ layers (Fig. 32). Comparing genome-wide transcriptional changes in the different *ex vivo* lineage reprogramming systems, namely mouse and human hepatic and non-hepatic cells, would provide valuable information about similarities or differences in how cell identity is locked across species and various lineages. Moreover, a thorough analysis of transcriptome, proteome and chromatin marks, ideally at the single cell level, would shed light on the identity of the cell products as well as the depth of reprogramming and frequency of incomplete reprogramming[438].

5.5 Pathways involved in the TGIF2-dependent acquisition of pancreatic identity

Previous RNA-seq analysis of hepatic and pancreatic progenitors performed by our group unveiled the important role of the non-canonical Wnt/PCP pathway in the liver *versus* pancreas lineage divergence[70]. Interestingly, several evidence point at a direct role of TGIF2 in the activation of this pathway to establish the pancreatic cell identity[71]. Components of the PCP core, which characterize the pancreatic progenitor state during endodermal patterning[70], were downregulated in embryonic pancreata lacking *Tgif(s)* (Fig. 20). Conversely, forced expression of TGIF2 in hepatic cells elicit the expression of members of the PCP pathway[71], including CELSR3, underlying a role of the remodelling of epithelial cell polarity in the liver-to-pancreas cell conversion. The fact that the expression of factors establishing planar polarization is induced by TGIF2 also in non-endodermal cell types (e.g. fibroblasts) undergoing pancreatic conversion (Fig. 29) underpin a general solid link between this pathway and the acquisition of the pancreatic fate. Overall, reprogramming experiments in mouse and human cells presented in this thesis are in accordance with the *in vivo* data and support the concept that the non-canonical Wnt/PCP pathway is a signature of pancreatic progenitor identity in a conserved way among vertebrates and its activation is dependent on TGIF2 during both pancreas development and lineage reprogramming. How TGIF2 modulates this pathway remains an open question that deserves closer investigation.

In addition, the impairment of the PCP pathway activity caused by *Tgif2* depletion might also explain defects in the endocrine cell formation observed in mutant embryonic pancreata (Fig. 16). Consistently, mutations of *Celsr2/3* have been shown to lead to decreased endocrine differentiation in the mouse[414]. Moreover, the correct apical localization of VANGL at the membrane of pancreatic ductal cells is crucial for epithelial integrity and proper tubulogenesis processes during pancreas development[439]. These findings highlight the essential role of PCP

components as regulators linking pancreas morphogenesis with cell fate. Of note, Wnt/PCP signalling underlies functional β-cell heterogeneity and induce β-cell maturation[440]. A new effector of the Wnt/PCP pathway, named *Fltp* (aka *Flattop* and *Cfap126*), has been recently identified as a marker of non-proliferative mature post-natal β-cells, also in the human endocrine cellular context[440].

In *Tgif2*-deficient pancreata, streams of endocrine clusters, mostly composed of glucagon+ cells, were found in the main pancreatic ducts (Fig. 16). This morphogenetic defect might be the result of a perturbation of extracellular cues, for example involving the Wnt5a/Frizzled-2 ligand receptor couple[441]. This pathway indeed provides the right environment and signals for INSULIN$^+$ cell migration during islet formation in the Zebrafish and mouse, since genetic ablation of *Wnt5a* in mouse embryos results in increased association of pancreatic islets with ducts[441]. Additionally, the axon guidance pathway Neuropilin/Plexin/Semaphorin, which has been shown to control endocrine cell migration into the surrounding mesenchyme[415], was found reduced in the RNA-seq of *Tgif2*-deficient pancreata (Fig. 20). This might constitute a molecular cause for the endocrinogenesis altered upon *Tgif2* genetic deletion.

Finally, the phenotypic characterization of embryonic *Tgif2* mutants revealed also epithelial structural abnormalities, such as the formation of dilated ductal structures or cysts within the pancreas (Fig. 19). This phenotype is generally associated with defects in ciliogenesis, as commonly observed in kidney polycystic diseases, in part as consequence of the disruption of the PCP pathway[412]. Indeed, PCP pathway controls morphogenetic processes such as oriented cell division, convergent extension and ciliary orientation[413], providing therefore the most likely explanation for the observed impaired tubulogenesis in the absence of *Tgif2*. Alternatively, it is known that primary cilia formation is regulated also by Hedgehog (HH) pathway[442]. A direct repressive regulation of TGIF2 on the HH effector *Gli3* gene has been reported in another study[443] in a different cellular context, however I did not detect any changes in *Gli* expression at the transcriptome level in the RNA-seq analysis. Nevertheless, an *in situ* approach would elucidate better if changes in gene expression are more obvious in specific cell types, such as the ductal or the endocrine compartments.

5.6 Mechanisms of action of TGIF2

My studies of TGIF2 in the mouse model system and *ex vivo* cultures of primary cells supported its activity as a repressor of the hepatic state and activator of the pancreatic progenitor program[71] (Fig. 33A). However, the molecular mechanisms underlying TGIF2 biological function within the endoderm and during reprogramming remain elusive. TALE homeoproteins can be activators or repressors in a context-dependent manner in several biological processes[423]. In particular, TGIF2 has been mainly described as a transcriptional

repressor of the TGFβ/BMP signalling[295]. It is conceivable that, during endoderm patterning, TGIF2 acts as an intrinsic inhibitor of BMP signalling by sequestering BMP-activated intracellular mediators SMADs (1/5/8) to exclude this signalling from the prospective pancreatic endoderm and, consequently, repress the liver fate[63,71]. However, the cell-autonomous role of BMP/SMADs in pancreas development is not fully understood[129,147,148]. Some studies reported TGFβ/SMAD2/3 as key regulator of β-cell differentiation[148,444], being expressed in the nuclei of endocrine cells where they promote proliferation and prevent premature maturation[445]. Instead, the common TGFβ/BMP mediator SMAD4 has been shown to control pancreatic islets size and maintain ductal differentiation in a model of pancreatic cancer[446,447]. Overall, the responses that BMP signalling mediate in the pancreas are very dynamic[129], making it difficult to infer a role for TGIF2 as BMP inhibitor at different stages of pancreas organogenesis.

TGIF2 also possesses a homeodomain for direct binding the DNA[293,295]. Comparing the consequences of TGIF2 ectopic expression in different biological contexts and the transcriptional hallmarks induced by TGIF2 in various cell types is fundamental to start unravelling the mechanism of action of this transcriptional regulator. For example, a common induction of master pancreatic genes, such as *PDX1*, in different cell types (e.g. hepatocytes and fibroblasts) suggest that TGIF2 acts as a transcriptional activator by possibly binding to specific gene regulatory regions or unlocking genomic loci through recruitment of chromatin modifiers (Fig. 33B). In line with this, preliminary results obtained in our laboratory reported a direct interaction of TGIF2 with *Pdx1* promoter in the mouse (Cerdá-Esteban N and Spagnoli FM, unpublished). Alternatively, the activation of *Pdx1* gene expression can be tuned in an indirect way, whereby TGIF2 might inhibit a yet unknown transcription repressor of *Pdx1* (Fig. 33B). In addition, TGIF2 could exert a direct repressive function over non-pancreatic genes (e.g. hepatic genes) by binding to regulatory regions in their genomic loci to limit gene expression (Fig. 33B).

It also remains to be addressed if the gene network guiding endocrine differentiation cascade and non-canonical Wnt/PCP pathway activation (Fig. 32A) is regulated by TGIF2 through similar mechanisms and/or in synergism with other TALE homeoproteins. Analysis on the rVista platform (https://rvista.dcode.org) showed putative TGIF2 binding sites (BS) in conserved regions in the genomic loci of *MafA* and *Ngn3*. A regulatory activity of TGIF2 on these predicted regions could provide a mechanistic explanation for the endocrine phenotype observed in the *in vivo* loss-of-function approach and the induction of these genes upon enforced expression of TGIF2 in *ex vivo* systems.

To investigate all these different hypotheses, genome-wide transcriptional DNA-binding sites of TGIF2 need to be characterized by chromatin immunoprecipitation combined with high-throughput sequencing (ChIP-Seq) approach. This type of analysis has been hampered so far by the lack of specific antibodies against TGIF2. For this purpose, I have tested and validated

by Western Blot many different commercially available anti-TGIF2 antibodies that will be used in future biochemistry assays (data not shown). As an alternative strategy for pulling down the TGIF2/chromatin complexes, an epitope tag (e.g. FLAG) sequence can be inserted at one extremity of the *Tgif2* coding sequence to be recognized by a suitable antibody. However, this approach does not allow to examine endogenous TGIF2 complexes. In summary, using ChIP-based approaches I will be able to decode the TGIF-dependent regulatory landscapes during reprogramming and pancreatic development.

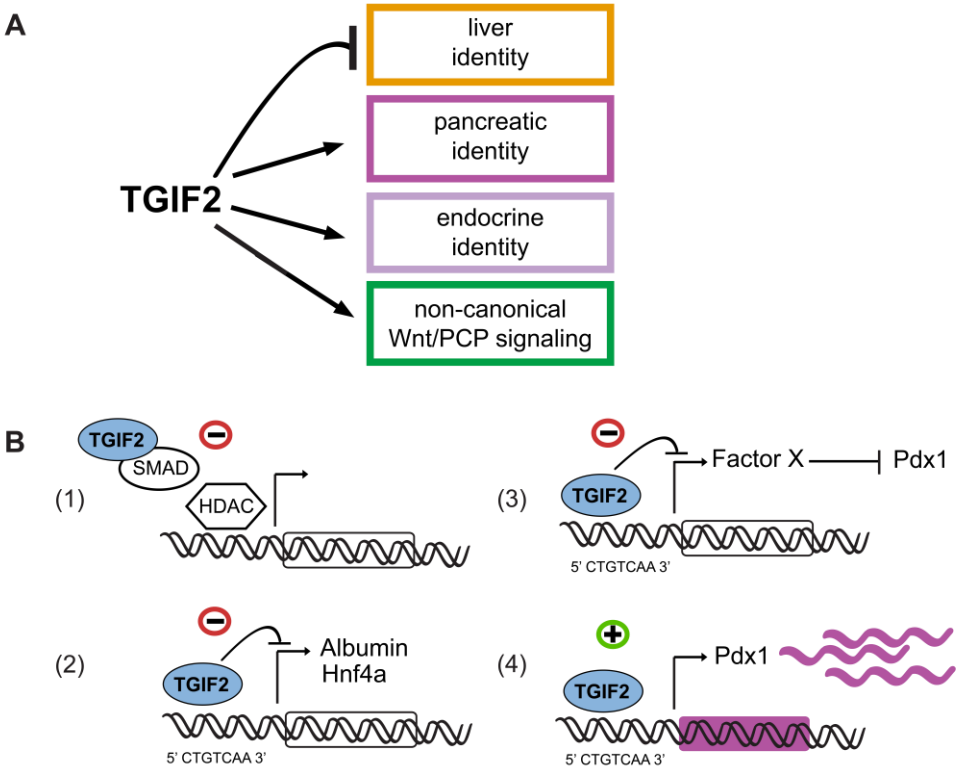

Figure 33. Mechanisms of action of TGIF2.
(A) Summary of the biological effects of TGIF2 in *in vivo* and *in vitro* systems. **(B)** Modes of activity of TGIF2. As previously reported, TGIF2 can sequester BMP-activated SMADs and recruit HDAC proteins to genomic loci with consequent inactivation of gene expression (1). In alternative scenarios, TGIF2 can directly bind to regulatory regions nearby hepatic genes and repress their activation (2), or TGIF2 can exert repression on yet undefined inhibitor factors of pancreatic specific genes (3). Last, TGIF2 can induce direct activation of *Pdx1* and other pro-pancreatic TFs by occupying regulatory regions in their proximity (4).

5.7 TGIF2-mediated reprogramming as cell replacement strategy to treat diabetes

Diabetes is a life-threatening multifactorial disorder caused by destruction or dysfunction of insulin-dependent β-cells[5,448]. The worldwide prevalence of diabetic patients is continuously rising and a definitive cure for them is not currently available. Hopes for regenerative medicine approaches for β-cell-replacement and glycaemic control restoration rely on the understanding of the ontogeny of pancreatic cell lineages[7,16,355]. Indeed, deciphering and recapitulating the developmental cues that govern transcriptional networks and establish the epigenetic state of the desired pancreatic cell type is instrumental to generate functional *de novo* therapeutic cell products. In particular, over the last decades, direct lineage reprogramming has emerged as a promising strategy to generate transplantable pancreatic β-cells *in vitro* or directly *in situ* in the host organism[16,355].

The liver and the pancreas share many similarities, including a close embryonic origin, the expression of a set of transcriptional factors and the control of the body metabolic function in the adult[16]. Moreover, the high regenerative ability and accessibility of the adult liver makes it an ideal renewable source for inducing β-cell properties[16,21], whereas the opposite is not conceivable, being the pancreas in an inaccessible anatomical location and with a very limited capacity for regeneration[13,320–325]. For all these reasons, the liver represents a privileged tissue to target for generating new β-like cells through direct lineage reprogramming. Harnessing replication potential of the hepatocytes and the intercellular plasticity between the two cell types to obtain therapeutic patient-specific β-cells through direct lineage reprogramming represents an advantageous route to ensure normoglycaemia avoiding long-term immunosuppression treatments of the recipients[16].

TGIF2 presents unique properties to convert cell identity. The results presented in this thesis indicate that the reprogramming activity of TGIF2 is conserved in the humans and might be the starting point to define a possible strategy to generate therapeutic pancreatic cell types. However, several considerations must be taken into account in order to turn these results into a real clinical approach. In view of a safe therapeutic approach, the delivery technique of the genetic material must be thoughtfully chosen. Among the most common strategies to achieve gene therapy are the lentiviral and the AAV methods. AAV vectors are considered safe systems, since they mediate efficient and tissue-specific gene transduction *in vivo* in various dividing and quiescent cell types, without causing any known pathogenicity[421]. Nevertheless, lentiviral vector-mediated gene therapy has also been employed to engineer hematopoietic stem cells to correct primary immunodeficiencies and red blood cell diseases [449]. For instance, it has been tested in phase I/II clinical trials to treat SCID-X1 disease[450].

