EPFL

Thèse n° 9201

In vitro endoderm emergence and self-organisation in the absence of extraembryonic tissues and embryonic architecture

Présentée le 11 février 2022

Faculté des sciences de la vie
Unité du Prof. Lutolf
Programme doctoral en biotechnologie et génie biologique

pour l'obtention du grade de Docteur ès Sciences

par

Stefano Davide VIANELLO

Acceptée sur proposition du jury

Prof. A. C. Oates, président du jury
Prof. M. Lütolf, directeur de thèse
Prof. A.-K. HADJANTONAKIS, rapporteuse
Prof. H. LICKERT, rapporteur
Prof. D. DUBOULE, rapporteur

École polytechnique fédérale de Lausanne

2022

École polytechnique fédérale de Lausanne

School of Life Sciences (SV) and School of Engineering (STI)
Institute of Bioengineering
Laboratory of Stem Cell Bioengineering (LSCB)

Doctoral Thesis

In vitro endoderm emergence and self-organisation in the absence of extraembryonic tissues and embryonic architecture

Stefano Davide Vianello

Ext. Reviewer	Anna-Katerina Hadjantonakis Developmental Biology Programme The Sloan Kettering Institute
Ext. Reviewer	Heiko Lickert Institute of Diabetes and Regeneration Research (IDR) Helmholtz Zentrum München
Int. Reviewer	Denis Duboule École Polytechnique Fédérale de Lausanne Laboratory of Developmental Genomics
Committee President	Andrew C. Oates (EPFL)
Thesis Director	Matthias Lutolf (EPFL)
Mentor	Maartje Basting (EPFL)

December 2021

Stefano Davide Vianello

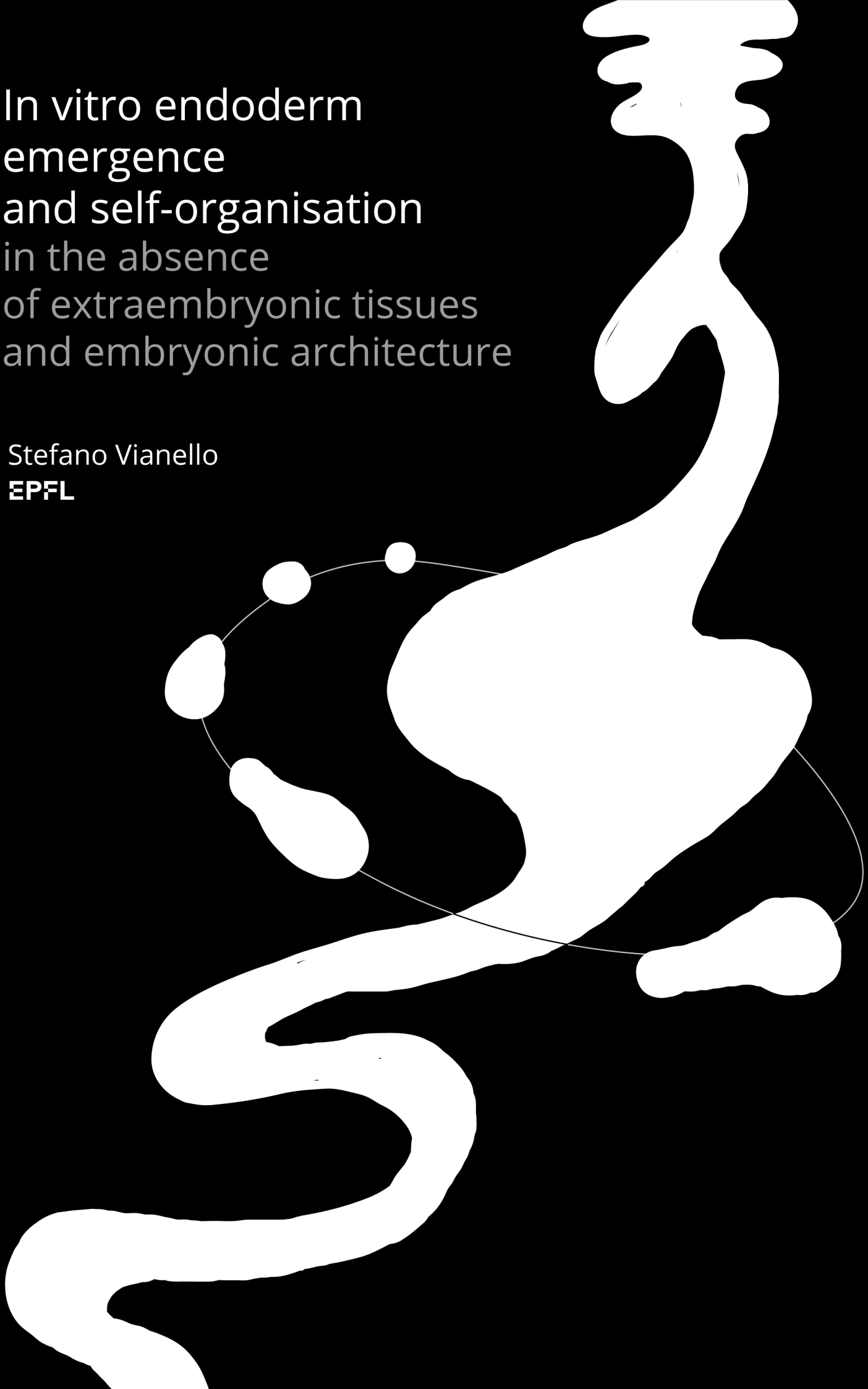
In vitro endoderm emergence and self-organisation in the absence of extraembryonic tissues and embryonic architecture

Doctoral Thesis, December 2021
Committee: Anna-Katerina Hadjantonakis, Heiko Lickert, and Denis Duboule
Committee President: Andrew C. Oates
Supervisors: Matthias Lutolf (PI) and Maartje Basting (mentor)

École polytechnique fédérale de Lausanne
Laboratory of Stem Cell Bioengineering (LSCB)
Institute of Bioengineering
School of Life Sciences (SV) and School of Engineering (STI)
Route Cantonale
1015 and Lausanne

In vitro endoderm emergence and self-organisation in the absence of extraembryonic tissues and embryonic architecture

Stefano Vianello
EPFL



-

> *Are you sure, sweetheart, that you want to be well?*
>
> **— from "The Salt Eaters" by T. C. Bambara**
> (Bambara, 1992)

To every PhD student (end to every person) that took the brave decision to choose to, indeed, be well. And everything that that entails in the too often violent reality of academia.

Acknowledgements

I am thankful to Matthias Lütolf for letting me work on this project, for granting me the freedom to explore and work across fields and projects, and for the trust put in my abilities and decisions. You have allowed me independence and you made this project feel as mine. Thank you to my mentor, Professor Maartje Bastings: thank you for your support and for the many times you went out of your way to ensure my academic, physical, emotional well-being.

I would like to thank the members of my jury committee: Professor Anna-Katerina Hadjantonakis, Professor Heiko Lickert, Professor Denis Duboule. Your work has defined early mouse developmental biology and embryonic endodermal biology: I am greatly indebted to each of you and to the time you took to read and evaluate this thesis before and during the defence, as well as for discussing endoderm and gastruloids with me. I would like to thank Professor Andy Oates for agreeing- and taking time to preside the defence. Thank you.

Acknowledgements go to the staff of all EPFL core facilities: thank you to Jessica Dessimoz, Gian-Filippo Mancini, and Nathalie Müller at the Histology Core Facility. Dr Dessimoz, your work on endoderm patterning is weaved into this very thesis. Thank you Sènan d'Almeida for the training at the Flow Cytometry Core Facility. Thank you to everyone at the BioImaging and Optics Platform: Arne Seitz, José Artacho, Nicolas Chiaruttini, Thierry Laroche, Romain Guiet, Olivier Burri. Thank you to Fabien Kuttler of the Biomolecular Screening Facility for training me at the automated slide scanner and thus for allowing me to set up the foundations of my quantification pipeline. Thank you to the entire team of the Glassware Washing Facility: Guillaume Burtin, Sergiy Ulyanov, Maria Sonia Hernandez Alba, Negjmije Mazreku, Dominique Violette Salir, Silvia Maria Duarte Martinho Antunes.

I thank Professor Alfonso Martinez-Arias. I learned in one year in your lab foundations that would drive me through four years of PhD and that would inform my approach to Developmental Biology. In your lab I lived a model of what science can be and how it can be done.

My PhD experience taught me, more than anything, that few things are more

important than a lab environment where one can feel valued and respected: it can make you or break you. I thank current and past members of the Laboratory of Stem Cell Bioengineering at EPFL that created or gave me the space to put the pieces back together. I thank Hao-An Hsiung, Hwanseok Jang, Si-Ying Lai, I look forwards to hear about your discoveries on gastruloids and I wish you all the best. Moritz Hofer, Tania Huebscher, Jakob Langer, Bilge Sen Elçi, I find your research fascinating and I wish you the best too. Thank you for inspiring me in many ways and I hope I might have left you a slightly easier way through your own PhD journey. Grazie Francesca Sgualdino, compagna di avventure, lotte, e rivoluzioni negli ultimi mesi. Per citare Montale, il vento ch' entra nel pomario. Abbiamo occupato spazio in spazi che non ci meritavano. Thank you Francisco Lorenzo-Martín and Sophia Li, and — from my earlier years in the lab — thank you Andrea Manfrin, Nathalie Brandenberg, Delphine Blondel, Sonja Giger, François Rivest. Thank you Jonathan Brassard, Lorenzo Mattolini, Roxane Kreuter, David Maycotte, Gabriele Gambardella, you have been help and support without even knowing it. Thank you Evangelos Panopoulos for guiding my very first steps in the lab and helping me settle in. Thank you Lucie Tillard for managing the lab during my final year here.

Then there are people for which written acknowledgements cannot do justice and for which not enough could probably ever be said. And maybe it is in the unspoken and in the unwritten that the true dimension of their significance to these four years of PhD of mine is best expressed. The work in this thesis speaks and breathes their names and their energy even in the absence of direct contribution. Jisoo Park. Saba Rezakhani. Antonius Chrisnandy. The ink of your very names on this page bleeds through the entire thickness of this thesis, it is the ink from which every single result in it was written. Thank you.

I met people that seemed to inhabit and define kinder universes within and around them. People able to locally warp the fabric of space and time, inhabiting and defending pockets shielded from the violent laws of academia. Pockets in which to escape whenever crossing paths: at times refuges, at time spaces of deep learning and empowerment. Again, thank you Jisoo, Saba, Antonius. I have learned a lot from you. You bring deep, dimensional, powerful universes to this world. Thank you Stefanie Boy, as a lab manager and as a person during and after your stay in the lab. You have been a safe space. A lit window in the dark. A warm and kind presence for me and others during my PhD. Thank you for sharing this space with me. Thank you Julia Tischler. You brought light, lightness, and warmth with you. I learned a lot from you and it was an honour to be able to help you back when I could. Your positivity ripples around you and changes the world; thank you for sharing this

space with me. Thank you Rita Sarkis. Thank you for creating a space of resistance, revolution, empowerment, and possibility at EPFL. Sara Ahmed writes: "In order to survive institutions, we need to transform them. But we still need to survive the institutions we are trying to transform". You allowed me and many other students at EPFL to survive attempts to change this institution. Thank you for sharing this space with me. Thank you Dominique Kolly, my model of what academia can be. You have saved lives. Thank you for sharing this space with me. Finally, thank you Julia Prébandier. We started and ended our time at EPFL at the same time. Thank you for supporting me throughout; thank you for sharing this space with me.

I thank all the other fellow students and researchers at EPFL (and beyond) with which I crossed paths and that I remember with affection. Grazie Alice Comberlato, and thank you Eva Kurisinkal: you made your office a welcoming and healing harbour in MX. Thank you Alex Mayran and, again, Dominique for sharing expertise, reagents, protocols, and insights throughout these years. Upstairs neighbours and collaborators that always put kindness above anything else. Thank you to my "second lab" and to the entire academic community at the European Molecular Biology Laboratory in Heidelberg, and specifically to all members of the Aulehla Lab and to Professor Alexander Aulehla. Silvija Svambaryte, Carina Vibe, Emilia Esposito, Hidenobu Miyazawa, Jona Rada, Takehito Tomita, Simona Gioe, Nobuko Tsuchida-Straeten, Christine Ho, you welcomed a stranger in your lab with nothing but warmth and kinship. It is in your lab that I saw mice embryos for the first time. More than anything, you showed me what a lab can be and what a lab can do. Mariana Alves, Samanta Seah, you are ever inspiring. Hui Ting Zhang, I wish you all the success in the world and I am so happy to have seen you again and again since we were classmates in Cambridge. Thank you for the time then and for the time in these four years talking hydrogels and mouse embryos. Thank you to Professor Alba Diz-Muñoz, Professor Robert Prevedel and Carlo Bevilacqua for hosting me at EMBL and teaching me about Brilluoin spectroscopy.

I would like to acknowledge the role of people and organisations that have been part of my PhD journey tangentially to what presented in my thesis. Thank you all involved in the e-Mouse Atlas Project, and specifically to Chris Armit for supporting my project and my vision for the EMAP models ever since we met at the VizBi conference in Heidelberg. Thank you Sarah Kearns of the Knowledge Futures Group for publishing my essay on Commonplace, you gave a platform to my words. Thank you to all members of the Knowledge Equity Lab: your work has not yet reached Developmental Biology and the field is poorer from it. Your projects and your writings have forever change my approach to open science, preprint curation,

and scientific practice at large.

Thank you Andrea Attardi, you are a big part of this thesis. Thank you Victoria Yan. Our discussions, sometimes with spotty internet, most of the time until too late at night, had that flavour of revolutionary and of undercommoning. Talking about what matters, building new presents and new futures in the face of so much destruction. More than any result discussed in this thesis it is in these conversations that I see my legacy to be, and from what has and will come out of it. You both have been and remain an inspiration. We exist, scattered around the world but we exist.

Thank you Paul Gerald Layague Sanchez. This thesis, and much of this research, would not have existed without you. This work is yours as much as it is mine. You are the paper fabric on which it is printed, you are the energy that allowed me to write it, you are the force that drove me through these four years, discussion partner of most of these experiments, the home I would come back to. You have been and will forever be my example of humanity, kindness, unconditional love, acceptance, support, knowledge in its multiplicity of dimensions, scientific practice. It is through you that I truly learned what a PhD is, regardless of diplomas, experiments, papers, and results. You have brought humanity there where I could find so little. Thank you and thanks to your entire family. Salamat.

Questa tesi esiste sopratutto grazie alla mia famiglia. Mamma, Francesca, Alvise. Non sarei qui senza di voi. A voi devo tutto. Mamma, a parole non ti ho mai ringraziata abbastanza, o fatto capire quanto tu mi abbia aiutato in questi quattro anni. Voglio scriverlo qui. Grazie per avermi chiamato e chiesto di come stavo. Grazie per essere stata lì per me e per esserci sempre. Ma è anche vero che questi quattro anni di dottorato ci hanno rubato tantissimo tempo insieme. Non sono sicuro che ne sia valsa la pena. Vorrei allora che questa tesi resti come ricordo di quanto io abbia bisogno di voi e di quanto siate importanti per me. E voglio che i vostri nomi siano qui accanto al mio. Anna Boldo, Francesca Vianello, Alvise Vianello, Francesco Vianello. Grazie per il vostro amore incondizionato.

Abstract (English)

In humans, mice, and other mammals key internal organs such as the gut, the lungs, the pancreas, and the liver all derive from the same embryonic tissue: the endoderm. The development of all of these structures thus depends on a same set of early cells, and on the developmental instructions provided to them by the embryo. One approach to study what these instructions might be is to study the behaviour of endoderm cells when they can *not* rely on embryonic instructions. That is, when they are generated and made to develop outside of the embryo. The rationale behind this approach, widely embraced by the field of synthetic embryology, is that the unguided development of specific embryonic cell types can provide a window to the observation of intrinsic characteristics of these cells and of fundamental developmental principles which might not be executed or manifested within the regulated environment of the embryo. Accordingly, in the work here described I use gastruloids — synthetic models of development made by aggregation of mouse embryonic stem cells *in vitro* — to characterise the developmental behaviour of endoderm cells, in the absence of embryonic cues and in the absence of the extraembryonic tissues that they would usually rely on *in vivo*.

In the **first chapter** of this thesis, I provide an introduction, overview, and review of our current understanding of the developmental biology of mouse endoderm as this develops in the embryo. I also introduce gastruloids as stem-cell-based experimental system to study developmental mechanisms *in vitro*, decoupled from the constraints of experimental work with animals and animal embryos.

Chapter 2 is dedicated to the earliest steps of endoderm development in gastruloids. I show when and where these cells appear, and how they sort as the gastruloid starts to mature. I find that endoderm cells follow similar spatial and temporal developmental patterns as those they would follow if they were in the embryo. I notably find that they always remain in an epithelial state, even as other cell types undergo mesenchymal transition.

In **Chapter 3** I describe surprising dynamics of endoderm cells over time, and how they self-organise into complex architectures at the core of the gastruloids. The epithelial domain that forms is dynamic and plastic, and partitions the gastruloids into interfacing compartments. Importantly, I reveal a significant endodermal component

to this *in vitro* system, usually considered to be unstructured and neural/mesodermal in character.

In **Chapter 4** I show how the core endoderm compartment that I see forming in gastruloids is patterned along the longitudinal axis of the aggregate, with different cellular identities at different anterior-posterior positions of the domain. By comparison with published embryonic datasets, I identify endodermal identities corresponding to the full spectrum of those that make up the gut tube in the embryo. Specifically, I describe how most endodermal cells adopt an early intestinal fate, and how many acquire anterior foregut identities.

The work described in this thesis presents gastruloids as promising *in vitro* systems for the study of endoderm biology, morphogenesis, and mature fate specification, filling a gap within the field of synthetic embryology. It also notably opens translational possibilities to the use of gastruloids to produce endodermal cell types which may be otherwise difficult to generate using conventional directed differentiation methods.

Keywords: endoderm, gut tube, gastruloid, stem cells, embryo, developmental biology, epithelium, FoxA2, Cdh1, mouse

Sommario

Nello sviluppo degli esseri umani, dei topi, e della maggior parte delle altre specie di mammiferi, organi di primaria importanza quali lo stomaco, i polmoni, il pancreas ed il fegato (per menzionarne solo alcuni) derivano tutti dallo stesso tessuto embrionale: l'endoderma. Lo sviluppo di queste strutture dipende quindi dalla stessa (ed unica) famiglia di cellule primigene, e dalle istruzioni che esse ricevono dall'embrione. Per elucidare quali possano essere tali istruzioni, una strategia investigativa consiste nello studiare il comportamento delle cellule endodermali quando queste sono *private* di istruzioni embrionali. In altre parole, quando sono lasciate a svilupparsi (o quando sono generate direttamente) all'esterno dell'embrione. Giustificazione di questo approccio, sempre più adottato nel campo della "synthetic embriology" [embriologia sintetica], è il fatto che il libero sviluppo di cellule embrionali apre una possibile finestra su caratteristiche intrinseche a queste cellule e su principî fondamentali di sviluppo che potrebbero invece non essere eseguiti (o non essere distinguibili) nel contesto altamente controllato e coreografato dell'embrione. Alla luce di queste considerazioni, il lavoro di tesi qui presentato fa uso di cosiddetti gastruloidi [gastruloids], modelli sintetici di sviluppo ottenuti dall'aggregazione di cellule staminali embrionali di topo *in vitro*, al fine di caratterizzare la biologia dello sviluppo delle cellule di endoderma, e quindi in assenza delle istruzioni che fornirebbe l'embrione e in assenza dei tessuti extraembrionali dai quali dipende lo sviluppo di queste cellule *in vivo*.

Il **Capitolo primo** di questa tesi presenta un'introduzione, riassunto, e recensione delle conoscenze attuali sulla biologia dello sviluppo dell'endoderma di topo così come si sviluppa nell'embrione. Il capitolo introduce anche i gastruloidi in quanto sistemi sperimentali — formati a partire da cellule staminali embrionali — per lo studio, *in vitro*, dei meccanismi di sviluppo embrionale, disaccoppiati dai vincoli associati alla sperimentazione animale o embrionale animale.

Il **Capitolo secondo** è dedicato alla descrizione delle primissime tappe dello sviluppo dell'endoderma nei gastruloidi. Descrivo dove e quando queste cellule appaiono, e le loro prime fasi di riorganizzazione a man mano che il gastruloide comincia a maturare. Mostro che le cellule endodermali seguono modelli di sviluppo simili, sia nello spazio che nel tempo, a quelli che seguirebbero se si trovassero nell'embrione. In particolare, mostro che queste cellule rimangono sempre in uno stato epiteliale, anche mentre altri tipi cellulari compiono transizioni verso stati mesenchimali.

Nel **Capitolo terzo**, descrivo sorprendenti dinamiche delle cellule endodermali, e come queste si riorganizzano per formare architetture complesse al centro dei gastruloidi. Il territorio epiteliale che si viene a formare è dinamico e plastico, e

suddivide il gastruloide in compartimenti interfaccianti. Notevolmente, evidenzio una componente endodermale non trascurabile in un modello *in vitro* (il gastruloide) fino a quì prevalentemente considerato in termini neurali e mesodermali e considerato privo di struttura.

Nel **Capitolo quarto** dimostro che diversi tipi di cellule endodermali sono distribuiti in modo ordinato lungo il compartimento al centro dei gastruloidi, e quindi organizzate in funzione del loro asse principale. Appoggiandomi su dataset embrionici pubblici, identifico identità cellulari corrispondenti a quelle dell'intero spettro di cellule che formano l'intestino primitivo dell'embrione. In particolare, mostro che la maggior parte delle cellule endodermali adotta un destino intestinale, e che molte adottano identità corrispondenti a quelle dei territori anteriori dell'intestino primitivo [foregut].

Il lavoro descritto in questa tesi presenta i gastruloidi come promettenti sistemi sperimentali per lo studio della biologia dello sviluppo dell'endoderma, la morfogenesi di questo tessuto, e la specificazine di indetità endodermali mature *in vitro*, colmando anche un vuoto nel campo della synthetic embryology. Specialmente, i risultati qui presentati aprono possibilità traslazionali nell'uso dei gastruloidi per la produzione di cellule endodermali mature di identità altrimenti difficilmente ottenibili tramite protocolli convenzionali di "directed differentiation" [differenziazione guidata].

Parole chiave: endoderma, intestino primitivo, gastruloide, cellule staminali, embrione, biologia dello sviluppo, epitelio, FoxA2, Cdh1, topo

Accessible Abstract

Note: Below is a summary, in more accessible terms, of the research presented in this thesis. It is written by using words from only the thousand most common words in the English language (see https://splasho.com/upgoer5/); i.e. without being able to use words like "endoderm","embryo", "cells" and "gastruloids".

In the first few days that animal babies are growing inside their mother, when they are still very very little, they have to go through very big changes. They are still missing a lot of important body parts! They will need a stomach to eat food, they will need a way to breathe, and they will need many other inside parts to allow their bodies to work best. We now know that many of these important body parts are built from the same family of building blocks. This work studies what this family of building blocks can do when they are by themselves, without any help from the outside. This kind of work helps us better understand what these building blocks look like, how they know what to do, how they talk with one another to build all the body parts we need, and what could have gone wrong when this does not work out.

Contents

List of Figures

List of Tables

1 Introduction

> *For greater precision in description, [the author] finds it necessary to give to these membranes special names, and he therefore employs [...] for the internal that of endoderm*
>
> **— Proceedings of the Royal Society of London**
> (On George James Allman's study of freshwater hydroids
> Allman and Forbes, 1853)

It is to distinguish the cellular layers that make up the body of a freshwater Hydrozoan (specifically, *Cordylophora lacustris*), that the term endoderm first sees the light within the scientific literature (Allman and Forbes, 1853). And with it the term came to define that innermost, "very distinctly cellular" layer lining the internal stomach cavity of these hollow organisms (Allman and Forbes, 1853). An "internal skin" (etymologically; Vaca-Merino et al., 2021) that in fact had already been seen to be present in other organisms at least since the description of the "mucous layer" of developing chicken embryos earlier in the century (Pander, 1817, as translated in Oppenheimer and Hamburger, 1976). From Allman's hydroids to Pander's chick embryos, the endoderm is in fact a defining feature of most (Ereskovsky and Dondua, 2006) animal life forms, and is accordingly now recognised as one of the three fundamental "morphological units" of all early embryos, the so-called germ layers (Oppenheimer and Hamburger, 1976). And while the digestive tube formed by the endoderm may well arguably be one of the organs with clearest homology across the animal tree of life (Mashanov et al., 2014), it can equally take the most disparate forms, be it the single, recursively bifurcating gut of some branched annelids (Ponz-Segrelles et al., 2021), or the showy, dangling, externally-floating appendage of cusk-eel fish (Fukui and Kuroda, 2007).

In mammalian embryos (*e.g.* mouse and human), the early endoderm first develops to form an internal structure called gut tube, generative platform for many of the internal organs of the body, and not just the digestive tube itself. During embryonic development, this single structure will thus give rise to the pharynx, the thyroid, the parathyroid, the thymus, the trachea, the oesophagus, the bronchi of the lungs, the stomach, the liver, the pancreas, the small intestine, the large intestine, and the colon

(from anterior to posterior position; Carlson, 1994; Nowotschin et al., 2019a; Lewis and Tam, 2006). Organs with clear vital importance, responsible for maintaining homeostatic function and to serve the metabolic needs of the organism (Ober and Grapin-Botton, 2015). In dysfunction, organs often named in association of severe congenital and non-congenital disease (cystic fibrosis, chronic hepatitis, diabetes), sites of cancer development, interfaces and targets of bacterial and parasitic burden. Underscored by the position of endoderm as the common denominator of function and dysfunction for such a wide variety of organs, the importance of studying the developmental biology of this germ layer and of its cells, especially given that many aspects of endoderm early development still remain outstanding (see Nowotschin et al., 2019a; Bardot and Hadjantonakis, 2020, and later discussion points).

In the work presented in this thesis, and as detailed in the following sections, I approach the topic of endoderm development from an *in vitro*, deconstructed standing point. That is, whereby key features of endoderm identity and behaviour (i.e. of endoderm developmental biology) can be revealed when these cells are taken out of normal developmental contexts, and characterised in their self-organisation deprived of external cues. From the dialectic comparison with what known to take place (or what yet to be explored) when these cells are in the embryo, what these cells are believed to require during development *in vivo*, insights into necessary and sufficiencies of endodermal biology, and into signature features and behaviours associated with endodermal identity (see also Steventon et al., 2021). In turn, a more focused picture of endoderm developmental biology able to inform critical targets of consideration and intervention when translated to other endodermal contexts, whether in normal development or in disease.

1.1 The endoderm and the embryo

Informed investigations of endoderm development *in vitro* need to be grounded to the knowns and unknowns of endoderm development *in vivo*, and what inferred from observation of normal and abnormal embryonic development. Accordingly, an overview of endoderm development in the mouse embryo is illustrated in Fig.1.1 A1. [Note: The embryonic models illustrated in this chapter and throughout the rest of the thesis are based on real-life embryonic volumetric data as described and distributed in Armit et al., 2017; Richardson et al., 2014; here employed for illustration through methods I developed during my PhD and published in Vianello, 2020, a copy of which is appended to this thesis.

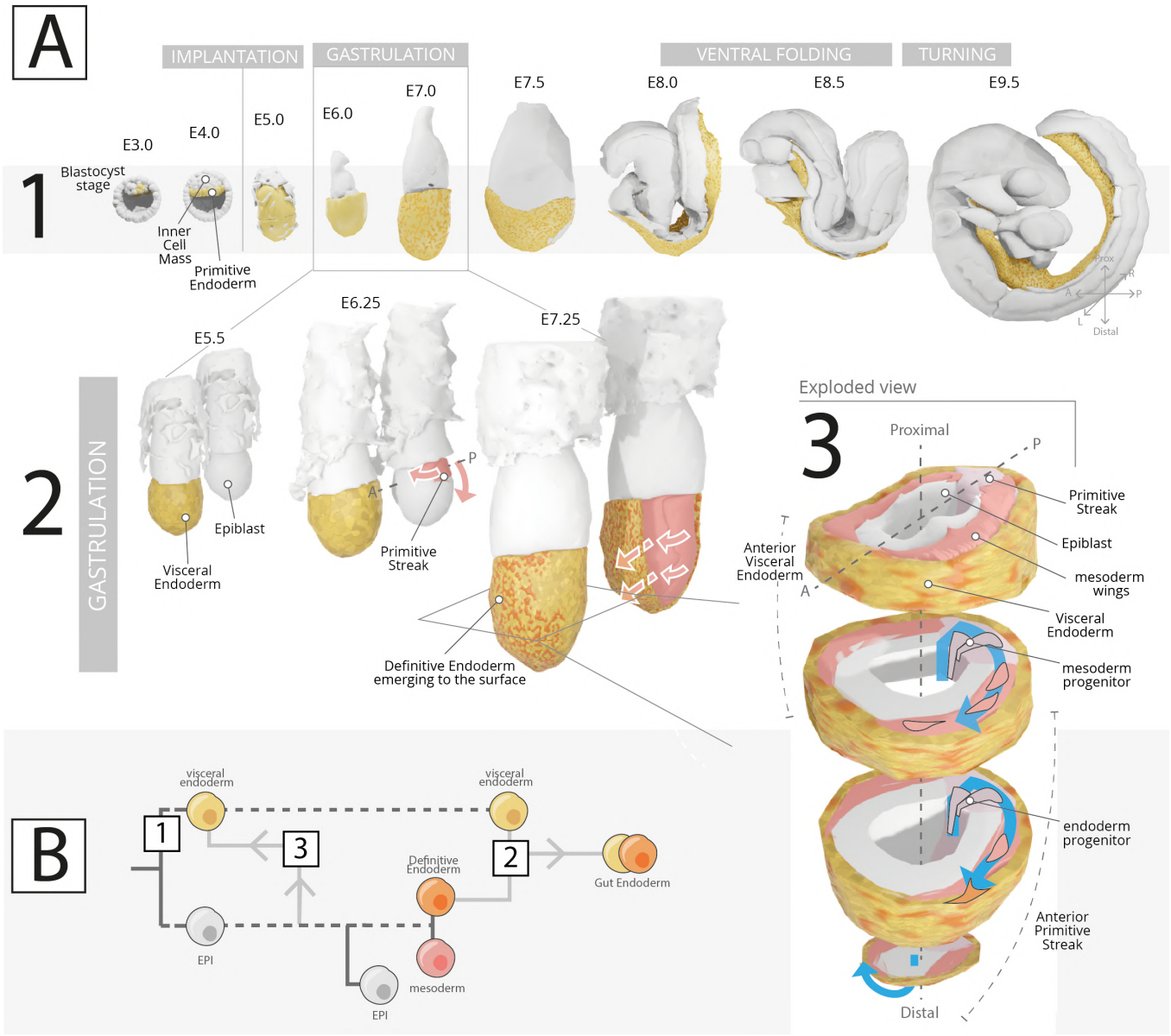


Fig. 1.1.: Summary of endoderm development in the peri-implantation mouse embryo. **A) 1:** Timeline of mouse development from pre-implantation to gut tube formation. For visualisation purposes, not all embryonic tissues are displayed. The endoderm lineage is highlighted in yellow. Relevant developmental events are labelled. **2:** Focused illustration of the moments at- and around gastrulation. The primitive streak and mesodermal cells emerging from it are highlighted in red. Definitive endoderm cells intercalating to the outer visceral endoderm layer (yellow) are highlighted in orange. **3:** Exploded view of the mid-gastrulation mouse embryo (E7.0). Illustrative paths of mesoderm and endoderm cells are indicated. **B)** Summary schematic of the different contributions to the gut endoderm in the mouse; adapted from Nowotschin et al., 2019a. A: Anterior, P: Posterior, L: Left, R: Right, EPI: Epiblast, E: Embryonic Day. Note: parts of this figure are also published in Vianello and Lutolf, 2020, CC-BY 4.0

As immediately noticeable, endoderm (highlighted in yellow in Fig.1.1 A) first assembles on the outer surface of the embryo, and only later (after embryo implantation, Embryonic day (E)4.5, and after gastrulation, E6.25) reaches its internal position to take the form of the gut tube (compare positions of endoderm at the beginning and the end of the timeline illustrated). In fact, the initial emergence of embryonic endoderm to the surface of the embryo and its later infolding both involve complex cell-level and embryo-level transformations that have been (and still are) objects of active investigation (Burtscher and Lickert, 2009; Kwon et al., 2008; Probst et al., 2021; Viotti et al., 2014, summarised in the present section).

Crucially, "endoderm" is already present in the embryo before implantation. This so-called Primitive Endoderm (PrE) is however extraembryonic in origin, and comes from one of the earliest fate segregation events of embryonic development: cells that are not fated to reserve their potency towards the building of the embryo proper (*i.e.* Epiblast (EPI) cells) are allocated to the PrE fate (Chazaud et al., 2006). The PrE will then expand as to sheathe embryonic cells while the conceptus implants into the uterus of the mother, and as the embryo grows into its characteristic cylindrical shape (Smith, 1985). It is within this outer extraembryonic envelope of endoderm (now referred to as Visceral Endoderm (VE)), that the small and potent epithelium of the EPI thus starts undergoing gastrulation (Arnold and Robertson, 2009; Takaoka and Hamada, 2012; Tam and Behringer, 1997; Tam and Loebel, 2007; Fig.1.1 A2). In fact, it is a signalling centre asymetrically established at one side of the VE, the Anterior Visceral Endoderm (AVE), that breaks the radial symmetry of the embryo and confines the site of gastrulation to the newly-defined posterior of the embryo (Stower and Srinivas, 2014; Thowfeequ et al., 2021; Zhang and Hiiragi, 2018; Takaoka and Hamada, 2012). Consequently to this breaking of simmetry, the initially multipotent and uncommitted cells of the early EPI commit to specific fates, generally classified into the broad germ layer categories of ectoderm (skin and neural types), mesoderm (heart, muscles, and mesenchyme), and endoderm (as already noted, internal organs, respiratory and digestive tract; Arnold and Robertson, 2009; Takaoka and Hamada, 2012; Tam and Behringer, 1997; Tam and Loebel, 2007). It is thus at the moment of gastrulation that, in the embryo, embryonic endoderm (called Definitive Endoderm (DE) as to distinguish it from the pre-existing, extraembryonic PrE; illustrated in orange and yellow respectively in Fig.1.1) first emerges, literally, out of the EPI epithelium (Burtscher and Lickert, 2009; Kwon et al., 2008; Probst et al., 2021; Viotti et al., 2014; Fig.1.1 A3).

Just as for that of the other germlayers, embryonic day (E)6.25 (gastrulation) thus becomes a crucial moment in the developmental history of endoderm. Exposed to the high Wingless/int1 (Wnt), Bone Morphogenetic Protein (BMP), and Fibroblast

Growth Factor (FGF) signalling activity restricted posteriorly by the AVE and by other extraembryonic tissues, posterior EPI cells respond by engaging so-called Epithelial-to-Mesenchymal Transition (EMT) programmes: they start losing attachment with the rest of the epithelium, they become motile and mesenchymal, they leave the EPI (Arnold and Robertson, 2009; Tam and Behringer, 1997; Tam and Loebel, 2007; Fig.1.1 A2 and A3). Morphologically, the so-called Primitive Streak (PS) appears: a distally-expanding zone of EMT leading to delamination of EPI cells and simultaneous commitment to embryonic fates (Hashimoto and Nakatsuji, 1989; Williams et al., 2012). As EPI cells undergo EMT and leave the EPI, they start circumnavigating its outer surface, sandwiched under the overlaying VE, forming wings of tissue converging towards the anterior of the embryo (Hashimoto and Nakatsuji, 1989; Saykali et al., 2019; Viotti et al., 2014; red cells in Fig.1.1 A3). Many of these cells will generate mesodermal derivatives, heart and muscles (Tam and Behringer, 1997). Yet, others only share this compartment for a short time (Burtscher et al., 2013) and instead intercalate into the overlying VE. These cells are DE progenitors differentiating to the endoderm fate (Fig.1.1 A3, bottom slice).

Clearly thus, and as illustrated in Fig.1.1 A, the mature endoderm, the layer of endoderm that will then fold and close to form the gut tube, does not come entirely from cells originating in the EPI (Kwon et al., 2008; Nowotschin et al., 2019b; Nowotschin et al., 2019a; Viotti et al., 2014). To the opposite, it derives from the folding and closing of an epithelium composed of a mixed contribution of "extraembryonic" cells (the cells of the original VE epithelium sheathing the EPI; yellow) and of embryo-derived DE cells that intercalated within it as they emerged from the PS at gastrulation (orange; Burtscher and Lickert, 2009; Kwon et al., 2008; Probst et al., 2021; Viotti et al., 2014). Originally believed to be displaced off the distal portion of the embryo, the outer envelope of VE instead quickly becomes a mosaic of endodermal cells of extraembryonic- and embryonic origin (Kwon et al., 2008; Viotti et al., 2014; Burtscher and Lickert, 2009), though both cell types eventually converge to very similar (yet not identical) transcriptional signatures (Nowotschin et al., 2019a). The multiple contributions to gut endoderm described above are summarised in Fig.1.1 B (as adapted from e.g. Nowotschin et al., 2019a). In addition to the first contribution from VE cells and to the later one from intercalating DE cells, it also highlights a third source of cells: EPI cells bypassing EMT and altogether bypassing transit within the mesodermal compartment of the embryo. These cells leave the EPI to directly intercalate into the VE, a contribution that has been documented to occur at the distal tip of the pre-gastrulation mouse embryo, zone of maximal mechanical stress (Hiramatsu et al., 2013; Matsuo and Hiramatsu, 2017), and that has found support from single-cell transcriptome analyses (Nowotschin

et al., 2019a). Direct EPI to endoderm transitions are particularly interesting as even endodermal progenitors that do classically egress from the EPI into the mesodermal space might do so by EMT processes different than those governing the egression and specification of mesoderm (Bardot and Hadjantonakis, 2020; Burtscher and Lickert, 2009; Probst et al., 2021; Scheibner et al., 2021). Currently, data suggests that egressing endodermal progenitors do not completely lose their epithelial character but instead transiently redistribute their surface adhesion molecules as they travel along the mesodermal compartment, until they contact their new epithelial niche, the VE, and fully repolarise (Bardot and Hadjantonakis, 2020; Kwon et al., 2008; Nowotschin et al., 2019a; Viotti et al., 2014). Indeed, recent transcriptional comparisons have confirmed that endodermal progenitors show reduced expression of EMT and migration determinants compared to their mesodermal counterparts, suggesting separate and distinct modes of delamination (Probst et al., 2021; Scheibner et al., 2021). A dedicated later section expands on these important new findings.

Regardless of the specific modes through which endoderm-fated cells leave the EPI compartment to intercalate within the VE, the resulting outer sheet of cells will later form pockets at the anterior and posterior of the embryo, and finally fold along its midline to close into a tube that will end up internalised within the embryo (Carlson, 1994; Lewis and Tam, 2006; Spence et al., 2011; Wells and Melton, 1999; Fig.1.1 A1). The gut tube has formed, and along its entire length progenitors of all endoderm-derived visceral organs will emerge and take shape (Carlson, 1994; McGrath and Wells, 2015).

1.1.1 Embryonic origins of the endoderm

Despite the consolidated validity of the general picture of mouse endoderm development described above (and illustrated in Fig.1.1 A), uncertainty still remains over many specificities of several of the steps within the process. One of these, the very identity (or rather, the effective potency) of the cells from which DE cells emerge at gastrulation. Given that embryonic endoderm (*i.e.* DE) delaminates out of the EPI in spatial and temporal continuity (and overlap) with mesoderm cells (Burtscher and Lickert, 2009), questions remain regarding the possible existence of a shared progenitor that would originate both cell types. Central to this point, the concept of so-called mesendodermal progenitors, seen to give rise to both mesoderm and endoderm *in vitro* (Pfendler et al., 2005; Tada et al., 2005; Kubo et al., 2004; D'Amour et al., 2005), as well as the fact that such a shared progenitor is indeed a common developmental mode in other living organisms (e.g. the textbook case of the *Caenorhabditis elegans (C. elegans)*' EMS cell (Sulston et al., 1983), but also in

vertebrate species; see Nowotschin and Hadjantonakis, 2020; Rodaway and Patient, 2001; McClay et al., 2021). The embryonic origin of endoderm would thus appear to be intertwined, cellularly, with that of mesoderm, rather than a separate, distinct specification process. A cell lineage would exist that would not give rise to mesoderm only, but rather to both mesoderm and endoderm cells (Lewis and Tam, 2006). In the mouse embryo however, evidence for the existence of such a mesendodermal progenitor population remains scarce (Tzouanacou et al., 2009; Lewis and Tam, 2006; Nowotschin and Hadjantonakis, 2020).

Where the location from which endoderm cells emerge out of the EPI has been investigated, endoderm progenitors have been seen to delaminate anteriorly to the expanding PS (through which mesodermal cells break away from the EPI epithelium), directly squeezing under the VE (Burtscher and Lickert, 2009; Scheibner et al., 2021). While DE cells will also be found intermingled amongst mesodermal cells circumnavigating the embryo (Viotti et al., 2014), there is thus clearly a spatial segregation between emerging endoderm and mesoderm cells within the embryo. Double positive cells do exist in the liminal space between EPI and VE (Burtscher and Lickert, 2009; Ivanovitch et al., 2021), but these have been demonstrated to be progenitors of the cardiac ventricle (a derivative classified as mesodermal) rather than a progenitor population that would give rise to endoderm (Bardot et al., 2017; Ivanovitch et al., 2021). In fact, DE cells already show clear commitment to either germlayer (in terms of marker expression) prior to their egression from the epitheliun (Burtscher and Lickert, 2009). Such an early commitment is further reflected by the majoritarily endoderm-restricted contribution of fate-mapped EPI cells in labelled embryos (Tzouanacou et al., 2009).

Searches for mesendodermal progenitors within the EPI itself, rather than within the PS, have not been successful either. Coordinated methylome, nucleosome, and transcriptome profiling of perigastrulation mouse embryos has shown that the epigenetic priming (here: increased chromatin accessibility and demethylation) of gene batteries associated with mesoderm and endoderm is, yes, synchronous and coordinated, but characterises later stages of differentiation (*i.e.* cells transiting through the PS; Argelaguet et al., 2019). The epigenetic landscape of pre-gastrulation EPI cells does not appear to be sub-dividable in groups with clear (joint) endomesodermal *vs* ectoderm commitment (Argelaguet et al., 2019), and in fact key endodermal determinants, such as the Transcription Factor (TF) FoxA2, have pioneering activity of their own (Cernilogar et al., 2019). Instead, most recent lineage tracing and single cell analysis of early-gastrulating mouse embryos have shown that endoderm cells are specified within the very PS and from unbiased progenitors (Probst et al., 2021), and that endomesodermal progenitors do not stably arise during early endoderm

development *in vivo* (Probst et al., 2021; Mittnenzweig et al., 2021). The "last common ancestor" of DE cells *in vivo* would thus effectively be the general category of unbiased (posterior) EPI cells, being independently fated to either mesoderm or endoderm (Probst et al., 2021).

Still, and as put forward *e.g.* in Nowotschin and Hadjantonakis, 2020, the absence of clear mesendodermal progenitor state *in vivo* could be a contingency of the specific signalling landscape existing within the time and spaces of mouse development at the posterior EPI, and at the PS (specifically, varying levels of Nodal; Lewis and Tam, 2006). Were mouse EPI cells bipotential, they are *de facto* not so when existing in a developmental landscape where the signals for either lineage are asymmetrically or chronologically distributed. Outside of this specific context, *i.e. in vitro*, such "mesendodermal" potential might be revealed and might perhaps be what indeed described in the literature (Nowotschin and Hadjantonakis, 2020; Wang and Chen, 2016; Sui et al., 2013).

Of note in light of the apparent separation of mesoderm and endoderm progenitors within the EPI, not necessarily conserved *in vitro*, is still the intertwined developmental history of cells of these two germ layers, not only in their spatial distribution within the gastrulating embryo (Burtscher and Lickert, 2009; Viotti et al., 2014), but also in the apparent plasticity that DE cells seem to retain at all times. This plasticity is executed for example in their ability to shunt to mesodermal identities upon perturbation(see also discussions in Nowotschin and Hadjantonakis, 2020; Lewis and Tam, 2006). Indeed, perturbation of canonical Wnt signalling levels by β-catenin deletion has been shown to induce loss of endoderm and transfating to mesoderm (Engert et al., 2013; Lickert et al., 2002). DE cells that fail to rintercalate into the epithelium of the VE (their final destination), downregulate endoderm markers and are lost within the mesodermal wings (Viotti et al., 2014; Kanai-Azuma et al., 2002; though their fate, mesoderm or otherwise, is unknown). Additionally, specific endoderm subsets have been found to transit through differentiation stages that include the expression of classic mesoderm markers, such as T/Bra (Brachyury) (Imuta et al., 2013), likely reflecting dominant transcription factors and Wnt-richness (Garriock et al., 2020) within the posterior embryo rather than necessarily germ layer identity, but possibly further confusing *in vitro* inferences on the necessity of T/Bra^{+} ("mesendodermal") transitions for DE development (Probst et al., 2021).

1.1.2 Endoderm, epithelia, and EMT

Second outstanding point of mouse endoderm developmental biology is its relationship with epithelial identity, both its own and that of the cell types that surround it during gastrulation. It has long been postulated that endoderm cells may undergo a sequence of mirrored EMT-Mesenchymal-to-Epithelial Transition (MET) processes as they first leave the EPI (EMT) and then integrate the overlaying VE (MET) (Viotti et al., 2014; Nowotschin et al., 2019a; Nowotschin and Hadjantonakis, 2020). Yet the short time taken by these cells to complete their journey across compartments (Burtscher et al., 2013) is unlikely to allow for such profound transciptional and proteomic changes (as suggested in Scheibner et al., 2021). Indeed, an increasing number of studies are finding evidence for alternative modes of epithelial egression rather than EMT when considering endoderm progenitors (Probst et al., 2021; Scheibner et al., 2021), with Viotti et al., 2014 having long observed that endoderm progenitors crossing compartments unexpectedly maintain E-cadherin (Cdh1) expression (whose downregulation has been generally considered a landmark of EMT; Campbell, 2018; Cano et al., 2000; Wheelock et al., 2008).

Focusing on the relationship between endoderm identity, epithelial properties, and changes in these properties (EMT), a recent combination of lineage labelling and single cell transcriptomics has in fact established that endoderm progenitors do not employ classical EMT programmes to leave the EPI, but rather delaminate by plastic rearrangement of their epithelial contacts, upregulating N-cadherin (Cdh2) while also maintaining E-cadherin (Scheibner et al., 2021). The study demonstrates that FoxA2, marker upregulated as endoderm progenitors leave the EPI, protects these cells from EMT by shielding them from its trigger, Wnt signalling. In fact, FoxA2 directly induces the expression of canonical Wnt signalling inhibitors (Scheibner et al., 2021), resulting in endoderm cells in the embryo *downregulating* classic EMT mediators (*e.g.* Snai1) and Wnt signalling effectors that are instead deployed within the mesoderm. Accordingly, it has been shown that Snai1 knockout embryos are not impaired in endoderm formation (Scheibner et al., 2021) but show defective mesoderm (Carver et al., 2001). Other single cell transcriptional analysis focused on these early embryonic cells also confirm that early endoderm progenitors express lower levels of EMT genes compared to mesoderm progenitors (Probst et al., 2021).

Where Snai1-independent endoderm egression has been documented in other systems, it has found to be mediated instead by Gata orthologues (Campbell et al., 2011). Whether some Gata TFs may be involved in the mouse is still to be disentangled but factors like Gata4 and Gata6 are well known to be expressed in

these cells (Nowotschin et al., 2019a). The T-box TF Eomes, other major marker of endodermal progenitors within the EPI (Probst et al., 2021; Arnold et al., 2008), may alternatively also be responsible of initiating EMT programmes in those cells of the posterior EPI that are not expressing Snai1 (Arnold et al., 2008; Bardot and Hadjantonakis, 2020).

Regardless of the precise identity of upstream inducers, FoxA2$^+$ endoderm cells are now known to express relatively high levels of EMT suppressors (GRHL2, Ovol2, Pofut2, Elf3) and of several metalloproteinases that may thus facilitate local Extracellular Matrix (ECM) remodelling (Scheibner et al., 2021). Cells would thus leave the EPI by a partial "EMT" likely involving basement membrane breakdown by matrix metalloproteinases, and transient downregulation of epithelial integrity and polarity determinants rather than loss of epithelial cadherin (Scheibner et al., 2021). In the mouse (as in other amniotes, Voiculescu et al., 2014) egressing EPI cells have been documented to constrict apically and displace their cell bodies towards the basal side (Ramkumar et al., 2016). Effective extrusion from the EPI columnar epithelium, once the basement membrane has been broken down, may be itself depend on myosin anisotropy differences between endoderm progenitors and their neighbours within the EPI (Ramkumar et al., 2016), and this could thus be a mechanism of endoderm progenitor egression upon basement membrane breakdown.

What might be the functional implication of the retention of E-cadherin by endoderm progenitors, and thus their double Cdh1$^+$/Cdh2$^+$ decoration once they have egressed from the EPI? References to sorting effects are now being most considered, whereby the double-positive cadherin set of endodermal cells (N-cadherin$^+$/E-cadherin$^+$) compared to N-cadherin$^+$/E-cadherin$^-$ mesoderm in which these cells end up transiently intermixed (Viotti et al., 2014) would mediate sorting out of these two populations. Indeed differences in surface cadherins have been able to drive sorting in simple *in vitro* models of engineered cells (Toda et al., 2018; Katsamba et al., 2009; Nose et al., 1988), sometimes through processes that appear size-dependent (Cachat et al., 2016), as well as in embryonic cells sorting *in vitro* (Ninomiya et al., 2012). In aggregates of mixed *Xenopus* cells, extensive sorting behaviour occurs spontaneously to regroup cells based on their germ layer of origin (Townes and Holtfreter, 1955), just as is the case for re-aggregated 2D stem cell micropatterns (Minn et al., 2020). Yet the extent to which this may be sufficient to drive sorting *in vivo* is uncertain (Ninomiya et al., 2012). Certainly, expression of different surface cadherins has been shown to influence active cell movement/migration processes, and E-Cadherin-retaining endoderm cells retain contacts and migrate collectively in the fly (Campbell and Casanova, 2015; Campbell, 2018). In fact, differences between cadherin types have been found to govern migration dynamics in neural

crest cells (through a process termed Contact inhibition of Locomotion, CiL), another embryonic cell type born from incomplete EMT (Campbell, 2018; Scarpa et al., 2015). Still, the movement/migration parameters of endoderm progenitors in the mouse have been less studied than, for example, those of mesoderm (Saykali et al., 2019). Endoderm movement may well not be an active migration process, but be mere "hitch-hiking" on migrating mesodermal cells (Nowotschin et al., 2019a).

Regardless of any possible functional involvement of their Cadherin set during compartment crossing, it is clear that the *retention* of epithelial features such as E-cadherin is essential for the seamless integration into their destination epithelium, which itself could explain the advantage of undergoing EMT through modes that do not require sacrificing Cdh1 expression. Here again, the "epithelial gatekeeper" (Scheibner et al., 2021) FoxA2 mediates the epithelialisaiton process as migrating endoderm cells start upregulating Sox17 to merge into the VE (Burtscher and Lickert, 2009; Viotti et al., 2014). Mutant cells that fail to integrate the VE remain trapped within the mesodermal compartment, seem to lose their identity, and the resulting embryos lack midgut and hindgut (Viotti et al., 2014; Kanai-Azuma et al., 2002).

1.1.3 Towards the formation of a patterned gut tube

Where do endoderm cells go after they have delaminated off the EPI? As mentioned previously, studies that have labelled and tracked endoderm progenitors as they emerge in the embryo describe them first emerging distally (and at EPI domains anterior to the morphological anterior/distal end of the PS; Burtscher and Lickert, 2009), then intermixing with mesodermal cells emerging from the PS and circumnavigating the embryo (Burtscher and Lickert, 2009; Viotti et al., 2014), and then intercalating into the VE that envelops the entire embryo (Viotti et al., 2014; Kwon et al., 2008; Fig.1.1 A.3).

As shown by a number of cell tracking and clonal analysis studies, DE exiting the EPI at progressively later timepoints (throughout gastrulation) populate the overlaying VE at progressively posterior positions (Franklin et al., 2008; Lawson and Pedersen, 1987; Lawson et al., 1986). As such, the earliest egressing DE cells will intercalate at the anterior of the VE and thus end up in the portion of the epithelium that will end up forming the foregut, while the latest egressing DE cells will intercalate at more posterior positions and end up within the midgut and hindgut (Tam et al., 2007). In fact, the intercalation of DE cells within the VE is not symmetrical and homogeneous, rather, it has been found to occur majoritarily within the anterior half of the embryo (Nowotschin et al., 2019b; Kwon et al., 2008; Pijuan-Sala et al., 2019).

Accordingly, anterior and posterior halves of the VE (which then folds to become the gut tube) end up being characterised by different proportions of embryonic (DE) cells (Nowotschin et al., 2019b; Kwon et al., 2008; Pijuan-Sala et al., 2019). The implications of this differential in embryonic/extraembryonic contributions remain unknown, yet — and despite the quick convergence of transcriptional signatures between cells of the two different origins once they become part of the gut tube — former VE remain notably characterised by the expression of epigenetic and metabolic determinants (Ttr, Rhox5, Trap1a, Xlr3a, Cdkn2a, Trf, PLA2G1B, FMR1nb; Nowotschin et al., 2019b).

Regardless, following the migration and intercalation of DE cells at different axial positions of the VE is a phase of drastic morphological transformations involving the "inwards" folding (i.e. invaginating within the amniotic cavity) of the EPI/Ectoderm, mesoderm, and VE layers to form a pocket anteriorly (the Foregut Pocket) and, 6-to-8h later, an equivalent pocket at the posterior (the Hindgut Pocket; Lewis and Tam, 2006; Spence et al., 2011; Fig.1.2 A). As these pockets deepen and as the sides of the embryo further fold around the midline, the embryo eventually closes on itself, internalising and folding its endoderm as to form the embryonic gut tube (Fig.1.2 A, right). In very coarse descriptive terms, the endoderm epithelium closes to form the cylindrical gut tube by ventral folding around the midline (such as that lateral endoderm cells will end up located ventrally, and midline endoderm cells will form the dorsal part of the gut tube), and by anterior and posterior expansion of the anterior and posterior pockets respectively (Spence et al., 2011; Franklin et al., 2008; Tam et al., 2004). Accordingly, cells towards the apex of the foregut pocket will end up in the anteriormost positions of the gut tube, cells at the side of the pocket facing the anterior of the embryo (the "floor" of the pocket, facing the heart domain) will end up forming ventral portions of the anterior gut tube, and cells at the side of the pocket facing the posterior of the embryo (the "roof" of the pocket, facing the neural ectoderm and notochord) will end up forming dorsal portions of the anterior gut tube (Tremblay and Zaret, 2005; Franklin et al., 2008). Endoderm cells situated at the lip of the gut pocket end up located at the posteriormost boundary of the foregut, as the lip propagates caudally towards the rostrally-advancing posterior pocket (Tremblay and Zaret, 2005; Zorn and Wells, 2009; Lawson et al., 1986; Fig.1.2 A). The closing edges of the endoderm sheet end up gathered together at the level of the midgut, resolving into the so-called vitelline (omphalomesenteric) duct (Kaufman and Bard, 1999; Cervantes et al., 2009). This set of transformations is described as Ventral Folding Morphogenesis (VFM) (detailed in Gavrilov and Lacy, 2013; Madabhushi and Lacy, 2011). The embryo has shifted from a lordose position to its more distinctive and recognisable foetal position, and the gut endoderm has

shifted from enveloping the conceptus as VE to being innermost to the embryo in the form of the gut tube (in fact, VFM inverts the relative outside-in position of all three germ layers; Lewis and Tam, 2006; Fig.1.2 A, right).

