

Identification of Protogenin as a novel surface marker for early cortical neural stem cells

Dissertation

to obtain the academic degree
Doctor of Philosophy in Natural Science (PhD)

submitted to the Department of Biology, Chemistry, Pharmacy of
Freie Universität Berlin

by
Amèlia Aragonés Hernández

2022

The research work for this dissertation was performed from December 2017 to November 2022 under the supervision of Dr. Yechiel Elkabetz at the Max Planck Institute for Molecular Genetics in Berlin, Germany. The dissertation was submitted in December 2022 to the Department of Biology, Chemistry and Pharmacy of the Freie Universität Berlin, Germany.

1st Reviewer: Prof. Dr. Katja Nowick

Institute of Biology - Zoology
Freie Universität Berlin

2nd Reviewer: Prof. Dr. Mathias F. Wernet

Institute of Biology
Freie Universität Berlin

Date of defense: 27.03.2023

Acknowledgements

I am very grateful to Dr. Yechiel Elkabetz for giving me the opportunity to work on this amazing project for my doctoral studies, as well as for all his helpful discussions. I am mostly thankful for the role he had in my professional and personal growth during these years.

I also wish to express my gratitude towards my TAC committee, Prof. Dr. Katja Nowick and Dr. Edda Schulz for their valuable suggestions and support throughout this project.

I would like to thank all the people that have made this project possible. I am very thankful to all the lab members during these years for their friendship, thoughtful discussions, and guidance through the rough times. I would like to thank Sneha Arora and Naresh Mutukula for showing me the ropes and Anastasios Balaskas for always willing to give a helpful hand. A special thanks to Daniel Rosebrock who was crucial for this work, and essential for creating a fun and enjoyable environment in the lab.

I would like to express my appreciation to all the members of the Meissner department for helping me in any and every possible way, and most importantly for creating a friendly atmosphere. I would specially like to thank Abhishek Sampath Kumar for involving me on his awesome spatial transcriptomics project.

Finally, I would also like to thank all those who have helped me along the way either at a professional or personal level. A big thank you for being part of my journey!

Declaration of Independence

Herewith I certify that I have prepared and written my thesis independently and that I have not used any sources and aids other than those indicated by me. Intellectual property of other authors has been marked accordingly. I also declare that I have not applied for an examination procedure at any other institution and that I have not submitted the dissertation in this or any other form to any other faculty as a dissertation.

Date: 05.12.2022

Table of Contents

List of figures	viii
List of tables	xi
Abbreviations	xii
Abstract	xv
Zusammenfassung	xvii
1. Introduction	19
1.1. The formation of the vertebrate nervous system.....	19
1.2. Early neural induction: How is the CNS specified?.....	20
1.2.1. First insights into embryonic neural induction: What makes the ectoderm decide to acquire a neural fate?.....	20
1.2.2. Modern insights: What are the major signaling pathways involved in embryonic neural induction?	21
1.3. Vertebrate telencephalic development	24
1.4. Telencephalic patterning: progressive molecular and cellular specification	25
1.4.1. Patterning along the AP axis	25
1.4.2. Patterning along the DV axis	26
1.5. The cerebral cortex	28
1.5.1. Composition of the cortical layers	30
1.6. How can we study cortical development?	32
1.6.1. Stem cell models of cortical development: derivation of neural rosettes	32
1.6.2. Derivation and characterization of human cortical NSCs in vitro	36
1.6.3. The emergence of induced pluripotent stem cells	38
1.7. Single-cell RNA sequencing technology: unprecedented insight into human embryonic development.....	38
2. Main aim	41
3. Materials and methods	42
3.1. Reagents	42

3.2. Equipment and instruments	43
3.3. Cell culture	45
3.3.1. hiPSC line	45
3.3.2. Cell propagation	45
3.3.3. Passaging hiPSCs	45
3.3.4. Freezing hiPSCs	45
3.4. Neural induction media	46
3.4.1. KSR medium	46
3.4.2. N2 medium	46
3.4.3. Neurobasal medium	46
3.5. Neural induction protocols	47
3.5.1. Monolayer neural differentiation	47
3.5.2. Generation of neural organoids	48
3.6. Immunofluorescence and confocal microscopy	48
3.7. Fluorescence-activated cell sorting of live cells	51
3.8. RNA extraction	53
3.9. cDNA synthesis	53
3.10. Real-time quantitative PCR	54
3.11. Preparation of RNA-seq libraries	56
3.11.1. Purification and fragmentation of mRNA	56
3.11.2. Synthesis of the first strand of cDNA	57
3.11.3. Synthesis of the second strand of cDNA	57
3.11.4. End repair	57
3.11.5. 3'-ends adenylation	58
3.11.6. Adapters ligation	58
3.11.7. Enrich DNA Fragments	58
3.12. RNA-Seq processing and analysis	59
3.13. Single cell RNA sequencing	59
3.13.1. Sample preparation	59
3.13.2. Library construction	60
3.13.3. GEM generation and barcoding	60
3.13.4. Post GEM-RT cleanup and cDNA amplification	60

3.13.5. 3'-Gene expression library construction.....	62
3.13.6. Sample indexing PCR.....	64
3.13.7. Single cell RNA sequencing data processing.....	65
4. Results	67
4.1. Characterization of hiPSCs-derived cortical progenitors by scRNA-seq.....	67
4.2. Identification of potential surface markers for early cortical NSCs	73
4.3. FACS analysis confirms early expression of PRTG and MCAM following neural induction.....	81
4.4. PRTG sorting on day 12 enriches for early cortical NSCs	82
4.5. Characterization of sorted populations further confirms PRTG validity as an early cortical marker	87
4.6. PRTG expression emerges after four days of cortical neural induction	94
4.7. Sorting for PRTG at day 5 enriches for anterior telencephalic identity	95
4.8. PRTG demarcates early specification of anterior identity and prospectively isolates distinct telencephalic subpopulations.....	100
5. Discussion	108
6. Bibliography	116
7. List of publications	128

List of figures

Figure 1.1. Scanning electron microscopy (SEM) of neurulation in chick embryos	19
Figure 1.2. SEM of medial sagittal views of mouse embryos depicting the brain vesicles	20
Figure 1.3. The TGFB/BMP signaling pathway.....	22
Figure 1.4. The canonical WNT signaling pathway	23
Figure 1.5. Embryonic vertebrate telencephalon.	25
Figure 1.6. Early telencephalic patterning.....	28
Figure 1.7. The developing human cortex	28
Figure 1.8. Cortical radial glia potential	30
Figure 1.9. Cortical neurogenesis	31
Figure 1.10. Rosette formation as a readout to measure cortical specification.....	33
Figure 1.11. Seminal neural induction protocols	34
Figure 1.12. Triple-i protocol to derive neural lineages from stem cells promotes cortical fates	35
Figure 1.13. Progression of NSC states throughout cortical differentiation <i>in vitro</i>	37
Figure 1.14. Droplet-based single-cell sequencing.....	39
Figure 4.1. Schematic overview of the neural induction protocol	67
Figure 4.2. scRNA-seq analysis of day 12 hiPSC-derived neural progenitors.....	69
Figure 4.3. scRNA-seq analysis of day 35 hiPSC-derived neural progenitors.....	71
Figure 4.4. scRNA-seq analysis of day 50 hiPSC-derived neural progenitors.....	72
Figure 4.5. scRNA-seq merged data depicting main clusters and cell compositions of the different time points	73
Figure 4.6. Stage-specific gene signature of neural stem cells and derivatives	74

Figure 4.7. Selected potential candidates for early cortical NSCs.....	76
Figure 4.8. Expression of candidate surface markers through 2D differentiation	78
Figure 4.9. Expression of PRTG and MCAM throughout 3D differentiation	79
Figure 4.10. PRTG and MCAM expression during human corticogenesis.....	80
Figure 4.11. Surface markers expression profiles in neural cell types	82
Figure 4.12. FACS analysis average cell frequencies of PRTG and MCAM	82
Figure 4.13. Gating strategy followed to sort and collect samples for RNA-seq	83
Figure 4.14. Correlation matrix of RNA-Seq datasets obtained from sorted cells	84
Figure 4.15. RNA-seq analysis identifies specific gene expression profiles based on PRTG sorting	86
Figure 4.16. PRTG expression is enriched in rosettes co-expressing PAX6 and FOXG1 on day 12	87
Figure 4.17. PRTG sorting strategy and phenotype of replated cells at day 12 of neural induction	89
Figure 4.18. PRTG sorting enriches for NSCs populations at expenses of neurons.....	90
Figure 4.19. PRTG sorting enriches for cortical NSCs	91
Figure 4.20. PRTG positive cells retain a higher neurogenic potential	92
Figure 4.21. Re-sorting of replated cells at day 35 confirms prospective isolation and enrichment of cortical lineages by means of PRTG sorting at day 12	93
Figure 4.22. PRTG expression analysis in early days of neural induction.....	95
Figure 4.23. PRTG sorting strategy and phenotype of replated cells at day 5 of neural induction	96
Figure 4.24. Sorting for PRTG at day 5 enriches for anterior neural identity	97
Figure 4.25. Sorting for PRTG at day 5 enriches for anterior forebrain	98
Figure 4.26. PRTG sorting at day 5 enriches for rosette forming cells.....	98
Figure 4.27. PRTG sorting at day 5 enriches for cortical identity.....	99
Figure 4.28. Sorting strategy at day 5 for scRNA-seq analysis.....	101
Figure 4.29. PRTG demarcates early specification of anterior neural identity	102

Figure 4.30. Differential gene expression analysis of day 5 PRTG sorted subpopulations .	104
Figure 4.31. PRTG prospectively isolates distinct telencephalic subpopulations	105
Figure 4.32. Differential gene expression analysis of day 10 prospectively sorted subpopulations	107
Figure 5.1. Experimental workflow of a screening strategy using PRTG as a readout	115

List of tables

Table 3.1. Primary antibodies	50
Table 3.2. Secondary antibodies	51
Table 3.3. Primary antibodies use for the pre-staining of cells for FACS analysis	52
Table 3.4. Secondary antibodies use for the detection of surface antigens for FACS analysis	52
Table 3.5. Brief protocol for cDNA synthesis.	54
Table 3.6. Brief protocol for RT-qPCR	55
Table 3.7. RT-qPCR oligonucleotides	55
Table 3.8. Brief protocol for cDNA amplification.	61
Table 3.9. Brief protocol for cDNA fragmentation.....	62
Table 3.10. Brief protocol for adaptor ligation.	63
Table 3.11. Brief protocol for the indexing PCR	64

Abbreviations

2D	Two-dimensional
3D	Three-dimensional
AEP	Anterior entopeduncular
AP	Anterior-posterior
BMP4	Bone morphogenetic protein 4
BPs	Basal progenitors
CEMIP2	Cell Migration Inducing Hyaluronidase 2
CH	Cortical hem
CNS	Central nervous system
CP	Cortical plate
CPN	Callosal projection neurons
CR	Cajal–Retzius cells
CThPN	Corticothalamic projection neurons
dLGE	dorsal Lateral ganglionic eminence
DP	Dorsal pallium
DV	Dorso-ventral
EBs	Embryoid bodies
E-RG	Early radial glial cells
FACS	Fluorescent-activated cell sorting
FGF	Fibroblast Growth Factor
FN	Fibronectin
FPKM	Fragments per kilo base of transcript per million mapped fragments
FSC	Forward scatter
GEMs	Gel Bead-In-Emulsions

GN	Granular neurons
(h)iPSCs	(Human) induced pluripotent stem cells
IGDCC3	Immunoglobulin Superfamily DCC Subclass Member 3
IPs	Intermediate progenitors
IZ	Intermediate zone
Lam	Laminin
LGR5	Leucine Rich Repeat Containing G Protein-Coupled Receptor 5
LNP	Long-term cultured progenitors
LP	Lateral pallium
LR	Left-right
L-RG	Late radial glial cells
MCAM	Melanoma Cell Adhesion Molecule
MEFs	Mouse embryonic fibroblasts
MGE	Medial ganglionic eminence
MP	Medial pallium
M-RG	Mid radial glial cells
MSCs	Mesenchymal stem cells
MZ	Marginal zone
NE	Neuroepithelial cells
NSCs	Neural stem cells
oRG	outer Radial glia
pcw	Post-conception week
PO	Polyornithine
POA	Preoptic area
PRTG	Protogenin
PSCs	Pluripotent stem cells

RG	Radial glia
RT-qPCR	Real time quantitative PCR
SCPN	Subcerebral projection neurons
scRNA-seq	Single-cell RNA sequencing
SD	Standard deviation
SDC1	Syndecan 1
SEM	Scanning electron microscopy
SFEB	Serum-free suspension culture
SHH	Sonic hedgehog
SP	Subplate
SPN	Subplate neurons
SSC	Side scatter
STAMPs	Single-cell transcriptomes attached to microparticles
STAMPs	Single-cell transcriptomes attached to microparticles
SVZ	Subventricular zone
TFs	Transcription factors
TGFB	Transforming Growth factor beta
Triple-i	Triple-inhibition
UMAP	Uniform manifold approximation and projection
UMIs	Unique molecular identifiers
vLGE	ventral Lateral ganglionic eminence
VP	Ventral pallium
VZ	Ventricular zone
WNT	Wingless/Integrated

Abstract

During mammalian corticogenesis, a wide diversity of neural stem cells (NSCs) orchestrate the development and organization of the cortex. The pool of NSCs initially expands through proliferative symmetric divisions, and sequentially starts dividing asymmetrically to give rise to the diverse cell types residing within the cortical layers. Throughout this process, cortical NSCs undergo extensive modifications in their transcriptomic profile and chromatin landscape contributing to the formation of heterogeneous progenitor populations. Although much progress has been made towards understanding cell-fate specification during human corticogenesis the mechanisms responsible for the temporal lineage specification of NSCs remain largely unknown. Understanding the variability of these distinct NSC populations is key for developing an *in vitro* system that allows for the homogeneous and unlimited culture of the desired NSC type which is crucial for cell replacement-based therapies. Hence, one of the main aims in our lab is to identify and discern these distinct NSC types which sequentially appear during cortical development with the objective to better understand these cell stages and, eventually, being able to manipulate them *in vitro*.

In order to address this question, my project is focused on developing a strategy to isolate the early cortical NSC population for its characterization and potential manipulation. The main approach is to identify a cell surface marker to enable the isolation of these cells from our *in vitro* culture by fluorescence-activated cell sorting. By profiling our hiPSC-derived cortical progenitors at different stages by means of single-cell RNA sequencing, we selected potential candidate markers that were validated using immunofluorescence and sequencing methods.

In this study, we identify Protogenin (PRTG) as a novel surface marker for early human cortical NSCs that can be used to isolate this population *in vitro*. We provide evidence that early expression of the novel marker correlates with cortical lineage specification. Furthermore, by sorting for such marker at early stages of neural induction we can prospectively isolate three distinct cortical subpopulations, resulting in highly pure subtype-specific NSC cultures.

These findings illustrate the utility of PRTG cell-surface sorting for enriching early cortical NSCs in culture and, thus, aiding to develop a more robust and homogenous differentiation protocol. Ultimately, such knowledge should facilitate the generation of highly pure stage- and region-specific NSC populations from patient-derived samples which would provide a reliable source for cell replacement and regenerative therapies.

Zusammenfassung

Während der Kortikogenese bei Säugetieren ist eine große Vielfalt an neuronalen Stammzellen (NSCs) für die Entwicklung und Organisation des Kortex verantwortlich. Der Pool der NSCs vergrößert sich zunächst durch proliferative symmetrische Teilungen und beginnt dann, sich asymmetrisch zu teilen, um die verschiedenen Zelltypen in den Kortikalschichten hervorzubringen. Während dieses Prozesses erfahren die kortikalen NSCs umfangreiche Veränderungen in ihrem transkriptionellen Profil und ihrer Chromatinlandschaft, die zur Bildung heterogener Progenitorpopulationen beitragen. Obwohl große Fortschritte beim Verständnis der Zellspezifikation während der menschlichen Kortikogenese erzielt wurden, sind die Mechanismen, die für die zeitliche Spezifikation der NSC verantwortlich sind, noch weitgehend unbekannt. Das Verständnis der Variabilität dieser unterschiedlichen NSC-Populationen ist der Schlüssel zur Entwicklung eines in-vitro Systems, das die homogene und unbegrenzte Kultur des gewünschten NSC-Typs ermöglicht, was für zellbasierte Therapien von entscheidender Bedeutung ist. Eines der Hauptziele dieser Arbeit ist daher die Identifizierung und Unterscheidung dieser unterschiedlichen NSC-Typen, die während der kortikalen Entwicklung nacheinander auftreten. Ziel ist es, diese Zellstadien besser zu verstehen und sie schließlich in-vitro manipulieren zu können.

Um diese Frage zu beantworten, konzentriert sich diese Studie auf die Entwicklung einer Strategie zur Isolierung einer frühen kortikalen NSC-Population, um sie zu charakterisieren und möglicherweise zu manipulieren.

In dieser Studie identifizieren wir mittels Einzelzell RNA-Sequenzierung Protogenin (PRTG) als einen neuartigen Oberflächenmarker für frühe humane kortikale NSCs, der zur Isolierung dieser Population in vitro verwendet werden kann. Wir konnten nachweisen, dass die frühe Expression dieses Markers mit der Spezifikation der kortikalen Abstammungslinie korreliert. Darüber hinaus können wir durch Fluoreszenz aktivierte Zellsortierung anhand dieses Markers in frühen Stadien der neuronalen Induktion prospektiv drei verschiedene kortikale Subpopulationen isolieren, was zu hochreinen subtypspezifischen NSC-Kulturen führt.

Diese Ergebnisse veranschaulichen den Nutzen der PRTG-basierten Zellsortierung für die Anreicherung früher kortikaler NSCs in Kultur und helfen somit bei der Entwicklung eines robusteren und homogeneren Differenzierungsprotokolls. Letztendlich ermöglicht dieses Wissen die Erzeugung hochreiner stadien- und regionenspezifischer NSC-Populationen aus Patientenproben, die eine zuverlässige Quelle für Zellersatz und regenerative Therapien darstellen würden.

1. Introduction

1.1. The formation of the vertebrate nervous system

Neurulation -the process by which the early nervous system is established- starts directly after gastrulation once the three germ layers have been formed: the ectoderm, the endoderm, and the mesoderm. At the dorsal midline of the embryo, the ectodermal germ layer specializes into a thickened pseudostratified epithelium- the neural plate- instructed by the anterior endoderm and the notochord. The neural plate starts bending and neural folds form bilaterally creating the neural groove. Finally, epithelial fusion of the tips of the neural folds culminates in the closure and formation of the neural tube.

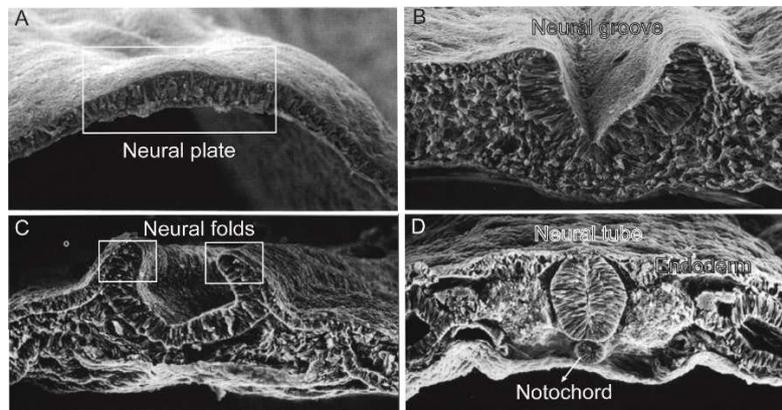


Figure 1.1. Scanning electron microscopy (SEM) of neurulation in chick embryos. The epidermis develops into a layer of pseudostratified columnar epithelium becoming the neural plate (A). In the midline of the neuroectoderm the neural groove forms and begins to elevate (B). Neural folds converge toward each other (C) to close and form the neural tube (D). Adapted from (Gilbert, 2000).

Upon closure, the neural tube undergoes a series of expansions and constrictions to form the primary brain vesicles: the Prosencephalon (Forebrain), the Mesencephalon (Midbrain), and the Rhombencephalon (Hindbrain) (*Figure 1.2*). These three primary brain vesicles further develop generating the five secondary brain vesicles: the Telencephalon and the Diencephalon (derived from the Prosencephalon); the Mesencephalon; and the Metencephalon and Myelencephalon (derived from the Rhombencephalon), which will give rise to the different structures of the adult central nervous system (CNS)(Ishikawa et al., 2012).

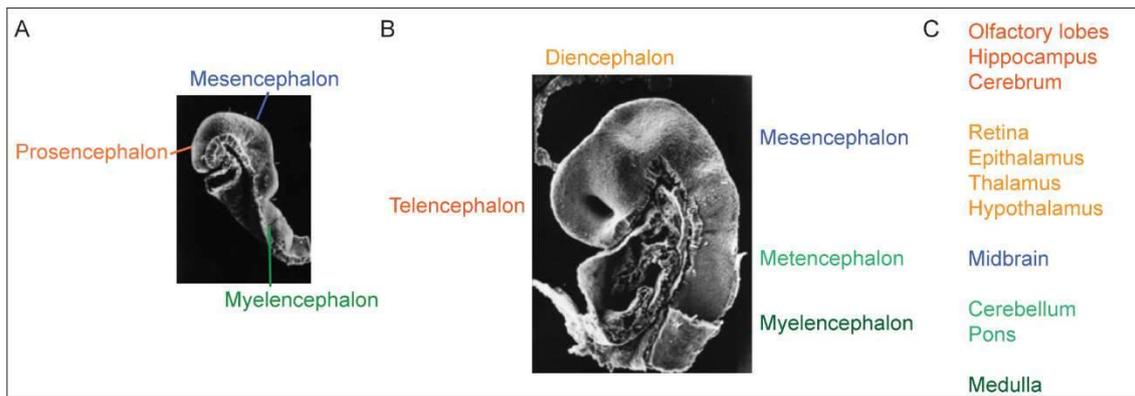


Figure 1.2. SEM of medial sagittal views of mouse embryos depicting the brain vesicles. A) E8.0 mouse section depicting the three primary vesicles. B) By E10.0 the five secondary vesicles have started to form. C) List of the final structures in the adult brain derived from each vesicle (color-coded) (SEM images from (Jacobson & Tam, 1982)).

1.2. Early neural induction: How is the CNS specified?

1.2.1. First insights into embryonic neural induction: What makes the ectoderm decide to acquire a neural fate?

Two models were developed in the mid 90's to explain the early patterning and specification of the CNS which begins at neural induction. The first insight into how the neural plate is established came from the famous experiments of Mangold and Spemann which led to the discovery of the so-called Spemann-Mangold organizer. In 1924, Mangold and Spemann postulated a first model based on their intra-species transplantation experiments involving developing amphibian embryos. These were based on dissecting and transplanting the dorsal portion of the developing embryo that was thought to serve as an 'organizing center' into the ventral portion of another embryo. This resulted in the formation of a second set of dorsal axial structures on the ventral side of the host embryo, including a well-organized secondary neural plate (Mangold & Spemann, 1927). Such seminal findings introduced the concept of an organizer center as a specific population of cells capable of inducing and assembling the various tissues in the embryo along the three body axes: 1) the anterior (rostral)–posterior (caudal) (AP) axis; 2) the dorso–ventral (DV) axis; and 3) the left–right (LR) axis. At the time, the neural inducing factors generated by the Spemann-Mangold organizer responsible for inducing neural tissues were unknown. Eventually a key player named Noggin was identified as the protein which was

secreted by the organizer and had the dorsalizing effect (Smith & Harland, 1992). Nevertheless, it was not until 1995 that the final piece of the puzzle was discovered. Wilson and Hemmati-Brivanlou showed that the bone morphogenetic protein 4 (BMP4) is secreted from the ventral side of the embryo, opposite to the organizer, and diffuses throughout the embryo inducing ectodermal tissues. BMP4 activity is inhibited by Noggin thus inducing ectodermal cells in the area of the organizer to develop into neural fate instead of becoming skin cells (Wilson & Hemmati-Brivanlou, 1995). However, a few years later, a novel model challenging this view was postulated by Peter Nieuwkoop. He observed that ectoderm can differentiate into anterior neural structures (forebrain) in the absence of any external factors, while posterior fates need to be actively induced by caudalization factors. He concluded that the formation of the CNS is inherent and the default differentiation path (Nieuwkoop, 1973). Even though a large body of growing evidence collected from the last seventy years favors Nieuwkoop's 'default' model, the postulation remains controversial. Currently, the most accepted and prevalent model is that in early development the dorsalizing Nieuwkoop signal (at that time unknown and later attributed to WNT signaling) produced in the dorsal cells of the embryo is required for establishing the Spemann–Mangold organizer.

Similar structures to the Spemann–Mangold organizer have been identified in other vertebrates- namely the Hensen's node in chick (Waddington & Schmidt, 1933) and in rabbit (Viebahn, 2001), the embryonic shield in zebrafish (Shih & Fraser, 1996) and the primitive node in most amniote embryo, discovered first in mouse (Camus & Tam, 1999)- proving that the organizer structure and its role is evolutionarily conserved.

1.2.2. Modern insights: What are the major signaling pathways involved in embryonic neural induction?

We have come a long way since the first description of the model and findings in the last years have emphasized the complexity of the process. After the discovery of Noggin as the first direct neural inducer, two other organizer-specific proteins, Chordin (Sasai et al., 1994) and Follistatin (Hemmati-Brivanlou et al., 1994), were identified as neural inducing. The further identification of these two proteins provided more insight into the molecular mechanisms underlying neural patterning. These three proteins have in

common that they are all antagonists of different branches of the Transforming Growth factor beta (TGFB) signaling pathway, binding with high affinity to the ligands and thus inhibiting the activation of the receptors. While the notion that BMP antagonists are the ones inducing neural patterning remains valid, more recent evidence suggests functional roles of Fibroblast Growth Factor (FGF) and Wingless/Integrated (WNT) signaling pathways.

On one hand, FGF signaling was also identified as a main player inhibiting the TGFB/BMP signaling pathway in the early embryo (Streit et al., 2000), thus potentially having a synergistic effect with the organizer-secreted BMP antagonists. FGF signaling reduces directly the activity of BMP signaling by promoting phosphorylation of the linker domain and degradation of SMAD1 (Pera et al., 2003), and indirectly by inducing the transcription factor ZEB2, which in turn binds to and represses the transcriptional activity of SMADS (Sheng et al., 2003).

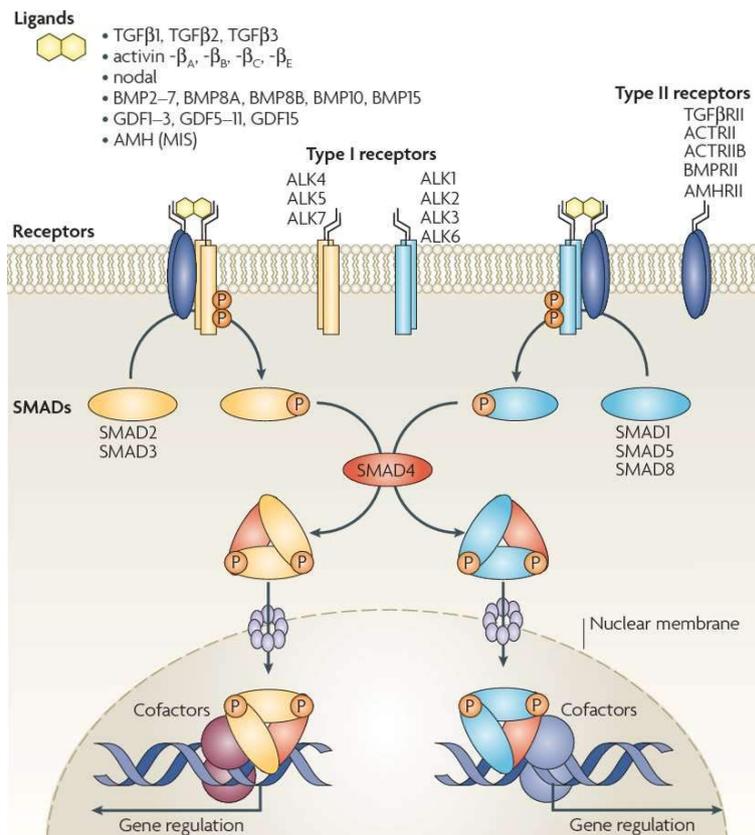


Figure 1.3. The TGFB/BMP signaling pathway. There are several TGFB ligands that can activate the TFGB pathway, including members of BMPs, GDFs, Activins, Nodal, as well as TGFβs. Upon binding of ligands, type II receptors are phosphorylated and activate

type I receptors (ALKs) forming a hetero-tetrameric complex. This leads to the propagation of signaling by at least two pathways involving SMADs (in the canonical pathway). Extracted from (Schmierer & Hill, 2007).

On the other hand, the canonical WNT pathway involvement in neural induction remains controversial. While in amphibian eggs WNT signaling has been shown to activate neural fates (Baker et al., 1999), in chick embryos it seems that WNT inhibition cooperates with FGF signaling in inducing early neural patterning (Wilson et al., 2001). Such disagreement can be explained to a certain extent when addressing the temporal component of such regulation. It appears that at very early stages WNT signaling is required to specify the dorsal fate and to promote the formation of the Spemann–Mangold organizer, becoming the main candidate for being the dorsalizing Nieuwkoop signal (Vonica & Gumbiner, 2007). However, at later stages WNT signaling seems to induce epidermis at expenses of neural fate by regulating the response of ectodermal cells to FGF signaling (Wilson et al., 2001).

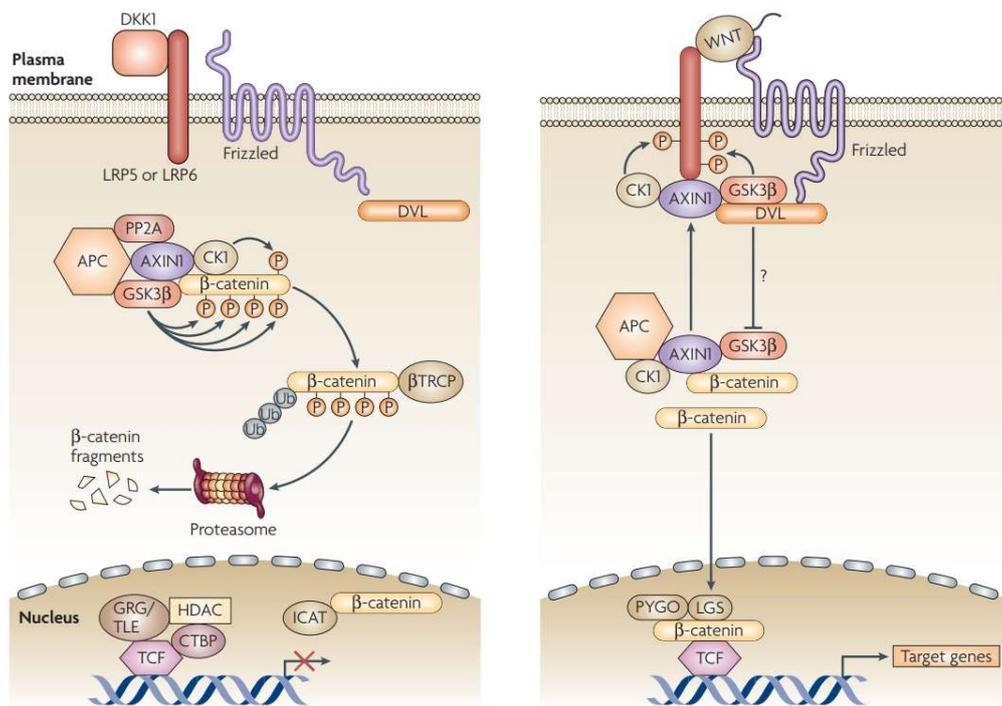


Figure 1.4. The canonical WNT signaling pathway. (Left) In the absence of a WNT ligand, β -catenin is targeted for degradation through the actions of the destruction complex: adenomatous polyposis coli (APC)/axis inhibition protein 1 (AXIN1)/casein kinase 1 (CK1)/glycogen synthase kinase 3 β (GSK3 β). This complex phosphorylates β -catenin creating

recognition sites for β -transducin-repeat-containing protein (β TRCP), leading to its ubiquitylation and proteasomal breakdown. With the absence of β -catenin in the nucleus, TCF assembles a transcriptional repressor complex to silence WNT target genes. (Right) Upon binding of a WNT ligand to the receptor complex, LRP is phosphorylated by CK1/GSK3 β /AXIN1 and recruited to the plasma membrane. Thus, β -catenin is not targeted for degradation and can translocate to the nucleus to form an active transcription factor complex with TCF, leading to the transcription of a large set of target genes. Extracted from (Staal et al., 2008).

In summary, intricate spatiotemporal modulation and interplay of WNT, TGFB/BMP and FGF signaling pathways is required for early neural induction and patterning of the AP identity. Additionally, Sonic hedgehog (SHH) plays a key role in establishing the DV axis. In early development, SHH is secreted from the ventral regions of the neural tube- the notochord and floor plate- generating a signaling gradient that is essential for proper DV patterning of the CNS, specifically the development of the ventral forebrain, midbrain and hindbrain (Ruiz i Altaba et al., 2002).

1.3. Vertebrate telencephalic development

Once the neural plate is established, early in its development the most anterior primordial sheet of cells give rise to the prosencephalon, which is subsequently subdivided into the telencephalon and the diencephalon, as above mentioned.

The telencephalon arises from the most anterior rostral end of the prosencephalon and it is further subdivided into two distinct regions across the DV axis: the pallium, the dorsal region that will primarily become the cerebral cortex; and the subpallium, the ventral region that, in turn, will primarily become the basal ganglia. In mammals, the pallium is further comprised by four major domains that form cortical structures: dorsal, medial, lateral and ventral pallium. The dorsal pallium will give rise to the neocortex (isocortex), while the medial pallium develops into the cortical hem that will give rise to the hippocampus (archicortex). In turn, the lateral pallium develops into the olfactory and some limbic areas (paleocortex), while the ventral pallium also gives rise olfactory areas (Medina & Abellan, 2009).

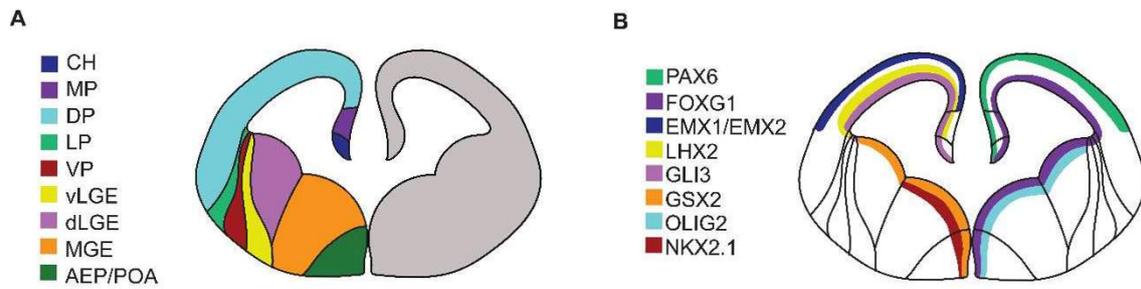


Figure 1.5. Embryonic vertebrate telencephalon. Schematic of a coronal section of the telencephalon highlighting the main regions (A) and key genes demarcating those regions (B). CH: cortical hem, MP: medial pallium, DP: dorsal pallium, LP: lateral pallium, VP: ventral pallium, vLGE: ventral lateral ganglionic eminence, dLGE: dorsal lateral ganglionic eminence, MGE: medial ganglionic eminence, AEP/POA: anterior entopeduncular/preoptic area.

1.4. Telencephalic patterning: progressive molecular and cellular specification

Early patterning of the developing telencephalon is orchestrated by an intricate interplay of morphogenetic gradients of growth factors working in concert with multiple components, including cell–cell interactions, to regulate regional identities. Mainly, SHH is produced ventrally, FGF8 is produced rostrally (most prominently in the cortex) and several BMP and WNT proteins are mainly produced caudo-medially (Charron & Tessier-Lavigne, 2005). These signaling molecules activate a spatially specific signaling code that activates a pattern of transcription factors (TFs) that control many aspects of the subsequent development. In turn, these TFs can also modulate the secretion of morphogens creating regulatory loops to instruct lineage commitment, thus establishing specific telencephalic cell types from a common primordium (Lee et al., 2014).

1.4.1. Patterning along the AP axis

FOXG1 (also known as BF1) (Tao & Lai, 1992), PAX6 (Haubst et al., 2004; Holm et al., 2007) and GLI3 (Aoto et al., 2002; Wang et al., 2014) are among the earliest TFs identified to be expressed in the neural plate cells destined to become telencephalon.

The telencephalic primordium is first characterized by the expression of FOXG1. At this stage the telencephalon is still a single-layered neuroepithelium, and subsequent to FOXG1 expression it becomes further subdivided into its several distinct regions (Hettige et al., 2022). It has been shown that FOXG1 acts in concert with FGF signaling and that

it is required for FGF8 expression. Conversely, studies also suggest that FGF8 induces and maintains FOXP1 expression in the anterior neural plate (Shimamura & Rubenstein, 1997). Moreover, FOXP1 then restricts expression of BMP4 to the telencephalic midline (Ohkubo et al., 2002). FOXP1 has also been shown to exert control over telencephalic progenitors to induce proliferation by cell autonomous mechanisms that include the regulation of PAX6 (Manuel et al., 2011). Null mutation of FOXP1 have been reported to cause hypoplasia of the mouse telencephalon and loss of ventral telencephalic structures (Martynoga et al., 2005). Additionally, lack of PAX6 in mice leads to the caudalization of the cortex through loss of anterior regions. And it has also been described that loss of both PAX6 and EMX2 in mice (another specific telencephalic dorsal marker) results in a drastic reduction of cortical structures in favor of subpallium structures (Muzio et al., 2002).

Further anteriorization of the telencephalic neuroepithelium is characterized by the expression of SIX3 and OTX2 (Acampora et al., 1999). On one hand, SIX3 shows a defined spatiotemporal expression pattern in the developing telencephalon, and together with PAX6 has a role in mediating the early regional subdivision of the prospective prosencephalon into the telencephalon and diencephalon (Appolloni et al., 2008; Ypsilanti & Rubenstein, 2016). On the other hand, OTX2 together with GBX2 are among the earliest genes expressed in the neuroectoderm, and have been suggested to determine the midbrain-hindbrain boundary in vertebrates, dividing the anterior versus posterior domains (Crossley et al., 2001). OTX2 plays essential roles in rostral brain development being required for the development of the forebrain and midbrain, and is counteracted by the effects of GBX2 which is necessary for anterior hindbrain development. *Otx2*-null mice lack forebrain and midbrain regions due to a defective anterior neuroectoderm specification during gastrulation (Acampora et al., 1995). Conversely, mice lacking *Gbx2* show developmental failure of the hindbrain development and display a caudal expansion of the midbrain (Inoue et al., 2012).

