

Embryonic stem cell-based in vitro models to study early mammalian development

Berenger-Currias, N.M.L.P.

DOI

[10.4233/uuid:83a1b783-67b5-4b93-8d6a-705f4d8edf65](https://doi.org/10.4233/uuid:83a1b783-67b5-4b93-8d6a-705f4d8edf65)

Publication date

2022

Document Version

Final published version

Citation (APA)

Berenger-Currias, N. M. L. P. (2022). *Embryonic stem cell-based in vitro models to study early mammalian development*. [Dissertation (TU Delft), Delft University of Technology].
<https://doi.org/10.4233/uuid:83a1b783-67b5-4b93-8d6a-705f4d8edf65>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

EMBRYONIC STEM CELL-BASED IN VITRO MODELS TO STUDY EARLY MAMMALIAN DEVELOPMENT

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology,
by the authority of the Rector Magnificus Prof.dr.ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates,
to be defended publicly on Wednesday 23 February 2022 at 15:00

by

Noémie Marie Louise Pauline BÉRENGER-CURRIAS

Master in Genome and Organisms Science,
Université d'Evry, France,
born in Longjumeau, France.

This dissertation has been approved by promotor

Composition of the doctoral committee:

Rector Magnificus,	chairperson
Dr. T. Idema,	Delft University of Technology, promotor
Prof.dr.ir. S.J. Tans,	Delft University of Technology, promotor
Dr. S. Semrau,	Leiden University, copromotor

Independent members:

Prof.dr. A. Martinez Arias	University of Cambridge, United Kingdom
Prof.dr. N.H. Dekker	Delft University of Technology
Prof.dr. M.E. Drukker	Leiden University
Dr. J. van Zon	AMOLF
Prof.dr. C. Joo,	Delft University of Technology, reserve member



Keywords: stem cells, mammalian embryonic development, assembloïds

Printed by: ...

Cover: ...

Copyright © 2022 by N.M.L.P. Bérenger-Currias. All rights reserved.

Casimir PhD Series, Delft-Leiden 2022-03

ISBN 978-90-8593-514-8

An electronic version of this dissertation is available at
<http://repository.tudelft.nl/>.

A scientist in his laboratory is not only a technician: he is also a child placed before natural phenomena which impress him like a fairy tale.

Marie Curie

To my family

CONTENTS

Summary	ix
Resumé en français	xi
1 How to study mammalian development	1
1.1 Introduction	2
1.2 Top-down approaches – observing and manipulating embryonic development <i>in vivo</i> and <i>ex vivo</i>	3
1.3 Bottom-up approaches – recreating embryogenesis <i>in vitro</i>	5
1.4 Summary and outlook	8
1.5 References	11
2 XEN enhanced-gastruloids, a new <i>in vitro</i> model for studying development beyond gastrulation	17
2.1 Introduction	18
2.2 Materials and methods.	19
2.2.1 Cell culture.	19
2.2.2 Gastruloids.	20
2.2.3 XEN enhanced-gastruloids (XEGs).	20
2.2.4 Cerebral organoids.	20
2.2.5 Immunostaining.	21
2.2.6 smFISH staining	22
2.2.7 Imaging	23
2.2.8 Image analysis	24
2.2.9 Signaling perturbation experiments	24
2.2.10 Single cell RNA sequencing	24
2.3 Results	28
2.3.1 XEN cells induce neural tube-like structures in gastruloids	28
2.3.2 Tubes in XEGs show partial anteroposterior and dorsoventral patterning.	32
2.3.3 Signaling perturbation experiments and differentiation to cortex-like tissue support neural tube-like character	33
2.3.4 Single-cell RNA-seq reveals the transcriptional profiles of XEG cells	37

2.4	Discussion	42
2.5	Supplementary information	43
2.6	References	49
3	Effect and differentiation of XEN cells in XEN enhanced-Gastruloids	53
3.1	Introduction	54
3.2	Materials and methods.	56
3.2.1	Cell culture.	56
3.2.2	Gastruloids.	56
3.2.3	XEN enhanced-gastruloids (XEGs)	56
3.2.4	Labelling and imaging	57
3.2.5	Single cell RNA sequencing	57
3.3	Results	58
3.3.1	XEN cells become VE-like in XEGs	58
3.3.2	XEN cells guide symmetry breaking by locally inhibiting the formation of a T-positive population	63
3.4	Discussion	68
3.5	Supplementary information	69
3.6	References	72
4	<i>In vitro</i> model systems for epiblast / primitive endoderm sorting	75
4.1	Introduction	76
4.2	Materials and methods.	77
4.2.1	Cell culture.	77
4.2.2	Chimeric embryoid bodies (Chim-EBs)	78
4.2.3	Retinoic acid driven differentiation in 3D	80
4.3	Results	82
4.3.1	Epi/PrE sorting in Chimeric Embryoid Bodies	82
4.3.2	RA differentiation of mESCs in 3D microstructures as a model for Epi/PrE sorting.	85
4.4	Discussion	92
4.5	References	94
5	Single-cell transcriptomics reveals gene expression dynamics of human fetal kidney development	97
5.1	Introduction	98
5.2	Materials and methods	100
5.2.1	Ethics statement	100

5.2.2	Experimental methods.	100
5.2.3	Quantification and statistical analysis	102
5.3	Results	107
5.3.1	Clustering and identification of cell types	107
5.3.2	Developmental flow	116
5.3.3	Comparison with existing single-cell transcriptomics data.	117
5.3.4	Marker identification	118
5.3.5	Comparison of different developmental ages	123
5.3.6	Heterogeneity in the nephrogenic niche	126
5.3.7	Podocyte development.	131
5.4	Discussion	136
5.4.1	The nephrogenic niche is heterogeneous	136
5.4.2	Proximal-distal patterning	138
5.4.3	Fetal podocytes may have varying degree of maturation	138
5.5	Supplementary information	140
5.6	References	143
6	Outlook and discussion	155
	Curriculum Vitæ	159
	List of Publications	161

SUMMARY

Embryogenesis is a fascinating mystery. How a single fertilized egg, the zygote, reads out its genetic information and turns into a fully formed, multicellular organism still defies detailed understanding.

An adult organism can consist of trillions of cells of different types, organized to form various complex tissues and organs. To obtain such a system starting from a single cell requires an incredibly sophisticated combination of three key processes: cell division, differentiation and migration. Cell division turns a parent cell into two genetically identical copies, the daughter cells. Differentiation is the process by which a cell transforms into a particular cell type with a highly specialized function, such as a neuron, a bone cell or a muscle cell. While cells of different types have identical genomes, they express distinct sets of proteins that define the cells' function and shape. Finally, directed cell migration is crucial to obtain proper three-dimensional organization and thereby correctly formed tissues and organs.

Over many decades, several model organisms have been used to study the development of insects (*Drosophila*), fish (*Zebrafish*), amphibians (*Xenopus*), avians (Chicken), and mammals (Mouse, Human). In this thesis, we present three studies of different stages of mammalian embryogenesis. As reviewed in chapter 1, two approaches are used to study embryogenesis.

The first approach, called “top-down”, focuses on the observation and manipulation of normal embryos to identify their constitutive components and discover the mechanisms underlying embryogenesis. Lately, this approach has benefited greatly from the development of sophisticated imaging and sequencing techniques that operate at the single-cell level. In particular, single-cell expression atlases are becoming invaluable for the understanding of the developing embryo. As an example, chapter 5 presents a single-cell RNA sequencing study of the human fetal kidney. The analysis identified several transient cell types present only at the fetal stage (podocytes precursors and nephron progenitor cells). Through imaging assays, these cells were localized in the developing kidney and their gene expression profile was explored to identify potential disease-associated genes. The interactive web application accompanying this study allows for easy exploration of the data. Top-down strategies have produced a wealth of fundamental insights into embryonic development. However, they face both biological and technical limitations, especially for human development. The fact that embryogenesis occurs mostly *in utero* and legal restrictions on the *in vitro* culture of human embryos make especially high-throughput studies unfeasible.

As an alternative to the observation and manipulation of embryos, the “bottom-up” approach aims to recreate major steps of embryonic development *in vitro*, using cell lines as building blocks. These systems are simpler than a real embryo, more easily controllable and can be produced in large numbers. Hence, cell-based *in vitro* systems are ideal

for high-throughput studies and can be used to identify the minimal set of conditions needed to achieve specific morphogenic processes. In the present thesis, chapters 2, 3 and 4 present *in vitro* studies of different stages of mammalian development. Chapters 2 and 3 describe an *in vitro* system made of mouse embryonic and extra-embryonic cells called XEN-enhanced gastruloids (XEGs). This system is based on an existing model, the gastruloid, which recapitulates key hallmarks of gastrulation using only embryonic cells. The addition of extra-embryonic cells in XEGs lead to the formation of neural tube-like structures, analyzed in detail in chapter 2. The role of the extra-embryonic cells in the formation of those tubular structures, as well as their own differentiation are described in chapter 3. This study highlights the importance of extra-embryonic cues, in particular with respect to achieving complex tissue organization. In chapter 4, we present two different *in vitro* systems composed of mouse embryonic and extra-embryonic cells. Those models mimic the differentiation and subsequent sorting of embryonic and extra-embryonic cells during the earliest stages of mammalian development.

Overall, the work presented in this thesis exemplifies current trends in mammalian developmental biology. Despite the significant acceleration of research by the latest technical advances, many key mechanisms remain poorly understood. Synthetic *in vitro* systems overcome some of the limitations faced by *in vivo* observations, and can be used to test functional hypotheses. On the other hand, *in vitro* studies are always motivated by *in vivo* observations and results have to be validated *in vivo*. In conclusion, top-down and bottom-up approaches complement each other and their combination will lead to a much deeper understanding of embryonic development.

RESUMÉ EN FRANÇAIS

Le développement embryonnaire est un mystère fascinant. Comment une cellule unique, le zygote, interprète son information génétique et devient un organisme multicellulaire fonctionnel n'est toujours pas entièrement compris.

Un organisme adulte est composé de milliards de cellules différentes, organisées de manière complexe pour former les tissus et les organes. Obtenir un tel système à partir d'une cellule unique requiert une combinaison sophistiquée de trois processus clés: la division cellulaire, la différenciation et la migration. La division cellulaire est le mécanisme par lequel deux cellules filles génétiquement identiques sont formées à partir d'une cellule mère. La différenciation est la spécialisation d'une cellule dans un type particulier avec une fonction précise, tel qu'un neurone, une cellule osseuse ou une cellule musculaire. Bien que les différents types de cellules contiennent le même génome, ils expriment différentes sélections de protéines qui vont définir la fonction et l'aspect des cellules. Enfin, la migration contrôlée de cellules est cruciale pour obtenir une organisation en trois dimensions précise et former correctement les tissus et organes.

À travers les années, différents organismes ont été utilisés pour étudier le développement des insectes (*Drosophila*), des poissons (*Zebrafish*), des oiseaux (poule) et des mammifères (souris et humain). Dans cette thèse, trois études de différentes étapes de l'embryogenèse mammalienne sont détaillées. Comme expliqué dans le chapitre 1, il existe deux approches pour étudier l'embryogenèse.

La première approche, appelée "top-down", se concentre sur l'observation et la manipulation d'embryons afin d'identifier leur composants et de découvrir les mécanismes à la base de l'embryogenèse. Récemment, cette approche a bénéficié de l'essor de techniques d'imagerie et de séquençage sophistiquées ayant une résolution à l'échelle des cellules individuelles. En particulier, les bases de données d'expression génétique des cellules d'embryons de stades différents deviennent inestimables pour la compréhension du développement embryonnaire. Dans cette idée, le chapitre 5 présente une étude de séquençage du rein du fœtus humain. L'analyse de ces données a permis d'identifier plusieurs types de cellules présents seulement au stade fœtal du développement (précuseurs de podocytes et progéniteurs de cellules néphronique). Grâce à des analyses d'imagerie, ces cellules ont été localisées dans le rein en cours de développement, et leur profil d'expression génétique a été exploré afin d'identifier de potentiels gènes associés à des maladies rénales. Le site web accompagnant cette étude permet d'explorer facilement ces données. Les stratégies top-down ont permis de nombreuses découvertes sur le développement embryonnaire. Cependant, cette approche est confrontée à des limitations biologiques et techniques, surtout dans le développement humain. Le fait que l'embryogenèse se passe principalement dans l'utérus, et les restrictions légales sur la culture d'embryons humains en laboratoire rendent les études à grande échelle particulièrement compliquées.

Comme alternative à l'observation et la manipulation d'embryons, l'approche 'bottom-up' a pour but de recréer certaines étapes majeures du développement embryonnaire en utilisant les cellules comme éléments de construction. Ces systèmes sont plus simples qu'un véritable embryon, plus faciles à contrôler et peuvent être produits en grande quantité. Ainsi, les systèmes *in vitro* à base de cellules sont idéales pour les études à grande échelle et peuvent être utilisés pour identifier la combinaison minimale de facteurs nécessaire pour obtenir certaines transformations morphogéniques. Dans cette thèse, les chapitres 2, 3 et 4 présentent deux études *in vitro* de différents stades du développement embryonnaire. Les chapitres 2, 3 décrivent un système *in vitro* composé de cellules embryonnaires et extra-embryonnaires appelé XEN-enhanced gastruloid (XEGs). Ce système se base sur un système existant, le gastruloid, qui reproduit des caractéristiques constitutives de la gastrulation à partir de cellules souches embryonnaires uniquement. L'addition de cellules extra-embryonnaires dans les XEGs induit la formation de structures rappelant le tube neural, étudiées en détails dans le chapitre 2. Le rôle des cellules extra-embryonnaires dans la formation de ces structures, ainsi que leur propre différenciation dans le système sont détaillés dans le chapitre 3. Cette étude met en lumière l'importance des signaux extra-embryonnaires, en particulier dans l'organisation de tissus complexes. Le chapitre 4, présente deux systèmes *in vitro* composés de cellules souches embryonnaires et de cellules extra-embryonnaires. Ces modèles reproduisent la différenciation et la séparation des cellules embryonnaires et extra-embryonnaires observées au début du développement embryonnaire.

Le travail présenté dans cette thèse illustre les tendances actuelles dans le domaine de la biologie du développement. Malgré l'importante accélération de la recherche grâce aux récentes avancées techniques, beaucoup de mécanismes essentiels restent peu compris. Les systèmes synthétiques *in vitro* résolvent certaines des limitations inhérentes aux manipulations *in vivo*, et peuvent être utilisés pour tester des hypothèses fonctionnelles. En revanche, les études *in vitro* sont toujours motivées par des observations *in vivo* et leurs résultats doivent être validés *in vivo*. En conclusion, les approches top-down et bottom-up sont complémentaires et leur combinaison conduira à une meilleure compréhension du développement embryonnaire. //

1

HOW TO STUDY MAMMALIAN DEVELOPMENT

1.1. INTRODUCTION

The formation of fully functional organisms through embryonic development has fascinated scientists for decades. This process requires cell growth, differentiation and organization to follow a strict spatio-temporal plan. The goal of developmental biology is to understand the mechanisms that ensure the reproducibility and precision with which this plan is executed. Over the years, several model organisms have been used to study the development of insects (*Drosophila*), fish (*Zebrafish*), amphibians (*Xenopus*), avians (Chicken), and mammals (Mouse, Human). The mouse embryo is so far the best characterized model system for studying mammalian development. Notably, humans and mice share a very important trait, uncommon among mammals: the deep interstitial implantation [1], or embedding of the early embryo in the endometrial tissue of the uterus wall. This shared mode of implantation, together with the short developmental timeline of the murine embryo, and the availability of efficient tools for genetic manipulation, explain why the mouse has been used extensively as a model. The experimental approaches used to study embryonic development, can be roughly separated into two categories (Fig. 1.1).

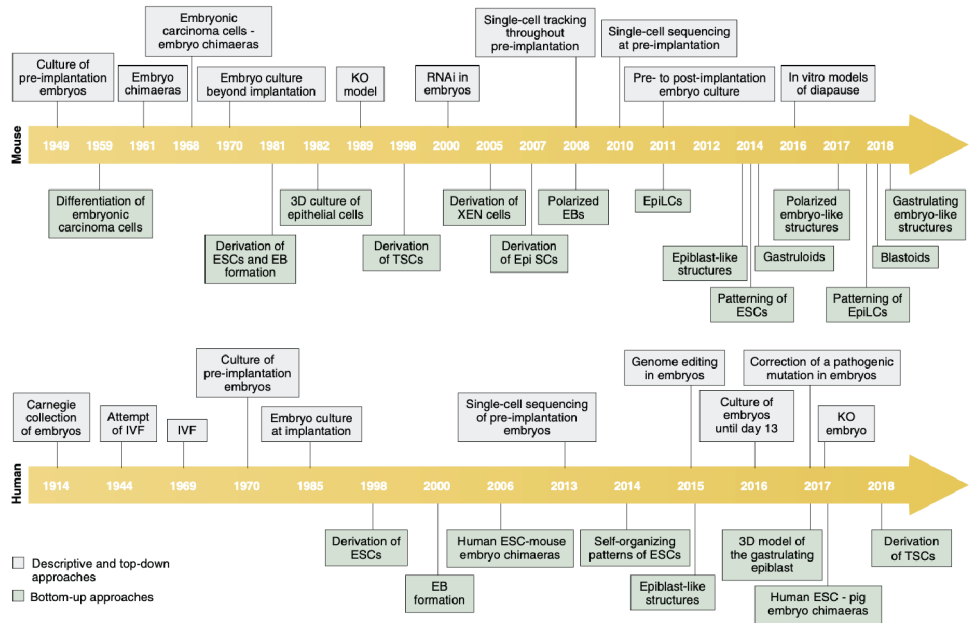


Figure 1.1: Timeline of technical breakthroughs in early mouse and human embryo research. Major milestones are divided into descriptive and top-down or bottom-up approaches. IVF: in vitro fertilization, ESC: embryonic stem cell, EB: embryoid body, KO: knock-out, TSC: trophoblast stem cell, XEN: extraembryonic endoderm-like, Epi SCs: epiblast stem cells, EpiLCS: epiblast-like cells. Figure from Shahbazi, M. N. & Zernicka-Goetz, M. (2018), *Nature Cell Biology*¹.

¹Figure reprinted with permission from Springer Nature. Licence number 5057681353757.

The first category, which was also the first explored historically, is focused on the observation of normal development and experimental manipulation of developing embryos. Those methods can be thought of as ‘top-down’ since they start from the complex system that is the developing embryo, and deconstruct it into its constitutive pieces. This approach has allowed us to acquire a lot of valuable, fundamental insights into the developing embryo and is still evolving in parallel with the latest technical advances in imaging, sequencing, and bioengineering. However, top-down strategies face limitations, due to ethical restrictions on the ways in which embryos, in particular human embryos, can be studied [2].

An alternative approach, which we might call synthetic developmental biology, works ‘bottom-up’: understanding is generated by recreating a major element or principle of embryonic development from the smallest number of simple components. Such minimal models can be controlled and tuned more easily than actual embryos and thus allow us to explore the role of many different factors, ranging from cell signaling to mechanical interactions, in great detail. This is especially interesting for studying mammalian development, and more precisely the early phases of post-implantation development when the embryo becomes less accessible to observation and experimentation. Lately, multiple models using mouse and human embryonic stem cell (ESC) were used to recreate specific milestones of mammalian development *in vitro* [3, 4, 5, 6]. Interestingly, the development of those ‘bottom-up’ systems raises new ethical questions, especially when human cells are involved.

In the coming sections, we will discuss both ‘top-down’ and ‘bottom-up’ approaches, as well as the advances their complementarity has allowed us to make.

1.2. TOP-DOWN APPROACHES – OBSERVING AND MANIPULATING EMBRYONIC DEVELOPMENT *in vivo* AND *ex vivo*

A top-down approach takes the whole embryo as the starting point and deconstructs it to identify its constitutive parts (i.e., limbs, organs, tissues, cell types, ...). This approach to mammalian development started in the 1950’s with *in vivo* observation and isolation of mouse embryos, rapidly followed by attempts to grow embryos *ex vivo* [7, 8, 9]. Those early experiments led to the development of current standard procedures to remove mouse embryos from the implantation chamber and grow them *ex vivo*, creating the possibility for diverse manipulations. Today, mouse embryos, separated from the mother as early as the blastocyst stage, can be grown *in vitro*, beyond the implantation stage and somite formation [10, 11, 12]. Among the first experimental manipulations of mouse embryos was the formation of chimeras, by extraction and transplantation of embryonic cells [13, 14]. This kind of experiment, also explored in many other model organisms, led to the discovery of cell clusters that have the ability to instruct differentiation and morphogenesis in adjacent cells. Those groups of cells are key actors in the developmental biology and are hence referred to as organizers [15, 16].

Over the years, multiple techniques of experimental embryology were developed to study mammalian development *in vitro*. Together with morphological characterization, staining of specific genes allowed the characterization of different cell types. With the advances in the field of genetic engineering, several techniques to manipulate gene expression were employed, such as homologous recombination [17], RNA interference [18], double stranded RNA [19], and knock-out through genetic engineering tools, like, most-recently CRISPR-Cas9 [20]. Knock-out experiments were particularly useful to identify essential genes in multiple key processes, such as, e.g., the epithelial to mesenchymal transition (EMT) [21, 22]. Those studies led to the discovery of specific developmental genes and their functions [23]. Genetic fluorescent reporters of specific markers were also developed to locate differentiating cell types over time and study their dynamics. Recently, progress in single cell techniques, especially RNA sequencing, brought new insights. Several organs and tissues [24, 25, 26] and even complete embryos [27, 28, 29] have been characterized by transcriptomics at the single-cell level. The constant evolution of imaging techniques has also brought a wealth of new information [30]. The development of high-throughput live imaging with confocal, and, more recently, light sheet microscopy [31], allows the imaging of single cell dynamics within the developing embryo [32, 33].

Among the many developmental processes that have been discovered by ‘top-down’ approaches, here we want to highlight the dynamics and role of the extraembryonic endoderm (Fig. 1.2), which provides mechanical [34] and biological [23] cues in the early stages of embryonic development.

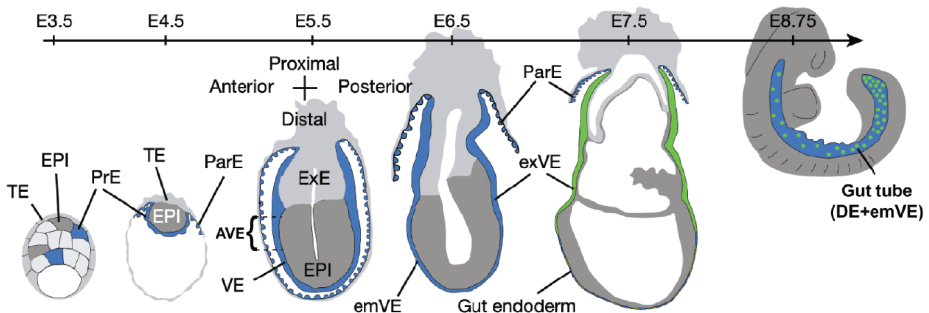


Figure 1.2: Schematic overview of the development of endodermal tissues in mice, from the blastocyst stage to the mid-gestation embryo. TE: trophectoderm, EPI: epiblast, PrE: primitive endoderm, ParE: parietal endoderm, VE: visceral endoderm, AVE: anterior visceral endoderm, ExE: extraembryonic ectoderm, exVE: extraembryonic visceral endoderm, emVE: embryonic visceral endoderm, DE: definitive (embryonic) endoderm. Adapted from Nowotschin, S. et al. (2019), *Nature*².

²Figure reprinted with permission from Springer Nature. Licence number 5057681177662.

The earliest precursor of this tissue, the primitive endoderm (PrE), differentiates in the inner cell mass (ICM), in a salt-and-pepper pattern. By day 4.5 of mouse embryonic development (E4.5), the cells have formed a single layer, which was first identified through morphological observations [35]. Further characterization of those cells identified specific marker genes, which enabled following PrE cells with live imaging [36, 37]. Knock-out experiments revealed essential factors for PrE differentiation into visceral endoderm (VE) and parietal endoderm (PE) around E5.0 [38, 39]. The specification of the distal visceral endoderm (DVE) subpopulation in the VE, and its migration to establish the anterior visceral endoderm (AVE) at E5.5 was observed with PrE marker gene fluorescent reporter [40, 41]. The AVE population was then shown to have a role in the establishment of the embryo's body axes [42, 43]. Originally, extraembryonic endoderm cells were not believed to contribute to the actual embryo. However, studies of embryonic endoderm formation using reporter cell lines and single cell RNA sequencing revealed that some VE cells become part of the developing gut, later in development [44, 24].

Ex vivo study and manipulation of embryos are an extremely valuable way of collecting information about mammalian development. Those techniques keep evolving and bring new insights about the developmental process, with constantly increasing precision. Yet, even with the latest technical advances, the *in vitro* work with mammalian embryos still faces limitations. Due to biological and technical reasons, high-throughput studies are not feasible. Moreover, besides the challenges linked to working with living animals, experimental manipulation of embryos faces ethical and legal restrictions. Those restrictions are even more extensive when working with human embryos, for which *in vitro* experimentation is typically limited to the first 14 days of development or until the formation of a primitive streak [2].

1.3. BOTTOM-UP APPROACHES – RECREATING EMBRYOGENESIS *in vitro*

An alternative to experiments with embryos, is the recreation of key developmental processes *in vitro* from one or multiple different cell lines. This 'bottom-up' approach is inspired by synthetic biology [45], where basic functional units are combined to build biological systems with defined functions. Synthetic developmental biology utilizes the self-organizing ability of pluripotent cells, such as mouse embryonic stem cell (mESC), human embryonic stem cell (hESC) or, more recently, induced pluripotent stem cell (iPSC) in two or three dimensional geometries (Fig. 1.3) [4, 5, 6]. These minimal models allow us to study features of embryonic development in a high-throughput, quantitative and easily tunable way.

The ability of embryonic stem cells (ESCs) to differentiate and self-organize in 3D was explored quickly after the isolation of ESCs from the epiblast (Epi) of mouse blastocysts [46], ESCs readily form a 3D aggregate called embryoid body (EB), which contain a variety of cell types [47]. Over time, two distinct ways of investigating and applying EBs

were developed.

Through the years, the EB model has been further refined and developed. Modern systems typically start with an undifferentiated population of ESCs, recreating an Epi-like system. Then, a minimal set of externally applied signaling cues drive the system through major developmental. In addition to the spatiotemporal regulation of cell differentiation, patterning and morphogenesis, those studies explore the importance of geometry, size and cell density. Some examples of mESC-based models, are 2D micropatterned systems [48], polarizing EBs [49], Inner Cell Mass organoids [50], and gastruloids [51, 52, 53]. Gastruloids are particularly remarkable in that they spontaneously display key hallmarks of peri-implantation mouse development. Recently, gastruloids were developed further to display indications of organogenesis under exogenous signaling inputs [54, 55, 56]. The derivation of hESCs from human blastocysts [57] and their study are more recent, yet multiple hESC *in vitro* models have already been developed. Since the geometry of the human embryo is significantly different from the mouse embryo, the geometry of those models is a crucial point. Recent models include 2D micropatterned systems [58], the synthetic human epiblast [59], as well as human gastruloids [60]. Elements of the post-implantation human embryo can be found in Post-implantation Amniotic Sac Embryoids (PASE) [61, 62].

More complex *in vitro* models can be created by combining multiple cell types, including, in particular, extraembryonic cells. This approach has so far been explored only with mouse cells. After the derivation of mESCs from the mouse blastocyst, extraembryonic cell lines were isolated in similar ways. The trophoblast stem cell (TSC) were derived from the polar embryonic trophectoderm [63], and extraembryonic endoderm-like (XEN) cells were cultured from primitive endoderm cells of the blastocysts [64]. Those cell lines offer new possibilities for patterning ESC-based model systems. An important model in this category is the blastoid, grown from mESCs and TSCs [65], which mimics the blastocyst stage *in vitro*. This model brings amazing opportunities to study the interactions between those two cell types, which are involved in implantation and axis establishment [66, 67]. Two other models, using co-culture of embryonic and extraembryonic cells, were developed to recreate the peri-implantation stage *in vitro*. By aggregation of mESCs and TSCs with XEN cells (ETX model) or without them (ETS model), those models create peri-implantation embryo-like structures displaying EMT and gastrulation initiation [68, 69].

All those *in vitro* models of the peri-implantation mammalian embryo remove several limitations of *in vivo* studies and allow us to perform scalable and reproducible experimental perturbations. The ‘bottom-up’ approach aims to identify the minimal set of conditions needed to achieve the major milestones of embryonic development. Multiple combinations of factors can be tested to recreate a particular process and verify the mechanisms gleaned from *in vivo* observations [45]. Extensive perturbation studies will help us to discriminate between causality and coincidence of complex events observed *in vivo*. An important example for the power of ‘bottom-up’ approaches is the observation of symmetry breaking events, reminiscent of body axis formation, in gastruloids. Such observations can reveal which developmental processes can occur independently of the extraembryonic environment.

Overall, *in vitro* systems can be a rich source of information, especially for human development, since experimentation with human embryos is restricted. However, the development of *in vitro* models, getting closer and closer to actual development, brings new ethical questions [70, 71]. It will be a difficult but important task to determine, when we cross the line between “an aggregate of cells resembling an embryo” and an actual embryo.

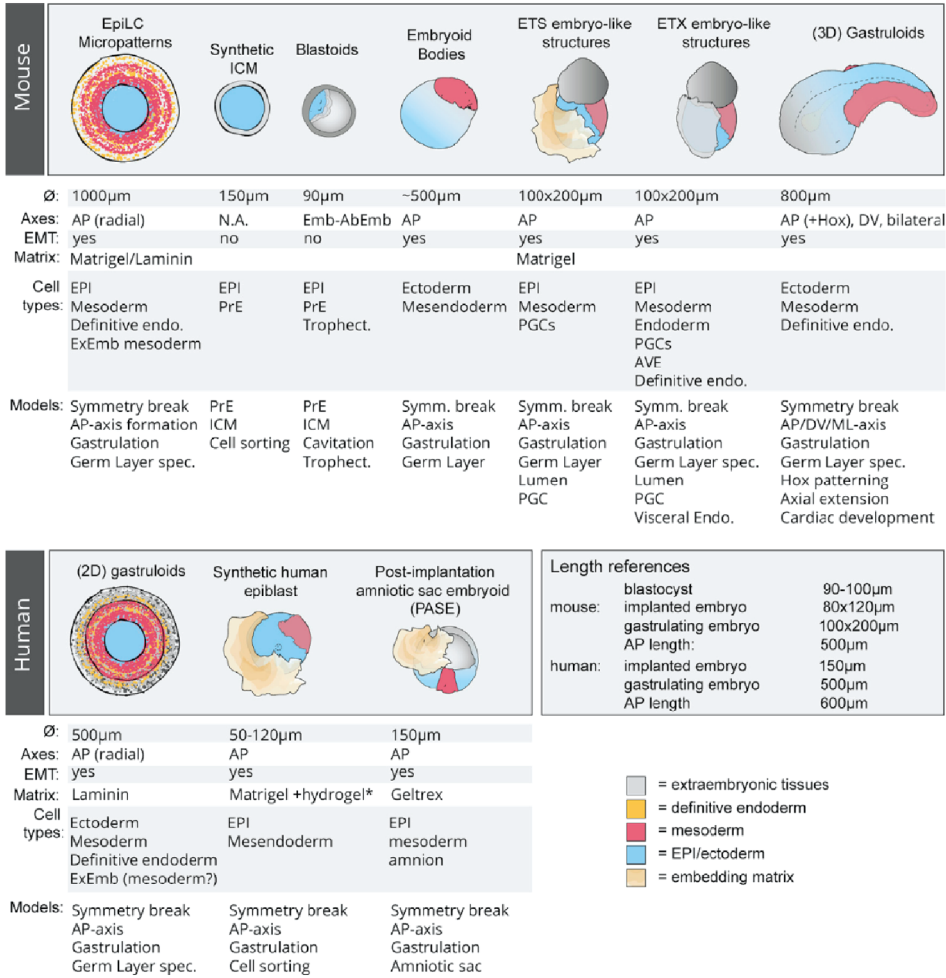


Figure 1.3: Current *in vitro* models of early mammalian development. Summary of the main self-organizing models of mouse (top row) and human (bottom row), and their characteristics. Figure from Vianello, S. & Lutolf, M. P. (2019), *Developmental Cell*³.

³Figure reprinted with permission from Elsevier. Licence number 5057690206560.

1.4. SUMMARY AND OUTLOOK

Top-down and bottom-up approaches constitute two different ways to study early mammalian development. *In vivo / ex vivo* observations of normal development bring a descriptive characterization of the embryo and experimental manipulation can reveal the molecular mechanisms. Despite their fundamental importance, such studies are restricted by ethical, biological, and technical limitations. In order to overcome those constraints, *in vitro* models based on pluripotent stem cells were developed as an alternative. The high level of control over *in vitro* systems brings better scalability, wider accessibility, and the possibility to tune a large number of parameters.

We believe that the two approaches of studying development are complementary and can work together towards a better understanding. The construction of synthetic *in vitro* systems mimicking embryos allows us to test and explore functional hypothesis informed by *in vivo* observations. Reciprocally, a set of factors identified as responsible for a certain process *in vitro* must be tested *in vivo*.

In the current thesis we will present several projects aiming at extending our understanding of different parts of mammalian development.

So far, most synthetic systems were focusing on the self-organizing and differentiating capacities of ESCs. Notwithstanding the importance of those remarkable abilities, the cells in a developing embryo also receive cues from extra-embryonic tissues, like the extra-embryonic endoderm. The addition of XEN cells to existing *in vitro* models like EBs or gastruloids might spur further ESC differentiation and/or organization. Chapter 2 and 3 present an *in vitro* model of mouse embryonic development, which combines mESCs and XEN cells in self-organizing aggregates, based on the gastruloid protocol. In chapter 2, we characterize those aggregates, called XEN enhanced-gastruloids (XEG), which display derivatives of all three germ layers with the striking organization of neural ectoderm-like cells in tubular structures. In chapter 3, we focus on the role and fate of the XEN cells in XEGs. We will demonstrate that XEN cells induce the formation of the neural tube-like structure by expression of the WNT-inhibitor DKK1, and the production of a basal membrane. We also report that XEN cells, when co-cultured with mESCs, differentiated themselves.

Model systems combining XEN cells and ESCs can also be used to better understand the mechanisms underlying XEN differentiation and localization *in vivo*. In chapter 4, we develop two different *in vitro* models using mESCs, which mimic the sorting of epiblast (Epi) and primitive endoderm (PrE) cells in the mouse blastocyst. The first model, chimeric embryoid body (Chim-EB) comprising mESCs and XEN cells, provides a simple system to explore the role of particular adhesion molecules in the sorting process. The second model, formed by the differentiation of PrE-like cells in EBs and embryoid disk (ED) of mESCs, allows us to track single PrE-like cells during the sorting process and characterize their movement. Finally, to draw any conclusion, results obtained from *in vitro* studies need to be tested and verified *in vivo*. Recently, single cell RNA sequencing has been used to characterize the cell type diversity of developing embryos. As an example, chapter 5 presents a detailed study of the human fetal kidney by single-cell

RNA sequencing and *in situ* marker expression analysis. It led us to the characterization of newly discovered progenitor cells, the pre-podocytes, and different subtypes of nephron progenitor cell (NPC).

ACRONYMS

Chim-EB chimeric embryoid body
E4.5 embryonic day 4.5
EB embryoid body
ED embryoid disk
EMT epithelial to mesenchymal transition
ESC embryonic stem cell

hESC human embryonic stem cell
iPSC induced pluripotent stem cell
mESC mouse embryonic stem cell
TSC trophoblast stem cell
XEG XEN enhanced-gastruloids
XEN extraembryonic endoderm-like

CELL TYPES

AVE anterior visceral endoderm
DVE distal visceral endoderm
Epi epiblast
ICM inner cell mass

NPC nephron progenitor cell
PrE primitive endoderm
VE visceral endoderm

1.5. REFERENCES

- [1] M. R. McGowen et al. “The evolution of embryo implantation.” In: *The International journal of developmental biology* 58.2-4 (2014), pp. 155–161.
- [2] M. F. Pera. “Human embryo research and the 14-day rule.” In: *Development* 144.11 (June 2017), pp. 1923–1925.
- [3] D. A. Turner, P. Baillie-Johnson, and A. Martinez Arias. “Organoids and the genetically encoded self-assembly of embryonic stem cells.” In: *BioEssays : news and reviews in molecular, cellular and developmental biology* 38.2 (Feb. 2016), pp. 181–191.
- [4] M. Simunovic and A. H. Brivanlou. “Embryoids, organoids and gastruloids: new approaches to understanding embryogenesis.” In: *Development* 144.6 (Mar. 2017), pp. 976–985.
- [5] M. N. Shahbazi and M. Zernicka-Goetz. “Deconstructing and reconstructing the mouse and human early embryo.” In: *Nature Cell Biology* 20.8 (Aug. 2018), pp. 878–887.
- [6] S. Vianello and M. P. Lutolf. “Understanding the Mechanobiology of Early Mammalian Development through Bioengineered Models.” In: *Developmental Cell* 48.6 (Mar. 2019), pp. 751–763.
- [7] J. Hammond. “Recovery and culture of tubal mouse ova.” In: *Nature* 163.4131 (Jan. 1949), p. 28.
- [8] W. K. Whitten. “Culture of tubal mouse ova.” In: *Nature* 177.4498 (Jan. 1956), pp. 96–96.
- [9] A. K. Tarkowski. “Experiments on the development of isolated blastomeres of mouse eggs.” In: *Nature* 184.4695 (Oct. 1959), pp. 1286–1287.
- [10] Y. C. Hsu. “In vitro development of individually cultured whole mouse embryos from blastocyst to early somite stage.” In: *Developmental biology* 68.2 (Feb. 1979), pp. 453–461.
- [11] I. Bedzhov et al. “In vitro culture of mouse blastocysts beyond the implantation stages.” In: *Nature protocols* 9.12 (Dec. 2014), pp. 2732–2739.
- [12] A. Aguilera-Castrejon et al. “Ex utero mouse embryogenesis from pre-gastrulation to late organogenesis.” In: *Nature* 593.7857 (May 2021), pp. 119–124.
- [13] R. L. Gardner. “Mouse chimeras obtained by the injection of cells into the blastocyst.” In: *Nature* 220.5167 (Nov. 1968), pp. 596–597.
- [14] A. K. Tarkowski. “Mouse chimaeras developed from fused eggs.” In: *Nature* 190.4779 (June 1961), pp. 857–860.
- [15] S. J. Kinder et al. “The organizer of the mouse gastrula is composed of a dynamic population of progenitor cells for the axial mesoderm.” In: *Development* 128.18 (Sept. 2001), pp. 3623–3634.

- [16] A. Martinez Arias and B. Steventon. “On the nature and function of organizers.” In: *Development* 145.5 (Mar. 2018).
- [17] B. H. Koller and O. Smithies. “Inactivating the beta 2-microglobulin locus in mouse embryonic stem cells by homologous recombination.” In: *Proceedings of the National Academy of Sciences* 86.22 (Nov. 1989), pp. 8932–8935.
- [18] P. Svoboda et al. “Selective reduction of dormant maternal mRNAs in mouse oocytes by RNA interference.” In: *Development* 127.19 (Oct. 2000), pp. 4147–4156.
- [19] F. Wianny and M. Zernicka-Goetz. “Specific interference with gene function by double-stranded RNA in early mouse development.” In: *Nature Cell Biology* 2.2 (Feb. 2000), pp. 70–75.
- [20] M. Jinek et al. “A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.” In: *Science* 337.6096 (Aug. 2012), pp. 816–821.
- [21] S. J. Arnold et al. “Pivotal roles for eomesodermin during axis formation, epithelium-to-mesenchyme transition and endoderm specification in the mouse.” In: *Development* 135.3 (Feb. 2008), pp. 501–511.
- [22] R. Hernández-Martínez, N. Ramkumar, and K. V. Anderson. “p120-catenin regulates WNT signaling and EMT in the mouse embryo.” In: *Proceedings of the National Academy of Sciences of the United States of America* 116.34 (Aug. 2019), pp. 16872–16881.
- [23] P. P. L. Tam and D. A. F. Loebel. “Gene function in mouse embryogenesis: get set for gastrulation.” In: *Nature Reviews Genetics* 8.5 (Mar. 2007), pp. 368–381.
- [24] S. Nowotschin et al. “The emergent landscape of the mouse gut endoderm at single-cell resolution.” In: *Nature* 569.7756 (Apr. 2019), pp. 361–367.
- [25] J. Delile et al. “Single cell transcriptomics reveals spatial and temporal dynamics of gene expression in the developing mouse spinal cord.” In: *Development* 146.12 (Mar. 2019).
- [26] M. Hochane et al. “Single-cell transcriptomics reveals gene expression dynamics of human fetal kidney development.” In: *PLoS biology* 17.2 (Feb. 2019), e3000152.
- [27] B. Pijuan-Sala et al. “A single-cell molecular map of mouse gastrulation and early organogenesis.” In: *Nature* 566.7745 (Feb. 2019), pp. 490–495.
- [28] J. Cao et al. “The single-cell transcriptional landscape of mammalian organogenesis.” In: *Nature* 566.7745 (Feb. 2019), pp. 496–502.
- [29] J. Cao et al. “A human cell atlas of fetal gene expression.” In: *Science* 370.6518 (Nov. 2020).
- [30] C. L. Gregg and J. T. Butcher. “Quantitative in vivo imaging of embryonic development: opportunities and challenges.” In: *Differentiation; research in biological diversity* 84.1 (July 2012), pp. 149–162.
- [31] P. Strnad et al. “Inverted light-sheet microscope for imaging mouse pre-implantation development.” In: *Nature Methods* 13.2 (Feb. 2016), pp. 139–142.
- [32] K. McDole et al. “In Toto Imaging and Reconstruction of Post-Implantation Mouse Development at the Single-Cell Level.” In: *Cell* 175.3 (Oct. 2018), 859–876.e33.

- [33] C. S. Simon et al. "Live Visualization of ERK Activity in the Mouse Blastocyst Reveals Lineage-Specific Signaling Dynamics." In: *Developmental Cell* (Oct. 2020).
- [34] R. Hiramatsu et al. "External mechanical cues trigger the establishment of the anterior-posterior axis in early mouse embryos." In: *Developmental Cell* 27.2 (Oct. 2013), pp. 131–144.
- [35] M. Nadjicka and N. Hillman. "Ultrastructural studies of the mouse blastocyst substages." In: *Journal of embryology and experimental morphology* 32.3 (Dec. 1974), pp. 675–695.
- [36] S.M. Meilhac et al. "Active cell movements coupled to positional induction are involved in lineage segregation in the mouse blastocyst". In: *Developmental biology* 331.2 (2009), pp. 210–221.
- [37] N. Saiz et al. "Atypical protein kinase C couples cell sorting with primitive endoderm maturation in the mouse blastocyst." In: *Development* 140.21 (Nov. 2013), pp. 4311–4322.
- [38] D. Mesnard, M. Guzman-Ayala, and D. B. Constam. "Nodal specifies embryonic visceral endoderm and sustains pluripotent cells in the epiblast before overt axial patterning." In: *Development* 133.13 (July 2006), pp. 2497–2505.
- [39] J. Artus, A. Piliszek, and A.-K. Hadjantonakis. "The primitive endoderm lineage of the mouse blastocyst: sequential transcription factor activation and regulation of differentiation by Sox17." In: *Developmental biology* 350.2 (Feb. 2011), pp. 393–404.
- [40] D. W. Stuckey et al. "Coordination of cell proliferation and anterior-posterior axis establishment in the mouse embryo." In: *Development* 138.8 (Apr. 2011), pp. 1521–1530.
- [41] T. Wu, A.-K. Hadjantonakis, and S. Nowotschin. "Visualizing endoderm cell populations and their dynamics in the mouse embryo with a Hex-tdTomato reporter." In: *Biology open* 6.5 (May 2017), pp. 678–687.
- [42] M. Madabhushi and E. Lacy. "Anterior visceral endoderm directs ventral morphogenesis and placement of head and heart via BMP2 expression." In: *Developmental Cell* 21.5 (Nov. 2011), pp. 907–919.
- [43] I. Bedzhov et al. "Development of the anterior-posterior axis is a self-organizing process in the absence of maternal cues in the mouse embryo". In: *Cell Research* 25.12 (Dec. 2015), pp. 1368–1371.
- [44] G. S. Kwon, M. Viotti, and A.-K. Hadjantonakis. "The endoderm of the mouse embryo arises by dynamic widespread intercalation of embryonic and extraembryonic lineages." In: *Developmental Cell* 15.4 (Oct. 2008), pp. 509–520.
- [45] J. Davies. "Using synthetic biology to explore principles of development." In: *Development* 144.7 (Apr. 2017), pp. 1146–1158.
- [46] G. R. Martin. "Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells." In: *Proceedings of the National Academy of Sciences* 78.12 (Dec. 1981), pp. 7634–7638.

- [47] T. C. Doetschman et al. “The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium.” In: *Journal of embryology and experimental morphology* 87 (June 1985), pp. 27–45.
- [48] S. M. Morgani et al. “Micropattern differentiation of mouse pluripotent stem cells recapitulates embryo regionalized cell fate patterning”. In: *eLife* 7 (Feb. 2018).
- [49] D. ten Berge et al. “Wnt Signaling Mediates Self-Organization and Axis Formation in Embryoid Bodies”. In: *Cell stem cell* 3.5 (Nov. 2008), pp. 508–518.
- [50] B. Mathew et al. “Mouse ICM Organoids Reveal Three-Dimensional Cell Fate Clustering.” In: *Biophysical journal* 116.1 (Jan. 2019), pp. 127–141.
- [51] S. C. van den Brink et al. “Symmetry breaking, germ layer specification and axial organisation in aggregates of mouse embryonic stem cells.” In: *Development* 141.22 (Nov. 2014), pp. 4231–4242.
- [52] D. Turner et al. “Gastruloids develop the three body axes in the absence of extraembryonic tissues and spatially localised signalling”. In: *bioRxiv* (2017), p. 104539.
- [53] L. Beccari et al. “Multi-axial self-organization properties of mouse embryonic stem cells into gastruloids.” In: *Nature* (Oct. 2018), p. 1.
- [54] S. C. van den Brink et al. “Single-cell and spatial transcriptomics reveal somitogenesis in gastruloids.” In: *Nature* 141 (Feb. 2020), pp. 4231–5.
- [55] J. V. Veenvliet et al. “Mouse embryonic stem cells self-organize into trunk-like structures with neural tube and somites.” In: *Science* 370.6522 (Dec. 2020).
- [56] G. Rossi et al. “Capturing Cardiogenesis in Gastruloids.” In: *Cell stem cell* (Nov. 2020).
- [57] J. A. Thomson et al. “Embryonic stem cell lines derived from human blastocysts.” In: *Science* 282.5391 (Nov. 1998), pp. 1145–1147.
- [58] A. Warmflash et al. “A method to recapitulate early embryonic spatial patterning in human embryonic stem cells.” In: *Nature Methods* 11.8 (Aug. 2014), pp. 847–854.
- [59] M. Simunovic et al. “A 3D model of a human epiblast reveals BMP4-driven symmetry breaking.” In: *Nature Cell Biology* 21.7 (July 2019), pp. 900–910.
- [60] N. Moris et al. “An in vitro model of early anteroposterior organization during human development.” In: *Nature* 582.7812 (June 2020), pp. 410–415.
- [61] Y. Shao et al. “A pluripotent stem cell-based model for post-implantation human amniotic sac development.” In: *Nature Communications* 8.1 (Aug. 2017), p. 208.
- [62] L. Yu et al. “Blastocyst-like structures generated from human pluripotent stem cells.” In: *Nature* 591.7851 (Mar. 2021), pp. 620–626.
- [63] S. Tanaka et al. “Promotion of trophoblast stem cell proliferation by FGF4.” In: *Science* 282.5396 (Dec. 1998), pp. 2072–2075.
- [64] T. Kunath et al. “Imprinted X-inactivation in extra-embryonic endoderm cell lines from mouse blastocysts.” In: *Development* 132.7 (Apr. 2005), pp. 1649–1661.

- [65] N. C. Rivron et al. “Blastocyst-like structures generated solely from stem cells”. In: *Nature* 557.7703 (May 2018), pp. 106–111.
- [66] B. Sozen et al. “Self-Organization of Mouse Stem Cells into an Extended Potential Blastoid.” In: *Developmental Cell* 51.6 (Dec. 2019), 698–712.e8.
- [67] J. Frias-Aldeguer et al. “Embryonic signals perpetuate polar-like trophoblast stem cells and pattern the blastocyst axis”. In: *bioRxiv* 129 (Jan. 2020), pp. 2733–29.
- [68] B. Sozen et al. “Self-assembly of embryonic and two extra-embryonic stem cell types into gastrulating embryo-like structures.” In: *Nature Cell Biology* 20.8 (Aug. 2018), pp. 979–989.
- [69] S. E. Harrison et al. “Assembly of embryonic and extra-embryonic stem cells to mimic embryogenesis in vitro.” In: *Science* (Mar. 2017), eaal1810.
- [70] M. Munsie, I. Hyun, and J. Sugarman. “Ethical issues in human organoid and gastruloid research.” In: *Development* 144.6 (Mar. 2017), pp. 942–945.
- [71] M. F. Pera et al. “What if stem cells turn into embryos in a dish?” In: *Nature Methods* 12.10 (Oct. 2015), pp. 917–919.

2

XEN ENHANCED-GASTRULOIDS, A NEW *in vitro* MODEL FOR STUDYING DEVELOPMENT BEYOND GASTRULATION

Recapitulating early mammalian embryonic development *in vitro* is a major challenge in biology. Recently developed *in vitro* models have exploited the self-organization potential of mouse ESCs (mESC), to successfully display hallmarks of gastrulation. However, these model systems fail to develop further or show tissue-level organization, unless specific signaling pathways are activated exogenously. In this chapter, we show that aggregates of mESC and extraembryonic endoderm (XEN) cells can self-organize to form neural tube-like structures and show signs of body axis formation. By single-cell RNA sequencing (scRNA-seq), we confirm the neural character of the tubes and demonstrate high cell type diversity of our model system. Imaging assays reveal partial patterning reminiscent of the anterior-posterior body axis. Our model system might thus be used to study the role of heterotypic cellular interactions between embryonic and extraembryonic cells in achieving complex, tissue-level organization.

2.1. INTRODUCTION

Organogenesis is a crucial and highly complex phase in mammalian development. One of the most intriguing examples of organogenesis is the formation of the neural tube, which occurs via two distinct mechanisms [1]. During primary neurulation, the main part of the neural tube is formed by the folding of the neural plate, an epithelial sheet of neural ectoderm. Secondary neurulation, which gives rise to the most posterior part of the neural tube, works differently: mesenchymal cells of the tail bud condense to form an epithelial rod which cavitates to form a tube [2, 3]. The two parts of the tube are then connected during junctional neurulation [4].

Much of the mechanistic understanding of this complex process has been derived from embryo cultures. Such cultures have been used extensively to study gastrulation [5, 6], as well as subsequent organogenesis [7]. Advances in light-sheet imaging of cultured embryos have started to reveal morphogenic events during neurulation in unprecedented detail [8]. Recently, stem cell-derived model systems have emerged, that recapitulate basic developmental processes *in vitro* with a minimal number of components [9, 10, 11]. Most of these systems model the earliest stages of development and are thus not suitable to study organogenesis. A prominent example is the gastruloid, which exhibits gastrulation-like events, such as embryo elongation and body axis formation [12, 13, 14, 15, 16, 17]. Despite the remarkable extent of symmetry breaking in gastruloids, these models do not yet produce tissue-level organization, unless specific cell fates are induced exogenously [18, 19].

Notably, gastruloids are made exclusively from embryonic stem cells (ESC). *In vivo*, mammalian embryos receive important patterning inputs from extra-embryonic cells [20]. Some existing *in vitro* models therefore incorporate such cells, but nevertheless only achieve resemblance with early embryonic stages that precede organogenesis [21, 18, 22, 23]. We hypothesized that adding extraembryonic cells to gastruloids might unlock additional developmental potential in this system. We therefore decided to add extraembryonic endoderm (XEN) cells. These cells are derived from the primitive endoderm (PrE), which overlays the developing epiblast before implantation (see Fig. 2.1 for a schematic of early mouse development.) After implantation, the PrE gives rise to the parietal endoderm (PE), which covers the inside of the blastocoel cavity [24] and the Visceral Endoderm (VE), which surrounds the embryo until the formation of the visceral yolk sac and its integration in the embryonic gut [25, 26]. A subpopulation of the VE, the Anterior Visceral Endoderm (AVE) has a role in the establishment of the embryo's body axes [27, 28]. Hence, XEN cells might have unexplored potential to pattern stem-cell based model systems.

Here we report that aggregates of mouse embryonic stem cells (mESC) and XEN cells, which we call XEN enhanced-gastruloids (XEG), can produce a central hallmark of organogenesis: columnar neural epithelia that resemble a neural tube. Using multiple markers, perturbation of the signaling pathways that play a role in neural tube development, and further differentiation to cortex-like tissue, we confirmed that the tubular epithelia are similar to neural tubes *in vivo*. Using single-cell RNA-seq (scRNA-seq), we

identified differences in composition and molecular profiles between XEGs and regular gastruloids. We then established that XEN-derived cells become visceral endoderm-like due to co-differentiation with the mESCs. Finally, we showed that XEN cells promote tube formation by local inhibition of primitive streak formation, likely via the WNT inhibitor DKK1, as well as through production of a basement membrane. Our study thus reveals a complex interplay between an embryonic and an extra-embryonic cell line and explores possible mechanisms underlying the great morphogenic potential of in vitro models of embryogenesis.

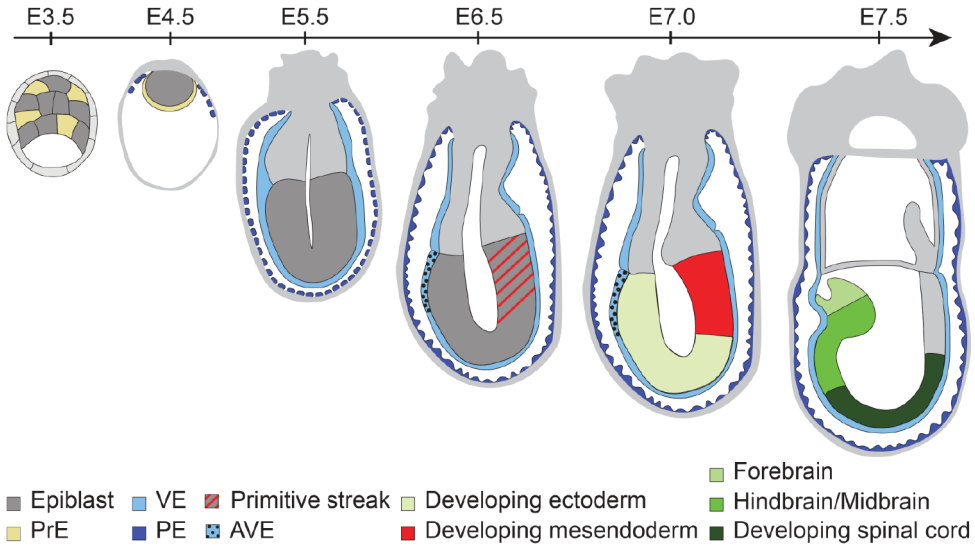


Figure 2.1: Schematic of early mouse embryonic development. Tissues modeled in XEGs are indicated with color. Tissues in grey are not present in XEGs. A: anterior, P: posterior, D: dorsal, V: ventral, L: left, R: right.

2.2. MATERIALS AND METHODS

2.2.1. CELL CULTURE

All cell lines were routinely cultured in KO DMEM medium (Gibco) supplemented with 10% ES certified FBS (Gibco), 0.1 mM 2-Mercaptoethanol (Sigma-Aldrich), 1×100 U/mL penicillin/streptomycin, $1 \times$ MEM Non-Essential Amino Acids (Gibco), 2 mM L-glutamine (Gibco), 1000 U/mL mouse LIF (ESGRO). Cells were passaged every other day and replated in tissue-culture treated dishes coated with gelatin. E14 mouse ES cells were provided by Alexander van Oudenaarden. The *Sox1^{GFPiresPac}* mouse ES cell line was created by Mario Stavridis and Meng Li in the group of Austin Smith [29] and provided by Sally Lowell. XEN and XEN-eGFP were provided by Christian Schröter [30]. All cell lines were regularly tested for mycoplasma infection. The ES mCherry-GPI cell line

was obtained by introducing a *mCherry-GPI* transgene in the *Pdgfra*^{H2B-GFP} cell line, provided by the group of Anna-Katerina Hadjantonakis [31].

2

2.2.2. GASTRULOIDS

The gastruloid differentiation protocol was adapted from van den Brink et al. [12]. ES cells were collected from tissue-culture treated dishes by trypsinization, gentle trituration with a pipet and centrifugation (1200 r.p.m., 3 min). After collection, cells were resuspended in 2 mL of freshly prepared, prewarmed N2B27 medium: DMEM/F12 (Life technologies) supplemented with 0.5 × N2 supplement (Gibco), 0.5 × B27 supplement (Gibco), 0.5 mM L-glutamine (Gibco), 1 × 100 U/mL penicillin/streptomycin (Gibco), 0.5 × MEM Non-Essential Amino Acids (Gibco), 0.1 mM 2-Mercaptoethanol (Sigma-Aldrich). Cells were counted to determine the cell concentration. For gastruloids, 200 ES cells were seeded in 40 µL of N2B27 in each well of a round-bottom low-adherence 96-well plate. 48 h after seeding, 150 µL of prewarmed N2B27 supplemented with 3 µM of GSK3 inhibitor (CHIR99021, Axon Medchem) was added to each well. 72 h after seeding, 150 µL of medium was removed from each well and replaced by 150 µL of preheated N2B27. Gastruloids were collected at 96 h after seeding and fixed with 4% paraformaldehyde (PFA, Alfa Aesar) overnight at 4°C.

2.2.3. XEN ENHANCED-GASTRULOIDS (XEGs)

ESCs and XEN cells were collected from tissue-culture treated dishes by trypsinization, gentle trituration with a pipet and centrifugation (1200 r.p.m., 3 min). After collection, cells were resuspended in 2 mL of fresh and prewarmed N2B27 medium. Cells were counted to determine cell concentration. For XEGs, several ratios of XEN and ESCs were tested (1:1, 1:2, 1:3, 1:4, 1:5) and compared with the regular gastruloid condition (0:1). The total number of cells was fixed at 200. Over two separate experiments, the proportion of organoids showing T staining and tubular structures was quantified (total number of embryonic organoids 1:1=179, 1:2=143, 1:3=143, 1:4=140, 0:1=130) and the optimal ratio was determined to be 1:3 (see Fig. 2.5a, b). A total of 200 cells (150 ES cells and 50 XEN cells) was seeded in 40 µL of N2B27 in each well of a round-bottom low-adherence 96-well plate. 48 h after seeding, 150 µL of prewarmed N2B27 supplemented with 3 µM of GSK3 inhibitor (CHIR99021, Axon Medchem) was added to each well. 72 h after seeding, 150 µL of medium was removed from each well and replaced by 150 µL of prewarmed N2B27. XEGs were collected at 96 h after seeding and fixed with 4% PFA overnight at 4°C.