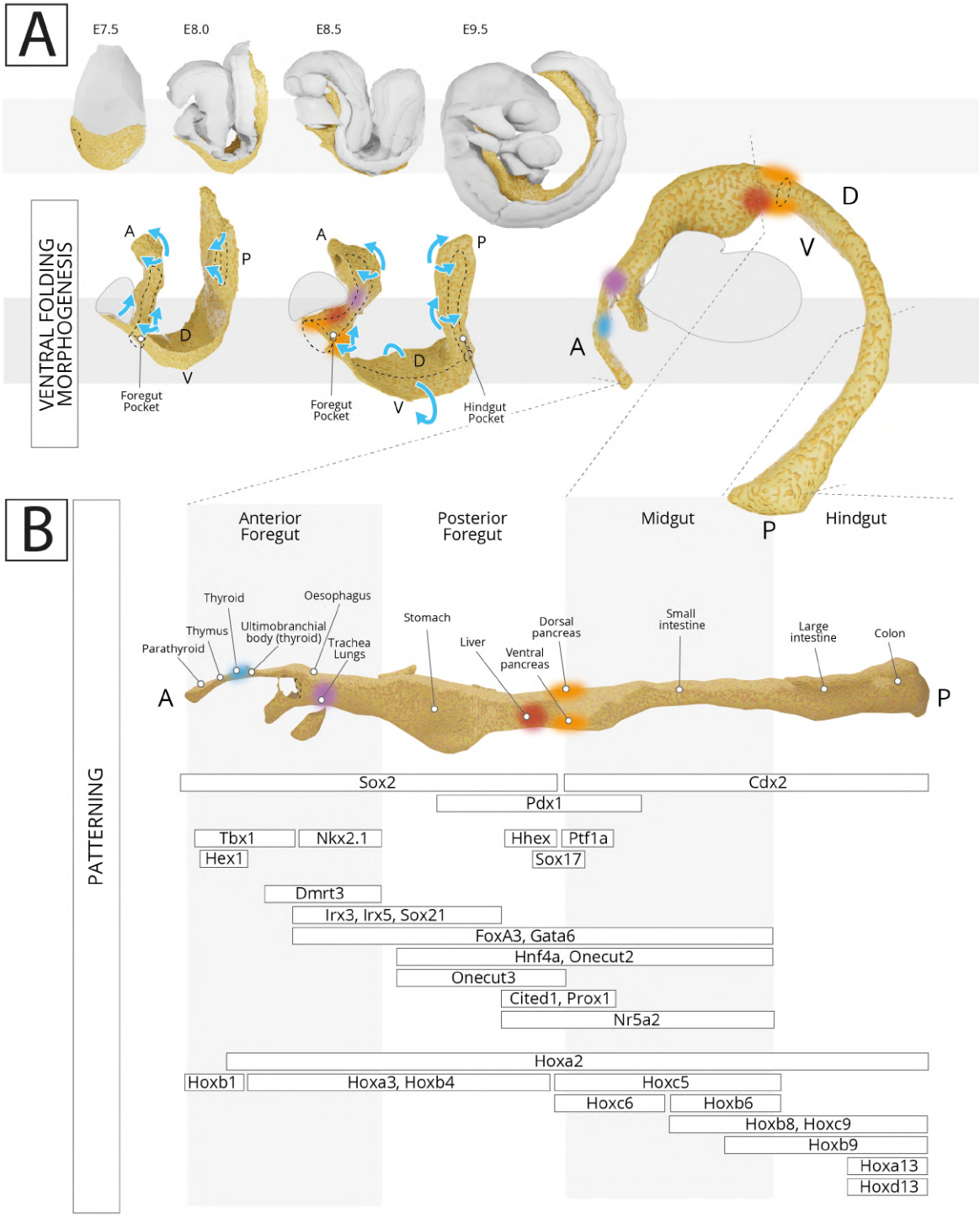


Fig. 1.2.: Ventral Folding Morphogenesis and patterning of the embryonic gut tube. See next page for the full caption to this figure ↪

Fig. 1.2.: ↪ continued from the previous page. **A)** Post-gastrulation development of the mouse embryo (top row) with a focus on the endoderm layer (bottom row, yellow. Blue arrows indicate tissue-level movement of the epithelium to produce the gut tube (rightmost) during Ventral Folding Morphogenesis. Dashed lines indicate internal cavities and the midline. The relative position of different primordium domains is color-coded (e.g. maroon: liver). Anterior structure in gray: cardiac tissue. **B)** Summary of the main gene expression domains patterning the early embryonic gut tube at around E9.0. The AP position of different organ domains is also indicated. A: Anterior, P: Posterior, D: Dorsal, V: Ventral. Illustration based on data from (Sherwood et al., 2007).

1.1.4 Embryonic patterning of the gut tube

Once it has been formed, the embryonic gut tube will act as the shared generative platform from which most visceral organ will then emerge (Lewis and Tam, 2006; Carlson, 1994). All along its length the tube swells, coils, buds to give rise — from anterior to posterior position — to the pharynx, the thyroid, the parathyroid, the thymus, the trachea, the oesophagus, the bronchi of the lungs, the stomach, the liver, the pancreas, the small intestine, the large intestine, and the colon (Carlson, 1994; Nowotschin et al., 2019a; Lewis and Tam, 2006). How are different regions of the gut tube specified to such different fates? How is positional information encoded, interpreted by, and relayed to endoderm cells at different positions along the embryonic gut tube?

Classic transplantation experiments (originally in rats, rather than mice) have long shown that endoderm at anterior and posterior positions of the embryo is already non-equivalent by the headfold stage (*i.e.* around E7.5). That is, it will differentiate into different tissues even when left to develop under the same patterning cues. Indeed, teratomas derived from anterior sections of endoderm, which thus neighboured the anterior neural tissues in the embryo, exclusively give rise to foregut (*i.e.* anterior) tissue, while teratomas derived from posterior sections of endoderm, and which thus neighboured the PS and the Node in the embryo, exclusively give rise to midgut and hindgut (*i.e.* posterior) tissue (Švajger and Levak-Švajger, 1974). Aiming to resolve global spatial and temporal transcriptional profiles that could discriminate different endodermal populations within the maturing gut tube endoderm, recent investigations have collected single-cell expression data from full-length mouse gut tubes (Nowotschin et al., 2019b; Li et al., 2021), as well as from parts of them (Han et al., 2020). These results not only confirm that non-equivalent endoderm cells with anterior and posterior bias can be detected already by E7.5, but that cells at different positions along the Anterior-Posterior (AP) axis of the gut tube are already specified to corresponding organ identities by E8.75 (Nowotschin et al., 2019b), that

is much prior to morphological differentiation of the different regions of the gut tube (which starts around E9.5; see *e.g.* Guiu and Jensen, 2015). These findings echo earlier observation that, at least for a subset of progenitors domains, organ-specific genes would already be expressed "pre-morphogenesis" (*e.g.* liver and pancreas; Gualdi et al., 1996; Gittes and Rutter, 1992).

Undoubtedly, the patterning and spatial information inputs of the gut tube endoderm are intimately linked to the local distribution of soluble signalling cues secreted by adjacent germ layers (mainly the surrounding mesoderm, Wells and Melton, 2000). Indeed, a variety of recent single-cell datasets of parts of the gut tube and associated tissues have consistenly highlighted signalling interplay between endodermal cells and their local non-endodermal neighbours (*e.g.* Han et al., 2020; Kishimoto et al., 2020; Rankin et al., 2021 and more). The functional importance of such interplay in fact explains numerous *in vitro* experimental approaches where endoderm identities are made to develop in conjunction with mesodermal tissue (*e.g.* Silva et al., 2020; Rossi et al., 2021a). Maybe most studied example of such an interaction is the role played by the cardiac mesoderm (and associated tissues and derivatives) in the patterning of the early anterior foregut pocket and thus of the anterior of the gut tube endoderm, which includes progenitor domains of identities spanning from the pharyngeal arches up to liver and pancreas at its posteriormost boundary (Fig.1.2 B). As highlighted in Fig.1.2 A, embryonic foregut tissues — and specifically the floor and the lip of the Foregut Pocket — develop in close spatial association with cardiac tissues (indicated in gray) and this throughout all stages of development. Accordingly, the cell populations in closest proximity to it will be exposed to FGF secreted by the cardiac mesoderm (Gualdi et al., 1996) and BMP from mesenchymal cells in the same region (Rossi et al., 2001) and thus be specified to a specific fate (in this case, hepatic fate; see *e.g.* summary in Duncan and Watt, 2001) and not to another one. If endoderm cells of the same region were not to receive such signals, they would eventually give rise to (ventral) pancreas instead (Rossi et al., 2001), and in fact that is the fate — to continue on this same example — of the cell of the anterior foregut lip, outside of the range of influence of hepatogenic FGFs and BMPs (Duncan and Watt, 2001; Deutsch et al., 2001).

Clearly, any of the other structures that surround the endoderm in the embryo could act as further positional cues through the secretion of signalling determinants. The notochord for example (not illustrated in Fig.1.2 but located dorsally to the endoderm from around E8.5 onwards; Balmer et al., 2016) has indeed been shown to inhibit Shh expression of adjacent (i.e. dorsally-located) endodermal domains (otherwise expressed throughout; Echelard et al., 1993), and thus instructs the specification of dorsal pancreas at the midgut-hindgut junction (Apelqvist et al.,

1997; Hebrok et al., 1998; Cleaver and Krieg, 2001). At the posteriormost of the gut tube, the high levels of FGF and Wnt signalling in the embryonic tailbud have been shown to otherwise maintain hindgut identity and actively repress foregut endoderm differentiation (Wells and Melton, 2000; Ameri et al., 2010).

Maps of the autonomous signalling landscape of the gut tube itself have also recently been drawn (Nowotschin et al., 2019b), confirming that the hindgut and midgut (*i.e.* future intestinal territories) are in a domain of high FGF and Wnt activity (as is more generally the posterior of the embryo; see *e.g.* Aulehla and Pourquié, 2010; Garriock et al., 2020; Dessimoz et al., 2006). Notch signalling responsive gene activity and Retinoic Acid determinants are instead mostly found in endoderm of the posterior hindgut. These studies also reveal that a variety of other signalling activities pattern the embryonic gut tube and even discontinuously so along its length (*e.g.* different Janus kinase/Signal Transducer and Activator of Transcription (JAK/STAT) inhibitors patterning posterior, anterior, and midgut domains; Nowotschin et al., 2019b; thus likely regulating different cytokine signalling contexts). With the exception of Nodal signalling, all major developmental signalling pathways (Gerhart, 1999) appear to potentially regulate endoderm biology within the gut tube, or at least able to provide positional information cues (Nowotschin et al., 2019b). In fact, studies on the topic have reached the state where extremely detailed hypothetical combinatorial signaling network roadmaps can be drawn (Han et al., 2020) but their functional relevance mostly remains to be investigated and validated.

The resulting gene expression landscape along the AP, Dorso-Ventral (DV), and Left-Right (LR) axes of the gut tube is complex (Sherwood et al., 2007), but ultimately combinatorially allows to define territories with unique gene-expression combinations that in turn translate precise positional information (as sketched in Fig.1.2 B). Ultimately, such fates are executed through the expression of region-specific transcription factors/target genes that there find permissive conditions for expression. In fact, the expression levels of a limited set of around 20 TFs has been recently shown to be able to carry positional information along the entire gut tube, whether these are the TFs that execute the corresponding fate or not (Nowotschin et al., 2019b). The Hox code, universally taken as laying down the spatial and temporal coordinate system of the vertebrate embryonic axes (Deschamps and Nes, 2005), does not instead seem to translate as well to the gut tube (Nowotschin et al., 2019b).

1.2 Synthetic embryology and *in vitro* models of embryonic development

How to investigate events that take place during the earliest days of embryonic development in a controlled, targeted, and high-throughput way given the constraints of implantation? One answer is clearly to remove the embryo from the implantation chamber and study it *in vitro*. Indeed, the mouse embryo can be grown *ex utero* across extensive developmental time-windows even when separated from the mother as early as the blastocyst stage (Hsu, 1979; Bedzhov and Zernicka-Goetz, 2014; Aguilera-Castrejon et al., 2021). Experiments such those of *e.g.* Wells and Melton, 2000 also show that insightful results can of course be obtained by direct *in vitro* manipulation of embryos extracted from the uterus at the appropriate stage, or of endoderm dissected away from these embryos. Yet, *in vitro* work with mouse embryos still suffers from the limitation of low-throughput, of work with live animals, of ethical and legal restrictions, as well as from *e.g.* the dependence on established mutant lines for genetic manipulation experiments. To obviate these problems, many have turned towards self-organising systems — based on Pluripotent Stem Cells (PSC)s such as mouse Embryonic Stem Cells (ESC)s, human PSCs, or induced PSCs rather than embryos — recapitulating, *in vitro*, key features of *in vivo* development. These include embryoid bodies, embryoids, 2D-gastruloids (micropatterns) and 3D gastruloids (reviewed in Simunovic and Brivanlou, 2017; Shahbazi and Zernicka-Goetz, 2018; Vianello and Lutolf, 2019). Crucially, the use of PSCs allows for high-throughput approaches to developmental investigations. Moreover the self-organising nature of the structures emerging favours the discrimination between real dependencies and mere contingencies between structures or events observed *in vivo* (Turner et al., 2016; Steventon et al., 2021; Shahbazi and Zernicka-Goetz, 2018). Additionally, and when these models are employed to study mouse development, a mutually-informative relationship can be established between *in vitro* observations and *in vivo* biology in virtue of the extensive knowledge currently existing about the mouse model (versus *e.g.* the human embryo; see comparative approach adopted *e.g* in Morgani et al., 2018)

In very recent years self-organising mESCs systems have been shown to be extremely interesting tools to recreate aspects of gastrulation *in vitro* and to study key features of the process (*i.e.* its induction, the EMT events associated with emergence of the PS, the patterning of emerging cell types; Morgani et al., 2018; Sozen et al., 2018; Turner et al., 2017). While some of these approaches aimed to reconstitute the conceptus from is individual components (*i.e.* trophectoderm, EPI, and PrE) and

indeed observe induction of gastrulation-like events in a fraction of the synthetic aggregates generated (Sozen et al., 2018; Amadei et al., 2021; Harrison et al., 2017), others adopted a more "deconstructed" approach, starting with unpatterned EPI-like systems and trying to identify the minimal set of conditions able to drive such a population through the major milestones of development (Morgani et al., 2018; Turner et al., 2017; Beccari et al., 2018). Arguably, this latter approach better fits bioengineering frameworks as it allows to ask directed questions about the types of cues read by cells during development, as well as to probe the outcome of the supply of alternative cues, not experienced by the embryo *in vitro*. Experimental control of specific cues provided in such minimal systems facilitates the dissection of the complex environment of *in utero* peri-implantation and allows to identify individual functional contributions by specific tissues or by specific mechanic/geometrical features. Specifically, *in vivo* cell fate specification and morphogenesis are heavily dependent on the 3-dimensional (3D) context of the developing embryo. Such features cannot all be recapitulated by monolayer systems. Among the 3D models of development available to date, one finds gastruloids, the experimental model employed in the work presented in this thesis.

1.2.1 Gastruloids: 3D embryonic organoids

Gastruloids are self-organising aggregates of mESCs employed as embryonic organoids because they spontaneously recapitulate *in vitro* key hallmarks of peri-implantation mouse development. The derivation of gastruloids has its origin in the observation that P19 embryonal carcinoma cells (pluripotent cells derived from an embryonic teratoma; McBurney, 2003) are able to aggregate and display spontaneous axial elongation if maintained for four days under hanging-drop conditions in the presence of serum and Dimethyl sulfoxide (DMSO). Elongation is accompanied by regionalised expression of posterior mesodermal markers such as T/Bra and Cdx2 and driven by the analogous signalling mechanisms than those operating in the embryo (*i.e.* canonical Wnt signalling; Marikawa et al., 2009). Molecular and genetic characterisation of these structures has showed that cells in this system are undergoing movements (*e.g.* convergent extension) and programmes of differentiation characteristic of gastrulation *in vivo* (Li and Marikawa, 2015; Lau and Marikawa, 2014; Marikawa et al., 2009). It was later shown that analogous behaviour could be replicated by mESCs if these are maintained in serum-free conditions for five days and treated with a pulse of CHIR99021 (CHIR) (an inhibitor of GSK3β, mimicking the downstream consequences of Wnt signalling activation) on the third day from aggregation (Brink et al., 2014). Just as the embryonic organoids generated from P19 embryonal carci-

noma cells, those generated by aggregation of mESCs were shown to recapitulate major transitions of early development such as symmetry-breaking, specification of the germ layers, axial organisation, elongation, and cell movements reminiscent of gastrulation (Brink et al., 2014; Turner et al., 2017; Beccari et al., 2018). For these reasons they have been called gastruloids, even though the specific sequence of events associated with EMT at the PS during gastrulation has never been formally documented.

The knowledge-base surrounding gastruloid biology — and the extent to which they recapitulate *in vivo* mouse development — has been increasing since, with detected patterns of gene expression indicating not only that they are able to develop along an AP axis marked by posterior T/Bra expression (just as in the embryo; Kispert and Herrmann, 1994), but that they also pattern dorso-ventrally and bilaterally (Turner et al., 2017; Beccari et al., 2018). Investigations by our and other labs have further shown that the emerging AP axis of gastruloids is progressively patterned by Hox gene expression in the same spatiotemporal sequence as embryos *in vivo* (Beccari et al., 2018). For the past three years, I have been curating a list of gastruloid-related publications on the community site of the Company of Biologists (consultable at https://prelights.biologists.com/prelists/3d-gastruloids/). Testifying to the explosive growth of the field, to date this list counts 26 entries in preprinted research alone, most of which were published just in the last couple of years. Accordingly, a summary timeline of the major milestones in contemporary gastruloid developmental biology is provided in Fig.1.3, with the topic having been most recently reviewed in Brink and Oudenaarden, 2021. Briefly, gastruloids have now been shown to develop mesodermal domains analogous to the embryonic cardiac crescent (Rossi et al., 2021a) and to capture some features of embryonic blood development (Rossi et al., 2021b), as well as long-missing anterior neural identities, either through co-culture with extraembryonic endoderm (Bérenger-Currias et al., 2020) or by growth in Wnt-suppressive conditions (Girgin et al., 2021). Most notably, recent variations on gastruloid culture, that embed them or expose them to the ECM extract Matrigel (Hughes et al., 2010), have found that the presence of ECM seems to be able to "unlock" morphogenic processes up to then unseen. Accordingly, gastruloids embedded in Matrigel develop into somitogenic models with visible segmentation and embryonic-tailbud-like patterning dynamics (Brink et al., 2020) and gastruloids exposed to ECM components during the last days of culture form so-called Trunk-Like Structures comprising a structured neural-tube-liked domain flanked by segmented somitic mesoderm. (Veenvliet et al., 2020). Importantly, culture conditions have now been found to generate gastruloids from human PSCs (Moris et al., 2020; Libby et al., 2021) and from human induced PSCs (Olmsted and Paluh, 2021). Most recent

applications have been using gastruloids as a screening platform for teratogenic compounds (Mantziou et al., 2021), and they have as well importantly gone back to the earliest steps of symmetry breaking to provide more accurate and comprehensive comparisons between the cell types that populate gastruloids and those that populate embryos (Anlaş et al., 2021).

While there are now a variety of gastruloid systems (and in fact not just made of mouse or human cells; see Cheng et al., 2021; Torres-Paz and Rétaux, 2021; Fulton et al., 2020), they all revolve around the same fundamental structure: an initial population of PSCs that self-organises to establish all three embryonic axes and to pattern these axes according to the blueprint of its (post-occipital) *in vivo* counterpart. While structures more anterior than the heart are not yet observed to form without external intervention, all these developmental events take place *in vitro* without the need or presence of extra-embryonic tissues (Turner et al., 2017; Brink and Oudenaarden, 2021).

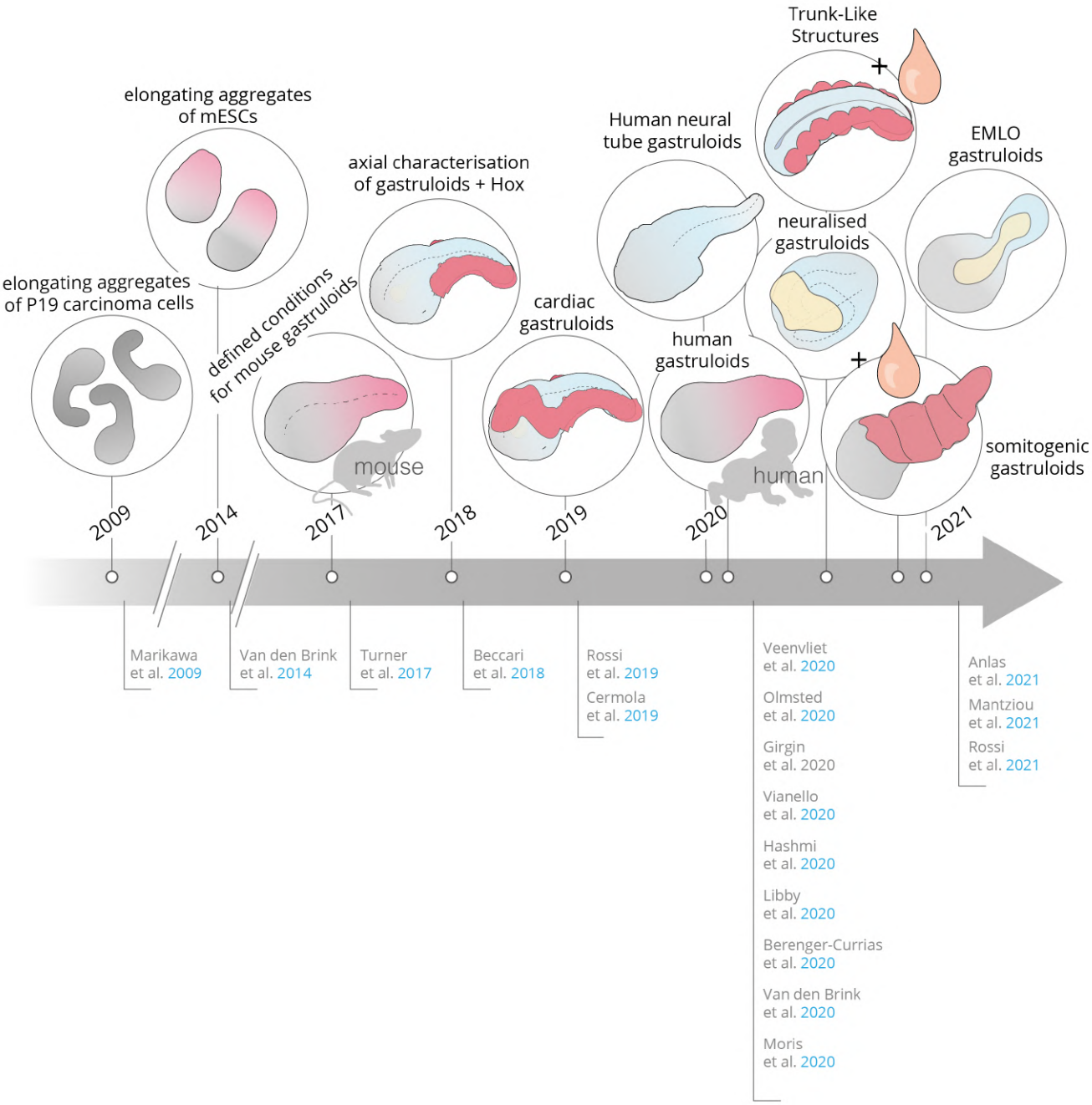


Fig. 1.3.: Timeline and progress in the field of gastruloid research. In the top half, major improvements and variations on the original gastruloid system (*i.e.* on the triaxial, Hox-patterned gastruloids of Beccari et al., 2018). Orange drop represents addition of Matrigel (ECM components). The papers referenced in the figure are listed as follows: Marikawa et al., 2009; Brink et al., 2014; Turner et al., 2017; Beccari et al., 2018; Rossi et al., 2021a; Cermola et al., 2021; Veenvliet et al., 2020; Olmsted and Paluh, 2021; Girgin et al., 2021; Vianello and Lutolf, 2020; Hashmi et al., 2020; Libby et al., 2021; Bérenger-Currias et al., 2020; Brink et al., 2020; Moris et al., 2020; Anlaş et al., 2021; Mantziou et al., 2021; Rossi et al., 2021b. The dates indicated in the figure are those when the studies were first published online (preprinted). EMLO: Elongating Multi-Lineage Organized

2 Endoderm Emergence in gastruloids

> *Because I must know what I am...and how can I know that if I do not know what I came from?*
>
> **— Nick Joaquin**
> "The woman who had two navels" (Joaquin, 2017)

In light of the points made in the Introduction, and to study endoderm development *in vitro*, I started by generating gastruloids. To build on previous reports and body of work established by our and other labs, and mindful of known differences between gastruloids obtained with different cell lines, I initially focused on gastruloids generated from the Sox1/TBra double reporter (SBR) cell line described in Deluz et al., 2016 and whose gastruloids have been extensively characterised in terms of axial patterning and gene expression dynamics (Beccari et al., 2018). Crucially, while a number of early and mature endodermal markers had been recovered, uncertainty remained on their spatial organisation within the gastruloids, on the identity of the cells expressing and possibly co-expressing such markers, and on the developmental origins and dynamics of such cells in the gastruloid (Beccari et al., 2018). In fact, while the presence of endoderm in gastruloids had been understood, gastruloids were mostly considered as unstructured or poorly-structured neuromesodermal entities, with restricted regions of endoderm either restricted at the posterior (Brink et al., 2014), or to undefined more anterior domains (Beccari et al., 2018; Turner et al., 2017).

The gastruloid-generation protocol, now published (10.17504/protocols.io.9j5h4q6), is illustrated in Fig.2.1 and is detailed in the Materials and Methods section at the end of this chapter. Briefly, mESCs maintained in ground pluripotency (Silva et al., 2008) are dissociated into a single cell suspension, which is then dispensed in individual wells of a low-adhesion 96well plate. The specific number of mESCs deposited in each well — here 300 — is then left to aggregate, out of pluripotency conditions and in N2B27 base medium, undisturbed. After 48h, the aminopyrimidine derivative CHIR is added to the medium. As CHIR is a GSK3βinhibitor and thus effectively recapitulates the downstream consequences of Wnt signalling activation

(*i.e* β-catenin stabilisation; Ye et al., 2012), this 24h-long pulse (pink in Fig.2.1) mimics the high Wnt signalling levels that characterise the posterior of the EPI *in vivo* and that trigger embryonic symmetry breaking through PS initiation (Arnold and Robertson, 2009; Tam and Behringer, 1997; Tam and Loebel, 2007). In gastruloids, pulsed (but here symmetric) delivery of CHIR leads to widespread TBra expression (Turner et al., 2017), followed by sorting and possible selective downregulation by a portion of the cells, eventually resulting in polarisation of TBra expression (Anlaş et al., 2021). The end of the gastruloid to which TBra ends up polarising is thus defined as the posterior in analogy to the expression domain of this marker in the embryo (Kispert and Herrmann, 1994), and TBra expression remains associated with the posterior tip of the now elongating gastruloid (Turner et al., 2017; Beccari et al., 2018; Brink et al., 2014). The gastruloid quickly transitions from a spherical aggregate of cells to a tear-drop shape, to an elongated structure with morphologically defined anterior and posterior (Fig. 2.1).

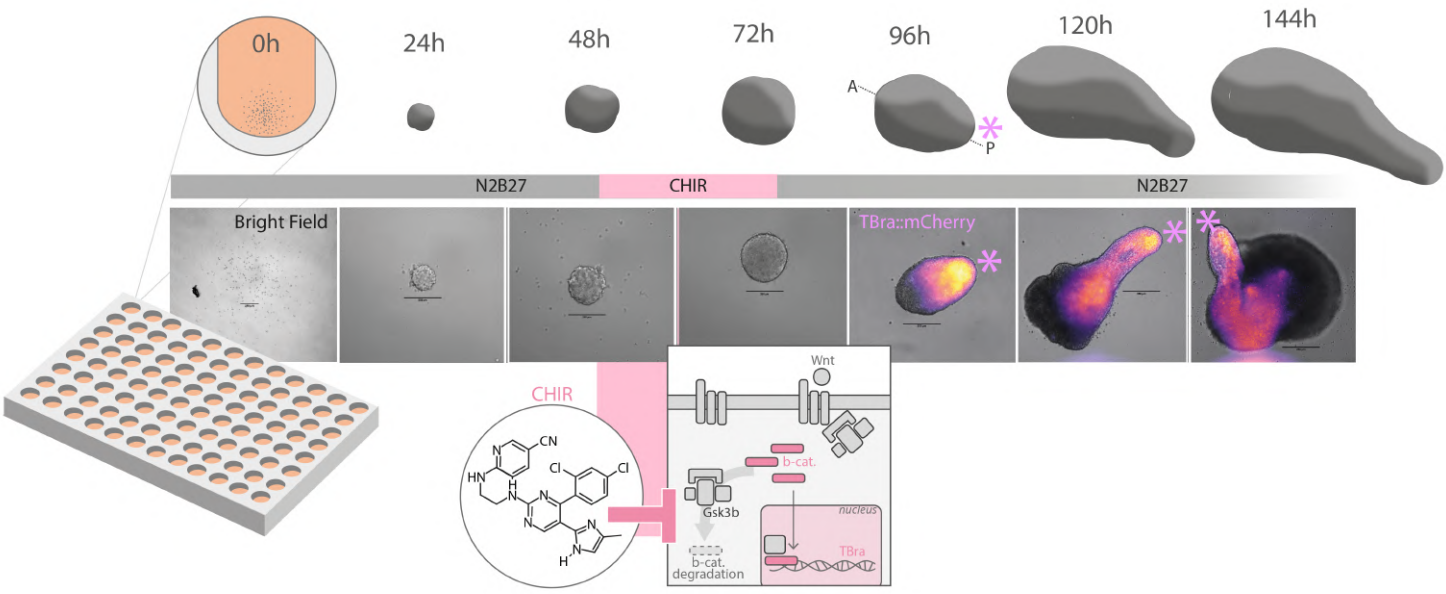


Fig. 2.1.: Gastruloids of mESCs break symmetry and elongate in response to pulsed CHIR: Illustration and Brightfield images of typical gastruloid development. Single cells initially left to aggregate in a well of a low-adhesion 96well plate (0h, leftmost) form a spherical aggregate by 24h, and then break symmetry and undergo morphological changes (elongation) in response to pulsed delivery of CHIR from 48h to 72h of development. Symmetry breaking is reported by polarised TBra expression (here imaged 96h onwards; TBra is homogeneously and lowly expressed at t=72h, see Turner et al., 2017), with the posterior marker remaining associated with the elongating posterior of the gastruloid. A simplified diagram of the mode of action of CHIR as a GSK3βinhibitor is also included. Scale bars: 250μm. A: Anterior, P/asterisk: Posterior. Note: parts of this figure are also published in Vianello and Lutolf, 2020, CC-BY 4.0

2.1 Emergence and early patterning of endoderm progenitors

When and where do endoderm cells first emerge in the gastruloid? To answer this question, I first carried out a series of timed immunostainings for key early endodermal markers FoxA2 and Sox17, before and after the morphological symmetry breaking of the gastruloid, *i.e.* around the time of gastruloid "gastrulation" (Fig.2.2 A). If gastruloids truly mimic mouse perigastrulation development, endodermal progenitors should be specified around the same timeframe as that when early mesodermal cell types are produced. Also for this reason, the expression of FoxA2 and Sox17 was analysed in relationship to that of TBra, taken as a marker of the posterior EPI, of the PS, and of mesodermal lineages as these emerge from it (Kispert and Herrmann, 1994). The choice of FoxA2 and Sox17 as endodermal markers is grounded both on the centrality of these markers in differentiating early stages of endodermal emergence in the mouse embryo *in vivo* (Burtscher and Lickert, 2009; Viotti et al., 2014), as well as them frequently being the markers used to impute endodermal identity in previous gastruloid studies (e.g. Brink et al., 2014), and thus to allow comparison with- (and building on) previous data in the field.

Having said that, and while co-expression of different combinations of these markers can indeed identify unique populations, neither FoxA2 nor Sox17 are unique markers of endoderm alone. Notably, Sox17 also marks cells of the early endothelial lineage (Choi et al., 2012), and FoxA2 can also marks the notochord (classified as mesoderm) and the neural floorplate (ectoderm; Monaghan et al., 1993) as well as, as noted previously, ventricular cardiac progenitors (Bardot et al., 2017; Ivanovitch et al., 2021; Gonzalez et al., 2021). The caveats of identity-inferences in gastruloids are further accentuated by the uncertain timing of gastruloid development (Beccari et al., 2018), and the possibility that cell type that emerge asynchronously in the embryo could instead be appearing within the same timeframe in gastruloids.

As shown in Fig.2.2 B.1, FoxA2$^+$ cells can be detected already at 72h, where they appear in a salt and pepper distribution intermingled with TBra$^+$ cells. Double-positive cells can also already be found (see green/magenta insets in Fig.2.2 B.1). 24h later, on the other hand, FoxA2$^+$ cells clearly segregate away from TBra$^+$ ones and appear to distribute all along the major (AP) axis of the now elongated gastruloid. FoxA2$^+$ cells form small clusters with like cells, while TBra$^+$ cells, as expected (Brink et al., 2014), segregate instead to one end of the aggregate (*de facto*, the posterior). Even more interestingly, and unlike FoxA2$^+$ cells, Sox17$^+$ cells cannot be detected at 72h, and instead only start emerging at the 96h timepoint

(Fig.2.2 B.2). When co-staining for FoxA2, these Sox17^{+} cells mostly appear as a subpopulation of the FoxA2^{+} cells (Fig. 2.2 B.3).

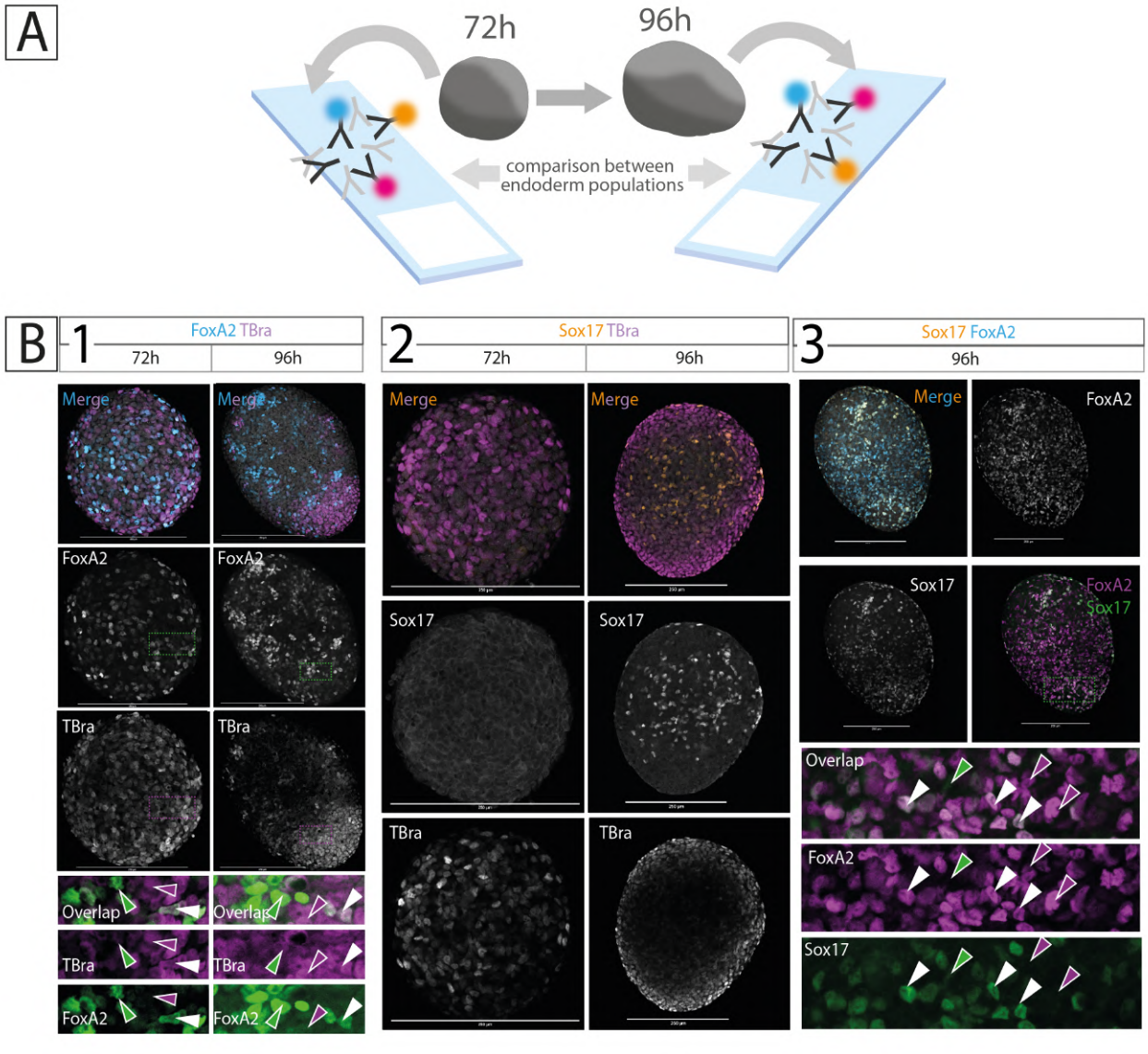


Fig. 2.2.: Emergence and patterning of endodermal markers in gastruloids. **A)** Schematic of the experimental approach. **B)** Immunostaining of gastruloid for markers of endoderm and PS in the embryo. **1.** Immunostaining of t=72h and t=96h gastruloids for TBra (magenta) and for early endoderm marker FoxA2 (cyan). **2.** Immunostaining of t=72h and t=96h gastruloids for TBra (magenta) and later endoderm marker Sox17 (orange). **3.** Immunostaining of t=96h gastruloids for FoxA2 (cyan) and Sox17 (orange). To show colocalisation between markers, green/magenta insets are also provided. Green/magenta arrowheads: single-positive cells for either marker; white arrowheads: double-positive cells. Scale bars: 250μm. Note: parts of this figure are also published in Vianello and Lutolf, 2020, CC-BY 4.0

To better assess co-expression of the markers, and the spatial distribution of cells expressing different levels of each marker under consideration, I developed a dedicated co-expression analysis pipeline. This pipeline is illustrated in Fig. 2.3 A, and detailed in the Materials and Methods section. Briefly, it uses the deep-learning based algorithm Cellpose (Stringer et al., 2021) to segments gastruloid cells expressing either or both markers under analysis, the segmentation output is converted to Region of Interest (ROI)s, and intensity values of each channel are thus measured on a cell-by-cell basis. Replotting single- and double-positive ROIs as an overlay of the DAPI channel (all nuclei) also provides a spatial map of the distribution of single and double-positive cells with respect to the AP- and centre-periphery axes of the gastruloid.

Furthermore, to quantify the robustness of the AP patterning of markers in immunostained gastruloids, at a population-level scale rather than just at the single gastruloids level, I also developed a second dedicated pipeline (*cfr.* Fig.2.3 B), detailed in the Materials and Methods section and now published (10.5281/zenodo.4899121). Briefly, the pipeline takes as inputs whole-population images and hand-drawn line ROIs corresponding to the AP axis of each gastruloid in the sample. By mapping pixels of each gastruloid to corresponding positions along this AP axis, the pipeline can output, in a variety of formats, ordered intensity data for each of the fluoresce channels under consideration. The spread of the data (or confidence interval, if displayed) at each given axial position thus also communicates a measure of variability of signal intensity at that position across the entire population.

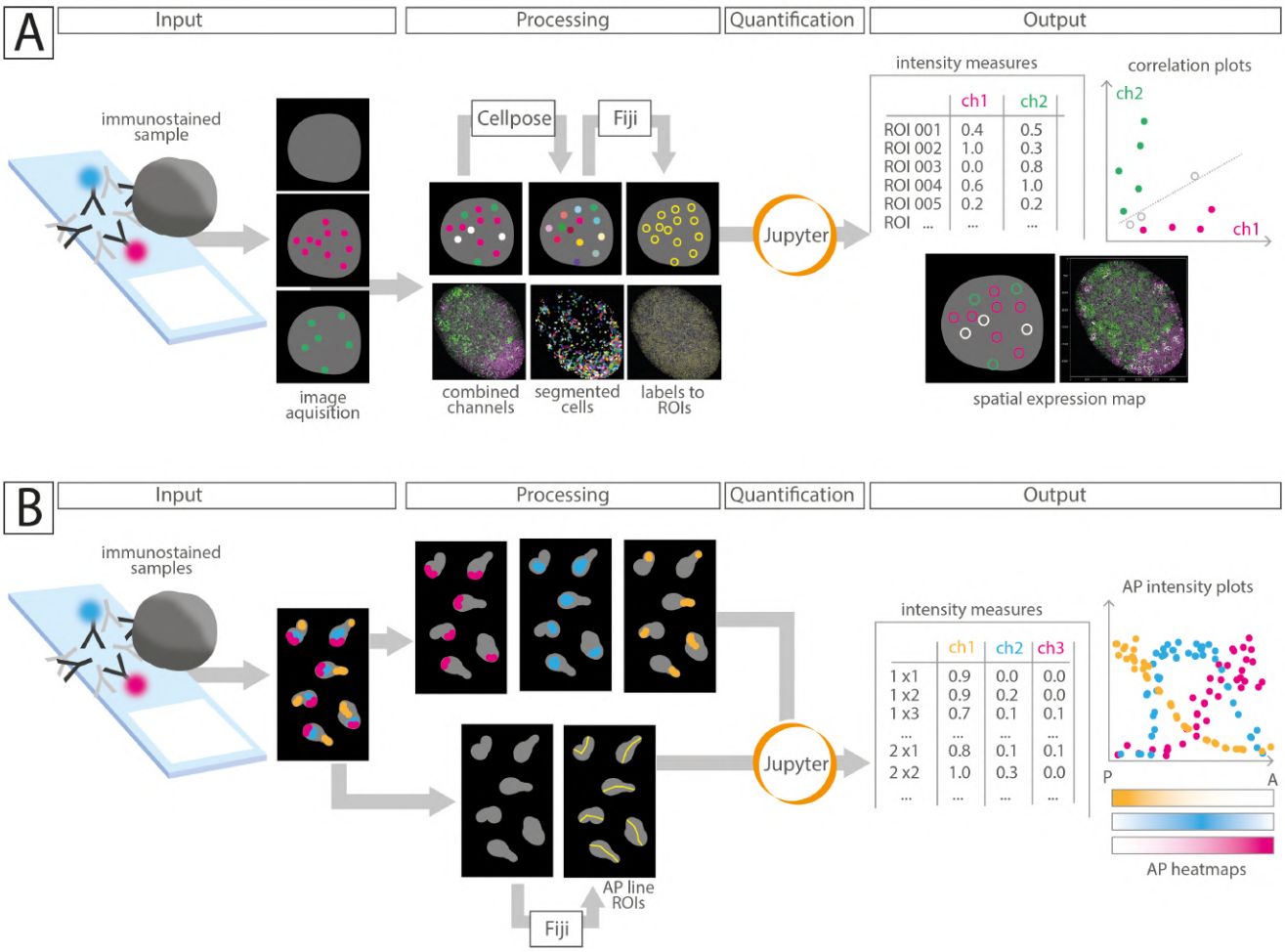


Fig. 2.3.: Pipelines developed for the quantification of coexpression and colocalisation data, and of AP patterning, from immunostaining data. **A)** Summary illustration of the colocalisation pipeline. Taking the midplane section of an immunostained gastruloid as input, cellpose is used to segment cells positive in either of the marker channels, and the labels are converted to ROIs through the Fiji plugin "LabelsToROIs" (see Waisman et al., 2021). Quantification is performed by an interactive Jupyter notebook, which outputs raw marker expression data for each cell, correlation plots, as well as a spatial visualisation of cells classified as single- or double-positive. **B)** Summary illustration of the AP patterning quantification pipeline. Taking marker expression data from entire populations of gastruloids, as well as hand-drawn line ROIs of the AP axis of each gastruloid, the interactive Jupyter notebook outputs raw marker expression data for each cell, and scatterplots, lineplots, and/or 1D heatmaps indicating the AP distribution of each marker expressed.

Application of the colocalisation quantification pipeline to gastruloids immunostained for endodermal markers at 72h and 96h, shows that while at the earlier timepoint gastruloids display a majority of FoxA2$^+$/TBra$^+$ double positive cells, all located within the peripheral TBra$^+$ domain of the aggregate (Fig. 2.4 A.1), these two markers quickly segregate by t=96h (Fig. 2.4 A.2). Crucially, the analysis highlights that those cells that are double positive at t=96h all reside at the interface between the TBra$^+$ and the FoxA2$^+$ domains (Fig. 2.4 A.2). Quantification also supports the observation that emerging Sox17$^+$ cells (at 96h) do so as a nearly exclusive subpopulation of the FoxA2$^+$ population (Fig. 2.4 B). Still, some Sox17 single-positive cells do also exist within the gastruloid, towards the outer perimeter of the double-positive zone (Fig. 2.4 B). Finally, population-wide quantification of AP patterning shows that the axial segregation of TBra and FoxA2 expression established in t=96h gastruloids is a robust property at the population level, with the TBra$^+$ posterior giving way to a more anterior FoxA2$^+$ domain just posteriorly to the mid of the AP axis (Fig. 2.4 C).

The relevance of these findings is significant. Here is a self-organising system where, despite it not representing the organisation, modes of gastrulation, or even the initial cell types of the developing embryo (Beccari et al., 2018; Anlaş et al., 2021), cells differentiate into endodermal lineages in temporal and spatial patterns analogous to what known to occur *in vivo*. Specifically, FoxA2$^+$ cells emerge in response to pulsed CHIR and within a TBra$^+$ context, to then segregate anteriorly to a posterior TBra$^+$ domain. Sox17$^+$ cells emerge only later, mostly as a subpopulation of FoxA2+ cells. In the temporal and spatial patterns of these events *in vitro*, many similarity can thus be drawn with the temporal and spatial modes of emergence of endoderm *in vivo* where, as detailed in the Introduction section, FoxA2$^+$ cells first emerge just prior to gastrulation intermingled with TBra$^+$ cells within the posterior EPI, and then can be found anteriorly (distally) to the TBra$^+$ domain once gastrulation is underway (Burtscher and Lickert, 2009). Sox17$^+$ cells, here too, emerge later and as a subpopulation of FoxA2$^+$ cells (Viotti et al., 2014).

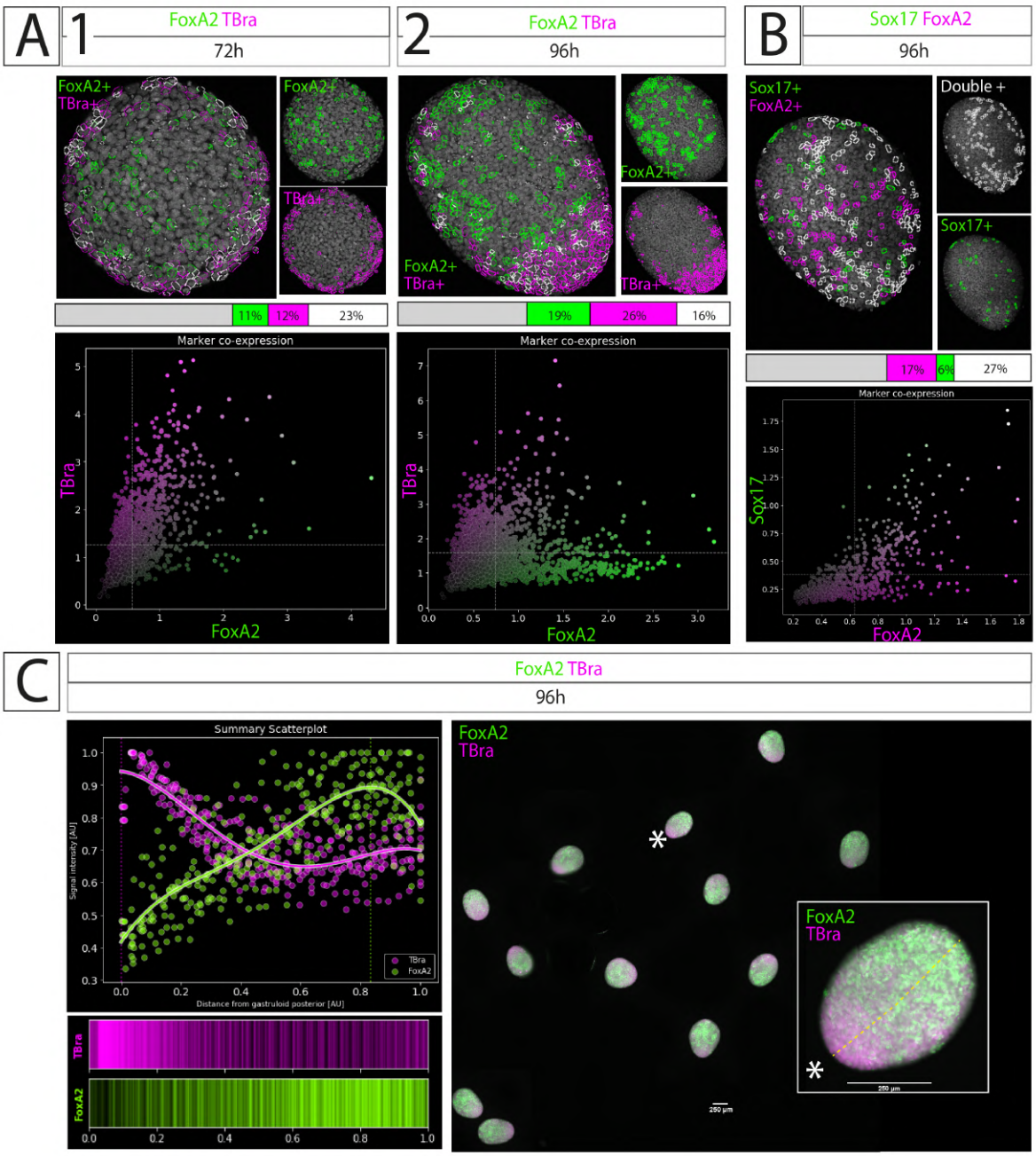


Fig. 2.4.: Quantification of endodermal markers to recover gastruloid- and population-level patterns of expression and co-expression **A)** Colocalisation analysis of the pattern of expression of early endodermal marker FoxA2 (green), and PS/early mesoderm marker TBra (magenta) in gastruloids at t=72h (1) and t=96h (2). **B)** Colocalisation analysis of early endodermal marker FoxA2 (magenta) and late endodermal marker Sox17 (green) in t=96h gastruloids. In both panels A and B, cells expressing higher-than-average levels of either marker are outlined in green or magenta, and cells expressing higher-than-average levels of both markers are outlined in white. **C)** Population-level analysis (summary scatterplot and 1D-heatmaps) of the AP distribution of FoxA2 (green) and TBra (magenta) in t=96h gastruloids. An AP position of 0 indicates the posterior of the gastruloid, with positions closer to 1 indicating more anterior positions. *n=11* gastruloids. Individual gastruloids are shown on the right. Asterisk indicates representative gastruloid shown in inset.

Where do these FoxA2$^+$/Sox17$^+$ cells go next? While in the embryo the readily intercalate within the overlaying VE (Kwon et al., 2008; Viotti et al., 2014), this is an interesting question in gastruloids, where extraembryonic tissues (thus including the VE) are not believed to be present.

2.1.1 Extraembryonic endoderm in Gastruloids

Gastruloids have been demonstrated to lack extraembryonic tissues (Turner et al., 2017; Brink et al., 2020). Indeed, where single-cell analyses have been performed at early or late timepoints and across cell lines, clear signatures of extraembryonic endoderm could not be detected (Brink et al., 2020; Rossi et al., 2021a). In fact, it is the very absence of extraembryonic endoderm, or maybe more specifically of cells that could supply AVE organising function, that is considered to be the reason gastruloids lack anterior structures and remain as posterioralised (*i.e.* post-occipital) embryonic models (Beccari et al., 2018). In turn, the absence of AVE explains why the development of anterior neural derivatives in gastruloids only appears to take place if the AVE is rescued externally by either culture in a global context of Wnt repression (Girgin et al., 2021), or by co-culture with supplied extraembryonic endoderm cells (Bérenger-Currias et al., 2020). The very observation that gastruloids invariably undergo symmetry breaking and polarise TBra expression despite the lack of extraembryonic endoderm and AVE function, has in fact been one of the earliest insights coming from gastruloid research (Brink et al., 2014; Turner et al., 2016).

Yet, most recent reports in so-called *embryoid* models of *in vitro* embryogenesis — in many ways analogous to gastruloids — have found evidence of extraembryonic endoderm within the aggregates based on recovery of Dolichos Biflorus Agglutinin (DBA) lectin signal (Xu et al., 2021). Curiously, and unlike what expected from embryonic development, these putative extraembryonic endoderm cells would emerge in response to the inductive cues of the posterior signalling centre that induces PS fates and triggers gastrulation-like fates in the system (Xu et al., 2021). To verify whether extraembryonic endoderm might have been missed or mischaracterised by previous literature in gastruloids, I stained gastruloids for DBA, as in Xu et al., 2021, and for the thyroid hormone transporter *Ttr* (Transthyretin/prealbumin), VE marker (Kwon and Hadjantonakis, 2009) and major discriminant of extraembryonic *vs* embryonic endoderm, as recovered from single-cell datasets (Nowotschin et al., 2019b). As can be seen in Fig. 2.5, DBA positivity is recovered throughout the cell aggregate at early timepoints, and in wide-spanning domains in late gastruloids. As DBA, traditional VE marker in the embryo, marks gastruloid domains that far exceed any realistic estimate of the possible VE makeup of these systems, it thus does not

appear specific for VE *in vitro*, or at least in contexts where embryonic stem cells are differentiating. In fact, no DBA^+ cells (and in general no cells at all) were positive for Ttr, used here as a confirmatory marker (Fig. 2.5 A).

Given the strong DBA reactivity of very early aggregates, and to try to understand the nature of DBA^+ cells, I further tested the 2D-cultured colonies of mESCs, grown in ground pluripotency conditions (see Materials and Methods), from which these aggregate are generated from. As shown in Fig. 2.5 B, adherent stem cell colonies are (heterogeneously) DBA^+, while clearly remaining pluripotent (morphology, $Sox2^+$) and not expressing any endodermal marker (Fig. 2.5 C). In light of these observations, it is interesting to notice that while the exact surface targets recognised by DBA on VE remain unknown (apart from the fact it binds terminal D-N-acetylgalactosamine on surface glycoproteins; Etzler and Kabat, 1970), large glycoproteins expressed by embryonic cells have been implicated (Muramatsu, 2017; Nash et al., 2007). Speculatively, the absence of differentiating stem cells in the peri-gastrulation mouse embryo might instead explain why DBA can instead been usable as a VE marker *in vivo* (Sherwood et al., 2007; Noguchi et al., 1982).

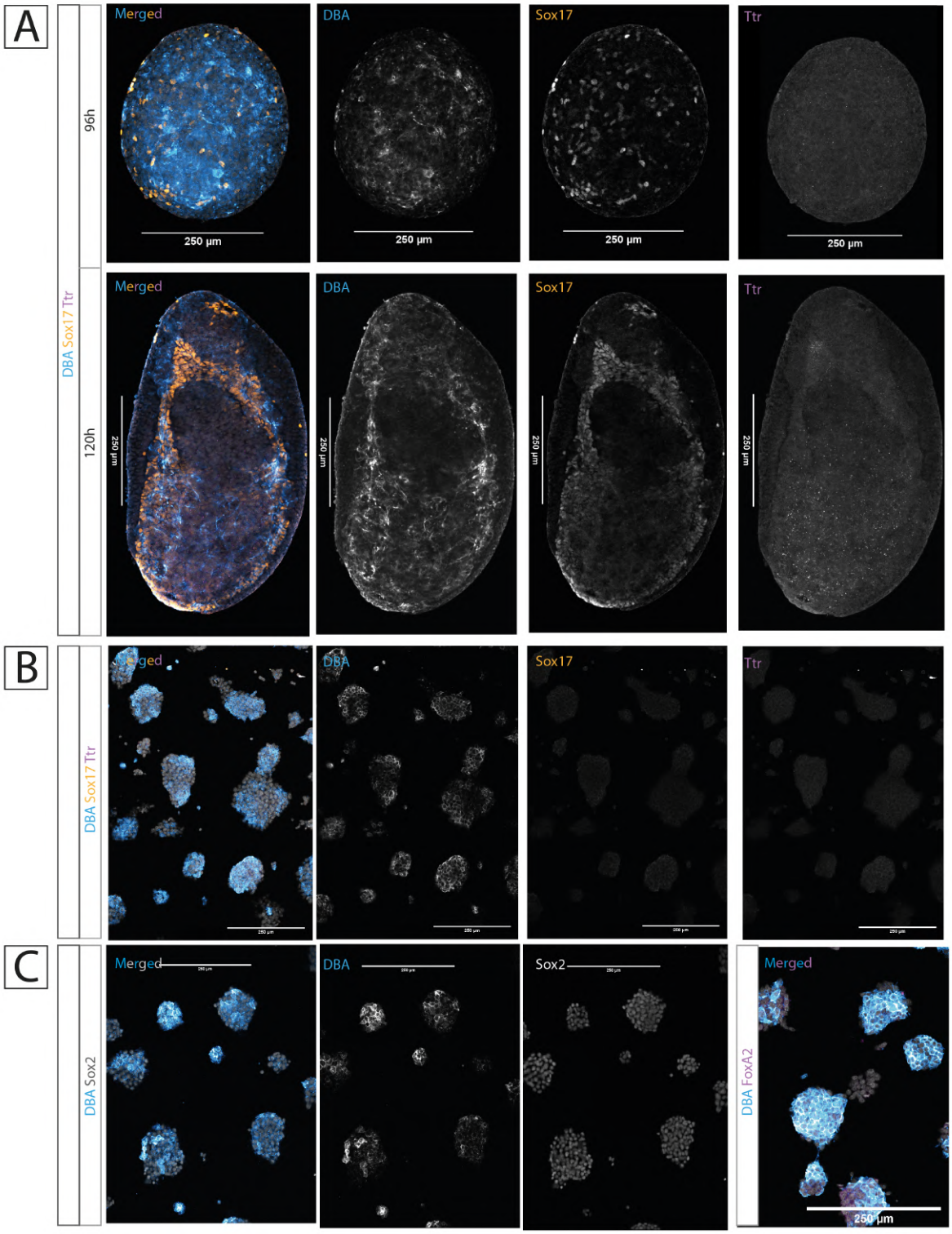


Fig. 2.5.: Both Gastruloids and naive pluripotent stem cells show strong DBA positivity but no Ttr expression: **A)** Gastruloids at t=96h and t=120h immunostained with DBA (cyan, traditional VE marker *in vivo*) and against mature endoderm marker Sox17 and VE marker Ttr. **B)** adherent colonies of mESCs grown for three days in ground pluripotency, immunostained for the same marker as gastruloids in A, or **C)** against pluripotency marker Sox2, or early endoderm marker FoxA2. Cells are negative for all markers except Sox2 and are decorated by heterogeneous levels of DBA targets. Single channel images are provided. Scale bars: 250μm.

