Spatial and Molecular Characterisation of the Cellular Niches Interacting with Intra-Embryonic Haemato-Endothelial Progenitors



David Grainger

St Hilda's College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hilary 2022

Acknowledgements

My first thanks must go to Catherine Porcher for her unwavering support and proactive mentorship throughout my project. She always championed my development as a scientist and taught me skills in writing, thinking and mentorship that will remain with me for the duration of my career. It has been a pleasure to learn from Catherine and I appreciate the substantial time and effort she has dedicated to me. I have also been fortunate enough to benefit from the support of Marella de Bruijn as my co-supervisor. Always happy to lend an ear to my progress and offer thoughtful, considered advice. From the very start, Hedia Chagraoui offered me training and insightful discussion. I will never forget the rule that if you need to set up something new, Hedia will be able to help and can give you pages of notes to get you started. Janhavi Apte was my first friend in Oxford and introduced me to the WIMM in the most warm and welcoming way. I will treasure the memories of us singing Arctic Monkeys in that hot clean room for many years. Hilal Şengül joining the lab brought me another life-long friend whose kindness is plentiful and scientific mind one of the sharpest I have seen. Together with Vivien Ho, we made a great team who knew how to have fun. Vivien is a great friend who I thank for her honesty in sharing her DPhil experiences and for the laughs we've had along the way. We have both been fortunate to have two great friends and mentors in Arif Kirmizitas and Aldo Ciau-Uitz. Nobody has taught me more about developmental biology and how to study it than those two. Arif is my coffee buddy for life, has offered me endless help and advice and I will look back fondly on our daily chats about science and life. Aldo has informed the direction of my project with his vast knowledge of blood development and has done so with such humility and kindness. Our many "short walks" have kept me happy and healthy. I am grateful to my friends in the Vyas lab for their advice and help. The WIMM facilities have been outstanding and taught me so much about their expertise. A special mention to Sally Anne-Clarke in the FACS facility, Jackie Sloane-Stanley in the transgenic facility, Ollie Burns in the MHU and Chris Lagerholm in the imaging facility. I'm grateful to Tom Milne and Liz Robertson for their insightful discussion and advice during my thesis committee meetings. Outside the lab, Sven Turkalj, Gerda Mickute and Tom Jackson have formed my Oxford family and have made it my home. To James Harris, thank you for the years of too many beers and kebabs, I look forward to the next wine club. To my Mum, Dad and Alice, your support and belief in me has been a great source of motivation and always kept my work in perspective. My greatest thanks must go to my partner, best friend, housemate and soulmate, Emma. She has provided incredible stability and support over the last four and a half years with her unrelenting ability to find the positives and keep a smile on my face. I cannot find the words to thank her enough. Every DPhil student should have an Emma.

Abstract

Haematopoietic stem cells (HSCs) are self-renewing cells sustaining production of the adult blood system throughout life. The entire pool of HSCs is produced during embryonic development, from a specialised arterial endothelium primarily in the dorsal aorta (DA). There is increasing evidence that blood fate is specified during mesoderm patterning, at a restricted anterior-posterior (A/P) embryonic position, before emergence of the first HSCs. Although essential when aiming to produce HSCs in vitro from pluripotent stem cells (PSCs), and for a complete understanding of HSC production, little is known about the signalling pathways and molecular mechanisms instructing a blood cell identity in mesodermal haemato-endothelial progenitors (HEPs). I have studied the micro-environment supporting some of the earliest haematopoietic specification mechanisms in 2 different ways. First, as intra-embryonic HEPs and intermediate mesoderm (IM)-derived mesonephric duct share restricted A/P co-ordinates, I differentiated human PSCs in culture conditions known to support the development of the mesonephric duct and permissive for endothelial differentiation. I report production of KDR+ endothelial progenitors in PAX8+ IM niches and their differentiation towards CD144+DLL4+ arterial endothelium. Testing four haemogenic induction regimes did not reproducibly impart a haematopoietic fate, suggesting our culture conditions were not optimum for initiation or maintenance of a blood cell fate. Secondly, I studied HEPs in their native environment in mouse embryos. After establishing a quantitative, multiplexed single molecule RNA FISH method, I described the complex cellular environment of distinctly located HEPs in the E8.0 mouse embryonic posterior region, including those forming the future DA. To explore how these spatially distinct neighbourhoods might govern blood cell fate acquisition, I generated scRNA-seq data from the same embryonic region and performed cell communication analyses. I identified many classes of signalling critical for HSC emergence, suggesting these signals could function earlier than previously described, and define thrombospondin-CD47 and Cgrp-Calcrl as novel signals for functional validation.

Contents

List of Figures xi

List of Abbreviations xv

1 Introduction 1

1.1 Gastrulation and early organogenesis . . . 2

1.1.1 Establishing asymmetries in the early embryo is critical for the initiation of gastrulation . . . 2

1.1.2 The formation of the embryonic germ layers during gastrulation 3

1.1.3 Atlasing mouse organogenesis with scRNA-seq . . . 6

1.1.4 Early organogenesis of the ectodermal and endodermal cells . . 7

1.1.5 Extraembryonic mesoderm forms the allantois and yolk sac . . . 8

1.1.6 Differentiation of the paraxial, intermediate and lateral plate mesoderm . . . 9

1.2 Embryonic haematopoiesis . . . 12

1.2.1 Primitive Wave 1 embryonic haematopoiesis . . . 13

1.2.2 Pro-definitive Wave 2 embryonic haematopoiesis . . . 13

1.2.3 Definitive Wave 3 embryonic haematopoiesis . . . 16

1.3 Stepwise development from pluripotent epiblast to multipotent HSC . . 17

1.3.1 Lateral plate mesoderm as the source of HSCs . . . 17

1.3.2 Transcriptional control of haemato-endothelial fate specification 20

1.3.3 From angioblasts to the arterial endothelium of the dorsal aorta 21

1.3.4 The endothelial to haematopoietic transition and HSC maturation 25

1.3.5 Cell signalling underlying the emergence of HSCs from HE . . . 26

1.4 *In vitro* differentiation of the haematopoietic lineages from hPSCs . . . 27

1.4.1 Motivation and rationale for developing hPSC differentiation systems . . . 27

1.4.2 Directing differentiation away from primitive and towards definitive haematopoiesis during mesoderm patterning . . . 29

1.4.3 The relationship between hPSC derived arterial and haemogenic endothelium . . . 32

1.4.4 The likely causes of the inability to form HSCs *de novo* from current hPSC differentiation protocols and new ways to address them . . . 34

1.5 Identifying a suitable spatial transcriptomic method depends on practical considerations . . . 37
1.5.1 Spatial transcriptomic methodologies using single molecule RNA FISH . . . 38
1.5.2 Spatial transcriptomic methodologies using spatial barcoding technologies with RNA-sequencing . . . 40
1.5.3 Spatial transcriptomic methodologies using laser capture microdissection (LCM) with RNA-sequencing . . . 41
1.6 Thesis aims and outline . . . 42

2 Materials and Methods 45
2.1 Thawing, freezing and maintenance of human embryonic stem cell lines 46
2.1.1 Mycoplasma testing . . . 46
2.1.2 Matrigel coating of culture plates . . . 47
2.1.3 Passaging human embryonic stem cells . . . 47
2.1.4 Thawing human embryonic stem cells . . . 47
2.1.5 Freezing human embryonic stem cells . . . 48
2.2 *In vitro* differentiation of human embryonic stem cells to intermediate mesoderm in defined medium . . . 48
2.3 *In vitro* differentiation of human embryonic stem cells to haemogenic endothelium in defined medium . . . 49
2.4 Freezing, thawing and passaging of human umbilical vein endothelial cells (HUVECs) . . . 49
2.5 Reverse transcription quantitative polymerase chain reaction (RT-qPCR) 50
2.6 Flow cytometry and fluorescent activated cell sorting (FACS) . . . 52
2.7 Tubule formation assay . . . 53
2.8 Immunofluorescence of hESC differentiation cultures . . . 53
2.9 Mouse colony maintenance and genotyping . . . 54
2.10 Timed matings and embryo collection . . . 55
2.11 Embryo preparation for cryosectioning . . . 55
2.12 Hybridisation chain reaction version 3.0 (HCR v3.0) . . . 56
2.12.1 Split-initiator probes, hairpin amplifiers and buffers . . . 56
2.12.2 Digital HCR (dHCR) . . . 56
2.12.3 Quantitative HCR (qHCR) . . . 58
2.13 Probe and signal stripping . . . 59
2.14 Cell membrane staining . . . 59
2.15 Image alignment . . . 60
2.16 3-D cell segmentation . . . 60
2.17 dHCR spot counting . . . 61

2.18 Hierarchical clustering and spatial remapping of clusters 63
2.19 Laser capture micro-dissection 63
2.19.1 Sectioning of fresh frozen mouse embryos 63
2.19.2 Section preparation and rapid staining for LCM 64
2.19.3 Laser capture micro-dissection and catapulting 64
2.19.4 Reverse transcription and pre-amplification of single cell lysates 65
2.20 10x Chromium scRNA-seq 66
2.20.1 Mouse embryo dissection and dissociation for 10x Chromium . . 66
2.20.2 10x Chromium library preparation 66
2.20.3 Illumina next generation sequencing 67
2.20.4 Sequencing data pre-processing and quality control 67
2.20.5 Normalisation and clustering 10x scRNA-seq 68
2.20.6 Removal of doublets from scRNA-seq data 69
2.21 CellChat cellular communication prediction 69

3 Directed differentiation of hESCs to intra-embryonic haemato-endothelium 71
3.1 Introduction 71
3.2 Results 73
3.2.1 Selection of intermediate mesoderm differentiation protocols for testing 73
3.2.2 Establishing intermediate mesoderm differentiation protocols . . 74
3.2.3 KDR^+ cells generated in intermediate mesoderm differentiation conditions are capable of endothelial differentiation 79
3.2.4 Differentiation cultures exhibit the correct HOX expression pattern 83
3.2.5 Angioblasts differentiated in high levels of VEGF give rise to arterial endothelium 85
3.2.6 Inducing a haematopoietic programme in arterial endothelium cultures is highly variable 86
3.3 Discussion 88

4 Defining the cellular neighbourhood of intra-embryonic haemato-endothelial progenitors 93
4.1 Introduction 93
4.2 Results 95
4.2.1 *Scl* positive angioblasts are widespread throughout the embryonic posterior 95
4.2.2 Establishing Spatial Genomics analysis 99
4.2.3 Selection of three embryos for spatial genomics analysis 110
4.2.4 Qualitative assessment of haemato-endothelial progenitors and neighbouring mesodermal cells by multiplexed HCR 111

4.2.5 Assessment of RNA degradation during multiplexed dHCR . . . 117
4.2.6 Quantitation of multiplexed HCR enables the spatial remapping of clusters . . . 120
4.2.7 Investigating the degree of heterogeneity in gene expression between spatially distinct haemato-endothelial progenitors . . . 125
4.3 Discussion . . . 127

5 Predicting cellular signalling to angioblasts in the E8.0 embryonic posterior 131
5.1 Introduction . . . 131
5.2 Results . . . 134
5.2.1 Establishing a rapid staining protocol to discern cell borders and identify haemato-endothelial progenitors on mouse embryo sections 134
5.2.2 LCM is able to cut and capture single cells but lysis results in degraded cDNA . . . 139
5.2.3 scRNA-seq of the caudal half of E8.0 . . . 140
5.2.4 Cell communication patterns received by HEPs in the caudal half of E8.0 embryos . . . 148
5.2.5 Cell signalling pathways recieved by the HEPs in the caudal half of E8.0 embryos . . . 152
5.3 Discussion . . . 158

6 Final conclusion and Discussion 163
6.1 Summary of key findings . . . 163
6.2 hPSC derived endothelium and the possibility of all young endothelium having blood-forming potential . . . 165
6.3 Using intermediate mesoderm A-P patterning to inform the signals required for intraembryonic HEP regionalisation . . . 166
6.4 Prediction of the early signals to intra-embryonic haemato-endothelial progenitors . . . 167
6.5 Application of newly developed technologies to understanding haemato-endothelial development . . . 167

Appendices

References 191

List of Figures

1.1 Mouse embryo development from E6.0 to E7.5 4
1.2 Emergence of the haematopoietic system during development occurs in three waves 14
1.3 Stepwise development from pluripotent epiblast to multipotent HSC . . 18
1.4 Critical transcription factors controlling haemato-endothelial fate specification 22
1.5 Two models for the development of primitive, pro-definitive and vascular endothelium from hPSCs 31
1.6 The progenitors of Wave 3 HE are restricted along the A-P axis and lie adjacent to the nephric duct 36
1.7 Comparison of the three classes of spatial transcriptomics methods . . 39

3.1 Assessment of the Xia differentiation protocol assessed by RT-qPCR . . 75
3.2 Intermediate mesoderm differentiation from hESCs 76
3.3 KDR and PAX8 are expressed in different cell subsets that exhibit organisation, structure and close proximity 80
3.4 KDR^{+} cells are capable of differentiation into endothelium 82
3.5 Haemato-endothelial progenitors differentiated in high levels of VEGF give rise to arterial endothelium 84
3.6 Addition of haematopoietic cytokines is unable to repeatably induce a haematopoietic programme in arterial endothelium 87

4.1 Whole mount imaging of the E8.0 mouse embryo posterior for location of Scl positive angioblasts 96
4.2 Quantification of angioblast numbers in the E8.0 posterior 98
4.3 Virtual sectioning of E8.0 mouse embryo posterior 100
4.4 Overview of the Spatial Genomics Analysis workflow. 102
4.5 DNaseI allows efficient removal of fluorescent signal and DNA probes. . 103
4.6 Cell volume segmentation on immunostained cell surface proteins . . . 105
4.7 Registration of imaging rounds is inaccurate with stitched images but accurate with single fields of view. 107
4.8 Schematic of smFISH quantification pipeline for SGA. 109
4.9 Imaging of haemato-endothelial gene expression in transverse sections of E8.0 embryos for selection of a representative region for quantification . 112

4.10 Haematoendothelial progenitors have distinct locations and varying expression of *Scl*, *Etv2* and *Tek* 114
4.11 smFISH of LPM, IM and pre-somitic mesoderm markers identifies overlapping domains of expression 116
4.12 Domains of *Tek*, *Lhx1*, *Etv2* and *Scl* staining in the first and last rounds of dHCR staining are the same 118
4.13 High magnification of *Tek*, *Lhx1*, *Etv2* and *Scl* staining in the first and last rounds of dHCR show a small loss in RNA molecule staining 119
4.14 Quantification of gene expression to enable cell type clustering and spatial remapping 122
4.15 Variation in HEP gene expression along the medial-lateral axis 126

5.1 Etv2-Cre can be used to mark major embryonic vessels and Wave 3 blood but is unsuitable for labelling cells for LCM-seq 136
5.2 Discrimination of cell borders and identification of angioblasts by rapid staining in preparation for cDNA library preparation 138
5.3 Overview and quality control of scRNA-seq of E8.0 embryo caudal half 141
5.4 Clustering and UMAP of scRNA-seq data of the caudal half of E8.0 embryos 144
5.5 Gene expression matrix for the top ten enriched genes for each cell type of the caudal half of E8.0 embryos 145
5.6 Identification of doublets with DoubletFinder and scDblFinder 147
5.7 CellChat predicts a complex cell signalling network 149
5.8 Identification of incoming cell signalling patterns 151
5.9 Summary of enriched signalling to haemato-endothelial cells 153

1 Additional genes profiled during Takasato protocol differentiation of hESCs 173
2 Variation in gene expression profiles during endothelial differentiation . 174
3 Variation in HOX gene expression profiles during Takasato protocol differentiation 175
4 Variation in gene expression profiles during arterial differentiation . . . 176
5 Brightfield images of 30µm sections from embryo 3 177
6 List of HCR probes and their initiators used 178
7 High magnification of Etv2 single positive cell in Embryo 2 179
8 All genes probed for embryo 3 180
9 All genes probed for embryo 2 181
10 All genes probed for embryo 3 182
11 Individually plotted gene profiles in embryo 1 HEPs ordered along the medial-ventral axis 183

12 Individually plotted gene profiles in embryo 2 HEPs ordered along the medial-ventral axis 184
13 FACS gating strategy for assessment of Etv2-Cre labelling efficiency in peripheral blood 185
14 Brightfield images of fresh frozen E8.0 mouse embryonic sections in decidua 186
15 Extrapolation of sequencing saturation 187
16 Identification of two clusters for removal 188
17 Cluster quality control 189

List of Abbreviations

AF . . . . . Alexa Fluor

AGM . . . . . Aorta-gonad-mesonephros region

AS . . . . . (3-aminopropyl)triethoxysilane

BMP . . . . . Bone morphogenic protein

BSA . . . . . Bovine serum albumin

CHIR CHIR99021

CO_2 . . . . . Carbon dioxide

DE . . . . . Definitive endoderm

DMSO . . . . . Dimethyl sulfoxide

DNA . . . . . Deoxyribonucleic acid

DPBS . . . . . Dulbecco's phosphate-buffered saline

EHT . . . . . Endothelial-to-haematopoietic transition

EDTA . . . . . Ethylenediaminetetraacetic acid

EMP . . . . . Erythro-myeloid progenitor

ExE . . . . . Extraembryonic ectoderm

ExM . . . . . Extraembryonic mesoderm

FGF . . . . . Fibroblast growth factor

FACS . . . . . Fluorescent activated cell sorting

FL . . . . . Foetal liver

HCR . . . . . Hybridisation Chain Reaction

HE . . . . . Haemogenic endothelium

HSC . . . . . Haematopoietic stem cell

hESC . . . . . Human embryonic stem cell

HLA . . . . . Human leukocyte antigen

hPSC . . . . . Human pluripotent stem cell

HUVEC . . . . . Human umbilical vein endothelial cell

IgG . . . . . Immunoglobulin G

IM	Intermediate mesoderm
IAHC	Intra-aortic haematopoietic cluster
KOSR	Knock-out serum replacement
LCM	Laser capture microdissection
LPM	Lateral plate mesoderm
mm	Millimetre
NA	Numerical aperture
nM	Nanomolar
nm	Nanometre
OCT	Optimal cutting temperature compound
PBS	Phosphate-buffered saline
PBST	Phosphate-buffered saline plus 0.2% Triton-X100
PFA	Paraformaldehyde
RNA	Ribonucleic acid
RT-qPCR	Quantitative reverse transcription PCR
SB	SB-431542
SSC	Saline Sodium Citrate buffer
SSCT	Saline Sodium Citrate plus 0.2% Triton-X100
scRNA-seq	Single cell RNA sequencing
TF	Transcription factor
VE	Visceral endoderm
YS	Yolk sac
µ L	Microlitre
µ m	Micrometre
℃	Degrees celcius

Chapter 1

Introduction

Contents

1.1 Gastrulation and early organogenesis 2
1.1.1 Establishing asymmetries in the early embryo is critical for the initiation of gastrulation 2
1.1.2 The formation of the embryonic germ layers during gastrulation 3
1.1.3 Atlasing mouse organogenesis with scRNA-seq 6
1.1.4 Early organogenesis of the ectodermal and endodermal cells 7
1.1.5 Extraembryonic mesoderm forms the allantois and yolk sac 8
1.1.6 Differentiation of the paraxial, intermediate and lateral plate mesoderm 9
1.2 Embryonic haematopoiesis 12
1.2.1 Primitive Wave 1 embryonic haematopoiesis 13
1.2.2 Pro-definitive Wave 2 embryonic haematopoiesis 13
1.2.3 Definitive Wave 3 embryonic haematopoiesis 16
1.3 Stepwise development from pluripotent epiblast to multipotent HSC 17
1.3.1 Lateral plate mesoderm as the source of HSCs 17
1.3.2 Transcriptional control of haemato-endothelial fate specification 20
1.3.3 From angioblasts to the arterial endothelium of the dorsal aorta 21
1.3.4 The endothelial to haematopoietic transition and HSC maturation 25
1.3.5 Cell signalling underlying the emergence of HSCs from HE 26
1.4 *In vitro* differentiation of the haematopoietic lineages from hPSCs 27
1.4.1 Motivation and rationale for developing hPSC differentiation systems 27
1.4.2 Directing differentiation away from primitive and towards definitive haematopoiesis during mesoderm patterning . 29
1.4.3 The relationship between hPSC derived arterial and haemogenic endothelium 32
1.4.4 The likely causes of the inability to form HSCs *de novo* from current hPSC differentiation protocols and new ways to address them 34
1.5 Identifying a suitable spatial transcriptomic method depends on practical considerations 37
1.5.1 Spatial transcriptomic methodologies using single molecule RNA FISH 38

1.5.2 Spatial transcriptomic methodologies using spatial barcoding technologies with RNA-sequencing 40
1.5.3 Spatial transcriptomic methodologies using laser capture microdissection (LCM) with RNA-sequencing 41
1.6 Thesis aims and outline 42

1.1 Gastrulation and early organogenesis

At the core of developmental biology lies the question of how cells of the embryo diversify from a single fertilised egg into the many distinct cell types of the adult organism? The diversification of cell identities requires a series of cell fate decisions. Cell fate can be described as "the developmental destination of a cell if left undisturbed in its environment" [1]. While cells can only have one fate, many cells exhibit plasticity and have multiple potentials. A key driver of understanding how cell fates are realised in cells with multiple potentials is to study the cell signalling that drives intracellular cascades of epigenetic and transcriptional change. Although stark differences in morphology and structure exist between animal species, the molecular signals that drive the series of cell state decisions and allocation to cell types during embryogenesis are highly conserved [2]. Here I focus primarily on murine development as the most studied mammalian embryo owing to its relative ease of genetic manipulation and *in vitro* embryonic stem (ES) cell differentiation.

1.1.1 Establishing asymmetries in the early embryo is critical for the initiation of gastrulation

By the time embryos have implanted into the uterine wall at embryonic day (E)4.5 of mouse development, the first cell state decision has already been made with the totipotent blastomere giving rise to the innner cell mass (ICM) and trophectoderm (TE). While TE contributes to the foetal compartment of the placenta, the ICM primarily gives rise to the embryo proper, as it differentiates in part to form the pluripotent epiblast. Pluripotency is defined by the ability of a cell to differentiate into all cells within the embryo proper. To begin, this diversification of embryonic cell types requires

a break in symmetry not present when the epiblast first forms at E4.5. At E5.5, the epiblast has adopted a cup shape and is wrapped in the visceral endoderm (VE) and attached proximally to the ectoplacental cone through the extra-embryonic ectoderm (ExE). The signalling to and from these three primitive tissues break both proximal-distal and anterior-posterior symmetries. Reciprocal Nodal, bone morphogenic protein (BMP), fibroblast growth factor (FGF) and Wnt signalling with the proximal ExE and distal VE induces a proximal-distal axis to form within the epiblast (Fig.1.1i)[2].

The morphogenic movement of the distal VE proximally across one side of the epiblast at E6.0 (Fig.1.1i) [3, 4] shifts this signalling centre to the prospective embryonic anterior and breaks radial symmetry in the epiblast to form an anterior-posterior axis [5]. The now anterior VE acts as a signalling centre to antagonise the expression of markers of the primitive streak such as *Nodal, Brachyury* and *Cripto* [6–8]. This establishes the axes, differences in gene expression and signalling gradients required to initiate gastrulation.

1.1.2 The formation of the embryonic germ layers during gastrulation

At E6.0, cells of the newly formed posterior epiblast migrate towards the proximal edge of the embryo, undergo a epithelial-to-mesenchymal transition (EMT) and form the primitive streak (Fig.1)[9, 10]. This structure is formed from the invagination of cells into the space between the epiblast and the VE. Nodal and Wnt signals from the epiblast intersect with ExE-derived BMP signals to pattern these cells as nascent mesoderm [11–15]. Following the initiation of primitive streak formation, it gradually extends distally along the posterior half of the epiblast until it reaches the distal tip at E7.75 (Fig.1.1). Lineage labelling of cells in the epiblast demonstrates that, although cells are pluripotent at that stage, a cells position within the epiblast is predictive of its fate. The position of a cell in the epiblast determines when and where it will transverse the streak and consequently the precise concentration of signalling it will receive [9, 16,

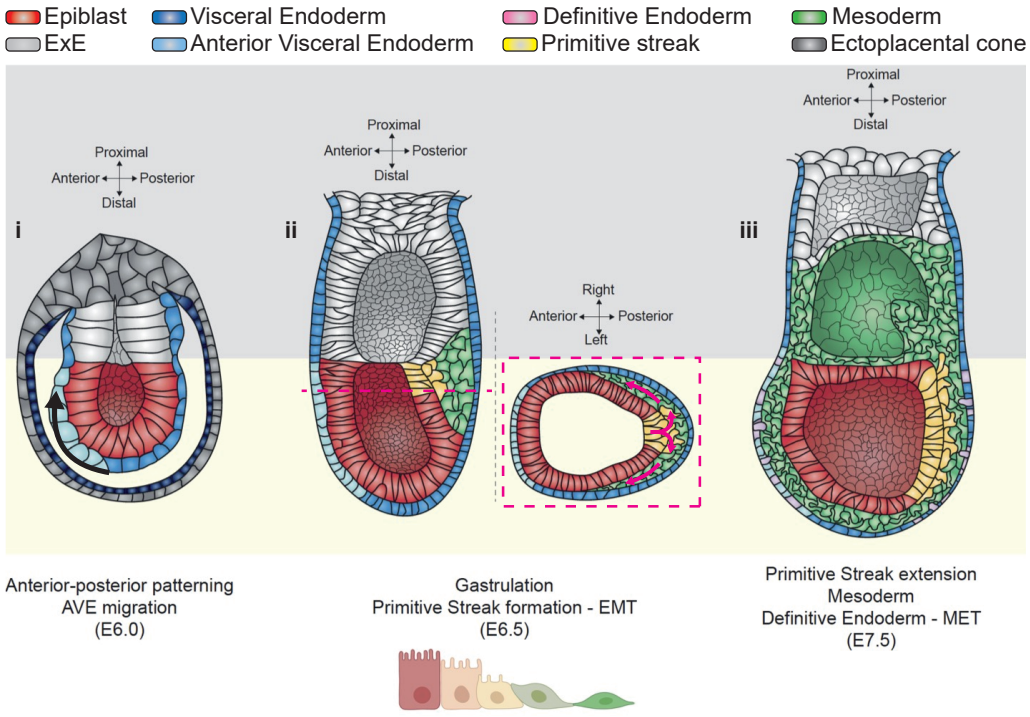


Figure 1.1: Mouse embryo development from E6.0 to E7.5
At E6.0 (**i**), the cylindrical embryo consists of the pluripotent epiblast surrounded by the visceral endoderm (VE). The embryo is connected to the ectoplacental cone by the extra-embryonic ectoderm (ExE). The distal-most endoderm moves anteriorly (curved black arrow) to form the anterior visceral endoderm to form the antieror-posterior axis. By E6.5 (**ii**), epiblast diametrically opposed to the AVE undergo an epithelial-to-mesenchymal transition (EMT, represented by change in cell shape in bottom of ii) and migrate between the epiblast and VE to form the mesodermal layer (transverse section in dashed magenta box with magenta arrows showing exit from the primitive streak and into the mesodermal wings) and integrate with into the VE to form the definitive endoderm. Between E6.5 and E7.5 (**iii**) the primitive streak extends distally. Adapted with permission from Nahaboo & Migeotte, eLS 2018. [18]

17].

Early in gastrulation, the primitive streak has not yet extended and so cells passing through will experience high levels of BMP4 from the ExE. The first cells that transverse the primitive streak form the extra-embryonic mesoderm (ExM) [19]. As the streak moves distally during the mid-streak stage of gastrulation, epiblast cells fated to form the lateral plate, intermediate, paraxial and cardiac mesoderm pass through, specified by Nodal (from the anterior streak) but independent of SMAD4 [13, 20]. Finally, as the streak reaches its most distal location, distal epiblast cells fated to the prechordal plate, the notochord, the node and the definitive endoderm (DE) pass through. This allocation of fates along the proximal-distal primitive streak is established primarily by the inverse gradients of BMP4 and Nodal. Nodal-SMAD2 signalling is critical for specification of the fates of the anterior primitive streak with graded levels of signalling conferring different fates. Increasing levels of SMAD activation are required to induce paraxial and lateral plate mesoderm versus the node with the DE and prechordal plate requiring the highest [20, 21]. Meanwhile in the posterior streak, cells experience high levels of BMP signalling from the ExE.

After emerging from the streak, nascent mesoderm is instructed with specific morphological directions such that cells migrate far across the embryo. Firstly, the ExM mesoderm leaves the posterior streak and migrates anteriorly and proximally. Conversely the intra-embryonic mesodermal lineages exiting during the mid-streak stage wrap around the epiblast on the left and the right of the embryo to form two symmetrical mesodermal wings (Fig.1.1ii, magenta arrows). These sheets of tissue are moving from posterior to anterior and towards their final embryonic coordinates. From E6.5, the nascent DE moves out of the streak and intercalates with the pre-existing VE with both cell types persisting into early organogenesis [22]. Finally, epiblast cells that lie on the most anterior aspect of the epiblast at the start of gastrulation never pass through the streak. These remaining cells take what is often referred to as the 'default differentiation path of epiblast cells' [2, 23, 24] the neuroectoderm (NE) [16]. In addition to not

undergoing EMT and passing through the streak, NE specification requires antagonistic inhibitory signals from the anterior VE [25].

Heterotypic transplantation of cells demonstrates that prior to gastrulation, cells remain very plastic with broad potency and an ability to adopt the fates of the cells around them [17]. Following gastrulation, this effect is much smaller with cells having a reduced potency [26] and distinct transcriptional programmes [27]. Furthermore, there is evidence that some fates are specified during gastrulation. Firstly, *Mesp1*-derived cardiovascular progenitors in the E6.25 primitive streak are already restricted to either a cardiomyocyte or epicardial fate [28, 29]. Secondly, the cells contributing to the left ventricle, right ventricle, outflow tract and atrium all exit the streak at different times and are therefore pre-patterned separately in the primitive streak [30]. Additionally, the reconstruction of lineage trajectories from single cell RNA-sequencing (scRNA-seq) data identified two routes to the early somites with a shared origin in the epiblast, but that quickly diverged soon after gastrulation [31]. Such early separation indicates these lineages are specified differently as early as E7.0-E7.5. These studies highlight the importance of studying the earliest stages of cell state segregation and motivate the investigation of early specification of other lineages at high resolution.

1.1.3 Atlasing mouse organogenesis with scRNA-seq

Following the assignment to one of three germ layers, cells further specialise to form the progenitors of tissues and organs. Traditionally, organogenesis is defined as beginning when the neural tube is formed and the first somites begin to form at E8.0. This definition excludes the formation of the first blood progenitors from occurring during organogenesis, likely due to blood often being defined as a connective tissue rather than a definitive organ. The first studies aiming to describe the diversification of these transcriptional programmes analysed the changes in transcript diversity in 6 hour measurements of whole-embryo micro-array based transcriptomics. Although averaging the transcriptome of thousands of cells has its limitations, it did identify

a significant increase in transcript diversity starting at E8.0 [32]. Refinement of this global approach to transcriptome diversity came with the advent of single cell RNA sequencing (scRNA-seq) as a way of measuring cellular heterogeneity [33–35]. These studies rely on clustering algorithms to group transcriptionally similar cells together that are then annotated by probing the expression of known marker genes and evaluating the enriched genes in each cluster. The annotation of cells in this manner is one of the most challenging aspects of such atlasing projects but these seminal studies now provide excellent references for future work to accurately define cell types more easily.

With the embryo developing very rapidly during this critical time period [36], these atlases needed to sample the embryo at regular (6 hour) intervals. Pijuan-Sala *et. al.,* profiled more than 116,000 cells from E6.5-E8.5 and annotate 37 major cell types [33]. Comparative analyses of these populations can inform the selection of marker genes and when combined with computational reconstruction of lineage trajectories can predict the change in gene expression during differentiation. The authors utilise this extensive dataset to investigate the transcriptional convergence of the VE and DE into the gut, the stepwise development of the blood and endothelium of the yolk-sac (YS) and consequence of *SCL* deletion on blood fate specification. These atlases required the dissociation of cells from the embryo and currently only measure the transcriptome. Increasingly, studies incorporating lineage history [37, 38], cell location [39, 40] and epigenetic information [27, 41] are being appended and incorporated into these atlases.

1.1.4 Early organogenesis of the ectodermal and endodermal cells

Following the integration of DE into the overlying VE, these two cell types undergo complex changes to form the gut tube [42]. Folding at both the anterior and posterior ends of the DE to form the anterior and caudal intestinal portals. Meanwhile the ventral edges of the DE move towards the midline to form a tube along the length of the embryo enclosing the DE. The DE is patterned along the A-P axis by the surrounding mesoderm to form discrete organ-specific molecular domains by E8.5 [43]. The gut

tube is separated into the anterior, mid- and hindgut. The anterior gut generates the pancreas and biliary system, lungs, thyroid, liver and oesophagus; the midgut forms the stomach and small intestine and; the hindgut gives rise to large intestines and the lining of the urogenital system.

The ectoderm differentiates into the neural tube, the neural crest and the surface ectoderm during primary neurulation. Inductive BMP signals from the notochord thicken the overlying ectoderm to form the neural plate. The edges of the neural plate move dorsally while the centre remains in place creating a neural groove. The neural folds (dorsal lips of the groove) move towards the midline and pinch together to form the neural tube. The neural tube is regionalised by the surrounding mesoderm to form the components of the central nervous system. Until recently, the prevalent view was this regionalisation occurred after neural induction [44]. Through measuring chromatin accessibility through neural induction, Metzis *et. al.,* demonstrate that the neueroectoderm is first regionalised along the A-P axis and subsequently acquire neural identity [45]. Meanwhile cells of the neural folds delaminate and migrate to form the neural crest, which differentiates into the peripheral nervous system, adrenal medulla, melanocytes, facial cartilage and dentine of teeth [46, 47]. The ectoderm lateral to the neural plate remains as a flat, thin layer of cells dorsal to the neural tube to form the surface ectoderm. The surface ectoderm differentiates to form the epidermis, hair, nails, sebaceous glands, olfactory epithelium, mouth epithelium, lens and cornea [47].

1.1.5 Extraembryonic mesoderm forms the allantois and yolk sac

The first cells to exit the streak are exposed to high levels of BMP4 signalling from the neighbouring ExE and form the ExM [48]. Following exit, these cells move proximally across the proximal edge of the embryo and into extra-embryonic tissues to form the allantois, YS vasculature and blood. Two cell types coalesce to form the allantoic bud: the mesothelium on the outside [49] and the allantoic core in the centre [50]. The bud extends out into the exocoelomic cavity before fusing with the chorion at around

the 6 somite stage (ss) [50]. Meanwhile, the allantoic core differentiates into vascular progenitors (angioblasts) which organise to form the allantoic vascular plexus [51]. This plexus continues to develop and grows into the chorion-derived placental villi to form the layrinthine layer of the placenta. Fate mapping of the allantoic, YS and amniotic mesoderm demonstrated that the ExM lineages derive from independent primitive streak populations and therefore are likely specified separately [9, 16, 49].

Meanwhile, the YS mesoderm differentiates into the YS endothelium and first primitive haematopoietic cells (discussed further in section 1.2.1). These two cell types develop in close proximity in structures called blood islands (BI), where clusters of haematopoietic cells develop surrounded by endothelium. These primitive structures led to the hypothesis of a shared mesodermal progenitor between haematopoietic and endothelial cells (termed the haemangioblast) [52, 53]. Through the identification of the blast colony-forming cell (BL-CFC) first *in vitro* from embryonic stem cell (ESC) differentiations [54, 55] and subsequently *in vivo* in the posterior primitive streak [56] this common progenitor was defined as FLK1 and Brachyury positive. However, these studies relied heavily on the culture of BL-CFC and as such were not studying cells in their undisturbed environment. This is critical, as a cells potential *in vitro* does not necessarily reflect its fate *in vivo*. Indeed, when studying the cell fates realised *in vivo* through clonal lineage tracing or chimeric embryo studies, YS blood and endothelial lineages were concluded to derive from independent epiblast populations [17, 57, 58]. Therefore, although cells with bi-potential may exist, the frequency of a common progenitor is likely low. Later in development, a specialised subset of the YS endothelium differentiates into multipotent haematopoietic progenitors (discussed in detail in section 1.2.2).

1.1.6 Differentiation of the paraxial, intermediate and lateral plate mesoderm

In contrast to ExM, intraembryonic mesoderm exits the streak and move anteriorly, forming two mesodermal wings wrapping around the epiblast underneath the overlying endoderm. As its name suggests, the paraxial mesoderm sits either side of the

midline (major axis) and undergoes a series of differentiation steps to form the axial musculoskeletal system. The presomitic mesoderm is protected from the lateralisation effects of BMP signals from the intermediate mesoderm (IM) and lateral plate mesoderm (LPM) [59] by notochord derived noggin [60]. Many transcription factors (TFs) have been shown to have essential roles in somite differentiation and formation [61]. Perturbation of expression levels or domains has demonstrated some of these factors act to suppress specification of the paraxial mesoderm away from neighbouring cell types. The T-box TF TBX6 is initially expressed in the primitive streak but is maintained in the presomitic mesoderm [62] by Msng1, the master regulator of paraxial mesoderm fate [63]. Loss of TBX6 function results in the conversion of paraxial mesoderm to ectopoic neural tubes [64] indicating its role in neural versus mesodermal fate. Similarly, the forkhead helix TFs Foxc1/2 are both required for somite formation but also have function to suppress the adoption of an intermediate mesoderm fate [65].

Two major cell populations are formed in the mature somite, the sclerotome and dermomyotome. The sclerotome specifically expresses Pax1, Pax9 and Sox9 in the ventromedial somite and differentiates to form the vertebrae and associated ribs and tendons. In zebrafish, the sclerotome contributes to the endothelium of the dorsal aorta but this has not been formally demonstrated in mammals. However, the contribution of $Pax3^{+}$ mesoderm-derived endothelial progenitors to the lymphatic vasculature has been shown in mice [66]. The dermamoyotome is characteriesed by MyoD, a TF required for muscle differentiation and it upstream regulators Pax3 and Myf5 [67]. This dorsolateral region of the somite splits further into the myotome (progenitor or skeletal muscle) and the dermatome (the dermis).

Lying lateral to the paraxial mesoderm but medial to the lateral plate mesoderm, a strip of mesoderm that forms the urogenital system develops. The IM extends caudally from approximately the 6th somite [68, 69] and is initially regionalised into the anterior and posterior intermediate mesoderm (aIM, pIM respectively) [70]. The anterior and posterior IM are spatially and transcriptionally defined regions sharing expression of TFs

critical for IM specification such as OSR1 [69, 71] but with the aIM expressing LHX1 and PAX2 and the pIM expressing HOXD11 [70, 72]. Each IM population differentiates to form a mesenchymal primordia with aIM forming the mesonephric mesenchyme and the pIM forming the metanephric mesenchyme. The aIM additionally gives rise to the nephric duct, the extension of which defines the three nephric stages [73]. These nephric stage are: the non-functional pronephros, a primitive mesonephros and the definitive metanephros that forms the adult kidney. As the nephric duct extends caudally, it reciprocally signals to the local mesenchyme to form excretory organs that consist of nephronic structures that drain into the cloaca. In females, the mesonephros, along with the anterior pronephros will completely degrade but in males forms the epididymas of the testis. The metanephric kidney continues to develop into the definitive adult kidney. Both of the mesenchymal and nephric duct components of these organs derive from *Osr1*$^+$ IM [74]. Together with another TF, LHX1, OSR1 mark the earliest IM fated cells with overlapping expression into the LPM [69, 71, 75]. It is not until around the 4ss that more specific markers have been identified in PAX2 and PAX8 [68, 76].

Much of our understanding of the signalling controlling IM specification and diversification comes from *ex vivo* culture of kidney rudiments and the differentiation of ESCs towards cell fates of the metanephric kidney [73]. hPSC differentiation protocols share many of their first stages of differentiation through the induction of a primitive streak state by WNT or BMP signalling followed by the use of FGF9. Comparison between induction of primitive streak (MIXL1-GFP+) by Wnt or BMP showed no difference. Following the observation that BMP4/WNT signalling alone could not differentiate *LHX1* or *PAX2* positive mesoderm (IM), Takasato *et. al.,* tested the ability of FGF signalling to direct differentiation [77]. FGF2, FGF8 and FGF9 were tested for independent reasons. Only the addition of 200 ng/mL of FGF9 to cells resembling the primitive streak at day 2 of differentiation saw the upregulation of *OSR1, LHX1* and *PAX2*. FGF9 was selected for testing because of two observations. Firstly, its expression in the IM and paraxial mesoderm [78] demonstrates its high levels in the region of interest. Secondly, double genetic knock out of FGF9 in conjunction with FGF20, results

in kidney agenesis due to their role in supporting metanephric mesenchyme survival [79]. It is, therefore, a little surprising that FGF9 alone is able to support the differentiation of aIM if it is not required for specification *in vivo*. This indicates FGF9 is sufficient but not necessary in this process and is likely part of a redundant signalling network established to form the earliest urogenital progenitors.

The most lateral intraembryonic mesoderm, the LPM exits from the middle of the streak and experiences high levels of BMP4 and low levels of noggin during its specification [59, 60, 80]. The LPM is regionalised in both the A-P axis forming the anterior (a)LPM and posterior (p)LPM and the dorsal-ventral (D-V) axis splitting into the somatic and splanchnic. Both domains differentiate in part to the mesothelium surrounding the body (parietal) and organs (visceral). Unlike the other intraembryonic mesoderm compartments, there is no known master regulator or marker gene that specifically and universally labels the LPM. As such, genes labelling a particular subset of the LPM or that label other tissues have primarily been used to map LPM fate [81]. Using *Wt1*-Cre, many groups have mapped the fate of a portion of the LPM to visceral fat, fibroblasts and the smooth muscle of the vascular system, gastrointestinal and respiratory tracts [81–85]. Similarly, the Hoxb6-Cre and Hoxb6-CreERT transgenic mouse lines have been used to trace the LPM and perform region-specific perturbation [86–89]. Hoxb6^{+} mesoderm contributes to the limbs [89], endothelium of the trunk DA [86, 87] and intra-embryonic haematopoietic clusters [87] and is often used to assign an LPM origin to these tissues. However, Hoxb6 fate mapping also labels the extra-embryonic mesoderm with broad labelling of the YS [88, 89] and so require further investigation.

1.2 Embryonic haematopoiesis

Embryonic haematopoiesis can be simply defined as the production of blood/haematopoietic cells during embryogenesis. This process is complex with the *de novo* generation of many independent blood progenitors at different stages of development. As the needs of the

embryo evolve and increase in complexity, so does the haematopoietic system. The study of multiple vertebrate species has demonstrated that these developmental processes are highly conserved and share striking similarity right down to the transcription factors and cellular signals orchestrating them. Though the main focus of this introduction is on murine development, other organisms are referenced when appropriate.

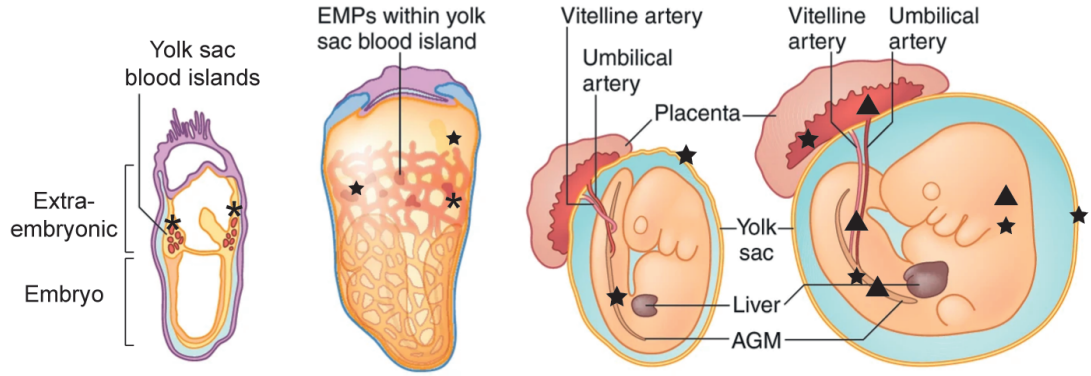
As detailed below, embryonic haematopoiesis can be divided into three temporally overlapping waves that support the developing needs of the embryo through development and into adulthood. Defined here as Wave 1, 2 and 3, each can be separated by cellular output, site of emergence, and function in development and adulthood (Fig.1.2).

1.2.1 Primitive Wave 1 embryonic haematopoiesis

Initiating at E7.25, Wave 1 is the most primitive in its cellular output, producing nucleated erythrocytes, megakaryocytes and macrophages (Fig.1.2, Wave 1) [90, 91]. Primitive erythrocytes are large, nucleated cells that express embryonic globin genes and serve to distribute oxygen around the embryo as it becomes too large to rely on diffusion alone [92]. As compared to their adult counterparts, the primitive macrophages in this wave do not pass through a monocyte precursor [93, 94] and primitive megakaryocytes have a lower ploidy and produce fewer platelets [95]. This first wave of primitive haematopoiesis occurs extra-embryonically in the yolk-sac (YS) of mammals and birds. From this early stage in development, haematopoietic cells form a close relationship with endothelial cells, the inner most cell lining of blood vessels.

1.2.2 Pro-definitive Wave 2 embryonic haematopoiesis

Soon after the initiation of the primitive Wave 1 of embryonic haematopoiesis, the pro-definitive Wave 2 begins. From E8.25, multi-potent haematopoietic progenitors with erythroid, megakaryocyte and myeloid potential (EMPs) are detectable in the YS [96] (Fig.1.2, Wave 2). These progenitors lack lymphoid potential and are distinct from Wave



	Embryonic stage			
	E7	**E8.25**	**E9**	**E10.5**
Wave 1 Primitive *	Primitive erythroid progenitors macrophages megakaryocytes	Primitive erythroid progenitors macrophages megakaryocytes	Primitive macrophages	Primitive macrophages
Wave 2 Pro-definitive ★		EMPs	EMPs B lymphoid cells T lymphoid cells	EMPs (yolk sac) B lymphoid cells (AGM, yolk sac) T lymphoid cells (AGM, yolk sac)
Wave 3 Definitive ▲				Hematopoietic stem cells (AGM)

Figure 1.2: Emergence of the haematopoietic system during development occurs in three waves

The emergence of the blood system can be defined into three waves: Wave 1 Primitive, Wave 2 Pro-defintiive and Wave 3 Definitive. The first wave produced primitive erythrocytes, macrophages and megakaryocytes from the blood islands of the extra embryonic yolk sac (Wave 1 locations indicated by asterisks). Wave 2 emerges from haemogenic endothelium first producing erythro-myeloid progenitors (EMPs) at E8.25 in the yolk sac and allantois (indicated by black solid star) and later (E9.0-E10.5) also generating lympoid progenitors with T- and B-cell potential. Wave 3 is the definitive wave with the largest contribution to adult haematopoiesis with the emergence of the haematopoietic stem cell (HSC) from the aorta gonad mesophros region (AGM) from haemogenic endothelium of the dorsal aorta as well as umbilical arteries and placenta (indicated by black triangles). Adapted with permission from Yoder, Nat. Biotech 2014 [106].

1 in surface phenotype, origin and function. Defined initially by their high expression of CD117 (ckit) and CD41, EMPs phenotypically mature during early organogenesis and, by E9.5, also express the granulocyte–monocyte marker CD16/32. Their contribution to the haematopoietic compartment extends far beyond early embryogenesis with EMPs being the main contributor to foetal erythrocytes before HSC-derived erythrocytes are produced [97, 98] and most tissue-resident macrophages of the adult being EMP-derived [99, 100]. Additionally, in contrast to primitive Wave 1, these YS multi-potent progenitors do not derive directly from mesoderm but rather are derived from a specialised subset of endothelium termed the haemogenic endothelium (HE) [101, 102]. Cells undergo an endothelial-to-haematopoietic transition (EHT) and bud out into the vessel lumen in a manner dependent on the transcription factor RUNX1 [103–105].

Wave 2 embryonic haematopoiesis is not limited to EMPs however. Shortly after EMPs are detected in the YS, distinct cells with lymphoid potential are identified at E8.5 (Fig.1.2, Wave 2) [107–109]. Harbouring both B- and T-cell potential, these progenitors contribute to the first adaptive immune system [110]. As with EMPs these progenitors contribute to adult haematopoiesis, with B1 cells,a subset of B-cells, which exclusively arises during Wave 2 haematopoiesis and cannot be replenished by HSCs [111–113]. Although these EMPs and lymphoid progenitors are typically seen as forming the same wave, become detectable at very similar times during development and both derive from a HE [114], their lineage relationship remains unclear and is limited by the *in vitro* assays used to measure potential. What determines whether Wave 2 HE will transition to a EMP or lymphoid progenitor and whether Wave 2 HE has bi-potential or is fate restricted requires further investigation. But increasingly, the pro-definitive, HSC-independent wave of embryonic haematopoiesis is shown to contribute far beyond embryogenesis, and has contributions unmatched by HSCs.

1.2.3 Definitive Wave 3 embryonic haematopoiesis

Although the contribution of HSC-independent haematopoiesis is seemingly greater than originally thought, it is unquestionable that HSCs are vital for adult haematopoiesis. Sitting at the apex of the haematopoietic differentiation hierarchy, these cells are unique in their ability to reconstitute all the necessary haematopoietic lineages in serial transplantation into sub-lethally irradiated mice. This property is reflective of their stem cell characteristics of self-renewal and multipotency [115–118] and grants HSCs their critical importance in bone marrow transplantation: the only curative treatment for patients suffering from a variety of haematological disorders. This has driven a large research focus in recent years to Wave 3 of embryonic haematopoiesis, as this is the only wave during which HSCs emerge.

Wave 3 initiates intra-embryonically from the major embryonic arteries (dorsal aorta, vitelline and umbilical artery) ((Fig.1.2, Wave 3) [119–121]. First detectable at E10.5, HSCs are autonomously generated with the most well studied site of emergence being the ventral wall of the dorsal aorta (DA). Reminiscent of Wave 2 embryonic haematopoiesis, a rare population of HE cells embedded in the vessel wall undergo the EHT and bud into the vessel lumen forming intra-aortic haematopoietic clusters (IAHCs) [122]. Evidence from *ex vivo* cultured reaggregates and transplantation of cells from IAHCs suggests HSCs undergo a multi-step process from pre-HSC through two stages of pro-HSC before maturing to definitive HSCs [123, 124]. IAHCs are heterogeneous containing pre-HSCs (I and II), committed progenitors and EMP-like cells [123, 125, 126]. These studies all rely on *in vitro* culture or computational reconstruction of lineage trajectories and so the progression of HSCs in an undisturbed environment remains to be determined. HSCs seed the foetal liver and placenta between E11-12, where they undergo a large expansion [127, 128]. Their final migration occurs shortly after birth with HSCs settling in the bone marrow niche where they will reside for the duration of the organisms life.

1.3 Stepwise development from pluripotent epiblast to multipotent HSC

1.3.1 Lateral plate mesoderm as the source of HSCs

Above I have primarily discussed the origins of each wave as being defined by where haematopoietic cells are first detectable. However, the emergence of HSCs from HE represents one of the last steps in the development of HSCs. Studies in model organisms from frogs and fish to mouse and humans have demonstrated a step-wise acquisition of cell state with each cell state decision reducing cell potency and increasing commitment (Fig.1.3). Understanding the cell state choices that are made and the molecules that regulate them is critical as it lends insight into the function of haematopoietic cells and is required to recapitulate the process *in vitro*. Furthermore, identifying any similarities and differences in the path HSC ancestors take in comparison to those of the YS waves can inform adjustments to *in vitro* culture conditions. It is useful to define the starting point of this journey when cells are pluripotent (defined by their ability to differentiate into any cell of the embryo proper) in the epiblast. The first patterning epiblast cells undergo on their road to the HSC is to adopt a mesodermal state (Fig.1.3). The mesodermal origins of mammalian HSCs remains controversial with the exact timing, location and signalling incompletely understood.

Much of our understanding of the early programming of the HSC lineage comes from studies in *Xenopus* [129]. Lineage labelling and diploid–triploid chimera studies clearly demonstrate that the ancestral cells of primitive and HSC haematopoiesis are separated very early in development [130, 131]. Transplantation of the trunk LPM selectively labelled the haematopoietic cells of the DA and adult haematopoietic system [131–134] but not the ventral blood islands (equivalent to YS waves in amniotes). This observation has been further validated by selective DA angioblast staining when labelling the LPM will lipophilic fluorescent stain DiI and, the absence of a DA when the LPM is removed [135]. Finally, fate mapping single blastomeres of the 32-cell stage xenopus embryo showed the

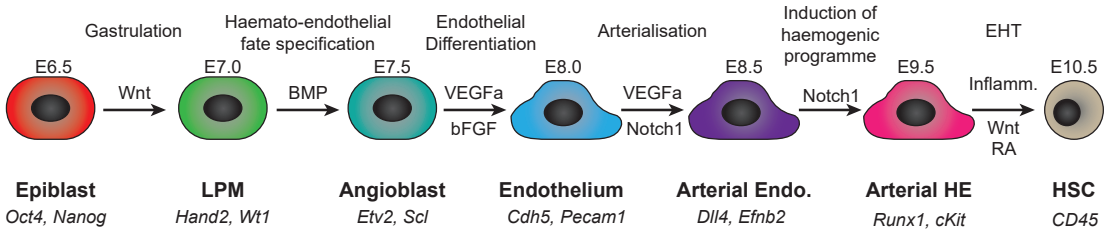


Figure 1.3: Stepwise development from pluripotent epiblast to multipotent HSC
From the pluripotent epiblast before gastrulation (E6.0) cells pass through the mid-primitive streak and are specified to lateral plate mesoderm (LPM) through exposure to Wnt signalling. A subset of the *Hoxb6*-positive LPM is specified into endothelial progenitors (angioblasts) through BMP signalling. These *Scl* and *Etv2* expressing cells migrate towards the midline and differentiate to endothelium under exposure to Vegf between E7.5-E8.0. Almost immediately after, signals such as Notch1 arterialise these endothelium, first discernible by the upregulation of the Notch ligand *Dll4*. Notch and other unknown signals induce a haemogenic programme in a subset of these cells that undergo the endothelial-to-haematopoietic transition (EHT) controlled by the transcription factor *Runx1* and bud into the lumen to form the haematopoietic stem cell. This process is controlled by a number of cell signals including but not limited to pro-inflammatory, Wnt and retinoic acid (RA) signals.

complete separation of these lineages had occurred before gastrulation [130]. The LPM origin of DA endothelium is conserved in amniotes as Scl^{+} angioblasts derived from the LPM have been observed migrating to form the paired DA [136–138]. Unfortunately the haematopoietic contribution of these LPM-derived angioblasts has not been reported.

As discussed earlier (section 1.1.6), the mesodermal origin of Wave 3 haematopoiesis in mammals has primarily been studied through the use of transgenic Cre lines such as Hoxb6 [86]. However, these transgenic lines also label the ExM as shown by labelling in the YS blood islands [86, 88]. Two studies using a haploinsufficient Runx1-MerCreMer transgenic mouse line claim YS derived haemato-endothelial progenitors significantly contribute to Wave 3. Induction of Cre-mediated labelling at E7.5 when Runx1 is exclusively expressed in YS blood islands labelled cells of the E11.5 DA, 7% of phenotypic HSCs and 1-10% of all peripheral blood cells in 9 month old adult mice. This study is controversial because of its use of the haploinsufficient Runx1-Cre model. *Runx1*$^{+/-}$ mESC differentiations have an acceleration of mesodermal commitment and specification to the BL-CFCs and to the hematopoietic lineages as compared to wildtype cells [139]. Consequently, HSC emergence is considerably temporally and spatially altered in *Runx1*$^{+/-}$ embryos [140]. Functional HSCs (serially transplantable) are detectable in the YS and aorta-gonad-mesonephros (AGM) of *Runx1*$^{+/-}$ embryos a day earlier than in wildtype embryos. The altered specification and emergence of HSCs in haploinsufficient embryos bring in to question whether the YS contribution of $Runx1^{+/Cre}$ is physiological or an artifact of altered haematopoietic specification. In attempt to address this question, a second study from the same lab generated a Runx1 transgenic fluorescent reporter that did not disrupt Runx1 function and imaged embryos *ex vivo* [141]. No Runx1-eGFP cells were identified migrating from the YS blood islands and into the embryo. Tanaka *et. al.,* conclude this likely means the extraembryonic contributions of $Runx1^{+}$ therefore reside in the allantois or chorion and cells migrate from there. Given the HSC specification phenotype caused by Runx1 haploinsufficiency, it is more likely these cells are in fact misprogrammed and either aberrantly migrate or emerge *de novo* in the YS unlike in

Runx1-wildtype embryos.

However, the use of Runx1-Cre is a peculiar choice for labelling haemato-endothelial progenitors as it functions in the HE. The phenotype observed in Runx1 haploinsufficient embryos is not reported to involve a change in Runx1 expression domains but rather a change in timing and progenitor output. Therefore, the contribution of YS progenitors observed in Runx1+/Cre fate mapping could label cells that do migrate in WT embryos but that are not yet Runx1 positive (and so not labelled in the Runx1-eYFP). This is supported by a recent study using a newly developed Scl-cerulean by live imaging of *ex vivo* cultured embryos [142]. Lightsheet and multiphoton imaging every 11 minutes for 9 hours reported the E7.5 embryo does not show any cerulean-positive cells but that the cerulean-positive cells of what they define as the distal YS crawl distally to form the paired DA and endocardium. However, how the authors defined the YS-embryo border is unclear and with only cells at the distal border contributing to the DA, where this border is drawn is critically important to conclude a YS origin. The identification of the YS-embryo border at this time is difficult but annotation of the amnion at this time could serve as an important structure to help determine the border. Secondly, although the authors identify the distance between the YS and DA-fated cerulean-positive cells increases, no counter-stain is used to measure the active migration of cerulean positive cells beyond the morphogenic movements of the embryo. Therefore, while contributions from the YS cannot be eliminated, further studies are required to convincingly change the dogma established by studying frog and chicken embryos that HSCs are LPM derived (Fig.1.3).

1.3.2 Transcriptional control of haemato-endothelial fate specification

A number of TFs have been implicated in the specification of the haemato-endothelial lineage from ExM and LPM (Fig.1.4). The master regulator identified as sitting at the apex of specification is ETV2/ER71 [143] with its ablation leading to the complete loss of blood and endothelium (Fig.1.4) [144]. ETV2 is transiently expressed during

embryogenesis between E7.0 and E9.5 with expression initiated through BMP, Notch and Wnt signalling [144]. Consistent with its transient expression, ETV2 is only required for endothelial specification as conditional deletion of *Etv2* with Flk1-Cre resulted in no vascular defects [145]. ETV2's function is to activate the endothelial gene regulatory network through direct binding to gene promoters [146, 147] in endothelial-fated cells but is not required for maintenance of endothelial identity [87]. One of the key downstream targets for blood specification is the the bHLH TF SCL/TAL1. In contrast to ETV2, SCL ablation does not entirely remove endothelial specification but rather results in the absence of all three waves of haematopoiesis [148]. Again, this specification occurs early in development in mesodermal cells, as embryos that have TIE2-Cre mediated excision of SCL showing no defects in haematopoietic emergence. The expression of SCL in *Etv2*$^{-/-}$ mesoderm rescues the loss in haematopoieitic differentiation showing ETV2's role in haematopoietic specification is primarily to activate SCL transcription [145]. Meanwhile, expression of GATA2 and FLI1 (act in a recursive loop with SCL to drive haematopoieitic specification [149]) could not fully restore haematopoietic differentiation [145]. Global analysis of SCL chromatin binding demonstrates it directly binds both endothelial and haematopoietic genes including the critical regulator of EHT, *Runx1* [150, 151]. In addition to activating the haematopoietic program, SCL also functions to actively suppress genes critical for cardiac and paraxial mesoderm specification [150].

1.3.3 From angioblasts to the arterial endothelium of the dorsal aorta

New blood vessels form from one of two mechanisms. Firstly, the arrangement of endothelial precursors (angioblasts) into primitive tube structures followed by lumenisation forms vessels in avascular regions such as the E7.75 embryo. This is in contrast to angiogenesis where new vessels sprout from existing ones. Following their emergence from mesoderm, angioblasts are recruited and organised into primitive vessel structures. The first intra-embryonic vessel to form is the DA between E7.75-E8.25. By E8.25, the DA is lumenised along its entire length [152]. Vascular endothelial growth factor A (VEGFA) has been identified as essential for the migration of angioblasts towards the

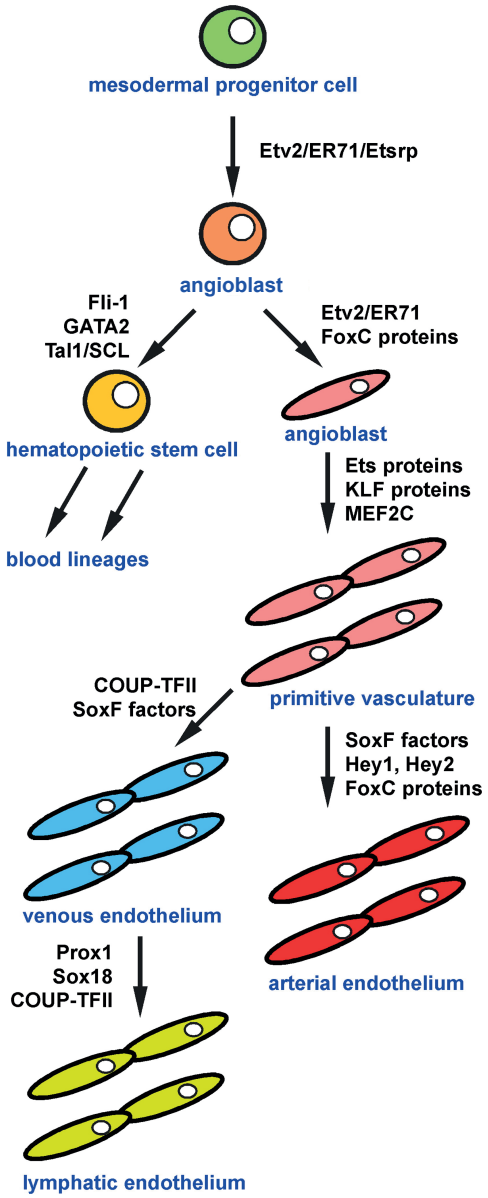


Figure 1.4: Critical transcription factors controlling haemato-endothelial fate specification

The stepwise acquisition and specialisation of endothelial cell state development is regulated by a series of transcription factors. This schematic represents a simplified view of the progression from mesodermal progentior cells, first specialised with ETV2 towards the haemato-endothelial lineage. This model demonstrates the importance of SCL in the establishment of haematopoietic fate while vascular endothelial (right) development is regulated by a different group of regulatory molecules. Reused with permission from De Val, Dev. Cell, 2006 [143].

midline to form the DA [135, 153] and for the formation of extracellular space between cells to form the lumen [152]. VEGFA signals to angioblasts through their expression of VEGF receptor 2 (VEGFR2/Flk1/Kdr) [154]. As cells are recruited to the DA, they differentiate from angioblast to endothelium with upregulation of VE-Cadherin (CD144), Pecam1 (CD31) and Tie2.

Almost immediately after their formation into the paired dorsal aortae, endothelium undergo arterio-venous specification [155]. The adoption of an arterial identity is strongly imparted by Notch signalling [156, 157], with Notch receptors enriched in arterial versus venous endothelium. The intracellular domains of Notch receptors are cleaved following ligand binding and subsequently trafficked to the nucleus to activate target genes such as *Hey1* and *Hey2* encoding the bHLH TFs (Fig.1.4) [158]. This activates the arterial transcriptional programme and seemingly represses expression of venous genes such as *Nr2f2* encoding the master regulator of venous fate COUPTF-II (Fig.1.4) [159]. In the mouse DA, the first signs of arterial commitment are the upregulation of the Notch ligand delta-like ligand 4 (Dll4) at approximately the 4 somite stage [155]. This indicates arterial identity in the DA is initiated before haemogenic endothelium.

Two models have emerged on how arterial and haemogenic endothelium are related. One proposes arterial fate acquisition is essential for HSC emergence, and the other suggests HSC specification is arterial independent and relies on a permissive 'arterial niche' (e.g. high Notch signalling) with HE representing a lineage distinct from than that of arterial endothelium [160]. The shared gene expression programmes favour the former model, but does not eliminate the possibility of two converging lineages of different origins. Additionally, Notch, Vegf and Hedgehog signalling are all implicated in both arterial and HSC development [157, 161–165]. Commonly, the perturbation of a signal will disrupt both arterial and haemogenic endothelium and in the case of Efnb2, specifically affects Wave 3 haematopoiesis [166]. This frequent co-occurrence of disruption could either be because of shared programmes or two independent functions

of the perturbed molecule.

The second, separate lineage model has formed from the observation that the specification of the lineages can be separated on the perturbation of some signalling pathways. For example, Notch1$^{-/-}$ mice exhibit defects in expression of arterial markers (Dll4) with dorsal aorta malformations [156]. However, while HSC activity is reduced in Notch1$^{-/-}$ mice, the expression of key haemogenic and haematopoietic genes were normal [165] and haematopoiesis could be restored *in vitro* by the overexpression of Runx1 [167]. These studies demonstrate that haematopoietic fate has been specified in the absence of normal arterialisation and therefore suggests it is a non-essential step in *de novo* HSC generation. Conversely, the Notch ligand Jag1 was demonstrated to have the reverse impact with normal arterialisation but impaired haematopoiesis [164]. Studies of *Jag1* in the zebrafish AGM has implicated TGFβ in haematopoietic specification separate from arterialisation. While not directly observed in mice, it is clear TGFβ effectors are crucial for Runx1 to orchestrate EHT [168]. These studies identifying signals to separate out the specification of arterial and haematopoietic fates support the notion the two programmes are in fact separate.

However, more recent work focusing on the lineage relationships between arterial and HE rather than the signals specifying them indicates HE derives directly from arterial endothelium. The removal of Runx1 function by genetic deletion or morpholino in *TgBAC(runx1P2:Citrine)* zebrafish embryos resulted in dll4^{+}citrine^{+} DA endothelium [169]. This demonstrates *runx1* transcription is initiated in arterial endothelium and in absence of translation arterial identity is maintained. This is supported in mammalian systems by the pseudotime ordering of scRNA-seq data of cells taken from the mouse and human DA. Again, cells were predicted to first acquire arterial identity before upregulating a haemogenic programme [168, 170, 171].

1.3.4 The endothelial to haematopoietic transition and HSC maturation

The endothelial origin of Wave 3 haematopoiesis is now well accepted thanks to numerous studies using multiple approaches. Landmark studies defined HSC emergence is restricted to the AGM region between E10.5 and E11.5 [120, 172]. Lymphoid progenitors are detectable from early as E8.0 in mouse and Carnegie stage (CS)19 in human embryos in parasplanchnopleura (pSp) explant culture [173, 174]. The first studies to identify HSCs as developing within the major embryonic arteries focused research into how these two lineages are related [119, 175]. Lineage tracing in the chick embryo provided the first evidence of the endothelial origin of HSCs [176]. Fate mapping using two Cre recombinase transgenic mouse models [105, 177] and elegant time-lapse imaging [122, 178–181] confirmed this in mice. These experiments confirmed HSCs underwent an endothelial-to-haematopoietic transition (EHT) as they bud out into the vessel lumen to form IAHCs. Essential for this transition is the core-binding factor family TF, Runx1 [105, 182]. Runx1 directly orchestrates the upregulation of haematopoietic genes by binding at loci primed by SCL, FLI1 and C/EBPβ and initiating a local increase in histone acetylation. Together, these factors induce the upregulation of target genes such as the myeloid master regulator PU.1 [183, 184]. Concomitantly, RUNX1 indirectly suppresses the endothelial programme through binding and activating transcription of *Gfi1* and *Gfi1b* [185].