In the reprogramming of human primary hepatocytes described here, I compared both gene transfer systems. Notably, the lentiviral approach, which ensures stable integration of the transgene in the genome of the host cell, conferred a more robust hepatic inhibition and pancreatic induction compared to the hepato-tropic AAV method, likely due to the lower amount of transgene induced in the latter case. Nevertheless, the possibility to temporally control the expression of the TF in a restricted window of time should be assessed to avoid undesired consequences deriving from the perpetual activity of the transgene beyond the reprogramming event.

Moreover, the reprogramming route needs to be refined, with the possibility to include multiple factors or eventually replace the transgenes with small molecules mimicking their activity. For this reason, it is important to understand the mode of action of TGIF2 at the molecular level. Future efforts will focus on the development of strategies to improve the functional maturation of converted cells. It is of primary importance indeed to investigate whether the reprogrammed cells integrate, expand and further mature in a proper *in vivo* environment, and are finally able to secrete insulin in response to glucose stimulation. Humanized models would be valuable to perform *in vivo* lineage reprogramming of human cells but also to assess the functionality of reprogrammed cells before any reprogramming strategy can be tested in patients.

Overall, experiments using human primary hepatocytes and fibroblasts described in this thesis endorse the conserved reprogramming activity of TGIF2 across mammalian species. These promising data could serve as a starting point to develop cell replacement strategies for diabetes based on lineage reprogramming of abundant and easily accessible autologous patient-derived cell types as source, like skin or liver biopsies, into pancreas.

6. ABBREVIATIONS

2A	Self-cleaving 2A peptide
3D	Three-dimension
AAV	Adeno-associated virus
Ad	Adenovirus
ALK5	Activin receptor-like kinase 5
ALK5i	ALK5 inhibitor
ATP	Adenosine triphosphate
Ascl1	Achaete-Scute Family BHLH Transcription Factor
Aka	also known as
A/P	Anterior-Posterior
Acvr2a	Activin receptor type IIA
Acvr2b	Activin receptor type IIB
AFP	α-fetoprotein
A/P	anterior-posterior
Arx	Aristaless-related homeobox
ATP	Adenosine triphosphate
BAML	Bipotential adult mouse liver cells
bFGF	basic Fibroblast Growth Factor
bHLH	basic helix-loop-helix
Bhlha15	basic helix-loop-helix family member a15
BMP	Bone morphogenetic protein
Bmpr1a	BMP receptor 1A
bp	base pairs
BP	Biological process
BS	Binding site
BSA	Bovine serum albumin
Ca^{2+}	Calcium
CAM	Cell Adhesion Molecule
CBP	CREB-binding protein
Cdh1	Cadherin-1
Cdx2	Caudal type homeobox 2
Celsr2	Cadherin EGF LAG seven-pass G-type receptor 2
Celsr3	Cadherin EGF LAG seven-pass G-type receptor 3
Cre	Cre recombinase
CtBP	C-terminal binding protein 1
Ctrl	Control
c-Myc	Myelocytomatosis oncogene (c-Myc)
ChIP	Chromatin immunoprecipitation
ChIP-Seq	Chromatin immunoprecipitation sequencing
Cpa1	Carboxypeptidase A1
CRBP II	Cellular Retinol-Binding Protein II
CXCR-4	C-X-C chemokine receptor type 4
CtBP	C-terminal-binding protein 1
CASD	Cell-activation and signaling-directed
DE	Definitive endoderm

DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic acid
Dnmt	DNA methyltransferase
dNTP	Deoxynucleotide triphosphate
DP	Dorsal pancreas
dpc	days post-coitum
DTT	Dithiothreitol
E	Embryonic day
e.g.	exempli gratia
E-cad	E-cadherin
ECM	Extracellular Matrix
EGF	Epidermal Growth Factor
EMT	Epithelial-to-mesenchymal transition
Eomes	Eomesodermin
ESC	Embryonic Stem Cells
F	Forward
FACS	Fluorescence-Activated Cell Sorting
FBS	Fetal bovine serum
FGF	Fibroblast Growth Factor
Fgf10	Fibroblast growth factor 10
FGFRB	Fibroblast Growth Factor Receptor 2B
FI	Fluorescence Intensity
Fig.	Figure
Foxa2	Forkhead box A2
FPKM	Fragments per kilobase of exon per million mapped fragments
Fzd	Frizzled
GAPDH	glyceraldehyde 3-phosphate dehydrogenase
GFP	Green Fluorescent Protein
GLIS3	GLIS family zinc finger 3
GLUT-2	glucose transporter 2
GO	Gene Ontology
hADF	Human Adult dermal fibroblasts
H2B	Histone H2B
HDAC	Histone deacetylase
HEK	Human Embryonic Kidney
HEP	Hepatic medium
HGF	Hepatocyte Growth Factor
Hes1	Hairy and enhancer of split
Hhex	Hematopoietically expressed homeobox gene
HLIM	Hepatic medium
Hnf1b	HNF1 homeobox B
Hnf4a	Hepatocyte nuclear factor 4-alpha
Hox	Homeobox
Iapp	Islet amyloid polypeptide
i.e.	id est (in other words)
IF	Immunofluorescence
IgG	Immunoglobulin G
Ins2	Insulin 2

Insm1	Insulinoma-associated 1
Ipf1	Insulin promoter factor 1
iPSCs	induced Pluripotent Stem Cells
IRX	Iroquois
ISH	In Situ Hybridization
Isl1	Islet1
Itga5	Integrin subunit α5
Itga6	Integrin subunit α6
ITR	
Klf4	Krüppel-like factor 4
KO	knock-out
LB	Luria Broth Base
LiCl	Lithium chloride
loxP	Locus of Crossover in P1
LV	Lentivirus
MAF	Mouse Adult Fibroblasts
Mafa	v-Maf musculoaponeurotic Fibrosarcoma oncogene family, protein A
Mafb	v-Maf musculoaponeurotic Fibrosarcoma oncogene family, protein B
MAPK	Mitogen-Activated Protein Kinase
MEIS	Myeloid Ecotropic Viral Integration Site
MET	mesenchymal-to-epithelial transition
$MgCl_2$	Magnesium chloride
Mist1	muscle, intestine and stomach expression 1
Mixl	Mix-like protein
Mnx1	Motor neuron and pancreas homeobox 1
MODY	Maturity-Onset Diabetes of the Young
MOI	Molteplicity Of Infection
MPCs	Multipotent pancreatic Progenitor Cells
mRNA	messenger Ribonucleic Acid
MYOD	Myogenic Differentiation 1
Myog	Myogenin
nd	not detected
NEAA	Non essential amino acids
NeuroD1	Neurogenic differentiation 1
Ngn3	Neurogenin-3
Nkx2.2	NK2 homeobox 2
Nkx3.2	NK3 homeobox 2
Nkx6.1	NK6 homeobox 1
Nkx6.2	NK6 homeobox 2
Nr5a2	Nuclear receptor subfamily 5, group A, member 2
Nrp	Neurpilin
Oct-4	Octamer-binding transcription factor 4
Onecut1	One cut domain 1
ORF	Open reading frame
Osr	Odd-Skipped Related
OSKM	Oct3/4, Sox2, Klf4 and c-Myc
Pax4	Paired box 4
Pax6	Paired box 6

PBX	Pre-B Cell Leukemia Homeobox
PCP	Planar Cell Polarity pathway
PCR	Polymerase Chain Reaction
Pdx1	Pancreatic and duodenal homeobox 1
PE	Pancreatic endoderm
PEC	Pancreatic endoderm cells
PEI	Polyethylenimine
PHH	Primary human hepatocytes
pHH3	Phosphorylated histone H3
PF	Posterior foregut
PMN	Pdx1, MafA, Ngn3
PP	Pancreatic progenitors
PREP	Pbx regulating protein
Prox1	Prospero homeobox 1
Ptf1a	Pancreatic specific Transcription Factor 1a
R	Reverse
RBPJ	Recombination signal binding protein for immunoglobulin kappa J region
RA	Retinoic Acid
RNA	Ribonucleic acid
RNA-seq	RNA-sequencing
Robo	Roundabout signaling pathway
ROCK	Rho-associated protein kinase
rpm	rotations per minute
RT	Reverse Transciption
RT-qPCR	Reverse Transcription-quantitative Polymerase Chain Reaction
sc-RNAseq	Single cell RNA sequencing
SEM	Standard error of the mean
SHH	Sonic Hedgehog
Slit	Slit guidance ligand
SMAD	Small body size Mothers Against Decapentaplegic
SMADs-R	Receptor regulated SMADs
SOX	Sex-determining region on Y box
Sox9	SRY-box 9
Sox17	SRY-box 17
SSC	Saline-sodium citrate buffer
STM	Septum transversum mesenchyme
T1D	Type 1 Diabetes
T1/T2 cdKO	Tgif1/Tgif2 compound double knock-out
T2D	Type 2 Diabetes
TALE	Three-Amino-Acid Loop Extension
Tbx3	T-box Transcription Factor 3
TF	Transcription Factor
TG	Transgene
TGFβ	Transforming Growth Factor Beta
TGFBR2	Transforming Growth Factor Beta Receptor
TGIF	TG-Interacting Factor
TGIF1	TG-Interacting Factor 1
TGIF2	TG-Interacting Factor 2

TSS	Transcription-Starting Site
TTR	Transthyretin
VP16	Herpes simplex virus protein vmw65
WHO	World Health Organization
Wnt	Wingless-type MMTV integration site family
wpc	weeks post-conception
WT	Wild Type
YAP	yes-associated protein

7. LIST OF FIGURES

8. LIST OF TABLES

9. BIBLIOGRAPHY

1. Pan, F. C. & Wright, C. Pancreas organogenesis: from bud to plexus to gland. *Dev. Dyn.* **240**, 530–65 (2011).
2. Shih, H. P., Wang, A. & Sander, M. Pancreas Organogenesis: From Lineage Determination to Morphogenesis. *Annu. Rev. Cell Dev. Biol.* **29**, 81–105 (2013).
3. Bonner-Weir, S., Sullivan, B. A. & Weir, G. C. Human Islet Morphology Revisited. *J. Histochem. Cytochem.* **63**, 604–612 (2015).
4. Nolan, C. J., Damm, P. & Prentki, M. Type 2 diabetes across generations: From pathophysiology to prevention and management. in *The Lancet* (2011). doi:10.1016/S0140-6736(11)60614-4
5. Weir, G. C. & Bonner-Weir, S. Islet β cell mass in diabetes and how it relates to function, birth, and death. *Ann. N. Y. Acad. Sci.* **1281**, 92–105 (2013).
6. Fu, Z., R. Gilbert, E. & Liu, D. Regulation of Insulin Synthesis and Secretion and Pancreatic Beta-Cell Dysfunction in Diabetes. *Curr. Diabetes Rev.* **9**, 25–53 (2012).
7. Zhou, Q. & Melton, D. A. Pancreas regeneration. *Nature* **557**, 351–358 (2018).
8. Sneddon, J. B. *et al.* Stem Cell Therapies for Treating Diabetes: Progress and Remaining Challenges. *Cell Stem Cell* **22**, 810–823 (2018).
9. Bakhti, M., Böttcher, A. & Lickert, H. Modelling the endocrine pancreas in health and disease. *Nature Reviews Endocrinology* **15**, 155–171 (2019).
10. Pociot, F. & Lernmark, Å. Genetic risk factors for type 1 diabetes. *The Lancet* **387**, 2331–2339 (2016).
11. Fajans, S. S., Bell, G. I. & Polonsky, K. S. Molecular Mechanisms and Clinical Pathophysiology of Maturity-Onset Diabetes of the Young. *N. Engl. J. Med.* **345**, 971–980 (2001).
12. Johannesson, B., Sui, L., Freytes, D. O., Creusot, R. J. & Egli, D. Toward beta cell replacement for diabetes. *EMBO J.* **34**, 841–855 (2015).
13. Ellis, C., Ramzy, A. & Kieffer, T. J. Regenerative medicine and cell-based approaches to restore pancreatic function. *Nature Reviews Gastroenterology and Hepatology* (2017). doi:10.1038/nrgastro.2017.93
14. Duncan, A. W., Dorrell, C. & Grompe, M. Stem Cells and Liver Regeneration. *Gastroenterology* **137**, 466–481 (2009).
15. Heinrich, C., Spagnoli, F. M. & Berninger, B. In vivo reprogramming for tissue repair. *Nature Cell Biology* (2015). doi:10.1038/ncb3108
16. Ruzittu, S., Willnow, D. & Spagnoli, F. M. Direct Lineage Reprogramming: Harnessing Cell Plasticity between Liver and Pancreas. *Cold Spring Harb. Perspect. Biol.* a035626 (2019). doi:10.1101/cshperspect.a035626

17. Si-Tayeb, K., Lemaigre, F. P. & Duncan, S. A. Organogenesis and Development of the Liver. *Developmental Cell* (2010). doi:10.1016/j.devcel.2010.01.011
18. Abdel-Misih, S. R. Z. & Bloomston, M. Liver anatomy. *Surg. Clin. North Am.* **90**, 643–53 (2010).
19. Gordillo, M., Evans, T. & Gouon-Evans, V. Orchestrating liver development. *Development (Cambridge)* **142**, 2094–2108 (2015).
20. Ben-Moshe, S. & Itzkovitz, S. Spatial heterogeneity in the mammalian liver. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 395–410 (2019).
21. Michalopoulos, G. K. Liver regeneration. *J. Cell. Physiol.* **213**, 286–300 (2007).
22. Yanger, K. *et al.* Adult hepatocytes are generated by self-duplication rather than stem cell differentiation. *Cell Stem Cell* **15**, 340–349 (2014).
23. Furuyama, K. *et al.* Continuous cell supply from a Sox9-expressing progenitor zone in adult liver, exocrine pancreas and intestine. *Nat. Genet.* **43**, 34–41 (2011).
24. Font-Burgada, J. *et al.* Hybrid Periportal Hepatocytes Regenerate the Injured Liver without Giving Rise to Cancer. *Cell* **162**, 766–779 (2015).
25. Wang, B., Zhao, L., Fish, M., Logan, C. Y. & Nusse, R. Self-renewing diploid Axin2 + cells fuel homeostatic renewal of the liver. *Nature* **524**, 180–185 (2015).
26. Raven, A. *et al.* Cholangiocytes act as facultative liver stem cells during impaired hepatocyte regeneration. *Nature* **547**, 350–354 (2017).
27. Lu, W. Y. *et al.* Hepatic progenitor cells of biliary origin with liver repopulation capacity. *Nat. Cell Biol.* **17**, 973–983 (2015).
28. Tarlow, B. D. *et al.* Bipotential adult liver progenitors are derived from chronically injured mature hepatocytes. *Cell Stem Cell* **15**, 605–18 (2014).
29. Deng, X. *et al.* Chronic Liver Injury Induces Conversion of Biliary Epithelial Cells into Hepatocytes. *Cell Stem Cell* **23**, 114-122.e3 (2018).
30. Rossant, J. & Tam, P. P. L. Blastocyst lineage formation, early embryonic asymmetries and axis patterning in the mouse. *Development* **136**, 701–713 (2009).
31. D'Amour, K. A. *et al.* Production of pancreatic hormone–expressing endocrine cells from human embryonic stem cells. *Nat. Biotechnol.* **24**, 1392–1401 (2006).
32. Stainier, D. Y. R. A glimpse into the molecular entrails of endoderm formation. *Genes and Development* **16**, 893–907 (2002).
33. Wells, J. M. & Melton, D. A. Vertebrate Endoderm Development. *Annu. Rev. Cell Dev. Biol.* **15**, 393–410 (1999).
34. Zorn, A. M. Liver development, Stembook, http://www.stembook.org. *Sci. Collab. Framew.* 1–26 (2008). doi:10.3824/stembook.1.25.1
35. Zorn, A. M. & Wells, J. M. Vertebrate Endoderm Development and Organ Formation. *Annu. Rev. Cell Dev. Biol.* **25**, 221–251 (2009).