1.4.2. Patterning along the DV axis

DV patterning of the telencephalon is established early in forebrain development. The initial subdivision that defines regional identity is regulated by SHH signaling which is

essential for ventral patterning and by the dorsalizing effects of GLI3 expression. GLI3 is initially expressed broadly throughout the telencephalic primordium and then is progressively downregulated in the ventral telencephalon (Gunhaga et al., 2003). In Gli3-deficient mice, they found that the development of the dorsal telencephalon is completely disrupted resulting in a compromised neocortex and the failed formation of the choroid plexus, the cortical hem and the hippocampus (Grove et al., 1998). In turn, SHH promotes ventral cell fates in the forebrain by antagonizing the dorsalizing effects of GLI3 (Hebert & Fishell, 2008).

Another crucial gene for ventral specification is the TF NKX2-1, which defines and delineates MGE from LGE progenitors (Butt et al., 2008). Mice lacking NKX2-1 display a ventral to dorsal change of fate within the basal telencephalon, with precursor cells generating LGE instead of MGE. In turn, GSX2 (*Gsh2* in mouse) accompanies the emergence LGE with a lower expression level in MGE. It has been demonstrated that GSX2 is a downstream target of SHH and that its function is required to repress pallial fates. Lack of GSX2 in mice results in profound defects in telencephalic development (Corbin et al., 2000). In the absence of *GSX2*, it has been observed a loss of DV regionalization in mice, shown by the expansion of MGE (Sussel et al., 1999).

In the dorsal telencephalon, PAX6 also plays an essential role in creating the sharp border between the pallium (dorsal) and subpallium (ventral), mainly being expressed in the prospective neocortex. In *Pax6-null* mouse embryos, there is a disruption of the pallium-subpallium boundary leading to patterning defects that include an expansion of the dorsal LGE at the expense of the ventral pallium. Cells normally restricted to the subpallial side might migrate across the boundary retaining their initial identity, thereby blurring and shifting the border. Dorsal expansion of subpallial markers such as GSX2 coinciding with a downregulation of pallial marker genes can also be observed (Georgala et al., 2011). Concomitantly to these roles, FOXG1 also plays a key role in inducing ventral subpallial identity as a reported downstream effector of SHH. It has been shown that FOXG1 inhibits WNT signaling through direct transcriptional repression of WNT ligands, thus restricting dorsal WNT signaling and delimiting pallial identities (Danesin et al., 2009).

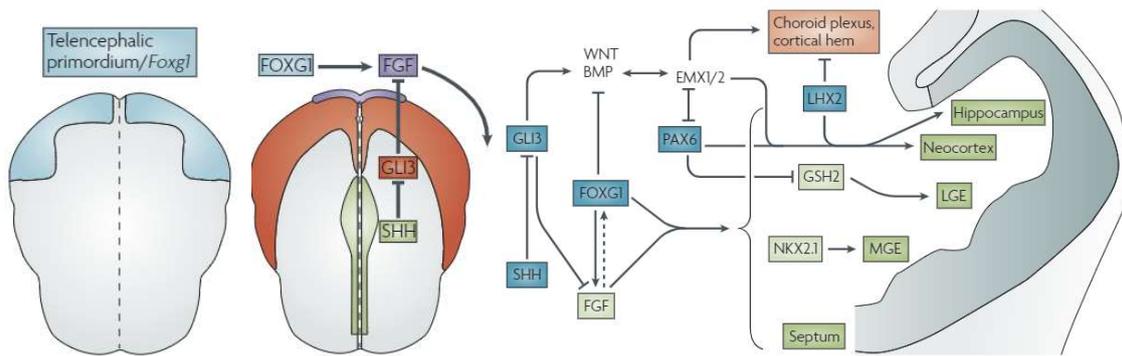


Figure 1.6. Early telencephalic patterning. (left) Schematics of the anterior neural plate in the mouse embryo showing the earliest factors to drive telencephalic specification. (right) Gene regulatory network highlighting key extrinsic and intrinsic cellular factors that establish the dorsal and ventral subdivisions of the telencephalon. Extracted from (Hebert & Fishell, 2008).

1.5. The cerebral cortex

The cerebral cortex derives from the dorsal pallium of the telencephalon. The cerebral cortex is the outer-most layer of grey matter that completely covers the surface of the two cerebral hemispheres. It is the brain region responsible for many of the high-level cognitive functions in humans, including language, perception, reasoning, decision making and consciousness (Adesnik & Naka, 2018).

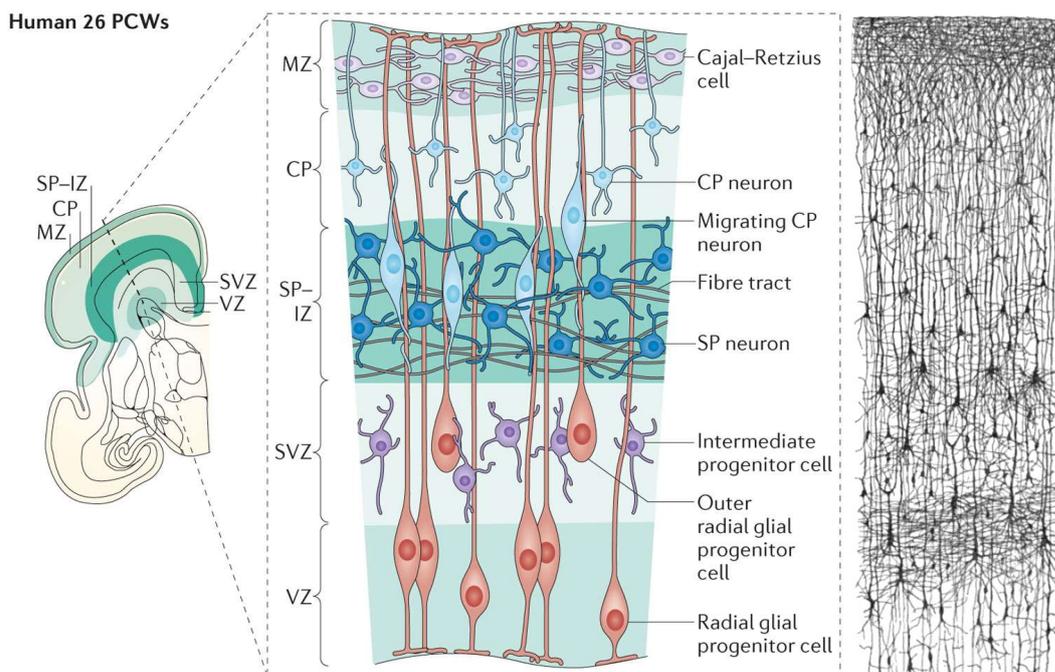


Figure 1.7. The developing human cortex. (on the right) Schematic of a human cortex during at 26 post-conception weeks (PCWs) showing the main compartments and cell types during development. The germinal zone consists of the ventricular and subventricular zones which contain the progenitor cells. The subplate (SP) and the intermediate zone (IZ) lie between the SVZ and the cortical plate (CP). The outermost layer is the marginal zone (MZ). Extracted from (Hoerder-Suabedissen & Molnar, 2015). (on the left) Drawing of a coronal section of the developing cortex, around the same stage, showing the cortical lamination done by Santiago Ramón y Cajal. Extracted from (Ramón y Cajal, 1995).

Cortical development is characterized by three main steps: 1) proliferation and differentiation of the neuronal stem cells (NSCs), 2) migration of neuronal precursors toward the cortical plate, and (3) neuronal population of the cortical plate into six layers. The process starts with the stem-like neuroepithelial cell population residing in the ventricular zone (VZ), which is the dorsolateral wall of the rostral neural tube. VZ progenitors undergo symmetric divisions to expand the pool of cortical progenitors. Then, neuroepithelial cells subsequently differentiate into neurogenic radial glia (RG), which undergo asymmetric divisions to generate the neurons of the cortical plate (CP), as well as generating intermediate progenitors (IPs). They also generate a distinct progenitor population known as the basal progenitors (BPs). BPs leave the VZ and aggregate with one another to establish an additional proliferative layer, the subventricular zone (SVZ), where they generate more neurons (Rakic, 1972). These processes are simultaneously occurring and dynamic. Any perturbation in one of these developmental events can cause a wide range of cortical malformations and diseases, such as microcephaly due to impaired proliferation capacity of cortical progenitors, or lissencephaly due to defective neuronal migration (Juric-Sekhar & Hevner, 2019). Abnormal distribution of cortical neurons can also derive into various psychiatric disorders and brain abnormalities, including epilepsy, autism, bipolar disorder, schizophrenia and attention deficit/hyperactivity disorder (Bush, 2011; Hoerder-Suabedissen & Molnar, 2015).

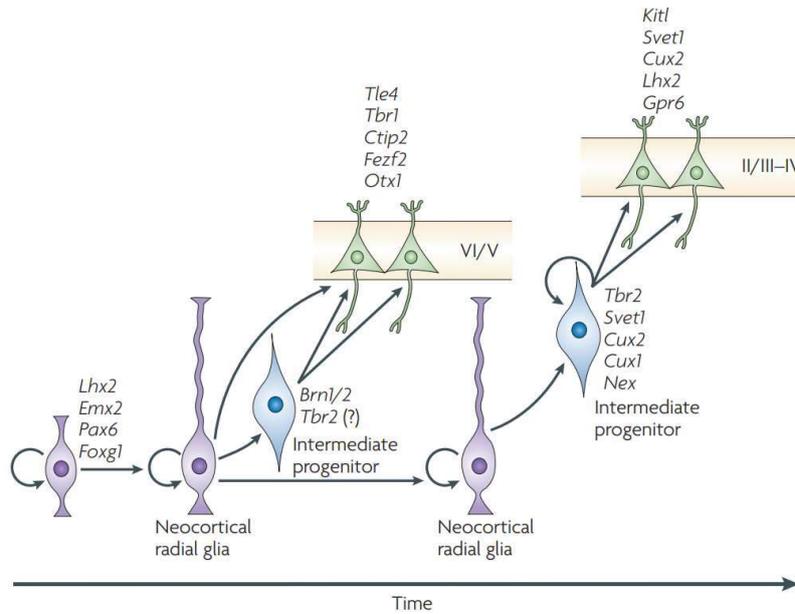


Figure 1.8. Cortical radial glia potential. Schematic representation of cortical progenitors generating the diverse neuronal cell types during the development of the cortex. Progenitors residing in the VZ (RG) and SVZ (IP and BP) produce the neurons of the different neocortical layers in a tightly controlled temporal order. Postmitotic neurons migrate radially and tangentially to position themselves, generating first the deep layers and consequently the upper layers of the cortex. Shown also specific gene markers that can be used for the identification of each cell type, from the radial glia to the layer-specific neurons. Extracted from (Molyneaux et al., 2007).

1.5.1. Composition of the cortical layers

Already in 1878, Bevan Lewis suggested a stratification plan for the cerebral cortex which remains accepted. Based on this structural plan, the cerebral cortex has been divided into: the isocortex and the allocortex. The isocortex or neocortex corresponds to the six stratified layers, while the allocortex, represented by the archicortex (hippocampus) and paleocortex (olfactory cortex), exhibit a laminar structure composed of 3 layers (Triarhou, 2021). It was at the beginning of the 20th century that many studies were undertaken to exclusively analyze the cytoarchitecture of the cortical areas, being Ramón y Cajal the one who described the intrinsic organization of the cerebral cortex in human and vertebrates.

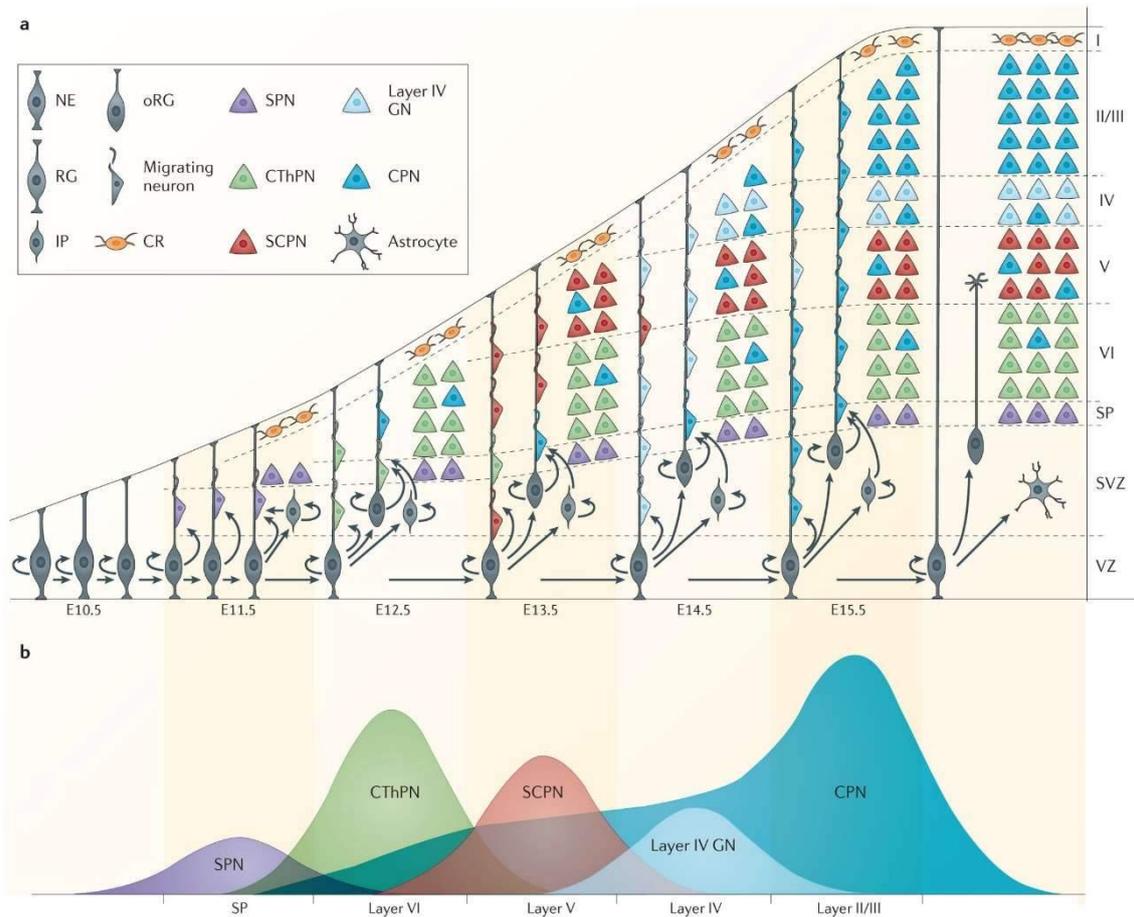


Figure 1.9. Cortical neurogenesis. Schematic depicting the sequential generation of neuronal subtypes and their migration to populate the distinct cortical layers in mice. There are two major waves of neuronal generation layers that separate the formation of deep layers (mostly between E12.5 and E13) and upper layers (between E14-E17) of the cortex. Legend: outer RG (oRG), Cajal–Retzius cells (CR), subplate neurons (SPN), corticothalamic projection neurons (CThPN), subcerebral projection neurons (SCP), granular neurons (GN), callosal projection neurons (CPN). Extracted from (Greig et al., 2013).

The six layers of neocortex are: layer I or plexiform layer -also known as molecular layer or marginal zone (MZ)- containing mainly nerve axons and a few scattered Cajal-Retzius cells; layer II or external granular layer composed of a varying density stellate (granular) cells and pyramidal cells; layer III or external pyramidal layer containing predominantly pyramidal cells of varying sizes; layer IV or internal granular layer consisting mostly of the stellate cells and a smaller portion of the pyramidal cells; layer V or internal pyramidal layer containing mainly medium-sized to large pyramidal cells; and layer VI

or fusiform/multiform layer composed by different types of neuron types, mostly fusiform cells with less dominant pyramidal cells and interneurons.

The deep layers (layer V/VI) are the first ones to be generated and to achieve their stratification and functional maturation of their neurons. Neuronal maturation proceeds successively from the deeper layers to the upper layers (layers II/III/IV), from the oldest generated neurons to most superficial and recently born neurons. Neurons mature into two main subtypes: pyramidal or non-pyramidal neurons. Neurons that retain their original contact with layer I become pyramidal neurons (which represent 70% the of cortical neurons), while neurons that lose that contact become stellate cells, non-pyramidal neurons or interneurons (Cadwell et al., 2019).

.

1.6. How can we study cortical development?

1.6.1. Stem cell models of cortical development: derivation of neural rosettes

Animal models have been the prevailing approach for studying developmental biology due to the high conservation of processes and the accessibility and manipulation ability. Nevertheless, it is well known that discrepancies in the nervous system development between species exist. For example, when comparing mouse and human development some differences can be identified (Copp et al., 2013), especially in regards of the cortical development and the diversity of neuronal cell types residing in the cortex (Cheung et al., 2007; Loomba et al., 2022). At the same time, efforts are also being made in order to reduce and eventually abolish animal experimentation. Such need promoted the extraordinary development of stem cell-based techniques aimed at inducing and culturing neural lineages to reproduce human brain development *in vitro*. Current methods for human neural differentiation protocols range from more homogenous two-dimensional (2D) systems to more complex three-dimensional (3D) systems, each one tailored to specific needs. However, all these protocols follow the same approach in applying the mechanisms revealed by classical embryological studies, and induce neural differentiation by means of multiple cytokines and growth factors.

In 2001, the first derivation of human neural rosettes from ESCs was done by generating embryoid bodies (EBs) – 3D aggregates of cells- that could be directed to differentiate into neural lineages. This resulted in a cluster of NE cells that self-organize to form the so-called neural rosettes (Zhang et al., 2001). Rosettes structures are characterized by their unique cytoarchitecture where cells are radially organized and apically constrained. Thus, forming an apical lumen resembling the VZ seen during embryonic development of the neural tube. Compared to other 2D systems the generation of neural rosettes better recapitulate the *in vivo* properties of cortical RG cells, as they generate intermediate populations and even a rough layering of progenitor zones similar to the *in vivo* counterparts, VZ and SVZ (Edri et al., 2015; Shi, Kirwan, Smith, et al., 2012). Given that the formation of neural rosettes is a critical morphogenic event that reflects the induction of cortical lineages during development it is widely used a readout for proper generation of hiPSC-derived cortical RG cells *in vitro*.

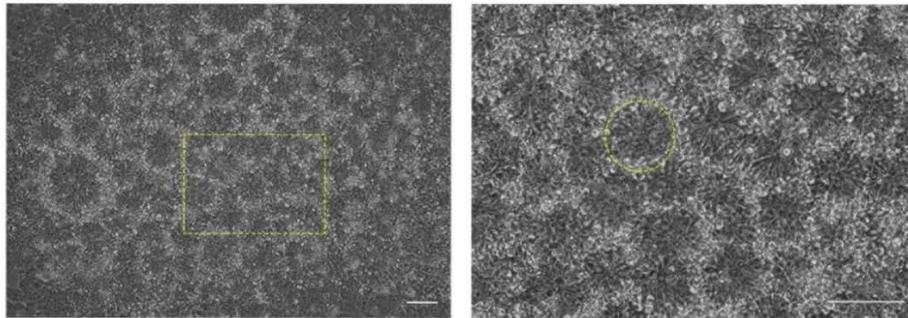


Figure 1.10. Rosette formation as a readout to measure cortical specification. Bright-field images of our own cells at day 12 of neural induction. On the right the magnified image highlighting a neural rosette (yellow circle). This very distinctive cell morphology resulting from the apico-basal polarity properties of cortical RG cells allows for the morphological identification of proper cortical induction from hiPSCs. Scale bar= 50.

Furthermore, it was later shown that these neural rosette structures corresponding to early anterior NSCs can be induced by using the BMP antagonist Noggin. More importantly, they can be propagated and expanded in culture under specific conditions thus allowing functional characterization of the NSCs and their progeny (Elkabetz et al.,

2008). Additional to EB formation capacity, it was shown that pluripotent stem cells (PSCs) in adherent monocultures are able to commit efficiently to a neural fate in the absence of serum or growth factors due to autocrine signaling (Ying et al., 2003). In 2005, pioneering work from Sasai's group demonstrated that combining EBs culture and serum-free conditions (SFEB), with the addition of Nodal and WNT pathway antagonists, could generate broad telencephalic neural precursors from PSCs when plated on coated dishes (Watanabe & Nakamura, 2012). Later, this method was significantly improved by adding TGFB inhibition to BMP inhibition in feeder-free cultures of hPSCs which rapidly differentiated them into early neuroectoderm- becoming the commonly used dual SMAD inhibition (Dual SMAD-i) protocol (Chambers et al., 2009).

A plethora of studies aimed to the optimization of neural induction protocols followed this seminal work by trying different combinations of TFGB, BMP, WNT pathway inhibition, with or without further FGF and SHH pathway modulations (Gaspard et al., 2008; Maroof et al., 2013). Other studies combined WNT, FGF and NOTCH inhibition to induce rapid production of early cortical neurons (Qi et al., 2017). Altogether, different strategies were developed to optimize the generation of specific telencephalic precursors in order to generate specific cortical cell types of interest.

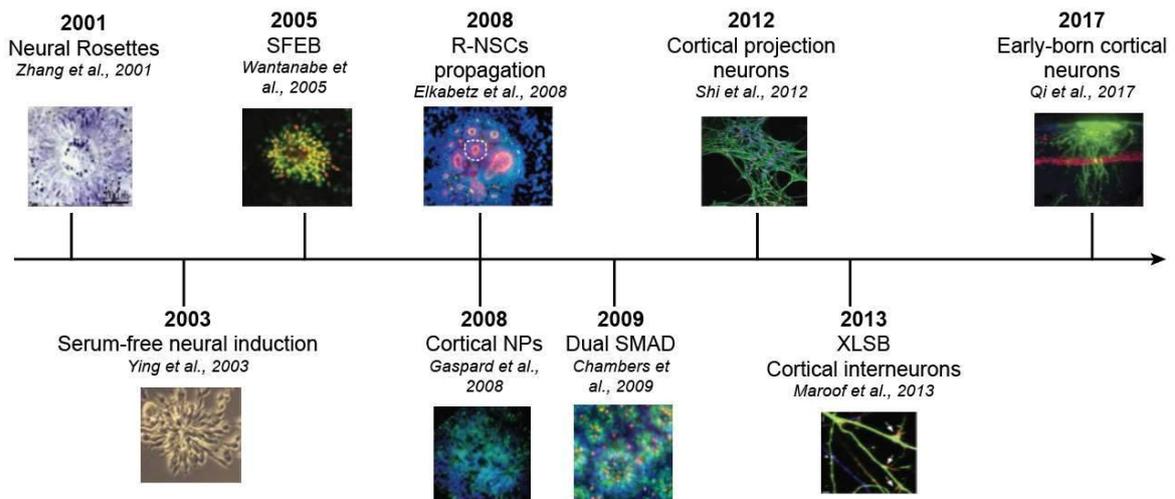


Figure 1.11. Seminal neural induction protocols. Timeline showing the generation and development of different neural induction protocols for the generation of neural/cortical tissue. Adapted and modified from (Kelava & Lancaster, 2016).

More recently and concomitant with the latter 2D differentiation protocols, 3D systems for recapitulating corticogenesis were developed giving rise to the cerebral organoid models (Eiraku et al., 2008; Lancaster & Knoblich, 2014). Nevertheless, methods for generating cortical organoids are also highly variable which results in differential neural patterning trajectories and high heterogeneity. This lack of standardization in the field emphasizes the need to keep on optimizing differentiation protocols to achieve a starting pure population of NCSs ('the founder' NSC population) that would give rise to a more homogenous cortical population.

To overcome such hindrance our lab has established a method to derive homogenous early cortical progenitors from PSCs. In the past years, our lab established a streamlined method known as Triple-i paradigm- combination of TFGB, BMP, WNT pathway inhibition- to derive homogenous starting cortical progenitors both in neural rosettes and cerebral organoids platforms (Rosebrock et al., 2022).

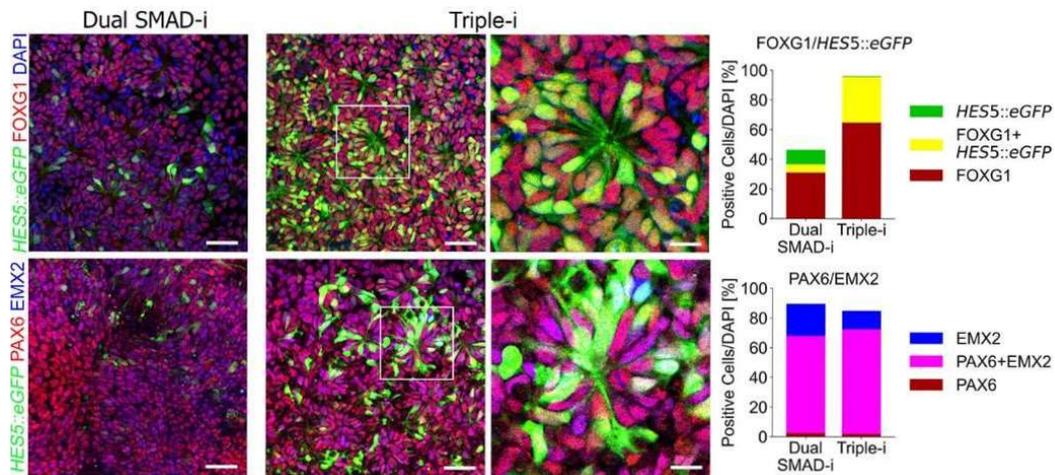


Figure 1.12. Triple-i protocol to derive neural lineages from stem cells promotes cortical fates. In a 2D monolayer differentiation setting, the protocol established in our lab, Triple-i,, is compared with one of the most used to derive cortical fates, Dual SMAD-i. (on the left) Immunostaining of cortical markers FOXG1 (top) and PAX6 and EMX2 (bottom) with respect to Notch activation (*HES5::eGFP*) together with (on the right) cell counts and co-localization analysis of markers. Differences between derivation methods are shown

highlighting an enrichment of cortical lineages in Triple-i as compared with Dual SMAD-i
Scale bar= 50 μ m. Extracted from (Rosebrock et al., 2022).

1.6.2. Derivation and characterization of human cortical NSCs *in vitro*

Already in 2015, our lab was able to derive and dissect the human cortical differentiation process *in vitro*, going from neuroepithelial cells towards distinct RG cell types, recapitulating *in vivo* development (Edri et al., 2015). This was achieved by prospectively isolating consecutively appearing PSC-derived progenitors based on their NOTCH activation state by using a reporter cell line tagging a downstream effector of the NOTCH signaling pathway (HES5::eGFP line). The stepwise isolation allowed for the dissection of the dynamic changes that lead to heterogeneity in cortical NSCs during long-term culture. Five distinct progenitor states were defined based on their cell morphology and NOTCH activation state: neuroepithelial cells (NE), early radial glial cells (E-RG), mid radial glial cells (M-RG), late radial glial cells (L-RG), and long-term cultured progenitors (LNP). Beyond being able to derive the *in vivo* counterparts of the diverse cortical NSCs, this study provides evidence of the broad heterogeneity found in these progenitor populations highlighting variable proliferation capacity and differentiation potential. The NE stage corresponds to the first NSC population which begin dividing symmetrically to expand the pool of progenitors and, consequently, become the E-RG population. E-RG rosettes contain highly proliferative NSCs exhibiting broad differentiation potential, and are able to divide asymmetrically to give rise to IP and deep layer neurons. As differentiation progresses, M-RG rosettes emerge becoming restricted from generating earlier fates, and will give rise to upper layer neurons. Generally, there is a decrease in stemness and an increase tendency to differentiate into neurons. As development progresses in culture, rosettes dismantle losing their characteristic cytoarchitecture, becoming L-RG which switch from neurogenic potential to astrogenic potential. Finally, the LNP stage is achieved where cells are no longer capable of forming rosette structures, reflecting loss of epithelial integrity due to accumulation of basal progenitors, neurons and cells with astroglial character. This indicates that LNP cells progress beyond RG fates towards adult-like progenitor identity.

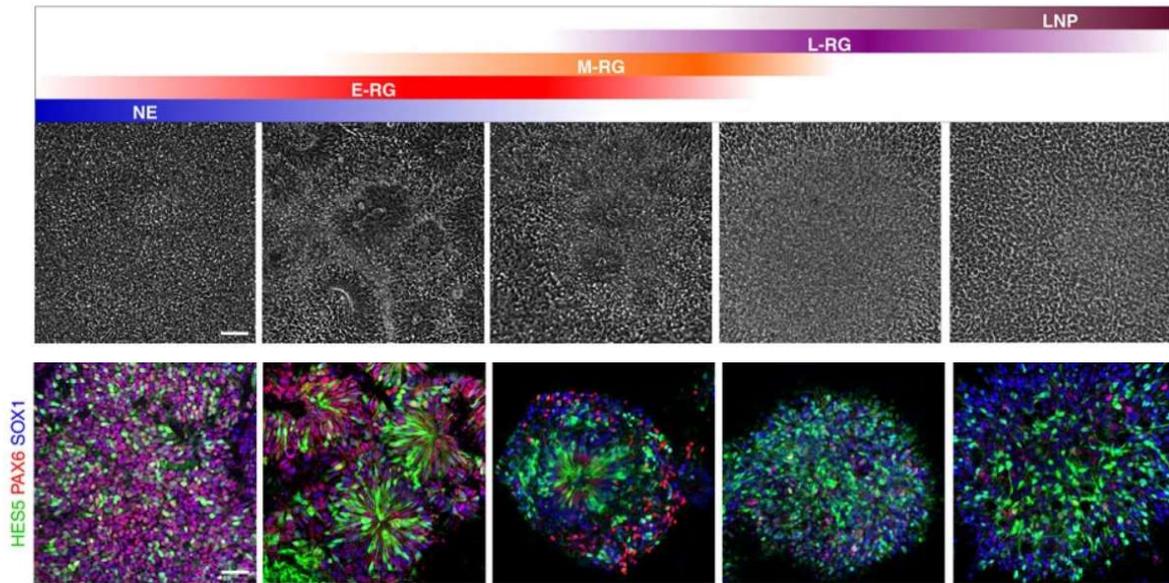


Figure 1.13. Progression of NSC states throughout cortical differentiation *in vitro*. (top) Bright-field images of progenitor cells during long-term differentiation shows dynamic morphological features. Naming conventions representing neuroepithelial (NE), early radial glial (E-RG), mid radial glial (M-RG), late radial glial (L-RG) and long-term cultured progenitors (LNP) are indicated (bottom) Immunostainings of progenitor cell throughout the progression period. Top: PAX6, SOX1 and HES5 reporter induction during early stages. Note that on E-RG rosettes there is high expression of PAX6 and HES5, marking neurogenic RG cells, and fitting the dorsal cortex molecular identity of E-RG cells, coinciding with PAX6 and HES5. Scale bar= 50 μ m. Extracted from (Edri et al., 2015).

Global transcriptional analysis and epigenetic characterization of the PSC-derived progenitor states revealed stage-specific molecular signatures and coordinated epigenetic changes that seem to be responsible for driving the transition through the distinct NSC competences (Ziller et al., 2015). However, there is still the need to unravel the underlying regulatory mechanisms driving the transition to better understand how different types of NSCs emerge during cortical development. Furthermore, such knowledge would allow us to manipulate these NSC stages. For example, it would allow to control the progression of these NSC populations thus generating an unlimited *in vitro* culture of the desired NSC type that could potentially be used for therapeutic approaches.

1.6.3. The emergence of induced pluripotent stem cells

In 2006, Yamanaka's group reported for the first time that the introduction of four transcription factors- Oct4, Sox2, c-Myc, and Klf4- were sufficient to reprogram mouse fibroblasts into pluripotent stem cells, termed induced pluripotent stem cells (iPSCs). Only one year later, human iPSCs (hiPSCs) were already being developed by two independent research groups. Yamanaka's group successfully differentiated human fibroblasts into iPSCs through the transduction of the same 4 transcription factors previously used in mice cells (Oct4, Sox2, Klf4 and c-Myc) by means of a retroviral system (Takahashi et al., 2007). While, concomitantly, Thomson's group used a different set of transcription factors- Oct4, Sox2, Nanog and Lin28- that were transduced by means of a lentiviral system (Yu et al., 2007).

Since then, human iPSC technology has hold great promise for regenerative therapies because patient-specific iPSCs can be derived from patient-somatic cells, such as fibroblasts, providing unprecedented human models for studying neurodevelopmental diseases and for personalized therapy. On one hand, this opens the possibility to study disease pathology in different and patient-distinct genetic backgrounds and their response to drugs (Costamagna et al., 2021; Sabitha et al., 2021). On the other hand, combining advancements in cortical differentiation methods with the iPSCs technology to derive cortical neurons from patient-somatic cells could potentially provide a source for isogenic cortical cells that can be used for transplantation, thus avoiding immunogenicity responses. Even though this is not yet a possibility, efforts are being made to promote such avenues for neurodevelopmental disorders that are caused by cortical abnormalities such as autism spectrum disorder (Nestor et al., 2016).

1.7. Single-cell RNA sequencing technology: unprecedented insight into human embryonic development

Recently, another great scientific advancement has been the development of the Single-cell RNA sequencing (scRNA-seq) technology. scRNA-seq has become the state-of-the-art approach for unfolding the heterogeneity and cellular complexity of human tissues at unprecedented single cell resolution. Since its first discovery in 2009 (Tang et al., 2009),

studies based on scRNA-seq are extensively used due to their power of overcoming many drawbacks arising from other sequencing technologies such as bulk RNA-seq. Mainly, the ability to comprehensively identify novel cell types and rare cell subsets avoiding the averaging phenomenon inherent to bulk analysis. We have come a long way since the first studies that covered the analysis of 10 to 100 cells (Kumar et al., 2014; Shalek et al., 2013), and advances have enabled us to profile gene expression in individual cells on a large scale - up to tens of thousands of individual cells (Klein et al., 2015; Macosko et al., 2015)- and it continues to grow. scRNA-seq has become a well-established technique and various methods have been developed with different single cell isolation and library preparation strategies. Currently, droplet-based methods, such as Drop-seq (*Figure 1.14*) or the more recently developed 10× Genomics platform (*for more details see methods section 3.13*), are considered the gold standard because they generally yield higher throughput while reducing cost and workload.

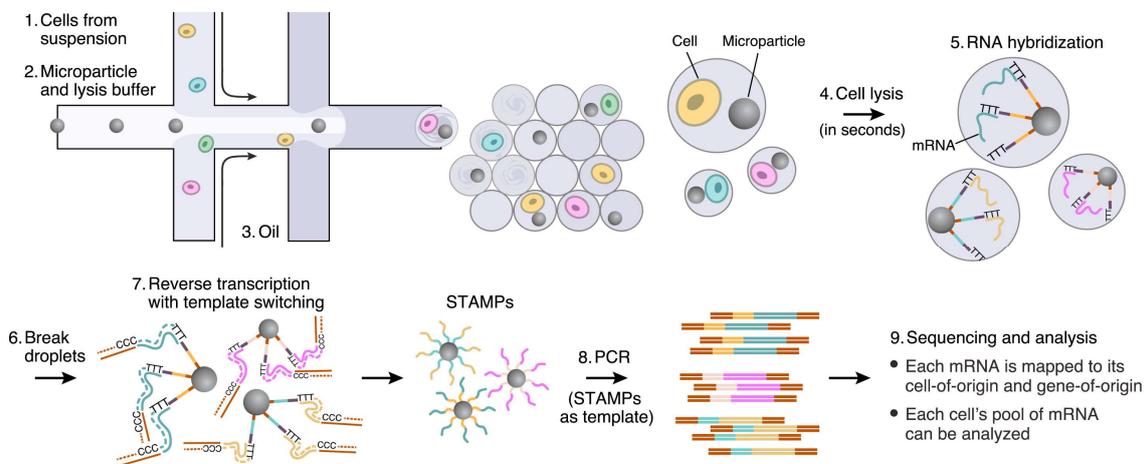


Figure 1.14. Droplet-based single-cell sequencing. Schematic workflow for library preparation for Drop-seq. The microfluidic device joins two aqueous flows before their compartmentalization into discrete droplets: one flow containing cells and the other containing barcoded primer beads. Immediately after droplet formation, the cell is lysed and releases its mRNAs which hybridize to the primers on the microparticle surface. The droplets are broken and the mRNAs are then reverse-transcribed in bulk, forming single-cell transcriptomes attached to microparticles (STAMPs). Finally, template switching is used to introduce a PCR handle downstream of the synthesized cDNA, which is then amplified and ready for library preparation and sequencing. Extracted from (Macosko et al., 2015).

Since scRNA-seq enables the study of biological properties of individual cells allowing to unravel cellular complexity of thousands of cells at once, it seems logical to want to apply such technology to the study the intricate cellular complexity of the human brain. Hence, plenty of studies in the recent years have made use of the single cell technology to study both fetal and adult brain in order to capture and decipher their cellular complexity (Feng et al., 2021; Hedlund & Deng, 2018).

Initial studies focused on exploring cell type identities in the developing brain by profiling between 100 to 500 cells, resulting in the identification of diverse neural cell types including RG, newborn and mature neurons, astrocytes and oligodendrocytes, among others (Darmanis et al., 2015; Pollen et al., 2014). More recent scRNA-seq studies are attempting to obtain a more comprehensive cell typing, especially in the cerebral cortex which comprises a vast diversity of neuronal cell types (Chang et al., 2020). Additionally, scRNA-seq is also used to profile hiPSC-derived cortical progenitors and cerebral organoids to study the viability of derivation protocols. Also allowing to shed light into the developmental process of neural differentiation and the molecular mechanisms governing cell fate specification by using a more controlled model system (Bhaduri et al., 2020; Velasco et al., 2019).

2. Main Aim

The motivation behind this project is to better understand the ontogeny of the cortical NSCs that sequentially appear during cortical development. This is of great importance since characterizing these distinct NSC populations is key for developing an *in vitro* system that allows for the homogeneous and unlimited culture of the desired NSC type, which is critical for potential cell replacement therapies.

In order to address this question, the main aim of my project is to develop a strategy to isolate the early cortical NSC population for its characterization and potential manipulation. To achieve this, the experimental pipeline is focused on identifying a cell surface marker to enable a cell-sorting approach for the isolation of these cells *in vitro*.