Following the budding into the vessel lumen, the HSC-fated cell is not immediately endowed with the function used to define definitive adult HSCs. The first HSCs must first undergo maturation in order to gain the ability to fully reconstitute the haematopoietic system of irradiated adult mice after serial transplantation. This stepwise maturation is associated with the acquisition of cell surface markers. Starting at E9.0, CD41+ pro-HSCs are detectable [186]. By E10 CD43+ pre-HSC type I and finally CD45+ pre-HSC type II emerge at E11 [124, 187]. These precursors require maturation in an embryonic or neonatal environment before gaining definitive HSC function.

1.3.5 Cell signalling underlying the emergence of HSCs from HE

Extensive research has been conducted into the cell signals that control HSC emergence from the ventral wall of the DA between E9.5 and E11.5 [188]. Almost all the major signalling pathways [189] have in some way been implicated in the emergence and/or maturation of HSCs in the murine AGM. Studying these regulators is crucial for the recapitulation of the process *in vitro*. Signalling pathways controlling Runx1 regulation, HSC emergence and maturation have all be identified. With signals being so numerous and their interactions so complex, only a brief summary of signalling is provided here.

The induction of *Runx1* in Wave 3 HE is at least in part controlled by Notch signalling with *Runx1* overexpression rescuing the loss of intraembryonic haematopoiesis in Notch1 mutant mice [165, 167, 190]. Furthermore, the Notch effector gene *Hes1* also complexes with RUNX1 in HE to mediate transactivation of RUNX1 target genes [167].

Many of the signals received by the DA HE integrate and regulate each other. For example, retinoic acid (RA) signalling directly promotes HSC development from HE in the mouse embryo but in doing so inhibits Wnt signalling [191]. β-catenin, the Wnt effector, is selectively localised to the nucleus in a small population of DA cells and its genetic deletion in endothelium (by VE-Cadherin-Cre) eliminates HSC emergence. However, haematopoietic-specific genetic deletion (with Vav-Cre) does not [192]. This suggests that Wnt signalling is required for the EHT but not thereafter and identifies two roles for β-catenin during haemato-endothelial development as it is also required for vasculogenesis [193, 194].

Similarly, the interaction between Notch signalling and pro-inflammatory signalling regulates HSC emergence. Modulating the expression of both interferon gamma (IFN-γ) and tumour necrosis factor alpha (TNFα) results in a change in the number of HSCs emerging in the DA of zebrafish. [195, 196]. In both cases, the DA was specified normally but time lapse imaging of fluorescent transgenic lines demonstrated the positive

regulation of IAHC emergence (and not proliferation or survival) by pro-inflammatory signals. The mechanism of action of the two signals was distinct however. IFN-γ was secreted by HE in a self-regulatory loop binding to IFNR1 on its own cell surface. Conversely, TNFα, was expressed by locally invaded primitive neutrophils with TNFR2 on the surface of non-HE endothelial cells. This consequently, upregulated the Notch ligand Jag1 which in turn activated Notch1 signalling in HE. The regulation of HSC emergence by inflammatory signals in the mouse AGM was shown later with primitive myeloid cells again contributing reguatory signals [197]. In conclusion, these signals highlight the importance in understanding the full repertoire of cellular neighbours around HEPs as they can mediate signalling to the HE and emit different signals that converge on the same outcome.

Finally, the role of BMP signalling has been unfolded in two stages. Firstly, through expression analysis of AGM stromal lines, BMP4 was identified as a potential regulator of HSC development. It was found to support the repopulating ability of HSCs in AGM explant cultures [198]. Examining the expression of Bmp4 in the E11.0 AGM revealed its expression in the sub-aortic mesenchyme [198, 199]. It wasn't until scRNA-seq analysis of the E9.5 and E10.5 AGMs that BMPER (a BMP signalling antagonist) was identified as important for HSC maturation [200]. These data could be reconciled with earlier observations of the positive effect of BMP signalling through the observation that BMPER expression was induced by BMP. This is a good example of self-regulation in the AGM and highlights the importance of studying signalling antagonists as well as activators and that the same signalling pathway can have vastly different roles in only a short developmental time frame.

1.4 *In vitro* differentiation of the haematopoietic lineages from hPSCs

1.4.1 Motivation and rationale for developing hPSC differentiation systems

While much of the scientific effort to understand the development of the haematopoietic lineages in model organisms comes from curiosity to understand our own development

and the cellular mechanisms controlling cell state diversification, much of this work also aims to provide the wherewithal to recapitulate these embryonic development from the *in vitro* differentiation of human pluripotent stem cells (hPSCs). As with many stem cell fields, research striving to generate adult cells and tissues *in vitro* hopes to offer new models of development but also serve a source of tissue for translational research and clinical therapeutics. hPSC derived tissues offer more faithful models of disease than animal models and offer platforms for high-throughput drug screening in cases where human tissue is scarce. Furthermore, increasing focus is being devoted to the applications of these tissues to regenerative medicine, offering potential sources of HLA-matched cells or tissue for the treatment and cure of complex, multi-factorial human diseases.

Researchers and clinicians studying and treating haematological diseases have long sought such sources of healthy human cells with HSC transplants offering curative potential to many diseases. However, these therapies are often limited by the rarity of HLA matched donors [201]. Although the HSC represents the long-term functional unit of regeneration during a bone marrow transplant, the role of other haematopoietic compartments should not be underestimated. For example, donor-derived immune cells perform critical graft-versus-disease functions [202]. While HSC expansion *in vitro* represents one solution to these issues [203], another possibility would be to generate HSCs *de novo* [201].

Therefore, three complementary approaches to generate HSCs *de novo* have been taken with varying degrees of success. One approach is to reprogram terminally differentiated or hPSC-derived embryonic cells through the overexpression of TFs critical for HSC emergence to produce so-called induced HSCs[204–209]. The second strategy is to inject pluripotent stem cells into a mouse to form teratomas that form HSCs [210, 211]. The final approach is to mimic the stepwise development of HSCs *in vivo* by performing directed differentiation of PSCs through the stepwise-addition of cytokines to their media [55, 212–216]. I focus on the third approach for two reasons. Firstly, to

produce HSCs for clinical purposes, direct differentiation is the safest with no forced expression or cross-species contamination required. Secondly, it offers the best model in which to study the extrinsic and intrinsic signals controlling blood specification.

1.4.2 Directing differentiation away from primitive and towards definitive haematopoiesis during mesoderm patterning

Directed differentiation of hPSCs is controlled by the cytokines and small molecules present in the culture media. Initially, many groups differentiated hPSCs on layers of supporting feeder cells in serum-containing media to stabilise cultures and promote cell survival. However, these components remove the ability of experimenters to control the differentiation conditions, with variability between batches of serum and inconsistencies in feeder cells. Therefore, many groups opted to establish feeder and serum-free protocols with all the required nutrients and cytokines provided in the media [217]. This level of control allows researchers to more accurately mimic the signalling steps required *in vivo* to specify HSCs. To date, no groups have described the genesis of functional (serial transplantation) HSCs in this manner. Each step of differentiation comes with difficult and sometimes contentious decisions on how to direct towards the HSC lineage. These decisions are informed by our knowledge of HSC development but importantly are influenced by how this is different from Wave 2 haematopoietic specification. In contrast to embryogenesis, Wave 2 and Wave 3 progenitors formed *in vitro* from hPSCs cannot be distinguished using anatomy and embryonic staging.

With increasing evidence that cell fates are specified soon after gastrulation [218] and with primitive and HSC lineages distinct from very early in development [130], it is critical that the first steps of differentiation set cells along the correct path. Conditions to induce a primitive streak state in hPSCs that accurately mimics the signalling experienced by LPM fated cells of the mid-primitive streak during gastrulation must be identified. Early work first defined the necessary steps to differentiate hPSCs to primitive haematopoietic cells [55, 219, 220] and tested the intersection of noncanonical Wnt, Nodal and BMP

signalling on primitive streak identity and mesoderm patterning [219].

Building on this, two seminal studies found primitive and definitive progenitors are differently specified during mesoderm induction (Fig.1.5a) [214, 221]. The inverse actions of Wnt and Activin/Nodal signalling during mesoderm patterning could be manipulated to control the proportion of primitive versus (pro-) definitive haematopoiesis later in differentiation. Inhibition of Nodal signalling from differentiation day (D)2 to D4 with the ACTIVIN receptor-like kinase (ALK) small molecule inhibitor SB-431542 (SB) selectively inhibited the development of primitive haematopoietic progenitors [221]. Conversely, the activation of canonical Wnt signalling with the GSK-3 inhibitor CHIR99021 (CHIR) inhibited the development of the mesodermal progenitors of the primitive wave and increased the prevalence of mesoderm with definitive potential. The manipulation of these signals is reminiscent of the graded Wnt and Nodal signalling of the early embryo [2]. Both studies identify mesodermal patterning conditions and conclude they form an essential step in the development of hPSC differentiation protocols directed towards definitive HSPCs. Both studies utilised the identification of T-lymphocyte potential as a measure of definitive haematopoiesis. Thus, mesoderm patterning will likely form a critical factor in HSC specification but whether these conditions are sufficient to robustly generate such cells is still to be determined. Indeed, the mesoderm generated here and the T-lymphocyte progenitors it produces could conceivably represent ExM that forms YS-type lymphoid progenitors.

Another report also noted the importance of Wnt and Nodal activation/inhibition during the mesodermal differentiation of hPSCs. Transcriptional profiling of $CD34^+$ cells from culture conditions without Wnt/Nodal manipulations [222] demonstrated they lacked expression of genes in the HOXA cluster as compared to $CD34^+$ cord blood progenitors [213]. With HOX expression initiated during mesodermal patterning, Ng *et. al.,* aimed to alter conditions during this time. Combining Wnt activation and Nodal inhibition with CHIR and SB at previously used concentrations [214, 221] greatly increased the expression of a number of HOXA genes in $CD34^+$ HE and haematopoietic

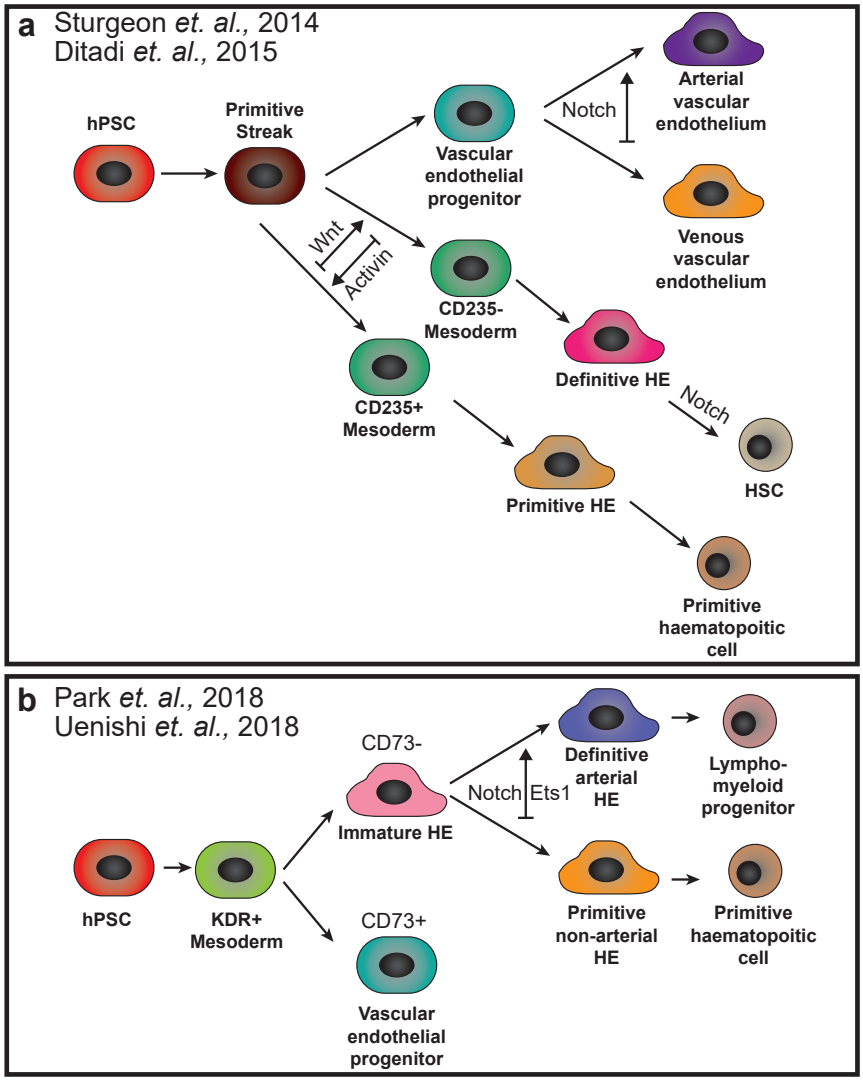


Figure 1.5: Two models for the development of primitive, pro-definitive and vascular endothelium from hPSCs

a Summary of the lineage relationships described by Sturgeon, *et. al.*, Nat. Biotech. 2014 [214] and Ditadi, *et. al.*, Nat. Cell Bio. 2015 [223]. Cell populations are labelled according to the authors model. Patterning of mesoderm with the inverse actions of Wnt and Activin signalling direct differentiation towards a CD235a+ or CD235- mesoderm. CD235a+ mesoderm has definitive HE potential that in the presence of Notch signalling can undergo EHT to definitive haematopoietic progenitors. This Notch activation is independent of arterial identity which in this model is described as a separate lineage with venous endothelium making up vascular endothelium. **b** Model of HE arterialisation based on Park, *et. al.*, Cell Rep. 2018 [224] and Uenishi, *et. al.*, Nat. Comms. 2018 [225]. In this model, immature HE can be separated from vascular endothelium by its CD73 negativity. This immature HE can be arterialised with either Notch signalling or Ets1 overexpression to form definitive HE with lympho-myeloid potential. In the absence of Notch signalling, HE does not acquire an arterial identity and has primitive blood potential only.

progenitors [213]. In contrast to previous studies however, T-lymphocyte is not used to define these as definitive progenitors. Instead, the authors use a novel SOX17 transgenic reporter line to report SOX17 expression as a marker of the aorta, show fetal-stage erythroid potential and show transcriptional similarity with human AGM and FL populations.

These experiments certainly demonstrate the generation of a pro-definitive programme but there are still questions as to the similarity with the human AGM. Although the mammalian aorta does express Sox17, it is not specific as the YS vasculature also widely expresses it, with important roles in arterial identity specification and maintenance widely throughout the embryo [226]. The transcriptional data comparing hPSC derived CD34+ HSPCs to the AGM of human embryos are more convincing. However, the transcriptional programmes of the Wave 2 YS and Wave 3 AGM HE are likely much more similar than they are different. Therefore the transcriptional similarity between *in vitro* and embryonically derived cells could represent this transcriptional core and so would also ideally be compared to the YS HE and haematopoietic progenitors. This comparison would aid in defining if the mesodermal patterning stages induce LPM rather than ExM and therefore the likely cause of the absence of HSC generation from these conditions can be assigned to other stages.

1.4.3 The relationship between hPSC derived arterial and haemogenic endothelium

Patterning mesoderm is thought to be the major step in separating the YS and AGM waves of embryonic haematopoiesis *in vitro*. However, this only represents the first stage of a multi-step differentiation process. Through modulating Wnt and Nodal signalling, mesoderm expressing the VEGF receptor 2 (VEGFR2)/FLK1/KDR is expanded and differentiated towards an endothelial fate. Through the addition of angiogenic cytokines VEGFA and bFGF [227] from approximately D2-4, differentiation from endothelial progenitor to endothelium can be induced. Endothelial differentiation is monitored through the flow cytometric measurement of the upregulation of endothelial markers

such as VE-Cadherin (CD144/CDH5) and PECAM1 (CD31), characteristic morphological changes, and functional assays of tubule formation. These differentiation steps are standardised across protocols with some groups also including stem cell factor (SCF) and BMP during this time. Endothelium is then either FACS sorted (often excluding $CD43^+$ primitive progenitors) [214, 223] or left unperturbed and unenriched for haemogenic induction [213]. FACS sorting is beneficial as it allows monitoring of differentiation progress between independent experiments, removes any primitive progenitors ($CD43^+$) that have been specified and gives finer control of cell signalling by removing non-endothelial niche cells. In contrast, keeping cultures intact ensures cells are not disturbed by dissociation and keeps the surrounding niche cells in culture that could provide crucial signalling unknown to the researcher. In either setting, endothelial cells are exposed to a broad cocktail of cytokines aiming to induce a haemogenic programme. As discussed for the equivalent step *in vivo*, the mechanism that grants a subset of these endothelial cells with haemogenic competency is unclear. However, taking a pragmatic approach of adding a wide spectrum of haemogenic and haematopoietic cytokines reportedly results in the budding of round haematopoietic cells from the adherent endothelium in tandem with the upregulation of RUNX1 after 4-5 days [213, 214, 223].

The first human HE derived from hPSCs did not exhibit arterial identity but could be separated from non-HE by its lack of CD73 expression and separated by the haematopoietic cells it produced by CD43 negativity. Therefore, human HE was defined by $CD144^+CD43^-CD73^-$ and was shown to give rise to unipotent haematopoietic cells [228]. With the aim of increasing the definitive HE output in cultures, this HE population was isolated and arterialised either through overexpression of ETS1 or through treatment with NOTCH ((Fig.1.5b) [224, 225]. Interestingly, this work reverses the current dogma on the order of arterialisation, with HE being arterialised rather than arterial endothelium becoming haemogenic (as discussed in section 1.3.3, Fig.1.3). Similarly, Ditadi *et., al*, observe the NOTCH dependent haematopoietic development in hESC derived HE. In contrast however, they do not identify arterial HE and propose the two lineages are separate (Fig.1.5a). However, DLL4 expression was not profiled across differentiation

and as such the definitive HE as defined by $CD34^{+}CD144^{+}CD73^{-}CD184^{-}DLL4^{-}CD45^{-}$ could conceivably represent a relatively mature HE that has already downregulated DLL4 in a RUNX1-dependent manner as observed in zebrafish, mouse and human [169–171, 216]. Whether arterialisation of endothelium occurs before or after the induction of the haemogenic programme is a consequence of *in vitro* differentiation, a YS associated mechanism or human specific phenomenon remains unknown.

1.4.4 The likely causes of the inability to form HSCs *de novo* from current hPSC differentiation protocols and new ways to address them

Studying human haematopoietic development using *in vitro* differentiation of hPSCs has provided great insight in the development of primitive versus (pro-)definitive waves. However, there has yet to be the development of a protocol that can form HSCs *de novo*. Representing the functional unit of transplantation and the main pool of human haematopoietic cells through life, *de novo* HSC generation remains the top priority for many in the field. Here I will summarise three of the possible causes for absence of HSC emergence *in vitro*. These are formed through the identification of differences or unknowns as compared to what we have learned from the embryo.

Firstly, is the possibility that hPSC-derived HE from current cultures does not accurately reflect the Wave 3 HE of the major arteries but is Wave 2 HE. Studying the embryo has demonstrated that HSCs derive exclusively from the HE of the embryonic arteries with no *de novo* contribution from the Wave 2 YS HE . Furthermore, the HE that gives rise to the EMP wave is distinguishable from the HE that forms HSCs [229]. While this has not been formally shown for the HE of lymphoid progenitors and HSCs, I hypothesise these HE are distinct in their developmental origin and potential. Therefore, new differentiation protocols would need to identify novel mesoderm patterning conditions that direct differentiation away from ExM and YS endothelium and towards LPM and intraembryonic endothelium. The definitive haemangioblasts identified in *Xenopus* originate from a restricted LPM region (Fig.1.6) [131, 133, 134, 230] showing the LPM is patterned along the A-P axis and that this is critical for HSC emergence [129].

Although the definitive haemangioblast has not identified in other organisms and so their A-P patterning cannot be determined, the IM that lies immediately medial to the definitive haemangioblasts (*Xenopus*) and DA progenitors (zebrafish and mouse) is patterned along the A-P axis [70, 231–233] (Fig.1.6). In *Xenopus*, the definitive haemangioblasts are only present in the region adjacent to the IM that produces the mesonephric duct in the trunk region. This results in HSC-competent HE restricted to the region directly adjacent to the mesonephric duct and not extending further anteriorly or posteriorly in the DA. Similarly in mice, the DA does not contain HE potential (haematopoietic clusters) along the entire A-P axis but is restricted to the trunk region (Fig.1.6) [234]. As discussed above (section 1.1.6), hPSC differentiation protocols towards IM and its derivatives have identified how to modulate signalling to pattern IM to anterior and posterior IM [70, 72, 77]. Whether these culture conditions also contain HEPs that would be correctly patterned along the A-P axis and directed away from ExM and YS waves remains to be determined.

A second possible cause of the absence of HSC generation from current hPSC differentiations is the order of arterialisation during differentiation protocols. Arterialisation and signals critical for this process such as NOTCH have been demonstrated as critical for the emergence of definitive versus primitive blood [160, 223–225]. These studies either do not define the order of arterilisation/Notch upregulation [223] or arterialise HE [224, 225] (Fig.1.5). As discussed above (section 1.4.3), this goes against the order identified *in vivo* in zebrafish, mouse and human DA HE [168–171]. The importance of this reversal in order is unclear but could feasibly alter the potential of a cell to give rise to HE. Therefore, studies aiming to more accurately follow the *in vivo* order of arterialisation could better understand the role of arterialisation and more exactly recapitulate the path taken by HSC ancestors.

Finally, there is the possibility that hPSC derived HE is correctly patterned and is a faithful recapitulation of Wave 3 HE but that the signals required for HSC emergence, maturation, expansion and self-replication are not mimiced *in vitro*. The faithful

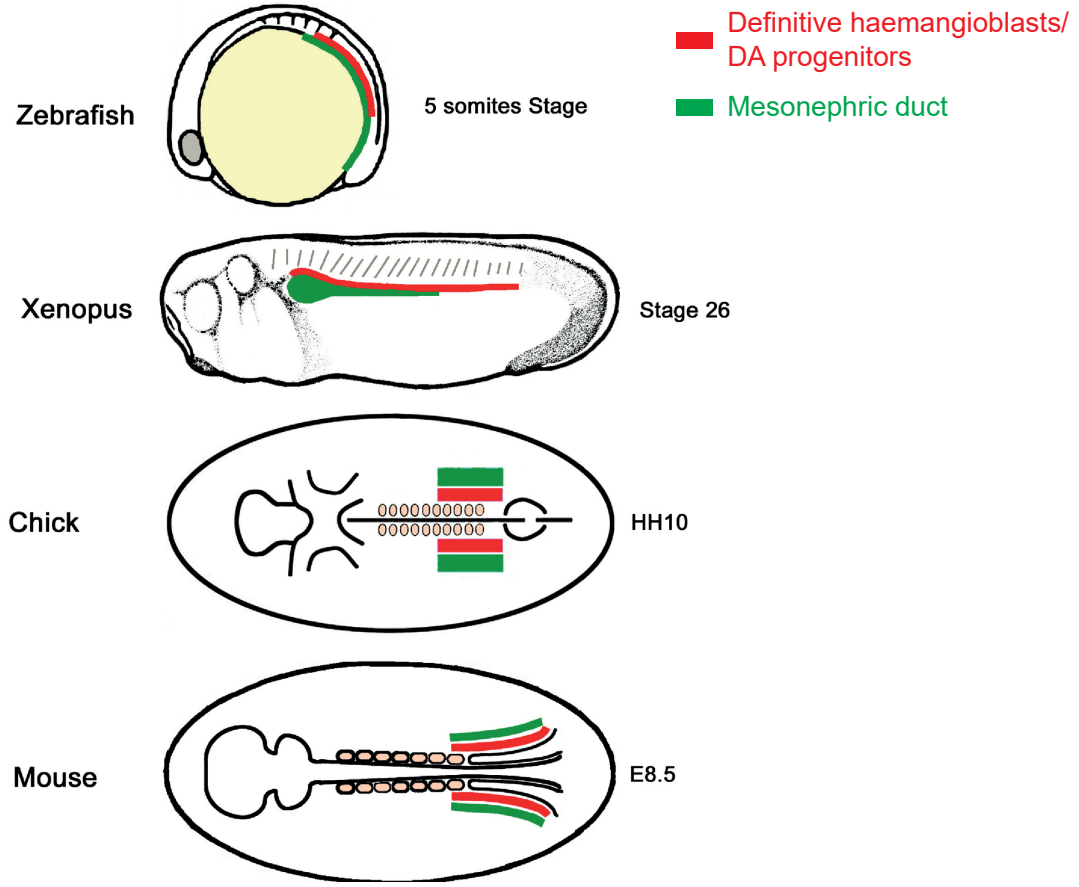


Figure 1.6: The progenitors of Wave 3 HE are restricted along the A-P axis and lie adjacent to the nephric duct
Shown are four embryos from four species of roughly the equivalent stage. Definitive haemangioblasts are in the dorsal lateral plate of Xenopus embryos (stage 26) and the progenitors of the trunkal DA are present as bilateral strips in zebrafish (5 somite stage), chick (Hamburger and Hamilton stage 10) and mouse (E8.5) (all shown in red). These regions are conserved across evolution and suggest this restricted region along the A-P axis is critical for Wave 3 programming. Lying adjacent to these populations in all four species is the nephric duct (wolffian duct) which has been experimentally demonstrated to be patterned along the A-P axis. Our knowledge of intermediate mesoderm and nephric duct A-P patterning could be used to derive suitably patterned angioblasts *in vitro*. Adapted with permission from Ciau-Uitz & Patient, FEBS Letters 2016 [129].

recapitulation is supported by the transcriptional correlation with human embryo derived AGM HE [213]. The AGM is a massively complex signalling niche with multiple cellular, extracellular matrix and mechanical components already identified. More global analyses using cell communication inference from transcriptomic data have the potential to find even more [235–237]. It is very difficult to recapitulate each of these factors at the correct concentration and timing and thus the commonly used one size fits all approach of current methods using static media containing unchanging cocktail of cytokines over several days cannot capture the dynamic environment of the AGM. This can in part by solved by relinquishing some control over signalling with multi-lineage organoids/embryoid body differentiations [171, 213] with niche cells providing some of the necessary signals to HEP differentiation. If HSCs could be generated by recreating the complex AGM environment in this way, it would provide a platform with which to study the necessary cells/signals in more accessible system.

1.5 Identifying a suitable spatial transcriptomic method depends on practical considerations

Although trying to recreate embryonic niches *in vitro* can help define the requirements for HSC production, this may not reveal all critical regulators/niches and, therefore, studies of the embryo remain essential. The use of single cell transcriptomics has revolutionised how researchers can study heterogeneity between individual cells. These methods rely on the dissociation of tissue and consequently cannot associate a cell with its location within a tissue/organ. In order to understand how haemato-endothelial lineages are specified during embryogenesis, it is critical to understand the niche in which they develop. A major component of the niche is the cellular neighbourhood. Dissociating tissues necessitates the removal of this fundamental aspect of cell identity. However, the next generation of high-throughput techniques promise to solve this issue by retaining the location of the collected transcriptomes. These collection of methodologies can be classified into groups depending on their underlying technology: 1) single molecule fluorescence *in situ* hybridisation (smFISH) 2) spatial barcoding combined

with Illumina sequencing and 3) laser capture microdissection (LCM) combined with Illumina sequencing (Fig.1.7). Each class of methods has varying levels of throughput, specificity, challenge and gene detection. In this section, I will introduce some of the most widely published methods, compare them and define some parameters with which to select an appropriate method.

1.5.1 Spatial transcriptomic methodologies using single molecule RNA FISH

smFISH methods differ from the other two classes in their method of gene detection. Rather than sequencing cDNA libraries, smFISH methods count the number of transcripts within a cell by imaging individual molecules as diffraction limited spots. The main variability in this class of methods comes from how single molecules are fluorescent labelled. Traditionally, smFISH used fluorescently conjugated RNA probes tiled along a transcript to label them. More recently, the next generation of smFISH methods use split DNA probes with signal amplification to overcome the low signal : noise ratio and frequent off target effects [238–240]. These methods are relatively easy to establish and can be performed on nearly any tissue from any organism. Using 5 different initiation sequences coded to 5 different fluorescent channels, up to 5 genes could be quantified by spot counting using these methods alone. If, however more genes are to be studied, multiplexing is required. The simplest form of multiplexing is to remove DNA probes and their fluorochromes and restain with different genes. Stripping can be performed either enzymatically with DNaseI [241] chemically with high formamide concentrations [242]. This raises the number of genes detectable in a single tissue section linearly with the number of rounds of imaging performed (Fig.1.7, column 1). Gene counts can then be generated through image analysis pipelines that count the number of genes in a defined cell volume as performed in the chicken neuroectoderm in spatial genomics analysis (SGA). This technique is great for its relative accessibility as it can be performed using standard scientific equipment such as laser scanning confocal microscopes and is suitable in cases where known cell types need to be mapped or when 6-40 genes are

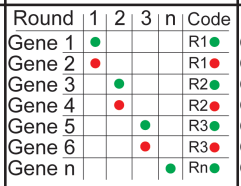
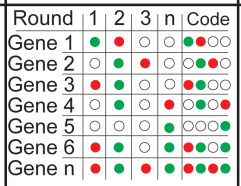
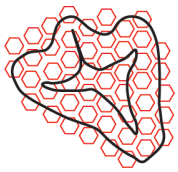
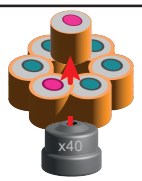
Class	smFISH		Spatial Barcoding	Laser capture microdissection
	Linear Multiplexed	Combinatorial		
Graphical summary				
Example	SGA, seqFISH	MERFISH,seqFISH+	10x Visium,Slide-seq	LCM-seq, Geo-seq
Genes Detected	1-40	100-Transcriptome	Transcriptome	Transcriptome
Sensitivity	>95%	50% - 85%	7% - 12%	NA
Difficulty to establish	+	+++	+	++
Spatial resolution	Subcellular	Subcellular	10 - 50 cells	1 - 10 cells
Advantages	+ No specialist equipment + High sensitivity + Subcellular resolution	+ Highly multiplexed + Subcellular resolution	+ Large fields of view + Whole transcriptome + Ease of setup & analysis	+ User control + Whole transcriptome
Disadvantages	- Low number of genes - Biased gene selection - Limited field of view	- Need specialist equipment - Cost - Limited field of view	- Low spatial resolution - Multiple cells per spot - Low sensitivity	- Technical expertise required - Low throughput - Requires staining to identify cells of interest
Good for	Few genes needed Low expressed genes	High number of genes Atlasing tissues	Defining tissue domains	Discovery analysis of targeted cells

Figure 1.7: Comparison of the three classes of spatial transcriptomics methods
Spatial transcriptomic methods can be grouped into three classes (smFISH based, spatial barcoding and laser capture mircodissection based) and vary on a number of practical considerations (rows). smFISH methods can be divided into low throughput, linear multiplexing where each gene is imaged once and the high throuput, combinatorial multiplexing where each gene is imaged multiple times with the colour and round of imaging combination unique to each gene (code). Spatial barcoding uses spots (red hexagons) each with their own unique molecular identifier overlayed on tissue sections (black outline) to label RNA molecules by their location within the tissue during reverse transcription. Laser capture microdissection techniques cut and catapult user defined regions for lysis and RNA-seq analysis. Difficulty to establish is scaled from relatively low (+) to relatively high (+++).

needed.

More advanced smFISH methods utilise microfluidic devices to perform staining on custom-built microscopes to significantly increase the number of genes detected. To overcome optical crowding (too many spots in a given space cannot be resolved due to overcrowding), these technologies utilise a sequential barcoding approach to identify genes where each gene will be imaged in multiple rounds, requiring single-molecule registration very difficult to achieve with computional registration alone. Both multiplexed error-robust FISH (MERFISH) [243] and sequential FISH plus (seqFISH+) [242] encode each gene with multiple initiator probes (Fig.1.7, column 2) that are imaged in different rounds and computationally error-corrected to account for imperfect hybridisation. This can cause significant drop in detection of genes from 51% for seqFISH+ [242] and 20% in MERFISH [243]. With no need to register single molecules in SGA, gene detection is much greater (close to 100% [244]) and so is more appropriate for the detection of lowly expressed genes. This much lower detection efficiency is traded for the ability to profile magnitudes higher number of genes (>10,000). Therefore, these methods are currently best suited for atlasing projects where large consortia can intersect expertise on cell biology, imaging and image analysis to map cell types with high confidence and the possibility to perform discovery scale analyses. Indeed, seqFISH has recently been applied to probe 387 genes in mouse embryos at the 8-12 somite stage (ss) [39]. These genes were selected to allow computational mapping of their scRNA-seq atlas [27] and generate a reference embryo for the community to interrogate the spatial distribution of cell types and selected genes.

1.5.2 Spatial transcriptomic methodologies using spatial barcoding technologies with RNA-sequencing

The use of smFISH to count transcripts provides subcellular resolution and so in addition to providing spatial relationship between cells, could be used to describe the spatial relationship between RNA species in the same cell [245]. This is in stark contrast to the second class of spatial transcriptomics methods - spatial barcoding. These methods

print spots of anchored cDNA barcodes (each spot has a different barcode) onto glass slides that complex with mRNA from a tissue section collected onto the cDNA spots. Reverse transcription is performed *in situ* before release and performing cDNA library preparation for Illumina sequencing. Reads can be remapped to spots using the cDNA barcode coordinate. With their high level of commercialisation (10x Visium being the most widely used) as compared to other spatial transcriptome methods and relative ease of use, these are likely the most widely used methods. Their limitation comes from increasingly low RNA capture efficiency (sensitivity) with smaller spot sizes. This necessitates large (55µm in the case of 10x visium) spots to capture sufficient material (Fig.1.7, column 3). A large array of spots can, however, allow a large field of view. Low resolution is in part being addressed computationally through the mapping of scRNA-seq datasets with programmes such as Cell2Location [246]. New generations of this concept are constantly improving sensitivity and resolution and so likely represents the commonest selection of future spatial transcriptomics studies [247, 248].

1.5.3 Spatial transcriptomic methodologies using laser capture microdissection (LCM) with RNA-sequencing

While these high-throughput but low resolution technologies continue to improve and become more accessible, others have used existing technologies to collect cells for transcriptomics. Although of lower throughput, using LCM to dissect tissue has the advantage of being user controlled and has the potential to selectively isolate single cells for transcriptome scale profiling (Fig.1.7, column 4). Following cutting, the desired region (1-cell up to large sheets of tissue) is catapulted into a tube ready for lysis and Smart-seq2 cDNA library preparation [249, 250]. The difficulty with this technology comes from the need to stain and identify the cells of interest without degrading RNA, but when properly optimised can detect 10,000 genes from single cells. This approach is best used for projects aiming to profile the heterogeneity of small regions or rare cells that benefit from unbiased gene selection.

In conclusion, a range of spatial transcriptomic methodologies are available and under development. These methods can be challenging to establish and expensive to run and so practical considerations must be made when selecting a method. For profiling haemato-endothelial development during mouse-organogenesis, multiplexed smFISH (e.g. SGA) offers excellent selection to map progenitors and their neighbours that can be defined by a small number of genes. Once mapped, LCM-seq could provide an excellent tool to selectively isolate differently located progenitors to discover any differences in transcriptional programmes and signalling from neighbouring cells.

1.6 Thesis aims and outline

The ability to generate HSCs *de novo* from the direct differentiation of hPSCs could provide a source of HSCs for treatment of haematological diseases, and act as an excellent experimental model with which to study human haematopoietic development. To date, no protocol has been developed for the de novo generation of HSCs from human pluripotent stem cells (hPSCs), likely due to incorrect anterior-posterior patterning, arterialisation kinetics and/or missing signals required for HSC emergence. This work focuses on the first two possibilities.

Work from multiple species has defined HSCs ancestors as residing within a restricted region along the anterior-posterior axis [129]. The intermediate mesoderm derived mesonephric duct is also specifically patterned within this region [129]. Therefore, to correctly pattern HEPs from hPSCs, I will first establish an anterior intermediate mesoderm differentiation protocol previously shown to form the mesonephric duct. Once established, I will systematically identify and differentiate HEPs produced in these cultures. As these HEPs will have experienced the same signalling conditions as the progenitors to the mesonephric duct, they should be patterned in the same way. This will allow us to direct the differentiation trajectory away from YS waves of haematopoiesis and towards Wave 3. My next aim will be to direct the differentiation of these HEPs towards haematopoietic cells in a stepwise manner, reflecting that of the embryo. I will

monitor the acquisition of endothelial and arterial endothelial fates before attempting to induce a haemogenic programme. The work addressing these aims is reported in Chapter 3.

The actions of TFs such as ETV2 and SCL, essential for haemato-endothelial fate specification, acting in mesodermal progenitors provides a strong argument for the early fate specification of haemato-endothelial progenitors. Reports detailing the expression patterns of these TFs [251] and the timely deletion of their function [252], suggest that haemato-endothelial fate specification occurs between E7.5 and E8.5 for intra-embryonic HEPs [218]. HEPs and their microenvironment are poorly described during this time in mouse embryogenesis. Therefore, my third aim is to characterise these HEPs and their cellular microenvironment *in situ* through establishing and applying a quantitative, multiplexed smRNA-FISH approach. These findings will describe the immediate cellular neighbourhood of intra-embryonic cells undergoing haemato-endothelial fate specification and will be presented in Chapter 4.

Finally, having described the variation in HEP location and the diversity in their cellular microenvironment, I aim to describe the cellular signalling received by HEPs at this time. Through establishing a high resolution spatial transcriptomics method in LCM-seq and combining it with scRNA-seq data of the E8.0 mouse embryonic posterior, I aim to perform cell signalling analysis. This will give us a global picture of the signalling environment of HEPs during haemato-endothelial fate specification and aims to identify novel candidates for functional validation. These data are presented in Chapter 5.

Chapter 2

Materials and Methods

Contents

2.1 Thawing, freezing and maintenance of human embryonic stem cell lines . . . 46
2.1.1 Mycoplasma testing . . . 46
2.1.2 Matrigel coating of culture plates . . . 47
2.1.3 Passaging human embryonic stem cells . . . 47
2.1.4 Thawing human embryonic stem cells . . . 47
2.1.5 Freezing human embryonic stem cells . . . 48
2.2 *In vitro* differentiation of human embryonic stem cells to intermediate mesoderm in defined medium . . . 48
2.3 *In vitro* differentiation of human embryonic stem cells to haemogenic endothelium in defined medium . . . 49
2.4 Freezing, thawing and passaging of human umbilical vein endothelial cells (HUVECs) . . . 49
2.5 Reverse transcription quantitative polymerase chain reaction (RT-qPCR) . . . 50
2.6 Flow cytometry and fluorescent activated cell sorting (FACS) . 52
2.7 Tubule formation assay . . . 53
2.8 Immunofluorescence of hESC differentiation cultures . . . 53
2.9 Mouse colony maintenance and genotyping . . . 54
2.10 Timed matings and embryo collection . . . 55
2.11 Embryo preparation for cryosectioning . . . 55
2.12 Hybridisation chain reaction version 3.0 (HCR v3.0) . . . 56
2.12.1 Split-initiator probes, hairpin amplifiers and buffers . . 56
2.12.2 Digital HCR (dHCR) . . . 56
2.12.3 Quantitative HCR (qHCR) . . . 58
2.13 Probe and signal stripping . . . 59
2.14 Cell membrane staining . . . 59
2.15 Image alignment . . . 60
2.16 3-D cell segmentation . . . 60
2.17 dHCR spot counting . . . 61
2.18 Hierarchical clustering and spatial remapping of clusters . . . 63
2.19 Laser capture micro-dissection . . . 63
2.19.1 Sectioning of fresh frozen mouse embryos . . . 63
2.19.2 Section preparation and rapid staining for LCM . . . 64
2.19.3 Laser capture micro-dissection and catapulting . . . 64
2.19.4 Reverse transcription and pre-amplification of single cell lysates . . . 65
2.20 10x Chromium scRNA-seq . . . 66

2.20.1 Mouse embryo dissection and dissociation for 10x Chromium 66
2.20.2 10x Chromium library preparation 66
2.20.3 Illumina next generation sequencing 67
2.20.4 Sequencing data pre-processing and quality control . . . 67
2.20.5 Normalisation and clustering 10x scRNA-seq 68
2.20.6 Removal of doublets from scRNA-seq data 69
2.21 CellChat cellular communication prediction 69

2.1 Thawing, freezing and maintenance of human embryonic stem cell lines

Human embryonic stem cell (hESC) line H9 was purchased from WiCell, USA, and hESC line SHEF-4 obtained from our collaborator. All hESC work was performed in accordance with MRC Code of Practice for the use of hESC lines and WiCell's MTA. The cells were maintained in complete mTesR1 media on Corning Matrigel coated wells at 37°C, 5% CO_2. All procedures were performed in a tissue culture hood using sterile techniques. Cell lines were regularly screened for mycoplasma.

2.1.1 Mycoplasma testing

Mycoplasma testing was performed on all cell lines approximately every three months using the MycoAlert Mycoplasma Detection Kit as per the manufactures protocol (Lonza, #LT07). A 100 µL aliquot of culture media was incubated with an equal volume of MycoAlert reagent for 5 minutes before measuring the luminescence (reading A). Then 100 µL of the MycoAlert Substrate was incubated with the samples for 10 minutes before reading the luminescence again (reading B). The ratio of reading B / reading A determines whether cell culture is contaminated by mycoplasma. A reading $>$1.2 indicates the culture is positive, 0.9 - 1.2 is borderline, and $<$0.9 is negative. Any cultures producing borderline results were quarantined and retested after 24 hours. While no positive cultures were detected during these studies, plans to safely dispose of them were in place.

2.1.2 Matrigel coating of culture plates

Corning Matrigel hESC-qualified matrix (Corning, #354277) was aliquoted according to manufacturer's instructions and stored at -20°C until use. Aliquots were diluted to working concentrations with DMEM/F12 (Stem Cell Tech., #36254) and added to six-well plates for two hours at room temperature or overnight at 4°C. Excess matrigel solution was aspirated and plates were stored for up to two weeks at 4°C in DMEM/F12.

2.1.3 Passaging human embryonic stem cells

hESCs were passaged when 80% confluent approximately every 5 days. mTesR1 complete media (Stem Cell Tech., #85850) was formulated through the addition of 100 mL mTesR1 supplement to 400 mL of mTesR1 basal medium. Matrigel coated plates were removed from 4°C, washed once with Dulbecco's PBS (DPBS, Thermo Fisher, #14190094) and pre-warmed with complete mTesR1 at 37°C, 5% CO_2 for 20 minutes. Complete mTesR1 was aspirated from confluent hESCs and cells were washed with DPBS and 0.02% EDTA in DPBS (Lonza, #BE17-711E) before incubation with 0.02% EDTA/DPBS for 5 minutes at 37°C, 5% CO_2. hESC colonies were then washed and gently scraped from the tissue culture dish in complete mTesR1 with a cell scraper and the cell suspension was dissociated into clusters of 6-8 cells by gentle pipetting with a p1000 pipette onto the wall of the well. A small proportion of the cell suspension was transferred (typically 1:10 dilution) into the pre-warmed matrigel coated plates, evenly distributed and cultured at 37°C, 5% CO_2. Complete mTesR1 medium was replaced every 24 hours.

2.1.4 Thawing human embryonic stem cells

Matrigel coated 6-well plates were removed from 4°C, washed once with DPBS and pre-warmed for 20 minutes at 37°C with complete mTesR1. Vials containing $1x10^6$ hESCs were transferred from liquid nitrogen to the tissue culture facility on dry ice. Cells were rapidly thawed in a 37°C water bath until only a small amount of ice remains.

The cell suspension was quickly diluted in 9 mL of pre-warmed, complete mTesR1 media and centrifuged at 300 x g for 3 minutes. Supernatant was aspirated and the cell pellet resuspended in 1 mL of fresh complete mTesr1. The cell suspension was split between two, pre-warmed wells and incubated at 37°C, 5% CO_2. Cell media was completely replaced every 24 hours.

2.1.5 Freezing human embryonic stem cells

hESC lines were harvested for freezing when actively dividing (60-80% confluent). mTesR1 was aspirated from the well and cells were washed once with DPBS, once with 0.02% EDTA in DPBS before incubation with 0.02% EDTA in DPBS for 5 minutes at 37°C, 5% CO_2 to loosen cell contacts. The EDTA solution was carefully aspirated, cells were washed with DPBS before the addition of 1x knock-out serum replacement (KOSR). hESC colonies were gently scraped from the tissue culture dish with a cell scraper. Cell suspension was dissociated into clusters of 6-8 cells by gentle pipetting with a p1000 pipette onto the wall of the well. Cell suspension was added to pre-cooled cryovials before the addition of 20% dimethyl sulfoxide (DMSO) in 1x KOSR. Cryovials were placed into a Mr. Frostie container and cooled to -80°C overnight before transfer to liquid nitrogen for long term storage.

2.2 *In vitro* differentiation of human embryonic stem cells to intermediate mesoderm in defined medium

The directed differentiation of hESC was first established by adapting two published protocols using defined medium to generate kidney tissue [253, 254]. These protocols were followed for the first seven days of differentiation to generate intermediate mesoderm. Both protocols started by splitting non-confluent hESCs using EDTA onto matrigel coated plates and allowed to reach 40% confluence (typically 2 days). Basal media was switched from maintenance mTesR1 to differentiation APEL (Albumin Polyvinylalcohol Essential Lipids)(StemCell Tech., #05270) medium.

Takasato *et al.* add 8 μM CHIR99021 (R&D, #4423/10) for the first 2 days before switching to 200 ng/mL FGF9 (R&D, #273-F9-025) and 1 μg/ml heparin (Sigma-Aldrich, #H4784-250MG) for the remaining 5 days [253]. Alternatively, Xia *et al.* supplement with 50 ng/mL bFGF (R&D, #233-FB-025/CF) and 30 ng/mL human BMP4 (R&D, #314-BP-050) for 2 days before switching to 1 μM all-trans retinoid acid (R&D, #0695/50), 10 ng/mL human activin A (R&D, #338-AC-050/CF) and 100 ng/mL human BMP2 (R&D, #355-BM-050/CF) for 2 days [254].

2.3 *In vitro* differentiation of human embryonic stem cells to haemogenic endothelium in defined medium

The first stages of differentiation were adapted from Takasato *et al.*[253] with 2 days in 10 μM CHIR99021 in APEL media followed by 200 ng/mL FGF9 and 1 μg/ml heparin (IM-media) for 3 days. To expand angioblasts and direct differentiation to endothelium, IM-media was further supplemented with 50 ng/mL recombinant hVEGF and 10ng/mL bFGF for 2 days. To isolate endothelium from cultures, cells were harvested using TrypLE, washed and stained with CD144-PE antibody and sorted into APEL media as described below in section 2.6. 50,000 CD144-positive cells were re-plated in matrigel coated 12-well plates in APEL media supplemented with 50 ng/mL recombinant hVEGF and 10 ng/mL bFGF for 1-5 days. To induce a haemogenic programme 10 ng/mL BMP4 (R&D, #314-BP), 100 ng/mL SCF (R&D, #255-SC), 30 ng/mL IL-3 (R&D, #203-IL), 10 ng/mL IL-6 (R&D, #205-IL), 30 ng/mL TPO ((R&D, #288-TP), 2 IU EPO (Peprotech, #100-64), 10 ng/mL Flt-3L (R&D, #308-FK) was added to the endothelial media for up to 7 days.

2.4 Freezing, thawing and passaging of human umbilical vein endothelial cells (HUVECs)

Human umbilical vein endothelial cells (HUVECs) were purchased from Lonza (#C2519A) and grown in Endothelial Growth Medium 2 (EGM-2) (#CC-3162) as per

their instructions. Cells were rapidly thawed at 37°C and immediately diluted in 9 mL of EGM-2 to dilute DMSO. Cells were spun at 400 x g for 7 minutes, resuspended in 1 mL of media and distributed into a T75 flask. Cells were cultured at 37°C , 5% CO_2 overnight before addition of fresh medium. Thereafter, medium was replaced every 48 hours. Cells were only cultured for up to 5 passages before re-establishing cultures from frozen aliquots of early passages ($<$3).

HUVECs were passaged at 80% confluence to avoid contact inhibition. Cells were first washed with PBS and treated with trypsin/EDTA for 4 minutes at 37°C before neutralising with an equal volume of FBS. The resulting cell suspension was spun at 400 x g for 7 minutes before re suspending in EGM-2 media. Cells were manually counted with a haemocytometer and trypan blue exclusion. 2,500 viable cells/cm^2 were transferred into a new flask containing pre-warmed EGM-2 media.

HUVECs were regularly frozen to maintain stocks of early passages. Cells were harvested in the same manner as for passaging (above) and resuspended in ice cold cryopreservation media (80% EGM-2, 10% DMSO, 10% FBS) before transferring to labelled cryovials on ice. Cryovials were placed into a Mr. Frostie container and cooled to -80°C overnight before transfer to liquid nitrogen for long-term storage.

2.5 Reverse transcription quantitative polymerase chain reaction (RT-qPCR)

To quantify expression of a select panel of genes, RT-qPCR was applied. RNA was extracted from cells using Qiagen RNeasy plus kits (#74134) as per the manufacturers instructions. Cells were lysed in RLT buffer and either snap frozen on dry ice or immediately processed. Cell lysates were first transferred to gDNA eliminator columns, spun for 30 seconds at 8500 x g before adding 70% ethanol to the flow through and adding to a RNeasy column. Columns were again spun at 8500 x g but for 15 seconds. Column membranes were thoroughly washed with RW1 and RPE buffer before eluting

in RNase-free water. RNA concentration and purity was assessed on the NanoDrop One spectrophotometer (Thermo Scientific, #ND-ONE-W).

1 µg of RNA was used in reverse transcription reactions using Omniscript reverse transcriptase (Qiagen, #205111) with random primers, RNase inhibitor, dNTPs and RT buffer at 37°C for 1 hour. Reverse transcription was always run alongside a RT-negative control where the same input and reaction was used but with the absence of the reverse transcriptase.

TaqMan primers for genes of interest were ordered from Thermo Fisher with selected probes designed to span exons (shouldn't amplify gDNA) and were well cited where possible. When fewer than 12 genes were examined, 50 ng of cDNA was loaded into each reaction with 1x TaqMan fast master mix (Thermo Fisher Scientific, #A44360) and 1x TaqMan probes in a 20 µL reaction. Each gene was run against no template control (NTC), RT-negative control, and a positive control. These reactions were run on a Quantstudio 3 Real-Time PCR machine (Applied Biosystems, #A28567) for 40 cycles. When 12 genes or more were examined, qPCR reactions were run on the Fluidigm Biomark (#BMKHD). In these instances, samples were first pre-amplified with pooled TaqMan probes and Platinum Taq DNA polymerase (Thermo Fisher Scientific, #10966018) for 14 cycles. The thermal cycler was set to 96°C for 2 minutes then cycled through 95°C for 15 seconds and 60°C for 4 minutes before resting at 4°C . Pre-amplified cDNA was added onto a primed 48:48 integrated fluidic circuit (IFC) in triplicate, loaded using the MX controller before running 40 cycles on the Biomark.

Data analysis was performed in custom Excel spreadsheets and R scripts. Firstly, each gene was validated against its controls with expected amplification in the positive control, and no amplification in the no template and RT-negative controls. Any probe set not meeting these requirements was removed. Ct values were calculated with a threshold set individually for each gene in the geometric phase. Subsequently, any of the technical triplicates performed for each condition (gene x sample) that were significantly

different from the other two ($\geq$0.5 cycles) were excluded from the analysis. The mean of the remaining technical replicates was used to calculate ΔCt against the average of three housekeeping genes (*B2M, ACTB, HPRT1*). ΔCt were exponentiated to base 2 and plotted on bar graphs or z-score normalised and plotted on heatmaps.

2.6 Flow cytometry and fluorescent activated cell sorting (FACS)

Cells for flow cytometric analysis from *in vitro* differentiation cultures or mouse embryos were harvested (described in relevant sections), isolated by centrifugation at 400 x g for 7 minutes and washed in FACS buffer (PBS supplemented with 5% FBS). When available, fluorescent conjugated primary antibodies raised against proteins of interest were added to 100 µL of FACS buffer for 20 minutes at 4°C . When directly conjugated primary antibodies were unavailable, antibodies were first conjugated in house using the Zenon conjugation kit (Invitrogen, #Z25060) and following manufacturers instructions. Controls relevant for each experiment and antibody panel (isotype controls, fluorescent minus one (FMOs), no stain controls) were set up in parallel. Cells were thoroughly washed in 1 mL of FACS buffer before centrifugation. Finally, cells were resuspended in FACS buffer at a concentration of 25 million cells per mL and filtered through a 40µm filter.

For flow cytometric acquisition (no sorting), cells were run at approximately 10,000 events per second on a BD LSRFortessa X-20 Cell Analyser (X20) equipped with 405 nm, 488 nm, 561 nm, and 647 nm lasers. When populations of interest required isolation by FACS, a BD FACSAria Fusion (Fusion) equipped with 405 nm, 488 nm, 561 nm, and 647 nm lasers was used at a flow rate of approximately 5,000 events. Population gates were typically set using internal negative controls where possible or FMOs when not available. Sort gates were set similarly but with tighter borders to limit contamination from neighbouring populations. Both machines were only used following running the BD Cytometer Setup and Tracking (CS&T) application to ensure optimal fluorescent

measurement. The Fusion additionally underwent BD Accudrop optimisation to allow accurate sorting of up to four populations. Purity sort mode was used where populations of interest were in abundance but changed to yield to collect rare cell populations. After sorting, purity was checked by running an aliquot of sorted populations back through the FACS machine. Data were analysed in FlowJo or BD FACSDiva.

2.7 Tubule formation assay

To validate the production of functional endothelial cells from hESC differentiation cultures tubule formation assays were performed. 12-well plates (Corning, #10253041) were coated with 200 µL of growth factor reduced (GFR) matrigel at 37°C for 30 minutes before adding 40,000 cells to each well. Each condition was tested in triplicate with HUVECs used as positive controls. Cells typically formed tubules after 6-12 hours with representative images being taken on a EVOS M5000 microscope <15 hours after seeding.

2.8 Immunofluorescence of hESC differentiation cultures

To visualise the organisation of hESC differentiation cultures, cells were stained for marker genes using antibodies. Before starting differentiation as described in section 2.3, ethanol-cleaned, glass coverslips were placed in the bottom of 6-well plates and Matrigel coated as described in section 2.1.2. At the time of interest, differentiation media was removed, cells were thoroughly washed with DPBS and fixed with 3% PFA for 15 minutes at room temperature. After DPBS washing, auto-florescence was quenched with 0.3 M glycine for 15 minutes at room temperature. Surface markers were first stained, cells were permeabilised, and intracellular markers stained.

Cells were blocked with DPBS plus 10% serum (same species as secondary antibody source) for 1 hour at room temperature before diluting primary antibody in blocking solution and incubating in a humid, sealed container overnight at 4°C . Excess primary was removed with thorough DPBS washes and fluorescent conjugated secondary antibodies

were incubated for 45 minutes at room temperature. After DPBS washes, cells were permeabilised with 0.3% triton for 15 minutes before an intracellular block for using 1 hour at room temperature in DPBS plus 10% serum (same species as secondary antibody source). Primary and secondary antibody staining were performed in the same fashion as surface markers. Finally, cells were briefly stained with 1 µg/mL of DAPI for 30 minutes at room temperature. Coverslips were briefly dipped in water before mounting with VectaShield and sealing with clear nail polish.

Stained hESC colonies were imaged on a DeltaVision Elite widefield microscope using a x20/0.75NA objective using the DAPI, FITC, and TRITC imaging channels. Tiling and z-stacks were perfomed using the DeltaVision built-in software.

2.9 Mouse colony maintenance and genotyping

Mouse colony maintenance and all regulated procedures were in accordance with the Operation of the Animals (Scientific Procedures) Act 1986, approved by the relevant authorities (UK Home Office and University of Oxford) and performed under project licence number PA7C92A40. All mice were maintained in a 12 hour light-dark cycle. Matings were set up at 8-12 weeks of age and kept until a maximum of 30 weeks old. Pups were weaned at 21 days old and ear-clipping was performed between 10-14 days old for identification and genotyping. Genomic DNA (gDNA) was extracted from ear clips using the Qiagen DNeasy Blood & Tissue Kit (Qiagen, #69504) following the manufacturers instructions. gDNA concentration and purity was assessed on the NanoDrop One spectrophotometer (Thermo Scientific, #ND-ONE-W). 90 ng of gDNA is loaded into genotyping PCR reactions using the Phire green hot start II DNA polymerase (Thermo Scientific, #F124L) and primers designed to amplify 100-250 base pair regions specific to individual alleles. PCR cycling was performed on a Proflex thermal cycler (Applied Biosystems, #4484073) at 98°C for 30 seconds followed by 32 cycles of 98°C for 5 seconds, 64°C for 5 seconds, 72°C for 5 seconds before a final extension of 72°C for 60 seconds and holding at 12°C . PCR products were run on a 1.5% agarose gel

supplemented with 0.5% ethidium bromide at 120 volts for 20 minutes alongside a 1 kb ladder and visualised using a Gel Doc XR+ (Bio-Rad, #1708195). Each mouse's genotype was compared to positive and negative controls and recorded in the University of Oxford mouse colony management system.

2.10 Timed matings and embryo collection

Timed matings between C57BL/6 mice were performed to collect embryos of the correct stage. Females identified as in oestrus by vaginal inspection were paired with stud males overnight or for 2 hours between 7am - 9am. Noon of the day when a vaginal plug was found was designated E0.5. At the time of interest, dams were culled and uteri were dissected into ice cold PBS + 10% FBS. Litters were staged according to the dissection of 2-3 embryos out of their decidua and comparison of anatomical features with published staging materials (EMAP) [255, 256] when retaining conceptus in decidua and/or by counting somite pairs when embryos dissected out of decidua.

2.11 Embryo preparation for cryosectioning

Embryos were fixed and sectioned in their decidua to limit physical damage and maintain structural integrity during preparation and sectioning. The myometrium were removed and whole decidua were fixed overnight in 4% para-formaldehyde at 4°C . All subsequent steps were performed on ice. Embryos were thoroughly washed with 1× PBS/0.2% Triton-X100 (PBST) followed by gradual dehydration into 100% methanol (0%, 25%, 50%, 75%, 100% methanol-PBT steps) and stored at -80°C for a minimum of 48 hours. To prepare for embedding, embryos were first gradually re-hydrated (100%, 75%, 50%, 25%, 0% methanol-PBST mix) and then slowly equilibrated into 20% sucrose in PBST for 4 hours. Single embryos were transferred into OCT-filled cryomolds before being snap-frozen in liquid nitrogen and stored at -80°C .

Blocks were allowed to equilibrate at -20°C before mounting onto a Leica CM3050S

cryostat sample holder. 10, 15, or 30 µm sections were collected onto (3-aminopropyl) triethoxysilane (AS) coated coverslips (thickness #1.5, 24 mm x 60 mm) at -20°C . Sections were allowed to air dry for 30 minutes before immediately proceeding to HCR staining.

2.12 Hybridisation chain reaction version 3.0 (HCR v3.0)

2.12.1 Split-initiator probes, hairpin amplifiers and buffers

Hybridisation chain reaction [239] probes, amplifiers and buffers were all purchased from Molecular Instruments. A summary of probes and amplifiers used can be found in Table 6. Hybridisation probe libraries were designed by Molecular Instruments against NCBI sequences they were provided. When gene length permitted, 40 split-probe-pairs per gene were designed. In all other cases, as many probe-pairs were designed as possible. Probe sets (one set per gene) were evenly spread across three initiator channels (B1, B4 and B5) with the exception of *Lhx1* which was designed to initiate B3 in order to test four-channel imaging. Amplifier hairpins were purchases in two colours each to permit the staining of each probe set in two channels.

2.12.2 Digital HCR (dHCR)

Digital HCR (for the imaging of diffraction-limited spots) [239], was always performed on freshly prepared cryosections. All reagents and buffers were of high molecular grade and nuclease free. Work surfaces, pipettes and gloves were treated with RNAse-zap to reduce RNase contamination. Sections were protected from drying out at all times through addition of excess buffer, storage in humid staining chambers and covering with parafilm when necessary.

Each section was enclosed in a 200 µL silicon SecureSeal chamber (#621505) to allow easy exchange of buffers and maintenance of small working volumes. Following a 30 minute air drying, OCT was washed off with PBS and sections were gently fixed for

15 minutes in 1% para-formaldehyde (PFA). Subsequently, sections were permeabilised with 0.5% sodium dodecyl sulfate (SDS) in PBS for five minutes followed by a brief 1% PFA fixation. To limit the auto-fluorescence caused by PFA fixation, sections were treated with 1% sodium borohydride (NaBH4) for 5 minutes before washing in 2x saline-sodium citrate (SSC) buffer. Blocking was then performed using 1 µM of random, 40 nucleotide-long oligonucleotide in probe hybridisation buffer for 30 minutes at 37°C in a hybridisation oven. To hybridise split-initiator probe pairs, 16 nM of up to four probe sets were added to sections in probe hybridisation buffer overnight (12-16 hours) at 37°C . Sections were subsequently washed with probe wash buffer with increasingly high 5x SSC + 0.1% Tween-20 (SSCT) buffer (0%, 25%, 50%, 75%, 100% SSCT) at 37°C for 15 minutes each.

Hairpin amplifiers corresponding to the cognate initiator sequence present on hybridisation probes were snap-heated at 95°C for 90 seconds before cooling for 30 minutes at room temperature. Hairpins conjugated to Alexa Fluor (AF)488, AF547, AF596, and AF647 were routinely used. 60 nM of each hairpin were combined in amplification buffer and added to sections for 45 minutes. Excess hairpins were removed with three, 5 minute washes in 5x SSCT. 1 µg/mL of 4,6-diamidino-2-phenylindole (DAPI) added during the second wash to counter-stain for nuclei. Finally, 200 µL of prolong gold anti-fade medium (#P36930) was added to the sections for imaging.

For single molecule quantification, sections were imaged on a Zeiss 780i inverted laser-scanning confocal with a x63/1.4 NA oil immersion lens and a pixel size of 0.22 nm x 0.22 nm. To enable higher-throughput imaging of stained sections, regions outside those being quantified were often imaged on a Zeiss spinning disc confocal in parallel using a x40/1.3 NA oil immersion lens with a pixel size of 0.43 nm x 0.43 nm. On both systems, z-stacks were acquired with a 0.5 µm interval and tiled regions were set with 10% overlap.

2.12.3 Quantitative HCR (qHCR)

qHCR was performed for relative gene expression quantitation was not required. In contrast to dHCR, samples were amplified overnight in order to increase their brightness for identification in whole mount imaging of embryos. Staining and imaging was performed as published [257]. Embryos were fixed in 4% PFA for 20 minutes on ice before washing with PBST and gradually dehydrating in a series of increasing methanol concentration (0%, 25%, 50%, 75%, 100% methanol-PBST mix). Embryos were kept at -20°C in 100% methanol for at least 2 days and up to six months. When ready to stain, embryos were re-hydrated in decreasing methanol concentration (100%, 75%, 50%, 25%, 0% methanol-PBST mix) and washed in PBST. Embryos were then treated with 10 µg/mL proteinase K solution for 5 minutes at room temperature before washing with PBST and post-fixation with 4% PFA for 20 minutes at room temperature.

After further washing in PBST, embryos were pre-hybridised in 50% probe hybridisation buffer/ 50% PBST for 2 hours at room temperature before incubation at 37°C for 30 minutes in probe hybridisation buffer. Probes were hybridised by incubating embryos in probe hybridisation buffer containing 2 pmol of each probe library overnight (12-16 hours) at 37°C. Embryos were thoroughly washed with pre-warmed probe wash buffer at 37°C followed by 5x SSCT buffer to remove excess probes before equilibrating in amplification buffer for 30 minutes at room temperature. Hairpin amplifiers (30 pmol) were prepared by snap cooling (95°C for 90 seconds, room temperature for 30 minutes), diluted in amplification buffer and added to embryos overnight at 37°C. Excess hairpins were removed with thorough 5x SSCT washes during which time embryos were stained with DAPI.

Finally, embryos were mounted in SlowFade Gold anti-fade (Thermo Fisher, #S36936) onto coverslips (thickness #1.5, 24 mm x 60 mm). Electrical tape with a small hole cut in it was added to coverslips to create a space as not to squish embryos when adding a second coverslip on top and sealing with clear nail varnish. Embryos were images using a Zeiss z900 laser scanning confocal microscope equipped with a x25/0.8NA water

immersion objective.

2.13 Probe and signal stripping

To allow multiplexing of HCR beyond four orthogonal channels, hybridisation probes, amplification hairpins, and fluorochromes were removed by DNaseI treatment before re-staining with different probe sets. Following imaging, sections were washed with 2x SSC before being treated with DNaseI (500 units/mL) for 1 hour. To remove excess enzyme and enable re-hybridisation, sections were thoroughly washed with 30% formamide/2x SSC and 2x SSC. Sections could then be re-stained by dHCR (section 2.12.2) or immunostained for cell membrane visualisation.

2.14 Cell membrane staining

Following the final round of HCR, probes and signal were stripped as above to allow immunostaining of membrane bound proteins to identify cell borders. Samples are first washed with 30% formamide/2x SSC and 2x SSC to remove excess enzyme before switching to 1x PBS. Sections were initially blocked with 5% bovine serum albumin (BSA, Sigma-Aldrich #A7906-50G) and 10% donkey serum (Sigma-Aldrich #G9023-10ML) in PBS + 0.2% Triton-X100 before staining with a cocktail of antibodies against E-cadherin (BD-Biosciences #610181), β -catenin (Abcam #ab6301), and N-cadherin (ThermoFisher #333900) overnight in blocking solution. Excess primary antibody was removed with thorough PBS washes before incubation with secondary antibodies conjugated to AF633 to allow single channel imaging of all markers simultaneously (goat anti-mouse IgG$_2a$-AF633 and goat anti-mouse IgG$_1\kappa$-AF633). Imaging was performed in an identical manner as for dHCR (section 2.12.2).

2.15 Image alignment

Image alignment is required to accurately overlap rounds of HCR imaging and membrane staining for accurate quantification and was performed in a semi-automated manner. When imaging each round, the field of view was set to be as close to the first round as possible. Images to be registered were checked to be the same dimensions and adjusted if different using a custom ImageJ macro. Alignment is composed of two factors - rotation and shift. To calculate rotation, the angle between two reference sites was measured in each round (recognisable, auto-florescent red blood cells for example). The difference in angle between each round and the reference round (typically membrane stain round) was used to rotate the respective round. The three-dimensional shift (in pixels) was calculated using the auto-fluorescence acquired in the 405 nm channel every round as a reference. Each round was measured relative to the membrane stained round. To measure the shift, images were loaded into MatLab (MathWorks, vR2019b), convoluted (rolling ball) and rough shift was calculated using a custom MatLab function (xcorr3fft) [258]. This function employs phase correlation to calculate the offset between two 3-D images. After applying this calculated shift, x-y and y-z maximum intensity projections were displayed as two-colour composite to assess overlap. If unsatisfactory, shift values were manually adjusted until overlap was acceptable.

2.16 3-D cell segmentation

3-D cell segmentation aims to define the volume contained within a cell's plasma membrane to enable counting of single mRNA molecules as a measure of gene expression. Immunostaining of membrane-bound proteins (see 2.14 for details) were used as input. To limit the effect of non-continuous staining, images were convoluted (rolling ball) and input into the machine learning software Ilastik for pixel classification. Two classifications were trained by the user (cell membrane and cell interior) by manually painting onto a small subset of data. The resulting model was tested by displaying the prediction

of each pixel classifier for three independent sections. Once the model could predict cell membranes and interiors to the same degree as the user, the classification model was applied to the convolved data to generate a prediction matrix for both membrane and interior classifiers. These matrices were threshold-ed to remove low confidence predictions (typically a threshold of 0.8 - 0.95) as these reflect sources of signal other than cell membrane staining (e.g. auto-fluorescent red blood cells). In this way, the data input into segmentation algorithms were of higher contrast and specificity.

