UNIVERSITY OF CALIFORNIA, SAN DIEGO

Defining Mesodermal Cell Types, Branching States, and Origins by Single Cell Genomics

A dissertation submitted in satisfaction of the
requirements for the degree Doctor of Philosophy

in

Biomedical Sciences

by

Jonathan David Grinstein

Committee in charge:

Professor Neil C. Chi, Chair
Professor Sylvia M. Evans
Professor Lawrence S. Goldstein
Professor Maike Sander
Professor Gene W. Yeo

2017

The Dissertation of Jonathan David Grinstein is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____
Chair

University of California, San Diego

2017

DEDICATION

In recognition of their unwavering support, this doctoral dissertation is dedicated to my parents Mami and Papi, my brother Gabi, my wife Jess, my family, and friends.

EPIGRAPH

For all sad words of tongue and pen, the saddest are these, ‘It might have been’.

John Greenleaf Whittier

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

A-P – Anterior-Posterior

bHLH – basic-Helix-Loop-Helix

CBCs – Crypt Base Columnar Cells

CCPM – Cranio-Cardial-Pharyngeal Mesoderm

cDNA – complementary DNA

CMs – Cardiomyocytes

DLTs – Developmental Lineage Trajectories

DNA – Deoxyribonucleic Acid

EBS – Early Bud Stage

ECs – Endothelial Cells

EMPs – Erythroid Myeloid Progenitors

ESS – Early Streak Stage

ExEM – Extra-Embryonic Mesoderm

FBs – Fibroblasts

GAM - Generalized Additive Model

GO – Gene Ontology

GRNs – Gene Regulatory Networks

HSCs – Hematopoietic Stem Cell

LERC – Lgr5:EGFP:ROSA:creERT2;R26R:tdT

LacZ – beta-galactosidase

Lgr5 – Leucine-rich repeat-containing G-coupled receptor 5

LPM – Lateral Plate Mesoderm

LSS – Late Streak Stage

ltHSCs – Long term HSCs

Mesp1 – Mesoderm Posterior 1

MSS – Mid Streak Stage

MST – Minimal Spanning Tree

NGS – Next Generation Sequencing

OFT – Outflow Tract

PAGODA – Pathway and Gene set Over-Dispersion Analysis

PCA – Principal Component Analysis

PCR – Polymerase Chain Reaction

qPCR – Quantitative PCR

PS – Primitive Streak

RNA – Ribonucleic Acid

Runx1 – Runt-related Transcription Factor 1

SCDE – Single Cell Differential Expression

ScRNAseq – Single Cell RNA-seq

shHSCs – Short term HSCs

SMCs – Smooth Muscle Cells

TFs – Transcription Factors

tSNE – t-distributed Stochastic Neighbor Embedding

TSS – Three Somite Stage

YS – Yolk Sac

YS-BI – Yolk Sac Blood Islands

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank the entire Neil Chi Laboratory. In addition, I would like to thank Dekker Deacon, my laboratory partner in crime who kept me hanging on through laughter and tears. I would like to thank Neil Tedeschi, thank you for all your humble support and never change who you are. I would like to thank Elie Farah for his unabashed support for my work, and I know you can achieve anything you put your mind towards. I would like to thank Dr. Josh Bloomekatz, for his collaborative, supportive, insightful, and kind role throughout my graduate career.

I would like to thank Dr. Paola Cattaneo of Professor Sylvia Evans' laboratory for her essential role in this project. Paola and I tried to do things that people said were impossible at the time we attempted them, and it wouldn't have been possible without her resolve, attitude, and wits. I would like to thank Dr. Nuno Camboa for his collaboration and expertise in histology and microscopy. I would like to thank Yan Song and Leen Jamal-Schafer of Professor Gene Yeo's laboratory for getting this project off the ground. I would like to thank Kristen Jepsen and Tala Heidari from the UCSD Institute for Genomic Medicine for troubleshooting and generating the experimental design for processing and sequencing of single cell RNA libraries, whom without this data would never even be available for analysis.

I would like to thank my entire Committee in charge. I would like to thank Dr. Sylvia Evans for her unwavering support and invaluable time spent accessing her mental encyclopedia of development.

I would like to thank Professor Neil Chi for his support as chair of my committee. Through many long nights, his guidance, support, and passion has proved to be

invaluable. Neil has challenged me to be the best scientist I can be, and to focus on the question at hand instead of technical and biological fads, which come and go. Neil has taught me to be true to myself and to believe in my capabilities, scientific and personal. You have given me the support and encouragement I needed to follow and find my interests. I will carry these lessons, and colloquialisms, for the rest of my life. I am forever thankful.

Chapters 2 and 3 are currently being prepared for submission for publication of the material. Jonathan D. Grinstein, Paola Cattaneo, Josh Bloomekatz, Neil Tedeschi, Elie Farah, Yan Song, Leen Jamal-Schafer, Nuno Camboa, Gene Yeo, Sylvia M. Evans, Neil C. Chi. The dissertation/thesis author, Jonathan D. Grinstein, was the primary investigator and author of this material.

Chapter 1 is an original document written by Jonathan D. Grinstein with oversight from Neil Chi and Sylvia Evans. Jonathan D. Grinstein was supported by pre-doctoral fellowships from the University of California, San Diego (UCSD) Pre-Doctoral fellowship and the California Institute of Regenerative (CIRM) Medicine Pre-Doctoral fellowship. Neil Chi is an Associate Professor of Medicine at UCSD and is supported by grants from the National Institutes of Health (NIH) and the American Heart Association.

Chapter 2 is an original document written by Jonathan D. Grinstein with oversight from Neil Chi and Sylvia Evans. The work presented in this chapter was led by Jonathan D. Grinstein with contributions from the laboratories of Neil Chi (Neil Tedeschi, Elie Farah, and Josh Bloomekatz), Sylvia Evans (Paola Cattaneo and Nuno Camboa), and Gene Yeo (Yan Song and Leen Jamal-Schafer). Jonathan D. Grinstein was supported by a pre-doctoral fellowship from the California Institute of Regenerative Medicine (CIRM).

Neil Chi is an Associate Professor of Medicine at UCSD and is supported by grants from the National Institutes of Health (NIH) and the American Heart Association.

Chapter 3 is an original document written by Jonathan D. Grinstein with oversight from Neil Chi and Sylvia Evans. The work presented in this chapter was led by Jonathan D. Grinstein in collaboration with Josh Bloomekatz, Paola Cattaneo, and Nuno Camboa. Jonathan D. Grinstein was supported by a pre-doctoral fellowship from the California Institute of Regenerative Medicine (CIRM). Neil Chi is an Associate Professor of Medicine at UCSD and is supported by grants from the National Institutes of Health (NIH) and the American Heart Association.

VITA

2005-2009	New York University Bachelor of Science Neural Science
2007	New York University, Center for Neural Science Research Assistant Suzuki Laboratory
2008	New York University, Center for Neural Science Research Assistant Klann Laboratory
2008	Salk Institute Research Technician Gage Laboratory of Genetics
2008-2010	New York University Langone Medical Center Neuroscience Institute Research Scientist Dasen Laboratory
2011-2017	University of California, San Diego, Biomedical Sciences Doctor of Philosophy Chi Laboratory

PUBLICATIONS

Grinstein JD, … Chi NC. TBD. *In Preparation*.

Panopoulos AD, D'Antonio M, Arias AD, Benaglio P, DeBoever C, Williams R, Garcia M, Nelson B, Harismendy O, **Grinstein JD**, Drees F, Okubo J, Diffenderfer KE, Hishida Y, Modesto V, Dargitz CT, Feiring R, Zhao C, Jepsen K, McGarry T, Matsui H, Reyna J, Aguirre A, Rao F, O'Connor DT, Chi NC, Goldstein LS, Izpisua Belmonte JC, Berggren WT, Adler E, D'Antonio-Chronowska A, Smith EN, Frazer KA. CARDiPS: A well-characterized resource of iPSC lines from 222 individuals for examining how genetic variation affects cellular traits and disease. *In Review, Cell*.

Han P, Bloomekatz J, Ren J, Zhang R, **Grinstein JD**, Zhao L, Burns CG, Burns CE, Anderson RM, Chi NC. *Nature. 2016 doi: 10.1038/nature18310.*

Lacombe J, Hanley O, Jung H, Philippidou P, Surmeli G, **Grinstein J**, Dasen JS. Dev. Cell. 2013. Genetic and functional modularity of Hox activities in the specification of limb-innervating motor neurons. *PLoS Genet. 2013 doi: 10.1371/journal.pgen.1003184.*

Yang L, Lin C, Liu W, Zhang J, Ohgi KA, **Grinstein JD**, Dorrestein PC, Rosenfeld MG. ncRNA- and Pc2 methylation-dependent gene relocation between nuclear structures mediates gene activation programs. *Cell. 2011 doi: 10.1016/j.cell.2011.08.054.*

Jung H, Lacombe J, Mazzoni EO, Liem KF Jr, **Grinstein J**, Mahony S, Mukhopadhyay D, Gifford DK, Young RA, Anderson KV, Wichterle H, Dasen JS. Global control of motor neuron topography mediated by the repressive actions of a single hox gene. *Neuron. 2010 doi: 10.1016/j.neuron.2010.08.008.*

FIELDS OF STUDY

Major Field: Biomedical Sciences

Studies in Genetics
Professors Neil Chi and Sylvia Evans

Studies in Cellular and Molecular Medicine
Professors Lawrence Goldstein, Maike Sander, and Gene Yeo

ABSTRACT OF THE DISSERTATION

Defining Mesodermal Cell Types, Branching States, and Origins by Single Cell Genomics

By

Jonathan David Grinstein

Doctor of Philosophy in Biomedical Sciences

University of California, San Diego, 2017

Professor Neil C. Chi, Chair

Every cell in the body arises from a single fertilized cell that, through many cell divisions and transitions from one type to another, travels a landscape of possible fate decisions during development. Though adult cell types have been well cataloged molecularly and functionally, the states a cell can pass through and the genes that regulate these choices remain unclear. Recent advances in quantitative cellular measurements for monitoring global gene regulation simultaneously in hundreds to thousands of single cells from an experiment is allowing for the discovery of cell types and cell states as well as the trajectories linking them. This dissertation is comprised of

investigation of, and applies towards, the developmental states and genetic regulation of mouse mesoderm.

The goal of the first study was to catalog cell types and states during mouse mesodermal developmental by gene expression at the single cell level (Chapter 2). We determined developmental trajectories for six mesodermal cell types, providing putative novel cellular intermediates with previously undocumented gene expression profiles. The goal of the second study was to functionally investigate a cellular intermediate identified in the first study that had not been previously described.

These studies describe new findings concerning the cellular and molecular hierarchy of mesodermal development that will enable further study of genetic regulation of developing cell types. Mapping developmental landscapes may promote the discovery of genes and pathways that govern cell fate decisions and transitions essential for advancements in understanding the etiology of disease and the development of medical therapies.

Chapter 1: Introduction

1.1 Abstract

This chapter serves as an introduction to the allocation and differentiation of mesodermal progenitors in the murine embryo. It begins with an introduction to the process of gastrulation, the outcome being the formation of the definitive germ layers and the organization of a general body plan body, and goes onto examine, specifically, mesoderm. Next, we explore next generation sequencing (NGS) and computational methods used to examine single cell genomics associated with developmental lineage trajectories (DLTs). Finally, this chapter concludes with remaining questions in the genetic regulation of progenitor specification and cell type potential in the mesodermal hierarchy.

1.2 Dissertation Introduction

All organs of the body originate from the three germ layers of the embryo: endoderm (inside layer), mesoderm (middle layer), and ectoderm (outside layer). Of these, the mesoderm will give rise to the following organs: axial skeleton, skeletal muscles, connective tissue of the skin, connective tissue of the body wall and limbs, urogenital structures, the cardiovascular system, and hematopoietic system (red blood cells, platelets, and immune cells). In addition, mesoderm also contributes outside the embryo to the fetal membranes. Extensive leaps have been made in the understanding of the mesodermal lineage tree, however many mesodermal intermediates and their potential lineage decisions remain unclear.

1.2.1 *Embryo Structure and Morphogenesis of Germ Layers during Gastrulation*

Recognized development biologist Lewis Wolpert was once quoted saying, "It is not birth, marriage or death, but gastrulation which is truly the most important time in your life" (Slack, 1999). This is because the formation of the germ layers and the organization of its derived organs occurs through a process known as gastrulation (Tam PP., 1997). Prior to gastrulation, the developing mouse embryo consists of three tissue lineages (trophectoderm, epiblast and primitive endoderm) known as the blastocyst. The blastocyst is built as a vesicular structure with an epithelialized trophectoderm enclosing a cavity (the blastocoel) and, attached to the wall on one side of the blastocoel, a cluster of cells that constitutes the epiblast and primitive endoderm. The epiblast, which gives rise to the entire embryo and some components of the fetal extraembryonic membranes, is a continuous epithelial layer of cells that is lined by primitive endoderm at the luminal surface.

During gastrulation, epiblast cells ingress through morphogenetic activity at the primitive streak (PS) to give rise to mesoderm and endoderm, with the remaining epiblast cell population acquires an ectodermal fate (Ang SL, 1993; Yamaguchi, 1994; Tam P. , 1989). The formation of the mesoderm and endoderm is accomplished by an epithelial–mesenchyme transition of epiblast cells at the PS, the organization of the ingressed mesoderm progenitors into a mesenchymal layer, and the incorporation of the endoderm progenitors into the pre-existing layer of primitive endoderm. Mesodermal cells recruited from the epiblast are organized into mesenchymal layers of mesodermal cells that constitute the embryonic and extraembryonic mesoderm.