Main objectives

- 1- Identification of a candidate surface marker for the isolation of early cortical NSCs
- 2- Validation of the candidate surface marker
- 3- Molecular and cellular characterization of the sorted populations

3. Materials and Methods

3.1. Reagents

- ❖ 2-Mercaptoethanol (Gibco™, cat. no.: 31350010)
- ❖ Accutase® solution (Sigma-Aldrich®, cat. no.: A6964)
- ❖ Apo-transferin human (Sigma-Aldrich®, cat. no.: T1147)
- ❖ B-27™ serum free (Gibco™, cat. no.: 17504044)
- ❖ B-27™ Supplement minus vitamin A (Gibco™, cat. no.: 12587010)
- ❖ Bovine Serum Albumin (Sigma-Aldrich®, cat. no.: A9418)
- ❖ D-(+)-Glucose (Sigma-Aldrich®, cat. no.: G8270)
- ❖ Dimethyl Sulphoxide (DMSO, Sigma-Aldrich®, cat. no.: D2650)
- ❖ DMEM/F-12 (Gibco™, cat. no.: 11320033)
- ❖ DMEM/F-12 (powder, Gibco™, cat. no.: 32500035)
- ❖ DMEM/F-12, HEPES (Gibco™, cat. no.: 31330038)
- ❖ Dimethylsulfoxid, DMSO (Sigma-Aldrich®, cat. no.: D2650)
- ❖ DNase I (STEMCELL technologies, cat.no.: 07469)
- ❖ DPBS, no calcium, no magnesium (Gibco™, cat. no.: 14190169)
- ❖ Fetal Bovine Serum, qualified, heat inactivated (FBS, Gibco™, cat. no.: 16140071)
- ❖ GlutaMAX™ Supplement (Gibco™, cat. no.: 35050061)
- ❖ Hank's Balanced Salt Solution (HBSS, Gibco™, cat. no.: 14025092)
- ❖ Insulin from bovine pancreas (Sigma-Aldrich®, cat. no.: I6634)
- ❖ Knockout™ DMEM (Gibco™, cat. no.: 10829018)
- ❖ KnockOut™ Serum Replacement (Gibco™, cat. no.: 10828028)
- ❖ L-Glutamine (Gibco™, cat. no.: 21051024)
- ❖ Matrigel® Membrane Matrix (Corning®, cat. no.: 354234)
- ❖ MEM Non-Essential Amino Acids (Gibco™, cat. no.: 11140050)
- ❖ mTeSR₁™ Basal Medium (Stem Cell™ Technologies, cat. no.: 85850)
- ❖ Neurobasal™ Medium (Gibco™, cat. no.: 21103049)
- ❖ Neutral protease (Dispase, Worthington, cat. no.: LS02100)
- ❖ Optimal Cutting Temperature compound, OCT (Tissue-Tek®, cat. no.: 4583)

- ❖ Paraformaldehyde (Sigma-Aldrich®), cat. no.: 158127)
- ❖ Penicillin-Streptomycin (Gibco™, cat. no.: 15140122)
- ❖ Progesterone (Sigma-Aldrich®), cat. no.: P0130)
- ❖ Putrescine dihydrochloride (Sigma-Aldrich®), cat. no.: P7505)
- ❖ Recombinant Human FGF basic/FGF2/bFGF (146 aa) Protein (R&D Systems, cat. no.: 233-FB)
- ❖ Recombinant Mouse Noggin Fc Chimera Protein, CF (R&D Systems, cat. no.: 719-NG)
- ❖ SB-431542 (Tocris, cat. no.: 1614)
- ❖ Sodium bicarbonate (Sigma-Aldrich®), cat. no.: S5761)
- ❖ Sodium Selenite (Sigma-Aldrich®), cat. no.: 214485)
- ❖ Sucrose (Sigma-Aldrich®), cat. no.: S0389)
- ❖ Triton™ X-100 (Sigma-Aldrich®), cat. no.: X100)
- ❖ Trypan Blue Solution (Gibco™, cat. no.: 15250061)
- ❖ UltraPure™ EDTA, pH 8.0 (Invitrogen™, cat. no.: 15575020)
- ❖ XAV 939 (Tocris, cat. no.: 3748)
- ❖ Y-27632 dihydrochloride (ROCK inhibitor, ROCKi, Tocris, cat. no.: 1254)

3.2. Equipment and instruments

- ❖ Portable Pipet-Aid® XP Pipette Controller (Drummond, cat. no.: 4-000-101)
- ❖ Pipettes (P1000, P200, P10, Eppendorf Research Plus 3-Pack Option 2, cat. no.: 3120000917)
- ❖ Serological pipettes (25ml, 10ml, 5ml, Sarstedt, cat. no.: 86.1685.020, 86.1254.025 and 86.1253.025)
- ❖ Sterile filter tips (1000, 200, 20, 10 ul, Biozym, SafeSeal SurPhob®, cat. no.: VT0270, VT0250, VT0220 and VT0200)
- ❖ 500 mL Vacuum Filter/Storage Bottle System, 0.22 µm Pore 33.2cm² PES Membrane, Sterile (Corning®), cat. no.: 431097)
- ❖ 250 mL Vacuum Filter/Storage Bottle System, 0.22 µm Pore 19.6cm² CN Membrane, Sterile (Corning®), cat. no.: 430756)
- ❖ Screw cap tube, 50 ml, (LxØ): 114 x 28 mm, PP, with print (Sarstedt, cat. no.: 62.547.254)

- ❖ Screw cap tube, 15 ml, (LxØ): 120 x 17 mm, PP, with print (Sarstedt, cat. no.: 62.554.502)
- ❖ 60 mm TC-treated Culture Dish (Corning®), cat. no.: 430166)
- ❖ 96-well Clear Round Bottom Ultra-Low Attachment Microplate, Individually Wrapped, with Lid, Sterile (Corning®), cat. no.: 7007)
- ❖ 24-well Clear Flat Bottom Ultra-Low Attachment Multiple Well Plates, Individually Wrapped, Sterile (Corning®), cat. no.: 3473)
- ❖ 6-well Clear Flat Bottom Ultra-Low Attachment Multiple Well Plates, Individually Wrapped, Sterile (Corning®), cat. no.: 3471)
- ❖ Counting Chamber (Marienfeld, cat. no.: 0610010)
- ❖ Cryotube (Greiner Cryo.s[™]vials, VWR, cat. no.: 122277)
- ❖ Fluid aspiration system BVC control (Vacuubrand, cat. no.: 20727200)
- ❖ Aqualine AL 12 Water bath (LAUDA, cat. no.: 92635)
- ❖ Heracell VIOS CO2 incubators (Thermo Scientific[™], cat. no.: 50145515)
- ❖ Herasafe[™]KS, Class II Biological Safety Cabinet (Thermo Scientific[™], cat. no.: 51022734)
- ❖ Celltron Orbital Shaker (Infors HT, cat. no.: 69455)
- ❖ Inverted microscope (Nikon SMZ1270, cat. no.: MNA52110)
- ❖ 4200 TapeStation System (Agilent, cat. no.: G2991BA)
- ❖ LSM 880 Confocal Laser Scanning Microscope (Zeiss)
- ❖ BD FACSAria[™]Fusion flow cytometer (BD Biosciences)

3.3. Cell culture

3.3.1. hiPSC line

The ZIP13K2 hiPSC line (a gift from Franz-Josef Müller, Zentrum für Integrative Psychiatrie, University Hospital Schleswig-Holstein, Kiel, Germany) was obtained by reprogramming human dermal fibroblast cell line (HDF51) derived from an aborted female fetus as previously described (Tandon et al., 2018).

3.3.2. Cell propagation

For propagation of hiPSC lines, cells were cultured in mTeSR₁ medium (85850, StemCell Technologies) in standard cell culture plates coated with Matrigel Basement Membrane Matrix (354234, BD Biosciences), and grown at 37 °C, 5 % CO₂ in a water vapor saturated atmosphere. Under the same conditions, HEK273T cells were grown in uncoated plates and cultured in self-made HEK medium containing KnockOut DMEM (10829018, Gibco) supplemented with 10% Fetal Bovine Serum (16140071, Gibco), 2mM Glutamine (35050038, Thermo Scientific), 50 µM 2-Mercaptoethanol (31350010, Gibco) and Penicillin/Streptomycin (15140122, Gibco).

3.3.3. Passaging hiPSCs

For passaging hiPSC lines, cells were dissociated by adding EDTA (15575020, Invitrogen) to the culture dish and incubating for 3-4 minutes at 37°C. After incubation, EDTA was removed carefully and the detachment of the colonies was achieved by flushing the cells with mTESR₁ medium. Cells were then resuspended in fresh mTESR₁ medium to achieve the desired density and re-seeded in matrigel-coated dishes. Cells were cultured until they reached c.a. 70-80% confluency.

3.3.4. Freezing hiPSCs

In order to create stocks of the generated cell lines, cells were frozen and stored in liquid nitrogen using the following protocol. The mTESR₁ medium was removed, cells were washed with DPBS and EDTA was added to the cells which were incubated at 37°C for 5 min. Detached cells were washed with DPBS and centrifuged at 14,000 rpm for 3 min. Then, cells were resuspended freezing mix containing 10% DMSO (D2650, Sigma-

Aldrich) in KSR media (description below). The desired number of cells in a total volume of 1 ml were transferred into a Cryotube (Greiner). Cryotubes were then placed in a slow freezing box, where one 1 °C drops per 1 minute which was stored at the -80°C. The next day, all Cryotubes were placed in the liquid nitrogen tank for long-term storage.

3.4. Neural induction media

3.4.1. KSR medium

The KSR medium used for monolayer neural differentiation was made by adding 75 ml of KSR supplement, 5 ml of GlutaMAX, 5 ml of MEM-NEAA, 5 ml of Penicillin-Streptomycin, and 0.5 ml of beta-mercaptoethanol to 409.5 ml of Knockout DMEM (for a total volume of 500 ml). The medium was then filtered by using a vacuum-driven 0.2- μ m filter unit and store at 4°C for up to one month.

3.4.2. N2 medium

The N2 medium used for neural differentiation was made by adding the following components to 490 ml of double distilled water: 6.5 g of DMEM/F-12 powder, 0.775 g of D-Glucose, 1 g of Sodium bicarbonate, 5 mg of Apo-transferrin, 12.5 mg of insulin, 30 μ l of 500 μ M Sodium selenite, 100 μ l of 830 nM Putrescine, 100 μ l of 100 μ M Progesterone and 5 ml of Penicillin-Streptomycin (for a total volume of 500 ml). The medium was stirred at room temperature until all components were dissolved, then filtered by using a vacuum-driven 0.2- μ m filter unit and store at 4°C for up to one month.

3.4.3. Neurobasal medium

The Neurobasal (NB) medium used for neural differentiation was made by adding, 5 ml of GlutaMAX, 5 ml of MEM-NEAA, 5 ml of Penicillin-Streptomycin, and 0.5 ml of beta-mercaptoethanol to 484.5 ml of NB (for a total volume of 500 ml). The medium was then filtered by using a vacuum-driven 0.22- μ m filter unit and store at 4°C for up to one month.

3.5. Neural induction protocols

3.5.1. Monolayer neural differentiation

hiPSC colonies were dissociated by adding EDTA to the culture dish and incubating for 2 minutes at 37°C. After incubation, EDTA was removed carefully Accutase (A6964, Sigma-Aldrich) was added to the culture dish and incubated for 3-4 minutes at 37°C. After incubation, the detachment of the cells was achieved by flushing the cells and resuspending them in fresh mTESR₁ medium containing 10 µM ROCK inhibitor. The single cell suspension was then centrifuged at 270 xg for 5 min at room temperature. After centrifugation, the supernatant was aspirated, and cells were washed once and then resuspended with ½ KSR + ½ N2+NB media containing 10 µM ROCK inhibitor. After counting, 750,000 cells were seeded per each well of a 6-well low-attachment plate in 2 ml of ½ KSR + ¼ N2 + ¼ NB media containing 1% B27 without retinoic acid and 10 µM ROCK inhibitor. Next day- day 1 of the differentiation protocol- EBs were gently scraped to avoid their settlement on the bottom of the dish and to allow them to grow in suspension. Neural induction was initiated on day 2 by changing the medium to ¼ KSR and ¾ N2+NB containing 1% B27 without retinoic acid, 10 µM of ROCK inhibitor, and adding the inhibitor molecules SB-431542 (10 µM), Noggin (250 ng/ml) and XAV939 (3.3 µM). On day 3, EBs are transferred from the low-attachment plates to the 6 cm dishes that had been previously coated with polyornithine (PO) (15 µg/ml), laminin (Lam) (1 µg/ml) and fibronectin (FN) (1 µg/ml). The pre-coated dishes were dried around the edges in order to limit and restrict the surface area to the center of the dish where EBs will be able to attach and grow. Then, EBs were scraped, collected and distributed into the PO/Lam/FN coated-dishes (1:2) to allow their flattening, growth and rosette-formation. On day 7, medium was change with fresh ½ N2 + ½ NB media containing 1% B27 without retinoic acid and SB-431542 (5 µM), Noggin (125 ng/ml) and XAV939 (3.3 µM). After 2 days, on day 9, medium was changed replacing all inhibitors for FGF8 (100 ng/ml) and BDNF (20 ng/ml). On day 12, for propagation of the neural stem cells, rosettes were harvested by picking and re-plated on PO/Lam/FN pre-coated dishes with fresh media (same as on day 9). This procedure was performed every 5-7 days- once confluency was reached- unless cells were collected for downstream analysis on the desired time point. From day 28 onwards, medium was substituted with of ½ N2

+ ½ NB media containing 1% B27 without retinoic acid and with FGF2 (20 ng/ml), EGF (20 ng/ml), and BDNF (20 ng/ml), which was changed every 2-3 days.

3.5.2. Generation of neural organoids

hiPSC colonies were dissociated and processed as mentioned above (see Monolayer neural differentiation section). After centrifugation, the supernatant was aspirated, and cells were washed twice and then resuspended in hESC medium containing 4 ng/ml FGF2 and 50 µM ROCK inhibitor. After counting, 9,000 cells were seeded per each well of a 96-well U-bottom low-attachment plate in a total volume of 150 µl resuspension medium. After 2 days, neural induction was initiated by exchanging 75 µl of the medium for 150 µl of fresh hESC medium containing 4 ng/ml FGF2 and 50 µM ROCK inhibitor, together with the inhibitor molecules SB-431542 (10 µM), Noggin (250 ng/ml) and XAV939 (3.3 µM). On day 4, 150 µl of the medium was exchanged for fresh one containing the same concentration of inhibitor molecules. If the organoid size was above 350 µm in diameter, FGF2 and ROCK inhibitor were not added to the medium. Once organoids achieved a size greater than 400 µm in diameter and showed the expected morphology (around day 6), they were transferred to a 24-well low-attachment plates in 500 µl of N2 medium supplemented with the inhibitor molecules SB-431542 (10 µM), Noggin (250 ng/ml) and XAV939 (3.3 µM). Every 2 days, 300 µl of media was replaced for fresh one. On day 11, organoids were embedded in matrigel drops (30 µl) and transferred into a 6-well low-attachment plates. For each well, 4 organoids were transferred with 2.5 ml of N2 medium containing 1% B27 without retinoic acid. Every 2 days media was changed. On day 15, organoids were transferred to an orbital shaker (at 86 rpm) to allow for better oxygenation, and N2 medium was supplemented with 1% B27 with retinoic acid. From day 50 onwards, for long term organoid differentiation, 1% matrigel was added into the medium every time the media was changed- every 2-3 days.

3.6. Immunofluorescence and confocal microscopy

The immunostaining protocol was done either directly on cells grown on matrigel-coated plastic plates or on cryopreserved organoid sections.

For cultures, cells were fixed for 30 minutes with 4% paraformaldehyde solution. After a couple washes in DPBS, cells were permeabilized for 30 minutes at room temperature in PBST: DPBS containing 0.3% Triton X-100 (9002-93-1, Sigma Aldrich), 1% bovine serum albumin (BSA) (A9647-100G, Sigma Aldrich), 10% Fetal Bovine Serum (FBS) (16140071, Gibco). Followed by blocking for 30 minutes at room temperature in PB (DPBS containing 1% BSA and 10% FBS). Then, cells were incubated overnight at 4°C with the desired primary antibody combination diluted in PB (*listed in table 3.1*). For signal detection, cells were washed three times during 10 minutes in PB followed by a 45 min incubation at room temperature with fluorescent conjugated secondary antibodies (*listed in table 3.2*), and a 5 minutes incubation of DAPI (10 µg/ml, Roche) in PBS for nuclear counterstaining. After washing twice with DPBS, fresh DPBS was added and plates were stored, covered from light, at 4°C until image acquisition.

Organoid samples were taken at the desired date and fixed with 4% paraformaldehyde solution (E15713-S, Science Service) for 30 minutes to 1 hour depending on the size. Following fixation, they were washed twice in PBS and infused in 30% (w/v) sucrose in DPBS for minimum 3 hours (to overnight at 4°C) for cryoprotection. The following day, organoids were submerged in OCT (4583, Tissue-Tek®) for embedding and stored at -80°C until processing. Prior to cryosectioning, the prepared blocks containing the organoids were placed at -20°C to allow the tissue to acclimate, and then were sectioned into 10-µm-thick slices. For the immunostaining process, circles around the organoid sections were traced with the hydrophobic PAP-pen to form a repellent barrier to keep the reagents localized on the organoid tissue and preventing the mixing of reagents. After a wash in DPBS, sections were permeabilized for 45 minutes at room temperature in PBST followed by blocking for 45 minutes at room temperature in PB. Then, sections were incubated with the desired antibody combination diluted in PB (*listed in table 3.1*) for 3 hours at 37°C. After the incubation, sections were washed three times during 10 minutes in PB followed by a 45 min incubation at room temperature with fluorescent conjugated secondary antibodies (*listed in table 3.2*), and a 5 minutes incubation of DAPI

(10 µg/ml, Roche) in PBS for nuclear counterstaining. After washing twice with DPBS, slides were dried and coverslips were mounted with Mowiol mounting solution (0713.2, Roth).

Image acquisition was done by using a confocal microscope LSM880 (Carl Zeiss Micro Imaging) and obtained data were processed and analysed with the Zeiss ZEN 2011 software (Carl Zeiss, Inc.).

Primary antibodies	Distributor	Cat. No.	Dilution
CEMP2	ThermoFisher Scientific	BS-8059R	1:200
DCX	Merck Millipore	AB2253	1:500
EMX1	Merck Millipore	HPA006421	1:100
FOXC1	Abcam	ab18259	1:400
GBX2	proteintech	21639-1-AP	1:200
IGDCC3 (PUNC)	Santa Cruz Biotechnology	sc-514023	1:200
LGR5	Origene	TA503316	1:200
MCAM	Santa Cruz Biotechnology	sc-18837	1:100
OTX2	R&D Systems	AF1979	1:500
PAX6	DSHB	Supernatant	1:22
PRTG	ThermoFisher Scientific	TA501394	1:200
SDC1	Santa Cruz Biotechnology	sc-390791	1:100
TFAP2A (3B5)	DSHB	Concentrate	1:100

Table 3.1. Primary antibodies. Table showing the primary antibodies and the concentration used for the immunostainings of cells.

Secondary antibodies	Distributor	Cat. No.
Goat anti-Mouse IgG1 Cross-Adsorbed Secondary Antibody, Alexa Fluor 546	Invitrogen	A-21123
Goat anti-Mouse IgG1 Cross-Adsorbed Secondary Antibody, Alexa Fluor 488	Invitrogen	A-21121
Goat anti-Mouse IgG2a Cross-Adsorbed Secondary Antibody, Alexa Fluor 546	Invitrogen	A-21143
Goat anti-Mouse IgG2b Cross-Adsorbed Secondary Antibody, Alexa Fluor 647	Invitrogen	A-21242
Goat anti-Mouse IgG2b Cross-Adsorbed Secondary Antibody, Alexa Fluor 488	Invitrogen	A-21141
Goat anti-Mouse IgM Heavy Chain Secondary Antibody, Alexa Fluor 647	Invitrogen	A-21238
Goat anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 546	Invitrogen	A-11010
Goat anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 647	Invitrogen	A-21244

Table 3.2. Secondary antibodies. Table showing the secondary antibodies together used for the immunostainings of cells. Concentration used for all was 1:700 dilution.

3.7. Fluorescence-activated cell sorting of live cells

Cells were harvested and gently dissociated using HBSS (14025092, Gibco) containing 7.5 ml HEPES 1M for a non-enzymatic dissociation to avoid disruption of surface antigens. Cells were collected and centrifuged at 300 xg for 5 minutes and resuspended in HBSS with 5% FBS and ROCKi (10 μ M). Then, cells were filtered through a 35- μ m mesh cell strainer caps and counted to obtain a single cell suspension of approximately 10×10^6 cells per ml for analysis. Cell viability was determined by trypan blue dye exclusion for a viability above 80% before use for analysis and sorting experiments. Surface antigens were labeled by incubating with MCAM and PRTG primary antibodies diluted in HBSS with 5% FBS and ROCKi (10 μ M) (*listed in table 3.3*) for 15 minutes

on ice, followed by two washes in HBSS and incubation for 10 minutes with the appropriate fluorescent secondary antibodies diluted in HBSS with 5% FBS and ROCKi (10 μ M) (*listed in table 3.4*). After incubation, cells were washed twice and resuspended in HBSS with ROCKi (10 μ M) for sorting. Cell sorting was done with a BD FACSAria™ Fusion flow cytometer (BD Biosciences) by using an 85 μ m nozzle and collecting the sorted cells in FBS with ROCKi (10 μ M).

For sorting we used the FACSDiva software (BD Biosciences) and fluorescence was determined by the analysis and gating against appropriate controls. Flow cytometry forward (FSC) versus side scatter (SSC) density plots were first used to exclude debris. Then, singlets were selected based on FSC-area versus FSC- width, and SSC-area versus SSC-width. Samples stained only with the secondary antibodies were used as controls to set the appropriate negative gates since there was some false positive background signal.

Primary antibodies	Distributor	Cat. No.	Dilution
MCAM	Santa Cruz Biotechnology	sc-18837	1:50
PRTG	ThermoFisher Scientific	TA501394	1:100

Table 3.3. Primary antibodies use for the pre-staining of cells for FACS analysis. Cell-surface staining was done by using the antibodies listed individually or in combination.

Secondary antibodies	Distributor	Cat. No.	Dilution
Goat anti-Mouse IgG1 Cross-Adsorbed Secondary Antibody, Alexa Fluor 488	Invitrogen	A-21121	1:2000
Goat anti-Mouse IgG2a Cross-Adsorbed Secondary Antibody, Alexa Fluor 546	Invitrogen	A-21143	1:4000

Table 3.4. Secondary antibodies use for the detection of surface antigens for FACS analysis. For MCAM the anti-Mouse IgG1 on the 488-emission wavelength was used, and for PRTG anti-Mouse IgG2a on the 546-emission wavelength.

After sorting, data were additionally analysed by using FlowJo software (Tree Star, Ashland, OR, <http://www.treestar.com>), and collected cells were either replated in conditioned medium or pelleted down and snap frozen for downstream analysis.

For neuronal terminal differentiation, sorted cells were replated at very high density (400,000 cells per cm²) and differentiated for 14 days with NB medium supplemented with BDNF (20 ng/ml), ascorbic acid (0.2 mM), GDNF (20 ng/ml) and DAPT (10 μ M).

3.8. RNA extraction

Total RNA was isolated using the miRNeasy[™]RNA Mini Kit (Qiagen) according to the manufacturer's instructions. Briefly, cell pellets were resuspended and homogenized in 700 μ l of QIAzol lysis reagent. After incubation at room temperature for 5 minutes, 140 μ l of chloroform was added and incubated again for 3 minutes. Next, the homogenate was then centrifuged at 12000 xg for 15 min and at 4°C. After centrifugation, 300 μ l of the aqueous phase was collected and mixed with 450 μ l of 100% ethanol. The sample was then transferred into a RNeasy mini column and centrifuged at 8000 xg for 15 seconds. After washing with 350 μ l of RWT buffer and centrifuged for 15s at 8000 xg, 80 μ l of DNase I mix was added directly into the RNeasy mini spin column membrane and incubated for 15 minutes at room temperature. After incubation, two washes were done with 500 μ l RWT and RPE buffers and centrifuging for 15 s at 8000 xg. The membrane of the RNeasy mini spin column was dried by additional centrifugation and, lastly, samples were eluted in 20 μ l of RNase free water. RNA quality assessment and concentration measurement was done by using the Nanodrop (for samples being used for RT-qPCR) or the TapeStation (for samples being used for library preparation).

3.9. cDNA synthesis

250 ng of total RNA was used for cDNA synthesis by using the high-capacity cDNA reverse transcription kit (4368814, Applied biosystems) as described by the manufacturer.

Briefly, RNA was added to the master mix and incubated on the thermocycler as described on *table 3.5*.

cDNA reaction (20 μ l)	Cycle n ^o	T [$^{\circ}$ C]	Time [h:min:sec]
2 μ l RT buffer x10	1	25	10:00
2 μ l Random Primers x10	2	37	2:00:00
0.8 μ l dNTPs mix x25 (100mM)	3	85	0:5
0.25 μ l RNase inhibitor	4	4	hold
0.25 μ l RT enzyme			
5 μ l RNA (250 ng)			
Up to 20 μ l Nuclease free water			

Table 3.5. Brief protocol for cDNA synthesis. Master mix reaction and thermocycler conditions for conducting the cDNA preparation.

3.10. Real-time quantitative PCR

Quantitative real-time PCR reaction was performed with the FastStart Universal SYBR green (Roche) on a QuantStudio 7 Flex Real-Time PCR System (Applied Biosystems) following instructions (*table 3.6*). Three biological replicates were used for each condition and normalized on HPRT expression levels. Oligonucleotides used to measure mRNA levels are listed in *table 3.7*. Melt and standard curves for each primer set were generated to confirm that only one amplicon was generated at the same efficiency as the housekeeping gene HPRT. Obtained data were analyzed using the $\Delta\Delta$ CT method described elsewhere (Livak & Schmittgen, 2001). For statistical analysis, the GraphPad Prism software was used.

RT-PCR reaction (12 µl)	Cycle n ^o	T [°C]	Time [min:sec]
6 µl FastStart Universal SYBR Green Master (ROX)	1	48	30:00
0.4 µl Forward Primer (30 µM)	2	95	10:00
0.4 µl Reverse Primer (30 µM)	[3	95	0:15
5 µl of cDNA lysate (6 ng)	4] x 40	60	1:00
Up to 12 µl Nuclease free water	5	95	0:15
	6	60	0:15
	7	95	0:15

Table 3.6. Brief protocol for RT-qPCR. Master mix reaction and thermocycler conditions for conducting the RT-qPCR.

Target	Forward oligonucleotide (5'-3')	Reverse oligonucleotide (5'-3')
DCX	CATCCCCAACACCTCAGAAGA	CGTTTGCTGAGTCAGCTGGA
EMX1	GAGACGCAGGTGAAGGTGTG	CACCGGTTGATGTGATGGGA
EMX2	GTCATCGCTTCCAAGGTAAAAGT	TGTTGCGAATCTGAGCCTTCT
EOMES	AACCACTGGCGCTTCCA	AACATACATTTTGTTGCCCTG
FOXG1	AGGAGGGCGAGAAGAAGAAC	TGAACTCGTAGATGCCGTTG
GBX2	GCGGTGACCTGGGGTTC	GAGAAGCTCTCCTCCTTGCC
HES5	ACCAGCCCAACTCCAAGCT	GGCTTTGCTGTGCTTCAGGTA
HPRT	TGACACTGGCAAAACAATGCA	GGTCCTTTTCACCAGCAAGCT
IRX3	GATCGCTGTAGTGCTTGGGA	CAGATGGTTCTGGGGCCG
LHX2	CAAAAGACGGGCCTCACCAA	TTCTGCCGTAAGAGGTTGC

OTX2	GGGAGTGAAGAGGGAAGGGA	GTGAGAGTTCAAAGCAGGGC
PAX6	CACACCGGTTTCCTCCTTCA	GGCAGAGCGCTGTAGGTGTTT
SIX3	CAGCAAGAAACGCGAACTGG	TGCTGGAGCCTGTTCTTGG
SOX2	GGCAATAGCATGGCGAGC	TTCATGTGCGCGTAACTGTC

Table 3.7. RT-qPCR oligonucleotides. Oligonucleotides used for conducting the quantitative Real-Time PCR experiments.

3.11. Preparation of RNA-seq libraries

RNA-seq libraries were generated by using the TruSeq RNA Library Preparation Kits (Illumina), as described by the manufacturer. Briefly, the main steps are described below.

3.11.1. Purification and fragmentation of mRNA

100 ng per sample of total RNA was diluted with nuclease-free ultra-pure water to a final volume of 50 μ l with the RBP barcode label. 50 μ l of RNA Purification Beads was added to each sample to bind the poly-A RNA to the oligo dT magnetic beads. Samples were added to the thermocycler for mRNA denaturation (65°C for 5 minutes, 4°C hold) to denature the RNA and facilitate binding of the poly-A RNA to the beads was done. Next, samples were incubated at room temperature for 5 minutes to allow the RNA to bind to the beads. Without disturbing the beads, the supernatant was discarded. Beads were washed by adding 200 μ l of Bead Washing Buffer per sample and incubated at room temperature for 5 minutes. Without disturbing the beads, the supernatant was discarded, and 50 μ l of Elution Buffer was added to each sample. Samples were placed on the pre-programmed thermal cycler (80°C for 2 minutes, 25°C hold) to elute the mRNA from the beads. 50 μ l of Bead Binding Buffer was added to each sample and were incubated at room temperature for 5 minutes. Supernatant was then removed and beads were washed by adding 200 μ l of Bead Washing Buffer. After 5 minutes of incubation at room temperature, the supernatant was discarded and 19.5 μ l of Elute, Prime, Fragment Mix was added per sample which were added to the thermocycler (94°C for 8 minutes, 4°C hold) to elute, fragment, and prime the RNA. The Elute, Prime, Fragment Mix contains

random hexamers for RT priming and serves as the 1st strand cDNA synthesis reaction buffer.

3.11.2. Synthesis of the first strand of cDNA

17 μ l of the supernatant (fragmented and primed mRNA) was transferred to new eppendorfs, and 50 μ l SuperScript II was added to the First Strand Master Mix tube (ratio: 1 μ l SuperScript II for each 7 μ l First Strand Master Mix). Samples were then incubated in the thermocycler using the 1st Strand program: 25°C for 10 minutes, 42°C for 50 minutes, 70°C for 15 minutes and hold at 4°C.

3.11.3. Synthesis of the second strand of cDNA

25 μ l of thawed Second Strand Master Mix was added to each sample and incubated in a pre-heated thermal cycler at 16°C for 1 hour. After incubation, 90 μ l of well-mixed AMPure XP beads was added to 50 μ l of ds cDNA and incubated at room temperature for 15 minutes. Then, 135 μ l of the supernatant was removed and beads were washed twice by adding 200 μ l of freshly prepared 80% EtOH for 30 seconds. Then, samples were air-dried for 15 minutes at room temperature and 52.5 μ l of Resuspension Buffer was added and incubated at room temperature for 2 minutes. Aplacing samples on the magnetic rack, 50 μ l of the supernatant (ds cDNA) was transferred to new tubes.

3.11.4. End repair

10 μ l of diluted End Repair Control (or 10 μ l of Resuspension Buffer if not using End Repair Control) was added to the 50 μ l of ds cDNA. Then, 40 μ l of End Repair Mix was added and samples were incubated on the pre-heated thermal cycler at 30°C for 30 minutes. After incubation, 160 μ l of well-mixed AMPure XP Beads were added to the samples and incubated at room temperature for 15 minutes. Then, 127.5 μ l of the supernatant was removed beads were washed twice by adding 200 μ l of freshly prepared 80% EtOH for 30 seconds. Then, samples were air-dried for 15 minutes at room temperature and resuspended in 17.5 μ l Resuspension Buffer. After incubating for 2 minutes room temperature and placing samples on the magnetic rack, 15 μ l of the clear supernatant was transferred to new tubes.

3.11.5. 3'-ends adenylation

2.5 μ l of diluted A-Tailing Control (or 2.5 μ l of Resuspension Buffer if not using A-Tailing Control) was added to each sample together with 12.5 μ l of A-Tailing Mix. Samples were then incubated on the pre-heated thermal cycler at 37°C for 30 minutes.

3.11.6. Adapters ligation

2.5 μ l of diluted Ligase Control (or 2.5 μ l of Resuspension Buffer if not using Ligase Control) was added to each sample together with 2.5 μ l of single RNA Adapter Indexes. Samples were then incubated on the pre-heated thermal cycler at 30°C for 10 minutes. To inactivate the ligation mix, 5 μ l of Stop Ligase Mix was added to each sample. After incubation, 42 μ l of well-mixed AMPure XP Beads were added to the samples and incubated at room temperature for 15 minutes. Then, 127.5 μ l of the supernatant was removed beads were washed twice by adding 200 μ l of freshly prepared 80% EtOH for 30 seconds. Then, samples were air-dried for 15 minutes at room temperature and resuspended in 52.5 μ l Resuspension Buffer. After incubating for 2 minutes room temperature and placing samples on the magnetic rack, 50 μ l of the clear supernatant was transferred to new tubes. Another clean-up round was done by adding 50 μ l of mixed AMPure XP Beads. After washes, 95 μ l of the supernatant was removed and discarded. After air-drying samples for 15 minutes at room temperature, 22.5 μ l of Resuspension Buffer was added. After incubating for 2 minutes room temperature and placing samples on the magnetic rack, 20 μ l of the clear supernatant was transferred to new tubes.

3.11.7. Enrich DNA Fragments

5 μ l of the PCR Primer Cocktail was added to each sample together with 25 μ l of PCR Master Mix. Library was then amplified by using the following thermocycler settings: 98°C for 30 seconds and 15 cycles of: 98°C for 10 seconds, 60°C for 30 seconds, 72°C for 30 seconds, 72°C for 5 minutes, and hold at 4°C. After incubation, 50 μ l of well-mixed AMPure XP Beads were added to the samples and incubated at room temperature for 15 minutes. Then, 95 μ l of the supernatant was removed beads were washed twice by adding 200 μ l of freshly prepared 80% EtOH for 30 seconds. Then, samples were air-dried for 15 minutes at room temperature and resuspended in 32.5 μ l Resuspension Buffer.

After incubating for 2 minutes room temperature and placing samples on the magnetic rack, 30 μ l of the clear supernatant was transferred to new tubes.

Finally, libraries were validated by using a TapeStation (Agilent 4200) to check for quality control and accurate quantification. The final product was a band at approximately 260 bp (for single-read libraries). Samples were then sequenced on the Illumina HiSeq 2500 sequencer as 100 bp paired-end reads.

3.12. RNA-Seq processing and analysis

Raw RNA-Seq reads were processed by trimming using Trimmomatic v0.36 (Bolger et al., 2014). The following parameters were used: leading:3; trailing:3; sliding window:4:15; minlen:36. Trimmed reads were then mapped to the human reference genome hg38 with gencode v29 as a reference transcriptome (https://www.encodegenes.org/human/release_29.html) using STAR v2.6.1d (Dobin et al., 2013). After mapping, FPKM values for each gene and corresponding isoforms were estimated with RSEM v1.3.1 (Li & Dewey, 2011) and aligned to the reference transcriptome by using the STAR aligned bam. Principal component analysis was run on the logged FPKM expression values (base 10 with a pseudocount of 1) using the top 10,000 genes with the highest variance. *The analysis was done by my colleague Daniel Rosebrock.*

3.13. Single cell RNA sequencing

3.13.1. Sample preparation

For generating the 2D monolayer differentiation dataset, neural rosettes were picked at the three different timepoints (day 12, day 35, and day 50). Cells were collected in 15 ml falcon tubes containing Accutase and were incubated in a water bath for 5 minutes at 37°C. 25 μ g/ml of DNase I (07469, STEMCELL technologies) was added to reduce viscosity.

After incubation, the detachment of the cells was achieved by flushing the cells and resuspending them in fresh mTESR₁ medium containing 10 μ M ROCK inhibitor. The single cell suspension was then centrifuged at 270 xg for 5 min at room temperature.

After centrifugation, the supernatant was aspirated, and cells were washed once with PBS containing 0.4% BSA and then counted to obtain the desired concentration.

3.13.2. Library construction

Chromium Next GEM Single Cell 3' GEM, Library & Gel Bead Kit (PN-1000121, 10xGenomics) with the v3.1 chemistry. Briefly, a cell suspension of 16500 cells per sample (1000/ μ l) was taken aiming at a recovery of 10000 cells per sample, as recommended in the protocol. Cells were then used to generate the Gel Bead-In-Emulsions (GEMs) and then followed by library preparation as suggested by v3.1 single cell kit protocol as mentioned by manufacturer. Briefly, the main steps are described below.

3.13.3. GEM generation and barcoding

70 μ l Master Mix was added to the cell suspension and was gently dispensed into the bottom center of each well in row labeled 1 without introducing bubbles. After vortexing, 50 μ l of Gel Beads were dispensed into the wells in row labeled 2 without introducing bubbles. 45 μ l Partitioning Oil was added into the wells in row labeled 3 from a reagent reservoir. And, 50% Glycerol was dispensed into all the unused chip wells, following the same volumes. Finally, the chip was covered with a gasket. Assembled chip with the gasket was then placed inside the Chromium Controller ensuring that the chip stays horizontal.

After the program was completed, 100 μ l GEMs were slowly aspirated from the lowest points of the recovery wells in the top row labeled 3 without creating a seal between the tips and the bottom of the wells. GEMs should appear opaque and uniform across all channels. Over the course of \sim 20 sec, GEMs were dispensed into a new tube strip on ice with the pipette tips against the sidewalls of the tubes. Then, the samples incubated in the thermocycler with the following program: 53°C for 45 minutes, 85°C for 5 minutes, and hold at 4°C.

3.13.4. Post GEM-RT cleanup and cDNA amplification

Add 125 μ l of Recovery Agent to each sample at room temperature without disrupting the biphasic mix. The resulting biphasic mixture contains Recovery Agent/Partitioning Oil (pink) and aqueous phase (clear), with no persisting emulsion (opaque). Then, 125 μ l

of the Recovery Agent/Partitioning Oil (pink) was slowly removed and discarded from the bottom of the tube, being careful not to aspirate any aqueous sample. 200 μ l of pre-vortexed Dynabeads MyOne SILANE were added to each sample and incubated for 10 min at room temperature. At the end of 10 min incubation, samples were placed on a 10x Magnetic Separator until the solution was cleared. 300 μ l of 80% ethanol was added to the pellet while on the magnet for 30 seconds and removed. Another 200 μ l 80% of ethanol was added repeating the previous step. Samples were then air-dried for 1 minute and immediately 35.5 μ l of Elution Solution was added to each sample. After incubation for 2 minutes at room temperature, samples were placed on a 10x Magnetic Separator until the solution was cleared. Then, 35 μ l of sample was transferred to a new tube strip for amplifying cDNA. 65 μ l of cDNA Amplification Reaction Mix was added to 35 μ l of each sample and incubated in a thermal cycler with the following protocol (*table 3.8*).

cDNA amplification reaction (100 μ l)	Cycle n ^o	T [°C]	Time [min:sec]
	1	98	03:00
50 μ l Amp Mix	[2	98	0:15
15 μ l cDNA Primers	3	63	0:20
35 μ l sample	4] x 9	72	1:00
	5	72	1:00
	6	4	Hold

Table 3.8. Brief protocol for cDNA amplification. Master mix reaction and thermocycler conditions for amplifying cDNA.