Many segmentation algorithms across analysis environments (Python, MatLab, Imaris) were tested. The most accurate was the Cells function in Imaris. Threshold-ed Ilastik predictions were loaded into Imaris and converted from .tif to .ims in Imaris 9.5.1 before opening in the Surpass 3-D viewer. A new Cells object was created and run with Detect Cell Only option selected, Cell Smallest Size set to 7.50 µm, Cell Membrane Detail set to 0.20 µm, and using a Local Contrast filter. These were then filtered with a cell membrane intensity threshold manually set at 34.864 and a quality threshold manually set at 0.502. To remove segmented regions that were not cells, another filter excluding cells below 70 μm^3 and above 3324 μm^3. In instances where regions outside the section were segmented, the user manually removed them. Cells were subsequently exported to a Surfaces Object to enable channel masking. Masked channels were saved as .tif in preparation for use during dot counting.

2.17 dHCR spot counting

In order to quantify the number of transcripts per cell as a measure of gene expression, single molecules were counted from multiplexed dHCR imaging data. This step requires image alignment (section 2.15) and cell segmentation (section 2.16) to have already been performed. Firstly, dHCR data was input into the machine learning software Ilastik to generate a matrix containing a prediction value (between 0 and 1) for each pixel to be a part of a mRNA spot. Firstly, a pixel classification model for each fluorescent channel (AF488, AF546, AF594, and AF647) was manually trained by the user by

painting pixels that form mRNA spots in one classification and pixels that are not spots in another classification on a small subset of dHCR data. The inclusion of a second 'non-spot' classification allowed the user to circle individual spots and create high levels of uncertainty between adjacent spots, making segmentation of individual spots simpler. The models were tested by displaying the prediction of each pixel classifier on three independent dHCR rounds. Training was stopped once the model could predict spot pixels as accurately as the user. dHCR images were split into individual fluorescent channels using a custom ImageJ macro before being input into their respective Ilastik model. Prediction matrices were exported as .h5 files.

Spot counting was performed using a custom MatLab pipeline heavily adapted from Lignell *et al* [258]. Cell segmentation masks, dHCR data, and spot prediction matrices were loaded into the MatLab environment. Firstly, rotation and shift were applied as calculated in 2.15 to align all dHCR rounds (and their respective probability matrices) to the cell segmentation mask. Next, spot prediction matrices were threshold-ed (set at 0.70 - 0.99) to remove low probability pixels. To identify single spots, touching pixels in the binary prediction map were grouped and given unique values. To create spot statistics (volume, centroid coordinate, mean intensity), these unique values were overlapped with the dHCR imaging data. These statistics were used to filter unique values so they only contain spots. Thresholds were manually calculated by plotting histograms of all spots and identifying outliers outside the gaussian distribution. The threshold was validated by plotting included spots in red and excluded spots in blue overlayed onto a maximum intensity projection of dHCR data. Threshold was adjusted so that all dHCR spots were overlayed with red circles but with no red circles present elsewhere. Subsequently, the number of unique values (spots) within a single cell mask was used to count the number of molecules. After repeating this process for every gene imaged by dHCR, a gene count matrix was generated with single cells as rows and genes as columns.

2.18 Hierarchical clustering and spatial remapping of clusters

To allow biological interpretation from the gene count matrix calculated by dHCR spot counting (Section 2.17), single cells were clustered using unbiased hierarchical clustering. Gene count matrices were loaded into MatLab and clusters were calculated with the *clustergram* function using euclidean distances. The resultant dendrograms were displayed on z-score normalised expression heatmaps. Clusters of biological significance were identified by the user through careful analysis of the expression of multiple marker genes. To show the location of these biological clusters, their constituent cells were coloured the same as the cluster and displayed in the section.

2.19 Laser capture micro-dissection

2.19.1 Sectioning of fresh frozen mouse embryos

Multiple studies have demonstrated the detrimental effect of fixation (particularly PFA) on RNA recovery and downstream sequencing. Therefore to recover material from cryosections of mouse embryos for sequencing, embryos were collected differently than for HCR. Timed matings were set in the same way as described in section 2.10). At the time of collection, pregnant dams were culled and uterine tissue was dissected into 10% FCS + PBS. Individual decidua were separated and cleaned of any fatty tissue that remained attached but, unlike for HCR, the myometrium was not removed. Keeping the myometrium significantly improved section integrity. Individual decidua were immediately embedded in pre-cooled OCT before snap-freezing in liquid nitrogen and storage at -80°C until sectioning.

OCT blocks were equilibrated to -20°C before mounting onto a cryostat sample holder. The cryostat chamber temperature was kept at -20°C but the object temperature was raised from -20°C to -18°C as this was shown to improve section integrity. For LCM,

12µm sections were collected onto pre-cooled, UV treated (20 minutes, 254 nm), PET-membrane coated slides (Carl Zeiss, #415190-9051-000). Sections were immediately frozen and stored at -20°C until processed for LCM.

2.19.2 Section preparation and rapid staining for LCM

To limit RNA degradation, sections were prepared and stained as quickly as possible immediately before cutting. All solutions were RNase-free and equipment was thoroughly cleaned with ethanol and RNaseZap before use. Sections were thawed at room temperature in a closed container (to limit condensation formed on slides) and air-dried for 10 minutes. OCT was washed off with 1x PBS for 30 seconds before rapidly fixing with 70% ethanol. To allow clear identification of angioblasts sections were stained with a polyclonal goat anti-mouse SCL antibody (Santa Cruz, #sc-12984) in PBS for 4 minutes. Sections were quickly rinsed in 1x PBS before staining with DAPI, 1 mg/mL wheat germ agglutinin (WGA) conjugated to AF555 (Thermo Fisher, #W32464), and AF488 conjugated secondary donkey anti-goat IgG antibody (Thermo Fisher, #A-21447) to discern the location of neighbouring cells, visualise cellular borders, and SCL staining. Sections were again quickly washed in PBS before serial dehydration in ethanol (0%, 25%, 50%, 75%, 100% ethanol-PBS). After air drying for 5 minutes, sections were immediately loaded on to the LCM microscope for cutting.

2.19.3 Laser capture micro-dissection and catapulting

Firstly, the LCM microscope (Zeiss PALM MicroBeam with Robomover) was optimised for clean cutting and catapulting with minimum laser power. The Calibrate Laser Cutting function was used in a region of empty membrane to finely adjust the power and focus of the laser for cutting. With these initial values in place, a region of tissue not relevant to the experiment (typically decidual tissue) was used to finely tune laser power and focus to also cut tissue and select suitable catapult delta values. To selectively dissect cells of interest, stained sections were loaded on the LCM microscope in preparation

for cutting. Sections were scanned in the DAPI channel at x20 to identify embryonic tissue within the decidua before switching to x40 objective for cutting. Subsequently, sections were scanned in the green channel to identify SCL-positive cells. Once a region of interest was identified, the red channel (WGA staining) was used to distinguish the borders of neighbouring cells. The RoboLPC function was used to manually draw round single cells of interest, cut them and defocus the laser to catapult them into an adhesive PCR-tube cap (Carl Zeiss, #415190-9161-000). The first and last cap of a session was visually checked for presence of catapulted cell(s)under the microscope with the Check Cap function. Once all cells of interest had been cut, caps were removed from the tube holder and 3 µL of Smart-seq3 lysis buffer (containing 40% PEG 800 solution (Sigma Aldrich, #89510-250G-F), 10% Triton X-100, RNAse Inhibitor (40U/µL) (Takarabio, #2313A), OligodT30VN, dNTPs (Thermo Fisher, #R0182)) [259]. Lysate was gently agitated before adding capless PCR tubes, spinning down and snap-freezing on dry ice. Lysates were stored at -80°C for up to 2 weeks.

2.19.4 Reverse transcription and pre-amplification of single cell lysates

All stages of cDNA library preparation were performed in a clean, PCR product-free environment regularly UV treated and cleaned with RNAseZap. To reverse transcribe RNA, lysates were thawed and then incubated at 72°Cfor ten minutes to release fully lyse the cells before adding the reverse transcription mix containing 25 mM Tris-HCl pH 8.3, 30mM NaCl, 2.5 mM $MgCl_2$, 1 mM GTP, 8 mM DTT, 0.5 µ/µL RNase inhibitor, 2 uM TSO, 2 u/uL maxima H-minus RT enzyme. Reverse transcription was performed in a thermal cycler at 43°C for 90 minutes, 10 cycles of 50°C for 2 minutes followed by 42°C for 2 minutes and finally 85°C for 5 minutes. cDNA from single cells was then pre-amplified by PCR for 8-15 cycles using 1x Kapa HiFi HotStart buffer, 0.3 mM each dNTPs, 0.5 mM $MgCl_2$, 0.5uM forward and reverse primers, 0.02 U/µL KAPA DNA polymerase using the following PCR cycling conditions: initial denaturation at 98°C for 3 minutes, cycling 20 seconds at 98°C 30 seconds at 65°C 4 minutes at 72°C before a final elongation at 72°C for 5 minutes. Typically, amplified cDNA was subsequently

purified using 22% PEG clean-up beads before being diluted 1:10 and run on the Agilent Tapestation High Sensitivity D5000 ScreenTape to check library quality, quantity and complexity.

2.20 10x Chromium scRNA-seq

2.20.1 Mouse embryo dissection and dissociation for 10x Chromium

Timed matings between C57BL/6 mice was performed as described above (section 2.10. Pregnant dams were sacrificed and embryos collected at E8.0 in PBS with 10% FBS. Individual embryos were dissected from their decidua and staged according to the number of somite pairs. Four somite-stage (ss) embryos were further dissected by cutting between the second and third somite, removal of the allantois and adjacent embryonic tissue to isolate the caudal half of the embryo. A total of seven 4ss embryos from two litters were dissected and pooled for further processing. Tissue was dissociated in 0.25% collagenase IV at 37°Cfor ten minutes. Cell clumps were gently dissociated by manual pipetting and gentle shaking. Dissociation was halted through the addition of 10% FBS. The resulting cell suspension was washed and filtered through 100 µm tip filters and counted using a haemocytometer. A minimum of one hundred live cells were counted with dead cells excluded by trypan blue staining. Dead cells constituted $<5\%$ of the cell suspension. Cells were diluted to 1000 cells per µL.

2.20.2 10x Chromium library preparation

To enable single cell barcoding, the 10x chromium system was utilised using the Chromium Next GEM single cell 3' reagent kit v3.1 (10x Genomics, #PN-1000127 and #PN-1000269). The manufacturer's instructions were followed throughout with the commonly performed adaptation of intentionally overloading with higher cell numbers.

Firstly, 20,000 cells were loaded into 10x chromium controller and individually partitioned into Gelbeads-in-Emulsions (GEMs), lysed and reversed transcribed such that each

polyadenylated mRNA molecule creates a cDNA molecule with a unique molecular identifier (UMI), 10x cell barcode, and Illumina TruSeq read 1 sequencing primer. After reverse transcription (RT), GEMs are broken, cell libraries are recovered and pooled. SPRIselect para-magnetic beads (Beckman Coulter, #B23318) were used to clean full length cDNA of excess RT and GEM reagents before running a small aliquot on the Agilent Tapestation High Sensitivity D5000 ScreenTape (Agilent, #5067- 5592 and #5067- 5593) to check the quality and quantity of the library. To maintain library complexity, 25% of the library was used for amplification by PCR (11 cycles) before libraries were enzymatically fragmented and size selected with SPRIselect beads to carefully select amplicon size. Fragments were end-repaired, A-tailed and ligated to add the TruSeq Read 2 before performing a sample index PCR to add the i5 and i7 sample indices (GTGGATCAAA and GCCAACCCTG respectively) from the Dual Index Plate TT Set A (10x genomics, #PN-3000431). A final post-PCR SPRIselect double side size selection was performed before running on the Agilent Tapestation High Sensitivity D5000 ScreenTape to ensure the correct library size was obtained and impurities were removed. Libraries were stored at -20°Cuntil they were sent for sequencing.

2.20.3 Illumina next generation sequencing

Next generation sequencing of the prepared 10x library was performed at Novogene using an Illumina Novaseq S4 flow cell with 150 base pair, paired end sequencing targeting 360 gigabytes (1.2 billion reads) for approximately 100,000 reads per cell (predicted 12,000 cells). Libraries were pooled and sequenced with other users and demultiplexed by Novogene. Demultiplexed fastq files were shared over the cloud with data integrity monitored by MD5 hash generator.

2.20.4 Sequencing data pre-processing and quality control

All analysis of single cell data was performed on the MRC WIMM Centre for Computational Biology computer clusters. Gene expression data were generated by running fastq

files through 10x CellRanger v6.0.1 [260] 'count' function using the genome reference consortium mouse build 38 (mm10). Default CellRanger filtering was applied to generate barcodes.tsv, genes.tsv, and matrix.mtx files containing counts of transcripts for each gene.

All subsequent analysis was performed in rstudio v1.4.1106 using R v4.0.1. To process, filter, and cluster cells, Seurat v3.2.1 was used [261]. Gene count matrix output from CellRanger was used to create a Seurat object. To remove empty droplets and low-quality cells, stringent filters were applied, only progressing with cells with $<$5% mitochondrial reads, $>$1500 detected genes, and $>$1250 unique reads. The cutoff for mitochondrial reads was set based upon widely accepted thresholds, while genes and unique reads were set following visualisation on violin plots.

2.20.5 Normalisation and clustering 10x scRNA-seq

Seurat objects containing filtered cells described above were initially subjected to a standard Seurat workflow. Firstly, gene counts are normalised by 'NormalizeData' function using the LogNormalize method whereby feature counts for each cell are divided by the total counts for that cell, multiplied by a scale factor and then natural-log transformed. The top 2000 most variable genes were identified by 'FindVariableFeatures' and used to run a principle component analysis (PCA). Visualisation of the variation capture by each principle component in a scree plot allowed identification of the number of principle components to take forward by selecting a value shortly after "the elbow". These principle components were used to generate a k-nearest neighbours graph for the subsequent 'FindClusters' function. Clusters could then be visualised on a low dimensional Uniform Manifold Approximation and Projection (UMAP). Clusters were manually annotated by assessing the top thirty enriched genes for each cluster ('VariableFeatures') and visualising expression of well-characterised marker genes.

2.20.6 Removal of doublets from scRNA-seq data

During encapsulation of cells within GEMs, there is a predictable rate at which two cells will enter the same GEM. Doublet identification aims to remove these from the analysis. This is a difficult task and so two independent analysis programmes (doubletFinder and scDblFinder) were run and the intersection was removed. Firstly, Seurat clusters 10 and 11 were removed from analysis as they had been identified as low-quality. In both instances, the estimated doublet rate was set to 9.6% to set the probability threshold. Cells who were identified as doublets by both programmes were removed from downstream analysis.

2.21 CellChat cellular communication prediction

CellChat uses gene expression matrices from scRNA-seq data to predict potential cellular signalling between clusters of cells [262]. The package was run on E8.0 data from the publicly available MGA and 10x chromium data from E8.0 posterior generated as described in section 2.20. Only cells types identified as constituting the angioblast niche (paraxial mesoderm, intermediate mesoderm, lateral plate mesoderm, definitive endoderm, and surface ectoderm) were used in CellChat analysis. After creating the CellChat object, 'over expressed' ligands and receptors for each cluster were calculated and the mouse ligand-receptor database was used to calculate the predicted interactions and their probabilities. These interactions were summarised in circle plots using the aggregateNet function. The predicted signalling pathways were each evaluated in hierarchy plots with the netVisualaggregate function to visualise the predicted sender and receiver with the contributing ligands and receptors displayed using netAnalysiscontribution function. Cellular signalling could be grouped into similar outgoing and incoming patterns using the identifyCommunicationPatterns with the number of patterns selected using the Cophenetic and Silhouette metrics generated by the selectK function.

Chapter 3

Directed differentiation of hESCs to intra-embryonic haemato-endothelium

Contents

3.1 Introduction 71
3.2 Results 73
3.2.1 Selection of intermediate mesoderm differentiation protocols for testing 73
3.2.2 Establishing intermediate mesoderm differentiation protocols 74
3.2.3 KDR^+ cells generated in intermediate mesoderm differentiation conditions are capable of endothelial differentiation 79
3.2.4 Differentiation cultures exhibit the correct HOX expression pattern 83
3.2.5 Angioblasts differentiated in high levels of VEGF give rise to arterial endothelium 85
3.2.6 Inducing a haematopoietic programme in arterial endothelium cultures is highly variable 86
3.3 Discussion 88

3.1 Introduction

HSCs are unique in their capacity to provide long-term blood production of all haematopoietic lineages. Their study has profound impact on human health and disease [201]. However, there are currently no *in vitro* models of human HSC development that do not require reprogramming [212]. Therefore, understanding the cellular signalling, gene regulatory networks, and sources of heterogeneity in their development is severely impeded. While many *in vitro* differentiation protocols exist to generate blood progenitors not differentiated from HSCs but directly from HE (HSC-independent progenitors), none have been capable of producing HSCs *de novo*. Three possible reasons for this are: 1) The HE produced in these cultures does not reflect Wave 3 HE and/or does not

have HSC potential, 2) the order of arterialisation is reversed as compared to *in vivo* development of HE and/or 3) the HE culture conditions do not contain the necessary cytokines and physical properties to enable HSC-EHT, maturation, maintenance and/or expansion. Here, I aim to address the first and second possibility as I hypothesise that current differentiation conditions only generate Wave 2 HE.

Wave 2 HE represents a distinct lineage [130] that has undergone considerably different signalling than Wave 3 HE. Wave 2 HE originates from extra-embryonic mesoderm (ExM) while Wave 3 is of lateral plate mesoderm (LPM) origin. Furthermore, HSC competent progenitors are only known to be specified in a well defined anterior-posterior (A-P) region and so *in vitro* cultures must recapitulate the signalling ongoing in that region (discussed in section 1.4.4). The mesonephric duct (wolffian duct) is derived from anterior (a)IM [70] and is patterned in the same restricted A-P region as HSC progenitor cells (Fig.1.6) [129]. In this chapter, I aim to use culture conditions shown to drive nephric duct formation and pattern mesoderm in this limited A-P region.

While these cultures are optimised to differentiate IM, they are only up to 80% efficient [254, 263]. We hypothesise the heterogeneity in culture will contain LPM derived haemato-endothelial progenitors. Secondly, I hypothesise the mesonephric duct could represent a signalling centre to Wave 3 HEPs. After establishing haematopoietic differentiation I then aim to describe any structure (e.g. cell types organised beyond random distribution) forming between the two cell types as an indication of signalling interaction. Finally, there are indications the mesonephric duct and Wave 3 HEPs share a close lineage history [130, 264, 265]. Lineage labelling the C3 blastomere in xenopus labels both the DA and the mesonephric duct [130]. Furthermore, computational reconstruction of lineage progression from transcriptomic data across stages of development predicts definitive haematopoietic progenitors and aIM diverge later than other mesodermal lineages in zebrafish [265] and xenopus [264] embryos. Therefore, using culture conditions demonstrated to robustly produce aIM and the mesonephric duct, I will recapitulate the mesoderm patterning signals critical for regionalisation

experienced by Wave 3 HEPs and aim to create a cellular niche mimicking the embryo. This approach requires the identification of an intermediate mesoderm differentiation protocol and screening for haemato-endothelial progenitors within these cultures.

3.2 Results

3.2.1 Selection of intermediate mesoderm differentiation protocols for testing

At the time of selection, multiple protocols to direct differentiation of hPSCs towards intermediate mesoderm and it's derivatives were available [70, 72, 77, 254, 266–271]. We screened each of these reports on a number of criteria for their suitability for adaptation to haematopoietic differentiation. Typically, the primary aim of these reports is to identify culture conditions to derive kidney structures resembling the metanephric kidney [272]. Identification that this requires induction of both anterior and posterior IM to derive the metanepheric duct (forming the uteric bud) and nephron progenitors respectively motivated research into how to control A-P patterning. Protocols from Mae 2013 [269], Narayanan 2013 [271], Kang 2014 [270], Taguchi 2014 [70], Takasato 2014 [77], Freedman 2015 [267] and colleagues were all excluded as they did not make the distinction between anterior versus posterior IM. As Wave 2 HEPs are derived from ExM and induced by posterior signalling, we wanted to avoid these conditions as far as possible. Xia, *et. al.*, 2013 reports specific induction of uteric bud progenitors (derived from the nephric duct) through timely addition of BMP4 and FGF2 for two days to induce primitive streak and then retinoic acid (RA), Activin A and BMP2 for 2 days (hereafter called Xia protocol). Takasato, *et. al.*, 2015 report hPSC derived IM can be regionalised along the A-P axis through changing the duration of WNT signalling (through CHIR99021 mediated GSK3 inhibition) before the addition of FGF9 (refered to as Takasato protocol). The Takasato protocol defines aIM is specified by FGF9 after just two days of CHIR exposure. This supports the Xia protocol's short exposure of BMP4 signalling to induce the same population. With contrasting use of cytokines, I decided to establish both protocols and measure the presence of HEPs in *in vitro* cultures.

3.2.2 Establishing intermediate mesoderm differentiation protocols

Both the Xia and Takasato protocol were followed as described (Figs.3.1a & 3.2b). Differentiation progression was monitored in the first instance with RT-qPCR across the course of differentiation. Firstly, the cultures exposed to the Xia protocol saw a steady decrease in pluripotency genes *POU5F1* and *NANOG* (Fig.3.1b). They were still detectable at D4, suggesting some undifferentiated hPSCs remained. As reported, *BRACHYURY* was sharply upregulated at D2 indicating the formation of primitive streak. However, *MIXL1* was not detectable at any point during differentiation. Genes used as early IM markers [254, 266] *LHX1* and *OSR1* showed differences from the initial paper [254]. Xia, *et. al.* report *LHX1* expression gradually increases and peak at D4. I, however saw an initial increase but peak at D3 with a sharp drop at D4. *OSR1* expression is also different. They report a gradual increase in *Osr1* expression, while I observe a peak at D2 but then a sharp drop off. Finally, the more specific and mature IM markers *PAX2* and *PAX8* were undetectable in my cultures (Fig.3.1b), while Xia, *et. al.* observe a gradual increase in expression. This suggests despite following their protocol as described, we are not directing differentiation in the same way and that, even though I see low expression of *KDR* (Fig.3.1b), these conditions are not suitable for generating aIM. I, therefore, focused on the Takasato protocol.

The Takasato protocol was followed up to the point of mature, specified IM (day 7) (Fig.3.2a), as measured by the expression of mature markers such as *PAX2, PAX8* and *ITGA8* [266]. Characteristic H9 hESC cell and colony morphology was lost in 24 hours with cells adopting more mesenchymal morphology and colonies losing sharp, well defined edges in exchange for rough borders (Fig.3.2b). By D3, after 24 hours in FGF9, cultures had expanded rapidly at 65% confluence. Additionally, although starting as a monolayer, small clusters of cells could be observed in the centre of colonies (Fig.3.2b, red arrows). These clusters became less apparent in D5 and D7. By D7, plates were confluent and any further differentiation would require splitting.

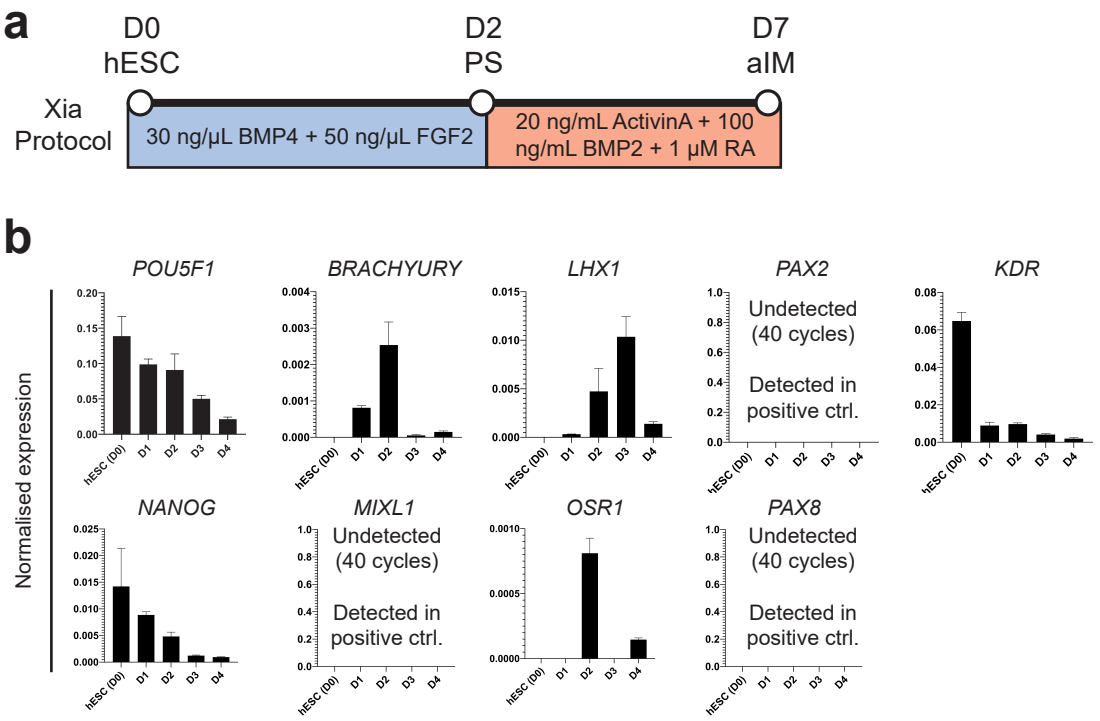


Figure 3.1: Assessment of the Xia differentiation protocol assessed by RT-qPCR
a Schematic representation of the first four days of Xia, *et. al.*, 2013 hPSC differentiation. hESC cells are differentiated for two days in BMP4 and FGF2 to induce a primitive streak state before direction to anterior intermediate mesoderm (aIM) with Activin A, BMP2 and Retinoic Acid (RA) for two days. **b** RT-qPCR was used to survey differentiation every 24 hours starting with hESCs at D0 up to D4. Whole cultures were taken for RNA extraction and cDNA synthesis. Gene expression is normalised to the average of three housekeeping genes (*GAPDH, ACTB, B2M*) and shown with bars showing mean ± SD (n = 3 independent differentiations). *MIXL1, PAX2* and *PAX8* were not detected in experimental samples after 40 PCR cycles despite being detected in the positive control (universal human cDNA).

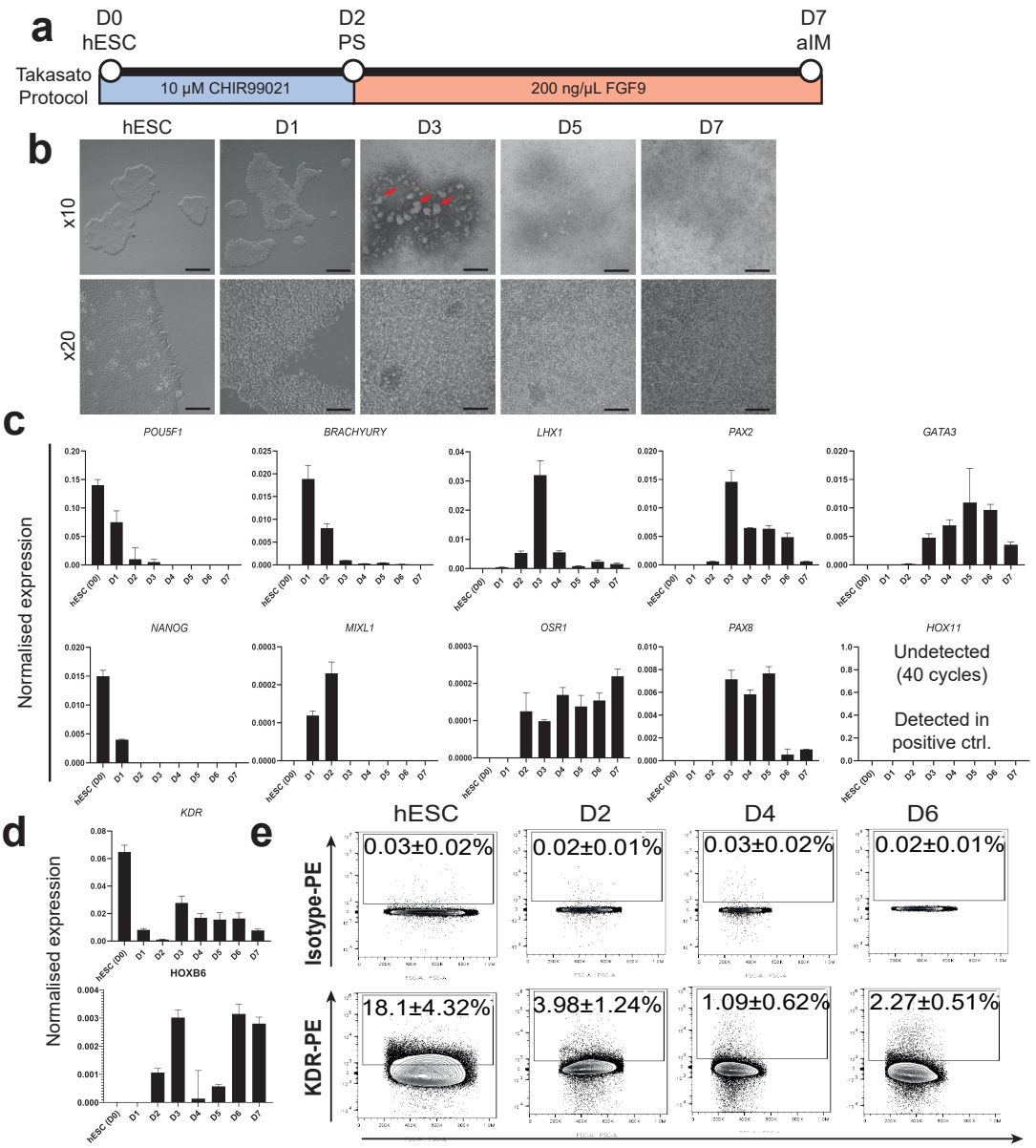


Figure 3.2: Intermediate mesoderm differentiation from hESCs
a Schematic representation of the first seven days of Takasato *et. al.*, 2016 hPSC differentiation. hESC cells are differentiated with 10 µM of the glycogen synthase kinase 3 inhibitor, CHIR99021 for two days. On the second day, media is switched to APEL supplemented with 200 ng/µL FGF9 for the remaining five days. **b** Brightfield images of colony (top, 10x) and cell (bottom, 20x) morphology starting from undifferentiated hESCs to D7 of differentiation. **c** RNA expression kinetics for a panel of genes of interest were monitored daily over the week differentiation protocol shown in **a**. Whole cultures were taken for RNA extraction and cDNA synthesis. Gene expression is normalised to the average of three housekeeping genes (*GAPDH, ACTB, B2M*) and shown with bars showing mean ± SD (n = 3 independent differentiations). **d** *KDR* gene expression during differentiation. **e** Flow cytometry plots pre-gated on live cells showing the kinetic of KDR protein expression along differentiation. Top row shows isotype controls and bottom row shows KDR-PE stained cells. Gates set using isotype control. Plots are representative of 3 differentiations with mean ± SD shown in the gate.

Differentiation progression was monitored daily by a panel of genes selected to detect the presence of a primitive streak state, ectoderm, endoderm, early and late IM by RT-qPCR (Fig.3.2c). Exit from pluripotency can be observed through the downregulation of pluripotency factors *POU5F1* and *NANOG* over the first two days of differentiation with complete absence by day three (D3). Subsequently, many of the cells transition through a primitive streak state as observed by the marked upregulation of *BRACHYURY* [273] and *MIXL1* [274] on D1. Also, *PAX6* and *FOXA1*, markers of ectoderm [77, 275] and endoderm [77, 276, 277] respectively are undetectable throughout differentiation showing no ectoderm or endoderm differentiation in culture (Appendix Fig.1). Together, these data show differentiation away from pluripotency through a primitive streak state and do not differentiate to alternative germ layers (Fig.3.2b).

To further document intermediate mesoderm development, markers of early and late IM were profiled. *LHX1* and *OSR1* have been shown to initially mark both LPM and IM but are subsequently restricted to the IM [74, 75, 269]. Both genes are upregulated shortly after the expression of primitive streak markers. *LHX1* expression peaks at D3 and is low afterwards. *OSR1* expression gradually rises and peaks at D7 (Fig.3.2c). This pattern mimics kinetics seen in published data reporting IM differentiation [253, 266]. Additionally, markers specific for mature IM: *PAX2* and *PAX8* are upregulated 24 hours after primitive streak markers and change of media from primitive streak induction to IM specification media (Fig.3.2c). Expression levels fluctuate over the subsequent five days, likely due to asynchronous differentiation. These data show the presence of IM, suggesting accurate reproduction of the original publication's differentiation protocol [253]. Finally we confirmed the IM differentiated was patterned as anterior IM and not posterior IM. *GATA3* marks the aIM and was expressed from D3, concomitant with the upregulation of *PAX2* and *PAX8*. Conversely, the posterior IM marker *HOXD11* was undetectable. Together these data show accurate production of anterior IM as previously reported [266].

With differentiation cultures established able to form IM, I next wanted to assess whether our cultures were permissive to production of LPM [129] and, specifically, to generation of haemato-endothelial progenitors. Identification of this population would suggest I could adapt the later stages of this protocol to form endothelium and blood. Concomitant with the upregulation of IM markers, the LPM marker *HOXB6*, used to trace the origin of HSCs to LPM was upregulated from D2 with two peaks, one at D3 and another at D6 (Fig.3.2d). Critically, the gene coding for the VEGF receptor, *KDR* (VEGFR2 or FLK-1), a fundamental receptor for the specification of haemato-endothelial progenitors [154, 278] was also expressed during mesoderm differentiation (Fig.3.2d). As seen in other hESC differentiation publications, expression was first seen in undifferentiated hESC [279–281] but this expression was lost as cells differentiate to a primitive streak state. *KDR* expression was subsequently upregulated from D3 and plateaued thereafter (Fig.3.2d).

To assess the proportion of cells receptive to VEGFA in differentiation cultures (positive for KDR protein), flow cytometric analysis throughout differentiation was performed (Fig.3.2e). This confirmed the *KDR* expression kinetics at the protein level. The highest percentage of KDR^+ cells was before differentiation started at 18.1% in hESCs as previously observed [279–281]. The percentage of KDR^+ did rapidly drop with the lowest at 1.09% at D4 before rising to 2.27% at D6. This small proportion is expected given culture conditions are optimised to give rise to KDR^- IM but will likely expand through addition of endothelial growth factors. Given the significant change in cell morphology, undetectable expression of pluripotency markers and upregulation of mesodermal markers, these KDR^+ cells are very unlikely to represent undifferentiated hESCs but rather the production of haemato-endothelial progenitors in culture.

A key aim of this work is to generate LPM in close proximity to IM as in the developing embryo [282]. To determine the formation of structure (indicative of interaction) between these two subsets, immunofluorescence microscopy for KDR and PAX8 was performed (Fig.3.3). KDR and PAX8 did not overlap in the vast majority of cells indicating only a

minority of KDR$^+$ were IM derived. Rare double positive cells were occasionally identified however (Fig.3.3i,iii red arrows) Additionally, KDR$^+$ cells formed spherical structures with KDR$^+$ cells seemingly aligned (Fig.3.3i) and interspersed with PAX8$^+$ cells. Most PAX8$^+$ cells formed groups around the edge of these KDR$^+$ spheres (Fig.3.3iii). The shape of cell groups and distribution of these two distinct cell types in these cultures exhibit signs of organisation. The interaction of these two cell types in this system requires further investigation. We decided to focus our efforts on differentiating these KDR$^+$ cells towards endothelium and blood.

3.2.3 KDR$^+$ cells generated in intermediate mesoderm differentiation conditions are capable of endothelial differentiation

KDR+ cells are not just restricted to HEPs but are also expressed in a subset of mesoderm [154]. Measuring the expression of TFs required for haemato-endothelial specification (*FLI1, ETV2 and SCL* indicated that there were at least some HEPs in these cultures (Fig.3.4b). Observations in model organisms and experiments from *in vitro* differentiation systems have demonstrated that HEPs activate an endothelial programme following VEGF signalling. Therefore, I adapted the differentiation protocol to direct the KDR$^+$ cells generated in intra-embryonic conditions by adding VEGF and bFGF, another potent endothelial inducer (Fig.3.4a) [283, 284]. Media was switched at D5 from mesoderm patterning media to endothelial inducing, as this is when *SCL* is upregulated (Fig.3.4b, likely representing the emergence of angioblasts from their mesodermal precursors. To assay for the emergence of endothelial cells, flow cytometric analysis for three endothelial markers: KDR, CD144 (VE-Cadherin) and CD31 (PECAM-1) was performed from D5 up to D7+5 with cells being replated on D7 to enrich for endothelial cells and split the confluent dish.

The addition of VEGF and bFGF at D5 dramatically increased the proportion of cells expressing endothelial markers (Fig.3.4b,c). Initially, cells transition from KDR$^+$CD144$^-$CD31$^-$, to KDR$^+$CD144$^+$CD31$^-$ and then finally upregulate CD31 to become triple positive (Fig.3.4b,c). The developmental relevance of these three stages

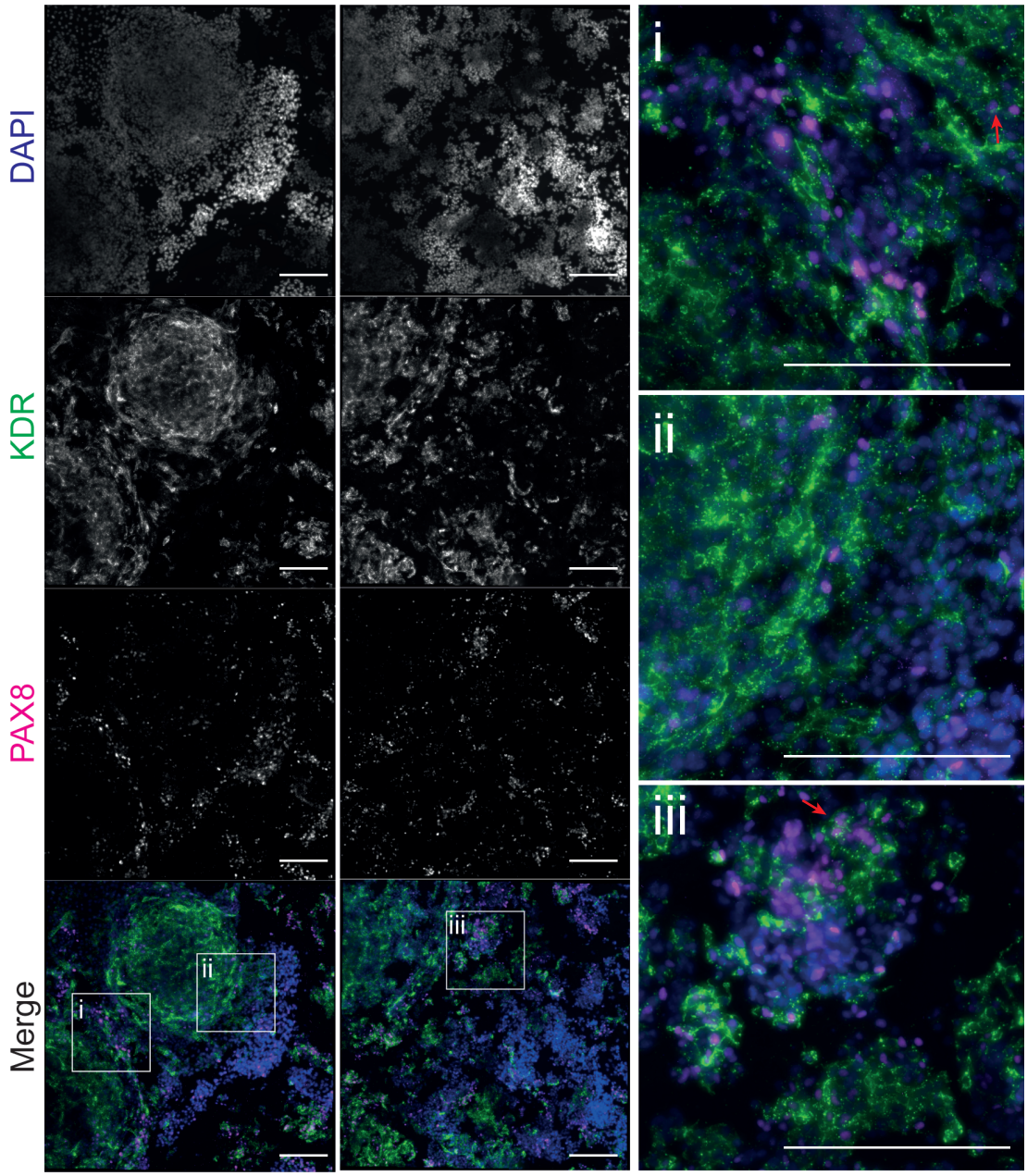


Figure 3.3: KDR and PAX8 are expressed in different cell subsets that exhibit organisation, structure and close proximity

Widefield fluorescent images of day 5 differentiation cultures from Takasato *et. al.*, 2016 protocol (Fig.3.2a) stained for DAPI (blue), KDR (green) and PAX8 (magenta). Three regions (i, ii and iii) are magnified where KDR and PAX8 are almost entirely mutually exclusive with only very rare cells co-expressing both (red arrows). KDR^+ cells primarily form spherical structures while $PAX8^+$ cells are sometimes intermingled within the spheres but the majority lie around the edge of the spheres. Scale bars represent 100 µm

are incompletely understood but could reflect maturation and commitment to an endothelial fate.

To determine whether the emergence of a $CD144^{+}CD31^{+}$ population reflects the upregulation of an endothelial programme, RT-qPCR was performed for many known endothelial genes (Fig.3.4d, Appendix Fig.2b). Concomitant with the upregulation of cell surface markers, most endothelial genes start with a low expression (as compared with human umbilical vein endothelial cells (HUVECs)) which rise by D7+3 (Fig.3.4d). The expression levels at D7+2 onwards in these genes is equal or greater than the expression in HUVECs, suggesting robust expression within the hESC-derived endothelial cells. In contrast, *TEK* and *VWF*, two alternative endothelial markers are much more highly expressed in HUVECs rather than our endothelium (Fig.3.4d). However, these genes do increase over the course of differentiation suggesting they are still expressed in our cultures but at a lesser extent, possibly due to each reflecting a different endothelial specialisation or immaturity.

In addition to a strong, well defined transcriptional programme endothelial cells have a distinct cellular morphology when grown on plastic. With the same morphology HUVECs, $CD144^{+}$ cells displayed a cobblestone morphology and displayed swirling patterns when confluent (Fig.3.4e). This is in stark contrast to $CD144^{-}$ cells which exhibited a range of cellular morphologies from small and round to mesenchymal and flat with multiple protrusions(Fig.3.4e). Additionally, endothelial cells have a strong propensity to form tubules when seeded in growth-factor reduced matrigel. hESC derived $CD144^{+}$ cells formed many tubules connecting adjacent cells and forming large networks in a manner indistinguishable from HUVECS (Fig.3.4f). Conversely, $CD144^{-}$ cells formed a small number of small protrusions in seemingly random directions but did not form tubules between cells and did not create an interlinked network (Fig.3.4f). Together, these data demonstrate successful generation of endothelial cells from angioblasts specified in conditions generating IM and thus possibly reflecting an intra-embryonic

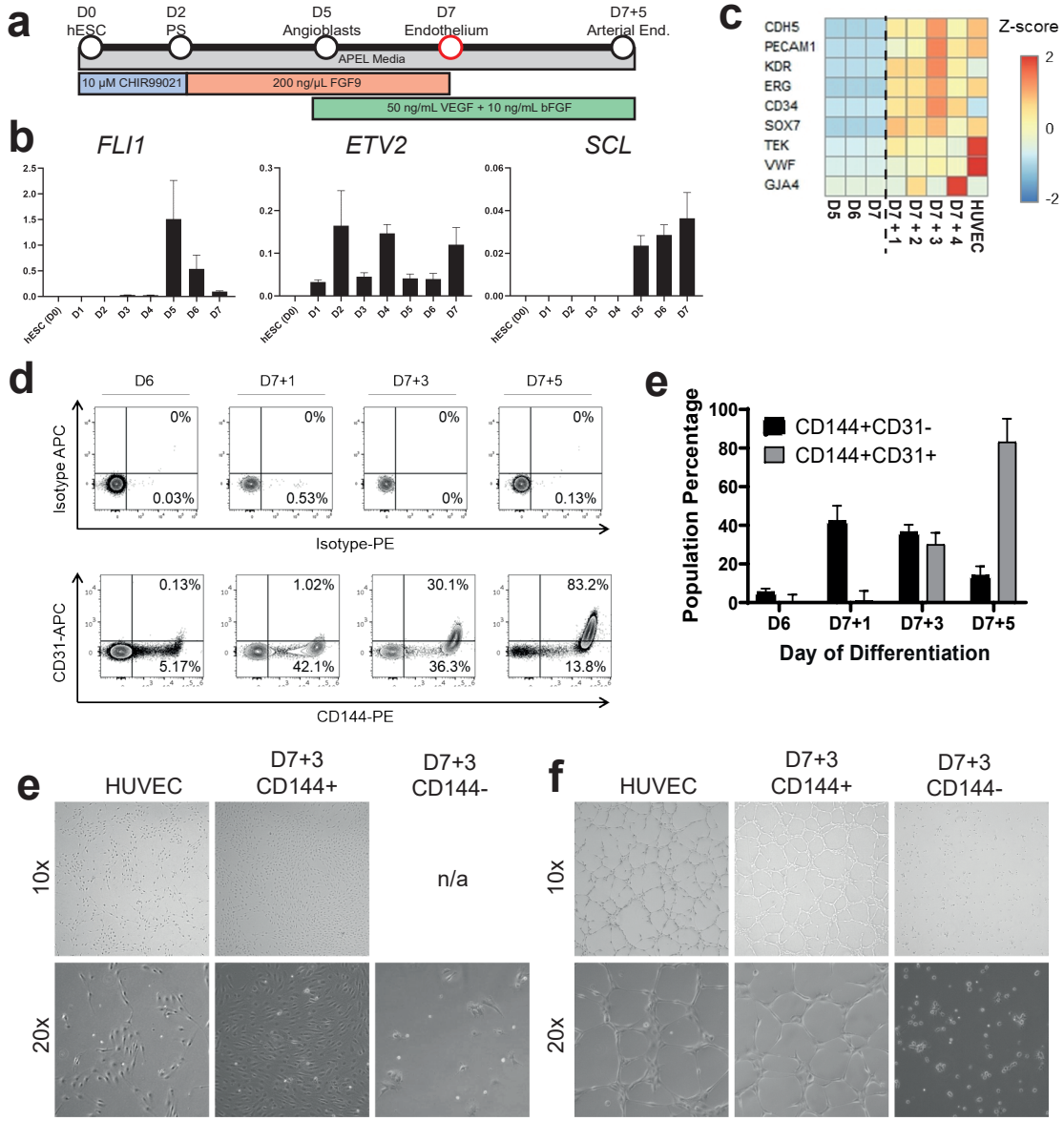


Figure 3.4: KDR^{+} cells are capable of differentiation into endothelium
a Schematic of the adaptation of Takasato *et. al.*, 2016 protocol (Fig.3.2a) through the addition of VEGF and bFGF from the fifth day of differentiation to direct KDR^{+} cells towards endothelium. White circles represent key stages in differentiation. Red circle indicates replating. **b** Gene expression kinetics showing mean ± SD of $2^{-\Delta Ct}$ values from Takasato protocol. **c** RT-qPCR measurement of endothelial gene expression kinetics from D5 to D7+4. cDNA derived from HUVECs was loaded as a positive control. $2^{-\Delta Ct}$ values are z-score normalised across rows (per gene). For standard deviation please refer to Appendix Fig.2.**d** Flow cytometric measurement of endothelial cell surface markers CD144 (VE-Cadherin) and CD31 (PECAM1) expression from D6 to D7+5. Representative plots show live cells with the top plots showing isotype controls for APC and PE and the bottom plots cells stained for CD31 and CD144. Percentages are mean of 4 independent differentiations. **e** Summary of 4 differentiations showing the mean ± SD of. **e** Morphology of cells following sorting cells based on CD144 expression. HUVECS are shown as example of endothelial morphology. **f** Tubule formation assay on cells sorted based on CD144 expression. CD144^{+} exhibit tubule formation ability, very similar to the positive control cells, HUVECs. CD144^{-} cells form cellular protrusions but not tubules.

environment.

3.2.4 Differentiation cultures exhibit the correct HOX expression pattern

A key difference between the extra- and intra-embryonic haematopoietic lineages are their positioning on the anterior-posterior (A-P) axis. The precursors of the haemogenic endothelium capable of generating HSCs *in vivo* is restricted to the trunk of the embryo across evolutionary time [282] whilst the precursors of the extra-embryonic lineage develop in the YS. However, anatomy cannot be used to judge positioning on the A-P axis *in vitro*. Therefore, a panel of HOX genes that mark the trunk of organisms were selected for expression analysis. Firstly, *HOXB4, HOXB5* and *HOXD4* expression (used to define the kidney morphogenic field [285]) show inconsistent expression kinetics (Fig.3.5a,). *HOXB4* and *HOXB5* are upregulated at D2, dip on D3, increase on D4 before being turned off by D5. Conversely, *HOXD4* sees very little expression for the first four days of differentiation but is sharply upregulated on D5. Similarly, *HOXA5 HOXA9*, and *HOXA10* (overexpressed to reprogram endothelial cells to HSCs [286]) are all sharply upregulated on D5 of differentiation and maintained over the subsequent two days (Fig.3.5a). Thirdly, *HOXC9* and *HOXC10*, shown to inhibit anterior limb development and thus mark the trunk show the same expression pattern [287]. Interestingly, the only HOX gene to show characteristics of both early and late expression is *HOXB6*, the HOX gene used to trace the LPM origin of HSCs (Fig.3.5a)[86, 87]. Finally, we assessed the expression of a postierior HOX gene, *HOXB13* which was undetectable after 40 PCR cycles. Although difficult to conclude exact position along the A-P axis, expression of many of the selected HOX genes indicates many of the cells formed in these conditions likely represent those found in the embryonic trunk and not the posterior, an evolutionary conserved pre-requisite to HSC generation [129].

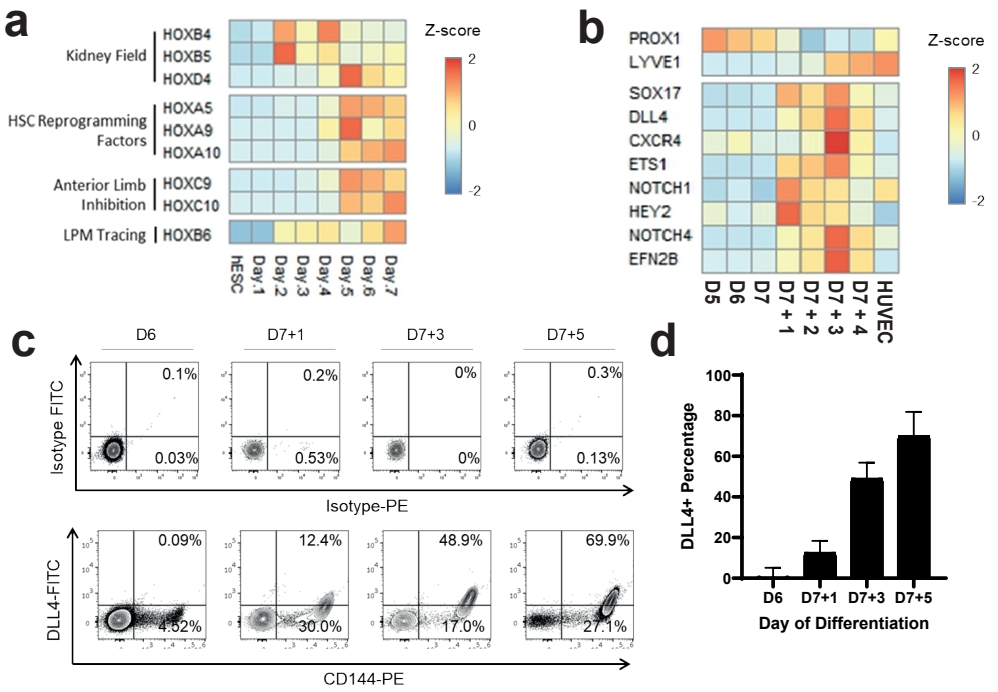


Figure 3.5: Haemato-endothelial progenitors differentiated in high levels of VEGF give rise to arterial endothelium

a RT-qPCR measurement of HOX genes across the first seven days of differentiation. HOX genes were selected and are grouped by four criteria. Those described as being critical for forming the kidney field, those used during reprogramming to HSCs, those important for inhibition of the anterior hind limb and HOXB6 which was used to fate map the mesodermal ancestors of HSCs. $2^{-\Delta Ct}$ values are z-score normalised across rows (per gene). Mean $\pm$ SD expression shown in Appendix Fig.3. **b** Flow cytometric measurement of the arterial cell surface marker Dll4 during endothelial differentiation (Fig.3.4a). Dll4 expression starts three days later than the endothelial marker CD144 but by D7+5, 69.9% of cells are expressing Dll4. **c** RT-qPCR measurement of arterial and venous marker gene expression kinetics across endothelial differentiation starting at D5 and ending at D7+5. HUVECs were used as a positive control for venous genes (top) and negative control for arterial genes (bottom). $2^{-\Delta Ct}$ values are z-score normalised across rows (per gene). Mean $\pm$ SD can be seen in Appendix Fig. 4.

3.2.5 Angioblasts differentiated in high levels of VEGF give rise to arterial endothelium

In recent years, the relationship between haemogenic and arterial endothelium has been controversial [160]. One school of thought suggests that arterial and haemogenic endothelium represent distinct lineages [223] while more recent studies suggest that the arterial programme is required for definitive blood production from a haemogenic endothelium [224, 225]. Therefore, I assessed the presence of an arterial programme in our cultures by flow cytometry and RT-qPCR. The NOTCH ligand and arterial marker delta-like 4 (DLL4) was initally absent in $CD144^+$ cells at D6. Expression levels were gradually increased upon replating on D7 with the proportion of cells expressing it increasing up to 70% five days after replating (Fig.3.5c). This initial negativity and gradual increase in expression levels and proportion reflects our stepwise model of HE differentiation built from observations *in vivo*.

While DLL4 expression is indicative of an arterial identity, RT-qPCR for a panel of genes was used to determine the expression of other common arterial markers (Fig.3.5b). Reflecting DLL4 expression, all arterial markers analysed increased during differentiation and crucially were highly enriched when compared to HUVECs as a negative, venous control(Fig.3.5b). This suggests a robust activation of the arterial programme. Meanwhile the lymphatic marker genes *PROX1* and *LYVE1* did not share the same patterns. *PROX1* was initially high in cultures at D6 but was quickly downregulated by D7+1 to a level far below HUVECs. This initial expression before replating likely reflects non-endothelial expression that is downregulated after replating cells. *LYVE1* on the other hand is absent up to D7+3 where there is a sudden increase in expression at levels similar to HUVECs. This could explain the 27% of $CD144^+DLL4^-$ in the culture (Fig.3.5c).

3.2.6 Inducing a haematopoietic programme in arterial endothelium cultures is highly variable

Thus far I have shown that the IM differentiation culture harbour endothelial progenitors that can be directed to arterial endothelium and form close contact with intermediate mesoderm. I aimed to further this differentiation through the induction of a haemogenic and subsequent haematopoietic programme within the endothelium. Many protocols have been developed to induce these programmes in hESC differentiation cultures that share cytokine cocktails [212, 213, 223], mostly derived from knowledge gained from *in vivo* experimentation. Therefore, I mirrored these protocols through the addition of cytokines to cultures sorted for endothelium at D7 and monitored for the emergence of CD45+ cells by flow cytometry.

The first two differentiation cultures subjected to haematopoietic cytokines (Fig.3.6a) were 25% positive for CD45 indicating endothelium had undergone EHT and formed haematopoietic cells. Reassuringly, these flow cytometric plots closely resembled those from other hESC differentiation conditions with five populations detectable using only CD144 and CD45 (Fig.3.6c, $CD144^{+}CD45^{-}$, $CD144^{-}CD45^{+}$, double negative, $CD144^{mid}CD45^{+}$ and $CD144^{lo}CD45^{+}$)[223]. A day later, the double positive populations had reduced but with 21% of cells still positive. However, in the subsequent six differentiations run with these conditions (Fig.3.6a), CD45 percentage was never detected over 4% (Fig.3.6d) with very rare cells scattered in $CD45^{+}$ gate and no discernible populations forming (Fig.3.6e). Many possible causes for this sudden difference in differentiation were tested including mycoplasma, cytokine batch or expiry, hESC passage number, basal media batch, seeding density and antibody staining. Unfortunately, this thorough investigation of multiple factors did not return the differentiation to its initial success. The inability to replicate these initial haematopoietic differentiation points to the high sensitivity of these protocols to small, unidentifiable change and requirement to identify more robust differentiation conditions.

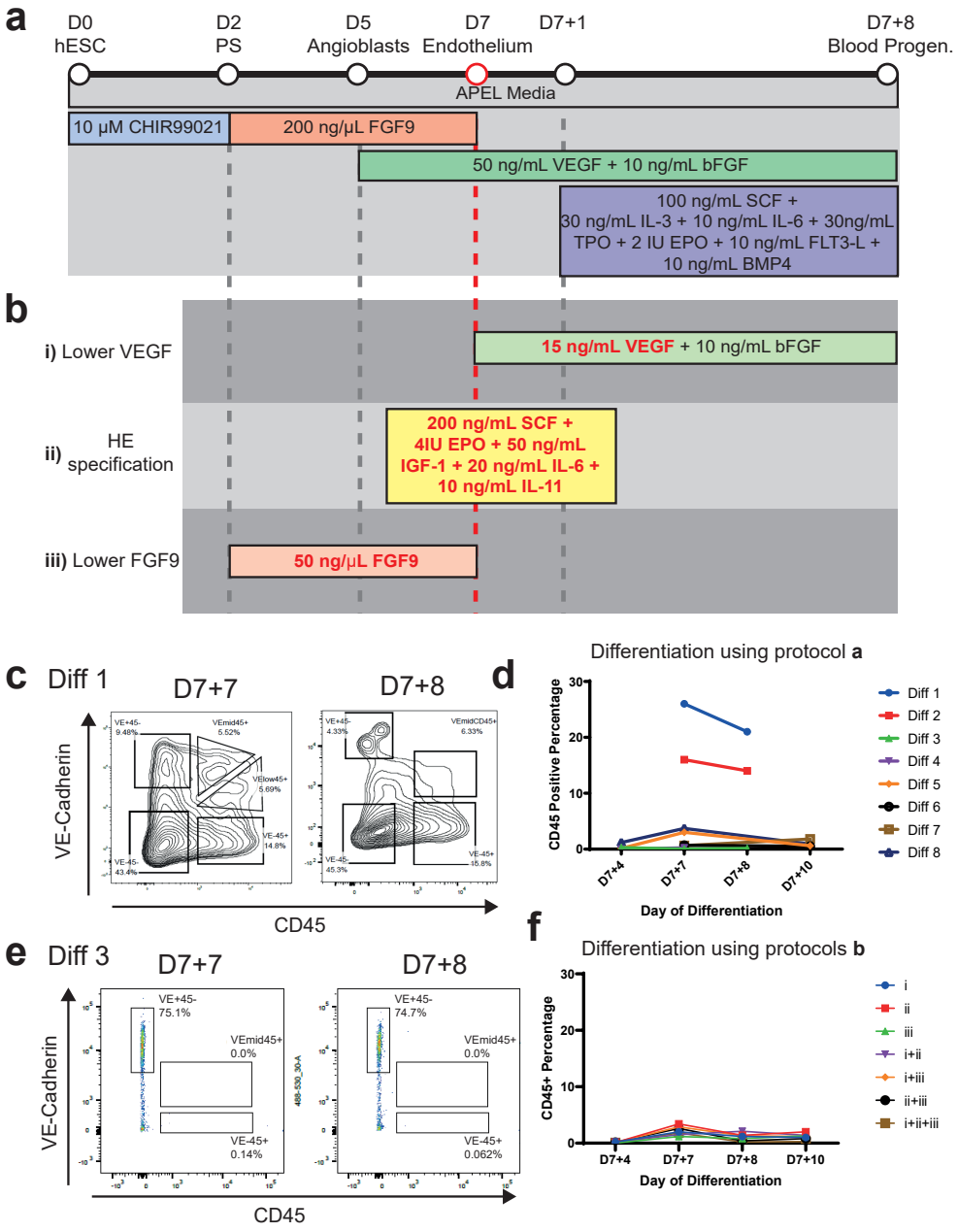


Figure 3.6: Addition of haematopoietic cytokines is unable to repeatably induce a haematopoietic programme in arterial endothelium

a Schematic showing the differentiation conditions hESCs were initially subjected to over 15 days. Red circle at D7 indicate the sorting of $CD144^+$ cells. **b** After the irreproducible blood differentiation using conditions in **a**, three adaptations were tested in independently and in combination. i) Lower VEGF concentration to 15 ng/mL was used in endothelial specification and maintenance. ii) A HE specification phase was added earlier in differentiation. iii) FGF9 concentration was lowered during IM patterning. **c** FACS plots showing the distribution of cells with VE-Cadherin (CD144) and CD45 expression at D7+1+7 and D7+1+8 of the successful blood differentiation (Diff1) using conditions in **a**. **d** A summary of $CD45^+$% across the eight differentiations run using conditions **a**. **e** Representative FACS plots of differentiations 3-8 showing VE-Cadherin and CD45 expression at D7+7 and D7+8. **f** Summary of CD45 percentage in conditions summerised in **b** at 4 time points.

Therefore, I aimed to find differentiation culture conditions capable of reliably inducing haematopoietic differentiation from the arterial endothelium. Three adaptations were tested on their own and in combination (Fig.3.6b). Firstly, the concentration of VEGF after sorting for endothelium at D7 was reduced from 50 ng/mL to 15 ng/mL (Fig.3.6bi) with the rationale that while high VEGF concentration robustly induces an arterial fate in endothelium (Fig.3.5), it could inhibit a haematopoietic programme if too robust [288]. Secondly, previous studies have used a short period of SCF, EPO, IGF-1, IL-6 and IL-11 to induce a haemogenic programme before switching to haematopoietic differentiation conditions (Fig.3.6bii) [223]. Finally, the concentration of FGF9 used to induce IM fate in differentiation cultures was dropped from 200 ng/µL to 50 ng/µL as the impact of FGF9 on haematopoietic fate specification is unknown (Fig.3.6biii).

Each of these three conditions (Fig.3.6b) were tested independently and in combination. Similarly to the initial differentiation culture conditions (Fig.3.6a), none of these conditions caused the increase in $CD45^+$ cells (Fig.3.6f), demonstrating a haematopoietic programme had not been induced. All of the conditions had fewer than 3.5% of $CD45^+$ cells. Such small percentages of cells make the profiling of them and their progenitors very difficult. Rather than study rare populations of cells or blindly test different conditions, I thought it best to reflect on our findings and assess the gaps in our understanding halting empirical adjustment of culture conditions.

3.3 Discussion

This chapter aimed to reproduce an IM differentiation protocol to establish an *in vitro* niche similar to that identified across evolution *in vivo* [129] and direct any haemato-endothelial progenitors that developed in this intra-embryonic environment first towards endothelium and then definitive HE. I first established an IM differentiation protocol and identified a KDR^+ cell population before modifying conditions to generate arterial endothelium. However, despite some initial blood differentiation, multiple protocols could not induce haematopoietic differentiation in the arterial endothelium.

The conditions tested were based on our understanding of blood development in the embryo and any further adjustments to the conditions would have been empirical and not founded in biological understanding of development. Rather than testing conditions in this way, it was a good time to put our results into context and see what was missing from our understanding to induce haematopoiesis in the generated arterial endothelium.

The relationship between blood and endothelial development is longstanding and widely accepted [289, 290]. However, the details of how and when different endothelial lineages segregate is still contentious. In particular the relationship between arterial and haemogenic endothelium. Definitive HE is almost exclusively found in arterial vessels [119, 291] and expresses many arterial genes (*EphnB2, Dll4, Notch4* and *Sox17*)[288, 292]. From D7+1, hESC cultures differentiated in these conditions were expressing a broad range of these arterial markers and were highly enriched for endothelium and as such in an arterial niche. Therefore, our cultures represent a good starting point from which to induce definitive haematopoietic development.

One possibility in the inability to subsequently induce this programme is that I produced vascular arterial endothelium with no haemogenic competency. Two models have emerged on how arterial and haemogenic endothelium are related. One proposes arterial fate acquisition is essential for HSC emergence, and the other suggests HSC specification is arterial independent and relies on a permissive 'arterial niche' (e.g. high Notch signalling) with HE representing a distinct lineage than that of arterial endothelium [160]. This could indicate therefore, that arterial specification in my hESC cultures does not necessarily demonstrate the progression along the correct differentiation trajectory and could indicate only vascular endothelium is being specified. However, work in zebrafish focusing on the initiation of *runx1* transcription in arterial cells and removing translation maintaining arterial identity would indicate this not likely the case [169]. This model is supported by the pseudotime ordering of scRNA-seq data from human and mouse DA showing a progression from arterial endothelium through an arterial HE

to HSC identity [168, 170, 171].

Many groups are working to translate the knowledge gained from *in vivo* work to replicate the process *in vitro* and better understand human-specific phenomena of blood development and one-day make haematopoietic progenitors with clinical applications (as discussed in section 1.4.3). To our knowledge, all hPSC derived definitive HE reporting arterial identity induces an arterial programme in the HE rather than a haemogenic programme in arterial endothelium [223–225]. This reverses the current dogma from *in vivo* studies [169–171] that the haemogenic transcriptional programme is initiated in an arterial endothelium (as discussed in section 1.3.3). Whether the arterialisation of HE rather than the reverse observed in hESC differentiation is a consequence of *in vitro* differentiation, a YS associated mechanism or human specific phenomenon remains unknown. If it is a human specific phenomenon, it could explain the inability of the arterial endothelium to induce a haemogenic programme even with adjusted VEGF levels or HE specification phase.

Although Wave 3 HE is detectable from E9.5, the specification of haematopoietic fate occurs much earlier between E7.5 and E8.5 [218]. One possibility is that blood specification was not enacted on angioblasts within hESC differentiation cultures and therefore cells were not receptive to the instructive cues of haematopoietic cytokines. While *SCL* was confirmed to be expressed (Fig.3.4b), it is not clear if this is sufficient to specify a blood fate or if other currently unknown factors are required. These could be cell extrinsic signals inducing the expression of required cofactors in a subset of cells or a permissive chromatin state within cells [218]. It is clear that the AGM is a complex developmental niche with many cell signalling and physical factors required for *de novo* HSC development. While I aimed to manually recapitulate one aspect of the evolving niche through IM induction, the microenvironment of our culture system still lacks many of these other components, some of which are likely required for the proper specification of the blood lineage in concert with SCL [218]. This is known to be true at E10.5 where the AGM is severely limited in HSC production without signalling from ventral tissues

(e.g. gut) [293–295].

Blood vessels are distributed across the entire organism and vary hugely in their gene expression, morphology, growth and function depending on their location [296–298]. The cause of this tissue-specific nature of endothelium is unclear but is likely caused by heterogeneity in the local environments in which they develop and their origins. Evidence of anterior-posterior patterning of endothelium during development has a yet unknown importance in the establishment of organ-specific functions [299]. The renal vasculature forms a crucial part of the functioning adult kidney and is highly specialised in its structure and function. While the precise origins of the renal vasculature are unclear, broadly, the embryonic kidney is seeded through both angiogeneic branching from major renal arteries and vasculogenesis from Flk1$^+$ angioblasts of the developing kidney [300]. Notably, Osr1$^+$ cells contribute to the renal vasculature between E7.5-E11.5. These Osr1-Cre labelled endothelial cells were not restricted to the kidney but did not contribute to the HSC-derived kidney macrophages, suggesting definitive HE does not originate from Osr1$^+$ mesoderm [74].