2.2.4. CEREBRAL ORGANOIDs

Cerebral cortex-like tissue was created according to a protocol adapted from Lancaster et al., [32]. Instead of collecting XEGs at 96 h, the medium was replaced by cerebral

organoid differentiation medium: DMEM-F12 (Life technologies), Neurobasal (Gibco), 0.5 × B27 supplement containing vitamin A (Gibco), 0.5 × N2 supplement (Gibco), 2.5 μM/mL Insulin, 2mM L-glutamine (Gibco), 0.5 × MEM-Non-Essential Amino Acids (Gibco), 1 × 100 U/mL penicillin-streptomycin and 0.05 mM 2-Mercaptoethanol (Sigma-Aldrich). At 168 h, aggregates were collected and transferred, with fresh medium, into 10 cm dishes on an orbital shaker installed in the incubator (85 r.p.m.). Aggregates were grown until 192 h (8 days) during which medium was refreshed every other day until collection. Collected aggregates were fixed with 4% PFA for 48 h at 4°C.

2.2.5. IMMUNOSTAINING

FIXATION AND BLOCKING

After collection, gastruloids and XEGs were fixed in 4% PFA at 4°C overnight. Cerebral organoids were fixed under the same conditions, but for 48 h. After fixation, samples were washed three times in washing solution (PBS, 1% bovine serum albumin bovine serum albumin (BSA)) and incubated at 4°C in blocking buffer (PBS, 1% BSA, 0.3% Triton-X-100) for a minimum of 16 h. Samples for smFISH were washed 3 times in PBS after fixation and stored in 70% ethanol at 4°C. To stain E14 cells for pluripotency markers, cells in suspension were fixed for 30 min in 4% PFA at 4°C, washed three times in washing solution at RT and incubated in blocking buffer for 1 h at 4°C.

WHOLE-MOUNT IMMUNOLABELING AND CLEARING

Immunolabeling and clearing of gastruloids and XEGs were based on the protocol described by Dekkers et al., [33]. Briefly, after fixation and blocking, samples were incubated with primary antibodies at 4°C overnight on a rolling mixer (30 r.p.m.) in organoid washing buffer organoid washing buffer (OWB) (PBS, 2% BSA, 0.1% Triton-X-100) supplemented with 0.02% sodium dodecyl sulfate sodium dodecyl sulfate (SDS), referred to as OWB-SDS. The following primary antibodies were used: rat anti-SOX2 (1:200, 14-9811-82, Thermo Fisher Scientific), goat anti-T (1:200, sc-17745, Santa Cruz Biotechnology), goat anti-T (1:100, AF2085, R&D systems), mouse anti-DAB2 (1:100, 610464, BD Biosciences). The next day, samples were washed three times for 2 h in OWB-SDS at RT, followed by incubation with secondary antibodies (donkey anti-goat Alexa Fluor 488 (1:200, A-11055, Thermo Fisher Scientific), donkey anti-rat Alexa Fluor 488 (1:200, A-21208, Thermo Fisher Scientific), donkey anti-goat Alexa Fluor 555 (1:200, A-21432, Thermo Fisher), donkey anti-mouse Alexa Fluor 555 (1:200, A-31570, Thermo Fisher Scientific), chicken anti-rat Alexa Fluor 647 (1:200, A-21472, Thermo Fisher Scientific)) and 4',6-diamidino-2-phenylindole (DAPI, 1 μg/mL, Merck) in OWB-SDS at 4°C overnight on a rolling mixer (30 r.p.m.), protected from light. Finally, samples were washed three times for 2 h in OWB-SDS at RT. Clearing was performed by incubation in fructose-glycerol clearing solution (60% vol/vol glycerol, 2.5 M fructose) for 20 min at RT. Samples were imaged directly after clearing or stored at 4°C in the dark.

CRYOSECTIONING AND IMMUNOLABELING OF SECTIONS

Prior to cryosectioning, fixed and blocked samples were incubated sequentially in sucrose solutions (10, 20 and 30%) for 30 min (gastruloids and XEGs) or 2 h (cerebral organoids) at 27°C, and embedded in optimal cutting temperature optimal cutting temperature compound (OCT) compound. Samples in OCT were placed on dry ice for rapid freezing, and stored at -80°C prior to cryosectioning. Samples were cut to cryosections (10 μm thickness) using a cryostat (Thermo Fisher Scientific, USA) and cryosections were placed on poly-L-lysine coated glass slides (Merck). The slides were stored directly at -80°C. For immunofluorescence staining, slides were thawed and rinsed with PBS for 10 min at RT to dissolve the OCT. Subsequently, slides were incubated overnight at 4°C with the following primary antibodies diluted in blocking buffer: rat anti-SOX2 (1:200, 14-9811-82, Thermo Fisher Scientific), goat anti-T (1:200, sc-17745, Santa Cruz Biotechnology), mouse anti-N-cadherin (1:200, 33-3900, Thermo Fisher Scientific), rabbit anti-E-cadherin (1:200, 3195, Cell Signaling Technology), rabbit anti-PAX6 (1:100 (cerebral organoids) or 1:200 (gastruloids, XEGs), 42-6600, Thermo Fisher Scientific), mouse anti-NKX6.1 (1:200, F55A12, Developmental Studies Hybridoma Bank), rabbit anti-NKX6.1 (1:200, HPA036774, Merck), mouse anti-TUJ1 (1:200, 801202, BioLegend), rabbit anti-CD31 (1:50, ab28364, Abcam), rabbit anti-GATA6 (1:200, PA1-104, Thermo Fisher Scientific), goat anti-GATA6 (1:200, AF1700, R&D Systems), rabbit anti-Laminin (1:200, PA1-16730, Thermo Fisher Scientific), mouse anti-OCT4 (1:200, MA1-104, Thermo Fisher Scientific). The next day, the slides were washed twice for 10 min in PBS at RT. Subsequently, the slides were incubated with secondary antibodies (donkey anti-goat Alexa Fluor 488 (1:200, A-11055, Thermo Fisher Scientific), donkey anti-rat Alexa Fluor 488 (1:200, A-21208, Thermo Fisher Scientific), donkey anti-goat Alexa Fluor 555 (1:200, A-21432, Thermo Fisher), donkey anti-mouse Alexa Fluor 555 (1:200, A-31570, Thermo Fisher Scientific), chicken anti-rat Alexa Fluor 647 (1:200, A-21472, Thermo Fisher Scientific), donkey anti-rabbit Alexa Fluor 647 (1:200, A-31573, Thermo Fisher Scientific)) and DAPI (1 $\mu\text{g}/\text{mL}$, Merck) in blocking buffer for 4 h at 4°C, and washed three times for 10 min at RT. Slides were mounted in ProLong™ Gold Antifade Mountant (Thermo Fisher Scientific) and imaged after 24-48 h.

IMMUNOLABELING OF E14 CELLS

After fixation and blocking, E14 cells were incubated with the following primary antibodies in blocking buffer overnight at 4°C: rat anti-SOX2 (1:200, 14-9811-82, Thermo Fisher Scientific) and mouse anti-OCT4 (1:200, MA1-104, Thermo Fisher Scientific). The next day, cells were washed three times in washing solution for 5 min at RT and incubated with secondary antibodies (donkey anti-rat Alexa Fluor 488 (1:200, A-21208, Thermo Fisher Scientific) and donkey anti-mouse Alexa Fluor 555 (1:200, A-31570, Thermo Fisher Scientific)) and DAPI (1 $\mu\text{g}/\text{mL}$, Merck) in blocking buffer for 3 h at 4°C. Finally, the cells were washed three times in washing solution for 5 min at RT and imaged directly.

2.2.6. SMFISH STAINING

smFISH was performed as described previously [34]. Briefly, samples were fixed

with PFA and stored in 70% ethanol, as described above. Custom designed smFISH probes for *Dab2*, *Fst*, *Hhex* and *Spink1* (BioCat), labeled with Quasar 570, CAL Fluor Red 610, or Quasar 670, were incubated with the samples overnight at 30°C in hybridization buffer (100 mg/mL dextran sulfate, 25% formamide, 2× SSC, 1 mg/mL E.coli tRNA, 1 mM vanadyl ribonucleoside complex, 0.25 mg/mL BSA; Thermo Fisher Scientific). Samples were washed twice for 30 min at 30 °C with wash buffer (25% formamide, 2× SSC). The wash buffer was supplemented with DAPI (1 μg/mL) in the second wash step. All solutions were prepared with RNase-free water. Finally, the samples were mounted in ProlongGold (Life Technologies) and imaged when hardened (sections) or immediately (ibidi dishes). All components are from Sigma-Aldrich unless indicated.

smFISH samples were imaged on a Nikon Ti-Eclipse epifluorescence microscope equipped with an Andor iXON Ultra 888 EMCCD camera and dedicated, custom-made fluorescence filter sets (Nikon) using a 100× / 1.3 Super Fluor oil-immersion objective (Nikon) were used.

Z-stacks were collected with a distance of 0.2 μm between planes in four fluorescence channels (DAPI, Quasar 570, CAL Fluor Red 610, Quasar 670) using a 100× /1.45 Plan Apo Lambda oil (Nikon) objective.

2.2.7. IMAGING

FIXED SAMPLES

Fixed and stained samples were imaged on a Nikon Ti-Eclipse epifluorescence microscope equipped with an Andor iXON Ultra 888 EMCCD camera and dedicated, custom-made fluorescence filter sets (Nikon). Primarily, a 10× / 0.3 Plan Fluor DLL objective, a 20× / 0.5 Plan Fluor DLL objective, or a 40× / 1.3 Super Fluor oil-immersion objective (Nikon) were used. To image complete sections of cerebral organoids, multiple adjacent fields of view were acquired and combined using the tiling feature of the NIS Elements software (Nikon). Z-stacks were collected of whole-mount gastruloids and XEGs with distances of 10 μm between planes.

LIVE CELL IMAGING

Time lapses to observe the tubular structures formation were performed 24 h and 48 h after cell seeding, on XEGs grown from the mCherry-GPI ES cell line. XEGs were transferred to a glass-bottom μ-Slide imaging chamber (ibidi) and imaged every 30 min for 24 h, with a Nikon Eclipse Ti C2+ confocal laser microscope (Nikon, Amsterdam, The Netherlands), equipped with lasers at wavelengths 408, 488 and 561, an automated stage and perfect focus system at 37°C and 5% CO₂. Images for 3D cultures were acquired with a Nikon 20× Dry Plan Apo VC NA 0.75 objective.

To track SOX1 expression in gastruloids and XEGs during the 24 h growth after the GSK3 inhibitor pulse, 72 h gastruloids and XEGs grown from the Sox1GFPiresPac ES cell line were transferred to a glass-bottom μ-Slide imaging chamber (ibidi) and imaged every 40 min for 24 h, while temperature and CO₂ levels were maintained at 37°C and 5%, respectively, by a stage top incubator (INUG2-TIZW-SET, Tokai Hit) mounted on the

Nikon Ti-Eclipse epifluorescence microscope.

2.2.8. IMAGE ANALYSIS

Image stacks of whole-mount immunostained gastruloids and XEGs, and images of immunostained sections were pre-processed by background subtraction (rolling ball, radius: 50 pixels = 65 μm (10 \times objective), 32 μm (20 \times objective) or 16 μm (40 \times objective)) in the channels that showed autofluorescent background using ImageJ [35]. When background subtraction in images of sections did not result in proper removal of autofluorescent background signal, the Enhance Local Contrast enhance local contrast (CLAHE) tool was used in ImageJ [35]. smFISH image stacks were pre-processed by applying a Laplacian of a Gaussian filter ($\alpha = 1$) over the smFISH channels using scikit-image (v0.16.1) [36]. For all image stacks, a maximum projection was used to obtain a 2D representation. To show a single object per image, images were cropped around the object of interest.

2.2.9. SIGNALING PERTURBATION EXPERIMENTS

In the signaling experiments with XEGs, aggregates were treated between 72 h and 96 h with either LDN193189 (BMPi, 100 nM, Reagents Direct), a potent BMP pathway inhibitor, Purmorphamine (1 μM , STEMCELL Technologies), a small molecule agonist of the hedgehog pathway, Retinoic acid (RA, 100 nM, Sigma-Aldrich) or DMSO (0.1% final concentration, Sigma Aldrich) as a vehicle control. For this experiment, the XEGs were allowed to grow for an additional 48 h before fixation (144 h total growth) and preparation for staining (see Immunostaining).

2.2.10. SINGLE CELL RNA SEQUENCING

LIBRARY PREPARATION AND SEQUENCING

For each replicate, 96 pooled gastruloids and 96 pooled XEGs were collected from a round-bottomed low-adherence 96-well plate in 15 mL Falcon tubes and pelleted by gentle centrifugation (500 r.p.m. for 2 min). No final aggregate was excluded from the collection. After washing with cold PBS, samples were resuspended in N2B27. Cells were then dissociated by 5 min incubation in TrypLE (Gibco) and gentle trituration with a pipet, centrifuged and resuspended in 1 mL of cold N2B27. Cells were counted to determine cell number and viability. For the first replicate, ES-mCherry-GPI were spiked in at a frequency of 5%. For the second replicate, E14 cells were collected from culture dishes and incubated for 30 min at 4°C with CITE-seq cell hashing [37] antibody Ab_CD15 (1:200) (Biolegend). XEN-eGFP were collected from culture plates and incubated for 30 min at 4°C with CITE-seq cell hashing antibody Ab_CD140 (1:200) (Biolegend). In the gastruloid sample, labeled E14 cells were spiked in at a frequency of 5%, whereas in the neuruloid sample labeled E14 and XEN-eGFP were spiked in, both at a

frequency of 5%. High viability of the cells in all samples was confirmed before 10X library preparation. Single-cell RNA-seq libraries were prepared using the Chromium Single Cell 3' Reagent Kit, Version 3 Chemistry (10× Genomics) according to the manufacturer's protocol. CITE-seq libraries were prepared according to the CITE-seq protocol from New York Genome Center version 2019-02-13. Libraries were sequenced paired end on an Illumina. Novaseq6000 at 150 base pairs.

SINGLE-CELL RNA-SEQ DATA PRUNING AND NORMALIZATION

Cells with a low number of transcripts were excluded from further analysis based on the histograms in Supplementary Fig. 2.13a (count < 1300 for replicate 1 of the neuruloid experiment and count < 2300 for the other datasets). Genes expressed in less than 2 cells (across merged replicates) were excluded from further analysis. The final neuruloid dataset contains 14286 genes and 4591 or 6857 cells for replicate 1 or 2, respectively. The gastruloid dataset contains 14384 genes and 4233 or 8363 cells per replicate. The two datasets were normalized using the scran R-package (V 1.10.2, [38]). Gene variabilities were calculated (improvedCV2, scran) for each replicate separately, after excluding ribosomal genes [Ribosomal Protein Gene Database, <http://ribosome.med.miyazaki-u.ac.jp/>], exogenously expressed genes and genes expressing the antibodies used for CITE-seq. The 10% most highly variable genes (HVG) were selected based on variability p-values.

DIMENSIONALITY REDUCTION

For each of the two datasets, the two replicates were batch corrected with the fast mutual nearest neighbors mutual nearest neighbors (MNN) method implemented in the scran R-package [39], using the union of the 10% HVG of the two replicates and log-transformed normalized counts with $d = 120$ (number of principal components) and $k = 50$ (number of nearest neighbours). For dimensionality reduction, a uniform manifold approximation and projection (UMAP) was calculated on the batch corrected data using the R-package umap (V 0.2.3.1, [40].) with $n = 50$, $\text{min_dist} = 0.7$ and using the cosine distance measure.

IDENTIFICATION OF SPIKE-IN CELLS

Cells with any expression of mCherry were annotated as ES (mCherry+). The remaining spike-in cells, E14 (CD15+) and XEN spike-in (CD140+) (see Single-cell RNA-seq library preparation and sequencing), could not be determined by the expression level of the antibody alone. We therefore chose to assign spike-ins based on clusters. For each of the two datasets, a shared nearest neighbor graph was constructed from the batch corrected data (see Dimensionality reduction) with scran using $k = 20$ and $d = 30$. Louvain clustering was performed on the constructed graphs with the R-package igraph (V1.2.4.1, Blondel:2008do), which resulted in 8 clusters for XEGs and 7 clusters for gastruloids (see Supplementary Fig. 2.13c). We identified 3 out of the 8 clusters in XEGs based on literature markers and spike-in gene expression. One cluster out of these three was mainly comprised of mESCs, due to high Ab_CD15 expression and mCherry positive cells. Cells that had an expression of Ab_CD15 > 50 and were part of this cluster were considered spiked-in E14 and annotated as E14 (CD15+). The other two clusters were both eGFP

positive, where one of them had a higher Ab_CD140 expression and was thus annotated as XEN spike-in (Ab_CD140+). The second cluster was annotated as XEN derived (Ab_CD140-). Similarly, for gastruloids, one of the 7 clusters was comprised of mainly mESCs based on literature markers and spike-in gene expression. Cells that had an expression of Ab_CD15 > 100 and were part of this cluster were considered spiked-in E14 and annotated as E14 (Ab_CD15+).

ANALYSIS OF CELL CYCLE AND STRESS-RELATED GENES

For each of the two datasets, cell cycle analysis was performed with the *scraper* package using the *cyclone* function [41] on the normalized counts. Cells in G2M phase were distributed evenly across all clusters and thus the clustering was not biased by cell cycle. No other separate cluster that consisted entirely of cell cycle related cells appeared.

For the analysis of stress-related genes, a list of known stress genes [42] was used to calculate the average standardized expression per cell based on normalized counts. Stress-related genes were mainly found within the spike-in cells and there was no other separate cluster that consisted entirely of highly stressed cells.

MAPPING TO *IN VIVO* DATASETS

The Pijuan-Sala et al. dataset [43], which was downloaded from https://content.cruk.cam.ac.uk/jmlab/atlas_data.tar.gz, consists of 9 timepoints from E6.5 to E8.5. The data was normalized by size factors provided by the authors. Cells with no cell type assignment were excluded from further analysis. The 10% HVG were calculated (improvedCV2, *scraper* package) on the remaining cells excluding sex genes, similar to Pijuan-Sala et al.'s method. Cells in the "mixed_gastrulation" cluster were also excluded. MNN mapping was applied to log-transformed normalized counts of the 10% HVG. First, *in vivo* timepoints were mapped to each other in decreasing order. Then, each of our four datasets was mapped separately to the combined Pijuan-Sala et al. dataset (MNN method with $d = 120$, $k = 50$). K-nearest-neighbor k-nearest neighbors (knn) assignment was performed in the batch corrected principal component space. For each cell in our datasets, the 50 nearest neighbors in the *in vivo* dataset, based on Euclidean distances, were calculated. Each cell was assigned the most abundant cell type within the knn, if certain distance and confidence score conditions were met. This confidence score was calculated for each cell as the number of the most abundant cell type divided by the total number of neighbors ($k=50$). A cell was annotated as "Not assigned" if either, the average distance to its nearest neighbor exceeded a certain threshold (determined by the long tail of the histogram of average distances for each of our datasets separately) or the assignment had a confidence score less than 0.5. Additionally, we placed cells in "Not assigned" if they were assigned to clusters with less than 10 cells, or to the cluster "Blood progenitors 2" (because this cluster did not show distinct expression of known literature markers). This resulted in 22 assigned clusters for XEGs and 15 assigned clusters for gastruloids. For each cell in our dataset we calculated the average and the standard deviation of the developmental age of the knn.

The Delile et al. dataset `Delile:2019go`, which was downloaded from <https://github.com/juliendelile/MouseSpinalCordAtlas>, consists of 5 timepoints from E9.5 to E13.5. Cells that had a cell type assignment of "Null" or "Outlier" were excluded from

further analysis. The data was normalized (scran) and the 10% HVG were calculated. First, MNN was applied to the Delile et al. dataset in order of increasing timepoints (log-transformed normalized counts of the 10% HVG, $d = 120$, $k = 50$). Then, we mapped “Spinal cord” cells to the MNN corrected Delile et al. dataset separately for each of our replicates. knn assignment was performed as described above and resulted in 3 clusters for XEGs and 3 clusters for gastruloids.

DIFFERENTIAL EXPRESSION ANALYSIS

For the differential expression test between “spike-in XENs” and “XENs in XEGs” a Welch t-test (implemented in findMarkers, scran R package) was conducted on the normalized log-transformed counts. The test was performed on XEGs from replicate 2. “spike-in XENs” were chosen as the 100 cells with highest Ab_CD140 expression and “XENs in XEGs” were the 100 cells with lowest Ab_CD140 expression within the XEN identified cells.

For the differential expression test between XEGs and gastruloids, a negative binomial regression was performed (R package edgeR V 3.24.3 [44]). Based on the knn assignment to the Pijuan-Sala et al. dataset, all cells annotated as “Spinal cord” were extracted from our four datasets (in XEGs 859 cells in replicate 1 and 166 cells in replicate 2, in gastruloids 2071 cells in replicate 1 and 1882 cells in replicate 2). Raw counts were used for the regression with these four subsets as dummy variables and a variable corresponding to the total number of counts per cell. P-values were obtained for the contrast between XEGs and gastruloids using the average regression coefficients among variables of both replicates. Similarly, for the differential expression test of the “Spinal cord” in XEGs, a negative binomial regression was used. Cells were excluded from the test if either their cell type occurred in less than 10 cells per replicate, or if the cells were annotated as “Not assigned”, leaving a total of 13 cell types (7742 cells) to be considered. For each cell type and each replicate a dummy variable was created and a variable corresponding to the total number of counts per cell. Then, p-values were obtained for the contrast between the average regression coefficients of the two replicates of the “Spinal cord” cluster and the average regression coefficients of all other variables considered in the test.

For all differential expression tests p-values were adjusted for multiple hypothesis testing with the Benjamini-Hochberg method.

DORSO-VENTRAL MARKERS SCATTER PLOTS

For the scatter plots, we used once the average log-transformed normalized expression of 10 canonical dorsal markers (*Pax7*, *Msx3*, *Pax3*, *Gsh2*, *Msx2*, *Olig3*, *Math1*, *Zic1*, *Msx1*, *Zic2*) and the average log-transformed normalized expression of 10 canonical ventral markers (*Olig2*, *Nkx6.1*, *Nkx2.2*, *Foxa2*, *Ncam1*, *Nrcam*, *Nkx6.2*, *Isl1*, *Fabp7*, *Hb9*). The smoothed line was calculated by sorting the average expression of ventral markers and then using a sliding window with a step size of 50 on the average expression of dorsal markers. The second scatterplot was calculated similarly by using 10 genes that correlated the most with *Msx1* based on spearman correlation and the 10 genes that correlated the most with *Nrcam*.

2.3. RESULTS

2.3.1. XEN CELLS INDUCE NEURAL TUBE-LIKE STRUCTURES IN GASTRULOIDS

We first implemented the original mouse gastruloid protocol [12] in which mESCs are aggregated in N2B27 media and exposed to a pulse of WNT signaling for 24 h. After 96 h, this protocol results in elongated gastruloids. As reported before [12, 13, 14], 96 h gastruloids contained localized primitive streak-like and neural progenitor-like compartments, marked by Brachyury (T) and SOX2, respectively (Fig. 2.2b, inset). Starting from the gastruloid protocol, we developed a new model system by aggregating mESCs and XEN cells, keeping all other experimental conditions the same (Fig. 2.2a). We call our mixed aggregates XEN enhanced-gastruloids (XEGs). Like gastruloids, 96 h XEGs showed an elongated morphology and localized T-positive and SOX2-positive compartments. However, unlike gastruloids, in XEGs SOX2⁺ cells were organized in columnar epithelia surrounding one or several lumina (Fig 2.2b).

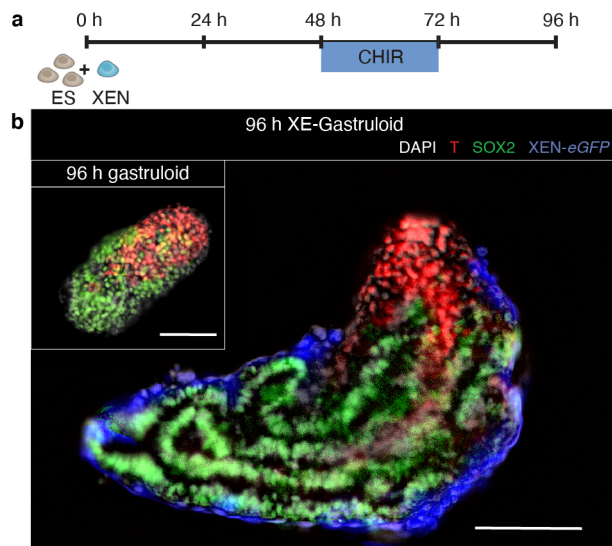


Figure 2.2: Co-aggregation of mESCs and XEN cells produce tubular structures. **a**, Schematic of the culture protocol: at 0 h, 200 cells (150 ESCs and 50 XEN cells) were aggregated; CHIR was added between 48 h and 72 h after cell seeding; cell aggregates were cultured until 96 h. **b**, T and SOX2 expression in aggregates at 96 h (z-projection of whole mount immunostaining). Inset: expression in aggregate resulting from the standard gastruloid protocol (without XEN cells). Scale bars 100 μ m. Cell nuclei were stained with DAPI.

Expression of the broadly expressed neural marker SOX2 and the striking morphology suggested that the observed structures resemble neural tubes. The lack of pluripo-

tency marker expression (Fig. 2.3a) excluded that the structures were formed by remaining undifferentiated cells. The presence of N-cadherin and absence of E-cadherin in the tubes (Fig. 2.3b, top) is consistent with the known switch from E- to N-cadherin during neural differentiation *in vivo* [45] and *in vitro* [46]. We could also observe that the cells in the tubes were polarized and expressed apical markers ZO-1 and α PKC, consistent with neural tubes *in vivo* [47] (Fig. 2.3b, bottom). Finally, we detected the neural progenitor markers PAX6 and NKX6.1 [48] in a subpopulation of tubular cells (Fig. 2.3c). Combined, these results suggest that the tubular structures in XEGs have neural tube characteristics.

To understand how neural tube-like structures formed, we used time-lapse microscopy of developing XEGs. Around 48 h after seeding, cells formed rosette-like shapes (Fig. 2.4a), which resembled structures found in Matrigel-embedded mESCs [49, 50] and indicated a mesenchymal-epithelial transition. Subsequently, a columnar epithelium was formed. Then lumina opened at different places and merged between 48 h and 72 h (Fig. 2.4b). During the final 24 h, the epithelium kept extending and differentiated further, as revealed by the expression of the neural progenitor marker SOX1 [51] (Fig. 2.4c). A SOX1 positive cell population also appeared in gastruloids within the same time frame, but, importantly, remained unorganized (Fig. 2.4c).

To explore the robustness of the protocol and identify optimal conditions for the formation of tubular structures, we tested different ratios of mESCs and XEN cells (Fig. 2.5a, b). Interestingly, even the smallest proportion of XEN cells tested (1:5), was able to induce some tubular structure formation. On the other hand, elongation and symmetry breaking were inhibited when the proportion of XEN cells exceeded 1:2. A ratio of 1:3 gave optimal results, with the concurrence of SOX2-positive tubes and T-positive cells in nearly all aggregates.

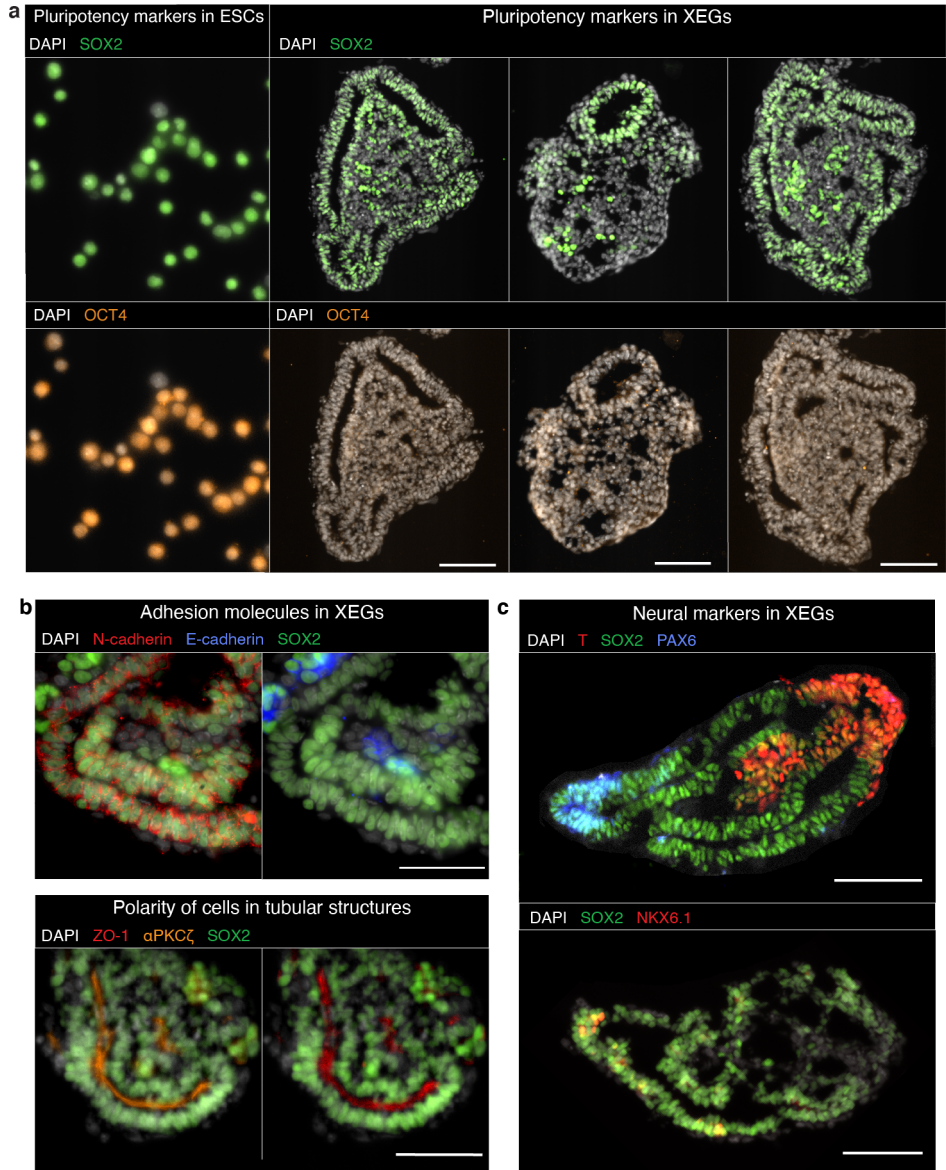


Figure 2.3: Tubular structures in XEGs have neural characteristics. **a**, Expression of SOX2 and OCT4 (immunostaining) in sections of XEGs at 96 h (left, scale bars: 100 μ m) and cultured ESCs (right, scale bars: 10 μ m). **b**, Expression of SOX2, E-cadherin, N-cadherin, ZO-1 and α PKC ζ in aggregates at 96 h (immunostaining of cryosections). Scale bar 50 μ m. **c**, Expression of T, SOX2, PAX6, MASH1 and NKX6.1 (immunostaining) in 96 h aggregates. Scale bars 100 μ m. Cell nuclei were stained with DAPI.

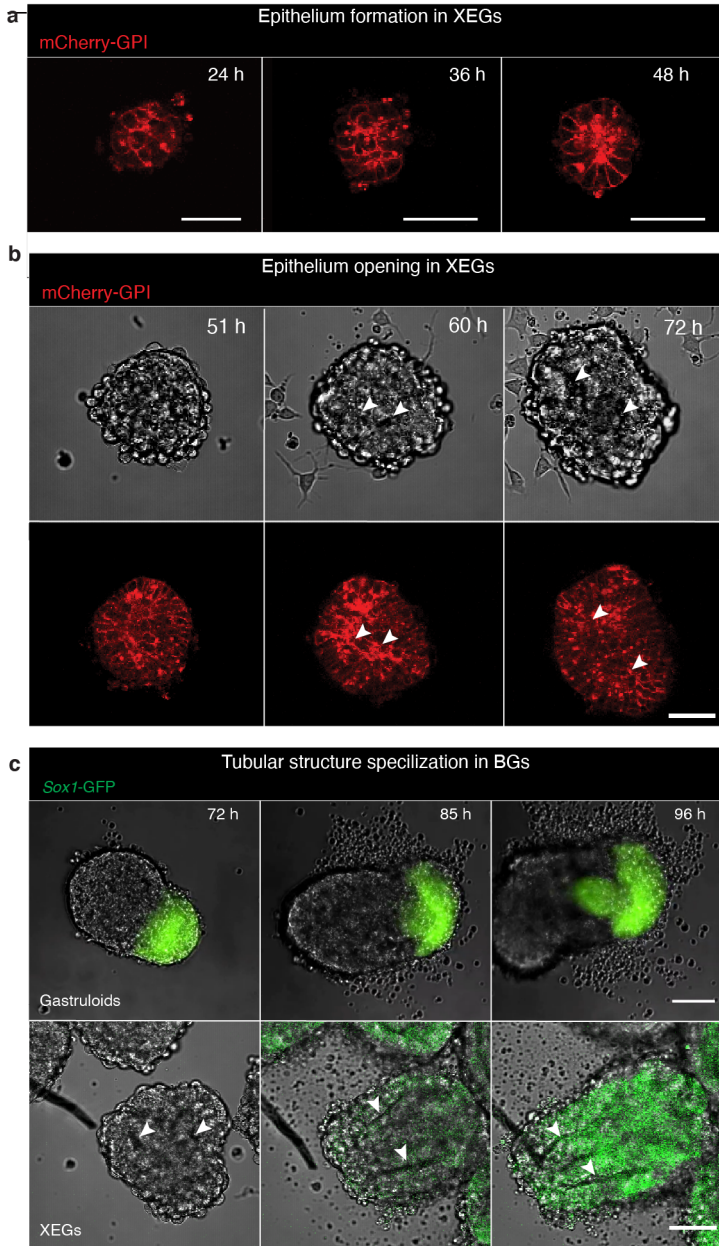


Figure 2.4: The formation of tubular structures in XEGs resemble secondary neurulation. **a,b,** Live cell imaging of morphological changes in XEGs grown from mCherry-GPI expressing mESCs. mCherry-GPI is localized to the cell membrane. **a,** Formation of rosettes **b,** Cavitation of rosettes. Arrowheads indicate cavitation events. Scale bars: 50 μm . **c,** Live cell imaging of SOX1 expression in gastruloids (top panel, scale bars 20 μm) and XEGs (lower panel, scale bars 50 μm) grown with *Sox1*-GFP mESCs. The arrows indicate the tubular structures. Scale bars 50 μm .

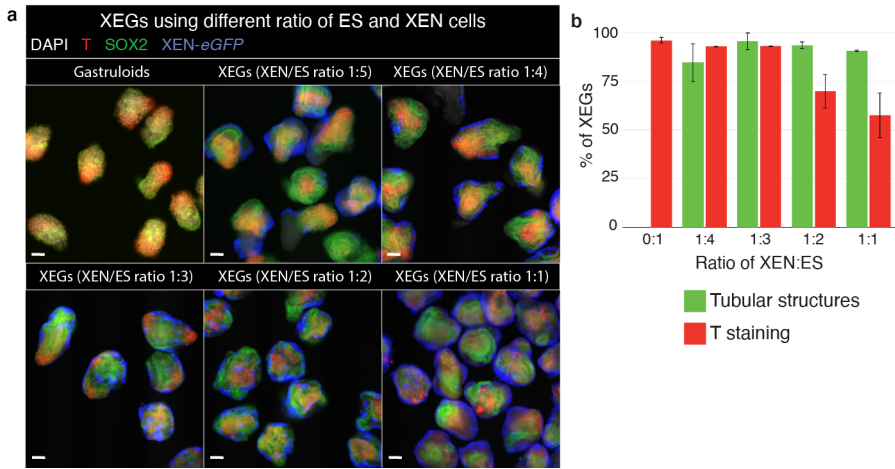


Figure 2.5: The formation rate of the tubular structures depends on the ratio of mESCs and XEN cells. a, SOX2 (neural progenitors-like cells) and T (primitive streak-like cells) expression in gastruloids and XEGs with different starting ratios of mESCs and XEN cells (z-projection of whole mount immunostaining). Scale bars: 100 μ m. Cell nuclei were stained with DAPI. **b,** Average fraction of aggregates showing tubular structures and T staining at 96 h for different starting ratios of mESCs and XEN cells (n = 2 experiments, error bars show standard deviation).

2.3.2. TUBES IN XEGs SHOW PARTIAL ANTEROPOSTERIOR AND DORSOVENTRAL PATTERNING

To establish, how closely XEGs mimic *in vivo* development, we looked for patterning related to body axis formation. Like in gastruloids, expression of T (Fig. 2.2b), revealed a posterior-like compartment. Localized expression of anterior markers (ASCL1, PAX6, see Fig. 2.6a and Fig. 2.3c) indicated an anterior-like compartment, although important canonical markers of the most anterior part of the embryo (OTX2, LEFTY1, EN1, ZIC1) could not be detected in XEGs (data not shown). The presence of anteroposterior patterning was supported by the expression of *Wnt4*, *Wnt8a* and *Fgf8* as gradients along the long axis of XEGs (Fig. 2.6b), which resembles the pattern found *in vivo* [52].

Tubes in XEGs also expressed markers of both the dorsal (PAX6, MSX1) and ventral (NKX6.1, NRCAM) neural tube (Fig. 2.6c). While PAX6 and MSX1 were found in cells close to the outside of the XEG (which are usually in contact with the XEN cells), NKX6.1 and NRCAM were expressed broadly throughout the tubes. In summary, 96 h XEGs show some indications of anteroposterior antero-posterior (AP) and dorsoventral dorso-ventral (DV) symmetry breaking.

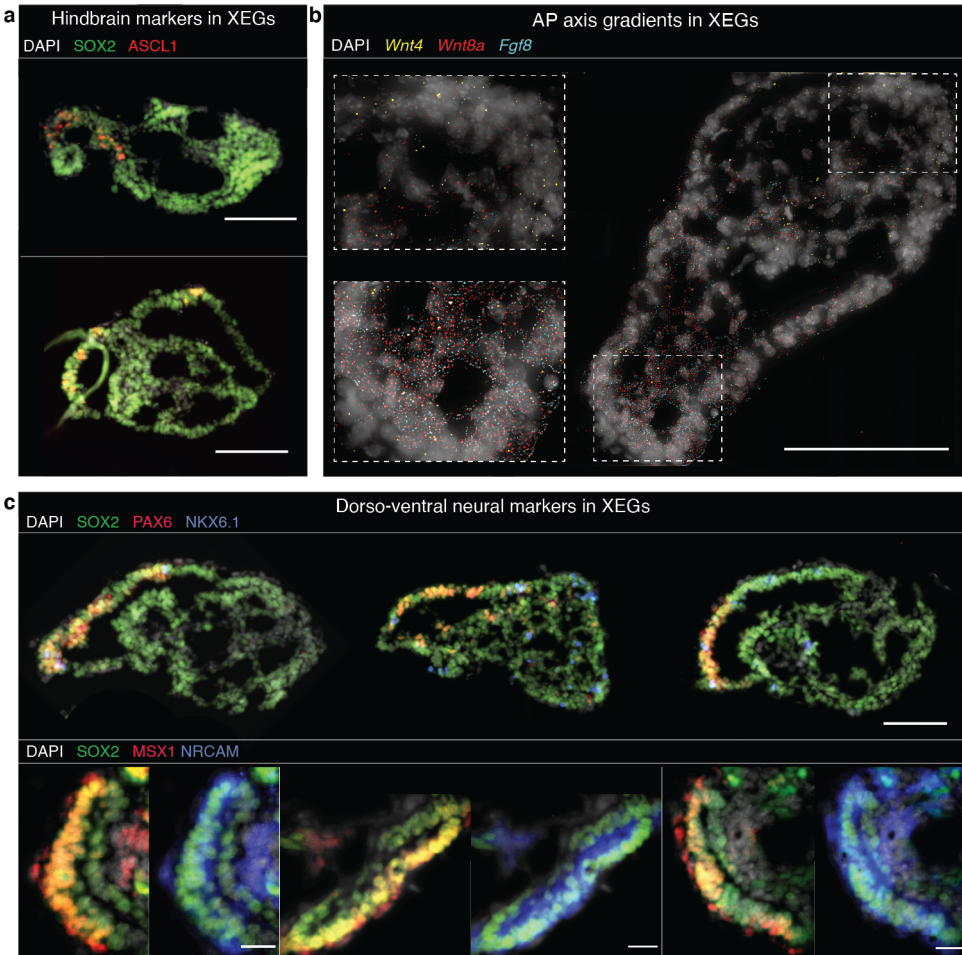


Figure 2.6: Tubes in XEGs show dorsal and ventral markers and partial AP patterning. **a**, Expression of SOX2 and ASCL1 in 96 h XEGs (immunostaining of sections). Scale bar: 100 μm . **b**, *Wnt4*, *Wnt8a* and *Fgf8* expression in XEGs at 96 h, visualized by smFISH on sections. Each diffraction limited dot is a single mRNA molecule. Scale bar: 100 μm . **c**, Expression of dorsal (PAX6, MSX1) and ventral (NKX6.1, NRCAM) neural markers in 96 h XEG (immunostaining of sections). Top, expression of PAX6 and NKX6.1 Scale bar: 100 μm . Bottom, zoomed pictures of tubes showing the expression of MSX1 and NRCAM. Scale bars: 20 μm . Cell nuclei were stained with DAPI.

2.3.3. SIGNALING PERTURBATION EXPERIMENTS AND DIFFERENTIATION TO CORTEX-LIKE TISSUE SUPPORT NEURAL TUBE-LIKE CHARACTER

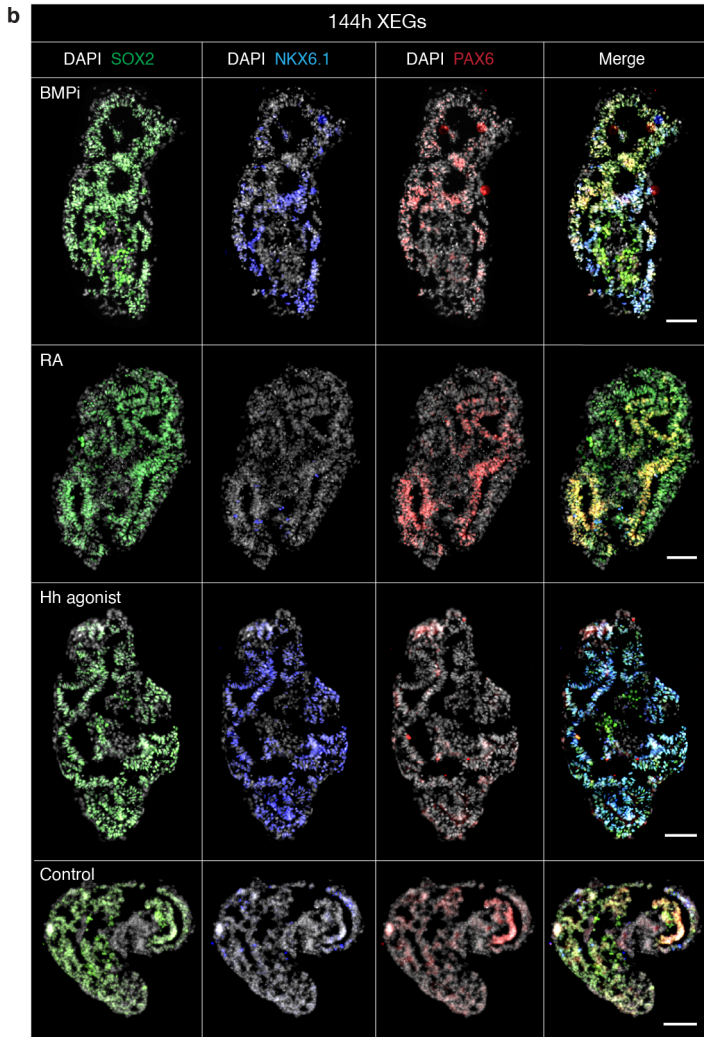
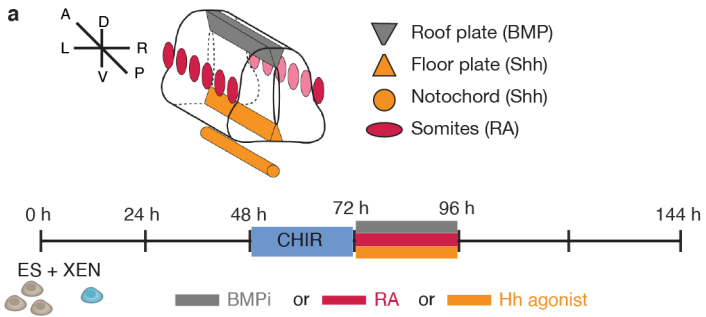
To further characterize the neural tube-like structures, we tested how they respond to signaling inputs found *in vivo*. Specifically, we explored the response to BMP, Sonic Hedgehog (Shh) and retinoic acid (RA) pathway activation (Fig. 2.7a, b). The inhibition of BMP signaling, known to prevent premature neural specification [53], resulted in an

increased number of neural progenitor-like cells (marked by SOX2, PAX6 and NKX6.1). Sonic hedgehog, produced *in vivo* by the notochord and the floor plate (see schematic in Fig. 2.7a), is known to be necessary for the patterning of the ventral part of the neural tube [54]. The activation of the Hh signaling pathway led to a higher frequency of cells expressing ventral markers (NKX6.1) in XEGs. RA, involved in anterior-posterior patterning [55], strongly increased the number of cells expressing PAX6, which are found specifically in the anterior part of the neural tube [56]. The neural tube-like structures in XEGs thus responded to signaling inputs as expected from *in vivo* development.

Even in the absence of RA, some hindbrain markers (ASCL1, PAX6) were sporadically expressed, typically concentrated at one pole of the XEG (Fig. 2.3c, Fig. 2.6a). This hint of anterior specialization suggested a potential to differentiate further into cerebral tissue. Indeed, within four days of additional culture in cerebral organoid differentiation media [32], XEGs developed a layered organization of neural progenitors (SOX2+/PAX6+) and neurons (TUJ1+/CTIP2+), surrounding cavities, reminiscent of cortex organization around ventricles (Fig. 2.8a, b, c). Interestingly, we also observed a population of cells expressing the endothelial marker CD31 (Fig. 2.8d). This might indicate that non-neural cells from XEGs remained and might have differentiated further. Those CD31+ cells could specifically represent an early stage of vasculature.

Figure 2.7 (following page): Signaling perturbation experiments confirm neural tube-like character. **a**, Top: schematic of signaling sources patterning the developing neural tube *in vivo*. A: anterior, P: posterior, D: dorsal, V: ventral, L: left, R: right. Bottom: time line of the signaling experiments. XEGs were treated from 72 h to 96 h, with either BMP pathway inhibitor (BMPi), retinoic acid (RA) or hedgehog pathway agonist (Hh agonist). The XEGs were then allowed to grow for an additional 48 h before staining. **b**, Expression of SOX2, NKX6.1 and PAX6 in XEGs at 144 h, treated with the indicated factors (immunostaining of sections). n = 3 experiments. Scale bars 100 μ m. Cell nuclei are stained with DAPI.

[H]



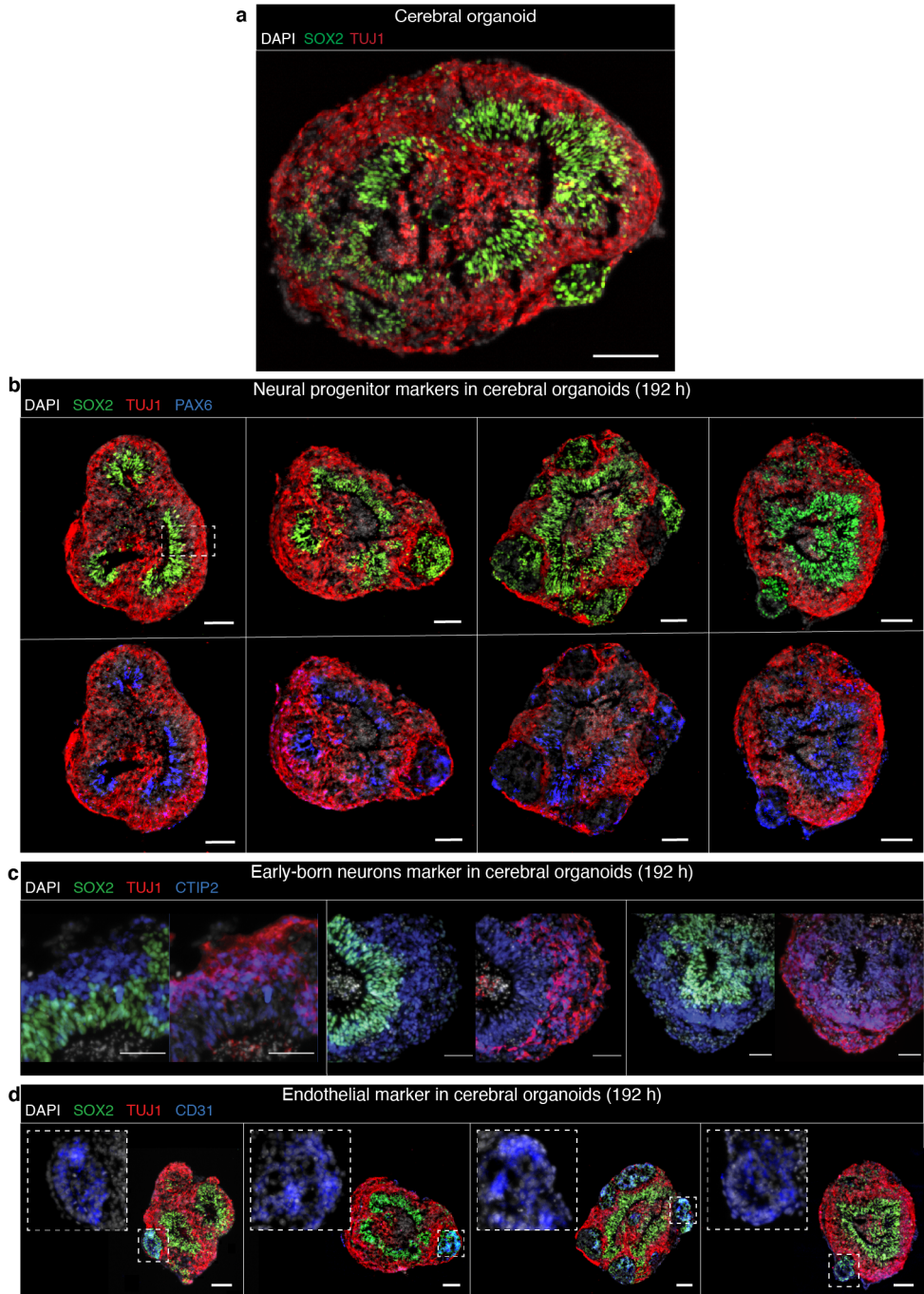


Figure 2.8: Neural tubes-like in XEGs can be further differentiated to cerebral fate. **a**, Expression of SOX2 and TUJ1 in XEG-derived cerebral cortex-like tissue, 8 days after cell seeding (differentiated from XEGs for 4 days) (immunostaining of sections). **b**, **c**, **d**, Immunostaining in sections of cerebral cortex-like tissue differentiated from XEGs at 8 days after cell seeding. **b**, TUJ1, SOX2 (top) and PAX6 (bottom). The dashed box highlights the layered, cortex-like organization adjacent to a ventricle-like cavity. **c**, TUJ1, SOX2 and CTIP2, a marker for early-born neurons. **d**, TUJ1, SOX2 and CD31. Insets show clusters of cells positive for the endothelial marker CD31. Scale bars: 100 μm . Cell nuclei are stained with DAPI.

2.3.4. SINGLE-CELL RNA-SEQ REVEALS THE TRANSCRIPTIONAL PROFILES OF XEG CELLS

Having characterized the most striking difference between gastruloids and XEGs, we wanted to take a more comprehensive approach to reveal additional differences between the two model systems. To that end, we used single-cell RNA-sequencing (scRNA-seq) (Supplementary Fig. 2.13). By mapping the data to single-cell transcriptomes of mouse embryos from E6.5 to E8.5 [43] (Fig. 2.9a, b) we classified the transcriptional identity of the cells (Fig. 2.10a, b). Except for the least abundant cell types, the distribution of cell types was consistent across two biological replicates (Fig. 2.10c). Expression of known markers confirmed the classification by mapping to *in vivo* data (Supplementary Fig. 2.14, Supplementary Fig. 2.15). Most cell types belonged to the E8.0 or E8.5 embryo (Fig. 2.10d), which might indicate that *in vitro* differentiation proceeded roughly with the same speed as *in vivo* development.

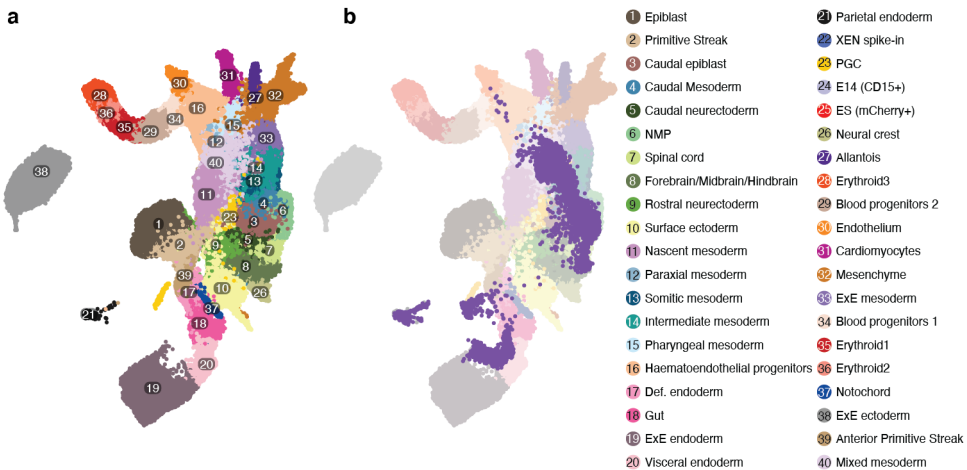


Figure 2.9: scRNA-seq data from XEGs were classified by mapping on an *in vivo* dataset. **a**, UMAP of the Pijuan-Sala et al. [43] dataset with cell types indicated by color. **b**, MNN mapping of XEG cells from replicate 2 (bright colors) to the Pijuan-Sala et al. [43] dataset (dim colors), as an example for the mapping procedure.

The neuro-mesodermal progenitors (NMP) and spinal cord-like cells were the most abundant in both model systems (Fig. 2.10c). Gastruloids thus already contain cells of the neural lineage, which, however, seem to lack organization (Fig. 2.2b, inset). To identify the cells forming tubular structures in XEGs, we used the neural tube markers *Sox2*, *Pax6* and *Nkx6.1* [48], which we had detected by immunostaining (Fig. 2.3c). We found these markers to be co-expressed in cells classified as “spinal cord” and “brain” in the scRNA-seq data (Fig. 2.10e, f), confirming their neural ectoderm identity. While NMPs

also expressed *Sox2* and *Nkx6.1*, tubular structures were clearly distinguishable by the presence of *Pax6* and the absence of *T*.

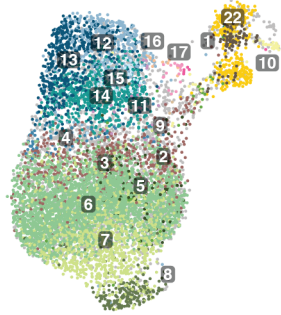
The scRNA-seq results also supported the existence of some anteroposterior axis, as well as dorsoventral patterning. Several anterior and posterior markers were expressed in separate subpopulations (Fig. 2.10e). Differential gene expression analysis between “spinal cord” cells and other cells in XEGs identified markers of both the dorsal and ventral neural tube (Fig. 2.11a, c, Supplementary Table 1). Two sets of genes that were either co-expressed with *Msx1* or *Nrcam*, respectively, showed strong anticorrelation (Fig. 2.11b, left), which confirmed our immunostaining results (Fig. 2.6c). Similarly, the averages of canonical dorsal and ventral markers were anti-correlated (Fig. 2.11b, right), despite individual markers being lowly expressed. A comparison between XEGs and gastruloids revealed that neural ectoderm-like cells expressed more dorsal markers in XEGs (Fig. 2.11d). This dorsal identity was confirmed by mapping the neural ectoderm-like cells to single-cell expression profiles of *in vivo* neural tube [57] (Fig. 2.11e). The majority of neural ectoderm-like cells from XEGs turned out to be more similar to dorsal progenitors *in vivo*. All in all, the scRNA-seq results support dorsoventral patterning in XEGs with a bias towards the dorsal identity, compared to gastruloids.

Overall, both model systems showed a diverse cell type distribution, comprising a variety of mesodermal as well as anterior cell types. Thus, XEG cells are not globally biased towards the neural fate, as occurring in other protocols for induction of neural epithelia [58, 50, 19]. On the contrary, XEGs even contained a bigger proportion of mesoderm-like cells. While paraxial, intermediate and somitic mesoderm were present in both model systems, only XEGs contained primitive streak, nascent mesoderm, pharyngeal mesoderm and haematoendothelial progenitors (Fig. 2.3c, Fig. 2.12a). To confirm the presence of mesodermal cell types in XEGs, we focused on two genes, *Tbx6* and *Pax2*, which are markers of nascent and intermediate mesoderm, respectively. Our single-cell RNA-seq data showed expression of both genes in subpopulations of XEG cells (Fig. 2.12b) and immunostaining confirmed their presence in XEGs (Fig. 2.12c). However, we did not observe any tissue-level organization of those cells. Taken together, these results suggest that the XEN-derived cells in XEGs might also have an effect on the mesodermal cell population, which, in turn, could potentially also contribute to neural tube formation.

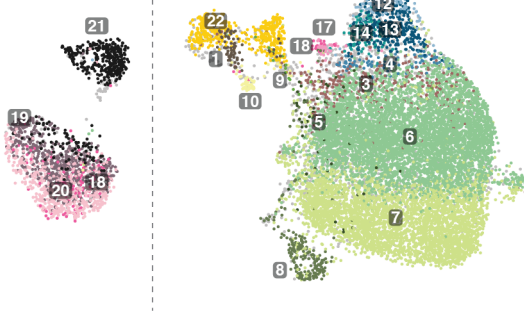
Figure 2.10 (following page): Single-cell RNA-seq reveals the transcriptional profiles of XEG cells. a, b, UMAP of cells in XEGs and gastruloids colored by cell type based on the mapping to *in vivo* data shown in Fig. 2.9. c, Cell type frequencies for both replicates in XEGs and gastruloids. d, Developmental age of cell types based on mapping to *in vivo* data. e, *Sox2*, *Pax6*, *Nkx6.1* and *T* log-expression levels indicated by color in UMAPs of XEGs. f, Expression of *Sox2*, *Pax6* and *Nkx6.1* and *T* in XEGs, as measured by single-cell RNA-seq.