At the end of the program, 60 μ l of SPRIselect reagent (0.6X) was added to each sample and incubated for 5 minutes at room temperature. Then, samples were placed on a 10x Magnetic Separator until the solution was cleared and the supernatant was removed.

Beads were washed twice by adding 200 μ l of 80% ethanol to the pellet while on the magnet for 30 seconds and removed. Samples were then air-dried for 1 minute and immediately 40.5 μ l of Elution Solution was added to each sample. After incubation for 2 minutes at room temperature, samples were placed on a 10x Magnetic Separator until the solution was cleared, and 40 μ l of sample was transferred to new tubes.

Final quality assessment and quantification of the amplified cDNA was done by using the TapeStation (Agilent 4200). Accurate quantification is essential since the total number of SI PCR cycles need to be optimized based on carrying forward a fixed proportion (10 μ l, 25%) of the total cDNA yield calculated during post cDNA amplification QC and quantification.

3.13.5. 3'-Gene expression library construction

10 μ l of the purified cDNA sample (25%) was used for generating the 3'-Gene Expression libraries. 25 μ l of Buffer EB and 15 μ l of Fragmentation Mix were added together to each sample. Samples were then transferred into the pre-cooled thermal cycler (4°C) and first step was skipped to initiate the protocol (*table 3.9*).

Fragmentation reaction (50 μ l)	Cycle n ^o	T [°C]	Time [min:sec]
5 μ l Fragmentation Buffer	1	4	Hold
10 μ l Fragmentation Enzyme	2	32	30:00
10 μ l purified cDNA	3	65	5:00
25 μ l Buffer EB	4	4	Hold

Table 3.9. Brief protocol for cDNA fragmentation. Master mix reaction and thermocycler conditions for fragmenting the cDNA.

At the end of the program, 30 μl of SPRIselect reagent (0.6X) was added to each sample and incubated for 5 minutes at room temperature. Then, samples were placed on a 10x Magnetic Separator until the solution was cleared and 75 μl of the supernatant was transferred to new tubes. Then, 10 μl of SPRIselect reagent (0.8X) was added to each sample and incubated for 5 minutes at room temperature. Then, samples were placed on a 10x Magnetic Separator until the solution was cleared and the supernatant was removed. Beads were washed twice by adding 200 μl of 80% ethanol to the pellet while on the magnet for 30 seconds and removed. Samples were then air-dried for 1 minute and immediately 50.5 μl of Elution Solution was added to each sample. After incubation for 2 minutes at room temperature, samples were placed on a 10x Magnetic Separator until the solution was cleared, and 50 μl of sample was transferred to new tubes. In order to add the adaptors, 50 μl of the Adaptor Ligation Mix was added to 50 μl of sample. Then, samples were then transferred into the thermocycler with the following protocol (*table 3.10*).

Adaptor ligation reaction (50 μl)	Cycle n ^o	T [°C]	Time [min:sec]
20 μl Ligation Buffer			
10 μl DNA Ligase	1	20	15:00
20 μl Adaptor Oligos	2	4	Hold
50 μl sample			

Table 3.10. Brief protocol for adaptor ligation. Master mix reaction and thermocycler conditions for adding the adaptors to the cDNA.

At the end of the program, 80 μl of SPRIselect reagent (0.8X) was added to each sample and incubated for 5 minutes at room temperature. Then, samples were placed on a 10x Magnetic Separator until the solution was cleared and the supernatant was removed. Beads were washed twice by adding 200 μl of 80% ethanol to the pellet while on the magnet for 30 seconds and removed. Samples were then air-dried for 1 minute and

immediately 30.5 µl of Elution Solution was added to each sample. After incubation for 2 minutes at room temperature, samples were placed on a 10x Magnetic Separator until the solution was cleared, and 30 µl of sample was transferred to new tubes.

3.13.6. Sample indexing PCR

Individual indexes from the Single Index Kit T Set A (PN-1000213, 10xGenomics) were used for sample barcoding. 60 µl of Sample Index PCR Mix was added to 30 µl of each sample and incubated in a thermal cycler with the following protocol (*table 3.11*).

Sample Index PCR reaction (100 µl)	Cycle n ^o	T [°C]	Time [min:sec]
	1	98	0:45
50 µl Amp Mix	[2	98	0:20
10 µl SI Primer	3	54	0:30
30 µl sample	4] x 11	72	0:20
	5	72	1:00
	6	4	Hold

Table 3.11. Brief protocol for the indexing PCR. Master mix reaction and thermocycler conditions indexing the samples.

At the end of the program, 60 µl of SPRIselect reagent (0.6X) was added to each sample and incubated for 5 minutes at room temperature. Then, samples were placed on a 10x Magnetic Separator until the solution was cleared and 150 µl of the supernatant was transferred to new tubes. Then, 20 µl of SPRIselect reagent (0.8X) was added to each sample and incubated for 5 minutes at room temperature. Then, samples were placed on a 10x Magnetic Separator until the solution was cleared and the supernatant was removed. Beads were washed twice by adding 200 µl of 80% ethanol to the pellet while on the magnet for 30 seconds and removed. Samples were then air-dried for 1 minute and

immediately 50.5 μ l of Elution Solution was added to each sample. After incubation for 2 minutes at room temperature, samples were placed on a 10x Magnetic Separator until the solution was cleared, and 35.5 μ l of sample was transferred to new tubes. In order to add the adaptors, 30 μ l of the Adaptor Ligation Mix was added to 50 μ l of sample.

Then, quality assessment and quantification of the libraries was done by using the TapeStation (Agilent 4200). We obtained an average library of 0.55 ng/ μ l in a total of 33 μ l (18.15 ng). Finally, libraries were pooled and sequenced using the 28/91 Illumina high output sequencing aiming at 500 million fragments per sample on the Illumina NovaSeq 6000 S2 flow cell.

3.13.7. Single cell RNA sequencing data processing

The scRNA-Seq data were processed using the Cellranger v3.1.0 software (Zheng et al., 2017) was used to cluster and determine valid cell barcodes, identify unique molecular identifier (UMI) corresponding to identify and quantify unique RNA molecules for each individual cell, and map reads to the reference genome hg38 and nd ensembl reference transcriptome version 93 (http://ftp.ensembl.org/pub/release-93/gtf/homo_sapiens/Homo_sapiens.GRCh38.93.gtf.gz). Cell barcodes that had at least 10,000 unique molecular identifiers (UMIs) or at least 40% mitochondrial UMIs were filtered out from the downstream analyses together with the detected doublets. Detection of doublets was done by running scrublet version 0.2.3 (Wolock et al., 2019) setting the input parameter to ‘expected_doublet_rate=0.05’ and applying a doublet score threshold of 0.2. For normalization of gene expression values, UMI gene counts per cell were divided by the library size, then multiplied by a scaling factor of 10000, and log transformed after adding a pseudocount of 1. Individual cell expression profiles were then clustered using Scanpy v1.5.1 (Wolf et al., 2018) and clusters were assigned to *in vivo* biological cell types based on relative expression levels of well-known marker genes for cell type, brain region, and cycling status. Finally, Principal Component Analysis was then performed on the normalized expression values of the top 2000 highly variable genes. Clusters were determined using the Louvain clustering procedure on the top 50 principal components with scanpy’s louvain function.

For the PRTG-sorted datasets at day 5 and at day 10 we ran differential expression analysis comparing the NSC populations. Differential expression analysis was done by using a t -test with scanpy's `rank_genes_group` function and the method `t-test_overestim_var`. Only genes expressed in at least 5% of cells within at least one cluster were tested, and they were labelled as significantly upregulated if they had a \log_2 -transformed fold change of at least one and q -value less than 0.05 after applying a Benjamini–Hochberg multiple hypothesis correction to the estimated p -values.

For all scRNA-seq datasets generated, my colleague Daniel Rosebrock did the main analysis, and together we annotated the clusters and explored the data.

4. Results

4.1. Characterization of hiPSCs-derived cortical progenitors by scRNA-seq

As a first step towards identifying novel surface markers for early cortical NSCs, we focused on better understanding our 2D monolayer differentiation paradigm (*Figure 4.1*). Neural induction of ZIP13K2 (hiPSC line) towards cortical lineages is conducted in a 2D setting using the Triple-inhibitor (Triple-i) protocol recently established in the lab (Rosebrock et al., 2022). The triple-i protocol combines WNT inhibition using XAV 939, and TGF- β and BMP inhibition using SB-431542 and Noggin, respectively. We employ these inhibitors starting on day 2 of differentiation, after EB formation, and continue until day 9 to promote expansion of anterior neuroectodermal fates. Around day 12, neural rosette formation- a hallmark of early cortical differentiation- is visible, representing the early cortical NSC stage exhibiting epithelial features. Such rosette structures are manually picked and replated weekly in order to enrich for and propagate cortical NSCs. Progression of the culture will eventually entail the dismantlement of the rosette structures in accordance with the advancement of differentiation towards cells with reduced epithelial characteristics, such as neuronal and glial progenitors

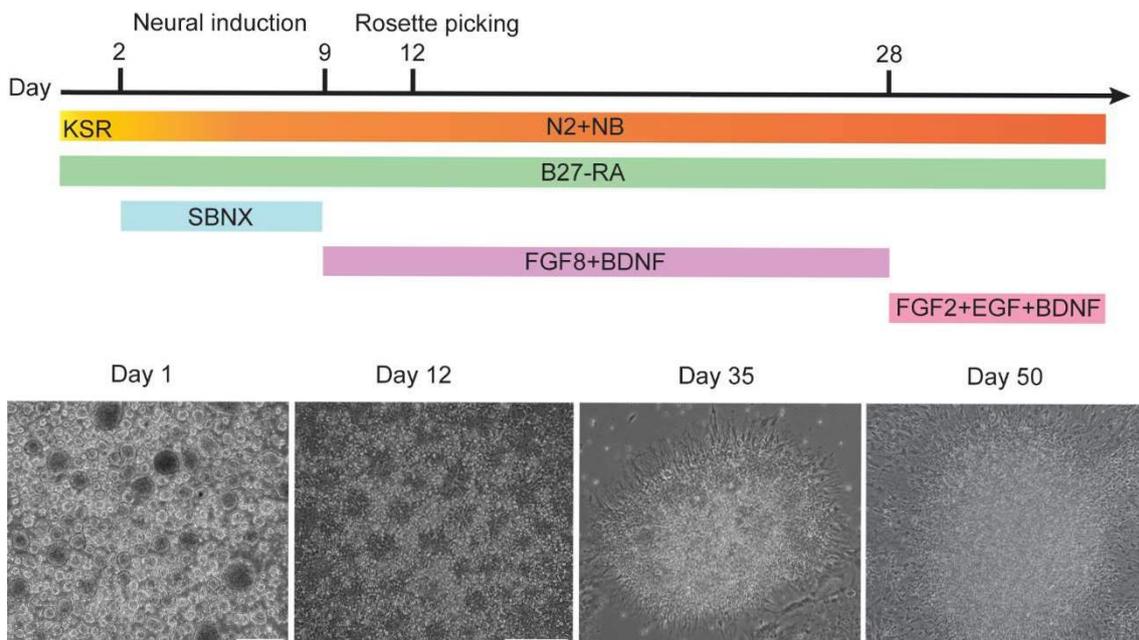


Figure 4.1. Schematic overview of the neural induction protocol. Differentiation of hiPSCs starts with the generation of EBs by short aggregation (day 1). On day 2 we start neural induction by adding the small molecule inhibitors and it continues until day 9. From

that day onwards, media is supplemented with additional factors to promote proliferation and survival of neural stem cells and derivatives. Bright-field images of some key time points of neural induction. Scale bar= 100 μ m.

To evaluate the transcriptional identity of the hiPSCs-derived cortical progenitors, we performed scRNA-seq on rosette cells collected from three relevant time points: day 12, day 35 and day 50, corresponding to early-, mid-, and late-NSC populations. Initially, we analyzed each time point individually. After clustering for the highly variable genes among the diverse cell populations we characterized each cluster based on the expression of well-known marker genes for cell state and brain regions, including dorsal pallium (neocortex), medial pallium, subpallium, and diencephalon, as well as more posterior regions such as mid/hindbrain and non-neural lineages (epithelial and mesenchymal) (*Figures 4.2, 4.3, and 4.4*).

At the earliest stage, day 12, we identify thirteen clusters which are primarily composed of dividing and non-dividing NSCs denoted by the expression of SOX2, SOX1, FABP7 and HES5. Moreover, ten out of these thirteen clusters appear to highly express anterior forebrain markers including PAX6, FOXG1, SP8, SIX3, EMX2, LHX2. However, there appears to be a clear dichotomy in the identity of these forebrain cells, where half of the population at day 12 shows anterior forebrain identity (neocortex and medial pallium) (clusters 1, 4, 6, 10, 7, and 8) while the other half shows posterior forebrain identity (clusters 0, 2, 3, and 5). Additionally, one cluster of NSCs shows expression of mid/hindbrain markers instead, such as GBX2, IRX2, IRX3 and EN1. And a small number of remaining cells form two additional clusters corresponding to pluripotent stem cells denoted by the expression of POU5F1 and NANOG, and early neurons marked by expression of DCX, ELAVL2 and ELAVL3 (*Figure 4.2*).

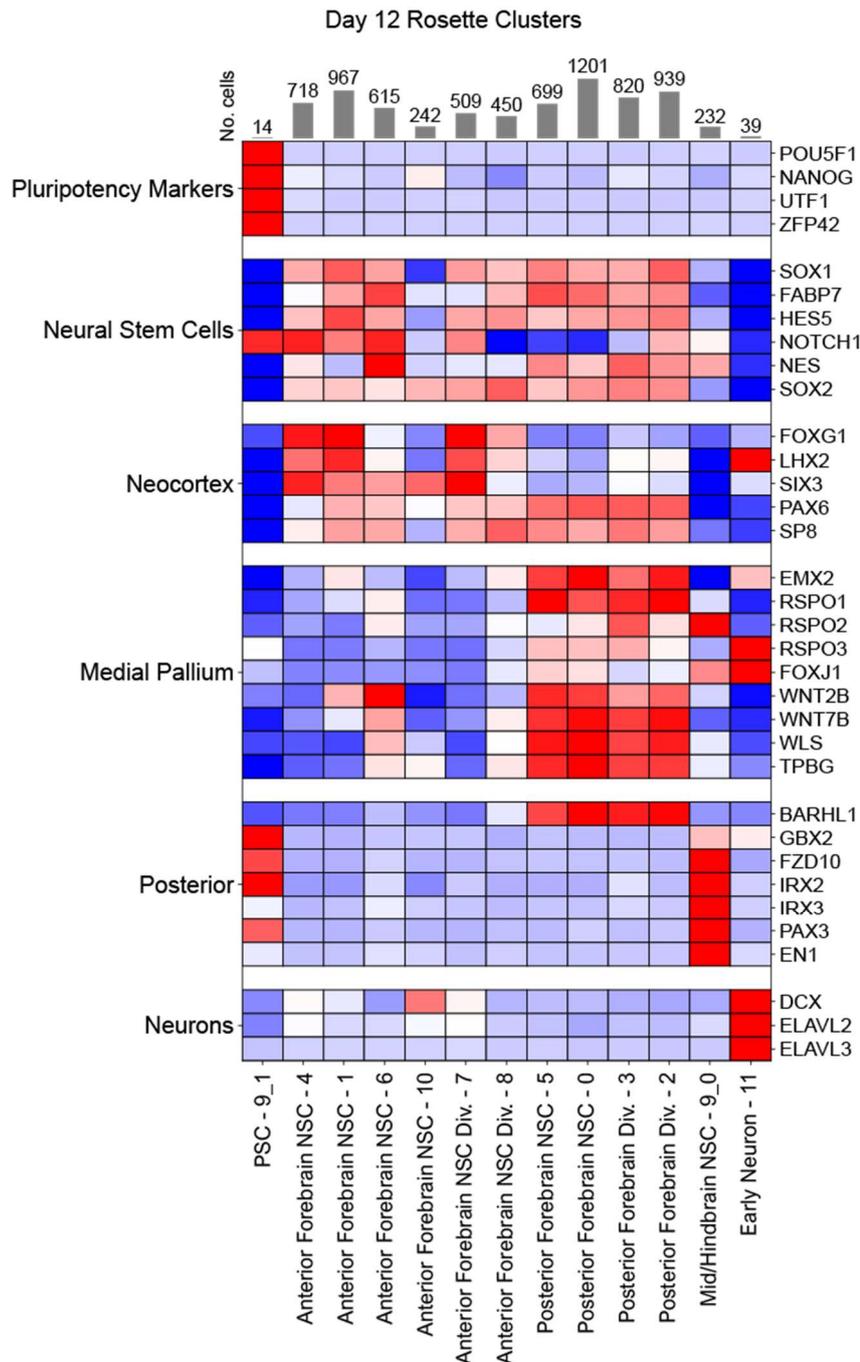


Figure 4.2. scRNA-seq analysis of day 12 hiPSC-derived neural progenitors. A heatmap representing day 12 clusters showing expression values for genes marking cell state as well as different brain regions (Neocortex, Medial pallium, Posterior). Note that there is a prevalence of forebrain NCSs identity which account for most part of the cell population: 96.15% (7,115 cells) forebrain, 3.14% (232 cells) mid/hindbrain, 0.53% (39 cells) neurons, and 0.19% (14 cells) PSCs. Bar plots above display the total number of cells within each cluster. The color scheme is based on the Z scores, with upregulation in red, downregulation in blue, and undetermined directionality in white.

At the mid stage, day 35, we observe an increase in cell type diversity as exemplified with the identification of twenty cell clusters. As expected, most of the population is composed of NSCs but we observe the appearance of more differentiated cells exemplified by the presence of intermediate progenitors (cluster 17) expressing EOMES and TBR1, and neuronal clusters of various regional identities (clusters 11, 14, 18, and 19) expressing DCX and TUBB3. Surprisingly, expression of posterior fate (diencephalon and mid/hindbrain) genes are more prominent, promoting tissues posterior to the cortex at the expense of cortical fates. Additionally, we also found clusters of choroid plexus and non-neural lineages (clusters 0, 1, 8, 13, 15, and 16) such as mesenchymal cells expressing KRT18 and KRT8, and epithelial cells expressing DCN and LUM (*Figure 4.3*).

Similarly, on day 50 we observe a larger diversity of cell types identifying 24 cell clusters. Again, we can appreciate the same cell identities that appear on day 35, but posterior identities gain even more prevalence (*Figure 4.4*). Altogether our analysis suggests that even though we induce a highly homogeneous forebrain population by day 12, the culture becomes more heterogeneous as differentiation progresses, highlighting the need for protocol improvement towards the generation of more homogeneous cortical lineages in monolayer cultures.

4.2. Identification of potential surface markers for early cortical NSCs

We made use of the scRNA-seq dataset to investigate possible surface markers for identifying early cortical NSCs in culture. In order to profile cell surface markers that identify specific regional and temporal populations in our culture system we integrated the tree time points and used the above-described clustering annotation (*Figure 4.5*).

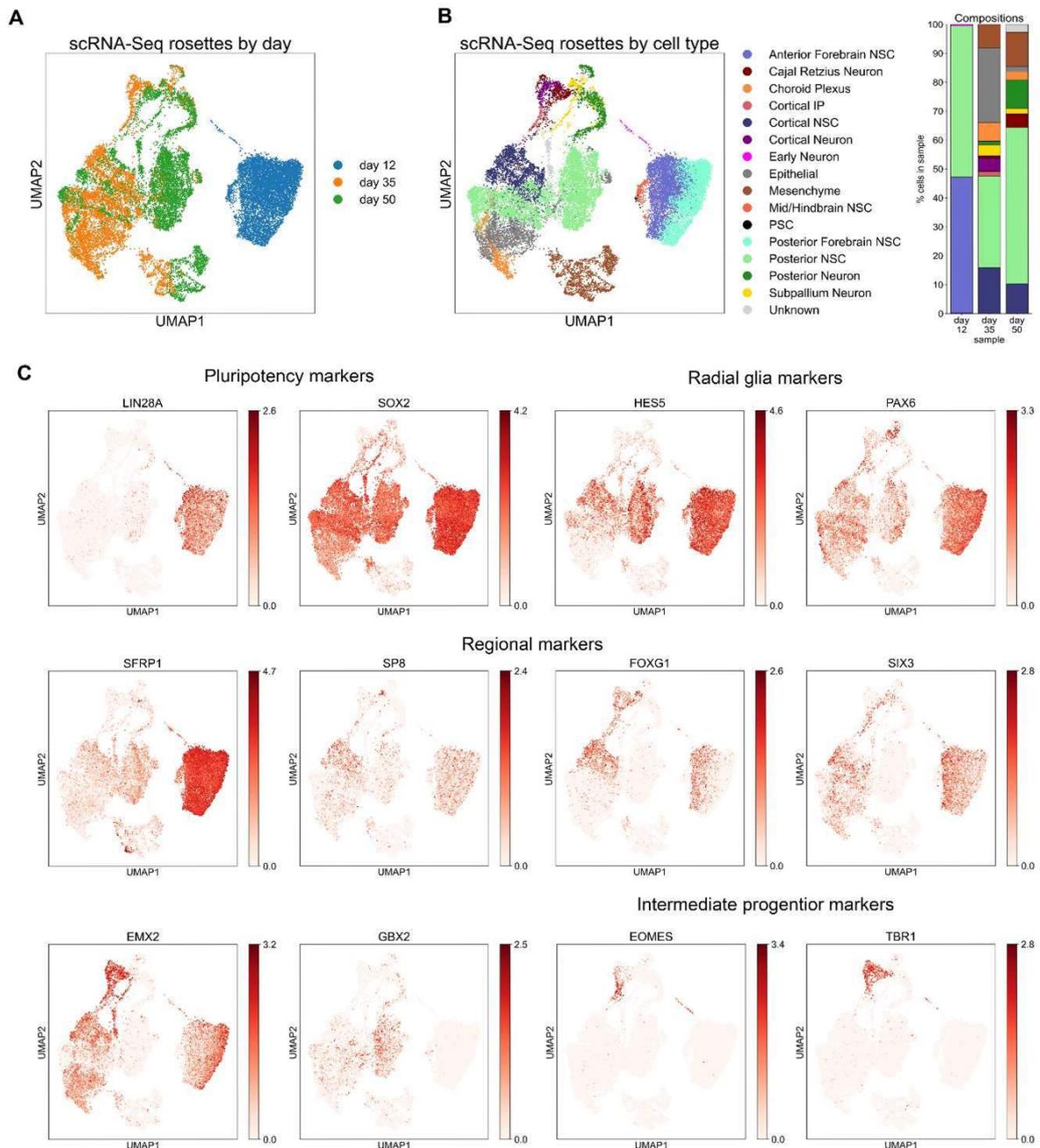


Figure 4.5. scRNA-seq merged data depicting main clusters and cell compositions of the different time points. A) Uniform Manifold Approximation and Projection (UMAP) plot of the merged scRNA-seq datasets showing the clustering of the different days

highlighted in different colors. B) UMAP plot of the merged scRNA-seq data with the distinct clusters based on cell types highlighted in different colors (right) and stacked plot of cell composition depicting the percentage of cell types within the different days (left). C) UMAP plots depicting the expression of selected marker genes which were used to annotate the scRNA-seq clusters for cells states and brain regions.

Upon merging the scRNA-seq data, we see that day 35 and day 50 either cluster together or are in great proximity as compared with day 12. This indicates how transcriptionally similar day 35 and day 50 are, while day 12 has the most distinct signature (most probably due to the high homogeneity in forebrain NSCs). We identify stage- and cell state-specific genes which exhibit differential expression patterns across the progression of neural induction (*Figure 4.6*), recapitulating *in vivo* development. Such results consolidate the fact that the progressively emerging and changing NSC populations have distinct molecular signatures, from which we sought to extract surface marker genes for a specific cell state.

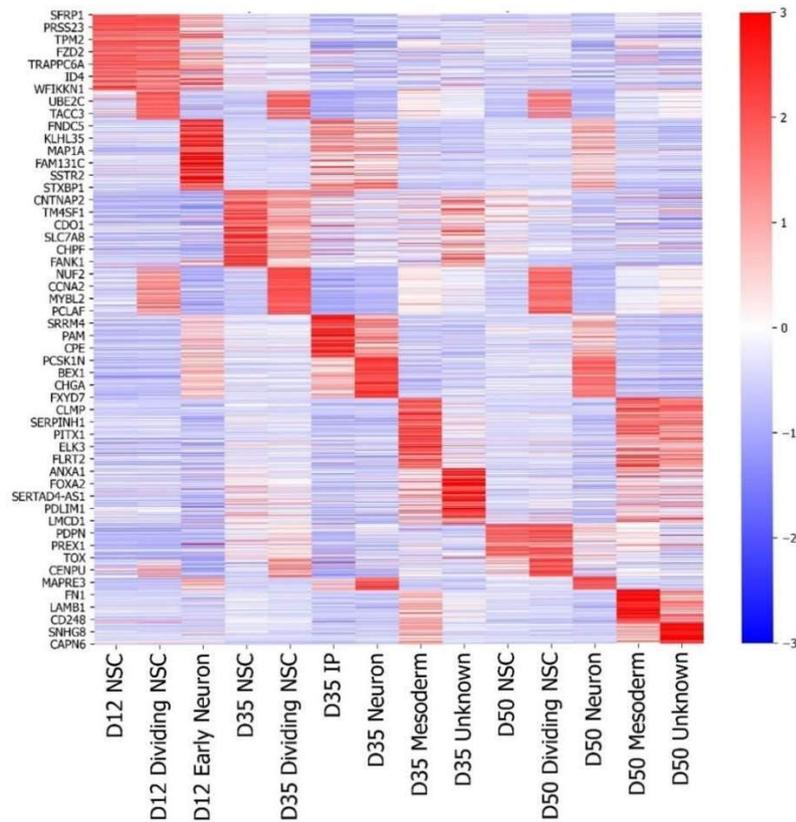


Figure 4.6. Stage-specific gene signature of neural stem cells and derivatives. Heatmap representing expression values for the top high variable genes categorized across

clusters of various differentiation days. Note the specific transcriptional signatures from where we can extract potential markers genes. Here we used a different cluster than previously shown and grouped together cell states without showing regional identity to create an overview of the temporal-specific signatures of NSCs. The color-coded scale represents relative expression levels of each gene (row) across clusters and it is based on the Z scores. Upregulation is shown in red, downregulation in blue, and undetermined directionality in white.

For selecting potential surface markers of early cortical NSCs from our scRNA-seq dataset, we examined such gene expression patterns and extracted those genes that followed five main criteria: 1) stem cell marker: must have a higher expression in the stem cell populations compared to the intermediate progenitor or neuronal cell populations; 2) cortical identity: must be regionally restricted and highly expressed in those cells with cortical identity (forebrain specific); 3) early marker: must have a clear peak expression in early days followed by a decrease in its expression at later stages; 4) representative: must be expressed in at least 50% of the cells in the target population; 5) surface marker: to be able to conduct fluorescence-activated single cell sorting (FACS) the candidate marker must be expressed in the cell membrane.

However, when analyzing the data, it becomes apparent that it is not straightforward to find a candidate that follows all the criteria. Hence, we shortlisted candidates that followed the above-mentioned criteria to the greatest extent (*Figure 4.7*), and from which we were able to obtain commercial antibodies. A total number of 6 candidate genes were selected to be validated: Melanoma Cell Adhesion Molecule (MCAM), Leucine Rich Repeat Containing G Protein-Coupled Receptor 5 (LGR5), Cell Migration Inducing Hyaluronidase 2 (CEMIP2), Immunoglobulin Superfamily DCC Subclass Member 3 (IGDCC3), Syndecan 1 (SDC1), and Protogenin (PRTG).

When looking at the global expression patterns, we see that early forebrain NSCs (day 12) have high expression of PRTG, CEMIP2, IGDCC3 and SDC1 compared to other cell states and later stages, while showing a lower expression of MCAM and LGR5. However, these latter genes show a more unique expression pattern being virtually absent in other

cell states and in the other time points. It should be noted that while we do detect high expression of some markers in mid/hindbrain NSCs on day 12, we need to consider that under our differentiation protocol there is less than 0.5% of cells accounting for this population at that stage. Most importantly, none of the genes show higher expression PSCs or neuronal populations.

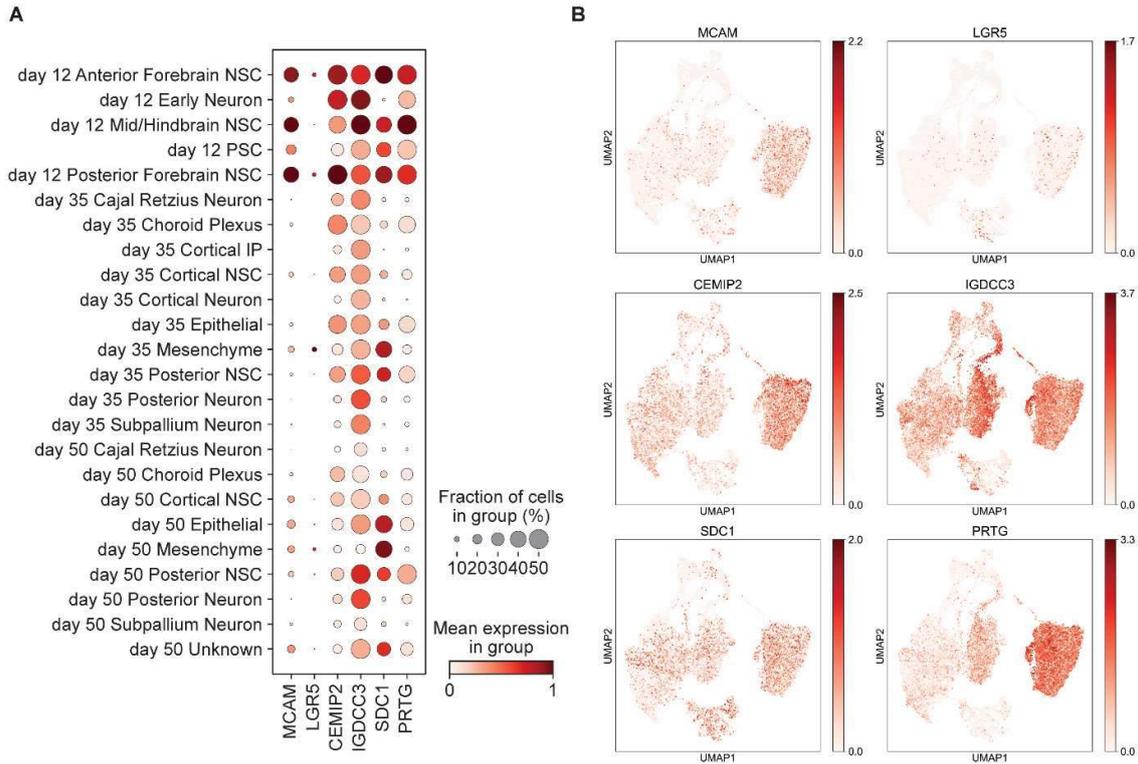


Figure 4.7. Selected potential candidates for early cortical NSCs. A) scRNA-seq dot plot depicting the expression profiles of the surface marker candidates. Each dot represents multiple features for each marker gene in each cell population. The size of each dot represents the percentage of cells expressing a given marker and the color represents the average scaled expression of a given marker. B) UMAP plots depicting expression of selected candidate surface markers on the merged scRNA-seq data.

In order to confirm the feasibility of the potential surface markers to isolate early cortical NCSs, we first checked the expression patterns of the candidate markers at a protein level by means of immunostainings at the various differentiation stages: day 0, 12, 35 and 50 -accounting for the key distinct stages in our neural induction protocol (*Figure 4.8*). In general, we observe that even though all markers seem to be expressed at day 12, they do not all follow the expected expression pattern based on the scRNA-seq results. For example, some are already being expressed in undifferentiated cells, such as CEMIP2 and SDC1, which would not allow us to see the emergence in expression of the marker by day 12, thus hindering the identification of newly emerging early cortical NCSs in culture. Other markers like LGR5, as well as CEMPI2, are expressed in higher levels at day 50 than at day 12, rendering these markers suboptimal for tracking possible reprogramming events when conducting such experiments.

The above findings result in MCAM, IGDC3 and PRTG being the only surface markers that show a high and specific enrichment in early stages (day 12), as they are present in early NSCs and are virtually absent in the cellular populations of the other time points. However, in the case of IGDC3, the staining pattern of the marker is the punctate type instead of being expressed in the cell membrane which hinders the possibility of conducting FACS. Taken together, these results show the potential of PRTG and MCAM for being used as markers of early NSC populations in cortical development.

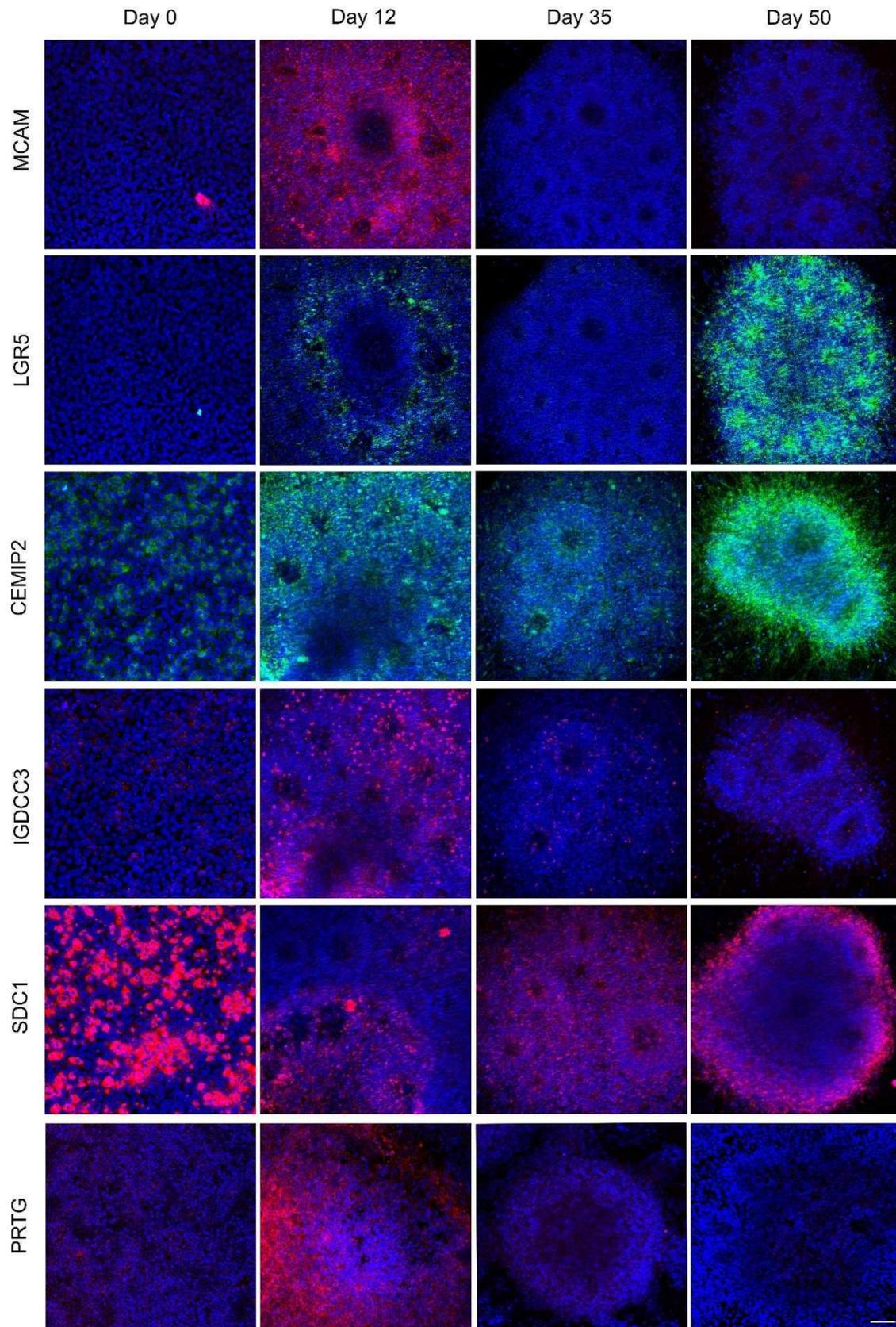


Figure 4.8. Expression of candidate surface markers through 2D differentiation. Undifferentiated (hiPSCs), ZIP13K2, were negative for most markers except for CEMIP2 and

SDC1. After 12 days, all markers show low-to-high expression levels which remain in later days in the case of LGR5, CEMIP2 and SDC1, while MCAM and PRTG show a specific enrichment on day 12. Although IGDC3 also follows the same pattern of unique early expression, its staining pattern follows a punctate type instead of a membrane type, rendering this marker unsuitable for FACS. Scale bar= 100 μ m.

To further assess the fidelity of both markers in identifying early cortical NSCs, we decided to check their expression in cerebral organoids derived under the same conditions, i.e. Triple-i protocol. In agreement with the expression of these markers in the 2D system, we find an early enrichment of PRTG and MCAM (day 15) compared to a complete absence (day 30) or lower expression on later days (day 50) (*Figure 4.9*). Whereas MCAM expression seems less specific on day 12, concomitant expression of these markers is higher in the organoid's vesicles- which represent an equivalent structure to the neural rosettes- indicating higher expression in cortical NSCs also in 3D cultures.

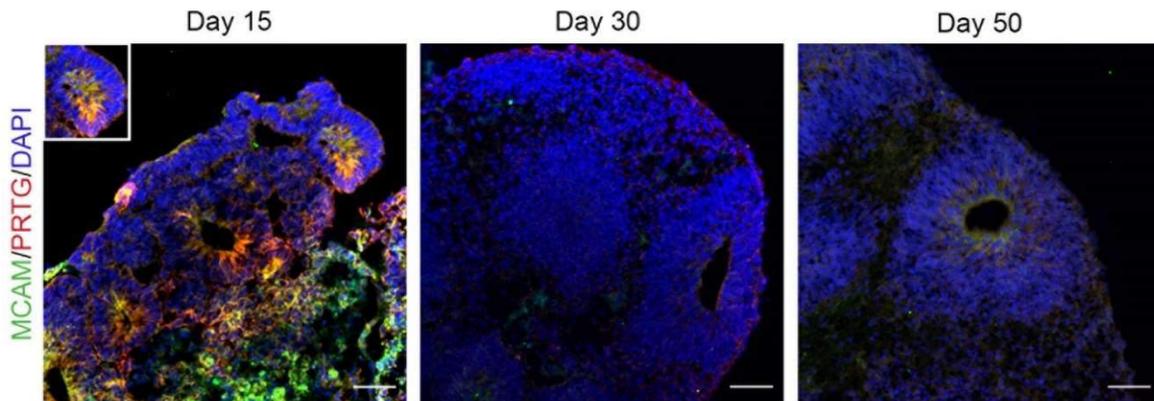


Figure 4.9. Expression of PRTG and MCAM throughout 3D differentiation. Immunostaining of PRTG and MCAM in cerebral organoids (vesicles areas were selected) show enrichment of surface markers in early days of neural induction. Scale bar= 50 μ m.