One hypothesis for the absence of haemogenic competence in commonly used haematopoietic cytokines is that the arterial endothelium generated is not equivalent to that of the major embryonic arteries described to contain HE [119] but instead renal arterial endothelium. This could be tested through measuring the coexpression of OSR1 and KDR in the same cell. I tested many antibodies against OSR1 for other purposes but could not identify specific staining and so single-cell RT-qPCR or FISH would be required. Coexpression would indicate the mesodermal cells produced in the tested conditions likely reflects renal rather than DA angioblasts. It is not clear whether this rationalisation of angioblasts is a requirement to induce haemogenic endothelium but could feasibly inhibit haemogenic competency. However, rather than test possibly endless hypotheses on why the arterial endothelium produced in these culture conditions would not further differentiate, I decided to better characterise angioblasts during development

in the mouse embryo.

Chapter 4

Defining the cellular neighbourhood of intra-embryonic haemato-endothelial progenitors

Contents

4.1 Introduction 93
4.2 Results 95
4.2.1 *Scl* positive angioblasts are widespread throughout the embryonic posterior 95
4.2.2 Establishing Spatial Genomics analysis 99
4.2.3 Selection of three embryos for spatial genomics analysis 110
4.2.4 Qualitative assessment of haemato-endothelial progenitors and neighbouring mesodermal cells by multiplexed HCR 111
4.2.5 Assessment of RNA degradation during multiplexed dHCR 117
4.2.6 Quantitation of multiplexed HCR enables the spatial remapping of clusters 120
4.2.7 Investigating the degree of heterogeneity in gene expression between spatially distinct haemato-endothelial progenitors 125
4.3 Discussion 127

4.1 Introduction

Haematopoietic stem cells emerge within the mouse embryo between E10.5 and E11.5, arising from HE along distinct portions of the major embryonic arteries [119, 120]. Functional HSCs have only been detected from these intra-embryonic vessels and not from the HE present in the extra-embryonic YS that contributes to the Wave 2 definitive haematopoiesis. What endows Wave 3 HE with this HSC competency is yet unknown. Much research focus has been paid to the gene regulatory networks and cellular signalling that control the transition from HE to pre-HSC to describe the cell states and environment required to generate these rare stem cells [165, 186, 190, 192, 195–197, 294, 295, 301–307]. I propose that, to understand how HSC-competent

HE is programmed, it is critical to study the developmental trajectory that cells take to their adult state [218]. In this context, trajectory means to define the step-wise cell state decisions a cell takes from pluripotent epiblast to naive mesoderm through to angioblasts, endothelium, arterial endothelium and finally HE. Understanding the history of a cell is critical with a handful of examples of a cell's origin influencing fate and function in the haematoendothelial lineages alone. For example, macrophages derived from Wave 2 haematopoiesis typically reside in tissues performing homeostatic functions while HSC-derived macrophages are typically responsible for inflammation [308]. Additionally, a recent study has demonstrated that transitioning through paraxial mesoderm is crucial to imprint lymphatic endothelium fate and suggests lineage history is critical in creating EC heterogeneity [66].

Following the inconsistency in haemogenic versus vascular differentiation from our novel hESC differentiation system, I wanted to explore the signals that grant haemogenic competency. There is strong evidence that blood fate is conferred days before the emergence of haematopoietic cells. SCL, the transcription factor essential for blood fate specification, is no longer required for blood fate specification after E8.5 as demonstrated by the Tie2-Cre dependent excision of *Scl* [252]. Given its essential role in blood fate specification and ability to rescue blood emergence from *Etv2*$^{-/-}$ mESC differentiations [145], SCL is often depicted as the earliest regulator between vascular and haemogenic endothelium [143, 309]. Drake and Fleming observe SCL to be widespread amongst embryonic angioblasts [251], questioning which other signals are required to impart a haematopoietic fate on HEPs.

Despite the importance of lineage history and the unique HSC-competency of Wave 3 HE, Wave 3 angioblasts are relatively poorly understudied. Particularly lacking is a clear picture of the heterogeneity in angioblasts during the window of Wave 3 blood fate specification (E7.5 - E8.5). The extent to which spatially separated progenitors are transcriptionally distinct, the variety in their cellular neighbourhoods and

the consequences of those differences on cellular signalling requires further exploration.

Addressing these gaps in our understanding requires the application of methods that can probe the expression of multiple genes simultaneously while retaining cellular location. Additionally, the selected methodology must be sensitive to allow detection of low expressed genes and subtle differences in expression as any spatially distinct programmes are likely slight. I chose to establish an updated version of Spatial Genomics Analysis (SGA) [258]. SGA quantifies smRNA-FISH that can be multiplexed through DNaseI-based stripping of signal and subsequent restaining [241]. Quantification is performed through counting the number of single mRNA molecules within cell borders defined by antibody staining of membrane-bound proteins. In this way, a gene-count matrix for tens of genes can be generated at single-cell resolution to perform hierarchical clustering and cell-type annotation before remapping these clusters back on to the section. This chapter aims to establish SGA on the caudal region of E8.0 mouse embryos to carefully map the ancestors of HSC-competent HE [129, 173] and sensitively measure the expression of known fate determinants while defining their neighbouring cell types.

4.2 Results

4.2.1 *Scl* positive angioblasts are widespread throughout the embryonic posterior

To gain a greater understanding of angioblasts undergoing blood fate specification, I first wanted to map their location in the embryonic posterior with high spatial resolution. 3D confocal imaging of the embryonic posterior (Fig.4.1a) stained for *Scl* reveals the location of each HEP. Additionally, markers for pre-somitic mesoderm (PSM, *Tbx6*), intermediate mesoderm (IM, *Lhx1*), and the earliest arterial specification (*Dll4*) [155] were stained to provide insight into the surrounding cells into different HEPs and monitor heterogeneity in arterial specification (Fig.4.1c).

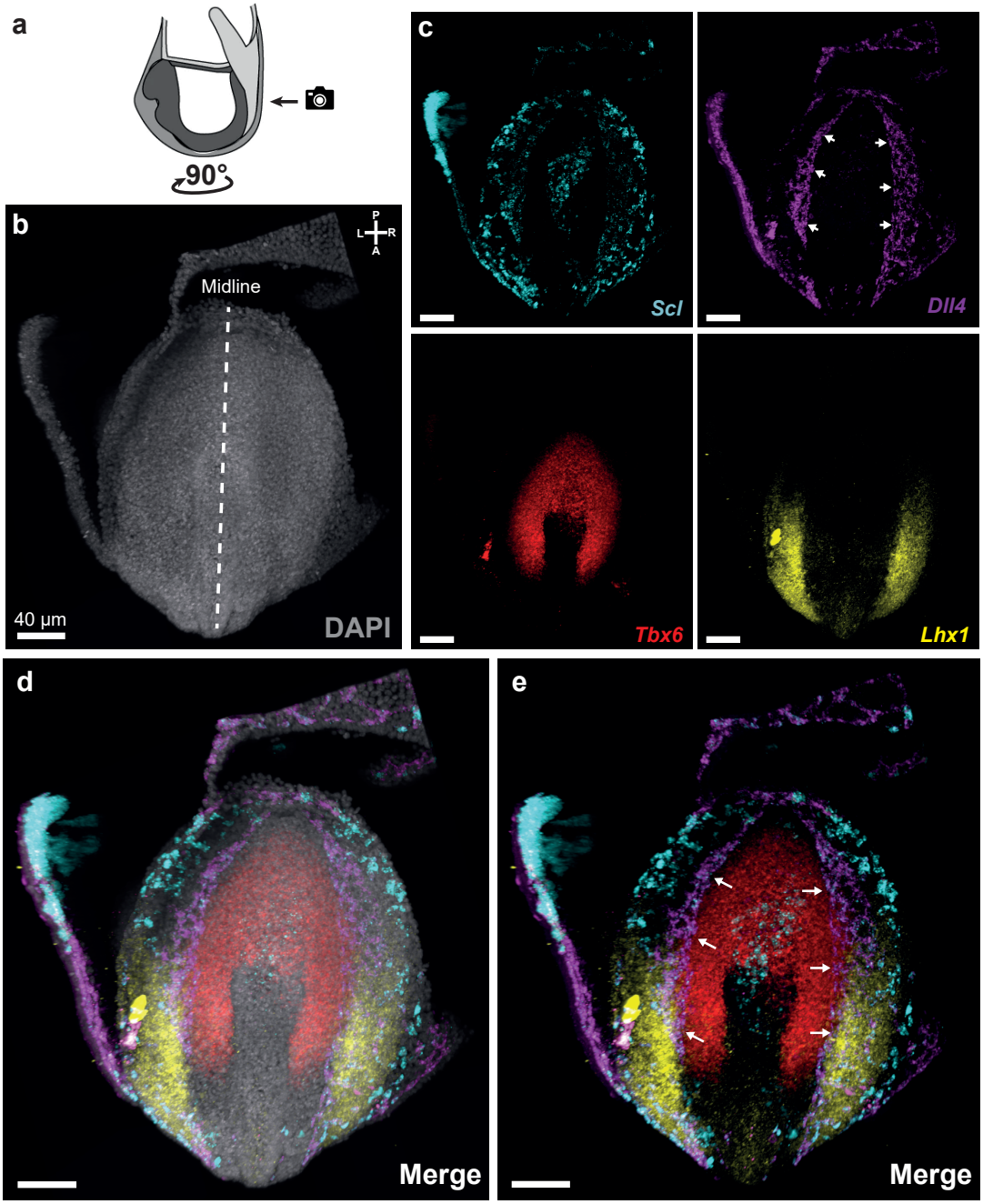


Figure 4.1: Whole mount imaging of the E8.0 mouse embryo posterior for location of Scl positive angioblasts.
E8.0 mouse embryos were dissected out of maternal tissue, stained with qHCR and imaged by 3D confocal microscopy. **c**-**e** show maximum intensity projections of the posterior view of one representative embryo (n=4) as shown schematically in **a**. **b** shows DAPI staining and indicates the midline of the embryo as well as a compass for the posterior (P), anterior (A), left (L), and right (R). Each gene stained is shown in **c** with *Scl, Dll4, Tbx6,* and *Lhx1* shown in cyan, magenta, red, and yellow respectively. The merge with and without DAPI is shown in **d** and **e**. White arrows in *Dll4* and right merge indicate location of paired dorsal aortae. Scale bar represents 40 µm

At E8.0, *Scl* is in many cells scattered throughout the caudal embryo and is not restricted to the developing dorsal aortae which, at this time, lie parallel to the embryonic midline in bilateral strips (Fig 4.1b). Many Scl^+ cells could be identified more laterally, outside of the dorsal aortae that did not form chords or seem to resemble primitive vessel structures. Furthermore, while I focused on the embryonic posterior as the site of HSC ancestor development [173], I also observed strong *Scl* staining in both the head mesoderm and all along the adjacent YS. Interestingly, the Scl^+ cells in the YS seemed continuous with those in the embryo (Fig.4.1).

By comparing the expression patterns of *Dll4*, I can see that it is more restricted, only present in the developing dorsal aortae and connected YS tissue. Additionally, when assessing the signal intensity of *Scl* (correlates with expression level [239]), in $Dll4^+$ cells, it seems *Scl* is dimmer, suggesting lower expression than in the more lateral and posterior angioblasts. *Tbx6* and *Lhx1* expression does not overlap with *Scl* as expected due to their association with other mesodermal cell types. *Tbx6* expression is primarily more posterior to that of *Lhx1* with an overlapping region of 60 µm (Fig.4.1e). Scl^+ cells were in close proximity to both populations with the DA sitting adjacent to the both.

Using Imaris to perform nuclei segmentation, the total number of cells can be quantified (Fig.4.2). To focus on the embryonic cells, regions were selected that excluded the YS and head cells contained within these data as shown in Fig. 4.2b. Segmented nuclei can be assigned as positive and negative for genes imaged by thresholding channel intensity in the surrounding 3 µm. The location of these quantified cells (Fig.4.2c) closely matches the expression domains of the genes probed (Fig.4.1c) giving confidence to the quantification method. A mean of 12,596 nuclei were segmented in the E8.0 posterior with an average of 4.58% detected as Scl^+ (Fig.4.2e). Of these Scl^+ cells, a mean of 10% are also $Dll4^+$ at this stage.

To better define the location of angioblasts along the dorsal-ventral axis within the E8.0 posterior, volumetric images were virtually sectioned at three transverse levels and

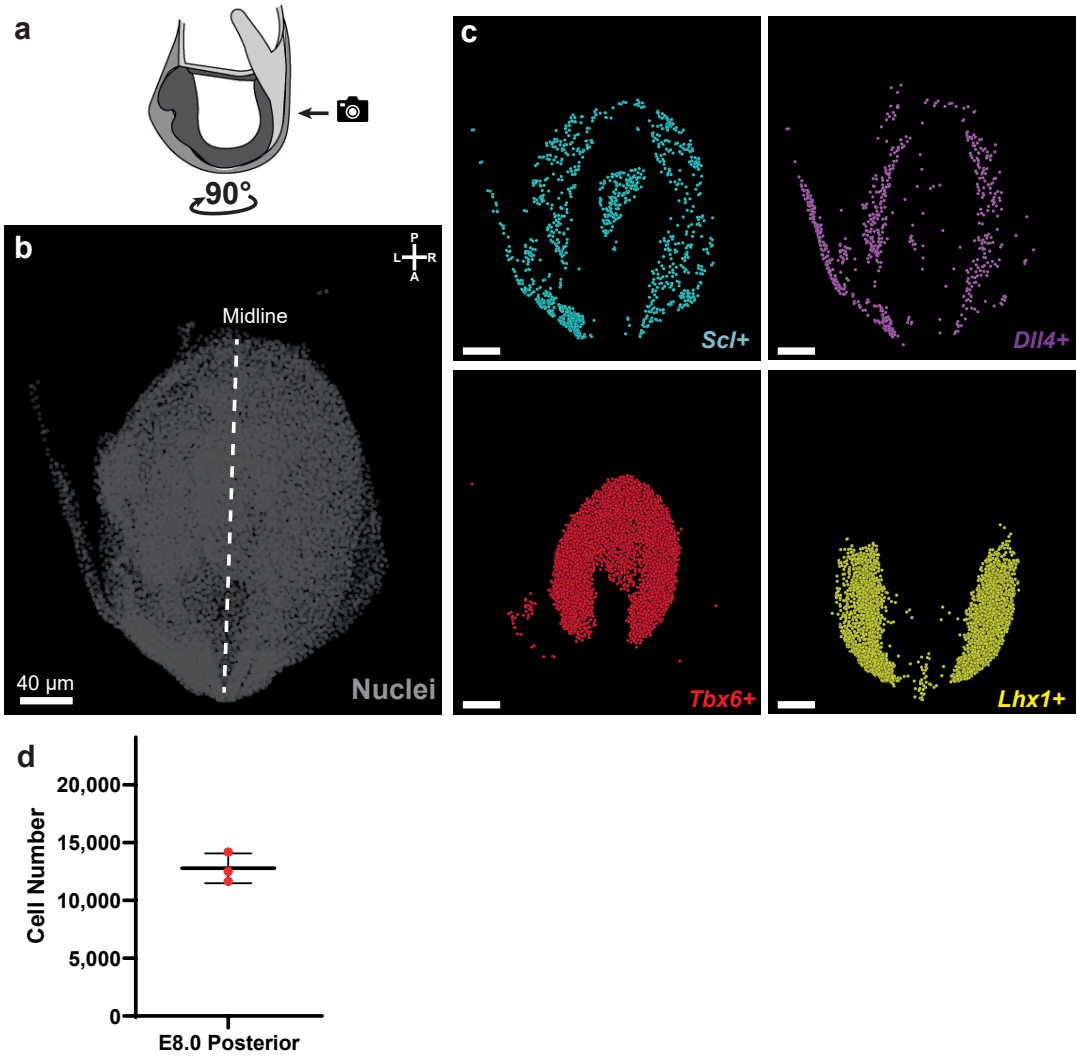


Figure 4.2: Quantification of angioblast numbers in the E8.0 posterior
Cells from wholemount imaging of E8.0 posterior (Fig.4.1) were segmented into individual cells (**b**) and assigned overlapping identities based upon expression of *Scl*, *Dll4*, *Tbx6*, and *Lhx1* (**c**). Total number of cells segmented in the posterior is shown in **d**.

one sagittal level using the oblique slicer in Imaris (Fig.4.3). *Scl*$^{+}$ angioblasts were notably only on the edges of the mesoderm close to either the surface ectoderm (orange arrow) or the definitive endoderm (pink arrow) (Fig.4.3b). Dorsal aortae angioblasts can be identified as the most medial *Scl*$^{+}$ cells on the ventral aspect of the mesoderm (white arrows). These DA angioblasts can be seen at all three transverse levels (Fig.4.3a, b and c) and along the entire length of the embryonic posterior (Fig.4.3d). These angioblasts reside at the junction between *Tbx6* and *Lhx1*. Similarly, the *Scl*$^{+}$ angioblasts on the ventral side that lie more laterally (pink arrows), only adjacent to *Lhx1*$^{+}$ mesoderm, are distributed along the length of the posterior but in a less structured form. At the most posterior levels (Fig.4.3a), there are no other angioblasts present. Conversely, halfway down the posterior, there are also angioblasts on the dorsal side situated directly next to the surface ectoderm (orange arrows). The most angioblast locations are seen at the most-distal posterior level. Here, I see angioblasts along the dorsal and ventral side but also adjacent to the neuroepithelium (green arrows) and surrounded by *Tbx6*$^{+}$ mesoderm (Fig.4.3c).

As seen from the posterior view (Fig.4.1), angioblasts of the future dorsal aortae and their most lateral *Scl*$^{+}$ neighbours express equally high levels of *Dll4*. However from the transverse sections I can also distinguish the expression in the dorsal angioblasts. Signal in these cells seems lower in these cells suggesting lower expression. Additionally, the two most medial angioblasts in Fig. 4.3c exhibit differential *Dll4* expression with dorsal cells seemingly absent of expression.

4.2.2 Establishing Spatial Genomics analysis

Having characterised where *Scl*$^{+}$ angioblasts are within the E8.0 embryonic posterior, I wanted to probe further their expression of other key endothelial and blood markers, while better defining their cellular neighbourhood with multiple markers for each cell of interest. SGA (summarised in Fig.4.4) provides a method to quantify the expression of tens of genes while retaining their spatial information. This has previously only been

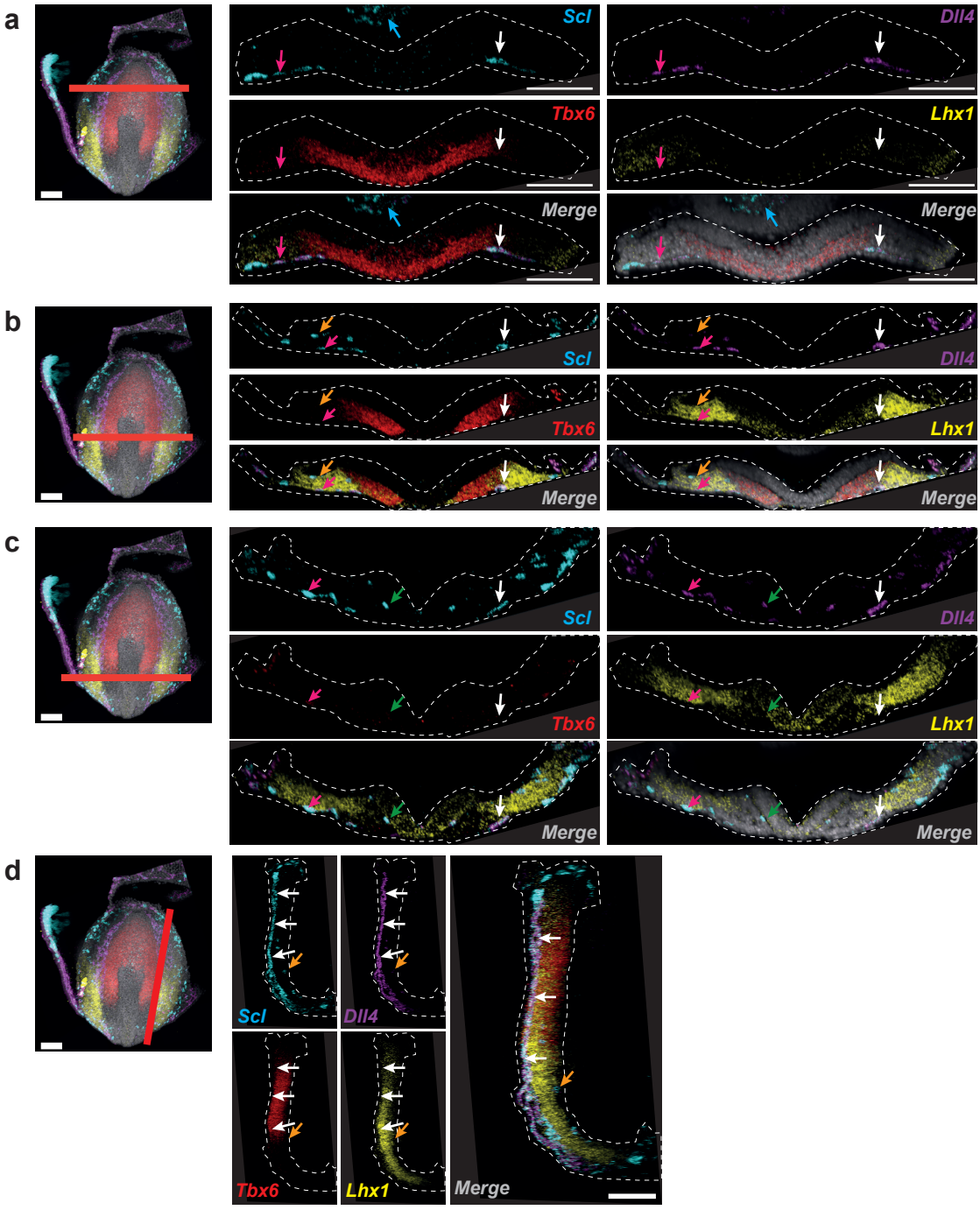


Figure 4.3: Virtual sectioning of E8.0 mouse posterior.
3D confocal data was 'virtually sectioned' using the oblique view in Imaris at three transverse levels (**a**, **b** and **c**), and one sagittal level (**d**). The plane of sectioning is represented by the red line on the posterior view on the left and maximum intensity projections of 15 µm are shown on the right. Each gene is shown separately and in a merge, with and without DAPI counter-stain. Representative images from 1 of 4 embryos. Scale bars represent 40 µm and dotted lines outline the border of posterior embryonic tissue. Blue arrows, *Scl*+ cells in head; white arrows, right DA; orange arrows, dorsolateral angioblast; pink arrows, ventrolateral angioblasts; green arrows, medial angioblasts adjacent to neural epithelium.

applied to the neuroepithelium of stage 9 chicken embryos [310] and so demonstrating applicability in sectioned mouse embryo was the first aim. Additionally, since SGA was first published, improved versions of HCR [239] and new computational tools have been released and so implementing these during set up could improve usability and accuracy.

Firstly, mouse embryos were fixed, frozen, dehydrated, rehydrated, equilibrated in sucrose, and embedded inside their decidua. I and others have observed this significantly improves cryosection structure and stability as compared to embryos dissected out [250]. Sections are then processed for digital HCR (details in section 2.12.2), imaged in 3D (z-step of 0.5 µm) before being stripped with DNaseI. Stripping with DNaseI must remove both the fluorochrome conjugated hairpins and the DNA initiator probes bound to mRNA molecules to remove signal and stop amplification in the next round.

To test this, sections were stained with 3 genes, stripped and then re-stained following the same HCR protocol but without the addition of new DNA probes. Imaging the section again would detect any signal remaining from round 1 and/or new signal initiated from remaining DNA probes. In round 1, diffraction limited spots, representing single mRNA molecules, are clearly above background and are easily distinguished from sources of autofluorescence such as red blood cells (Fig.4.5a). Conversely, after stripping and re-amplification, there is no detectable signal above background (Fig.4.5b) (note: maximum display value lowered in 4.5b to show all sources of signal). Also, DAPI staining is no longer limited to cell nuclei but is more widespread due to DNA digestion. Highly compacted DNA (bright DAPI foci) remain after DNaseI treatment This shows that both DNA hairpins and probes have been sufficiently removed by DNaseI treatment.

After running all the rounds of HCR on embryonic sections, signal is stripped for the final time before staining cell membranes. Given the processing required for *in situ* hybridisation, cell membrane dyes do not stain the membrane. Therefore, antibody staining against membrane-bound proteins are employed to label cell surface membranes. This allows cell borders to be defined for cell volume segmentation. This is one of the

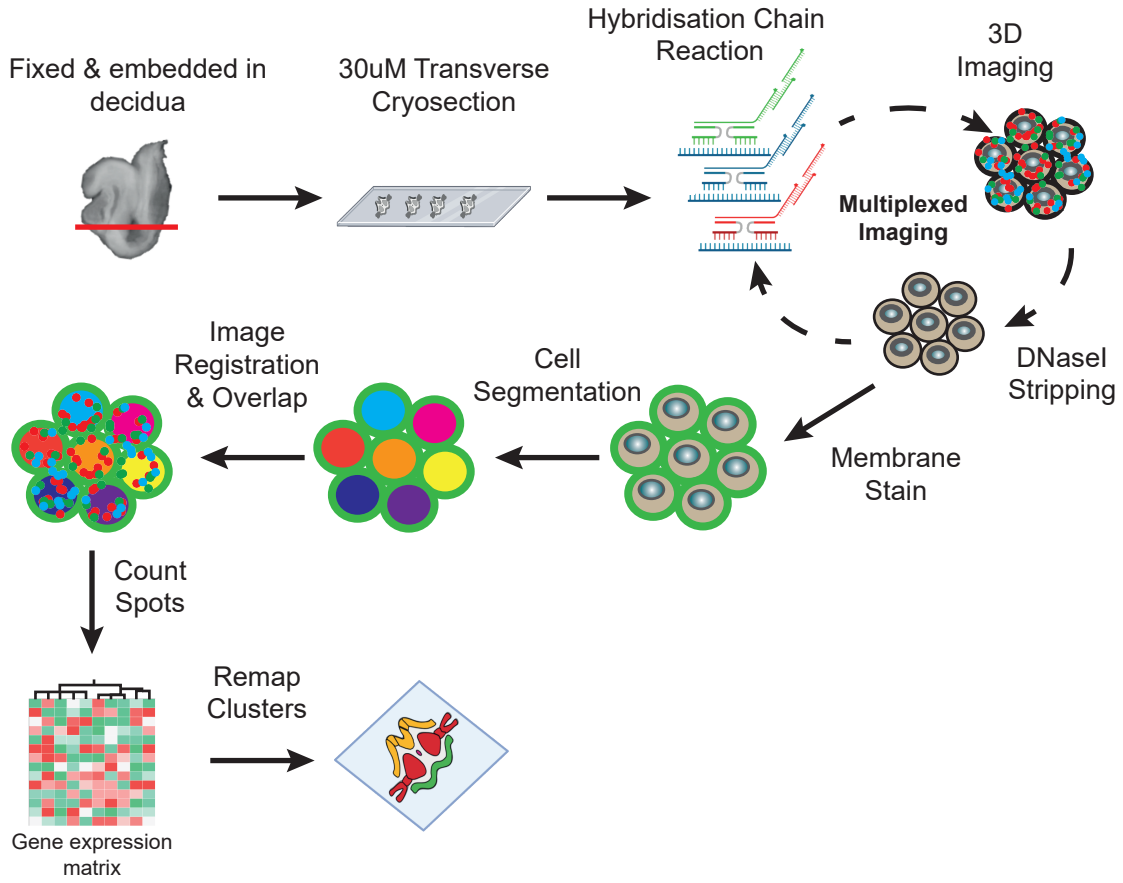


Figure 4.4: Overview of the Spatial Genomics Analysis workflow.
Schematic showing the laboratory steps of SGA from collecting embryos for embedding and sectioning to performing version 3 of hybridisation chain reaction. Hybridisation chain reaction can be multiplexed by removing signal with DNaseI and restaining with different probes up to nine times. After the final DNaseI strip, cell borders can be identified through the immunostaining and imaging of membrane bound proteins. Cell segmentation uses these data to define cell volumes that are registered and overlapped with smFISH data. By counting the number of spots within a defined cell volume as a measure of gene expression, a gene count matrix is created with genes as rows and single cells as columns. These data can be used for hierarchical clustering. Clusters are annotated based upon their expression of multiple marker genes and an subsequently be spatially remapped back onto the clusters to define cell location.

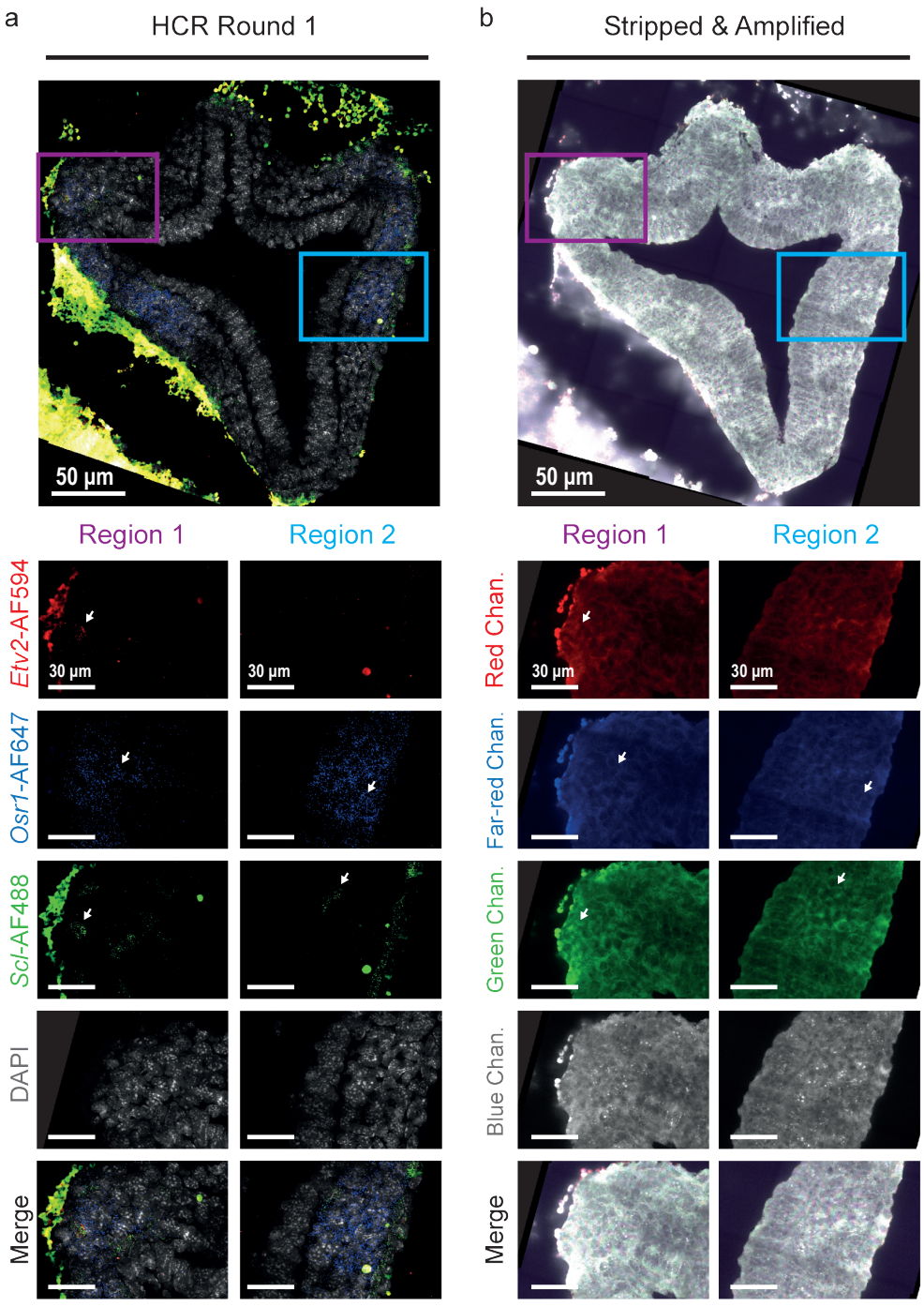


Figure 4.5: DNaseI allows efficient removal of fluorescent signal and DNA probes. In **a**, HCR staining on unstained transverse section of an E8.0 mouse embryo for three genes is shown. Two regions are highlighted in the magenta and cyan boxes with magnified versions of each channel and a merge shown below. Sections were treated with DNaseI and washed before running the hybridisation protocol with no new probes and amplified with B1-AF488, B4-AF594, B5-AF647 hairpins to test for residual DNA probes. Imaging was performed with the same settings as for HCR round 1. **b** shows the same section as in **a** after stripping and amplification. The maximum display value has been lowered ensure no dim spots were present in the sections. All channels in the entire section contain no smFISH signal. White arrows indicate sources of signal before stripping.

most challenging aspects for smFISH-based spatial expression methodologies, particularly when cell morphology is mesenchymal as with mesoderm. In the original SGA method, antibodies raised against E-cadherin and β-catenin are used to demarcate cell borders in the dorsal neural tube [258, 310]. To improve cell border discrimination in mesodermal tissue, I also included N-cadherin.

Combining the three antibodies in the same channel allowed reasonably even membrane staining with cell border discernible in all three germ layers (Fig.4.6a). As expected, mesodermal borders were less sharp and discontinuous as compared to ectoderm and endodermal cells but the large majority of cell borders could be determined by the experimenter. Cell segmentation algorithms using cell surface labelling benefit from high contrast staining. Additionally, staining proteins in the cell membrane creates punctate staining along the length of cell membranes with signal detectable as small dots along the membrane rather than continuous staining. While this can be limited by image processing, SGA employs machine learning to increase contrast, limit discontinuous staining and reduce the impact of autofluorescence.

Three image classifiers are trained by the experimenter (cell membrane, interior and autofluorescence) and used to create prediction matrices with each pixel assigned a value between 0 and 1 based on the confidence the model has assigning a pixel to a particular class. Membrane probability matrices can be displayed with 0 to 1 represented from black to white respectively (Fig.4.6b). This massively increases the contrast between cell border and interior, improves the continuity of signal along cell borders, reduces sources of autofluorescence and reduces the impact of shading at the border between tiles. The prediction map is used instead of the raw membrane staining to perform cell volume segmentation.

In the first version of SGA, the membrane probability matrices were imported into MatLab and input into the Watershed algorithm in the Image Analysis package. This was tested in our system but led to unsatisfactory results (Fig.4.6c). While most cells

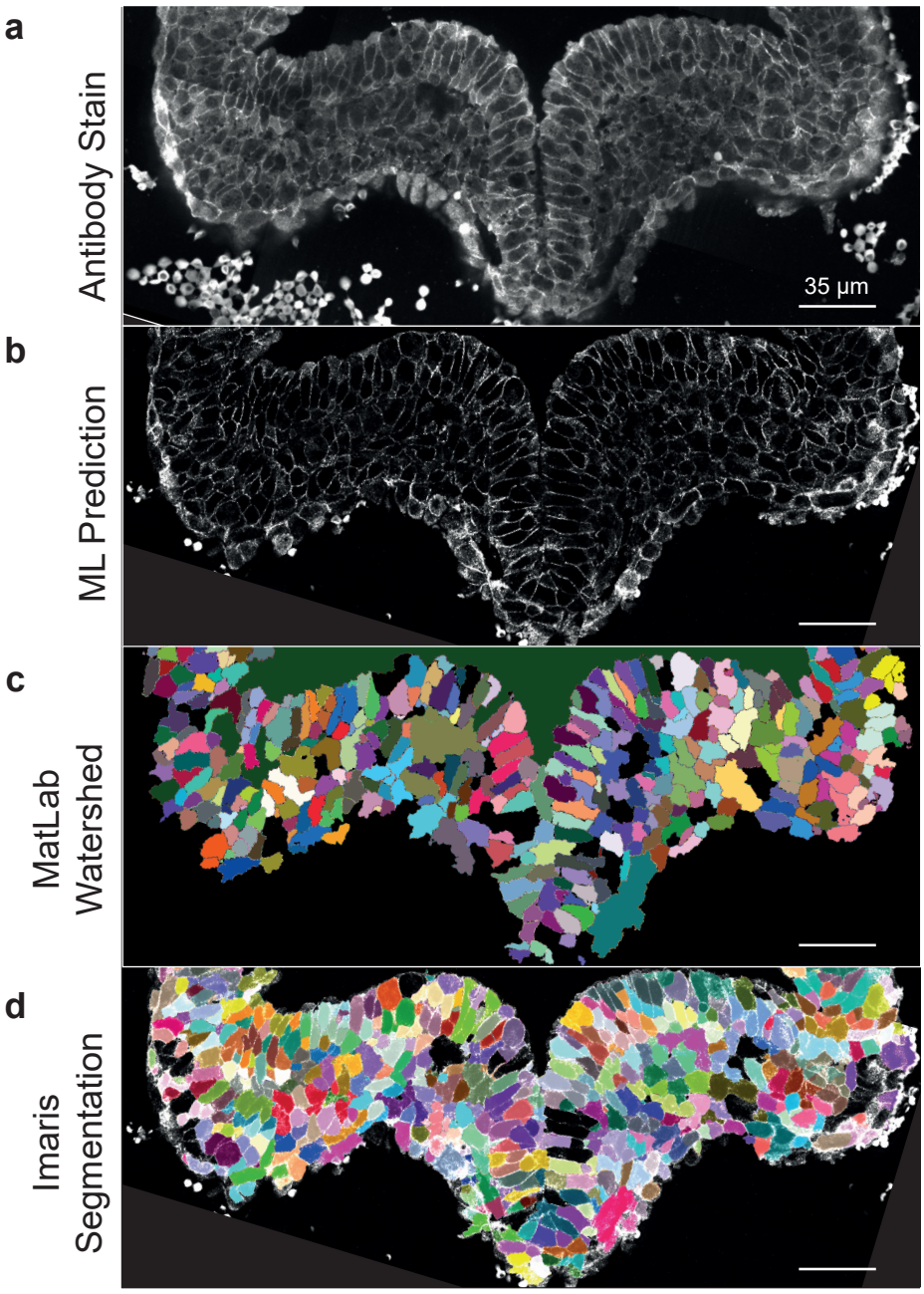


Figure 4.6: Cell volume segmentation on immunostained cell surface proteins
Shown is four stages of cell segmentation on a transverse section of an E8.0 mouse embryo. **a** Antibodies targeting the cell surface proteins β-catenin, E-cadherin and N-cadherin were stained in the same channel to label cell borders. **b** Ilastik pixel classification was used to train two machine learning (ML) classifiers for cell membrane and cell interior. A prediction map for the membrane staining in **a** is shown with values scaled from 0 to 1 represented as black to white respectively. **c** Ilastik membrane predictions from **b** were thresholded (at 0.9) and run through MatLab watershed algorithm to segment cell volumes. A random look-up-table (LUT) has been used to assign a random colour to each predicted cell volume. **d** Using non-thresholded Ilastik prediction data, a cell volumes were determined by Imaris' 'Cells' function. Shown is the cell mask randomly coloured before any manual intervention.

of the neural tube with their clean, sharp border staining segmented well, cells of the surface ectoderm and particularly the mesoderm were improperly segmented. In regions where border prediction confidence was a little lower (0.6 - 0.8), cells were often under-segmented with two adjacent cells being assigned to the same mask. This is a problem for SGA as it would count transcripts in two separate cells as one. Therefore, alternative methods of segmentation were sought.

Using the commercial Imaris software to perform segmentation on membrane prediction matrices proved to have many benefits. Firstly, the segmentation with no manual intervention was much more accurate than MatLab Watershed (Fig.4.6d). Even in the harder-to-segment mesoderm, the majority of cell masks would match user performed segmentation. There are regions of over-segmentation where one cell has been split into two masks but this can be rectified by the second benefit of Imaris: it's user interaction allows manual adjustment of segmentation. Two cell masks can be merged manually and thus segmentation vastly improved by the user. Therefore, analysis deviated from the original SGA publication to improve segmentation in this new tissue context.

With multi-round HCR staining, membrane staining and segmentation established, it was critical to assess the accuracy of overlapping different round of imaging (image registration). Initially, the aim was to perform tiled imaging of whole embryonic sections for analysis to allow sufficient resolution and sensitivity to be achieved with a high magnification and numerical aperture objective while retaining large field of analysis (Fig.4.7a). This required a few additional steps not used in the original SGA work as their field of view was large enough without tiling.

When overlaying stitched images, the image registration algorithm (cross-correlation) could not create sufficient overlap. Even with manual registration, two rounds could never be accurately overlaid (Fig.4.7b). Confusingly, different areas of the sections indicated different translations were necessary. For example in the region magnified in magenta in

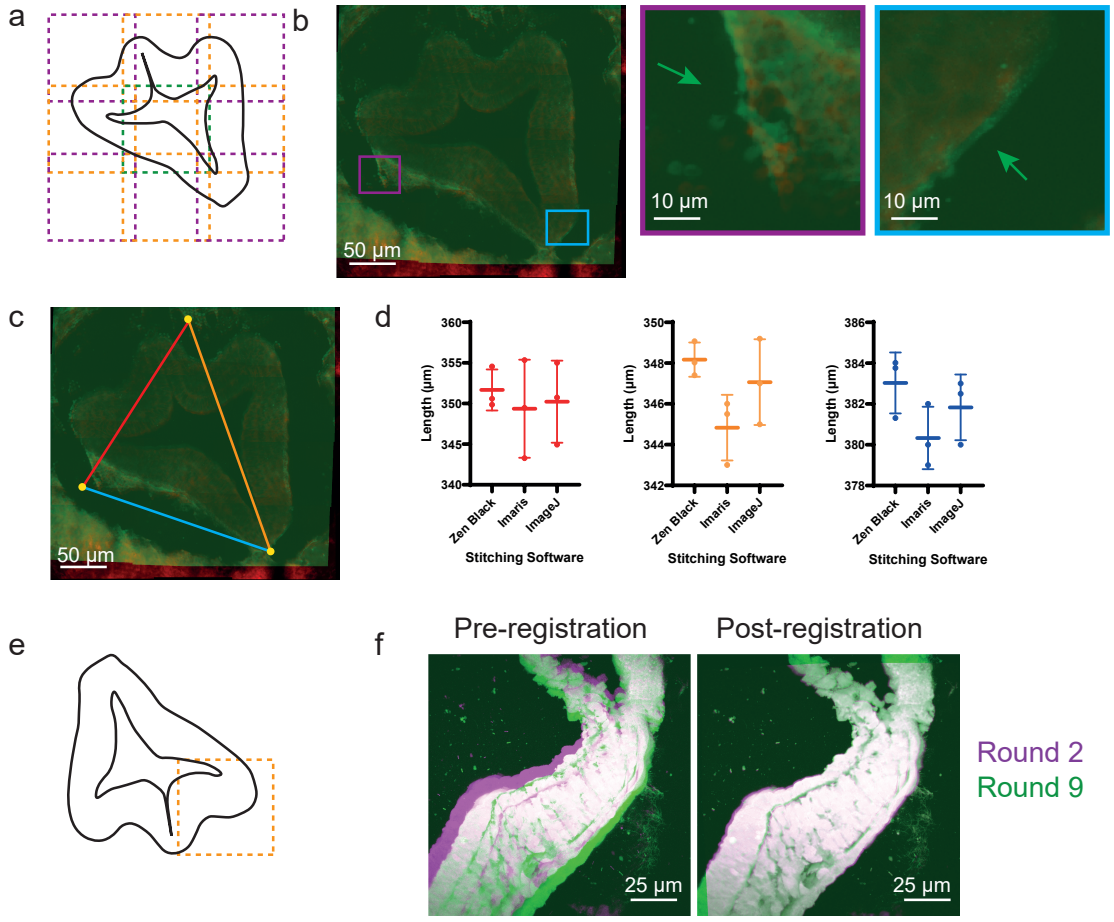


Figure 4.7: Registration of imaging rounds is inaccurate with stitched images but accurate with single fields of view.
a Schematic representing the overlapping tiles of multiple fields of view required to image the entire embryonic transverse section. Each image represents a maximum intensity projection of tissue auto fluorescence. A single field of view is represented as a dashed coloured square. Each tile has an overlap of approximately 10%. **b** Two rounds of imaging of the same embryonic sectioned overlayed with round 2 in green and round 4 in red. Two regions are magnified in the magenta and cyan boxes highlighting the inconsistencies in shift required. Green arrows indicate the direction of shift required by green image relative to red. Tiled images cannot be accurately registered because of inconsistencies in stitching algorithms. Three lengths (shown in **c**) were measured in stitched images in three serial rounds of imaging. The distance of each measurement was plotted in **d** with the colour of the plot representing the corresponding line in **c**. Three stitching software (Zen Black, Imaris and ImageJ) were compared for each measurement. **e** Schematic of transverse embryonic section showing the region visible in one field of view in **f**. **f** Two rounds of imaging overlayed with round 2 in magenta and round 9 in green. Each image represents a maximum intensity projection of tissue auto fluorescence. Before image registrations image rotation and transformation is required. After registration, green and magenta are well overlayed as represented the combined white colour.

figure 4.7b, the green channel (from imaging round 2) should be translated down and to the right (as indicated by the green arrow). Conversely, the region magnified in cyan, requires the green round to be shifted up and to the left. This indicated that the images of the same section were different between rounds. To test this, the length between three fixed points was taken to assess change (Fig.4.7c). Relatively large differences in length were observed between rounds for all the distances measured (Fig.4.7d). The same approach on low-magnification images sections (no tiling and stitching) demonstrated this was not due to a change in the section's physical dimensions between rounds. When stitching tiled images, there is a small error in stitching accuracy with every tile added [311]. While small, this error is additive and so large tiled images accumulate larger inaccuracy. Three stitching software were tested for their consistency between rounds but even the best (Zeiss Zen Black) had a maximum range of 2.5μm across three rounds.

This inconsistency in stitching between rounds makes accurate image registration impossible. Other FISH-based multiplexed imaging techniques that perform tiled image analysis rely on accurate overlay of individual fields of view (FOVs) using automated microscopes [39, 312]. Aiming to use existing confocal microscopes, I instead reverted to focusing on single FOVs (Fig.4.8e). As expected, after manually calculating rotation angle, the image registration algorithm could accurately translate images to overlay with a reference (Fig.4.7f). This enables the overlaying of multiple rounds of dHCR imaging with cell segmentation, critical for the quantification of smFISH data (Fig.4.8)

SGA analysis takes raw spot data and first improves contrast and accuracy of spot detection by using Ilastik machine learning pixel classification (Fig.4.8a). Two classifiers are trained for each fluorescent channel by manually training to recognise spots and not-spots. Application of these classifiers on each gene imaged yields a 3D probability matrix with 0 representing confidence in no spots and 1 representing certainty. Values are thresholded at a high probability (typically 0.7 - 0.95), setting low confidence pixels to 0 and high confidence to 1. Touching voxels (3D pixels) are grouped and assigned a unique identification number (Fig.4.8b). Overlaying this spot I.D. mask with raw imaging

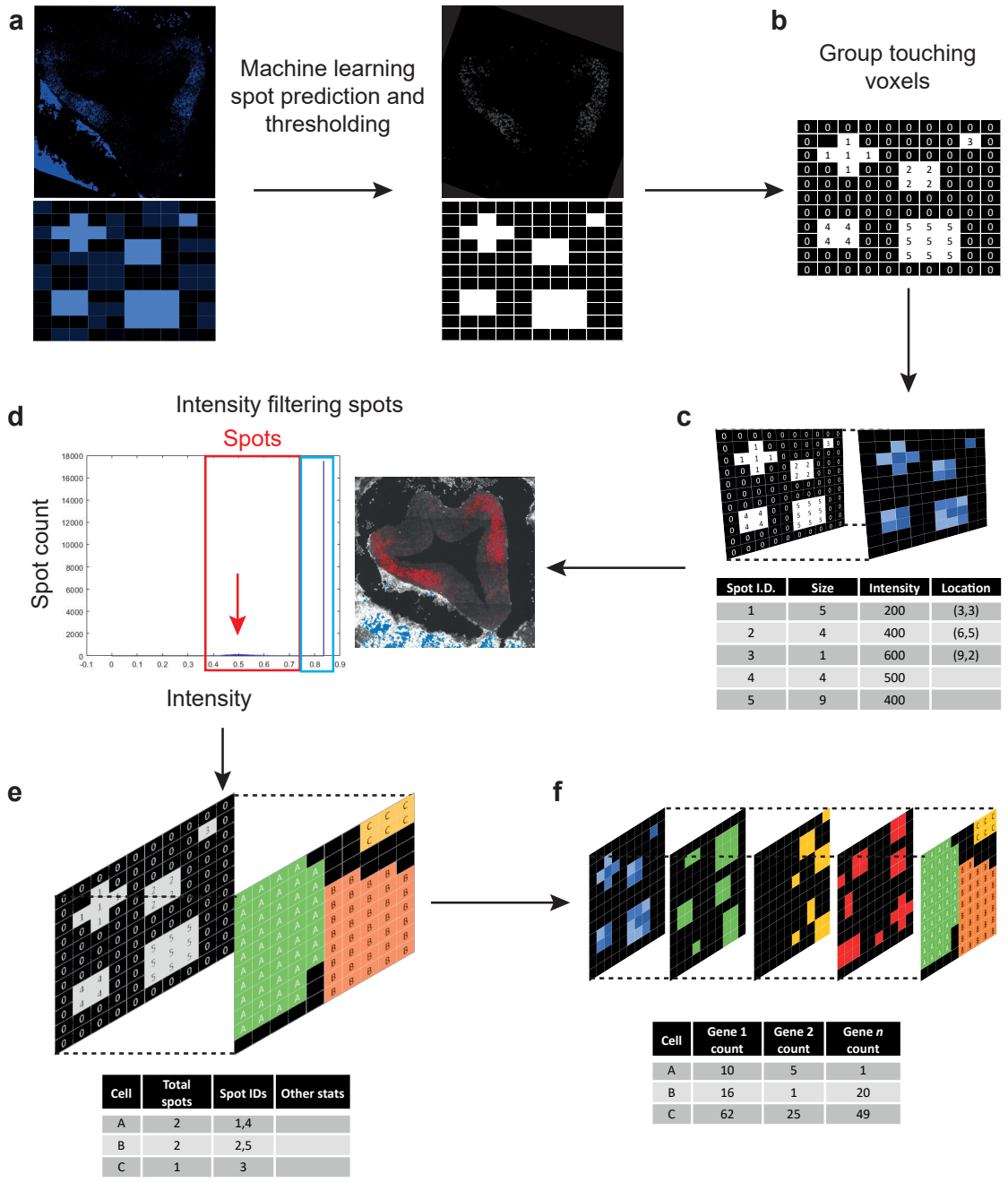


Figure 4.8: Schematic of smFISH quantification pipeline for SGA.
Multiplexed smFISH data generated using SGA (Fig.4.4 was quantified using custom MatLab software heavily adapted from [258]. Ilastik was used to predict the likelihood each voxel (3D pixel) formed part of a spot and low confidence voxels were removed by thresholding (all voxel $>$ threshold set to 1, and those below set to 0) (**a**). Touching voxels were grouped to create a spot ID (**b**). These spot IDs were overlapped with raw fluorescence data to retrieve data such as mean intensity (**c**). Such data was used to plot and filter spots based on intensity and size (**d**). Filtered spots were counted by overlapping with cell segmentation data (**e**) for each gene measured to create a gene count matrix (**f**). These data were used for hierarchical clustering and cell type annotation.

data enables completion of a spot statistics table consisting of I.D., size, intensity and location for downstream filtering of spots (Fig.4.8c). Primarily, spots are filtered on median intensity to exclude any false positive spots originating from autofluorescence (typically far brighter). Once satisfied with the spot identification and statistics, these data were integrated with the cell segmentation mask previously calculated. This allowed the counting of spots within cells as a measure of gene expression (Fig.4.8e) This process was repeated for all genes measured and taken forward for downstream, gene expression analysis (Fig.4.8f).

4.2.3 Selection of three embryos for spatial genomics analysis

I selected three embryos for SGA, all of which were in the window of blood fate specification [218] at approximately E8.0. The three embryos were selected as they represented slightly different stages of DA formation and could be ordered from least-to-most mature (E1 - E3). The three embryos could be ordered according to section morphology (Appendix Fig.5) and expression of *Pax8* in the intermediate mesoderm (Fig.4.11 and Appendix Figs.8, 9 & 10). Embryo 1 is the youngest with the lowest expression of the mature IM marker *Pax8* in the intermediate mesoderm, least defined DA and wider neural groove. In contrast, Embryo 3 is the most developed with high *Pax8* expression in the intermediate mesoderm, a circular group of HEPs forming the DA and narrower neural groove. Finally, embryo 2 lies in the middle on all these aspects and so was labelled as representing a stage in between.

Our primary interest is in Wave 3 blood fated HEPs known to reside within the E8.0 splanchnopleura [173]. Therefore, I focused on the caudal half of the embryo and at a level based on anatomical features on the anterior side of the section (just below the heart) (Fig.4.9 and Appendix Fig. 5). In the first round, I probed for genes that allowed the identification of haemao-endothelial progenitors (HEPs) so I could select sections that were representative of that region of the embryo and compare (Fig.4.9). Reassuringly, I saw the same distribution of HEPs across the caudal embryo as I saw

by volumetric imaging (Fig.4.3). *Scl* positive cells were detectable on the ventral and dorsal side of the caudal mesoderm.

By comparing the morphology on our sections to the 3D reconstruction on the eMouseAtlas community resource [313, 314], I could select sections in the relevant regions for further investigation. I aimed to select a section lying midway up the posterior half with the presence of gut endoderm on the anterior side not yet meeting to form the gut lumen and without any cardiac or somitic tissue visible. For example section i in figure 4.9 is too distal with visible somites on the posterior side (white arrow) and no protrusions in the def. endoderm on the anterior side. Meanwhile section vi of figure 4.9 lies proximal to the selected region with a formed gut lumen (labelled with G) on the anterior side. This left the selection between ii, iii, iv and v. Selection of the region to focus on was based on both the presence of Scl^+ cells in the section and practical considerations of being able to image the whole posterior embryonic mesoderm from neuroepithelium to the lateral edge in single field of view. The left side of ii, iv and v and the right of iii were all contained numerous Scl^+ cells but ii and iii did not contain Scl^+ cells on the dorsal side (Fig.4.9). I wanted to profile cells on both the dorsal and ventral side of the mesoderm and so excluded ii and iii. Finally, iv could be easily imaged in a single field of view with appropriate objectives and zoom settings and so was selected as the focus of embryo 3. This process of informed selection of regions of interest was also performed for embryo 1 and 2 to provide three independent regions of interest for SGA.

4.2.4 Qualitative assessment of haemato-endothelial progenitors and neighbouring mesodermal cells by multiplexed HCR

Looking more closely at HEPs in the selected region, they are always on the edge of the mesodermal wings either directly adjacent to the def. endoderm or surface ectoderm (Fig.4.10). HEPs can be split into those on the ventral side of the mesodermal wing and those on the dorsal side. The ventral HEPs typically formed a single non-continuous layer of cells apart from in the DA of embryo 3 where the forming chord can be identified

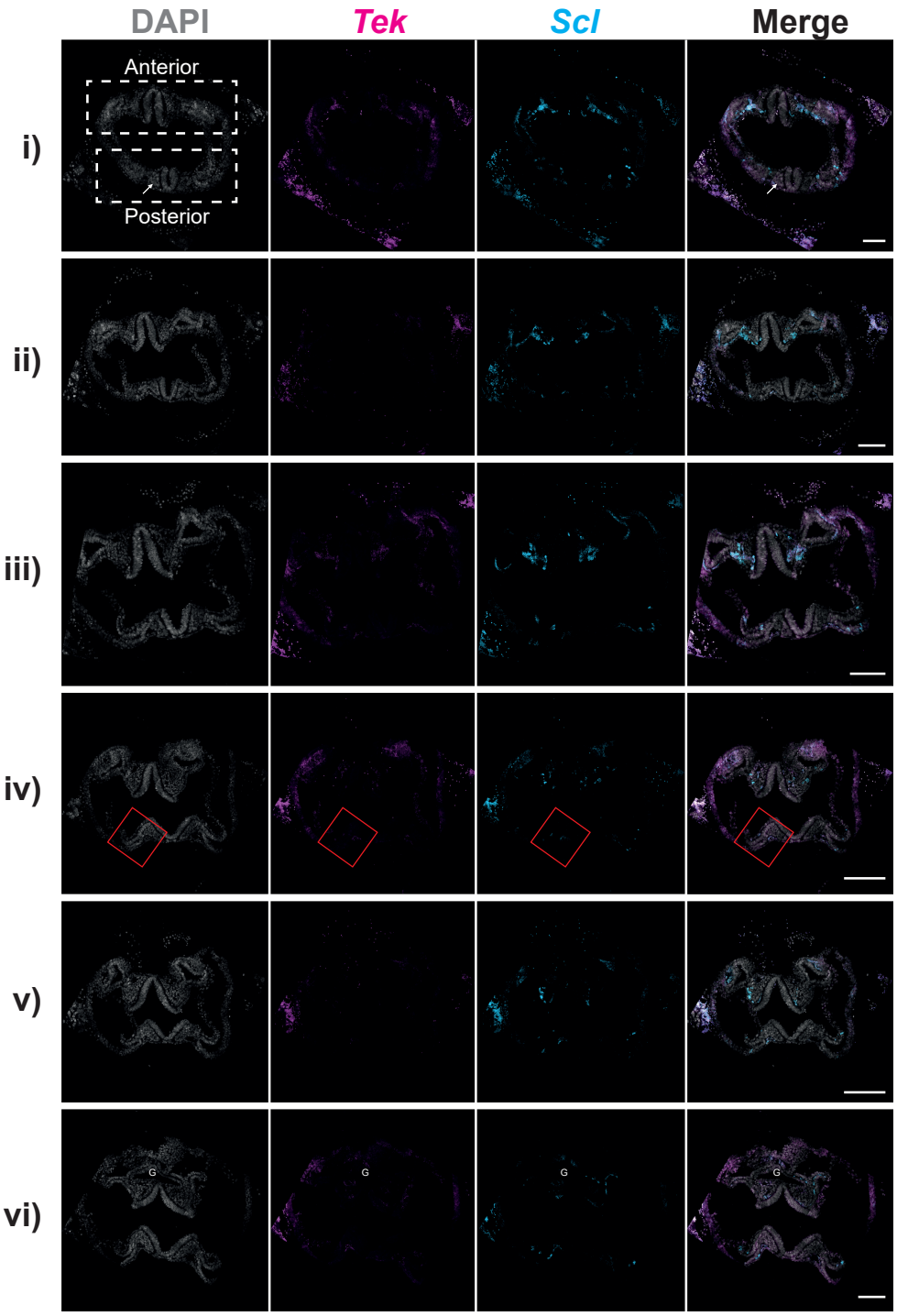


Figure 4.9: Imaging of haemato-endothelial gene expression in transverse sections of E8.0 embryos for selection of a representative region for quantification
Two genes from the first round of dHCR (*Scl* and *Tek*) were imaged on all sections collected from each embryo a spinning disc confocal for speed. Shown are three sections from embryo 3 to highlight the selection process. Sections (i, ii, iii, iv, v, vi) are 60 µm apart. Anterior and posterior side of the embryo are indicated in i alone. Boxed in red was the region in embryo 3 taken forward for more thorough analysis and quantification. White arrow, somite; G, gut lumen.

as a circular group of HEPs. The ventral HEPs lateral to the forming DA (green arrow head in Fig.4.10) had a recognisable morphology typically elongated along the medial-lateral axis and slender in the dorsal-ventral axis. This was reflected in their nucleus morphology which was elongated in the same orientation. When compared to their surrounding cells these ventral lateral HEPs were more aligned with their ventral neighbours the def. endoderm than the dorsal mesodermal cells. Additionally, these cells are numerous in embryo 1 and 2 but seemingly absent in the section of embryo 3. The simplest explanation for this is that the ventral lateral HEPs are moving medially to form the DA and is supported by recent live imaging of an Scl reporter line showing Scl-positive cells migrating down from the YS through the lateral edges of the embryo and towards the midline to form the DA [142]. In contrast to the ventral HEPs, the dorsal HEPs displayed a more spherical morphology with a rounded nuclei (magenta arrows in Fig.4.10). In addition to there being fewer cells on the dorsal side, they were present in all three embryos and did not seem to expand during this short window of development.

Qualitative assessment of expression levels of these markers suggests that in embryo 1 and 2 the majority of HEPs are expressing similar levels of these three markers. In embryo 3, many of the angioblasts are *Etv2* negative, suggesting they had already downregulated it. Meanwhile, all cells expressing *Tek* were also *Scl* positive. In all the sections and embryos analysed in this study, no cells positive for *Scl* and/or *Tek* were ever observed in the centre of the mesodermal wing. However, I occasionally observed *Etv2* single positive cells that displayed a cell and nuclear morphology more similar to that of non-HEP mesodermal cells than double or triple positive HEPs (white arrow head in embryo 2 of Fig.4.10). These were the only cells expressing HEP markers that were not directly adjacent to surface ectoderm or def. endoderm but more integrated within the mesodermal wing. A cell expressing *Etv2* but not *Scl* or *Tek* is likely in the very early stages of haemato-endothelial specification. Qualitative assessment of the frequency of these cells would indicate they are rare, with only 1-2 cells per embryo observed.

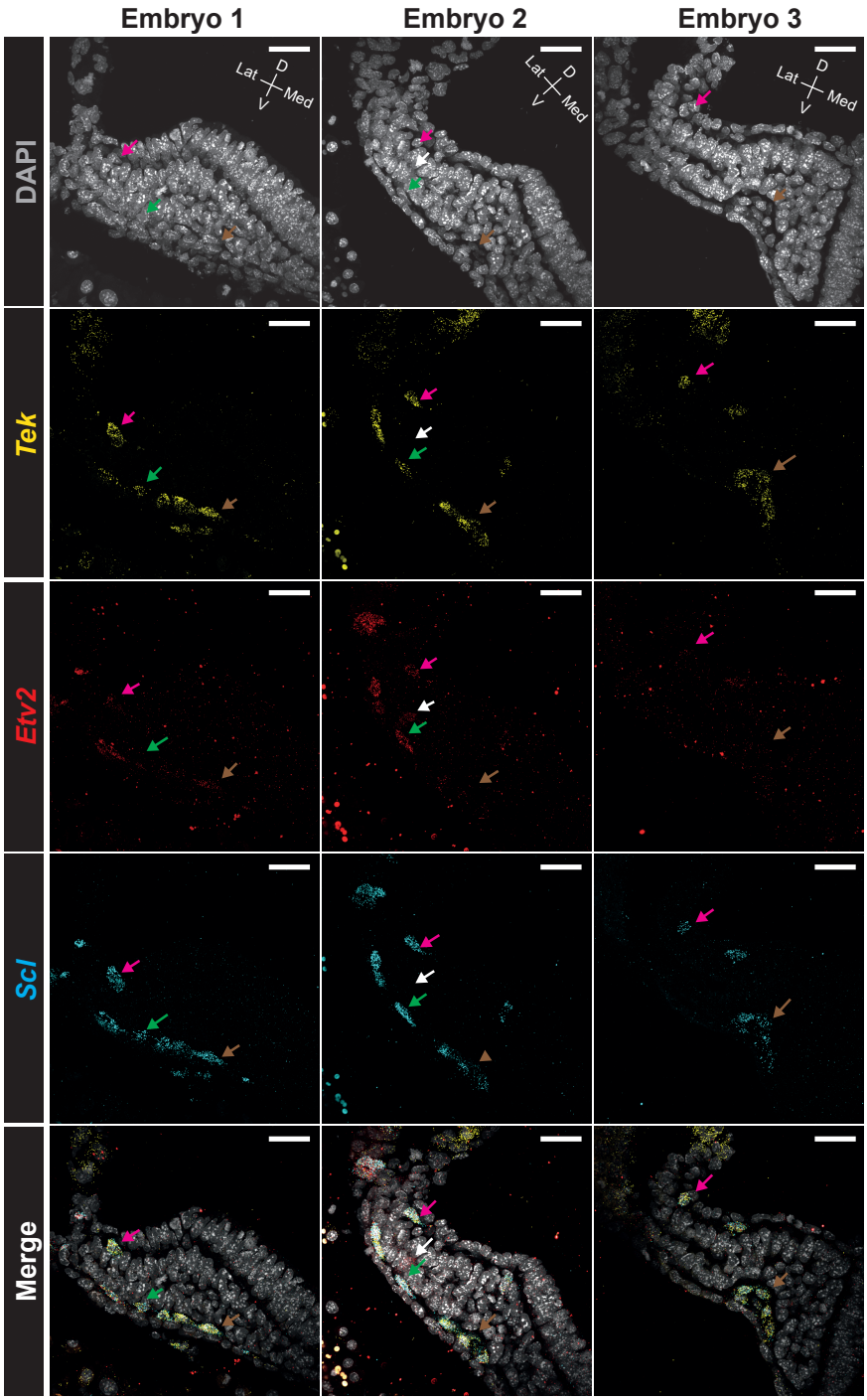


Figure 4.10: Haematoendothelial progenitors have distinct locations and varying expression of *Scl*, *Etv2* and *Tek*

Three sections from the three embryos (embryo 1, 2 and 3) selected for SGA analysis stained for key HEP markers. Images show maximum intensity projections from 30µm sections stained for DAPI and dHCR for *Scl*, *Etv2* and *Tek*. These three markers allow the identification of all HEPs present in the caudal embryo at this level. The future DA (brown arrow head), ventral lateral HEPs (green arrow head), dorsal lateral (magenta arrow head) and *Etv2* single positive HEP (white arrow head, enlarged in Appendix Fig.7) are identified when present. Compass shows orientation with D, dorsal; V, ventral; Lat, lateral; Med, medial. Scale bars represent 30µm.

In order to probe the expression of more HEP/endothelial genes and markers of other cell types, I stripped these probes (Fig.4.10) and ran our panel of genes. Before quantifying the expression of these genes with SGA as described (Fig. 4.8), I qualitatively assessed each one. Each gene for the panel of genes can be seen individually for embryo 1, 2 and 3 in Appendix Figures 8, 9 and 10 respectively. Here I focus on some genes of interest. Focusing on *Foxc2* (presomitic), *Hand2* (LPM), *Osr1* and *Pax8* (IM) (Fig.4.11) shows these markers identify the tissues of interest but display overlapping domains of expression. Comparing the three embryos I see the expression of these markers overlaps most in the youngest embryo (embryo 1) but has more restricted domains of expression in the oldest (embryo 3). This overlap is particularly discernible for *Osr1* and *Hand2* expression, with cells at the border between the lateral and intermediate mesoderm expressing both markers suggesting these cells are actively undergoing cell state decision between IM and LPM with Hand2 negatively regulating Osr1 in the LPM [315].

Both Lhx1 and Osr1 have been previously described as being expressed in both the LPM and IM and then becoming restricted to the intermediate mesoderm as development proceeds [69, 71, 75]. It is not until the expression of Pax2 and Pax8 that there are any truly specific markers of the intermediate mesoderm [68, 316]. Embryo 1 shows no staining of *Pax8* in the intermediate mesoderm but embryo 2 has low expression in 1-2 cells on the ventral side of *Osr1*$^+$ mesoderm (Fig.4.11). Embryo 3 has 5-10 cells with higher expression than those in embryo 2. The *Pax8* and *Osr1* expression domains highly correlate and are sharply distinct from that of pre-somitic mesoderm expression of *Foxc2*. This correlates with the observation that Foxc1 and Foxc2 regulate paraxial versus intermediate mesoderm fate [65]. Cells at the border of intermediate and paraxial mesoderm in embryo 1 and 2 on the other hand express both *Osr1* and *Foxc2* suggesting this decision has not fully resolved. This shows that these three embryos capture the window of development in which mesodermal identities become spatially distinct.