36. Spagnoli, F. M. From endoderm to pancreas: a multistep journey. *Cell. Mol. Life Sci.* **64**, 2378–90 (2007).
37. Mcgrath, P. S. & Wells, J. M. SnapShot: GI tract development. *Cell* (2015). doi:10.1016/j.cell.2015.03.014
38. Poulain, M., Fürthauer, M., Thisse, B., Thisse, C. & Lepage, T. Zebrafish endoderm formation is regulated by combinatorial Nodal, FGF and BMP signalling. *Development* **133**, 2189–2200 (2006).
39. Wells, J. M. & Melton, D. A. Early mouse endoderm is patterned by soluble factors from adjacent germ layers. *Development* **127**, 1563–1572 (2000).
40. Tremblay, K. D. & Zaret, K. S. Distinct populations of endoderm cells converge to generate the embryonic liver bud and ventral foregut tissues. *Dev. Biol.* **280**, 87–99 (2005).
41. Spagnoli, F. M. From endoderm to pancreas: A multistep journey. *Cellular and Molecular Life Sciences* (2007). doi:10.1007/s00018-007-7184-x
42. Jørgensen, M. C. *et al.* An illustrated review of early pancreas development in the mouse. *Endocr. Rev.* **28**, 685–705 (2007).
43. Hebrok, M., Kim, S. K. & Melton, D. A. Notochord repression of endodermal sonic hedgehog permits pancreas development. *Genes Dev.* **12**, 1705–1713 (1998).
44. Kim, S. K., Hebrok, M. & Melton, D. A. Notochord to endoderm signaling is required for pancreas development. *Development* **124**, 4243–4252 (1997).
45. Stafford, D. & Prince, V. E. Retinoic acid signaling is required for a critical early step in zebrafish pancreatic development. *Curr. Biol.* **12**, 1215–1220 (2002).
46. Martín, M. *et al.* Dorsal pancreas agenesis in retinoic acid-deficient Raldh2 mutant mice. *Dev. Biol.* **284**, 399–411 (2005).
47. Thomas, M. K., Lee, J. H., Rastalsky, N. & Habener, J. F. Hedgehog signaling regulation of homeodomain protein islet duodenum homeobox-1 expression in pancreatic β-cells. *Endocrinology* **142**, 1033–1040 (2001).
48. Lau, J. & Hebrok, M. Hedgehog signaling in pancreas epithelium regulates embryonic organ formation and adult β-cell function. *Diabetes* **59**, 1211–1221 (2010).
49. Lammert, E., Cleaver, O. & Melton, D. Induction of pancreatic differentiation by signals from blood vessels. *Science (80-.).* **294**, 564–567 (2001).
50. Villasenor, A. & Cleaver, O. Crosstalk between the developing pancreas and its blood vessels: An evolving dialog. *Seminars in Cell and Developmental Biology* **23**, 685–692 (2012).
51. Jacquemin, P. *et al.* An endothelial-mesenchymal relay pathway regulates early phases of pancreas development. *Dev. Biol.* **290**, 189–199 (2006).
52. Ober, E. A. & Lemaigre, F. P. Development of the liver: Insights into organ and tissue

morphogenesis. *Journal of Hepatology* **68**, 1049–1062 (2018).

53. Sosa-Pineda, B., Wigle, J. T. & Oliver, G. Hepatocyte migration during liver development requires Prox1. *Nat. Genet.* **25**, 254–255 (2000).
54. Zaret, K. S. From Endoderm to Liver Bud: Paradigms of Cell Type Specification and Tissue Morphogenesis. in *Current Topics in Developmental Biology* **117**, 647–669 (Academic Press Inc., 2016).
55. Bort, R., Martinez-Barbera, J. P., Beddington, R. S. P. & Zaret, K. S. Hex homeobox gene-dependent tissue positioning is required for organogenesis of the ventral pancreas. *Development* **131**, 797–806 (2004).
56. Watt, A. J., Zhao, R., Li, J. & Duncan, S. A. Development of the mammalian liver and ventral pancreas is dependent on GATA4. *BMC Dev. Biol.* **7**, 37 (2007).
57. Burke, Z. & Oliver, G. Prox1 is an early specific marker for the developing liver and pancreas in the mammalian foregut endoderm. *Mech. Dev.* **118**, 147–155 (2002).
58. Deutsch, G., Jung, J., Zheng, M., Lóra, J. & Zaret, K. S. A bipotential precursor population for pancreas and liver within the embryonic endoderm. *Development* **128**, 871–881 (2001).
59. Angelo, J. R., Guerrero-Zayas, M.-I. & Tremblay, K. D. A fate map of the murine pancreas buds reveals a multipotent ventral foregut organ progenitor. *PLoS One* **7**, e40707 (2012).
60. Chung, W. S., Shin, C. H. & Stainier, D. Y. R. Bmp2 Signaling Regulates the Hepatic versus Pancreatic Fate Decision. *Dev. Cell* **15**, 738–748 (2008).
61. Rossi, J. M., Dunn, N. R., Hogan, B. L. M. & Zaret, K. S. Distinct mesodermal signals, including BMPs from the septum, transversum mesenchyme, are required in combination for hepatogenesis from the endoderm. *Genes Dev.* **15**, 1998–2009 (2001).
62. Kumar, M., Jordan, N., Melton, D. & Grapin-Botton, A. Signals from lateral plate mesoderm instruct endoderm toward a pancreatic fate. *Dev. Biol.* **259**, 109–22 (2003).
63. Spagnoli, F. M. & Brivanlou, A. H. The Gata5 target, TGIF2, defines the pancreatic region by modulating BMP signals within the endoderm. *Development* **135**, 451–61 (2008).
64. Jung, J., Zheng, M., Goldfarb, M. & Zaret, K. S. Initiation of mammalian liver development from endoderm by fibroblast growth factors. *Science (80-.).* **284**, 1998–2003 (1999).
65. Wandzioch, E. & Zaret, K. S. Dynamic signaling network for the specification of embryonic pancreas and liver progenitors. *Science (80-.).* **324**, 1707–1710 (2009).
66. Bhushan, A. *et al.* Fgf10 is essential for maintaining the proliferative capacity of epithelial progenitor cells during early pancreatic organogenesis. *Development* **128**,

5109–5117 (2001).
67. Dong, P. D. S. *et al.* Fgf10 regulates hepatopancreatic ductal system patterning and differentiation. *Nat. Genet.* **39**, 397–402 (2007).
68. Sumazaki, R. *et al.* Conversion of biliary system to pancreatic tissue in Hes1-deficient mice. *Nat. Genet.* **36**, 83–87 (2004).
69. Fukuda, A. *et al.* Ectopic pancreas formation in Hes1-knockout mice reveals plasticity of endodermal progenitors of the gut, bile duct, and pancreas. *J. Clin. Invest.* **116**, 1484–1493 (2006).
70. Rodríguez-Seguel, E. *et al.* Mutually exclusive signaling signatures define the hepatic and pancreatic progenitor cell lineage divergence. *Genes Dev.* **27**, 1932–46 (2013).
71. Cerdá-Esteban, N. *et al.* Stepwise reprogramming of liver cells to a pancreas progenitor state by the transcriptional regulator Tgif2. *Nat. Commun.* **8**, 14127 (2017).
72. Jennings, R. E. *et al.* Laser Capture and Deep Sequencing Reveals the Transcriptomic Programmes Regulating the Onset of Pancreas and Liver Differentiation in Human Embryos. *Stem cell reports* **12**, 224–237 (2017).
73. Escot, S., Willnow, D., Naumann, H., Di Francescantonio, S. & Spagnoli, F. M. Robo signalling controls pancreatic progenitor identity by regulating Tead transcription factors. *Nat. Commun.* (2018). doi:10.1038/s41467-018-07474-6
74. Zaret, K. S. Regulatory phases of early liver development: Paradigms of organogenesis. *Nature Reviews Genetics* **3**, 499–512 (2002).
75. Bort, R., Signore, M., Tremblay, K., Barbera, J. P. M. & Zaret, K. S. Hex homeobox gene controls the transition of the endoderm to a pseudostratified, cell emergent epithelium for liver bud development. *Dev. Biol.* **290**, 44–56 (2006).
76. Shiojiri, N. & Sugiyama, Y. Immunolocalization of extracellular matrix components and integrins during mouse liver development. *Hepatology* **40**, 346–355 (2004).
77. Barbera, J. P. M. *et al.* The homeobox gene Hex is required in definitive endodermal tissues for normal forebrain, liver and thyroid formation. *Development* **127**, 2433–2445 (2000).
78. Keng, V. W. *et al.* Homeobox gene Hex is essential for onset of mouse embryonic liver development and differentiation of the monocyte lineage. *Biochem. Biophys. Res. Commun.* **276**, 1155–1161 (2000).
79. Beaudry, J.-B. *et al.* Threshold levels of hepatocyte nuclear factor 6 (HNF-6) acting in synergy with HNF-4 and PGC-1alpha are required for time-specific gene expression during liver development. *Mol. Cell. Biol.* **26**, 6037–46 (2006).
80. Margagliotti, S. *et al.* Role of metalloproteinases at the onset of liver development. *Dev. Growth Differ.* **50**, 331–338 (2008).
81. Li, J., Ning, G. & Duncan, S. A. Mammalian hepatocyte differentiation requires the

transcription factor HNF-4α. *Genes Dev.* **14**, 464–474 (2000).

82. Matsumoto, K., Yoshitomi, H., Rossant, J. & Zaret, K. S. Liver organogenesis promoted by endothelial cells prior to vascular function. *Science (80-.).* **294**, 559–563 (2001).
83. Gittes, G. K. Developmental biology of the pancreas: A comprehensive review. *Developmental Biology* **326**, 4–35 (2009).
84. Uchida, T., Takada, T., Ammori, B. J., Suda, K. & Takahashi, T. Three-dimensional reconstruction of the ventral and dorsal pancreas: A new insight into anatomy and embryonic development. *J. Hepatobiliary. Pancreat. Surg.* **6**, 176–180 (1999).
85. Rieck, S., Bankaitis, E. D. & Wright, C. V. E. Lineage determinants in early endocrine development. *Semin. Cell Dev. Biol.* **23**, 673–684 (2012).
86. Cano, D. A., Soria, B., Martín, F. & Rojas, A. Transcriptional control of mammalian pancreas organogenesis. *Cellular and Molecular Life Sciences* **71**, 2383–2402 (2014).
87. Zhou, Q. *et al.* A Multipotent Progenitor Domain Guides Pancreatic Organogenesis. *Dev. Cell* **13**, 103–114 (2007).
88. Stanger, B. Z., Tanaka, A. J. & Melton, D. A. Organ size is limited by the number of embryonic progenitor cells in the pancreas but not the liver. *Nature* **445**, 886–891 (2007).
89. Villasenor, A., Chong, D. C., Henkemeyer, M. & Cleaver, O. Epithelial dynamics of pancreatic branching morphogenesis. *Development* (2010). doi:10.1242/dev.052993
90. Kesavan, G. *et al.* Cdc42-Mediated Tubulogenesis Controls Cell Specification. *Cell* **139**, 791–801 (2009).
91. Marty-Santos, L. & Cleaver, O. Progenitor Epithelium: Sorting Out Pancreatic Lineages. *J. Histochem. Cytochem.* **63**, 559–574 (2015).
92. Gu, G., Dubauskaite, J. & Melton, D. A. Direct evidence for the pancreatic lineage: NGN3+ cells are islet progenitors and are distinct from duct progenitors. *Development* **129**, 2447–2457 (2002).
93. Kopp, J. L. *et al.* Sox9+ ductal cells are multipotent progenitors throughout development but do not produce new endocrine cells in the normal or injured adult pancreas. *Development* **138**, 653–665 (2011).
94. Larsen, H. L. & Grapin-Botton, A. The molecular and morphogenetic basis of pancreas organogenesis. *Seminars in Cell and Developmental Biology* (2017). doi:10.1016/j.semcdb.2017.01.005
95. Petzold, K. M., Naumann, H. & Spagnoli, F. M. Rho signalling restriction by the RhoGAP Stard13 integrates growth and morphogenesis in the pancreas. *Dev.* **140**, 126–135 (2013).
96. Pan, F. C. *et al.* Spatiotemporal patterns of multipotentiality in Ptf1aexpressing cells

during pancreas organogenesis and injuryinduced facultative restoration. *Dev.* **140**, 751–764 (2013).