[H]

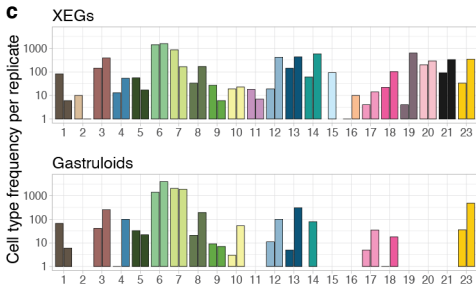
a XEGs



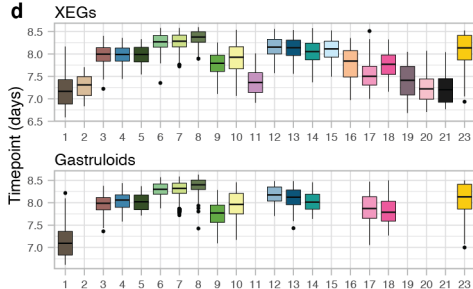
b Gastruloids



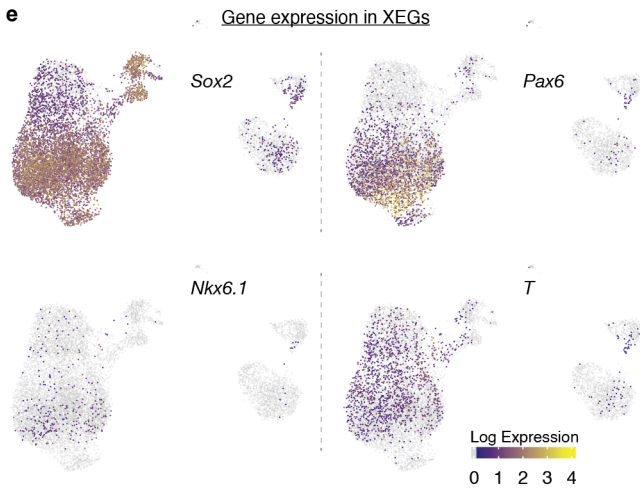
c



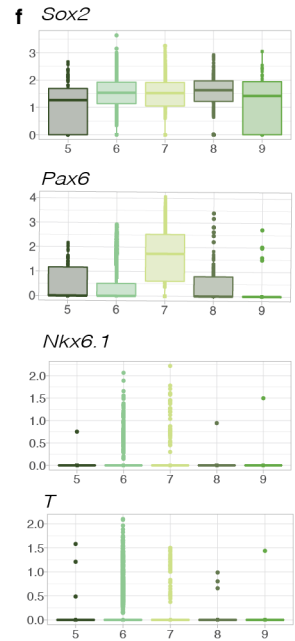
d



e



f



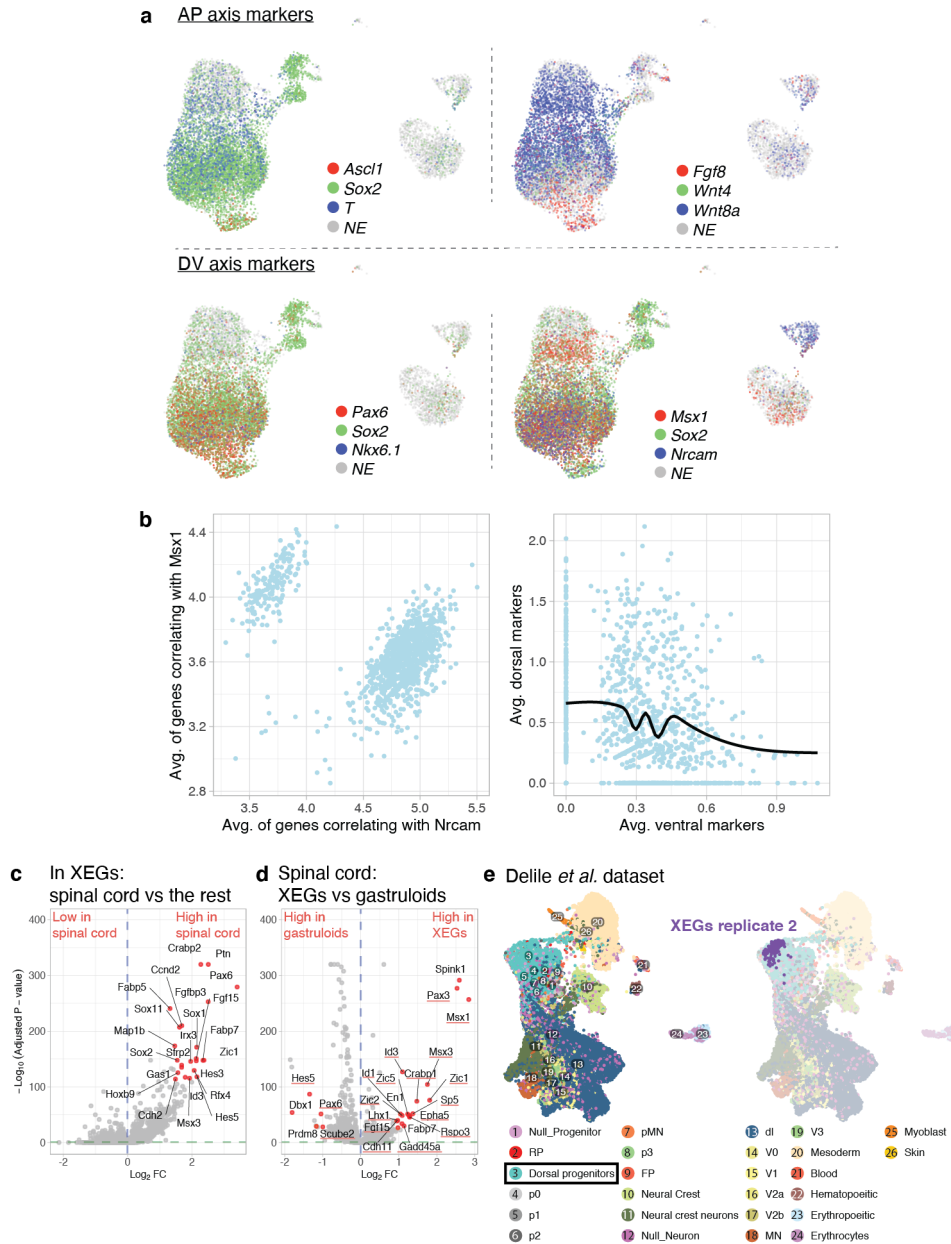


Figure 2.11: Single-cell RNA-seq confirm the axis patterning of the XEGs. **a**, Top, UMAP of cells in XEGs with log expression of genes expressed along the AP axis indicated by colors. Left: *Ascl1*, *Sox2* and *T*, right: *Wnt4*, *Wnt8a* and *Fgf8*. Bottom, UMAP of cells in XEGs with log expression of genes expressed along the DV axis indicated by colors. Left: *Pax6*, *Sox2* and *Nkx6.1*, right: *Msx1*, *Sox2* and *Nrcam*. All UMAP show both replicates, batch corrected. **b**, Left, scatter plot of average expression of the top10 genes correlating with *Msx1* and *Nrcam* in our scRNA-seq data. Right, scatter plot of average expression of 10 canonical marker genes for dorsal and ventral neural progenitors, with the linear smoothing indicated by the black curve. **c**, Gene expression differences between cells classified as “spinal cord” and all other cells in XEGs (fold-change vs p-value). Named genes are expressed in the neural tubes (Supplementary Table 1). **d**, Gene expression differences between cells classified as “spinal cord” in gastruloids and XEGs (fold-change vs p-value). Underlined genes are expressed in the dorsal part of the neural tube (Supplementary Table 2). **e**, Left, UMAP of the cells in the Delile *et al.* dataset [57] colored by cell type. Right, MNN mapping of cells classified as “spinal cord” in replicate 2 XEGs (bright colors).

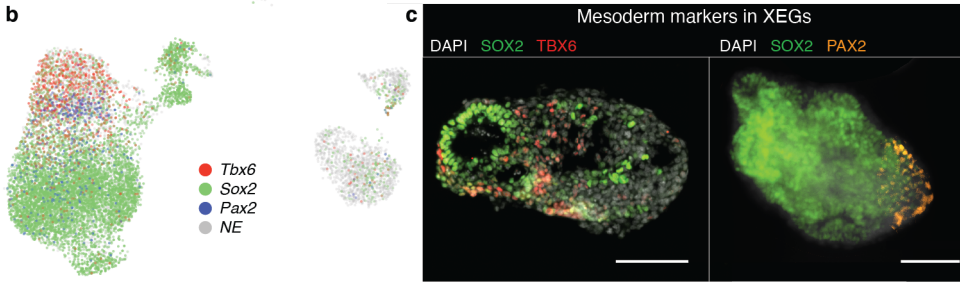
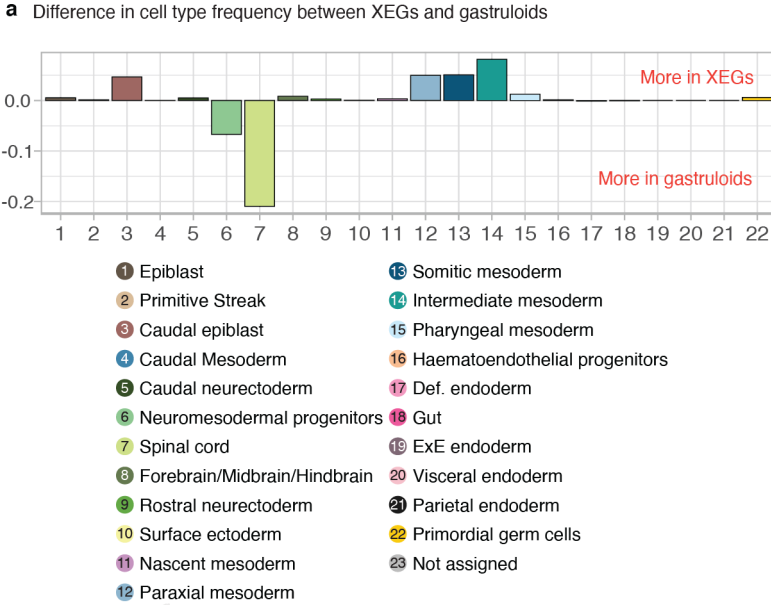


Figure 2.12: XEGs and gastruloids differ in cell type distribution. **a**, Differences between relative frequencies of cell types in XEGs and gastruloids. **b**, UMAP of cells in XEGs (both replicates, batch corrected) with log expression of *Tbx6*, *Sox2* and *Pax2* indicated by color. **c**, Expression of mesoderm markers. Left, immunostaining of TBX6 in a 96 h XEG section. Right, wholemount immunostaining of PAX2 in a 96 h XEG. Scale bars: 100 μ m. Cell nuclei were stained with DAPI.

2.4. DISCUSSION

In this study we explored the interaction between embryonic and extraembryonic cells in a new *in vitro* model of embryogenesis that achieves tissue-level organization.

While we did not observe cell rearrangements characteristic of primary neurulation, the rosette formation seen in XEGs was reminiscent of secondary neurulation [1], which gives rise to the posterior neural tube. Nevertheless, we were able to differentiate XEGs towards cortex-like tissue, which derives from the anterior neural tube. The neural tube-like structures in 96 h XEGs are therefore likely still plastic and would need additional time to become fully specified. Our model might thus be most useful for studying the minimal cues necessary for initiating neural tube formation *in vitro*. The recently reported Trunk-Like Structures (TLS) [59] also produce neural tube-like tissues in aggregates derived from gastruloids, together with mesodermal tissues resembling somites. Notably, TLS are formed exclusively from mESCs and are grown in Matrigel with additional exogenous signaling cues (BMP signaling inhibitor LDN193189 and CHIR). Together with our results, TLS showcase the potential of stem cell-derived aggregates to explore the different factors involved in organogenesis.

Finally, with their large diversity of cell types, XEGs could be a starting point for developing more complex models containing all three germ layers as well as extraembryonic cells. Specifically, the CD31 positive endothelial cells observed in the cerebral organoids obtained from XEGs might be able to form a vascular network if additional signaling cues are given [60].

In conclusion, in this study we established a new *in vitro* model that recapitulates elements of mammalian development, leveraging the morphogenic potential of heterotypic cell-cell interactions.

2.5. SUPPLEMENTARY INFORMATION

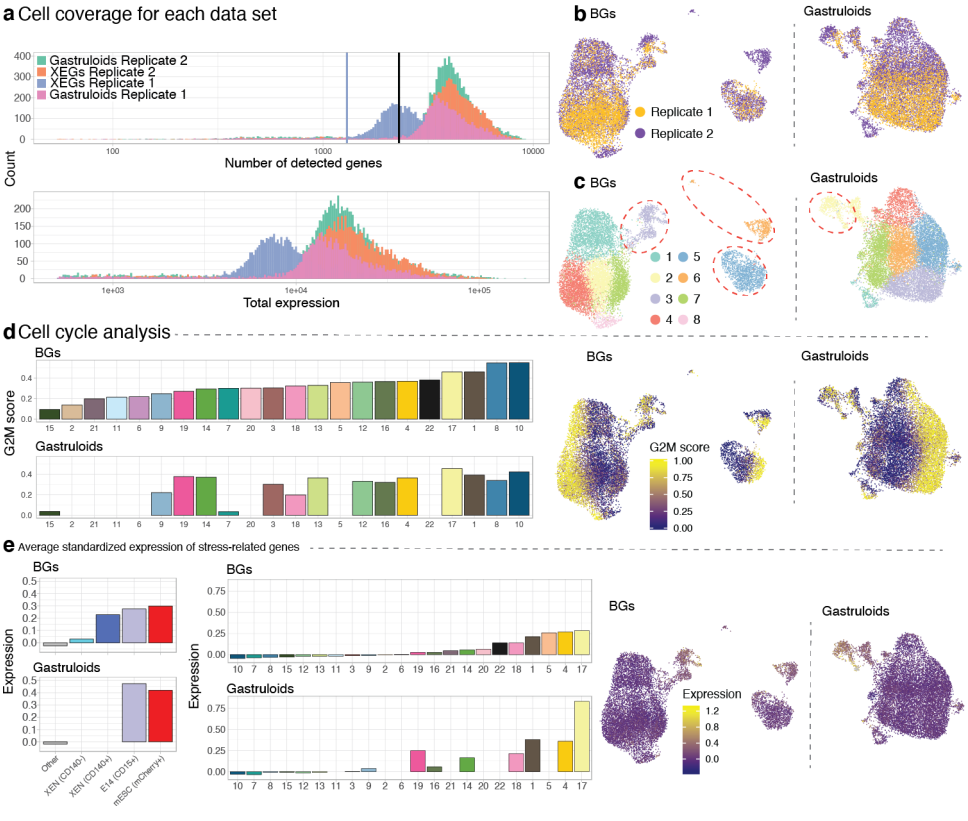


Figure 2.13: Quality control of single-cell RNA-seq data. **a**, Top, number of detected genes per cell in each replicate; the blue line indicates a quality control threshold for XEGs from replicate 1 and the black line for the remaining datasets. Bottom, total expression per cell for each dataset. **b**, UMAP of cells in XEGs and gastruloids, colored by replicate. **c**, UMAP of cells in XEGs and gastruloids, colored by Louvain clustering. The encircled clusters contain the spiked-in cells. **d**, Left, average G2M scores for each cell type. Right, UMAPs of cells in XEGs and gastruloids colored by G2M score. **e**, Left, average standardized expression of stress-related genes in spike-in cells. Middle, expression of stress-related genes by cell type. Right, UMAPs of cells in XEGs and gastruloids with expression of stress-related genes indicated by color. **b-e**, UMAPs contain both replicates, batch corrected.

Marker genes in XEGs

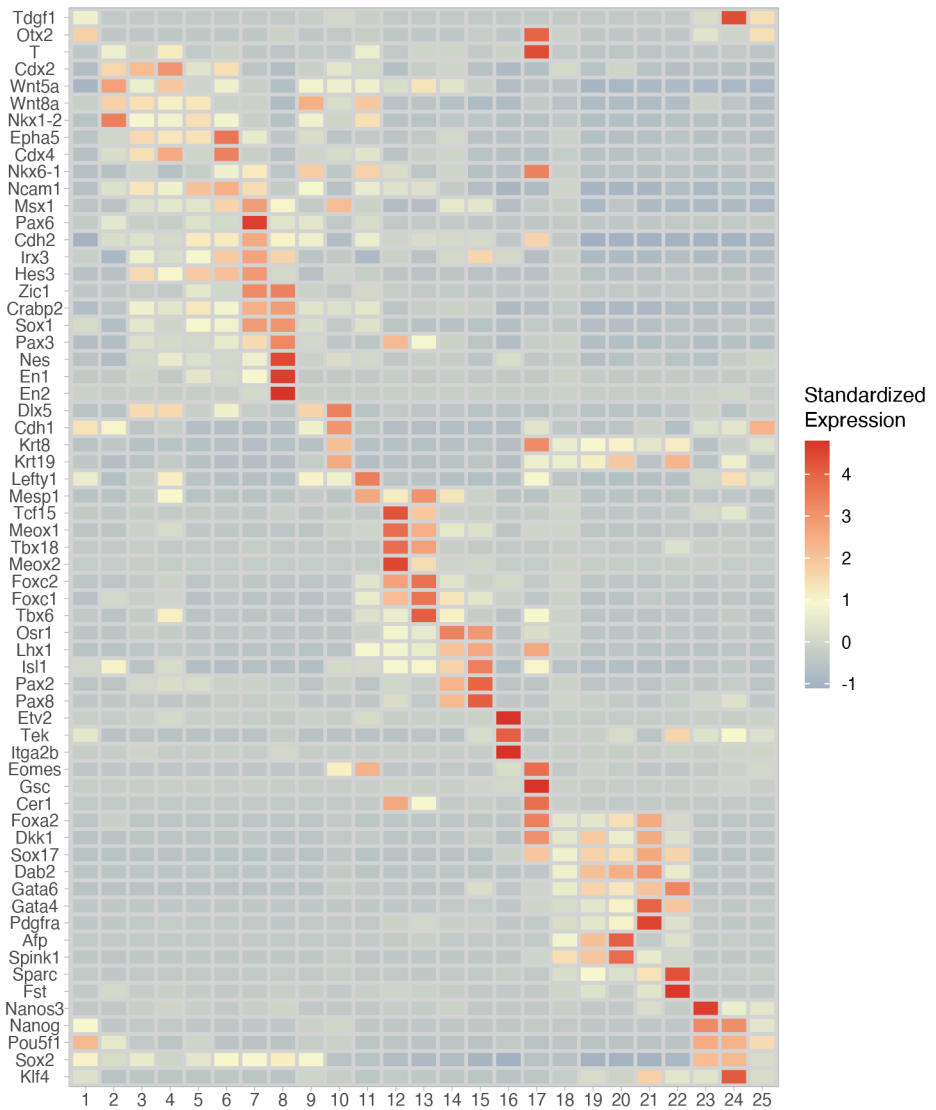


Figure 2.14: Heat map of standardized expression of genes associated with mouse embryonic development in XEGs.

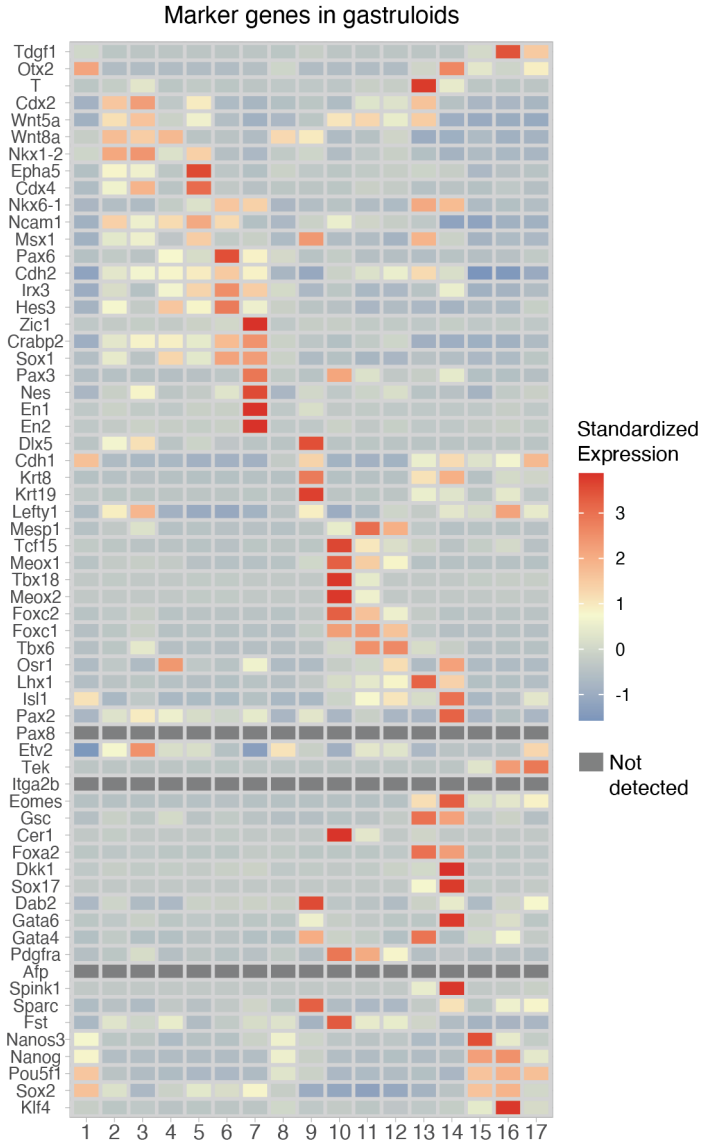


Figure 2.15: Heat map of standardized expression of genes associated with mouse embryonic development in gastruloids.

Gene	Log FC	p-adjust	Reference
Ccnd2	1.684190185	1.37E-210	Lacomme, M., et al., (2012). <i>Molecular and Cellular Biology</i>
Crabp2	2.27328271	0	Ruberte, E., et al., (1992). <i>Development</i>
Id3	1.789970808	5.65E-118	Wine-Lee, L., et al., (2004). <i>Development</i>
Fabp7 (BLBP)	2.339513279	2.65E-148	Feng, L., Hatten, M. E., & Heintz, N. (1994). <i>Neuron</i>
Rfx4	2.066843727	3.40E-130	Ashique, A. M., et al. (2009). <i>Science Signaling</i>
Hoxb9	1.561341789	4.07E-126	Graham, A., Maden, M., & Krumlauf, R. (1991). <i>Development</i>
Cdh2	1.486458795	1.15E-114	Radice, G. L., et al., (1997). <i>Developmental Biology</i>
Msx3	1.925035016	3.06E-116	Misra, K., Luo, H., Li, S., Matisse, M., & Xiang, M. (2014). <i>Development</i>
Pax6	3.403368474	4.04E-280	Jeong, J., & McMahon, A. P. (2005). <i>Development</i>
Fabp5	1.316127675	1.97E-241	Shimamoto, C., et al. (2014). <i>Human Molecular Genetics</i>
Sfrp2	1.677002264	9.74E-140	Misra, K., & Matisse, M. P. (2010). <i>Developmental Biology</i>
Hes3	2.132129411	2.12E-147	Lobe, C. G. (1997). <i>Mechanisms of Development</i>
Ptn	2.50622717	0	Magdaleno, S., et al. (2006). <i>PLoS Biology</i>
Fgf15	2.131102328	2.17E-151	McWhirter, J. R., et al., (1997). <i>Development</i>
Irx3	1.957663655	1.71E-146	Lebel, M., et al. (2003). <i>Molecular and Cellular Biology</i>
Zic1	2.371434543	1.56E-148	Nagai, T., et al., (1997). <i>Developmental Biology</i>
Fgfbp3	2.502793725	1.10E-253	Yu, K., McGlynn, S., & Matisse, M. P. (2013). <i>Development</i>
Hes5	2.152139781	9.35E-119	Hatakeyama, J., et al., (2004). <i>Development</i>
Map1b	1.46443751	3.03E-174	Fawcett, J. W., et al., (1994). <i>Neuroscience</i>
Gas1	1.676359071	1.03E-135	Allen, B. L., Tenzen, T., & McMahon, A. P. (2007). <i>Genes & Development</i>
Sox11	1.610646468	6.85E-208	Tanaka, S., et al., (2004). <i>Molecular and Cellular Biology</i>
Sox2	1.54213575	1.78E-148	Hoffmann, S. A., et al., (2014). <i>Development</i>
Sox1	2.143957789	5.90E-172	Hoffmann, S. A., et al., (2014). <i>Development</i>

Supplementary Table 1: Genes differentially expressed between neural ectoderm-like cells and the other cell types in XEGs.

Gene	Log FC	Expression summary	BMP signalling response	Reference
Msx1	2.84	Dorsal neural tube + small ventral domain	BMP (Lee, K. J., Jessell, T. M. (1999))	Misra, K., et al., (2014). Duval, N., et al. (2014).
Spink1	2.59			
Pax3	2.53	Dorsal neural tube	BMP (Lee, K. J., Jessell, T. M. (1999))	Kriks, S., et al., (2005). Nagai, T., et al., (1997).
Zic1	1.81	Dorsal neural tube	BMP (Lee, K. J., Jessell, T. M. (1999))	Aruga, J., et al., (2002).
Msx3	1.75	Dorsal neural tube	BMP4 (Shimeld, S. et al., (1996))	Misra, K., et al., (2014). Duval, N., et al. (2014).
Crabp1	1.47	Forebrain, Neural crest + dorsal neural tube	-	Ruberte, E., et al., (1991).
Sp5	1.37	Midbrain and dorsal neural tube	-	Andoniadou, C. L., et al. (2011). Dunty, W. C., et al., (2014).
Epha5	1.29	Dorsal and ventral	BMP2 (Yamada, T., et al., (2016))	Abdul-Aziz, N. M., et al., (2009). Ono, K., et al. (2014).
Rspo3	1.28	Forebrain, dorsal neural tube	-	Nam, J.-S., et al., (2007).
En1	1.27	Hinbrain/Midbrain Ventral neural tube	Not affected by BMP (Timmer, J. R., et al., (2002))	Ericson, J., et al. (1997). Sgaier, S. K., et al., (2007).
Zic5	1.23	Dorsal neural tube and head fold	BMP (Nakata, K., et a., (2000))	Inoue, T., et al., (2004).
Fabp7 (BLBP)	1.14	Ventral neural tube Midbrain	-	Delile, J., et al., (2019). Feng, L., et al., (1994).
Id3	1.1	The roof plate and dorsal neural ectoderm + small ventral domain	BMP (Wine-Lee, L., et al., (2004))	Wine-Lee, L., et al., (2004).
Zic2	1.09	Dorsal neural tube	BMP6/7 (Lee, K. J., Jessell, T. M. (1999))	Ybot-Gonzalez, P., et al., (2007). Sanchez-Ferras, O., et al., (2014).
Gadd45a	1.08	Mostly dorsal neural tube Residual expression in ventral neural tube	BMP2 (Ijiri, K., et al. (2005))	Kaufmann, L. T., et al., (2011).
Id1	1.04	Roof plate and dorsal neural ectoderm + small ventral domain	BMP (Wine-Lee, L., et al., (2004))	Wine-Lee, L., et al., (2004).
Cdh11	0.98	Neural crest + dorsal neural tube	BMP7 (Awazu, M., et al., (2017))	Tondeleir, D., et al., (2014). Kashef, J., et al., (2009).
Lhx1	0.97	Dorsal interneurons	Not affected by BMP4/7 (Le Dréau, G., et al., (2012))	Le Dréau, G., et al., (2012).
Egfl5	0.95	Dorsal regions of the di-, mes and metencephalon	-	McWhirter, J. R., et al., (1997). Fischer, T., et al., (2011).

Supplementary Table 2: Genes differentially expressed between neural ectoderm-like cells in XEGs and gastruloids.

ACRONYMS

AP	antero-posterior	OWB	organoid washing buffer
BSA	bovine serum albumin	PFA	paraformaldehyde
CITE-seq	cellular indexing of transcriptomes and epitopes by sequencing	RT	room temperature
CLAHE	enhance local contrast	scrNA	single cell RNA
DV	dorso-ventral	scrNA-seq	single-cell RNA sequencing
E3.5	embryonic day 3.5	SDS	sodium dodecyl sulfate
ESC	embryonic stem cell	smFISH	single-molecule FISH
HVG	highly variable genes	tRNA	transfer RNA
knn	k-nearest neighbors	UMAP	uniform manifold approximation and projection
mESC	mouse embryonic stem cell	XEG	XEN enhanced-gastruloids
MNN	mutual nearest neighbors	XEN	extraembryonic endoderm-like
OCT	optimal cutting temperature compound		

CELL TYPES

AVE	anterior visceral endoderm	NMP	neuro-mesodermal progenitors
DVE	distal visceral endoderm	PE	parietal endoderm
Epi	epiblast	PrE	primitive endoderm
ICM	inner cell mass	VE	visceral endoderm

2.6. REFERENCES

- [1] L. A. Lowery and H. Sive. “Strategies of vertebrate neurulation and a re-evaluation of teleost neural tube formation.” In: *Mechanisms of Development* 121.10 (Oct. 2004), pp. 1189–1197.
- [2] G. C. Schoenwolf. “Histological and ultrastructural studies of secondary neurulation in mouse embryos.” In: *The American journal of anatomy* 169.4 (Apr. 1984), pp. 361–376.
- [3] J. F. Colas and G. C. Schoenwolf. “Towards a cellular and molecular understanding of neurulation.” In: *Developmental dynamics : an official publication of the American Association of Anatomists* 221.2 (June 2001), pp. 117–145.
- [4] A. Dady et al. “Junctional neurulation: a unique developmental program shaping a discrete region of the spinal cord highly susceptible to neural tube defects.” In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 34.39 (Sept. 2014), pp. 13208–13221.
- [5] K. A. Lawson, J. J. Meneses, and R. A. Pedersen. “Clonal analysis of epiblast fate during germ layer formation in the mouse embryo.” In: *Development* 113.3 (Nov. 1991), pp. 891–911.
- [6] I. Bedzhov et al. “In vitro culture of mouse blastocysts beyond the implantation stages.” In: *Nature protocols* 9.12 (Dec. 2014), pp. 2732–2739.
- [7] P. Thomas and R. Beddington. “Anterior primitive endoderm may be responsible for patterning the anterior neural plate in the mouse embryo.” In: *Current biology : CB* 6.11 (Nov. 1996), pp. 1487–1496.
- [8] K. McDole et al. “In Toto Imaging and Reconstruction of Post-Implantation Mouse Development at the Single-Cell Level.” In: *Cell* 175.3 (Oct. 2018), 859–876.e33.
- [9] S. Vianello and M. P. Lutolf. “Understanding the Mechanobiology of Early Mammalian Development through Bioengineered Models.” In: *Developmental Cell* 48.6 (Mar. 2019), pp. 751–763.
- [10] M. N. Shahbazi, E. D. Siggia, and M. Zernicka-Goetz. “Self-organization of stem cells into embryos: A window on early mammalian development.” In: *Science* ().
- [11] M. Simunovic and A. H. Brivanlou. “Embryoids, organoids and gastruloids: new approaches to understanding embryogenesis.” In: *Development* 144.6 (Mar. 2017), pp. 976–985.
- [12] S. C. van den Brink et al. “Symmetry breaking, germ layer specification and axial organisation in aggregates of mouse embryonic stem cells.” In: *Development* 141.22 (Nov. 2014), pp. 4231–4242.
- [13] D. A. Turner et al. “Anteroposterior polarity and elongation in the absence of extra-embryonic tissues and of spatially localised signalling in gastruloids: mammalian embryonic organoids.” In: *Development* 144.21 (Nov. 2017), pp. 3894–3906.

- [14] L. Beccari et al. “Multi-axial self-organization properties of mouse embryonic stem cells into gastruloids.” In: *Nature* (Oct. 2018), p. 1.
- [15] S. C. van den Brink et al. “Single-cell and spatial transcriptomics reveal somitogenesis in gastruloids.” In: *Nature* 141 (Feb. 2020), pp. 4231–5.
- [16] A. Warmflash et al. “A method to recapitulate early embryonic spatial patterning in human embryonic stem cells.” In: *Nature Methods* 11.8 (Aug. 2014), pp. 847–854.
- [17] M. Simunovic et al. “A 3D model of a human epiblast reveals BMP4-driven symmetry breaking.” In: *Nature Cell Biology* 21.7 (July 2019), pp. 900–910.
- [18] B. Sozen et al. “Self-assembly of embryonic and two extra-embryonic stem cell types into gastrulating embryo-like structures.” In: *Nature Cell Biology* 20.8 (Aug. 2018), pp. 979–989.
- [19] T. Haremake et al. “Self-organizing neuruloids model developmental aspects of Huntington’s disease in the ectodermal compartment.” In: *Nature Biotechnology* 37.10 (Oct. 2019), pp. 1198–1208.
- [20] S. Pfister, K. A. Steiner, and P. P. L. Tam. “Gene expression pattern and progression of embryogenesis in the immediate post-implantation period of mouse development”. In: *Gene Expression Patterns* 7.5 (Apr. 2007), pp. 558–573.
- [21] S. E. Harrison et al. “Assembly of embryonic and extra-embryonic stem cells to mimic embryogenesis in vitro.” In: *Science* (Mar. 2017), eaal1810.
- [22] N. C. Rivron et al. “Blastocyst-like structures generated solely from stem cells”. In: *Nature* 557.7703 (May 2018), pp. 106–111.
- [23] B. Mathew et al. “Mouse ICM Organoids Reveal Three-Dimensional Cell Fate Clustering.” In: *Biophysical journal* 116.1 (Jan. 2019), pp. 127–141.
- [24] B. L. Hogan, A. R. Cooper, and M. Kurkinen. “Incorporation into Reichert’s membrane of laminin-like extracellular proteins synthesized by parietal endoderm cells of the mouse embryo.” In: *Developmental biology* 80.2 (Dec. 1980), pp. 289–300.
- [25] G. S. Kwon, M. Viotti, and A.-K. Hadjantonakis. “The endoderm of the mouse embryo arises by dynamic widespread intercalation of embryonic and extraembryonic lineages.” In: *Developmental Cell* 15.4 (Oct. 2008), pp. 509–520.
- [26] S. Nowotschin, A.-K. Hadjantonakis, and K. Campbell. “The endoderm: a divergent cell lineage with many commonalities.” In: *Development* 146.11 (June 2019).
- [27] M. Madabhushi and E. Lacy. “Anterior visceral endoderm directs ventral morphogenesis and placement of head and heart via BMP2 expression.” In: *Developmental Cell* 21.5 (Nov. 2011), pp. 907–919.
- [28] I. Bedzhov et al. “Development of the anterior-posterior axis is a self-organizing process in the absence of maternal cues in the mouse embryo”. In: *Cell Research* 25.12 (Dec. 2015), pp. 1368–1371.

- [29] Q.-L. Ying et al. *Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture*. Tech. rep. Institute for Stem Cell Research, University of Edinburgh, King's Buildings, West Mains Road, Edinburgh EH9 3JQ, United Kingdom., Feb. 2003.
- [30] T. Kunath et al. "Imprinted X-inactivation in extra-embryonic endoderm cell lines from mouse blastocysts." In: *Development* 132.7 (Apr. 2005), pp. 1649–1661.
- [31] J. Artus, J.-J. Panthier, and A.-K. Hadjantonakis. "A role for PDGF signaling in expansion of the extra-embryonic endoderm lineage of the mouse blastocyst." In: *Development* 137.20 (Oct. 2010), pp. 3361–3372.
- [32] M. A. Lancaster et al. "Cerebral organoids model human brain development and microcephaly." In: *Nature* 501.7467 (Sept. 2013), pp. 373–379.
- [33] J. F. Dekkers et al. "High-resolution 3D imaging of fixed and cleared organoids." In: *Nature protocols* 14.6 (June 2019), pp. 1756–1771.
- [34] S. Semrau et al. "FuseFISH: robust detection of transcribed gene fusions in single cells." In: *Cell Reports* 6.1 (Jan. 2014), pp. 18–23.
- [35] C. A. Schneider, W. S. Rasband, and K. W. Eliceiri. "NIH Image to ImageJ: 25 years of image analysis." In: *Nature Methods* 9.7 (July 2012), pp. 671–675.
- [36] S. van der Walt et al. "scikit-image: image processing in Python." In: *PeerJ* 2.2 (2014).
- [37] M. Stoeckius et al. "Simultaneous epitope and transcriptome measurement in single cells." In: *Nature Methods* 14.9 (Sept. 2017), pp. 865–868.
- [38] A. T. L. Lun, K. Bach, and J. C. Marioni. "Pooling across cells to normalize single-cell RNA sequencing data with many zero counts." In: *Genome biology* 17.1 (Apr. 2016), pp. 75–14.
- [39] L. Haghverdi et al. "Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors." In: *Nature Biotechnology* 36.5 (June 2018), pp. 421–427.
- [40] L. McInnes, J. Healy, and J. Melville. "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction". In: (Feb. 2018). arXiv: 1802.03426.
- [41] A. Scialdone et al. "Computational assignment of cell-cycle stage from single-cell transcriptome data." In: *Methods* 85 (Sept. 2015), pp. 54–61.
- [42] S. C. van den Brink et al. "Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations." In: *Nature Methods* 14.10 (Sept. 2017), pp. 935–936.
- [43] B. Pijuan-Sala et al. "A single-cell molecular map of mouse gastrulation and early organogenesis." In: *Nature* 566.7745 (Feb. 2019), pp. 490–495.
- [44] M. D. Robinson, D. J. McCarthy, and G. K. Smyth. "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data." In: *Bioinformatics (Oxford, England)* 26.1 (Jan. 2010), pp. 139–140.

- [45] K. Hatta and M. Takeichi. "Expression of N-cadherin adhesion molecules associated with early morphogenetic events in chick development." In: *Nature* 320.6061 (Apr. 1986), pp. 447–449.
- [46] K. Punovuori et al. "N-cadherin stabilises neural identity by dampening anti-neural signals." In: *Development* 146.21 (Nov. 2019).
- [47] S. Tsuda et al. "FAK-mediated extracellular signals are essential for interkinetic nuclear migration and planar divisions in the neuroepithelium." In: *Journal of Cell Science* 123 (Feb. 2010), pp. 484–496.
- [48] T. M. Jessell. "Neuronal specification in the spinal cord: inductive signals and transcriptional codes." In: *Nature Reviews Genetics* 1.1 (Oct. 2000), pp. 20–29.
- [49] I. Bedzhov and M. Zernicka-Goetz. "Self-Organizing Properties of Mouse Pluripotent Cells Initiate Morphogenesis upon Implantation". In: *Cell* 156.5 (Feb. 2014), pp. 1032–1044.
- [50] A. Meinhardt et al. "3D Reconstitution of the Patterned Neural Tube from Embryonic Stem Cells". In: *Stem Cell Reports* 3.6 (Dec. 2014), pp. 987–999.
- [51] L. H. Pevny et al. "A role for SOX1 in neural determination." In: *Development* 125.10 (May 1998), pp. 1967–1978.
- [52] I. Olivera-Martinez et al. "Major transcriptome re-organisation and abrupt changes in signalling, cell cycle and chromatin regulation at neural differentiation *in vivo*." In: *Development* 141.16 (Aug. 2014), pp. 3266–3276.
- [53] A. Di-Gregorio et al. "BMP signalling inhibits premature neural differentiation in the mouse embryo." In: *Development* 134.18 (Sept. 2007), pp. 3359–3369.
- [54] E. Dessaud, A. P. McMahon, and J. Briscoe. "Pattern formation in the vertebrate neural tube: a sonic hedgehog morphogen-regulated transcriptional network." In: *Development* 135.15 (Aug. 2008), pp. 2489–2503.
- [55] M. Maden. "Retinoid signalling in the development of the central nervous system." In: *Nature reviews. Neuroscience* 3.11 (Nov. 2002), pp. 843–853.
- [56] J. Ericson et al. "Pax6 controls progenitor cell identity and neuronal fate in response to graded Shh signaling." In: *Cell* 90.1 (July 1997), pp. 169–180.
- [57] J. Delile et al. "Single cell transcriptomics reveals spatial and temporal dynamics of gene expression in the developing mouse spinal cord." In: *Development* 146.12 (Mar. 2019).
- [58] A. Ranga et al. "Neural tube morphogenesis in synthetic 3D microenvironments." In: *Proceedings of the National Academy of Sciences of the United States of America* 113.44 (Nov. 2016).
- [59] J. V. Veenvliet et al. "Mouse embryonic stem cells self-organize into trunk-like structures with neural tube and somites." In: *Science* 370.6522 (Dec. 2020).
- [60] G. Rossi et al. "Capturing Cardiogenesis in Gastruloids." In: *Cell stem cell* (Nov. 2020).

3

EFFECT AND DIFFERENTIATION OF XEN CELLS IN XEN ENHANCED-GASTRULOIDS

In Chapter 2, we presented our new *in vitro* model system, called XEN enhanced-gastruloid (XEG), which is composed of mouse embryonic stem cells (mESC) and extraembryonic endoderm (XEN) cells. We showed that this model system is able to self-organize to display neural tube-like structures and exhibit signs of body axis formation. In this chapter, we further explore the role and fate of the XEN cells in XEGs.

We demonstrate that XEN cells promote the formation of neural tissues in XEGs by local inhibition of primitive streak formation, likely involving WNT signaling, and by producing a basal membrane. In turn, mESCs guide XEN differentiation to a visceral endoderm-like state. Our model system can thus be used to study fundamental mechanisms of *in vivo* development, in particular to identify the minimal set of factors required to achieve neural tube-like tissue organization.

3.1. INTRODUCTION

The extraembryonic endoderm has been the object of intense study for many decades [1] and it has become clear that this tissue has a crucial and multifaceted role in mammalian development [2].

In the mouse embryo, the primitive endoderm (PrE) differentiates from the Inner Cell Mass (ICM) around embryonic day 3.5 (E3.5) [3, 4]. Shortly after, it forms a single layer of cells surrounding the rest of the ICM, which differentiated into epiblast (Epi) at E4.5 [5, 6]. After implantation, the PrE will segregate [7] and give rise to the Parietal Endoderm (PE), covering the inside of the blastocoel cavity [8] and the Visceral Endoderm (VE), surrounding the embryo as the ICM is elongating to form the egg cylinder, until the formation of the visceral yolk sac (Fig. 3.1).

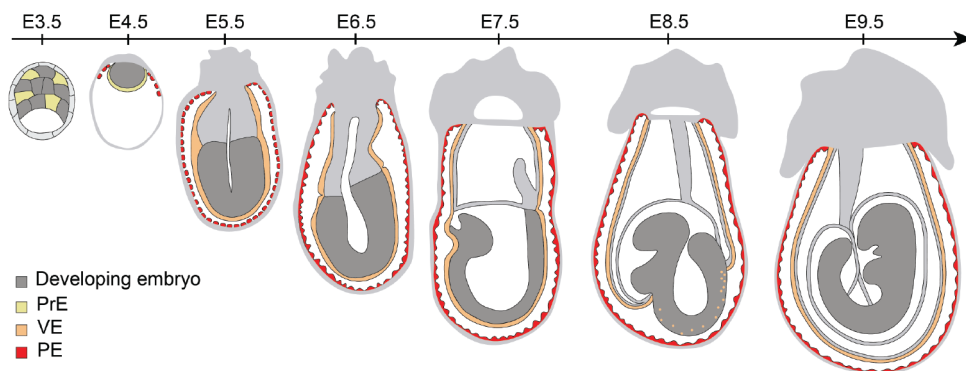


Figure 3.1: Extraembryonic endoderm lineages in early stages of mouse embryo development.

Around E5.5, a morphologically different domain arises within the VE, at the distal tip of the egg cylinder, called Distal Visceral Endoderm (DVE). From this domain, some cells migrate towards the embryonic/extra-embryonic junction and thereby determine the anterior side of the embryo. These cells, now called Anterior Visceral Endoderm (AVE), express specific markers such as DKK1, CER1, LEFTY1, which modulate the WNT, BMP and Nodal signaling pathways, respectively. By limiting the expansion of the primitive streak [9], and maintaining the anterior Epi in a naïve state, the signaling cues emanating from the AVE are essential for a proper development of the anterior ectoderm [10, 11]. The AVE was also shown to be a source of instructions for organ morphogenesis and placement through the expression of BMP2 [12]. The rest of the VE has a different, but similarly important role. Recent *in vivo* studies of cell movement and single-cell sequencing have revealed that the developing gut is formed by a mixture of embryonic and extraembryonic endoderm. Cells originating from the VE are integrated in the early gut tube around E8.0 [13, 14, 15].

Most of our knowledge of the extraembryonic endoderm has been obtained from *in*

in vivo experiments. Additionally, an *in vitro* model has been derived by adapting extraembryonic endoderm cells from mouse blastocysts to *in vitro* culture [16]. These XEN cells have properties reminiscent of the PE [17], but can be differentiated towards a VE-like state when exposed to the appropriate signaling cues [18, 19]. Aggregate of these XEN cells and mouse embryonic stem cells (mESC) give rise to a new model system, the XEN enhanced-gastruloid (XEG) (Fig. 3.2), which we introduced in detail in chapter 2. Many existing models of mammalian development, such as the gastruloid, are exclusively derived from embryonic stem cells (ESC). However, as the embryo receives crucial signaling cues from extraembryonic tissues [20], other models do incorporate extraembryonic cells [21, 22, 23, 24]. In this chapter we explore the signalling cues produced by the XEN cells that influence XEG formation and organization, as well as their transcriptional dynamics in response to the co-culture with mESCs.

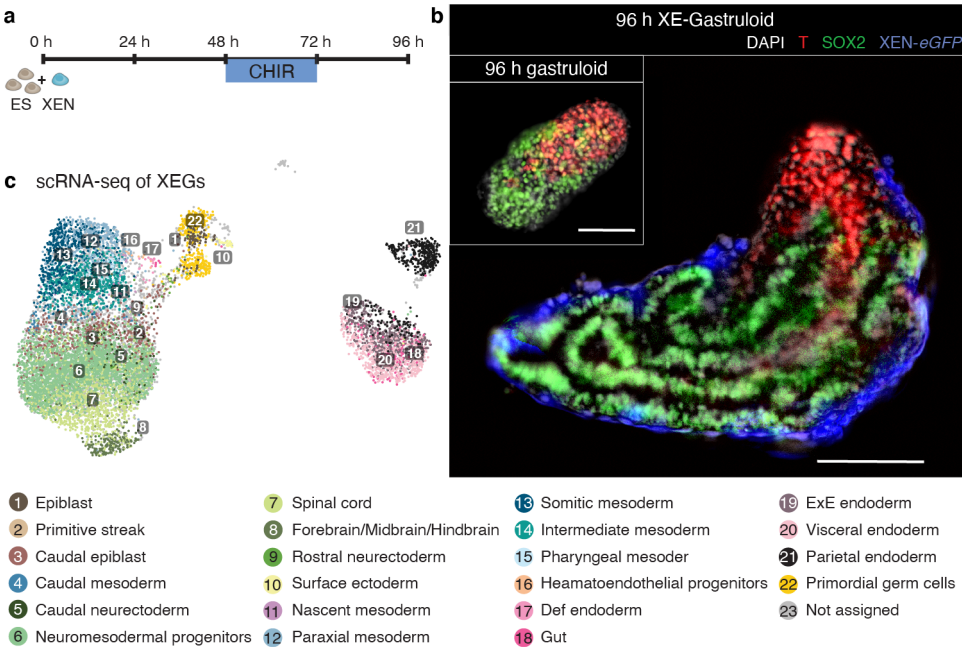


Figure 3.2: XEN enhanced-gastruloids (XEGs) *in vitro* model. **a**, Schematic of the culture protocol: at 0 h, 200 cells (150 ESCs and 50 XEN cells) were aggregated; CHIR was added between 48 h and 72 h after cell seeding; cell aggregates were cultured until 96 h. **b**, T and SOX2 expression in aggregates at 96 h (z-projection of whole mount immunostaining). Inset: expression in aggregate resulting from the standard gastruloid protocol (without XEN cells). Scale bars 100 μm . **c**, UMAPs of cells in XEGs, colored by cell type based on the mapping to *in vivo* data from Pijuan-Sala et al., [25]. Figures from chapter 2.

We reveal that the XEN cells induce the formation of neural tubular structures in XEGs by local inhibition of primitive streak formation, through production of a basement membrane as well as the WNT inhibitor DKK1. We further show that XEN cells in XEGs become visceral endoderm-like due to co-differentiation with the mESCs. Our

study thus reveals a complex interplay between two major embryonic cell types and highlights the possible mechanisms underlying the great morphogenetic potential of *in vitro* models of embryogenesis.

3.2. MATERIALS AND METHODS

3

3.2.1. CELL CULTURE

Routine maintenance of the cells is described in chapter 2, section 2.2. For control of the smFISH experiments, XEN cells were seeded at low density in N2B27 medium. At 48 h the medium was replaced by prewarmed N2B27 supplemented with 3 μ M of GSK3 inhibitor. 72 h after seeding, the medium was replaced with prewarmed N2B27. Cells were fixed at 96 h with 4% PFA for 1 h at 4°C.

3.2.2. GASTRULOIDS

The standard gastruloids protocol is described in chapter 2, section 2.2.

GASTRULOIDS GROWN IN GELTREX

For the experiments with gastruloids grown in Geltrex, cell aggregates were collected at 24 h, 48 h and 72 h and embedded into LDEV-Free, hESC-Qualified, reduced growth factor Geltrex (Gibco) in culture dishes for the rest of the procedure. Only the gastruloids transferred at 72 h showed robust growth. At 96 h, culture dishes were covered with ice-cold PBS and placed on a shaker at 4°C for 10 min. Gastruloids were gently collected by pipetting and washed three times by centrifugation in ice-cold PBS to remove the gel, then fixed with 4% PFA overnight at 4°C.

GASTRULOIDS GROWN WITH DKK1

For gastruloids grown with DKK1, cells were seeded according to the usual protocols. At 24 h, 40 μ L of N2B27 supplemented with various concentration of DKK1 (Sigma-Aldrich), were added to each well. Next steps of the protocol were performed using N2B27 supplemented with DKK1. Aggregates were fixed at 96 h with 4% PFA overnight at 4°C.

3.2.3. XEN ENHANCED-GASTRULOIDS (XEGs)

The standard XEG protocol is described in chapter 2, section 2.2.

XEGs GROWN WITH XEN CONDITIONED MEDIUM

For the experiments with gastruloids grown in XEN-conditioned medium, N2B27 medium was conditioned by XEN cells for 24 h, collected and filtrated with 0.2 μ m low protein

binding filter and stored at -20°C . Gastruloids were seeded according to the usual protocol. After 24 h, $40\ \mu\text{L}$ of XEN-conditioned medium was added to each well. Next steps of the protocol were performed using XEN-conditioned medium. Gastruloids were fixed at 96 h with 4% PFA overnight at 4°C .

XEGS GROWN WITH DKK1 INHIBITOR

XEGs were seeded according to the usual protocols. At 24 h, $40\ \mu\text{L}$ of N2B27 supplemented with various concentration of Way262611 (Sigma-Aldrich) respectively, were added to each well. Next steps of the protocol were performed using N2B27 supplemented with Way262611. Aggregates were fixed at 96 h with 4% PFA overnight at 4°C .

3.2.4. LABELLING AND IMAGING

The samples preparation, immunostaining, smFISH, imaging and image analysis protocol of whole-mount and sections samples are described in chapter 2, section 2.2.

3.2.5. SINGLE CELL RNA SEQUENCING

The single cell RNA (scRNA) library preparation, sequencing and normalization, as well as the identification of the spike-in cells were performed as explained in chapter 2, section 2.2.

MAPPING TO *in vivo* ENDODERM DATASET

The Nowotschin et al. dataset [15], which was downloaded from <https://endoderm-explorer.com/>, consists of mouse embryos from 6 timepoints from E3.5 to E8.75. The data was normalized (scran) and the 10% HVG were calculated (improvedCV2, scran package). First, MNN was applied to the Nowotschin et al. dataset in increasing order of the timepoints (using log-transformed normalized counts of the 10% HVG, $d = 150$, $k = 50$). Then, XEN cells from our XEG dataset (XEN spike-ins (CD140+) and XEN derived (CD140-)) were mapped to the MNN-corrected Nowotschin et al. dataset. knn assignment was performed as described above and resulted in 7 assigned clusters.

DIFFERENTIAL EXPRESSION ANALYSIS

For the differential expression test between “spike-in XENs” and “XENs in XEGs” a Welch t-test (implemented in findMarkers, scran R package) was conducted on the normalized log-transformed counts. The test was performed on XEGs from replicate 2. “spike-in XENs” were chosen as the 100 cells with highest Ab_CD140 expression and “XENs in XEGs” were the 100 cells with lowest Ab_CD140 expression within the XEN identified cells. For all differential expression tests p-values were adjusted for multiple hypothesis testing with the Benjamini-Hochberg method.

3.3. RESULTS

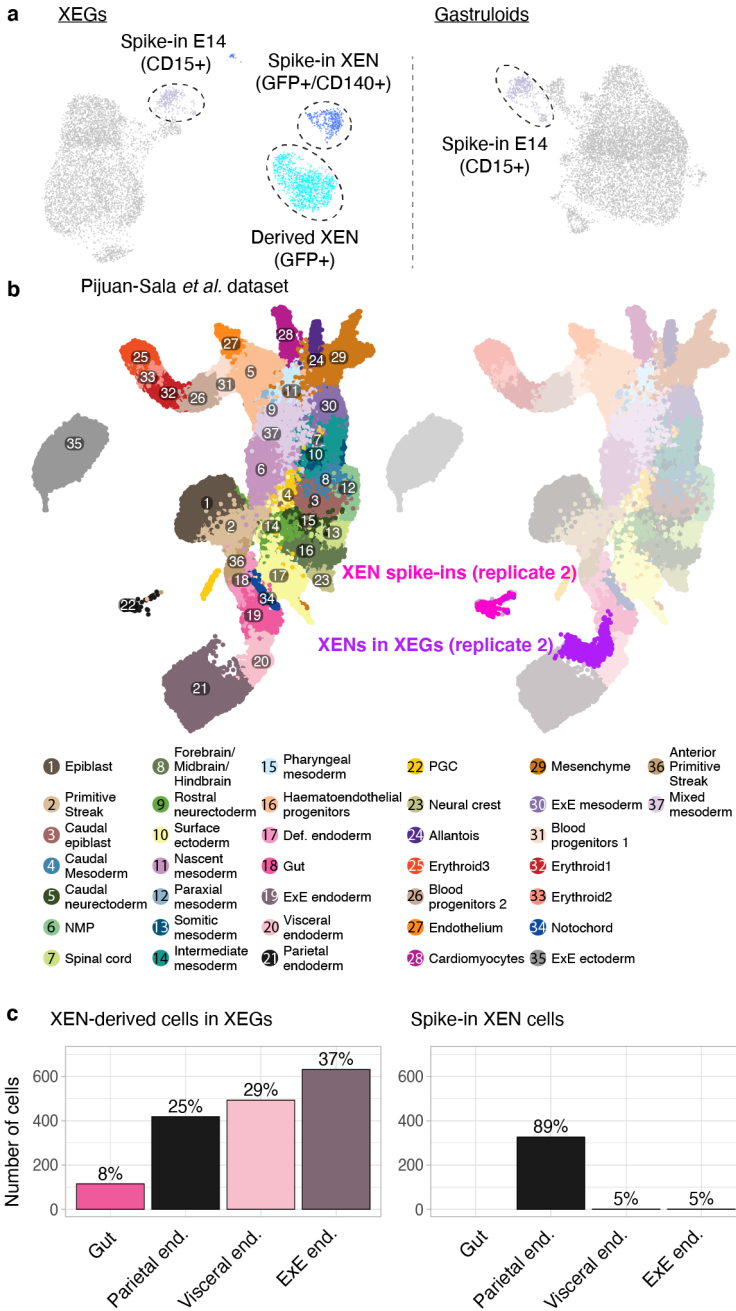
3.3.1. XEN CELLS BECOME VE-LIKE IN XEGS

Compared to gastruloids, XEGs additionally contained extraembryonic endoderm cell types. By using GFP-expressing XEN cells in XEGs (Fig. 3.3a), we established that those cell types were exclusively differentiated from XEN cells.

By comparison to undifferentiated XEN cells, which were spiked into the scRNA-seq, we studied the transcriptional changes in XEN-derived cells. Undifferentiated XEN cells mostly mapped to PE *in vivo* [25] (Fig. 3.3b, c), consistent with a previous study [17, 16]. Their derivatives in XEGs mapped to multiple kinds of extra-embryonic endoderm: PE, embryonic and extraembryonic VE. Interestingly, some also mapped to gut, reminiscent of the contribution of VE to the gut *in vivo* [13, 15].

The identification of those cell types was confirmed by mapping our scRNA-seq data to an endoderm-focused scRNA-seq dataset [15] (Fig. 3.4a, b). Quantification revealed that, on average, 8% of the initially PE-like XEN cells acquired a gut-like and 66% a VE-like transcriptomic profile (29% embryonic VE, 37% extraembryonic VE) when co-cultured in XEGs (Fig. 3.3c). However, 25% retained a PE-like transcriptome (Fig. 3.3c). Differential gene expression analysis between undifferentiated XEN cells and XEN derivatives revealed several differences (Fig. 3.5, Supplementary Table 1). PE markers were less expressed in XEN-derivatives, while VE markers were highly expressed, suggesting that XEN cells differentiate from at PE to a VE-like state in XEGs.

Figure 3.3 (following page): XEN-derived cells are differentiating to VE in XEGs. **a**, UMAPs of cells in XEGs and gastruloids with spike-in cells and XEN derived cells highlighted by color and circled (replicate 2). **b**, UMAPs of the Pijuan-Sala dataset [25]. XEN spike-ins and XEN-derived cells from XEGs replicate 2 (bright colors) are mapped to the *in vivo* datasets (dim colors). Framed cell types correspond to the ones on which the XEN cells mapped. **c**, Left, cell types of XEN-derived cells in XEGs. Cells were classified as gut, PE (“parietal end.”), embryonic VE (“visceral end.”) or extraembryonic VE (“ExE end.”) by mapping to the data set from Pijuan-Sala dataset [25]. Right, cell types of spiked-in XEN cells.



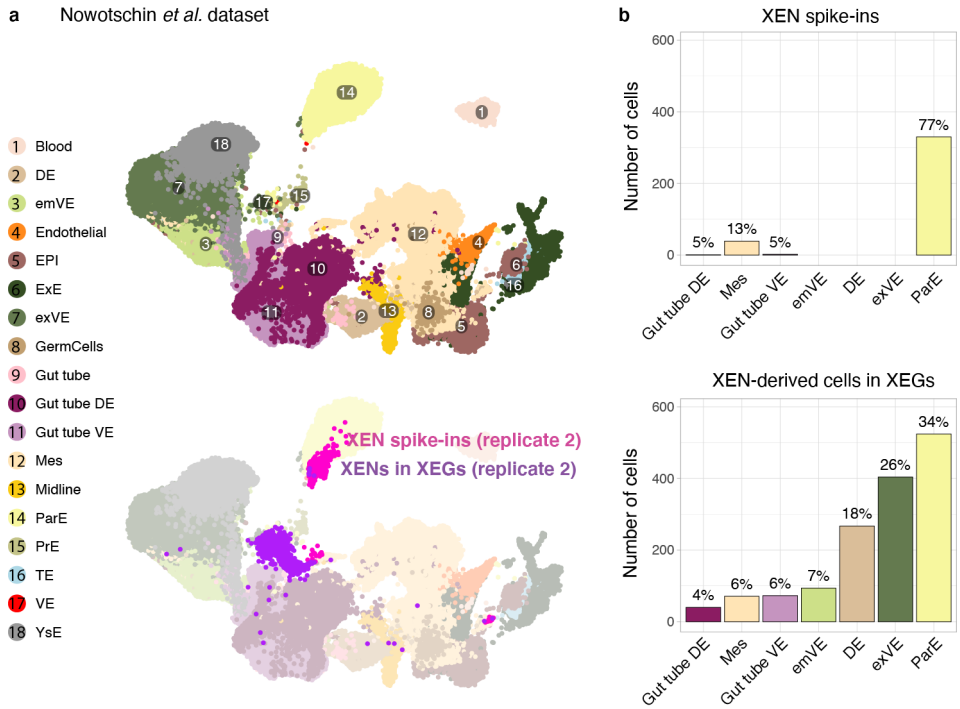


Figure 3.4: Mapping on endoderm-focused scRNA-seq dataset confirmed VE-like identity of XEN-derived cells. **a**, UMAPs of the Nowotschin dataset [15]. XEN spike-ins and XEN-derived cells from XEGs replicate 2 (bright colors) are mapped to the *in vivo* datasets (dim colors). Framed cell types correspond to the ones on which the XEN cells mapped. **b**, Cell type frequencies of XEN spike-ins and XEN derived cells in XEGs, resulting from knn assignments based on the mapping in (a). DE: definitive endoderm, emVE: embryonic visceral endoderm, Epi: epiblast, ExE: extraembryonic ectoderm, exVE: extraembryonic visceral endoderm, Mes: mesoderm, ParE: parietal endoderm, PrE: primitive endoderm, TE: trophoctoderm, VE: visceral endoderm, YsE: yolk sac endoderm.