2.2 Travelling between compartments

Endoderm cells that fail to intercalate within their target epithelium *in vivo* are believed to be lost within the mesodermal compartment, and to possibly also lose their endodermal identity completely (Viotti et al., 2014; Kanai-Azuma et al., 2002). In the absence of VE, or at least of a distinct epithelial layer of VE in which to intercalate, where do the FoxA2$^+$/Sox17$^+$ cells that appear in t=96h gastruloids go? In which compartment of the gastruloid are FoxA2$^+$ cells emerging in the first place? Does the gastruloid maintain epithelial EPI-like, mesenchymal mesoderm-wings-like, and epithelial VE-like compartments as is the case in the embryo? To address these questions, I performed immunostainings for the epithelial marker Cdh1 (E-cadherin, component of adherens junctions; Niessen and Gottardi, 2008), at the same early timepoints where cells expressing endoderm markers had been found to emerge.

Conflicting with the idea of gastruloids as a mostly mesenchymal aggregate, and as shown in Fig. 2.6 A.1, A.2, gastruloids at t=72h can be seen to be entirely Cdh1$^+$, even though cells are clearly not organised as an epithelium. They express epithelial markers but do not display epithelial architecture. In fact, the widespread expression of Cdh1 throughout early gastruloids likely simply reflect the fact that at this stage they are little more than aggregates of mESCs in their transition to an EpiSC stage (Turner et al., 2017; Hamidi et al., 2020). Consequence of this is that the FoxA2$^+$ cells emerging at t=72h intermingled with TBra$^+$ cells are emerging within a substrate of Cdh1$^+$ cells (as confirmed in Fig. 2.6 A.1 (72h), and A.2 (72h) for TBra and FoxA2 respectively). This landscape could be considered analogous, minus its columnar architecture, to the epithelial context of the embryonic EPI where TBra$^+$ and FoxA2$^+$ progenitors first emerge.

Just one day later however, the Cdh1 landscape of gastruloids is drastically different: the Cdh1 distribution is now patchy and fragmented. While still forming some sort of continuum throughout the gastruloid, especially at the TBra$^+$ posterior (Fig. 2.6 A.1 96h) Cdh1$^+$ tissue is now interrupted anteriorly at multiple points by intervening Cdh1$^-$ cells (Fig. 2.6 A.1, A.2; Cdh1$^-$ cells are interpreted as being mesoderm). Notably, all FoxA2$^+$ cells of the t=96h gastruloid fall within the Cdh1$^+$ patches (Fig. 2.6 A.2). Sox17$^+$ cells, as mostly a subpopulation of FoxA2$^+$ cells, also fall within the Cdh1$^+$domain (Fig. 2.6 A.3). From t=72h to t=96h endoderm cells thus always maintain Cdh1 positivity.

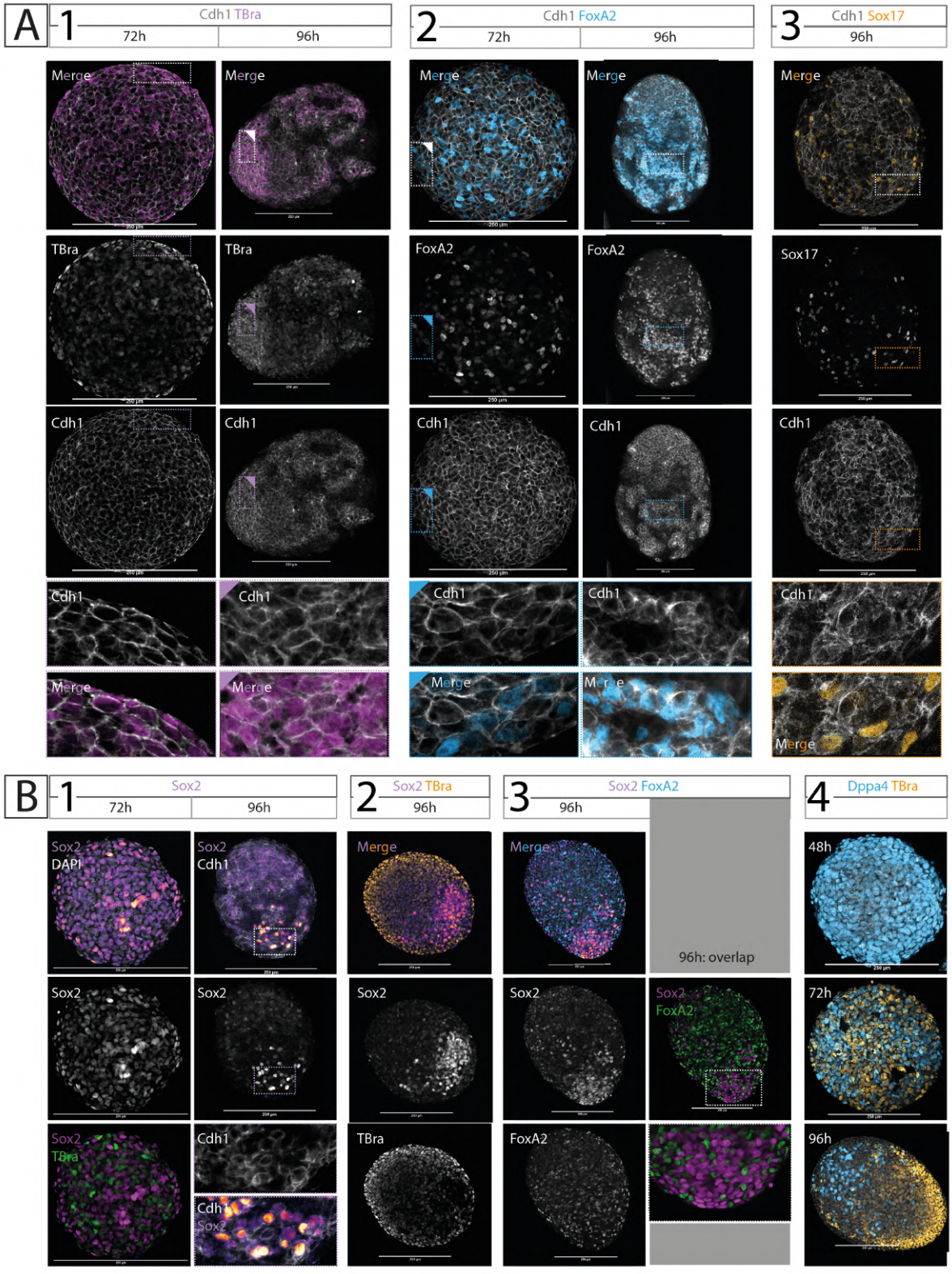


Fig. 2.6.: Endoderm cells reside within an epithelioid domain that undergoes fragmentation. **A):** Immunostainings of gastruloids at t=72h and t=96h for epithelial marker Cdh1 and TBra (**1**), FoxA2 (**2**) and Sox17 (**3**; Sox17 is not expressed at t=72h). **B)** Immunostaining of gastruloids at t=72h for pluripotency marker Sox2 with TBra or Cdh1 (**1**). Immunostaining of t=96h gastruloids for Sox2 and TBra (**2**) or Sox2 and FoxA2 (**3**). Immunostaining for TBra and Dppa4 of gastruloids from t=48h to t=96h (**4**). Scale bars: 250μm. Note: parts of this figure are also published in Vianello and Lutolf, 2020, CC-BY 4.0

To understand whether the Cdh1^{+} cells in which FoxA2^{+} cells emerge could be compared to the potent substrate in which they are known to emerge *in vivo* (*i.e.* the EPI), and to understand whether endodermal cells, in the absence of VE in which to integrate, aggregate instead amongst themselves or with other Cdh1^{+} cells in their immediate local environment, I immunostained for the pluripotency and EPI marker Sox2. Fig. 2.6 B.1 shows that t=72h gastruloids indeed appear to be a mix of intermingled and mutually exclusive Sox2^{+} and TBra^{+} cells. By 96h however, the Cdh1^{+} compartment can be subdivided in a wider territory of Sox2^{LOW} cells, and a more restricted pole of Sox2^{HIGH} cells (Fig. 2.6 B.1). Co-staining for the posterior marker TBra shows that this Sox2^{HIGH} pole resides at the anterior of the gastruloid (Fig. 2.6 B.2). FoxA2^{+} cells that occupy the fragmenting Cdh1^{+} archipelago of the gastruloid at t=96h, are nonetheless excluded from the Cdh1^{+}/Sox2^{HIGH} anteriormost pole (Fig. 2.6 B.3). While it is thus very possible that this cluster of Sox2^{HIGH} cells represents *de novo* Sox2 expression by cells that were Sox2^{-} at earlier timepoints, the pattern of expression is very similar to the pattern observed from timed immunostaining of pluripotency marker Dppa4 , which also marks extensive regions of the aggregates at 48h, is then distributed in a salt and pepper fashion with TBra at t=72h, and is then preserved anteriorly at t=96h (Fig. 2.6 B.4). As surprising as this maintenance of pluripotency is — amongst such an environment of active cell differentiation — it thus appears that FoxA2^{+} cells are emerging within a pluripotent epithelial substrate, and are later segregating from it while maintaining interactions with like-cells as they differentiate.

Of notice, recent single-cell analyses have indeed highlighted that the "EPI" is amongst the least conserved cell identities in gastruloids, pointing towards the possibility that the starting substrate of gastruloid "gastrulation" and developmental biology is in fact a different one than the one in the embryo, even if the following cell type specification processes may well converge to the same outcomes (mesoderm, endoderm, ectoderm; (Anlaş et al., 2021)).

2.3 EMT in Gastruloids

Is the continued association of endodermal cells with Cdh1^{+} territories due to a failure of undergoing EMT, and as such self-sorting and retention within the original EPI-like compartment? Or are these endoderm progenitors indeed undergoing EMT and, unable to find a VE layer in which to intercalate, re-entering the Cdh1^{+} tissue that they have left?

To investigate the modes of EMT taking place in "gastrulating" gastruloids, I immunostained the aggregates before and after the observed Cdh1 fragmentation (*i.e* at t=72h and at t=96h, respectively). While nuclear Snai1 cannot be detected in t=72h (epithelioid) aggregates, cells in the $Cdh1^-$ patches that start fragmenting the epithelioid compartmet of t=96h gastruloids all express Snai1 (Fig. 2.7 A.1). $Snai1^+$ cells distribute in a complementary fashion to $Cdh1^+$ patches (in which $FoxA2^+$ cells reside) and Snai1 can be detected within some $Cdh1^+$ cells. At the gastruloid-level, $Snai1^+$ cells appear to form the anterior (and cortical) extreme of a continuum with cells expressing TBra at the posterior of the gastruloid (Fig. 2.7 A.2). Based on these results, it appears that EMT is a diffuse process happening throughout the gastruloid, at the edges of the fragmenting Cdh1 mass. Cells that express Snai1 are associated with the $Cdh1^-$ compartment. Consequently, $FoxA2^+$ are likely not undergoing EMT (or, at least, Snai1-mediated EMT).

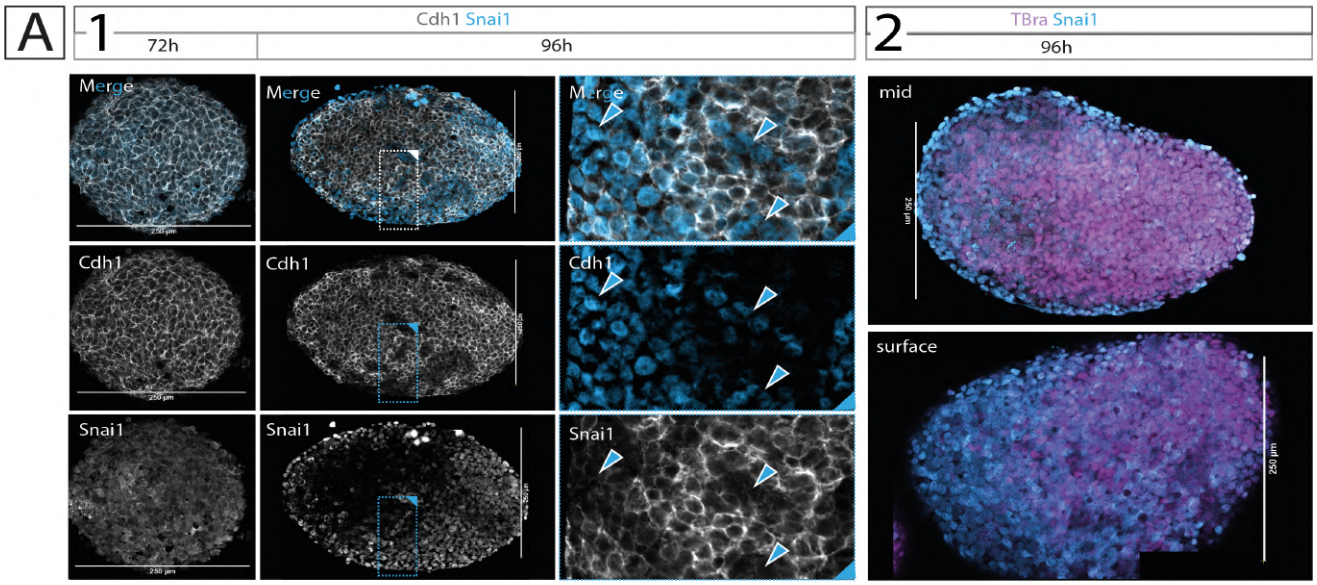


Fig. 2.7.: Fragmentation of gastruloid $Cdh1^+$ cells is associated with widespread Snai1-mediated EMT. **A) 1:** Immunostaining of t=72h and t=96h gastruloids for epithelial marker Cdh1 (white) and for classic EMT mediator Snai1 (cyan). Cyan arrowheads: $Snai1^+/Cdh1^-$ cells**2:** Midplane (top panel) and surface (bottom panel) optical slices of t=96h gastruloid immunostained for posterior marker TBra (magenta) and for Snai1 (cyan). Scale bars: 250μm. Note: parts of this figure are also published in Vianello and Lutolf, 2020, CC-BY 4.0

Conclusions:

- Early endoderm cells ($FoxA2^+$) emerge at t=72h in gastruloids exposed to pulsed CHIR. $Sox17^+$ cells emerge at t=96h as a subpopulation of $FoxA2^+$ cells.
- Early endoderm cells first emerge intermingled amongst $TBra^+$ cells. At t=96h, they reproducibly segregate/pattern to the anterior of the gastruloid, anteriorly to the posterior $TBra^+$ domain. $FoxA2^+/TBra^+$ double positive cells are located at the interface between the two cell populations.
- Endoderm cells emerging in gastruloids do so in a system deprived of VE (*i.e.* of extraembryonic endoderm). Incidentally, DBA may not be a reliable VE marker *in vitro*.
- Gastruloids starts as spherical $Cdh1^+$ epithelioid aggregates that could be considered analogous to the potent embryonic EPI. This compartment then fragments by t=96h as $Cdh1^-$ cells emerge. $FoxA2^+$ and ($FoxA2^+$/)$Sox17^+$ cells are persistently $Cdh1^+$.
- $FoxA2^+$, as they are consistently $Cdh1^+$, are likely spared from Snai1-mediated EMT that generates $Cdh1^-$ domains in the rest of the gastruloid.
- The results are consistent with a model whereby endodermal cells either never undergo classical EMT, or undergo Snai1-independent EMT without losing Cdh1 expression and — in the absence of VE — either re-intercalate within the very tissue they have left, or associate with other $Cdh1^+$ endodermal cells in their local neighbourhood.
- Endoderm cells emerge, *in vitro*, in time and spatial patterns analogous to what known to occur in the mouse embryo.

Materials and methods

Note: relevant Materials and Methods are here reported from my own work in Vianello and Lutolf, 2020, licensed under CC-BY 4.0 International license and for which I hold the rights of reproduction.

Cell culture: mESCs ("SBR" Sox1/TBra double reporter cell line described in (Deluz et al., 2016); CRG8 cells of 129P2 background (RRID:CVCL_3987, Mountford et al., 1994) were cultured in tissue-culture-treated, non-gelatinised, 6well plates, in 10%Serum Medium with added 2i and LIF. Cells were split every third day, by washing in PBS-/-, adding Accutase for around 3min RT, and collecting the resulting cell suspension in a clean centrifuge tube. The Accutase of the cell suspension was then diluted out 1:10 in 10% Serum Medium, and cells were pelleted by centrifuging 200xg, 4min, 4°C. After aspirating out the supernatant, the pellet was resuspended in 1mL 10% Serum Medium to a single cell suspension, and cell density was counted with a haemocytometer. Around 65000-75000 cells were transferred to a new well with 2mL pre-equilibrated 10% Serum Medium (6750-7800 cells/cm2). Cells were then left in a humidified incubator, 37°C, 5% CO2, until use or until further splitting 3 days later. NOTE: In most cases, splitting was coupled to Gastruloid generation. In those cases, the cell pellet was resuspended in N2B27 rather than 10% Serum Medium + 2i and LIF. A complete, step-by-step, protocol is available at: https://dx.doi.org/10.17504/protocols.io.7xbhpin. **Recipes**: 10% Serum Medium: 86.8% DMEM, high glucose, with Glu-taMAX (L-Alanyl-Glutamine, final concentration: 3.45mM), 10% ES-grade Foetal Bovine Serum, 100U/mL Penicillin, 100ug/mL Streptomycin, 0.1mM Non Essential Amino Acids, 1mM Sodium Pyruvate, 0.1mM beta-mercaptoethanol 10% Serum Medium + 2i and LIF: add 3uM CHIR99021, 1uM PD0325901,and 100u/mL LIF.

Gastruloid generation: mESCs were washed in PBS-/-, detached from adherent culture with Accutase (around 3min, RT), and collected in a centrifuge tube. The Accutase in the cell suspension was then diluted out 1:10 in 10% Serum Medium, and cells were pelleted by centrifuging 200xg, 4min, 4C. The supernatant was removed, and the pellet was washed by resuspension in 10mL PBS-/-. Cells were re-pelleted by centrifuging 200xg, 4min, 4C and washed once more in 10mL fresh PBS-/-. After re-pelleting once more (200xg, 4min, 4C), the pellet was dissociated as a single-cell suspension in 1mL N2B27 Medium. Cells were counted with a haemocytometer, and, for each plate of Gastruloids made, 37500 cells (SBR line) or 93750 cells (TFoxA2 line) were transferred to 5mL fresh N2B27 (7.5 cells/uL or 18.75 cells/ul final concentration, respectively). The cell suspension was distributed as 40uL droplets (=300 SBR cells/droplet, or 750 TFoxA2 cells/droplet)

in wells of a U-bottomed, low-adhesion, 96 well plate, and the plates were left for 120h in a humidified incubator, 5% CO2, 37°C. At 48h after plating, 150uL of 3uM CHIR99021 N2B27 were added to each well, and this solution was substituted with fresh N2B27 (no CHIR) every 24h after that. A step-by-step detailed protocol is available at: https://dx.doi.org/10.17504/protocols.io.9j5h4q6. **Recipes**: 10% Serum Medium: 86.8% DMEM, high glucose, with 3.97mM GlutaMAX (L-Alanyl-Glutamine, final concentration: 3.45mM), 10% ES-grade FBS, 100U/mL Penicillin, 100ug/mL Streptomycin, 0.1mM Non Essential Amino Acids, 1mM Sodium Pyruvate, 1mM beta-mercaptoethanol. N2B27: 47.4% Neurobasal Medium, 47.4% DMEM/F-12, with 2.50mM GlutaMAX (L-Alanyl-Glutamine, final concentration: 1.18mM), 1mM GlutaMAX Supplement (total concentration: 2.18mM), 100U/mL Penicillin, 100ug/mL Streptomycin, 0.1mM Non Essential AminoAcids, 1mM Sodium Pyruvate, 1mM beta-mercaptoethanol, 1% B27Supplement, serum-free, 0.5% N-2 Supplement.

Gastruloid immunostaining: Gastruloids were collected at every given timepoint, washed in PBS-/-, and fixed in 4% PFA in PBS-/-, for 2h, 4C, on a low-speed or-bital shaker; or 45min, RT, static. Gastruloids were then washed in PBS+FT (PBS-/-, 10% ES-grade Foetal Bovine Serum, 0.2% Triton-X100), and blocked and permeabilisedin PBS+FT for 1h, RT, static. Primary antibody solutions were then prepared in PBS+FT, with 2ug/mL DAPI. Samples were stained overnight, 4C, on a low-speed orbital shaker. Similarly, secondary antibody solutions were prepared in PBS+FT, 2ug/mL DAPI, and samples were stained overnight, 4C, on a low-speed orbital shaker. Gastruloids were mounted in Fluoromount-G mounting medium (no spacers), and slides kept at 4C long term. All antibody solutions were washed away after incubation by washes in PBS+FT. A detailed, step-by-step protocol is available at: https://dx.doi.org/10.17504/protocols.io.7tzhnp6. A list of the primary antibodies used is provided in Table A.6. Secondary antibodies used were all from Thermo Fisher Scientific: donkey anti-mouse IgG Alexa 647 (CAT#A-31571, RRID: AB_162542); donkey anti-rabbit IgG Alexa 488 (CAT#A-21206, RRID: AB_2535792), Alexa 568 (CAT#A-10042, RRID: AB_2534017), or Alexa 647 (CAT#A-31573, RRID: AB_2536183); donkey anti-rat IgG Alexa 488 (CAT#A-21208, RRID: AB_2535794), goat anti-rat Alexa 568 (CAT#A-11077, RRID: AB_2534121), or Alexa 647 (CAT#A-21247, RRID: AB_141778); donkey anti-goat IgG Alexa 488 (CAT#A-11055, RRID: AB_2534102), or Alexa 568 (CAT#A-11057 RRID: AB_2534104).

mESCs colony immunostaining: mESCs were grown under normal cell culture conditions (see "Cell culture" section above) but seeded on wells of a 6-chamber Ibidi slide. After three days of growth, the culture medium was removed and cells were fixed in 4% PFA in PBS-/-, for 45min RT. The rest of the immunostaining protocol is as reported for gastruloids above.

DBA staining: since the DBA staining protocol used by Xu et al., 2021 was not available, DBA staining of mESCs colonies and gastruloids was done accordingly to what reported in Scheibner et al., 2021. Briefly, gastruloids and mESCs colonies were processed as per immunostaining (see corresponding sections above). DBA lectin was applied dissolved 1:500 (5 ug/mL) in PBS+FT (PBS-/-, 10% ES-grade Foetal Bovine Serum, 0.2% Triton-X100) in place of the primary antibodies; and a solution of Alexa-Fluor647-streptavidin 1:1000 (1ug/mL) in PBS+FT was applied in place of the secondary antibodies.

Imaging and image processing: Bright-field images of Gastruloids were taken on either a Nikon Ti inverted spinning-disk microscope or an Olympus CellR inverted widefield microscope (UPLAN S APO 10x/0.40 air objective, CCD Grayscale Hamamatsu ORCA ER B7W Camera; Olympus XCellence software for data capture). Both microscope setups had CO2 and temperature control (37°C and 5% CO2). Immunostained Gastruloids were imaged on a Zeiss LSM700 inverted confocal microscope (Plan-Apochromat 20x/0.80 air objective, motorized stage, LED Lumencor SOLA Illumination, CCD Grayscale Axiocam MRm (B/W) Camera; ZEN 2009 software for data capture) or on a Zeiss LSM780 inverted confocal microscope (for the CDH1 projection, Plan-Apochromat 20x/0.80 air objective, motorized stage). Images were opened, stitched, and processed for publication (LUT assignment, channel display, min and max intensity thresholding based on no-primary control) using the Fiji ImageJ distribution (Rueden et al., 2017; Schindelin et al., 2012), and the "Grid/Collection Stitching" plugin therein (Preibisch et al., 2009).

Quantification of AP patterning: Batch quantification of immunostaining signal intensity along the AP axis of the Gastruloids was performed through a custom processing pipeline available as a Jupyter notebook at https://doi.org/10.5281/zenodo.4899121 and outlined as follows (step-by-step walk through provided in the notebook itself). The pipeline takes two inputs: i) the multichannel raw image resulting from the scan of an entire microscope slide of immunostained and mounted Gastruloids (here acquired on a GE Healthcare IN Cell Analyzer 2200 automated microscope) and ii) hand-traced line coordinates defining the central axis of each Gastruloid on the slide (starting from the posterior). At early timepoints where the posterior of the Gastruloid is not distinguishable morphologically, the area of TBra polarisation is to be used instead. The script then subdivides each interval of the line ROI provided into n finer intervals of equal length (thus avoiding to have to manually draw a line with high number of points; here n=10), and for each point along the line it defines a non-overlapping polygon mask covering an area of thickness N (here N=500px) across the line and whose lateral edges are orthogonal to the line itself at each side of the point. Having computed the mask, the script then assigns the

total signal intensity recovered in the area to the point of the line ROI around which the polygon was constructed, thus effectively assigning signal intensities to points that can be ordered along an x-axis. These raw values are then normalised by the number of cells in the area (using the DAPI nuclear intensity as a proxy) and both position along the length of the Gastruloid and signal intensity are normalised to the absolute length of the Gastruloid and to the maximal DAPI-normalised intensity value. The script outputs lineplots and scatterplots for each Gastruloid analysed, summary lineplots and scatterplots with collated data of all gastruloids analysed, and the tabulated raw data for re-use.

Quantification of co-expression from immunostaining data: processing of immunostaining signal intensity to quantify and map co-expression of marker pairs was performed through a custom processing pipeline outlined as follows. Initially, the combined intensity data of the two channels stained for each marker was run through the deep-learning based algorithm Cellpose (Stringer et al., 2021) to obtain a mask of individually labelled segmented cells. These cells (considered for analysis) are thus all cells where either marker (or both markers) was found to be expressed by immunostaining. A more comprehensive segmentation of all cells of the gastruloids (including cells expressing neither marker) was not possible due to the poor performance of Cellpose on the DAPI channel of gastruloids (especially at later timepoints). The cell labels were then converted to ROIs through the Fiji plugin "`LabelsToROIs`" (see Waisman et al., 2021). These coordinates, along with the original immunostaining image data, were then passed to a custom Jupyter notebook for quantification. The notebook extracts intensity values on each channel for each cell ROI, and then normalises these values with respect to the highest intensity found within each marker channel, as well as with respect to the intensity measured from the DAPI channel for any given cell. The code outputs tabulated raw data files, and colour-coded scatterplots of expression values of each cell with respect to both markers. The pipeline finally classifies cells as "double-negative" if they express both markers at lower levels than average, "single-positive" for either marker if they express one marker at higher level than average and the other marker at lower level than average, and "double-positive" if they express both markers at higher level than average. The proportions of each of these categories are calculated and returned as output, as well as visual maps where each ROI, colour-coded based on its assigned category, is overlayed to the DAPI image of the sample.

Morphogenesis

3

> *However, in my hands, and as described by others, exogastrulation was a notoriously capricious phenomenon with no well understood pattern of early cell movement*
>
> **— Linda A. Barlow**
> (Barlow, 2001)

3.1 Evolution of the epithelial primordium during gastruloid development

While fragmentation of the original Cdh1 makeup of the gastruloid already starts between t=72h and t=96h (see the previous section), timed immunostaining for Cdh1 in gastruloids cultured for longer timespans (*i.e.* past 96h) documents even more drastic reorganisation of Cdh1$^+$ cells. As shown in Fig.3.1 A, the patchy, fragmented Cdh1 landscape of the t=96h gastruloid appears to converge to the core of the elongated gastruloid to form a "whisk-shaped" structure by t=120h (*i.e.* a mass displaying multiple branches and interrupted by intervening tissue), which then consolidates/resolves into a solid epithelial (Cdh1$^+$) primordium at 144h. At 120h, some of the Cdh1$^+$ branches can be seen emerging to the surface of the gastruloid in the form of rosette openings within the posterior and neck-region (Fig.3.1 A, right). Notably, all cells of this internal Cdh1$^+$ core are FoxA2$^+$ (Fig.3.1 A, bottom row). Highlighted by these immunostainings is thus also the noticeable expansion of the FoxA2$^+$ (and thus possibly endodermal) domain between 96h and 120h (see relevant panels in Fig.3.1 A) in conjunction with the known expansion of mesodermal types at this timepoint (Beccari et al., 2018).

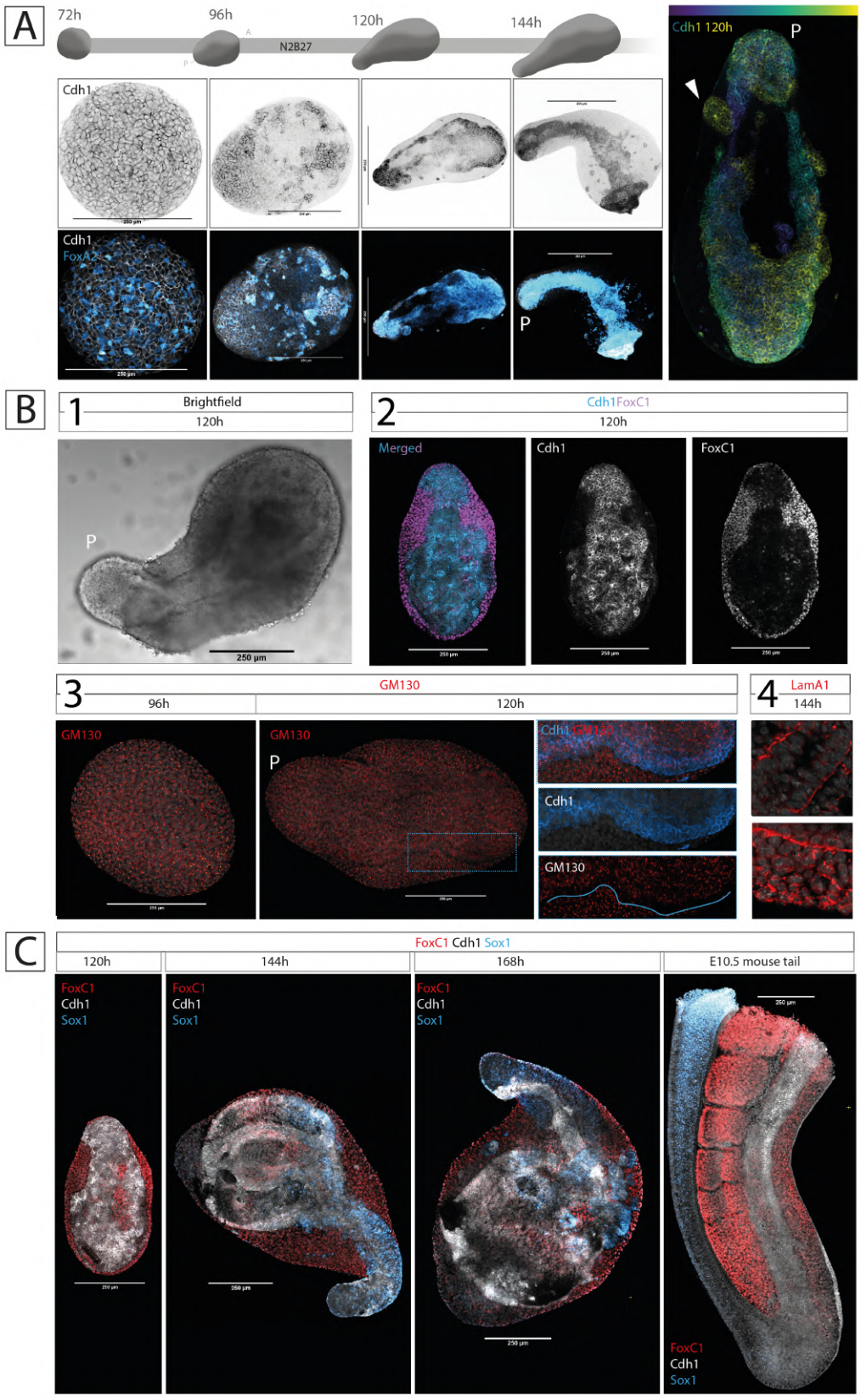


Fig. 3.1.: An epithelial primordium forms at the core of elongating gastruloids. **A)** Illustrative time series of gastruloids immunostained for epithelial marker Cdh1 (top row) and FoxA2 (bottom row, with Cdh1) from t=72h to t=144h. A color-coded Z-projection of the Cdh1 primordium at t=120h is also provided (right). White arrowhead: surface contact of one of the Cdh1$^+$ branches. Caption continues on the next page ↪

Fig. 3.1.: ↪ continued from the previous page. **B) 1:** Brightfield imaging of a t=120h gastruloid. The internal primordium, surrounded by wings of cells is clearly visible. **2:** Immunostaining for the pan-mesodermal marker FoxC1 highlights these wings as mesoderm (surrounding the $Cdh1^+$ primordium). **3:** Immunostaining for Golgi marker GM130 shows polarisation of the primordium by t=120h. **4:** Laminin deposition can be detected in stretches on the basal side of the epithelial primordium (LamA1 staining, red). **C)** Situation of the $Cdh1^+$ primordium (white) with respect to the other germ layers from t=120 to t=168h. The posterior of a E10.5 mouse embryo tail is provided as reference. FoxC1 (red) is here used as a mesodermal marker, Sox1 (cyan) is here used as a neural/neurectoderm marker. Scale bars: 250μm. P: posterior. Note: parts of this figure are also published in Vianello and Lutolf, 2020, CC-BY 4.0

With respect with the global architecture of the gastruloid, the $FoxA2^+/Cdh1^+$ core is enveloped by thick wings of cells at all sides (except at the very posterior) as noticeable even by simple brightfield imaging (Fig.3.1 B.1). These cells can be identified as mesoderm as they result positive for the pan-mesodermal marker FoxC1 (Sasaki and Hogan, 1993; Fig.3.1 B.2). These cells are in fact those that will give rise to the variety of trunk (e.g. somites) and cardiac structures that have indeed been described in other gastruloid work (Rossi et al., 2021a; Brink et al., 2020; Veenvliet et al., 2020).

Structurally, the core $Cdh1^+$ domain appears to develop apico-basal polarity and thus undergo epithelial maturation over time. Whereas $Cdh1^+$ cells of the t=72h Gastruloid represented more an epithelioid state, with expression of epithelial markers but without epithelial architecture, the t=120h and onward $Cdh1^+$ mass shows signs of apico-basal polarity with polarised arrangement of GM130 (Fig. 3.1 B.3), and gradual deposition of discontinuous stretches of laminin (LAMA1 subunit) at the interface with the overlaying mesoderm (*i.e.* its basal side; Fig. 3.1 B.4). Immunostaining for neural/neural ectoderm marker Sox1 (Wood and Episkopou, 1999), in combination with FoxC1, allows to contextualise the location and development of the $Cdh1^+$ primordium with respect to the other germ layers emerging within the gastruloid. As shown in Fig. 3.1 C, $FoxC1^+$ cells keep enveloping the $Cdh1^+$ primordium at all sides at all stages of development (the anteriormost domain, forming a $FoxC1^-$ bump is likely cardiac mesoderm as identified in Rossi et al., 2021a). Sox1 on the other hand only starts being expressed after t=120h, and marks, at 144h, cells of the mid-posterior portion of the $Cdh1^+$ primordium and at t=168h cells that are now $Cdh1^-$ and surround it. The relative positioning of the germ layers is thus not properly conserved in gastruloids (an embryonic reference, if a bit delayed, is provided in the rightmost panel of Fig. 3.1 C) and the posterior part of the $Cdh1^+$ primordium seems to be a shared generative source of cell types other than endoderm, possibly akin to a posterior epiblast equivalent. It appears

very possible that multiple cell types, across germ layers, may be aggregated (and organised along) the Cdh1$^+$ primordium just in virtue of their homotypic surface decoration (*i.e.* all being Cdh1$^+$).

For a better perception of the 3D spatial organisation of the Cdh1$^+$ gastruloid core, Z-projections, individual z-slices, and orthogonal views of gastruloids at both t=120h and t=168h are provided in Fig.3.2. Crucially, orthogonal views show that the Cdh1$^+$ primordium is indeed hollow at the anterior already at t=120h. In fact, while there are cells within the limits of the primordium, these appear to be Cdh1$^-$.

Finally, the internal epithelial structure is consistently seen in all ($n = 97/99$ imaged gastruloids, $N = 6$ independent experiments) samples (examples of different populations can be seen in Fig. 3.3 4). Consistently, t=120h gastruloids develop a central Cdh1$^+$ domain enveloped by FoxC1$^+$ mesenchymal cells (Fig. 3.3 4.A) emerging at the posterior as a Cdh1$^+$/TBra$^+$ posterior-EPI-like domain (Fig. 3.3 4.B). Analogous structures have now also been documented in gastruloids obtained with different cell lines (Gharibi et al., 2020). To provide a better visualisation of the varied branched structures that this Cdh1 domain takes in t=120h gastruloids, z-coded projections are provided in Fig. 3.3 4.C. Notice how all Cdh1$^+$ are also FoxA2$^+$. The domain is maintained in t=144h gastruloids (Fig. 3.3 4.D), with Cdh1$^+$FoxA2$^+$ cells spanning the entire length of the gastruloids, away from a TBra$^+$ posterior.

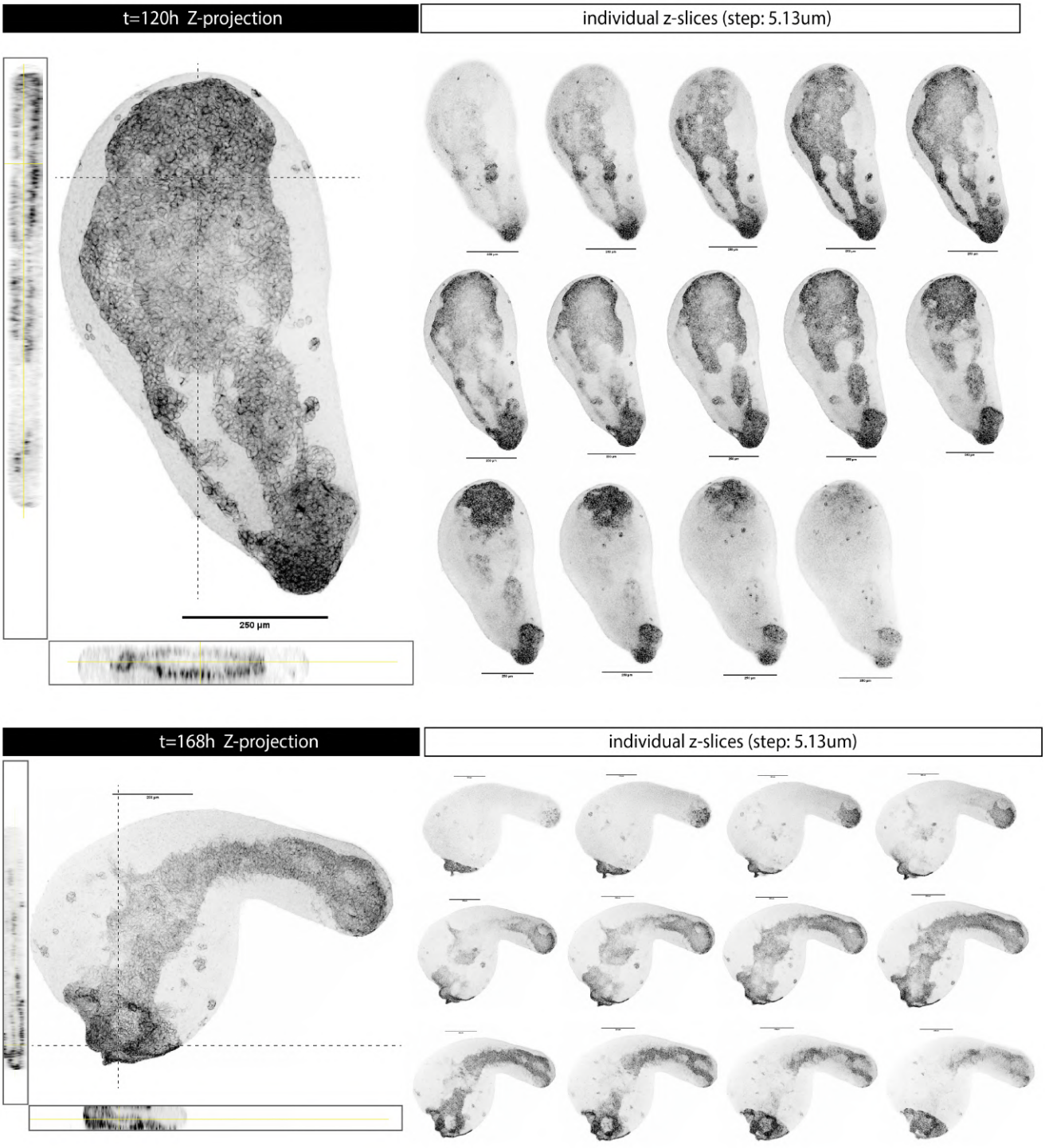


Fig. 3.2.: The complex architecture of the Cdh1^{+} primordium of mature gastruloids. Maximal Intensity Z-projection (left) and individual z-slices (right) of a t=120h gastruloid (**top**) and a t=168h gastruloid (**bottom**) immunostained for epithelial marker Cdh1. Orthogonal views shows that extensive portions of the primordium are hollow (in fact, Cdh1^{-}), specifically at the anterior. Scale bars: 250μm. Z-step: 5.13μm. Note also that the gastruloids are compressed under a coverslip during imaging.

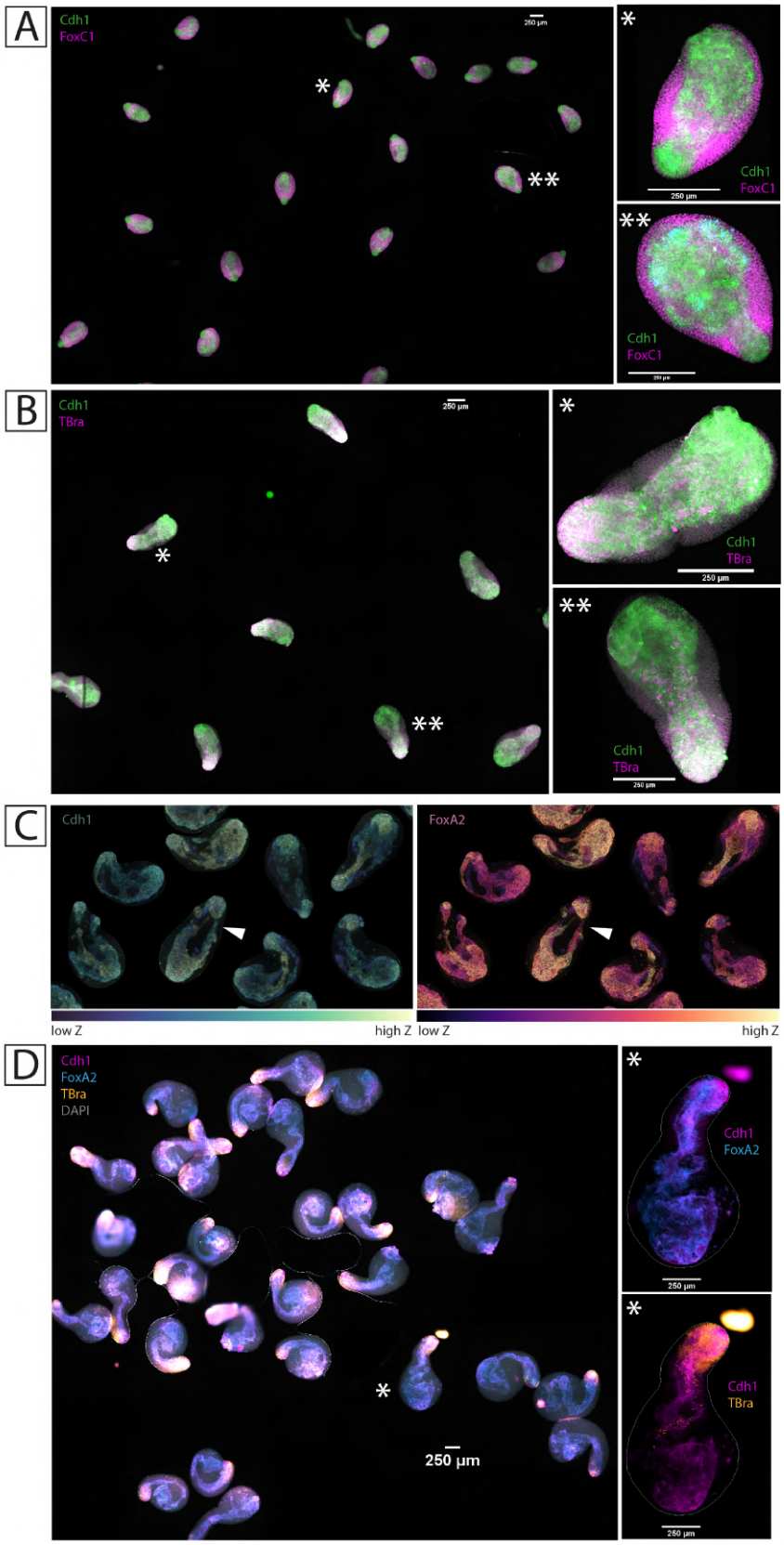


Fig. 3.3.: The Cdh1 epithelial primordium is consistently observed in all TBra/Sox1 double reporter Gastruloids (n=97/99, N=6). See next page for the full caption to this figure ↪

Fig. 3.3.: ↪ continued from the previous page. **A)** Slide scan of t=120h gastruloids immunostained for the epithelial marker Cdh1 (green) and pan-mesodermal marker FoxC1 (magenta). **B)** Slide scan of t=120h gastruloids immunostained for Cdh1 (green) and posterior marker TBra (magenta). **C)** Montage of color-coded Z-projections of t=120h gastruloids immunostained for Cdh1 and endodermal marker FoxA2. White arrowhead indicates the sample shown in Fig.3.1 A. **D)** Slide scan of t=168h gastruloids immunostained for Cdh1, FoxA2, and TBra. Scale bars: 250μm. Asterisks: gastruloids magnified in the insets on the right of each panel. Note: parts of this figure are also published in Vianello and Lutolf, 2020, CC-BY 4.0

3.2 Dynamics of endoderm primordium formation

To confirm the spatial dynamics of endoderm cells, I tracked live FoxA2$^+$ cells by imaging gastruloids formed by aggregation of FoxA2/TBra double reporter cells (TFoxA2; T/Bra:GFP, FoxA2:tagRFP; Yang, 2015). As confirmed by both immunostaining (Fig. 3.4 A) and live reporter imaging (Fig. 3.4 B), this cell line too generates gastruloids that elongate upon pulsed CHIR delivery and that display a TBra$^+$ posterior and an internal core domain that is FoxA2$^+$ and Cdh1$^+$. Following the dynamics of the FoxA2$^+$ population by live imaging throughout its emergence and reorganisation, shows FoxA2$^+$ cells populating and extending away from the t=96h TBra$^+$ posterior, gradually colonising the anteriormost domains of the gastruloid, and then coalescing to form a single mass at the core of the gastruloid (Fig. 3.4 C).The formation of a single, compact mass of cells by 144h thus appears the result of cells clustering to each other as they move from posterior to anterior, and as they move and divide locally within the anterior domain. FoxA2$^+$ display surprising degrees of motility, especially given their (inferred) Cdh1$^+$ makeup.

When quantified through Flow Analysis Cytometry (FC), within a timeframe ranging from t=48h to endodermal specification (t=72h, t=96h) and reorganisation (t=120h), the FoxA2$^+$ population can be seen to readily increase, especially from t=96h to t=120h (Fig. 3.4 D); confirming what observed through live imaging and timed immunostaining. Few TBra$^+$/FoxA2$^+$ double positive cells can be detected however, and only at later timepoints. In general, the incongruence between FC and live imaging in the evolution of the TBra$^+$ population over time (steady increase in FC, extinction in live imaging), may be due to a phototoxic effect of the light intensities used during timelapse acquisition. Indeed, a strong TBra$^+$ pole can clearly be seen if gastruloid are only imaged at the end of their development (see Fig. 3.4 B).

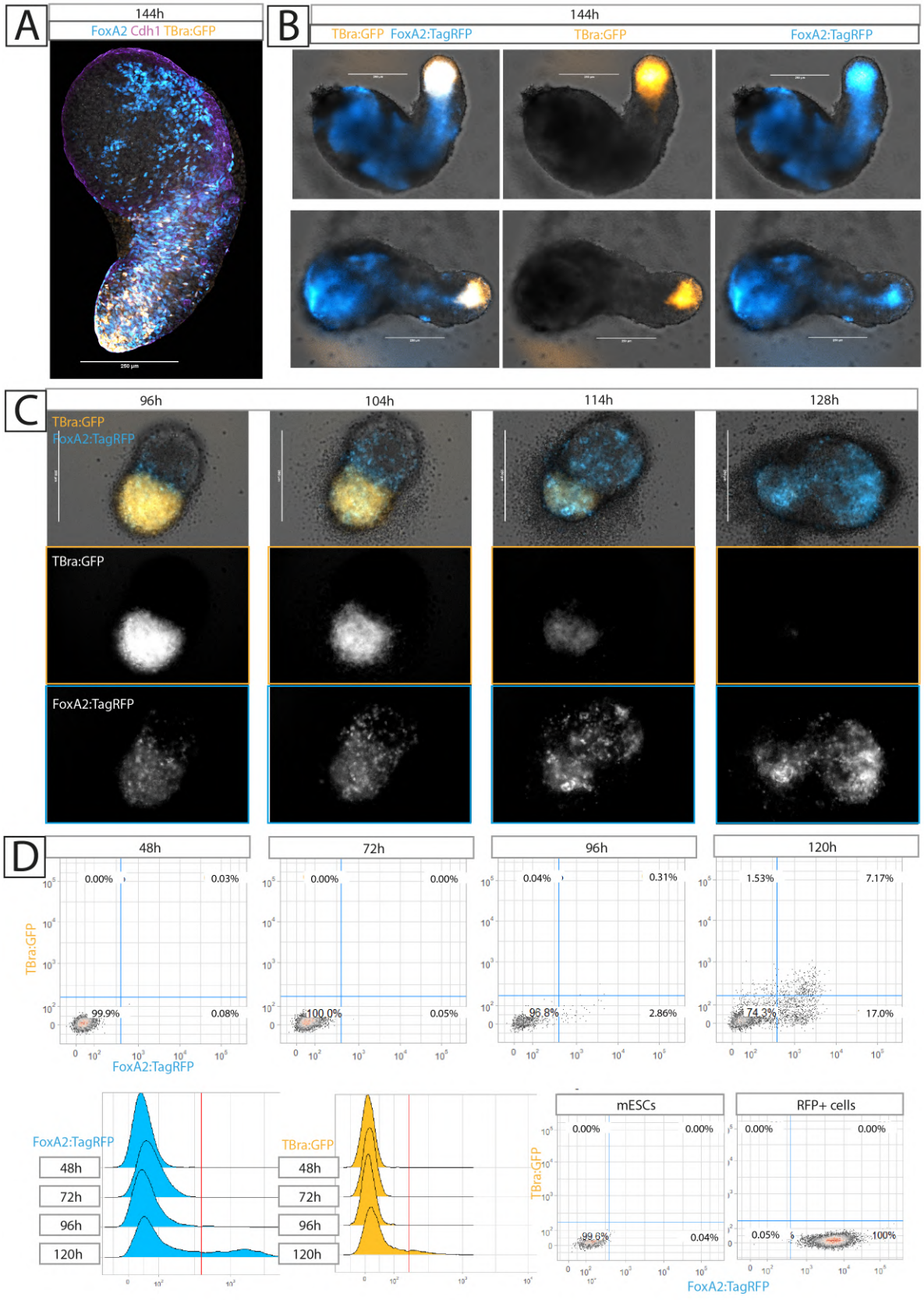


Fig. 3.4.: Gastruloids made from TBraGFP/FoxA2tagRFP cells highlight dynamics of endoderm primordium morphogenesis See next page for the full caption to this figure ↪

Fig. 3.4.: ↪ continued from the previous page. **A)** t=144h gastruloid made from the TBraGFP/FoxA2tagRFP cell line, immunostained against Cdh1 (magenta) and FoxA2 (cyan). The endogenous TBra signal is shown in gold. As in SBR gastruloids, FoxA2$^+$ cells populate a central epithelial domain. **B)** Live imaging of t=144h gastruloids. The endogenous FoxA2 reporter (cyan) is distributed in patterns consistent with the presence of an internal FoxA2$^+$ primordium. The endogenous TBra reporter, as expected, associates with the elongating tip of the gastruloid. **C)** Stills from a timelapse of gastruloid development, from t=96h to t=128h. FoxA2$^+$ cells can be seen emerging at the posterior and gradually occupying anterior domains while coalescing to the core of the aggregate. **D):** FC quantification of the TBra$^+$ and FoxA2$^+$ populations as gastruloids progress from t=48h to t=120h. Undifferentiated mESCs and constitutively-RFP$^+$ cells are taken as references. Scale bars: 250μm. Note: parts of this figure are also published in Vianello and Lutolf, 2020, CC-BY 4.0

Conclusions:

- Extensive expansion of the FoxA2$^+$/Cdh1$^+$ population is seen in gastruloids developing past t=96h (*i.e.* from t=120h onwards). This population reproducibly forms a distinctive structure at the core of mature gastruloids, and spans their entire longitudinal axis.
- The Cdh1$^+$ primordium matures a polarised epithalial architecture over time, with a basal side facing the periphery of the gastruloid, and an apical side facing inner (Cdh1$^-$) cells. It is surrounded basally by a thick envelope of mesodermal cells.
- Posteriorly, the Cdh1$^+$ primordium seems to be a mix of epithelial cell types. While it is indeed FoxA2$^+$, Sox1$^+$ cells can be transiently seen to occupy parts of the domain.
- When followed by live imaging, FoxA2$^+$ cells can be seen emerging from the TBra$^+$ posterior of the early, and migrating towards the anterior domain. They display high motility and appear to form the core primordium by coalescence and local proliferation.

Materials and methods

Note: relevant Materials and Methods are here reported from my own work in Vianello and Lutolf, 2020, licensed under CC-BY 4.0 International license and for which I hold the rights of reproduction.

Cell culture: mESCs ("SBR" Sox1/TBra double reporter cell line described in (Deluz et al., 2016); CRG8 cells of 129P2 background (RRID:CVCL_3987, Mountford et al., 1994); or "TFoxA2" FoxA2/TBra double reporter cell line described in (Yang, 2015); E14 cells of 129P2 background (RRID:CVCL_C320, Fehling et al., 2003; Hooper et al., 1987) were cultured in tissue-culture-treated, non-gelatinised, 6well plates, in 10%Serum Medium with added 2i and LIF. Cells were split every third day, by washing in PBS-/-, adding Accutase for around 3min RT, and collecting the resulting cell suspension in a clean centrifuge tube. The Accutase of the cell suspension was then diluted out 1:10 in 10% Serum Medium, and cells were pelleted by centrifuging 200xg, 4min, 4°C. After aspirating out the supernatant, the pellet was resuspended in 1mL 10% Serum Medium to a single cell suspension, and cell density was counted with a haemocytometer. Around 65000-75000 cells were transferred to a new well with 2mL pre-equilibrated 10% Serum Medium (6750-7800 cells/cm2). Cells were then left in a humidified incubator, 37°C, 5% CO2, until use or until further splitting 3 days later. NOTE: In most cases, splitting was coupled to Gastruloid generation. In those cases, the cell pellet was resuspended in N2B27 rather than 10% Serum Medium + 2i and LIF. A complete, step-by-step, protocol is available at: https://dx.doi.org/10.17504/protocols.io.7xbhpin. **Recipes**: 10% Serum Medium: 86.8% DMEM, high glucose, with Glu-taMAX (L-Alanyl-Glutamine, final concentration: 3.45mM), 10% ES-grade Foetal Bovine Serum, 100U/mL Penicillin, 100ug/mL Streptomycin, 0.1mM Non Essential Amino Acids, 1mM Sodium Pyruvate, 0.1mM beta-mercaptoethanol 10% Serum Medium + 2i and LIF: add 3uM CHIR99021, 1uM PD0325901,and 100u/mL LIF.

Gastruloid generation: mESCs were washed in PBS-/-, detached from adherent culture with Accutase (around 3min, RT), and collected in a centrifuge tube. The Accutase in the cell suspension was then diluted out 1:10 in 10% Serum Medium, and cells were pelleted by centrifuging 200xg, 4min, 4C. The supernatant was removed, and the pellet was washed by resuspension in 10mL PBS-/-. Cells were re-pelleted by centrifuging 200xg, 4min, 4C and washed once more in 10mL fresh PBS-/-. After re-pelleting once more (200xg, 4min, 4C), the pellet was dissociated as a single-cell suspension in 1mL N2B27 Medium. Cells were counted with a haemocytometer, and, for each plate of Gastruloids made, 37500 cells (SBR line) or 93750 cells (TFoxA2 line) were transferred to 5mL fresh N2B27 (7.5 cells/uL

or 18.75 cells/ul final concentration, respectively). The cell suspension was distributed as 40uL droplets (=300 SBR cells/droplet, or 750 TFoxA2 cells/droplet) in wells of a U-bottomed, low-adhesion, 96 well plate, and the plates were left for 120h in a humidified incubator, 5% CO2, 37°C. At 48h after plating, 150uL of 3uM CHIR99021 N2B27 were added to each well, and this solution was substituted with fresh N2B27 (no CHIR) every 24h after that. A step-by-step detailed protocol is available at: https://dx.doi.org/10.17504/protocols.io.9j5h4q6. **Recipes**: 10% Serum Medium: 86.8% DMEM, high glucose, with 3.97mM GlutaMAX (L-Alanyl-Glutamine, final concentration: 3.45mM), 10% ES-grade FBS, 100U/mL Penicillin, 100ug/mL Streptomycin, 0.1mM Non Essential Amino Acids, 1mM Sodium Pyruvate, 1mM beta-mercaptoethanol. N2B27: 47.4% Neurobasal Medium, 47.4% DMEM/F-12, with 2.50mM GlutaMAX (L-Alanyl-Glutamine, final concentration: 1.18mM), 1mM GlutaMAX Supplement (total concentration: 2.18mM), 100U/mL Penicillin, 100ug/mL Streptomycin, 0.1mM Non Essential AminoAcids, 1mM Sodium Pyruvate, 1mM beta-mercaptoethanol, 1% B27Supplement, serum-free, 0.5% N-2 Supplement.