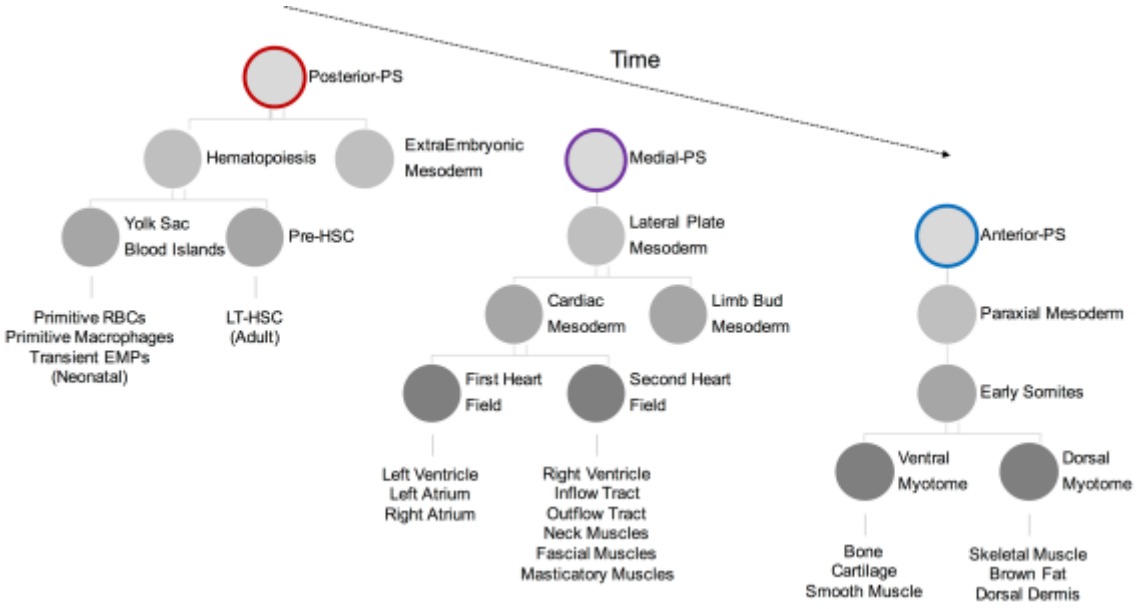


Figure 1.1: Developmental Hierarchy of Mesodermal Lineages.
Spatiotemporal emergence of mesodermal progenitors along the primitive streak and known, respective cellular hierarchies. Gastrulation begins at the intersection of the embryo proper and the ExEM at the Posterior-PS at ESS. As the primitive streak extends, cells begin to ingress at the Medial-PS around MSS, and continues distally to the Anterior-PS by LSS. Fate mapping of cells at different spatiotemporal locations in the epiblast and PS during gastrulation has identified how these cells contribute to different mesodermal cell types .

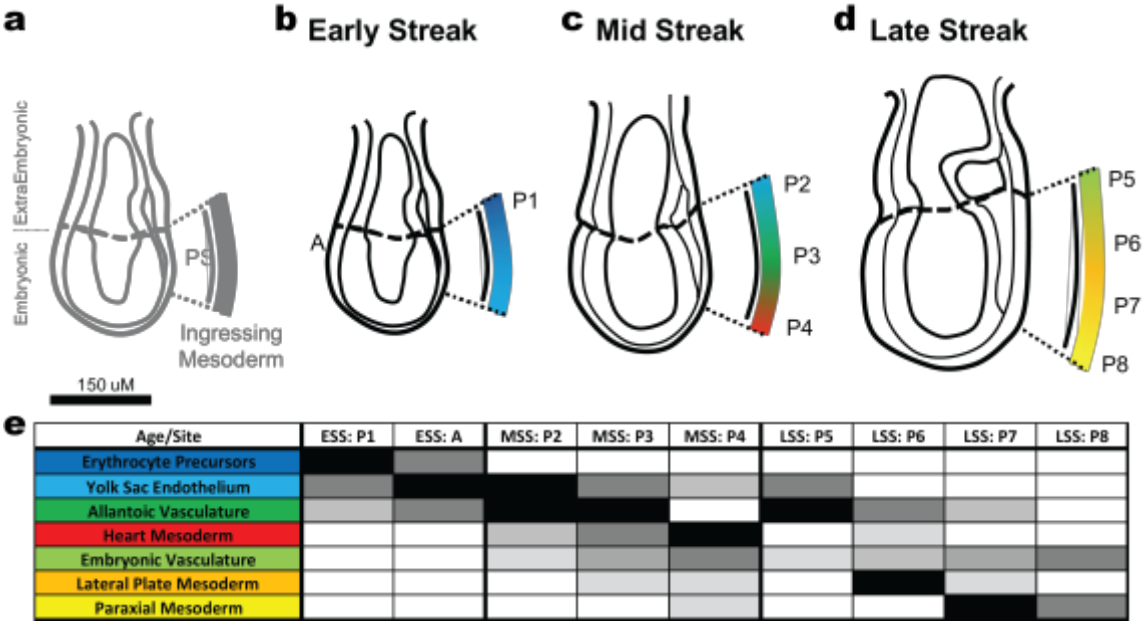
1.2.2 *Mesoderm Formation during Mouse Gastrulation*

In the 1990s, mouse fate mapping studies tracking the fate of epiblast cells provided the first real insight into the spatiotemporal organization allocating mesodermal cells during murine gastrulation (Lawson KA, 1991; Parameswaran M, 1995). These studies revealed that when gastrulation begins (Early Streak, ESS), cells that ingress contribute to mesodermal tissues of the extra-embryonic mesoderm (ExEM), YS, amnion, and hematopoiesis. As the PS extends, cells of embryonic mesoderm begin to ingress (Mid Streak, MSS). Of these, the first to ingress are destined for cardiac mesoderm, cranial mesoderm, and lateral plate mesoderm (LPM) that contributes to the anterior (upper) body (Kinder SJ L. D., 2001). As gastrulation begins to cease (Late Streak, LSS) and the node reaches its distal apex, cells ingress that contribute to LPM and paraxial mesoderm of the trunk (Figure 1.1) (Tam PP G. D., 2000).

Extensive fate mapping studies of cells in different locations of the PS at different stages revealed the further resolved the spatiotemporal organization of mesodermal allocation. The first cells ingress at the forming PS in the posterior intersection of the epiblast and visceral endoderm and cells that contribute to YS-endothelium followed by erythrocyte precursors (Figure 1.2 b,e). At MSS, the primitive streak extends and the remaining precursors to YS-endothelium and precursors to allantoic vasculature ingress at the most posterior portion of the PS (P2) (Figure 1.2 c,e). Simultaneously, cells at the medial portion of the PS (P3) also contribute to allantoic vasculature and the first precursors to cardiac mesoderm. The most distal portion of the MSS PS (P4) contains the largest the bulk of cells contributing to cardiac mesoderm. At LSS, the remaining YS and allantoic vasculature ingress in the rostral most part of the PS (P5), and the remaining

precursors to cardiac mesoderm and embryonic vasculature ingress in the adjacent PS (P6) (Figure 1.2 c,e). Cells contributing to cephalic LPM and paraxial mesoderm ingress at LS. Precursors to cephalic LPM following the remaining precursors to cardiac mesoderm in P6, and paraxial mesoderm further down the PS (P7). The distal most part of the streak contains cells contributing to embryonic vasculature and paraxial mesoderm.

Not surprisingly, fate mapping within different kinds of mesoderm discussed here have been resolved to individual cell types. For instance, different populations of ExEM-vasculature have been showed to have distinct spatiotemporal recruitment to the PS, where the majority of cells contributing to the vitelline vessels ingress at ESS in contrast to allantoic vasculature at MSS and LSS. In addition, retrospective lineage tracing suggests that the spatiotemporal allocation of cardiovascular progenitor cell types such that different portions of the heart arise from different progenitor pools. These findings highlight the complexity of gastrulation and mesodermal allocation through highly regulated morphological movements.



e

Age/Site	ESS: P1	ESS: A	MSS: P2	MSS: P3	MSS: P4	LSS: P5	LSS: P6	LSS: P7	LSS: P8
Erythrocyte Precursors									
Yolk Sac Endothelium									
Allantoic Vasculature									
Heart Mesoderm									
Embryonic Vasculature									
Lateral Plate Mesoderm									
Paraxial Mesoderm									

Figure. 1.2: Spatiotemporal Contribution of Murine Primitive Streak Cells to Mesodermal Lineages.

a, The gastrula stage embryo is split into the embryonic and extra-embryonic by the amnion (the dotted line across the middle of the embryo). The posterior region of the embryo (right) contains the primitive streak (thin grey bar). The embryonic compartment is further sub-divided into proximal, middle, and distal regions according to their position relative to the amnion for describing the regionalization of mesodermal progenitors in the epiblast. The mesoderm layer (thick grey bar), which lies on top of the epiblast (dotted outline), is pulled away from the embryo to illustrate cell fate in b-d. **b-d.** A developmental series of fate maps showing the localization of mesodermal tissue progenitors in the epiblast, the primitive streak (thin black bar) and the mesoderm (thick colored bar). The different types of mesodermal progenitors are color-coded (see the color key). The tissue composition of the mesoderm reflects the types of progenitors that have been recruited from the epiblast through the primitive streak in the immediately preceding developmental stage, but not those that are currently ingressing into the primitive streak. Data is adapted from Kinder SJ, 1999.

1.2.3 *Defining Cell Types and States with Single Cell Genomics*

We have a vast understanding of the movements of the germ layers, defined differentiated cell types –it is commonly stated that there are 210 different cell types in the adult human body-, and the location of precursors to these cell types in the epiblast. However, much remains unknown about the cellular and molecular identities of precursors and intermediates with differing fate potentials to differentiated cell types, let alone the molecular mechanisms and principles regulating these developmental states and decisions. Furthermore, the current models for developmental hierarchies relied heavily on fate mapping using dyes to label cells in a spatial location or few genes to label cells expressing such genes, a fundamental limitation in the complexity of characterizing cell types and states and the many genes that define them.

In addition, most quantitative measures of gene expression, chromatin state, and protein profiles have, until recently, relied on large cell numbers and are processed in bulk, raising another fundamental limitation to resolve cellular heterogeneity in cellular states (Trapnell, 2015). This limitation, known as the Yule-Simpson effect, states the following paradox:

> The fictitious association caused by mixing records finds its counterpart in the spurious correlation to which the same process may give rise in the case of continuous variables, a case to which attention was drawn and which was fully discussed by Professor Pearson in a recent memoir. If two separate records, for each of which the correlation is zero, be pooled together, a spurious correlation will necessarily be created unless the mean of one of the variables, at least, be the same in the two cases. (Yule, 1903)

As introduced in the previous section, quantitative cell measurements have taken a huge leap forward in the past few years in the field on single cell genomics, allowing for the

analysis of total gene expression in hundreds to thousands of individual cells. These tools provide unprecedented insight into mixed populations through unbiased statistical models without the need for *a posteriori* markers. Application of single cell genomics to developmental and adult stem cell populations may provide insight into previously undocumented cell types and states and the genes regulating cell fate decisions.

1.3 Remaining Questions

Although Waddington depicted the developmental landscape of differentiating stem cells splitting off to different lineages at branch points nearly 80 years ago, several fundamental questions remain about developmental and stem cell biology. In order to recreate this branching landscape, it is essential to identify all cell types and intermediates in the body of the organism of interest. Only then, can one begin to map the branching lineage choices for cellular intermediates and the genes molding and signaling cues effecting their fate decisions.

1.4 Dissertation Organization

The purpose of this dissertation is to apply single cell genomics and robust statistical models to define the cell types and intermediate states within the mouse mesoderm and the genes regulating each cell fate decision. Specifically, we seek to identify mesodermal cell types by known markers and *de novo* gene sets in order to model cell type lineage trajectories temporally. These analyses then provide testable predictions of different putative mesodermal intermediates for experimental validation.

Chapter 2 details work led by Jonathan D. Grinstein on reconstructing the murine mesodermal hierarchy via single cell expression trajectories. This work was done in collaboration with Paola Cattaneo, Yan Song, Leen Jamal-Schafer, Neil Tedeschi, and Elie Farah.

Chapter 3 details work led by Jonathan D. Grinstein on defining the molecular signature and trajectory of hemogenic cellular intermediates. This work was complemented through collaboration with Paola Cattaneo, Nuno Camboa, and Josh Bloomekatz on the role of Lgr5 in murine developmental hematopoiesis.

1.5 Acknowledgements

This chapter is an original document written by Jonathan D. Grinstein with oversight from Neil Chi and Sylvia Evans. Jonathan D. Grinstein was supported by a pre-doctoral fellowship from the California Institute of Regenerative Medicine (CIRM). Neil Chi is an Associate Professor of Medicine at UCSD and is supported by grants from the National Institutes of Health (NIH) and the American Heart Association.

Chapter 2: Deconstructing the Mesodermal Hierarchy via Single Cell Expression Trajectories

2.1 Abstract

This chapter is adapted from work led by Jonathan D. Grinstein under the supervision of Neil Chi. This chapter combines genetic, NGS, and computational experimental strategies described in previous chapters to complete an in-depth examination of the mesodermal hierarchy. Here we analyze the transcriptomes of single, purified cells from gastrulating mesoderm at during MSS, LSS, and TSS using scRNAseq. We computationally applied robust statistical analyses to spatiotemporally classify mesodermal cell types to define a roadmap of mesodermal development and propose a hierarchy of cell developmental lineage trajectories (DLTs) with branching intermediate cell fate potentials in mesodermal development. These data suggest that mesoderm develops through three main trajectories with differing cell fate potentials: (1) blood and endothelium, (2) CMs, cranio-cardial-pharyngeal mesoderm (CCPM), and muscle, and (3) ExEM. Together, these studies provide a model for the mesodermal developmental hierarchy, and a resource for identifying novel mesodermal cellular intermediates and genetic regulators of these cell fate decisions.

2.2 Introduction

Mesoderm Posterior 1 (Mesp1) is one of the earliest genes expressed during gastrulation by nascent mesoderm at the PS (Saga Y M.-T. S., 1999). Over twenty years ago, Mesp1 was cloned from the base of the allantois at LSS (Saga Y H. N., 1996). The Mesp1 gene encodes a basic helix-loop-helix (bHLH) domain –it originally was named bHLHc5- that is evolutionarily related to master regulators of muscle, myogenic differentiation1 (Myod1/bHLHc1) and myogenic factor 4 (Myog/ bHLHc3). Mesp1 has

been established as a conserved master regulator of cardiovascular development from the invertebrate chordate *Ciona intestinalis* to humans (Davidson B, 2003; Diogo R, 2015). Studies using genetic models in mouse and zebrafish and over-expression experiments in hESCs has shown that Mesp1 expressing cells contribute to all layers of the heart (myocardium, endocardium, and epicardium) and the cardiovascular cell types that compose them: cardiomyocytes (CMs), endothelial cells (ECs), smooth muscle cells (SMCs), and fibroblasts (FBs).