Based on these initial immunostaining results we decided to only continue with PRTG and MCAM for further characterization as they stand out as the most promising surface markers. In order to provide further robustness to our results, we used a publicly available human cortical development dataset to check the expression of these two markers in primary cortical cells (Bhaduri et al., 2020). Upon examination of the dataset, we find both markers being enriched in the early stages of human cortical development. Both

PRTG and MCAM are highly expressed in early radial glia (RG) cells, with PRTG more highly co-expressed with other known early markers such as LIN28A and DLK1 (Figure 4.10). This analysis provides higher confidence in selecting PRTG and MCAM as promising candidate for early cortical NSC markers.

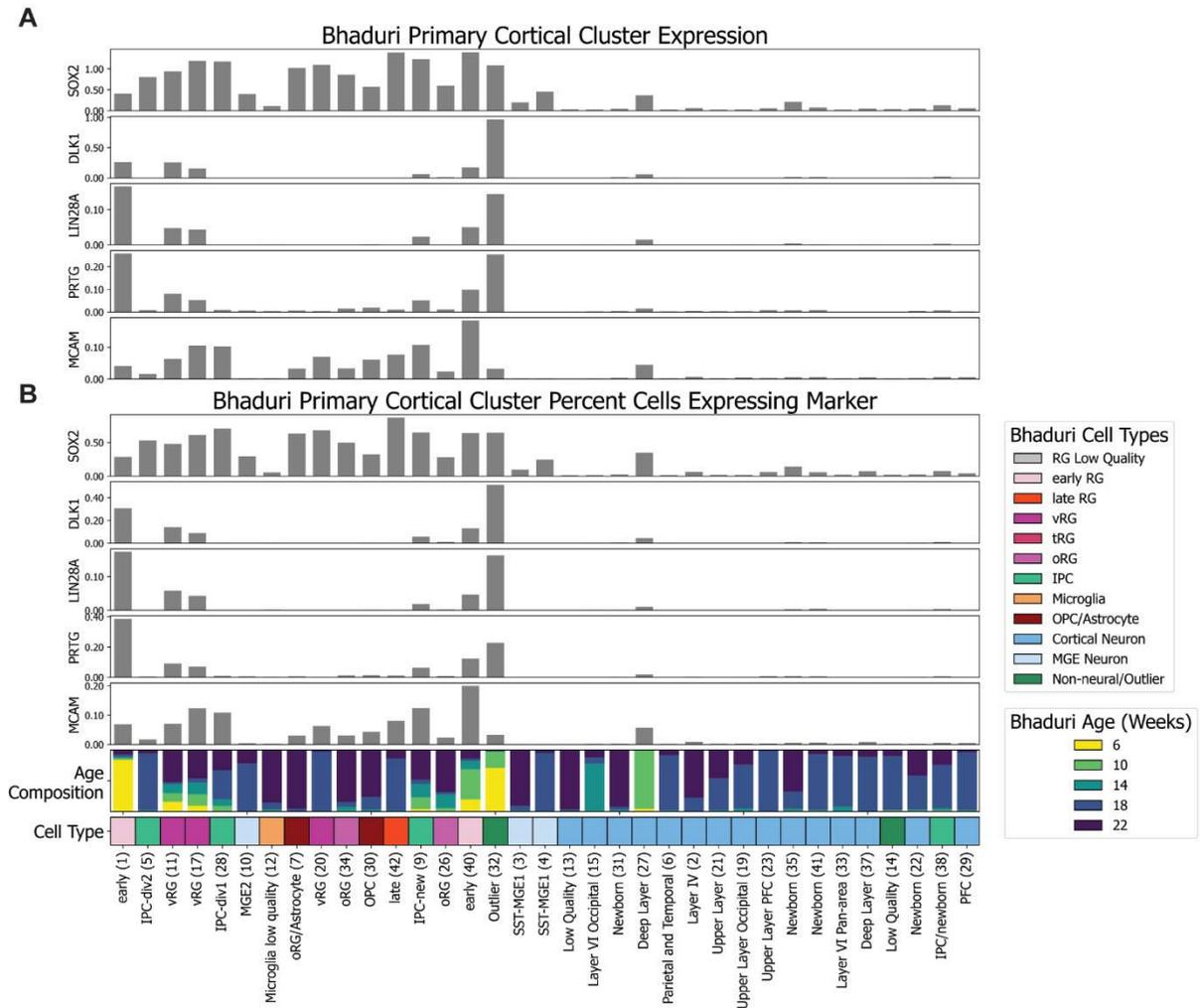


Figure 4.10. PRTG and MCAM expression during human corticogenesis. Cross-reference with *in vivo* data shows enrichment of both marker genes in the early stages of human cortical development (Bhaduri et al., 2020). Relative expression levels (A) and percentage of expressing cells (B) across primary clusters of scRNA-seq data of dissociated cells from cortical samples collected at 6–22 gestational weeks. Both PRTG and MCAM are highly enriched in early RG cells being co-expressed with other known early markers such as LIN28A and DLK1. Note that MCAM has a wider expression than PRTG, being expressed also in late RG cells.

4.3. FACS analysis confirms early expression of PRTG and MCAM following neural induction

In order to validate our approach once we determined the early expression of PRTG and MCAM, we tested whether these surface markers can serve as readout for detecting and isolating early cortical NSCs. To this end, we conducted FACS analysis across various differentiation stages followed by RNA-seq of the isolated populations (*Figure 4.11*). Co-expression of PRTG and MCAM transitions from almost undetectable to low expression in undifferentiated hiPSCs (day 0) followed by a peak expression on day 12, and a slow downregulation of both markers as differentiation progresses (day 35 and day 50). Before neural induction, undifferentiated hiPSCs present very low levels of double positive cells, 1.14% (3.5% on average). By day 12, we detect the highest double positive population accounting for 84.1% of the total cell population (89.3% on average) (*Figure 4.12*). When moving towards day 35 of differentiation, we detect a steep reduction in double positive cells only accounting for 8.21% of the total population (10.3% on average). At the latter stage, the number of double positive cells drops lower than on day 0, being 1.82% of the total cell population (1.87% on average). However, when looking individually at each marker, we do see relevant differences in their temporal specificity. This was unexpected since we did not detect it in the immunostainings but could be explained by the fact that the FACS method is more sensitive in detecting fluorescent signal. When examining MCAM individually, we see a higher level of positive cells on day 0, 35 and 50, reflecting differences in RNA and protein expression due to differential membrane protein recycling turnover. This could potentially hinder the specific detection of early NSCs if we would base it solely on MCAM sorting. However, PRTG follows more closely the pattern seen when being coimmunostained with MCAM, indicating that PRTG is more specific in its temporal expression.

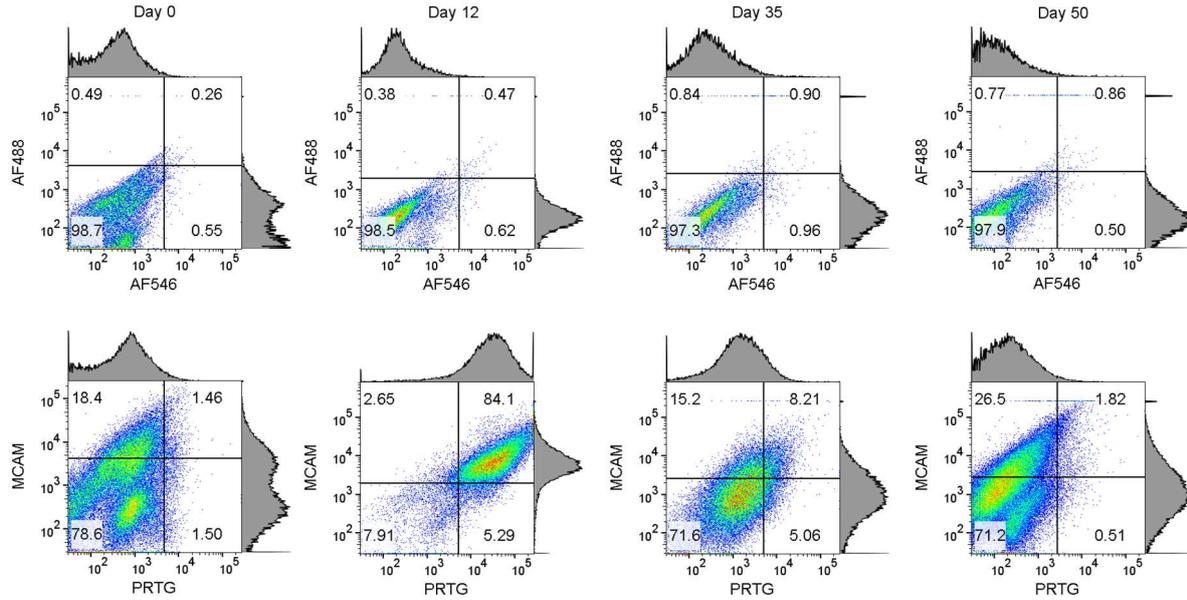


Figure 4.11. Surface markers expression profiles in neural cell types. Representative FACS analysis of PRTG and MCAM expression in hiPSCs-derived neural stem cells at various stages of differentiation. Samples incubated with the secondary antibody were used as negative controls to set the appropriate negative gates (<1%) (top). Scatter plots showing coimmunostaining of both cell surface markers and the frequency of each cell population (bottom).

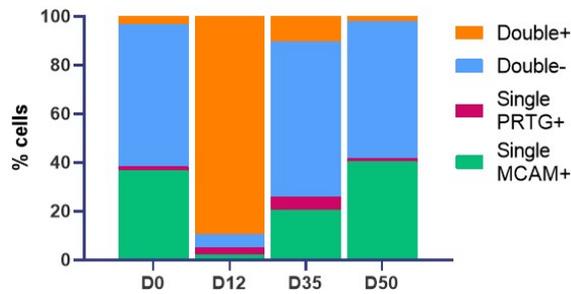


Figure 4.12. FACS analysis average cell frequencies of PRTG and MCAM. Since there is certain degree of variability with working with hiPSCs and their neural induction, here we show a stacked bar plot indicating percentage frequencies of each subpopulation across the various neural

differentiation days (n= 3 per time point).

4.4. PRTG sorting on day 12 enriches for early cortical NSCs

We next employed bulk RNA-seq to determine the cortical identity of the sorted populations in order to validate our approach in isolating early cortical NSCs within our culture. We assessed the transcriptional signature of the sorted populations at day 12 and at day 35. We collected single negative and positive cells for both time points as well as collecting double positive and double negative cells at day 12, in order to potentially

increase accuracy in detecting and isolating our population of interest. We also included undifferentiated hiPSCs to use as control.

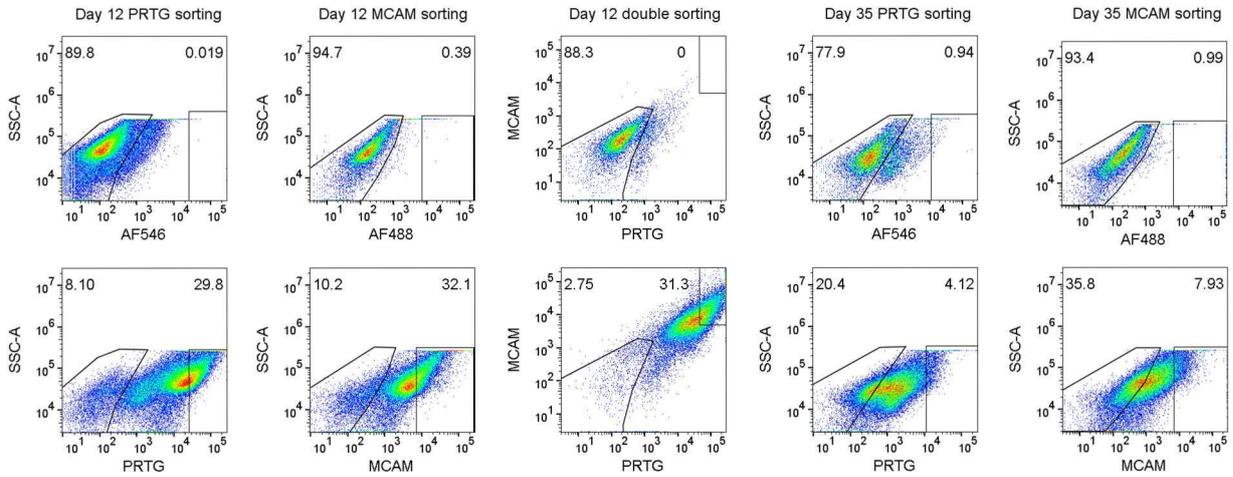


Figure 4.13. Gating strategy followed to sort and collect samples for RNA-seq. FACS density plots showing PRTG and MCAM single or combined expression based on SSC-area versus secondary antibody staining (top) or surface marker staining (bottom). Control gates as indicated were set against the corresponding negative controls based on secondary antibody staining background (<1%). The frequency of cells within each selected gate is shown (percentage). Note that for sorting we collected only the high positive PRTG, MCAM, or double stained, and not the whole positive population.

Firstly, we assessed intragroup variability by performing a Pearson’s correlation analysis across the top 2000 highly variable genes (measured using variance of log10 FPKM values across all samples) to see the extent to which the sorted populations differ or relate among each other (*Figure 4.14*).

The correlation analysis first confirms a general trend of samples grouping together based on time point which falls in line with the idea of there being different NSC populations- building blocks- at the different stages of differentiation, changing in their transcriptional identity and developing distinct commitment capabilities as differentiation progresses. However, day 35 PRTG negative is an exception, sharing a higher correlation with day 12 samples. This could suggest the ability of PRTG to segregate between a more advanced and committed subpopulation versus one that contains “early” characteristics

on day 35. Similarly, the PRTG sorted negative subpopulation shares a higher correlation than expected with the undifferentiated hiPSCs (day 0 unsorted). Whereas the PRTG positive, the MCAM positive and the MCAM negative subpopulations of day 12 are highly correlated and segregate from the day 0 undifferentiated cells, reflecting expected differences in general transcriptional identity. While such analysis provides a first glimpse into the data it does not give any information about the transcriptional identity of each subpopulation. So, we next analyzed the differences in expression levels of highly variable genes distinguishing cell state such as pluripotent or neural stem cells, as well as different brain regions (Neocortex, Subpallium, Medial pallium, Diencephalon and Mid/Hindbrain) (*Figure 4.15*).

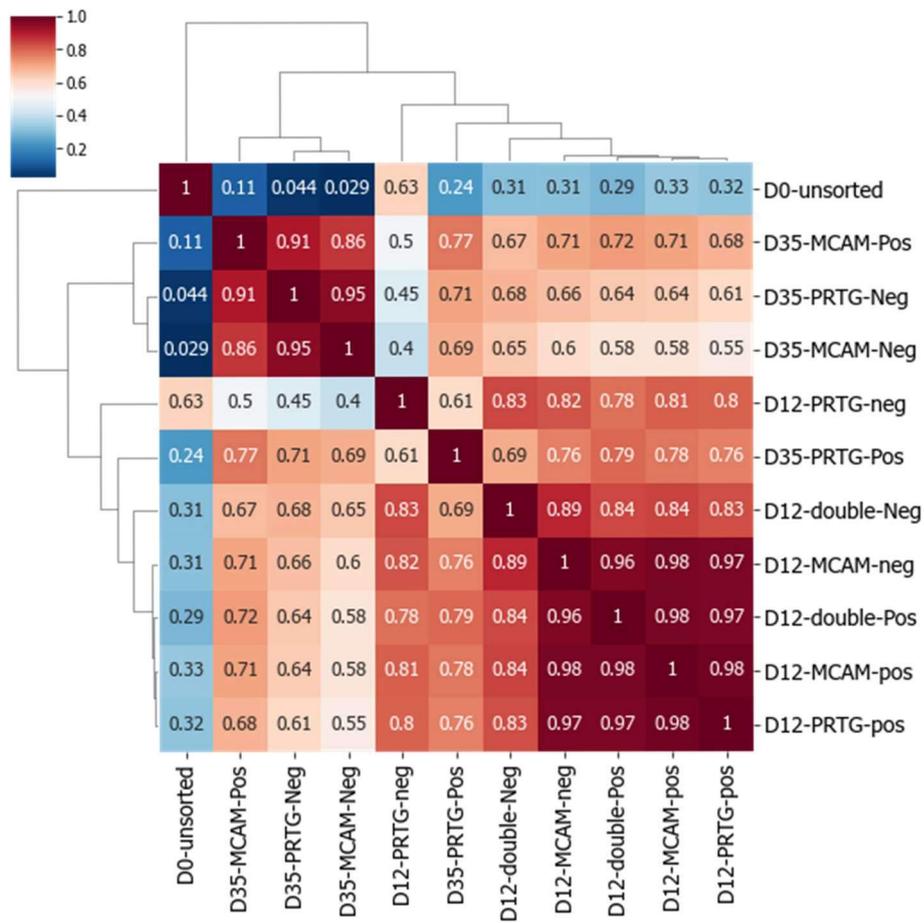


Figure 4.14. Correlation matrix of RNA-Seq datasets obtained from sorted cells. Pearson's correlation plot across log₁₀ FPKM of top 2000 highly variable genes visualizing the correlation values between samples to provide an overview of the variation between samples. A hierarchical clustering with Euclidean distance metric was used to generate the dendrogram. The scale bar represents the range of the correlation coefficients displayed from 0 to 1 (color-coded).

At day 12, sorted populations show a quite homogenous identity except for the PRTG negative subpopulation as hinted by the correlation analysis. Commonly they all express high levels of NSC markers such as FABP7, HES5, NES and SOX2, and telencephalic markers (Neocortex and medial Pallium). However, the PRTG positive sorted subpopulation, compared to the PRTG negative subpopulation, shows higher expression levels of cortical markers including FOXG1, SP8, LHX2 and SIX3, together with lower expression levels of posterior markers such as BARHL1 and GBX2. Additionally, it presents lower expression levels of pluripotency - POU5F1 (OCT4) and NANOG- and neuronal markers- DCX, STMN2 and TUBB3. This indicates that by sorting for PRTG expression we are able to purify our cortical culture by sorting out pluripotent cells and neurons, as well as posterior NSCs which remain in the PRTG negative subpopulation. Unexpectedly, in the early neural induction days, MCAM positive and MCAM negative populations seem to share to a large extent their transcriptional signature meaning that sorting for MCAM at day 12 is insufficient to segregate distinct subpopulations. Hence, when comparing the double sorted populations to the PRTG single sorted populations comes as no surprise that they share their general transcriptional identity.

Remarkably, on day 35 we see that only the PRTG negative, but not the positive population, expresses cortical markers indicating a shift in PRTG expression as a readout from cortical identity to non-cortical identity. A clear dichotomy is drawn on day 35, by PRTG positive cells having posterior identity (exclusively expressing posterior markers such as GBX2, IRX3, PAX3 and EN2), and PRTG negative having cortical identity (exclusively expressing telencephalic markers).

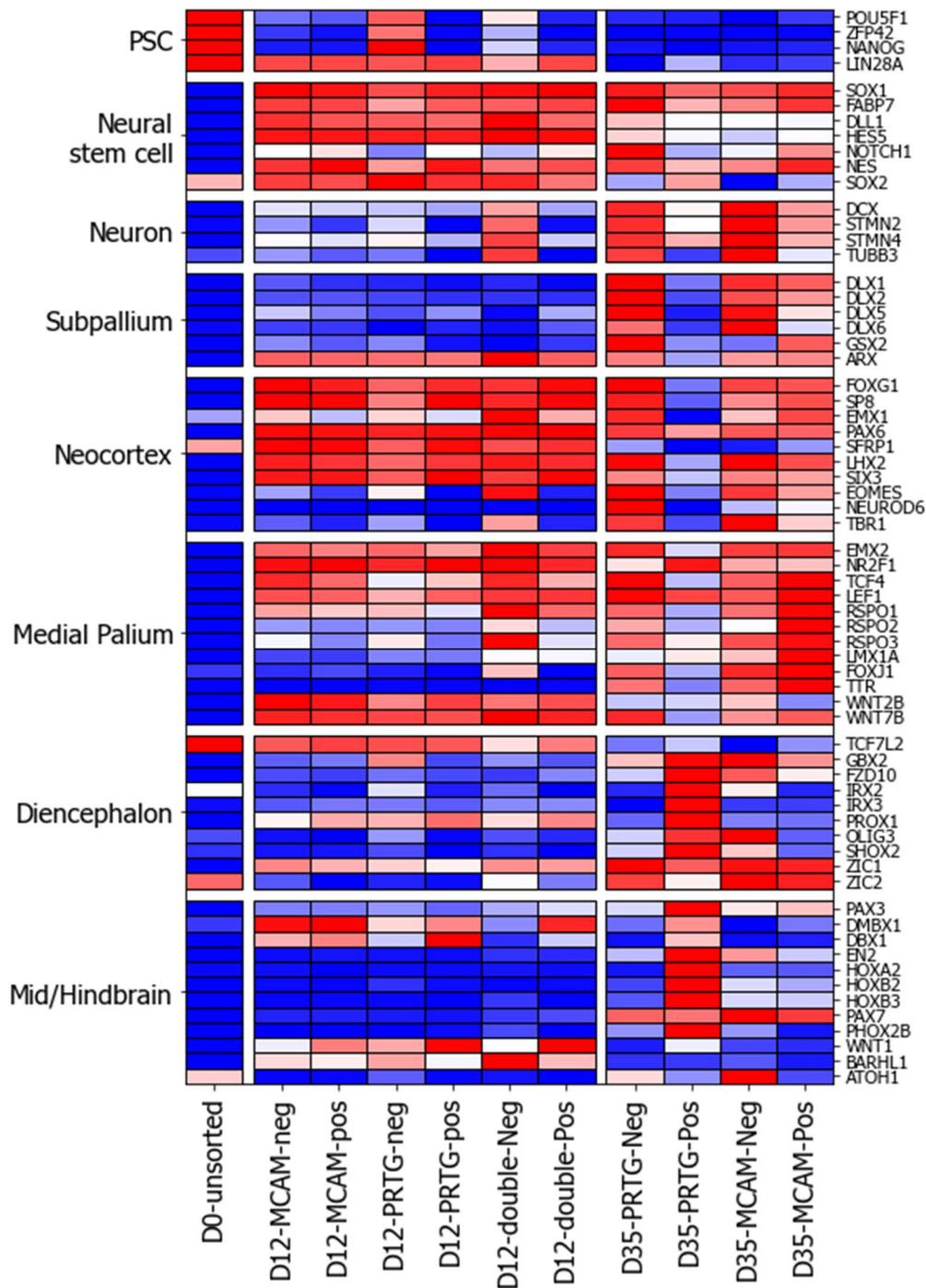


Figure 4.15. RNA-seq analysis identifies specific gene expression profiles based on PRTG sorting. Heatmap representing expression of marker genes for pluripotent, neural stem cells, and differentiated cells, as well as different brain regions (Subpallium, Neocortex, Medial pallium, Diencephalon and Mid/Hindbrain). Color-coded scale represents relative expression levels of each gene (row) across clusters. Note the increased cortical cell identity in day 12 PRTG sorted cells. Surprisingly, at day 35 PRTG detection demarcates non-cortical stem cells, including diencephalic and mid/hindbrain NSCs, instead of cortical lineages.

In summary, while at day 12 PRTG alone efficiently labels NSC populations enriching for cortical NSCs, MCAM does not show a clear segregation of subpopulations based on regional identity. Thus, MCAM does not provide a clear additional enrichment when sorting for both markers concomitantly. Similarly on day 35, PRTG expression demarcates different transcriptional identities whereas MCAM positive and negative sorted cells seem to have a very similar transcriptomic signature. These results indicate that MCAM sorting does not provide any additional enrichment beyond the one provided by PRTG sorting, hence we decided to focus on PRTG as the most promising surface marker and continued towards its validation and characterization.

4.5. Characterization of sorted populations further confirms PRTG validity as an early cortical marker

In order to further strengthen the finding that PRTG is an early cortical NSC marker, we conducted a series of downstream analysis after sorting neural induced cells on day 12. We first have a look at our hiPSCs-derived cortical progenitor cells on day 12 (before sorting) and check how PRTG expression is distributed within the culture (*Figure 4.16*).

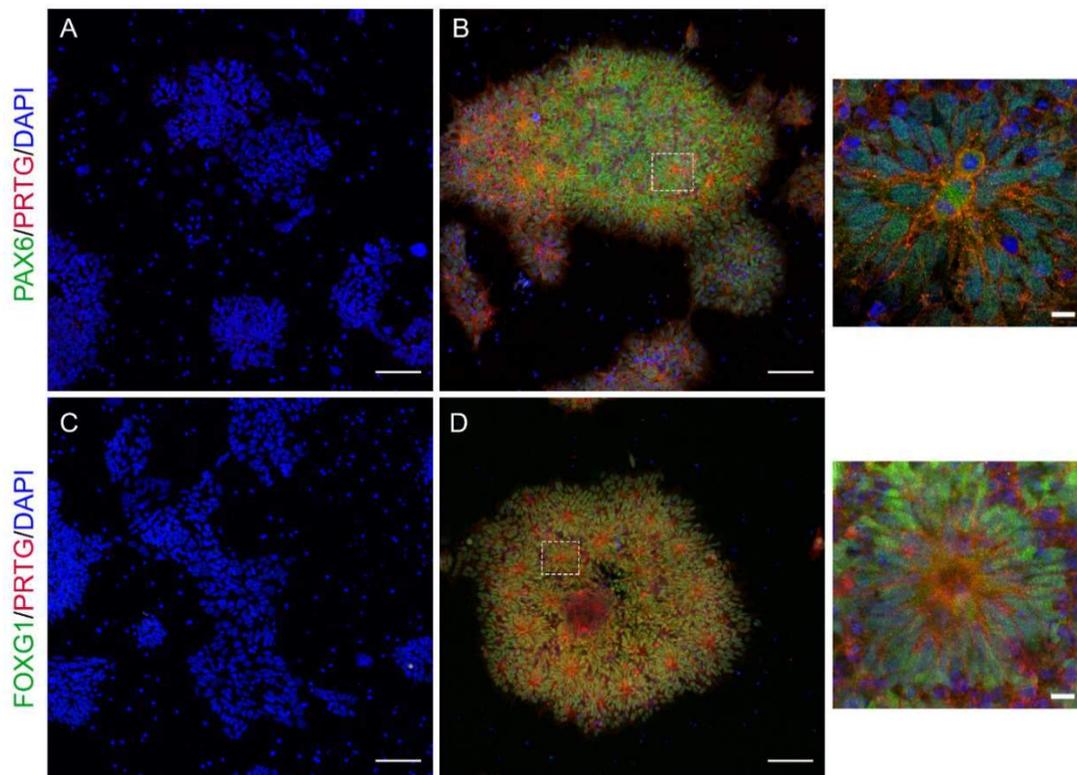


Figure 4.16. PRTG expression is enriched in rosettes co-expressing PAX6 and FOXG1 on day 12. Same neural induced culture on day 12 showing the expression of PRTG together with cortical markers PAX6 (A, B) or FOXG1 (C, D) in areas containing rosettes (B, D) and non-rosettes (A, C). Scale bar= 100 μ m. The right images represent magnified rosette structures from (B) and (D) images (dashed square). Scale bar= 10 μ m.

As previously mentioned, at this early time point of neural induction, the culture is very homogenous and abundant in cortical NSCs which generate neural rosettes expressing high levels of PAX6 and FOXG1 (*Figure 4.16, B and D*). Remarkably, PRTG expression is highly correlated with these structures, being almost exclusively expressed at their apical site (lumen) while areas of the same culture that do not form rosettes due to cells not having cortical identity (PAX6 and FOXG1 negative cells) do not possess PRTG expression (*Figure 4.16, A and C*). These findings provide further evidence that PRTG is a surface marker for early cortical NSCs.

To get initial insights into the properties and potential of PRTG sorted cells, we performed FACS at day 12 and replated the cells for downstream analysis. Following stringent gating for sorting, we collected the PRTG negative subpopulation and only the high PRTG positive cells -accounting for 33.5% of the population (*Figure 4.17, A*)- and subsequently expanded them in culture under the standard culture conditions until confluency was reached.

Notably, PRTG negative cells present low survival rate after replating. We notice that when replating both subpopulations at the same high density, only the positive population visibly attaches to the plate already 4h after sorting. This is a result of low cell viability of the PRTG negative subpopulation which is even more noticeable 18h after sorting, with only few cells attaching to the culture plate (*Figure 4.17, B*). Although being a factor that hinders the culture of this subpopulation for downstream analysis, in itself is indicative of phenotypic differences between the two sorted populations. Since we know, based on the bulk RNA-seq analysis, that the PRTG negative population contains pluripotent cells and neurons, a possible explanation is that viability of the sorted subpopulation was affected by the mere sorting technique or due to the lack of appropriate environmental cues. On one hand, differentiated neurons are known to be more susceptible to sorting procedures due to the possible disruption of their processes.

On the other hand, replating pluripotent cells in neural induction media might prove detrimental for their survival without their standard cellular environment, possibly providing necessary growth factors that should be secreted by neighboring cells.

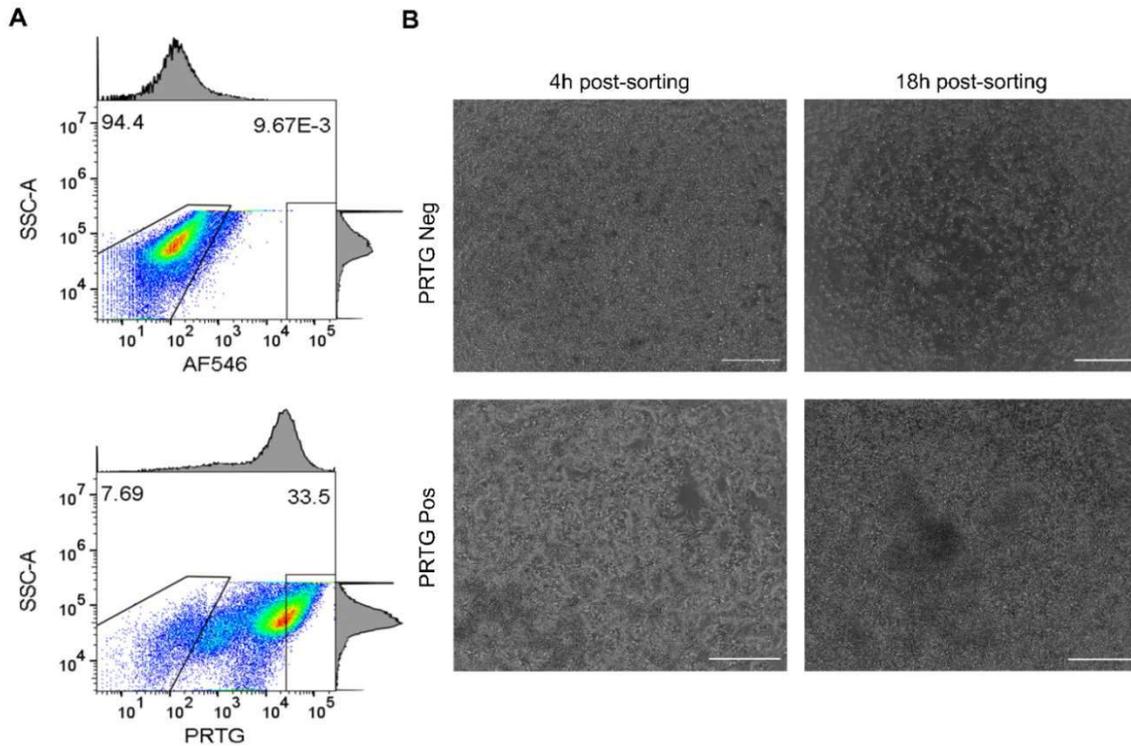


Figure 4.17. PRTG sorting strategy and phenotype of replated cells at day 12 of neural induction. A) FACS density plots showing PRTG expression on day 12 based on SSC-area versus secondary antibody staining (top) or PRTG staining (bottom). Control gates as indicated were set against the corresponding negative controls based on secondary antibody staining background (<1%). The frequency of cells sorted and collected for downstream analysis within each selected gate is shown (percentage). B) Bright-field images of the replated cultures a few hours after sorting. Note the inability of most PRTG negative cells to attach to the culture plate, indicating a high cell death rate of this subpopulation. Scale bar= 200 μ m.

Next, we dissected the cortical identity by means of immunostainings of the sorted subpopulations. Compared with the PRTG negative subpopulation, FACS-purified PRTG positive cells display higher levels of cortical markers, PAX6 and EMX1 (*Figure 4.18, A; and 4.19, A; top*), while simultaneously showing a greater ability to generate

rosettes (*Figure 4.18, C; and 4.19, C*) and displaying lower levels of the neuronal marker DCX (*Figure 4.18, A; bottom*).

Interestingly, at day 18 after replating the positive sorted cells, we observe lower levels of PRTG expression remaining in rosettes cells (lumen) in correlation with decrease of PRTG with time in cortical NSCs at advanced stages. However, a higher expression level of PRTG is found in non-cortical cells lacking EMX1 expression within the PRTG positive sorted subpopulation (*Figure 4.19, A; top right*). Moreover, we show that PRTG fluorescence intensity is anticorrelated with the one of EMX1 (*Figure 4.19, E*), supporting the bulk RNA-seq data from day 35 where cells that are PRTG positive have a non-cortical identity.

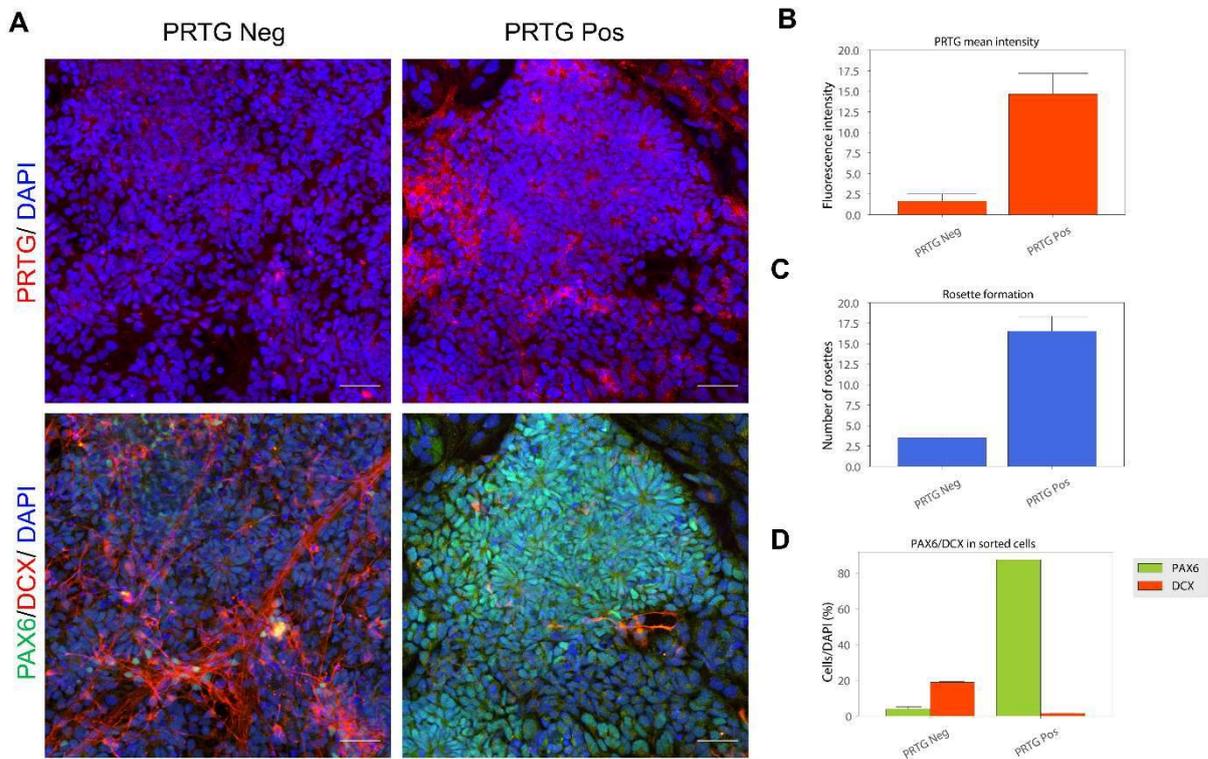


Figure 4.18. PRTG sorting enriches for NSCs populations at expenses of neurons.

A) Immunostaining images of PRTG and PAX6/DCX on day 18 monolayer neural progenitors derived from PRTG sorted populations at day 12. Scale bar= 50 μ m. B) Mean PRTG fluorescence intensity used as a proxy for PRTG expression. C) Quantification of rosette structures per replated population. D) Cell counts of markers in relation to DAPI positive cells. Box and whisker depicting mean and SD, respectively; n=2.