Gastruloid (and embryonic tailbud) immunostaining: Gastruloids were collected at every given timepoint, washed in PBS-/-, and fixed in 4% PFA in PBS-/-, for 2h, 4C, on a low-speed or-bital shaker; or 45min, RT, static. Gastruloids were then washed in PBS+FT (PBS-/-, 10% ES-grade Foetal Bovine Serum, 0.2% Triton-X100), and blocked and permeabilisedin PBS+FT for 1h, RT, static. Primary antibody solutions were then prepared in PBS+FT, with 2ug/mL DAPI. Samples were stained overnight, 4C, on a low-speed orbital shaker. Similarly, secondary antibody solutions were prepared in PBS+FT, 2ug/mL DAPI, and samples were stained overnight, 4C, on a low-speed orbital shaker. Gastruloids were mounted in Fluoromount-G mounting medium (no spacers), and slides kept at 4C long term. The E10.5 mouse embryonic tailbud, fixed in 4% PFA, dehydrated in methanol, and stored at -20C, was a kind gift from Paul Gerald Layague Sanchez (Aulehla lab, EMBL Heidelberg, Germany; supernumerary tissue from own research). It was rehydrated in a graded series of methanol-PBST solutions (100%-0%, 75%-25%, 50%-50%, 25%-75%, 100% PBST) and processed as the gastruloid samples. All antibody solutions were washed away after incubation by washes in PBS+FT. A detailed, step-by-step protocol is available at: https://dx.doi.org/10.17504/protocols.io.7tzhnp6. A list of the primary antibodies used is provided in Table A.6. Secondary antibodies used were all from Thermo Fisher Scientific: donkey anti-mouse IgG Alexa 647 (CAT#A-31571, RRID: AB_162542); donkey anti-rabbit IgG Alexa 488 (CAT#A-21206, RRID: AB_2535792), Alexa 568 (CAT#A-10042, RRID: AB_2534017), or Alexa 647 (CAT#A-31573, RRID: AB_2536183); donkey anti-rat IgG Alexa 488 (CAT#A-21208, RRID: AB_2535794), goat anti-rat Alexa 568 (CAT#A-11077, RRID: AB_2534121), or Alexa 647 (CAT#A-

21247, RRID: AB_141778); donkey anti-goat IgG Alexa 488 (CAT#A-11055, RRID: AB_2534102), or Alexa 568 (CAT#A-11057 RRID: AB_2534104).

Gastruloid imaging and image processing: Bright-field images of Gastruloids were taken on either a Nikon Ti inverted spinning-disk microscope or an Olympus CellR inverted widefield microscope (UPLAN S APO 10x/0.40 air objective, CCD Grayscale Hama-matsu ORCA ER B7W Camera; Olympus XCellence software for data capture). Both microscope setups had CO2 and temperature control (37°C and 5% CO2). Live imaging of TFoxA2 reporter Gastruloids was done on the Olympus CellR inverted wide-field microscope described above, with acquisition every 30min (5 z-slices, 27.5um spacing) Immunostained Gastruloids were imaged on a Zeiss LSM700 inverted confocal microscope (Plan-Apochromat 20x/0.80 air objective, motorized stage, LED Lumencor SOLA Illumination, CCD Grayscale Axiocam MRm (B/W) Camera; ZEN 2009 software for data capture) or on a Zeiss LSM780 inverted confocal microscope (for the CDH1 projection, Plan-Apochromat 20x/0.80 air objective, motorized stage). Images were opened, stitched, and processed for publication (LUT assignment, channel display, min and max intensity thresholding based on no-primary control) using the Fiji ImageJ distribution (Rueden et al., 2017; Schindelin et al., 2012), and the “Grid/Collection Stitching” plugin therein (Preibisch et al., 2009). The depth-coded reconstruction in Fig.3.1 A was generated using the "Temporal-Color-Code" (https://imagej.net/Temporal-Color_Code) function. The blue, orange, and purple LUTs used throughout the figures were designed by Christophe Leterrier (https://github.com/cleterrier/ChrisLUTs, "BOP" palette).

Gastruloid FC: TFoxA2 Gastruloids were grown according to the protocol described above, and two/four 96well plates of Gastruloids at each timepoint were used and processed for FACS. Briefly, Gastruloids were collected at every given timepoint, washed in PBS-/-, and digested 8min, 37C, in Digestion Solution (Collagenase IV [3mg/mL],Dispase [4mg/mL], DNAseI [100ug/mL], in PBS). Working on ice, the cell suspension was then strained through the filter cap of a FACS tube, and an excess of cold Staining Buffer (10%ES-FBS, Pen-Strep [100U/mL], EDTA [1mM], in PBS) was added to stop the digestion. Cells were then stained with DAPI ([0.2ug/mL] DAPI, in Staining Buffer), 10min, 4C, fixed in 2%PFA, 4C, 10min, and stored in Staining Buffer, 4C, in the dark, until use. Standard 2D cultures of TFoxA2 mESCs and RFP^+ mESCs were used as negative and positive references, respectively. These were detached in Accutase, 4min, RT, and DAPI-stained and fixed as done for filtered gastruloid cells and as described above. GFP BrightComp eBeads™ (Invitrogen/Thermo Fisher Scientific, CAT#A10514) were used as GFP^+ positive reference, according to manufacturer protocol (1 drop of beads resuspended in 1mL Staining Buffer). A step-by-step detailed protocol is available at https://dx.doi.org/10.17504/protocols.io.bvgrn3v6.

Samples were analysed on a Becton Dickinson LSR-Fortessa™ Flow Cytometer, with optical configuration 355nm[450/50], 488nm[530/30], 561nm[585/15], using BD FACSDivaTM software, with applied compensation. Exported FCS files were analysed in RStudio (ggcyto library, Van et al., 2018, and flowCore library, Ellis et al., 2019). The annotated notebook, with a step by step walkthrough the entire analysis pipeline, is available at https://doi.org/10.5281/zenodo.4894122.

4 Identity and patterning of the endoderm primordium

> *[Imset's] jar contained the liver. Hapy [...] guarded the jar that held the lungs. Qebhseneuf watched over the intestines [...]. Duwamutaf protected the stomach [...]*
>
> **— Burial practices of Ancient Egypt**
> (from the Encyclopedia of Time; Birx, 2009)

The observation of a compact endodermal primordium at the core of the gastruloid from day 5 of culture (t=120h; and onwards; refer back to the previous chapter) opens the possibility that this domain may host cells which, over such time, have had the opportunity to mature towards more advanced endodermal identities. Specifically, and given that the primordium aligns with the AP axis of the gastruloid and spans the entirety of it, we questioned whether these identities may be patterned according to their position along the axis. Gastruloids have indeed been shown to maintain Hox colinearity (Beccari et al., 2018) and pattern correct AP identities at least for what concerns their paraxial mesoderm compartment (Brink et al., 2020). Does the axial organisation of the gastruloid translate to the Cdh1$^+$ core? Similarly to the way progenitors of different visceral organs are specified as a function of their AP position along the embryonic gut tube (Zorn and Wells, 2009), caudal endoderm cells would populate the posterior of the gastruloid primordium, and (more) anterior progenitors would populate (more) anterior regions of the gastruloid primordium. Given the posteriorly-biased character of gastruloid tissues (at least for what concerns their neural/ectodermal tissues, Beccari et al., 2018; Brink et al., 2020), we also wondered whether the endoderm primordium would contain cell identities spanning the entirety of the embryonic gut tube, or only those of a posterior subset of it. Finally, we wondered whether the absence of endoderm with extraembryonic origin in gastruloids would limit the range of endodermal identities obtained, especially at the posterior where the highest representation of VE cells is found *in vivo* (Nowotschin et al., 2019b; Kwon et al., 2008; Pijuan-Sala et al., 2019).

4.1 Immunostaining of basic endodermal markers

To start contextualising the endodermal identity of the internal Cdh1$^+$ of mature gastruloids (t=120h), and to paint broad coordinates of its AP patterning, I immunostained t=120h gastruloids for the PS and mesodermal marker TBra (Kispert and Herrmann, 1994; Fig. 4.1 A.1), for the two classical endodermal markers FoxA2 (Burtscher et al., 2013; Fig. 4.1 A.2) and Sox17 (Viotti et al., 2014; Burtscher et al., 2012; Fig. 4.1 A.3), and for the posterior marker Cdx2 (Beck et al., 1995; Fig. 4.1 A.4). As can be seen (and already illustrated in Fig. 3.3 B), TBra marks the posteriormost end of the Cdh1$^+$ primordium. TBra$^+$ cells also extend deep within the more anterior portions of the primordium (Fig. 4.1 A.1). In fact, since FoxA2$^+$ cells can be seen to make up the entire portion of the Cdh1$^+$ domain anterior to the TBra$^+$ pole (Fig. 4.1 A.2, A.5, and quantification in A.8), these more anterior TBra$^+$ cells result TBra$^+$/FoxA2$^+$ double positive ((Fig. 4.1 A.5) and as such present a Cdh1$^+$/TBra$^+$/FoxA2$^+$ signature, and axial distribution, consistent with that of midline embryonic structures (Burtscher and Lickert, 2009; Yamanaka et al., 2007).

Sox17 marks large domains of the primordium's anterior and finds its posterior limit to the neck region of the gastruloid, often marking the rosette-like contacts of the Cdh1$^+$ network with the surface (Fig. 4.1 A.3). Cdx2, recapitulating its expression pattern in the embryo (Beck et al., 1995), marks the entire posterior of the gastruloid: that is, the posterior of the Cdh1$^+$ domain, but also Cdh1$^-$ cells of the posteriormost part of the mesodermal wings enveloping the primordium (Fig. 4.1 A.4). Notably, Cdx2 also marks (Cdh1$^+$/)Sox17$^+$ cells at the posteriormost limit of the Sox17 expression domain (Fig. 4.1 A.7), a signature that has been likened — in other gastruloid studies — to the caudal intestinal portal forming during *in vivo* endoderm development (Beccari et al., 2018). Around E8.0-E8.5 (but not after E9.0) Sox17 indeed marks the embryonic hindgut (Kanai-Azuma et al., 2002), before playing more targeted roles in pancreatic specification at the posterior foregut boundary (Spence et al., 2009).

Interestingly, Dppa4$^+$ cells intermingle with FoxA2$^+$ endoderm at the anterior (Fig. 4.1 A.9), possibly representing a surprising maintenance of pluripotency from the earliest timepoints of Gastruloid differentiation (giving their continuity with Dppa4+ cells at all previous timepoints, refer back to Fig. 2.6 B.4). In this regard, other groups have interestingly reported the presence of Primordial Germ Cells (PGC)-like cells, marked by Dppa3, in association with the endodermal component (Veenvliet et al., 2020). PGCs indeed transit through the posteriormost hindgut of the embryo

between E7.5 until E8.5/E9.0, prior to their further migration to the gut mesentery and genital ridges (Hara et al., 2009). Here, the anteriormost position of this cluster of cells (rather than posterior), as well as its presence since the earliest steps of gastruloid aggregation (pre-"gastrulation"), argue against a PGC interpretation and favour instead its framing as an idiosyncratic gastruloid behaviour involving entrenchment of a subset of the original gastruloid cells in the pluripotent state upon exposure to CHIR (collaboration; in press).

Finally, I identify the cell surface protein EpCAM as another marker of the entire primordium, with expression almost completely overlapping that of Cdh1 (Fig. 4.1 A.10), yet with an apparent enrichment towards the anterior. The expression of EpCAM distinguishes the ($Cdh1^{+}$/)$Sox17^{+}$ cells we observe as being indeed endodermal, given that this same marker also characterises endothelial progenitors (which would however be $EpCAM^{-}$) at around the same developmental timepoints (Choi et al., 2012). Interestingly, EpCAM staining appears enriched in the region of the Cdh1 primordium occupied by $Sox17^{+}$ cells, hinting at the possibility that combinations of cell-surface markers might drive further sub-sorting of different epithelial combinations within this same $Cdh1^{+}$ core. Here, the posterior of the primordium would represent a “posterior epiblast”-like, $Cdh1^{+}/EpCAM^{LOW}/TBra^{+}$ domain; and the anterior a $Cdh1^{+}/EpCAM^{HIGH}/FoxA2^{+}/Sox17^{+}$ “endoderm”-like domain.

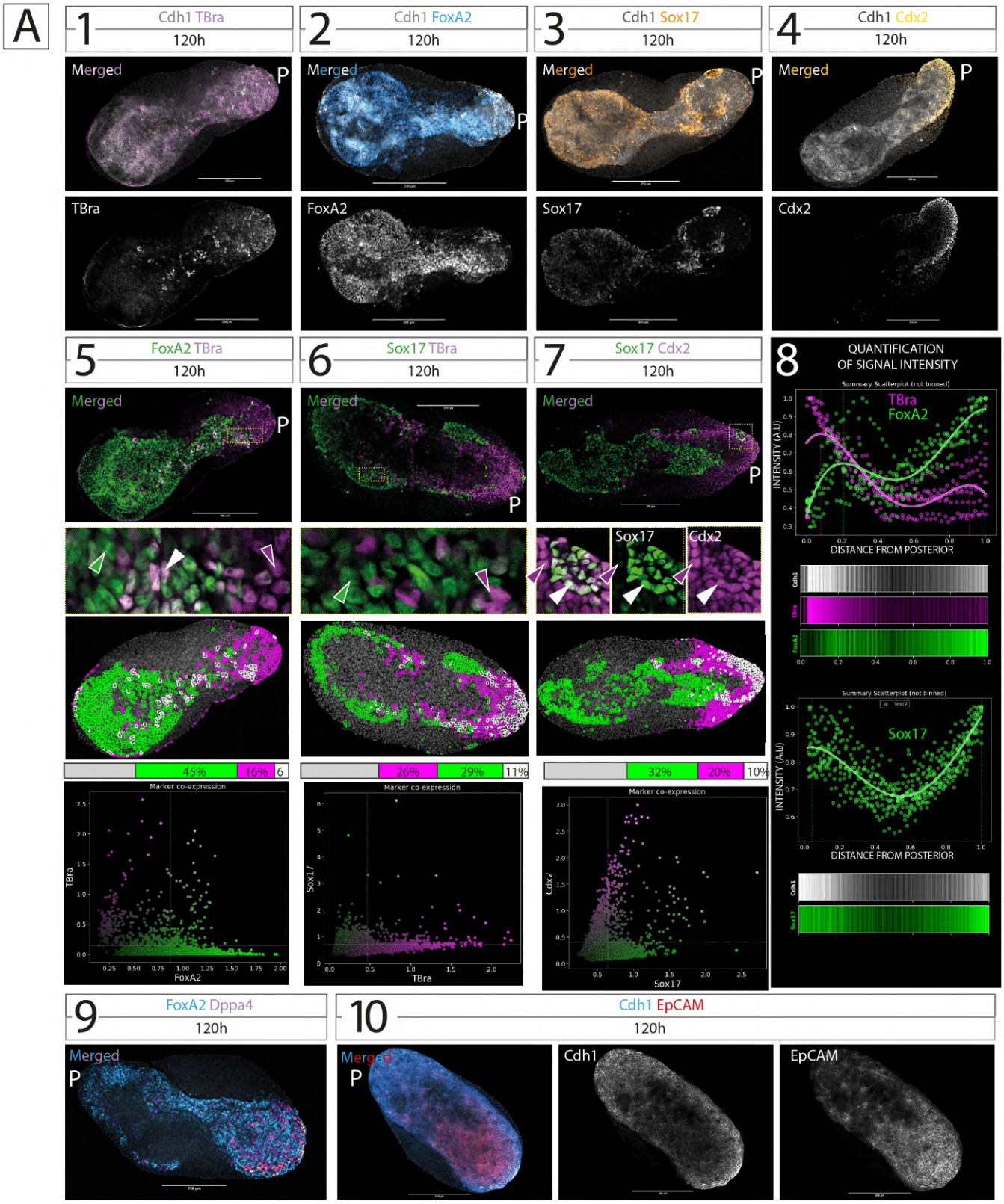


Fig. 4.1.: The epithelial primordium at the core of elongating gastruloids shows patterned expression of endodermal markers. See next page for the full caption to this figure ↪

Fig. 4.1.: ↪ continued from the previous page. **A)** Immunostaining of t=120h gastruloids stained for Cdh1 (core epithelial primordium) and markers TBra (**1**), FoxA2 (**2**), Sox17 (**3**), and Cdx2 (**4**). The expression pattern of endodermal markers FoxA2 and Sox17 is further explored with respect to embryonic tailbud/posterior markers TBra (**5, 6**) and/or Cdx2 (**7**). The proportions and the graphical map of single- and double-positive cells in these immunostainings is also provided, as obtained through the pipeline described in the text. The population-level AP patterning of FoxA2 and Sox17 is quantified ($n=6$ and $n=13$, respectively; **8**). Additional immunostainings show Dppa4+ cells associating with the anterior of the FoxA2+ primordium (**9**) and identify EpCAM as another marker of the entire domain, with preferential accumulation at the anterior (**10**). Scale bars: 250μm. Note: parts of this figure are also published in Vianello and Lutolf, 2020, CC-BY 4.0

4.2 Comparison of single-cell RNAseq datasets

To move beyond the manual validation of well-known, general endodermal markers, and in an attempt to better resolve the identity of the cells contained within the epithelial primordium of the gastruloids, I leveraged the recent availability of Single-cell RNA sequencing (scRNA-seq) data for both gastruloids (Rossi et al., 2021a), and peri and post-gastrulation mouse embryos (specifically, Nowotschin et al., 2019b). We reasoned that embryonic gut tube datasets could be used as a reference map on which to organise endodermal cells from the gastruloid, and as such could help resolve the possible variety of anterior-posterior identities of such cells (an approach more fully elaborated upon in Tan et al., 2021).

Analysis of the gastruloid ScRNA-seq dataset generated in our lab (Rossi et al., 2021a), spanning timepoints t=96h to t=168h, identified two major endodermal clusters characterised by the expression of FoxA2, Sox17, Cdh1, and EpCAM, and lower numbers of cells also expressing TBra and Eomes (see Fig. 4.2 A, lower). These two clusters were as such interpreted as to represent endoderm. Cells of one of such clusters are marked by the expression of genes such as Otx2, Sox17, Hhex, Gata6, Gsc (cluster 13 in Fig. 4.2 A, light yellow), and were here labelled as "early endoderm" given that these cells present a signature of anterior mesendoderm and definitive endoderm (Costello et al., 2015; Thomas et al., 1998; (*cfr.* Table A.1 for a more comprehensive list of markers of this cluster). The second cluster was instead demarcated by the expression of genes such as Cldn4,6,7, Krt7, Crb3 (cluster 4 in Fig. 4.2 A, beige) and was here labelled as "mature endoderm", given the strong expression of epithelial markers shown to characterise later (rather than early) stages of endoderm maturation *in vivo* (*i.e.* gut tube; Anderson et al., 2008; Ogaki et al., 2011; *cfr.* Table A.2 for a more comprehensive list of markers of this cluster). Indeed, performing differential expression analysis between the two clusters (Fig.

4.2 A, side-by-side grid) reveals that the “early endoderm” can be distinguished by higher expression of genes such as Sfrp1, Lhx1, Hesx1, Fgf5 (consistent with an anterior endoderm/mesendoderm character; Costello et al., 2015; Finley et al., 2003; Khoa et al., 2016), while “mature endoderm” distinctively expresses higher levels of e.g. Igfbp5 and Frem2 (expressed in the gut tube; Green et al., 1994; Timmer et al., 2005) as well as additional epithelial markers such as Krt19 and Cldn4,9 (*cfr.* Table A.3 for a more comprehensive list of markers differentially expressed in each cluster).

Reduction of the dimensionality of the data to 3- rather than 2 dimensions (Fig. 4.2 B), while keeping in mind the limitations of the visual inferences that can be made from such maps (Chari et al., 2021), shows the "early endoderm" cluster to be in closest proximity to cells annotated as anterior mesoderm (cluster 3; markers: Gata6, Hand2, Myl7, Gata4, Lhfp), and reveals the "late endoderm" cluster to be closest to the cluster labelled as ectoderm (cluster 5; markers: Gjb3, Epcam, Tfap2c, Cdh1, Krt18), as well with the cluster annotated as notochord/axial mesoderm (cluster 1; markers: TBra, Cobl, Shh, Cdx2, Noto; not evident from the 2-dimensional (2D) Uniform Manifold Approximation and Projection (UMAP)). For reference, a table with the top 10 markers of each of the gastruloid clusters identified is appended (Table A.4). Supporting the "early"/"late" annotation of the two endodermal clusters, the "early endoderm" cluster appears to be majoritarily composed of cells from early gastruloid timepoints (t=96h to t=144h), while the "mature endoderm" cluster appears mostly composed of cells from later timepoint gastruloids (t=120h to t=168h; Fig. 4.2 C).

Notably, the “mature endoderm” cluster expresses all markers recovered by immunostaining (see the previous section and Fig. 4.2 A, lower), and many markers associated with foregut and anterior foregut (e.g. Nkx2.3, Isl1, Otx2; Biben et al., 2004; Zhuang et al., 2013; Nowotschin et al., 2019b; Fig. 4.2 D). Still, markers identifying *all* positions along the AP axis of the embryonic gut tube (Nowotschin et al., 2019b) could be recovered within Gastruloid endoderm clusters (Fig. 4.2 D and Fig. 4.3).

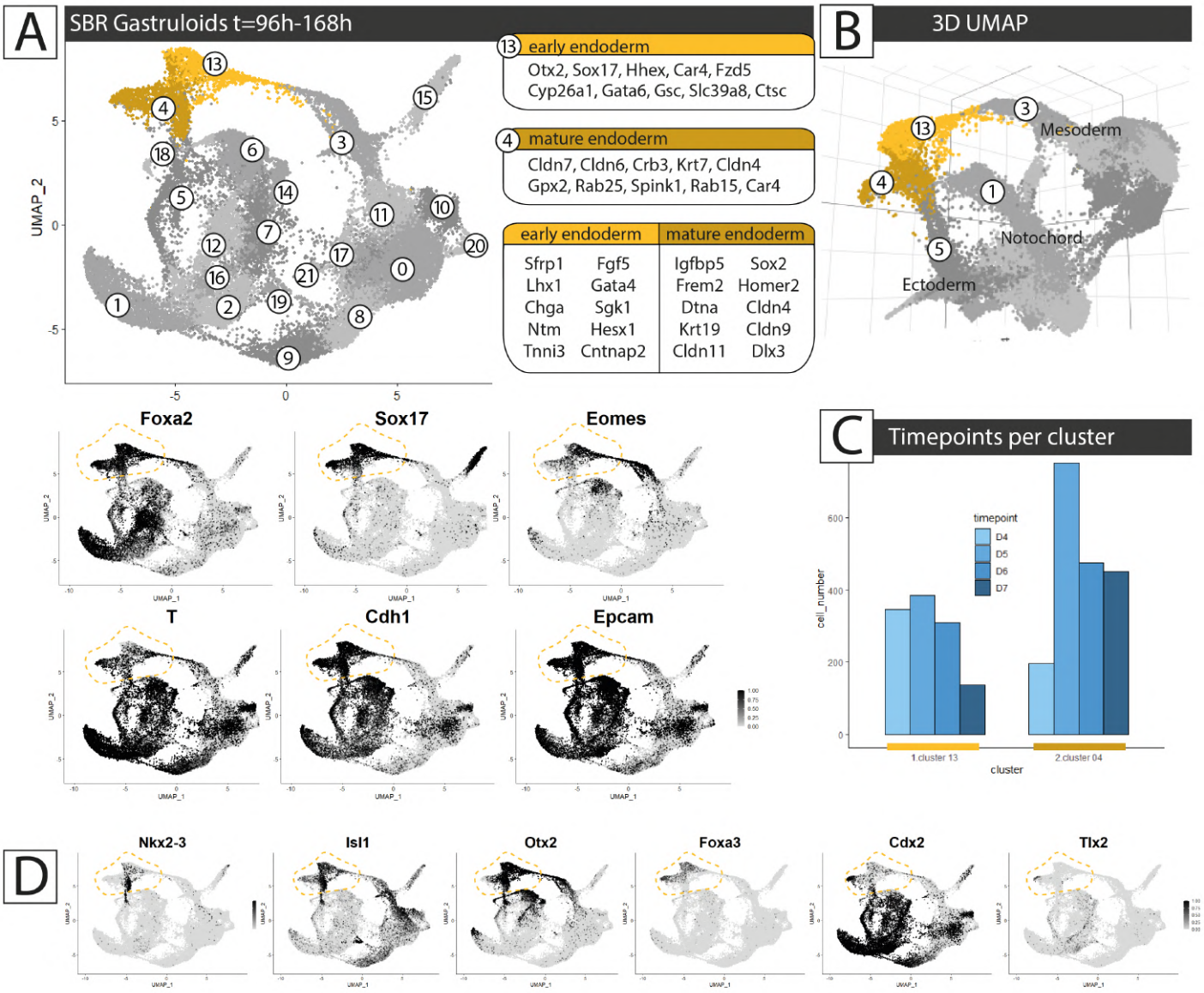


Fig. 4.2.: Analysis of single-cell RNAseq datasets from SBR Gastruloids highlights two "endoderm" clusters. **A) Top:** UMAP of the re-processed scRNAseq dataset generated in Rossi et al., 2021a and comprising expression data from gastruloids spanning t=96h to t=168h, at 24h intervals. Highlighted in gold and beige are clusters classified as "early" and "mature" endoderm (clusters 13 and 4), respectively. The top 10 marker genes for each cluster are indicated, as well as the top 10 differentialy expressed genes distinguishing one from the other. **Bottom:** Expression of classic endoderm markers in the dataset. **B)** 3D UMAP of the same dataset as in A. Clusters neighbouring the two endoderm ones are highlighted. **C)** Number of cells from each gastruloid timepoint falling in each of the two clusters labelled as endoderm. **D)** Expression of anterior and posterior (left to right) gut tube markers (see Nowotschin et al., 2019b) within the gastruloid dataset. "Endoderm" clusters are circled in gold throughout. Note: This figure is also published in Vianello and Lutolf, 2020, CC-BY 4.0

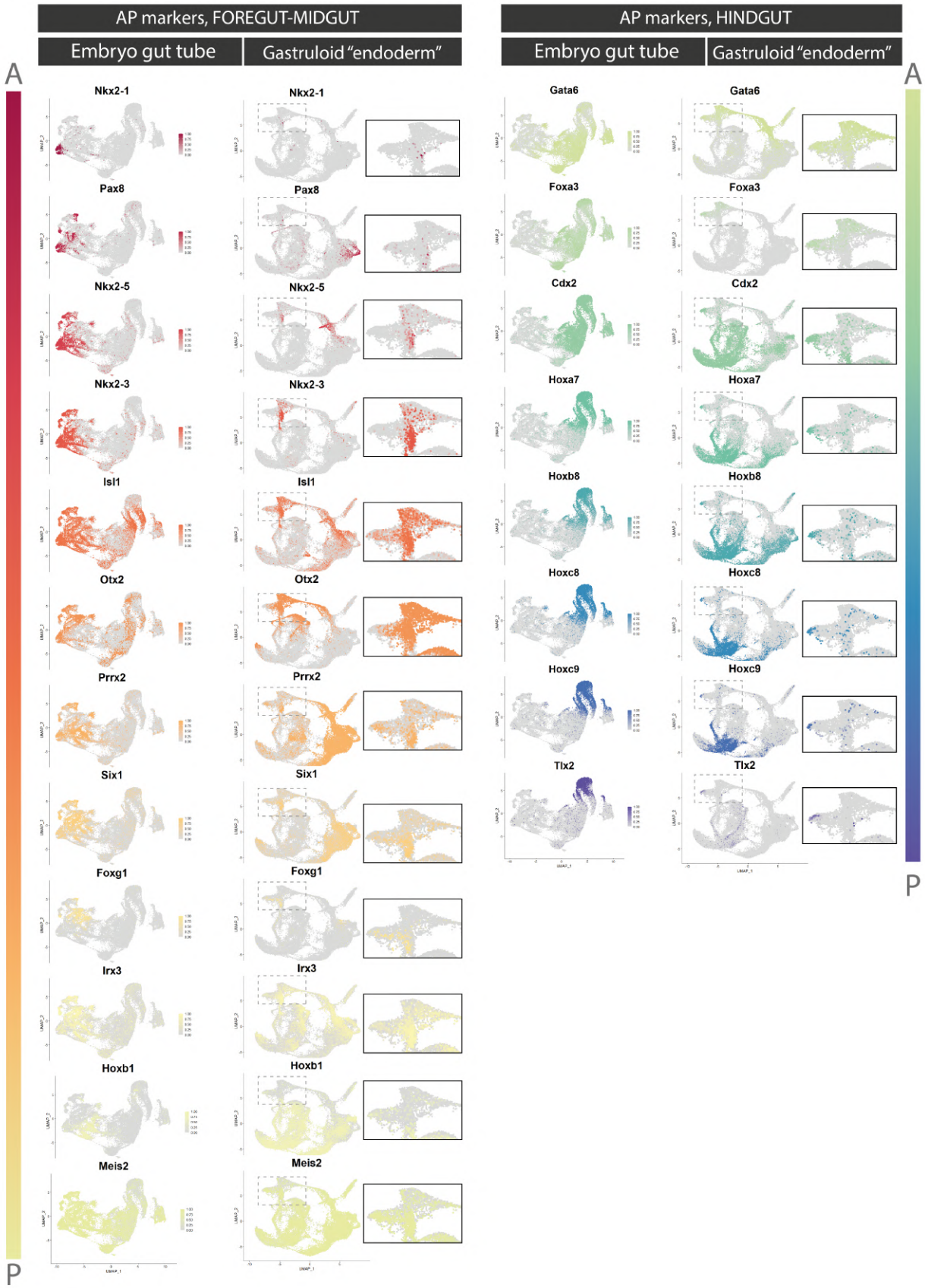


Fig. 4.3.: Expression-pattern of gut endoderm Anterior-Posterior markers in *in vivo* and *in vitro* datasets. See next page for the full caption to this figure ↪

Fig. 4.3.: ↪ continued from the previous page. Markers of AP position along the embryonic gut tube (Nowotschin et al., 2019b) are plotted from anteriormost foregut (top, Nkx2-1), to posteriormost hindgut (bottom, Tlx2). For each gene, the expression pattern is shown for both the embryonic dataset (E8.75 gut tube, Nowotschin et al., 2019b); left) and the Gastruloid dataset (SBR gastruloids, Rossi et al., 2021a, right). Validating the reprocessing of the embryonic dataset, increasingly posterior markers define continuous domains from one extremity of the UMAP to the other. For the gastruloid dataset, an inset focusing on the two "endoderm" clusters is also provided. A = Anterior, P = Posterior. Note: this figure has also been published in Vianello and Lutolf, 2020, CC-BY 4.0

Parallelly, processing of the embryonic dataset from Nowotschin et al., 2019b, and maintaining only cells of the gut tube (main cluster, marked by FoxA2 and Cdh1 expression), allows to use the resulting UMAP (shown in Fig. 4.4 A and throughout) as an approximate visual map of the gut tube as it extends from anterior foregut (leftmost in Fig. 4.4 A) to posterior hindgut (up-right in Fig. 4.4 A). Accordingly, plotting known markers of anterior (*e.g.* Nkx2.5, Sox2; Nowotschin et al., 2019b; Wei and Condie, 2011; Wood and Episkopou, 1999; Zhang et al., 2005) and posterior gut (*e.g.* Cdx2, TBra; Kispert and Herrmann, 1994; Beck et al., 1995) marks corresponding left and right regions in the UMAP (Fig. 4.4 A). In fact, plotting the expression of the core set of 20 TFs identified by Nowotschin et al., 2019b as predictors of anterior–posterior pseudo-space ordering in the dataset defines, in our re-processed dataset too, a continuum from one end of the UMAP to the other (Fig. 4.3).

When integrating cells from the two gastruloid clusters classified as endoderm with gut tube cells from the embryonic dataset, the former can be seen to fall along the entire width of the UMAP (Fig. 4.4 B). Notably however, quantification of the number of Gastruloid cells falling in each embryonic cluster shows differential distribution of cells from each of the two identified endoderm clusters with respect to the embryonic AP axis. Cells of the "early endoderm" cluster (in light blue in Fig. 4.4 B) preferentially populate posterior clusters, while cells of the "mature endoderm" cluster (in coral red in Fig. 4.4 B) distribute more broadly along the entire embryonic gut tube, thus covering the anterior clusters under-populated by "early endoderm". Specifically, "mature endoderm" gastruloid cells appear to be over-represented in embryonic anterior clusters with anterior foregut assigned identity (Fig. 4.4 C, leftmost; thyroid, thymus, lung territories) as well as posterior hindgut. "Early endoderm" gastruloid cells appear to be over-represented instead in territories of the embryonic foregut-midgut junction, and thus with liver and pancreas assigned identity (Fig. 4.4 C, rightmost). The midgut cluster with small intestine assigned identity (Fig. 4.4 C, mid) was instead populated by equal amount

of cells from both gastruloid clusters, and in fact was populated by the majority of gastruloid cells classified as endoderm, in both conditions (Fig. 4.4 C, mid). A table with the top 20 markers of each of the embryonic gut tube clusters highlighted is appended for reference (Table A.5).

It is important to notice that a similar comparison strategy (Tan et al., 2021) has shown that integration and batch effect correction can coerce *in vitro* cells to overlap with their most similar *in vivo* counterpart in a low dimensional embedding, even though the cells may in fact display low similarity. Quantification of transcriptomics differences using cell typing approaches using the *in vivo* dataset as a reference may be more appropriate (Tan et al., 2021) and indeed were attempted, but could not achieve high enough resolution (not shown).

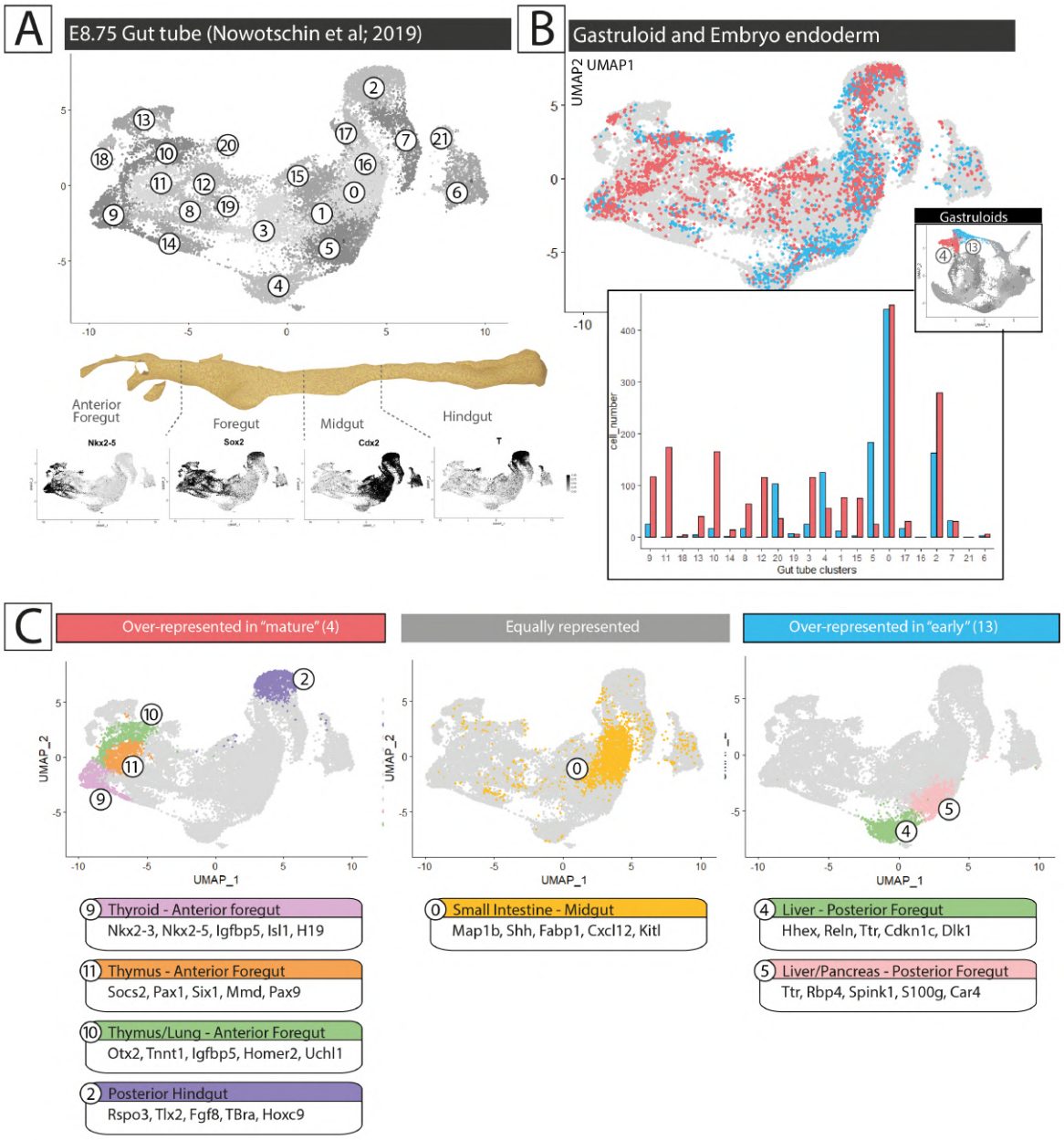


Fig. 4.4.: Gastruloid "mature endoderm" aligns with identities across the entire length of the embryonic gut tube, including Anterior Foregut. **A)Top:** UMAP representation of the re-processed dataset fron Nowotschin et al., 2019b. Only the cluster corresponding to the embryonic gut tube is shown. **Bottom:** Expression of selected anterior foregut, foregut, midgut, and hindgut markers highlight UMAP dimension 1 (x-axis) as carrying information of position along the embryonic gut tube AP-axis (leftmost: anterior; upper-right: posterior) **B)** UMAP of the dataset in A, integrated with cells from gastruloid "endoderm" cluster 13 ("early endoderm", light blue) and cluster 4 ("mature endoderm", coral red). **Bottom:** quantification of the number of cells from each endoderm cluster falling within each of the embryonic gut tube clusters (ordered from anteriormost to posteriormost). **C)** Embryonic gut tube clusters in which gastruloid "mature endoderm" cells are over-represented (left), in which gastruloid "early endoderm" cells are over-represented (right), or in which gastruloid endoderm cells are equally represented (mid). The top 5 genes for each of the embryonic clusters highlighted are indicated. Note: a version of this figure has also been published in Vianello and Lutolf, 2020, CC-BY 4.0

4.3 Spatial patterning of endodermal identities

Alignment of the molecular phenotype (transcript expression) of gastruloid cells with that of cells along varying position of the gut tube of the embryo does not in fact necessarily imply that these cells are spatially patterned within the gastruloid. An equally probable possibility is indeed that these different AP cell identities are intermingled throughout the Cdh1$^+$ primordium. To verify whether gastruloids show patterned distribution of AP identities, and to circumvent the failure of a number of antibodies to detect a number of markers (Onecut1, Nkx2.5, Tbx1; data not shown), I developed a protocol for Hybridisation Chain Reaction (HCR) on gastruloids, based on dx.doi.org/10.17504/protocols.io.7pyhmpw, as well as a pipeline to design custom HCR probes (see Methods).

HCR for *Pax9* (marker of the anterior foregut in the embryo; Peters et al., 1998; Neubüser et al., 1995) and *Pax1* (marker of pharyngeal endoderm in the embryo; Wallin et al., 1996) validates the expression patterns highlighted by single-cell analysis (Fig. 4.5). Both markers are expressed within the endodermal domain, and are specific to it. Still, notice Pax1 also marking mesodermal cells of the outer gastruloid envelope, consistent with the known expression of Pax9 in the embryonic sclerotome and tailbud mesoderm (Neubüser et al., 1995). Most notably, the expression of Pax9 and Pax1 in the Gastruloid is not uniform within the endodermal domain, but is restricted to one part of it, the anterior (Fig. 4.5 B,C). At the edge of a broader Pax9$^+$ anterior domain, Pax1$^+$ cells — with distinctive epithelial organisation — further define the ateriormost limit of the Gastruloid endoderm primordium (Fig. 4.5 B, bottom).

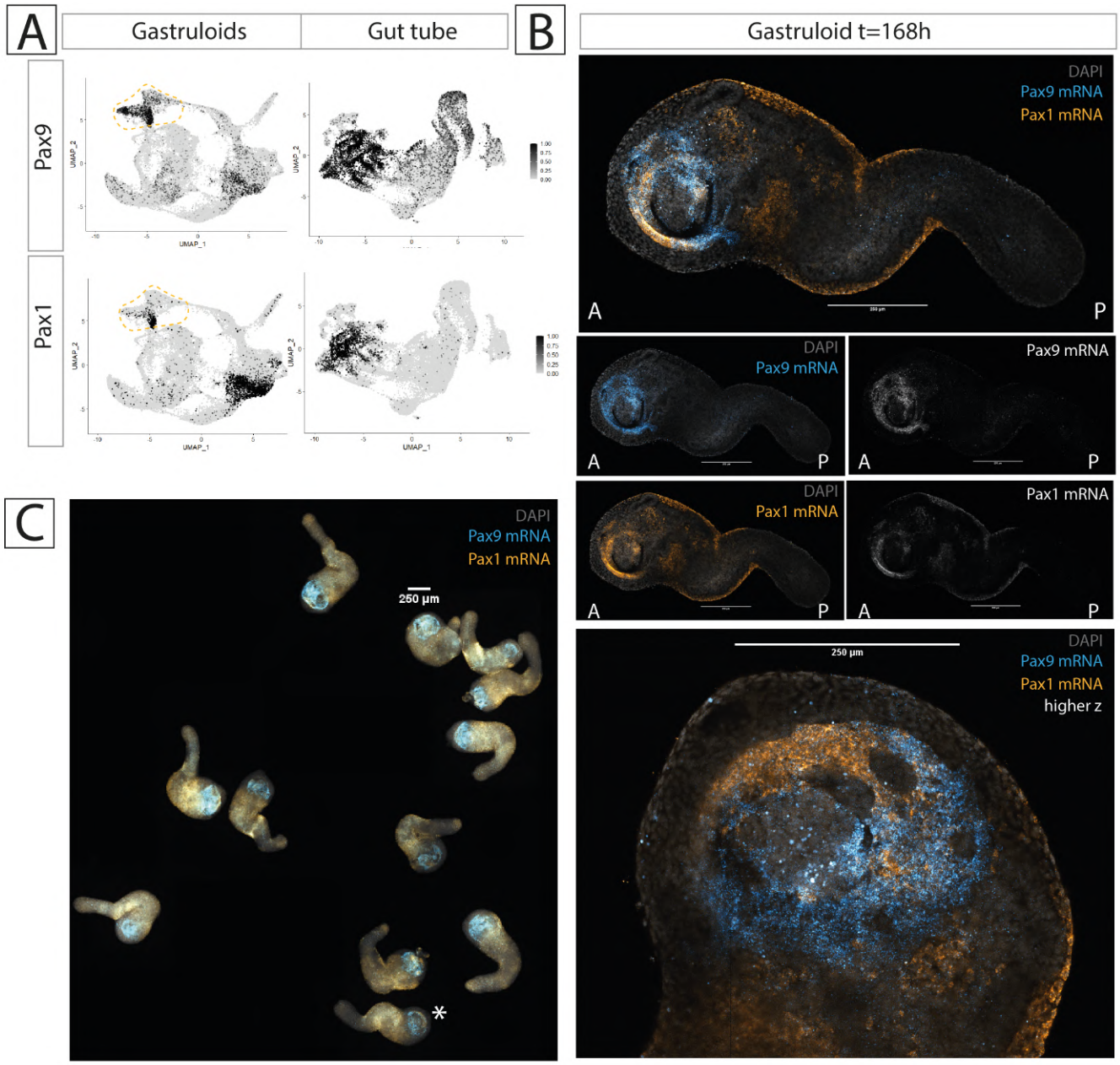


Fig. 4.5.: Anterior foregut endoderm markers are correctly patterned in late gastruloids. **A)** Expression pattern of Pax9 and Pax1 in the reprocessed gastruloid dataset from Rossi et al., 2021a (left) and in the reprocessed embryonic gut tube dataset from Nowotschin et al., 2019b (right). Not the expression in the two gastruloid endoderm clusters (dashed yellow boundary). **B) Top:** HCR of late gastruloids (t=168h) for Pax9 mRNA (cyan) and Pax1 mRNA (gold). **Bottom:** Different z-plane of the anterior portion of the same gastruloid, to highlight how the anterior Pax9+ domain gives way to an anteriormost Pax1+ domain. **C)** Robustness of the Pax9 and Pax1 patterning at the population level (*n=12*). The majority of the Pax1 signal is detected in the presomitic mesoderm, as also shown in A. Asterisk: gastruloid shown in B. Note: part of this figure has also been published in Vianello and Lutolf, 2020, CC-BY 4.0

Conclusions:

- The t=120h $Cdh1^+$ primordium displays coarse domains of endodermal cells, patterned along the AP axis. In a $FoxA2^+/Cdh1^+$ general context, $Sox17^+/EpCAM^+$ cells occupy most of the anterior, extendingly posteriorly up to the neck of the gastruloid, and $TBra^+$ and $Cdx2^+$ cells mark the posterior.
- Analysis of scRNAseq data of gastruloids spanning the t=96h to t=168h interval identifies two endoderm clusters can be differentiated by the expression of early and late endodermal markers, as well as early and late timepoint cell contributions. These clusters express all markers identified by immunostaining, as well as markers of all AP positions of the embryonic gut tube.
- Comparison with scRNAseq data from the gut tube of E8.75 mouse embryos, shows that a high proportion of the gastruloid endoderm falls within mid-gut identities, gastruloid "early endoderm" preferentially accumulates in midgut/posterior-foregut clusters, and "mature endoderm" additionally populates anterior foregut domains.
- As confirmed for anterior foregut markers Pax9 and Pax1 by HCR, and hindgut marker Cdx2 by immunostaining, it appears that the cell identities identified by scRNAseq analysis are correctly patterned along the AP axis of the gastruloid, within the $Cdh1^+$ primordium.
- Incidentally, the recovery of anteriormost endodermal identities in gastruloids contrast with the known post-occipital restriction of their mesodermal and neurectodermal compartments.

Materials and methods

Note: some of the Methods described here are reported from my own work in Vianello and Lutolf, 2020, licensed under CC-BY 4.0 International license and for which I hold the rights of reproduction.

Cell culture: mESCs ("SBR" Sox1/TBra double reporter cell line described in (Deluz et al., 2016); CRG8 cells of 129P2 background (RRID:CVCL_3987, Mountford et al., 1994) were cultured in tissue-culture-treated, non-gelatinised, 6well plates, in 10%Serum Medium with added 2i and LIF. Cells were split every third day, by washing in PBS-/-, adding Accutase for around 3min RT, and collecting the resulting cell suspension in a clean centrifuge tube. The Accutase of the cell suspension was then diluted out 1:10 in 10% Serum Medium, and cells were pelleted by centrifuging 200xg, 4min, 4°C. After aspirating out the supernatant, the pellet was resuspended in 1mL 10% Serum Medium to a single cell suspension, and cell density was counted with a haemocytometer. Around 65000-75000 cells were transferred to a new well with 2mL pre-equilibrated 10% Serum Medium (6750-7800 cells/cm2). Cells were then left in a humidified incubator, 37°C, 5% CO2, until use or until further splitting 3 days later. NOTE: In most cases, splitting was coupled to Gastruloid generation. In those cases, the cell pellet was resuspended in N2B27 rather than 10% Serum Medium + 2i and LIF. A complete, step-by-step, protocol is available at: https://dx.doi.org/10.17504/protocols.io.7xbhpin. **Recipes**: 10% Serum Medium: 86.8% DMEM, high glucose, with Glu-taMAX (L-Alanyl-Glutamine, final concentration: 3.45mM), 10% ES-grade Foetal Bovine Serum, 100U/mL Penicillin, 100ug/mL Streptomycin, 0.1mM Non Essential Amino Acids, 1mM Sodium Pyruvate, 0.1mM beta-mercaptoethanol 10% Serum Medium + 2i and LIF: add 3uM CHIR99021, 1uM PD0325901,and 100u/mL LIF.

Gastruloid generation: mESCs were washed in PBS-/-, detached from adherent culture with Accutase (around 3min, RT), and collected in a centrifuge tube. The Accutase in the cell suspension was then diluted out 1:10 in 10% Serum Medium, and cells were pelleted by centrifuging 200xg, 4min, 4C. The supernatant was removed, and the pellet was washed by resuspension in 10mL PBS-/-. Cells were re-pelleted by centrifuging 200xg, 4min, 4C and washed once more in 10mL fresh PBS-/-. After re-pelleting once more (200xg, 4min, 4C), the pellet was dissociated as a single-cell suspension in 1mL N2B27 Medium. Cells were counted with a haemocytometer, and, for each plate of Gastruloids made, 37500 cells (SBR line) or 93750 cells (TFoxA2 line) were transferred to 5mL fresh N2B27 (7.5 cells/uL or 18.75 cells/ul final concentration, respectively). The cell suspension was distributed as 40uL droplets (=300 SBR cells/droplet, or 750 TFoxA2 cells/droplet)

in wells of a U-bottomed, low-adhesion, 96 well plate, and the plates were left for 120h in a humidified incubator, 5% CO2, 37°C. At 48h after plating, 150uL of 3uM CHIR99021 N2B27 were added to each well, and this solution was substituted with fresh N2B27 (no CHIR) every 24h after that. A step-by-step detailed protocol is available at: https://dx.doi.org/10.17504/protocols.io.9j5h4q6. **Recipes**: 10% Serum Medium: 86.8% DMEM, high glucose, with 3.97mM GlutaMAX (L-Alanyl-Glutamine, final concentration: 3.45mM), 10% ES-grade FBS, 100U/mL Penicillin, 100ug/mL Streptomycin, 0.1mM Non Essential Amino Acids, 1mM Sodium Pyruvate, 1mM beta-mercaptoethanol. N2B27: 47.4% Neurobasal Medium, 47.4% DMEM/F-12, with 2.50mM GlutaMAX (L-Alanyl-Glutamine, final concentration: 1.18mM), 1mM GlutaMAX Supplement (total concentration: 2.18mM), 100U/mL Penicillin, 100ug/mL Streptomycin, 0.1mM Non Essential AminoAcids, 1mM Sodium Pyruvate, 1mM beta-mercaptoethanol, 1% B27Supplement, serum-free, 0.5% N-2 Supplement.

Gastruloid immunostaining: Gastruloids were collected at every given timepoint, washed in PBS-/-, and fixed in 4% PFA in PBS-/-, for 2h, 4C, on a low-speed or-bital shaker; or 45min, RT, static. Gastruloids were then washed in PBS+FT (PBS-/-, 10% ES-grade Foetal Bovine Serum, 0.2% Triton-X100), and blocked and permeabilisedin PBS+FT for 1h, RT, static. Primary antibody solutions were then prepared in PBS+FT, with 2ug/mL DAPI. Samples were stained overnight, 4C, on a low-speed orbital shaker. Similarly, secondary antibody solutions were prepared in PBS+FT, 2ug/mL DAPI, and samples were stained overnight, 4C, on a low-speed orbital shaker. Gastruloids were mounted in Fluoromount-G mounting medium (no spacers), and slides kept at 4C long term. All antibody solutions were washed away after incubation by washes in PBS+FT. A detailed, step-by-step protocol is available at: https://dx.doi.org/10.17504/protocols.io.7tzhnp6. A list of the primary antibodies used is provided in Table A.6. Secondary antibodies used were all from Thermo Fisher Scientific: donkey anti-mouse IgG Alexa 647 (CAT#A-31571, RRID: AB_162542); donkey anti-rabbit IgG Alexa 488 (CAT#A-21206, RRID: AB_2535792), Alexa 568 (CAT#A-10042, RRID: AB_2534017), or Alexa 647 (CAT#A-31573, RRID: AB_2536183); donkey anti-rat IgG Alexa 488 (CAT#A-21208, RRID: AB_2535794), goat anti-rat Alexa 568 (CAT#A-11077, RRID: AB_2534121), or Alexa 647 (CAT#A-21247, RRID: AB_141778); donkey anti-goat IgG Alexa 488 (CAT#A-11055, RRID: AB_2534102), or Alexa 568 (CAT#A-11057 RRID: AB_2534104).

Gastruloid imaging and image processing: Bright-field images of Gastruloids were taken on either a Nikon Ti inverted spinning-disk microscope or an Olympus CellR inverted widefield microscope (UPLAN S APO 10x/0.40 air objective, CCD Grayscale Hama-matsu ORCA ER B7W Camera; Olympus XCellence software for data capture). Both microscope setups had CO2 and temperature control (37°C and

5% CO2). Immunostained Gastruloids were imaged on a Zeiss LSM700 inverted confocal microscope (Plan-Apochromat 20x/0.80 air objective, motorized stage, LED Lumencor SOLA Illumination, CCD Grayscale Axiocam MRm (B/W) Camera; ZEN 2009 software for data capture) or on a Zeiss LSM780 inverted confocal microscope (for the CDH1 projection, Plan-Apochromat 20x/0.80 air objective, motorized stage). Images were opened, stitched, and processed for publication (LUT assignment, channel display, min and max intensity thresholding based on no-primary control) using the Fiji ImageJ distribution (Rueden et al., 2017; Schindelin et al., 2012), and the "Grid/Collection Stitching" plugin therein (Preibisch et al., 2009).

Quantification of AP patterning: Batch quantification of immunostaining signal intensity along the AP axis of the Gastruloids was performed through a custom processing pipeline available as a Jupyter notebook at https://doi.org/10.5281/zenodo.4899121 and outlined as follows (step-by-step walk through provided in the notebook itself). The pipeline takes two inputs: i) the multichannel raw image resulting from the scan of an entire microscope slide of immunostained and mounted Gastruloids (here acquired on a GE Healthcare IN Cell Analyzer 2200 automated microscope) and ii) hand-traced line coordinates defining the central axis of each Gastruloid on the slide (starting from the posterior). At early timepoints where the posterior of the Gastruloid is not distinguishable morphologically, the area of TBra polarisation is to be used instead. The script then subdivides each interval of the line ROI provided into n finer intervals of equal length (thus avoiding to have to manually draw a line with high number of points; here n=10), and for each point along the line it defines a non-overlapping polygon mask covering an area of thickness N (here N=500px) across the line and whose lateral edges are orthogonal to the line itself at each side of the point. Having computed the mask, the script then assigns the total signal intensity recovered in the area to the point of the line ROI around which the polygon was constructed, thus effectively assigning signal intensities to points that can be ordered along an x-axis. These raw values are then normalised by the number of cells in the area (using the DAPI nuclear intensity as a proxy) and both position along the length of the Gastruloid and signal intensity are normalised to the absolute length of the Gastruloid and to the maximal DAPI-normalised intensity value. The script outputs lineplots and scatterplots for each Gastruloid analysed, summary lineplots and scatterplots with collated data of all gastruloids analysed, and the tabulated raw data for re-use.

Quantification of co-expression from immunostaining data: processing of immunostaining signal intensity to quantify and map co-expression of marker pairs was performed through a custom processing pipeline outlined as follows. Initially, the combined intensity data of the two channels stained for each marker was run

through the deep-learning based algorithm Cellpose (Stringer et al., 2021) to obtain a mask of individually labelled segmented cells. These cells (considered for analysis) are thus all cells where either marker (or both markers) was found to be expressed by immunostaining. A more comprehensive segmentation of all cells of the gastruloids (including cells expressing neither marker) was not possible due to the poor performance of Cellpose on the DAPI channel of gastruloids (especially at later timepoints). The cell labels were then converted to ROIs through the Fiji plugin "`LabelsToROIs`" (see Waisman et al., 2021). These coordinates, along with the original immunostaining image data, were then passed to a custom Jupyter notebook for quantification. The notebook extracts intensity values on each channel for each cell ROI, and then normalises these values with respect to the highest intensity found within each marker channel, as well as with respect to the intensity measured from the DAPI channel for any given cell. The code outputs tabulated raw data files, and colour-coded scatterplots of expression values of each cell with respect to both markers. The pipeline finally classifies cells as "double-negative" if they express both markers at lower levels than average, "single-positive" for either marker if they express one marker at higher level than average and the other marker at lower level than average, and "double-positive" if they express both markers at higher level than average. The proportions of each of these categories are calculated and returned as output, as well as visual maps where each ROI, colour-coded based on its assigned category, is overlayed to the DAPI image of the sample.

Gastruloid Hybridisation Chain Reaction (HCR): Gastruloids were collected at t=168h, washed in PBS-/-, and fixed in 4% PFA in PBS-/-, overnight, 4C, on a low-speed orbital shaker. Gastruloids were then washed in PBS-/-, and then dehydrated in a graded series of methanol-PBST solutions (0%-100%, 25%-75%, 50%-50%, 75%-25%, 100% methanol). Gastruloids were then stored in 100% methanol, -20C, until use (and at least overnight). When needed, gastruloids were rehydrated in a graded series of methanol-PBST solutions (100%-0%, 75%-25%, 50%-50%, 25%-75%, 100% PBST), digested in 25ug/mL Proteinase K in PBST, 4min, RT, washed in PBST, and re-fixed in 4%PFA in PBS-/-for 20min, RT. For the probe hybridisation step, samples were washed in PBST, pre-incubated 1h30min in warm Probe Hybridisation Buffer, 37C, and then incubated for 16-20h with 4pmol of odd HCR probes and 4pmol of even HCR probes mixed in Probe Hybridisation Buffer, 37C. For the amplification step, samples were washed in warm Probe Wash Buffer, 37C, further washed in RT 5XSSCT, and then left to incubate for 16-20h with 48pmol of hairpin 1 and 48pmol of hairpin h2 (for each colour used) mixed in Probe Amplification Buffer with 2ug/mL DAPI, RT. Each hairpin was heated to 95C for 1min30s, and snap-cooled at RT for at least 30min before use. After amplification, samples were incubated for

1h15min in 5XSCCT with 2ug/mL DAPI, washed in 5XSCCT, and then mounted on microscope coverslips in Fluoromount G mounting medium. A step-by-step detailed protocol is now published at https://dx.doi.org/10.17504/protocols.io.bcwfixbn The sequences of the Pax9 probe set used (coupled to amplifier B5) is provided in Table A.7. The Pax1 probe set used (coupled to amplifier B3) was custom bought from Molecular Instruments Inc. and its sequences were unfortunately not disclosed by the company.