However, for how extensively Mesp1 has been studied in the cardiovascular system, many questions remain about how it mechanistically functions to specify cell fates. In addition, Mesp1 is not only expressed in cells during gastrulation that give rise to the cardiovascular system, but also contribute to a variety of mesodermal lineages and cell types including muscles of the head, neck, and limbs, liver, and extra-embryonic tissues such as the umbilical cord and YS. The purpose of this chapter is to investigate the entire Mesp1 developmental hierarchy by characterizing the identity of cell types, states, and origins of Mesp1 expressing cells by single cell genomics.

2.2.1 *Mesp1 in Murine Cardiovascular Development*

Mesp1 has been extensively studied for its role in regulating the development of the cardiovascular system by deriving major cardiovascular cell types (Kitajima S, 2000; Bondue A, 2008; Lescroart F, 2014; Devine WP, 2014). During gastrulation, Mesp1 expressing cells migrate through the PS and begin to lose expression of Mesp1 as they migrate anteriorly. The first lineage tracing results performed by genetically placing a Cre recombinase into the Mesp1 locus showed that cells that express Mesp1 during

gastrulation, when genetically labelled, eventually make up the entirety of the heart and the amnion. Genetical disruption of both Mesp1 alleles in mice causes death during development at E9.5 due to failure of the bilateral myocardial cells to coalesce into a single central heart tube resulting in the presence of two independent hearts, a genetic disease named *cardia bifida*.

Mesp1 is physically located in the genome near its paralog Mesp2, which has a similar but delayed expression pattern to Mesp1 and, together with Mesp1 later in development, regulates somitogenesis. Although individual genetic disruptions of either Mesp paralog results in embryonic lethality, disruption of both leads to the complete loss of the heart and somite-derived and non-derived (cephalic mesoderm) muscle. However, in experiments using chimeras composed of wild-type and Mesp1/Mesp2 double mutant cells, the mutant cells contribute to somite-derived muscle but not to cardiac and cephalic mesoderm derived structures including the heart and some head and neck muscles. Perhaps it is these phenotypes highlighting the cell-autonomous function of Mesp1/Mesp2 in the cardiovascular system that has led to investigation into its function primarily in this organ and its cell types.

Although retrospective lineage tracing had shown that distinct regions of the heart arise from separate progenitor pools, it was only until recent single cell clonal analysis of Mesp1-expressing cells that it was determined when cells became destined to specific cardiovascular fates (Lescroart F, 2014; Devine WP, 2014). These studies suggest that cardiovascular mesoderm -and perhaps all mesoderm- is specified into distinct potentials for cardiovascular progenitors and region before and during gastrulation. However, whether these cardiovascular progenitors are pre-patterened or molded by the

morphogenetic events and signaling cues during and immediately subsequent to gastrulation. Single cell qPCR of Mesp1 expressing cells revealed molecular heterogeneity of cardiovascular progenitors during gastrulation indicative of spatiotemporally distinct populations, yet the identity of these progenitors and the genes that regulate their cell fate remain elusive.

2.2.2 *Mesp1 in Murine Hematopoietic Development*

Saga's initial work on Mesp1 did observe that, in addition to the primitive streak, Mesp1 was expressed in the ExEM, and that genetically labelled Mesp1 expressing cells in the ExEM could be later be observed in YS-BI, which is generally known as the main site of embryonic hematopoiesis. Although the ExEM has been shown to be the site harboring cells that contribute to both embryonic and adult hematopoiesis and much of the adult hematopoietic cellular and molecular hierarchy has been defined, much of what happens from gastrulation to the onset of adult hematopoiesis with the seeding of liver and, ultimately, bone marrow with hematopoietic stem cells (HSCs) remains unclear.

Recent work has shown that, in the absence of lineage-inducing factors, over-expression of Mesp1 early in its window of expression in hESC differentiations resulted in the initiation of hematopoiesis *in vitro* and, in contrast, to cardiogenesis when over-expressed late in its expression window (Chan SS, 2013; Kouskoff V, 2005). This context-dependent function of Mesp1 is not entirely surprising given our understanding of the spatiotemporal sequential ingression of mesodermal cell types through the streak with the ingression of blood preceding cardiovascular cell types. Furthermore, analysis of genetically labelled Mesp1+ cells can be observed in the liver at the onset of adult

hematopoiesis, and, presumably, these cells then contribute to one third of adult bone marrow and generates HSCs capable of repopulating irradiated hosts –the gold standard for HSC transplant function. Therefore, Mesp1-expressing cells contribute to essentially the entire hematopoietic hierarchy, from the first circulating primitive erythrocytes from YS-BI to gully functional HSCs.

2.2.3 *Mesp1 in Murine Muscle, Mesenchymal, and Mesothelial Development*

Mesp1 acts on a vast number of mesodermal fate decisions, including the cardiovascular system, primitive and definitive blood, and skeletal muscle, through both cell-autonomous and non-cell autonomous mechanisms (Chan SS, 2013; Chiapparo G, 2016; David R, 2008). Skeletal muscle ingresses spatially posterior and temporally after hematopoietic and cardiac mesoderm, and, consistently, over-expression of Mesp1 at an even later window than that which induced hematopoitic and cardiac cell types resulted in the induction of skeletal muscle progenitors and, ultimately, skeletal muscle and satellite cells (Yoshida T, 2008; Chan SS, 2013). In addition, over-expression of Mesp1 in a window between that which induced cardiac and skeletal muscle generated a bi-potent cardiac and skeletal muscle progenitor, which was validated *in vivo* by clonal lineage tracing of individual Mesp1-labelled cells in a population commonly refered to as cardio-pharyngeal mesoderm (Chan SS., 2016). Lineage tracing of Mesp1 cells is observed in satellite cell populations in craniofacial skeletal muscles such as the masseter and is absent from the trunk.

In addition to craniofacial skeletal muscle populations, Mesp1 labelled cells contribute to muscles, endothelium, and bone of the head and neck. These craniofacial

mesenchymal cells originate from cranial-mesoderm or neural crest. It has been shown that Mesp1 is critical to craniofacial morphogenesis through both cell-autonomous and cell non-autonomous mechanisms.

Finally, Mesp1 has been shown to label mesothelium, a single cell layer that functions to cover the surface of body cavities (Lua I, 2015). Mesothelium functions to expedite movement of fluids between organs and as a source of myofibroblasts, which function in fibrosis in response to injury (Li Y, 2013). Mesp1 labelled cells contribute to mesothelial cells (MCs) of the liver but not the abdomen (peritoneal mesothelium), where they differentiate into hepatic stem cells as well as portal fibroblasts that convert into myofibroblasts during fibrosis (Lua I., 2014).

2.2.4 *Chapter 2 Research Strategy*

Although the role of Mesp1 atop the hierarchy in specifying mesodermal cell fates as early as the epiblast stage has been extensively studied, the identity of the cell states and how genes regulate fate decisions towards different cell types remain unclear. Here we identify cell types and states that arise from Mesp1 expressing cells during and following gastrulation. We integrate robust statistical analysis to single cell RNA expression data to define a developmental trajectory for these cell types.

2.3 Results

2.3.1 *Experimental Strategy for Investigating Mesodermal DLTs by ScRNAseq*

We used single-cell transcriptomics to investigate mesodermal sub-populations in Mesp1-labelled single cells from individual mouse embryos. We purified *Mesp1*-labelled

cells at various points across a developmental window from the middle of *Mesp1* expression in the PS (E7.0) to when Mesp1 is no longer expressed and labelling new cells (E8.0). Embryos were dissected away from the decidua, parietal YS, Reichert's membrane, and the ectoplacental cone, and then staged according to anatomical features (Methods) as E7.0/MSS, E7.5/LSS, and E8.0/TSS (Figure 2.1 a). In addition, most the visceral YS was manually dissected away from the TSS embryo due its over-representation at this stage. Single-cell transcriptomes were obtained from individual E7.0, E7.5, and E8.0 Mesp1:Cre/+;R26R:tdTomato/+ embryos separately dissociated and FACS-purified single cells positive for Mesp1-labelling by tdTomato (tdT) and viability (DAPI- events) (Figure 2.1 b). We purified 1260 tdT+/DAPI- cells out of a total of 12571 total events from E7.0/MSS, 1208 tdT+/DAPI- cells out of 15957 total events from E7.5/LSS, and 5444 tdT+/DAPI- cells out of32000 total events from E8.0/TSS.

ScRNAseq was carried out on FACS-purified cells using the Fluidigm C1 microfluidics cell capture platform followed by library preparation and Illumina sequencing. Following visual inspection of individual capture sites for single tdT+/DAPI- cells, we captured 65 tdT+/DAPI- cells from MSS, 67 tdT+/DAPI- cells from LSS, and 79 tdT+/DAPI- cells from TSS. Using in house metrics (Methods), we observed that 64 single cells from E7.0/MSS, 63 cells from E7.5/LSS, and 70 cells from E8.0/TSS generated cDNA sequencing libraries that were of high quality, yielding 197 Mesp1-labelled single cell transcriptomes for sequencing and further analysis (Figure 2.1 a ii,v,viii). We sequenced our cells to an average depth of $2x10^6$ reads.

To fit error individual error models for our scRNAseq measurements from all embryos, we used SCDE, a computational package that implements a set of statistical

methods for analyzing scRNAseq by building a cell-specific error models with estimates of drop-out and amplification biases on gene expression magnitude (Kharchenko PV, 2014). We next used PAGODA framework from SCDE to perform weighted principal component analysis (PCA) of individual pre-defined gene sets, such as Gene Ontology (GO) categories, as well as 'de novo' gene sets to identify significantly over-dispersed gene sets to generate aspects of transcriptional heterogeneity that characterize populations of single cells (Fan J, 2016).

Significant aspects were used to cluster cells into subpopulations by hierarchical clustering. Hierarchical clustering yielded fourteen robust clusters (Figure 2.2 a, upper bar), with five clusters in one branch and eleven in the other. Most clusters received contributions primarily from one embryonic stage (Figure 2.2 a, lower bar) with the five clusters branch containing cells mostly from MSS and the eleven clusters branch from LSS and TSS. We analyzed aspects of transcriptional heterogeneity and the gene sets they contain generated by PAGODA (Figure 2.2 b).

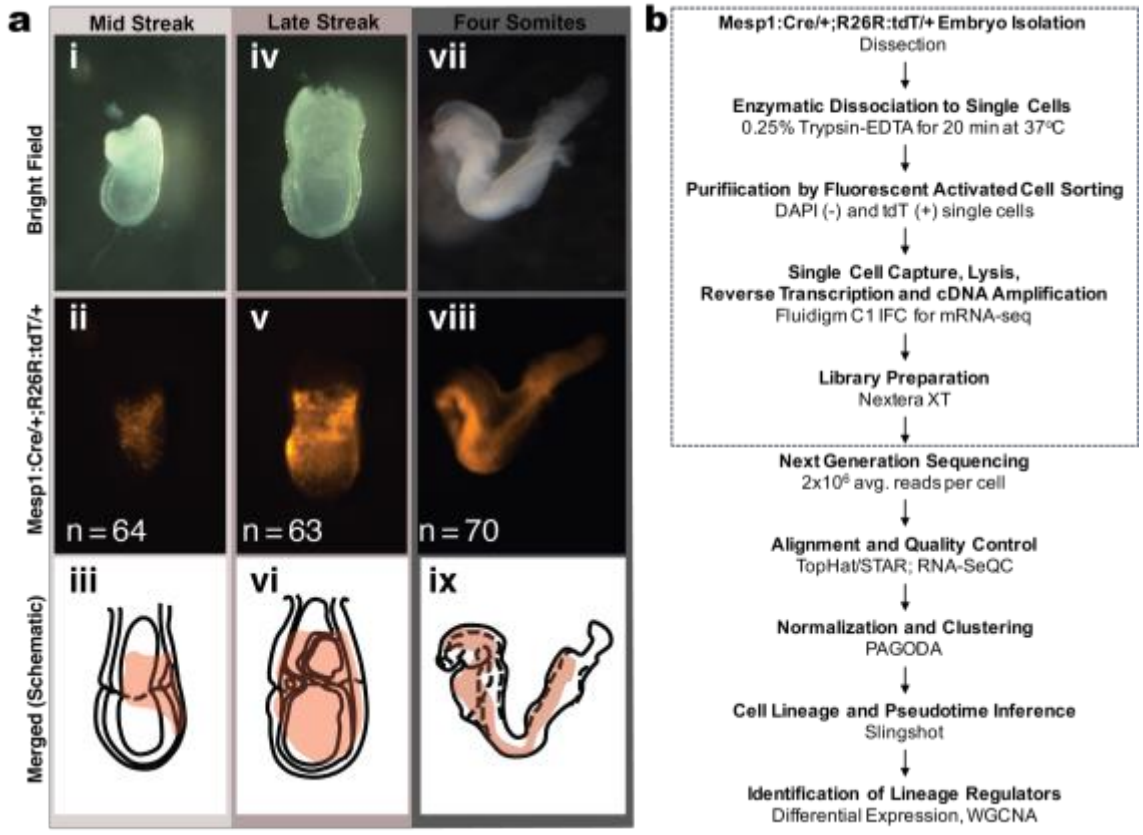


Figure 2.1: Experimental Strategy for Single Cell Lineage Trajectory Analysis of Mesodermal Progenitors using scRNAseq.

a, Whole-mount images of Mesp1:Cre/+;R26R:tdT/+ embryos used for scRNAseq experiments. Top, bright field images; middle, red-channel images; bottom, representative schematic of merge. Colors indicate stage of embryo. Light grey, Mid Streak (i-iii); mid grey, Late Streak (iv-vi); and dark grey, Three-somites (vii-ix). *n*, Number of sequenced cells with libraries of high quality. **b,** Experimental design for each embryo prepared and processed. Boxed area indicates independent steps performed on embryos.

2.3.2 *Clustering and Assignment of Mesodermal Cells to Branching Cell Lineages*

First, we analyzed which pre-defined pathways and de-novo gene sets were the most over-dispersed across all cells. The top most over- dispersed gene set was a de-novo gene set, GeneCluster-130, comprised of known cardiac contractile proteins (e.g. Tnnt2, Myh6, and Myh7), myofibrilar proteins (e.g. Actc1, Ttn), and transcription factors (TFs) critical to heart formation (e.g. Mef2c, Myocd, and Ankrd1). GeneCluster-130 highly correlates with cluster thirteen (pink), which contains all cells from TSS (D). The following nineteen over-dispersed gene sets were pre-defined pathways that all described biology of contractile cells except for one de-novo gene set consisting of mitochondrial proteins. All the pre-defined gene sets correlated with most TSS cells. In addition, the aspect explaining the most variance across our data set positively correlates only with cluster thirteen and contained pre-defined gene sets for contractile fiber part (GO:0044449) and sarcomere (GO:0030017), which are defined as a fiber composed of actin, myosin, and associated proteins found in cells of smooth or striated muscle and the repeating unit of a myofibril in a muscle cell. These data are consistent with the developmental timing for differentiation of the first CMs in the forming primitive heart and indicate that cluster thirteen contains CMs.