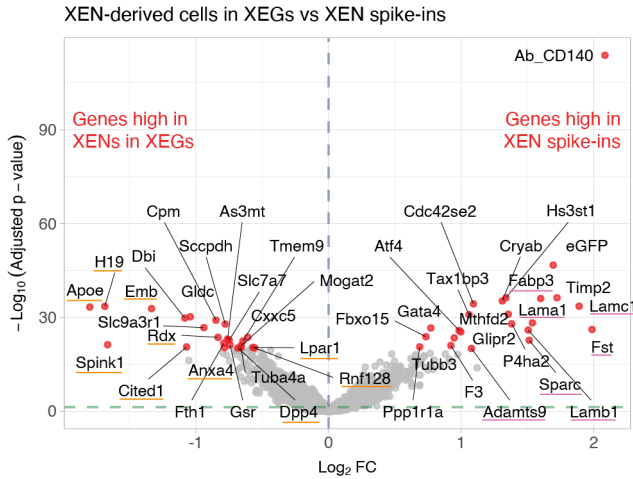


Figure 3.5: Gene expression differences between XEN spike-ins and XEN-derived cells in XEGs. (fold-change vs p-value). Orange and pink lines indicate genes with PE-like and VE-like identity, respectively (see Supplementary Table 1).

To validate this finding experimentally, we performed *smfish* of the PE marker *Fst* [26], the VE marker *Spink1* [27] and the pan-extraembryonic endoderm marker *Dab2* [28] (Fig. 3.6a). This measurement showed that XEN-derived cells in XEGs only expressed *Dab2* and *Spink1*, while undifferentiated XEN cells broadly co-expressed all markers. Some XEN cells in XEGs were also highly expressing E-cadherin, known to be expressed in VE [29] (Fig. 3.6b). However, the more anterior VE marker *Hhex* [30] was not detected by smFISH (Fig. 3.7). Exposing undifferentiated XEN cells to CHIR in the same way as XEGs did not cause differentiation (Fig. 3.6), which suggests that the interaction with mESCs plays a role.

Taken together, these results suggest that XEGs mimic elements of embryo organization *in vivo*, where the VE is in direct contact with the embryo proper. While undifferentiated XEN cells have both PE and VE characteristics, the majority (66%) of these cells becomes more VE-like, due to the presence of mESCs.

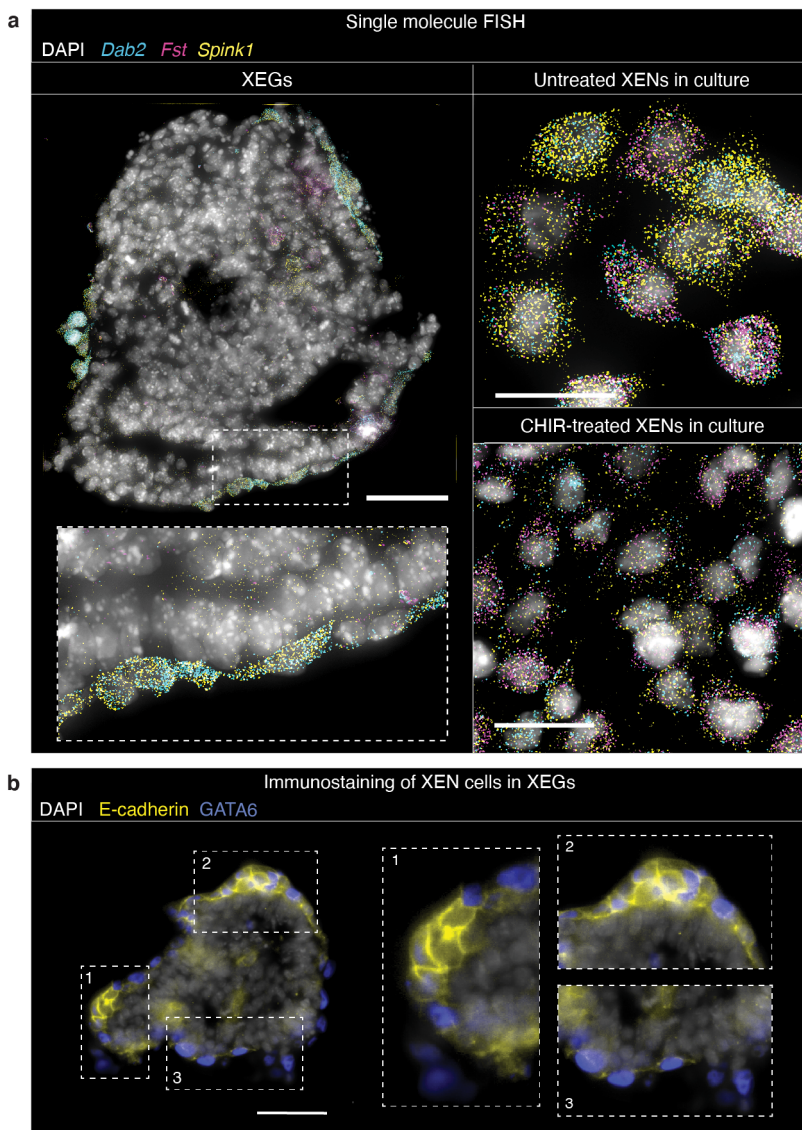


Figure 3.6: XEN-derived cells express VE markers. **a**, *Dab2*, *Spink1* and *Fst* expression in section of a XEG at 96 h (left, scale bar: 50 μm), in XEN cells cultured under standard maintenance conditions (top right, scale bar 20 μm) and XEN cells treated with CHIR according to the XEG protocol (bottom right, scale bar 20 μm). Expression was visualized by smFISH. Each diffraction limited dot is a single mRNA molecule. **b**, Expression of E-cadherin in XEGs at 96 h (immunostaining of sections). XEN cells were localized by immunostaining of GATA6. Scale bars: 50 μm .

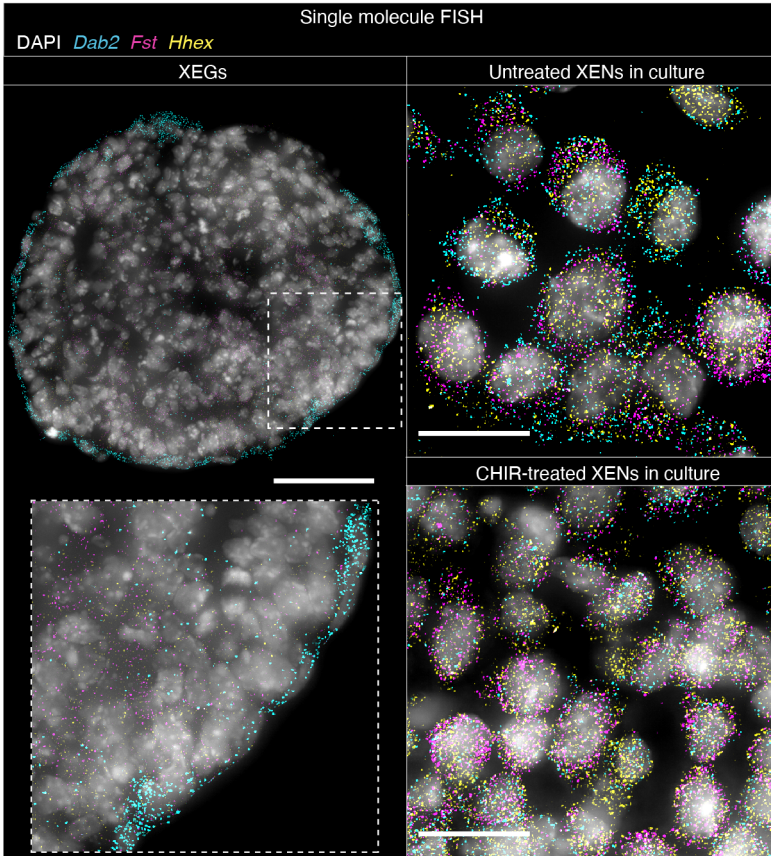


Figure 3.7: XEN-derived cells do not express most anterior VE markers. *Dab2*, *Fst* and *Hhex* expression in a section of a XEG at 96 h (left, scale bar: 50 μm), XEN cells cultured under standard maintenance conditions (top right, scale bar: 20 μm) and XEN cells treated with CHIR according to the XEG protocol (bottom right, scale bar: 20 μm). Expression was visualized by smFISH. Each diffraction limited dot is a single mRNA molecule.

3.3.2. XEN CELLS GUIDE SYMMETRY BREAKING BY LOCALLY INHIBITING THE FORMATION OF A T-POSITIVE POPULATION

Although XEN-derived cells did not express canonical AVE markers (Supplementary Fig. 2.14), we were wondering if they might effectively carry out an AVE-like patterning function in XEGs. XEN cells always formed the outermost layer of XEGs (Fig. 3.2b), resembling *in vivo* organization. Focusing on XEGs partially covered with XEN cells, we observed that tubular structures were always adjacent to the XEN cells, while the T-positive population was on the opposite side (3.8a, b). Notably, this organization was already established at 72 h, when aggregates are still spherical (Fig. 3.8c). This observation suggested that XEN cells guide the symmetry breaking by a local effect on the adjacent mESCs.

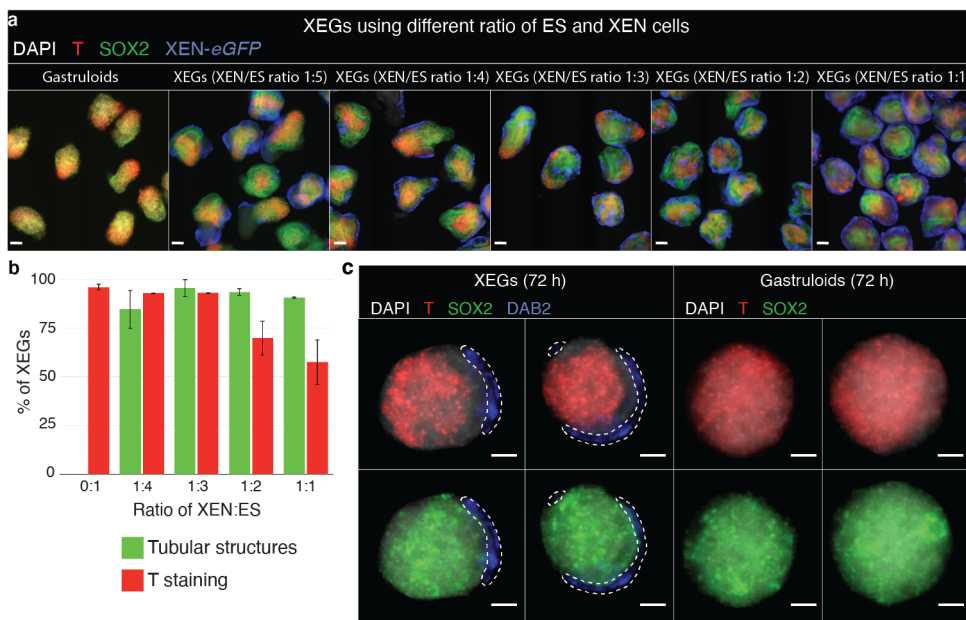


Figure 3.8: T+ compartment is coupled with XEN positioning. **a**, SOX2 (neural progenitors-like cells) and T (primitive streak-like cells) expression in gastruloids and XEGs with different starting ratios of ESCs and XEN cells (z-projection of whole mount immunostaining). **b**, Average fraction of aggregates showing tubular structures and T staining at 96 h for different starting ratios of ESCs and XEN cells (n = 2 experiments, error bars show standard deviation). **c**, T and SOX2 expression in XEGs (left) and gastruloids (right) at 72 h (z-projection of whole mount immunostaining). XEN cells were localized by expression of DAB2 and are indicated by a dashed outline. Scale bars: 100 μm . Cell nuclei were stained with DAPI.

We speculated that this effect might be mediated by a basement membrane produced by the XEN cells. As established in chapter 2, cells were polarized early during XEG development, prior to forming a columnar epithelium (Fig. 2.4). This epithelium was supported by a basement membrane containing laminin and collagen (Fig. 3.9a, b) which were mostly produced by the XEN cells (Fig. 3.9c). It has been demonstrated previously, for small aggregates of mESCs, that the presence of an extracellular matrix can be sufficient for polarization and lumen formation [31, 32, 33]. Growing gastruloids in Geltrex did induce the formation of inner cavities; however, we did not observe a columnar epithelium (Fig. 3.9d).

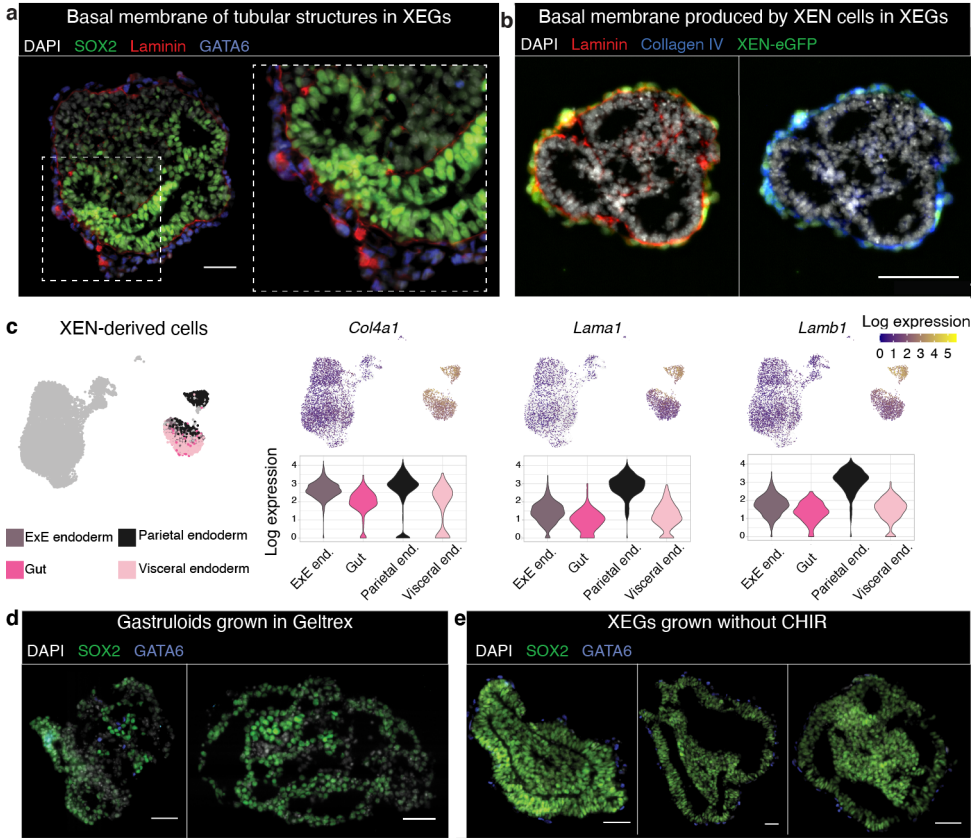


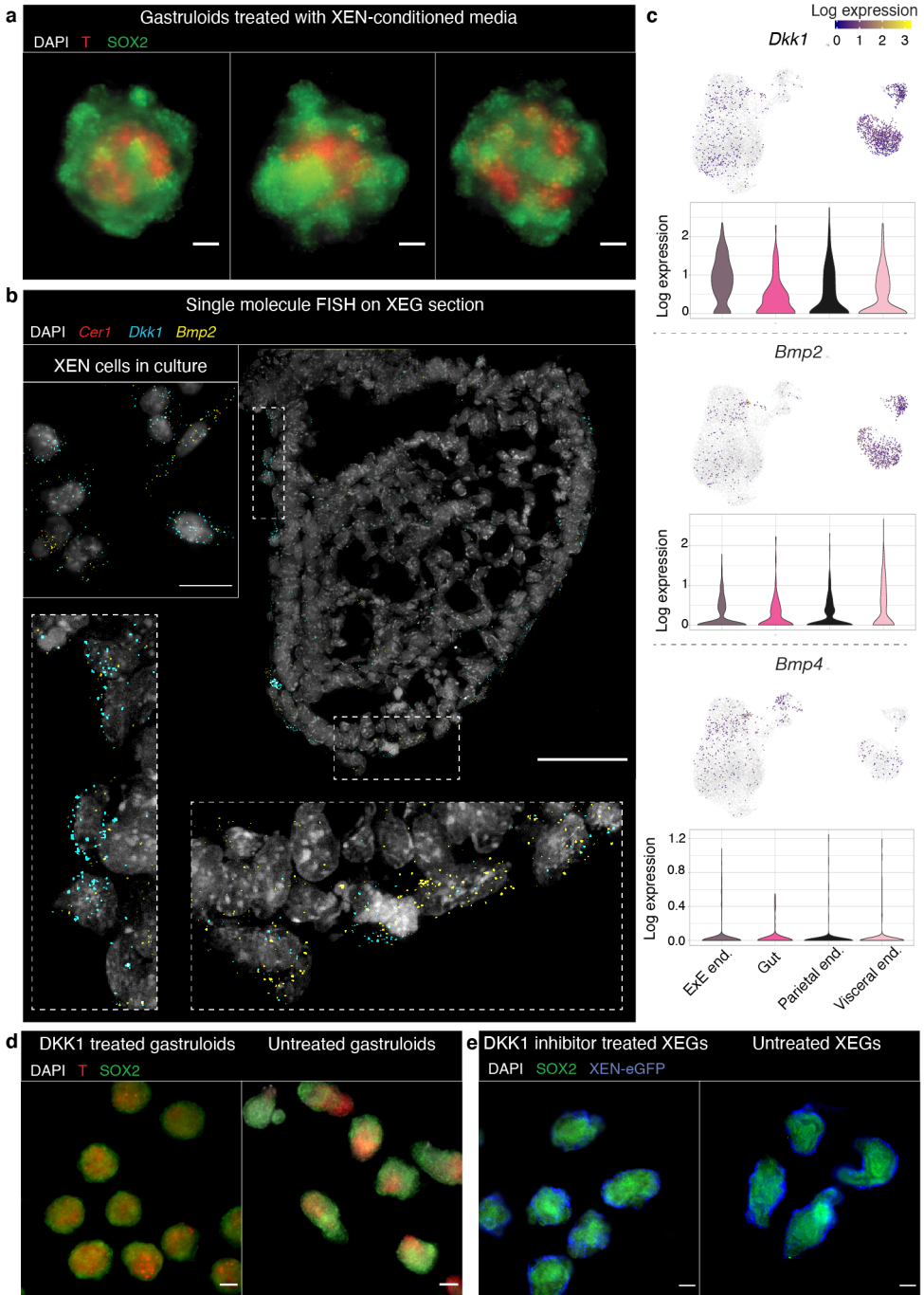
Figure 3.9: XEN cells induce columnar epithelium and lumen formation. **a**, Immunostaining of SOX2 and laminin on 96 h XEGs sections. XEN cells were localized by expression of GATA6. The dashed boxes are shown at a higher magnification in the insets. Scale bars: 50 μm . **b**, Expression of Collagen IV and laminin in XEGs at 96 h (immunostaining of sections). XEN cells were localized by their endogenous expression of H2B-GFP. Scale bars: 100 μm . **c**, Expression of genes in XEN-derived cells. Left, UMAP of cells in XEGs with XEN-derived cells colored by cell type (gut, parietal endoderm (parietal end.), embryonic VE (visceral end.) or extraembryonic VE (ExE end.)). Right, expression of basement membrane components Collagen 4 (*Col4a1*), Laminin alpha 1 (*Lama1*) and Laminin beta 1 (*Lamb1*). UMAP for each gene indicate log expression by color and contain both replicates, batch corrected. A violin plot of log expression in XEN-derived cell types is shown below the UMAP for each gene. **d**, Expression of SOX2 and GATA6 (immunostaining) in sections of gastruloids grown in Geltrex at 96 h. Scale bars 50 μm . **e**, Expression of SOX2 (immunostaining) in sections of XEGs grown without CHIR. No specific T staining could be detected (data not shown). XEN cells were localized by expression of GATA6. Scale bars 50 μm . Cell nuclei were stained with DAPI.

Since an extracellular matrix was not sufficient to obtain neural tube-like structures, we were wondering, which other mechanisms might play a role. We observed that gastruloids grown in medium conditioned by XEN cells, did not elongate and had a T-positive cell population that was restricted to the center of the aggregate (Fig. 3.10a). Thus, we reasoned that diffusible factors produced by the XEN cells might have a role in the formation of the neural tube-like structures. We hypothesized that the WNT inhibitor DKK1,

which is expressed by the XEN cells (Fig. 3.10b, c), might be one of those factors. *In vivo*, DKK1 is expressed by the AVE and limits the growth of the primitive streak [34]. Notably, other factors known to act in this process (CER, LEFTY1) are not expressed in XEGs (Fig. 3.10b, Supplementary Fig. 2.14). Growing gastruloids in the presence of DKK1 resulted in a round morphology, with the T-positive cells confined to the center, as observed for XEN-conditioned medium (Fig. 3.10d, Supplementary Fig. 3.11a). Factors limiting the primitive streak expansion *in vivo* are also known to preserve the anterior part of the epiblast and are thereby necessary for proper ectoderm domain differentiation [11]. Thus, we wanted to explore if DKK1 could have a similar role in XEGs and bias differentiation towards the ectodermal lineage. Growing XEGs with the DKK1 inhibitor WAY-262611 [35] led to a suppression of tubular structures (Fig. 3.10e, Supplementary Fig. 3.11b). Notably, XEGs treated with the DKK1 inhibitor resembled gastruloids, showing elongated shapes but no tubular structures. Since growing XEGs without CHIR resulted in similar epithelial structures as in regular XEGs (Fig. 3.9e), XEN cells likely suppress pre-existing, low-level endogenous WNT activity [36, 37]. Finally, we noticed that XEN-derived cells highly express BMPs (Fig. 3.10c) and that several of the dorsal neural tube markers expressed in XEGs, are induced by BMP signaling (Supplementary Table 2 in chapter 2). Thus, XEN-derived cells might also be responsible for the dorsal patterning of the neural progenitors-like cells in XEGs.

All combined, our experiments suggest that XEN cells guide symmetry-breaking by local inhibition of cell differentiation into a T-positive population. Diffusible factors, likely including DKK1, and the presence of a basement membrane both seem to contribute to the formation of neural tube-like structures.

Figure 3.10 (following page): XEN cells guide symmetry breaking by locally inhibiting primitive streak formation. **a**, T and SOX2 expression in gastruloids grown in XEN-conditioned media at 96 h (z-projection of whole mount immunostaining). **b**, *Cer1*, *Dkk1* and *Bmp2* expression in a section of a XEG at 96 h (scale bar: 50 μm) or XEN cells cultured under standard maintenance conditions (inset, scale bars: 20 μm). Expression was visualized by smFISH. Each diffraction limited dot is a single mRNA molecule. **c**, Expression of genes in XEN-derived cells. From top to bottom: expression of signaling factors DKK1 (*Dkk1*), BMP2 (*Bmp2*) and BMP4 (*Bmp4*). UMAP for each gene indicate log expression by color and contain both replicates, batch corrected. A violin plot of log expression in XEN-derived cell types is shown below the UMAP for each gene. **d**, T and SOX2 expression in 96 h gastruloids treated with 200 ng/mL DKK1 and untreated (z-projection of whole mount immunostaining). Scale bars: 100 μm . **e**, T and SOX2 expression in 96 h XEGs treated with 0.25 μM DKK1 inhibitor WAY-262611 and untreated (z-projection of whole mount immunostaining). Scale bars: 100 μm . Cell nuclei were stained with DAPI.



3.4. DISCUSSION

In this study we developed a new *in vitro* model of embryogenesis that exploits the interaction between embryonic and extraembryonic cells to achieve tissue-level organization. Our results demonstrate that the XEN cells induce the formation of the neural tube-like structures in XEGs (described and characterized in chapter 2) through, at least, two distinct mechanisms.

On one hand, XEN cells are known to produce large quantities of basement membrane components, like laminin and collagen IV [16], abundant in the Reichert's membrane produced *in vivo* by the PE. This is consistent with the state of the XEN cells in culture, mapping to PE in our scRNA-seq data analysis, and previously described as closer to PE than VE [17, 16]. In XEGs, XEN cells produce a thick layer of extracellular matrix. This basement membrane-like structure likely contributes to the epithelialization of the mESCs at the early stages of XEG growth. Similar morphological changes were observed on mESCs embedded in Matrigel [33].

On the other hand, XEN cells in XEGs highly express the WNT antagonist DKK1 [17]. Our signaling perturbation experiments using DKK1 on gastruloids confirmed its inhibiting effect on elongation and positioning of the T+ population. The use of DKK1 inhibitor on XEGs also confirmed that DKK1 was necessary for neural tube-like structures formation, reminiscent of the role of DKK1 produced by the AVE in the egg cylinder. Such an effect is surprising, given that DKK1 acts on the WNT receptor, but WNT signaling is activated downstream by Gsk3 inhibition in our protocol (Supplementary Fig. 3.12). To explain the effect of DKK1, two aspects have to be considered. First, *Dkk1* is expressed throughout XEG formation, while CHIR is only added after 48 h. We could observe that growing XEGs without CHIR still results in the formation of epithelial structures as in regular XEGs, thus DKK1 produced by the XEN cells likely suppresses low-level endogenous WNT activity in the first 48 h of aggregate growth. Secondly, DKK1 has been shown to not only act on the WNT receptors, but also on the distribution of cellular β -Catenin pools [38]. It was shown that, by retaining the β -Catenin at the plasma membrane, and thus preventing its transfer to the nucleus, DKK1 was able to tune cell polarization and cell-cell adhesion.

In conclusion, the fact that XEN cells are differentiating under the influence of mESCs might suggest that the developing Epi might impact VE differentiation *in vivo*. Our observations in XEGs support the notion that reciprocal interactions between co-differentiating embryonic and extraembryonic cells are crucial for proper development. Thus, XEN cells could represent a way to augment existing developmental organoid systems, by providing a basement membrane and extraembryonic signaling inputs.

3.5. SUPPLEMENTARY INFORMATION

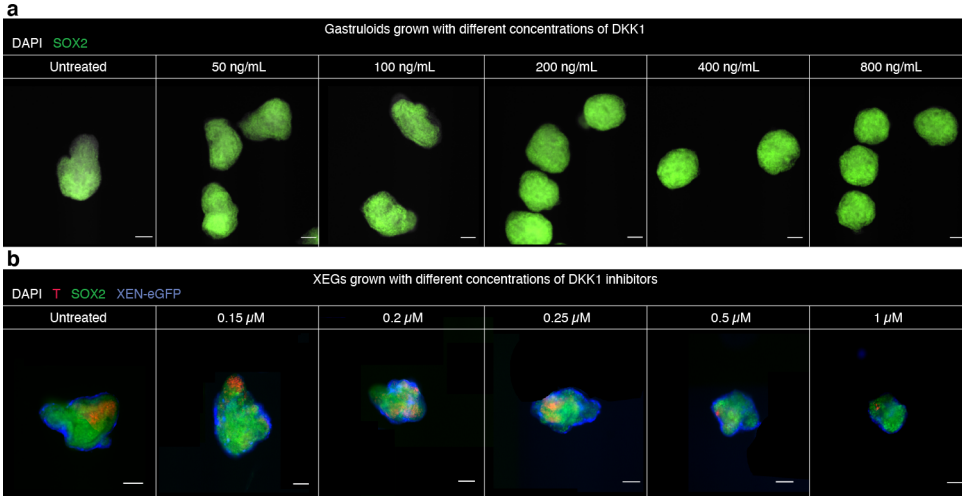


Figure 3.11: Dose response effect of DKK1 and WAY-262611 in gastruloids and XEGs. **a**, Expression of SOX2 (immunostaining) in 96 h gastruloids treated with various concentration of DKK1 between 24 and 96 h. Scale bars: 100 μ m. **b**, Expression of SOX2 and T (immunostaining) in 96 h XEGs treated with various concentrations of DKK1 inhibitor WAY-262611 between 24 h and 96 h. XEN cells were localized by endogenous expression of H2B-GFP. Scale bars: 100 μ m. Cell nuclei were stained with DAPI.

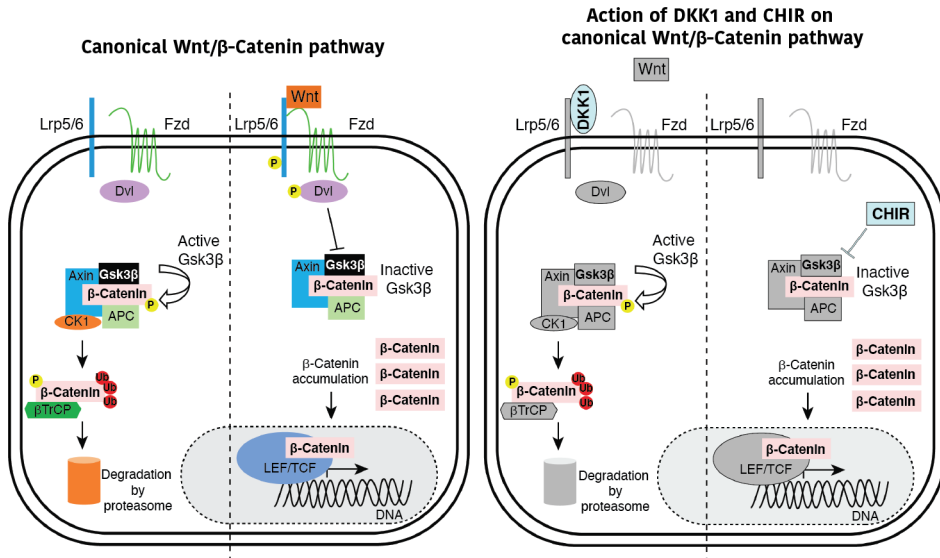


Figure 3.12: Canonical Wnt pathway and effects of CHIR and DKK1.

Gene	Log FC	Expression summary	Reference
Apoe	-1.80	VE PE	Basheeruddin, K., et al., (1987)
H19	-1.69	VE PE	Long, L., Spear, B. T. (2004).
Spink1	-1.66	VE	Goh, H. N., et al. (2014).
Emb	-1.33	pan-endoderm	Brown, K., et al., (2010).
Cited1 (Mgs1)	-1.07	exVE	Dunwoodie, S. L., et al., (1998).
Rdx (Radixin)	-0.83	VE	Igarashi, H., et al. (2018). McClatchey, A. I., et al.,(1997).
Anxa4	-0.78	pan-endoderm	Brown, K., et al., (2010).
Dpp4	-0.68	VE	Sherwood, R. I., et al. (2007).
Lapr1 (Lpa1)	-0.57	VE	Koike, S., et al., (2009).
Rnf128 (Greul)	-0.56	VE	Borchers, A. G. M., et al., (2002).
Gata4	0.77	pan-endoderm	Morrissey, E. E., et al.,(1998).
P4ha2	0.79	PrE	Ohnishi, Y., et al. (2014).
Adamts9	1.08	PE	Jungers, K. A., et al., (2005).
Lamb1	1.51	PE	Miner, J. H., et al., (2004).
Sparc	1.51	PE	Mason, I. J., et al., (1986).
Lama1	1.54	PE	Hogan, B. L., et al., (1980).
Fabp3	1.60	exVE - PE	Cheng, S., et al. (2019). Futaki, S., et al. (2003).
Lamc1	1.89	PE	Smyth, N., et al., (1999).
Fst	1.99	PE	Feijen, A., et al. (1994).

Supplementary Table 1: Genes differentially expressed between XEN-derived cells in XEGs and cultured XEN cells.

ACRONYMS

AP	antero-posterior	OWB	organoid washing buffer
BSA	bovine serum albumin	PFA	paraformaldehyde
CITE-seq	cellular indexing of transcriptomes and epitopes by sequencing	scrNA	single cell RNA
CLAHE	enhance local contrast	scrNA-seq	single-cell RNA sequencing
DV	dorso-ventral	SDS	sodium dodecyl sulfate
E3.5	embryonic day 3.5	tRNA	transfer RNA
HVG	highly variable genes	UMAP	uniform manifold approximation and projection
MNN	mutual nearest neighbors	XEG	XEN enhanced-gastruloids
OCT	optimal cutting temperature compound	XEN	extraembryonic endoderm-like

CELL TYPES

AVE	anterior visceral endoderm	PE	parietal endoderm
DVE	distal visceral endoderm	PrE	primitive endoderm
Epi	epiblast	VE	visceral endoderm
ICM	inner cell mass		

3.6. REFERENCES

- [1] M. Nadjicka and N. Hillman. "Ultrastructural studies of the mouse blastocyst substages." In: *Journal of embryology and experimental morphology* 32.3 (Dec. 1974), pp. 675–695.
- [2] S. Nowotschin, A.-K. Hadjantonakis, and K. Campbell. "The endoderm: a divergent cell lineage with many commonalities." In: *Development* 146.11 (June 2019).
- [3] C. Chazaud et al. "Early Lineage Segregation between Epiblast and Primitive Endoderm in Mouse Blastocysts through the Grb2-MAPK Pathway". In: *Developmental Cell* 10.5 (May 2006), pp. 615–624.
- [4] J. Rossant, C. Chazaud, and Y. Yamanaka. "Lineage allocation and asymmetries in the early mouse embryo." In: *Phil. Trans. R. Soc. B* 358.1436 (Aug. 2003), pp. 1341–1349.
- [5] B. Plusa et al. "Distinct sequential cell behaviours direct primitive endoderm formation in the mouse blastocyst." In: *Development* 135.18 (Sept. 2008), pp. 3081–3091.
- [6] F. Gerbe et al. "Dynamic expression of Lrp2 pathway members reveals progressive epithelial differentiation of primitive endoderm in mouse blastocyst." In: *Developmental biology* 313.2 (Jan. 2008), pp. 594–602.
- [7] R. L. Gardner. "Investigation of cell lineage and differentiation in the extraembryonic endoderm of the mouse embryo." In: *Journal of embryology and experimental morphology* 68 (Apr. 1982), pp. 175–198.
- [8] B. L. Hogan, A. R. Cooper, and M. Kurkinen. "Incorporation into Reichert's membrane of laminin-like extracellular proteins synthesized by parietal endoderm cells of the mouse embryo." In: *Developmental biology* 80.2 (Dec. 1980), pp. 289–300.
- [9] A. Perea-Gomez et al. "Nodal antagonists in the anterior visceral endoderm prevent the formation of multiple primitive streaks." In: *Developmental Cell* 3.5 (Nov. 2002), pp. 745–756.
- [10] P. Thomas and R. Beddington. "Anterior primitive endoderm may be responsible for patterning the anterior neural plate in the mouse embryo." In: *Current biology : CB* 6.11 (Nov. 1996), pp. 1487–1496.
- [11] R. M. Arkell and P. P. L. Tam. "Initiating head development in mouse embryos: integrating signalling and transcriptional activity." In: *Open Biology* 2.3 (Mar. 2012), pp. 120030–120030.
- [12] M. Madabhushi and E. Lacy. "Anterior visceral endoderm directs ventral morphogenesis and placement of head and heart via BMP2 expression." In: *Developmental Cell* 21.5 (Nov. 2011), pp. 907–919.
- [13] G. S. Kwon, M. Viotti, and A.-K. Hadjantonakis. "The endoderm of the mouse embryo arises by dynamic widespread intercalation of embryonic and extraembryonic lineages." In: *Developmental Cell* 15.4 (Oct. 2008), pp. 509–520.

- [14] M. Viotti, S. Nowotschin, and A.-K. Hadjantonakis. “SOX17 links gut endoderm morphogenesis and germ layer segregation.” In: *Nature Cell Biology* 16.12 (Dec. 2014), pp. 1146–1156.
- [15] S. Nowotschin et al. “The emergent landscape of the mouse gut endoderm at single-cell resolution.” In: *Nature* 569.7756 (Apr. 2019), pp. 361–367.
- [16] T. Kunath et al. “Imprinted X-inactivation in extra-embryonic endoderm cell lines from mouse blastocysts.” In: *Development* 132.7 (Apr. 2005), pp. 1649–1661.
- [17] K. Brown et al. “A comparative analysis of extra-embryonic endoderm cell lines.” In: *PLoS ONE* 5.8 (Aug. 2010).
- [18] J. Artus et al. “BMP4 signaling directs primitive endoderm-derived XEN cells to an extraembryonic visceral endoderm identity”. In: *Developmental biology* 361.2 (Jan. 2012), pp. 245–262.
- [19] A. Paca et al. “BMP signaling induces visceral endoderm differentiation of XEN cells and parietal endoderm”. In: *Developmental biology* 361.1 (Jan. 2012), pp. 90–102.
- [20] P. P. L. Tam and D. A. F. Loebel. “Gene function in mouse embryogenesis: get set for gastrulation”. In: *Nature Reviews Genetics* 8.5 (Mar. 2007), pp. 368–381.
- [21] S. E. Harrison et al. “Assembly of embryonic and extra-embryonic stem cells to mimic embryogenesis in vitro.” In: *Science* (Mar. 2017), eaal1810.
- [22] B. Sozen et al. “Self-assembly of embryonic and two extra-embryonic stem cell types into gastrulating embryo-like structures.” In: *Nature Cell Biology* 20.8 (Aug. 2018), pp. 979–989.
- [23] N. C. Rivron et al. “Blastocyst-like structures generated solely from stem cells”. In: *Nature* 557.7703 (May 2018), pp. 106–111.
- [24] B. Mathew et al. “Mouse ICM Organoids Reveal Three-Dimensional Cell Fate Clustering.” In: *Biophysical journal* 116.1 (Jan. 2019), pp. 127–141.
- [25] B. Pijuan-Sala et al. “A single-cell molecular map of mouse gastrulation and early organogenesis.” In: *Nature* 566.7745 (Feb. 2019), pp. 490–495.
- [26] A. Feijen, M. J. Goumans, and A. J. van den Eijnden-van Raaij. “Expression of activin subunits, activin receptors and follistatin in postimplantation mouse embryos suggests specific developmental functions for different activins.” In: *Development* 120.12 (Dec. 1994), pp. 3621–3637.
- [27] J. Hou et al. “A systematic screen for genes expressed in definitive endoderm by Serial Analysis of Gene Expression (SAGE).” In: *BMC Developmental Biology* 7.1 (Aug. 2007), pp. 92–13.
- [28] D.-H. Yang et al. “Disabled-2 Is Essential for Endodermal Cell Positioning and Structure Formation during Mouse Embryogenesis”. In: *Developmental biology* 251.1 (Nov. 2002), pp. 27–44.
- [29] A. Wang et al. “Nonmuscle myosin II isoform and domain specificity during early mouse development.” In: *Proceedings of the National Academy of Sciences of the United States of America* 107.33 (Aug. 2010), pp. 14645–14650.

- [30] P. Q. Thomas, A. Brown, and R. S. Beddington. "Hex: a homeobox gene revealing peri-implantation asymmetry in the mouse embryo and an early transient marker of endothelial cell precursors." In: *Development* 125.1 (Jan. 1998), pp. 85–94.
- [31] A. Meinhardt et al. "3D Reconstitution of the Patterned Neural Tube from Embryonic Stem Cells". In: *Stem Cell Reports* 3.6 (Dec. 2014), pp. 987–999.
- [32] A. Ranga et al. "Neural tube morphogenesis in synthetic 3D microenvironments." In: *Proceedings of the National Academy of Sciences of the United States of America* 113.44 (Nov. 2016).
- [33] I. Bedzhov and M. Zernicka-Goetz. "Self-Organizing Properties of Mouse Pluripotent Cells Initiate Morphogenesis upon Implantation". In: *Cell* 156.5 (Feb. 2014), pp. 1032–1044.
- [34] C. Kimura-Yoshida et al. "Canonical Wnt signaling and its antagonist regulate anterior-posterior axis polarization by guiding cell migration in mouse visceral endoderm." In: *Developmental Cell* 9.5 (Nov. 2005), pp. 639–650.
- [35] J. C. Pelletier et al. "(1-(4-(Naphthalen-2-yl) pyrimidin-2-yl) piperidin-4-yl) methanamine: a wingless beta-catenin agonist that increases bone formation rate." In: *Journal of medicinal chemistry* 52.22 (Nov. 2009), pp. 6962–6965.
- [36] D. Turner et al. "Gastruloids develop the three body axes in the absence of extraembryonic tissues and spatially localised signalling". In: *bioRxiv* (2017), p. 104539.
- [37] D. ten Berge et al. "Wnt Signaling Mediates Self-Organization and Axis Formation in Embryoid Bodies". In: *Cell stem cell* 3.5 (Nov. 2008), pp. 508–518.
- [38] M. Johansson et al. "Dkk1 Controls Cell-Cell Interaction through Regulation of Non-nuclear β -Catenin Pools." In: *Developmental Cell* 51.6 (Dec. 2019), pp. 775–786.

4

In vitro MODEL SYSTEMS FOR EPIBLAST / PRIMITIVE ENDODERM SORTING

Proper spatial organization of differentiating cell types is key to achieve normal embryonic development. Lately, the self-organization potential of mouse embryonic stem cells (mESC) has been used to reproduce some milestones of mammalian development *in vitro*. However, one of the earliest organizational processes in embryonic development, the sorting between epiblast (Epi) and primitive endoderm (PrE), is still not fully understood. In this chapter, we present two mESC-based *in vitro* models, reproducing the Epi/PrE sorting process in a tunable way. The first model, based on the co-aggregation of mESCs and PrE-like cells, offers a convenient way to study the effect of knocking-down specific adhesion molecules. The second model consists into inducing the differentiation of PrE-like cells in aggregates of mESCs. The growth of mESC into cylinder shape of controlled size allowed us to optimize the imaging of the sorting process and quantify the movement of single cells. Together, those two *in vitro* systems could bring significant new insight on the dynamics of the Epi/PrE sorting process. By performing high throughput experiments they will allow us to identify key actors of the sorting process and characterise the dynamics of the cells.

4.1. INTRODUCTION

Proper spatial organization of the different cell types within the embryo is necessary for normal embryonic development. During development, multiple cell types have to rearrange, while differentiating, to create specific tissue structures [1, 2, 3].

One of the first, and simplest, spatial rearrangement of cells in mammalian development is the formation of the epiblast (Epi) and primitive endoderm (PrE). Failure of this crucial process will stop the development of the embryo [4, 5].

Epi and PrE differentiate from the Inner Cell Mass (ICM) of the embryo, around the embryonic day 4.5 (E4.5). The Epi gives rise to all embryonic tissues, while the PrE will mostly contribute to extra-embryonic tissues, such as the yolk sac [6]. Some cells originating in the PrE will eventually be incorporated in the embryonic gut [7, 8, 6]. The differentiation of the ICM into Epi and PrE occurs in a salt-and-pepper pattern, that can be observed from E3.5 [9, 10], quickly followed by the spatial separation of the two cell types. Within less than a day after their differentiation, PrE form a monolayer epithelium of cells at the surface of the Epi cells (Fig. 4.1) [11, 12, 13, 14].

4

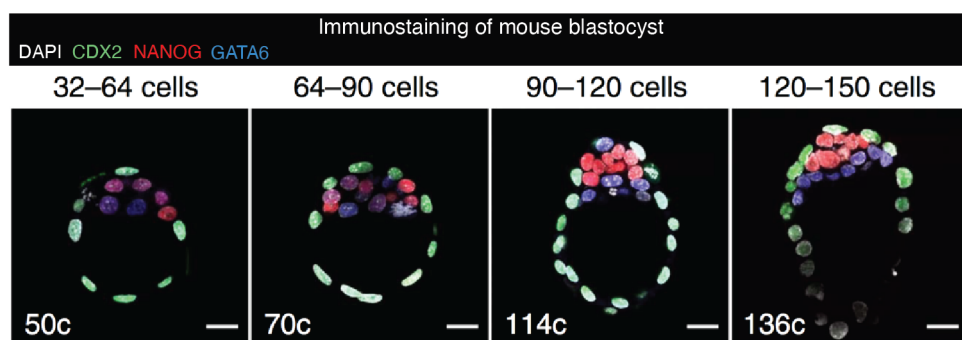


Figure 4.1: Epi/PrE sorting *in vivo*. Representative immunofluorescence images of mouse blastocysts collected at sequential developmental stages from Saiz et al. (2016), *Nature Communication* [15]. Epi and PrE cells are identified by NANOG and GATA6 expression respectively. CDX2 marks the trophoblast cells. Total number of cells ('c') of the embryo shown is indicated. Scale bars, 20 μm .¹

As the sorting process happens at the very beginning of the embryonic development and is completed in a very short time, it has been challenging to study it *in vivo*. Several hypotheses were developed about the mechanism underlying the sorting. The first hypothesis is that differentiated Epi and PrE cells are actively and randomly moving in the ICM, and once the PrE cells reach the surface of the structure, they established an apical polarization and stay fixed [16]. Another hypothesis is that the PrE cells move actively to the surface of the ICM/Epi, directed by an external signal [17]. Alternatively, sorting could be a partially passive process, resulting from a difference in the adhesive properties of both cell types [18]. The differential adhesive affinity hypothesis [19, 20, 21],

¹Figure under open access CC-BY 4.0 Creative Common licence.

postulates that, in an aggregate of cells, the less adhesive ones will spontaneously sort to the outside and envelope the highly adhesive cells in the center. Based on this theory, if the Epi cells have a higher adhesive affinity with themselves than PrE cells, it would result in the exclusion of the PrE cells to the outside. Several experiments have shown a difference in the expression of adhesion molecules between the cell types, supporting the hypothesis that those molecules have a role in the sorting process [16, 22]. However, in other studies the knock-out of specific single adhesion molecules did not prevent the sorting [23]. Those results might indicate that if the Epi/PrE sorting process is based on cellular adhesion, it is not dependent on a single subtype of adhesion molecules but the result of a global difference in the cells adhesion profile.

In this chapter, we explore a way to study the Epi/PrE sorting *in vitro* using mouse embryonic stem cells (mESC). We developed two systems mimicking the process. The first system we developed uses mESCs, to model the Epi, and PrE-like cells, differentiated from mESCs using retinoic acid (RA) to model PrE. The two cell types are combined to form aggregates which we called Chimeric Embryoid Bodies (Chim-EB). This system allowed us to precisely control the state of the cells right before aggregation and, using a GFP reporter of PrE identity, watch the positioning of the PrE-like cells. With this model system we successfully recapitulated the sorting process. However, since sorting occurred concomitantly with aggregation, it was impossible to follow the movement of individual cells.

To transcend the limitations of Chim-EB, we developed another *in vitro* model system in which aggregation and sorting are decoupled. In this system we seed mESCs in cylindrical microwells and differentiate a subpopulation to PrE-like cells using RA. These embryonic disks (ED) thus mimic the *in vivo* differentiation and sorting process more closely than Chim-EB. A fluorescent reporter of PrE allowed us to track PrE-like differentiating cells and reveal sorting dynamics.

4.2. MATERIALS AND METHODS

4.2.1. CELL CULTURE

MOUSE EMBRYONIC STEM CELLS (MESCs) CULTURE

All mESC lines were grown routinely at 37 °C with 5% CO₂ in 2i medium plus LIF (basal DMEM/F12 (Life technologies) supplemented with 0.5× N2 supplement, 0.5 ×B27 supplement, 0.5 mM L-glutamine (Gibco), 20 μg/ml human insulin (Sigma-Aldrich), 1 × 100 U/ml penicillin/streptomycin (Gibco), 0.5× MEM Non-Essential Amino Acids (Gibco), 0.1 mM 2-Mercaptoethanol (Sigma-Aldrich), 1 μM MEK inhibitor (PD0325901, Stemgent), 3 μM GSK3 inhibitor (CHIR99021, Stemgent), 1000 U/ml mouse LIF (ESGRO)). E14 cells were grown on gelatin coating and *Pdgfra*^{H2B-GFP} cells were grown on laminin (from Engelbreth-Holm-Swarm murine sarcoma basement membrane, Sigma-Aldrich) coating. All mESCs were passaged every other day with Accutase (Life technologies) and

replated in tissue-culture treated dishes with the proper coating. E14 mESCs were provided by Alexander van Oudenaarden, the *Pdgfra*^{H2B-GFP} mESC line was provided by the group of Anna-Katerina Hadjantonakis [24]. Those cells express the fusion protein H2B-GFP controlled by the promoter of the PrE marker gene *Pdgfra*. To visualize all the cells during differentiation experiment, this cell line was transfected to express the nuclear fluorescent marker mOrange-H1-10 controlled by the chicken β -actin promoter. For aggregation test, the cell line was transfected to express or the membrane fluorescent marker mCherry-GPI, a fusion protein of mCherry and the anchor peptide GPI. All cell lines were regularly tested for mycoplasma infection.

PRE-LIKE CELLS DIFFERENTIATION AND CULTURE

To obtain PrE-like cells, mESC were differentiated by following the protocol published by Semrau et al. [25]. Briefly, cells were seeded at a low density and grown over night (12 h). After washing cells twice with PBS, differentiation was initiated by replacing the medium by basal N2B27 medium (DMEM/F12 (Life technologies) supplemented with 0.5 \times N2 supplement, 0.5 \times B27 supplement, 0.5mM L-glutamine (Gibco), 1 \times 100U/ml penicillin/streptomycin (Gibco), 0.5 \times MEM Non-Essential Amino Acids (Gibco), 0.1 mM 2-Mercaptoethanol (Sigma-Aldrich)) supplemented with 0.25 μ M of all-trans retinoic acid (Sigma-Aldrich). Differentiation medium was exchanged with fresh one after 48 h. Cells were collected after 96 h. The 96 h collected cells were incubated with antibodies, targeting PDGFRA and coupled with magnetic beads (Miltenyi Biotech) for 30min at 4 °C. The cells solution was then passed through a magnetic MiniMACS MS column (Miltenyi Biotech), to specifically purify PrE-like cells. Purified PrE-like cells were routinely cultured at 37 °C with 5% CO₂ on Mouse Embryonic Fibroblasts (MEFs), in Serum+LIF medium (KO DMEM medium (Gibco) supplemented with 10% ES certified FBS (Gibco), 0.1 mM 2-Mercaptoethanol (Sigma-Aldrich), 1 \times 100 U/mL penicillin/streptomycin, 1 \times MEM Non-Essential Amino Acids (Gibco), 2 mM L-glutamine (Gibco), 1000 U/mL mouse LIF (ESGRO)). PrE-like cells were passaged every other day with trypsin, and replated in tissue-culture treated dishes containing MEFs.

4.2.2. CHIMERIC EMBRYOID BODIES (CHIM-EBS)

AGAROSE MICROWELLS

To grow Chimeric Embryoid bodies, round bottom microwells of 300 μ m of diameter and 800 μ m height were made using agarose and 3D petri dish micro-molds (MicroTissues, 12-256) (Fig. 4.2a). Molds were autoclaved prior each use. To make the microwells, the molds were filled with 500 μ L of 5% sterile low-melting agarose (Sigma-Aldrich) and covered with a petri dish to obtain a flat bottom. After 20 min, the microwells dishes were gently separated from the mold, and use directly for experiments. Molds were cleaned with 70% ethanol and stored before the next used at room temperature.

CHIM-EBS

Before seeding cells, agarose microwells were placed in 3.5 cm dishes and immersed overnight in N2B27 medium to equilibrate.

mESC and PrE-like cells were collected from culture dish using PBS+EDTA 1,0 mM incubated for 10 min or 0.25 % trypsin incubated for 3 min. Most MEFs supporting PrE-like cells growth stayed in the culture dish. To discard the remaining MEFs, PrE-like cells were purified using PDGFRA targeting antibodies coupled to magnetic beads and Mini-MACS MS columns (Miltenyi Biotech).

A solution of 2.0×10^5 cells per mL (10% PrE-like cells, 90% mESCs) in 200 μ L of N2B27 medium was deposited in each agarose microwells structure (~ 150 cells per well)(Fig. 4.2b). The dishes were centrifuged for 5 min at 500 r.p.m., then 1.5 mL of N2B27 medium was added into the dish, around the microwells. Cell were let aggregating and the resulting aggregate was imaged after 16 h (overnight) with a Nikon Ti-Eclipse epifluorescence microscope equipped with an Andor iXON Ultra 888 EMCCD camera and custom-made fluorescence filter sets (Nikon).

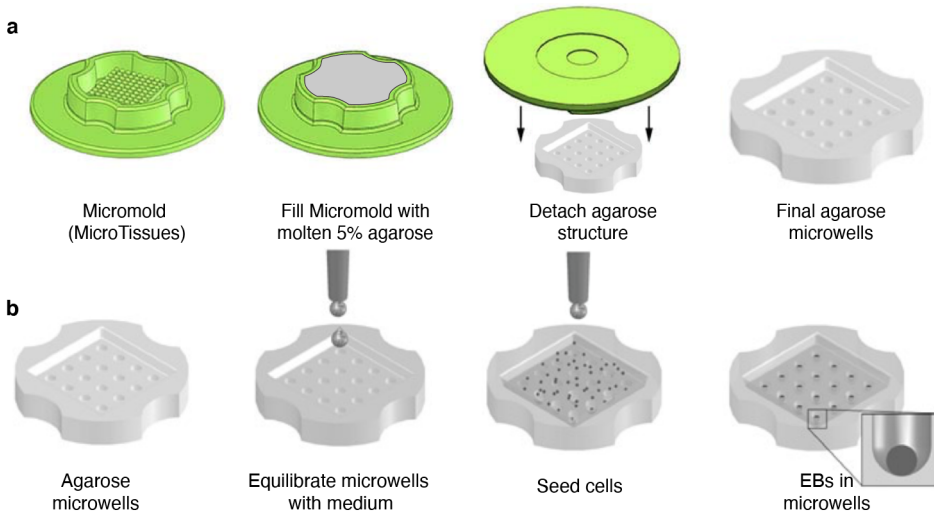


Figure 4.2: Agarose microwells formation and EBs growth. **a**, Schematic of agarose microwell formation. 5% molten agarose was poured into a micro-mold from MicroTissues. After 20 min, the agarose pieces were separated from the molds. **b**, Schematic of EBs formation and growth in agarose microwells. Agarose microwells were placed in culture dish and equilibrated with medium overnight. Cells were pipetted onto the agarose microwells, the next day EBs were formed.

CHIM-EBs DATA ANALYSIS

Chim-EBs pictures were cropped to isolate single aggregate and saved as composite of BrightField pictures of EBs and fluorescence pictures of the *Pdgfra*^{H2B-GFP} PrE-like cells with Fiji [26]. Analysis of the composite pictures was performed with a home-made segmentation pipeline made in CellProfiler [27] creating masks of Chim-EBs and extracting PrE-like cells positions on those masks. Those positions values were plotted and statistically analyzed (unpaired samples t-test) with R (V 3.5.1) and RStudio (V 1.1.463).

4.2.3. RETINOIC ACID DRIVEN DIFFERENTIATION IN 3D

EMBRYOID BODIES (EBs)

To grow embryoid bodies (EBs), we used the same round bottom agarose microwells (300 μ m of diameter and 800 μ m height) as for the Chimeric Embryoid bodies (Method Fig. 1a). After separation from the mold, microwells were placed in 3.5 cm cell culture dishes and immersed overnight in 2i medium to equilibrate.

Cells were seeded at a concentration of 2.0×10^5 cells per mL (~150 cells per well) in 200 μ L of 2i medium per microwells structure. The dishes were centrifuged for 5 min at 500 r.p.m., then 1.5 mL of 2i medium was added into the dish, around the microwells. Cells were let aggregating overnight before starting the differentiation. After washing cells with PBS, differentiation was initiated, by changing the medium for basal N2B27 medium supplemented with 4 μ M of all-trans retinoic acid (Sigma-Aldrich). Differentiation medium was exchanged with fresh one after 48 h. For time lapses, differentiating EBs were imaged from 24 h to 48 h of differentiation (time step of 30 min) with a Nikon Ti-Eclipse epifluorescence microscope equipped with an Andor iXON Ultra 888 EMCCD camera and custom-made fluorescence filter sets (Nikon).

PDMS MICROWELLS

To grow embryoid disk, glass bottom microwells of 100 μ m diameter and 150 – 180 μ m height were made using PDMS and home-made silicon wafers (Fig. 4.3a).

To create a wafer with pillars of desired dimensions, we designed a mask of 100 μ m diameter round shapes. Silicon wafer were cleaned in a plasma chamber for 10 min and coated with SU8-50 photoresist spun at 1250 r.p.m. for 1 min. The photoresist on the wafer was pre-baked for 30 min at 95 °C and let cool down at room temperature. The wafer was then exposed to a nitrogen lamp at 15 mW/cm² for 45 sec through the designed mask. Photoresist structures were baked again for 1 min at 65 °C, then 10 min at 95 °C. After cooling down, the wafer was developed for 10 min in SU8 developer, rinsed with isopropanol and dried with nitrogen. Final baking of the photoresist was performed for 30 min at 150 °C. The resulting wafer was let to cool down at room temperature overnight before the first use. After each use, wafers were cleaned with sequentially with MilliQ, and twice with isopropanol. They were finally dried with nitrogen and stored at room temperature until next use.

To create the microwells, wafers were pre-coated using 10 mL of 10% PVA solution (Sigma-Aldrich), distributed evenly with spinning (1 min, 1200 r.p.m.) and baked at 95 °C for 10 min. Wafers were then coated with 10 mL of 1:10 PDMS, distributed evenly with spinning (2 min, 2000 r.p.m.), and baked overnight at 65 °C. After cooling down at room temperature, PDMS was gently separated from the wafer with tweezers after immersion in milliQ water to dissolve the PVA layer. The obtained PDMS structures were then put on UV-ozone treated glass coverslips. The resulting microwells were baked 15 min at 120 °C to fix the whole structure. PDMS microwells could be stored at room temperature in sterile conditions until use.

EMBRYOID DISKS (EDS)

Before seeding cells, microwells were equilibrate at least 2 h with 2i medium, then briefly spun to remove the remaining air bubbles. Cells were seeded at a concentration of 3.0×10^5 mESCs per mL in 2i medium and let grow for 48 h to create colonies in the microwells (Fig. 4.3b). After washing cells with PBS, differentiation was initiated, by changing the medium for basal N2B27 medium supplemented with $4 \mu\text{M}$ of all-trans retinoic acid (Sigma-Aldrich). Differentiation medium was exchanged with fresh one after 48 h. For time lapses, differentiating EDs were imaged from 24 h to 48 h of differentiation (time step of 30 min) with a Nikon Ti-Eclipse epifluorescence microscope equipped with an Andor iXON Ultra 888 EMCCD camera and custom-made fluorescence filter sets (Nikon).