HCR probe design: Pax9 HCR probes were designed using a custom R-based Jupyter Notebook developed from an initial pipeline authored by T. Fulton (Steventon Lab, University of Cambridge, UK). Briefly, the pipeline downloads and reformats the NCBI FATSA mucleotide sequence of the mRNA of the gene of interest, subidivides it in all possible 25 nucleotide-long fragments (the length of an HCR probe), and only keeps those that pass quality control metrics (GC content between 40% and 70%; melting temperature between 55C and 77C). Finally, of the fragments that passed quality control, the pipeline only keeps those that have a quality partner exactly 2 nucleotides downstream (HCR hairpin binding requires a 2 nucleotide spacing between elements of a pair). The position of these probe candidates along the mRNA is then visualised as to allow to select 10 pairs that are equally spaced along it. The desired HCR amplifier sequence is then selected and the final probe sequence is outputted.

Processing of scRNAseq datasets: All data analysis was done on R, with the Seurat v4.0 library (Hao et al., 2021). Gastruloid dataset: scRNAseq data corresponding to Gastruloids spanning timepoints t=96-168h was taken from Rossi et al., 2021a. Raw count matrices for both batches of each timepoint and for different timepoints were merged, and filtered based on the following quality control parameters: number of Unique molecular identifiers > 10000, number of Genes > 2000, Complexity > 0.75, Percentage of mitochondrial genes < 15%. Genes expressed in 0 or less than 5 cells of the dataset were discarded. The data underwent normalisation, variance stabilisation, and differences due to mitochondrial content and cell cycle phase were regressed out via Seurat's `NormalizeData` and `SCTransform` functions. Data from different timepoints was integrated using `SelectIntegrationFeatures` on the top 3000 genes, `Find-IntegrationAnchors`, and `IntegrateData`. PCA and UMAP (on the first 40 dimensions) were then calculated through `RunPCA` and `RunUMAP`, respectively. Clustering was done via the functions `FindNeighbours` and `FindClusters`, by a shared nearest neighbor (SNN) modularity optimization based clustering algorithm on a SNN graph based on the 20 nearest neighbours (default). A resolution of 0.8 was chosen to proceed with analysis. Cluster identities were assigned based on the patterns of expression of selected marker genes, along with an analysis of the genes

marking each cluster (`FindaAllMarkers` function, limiting testing to genes which showed, on average, at least 0.25-fold difference (log-scale) between the two groups of cells; default). Top genes were ranked based on the difference between their pct.1 and pct.2 values (i.e. between the percentage of cells expressing a given gene in the cluster of interest versus in the other clusters combined). To find markers differentiating close clusters, the function `FindMarkers` was used instead. The annotated notebook, with a step by step walkthrough the entire analysis pipeline, is available at my GitHub repository https://github.com/StefanoVianello/Endoderm_scRNAseq, "Gastruloid_scRNAseq_preprocessing_RNotebook.Rmd".

Gut endoderm dataset: scRNAseq data corresponding to Gut endoderm cells at Embryonic Day 8.75 was taken from Nowotschin et al., 2019b. The raw count matrix was imported as a Seurat object and filtered based on the following quality control parameters: number of Unique molecular identifiers > 5000, number of Genes > 3000, Complexity > 0.75, Percentage of mitochondrial genes < 20%. Genes expressed in 0 or less than 5 cells of the dataset were discarded. The data underwent normalisation, variance stabilisation, and differences due to mitochondrial content and cell cycle phase were regressed out via Seurat's `NormalizeData` and `SCTransform` functions. PCA and UMAP (on the first 30 dimensions) were then calculated through `RunPCA` and `RunUMAP`, respectively. *In vivo* and *in vitro* data was integrated using `SelectIntegrationFeatures` on the top 3000 genes, `FindIntegrationAnchors`, and `IntegrateData`. PCA and UMAP (on the first 40 dimensions) were then calculated through `RunPCA` and `RunUMAP`, respectively. Clustering was done via the functions `FindNeighbours` and `FindClusters`, by a shared nearest neighbor (SNN) modularity optimization based clustering algorithm on a SNN graph based on the 20 nearest neighbours (default). A resolution of 0.8 was chosen to proceed with analysis. Cluster identities were assigned based on the patterns of expression of selected marker genes, along with an analysis of the genes marking each cluster (`FindMarkers` function, limiting testing to genes which showed, on average, at least 0.25-fold difference (log-scale) between the two groups of cells; default). Top genes were ranked based on the difference between their pct.1 and pct.2 values (see explanation above). The annotated notebook, with a step by step walkthrough the entire analysis pipeline, is available at my GitHub repository https://github.com/StefanoVianello/Endoderm_scRNAseq, "GutTube_scRNAseq_preprocessing_RNotebook.Rmd".

Alignment of Gastruloid and Gut tube cells: Cells corresponding to endoderm clusters were subsetted from the Rossi et al., 2021a dataset, processed as described above. The identification of the endoderm clusters is justified in the text of this chapter. Cells corresponding to the embryonic gut tube were subsetted from the Nowotschin et al.,

2019b dataset, processed as described above. The gut tube cluster is the biggest cluster in the dataset, and its endodermal identitiy was confirmed based on the expression of classic endodermal markers. The count matrices of the two subsetted datasets (*in vitro* endoderm and *in vivo* endoderm) were merged into a single object and processed according to standard pipeline. Genes expressed in 0 or less than 5 cells of the dataset were discarded. The data underwent normalisation, variance stabilisation, and differences due to mitochondrial content and cell cycle phase were regressed out via Seurat's `Normalize-Data` and `SCTransform` functions. PCA and UMAP (on the first 30 dimensions) were then calculated through `RunPCA` and `RunUMAP`, respectively. Clustering was done via the functions `FindNeighbours` and `FindClusters`, by a shared nearest neighbor (SNN) modularity optimization based clustering algorithm on a SNN graph based on the 20 nearest neighbours (default). A resolution of 0.8 was chosen to proceed with analysis. Cluster identities were assigned based on the patterns of expression of selected marker genes, along with an analysis of the genes marking each cluster (`FindaAll-Markers` function, limiting testing to genes which showed, on average, at least 0.25-fold difference (log-scale) between the two groups of cells; default). Top genes were ranked based on the difference between their pct.1 and pct.2 values (see explanation above). To find markers differentiating close clusters, the function `FindMarkers` was used instead. The annotated notebook, with a step by step walkthrough the entire analysis pipeline, is available at my GitHub repository https://github.com/StefanoVianello/Endoderm_scRNAseq, "Endoderm_comparison_RNotebook.Rmd".

Conclusion and Outlook

5

The work presented in this thesis describes the developmental progression of endoderm (here taken as to mean "cells expressing endodermal markers") *in vitro*, and specifically in gastruloids. In other words, endoderm behaviour was studied in a a 3D context, and in a 3D context known to recapitulate many features of *in vivo* embryonic development, yet lacking the maternal and extraembryonic cues available to the embryo. As outlined in *e.g.* Steventon et al., 2021, gastruloids provide an experimental platform to probe the intrinsic features and behaviours of embryonic populations in dialectic comparison with what known to happen *in vivo*. The field of experimental/synthetic embryology, more immediately focused on neurectoderm and mesoderm development (with great insight in such areas, see *e.g.* Veenvliet et al., 2020; Beccari et al., 2018; Rossi et al., 2021a; Brink et al., 2020; Bérenger-Currias et al., 2020), has in fact generally lagged in terms of detailed investigation of the modes and capabilities of endoderm development in these same systems. At the same time, endoderm is likely the germ layer with the most intimate relationship with its extraembryonic counterpart and for which such a decoupled system offers most interesting terms of comparison.

After providing an overview, review, and summary of our current understanding of mouse endoderm developmental biology *in vivo* in Chapter 1, I demonstrate in Chapter 2 that endoderm progenitors emerge, *in vitro*, in temporal and spatial patterns comparable to those of the mouse embryo. I identify steps of endoderm specification that occur throughout an uninterrupted epitheliod state, and that can progress in the absence of extreaembryonic tissues or of organised epithelia in which to integrate. In Chapter 3 I describe dramatic morphogenetic events involving endoderm in gastruloids, involving the reproducible formation of complex architectures and epithelial maturation at the interface of other germ layers. Through the study of live reporters, I further describe highly motile behaviour of endoderm cells and aggregative behaviour underlying the morphogenesis of this cell type. Finally, in Chapter 4, I show that the endoderm domain formed in gastruloids is patterned along the AP axis. In fact, drawing from single-cell transcriptomics data, I show that gut-tube-like endoderm matures within gastruloids, and that this endoderm generates cell identities spanning the entire breadth of cell identities found in the embryonic gut tube. I notably describe the spontaneous emergence, *in*

vitro, of anterior foregut endoderm types, and how these are correctly patterned at the anterior of the gastruloid. Overall, this work builds gastruloids as a valid and promising platform for the study of endoderm behaviour *in vitro*. Building on these results may be studies aimed to investigate the developmental interaction between endoderm and surrounding germ layers, as well as — potentially — more applied approaches to derive endodermal types that may be non-trivial to derive *in vitro* through other methods.

The contribution of the work presented in this thesis is thus to be positioned within the emerging literature in the field of synthetic embryology and *in vitro* models of embryonic development, and more specifically alongside most recent reports that have similarly started to approach the question of gastruloid endoderm development in these systems, if from alternative vantage points. These results can also be inserted within the broader context of mouse endoderm biology, especially given recent findings on endoderm development *in vivo* that have picked up similar endoderm behaviours as those observed here *in vitro* (e.g. Scheibner et al., 2021; Probst et al., 2021).

5.1 Contextualisation with current gastruloid literature

Cells expressing endodermal markers, and gene expression patterns consistent with endodermal identities, have been described since the initial characterisations of gastruloid development (Beccari et al., 2018; Turner et al., 2017; Brink et al., 2014). With few exceptions however (detailed in what follows), these descriptions had been limited to the simple identification of endoderm as one of the germ layers derivatives of the system rather than as an investigative focus *per se*. Accordingly, the earliest gastruloid work in Brink et al., 2014 already described polarised emergence of cells expressing both FoxA2 and Sox17: a compact FoxA2^{+} domain clustering at the posterior of late stage gastruloids, occupied by Sox17^{+}/TBra^{-} cells, and internalising as part of epithelial vesicles. This posteriorly-restricted pattern of expression is difficult to reconcile with what described in this thesis, but it is also true that in those cell lines where the Cdh1^{+} primordium does not extend throughout the length of the aggregate it does in fact restrict posteriorly (data not shown, but see FGF4-treated deficient gastruloids in Gharibi et al., 2020). The described invagination of Cdh1^{+}/Sox17^{+} cells may however indeed explain the numerous surface connections of the Cdh1^{+} primordium observed at the posterior of the gastruloids and as shown in this work.

Early work by Turner et al., 2017 had also been describing aspects of endoderm development in gastruloids. When imaging a Sox17 reporter, midline tubular Sox17$^+$ patterns of expression could be detected in elongating gastruloids, and were in fact likened to ventral endodermal cells of the E8.5 mouse embryo. Based on the characterisation of Sox17$^+$ cells presented in this thesis, and the extensive Sox17 positivity of the Cdh1$^+$ primordium, one would expect the reporter line used in Turner et al., 2017 to be equally giving rise to an internal endodermal primordium, even if such compartment was not explicitly identified at the time. Interestingly what is here inferred by immunostaining seems to be consistent with the early Sox17 dynamics described in the paper: Sox17$^+$ cells emerging towards the anterior pole of the early gastruloid and then expanding to occupy a relatively more posterior domain.

Sox17$^+$ cells were also described as part of the extensive characterisation of gastruloids published in Beccari et al., 2018. These cells were further described as forming tubular structures based on 4,6-diamidino-2-phenylindole (DAPI) counterstaining. More importantly, the publication described gene expression signatures suggesting the presence of maturing endoderm: the early upregulation of markers such as Gsc and Cdx2, upregulation of Cer1 during elongation (and which indeed appear in the "early endoderm" cluster discussed in Chapter 4), and expression of gut endoderm markers during later development (Nedd9, Sorcs2, Pax9, Pyy, Shh, Krt18). *In situ* hybridisation patterns for some of these markers are again consistent with the presence of an internal endodermal structure. Validation of the maturation of endodermal identities can also be found in the single-cell transcriptomics data generated by Brink et al., 2020, which detect a cluster of cells postulated to represent definitive endoderm as expressing markers Epcam, Col4a1, Sox17. All of these markers are indeed recovered within the endoderm clusters considered here. Still, detailed spatial characterisation had been lacking: gastruloids remained framed mainly as mesenchymal neuromesodermal aggregates.

Only more recently, two studies have instead approached gastruloids from an explicitly endodermal perspective. Pour et al., 2019 attentively detailed the earliest steps of endoderm progenitor emergence through the use of a TBra/Sox17 double reporter cell line. The (relative) temporal dynamics they describe in the paper are consistent with those recovered by timed immunostaining in Chapter 2: intermingled amongst TBra$^+$ cells, Sox17$^+$ cells first appear around 1 day after CHIR exposure, and are found to be Cdh1$^+$. Because the authors do identify Sox17$^+$/Cdh1$^-$ cells they ultimately favour an interpretation supporting a mesendodermal origin to endoderm cells. Second study of immediate relevance is that of Hashmi et al., 2020, also entirely dedicated to the developmental morphogenesis of endoderm in gastruloids,

and which takes a focused approach to the early steps of endodermal morphogenesis. Their findings that endoderm cells always seem to retain Cdh1 expression, and their description of endoderm island flowing and streaming across the gastruloid to then form a segregated domain are extremely coherent with what shown and described in Chapter 2. The authors attribute these sorting dynamics to differences in interfacial tension across cell types (Hashmi et al., 2020).

Finally, a number of recent studies have started to observe the generation of internal tubular endodermal structures within gastruloids (and in fact, of structure in itself, something long considered to lack in gastruloids). Accordingly, Veenvliet et al., 2020 describes tubular structures populated by FoxA2$^+$ and few Sox17$^+$ and TBra$^+$ cells. Associated with this gut-like domain, Dppa3$^+$ cells (postulated to represent Primoridal Germ Cells) known indeed to migrate along the gut tube during *in vivo* development. The observations of analogous structures in "free-floating" gastruloids (presented in Chapter 3) challenge however the apparent necessity of ECM embedding to trigger endoderm morphogenesis, as put forwards by the authors. Unlike mesoderm thus, which seems to require in vitro ECM cues for productive MET (Brink et al., 2020; Veenvliet et al., 2020), morphogenesis of the endoderm appears to be intrinsic to the tissue. Endoderm cells in fact seem to be able to organise their own extracellular environment. Recent work on human gastruloids (gastruloids obtained from human induced pluripotent cells) by Olmsted and Paluh, 2021 further showed that such a morphogenetic behaviour appears to be conserved: epithelial "primitive gut tubes" form in human gastruloids under shaking culture (*i.e.* without embedding). These structures expresses classic endodermal markers (FoxA2, Sox17, Gata6, Sox2, Cdh1, and Epcam), as well as markers of more specialised cell types such as Lgr5, Nkx2.1, Lyz, and Vil1 over 16 days of culture. No information is however provided on the extent to which such markers are patterned (Olmsted and Paluh, 2021).

Most recently Xu et al., 2021 reported descriptions of patterned endodermal domains in an embryoid model in many ways analogous to gastruloids, underscoring the extent to which such morphogenetic and patterning behaviour does indeed seem an intrinsic feature of endodermal cells regardless of the developmental context in which they are located. Accordingly, Xu et al., 2021 describe the formation of tubular, lumenised Sox17$^+$ epithelia by vesicular fusion. More strikingly, these endodermal domain show show both molecular patterning (anterior Pyy, mid Nepn, posterior Rnf128) and morphological patterning (domains the authors liken to anterior and posterior embryonic gut pockets, and a posterior branch connecting to the surface of the embryoid likened to the embryonic yolk stalk). As discussed in Chapter 2, the claim that these endodermal domains are co-developing alongside VE cells questions

the demonstrated absence of extraembryonic endoderm types within gastruloids. Still, we could not unambiguously confirm the presence of extraembryonic cell types in gastruloids and in fact find inconsistencies between *in vivo* and *in vitro* markers (*cfr* DBA results). The structuring and patterning processes described in this thesis thus reveal modes of endoderm development deployed in the absence of extraembryonic types and embryonic structure, underscoring remarkable intrinsic morphogenic and sorting/patterning behaviours of this cell type.

5.2 Points of discussion

5.2.1 Mesendodermal progenitors and modes of EMT

While the current evidence (for and) against the functional existence of mesendodermal progenitor intermediates to embryonic endoderm development has already been presented in an earlier section dedicated to the matter, the possibility always remains that — away from the tight spatiotemporal constraints of the embryonic environment — the same populations might reveal a mesendodermal potential *in vitro* (Nowotschin and Hadjantonakis, 2020; Wang and Chen, 2016; Sui et al., 2013). While not focused on the identification of mesendoderm itself, the investigations of the earliest steps of endoderm emergence presented in Chapter 2 provide some elements of relevance to the matter.

Mesendodermal progenitors would be expected to be identified as $TBra^{+}/FoxA2^{+}$ double positive cells within the EPI compartment, then shown to be able to produce both mesodermal and endodermal cell contributions. Certainly, the typing of mesendodermal identity in gastruloids is complicated by the fact that a distinction between EPI-equivalent compartments and non-EPI compartments in is not immediate and as such $T^{+}/FoxA2^{+}$ double positive cells could well represent the known double-positivity of cardiac mesoderm progenitors as well as that of axial mesoderm structures. We do observe $T^{+}/FoxA2^{+}$ double positive cells within the $Cdh1^{+}$ epithelioid context of the early gastruloid, at the time of "gastrulation". Imputing mesendodermal identity to it would however require the tracking of these cells to ascertain their fate. In turn, this would require the tracking of individual cells of the TBra/FoxA2 double reporter cell line used in Chapter 3 at higher resolution to those immediately available to us, and the identification of *e.g.* $TBra^{+}/FoxA2^{+}$ cells then resolving into a $FoxA2^{+}$ single positive lineage (that is, of a double positive intermediate to a restricted endodermal lineage).

Other literature that has described mesendodermal progenitors in gastruloids or embryoids has used Sox17 (and TBra) as a mesendodermal discriminants and comes to contrasting results (Xu et al., 2021; Pour et al., 2019). Xu et al., 2021, not finding any $TBra^+/Sox17^+$ cells at the time of "gastrulation" concludes absence of mesendoderm. Pour et al., 2019, on the other hand, finds such double positive cells and ultimately takes these cells as mesendoderm representatives. Sox17 is however a relatively late marker of endodermal lineage (Viotti et al., 2014), and thus maybe lesser informative of shared early developmental steps that would or would not be intertwined with mesodermal ones.

By consistently observing $FoxA2^+$ cells within a $Cdh1^+$ context, in domains that appear spared from classical Snai1-mediated EMT, we favour the interpretation also put forward by Hashmi et al., 2020: endoderm in the gastruloid forms through separate mechanisms than mesoderm, and in ways that do not depened on EMT but rather from the active or passive fragmentation of $Cdh1^+$ tissues, and continued association of $FoxA2^+/Sox17^+$ with such cell domains.

5.2.2 Epithelial plasticity and rearrangement

Second feature clearly highlighted by the endoderm behaviours described in this thesis is the intimate and persistent association between endoderm identity and epitheliality, in fact already appreciated from embryonic studies (if only in the notion that endoderm cells egress from the primitive streak while maintaining Cdh1 expression; Viotti et al., 2014) and reflected in the embedding of both FoxA2 and Sox17 within gene regulatory networks associated with epithelial identity (Burtscher and Lickert, 2009; Viotti et al., 2014; Scheibner et al., 2021). What gastruloids reveal is a high degree of plasticity and morphogenic potential of $Cdh1^+$ cells in forming dynamically remodelling aggregative structures within a surrounding mesenchymal context. Endoderm self-organisation in gastruloids also raises important questions regarding the necessity of VE (or target epithelia in general) for endoderm development and maturation. Clearly integration of DE cells into the VE *in vivo* is necessary *de facto* (Kanai-Azuma et al., 2002). Yet if these cells could reintegrate within the epithelium of the EPI (which they do not seem to be able to do) would they be able to mature regardless? Can (and do) endoderm cells that would find each other within the mesodermal wings of the embryo mature into a self-established local epithelial context, without the need for integration within the overlaying VE? The observations reported herw in Chapter 3 and4 suggest that this would be the case.

Generally, endoderm behaviour in gastruloids may be expose "backup" or "emergency" developmental strategies endodermal cells may hold the potential for but which they do not need to execute during normal development. Anecdotally, the highly plastic dynamic fragmentation and coalescence of Cdh1$^+$ endoderm in gastruloids is reminiscent of *e.g.* behaviours recently described in other developmental contexts (Teng et al., 2021). It happens that upon tissue fusion of opposing shelves of the secondary palate the epithelium at the seam must then be removed, and that during this removal tissue forms trails and islands as it streams away through the surrounding mesenchymal tissue, while also undergoing epithelial maturation. The imaging and morphological patterns of Cdh1$^+$ tissue in Teng et al., 2021 is extremely reminiscent of those of Cdh1$^+$ endoderm described here. In fact studies of different migration strategies of epithelial and epithelioid collectives are not new (*cfr.* *e.g.* the role of cadherins in neural crest cells already discussed in the Introduction). Given that little is known about the migratory parameters of endoderm *in vivo*, could gastruloids be offering a window to signature aggregative epithelioid tendencies that may be a big component of such migration dynamics (and maybe extremised in the gastruloid context)?

Finally, notable is also the observation that the gastruloid endodermal primordium undergoes some degree of epithelial maturation, transitioning from an epitheliod state to developoing apicobasal polarity and forming a lumenised monolayer epithelium. Since such an internal lumen has since been reported in other gastruloid/embryoid endodermal descriptions (Xu et al., 2021; Veenvliet et al., 2020; Olmsted and Paluh, 2021) this seems again to be a recurring feature of endoderm development *in vitro* and thus likely an intrinsic feature of it. In the embryo the tube-like structure of the gut tube is formed by embryo/tissue-level topological transformations associated with ventral folding morphogenesis, and not exactly by the movement of endoderm cells within the epithelium *per se* (Gavrilov and Lacy, 2013). Expansion of the luminal space by gut tube cells in the mouse embryo has mainly been studied with respect to intestinal morphogenesis, but takes place at much later developmental stages than those that could be expected to be recapitulated by gastruloids (Walton et al., 2016; Guiu and Jensen, 2015). Could however endoderm cells in the gastruloid here be actuating modes of development that have been described in *e.g* zebrafish, where the gut tube cavity does indeed form by active lumenogenesis (Alvers et al., 2014)?

5.2.3 Patterning of anterior foregut identities

Third important feature emerging from the results presented in this work, and specifically from Chapter 4, is that — by developing through a gastruloid context — endoderm seems to be able to differentiate into multiple mature identities concomitantly, and in a spatially patterned way as such. That is, gastruloids can be employed through undirected differentiation as a generative platform for the maturation o gut endoderm cell types spanning the entire length of the embryonic AP axis. Counter to this is of course the fact that, aside from the possible difficulties of generating specific endodermal types *in vitro*, translational applications most often struggle with generating pure population of single endodermal types of interest. The gastruloid approach does not address this issue and in fact complicates it due to the presence of adjacent cell types not only of other endodermal identities, but of other germ layers. One can however imagine to apply surface-marker-based sorting and isolation strategies to late stage gastruloids and through this obviate to the problem while simultaneously obtain endodermal identities obtained through embryo-like developmental processes.

Of the identities maturing within the Cdh1^{+} primordium, those maybe most surprising are those corresponding to the anterior foregut (and in fact to the anterior anterior foregut), indeed the focus of the HCR staining performed in Chapter4 to confirm their correct anterior spatial patterning. As mentioned in previous sections Pax9 marks the anterior foregut of E8.5 and the pharyngeal puches of E9.5 mouse embryos (Peters et al., 1998; Neubüser et al., 1995) and Pax1 marks the pharyngeal endoderm and the epithelium of the 3rd pouch specifically (Wallin et al., 1996). Accordingly, Pax9 mutant mice lack the derivatives of the third and fourth pharyngeal pouches (Peters et al., 1998). Yet most of the surprise comes from the contrast between the robust development of anteriormost endoderm identities, and the post-occipital restriction of gastruloids in both their mesodermal and neural compartments (Beccari et al., 2018). Of note, in a recent report of gastruloids generated from zebrafish cells, the only type of endoderm produced is indeed pharyngeal endoderm (Cheng et al., 2021). In monolayer differentiation approaches, (EpCAM^{+}/FoxA2^{+}/)Pax9^{+} anterior foregut cells can be generated by stimultating pluripotent stem cells with Wnt- and Transforming growth factor beta (TGFβ) agonists to induce DE differentiation, optionally followed by prolonged BMP and TGFβ inhibition (Kearns et al., 2013). More generally, anteriorisation of endodermal fates obtained by 2D differentiation approaches has consistently been achieved by supply of BMP and TGFβ inhibitors, which favour foregut at the expense of hindgut commitment (Green et al., 2011; Mou et al., 2012). It would thus seem likely that

the anteriormost domain of the Cdh1^{+} primordium is maintained under such an inhibitory context (and in fact the apex of the foregut pocket *in vivo* is nested within the BMP-suppressed environment of the headfolds of the early embryo; Madabhushi and Lacy, 2011).

Importantly however, we could not detect, even within Pax9^{+} gastruloid cells, transcriptional signatures of more advanced pharyngeal (*e.g.* thyroid Pax8) or of specifically 3rd pharyngeal pouch derivatives such as the parathyroid (*Gcm2*) and, even more relevantly for the research done in our lab, thymus marker *Foxn1*. This suggests that either the more regionally-specific signalling context required for the specification of these mature structures is not spontaneously recapitulated in gastruloids, or that gastruloids may need to be cultured for longer than t=168h to allow the maturation of such identities.

Importantly, the majority of endodermal cells in gastruloids are most transcriptionally similar to midgut identities and to progenitors of the intestine in the gut tube. These cells express genes such as *Map1b*, *Fabp1*, *Cxcl12*, *Kitl1*. Importantly, they express *Shh* and indeed *in situ* hybridisation against this signalling molecule had shown to mark gastruloid domains consistent to the internal Cdh1^{+} primordium I have here shown to exist in gastruloids (Beccari et al., 2018). The predominance of such identity is likely resulting from the high levels of Wnt and FGF signalling known to dominate the gastruloid posterior (Beccari et al., 2018; Brink et al., 2014) and known to have a posteriorising effect on endoderm cells *in vivo* and *in vitro* (Wells and Melton, 2000; Ameri et al., 2010).

5.3 Contextualisation with mouse endoderm literature

Within the three years the work presented in this thesis has been carried out, research in mouse embryo endoderm biology has been marked by important contributions that have greatly advanced our understanding of the modes of development of this cell type *in vivo*. As abundantly referenced and leveraged throughout the thesis, transcriptomic characterisation of the embryonic gut tube at single cell level has not only provided valuable reference datasets to type endoderm identities along the gut tube, but has in itself greatly advanced our understanding on the cellular and molecular heterogeneities within the gut tube and their relationship with earlier stages of endoderm specification (Nowotschin et al., 2019b). Work by Probst et al., 2021 has delved into the specificities of mesendodermal transitions as endoderm is specificed during gastrulation. Investigations by Scheibner et al., 2021 have most

recently highlighted the details of the molecular underpinnings that explain the alternative EMT modes deployed by endoderm cells as they egress from the EPI (all detailed in the Introduction).

When working with *in vitro* models of embryonic development, one would aim to be able to continuously draw comparisons between what observed to happen *in vitro* and what observed/known to happen *in vivo*(Steventon et al., 2021). In fact, an important part of the validity of *in vitro* models of embryonic development rests in their potential to reveal underlying features of developmental biology that may in fact be those deployed in the embryo (with insight coming from either agreement or disagreement with *in vivo* observations). Under this perspective, the recentmost demonstration that nascent endoderm expresses epithelial markers at all stages of development *in vivo*, from specification within the EPI to maturation as DE, and that its formation relies on plastic epithelial transitions across compartments rather than mirrored EMT/MET events (Scheibner et al., 2021), highlight the value of gastruloids as an *in vitro* study platform, and a platform where such features clearly come to light. The concordance and mutual support between *in vitro* developmental models and *in vivo* embryo work are extremely encouraging and build a model of scientific investigation that can confidently exploit the synergistic effect of the two to the benefit of Developmental Biology and of a fuller understanding of early embryonic developmental strategies.

5.4 Outlook

The results reported in this thesis bring centre stage the question of the epithelial character of endodermal precursors, and its link to the fate (both in terms of identity and of location) of these cells (Ferrer-Vaquer et al., 2010; Nowotschin et al., 2019a; Viotti et al., 2014). They expand on previous gastruloid studies by taking into consideration multiple markers of endodermal identity and documenting their dynamics in relation to one another, and notably by also tracking these cells over time as they undergo morphogenesis. They further also describe sorting of these cells into a complex primordium with maturing epithelial architecture and coarsely patterned domains of gene expression along the AP axis of the organoid. By transcriptional comparison with relevant embryonic datasets (Nowotschin et al., 2019b), they highlight the emergence of cell identities corresponding to the entire length of the gut tube, with notable representation of anterior foregut, midgut, and posterior hindgut types, and show these to be patterned along the AP axis of the system.

The observations are consistent with a model whereby Sox17$^+$ cells never leave their epithelial environment (initially, the "epiblast"-like domain) and do not need to transition through a mesenchymal state, at least not through classic Snai1-mediated EMT. We speculate that, if Sox17$^+$ endodermal cells are not being directly specified within the FoxA2$^+$ epiblast, the Sox17$^+$ cells that we incongruously see in this compartment are a result of these cells sticking to neighbours with the same epithelioid character. While in the embryo the isotropic relocalisation of Cdh1 associated with egressed endodermal cells might be compatible with segregation from both epiblast (columnar epithelium) and visceral endoderm (squamous epithelium), and indeed reintegration of Sox17$^+$ cells requires re-polarisation, the situation is different in gastruloids. In these systems, FoxA2$^+$ and Sox17$^+$ cells are emerging not in a polarised columnar epithelium, but in a context that already displays isotropic Cdh1 localisation, such that these cells may remain stuck in their original compartment just by virtue of homotypic interactions (see also Hashmi et al., 2020). Described in the present work is also the expansion of the endodermal population and its internalisation within the core of the gastruloid as the surface of the aggregate start being populated by an increasing number of mesodermal cells. Interestingly, the relative position of different epithelial populations may here again be explained by the expression of different combinations of cell-surface adhesion molecules, a common sorting mechanisms (Toda et al., 2018) that sees here some support from the biased EpCAM distribution within our CDH1 primordium, enriched in the domains occupied by endodermal cells.

The identification of a maturing epithelial structure throughout late gastruloid development, contrasting with the overlaying mesenchymal mesodermal tissues enveloping it, reframes expectations regarding the extent to which fate and morphogenesis can spontaneously arise *in vitro*. While gastruloids have traditionally been pictured as aggregates of fates without corresponding organisation, increasing examples are showing that such missing morphogenesis does occur. As already shown by (Bérenger-Currias et al., 2020; Brink et al., 2020; Veenvliet et al., 2020) this can be transformatively brought about by the addition of diluted ECM components or extraembryonic cell types. For precision, one should not that such high degree architectural structure is not immediately equivalent to that of equivalent compartments in the embryo, so that the alternative architectures that surprisingly form so robustly testify of intrinsic cellular behaviours worth of consideration. The results in Chapter 3 show that differences in epithelial identities between emerging cell types may already be sufficient to generate simple architectures, and that complex epithelia may spontaneously organise *de novo* in gastruloids. Still, the extent to which the elaborate morphogenesis seen here is cell-line specific remains to be defined. Current

literature contains examples of gastruloids from a variety of cell lines and mouse backgrounds that do indeed originate such internal epithelial primordia (Beccari et al., 2018; Rossi et al., 2021a), or where one would infer such primordium to be what described (Gharibi et al., 2020; Turner et al., 2017; Veenvliet et al., 2020), as well as of gastruloids where such structures do not seem to appear (Brink et al., 2014; Brink et al., 2020) and where Cdh1$^+$ endodermal tissue segregates to the posteriormost tip of the aggregates instead. These differences likely correlate to intrinsic cell-line biases in core signalling pathways, and/or on the very degree of epitheliality maintained by these cells by the time they are exposed to CHIR (in turn possibly relating to differences in the 2D culture conditions of these cells). A systematic assessment of the cell-line specific differences in these key parameters remains to be performed, yet this hypothesis seems to be supported by the fact that pretreatment with different signalling factors can allow/prevent formation of the primordium as shown in Gharibi et al., 2020. As is the case with the cell line used in this study, we nonetheless expect all gastruloids generated from a given cell line to consistently produce the same endoderm phenotype in all gastruloids generated. Regardless, the stratified nature of endodermal and mesodermal tissues we here observed is at least broadly comparable to the configuration of endoderm and mesoderm in the embryo, and it is interesting to speculate that this interfacing may favour the development of more advanced cell fates by reciprocal signalling interactions, as *in vivo* (Bardot and Hadjantonakis, 2020; Han et al., 2020).

Having demonstrated that gastruloids generate and extensive endodermal compartment, and having characterised both the developmental events that lead to such processes and their output (a patterned endodermal domain), where does this work lead next? In its most immediate implications, the presentation of gastruloids as developmental systems with an extensive endodermal component allows their consideration for targeted studies on any of the still unresolved sides of the developmental biology of this germ layer (just as the investigation of their mesodermal component has lead to insights on the mechanisms of e.g. somitogenesis or early cardiac specification; Veenvliet et al., 2020; Brink et al., 2020; Rossi et al., 2021a). Accordingly, and as mentioned earlier, one could envision more targeted and higher resolution investigations of mesendodermal cell lineage choice and commitment at the earliest stages of endoderm specification. Knowing that gastruloids generate and pattern AP gut endoderm identities, one could in turn use to system to investigate *e.g.* the relationship between the specification of such identities and the temporal parameters of the generation of each of these cell types as endodermal cells appear at the time of gastruloid "gastrulation" (*i.e.* the link between the time of specification *in vitro* and the mature AP fate acquired within the Cdh1$^+$ primordium). Furthermore,

having shown that the central endodermal domain structures the gastruloid into a series of interfacing compartment, across germ layer, one could build on the system to study the functional dependencies between such compartments, and how the interaction between them may allow the emergence of cell identities that would not be specified otherwise. In fact, this is an avenue that has been taken up and explored by other work in the lab and that indeed did allow considerations on how the endodermal domain described here could influence- and be influenced by the overlaying cardiac mesoderm (Rossi et al., 2021a).

On a second level, there is clear scope to push the current limitations of the system towards the generation of more mature cell types. As mentioned in previous sections, by t=168h the gastruloid Cdh1$^+$ primordium expresses markers of the embryonic gut tube and early territories, but not of more advanced markers of functional specialised cellular derivatives. Attempts to obtain such specialised cellular derivatives will likely depend on the non-trivial feat of extending gastruloid culture beyond 1 week, at which the system typically collapses, and/or the supply of small-molecule steering regimens to push the spontaneous endodermal differentiation taking place within the Cdh1$^+$ primordium towards the desired endodermal identity (*e.g.* thymus, provided such differentiation protocols are known). Following up from such type of work, one could conceive more applied/translational applications of endodermal gastruloids as a generative platform sourcing endoderm identities that may difficult to obtain in DE, or that may be wanted to be derived through embryonic-like differentiation processes, and their isolation/purification by selection of appropriate markers.

Glossary

2D 2-dimensional. 64, 88, 92

3D 3-dimensional. 19, 48, 81

AP Anterior-Posterior. 15, 17, 20, 26, 28, 30, 42, 59, 60, 67, 70, 72, 75, 81, 88, 90, 92

AVE Anterior Visceral Endoderm. 4, 5, 32

BMP Bone Morphogenetic Protein. 4, 16, 88, 89

C. elegans Caenorhabditis elegans. 6

CHIR CHIR99021. 19, 23–25, 30, 39, 51, 61, 83, 92

DAPI 4,6-diamidino-2-phenylindole. 83

DBA Dolichos Biflorus Agglutinin. 32, 33, 39, 42, 85

DE Definitive Endoderm. 4–8, 11, 12, 86, 88, 90, 93

DMSO dimethyl sulfoxide. 19

DV Dorso-Ventral. 17

E Embryonic day. 4, 15, 16, 60, 61, 72, 83, 88

ECM Extracellular Matrix. 10, 20, 84, 91

EMT Epithelial-to-Mesenchymal Transition. 5, 6, 9–11, 18, 20, 37–39, 86, 90, 91

EPI Epiblast. 4–12, 18, 19, 24, 26, 30, 35, 37, 39, 48, 85, 86, 90

ESC Embryonic Stem Cells. 18–20, 23, 25, 33, 35, 41, 42

FC Flow Analysis Cytometry. 51

FGF Fibroblast Growth Factor. 4, 16, 17, 82, 89

HCR Hybridisation Chain Reaction. 70, 88

JAK/STAT Janus kinase/Signal Transducer and Activator of Transcription. 17

LR Left-Right. 17

MET Mesenchymal-to-Epithelial Transition. 9, 84, 90

PGC Primordial Germ Cells. 60, 61

PrE Primitive Endoderm. 4, 18

PS Primitive Streak. 5, 7, 8, 11, 15, 18, 20, 24, 26, 32, 60

PSC Pluripotent Stem Cells. 18, 20, 21

ROI Region of Interest. 28, 43, 76

scRNA-seq single-cell RNA sequencing. 63

TF Transcription Factor. 7, 9, 10, 17, 67

TGFβ Transforming growth factor beta. 88

UMAP Uniform Manifold Approximation and Projection. 64, 67

VE Visceral Endoderm. 4–9, 11–13, 32, 33, 35, 37, 39, 59, 84, 86

VFM Ventral Folding Morphogenesis. 12, 13

Wnt Wingless/int1. 4, 8, 9, 17, 20, 23, 24, 32, 88, 89

Appendix A

A.1 Tables

Tab. A.1.: Top 25 markers of "early endoderm" cluster 13. **avg_logFC:** average log fold change of the gene between the expression in the cluster and the expression within the rest of the dataset. **pct.1:** proportions of cells expressing the gene in the cluster. **pct.2:** proportion of cells expressing the gene in the rest of the dataset. **p_val_adj:** adjusted p-value, based on Bonferroni correction using all features in the dataset. **Gene:** name of the gene under consideration. **Description:** full gene name. **Delta:** difference between pct.1 and pct.2 (here used as a ranking criterion)

avg_logFC	pct.1	pct.2	p_val_adj	Gene	Description	Delta
1.11973391	0.952	0.161	0	Otx2	orthodenticle homeobox 2	0.791
0.940700012	0.804	0.09	0	Sox17	SRY (sex determining region Y)-box 17	0.714
0.797630264	0.869	0.195	0	Hhex	hematopoietically expressed homeobox	0.674
0.979970122	0.871	0.207	0	Car4	carbonic anhydrase 4	0.664
0.672261308	0.754	0.097	0	Fzd5	frizzled class receptor 5	0.657
0.937266348	0.864	0.212	0	Cyp26a1	cytochrome P450, family 26	0.652
0.472286654	0.823	0.184	0	Gata6	GATA binding protein 6	0.639
0.68337533	0.701	0.078	0	Gsc	goosecoid homeobox	0.623
0.448296177	0.818	0.197	0	Slc39a8	solute carrier family 39 (metal ion transporter)	0.621
0.754273624	0.822	0.206	0	Ctsc	cathepsin C	0.616
0.594795617	0.913	0.299	0	Foxa2	forkhead box A2	0.614
0.811482781	0.738	0.127	0	Flt1	FMS-like tyrosine kinase 1	0.611
2.184835634	0.974	0.364	0	Trh	thyrotropin releasing hormone	0.61
2.115945621	0.701	0.092	0	Spink1	serine peptidase inhibitor, Kazal type 1	0.609
0.547692147	0.776	0.172	0	Nptx2	neuronal pentraxin 2	0.604
0.54428926	0.851	0.247	0	Podxl	podocalyxin-like	0.604
0.882625042	0.862	0.27	0	Lhx1	LIM homeobox protein 1	0.592
0.3637332	0.729	0.141	0	Atp8a1	ATPase, aminophospholipid transporter (APLT)	0.588
0.730160593	0.743	0.164	0	Fgf5	fibroblast growth factor 5	0.579
0.353270453	0.703	0.125	0	Chst15	carbohydrate sulfotransferase 15	0.578
0.922676996	0.869	0.299	0	Cldn6	claudin 6	0.57
0.386290921	0.76	0.192	0	Dgkk	diacylglycerol kinase kappa	0.568
0.526010411	0.728	0.165	0	Cfc1	cripto, FRL-1, cryptic family 1	0.563
0.468991713	0.632	0.08	0	Eomes	eomesodermin	0.552
0.306809924	0.689	0.156	0	Slc39a4	solute carrier family 39 (zinc transporter)	0.533
0.863432261	0.757	0.227	0	Gpx2	glutathione peroxidase 2	0.53

Tab. A.2.: Top 25 markers of "mature endoderm" cluster 4. **avg_logFC:** average log fold change of the gene between the expression in the cluster and the expression within the rest of the dataset. **pct.1:** proportions of cells expressing the gene in the cluster. **pct.2:** proportion of cells expressing the gene in the rest of the dataset. **p_val_adj:** adjusted p-value, based on Bonferroni correction using all features in the dataset. **Gene:** name of the gene under consideration. **Description:** full gene name. **Delta:** difference between pct.1 and pct.2 (here used as a ranking criterion)

avg_logFC	pct.1	pct.2	p_val_adj	Gene	Description	Delta
1.032107387	0.968	0.24	0	Cldn7	claudin 7	0.728
1.114100133	0.986	0.277	0	Cldn6	claudin 6	0.709
0.521786874	0.818	0.144	0	Crb3	crumbs family member 3	0.674
0.819381405	0.778	0.109	0	Krt7	keratin 7	0.669
0.482522017	0.792	0.134	0	Cldn4	claudin 4	0.658
1.33187761	0.845	0.208	0	Gpx2	glutathione peroxidase 2	0.637
0.512880481	0.848	0.224	0	Rab25	RAB25, member RAS oncogene family	0.624
1.794684935	0.691	0.078	0	Spink1	serine peptidase inhibitor, Kazal type 1	0.613
0.405902606	0.757	0.176	0	Rab15	RAB15, member RAS oncogene family	0.581
0.73017019	0.772	0.197	0	Car4	carbonic anhydrase 4 [	0.575
0.765017328	0.913	0.339	0	Espn	espin	0.574
0.826122719	0.756	0.184	0	Krt19	keratin 19	0.572
0.608862689	0.941	0.369	0	Cdh1	cadherin 1	0.572
0.584080161	0.902	0.339	0	Cmtm8	CKLF-like MARVEL 8	0.563
1.308183109	0.999	0.437	0	Epcam	epithelial cell adhesion molecule	0.562
0.464828782	0.88	0.344	0	Stard10	START domain containing 10	0.536
0.439110706	0.768	0.238	0	Podxl	podocalyxin-like	0.53
0.300189385	0.67	0.142	0	Prss8	protease, serine 8 (prostasin)	0.528
1.871674102	0.999	0.48	0	Krt18	keratin 18	0.519
0.405808769	0.633	0.117	0	Tmprss2	transmembrane protease, serine 2	0.516
0.56602433	0.59	0.081	0	Cpn1	carboxypeptidase N, polypeptide 1	0.509
0.623333157	0.618	0.11	0	Cpm	carboxypeptidase M	0.508
0.419029216	0.858	0.353	0	Spint1	serine protease inhibitor, Kunitz type 1	0.505
0.37020389	0.692	0.19	0	Slc39a8	solute carrier family 39 (metal ion transporter)	0.502
0.301659167	0.662	0.181	0	Fermt1	fermitin family member 1	0.481

Tab. A.3.: Top 25 markers differentially expressed between "early" (13) and "mature" (4) endoderm. Cluster: cluster whose most differentially expressed markers are being listed. **avg_logFC:** average log fold change of the gene between the expression in the cluster and the expression within the rest of the dataset. **pct.1:** proportions of cells expressing the gene in the first cluster. **pct.2:** proportion of cells expressing the gene in the second cluster. **p_val_adj:** adjusted p-value, based on Bonferroni correction using all features in the dataset. **Gene:** name of the gene under consideration. **Delta:** difference between pct.1 and pct.2 (here used as a ranking criterion)

Cluster	avg_logFC	pct.1	pct.2	p_val_adj	Gene	Delta
13 >4	12.00479255	0.206	0.902	0	Sfrp1	0.696
13 >4	7.673819133	0.219	0.879	3.59E-288	Lhx1	0.66
13 >4	7.291439226	0.171	0.775	9.96E-193	Chga	0.604
13 >4	7.830649151	0.13	0.71	4.02E-195	Ntm	0.58
13 >4	3.775721422	0.225	0.804	2.32E-228	Tnni3	0.579
13 >4	7.986047378	0.236	0.814	8.48E-224	Fgf5	0.578
13 >4	5.644605462	0.158	0.734	2.57E-174	Gata4	0.576
13 >4	2.88835479	0.222	0.795	6.57E-204	Sgk1	0.573
13 >4	7.599083553	0.308	0.876	4.08E-257	Hesx1	0.568
13 >4	3.117354286	0.297	0.858	1.54E-231	Cntnap2	-0.561
13 >4	9.819981057	0.25	0.802	5.24E-200	Gpc4	0.552
13 >4	9.85825176	0.144	0.695	1.28E-164	Kcnj5	0.551
13 >4	3.866611269	0.061	0.611	6.90E-123	Pilrb1	0.55
13 >4	11.73820458	0.076	0.625	6.62E-118	Pilra	0.549
13 >4	5.359637576	0.151	0.696	1.75E-147	Eomes	0.545
13 >4	7.120114343	0.332	0.87	5.21E-214	Col9a3	0.538
13 >4	6.206813349	0.199	0.729	4.35E-153	Cer1	0.53
13 >4	2.18599745	0.1	0.63	1.16E-122	Pilrb2	0.53
13 >4	3.309755276	0.389	0.916	2.43E-232	Cyp26a1	0.527
13 >4	2.742289532	0.265	0.791	2.68E-198	Rax	0.526
13 >4	5.943210565	0.356	0.879	3.14E-241	Mum1l1	0.523
13 >4	6.183079916	0.231	0.754	4.87E-158	Prss12	0.523
13 >4	3.926880117	0.188	0.704	2.83E-175	Myl7	0.516
13 >4	6.78484049	0.296	0.798	1.52E-169	Slc7a7	0.502
13 >4	7.862651134	0.131	0.626	8.85E-120	Bmper	0.495
4 >13	4.525065139	0.764	0.192	3.21E-236	Igfbp5	0.572
4 >13	4.995097459	0.796	0.331	2.97E-159	Frem2	0.465
4 >13	6.919877288	0.741	0.276	8.08E-131	Dtna	0.465
4 >13	6.804050967	0.816	0.352	1.31E-189	Krt19	0.464
4 >13	7.021859436	0.575	0.115	6.43E-97	Cldn11	0.46
4 >13	8.474708211	0.526	0.077	3.07E-154	Sox2	0.449
4 >13	9.11899314	0.776	0.331	9.30E-164	Homer2	0.445
4 >13	5.586957334	0.907	0.497	3.32E-196	Cldn4	0.41
4 >13	0.661763993	0.693	0.293	2.28E-90	Cldn9	0.4
4 >13	2.828425897	0.528	0.129	1.84E-102	Dlx3	0.399
4 >13	3.724003768	0.656	0.257	5.13E-96	Prss22	0.399
4 >13	13.01909209	0.427	0.039	1.48E-109	Igf1	0.388

4 >13	4.429582174	0.643	0.257	6.61E-88	Id4	0.386
4 >13	3.610619009	0.792	0.417	9.57E-145	Pax9	0.375
4 >13	5.305923963	0.554	0.183	9.38E-132	Irx2	0.371
4 >13	5.176651422	0.842	0.475	1.38E-147	Uchl1	0.367
4 >13	5.110825887	0.583	0.218	4.76E-82	Efna1	0.365
4 >13	8.872967293	0.412	0.047	5.67E-65	AI646519	0.365
4 >13	4.539226159	0.587	0.223	2.21E-94	Smoc2	0.364
4 >13	1.063295512	0.465	0.109	4.97E-97	Dkk2	0.356
4 >13	4.226766167	0.538	0.184	2.22E-120	Foxp1	0.354
4 >13	3.487894224	0.939	0.586	1.92E-158	Espn	0.353
4 >13	1.634367043	0.77	0.42	1.32E-118	Ppp2r2b	0.35
4 >13	5.584797134	0.758	0.41	9.96E-109	Sdc4	0.348
4 >13	6.937313901	0.589	0.242	1.90E-79	Hmgn3	0.347

Tab. A.4.: Top 11 markers of each of the 22 clusters identified in the reprocessed gastruloid scRNAseq dataset from Rossi et al., 2021a. Cluster: number of the cluster as labelled in Fig. 4.2. **avg_logFC:** average log fold change of the gene between the expression in the cluster and the expression within the rest of the dataset. **pct.1:** proportions of cells expressing the gene in the cluster. **pct.2:** proportion of cells expressing the gene in the rest of the dataset. **p_val_adj:** adjusted p-value, based on Bonferroni correction using all features in the dataset. **Gene:** name of the gene under consideration. **Delta:** difference between pct.1 and pct.2 (here used as a ranking criterion)

Cluster #	avg_logFC	pct.1	pct.2	p_val_adj	Gene	Delta
0	0.780211353	0.804	0.266	0	Meox1	0.538
0	0.543179411	0.766	0.271	0	Col26a1	0.495
0	0.556769706	0.794	0.346	0	Ebf1	0.448
0	0.429952605	0.565	0.137	0	Ebf2	0.428
0	0.605232483	0.658	0.248	0	Cxcl12	0.41
0	0.409830936	0.653	0.262	0	Cdh11	0.391
0	0.557978049	0.636	0.256	0	Nr2f1	0.38
0	0.368174188	0.78	0.418	0	Foxc1	0.362
0	0.420384547	0.759	0.399	0	Nr2f2	0.36
0	0.528196145	0.846	0.487	0	Twist1	0.359
0	0.452580034	0.809	0.452	0	Prrx2	0.357
1	1.500633721	0.998	0.393	0	T	0.605
1	0.577730297	0.701	0.12	0	Cobl	0.581
1	1.195519379	0.846	0.269	0	Cthrc1	0.577
1	0.409111314	0.666	0.101	0	Mnx1	0.565
1	1.155788901	0.604	0.059	0	Shh	0.545
1	0.676712218	0.874	0.344	0	Cdx2	0.53
1	0.599299736	0.644	0.117	0	Atp1a2	0.527
1	0.516175348	0.655	0.13	0	Samd3	0.525
1	0.5074068	0.806	0.282	0	Foxa2	0.524
1	1.076555	0.647	0.128	0	Defa30	0.519
1	0.516872566	0.52	0.014	0	Noto	0.506
2	0.763517155	0.929	0.206	0	Hoxb9	0.723
2	0.756391817	0.964	0.282	0	Hoxb8	0.682
2	0.907660463	0.768	0.118	0	Hoxc8	0.65
2	0.56065456	0.778	0.141	0	Epha5	0.637
2	0.818377169	0.812	0.196	0	Hoxa7	0.616
2	0.567778001	0.822	0.221	0	Hoxb5os	0.601
2	0.619557467	0.653	0.083	0	Hoxc9	0.57
2	0.383963076	0.676	0.123	0	Hoxc6	0.553
2	1.177583298	0.947	0.405	0	Hoxaas3	0.542
2	0.275873283	0.752	0.213	0	Hoxb7	0.539
2	0.630581031	0.865	0.35	0	Cdx2	0.515
3	1.00027499	0.905	0.161	0	Gata6	0.744
3	0.795801291	0.695	0.028	0	Hand2	0.667
3	2.102460724	0.766	0.108	0	Myl7	0.658

3	0.800064627	0.802	0.16	0	Capn6	0.642
3	0.717579304	0.652	0.112	0	Asb4	0.54
3	0.308176855	0.571	0.043	0	Gata4	0.528
3	0.696457865	0.66	0.14	0	Mef2c	0.52
3	1.523091147	0.594	0.079	0	Tnni1	0.515
3	0.584834035	0.63	0.13	0	Runx1t1	0.5
3	0.308119842	0.618	0.187	0	Pcdh7	0.431
3	0.430023055	0.705	0.275	0	Lhfp	0.43
4	1.032107387	0.968	0.24	0	Cldn7	0.728
4	1.114100133	0.986	0.277	0	Cldn6	0.709
4	0.521786874	0.818	0.144	0	Crb3	0.674
4	0.819381405	0.778	0.109	0	Krt7	0.669
4	0.482522017	0.792	0.134	0	Cldn4	0.658
4	1.33187761	0.845	0.208	0	Gpx2	0.637
4	0.512880481	0.848	0.224	0	Rab25	0.624
4	1.794684935	0.691	0.078	0	Spink1	0.613
4	0.405902606	0.757	0.176	0	Rab15	0.581
4	0.73017019	0.772	0.197	0	Car4	0.575
4	0.765017328	0.913	0.339	0	Espn	0.574
5	0.400487154	0.699	0.128	0	Gjb3	0.571
5	0.688736371	0.986	0.439	0	Epcam	0.547
5	0.386885891	0.686	0.154	0	Tfap2c	0.532
5	0.27862073	0.729	0.224	0	Esrp1	0.505
5	0.382512381	0.755	0.255	0	Cldn7	0.5
5	0.281878933	0.867	0.375	0	Cdh1	0.492
5	0.355165106	0.716	0.234	0	Rab25	0.482
5	0.390542271	0.774	0.293	0	Cldn6	0.481
5	0.589096937	0.947	0.485	0	Krt18	0.462
5	0.25461672	0.676	0.226	0	Mapk13	0.45
5	0.327276602	0.592	0.149	0	Cldn4	0.443
6	1.168962928	0.987	0.233	0	Pim2	0.754
6	0.996666936	0.898	0.15	0	Utf1	0.748
6	0.964207063	0.891	0.177	0	L1td1	0.714
6	1.196945537	0.992	0.346	0	Pou5f1	0.646
6	0.632374187	0.826	0.185	0	Pou3f1	0.641
6	0.331571176	0.726	0.09	0	Pdzd4	0.636
6	0.547297887	0.938	0.314	0	Gng3	0.624
6	0.671668868	0.907	0.305	0	Rasgrp2	0.602
6	0.441414821	0.818	0.218	0	Slc7a3	0.6
6	0.675352365	0.979	0.388	0	Bst2	0.591
6	0.404534715	0.912	0.333	0	Tmem238	0.579
7	0.308620971	0.688	0.277	1.20E-303	Sox3	0.411
7	0.250147154	0.7	0.328	4.01E-205	Stmn2	0.372
7	0.295739958	0.798	0.439	2.89E-178	Sox2	0.359
7	0.310579148	0.571	0.222	8.71E-255	Fgfbp3	0.349

7	0.569321648	0.781	0.449	1.27E-207	Fst	0.332
7	0.273367504	0.528	0.207	6.64E-213	Wnt8a	0.321
7	0.273447252	0.569	0.259	1.08E-162	Pim2	0.31
7	0.490464562	0.656	0.367	2.43E-130	Pou5f1	0.289
7	0.879014362	0.925	0.654	8.90E-269	Crabp2	0.271
7	0.28349457	0.692	0.44	5.35E-108	Hoxb1	0.252
7	0.254230025	0.482	0.231	1.94E-127	Nkx1-2	0.251
8	1.196139139	0.856	0.208	0	Aldh1a2	0.648
8	0.623841327	0.853	0.233	0	Foxc2	0.62
8	0.809154041	0.76	0.171	0	Pcdh19	0.589
8	0.732173323	0.906	0.324	0	Meox1	0.582
8	0.573079702	0.96	0.452	0	Foxc1	0.508
8	0.396869136	0.588	0.128	0	Pax3	0.46
8	0.479787116	0.943	0.487	0	Prrx2	0.456
8	1.315049716	0.875	0.432	0	Tcf15	0.443
8	0.563226908	0.878	0.438	0	Cited1	0.44
8	0.329918592	0.501	0.073	0	Fgf18	0.428
8	0.392201224	0.863	0.456	0	Tenm4	0.407
9	0.830852769	0.957	0.151	0	Tbx6	0.806
9	0.864619848	0.975	0.285	0	Dll1	0.69
9	0.823631683	0.956	0.278	0	Hes7	0.678
9	0.833750364	0.843	0.167	0	Pcdh19	0.676
9	0.772826114	0.878	0.213	0	Gm45889	0.665
9	0.434229078	0.667	0.056	0	2610528A11Rik	0.611
9	0.867821367	0.872	0.268	0	Lmo2	0.604
9	0.43076805	0.757	0.168	0	Rnf213	0.589
9	0.707758913	0.97	0.382	0	Dll3	0.588
9	1.449318601	0.985	0.432	0	Cited1	0.553
9	0.329436324	0.769	0.239	0	Rftn1	0.53
10	1.015790154	0.909	0.264	0	Cdkn1c	0.645
10	0.670352964	0.728	0.218	0	Grb10	0.51
10	0.254548486	0.625	0.292	1.34E-186	Eif2s3y	0.333
10	0.828976697	0.512	0.208	1.04E-225	Dlk1	0.304
10	0.31818709	0.835	0.619	1.34E-224	H13	0.216
10	0.36274132	0.544	0.342	8.71E-88	Col26a1	0.202
10	0.300861516	0.536	0.344	1.09E-60	Meox1	0.192
10	0.429491116	0.543	0.369	2.75E-79	Igfbp3	0.174
10	0.304110562	0.505	0.347	2.55E-68	Tshz2	0.158
10	0.409077016	0.27	0.129	2.34E-53	Sprr2a3	0.141
10	0.372451155	0.639	0.503	7.25E-49	Prrx2	0.136
11	0.741574373	0.812	0.215	0	Prrx1	0.597
11	0.526772852	0.873	0.326	0	Col26a1	0.547
11	1.171219261	0.974	0.487	0	Prrx2	0.487
11	0.447280833	0.787	0.304	0	Cdh11	0.483
11	0.543858858	0.867	0.398	0	Ebf1	0.469