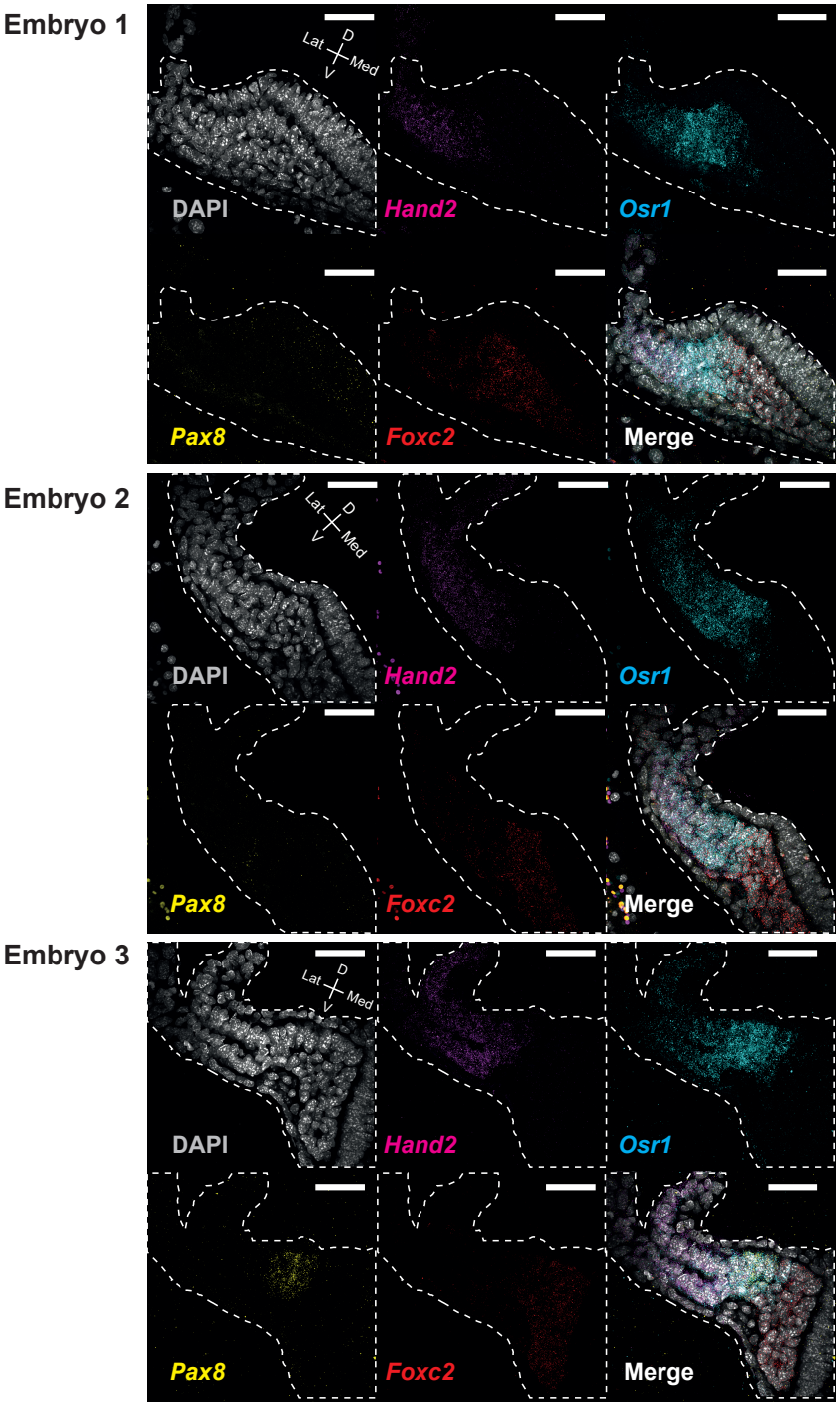


Figure 4.11: smFISH of LPM, IM and pre-somitic mesoderm markers identifies overlapping domains of expression
Maximum intensity projections of 30µm transverse sections through three E8.0 embryos showing the left side of the posterior embryo. Staining from from three rounds of imaging has been registered with DAPI from round 1, Foxc2 from round 3, Osr1 and Hand2 from round 4. Compass shows orientation with D, dorsal; V, ventral; Lat, lateral; Med, medial. Scale bars represent 30µm.

4.2.5 Assessment of RNA degradation during multiplexed dHCR

I ran the full panel of our selection of genes on these three embryos across seven rounds of staining and stripping. After collecting these data but before running antibody-based membrane staining I stained the same genes as in round 1 (*Tek*, *Lhx1*, *Etv2* and *Scl*) to determine the extent of RNA degradation (Fig.4.12 and Fig.4.13). Broadly, all cells positive for each of these genes can be identified in the final round of staining too (Fig.4.12). This indicates in cells expressing a gene at a high level that any loss in single molecule staining will only reduce the expression level rather than convert a positive cell into a negative cell. Additionally, staining between round 1 and round 8 was dimmer in every channel. This cross-channel observation suggests this is unlikely due to a change in microscope function (e.g. laser power) but rather a reduction in the number of fluorochromes per mRNA molecule. This could be caused by a slight difference in the concentration of probes or hairpins but more likely reflects a reduction in the number of initiator probes binding to single molecules. If an mRNA molecule is partially degraded, fewer split-initiator probes will bind. This makes genes where relatively few split-initiator probes can be designed (e.g. *Etv2* short length limits probe library to just 14 pairs) particularly susceptible to loss of signal. This is reflected in the relative decrease in *Etv2* staining in round 1 versus round 8 as compared to a gene like *Tek* with 40 probe pairs (Fig.4.12 and Fig.4.13).

Enlarging regions containing single-molecule staining and qualitative assessment of the number of molecules stained per cell shows that the majority of molecules are stained in the both round 1 and round 8 (Fig.4.13), validating this stripping and restaining method for multiplexing smFISH in murine tissue. As noted in the overview figures, although most molecules remain, their average brightness is reduced. This is particularly noticeable again in *Etv2* with many spots becoming dimmer. This was accounted for in quantification through manually setting the fluorescence intensity filter manually for each gene and round. Finally, in embryos 1 and 3, *Scl* staining has been severely impeded by the incomplete removal of *Xist* staining from the previous round. *Xist* was stained to test a hypothesis not discussed in this thesis. Our hypothesis is that the

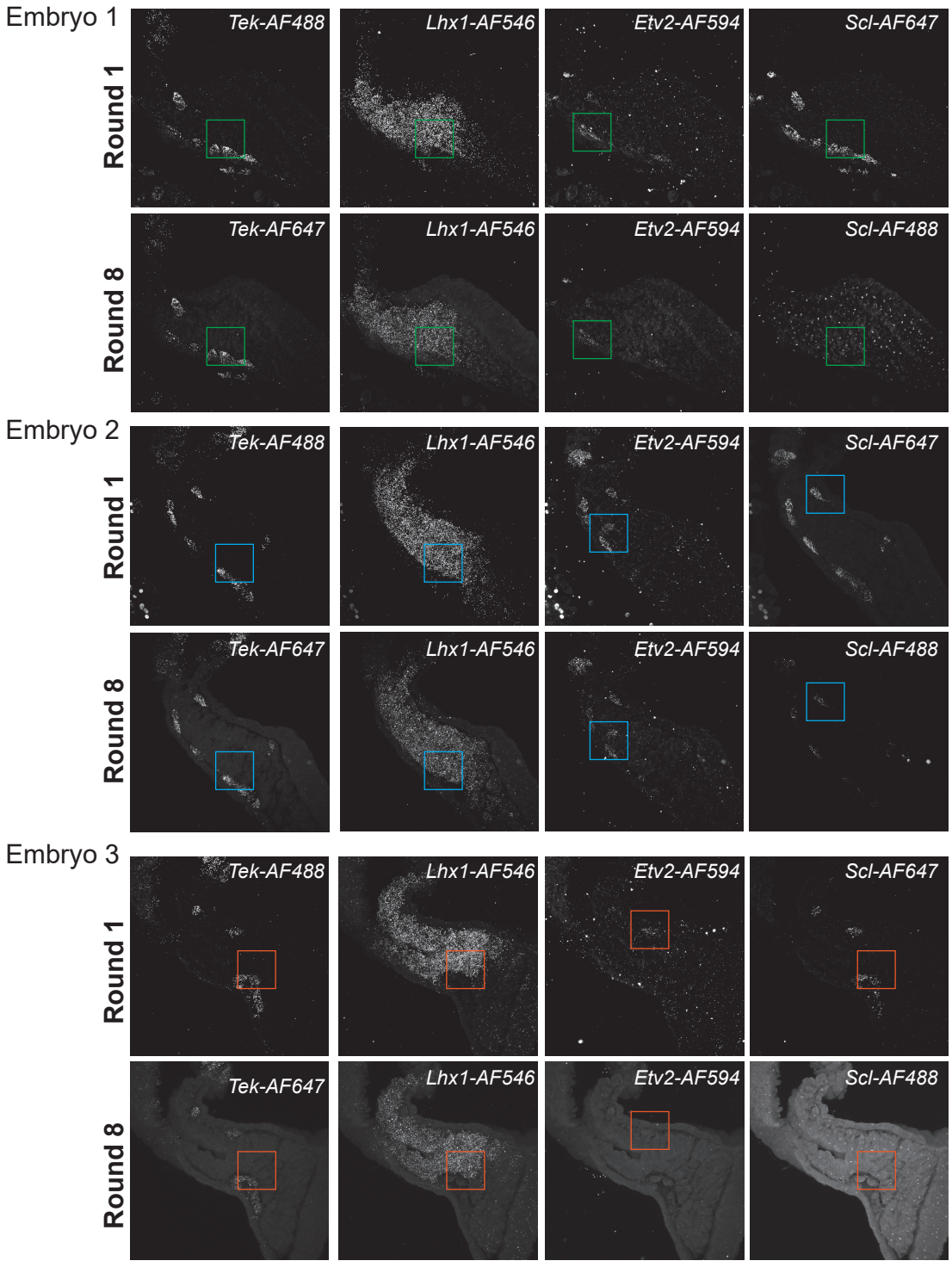


Figure 4.12: Domains of *Tek*, *Lhx1*, *Etv2* and *Scl* staining in the first and last rounds of dHCR staining are the same

The same genes (*Tek*, *Lhx1*, *Etv2* and *Scl*) were stained in the first (round 1) and last round (round 8) of dHCR with maximum intensity projections for each gene in both rounds for each embryo shown. *Scl* staining has been severely affected by remaining *Xist* probes in embryos 1 and 3.*Lhx1* and *Etv2* were stained in the same fluorescent channels (AF546 and Af594 respectively) while *Tek* and *Scl* were switched between round 1 and round 8. Regions containing the future DA and $Lhx1^{+}$ mesoderm are boxed in this figure and enlarged in Fig.4.13.

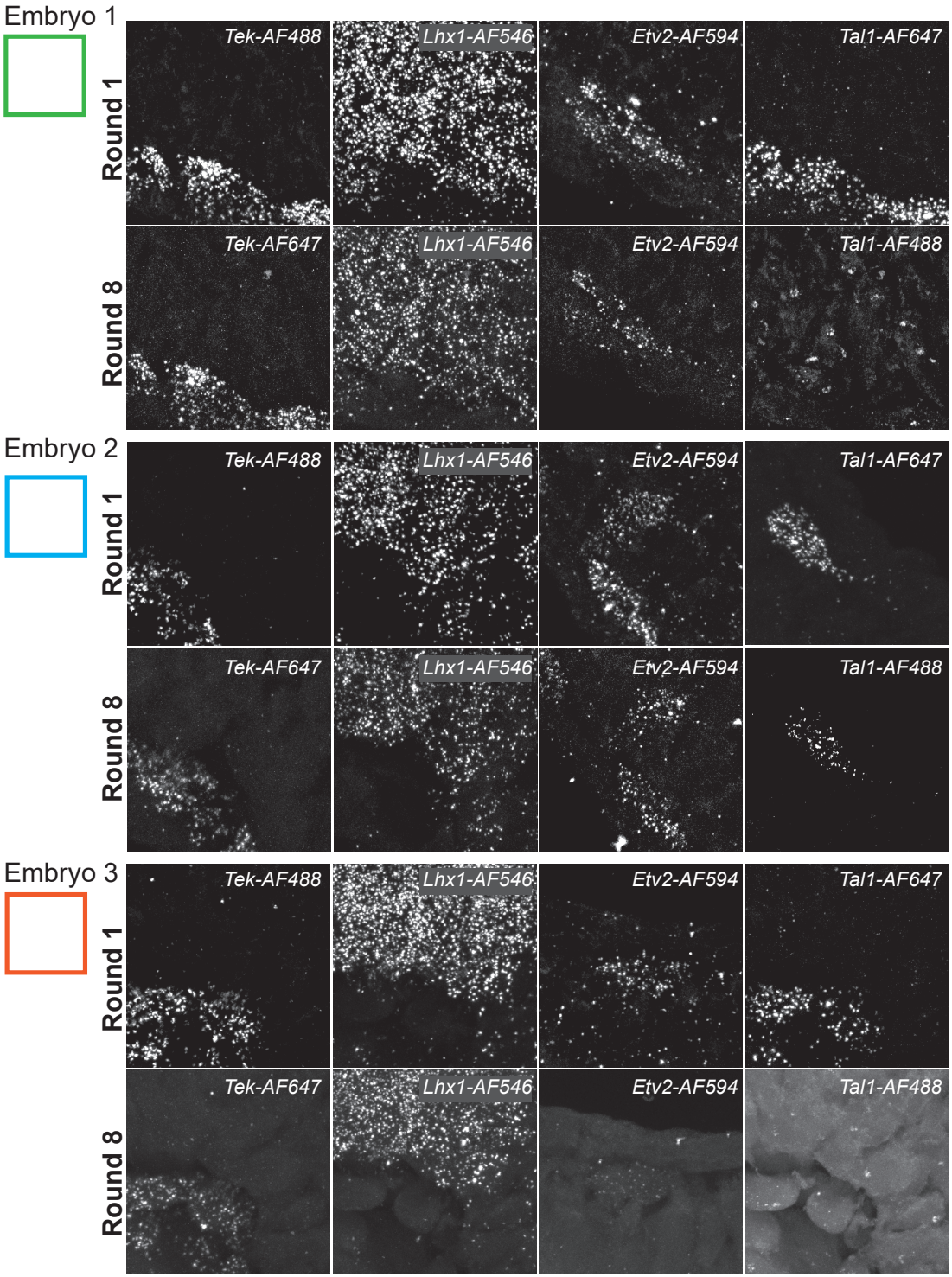


Figure 4.13: High magnification of *Tek*, *Lhx1*, *Etv2* and *Scl* staining in the first and last rounds of dHCR show a small loss in RNA molecule staining
Enlarged regions from Fig.4.12 showing the first (round 1) and last round (round 8) of dHCR staining for *Tek*, *Lhx1*, *Etv2* and *Scl*. *Scl* staining has been severely affected by remaining *Xist* probes in embryos 1 and 3.*Lhx1* and *Etv2* were stained in the same fluorescent channels (AF546 and Af594 respectively) while *Tek* and *Scl* were switched between round 1 and round 8.

high density of *Xist* transcripts to the X-chromosome causes a high density of probes and fluorochromes that cannot be fully digested by DNaseI using the current protocol. As a result, the re-staining of *Scl* in the subsequent round is hindered by re-initiation on the *Xist* split-initator probes. This is the only instance of carry over of staining from previous rounds and shows current DNaseI stripping is only suitable for mRNA and not lncRNA. Embryo 2 showed no *Xist* staining in round 7 (likely a male embryo) demonstrated *Scl* staining consistent with the re-staining features noted in the other channels showing the lack of *Scl* staining in embryos 1 and 3 is caused by residual *Xist* staining.

Secondly, I tested whether the fluorochrome used impacted the detection of spots by switching the fluorochrome used for *Tek* and *Scl* staining between round 1 and round 8. When compared to *Lhx1* and *Etv2* staining where the fluorochrome was kept consistent, changing channel for *Tek* and *Scl* staining did not seem to impact staining and any changes are resulting from RNA degradation (Fig.4.12 and Fig.4.13). This suggests the fluorochrome selection should not impact gene counts. Finally in all three embryos there was an increase in background fluorescence. Although not a problem as this signal was discernible from the bright, punctate smFISH signal, it did decrease the signal:noise ratio.

4.2.6 Quantitation of multiplexed HCR enables the spatial remapping of clusters

Satisfied our dHCR data were of high quality and RNA degradation was limited, I quantified gene expression by counting transcripts within segmented cell in three dimensions (Fig.4.8). This generated a single cell gene count matrix for each embryo. Cells were clustered into similar groups by hierarchical clustering. I manually annotated these clusters based on their expression or absence of marker genes. Since genes were selected to specifically label cell types, I can easily assign a cell identity to all clusters (Fig.4.14). Transcription factors critical for haemato-endothelial development such as *Scl* and *Etv2* were highly enriched in a small subset of cells in each embryo. These genes were high in all cells of that cluster (red dendrograms) while other marker genes

were expressed only in a subset. These are endothelial genes such as *Cdh5* and *Tek* and suggests this cluster is composed of cells representing a continuum of cellular maturation as cells start by expressing *Etv2* and *Scl* and upregulate *Cdh5* as they mature to endothelium. A small subset of these *Cdh5*$^+$ cells (10%) also expressed the arterial marker *Dll4*. I also included the commonly used venous marker *Nr2f2*. Both *Dll4* and *Nr2f2* were expressed in cells other than HEPs, highlighting the importance of co-staining with other endothelial markers.

Within the HEP clusters of each embryo were cells that co-expressed *Dll4* and *Nr2f2*, single *Dll4*$^+$ but no single *Nr2f2*$^+$ cells (Fig.4.14). This indicates that there are endothelial cells committed to an arterial fate (single Dll4$^+$), cells undertaking the arterio-venous decision (double positive) and immature progenitors not yet expressing marker of either fate (double negative).

In addition to genes labelling the haemato-endothelial lineage, I included genes to label the other mesodermal cell types. Qualitative assessment of these genes found that they had overlapping patterns of expression (Fig.4.14). This is observed in the gene count matrices with *Osr1* and *Lhx1* being broadly expressed across IM and LPM clusters in all three embryos and in a subset of cells in the PSM cluster (Fig.4.14). In contrast, *Hand2* is most enriched in LPM, sparsely expressed in some IM cells but absent in all cells of the PSM. The reverse pattern is observed for PSM markers *Tbx6* and *Foxc2* with highest expression in PSM, sparse in IM and rare in LPM (Fig.4.14). Finally, I could identify the spatially distinct ectoderm and endodermal cells present in these sections through gene expression alone. Def. endoderm was the clearest of the two as it had high expression of Sox17 but could be distinguished from HEPs by undetectable levels of *Etv2, Scl, Cdh5* and *Tek* in all three embryos. Def. endoderm was also consistently high for the expression of *Vegfa* (Fig.4.14). The role of Vegfa in this context is incompletely understood but is reminiscent for the requirement of VEGFA expression in the visceral endoderm for YS vasculature [317]. Additionally, VEGFA has been suggested as a potential attraction signal to mediate directed migration of *Scl*$^+$

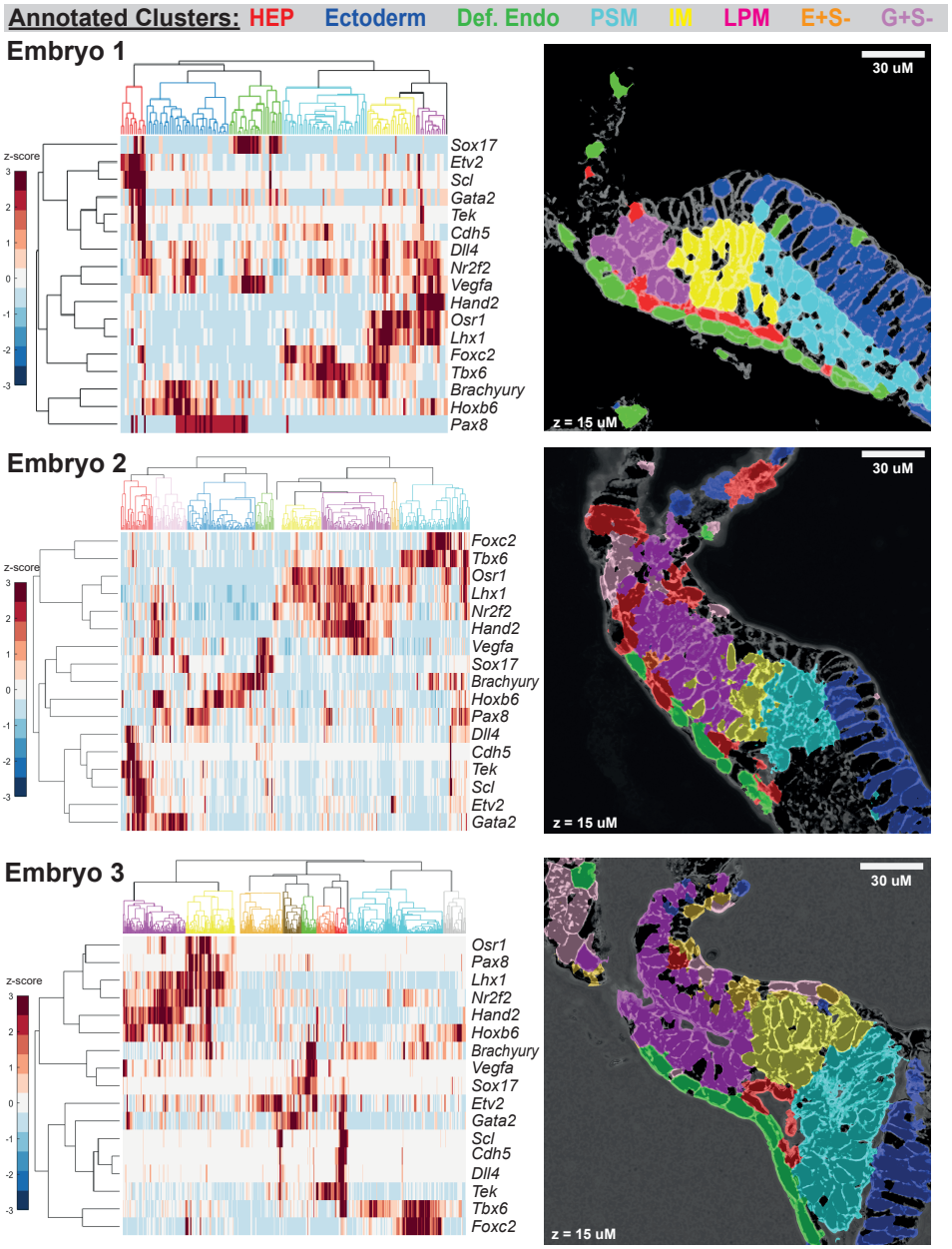


Figure 4.14: Quantification of gene expression to enable cell type clustering and spatial remapping

Quantification of eighteen genes stained in seven rounds of dHCR imaging is shown in row (gene) normalised heatmaps for three regions from three embryos. Z-score is shown for each gene (row) for single cells (columns). The result of hierarchical clustering of genes and cells is shown in the dendrograms on the left and above the heatmap respectively. Clusters were manually annotated by inspecting co-expression or absence of key marker genes and their dendrograms coloured according to the annotation at the top of the figure. These clusters were spatially remapped onto the sections with the same colour. Shown on the right is the remapped clusters at a depth of 16µm. HEP

HEPs from the YS [142].

Although I did not include markers of ectoderm, I could identify cells clustering together that mapped primarily to the neuroepithelium (Fig.4.14). The cells clustered together through their expression of *Brachyury* and *Hoxb6*. These are not markers of this tissue but, combined with their absence of other marker genes, could be clustered together. In embryo 1 and 2, I also detected relatively high expression of *Pax8* as compared to other cell types in the ectoderm but not in embryo 3. This will be emphasised by the row normalisation performed in the heatmap, as the IM does not yet robustly express *Pax8* in embryo 1 or 2.

Finally, in embryos 2 and 3 two clusters were identified that did not match a defined cell type. I have labelled these as *Etv2*$^{+}$*Scl*$^{-}$ (E+S-, orange) and *Gata2*$^{+}$*Scl*$^{-}$ (G+S-, pink) (Fig.4.14). The E+S- cluster is characterised by high levels of *Etv2* expression but absence of all other genes high in the HEP cluster (e.g. *Scl, Gata2, Tek, Cdh5*). These cells also express *Lhx1* but no other genes probed at a consistent level. I suspect these represent the rare *Etv2*$^{+}$ cells observed during qualitative assessment of HEP markers (Fig.4.10). The contribution of these cells to haemato-endothelial development is unknown. I do not focus on these cells here as their lack of expression of *Scl* in this window of development when SCL is required, suggesting they are not required for intra-embryonic blood development. Secondly, the G+S- cells were more numerous than E+S- and had a consistent location. These cells were detected in embryo 2 and 3 primarily in the most lateral regions of the section at the border between embryonic and extra-embryonic tissue. Most of these cells likely represent YS mesoderm previously observed to be high in *Gata2* [318].

With clusters annotated, I spatially remapped these clusters onto the membrane stained images. The remapping could be used to validate the accuracy of cell clustering and gene quantification (Fig.4.14). Broadly, a cells annotation reflected its anatomy with the vast majority of cells in a cluster assigned to the correct region. There are

rare instances of cells annotated as forming a part of one cluster spatially remapping to the incorrect region (<10%). This is most common at the borders of two cell types (e.g. HEP annotation in the def. endoderm) or in clusters where not many marker genes are available (e.g. def. endoderm). Furthermore, the most medial PSM cells in embryo 2 are not segmented (and therefore quantified) because of poor antibody staining (Fig.4.14). The majority of cells however map correctly.

Embryo 1 at a z depth of 15µm has HEPs lined along the ventral aspect of the mesodermal as observed by the expression of *Scl* and *Tek* 4.14. Mapping the clusters allows us to visualise the different cellular neighbours these ventral HEPs have. While they all share the def. endoderm on their ventral aspect, the most medial ventral HEPs are in contact with pre-somitic mesoderm (Fig.4.14). As I move to the lateral edge there are HEPs in contact with both the IM and LPM. Furthermore, the HEPs on the dorsal side of the mesodermal wing do not associate with the PSM or IM at this transverse level, and are observed in direct contact with the LPM only (Fig.4.14). To what extent these different cellular neighbours may send distinct cellular signalling and whether this generates heterogeneity within the HEPs is unknown.

Focusing on the HEPs of the more mature embryo 3, I see this heterogeneity in cellular neighbourhoods is maintained even when cells of the dorsal aortae are forming chords and lumenising (Fig.4.14). It is widely accepted that within the DA of E9.5 - E11.5 embryos that HE only resides in the ventral wall. Studies have focused on the signals during these stages that may induce or maintain this dorsal-ventral polarisation in HE location. However, whether when this heterogeneity is established and whether signals at stages prior to HE emergence could pattern endothelial cells of the DA remains unexplored. In embryo 3, I observe that already cells of the DA are all contacting different cell types (Fig.4.14). The ventral most HEPs are in contact almost exclusively with the def. endoderm, while those on the lateral edge primarily contact the LPM. The most dorsal HEPs are in immediate proximity with the IM and the medial HEPs with the PSM. Although less complex than the cellular component of the E9.5 - E11.5 niche,

this diversity in cellular neighbours could conceivably impose a difference in cellular signalling and induce heterogeneity.

4.2.7 Investigating the degree of heterogeneity in gene expression between spatially distinct haemato-endothelial progenitors

Spatially remapping clusters allows us to evaluate the variation in expression of key genes based upon location. Having identified a line of HEPs along the medial-lateral axis on the ventral aspect of the mesodermal wing, I hypothesised that this may represent cell migrating towards the midline and maturing as they move. To test this I evaluated the expression of key HEP maturation genes in cells ordered along the medio-lateral axis (Fig.4.15. Genes plotted individually in Appendix Figs.11 & 12).

Since I only observed this line of cells in embryo 1 and 2, I only quantified the variation from these embryos. I ordered cells according to their distance (in three-dimensions) from the most medial HEP. The graphs in figure 4.15 show the number of transcripts per unit volume (tpv) for each gene of interest. If more mature endothelium was indeed more medial, I would expect higher expression of *Cdh5* as a marker of endothelium and lower expression of *Etv2*, as this is downregulated from endothelial progenitor to endothelium. In both embryos, there were two distinct regions. The most medial cells (on the right side of the graph in the first 550 pixels) are cells forming the developing DA and can be characterised by higher expression of *Cdh5* ($>$13 tpv) and in some cells higher expression of the notch ligand *Dll4* ($>$5 tpv) indicating the first adoption of an arterial identity. Lateral to the DA, *Cdh5* and *Dll4* expression is lower and is particularly clear when each gene is plotted individually (Appendix Figs.11 & 12). In embryo 2, this coincides with the higher expression of *Etv2* and *Sox17*. In embryo 1, this effect is less apparent and only the most lateral HEPs have high *Etv2* expression. In conclusion, ordering cells along the medial-lateral axis shows that the most medial cells (forming the DA) are more mature and the further from the DA a HEP is, the more immature its expression pattern of these key genes is.

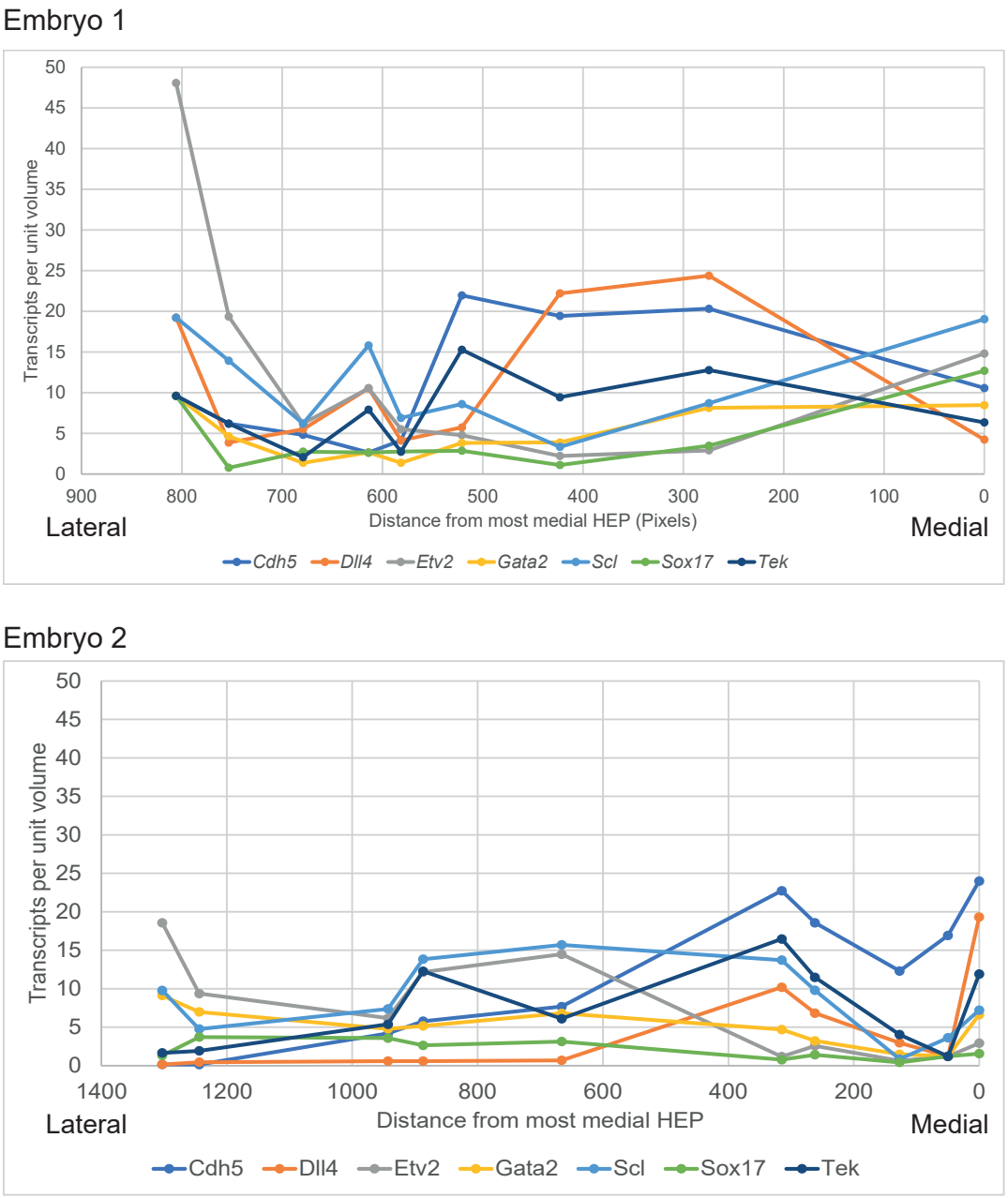


Figure 4.15: Variation in HEP gene expression along the medial-lateral axis
Cells defined as HEPs in hierarchical clustering of gene expression data who were identified to be on the lateral edge of the mesodermal wing were ordered according to their distance from the most medial HEP (along the medial-lateral axis). Distance is measured in pixels. The transcript count per unit volume ((transcript count / cell volume) x average cell volume) was plotted for each cell. Each data point is the normalised transcript count with lines joining cells to show general trends. Points and lines are coloured accoring to gene.

4.3 Discussion

Having attempted to partially recapitulate the *in vivo* niche of Wave 3 blood progenitors *in vitro* but being unable to consistently induce a blood programme, I wanted to improve the characterisation of cells undergoing blood fate specification. Through reviewing studies on SCL, the master regulator of blood fate specification, I identified E8.0 as a critical period during which to study the earliest progenitors of Wave 3 haematopoiesis. This time in development is relatively poorly characterised for HEPs as compared to later stages. Additionally, I wanted to characterise the cells *in situ*, to define their niche, and how it may differ between different HEP populations. Volumetric imaging of E8.0 embryos showed that *Scl* was widespread in the caudal half of the embryo as described previously [251]. These cells were distributed along the length of the dorsal aortae, but also present more laterally and are scattered on the dorsal side of the mesodermal wing. To better understand the HEPs at this time I wanted to study to what extent there is heterogeneity in the expression of key haemato-endothelial genes in these differently located populations, and more accurately define the cellular components of their niche.

To this end, I established a quantitative and multiplexed smRNA-FISH method only previously utilised in the neuroepithelium of chicken embryos [258, 310]. I have demonstrated that cryosections of E8.0 mouse embryos can be stained using dHCR and subsequently stripped with DNaseI at least eight times with limited RNA degradation. Although restricted to the quantification of single fields-of-view, I was able to quantify the expression of 18 genes in the same tissue section. Quantitative assessment of multiplexed dHCR has enabled us to confidently map 5 cell types (HEPs, PSM, IM, LPM, Def.endoderm) using transcript counts from just 18 genes.

SGA has provided insight into the heterogeneity in HEPs during this critical window. Firstly, by mapping clusters I am able to categorically assign cell identity to the cells in

direct contact with HEPs. Through mapping all HEPs using a broad range of markers, I can conclude that Scl is expressed in almost all progenitors of the haematoendothelial lineage at this time. This broad expression in posterior HEPs during the window of blood fate specification, drove us to develop a Scl-CreERT2 fate mapping mouse line. We are in the process of validating a mouse line with Cre-ERT2 inducible recombinase knocked into the 3' UTR of the Scl locus. Although I had hoped to use these mice for this project, difficulty with chimera generation and germline transmission of the transgenic allele delayed progress such that the line is only now ready for validation and characterisation. The WIMM transgenic facility performed 12 blastocyst injection sessions resulting in only 3 chimeric mice. The first two chimeras each had $>$60 pups and did not transmit the transgenic allele once. The third transmitted after 16 pups. Since outcrossing, I now aim to describe the labelling of blood and endothelial populations with induction at various times during development. This will test our hypothesis that although all HEPs are expressing Scl that not all these cells are blood fated.

Such broad expression of Scl during the window of blood fate specification likely means that not all cells that express Scl will be blood fated and that Scl's role is to establish a cell state receptive to external signals that commit cells to the blood lineage. This highlights the importance of studying cells *in situ* as these signals will be sent from the cells in the immediate vicinity. Another possibility is that Scl is not performing the same role in all cells, and that association with different protein partners or varied chromatin accessibility could specify a subset of Scl-expressing cells to a blood fate [218]. These two mechanisms are not mutually exclusive and so both require further investigation.

The ability to quantitate spatially resolved gene expression data provides the possibility to phenotype cells *in situ*, where their identity may be perturbed. Having defined cell types with SGA in wildtype embryos, I would like to use this to map cell types in the context of *Scl* null embryos. In addition to their complete absence of blood [148, 319], HEPs of Scl-null embryos adopt alternative fates with cardiomyocytes

forming in the YS and adoption of cardiac and paraxial fates *in vitro*[150]. A complete description of the fates adopted and the location of these cells *in vivo* is lacking.

I had hoped to use SGA to accurately map the location and identity of these cells by fate mapping Scl^{+} progenitors using a novel Scl knockin-KO line where Cre-ERT2 was knocked in to exon 4 of Scl and thus rendering it non functional [320]. When bred to homozygosity, Scl-null embryos could have Scl-expressing cells fate mapped and thus the full range of their alternative fates characterised. Using this mouse in combination with identity mapping with SGA I hoped to describe where alternative fates arise. Understanding whether the fate adopted is random or different fates map to different locations would inform us on the heterogeneity of HEPs during WT haemato-endothelial development and provide insight into the possible different cellular contexts in which Scl is functioning. SGA provides a great platform with which to do this as cells could be confidently mapped with multiple genes per cell type. Unfortunately, 12 rounds of blastocyst injections resulted in only 2 chimeras and no germline transmission from $>$60 pups each. With difficulty in producing this fate mapping line I hypothesised I may be able to identify these alternative fates through comparison with WT embryos alone so I attempted to revive previously published Scl-null lines. Unfortunately, the frozen sperm in storage did not result in viable pups. I have since recieved these lines from elsewhere but delays in shipping and out-crossing have also delayed these experiments.

Chapter 5

Predicting cellular signalling to angioblasts in the E8.0 embryonic posterior

Contents

5.1 Introduction 131
5.2 Results 134
5.2.1 Establishing a rapid staining protocol to discern cell borders and identify haemato-endothelial progenitors on mouse embryo sections 134
5.2.2 LCM is able to cut and capture single cells but lysis results in degraded cDNA 139
5.2.3 scRNA-seq of the caudal half of E8.0 140
5.2.4 Cell communication patterns received by HEPs in the caudal half of E8.0 embryos 148
5.2.5 Cell signalling pathways recieved by the HEPs in the caudal half of E8.0 embryos 152
5.3 Discussion 158

5.1 Introduction

In Chapter 4, SGA demonstrated that haemato-endothelial progenitors are present in a wide variety of cellular niches with cells forming the same vessel in contact with different cellular neighbours. To what extent these heterogeneous environments translate to distinct signalling and differential fates is poorly understood. The identification of signals currently unknown to specify haemato-endothelial fate can be performed through coincidental perturbation in a genetic model, screening potential ligands and receptors by *in situ* hybridisation (ISH)/ immunohistochemistry (IHC) or by computational prediction from transcriptomic data. Transcriptomic data has the added benefit of allowing other analyses such as differential gene expression and was selected as the approach for this study.

With scRNA-seq and its analysis becoming increasingly accessible, many systems have been studied at high throughput and resolution (single cells) without retaining spatial information. These data have been used to discover heterogeneity in seemingly homogeneous cell populations, sensitive measurement of the transcriptomic consequences of perturbation, and even inference of lineage trajectories [33, 321]. There is an increasing effort to integrate spatial information within these analyses either through computational integration of data sets or high-throughput spatial transcriptomics. The cells of interest can be studied *in situ*, providing a platform to study cell-cell interactions more faithfully and also, allowing cells whose transcriptome would be indistinguishable by current clustering methodologies to be separated on location alone. This is advantageous as it enables more sensitive discovery analysis. For example, the transcriptomes of haemato-endothelial progenitors of the E8.0 caudal region are all much more similar than they are different, sharing a core transcriptional programme [143]. Therefore, they cluster as one and cells that lie dorsally versus ventrally cannot be distinguished and subsequently compared. However, if these transcriptomes could be linked to their location, it negates the need for clustering and differential gene expression analysis could detect more subtle differences. These differences could be the critical difference between cells with blood differentiation potential and those without.

Therefore, we aimed to establish a spatial transcriptomic method. Typically, spatial transcriptomics methods can be split into those that use highly multiplexed FISH (e.g. osmFISH [312], seqFISH+ [242], MERFISH [243]) and those using Illumina next generation sequencing (e.g. 10x Visium, Slide-seq [322], LCM-seq [249]) (Fig.1.7). Based on our aims and available equipment, we required single cell resolution and unbiased gene detection, so opted for LCM-seq [249, 323]. This method relies on the laser microdissection of cells of interest from tissue sections into collection tubes, before performing Smart-seq2 library preparation and Illumina NGS [249]. While relatively low throughput, this method allows unbiased detection of a high number of genes ($>$10,000 detected from single cell [323]) with high spatial accuracy. Ideally, single cells can be

processed but with the possibility of pooling equivalent cells from serial sections if needed.

LCM-seq is performed from a limited number of sections and is therefore low throughput. scRNA-seq data from the embryonic region of interest are often used to supplement LCM-seq data (using the spatially resolved datasets as anchors to spatially map relevant cells and clusters [246]), thus increasing the number of cells analysed and genes identified, as well as the statistical power of downstream analyses. Since we are focusing on E8.0 embryos, the mouse gastrulation atlas could be used for this purpose [33]. However, this resource was designed to map all cells within the embryo without distinction between intra- and extra-embryonic tissues, and no accurate way to order cells along the anterior-posterior axis. Both of these variables are critical to focus on the intra-embryonic angioblasts demonstrated to have HSC potential [173]. Therefore, we also aimed to perform scRNA-seq on dissected caudal E8.0 embryos to support our SGA and LCM-seq analysis.

Ultimately, transcriptomic data can be used to predict signalling using ligand-receptor interaction programmes [237, 262, 324]. These software use manually curated databases of known ligand-receptor interactions and gene expression across pre-calculated clusters to predict possible signalling. Such methods have improved through extension of their ligand-receptor databases and addition of signalling nuances such as heterodimeric receptors and co-receptors. CellChat is a second generation method that integrates information on heterodimers and co-receptors to predict inter-cellular interactions, and use pattern recognition to classify signals [262]. Therefore, CellChat was selected for our analysis. Ideally, LCM-seq data would be integrated with scRNA-seq to identify the haemato-endothelial progenitor neighbouring cell types identified by SGA for prediction of inter-cellular interaction.

5.2 Results

5.2.1 Establishing a rapid staining protocol to discern cell borders and identify haemato-endothelial progenitors on mouse embryo sections

LCM-seq relies on first cutting and catapulting cells from embryonic sections into collection tubes, using a LCM-microscope. The user first defines the path of the cutting laser manually by drawing over an image of the tissue. For accurate micro-dissection, HEPs and their neighbouring cells must be identifiable and their cell borders distinguishable. Importantly, care should be taken that the labelling approaches do not severely impact the quality of RNA collected. To minimise this, embryos are dehydrated and kept frozen whenever possible. To remove the need for any staining protocols (that require incubation in acqueous solutions), I tested labelling cell borders and HEPs using available transgenic mouse lines.

ETV2 is the master regulator of haemato-endothelial development and sits at the apex of the transcriptional regulation of the endothelial programme [143]. Homozygous deletion of *Etv2* results in the absence of all blood vessels and blood cells [144, 325]. Of particular interest to us, Etv2-Cre lines crossed with R26R-LacZ [326, 327] or R26R-EYFP [327] Cre-reporter lines exhibit high levels of labelling of the earliest endothelial progenitors. We aimed to use this system to specifically label HEPs for identification and cutting by LCM. We obtained the transgenic heterozygous Etv2-Cre strain carrying the Cre recombinase gene under control of a 3.3kb-fragment of the upstream control region of Etv2 that was shown to faithfully reproduce Etv2 expression [326]. We first validated the reported high level of recombination and labelling by crossing the Etv2-Cre line with the Cre activity reporter strain Ai9 [328]. This line harbours a floxed stop cassette upstream of the tdTomato (tdT) gene in the Rosa26 locus.

Firstly, to confirm that Etv2-Cre labels the intra-embryonic endothelial progenitors giving rise to the adult blood system, we collected blood from adult mice obtained from

Etv2-Cre x Ai9 crosses. Using flow cytometry on red-blood cell depleted peripheral blood from 8 week-old animals to assess the percentage of cells labelled confirmed only animals carrying the Etv2-Cre allele underwent recombination with 95.65% of blood cells labelled with tdT (Fig.5.1a & Appendix Fig.13). The unlabelled blood cells (4.28%) likely reflects slight inefficiency in labelling or cells that have silenced the Rosa26 locus. To confirm that Etv2-Cre labels the major embryonic vessels and specifically, the DA we then collected E10.5 embryos from Etv2-Cre x Ai9 crosses for whole mount imaging of embryonic vessels, using anti-CD31 antibodies to label endothelial cells. All $CD31^{+}$ embryonic vessels including the DA (green arrows) were labelled by tdT demonstrating the suitability of this model to label endothelial cells (Fig.5.1b). Additionally, rounded cells in the vessel lumen seemingly budding out from the endothelium were all also tdT^{+} suggesting the earliest intra-embryonic haematopoietic cells were labelled (Fig.5.1b). These data confirm the Etv2-Cre model behaves as published and is suitable to label haemato-endothelial progenitors.

For LCM-seq experiments, we initially chose to breed the Etv2-Cre line with the mTmG Cre-activity reporter ($Gt(ROSA)26Sor^{tm4(ACTB\text{-}tdTomato,\text{-}EGFP)Luo}$) [329] as this results in embryos whose cells express a membrane-localised red fluorescent protein (RFP, tdTomato) in all non-recombined cells and a membrane-localised -green fluorescent protein (GFP, eGFP) in those that have had the mTmG locus recombined by Cre. Since Cre is under the promoter of Etv2, HEPs should have a GFP plasma membrane while their neighbours should have RFP borders. Labelling the cell borders in this way would allow us to identify where to cut with the laser without capturing neighbouring cell cytoplasm (and RNA). However, when E8.0 embryos were harvested from Etv2-Cre x mTmG breedings, all embryonic cells have both red and green localised fluorescence in the cell membrane (Fig.5.1c). This is in sharp contrast to cells imaged using the same microscope settings in wildtype embryos processed in the same manner who had no detectable fluorescence (Fig.5.1d). Furthermore, when imaging embryos sectioned in decidua (maternal tissue) from dams not carrying the mTmG transgene, no fluorescence could be detected in decidual tissue suggesting the source of signal is indeed from the

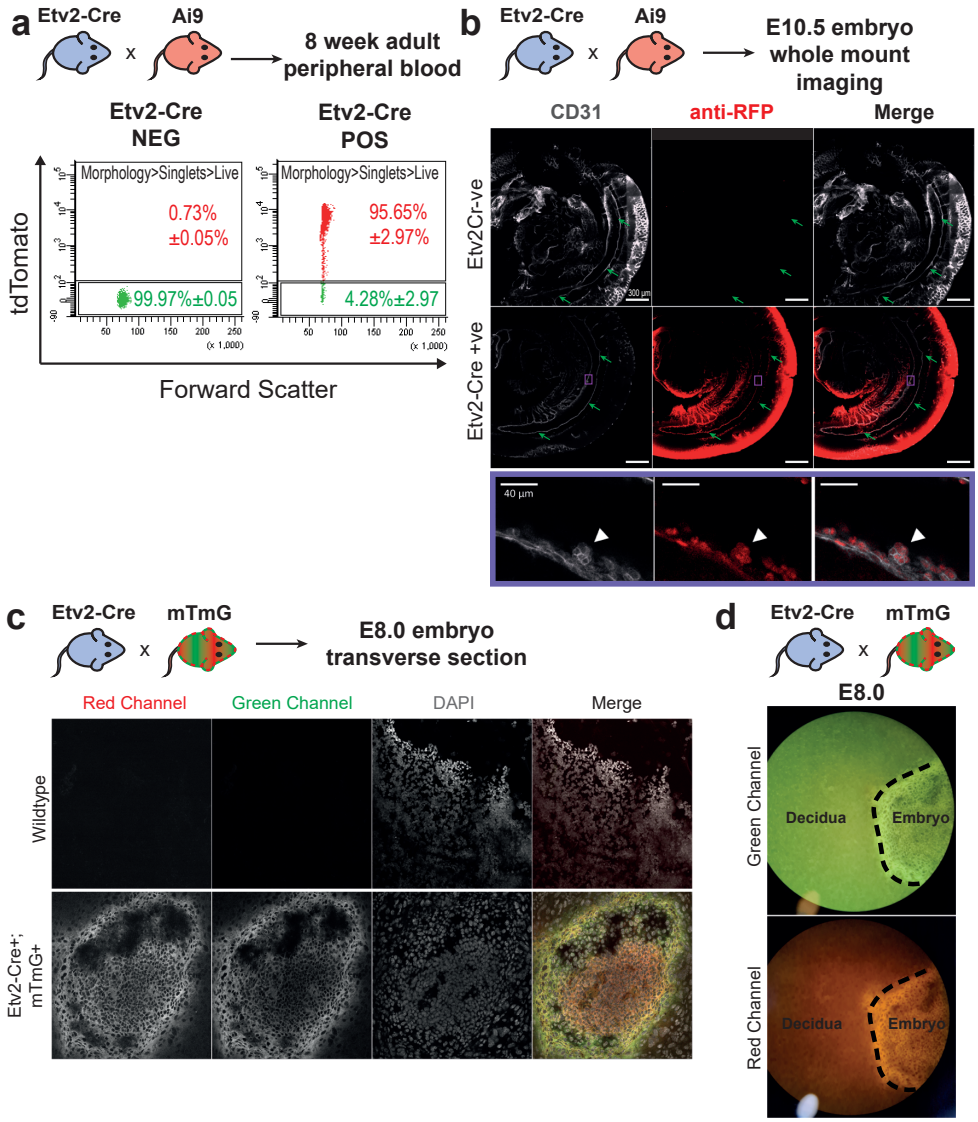


Figure 5.1: Etv2-Cre can be used to mark major embryonic vessels and Wave 3 blood but is unsuitable for labelling cells for LCM-seq

a Red blood cell depleted peripheral blood from 8 week old mice resulting from Etv2-Cre x Ai9 crossing, was gated on classical FSC-SSC morphology and for live singlets (Appendix Fig. 13) before assessing the percentage of tdTomato positive cells. Mice carrying Etv2-Cre allele (POS) exhibit a high level of Cre-mediated recombination as compared to those who did not inherit it (NEG). Percentages represent mean±SEM from 3 mice. Full gating strategy shown in appendix (Fig.13). **b** Whole mount imaging of E10.5 embryos obtained from Etv2-Cre x Ai9 crosses. One z slice at 100 µm is shown. Endothelial cells are labelled with CD31 and, tdTomato positivitiy is detected by anti-RFP antibody. Green arrows indicate the DA. Purple region highlighted shows intra-aortic clusters closely associated with DA endothelium is also tdTomato-positive (white arrow head). Representative images from 1 of 3 embryos imaged per genotype. **c** Transverse section through E8.0 mouse embryos from either wildtype or Etv2-Cre x mTmG breeding showing high green and red fluorescence in Etv2-Cre+mTmG+ embryo only. **d** Image captured down the eyepiece of a microscope showing the high fluorescence caused by Etv2-Cre x mTmG (in **c**) is specific to embryonic tissue and absent from maternally derived decidua in embryos from mTmG-/- dams.

mTmG locus. Finally, fluorescence could be visualised by eye showing fluorescence was bright and that the microscope is performing as expected as the eyepiece uses different excitation and emission sources and filters (Fig.5.1d). These data were confusing as the mTmG reporter locus is designed to only express one colour at a time and is widely published. This could reflect a delay in tdT degradation while eGFP is expressed but does not explain why it would be in all cells.

With transgenic mouse lines providing unexpected results, we instead developed a protocol for the rapid and clean staining of cell borders and HEPs in preparation for LCM-seq (Fig.5.2a). This protocol is designed to label cell borders with fluorescently conjugated wheat germ agglutinin (WGA), and HEPs with a directly conjugated SCL antibody in a short period of time (10 minutes) as to minimise RNA degradation in the aqueous solutions required for staining. Firstly, WGA staining in 30 seconds at a concentration of 5 µg/mL on transverse sections of E8.0 mouse embryos. Staining was bright and with good signal:noise ratio showing the plasma membrane of cells throughout embryonic sections (Fig.5.2b). An antibody against SCL was directly conjugated in-house and added to sections at high concentrations (1:50, conventional is 1:500) for 4 minutes before washing. Staining performed in this way was dimmer than conventional staining protocols but could be accounted for by adjusting imaging conditions (Fig.5.2c). Finally, sections are normally dehydrated before being cut on the LCM [249]. To assess the consequence of dehydration on staining, sections were imaged before and after dehydration. SCL staining was noticeably dimmer but still easily detectable after dehydration (Fig.5.2c). In short rehydration, WGA and SCL staining, washes and dehydration can be performed in approximately 10 minutes allowing cell borders and HEPs to be identified. To test the effect on RNA quality, the LCM process needed to be established.

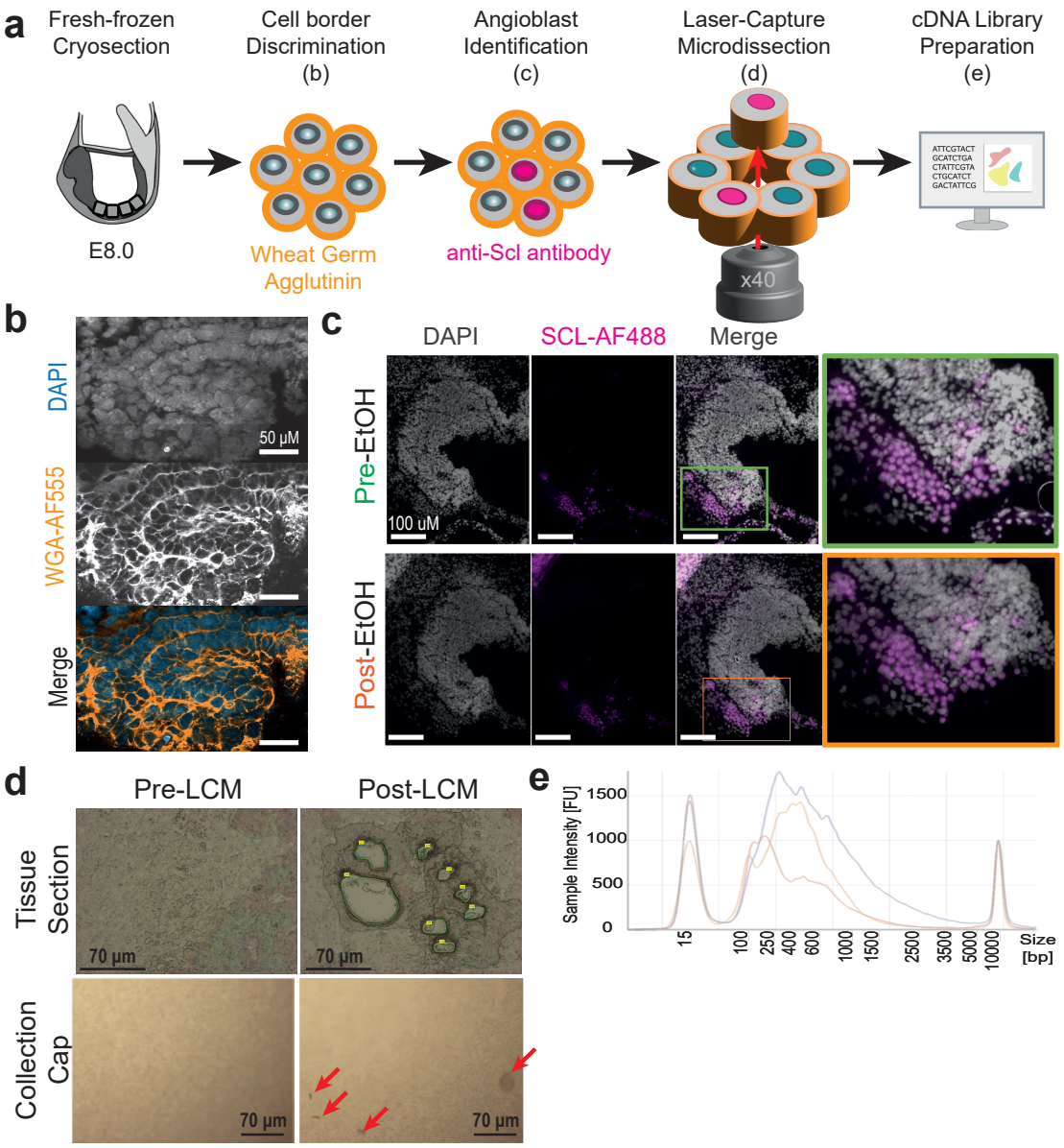


Figure 5.2: Discrimination of cell borders and identification of angioblasts by rapid staining in preparation for cDNA library preparation

a Schematic showing the overview of the LCM-seq workflow. **b** and **c** Fresh-frozen sections obtained from E8.0 wild-type mouse embryos are stained with fluorochrome-conjugated wheat germ agglutinin (WGA) to discriminate cell borders (**b**) or immunostained for angioblast marker SCL (**c**). The effect of dehydration with ethanol (EtOH) on antibody staining on fluorescence was qualitatively measured. Display values are equal in both images. **d** The ability of an LCM microscope to cut, catapult (top panel shows tissue section) and capture (bottom panel shows capture cap) decidua tissue was tested. User defined cut regions are shown in green (top right) and the location of captured tissue is highlighted with red arrows). Captured single cells are lysed to release RNA for reverse transcription and cDNA preamplification. **e** Representative quality control graphs of pre-amplified cDNA libraries from single cells run on a HS D5000 Tapestation (n=30 cells). Lower and upper markers are shown.

5.2.2 LCM is able to cut and capture single cells but lysis results in degraded cDNA

LCM is broadly applicable to many tissues for use with many downstream applications [330]. We aimed to apply LCM-seq to fresh-frozen E8.0 mouse embryos. LCM-seq requires the collection and sectioning of fresh-frozen tissue [249, 250] for efficient RNA lysis and reverse transcription so embryos cannot be PFA fixed as per our usual sectioning protocol. Sectioning fresh frozen E8.0 mouse embryos was challenging and collection of intact, morphologically normal embryonic tissue sections was difficult. Encouraged by the work of Chen, *et. al.,* who perform LCM on E7.5 fresh frozen mouse embryos [40] using Geo-seq [250], this was improved by retaining the myometrial layer around the decidua (Appendix Fig.14a). This allowed the collection of morphologically normal sections (Appendix Fig.14b) but with much lower frequency than from PFA fixed embryos. Therefore, LCM cutting and catapulting was first confirmed on decidual tissue.I acquired images before and after cutting multiple single cells and larger regions of decidual tissue (Fig.5.2d, top) before checking the cap for the presence of catapulted material (Fig.5.2d, bottom). Cell capture rates were estimated to be 80% (approximate as identifying a cell in the cap can be challenging).

Subsequently, single cells were cut and catapulted from mouse embryo sections having undergone the staining protocol before immediate lysis in Smart-seq3 lysis buffer and snap freezing on dry ice. I collected a total of 64 cells across three LCM sessions into their own tubes, cDNA libraries were generated from each cell with RNA degradation assessed following cDNA amplification. Unfortunately, all cells profiled exhibited a TapeStation profile consistent with degraded RNA (Fig.5.2e). Peaks for Smart-seq3 libraries should peak at 1000 base pairs while LCM-derived cells peaked 300. Cells sorted by FACS into the same lysis buffer used for LCM collected cells meanwhile demonstrated traces expected from non-degraded cells. Cells collected from unstained sections exhibited the same profile, indicating that degradation is not caused by the staining protocol. Repetition of the experiment with more thorough RNase treatment did not alleviate the issue. In order to perform LCM-seq, further optimisation of cutting

and sample preparation conditions are required.

5.2.3 scRNA-seq of the caudal half of E8.0

Our next aim was to analyse scRNA-sequencing datasets, initially for the mapping to LCM-seq data. To map to LCM-seq data and compare with our multiplexed smRNA-FISH data, cells should be from the same embryonic region. The mouse gastrulation atlas (MGA) performed scRNA-seq on cells from the whole embryo and its yolk sac. We wanted to be confident in the location of our scRNA-seq dataset and the MGA did not allow this. Therefore, we aimed to harvest and sequence cells from the caudal half of the embryo proper, posterior to the second somite and remove the allantois and it's embryonic connections (Fig.5.3a). In this way, we can be more confident that any HEPs and neighbouring cells we collect reflect the region described by SGA and containing the intraembryonic Wave 3 HEPs [173]. We decided not to dissect out the splachnopleura, the precise location of Wave 3 HEPs, as we also wanted to include HEPs on the dorsal edge of mesoderm so that when LCM-seq is established and we can spatially map these cells, we may compare them to those of the dorsal aorta.

Seven, 4 somite stage embryos were dissected from their YS, cut posterior to the second somite and had allantois and associated embryonic tissues removed (Fig.5.3a). Dissected regions were dissociated into single cell suspensions and counted. 20,000 cells were loaded into the 10x Chromium controller with the aim of capturing 12,000 cells. This should allow for the capture of approximately 1-embryo equivalent as the whole posterior contains 12,596 as quantified from whole-mount imaging data (Fig.4.2d). After unloading the controller, 10x libraries were prepared according to the manufacturers instructions and resulted in pooled cDNA libraries of the expected size (400bp) for a 10x library (Fig.5.3b). The library was sequenced to a depth of 1.2 billion reads, aiming for 100,000 reads per cell. Running the data through 10x CellRanger, 11,598 cell barcodes were detected with a mean read count per cell of 100,232 and median genes detected of 4,734. Looking at sequencing saturation, approximately 40% of the library has been

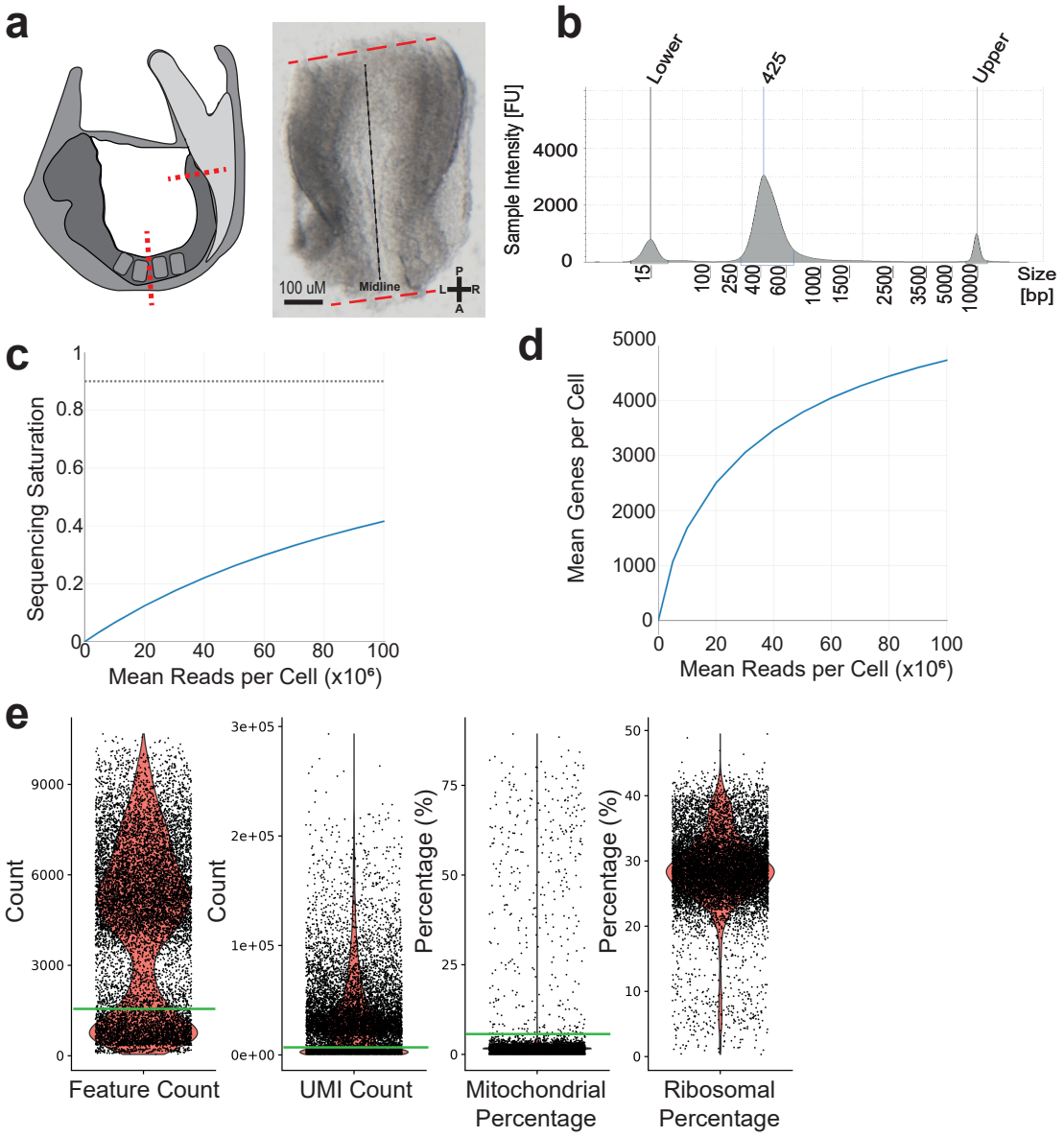


Figure 5.3: Overview and quality control of scRNA-seq of E8.0 embryo caudal half
a Schematic demonstrating the cut sites (dotted red lines) of the lateral view of a 4 somite-stage embryo. The region posterior to the second somite but anterior to the allantois and its connecting embryonic tissue was dissected. A ventral view of one of the seven embryos dissected is shown on the right with the midline indicated by the black dotted line. A-P and L-R axes are shown in the compass. **b** Tapestation trace derived from the electropherogram of the prepared 10x library. Lower and upper markers are shown at 15 and 1000 base pairs respectively with the cDNA library peak at 425. **c** Sequencing saturation as a function of mean reads per cell. As mean read per cell increases, so too does sequencing saturation but with diminishing returns. **d** Mean genes per cell as a function of mean reads per cell. **e** Key parameters for 10x data. Each black dot represents a single cell with the violins showing distributions of cells with width indicating frequency. Green lines on feature (gene) counts, UMI counts and mitochondrial percentage represent thresholds with which cells were filtered. A, anterior; P, posterior; L, left; R, right; FU, fluorescent units; bp, base pairs; UMI, unique molecular identifier.