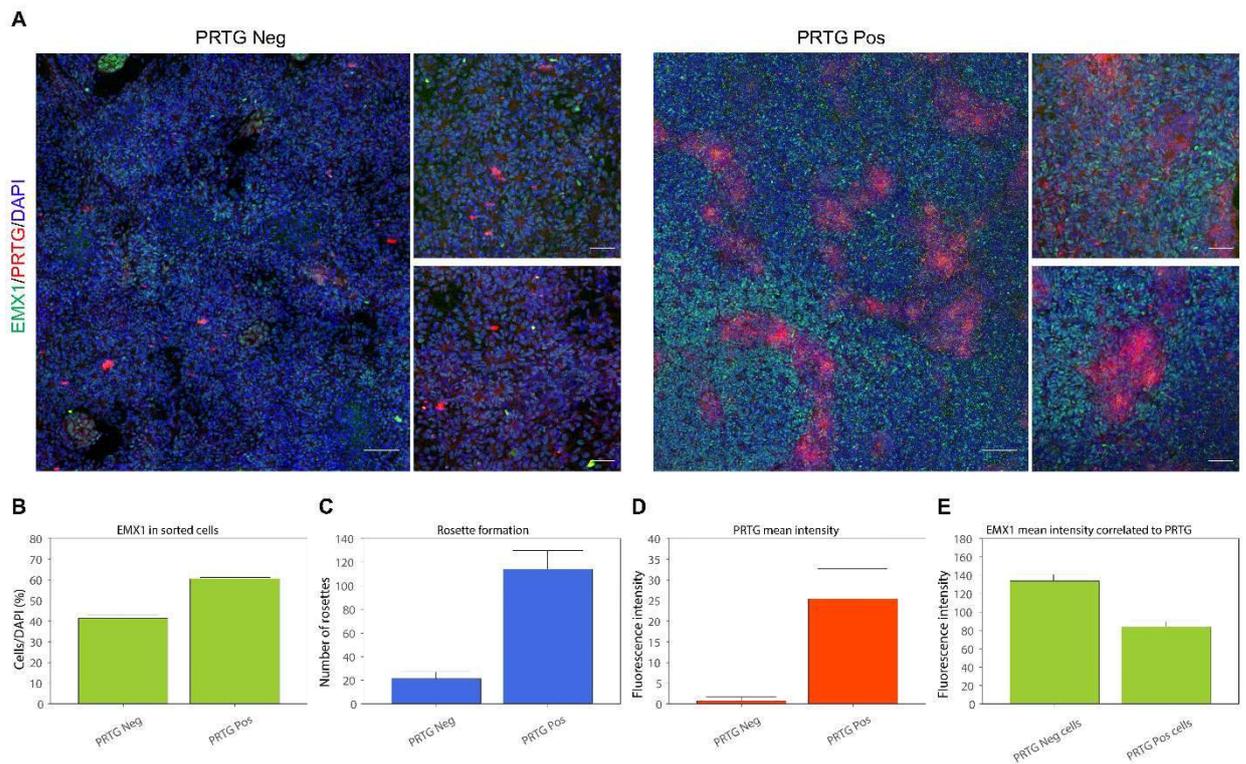


Figure 4.19. PRTG sorting enriches for cortical NSCs. A) Immunostaining images of PRTG and cortical marker EMX1 in day 18 monolayer neural progenitors derived from PRTG sorted populations at day 12. Scale bar= 50 μ m. B) Cell counts of EMX1 positive cells in relation to DAPI positive cells. C) Quantification of rosette structures. D) Mean PRTG fluorescence intensity used as a proxy for PRTG expression. E) Mean EMX1 fluorescence intensity used as a proxy for EMX1 expression in relation with PRTG positive and negative areas found in the PRTG positive sample. Note that areas that have higher PRTG intensity on day 18 have lower levels of EMX1 expression. Box and whisker depicting mean and SD, respectively; n=3.

Sorted cells at day 12 were also cultured in neuron differentiation medium to induce terminal differentiation (for 14 days post FACS), and subsequently stained for FOXP1 and DCX markers. Upon inducing neuronal differentiation of sorted cells, we see a higher differentiation rate in PRTG positive cells exemplified by the higher number of DCX positive cells (*Figure 4.20, A and C*). This supports the idea of the positive subpopulation being mainly composed of NSCs since they have the plasticity to become neurons when instructed. Also, there seem to be morphological differences among the neurons generated by both subpopulations such as shape and length of their axons. This could be indicative of distinct neuronal subtypes being generated, although staining with specific neuronal markers would be needed to be sure. Notably, we see comparable high levels of FOXP1

indicating a general telencephalic identity in both subpopulations (*Figure 4.20, A and C*).

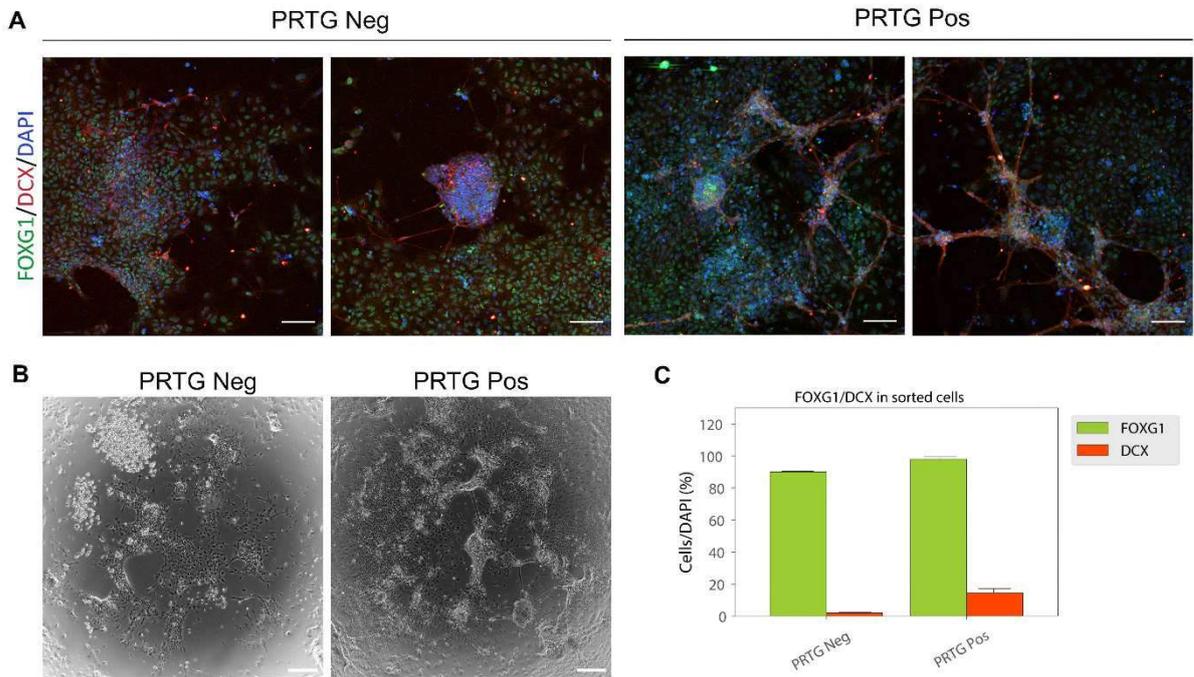


Figure 4.20. PRTG positive cells retain a higher neurogenic potential. A) Immunostaining images of PRTG and FOXG1/DCX in PRTG sorted populations induced towards neuronal differentiation. Note the comparable levels of FOXG1 irrespective of the PRTG sorting but higher DCX expression levels indicating a higher neurogenic potential of the PRTG positive subpopulation. Scale bar= 100 μ m. B) Bright-field images showing clear morphological differences between both cultures. Scale bar= 200 μ m. C) Cell counts of markers in relation to DAPI positive cells. Box and whisker depicting mean and SD, respectively; n=2.

Finally, to provide further validation, we grew the sorted PRTG positive cells until day 35 and re-sort (D35 re-sorted) to compare them to an unsorted culture. Before sorting, we can appreciate that there are more NSC clusters in the pre-sorted culture already indicating higher homogeneity (*Figure 4.21, A*). FACS analysis further shows a shift in PRTG expression levels being lower in the re-sorted population (going from 11.1% PRTG positive cells to 2.18%) (*Figure 4.21, B*), highlighting an enrichment towards cortical NSCs based on our previous knowledge.

Upon further transcriptional analysis of the sorted populations at day 35 by means of RT-qPCR, we see a higher expression of the pan-telencephalic marker OTX2 as well as

cortical markers PAX6, FOXG1 and SIX3 in the re-sorted PRTG negative cells versus the positive (*Figure 4.21, C*). It is important to address that we detect lower expression levels of these markers when we compare the re-sorted PRTG positive subpopulation with the D35 PRTG positive (unsorted at day 12). However, we find that expression of the posterior marker GBX2 is completely absent in the re-sorted culture (both in positive and negative subpopulations at day 35), supporting the idea of higher purity of cortical identity in the pre-sorted culture.

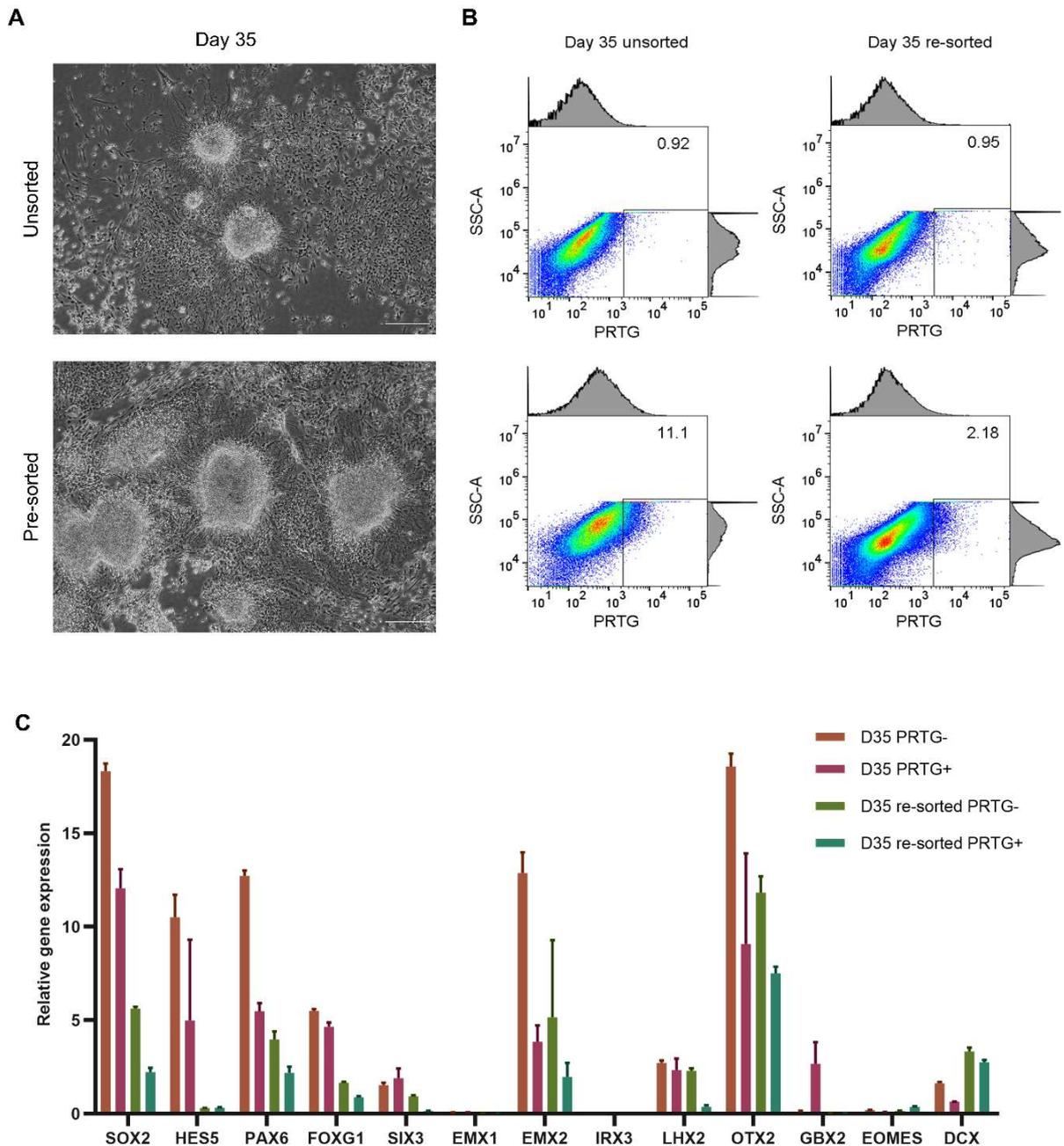


Figure 4.21. Re-sorting of replated cells at day 35 confirms prospective isolation and enrichment of cortical lineages by means of PRTG sorting at day 12. A) Bright-field images of unsorted and PRTG positive cells (pre-sorted at day 12) before collecting for sorting at day 35. Scale bar= 200 μm . B) FACS density plots showing the PRTG expression pattern of the two cultures on day 35, based on SSC-area versus secondary antibody staining (top) or PRTG staining (bottom). The frequency of cells within the positive gates is shown (percentage). Note that for downstream RT-qPCR analysis we sorted and collected 15% of the negative population and 100% of the positive populations. C) Evaluation of expression levels of the re-sorted population compared to the unsorted cells on day 35 by means of RT-qPCR.

Taken together, our results show that sorting for PRTG high positive cells at day 12 allows for the prospective isolation of early cortical NSCs while removing unwanted cell lineages, including posterior NSCs, pluripotent, and more differentiated cells, thus increasing the purity of cortical fates in culture.

4.6. PRTG expression emerges after four days of cortical neural induction

Given that PRTG expression peaks at day 12, distinguishing a more cortical NSC population, we were interested to know at what time point PRTG starts being expressed to better understand if its emergence correlates with early cell type specification of cortical lineages. In order to address this question, we conducted FACS analysis at various early time points after neural induction. We began by examining the PRTG expression pattern from day 2 to day 10 (here shown 4-6, *Figure 4.22*). We see that expression of PRTG commences at day 4, accounting for 16.9% of the total population. By day 5, PRTG positive cells present in culture double in numbers becoming a clear PRTG positive subpopulation, being 39.2% of the total population. And it keeps increasing being 56.1% of the total population by day 6.

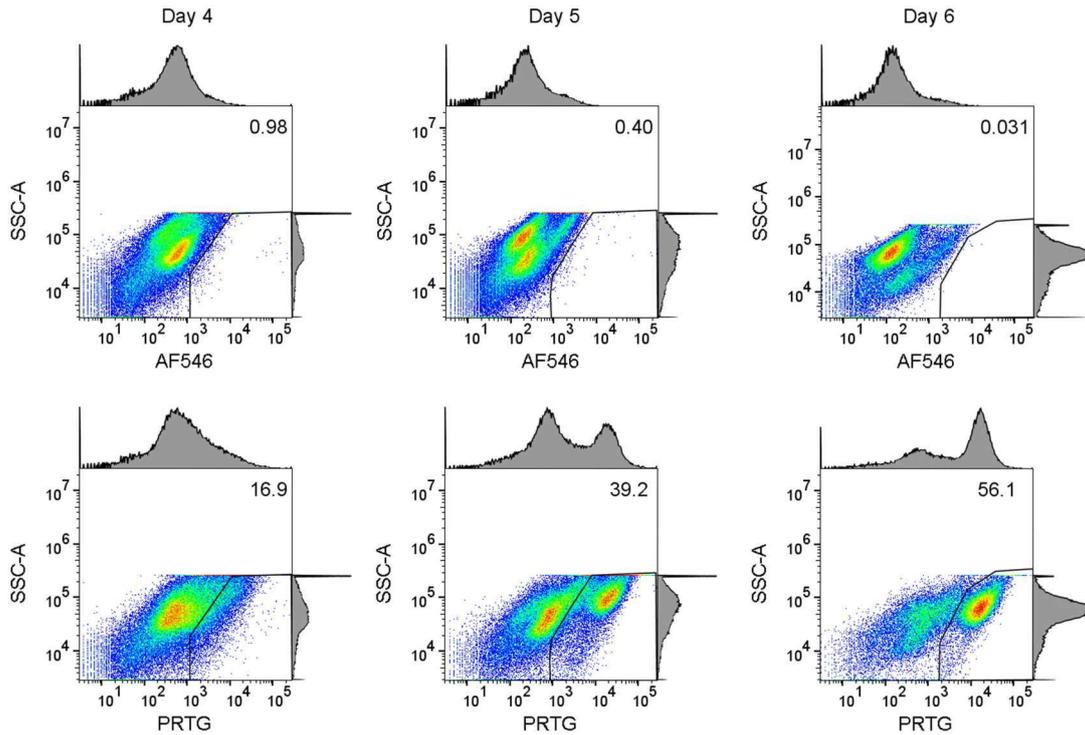


Figure 4.22. PRTG expression analysis in early days of neural induction. Representative FACS analysis of PRTG expression in hiPSCs-derived neural stem cells at early stages of differentiation. Samples incubated with the secondary antibody were used as negative controls to set the appropriate negative gates (<1%) (top). Scatter plots showing PRTG cell surface expression and the frequency of the positive populations (bottom).

4.7. Sorting for PRTG at day 5 enriches for anterior telencephalic identity

In order to investigate whether PRTG expression relates to the first signs of telencephalic specification in our culture, we proceeded to FACS our neural induced cells at day 5. Following stringent gating for sorting on day 5, we took the whole negative subpopulation (12.5% of the total population) and the 12.6% highest PRTG positive cells (*Figure 4.23, A*). Collected cells were subsequently expanded under the standard culture conditions until reaching confluency, and some were acute fixated a few hours after sorting.

Upon replating the cells after sorting, we found impaired cell viability of the PRTG negative cells, very similar to the phenotype seen on day 12 (*Figure 4.23, B*). Again, this could be explained by the lack of paracrine cell communication required for the survival of these cells, hinting to the possible identity of the PRTG negative subpopulation.

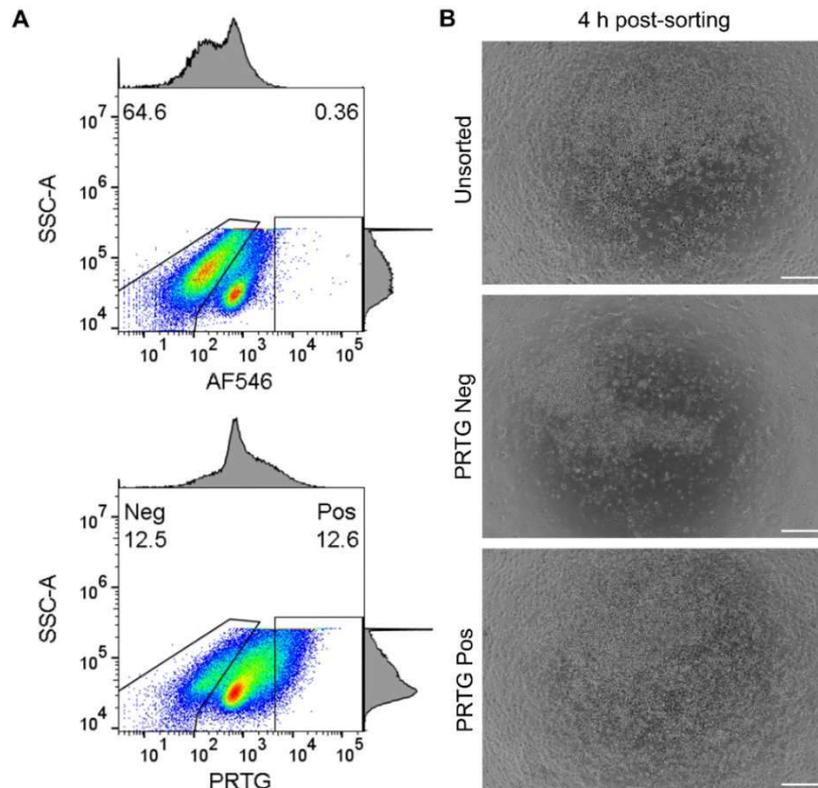


Figure 4.23. PRTG sorting strategy and phenotype of replated cells at day 5 of neural induction. A) FACS density plots showing PRTG expression on day 5 based on SSC-area versus secondary antibody staining (top) or PRTG staining (bottom). Control gates as indicated were set against the corresponding negative controls based on secondary antibody staining background (<1%). The frequency of cells sorted and collected for downstream analysis within each selected gate is shown (percentage). B) Bright-field images of the replated cultures, together with the unsorted control, a few hours after sorting. Note the high cell death rate of the PRTG negative subpopulation similar to what we see on day 12. Scale bar= 200 μm .

By sorting for PRTG at these early stages of neural induction, we can already see differences in the sorted subpopulations upon replating them and staining them with a set of ectodermal lineage markers. Therefore, these results indicate that we can isolate distinct subpopulations present in our culture based solely on PRTG expression.

Compared with the PRTG negative subpopulation, FACS-purified PRTG positive cells display higher levels of the anterior marker OTX2 and lower levels of the posterior marker GBX2 (*Figure 4.24, A*). Upon comparing both sorted subpopulations with the unsorted population and quantifying OTX2- and GBX2-expressing cells (*Figure 4.24, B*), we

clearly see that GBX2 positive cells present in the unsorted culture are enriched in the negative subpopulation, whereas OTX2 positive cells are enriched in the positive subpopulation. These data show that PRTG expression on day 5 identifies those cells undergoing anterior neural plate specification.

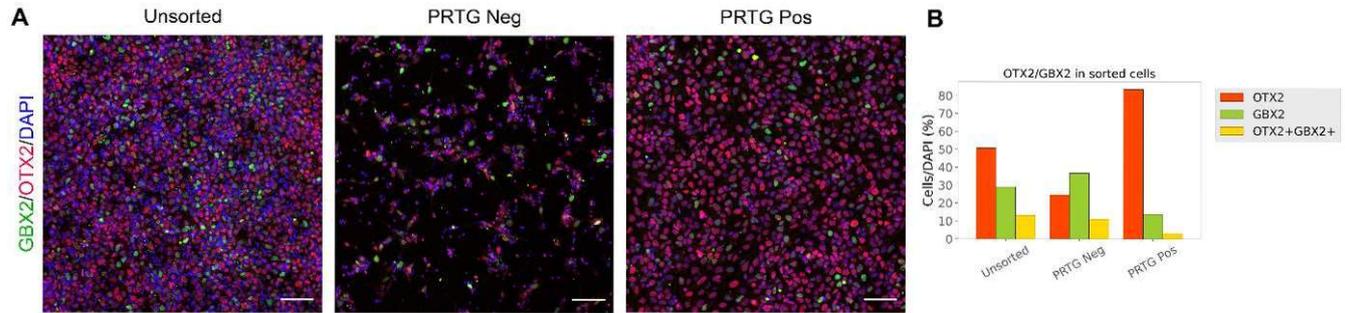


Figure 4.24. Sorting for PRTG at day 5 enriches for anterior neural identity. A) Immunostaining images of sorted populations at day 5 for OTX2 and GBX2. Scale bar= 50 μ m. B) Cell counts and co-localization analysis of markers are presented as a bar plot panel (n=1).

We next sought to investigate whether we identify segregation of other lineages that are specified in early stages of differentiation within our culture. Thus, we stained the day 5 sorted subpopulations with the pan-neural marker SOX2, the dorsal telencephalic marker PAX6, and the TFAP2A marker which is highly expressed in neural crest and non-neural ectoderm. Again, we see a mixed population in the unsorted culture comprised by cells expressing all three markers, being SOX2 and PAX6 more prominent, as expected. FACS-purified PRTG positive cells show equally high levels of SOX2 and PAX6 positive cells but almost complete lack of TFAP2A positive cells. Oppositely, all cells found in the PRTG negative subpopulation are triple positive, compatible with cranial placode identity (*Figure 4.25*). These results indicate that based on PRTG expression we are able to segregate neural ectodermal cells from placodal ectodermal cells, additionally purifying our culture from unwanted lineages.

Taken together, these immunostainings suggest that PRTG expression within early stages of neural induction marks those cells that have been already specified to anterior neural ectoderm. Therefore, FACS-purification of PRTG positive cells allows for an enrichment of anterior CNS identity providing a more homogeneous culture.

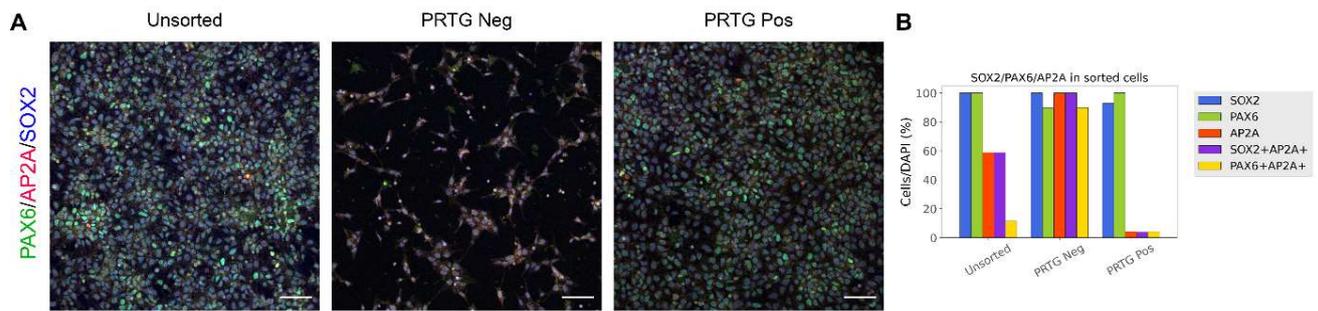


Figure 4.25. Sorting for PRTG at day 5 enriches for anterior forebrain. A) Immunostaining images of sorted populations at day 5 for PAX6 and TFAP2A. Scale bar= 50 μ m. B) Cell counts and co-localization analysis of markers are presented as a bar plot panel (n=1).

Following 10 days in culture, after replating the cells and allowing them to grow under the standard neural induction media, we can immediately observe morphological differences between populations. The most striking one is the ability of the PRTG positive subpopulation to generate rosettes compared to the unsorted population. As seen in *Figure 4.26*, the faster appearance of rosettes in the PRTG positive culture is either indicative of an accelerated or a more efficient neural induction, which could be explained by having a more homogenous cortical NSC identity within the culture.

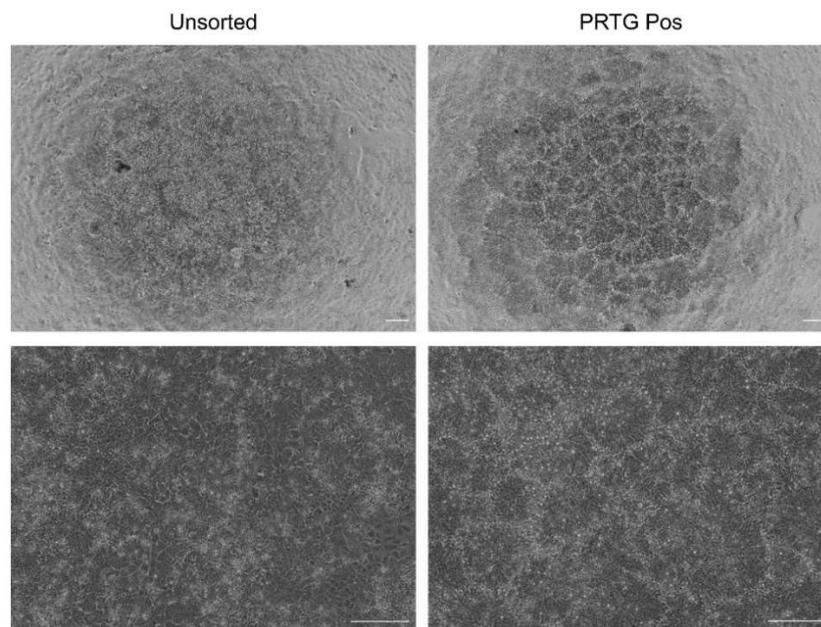


Figure 4.26. PRTG sorting at day 5 enriches for rosette forming cells. Bright-field images of the day 5 sorted replated cultures grown until day 10. At the bottom magnification

of the top culture to better appreciate the distinct cell morphology of PRTG positive sorted cells compared to the unsorted control. Note that neural rosette structures are less frequently observed in the unsorted population compared to the PRTG positive sorted subpopulation. Scale bar= 100 μ m.

To further confirm cell identity, we stained for cortical markers together with PRTG. Not surprisingly, we find a higher PRTG intensity in the positive subpopulation which beautifully marks the apical site of rosettes. Clusters of cells co-expressing PAX6 and FOXG1, mainly in rosette structures, were observed in both cultures. Interestingly, PRTG positive cells exhibit increased rosette homogeneity compared to unsorted cells, the latter underscoring for higher heterogeneity within forebrain identity of our standard culture. Our results show that sorting for PRTG high positive cells enriches for a more homogeneous cortical population shown by the higher PAX6 and FOXG1 expression correlating with radial organization (rosettes) (*Figure 4.27, A*).

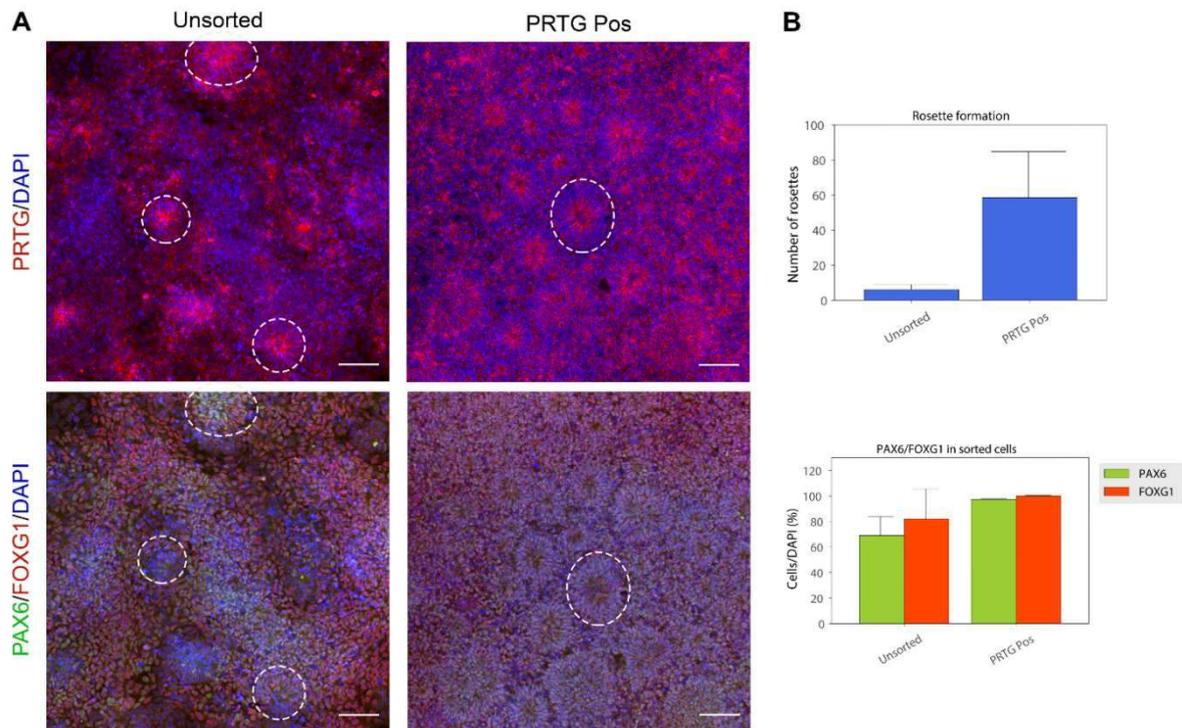


Figure 4.27. PRTG sorting at day 5 enriches for cortical identity. A) Immunostaining images of PRTG and PAX6/FOXG1 on day 10 monolayer neural progenitors derived from PRTG sorted populations at day 5. Note the higher presence of

rosette structures (examples marked by dashed circles) in the PRTG positive subpopulation compared to the unsorted. Scale bar= 100 μ m. B) Quantification of rosette structures per replated population. C) Cell counts of markers in relation to DAPI positive cells. Box and whisker depicting mean and SD, respectively; n=2.

4.8. PRTG demarcates early specification of anterior identity and prospectively isolates distinct telencephalic subpopulations

Given that PRTG expression clearly defines distinct lineages within our culture, we wanted to further characterize the sorted populations in order address whether we can prospectively isolate our population of interest.

Hence, three subpopulations were sorted at day 5, including the PRTG negative, the PRTG high positive, and the PRTG low positive (referred as middle subpopulation). We thought it would be interesting to also sort the middle population to decipher if it represents a transition state from negative to positive, a mixed population or simply a completely different third subpopulation within our culture. Stringent gating was performed in order to avoid contamination of subpopulations. For the PRTG high positive, cells at 24.1% of the highest fluorescent intensity were sorted. Each population was collected for scRNA-seq after sorting. Simultaneously, sorted cells were replated into wells for further culture under standard conditions, and collected on day 10 for scRNA-seq analysis. Given that most cells in unsorted cultures on day 10-12 are composed of cortical NSCs, we tested here whether non-expressors, early mid expressors and early high expressors of PRTG on day 5 would develop different fates when grown separately under the same conditions.

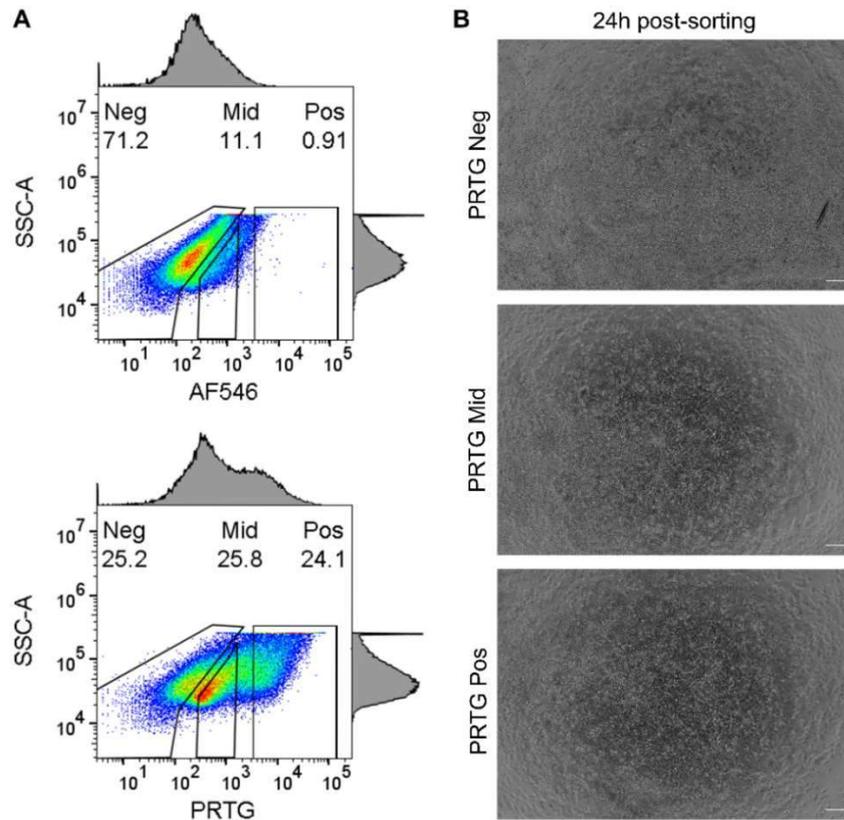


Figure 4.28. Sorting strategy at day 5 for scRNA-seq analysis. A) FACS density plots showing the gating strategy used to collect the three subpopulations for scRNA-seq based on SSC-area versus secondary antibody staining (top) or PRTG staining (bottom). Control gates as indicated were set against the corresponding negative controls based on secondary antibody staining background (<1%). The frequency of cells sorted and collected for downstream analysis within each selected gate is shown (percentage). B) Bright-field images of the replated cultures a few hours after sorting. Scale bar= 100 µm.

For each day, 5 and 10, we integrated the three subpopulations- positive (P), middle (M) and negative (N)- and we clustered the cells based on the expression of well-known marker genes as previously done.

On day 5, we observe three main clusters based on cell identity: i) early NSCs, marked by expression the of SOX2, NES and GLI3; ii) PSCs, mainly marked by POU5F1 (OCT4) and high SOX2 and IRX2; iii) early neurons, expressing STMN2 and DCX (*Figure 4.29, C*).

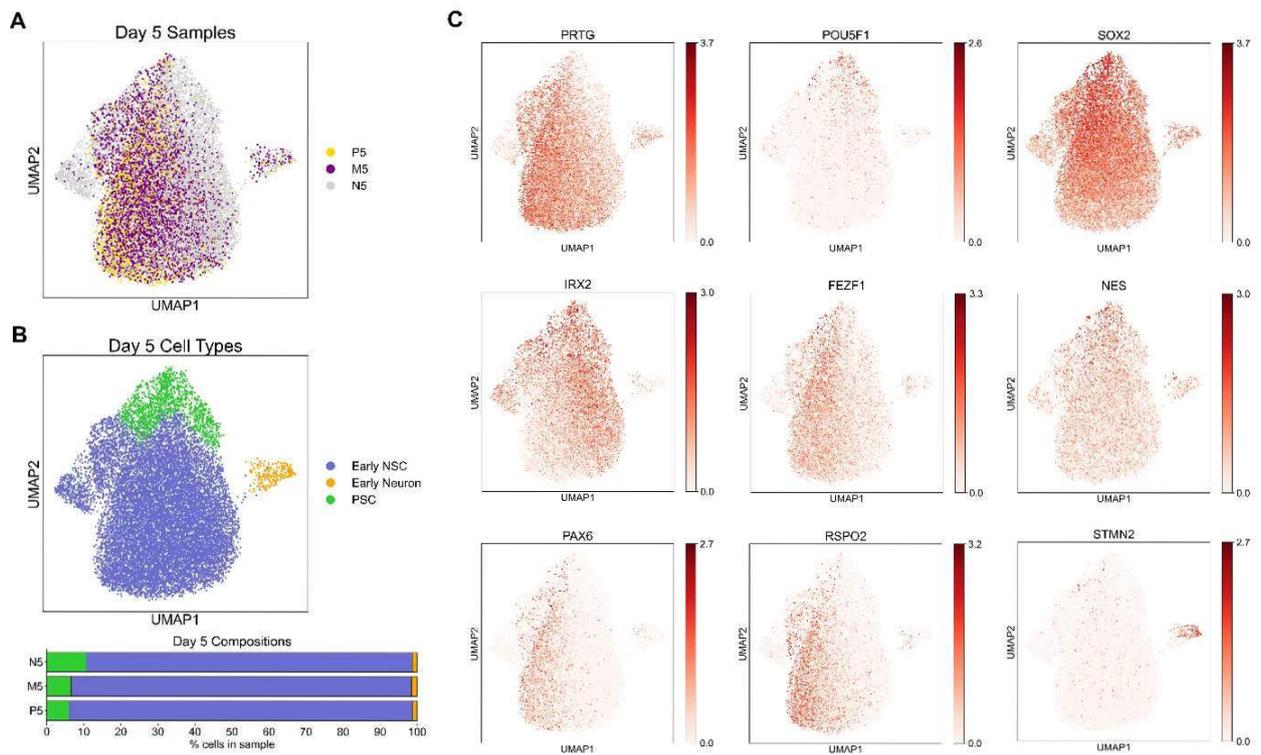


Figure 4.29. PRTG demarcates early specification of anterior neural identity. A) UMAP plot of the merged day 5 scRNA-seq datasets showing the clustering of the different subpopulations based on PRTG sorting. B) UMAP plot of the merged day 5 scRNA-seq data with the distinct clusters based on cell types highlighted in different colors (up) and stacked plot of cell composition depicting the percentage of cell types within the different subpopulations (left). C) UMAP plots depicting expression of selected marker genes which were used to annotate the scRNA-seq clusters.

These three clusters are formed by cells coming from the three sorted populations discarding a clear enrichment of lineages based on PRTG sorting. However, there are differences in proportions making the positive population the most homogeneous one among them, and being the negative the most heterogeneous. A total of 93% of cells in the positive-sorted population are early NSCs while in the negative-sorted sample they only account for 88% of the total population. Only 6% and 1% of the cells in the positive-sorted population is assigned to PSCs and neurons, respectively (*Figure 4.29, B*), whereas the negative population contains 11% of PSCs. In between, we find the middle population which is very similar to the PRTG high positive population, assumingly representing cells in transition.