11	0.353343752	0.68	0.262	0	Mfap4	0.418
11	0.435276255	0.806	0.393	0	Serpinf1	0.413
11	0.296719783	0.647	0.242	0	Pdgfrb	0.405
11	0.524708592	0.749	0.345	0	Mfap2	0.404
11	0.265226411	0.489	0.093	0	Cped1	0.396
11	0.329430248	0.585	0.189	1.76E-291	Ebf2	0.396
12	0.772440284	0.899	0.296	0	Foxa2	0.603
12	0.631364071	0.612	0.102	0	Nkx6-1	0.51
12	0.465019198	0.673	0.169	0	Sox21	0.504
12	0.628251679	0.924	0.437	0	Sox2	0.487
12	0.412126373	0.513	0.076	0	Rfx4	0.437
12	0.374377597	0.803	0.369	2.89E-259	Apela	0.434
12	0.71522566	0.707	0.292	2.12E-302	Calca	0.415
12	0.256957966	0.503	0.112	0	Adgrv1	0.391
12	0.366972674	0.724	0.358	9.18E-212	Cmtm8	0.366
12	0.355506671	0.756	0.391	1.53E-242	Kif21a	0.365
12	0.499557618	0.58	0.216	1.42E-243	Car4	0.364
13	1.11973391	0.952	0.161	0	Otx2	0.791
13	0.940700012	0.804	0.09	0	Sox17	0.714
13	0.797630264	0.869	0.195	0	Hhex	0.674
13	0.979970122	0.871	0.207	0	Car4	0.664
13	0.672261308	0.754	0.097	0	Fzd5	0.657
13	0.937266348	0.864	0.212	0	Cyp26a1	0.652
13	0.472286654	0.823	0.184	0	Gata6	0.639
13	0.68337533	0.701	0.078	0	Gsc	0.623
13	0.448296177	0.818	0.197	0	Slc39a8	0.621
13	0.754273624	0.822	0.206	0	Ctsc	0.616
13	0.594795617	0.913	0.299	0	Foxa2	0.614
14	0.731100931	0.825	0.255	0	Pim2	0.57
14	0.398323485	0.713	0.173	0	Utf1	0.54
14	0.261620999	0.814	0.275	5.55E-293	Cdkn1c	0.539
14	1.333618649	0.887	0.364	0	Pou5f1	0.523
14	0.505608141	0.849	0.331	0	Gng3	0.518
14	0.304849862	0.793	0.281	4.91E-297	Sox3	0.512
14	0.462975374	0.83	0.33	1.26E-300	Stmn2	0.5
14	0.558201668	0.69	0.2	0	L1td1	0.49
14	0.438805149	0.684	0.207	0	Wnt8a	0.477
14	0.345965698	0.879	0.404	2.03E-232	Bst2	0.475
14	0.293663554	0.784	0.318	5.02E-253	H2-D1	0.466
15	1.707180062	0.978	0.102	0	Kdr	0.876
15	1.122577723	0.928	0.053	0	Rasip1	0.875
15	1.190559331	0.907	0.035	0	Sox18	0.872
15	1.063328269	0.869	0.016	0	Esam	0.853
15	1.456022232	0.851	0.011	0	Cdh5	0.84
15	0.884632395	0.922	0.092	0	Rhoj	0.83

15	2.211296278	0.86	0.041	0	Cldn5	0.819
15	1.325026019	0.939	0.124	0	Plvap	0.815
15	0.942090782	0.894	0.087	0	Fli1	0.807
15	0.598851008	0.841	0.036	0	She	0.805
15	0.667953997	0.814	0.012	0	Tek	0.802
16	1.118837534	0.896	0.147	0	Hoxc8	0.749
16	0.654954231	0.886	0.17	0	Epha5	0.716
16	0.766511993	0.949	0.241	0	Hoxb9	0.708
16	0.590021702	0.778	0.089	0	Cdx4	0.689
16	0.901617096	0.891	0.224	0	Hoxa7	0.667
16	0.790667678	0.907	0.249	0	Hoxb5os	0.658
16	0.712219734	0.968	0.315	0	Hoxb8	0.653
16	0.414720271	0.784	0.147	0	Hoxc6	0.637
16	0.701691866	0.742	0.109	0	Hoxc9	0.633
16	0.365156069	0.866	0.233	0	Hoxc4	0.633
16	0.730716174	0.832	0.283	5.07E-220	Cdkn1c	0.549
17	0.662926903	0.865	0.286	4.44E-174	Cdkn1c	0.579
17	0.463324143	0.813	0.345	5.03E-108	Col26a1	0.468
17	0.510109147	0.761	0.32	1.65E-105	Cdh11	0.441
17	0.813913514	0.645	0.216	1.73E-123	Dlk1	0.429
17	0.56711667	0.761	0.348	1.00E-82	Meox1	0.413
17	0.481368605	0.827	0.414	3.66E-78	Ebf1	0.413
17	0.40052666	0.879	0.472	1.79E-81	Foxc1	0.407
17	0.329847604	0.664	0.259	4.11E-83	Foxc2	0.405
17	0.576567775	0.868	0.505	3.42E-77	Prrx2	0.363
17	0.48173405	0.674	0.313	3.30E-59	Nr2f1	0.361
17	0.550364562	0.901	0.54	1.85E-89	Twist1	0.361
18	0.757686793	0.92	0.03	0	AC158554.1	0.89
18	0.546469595	0.897	0.022	0	Dppa2	0.875
18	0.838892826	0.973	0.126	0	Dppa4	0.847
18	1.184650848	0.904	0.06	0	Zfp42	0.844
18	0.430834368	0.85	0.021	0	Gdf3	0.829
18	1.148755821	0.9	0.08	0	Tdh	0.82
18	1.666829565	1	0.184	0	Utf1	0.816
18	0.660638812	0.857	0.045	0	Fbxo15	0.812
18	0.471389599	0.854	0.045	0	Gtsf1l	0.809
18	0.60037135	0.87	0.062	0	Tex19.1	0.808
18	0.540709578	0.807	0.01	0	Morc1	0.797
19	0.697221649	0.928	0.407	2.32E-110	Dll3	0.521
19	0.549264468	0.544	0.051	1.24E-289	Srrm4	0.493
19	1.083779257	0.802	0.32	8.15E-103	Rtn1	0.482
19	1.880778904	0.601	0.121	3.89E-147	Tubb3	0.48
19	0.993278427	0.525	0.049	7.36E-278	Elavl3	0.476
19	1.029488869	0.665	0.192	2.87E-105	Isl1	0.473
19	1.388927047	0.646	0.195	3.53E-99	Gadd45g	0.451

19	0.714199661	0.452	0.004	0	Neurod4	0.448
19	1.019610831	0.548	0.115	5.84E-125	Btbd17	0.433
19	0.577988042	0.745	0.316	3.16E-70	Dll1	0.429
19	0.642141503	0.532	0.115	2.53E-117	Plk3	0.417
20	2.093902724	0.995	0.168	2.63E-288	Pcp4	0.827
20	0.912151543	0.834	0.063	0	Pax2	0.771
20	0.593500222	0.796	0.027	0	Pax8	0.769
20	0.461741304	0.664	0.098	3.39E-178	Fgf9	0.566
20	1.19143691	0.905	0.373	4.04E-104	Wfdc2	0.532
20	0.370837841	0.72	0.19	5.17E-98	Scube1	0.53
20	0.583086322	0.649	0.127	1.08E-113	Uncx	0.522
20	0.568480404	0.82	0.307	2.92E-86	Arid5b	0.513
20	0.327716411	0.583	0.077	8.89E-168	Emx2	0.506
20	0.370853665	0.682	0.18	2.75E-94	Hs3st3b1	0.502
20	0.3535677	0.512	0.016	0	Pou3f3	0.496
21	0.568502514	0.934	0.86	1.49E-12	Ube2c	0.074
21	0.630337073	0.686	0.619	0.022294745	Crabp1	0.067
21	0.286230128	0.92	0.863	1.03E-05	Cdk1	0.057
21	0.314071353	0.964	0.908	9.49E-12	Birc5	0.056
21	0.446383699	0.818	0.782	3.91E-13	AY036118	0.036
21	0.393201898	0.993	0.965	6.88E-12	Cks2	0.028
21	0.420286176	0.993	0.975	4.01E-12	H2afx	0.018
21	0.38573409	0.985	0.971	3.31E-16	Eif4ebp1	0.014
21	0.338891568	1	0.988	5.09E-19	Bola2	0.012
21	0.311022844	1	0.988	1.12E-15	Lsm3	0.012
21	0.418660035	1	0.992	1.43E-29	Ndufa6	0.008

Tab. A.5.: Top 20 markers of each of the embryonic gut tube clusters highlighted in Fig. 4.4 Cluster: number of the cluster as labelled in Fig. 4.4. **avg_logFC:** average log fold change of the gene between the expression in the cluster and the expression within the rest of the dataset. **pct.1:** proportions of cells expressing the gene in the cluster. **pct.2:** proportion of cells expressing the gene in the rest of the dataset. **p_val_adj:** adjusted p-value, based on Bonferroni correction using all features in the dataset. **Gene:** name of the gene under consideration. **Delta:** difference between pct.1 and pct.2 (here used as a ranking criterion)

Cluster #	avg_logFC	pct.1	pct.2	p_val_adj	Gene	Delta
0	-8.109348173	0.787	0.194	0	2610528A11Rik	0.593
0	-5.006721714	0.827	0.259	0	Map1b	0.568
0	-4.178377867	0.768	0.232	0	Shh	0.536
0	2.108570186	0.655	0.147	0	Fabp1	0.508
0	-0.812178937	0.748	0.247	0	Cxcl12	0.501
0	-4.069909099	0.751	0.295	0	Kitl	0.456
0	-6.032116028	0.724	0.286	0	Rnase4	0.438
0	4.41497132	0.649	0.235	0	Ang	0.414
0	-3.367682135	0.809	0.402	0	Tmsb10	0.407
0	-7.483942937	0.732	0.326	0	Krt19	0.406
0	-1.38288923	0.489	0.085	0	Cdx4	0.404
0	-0.984449897	0.76	0.36	0	Fth1	0.4
0	1.134812795	0.679	0.281	0	Gm21814	0.398
0	-13.65371056	0.717	0.325	0	Emb	0.392
0	0.570685626	0.509	0.122	2.72E-226	Aqp7	0.387
0	-0.734541876	0.764	0.384	0	Tceal8	0.38
0	-3.001256535	0.684	0.306	0	Pura	0.378
0	-6.382226168	0.67	0.298	0	Ihh	0.372
0	2.442824499	0.565	0.201	2.52E-251	Krtap17-1	0.364
0	1.001296601	0.689	0.333	0	Cystm1	0.356
0	1.283476001	0.435	0.085	4.17E-255	Aqp3	0.35
2	4.620978185	0.97	0.109	0	Rspo3	0.861
2	5.787311089	0.938	0.097	0	Tlx2	0.841
2	5.799500753	0.961	0.131	0	Fgf8	0.83
2	1.320306059	0.941	0.123	0	T	0.818
2	4.611793463	0.955	0.14	0	Hoxc9	0.815
2	4.428307187	0.906	0.113	0	Lmo2	0.793
2	3.892004239	0.954	0.161	0	Hoxc8	0.793
2	-3.188951998	0.924	0.149	0	Hoxd9	0.775
2	13.99623227	0.908	0.146	0	Hoxb6	0.762
2	4.647830671	0.916	0.159	0	Hoxa7	0.757
2	3.748607988	0.899	0.167	0	Hoxa9	0.732
2	12.79236079	0.923	0.202	0	Wnt5b	0.721
2	14.97854934	0.903	0.187	0	Hoxb8	0.716
2	16.32216193	0.905	0.201	0	Sp5	0.704
2	8.172756465	0.901	0.198	0	Nkd1	0.703
2	4.810393383	0.825	0.123	0	Hoxc6	0.702
2	-1.754517684	0.874	0.173	0	Hoxb9	0.701

2	2.509135577	0.96	0.27	0	Apela	0.69
2	-1.586261339	0.895	0.207	0	Xlr3a	0.688
2	1.995215777	0.86	0.179	0	Trap1a	0.681
2	6.211544721	0.76	0.085	0	Hes7	0.675
4	5.816817711	0.973	0.12	0	Hhex	0.853
4	22.31934201	0.793	0.056	0	Reln	0.737
4	2.995923678	0.956	0.221	0	Ttr	0.735
4	6.71335635	0.978	0.258	0	Cdkn1c	0.72
4	13.70302769	0.946	0.251	0	Dlk1	0.695
4	9.063207339	0.818	0.137	0	Prox1	0.681
4	3.603857761	0.952	0.322	0	Ddt	0.63
4	2.846236713	0.859	0.232	0	Apoe	0.627
4	2.508081503	0.904	0.291	0	Cald1	0.613
4	-12.03271069	0.882	0.271	0	Vtn	0.611
4	-0.379826888	0.731	0.16	8.17E-303	Tgm2	0.571
4	11.27991932	0.731	0.162	1.26E-291	Asb4	0.569
4	4.872612806	0.658	0.111	3.73E-229	Fgb	0.547
4	6.561506168	0.742	0.202	2.28E-281	Maob	0.54
4	5.19407635	0.943	0.409	0	Bex4	0.534
4	-1.041621384	0.667	0.137	5.01E-237	Col1a2	0.53
4	3.322616646	0.877	0.352	0	Psat1	0.525
4	11.73871437	0.652	0.127	1.26E-224	Serpina1c	0.525
4	7.59791777	0.726	0.227	1.11E-265	Rasgrp3	0.499
4	4.415519437	0.689	0.196	3.85E-236	Lpl	0.493
4	16.22805861	0.706	0.221	3.55E-241	Ambp	0.485
5	2.587866973	0.984	0.219	0	Ttr	0.765
5	3.609446057	0.969	0.205	0	Rbp4	0.764
5	3.968986022	0.995	0.292	0	Spink1	0.703
5	4.863548124	0.865	0.181	0	S100g	0.684
5	3.953482823	0.938	0.267	0	Car4	0.671
5	10.24737723	0.974	0.327	0	Emb	0.647
5	10.02340072	0.921	0.286	0	Lrpap1	0.635
5	5.331227453	0.783	0.166	0	Lgals2	0.617
5	-0.411753929	0.894	0.281	0	Smim14	0.613
5	6.920890417	0.932	0.325	0	Rbbp7	0.607
5	1.371900635	0.853	0.291	0	Ociad2	0.562
5	4.543912133	0.698	0.145	6.16E-294	Tff3	0.553
5	8.32734117	0.667	0.115	6.09E-297	Tfcp2l1	0.552
5	1.116743428	0.745	0.203	0	Cst3	0.542
5	-3.935061253	0.827	0.299	0	Nudt11	0.528
5	4.576234229	0.715	0.193	0	Apoa1	0.522
5	2.320194975	0.777	0.26	0	Tdh	0.517
5	-8.533127349	0.751	0.24	0	Pth1r	0.511
5	14.22695964	0.782	0.272	0	Dpp4	0.51
5	2.221026336	0.904	0.399	0	Krt18	0.505

5	7.325965393	0.77	0.265	0	Nrp2	0.505
9	6.584008903	0.947	0.139	0	Nkx2-3	0.808
9	4.871054921	0.888	0.126	0	Nkx2-5	0.762
9	1.823766973	0.926	0.264	0	Igfbp5	0.662
9	9.844838741	0.933	0.277	0	Isl1	0.656
9	3.174040479	0.955	0.324	0	H19	0.631
9	11.62884426	0.664	0.048	0	Nrxn3	0.616
9	11.09563252	0.738	0.157	0	Dok5	0.581
9	2.890448827	0.854	0.279	0	Mest	0.575
9	18.34938404	0.843	0.269	0	Shisa2	0.574
9	2.64157084	0.878	0.326	0	Igf2	0.552
9	1.113240035	0.924	0.373	0	Csrp2	0.551
9	15.10881535	0.841	0.331	0	Capn6	0.51
9	-0.357619887	0.848	0.342	9.45E-283	Tmsb4x	0.506
9	7.990025033	0.776	0.277	0	Tbx3	0.499
9	13.8041693	0.72	0.223	0	Pcdh7	0.497
9	27.94362627	0.526	0.044	2.39E-191	Pnpla3	0.482
9	11.59216647	0.824	0.345	0	Nuak1	0.479
9	0.808577513	0.831	0.353	0	Nnat	0.478
9	5.534045334	0.837	0.369	2.09E-301	Malat1	0.468
9	13.16005394	0.529	0.066	8.41E-220	Ptpre	0.463
9	9.91943634	0.475	0.023	8.53E-162	Nkx2-1	0.452
10	10.83643139	0.827	0.149	0	Otx2	0.678
10	7.396667647	0.696	0.135	8.11E-292	Tnnt1	0.561
10	3.514116809	0.822	0.271	0	Igfbp5	0.551
10	14.38510272	0.81	0.267	5.28E-296	Homer2	0.543
10	14.77542287	0.886	0.358	0	Uchl1	0.528
10	-2.978975671	0.775	0.267	8.62E-267	Socs2	0.508
10	10.8332394	0.84	0.348	0	Ccnd2	0.492
10	6.248061722	0.616	0.147	1.96E-192	Col11a1	0.469
10	3.789228899	0.543	0.09	4.63E-120	Foxg1	0.453
10	7.503638501	0.73	0.28	2.17E-181	Sox2	0.45
10	5.488832809	0.661	0.214	5.35E-194	Ehbp1	0.447
10	3.131009604	0.779	0.334	1.26E-232	H19	0.445
10	1.96446449	0.695	0.253	3.60E-172	Pax9	0.442
10	9.205403703	0.683	0.241	3.01E-217	Sulf1	0.442
10	1.049441485	0.678	0.239	2.41E-174	Six1	0.439
10	-3.613370532	0.646	0.211	5.17E-137	Has2	0.435
10	3.110893688	0.775	0.347	8.86E-241	Tmsb4x	0.428
10	15.60634075	0.503	0.081	4.69E-121	Lypd2	0.422
10	3.490421605	0.576	0.157	2.90E-108	Vgll2	0.419
10	1.56771571	0.75	0.334	5.46E-181	Igf2	0.416
10	-2.812010924	0.721	0.311	1.48E-162	Col2a1	0.41
11	8.732548199	0.88	0.263	0	Socs2	0.617
11	5.033534444	0.684	0.123	3.63E-206	Pax1	0.561
11	4.620484036	0.786	0.234	2.61E-265	Six1	0.552

11	11.16455862	0.862	0.318	0	Mmd	0.544
11	6.143585815	0.748	0.251	1.45E-195	Pax9	0.497
11	-2.595727726	0.684	0.209	4.57E-156	Has2	0.475
11	0.829639468	0.805	0.355	1.25E-211	Npm3	0.45
11	-2.371498682	0.755	0.33	7.89E-200	Crabp2	0.425
11	-1.276294433	0.584	0.16	1.02E-73	Nkx2-3	0.424
11	2.290420655	0.526	0.11	1.39E-78	Cldn10	0.416
11	6.486929129	0.633	0.218	1.11E-148	Jag1	0.415
11	4.716677093	0.76	0.348	3.73E-238	Tmsb4x	0.412
11	5.438447226	0.51	0.104	3.72E-66	Fgf3	0.406
11	-5.18626846	0.741	0.342	8.06E-156	Meis2	0.399
11	5.415783196	0.702	0.305	1.44E-142	Efnb2	0.397
11	0.971492324	0.851	0.463	3.59E-179	Rpl8	0.388
11	3.978746195	0.649	0.262	2.44E-105	Irx3	0.387
11	-3.648753649	0.6	0.219	6.02E-103	Eya1	0.381
11	5.601559625	0.731	0.369	1.46E-149	Adgrl2	0.362
11	1.056958541	0.857	0.502	3.53E-153	Rps8	0.355
11	4.340731192	0.427	0.074	2.23E-40	Tmem100	0.353

A.2 Reagent details

Tab. A.6.: Details of the antibodies used in this work. The concentration of the Cdx2 antibody was not disclosed by the supplier.

Target	Host	Clonality	Supplier	CAT#	RRID	Concentration
FoxA2	Rabbit	monoclonal (EPR4466)	Abcam	ab108422	AB_11157157	1:500 (4ug/mL)
TBra	Rabbit	monoclonal (EPR18113)	Abcam	ab209665	AB_2750925	1:500 (0.8ug/mL)
TBra	Goat	polyclonal	R&D Systems	AF2085	AB_2200235	1:250 (0.8ug/mL)
Sox17	Goat	polyclonal	R&D Systems	AF1924	AB_355060	1:250 (0.8ug/mL)
Cdh1	Rat	monoclonal (DECMA-1)	Abcam	ab11512	AB_298118	1:500 (2ug/mL)
Cdh1	Rabbit	monoclonal (24E10)	CST	3195	AB_2291471	1:500 (0.1ug/mL)
Sox2	Rabbit	polyclonal	Abcam	ab97959	AB_2341193	1:500 (2ug/mL)
Sox2	Rat	monoclonal (Btjce)	eBioscience	14-9811-82	AB_11219471	1:500 (1ug/mL)
Dppa4	Goat	polyclonal	R&D Systems	AF3730	AB_2094166	1:250 (0.8ug/mL)
Snai1	Goat	polyclonal	R&D Systems	AF3639	AB_2191738	1:250 (0.8ug/mL)
FoxC1	Rabbit	monoclonal (EPR20678)	Abcam	ab223850	AB_2848202	1:500 (1ug/mL)
GM130	Mouse	monoclonal (Clone 35)	BD Laboratories	610822	AB_398141	1:500 (0.5ug/mL)
Lama1	Rabbit	polyclonal	Sigma Aldrich	L9393	AB_477163	1:500 (1ug/mL)
Sox1	Goat	polyclonal	R&D Systems	AF3369	AB_2239879	1:250 (0.8ug/mL)
Cdx2	Rabbit	monoclonal (EPR2764Y)	Thermo Fisher	MA5-14494	AB_11004783	1:250 (N.A.*)
EpCAM	Rat	monoclonal (G8.8)	eBioscience	17-5791-82	AB_1659714	1:250 (0.8ug/mL)

Tab. A.7.: Sequences of the Pax9 HCR probe sets used (coupled to amplifier B5). For each probe, the universal sequence that binds B5 hairpins is indicated in lowercase, while the sequence targeting Pax9 mRNA is indicated in uppercase.

Sequence	Label
ctcactcccaatctctataaTAATTGGTGCAGGCCGGTGAGTGAC	Pax9 B5 Odd_01
AGGTGGGCCCGCGAGCAATCTGCGTaactaccctacaaatccaat	Pax9 B5 Even_01
ctcactcccaatctctataaTCACAAGGTCGGATGCCCAGTTGGG	Pax9 B5 Odd_02
TGCGAGACCCGTAGCTGGCGGCTGAaactaccctacaaatccaat	Pax9 B5 Even_02
ctcactcccaatctctataaACGGAGTGAGAGGAGGGCCAGGTGC	Pax9 B5 Odd_03
ATGGAGCGGATGCCCAGAATGTCCGaactaccctacaaatccaat	Pax9 B5 Even_03
ctcactcccaatctctataaCCTGGCTGCCTGCATTCCCTTGAAA	Pax9 B5 Odd_04
AGAAGCGGTCACAGAATGGCTGCCTaactaccctacaaatccaat	Pax9 B5 Even_04
ctcactcccaatctctataaCAACACAAATGCCTCATTCCACCTT	Pax9 B5 Odd_05
CTTCTCTGTGTTAAAACAGTGTGCGaactaccctacaaatccaat	Pax9 B5 Even_05
ctcactcccaatctctataaGAATACACACGACCTGGGGCAAGTC	Pax9 B5 Odd_06
CCTTCATGAGATGCTGAAGCCTCTAaactaccctacaaatccaat	Pax9 B5 Even_06
ctcactcccaatctctataaGAAACATGCTTCAGTACGCACTGTA	Pax9 B5 Odd_07
GATGCCAACACCAATTGACCTACTCaactaccctacaaatccaat	Pax9 B5 Even_07
ctcactcccaatctctataaAGCCATGACTACACAGAGAAAGCCT	Pax9 B5 Odd_08
CCTTGTCTACAGAGTGAGTTTCAGGaactaccctacaaatccaat	Pax9 B5 Even_08
ctcactcccaatctctataaGTCCCCACCAAAATATTGGATTGGA	Pax9 B5 Odd_09
GCCTTCAGTCAGATTCTCTGGATAGaactaccctacaaatccaat	Pax9 B5 Even_09
ctcactcccaatctctataaCTGGAGGAGGCTGGTGTTAACAGTA	Pax9 B5 Odd_10

A.3 Other publications

Note: The following work has already been published, and under CC-BY 4.0 International license (Vianello, 2020). I (Stefano Vianello) hold the rights of reproduction and as such am including it as an appendix to the work presented in this thesis. The licensing terms that apply to the thesis (see colophon) are not to be considered to overwrite those of the publication that is reproduced here, which remains under CC-BY 4.0 International license

Exploring and illustrating the mouse embryo: virtual objects to think and create with

Stefano Vianello[1,✉]

[1]Laboratory of Stem Cell Bioengineering, Institute of Bioengineering, School of Life Sciences (SV) and School of Engineering (STI), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

The teaching, learning, communication, and practice of Developmental Biology require interested parties to be at ease with the considerable spatial complexity of the embryo, and with its evolution over time as it undergoes morphogenesis. In practice, the four dimensionality of embryonic development (space and time) calls upon strong visual-spatial literacy and mental manipulation skills, generally expected to be innate or to come through experience. Yet it has been argued that Developmental Biology suffers the most from available traditional media of communication and representation. To date, few resources exist to engage with the embryo in its 3D and 4D aspects, to communicate such aspects in one's work, and to facilitate their exploration in the absence of live observations. I here provide a collection of readily-usable volumetric models for all tissues and stages of mouse peri-implantation development as extracted from the eMouse Atlas Project (E5.0 to E9.0), as well as custom-made models of all pre-implantation stages (E0 to E4.0). These models have been converted to a commonly used 3D format (.stl), and are provided in ready-made files for digital exploration and illustration. Further provided is a step-by-step walkthrough on how to practically use these models for exploration and illustration using the free and open source 3D creation suite Blender. I finally outline possible further uses of these very models in outreach initiatives of varying levels, virtual and augmented reality applications, and 3D printing.

EMAP | 3D models | mouse embryo | Developmental Biology
Correspondence: ***stefano.vianello@epfl.ch***

Spatial thinking in Developmental Biology: understanding and communicating four dimensions.

During mouse embryonic development a single, symmetric, undifferentiated cell transforms into a multicellular, differentiated organism populated by multiple specialised cell types (Pijuan-Sala et al., 2019). As new cells emerge, they migrate across the embryo, they squeeze between other tissues, they break down and deposit basement membranes, they traverse and integrate surrounding tissue layers (cfr. e.g. Arnold & Robertson (2009); Hashimoto & Nakatsuji (1989); Nahaboo & Migeotte (2018); Saykali et al. (2019); Viotti et al. (2014)). The embryo as a whole undergoes dramatic shape changes (McDole et al., 2018; Rivera-Pérez & Hadjantonakis, 2014; Snow, 1977). Students and researchers in mouse Developmental Biology are required to be familiar with such fundamental processes (gastrulation, cell migration out of the primitive streak, formation of the node, endoderm pocketing, gut morphogenesis, notochord deposition...). Yet, this requires them to build complex mental pictures involving simultaneous changes in multiple embryonic tissues, movement of different cell populations in three dimensions, and to entertain such complex mental representations as they evolve over time. Building such pictures is neither immediate nor intuitive, and for the many that do not have direct access to mouse embryos it usually only comes from the prolonged exposure to literature and the uncertain piecing together of a heterogeneous range of visual data formats. Here are then students and early career researchers scouring the literature for good representations of their structure of interest, at the right developmental timepoint, seen from that specific angle they need, with or without such and such other tissue. The outcome is rarely a single picture, and is typically a collage of disparate histological sections, immunostainings, electron microscopy images, diagrams.

Crucially to this point, the mental manipulation of 3D objects, over time (i.e. 4D), actively calls upon cognitive skills whose teaching has long been the focus of psychological and education research; notions of representational competence, spatial visualisation, spatial perception, visual-spatial literacy (see e.g. Milner-Bolotin & Nashon (2011); NRC (2012)). The importance of "the skill for thinking in three dimensions", "for visualising shapes in the mind's eye, rotating, translating, and shearing them", and "imaging complex changes over time in the form of a cinematographic visual image" (Chadwick, 1978) will resonate strongly with most developmental biologists, just as that of so-called "penetrative thinking", i.e. the capability to infer the cross-section of a layered object from surface features alone (Kali & Nir, 1996). While little research has focused on the importance of such skills in Developmental Biology Education (but see references in Hardin (2008)), foundational studies have been done in geoscience (and indeed compare the cross-section of a gastrulating embryo with the multi-layered landscape of earth crust sections). These studies have crucially shown that spatial thinking abilities can be i) taught, and ii) improved by practice (Kali & Nir, 1996; Titus & Horsman, 2009). Whether this transcends geological sciences or not, tools to foster visual-spatial literacy in Developmental Biology still remain scarce and not widely available (Hardin, 2008). It could be argued that the highly dynamic and dimensional nature of Embryology makes this discipline particularly penalised by traditional media of scientific dissemination.

When dealing with developmental data and with the inherent flatness (and stillness) of the printed page, the problem is not just the the extraction of information out of it (i.e. "understanding"), but also its deposition (i.e. "communicating"). Note the insidiousness of the problem: that is, even where one did successfully manage to mentally reconstruct complex representations of the system they want to illustrate, then its translation on paper still requires considerable artistic and technical ability. Regardless of the success of such translation process, it is this output that will serve as the basis for other scientists to inform their own mental pictures (cfr. Figure 1). The circle is vicious, and difficult to escape. A good illustration requires understanding, and a good understanding requires good summary references. If one does not have the possibility to see and explore real mouse embryos, dependence is built. Dependence on either luck and the pre-existence of illustrations of the embryonic context one wanted to illustrate, dependence on the pre-existing/solicited/paid work of rare scientists-artists, or dependence on the difficult investment of time and resources by the scientists themselves as to reach full illustrative autonomy.

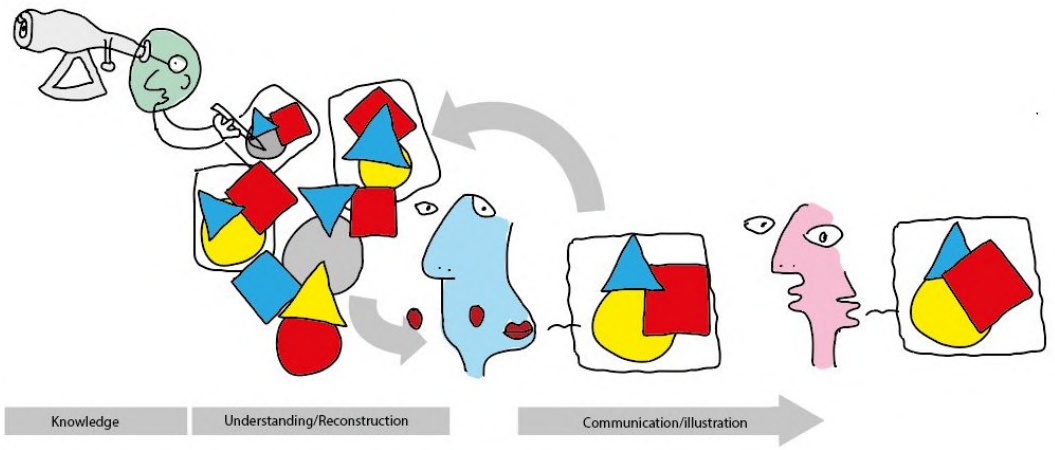


Fig. 1. The vicious cycle of scientific illustration. Blue does not have access to real mouse embryos, and thus their entire understanding depends on the mental reconstruction of pre-existing representations from the literature. Blue's output representations, regardless of the fidelity with which they match Blue's mental reconstructions, will in turn serve as the basis of understanding for other scientist (e.g. Pink, and Blue themselves). Blue is thus entirely dependent on the output of direct observers such as Green, and has to wait for Green to draw the specific scenario needed.

How then to facilitate and democratise both the visualisation and the illustrations of mouse embryos? An important resource has been, and still remains, the Edinburgh Mouse Atlas Project (EMAP), the freely-accessible, annotated, online collection of 3D volumetric data hosted at https://www.emouseatlas.org/emap/ema/home.php, covering mouse development from pre-implantation, to gastrulation, to organogenesis and late post-implantation (Armit et al., 2017; Richardson et al., 2013). 3D models are easily navigable and allow users to define cutting planes as to see virtual histological sections at any desired angle. The value of such resource is maybe understated, but here are on-demand, user-explorable, customisable, 3D (+ timepoints) visualisations of mouse embryos at almost all stages of development. Clearly, this is a complete subversion of the normal illustration-spectator dependant relationship built by traditional developmental visualisation, and a powerful enabler of understanding (cfr. Figure 2) . Crucial in this, is the role played by the actual data-format (i.e. a 3D object in space) and how it allows to completely evade the two-dimensional constraints of the printed page. Put very simply, an interactive, explorable 3D model condenses the informational potential of as many 2D pictures as there are angles to visualise that model.

With the evolution of our formats of scholarly communication, published data is also slowly starting to take life out of the page. Concomitant with the increasing use of experimental strategies outputting volumetric microscopy data, and with their application to the mouse embryo, such data is increasingly being presented as videos accompanying the online version of manuscripts. In these videos, the models are usually animated to rotate around their main axes, they might toggle specific structures on and off to reveal internal structure, or show internal cross-sections. Here is the dimensionality, here is the change over time, features so intimately associated with developmental biology (Hardin, 2008). The models live and are communicated in their 3D environment, if they are not user-explorable they are - at least - explored (cfr. Figure 2) . Even if just provided in their printed version, 3D models still remain potent embryonic representations. One could say that the understanding of e.g. tailbud structure, embryonic heart development, and uterine architecture is almost immediate when shown, respectively, as in (Arora et al., 2016; Dias et al., 2020; Ivanovitch et al., 2017), even without the animated videos that may accompany such publications. Still, access to these models and thus the ability to explore them, is entirely dependent on the sharing practices of each individual set of authors, or of the venue of publication. Even in the case of the EMAP database, where volume data for every single embryonic structure is indeed downloadable (most updated version of the collection available at https://datashare.is.ed.ac.uk/handle/10283/2805), this does not come in a format of easy transfer to common visualisation and

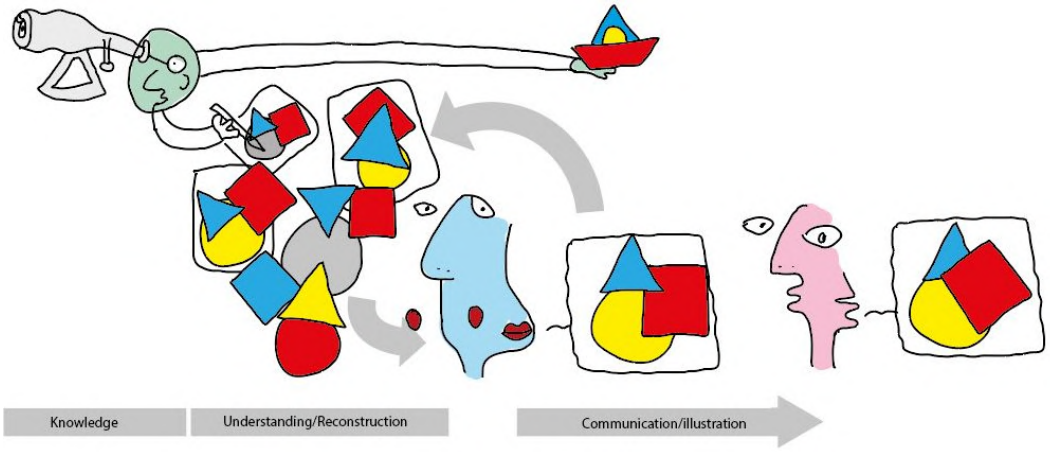


Fig. 2. The power of 3D models. Direct observers like Green do not just illustrate the object from a specific point of view, but also decide to provide a 3D model of the original object (here, a boat). Blue can now observe the object from the specific angle they needed (becomes Green by proxy): the cycle of dependence is bypassed. Notice however that Blue cannot necessarily use the model provided by Green (e.g. only available as a video, or as 2.5D) when communicating with Pink. The cycle of dependence is broken at the understanding/reconstruction step, but not at the communication step.

illustration softwares. These conversion steps, and the digital reuse of 3D objects, requires skills that - while not difficult - are yet not mainstream compared e.g. to the relative familiarity one is nowadays expected to have for 2D illustration software.

In the light of the "illustration -> understanding -> new illustration" framework outlined above , we finally also have to reflect on the usability of our visualisations. That is, to what extent do they allow the reader to transition from passive "visualiser" and "learner", to active user, "illustrator", "teacher", "communicator"? To what extent do they empower them? To what extent to they break this cycle of dependency? Is it possible to imagine a framework where illustrations not only catalyse understanding, but they simultaneously facilitate creative experimentation? Where the lifecycle of an illustration/model does not end upon publication but gets new life in the hands of the reader? There is an untapped value in 3D models, probably uniquely so for the mouse developmental biology community given the existence of repositories such as EMAP. Tapping into it only requires easy access to the models themselves and the technical know-how on how to use and explore such models digitally. And here are scientist that will be able to explore these models on their own. Scientists and students that will be able to incorporate them into their own illustrations. Outreach officers that will be able to deploy them in e.g. virtual or augmented reality applications. 3D printers that will be able to bring these structure to the hands of students and of the public (Figure 3).

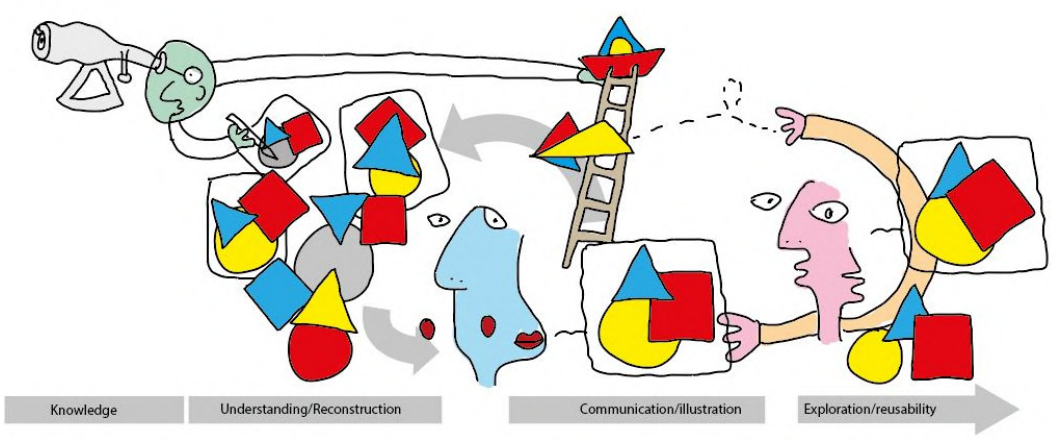


Fig. 3. Visualisations become data. Green not only provides models, but makes them accessible too (e.g. copies of the model can be downloaded and manipulated). Blue can actively explore the object and is even more likely to find the point of view they needed. Pink developed the technical skills to use the model itself, and can now transform/reinterpret it, 3D print it, or use it in new visualisations. All of these elements feed back into the visual data formats available to the community, and thus further catalyse understanding.

I here provide a collection of readily-usable volumetric models for all tissues and stages of mouse peri-implantation development as extracted from the eMouse Atlas Project (E5.0 to E9.0), as well as custom-made models of all pre-implantation stages (E0 to E4.0). These models have been converted to a commonly used 3D format (.stl), and are provided in ready-made files for digital exploration and illustration. I further provide a step-by-step walkthrough on how to practically use these models using the free and open source 3D creation suite Blender. I finally outline possible further uses of these very models in outreach initiatives of varying levels, virtual and augmented reality applications, and 3D printing. I hope these resources and instructions may add to those already available to the mouse developmental biology community, serve as a soft introduction to more members of our community to the world of 3D illustration, and encourage creative exploration and experimentation in such a visually rich and stimulating field that is that of embryology.

A collection of volumetric models of embryonic stages.

An overview of all 16 models available for download is provided in Figure 4. These models cover Theiler Stage (TS) 01 to TS14, and include 8 custom-made models (covering TS01 to TS06, or E0.5 to E4.5), as well as 8 models assembled from EMAP data (covering TS07 to TS14, or E5.0 to E9.0). Also provided are two additional models corresponding to cross-sectional views of the early and late blastocyst models (TS04 and TS05, or E3.0 and E4.0), as to reveal the inner cell mass. As each of the 18 embryonic models provided is given as an assembly of individual subcomponents, a total of 251 individual volumes are thus available for download, exploration, and reuse (see Supplementary Data Table for a complete list).

The models have been deposited at http://doi.org/10.5281/zenodo.4284380 and will be found as 18 separate .blend files (one for each of the 16 stages, +2 open blastocysts). These files are to be opened with the free and open source 3D creation suite Blender (https://www.blender.org/) and have been prepared as to be immediately usable for exploration and illustration (see later paragraphs). Also provided are the 251 separate volumes of each embryonic structure represented, grouped by embryonic stage. These are provided as .stl format, as to be readily used not only for illustration, but also in 3D printing, gaming, and virtual/augmented reality environments. Finally, an additional .blend file with the full embryonic collection is also provided (as shown in Figure 4), as well as a "starting-pack" of pre-made materials to apply to the models (see Figure 5, and **Glossary**).

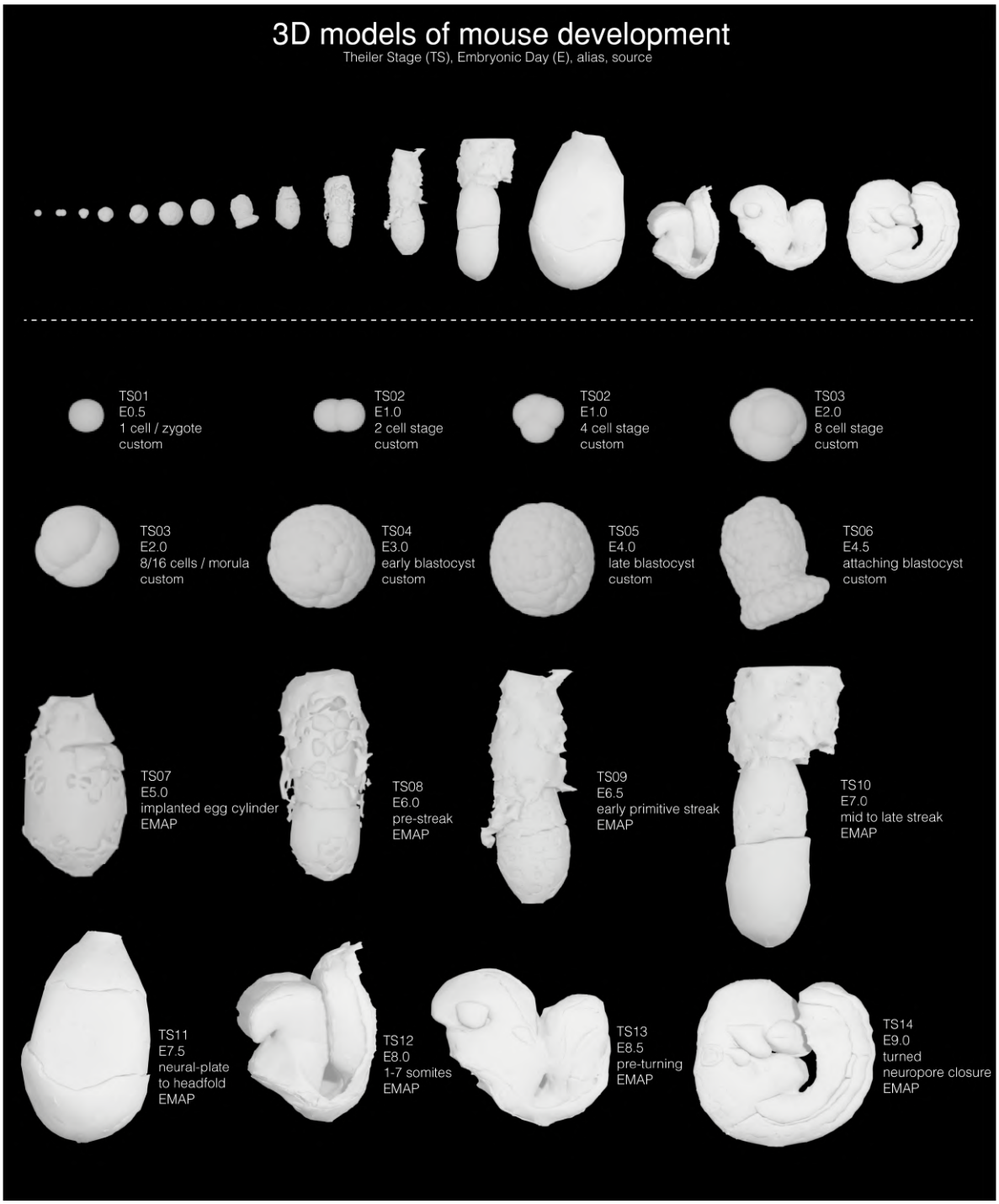


Fig. 4. Summary of the 18 assembled models available for download. for a total of 251 individual volume models. Models for early and late blastocysts are also given in an open version showing the inner cell mass (not shown here).

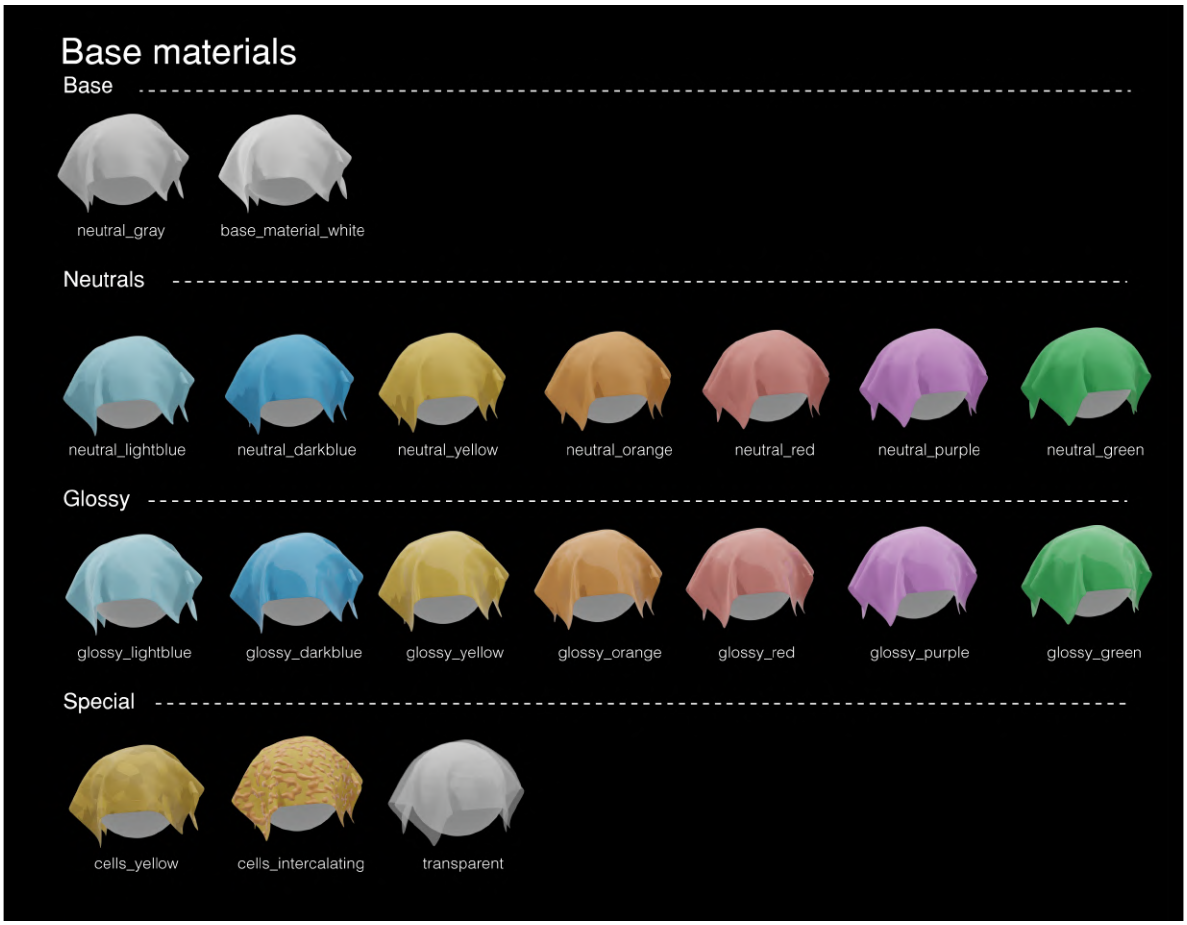


Fig. 5. List of pre-made materials available for download. By default, 3D models will be displayed as grayish-white, but the visual appearance of a model can be controlled by applying materials/shaders to it. To facilitate the process, the above materials can be downloaded and used. These correspond to all the main colours, in a neutral and in a glossy ("wet") version. "Special" materials include one made to mimic the surface of an epithelium, one showing mesenchymal cells migrating above it, and one making the structure transparent. Technical details on how these materials were created are provided in the **Methods section**.

Use of the models for exploration/understanding.

As mentioned in the introduction, one of the biggest advantages of having access to 3D models is that they allow the user to take control over the exploration of these very models (as per the paradigm illustrated in Figure 3). A big part of the models provided here have been long accessible and explorable on the EMAP portal, but they come in a dedicated file format to be open and read with the Java-based viewer JAtlasViewer. Through this viewer, the user can relate 3D structure and histological cross-sections at any cutting plane defined in the application. At the cost of losing the possibility to view cross-section histology (which would however still be available on the EMAP website), Blender provides an alternative way of manipulating and exploring 3D objects, while also allowing easy customisation of the models, rendering to 2.5D illustrations, the creation of scenes with multiple models, sculpting, and more advanced functionalities. Furthermore, and for 3D models that do not benefit from curated interactive visualisation platforms such as those of EMAP, Blender simply provides an easy-to use interaction and exploration platform.

To explore the embryonic models provided, just download the corresponding .blend file. If Blender has been installed on the computer, the file should then be openable just by double-clicking on it (Figure 6.1, here TS13 as an example). One can now freely explore the embryo. Unwanted tissues and structures (e.g. outer layers hiding internal structures) can be toggled off by clicking on the eye icon next to their name in the lateral menu (Figure 6.2). If one is unsure of the name of the tissue to be removed, right clicking on it will highlight its name in orange in the lateral list (later stage models are made up of up to 80 components!). As a rule of thumb, one will likely want to toggle off all extraembryonic tissues: these can be easily identified because their name is preceded by an "X" (see Supplementary Data Table). In the example provided here, most extraembryonic tissues have been removed, as well as mesenchmyal and cardiovascular structures around the tailbud to get a better view of gut tube and notochord.

Exploration of the model can then proceed by scrolling in/out with the centerwheel of the mouse (zoom in/out), moving the mouse while holding the wheel down (rotation around the model), and moving the mouse while holding SHIFT and the wheel down (panning vertically and horizontally). Alternatively, and from Blender 2.8 onward, one can more simply interact with either of the three icons on the top right of the viewer (see Figure 6.3). In the example, this has been done to get a better view of how the endoderm pockets into the anterior of the embryo (left), and to track the notochord as it emerges from the tailbud (right).

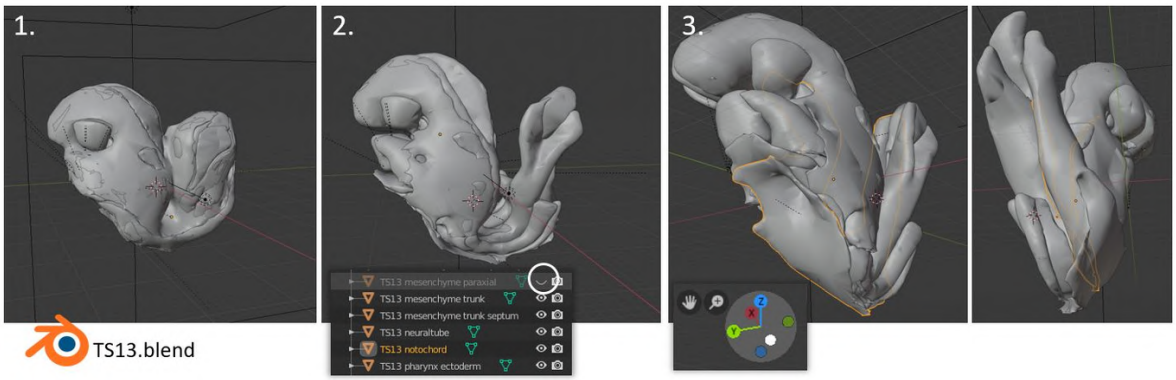


Fig. 6. Exploring 3D embryonic models in Blender: moving perspective.

An alternative way to explore the embryo models provided is to move and rotate the model itself, and to move structures in and out of them (see Figure 7, TS09 as an example). To move the whole embryo at once, select the "HANDLE" object in the list on the right (left-click on the name, Figure 7, bottom). The model can be now moved along the three axes, or rotated around them, by clicking on the corresponding icons at the left of the screen and then interacting with the widgets that appear on top of the model (Figure 7.2, move icon; Figure 7.3, rotate icon). The same actions can be performed with the keyboard shortcuts G (for "grab", followed by X, Y, or Z to lock the movement along these axes), and R (for "rotate", followed by X, Y, or Z to lock the rotation around these axes). The same can be done on individual components, rather than on the whole embryo, by selecting them first (right click on them in the viewer, or left click on their name on the right panel). In Figure 7.4 this has been done on the (visceral) endoderm to move it down and reveal the epiblast and the primitive streak at one side of the embryo (some extraembryonic structures were also toggled off as explained before).

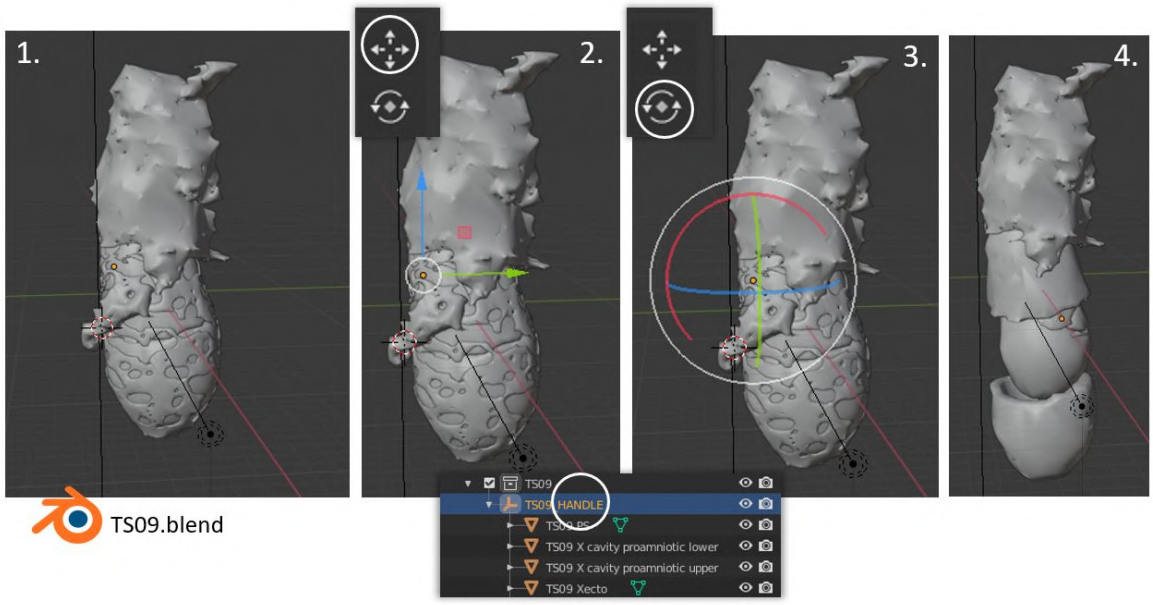


Fig. 7. Exploring 3D embryonic models in Blender: moving the model

Thus, the simple use of basic Blender functions (movement in space, on/off visibility toggling, object interaction and manipulation) allows for unbounded exploration of any of the embryonic models provided, allowing users to easily visualise and understand the spatial relationship of their structure of interest in the context of the whole embryo. In fact, it would be interesting to explore whether such exploration could improve visual-spatial literacy in the users, and whether this could represent a valid and beneficial tool within developmental biology pedagogical settings. To quote Hardin (2008): "the technology already exists to depict embryos on the computer as true 3D objects in 4D space. What is needed is the application of instructional materials development resources toward the production of such models. If such models become widely available, it should be possible to reclaim all four dimensions of the embryo in the undergraduate developmental biology curriculum ".