97. Bastidas-Ponce, A., Scheibner, K., Lickert, H. & Bakhti, M. Cellular and molecular mechanisms coordinating pancreas development. *Development* **144**, 2873–2888 (2017).
98. Kopinke, D. *et al.* Lineage tracing reveals the dynamic contribution of Hes1+ cells to the developing and adult pancreas. *Development* **138**, 431–441 (2011).
99. Solar, M. *et al.* Pancreatic Exocrine Duct Cells Give Rise to Insulin-Producing β Cells during Embryogenesis but Not after Birth. *Dev. Cell* **17**, 849–860 (2009).
100. Schaffer, A. E., Freude, K. K., Nelson, S. B. & Sander, M. Nkx6 transcription factors and Ptf1a function as antagonistic lineage determinants in multipotent pancreatic progenitors. *Dev. Cell* (2010). doi:10.1016/j.devcel.2010.05.015
101. Masui, T., Long, Q., Beres, T. M., Magnuson, M. A. & MacDonald, R. J. Early pancreatic development requires the vertebrate Suppressor of Hairless (RBPJ) in the PTF1 bHLH complex. *Genes Dev.* **21**, 2629–2643 (2007).
102. Hale, M. A. *et al.* The nuclear hormone receptor family member NR5A2 controls aspects of multipotent progenitor cell formation and acinar differentiation during pancreatic organogenesis. *Dev.* **141**, 3123–3133 (2014).
103. Beres, T. M. *et al.* PTF1 is an organ-specific and Notch-independent basic helix-loop-helix complex containing the mammalian Suppressor of Hairless (RBP-J) or its paralogue, RBP-L. *Mol. Cell. Biol.* **26**, 117–30 (2006).
104. Masui, T. *et al.* Replacement of Rbpj With Rbpjl in the PTF1 Complex Controls the Final Maturation of Pancreatic Acinar Cells. *Gastroenterology* **139**, 270–280 (2010).
105. Pin, C. L., Rukstalis, J. M., Johnson, C. & Konieczny, S. F. The bHLH transcription factor Mist1 is required to maintain exocrine pancreas cell organization and acinar cell identity. *J. Cell Biol.* **155**, 519–530 (2001).
106. Jia, D., Sun, Y. & Konieczny, S. F. Mist1 Regulates Pancreatic Acinar Cell Proliferation Through p21CIP1/WAF1. *Gastroenterology* **135**, 1687–1697 (2008).
107. Cleveland, M. H., Sawyer, J. M., Afelik, S., Jensen, J. & Leach, S. D. Exocrine ontogenies: On the development of pancreatic acinar, ductal and centroacinar cells. *Semin. Cell Dev. Biol.* **23**, 711–719 (2012).
108. Thompson, N., Scheinert, P., Grapin-Botton, A., Bucher, P. & Gesina, E. RNA Profiling and Chromatin Immunoprecipitation-Sequencing Reveal that PTF1a Stabilizes Pancreas Progenitor Identity via the Control of MNX1/HLXB9 and a Network of Other Transcription Factors. *Mol. Cell. Biol.* (2012). doi:10.1128/mcb.06318-11
109. Benitez, C. M., Goodyer, W. R. & Kim, S. K. Deconstructing pancreas developmental biology. *Cold Spring Harb. Perspect. Biol.* **4**, 1–17 (2012).

110. MacDonald, R. J., Swift, G. H. & Real, F. X. Transcriptional control of acinar development and homeostasis. in *Progress in Molecular Biology and Translational Science* **97**, 1–40 (Elsevier B.V., 2010).
111. Miralles, F., Czernichow, P. & Scharfmann, R. Follistatin regulates the relative proportions of endocrine versus exocrine tissue during pancreatic development. *Development* **125**, 1017–1024 (1998).
112. Ye, F., Duvillié, B. & Scharfmann, R. Fibroblast growth factors 7 and 10 are expressed in the human embryonic pancreatic mesenchyme and promote the proliferation of embryonic pancreatic epithelial cells. *Diabetologia* **48**, 277–281 (2005).
113. Elghazi, L., Cras-Méneur, C., Czernichow, P. & Scharfmann, R. Role for FGFR2IIIb-mediated signals in controlling pancreatic endocrine progenitor cell proliferation. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 3884–3889 (2002).
114. Miralles, F., Lamotte, L., Couton, D. & Joshi, R. L. Interplay between FGF10 and Notch signalling is required for the self-renewal of pancreatic progenitors. *Int. J. Dev. Biol.* **50**, 17–26 (2006).
115. Norgaard, G. A., Jensen, J. N. & Jensen, J. FGF10 signaling maintains the pancreatic progenitor cell state revealing a novel role of Notch in organ development. *Dev. Biol.* **264**, 323–338 (2003).
116. Hart, A., Papadopoulou, S. & Edlund, H. Fgf10 maintains notch activation, stimulates proliferation, and blocks differentiation of pancreatic epithelial cells. *Dev. Dyn.* **228**, 185–193 (2003).
117. Ghosh, B. & Leach, S. D. Interactions between Hairy/Enhancer of Split-related proteins and the pancreatic transcription factor Ptf1-p48 modulate function of the PTF1 transcriptional complex. *Biochem. J.* **393**, 679–685 (2006).
118. Herrera, P. L. Adult insulin- and glucagon-producing cells differentiate from two independent cell lineages. *Development* **127**, 2317–2322 (2000).
119. Johansson, K. A. *et al.* Temporal Control of Neurogenin3 Activity in Pancreas Progenitors Reveals Competence Windows for the Generation of Different Endocrine Cell Types. *Dev. Cell* **12**, 457–465 (2007).
120. Gradwohl, G., Dierich, A., LeMeur, M. & Guillemot, F. neurogenin3 is required for the development of the four endocrine cell lineages of the pancreas. *Proc. Natl. Acad. Sci.* **97**, 1607–1611 (2000).
121. Miyatsuka, T., Kosaka, Y., Kim, H. & German, M. S. Neurogenin3 inhibits proliferation in endocrine progenitors by inducing Cdkn1a. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 185–190 (2011).
122. Gouzi, M., Kim, Y. H., Katsumoto, K., Johansson, K. & Grapin-Botton, A. Neurogenin3 initiates stepwise delamination of differentiating endocrine cells during pancreas

development. *Dev. Dyn.* (2011). doi:10.1002/dvdy.22544

123. Wang, S. *et al.* Neurog3 gene dosage regulates allocation of endocrine and exocrine cell fates in the developing mouse pancreas. *Dev. Biol.* **339**, 26–37 (2010).
124. Beucher, A. *et al.* Competence of failed endocrine progenitors to give rise to acinar but not ductal cells is restricted to early pancreas development. *Dev. Biol.* **361**, 277–285 (2012).
125. Mamidi, A. *et al.* Mechanosignalling via integrins directs fate decisions of pancreatic progenitors. *Nature* **564**, 114–118 (2018).
126. Desgraz, R. & Herrera, P. L. Pancreatic neurogenin 3-expressing cells are unipotent islet precursors. *Development* **136**, 3567–3574 (2009).
127. Pierreux, C. E. *et al.* The Transcription Factor Hepatocyte Nuclear Factor-6 Controls the Development of Pancreatic Ducts in the Mouse. *Gastroenterology* **130**, 532–541 (2006).
128. Golosow, N. & Grobstein, C. Epitheliomesenchymal interaction in pancreatic morphogenesis. *Dev. Biol.* **4**, 242–255 (1962).
129. Cozzitorto, C. & Spagnoli, F. M. Pancreas organogenesis: The interplay between surrounding microenvironment(s) and epithelium-intrinsic factors. in *Current Topics in Developmental Biology* **132**, 221–256 (Academic Press Inc., 2019).
130. Sneddon, J. B., Borowiak, M. & Melton, D. A. Self-renewal of embryonic-stem-cell-derived progenitors by organ-matched mesenchyme. *Nature* **491**, 765–8 (2012).
131. Landsman, L. *et al.* Pancreatic mesenchyme regulates epithelial organogenesis throughout development. *PLoS Biol.* **9**, (2011).
132. Kim, S. K. & Hebrok, M. Intercellular signals regulating pancreas development and function. *Genes and Development* **15**, 111–127 (2001).
133. Celli, G., Larochelle, W. J., Mackem, S., Sharp, R. & Merlino, G. Soluble dominant-negative receptor uncovers essential roles for fibroblast growth factors in multi-organ induction and patterning. *EMBO J.* **17**, 1642–1655 (1998).
134. Seymour, P. A. *et al.* A Sox9/Fgf feed-forward loop maintains pancreatic organ identity. *Dev.* **139**, 3363–3372 (2012).
135. Apelqvist, Å. *et al.* Notch signalling controls pancreatic cell differentiation. *Nature* **400**, 877–881 (1999).
136. Jensen, J. *et al.* Control of endodermal endocrine development by Hes-1. *Nat. Genet.* **24**, 36–44 (2000).
137. Bonfanti, P. *et al.* Ex Vivo Expansion and Differentiation of Human and Mouse Fetal Pancreatic Progenitors Are Modulated by Epidermal Growth Factor. doi:10.1089/scd.2014.0550
138. Crisera, C. A. *et al.* Transforming growth factor-β1 in the developing mouse pancreas:

a potential regulator of exocrine differentiation. *Differentiation* **65**, 255–259 (2000).

139. Crisera, C. A. *et al.* The ontogeny of TGF-β1, -β2, -β3, and TGF-β receptor-II expression in the pancreas: Implications for regulation of growth and differentiation. in *Journal of Pediatric Surgery* **34**, 689–694 (W.B. Saunders, 1999).
140. Tulachan, S. S. *et al.* TGF-β isoform signaling regulates secondary transition and mesenchymal-induced endocrine development in the embryonic mouse pancreas. *Dev. Biol.* **305**, 508–521 (2007).
141. Miralles, F., Battelino, T., Czernichow, P. & Scharfmann, R. TGF-β plays a key role in morphogenesis of the pancreatic islets of langerhans by controlling the activity of the matrix metalloproteinase MMP-2. *J. Cell Biol.* **143**, 827–836 (1998).
142. Sanvito, F. *et al.* TGF-beta 1 influences the relative development of the exocrine and endocrine pancreas in vitro. *Development* **120**, 3451–62 (1994).
143. Böttinger, E. P. *et al.* Expression of a dominant-negative mutant TGF-β type II receptor in transgenic mice reveals essential roles for TGF-β in regulation of growth and differentiation in the exocrine pancreas. *EMBO J.* **16**, 2621–2633 (1997).
144. Kim, S. K. *et al.* Activin receptor patterning of foregut organogenesis. *Genes Dev.* **14**, 1866–1871 (2000).
145. Demeterco, C., Beattie, G. M., Dib, S. A., Lopez, A. D. & Hayek, A. A Role for Activin A and Betacellulin in Human Fetal Pancreatic Cell Differentiation and Growth [1]. *J. Clin. Endocrinol. Metab.* **85**, 3892–3897 (2000).
146. Ritvos, O. *et al.* Activin disrupts epithelial branching morphogenesis in developing glandular organs of the mouse. *Mech. Dev.* **50**, 229–45 (1995).
147. Ahnfelt-Rønne, J., Ravassard, P., Pardanaud-Glavieux, C., Scharfmann, R. & Serup, P. Mesenchymal bone morphogenetic protein signaling is required for normal pancreas development. *Diabetes* **59**, 1948–1956 (2010).
148. Goulley, J., Dahl, U., Baeza, N., Mishina, Y. & Edlund, H. BMP4-BMPR1A Signaling in β Cells Is Required for and Augments Glucose-Stimulated Insulin Secretion. *Cell Metab.* **5**, 207–219 (2007).
149. Dichmann, D. S., Miller, C. P., Jensen, J., Heller, R. S. & Serup, P. Expression and misexpression of members of the FGF and TGFβ families of growth factors in the developing mouse pancreas. *Dev. Dyn.* **226**, 663–674 (2003).
150. Kawaguchi, Y. *et al.* The role of the transcriptional regulator Ptf1a in converting intestinal to pancreatic progenitors. *Nat. Genet.* **32**, 128–134 (2002).
151. Que, J. *et al.* Multiple dose-dependent roles for Sox2 in the patterning and differentiation of anterior foregut endoderm. *Development* **134**, 2521–2531 (2007).
152. Gao, N., White, P. & Kaestner, K. H. Establishment of Intestinal Identity and Epithelial-Mesenchymal Signaling by Cdx2. *Dev. Cell* **16**, 588–599 (2009).

153. Spence, J. R. *et al.* Sox17 Regulates Organ Lineage Segregation of Ventral Foregut Progenitor Cells. *Dev. Cell* **17**, 62–74 (2009).
154. Shih, H. P. *et al.* A Gene Regulatory Network Cooperatively Controlled by Pdx1 and Sox9 Governs Lineage Allocation of Foregut Progenitor Cells. *Cell Rep.* **13**, 326–336 (2015).
155. Willet, S. G. *et al.* Dominant and context-specific control of endodermal organ allocation by Ptf1a. *Dev.* **141**, 4385–4394 (2014).
156. Ohlsson, H., Karlsson, K. & Edlund, T. IPF1, a homeodomain-containing transactivator of the insulin gene. *EMBO J.* **12**, 4251–4259 (1993).
157. Ahlgren, U., Jonsson, J. & Edlund, H. The morphogenesis of the pancreatic mesenchyme is uncoupled from that of the pancreatic epithelium in IPF1/PDX1-deficient mice. *Development* **122**, 1409–1416 (1996).
158. Gannon, M. *et al.* pdx-1 function is specifically required in embryonic β cells to generate appropriate numbers of endocrine cell types and maintain glucose homeostasis. *Dev. Biol.* **314**, 406–417 (2008).
159. Gao, T. *et al.* Pdx1 Maintains β Cell Identity and Function by Repressing an α Cell Program. *Cell Metab.* **19**, 259–271 (2014).
160. Staffers, D. A., Ferrer, J., Clarke, W. L. & Habener, J. F. Early-onset type-ll diabetes mellitus (Mody4) linked to ipf1. *Nat. Genet.* **17**, 139 (1997).
161. Stoffers, D. A., Zinkin, N. T., Stanojevic, V., Clarke, W. L. & Habener, J. F. Pancreatic agenesis attributable to a single nucleotide deletion in the human IPF1 gene coding sequence. *Nat. Genet.* **15**, 106–110 (1997).
162. Yang, Y. P., Magnuson, M. A., Stein, R. & Wright, C. V. E. The mammal-specific Pdx1 area II enhancer has multiple essential functions in early endocrine cell specification and postnatal β-cell maturation. *Dev.* **144**, 248–257 (2017).
163. Fujitani, Y. *et al.* Targeted deletion of a cis-regulatory region reveals differential gene dosage requirements for Pdx1 in foregut organ differentiation and pancreas formation. *Genes Dev.* **20**, 253–266 (2006).
164. Cox, A. R. & Kushner, J. A. Area IV Knockout Reveals How Pdx1 Is Regulated in Postnatal β-Cell Development. *Diabetes* **66**, 2738–2740 (2017).
165. Gannon, M., Gamer, L. W. & Wright, C. V. E. Regulatory regions driving developmental and tissue-specific expression of the essential pancreatic gene pdx1. *Dev. Biol.* **238**, 185–201 (2001).
166. Vanhoose, A. M. *et al.* MafA and MafB regulate Pdx1 transcription through the area II control region in pancreatic β cells. *J. Biol. Chem.* **283**, 22612–22619 (2008).
167. Gao, N. *et al.* Dynamic regulation of Pdx1 enhancers by Foxa1 and Foxa2 is essential for pancreas development. *Genes Dev.* **22**, 3435–48 (2008).