The expression of key marker genes by heatmap allowed us to assign preliminary identities to each cluster (Figure 2 c, and Table 2.1): YS-Blood Islands, Blood Progenitors, Mesoderm (developing-YS), Ingressed Mesoderm, Primitive Streak, Allantois, CCPM-derived Skeletal/Myogenic Progenitors, YS-Mesenchyme, Muscle Progenitors, CCPM, Endothelium (developing YS-Vasculature), Posterior (2nd wave) Cranio-Cardio-Pharyngeal Mesoderm, Cardiomyocytes, and Muscle. These data are

consistent with the developmental timing of ingression and differentiation of these mesodermal lineages by the expression of key marker genes.

Using 3D tSNE dimensionality reduction to visualize the data, five major groups were observed: (A) one large group (bottom left) and (B) one small group (bottom right) comprising of MSS cells and some LSS cells, (C) one medium group (top right) and (D) one small group (top left) comprised of all TSS cells, and (E) one group of LSS cells (middle) (Figure 2.3 a). Visualization of key marker gene expression by tSNE identifies these major groups: (A) Eomes, primitive streak and ingressed mesoderm (Figure 2.3 b,c); (B) Runx1 and Tal1, blood and endothelium (Figure 2.3 e,f); (C) Isl1, Tcf21, Meox1, and Aldh1a2, CCPM and Muscle (Figure 2.3 i-l); (D) Tnnt2 and Myh6 CMs (Figure 2.3 g,h); and (E) Hand1, mesodermal intermediates (Figure 2.3 d). Importantly, clusters and their organization were coherent between the t-SNE visualization and hierarchical clustering.

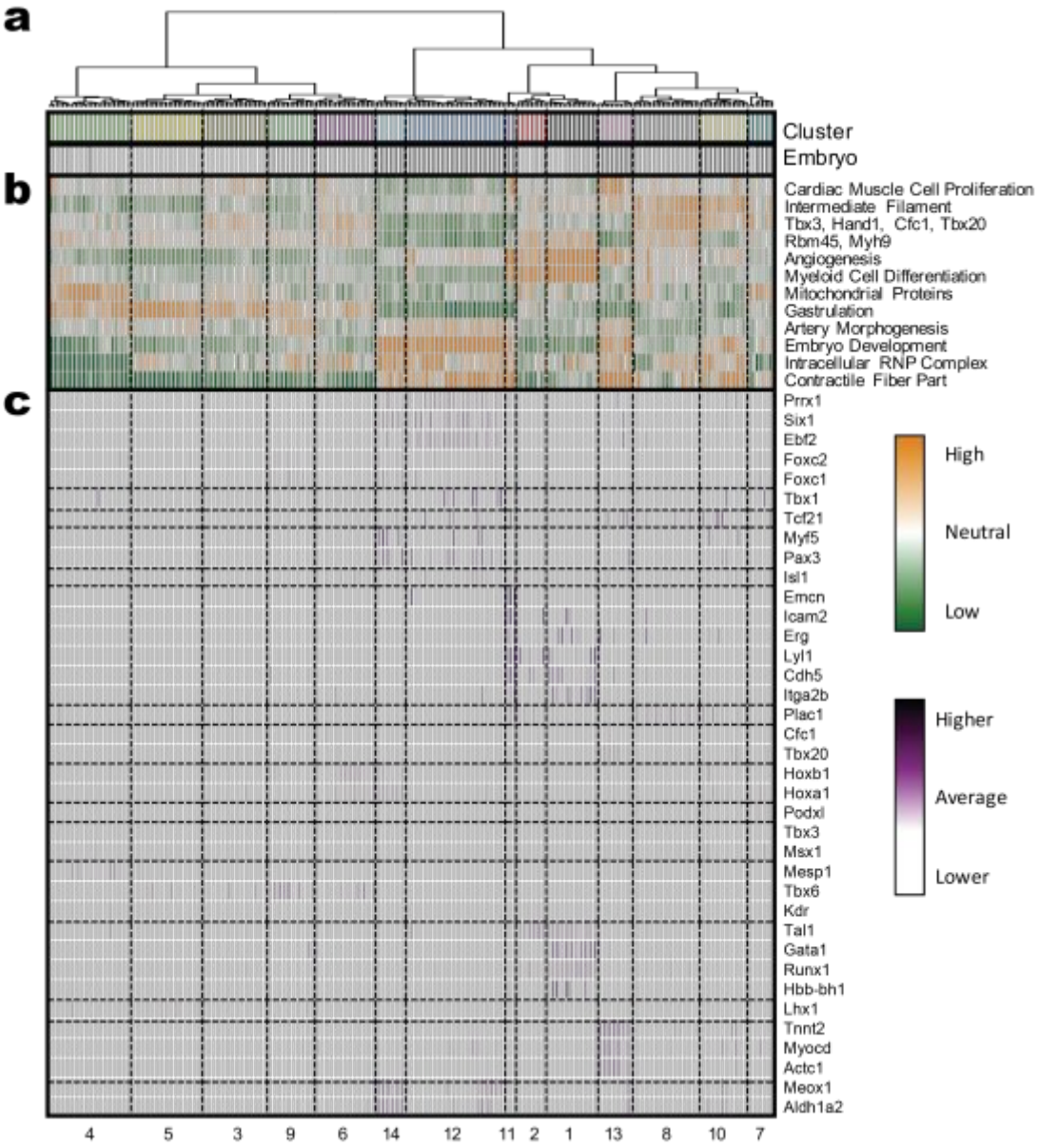


Figure 2.2: Single Cell Transcriptomics Identifies Fourteen Populations Relevant to Early Mesodermal Development.

a, Hierarchical clustering (k=14) of all cells by PAGODA. Colored bars indicate assigned cluster (top) and embryo (bottom: Light grey, Mid Streak; mid grey, Late Streak; and dark grey, Three Somites). **b,** Heatmap of top GO term or de novo gene cluster for each aspect of variability from PAGODA (12); $\mu=0$ $\sigma^2=0.2$. **c,** Heatmap of key genes distinguishing the fourteen clusters; log normalized: $1e^0$ - $1e^3$. Bottom, cluster number; right, respective scale bars.

Table 2.1: Summary of Cell Cluster Identities:

Cluster Number	Number of Cells	Cluster Color	Main Stage	Assigned ID	Key Genes and Evidence
1	16		LSS	Yolk Sac Blood Islands	Runx1 (Tanaka Y H. M., 2012), Itga2b (Ferkowicz MJ, 2003), Gata1 (Tanaka Y S. V., 2014), Hbb-bh1
2	9		MSS	Blood Progenitors	Tal1
3	20		MSS	Mesoderm (developing Yolk Sac)	Kdr (Yamaguchi TP, 1993), Cfc1 (Shen MM, 1997), Msx1
4	26		MSS	Ingressed Mesoderm	Mesp1 (Saga Y M.-T. S., 1999), Lhx1 (Barnes JD, 1994)
5	22		MSS	Primitive Streak	Mesp1 (Kitajima S, 2000), Tbx6 (Chapman DL, 1996)
6	19		LSS	Allantois	Tbx20 (Kraus F, 2001), Cfc1, Hoxb6 (Becker D, 1996)
7	8		TSS	CCPM-derived Skeletal/Myogenic Progenitors	Podxl (Chan SS., 2016)
8	21		LSS	Yolk Sac Mesenchyme	Plac1, Tbx20 (Carson CT, 2000), Msx1 (Catron KM, 1996)
9	15		LSS	Late Primitive Streak	Tbx6
10	15		TSS	Lateral / Paraxial Mesoderm	Isl1, Tcf21 (Robb L, 1998), Pax3
11	4		TSS	Endothelium (developing Yolk Sac-Vasculature)	Erg (Vlaeminck-Guillem V, 2000), Emcn (Brachtendorf G, 2001), Cdh5 (Tanaka Y H. M., 2012)
12	31		TSS	Posterior Cranio-Cardio-Pharyngeal Mesoderm	Tbx1 (Simrick S, 2012), Prrx1 (Leussink B, 1995), Ebf2, Hoxa1 (Dupé V, 1997)
13	11		TS	Cardiomyocytes	Myocd (Espinoza-Lewis RA, 2014), Actc1 (Smart N, 2002), Tnnt2 (Tamplin OJ, 2008)
14	9		TS	Muscle	Meox1, Myf5

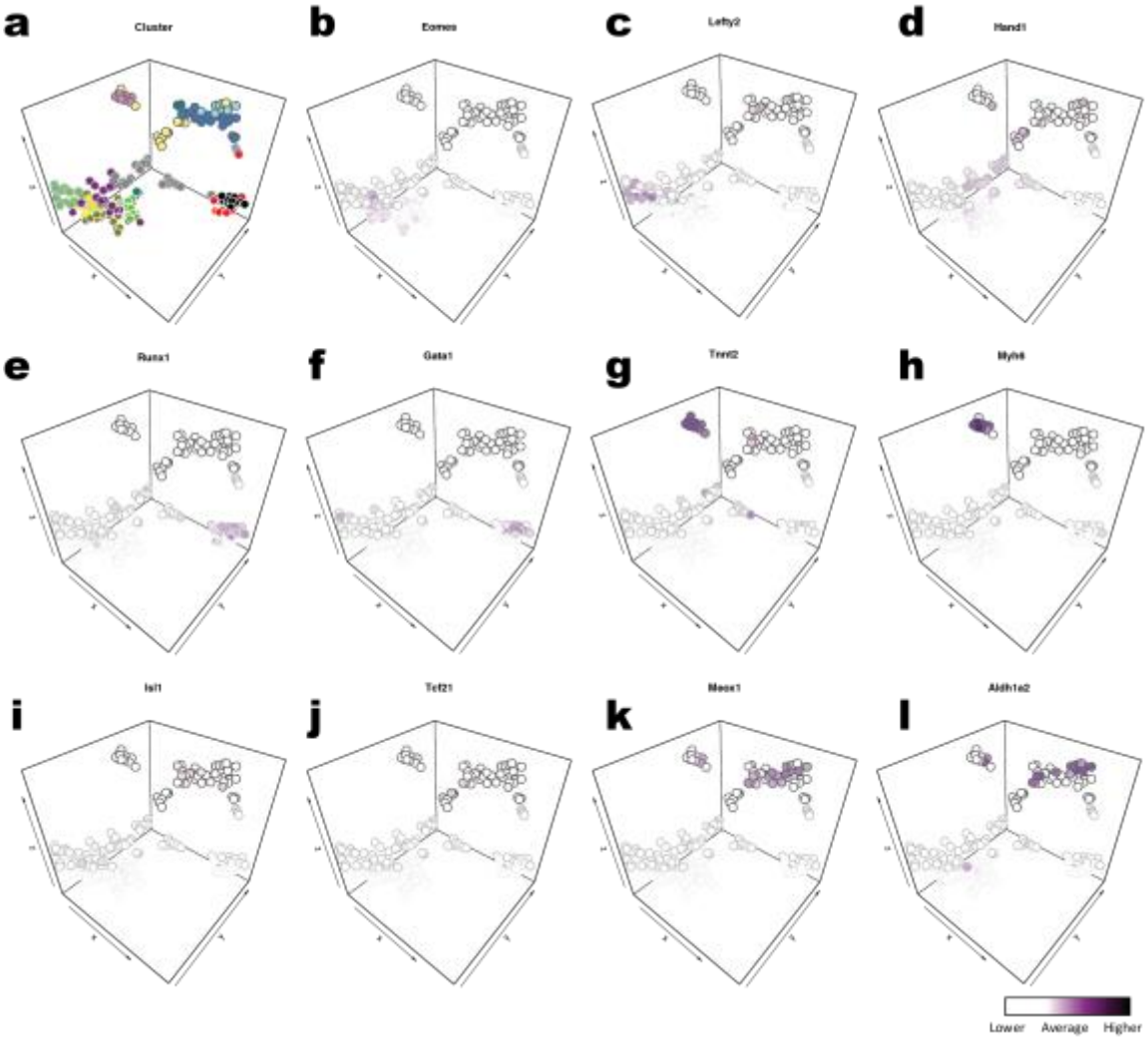


Figure 2.3: Dimensionality Reduction Reveals Transcriptional Profiles Associated with Spatiotemporal Position in the Embryo.

3D tSNE plots (perplexity = 10) based on all genes showing the separation of the cells into discrete groups congruent with the clustering. Each circle represents a cell. Border of each cell indicates embryo of origin. Cells colored by cluster (a) or spatiotemporal marker gene expression (b-l): Eomes (primitive streak), Lefty2 (ingressing mesoderm), Hand1 (extraembryonic mesoderm), Runx1 (hemogenic endothelium), Gata1 (yolk sac blood islands), Tnnt2 (cardiogenic cells), Myh6 (cardiomyocytes), Isl1 (second wave cranial-cardio-pharyngeal cells), Tcf21 (brachial arches), Meox1 (myogenic cells), and Aldh1a2 (posterior mesoderm). Bottom-right: scale bar.

2.3.3 *Developmental Ordering of Cells of Mesodermal Cell Lineages*

We next sought to order all cells and analyze transitions in their transcriptional states as they differentiate to identified cell types. Slingshot assigned developmental positions of cells along the DLTs (analogous to pseudotime (Trapnell C, 2014) by orthogonal projection of each cell's principal coordinates onto its respective curve (Figure 2.4 a-c) (Fletcher RB, 2017). From this analysis, it is evident that there are six major DLTs, in the following order: blood, endothelium, CCPM, CMs, muscle, and ExEM (Table 2.2). In addition, there appears to be three major DLTs: (1) blood and endothelium, (2) CCPM, CMs, and muscle, and (3) ExEM. These major DLTs may represent mesodermal intermediates with differing lineage potentials. These data are consistent with fate mapping studies of mesodermal progenitors.