While these models have been here made available as pre-prepared .blend file, one can exploit these same basic functions to explore any type of volume data (.obj, .wrl, .stl; imported in a Blender scene via File>Import). This has been e.g. recently exploited for 3D exploration of cell migration tracking data in Samal et al. (2020), and can also be applied to explore .obj/.stl meshes generated from e.g. light-sheet imaging.

Use of the models for illustration.

The second advantage allowed by having access to 3D models in Blender is the possibility to create 2D (2.5D) illustrations out of them. This is done by taking snapshots of them at the angle needed: this creates an output image, a process called rendering (see **Glossary**), where the illustration software applies materials to the models (colour, roughness, refractivity...) and light to the scene. Regardless of drawing or artistic skills, one can thus generate reference figures for publications, scientific illustrations, material for outreach purposes, or even art. The illustration process thus becomes focused on the concept and message one wants to illustrate rather than on the technical aspects of it: accurate embryo models are already provided. Furthermore, illustrations of embryos from non-traditional perspectives do not require to mentally recalculate what the embryo would look like (think about how few illustrations show mouse embryos from the ventral or anterior side, compared to the lateral or dorsal sides).

To explain how the models provided here can be used for illustration, we will take a simple case study: that of a scientist wanting to illustrate endoderm development after gastrulation. This process involves extensive shape changes as the endoderm starts as a curved sheet on the surface of the embryo, later makes deep pockets at its two extremities, and finally closes to form a tube (the gut) while the whole embryo is closing on itself (Lewis & Tam, 2006). Illustration is made especially challenging in our case because the scientist is not confident they fully understand how the terminal endoderm pockets fit with other structures at the anterior and posterior of the embryo.

To create an illustration, one first need to set up a scene. In our specific case, the single-stage .blend files available for download are not sufficient: we would like to show models TS09 to TS14 within the same image. We will thus open TS09, and import the other models from the other files by selecting File>Append (Figure 8.1), navigating to the .blend file with the model needed (here TS10), opening its Collections folder, and selecting the TS10 entry. Appended models will be imported within the existing hierarchy, so make sure that you are in the topmost collection (Scene Collection, see Figure 8.1) before importing any model. Once the model has been imported and positioned (Figure 8.2; see previous section for how to move models in Blender), one can repeat the Append process to add each of the following embryonic stages (Figure 8.3). As explained before, specific structures and layers can be then toggled off (eye icon) to expose the structures one is interested in (here, the endoderm) (Figure 8.4).

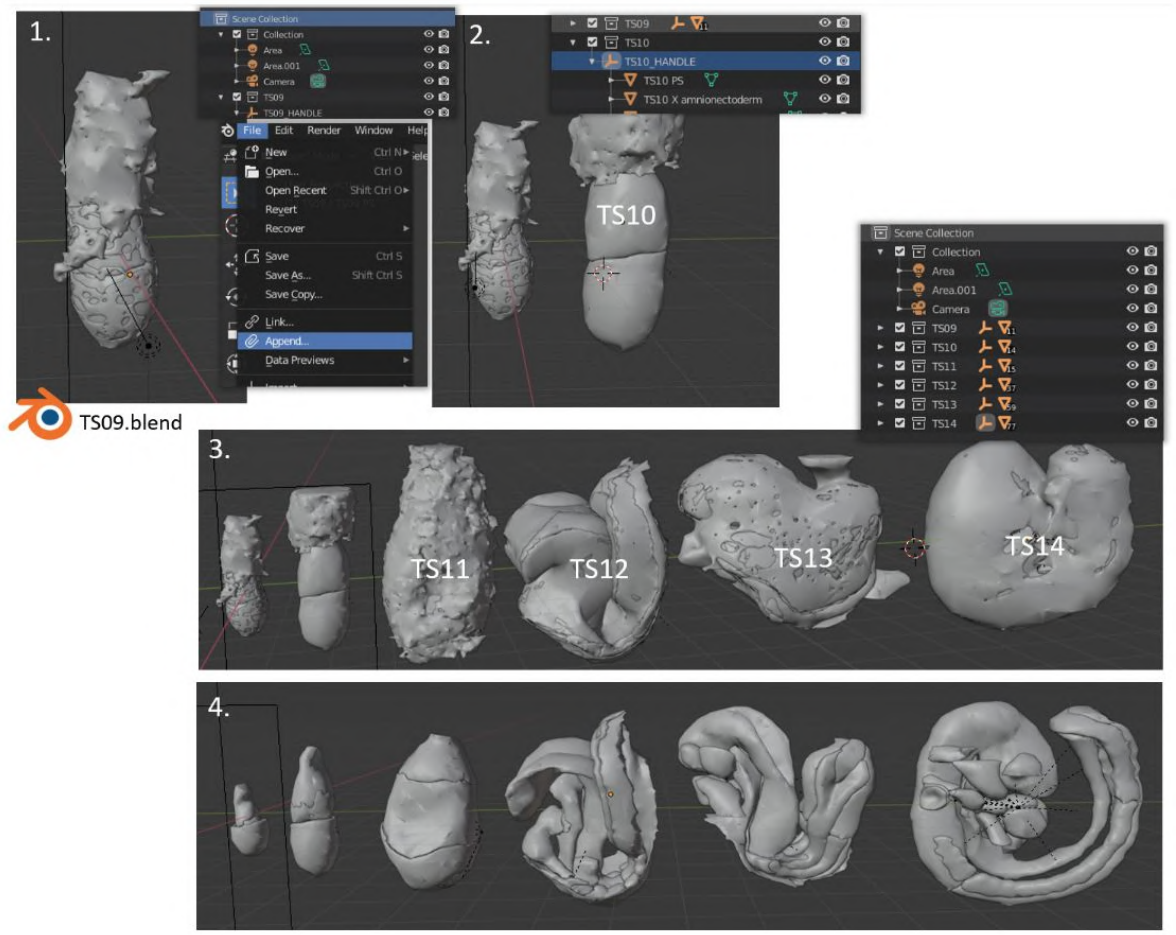


Fig. 8. Importing models from other files by using File>Append>Collection

Once the scene is set with all the models in the right place, it is time to add shaders/materials. Materials determine the final appearance (in its simplest form, the colour) of the models in the output illustration, and as such have a heavy influence on its success. To get an idea of what the final image will look like one can toggle the Rendered Viewport Shading icon at the top right of the viewport (Figure 9.2; allow some processing time). Since all models provided here come with a generic grayish shader, this is indeed how the models will look like (Figure 9.2). An important note is that, depending on where you placed your models into the scene, there might nor be enough light to actually see the real final colour of the model. This case is shown in Figure 9.3B, and corrected in Figure 9.3C by providing more light to the rightmost model. In Blender, light is extremely important, and every .blend model provided here comes with it's own set of lamps. These are the objects highlighted in orange in Figure 9.3A, and can be imagined as literal rectangles shining light out of their surface: one illuminates the front of the model and the other the top. To add lights, select both of them (if using the name list on the right, Shift+LeftClick; if clicking directly on the objects, Shift+RightClick), press Shift+D to duplicate, and then move them to your next model (in Figure 9.2 you can indeed see how additional lights have been placed all along the lineup of models).

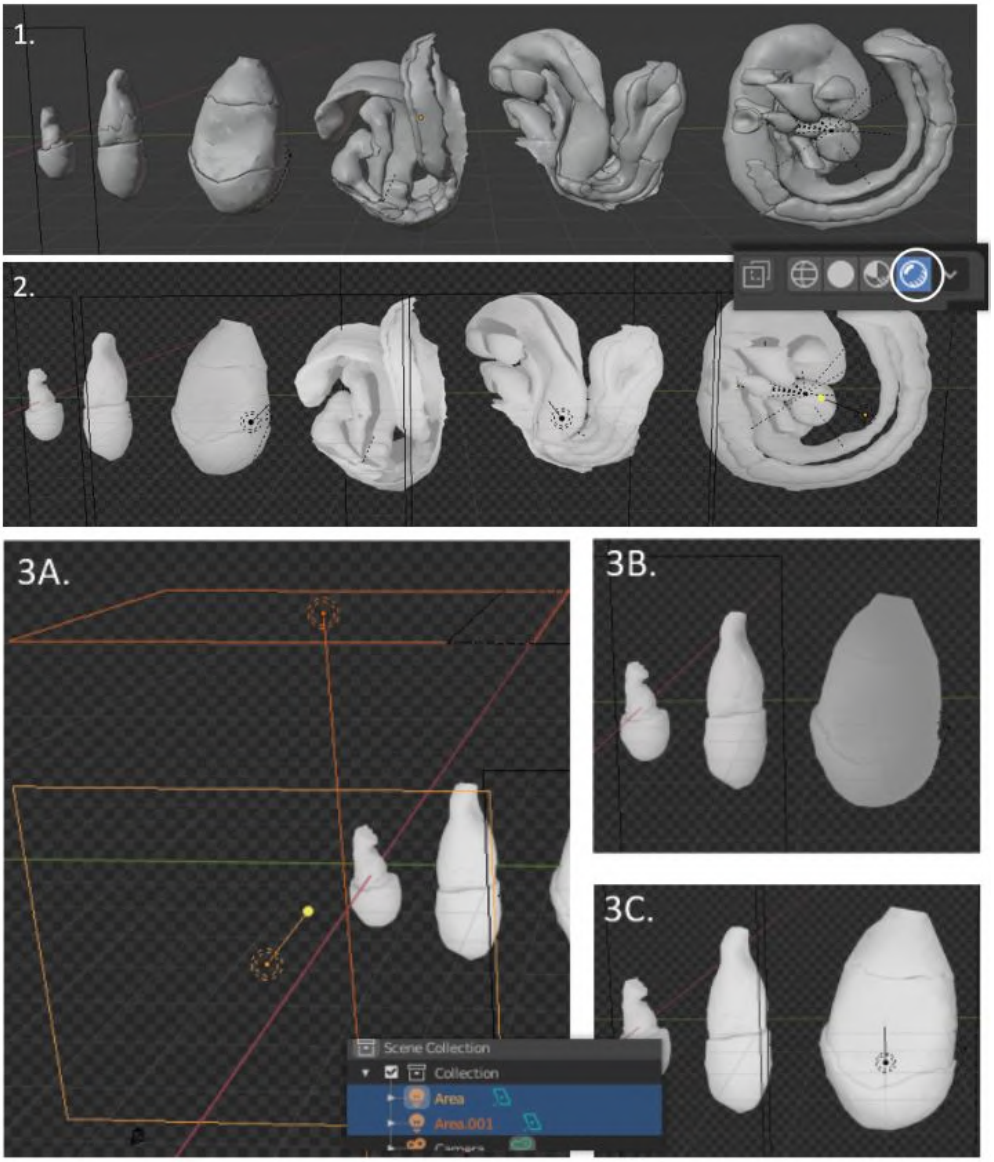


Fig. 9. Rendered viewport and lights

It's then now time to apply materials to the structures we want to highlight (endoderm in this case). For a 2D illustration analogy, this is equivalent to having drawn a shape on paper and now having to colour it in. While Blender allows the creation of virtually any material one can imagine, we facilitated the task here by providing a starter-pack of pre-made ones ("materials.blend", see Figure 5) alongside the embryo models. To import any of these materials into the scene, repeat the same process used to import additional embryo models (File>Append) but this time select the "materials.blend" file, and pick from the "Materials" folder. In this scene we will start by using the "cells_yellow" material. Once it has been imported, apply it by selecting the component you want to colour (here the visceral endoderm, Figure 10.1), clicking on the red checkered sphere icon in the panel on the right (this is the materials icon), and use the drop-down menu to select the material you just imported (Figure 11.1, right). Provided you are still in Rendered View mode (toggled on previously), the model will now look as in Figure 10.2. Proceed for all other models (materials only needed to be appended once) until satisfied (Figure 10.3). Since the endoderm soon starts being populated by cells intercalating from the mesoderm (Viotti et al., 2014), we have switched to the "cells_intercalating" material for all later models to better illustrate this phenomenon.

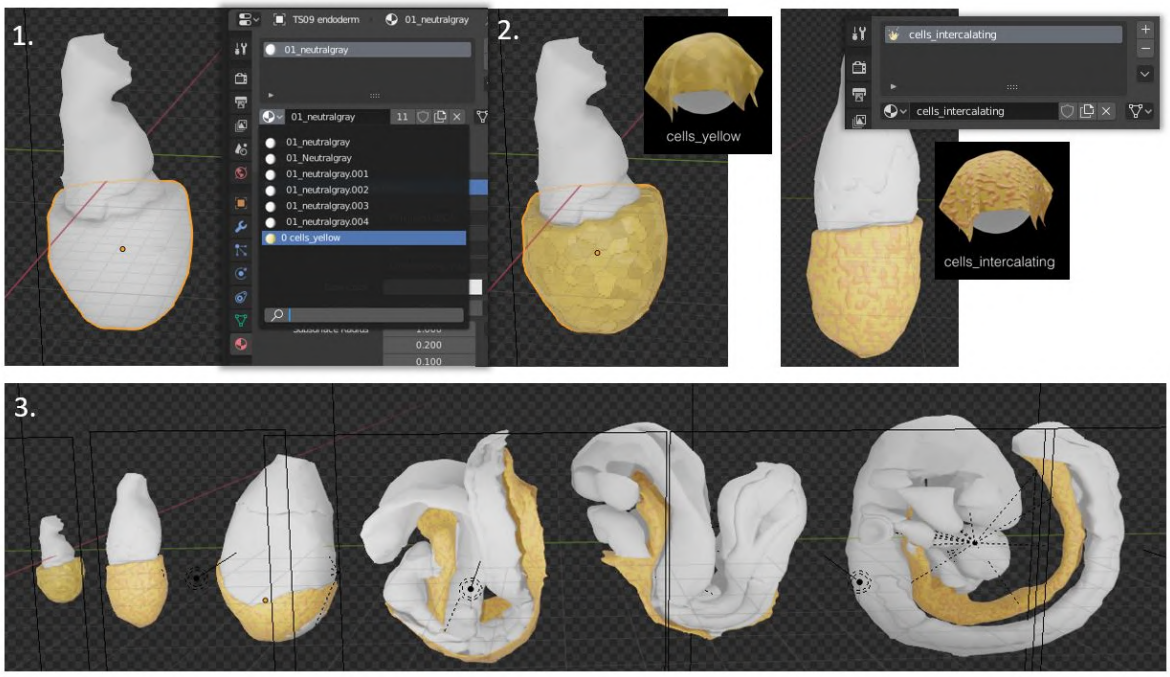


Fig. 10. Applying materials

The models are in place, the scene is illuminated, and materials have been applied. We are now ready to transform this 3D scene into a 2D image. This is done by taking a virtual snapshot of our models, a process called rendering. This "snapshot" is taken with a Camera, which is indeed provided with all Blender scenes, just as lamps and lights are. To place the camera in position, start by positioning your view as to see your models as you want them in the final image (Figure 11.1). With the camera object selected (left-click on its name in the panel on the right), click on View>Align View>Align Active Camera to View (Figure 11.2): this will move the camera in position, and make you look through it. If all your models disappear at this point, this is because they happen to be out of the working distance range of the camera. To correct this, and with the camera object still selected, click on the green camera tab in the bottom right panel, and increase the End value until the models appear again (Figure 11.3, 10000 meters in this case). As in Figure 11.4A, the camera will not exactly capture the whole scene and will likely need to be moved "in" or "out" as to capture a smaller/bigger area. This is done by moving the camera along its relative Z axis, which is set through the keyboard shortcut "G" (for "grab"), followed by "Z" (lock movement to absolute Z axis), followed by "Z" again (lock movement to relative Z axis). Adjust the camera position until satisfied (Figure 11.4B): what you see through the camera is the snapshot that will be taken.

Once the camera is in place, it is time to launch the rendering process. To avoid any surprises, start by making sure that all the objects that had been hidden from view (eye icon toggled off), are also hidden from the final render. Just as you toggled off the eye icon next to the name of each object, toggle off the camera icon too (Figure 11.5A; if this icon is not shown, make it available from the Filters submenu as shown). Another aspect to double-check is that the "Transparent" Film checkbox is ticked in the display settings in the bottom left panel (tab with the camera icon, Figure 11.5B) because this will make the

background transparent when the image is then saved as .png. Finally, the size/resolution of the final image can be set in the rendering tab (Figure 11.5C, tab with printer icon, here 1920px x 1080px). Notice that if you change the ratio of these values, your camera will also resize and you will need to realign it. Once all is set, launch the rendering via Render>Render Image (Figure 11.5C, bottom). A new window will open, and after some processing time, the final image will appear (Figure 11.6C). Save the image via Image>Save As. The image is done (Figure 11.7).

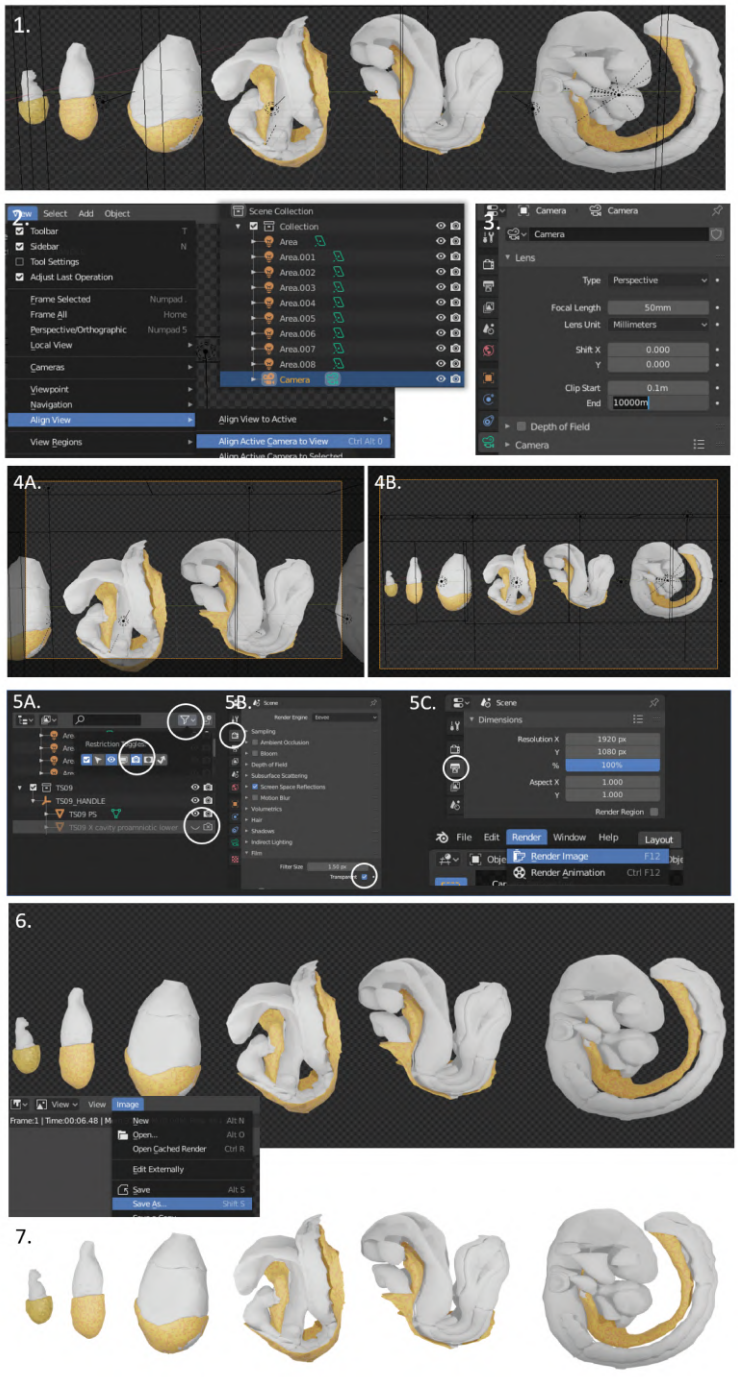


Fig. 11. Camera positioning and rendering.

Illustrating cross-sections.

As a final illustration technique, a developmental biologist might often need to show cross-sections or cut-outs of a specific tissue: that is, not toggle the tissue off view completely, but just remove a section of it as to show underlying structures. This is achieved in Blender through the use of so-called Boolean modifiers, which essentially allow to substract/add shapes from/to one another. A cube or a parallelepiped can thus be intersected with our tissue of interest and the intersection deleted out. Upon removal of the cube, the tissue is left open: a process illustrated in Figure 12.

As a first step, one needs to create the shape to be subtracted. Making sure that one is in the topmost hierarchy (Scene Collection), add a cube to the scene by selecting Add>Mesh>Cube (keyboard shortcut Shift+A) (Figure 12.1). To scale the cube right-click on it (Figure 12.2) and select the Scale icon on the left, or use keyboard shortcut "S", followed by X, Y, or Z to lock the scaling to either of the three axes. Since we here want to cut out the visceral endoderm both in the embryonic and in the extraembryonic region, we make a parallelepiped that is as tall as the embryo (Figure 12.3). Once satisfied with the shape, move it as to intersect your model over the regions you want to carve out (Figure 12.4), select the tissue that needs to be cut (here starting with the embryonic visceral endoderm), and select the wrench icon of the bottom right menu (modifiers tab, Figure 12.5). Under "Add Modifier", select "Boolean", and then make sure that the Operation is set to "Difference", and the Object field displays the shape you want to subtract (our parallelepiped is called "Cube" in this example; Figure 12.5). Click on "Apply" and then repeat the whole process for the extraembryonic visceral endoderm. Once both tissues have been processed, delete the parallelepiped by pressing the keyboard shortcut "X" (Figure 12.6) to reveal what is left of the model (Figure 12.7). We can now clearly see epiblast and primitive streak under this newly created window through the visceral endoderm.

Because the models provided here are not "solid" (they are actually hollow, thin-layered envelopes), the kind of editing described above might reveal exposed holes in the cross-sections (Figure 12.8A, the cut visceral endoderm is hollow). These are problematic because one wants to hide this in the final illustration and give the impression of a solid model. To correct this, switch to Edit Mode (keyboard shortcut "Tab", or as in Figure 12.8A, right) to enable interaction with the mesh of the objects (Figure 12.8B). After deselecting everything (keyboard shortcut Ctrl+A), toggle the edge-selection icon, and select any edge of the mesh making up the border of the exposed side (Figure 12.8C). Now use Select>Select Loops>Edge Loops to automatically select the entire border (Figure 12.8D), and press keyboard shortcut "F" to create a new face within the perimeter defined by this border (Figure 12.8E). By leaving Edit Mode and going back to Object Mode (keyboard shortcut "Tab") one can see that this gives the illusion of a solid model (Figure 12.8F), which can now be processed for illustration as described above.

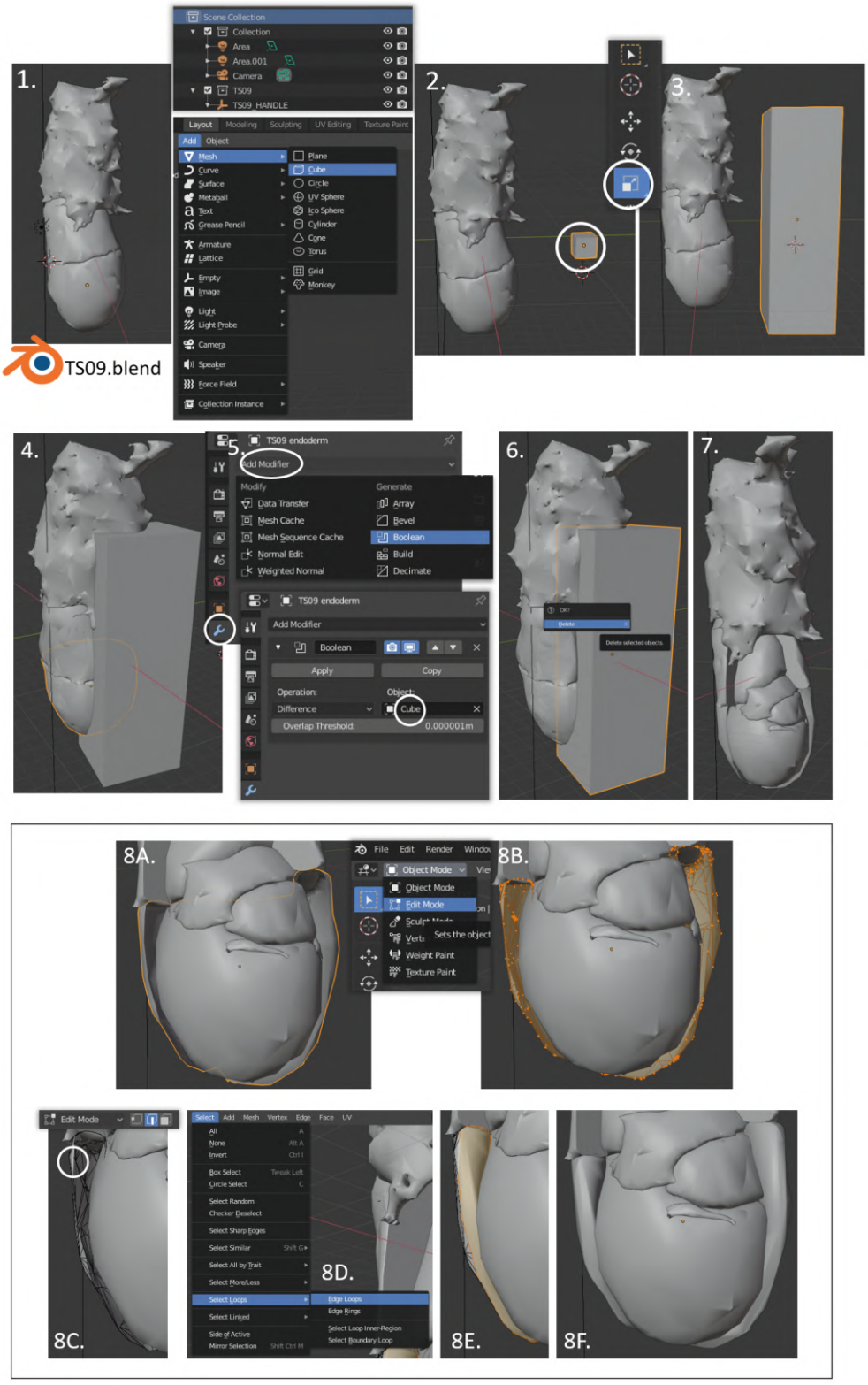


Fig. 12. Creating cut-outs and using Edit Mode

Non-conventional uses and outreach

Crucially to the paradigm outlined in Figure 3, and even though the status-quo (bias) of academia might suggest otherwise, scientific communication does not take its only form as that of a published image in a journal, nor does creative experimentation have to falter in the constraints of the printed page. As such, while 3D models do allow the generation of conventional 2D figures as described above, they also uniquely open up the doors to a wide variety of new forms of user data-engagement which are unthinkable if only dealing with pen and paper. Notably, non-conventional communication media and illustrations tools such as Augmented Reality, Virtual Reality, and 3D printing (see **Glossary**, and Figure 13), all require at their source the availability of 3D models of the objects to be brought to life (usually in either .obj or .stl format). Up to now, and in the absence of readily available compatible models, the use of these technologies to communicate mouse embryology thus required the sculpting of each of these models from scratch in software like Blender, or the generation of appropriate microscopy volume data in the lab. These requirements are unrealistic for many, either in terms of skills or opportunity, and especially since communicators and data generators might not always be the same person. Indeed examples where such approaches have been deployed are few. Yet because of their novelty, and because they engage different senses in addition than just sight, these communication strategies hold extraordinary didactic potential, whether within academia or for outreach.

The goal of this section is not to provide a detailed guide on how to deploy embryo models in virtual reality, augmented reality, or 3D printing, but just to highlight that this is possible with free and open source software. Hopefully, the embryonic models provided here can at least jump start experimentation in these avenues by interested parties. Examples of potential applications include the projection of rotating models out of posters/papers to allow live user exploration of the specific embryonic stages discussed (Figure 13.1), deployment of mouse embryos in virtual reality environments as to captivate unfamiliar but interested audiences for outreach (Figure 13.2), and multimaterial 3D printing of replica models for didactic purposes (Figure 13.3; bones in this example from https://vizbi.org/Posters/2019/B02). While a variety of different solution exist for each of these applications, augmented reality apps can be built for example with a combination of Unity and Vuforia, or the web-based AR.js. A basic script and setup behind the AR triggered by Figure 13.1 is provided at: https://github.com/StefanoVianello/Augmented_Reality. Virtual reality scenes can similarly be created by importing models in Unity and coupling this to a e.g. Google Cardboard interface; and 3D printers only require the user to supply one of the models available here. 3D printers are increasingly common in universities or in so-called community makerspaces, and we recommend considering printers with a dual nozzle setup to allow multimaterial fabrication.

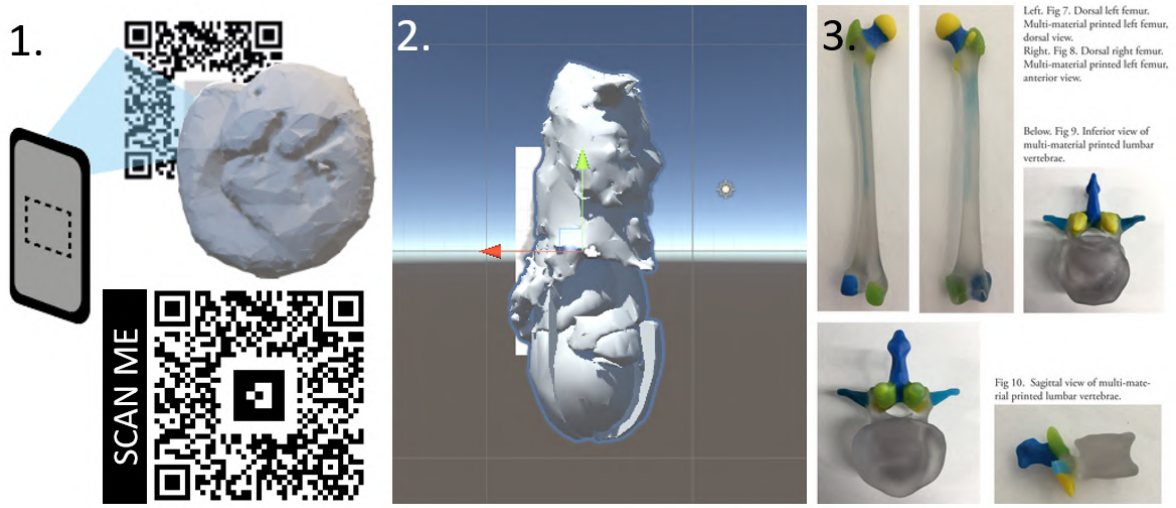


Fig. 13. Non-conventional uses of 3D models for communication 1. Augmented Reality to bring to life posters and papers. 2. Example of Virtual Reality setup (Unity software) 3. 3D printed models from Inoue, Leevy 2019 (https://vizbi.org/Posters/2019/B02)

Conclusion

By converting EMAP models to a more versatile format, and by sharing them alongside models of earlier embryonic stages as ready-to-use .blend files, I hope to instill new life in such an important resource and to make its models even more accessible to the community. Undoubtedly, many researchers in the field of mouse developmental biology are in possession of 3Dmodels of various structures and embryonic stages, or generate these models as part of their research. I hope this guide can serve as a soft introduction to software (such as Blender) that allows exploration and use of these models for illustration, software that is otherwise associated with a steep learning curve and might appear overwhelming in its variety of functionalities. Hopefully, the considerations made here will also serve as an encouragement to authors to make future 3D models models they might generate as accessible as possible to the public, as their value transcend that of publication. For scientists, creators, communicators, teachers, and artists reading this, I hope these models can act as fertile substrate for creative exploration, and ultimately encourage new paradigms through which to live, communicate, and teach Developmental Biology.

Methods

Data availability. All data has been deposited on Zenodo (http://doi.org/10.5281/zenodo.4284380). These include: i) individual .stl models of embryonic subcomponents for each developmental stage (e.g. epiblast, visceral endoderm; 1 folder per embryonic stage); ii) ready-made .blend files where such components have been reassembled as a full embryo model, and in a scene with pre-prepared light sources and aligned camera (1 file per embryonic stage); iii) a .blend file with all embryonic models provided arranged in a temporal lineup; iv) a .blend file with a "starter-pack" of pre-made materials to use for easy illustration.

Processing of EMAP models. EMAP models of each Theiler Stage (TS07 to TS14) were downloaded from the EMAP website (https://www.emouseatlas.org/emap/ema/home.php; also available at https://datashare.is.ed.ac.uk/handle/10283/2805) as individual .zip files. These files contain .vtk and .wlz models of i) the entire embryo and of ii) each individual embryonic subcomponent. Each folder was then decompressed (extracted), .wlz files were removed, and .vtk files were renamed according to the key provided in Supplementary Data Table. These files were then converted into .stl files through the script provided at https://doi.org/10.5281/zenodo.4284367 (Note, an alternative conversion script is also available at https://github.com/ma-tech/PyWoolzScripts/blob/master/WlzDomainToVTKSurf.py, from Bill Hill, EMAP team). The resulting objects were finally imported into Blender 2.80, and reassembled to form a combined model of the entire conceptus for each Theiler Stage. Models corresponding to Theiler stages 01 to 06 are not available from the EMAP website and were custom made in Blender. See dedicated Methods section for more details on how these were made.

Blender assembly of full embryo models. For each Theiler stage (TS07 to TS14), individual .stl components were imported into Blender 2.80 (File>Import>stl), shaded smooth (Object>Shade Smooth), scaled (scale: TS12 TS14 0.2. TS07-11 0.09). Origins were set to the centre of mass of each volume (Object>Set Origin>Origin to centre of mass (volume)), and an "empty" was created (Add > New > Empty) and parented to all imported pieces (control+P > Parent to Object). The whole conceptus was then assigned to a dedicated collection (group of objects, Object>Collection>Move to collection > New Collection), moved to [0,0,0], and rotated 180° with respect to its y-axis as to align the top of the model to the positive z-axis. Models were then rotated with respect to their z axis to align their posterior to the positive y axis (right side of the viewer). Finally, models were all assigned the material "base_material_white", Color Management was kept as default (Filmic), Exposure 3, Gamma 0.75. "Film" was set as "Transparent" so that renders are created as .png by default. Note about scale: all models provided inherited size parameters from the files downloaded from EMAP. This means that TS07 to TS11 are all in scale with respect to each other and proportional to each other because they were all here rescaled by the same factor. Other models are provided unscaled and may need manual adjustment based on user needs (realism vs illustrative needs).

Material preparation. A set of pre-made materials is available for download as to avoid having to create them from scratch. Base and Neutral materials (see Figure 5) were constructed from a single "Principled BSDF" node set on the appropriate base color (e.g. lightblue for the "neutral_lightblue" material), all other settings left to default (GGX, Christensen-Burley, Subsurface: 0, Subsurface Radius: 1-0.2-0.1, Subsurface Color: E7E7E7, Metallic: 0, Specular: 0.5, Specular Tint: 0, Roughness: 0.5, Anisotropic: 0, Anisotropic Rotation: 0, Sheen: 0, Sheen Tint: 0.5, Clearcoat: 0, Clearcoat Roughness: 0.3, IOR: 1.450, Transmission: 0, Transmission Roughness: 0, Emission: 000000, Alpha: 1). Glossy materials will give a wet-like appearance to the model, and were made just as Neutral materials but with Roughness: 0, and Clearcoat: 1. The "cells_yellow" material was created by mixing ("Mix Shader") two "Principled BSDF" nodes with the default settings listed above for Base and Neutral Materials, with two different shades of yellow as Base Color. The "Normal" imput of the node with the lighter shade was plugged to a "Bump" node (Invert: unchecked, Strength: 1, Distance: 1) whose "Height" input was linked to the "Color" output of a "Voronoi Texture" node (Cells, Distance, Closest, Scale: 9.6). The "Fac" output of this same node was linked to a "ColorRamp" node, whose "Color" output was linked to the "Fac" input of the Mix Shader mixing the two original Principled BSDF nodes. The "cells_intercalating" material was created with the same basic setup as "cells_yellow", but a "Musgrave

Texture" node was used instead of a "Voronoi Texture" (fBM, Scale: 18.2, Detail: 4.6, Dimension: 1.614, Lacunarity: 10.5, Offse: 0.5, Gain: 0). Altering the Scale value of this node will allow to change the size of the cells as a function of the size of the tissue of interest. The "Bump" node had values= Invert: unchecked, Strength: 1, Distance: 0.2. The "transparent" material was created by mixing ("Mix Shader") a "Principled BSDF" node set as for Base and Neutral Materials (Base Color: FFFFFF) and a "Transparent BSDF" node (Color: FFFFFF). The "Fac" value of the Mix Shader was set at 0.225 (for the transparent input).

Preparation of pre-implantation models. Models for the 1-, 2-, and 4-cell stages (TS01.blend, TS02-2cell.blend, TS02-4cell.blend) were created from simple sphere shapes (Add>Mesh>UV Sphere). Models for the 8-cell stage, morula, early blastocyst, late blastocyst, and implanting blastocyst stages (TS03-8cell.blend, TS03-16cell.blend, TS04.blend, TS05.blend, TS06.blend) were created by using the Particle Emitter function in Blender. Briefly, an object of the desired shape (a sphere for most models; a sculpted "implanting" mesh for TS06.blend) is added (Add>Mesh>UV Sphere) and selected as an "emitter" object ("Particles" properties tab). Another object (the 1-cell model) is imported (File>Append) and selected as the "emitted" object. If the particle settings of the emitter object is set as "Hair", running the simulation (keyboard shortcut: spacebar) will lead to individual cells emerging out of (and thus covering) the faces of the object. The entire system was then saved as a single model by applying the particle modifier ("Convert") of the emitter object. The entire setup used to create each of the models is included in the .blend files provided, for reference and hidden from view in a dedicated collection called "Emitter_system". The open versions of the early and mid-blastocyst models (TS04_HALF.blend, TS05_HALF.blend) were generated as above, but the emitter object used was a hollowed out hemisphere for the surface that will be covered by trophectoderm cells, and a separate shape for the surface that will be covered by inner cell mass cells (TS04_HALF.blend). For the latter particle system, the "emitted" object was a collection of two different cell objects: a sphere to represent epiblast cells, and another sphere to represent primitive endoderm cells. To instead avoiding having intermingled cells in TS05.blend, two separate inner emitters were used (each producing one type of cell).

ACKNOWLEDGEMENTS

This work would have not been possible if it were not for the existence of the Edinburgh Mouse Atlas Project. I am therefore indebted to the entire EMAP team (https://www.emouseatlas.org/emap/about/people.html) and would like to specifically thank Chris Armit for its support and encouragement at early and late stages of this project, and for his feedback on the manuscript. The guide and resources presented here are indeed intended to give new lives to EMAP models, and make them even more accessible to the community.Thanks also go to the Stack Exchange community (and specifically, user Normanius) for the code to convert volumes into .stl. I would like to thank André Dias (Instituto Gulbenkian de Ciência) for his support, correspondence, comments on the final versions of the manuscript, and for being so generous with his own 3D models. I would finally like to thank Francesca Vianello for introducing me to ARjs and for essentially setting up the entire ARjs Augmented Reality experience deployed here.

Bibliography

Armit, C., Richardson, L., Venkataraman, S., Graham, L., Burton, N., Hill, B., Yang, Y., & Baldock, R. A. (2017). eMouseAtlas: An atlas-based resource for understanding mammalian embryogenesis. *Developmental Biology*, *423*(1), 1–11. URL https://doi.org/10.1016%2Fj.ydbio.2017.01.023

Arnold, S. J., & Robertson, E. J. (2009). Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. *Nature Reviews Molecular Cell Biology*, *10*(2), 91–103. URL https://doi.org/10.1038%2Fnrm2618

Arora, R., Fries, A., Oelerich, K., Marchuk, K., Sabeur, K., Giudice, L. C., & Laird, D. J. (2016). Insights from imaging the implanting embryo and the uterine environment in three dimensions. *Development*, *143*(24), 4749–4754. URL https://doi.org/10.1242%2Fdev.144386

Chadwick (1978). Some aspects of the development of geological thinking. *Journal of Geology Teaching*, *3:142-8.*.

Dias, A., Lozovska, A., Wymeersch, F. J., Nóvoa, A., Binagui-Casas, A., Sobral, D., Martins, G. G., Wilson, V., & Mallo, M. (2020). A TgfRI/Snai1-dependent developmental module at the core of vertebrate axial elongation. URL https://doi.org/10.1101%2F2020.03.09.983809

Hardin, J. (2008). The Missing Dimension in Developmental Biology Education. *CBE—Life Sciences Education*, *7*(1), 13–16. URL https://doi.org/10.1187%2Flse.7.1.cbe13

Hashimoto, K., & Nakatsuji, N. (1989). Formation of the Primitive Streak and Mesoderm Cells in Mouse Embryos-Detailed Scanning Electron Microscopical Study. (primitive streak/cell migration/extracellular matrix/mouse gastrulation/scanning electron microscopy). *Development Growth and Differentiation*, *31*(3), 209–218. URL https://doi.org/10.1111%2Fj.1440-169x.1989.00209.x

Ivanovitch, K., Temiño, S., & Torres, M. (2017). Live imaging of heart tube development in mouse reveals alternating phases of cardiac differentiation and morphogenesis. *eLife*, *6*. URL https://doi.org/10.7554%2Felife.30668

Kali, Y., & Nir, O. (1996). Spatial abilities of high-school students in the perception of geologic structures. *Journal of Research in Science Teaching*, *33*.

Lewis, S. L., & Tam, P. P. (2006). Definitive endoderm of the mouse embryo: Formation cell fates, and morphogenetic function. *Developmental Dynamics*, *235*(9), 2315–2329. URL https://doi.org/10.1002%2Fdvdy.20846

McDole, K., Guignard, L., Amat, F., Berger, A., Malandain, G., Royer, L. A., Turaga, S. C., Branson, K., & Keller, P. J. (2018). In Toto Imaging and Reconstruction of Post-Implantation Mouse Development at the Single-Cell Level. *Cell*, *175*(3), 859–876.e33. URL https://doi.org/10.1016%2Fj.cell.2018.09.031

Milner-Bolotin, M., & Nashon, S. M. (2011). The essence of student visual–spatial literacy and higher order thinking skills in undergraduate biology. *Protoplasma*, *249*(S1), 25–30. URL https://doi.org/10.1007%2Fs00709-011-0346-6

Nahaboo, W., & Migeotte, I. (2018). Cleavage and Gastrulation in the Mouse Embryo. *eLS*. Accessed on Sat, April 11, 2020. URL https://onlinelibrary.wiley.com/doi/10.1002/9780470015902.a0001068.pub3

NRC, N. R. C. (2012). *Discipline-Based Education Research*. National Academies Press. URL https://doi.org/10.17226%2F13362

Pijuan-Sala, B., Griffiths, J. A., Guibentif, C., Hiscock, T. W., Jawaid, W., Calero-Nieto, F. J., Mulas, C., Ibarra-Soria, X., Tyser, R. C. V., Ho, D. L. L., Reik, W., Srinivas, S., Simons, B. D., Nichols, J., Marioni, J. C., & Göttgens, B. (2019). A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature*, *566*(7745), 490–495. URL https://doi.org/10.1038%2Fs41586-019-0933-9

Richardson, L., Venkataraman, S., Stevenson, P., Yang, Y., Moss, J., Graham, L., Burton, N., Hill, B., Rao, J., Baldock, R. A., & Armit, C. (2013). EMAGE mouse embryo spatial gene expression database: 2014 update. *Nucleic Acids Research*, *42*(D1), D835–D844. URL https://doi.org/10.1093%2Fnar%2Fgkt1155

Rivera-Pérez, J. A., & Hadjantonakis, A.-K. (2014). The Dynamics of Morphogenesis in the Early Mouse Embryo. *Cold Spring Harbor Perspectives in Biology*, *7*(11), a015867. URL https://doi.org/10.1101%2Fcshperspect.a015867

Samal, P., Maurer, P., Blitterswijk, C., Truckenmüller, R., & Giselbrecht, S. (2020). A New Microengineered Platform for 4D Tracking of Single Cells in a Stem-Cell-Based In Vitro Morphogenesis Model. *Advanced Materials*, (p. 1907966). URL https://doi.org/10.1002%2Fadma.201907966

Saykali, B., Mathiah, N., Nahaboo, W., Racu, M.-L., Hammou, L., Defrance, M., & Migeotte, I. (2019). Distinct mesoderm migration phenotypes in extra-embryonic and embryonic regions of the early mouse embryo. *eLife*, *8*. URL https://doi.org/10.7554%2Felife.42434

Snow, M. H. L. (1977). Gastrulation in the mouse: Growth and regionalization of the epiblast. *Development*, *42: 293-303*. Accessed on Sat, April 11, 2020. URL https://dev.biologists.org/content/42/1/293

Titus, S., & Horsman, E. (2009). Characterizing and Improving Spatial Visualization Skills. *Journal of Geoscience Education*, *57*(4), 242–254. URL https://doi.org/10.5408%2F1.3559671

Viotti, M., Foley, A. C., & Hadjantonakis, A.-K. (2014). Gutsy moves in mice: cellular and molecular dynamics of endoderm morphogenesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1657), 20130547. URL https://doi.org/10.1098%2Frstb.2013.0547

Bibliography and Back matter

> *Citations are feminist memory [...] Feminist bricks: they are the materials through which, from which, we create our dwellings*
>
> — **Sara Ahmed**
> (Ahmed, 2016)

Bibliography

Aguilera-Castrejon, Alejandro, Bernardo Oldak, Tom Shani, et al. (2021). “Ex utero mouse embryogenesis from pre-gastrulation to late organogenesis”. In: *Nature* 593.7857, pp. 119–124 (cit. on p. 18).

Ahmed, Sara (2016). *Living a feminist life*. duke university Press (cit. on p. 133).

Allman, George James and Edward Forbes (1853). “XV. On the anatomy and physiology of cordylophora, a contribution to our knowledge of the Tubularian zoophytes”. In: *Philosophical Transactions of the Royal Society of London* 143, pp. 367–384 (cit. on p. 1).

Alvers, Ashley L, Sean Ryan, Paul J Scherz, Jan Huisken, and Michel Bagnat (2014). “Single continuous lumen formation in the zebrafish gut is mediated by smoothened-dependent tissue remodeling”. In: *Development* 141.5, pp. 1110–1119 (cit. on p. 87).

Amadei, Gianluca, Kasey YC Lau, Joachim De Jonghe, et al. (2021). “Inducible stem-cell-derived embryos capture mouse morphogenetic events in vitro”. In: *Developmental cell* 56.3, pp. 366–382 (cit. on p. 19).

Ameri, Jacqueline, Anders Ståhlberg, Jesper Pedersen, et al. (2010). “FGF2 specifies hESC-derived definitive endoderm into foregut/midgut cell lineages in a concentration-dependent manner”. In: *Stem Cells* 28.1, pp. 45–56 (cit. on pp. 17, 89).

Anderson, William J, Qiao Zhou, Victor Alcalde, et al. (2008). “Genetic targeting of the endoderm with claudin-6CreER”. In: *Developmental Dynamics* 237.2, pp. 504–512 (cit. on p. 63).

Anlaş, Kerim, Nicola Gritti, David Oriola, et al. (2021). “Dynamics of anteroposterior axis establishment in a mammalian embryo-like system”. In: *bioRxiv* (cit. on pp. 21, 22, 24, 30, 37).

Apelqvist, Åsa, Ulf Ahlgren, and Helena Edlund (1997). “Sonic hedgehog directs specialised mesoderm differentiation in the intestine and pancreas”. In: *Current Biology* 7.10, pp. 801–804 (cit. on p. 16).

Argelaguet, Ricard, Stephen J Clark, Hisham Mohammed, et al. (2019). “Multi-omics profiling of mouse gastrulation at single-cell resolution”. In: *Nature* 576.7787, pp. 487–491 (cit. on p. 7).

Armit, Chris, Lorna Richardson, Shanmugasundaram Venkataraman, et al. (2017). "eMouseAtlas: An atlas-based resource for understanding mammalian embryogenesis". In: *Developmental biology* 423.1, pp. 1–11 (cit. on p. 2).

Arnold, Sebastian J, Ulf K Hofmann, Elizabeth K Bikoff, and Elizabeth J Robertson (2008). "Pivotal roles for eomesodermin during axis formation, epithelium-to-mesenchyme transition and endoderm specification in the mouse". In: *Development* 135.3, pp. 501–511 (cit. on p. 10).

Arnold, Sebastian J and Elizabeth J Robertson (2009). "Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo". In: *Nature reviews Molecular cell biology* 10.2, pp. 91–103 (cit. on pp. 4, 5, 24).

Aulehla, Alexander and Olivier Pourquié (2010). "Signaling gradients during paraxial mesoderm development". In: *Cold Spring Harbor perspectives in biology* 2.2, a000869 (cit. on p. 17).

Balmer, Sophie, Sonja Nowotschin, and Anna-Katerina Hadjantonakis (2016). "Notochord morphogenesis in mice: Current understanding & open questions". In: *Developmental Dynamics* 245.5, pp. 547–557 (cit. on p. 16).

Bambara, Toni Cade (1992). *The salt eaters*. Vintage (cit. on p. vii).

Bardot, Evan, Damelys Calderon, Francis Santoriello, et al. (2017). "Foxa2 identifies a cardiac progenitor population with ventricular differentiation potential". In: *Nature communications* 8.1, pp. 1–15 (cit. on pp. 7, 26).

Bardot, Evan S and Anna-Katerina Hadjantonakis (2020). "Mouse gastrulation: Coordination of tissue patterning, specification and diversification of cell fate". In: *Mechanisms of development* 163, p. 103617 (cit. on pp. 2, 6, 10, 92).

Barlow, Linda A (2001). "Specification of pharyngeal endoderm is dependent on early signals from axial mesoderm". In: *Development* 128.22, pp. 4573–4583 (cit. on p. 45).

Beccari, Leonardo, Naomi Moris, Mehmet Girgin, et al. (2018). "Multi-axial self-organization properties of mouse embryonic stem cells into gastruloids". In: *Nature* 562.7726, pp. 272–276 (cit. on pp. 19, 20, 22–24, 26, 30, 32, 45, 59, 60, 81–83, 88, 89, 92).

Beck, F, T Erler, A Russell, and R James (1995). "Expression of Cdx-2 in the mouse embryo and placenta: possible role in patterning of the extra-embryonic membranes". In: *Developmental Dynamics* 204.3, pp. 219–227 (cit. on pp. 60, 67).

Bedzhov, Ivan and Magdalena Zernicka-Goetz (2014). "Self-organizing properties of mouse pluripotent cells initiate morphogenesis upon implantation". In: *Cell* 156.5, pp. 1032–1044 (cit. on p. 18).

Bérenger-Currias, Noémie MLP, Maria Mircea, Esmée Adegeest, et al. (2020). "Extraembryonic endoderm cells induce neuroepithelial tissue in gastruloids". In: *bioRxiv*, pp. 2020–02 (cit. on pp. 20, 22, 32, 81, 91).

Biben, Christine, Cheng-Chun Wang, and Richard P Harvey (2004). "NK-2 class homeobox genes and pharyngeal/oral patterning: Nkx2-3 is required for salivary gland and tooth morphogenesis." In: *International Journal of Developmental Biology* 46.4, pp. 415–422 (cit. on p. 64).

Birx, H James (2009). *Encyclopedia of time: Science, philosophy, theology, & culture*. Vol. 1. Sage (cit. on p. 59).

Brink, Susanne C van den, Anna Alemany, Vincent van Batenburg, et al. (2020). "Single-cell and spatial transcriptomics reveal somitogenesis in gastruloids". In: *Nature* 582.7812, pp. 405–409 (cit. on pp. 20, 22, 32, 47, 59, 81, 83, 84, 91, 92).

Brink, Susanne C Van den, Peter Baillie-Johnson, Tina Balayo, et al. (2014). "Symmetry breaking, germ layer specification and axial organisation in aggregates of mouse embryonic stem cells". In: *Development* 141.22, pp. 4231–4242 (cit. on pp. 19, 20, 22–24, 26, 32, 82, 89, 92).

Brink, Susanne C van den and Alexander van Oudenaarden (2021). "3D gastruloids: a novel frontier in stem cell-based in vitro modeling of mammalian gastrulation". In: *Trends in Cell Biology* (cit. on pp. 20, 21).

Burtscher, Ingo, Wenke Barkey, and Heiko Lickert (2013). "Foxa2-venus fusion reporter mouse line allows live-cell analysis of endoderm-derived organ formation". In: *genesis* 51.8, pp. 596–604 (cit. on pp. 5, 9, 60).

Burtscher, Ingo, Wenke Barkey, Michael Schwarzfischer, Fabian J Theis, and Heiko Lickert (2012). "The Sox17-mCherry fusion mouse line allows visualization of endoderm and vascular endothelial development". In: *Genesis* 50.6, pp. 496–505 (cit. on p. 60).

Burtscher, Ingo and Heiko Lickert (2009). "Foxa2 regulates polarity and epithelialization in the endoderm germ layer of the mouse embryo". In: *Development* 136.6, pp. 1029–1038 (cit. on pp. 4–8, 11, 26, 30, 60, 86).

Cachat, Elise, Weijia Liu, Kim C Martin, et al. (2016). "2-and 3-dimensional synthetic large-scale de novo patterning by mammalian cells through phase separation". In: *Scientific reports* 6.1, pp. 1–8 (cit. on p. 10).

Campbell, Kyra (2018). "Contribution of epithelial-mesenchymal transitions to organogenesis and cancer metastasis". In: *Current opinion in cell biology* 55, pp. 30–35 (cit. on pp. 9–11).

Campbell, Kyra and Jordi Casanova (2015). "A role for E-cadherin in ensuring cohesive migration of a heterogeneous population of non-epithelial cells". In: *Nature communications* 6.1, pp. 1–11 (cit. on p. 10).

Campbell, Kyra, Gavin Whissell, Xavier Franch-Marro, Eduard Batlle, and Jordi Casanova (2011). "Specific GATA factors act as conserved inducers of an endodermal-EMT". In: *Developmental cell* 21.6, pp. 1051–1061 (cit. on p. 9).

Cano, Amparo, Mirna A Pérez-Moreno, Isabel Rodrigo, et al. (2000). "The transcription factor snail controls epithelial–mesenchymal transitions by repressing E-cadherin expression". In: *Nature cell biology* 2.2, pp. 76–83 (cit. on p. 9).

Carlson, BM (1994). "Digestive and respiratory systems and body cavities". In: *Human embryology and development biology. St Louis: Mosby Year Book Inc*, pp. 307–39 (cit. on pp. 2, 6, 15).

Carver, Ethan A, Rulang Jiang, Yu Lan, Kathleen F Oram, and Thomas Gridley (2001). "The mouse snail gene encodes a key regulator of the epithelial-mesenchymal transition". In: *Molecular and cellular biology* 21.23, pp. 8184–8188 (cit. on p. 9).

Cermola, Federica, Cristina D'Aniello, Rosarita Tatè, et al. (2021). "Gastruloid development competence discriminates different states of pluripotency". In: *Stem cell reports* 16.2, pp. 354–369 (cit. on p. 22).

Cernilogar, Filippo M, Stefan Hasenöder, Zeyang Wang, et al. (2019). "Pre-marked chromatin and transcription factor co-binding shape the pioneering activity of Foxa2". In: *Nucleic acids research* 47.17, pp. 9069–9086 (cit. on p. 7).

Cervantes, Sara, Terry P Yamaguchi, and Matthias Hebrok (2009). "Wnt5a is essential for intestinal elongation in mice". In: *Developmental biology* 326.2, pp. 285–294 (cit. on p. 12).

Chari, Tara, Joeyta Banerjee, and Lior Pachter (2021). "The specious art of single-cell genomics". In: *bioRxiv* (cit. on p. 64).

Chazaud, Claire, Yojiro Yamanaka, Tony Pawson, and Janet Rossant (2006). "Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway". In: *Developmental cell* 10.5, pp. 615–624 (cit. on p. 4).

Cheng, Tao, Yanyi Xing, Yunfei Li, et al. (2021). "Single cell response landscape of graded Nodal signaling in zebrafish explants". In: *bioRxiv* (cit. on pp. 21, 88).

Choi, Eunyoung, Marine R-C Kraus, Laurence A Lemaire, et al. (2012). "Dual lineage-specific expression of Sox17 during mouse embryogenesis". In: *Stem cells* 30.10, pp. 2297–2308 (cit. on pp. 26, 61).

Cleaver, Ondine and Paul A Krieg (2001). "Notochord patterning of the endoderm". In: *Developmental biology* 234.1, pp. 1–12 (cit. on p. 17).

Costello, Ita, Sonja Nowotschin, Xin Sun, et al. (2015). "Lhx1 functions together with Otx2, Foxa2, and Ldb1 to govern anterior mesendoderm, node, and midline development". In: *Genes & development* 29.20, pp. 2108–2122 (cit. on pp. 63, 64).

D'Amour, Kevin A, Alan D Agulnick, Susan Eliazer, et al. (2005). "Efficient differentiation of human embryonic stem cells to definitive endoderm". In: *Nature biotechnology* 23.12, pp. 1534–1541 (cit. on p. 6).

Deluz, Cédric, Elias T Friman, Daniel Strebinger, et al. (2016). “A role for mitotic bookmarking of SOX2 in pluripotency and differentiation”. In: *Genes & development* 30.22, pp. 2538–2550 (cit. on pp. 23, 40, 54, 73).

Deschamps, Jacqueline and Johan van Nes (2005). “Developmental regulation of the Hox genes during axial morphogenesis in the mouse”. In: (cit. on p. 17).