As already mentioned, this early stage is characterized by high cell homogeneity given that 91.1% of the total amount cells (averaging all populations) are early NSCs. However, we detect some signs of regional specification such as the specific expression of PAX6 (and RSPO2) in the positive population (*Figure 4.29, C*), in agreement with what we have previously observed in our immunostainings.

In order to gain a better insight into the transcriptional identity of these early NSCs and to address whether we can detect early specification based on PRTG expression, we conducted differential gene expression analysis on pairwise comparisons of the different early NSCs populations i.e., P5 versus N5 (*Figure 4.30, A*), P5 versus M5 (*Figure 4.30, B*), and M5 versus N5 (*Figure 4.30, C*).

As expected, PRTG is enriched in the positive population compared to the negative population, and it is also highly expressed in the middle one. A striking difference is the exclusive expression of CDH1 (E-cadherin) in the negative population versus CDH2 (N-cadherin) in the positive and middle. It is known that, during neural induction, cells of the neural tube lose CDH1 and acquire CDH2, hinting at the progression from PRTG negative neuroectodermal cells towards PRTG positive neuroepithelial cells. Neuroepithelial cells - and their progeny, the radial glial cells - are also characterized by the expression of VIM and PAX6, both of which are enriched in the positive and middle populations. Additionally, we detect enrichment of other forebrain markers including FEZF1, LRP2, RSPO2 and SOX5- in the positive and middle populations highlighting their early specification towards forebrain identity. However, at this stage we are not able to detect FOXG1 expression yet.

Beyond the differential expression of marker genes, we also see high variability in signaling molecules such as WNT5B and guidance molecules like SEMA3A, which can contribute to specification of regional identity. While WNT5B has been reported to have a restricted expression within the telencephalon together with other WNTs (Harrison-Uy & Pleasure, 2012; Quinlan et al., 2009), SEMA3A is also found to be expressed in the developing neocortex, specifically within the upper layers (Chen et al., 2008; Polleux et al., 2000). Interestingly, both molecules are highly differentially expressed in a gradual manner, from positive high to middle low and absent in the negative population.

This analysis reveals that expression of PRTG at day 5 (P5 and M5) demarcates those cells that have already transitioned from neuroectoderm (N5) to neuroepithelial (or early radial glial) cells (P5, and to lesser extent M5) and that have already acquired an early regional identity towards forebrain lineages (P5, and to lesser extent M5).

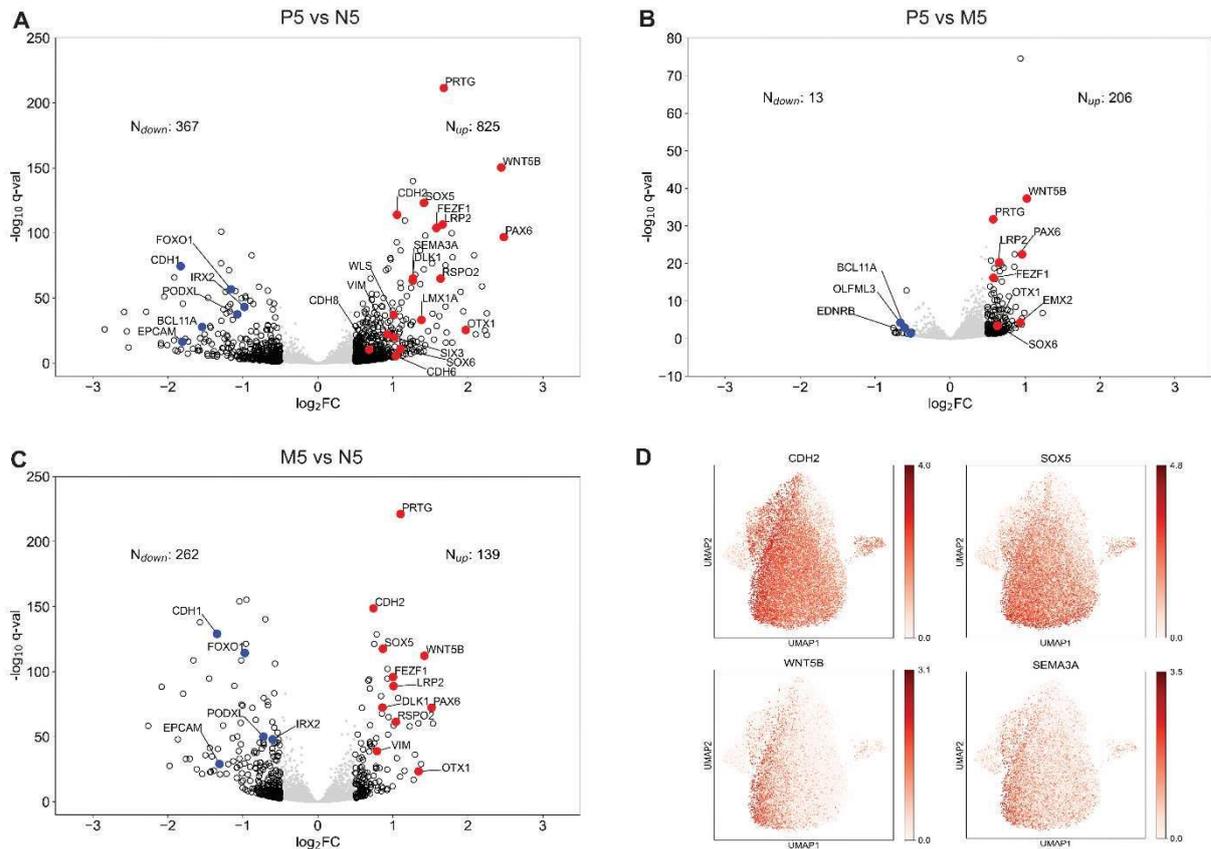


Figure 4.30. Differential gene expression analysis of day 5 PRTG sorted subpopulations. A) Volcano plot depicting the differential gene expression analysis comparing early NSCs found in the positive subpopulation versus the middle population. B) Volcano plot depicting the differential gene expression analysis comparing early NSCs found in the positive subpopulation versus the negative population. C) Volcano plot depicting the differential gene expression analysis comparing early NSCs found in the middle subpopulation versus the negative population. The \log_2 -transformed fold-changes and adjusted P values from a t -test with overestimation of variance after Benjamini–Hochberg correction ($-\log_{10}(Q\text{-value})$) are plotted on the x- and y-axis, respectively. Total number of significantly differentially expressed genes are shown. Upregulated genes of interest are shown in red and the downregulated ones in blue. D) UMAP plots depicting expression of selected differentially expressed genes of interest.

When analyzing day 10, we observe six distinct clusters based on cell identity (*Figure 4.31, C*): i) Telencephalic NSCs, marked by the expression of SOX2, NES and GLI3; ii) Mid/Hindbrain NSCs, mainly expressing IRX2; iii) PSCs, marked by POU5F1 (OCT4), and high SOX2 and IRX2; iv) early neurons expressing STMN2 and DCX; v) Cortical Hem cells specifically expressing WNT2B; vi) Neural crest, marked by the expression of SOX10. At this stage, there is increased heterogeneity but it is important to point out that the middle and the positive populations are more homogenous than the negative. While the negative population only contains 84.3% of telencephalic NSCs, the positive contains 97% and the middle contains 98.5%. Remarkably, the middle population and the positive population lack cells assigned to Cortical Hem or PSC identity. Additionally, the middle population lacks cells with Neural crest identity (*Figure 4.31, B*). These observations indicate that cultures derived from PRTG expressing cells on day 5 become purer in telencephalic lineages.

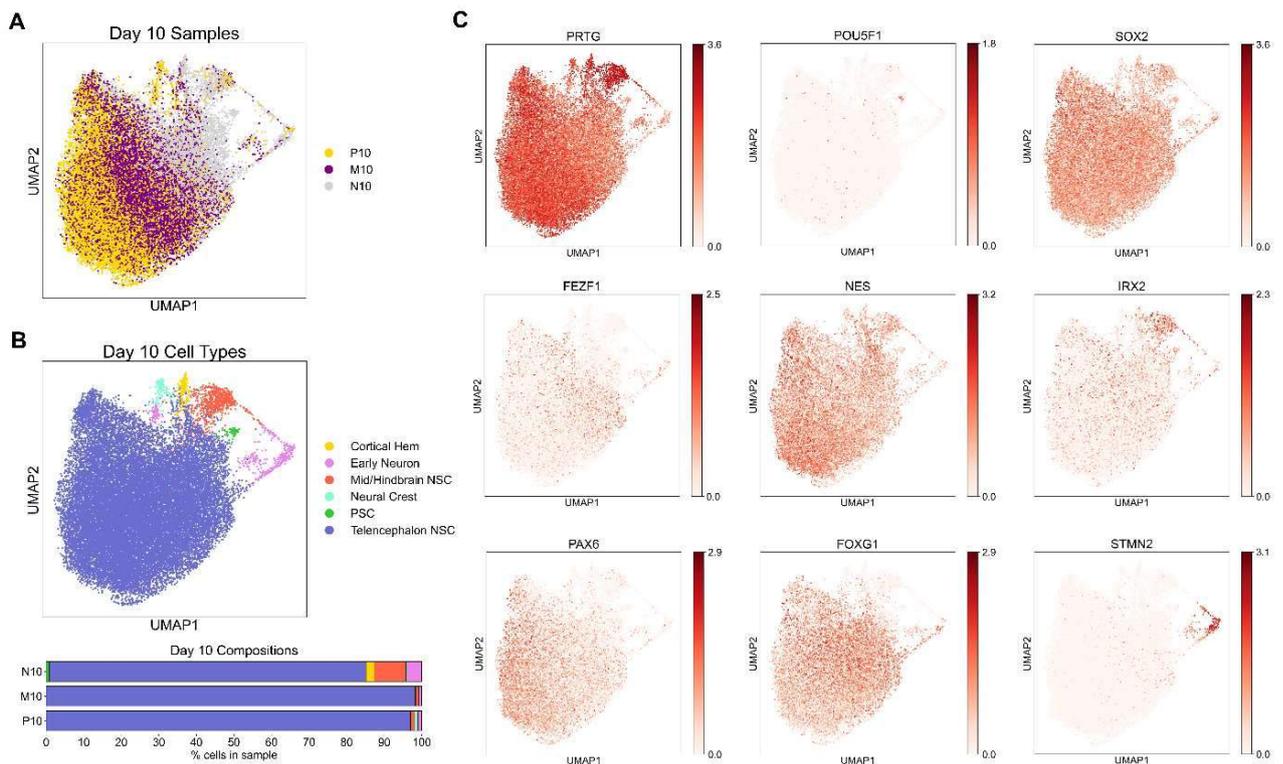


Figure 4.31. PRTG prospectively isolates distinct telencephalic subpopulations.

A) UMAP plot of the merged day 10 scRNA-seq datasets showing the clustering of the different subpopulations based on PRTG sorting. B) UMAP plot of the merged day 10 scRNA-seq data with the distinct clusters based on cell types highlighted in different colors

(up) and stacked plot of cell composition depicting the percentage of cell types within the different subpopulations (left). C) UMAP plots depicting expression of selected marker genes which were used to annotate the scRNA-seq clusters.

We were surprised to discover that even the PRTG negative-derived population acquires PRTG expression by day 10 (*Figure 4.31, C*), although it remains higher in the positive and middle populations (*Figure 4.31, A and C*). This implies that also cells that are negative for PRTG on day 5 are in fact presumptive telencephalic cells. Indeed, at this stage, we are able to detect FOXP1 expression in all three populations. Interestingly, we also detect higher expression of PAX6 in the positive population, reminiscent of day 5 expression pattern (*Figure 4.31, C*). This population also presents specific expression of Clorf61 which marks a subset of early cortical progenitors in the developing brain (*Figure 4.32, D*).

As we did for day 5, in order to gain a better insight into the transcriptional identity of these subpopulations, we conducted differential gene expression analysis on pairwise comparisons of the different telencephalic NSCs populations i.e., P10 versus N10 (*Figure 4.32, A*), P10 versus M10 (*Figure 4.32, B*), and M10 versus N10 (*Figure 4.32, C*).

Remarkably, this analysis allowed us to identify unique expression patterns associated with cortical areal specification. On one hand, within the positive population we find high expression of NR2F1 and EMX2 together with PAX6 indicating dorso-medial identity (*Figure 4.32, A and B*). On the other hand, in the middle population we find exclusive expression of SP8 and SOX6, known to be expressed in a slight dorsal-high ventral-low gradient (Azim et al., 2009). Combined with higher expression of FZD5 and SEMA5A assigns the middle population to a rostro-dorsal identity (*Figure 4.32, B and C*) (Carulli et al., 2021; Fischer et al., 2007). Finally, the negative population seems to represent the rostro-ventral identity marked by its unique expression of the ventral marker NKX2-1 together with anterior rostral marker SIX3. This is further supported by the higher expression of FZD8 which is known to be ventrally expressed in the telencephalon (*Figure 4.32, A and C*) (Fischer et al., 2007).

To conclude, our results show that sorting for PRTG at early stages allows to prospectively isolate distinct cortical fates, thus promoting high homogeneity in our culture. We report that within the telencephalon, we are able to derive highly pure NSCs corresponding to rostro-dorsal, dorso-medial and rostro-ventral cortical areas based solely on PRTG expression levels.

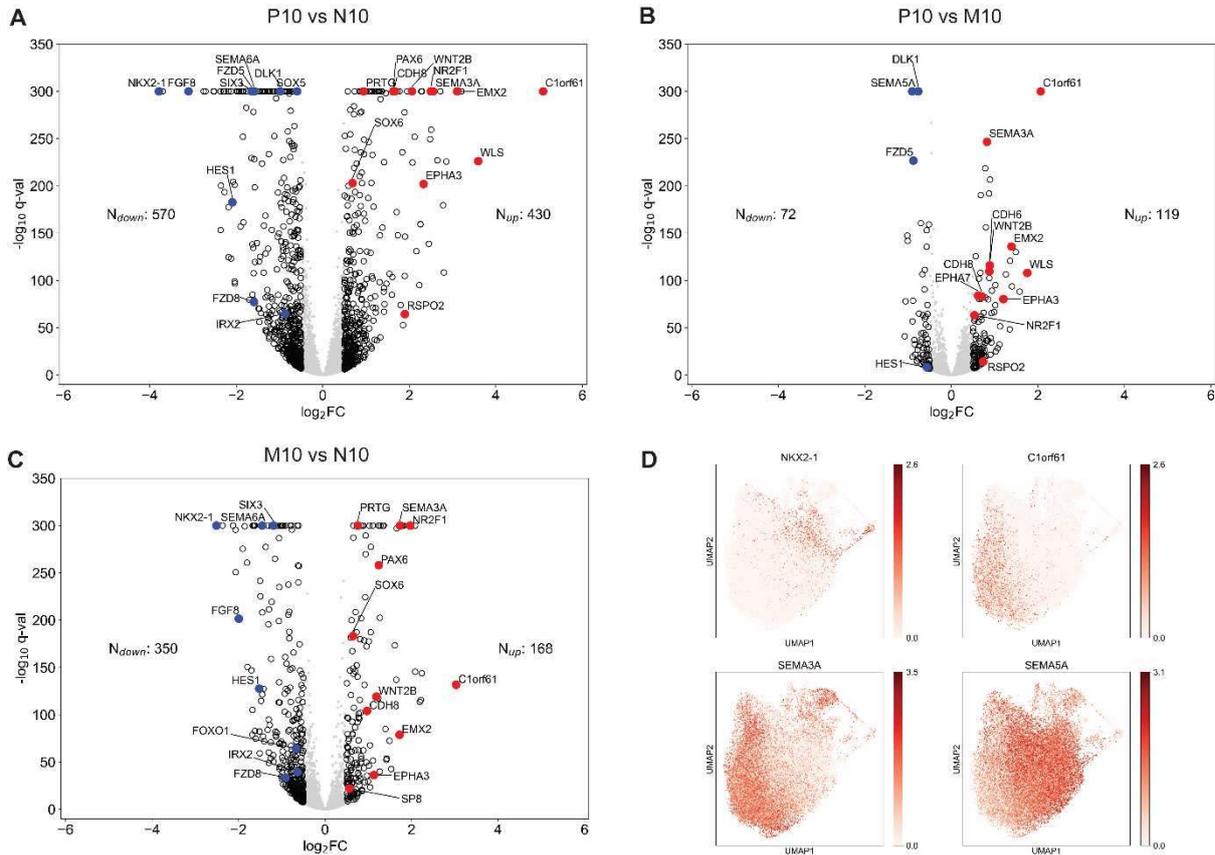


Figure 4.32. Differential gene expression analysis of day 10 prospectively sorted subpopulations. A) Volcano plot depicting the differential gene expression analysis comparing all telencephalic NSCs found in the positive subpopulation versus the middle population. B) Volcano plot depicting the differential gene expression analysis comparing all telencephalic NSCs found in the positive subpopulation versus the negative population. C) Volcano plot depicting the differential gene expression analysis comparing all telencephalic NSCs found in the middle subpopulation versus the negative population. The \log_2 -transformed fold-changes and adjusted P values from a t -test with overestimation of variance after Benjamini–Hochberg correction ($-\log_{10}(Q\text{-value})$) are plotted on the x- and y-axis, respectively. Total number of significantly differentially expressed genes are shown. Upregulated genes of interest are shown in red and the downregulated ones in blue. D) UMAP plots depicting expression of selected differentially expressed genes of interest.

5. Discussion

Stem cell-derived neural cultures hold great promise as a tool for functional studies to understand *in vivo* cortical brain development. Even more importantly, the development of methods to reprogram adult somatic cells to generate pluripotent cells (Takahashi et al., 2007; Yu et al., 2007) makes possible to generate patient-derived hiPSCs that provide the platform for *in vitro* disease modelling and drug discovery, as well as therapeutic advents to tackle a wide range of neurological diseases (Lindvall & Kokaia, 2010; Park et al., 2008). In the past years, research on stem cells has expanded greatly and there have been numerous efforts in order to optimize protocols for the generation and differentiation of cortical progenitors and their neuronal progeny (Bibel et al., 2004; Hansen et al., 2011; Hu & Zhang, 2010; Mariani et al., 2012; Saurat et al., 2016; Shi, Kirwan, & Livesey, 2012; Shi, Kirwan, Smith, et al., 2012; Ying et al., 2003).

During mammalian corticogenesis, a wide diversity of NSCs orchestrate the development and the organization of the cortex. Initially, there is the expansion of the NSC pool through proliferative symmetric divisions, and later through differentiative asymmetric divisions they give rise to the diverse cell populations that reside within the cortical layers (Miyata et al., 2004; Noctor et al., 2004). Throughout this process, NSCs undergo extensive modifications in their transcriptomic profile and chromatin landscape contributing to the formation of heterogeneous progenitor populations. These NSC subtypes are more restricted in their differentiation capacity, and thus more limited in the types of neurons they can generate. In our lab, such NSC populations were derived and studied to dissect the differentiation process from neuroepithelial cells towards the diverse cortical cell types, recapitulating *in vivo* development (Edri et al., 2015; Elkabetz et al., 2008; Ziller et al., 2015). Such studies provided the first glimpse into cell-fate decisions and specification during the ontogeny of *in vitro* derived cortical neural stem cells. However, the regulatory mechanisms that orchestrate the stage-specific differentiation process remain poorly understood.

More recently, efforts to study the development and function of the human cerebral cortex in health and disease have promoted the establishment and optimization of multiple protocols, especially involving 3D systems, for better mimicking *in vivo* development (Camp et al., 2015; Lancaster & Knoblich, 2014; Pasca et al., 2015; Quadrato et al., 2017; Velasco et al., 2019). Additionally, an improved *in vitro* model has been established by the lab relying on Triple-inhibition (dual SMAD-i and WNT-i) to promote cortical fates in a higher yield (Rosebrock et al., 2022). However, given the array of high diversity of neuronal cell types being generated during cortical development, one of the biggest challenges in neural differentiation is to reliably generate specific neuronal cell types which is a prerequisite for cell-based therapeutical approaches. Thus, differentiation protocols are continuously being developed and improved to work towards high pure cortical NSC cultures with a specific and limited differentiation potential. One key aspect to achieve this is to understand the variability in NSC populations during cortical development and the mechanisms driving the transition through these different NSC subtypes. This knowledge will allow to generate a homogeneous and unlimited culture of the desired early-onset NSC subtype ('the founder' NSC population).

To this end, the main aim of my project was to develop a strategy to isolate the early cortical NSC population, derived under our differentiation paradigm, for its characterization and potential manipulation *in vitro*.

Over the years, prospective methods for identifying and isolating NSCs have been developed mainly based on cell-surface antigen-based selection (Panchision et al., 2007). This method of enrichment allows prospective isolation of NSCs and provides a valuable tool to purify neural subpopulations under defined conditions from embryonic tissue and *in vitro* cultures. For example, there are several pan surface markers that identify NSCs, such as Prominin-1 (CD133) (Schwartz et al., 2003; Uchida et al., 2000), NCAM (CD56) (Butenschon et al., 2016) and NGFR (CD271) (Vishwakarma et al., 2014). However, no specific surface markers have emerged for identifying cortical NSCs. Years ago, our lab showed that Forse-1 was highly efficient in isolating early rosette cells corresponding to early cortical NSCs (Elkabetz et al., 2008). However, Forse-1 is a sugar moiety and not a concrete gene marker which was also expressed at later time points and therefore was

not trivial to use. Hence, our primary goal was to find a surface marker which positively identifies the early cortical NSCs in our culture. To do so, we first derived cortical NSC from hiPSCs under our Triple-i protocol and propagated them long-term (until day 50). We collected the early-stage NSCs (day 12), mid-stage NSCs (day 35) and late-stage NSCs (day 50) by specifically picking the rosette structures on every stage and conducted scRNA-seq. In alignment with the previous studies from our lab (Edri et al., 2015; Ziller et al., 2015) and mimicking *in vivo* development, we identify stage- and cell state-specific genes which exhibit differential expression patterns in the diverse cell types across the progression of neural induction. From this dataset we analyzed the highest differentially expressed genes in early NSCs compared to late NSCs, and we initially identify 6 potential candidates for early cortical NSCs that followed five main criteria: 1) being a stem cell marker; 2) having cortical identity; 3) being an early marker; 4) being representative of the population; and 5) being expressed in the cell membrane to be able to conduct FACS. After initial immunostaining analysis we show that only two markers, PRTG and MCAM, are specifically enriched in early NSCs. In line with these observations, we also show their enrichment in early cerebral organoids compared to late-stage organoids, derived under the same conditions.

Additionally, upon examining a publicly available dataset containing primary human cortical samples from embryonic development, we find both markers being enriched in the early stages of human cortical development (pcw 6), predominantly in early cortical RG cells (Bhaduri et al., 2020). While the above findings support the potential of both markers being used to isolate early NSC populations in cortical development, there are no studies reporting a specific expression of these markers in cortical NSCs. Remarkably, there is published evidence that PRTG is expressed in early CNS development, but there is no literature that connects MCAM expression to any NSC population.

PRTG is a cell adhesion molecule, part of the immunoglobulin superfamily, that was first identified during embryonic development in chick (Toyoda et al., 2005) and soon after in mouse (Vesque et al., 2006). One of the more recent studies reports that expression of PRTG emerges during mouse development in the neural tube by day E7.75, being strong until day E9.5 and starting to decrease after E10.5. Co-expression of PRTG together

with Sox2 during E7.5–E10.5 in mouse embryos confirms expression of PRTG in early neural progenitors (Wong et al., 2010). This sharp downregulation that PRTG undergoes during mouse embryonic neurogenesis falls in line with our results that show an early and transitional stage of PRTG expression during neural differentiation. However, there are no studies that provide any evidence of PRTG demarcating cortical lineages.

MCAM, also known as CD146, was originally identified as an endothelial cell marker with a role in cell-matrix interaction and angiogenesis, and highly expressed in many tumors and mesenchymal cells (Lehmann et al., 1987). Multiple studies identify and demonstrate that MCAM defines a subpopulation of mesenchymal stem cells (MSCs) that are capable of bone formation and in vivo trans-endothelial migration (Harkness et al., 2016). Also, it has been reported that high expression of MCAM can be used to detect a specific subpopulation of self-renewing MSCs from the bone marrow (Sacchetti et al., 2007) and from the placenta (Ulrich et al., 2015). Higher MCAM expression in these cells correlates with their robust osteogenic differentiation potential (Tormin et al., 2011). However, in the neural context, the only reported evidence connecting MCAM to the CNS is in neural stem cell vascular niche regulation. It has been shown that MCAM is expressed in endothelial cells composing the SVZ niche, and that physical binding between neural matriptase expressed in NSCs and MCAM induces a signaling cascade that regulates NSC behavior (Tung & Lee, 2017).

Based on the above-described studies, there is some evidence to further support the idea that both PRTG and MCAM are potential markers for progenitor cells. However, discrepancies appear upon further assessing the fidelity of both markers by means of FACS and subsequent RNA-seq of the sorted subpopulations in our culture. We discover that MCAM does not show a clear segregation of subpopulations based on regional identity, while sorting for PRTG expression alone efficiently enriches for cortical NSCs since pluripotent cells, neurons, and posterior NSCs remain in the PRTG negative subpopulation. Remarkably, on day 35, PRTG expression clearly demarcates non-cortical identity instead of cortical fates. This shift in PRTG expression falls in line with the fact that on day 35 only about 10% of the total population is PRTG positive, since our

protocol promotes cortical fates at expenses of non-cortical fates. Considering these results, we moved forward with PRTG as the most promising surface marker.

To further validate PRTG we characterized the day 12 sorted subpopulations by means of immunostainings. While both subpopulations show similarly high levels of FOXP1 expression, FACS-purified PRTG positive cells display higher levels of cortical markers, PAX6 and EMX1. The higher expression correlates with enhanced radial organization while displaying lower levels of the neuronal marker DCX, indicating differences in regional identity within telencephalic fates. Interestingly, we also observe that PRTG expression at day 18 is already reduced. While low expression remains in rosette cells, a higher expression level of PRTG is found in non-cortical cells that lack EMX1 expression. This supports the idea that the remaining expression of PRTG on later days will eventually be marking non-cortical cells in culture. Finally, when terminally differentiating both sorted subpopulations, we find that PRTG positive cells are more prone to differentiate into neurons as indicated by the higher number of DCX positive cells. These results further support the model by which the positive subpopulation is mainly composed of NSCs since they have more plasticity and differentiation potential to become neurons.

Finally, we also decided to grow the sorted PRTG positive cells and re-sort them on day 35. When comparing them to an unsorted culture, we see a downregulation in PRTG expression going from 11.1% of total positive cells (control) to 2.18% in the re-sorted population. Based on our previous knowledge, this indicates an enrichment towards cortical identity when pre-sorting on day 12. This is confirmed by RT-qPCR analysis of the sorted populations. We show the high expression of anterior telencephalic markers such as OTX2, PAX6, FOXP1 and SIX3 and complete absence of posterior marker GBX2 in the re-sorted populations. In summary, we show evidence that sorting for PRTG high positive cells at day 12 allows for the prospective isolation of early cortical NSCs while excluding unwanted lineages such as posterior NSCs.

Next, we investigated whether the emergence of PRTG in culture correlates with the specification of telencephalic lineages during early neural induction. First, we identify

that the earliest prominent PRTG positive population appears after 5 days of neural induction, and that its expression demarcates a distinct population. FACS-purified PRTG positive cells at day 5 display higher levels of the anterior marker OTX2 and lower levels of the posterior marker GBX2, as well as lower levels of neural crest/placodal marker TFAP2A (Dincer et al., 2013) compared to the negative subpopulation. This initial immunostaining analysis of the sorted populations show clear differences indicating PRTG's ability in segregating anterior neural ectodermal cells from other lineages such as placodal ectodermal cells.

Finally, by using scRNA-seq we further provide evidence that the establishment of cortical NSC identity in culture can be detected by the emergence of high PRTG expression as early as day 5. Furthermore, the analysis of re-plated cells grown until day 10, shows unique expression patterns of each subpopulation associated with cortical areal specification. We report that sorting for PRTG expression at early stages allows to prospectively isolate distinct cortical identities within the telencephalon in a highly homogenous manner.

In this study, we did not address the functional role of PRTG since it was beyond the scope of our main aim. However, we believe that it would be interesting to examine its potential role in telencephalic development since, in view of our results, it is not unreasonable to think that PRTG could be implicated in the initial specification of cortical identity and arealization. To date, there is little information about the biological function of PRTG. One study has reported that PRTG might act as a cellular receptor by interacting with DNAJB11 (also known as ERdj3). By extrapolating experimental results that they obtained from *in vitro* cultures and chick models, they propose that DNAJB11/PRTG signaling might play a role in maintaining the stemness potential of neural progenitors and suppressing premature neuronal differentiation during development (Wong et al., 2010). Even though the exact mechanism by which PRTG exerts its role is not clear, it is not surprising that it can act as a receptor since it has been previously demonstrated that cell adhesion molecules can function as signaling receptors, as is the case of NCAM (Paratcha et al., 2003). Moreover, recent studies report that mutations in the PRTG gene are linked to disease that arise from cortical

developmental abnormalities such as autism (Chen et al., 2022) and attention deficit/hyperactivity disorder (Wigg et al., 2008), proposing a key role of PRTG during cortical development.

To conclude, in this study we identify a novel surface marker for early human cortical NSCs. We also provide proof that PRTG can be used as a marker to cell sort and enrich for specific cortical progenitors in culture. This knowledge opens the possibility to generate more homogenous cultures of cortical NSCs derived from hiPSCs that can be used for potential therapeutic approaches.

Additionally, we propose that this knowledge could be used for developing a screening platform to unravel regulatory mechanisms driving the transition throughout the cortical NSC stages (*Figure 5.1*). We propose that PRTG could be used as a readout in a FACS-based screening platform to identify and isolate early cortical NSCs upon perturbation. For example, we could perturb our differentiation paradigm at different stages with a TF lentiviral library. PRTG would then be used in order to detect and isolate those cells that maintain or achieve an early cortical NSC identity upon overexpression of a specific TF. Sorted cells would then be analyzed by means of DNA-sequencing to identify the integrated TF responsible for the effect, and by means of RNA-sequencing to corroborate the early cortical NSC identity. Such strategy would allow us to identify TFs involved in the establishment of early cortical NSCs and possibly TFs that regulate the transition towards other NSC stages. We believe that such knowledge could provide the tools to manipulate and generate highly homogeneous stage- and region-specific NSCs which could eventually be derived from patient samples providing a reliable source for cell replacement therapies.

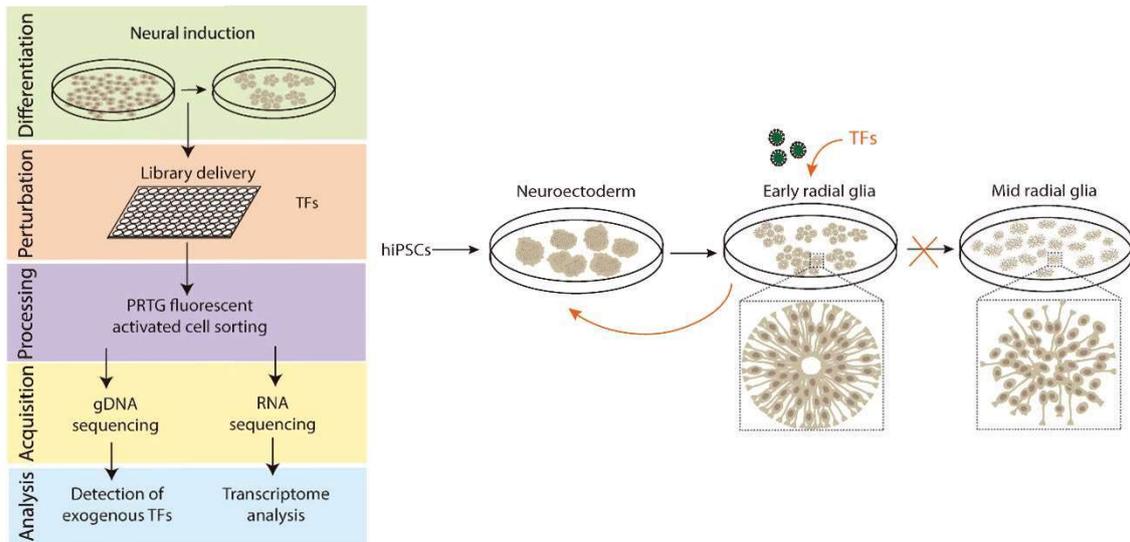


Figure 5.1. Experimental workflow of a screening strategy using PRTG as a readout. The main approach would be to transduce the early-stage cortical NSCs with a TF library and check the maintenance of PRTG positive cells along the progression of differentiation (until day 35 or day 50). This would indicate the maintenance of early NSCs on later stages of differentiation which would be sorted and analysed by bulk RNA and scRNA sequencing. By doing this, the hope would be to identify the TFs responsible for the reprogramming of the cells and simultaneously validate their early cortical NSC identity.

6. Bibliography

- Acampora, D., Barone, P., & Simeone, A. (1999). Otx genes in corticogenesis and brain development. *Cereb Cortex*, *9*(6), 533-542. <https://doi.org/10.1093/cercor/9.6.533>
- Acampora, D., Mazan, S., Lallemand, Y., Avantaggiato, V., Maury, M., Simeone, A., & Brulet, P. (1995). Forebrain and midbrain regions are deleted in Otx2^{-/-} mutants due to a defective anterior neuroectoderm specification during gastrulation. *Development*, *121*(10), 3279-3290. <https://doi.org/10.1242/dev.121.10.3279>
- Adesnik, H., & Naka, A. (2018). Cracking the Function of Layers in the Sensory Cortex. *Neuron*, *100*(5), 1028-1043. <https://doi.org/10.1016/j.neuron.2018.10.032>
- Aoto, K., Nishimura, T., Eto, K., & Motoyama, J. (2002). Mouse GLI3 regulates Fgf8 expression and apoptosis in the developing neural tube, face, and limb bud. *Dev Biol*, *251*(2), 320-332. <https://doi.org/10.1006/dbio.2002.0811>
- Appolloni, I., Calzolari, F., Corte, G., Perris, R., & Malatesta, P. (2008). Six3 controls the neural progenitor status in the murine CNS. *Cereb Cortex*, *18*(3), 553-562. <https://doi.org/10.1093/cercor/bhm092>
- Azim, E., Jabaudon, D., Fame, R. M., & Macklis, J. D. (2009). SOX6 controls dorsal progenitor identity and interneuron diversity during neocortical development. *Nat Neurosci*, *12*(10), 1238-1247. <https://doi.org/10.1038/nn.2387>
- Baker, J. C., Beddington, R. S., & Harland, R. M. (1999). Wnt signaling in Xenopus embryos inhibits bmp4 expression and activates neural development. *Genes Dev*, *13*(23), 3149-3159. <https://doi.org/10.1101/gad.13.23.3149>
- Bhaduri, A., Andrews, M. G., Mancia Leon, W., Jung, D., Shin, D., Allen, D., Jung, D., Schmunk, G., Haeussler, M., Salma, J., Pollen, A. A., Nowakowski, T. J., & Kriegstein, A. R. (2020). Cell stress in cortical organoids impairs molecular subtype specification. *Nature*, *578*(7793), 142-148. <https://doi.org/10.1038/s41586-020-1962-0>
- Bibel, M., Richter, J., Schrenk, K., Tucker, K. L., Staiger, V., Korte, M., Goetz, M., & Barde, Y. A. (2004). Differentiation of mouse embryonic stem cells into a defined neuronal lineage. *Nat Neurosci*, *7*(9), 1003-1009. <https://doi.org/10.1038/nm1301>
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114-2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Bush, G. (2011). Cingulate, frontal, and parietal cortical dysfunction in attention-deficit/hyperactivity disorder. *Biol Psychiatry*, *69*(12), 1160-1167. <https://doi.org/10.1016/j.biopsych.2011.01.022>
- Butenschon, J., Zimmermann, T., Schmarowski, N., Nitsch, R., Fackelmeier, B., Friedemann, K., Radyushkin, K., Baumgart, J., Lutz, B., & Leschik, J. (2016). PSA-NCAM positive neural progenitors stably expressing BDNF promote

- functional recovery in a mouse model of spinal cord injury. *Stem Cell Res Ther*, 7, 11. <https://doi.org/10.1186/s13287-015-0268-x>
- Butt, S. J., Sousa, V. H., Fuccillo, M. V., Hjerling-Leffler, J., Miyoshi, G., Kimura, S., & Fishell, G. (2008). The requirement of Nkx2-1 in the temporal specification of cortical interneuron subtypes. *Neuron*, 59(5), 722-732. <https://doi.org/10.1016/j.neuron.2008.07.031>
- Cadwell, C. R., Bhaduri, A., Mostajo-Radji, M. A., Keefe, M. G., & Nowakowski, T. J. (2019). Development and Arealization of the Cerebral Cortex. *Neuron*, 103(6), 980-1004. <https://doi.org/10.1016/j.neuron.2019.07.009>
- Camp, J. G., Badsha, F., Florio, M., Kanton, S., Gerber, T., Wilsch-Brauninger, M., Lewitus, E., Sykes, A., Hevers, W., Lancaster, M., Knoblich, J. A., Lachmann, R., Paabo, S., Huttner, W. B., & Treutlein, B. (2015). Human cerebral organoids recapitulate gene expression programs of fetal neocortex development. *Proc Natl Acad Sci U S A*, 112(51), 15672-15677. <https://doi.org/10.1073/pnas.1520760112>
- Camus, A., & Tam, P. P. (1999). The organizer of the gastrulating mouse embryo. *Curr Top Dev Biol*, 45, 117-153. [https://doi.org/10.1016/s0070-2153\(08\)60315-4](https://doi.org/10.1016/s0070-2153(08)60315-4)
- Carulli, D., de Winter, F., & Verhaagen, J. (2021). Semaphorins in Adult Nervous System Plasticity and Disease. *Front Synaptic Neurosci*, 13, 672891. <https://doi.org/10.3389/fnsyn.2021.672891>
- Chambers, S. M., Fasano, C. A., Papapetrou, E. P., Tomishima, M., Sadelain, M., & Studer, L. (2009). Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nat Biotechnol*, 27(3), 275-280. <https://doi.org/10.1038/nbt.1529>
- Chang, E., Ruan, X., Zhu, R., Wang, Y., & Zhang, J. (2020). Values of Single-Cell RNA Sequencing in Development of Cerebral Cortex. *Adv Exp Med Biol*, 1255, 231-247. https://doi.org/10.1007/978-981-15-4494-1_19
- Charron, F., & Tessier-Lavigne, M. (2005). Novel brain wiring functions for classical morphogens: a role as graded positional cues in axon guidance. *Development*, 132(10), 2251-2262. <https://doi.org/10.1242/dev.01830>
- Chen, G., Sima, J., Jin, M., Wang, K. Y., Xue, X. J., Zheng, W., Ding, Y. Q., & Yuan, X. B. (2008). Semaphorin-3A guides radial migration of cortical neurons during development. *Nat Neurosci*, 11(1), 36-44. <https://doi.org/10.1038/nn2018>
- Chen, S., Xiong, J., Chen, B., Zhang, C., Deng, X., He, F., Yang, L., Chen, C., Peng, J., & Yin, F. (2022). Autism spectrum disorder and comorbid neurodevelopmental disorders (ASD-NDDs): Clinical and genetic profile of a pediatric cohort. *Clin Chim Acta*, 524, 179-186. <https://doi.org/10.1016/j.cca.2021.11.014>
- Cheung, A. F., Pollen, A. A., Tavare, A., DeProto, J., & Molnar, Z. (2007). Comparative aspects of cortical neurogenesis in vertebrates. *J Anat*, 211(2), 164-176. <https://doi.org/10.1111/j.1469-7580.2007.00769.x>