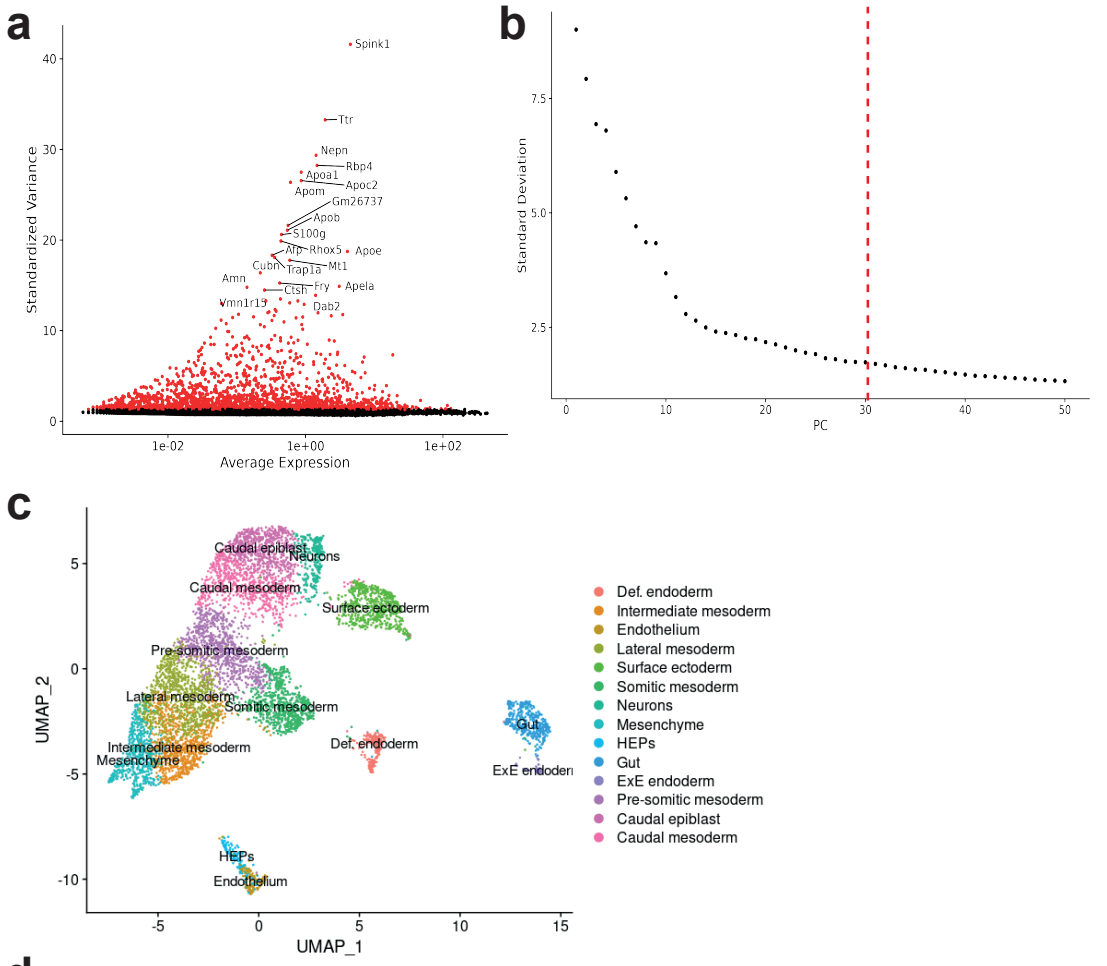
sequenced (Fig.5.3c) with the number of genes starting to plateu (Fig.5.3d). This is an acceptable level as, if we extrapolate the data then its predicted that even doubling the sequencing depth is only likely to increase this by approximately 16% and increase the number of genes per cell by 850 (Appendix Fig.15). Therefore, this sequencing depth is a good balance between sensitivity and cost. To perform downstream analysis, cells of low quality are removed by placing a threshold on a number of variables. The exact threshold values are somewhat subjective and are based upon the distribution of all cells (Fig.5.3e) on a number of key parameters. Cells with gene counts below 1500, UMI counts below 1250 or a percentage of reads mapping to mitochondrial genes over 10% are all removed. This leaves 8604 cells (74%) with a median of 5456 genes and 33370 UMIs detected per cell. This provides another advantage of using these data over the MGA where a median of 3,200 genes and 15,000 UMIs were detected in E8.0 embryos, likely due to the use of v1 of the 10x chromium chemistry [33].

Following normalisation and scaling of the data using a standard Seurat analysis workflow, cells were clustered and projected into low-dimensional space to facilitate cell-type annotation. Firstly, the top 2000 most variable genes across the cell population were calculated and plotted (Fig.5.4a). The most variable genes (labelled in Fig.5.4a) are commonly associated with the extraembryonic (ExE) endoderm (e.g. *Ttr, Apoa1, and Apom*, likely reflecting its distinctive transcriptional signature. The top 2000 variable genes are used to run a principle component analysis (PCA). The contribution (standard deviation) of each principle component (PC) is plotted on an elbow plot to help assess the number of PCs required for downstream analysis (Fig.5.4b). The standard deviation fell with each PC and an elbow formed at PCs 25-30. 30 PCs were used to find neighbours in low dimensional space, cluster cells and generate a Uniform Manifold Approximation and Projection (UMAP) (Fig.5.4c). Initially, these data contained two clusters that were enriched for low gene and UMI counts as compared to other clusters (2700 genes versus 5456 genes) (Appendix Fig.16a,b). Their G2M and S-phase scores were normal demonstrating they are not just cycling cells (Appendix Fig.16c). Furthermore these clusters were not enriched for genes defining known cell

types but rather ribosomal genes. Together, these data suggested that these clusters represented low quality cells not excluded in the first round of filtering and so were subsequently removed. This left 14 clusters for cell type annotation. These clusters all displayed the expected range of UMI and gene counts with no issues with mitochondiral or ribosomal percentage (Appendix Fig.17).

The annotation of scRNA-seq data represents a significant step in accurate analysis. Clusters were manually annotated through investigation of the top enriched genes in each cluster and displaying expression levels of known marker genes. Significantly enriched genes were calculated for each cluster with the top ten for each cluster used to aid annotation (Fig.5.5). Each cluster displayed a distinct expression pattern of the top 130 enriched genes, each with high expression in the 10 genes enriched in their cluster and primarily low expression in other clusters' enriched genes with a few exceptions. Cells forming the presumptive mid- and hindgut were identifiable by their enriched expression of *Apela* [331, 332], *Nepn* [333] and *Sox17* [334]. Meanwhile, the intermediate mesoderm could be identified by its expression of *Lhx1* [75, 335] and *Twist1* [336, 337] (Fig.5.5). As seen in my FISH profiling in Chapter 4, many of these genes are also expressed in lateral mesoderm at this time. However, differential expression of *Hand1* and *Foxf1* (Fig.5.5) allows accurate calling of these cell types as Hand1-Cre does not label IM derivatives [338] and Foxf1 knockout impacts only lateral plate and extra-embryonic mesoderm. It should be stated that the lateral mesoderm cluster likely contains both embryonic lateral plate and extra-embryonic mesoderm as they cannot be separated by gene expression alone.

The two paraxial mesodermal clusters pre-somitic and somitic mesoderm, were easily distinguishable. Somitic mesoderm can be characterised by the expression of *Tcf15* [339, 340], *Meox1* [341] and *Foxc2* [342, 343] which show much lower expression in pre-somitic mesoderm (Fig.5.5). Additionally, somitic mesoderm can be distinguished from pre-somitic mesoderm by lower expression of *Hes7* [344], *Tbx6* [62] and *Dll1* [345] (Fig.5.5). Clusters of caudal epiblast and caudal mesoderm were identified (Fig.5.5).



	Def. endoderm	Intermediate mesoderm	Endothelium	Lateral mesoderm	Surface ectoderm	Somitic mesoderm	Neurons	Mesenchyme	HEPs	Gut	ExE endoderm	Pre-somitic mesoderm	Caudal epiblast	Caudal mesoderm
Freq	227	802	143	975	610	720	265	595	131	403	58	861	676	668
Percentage	3.2	11.2	2.0	13.7	8.6	10.1	3.7	8.3	1.8	5.6	0.8	12.1	9.5	9.4

Figure 5.4: Clustering and UMAP of scRNA-seq data of the caudal half of E8.0 embryos

a The top 2000 most variable genes across all cells were calculated using Seurat FindVariableFeatures and plotted as red dots. Each dot represents a gene with average expression level and standard variance shown. The top 25 gene names are shown. **b** An elbow plot showing the contribution of each principle component (PC) to the standard deviation. The red dotted line at PC30 illustrates the number of PCs used for clustering and projection into low dimensional space. **c** Uniform manifold approximation and projection (UMAP) plot showing all cells (7,134 cells). Cells are coloured by cluster and labelled with their cell-type annotation. **d** Table displaying the number of cells assigned to each cluster. Cell types are coloured according to the UMAP in **c**. ExE, extraembryonic; HEP, haemato-endothelial progenitor; Def., definitive.

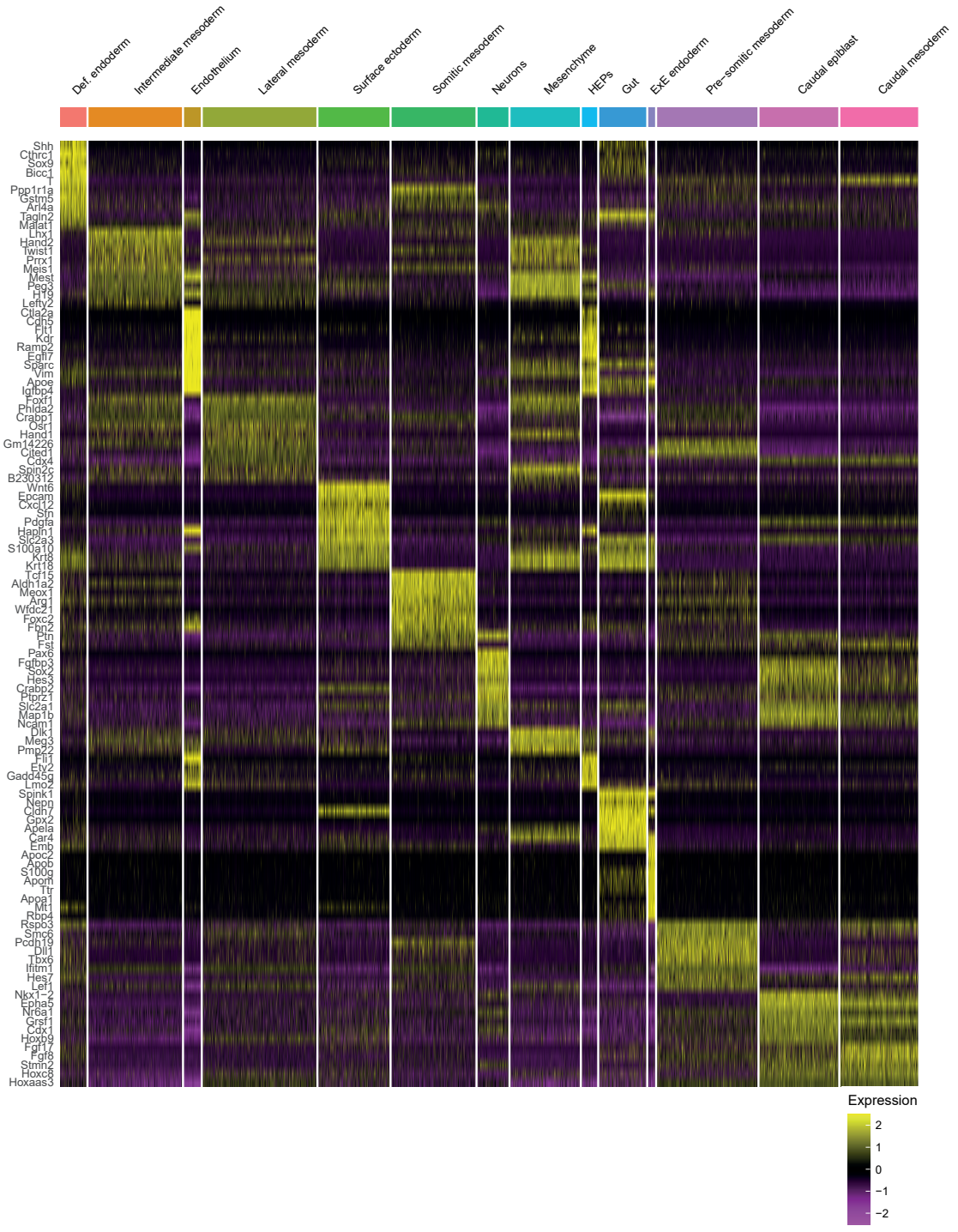


Figure 5.5: Gene expression matrix for the top ten enriched genes for each cell type of the caudal half of E8.0 embryos

A gene expression matrix with each column as a single cell and each row as a gene. The top ten enriched genes for each cluster was calculated with Seurat FindAllMarkers function, plotted as row normalised z scores. Cells are grouped by cluster and identifiable by the colour (matching Fig.5.4c) and label given above.

These clusters share expression of many enriched genes. However, the caudal epiblast has significantly higher *Nkx1-2* [346] and the Drosophila caudal homologue *Cdx1* [33, 347] and lower *T* expression as compared to the caudal mesoderm. Also the caudal mesoderm had higher expression of *Fgf17* and *Fgf8* as well as *Tbx6* and *Dll1*.

As observed on FISH stained sections, the HEPs and endothelial cells had the most distinct transcriptional signature of all mesodermal clusters. Both express the Vegf receptors *Kdr* and *Flt1* with endothelium being enriched for *Cdh5* expression but depleted in *Etv2*. With the mesoderm and endoderm clusters annotated, the remaining clusters were of ectodermal origin. Neurons of the spinal cord had an expression pattern in enriched genes similar to that of caudal epiblast expression with high levels of *Sox2, Hes3* and *Ncam1*. Although similar in expression levels of these genes, they are likely performing different roles as exemplified by the role of Sox2 in both pluripotency and neuronal differentiation [348, 349]. Additionally, this cluster had much higher expression of the neuronal transcription factor *Pax6*, a key determinant in neuronal fate acquisition [275, 350]. In contrast, surface ectoderm exhibited a very distinct set of enriched genes (Fig.5.5). For example, *Wnt6* is very specifically expressed in this cluster as observed in surface ectoderm *in situ* stainings [351, 352] and Krt18 highly enriched in PSC differentiation derived surface ectoderm [353].

Confident with cell type annotations, we next aimed to remove doublets. These are common place and particularly frequent in droplet-based scRNA-seq as with 10x Chromium. Fortunately, species mixing and cell hashing studies have determined the doublet frequency to be predictable based upon the number of cells loaded. Without cell hashing, doublets can only be predicted based on their gene expression. Homotypic doublets (two cells of the same transcriptome/type) are very difficult to identify but are unlikely to affect my analysis of cell-cell communication. Conversely, heterotypic doublets could feasibly introduce problems with accurate calling of ligands and receptors from cell clusters and so should be removed. Prediction of cell doublets can be performed through computationally creating doublets from scRNA-seq data and using this model to predict

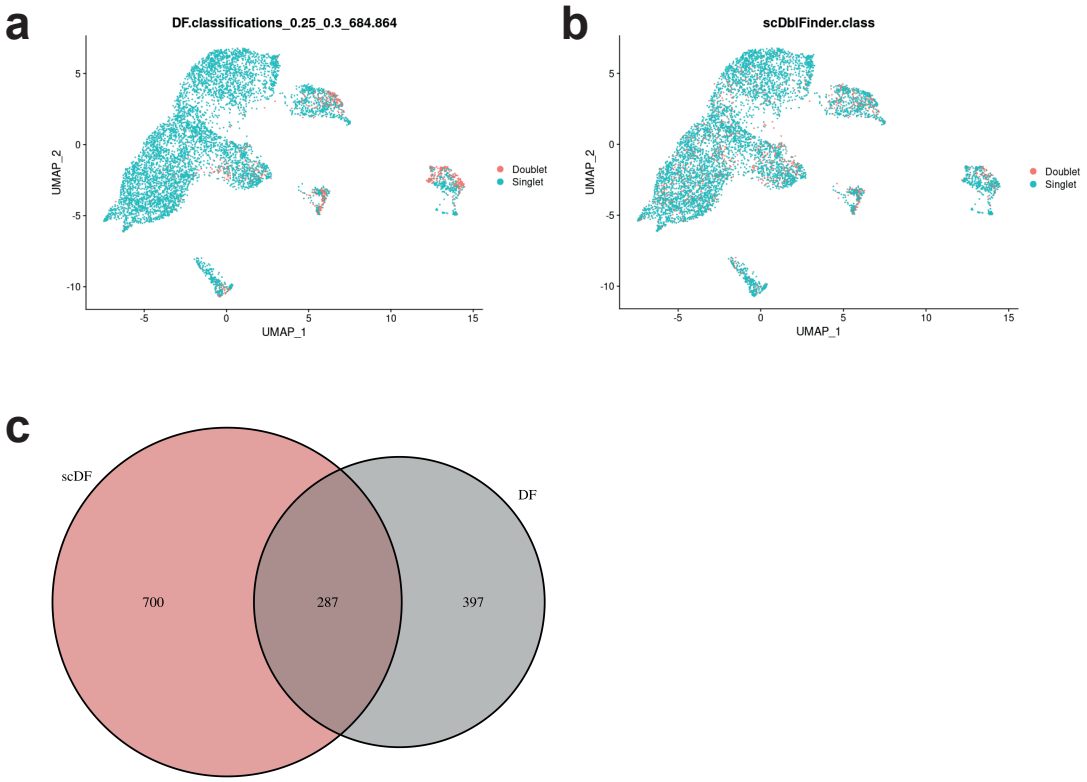


Figure 5.6: Identification of doublets with DoubletFinder and scDblFinder
Doublets were identified by two independent algorithms and integrated for removal. **a** and **b** show UMAP of cells as in Fig.5.4c coloured by singlet-doublet classification by DoubletFinder (DF) [354] and scDblFinder (scDF) [355] respectively. **c** Venn diagram of the cells classified as doublets by scDF and DF.

the probability of each cell being a doublet. Doublet probabilities are then thresholded at a value that gives the number of expected doublets based on the number of cells loaded. Since these models are only predictors, we wanted to integrate information from two doublet identification programmes: DoubletFinder (DF)[354] and scDblFinder (scDF)[355]. These programmes were selected based on their relative accuracy [356] and discussion with experienced colleagues. Both programmes predicted doublets scattered around each cluster with higher proportions in the Gut, Def. endoderm and Surface ectoderm. Although requested to predict the same proportion of doublets, DF (987) predicted notably more doublets than scDF (684). Surprisingly, overlapping the outputs of both programmes saw only a small percentage agreement between the clusters. We opted to remove only the cells identified as doublets in both programmes. These are the high confidence doublets while the others are more likely to be singlets. Since CellChat analysis take a cluster-average of ligand and receptor expression, any remaining doublets should have minimal impact.

5.2.4 Cell communication patterns received by HEPs in the caudal half of E8.0 embryos

Our initial plan was to perform cell communication analyses from spatially resolved transcriptomic data. However, given the issues with RNA degradation during LCM-seq, we adapted our workflow to use our scRNA-seq data to predict cell-cell communication. Identified cell signalling will then be functionally validated by a number of methods. Rather than use all cell clusters identified in scRNA-seq data, we filtered cells constituting only clusters identified being in close association with HEPs by SGA (lateral mesoderm, intermediate mesoderm, pre-somitic mesoderm, surface ectoderm and def. endoderm/gut)(Fig.5.7a, Fig.4.14). These clusters were input into CellChat for cell-cell communication analysis [262].

Looking broadly at cell-cell communication, multiple signals were predicted to be sent and received from each cluster (Fig.5.7b). In addition to the identification of autocrine signalling predicted within each cluster (represented by arrows originating and

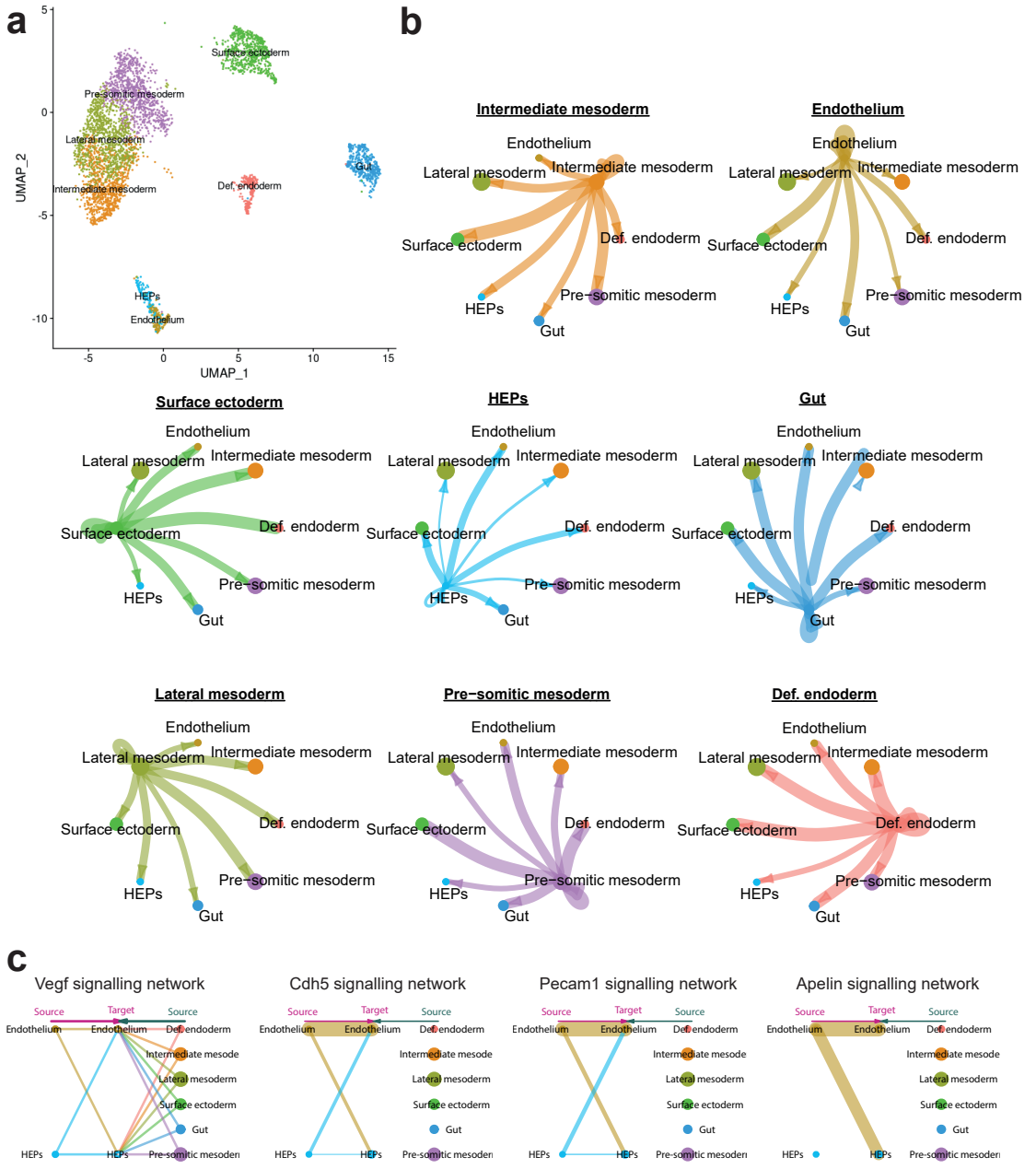


Figure 5.7: CellChat predicts a complex cell signalling network
a Uniform manifold approximation and projection (UMAP) plot showing cells selected for CellChat analysis. Cells are coloured by cluster and labelled with their cell-type annotation. **b** Circle plots showing the predicted interaction strengths between each cluster. Each plot is separated by the sender cluster labelled above and coloured according to the UMAP in **a**. Line width is proportional to the number of signals sent between the two clusters. **c** Cell signalling network for Vegf, Cdh5, Pecam1 and Apelin showing sender clusters on the left and right with the target clusters (endothelium top, HEPs bottom) in the centre. Line width is proportional to the strength of signalling and circle is portional to cluster size (cell number).

pointing to same cluster), all clusters were predicted to interact with each other to some extent. The widespread prediction of signals highlighted the importance of filtering signals in the selection of signals for validation. Since we are focusing on haemato-endothelial development, only those predicted to be received by HEPs or endothelium were studied. However, this dataset and analysis could also study interactions between non-haemato-endothelial tissues. While no one cluster dominated as sending signals, HEPs were predicted to send relatively fewer signals than other clusters. Conversely, endothelium as a source had relatively wide interaction lines suggesting an upregulation of signalling molecules as HEPs mature into endothelium. Both HEPs and endothelium receives signals from all clusters.

When manually searching the list of predicted signals, a number of known cell signalling pathways critical to haemato-endothelial development can be identified (Fig.5.7c). This includes Vegf signalling that is predicted to originate from all clusters including HEP/endothelium, the endothelial adhesion molecules Cdh5 and Pecam1 and Apelin signalling. The identification of these well characterised signalling molecules provides reassurance that this pipline is able to predict signalling pathways with functions on haemato-endothelial specification and function.

To gain a better understanding of the signals that HEPs/endothelium are receiving during this critical stage of their development, we organised signalling pathways into patterns using pattern recognition. These patterns can help determine how multiple cell types and signalling pathways can combine to impose function [262]. The number of patterns can be determined through the calculation of Cophenetic and Silhouette values which both measure the stability for a particular number of patterns based on a hierarchical clustering of the consensus matrix [357]. Measurement of these values with increasing numbers of patterns typically leads to a sudden drop when redundantly splitting patterns. Plotting these values for incoming signalling patterns revealed that there four patterns of incoming signalling (Fig.5.8a).

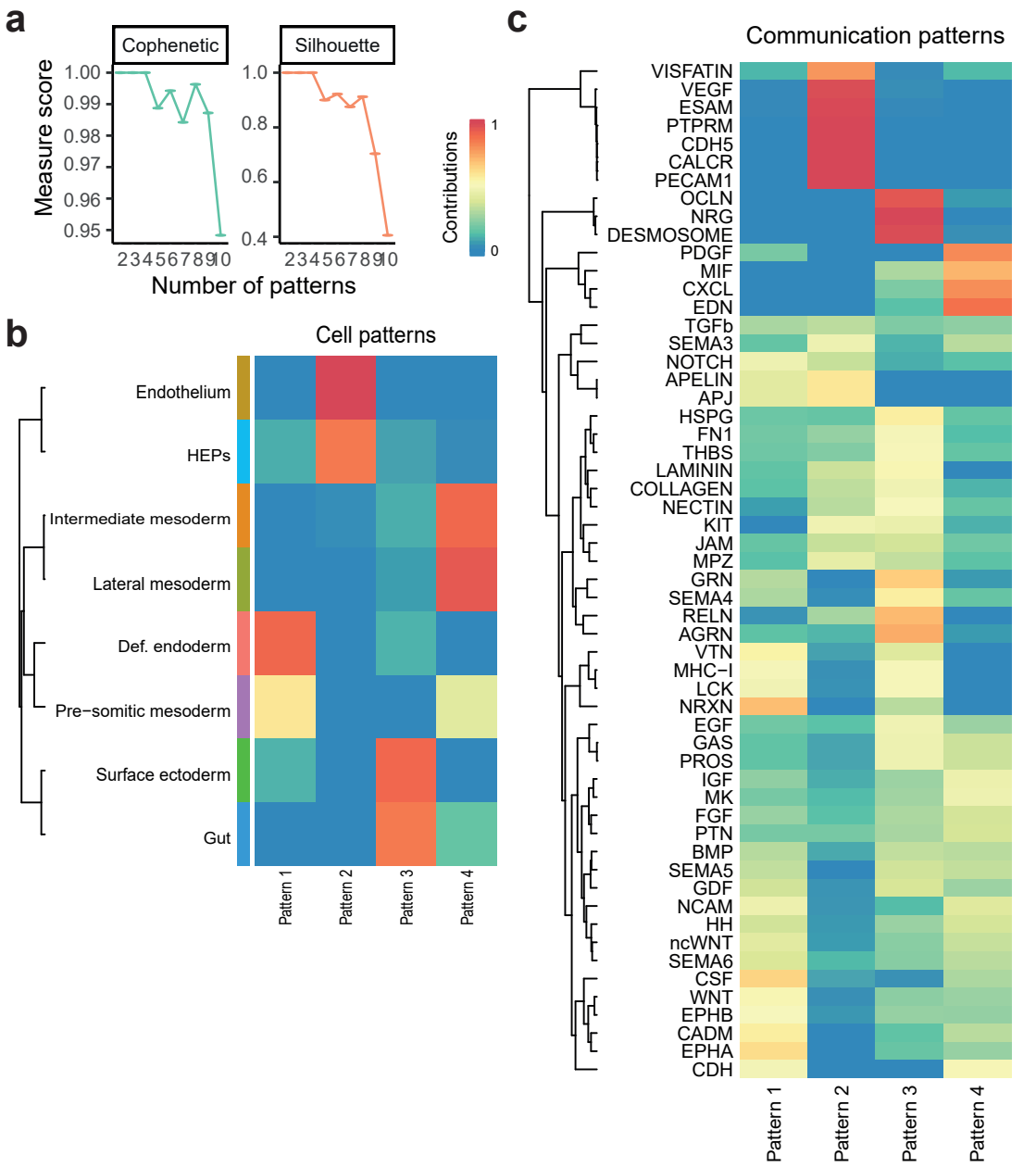


Figure 5.8: Identification of incoming cell signalling patterns
a Assessment of the Cophenetic and Silhouette values with increasing numbers of incoming patterns to determine the number of incoming patterns present amongst cell signalling networks. **b** A heatmap showing the contribution (red is high, blue is low) of each cell type to the four identified incoming signalling patterns. Cell type is identifiable by label and cluster colour. A dendrogram on the left hand side shows the result of hierarchical clustering of cell types. **c** A heatmap showing the top contributing signals to each incoming pattern of signalling. Each column is a pattern and each row a class of cell signalling. A dendrogram on the left shows the result of hierarchical clustering.

Examining the contribution of each cell type to these patterns revealed that HEPs and endothelium constituted their own pattern (pattern 2) that clustered far apart from the other clusters (Fig.5.8b). The lateral and intermediate mesoderm cells also formed a distinct pattern (pattern 4) with the final mesodermal cluster forming pattern 1 with def. endoderm. Surprisingly, def. endoderm and gut did not form a pattern together and clustering separated them by a significant distance (Fig.5.8b). This suggests that maturation of def. endoderm into gut changes the receptor profile considerably. When examining the signalling pathways that contribute to these patterns, pattern 2 is enriched for many well characterised endothelial signals (Vegf, Esam, Cdh5, Pecam1) (Fig.5.8c). None of the enriched signals in pattern 2 are novel signals that require validation and so a more extensive search of signalling was performed.

5.2.5 Cell signalling pathways recieved by the HEPs in the caudal half of E8.0 embryos

Signals were filtered on a number of criteria aimed to select signals previously unidentified in haematoendothelial development at this stage of development. Notably, the filtering parameters used here are not designed to keep all potentially interesting signals but rather select those most likely to direct haemato-endothelial fate. Therefore, it is likely some of the signals excluded from downstream analysis could be functional. Firstly, signals that are well characterised during HEP development were removed. Secondly, signals widely received by other (>2) clusters other than angioblasts and endothelium were removed. While the function of cell signalling is highly context-dependent and therefore could function differently in contrasting receiver cells, those received by many cell types are more likely to have broader applications and are less likely to influence the specification of blood fate. Thirdly, the remaining signals were reviewed in the literature for known function in other biological contexts, for example, in tumour endothelium or fate specification in other tissues. This allowed us to construct a list of signals that are enriched in signalling to the haemato-endothelial lineage at this time and classify them according to their novelty, expression and selection for validation (5.9).

Group	Ligand(Receptor)	Sender Count	Sender					
			IM	Lateral	Pre-Som	Ectoderm	Gut	Endoderm
Collagen	Col9a1(Itga9/1+Itgb1)	4	✓	✓	✓	-	-	✓
	Col4a6(Itga9/1+Itgb1)	4	✓	-	-	✓	✓	✓
	Col4a5(Itga9/1+Itgb1)	5	✓	-	✓	✓	✓	✓
	Col4a2(Itga9/1+Itgb1)	6	✓	✓	✓	✓	✓	✓
	Col4a1(Itga9/1+Itgb1)	5	✓	-	✓	✓	✓	✓
	Col2a1(Itga9/1+Itgb1)	2	-	-	-	-	✓	✓
	Col1a2(Itga9/1+Itgb1)	3	✓	-	✓	-	-	✓
	Col1a1(Itga9/1+Itgb1)	1	✓	-	-	-	-	-
Laminin	Lama1(Itga9/1+Itgb1)	4	✓	-	✓	-	✓	✓
	Lama5(Itga9/1+Itgb1)	5	✓	-	✓	✓	✓	✓
	Lamb1(Itga9/1+Itgb1)	6	✓	✓	✓	✓	✓	✓
	Lamc1(Itga9/1+Itgb1)	6	✓	✓	✓	✓	✓	✓
Vegf	Vegfa(Vegfr1/2)	6	✓	✓	✓	✓	✓	✓
	Vegfb(Vegfr1/2)	6	✓	✓	✓	✓	✓	✓
	Vegfc(Vegfr3)	3	✓	-	✓	-	-	✓
	Pgf (Vegfr1)	3	✓	✓	✓	-	-	-
Wnt	Wnt5a(Fzd4)	4	✓	✓	✓	-	-	✓
	Wnt5a(Mcam)	6	✓	✓	✓	✓	✓	✓
	Wnt3a(Fzd4+Lrp6)	2	-	-	✓	-	-	✓
	Wnt6(Fzd4+Lrp6)	3	-	-	-	✓	✓	✓
	Bmp7 (Acvr1 + Bmpr2)	4	-	-	✓	✓	✓	✓
	Sema3e(Plxnd1)	2	-	-	-	✓	-	✓
	Thbs3(Cd47)	3	✓	-	✓	-	-	✓
	Calca(Calcrl)	2	-	-	-	-	✓	✓
	Apela(Aplnr)	4	✓	✓	-	-	✓	✓
	Dlk1 (Notch1)	5	✓	✓	✓	✓	✓	-
	Fn1(Itga5+Itgb1)	6	✓	✓	✓	✓	✓	✓

Figure 5.9: Summary of enriched signalling to haemato-endothelial cells
A table summarising CellChat prediction to the haemato-endothelial lineage (HEPs and endothelium). When applicable signalling has been grouped into classes. Ligand and receptor pairs for validation are highlighted in blue. Adjacent to the signalling molecules are the sender cell types with a green ticks indicating a sender and red minus indicating no signalling detected. The number of sender tissues is shown in the Sender Count column.

This next section summarises the results of my literature search to select cell signalling pathways for validation, the plans for which are summarised in the discussion of this chapter. Among all the studied sender cells, groups of signals that were commonly predicted (Collagen, Laminin, Vegf and Wnt). Collagen and laminin are components of the extracellular matrix and signal to haemato-endothelial cells through their expression of 3 select intergrins : *Itga9, Itga1* and *Itgb1*. Integrins are adhesion molecules that form dimers consisting of α and β subunits. Lineage specific deletion of Itgab1 (predicted here to dimerise with Itga9 and Itga1) with a Tie2-Cre demonstrated that endothelium is specified but was disorganised with discontinuous vessels [358]. While Wave 3 haematopoiesis was not assessed in these mice, qualitative assessment of the YS blood islands showed no impairment. Additionally, endothelial cells cultured *in vitro* show differences in motility and angiogenesis when cultured on collagen and laminin but not fibronectin suggesting these are indeed the critical mediators of signal transduction in endothelial cells [358]. These data validate the enriched interaction between integrins and the cell matrix components collagen and laminin at this stage of development.

Interestingly, cells clusters did not all express the same profile or laminin and collagen molecules. While some ECM components were broadly expressed (e.g. *Col4a2* and all laminins), others were expressed quite specifically in select clusters (e.g. intermediate mesoderm being the only cluster to express *Col1a1* and*Col2a1* only expressed by def. endoderm/gut). While different collagens are known to have site-specific functions in adults [359], whether location-specific deposition of different collagen molecules during development could introduce heterogeneity is unknown. This is an enticing prospect as it could form a basis for heterogeneity between the dorsal and ventral aspect of the mesoderm with def.endoderm/gut expressing different collagen components to the dorsal cells.

In addition to integrin signalling, other previously studied signals were identified. As discussed above, Vegf signalling was easily identifiable as an enriched signalling group (Fig.5.9). Three of the Vegf ligand family were identified (A, B and C) as signalling

through all three Vegf receptors. As also seen in our FISH data in Chapter 4, Vegfa is very broadly expressed in this region of the embryo at this stage with all cell types identified as being a source of Vegfa. Alternatively, Vegfb is not expressed in lateral mesoderm, ectoderm or gut. In contrast to Vegfa, Vegfb does not have a direct impact on endothelial growth and sprouting and is thought to instead enhance Vegfa function [360]. Indeed, genetic deletion of Vegfb in mice is non-lethal and only impacts heart function but has no consequence on the development of the cardiovascular system [361, 362]. Finally, Vegfc, thought to primarily function as a pro-lymphangioenic factor, is predicted to signal from the fewest cell clusters (intermediate mesoderm, pre-somitic mesoderm and def. endoderm). Given how well studied these factors are in vascular development, we would focus on other signals predicted in our pipeline.

The third family of signalling predicted was the Wnt family (Fig.5.9). A number of Wnt molecules were predicted to be signalling to the HEP/endothelium clusters. Many of these Wnt molecules were predicted to interact with the Fzd4 and Lrp6 heterodimeric receptor. Fzd4 is well studied in its function to promote endothelial growth and stability during embryogenesis through β-catenin signalling [193]. Tie2-Cre mediated deletion of β-catenin displayed similar endothelial patterning defects [194]. These studies highlight the critical nature of canonical Wnt signalling during this time in vasculogenesis. Therefore, all the Wnt ligands predicted to engage with Fzd4 and Lrp6 do not match our criteria for further study here. However, Wnt5a sent from all cell clusters is predicted to interact with Mcam. Also known as CD146, Mcam is classically known for its role as a cell adhesion molecule in endothelial cells but has more recently been shown to be a co-receptor for Vegfr2 in tumour angiogenesis [363] and a receptor for Wnt5a [364, 365]. Through activation of non-canonincal Wnt signalling, Mcam increases endothelial motility and formation of protrusions *in vitro* [364]. The function of Mcam at 4ss of mouse embryogenesis is unknown and could lend a mechanism by which endothelial cells and their progenitors migrate to the site of vessel formation. Morpholino mediated knockdown in zebrafish embryos demonstrated a requirement of Mcam during gastrulation so validation in haemato-endothelial development would likely

require a lineage-specific (e.g. Tie2-Cre) knockout.

The remaining signalling pathways could not be grouped together into families. There were however, still well characterised signalling pathways. For example Apela, discussed above, and Notch (Dlk1-Notch1). Notch signalling is well studied because of its critical role in arterialisation of vessels and its requirement for definitive haematopoiesis. The expression of Notch1 in the endothelium at this time, just before Runx1 is detectable in the embryo is crucial as Notch signalling is at least partly responsible for the activation of *Runx1* transcription [167, 190]. What remains to be understood is the other signalling pathways that selectively activate Runx1 in only a subset of cells.

BMP signalling during embryonic haematopoiesis has also been focussed on during the window of EHT between E9.5 and E11.5 [198–200]. These investigations focus on how Bmp signalling and its inhibition are balanced during the emergence of HSCs. It is interesting therefore that we see Bmp7 also signalling to the haemato-endothelial linage long before the presence of haemogenic endothelium in the DA. Bmp7 knock out mice have not been reported to have blood or endothelial defects but die shortly after birth with severe eye [366], kidney [366], skeletal patterning [367] and corticogenic [368] defects. However, we are unaware of any detailed analysis focussing on Wave 3 haematopoiesis so it is feasible thatthere are indeed defects.

One of the prime candidates for further validation is the receptor Plxnd1. Here it is predicted to act as a receptor for Sema3e which has been widely described as being anti-angiogenenic by inhibiting Vegf signalling, disrupting cell adhesion and decreasing migration [369, 370]. The genetic deletion of these signalling components do not inhibit vasculogenesis but do have impacts on intersomitic vessel sprouting and patterning [370]. However, Plxnd1 also exhibits a Sema3e independent function as a mechanosensor [371] in endothelial cells. This function could be of interest as mechanical forces promote the emergence of HSCs [372–375]. One mechanosensor, Rho-GTPase was shown to act through YAP in human aorta culture to regulate haematopoietic stem cell formation.

The presence of other mechanosensors in blood stem cell emergence is unclear but Plxnd1 could be a good candidate.

Finally, this left the two most promising candidates to take forward for validation. Firstly was the thrombospondin-1 from intermediate mesoderm, pre-somitic mesoderm and def. endoderm signalling to Cd47. Thrombospondin has been studied for its inhibitory effect on angiogenesis but not during vasculogenesis. In contrast, CD47 has been identified in embryonic haematopoiesis as enriching for HSC-competent endothelial cells in the E10.0 mouse AGM [126]. Zhou *et. al.,* demonstrated that HSC-competent endothelium, pre-HSCs and mature HSCs are all contained with the $CD47^{+}$ fraction. As with many markers used to phenotypically define populations, their function is often not considered as crucial and so the function of CD47 has not been investigated further. However, with the prediction of thrombospondin signalling to CD47, this could provide a possible explanation.

Studies focusing on thrombospondin signalling through CD47 in *in vitro* cultured endothelium showed it inhibited the cell cycle and induced senescence [376] and is a negative regulator of angiogenesis [377]. This function could be linked to HSC emergence through considering the cell cycle changes that might be required during EHT. Indeed hESC-derived HE must tightly regulate its cell cycle to undergo EHT and is slow cycling or even quiescent as compared to non-haemogenic endothelium [378]. Whether CD47 could mediate this required transition to slow-cycling and allow EHT in a subset of endothelium is a question that requires further investigation.

Secondly, the calcitonin gene related peptide exclusively derived from the def. endoderm and gut is predicted to signal through the calcitonin receptor like receptor expressed on endothelium. This signalling pathway is understudied during development and little is known about its function. *Calcrl*$^{-/-}$ mice die *in utero* by E14.5 with severe peritoneal oedema, thin vascular smooth muscle walls, small hearts, with increased cell death [379]. There are no reports of endothelial or haematopoietic defects. However,

given the restricted expression of the ligand in a selection of def. endoderm/gut cells it is an attractive candidate to induce heterogeneity in dorsal versus ventrally located HEPs/endothelium. Therefore, we would also like to follow this pathway up with validation and subsequently evaluate its influence on Wave 3 embryonic haematopoiesis.

5.3 Discussion

The aim of this chapter was to predict cell-cell communication between angioblasts and their cellular neighbours identified in Chapter 4. Ideally, the predictions would be performed on spatially-resolved data to increase confidence in the predicted signals being functional. Retaining information on the location of cells within the embryo and, relative to each other significantly improves any cell communication analysis. It allows the calculation of cell-cell interactions between neighbouring cells rather than the averaging of clusters as necessary in CellChat. Not only would this reduce the number of false positive predictions but it also allows the identification of signalling that may be different between the same cell type in different locations, which could certainly be the case for HEPs in the caudal half of E8.0 embryos. In Chapter 4 we identified that HEPs have distinct cellular neighbours (Fig.4.14). Whether this difference in cellular neighbourhood translates to differences in cell signalling remains unknown. Additionally, any subtle transcriptomic differences between differently located HEPs (e.g. dorsal versus ventral) could give insight into their different fates at this critical time in development [218]. To this end, we aimed to establish LCM-seq that combines cutting, catapulting and capturing cells of interest using LCM with Smart-seq3 transcriptomics. The unexpected behaviour of the Etv2-Cre x mTmG embryos labelling all cells with both red and green fluorescence meant rapid staining with fluorescently labelled WGA and SCL immunolabelling was used to define cell borders and label angioblasts respectively. Single cells could be reliably cut, catapulted and captured using LCM with RNA successfully released during cell lysis and subsequently reverse transcribed to cDNA. Unfortunately, cDNA library size was notably smaller than expected, suggesting significant RNA degradation.

Therefore, the first steps going forward must be to improve sample preparation conditions. Cutting unstained sections suffered from the same problems as stained sections, therefore staining can be eliminated as the cause of RNA degradation for now. Dissecting embryos in media (e.g. M2-media) could be tried instead of PBS containing serum, as serum contains high quantities of RNases. This is not a problem for RNA-FISH as embryos are quickly PFA fixed and dehydrated, removing any possible RNase activity derived from the serum. However, since we require fresh-frozen tissue for LCM-seq, RNases could potentially remain and degrade RNA within sections during processing. In parallel, snapfreezing embryos in their decidua by plunging directly into liquid nitrogen before embedding in OCT should be tried to enhance RNA integrity. This is common practice for brain tissue, and would limit the time during which embryos are vulnerable. Typically, embryos are embedded and subsequently stored in the -80before sectioning. An additional optimisation could be performing all the steps in rapid succession, with minimal time in storage, and could be further optimised by laser microdissecting cells immediately after sectioning. Sectioning is already performed in the cleanest possible conditions, including the insertion of new blades and thorough anti-RNase treatment; this step cannot be optimised further. Although not currently available in our facility, one final enhancement could be to use a LCM-microscope in a high-level fumigation hood.

Hopefully, these practical steps would improve RNA quality and allow the production of full-length cDNA libraries. However, alternative options include trying to generate cDNA libraries from PFA fixed tissue. This has the additional benefit of improving sectioning stability, and would increase the number of sections collected per embryo. Although difficult, an increasing number of protocols are under development to prepare RNA-seq libraries from fixed tissue. Alternatively, protocols developed to sequence cDNA libraries from degraded RNA could be implemented, such as SMART3-seq [380]. However, this may reduce sensitivity with typical detection of only 2000-3000 genes per cell [380]. This would likely severely impede the discovery of cell-cell communication which requires the sensitive detection of receptors and ligands in receiver and sender cells, respectively. However, it could serve as spatial anchor points for the remapping of

scRNA-seq data.

I have generated a scRNA-seq dataset from the caudal half of the 4ss embryo, complimenting our multiplexed dataset from Chapter 4. Our approach provides confidence that we have enriched for the HEPs and endothelium contributing to Wave 3 haematopoiesis by selectively dissecting the region known to contain definitive haematopoietic progenitors [173]. Cell signalling identification was performed using CellChat. Numerous cell signals recieved by HEPs and endothelium have been predicted and one major task has been to appropriately filter the signals. We opted for a filtering strategy which would maximise the chance of identifying cell signalling with an unknown role in haemato-endothelial specification. CellChat predicts any interaction that is significantly enriched between two clusters [262], therefore the predictions discussed can be described as being in one of three classes of signalling. The first class consists of communication patterns that are known to specify the haemato-endothelial lineage, validating the workflow of ligand-receptor analysis. The second and third classes of predicted signals consist of true positives and false positives, where true positives have a functional effect on haemato-endothelial specification and false positives do not. False positives can arise in a few ways. Firstly, CellChat performs these predictions using transcriptomic data which does not necessarily correlate with functional, protein levels. Secondly, my analysis also does not account for the actions of receptor antagonists or a ligand threshold requirement, whereby the ligand must be present above a certain level to be functional. Therefore a robust, multi-layered validation process is crucial to first shortlist signals for validation of expression and then functional analysis.

Firstly, signals were screened for their predicted signalling to all non-HEP cells. Those signals predicted to specifically target HEPs were prioritised and those with broad activity discarded to reduce the number of signals universally required for essential cellular processes. Secondly, I performed a literature search to identify signals known to impact haemato-endothelial development. Those that are well described (e.g. Vegfa) were noted as validation of our analysis pipeline but excluded from the shortlist of novel

signals. It should be noted that many of these signalling pathways (e.g. Wnt, Bmp and Notch) have been characterised in haemato-endothelial development during HSC emergence between E9.5 and E10.5. It is clear however, through studying the master regulator of blood-fate specification, SCL, that haematopoietic fate is specified earlier, between E7.5-E8.5 [150, 218, 381]. My data indicate these signals are functioning at this earlier stage, before Wave 3 HE can be identified. Resolving the signals that are required for blood fate specification versus its emergence is challenging and highlights the importance of work characterising the order of molecular events underlying specification and commitment.

CellChat also identified signals to HEPs that are well studied in development but don't report a haematoendothelial phenotype. This could be because of three reasons. Firstly, they are false negatives as discussed above. Secondly, they could signal to haemato-endothelial progenitors but have a redundant role, compensated for by another signal when perturbed. Finally however, there are signals that are well studied in other contexts and do not describe HSC phenotypes upon perturbation because of a primary phenotype either masking or drawing attention away from a haematopoietic phenotype. This could be particularly pervasive in cell signalling which impacts endothelial remodelling where the cell types develop so closely.

Future work will focus on achieving two aims. Firstly, validating the expression of signalling components predicted by CellChat, including confirmation of RNA expression and spatial mapping of cells expressing each ligand and receptor. We will initially perform HCR-FISH on embryonic sections, which also limits the need for antibody identification. Upon HCR-FISH confirmation that both HEPs/endothelium express receptors and neighbouring cells express the cognate ligand, we would then validate at a protein level by immunostaining where possible. We have shortlisted a subset of ligands and receptors for expression validation (highlighted in blue in Fig.5.9). These selected ligand-receptor pairings represent five classes of signalling: BMP, WNT, CALCA, Thrombospondin and Semaphorin. The top candidates for validation and further functional study include

Thrombospondin-CD47 and Calcrl-Calcrl signalling pathways, due to their limited description in haemato-endothelial development. Some functional validation could be done using *in vitro* differentiations of mESC cells. For example, after validating CD47 expression on endothelium and haematopoietic cells, thrombospondin could be added to cultures to assess the effect on EHT. This would provide a system in which to easily study the effects on cell cycle for example. However, this is not a perfect system as it does not recapitulate Wave 3 haematopoiesis so we would also like to perform some functional validation *in vitro*. For CD47 we might compare the cell cycle of CD47+ versus CD47- cells to test our hypothesis of its regulation of cell cycle and try to perturb signalling and measure the effect on HSC number and location.

Chapter 6

Final conclusion and Discussion

Contents

6.1 Summary of key findings 163
6.2 hPSC derived endothelium and the possibility of all young endothelium having blood-forming potential 165
6.3 Using intermediate mesoderm A-P patterning to inform the signals required for intraembryonic HEP regionalisation 166
6.4 Prediction of the early signals to intra-embryonic haemato-endothelial progenitors 167
6.5 Application of newly developed technologies to understanding haemato-endothelial development 167

6.1 Summary of key findings

This work is driven by wanting to understand how HSCs develop during embryogenesis, to inform the recapitulation of that process *in vitro*. Comparing existing hPSC differentiation protocols to our current understanding of HSC development *in vivo*, we identified the possibility that current protocols had no way of determining the signalling required to pattern the restricted region of LPM required for HSC development [129]. We identified the mesonephric duct as being patterned in the same restricted region as Wave 3 precursors [129] and hypothesised we could use this as a way to more faithfully recapitulate the signals required to specify this region of LPM. Using a protocol that produces anterior intermediate mesoderm and its derivative the nephric duct [77], we identified KDR+ mesoderm specified using the same mesodermal patterning conditions. Multi-modal measurement of differentiation allowed us to characterise the stepwise differentiation of this population to arterial endothelium after adding VEGFA and bFGF. Unlike previous studies, we aimed to induce a haemogenic programme in this arterial endothelium, more accurately reflecting *in vivo* progression. Despite some initial success,

a haemogenic programme could not be reproducibly inducted using three commonly used culture conditions individually or in combination. This drove us to better understand the signals that drive the induction of a haemogenic programme in Wave 3 HEPs.

We selected to focus on the distribution of HEPs in the posterior of the embryo at E8.0 as it contains Wave 3 progenitors [173], sits in the centre of the window for blood fate specification (as defined by Scl expression and Tek-Cre mediated KO), and is just before *Runx1* expression is first initiated within the embryo [382]. Previous work at this time typically used single colour ISH or IF on multiple embryos to understand the distribution of HEPs [155, 251]. Using the newly established SGA, we described the expression of many of these key marker and specification genes in the same cell and correlate this with the cell's location within the embryo. This allowed us to identify at least three HEP locations at this time (ventromedial, ventrolateral and dorsolateral). The gene expression patterns of HEPs in each of these locations differed only mildly based on the genes we probed, with a trend towards lower endothelial expression (*Cdh5,Tek*) and higher progenitor gene (*Etv2, Tal1*) in the lateral HEPs as compared with the medial cells (Fig.4.15). Quantification and remapping of all cell types did, however, reveal variation in cellular neighbourhoods between and within these locations.

To identify how different cellular neighbours could impact on cellular signalling and gene expression patterns beyond those probed by SGA, we aimed to integrate the spatially resolved transcriptomics with high throughput scRNA-seq to predict location specific signalling. As the section morphology and RNA integrity were significantly compromised using our current protocol, we utilised the ligand-receptor analysis package CellChat on scRNA-seq data alone. This has identified that angioblasts and endothelium in the E8.0 posterior receive their own pattern of cell signalling as compared to the other mesoderm subsets and are predicted to integrate 50 signals, of which 24 are unique to HEPs. The majority of these signals are already implicated in haemato-endothelial development but their stage-specific roles are yet to be defined. Finally, we have identified two novel signals that, to our knowledge, have not been functionally studied in

haemato-endothelial development. The signalling of thrombospondin from the IM, PSM and defintive endoderm to CD47 on angioblasts and endothelium represents an exciting candidate for validation. CD47 has been defined as a marker for HSC-competent HE at E10.5 [321] but not investigated at earlier stages of development. Without the availability of a CD47 fate mapping line, it would be difficult to test this. However, thrombospondin signalling through CD47 in endothelium has been shown to influence cell cycle progression [376], a previously identified driver of EHT [378]. Secondly, the signalling of calcitonin gene related peptide from the Def. endoderm to Calcrl represents another candidate for validation. Although no haemato-endothelial defects were reported in *Calcrl*$^{-/-}$ mice, the restricted source of expression from the def.endoderm represents an attractive candidate for induction of dorsal-ventral heterogeneity in differently located HEPs. In conclusion, the L-R analysis from scRNA-seq data alone has provided a global picture of the summary received by HEPs in the E8.0 posterior and has highlighted some exciting candidates for validation. Future work establishing LCM-seq will provide higher spatial resolution than just associating clusters with SGA data and could identify differentially expressed genes between spatially segregated HEPs. Furthermore, studying the regulation of the expression receptors in HEPs through transcription factor binding for example, could provide insight into the transcriptional control of haemato-endothelial fate specfication.

6.2 hPSC derived endothelium and the possibility of all young endothelium having blood-forming potential

Following studies reporting the presence of HE in sites other than the ventral wall of the DA such as the placenta [383], head [384], endocardium [385], vitelline and umbilical arteries [121, 386], Swiers, *et. al.*, 2013 [387] reviewed the suggestion made 100 years earlier that haemogenic competency is a property of all young endothelium [388, 389]. At the time, the *de novo* generation of definitive haematopoietic cells from the endocardium and head vasculature was still questionable [387]. However, more recent live imaging of cells budding from the endocardium [142] indicates their haemogenic potential. These

studies show HE is widespread through embryonic and extraembryonic vessels and feeds into the idea of haemogenic competency being widespread in young endothelium. This is supported in part by the overexpression of *Runx1* specifying intraembryonic and YS endothelium as haemogenic between E7.5 and E8.5 only [390]. In this study, we have identified hPSC derived arterial endothelium that cannot differentiate into $CD45^{+}$ haematopoietic cells using the widely used culture conditions for inducing EHT (Fig.3.6). This indicates at the very least that hPSC derived endothelium are not equal in their ability to respond to haematopoietic signalling and could indicate that there are endothelial populations void of blood-forming potential. What grants this difference in responsiveness to haematopoietic cytokines in endothelium in our cultures versus others is unknown.

6.3 Using intermediate mesoderm A-P patterning to inform the signals required for intraembryonic HEP regionalisation

Using mesodermal patterning conditions that form the anterior IM of the mesonephric duct allowed us to generate arterial endothelium patterned in the same way. Although these endothelial cells could not be made haemogenic, it does not necessarily detract from the merit of the hypothesis. In particular, through studying the cellular neighbourhood of haematoendothelial progenitors we have confirmed that the IM forms in very close proximity with a subset of the HEPs and is particularly close with the DA (Fig.4.14). It likely therefore forms at least part complex signalling environment of the forming DA with 86 L-R interactions predicted using CellChat. Additionally, IMs A-P regionalisation could itself act as a source of signalling. This hypothesis should be tested *in vivo* and could be performed using LCM-seq with IM cells dissected at different A-P levels to test variability in signalling. However, detailed description of the DA during this crucial time in blood-fate specification has also demonstrated the niche is far more complex than we produced *in vitro* (Fig.3.3). This gives strength to the protocols that relinquish some control over differentiation in order to form multi-lineage embryoid bodies as they more accurately recapitulate the complexity we don't yet understand enough to replicate with

non-cellular components (cytokines, matrix, mechanical etc.).

6.4 Prediction of the early signals to intra-embryonic haemato-endothelial progenitors

It is notable that a large number of the signals predicted from our L-R analysis of the posterior of E8.0 embryos have been reported to influence HSC emergence later in development. While this is in part due to the large number of signals that have been characterised during this time, it also indicates that when performing perturbation, researchers should carefully assess when the signalling components are expressed. This is especially poignant when the readout is later (e.g. E10.5) as it is easy to conflate the time of measurement with the time of perturbation. The signals predicted by CellChat require validation, first through validation of the expression domains of signalling components followed by functional validation. Those confirmed to be expressed as predicted by CellChat could play a number of roles during this time in development. At this time, cells are undergoing arterialisation [155], will soon express Runx1 [382] and are undergoing Scl-controlled blood fate specification [218]. Uncoupling these processes will be difficult, with each programme sharing transcriptional and signalling mechanisms.

6.5 Application of newly developed technologies to understanding haemato-endothelial development

Our understanding of haemato-endothelial development has already been improved through the recent development of new technologies. scRNA-seq has allowed the computational reconstruction of developmental trajectories and identification of marker genes and signalling [33, 321]. What these techniques lack, however, is the ability to retain cell location. It is no wonder then that spatial transcriptomics is expanding and advancing at an incredible pace [391]. SGA has many strengths including very high sensitivity, relative ease of establishment and high spatial resolution. It does, however, have limitations that newer technologies can address and therefore provide greater biological insight. One of the greatest challenges in spatial transcriptomics is the

accurate assignment of transcripts to individual cells. This requires cell segmentation of varied cell shapes and sizes often from imperfect, noisy data. Machine learning approaches have greatly improved segmentation [392–394] with larger training datasets and new algorithms improving accuracy and generalisation of methods. The use of Ilastik with membrane staining greatly improved cell segmentation but significant user intervention was required for satisfactory segmentation. This level of user intervention is only possible for cases where relatively few cells are profiled and not appropriate for atlasing projects. Some of these issues originate from inconsistent membrane staining (e.g. medial paraxial mesoderm in embryo 3, Fig.4.14). This is problematic because even the most perfect segmentation algorithm could not define cell borders in this case. Therefore, there is an unmet need for membrane staining methods that can be used on fixed and permeabilised cells.

A major goal of chapter 3 was to direct differentiation away from the YS waves of haematopoiesis through the patterning of intraembryonic mesoderm. The importance of this distinction is still contentious, with some groups describing the production of HE with non-HSC (principally T-cell) definitive potential as a necessary step to *de novo* HSC production [217]. This school of thought rests on the assumption that the differentiation potential of HE is not determined by its lineage history but rather by the micro environment. Evidence from xenopus embryos demonstrates the primitive and definitive waves segregate very early in development [130]. This early segregation inevitably leads to different cellular signalling and programming. While lineage convergence could result in equally programmed HE with equal potential, my opinion is this is unlikely and that the path taken by HSC ancestors is crucial for their emergence. The concept of developmental origin being important for terminal cell fate is observed in the imprinting of lymphatic fate in a subset of paraxial mesoderm [66], and the different roles in tissue homeostasis and infection of YS and HSC derived macrophages [308]. A unified picture of HE heterogeneity is complicated by the change in HE population through developmental time in both the YS and HE and so would require multi-staged analyses. Although more challenging than comparing static populations, comparison of HSCs and

YS-derived haematopoietic progenitors remains a large gap in our understanding. The comparison of hPSC-derived populations to the AGM but not YS could conceivably lead to misleading conclusions as there will be many more transcriptional similarities than differences between the lineages. Finally, measuring the heterogeneity in cellular chromatin state with single cell epigenomic profiling could reveal underlying mechanisms by which different HE populations will have context dependent roles for shared TFs. This has recently been demonstrated with the ability of RUNX1 to confer haemogenic capacity to an endothelial cell being dependent on the accessibility of TGF-related genes [168]. Together, the application of these ground-breaking technologies can bring us closer to a unified view of HE heterogentity and its importance in HSC emergence.

Appendices

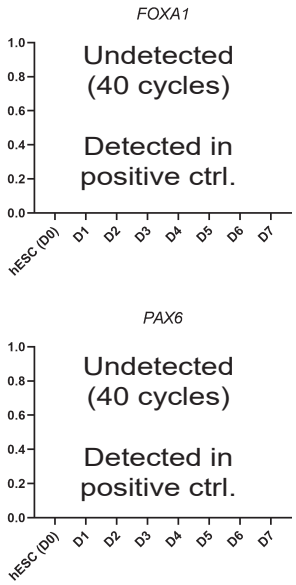


Figure 1: Additional genes profiled during Takasato protocol differentiation of hESCs RNA expression kinetics for a panel of genes of interest were monitored daily for the Takasato differentiation protocol. Whole cultures were taken for RNA extraction and cDNA synthesis. Gene expression is normalised to the average of three housekeeping genes (*GAPDH*, *ACTB*, *B2M*) and shown with bars showing mean $\pm$ SD (n $=$ 3 independent differentiations).

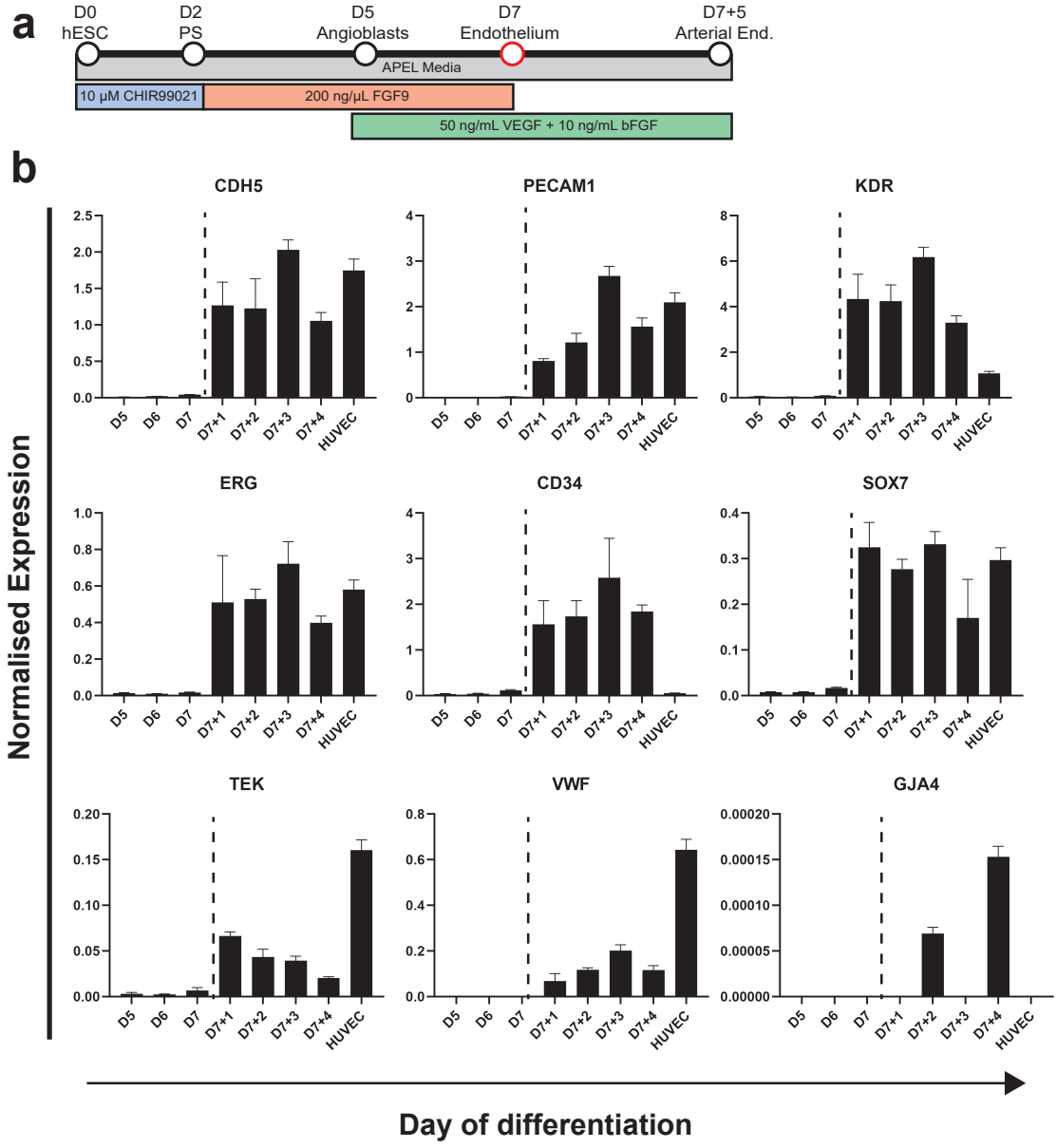


Figure 2: Variation in gene expression profiles during endothelial differentiation
This figure relates to Fig.3.4d.**a** Differentiation protocol used. **b** RT-qPCR of endothelial marker genes with the mean ± SD (n = 4 independent differentiations) for each gene over the course of differentiation plotted. Black dotted line indicates cells were replated. Whole cultures were taken for RNA extraction and cDNA synthesis. Gene expression is normalised using $2^{-\Delta Ct}$ to the average of three housekeeping genes (*GAPDH*, *ACTB*, *B2M*).