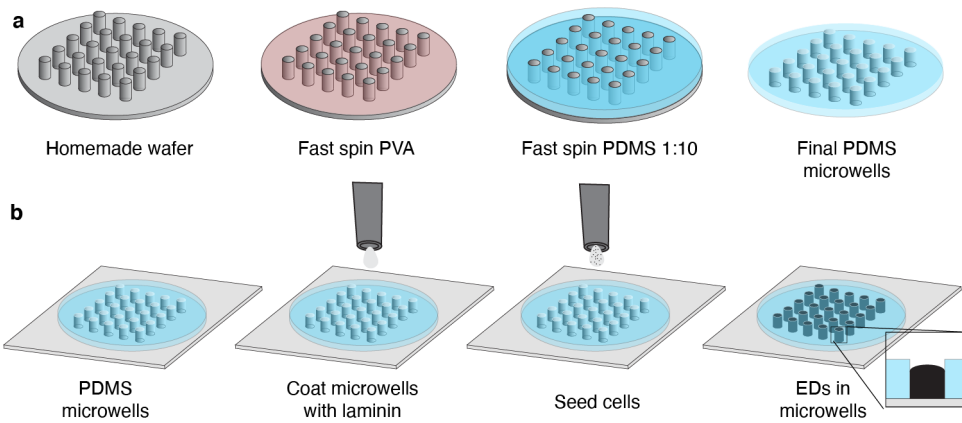


Figure 4.3: PDMS microwells formation and EDs growth. **a**, Schematic of PDMS microwell formation. A template on a Silicon wafer was produced by photolithography, containing pillars of desired dimensions (diameter $100 \mu\text{m}$ and height range $150 - 180 \mu\text{m}$). The wafer was coated with PVA by pipetting a PVA solution onto the wafer and fast spinning (1 min, 1200 r.p.m.). PVA coated wafer was baked 10 min at 95°C to dry the PVA layer. To create the microwells, PDMS was poured on the wafer, fast-spun (2 min, 2000 r.p.m.) and baked overnight at 65°C . After baking, the PDMS was detached from the wafer by dissolving the PVA layer, and placed on glass coverslip. **b**, Schematic of EDs growth in PDMS microwells. PDMS microwells mounted on glass coverslip were coated with laminin and mESC were seeded. After 48 h, most of the wells were filled with a mESC colony.

EDS DATA ANALYSIS

EDs time lapses pictures were cropped using Fiji [26]. For each cell, the focus plan at every time point was manually identified and isolated. The resulting data were saved as sequences of single TIFF files. The segmentation of the PrE-like cells in EDs time lapses data was performed with a custom Matlab (V R2014b) code, based on the existing code Schnitzcells [28]. Briefly, the segmentation step was replaced by a homemade adaptation of the WaterShed segmentation method, excluding all objects outside the microwell area. The tracking code stayed the same. For the analysis of the tracks, the Matlab package @msdanalyzer [29] was used to calculate and fit mean square displacement (MSD) and velocities. Statistical tests were performed in R (V 3.5.1) with Rstudio (V 1.1.463).

4.3. RESULTS

4.3.1. EPI/PRE SORTING IN CHIMERIC EMBRYOID BODIES

To study the Epi/PrE sorting *in vitro*, we first turned to unstructured 3D aggregates of cells cultured *in vitro*. To mimic the Epi (E4.0 - E6.0), we used mESCs which are extracted from the ICM of the embryo around E3.5, but resemble E4.5 Epi after adaptation to *in vitro* culture conditions [30]. mESCs self-renew *in vitro* while remaining pluripotent, which means they can in principle be differentiated to all cell types in the adult body. To model the PrE, we differentiated mESCs with RA, following a protocol which results in a subpopulation of cells that resemble the PrE transcriptionally [25, 31]. To identify those cells, we used a cell line harboring a reporter for the PrE specific marker PDGFRa (*Pdgfra*^{H2B-GFP}). After differentiation, the cells were collected from the dish and PrE-like cells were purified using magnetic beads coated with anti-PDGFRa antibodies.

In order to study the sorting process with mESCs and PrE-like cells, we created aggregates of these two cell types, which we call Chimeric Embryoid Bodies (Chim-EB). To create Chim-EBs in large numbers we deposited mixtures of various concentrations of cells into agarose microwells, by centrifugation (Fig. 4.4a). This is a simple way to bring the cells in close spatial proximity but allows only rough control over the number of cells per aggregate.

After overnight incubation (16h), we observed that PrE-like cells and mESCs already sorted, as expected, with PrE-like cells forming a single layer on the outside of the Chim-EBs. This organization was observed independently of the initial number of cells and the size of the aggregates (Fig. 4.4b). *In vivo* sorting occurs within a day after differentiation of the cells and ends up with the PrE forming an epithelium around Epi cells [11]. Thus, this rearrangement in Chim-EBs happened in less than a day, resembling the timing of the sorting in embryos.

We then performed time lapse experiments to follow the movement of the cells during the sorting process. These experiments revealed that sorting occurred concomitantly with aggregation (Fig. 4.4c, d). Thus, it was difficult to identify the interaction and movement of individual cells in the forming aggregate. In an attempt to decouple aggregation from sorting, we sought to accelerate aggregation. To that end we explored different methods to better preserve adhesion molecules, which get partially lost due to enzymatic dissociation from the culture dish (Fig. 4.4c). Using a non-enzymatic method to detach cells dramatically reduced aggregation time from 8 h to 1 h 20 (Fig. 4.4d). However, the sorting was accelerated to a similar degree, so it was still impossible to separate the two processes.

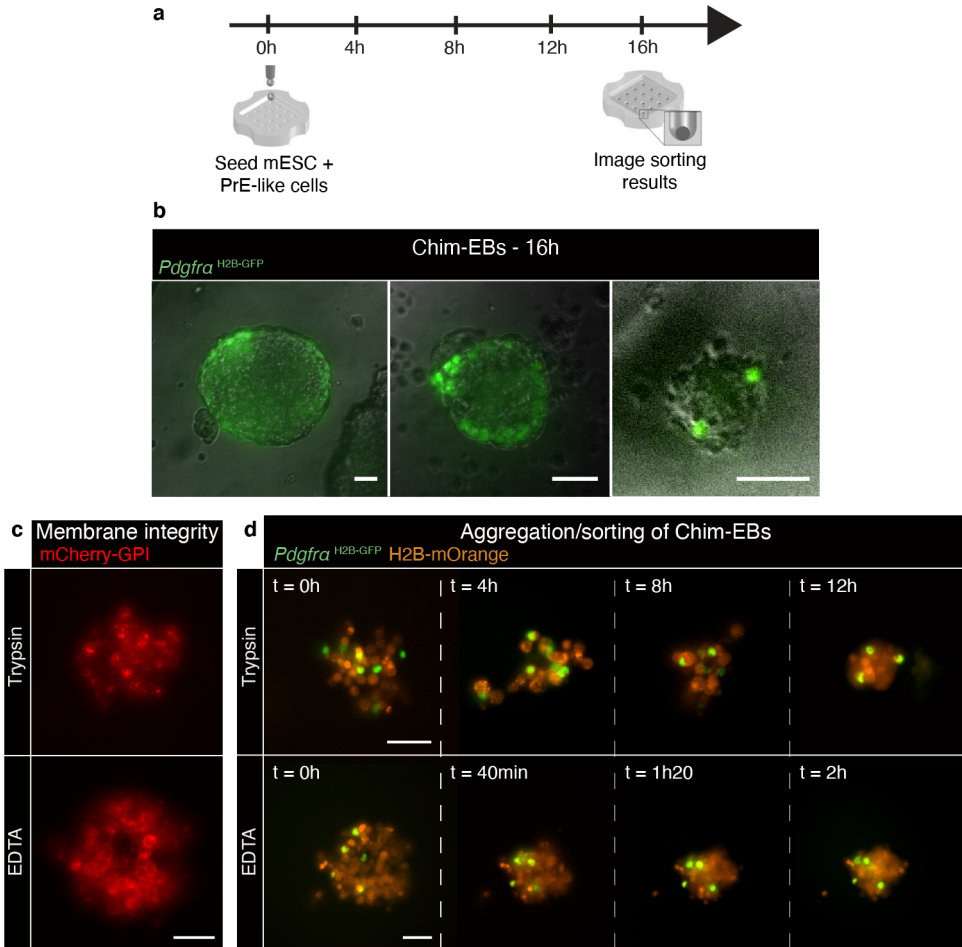


Figure 4.4: Epi/PrE sorting is recapitulated during aggregation of Chim-EBs. **a**, Schematic of the Chim-EB protocol: PrE-like cells were differentiated from mESCs by RA treatment. At 0 h, PrE-like cells and mESCs were deposited into agarose microwells and aggregated overnight to form Chim-EBs. After 16 h, Chim-EBs were imaged. mESCs and PrE-like cells reorganized, with PrE-like cells forming a single layer around the mESCs. **b**, mESC/PrE-like sorting pattern in Chim-EBs of different sizes, 16 h after seeding. PrE-like cells are identified by *Pdgfra*^{H2B-GFP} expression. **c**, Plasma membrane integrity after detachment from the culture dish with trypsin (top) or EDTA (bottom). Cell membranes were marked by expression of an mCherry-GPI fusion protein. **d**, Live cell imaging of Chim-EB aggregation after cell seeding. PrE-like cells and mESCs were dissociated using trypsin (top) or EDTA (bottom). Scale bars: 50 μ m.

As it became clear that Chim-EBs cannot be used to study the dynamics of sorting, we wanted to explore whether they can still be used to study qualitative effects of adhesion molecule knock-down. Adhesion molecules are considered as key molecular players in cell sorting mechanisms [1, 2, 3]. Earlier studies investigated the role of specific molecules in the Epi/PrE sorting process *in vivo* [4, 22], and *in vitro* in cultured embryos [23], or EBs [22, 16], but comprehensive and complete screening has not been carried

out so far. Due to the high level of control and accessibility of our *in vitro* Chim-EB system, we could in principle test the role of a large set of adhesion molecules using siRNA libraries or other knockdown strategies.

To demonstrate the feasibility of this approach, we studied the effect of integrin $\beta 1$ knock-down. Integrin $\beta 1$ subunits are expressed in both Epi and PrE, in which they dimerize with $\alpha 6$ and $\alpha 5$ subunits respectively [32]. Its knock-out has been shown to be lethal for the embryo during peri-implantation development, at the blastocyst stage (E4.5) [5]. Integrin $\beta 1$ -null embryos showed a retarded growth of the ICM and defective PrE migration. A similar effect was observed *in vitro*: integrin $\beta 1$ -null embryoid bodies contain PrE-like cells but they fail to migrate to the surface and form a proper layer [32]. To abrogate integrin $\beta 1$ function with maximal specificity and minimal perturbation of unrelated cell physiology, we exposed the mESC/PrE-like cells mixtures to a blocking antibody (HMb1-1), just before aggregation (Fig. 4.5a).

4

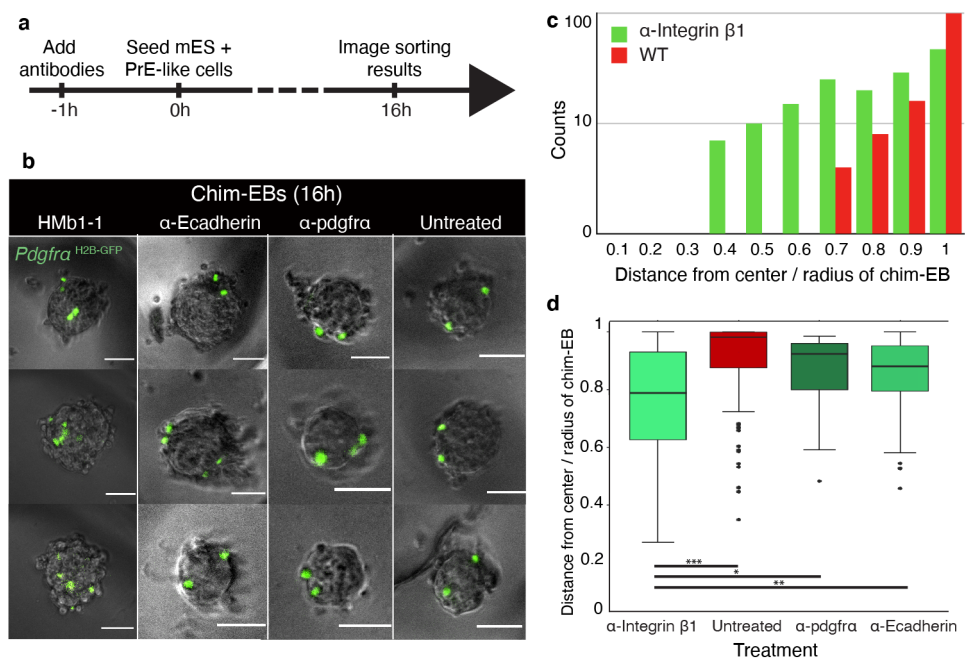


Figure 4.5: Blocking Integrin $\beta 1$ affects Chim-EBs sorting pattern. **a**, Schematic of the Chim-EBs antibodies treatment protocol. Before being seeded in agarose microwells, mESCs and PrE-like cells were treated with the antibody of interest for 1 h. Treated cells were then deposited in the agarose microwells and aggregated overnight. Chim-EBs were imaged after 16 h. **b**, Examples of mESC/PrE-like sorting pattern in Chim-EBs treated with different antibodies. Each column show Chim-EBs treated with a specific antibody or WT. HMb1-1 is functionally blocking integrin $\beta 1$. PrE-like cells are identified by *Pdgfra*^{H2B-GFP} expression. Scale bars: 50 μ m. **c**, Distribution of PrE-like cell positions according to the aggregate size of Chim-EBs treated with HMb1-1 and untreated. Radius of EBs were normalized so 0 is the center of the aggregate and 1 the edge. **d**, Distribution of PrE-like cell positions in Chim-EBs treated with different antibodies and untreated. Radius of aggregates were normalized so 0 is the edge (p-values: * < 0.05 , ** < 0.0005 , *** $< 5e-14$).

This antibody is known to block integrin $\beta 1$ activity by binding at the N-terminus [33, 34]. When HMB1-1 was used, PrE-like cells remained distributed within the Chim-EBs instead of sorting to the surface (Fig. 4.5b, c). Antibodies binding to other membrane proteins (E-cadherin, PDGFR α) had no such effect (Fig. 4.5b, d). The observed effect was thus related to integrin $\beta 1$ function.

Time lapse imaging of Chim-EB formation with HMB1-1 present revealed similar aggregation dynamics as observed before. However, PrE-like cells did not move to the outside but remained relatively static (Fig. 4.6). These results confirm that integrin $\beta 1$ plays an important role in the interaction between the cells and the organization of the two cell types.

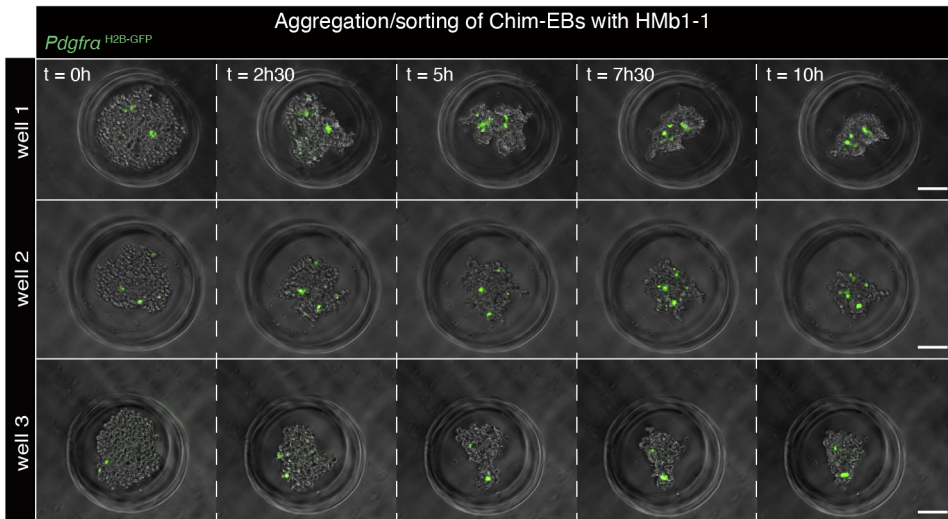


Figure 4.6: Live cell imaging of Chim-EB aggregation under treatment with HMB1-1. Scale bars: 100 μ m.

All in all, our results show that Chim-EBs can be used as an easy and quick assay to screen adhesion molecules and investigate their role in Epi/PrE sorting.

4.3.2. RA DIFFERENTIATION OF mESCs IN 3D MICROSTRUCTURES AS A MODEL FOR EPI/PrE SORTING

Our experiments with Chim-EBs made it clear that cell sorting occurs concomitantly with cell aggregation and aggregate compaction, which makes it difficult to observe the sorting process. Hence, we wanted to explore an alternative where we first aggregate the cells and then induce differentiation in a subset. Such an approach would also mimic *in vivo* Epi/PrE sorting more closely than the previously developed Chim-EBs.

We first tested spherical aggregates of mESCs harboring a nuclear, fluorescent reporter of PDGFR α (*Pdgfra*^{H2B-GFP}), a marker of PrE cells. Similarly to Chim-EBs, those

aggregates were formed by depositing cells into agarose microwells by centrifugation (Fig. 4.2). Treatment of aggregates with RA 24 h after cell seeding resulted in a subpopulation of PDGFR α positive cells. A prior 2D study of RA differentiation has characterized these cells in details and shown that they have a PrE-like profile [25]. 48 h after the start of RA differentiation, the first PDGFR α positive, PrE-like cells were already positioned on the outside of the aggregate. Thus, we performed time lapse experiments between 24 h and 48 h of RA treatment to observe cell movements (Fig. 4.7a). These measurements revealed that PrE-like cells do not preferentially appear on the outside of the aggregate, but throughout the aggregate (Fig. 4.7b, see $t = 0$ h and 5 h). Shortly afterwards, the cells move to the outside. Spherical aggregates of mESCs under RA thus recapitulate sorting between PrE-like and Epi-like cells.

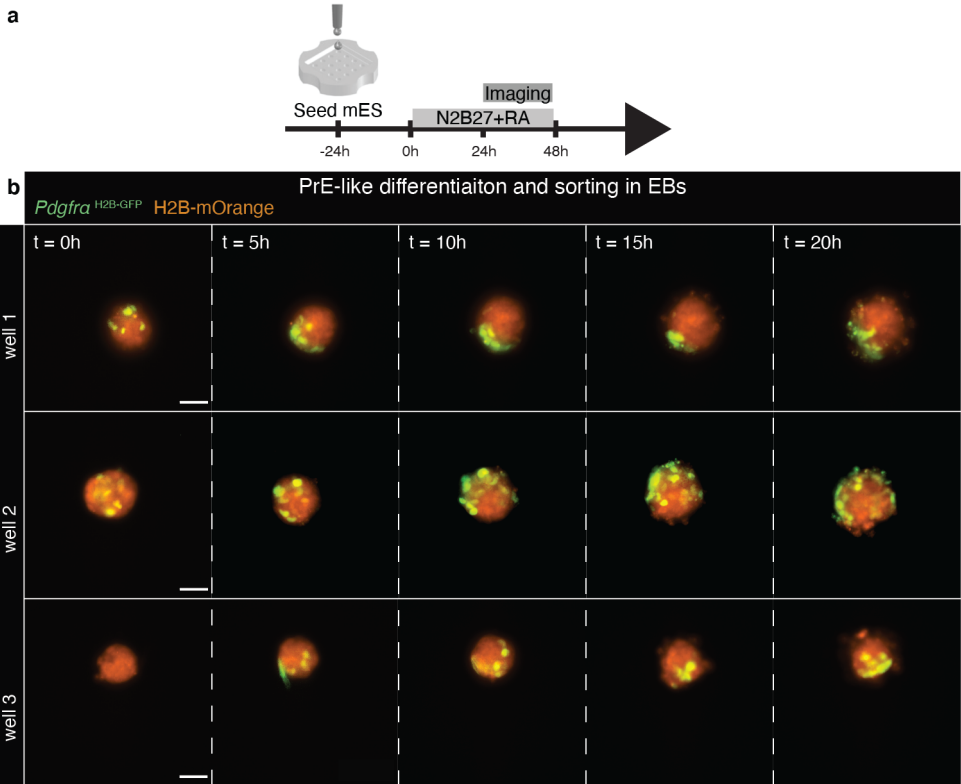


Figure 4.7: Epi/PrE sorting is recapitulated in EBs differentiated with RA. **a**, Schematic of the RA induced PrE-like differentiation and sorting in EBs. mESCs were seeded and aggregated in agarose microwells. The next day, the medium was supplemented with $4\mu\text{M}$ RA. 24 h after RA addition, EBs were imaged to follow the differentiation and migration of PrE-like cells (identified by *Pdgfra*^{H2B-GFP} expression). **b**, Live cell imaging of PrE-like cell differentiation and migration in EBs treated with RA. All cells nuclei are labelled with the fluorescent protein mOrange linked to the histone protein H1-10 (in orange in those pictures). PrE-like cells nuclei are additionally labelled with the genetic reporter *Pdgfra*^{H2B-GFP} (in green in those pictures). Scale bars: $50\mu\text{m}$.

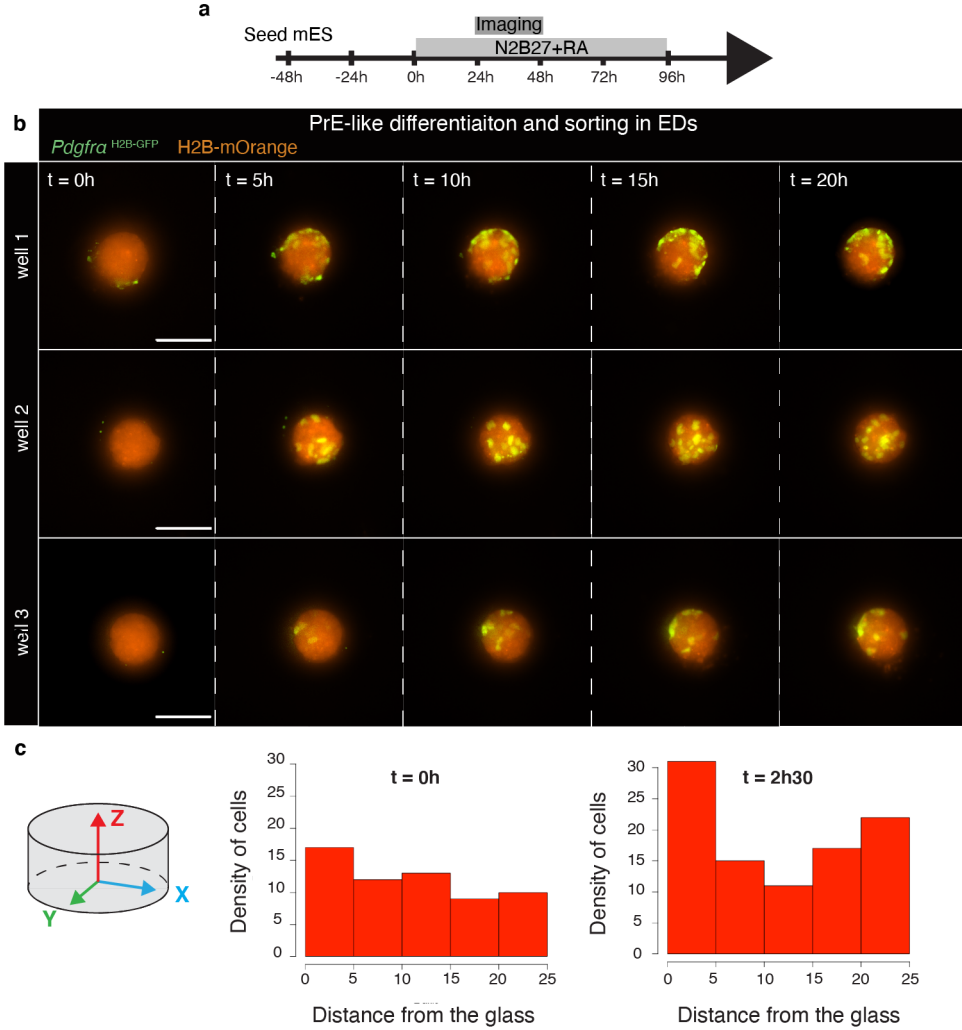


Figure 4.8: EDs allow better characterization of RA induced Epi/PrE-like sorting. **a**, Schematic of RA induced PrE-like differentiation and sorting in EDs. mESC were seeded in PDMS microwells and grown for 48 h. Once aggregates were formed in the microwells, the medium was supplemented with 4 μ M RA. 24 h after RA addition, EDs were imaged to follow the differentiation and migration of PrE-like cells (identified by *Pdgfra*^{H2B-GFP} expression). **b**, Live cell imaging of PrE-like cell differentiation and migration in EDs treated with RA. Scale bars: 100 μ m. **c**, Distribution of PrE-like cells along the z-axis in RA treated EDs at different time points.

While we were able to demonstrate cell sorting qualitatively in spherical aggregates, quantitative imaging and analysis turned out to be difficult due to Brownian motion of the free-floating aggregates and challenging optical properties of the agarose microwells. Furthermore, a sphere does not correspond well to the actual shape of the ICM *in vivo*, which has a half in direct contact with the basal membrane of the trophectoderm (Fig.

4.1). In order to improve imaging conditions and approximate the adhesion of the ICM to the trophectoderm, we developed 3D microstructures with cylinder-shaped growing on laminin-coated surface. To form these aggregates, cells were seeded in cylindrical microwells created by bonding a thin slab of PDMS with circular holes to a glass coverslip (Fig. 4.3). The glass was coated with laminin to enable cell attachment. This new system, which we called embryoid disk (ED), has the additional advantage that the extracellular matrix components to which the mESCs attach can be easily varied. Imaging of EDs is convenient since they are immobile and can be imaged through coverslips of standard thickness. After aggregation in the microwells, cells were exposed to RA to induce the formation of PrE-like cells (Fig. 4.8a). Time lapse imaging between 24 h and 48 h of RA exposure showed that PrE-like cells were appearing throughout the whole ED and ended up accumulating at the top and the bottom of the ED (Fig. 4.8b, c).

4

The optimized imaging conditions allowed us to extract the trace of PrE-like cells in the EDs. Interestingly, PrE-like cells seem to exhibit different behaviors according to their starting position along the z-axis of the well, i.e. the distance from the glass bottom (Fig. 4.9). Cells differentiating in the middle of the aggregate showed a clear movement in the z-direction. On the other hand, differentiating cells near the top or the bottom of the EDs did not move significantly in Z and seemed restricted to a limited XY plane.

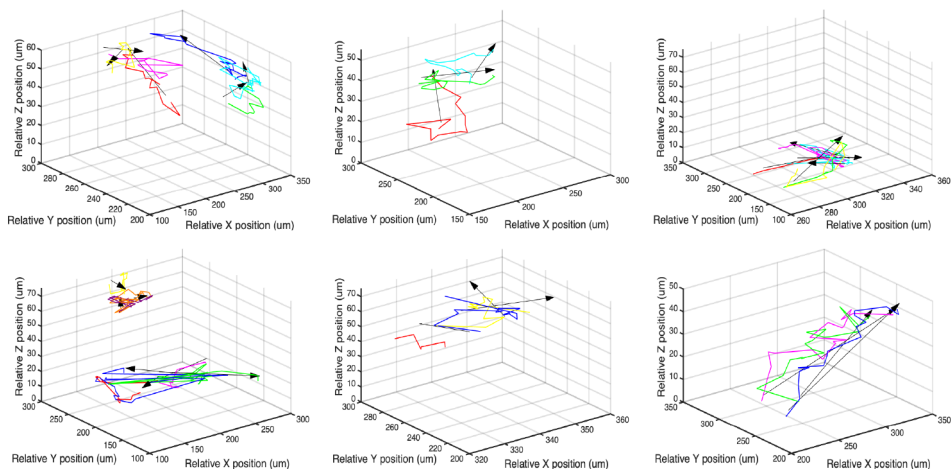


Figure 4.9: PrE-like cells exhibit different kinds of motion in EDs. Traces of PrE-like cells moving in RA treated EDs. Arrows show the global trend of cell movement, from starting point to end point.

To analyze the cell movement in greater detail we calculated the average instantaneous velocities as well as the mean-squared displacements (MSDs) for all traces (Fig. 4.10a). Due to the observed differences in cell movement depending on the z-position, we calculated separate velocities for movement in z and movement in the xy-plane. Since the cell traces were obtained in multiple experiments, the colonies in the wells

have slightly different sizes (from 50 to 75 μm along the z-axis). To perform combined analysis of cells at similar relative z-positions within an ED, we normalized the z-coordinates to the thickness of the colony, such that 0 corresponds to the glass surface and 1 to the top of the EDs.

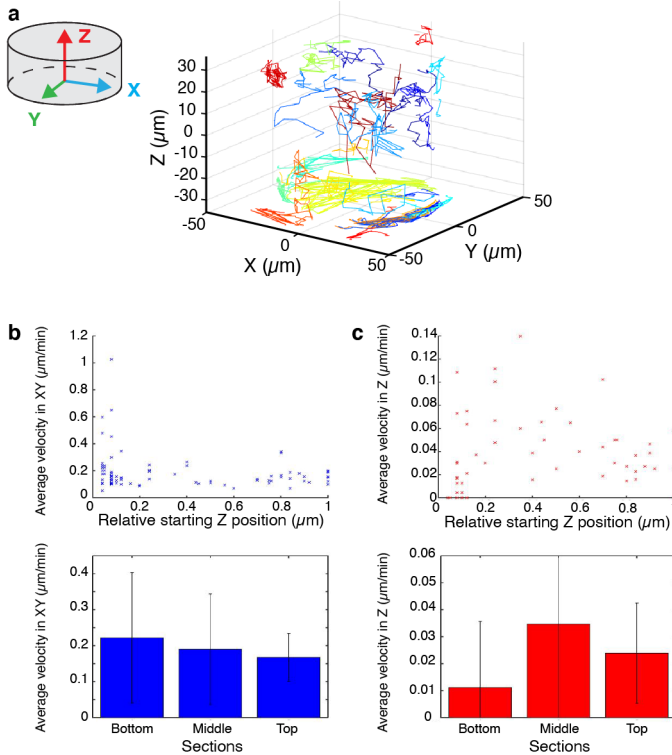


Figure 4.10: The behavior of PrE-like cells in EDs depends on their position along the Z-axis. **a**, Plot of all PrE-like cell movement traces. **b**, Top, average speed per trajectory in the xy direction according to their starting position along the z-axis. Bottom, speed of cells in xy, averaged over all cells in each indicated section of the ED. **c**, Top, average speed per trajectory in the z-direction according to their starting position along the z-axis. Bottom, speed of cells in z, averaged over all cells in each indicated section of the ED. The z coordinates were normalized so 0 corresponds to the glass surface and 1 to the top of the ED. The bottom section is defined between [0, 0.15], the middle section between [0.15, 0.75] and the top section between [0.75, 1].

The average speed per trajectory of PrE-like cells in the xy-plane (0.1 – 0.3 $\mu\text{m}/\text{min}$, Fig. 4.10b) was roughly independent of z-position, with the exception of the cells closest to the glass surface. There, some cells exhibited faster movement (up to 1.02 $\mu\text{m}/\text{min}$). The speed of cells in the z-direction was variable at the bottom of the ED (Fig. 4.10c) but most cells showed very little movement ($< 0.03 \mu\text{m}/\text{min}$). Cells differentiating in the middle section of the well ([0.15, 0.75] relative position) had higher speeds in the z-direction, than cells at the top or bottom. Finally, cells in the top section of the well hardly move in either z or the xy plane.

To further characterized cell dynamics, we calculated the MSD for the movement in xy or z, respectively (Fig. 4.11)). The averaged MSD of traces in different sections of the EDs, quantify the dependence of cell movement dynamics on z-position (Fig. 4.11)). Most of the movement in the z-direction is concentrated in the middle section of the ED. The top section of the ED showed very little movement in all directions. Finally, the cells in the bottom section showed a lot of movement in xy but almost none in the z-direction. Diffusion coefficients were estimated by fitting a linear function.

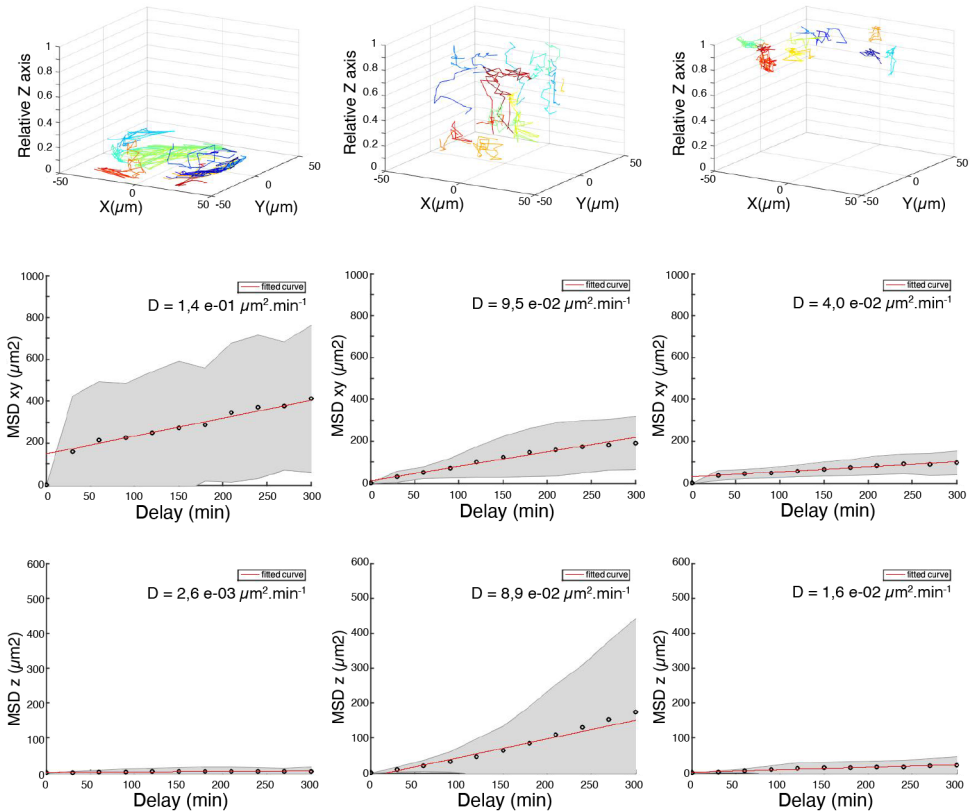


Figure 4.11: The behavior of PrE-like cells in EDs depends on their position along the Z-axis. Calculation of mean MSD of cells in the different sections of the colonies ([0.75, 1] at the top, [0.15, 0.75] in the middle, [0, 0.15] at the bottom). Mean MSD were calculated separately for xy plane (middle plots), and z-direction (bottom plots).

In several instances, the observed MSD curves were not fit well by a linear model. We therefore fit the same data again, this time with a power law. The inferred exponent α can give us an indication of the type of movement of the cells. An unbiased random walk corresponds to $\alpha = 1$, subdiffusion is characterized by $\alpha < 1$ and $\alpha > 1$ could indicate the presence of active, directed motion). For the movement in xy, all the exponents have a

value < 1 , probably due to confinement in the well (Fig. 4.12a). However, the fits of the MSDs for the z-direction (Fig. 4.12b), showed directed motion of the cells in the middle of the well ($\alpha > 1$), while the cells at the top and bottom showed subdiffusive behavior. This confirmed that the cells actively move along the z axis to reach the top of the ED (90% of cells moving in z are moving towards the top) and stop moving once they reach the surface. Cells differentiating at the bottom of the ED (below $50 \mu\text{m}$ above the glass) typically do not move to the top, possibly due to adhesion to the laminin coated surface, or low concentration of a morphogenic factor supplied by the media overlaying the ED. Cells differentiating at the top of the colony show very limited movement and mostly stay at the same place. Together, these behaviors explain the distribution of the PrE-like cells observed several hours after differentiation initiation (Fig. 4.8c).

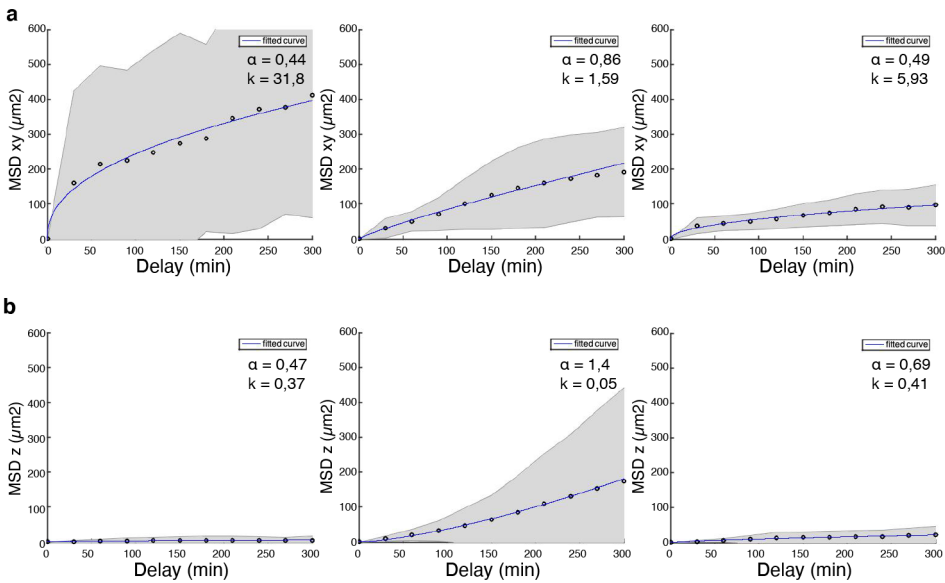


Figure 4.12: PrE-like cells have different type of movement along the z-axis. **a, b**, Analysis of the movement of PrE-like cells in xy (a) and z (b) direction. Mean MSD of the different sections were fitted with a power law ($y = kx^\alpha$). The value of α indicates the type of movement ($\alpha = 1$ random motion, $\alpha > 1$ directed motion, $\alpha < 1$ subdiffusion).

In conclusion, EDs allowed us to mimic the Epi/PrE sorting process *in vitro* and observe the movement of individual PrE-like cells. Several parameters of this system (coating, geometry of the well, size of the structure) can be easily tuned in the future to explore their influence on the dynamics of the cells. Such detailed measurements of cell movement could inform a mathematical model of cell sorting.

4.4. DISCUSSION

If a lot of studies about Epi/PrE sorting were so far performed on embryos *in vivo* or *in vitro*, mESCs-based *in vitro* models still represent an alternative way to study the process in a controlled, and easily tunable way. In this chapter we discussed two different models that we developed to study Epi/PrE sorting process *in vitro*.

So far, the default model used was the EB, made from genetically modified mESCs which allowed to observe qualitatively the sorting between cells of different adhesive properties [22] with sometimes differentiation of endoderm-like cells [18, 16]. Like for those previous models, our Chim-EBs cannot be used to study the dynamics of the cell's movement during the sorting, due to the concomitance of the aggregation process. However, they have the advantage to be an accessible system observe the effects of cells modification of both or one of two cell types on the sorting. Our protocol is short and allows to grow a large number of homogeneous aggregates at once, with a complete control of the state of the cells. Moreover, we are using well characterized PrE-like cells, and having a fluorescent PrE marker reporter (*Pdgfra*^{H2B-GFP}) allows a clear and easy read-out. Preliminary experiments with a siRNA knock-down screening assay revealed that knock-down of integrin $\beta 1$ in both cell types prevented sorting. A complete knock-down screening assay of all adhesion molecules in the two cell types separately would certainly give us more information about what adhesion molecules are required, and how different the adhesion between the two cell types need to be for the sorting to happen.

Among the different systems tested in this chapter to study Epi/PrE sorting, RA driven differentiation of EDs appears to be the best way to analyze the movement of the cells in a quantitative way. Using a *Pdgfra*-GFP reporter, each PrE-like cell differentiating in the system can be tracked. By adding another nucleus fluorescent reporter expressed in all cells, we could also study the dynamics of the other cells. While Chim-EBs would still be preferably used for large knock-down screening assay, this system can be used to look at the effect of a candidate gene on the sorting dynamics with a stable knock-out cell line.

Our preliminary experiments using PrE-like cells and WT mESCs, allowed us to already see that the movement of the PrE-like cells is directed. This already clear out the hypothesis that the cells are randomly moving within the IMC and forming the sorted pattern only with polarization establishment. The further movement characteristics extracted from the cell traces analysis could be used to develop an *in-silico* model. Such a model would help us to explore if differential adhesion is enough to explain the sorting or if it still requires an external signal to guide the PrE movement of the cells. Some questions are still remaining about the ED model, specially about the state the rest of the mESCs that do not differentiate into PrE-like. If from previous characterization mESCs are known to be close to Epi phenotype, the RA differentiation we are performing in EDs must also affect the cells that do not turn into PrE-like cells. Further characterization of those cells during and after the sorting must be performed to know if they still resemble Epi, before using our system to extrapolate about the sorting in the ICM.

ACRONYMS

Chim-EB chimeric embryoid body

E4.5 embryonic day 4.5

EB embryoid body

ED embryoid disk

ESC embryonic stem cell

mESC mouse embryonic stem cell

MSD mean square displacement

PDMS polydimethylsiloxane

RA retinoic acid

siRNA small interfering RNA

CELL TYPES

Epi epiblast

ICM inner cell mass

PrE primitive endoderm

4.5. REFERENCES

- [1] F. Fagotto. “The cellular basis of tissue separation.” In: *Development* 141.17 (Sept. 2014), pp. 3303–3318.
- [2] F. Fagotto. “Regulation of cell adhesion and cell sorting at embryonic boundaries.” In: *Current topics in developmental biology* 112 (2015), pp. 19–64.
- [3] J. D. Amack and M. L. Manning. “Knowing the Boundaries: Extending the Differential Adhesion Hypothesis in Embryonic Cell Sorting”. In: *Science* ().
- [4] D.-H. Yang et al. “Disabled-2 Is Essential for Endodermal Cell Positioning and Structure Formation during Mouse Embryogenesis”. In: *Developmental biology* 251.1 (Nov. 2002), pp. 27–44.
- [5] L. E. Stephens et al. “Deletion of beta 1 integrins in mice results in inner cell mass failure and peri-implantation lethality.” In: *Genes & Development* 9.15 (Aug. 1995), pp. 1883–1895.
- [6] S. Nowotschin et al. “The emergent landscape of the mouse gut endoderm at single-cell resolution.” In: *Nature* 569.7756 (Apr. 2019), pp. 361–367.
- [7] G. S. Kwon, M. Viotti, and A.-K. Hadjantonakis. “The endoderm of the mouse embryo arises by dynamic widespread intercalation of embryonic and extraembryonic lineages.” In: *Developmental Cell* 15.4 (Oct. 2008), pp. 509–520.
- [8] M. Viotti, S. Nowotschin, and A.-K. Hadjantonakis. “SOX17 links gut endoderm morphogenesis and germ layer segregation.” In: *Nature Cell Biology* 16.12 (Dec. 2014), pp. 1146–1156.
- [9] C. Chazaud et al. “Early Lineage Segregation between Epiblast and Primitive Endoderm in Mouse Blastocysts through the Grb2-MAPK Pathway”. In: *Developmental Cell* 10.5 (May 2006), pp. 615–624.
- [10] J. Rossant, C. Chazaud, and Y. Yamanaka. “Lineage allocation and asymmetries in the early mouse embryo.” In: *Phil. Trans. R. Soc. B* 358.1436 (Aug. 2003), pp. 1341–1349.
- [11] C. Chazaud and Y. Yamanaka. “Lineage specification in the mouse preimplantation embryo.” In: *Development* 143.7 (Apr. 2016), pp. 1063–1074.
- [12] S. Hermitte and C. Chazaud. “Primitive endoderm differentiation: from specification to epithelium formation.” In: *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 369.1657 (Dec. 2014), pp. 20130537–20130537.
- [13] N. Saiz et al. “Atypical protein kinase C couples cell sorting with primitive endoderm maturation in the mouse blastocyst.” In: *Development* 140.21 (Nov. 2013), pp. 4311–4322.
- [14] F. Gerbe et al. “Dynamic expression of Lrp2 pathway members reveals progressive epithelial differentiation of primitive endoderm in mouse blastocyst.” In: *Developmental biology* 313.2 (Jan. 2008), pp. 594–602.

- [15] N. Saiz et al. “Asynchronous fate decisions by single cells collectively ensure consistent lineage composition in the mouse blastocyst.” In: *Nature Communications* 7.1 (Nov. 2016), pp. 13463–14.
- [16] R. Moore et al. “Cell adhesion and sorting in embryoid bodies derived from N- or E-cadherin deficient murine embryonic stem cells”. In: *bio.biologists.org* 3.2 (Jan. 2014).
- [17] S. M Meilhac et al. “Active cell movements coupled to positional induction are involved in lineage segregation in the mouse blastocyst.” In: *Developmental biology* 331.2 (July 2009), pp. 210–221.
- [18] M. E. Rula et al. “Cell autonomous sorting and surface positioning in the formation of primitive endoderm in embryoid bodies”. In: *genesis* 45.6 (June 2007), pp. 327–338.
- [19] M. S. Steinberg. “Differential adhesion in morphogenesis: a modern view.” In: *Current Opinion in Genetics & Development* 17.4 (Aug. 2007), pp. 281–286.
- [20] M. S. Steinberg. “Reconstruction of tissues by dissociated cells. Some morphogenetic tissue movements and the sorting out of embryonic cells may have a common explanation.” In: *Science* ().
- [21] M. S. Steinberg. “Does differential adhesion govern self-assembly processes in histogenesis? Equilibrium configurations and the emergence of a hierarchy among populations of embryonic cells.” In: *The Journal of experimental zoology* 173.4 (Apr. 1970), pp. 395–433.
- [22] R. Moore et al. “The Primitive Endoderm Segregates from the Epiblast in 1 Integrin-Deficient Early Mouse Embryos”. In: *Molecular and cellular biology* 34.3 (Jan. 2014), pp. 560–572.
- [23] Néstor Filimonow K. and Saiz et al. “No evidence of involvement of E-cadherin in cell fate specification or the segregation of Epi and PrE in mouse blastocysts.” In: *PLoS ONE* 14.2 (2019).
- [24] J. Artus, J.-J. Panthier, and A.-K. Hadjantonakis. “A role for PDGF signaling in expansion of the extra-embryonic endoderm lineage of the mouse blastocyst.” In: *Development* 137.20 (Oct. 2010), pp. 3361–3372.
- [25] S. Semrau et al. “Dynamics of lineage commitment revealed by single-cell transcriptomics of differentiating embryonic stem cells.” In: *Nature Communications* 8.1 (Oct. 2017), pp. 1096–16.
- [26] J. Schindelin et al. “Fiji: an open-source platform for biological-image analysis.” In: *Nature Methods* 9.7 (June 2012), pp. 676–682.
- [27] T. R. Jones et al. “CellProfiler Analyst: data exploration and analysis software for complex image-based screens.” In: *BMC bioinformatics* 9.1 (Nov. 2008), pp. 482–16.
- [28] J. W. Young et al. “Measuring single-cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy.” In: *Nature protocols* 7.1 (Dec. 2011), pp. 80–88.

- [29] N. Tarantino et al. "TNF and IL-1 exhibit distinct ubiquitin requirements for inducing NEMO-IKK supramolecular structures." In: *The Journal of Cell Biology* 204.2 (Jan. 2014), pp. 231–245.
- [30] T. Boroviak et al. "The ability of inner-cell-mass cells to self-renew as embryonic stem cells is acquired following epiblast specification." In: *Nature Cell Biology* 16.6 (June 2014), pp. 516–528.
- [31] K. Brown et al. "A comparative analysis of extra-embryonic endoderm cell lines." In: *PLoS ONE* 5.8 (Aug. 2010).
- [32] J. Liu et al. "Integrins are required for the differentiation of visceral endoderm." In: *Journal of Cell Science* 122 (Jan. 2009), pp. 233–242.
- [33] V. C. Ridger et al. "Differential effects of CD18, CD29, and CD49 integrin subunit inhibition on neutrophil migration in pulmonary inflammation." In: *Journal of immunology (Baltimore, Md. : 1950)* 166.5 (Mar. 2001), pp. 3484–3490.
- [34] H. Chen et al. "In vivo beta1 integrin function requires phosphorylation-independent regulation by cytoplasmic tyrosines." In: *Genes & Development* 20.8 (Apr. 2006), pp. 927–932.

5

SINGLE-CELL TRANSCRIPTOMICS REVEALS GENE EXPRESSION DYNAMICS OF HUMAN FETAL KIDNEY DEVELOPMENT

This chapter is based on:

Mazène Hochane et al. “Single-cell transcriptomics reveals gene expression dynamics of human fetal kidney development”. In: *PLOS Biology* 17.2 (Feb. 2019), e3000152

In vitro models are a convenient way to study mechanisms of development, however, the results of those studies have to be validated by *in vivo* observations. Single-cell transcriptomics has emerged as one of the most useful tools to reveal complex developmental dynamics in a comprehensive way. Single-cell expression atlases can also serve as benchmarks for the development of more realistic *in vitro* systems, in particular organoids.

As an example of such a benchmark, this chapter presents a single-cell transcriptomics study of the human fetal kidney. Several fetal kidney samples of different ages (w9, w11, w13, w16, w18) were sequenced to characterize cell type diversity. The analysis identified 22 cell types and their marker genes. Among those cell types are newly discovered podocyte precursors, and several nephron progenitor cell (NPCs) subtypes that populate the nephrogenic niche. These cell types were localized in fetal kidney sections, and the extensive information contained in their transcriptomics allowed us to explore disease-associated genes expressed in those transient cell types. This study provides a rich resource for the understanding of human kidney development, easily accessible through an interactive web application.

5.1. INTRODUCTION

Mammalian kidney development initiates in the intermediate mesoderm through crosstalk between the metanephric mesenchyme (MM) and the ureteric bud (UB). The UB originates from the nephric duct, invades the MM, and starts to subdivide progressively into multiple ramifications. The UB tip cells, which make the first contact with the MM, become enveloped by an assembly of mesenchymal cells, the cap mesenchyme (CM) (Fig. 5.1A-B). The CM contains NPCs, which give rise to the whole nephron epithelium through tightly regulated morphogenic transformations [2]. Self-renewal of (mouse) NPCs is governed by key transcription factors, such as *Six2* and *Meox1*, which mark the nephrogenic zone of the kidney [3]. Signaling between UB tip cells and NPCs regulates the balance between self-renewal and differentiation of the NPCs [4]. In humans, about 1 million nephrons are produced before the NPC population is irrevocably exhausted a few weeks before birth [5]. During nephrogenesis, NPCs undergo mesenchymal-epithelial transition and differentiate into a succession of intermediate structures: the pretubular aggregate (PTA), renal vesicle (RV), comma-shaped body (CSB) and s-shaped body (SSB). Then, via the capillary loop stage, mature and functional glomerular and tubular structures are eventually formed. In contrast to the nephron epithelium, the collecting duct system originates from the UB. GDNF/RET signaling between the UB and CM critically regulates proliferation of UB tip cells and branching morphogenesis of the UB [6]. Stromal cells—such as interstitial cells (ICs), mesangial cells, juxtaglomerular cells, smooth muscle cells, fibroblasts, and pericytes—derive from a common interstitial progenitor [7, 8]. Finally, vascular endothelial cells and the highly specified glomerular endothelium originate from the MM [9], and leukocytes and erythrocytes enter with the blood stream. The current understanding of mammalian kidney development is largely based on mouse studies, although it is clear that human and mouse kidneys are morphologically different. Three recent landmark studies have revealed, in great detail, a significant divergence between mouse and human renal embryogenesis in terms of morphology as well as gene expression [10, 11, 12]. These studies underline that the prevailing lack of data on human kidney development severely hinders the detailed understanding of human kidney development and possible developmental origins of kidney disease.

In the study described here, we used scRNA-seq to study gene expression dynamics in human fetal kidney development. Analysis of a fetal kidney from week 16 (w16) of gestation revealed 22 cell types, which we identified by known marker genes. Pseudotime analysis clarified their temporal relationship. We further defined specifically expressed cell type marker genes of which many have not been implied in kidney development. Comparison to four additional samples (from w9, w11, w13, and w18) suggested that most cell types have a constant expression pattern, with the notable exception of podocytes. To highlight two ways in which our dataset can be interrogated, we then explored the nephrogenic niche and the development of podocytes. Gene expression differences between four NPC clusters were related to spatial heterogeneity by immunostaining and smFISH. Expression of the disease-associated gene *UNCX* was localized to NPCs and their early derivatives. Finally, we focused on podocyte development, which proceeds via a distinct precursor state. By immunostaining and smFISH, we localized

these precursors in situ and confirmed the disease-associated gene *OLFM3* as a marker.

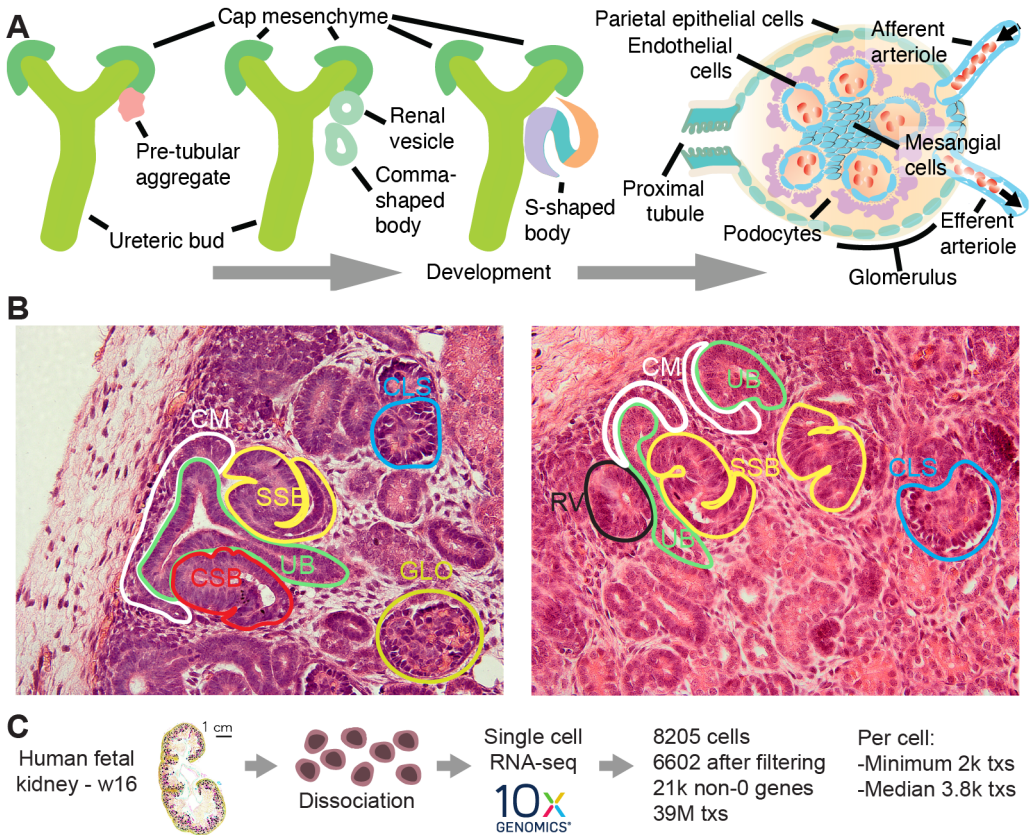


Figure 5.1: Overview of kidney development and the experimental design. (A) Schematic of kidney epithelium development. (B) Several morphologically distinct stages of nephrogenesis are highlighted by colored lines in images of human fetal kidney sections stained with hematoxylin-eosin. (C) Overview of the scRNA-seq experiment. Ureteric bud (UB), cap mesenchyme (CM), pretubular aggregate (PTA), renal vesicle (RV) (Stage I), comma-shaped body (CSB) (Stage II), s-shaped body (SSB) (Stage II), capillary-loop stage (CLS) (Stage III), glomerulus (GLO) (Stage IV)

My role was to characterize the S-shaped body podocytes (SSBpod) population, identified in the scRNA-seq clustering. Those cells are a novel, transient state in the development of podocytes (pods), simultaneously discovered in this study and one from Lindstrom et al.[13]. This new cell state was characterized by investigating its expression pattern, according to known literature markers and novel markers identified in our scRNA-seq data. By cross-referencing the obtained information with genome-wide association studies, we found that the SSBpod marker *OLFM3* was associated with glomerular filtration rate. The SSBpods were located in developing kidneys using immunostain-

ing and smFISH. Finally, the exploration of signaling factors differentially expressed in SSBpods led to nominate potential targets for perturbation studies.

5.2. MATERIALS AND METHODS

5.2.1. ETHICS STATEMENT

The collection and use of human material in this study was approved by the Medical Ethics Committee from the Leiden University Medical Center (P08.087). The gestational age was determined by ultrasonography, and the tissue was obtained by vacuum aspiration from women undergoing elective abortion. The material from six embryos (w9, male; w11, male; w13, female; w15, female; w16, male; and w18, female) was donated with written informed consent. Questions about the human material should be directed to S. M. Chuva de Sousa Lopes (Lopes@lumc.nl).

5.2.2. EXPERIMENTAL METHODS

SINGLE-CELL DISSOCIATION OF HUMAN FETAL KIDNEY

One human embryo of w16 (male) was isolated and the kidney dissected in cold saline solution (0.9% NaCl, Versylene Fresenius). For sex genotyping, polymerase chain reaction (PCR) for AMELX/Y was used as previously described [14]. The obtained kidney was decapsulated and kept on ice in dissociation buffer (DPBS + Penicillin 100 U/mL + Streptomycin 0.1 mg/mL; all from Life Technologies) before cutting it into 1-2 mm pieces. The pieces were washed three times with washing solution (Advanced DMEM F12 supplemented with ITS commercial solution (Insulin-Transferrin-Selenium; ThermoFisher), Glutamax, Penicillin 100 U/mL, and Streptomycin 0.1 mg/mL) with brief centrifugation (160 g) in order to remove as many red blood cells as possible. The washed kidney tissue was then incubated with digestion solution (Trypsin/EDTA solution 0.25% and Collagenase-II 280 U/mL) and incubated overnight at 4°C. The next day, the digestion solution was removed, and the kidney was rinsed with washing solution and incubated with washing solution for 30min at 37°C with agitation. Subsequently, the sample was sequentially passed through sterile cell strainers of 100, 70, and 40 μm pore size with the help of washing solution. The cells were then centrifuged and counted, and viability was measured to be 78% (trypan blue assay) before proceeding with scRNA-seq library preparation. Four additional human fetal kidneys (w9, male; w11, male; w13, female; and w18, female) were dissected as described above, but, additionally, live cells were purified by FACS before library preparation [15].

scRNA-SEQ LIBRARY PREPARATION AND SEQUENCING

scRNA-seq libraries were prepared using the Chromium Single Cell 3' Reagent Kit, Version 2 Chemistry (10X Genomics) according to the manufacturer's protocol. Libraries

were sequenced on a NextSeq500 in Mid Output mode using a version 2, 150-cycle kit (Illumina).

IMMUNOSTAINING

A paraffin-embedded w15 human kidney (female) was sectioned (5 μ m) using a RM2255 microtome (Leica Microsystems GmbH) and mounted on StarFrost slides (Waldemar Knittel).