168. Wiebe, P. O. *et al.* Ptf1a binds to and activates area III, a highly conserved region of the Pdx1 promoter that mediates early pancreas-wide Pdx1 expression. *Mol. Cell. Biol.* **27**, 4093–104 (2007).
169. Seymour, P. A. *et al.* SOX9 is required for maintenance of the pancreatic progenitor cell pool. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 1865–1870 (2007).
170. Seymour, P. A. *et al.* A dosage-dependent requirement for Sox9 in pancreatic endocrine cell formation. *Dev. Biol.* **323**, 19–30 (2008).
171. Shih, H. P. *et al.* A Notch-dependent molecular circuitry initiates pancreatic endocrine and ductal cell differentiation. *Dev.* **139**, 2488–2499 (2012).
172. Krapp, A. *et al.* The bHLH protein PTF1-p48 is essential for the formation of the exocrine and the correct spatial organization of the endocrine pancreas. *Genes Dev.* **12**, 3752–3763 (1998).
173. Villasenor, A., Chong, D. C. & Cleaver, O. Biphasic Ngn3 expression in the developing pancreas. *Dev. Dyn.* **237**, 3270–3279 (2008).
174. Sommer, L., Ma, Q. & Anderson, D. J. neurogenins, a novel family of atonal-related bHLH transcription factors, are putative mammalian neuronal determination genes that reveal progenitor cell heterogeneity in the developing CNS and PNS. *Mol. Cell. Neurosci.* **8**, 221–241 (1996).
175. Jenny, M. *et al.* Neurogenin3 is differentially required for endocrine cell fate specification in the intestinal and gastric epithelium. *EMBO J.* **21**, 6338–6347 (2002).
176. Gehart, H., Van Es, J. H., Dekkers, J. F., Rios, A. & Clevers Correspondence, H. Identification of Enteroendocrine Regulators by Real-Time Single-Cell Differentiation Mapping Data Resources GSE113561 Gehart et al. *Cell* **176**, 1158-1173.e16 (2019).
177. Qu, X. *et al.* Notch-mediated post-translational control of Ngn3 protein stability regulates pancreatic patterning and cell fate commitment. *Dev. Biol.* **376**, 1–12 (2013).
178. Azzarelli, R. *et al.* Neurogenin3 phosphorylation controls reprogramming efficiency of pancreatic ductal organoids into endocrine cells. *Sci. Rep.* **8**, 15374 (2018).
179. Krentz, N. A. J. *et al.* Phosphorylation of NEUROG3 Links Endocrine Differentiation to the Cell Cycle in Pancreatic Progenitors. *Dev. Cell* **41**, 129-142.e6 (2017).
180. Rukstalis, J. M. & Habener, J. F. Snail2, a mediator of epithelial-mesenchymal transitions, expressed in progenitor cells of the developing endocrine pancreas. *Gene Expr. Patterns* **7**, 471–479 (2007).
181. Bechard, M. E. *et al.* Precommitment low-level neurog3 expression defines a long-lived mitotic endocrine-biased progenitor pool that drives production of endocrine-committed cells. *Genes Dev.* (2016). doi:10.1101/gad.284729.116
182. Schwitzgebel, V. M. *et al.* Expression of neurogenin3 reveals an islet cell precursor population in the pancreas. *Development* **127**, 3533–3542 (2000).

183. Ahlgren, U., Pfaff, S. L., Jessell, T. M., Edlund, T. & Edlund, H. Independent requirement for ISL1 in formation of pancreatic mesenchyme and islet cells. *Nature* **385**, 257–260 (1997).
184. Du, A. *et al.* Islet-1 is required for the maturation, proliferation, and survival of the endocrine pancreas. *Diabetes* **58**, 2059–2069 (2009).
185. Mellitzer, G. *et al.* IA1 is NGN3-dependent and essential for differentiation of the endocrine pancreas. *EMBO J.* **25**, 1344–1352 (2006).
186. Gierl, M. S., Karoulias, N., Wende, H., Strehle, M. & Birchmeier, C. The Zinc-finger factor Insm1 (IA-1) is essential for the development of pancreatic β cells and intestinal endocrine cells. *Genes Dev.* **20**, 2465–2478 (2006).
187. Naya, F. J. *et al.* Diabetes, defective pancreatic morphogenesis, and abnormal enteroendocrine differentiation in BETA2/NeuroD-deficient mice. *Genes Dev.* **11**, 2323–2334 (1997).
188. Mastracci, T. L., Anderson, K. R., Papizan, J. B. & Sussel, L. Regulation of Neurod1 Contributes to the Lineage Potential of Neurogenin3+ Endocrine Precursor Cells in the Pancreas. *PLoS Genet.* **9**, (2013).
189. Soyer, J. *et al.* Rfx6 is an Ngn3-dependent winged helix transcription factor required for pancreatic islet cell development. *Development* **137**, 203–212 (2010).
190. Wang, S. *et al.* Myt1 and Ngn3 form a feed-forward expression loop to promote endocrine islet cell differentiation. *Dev. Biol.* **317**, 531–540 (2008).
191. Sussel, L. *et al.* Mice lacking the homeodomain transcription factor Nkx2.2 have diabetes due to arrested differentiation of pancreatic ?? cells. *Development* **125**, 2213–2221 (1998).
192. Sander, N. *et al.* Homeobox gene Nkx6.1 lies downstream of Nkx2.2 in the major pathway of β-cell formation in the pancreas. *Development* **127**, 5533–5540 (2000).
193. Taylor, B. L., Liu, F. F. & Sander, M. Nkx6.1 Is Essential for Maintaining the Functional State of Pancreatic Beta Cells. *Cell Rep.* (2013). doi:10.1016/j.celrep.2013.08.010
194. Henseleit, K. D. *et al.* NKX6 transcription factor activity is required for α- and β-cell development in the pancreas. *Development* **132**, 3139–3149 (2005).
195. Schaffer, A. E. *et al.* Nkx6.1 Controls a Gene Regulatory Network Required for Establishing and Maintaining Pancreatic Beta Cell Identity. *PLoS Genet.* **9**, (2013).
196. Papizan, J. B. *et al.* Nkx2.2 repressor complex regulates islet β-cell specification and prevents β-to-α-cell reprogramming. *Genes Dev.* **25**, 2291–2305 (2011).
197. Sander, M. *et al.* Genetic analysis reveals that PAX6 is required for normal transcription of pancreatic hormone genes and islet development. *Genes Dev.* **11**, 1662–1673 (1997).
198. Sosa-Pineda, B., Chowdhury, K., Torres, M., Oliver, G. & Gruss, P. The Pax4 gene is

essential for differentiation of insulin-producing β cells in the mammalian pancreas. *Nature* **386**, 399–402 (1997).

199. Wang, J. *et al.* The concerted activities of Pax4 and Nkx2.2 are essential to initiate pancreatic β-cell differentiation. *Dev. Biol.* **266**, 178–189 (2004).
200. St-Onge, L., Sosa-Pineda, B., Chowdhury, K., Mansouri, A. & Gruss, P. Pax6 is required for differentiation of glucagon-producing α-cells in mouse pancreas. *Nature* **387**, 406–409 (1997).
201. Hart, A. W., Mella, S., Mendrychowski, J., van Heyningen, V. & Kleinjan, D. A. The Developmental Regulator Pax6 Is Essential for Maintenance of Islet Cell Function in the Adult Mouse Pancreas. *PLoS One* **8**, e54173 (2013).
202. Swisa, A. *et al.* PAX6 maintains β cell identity by repressing genes of alternative islet cell types. *J. Clin. Invest.* **127**, 230–243 (2017).
203. Collombat, P. *et al.* The Ectopic Expression of Pax4 in the Mouse Pancreas Converts Progenitor Cells into α and Subsequently β Cells. *Cell* **138**, 449–462 (2009).
204. Collombat, P. *et al.* Opposing actions of Arx and Pax4 in endocrine pancreas development. *Genes Dev.* **17**, 2591–2603 (2003).
205. Collombat, P. *et al.* The simultaneous loss of Arx and Pax4 genes promotes a somatostatin-producing cell fate specification at the expense of the α- and β-cell lineages in the mouse endocrine pancreas. *Development* **132**, 2969–2980 (2005).
206. Al-Hasani, K. *et al.* Adult duct-lining cells can reprogram into β-like cells able to counter repeated cycles of toxin-induced diabetes. *Dev. Cell* **26**, 86–100 (2013).
207. Courtney, M. *et al.* The Inactivation of Arx in Pancreatic α-Cells Triggers Their Neogenesis and Conversion into Functional β-Like Cells. *PLoS Genet.* **9**, e1003934 (2013).
208. Nishimura, W. *et al.* A switch from MafB to MafA expression accompanies differentiation to pancreatic β-cells. *Dev. Biol.* **293**, 526–539 (2006).
209. Zhang, C. *et al.* MafA Is a Key Regulator of Glucose-Stimulated Insulin Secretion. *Mol. Cell. Biol.* **25**, 4969–4976 (2005).
210. Artner, I. *et al.* MafB is required for islet β cell maturation. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 3853–3858 (2007).
211. Artner, I. *et al.* MafA and MafB regulate genes critical to β-cells in a unique temporal manner. *Diabetes* **59**, 2530–2539 (2010).
212. Artner, I. *et al.* An activator of the glucagon gene expressed in developing islet α- and β-cells. *Diabetes* **55**, 297–304 (2006).
213. Matsuoka, T. A. *et al.* The MafA transcription factor appears to be responsible for tissue-specific expression of insulin. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 2930–2933 (2004).

214. Matsuoka, T. A. *et al.* Mafa enables Pdx1 to effectively convert pancreatic islet progenitors and committed islet α-cells into β-cells in vivo. *Diabetes* **66**, 1293–1300 (2017).
215. Xiao, X. *et al.* Endogenous Reprogramming of Alpha Cells into Beta Cells, Induced by Viral Gene Therapy, Reverses Autoimmune Diabetes. *Cell Stem Cell* **22**, 78-90.e4 (2018).
216. Zhou, Q., Brown, J., Kanarek, A., Rajagopal, J. & Melton, D. A. In vivo reprogramming of adult pancreatic exocrine cells to β-cells. *Nature* **455**, 627–632 (2008).
217. Kofent, J. & Spagnoli, F. M. Xenopus as a model system for studying pancreatic development and diabetes. *Seminars in Cell and Developmental Biology* (2016). doi:10.1016/j.semcdb.2016.01.005
218. Jennings, R. E., Berry, A. A., Strutt, J. P., Gerrard, D. T. & Hanley, N. A. Human pancreas development. *Development* **142**, 3126–37 (2015).
219. Miguel-Escalada, I. *et al.* Human pancreatic islet three-dimensional chromatin architecture provides insights into the genetics of type 2 diabetes. *Nat. Genet.* (2019). doi:10.1038/s41588-019-0457-0
220. Ramond, C. *et al.* Understanding human fetal pancreas development using subpopulation sorting, RNA sequencing and single-cell profiling. *Development* **145**, dev165480 (2018).
221. Piper, K. *et al.* Beta cell differentiation during early human pancreas development. *J. Endocrinol.* **181**, 11–23 (2004).
222. Jennings, R. E. *et al.* Development of the human pancreas from foregut to endocrine commitment. *Diabetes* **62**, 3514–3522 (2013).
223. Sellick, G. S. *et al.* Mutations in PTF1A cause pancreatic and cerebellar agenesis. *Nat. Genet.* **36**, 1301–1305 (2004).
224. Weedon, M. N. *et al.* Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis. *Nat. Genet.* **46**, 61–64 (2014).
225. Bonal, C. & Herrera, P. L. Genes controlling pancreas ontogeny. *International Journal of Developmental Biology* **52**, 823–835 (2008).
226. Salisbury, R. J. *et al.* The window period of NEUROGENIN3 during human gestation. *Islets* (2014). doi:10.4161/19382014.2014.954436
227. Jeon, J., Correa-Medina, M., Ricordi, C., Edlund, H. & Diez, J. A. Endocrine cell clustering during human pancreas development. *J. Histochem. Cytochem.* **57**, 811–824 (2009).
228. McGrath, P. S., Watson, C. L., Ingram, C., Helmrath, M. A. & Wells, J. M. The basic helix-loop-helix transcription factor neurog3 is required for development of the human endocrine pancreas. *Diabetes* (2015). doi:10.2337/db14-1412