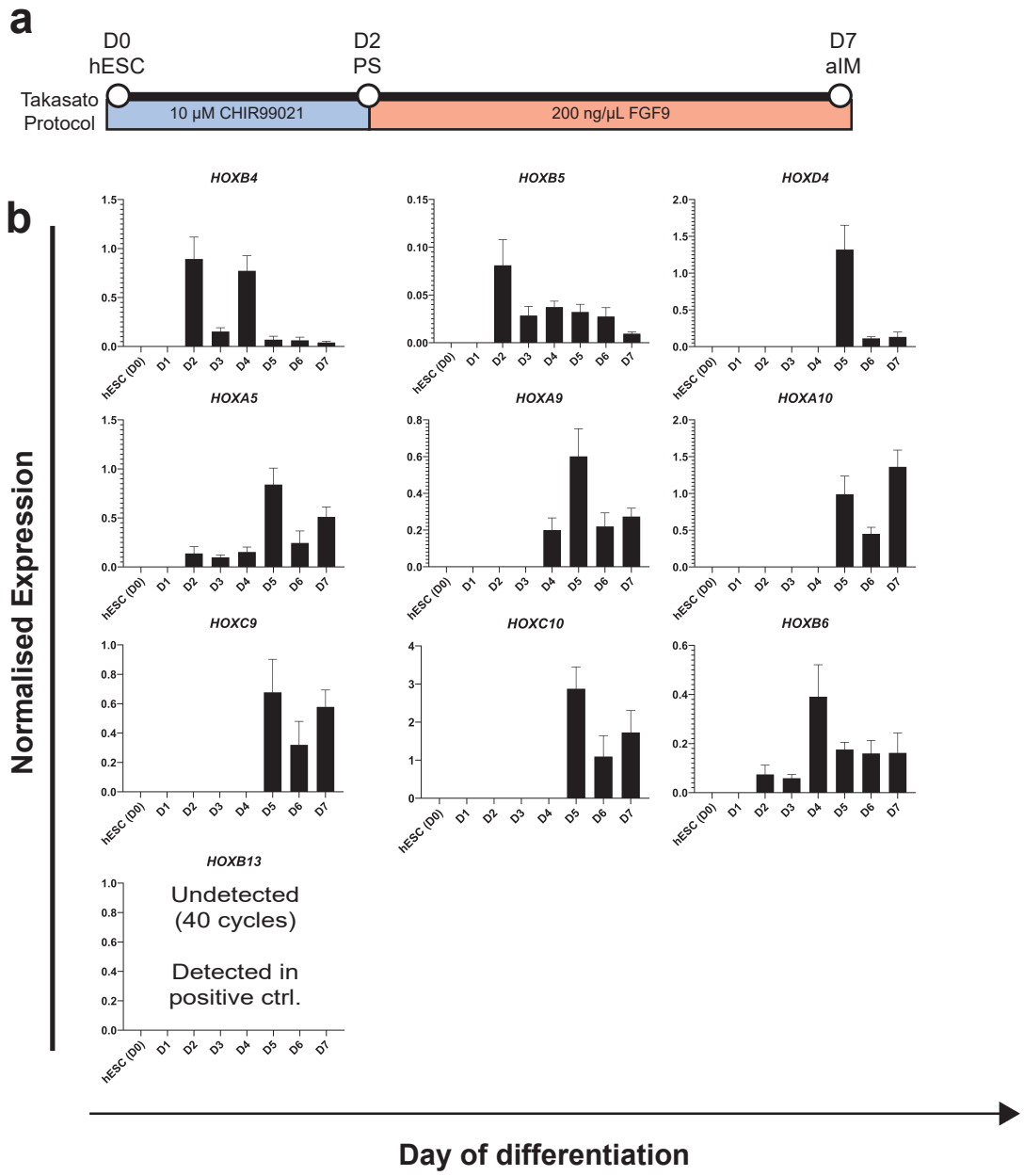


Figure 3: Variation in HOX gene expression profiles during Takasato protocol differentiation

This figure relates to Fig.3.5a.**a** Differentiation protocol used. **b** RT-qPCR of HOX genes with the mean ± SD (n = 4 independent differentiations) for each gene over the course of differentiation plotted. Whole cultures were taken for RNA extraction and cDNA synthesis. Gene expression is normalised using $2^{-\Delta Ct}$ to the average of three housekeeping genes (*GAPDH*, *ACTB*, *B2M*).

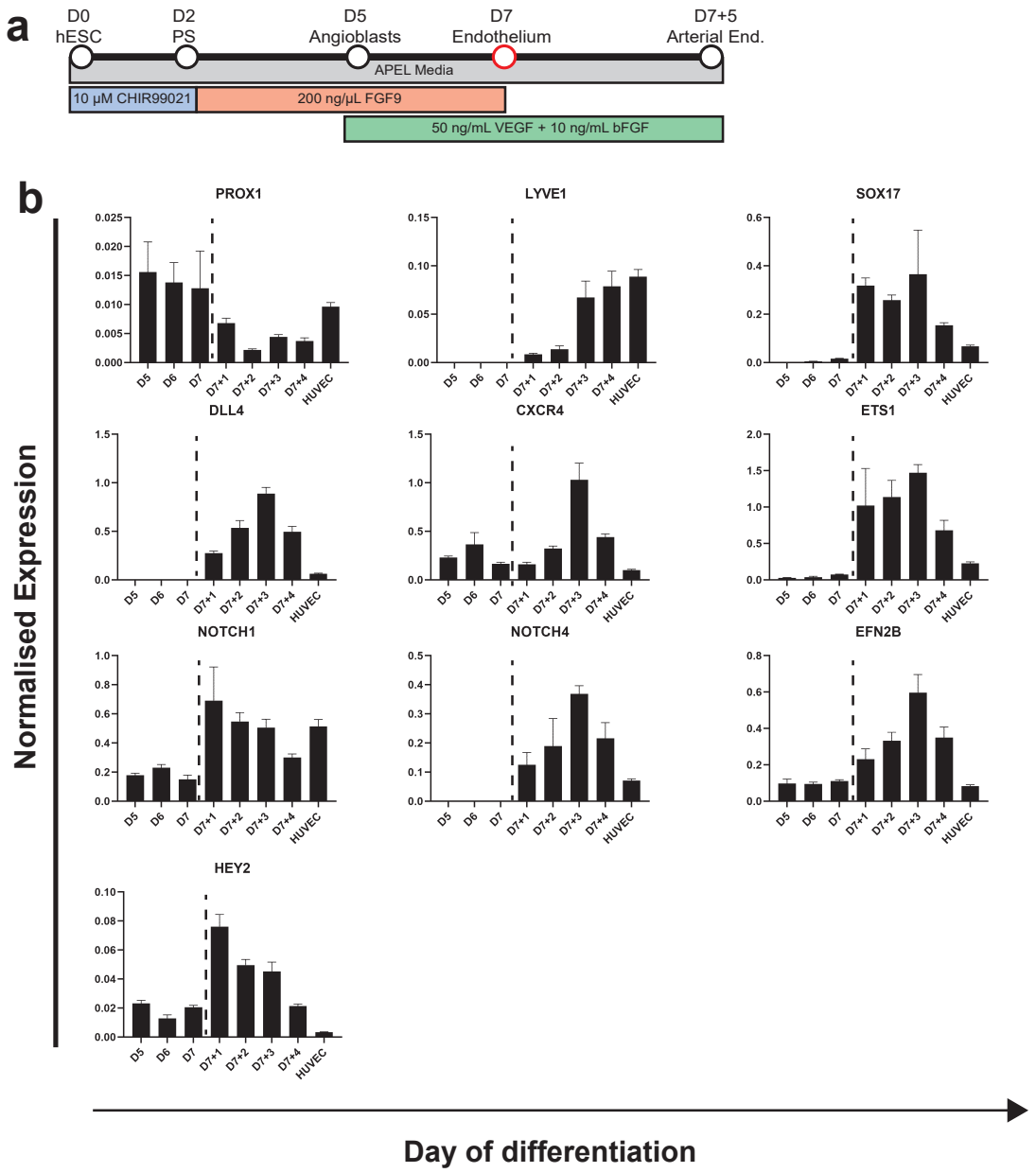


Figure 4: Variation in gene expression profiles during arterial differentiation
This figure relates to Fig.3.5c.**a** Differentiation protocol used. **b** RT-qPCR of arterial endothelium marker genes with the mean ± SD (n = 4 independent differentiations) for each gene over the course of differentiation plotted. Black dotted line indicates cells were replated. Whole cultures were taken for RNA extraction and cDNA synthesis. Gene expression is normalised using $2^{-\Delta Ct}$ to the average of three housekeeping genes (*GAPDH*, *ACTB*, *B2M*).

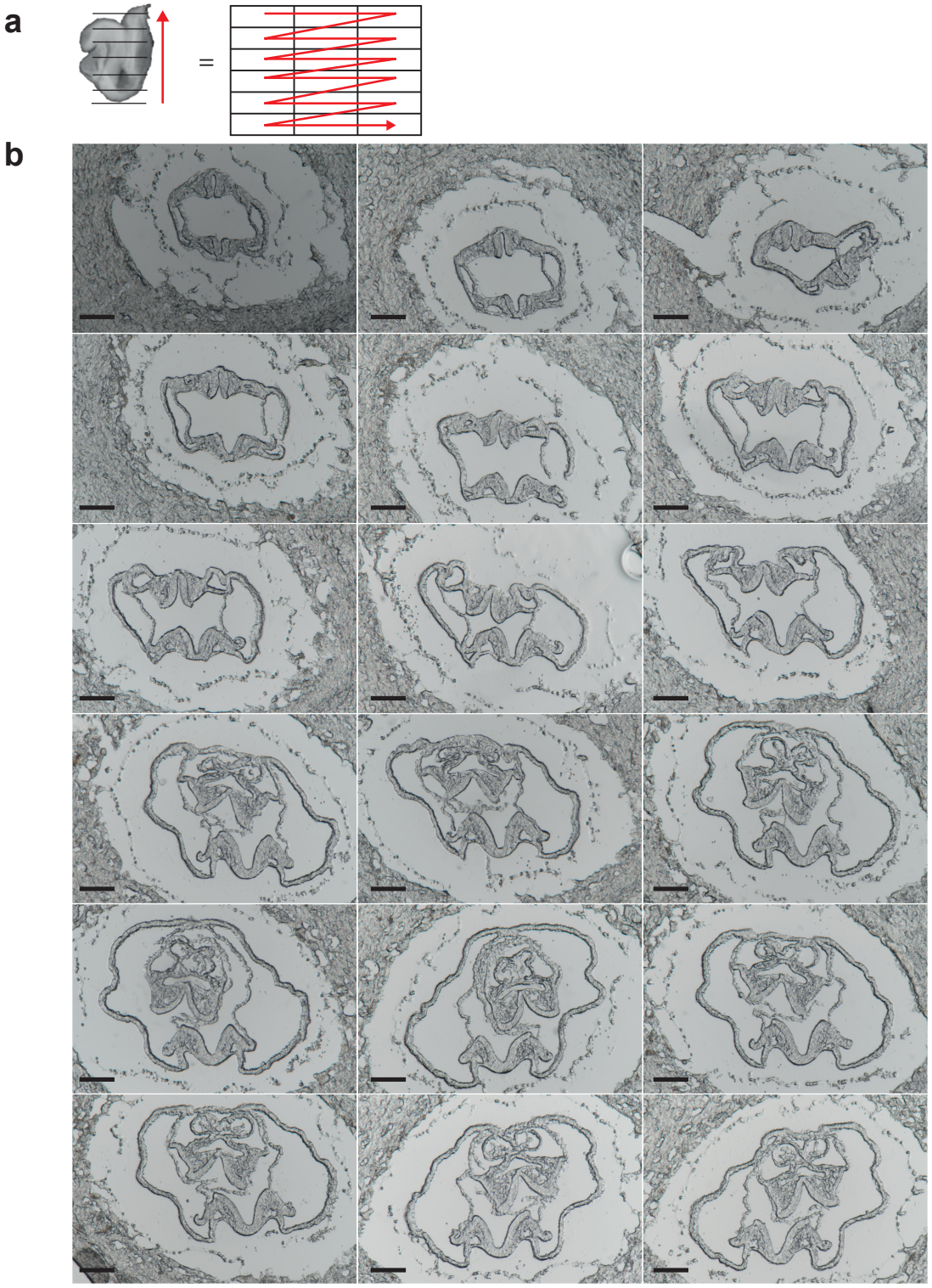


Figure 5: Brightfield images of 30µm sections from embryo 3
a Schematic showing the order of sections in the embryo. Sectioning was performed from the base of the trunk up towards the head and tail. **b** Every transverse section collected from E8.0 embryo 3 for dHCR with anterior at the top and posterior at the bottom. Scale bars represent 150 µm.

Channel	Gene	NCBI Accession Number	Probe Set Size
B1	Foxc2	NM_013519.2	32
	Gata2	NM_008090.5	20
	Hand2	NM_010402.4	18
	Hoxb6	NM_008269.2	16
	Nkx2-5	NM_008700.2	17
	Tal1	NM_011527.3	40
	Xist	NR_001463.3	40
B3	Lhx1	NM_008498.3	20
B4	Cdh5	NM_009868.4	40
	Etv2	NM_007959.2	14
	Mef2c	NM_001170537.1	40
	Nfatc1	NM_016791	30
	Nr2f2	NM_009697.3	33
	T	NM_009309.2	30
B5	Dll4	NM_019454.3	40
	Msgn1	NM_019544	8
	Osr1	NM_011859	20
	Pax8	NM_011040.4	20
	Runx1	NM_009821.3	40
	Sox17	NM_011441.5	20
	Tek	NM_013690.3	40
	Vegfa	NM_001287056.1	38

Figure 6: List of HCR probes and their initiators used
The initiator channel, NCBI accession number used to generate probe sets and the number of probe pairs per HCR probe set used during my thesis.

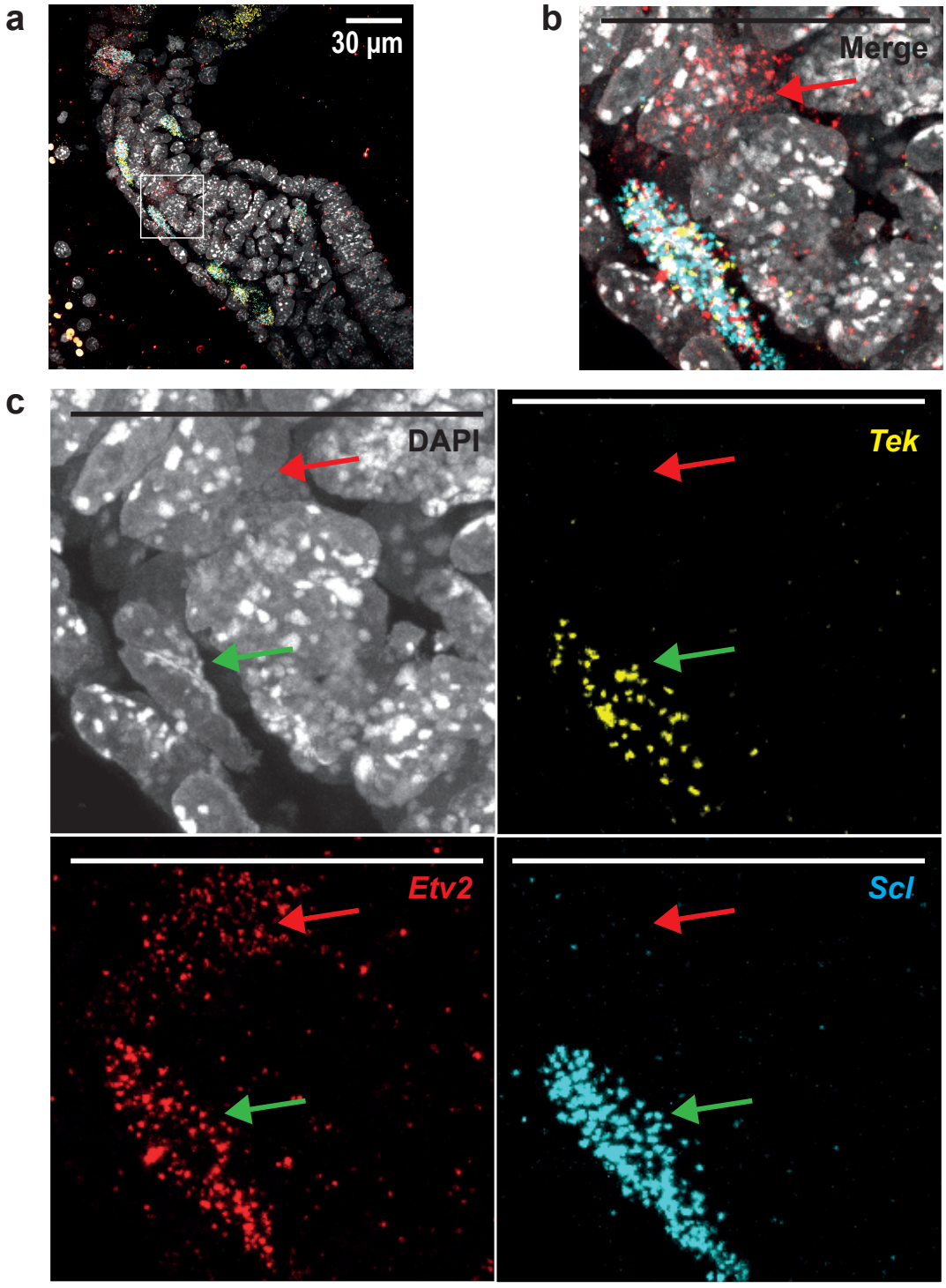


Figure 7: High magnification of Etv2 single positive cell in Embryo 2
a Left side of the posterior of E8.0 embryo 2 as displayed in Fig.4.10. White boxed region is magnified in **b** (merge) and **c** (single channels). Red arrow, rare Etv2-single positive cell; Green arrow, triple positive cell. Scale bars represent 30 µm.

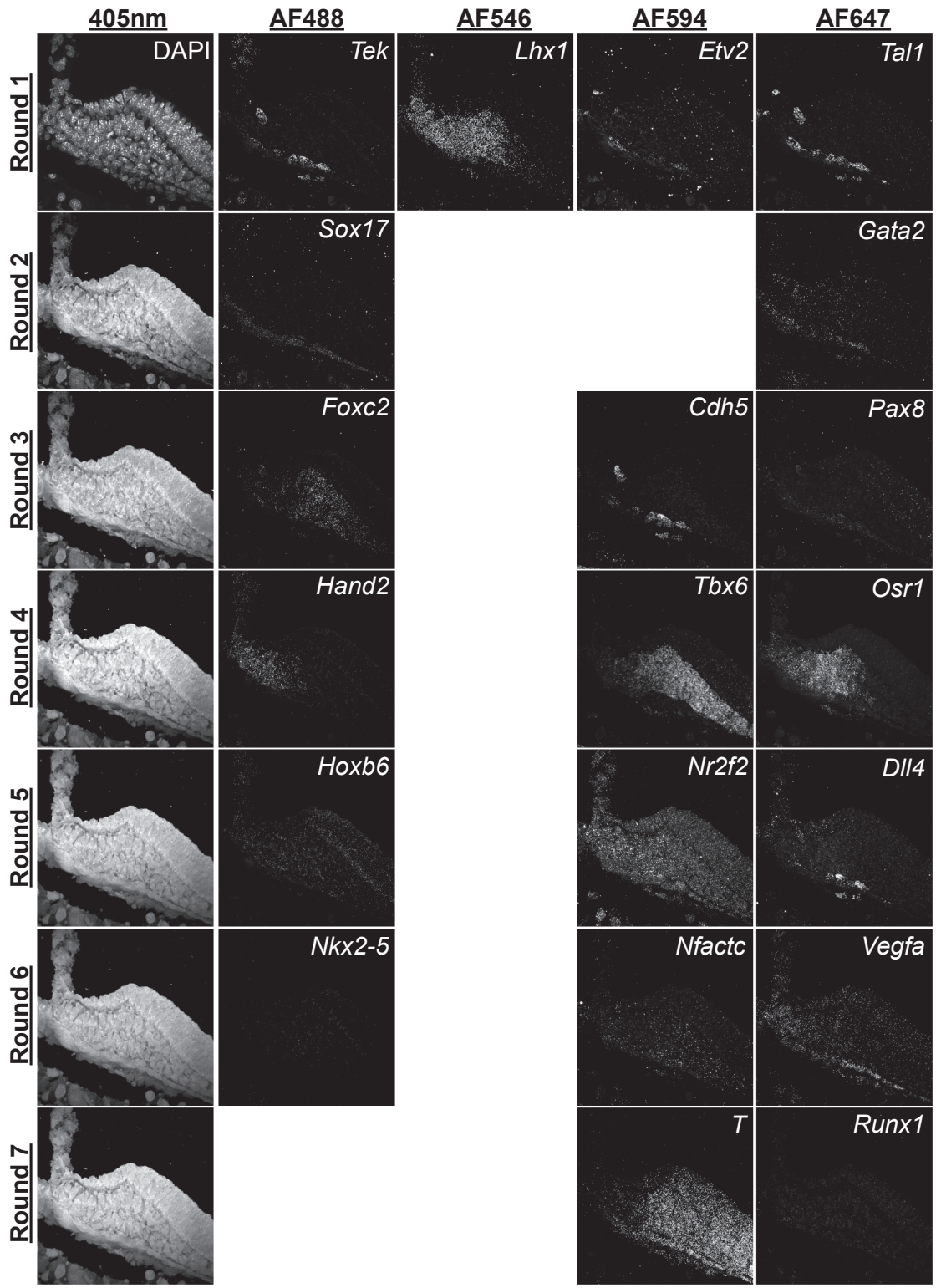


Figure 8: All genes probed for embryo 3

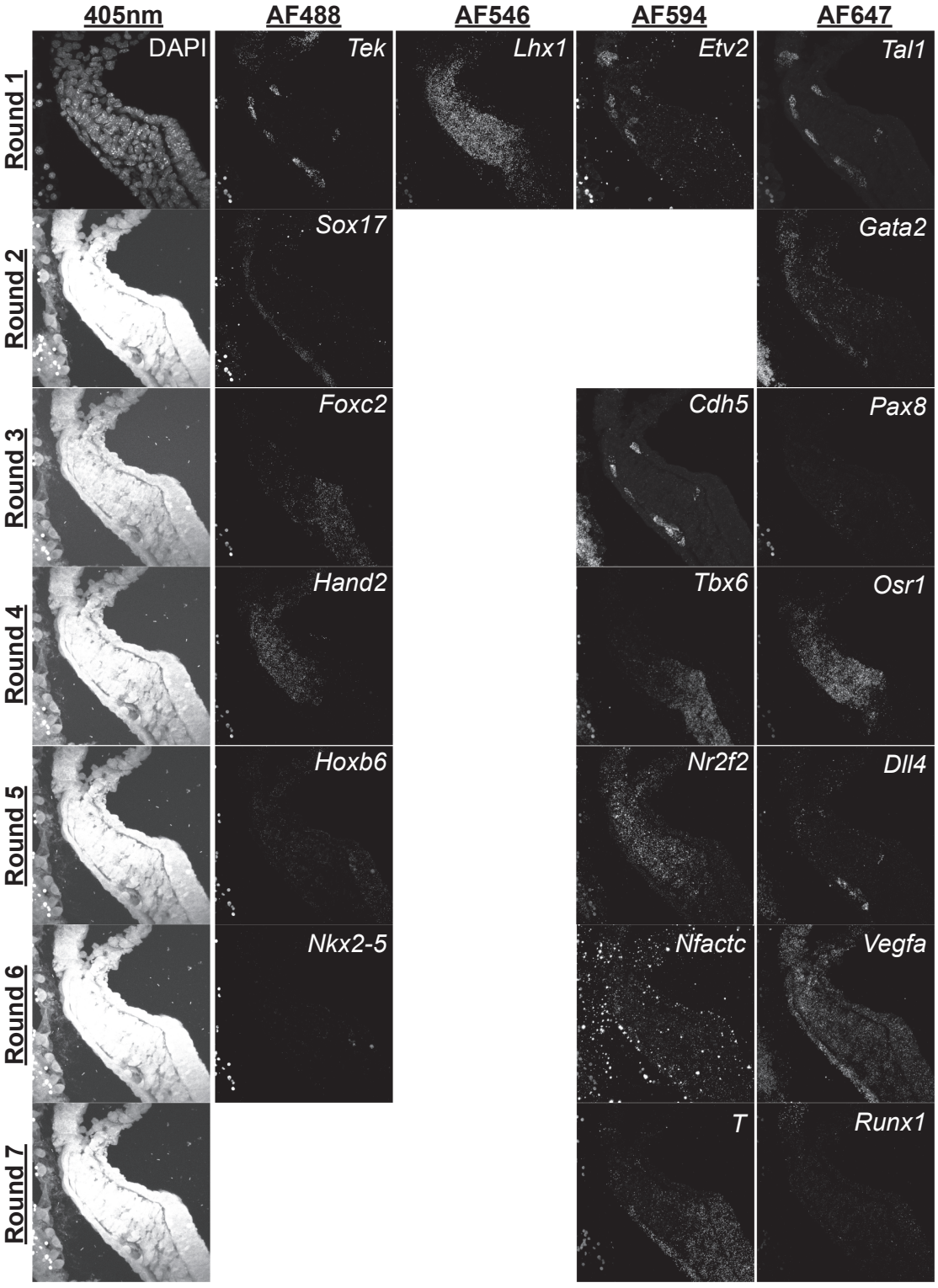


Figure 9: All genes probed for embryo 2

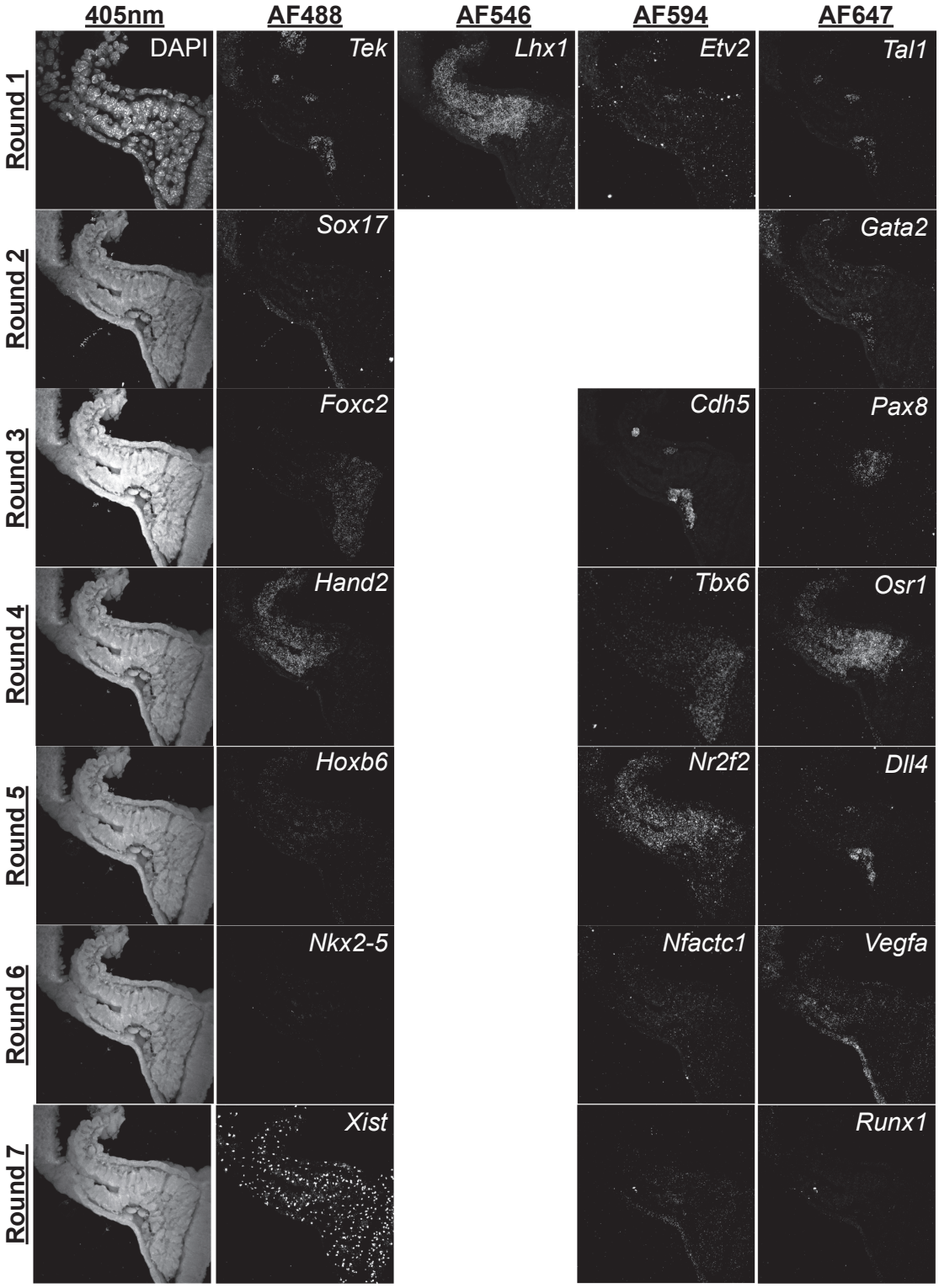


Figure 10: All genes probed for embryo 3

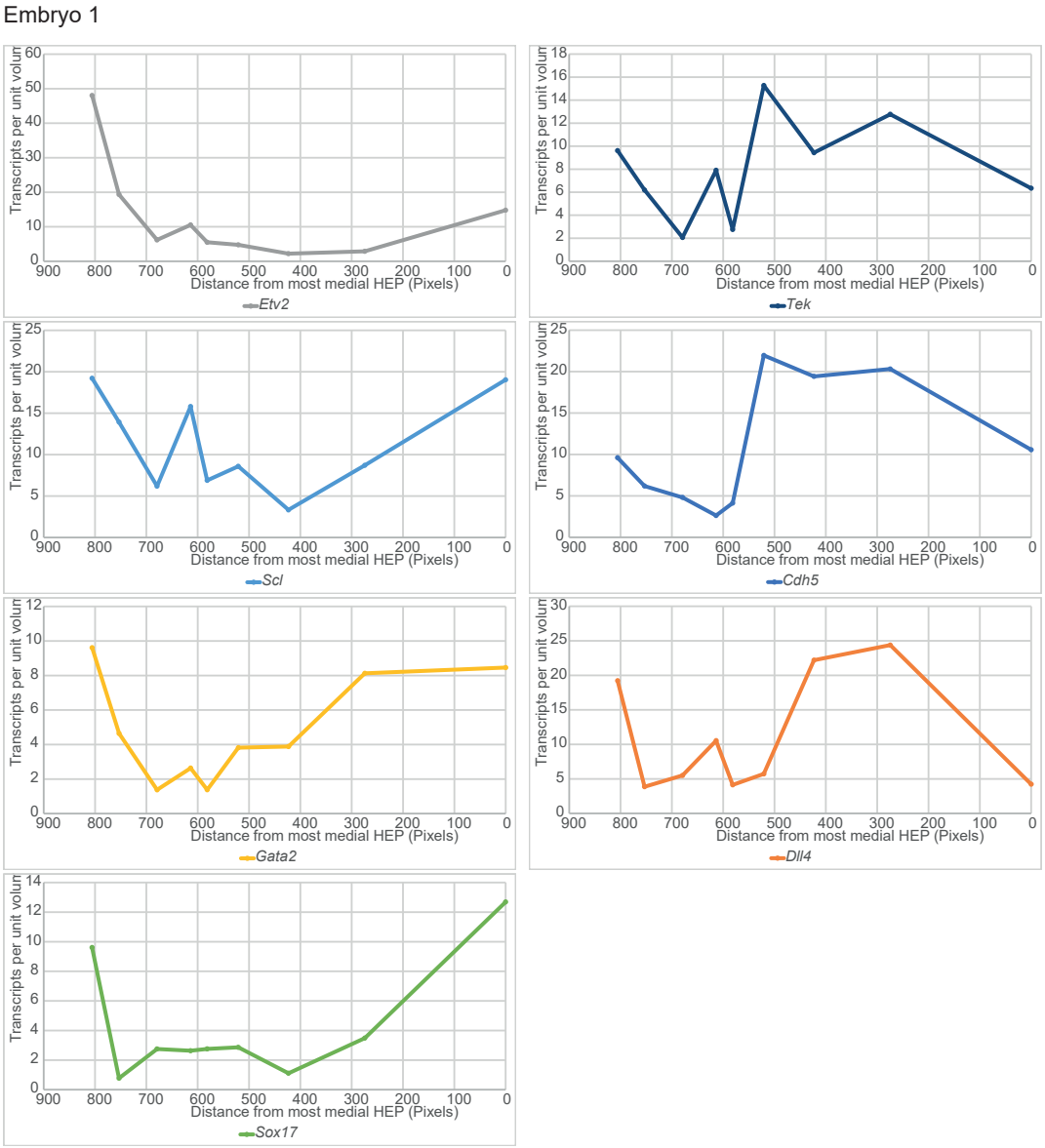


Figure 11: Individually plotted gene profiles in embryo 1 HEPs ordered along the medial-ventral axis

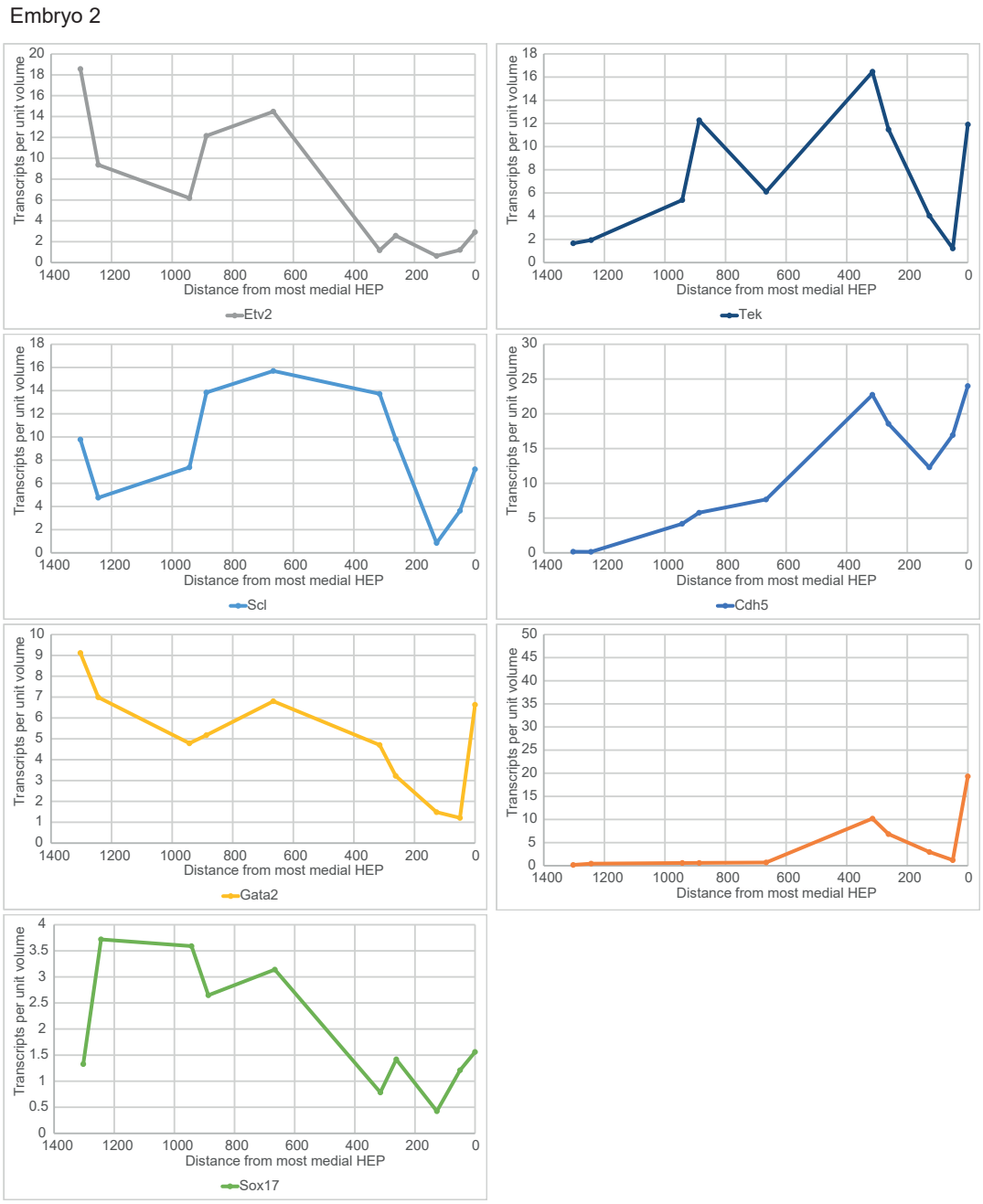


Figure 12: Individually plotted gene profiles in embryo 2 HEPs ordered along the medial-ventral axis

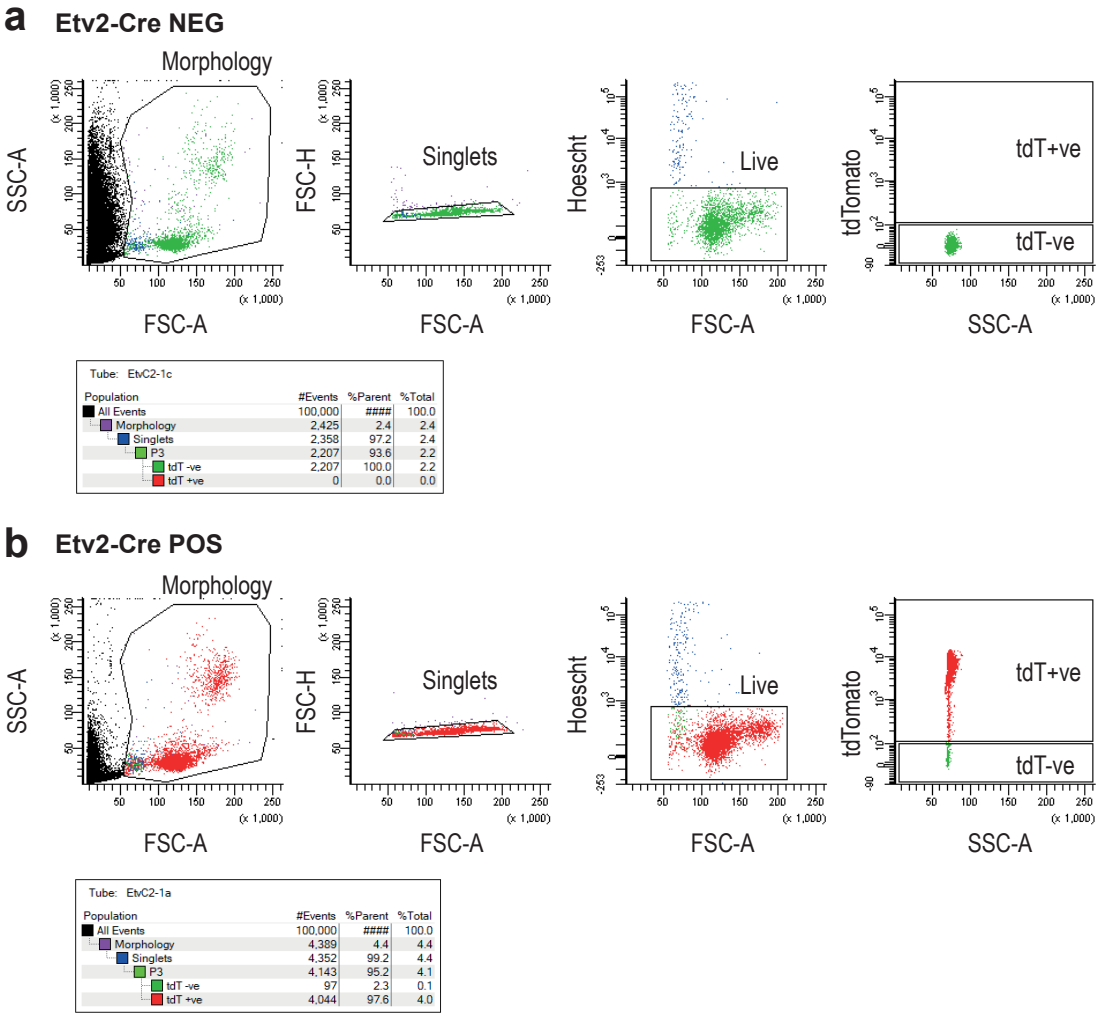


Figure 13: FACS gating strategy for assessment of Etv2-Cre labelling efficiency in peripheral blood

Red blood cell depleted blood from 8 week old animals (from Etv2-Cre cross with Ai9) was stained with Hoescht (live-dead descrimination) and run on FACS. Distinctive peripheral blood cell populations were identified by their FSC-SSC profile (Morphology), filtered for singlets (FSC-A,FSC-H), live cells (hoescht-negative) and assessed for tdTomato levels. Both Etv2-Cre negative (NEG, **a**) and positive (POS, **b**) animals are shown. Representative of 3 negative animals and 4 posiive animals. FSC, forward scatter; SSC, side scatter.

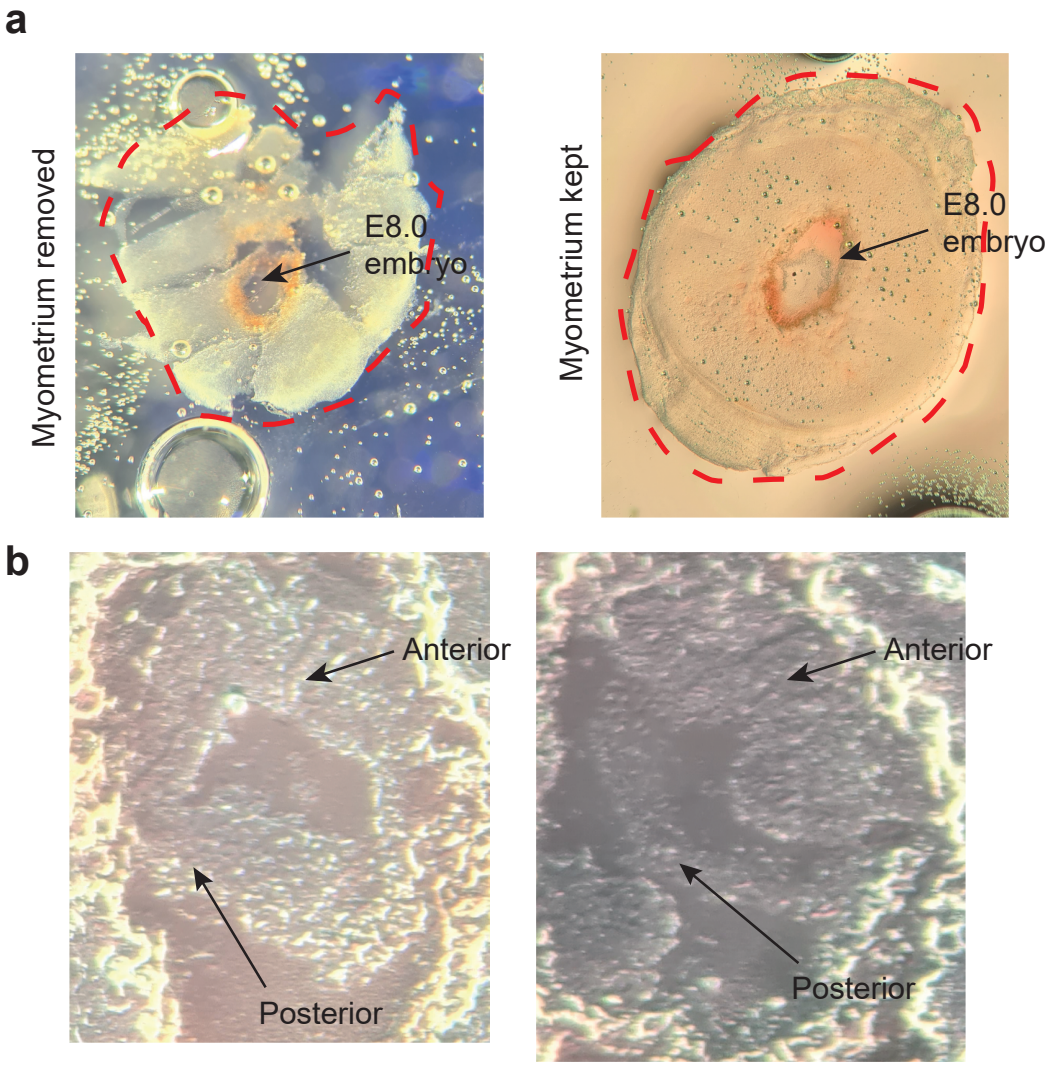


Figure 14: FACS gating strategy for assessment of Etv2-Cre labelling efficiency in peripheral blood

a Fresh frozen sections of E8.0 mouse embryos in decidua with the myometrium removed (left) and kept (right). Keeping the myometrium greatly improves section morphology. Red dotted line indicates the edge of the decidua. **b** Two examples of fresh frozen sections with undistrurbed structure and morphology.

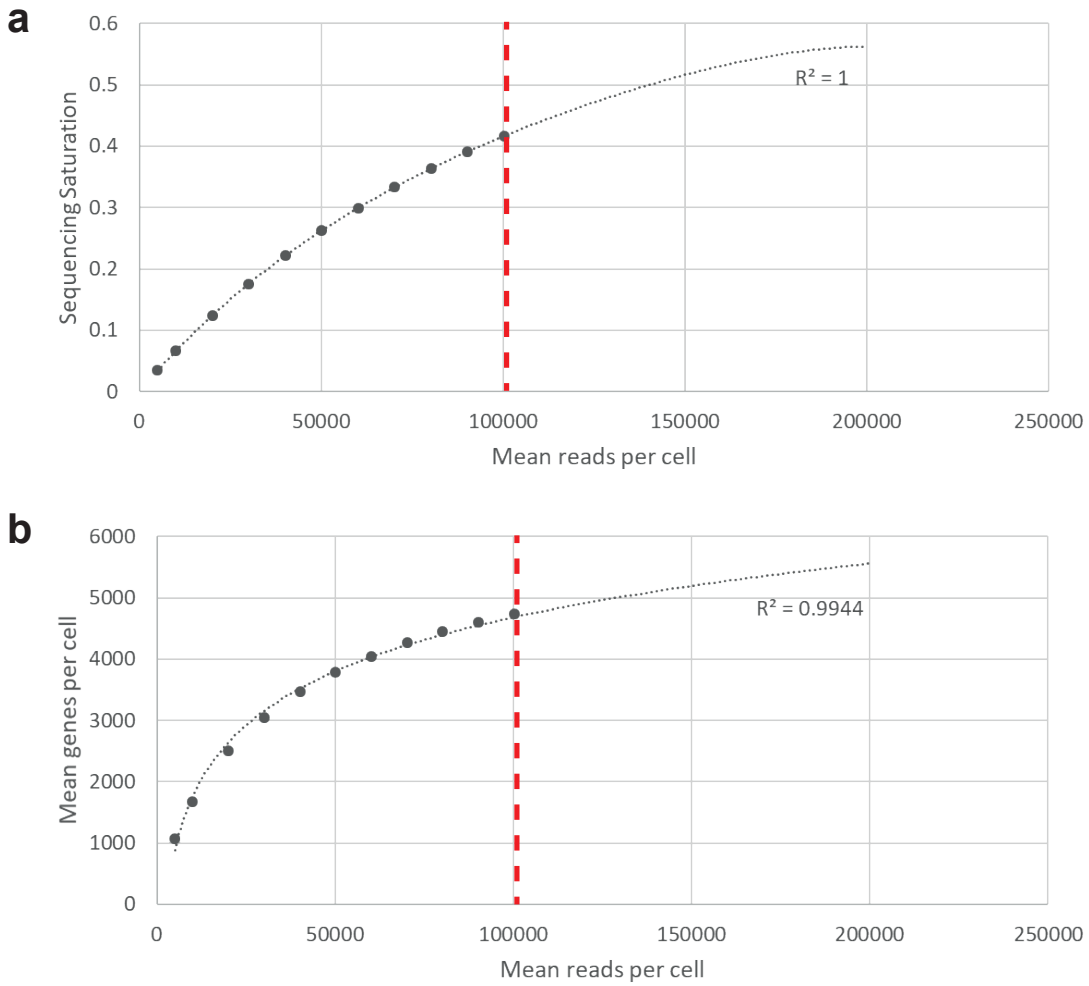


Figure 15: Extrapolation of sequencing saturation
Sequencing saturation data from the Illumina sequencing of 10x cDNA library was extrapolated to estimate the sequencing saturation (**a**) and mean genes per cell (**b**) if libraries were sequenced further. Each data-point is taken from sequencing data. Black dotted line is a line of best fit and the red dotted line represents current sequencing depth.

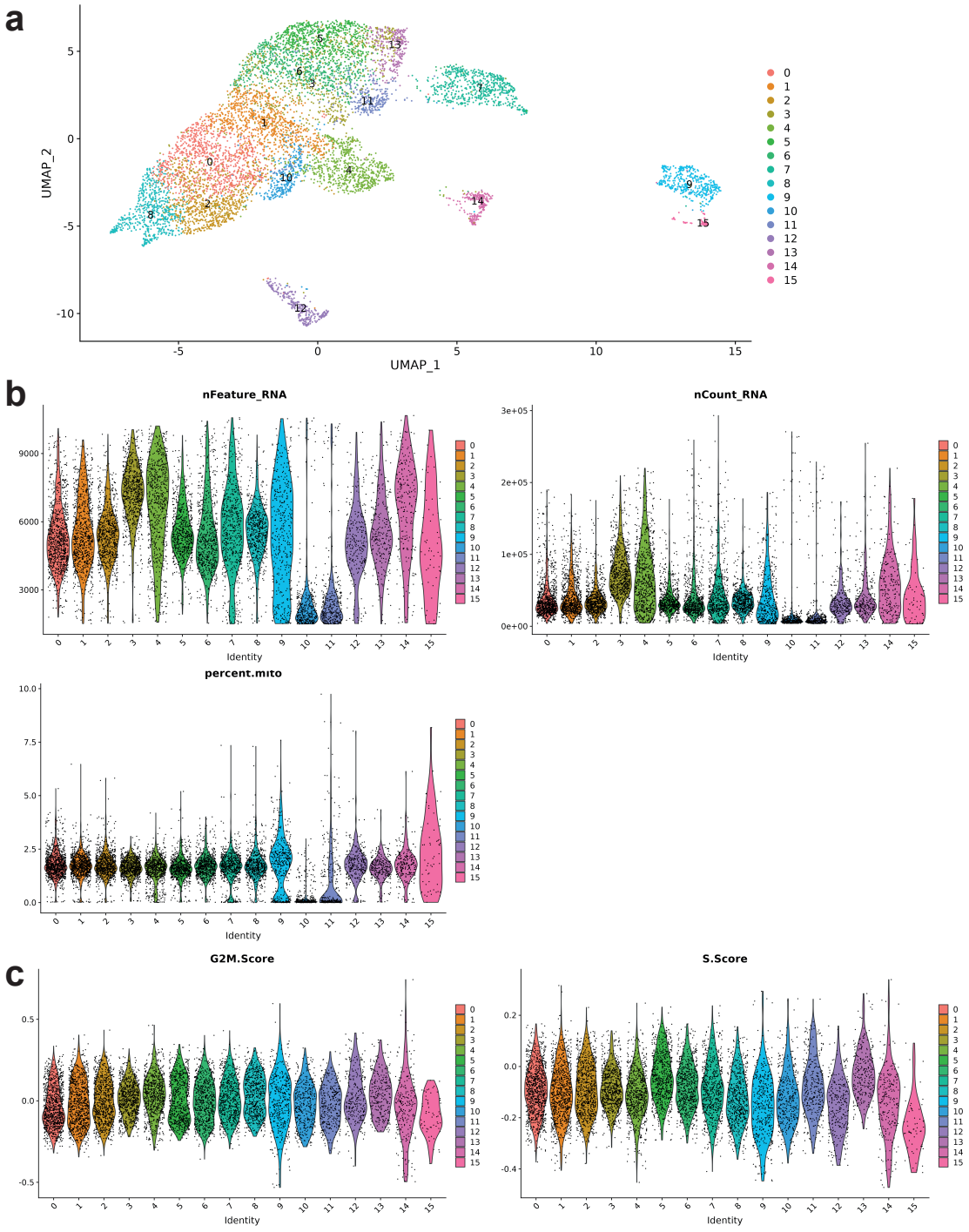


Figure 16: Identification of two clusters for removal

a Uniform manifold approximation and projection (UMAP) plot showing cells before removal of two low-quality clusters (10 and 11). Cells are coloured by cluster. **b** Key cluster metrics on violin plots used to identify clusters 10 and 11 as low-quality cells for removal. Each point is a single cell and the width of the violin is proportional to frequency. Violins are coloured according to cluster identity. **c** Assessment of cell-cycle scoring in each cluster.

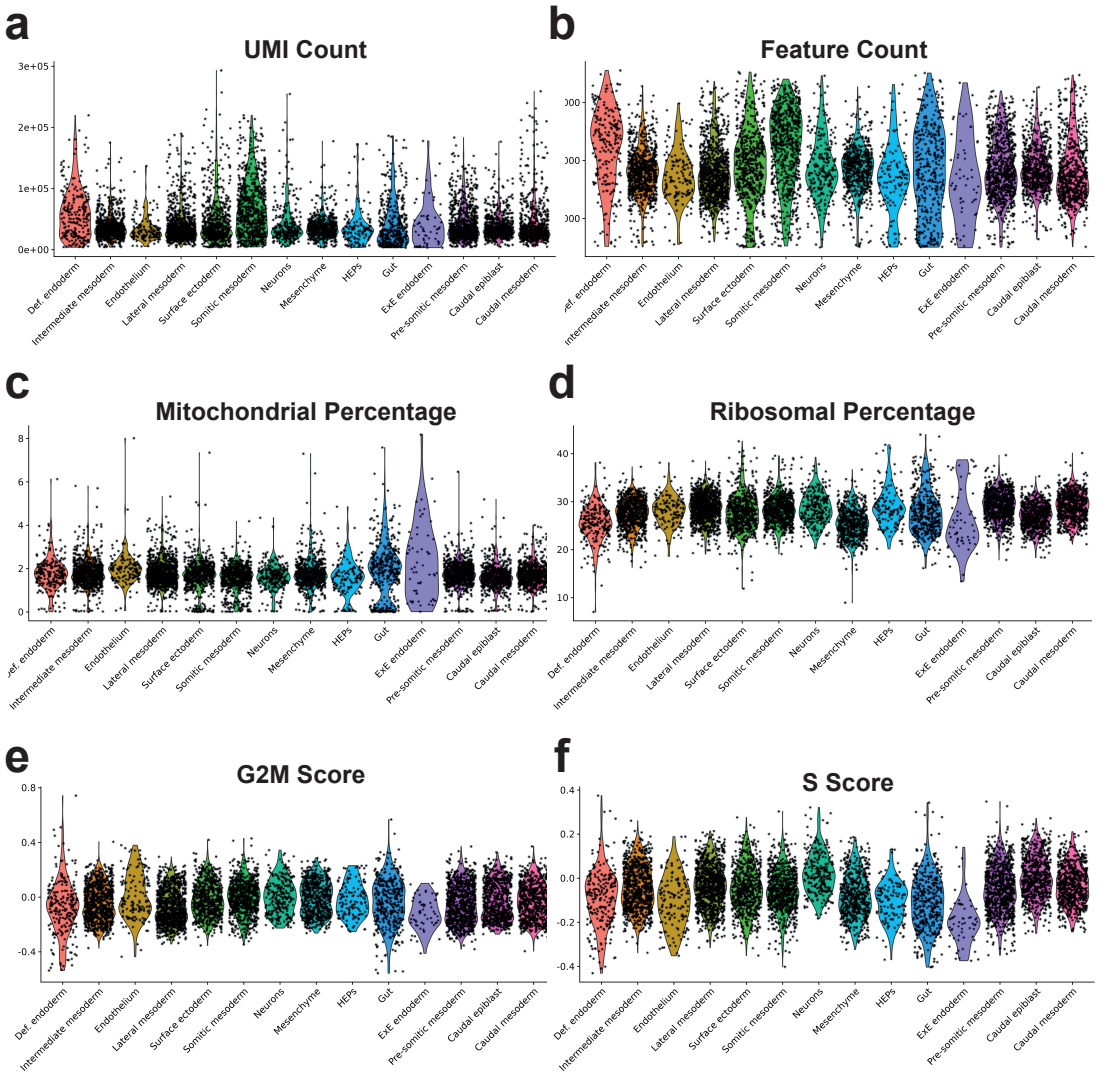


Figure 17: Cluster quality control

A number of key metrics were plotted on violin plots to show the distribution of cells in and between clusters to check the quality of identified clusters after removal of . Black dots represent single cells with width of the violin indicating frequency. Violins are coloured according to cluster identity which is written on the x-axis. UMI counts (**a**), feature count (**b**), mitochondrial percentage (**c**) and ribosomal percentage (**d**) can be used to assess the health and quality of capture cells. Meanwhile G2M (**e**) and S (**f**) score are used to assess cluster that may represent cycling cells instead of cell identity.

References

[1] Naomi Moris, Cristina Pina, and Alfonso Martinez Arias. "Transition states and cell fate decisions in epigenetic landscapes". In: *Nature Reviews Genetics* 17.11 (2016), pp. 693–703. URL: http://dx.doi.org/10.1038/nrg.2016.98.

[2] Sebastian J. Arnold and Elizabeth J. Robertson. "Making a commitment: Cell lineage allocation and axis patterning in the early mouse embryo". In: *Nature Reviews Molecular Cell Biology* 10.2 (2009), pp. 91–103.

[3] P. Q. Thomas, A. Brown, and R. S.P. Beddington. "Hex: A homeobox gene revealing peri-implantation asymmetry in the mouse embryo and an early transient marker of endothelial cell precursors". In: *Development* 125.1 (1998).

[4] Shankar Srinivas et al. "Active cell migration drives the unilateral movements of the anterior visceral endoderm". In: *Development* 131.5 (2004).

[5] Paul Thomas and Rosa Beddington. "Anterior primitive endoderm may be responsible for patterning the anterior neural plate in the mouse embryo". In: *Current Biology* 6.11 (1996).

[6] Jixiang Ding et al. "Cripto is required for correct orientation of the anterior-posterior axis in the mouse embryo". In: *Nature* 395.6703 (1998).

[7] Dominic P. Norris et al. *The Foxh1-dependent autoregulatory enhancer controls the level of Nodal signals in the mouse embryo*. 2002.

[8] Aitana Perea-Gomez et al. "Initiation of Gastrulation in the Mouse Embryo Is Preceded by an Apparent Shift in the Orientation of the Anterior-Posterior Axis". In: *Current Biology* 14.3 (2004).

[9] Kirstie A Lawson, Juanito J Meneses, and Roger A Pedersen. "Clonal analysis of epiblast fate during germ layer formation in the mouse embryo". In: *Development* 113 (1991), pp. 891–911.

[10] Koichiro Hashimoto and Norio Nakatsuji. "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 (1989).

[11] Yuji Mishina et al. "Bmpr encodes a type I bone morphogenetic protein receptor that is essential for gastrulation during mouse embryogenesis". In: *Genes and Development* 9.24 (1995).

[12] Frank L. Conlon et al. "A primary requirement for nodal in the formation and maintenance of the primitive streak in the mouse". In: *Development* 120.7 (1994).

[13] N. Ray Dunn et al. "Combinatorial activities of Smad2 and Smad3 regulate mesoderm formation and patterning in the mouse embryo". In: *Development* 131.8 (2004).

[14] Joerg Huelsken et al. "Requirement for β-catenin in anterior-posterior axis formation in mice". In: *Journal of Cell Biology* 148.3 (2000).

[15] Pentao Liu et al. "Requirement for Wnt3 in vertebrate axis formation". In: *Nature Genetics* 22.4 (1999).

[16] Kirstie A. Lawson. *Fate mapping the mouse embryo*. 1999.

[17] Simon J. Kinder, David A F Loebel, and Patrick P L Tam. "Allocation and early differentiation of cardiovascular progenitors in the mouse embryo". In: *Trends in Cardiovascular Medicine* 11.5 (2001), pp. 177–184.

[18] Wallis Nahaboo and Isabelle Migeotte. "Cleavage and Gastrulation in the Mouse Embryo". In: *eLS* (Feb. 2018), pp. 1–6. URL: https://onlinelibrary.wiley.com/doi/full/10.1002/9780470015902.a0001068.pub3%20https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470015902.a0001068.pub3%20https://onlinelibrary.wiley.com/doi/10.1002/9780470015902.a0001068.pub3.

[19] Maala Parameswaran and Patrick P.L. Tam. *Regionalisation of cell fate and morphogenetic movement of the mesoderm during mouse gastrulation*. 1995.

[20] Gerald C. Chu et al. "Differential requirements for Smad4 in TGFβ-dependent patterning of the early mouse embryo". In: *Development* 131.15 (2004).

[21] Stéphane D. Vincent et al. "Cell fate decisions within the mouse organizer are governed by graded Nodal signals". In: *Genes and Development* 17.13 (2003).

[22] Gloria S. Kwon, Manuel Viotti, and Anna Katerina Hadjantonakis. "The Endoderm of the Mouse Embryo Arises by Dynamic Widespread Intercalation of Embryonic and Extraembryonic Lineages". In: *Developmental Cell* 15.4 (2008).

[23] Aida Di-Gregorio et al. "BMP signalling inhibits premature neural differentiation in the mouse embryo". In: *Development* 134.18 (2007).

[24] Anne Camus et al. "Absence of Nodal signaling promotes precocious neural differentiation in the mouse embryo". In: *Developmental Biology* 295.2 (2006).

[25] Aitana Perea-Gomez et al. "Nodal antaginists in the anterior visceral endoderm prevent the formation of multiple primitive streaks". In: *Developmental Cell* 3.5 (2002).

[26] P P Tam et al. "The allocation of epiblast cells to the embryonic heart and other mesodermal lineages: the role of ingression and tissue movement during gastrulation." In: *Development (Cambridge, England)* 124.9 (1997), pp. 1631–42. URL: http://www.ncbi.nlm.nih.gov/pubmed/9165112.

[27] Blanca Pijuan-Sala et al. "Single-cell chromatin accessibility maps reveal regulatory programs driving early mouse organogenesis". In: *Nature Cell Biology* (Mar. 2020), pp. 1–11. URL: http://www.nature.com/articles/s41556-020-0489-9.

[28] Fabienne Lescroart et al. "Early lineage restriction in temporally distinct populations of Mesp1 progenitors during mammalian heart development". In: *Nature Cell Biology* 16.9 (2014).

[29] Fabienne Lescroart et al. "Defining the earliest step of cardiovascular lineage segregation by single-cell RNA-seq". In: *Science* 359.6380 (2018), pp. 1177–1181.

[30] Kenzo Ivanovitch et al. "Ventricular, atrial, and outflow tract heart progenitors arise from spatially and molecularly distinct regions of the primitive streak". In: *PLoS Biology* 19.5 (2021).

[31] Carolina Guibentif et al. "Diverse Routes toward Early Somites in the Mouse Embryo". In: *Developmental Cell* 56.1 (Jan. 2021), pp. 141–153.

[32] Nesanet Mitiku and Julie C. Baker. "Genomic Analysis of Gastrulation and Organogenesis in the Mouse". In: *Developmental Cell* 13.6 (2007).

[33] Blanca Pijuan-Sala et al. "A single-cell molecular map of mouse gastrulation and early organogenesis". In: *Nature* 566.7745 (Feb. 2019), pp. 490–495. URL: http://www.nature.com/articles/s41586-019-0933-9.

[34] Markus Mittnenzweig et al. "A single-embryo, single-cell time-resolved model for mouse gastrulation". In: *Cell* 184.0 (Apr. 2021), pp. 1–18. URL: https://linkinghub.elsevier.com/retrieve/pii/S0092867421004396.

[35] Junyue Cao et al. "The single-cell transcriptional landscape of mammalian organogenesis". In: *Nature* 566.7745 (2019), pp. 496–502. URL: https://doi.org/10.1038/s41586-019-0969-x.

[36] Yoji Kojima, Oliver H. Tam, and Patrick P.L. Tam. *Timing of developmental events in the early mouse embryo*. 2014.

[37] Caleb Weinreb et al. "Lineage tracing on transcriptional landscapes links state to fate during differentiation". In: *Science* 367.6479 (Jan. 2020), eaaw3381.

[38] Sadie Vanhorn and Samantha A Morris. "Next-Generation Lineage Tracing and Fate Mapping to Interrogate Development". In: *Developmental Cell* 0.0 (2020). URL: https://doi.org/10.1016/j.devcel.2020.10.021.

[39] T. Lohoff et al. "Integration of spatial and single-cell transcriptomic data elucidates mouse organogenesis". In: *Nature Biotechnology* 40.1 (2022).

[40] Guangdun Peng et al. "Spatial Transcriptome for the Molecular Annotation of Lineage Fates and Cell Identity in Mid-gastrula Mouse Embryo". In: *Developmental Cell* 36.6 (2016), pp. 681–697. URL: http://dx.doi.org/10.1016/j.devcel.2016.02.020.

[41] Ricard Argelaguet et al. "Multi-omics profiling of mouse gastrulation at single-cell resolution". In: *Nature* 576.7787 (2019), pp. 487–491. URL: http://dx.doi.org/10.1038/s41586-019-1825-8.

[42] Jason R. Spence, Ryan Lauf, and Noah F. Shroyer. *Vertebrate intestinal endoderm development*. 2011.

[43] Richard I. Sherwood, Tzong Yang Albert Chen, and Douglas A. Melton. "Transcriptional dynamics of endodermal organ formation". In: *Developmental Dynamics* 238.1 (2009).

[44] Claudio D. Stern. *Neural induction: 10 years on since the 'default model'*. 2006.

[45] Vicki Metzis et al. "Nervous System Regionalization Entails Axial Allocation before Neural Differentiation". In: *Cell* 175.4 (2018).

[46] Anne K. Knecht and Marianne Bronner-Fraser. *Induction of the neural crest: A multigene process*. 2002.

[47] Chandra S. Mayanil. *Transcriptional and epigenetic regulation of neural crest induction during neurulation*. 2013.

[48] G. Winnier et al. "Bone morphogenetic protein-4 is required for mesoderm formation and patterning in the mouse". In: *Genes and Development* 9.17 (1995).

[49] Karen M. Downs et al. *Investigation into a role for the primitive streak in development of the murine allantois*. 2004.

[50] K. M. Downs and R. L. Gardner. "An investigation into early placental ontogeny: Allantoic attachment to the chorion is selective and developmentally regulated". In: *Development* 121.2 (1995).

[51] Karen M. Downs et al. "Vascularization in the murine allantois occurs by vasculogenesis without accompanying erythropoiesis". In: *Development (Cambridge, England)* 125.22 (1998), pp. 4507–4520. URL: https://pubmed.ncbi.nlm.nih.gov/9778509/.

[52] Florence Sabin. "Studies on the origin of blood-vessels and of red blood-corpuscles as seen in the living blastoderm of chicks during the second day of incubation." In: *Contributions to embryology.* (1920).

[53] P Murray. " The development in vitro of the blood of the early chick embryo ". In: *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character* 111.773 (1932).

[54] Kyunghee Choi et al. "A common precursor for hematopoietic and endothelial cells". In: *Development* 125.4 (1998).

[55] Marion Kennedy et al. "Development of the hemangioblast defines the onset of hematopoiesis in human ES cell differentiation cultures." In: *Blood* 109.7 (Apr. 2007), pp. 2679–87. URL: http://www.ncbi.nlm.nih.gov/pubmed/17148580%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1852226.

[56] Tara L. Huber et al. "Haemangioblast commitment is initiated in the primitive streak of the mouse embryo". In: *Nature* 432.7017 (2004).

[57] Laura Padrón-Barthe et al. "Clonal analysis identifies hemogenic endothelium as the source of the blood-endothelial common lineage in the mouse embryo". In: *Blood* 124.16 (2014).

[58] Hiroo Ueno and Irving L. Weissman. "Clonal Analysis of Mouse Development Reveals a Polyclonal Origin for Yolk Sac Blood Islands". In: *Developmental Cell* 11.4 (2006).

[59] Olivier Pourquié et al. "Lateral and axial signals involved in avian somite patterning: A role for BMP4". In: *Cell* 84.3 (1996).

[60] Akane Tonegawa and Yoshiko Takahashi. "Somitogenesis controlled by Noggin". In: *Developmental Biology* 202.2 (1998).

[61] Shoichiro Tani et al. *Understanding paraxial mesoderm development and sclerotome specification for skeletal repair*. 2020.

[62] Deborah L. Chapman et al. "Tbx6, a mouse T-box gene implicated in paraxial mesoderm formation at gastrulation". In: *Developmental Biology* 180.2 (1996).

[63] Ravindra B. Chalamalasetty et al. "Mesogenin 1 is a master regulator of paraxial presomitic mesoderm differentiation". In: *Development (Cambridge)* 141.22 (2014).

[64] Tatsuya Takemoto et al. "Tbx6-dependent Sox2 regulation determines neural or mesodermal fate in axial stem cells". In: *Nature* 470.7334 (2011).

[65] Bettina Wilm et al. "The forkhead genes, Foxc1 and Foxc2, regulate paraxial versus intermediate mesoderm cell fate". In: *Developmental Biology* 271.1 (2004).

[66] Oliver A. Stone and Didier Y.R. Stainier. "Paraxial Mesoderm Is the Major Source of Lymphatic Endothelium". In: *Developmental Cell* 50.2 (2019).

[67] Shahragim Tajbakhsh et al. "Redefining the genetic hierarchies controlling skeletal myogenesis: Pax- 3 and Myf-5 act upstream of MyoD". In: *Cell* 89.1 (1997).