For immunofluorescence, sections were deparaffinized and rehydrated by standard procedures, starting with xylene (twice for 20 min), followed by ethanol with sequential dilution and ending with distilled water, all at RT. Antigen retrieval was performed by a double treatment of 10 min in a microwave (97°C) with 0.01M sodium citrate buffer (pH 6.0). The sample was then allowed to cool down, rinsed three times with PBS, and blocked for 1h at RT in blocking buffer (1% BSA, 0.05% Tween-20 in PBS). Subsequently, sections were incubated overnight with the following antibodies diluted in blocking buffer—rabbit anti-SIX2 (1:100, 11562-1-AP; Proteintech), mouse anti-CITED1 (1:500, H00004435-M03; Novus Biologicals), mouse anti-MAFB (1:200, LS-C336952; LifeSpan Biosciences), rabbit anti-ACTA2 (1:200, ab5694; Abcam), goat anti-PODXL (1:200, AF1658; R&D Systems), and rabbit anti-UNCX (1:10, PA5-69485; Thermo Fisher Scientific), rabbit anti-CKS2 (HPA003424, 1:100; Sigma Aldrich), rabbit anti-NURR77 (NR4A1) (ab13851, 1:50; Abcam Biochemicals), and mouse anti-HSP70 (HSPA1A) (ab2787, 1:50; Abcam Biochemicals). The secondary antibodies were diluted in blocking buffer and applied at RT for 1h followed by nuclear counterstaining with 4,6-diamidino-2-phenylindole (DAPI; Life Technologies). The secondary antibodies used were Alexa Fluor 647 donkey anti-rabbit (1:500, A-31573; Life Technologies), Alexa Fluor 594 donkey anti-mouse (1:500, A-21203; Life Technologies), and Alexa Fluor 555 donkey anti-goat (1:500, A32727; Life Technologies). The sections were then mounted using ProLong Gold (Life Technologies).

For immunohistochemistry, sections were deparaffinized and blocked as above. After overnight incubation with primary antibodies rabbit anti-UNCX (1:10, PA5-69485; Thermo Fisher Scientific) and mouse anti-CITED1 (1:500, H00004435-M03; Novus Biologicals) in blocking buffer, 0.3% H₂O₂ was used to quench endogenous peroxidase activity for 20min. Next, the sections were incubated with biotin-labeled goat anti-rabbit IgG (1:200, BA-1000; Vector Laboratories) diluted in normal goat serum (1:66, S-1000; Vector Laboratories) or biotin-labeled horse anti-mouse (1:200, BA-2000; Vector Laboratories) diluted in normal horse serum (1:66, S-2000; Vector Laboratories) for 40min. Sections were then treated for 40min with avidin-biotin-peroxidase complex (VECTASTAIN Elite ABC HRP Kit, #PK-6100; Vector Laboratories) following the manufacturer's instructions, followed by DAB (D5637; Sigma-Aldrich) and hematoxylin (1043020025; Merck) and were mounted with Entellan (1079610100; Merck).

SINGLE-MOLECULE FISH

Paraffin embedded sections from the w15 human fetal kidney (female) used for immunostaining were also used for smFISH experiments. Paraffin was removed by immersion in xylene twice for 10 min at RT. The sections were then rehydrated by sequential immersion in ethanol solutions (100% (2x, 10min), 85% (2x, 5 min), and 70% (2x, 3 min)).

Subsequently, sections were permeabilized in 70% ethanol for 5h before incubation with proteinase-K (P4850; Sigma Aldrich) for 15 min at 37°C (23 $\mu\text{g}/\text{mL}$ in TE buffer at pH = 8) and a wash in RNase-free water (3x, 5 min). smFISH was performed as described previously [16]. Briefly, custom designed smFISH probes (BioCat, Hochane et al. [1] S5 Table), labeled with Quasar 570, CAL FLuor Red 610, or Quasar 670, were incubated with the samples for 16h at 30°C in hybridization buffer (100 mg/mL dextran sulfate, 25% formamide, 2X SSC, 1 mg/mL E.coli tRNA, 1 mM vanadyl ribonucleoside complex, 0.25 mg/mL BSA). Samples were washed twice for 30 min at 30°C with wash buffer (25% formamide, 2x SSC) containing DAPI (1 $\mu\text{g}/\text{mL}$, D9542; Sigma). All solutions were prepared with RNase-free water. Finally, the sections were mounted using ProlongGold (P36930; Life Technologies) and imaged the next day.

IMAGING

Immunostained and smFISH-treated kidney sections were imaged on a Nikon Ti-Eclipse epifluorescence microscope equipped with an Andor iXON Ultra 888 EMCCD camera, using a 100x /1.45 Plan Apo Lambda oil objective (Nikon) and dedicated, custom-made fluorescence filter sets (Nikon). To cover large areas of the sectioned kidney, images of multiple adjacent areas were taken and combined using the tiling feature of the NIS Elements software (Nikon). For imaging of smFISH signals, z-stacks were collected with distances of 0.3-0.5 μm between planes in four fluorescence channels (GFP, Quasar 570, CAL FLuor Red 610, Quasar 670).

5.2.3. QUANTIFICATION AND STATISTICAL ANALYSIS

SCRNA-SEQ DATA PRUNING AND NORMALIZATION

Single-cell expression for the w16 sample was quantified using unique molecular identifiers (UMIs) by 10X Genomics' *Cell Ranger* software. After removing cells with less than 2,000 transcripts per cell, 8,503 cells were retained for further analysis. On average, 1,789 genes were detected per cell and a median of 4,805 transcripts per cell (Fig. 5.2A). Given the recent report that dissociation can have a significant influence on the single-cell transcriptome [17] and that the kidney is notoriously difficult to dissociate, special attention was paid to dissociation-related artifacts. The amount of 1,859 cells with signs of stress were removed from the dataset (Fig. 5.2B). These cells had more than 10% of their expression come from mitochondrial genes (*MT-ND1*, *MT-ND2*, *MT-CO1*, *MT-CO2*, *MT-ATP8*, *MT-ATP6*, *MT-CO3*, *MT-ND3*, *MT-ND4L*, *MT-ND4*, *MT-ND5*, *MT-ND6*, *MT-CYB*) or more than 5% from stress markers. Stress markers were defined as those genes that were significantly up-regulated upon prolonged enzymatic incubation of mouse kidney tissue in the study by Adam, Potter, and Potter [18] (Hochane et al. [1] S2 Table, Fig. 5.2B). Mouse genes from this list were converted to human genes using biomart [19]. Genes of the literature set (Table 5.1) only showed small differences between stressed and nonstressed cells (Fig. 5.2C-D), and stressed cells did not form a separate cluster (Fig. 5.2E-F). Therefore, removing stressed cells did not reduce the cell type diversity in the sample. Additionally, 42 cells had more than 1% of their expression coming from *HBB*, *HBA1*, and *HBA2* and were therefore classified as red blood cells

and discarded from any further analysis (Fig. 5.2E). Sporadic expression of hemoglobin genes in other cells was likely due to red blood cells that burst before isolation. The same filtering approach was applied to the samples from the other developmental ages as well as the data from Lindström et al. [13]. Raw UMI counts were smoothed by k-nearest neighbors smoothing version 2.1 [20]. This procedure reduces technical noise by sharing information between transcriptionally similar cells, which likely belong to the same cell type. Briefly, the expression profiles of each cell and its knn were summed ($k = 10$; distance metric: Euclidean distance of the first 10 principal components (PCs) with a dither of 0.05). The resulting smoothed count matrix had a higher total count than the original and was therefore scaled back to the original matrix by a global factor. Expression was normalized by the method developed by Lun and colleagues [21] (as implemented in the *scran* (version 1.10.1) R package using the functions *quickCluster* and *computeSumFactors*). Normalized gene expression was Freeman-Tukey (FT) transformed in further analyses unless stated otherwise.

REDUCTION OF DIMENSIONALITY

Variability of gene expression was calculated using the *improvedCV2*-function from the *scran* R package. Intercell distances were calculated using the 5% most highly variable genes (HVGs) excluding stress markers [18] (Hochane et al. [1] S2 Table) and ribosomal genes (obtained from the HGNC website) without any filter for minimum mean expression. For maps of individual samples, we used (1-Pearson correlation) as distance measure. For maps of combined samples, we used Euclidean distance in the MNN-corrected PC space. All t-distributed stochastic neighbor embedding (tSNE) maps used a perplexity setting of 500. For the DDRTree embedding used with pseudotime analysis, see Pseudotime analysis sections.

CLUSTERING

Hierarchical cluster analysis was performed using Ward linkage and the same intercell distances as for the reduction of dimensionality. The dendrogram of this clustering was cut at height 0.6 to yield 29 clusters of cells (Fig. 5.4). The cut off was chosen such that the number of resulting clusters was comparable to the number of cell types expected from the literature on mouse development [22] and other scRNA-seq studies of the human fetal kidney [11, 13, 23, 24, 25]. We estimated the number of cell types to be around 20 but created slightly more as a starting point to allow for the discovery of new cell types. On the other hand, we did not want to use a much higher number to avoid overclustering (i.e., creating many clusters that are merely driven by noise, which would then have to be merged manually). The presence of known markers of the different cell types in the kidney (literature set, Table 5.1) was then used to identify cell types (Fig. 5.4). Based on this analysis, some adjacent clusters (clusters 4 and 5, 11 and 12, 15 and 16, and 17 and 18) showed very similar expression of known marker genes of podocytes, ICs, UB, and collecting duct and proximal tubule cells, respectively. In addition, the aforementioned clusters were in close proximity, both in the clustering dendrogram as well as in tSNE space (Fig. 5.4). Consequently, these clusters were merged. For example, clusters 4 and 5 had similar expression of genes known to be expressed in mature podocytes (*NPHS2*,

PTPRO, *PODXL*) compared to cluster 6, which showed very weak expression of these genes and had distinctive expression of *OLFM3*, which has been shown to be specifically expressed in podocytes precursors [13, 23]. Furthermore, we also merged clusters 7 and 25, which were more distant in the dendrogram of the hierarchical clustering but had very similar literature marker profiles (e.g., *LHX1*, *WNT4*, *CCND1*, *JAG1*, *PAX2*, and *PAX8*) and appeared in close proximity in tSNE space. Finally, clusters 26 and 29 were also merged. Cluster 26 was a heterogeneous cluster of only 56 cells that were spread in tSNE space between multiple other clusters. This cluster was closest to PTA (cluster 29) in terms of literature marker expression (*WNT4*, *LHX1*, and *CCND1*) and differed from it with respect to proliferative state, which may account for the heterogeneous distribution.

COMBINING DIFFERENT DATASETS

To compare cells from multiple scRNA-seq datasets, we used the fastMNN function [26] implemented in scran (version 1.10.1) on the first 50 PCs of the 5% HVGs without stress markers or ribosomal genes. We used a knn approach to infer the cell types of unclassified cells from already classified cells. For each unclassified cell, the 20 nearest neighbors in batch-corrected PC space (Euclidean distance) were determined. The most common cell type among these neighbors was then assigned to the unclassified cell. For the comparison with the dataset from Lindström et al. [13], we restricted our dataset to the nephrogenic niche. The cluster identities for the Lindström et al. [13] dataset were kindly provided to us by the group of Andrew D. Smith.

CELL CYCLE AND PROLIFERATION

Cell cycle scores were calculated using the Cyclone tool [27] from the scran (version 1.10.1) R package. A list of proliferation markers was adopted from a publication by Whitfield et al. [28].

PSEUDOTIME ANALYSIS

We used the *Monocle 2* algorithm [29] to perform embedding and pseudotime analyses on the 2,594 cells of the nephron epithelium, starting from the PTA (cells classified as PTA, renal vesicle/comma-shaped body (RVCSB)a, RVCSBb, SSB medial/distal (SSBm/d), SSB proximal precursor cell (SSBpr), SSBpod, distal tubule/loop of Henle (DTLH), early proximal tubule (ErPrT), or podocytes), and separately on the 2,153 cells of the nephrogenic niche (NPC) and the PTA. The 5% HVGs (without stress or ribosomal genes) were used as input to the algorithm. We used the *reduceDimension* function (*max_components* = 3 for Fig. 5.9; *max_components* = 2 for Fig. 5.16A) to run the *DDRTree* algorithm [30]. The root of the graph learned by *DDRTree* was placed on the branch that starts with the PTA to obtain the pseudotime shown in Fig. 5.9B.

MARKER GENES AND KEYGENES

For each gene, the cluster of interest (COI) was defined as the cluster that had the highest mean expression of the gene. Then, a binary classifier based on an expression threshold was defined: cells with expression above that threshold were considered to be part of the COI. We systematically varied this threshold to create a receiver operating characteristic (ROC) based on the cells' true cluster identities. The area under the ROC (AUROC) was then used to determine the usefulness of this gene as a marker (rather than the specificity or sensitivity at a specific threshold). Genes that had an AUROC exceeding 0.8 were detected in at least 80% of the cells in the COI, had a minimum mean expression of 1.5 in the COI, and those for which maximally 25% of the cells outside the COI had significant expression, were defined as marker set candidates (Hochane et al. [1] S3 Table). Significant expression was defined here as an expression level higher than the 25th percentile of expression in the COI. Subsequently, the top four candidate marker genes per cluster, as ranked by the AUROC, resulted in a final set of 88 marker genes (marker set, Hochane et al. [1] S3 Table).

To apply the KeyGenes prediction algorithm [31], two-thirds of the cells were assigned to the training set and one-third to the test set. A multinomial logistic regression model was trained on the training set with LASSO shrinkage using the 500 most HVGs, filtered for stress markers and ribosomal genes. The shrinkage parameter was determined by 20-fold cross validation. To apply the KeyGenes method to single cells, each cell was treated as a sample, and cross validation was used to control for overfitting. The model obtained a list of 95 classifier genes with nonzero weights (KeyGenes set, Hochane et al. [1] S3 Table). Thereafter, the cells in the test set were assigned to the cell type with the highest identity (id) score; 84% of the cells in the test set were classified correctly (16% test error). On average, the id score was 0.59, and 24% of the cells in the test set obtained an id score higher than 0.8.

DIFFERENTIAL EXPRESSION ANALYSIS

For all differential expression analyses, we used EdgeR (version 3.24.0) [32] on raw counts. Normalization and dispersion estimates were calculated by `calcNormFactors` and `estimateDisp`, respectively. We modeled gene expression with a negative binomial generalized linear model with `glmQLFit`. Besides the conditions to be compared, a detection rate for each gene was added to the design matrix. The detection rate is defined as the fraction of cells with nonzero expression. In the comparison of different ages, we excluded the w9 and w16 samples. The w9 sample contained only a few cells, which results in high uncertainty for average gene expression levels. The w16 sample was created separately from the other samples. Therefore, to avoid batch effects, which are not corrected for in the differential expression analysis, we therefore also excluded the w16 sample.

GENOME-WIDE ASSOCIATION STUDIES (GWAS) ANALYSIS

The NHGRI-EBI GWAS catalog was used to retrieve genes associated with traits related to kidney diseases. Specifically, we selected the following kidney traits: *kidney stone*, *kidney disease*, *rapid kidney function decline*, *chronic kidney disease*, *kidney amyloid deposition measurement*, *acute kidney injury*, *type 1 diabetes nephropathy*, *nephrolithiasis*, *diabetic*

nephropathy, proteinuria, GFR change measurement, renal cell carcinoma, serum creatinine measurement, cystatin C measurement, type 2 diabetes nephropathy, immunosuppressive agent, tacrolimus measurement, focal segmental glomerulosclerosis, nephrotic syndrome, membranous glomerulonephritis, lupus nephritis, IgA glomerulonephritis, renal system measurement, and Wegener's granulomatosis. This selection resulted in a list of 560 genes (Hochane et al. [1] S4 Table, Kidney GWAS genes). As a negative control, we also obtained a list of 1,508 genes associated with traits related to lung diseases (Hochane et al. [1] S4 Table, Lung GWAS genes) in which we selected the following traits: lung adenocarcinoma, lung carcinoma, interstitial lung disease, squamous cell lung carcinoma, lung disease severity measurement, family history of lung cancer, non-small cell lung carcinoma, diffusing capacity of the lung for carbon monoxide, pulmonary function measurement, vital capacity, emphysema, idiopathic pulmonary fibrosis, chronic bronchitis, chronic obstructive pulmonary disease, pneumonia, and asthma. We performed a one-sided Fisher's exact test to determine whether the genes in the GWAS lists were significantly enriched in the genes that were differentially expressed in our clusters of interest.

5

MULTIPLE HYPOTHESIS TESTING

In all cases in which significance is reported, p-values were adjusted for multiple hypothesis testing using the Benjamini-Hochberg method.

FUNCTIONAL ANNOTATION ENRICHMENT

To look for enrichment of gene ontology (GO) terms or transcription factor binding sites we use the DAVID Functional Annotation tool [33], version 6.8 <https://david.ncifcrf.gov/> with all genes in the human genome as background gene set. For enrichment of transcription factor binding sites, we used the *UCSC_TFBS* category.

IMAGE ANALYSIS

smFISH image stacks were processed with 3D deconvolution and background correction (rolling ball, radius: 3 pixels = 0.39 μ m), using the NIS Elements software. Subsequently, maximum projection was used to create a 2D representation of the image stack. The resulting smFISH images were analyzed with homemade MATLAB scripts. First, autofluorescent background was removed by subtracting the appropriately scaled signal of the GFP channel from each of the other channels. Then a region of interest (ROI) containing the structure of interest was defined manually, and candidate smFISH signals were detected by binarizing the image using a global threshold. Connected components were then counted as smFISH signals if they fulfilled two criteria: their average intensity was bigger than the third quartile of individual pixel intensities and they had an area of three pixels or bigger. The density of smFISH signals in the ROI was calculated as the number of retained connected components divided by the area of the ROI.

Images of immunostaining were pre-processed by background subtraction (rolling ball, radius: 100 pixels = 13 μ m) using ImageJ [34]. Quantification of the immunostaining signal was carried out using homemade MATLAB scripts. First, the CM region was

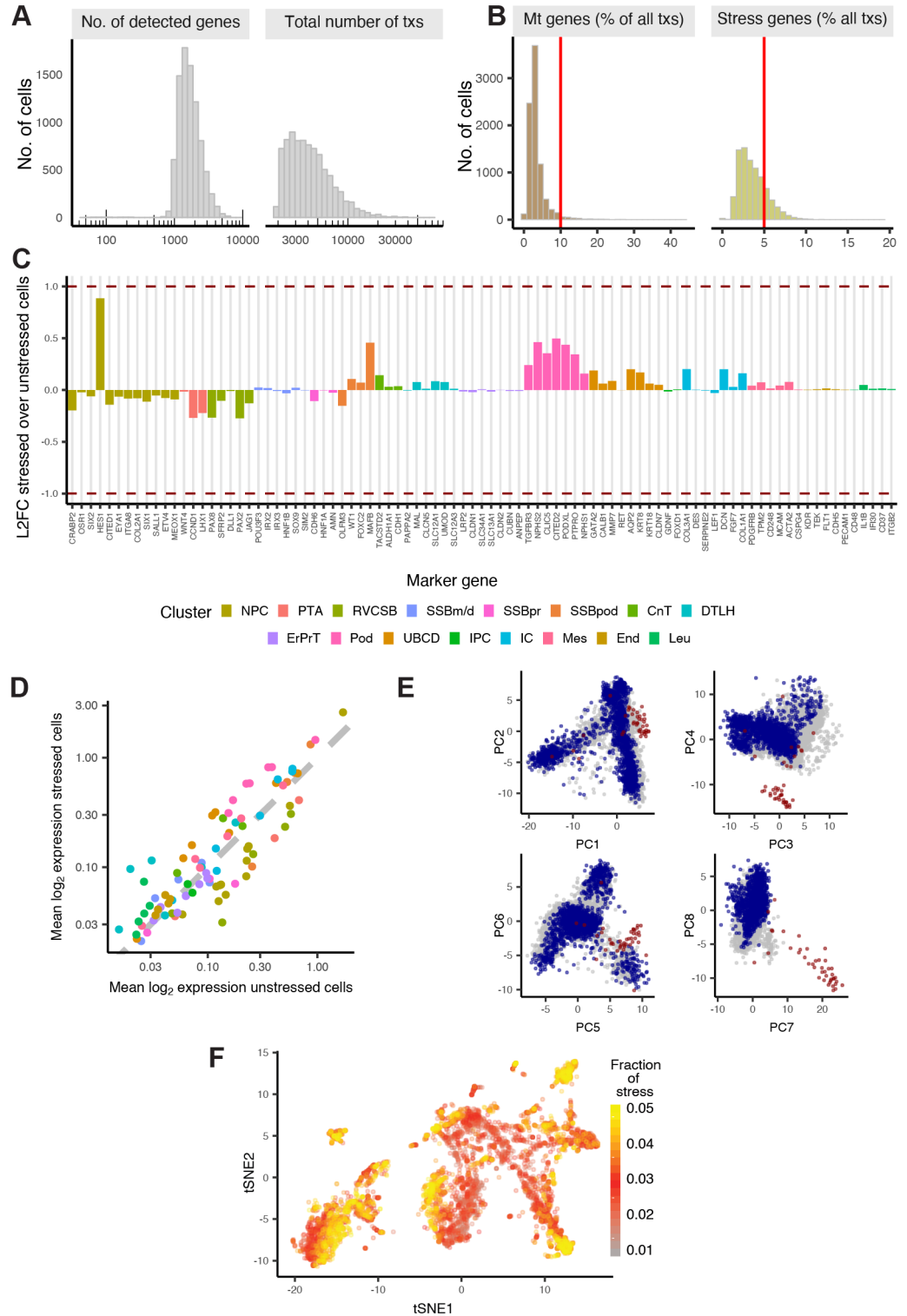
segmented manually. Then, cross-sections of the CM, roughly perpendicular to the outline of the UB, were drawn by hand, approximately 30 pixels apart. For the starting point of each cross-section, the contour length s along the UB starting from the top of the CM (close to the edge of the cortex) was determined (see Fig. 5.16B). The distance s was expressed relative to the total contour length (from top to bottom of the UB). The distance from the UB along the cross-sections was defined to be the distance d (see Fig. 5.16B). Fluorescence intensities were then averaged over lines of 30 pixels length perpendicular to the drawn cross-section. The resulting intensity profiles (which depend on d and s) were then averaged over multiple images and either s or d to get average intensity profiles depending only on d or s . Normalization to the maximum intensity of each profile resulted in the intensity profiles reported in Fig. 5.16D-E. Division of the CITED1 intensity profile by the SIX2 intensity profile gave the ratio plotted in Fig. 5.16D-E. Accuracy, indicated by error bars in the plots, was quantified as the standard error of the mean calculated over all evaluated profiles.

5.3. RESULTS

5.3.1. CLUSTERING AND IDENTIFICATION OF CELL TYPES

We performed single-cell transcriptomics on a human fetal kidney from w16 of gestation, equivalent to 14 weeks of development (Fig. 5.1C). After data pruning and stringent removal of cells affected by stress (Fig. 5.2, Materials and methods), 6,602 cells were retained for further analysis. Clusters of cells were identified by hierarchical clustering after knn smoothing [20]. We assigned cell types (see) to these clusters by expression of marker genes from the literature on mouse kidney development. The studies that linked the genes of this literature set to particular cell types are referenced in Fig. 5.1. After merging similar clusters (Fig. 5.4, Fig. 5.5, Materials and methods), we obtained 22 cell types (Hochane et al. [1] S4 Fig) and visualized the single-cell transcriptomes in a two-dimensional tSNE map [35] (Fig. 5.3).

Figure 5.2 (following page): Removing stressed cells did not bias the scRNA-seq results in the w16 sample. (A) Number of detected genes and total number of transcripts per cell. tx = transcript (B) Relative expression of mitochondrial and stress marker genes per cell. Red line indicates the threshold used to define stressed cells. See Materials and methods for the list of mitochondrial genes and Hochane et al. [1] S2 Table for the list of stress markers. (C and D) \log_2 fold change (L2FC) and scatter plot of the literature set genes (5.1). Red dashed lines indicate fold-change of 0.5 and 2. (E) PCs one to eight of the top 5% most HVGs for all cells. Blue and red points indicate stressed cells and red blood cells, respectively. Expression values in C-E are normalized to library size and log-transformed with a pseudocount of 1. (F) Fraction of stress markers in the 6,602 remaining cells. tSNE map corresponds to 5.3.



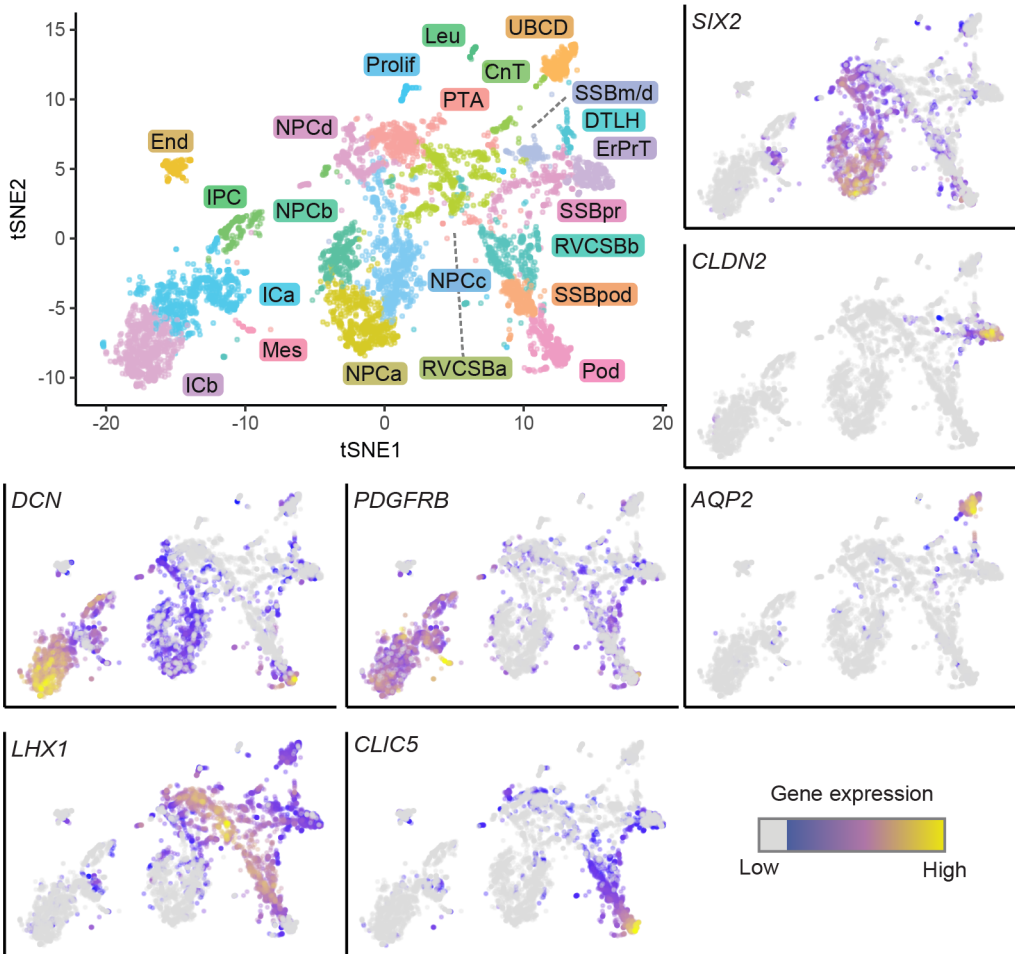
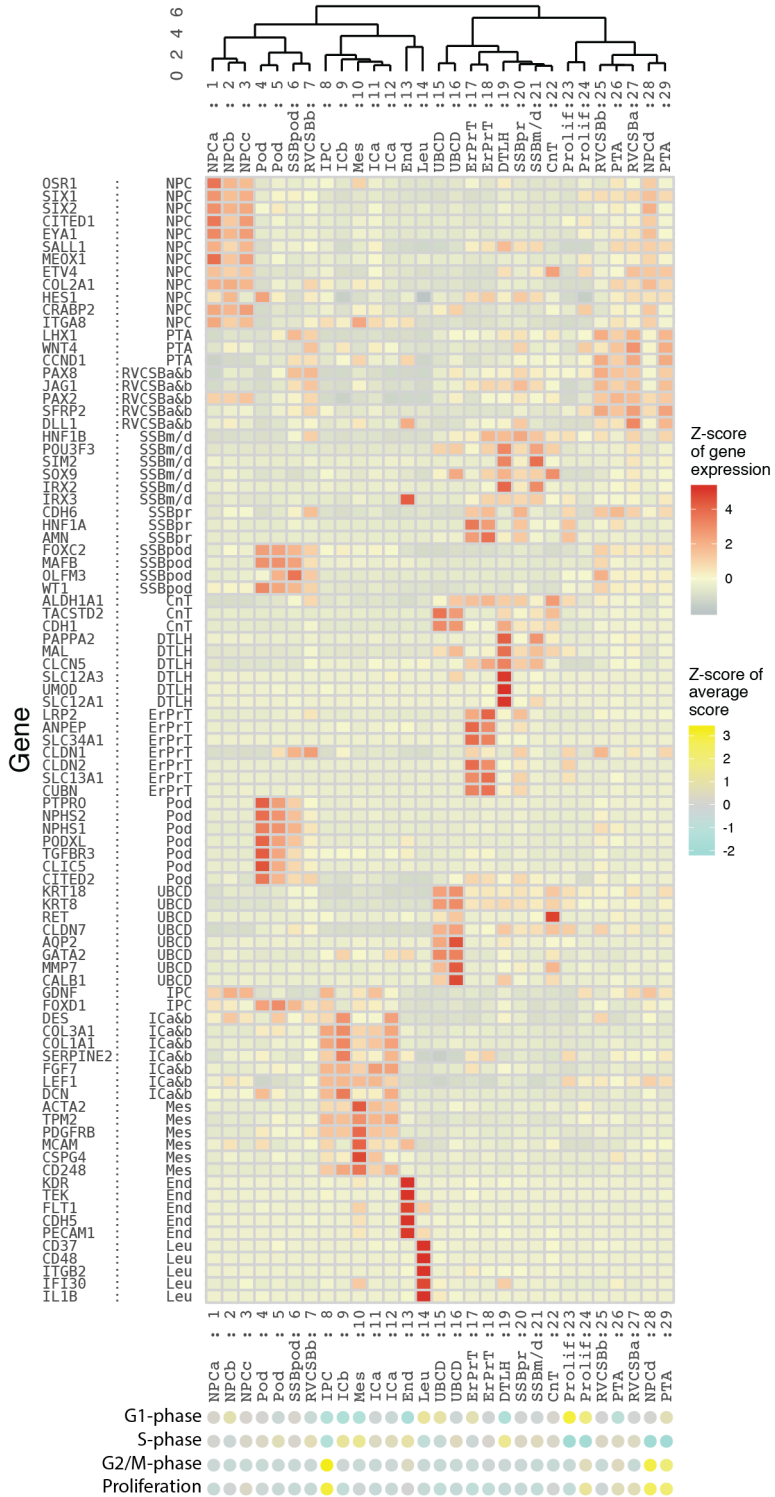


Figure 5.3: Single-cell transcriptomics identified 22 unique cell types in the human fetal kidney. Top left: 2D tSNE map of 6,602 human fetal kidney cells. Colors and labels indicate the assigned cell type. (Other panels) tSNE maps indicating expression of *SIX2*, *LHX1*, *CLDN2*, *CLIC5*, *DCN*, *PDGFRB*, and *AQP2*. Expression is indicated by color; expression values of 1 are plotted in gray.

Figure 5.4 (following page): Adjacent clusters were merged based on similarity in literature set gene expression. Heat map of literature set gene expression. Expression was Freeman-Tukey (FT) transformed, averaged over all cells in the 29 clusters found by hierarchical clustering (indicated by the dendrogram on top of the heat map) and standardized gene-wise. Cluster average cell cycle scores, calculated by Cyclone [27] as well as average expression of proliferation markers [28], are indicated by colored circles below each cluster (Z-score of the mean score or mean expression).



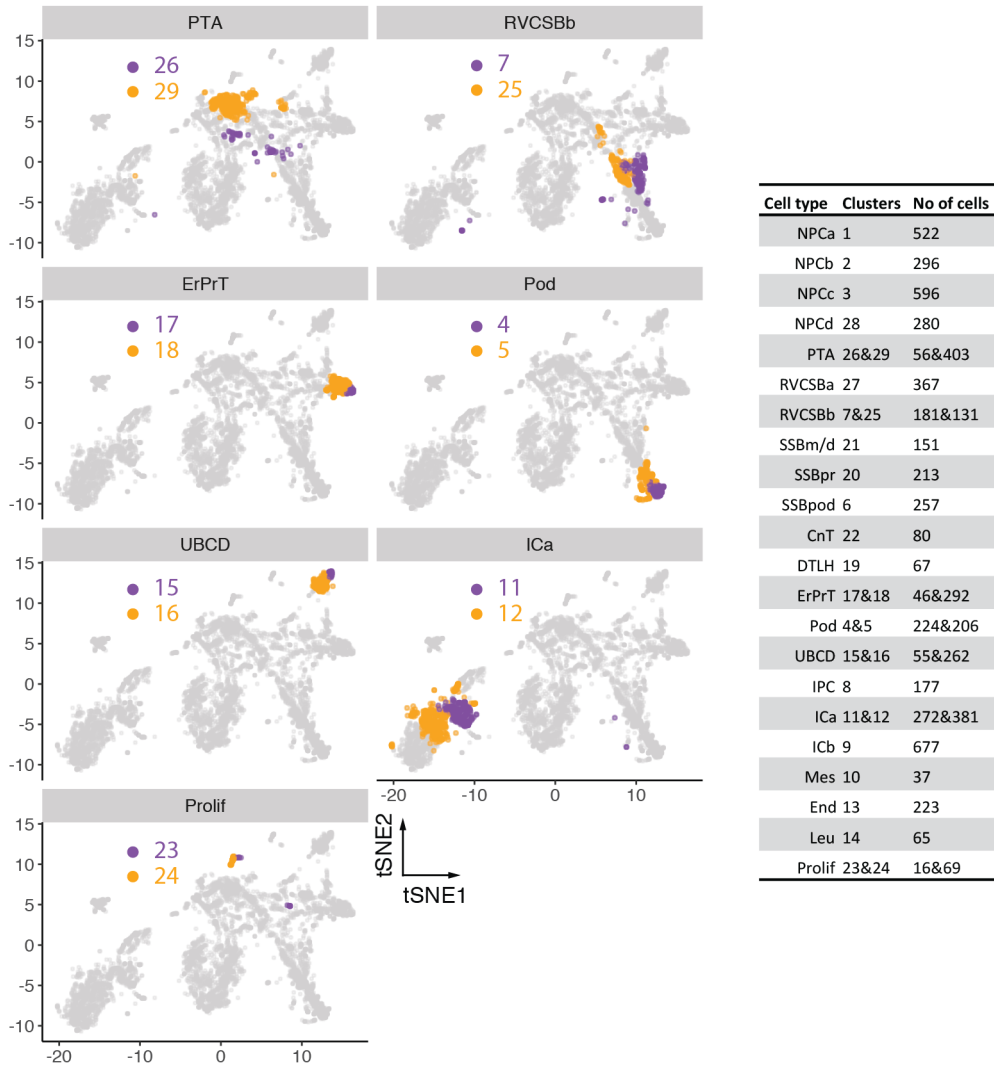
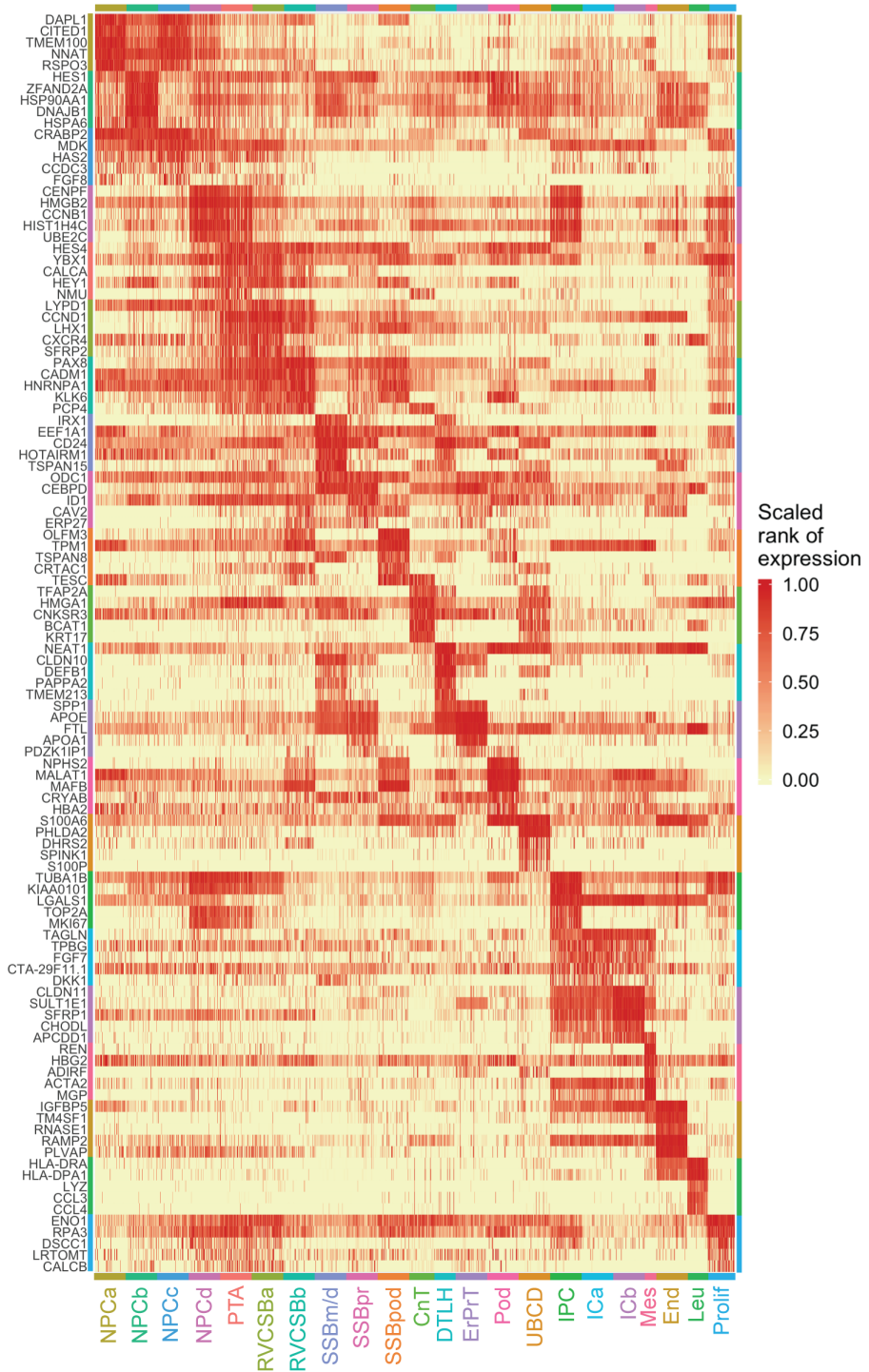


Figure 5.5: Merged hierarchical clusters were similar on the tSNE map. tSNE maps highlighting the clusters that were merged to give the cell types indicated in the titles of each map. (Inset right) Table listing the numbers of cells in each of the 29 original clusters. tSNE map corresponds to 5.3.

Figure 5.6 (following page): HVGs adequately described all cell clusters. Heat map of 2,034 randomly chosen cells (maximum 100 per cluster) and the five most HVGs with a minimum mean expression of 0.01 excluding stress markers (Hochane et al. [1] S2 Table) and ribosomal genes. Genes were assigned to clusters based on highest mean expression within that cluster. Values shown are the ranks of nonzero cells (cells with no expression receive rank 0) divided by the highest rank per gene.

5



The mean expression levels of the literature set genes showed clear differences between cell types (Fig. 5.7). NPCs, which were distributed over four distinct clusters (NPCa-d), were marked by the established markers *SIX2* (Fig. 5.3), *CITED1*, *MEOX1*, and *EYA1*. Expression of these progenitor markers was highest in NPCa, which we hence considered *bona fide* self-renewing NPCs. NPCb showed lower levels of *CITED1* and *SALL1* and higher levels of *GDNF* and *HES1* compared to the other NPC clusters. *HES1*, a transcription factor downstream of Notch signaling, is important for further renal cell differentiation. Compared to the other NPC clusters, NPCc showed higher expression of *CRABP2*, which is related to retinoic acid (RA) signaling [36]. NPCd exhibited low *OSRI*, *CITED1*, and *MEOX1* expression and increased levels of *LEF1*, a known indicator of NPC induction towards differentiation. Compared to the other NPC subtypes, NPCd was also marked by a larger fraction of cells in G2/M-phase (Fig. 5.8A-B) and a higher expression of proliferation markers (Fig. 5.8C), which indicated faster proliferation. We will discuss the relationship between the various NPC clusters in more detail below (see Heterogeneity in the nephrogenic niche).

Nephrogenesis continues with the creation of PTA cells, which in turn develop into the RV and CSB cells. In our data, PTA cells were identified based on high expression of *LHX1* (Fig. 5.3), *JAG1*, *WNT4*, and *CCND1*. Because RV and CSB are mainly distinguishable by morphology, cells belonging to these two structures were grouped in our analysis (RVCSB). RVCSBs were marked by the same genes as PTA cells but they appeared to proliferate less (Fig. 5.8). Furthermore, they expressed markers reflecting more advanced regional patterning, which allowed us to discriminate between two subtypes (a and b). RVCSBa had a higher expression of genes that were recently associated with the distal RV (*SFRP2*, *DLL1*, *LHX1*), whereas RVCSBb expressed genes that indicate the proximal RV (*CDH6*, *FOXC2*, *MAFB*, *CLDN1*, *WT1*).

The next step in development is the formation of the SSB. In our dataset, this structure was represented by three clusters, named according to the part of tubule and glomerular epithelium they are known to give rise to SSBpr, proximal tubule; SSBm/d; SSBpod, podocytes. SSBpr were identified in our data by markers of the ErPrT (such as *NFIA*, *CDH6*, *AMN*), as well as low expression of *SLC3A1*, *LRP2*, and *SLC13A1*, which are known to be found in more mature proximal tubule cells. Therefore, SSBpr were likely precursors of the ErPrT cells, which expressed higher levels of early proximal markers together with *CLDN2* (Fig. 5.3), *ANPEP*, and *SLC34A1*. Another cluster accounted for the precursor cells of the loop of Henle (LOH) and the distal tubule in the SSB (SSBm/d). This cluster could be identified by the presence of *IRX1*, *IRX2*, *SIM2*, *SOX9*, *POU3F3*, and *HNF1B*, together with low expression of *PAPPA2* and *MAL* and the absence of *CDH6* and *HNF1A*. Cell types that are known to develop from the SSBm/d were found together in one cluster (DTLH). This cluster showed high expression of the distal markers *MAL*, *CLCN5*, *SLC12A3*, and *POU3F3*, which are specific to the distal tubule, as well as *SLC12A1*, *PAPPA2*, and *UMOD*, which are found in the LOH. Finally, cells that likely gave rise to podocytes, SSBpod, clustered separately. These cells showed high expression of *MAFB* and *FOXC2*, both transcription factors necessary for the development of podocyte identity, and low levels of the mature podocytes markers *CLIC5* (Fig. 5.3), *PTPRO*, *NPHS1*,

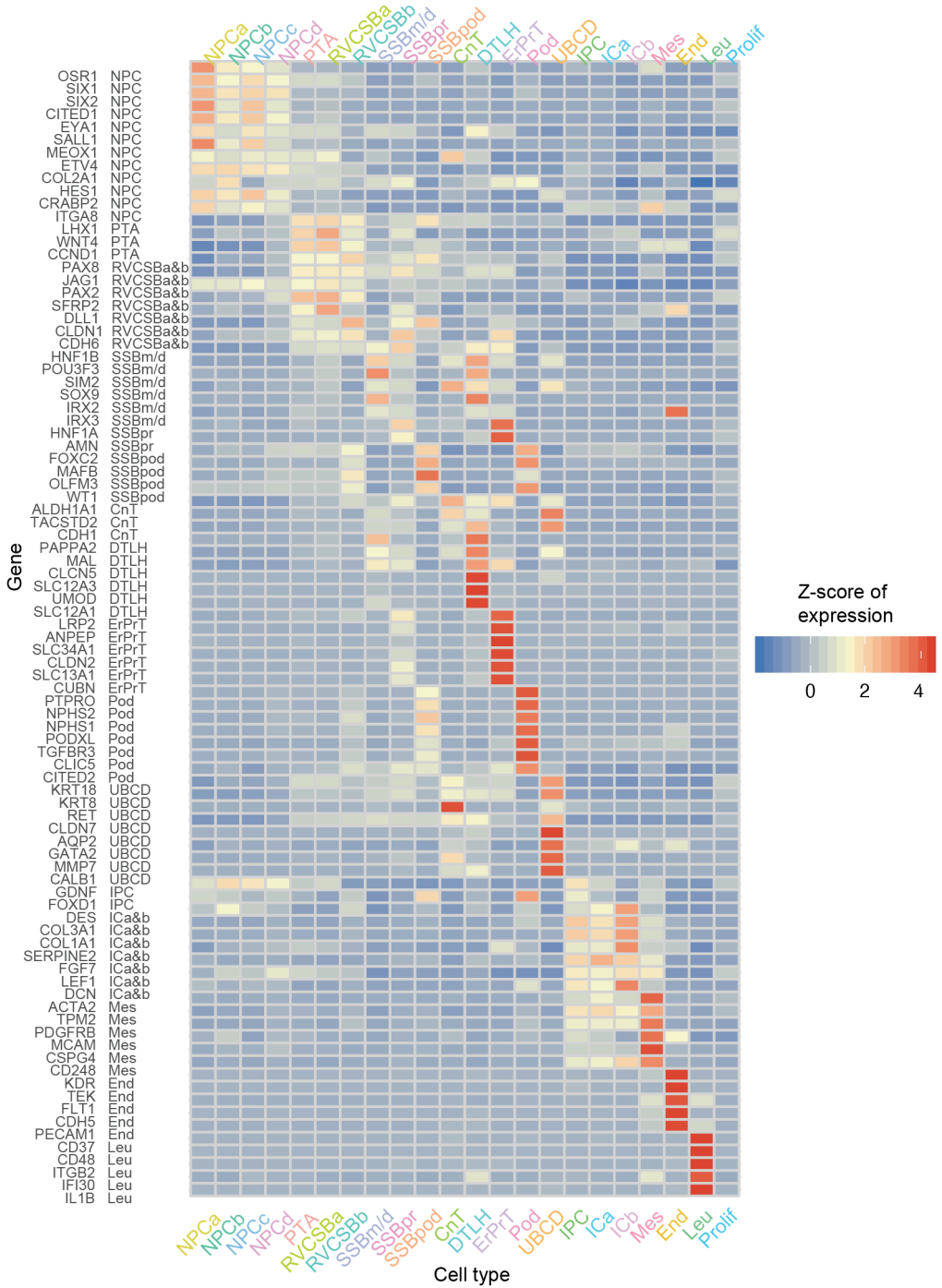


Figure 5.7: Known markers elucidated the cell types corresponding to each cluster. Heatmap of literature set gene expression in the 22 identified cell types. Expression was FT transformed, averaged over all cells in a cluster and standardized gene-wise.

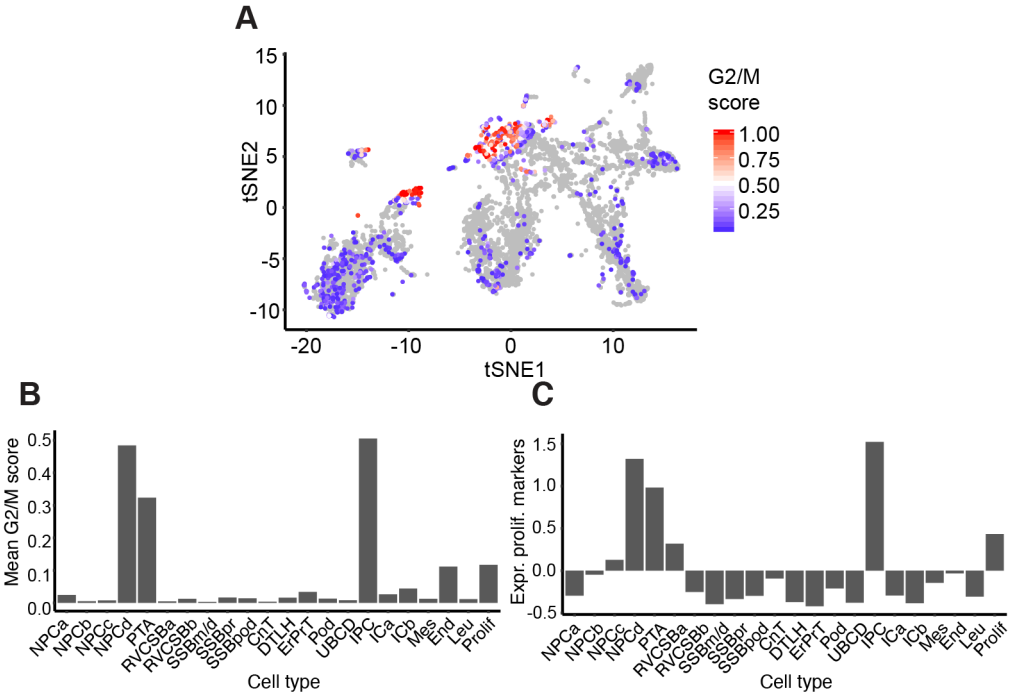


Figure 5.8: Proliferation states varied per cell type. (A) tSNE map of all cells with color indicating the G2/M score (calculated by the Cyclone tool [27]). This score reflects the likelihood that a cell is in G2/M phase. (B) G2/M scores from panel A averaged over the cells in each cell type. (C) Mean expression of proliferation markers [28] (Z-scores) per cell type.

and *NPHS2*. This cluster also showed the highest expression of *OLFM3*, previously identified as a specific marker of podocyte precursors residing in the visceral part of the proximal segment of the SSB. In contrast to SSBpod, podocytes showed higher expression of mature podocyte markers and lower levels of *MAFB*. Differences between SSBpod and podocytes will be studied in more detail below (see Podocyte development). Because of the high similarity in gene expression between SSB and capillary loop stage, we could not exclude that the SSB clusters also contained cells from the capillary loop stage.

Cells of the connecting tubule (CnT), which connects the distal tubule to the collecting duct, could also be identified in the data. They shared markers with the collecting duct (such as *ALDH1A1*, *TACSTD2*, and *CDH1*), distal tubule (*SOX9*, *POU3F3*), and UB (*RET*, *KRT8*, *KRT18*, *MMP7*). Cells of the ureteric bud/collecting duct (UBCD) were strongly marked by well-known genes like *AQP2* (Fig. 5.3), *CALB1*, *KRT8*, *KRT18*, *RET*, and *GATA2*, found in the collecting duct as well as the stalk and tip of the UB.

The developing nephrons are surrounded by interstitial tissue, a separate lineage that originates in interstitial progenitor cells (IPCs). We identified IPCs by coexpression of *FOXD1* and *GDNF*. These cells also expressed lower levels of markers known to be found in more mature cells like *PDGFRA* for ICs or *PDGFRB* (Fig. 5.3) and *ACTA2* for mesangial cells. We identified two subtypes of IC (a and b), which were similar in their marker gene profile. Compared to IPCs, they lacked *FOXD1* and expressed less (ICa) or no (ICb) *GDNF*. ICa showed high levels of *FGF7*, which has been localized to the renal fibroblasts or stroma surrounding the ureter and the collecting system. ICa also showed high levels of *TPM2* and *ACTA2*, markers of smooth muscle-like cells. ICb, on the other hand, expressed genes like *DCN* (Fig. 5.3), *DES*, *SERPINE2*, and *COL3A1*, which are known to mark cortical stromal cells. Endothelial cells were identified by markers such as *KDR* and *TEK*, whereas leukocytes showed many specifically expressed genes, such as *CD37* or *CD48*. Finally, one cluster of cells (proliferating cells) had a higher expression of proliferation markers compared to most other cell types (Fig. 5.8C) but lacked discernible cell type markers.

5.3.2. DEVELOPMENTAL FLOW

The literature-based analysis of the found clusters seemed to suggest that cells cluster by developmental progression (e.g., NPCs versus PTA cells), as well as location (e.g., RVCSBa, distal, versus RVCSBb, proximal). Because the interpretation of clusters is sometimes based on genes that are expressed in multiple developmental stages, we wanted to retrieve the developmental flow with an independent method. We used *Monocle 2* [29] to learn a graph that represents the developmental hierarchy of the cell types from the PTA on (Fig. 5.9A). Subsequently, cells were placed on a pseudotime scale rooted in the PTA (Fig. 5.9B-C). This analysis showed that PTA cells were followed by RVCSBa and RVCSBb, the SSB clusters, and finally the clusters identified as more mature types (DTLH, ErPrT, podocytes). Therefore, the clustering was strongly driven by developmental progression. RVCSBa cells were distributed over a fairly broad period of pseudotime and already occurred before branch point 1, which separates proximal from distal cell fates (Fig. 5.9A). This might indicate that some of these cells preceded the RVCSBb, whereas others were primed to develop into distal fates. RVCSBb cells, however, only appeared after branch

point 1, which confirmed that they were likely progenitors of proximal cell fates. On three separate branches, SSBm/d preceded DTLH, SSBpr preceded ErPrT, and SSBpod preceded podocytes, which confirmed the identity of the SSB clusters. The temporal relationship of the NPC subtypes will be discussed in detail below (see Heterogeneity in the nephrogenic niche).

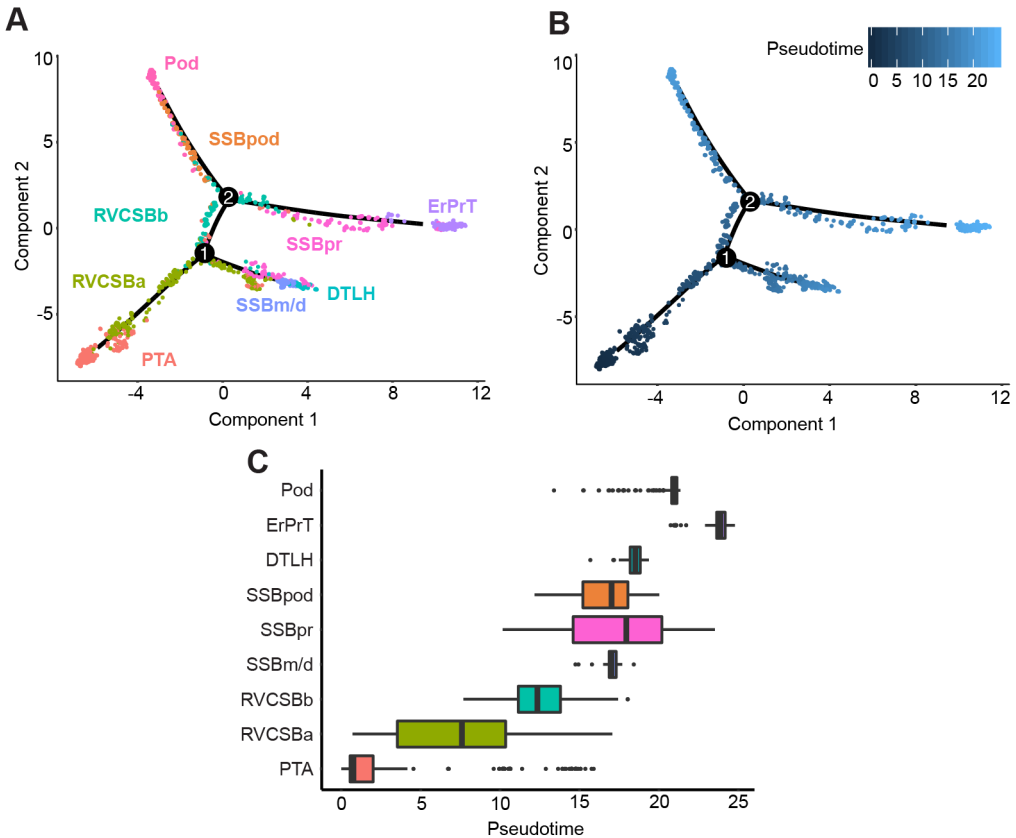


Figure 5.9: Pseudotime analysis clarified the developmental relationship of the cell clusters. (A) Two-dimensional embedding (with the *DDRTree* algorithm [30]) of all w16 kidney cells, calculated by *Monocle 2* [29]. The graph learned by the algorithm is shown as a black line. Colors and labels indicate cell types. (B) Same embedding as in panel A. Color indicates pseudotime calculated by *Monocle 2*. (C) Box plots of cell type distribution over pseudotime.

5.3.3. COMPARISON WITH EXISTING SINGLE-CELL TRANSCRIPTOMICS DATA

To further confirm the interpretation of the cell clusters, we wanted to compare our data with an existing single-cell transcriptomics study of a w17 fetal kidney by Lindström et al. [13]. To that end, we first corrected for batch effects, using a method based on matching mutual nearest neighbors in the two datasets [26]. After correction, the two

5

datasets showed a large degree of overlap (Fig. 5.10, Fig. 5.11). This allowed us to use the cell types found by Lindström et al. [13] to classify the cell clusters found here, using a knn approach (see Materials and methods). NPCa-c were also classified as NPC by Lindström et al. [13], whereas NPCb were considered *primed NPC*, which supports the notion that NPCb were primed to differentiate. The NPCd cluster was classified as *proliferating cells* This classification is in agreement with our observation that NPCd seemed to proliferate more than other NPC subtypes (Fig. 5.8). Because NPCd cells expressed low levels of NPC markers (such as *SIX2* and *CITED1*), these cells were likely in a transition state between NPCs and PTA cells. Although the majority of PTA cells identified here were considered *PTA/RV I* by Lindström et al. [13], RVCSBa cells were spread over multiple cell types. This spread was likely due to the fact that transitory cell types are transcriptionally similar, and their clustering is therefore less robust. Nevertheless, the *PTA/RV II* cluster received most of the RVCSBa cells. RVCSBb cells were called *podocyte precursors* in the Lindström et al. [13], whereas SSBpod as well as podocytes were classified as *podocytes*. In our dataset, RVCSBb directly preceded SSBpod (Fig. 5.9A), so they could indeed be considered podocyte progenitors. Below, we will show that SSBpod did form a cell state separate from podocytes and should not be grouped with them (see Podocyte development). In agreement with our analysis, the majority of SSBpr were classified as *proximal precursor* or *proximal tubule*, and all ErPrT were considered *proximal tubule* by Lindström et al. [13]. CnT and DTLH were both classified as *distal/LOH precursor*. The fact that two cell types in the study by Lindström et al. [13] (*podocytes* and *distal/LOH precursor*) were split in multiple subclusters in our study likely reflects differences in sample preparation. Whereas Lindström et al. [13] preferentially released single cells from the nephrogenic niche, here, the whole kidney was used. Consequently, the Lindström et al. [13] dataset has a finer resolution of NPCs and early, proliferating cell types, whereas our dataset allowed us to resolve more mature cell types. The two datasets therefore complement each other.

5.3.4. MARKER IDENTIFICATION

To confirm the inferred cell types and also identify novel markers, we pursued two complementary strategies. First, we determined a set of marker genes based on their usefulness as classifiers for individual cell types: for each gene, the performance of a binary classifier was evaluated by the AUROC and combined with expression level filtering (see Materials and methods). This resulted in 88 marker genes (See Fig. 5.12, marker set, Hochane et al. [1] S3 Table). Only 11 of these markers overlapped with the 89 genes in the literature set (See Fig. 5.13C). To our knowledge, many of the remaining markers had not been associated with kidney development in previous studies. As an independent approach, we used the KeyGenes algorithm [31] to identify classifier genes among the 500 most HVGs, using two-thirds of all cells as a training set. Based on the classifier genes determined by KeyGenes, we next predicted the cell types of the remaining one-third of the cells (test set). Cell types could be predicted with an average certainty (id score) of 0.59; 24% of the cells in the test set obtained an id score higher than 0.8. Of the 95 classifier genes (See Fig. 5.12A, KeyGenes set, Hochane et al. [1] S3 Table), 24 were the same as in the marker set, and 14 were common with the literature set (See Fig. 5.13B).