- Copp, A. J., Stanier, P., & Greene, N. D. (2013). Neural tube defects: recent advances, unsolved questions, and controversies. *Lancet Neurol*, *12*(8), 799-810. [https://doi.org/10.1016/S1474-4422\(13\)70110-8](https://doi.org/10.1016/S1474-4422(13)70110-8)
- Corbin, J. G., Gaiano, N., Machold, R. P., Langston, A., & Fishell, G. (2000). The Gsh2 homeodomain gene controls multiple aspects of telencephalic development. *Development*, *127*(23), 5007-5020. <https://doi.org/10.1242/dev.127.23.5007>
- Costamagna, G., Comi, G. P., & Corti, S. (2021). Advancing Drug Discovery for Neurological Disorders Using iPSC-Derived Neural Organoids. *Int J Mol Sci*, *22*(5). <https://doi.org/10.3390/ijms22052659>
- Crossley, P. H., Martinez, S., Ohkubo, Y., & Rubenstein, J. L. (2001). Coordinate expression of Fgf8, Otx2, Bmp4, and Shh in the rostral prosencephalon during development of the telencephalic and optic vesicles. *Neuroscience*, *108*(2), 183-206. [https://doi.org/10.1016/s0306-4522\(01\)00411-0](https://doi.org/10.1016/s0306-4522(01)00411-0)
- Danesin, C., Peres, J. N., Johansson, M., Snowden, V., Cording, A., Papalopulu, N., & Houart, C. (2009). Integration of telencephalic Wnt and hedgehog signaling center activities by Foxg1. *Dev Cell*, *16*(4), 576-587. <https://doi.org/10.1016/j.devcel.2009.03.007>
- Darmanis, S., Sloan, S. A., Zhang, Y., Enge, M., Caneda, C., Shuer, L. M., Hayden Gephart, M. G., Barres, B. A., & Quake, S. R. (2015). A survey of human brain transcriptome diversity at the single cell level. *Proc Natl Acad Sci U S A*, *112*(23), 7285-7290. <https://doi.org/10.1073/pnas.1507125112>
- Dincer, Z., Piao, J., Niu, L., Ganat, Y., Kriks, S., Zimmer, B., Shi, S. H., Tabar, V., & Studer, L. (2013). Specification of functional cranial placode derivatives from human pluripotent stem cells. *Cell Rep*, *5*(5), 1387-1402. <https://doi.org/10.1016/j.celrep.2013.10.048>
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15-21. <https://doi.org/10.1093/bioinformatics/bts635>
- Edri, R., Yaffe, Y., Ziller, M. J., Mutukula, N., Volkman, R., David, E., Jacob-Hirsch, J., Malcov, H., Levy, C., Rechavi, G., Gat-Viks, I., Meissner, A., & Elkabetz, Y. (2015). Analysing human neural stem cell ontogeny by consecutive isolation of Notch active neural progenitors. *Nat Commun*, *6*, 6500. <https://doi.org/10.1038/ncomms7500>
- Eiraku, M., Watanabe, K., Matsuo-Takasaki, M., Kawada, M., Yonemura, S., Matsumura, M., Wataya, T., Nishiyama, A., Muguruma, K., & Sasai, Y. (2008). Self-organized formation of polarized cortical tissues from ESCs and its active manipulation by extrinsic signals. *Cell Stem Cell*, *3*(5), 519-532. <https://doi.org/10.1016/j.stem.2008.09.002>
- Elkabetz, Y., Panagiotakos, G., Al Shamy, G., Socci, N. D., Tabar, V., & Studer, L. (2008). Human ES cell-derived neural rosettes reveal a functionally distinct early

- neural stem cell stage. *Genes Dev*, 22(2), 152-165.
<https://doi.org/10.1101/gad.1616208>
- Feng, H., Moakley, D. F., Chen, S., McKenzie, M. G., Menon, V., & Zhang, C. (2021). Complexity and graded regulation of neuronal cell-type-specific alternative splicing revealed by single-cell RNA sequencing. *Proc Natl Acad Sci U S A*, 118(10). <https://doi.org/10.1073/pnas.2013056118>
- Fischer, T., Guimera, J., Wurst, W., & Prakash, N. (2007). Distinct but redundant expression of the Frizzled Wnt receptor genes at signaling centers of the developing mouse brain. *Neuroscience*, 147(3), 693-711.
<https://doi.org/10.1016/j.neuroscience.2007.04.060>
- Gaspard, N., Bouschet, T., Hourez, R., Dimidschstein, J., Naeije, G., van den Aemele, J., Espuny-Camacho, I., Herpoel, A., Passante, L., Schiffmann, S. N., Gaillard, A., & Vanderhaeghen, P. (2008). An intrinsic mechanism of corticogenesis from embryonic stem cells. *Nature*, 455(7211), 351-357.
<https://doi.org/10.1038/nature07287>
- Georgala, P. A., Carr, C. B., & Price, D. J. (2011). The role of Pax6 in forebrain development. *Dev Neurobiol*, 71(8), 690-709. <https://doi.org/10.1002/dneu.20895>
- Gilbert SF. (2000) Developmental Biology. 6th edition. Sunderland (MA): Sinauer Associates. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK9983/>
- Greig, L. C., Woodworth, M. B., Galazo, M. J., Padmanabhan, H., & Macklis, J. D. (2013). Molecular logic of neocortical projection neuron specification, development and diversity. *Nat Rev Neurosci*, 14(11), 755-769.
<https://doi.org/10.1038/nrn3586>
- Grove, E. A., Tole, S., Limon, J., Yip, L., & Ragsdale, C. W. (1998). The hem of the embryonic cerebral cortex is defined by the expression of multiple Wnt genes and is compromised in Gli3-deficient mice. *Development*, 125(12), 2315-2325.
<https://doi.org/10.1242/dev.125.12.2315>
- Gunhaga, L., Marklund, M., Sjodal, M., Hsieh, J. C., Jessell, T. M., & Edlund, T. (2003). Specification of dorsal telencephalic character by sequential Wnt and FGF signaling. *Nat Neurosci*, 6(7), 701-707. <https://doi.org/10.1038/nn1068>
- Hansen, D. V., Rubenstein, J. L., & Kriegstein, A. R. (2011). Deriving excitatory neurons of the neocortex from pluripotent stem cells. *Neuron*, 70(4), 645-660.
<https://doi.org/10.1016/j.neuron.2011.05.006>
- Harkness, L., Zaher, W., Ditzel, N., Isa, A., & Kassem, M. (2016). CD146/MCAM defines functionality of human bone marrow stromal stem cell populations. *Stem Cell Res Ther*, 7, 4. <https://doi.org/10.1186/s13287-015-0266-z>
- Harrison-Uy, S. J., & Pleasure, S. J. (2012). Wnt signaling and forebrain development. *Cold Spring Harb Perspect Biol*, 4(7), a008094.
<https://doi.org/10.1101/cshperspect.a008094>
- Haubst, N., Berger, J., Radjendirane, V., Graw, J., Favor, J., Saunders, G. F., Stoykova, A., & Gotz, M. (2004). Molecular dissection of Pax6 function: the specific roles of

- the paired domain and homeodomain in brain development. *Development*, *131*(24), 6131-6140. <https://doi.org/10.1242/dev.01524>
- Hebert, J. M., & Fishell, G. (2008). The genetics of early telencephalon patterning: some assembly required. *Nat Rev Neurosci*, *9*(9), 678-685. <https://doi.org/10.1038/nrn2463>
- Hedlund, E., & Deng, Q. (2018). Single-cell RNA sequencing: Technical advancements and biological applications. *Mol Aspects Med*, *59*, 36-46. <https://doi.org/10.1016/j.mam.2017.07.003>
- Hemmati-Brivanlou, A., Kelly, O. G., & Melton, D. A. (1994). Follistatin, an antagonist of activin, is expressed in the Spemann organizer and displays direct neuralizing activity. *Cell*, *77*(2), 283-295. [https://doi.org/10.1016/0092-8674\(94\)90320-4](https://doi.org/10.1016/0092-8674(94)90320-4)
- Hettige, N. C., Peng, H., Wu, H., Zhang, X., Yerko, V., Zhang, Y., Jefri, M., Soubannier, V., Maussion, G., Alsuwaidi, S., Ni, A., Rocha, C., Krishnan, J., McCarty, V., Antonyan, L., Schuppert, A., Turecki, G., Fon, E. A., Durcan, T. M., & Ernst, C. (2022). FOXP1 dose tunes cell proliferation dynamics in human forebrain progenitor cells. *Stem Cell Reports*, *17*(3), 475-488. <https://doi.org/10.1016/j.stemcr.2022.01.010>
- Hoerder-Suabedissen, A., & Molnar, Z. (2015). Development, evolution and pathology of neocortical subplate neurons. *Nat Rev Neurosci*, *16*(3), 133-146. <https://doi.org/10.1038/nrn3915>
- Holm, P. C., Mader, M. T., Haubst, N., Wizenmann, A., Sigvardsson, M., & Gotz, M. (2007). Loss- and gain-of-function analyses reveal targets of Pax6 in the developing mouse telencephalon. *Mol Cell Neurosci*, *34*(1), 99-119. <https://doi.org/10.1016/j.mcn.2006.10.008>
- Hu, B. Y., & Zhang, S. C. (2010). Directed differentiation of neural-stem cells and subtype-specific neurons from hESCs. *Methods Mol Biol*, *636*, 123-137. https://doi.org/10.1007/978-1-60761-691-7_8
- Inoue, F., Kurokawa, D., Takahashi, M., & Aizawa, S. (2012). Gbx2 directly restricts Otx2 expression to forebrain and midbrain, competing with class III POU factors. *Mol Cell Biol*, *32*(13), 2618-2627. <https://doi.org/10.1128/MCB.00083-12>
- Ishikawa, Y., Yamamoto, N., Yoshimoto, M., & Ito, H. (2012). The primary brain vesicles revisited: are the three primary vesicles (forebrain/midbrain/hindbrain) universal in vertebrates? *Brain Behav Evol*, *79*(2), 75-83. <https://doi.org/10.1159/000334842>
- Jacobson, A. G., & Tam, P. P. (1982). Cephalic neurulation in the mouse embryo analyzed by SEM and morphometry. *Anat Rec*, *203*(3), 375-396. <https://doi.org/10.1002/ar.1092030308>
- Juric-Sekhar, G., & Hevner, R. F. (2019). Malformations of Cerebral Cortex Development: Molecules and Mechanisms. *Annu Rev Pathol*, *14*, 293-318. <https://doi.org/10.1146/annurev-pathmechdis-012418-012927>

- Kelava, I., & Lancaster, M. A. (2016). Stem Cell Models of Human Brain Development. *Cell Stem Cell*, *18*(6), 736-748. <https://doi.org/10.1016/j.stem.2016.05.022>
- Klein, A. M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D. A., & Kirschner, M. W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell*, *161*(5), 1187-1201. <https://doi.org/10.1016/j.cell.2015.04.044>
- Kumar, R. M., Cahan, P., Shalek, A. K., Satija, R., DaleyKeyser, A., Li, H., Zhang, J., Pardee, K., Gennert, D., Trombetta, J. J., Ferrante, T. C., Regev, A., Daley, G. Q., & Collins, J. J. (2014). Deconstructing transcriptional heterogeneity in pluripotent stem cells. *Nature*, *516*(7529), 56-61. <https://doi.org/10.1038/nature13920>
- Lancaster, M. A., & Knoblich, J. A. (2014). Generation of cerebral organoids from human pluripotent stem cells. *Nat Protoc*, *9*(10), 2329-2340. <https://doi.org/10.1038/nprot.2014.158>
- Lee, H. K., Lee, H. S., & Moody, S. A. (2014). Neural transcription factors: from embryos to neural stem cells. *Mol Cells*, *37*(10), 705-712. <https://doi.org/10.14348/molcells.2014.0227>
- Lehmann, J. M., Holzmann, B., Breitbart, E. W., Schmiegelow, P., Riethmuller, G., & Johnson, J. P. (1987). Discrimination between benign and malignant cells of melanocytic lineage by two novel antigens, a glycoprotein with a molecular weight of 113,000 and a protein with a molecular weight of 76,000. *Cancer Res*, *47*(3), 841-845. <https://www.ncbi.nlm.nih.gov/pubmed/3542195>
- Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, *12*, 323. <https://doi.org/10.1186/1471-2105-12-323>
- Lindvall, O., & Kokaia, Z. (2010). Stem cells in human neurodegenerative disorders--time for clinical translation? *J Clin Invest*, *120*(1), 29-40. <https://doi.org/10.1172/JCI40543>
- Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, *25*(4), 402-408. <https://doi.org/10.1006/meth.2001.1262>
- Loomba, S., Straehle, J., Gangadharan, V., Heike, N., Khalifa, A., Motta, A., Ju, N., Sievers, M., Gempt, J., Meyer, H. S., & Helmstaedter, M. (2022). Connectomic comparison of mouse and human cortex. *Science*, *377*(6602), eabo0924. <https://doi.org/10.1126/science.abo0924>
- Macosko, E. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A. R., Kamitaki, N., Martersteck, E. M., Trombetta, J. J., Weitz, D. A., Sanes, J. R., Shalek, A. K., Regev, A., & McCarroll, S. A. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell*, *161*(5), 1202-1214. <https://doi.org/10.1016/j.cell.2015.05.002>

- Mangold, O., & Spemann, H. (1927). [Not Available]. *Wilhelm Roux Arch Entwickl Mech Org*, *111*(1), 341-422. <https://doi.org/10.1007/BF02080953> (Über Induktion von Medullarplatte durch Medullarplatte im Jungeren Keim, ein Beispiel homoogenetischer oder assimilatorischer Induktion.)
- Manuel, M. N., Martynoga, B., Molinek, M. D., Quinn, J. C., Kroemmer, C., Mason, J. O., & Price, D. J. (2011). The transcription factor Foxg1 regulates telencephalic progenitor proliferation cell autonomously, in part by controlling Pax6 expression levels. *Neural Dev*, *6*, 9. <https://doi.org/10.1186/1749-8104-6-9>
- Mariani, J., Simonini, M. V., Palejev, D., Tomasini, L., Coppola, G., Szekely, A. M., Horvath, T. L., & Vaccarino, F. M. (2012). Modeling human cortical development in vitro using induced pluripotent stem cells. *Proc Natl Acad Sci U S A*, *109*(31), 12770-12775. <https://doi.org/10.1073/pnas.1202944109>
- Maroof, A. M., Keros, S., Tyson, J. A., Ying, S. W., Ganat, Y. M., Merkle, F. T., Liu, B., Goulburn, A., Stanley, E. G., Elefanty, A. G., Widmer, H. R., Eggan, K., Goldstein, P. A., Anderson, S. A., & Studer, L. (2013). Directed differentiation and functional maturation of cortical interneurons from human embryonic stem cells. *Cell Stem Cell*, *12*(5), 559-572. <https://doi.org/10.1016/j.stem.2013.04.008>
- Martynoga, B., Morrison, H., Price, D. J., & Mason, J. O. (2005). Foxg1 is required for specification of ventral telencephalon and region-specific regulation of dorsal telencephalic precursor proliferation and apoptosis. *Dev Biol*, *283*(1), 113-127. <https://doi.org/10.1016/j.ydbio.2005.04.005>
- Medina, L., & Abellan, A. (2009). Development and evolution of the pallium. *Semin Cell Dev Biol*, *20*(6), 698-711. <https://doi.org/10.1016/j.semcdb.2009.04.008>
- Miyata, T., Kawaguchi, A., Saito, K., Kawano, M., Muto, T., & Ogawa, M. (2004). Asymmetric production of surface-dividing and non-surface-dividing cortical progenitor cells. *Development*, *131*(13), 3133-3145. <https://doi.org/10.1242/dev.01173>
- Molyneaux, B. J., Arlotta, P., Menezes, J. R., & Macklis, J. D. (2007). Neuronal subtype specification in the cerebral cortex. *Nat Rev Neurosci*, *8*(6), 427-437. <https://doi.org/10.1038/nrn2151>
- Muzio, L., Di Benedetto, B., Stoykova, A., Boncinelli, E., Gruss, P., & Mallamaci, A. (2002). Conversion of cerebral cortex into basal ganglia in Emx2(-/-) Pax6(Sey/Sey) double-mutant mice. *Nat Neurosci*, *5*(8), 737-745. <https://doi.org/10.1038/nn892>
- Nestor, M. W., Phillips, A. W., Artimovich, E., Nestor, J. E., Hussman, J. P., & Blatt, G. J. (2016). Human Inducible Pluripotent Stem Cells and Autism Spectrum Disorder: Emerging Technologies. *Autism Res*, *9*(5), 513-535. <https://doi.org/10.1002/aur.1570>
- Nieuwkoop, P. D. (1973). The organization center of the amphibian embryo: its origin, spatial organization, and morphogenetic action. *Adv Morphog*, *10*, 1-39. <https://doi.org/10.1016/b978-0-12-028610-2.50005-8>

- Noctor, S. C., Martinez-Cerdeno, V., Ivic, L., & Kriegstein, A. R. (2004). Cortical neurons arise in symmetric and asymmetric division zones and migrate through specific phases. *Nat Neurosci*, 7(2), 136-144. <https://doi.org/10.1038/nn1172>
- Ohkubo, Y., Chiang, C., & Rubenstein, J. L. (2002). Coordinate regulation and synergistic actions of BMP4, SHH and FGF8 in the rostral prosencephalon regulate morphogenesis of the telencephalic and optic vesicles. *Neuroscience*, 111(1), 1-17. [https://doi.org/10.1016/s0306-4522\(01\)00616-9](https://doi.org/10.1016/s0306-4522(01)00616-9)
- Panchision, D. M., Chen, H. L., Pistollato, F., Papini, D., Ni, H. T., & Hawley, T. S. (2007). Optimized flow cytometric analysis of central nervous system tissue reveals novel functional relationships among cells expressing CD133, CD15, and CD24. *Stem Cells*, 25(6), 1560-1570. <https://doi.org/10.1634/stemcells.2006-0260>
- Paratcha, G., Ledda, F., & Ibanez, C. F. (2003). The neural cell adhesion molecule NCAM is an alternative signaling receptor for GDNF family ligands. *Cell*, 113(7), 867-879. [https://doi.org/10.1016/s0092-8674\(03\)00435-5](https://doi.org/10.1016/s0092-8674(03)00435-5)
- Park, I. H., Arora, N., Huo, H., Maherali, N., Ahfeldt, T., Shimamura, A., Lensch, M. W., Cowan, C., Hochedlinger, K., & Daley, G. Q. (2008). Disease-specific induced pluripotent stem cells. *Cell*, 134(5), 877-886. <https://doi.org/10.1016/j.cell.2008.07.041>
- Pasca, A. M., Sloan, S. A., Clarke, L. E., Tian, Y., Makinson, C. D., Huber, N., Kim, C. H., Park, J. Y., O'Rourke, N. A., Nguyen, K. D., Smith, S. J., Huguenard, J. R., Geschwind, D. H., Barres, B. A., & Pasca, S. P. (2015). Functional cortical neurons and astrocytes from human pluripotent stem cells in 3D culture. *Nat Methods*, 12(7), 671-678. <https://doi.org/10.1038/nmeth.3415>
- Pera, E. M., Ikeda, A., Eivers, E., & De Robertis, E. M. (2003). Integration of IGF, FGF, and anti-BMP signals via Smad1 phosphorylation in neural induction. *Genes Dev*, 17(24), 3023-3028. <https://doi.org/10.1101/gad.1153603>
- Pollen, A. A., Nowakowski, T. J., Shuga, J., Wang, X., Leyrat, A. A., Lui, J. H., Li, N., Szpankowski, L., Fowler, B., Chen, P., Ramalingam, N., Sun, G., Thu, M., Norris, M., Lebofsky, R., Toppani, D., Kemp, D. W., 2nd, Wong, M., Clerkson, B., . . . West, J. A. (2014). Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nat Biotechnol*, 32(10), 1053-1058. <https://doi.org/10.1038/nbt.2967>
- Polleux, F., Morrow, T., & Ghosh, A. (2000). Semaphorin 3A is a chemoattractant for cortical apical dendrites. *Nature*, 404(6778), 567-573. <https://doi.org/10.1038/35007001>
- Qi, Y., Zhang, X. J., Renier, N., Wu, Z., Atkin, T., Sun, Z., Ozair, M. Z., Tchieu, J., Zimmer, B., Fattahi, F., Ganat, Y., Azevedo, R., Zeltner, N., Brivanlou, A. H., Karayiorgou, M., Gogos, J., Tomishima, M., Tessier-Lavigne, M., Shi, S. H., & Studer, L. (2017). Combined small-molecule inhibition accelerates the derivation of functional cortical neurons from human pluripotent stem cells. *Nat Biotechnol*, 35(2), 154-163. <https://doi.org/10.1038/nbt.3777>

- Quadrato, G., Nguyen, T., Macosko, E. Z., Sherwood, J. L., Min Yang, S., Berger, D. R., Maria, N., Scholvin, J., Goldman, M., Kinney, J. P., Boyden, E. S., Lichtman, J. W., Williams, Z. M., McCarroll, S. A., & Arlotta, P. (2017). Cell diversity and network dynamics in photosensitive human brain organoids. *Nature*, *545*(7652), 48-53. <https://doi.org/10.1038/nature22047>
- Quinlan, R., Graf, M., Mason, I., Lumsden, A., & Kiecker, C. (2009). Complex and dynamic patterns of Wnt pathway gene expression in the developing chick forebrain. *Neural Dev*, *4*, 35. <https://doi.org/10.1186/1749-8104-4-35>
- Rakic, P. (1972). Mode of cell migration to the superficial layers of fetal monkey neocortex. *J Comp Neurol*, *145*(1), 61-83. <https://doi.org/10.1002/cne.901450105>
- Ramón y Cajal, S. (1995). *Histology of the Nervous System of Man and Vertebrates*. Oxford Univ. Press, New York.
- Rosebrock, D., Arora, S., Mutukula, N., Volkman, R., Gralinska, E., Balaskas, A., Aragonés Hernández, A., Buschow, R., Brandl, B., Müller, F. J., Arndt, P. F., Vingron, M., & Elkabetz, Y. (2022). Enhanced cortical neural stem cell identity through short SMAD and WNT inhibition in human cerebral organoids facilitates emergence of outer radial glial cells. *Nat Cell Biol*, *24*(6), 981-995. <https://doi.org/10.1038/s41556-022-00929-5>
- Ruiz i Altaba, A., Palma, V., & Dahmane, N. (2002). Hedgehog-Gli signalling and the growth of the brain. *Nat Rev Neurosci*, *3*(1), 24-33. <https://doi.org/10.1038/nrn704>
- Sabitha, K. R., Shetty, A. K., & Upadhyaya, D. (2021). Patient-derived iPSC modeling of rare neurodevelopmental disorders: Molecular pathophysiology and prospective therapies. *Neurosci Biobehav Rev*, *121*, 201-219. <https://doi.org/10.1016/j.neubiorev.2020.12.025>
- Sacchetti, B., Funari, A., Michienzi, S., Di Cesare, S., Piersanti, S., Saggio, I., Tagliafico, E., Ferrari, S., Robey, P. G., Riminucci, M., & Bianco, P. (2007). Self-renewing osteoprogenitors in bone marrow sinusoids can organize a hematopoietic microenvironment. *Cell*, *131*(2), 324-336. <https://doi.org/10.1016/j.cell.2007.08.025>
- Sasai, Y., Lu, B., Steinbeisser, H., Geissert, D., Gont, L. K., & De Robertis, E. M. (1994). Xenopus chordin: a novel dorsalizing factor activated by organizer-specific homeobox genes. *Cell*, *79*(5), 779-790. [https://doi.org/10.1016/0092-8674\(94\)90068-x](https://doi.org/10.1016/0092-8674(94)90068-x)
- Saurat, N. G., Livesey, F. J., & Moore, S. (2016). Cortical Differentiation of Human Pluripotent Cells for In Vitro Modeling of Alzheimer's Disease. *Methods Mol Biol*, *1303*, 267-278. https://doi.org/10.1007/978-1-4939-2627-5_16
- Schmierer, B., & Hill, C. S. (2007). TGFbeta-SMAD signal transduction: molecular specificity and functional flexibility. *Nat Rev Mol Cell Biol*, *8*(12), 970-982. <https://doi.org/10.1038/nrm2297>

- Schwartz, P. H., Bryant, P. J., Fuja, T. J., Su, H., O'Dowd, D. K., & Klassen, H. (2003). Isolation and characterization of neural progenitor cells from post-mortem human cortex. *J Neurosci Res*, *74*(6), 838-851. <https://doi.org/10.1002/jnr.10854>
- Shalek, A. K., Satija, R., Adiconis, X., Gertner, R. S., Gaublomme, J. T., Raychowdhury, R., Schwartz, S., Yosef, N., Malboeuf, C., Lu, D., Trombetta, J. J., Gennert, D., Gnirke, A., Goren, A., Hacohen, N., Levin, J. Z., Park, H., & Regev, A. (2013). Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature*, *498*(7453), 236-240. <https://doi.org/10.1038/nature12172>
- Sheng, G., dos Reis, M., & Stern, C. D. (2003). Churchill, a zinc finger transcriptional activator, regulates the transition between gastrulation and neurulation. *Cell*, *115*(5), 603-613. [https://doi.org/10.1016/s0092-8674\(03\)00927-9](https://doi.org/10.1016/s0092-8674(03)00927-9)
- Shi, Y., Kirwan, P., & Livesey, F. J. (2012). Directed differentiation of human pluripotent stem cells to cerebral cortex neurons and neural networks. *Nat Protoc*, *7*(10), 1836-1846. <https://doi.org/10.1038/nprot.2012.116>
- Shi, Y., Kirwan, P., Smith, J., Robinson, H. P., & Livesey, F. J. (2012). Human cerebral cortex development from pluripotent stem cells to functional excitatory synapses. *Nat Neurosci*, *15*(3), 477-486, S471. <https://doi.org/10.1038/nn.3041>
- Shih, J., & Fraser, S. E. (1996). Characterizing the zebrafish organizer: microsurgical analysis at the early-shield stage. *Development*, *122*(4), 1313-1322. <https://doi.org/10.1242/dev.122.4.1313>
- Shimamura, K., & Rubenstein, J. L. (1997). Inductive interactions direct early regionalization of the mouse forebrain. *Development*, *124*(14), 2709-2718. <https://doi.org/10.1242/dev.124.14.2709>
- Smith, W. C., & Harland, R. M. (1992). Expression cloning of noggin, a new dorsalizing factor localized to the Spemann organizer in *Xenopus* embryos. *Cell*, *70*(5), 829-840. [https://doi.org/10.1016/0092-8674\(92\)90316-5](https://doi.org/10.1016/0092-8674(92)90316-5)
- Staal, F. J., Luis, T. C., & Tiemessen, M. M. (2008). WNT signalling in the immune system: WNT is spreading its wings. *Nat Rev Immunol*, *8*(8), 581-593. <https://doi.org/10.1038/nri2360>
- Streit, A., Berliner, A. J., Papanayotou, C., Sirulnik, A., & Stern, C. D. (2000). Initiation of neural induction by FGF signalling before gastrulation. *Nature*, *406*(6791), 74-78. <https://doi.org/10.1038/35017617>
- Sussel, L., Marin, O., Kimura, S., & Rubenstein, J. L. (1999). Loss of Nkx2.1 homeobox gene function results in a ventral to dorsal molecular respecification within the basal telencephalon: evidence for a transformation of the pallidum into the striatum. *Development*, *126*(15), 3359-3370. <https://doi.org/10.1242/dev.126.15.3359>
- Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., & Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*, *131*(5), 861-872. <https://doi.org/10.1016/j.cell.2007.11.019>

- Tandon, R., Brandl, B., Baryshnikova, N., Landshammer, A., Steenpass, L., Keminer, O., Pless, O., & Muller, F. J. (2018). Generation of two human isogenic iPSC lines from fetal dermal fibroblasts. *Stem Cell Res*, *33*, 120-124. <https://doi.org/10.1016/j.scr.2018.10.004>
- Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B. B., Siddiqui, A., Lao, K., & Surani, M. A. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods*, *6*(5), 377-382. <https://doi.org/10.1038/nmeth.1315>
- Tao, W., & Lai, E. (1992). Telencephalon-restricted expression of BF-1, a new member of the HNF-3/fork head gene family, in the developing rat brain. *Neuron*, *8*(5), 957-966. [https://doi.org/10.1016/0896-6273\(92\)90210-5](https://doi.org/10.1016/0896-6273(92)90210-5)
- Tormin, A., Li, O., Brune, J. C., Walsh, S., Schutz, B., Ehinger, M., Ditzel, N., Kassem, M., & Scheding, S. (2011). CD146 expression on primary nonhematopoietic bone marrow stem cells is correlated with in situ localization. *Blood*, *117*(19), 5067-5077. <https://doi.org/10.1182/blood-2010-08-304287>
- Toyoda, R., Nakamura, H., & Watanabe, Y. (2005). Identification of protogenin, a novel immunoglobulin superfamily gene expressed during early chick embryogenesis. *Gene Expr Patterns*, *5*(6), 778-785. <https://doi.org/10.1016/j.modgep.2005.04.001>
- Triarhou, L. C. (2021). Pre-Brodmann pioneers of cortical cytoarchitectonics I: Theodor Meynert, Vladimir Betz and William Bevan-Lewis. *Brain Struct Funct*, *226*(1), 49-67. <https://doi.org/10.1007/s00429-020-02168-6>
- Tung, H. H., & Lee, S. L. (2017). Physical Binding of Endothelial MCAM and Neural Transmembrane Protease Matriptase-Novel Cell Adhesion in Neural Stem cell Vascular Niche. *Sci Rep*, *7*(1), 4946. <https://doi.org/10.1038/s41598-017-05131-4>
- Uchida, N., Buck, D. W., He, D., Reitsma, M. J., Masek, M., Phan, T. V., Tsukamoto, A. S., Gage, F. H., & Weissman, I. L. (2000). Direct isolation of human central nervous system stem cells. *Proc Natl Acad Sci U S A*, *97*(26), 14720-14725. <https://doi.org/10.1073/pnas.97.26.14720>
- Ulrich, C., Abruzzese, T., Maerz, J. K., Ruh, M., Amend, B., Benz, K., Rolauffs, B., Abele, H., Hart, M. L., & Aicher, W. K. (2015). Human Placenta-Derived CD146-Positive Mesenchymal Stromal Cells Display a Distinct Osteogenic Differentiation Potential. *Stem Cells Dev*, *24*(13), 1558-1569. <https://doi.org/10.1089/scd.2014.0465>
- Velasco, S., Kedaigle, A. J., Simmons, S. K., Nash, A., Rocha, M., Quadrato, G., Paulsen, B., Nguyen, L., Adiconis, X., Regev, A., Levin, J. Z., & Arlotta, P. (2019). Individual brain organoids reproducibly form cell diversity of the human cerebral cortex. *Nature*, *570*(7762), 523-527. <https://doi.org/10.1038/s41586-019-1289-x>
- Vesque, C., Anselme, I., Couve, E., Charnay, P., & Schneider-Maunoury, S. (2006). Cloning of vertebrate Protogenin (Prtg) and comparative expression analysis

- during axis elongation. *Dev Dyn*, 235(10), 2836-2844. <https://doi.org/10.1002/dvdy.20898>
- Viebahn, C. (2001). Hensen's node. *Genesis*, 29(2), 96-103. [https://doi.org/10.1002/1526-968x\(200102\)29:2<96::aid-gene1010>3.0.co;2-h](https://doi.org/10.1002/1526-968x(200102)29:2<96::aid-gene1010>3.0.co;2-h)
- Vishwakarma, S. K., Bardia, A., Tiwari, S. K., Paspala, S. A., & Khan, A. A. (2014). Current concept in neural regeneration research: NSCs isolation, characterization and transplantation in various neurodegenerative diseases and stroke: A review. *J Adv Res*, 5(3), 277-294. <https://doi.org/10.1016/j.jare.2013.04.005>
- Vonica, A., & Gumbiner, B. M. (2007). The *Xenopus* Nieuwkoop center and Spemann-Mangold organizer share molecular components and a requirement for maternal Wnt activity. *Dev Biol*, 312(1), 90-102. <https://doi.org/10.1016/j.ydbio.2007.09.039>
- Waddington, C. H., & Schmidt, G. A. (1933). Induction by heteroplastic grafts of the primitive streak in birds. *Wilhelm Roux Arch Entwickl Mech Org*, 128(3), 522-563. <https://doi.org/10.1007/BF00649863>
- Wang, H., Kane, A. W., Lee, C., & Ahn, S. (2014). Gli3 repressor controls cell fates and cell adhesion for proper establishment of neurogenic niche. *Cell Rep*, 8(4), 1093-1104. <https://doi.org/10.1016/j.celrep.2014.07.006>
- Watanabe, Y., & Nakamura, H. (2012). Nuclear translocation of intracellular domain of Protogenin by proteolytic cleavage. *Dev Growth Differ*, 54(2), 167-176. <https://doi.org/10.1111/j.1440-169X.2011.01315.x>
- Wigg, K. G., Feng, Y., Crosbie, J., Tannock, R., Kennedy, J. L., Ickowicz, A., Malone, M., Schachar, R., & Barr, C. L. (2008). Association of ADHD and the Protogenin gene in the chromosome 15q21.3 reading disabilities linkage region. *Genes Brain Behav*, 7(8), 877-886. <https://doi.org/10.1111/j.1601-183X.2008.00425.x>
- Wilson, P. A., & Hemmati-Brivanlou, A. (1995). Induction of epidermis and inhibition of neural fate by Bmp-4. *Nature*, 376(6538), 331-333. <https://doi.org/10.1038/376331a0>
- Wilson, S. I., Rydstrom, A., Trimborn, T., Willert, K., Nusse, R., Jessell, T. M., & Edlund, T. (2001). The status of Wnt signalling regulates neural and epidermal fates in the chick embryo. *Nature*, 411(6835), 325-330. <https://doi.org/10.1038/35077115>
- Wolf, F. A., Angerer, P., & Theis, F. J. (2018). SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol*, 19(1), 15. <https://doi.org/10.1186/s13059-017-1382-0>
- Wolock, S. L., Lopez, R., & Klein, A. M. (2019). Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst*, 8(4), 281-291 e289. <https://doi.org/10.1016/j.cels.2018.11.005>
- Wong, Y. H., Lu, A. C., Wang, Y. C., Cheng, H. C., Chang, C., Chen, P. H., Yu, J. Y., & Fann, M. J. (2010). Protogenin defines a transition stage during embryonic

- neurogenesis and prevents precocious neuronal differentiation. *J Neurosci*, *30*(12), 4428-4439. <https://doi.org/10.1523/JNEUROSCI.0473-10.2010>
- Ying, Q. L., Stavridis, M., Griffiths, D., Li, M., & Smith, A. (2003). Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nat Biotechnol*, *21*(2), 183-186. <https://doi.org/10.1038/nbt780>
- Ypsilanti, A. R., & Rubenstein, J. L. (2016). Transcriptional and epigenetic mechanisms of early cortical development: An examination of how Pax6 coordinates cortical development. *J Comp Neurol*, *524*(3), 609-629. <https://doi.org/10.1002/cne.23866>
- Yu, J., Vodyanik, M. A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J. L., Tian, S., Nie, J., Jonsdottir, G. A., Ruotti, V., Stewart, R., Slukvin, II, & Thomson, J. A. (2007). Induced pluripotent stem cell lines derived from human somatic cells. *Science*, *318*(5858), 1917-1920. <https://doi.org/10.1126/science.1151526>
- Zhang, S. C., Wernig, M., Duncan, I. D., Brustle, O., & Thomson, J. A. (2001). In vitro differentiation of transplantable neural precursors from human embryonic stem cells. *Nat Biotechnol*, *19*(12), 1129-1133. <https://doi.org/10.1038/nbt1201-1129>
- Zheng, G. X., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., Ziraldo, S. B., Wheeler, T. D., McDermott, G. P., Zhu, J., Gregory, M. T., Shuga, J., Montesclaros, L., Underwood, J. G., Masquelier, D. A., Nishimura, S. Y., Schnall-Levin, M., Wyatt, P. W., Hindson, C. M., . . . Bielas, J. H. (2017). Massively parallel digital transcriptional profiling of single cells. *Nat Commun*, *8*, 14049. <https://doi.org/10.1038/ncomms14049>
- Ziller, M. J., Edri, R., Yaffe, Y., Donaghey, J., Pop, R., Mallard, W., Issner, R., Gifford, C. A., Goren, A., Xing, J., Gu, H., Cachiarelli, D., Tsankov, A., Epstein, C., Rinn, J. R., Mikkelsen, T. S., Kohlbacher, O., Gnirke, A., Bernstein, B. E., . . . Meissner, A. (2015). Dissecting neural differentiation regulatory networks through epigenetic footprinting. *Nature*, *518*(7539), 355-359. <https://doi.org/10.1038/nature13990>

7. List of publications

Rosebrock, D.* , Arora, S.* , Mutukula, N. †, Volkam, R. †, Gralinska, E., Balaskas, A., **Aragonés Hernández, A.**, ...Elkabetz, Y. Enhanced cortical neural stem cell identity through short SMAD and WNT inhibition in human cerebral organoids facilitates emergence of outer radial glial cells. *Nat Cell Biol* **24**, 981–995 (2022). <https://doi.org/10.1038/s41556-022-00929-5>

Elkabetz, Y., Arora, S., Balaskas, A., **Aragonés Hernández, A.**, Rosebrock, D. Generation of Cerebral Organoids with Enriched Cortical Cellular Diversity and Outer Radial Glial Cell Identity from Human Pluripotent Stem Cells. *Nat Protoc* **17**, 1–12 (2022). <https://doi.org/10.21203/rs.3.pex-1877/v1>

Sampath Kumar, A.* , Tian, L.* , Bolondi, A.* , **Aragonés Hernández, A.**, Stickels, R., Kretzmer, H., Murray, E., Wittler, L., Walther, M., Haut, L., Barakat, G., Macosko, E., Elkabetz, Y., Guignard, L., Chen, F., Meissner, A. Spatial transcriptomic maps of whole mouse embryos. *Manuscript under revision in Nat Biotechnol* (2022).