229. Cyphert, H. A. *et al.* Examining how the MAFB transcription factor affects islet β-cell function postnatally. *Diabetes* **68**, 337–348 (2019).
230. Baron, M. *et al.* A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. *Cell Syst.* **3**, 346-360.e4 (2016).
231. Muraro, M. J. *et al.* A Single-Cell Transcriptome Atlas of the Human Pancreas. *Cell Syst.* (2016). doi:10.1016/j.cels.2016.09.002
232. Dorrell, C. *et al.* Human islets contain four distinct subtypes of β cells. *Nat. Commun.* **7**, 11756 (2016).
233. Mahajan, A. *et al.* Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat. Genet.* **46**, 234–244 (2014).
234. Malecki, M. T. *et al.* Mutations in NEUROD1 are associated with the development of type 2 diabetes mellitus. *Nat. Genet.* **23**, 323–328 (1999).
235. Hattersley, A. T. Maturity-onset diabetes of the young. in *Joslin's Diabetes Mellitus: Fourteenth Edition* 463–476 (Wolters Kluwer Health Adis (ESP), 2011). doi:10.1007/BF00257348
236. Hattersley, A. T. & Tooke, J. E. The fetal insulin hypothesis: An alternative explanation of the association of low birthweight with diabetes and vascular disease. *Lancet* **353**, 1789–1792 (1999).
237. Hales, C. & Barker, D. Type 2 (non-insulin-dependent) diabetes mellitus: the thrifty phenotype hypothesis. *Int. J. Epidemiol.* **42**, 1215–1222 (2013).
238. Pagliuca, F. W. *et al.* Generation of Functional Human Pancreatic β Cells In Vitro. *Cell* **159**, 428–439 (2014).
239. Rezania, A. *et al.* Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells. *Nat. Biotechnol.* **32**, 1121–1133 (2014).
240. Russ, H. a *et al.* Controlled induction of human pancreatic progenitors produces functional beta-like cells in vitro. *EMBO J.* (2015). doi:10.15252/embj.201591058
241. Zhu, Z. *et al.* Genome Editing of Lineage Determinants in Human Pluripotent Stem Cells Reveals Mechanisms of Pancreatic Development and Diabetes. *Cell Stem Cell* **18**, 755–768 (2016).
242. Lee, K. *et al.* FOXA2 Is Required for Enhancer Priming during Pancreatic Differentiation. *Cell Rep.* **28**, 382-393.e7 (2019).
243. Mann, R. S. & Affolter, M. Hox proteins meet more partners. *Curr. Opin. Genet. Dev.* **8**, 423–9 (1998).
244. Yang, Y. *et al.* Three-amino acid extension loop homeodomain proteins Meis2 and TGIF differentially regulate transcription. *J. Biol. Chem.* **275**, 20734–41 (2000).
245. Mukherjee, K. & Bürglin, T. R. Comprehensive analysis of animal TALE homeobox genes: New conserved motifs and cases of accelerated evolution. *J. Mol. Evol.* **65**,

137–153 (2007).
246. Cerdá-Esteban, N. & Spagnoli, F. M. Glimpse into Hox and tale regulation of cell differentiation and reprogramming. *Dev. Dyn.* **243**, 76–87 (2014).
247. Moens, C. B. & Selleri, L. Hox cofactors in vertebrate development. *Dev. Biol.* **291**, 193–206 (2006).
248. Morgan, R., El-Tanani, M., Hunter, K. D., Harrington, K. J. & Pandha, H. S. Targeting HOX/PBX dimers in cancer. *Oncotarget* **8**, 32322–32331 (2017).
249. McGinnis, W. & Krumlauf, R. Homeobox genes and axial patterning. *Cell* **68**, 283–302 (1992).
250. Pearson, J. C., Lemons, D. & McGinnis, W. Modulating Hox gene functions during animal body patterning. *Nature Reviews Genetics* **6**, 893–904 (2005).
251. Gehring, W. J. The animal body plan, the prototypic body segment, and eye evolution. *Evol. Dev.* **14**, 34–46 (2012).
252. Krumlauf, R. Hox genes in vertebrate development. *Cell* **78**, 191–201 (1994).
253. Lumsden, A. & Krumlauf, R. Patterning the vertebrate neuraxis. *Science* **274**, 1109–1115 (1996).
254. Kmita, M. & Duboule, D. Organizing axes in time and space; 25 years of colinear tinkering. *Science* **301**, 331–333 (2003).
255. Chang, C. P. *et al.* Pbx proteins display hexapeptide-dependent cooperative DNA binding with a subset of Hox proteins. *Genes Dev.* **9**, 663–674 (1995).
256. Knoepfler, P. S. & Kamps, M. P. The pentapeptide motif of Hox proteins is required for cooperative DNA binding with Pbx1, physically contacts Pbx1, and enhances DNA binding by Pbx1. *Mol. Cell. Biol.* **15**, 5811–5819 (1995).
257. Phelan, M. L., Rambaldi, I. & Featherstone, M. S. Cooperative interactions between HOX and PBX proteins mediated by a conserved peptide motif. *Mol. Cell. Biol.* **15**, 3989–3997 (1995).
258. Peltenburg, L. T. & Murre, C. Engrailed and Hox homeodomain proteins contain a related Pbx interaction motif that recognizes a common structure present in Pbx. *EMBO J.* **15**, 3385–93 (1996).
259. Merabet, S. & Mann, R. S. To Be Specific or Not: The Critical Relationship Between Hox And TALE Proteins. *Trends in Genetics* **32**, 334–347 (2016).
260. Ferretti, E., Schulz, H., Talarico, D., Blasi, F. & Berthelsen, J. The PBX-regulating protein PREP1 is present in different PBX-complexed forms in mouse. *Mech. Dev.* **83**, 53–64 (1999).
261. Knoepfler, P. S., Calvo, K. R., Chen, H., Antonarakis, S. E. & Kamps, M. P. Meis1 and pKnox1 bind DNA cooperatively with Pbx1 utilizing an interaction surface disrupted in oncoprotein E2a-Pbx1. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 14553–14558 (1997).

262. Berthelsen, J., Zappavigna, V., Ferretti, E., Mavilio, F. & Blasi, F. The novel homeoprotein Prep1 modulates Pbx-Hox protein cooperativity. *EMBO J.* **17**, 1434–1445 (1998).
263. Mann, R. S. & Affolter, M. Hox proteins meet more partners. *Curr. Opin. Genet. Dev.* **8**, 423–429 (1998).
264. Ryoo, H. D., Marty, T., Casares, F., Affolter, M. & Mann, R. S. Regulation of Hox target genes by a DNA bound homothorax/Hox/extradenticle complex. *Development* **126**, 5137–5148 (1999).
265. Mann, R. S., Lelli, K. M. & Joshi, R. Chapter 3 Hox Specificity. Unique Roles for Cofactors and Collaborators. *Current Topics in Developmental Biology* **88**, 63–101 (2009).
266. Jacobs, Y., Schnabel, C. A. & Cleary, M. L. Trimeric Association of Hox and TALE Homeodomain Proteins Mediates Hoxb2 Hindbrain Enhancer Activity . *Mol. Cell. Biol.* **19**, 5134–5142 (1999).
267. Longobardi, E. *et al.* Biochemistry of the tale transcription factors PREP, MEIS, and PBX in vertebrates. *Dev. Dyn.* **243**, 59–75 (2014).
268. Selleri, L., Zappavigna, V. & Ferretti, E. 'Building a perfect body': Control of vertebrate organogenesis by PBX-dependent regulatory networks. *Genes and Development* **33**, 258–275 (2019).
269. Bürglin, T. R. The PBC domain contains a MEINOX domain: Coevolution of Hox and TALE homeobox genes? *Dev. Genes Evol.* **208**, 113–116 (1998).
270. Fognani, C. Characterization of PREP2, a paralog of PREP1, which defines a novel sub-family of the MEINOX TALE homeodomain transcription factors. *Nucleic Acids Res.* **30**, 2043–2051 (2002).
271. Wagner, K., Mincheva, A., Korn, B., Lichter, P. & Pöpperl, H. Pbx4, a new Pbx family member on mouse chromosome 8, is expressed during spermatogenesis. *Mech. Dev.* **103**, 127–31 (2001).
272. Selleri, L. *et al.* Requirement for Pbx1 in skeletal patterning and programming chondrocyte proliferation and differentiation. *Development* **128**, 3543–3557 (2001).
273. Rhee, J. W. *et al.* Pbx3 deficiency results in central hypoventilation. *Am. J. Pathol.* **165**, 1343–1350 (2004).
274. Selleri, L. *et al.* The TALE homeodomain protein Pbx2 is not essential for development and long-term survival. *Mol. Cell. Biol.* **24**, 5324–31 (2004).
275. Brendolan, A. *et al.* A Pbx1-dependent genetic and transcriptional network regulates spleen ontogeny. *Development* **132**, 3113–3126 (2005).
276. Sanyal, M. *et al.* B-cell development fails in the absence of the Pbx1 proto-oncogene. *Blood* **109**, 4191–4199 (2007).

277. Capellini, T. D., Zappavigna, V. & Selleri, L. Pbx homeodomain proteins: TALEnted regulators of limb patterning and outgrowth. *Developmental Dynamics* **240**, 1063–1086 (2011).
278. Kim, S. K. *et al.* Pbx1 inactivation disrupts pancreas development and in Ipf1-deficient mice promotes diabetes mellitus. *Nat. Genet.* **30**, 430–435 (2002).
279. Losa, M. *et al.* Face morphogenesis is promoted by Pbx-dependent EMT via regulation of snail1 during frontonasal prominence fusion. *Dev.* **145**, (2018).
280. Grebbin, B. M. & Schulte, D. PBX1 as Pioneer Factor: A Case Still Open. *Front. cell Dev. Biol.* **5**, 9 (2017).
281. Choe, S. K., Ladam, F. & Sagerström, C. G. TALE Factors Poise Promoters for Activation by Hox Proteins. *Dev. Cell* **28**, 203–211 (2014).
282. Berkes, C. A. *et al.* Pbx marks genes for activation by MyoD indicating a role for a homeodomain protein in establishing myogenic potential. *Mol. Cell* **14**, 465–477 (2004).
283. de la Serna, I. L. *et al.* MyoD Targets Chromatin Remodeling Complexes to the Myogenin Locus Prior to Forming a Stable DNA-Bound Complex. *Mol. Cell. Biol.* **25**, 3997–4009 (2005).
284. Maves, L. *et al.* Pbx homeodomain proteins direct Myod activity to promote fast-muscle differentiation. *Development* **134**, 3371–3382 (2007).
285. Penkov, D. *et al.* Analysis of the DNA-Binding Profile and Function of TALE Homeoproteins Reveals Their Specialization and Specific Interactions with Hox Genes/Proteins. *Cell Rep.* **3**, 1321–1333 (2013).
286. Chang, C. P. *et al.* Meis proteins are major in vivo DNA binding partners for wild-type but not chimeric Pbx proteins. *Mol. Cell. Biol.* **17**, 5679–5687 (1997).
287. Shen, W.-F. *et al.* HOXA9 Forms Triple Complexes with PBX2 and MEIS1 in Myeloid Cells. *Mol. Cell. Biol.* **19**, 3051–3061 (1999).
288. Azcoitia, V., Aracil, M., Martínez-A, C. & Torres, M. The homeodomain protein Meis1 is essential for definitive hematopoiesis and vascular patterning in the mouse embryo. *Dev. Biol.* **280**, 307–320 (2005).
289. Hisa, T. *et al.* Hematopoietic, angiogenic and eye defects in Meis1 mutant animals. *EMBO J.* **23**, 450–459 (2004).
290. Zhang, X., Friedman, A., Heaney, S., Purcell, P. & Maas, R. L. Meis homeoproteins directly regulate Pax6 during vertebrate lens morphogenesis. *Genes Dev.* **16**, 2097–2107 (2002).
291. Wamstad, J. A. *et al.* Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. *Cell* **151**, 206–220 (2012).
292. Paige, S. L. *et al.* A temporal chromatin signature in human embryonic stem cells

identifies regulators of cardiac development. *Cell* **151**, 221–232 (2012).
293. Berger, M. F. *et al.* Variation in Homeodomain DNA Binding Revealed by High-Resolution Analysis of Sequence Preferences. *Cell* **133**, 1266–1276 (2008).
294. Wotton, D. & Taniguchi, K. Functions of TGIF homeodomain proteins and their roles in normal brain development and holoprosencephaly. *American Journal of Medical Genetics, Part C: Seminars in Medical Genetics* **178**, 128–139 (2018).
295. Melhuish, T. A., Gallo, C. M. & Wotton, D. TGIF2 Interacts with Histone Deacetylase 1 and Represses Transcription. *J. Biol. Chem.* **276**, 32109–32114 (2001).
296. Lo, R. S., Wotton, D. & Massagué, J. Epidermal growth factor signaling via Ras controls the Smad transcriptional co-repressor TGIF. *EMBO J.* **20**, 128–136 (2001).
297. Melhuish, T. A. & Wotton, D. The Tgif2 gene contains a retained intron within the coding sequence. *BMC Mol. Biol.* **7**, (2006).
298. Bartholin, L. *et al.* TGIF inhibits retinoid signaling. *Mol. Cell. Biol.* **26**, 990–1001 (2006).
299. Bertolino, E., Reimund, B., Wildt-Perinic, D. & Clerc, R. G. A novel homeobox protein which recognizes a TGT core and functionally interferes with a retinoid-responsive motif. *J. Biol. Chem.* **270**, 31178–31188 (1995).
300. Heldin, C. H., Miyazono, K. & Ten Dijke, P. TGF-β signalling from cell membrane to nucleus through SMAD proteins. *Nature* **390**, 465–471 (1997).
301. Feng, X. H., Zhang, Y., Wu, R. Y. & Derynck, R. The tumor suppressor Smad4/DPC4 and transcriptional adaptor CBP/p300 are coactivators for Smad3 in TGF-β-induced transcriptional activation. *Genes Dev.* **12**, 2153–2163 (1998).
302. Janknecht, R., Wells, N. J. & Hunter, T. TGF-β-stimulated cooperation of Smad proteins with the coactivators CBP/p300. *Genes Dev.* **12**, 2114–2119 (1998).
303. Melhuish, T. A. & Wotton, D. The interaction of the carboxyl terminus-binding protein with the Smad corepressor TGIF is disrupted by a holoprosencephaly mutation in TGIF. *J. Biol. Chem.* **275**, 39762–39766 (2000).
304. Gripp, K. W. *et al.* Mutations in TGIF cause holoprosencephaly and link NODAL signalling to human neural axis determination. *Nat. Genet.* **25**, 205–208 (2000).
305. Taniguchi, K., Anderson, A. E., Sutherland, A. E. & Wotton, D. Loss of tgif function causes holoprosencephaly by disrupting the Shh signaling pathway. *PLoS Genet.* **8**, (2012).
306. Kuang, C. *et al.* Intragenic deletion of Tgif causes defects in brain development. *Hum. Mol. Genet.* **15**, 3508–3519 (2006).
307. Jin, J.-Z., Gu, S., McKinney, P. & Ding, J. Expression and functional analysis ofTgif during mouse midline development. *Dev. Dyn.* **235**, 547–553 (2006).
308. Powers, S. E. *et al.* Tgif1 and Tgif2 regulate Nodal signaling and are required for gastrulation. *Development* **137**, 249–259 (2010).