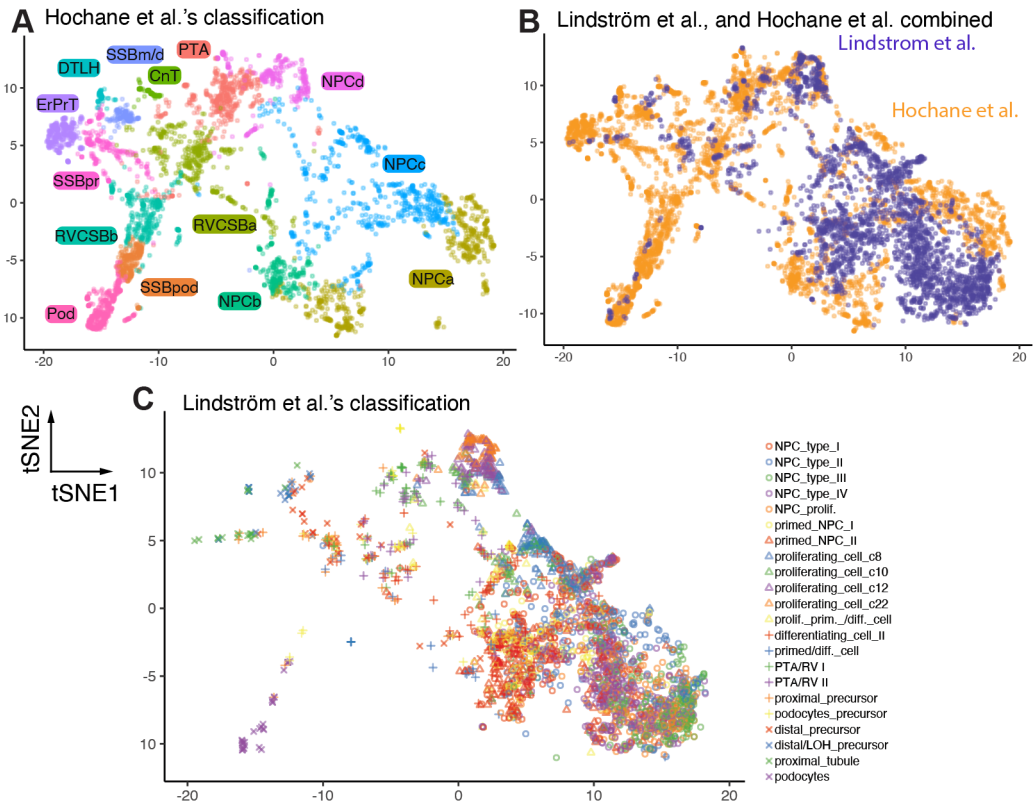


Figure 5.10: Comparison with an existing single-cell transcriptomics dataset showed congruent expression profiles. Two-dimensional tSNE maps comparing the data presented here with the data from Lindström et al. [13] both restricted to the nephrogenic niche by their own classification. The map was calculated using both data sets after batch correction [26]. (A) Only cells measured in this study are shown. Color and labels indicate the classification developed in this study. (B) Same tSNE map as above. Color indicates the data set. (C) Same tSNE map as (B). Only cells measured by Lindström et al. are shown. Color and labels indicate the classification by Lindström et al. See Fig. 5.1.1.

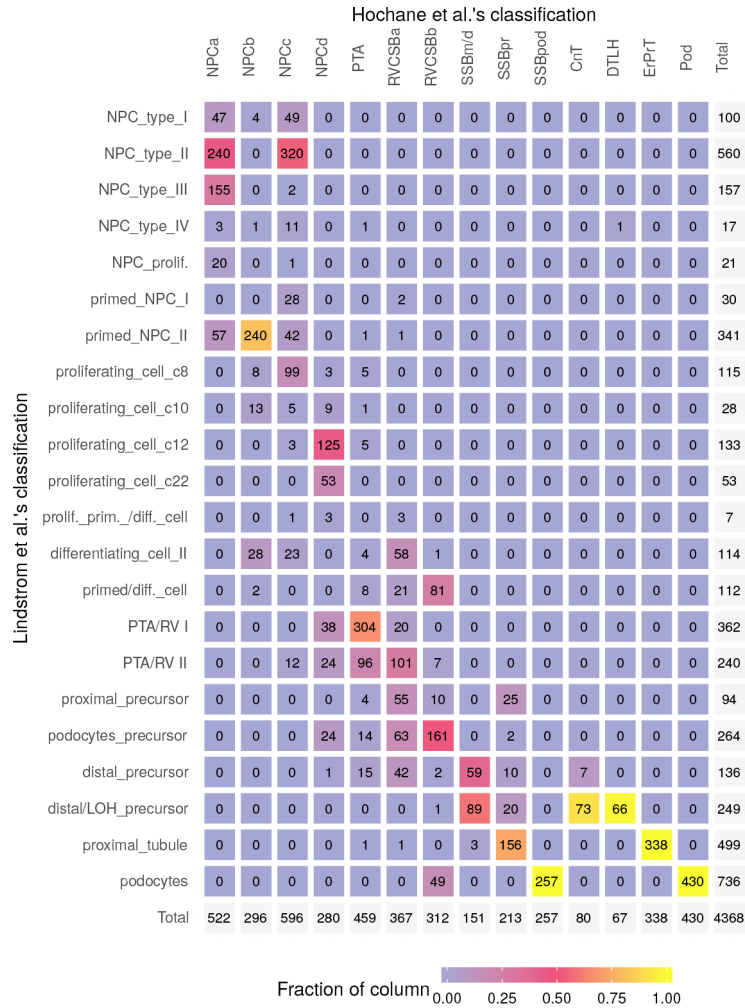


Figure 5.11: Comparison with an existing single-cell transcriptomics dataset showed differences in cell type distribution. Confusion matrix relating the cells measured in this study to the classification by Lindström et al. [13]. After batch correction, cells measured here were mapped on the cells in the Lindström et al. data set using a nearest neighbors-based approach (see Materials and methods). See Fig. 5.10.

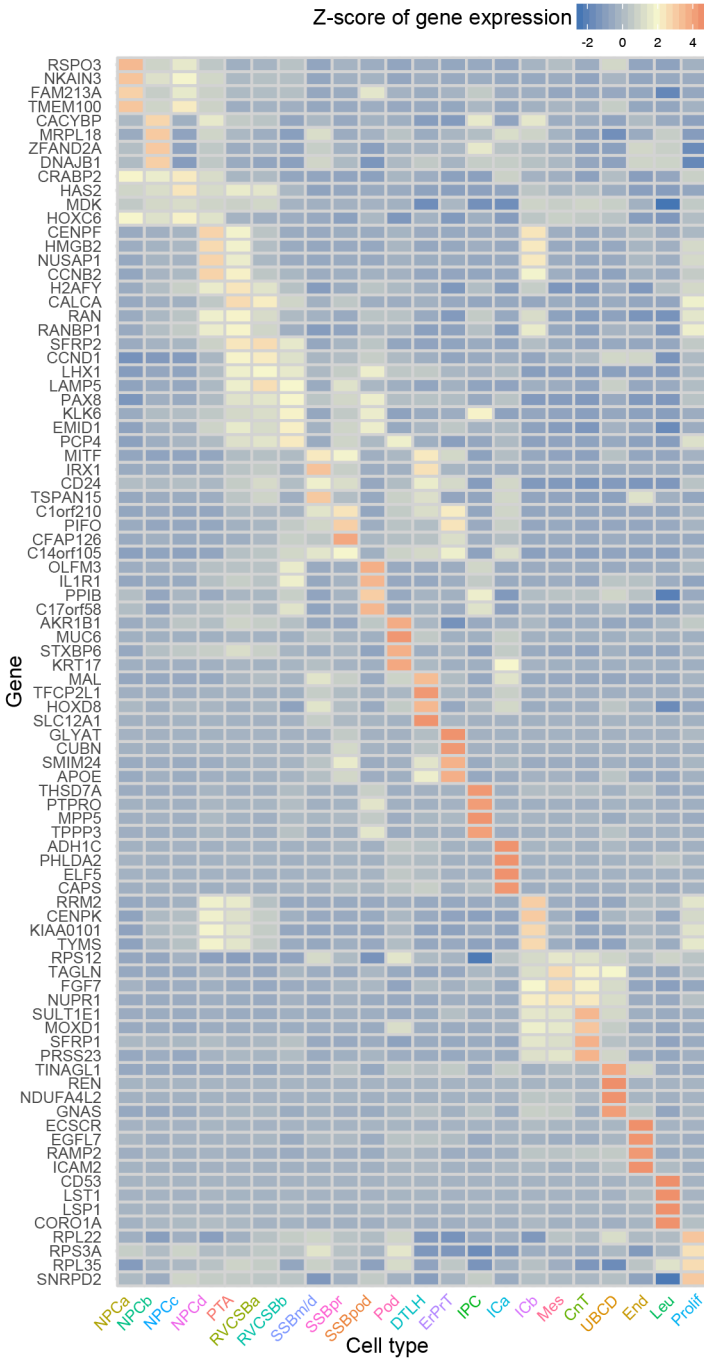


Figure 5.12: An ROC-based method identified novel marker genes. Expression heat map of the 88 genes identified by a method that evaluates the ROC for each gene (marker set, Hochane et al. [1] S3 Table). Expression was FT transformed, averaged over all cells in a cluster, and standardized gene-wise. See Fig. 5.13.

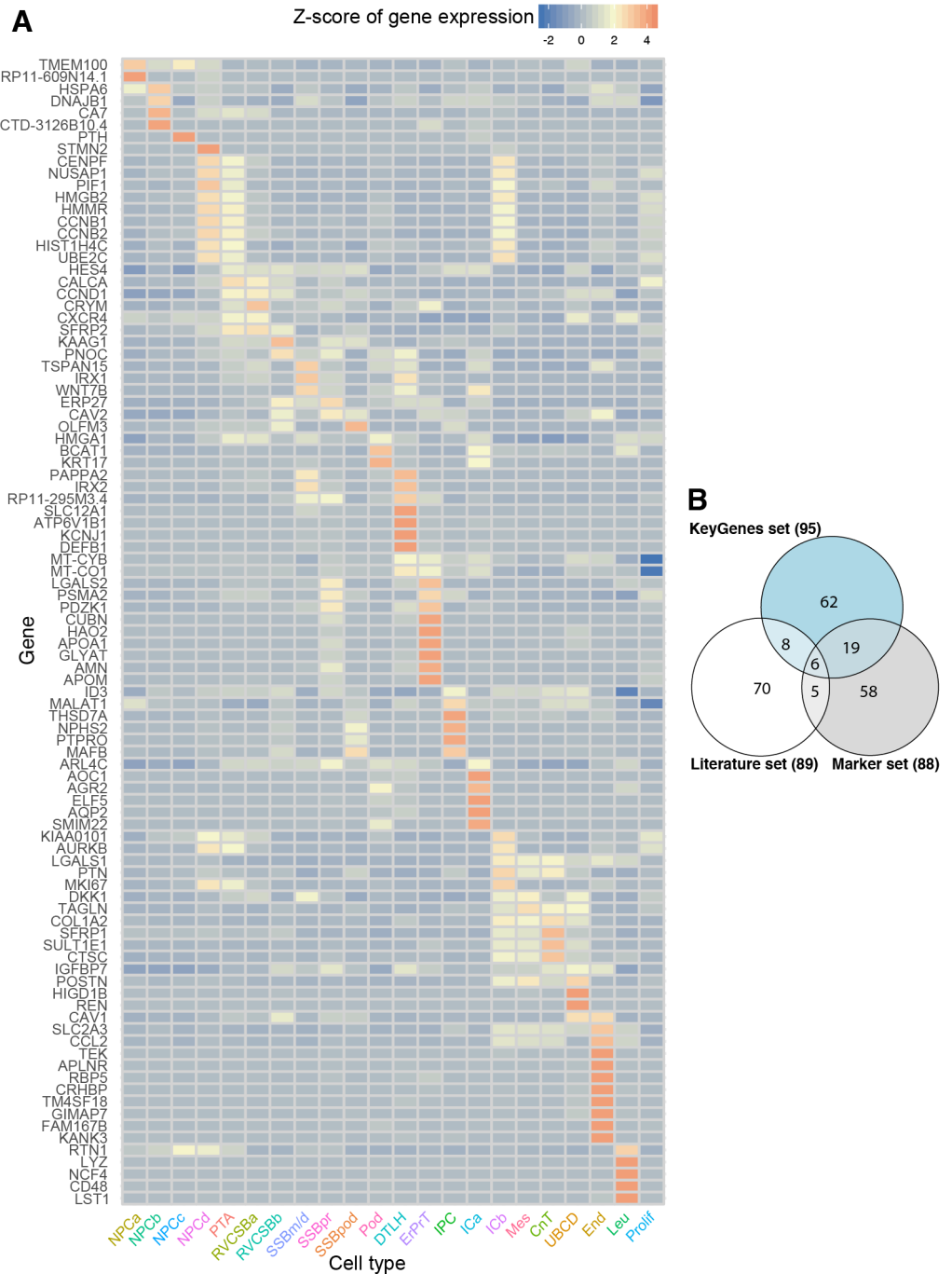


Figure 5.13: The KeyGenes method identified novel marker genes. (A) Expression heat map of the 95 genes identified by the KeyGenes algorithm (KeyGenes set, Hochane et al. [1] S3 Table). Expression was FT transformed, averaged over all cells in a cluster, and standardized gene-wise. (B) Euler diagram of the literature set, marker set, and KeyGenes set (Hochane et al. [1] S3 Table). See Fig. 5.12.

Because the interpretation of the found cell clusters was largely based on markers identified in mouse development, we were wondering whether the new markers identified here were informative for the classification of cell types in the mouse kidney. Using a scRNA-seq measurement of cells from a whole P1 mouse kidney [18], we plotted the expression of the newly identified marker genes in single cells (Hochane et al. [1] S7 Fig). In many cases, markers that were found to label a particular cell type in the human fetal kidney were coexpressed in the same subset of mouse cells. A few markers, however, were either ubiquitously expressed or almost completely absent. This might be due to interspecies differences.

5.3.5. COMPARISON OF DIFFERENT DEVELOPMENTAL AGES

By establishing the identity of cell clusters at w16, we obtained a snapshot of cell type diversity in the fetal kidney. To explore whether the identified expression patterns change dynamically throughout development, we analyzed four additional samples from different developmental ages (w9, w11, w13, and w18), which together contained 11,359 usable cells. Using, again, batch correction based on mutual nearest neighbors [26], we visualized all samples in a common tSNE map (5.15, 5.14). Overall, gene expression in the different samples was largely overlapping for the majority of cell types. For example, proximal tubules cells (ErPrT) appeared at the same positions in the tSNE map in all samples (5.15B). The position of podocytes shifted systematically across different ages, which corresponds to a continuing change in expression pattern (5.15C). This observation might suggest that podocytes further matured in terms of their expression pattern after being specified.

Differential expression analysis of podocytes of different ages revealed 109 differentially expressed genes (fold change > 2 in any comparison, FDR < 0.05, Hochane et al. [1] S4 Table). Functional annotation analysis of these genes showed significant enrichment of two gene ontology (GO) terms: *proteinaceous extracellular matrix* (adjusted p-value = $1.9 \cdot 10^{-3}$, including *SPON2*, *BGN*, *COL1A2*, and *CTGF*) and *extracellular exosomes* (adjusted p-value = $1.4 \cdot 10^{-3}$, including *NPNT*, *S100A10*, *ANXA1*, and *EPCAM*). Some of the differentially expressed genes have been shown to be important for kidney development. For example, *NPNT* and *DCN* showed increasing expression from w11 to w18. Knock-out of the extracellular matrix protein *NPNT* in mice decreases the invasion of the UB and causes agenesis or hypoplasia [37]. *NPNT* was further shown to be expressed in the glomerular basement membrane and to be necessary for podocyte adhesion in mice [38]. Ablation of this gene in mice causes podocyte effacement. As in the case of *NPNT*, *DCN* has been reported to be part of the glomerular basement membrane proteins [39]. This gene appeared strongly up-regulated in podocytes between w11 and w13 or w18 (fold changes of 3.25 and 4.6, respectively). The increase of *NPNT* and *DCN* expression over time in our data set could reflect an increase in adhesion between podocytes and glomerular basement membrane. Podocytes further showed significant differential expressions of genes related to stress, like *HSPA1A* and *HSPA1B* or *NFKB* genes (*NFKB2*, *NFKB1A*, and *REL*), with the highest levels at w18. This might suggest that dissociation-related stress increases with age for podocytes, maybe related to stronger adhesion of

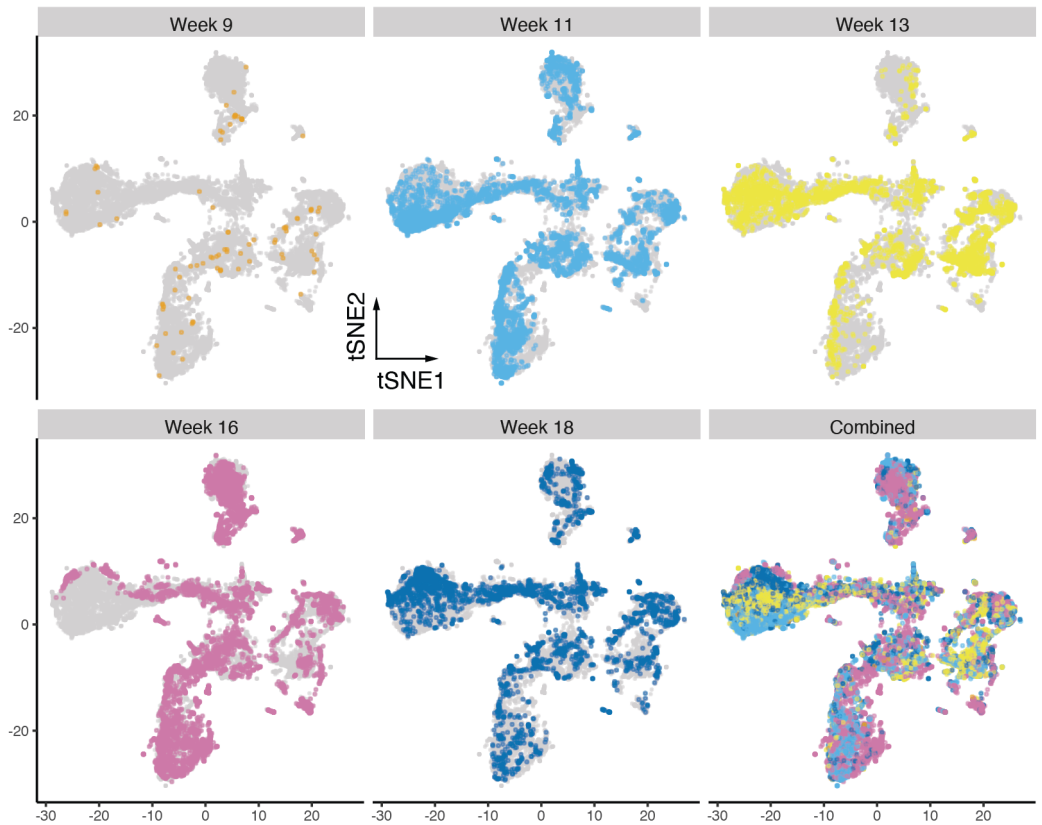


Figure 5.14: Samples of different developmental ages had a similar cell type diversity. tSNE map calculated for all five samples (w9, w11, w13, w16, w18) combined after batch correction [26]. Developmental age is indicated by color.

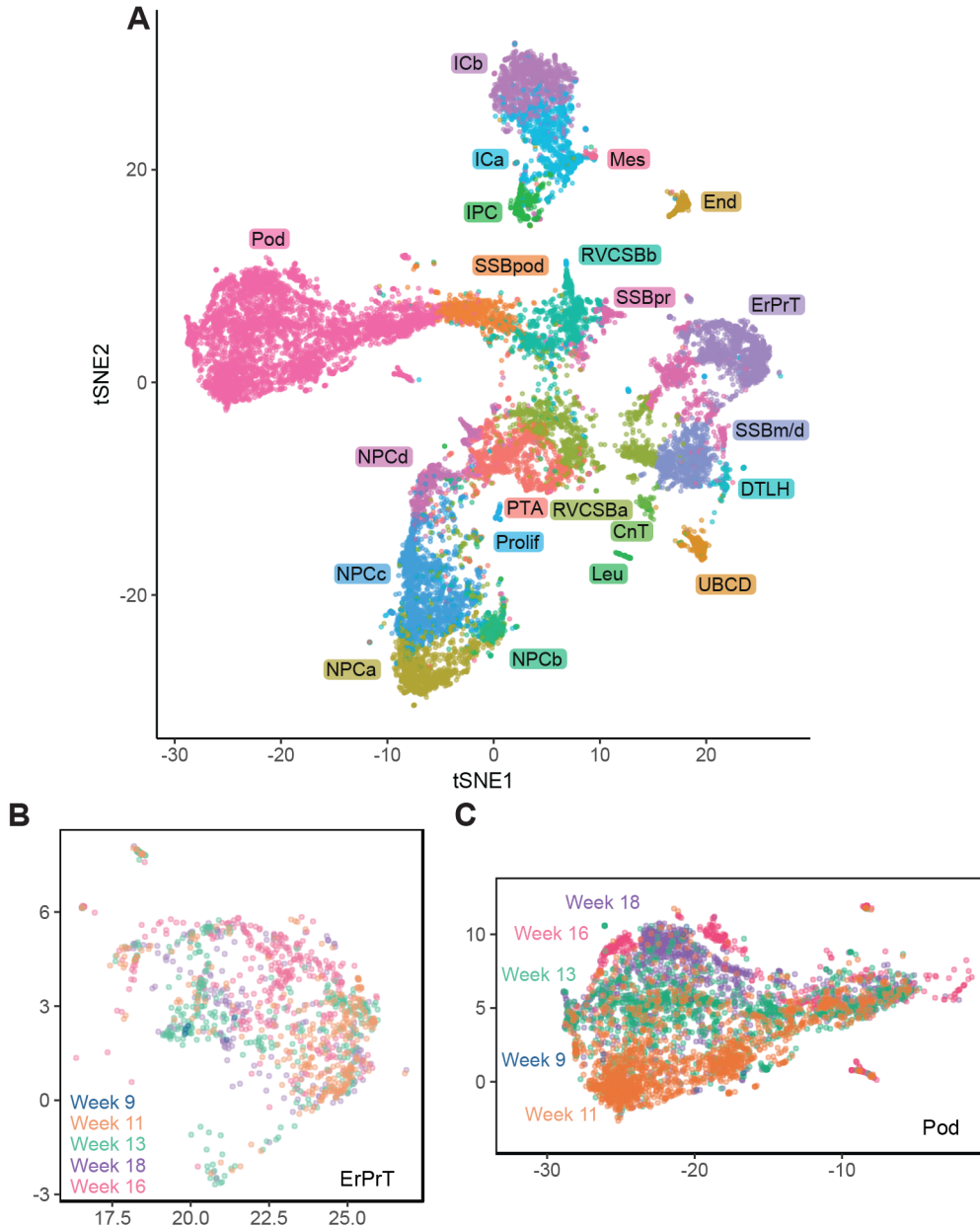


Figure 5.15: Comparison of different developmental ages suggested continued expression changes in podocytes. (A) tSNE map combining all five samples (w9, w11, w13, w16, w18). Samples were corrected for batch effects by matching mutual nearest neighbors [26]. Cells in the w9, w11, w13, and w18 samples were classified by comparing to the w16 sample using a knn-based approach (see Materials and methods). (B) tSNE map of all ages restricted to ErPrT. Labels and colors indicate ages. Six outlier cells were omitted from this plot to improve visualization. (C) tSNE map of all ages restricted to podocytes. Labels and colors indicate ages.

the cells, or that stress-related genes have another, physiological role in development.

We would like to emphasize that the observed gene expression changes with age should be considered with caution because they might be related to the differences in genotype between the samples. A much larger number of samples would be necessary to rule out such interindividual differences as a cause.

Having established the identity of the cell clusters, we next wanted to demonstrate how the dataset can be used to explore different aspects of kidney development. We specifically focused on the nephrogenic niche, which showed pronounced heterogeneity, and the development of podocytes, which progressed via a distinct, intermediate cell state (SSBpod).

5.3.6. HETEROGENEITY IN THE NEPHROGENIC NICHE

The formation of the nephron epithelium starts with the NPCs that differentiate and form the PTA, RV, and CSB. Studies in the mouse suggest that cells in the NPC compartment are not biased towards a particular lineage and patterning is first detectable in the PTA [40]. Nevertheless, the w16 scRNA-seq indicated the presence of several nephron progenitor subpopulations, NPCa-d. To clarify the temporal relationship of these clusters, we employed *Monocle 2* again to arrange them together with the PTA cells on a pseudotime scale (5.16A). NPCa clearly preceded NPCb and c, which seemed to appear around the same pseudotime. NPCd cells followed NPCb and c and preceded PTA. This analysis suggested that NPCa are the *bona fide* NPCs and give rise to NPCb and c. NPCd, which were likely more proliferative than the other NPCs (5.8), seemed to be an intermediate state between (slowly cycling) NPCa-c and the PTA.

To localize the NPC clusters in the tissue, we made use of the fact that they expressed various levels of *CITED1* and *SIX2* (5.7): although NPCa and NPCc exhibited roughly similar levels of these markers, NPCb and NPCd had lost *CITED1* almost completely, while retaining some *SIX2* expression. In an immunostaining of a w15 kidney, *CITED1* and *SIX2* appeared overlapping in a subset of cells (5.16C). Quantification of the fluorescence signal (see 5.2) revealed clear differences between their expression patterns. Although *SIX2* expression was approximately constant throughout the CM, *CITED1* expression decreased, relative to *SIX2*, with increasing (radial) distance from the UB (5.16D-E). A marked drop of *CITED1* was visible between 10 and 20 μm from the UB, which approximately corresponds to the first layer of cells. To exclude that the observed difference between *SIX2* and *CITED1* expression was due to the different fluorophores on the secondary antibodies, we repeated the experiment with swapped fluorophores. This measurement produced a very similar expression gradient (5.16F). To exclude that the observed effect was influenced by PTA found in the CM towards the stalk of the UB, the analysis was repeated, taking only the 20% of CM cells closest to the edge of the cortex into account.

A similar expression gradient was observed (5.16H). This result implies the existence of a *CITED1* low/*SIX2* high subpopulation of cells, which are not in contact with the UB. Secondly, we observed that *CITED1* decreased relative to *SIX2* towards the interface with the PTA and the stalk of the UB (5.16D-E). A similar observation was made when

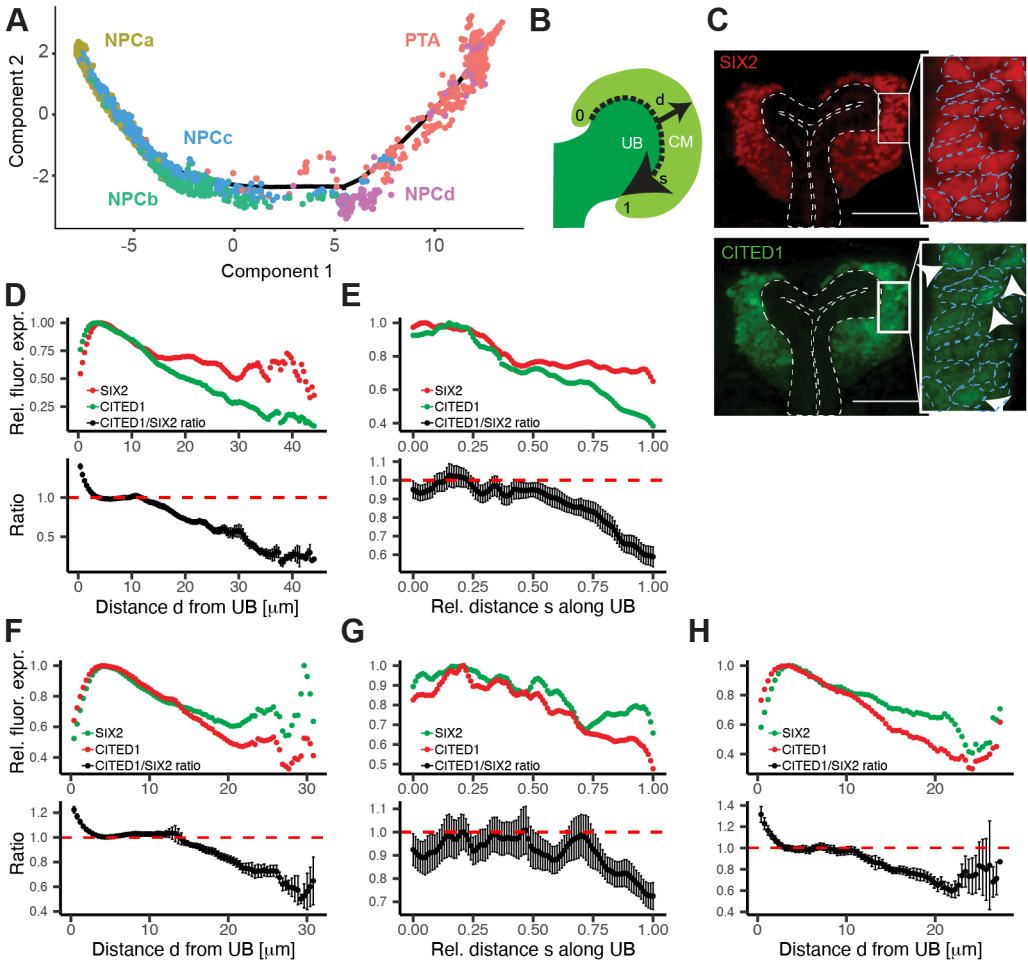


Figure 5.16: The nephrogenic niche exhibited a complex spatial organization. (A) Pseudotime analysis of the nephrogenic niche and the PTA. Two-dimensional *DDRTree* [30] embedding and the learned graph (shown as a black line) were calculated with *Monocle2* [29]. Labels and colors indicate cell types. (B) Schematic sketch of the CM indicating the distance d from the UB to the edge of the CM (solid arrow) and the relative distance s along the UB (dashed arrow), in which 0 and 1 represent the top and bottom of the CM, respectively. (C) Representative image of SIX2 and CITED1 immunostaining in a w15 human fetal kidney. Dashed lines in the insets indicate the outline of the nuclei, based on DAPI signal. Arrows in the inset point to cells in which CITED1 is concentrated in the nucleus. Scale bar = $50\mu m$. (D and E) Quantification of SIX2 and CITED1 immunostaining with respect to the distance d from UB or distance s along the UB; see panel A. Error bars indicate the standard error of the mean standard error of the mean calculated over all evaluated profiles ($n = 24$). (F and G) Same as D and E, but the fluorophores on the secondary antibodies were swapped. ($n = 19$). (H) Quantification of SIX2 and CITED1 immunostaining with respect to the distance d from UB in which only cells with a relative distance s (along the UB) < 0.2 were taken into account. Error bars indicate the SEM calculated over all evaluated profiles ($n = 19$).

the experiment was repeated with swapped fluorophores (5.16G). Taken together, these results suggested that NPCa and NPCc were located closer to the surface of the UB and closer to the tip of the UB compared to the other NPC subtypes. Additionally, we also observed differences in subcellular localization of CITED1 protein within the CITED1 high compartment. Although for the majority of cells CITED1 was found in the cytoplasm, in several cells it was concentrated in the nucleus (right inset in 5.16C). In contrast, SIX2 was always found restricted to the nucleus (left inset in 5.16C). This observation might indicate that CITED1 was only active in a small population of cells, which would constitute another layer of cell-cell heterogeneity.

In addition to the observed heterogeneity in CITED1 and SIX2, differences between the NPC clusters could also be gleaned from the set of novel markers (marker set, Hochane et al. [1] S3 Table).

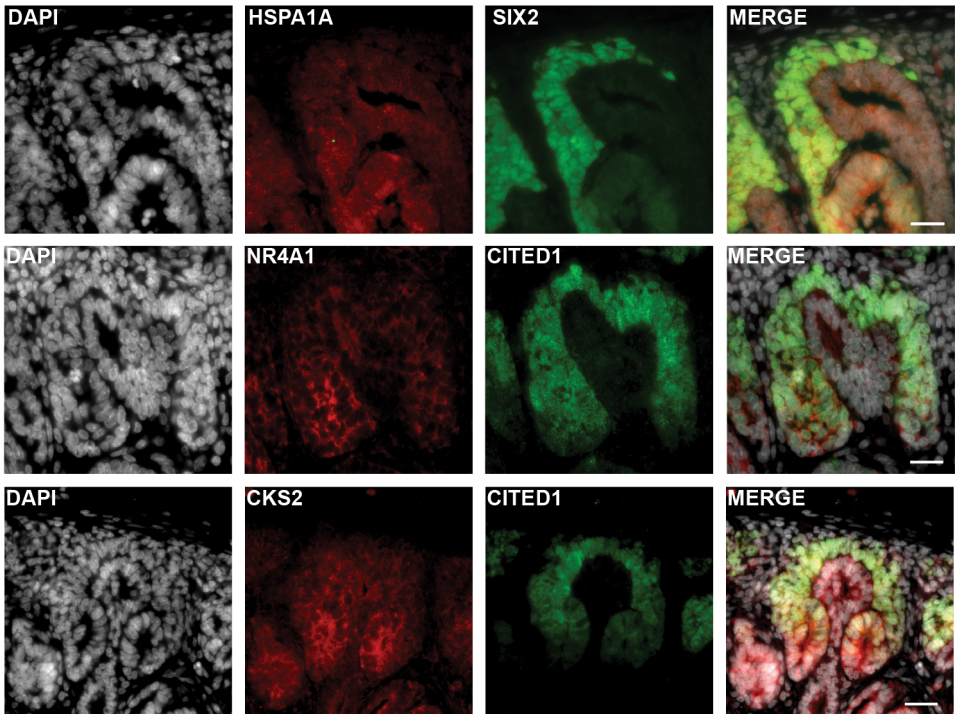


Figure 5.17: Expression variability in the nephrogenic niche. (Representative image of HSPA1A, NR4A1, and CKS2 immunostaining in a w15 human fetal kidney. Scale bar = 20 μ m.)

TMEM100 and *RSPO3* specifically marked NPCa. *RSPO3* is an activator of the canonical WNT signaling pathway [41], suggesting a role of WNT either in NPC self-renewal or UB branching morphogenesis. Notably, all markers of NPCb (*CACYBP*, *MRPL18*, *ZFAND2A*, *DNAJB1*) were related to the stress response in some form. The markers of NPCc (*CRABP2*,

HAS2, *MDK*, *HOXC6*), which were also expressed in the other NPC types, are all either targets of RA or binding it [36, 42, 43]. *MDK* has been shown to be expressed in the CM of the developing rat kidney, and its neutralization reduced the number of formed nephrons *in vitro* [44, 45]. Finally, the NPCd markers *CENPF*, *HMGB2*, *CCNB2*, and *NUSAP1* all have a role in cell cycle regulation or proliferation [28]. *HMGB2* was recently implied in the activation of quiescent adult neural stem cells [46].

The observation that markers of NPCb were related to the stress response seemed to suggest that this cluster was created as an artifact of cell dissociation [17], despite our best efforts to remove stressed cells (see Materials and methods). On the other hand, the vast majority of NPCb cells were classified as *primed NPC II* in the Lindström et al. [13] dataset (5.10, 5.11). The fact that NPCb cells were only detected in the w16 and w18 kidneys is consistent with single-cell dissociation becoming increasingly difficult with fetal age, or alternatively, with a progenitor cell aging phenomenon.

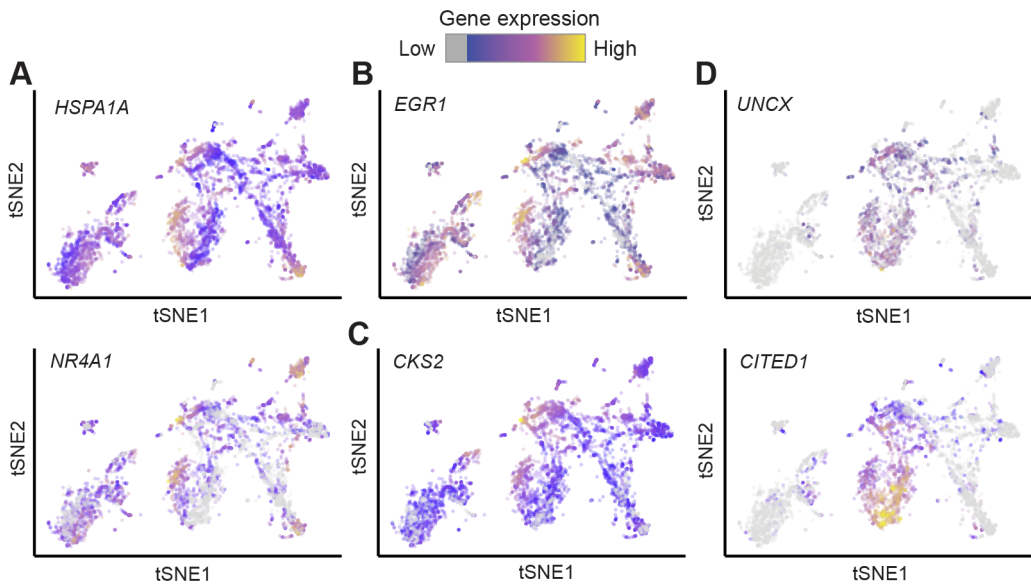


Figure 5.18: Markers of heterogeneity in the nephrogenic niche. (A [top], A [bottom], B, C, D [top] and D [bottom]) tSNE maps showing expression of *HSPA1A*, *NR4A1*, *EGR1*, *CKS2*, *CITED1* and *UNCX* respectively. Expression is indicated by color; expression values of 1 are plotted in gray.

To explore the differences between NPCb and the other NPC clusters further, we immunostained *HSPA1A* and *NR4A1*, both known stress-response genes, in w15 kidney sections (5.17). *HSPA1A* was identified as a marker of NPCb (5.18A, Hochane et al. [1] S3 Table), whereas *NR4A1* was expressed in multiple NPC clusters but highest in NPCb (5.18A). Furthermore, *NR4A1* was also identified in the study by Adam, Potter, and Potter [18] to be up-regulated in response to elevated temperatures during enzymatic dissociation. *HSPA1A* and *NR4A1* were both observable in the nephrogenic niche at the level of

the stalk of the UB and at the transition to the PTA or RV. Additionally, we studied the expression of *EGR1*, another stress-related gene that marked NPCb, with smFISH (5.18B). *EGR1* was mainly found toward the stalk of the UB and in a few cells around the tip of the UB, whereas *SIX2* and *CITED1* transcripts were visible throughout the CM (5.19A). Because results obtained in fixed tissue sections are not confounded by dissociation-related artifacts, these immunostainings and smFISH measurements supported the existence of NPCb cells in the fetal kidney.

The fourth NPC cluster, NPCd, was clearly distinguished by proliferation markers (5.8). To locate NPCd in the tissue, we immunostained the cell cycle regulator and NPCd marker *CKS2* (5.18C). *CKS2* signal could be observed around the stalk of the UB and in RV (5.17). This result supported the interpretation that NPCd were a proliferating transitory state between NPCa-c and PTA.

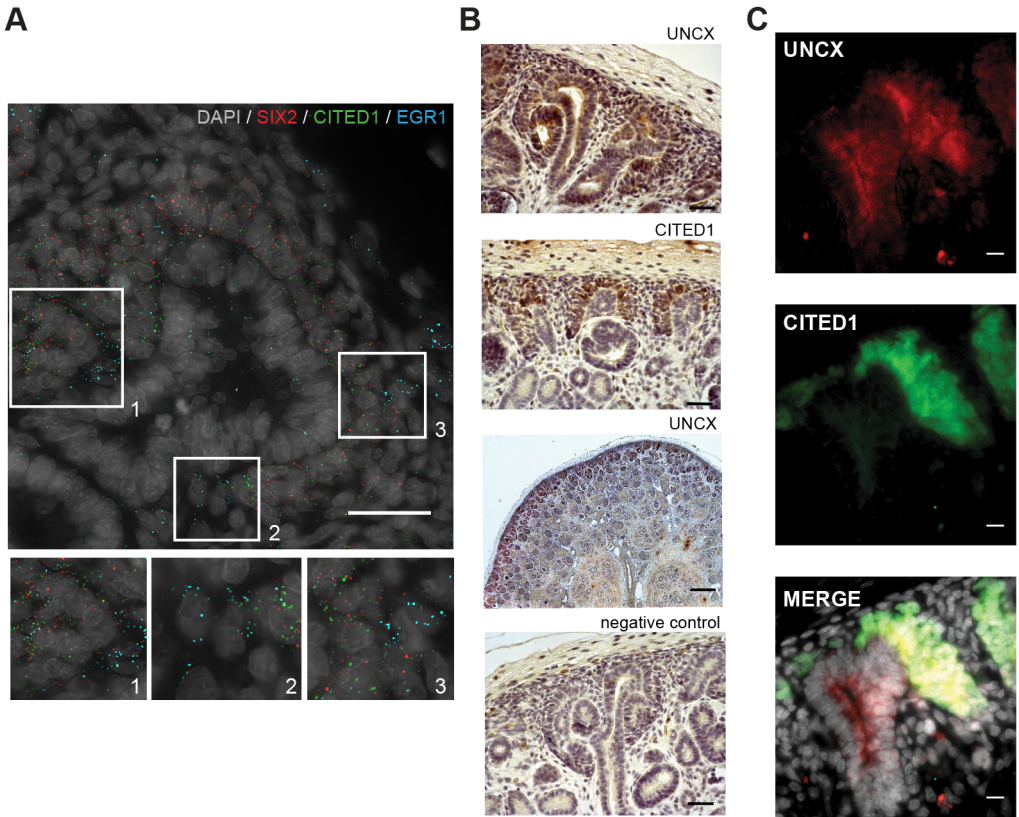


Figure 5.19: In situ expression of NPC markers. (A) smFISH of *SIX2*, *CITED1*, and *EGR1*. The three insets at the bottom correspond to the three areas marked by solid boxes in the main image. Scale bar = 25 μ m. (B) Representative image of UNCX and CITED1 immunostaining. Arrowheads indicate the presence of immunostaining signal. Scale bar = 100 μ m. (C) Immunostaining of CITED1 and UNCX. Scale bar = 10 μ m.

Given the crucial role of the nephrogenic niche in the development of nephrons, it is likely that misexpression or mutation of genes that are specifically expressed in NPC affect kidney function. Mining a database of GWAS revealed that genes that were differentially expressed in NPCs were significantly enriched for association with kidney disease (p -value = $1.7 \cdot 10^{-3}$, one-sided Fisher's exact test). No enrichment was found for lung diseases (p -value = 0.21), one-sided Fisher's exact test (see Materials and methods, 5.20A, Hochane et al. [1] S4 Table). Unsurprisingly, several of the disease-associated genes are known regulators of kidney development, such as *SALL1*, *SOX11*, and *HAS2* [47, 48, 49, 50]. The other identified genes had not been previously associated with kidney development. For example, *DDX1*, which was differentially expressed in NPCs as well as SSBpod, is an RNA helicase that promotes microRNA maturation [51]. *UNCX*, which was broadly expressed in all NPC clusters, is a homeobox transcription factor involved in somitogenesis and neurogenesis [52] and has also been found to be up-regulated in the induced mouse nephrogenic mesenchyme in culture [53]. It was recently associated with renal-function-related traits [54] as well as glomerular filtration rate [55, 56, 57]. In our data, the expression profile of *UNCX* was similar to that of *CITED1* (5.18D). Immunostaining of *UNCX* confirmed the scRNA-seq results and showed expression of *UNCX* in the nephrogenic zone, as marked by *CITED1* (5.19C-D). These findings suggested *UNCX* as a novel potential regulator of early nephrogenesis.

5.3.7. PODOCYTE DEVELOPMENT

Another cell type of high relevance for kidney function is the podocyte. This cell type is critical for filtration and is implied in several forms of kidney disease [38]. podocyte (Pod)ocytes wrap around the glomerular basement membrane (capillary bed) inside Bowman's capsule (5.21A). Clustering (5.3) and pseudotime analysis (5.9) of the w16 kidney dataset had indicated that development into podocytes occurs via a distinct intermediate state that we dubbed SSBpod here. This cell state was likely related or even identical to previously discovered podocyte precursors [13, 23]. In the Lindström et al. [13] dataset, SSBpod and podocytes were both classified as *podocytes*, and the RVCSB were considered *podocyte precursors* (5.11). To show that SSBpod cells were indeed a localizable cell state distinct from podocytes, we further investigated their expression pattern (5.21B), focusing on known literature markers (literature set) and the marker set (Hochane et al. [1] S3 Table). Compared to RVCSB, SSBpod showed higher expression of *MAFB* and *FOXC2*, which are necessary for the determination of podocyte identity [58, 59]. On the other hand, compared to podocyte cells, they exhibited lower expression of genes typically associated with more mature podocytes, like *CLIC5*, *PODXL*, and *PTPRO*. Filtration function-related genes like *NPHS1*, *NPHS2*, and *PTPRO* were expressed at intermediate levels in SSBpod compared to RVCSBb, where they were absent, and podocytes, where they are highly expressed. A similar pattern could be observed for genes associated with podocyte polarization or structural organization as well as pedicel growth and patterning. Finally, podocytes showed the expression of genes that negatively regulate the cell cycle and support long term survival, consistent with their post-mitotic nature [60]. In contrast, SSBpod specifically expressed *ORC4*, which has a func-

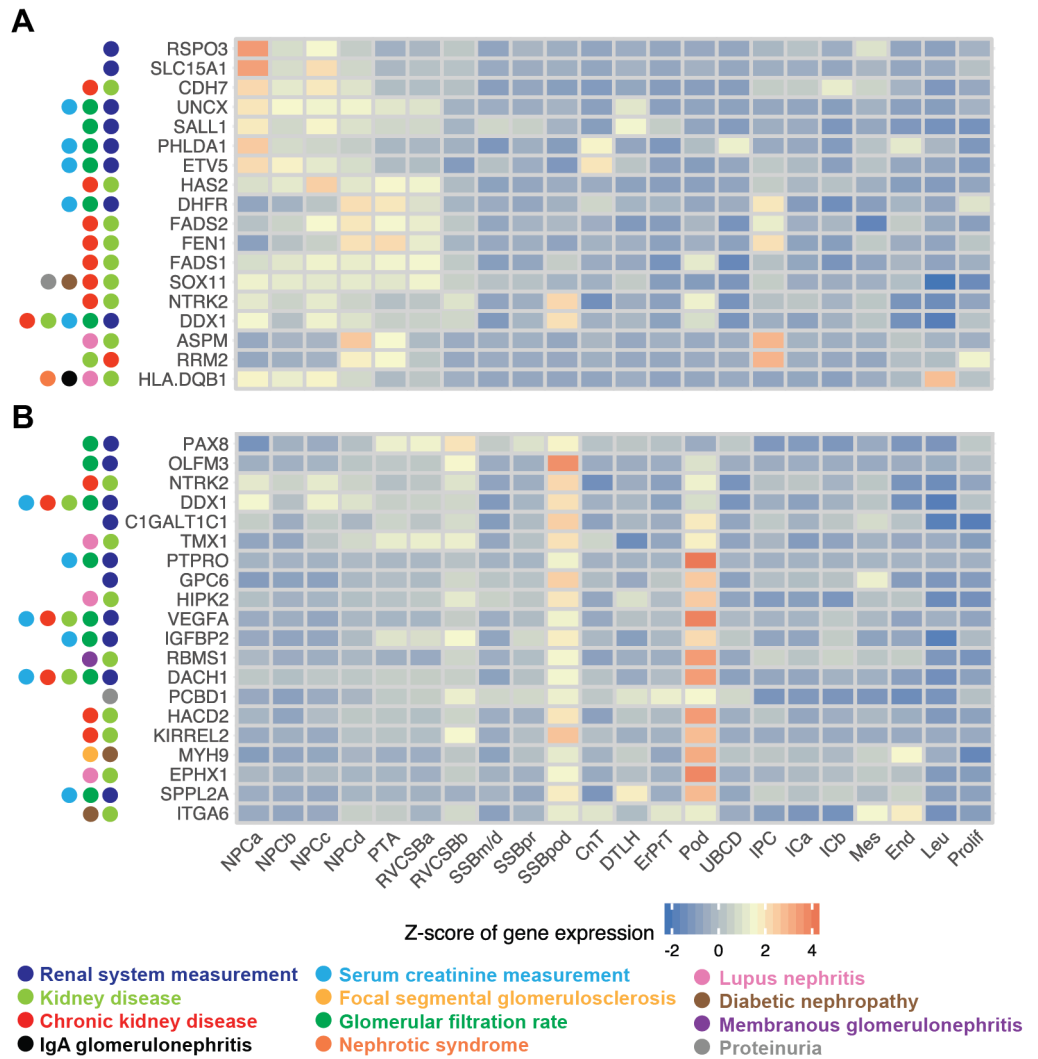


Figure 5.20: Disease-associated genes were specifically expressed in transient cell types. Expression of genes from GWAS traits related to kidney disease. Disease phenotypes associated with these genes are indicated by color; genes were filtered for high expression in cluster(s) of interest relative to all other cell types. Expression was FT transformed, averaged over all cells in a cluster, and standardized gene-wise. (A) Disease-associated genes expressed in early nephron progenitor state (NPC to PTA). (B) Disease-associated genes expressed in SSBpod.

tion in DNA replication. However, proliferation markers were lowly expressed in both SSBpod and podocytes (5.8C), which suggested low proliferative potential in both cell types. In contrast to NPCs, association with kidney disease was not significantly enriched among genes differentially expressed in SSBpod (p -value = 0.1, one-sided Fisher's exact test). One of the disease-associated genes was *OLFM3*, which has been associated with glomerular filtration rate (5.20B) [61]. *OLFM3*, a secreted glycoprotein, has a known function in brain and retina development [62] and has been identified as a marker for podocyte precursors in two independent studies [13, 23]. In our dataset, it was specifically expressed in SSBpod (5.21C) and was a marker for this cell type in the marker set and KeyGenes set (Hochane et al. [1] S3 Table).

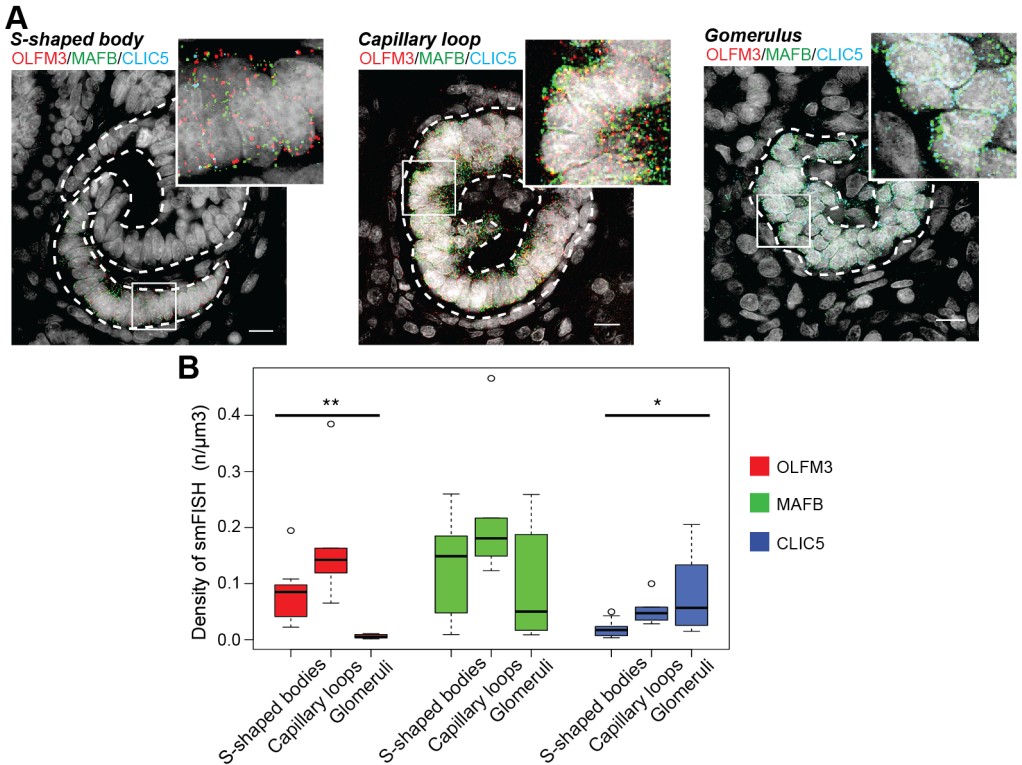


Figure 5.22: SSBpod is localized in the visceral proximal SSB. (A) Representative images of smFISH of *OLFM3*, *MAFB*, and *CLIC5* in SSBpod and Pod. w15 female kidney. Scale bar = 10 μm. (B) Box plots of smFISH signal densities in SSB ($n = 10$), capillary loop ($n = 4$), and glomeruli ($n = 8$), for *OLFM3*, *MAFB* and *CLIC5* (* adjusted $p < 0.05$, ** adjusted $p < 0.0005$).

In order to localize SSBpod, podocytes and mesangial cells in situ, we immunostained w15 kidney sections with antibodies for MAFB, PODXL, and ACTA2 (5.21D). As expected, PODXL and MAFB were found in podocytes at the capillary loop stage and in

more mature glomeruli. MAFB staining extended to the proximal segment of the SSB, which indicated that SSBpod may be part of this structure. To locate the SSBpod cells more precisely, we performed smFISH on *CLIC5* and *MAFB*, expressed both in podocytes and SSBpod (5.22A). We observed a subpopulation of *MAFB*+/*CLIC5*- cells outside the glomeruli, which we identified as the SSBpod. These cells could be found predominantly in the visceral part of the proximal segment of the SSB but also at the capillary loop stage. This result supported the notion that SSBpod were transient cells that preceded (mature) podocytes. Having localized the SSBpod, we next wanted to confirm *OLFM3* as a marker of this cell type. smFISH of *OLFM3*, *MAFB*, and *CLIC5* showed *OLFM3* to be coexpressed with *MAFB* but absent in cells that were positive for *CLIC5* (5.22A), a marker that persists in podocytes in the adult kidney. Quantification of the density of smFISH signals (5.22B) showed that *OLFM3* was absent in glomeruli but could be detected in the subpopulation we identified as SSBpod (*MAFB*+/*CLIC5*-). In summary, these results supported *OLFM3* as a robust marker of podocyte precursors.

Finally, we were wondering whether our dataset would also allow us to identify candidate mechanisms that drive development from SSBpod to podocytes. Differential expression analysis revealed 228 genes that had a significant, bigger than 2-fold changes between SSBpod and podocytes (5.23, Hochane et al. [1] S4 Table). Among these we found factors belonging to multiple signaling pathways, such as *FGF1*, *VEGFA*, *HES1*, and *EGF1*. *Vegfa* and *Fgf1* are known to have a homeostatic function in podocytes [63, 64, 65], whereas *Hes1*, a target of the Notch signaling pathway, seems to be necessary for the synthesis of extracellular matrix proteins in these cells [66]. Binding sites for the transcription factor AP-1 were strongly enriched in this set of genes (145 out of 228 genes, adjusted p -value = $1.3 \cdot 10^{-5}$). *AP-1* would therefore be an interesting target for perturbation studies in mouse models.

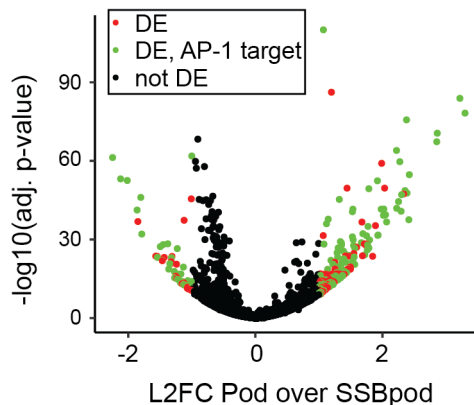


Figure 5.23: AP-1 targets are enriched in the differentially expressed genes of Pod versus SSBpod. Volcano plot of differential gene expression between SSBpod and Pod. \log_2 fold change Pod over SSBpod versus $-\log_{10}$ (adjusted p -value). Genes with an adjusted $p < 0.05$ and \log_2 fold change > 1 were considered significant (colored data points). Genes with an AP-1 binding site are shown in red.

All in all, the results presented here complement other, recent, single-cell transcriptomics studies of the fetal kidney. We demonstrated how the data can be interrogated to find expression patterns that will improve our understanding of human kidney development.

5.4. DISCUSSION

5.4.1. THE NEPHROGENIC NICHE IS HETEROGENEOUS

Building an organ during development requires the careful balance between two fundamental processes—growth and the creation of structure. In many organs, these two functions are reconciled by self-renewing progenitor cells that can be induced to differentiate. In kidney development, NPCs give rise to the epithelium of the nephron, the functional unit of the kidney. To balance growth with patterning, self-renewal and differentiation of NPCs have to be tightly controlled. It is well established that the niche of the NPC plays an important role in this control, but the precise mechanisms are not well understood. In particular, it is not clear how the position and movement of NPCs in the niche might impact the induction towards differentiation.

Heterogeneity in the nephrogenic niche was brought to light first by Mugford et al. [40] in 2009 and has been confirmed by multiple recent studies [11, 13, 23, 24, 25]. Mugford et al. [40] used *in situ* hybridization to study the localization of transcriptional regulators in E15.5 mouse kidney. Three distinct compartments were defined in the CM—inner capping mesenchyme (which lies closest to the cleft of the UB), outer capping mesenchyme (at the tip of the UB), and induced mesenchyme (at the level of the stalk of the UB). Although all compartments express *Six2*, only inner and outer capping mesenchyme express *Cited1*. The induced CM was distinguished by Wnt pathway activity, as evidenced by *Wnt4* expression. Several recent scRNA-seq studies confirmed heterogeneity in the nephrogenic niche. Brunskill et al. [24] studied an E12.5 mouse kidney and found two subpopulations in the CM, which they classified as uninduced (*Six2* positive, *Cited1* positive) and induced (*Six2* positive, *Cited1* negative). Among the hundreds of genes that were differentially expressed between these two populations, they found genes related to the Wnt signaling pathway as well as protein vesicular trafficking and degradation. Wang et al. [25] also found two subclusters in the CM of the human fetal kidney. They interpreted one subcluster as the self-renewing compartment due to higher expression of markers for cell division. The other subcluster, which showed gene expression related to Notch signaling (*HES1*, *HEY1*), was considered induced. Two studies by [11, 13] also explored NPC heterogeneity. The first study [11] identified four NPC clusters (self-renewing, primed, differentiating, and proliferating), whereas the second [13] revealed four clusters of NPCs (I-IV), two clusters of primed NPCs (I-II), as well as several clusters of proliferating cells.