Dessimoz, Jessica, Robert Opoka, Jennifer J Kordich, Anne Grapin-Botton, and James M Wells (2006). “FGF signaling is necessary for establishing gut tube domains alongthe anterior–posterior axis in vivo”. In: *Mechanisms of development* 123.1, pp. 42–55 (cit. on p. 17).

Deutsch, Gail, Joonil Jung, Minghua Zheng, José Lóra, and Kenneth S Zaret (2001). “A bipotential precursor population for pancreas and liver within the embryonic endoderm”. In: *Development* 128.6, pp. 871–881 (cit. on p. 16).

Duncan, Stephen A and Alistair J Watt (2001). “BMPs on the road to hepatogenesis”. In: *Genes & development* 15.15, pp. 1879–1884 (cit. on p. 16).

Echelard, Yann, Douglas J Epstein, Benoit St-Jacques, et al. (1993). “Sonic hedgehog, a member of a family of putative signaling molecules, is implicated in the regulation of CNS polarity”. In: *Cell* 75.7, pp. 1417–1430 (cit. on p. 16).

Ellis, B, P Haaland, F Hahne, et al. (2019). “flowCore: Basic structures for flow cytometry data”. In: *R package version* 1.0 (cit. on p. 57).

Engert, Silvia, Ingo Burtscher, W Perry Liao, et al. (2013). “Wnt/beta-catenin signalling regulates Sox17 expression and is essential for organizer and endoderm formation in the mouse”. In: *Development* 140.15, pp. 3128–3138 (cit. on p. 8).

Ereskovsky, Alexander V and Archil K Dondua (2006). “The problem of germ layers in sponges (Porifera) and some issues concerning early metazoan evolution”. In: *Zoologischer Anzeiger-A Journal of Comparative Zoology* 245.2, pp. 65–76 (cit. on p. 1).

Etzler, Marilynn E and Elvin A Kabat (1970). “Purification and characterization of a lectin (plant hemagglutinin) with blood group A specificity from Dolichos biflorus”. In: *Biochemistry* 9.4, pp. 869–877 (cit. on p. 33).

Fehling, Hans Joorg, Georges Lacaud, Atsushi Kubo, et al. (2003). “Tracking mesoderm induction and its specification to the hemangioblast during embryonic stem cell differentiation”. In: (cit. on p. 54).

Ferrer-Vaquer, Anna, Anna Piliszek, Guangnan Tian, et al. (2010). “A sensitive and bright single-cell resolution live imaging reporter of Wnt/ss-catenin signaling in the mouse”. In: *BMC developmental biology* 10.1, pp. 1–18 (cit. on p. 90).

Finley, Kenneth R, Jason Tennessen, and William Shawlot (2003). “The mouse secreted frizzled-related protein 5 gene is expressed in the anterior visceral endoderm and foregut endoderm during early post-implantation development”. In: *Gene Expression Patterns* 3.5, pp. 681–684 (cit. on p. 64).

Franklin, Vanessa, Poh Lynn Khoo, Heidi Bildsoe, et al. (2008). “Regionalisation of the endoderm progenitors and morphogenesis of the gut portals of the mouse embryo”. In: *Mechanisms of development* 125.7, pp. 587–600 (cit. on pp. 11, 12).

Fukui, Atsushi and Hiroyuki Kuroda (2007). “Larvae of Lamprogrammus shcherbachevi (Ophidiiformes: Ophidiidae) from the western North Pacific Ocean”. In: *Ichthyological Research* 54.1, pp. 74–80 (cit. on p. 1).

Fulton, Timothy, Vikas Trivedi, Andrea Attardi, et al. (2020). “Axis specification in zebrafish is robust to cell mixing and reveals a regulation of pattern formation by morphogenesis”. In: *Current Biology* 30.15, pp. 2984–2994 (cit. on p. 21).

Garriock, Robert J, Ravindra B Chalamalasetty, JianJian Zhu, et al. (2020). “A dorsal-ventral gradient of Wnt3a/β-catenin signals controls mouse hindgut extension and colon formation”. In: *Development* 147.8, dev185108 (cit. on pp. 8, 17).

Gavrilov, Svetlana and Elizabeth Lacy (2013). “Genetic dissection of ventral folding morphogenesis in mouse: embryonic visceral endoderm-supplied BMP2 positions head and heart”. In: *Current opinion in genetics & development* 23.4, pp. 461–469 (cit. on pp. 12, 87).

Gerhart, John (1999). “1998 Warkany lecture: signaling pathways in development”. In: *Teratology* 60.4, pp. 226–239 (cit. on p. 17).

Gharibi, Borzo, Emanuel Goncalves, Buhe Nashun, et al. (2020). “A FGF2-mediated incoherent feedforward loop induces Erk inhibition and promotes naïve pluripotency”. In: *BioRxiv* (cit. on pp. 48, 82, 92).

Girgin, Mehmet U, Nicolas Broguiere, Lorenzo Mattolini, and Matthias P Lutolf (2021). “Gastruloids generated without exogenous Wnt activation develop anterior neural tissues”. In: *Stem cell reports* 16.5, pp. 1143–1155 (cit. on pp. 20, 22, 32).

Gittes, George K and William J Rutter (1992). “Onset of cell-specific gene expression in the developing mouse pancreas.” In: *Proceedings of the National Academy of Sciences* 89.3, pp. 1128–1132 (cit. on p. 16).

Gonzalez, David M, Nadine Schrode, Tasneem Ebrahim, et al. (2021). “Understanding Mechanisms of Chamber-Specific Differentiation Through Combination of Lineage Tracing and Single Cell Transcriptomics”. In: *bioRxiv* (cit. on p. 26).

Green, Barrett N, Stephanie B Jones, Randal D Streck, et al. (1994). “Distinct expression patterns of insulin-like growth factor binding proteins 2 and 5 during fetal and postnatal development”. In: *Endocrinology* 134.2, pp. 954–962 (cit. on p. 64).

Green, Michael D, Antonia Chen, Maria-Cristina Nostro, et al. (2011). “Generation of anterior foregut endoderm from human embryonic and induced pluripotent stem cells”. In: *Nature biotechnology* 29.3, pp. 267–272 (cit. on p. 88).

Gualdi, Rossana, Pascale Bossard, Minghua Zheng, et al. (1996). “Hepatic specification of the gut endoderm in vitro: cell signaling and transcriptional control.” In: *Genes & development* 10.13, pp. 1670–1682 (cit. on p. 16).

Guiu, Jordi and Kim B Jensen (2015). “From definitive endoderm to gut—a process of growth and maturation”. In: *Stem cells and development* 24.17, pp. 1972–1983 (cit. on pp. 16, 87).

Hamidi, Sofiane, Yukiko Nakaya, Hiroki Nagai, et al. (2020). “Mesenchymal-epithelial transition regulates initiation of pluripotency exit before gastrulation”. In: *Development* 147.3, dev184960 (cit. on p. 35).

Han, Lu, Praneet Chaturvedi, Keishi Kishimoto, et al. (2020). “Single cell transcriptomics identifies a signaling network coordinating endoderm and mesoderm diversification during foregut organogenesis”. In: *Nature communications* 11.1, pp. 1–16 (cit. on pp. 15–17, 92).

Hao, Yuhan, Stephanie Hao, Erica Andersen-Nissen, et al. (2021). “Integrated analysis of multimodal single-cell data”. In: *Cell* (cit. on p. 77).

Hara, Kenshiro, Masami Kanai-Azuma, Mami Uemura, et al. (2009). “Evidence for crucial role of hindgut expansion in directing proper migration of primordial germ cells in mouse early embryogenesis”. In: *Developmental biology* 330.2, pp. 427–439 (cit. on p. 61).

Harrison, Sarah Ellys, Berna Sozen, Neophytos Christodoulou, Christos Kyprianou, and Magdalena Zernicka-Goetz (2017). “Assembly of embryonic and extraembryonic stem cells to mimic embryogenesis in vitro”. In: *Science* 356.6334 (cit. on p. 19).

Hashimoto, Koichiro and Norio Nakatsuji (1989). “Formation of the Primitive Streak and Mesoderm Cells in Mouse Embryos—Detailed Scanning Electron Microscopical Study: (primitive streak/cell migration/extracellular matrix/mouse gastrulation/scanning electron microscopy)”. In: *Development, growth & differentiation* 31.3, pp. 209–218 (cit. on p. 5).

Hashmi, Ali, Sham Tlili, Pierre Perrin, Alfonso Martinez-Arias, and Pierre-François Lenne (2020). “Cell-state transitions and collective cell movement generate an endoderm-like region in gastruloids”. In: *BioRxiv* (cit. on pp. 22, 83, 84, 86, 91).

Hebrok, Matthias, Seung K Kim, and Douglas A Melton (1998). “Notochord repression of endodermal Sonic hedgehog permits pancreas development”. In: *Genes & development* 12.11, pp. 1705–1713 (cit. on p. 17).

Hiramatsu, Ryuji, Toshiki Matsuoka, Chiharu Kimura-Yoshida, et al. (2013). “External mechanical cues trigger the establishment of the anterior-posterior axis in early mouse embryos”. In: *Developmental cell* 27.2, pp. 131–144 (cit. on p. 5).

Hooper, Martin, Kate Hardy, Alan Handyside, Susan Hunter, and Marilyn Monk (1987). “HPRT-deficient (Lesch–Nyhan) mouse embryos derived from germline colonization by cultured cells”. In: *Nature* 326.6110, pp. 292–295 (cit. on p. 54).

Hsu, Yu-Chih (1979). “In vitro development of individually cultured whole mouse embryos from blastocyst to early somite stage”. In: *Developmental biology* 68.2, pp. 453–461 (cit. on p. 18).

Hughes, Chris S, Lynne M Postovit, and Gilles A Lajoie (2010). “Matrigel: a complex protein mixture required for optimal growth of cell culture”. In: *Proteomics* 10.9, pp. 1886–1890 (cit. on p. 20).

Imuta, Yu, Hiroshi Kiyonari, Chuan-Wei Jang, Richard R Behringer, and Hiroshi Sasaki (2013). “Generation of knock-in mice that express nuclear enhanced green fluorescent protein and tamoxifen-inducible Cre recombinase in the notochord from Foxa2 and T loci”. In: *genesis* 51.3, pp. 210–218 (cit. on p. 8).

Ivanovitch, Kenzo, Pablo Soro-Barrio, Probir Chakravarty, et al. (2021). “Ventricular, atrial, and outflow tract heart progenitors arise from spatially and molecularly distinct regions of the primitive streak”. In: *PLoS biology* 19.5, e3001200 (cit. on pp. 7, 26).

Joaquin, Nick (2017). *The Woman Who Had Two Navels and Tales of the Tropical Gothic*. Penguin (cit. on p. 23).

Kanai-Azuma, Masami, Yoshiakira Kanai, Jacqueline M Gad, et al. (2002). “Depletion of definitive gut endoderm in Sox17-null mutant mice”. In: *Development* 129.10, pp. 2367–2379 (cit. on pp. 8, 11, 35, 60, 86).

Katsamba, P, K Carroll, G Ahlsen, et al. (2009). “Linking molecular affinity and cellular specificity in cadherin-mediated adhesion”. In: *Proceedings of the National Academy of Sciences* 106.28, pp. 11594–11599 (cit. on p. 10).

Kaufman, Matthew H and Jonathan BL Bard (1999). *The anatomical basis of mouse development*. Gulf Professional Publishing (cit. on p. 12).

Kearns, Nicola A, Ryan MJ Genga, Michael Ziller, et al. (2013). “Generation of organized anterior foregut epithelia from pluripotent stem cells using small molecules”. In: *Stem cell research* 11.3, pp. 1003–1012 (cit. on p. 88).

Khoa, Le Tran Phuc, Takuya Azami, Tomoyuki Tsukiyama, et al. (2016). “Visualization of the epiblast and visceral endodermal cells using Fgf5-P2A-Venus BAC transgenic mice and epiblast stem cells”. In: *PLoS One* 11.7, e0159246 (cit. on p. 64).

Kishimoto, Keishi, Kana T Furukawa, Agustin Luz-Madrigal, et al. (2020). “Bidirectional Wnt signaling between endoderm and mesoderm confers tracheal identity in mouse and human cells”. In: *Nature communications* 11.1, pp. 1–12 (cit. on p. 16).

Kispert, Andreas and Bernhard G Herrmann (1994). “Immunohistochemical analysis of the Brachyury protein in wild-type and mutant mouse embryos”. In: *Developmental biology* 161.1, pp. 179–193 (cit. on pp. 20, 24, 26, 60, 67).

Kubo, Atsushi, Katsunori Shinozaki, John M Shannon, et al. (2004). “Development of definitive endoderm from embryonic stem cells in culture”. In: (cit. on p. 6).

Kwon, Gloria S and Anna-Katerina Hadjantonakis (2009). “Transthyretin mouse transgenes direct RFP expression or Cre-mediated recombination throughout the visceral endoderm”. In: *genesis* 47.7, pp. 447–455 (cit. on p. 32).

Kwon, Gloria S, Manuel Viotti, and Anna-Katerina Hadjantonakis (2008). “The endoderm of the mouse embryo arises by dynamic widespread intercalation of embryonic and extraembryonic lineages”. In: *Developmental cell* 15.4, pp. 509–520 (cit. on pp. 4–6, 11, 12, 32, 59).

Lau, Caroline GY and Yusuke Marikawa (2014). “Morphology-based mammalian stem cell tests reveal potential developmental toxicity of donepezil”. In: *Molecular reproduction and development* 81.11, pp. 994–1008 (cit. on p. 19).

Lawson, Kirstie A, Juanito J Meneses, and Roger A Pedersen (1986). “Cell fate and cell lineage in the endoderm of the presomite mouse embryo, studied with an intracellular tracer”. In: *Developmental biology* 115.2, pp. 325–339 (cit. on pp. 11, 12).

Lawson, KIRSTIE A and RA Pedersen (1987). “Cell fate, morphogenetic movement and population kinetics of embryonic endoderm at the time of germ layer formation in the mouse”. In: *Development* 101.3, pp. 627–652 (cit. on p. 11).

Lewis, Samara L and Patrick PL Tam (2006). “Definitive endoderm of the mouse embryo: formation, cell fates, and morphogenetic function”. In: *Developmental dynamics: an official publication of the American Association of Anatomists* 235.9, pp. 2315–2329 (cit. on pp. 2, 6–8, 12, 13, 15).

Li, Aileen SW and Yusuke Marikawa (2015). “An in vitro gastrulation model recapitulates the morphogenetic impact of pharmacological inhibitors of developmental signaling pathways”. In: *Molecular reproduction and development* 82.12, pp. 1015–1036 (cit. on p. 19).

Li, Lin-Chen, Xin Wang, Zi-Ran Xu, et al. (2021). “Single-cell patterning and axis characterization in the murine and human definitive endoderm”. In: *Cell Research* 31.3, pp. 326–344 (cit. on p. 15).

Libby, Ashley RG, David A Joy, Nicholas H Elder, et al. (2021). “Axial elongation of caudalized human organoids mimics aspects of neural tube development”. In: *Development* 148.12, dev198275 (cit. on pp. 20, 22).

Lickert, Heiko, Stefanie Kutsch, Benoit Kanzler, et al. (2002). “Formation of multiple hearts in mice following deletion of beta-catenin in the embryonic endoderm”. In: *Developmental cell* 3.2, pp. 171–181 (cit. on p. 8).

Madabhushi, Mary and Elizabeth Lacy (2011). “Anterior visceral endoderm directs ventral morphogenesis and placement of head and heart via BMP2 expression”. In: *Developmental cell* 21.5, pp. 907–919 (cit. on pp. 12, 89).

Mantziou, Veronika, Peter Baillie-Benson, Manuela Jaklin, et al. (2021). “In vitro teratogenicity testing using a 3D, embryo-like gastruloid system”. In: *bioRxiv* (cit. on pp. 21, 22).

Marikawa, Yusuke, Dana Ann A Tamashiro, Toko C Fujita, and Vernadeth B Alarcón (2009). “Aggregated P19 mouse embryonal carcinoma cells as a simple in vitro model to study the molecular regulations of mesoderm formation and axial elongation morphogenesis”. In: *Genesis* 47.2, pp. 93–106 (cit. on pp. 19, 22).

Mashanov, Vladimir S, Olga Zueva, and José E García-Arrarás (2014). “Postembryonic organogenesis of the digestive tube: why does it occur in worms and sea cucumbers but fail in humans?” In: *Current topics in developmental biology* 108, pp. 185–216 (cit. on p. 1).

Matsuo, Isao and Ryuji Hiramatsu (2017). “Mechanical perspectives on the anterior-posterior axis polarization of mouse implanted embryos”. In: *Mechanisms of development* 144, pp. 62–70 (cit. on p. 5).

McBurney, MW (2003). “P19 embryonal carcinoma cells.” In: *International journal of developmental biology* 37.1, pp. 135–140 (cit. on p. 19).

McClay, David R, Jenifer C Croce, and Jacob F Warner (2021). “Conditional specification of endomesoderm”. In: *Cells & development*, p. 203716 (cit. on p. 7).

McGrath, Patrick S and James M Wells (2015). “SnapShot: GI tract development”. In: *Cell* 161.1, 176–176e171 (cit. on p. 6).

Minn, Kyaw Thu, Yuheng C Fu, Shenghua He, et al. (2020). “High-resolution transcriptional and morphogenetic profiling of cells from micropatterned human ESC gastruloid cultures”. In: *Elife* 9, e59445 (cit. on p. 10).

Mittnenzweig, Markus, Yoav Mayshar, Saifeng Cheng, et al. (2021). “A single-embryo, single-cell time-resolved model for mouse gastrulation”. In: *Cell* 184.11, pp. 2825–2842 (cit. on p. 8).

Monaghan, A Paula, Klaus H Kaestner, Evelyn Grau, and G Schutz (1993). “Postimplantation expression patterns indicate a role for the mouse forkhead/HNF-3 alpha, beta and gamma genes in determination of the definitive endoderm, chordamesoderm and neuroectoderm”. In: *Development* 119.3, pp. 567–578 (cit. on p. 26).

Morgani, Sophie M, Jakob J Metzger, Jennifer Nichols, Eric D Siggia, and Anna-Katerina Hadjantonakis (2018). “Micropattern differentiation of mouse pluripotent stem cells recapitulates embryo regionalized cell fate patterning”. In: *Elife* 7, e32839 (cit. on pp. 18, 19).

Moris, Naomi, Kerim Anlas, Susanne C van den Brink, et al. (2020). “An in vitro model of early anteroposterior organization during human development”. In: *Nature* 582.7812, pp. 410–415 (cit. on pp. 20, 22).

Mou, Hongmei, Rui Zhao, Richard Sherwood, et al. (2012). “Generation of multipotent lung and airway progenitors from mouse ESCs and patient-specific cystic fibrosis iPSCs”. In: *Cell stem cell* 10.4, pp. 385–397 (cit. on p. 88).

Mountford, Peter, Branko Zevnik, A Düwel, et al. (1994). “Dicistronic targeting constructs: reporters and modifiers of mammalian gene expression”. In: *Proceedings of the National Academy of Sciences* 91.10, pp. 4303–4307 (cit. on pp. 40, 54, 73).

Muramatsu, Takashi (2017). “Embryoglycan: a highly branched poly-N-acetyllactosamine in pluripotent stem cells and early embryonic cells”. In: *Glycoconjugate journal* 34.6, pp. 701–712 (cit. on p. 33).

Nash, Rodney, Lori Neves, Renate Faast, Michael Pierce, and Stephen Dalton (2007). “The lectin Dolichos biflorus agglutinin recognizes glycan epitopes on the surface of murine embryonic stem cells: a new tool for characterizing pluripotent cells and early differentiation”. In: *Stem Cells* 25.4, pp. 974–982 (cit. on p. 33).

Neubüser, Annette, Haruhiko Koseki, and Rudi Balling (1995). “Characterization and developmental expression of Pax9, a paired-box-containing gene related to Pax1”. In: *Developmental biology* 170.2, pp. 701–716 (cit. on pp. 70, 88).

Niessen, Carien M and Cara J Gottardi (2008). “Molecular components of the adherens junction”. In: *Biochimica et Biophysica Acta (BBA)-Biomembranes* 1778.3, pp. 562–571 (cit. on p. 35).

Ninomiya, Hiromasa, Robert David, Erich W Damm, et al. (2012). “Cadherin-dependent differential cell adhesion in Xenopus causes cell sorting in vitro but not in the embryo”. In: *Journal of Cell Science* 125.8, pp. 1877–1883 (cit. on p. 10).

Noguchi, Motoko, Takehiko Noguchi, Makoto Watanabe, and Takashi Muramatsu (1982). “Localization of receptors for Dolichos biflorus agglutinin in early post implantation embryos in mice.” In: (cit. on p. 33).

Nose, Akinao, Akira Nagafuchi, and Masatoshi Takeichi (1988). “Expressed recombinant cadherins mediate cell sorting in model systems”. In: *Cell* 54.7, pp. 993–1001 (cit. on p. 10).

Nowotschin, Sonja and Anna-Katerina Hadjantonakis (2020). “Guts and gastrulation: Emergence and convergence of endoderm in the mouse embryo”. In: *Current topics in developmental biology*. Vol. 136. Elsevier, pp. 429–454 (cit. on pp. 7–9, 85).

Nowotschin, Sonja, Anna-Katerina Hadjantonakis, and Kyra Campbell (2019a). “The endoderm: a divergent cell lineage with many commonalities”. In: *Development* 146.11, dev150920 (cit. on pp. 2, 3, 5, 6, 9–11, 15, 90).

Nowotschin, Sonja, Manu Setty, Ying-Yi Kuo, et al. (2019b). “The emergent landscape of the mouse gut endoderm at single-cell resolution”. In: *Nature* 569.7756, pp. 361–367 (cit. on pp. 5, 11, 12, 15, 17, 32, 59, 63–65, 67, 69, 71, 78, 89, 90).

Ober, Elke A and Anne Grapin-Botton (2015). “At new heights–endodermal lineages in development and disease”. In: *Development* 142.11, pp. 1912–1917 (cit. on p. 2).

Ogaki, Soichiro, Seiko Harada, Nobuaki Shiraki, Kazuhiko Kume, and Shoen Kume (2011). “An expression profile analysis of ES cell-derived definitive endodermal cells and Pdx1-expressing cells”. In: *BMC developmental biology* 11.1, pp. 1–15 (cit. on p. 63).

Olmsted, Zachary T and Janet L Paluh (2021). “Co-development of central and peripheral neurons with trunk mesendoderm in human elongating multi-lineage organized gastruloids”. In: *Nature communications* 12.1, pp. 1–19 (cit. on pp. 20, 22, 84, 87).

Oppenheimer, Jane M and Viktor Hamburger (1976). “The non-specificity of the germ-layers”. In: *The Quarterly Review of Biology* 51, pp. 96–124 (cit. on p. 1).

Pander, Christian Heinrich (1817). *Dissertatio inauguralis sistens historiam metamorphoseos, quam ovum incubatum prioribus quinque diebus subit*. Nitribitt (cit. on p. 1).

Peters, Heiko, Annette Neubüser, Klaus Kratochwil, and Rudi Balling (1998). “Pax9-deficient mice lack pharyngeal pouch derivatives and teeth and exhibit craniofacial and limb abnormalities”. In: *Genes & development* 12.17, pp. 2735–2747 (cit. on pp. 70, 88).

Pfendler, Kristina C, Carmina S Catuar, Juanito J Meneses, and Roger A Pedersen (2005). “Overexpression of Nodal promotes differentiation of mouse embryonic stem cells into mesoderm and endoderm at the expense of neuroectoderm formation”. In: *Stem cells and development* 14.2, pp. 162–172 (cit. on p. 6).

Pijuan-Sala, Blanca, Jonathan A Griffiths, Carolina Guibentif, et al. (2019). “A single-cell molecular map of mouse gastrulation and early organogenesis”. In: *Nature* 566.7745, pp. 490–495 (cit. on pp. 11, 12, 59).

Ponz-Segrelles, Guillermo, Christopher J Glasby, Conrad Helm, et al. (2021). “Integrative anatomical study of the branched annelid Ramisyllis multicaudata (Annelida, Syllidae)”. In: *Journal of Morphology* 282.6, pp. 900–916 (cit. on p. 1).

Pour, Maayan, Abhishek Sampath Kumar, Naama Farag, et al. (2019). “Emergence and patterning dynamics of mouse definitive endoderm”. In: *Naama and Walther, Maria and Wittler, Lars and Meissner, Alexander and Nachman, Iftach, Emergence and Patterning Dynamics of Mouse Definitive Endoderm* (cit. on pp. 83, 86).

Preibisch, Stephan, Stephan Saalfeld, and Pavel Tomancak (2009). "Globally optimal stitching of tiled 3D microscopic image acquisitions". In: *Bioinformatics* 25.11, pp. 1463–1465 (cit. on pp. 42, 56, 75).

Probst, Simone, Jelena Tosic, Carsten Schwan, Dominic Grun, Sebastian J Arnold, et al. (2021). "Spatiotemporal sequence of mesoderm and endoderm lineage segregation during mouse gastrulation". In: *Development* 148.1 (cit. on pp. 4–10, 82, 89).

Ramkumar, Nitya, Tatiana Omelchenko, Nancy F Silva-Gagliardi, et al. (2016). "Crumbs2 promotes cell ingression during the epithelial-to-mesenchymal transition at gastrulation". In: *Nature cell biology* 18.12, pp. 1281–1291 (cit. on p. 10).

Rankin, Scott A, Jeffrey D Steimle, Xinan H Yang, et al. (2021). "TBX5 drives Aldh1a2 expression to regulate a RA-Hedgehog-Wnt gene regulatory network coordinating cardiopulmonary development". In: *bioRxiv* (cit. on p. 16).

Richardson, Lorna, Shanmugasundaram Venkataraman, Peter Stevenson, et al. (2014). "EMAGE mouse embryo spatial gene expression database: 2014 update". In: *Nucleic acids research* 42.D1, pp. D835–D844 (cit. on p. 2).

Rodaway, Adam and Roger Patient (2001). "Mesendoderm: an ancient germ layer?" In: *Cell* 105.2, pp. 169–172 (cit. on p. 7).

Rossi, Giuliana, Nicolas Broguiere, Matthew Miyamoto, et al. (2021a). "Capturing cardiogenesis in gastruloids". In: *Cell stem cell* 28.2, pp. 230–240 (cit. on pp. 16, 20, 22, 32, 47, 63, 65, 67, 71, 77, 78, 81, 92, 93, 102).

Rossi, Giuliana, Sonja Giger, Tania Hübscher, and Matthias P Lutolf (2021b). "Gastruloids as in vitro models of embryonic blood development with spatial and temporal resolution". In: *bioRxiv* (cit. on pp. 20, 22).

Rossi, Jennifer M, N Ray Dunn, Brigid LM Hogan, and Kenneth S Zaret (2001). "Distinct mesodermal signals, including BMPs from the septum transversum mesenchyme, are required in combination for hepatogenesis from the endoderm". In: *Genes & development* 15.15, pp. 1998–2009 (cit. on p. 16).

Rueden, Curtis T, Johannes Schindelin, Mark C Hiner, et al. (2017). "ImageJ2: ImageJ for the next generation of scientific image data". In: *BMC bioinformatics* 18.1, pp. 1–26 (cit. on pp. 42, 56, 75).

Sasaki, Hiroshi and BL Hogan (1993). "Differential expression of multiple fork head related genes during gastrulation and axial pattern formation in the mouse embryo". In: *Development* 118.1, pp. 47–59 (cit. on p. 47).

Saykali, Bechara, Navrita Mathiah, Wallis Nahaboo, et al. (2019). "Distinct mesoderm migration phenotypes in extra-embryonic and embryonic regions of the early mouse embryo". In: *Elife* 8, e42434 (cit. on pp. 5, 11).

Scarpa, Elena, András Szabó, Anne Bibonne, et al. (2015). “Cadherin switch during EMT in neural crest cells leads to contact inhibition of locomotion via repolarization of forces”. In: *Developmental cell* 34.4, pp. 421–434 (cit. on p. 11).

Scheibner, Katharina, Silvia Schirge, Ingo Burtscher, et al. (2021). “Epithelial cell plasticity drives endoderm formation during gastrulation”. In: *Nature Cell Biology* (cit. on pp. 6, 7, 9–11, 42, 82, 86, 89, 90).

Schindelin, Johannes, Ignacio Arganda-Carreras, Erwin Frise, et al. (2012). “Fiji: an open-source platform for biological-image analysis”. In: *Nature methods* 9.7, pp. 676–682 (cit. on pp. 42, 56, 75).

Shahbazi, Marta N and Magdalena Zernicka-Goetz (2018). “Deconstructing and reconstructing the mouse and human early embryo”. In: *Nature cell biology* 20.8, pp. 878–887 (cit. on p. 18).

Sherwood, Richard I, Cristian Jitianu, Ondine Cleaver, et al. (2007). “Prospective isolation and global gene expression analysis of definitive and visceral endoderm”. In: *Developmental biology* 304.2, pp. 541–555 (cit. on pp. 15, 17, 33).

Silva, Ana C, Oriane B Matthys, David A Joy, et al. (2020). “Developmental co-emergence of cardiac and gut tissues modeled by human iPSC-derived organoids”. In: *bioRxiv* (cit. on p. 16).

Silva, Jose, Ornella Barrandon, Jennifer Nichols, et al. (2008). “Promotion of reprogramming to ground state pluripotency by signal inhibition”. In: *PLoS biology* 6.10, e253 (cit. on p. 23).

Simunovic, Mijo and Ali H Brivanlou (2017). “Embryoids, organoids and gastruloids: new approaches to understanding embryogenesis”. In: *Development* 144.6, pp. 976–985 (cit. on p. 18).

Smith, LJ (1985). “Embryonic axis orientation in the mouse and its correlation with blastocyst relationships to the uterus: II. Relationships from $41/4$ to $91/2$ days”. In: (cit. on p. 4).

Sozen, Berna, Gianluca Amadei, Andy Cox, et al. (2018). “Self-assembly of embryonic and two extra-embryonic stem cell types into gastrulating embryo-like structures”. In: *Nature cell biology* 20.8, pp. 979–989 (cit. on pp. 18, 19).

Spence, Jason R, Alex W Lange, Suh-Chin J Lin, et al. (2009). “Sox17 regulates organ lineage segregation of ventral foregut progenitor cells”. In: *Developmental cell* 17.1, pp. 62–74 (cit. on p. 60).

Spence, Jason R, Ryan Lauf, and Noah F Shroyer (2011). “Vertebrate intestinal endoderm development”. In: *Developmental dynamics* 240.3, pp. 501–520 (cit. on pp. 6, 12).

Steventon, Benjamin, Lara Busby, and Alfonso Martinez Arias (2021). “Establishment of the vertebrate body plan: Rethinking gastrulation through stem cell models of

early embryogenesis". In: *Developmental Cell* 56.17, pp. 2405–2418 (cit. on pp. 2, 18, 81, 90).

Stower, Matthew J and Shankar Srinivas (2014). "Heading forwards: anterior visceral endoderm migration in patterning the mouse embryo". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 369.1657, p. 20130546 (cit. on p. 4).

Stringer, Carsen, Tim Wang, Michalis Michaelos, and Marius Pachitariu (2021). "Cellpose: a generalist algorithm for cellular segmentation". In: *Nature Methods* 18.1, pp. 100–106 (cit. on pp. 28, 43, 76).

Sui, Lina, Luc Bouwens, and Josué K Mfopou (2013). "Signaling pathways during maintenance and definitive endoderm differentiation of embryonic stem cells". In: *International Journal of Developmental Biology* 57.1, pp. 1–12 (cit. on pp. 8, 85).

Sulston, John E, Einhard Schierenberg, John G White, and J Nichol Thomson (1983). "The embryonic cell lineage of the nematode Caenorhabditis elegans". In: *Developmental biology* 100.1, pp. 64–119 (cit. on p. 6).

Švajger, A and B Levak-Švajger (1974). "Regional developmental capacities of the rat embryonic endoderm at the head-fold stage". In: *Development* 32.2, pp. 461–467 (cit. on p. 15).

Tada, Shinsuke, Takumi Era, Chikara Furusawa, et al. (2005). "Characterization of mesendoderm: a diverging point of the definitive endoderm and mesoderm in embryonic stem cell differentiation culture". In: (cit. on p. 6).

Takaoka, Katsuyoshi and Hiroshi Hamada (2012). "Cell fate decisions and axis determination in the early mouse embryo". In: *Development* 139.1, pp. 3–14 (cit. on p. 4).

Tam, Patrick PL and Richard R Behringer (1997). "Mouse gastrulation: the formation of a mammalian body plan". In: *Mechanisms of development* 68.1-2, pp. 3–25 (cit. on pp. 4, 5, 24).

Tam, Patrick PL, Poh-Lynn Khoo, Samara L Lewis, et al. (2007). "Sequential allocation and global pattern of movement of the definitive endoderm in the mouse embryo during gastrulation". In: (cit. on p. 11).

Tam, Patrick PL, Poh-Lynn Khoo, Nicole Wong, Tania E Tsang, and Richard R Behringer (2004). "Regionalization of cell fates and cell movement in the endoderm of the mouse gastrula and the impact of loss of Lhx1 (Lim1) function". In: *Developmental biology* 274.1, pp. 171–187 (cit. on p. 12).

Tam, Patrick PL and David AF Loebel (2007). "Gene function in mouse embryogenesis: get set for gastrulation". In: *Nature Reviews Genetics* 8.5, pp. 368–381 (cit. on pp. 4, 5, 24).

Tan, Yuqi, Abby Spangler, Michael Farid, Da Peng, and Patrick Cahan (2021). “Quantitative comparison of in vitro and in vivo embryogenesis at a single cell resolution”. In: *BioRxiv* (cit. on pp. 63, 68).

Teng, Teng, Camilla Teng, Vesa Kaartinen, and Jeffrey O Bush (2021). “A unique form of collective epithelial migration is crucial for tissue fusion in the secondary palate and can overcome loss of epithelial apoptosis”. In: *bioRxiv* (cit. on p. 87).

Thomas, PQ, A Brown, and RS Beddington (1998). “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, pp. 85–94 (cit. on p. 63).

Thowfeequ, Shifaan, Jonathan Fiorentino, Di Hu, et al. (2021). “Characterisation of the transcriptional dynamics underpinning the function, fate, and migration of the mouse Anterior Visceral Endoderm”. In: *bioRxiv* (cit. on p. 4).

Timmer, John R, Tracy W Mak, Katia Manova, Kathryn V Anderson, and Lee Niswander (2005). “Tissue morphogenesis and vascular stability require the Frem2 protein, product of the mouse myelencephalic blebs gene”. In: *Proceedings of the National Academy of Sciences* 102.33, pp. 11746–11750 (cit. on p. 64).

Toda, Satoshi, Lucas R Blauch, Sindy KY Tang, Leonardo Morsut, and Wendell A Lim (2018). “Programming self-organizing multicellular structures with synthetic cell-cell signaling”. In: *Science* 361.6398, pp. 156–162 (cit. on pp. 10, 91).

Torres-Paz, Jorge and Sylvie Rétaux (2021). “Pescoids and chimeras to probe early evo-devo in the fish Astyanax mexicanus”. In: *Frontiers in cell and developmental biology* 9, p. 927 (cit. on p. 21).

Townes, Philip L and Johannes Holtfreter (1955). “Directed movements and selective adhesion of embryonic amphibian cells”. In: *Journal of experimental zoology* 128.1, pp. 53–120 (cit. on p. 10).

Tremblay, Kimberly D and Kenneth S Zaret (2005). “Distinct populations of endoderm cells converge to generate the embryonic liver bud and ventral foregut tissues”. In: *Developmental biology* 280.1, pp. 87–99 (cit. on p. 12).

Turner, David A, Peter Baillie-Johnson, and Alfonso Martinez Arias (2016). “Organoids and the genetically encoded self-assembly of embryonic stem cells”. In: *BioEssays* 38.2, pp. 181–191 (cit. on pp. 18, 32).

Turner, David A, Mehmet Girgin, Luz Alonso-Crisostomo, et al. (2017). “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, pp. 3894–3906 (cit. on pp. 18–25, 32, 35, 82, 83, 92).

Tzouanacou, Elena, Amélie Wegener, Filip J Wymeersch, Valerie Wilson, and Jean-François Nicolas (2009). “Redefining the progression of lineage segregations during mammalian embryogenesis by clonal analysis”. In: *Developmental cell* 17.3, pp. 365–376 (cit. on p. 7).

Vaca-Merino, Víctor, Ruth Maldonado-Rengel, Cristopher Nicholson, and Mariano del Sol (2021). "Uso del Lexema derma en Terminologia Embryologica". In: *International Journal of Morphology* 39.1, pp. 231–234 (cit. on p. 1).

Van, Phu, Wenxin Jiang, Raphael Gottardo, and Greg Finak (2018). "ggCyto: next generation open-source visualization software for cytometry". In: *Bioinformatics* 34.22, pp. 3951–3953 (cit. on p. 57).

Veenvliet, Jesse V, Adriano Bolondi, Helene Kretzmer, et al. (2020). "Mouse embryonic stem cells self-organize into trunk-like structures with neural tube and somites". In: *Science* 370.6522 (cit. on pp. 20, 22, 47, 60, 81, 84, 87, 91, 92).

Vianello, Stefano and Matthias P Lutolf (2019). "Understanding the mechanobiology of early mammalian development through bioengineered models". In: *Developmental cell* 48.6, pp. 751–763 (cit. on p. 18).

Vianello, Stefano Davide (2020). "Exploring and illustrating the mouse embryo: virtual objects to think and create with". In: *bioRxiv* (cit. on pp. 2, 113).

Vianello, Stefano Davide and Matthias Lutolf (2020). "In vitro endoderm emergence and self-organisation in the absence of extraembryonic tissues and embryonic architecture". In: *BioRxiv* (cit. on pp. 3, 22, 25, 27, 36, 38, 40, 47, 51, 53, 54, 63, 65, 67, 69, 71, 73).

Viotti, Manuel, Sonja Nowotschin, and Anna-Katerina Hadjantonakis (2014). "SOX17 links gut endoderm morphogenesis and germ layer segregation". In: *Nature cell biology* 16.12, pp. 1146–1156 (cit. on pp. 4–11, 26, 30, 32, 35, 60, 86, 90).

Voiculescu, Octavian, Lawrence Bodenstein, I-Jun Lau, and Claudio D Stern (2014). "Local cell interactions and self-amplifying individual cell ingression drive amniote gastrulation". In: *Elife* 3, e01817 (cit. on p. 10).

Waisman, Ariel, Alessandra Marie Norris, Martín Elías Costa, and Daniel Kopinke (2021). "Automatic and unbiased segmentation and quantification of myofibers in skeletal muscle". In: *Scientific reports* 11.1, pp. 1–14 (cit. on pp. 29, 43, 76).

Wallin, Johan, Hermann Eibel, A Neubuser, et al. (1996). "Pax1 is expressed during development of the thymus epithelium and is required for normal T-cell maturation". In: *Development* 122.1, pp. 23–30 (cit. on pp. 70, 88).

Walton, Katherine D, Andrew M Freddo, Sha Wang, and Deborah L Gumucio (2016). "Generation of intestinal surface: an absorbing tale". In: *Development* 143.13, pp. 2261–2272 (cit. on p. 87).

Wang, Lu and Ye-Guang Chen (2016). "Signaling control of differentiation of embryonic stem cells toward mesendoderm". In: *Journal of molecular biology* 428.7, pp. 1409–1422 (cit. on pp. 8, 85).

Wei, Qiaozhi and Brian G Condie (2011). "A focused in situ hybridization screen identifies candidate transcriptional regulators of thymic epithelial cell development and function". In: *PLoS One* 6.11, e26795 (cit. on p. 67).

Wells, James M and Douglas A Melton (2000). “Early mouse endoderm is patterned by soluble factors from adjacent germ layers”. In: *Development* 127.8, pp. 1563–1572 (cit. on pp. 16–18, 89).

– (1999). “Vertebrate endoderm development”. In: *Annual review of cell and developmental biology* 15.1, pp. 393–410 (cit. on p. 6).

Wheelock, Margaret J, Yasushi Shintani, Masato Maeda, Yuri Fukumoto, and Keith R Johnson (2008). “Cadherin switching”. In: *Journal of cell science* 121.6, pp. 727–735 (cit. on p. 9).

Williams, Margot, Carol Burdsal, Ammasi Periasamy, Mark Lewandoski, and Ann Sutherland (2012). “Mouse primitive streak forms in situ by initiation of epithelial to mesenchymal transition without migration of a cell population”. In: *Developmental Dynamics* 241.2, pp. 270–283 (cit. on p. 5).

Wood, Heather B and Vasso Episkopou (1999). “Comparative expression of the mouse Sox1, Sox2 and Sox3 genes from pre-gastrulation to early somite stages”. In: *Mechanisms of development* 86.1-2, pp. 197–201 (cit. on pp. 47, 67).

Xu, Peng-Fei, Ricardo Moraes Borges, Jonathan Fillatre, et al. (2021). “Construction of a mammalian embryo model from stem cells organized by a morphogen signalling centre”. In: *Nature communications* 12.1, pp. 1–22 (cit. on pp. 32, 42, 84, 86, 87).

Yamanaka, Yojiro, Owen J Tamplin, Anja Beckers, Achim Gossler, and Janet Rossant (2007). “Live imaging and genetic analysis of mouse notochord formation reveals regional morphogenetic mechanisms”. In: *Developmental cell* 13.6, pp. 884–896 (cit. on p. 60).

Yang, Dapeng (2015). “Novel mechanisms of endoderm and mesoderm formation”. PhD thesis. Technische Universität München (cit. on pp. 51, 54).

Ye, Shoudong, Li Tan, Rongqing Yang, et al. (2012). “Pleiotropy of glycogen synthase kinase-3 inhibition by CHIR99021 promotes self-renewal of embryonic stem cells from refractory mouse strains”. In: *PloS one* 7.4, e35892 (cit. on p. 24).

Zhang, Hui Ting and Takashi Hiiragi (2018). “Symmetry breaking in the mammalian embryo”. In: *Annual review of cell and developmental biology* 34, pp. 405–426 (cit. on p. 4).

Zhang, Zhen, Fabiana Cerrato, Huansheng Xu, et al. (2005). “Tbx1 expression in pharyngeal epithelia is necessary for pharyngeal arch artery development”. In: (cit. on p. 67).

Zhuang, Shaowei, Qingquan Zhang, Tao Zhuang, et al. (2013). “Expression of Isl1 during mouse development”. In: *Gene Expression Patterns* 13.8, pp. 407–412 (cit. on p. 64).

Zorn, Aaron M and James M Wells (2009). “Vertebrate endoderm development and organ formation”. In: *Annual Review of Cell and Developmental* 25, pp. 221–251 (cit. on pp. 12, 59).

Licensing and distribution

STEFANO D. VIANELLO

ORCID:0000-0002-4152-8990

EPFL SV IBI-SV UPLUT, AI 1.112, 1015 Lausanne, CH
[+41 21 69]31614
stefano.vianello@epfl.ch
https://stefanovianello.github.io/

WHO AM I?

I am a PhD student with 6 years of experience in stem cell and mouse developmental biology, with expertise in embryonic stem cell culture, synthetic embryology, and bioengineering. My main academic interests resides in endoderm and gut tube development, as well as mechanobiology and mechanical inputs to development. I am an advocate for intersectional open science, preprints, and equitable publishing. Aside from my research projects, I am interested in data communication and visual storytelling in developmental biology. I resonate with knowledge equity movements of embodied knowledge and pedagogy.

EDUCATION

2017 – present **ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE (EPFL)** **Lausanne, CH**

- **PhD Doctoral Program** in Biotechnology and Bioengineering, with Teaching Assistantship
- Laboratory of Stem Cell Bioengineering, with Prof M. Lütolf
- "Endoderm development and morphogenesis in self-organising stem cell-based models of mouse embryogenesis"
- Swiss National Science Foundation Synergia Grant

2016-2017 **GIRTON COLLEGE, UNIVERSITY OF CAMBRIDGE** **Cambridge, UK**

- **MPhil in Biological Science (Genetics)**
- "Wnt and Notch (Wntch) interactions in in vitro models of preimplantation embryonic development", with Prof A. Martinez-Arias

2013 – 2016 **GIRTON COLLEGE, UNIVERSITY OF CAMBRIDGE** **Cambridge, UK**

- **BA Hons Natural Sciences**
- **3d year** (First Class Honours): Genetics (Human Genome, Genomics, SysBio; Developmental Genetics; Chromosomes & Cell cycle; Evolutionary Genetics; Plant & Microbial Genetics)
- **2nd year** (First Class Honours): Biochemistry and Molecular Biology, Cell and Developmental Biology, Pathology
- **1st year** (First Class Honours): Biology of Cells, Physiology of Organisms, Evolution and Behaviour, Mathematical Biology

PUBLICATIONS

doi: https://doi.org/10.1101/2020.06.07.138883
Vianello, Lutolf; "In vitro endoderm emergence and self-organisation in the absence of extraembryonic tissues and embryonic architecture", 2021

doi: https://doi.org/10.1101/2020.11.23.393991
Vianello; "Exploring and illustrating the mouse embryo: virtual objects to think and create with", 2020

PMID: 30913407
Vianello, Lutolf; "Understanding the Mechanobiology of Early Mammalian Development through Bioengineered Models", 2019

PUBLICATIONS (PROTOCOLS, DATA, SCRIPTS)

Protocol **doi: https://dx.doi.org/10.17504/protocols.io.bvgrn3v6**
Vianello; "Protocol to process Gastruloids for FACS". 2021

Protocol **doi: https://dx.doi.org/10.17504/protocols.io.9j5h4q6**
Vianello, Girgin, Rossi, Lutolf; "Protocol to generate Gastruloids (LSCB, EPFL)". 2020

Protocol

doi: https://dx.doi.org/10.17504/protocols.io.7tzhnp6

Vianello, Girgin, Rossi, Lutolf; "Protocol to immunostain Gastruloids (LSCB, EPFL)". 2020

Protocol

doi: https://dx.doi.org/10.17504/protocols.io.7xbhpin

Vianello, Girgin, Rossi, Lutolf; "Protocol to culture mESCs (LSCB, UPLUT)". 2020

Script

doi: https://doi.org/10.5281/zenodo.4894121

Vianello; "RNotebook FACS pipeline". 2021

Script

doi: https://doi.org/10.5281/zenodo.3884237

Vianello, Sanchez, Bercowsky-Rama, Lutolf; "Gastruloid Intensity Profiler". 2020

Dataset

doi: https://doi.org/10.5281/zenodo.4284379

Vianello; "3D models of mouse embryonic development". 2020

PUBLICATIONS (OTHER)

Published essays

https://doi.org/10.21428/6ffd8432.5de25622

Vianello; "The "Pre" in [my] "Preprint" is for Pre-figurative". Commonplace, Knowledge Futures Group.

PreLight posts

doi: https://doi.org/10.1242/prelights.29514

Vianello; ""English only please!" Real world effects of English-centred literature searches.". preLights, 2021

doi: https://doi.org/10.1242/prelights.29577

Vianello; "Bitter bites for a chocolate lover. Searching for RNAi targets to fight cocoa crop pests". preLights, 2021

doi: pending

Vianello*, Sanchez*; "Say "Aaaah!". Foregut development and toothed tongues in the black Katy chiton". preLights, 2021

doi: https://doi.org/10.1242/prelights.26953

Vianello*, Sanchez*; "Navigating change: sentinels of the sea tell about ocean health and disease.". preLights, 2021

doi: https://doi.org/10.1242/prelights.25860

Vianello*, Sanchez*; "Gastruloids, pescoids, caveoids, surfoids….In vitro embryonic models to study evo-eco-devo. New experimental approaches to cavefish development.". preLights, 2020

doi: https://doi.org/10.1242/prelights.16775

Vianello*, Sanchez*; "On the (h)edge: the germline precursors of a basal metazoa are induced at the interface between Hedgehog signalling domains". preLights, 2020

doi: https://doi.org/10.1242/prelights.9812

Vianello*, Sanchez*; "(Transiently) Comfortable in its own "skin": formation of epithelium-like multicellular structures in a unicellular organism through conserved actomyosin-dependent mechanisms". preLights, 2019

doi: https://doi.org/10.1242/prelights.6820

Vianello*, Sanchez*; "Mind the gap: epiblast geometry at its extraembryonic boundary constrains BMP localization and ensures robust gradient formation". preLights, 2019

Blogposts

tinyurl.com/yfsycuhb

"The "pre" in (my) "preprint" is for pre-figurative". September 2021

tinyurl.com/yeu28vja

"Preprint highlighting in a haunted house [Part 1]". August 2021

POSTERS

04/2020 **BIRS-CMO WORKSHOP: "MODELLING AND ENGINEERING THE MOUSE EMBRYO"** **Oaxaca, MX**

Cancelled due pandemic

09/2019 **EUROPEAN SUMMER SCHOOL ON STEM CELLS AND REGENERATIVE MEDICINE** **Hydra, EL**
"Studying the mechanobiology of early mammalian development using self-organising embryonic organoids"

07/2019 **EMBO SYMPOSIUM: MECHANICAL FORCES IN DEVELOPMENT** **Heidelberg, DE**
"Altering geometry and mechanics to coax the development of self-organising embryonic organoids"

03/2019 **EMBO SYMPOSIUM: SYNTHETIC MORPHOGENSIS** **Heidelberg, DE**
"Altering geometry and mechanics to coax the development of self-organising embryonic organoids"

03/2019 **EMBO WORKSHOP: VISUALIZING BIOLOGICAL DATA** **Heidelberg, DE**

- "Illustrating mouse development through 3D volumetric models" [https://vizbi.org/Posters/2019/D07]
- "Potential" (Art & Biology entry) [https://vizbi.org/Posters/2019/Y02]

TALKS

2021 **PREPRINTS IN MOTION PODCAST** **virtual**
Invited speaker: "The "pre" in preprint is for prefigurative. A conversation on prefigurative politics in academia"

2021 **ASAPbio FELLOWS TRAINING PROGRAMME** **virtual**
Invited speaker: "Preprint highlighting in a haunted house: Matthew effect, bias, preprint curation"

2019 **PHYSICS OF LIVING SYSTEMS (INTERNAL)** **Lausanne, CH**
"Squeezing, pressing, and bounding Gastruloids: mechanics and symmetry-breaking in vitro"

2019 **EUROTECH SUMMER SCHOOL: OPEN SCIENCE IN PRACTICE** **Lausanne, CH**
Invited speaker: "PreLights: a community-driven effort to highlighting preprints"

2019 **EMBO WORKSHOP: IMAGING MOUSE DEVELOPMENT** **Heidelberg, DE**
"Elucidating the role of mechanical cues during peri-implantation mouse development"

OTHER WORK AND RESEARCH EXPERIENCE

01–03/2016 **UNDERGRADUATE RESEARCH**
DEPARTMENT OF GENETICS, UNIVERSITY OF CAMBRIDGE **Cambridge, UK**

- Wnt/Notch interactions in the control of mouse pre- and post-implantation development"; with Prof A. Martinez-Arias
- Investigated the role of Wnt- and Notch-signalling in early mouse embryonic development, and highlighted possible β-catenin dependent, CSL-independent, non-canonical interactions between the two pathways. Confirmed canonical roles in axial elongation and axial determination.
- Used small molecule inhibitors to generate GOF and LOF backgrounds on which to investigate signalling epistasis; strengthened expertise in 2D and 3D tissue culture, mouse Embryonic Stem Cells, live imaging of gene expression, in vitro models of early development, data processing and analysis

08 –10/2015 **SUMMER RESEARCH**
DEPARTMENT OF GENETICS, UNIVERSITY OF CAMBRIDGE **Cambridge, UK**

- "Role of signalling in tissue specification in patterned aggregates of mouse Embryonic Stem cells"; with Prof A. Martinez-Arias
- Investigated the validity of 2i-maintained stem cell population as representative models of embryonic populations
- Awarded Wellcome Trust Biomedical Vacation Scholarship (7 weeks)
- Gained familiarity with 2D and 3D tissue culture, mouse Embryonic Stem Cells, live imaging of gene expression, in vitro models ofearly development (i.e. gastruloids), Fluorescence-Activated Cell Sorting, data processing and analysis

2014 – 2017

VOLUNTEERING
UNIVERSITY MUSEUM OF ZOOLOGY CAMBRIDGE **Cambridge, UK**

- Actively participated in the packing, transfer, relocation, and restoration of more than 500 specimens of the museum's collections to allow for the large-scale renovation of the building.
- Effectively utilized time-management skills to evenly distribute the academic workload around more than 90 hours of volunteering, in order to successfully complete all extra-curricular tasks while obtaining First Class Honours
- Devised and introduced innovative packing techniques to deal with particularly delicate specimens of high historical, scientific, and economical value

COMMUNITY INVOLVEMENT & RESPONSIBILITIES

2020 – 2021

PREPRINT CURATOR
THE COMPANY OF BIOLOGISTS

- Created and curated the PreList (list of preprints) dedicated to Gastruloids on the community site of the Company of Biologists [tinyurl.com/ydvv59or]

2019 – 2021

PRELIGHTS CONTRIBUTOR
THE COMPANY OF BIOLOGISTS

- Highlighted preprints posted on bioRxiv, in collaboration with Paul Gerald Layague Sanchez (EMBL Heidelberg)
- Articles are posted on the Node, the community site for Developmental Biology

2016 – 2017

PROJECT LEADER
CAMBRIDGE UNIVERSITY SYNTHETIC BIOLOGY SOCIETY (CUSBS) **Cambridge, UK**

- Designed and planned experimental synthetic-biology projects for students from both scientific and non-scientific backgrounds
- Currently coordinating and supervising 20 other society student members over a one-year practical project (design and introduction of synthetic genetic networks in cell-free transcription-translation systems)
- Collaborated with local community to establish a Biomakespace (innovation space for biology and biological engineering) in the old Laboratory of Molecular Biology, Cambridge;
- Liaised with University and Departmental staff for laboratory space and other administrative matters

2016

MEMBER
CAMBRIDGE UNIVERSITY SYNTHETIC BIOLOGY SOCIETY (CUSBS) **Cambridge, UK**

- Participated in the design, construction, and assembly of a scanning automated microscope for high-throughput screening of multiple biological samples
- Built a Computerised Numerical Control (CNC) machine, designed 3D-printed microscope components
- Participated in the integration of Raspberry Pi/Arduino electronics, optimisation of existing software for image acquisition, and production of comprehensive hardware documentation for open access and reproducibility

2016; 2017

VOLUNTEER
CAMBRIDGE SCIENCE FESTIVAL, DEPARTMENT OF GENETICS **Cambridge, UK**

- 2017: Designed and introduced in the festival a new stand dedicated to Developmental Biology (zebrafish). Introduced Virtual Reality and Augmented Reality as pedagogical tools for outreach and public engagement (virtual mouse development).
- 2016: Worked as part of a team to organise and supervise activities for both children and adults on the theme "Genes and Heredity" (building DNA models with paper or gummy bears, climbing competitions between Drosophila strains, etc…)
- Engaged with the public and presented scientific posters on Human Evolution and Comparative Genomics
- Introduced children to the field of Genetics and to the wonders of Drosophila development and mutant phenotypes

TEACHING TRAINING/EXPERIENCE

2021 **EPFL ENG-629: LECTURING AND PRESENTING IN ENGINEERING** **Lausanne, CH**

- Practical course/workshop with theoretical component
- Informed by contemporary research on teaching engineering, designed lesson plans and practiced teaching in small classroom-like setting
- Articulation of own teaching philosophy

2020 **EPFL ENG-624: SCIENCE & ENGINEERING TEACHING AND LEARNING** **Lausanne, CH**

- Introduction of research-informed approaches to teaching and learning, and concrete strategies appropriate for higher education science and technology contexts

2021 **TEACHING ASSISTANT**
EPFL BIO-378 PHYSIOLOGY LAB 1 **Lausanne, CH**

- Practical course in animal physiology for second year undergraduate students. Taught and facilitated experiments on measures of cellular oxygen consumption, in small group settings.

2018-2021 **TEACHING ASSISTANT**
EPFL BIOENG-110 GENERAL BIOLOGY **Lausanne, CH**

- Introductory biology course, led weekly exercise sessions for first year students

24,25/10/2019 **COURSE INSTRUCTOR**
CAMBIOSCIENCE PRACTICAL COURSE: DISCOVERING GASTRULOIDS **Cambridge, UK**

- Invited instructor: gave lectures to workshop participants about fundamentals of mammalian embryonic development and current in vitro models used to study it
- Demonstrated mouse embryonic stem cell culture and Gastruloid generation protocol during wet lab practical sessions

10/2016; 02/2017 **DEMONSTRATOR**
UNIVERSITY OF CAMBRIDGE **Cambridge, UK**

- 2017 Introduced around 100 students over one week to zebrafish developmental biology (Biology of Cells course)
- 2017: In charge of first year students over one month of laboratory lessons (Biology of Cells course)
- 2016: Supervised undergraduate cohort during two sessions of practical experimental work (Cell & Developmental Biology course)
- Answered student questions and marked submitted work (lab write ups, reports, and answers)

LANGUAGES

Italian - native
French - fluent
English - fluent

AFFILIATIONS

PSDB (Philippine Society for Developmental Biology) member, **DORA** signatory

HOBBIES

Capoeira (acrobatic martial art; 9+ years; qualified to teach), **piano** (14 years), **swimming** (14 years)

IT SKILLS

Web Development: PHP, HTML, SQL, CSS **Programming**: R (e.g. Seurat scRNAseq analysis; FACS processing pipeline), Python/Jupyter notebooks, Binder **Software**: MS Office, OpenOffice, GIMP/Photoshop, Adobe Illustrator **Virtual Reality**: Unity, Blender, Google Cardboard (all self-taught) Familiarity with Unix environment