309. Mar, L. & Hoodless, P. A. Embryonic fibroblasts from mice lacking Tgif were defective in cell cycling. *Mol. Cell. Biol.* **26**, 4302–10 (2006).
310. Zerlanko, B. J., Bartholin, L., Melhuish, T. A. & Wotton, D. Premature senescence and increased TGFβ signaling in the absence of Tgif1. *PLoS One* **7**, (2012).
311. Shen, J. & Walsh, C. A. Targeted disruption of Tgif, the mouse ortholog of a human holoprosencephaly gene, does not result in holoprosencephaly in mice. *Mol. Cell. Biol.* **25**, 3639–47 (2005).
312. Lee, B. K. *et al.* Tgif1 Counterbalances the Activity of Core Pluripotency Factors in Mouse Embryonic Stem Cells. *Cell Rep.* **13**, 52–60 (2015).
313. Offield, M. F. *et al.* PDX-1 is required for pancreatic outgrowth and differentiation of the rostral duodenum. *Development* **122**, 983–95 (1996).
314. Swift, G. H. *et al.* 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.* **18**, 5109–5120 (1998).
315. Goudet, G., Delhalle, S., Biemar, F., Martial, J. A. & Peers, B. Functional and cooperative interactions between the homeodomain PDX1, Pbx, and Prep1 factors on the somatostatin promoter. *J. Biol. Chem.* **274**, 4067–4073 (1999).
316. Liu, Y., MacDonald, R. J. & Swift, G. H. DNA Binding and Transcriptional Activation by a PDX1·PBX1b· MEIS2b Trimer and Cooperation with a Pancreas-specific Basic Helix-Loop-Helix Complex. *J. Biol. Chem.* **276**, 17985–17993 (2001).
317. Zhang, X. *et al.* Pax6 is regulated by Meis and Pbx homeoproteins during pancreatic development. *Dev. Biol.* **300**, 748–757 (2006).
318. diIorio, P., Alexa, K., Choe, S. K., Etheridge, L. & Sagerström, C. G. TALE-Family homeodomain proteins regulate endodermal sonic hedgehog expression and pattern the anterior endoderm. *Dev. Biol.* **304**, 221–231 (2007).
319. Xie, R. *et al.* Dynamic chromatin remodeling mediated by polycomb proteins orchestrates pancreatic differentiation of human embryonic stem cells. *Cell Stem Cell* **12**, 224–237 (2013).
320. Dor, Y., Brown, J., Martinez, O. I. & Melton, D. A. Adult pancreatic β-cells are formed by self-duplication rather than stem-cell differentiation. *Nature* **429**, 41–46 (2004).
321. Xu, X. *et al.* β Cells Can Be Generated from Endogenous Progenitors in Injured Adult Mouse Pancreas. *Cell* **132**, 197–207 (2008).
322. Stanger, B. Z. & Hebrok, M. Control of cell identity in pancreas development and regeneration. *Gastroenterology* **144**, 1170–1179 (2013).
323. Yi, P., Park, J. S. & Melton, D. A. Betatrophin: A hormone that controls pancreatic β cell proliferation. *Cell* **153**, 747–758 (2013).
324. Nir, T., Melton, D. A. & Dor, Y. Recovery from diabetes in mice by β cell regeneration.

J. Clin. Invest. **117**, 2553–2561 (2007).

325. Teta, M., Rankin, M. M., Long, S. Y., Stein, G. M. & Kushner, J. A. Growth and Regeneration of Adult β Cells Does Not Involve Specialized Progenitors. *Dev. Cell* **12**, 817–826 (2007).
326. Cogger, K. & Nostro, M. C. Recent advances in cell replacement therapies for the treatment of type 1 diabetes. *Endocrinology* **156**, 8–15 (2015).
327. Shapiro, A. M. J., Pokrywczynska, M. & Ricordi, C. Clinical pancreatic islet transplantation. *Nat. Rev. Endocrinol.* **13**, 268–277 (2017).
328. Thomson, J. A. Embryonic stem cell lines derived from human blastocysts. *Science (80-.).* **282**, 1145–1147 (1998).
329. Martin, G. R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **78**, 7634–7638 (1981).
330. Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154–156 (1981).
331. Briggs, R. & King, T. J. Transplantation of living nuclei from blastula cells into enucleated frogs' eggs. *Proc. Natl. Acad. Sci.* **38**, 455–463 (1952).
332. Gurdon, J. B. *The Developmental Capacity of Nuclei taken from Intestinal Epithelium Cells of Feeding Tadpoles*.
333. Takahashi, K. & Yamanaka, S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* **126**, 663–676 (2006).
334. Kroon, E. *et al.* Pancreatic endoderm derived from human embryonic stem cells generates glucose-responsive insulin-secreting cells in vivo. *Nat. Biotechnol.* **26**, 443–452 (2008).
335. Velazco-Cruz, L. *et al.* Acquisition of Dynamic Function in Human Stem Cell-Derived β Cells. *Stem Cell Reports* **12**, 351–365 (2019).
336. Zhang, Y. *et al.* Small Molecules Facilitate the Reprogramming of Mouse Fibroblasts into Pancreatic Lineages. *Cell Stem Cell* (2014). doi:10.1016/j.stem.2014.01.006
337. Zhu, S. *et al.* Human pancreatic beta-like cells converted from fibroblasts. *Nat. Commun.* **7**, 10080 (2016).
338. Efe, J. A. *et al.* Conversion of mouse fibroblasts into cardiomyocytes using a direct reprogramming strategy. *Nat. Cell Biol.* **13**, 215–222 (2011).
339. Kim, J. *et al.* Direct reprogramming of mouse fibroblasts to neural progenitors Normal and neoplastic nonstem cells can spontaneously convert to a stem-like state. *Pnas* **108**, 7838–7843 (2011).
340. Ladewig, J., Koch, P. & Brüstle, O. Leveling Waddington: The emergence of direct

programming and the loss of cell fate hierarchies. *Nature Reviews Molecular Cell Biology* **14**, 225–236 (2013).

341. Graf, T. Historical origins of transdifferentiation and reprogramming. *Cell Stem Cell* **9**, 504–516 (2011).
342. Jopling, C., Boue, S. & Belmonte, J. C. I. Dedifferentiation, transdifferentiation and reprogramming: Three routes to regeneration. *Nature Reviews Molecular Cell Biology* **12**, 79–89 (2011).
343. Knapp, D. & Tanaka, E. M. Regeneration and reprogramming. *Current Opinion in Genetics and Development* **22**, 485–493 (2012).
344. Spechler, S. J. Barrett's Esophagus. *N. Engl. J. Med.* **346**, 836–842 (2002).
345. Thowfeequ, S., Myatt, E. J. & Tosh, D. Transdifferentiation in developmental biology, disease, and in therapy. *Developmental Dynamics* **236**, 3208–3217 (2007).
346. Xu, J., Du, Y. & Deng, H. Direct lineage reprogramming: Strategies, mechanisms, and applications. *Cell Stem Cell* **16**, 119–134 (2015).
347. Morris, S. A. Direct lineage reprogramming via pioneer factors; a detour through developmental gene regulatory networks. *Development* (2016). doi:10.1242/dev.138263
348. Vierbuchen, T. & Wernig, M. Direct lineage conversions: Unnatural but useful? *Nature Biotechnology* **29**, 892–907 (2011).
349. Davis, R. L., Weintraub, H. & Lassar, A. B. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* **51**, 987–1000 (1987).
350. Vierbuchen, T. *et al.* Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* **463**, 1035–1041 (2010).
351. Zhou, Q. & Melton, D. A. Extreme Makeover: Converting One Cell into Another. *Cell Stem Cell* **3**, 382–388 (2008).
352. Nashun, B., Hill, P. W. & Hajkova, P. Reprogramming of cell fate: epigenetic memory and the erasure of memories past. *EMBO J.* **34**, 1296–1308 (2015).
353. Cohen, D. E. & Melton, D. Turning straw into gold: directing cell fate for regenerative medicine. *Nat. Rev. Genet.* **12**, 243–252 (2011).
354. Cavelti-Weder, C., Zumsteg, A., Li, W. & Zhou, Q. Reprogramming of pancreatic acinar cells to functional beta cells by in vivo transduction of a polycistronic construct containing Pdx1, Ngn3, MafA in mice. *Curr. Protoc. Stem Cell Biol.* **2017**, 4A.10.1-4A.10.12 (2017).
355. Heinrich, C., Spagnoli, F. M. & Berninger, B. In vivo reprogramming for tissue repair. *Nat. Cell Biol.* **17**, 204–11 (2015).
356. Wang, Y. *et al.* Long-Term Correction of Diabetes in Mice by In Vivo Reprogramming of Pancreatic Ducts. *Mol. Ther.* **0**, (2018).

357. Chen, Y.-J. *et al.* De novo formation of insulin-producing 'neo-β cell islets' from intestinal crypts. *Cell Rep.* **6**, 1046–58 (2014).
358. Banga, A., Akinci, E., Greder, L. V, Dutton, J. R. & Slack, J. M. W. In vivo reprogramming of Sox9+ cells in the liver to insulin-secreting ducts. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 15336–41 (2012).
359. Wang, Y. *et al.* Efficient generation of pancreatic β-like cells from the mouse gallbladder. *Stem Cell Res.* **17**, 587–596 (2016).
360. Galivo, F. *et al.* Reprogramming human gallbladder cells into insulin-producing β-like cells. *PLoS One* **12**, e0181812 (2017).
361. Sancho, R., Gruber, R., Gu, G. & Behrens, A. Loss of Fbw7 reprograms adult pancreatic ductal cells into α, δ, and β cells. *Cell Stem Cell* **15**, 139–53 (2014).
362. Lee, J. *et al.* Expansion and conversion of human pancreatic ductal cells into insulin-secreting endocrine cells. *Elife* **2**, (2013).
363. Romer, A. I. & Sussel, L. Pancreatic islet cell development and regeneration. *Current Opinion in Endocrinology, Diabetes and Obesity* **22**, 255–264 (2015).
364. van der Meulen, T. & Huising, M. O. Role of transcription factors in the transdifferentiation of pancreatic islet cells. *Journal of Molecular Endocrinology* **54**, R103–R117 (2015).
365. Thorel, F. *et al.* Conversion of adult pancreatic α-cells to B-cells after extreme B-cell loss. *Nature* **464**, 1149–1154 (2010).
366. Furuyama, K. *et al.* Diabetes relief in mice by glucose-sensing insulin-secreting human α-cells. *Nature* 1 (2019). doi:10.1038/s41586-019-0942-8
367. Bramswig, N. C. *et al.* Epigenomic plasticity enables human pancreatic α to β cell reprogramming. *J. Clin. Invest.* **123**, 1275–1284 (2013).
368. Wilcox, C. L., Terry, N. A., Walp, E. R., Lee, R. A. & May, C. L. Pancreatic α-Cell Specific Deletion of Mouse Arx Leads to α-Cell Identity Loss. *PLoS One* **8**, e66214 (2013).
369. Chakravarthy, H. *et al.* Converting Adult Pancreatic Islet α Cells into β Cells by Targeting Both Dnmt1 and Arx. *Cell Metab.* **25**, 622–634 (2017).
370. Ariyachet, C. *et al.* Reprogrammed Stomach Tissue as a Renewable Source of Functional β Cells for Blood Glucose Regulation. *Cell Stem Cell* **18**, 410–421 (2016).
371. Talchai, C., Xuan, S., Kitamura, T., DePinho, R. A. & Accili, D. Generation of functional insulin-producing cells in the gut by Foxo1 ablation. *Nat. Genet.* **44**, 406–412 (2012).
372. Ferber, S. *et al.* Pancreatic and duodenal homeobox gene 1 induces expression of insulingenes in liver and ameliorates streptozotocin-induced hyperglycemia. *Nat. Med.* **6**, 568–572 (2000).

373. Sapir, T. *et al.* Cell-replacement therapy for diabetes: Generating functional insulin-producing tissue from adult human liver cells. *Proc. Natl. Acad. Sci.* **102**, 7964–7969 (2005).
374. Ber, I. *et al.* Functional, persistent, and extended liver to pancreas transdifferentiation. *J. Biol. Chem.* **278**, 31950–7 (2003).
375. Horb, M. E., Shen, C. N., Tosh, D. & Slack, J. M. W. Experimental conversion of liver to pancreas. *Curr. Biol.* **13**, 105–15 (2003).
376. Cao, L.-Z., Tang, D.-Q., Horb, M. E., Li, S.-W. & Yang, L.-J. High glucose is necessary for complete maturation of Pdx1-VP16-expressing hepatic cells into functional insulin-producing cells. *Diabetes* **53**, 3168–78 (2004).
377. Kaneto, H. *et al.* PDX-1/VP16 fusion protein, together with neurod or Ngn3, markedly induces insulin gene transcription and ameliorates glucose tolerance. *Diabetes* **54**, 1009–1022 (2005).
378. Teo, A. K. K. *et al.* PDX1 binds and repreßes hepatic genes to ensure robust pancreatic commitment in differentiating human embryonic stem cells. *Stem Cell Reports* **4**, 578–590 (2015).
379. Kojima, H. *et al.* NeuroD-betacellulin gene therapy induces islet neogenesis in the liver and reverses diabetes in mice. *Nat. Med.* **9**, 596–603 (2003).
380. Miyatsuka, T. *et al.* Ectopically expressed PDX-1 in liver initiates endocrine and exocrine pancreas differentiation but causes dysmorphogenesis. *Biochem. Biophys. Res. Commun.* **310**, 1017–25 (2003).
381. Wang, A. Y., Ehrhardt, A., Xu, H. & Kay, M. A. Adenovirus Transduction is Required for the Correction of Diabetes Using Pdx-1 or Neurogenin-3 in the Liver. *Mol. Ther.* **15**, 255–263 (2007).
382. Yechoor, V. *et al.* Neurogenin3 is sufficient for transdetermination of hepatic progenitor cells into neo-islets in vivo but not transdifferentiation of hepatocytes. *Dev. Cell* **16**, 358–73 (2009).
383. Nagasaki, H. *et al.* Generation of insulin-producing cells from the mouse liver using β cell-related gene transfer including Mafa and Mafb. *PLoS One* **9**, e113022 (2014).
384. Li, R. *et al.* Gene therapy with neurogenin3, betacellulin and SOCS1 reverses diabetes in NOD mice. *Gene Ther.* **22**, 876–882 (2015).
385. Yatoh, S. *et al.* NeuroD and reaggregation induce beta-cell specific gene expression in cultured hepatocytes. *Diabetes. Metab. Res. Rev.* **23**, 239–49 (2007).
386. Yang, Y., Akinci, E., Dutton, J. R., Banga, A. & Slack, J. M. W. Stage specific reprogramming of mouse embryo liver cells to a beta cell-like phenotype. *Mech. Dev.* **130**, 602–612 (2013).
387. Yang, L. *et al.* In vitro trans-differentiation of adult hepatic stem cells into pancreatic

endocrine hormone-producing cells. *Proc. Natl. Acad. Sci.* **99**, 8078–8083 (2002).