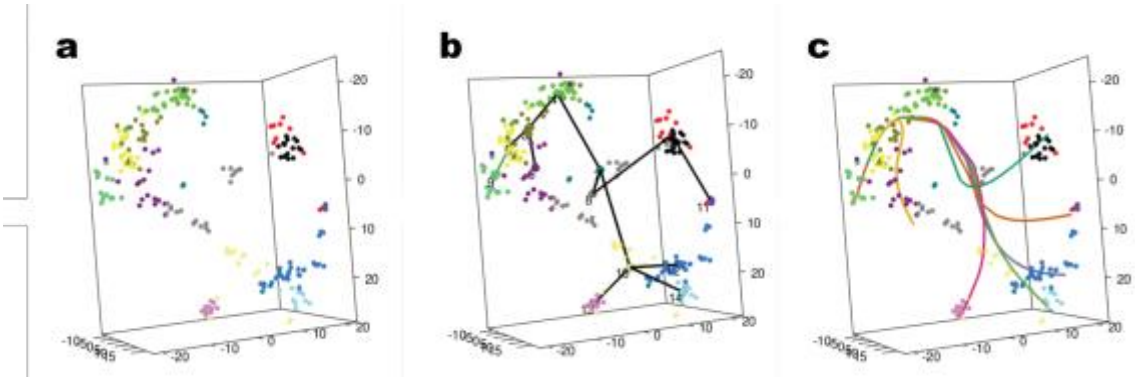


Figure 2.4: Statistical Analysis of ScRNAseq Data Predicts Cell Trajectories and Branch Points Relevant to Mesodermal Development.

a, 3D t-SNE plot (perplexity = 10) based on all genes showing the separation of the cells into discrete groups congruent with the clustering. Each circle represents a cell. **b,** Unsupervised cluster-based Minimum Spanning Tree (MST) to stably identify the key elements of the global lineage structure, i.e., the number of lineages and where they branch. Cluster medioids are displayed as larger circles with initial assignments of cluster identity based on the expression of a small number of marker genes represented by cluster number. **c,** Conversion of MST into smooth lineages represented by one-dimensional variables called "pseudotime". Colored lines represent different trajectories.

Table 2.2: Summary of Developmental DLTs Identified by Slingshot

Trajectory Number	Final Cluster Number	Final Cluster Name	Lineage Path (by Cluster No.)	Trajectory Color	Main Stage	Assigned Trajectory	Key Genes and Evidence
1	1	YS-Blood Islands	"9" "5" "3" "4" "7" "8" "2" "1"		LS	YS-Blood Islands	Runx1, Fli1, Tal1, Lmo2, Sox7
2	11	Endothelium (developing YS-Vasculature)	"9" "5" "3" "4" "7" "8" "2" "11"		FS	Endothelium	Egr1, Lyl1, Tie1, Cd38, Cd34
3	12	Posterior CranioPharyngeal Mesoderm	"9" "5" "3" "4" "7" "10" "12"		FS	Cardiopharyngeal Mesoderm	Prrx1, Ebf2, Tbx1, Col2a1
4	13	Cardiomyocytes	"9" "5" "3" "4" "7" "10" "13"		FS	Cardiomyocytes	Myocd, Tbx5, Actc1, Tnnt2, Myh6
5	14	Muscle	"9" "5" "3" "4" "7" "10" "14"		FS	Paraxial Mesoderm	Meox1, Pax3
6	8	YS-Mesenchyme	"9" "5" "3" "6"		LS	ExtraEmbryonic Mesoderm	Tbx3, Hand1, Bmp4, Tbx20

2.3.4 *Patterns of Coordinated Gene Regulation in Mesodermal Cell Lineages Reveal Different Strategies for Specification and Differentiation*

To gain insight into the coordinated patterns of gene expression that underlie the cell fate transitions in these DLTs, we identified the most differentially expressed genes across pseudotime within each trajectory lineage. The expression profiles for the twenty most differentially expressed genes are displayed using heatmaps (Figure 2.5 a-f) are presented in developmental order according to the predictions made by slingshot (Table 2.2). This comparison highlights the dramatic difference in coordinated gene expression through developmental progression in these six lineages.

This analysis identified canonical marker genes of each lineage towards the end of each pseudotime with differing dynamics. Since the most differentially expressed genes tend to identify genes at the beginning and end of each lineage trajectory, further analyses will be required to identify genes expressed prior to differentiation of each cell type. Nevertheless, due to the identification of canonical lineage markers in all six DLTs by differential expression within each trajectory, these data provide a valuable resource for identifying and investigating newly implicated genes in mesodermal lineage development.

To gain a more comprehensive view of the six DLTs and how they relate to one another, cell cluster medioids used to generate DLTs and the overall DLTs were further clustered (Fig. 7a). Analysis of the dendogram for DLTs (rows) further suggest that there are three major branches DLTs: (1) ExEM, (2) blood and endothelium, and (3) CCPM, CMs, and muscle. In addition, these data highlight the presence of cell groups (columns) specific to DLTs that may contain cells representative of branch point intermediates to

different DLTs, such as cluster eight (gray) for blood and endothelium and cluster ten (light yellow) for CCPM, cardiomyocytes, and muscle. These data have been summarized in a model of the mesodermal hierarchy (Fig. 7b).

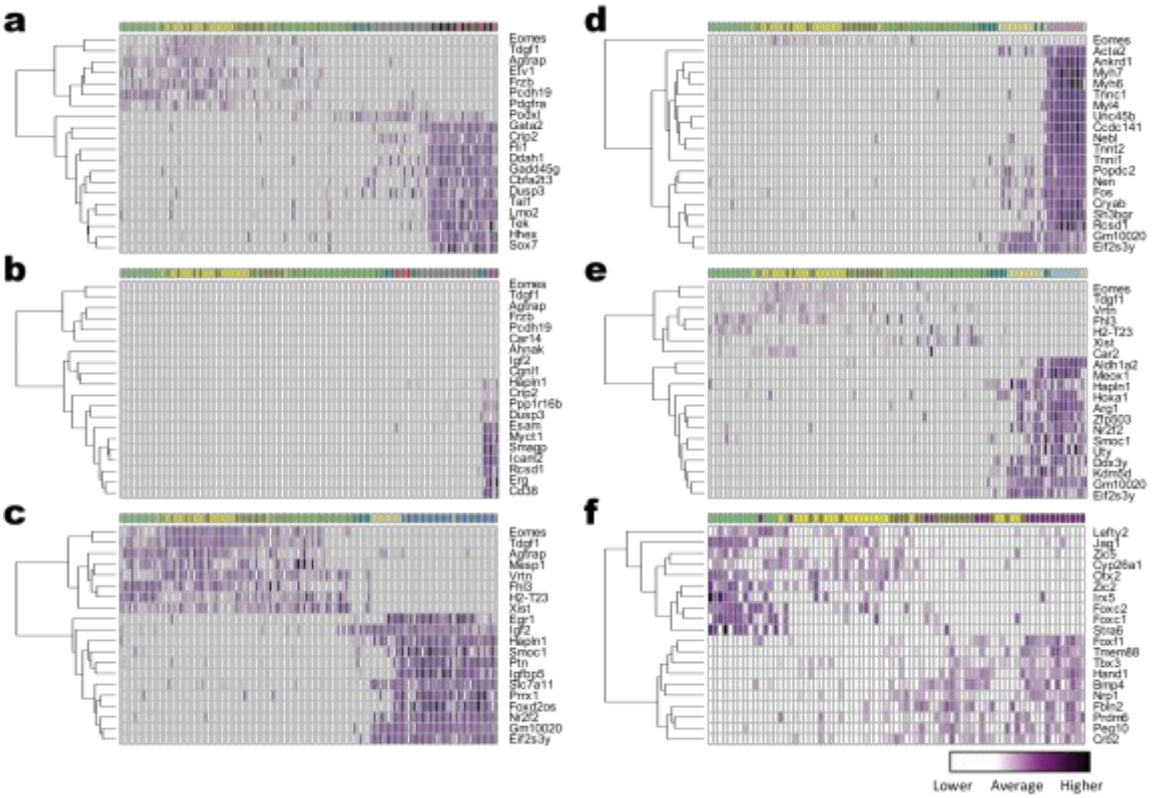


Figure 2.5: Pseudotime-Dependent Differential Expression within Trajectories Identifies Genes Implicated in Mesodermal Lineage Development.
Heatmaps displaying the average scaled expression profile for the twenty most differentially expressed genes as a function of pseudotime (rows) along each trajectory ordered according to their developmental positions (columns). **a,** Blood; **b,** Endothelium; **c,** Second Wave Cranio-Cardio-Pharyngeal Mesoderm; **d,** Cardiomyocytes; **e,** Muscle; **f,** Extra-Embryonic Mesoderm.

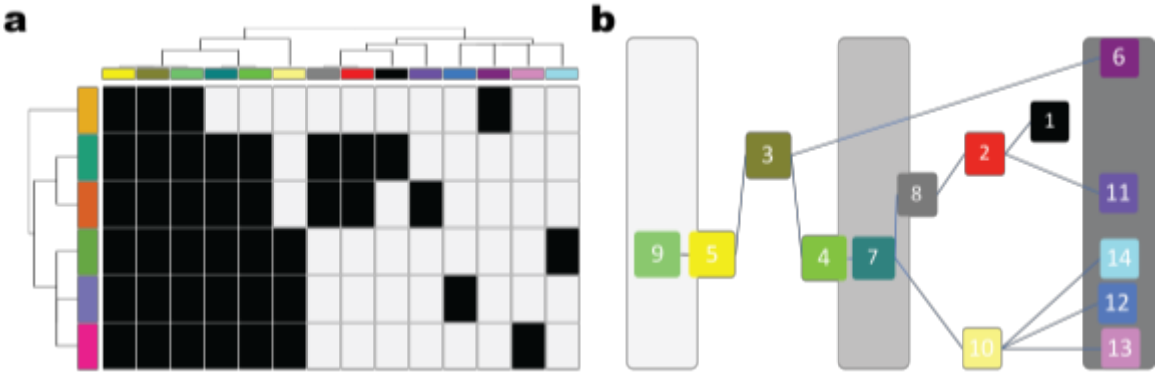


Figure 2.6: Computational Model of Developmental Lineage Decisions in the Mesodermal Hierarchy.
a, Heatmap of the cell clusters present in each slingshot trajectory. **b,** Schematic interpretation of global structure of the mesodermal hierarchy. Colored bars are representative of embryo stage.

2.4 Conclusions and Discussion

2.4.1 *Summary*

The studies detailed in this chapter provide insight into murine mesoderm development and diversification. We used scRNAseq to systematically catalog the diversity of intermediate cell states formed during gastrulation and how they relate to differentiated mesodermal cell types. We molecularly define six major cell types including blood, endothelium, CCPM, CMs, muscle, and ExEM, and the cellular trajectory for each cell type. We define three main cellular intermediates: (1) blood and endothelium, (2) CCPMs, CMs, and muscle, and (3) ExEM. The ordering of DLTs by spatiotemporal timing corroborates existing experimental data generated by lineage tracing and clonal analysis of mesodermal progenitors during gastrulation. Further analyses on these developmental trajectories and putative cellular intermediates to define the molecular trajectory of mesodermal cell types and intermediates.

2.4.2 *Testable Predictions from In Silico Branching Lineage Assignments*

Several testable predictions can be made from the in silico branching lineage assignments and developmental ordering of differentiating mesodermal cell types. First, the initial branching of ExEM from the remaining mesoderm fated for the embryo proper through as early as MSS suggests that there is a cellular intermediate fated for ExEM. In this scenario, there exists two multipotent progenitor populations with potential only to either the embryo proper or the ExEM, except for blood and endothelium that occur in both. This hypothesis can be tested through analysis of single cell lineage tracing or

multi-colored clonal analysis (e.g. brainbow) of Mesp1-expressing cells for their contribution to either embryo proper or ExEM.

Second, the subsequent branching of blood and endothelial lineages from CCPM, CMs, and muscle at LSS suggests that there are two cellular populations at this stage with differing potency. The functional evidence for a multi-potent intermediate for CCPM, CMs, and muscle at LSS would identify a previously undescribed progenitor, while a bi-potent blood and endothelial intermediate (i.e. hemangioblast) would be highly controversial. Analysis into genes specifically expressed by these two populations is critical to providing molecular markers to test *in vivo* to validate the hypothesis of such an intermediate existing. These hypotheses can be tested through analysis of single cell lineage tracing or multi-colored clonal analysis (e.g. brainbow) of individual cells expressing branch point specific genes for their clonal contribution to only either blood and endothelium or CCPM, CMs, and muscle.

2.4.3 *Future Directions*

Future analysis on the dynamics of coordinated TF interactions by examining lineage-specific cellular intermediates will provide greater insight into the genetic regulation of mesodermal specification. However, the possibility that subtle, additional intermediates in cell states exist that were not resolved by the present study due to that we analyzed only 226 cells from three stages. Future studies correlating these trajectory-specific gene expression data with available ChIP-seq data for histone states and TFs during murine mesodermal development at corresponding stages and cell types *in vitro* and *in vivo* would provide further insight into the genetic regulation of mesodermal

specification at branch points. Nevertheless, the approach presented in this chapter enabled an analysis of mesodermal progenitor potential and the mapping of branch points along trajectories with cellular and molecular resolution not possible by either *in vivo* lineage tracing alone.

2.5 Acknowledgements

This chapter is an original document written by Jonathan D. Grinstein with oversight from Neil Chi and Sylvia Evans. The work presented in this chapter was led by Jonathan D. Grinstein with contributions from the laboratories of Neil Chi (Neil Tedeschi and Elie Farah), Sylvia Evans (Paola Cattaneo), and Gene Yeo (Yan Song and Leen Jamal-Schafer. Jonathan D. Grinstein was supported by a pre-doctoral fellowship from the California Institute of Regenerative Medicine (CIRM). Neil Chi is an Associate Professor of Medicine at UCSD and is supported by grants from the National Institutes of Health (NIH) and the American Heart Association.

Chapters 2 is currently being prepared for submission for publication of the material. Jonathan D. Grinstein, Paola Cattaneo, Josh Bloomekatz, Neil Tedeschi, Elie Farah, Yan Song, Leen Jamal-Schafer, Nuno Camboa, Gene Yeo, Sylvia M. Evans, Neil C. Chi. The dissertation/thesis author, Jonathan D. Grinstein, was the primary investigator and author of this material.