[68] Maxime Bouchard et al. "Nephric lineage specification by Pax2 and Pax8." In: *Genes & development* 16.22 (Nov. 2002), pp. 2958–70. URL: http://www.ncbi.nlm.nih.gov/pubmed/12435636%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC187478.

[69] G. R. Dressler. "Advances in early kidney specification, development and patterning". In: *Development* 136.23 (2009), pp. 3863–3874. URL: http://dev.biologists.org/cgi/doi/10.1242/dev.034876.

[70] Atsuhiro Taguchi et al. "Redefining the in vivo origin of metanephric nephron progenitors enables generation of complex kidney structures from pluripotent stem cells". In: *Cell Stem Cell* 14.1 (2014), pp. 53–67. URL: http://dx.doi.org/10.1016/j.stem.2013.11.010.

[71] Richard G James. "Odd-skipped related 1 is required for development of the metanephric kidney and regulates formation and differentiation of kidney precursor cells". In: *Development* 133.15 (Aug. 2006), pp. 2995–3004. URL: http://www.ncbi.nlm.nih.gov/pubmed/16790474%20http://dev.biologists.org/cgi/doi/10.1242/dev.02442.

[72] Atsuhiro Taguchi and Ryuichi Nishinakamura. "Higher-Order Kidney Organogenesis from Pluripotent Stem Cells". In: *Cell Stem Cell* 21.6 (2017), pp. 730–746. URL: https://doi.org/10.1016/j.stem.2017.10.011.

[73] Navin Gupta and Ryuji Morizane. "Kidney development to kidney organoids and back again". In: *Seminars in Cell & Developmental Biology* (Oct. 2021).

[74] Joshua W. Mugford et al. "Osr1 expression demarcates a multi-potent population of intermediate mesoderm that undergoes progressive restriction to an Osr1-dependent nephron progenitor compartment within the mammalian kidney". In: *Developmental Biology* 324.1 (2008), pp. 88–98. URL: http://dx.doi.org/10.1016/j.ydbio.2008.09.010.

[75] Tania E. Tsang et al. "Lim1 activity is required for intermediate mesoderm differentiation in the mouse embryo". In: *Developmental Biology* 223.1 (2000), pp. 77–90.

[76] Miguel Torres et al. "Pax-2 controls multiple steps of urogenital development". In: *Development* 121.12 (1995).

[77] M. Takasato et al. "Directing human embryonic stem cell differentiation towards a renal lineage generates a self-organizing kidney". In: *Nature Cell Biology* 16.1 (2014), pp. 118–126. URL: http://dx.doi.org/10.1038/ncb2894.

[78] Jennifer S. Colvin et al. "Genomic organization and embryonic expression of the mouse fibroblast growth factor 9 gene". In: *Developmental Dynamics* 216.1 (Sept. 1999), pp. 72–88. URL: http://doi.wiley.com/10.1002/%28SICI%291097-0177%28199909%29216%3A1%3C72%3A%3AAID-DVDY9%3E3.0.CO%3B2-9.

[79] Hila Barak et al. "FGF9 and FGF20 Maintain the Stemness of Nephron Progenitors in Mice and Man". In: *Developmental Cell* 22.6 (2012).

[80] Patrick P L Tam and Richard R. Behringer. "Mouse gastrulation:the formation of the mammalian body plan". In: *Mechanisms of development* 68.1-2 (1997), pp. 3–25.

[81] Karin D. Prummel, Susan Nieuwenhuize, and Christian Mosimann. "The lateral plate mesoderm". In: *Development (Cambridge)* 147.12 (June 2020).

[82] Yuval Rinkevich et al. "Identification and prospective isolation of a mesothelial precursor lineage giving rise to smooth muscle cells and fibroblasts for mammalian internal organs, and their vasculature". In: *Nature Cell Biology* 14.12 (2012).

[83] Jianwen Que et al. "Mesothelium contributes to vascular smooth muscle and mesenchyme during lung development". In: *Proceedings of the National Academy of Sciences of the United States of America* 105.43 (2008).

[84] Rita Carmona et al. "Cells Derived from the Coelomic Epithelium Contribute to Multiple Gastrointestinal Tissues in Mouse Embryos". In: *PLoS ONE* 8.2 (2013).

[85] You Ying Chau et al. "Visceral and subcutaneous fat have different origins and evidence supports a mesothelial source". In: *Nature Cell Biology* 16.4 (2014).

[86] Ann C. Zovein et al. "Vascular remodeling of the vitelline artery initiates extravascular emergence of hematopoietic clusters". In: *Blood* 116.18 (2010), pp. 3435–3444.

[87] Hiroshi Kataoka et al. "Region-specific Etv2 ablation revealed the critical origin of hemogenic capacity from Hox6-positive caudal-lateral primitive mesoderm". In: *Experimental Hematology* 41.6 (2013), pp. 567–581. URL: http://dx.doi.org/10.1016/j.exphem.2013.02.009.

[88] Linda A. Lowe, Satoru Yamada, and Michael R. Kuehn. "HoxB6-Cre transgenic mice express Cre recombinase in extra-embryonic mesoderm, in lateral plate and limb mesoderm and at the midbrain/hindbrain junction". In: *Genesis* 26.2 (2000).

[89] Minh Thanh Nguyen et al. "Tamoxifen-dependent, inducible Hoxb6CreERT recombinase function in lateral plate and limb mesoderm, CNS isthmic organizer, posterior trunk neural crest, hindgut, and tailbud". In: *Developmental Dynamics* 238.2 (2009).

[90] J Palis et al. "Development of erythroid and myeloid progenitors in the yolk sac and embryo proper of the mouse." In: *Development (Cambridge, England)* 126.22 (1999), pp. 5073–5084. URL: http://dev.biologists.org/content/develop/126/22/5073.full.pdf%20http://www.ncbi.nlm.nih.gov/pubmed/10529424.

[91] Joanna Tober et al. "The megakaryocyte lineage originates from hemangioblast precursors and is an integral component both of primitive and of definitive hematopoiesis". In: *Blood* 109.4 (2007).

[92] Jane E. Barker. "Development of the mouse hematopoietic system. I. Types of hemoglobin produced in embryonic yolk sac and liver". In: *Developmental Biology* 18.1 (1968).

[93] K. Takahashi, F. Yamamura, and M. Naito. "Differentiation, maturation, and proliferation of macrophages in the mouse yolk sac: A light-microscopic, enzyme-cytochemical, immunohistochemical, and ultrastructural study". In: *Journal of Leukocyte Biology* 45.2 (1989).

[94] Makoto Naito et al. *Development, differentiation, and phenotypic heterogeneity of murine tissue macrophages*. 1996.

[95] Ming Jiang Xu et al. "Evidence for the presence of murine primitive megakaryocytopoiesis in the early yolk sac". In: *Blood* 97.7 (Apr. 2001), pp. 2016–2022. URL: https://pubmed.ncbi.nlm.nih.gov/11264166/.

[96] Kathleen E. McGrath et al. "Distinct Sources of Hematopoietic Progenitors Emerge before HSCs and Provide Functional Blood Cells in the Mammalian Embryo". In: *Cell Reports* 11.12 (2015).

[97] Paul D. Kingsley et al. "Yolk sac-derived primitive erythroblasts enucleate during mammalian embryogenesis". In: *Blood* 104.1 (2004).

[98] Jenna M. Frame, Kathleen E. McGrath, and James Palis. *Erythro-myeloid progenitors: "Definitive" hematopoiesis in the conceptus prior to the emergence of hematopoietic stem cells*. 2013.

[99] Elisa Gomez Perdiguero et al. "Tissue-resident macrophages originate from yolk-sac-derived erythro-myeloid progenitors". In: *Nature* 518.7540 (2015).

[100] Guillaume Hoeffel et al. "C-Myb+ Erythro-Myeloid Progenitor-Derived Fetal Monocytes Give Rise to Adult Tissue-Resident Macrophages". In: *Immunity* 42.4 (2015).

[101] Jenna M. Frame et al. "Definitive Hematopoiesis in the Yolk Sac Emerges from Wnt-Responsive Hemogenic Endothelium Independently of Circulation and Arterial Identity". In: *Stem Cells* 34.2 (2016).

[102] Tomomasa Yokomizo et al. "Requirement of Runx1/AML1/PEBP2αB for the generation of haematopoietic cells from endothelial cells". In: *Genes to Cells* 6.1 (2001).

[103] Tsukasa Okuda et al. "AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis". In: *Cell* 84.2 (1996).

[104] K. Sasaki et al. "Absence of fetal liver hematopoiesis in mice deficient in transcriptional coactivator core binding factor β". In: *Proceedings of the National Academy of Sciences of the United States of America* 93.22 (1996).

[105] Michael J. Chen et al. "Runx1 is required for the endothelial to haematopoietic cell transition but not thereafter". In: *Nature* 457.7231 (2009).

[106] Mervin C. Yoder. "Inducing definitive hematopoiesis in a dish". In: *Nature Biotechnology* 32.6 (2014), pp. 539–541. URL: http://dx.doi.org/10.1038/nbt.2929.

[107] C. P. Liu and R. Auerbach. "In vitro development of murine T cells from prethymic and preliver embryonic yolk sac hematopoietic stem cells". In: *Development* 113.4 (1991).

[108] Isabelle Godin, Françoise Dieterlen-Lièvre, and Ana Cumano. "Emergence of multipotent hemopoietic cells in the yolk sac and paraaortic splanchnopleura in mouse embryos, beginning at 8.5 days postcoitus". In: *Proceedings of the National Academy of Sciences of the United States of America* 92.3 (1995).

[109] Charlotta Böiers et al. "Lymphomyeloid contribution of an immune-restricted progenitor emerging prior to definitive hematopoietic stem cells." In: *Cell stem cell* 13.5 (2013).

[110] Tiago C. Luis et al. "Initial seeding of the embryonic thymus by immune-restricted lympho-myeloid progenitors". In: *Nature Immunology* 17.12 (2016).

[111] Eliver Eid Bou Ghosn et al. "Distinct B-cell lineage commitment distinguishes adult bone marrow hematopoietic stem cells". In: *Proceedings of the National Academy of Sciences of the United States of America* 109.14 (2012).

[112] A. B. Kantor et al. "Differential development of progenitor activity for three B-cell lineages". In: *Proceedings of the National Academy of Sciences of the United States of America* 89.8 (1992).

[113] Kyoko Hayakawa et al. "Progenitors for Ly-1 b cells are distinct from progenitors for other b cells". In: *Journal of Experimental Medicine* 161.6 (1985).

[114] Momoko Yoshimoto et al. "Embryonic day 9 yolk sac and intra-embryonic hemogenic endothelium independently generate a B-1 and marginal zone progenitor lacking B-2 potential." In: *Proceedings of the National Academy of Sciences of the United States of America* 108.4 (2011), pp. 1468–1473. URL: www.pnas.org/cgi/doi/10.1073/pnas.1015841108.

[115] Dirk Loeffler and Timm Schroeder. *Symmetric and asymmetric activation of hematopoietic stem cells*. 2021.

[116] Sean J. Morrison and Judith Kimble. *Asymmetric and symmetric stem-cell divisions in development and cancer*. 2006.

[117] Stuart H. Orkin and Leonard I. Zon. *Hematopoiesis: An Evolving Paradigm for Stem Cell Biology*. 2008.

[118] Adam C. Wilkinson, Kyomi J. Igarashi, and Hiromitsu Nakauchi. *Haematopoietic stem cell self-renewal in vivo and ex vivo*. 2020.

[119] M. F.T.R. de Bruijn. "Definitive hematopoietic stem cells first develop within the major arterial regions of the mouse embryo". In: *The EMBO Journal* 19.11 (2000), pp. 2465–2474. URL: http://emboj.embopress.org/content/embojnl/19/11/2465.full.pdf%20http://emboj.embopress.org/cgi/doi/10.1093/emboj/19.11.2465.

[120] Alexander Medvinsky and Elaine Dzierzak. "Definitive Hematopoiesis Is Autonomously Initiated by the AGM Region". In: *Cell* 86.6 (Sept. 1996), pp. 897–906. URL: https://www.sciencedirect.com/science/article/pii/S0092867400801658?via%3Dihub.

[121] Tomomasa Yokomizo and Elaine Dzierzak. "Three-dimensional cartography of hematopoietic clusters in the vasculature of whole mouse embryos". In: *Development* 137.21 (2010).

[122] Jean Charles Boisset et al. "In vivo imaging of haematopoietic cells emerging from the mouse aortic endothelium". In: *Nature* 464.7285 (2010).

[123] Jean Charles Boisset et al. "Progressive maturation toward hematopoietic stem cells in the mouse embryo aorta". In: *Blood* 125.3 (2015).

[124] Samir Taoudi et al. "Extensive hematopoietic stem cell generation in the AGM region via maturation of VE-Cadherin+CD45+ pre-definitive HSCs". In: *Cell Stem Cell* 3.1 (2008).

[125] Chloé S. Baron et al. "Single-cell transcriptomics reveal the dynamic of haematopoietic stem cell production in the aorta". In: *Nature Communications* 9.1 (Dec. 2018), p. 2517. URL: http://www.nature.com/articles/s41467-018-04893-3.

[126] Fan Zhou et al. "Tracing haematopoietic stem cell formation at single-cell resolution". In: *Nature* 533.7604 (2016), pp. 487–492. URL: http://dx.doi.org/10.1038/nature17997.

[127] Christos Gekas et al. "The placenta is a niche for hematopoietic stem cells". In: *Developmental Cell* 8.3 (2005).

[128] Hideo Ema and Hiromitsu Nakauchi. "Expansion of hematopoietic stem cells in the developing liver of a mouse embryo". In: *Blood* 95.7 (2000).

[129] Aldo Ciau-Uitz and Roger Patient. "The embryonic origins and genetic programming of emerging haematopoietic stem cells". In: *FEBS Letters* 590.22 (2016), pp. 4002–4015.

[130] Aldo Ciau-Uitz, Maggie Walmsley, and Roger Patient. "Distinct Origins of Adult and Embryonic Blood in Xenopus". In: *Cell* 102.6 (Sept. 2000), pp. 787–796. URL: https://www.sciencedirect.com/science/article/pii/S0092867400000672?via%3Dihub.

[131] M. Maéno, S. Tochinai, and Ch Katagiri. "Differential participation of ventral and dorsolateral mesoderms in the hemopoiesis of Xenopus, as revealed in diploid-triploid or interspecific chimeras". In: *Developmental Biology* 110.2 (1985).

[132] MITSUGU MAÉNO, ASAHI TODATE, and CHIAKI KATAGIRI. "The Localization of Precursor Cells for Larval and Adult Hemopoietic Cells of Xenopus Laevis in two Regions of Embryos: Amphibian embryo/Hemopoietic precursor cells/Chimera)". In: *Development, Growth & Differentiation* 27.2 (1985).

[133] X D Chen and J B Turpen. "Intraembryonic origin of hepatic hematopoiesis in Xenopus laevis." In: *Journal of immunology (Baltimore, Md. : 1950)* 154.6 (1995).

[134] C L Kau and J B Turpen. "Dual contribution of embryonic ventral blood island and dorsal lateral plate mesoderm during ontogeny of hemopoietic cells in Xenopus laevis." In: *Journal of immunology (Baltimore, Md. : 1950)* 131.5 (1983).

[135] Ondine Cleaver and Paul A. Krieg. "VEGF mediates angioblast migration during development of the dorsal aorta in Xenopus". In: *Development* 125.19 (1998).

[136] Michael Bressan et al. "Notochord-derived BMP antagonists inhibit endothelial cell generation and network formation". In: *Developmental Biology* 326.1 (2009).

[137] David E. Reese, Christopher E. Hall, and Takashi Mikawa. "Negative regulation of midline vascular development by the notochord". In: *Developmental Cell* 6.5 (2004).

[138] Steven A. Vokes and Paul A. Krieg. "Endoderm is required for vascular endothelial tube formation, but not for angioblast specification". In: *Development* 129.3 (2002).

[139] Georges Lacaud et al. "Haploinsufficiency of Runx1 results in the acceleration of mesodermal development and hemangioblast specification upon in vitro differentiation of ES cells". In: *Blood* 103.3 (2004).

[140] Zhongling Cai et al. "Haploinsufficiency of AML1 Affects the Temporal and Spatial Generation of Hematopoietic Stem Cells in the Mouse Embryo". In: *Immunity* 13.4 (Oct. 2000), pp. 423–431. URL: https://www.sciencedirect.com/science/article/pii/S107476130000042X.

[141] Yosuke Tanaka et al. "Circulation-independent differentiation pathway from extraembryonic mesoderm toward hematopoietic stem cells via hemogenic angioblasts". In: *Cell Reports* 8.1 (2014).

[142] C. Collart et al. "The migratory pathways of the cells that form the endocardium, dorsal aortae, and head vasculature in the mouse embryo". In: *BMC Developmental Biology* 21.1 (Dec. 2021), pp. 1–18. URL: https://doi.org/10.1186/s12861-021-00239-3.

[143] Sarah De Val and Brian L. Black. *Transcriptional Control of Endothelial Cell Development*. Feb. 2009. URL: http://www.cell.com/article/S1534580709000422/fulltext%20http://www.cell.com/article/S1534580709000422/abstract%20https://www.cell.com/developmental-cell/abstract/S1534-5807(09)00042-2.

[144] Dongjun Lee et al. "ER71 Acts Downstream of BMP, Notch, and Wnt Signaling in Blood and Vessel Progenitor Specification". In: *Cell Stem Cell* 2.5 (2008).

[145] Sarah Wareing et al. "The Flk1-Cre-Mediated Deletion of ETV2 Defines Its Narrow Temporal Requirement during Embryonic Hematopoietic Development". In: *Stem Cells* 30.7 (2012).

[146] Fang Liu et al. " Induction of hematopoietic and endothelial cell program orchestrated by ETS transcription factor ER 71/ ETV 2 ". In: *EMBO reports* 16.5 (2015).

[147] Ann N. Behrens et al. "Sox7 is regulated by ETV2 during cardiovascular development". In: *Stem Cells and Development* 23.17 (2014).

[148] Catherine Porcher et al. "The T Cell Leukemia Oncoprotein SCL/tal-1 Is Essential for Development of All Hematopoietic Lineages". In: *Cell* 86.1 (July 1996), pp. 47–57. URL: https://www.sciencedirect.com/science/article/pii/S0092867400800768.

[149] John E Pimanda et al. "Gata2, Fli1, and Scl form a recursively wired gene-regulatory circuit during early hematopoietic development". In: *Proceedings of the National Academy of Sciences* 104.45 (2007), pp. 17692–17697. URL: www.pnas.org/cgi/content/full/%20http://www.pnas.org/cgi/doi/10.1073/pnas.0707045104.

[150] Hedia Chagraoui et al. "SCL/TAL1 cooperates with Polycomb RYBP-PRC1 to suppress alternative lineages in blood-fated cells". In: *Nature Communications* 9.1 (Dec. 2018), p. 5375. URL: http://www.nature.com/articles/s41467-018-07787-6.

[151] Tõnis Org et al. "Scl binds to primed enhancers in mesoderm to regulate hematopoietic and cardiac fate divergence". In: *The EMBO Journal* 34.6 (2015).

[152] Boris Strilić et al. "The Molecular Basis of Vascular Lumen Formation in the Developing Mouse Aorta". In: *Developmental Cell* 17.4 (Oct. 2009), pp. 505–515.

[153] W. Scott Argraves et al. "VEGF signaling is required for the assembly but not the maintenance of embryonic blood vessels". In: *Developmental Dynamics* 225.3 (2002).

[154] Fouad Shalaby et al. "A requirement for Flk1 in primitive and definitive hematopoiesis and vasculogenesis". In: *Cell* 89.6 (1997).

[155] Diana C. Chong et al. "Stepwise arteriovenous fate acquisition during mammalian vasculogenesis". In: *Developmental Dynamics* 240.9 (Sept. 2011), pp. 2153–2165. URL: http://doi.wiley.com/10.1002/dvdy.22706.

[156] Luke T. Krebs et al. "Notch signaling is essential for vascular morphogenesis in mice". In: *Genes and Development* 14.11 (2000).

[157] N D Lawson et al. "Notch signaling is required for arterial-venous differentiation during embryonic vascular development." In: *Development (Cambridge, England)* 128.19 (Oct. 2001), pp. 3675–83. URL: http://www.ncbi.nlm.nih.gov/pubmed/11585794.

[158] Sarah J. Bray. *Notch signalling: A simple pathway becomes complex*. 2006.

[159] Li-Ru You et al. "Suppression of Notch signalling by the COUP-TFII transcription factor regulates vein identity". In: *Nature* 435.7038 (May 2005), pp. 98–104. URL: http://www.nature.com/articles/nature03511.

[160] Igor I. Slukvin and Gene I. Uenishi. "Arterial identity of hemogenic endothelium: a key to unlock definitive hematopoietic commitment in human pluripotent stem cell cultures". In: *Experimental Hematology* 71 (2019).

[161] Nathan D. Lawson, Andreas M. Vogel, and Brant M. Weinstein. "Sonic hedgehog and vascular endothelial growth factor act upstream of the Notch pathway during arterial endothelial differentiation". In: *Developmental Cell* 3.1 (2002).

[162] Martin Gering and Roger Patient. "Hedgehog signaling is required for adult blood stem cell formation in zebrafish embryos". In: *Developmental Cell* 8.3 (2005).

[163] Caroline E. Burns et al. "A genetic screen in zebrafish defines a hierarchical network of pathways required for hematopoietic stem cell emergence". In: *Blood* 113.23 (2009).

[164] Àlex Robert-Moreno et al. "Impaired embryonic haematopoiesis yet normal arterial development in the absence of the Notch ligand Jagged1". In: *EMBO Journal* 27.13 (2008).

[165] Keiki Kumano et al. "Notch1 but not Notch2 is essential for generating hematopoietic stem cells from endothelial cells". In: *Immunity* 18.5 (2003).

[166] Inn Inn Chen et al. "EphrinB2 regulates the emergence of a hemogenic endothelium from the aorta". In: *Scientific Reports* 6.January (2016), pp. 1–16.

[167] Masahiro Nakagawa et al. "AML1/Runx1 rescues Notch1-null mutation-induced deficiency of para-aortic splanchnopleural hematopoiesis". In: *Blood* 108.10 (2006).

[168] Elizabeth D. Howell et al. "Efficient hemogenic endothelial cell specification by RUNX1 is dependent on baseline chromatin accessibility of RUNX1-regulated TGFβ target genes". In: *Genes and Development* 35.21-22 (2021).

[169] Florian Bonkhofer et al. "Blood stem cell-forming haemogenic endothelium in zebrafish derives from arterial endothelium". In: *Nature Communications 2019 10:1* 10.1 (Aug. 2019), pp. 1–14. URL: https://www.nature.com/articles/s41467-019-11423-2.

[170] Siyuan Hou et al. "Embryonic endothelial evolution towards first hematopoietic stem cells revealed by single-cell transcriptomic and functional analyses". In: *Cell Research 2020 30:5* 30.5 (Mar. 2020), pp. 376–392. URL: https://www.nature.com/articles/s41422-020-0300-2.

[171] Muhammad Zaki Hidayatullah Fadlullah et al. "Murine AGM single-cell profiling identifies a continuum of hemogenic endothelium differentiation marked by ACE". In: *Blood* 139.3 (Jan. 2022), pp. 343–356.

[172] Mariá José Sánchez et al. "Characterization of the first definitive hematopoietic stem cells in the AGM and liver of the mouse embryo". In: *Immunity* 5.6 (1996).

[173] Ana Cumano, Françoise Dieterlen-Lievre, and Isabelle Godin. "Lymphoid potential, probed before circulation in mouse, is restricted to caudal intraembryonic splanchnopleura". In: *Cell* 86.6 (1996), pp. 907–916.

[174] Manuela Tavian et al. "The human embryo, but not its yolk sac, generates lympho-myeloid stem cells: Mapping multipotent hematopoietic cell fate in intraembryonic mesoderm". In: *Immunity* 15.3 (2001), pp. 487–495.

[175] Marella F.T.R de Bruijn et al. "Hematopoietic Stem Cells Localize to the Endothelial Cell Layer in the Midgestation Mouse Aorta". In: *Immunity* 16.5 (May 2002), pp. 673–683. URL: https://www.sciencedirect.com/science/article/pii/S1074761302003138?via%3Dihub#bBIB27.

[176] Thierry Jaffredo et al. "Intraaortic hemopoietic cells are derived from endothelial cells during ontogeny". In: *Development* 125.22 (1998).

[177] Ann C. Zovein et al. "Fate Tracing Reveals the Endothelial Origin of Hematopoietic Stem Cells". In: *Cell Stem Cell* 3.6 (2008), pp. 625–636. URL: http://dx.doi.org/10.1016/j.stem.2008.09.018.

[178] Hanna M. Eilken, Shin Ichi Nishikawa, and Timm Schroeder. "Continuous single-cell imaging of blood generation from haemogenic endothelium". In: *Nature* 457.7231 (2009).

[179] Karima Kissa and Philippe Herbomel. "Blood stem cells emerge from aortic endothelium by a novel type of cell transition". In: *Nature* 464.7285 (Mar. 2010), pp. 112–115. URL: http://www.nature.com/articles/nature08761.

[180] Julien Y. Bertrand et al. "Haematopoietic stem cells derive directly from aortic endothelium during development". In: *Nature* 464.7285 (Mar. 2010), pp. 108–111. URL: http://www.nature.com/articles/nature08738.

[181] Enid Yi Ni Lam et al. "Live imaging of Runx1 expression in the dorsal aorta tracks the emergence of blood progenitors from endothelial cells". In: *Blood* 116.6 (2010).

[182] Trista North et al. "Cbfa2 is required for the formation of intra-aortic hematopoietic clusters". In: *Development* 126.11 (1999).

[183] Monika Lichtinger et al. "RUNX1 reshapes the epigenetic landscape at the onset of haematopoiesis". In: *EMBO Journal* 31.22 (2012).

[184] Gang Huang et al. "PU.1 is a major downstream target of AML1 (RUNX1) in adult mouse hematopoiesis". In: *Nature Genetics* 40.1 (2008).

[185] Christophe Lancrin et al. "GFI1 and GFI1B control the loss of endothelial identity of hemogenic endothelium during hematopoietic commitment". In: *Blood* 120.2 (2012).

[186] Stanislav Rybtsov et al. "Tracing the origin of the HSC hierarchy reveals an SCF-dependent, IL-3-independent CD43- embryonic precursor". In: *Stem Cell Reports* 3.3 (2014).

[187] Stanislav Rybtsov et al. "Hierarchical organization and early hematopoietic specification of the developing HSC lineage in the AGM region". In: *Journal of Experimental Medicine* 208.6 (2011).

[188] Elaine Dzierzak and Anna Bigas. *Blood Development: Hematopoietic Stem Cell Dependence and Independence*. 2018.

[189] Norbert Perrimon, Chrysoula Pitsouli, and Ben Zion Shilo. "Signaling mechanisms controlling cell fate and embryonic patterning". In: *Cold Spring Harbor Perspectives in Biology* 4.8 (2012).

[190] Caroline Erter Burns et al. "Hematopoietic stem cell fate is established by the Notch-Runx pathway". In: *Genes and Development* 19.19 (2005).

[191] Bhaskar Chanda et al. "XRetinoic acid signaling is essential for embryonic hematopoietic stem cell development". In: *Cell* 155.1 (2013), pp. 215–227.

[192] Cristina Ruiz-Herguido et al. "Hematopoietic stem cell development requires transient Wnt/β-catenin activity". In: *Journal of Experimental Medicine* 209.8 (2012).

[193] Graeme M. Birdsey et al. "The endothelial transcription factor erg promotes vascular stability and growth through Wnt/β-catenin signaling". In: *Developmental Cell* 32.1 (2015).

[194] Anna Cattelino et al. "The conditional inactivation of the β-catenin gene in endothelial cells causes a defective vascular pattern and increased vascular fragility". In: *Journal of Cell Biology* 162.6 (2003).

[195] Raquel Espín-Palazón et al. "Proinflammatory signaling regulates hematopoietic stem cell emergence". In: *Cell* 159.5 (2014).

[196] Suphansa Sawamiphak, Zacharias Kontarakis, and Didier Y.R. Stainier. "Interferon Gamma Signaling Positively Regulates Hematopoietic Stem Cell Emergence". In: *Developmental Cell* 31.5 (2014).

[197] Yan Li et al. "Inflammatory signaling regulates embryonic hematopoietic stem and progenitor cell production". In: *Genes and Development* 28.23 (2014).

[198] Charles Durand et al. "Embryonic stromal clones reveal developmental regulators of definitive hematopoietic stem cells". In: *Proceedings of the National Academy of Sciences of the United States of America* 104.52 (2007).

[199] Caroline J. Marshall, Christine Kinnon, and Adrian J. Thrasher. "Polarized expression of bone morphogenetic protein-4 in the human aorta-gonad-mesonephros region". In: *Blood* 96.4 (2000).

[200] Alison C. McGarvey et al. "A molecular roadmap of the AGM region reveals BMP ER as a novel regulator of HSC maturation". In: *Journal of Experimental Medicine* 214.12 (2017).

[201] Linda T. Vo and George Q. Daley. "De novo generation of HSCs from somatic and pluripotent stem cell sources". In: *Blood* 125.17 (2015).

[202] Connor Sweeney and Paresh Vyas. *The Graft-Versus-Leukemia Effect in AML*. 2019.

[203] Adam C. Wilkinson et al. "Long-term ex vivo haematopoietic-stem-cell expansion allows nonconditioned transplantation". In: *Nature* 571.7763 (2019).

[204] Vladislav M. Sandler et al. "Reprogramming human endothelial cells to haematopoietic cells requires vascular induction". In: *Nature* 511.7509 (2014).

[205] Sergei Doulatov et al. "Induction of multipotential hematopoietic progenitors from human pluripotent stem cells via respecification of lineage-restricted precursors". In: *Cell Stem Cell* 13.4 (2013).

[206] Kiran Batta et al. "Direct Reprogramming of Murine Fibroblasts to Hematopoietic Progenitor Cells". In: *Cell Reports* 9.5 (2014).

[207] Carlos Filipe Pereira et al. "Induction of a hemogenic program in mouse fibroblasts". In: *Cell Stem Cell* 13.2 (2013).

[208] Michael Kyba, Rita C.R. Perlingeiro, and George Q. Daley. "HoxB4 confers definitive lymphoid-myeloid engraftment potential on embryonic stem cell and yolk sac hematopoietic progenitors". In: *Cell* 109.1 (2002).

[209] Irina Elcheva et al. "Direct induction of haematoendothelial programs in human pluripotent stem cells by transcriptional regulators". In: *Nature Communications* 5 (2014).

[210] Giovanni Amabile et al. "In vivo generation of transplantable human hematopoietic cells from induced pluripotent stem cells". In: *Blood* 121.8 (2013).

[211] Nao Suzuki et al. "Generation of engraftable hematopoietic stem cells from induced pluripotent stem cells by way of teratoma formation". In: *Molecular Therapy* 21.7 (2013).

[212] Andrea Ditadi and Christopher M. Sturgeon. "Directed differentiation of definitive hemogenic endothelium and hematopoietic progenitors from human pluripotent stem cells". In: *Methods* 101 (2016), pp. 65–72.

[213] Elizabeth S. Ng et al. "Differentiation of human embryonic stem cells to HOXA + hemogenic vasculature that resembles the aorta-gonad-mesonephros". In: *Nature Biotechnology* 34.11 (2016), pp. 1168–1179.

[214] Christopher M. Sturgeon et al. "Wnt signaling controls the specification of definitive and primitive hematopoiesis from human pluripotent stem cells". In: *Nature Biotechnology* 32.6 (2014), pp. 554–561. URL: http://dx.doi.org/10.1038/nbt.2915.

[215] Stephanie A. Luff et al. "Identification of a retinoic acid-dependent haemogenic endothelial progenitor from human pluripotent stem cells". In: *Nature Cell Biology 2022* (Apr. 2022), pp. 1–9. URL: https://www.nature.com/articles/s41556-022-00898-9.

[216] Vincenzo Calvanese et al. "Mapping human haematopoietic stem cells from haemogenic endothelium to birth". In: *Nature 2022* (Apr. 2022), pp. 1–7. URL: https://www.nature.com/articles/s41586-022-04571-x.

[217] Andrea Ditadi, Christopher M. Sturgeon, and Gordon Keller. "A view of human haematopoietic development from the Petri dish". In: *Nature Reviews Molecular Cell Biology* 18.1 (2016), pp. 56–67. URL: http://dx.doi.org/10.1038/nrm.2016.127.

[218] Vivien W. Ho et al. "Specification of the haematopoietic stem cell lineage: From blood-fated mesodermal angioblasts to haemogenic endothelium". In: *Seminars in Cell & Developmental Biology* (Feb. 2022). URL: https://linkinghub.elsevier.com/retrieve/pii/S1084952122000167.

[219] Kausalia Vijayaragavan et al. "Noncanonical Wnt Signaling Orchestrates Early Developmental Events toward Hematopoietic Cell Fate from Human Embryonic Stem Cells". In: *Cell Stem Cell* 4.3 (2009).

[220] Elizabeth S Ng et al. "Forced aggregation of defined numbers of human embryonic stem cells into embryoid bodies fosters robust , reproducible hematopoietic differentiation Brief report Forced aggregation of defined numbers of human embryonic stem cells into embryoid bodies fost". In: 106.5 (2011), pp. 1601–1603.

[221] Marion Kennedy et al. "T Lymphocyte Potential Marks the Emergence of Definitive Hematopoietic Progenitors in Human Pluripotent Stem Cell Differentiation Cultures". In: *Cell Reports* 2.6 (Dec. 2012), pp. 1722–1735. URL: https://www.sciencedirect.com/science/article/pii/S2211124712003841?via%3Dihub.

[222] Patrycja Sroczynska et al. "The differential activities of Runx1 promoters define milestones during embryonic hematopoiesis". In: *Blood* 114.26 (2009).

[223] Andrea Ditadi et al. "Human definitive haemogenic endothelium and arterial vascular endothelium represent distinct lineages". In: *Nature Cell Biology* 17.5 (2015), pp. 580–591.

[224] Mi Ae Park et al. "Activation of the Arterial Program Drives Development of Definitive Hemogenic Endothelium with Lymphoid Potential". In: *Cell Reports* 23.8 (2018), pp. 2467–2481. URL: https://doi.org/10.1016/j.celrep.2018.04.092.

[225] Gene I. Uenishi et al. "NOTCH signaling specifies arterial-type definitive hemogenic endothelium from human pluripotent stem cells". In: *Nature Communications* 9.1 (Dec. 2018), p. 1828. URL: http://www.nature.com/articles/s41467-018-04134-7.

[226] Monica Corada et al. "Sox17 is indispensable for acquisition and maintenance of arterial identity". In: *Nature Communications* 4 (2013), pp. 1–14. URL: http://dx.doi.org/10.1038/ncomms3609.

[227] Thomas J. Poole, Eric B. Finkelstein, and Christopher M. Cox. *The role of FGF and VEGF in angioblast induction and migration during vascular development*. 2001.

[228] Kyung Dal Choi et al. "Identification of the Hemogenic Endothelial Progenitor and Its Direct Precursor in Human Pluripotent Stem Cell Differentiation Cultures". In: *Cell Reports* 2.3 (2012), pp. 553–567. URL: http://dx.doi.org/10.1016/j.celrep.2012.08.002.

[229] Michael J. Chen et al. "Erythroid/myeloid progenitors and hematopoietic stem cells originate from distinct populations of endothelial cells". In: *Cell Stem Cell* 9.6 (2011).

[230] Kenna R. Mills, Deborah Kruep, and Margaret S. Saha. "Elucidating the origins of the vascular system: A fate map of the vascular endothelial and red blood cell lineages in Xenopus laevis". In: *Developmental Biology* 209.2 (1999).

[231] Daniela Raciti et al. "Organization of the pronephric kidney revealed by large-scale gene expression mapping". In: *Genome Biology* 9.5 (2008).

[232] F. C. Serluca and M. C. Fishman. "Pre-pattern in the pronephric kidney field of zebrafish". In: *Development* 128.12 (2001).

[233] Rebecca A. Wingert et al. "The cdx genes and retinoic acid control the positioning and segmentation of the zebrafish pronephros". In: *PLoS Genetics* 3.10 (2007).

[234] Juan A. Garcia-Porrero et al. "Antigenic profiles of endothelial and hemopoietic lineages in murine intraembryonic hemogenic sites". In: *Developmental and Comparative Immunology* 22.3 (1998).

[235] Edroaldo Lummertz da Rocha et al. "CellComm infers cellular crosstalk that drives haematopoietic stem and progenitor cell development". In: *Nature Cell Biology* 24.4 (Apr. 2022), pp. 579–589. URL: https://www.nature.com/articles/s41556-022-00884-1.

[236] Axel A. Almet et al. *The landscape of cell–cell communication through single-cell transcriptomics*. 2021.

[237] Erick Armingol et al. *Deciphering cell–cell interactions and communication from gene expression*. 2021.

[238] Harry M.T. Choi, Victor A. Beck, and Niles A. Pierce. "Next-generation in situ hybridization chain reaction: Higher gain, lower cost, greater durability". In: *ACS Nano* 8.5 (2014).

[239] Harry M.T. Choi et al. "Third-generation in situ hybridization chain reaction: Multiplexed, quantitative, sensitive, versatile, robust". In: *Development (Cambridge)* 145.12 (June 2018).

[240] Fay Wang et al. "RNAscope: A novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues". In: *Journal of Molecular Diagnostics* 14.1 (2012).

[241] Eric Lubeck et al. *Single-cell in situ RNA profiling by sequential hybridization*. 2014.

[242] Chee Huat Linus Eng et al. "Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+". In: *Nature* 568.7751 (2019), pp. 235–239.

[243] Kok Hao Chen et al. "Spatially resolved, highly multiplexed RNA profiling in single cells". In: *Science* 348.6233 (2015).

[244] Michaela Asp, Joseph Bergenstråhle, and Joakim Lundeberg. *Spatially Resolved Transcriptomes—Next Generation Tools for Tissue Exploration*. 2020.

[245] Malgorzata Kloc, N. Ruth Zearfoss, and Laurence D. Etkin. *Mechanisms of subcellular mRNA localization*. 2002.

[246] Vitalii Kleshchevnikov et al. "Cell2location maps fine-grained cell types in spatial transcriptomics". In: *Nature Biotechnology* (2022).

[247] Yang Liu et al. "High-Spatial-Resolution Multi-Omics Sequencing via Deterministic Barcoding in Tissue". In: *Cell* 183.6 (2020).

[248] Ao Chen et al. "Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball patterned arrays". In: *bioRxiv* (2021).

[249] Susanne Nichterwitz et al. "LCM-seq: A method for spatial transcriptomic profiling using laser capture microdissection coupled with PolyA-based RNA sequencing". In: *Methods in Molecular Biology*. Vol. 1649. 2018.

[250] Jun Chen et al. "Spatial transcriptomic analysis of cryosectioned tissue samples with Geo-seq". In: *Nature Protocols* 12.3 (2017).

[251] Christopher J. Drake and Paul A. Fleming. "Vasculogenesis in the day 6.5 to 9.5 mouse embryo". In: *Blood* 95.5 (2000).

[252] Thorsten M. Schlaeger et al. "Tie2Cre-mediated gene ablation defines the stem-cell leukemia gene (SCL/tal1)-dependent window during hematopoietic stem-cell development". In: *Blood* 105.10 (2005).

[253] Minoru Takasato et al. "Generation of kidney organoids from human pluripotent stem cells". In: *Nature Protocols* 11.9 (2016), pp. 1681–1692. URL: http://dx.doi.org/10.1038/nprot.2016.098.

[254] Yun Xia et al. "Directed Differentiation of Human Pluripotent Cells to Ureteric Bud Kidney Progenitor-Like Cells". In: *Nature Cell Biology* 15.12 (2013), pp. 1507–1515. URL: http://dx.doi.org/10.1038/ncb2872.

[255] Jonathan B.l. Bard et al. "An internet-accessible database of mouse developmental anatomy based on a systematic nomenclature". In: *Mechanisms of Development* 74.1-2 (1998).

[256] Terry F. Hayamizu et al. "EMAP/EMAPA ontology of mouse developmental anatomy: 2013 update". In: *Journal of Biomedical Semantics* 4.1 (2013).

[257] Richard C. V. Tyser et al. "Characterization of a common progenitor pool of the epicardium and myocardium". In: *Science* 2986.January (Jan. 2021), eabb2986. URL: https://www.sciencemag.org/lookup/doi/10.1126/science.abb2986.

[258] Antti Lignell and Laura Kerosuo. "Spatial genomic analysis: A multiplexed transcriptional profiling method that reveals subpopulations of cells within intact tissues". In: *Methods in Molecular Biology*. Vol. 2002. Humana Press Inc., 2019, pp. 151–163.

[259] Michael Hagemann-Jensen et al. "Single-cell RNA counting at allele and isoform resolution using Smart-seq3". In: *Nature Biotechnology* 38.6 (2020).

[260] Grace X.Y. Zheng et al. "Massively parallel digital transcriptional profiling of single cells". In: *Nature Communications* 8 (2017).

[261] Tim Stuart et al. "Comprehensive Integration of Single-Cell Data". In: *Cell* 177.7 (2019).

[262] Suoqin Jin et al. "Inference and analysis of cell-cell communication using CellChat". In: *Nature Communications* 12.1 (2021).

[263] Minoru Takasato and Melissa H. Little. "A strategy for generating kidney organoids: Recapitulating the development in human pluripotent stem cells". In: *Developmental Biology* 420.2 (2016), pp. 210–220. URL: http://dx.doi.org/10.1016/j.ydbio.2016.08.024.

[264] James A Briggs et al. "The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution". In: *Science* 360.6392 (Apr. 2018), eaar5780. URL: http://www.ncbi.nlm.nih.gov/pubmed/29700227%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6038144.

[265] Daniel E. Wagner et al. "Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo". In: *Science* 360.6392 (2018), pp. 981–987.

[266] Minoru Takasato et al. "Kidney organoids from human iPS cells contain multiple lineages and model human nephrogenesis". In: *Nature* 526.7574 (2015), pp. 564–568.

[267] Benjamin S. Freedman et al. "Modelling kidney disease with CRISPR-mutant kidney organoids derived from human pluripotent epiblast spheroids". In: *Nature Communications* 6 (2015).

[268] Atsuhiro Taguchi and Ryuichi Nishinakamura. *Nephron reconstitution from pluripotent stem cells*. 2015.

[269] Shin Ichi Mae et al. "Monitoring and robust induction of nephrogenic intermediate mesoderm from human pluripotent stem cells". In: *Nature Communications* 4 (2013).

[270] Minyong Kang and Yong Mahn Han. "Differentiation of human pluripotent stem cells into nephron progenitor cells in a serum and feeder free system". In: *PLoS ONE* 9.4 (2014).

[271] Karthikeyan Narayanan et al. "Human embryonic stem cells differentiate into functional renal proximal tubular-like cells". In: *Kidney International* 83.4 (2013).

[272] Ryuichi Nishinakamura. "Human kidney organoids: progress and remaining challenges". In: *Nature Reviews Nephrology 2019 15:10* 15.10 (Aug. 2019), pp. 613–624. URL: https://www.nature.com/articles/s41581-019-0176-x.

[273] David G. Wilkinson, Sangita Bhatt, and Bernhard G. Herrmann. "Expression pattern of the mouse T gene and its role in mesoderm formation". In: *Nature* 343.6259 (1990).

[274] Jonathan J.H. Pearce and Martin J. Evans. "Mml, a mouse Mix-like gene expressed in the primitive streak". In: *Mechanisms of Development* 87.1-2 (1999).

[275] Xiaoqing Zhang et al. "Pax6 is a human neuroectoderm cell fate determinant". In: *Cell Stem Cell* 7.1 (July 2010), pp. 90–100. URL: http://www.cell.com/article/S1934590910001724/fulltext%20http://www.cell.com/article/S1934590910001724/abstract%20https://www.cell.com/cell-stem-cell/abstract/S1934-5909(10)00172-4.

[276] D C Weinstein et al. "The winged-helix transcription factor HNF-3 beta is required for notochord development in the mouse embryo". In: *Cell* 78.4 (1994).

[277] S. L. Ang et al. "The formation and maintenance of the definitive endoderm lineage in the mouse: Involvement of HNF3/forkhead proteins". In: *Development* 119.4 (1993).

[278] Fouad Shalaby et al. *Failure of blood-island formation and vasculogenesis in Flk-1-deficient mice*. 1995.

[279] Basharat Jahan and Kara E. McCloskey. "Differentiation and expansion of endothelial cells requires pre-optimization of KDR+ expression kinetics". In: *Stem Cell Research* 42 (2020).

[280] Tong Chen et al. "Stromal Cell-Derived Factor-1/CXCR4 Signaling Modifies the Capillary-Like Organization of Human Embryonic Stem Cell-Derived Endothelium In Vitro". In: *Stem Cells* 25.2 (2007).

[281] Denis Evseenko et al. "Mapping the fi rst stages of mesoderm commitment during differentiation of human embryonic stem cells". In: *Pnas* 3.31 (2010), pp. 13742–13747.

[282] Aldo Ciau-Uitz et al. "Developmental hematopoiesis: Ontogeny, genetic programming and conservation". In: *Experimental Hematology* 42.8 (2014), pp. 669–683. URL: http://dx.doi.org/10.1016/j.exphem.2014.06.001.

[283] Y. Shing et al. "Heparin affinity: Purification of a tumor-derived capillary endothelial cell growth factor". In: *Science* 223.4642 (1984).

[284] Yng Jyh Jih et al. "Distinct regulation of genes by bFGF and VEGF-A in endothelial cells". In: *Angiogenesis* 4.4 (2001).

[285] E. Preger-Ben Noon et al. "Interplay between activin and Hox genes determines the formation of the kidney morphogenetic field". In: *Development* 136.12 (2009), pp. 1995–2004. URL: http://dev.biologists.org/cgi/doi/10.1242/dev.035592.

[286] Ryohichi Sugimura et al. "Haematopoietic stem and progenitor cells from human pluripotent stem cells". In: *Nature* 545.7655 (2017), pp. 432–438. URL: http://dx.doi.org/10.1038/nature22370.

[287] Satoko Nishimoto and Malcolm P.O. Logan. "Subdivision of the lateral plate mesoderm and specification of the forelimb and hindlimb forming domains". In: *Seminars in Cell and Developmental Biology* 49 (2016), pp. 102–108.

[288] Carlos O. Lizama et al. "Repression of arterial genes in hemogenic endothelium is sufficient for haematopoietic fate acquisition". In: *Nature Communications* 6.May (2015), pp. 1–10. URL: http://dx.doi.org/10.1038/ncomms8739.

[289] S-i Nishikawa. *DEV125-9*. Tech. rep. 1998. URL: http://dev.biologists.org/content/develop/125/9/1747.full.pdf.

[290] Stuart T. Fraser et al. "Definitive hematopoietic commitment within the embryonic vascular endothelial-cadherin+ population". In: *Experimental Hematology* 30.9 (2002).

[291] Tomomasa Yokomizo et al. "Whole-mount three-dimensional imaging of internally localized immunostained cells within mouse embryos". In: *Nature Protocols* 7.3 (Mar. 2012), pp. 421–423. URL: https://www.nature.com/articles/nprot.2011.441.

[292] Charlotte Richard et al. "Endothelio-Mesenchymal Interaction Controls runx1 Expression and Modulates the notch Pathway to Initiate Aortic Hematopoiesis". In: *Developmental Cell* 24.6 (2013).

[293] Marian Peeters et al. "Ventral embryonic tissues and Hedgehog proteins induce early AGM hematopoietic stem cell development". In: *Development* 136.15 (2009).

[294] Céline Souilhol et al. "Inductive interactions mediated by interplay of asymmetric signalling underlie development of adult haematopoietic stem cells". In: *Nature Communications* 7 (2016), pp. 1–13.

[295] Samir Taoudi and Alexander Medvinsky. "Functional identification of the hematopoietic stem cell niche in the ventral domain of the embryonic dorsal aorta". In: *Proceedings of the National Academy of Sciences of the United States of America* 104.22 (2007).

[296] Andrew Przysinda, Wei Feng, and Guang Li. "Diversity of Organism-Wide and Organ-Specific Endothelial Cells". In: *Current Cardiology Reports* 22.4 (Apr. 2020), pp. 1–6. URL: https://link.springer.com/article/10.1007/s11886-020-1275-9.

[297] Linden E. Craig et al. "Endothelial Cells from Diverse Tissues Exhibit Differences in Growth and Morphology". In: *Microvascular Research* 55.1 (Jan. 1998), pp. 65–76.

[298] Jen Tsan Chi et al. "Endothelial cell diversity revealed by global expression profiling". In: *Proceedings of the National Academy of Sciences of the United States of America* 100.19 (Sept. 2003), pp. 10623–10628. URL: https://pubmed.ncbi.nlm.nih.gov/12963823/.

[299] Clémence Carron and De Li Shi. "Specification of anteroposterior axis by combinatorial signaling during Xenopus development". In: *Wiley Interdisciplinary Reviews: Developmental Biology* 5.2 (Mar. 2016), pp. 150–168. URL: https://onlinelibrary.wiley.com/doi/full/10.1002/wdev.217%20https://onlinelibrary.wiley.com/doi/abs/10.1002/wdev.217%20https://wires.onlinelibrary.wiley.com/doi/10.1002/wdev.217.

[300] Vincent H. Gattone and Daniel Goldowitz. "The Renal Glomerulus and Vasculature in 'Aggregation' Chimeric Mice". In: *Nephron* 90.3 (2002), pp. 267–272. URL: https://www.karger.com/Article/FullText/49062%20https://www.karger.com/Article/Abstract/49062.

[301] Andrejs Ivanovs et al. "Identification of the niche and phenotype of the first human hematopoietic stem cells". In: *Stem Cell Reports* 2.4 (2014), pp. 449–456. URL: http://dx.doi.org/10.1016/j.stemcr.2014.02.004.

[302] Lili Jing et al. "Adenosine signaling promotes hematopoietic stem and progenitor cell emergence". In: *Journal of Experimental Medicine* 212.5 (2015).

[303] Peter Geon Kim et al. "Signaling axis involving Hedgehog, Notch, and Scl promotes the embryonic endothelial-to-hematopoietic transition". In: *Proceedings of the National Academy of Sciences of the United States of America* 110.2 (2013).

[304] Anna Bigas, Jordi Guiu, and Leonor Gama-Norton. "Notch and Wnt signaling in the emergence of hematopoietic stem cells". In: *Blood Cells, Molecules, and Diseases* 51.4 (2013).

[305] Kathrina L. Marcelo et al. "Hemogenic endothelial cell specification requires c-Kit, notch signaling, and p27-mediated cell-cycle control". In: *Developmental Cell* 27.5 (2013).

[306] Elena Cano et al. *Signaling by retinoic acid in embryonic and adult hematopoiesis*. 2014.

[307] Catherine Robin et al. "An Unexpected Role for IL-3 in the Embryonic Development of Hematopoietic Stem Cells". In: *Developmental Cell* 11.2 (2006).

[308] Florent Ginhoux and Martin Guilliams. *Tissue-Resident Macrophage Ontogeny and Homeostasis*. 2016.

[309] Sara Menegatti et al. "Transcriptional control of blood cell emergence". In: *FEBS Letters* 593.23 (Dec. 2019), pp. 3304–3315. URL: https://onlinelibrary.wiley.com/doi/full/10.1002/1873-3468.13585%20https://onlinelibrary.wiley.com/doi/abs/10.1002/1873-3468.13585%20https://febs.onlinelibrary.wiley.com/doi/10.1002/1873-3468.13585.

[310] Antti Lignell et al. "Identification of a neural crest stem cell niche by Spatial Genomic Analysis". In: *Nature Communications* 8.1 (Dec. 2017), p. 1830. URL: http://www.nature.com/articles/s41467-017-01561-w.

[311] Joe Chalfoun et al. "MIST: Accurate and Scalable Microscopy Image Stitching Tool with Stage Modeling and Error Minimization". In: *Scientific Reports* 7.1 (2017).

[312] Simone Codeluppi et al. "Spatial organization of the somatosensory cortex revealed by osmFISH". In: *Nature Methods* 15.11 (Nov. 2018), pp. 932–935.

[313] Lorna Richardson et al. "EMAGE mouse embryo spatial gene expression database: 2014 update". In: *Nucleic Acids Research* 42.D1 (2014).

[314] Elizabeth Graham et al. "The atlas of mouse development eHistology resource". In: *Development (Cambridge)* 142.11 (2015).

[315] Elliot A Perens et al. "Hand2 inhibits kidney specification while promoting vein formation within the posterior mesoderm". In: *eLife* 5.NOVEMBER2016 (Nov. 2016). URL: https://elifesciences.org/articles/19941.

[316] Gregory R. Dressler. "The Cellular Basis of Kidney Development". In: *Annual Review of Cell and Developmental Biology* 22.1 (2006), pp. 509–529. URL: http://www.annualreviews.org/doi/10.1146/annurev.cellbio.22.010305.104340.

[317] Siyeon Rhee et al. "Visceral Endoderm Expression of Yin-Yang1 (YY1) Is Required for VEGFA Maintenance and Yolk Sac Development". In: *PLoS ONE* 8.3 (2013).

[318] Louise Silver and James Palis. "Initiation of murine embryonic erythropoiesis: A spatial analysis". In: *Blood* 89.4 (1997).

[319] Lorraine Robb et al. *Absence of yolk sac hematopoiesis from mice with a targeted disruption of the sci gene (tal-l gene/gene targeting/homologous recombination)*. Tech. rep. 1995, pp. 7075–7079.

[320] L Robb et al. "Absence of yolk sac hematopoiesis from mice with a targeted disruption of the scl gene." In: *Proceedings of the National Academy of Sciences of the United States of America* 92.15 (July 1995), pp. 7075–9. URL: http://www.ncbi.nlm.nih.gov/pubmed/7624372%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC41474.

[321] Wen Zhou et al. "Single-Cell Analysis Reveals Regulatory Gene Expression Dynamics Leading to Lineage Commitment in Early T Cell Development". In: *Cell Systems* 9.4 (Oct. 2019), pp. 321–337.

[322] Samuel G. Rodriques et al. "Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution". In: *Science* 363.6434 (2019).

[323] Susanne Nichterwitz et al. "Laser capture microscopy coupled with Smart-seq2 for precise spatial transcriptomic profiling". In: *Nature Communications* 7 (2016).

[324] Mirjana Efremova et al. "CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes". In: *Nature Protocols* 15.4 (2020).

[325] Anwarul Ferdous et al. "Nkx2-5 transactivates the Ets-related protein 71 gene and specifies an endothelial/endocardial fate in the developing embryo". In: *Proceedings of the National Academy of Sciences of the United States of America* 106.3 (2009).

[326] Stefan C. Materna et al. "Cardiovascular development and survival require Mef2c function in the myocardial but not the endothelial lineage". In: *Developmental Biology* 445.2 (2019).

[327] Tara L. Rasmussen et al. "ER71 directs mesodermal fate decisions during embryogenesis". In: *Development* 138.21 (Nov. 2011), pp. 4801–4812.

[328] Linda Madisen et al. "A robust and high-throughput Cre reporting and characterization system for the whole mouse brain". In: *Nature Neuroscience* 13.1 (2010).

[329] Mandar Deepak Muzumdar et al. "A global double-fluorescent cre reporter mouse". In: *Genesis* 45.9 (2007).

[330] Soma Datta et al. *Laser capture microdissection: Big data from small samples*. Nov. 2015. URL: /pmc/articles/PMC4665617/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4665617/.

[331] Ali S. Hassan et al. "Expression of two novel transcripts in the mouse definitive endoderm". In: *Gene Expression Patterns* 10.2-3 (2010).

[332] Laina Freyer et al. "Loss of Apela Peptide in Mice Causes Low Penetrance Embryonic Lethality and Defects in Early Mesodermal Derivatives". In: *Cell Reports* 20.9 (2017).

[333] Juan Hou et al. "A regulatory network controls nephrocan expression and midgut patterning". In: *Development (Cambridge)* 141.19 (2014).

[334] Masami Kanai-Azuma et al. "Depletion of definitive gut endoderm in Sox17-null mutant mice". In: *Development* 129.10 (2002).

[335] Tetsuya Fujii et al. "Expression patterns of the murine LIM class homeobox gene lim1 in the developing brain and excretory system". In: *Developmental Dynamics* 199.1 (1994).

[336] Catherine Wolf et al. "The M-twist gene of Mus is expressed in subsets of mesodermal cells and is closely related to the Xenopus X-twi and the Drosophila twist genes". In: *Developmental Biology* 143.2 (1991).

[337] Gare Hoon Yeo et al. "Zebrafish twist1 is expressed in craniofacial, vertebral, and renal precursors". In: *Development Genes and Evolution* 217.11-12 (2007).

[338] Ralston M. Barnes et al. "Analysis of the Hand1 cell lineage reveals novel contributions to cardiovascular, neural crest, extra-embryonic, and lateral mesoderm derivatives". In: *Developmental Dynamics* 239.11 (2010).

[339] Michael A. Blanar et al. "Meso1, a basic-helix-loop-helix protein involved in mammalian presomitic mesoderm development". In: *Proceedings of the National Academy of Sciences of the United States of America* 92.13 (1995).

[340] Rob Burgess et al. "Requirement of the paraxis gene for somite formation and musculoskeletal patterning". In: *Nature* 384.6609 (1996).

[341] Susan Skuntz et al. "Lack of the mesodermal homeodomain protein MEOX1 disrupts sclerotome polarity and leads to a remodeling of the cranio-cervical joints of the axial skeleton". In: *Developmental Biology* 332.2 (2009).

[342] T. Kume et al. "The murine winged helix transcription factors, Foxc1 and Foxc2, are both required for cardiovascular development and somitogenesis". In: *Genes and Development* 15.18 (2001).

[343] Glenn E. Winnier, Linda Hargett, and Brigid L.M. Hogan. "The winged helix transcription factor MFH1 is required for proliferation and patterning of paraxial mesoderm in the mouse embryo". In: *Genes and Development* 11.7 (1997).

[344] Y. Bessho et al. "Hes7: A bHLH-type repressor gene regulated by Notch and expressed in the presomitic mesoderm". In: *Genes to Cells* 6.2 (2001).

[345] Martin Hrabě De Angelis, Joseph McIntyre, and Achim Gossler. "Maintenance of somite borders in mice requires the Delta homologue Dll1". In: *Nature* 386.6626 (1997).

[346] Aida Rodrigo Albors, Pamela A. Halley, and Kate G. Storey. "Lineage tracing of axial progenitors using nkx1-2creert2 mice defines their trunk and tail contributions". In: *Development (Cambridge)* 145.19 (2018).

[347] A. Frumkin et al. "A chicken caudal homologue, CHox-cad, expressed in the epiblast with posterior localization and in the early endodermal lineage". In: *Development* 112.1 (1991).

[348] Heather B. Wood and Vasso Episkopou. "Comparative expression of the mouse Sox1, Sox2 and Sox3 genes from pre-gastrulation to early somite stages". In: *Mechanisms of Development* 86.1-2 (1999).

[349] Shuchen Zhang. "Sox2, a key factor in the regulation of pluripotency and neural differentiation". In: *World Journal of Stem Cells* 6.3 (2014).

[350] J. Ericson et al. "Pax6 controls progenitor cell identity and neuronal fate in response to graded Shh signaling". In: *Cell* 90.1 (1997).

[351] Poongodi Geetha-Loganathan et al. "Regulation of ectodermal Wnt6 expression by the neural tube is transduced by dermomyotomal Wnt11: A mechanism of dermomyotomal lip sustainment". In: *Development* 133.15 (2006).

[352] Marc Rodríguez-Niedenführ et al. "Spatial and temporal pattern of Wnt-6 expression during chick development". In: *Anatomy and Embryology* 206.6 (2003).

[353] Ying Qu et al. "Transcriptome and proteome characterization of surface ectoderm cells differentiated from human iPSCs". In: *Scientific Reports* 6 (2016).

[354] Christopher S. McGinnis, Lyndsay M. Murrow, and Zev J. Gartner. "DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors". In: *Cell Systems* 8.4 (2019).

[355] Pierre-Luc Germain et al. "Doublet identification in single-cell sequencing data using scDblFinder". In: *F1000Research* 10 (2021).

[356] Nan Miles Xi and Jingyi Jessica Li. "Benchmarking Computational Doublet-Detection Methods for Single-Cell RNA Sequencing Data". In: *Cell Systems* 12.2 (2021).

[357] Renaud Gaujoux and Cathal Seoighe. "A flexible R package for nonnegative matrix factorization". In: *BMC Bioinformatics* 11 (2010).

[358] Timothy R. Carlson et al. "Cell-autonomous requirement for β1 integrin in endothelial cell adhesion, migration and survival during angiogenesis in mice". In: *Development* 135.12 (June 2008), pp. 2193–2202. URL: https://journals.biologists.com/dev/article/135/12/2193/64753/Cell-autonomous-requirement-for-1-integrin-in.

[359] Sylvie Ricard-Blum. "The Collagen Family". In: *Cold Spring Harbor Perspectives in Biology* 3.1 (2011).

[360] Nathaniel Lal, Karanjit Puri, and Brian Rodrigues. *Vascular Endothelial Growth Factor B and Its Signaling*. 2018.

[361] Karin Aase et al. "Vascular endothelial growth factor-B-deficient mice display an atrial conduction defect". In: *Circulation* 104.3 (2001).

[362] D. Bellomo et al. "Mice lacking the vascular endothelial growth factor-B gene (Vegfb) have smaller hearts, dysfunctional coronary vasculature, and impaired recovery from cardiac ischemia." In: *Circulation research* 86.2 (2000).

[363] Tianxia Jiang et al. "CD146 is a coreceptor for VEGFR-2 in tumor angiogenesis". In: *Blood* 120.11 (2012).

[364] Zhongde Ye et al. "Wnt5a uses CD146 as a receptor to regulate cell motility and convergent extension". In: *Nature Communications* 4 (2013).

[365] Xiaomei Li et al. "Wnt5a promotes renal tubular inflammation in diabetic nephropathy by binding to CD146 through noncanonical Wnt signaling". In: *Cell Death and Disease* 12.1 (2021).

[366] Andrew T. Dudley, Karen M. Lyons, and Elizabeth J. Robertson. "A requirement for bone morphogenetic protein-7 during development of the mammalian kidney and eye". In: *Genes and Development* 9.22 (1995).

[367] Guangbin Luo et al. "BMP-7 is an inducer of nephrogenesis, and is also required for eye development and skeletal patterning". In: *Genes and Development* 9.22 (1995).

[368] Aikaterini Segklia et al. "Bmp7 regulates the survival, proliferation, and neurogenic properties of neural progenitor cells during corticogenesis in the mouse". In: *PLoS ONE* 7.3 (2012).

[369] Toshihiko Toyofuku et al. "Semaphorin-4A, an activator for T-cell-mediated immunity, suppresses angiogenesis via Plexin-D1". In: *EMBO Journal* 26.5 (2007).

[370] Chenghua Gu et al. "Semaphorin 3E and plexin-D1 control vascular pattern independently of neuropilins". In: *Science* 307.5707 (2005).

[371] Vedanta Mehta et al. "The guidance receptor plexin D1 is a mechanosensor in endothelial cells". In: *Nature* 578.7794 (2020).

[372] Emanuele Azzoni et al. "The onset of circulation triggers a metabolic switch required for endothelial to hematopoietic transition". In: *Cell Reports* 37.11 (Dec. 2021), p. 110103.

[373] Miguel F. Diaz et al. "Biomechanical forces promote blood development through prostaglandin E2 and the cAMP-PKA signaling axis". In: *Journal of Experimental Medicine* 212.5 (2015).

[374] Peter Geon Kim et al. "Flow-induced protein kinase A-CREB pathway acts via BMP signaling to promote HSC emergence". In: *Journal of Experimental Medicine* 212.5 (2015).

[375] Vanessa Lundin et al. "YAP Regulates Hematopoietic Stem Cell Formation in Response to the Biomechanical Forces of Blood Flow". In: *Developmental Cell* 52.4 (2020).

[376] Qi Gao et al. "Thrombospondin-1 signaling through CD47 inhibits cell cycle progression and induces senescence in endothelial cells". In: *Cell Death and Disease* 7.9 (2016).

[377] Alexander Zaslavsky et al. "Platelet-derived thrombospondin-1 is a critical negative regulator and potential biomarker of angiogenesis". In: *Blood* 115.22 (2010).

[378] Giovanni Canu et al. "Analysis of endothelial-to-haematopoietic transition at the single cell level identifies cell cycle regulation as a driver of differentiation". In: *Genome Biology* 21.1 (2020).

[379] Ryan T. Dackor et al. " Hydrops Fetalis, Cardiovascular Defects, and Embryonic Lethality in Mice Lacking the Calcitonin Receptor-Like Receptor Gene ". In: *Molecular and Cellular Biology* 26.7 (2006).

[380] Joseph W. Foley et al. "Gene expression profiling of single cells from archival tissue with laser-capture microdissection and Smart-3SEQ". In: *Genome Research* 29.11 (2019).

[381] Catherine Porcher, Hedia Chagraoui, and Maiken S Kristiansen. *SCL/TAL1: A multifaceted regulator from blood development to disease*. Apr. 2017. URL: http://www.ncbi.nlm.nih.gov/pubmed/28179281.

[382] Marella de Bruijn and Elaine Dzierzak. "Runx transcription factors in the development and function of the definitive hematopoietic system." In: *Blood* 129.15 (2017), pp. 2061–2069. URL: www.bloodjournal.org%20http://www.ncbi.nlm.nih.gov/pubmed/28179276.

[383] Katrin E. Rhodes et al. "The Emergence of Hematopoietic Stem Cells Is Initiated in the Placental Vasculature in the Absence of Circulation". In: *Cell Stem Cell* 2.3 (2008).

[384] Zhuan Li et al. "Mouse embryonic head as a site for hematopoietic stem cell development". In: *Cell Stem Cell* 11.5 (2012).

[385] Haruko Nakano et al. "Haemogenic endocardium contributes to transient definitive haematopoiesis". In: *Nature Communications* 4 (2013).

[386] Sabrina Gordon-Keylock et al. "Mouse extraembryonic arterial vessels harbor precursors capable of maturing into definitive HSCs". In: *Blood* 122.14 (Oct. 2013), pp. 2338–2345. URL: https://ashpublications.org/blood/article/122/14/2338/32195/Mouse-extraembryonic-arterial-vessels-harbor.

[387] Gemma Swiers et al. *A short history of hemogenic endothelium*. 2013. URL: http://dx.doi.org/10.1016/j.bcmd.2013.09.005.

[388] H. E. Jordan. "Aortic Cell Clusters in Vertebrate Embryos". In: *Proceedings of the National Academy of Sciences* 3.3 (1917).

[389] V. E. Emmel. "The cell clusters in the dorsal aorta of mammalian embryos". In: *American Journal of Anatomy* 19.3 (1916).

[390] Amanda D. Yzaguirre et al. "Runx1 is sufficient for blood cell formation from non-hemogenic endothelial cells in vivo only during early embryogenesis". In: *Development (Cambridge)* 145.2 (2018).

[391] Vivien Marx. "Method of the Year: spatially resolved transcriptomics". In: *Nature Methods* 18.1 (2021).

[392] Christoph Sommer et al. "Ilastik: Interactive learning and segmentation toolkit". In: *Proceedings - International Symposium on Biomedical Imaging*. 2011.

[393] Stuart Berg et al. "ilastik: interactive machine learning for (bio)image analysis". In: *Nature Methods* 16.12 (2019).

[394] Carsen Stringer, Michalis Michaelos, and Marius Pachitariu. "Cellpose: a generalist algorithm for cellular segmentation". In: (). URL: https://doi.org/10.1101/2020.02.02.931238.