In the dataset presented here, we identified four clusters of NPCs. Among these, NPCa are most likely the self-renewing compartment. In agreement with the studies by Lindström et al. [13] [11, 13], they expressed the highest levels of *CITED1* and *TMEM100*

compared with the other NPCs. Furthermore, they preceded all other NPC clusters in pseudotime analysis. NPCb showed expression of several genes that modulate Notch, BMP, and TGF- β pathway activity, as well as low levels of *LEF1*, which has been shown to indicate induction towards differentiation [11, 13]. The classification of NPCb as *primed NPC* by comparison to the Lindström et al. [13] dataset supported the interpretation of NPCb as a state distinct from NPCa that is primed to differentiate. The fact that we detected NPCb only at w16 and w18 leads us to speculate that NPCb could be the result of continuous changes in the nephrogenic niche over the course of development. The third NPC cluster, NPCc, appeared together with NPCb in pseudotime and was distinguished from the other NPCs by higher expression of genes involved in or regulated by RA signaling. The RA binding protein CRABP2 has been identified as an NPC marker in other reports [11, 25]. We speculate that NPCc are the result of spatially varying concentrations of RA, which is produced in the cortical interstitium [67]. Finally, NPCd appeared between NPCb-c and PTA in pseudotime and were clearly distinguished from the other NPC clusters by increased proliferation, at least as far as that can be inferred from gene expression data. In agreement with our analysis, NPCd were classified as *proliferating cells* by comparison with the Lindström et al. [13] dataset. NPCd cells also lowly expressed markers of induction towards differentiation (such as *LEF1*, *LHX1*, *WNT4*), which indicates a transitory state between induced and/or primed NPCs and PTA. The suggested developmental flow from NPCa via NPCb-c to NPCd was supported by a gradual decrease of *OSR1*, which is a well-known marker of the early CM.

By in situ detection of *CITED1*, *SIX2*, and other genes, we also explored the spatial localization of the different NPC clusters. NPCa seemed to reside closest to the tip of the UB, the induced and/or primed NPCb and c were situated closer to the stalk, and NPCd were closest to the PTA. This finding is consistent with the recent report of NPCs streaming from their niche at the UB tip towards the UB branch point to form the PTA and RV [13]. On this path, the cells gradually lose the NPC transcriptional program, and differentiation is induced. In the mouse, trajectories of NPCs also seem to have a large stochastic component: NPCs repeatedly detach from the UB and attach again and also shuttle back and forth between the *uninduced region* at the UB tip and the *committed region* around the stalk of the UB [68]. This observation could indicate that varying expression levels of genes such as *CITED1* occur as a consequence of cell migration and are not necessarily functionally relevant. Indeed, *Cited1* knockout has no adverse effects on kidney development in the mouse [69]. Taken together, our results support a model in which (self-renewing) NPCa reside at the tip of the UB, probably in close proximity or even in contact with the UB. Movement away from the UB tip, toward the stalk, is accompanied by decreased *CITED1* expression and transformation to the (induced and/or primed) NPCb-c states. Arrival at the stalk of the UB is characterized by the NPCd expression state, increased proliferation, and eventually transformation to the PTA. It is conceivable that cells sometimes visit the different NPC states in reverse order, which would reconcile this model with the observed high, multidirectional motility of NPCs [68].

5.4.2. PROXIMAL-DISTAL PATTERNING

In the prevailing model of mammalian kidney development, self-renewing NPCs are not prepatterned to develop into a certain lineage. When the developing nephron first displays signs of proximal-distal patterning is an important, outstanding question. Muggford et al. [40] have found evidence that the PTA, which succeeds the NPC, is already polarized. A recent study by Lindström et al. [13] proposed an intriguing mechanism that couples temporal and spatial cues: although NPCs that are recruited to the PTA early develop into distal cell types, NPCs that are integrated later contribute to the proximal compartment. In our study, we identified the PTA by known marker genes (*CCND1*, *LHX1*, and *WNT4*) and high proliferation. We were unable to detect any substructure within the PTA, which might be due to the limited resolution of our scRNA-seq method. RV/CSB, the next developmental stage, however, was split in two clusters (a and b). Pseudotime analysis suggested that RVCSBa was a heterogeneous cluster comprising early RV cells (which appeared before RVCSBb) and the distal segment of the RVCSB. This observation is consistent with the time-dependent recruitment model by Lindström et al. [13] in the sense that in that model, distal specification precedes proximal patterning.

5.4.3. FETAL PODOCYTES MAY HAVE VARYING DEGREE OF MATURATION

Single-cell transcriptomics studies of various organs have brought to light many new, intermediate cell states. This has provoked the question of whether we should consider gene expression in complex tissues as a continuum rather than a collection of distinct expression profiles. In developmental systems, it is certainly useful to think about gene expression change as a continuous process. Nevertheless, there are clearly distinct intermediate cell states even within linear developmental paths. In our study, we observed that RVCSBb gave rise to podocytes via an intermediate state, the SSBpod, which directly preceded the podocytes in pseudotime. Specific expression of *OLFM3* made it likely that this cluster is identical to previously identified podocyte precursors, which were marked by this gene [13, 23]. Menon et al. [23] defined a cluster of *immature* or *early* podocytes, characterized by high *OLFM3* and low *MAFB* expression. In that study, podocytes showed increased *MAFB* expression but loss of *OLFM3*. Lindström et al. [13] located *OLFM3* positive cells to the proximal part of the SSB. In our study, we confirmed all of these observations: *OLFM3* was localized to the visceral part of the proximal segment of the SSB, and *OLFM3* negative podocytes showed higher expression of mature podocyte markers compared to the *OLFM3* positive SSBpod. Functional annotation analysis of genes that were differentially expressed between SSBpod and podocytes revealed enrichment of a binding site for AP-1. This transcription factor has been found to be important for the development of the skin [70], neural precursor cells [71], and the heart valve [72] in mice. A role of *AP-1* in kidney development has not been described yet, and further research is needed to elucidate its potential function. The analysis of the kidneys from different gestational ages showed high similarity of cell types across different ages with the exception of podocytes. These displayed a systematic change in expression pattern, which might indicate the continued maturation of podocytes over time. This observation is in agreement with a study by Brunskill et al. [73] in the mouse,

which compared embryonic (E13.5 and E15.5) with adult podocytes (defined as *Mafb* positive cells). That study found hundreds of genes that were differentially expressed between embryonic and adult podocytes. Furthermore, targeted experiments are needed to demonstrate the possible maturation of podocytes in human kidney development.

In summary, we have leveraged a combination of single-cell transcriptomics and in situ imaging to study the intricate structure of the developing human kidney. The transcriptomics data, accessible via a web application <http://www.semraulab.com/kidney>, will be a valuable starting point for discovering gene regulatory mechanisms or finding new disease mechanisms.

DATA AVAILABILITY

The scRNA-seq data have been deposited in the GEO database under accession number GSE114530. An interactive web application accompanying this paper, which provides convenient access to the data, can be found here: <http://www.semraulab.com/kidney>.

ACKNOWLEDGMENTS

We are thankful to Gynaikon Clinic in Rotterdam for their efforts in collecting and providing the fetal material; Susan Kloet and Emile Meijer from the Leiden Genome Technology Center for cell encapsulation, library preparation, single-cell sequencing, primary data mapping, and quality control; Vanessa Torrens-Juaneda and Ioannis Moustakas for primary and secondary data analysis and discussions; GenomseScan for technical support; and the group of Andrew D. Smith for providing the cluster identities of the Lindström et al. dataset.

5.5. SUPPLEMENTARY INFORMATION

Abbreviation	Cell type	Gene	Reference
NPCs	nephron progenitor cells	OSR1	[47]
		SIX1	[74]
		SIX2	[75, 76]
		CITED1	[77, 78]
		EYA1	[79]
		SALL1	[80, 69]
		MEOX1	[3]
		GDNF	[81]
		ETV4	[82]
		COL2A1	[83]
		HES1	[84, 78]
		CRABP2	[13]
LEF1	[13]		
ITGA8	[23]		
PTA	pretubular aggregate	LHX1	[13]
		WNT4	[85]
		CCND1	[86]
RVCSB a and b	renal vesicle/comma-shaped body	LHX1	[87]
		PAX8	[85]
		JAG1	[88]
		PAX2	[53]
		WNT4	[89]
		SFRP2	[90]
DLL1	[4]		
SSBm/d	SSB medial/distal	HNF1B	[91]
		POU3F3	[92]
		SIM2	[93]
		SOX9	[13]
		IRX2	[23]
SSBpr	SSB proximal precursor cell	CDH6	[94, 95]
		HNF1A	[96, 97]
		AMN	[93]
SSBpod	SSB podocyte precursor cell	FOXC2	[98]
		MAFB	[59]
		OLFM3	[13]
		WT1	[99]
Pod	podocytes	PTPRO	[100]
		NPHS2	[101]
		NPHS1	[102]
		PODXL	[103]
		TGFBR3	[103]
		CLIC5	[104]
		CITED2	[98]
DTLH	distal tubule/loop of Henle	PAPPA2	[105]
		MAL	[106]
		CLCN5	[107]
		SLC12A3	[108]
		UMOD	[13]
		SLC12A1	[109]
		LRP2	[110]
		ANPEP	[111]

Continued on next page

ErPrT early proximal tubule

<i>Continued from previous page</i>			
Abbreviation	Cell type	Gene	Reference
		SLC34A1 CLDN1 CLDN2 SLC13A1 CUBN	[112] [113] [114] [115] [116]
CnT	connecting tubule	ALDH1A1 TACSTD2 CDH1	[13] [13] [117]
IPC	interstitial progenitor cell	GDNF FOXD1	[118] [119]
ICs a and b	interstitial cells a and b	DES COL3A1 COL1A1 SERPINE2 FGF7 LEF1 DCN	[120] [121] [122] [123] [53, 124] [125] [126]
Mes	mesangial cells	ACTA2 TPM2 PDGFRB MCAM CSPG4 CD248	[123] [127] [128] [123] [123] [129]
UBCD	ureteric bud/collecting duct	KRT18 KRT8 RET CLDN7 AQP2 GATA2 MMP7 CALB1	[130] [130] [131] [132] [133] [133] [134] [117]
End	endothelial cell	KDR TEK FLT1 CDH5 PECAM1	[135] [136] [135] [137] [138]
Leu	leukocytes	CD37 CD48 ITGB2 IFI30 IL1B	[139] [140] [141] [142] [143]

Table 5.1: Literature marker set and references.

ACRONYMS

AUROC area under the ROC	PCR polymerase chain reaction
BSA bovine serum albumin	RA retinoic acid
COI cluster of interest	RET ret proto-oncogene
FISH fluorescence in situ hybridization	ROC receiver operating characteristic
FT Freeman-Tukey	ROI region of interest
GDNF glial cell-derived neurotrophic factor	RT room temperature
GO gene ontology	scrRNA-seq single-cell RNA sequencing
GWAS genome-wide association studies	smFISH single-molecule FISH
HVG highly variable gene	tSNE t-distributed stochastic neighbor embedding
knn k-nearest neighbors	UMI unique molecular identifier
PC principal component	w16 week 16

5

CELL TYPES

CM cap mesenchyme	NPC nephron progenitor cell
CnT connecting tubule	Pod podocyte
CSB comma-shaped body	PTA pretubular aggregate
DTLH distal tubule/loop of Henle	RV renal vesicle
End endothelial cell	RVCSB renal vesicle/comma-shaped body
ErPrT early proximal tubule	SSB s-shaped body
IC interstitial cell	SSBm/d SSB medial/distal
IPC interstitial progenitor cell	SSBpod SSB podocyte precursor cell
Leu leukocytes	SSBpr SSB proximal precursor cell
LOH loop of Henle	UB ureteric bud
Mes mesangial cells	UBCD ureteric bud/collecting duct
MM metanephric mesenchyme	

5.6. REFERENCES

- [1] Mazène Hochane et al. “Single-cell transcriptomics reveals gene expression dynamics of human fetal kidney development”. In: *PLOS Biology* 17.2 (Feb. 2019), e3000152.
- [2] Frank Costantini and Raphael Kopan. “Patterning a Complex Organ: Branching Morphogenesis and Nephron Segmentation in Kidney Development”. In: *Developmental Cell* 18.5 (May 2010), pp. 698–712. ISSN: 1534-5807. DOI: 10.1016/j.devcel.2010.04.008.
- [3] Alexander N. Combes et al. “Haploinsufficiency for the Six2 gene increases nephron progenitor proliferation promoting branching and nephron number”. English. In: *Kidney International* 93.3 (Mar. 2018), pp. 589–598. ISSN: 0085-2538. DOI: 10.1016/j.kint.2017.09.015.
- [4] Melissa H. Little and Andrew P. McMahon. “Mammalian kidney development: principles, progress, and projections”. eng. In: *Cold Spring Harb Perspect Biol* 4.5 (May 2012). ISSN: 1943-0264. DOI: 10.1101/cshperspect.a008300.
- [5] John F. Bertram et al. “Why and how we determine nephron number”. eng. In: *Pediatr. Nephrol.* 29.4 (Apr. 2014), pp. 575–580. ISSN: 1432-198X. DOI: 10.1007/s00467-013-2600-y.
- [6] Frank Costantini. “Genetic controls and cellular behaviors in branching morphogenesis of the renal collecting system”. eng. In: *Wiley Interdiscip Rev Dev Biol* 1.5 (Oct. 2012), pp. 693–713. ISSN: 1759-7692. DOI: 10.1002/wdev.52.
- [7] Akio Kobayashi et al. “Identification of a multipotent self-renewing stromal progenitor population during mammalian kidney organogenesis”. eng. In: *Stem Cell Reports* 3.4 (Oct. 2014), pp. 650–662. ISSN: 2213-6711. DOI: 10.1016/j.stemcr.2014.08.008.
- [8] Maria Luisa S. Sequeira-Lopez et al. “The earliest metanephric arteriolar progenitors and their role in kidney vascular development”. eng. In: *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 308.2 (Jan. 2015), R138–149. ISSN: 1522-1490. DOI: 10.1152/ajpregu.00428.2014.
- [9] B. Robert et al. “Evidence that embryonic kidney cells expressing flk-1 are intrinsic, vasculogenic angioblasts”. eng. In: *Am. J. Physiol.* 271.3 Pt 2 (Sept. 1996), F744–753. ISSN: 0002-9513. DOI: 10.1152/ajprenal.1996.271.3.F744.
- [10] Nils O. Lindström et al. “Conserved and Divergent Features of Human and Mouse Kidney Organogenesis”. en. In: *JASN* (Feb. 2018), ASN.2017080887. ISSN: 1046-6673, 1533-3450. DOI: 10.1681/ASN.2017080887.
- [11] Nils O. Lindström et al. “Conserved and Divergent Features of Mesenchymal Progenitor Cell Types within the Cortical Nephrogenic Niche of the Human and Mouse Kidney”. en. In: *JASN* (Feb. 2018), ASN.2017080890. ISSN: 1046-6673, 1533-3450. DOI: 10.1681/ASN.2017080890.

- [12] Nils O. Lindström et al. “Conserved and Divergent Molecular and Anatomic Features of Human and Mouse Nephron Patterning”. en. In: *JASN* (Feb. 2018), ASN.2017091036. ISSN: 1046-6673, 1533-3450. DOI: 10.1681/ASN.2017091036.
- [13] Nils O Lindström et al. “Progressive Recruitment of Mesenchymal Progenitors Reveals a Time-Dependent Process of Cell Fate Acquisition in Mouse and Human Nephrogenesis”. In: *Developmental Cell* 45.5 (June 2018), 651–660.e4.
- [14] A Heeren et al. “Development of the follicular basement membrane during human gametogenesis and early folliculogenesis”. In: *BMC Developmental Biology* 15.1 (2015), p. 4.
- [15] Ábel Vértesy et al. “Parental haplotype-specific single-cell transcriptomics reveal incomplete epigenetic reprogramming in human female germ cells”. eng. In: *Nat Commun* 9.1 (2018), p. 1873. ISSN: 2041-1723. DOI: 10.1038/s41467-018-04215-7.
- [16] Stefan Semrau et al. “FuseFISH: robust detection of transcribed gene fusions in single cells”. eng. In: *Cell Rep* 6.1 (Jan. 2014), pp. 18–23. ISSN: 2211-1247. DOI: 10.1016/j.celrep.2013.12.002.
- [17] Susanne C. van den Brink et al. “Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations”. en. In: *Nature Methods* 14.10 (Oct. 2017), pp. 935–936. ISSN: 1548-7105. DOI: 10.1038/nmeth.4437.
- [18] Mike Adam, Andrew S. Potter, and S. Steven Potter. “Psychrophilic proteases dramatically reduce single-cell RNA-seq artifacts: a molecular atlas of kidney development”. eng. In: *Development* 144.19 (2017), pp. 3625–3632. ISSN: 1477-9129. DOI: 10.1242/dev.151142.
- [19] Steffen Durinck et al. “Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt”. eng. In: *Nat Protoc* 4.8 (2009), pp. 1184–1191. ISSN: 1750-2799. DOI: 10.1038/nprot.2009.97.
- [20] Florian Wagner, Yun Yan, and Itai Yanai. “K-nearest neighbor smoothing for high-throughput single-cell RNA-Seq data”. en. In: *bioRxiv* (Apr. 2018), p. 217737. DOI: 10.1101/217737.
- [21] Aaron T. L. Lun, Karsten Bach, and John C. Marioni. “Pooling across cells to normalize single-cell RNA sequencing data with many zero counts”. eng. In: *Genome Biol.* 17 (Apr. 2016), p. 75. ISSN: 1474-760X. DOI: 10.1186/s13059-016-0947-7.
- [22] Andrew P. McMahon. “Development of the Mammalian Kidney”. eng. In: *Curr. Top. Dev. Biol.* 117 (2016), pp. 31–64. ISSN: 1557-8933. DOI: 10.1016/bs.ctdb.2015.10.010.
- [23] Rajasree Menon et al. “Single-cell analysis of progenitor cell dynamics and lineage specification in the human fetal kidney”. In: *Development* 145.16 (Aug. 2018). ISSN: 0950-1991. DOI: 10.1242/dev.164038.
- [24] Eric W. Brunskill et al. “Single cell dissection of early kidney development: multilineage priming”. en. In: *Development* 141.15 (Aug. 2014), pp. 3093–3101. ISSN: 0950-1991, 1477-9129. DOI: 10.1242/dev.110601.

- [25] Ping Wang et al. “Dissecting the Global Dynamic Molecular Profiles of Human Fetal Kidney Development by Single-Cell RNA Sequencing”. eng. In: *Cell Rep* 24.13 (Sept. 2018), 3554–3567.e3. ISSN: 2211-1247. DOI: 10 . 1016 / j . celrep . 2018 . 08 . 056.
- [26] Laleh Haghverdi et al. “Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors”. eng. In: *Nat. Biotechnol.* 36.5 (2018), pp. 421–427. ISSN: 1546-1696. DOI: 10 . 1038 / nbt . 4091.
- [27] Antonio Scialdone et al. “Computational assignment of cell-cycle stage from single-cell transcriptome data”. eng. In: *Methods* 85 (Sept. 2015), pp. 54–61. ISSN: 1095-9130. DOI: 10 . 1016 / j . ymeth . 2015 . 06 . 021.
- [28] Michael L. Whitfield et al. “Common markers of proliferation”. eng. In: *Nat. Rev. Cancer* 6.2 (Feb. 2006), pp. 99–106. ISSN: 1474-175X. DOI: 10 . 1038 / nrc1802.
- [29] Xiaojie Qiu et al. “Reversed graph embedding resolves complex single-cell trajectories”. eng. In: *Nat. Methods* 14.10 (Oct. 2017), pp. 979–982. ISSN: 1548-7105. DOI: 10 . 1038 / nmeth . 4402.
- [30] Qi Mao et al. “Dimensionality Reduction Via Graph Structure Learning”. In: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery, 2015, pp. 765–774. ISBN: 9781450336642. DOI: 10 . 1145 / 2783258 . 2783309.
- [31] Matthias S. Roost et al. “KeyGenes, a Tool to Probe Tissue Differentiation Using a Human Fetal Transcriptional Atlas”. In: *Stem Cell Reports* 4.6 (May 2015), pp. 1112–1124. ISSN: 2213-6711. DOI: 10 . 1016 / j . stemcr . 2015 . 05 . 002.
- [32] Mark D. Robinson, Davis J. McCarthy, and Gordon K. Smyth. “edgeR: a Bioconductor package for differential expression analysis of digital gene expression data”. eng. In: *Bioinformatics* 26.1 (Jan. 2010), pp. 139–140. ISSN: 1367-4811. DOI: 10 . 1093 / bioinformatics / btp616.
- [33] Da Wei Huang, Brad T. Sherman, and Richard A. Lempicki. “Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources”. en. In: *Nature Protocols* 4.1 (Jan. 2009), pp. 44–57. ISSN: 1750-2799. DOI: 10 . 1038 / nprot . 2008 . 211.
- [34] Caroline A. Schneider, Wayne S. Rasband, and Kevin W. Eliceiri. “NIH Image to ImageJ: 25 years of image analysis”. eng. In: *Nat. Methods* 9.7 (July 2012), pp. 671–675. ISSN: 1548-7105.
- [35] Laurens van der Maaten and Geoffrey Hinton. “Visualizing Data using t-SNE”. In: *Journal of Machine Learning Research* 9.Nov (2008), pp. 2579–2605. ISSN: ISSN 1533-7928.
- [36] Joseph L. Napoli. “Cellular retinoid binding-proteins, CRBP, CRABP, FABP5: Effects on retinoid metabolism, function and related diseases”. eng. In: *Pharmacol. Ther.* 173 (May 2017), pp. 19–33. ISSN: 1879-016X. DOI: 10 . 1016 / j . pharmthera . 2017 . 01 . 004.

- [37] J M Linton, G R Martin, and L F Reichardt. “The ECM protein nephronectin promotes kidney development via integrin $\alpha 8$ 1-mediated stimulation of Gdnf expression”. In: *Development* 134.13 (May 2007), pp. 2501–2509.
- [38] Janina Müller-Deile and Mario Schiffer. “Podocytes from the diagnostic and therapeutic point of view”. eng. In: *Pflugers Arch.* 469.7-8 (Aug. 2017), pp. 1007–1015. ISSN: 1432-2013. DOI: 10.1007/s00424-017-1993-z.
- [39] Liliana Schaefer et al. “Small proteoglycans of normal adult human kidney: Distinct expression patterns of decorin, biglycan, fibromodulin, and lumican”. In: *Kidney International* 58.4 (Oct. 2000), pp. 1557–1568. ISSN: 0085-2538. DOI: 10.1046/j.1523-1755.2000.00317.x.
- [40] Joshua W. Mugford et al. “High-resolution gene expression analysis of the developing mouse kidney defines novel cellular compartments within the nephron progenitor population”. eng. In: *Dev. Biol.* 333.2 (Sept. 2009), pp. 312–323. ISSN: 1095-564X. DOI: 10.1016/j.ydbio.2009.06.043.
- [41] Kyung-Ah Kim et al. “R-Spondin Family Members Regulate the Wnt Pathway by a Common Mechanism”. In: *MBoC* 19.6 (Apr. 2008), pp. 2588–2596. ISSN: 1059-1524. DOI: 10.1091/mbc.e08-02-0187.
- [42] C. Pedraza, S. Matsubara, and T. Muramatsu. “A retinoic acid-responsive element in human midkine gene”. eng. In: *J. Biochem.* 117.4 (Apr. 1995), pp. 845–849. ISSN: 0021-924X.
- [43] Katri M. Makkonen et al. “Regulation of the hyaluronan synthase 2 gene by convergence in cyclic AMP response element-binding protein and retinoid acid receptor signaling”. eng. In: *J. Biol. Chem.* 284.27 (July 2009), pp. 18270–18281. ISSN: 1083-351X. DOI: 10.1074/jbc.M109.012492.
- [44] José Vilar et al. “Midkine is involved in kidney development and in its regulation by retinoids”. eng. In: *J. Am. Soc. Nephrol.* 13.3 (Mar. 2002), pp. 668–676. ISSN: 1046-6673.
- [45] Waichi Sato et al. “Midkine expression in the course of nephrogenesis and its role in ischaemic reperfusion injury”. eng. In: *Nephrol. Dial. Transplant.* 17 Suppl 9 (2002), pp. 52–54. ISSN: 0931-0509.
- [46] Ayaka Kimura et al. “HMGB2 expression is associated with transition from a quiescent to an activated state of adult neural stem cells”. eng. In: *Dev. Dyn.* 247.1 (2018), pp. 229–238. ISSN: 1097-0177. DOI: 10.1002/dvdy.24559.
- [47] Joshua W. Mugford et al. “Osr1 expression demarcates a multi-potent population of intermediate mesoderm that undergoes progressive restriction to an Osr1-dependent nephron progenitor compartment within the mammalian kidney”. eng. In: *Dev. Biol.* 324.1 (Dec. 2008), pp. 88–98. ISSN: 1095-564X. DOI: 10.1016/j.ydbio.2008.09.010.
- [48] Yasmine Neirijnck et al. “Sox11 gene disruption causes congenital anomalies of the kidney and urinary tract (CAKUT)”. eng. In: *Kidney Int.* (Feb. 2018). ISSN: 1523-1755. DOI: 10.1016/j.kint.2017.11.026.

- [49] Priscilla Soulié et al. “Spatially restricted hyaluronan production by Has2 drives epithelial tubulogenesis in vitro”. eng. In: *Am. J. Physiol., Cell Physiol.* 307.8 (Oct. 2014), pp. C745–759. ISSN: 1522-1563. DOI: 10.1152/ajpcell.00047.2014.
- [50] Tomoko Ohmori et al. “Sall1 in renal stromal progenitors non-cell autonomously restricts the excessive expansion of nephron progenitors”. eng. In: *Sci Rep* 5 (Oct. 2015), p. 15676. ISSN: 2045-2322. DOI: 10.1038/srep15676.
- [51] Cecil Han et al. “The RNA-binding protein DDX1 promotes primary microRNA maturation and inhibits ovarian tumor progression”. eng. In: *Cell Rep* 8.5 (Sept. 2014), pp. 1447–1460. ISSN: 2211-1247. DOI: 10.1016/j.celrep.2014.07.058.
- [52] Neeraja Sammeta, Debra L. Hardin, and Timothy S. McClintock. “Uncx regulates proliferation of neural progenitor cells and neuronal survival in the olfactory epithelium”. eng. In: *Mol. Cell. Neurosci.* 45.4 (Dec. 2010), pp. 398–407. ISSN: 1095-9327. DOI: 10.1016/j.mcn.2010.07.013.
- [53] Minoru Takasato et al. “Identification of kidney mesenchymal genes by a combination of microarray analysis and Sall1-GFP knockin mice”. eng. In: *Mech. Dev.* 121.6 (June 2004), pp. 547–557. ISSN: 0925-4773. DOI: 10.1016/j.mod.2004.04.007.
- [54] Yukinori Okada et al. “Meta-analysis identifies multiple loci associated with kidney function-related traits in east Asian populations”. eng. In: *Nat. Genet.* 44.8 (July 2012), pp. 904–909. ISSN: 1546-1718. DOI: 10.1038/ng.2352.
- [55] Mathias Gorski et al. “1000 Genomes-based meta-analysis identifies 10 novel loci for kidney function”. eng. In: *Sci Rep* 7 (2017), p. 45040. ISSN: 2045-2322. DOI: 10.1038/srep45040.
- [56] Anubha Mahajan et al. “Trans-ethnic Fine Mapping Highlights Kidney-Function Genes Linked to Salt Sensitivity”. eng. In: *Am. J. Hum. Genet.* 99.3 (Sept. 2016), pp. 636–646. ISSN: 1537-6605. DOI: 10.1016/j.ajhg.2016.07.012.
- [57] Cristian Pattaro et al. “Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function”. eng. In: *Nat Commun* 7 (Jan. 2016), p. 10023. ISSN: 2041-1723. DOI: 10.1038/ncomms10023.
- [58] Masaru Motojima, Tsutomu Kume, and Taiji Matsusaka. “Foxc1 and Foxc2 are necessary to maintain glomerular podocytes”. eng. In: *Exp. Cell Res.* 352.2 (Mar. 2017), pp. 265–272. ISSN: 1090-2422. DOI: 10.1016/j.yexcr.2017.02.016.
- [59] Virginia Sadl et al. “The mouse Kreisler (Krrml1/MafB) segmentation gene is required for differentiation of glomerular visceral epithelial cells”. eng. In: *Dev. Biol.* 249.1 (Sept. 2002), pp. 16–29. ISSN: 0012-1606.
- [60] Kevin V. Lemley. “Mechanical challenges to the glomerulus and podocyte loss: evolution of a paradigm”. eng. In: *Pflugers Arch.* 469.7-8 (Aug. 2017), pp. 959–963. ISSN: 1432-2013. DOI: 10.1007/s00424-017-2012-0.
- [61] S. Nanayakkara et al. “An integrative study of the genetic, social and environmental determinants of chronic kidney disease characterized by tubulointerstitial damages in the North Central Region of Sri Lanka.” eng. In: *J Occup Health* 56.1 (2014), pp. 28–38. ISSN: 1341-9145. DOI: 10.1539/joh.13-0172-0A.

- [62] T. A. Moreno and M. Bronner-Fraser. “The secreted glycoprotein Noelin-1 promotes neurogenesis in *Xenopus*”. eng. In: *Dev. Biol.* 240.2 (Dec. 2001), pp. 340–360. ISSN: 0012-1606. DOI: 10.1006/dbio.2001.0472.
- [63] Diangeng Li et al. “Mesenchymal stem cells protect podocytes from apoptosis induced by high glucose via secretion of epithelial growth factor”. eng. In: *Stem Cell Res Ther* 4.5 (2013), p. 103. ISSN: 1757-6512. DOI: 10.1186/scrt314.
- [64] Guang Liang et al. “Fibroblast growth factor 1 ameliorates diabetic nephropathy by an anti-inflammatory mechanism”. eng. In: *Kidney Int.* 93.1 (2018), pp. 95–109. ISSN: 1523-1755. DOI: 10.1016/j.kint.2017.05.013.
- [65] Christina S. Bartlett, Marie Jeansson, and Susan E. Quaggin. “Vascular Growth Factors and Glomerular Disease”. eng. In: *Annu. Rev. Physiol.* 78 (2016), pp. 437–461. ISSN: 1545-1585. DOI: 10.1146/annurev-physiol-021115-105412.
- [66] Min Yao et al. “The Notch pathway mediates the angiotensin II-induced synthesis of extracellular matrix components in podocytes”. eng. In: *Int. J. Mol. Med.* 36.1 (July 2015), pp. 294–300. ISSN: 1791-244X. DOI: 10.3892/ijmm.2015.2193.
- [67] Carolina Rosselot et al. “Non-cell-autonomous retinoid signaling is crucial for renal development”. In: *Development* 137.2 (Jan. 2010), pp. 283–292. ISSN: 0950-1991. DOI: 10.1242/dev.040287.
- [68] Alexander N. Combes et al. “Cap mesenchyme cell swarming during kidney development is influenced by attraction, repulsion, and adhesion to the ureteric tip”. eng. In: *Dev. Biol.* 418.2 (2016), pp. 297–306. ISSN: 1095-564X. DOI: 10.1016/j.ydbio.2016.06.028.
- [69] Scott Boyle et al. “Cited1 and Cited2 are differentially expressed in the developing kidney but are not required for nephrogenesis”. en. In: *Dev. Dyn.* 236.8 (Aug. 2007), pp. 2321–2330. ISSN: 1097-0177. DOI: 10.1002/dvdy.21242.
- [70] Christina A. Young et al. “Embryonic AP1 Transcription Factor Deficiency Causes a Collodion Baby-Like Phenotype”. eng. In: *J. Invest. Dermatol.* 137.9 (2017), pp. 1868–1877. ISSN: 1523-1747. DOI: 10.1016/j.jid.2017.04.032.
- [71] Fumiaki Kawashima et al. “c-jun is differentially expressed in embryonic and adult neural precursor cells”. eng. In: *Histochem. Cell Biol.* 147.6 (June 2017), pp. 721–731. ISSN: 1432-119X. DOI: 10.1007/s00418-016-1536-2.
- [72] Victoria C. Garside et al. “SOX9 modulates the expression of key transcription factors required for heart valve development”. eng. In: *Development* 142.24 (Dec. 2015), pp. 4340–4350. ISSN: 1477-9129. DOI: 10.1242/dev.125252.
- [73] Eric W. Brunskill et al. “Defining the molecular character of the developing and adult kidney podocyte”. eng. In: *PLoS ONE* 6.9 (2011), e24640. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0024640.
- [74] Pin-Xian Xu et al. “Six1 is required for the early organogenesis of mammalian kidney”. eng. In: *Development* 130.14 (July 2003), pp. 3085–3094. ISSN: 0950-1991.
- [75] Michael Marcotte, Richa Sharma, and Maxime Bouchard. “Gene regulatory network of renal primordium development”. en. In: *Pediatr Nephrol* 29.4 (Apr. 2014), pp. 637–644. ISSN: 0931-041X, 1432-198X. DOI: 10.1007/s00467-013-2635-0.

- [76] Akio Kobayashi et al. "Six2 defines and regulates a multipotent self-renewing nephron progenitor population throughout mammalian kidney development". eng. In: *Cell Stem Cell* 3.2 (Aug. 2008), pp. 169–181. ISSN: 1875-9777. DOI: 10.1016/j.stem.2008.05.020.
- [77] Scott Boyle et al. "Fate mapping using Cited1-CreERT2 mice demonstrates that the cap mesenchyme contains self-renewing progenitor cells and gives rise exclusively to nephronic epithelia". eng. In: *Dev. Biol.* 313.1 (Jan. 2008), pp. 234–245. ISSN: 1095-564X. DOI: 10.1016/j.ydbio.2007.10.014.
- [78] Sayoko Fujimura et al. "Notch2 activation in the embryonic kidney depletes nephron progenitors". eng. In: *J. Am. Soc. Nephrol.* 21.5 (May 2010), pp. 803–810. ISSN: 1533-3450. DOI: 10.1681/ASN.2009040353.
- [79] P. X. Xu et al. "Eya1-deficient mice lack ears and kidneys and show abnormal apoptosis of organ primordia". eng. In: *Nat. Genet.* 23.1 (Sept. 1999), pp. 113–117. ISSN: 1061-4036. DOI: 10.1038/12722.
- [80] Ryuichi Nishinakamura et al. "Murine homolog of SALL1 is essential for ureteric bud invasion in kidney development". In: *Development* 128.16 (2001), pp. 3105–3115. ISSN: 0950-1991.
- [81] Alexander N. Combes, Jamie A. Davies, and Melissa H. Little. "Cell-cell interactions driving kidney morphogenesis". eng. In: *Curr. Top. Dev. Biol.* 112 (2015), pp. 467–508. ISSN: 1557-8933. DOI: 10.1016/bs.ctdb.2014.12.002.
- [82] Benson C. Lu et al. "Etv4 and Etv5 are required downstream of GDNF and Ret for kidney branching morphogenesis". eng. In: *Nat. Genet.* 41.12 (Dec. 2009), pp. 1295–1302. ISSN: 1546-1718. DOI: 10.1038/ng.476.
- [83] Masahito Miyamoto et al. "In-depth proteomic profiling of the normal human kidney glomerulus using two-dimensional protein prefractionation in combination with liquid chromatography-tandem mass spectrometry". eng. In: *J. Proteome Res.* 6.9 (Sept. 2007), pp. 3680–3690. ISSN: 1535-3893. DOI: 10.1021/pr070203n.
- [84] Sophie Jarriault et al. "Signalling downstream of activated mammalian Notch". en. In: *Nature* 377.6547 (Sept. 1995), pp. 355–358. ISSN: 1476-4687. DOI: 10.1038/377355a0.
- [85] K. Stark et al. "Epithelial transformation of metanephric mesenchyme in the developing kidney regulated by Wnt-4". eng. In: *Nature* 372.6507 (Dec. 1994), pp. 679–683. ISSN: 0028-0836. DOI: 10.1038/372679a0.
- [86] Thomas F. Gallegos et al. "A Protein Kinase A and Wnt-dependent network regulating an intermediate stage in epithelial tubulogenesis during kidney development". In: *Dev Biol* 364.1 (Apr. 2012), pp. 11–21. ISSN: 0012-1606. DOI: 10.1016/j.ydbio.2012.01.014.
- [87] Kimmo Halt and Seppo Vainio. "Coordination of kidney organogenesis by Wnt signaling". In: *Pediatr Nephrol* 29.4 (2014), pp. 737–744. ISSN: 0931-041X. DOI: 10.1007/s00467-013-2733-z.

- [88] Madhulika Sharma et al. “Coexpression of Cux-1 and Notch signaling pathway components during kidney development”. eng. In: *Dev. Dyn.* 231.4 (Dec. 2004), pp. 828–838. ISSN: 1058-8388. DOI: 10.1002/dvdy.20175.
- [89] S. Steven Potter et al. “Laser capture-microarray analysis of Lim1 mutant kidney development”. eng. In: *Genesis* 45.7 (July 2007), pp. 432–439. ISSN: 1526-954X. DOI: 10.1002/dvg.20309.
- [90] Cornelia Leimeister, Alexandra Bach, and Manfred Gessler. “Developmental expression patterns of mouse sFRP genes encoding members of the secreted frizzled related protein family”. In: *Mechanisms of Development* 75.1 (July 1998), pp. 29–42. ISSN: 0925-4773. DOI: 10.1016/S0925-4773(98)00072-0.
- [91] Claire Heliot et al. “HNF1B controls proximal-intermediate nephron segment identity in vertebrates by regulating Notch signalling components and Irx1/2”. In: *Development* 140.4 (2013), pp. 873–885. ISSN: 0950-1991. DOI: 10.1242/dev.086538. eprint: <https://dev.biologists.org/content/140/4/873.full.pdf>.
- [92] Alexandra Rieger et al. “Missense Mutation of POU Domain Class 3 Transcription Factor 3 in Pou3f3L423P Mice Causes Reduced Nephron Number and Impaired Development of the Thick Ascending Limb of the Loop of Henle”. en. In: *PLOS ONE* 11.7 (July 2016), e0158977. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0158977.
- [93] Eric W. Brunskill et al. “Atlas of Gene Expression in the Developing Kidney at Microanatomic Resolution”. In: *Dev Cell* 15.5 (Nov. 2008), pp. 781–791. ISSN: 1534-5807. DOI: 10.1016/j.devcel.2008.09.007.
- [94] Steven P. Mah et al. “Kidney Development in Cadherin-6 Mutants: Delayed Mesenchyme-to-Epithelial Conversion and Loss of Nephrons”. In: *Developmental Biology* 223.1 (July 2000), pp. 38–53. ISSN: 0012-1606. DOI: 10.1006/dbio.2000.9738.
- [95] Liwei Jia et al. “Distinct roles of cadherin-6 and E-cadherin in tubulogenesis and lumen formation”. In: *Mol Biol Cell* 22.12 (June 2011), pp. 2031–2041. ISSN: 1059-1524. DOI: 10.1091/mbc.E11-01-0038.
- [96] Filippo Massa et al. “Hepatocyte nuclear factor 1 β controls nephron tubular development”. eng. In: *Development* 140.4 (Feb. 2013), pp. 886–896. ISSN: 1477-9129. DOI: 10.1242/dev.086546.
- [97] M. Pontoglio et al. “Hepatocyte nuclear factor 1 inactivation results in hepatic dysfunction, phenylketonuria, and renal Fanconi syndrome”. eng. In: *Cell* 84.4 (Feb. 1996), pp. 575–585. ISSN: 0092-8674.
- [98] Minoru Takemoto et al. “Large-scale identification of genes implicated in kidney glomerulus development and function”. eng. In: *EMBO J.* 25.5 (Mar. 2006), pp. 1160–1174. ISSN: 0261-4189. DOI: 10.1038/sj.emboj.7601014.
- [99] A. J. Buckler et al. “Isolation, characterization, and expression of the murine Wilms’ tumor gene (WT1) during kidney development”. eng. In: *Mol. Cell. Biol.* 11.3 (Mar. 1991), pp. 1707–1712. ISSN: 0270-7306.

- [100] Pedro J. Beltran, John L. Bixby, and Brian A. Masters. “Expression of PTPRO during mouse development suggests involvement in axonogenesis and differentiation of NT-3 and NGF-dependent neurons”. eng. In: *J. Comp. Neurol.* 456.4 (Feb. 2003), pp. 384–395. ISSN: 0021-9967. DOI: 10.1002/cne.10532.
- [101] Natalya V. Kaverina et al. “Partial podocyte replenishment in experimental FSGS derives from nonpodocyte sources”. eng. In: *Am. J. Physiol. Renal Physiol.* 310.11 (June 2016), F1397–1413. ISSN: 1522-1466. DOI: 10.1152/ajprenal.00369.2015.
- [102] Ellen F. Carney. “Podocytes: ShcA regulates nephrin turnover”. eng. In: *Nat Rev Nephrol* (Oct. 2017). ISSN: 1759-507X. DOI: 10.1038/nrneph.2017.153.
- [103] C. Schell, N. Wanner, and T. B. Huber. “Glomerular development—shaping the multi-cellular filtration unit”. eng. In: *Semin. Cell Dev. Biol.* 36 (Dec. 2014), pp. 39–49. ISSN: 1096-3634. DOI: 10.1016/j.semcdb.2014.07.016.
- [104] Brian A. Pierchala, Maura R. Muñoz, and Cynthia C. Tsui. “Proteomic analysis of the slit diaphragm complex: CLIC5 is a protein critical for podocyte morphology and function”. English. In: *Kidney International* 78.9 (Nov. 2010), pp. 868–882. ISSN: 0085-2538. DOI: 10.1038/ki.2010.212.
- [105] Allen W. Cowley et al. “Pappa2 is linked to salt-sensitive hypertension in Dahl S rats”. In: *Physiol Genomics* 48.1 (Jan. 2016), pp. 62–72. ISSN: 1094-8341. DOI: 10.1152/physiolgenomics.00097.2015.
- [106] Monica Carmosino et al. “MAL/VIP17, a New Player in the Regulation of NKCC2 in the Kidney”. en. In: *Mol. Biol. Cell* 21.22 (Nov. 2010), pp. 3985–3997. ISSN: 1059-1524, 1939-4586. DOI: 10.1091/mbc.E10-05-0456.
- [107] J. Christopher Hennings et al. “The ClC-K2 Chloride Channel Is Critical for Salt Handling in the Distal Nephron”. In: *J Am Soc Nephrol* 28.1 (Jan. 2017), pp. 209–217. ISSN: 1046-6673. DOI: 10.1681/ASN.2016010085.
- [108] Catherina A. Cuevas et al. “Potassium Sensing by Renal Distal Tubules Requires Kir4.1”. eng. In: *J. Am. Soc. Nephrol.* 28.6 (June 2017), pp. 1814–1825. ISSN: 1533-3450. DOI: 10.1681/ASN.2016090935.
- [109] Hayo Castrop and Ina Maria Schießl. “Physiology and pathophysiology of the renal Na-K-2Cl cotransporter (NKCC2)”. eng. In: *Am. J. Physiol. Renal Physiol.* 307.9 (Nov. 2014), F991–F1002. ISSN: 1522-1466. DOI: 10.1152/ajprenal.00432.2014.
- [110] Christopher P. Larsen et al. “LDL Receptor-Related Protein 2 (Megalin) as a Target Antigen in Human Kidney Anti-Brush Border Antibody Disease”. eng. In: *J. Am. Soc. Nephrol.* (Oct. 2017). ISSN: 1533-3450. DOI: 10.1681/ASN.2017060664.
- [111] Kumar Kotlo et al. “Aminopeptidase N reduces basolateral Na⁺-K⁺-ATPase in proximal tubule cells”. eng. In: *Am. J. Physiol. Renal Physiol.* 293.4 (Oct. 2007), F1047–1053. ISSN: 1931-857X. DOI: 10.1152/ajprenal.00074.2007.
- [112] Daniel Caballero et al. “Intraperitoneal pyrophosphate treatment reduces renal calcifications in Npt2a null mice”. eng. In: *PLoS ONE* 12.7 (2017), e0180098. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0180098.

- [113] Yumiko Kiuchi-Saishin et al. “Differential Expression Patterns of Claudins, Tight Junction Membrane Proteins, in Mouse Nephron Segments”. en. In: *JASN* 13.4 (Apr. 2002), pp. 875–886. ISSN: 1046-6673, 1533-3450.
- [114] Takamoto Ohse et al. “Establishment of Conditionally Immortalized Mouse Glomerular Parietal Epithelial Cells in Culture”. In: *J Am Soc Nephrol* 19.10 (Oct. 2008), pp. 1879–1890. ISSN: 1046-6673. DOI: 10.1681/ASN.2007101087.
- [115] Daniel Markovich. “Na⁺-sulfate cotransporter SLC13A1”. en. In: *Pflugers Arch - Eur J Physiol* 466.1 (Jan. 2014), pp. 131–137. ISSN: 0031-6768, 1432-2013. DOI: 10.1007/s00424-013-1388-8.
- [116] Erik Ilso Christensen, Pierre J. Verroust, and Rikke Nielsen. “Receptor-mediated endocytosis in renal proximal tubule”. eng. In: *Pflugers Arch*. 458.6 (Oct. 2009), pp. 1039–1048. ISSN: 1432-2013. DOI: 10.1007/s00424-009-0685-8.
- [117] Kylie Georgas et al. “Analysis of early nephron patterning reveals a role for distal RV proliferation in fusion to the ureteric tip via a cap mesenchyme-derived connecting segment”. eng. In: *Dev. Biol.* 332.2 (Aug. 2009), pp. 273–286. ISSN: 1095-564X. DOI: 10.1016/j.ydbio.2009.05.578.
- [118] Bliss Magella et al. “Cross-platform single cell analysis of kidney development shows stromal cells express Gdnf”. eng. In: *Dev. Biol.* 434.1 (2018), pp. 36–47. ISSN: 1095-564X. DOI: 10.1016/j.ydbio.2017.11.006.
- [119] Benjamin D. Humphreys et al. “Fate tracing reveals the pericyte and not epithelial origin of myofibroblasts in kidney fibrosis”. eng. In: *Am. J. Pathol.* 176.1 (Jan. 2010), pp. 85–97. ISSN: 1525-2191. DOI: 10.2353/ajpath.2010.090517.
- [120] Scott C. Boyle, Zhenyi Liu, and Raphael Kopan. “Notch signaling is required for the formation of mesangial cells from a stromal mesenchyme precursor during kidney development”. eng. In: *Development* 141.2 (Jan. 2014), pp. 346–354. ISSN: 1477-9129. DOI: 10.1242/dev.100271.
- [121] Lino Muñoz Cuellar et al. “Identification and localization of novel genes preferentially expressed in human kidney glomerulus”. eng. In: *Nephrology (Carlton)* 14.1 (Feb. 2009), pp. 94–104. ISSN: 1440-1797. DOI: 10.1111/j.1440-1797.2008.01009.x.
- [122] Prashant S. Patole et al. “Toll-like receptor-4: renal cells and bone marrow cells signal for neutrophil recruitment during pyelonephritis”. eng. In: *Kidney Int.* 68.6 (Dec. 2005), pp. 2582–2587. ISSN: 0085-2538. DOI: 10.1111/j.1523-1755.2005.00729.x.
- [123] Yuqiu Lu et al. “Single-cell RNA-sequence analysis of mouse glomerular mesangial cells uncovers mesangial cell essential genes”. eng. In: *Kidney Int.* 92.2 (Aug. 2017), pp. 504–513. ISSN: 1523-1755. DOI: 10.1016/j.kint.2017.01.016.
- [124] J. Qiao et al. “FGF-7 modulates ureteric bud growth and nephron number in the developing kidney”. eng. In: *Development* 126.3 (Feb. 1999), pp. 547–554. ISSN: 0950-1991.

- [125] Naoki Nakagawa et al. "Dicer1 activity in the stromal compartment regulates nephron differentiation and vascular patterning during mammalian kidney organogenesis". eng. In: *Kidney Int.* 87.6 (June 2015), pp. 1125–1140. ISSN: 1523-1755. DOI: 10.1038/ki.2014.406.
- [126] Jennifer L. Fetting et al. "FOXD1 promotes nephron progenitor differentiation by repressing decorin in the embryonic kidney". In: *Development* 141.1 (Jan. 2014), pp. 17–27. ISSN: 0950-1991. DOI: 10.1242/dev.089078.
- [127] Konstantinos Stamatakis, Francisco J. Sánchez-Gómez, and Dolores Pérez-Sala. "Identification of novel protein targets for modification by 15-deoxy-Delta12,14-prostaglandin J2 in mesangial cells reveals multiple interactions with the cytoskeleton". eng. In: *J. Am. Soc. Nephrol.* 17.1 (Jan. 2006), pp. 89–98. ISSN: 1046-6673. DOI: 10.1681/ASN.2005030329.
- [128] Taizo Nakagawa et al. "Roles of PDGF receptor-beta in the structure and function of postnatal kidney glomerulus". eng. In: *Nephrol. Dial. Transplant.* 26.2 (Feb. 2011), pp. 458–468. ISSN: 1460-2385. DOI: 10.1093/ndt/gfq468.
- [129] Stuart W. Smith et al. "CD248+ stromal cells are associated with progressive chronic kidney disease". eng. In: *Kidney Int.* 80.2 (July 2011), pp. 199–207. ISSN: 1523-1755. DOI: 10.1038/ki.2011.103.
- [130] Sonja Djurdjaj et al. "Keratins are novel markers of renal epithelial cell injury". eng. In: *Kidney Int.* 89.4 (Apr. 2016), pp. 792–808. ISSN: 1523-1755. DOI: 10.1016/j.kint.2015.10.015.
- [131] Frank Costantini. "GDNF/Ret signaling and renal branching morphogenesis: From mesenchymal signals to epithelial cell behaviors". eng. In: *Organogenesis* 6.4 (Dec. 2010), pp. 252–262. ISSN: 1555-8592. DOI: 10.4161/org.6.4.12680.
- [132] Halim Khairallah et al. "Claudin-7, -16, and -19 during mouse kidney development". In: *Tissue Barriers* 2.4 (Aug. 2014). ISSN: 2168-8362. DOI: 10.4161/21688362.2014.964547.
- [133] Lei Yu et al. "GATA2 Regulates Body Water Homeostasis through Maintaining Aquaporin 2 Expression in Renal Collecting Ducts". In: *Mol Cell Biol* 34.11 (June 2014), pp. 1929–1941. ISSN: 0270-7306. DOI: 10.1128/MCB.01659-13.
- [134] John K. McGuire et al. "Matrilysin (MMP-7) inhibition of BMP-7 induced renal tubular branching morphogenesis suggests a role in the pathogenesis of human renal dysplasia". eng. In: *J. Histochem. Cytochem.* 60.3 (Mar. 2012), pp. 243–253. ISSN: 1551-5044. DOI: 10.1369/0022155411435152.
- [135] M. Shibuya. "Structure and dual function of vascular endothelial growth factor receptor-1 (Flt-1)". eng. In: *Int. J. Biochem. Cell Biol.* 33.4 (Apr. 2001), pp. 409–420. ISSN: 1357-2725.
- [136] S. Davis et al. "Isolation of angiopoietin-1, a ligand for the TIE2 receptor, by secretion-trap expression cloning". eng. In: *Cell* 87.7 (Dec. 1996), pp. 1161–1169. ISSN: 0092-8674.

- [137] P. Carmeliet et al. "Targeted deficiency or cytosolic truncation of the VE-cadherin gene in mice impairs VEGF-mediated endothelial survival and angiogenesis". eng. In: *Cell* 98.2 (July 1999), pp. 147–157. ISSN: 0092-8674.
- [138] Qi Ren et al. "Platelet endothelial cell adhesion molecule-1 (PECAM1) plays a critical role in the maintenance of human vascular endothelial barrier function". eng. In: *Cell Biochem. Funct.* 33.8 (Dec. 2015), pp. 560–565. ISSN: 1099-0844. DOI: 10.1002/cbf.3155.
- [139] R. Schwartz-Albiez et al. "The B cell-associated CD37 antigen (gp40-52). Structure and subcellular expression of an extensively glycosylated glycoprotein". eng. In: *J. Immunol.* 140.3 (Feb. 1988), pp. 905–914. ISSN: 0022-1767.
- [140] S. J. Davis and P. A. van der Merwe. "The structure and ligand interactions of CD2: implications for T-cell function". eng. In: *Immunol. Today* 17.4 (Apr. 1996), pp. 177–187. ISSN: 0167-5699.
- [141] Sam W. Moore, Daniel Sidler, and Monique G. Zaahl. "The ITGB2 immunomodulatory gene (CD18), enterocolitis, and Hirschsprung's disease". eng. In: *J. Pediatr. Surg.* 43.8 (Aug. 2008), pp. 1439–1444. ISSN: 1531-5037. DOI: 10.1016/j.jpedsurg.2007.12.057.
- [142] Priya Srinivasan and Maja Maric. "Signal transducer and activator of transcription 1 negatively regulates constitutive gamma interferon-inducible lysosomal thiol reductase expression". In: *Immunology* 132.2 (Feb. 2011), pp. 209–216. ISSN: 0019-2805. DOI: 10.1111/j.1365-2567.2010.03355.x.
- [143] øyvind Salvesen et al. "Activation of innate immune genes in caprine blood leukocytes after systemic endotoxin challenge". In: *BMC Vet Res* 12 (Oct. 2016). ISSN: 1746-6148. DOI: 10.1186/s12917-016-0870-x.

6

OUTLOOK AND DISCUSSION

Our current understanding of embryogenesis is the result of decades of work by many generations of scientists. Despite this tremendous effort, there is still much to discover and unravel. Importantly, the continuing interest in developmental biology is not only driven by the desire to achieve fundamental understanding, but also applications in the medical field. The treatment of conditions caused by developmental defects and leveraging the great potential of regenerative approaches both rely on a detailed understanding of embryogenesis and cell differentiation. Lately, the combination of top-down and bottom-up approaches, as well as the development of new technologies in imagining, sequencing and bioengineering, have delivered new insights at unprecedented speed. In this thesis, we have leveraged state-of-the-art techniques to delve into three different pieces of the embryogenesis puzzle. In each example, new questions and exciting avenues for further exploration were unveiled.

In chapter 2 and 3, we studied how the interaction between embryonic and extraembryonic cells leads to tissue-level organization. In a new *in vitro* model of embryogenesis, the XEN-enhanced gastruloid (XEG) that combines mouse embryonic stem cells and mouse extraembryonic endoderm-like cells (XEN) called XE-gastruloids (XEGs), we observed the formation of tubular structures consisting of neural progenitors-like cells. Notably, this tissue-level organization is achieved without the external addition of signaling cues, or the use of finely tuned growth conditions. The experiments presented in chapter 2 demonstrated the neural-like characteristics of the tubular structures in XEGs. They also showed that cell rearrangements in XEGs are reminiscent of secondary neurulation. This system could thus be used to identify the minimal parameters necessary for neural tube formation *in vitro*. Finally, since XEGs contain derivatives of the three germ layers, as well as extraembryonic cells, they could be the starting point for developing complex model systems of other stages of the development. For example, we were able to differentiate XEGs further to cortex-like tissue and observed the presence of cells with endothelial characteristics, which might be able to form a vascular network.

In chapter 3, we focused on the role and fate of the XEN cells in XEGs. Our ex-

periments showed that the XEN cells induce the formation of neural tubular structures through at least two different mechanisms. On one hand, XENs produce large amounts of extracellular matrix, thereby creating a basal membrane. This might induce the epithelialization of adjacent cells. On the other hand, XEN cells secrete DKK1, an antagonist of WNT signaling. Our signaling perturbation experiments showed that DKK1 was directly connected to the positioning of primitive streak-like cells within XEGs and the elongation axis of the aggregates. It also appears to be necessary for the formation of the neural tube-like structures. Finally, a scRNA-seq measurement revealed that the XEN cells in XEGs differentiate toward a more visceral endoderm (VE)-like state. VE is the extraembryonic endoderm directly surrounding the developing embryo. While it was already established that VE cells and their derivatives influence the differentiation and patterning of the developing embryo, our results suggest that there is also a reciprocal effect of the developing epiblast on the differentiation of the VE cells. This observation supports the idea that interaction between embryonic and extraembryonic cells are crucial for proper embryonic development. Adding extraembryonic cells could thus bring existing embryonic *in vitro* models or even organoids to another level of complexity.

In chapter 4, we proposed new *in vitro* systems to study the sorting between epiblast (EPI) and primitive endoderm (PrE) cells. This important morphogenic event occurs early in development (for example, on day 3 in mouse embryos), when the embryo is still fairly simple. This process is therefore an ideal candidate for minimal *in vitro* modeling. The first system, the chimeric embryoid body (chim-EBs) is perfect for high throughput studies that do not require high-resolution imaging. Chim-EBs could be used for the screening experiments that can identify the key factors of the sorting process. Preliminary experiments with chim-EBs revealed, that blocking the adhesion molecule integrin- β 1 prevents sorting. The second *in vitro* approach, the embryonic disk (ED), approximates the actual geometry of the embryo more closely and has the great advantage of providing optimal conditions to image cell movement. This system allowed us to characterize EPI/PrE sorting at the single-cell level. In the future this system could be used to study the effects of particular candidate genes, identified in screening assays with chim-EBs. The detailed migration dynamics extracted from cell tracking experiments in EDs could also be used to constrain the parameters of a mathematical model of cell sorting. We envision that the combination of our two minimal *in vitro* systems and mathematical modeling will help achieve a detailed understanding of cell sorting in the early mammalian embryo.

Finally, in Chapter 5 we presented a single-cell transcriptomics atlas of the human fetal kidney. We focused specifically on investigating the heterogeneity of nephron progenitor cell types, and the intermediates of podocyte maturation. Those cell types are essential for the proper development of the kidney epithelium and thereby optimal function. Our measurements revealed several transient cell types, present only during development, which were nevertheless associated with kidney disease. As an example, we found the gene OLFM3, which was associated with glomerular filtration rate in a genome-wide association study to be specifically expressed in podocytes precursors. Our data is accessible through the GEO database, as well as an interactive web appli-

cation. An accessible single-cell atlas constitutes a very valuable source of information that might support our understanding of developing organs.

All in all, we hope that this thesis has demonstrated, that minimal model systems and quantitative tools can drive new insights in developmental biology. We are excited for the unprecedented level of understanding that these new approaches will help us achieve in the future.

CURRICULUM VITÆ

Noémie Marie Louise Pauline BÉRENGER-CURRIAS

08-09-1992 Born in Longjumeau, France.

EDUCATION

1998–2003 Primary school
Courcouronnes, France

2003–2007 Secondary school
Courcouronnes, France

2007–2010 High school
Bondoufle, France

2010–2013 Bachelor in Biology
Université d'Evry, France
Thesis: Pathological and pronostic markers for DIV
Supervisor: Prof.dr. P. Manivet

2013–2015 Master in Biology
Université d'Evry, France
Thesis: Correlation between gene expression and growth in
E-coli
Supervisor: Dr. S. van Teeffelen

2015–2021 PhD. Biophysics
Leiden University and Delft University of Technology (Netherlands)
Thesis: Embryonic stem cell-based in vitro models to study
early mammalian development
Promoters: Prof.dr.ir. S.J. Tans, Dr. T. Idema and Dr. S. Semrau

LIST OF PUBLICATIONS

3. **N. M. L. P. Bérenger-Currias**, T. Idema, S. J. Tans, S. Semrau, *Towards modeling epiblast/primitive endoderm segregation in vitro*, Manuscript in preparation.

2. **N. M. L. P. Bérenger-Currias**, M. Mircea, E. Adegeest, P. R. van den Berg, M. Feliksik, M. Hochane, T. Idema, S. J. Tans, S. Semrau, *Extraembryonic endoderm cells induce neuroepithelial tissue in gastruloids*, Manuscript under review.

1. M. Hochane, P. R. van den Berg, X. Fan, **N. M. L. P. Bérenger-Currias**, E. Adegeest, M. Bialecka, M. Nieveen, M. Menschaart, S. M. Chuva de Sousa Lopes, S. Semrau, *Single-cell transcriptomics reveals gene expression dynamics of human fetal kidney development*, PLOS Biology **17**, 2 (2019).