Chapter 3: Reconstructing the Hematopoietic Hierarchy via *In Vivo* Lineage Tracing of Lgr5

3.1 Abstract

This chapter is adapted from work led by Jonathan D. Grinstein under the supervision of Neil Chi. This chapter combines genetic, NGS, and computational experimental strategies described in previous chapters to complete an in-depth examination of the hematopoietic developmental hierarchy. Here we further analyze the most variant genes in hematopoietic lineage pseudotime to define the hematopoietic molecular trajectory by key genes. We define a transient hematopoietic intermediate during gastrulation labelled by master regulator of hematopoiesis Runx1, cardiac structural protein Actc1, and a previously unassociated stem cell marker Lgr5. We demonstrate by *in vivo* lineage tracing that Lgr5-expressing cells are located primarily in ExEM YS-BI and contribute to several circulating erythrocytes and sites of transient developmental hematopoiesis including the outflow tract, brain, and umbilical cord. Together, these studies provide evidence for a novel precursor to transient hematopoiesis, thus challenging the hierarchy of developmental hematopoiesis.

3.2 Introduction

As gastrulation begins, cells at the intersection between embryo proper and ExEM ingress and migrate around the surrounding prospective ectoderm contributing either to the embryo proper or the ExEM region to form the YS, umbilical cord and placenta (Kinder SJ T. T., 1999). Fate mapping studies have identified the specific regions blood lineages initiate from in the pregastrula epiblast (Parameswaran M, 1995; Tam PP P. M.,

1997), but the identity of these cells within the embryo, the exact cell types they give rise to, and the function of key cell-type-specific TFs remain elusive.

3.2.1 *Temporal Waves of Developmental Hematopoiesis*

Hematopoietic cells emerge in the embryo during development in temporal waves with differing cellular potentials (Tober J, 2016). First, primitive hematopoietic progenitors emerge from YS-BI as early as E7.0/MS and produces a large pool of unipotent progenitors to primitive erythrocytes and a small pool of bi-potent primitive megakaryocyte/erythrocyte progenitors and primitive macrophages (Palis J. R., 1999). Primitive erythrocytes are larger and different hemoglobin proteins than adult erythrocytes (Kingsley, 2004; Palis J. , 2014). In addition, primitive megakaryocytes are diploid, which are distinct from polyploid adult megakaryocytes (Potts, 2014). As circulation of blood begins, progenitors to macrophages migrate from YS to differentiate, a process that remains unclear, and become dispersed throughout the embryo (Herbomel, 1999). These primitive macrophages are the only known cells from this first, primitive hematopoietic wave that have been shown to continue to contribute to hematopoiesis in adults, whereby they supply microglia in the brain and tissue resident macrophages (Ginhoux, 2010; Gomez Perdiguero, 2015).

The emergence of cells that contribute to HSCs begins with and is defined by the second wave of hematopoiesis (Nakano H, 2013). This second wave can be first observed at around E8.25 in the YS with the emergence of multipotent erythroid-myeloid progenitors (EMPs). At E9.5, lymphoid progenitors emerge from the YS and caudal para-aortic splanchnopleura (PSp), which is LPM derived. Both EMPs and lymphoid

progenitors are distributed in the dorsal aorta connecting to the YS and placenta, the placenta itself, and the umbilical and vitelline arteries. These progenitors do not have the ability to repopulate irradiated adult mice or even embryonic development. At around E9.0 the second wave produces neonatal HSCs, defined by their ability to engraft in neonatal but not adult mice, and pre-HSCs, which cannot functionally engraft into bone barrow but can give rise to adult HSCs, and emerge from the PSp and umbilical and vitelline arteries. Many of these preHSCs, although unable to engraft an adult mouse, will migrate to the liver where they will differentiate into functional, repopulating HSCs.

The third and final wave of hematopoiesis is defined by the emergence of the long term HSC (ltHSC), the self-renewing, repopulating HSC capable of engrafting and suppling multi-lineange hematopoietic cell types (Yoder MC., 2014). HSC emergence is first observed in the aorta-gonad-mesenephros (AGM), umbilical and vitelline arteries, placenta, and head in mice at around E10.5. These HSCs will colonize the liver and, ultimately, bone marrow where they will provide the entirety of blood, except a macrophage subtype, for all of post-neonatal life.

3.2.2 *The Requirement of Runx1 for Hematopoiesis*

Runx1 is required for the emergence of all hematopoietic cell types except for primitive erythrocytes and megakaryocytes from first wave hematopoiesis (Cai Z, 2000; Chen MJ Y. T., 2009; Chen MJ L. Y.-I., 2011). Interestingly, genetic deletion of Runx1 causes loss of definitive hematopoiesis, but the formed vasculature is, for all intents and purposes, intact. Investigation of Runx1 expression using transgenic mice expressing β-galactosidase (LacZ) upon endogenous Runx1 expression identified LacZ+ cells in blood

progenitors in the vessel lumen connected to endothelial cells at the same sites where definitive hematopoiesis had been observed. Subsequently, molecular and functional experiments went on to identify that these Runx1+ endothelial cells were precursors to blood and, thus, termed hemogenic endothelium.

3.2.3 *The Role of Lgr5 in Wnt Signaling and Stem Cell Identity*

Since the first description of "tissue renewal" by Leblond and Walker, in which they observed tissues that maintained size regardless of constant cellular proliferation (Leblond CP W. B., 1956), it has become evident that, akin to development, stem cells are required for not only growth but tissue homeostasis. There are several adult tissues such as the skin, blood, and gut with remarkable cellular turnover that are repopulated daily by self-renewing adult stem cells. These adult stem cells reside in structured microenvironments (the niche) at the interface of responding to developmental signals to endure the demands of homeostasis and regeneration. One class of these developmental signals known as Wnt signaling is critical to stem cell identity, function, and self-renewal for multiple tissues (Zeng YA, 2010).

The Wnt signaling system is composed of secreted Wnt signaling proteins that are biochemically tetherable to cell membranes and generally work over short distances (Najdi R, 2012). Wnt signaling proteins bind cognate receptors Frizzled and co-receptors Lrp5/6 to form a complex (Janda CY, 2012). In the absence of Wnt signaling proteins, the transcriptional co-activator beta-catenin is continually targeted to the proteasome for degradation by glycogen synthase kinase (GSK3) (Stamos JL, 2014). In the presence of Wnt signaling proteins, GSK3 is sequestered and thereby inhibited to degrade beta-

catenin, which accumulates in the nucleus where is binds Tcf/Lef TFs to engage expression modules related to proliferation, genome integrity and maintenance (Hoffmeyer K, 2012; Schepers AG, 2011), and survival. In addition, Wnt signaling is not just a binary "on/off" signaling cascade, but can be tuned by another class of secreted Wnt proteins named R-Spondins that bind to Lgr family receptors to inhibit the degradation and enhance the stabilization of Frizzled receptors (Glinka A, 2011; de Lau W, 2014).

The fastest proliferating tissue in adult mice is the small intestine epithelium, which is lined with villi that constantly shed differentiated cells from their apexes that replenishes about every five days (Leblond CP S. C., 1948). To replenish this constant loss of cells, a microenvironment at the base of each called the intestinal crypt contains stem cells. This microenvironment is dependent on Wnt signaling as shown by the loss of tissue renewal upon genetic disruption of Wnt signaling components in the intestine. At the bottom of the crypt resides a ring of cycling, stem cells called crypt base columnar cells (CBCs) that have the potential to differentiate into and maintain all intestinal cell types. Expression profiling of CBCs drew attention to the receptor Lgr5, which facilitated the lineage tracing of CBCs by Lgr5 to genetically demonstrate the mechanism of intestinal self-renewal (Muñoz J, 2012).

These findings, which put Lgr5 in the spot light, led to an explosion of investigation in adult stem cell populations identified and regulated by Wnt signaling and augmentation by Lgr receptors. Many different adult stem cell populations have been identified as Wnt signaling responsive and Lgr5 receptor mediated including but not limited to the hair follicle, kidney, and cochlea (Clevers H, 2014; Kinzel B, 2014).

Genetic disruption of Lgr5 display 100% neo-natality that has been attributed craniofacial disorder ankyloglossia leading to the absence of ingestion of milk, exhibit gastrointestinal distention and dilation, and cyanosis elucidated the role of Lgr5 in embryonic development (Morita H, 2004). Lgr5 expression profiling during murine embryonic development detected expression in proliferating hematopoietic stem and progenitor cells (HSPCs) in the AGM and fetal liver, sites of definitive hematopoiesis (Liu D, 2014). Transplantation of these Lgr5-expressing HSPCs into irradiated animals revealed the functional limitation of these cells as short term hematopoietic stem cells (stHSCs) of second wave hematopoiesis (Liu D, 2014).

3.2.4 *Chapter 3 Research Strategy*

There is extensive data on the anatomical location of hematopoietic cell emergence and the identity of cell types generated by the three waves of hematopoiesis, however the molecular identity of precursors to each hematopoietic wave and the genetic mechanism regulating these fate decisions remain unclear. Here we use computational and genetic strategies to investigate co-expression of hemogenic endothelial marker Runx1, hematopoietic marker Gata1, and candidate marker Lgr5 in developmental hematopoiesis. We identify a molecular signature for precursors to each hematopoietic developmental wave based on the combinatorial expression of Runx1, Gata1, and Lgr5, and provide suggestive lineage tracing evidence that Lgr5-expressing cells during gastrulation contribute to first and second wave, but not third wave, hematopoiesis.

3.3 Results

3.3.1 *Lineage Tracing of Lgr5 In Vivo Validates Hematopoietic Branching Lineage Assignment Predictions*

We next sought to identify putative markers specific to cell fate transitions in these DLTs by examining the differential expression results along lineage pseudotime. Differential expression analysis along the hematopoietic lineage trajectory identified several putative marker genes including microrchidia 4 (Morc4), nuclear receptor interacting protein (*Nrip3*), and leucine rich repeat containing G protein coupled receptor 5 (Lgr5) (Figure 3.1a).

For validation studies, we focused on Lgr5 for a variety of practical and reasons: (1) Lgr5 has been extensively studied as a regulator of both embryonic and adult stem cell populations in various tissues from all germ layers including such as the gut, kidney, skin, and ear (Barker N, 2013); (2) the availability of validated research tools including transgenic mice and molecular tools to study the expression and function of *Lgr5* due to its popularity (Grün D, 2015); (3) *Lgr5* is a member of the Wnt signaling pathway, which has been shown to be required for embryonic hematopoiesis (Sturgeon CM, 2014); (4) Lgr5 is a cell surface receptor that can be used for cell purification . In addition, we found that Lgr5 was not just co-expressed with canonical hematopoietic regulators such as Runx1 (Fig 3.1 b), Tal1, Hhex, Fli1, Lmo2, Egfl7, Sox7, and Gata2, it was co-expressed with smoothened-like 2 (Smtnl2) and Crip2 (Figure 3.1a), which both have been shown to label endocardial cells at a known site of embryonic hematopoiesis in the outflow tract *in vivo* (Narumiya H, 2007) (Wei TC, 2011), and canonical cardiac muscle marker Tnnt2 (Figure 3.1 c,d).

To test the prediction that Lgr5 labels the hematopoietic branching trajectory by contributing to primitive or definitive hematopoiesis, we performed lineage tracing experiments using transgenic mice carrying alleles for an Lgr5-dependent doxycycline-inducible Cre (Lgr5:EGFP:ROSA:creERT2, here on abbreviated as LERC). Lgr5 expressing cells were labeled by injecting doxycycline at E6.75 to activate Cre recombinase activity prior to our first observed expression of Lgr5 in our scRNAseq data at MSS in LERC;R26R:tdT mice. Mice were sacrificed and harvested labelled embryos at E10.25 when both primitive and definitive sites of hematopoiesis have been initiated (Figure 3.2 a). Consistent with the prediction by differential expression along the hematopoietic branching lineage, we observed tdT+ erythrocytes in circulation (Figure 3.2 a-g) and in the outflow tract region (Figure 3.2 h-o) of embryos at E10.25 when *Lgr5* cells were labeled as early as E6.75. In addition, we observed tdT+ staining in the head region in cell types with either endothelial or red blood cell morphologies. Further analysis on tissue sections stained with canonical endothelial and blood markers are required to confirm the exact cell type of these tdT+ cells.

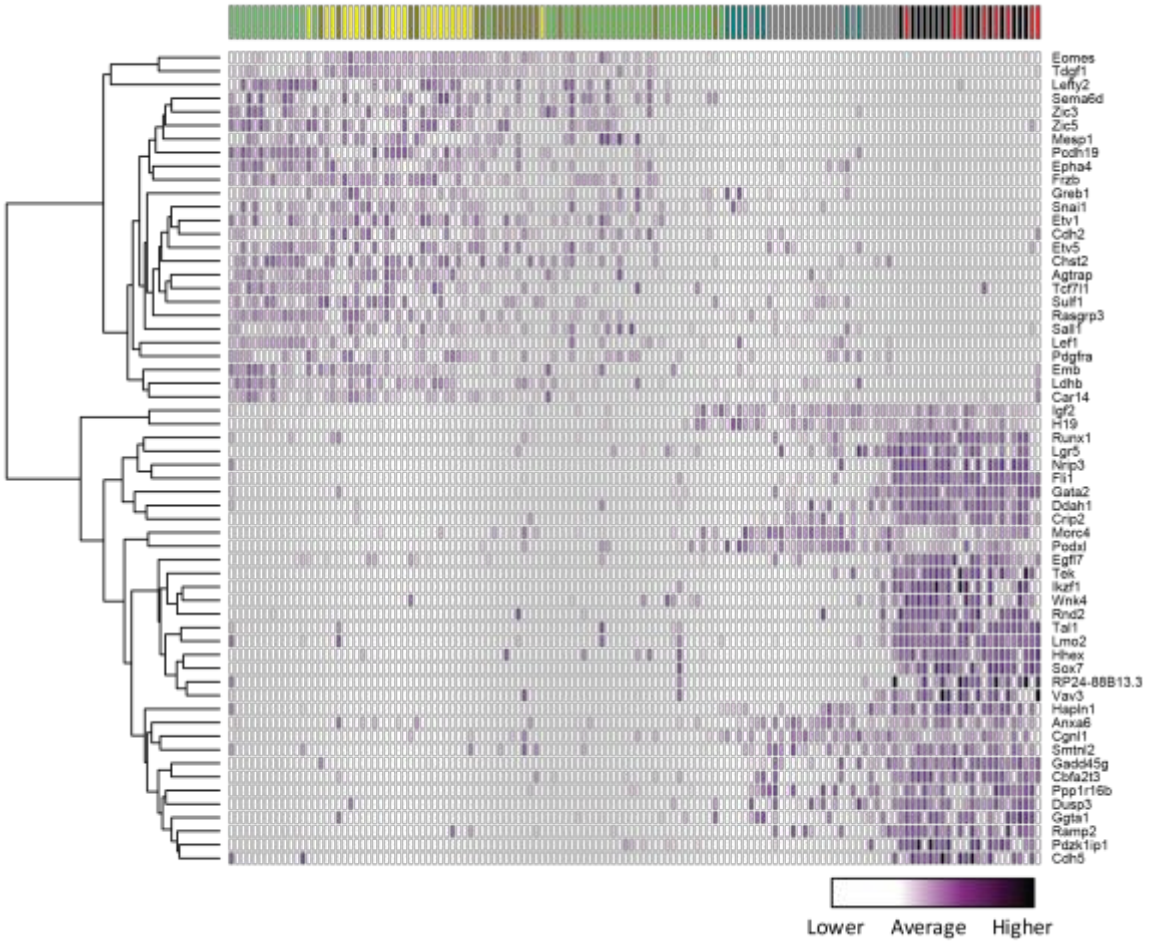


Figure 3.1 Pseudotime-Dependent Differential Expression within the Blood Trajectory Identifies Lgr5 as a Putative Marker of Hematopoiesis.
Differential expression of genes as a function of pseudotime along the blood trajectory. Top sixty most differentially expressed are shown. Representative hemogenic markers *Runx1* and *Tal1* are co-expressed with putative marker Lgr5. Bottom right, scale bar for all plots.