388. Tang, D.-Q. *et al.* Role of Pax4 in Pdx1-VP16-mediated liver-to-endocrine pancreas transdifferentiation. *Lab. Invest.* **86**, 829–41 (2006).
389. Liu, J. *et al.* Direct differentiation of hepatic stem-like WB cells into insulin-producing cells using small molecules. *Sci. Rep.* **3**, 1185 (2013).
390. Zalzman, M. *et al.* Reversal of hyperglycemia in mice by using human expandable insulin-producing cells differentiated from fetal liver progenitor cells. *Proc. Natl. Acad. Sci.* **100**, 7253–7258 (2003).
391. Gefen-Halevi, S. *et al.* NKX6.1 promotes PDX-1-induced liver to pancreatic β-cells reprogramming. *Cell. Reprogram.* **12**, 655–64 (2010).
392. Meivar-Levy, I. *et al.* Human Liver Cells Expressing Albumin and Mesenchymal Characteristics Give Rise to Insulin-Producing Cells. *J. Transplant.* **2011**, 1–12 (2011).
393. Donelan, W. *et al.* Pancreatic and duodenal homeobox gene 1 (Pdx1) down-regulates hepatic transcription factor 1 alpha (HNF1α) expression during reprogramming of human hepatic cells into insulin-producing cells. *Am. J. Transl. Res.* **7**, 995–1008 (2015).
394. Ham, D.-S. *et al.* Generation of Functional Insulin-Producing Cells from Neonatal Porcine Liver-Derived Cells by PDX1/VP16, BETA2/NeuroD and MafA. *PLoS One* **8**, e79076 (2013).
395. Lee, J. *et al.* Activation of innate immunity is required for efficient nuclear reprogramming. *Cell* **151**, 547–558 (2012).
396. Hingorani, S. R. *et al.* Preinvasive and invasive ductal pancreatic cancer and its early detection in the mouse. *Cancer Cell* **4**, 437–50 (2003).
397. Mar, L. & Hoodless, P. A. Embryonic Fibroblasts from Mice Lacking Tgif Were Defective in Cell Cycling. *Mol. Cell. Biol.* **26**, 4302–4310 (2006).
398. Abe, T. *et al.* Establishment of conditional reporter mouse lines at ROSA26 locus for live cell imaging. *Genesis* **49**, 579–590 (2011).
399. Fougère-Deschatrette, C. *et al.* Plasticity of Hepatic Cell Differentiation: Bipotential Adult Mouse Liver Clonal Cell Lines Competent to Differentiate In Vitro and In Vivo. *Stem Cells* **24**, 2098–2109 (2006).
400. Zhang, K. *et al.* In Vitro Expansion of Primary Human Hepatocytes with Efficient Liver Repopulation Capacity. *Cell Stem Cell* **23**, 806-819.e4 (2018).
401. Philippeos, C. *et al.* Spatial and Single-Cell Transcriptional Profiling Identifies Functionally Distinct Human Dermal Fibroblast Subpopulations. *J. Invest. Dermatol.* **138**, 811–825 (2018).
402. Chen, G. *et al.* Chemically defined conditions for human iPSC derivation and culture. *Nat. Methods* **8**, 424–429 (2011).

403. Pennarossa, G. *et al.* Brief demethylation step allows the conversion of adult human skin fibroblasts into insulin-secreting cells. *Proc. Natl. Acad. Sci.* (2013). doi:10.1073/pnas.1220637110
404. Ye, Z., Yu, X. & Cheng, L. Lentiviral gene transduction of mouse and human stem cells. *Methods Mol. Biol.* **430**, 243–253 (2007).
405. Szymczak, A. L. & Vignali, D. A. A. Development of 2A peptide-based strategies in the design of multicistronic vectors. *Expert Opinion on Biological Therapy* **5**, 627–638 (2005).
406. Zufferey, R. *et al.* Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery. *J. Virol.* **72**, 9873–80 (1998).
407. Pfaffl, M. W., Horgan, G. W. & Dempfle, L. Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res.* **30**, e36 (2002).
408. Schaeren-Wiemers, N. & Gerfin-Moser, A. A single protocol to detect transcripts of various types and expression levels in neural tissue and cultured cells: in situ hybridization using digoxigenin-labelled cRNA probes. *Histochemistry* **100**, 431–440 (1993).
409. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
410. Powers, S. E. *et al.* Tgif1 and Tgif2 regulate Nodal signaling and are required for gastrulation. *Development* **137**, 249–259 (2010).
411. Magnuson, M. A. & Osipovich, A. B. Pancreas-specific Cre driver lines and considerations for their prudent use. *Cell Metab.* **18**, 9–20 (2013).
412. Papakrivopoulou, E., Dean, C. H., Copp, A. J. & Long, D. A. Planar cell polarity and the kidney. *Nephrology Dialysis Transplantation* **29**, 1320–1326 (2014).
413. Yang, Y. & Mlodzik, M. Wnt-Frizzled/Planar Cell Polarity Signaling: Cellular Orientation by Facing the Wind (Wnt). *Annu. Rev. Cell Dev. Biol.* **31**, 623–646 (2015).
414. Cortijo, C., Gouzi, M., Tissir, F. & Grapin-Botton, A. Planar Cell Polarity Controls Pancreatic Beta Cell Differentiation and Glucose Homeostasis. *Cell Rep.* (2012). doi:10.1016/j.celrep.2012.10.016
415. Pauerstein, P. T. *et al.* A radial axis defined by semaphorin-to-neuropilin signaling controls pancreatic islet morphogenesis. *Development* **144**, 3744–3754 (2017).
416. Helker, C. S. M. *et al.* A whole organism small molecule screen identifies novel regulators of pancreatic endocrine development. *Dev.* **146**, (2019).
417. Zeilinger, K., Freyer, N., Damm, G., Seehofer, D. & Knöspel, F. Cell sources for in vitro human liver cell culture models. *Exp. Biol. Med.* **241**, 1684–1698 (2016).
418. Richert, L. *et al.* Cytotoxicity evaluation using cryopreserved primary human

hepatocytes in various culture formats. *Toxicol. Lett.* **258**, 207–215 (2016).

419. Baker, T. K. *et al.* Temporal gene expression analysis of monolayer cultured rat hepatocytes. *Chem. Res. Toxicol.* **14**, 1218–1231 (2001).
420. Lisowski, L. *et al.* Selection and evaluation of clinically relevant AAV variants in a xenograft liver model. *Nature* (2014). doi:10.1038/nature12875
421. Ellis, B. L. *et al.* A survey of ex vivo/in vitro transduction efficiency of mammalian primary cells and cell lines with Nine natural adeno-associated virus (AAV1-9) and one engineered adeno-associated virus serotype. *Virology Journal* **10**, (2013).
422. Mall, M. *et al.* Myt1l safeguards neuronal identity by actively repressing many non-neuronal fates. *Nature* **544**, 245–249 (2017).
423. Cerdá-Esteban, N. & Spagnoli, F. M. Glimpse into Hox and tale regulation of cell differentiation and reprogramming. *Dev. Dyn.* **243**, 76–87 (2014).
424. Cerdá-Esteban, N. & Spagnoli, F. M. Glimpse into Hox and tale regulation of cell differentiation and reprogramming. *Dev. Dyn.* **243**, 76–87 (2014).
425. Huang, P. *et al.* Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors. *Nature* **475**, 386–9 (2011).
426. Zhu, S. *et al.* Human pancreatic beta-like cells converted from fibroblasts. *Nat. Commun.* **7**, (2016).
427. Mullen, A. C. & Wrana, J. L. TGF-β family signaling in embryonic and somatic stem-cell renewal and differentiation. *Cold Spring Harb. Perspect. Biol.* **9**, (2017).
428. Pennarossa, G. *et al.* Brief demethylation step allows the conversion of adult human skin fibroblasts into insulin-secreting cells. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 8948–53 (2013).
429. Pennarossa, G. *et al.* Epigenetic Erasing and Pancreatic Differentiation of Dermal Fibroblasts into Insulin-Producing Cells are Boosted by the Use of Low-Stiffness Substrate. *Stem Cell Rev. Reports* **14**, 398–411 (2018).
430. Tanabe, K., Haag, D. & Wernig, M. Direct somatic lineage conversion. *Philosophical Transactions of the Royal Society B: Biological Sciences* **370**, (2015).
431. Johnson, J. D. The quest to make fully functional human pancreatic beta cells from embryonic stem cells: climbing a mountain in the clouds. *Diabetologia* **59**, 2047–2057 (2016).
432. Iwafuchi-Doi, M. & Zaret, K. S. Pioneer transcription factors in cell reprogramming. *Genes and Development* (2014). doi:10.1101/gad.253443.114
433. Sekiya, S. & Suzuki, A. Direct conversion of mouse fibroblasts to hepatocyte-like cells by defined factors. *Nature* (2011). doi:10.1038/nature10263
434. Guguen-Guillouzo, C., Corlu, A. & Guillouzo, A. Stem cell-derived hepatocytes and their use in toxicology. *Toxicology* **270**, 3–9 (2010).

435. Zakikhan, K., Pournasr, B., Vosough, M. & Nassiri-Asl, M. In Vitro Generated Hepatocyte-Like Cells: A Novel Tool in Regenerative Medicine and Drug Discovery. *Cell J.* **19**, 204–217
436. Shinozawa, T., Yoshikawa, H. Y. & Takebe, T. Reverse engineering liver buds through self-driven condensation and organization towards medical application. *Developmental Biology* **420**, 221–229 (2016).
437. Wapinski, O. L. *et al.* XHierarchical mechanisms for direct reprogramming of fibroblasts to neurons. *Cell* **155**, 621 (2013).
438. Mikkelsen, T. S. *et al.* Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454**, 49–55 (2008).
439. Flasse, L. *et al.* Apical restriction of the planar cell polarity component VANGL in pancreatic ducts is required to maintain epithelial integrity. *bioRxiv* 778332 (2019). doi:10.1101/778332
440. Bader, E. *et al.* Identification of proliferative and mature β-cells in the islets of langerhans. *Nature* **535**, 430–434 (2016).
441. Kim, H. J. *et al.* Wnt5 signaling in vertebrate pancreas development. *BMC Biol.* **3**, (2005).
442. Cano, D. A., Sekine, S. & Hebrok, M. Primary Cilia Deletion in Pancreatic Epithelial Cells Results in Cyst Formation and Pancreatitis. *Gastroenterology* **131**, 1856–1869 (2006).
443. Taniguchi, K. *et al.* Genetic and Molecular Analyses indicate independent effects of TGIFs on Nodal and Gli3 in neural tube patterning. *Eur. J. Hum. Genet.* (2017). doi:10.1038/ejhg.2016.164
444. Smart, N. G. *et al.* Conditional expression of Smad7 in pancreatic β cells disrupts TGF-β signaling and induces reversible diabetes mellitus. *PLoS Biol.* **4**, 200–209 (2006).
445. El-Gohary, Y. *et al.* Smad signaling pathways regulate pancreatic endocrine development. *Dev. Biol.* **378**, 83–93 (2013).
446. Bardeesy, N. *et al.* Smad4 is dispensable for normal pancreas development yet critical in progression and tumor biology of pancreas cancer. *Genes Dev.* **20**, 3130–3146 (2006).
447. Simeone, D. M. *et al.* Islet hypertrophy following pancreatic disruption of Smad4 signaling. *Am. J. Physiol. - Endocrinol. Metab.* **291**, (2006).
448. Nolan, C. *et al.* Type 2 diabetes across generations: from pathophysiology to prevention and management (2011). doi:10.1016/S0140-6736(11)60614-4
449. Milone, M. C. & O'Doherty, U. Clinical use of lentiviral vectors. *Leukemia* **32**, 1529–1541 (2018).

450. Poletti, V. *et al.* Preclinical Development of a Lentiviral Vector for Gene Therapy of X-Linked Severe Combined Immunodeficiency. *Mol. Ther. - Methods Clin. Dev.* **9**, 257–269 (2018).

10. PUBLICATIONS

1. Cerdá-Esteban N., Naumann H., **Ruzittu S**., Mah N., Pongrac I.M., Cozzitorto C., Hommel A., Andrade-Navarro M.A., Bonifacio E., Spagnoli F.M. Stepwise reprogramming of liver cells to a pancreas progenitor state by the transcriptional regulator Tgif2. *Nat Commun*; 8:14127 (2017).

2. **Ruzittu S.**, Willnow D. & Spagnoli F.M. Direct lineage reprogramming: harnessing cell plasticity between liver and pancreas. *Cold Spring Harb Perspect Biol*. a035626 (2019).

3. Cozzitorto C., Mueller L., **Ruzittu S**., Mah N., Willnow D., Risolino M., Selleri L., Spagnoli F.M. A specialized niche in the pancreatic microenvironment regulates endocrine differentiation (second round revision).