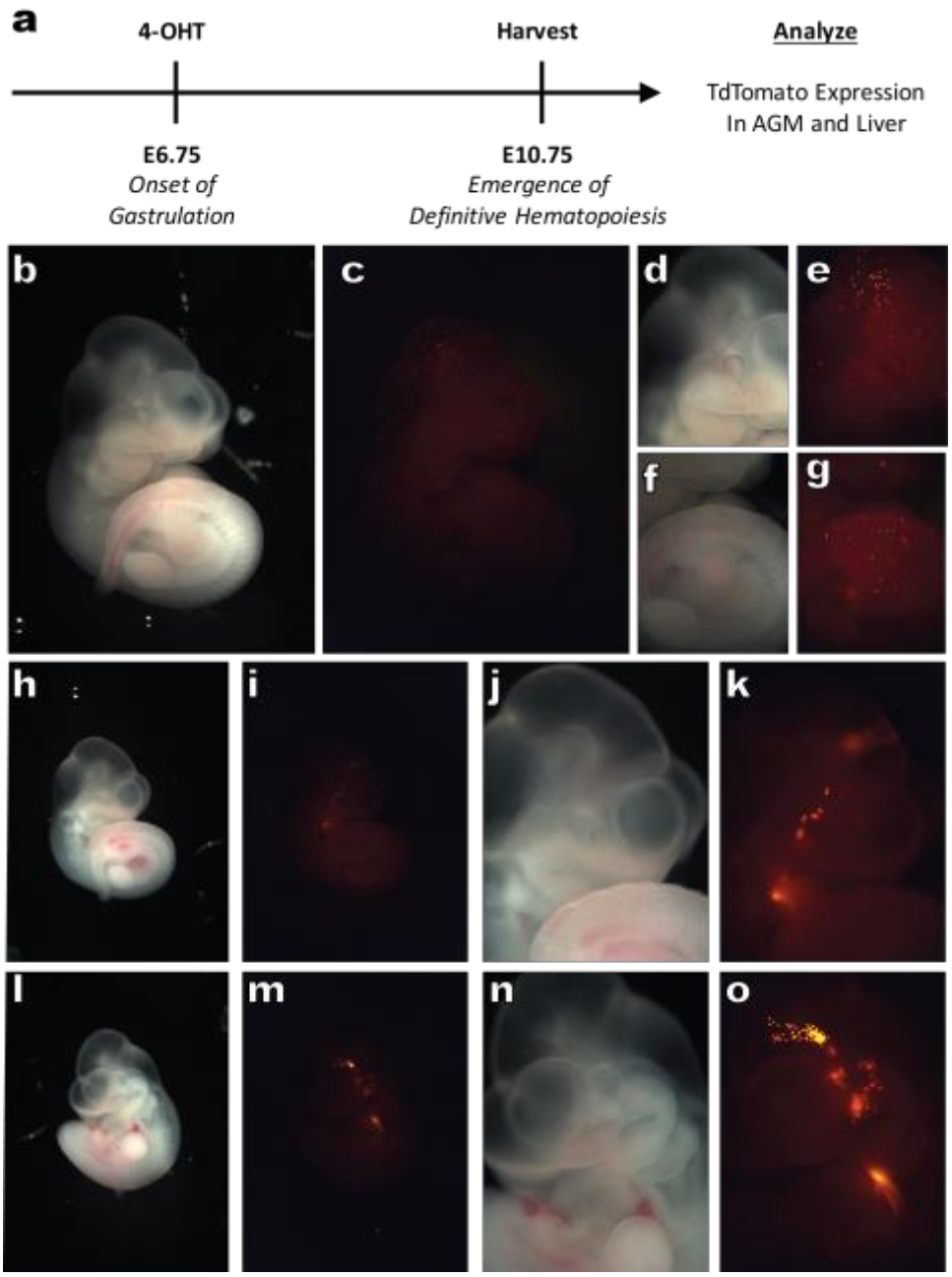


Figure 3.2: Lgr5-Expressing Cells during Gastrulation are Precursors to Circulating Erythrocytes and Endothelium at Putative Sites of Transient Hematopoiesis.

A, Experimental strategy. **b-o** Whole mount images of two representative embryos (b-g and h-o) from one litter of Lgr5:EGFP:ROSA:creERT2/+;R26R:tdT/R26R:tdT mice injected with tamoxifen at E6.75 and harvested at E10.75. Lgr5 labels erythrocytes (e, g) and endothelium in the head (k), outflow tract (o), and umbilical cord (data not shown), which are putative regions of transient hematopoiesis. b-g, lateral-right; h-k, lateral-right; l-o, lateral-left.

3.3.2 *Molecular Signatures of Hemogenic Populations During Gastrulation*

We next sought to gain further insight into the cell type expressing *Lgr5* during gastrulation. Since *Lgr5* was shown unbiasedly to be co-expressed with hemogenic and cardiovascular markers, we further analyzed the co-expression of *Lgr5* with known hemogenic and cardiovascular markers in our LS embryo only, which contains the most *Lgr5* positive cells. To do so, cells from LS were independently put through the same SCDE and PAGODA pipelines used on all cells (Methods). Hierarchical clustering resulted in seven clusters (Figure 3.3 a) with three major branches (from left to right): (1) dark blue, light yellow, and grey; (2) bronze; and (3) dark yellow, light blue, and black. *Lgr5* was expressed robustly in the bronze and the dark yellow clusters. Further analysis of these two clusters reveals that in the bronze cluster *Lgr5* is co-expressed with markers of hemogenic endothelium (*Runx1)* and primitive erythropoiesis *(Tal1*, *Gata1*, and *Itga2b/Cd41)*. In the dark yellow cluster (Figure 3.3 b), *Lgr5* was also co-expressed with hemogenic endothelial marker *Runx1* (Figure 3.3 c,d), but, instead of co-expressing markers of primitive erythropoiesis, *Lgr5* was co-expressed with markers of CMs (*Tnnt2* and *Actc1*; Figure 3.3 e,f) and ExEM (*Plac1*; Figure 3.3 g). We also observed that there were various cells in the last cluster (black) that express *Runx1* but not *Lgr5*.

In order to gain insight into the expression of these two cell types, we interrogated the expression and contribution of *Lgr5*-expressing cells in LERC;R26R:tdT mice. *Lgr5* expressing cells were labeled by injecting doxycycline at E6.75 to activate Cre recombinase activity prior to our first observed expression of *Lgr5* in our scRNAseq data at MS in LERC;R26R:tdT mice. Mice were sacrificed and harvested labelled embryos at E8.25 (Figure 3.4 a). We did not observe any EGFP expression, most likely due to the

low and transient expression of *Lgr5* such that it may no longer be expressed at E8.25. However, we did observe tdT+ cells in the yolk sac in a pattern reminiscent of yolk sac blood islands. In addition, there is some labelling in the apparent gut region, however it is possible that these cells are hemogenic as the dorsal aorta and AGM lie in a similar region (Figure 3.4 b-g). Further analyses on sections of these embryos by co-expression studies of *Lgr5* with markers is required to validate these cell types in vivo during and after gastrulation.

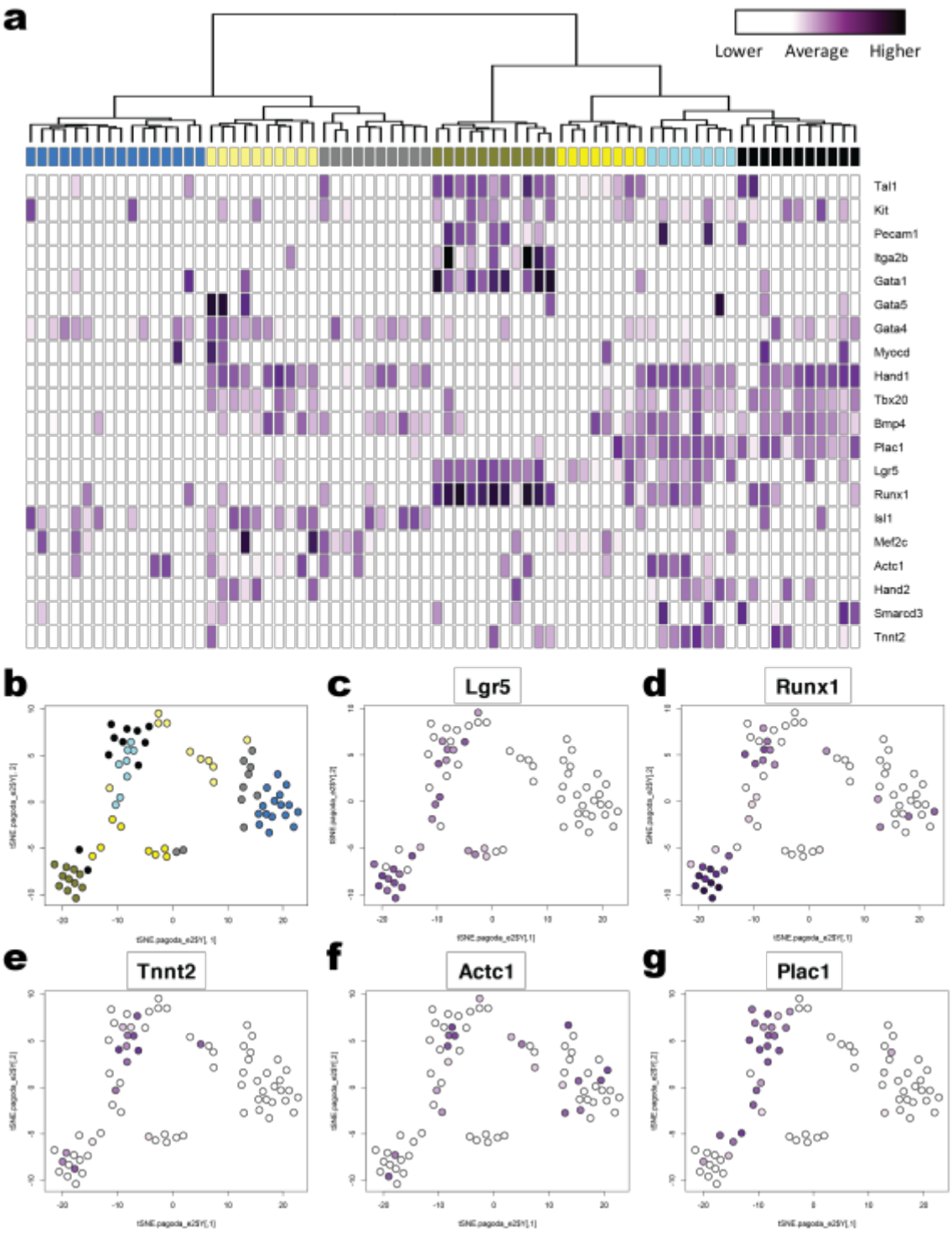


Figure 3.3: Lgr5 is Co-Expressed with Runx1 in Two Populations of *Mesp1*-labelled cells during Gastrulation.

a, Hierarchical clustering of LS cells with a heatmap of blood, cardiac, and ExEM canonical markers. **b-g,** 2D tSNE plots. Bottom right, scale bar for all plots.

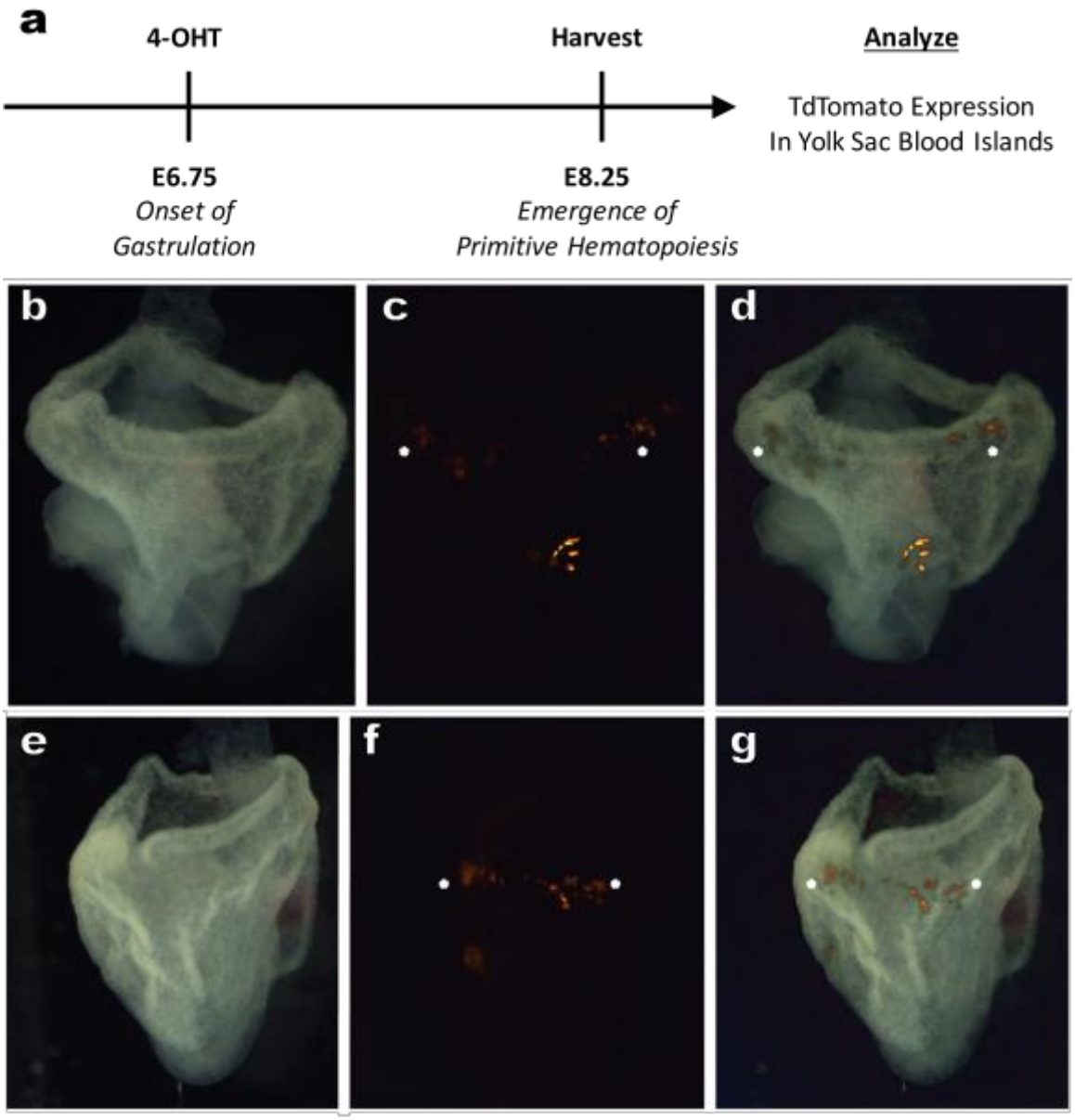


Figure 3.4: Lgr5-Expressing Cells during Gastrulation Contribute to the Yolk Sac. Whole mount images of a representative embryo from one litter of Lgr5:EGFP:ROSA:creERT2/+;R26R:TdT/R26R:TdT mice injected with tamoxifen at E6.75 and harvested at E8.25. Lgr5 labels a band of cells in the yolk sac in a pattern like YS-BI, as well as some cells in the future gut. a-c, anterior view; d-e, lateral view. **a** and **d,** bright-field images. **b** and **e,** red-channel images. **c** and **f,** merged images.

Table 3.1: Proposed Molecular Signatures for Runx1+ Hemogenic Cellular Intermediates During Gastrulation

	Runx1	Gata1	Lgr5	Actc1	Plac1
Primitive Hematopoiesis	■	■	■		
Transient Hematopoiesis	■		■	■	■
Definitive Hematopoiesis	■				■

3.4 Conclusions and Discussion

3.4.1 *Summary*

The studies detailed in this chapter provide insight into the specification of murine developmental hematopoiesis through validation of our hematopoietic lineage trajectory predictions. Specifically, we validate predictions that Lgr5 is expressed during gastrulation in the hematopoietic lineage trajectory. In addition, we find cells expressing Lgr5 during gastrulation label circulating erythrocytes and various sites of murine developmental hematopoiesis that so far include YS-BI, OFT, umbilical cord, and the brain. In addition, we found that Lgr5 is expressed in a YS-BI pattern during LS, implicating its expression at a main site of developmental hematopoiesis. Further characterization of Lgr5 expressing- and labelled-cells for co-expression of hematopoietic and endothelial cell type markers by multi-color ISH or immunofluorescence is required to define these cell types.

We show that Lgr5-expressing cells during gastrulation co-express Runx1+. This observation is interesting considering: (1) differentiation strategies for the generation of primitive or definitive hematopoietic progenitors require Wnt signaling manipulation, (2) Wnt signaling increases total Runx1 mRNA levels in human CD34+ hematopoietic progenitors, (3) a distal Runx1 promoter essential for HSCs is a direct transcriptional target of Wnt signaling important for normal hematopoiesis, and (4) Lgr5 modulates Wnt signaling by amplifying the signal in the presence of the ligand R-spondin (RSPO). This suggests that Lgr5 may function as a tuning dial to amplify the Wnt signaling gradient for specification of hematopoietic progenitor types.

In addition, these data are mostly consistent with data on the expression of two different Runx1+ populations differentially labelled by Gata1 in which Runx1+/Gata1+ cells give rise to primitive erythropoiesis and Runx1+/Gata1- cells contribute to transient and definitive hematopoiesis. However, we observe three Runx1+ populations: (1) Runx1+/Lgr5+/Gata1+ population, primitive erythropoiesis, (2) Runx1+/Lgr5+/Gata1-, transient hematopoiesis, and (3) Runx1+/Lgr5-/Gata1- definitive hematopoiesis. Future studies are required to confirm this hypothesis including validating the co-expression of Runx1, Lgr5, and Gata1 during gastrulation by multi-color ISH or immunofluorescence and characterization of extensive lineage tracing experiments of these genes by hematopoietic cell type markers.

3.4.2 *Origin of Hemogenic Endothelium in the Outflow Tract*

We show that Runx1+/Lgr5+/Gata1- cells at LS, presumably representing transient hematopoiesis, co-express various canonical CM and ExEM TFs and structural proteins. This is intriguing considering recent studies identifying a hemogenic angioblast cell lineage characterized by transient Nkx2.5 expression, a canonical cardiac TFs, post-gastrulation in the ExEM that contributes to hemogenic endothelium and endocardium, suggesting a novel role for Nkx2.5 in hemoangiogenic lineage specification and diversification (Zamir L, 2017). Our observation that Lgr5-expressing cells at LS, which is prior to Nkx2-5 expression (data not shown), appear to label to outflow tract endothelium, at the hemogenic site and embryonic age described in Zamir et al., suggest that Runx1+/Lgr5+/Gata1- cells maybe are precursors to transient hemogenic endothelium in the OFT.

3.4.3 *Reconstructing the Hematopoietic Hierarchy*

The identification of three hematopoietic populations differentially expressing Runx1, Gata1, and Lgr5 brings into question a generally accepted model of the cellular potential of Runx1 and Gata1 expressing cells, which postulates that Runx1+/Gata1+ cells contribute to developmental hematopoiesis and Runx1+/Gata1- cells to adult hematopoiesis. Our data suggest further resolution into this model, proposing that Runx1+/Gata1+/Lgr5+ cells contribute to first-wave developmental hematopoiesis, Runx1+/Gata1-/Lgr5+ cells contribute to transient definitive hematopoiesis, and Runx1+/Gata1+/Lgr5- cells contribute to adult hematopoiesis. In addition, our data suggest a potential spatial location for Runx1+/Gata1-/Lgr5+ cells in the ExEM that is overlapping with expression of cardiac structural proteins such as Actc1 and Tnnt2. This hypothesis would suggest an updated model of hematopoiesis including a novel Runx1+ cellular intermediate and improved molecular signature for Runx1+ intermediates cells with differing cellular potentials.

3.5 Acknowledgements

This chapter is an original document written by Jonathan D. Grinstein with oversight from Neil Chi and Sylvia Evans. The bioinformatics data for this chapter were generated by Jonathan D. Grinstein, and the experimental data were generated by Paola Cattaneo, Nuno Camboa, and Josh Bloomekatz. Jonathan D. Grinstein was supported by a pre-doctoral fellowship from the California Institute of Regenerative Medicine (CIRM). Neil Chi is an Associate Professor of Medicine at UCSD and is supported by grants from the National Institutes of Health (NIH) and the American Heart Association.

Chapters 3 is currently being prepared for submission for publication of the material. Jonathan D. Grinstein, Paola Cattaneo, Josh Bloomekatz, Neil Tedeschi, Elie Farah, Yan Song, Leen Jamal-Schafer, Nuno Camboa, Gene Yeo, Sylvia M. Evans, Neil C. Chi. The dissertation/thesis author, Jonathan D. Grinstein, was the primary investigator and author of this material.

Methods

No statistical methods were used to predetermine sample size.

Transgenic Mice and Timed Matings

All animal care was in compliance with the *Guide for the Care and Use of Laboratory Animals* published by the US National Institutes of Health, as well as institutional guidelines at the University of California, San Diego. All transgenic lines used were kept on an outbred background (Black-Swiss, Charles River laboratories). Mice were maintained in disposable plastic cages with filtered air intake ports (Innovive Inc.) on a 12-hr. light cycle and fed Teklad LM-485 irradiated diet (Harlan Laboratories, catalog number 7912). Embryos were staged according to the embryonic day (E) on which dissection took place, with noon of the vaginal plug day being considered as E0.5. Embryos were staged according to the morphological criteria of Downs and Davies, and classified as mid-streak, late-streak or three-somites.

Mesp1:Cre/+ mice were obtained from Dr. Yumiko Saga. Timed matings were set up between heterozygous Mesp1:Cre/+ and homozygous R26R:tdT/R26R:tdT mice.

Lgr5:EGFP:ROSA:creERT2/+ mice were obtained from The Jackson Laboratory (B6.129P2-Lgr5$^{tm1(cre/ERT2)Cle}$/J) . Timed matings were set up between heterozygous Lgr5:EGFP:ROSA:creERT2/+ and homozygous R26R:tdT/R26R:tdT mice. Pregnant females were injected with tamoxifen at E6.75. Embryos were staged according to the morphological criteria of Downs and Davies, and classified as E8.25 or E10.75.

Fluorescence Activated Cell Sorting (FACS)

Suspensions of cells from individual embryos, both tdT+ and tdT-, were prepared independently by incubating with 0.25% Trypsin-EDTA dissociation reagent (Life Technologies) at 37°C for 20 min and quenching with heat-inactivated serum. All cells were stained with DAPI for viability. Cells from tdT- mice were used for gating controls. Live, tdT+ cells were sorted from individual mice at each stage into separate wells of a 96-well plate coated with heat-inactivated serum. Cell sorting was performed with a BD Influx cell sorter. Cells were visually inspected to confirm the presence of tdT+ events. 96-well plates were centrifuged at 200 g for 5 min, the supernatant was then discarded and cells were re-suspended in Resuspension Buffer (Fluidigm) with cell viability dye Calcein AM (ThermoFisher). Cells were visually inspected again to confirm the presence of tdT+ events.

Cell Capture for Single Cell RNA Sequencing

Upon priming a 96-chamber medium mRNAseq IFC (Fluidigm), cells from one well were loaded for single cell capture. Following cell capture, each capture site was characterized for the presence of a single, live (green), and *Mesp1*-labelled (red) cell by fluorescence microscopy. Single cell cDNA libraries were generated with Smart-seq2 per the manufacturer's instructions. Single cell cDNA libraries were harvested and, prior to dilution, were analyzed for cDNA concentration using High Sensitivity dsDNA Qubit (Invitrogen).

Single Cell RNA Sequencing Library Preparation and Mapping of reads.

Libraries were prepared using the Illumina Nextera XT DNA preparation kit. Random libraries were assayed for DNA quality and fragment size by Tape Station. Pooled libraries of 96-cells were sequenced on the Illumina Hi-Seq 4000. Reads were mapped simultaneously to the *Mus musculus* genome (Ensembl version 38.77).

Computational Quality Control of Single Cell Transcriptomes and Genes for Analysis:

To assess data quality, the following metrics were used: (1) filter out low-gene cells (often dead or dying cells) (< 1.8e3 reads total), (2) remove genes that don't have many reads (<10 reads total), and (3) remove genes that are not seen in enough cells (<5 cells).

Single cell error model generation, Normalization of read counts, clustering, and identification of highly variable genes and gene sets via PAGODA.

We follow the approaches described in Kharchenko et al. 2014 and Fan et al., 2016.

t-Distributed Stochastic Neighbor Embedding

To display the relative distances between cells in a lower-dimension representation of gene expression space, we employed tSNE. We use it here to visualize our data independently from how we generated the cell clusters. We use the Barnes-Hut implementation of tSNEs available in the R package *Rtsne*. We chose as to input the entire expression matrix and used all default parameters (initial dimensions, 50;

perplexity, 10; iterations, 1000). Three dimensional tSNEs were generated using the first three tSNE dimensions and plotted with the R package *plot3D* using the function 'scatter3D'.

Cell Lineages and Developmental Distance

We used a recently developed cell lineage inference algorithm, Slingshot (Version 0.0.0.9005, available as an open-source R package slingshot at https://github.com/kstreet13/slingshot, to identify DLTs and bifurcations and to order cells along trajectories.

Slingshot takes as input a matrix of reduced dimension normalized expression measures (e.g., tSNE) and cell clustering assignments. It infers DLTs and branch points by connecting the cluster medoids using a MST and identifying the starting cluster or root node. Lineages are defined by ordered sets of clusters beginning with the root node and terminating in the most distal cluster(s) with only one connection. Next, principal curves are fit to the subsets of cells making up each lineage, providing a smooth, nonlinear summary of each trajectory. Individual cells are then orthogonally projected to each curve and thereby ordered in a space reflecting developmental distance. The ordering provided by Slingshot, analogous to pseudotime, is referred to herein as developmental order.

The root and end nodes were unbiasedly inferred by slingshot. Slingshot then generated principal curves and cell developmental distances for each lineage. Slingshot was applied to the first three tSNE dimensions of the normalized expression matrix.

Differential Expression of Genes in Pseudotime

We used a generalized additive model (GAM) to regress each gene on generated pseudotime vectors in order to detect non-linear patterns in gene expression over developmental time. We use the generalized additive models with integrated smoothness estimation available in the R package *gam*. We then pick out the top genes for each lineage and visualize their expression over developmental time with a heatmap.

Code availability

All data were analyzed with standard programs and packages, as detailed above. Code is available on request.

References

Ang SL, R. J. (1993). *Development*, 139-149.

Barker N, T. S. (2013). Lgr proteins in epithelial stem cell biology. *Development*, 2484-94.

Barnes JD, C. J. (1994). Embryonic expression of Lim-1, the mouse homolog of Xenopus Xlim-1, suggests a role in lateral mesoderm differentiation and neurogenesis. *Dev Biol.* , 168-78.

Becker D, J. Z. (1996). Conserved regulatory element involved in the early onset of Hoxb6 gene expression. *Dev Dyn.*, 73-81.

Bondue A, L. G. (2008). Mesp1 acts as a master regulator of multipotent cardiovascular progenitor specification. . *Cell Stem Cell*, 69-84.

Brachtendorf G, K. A. (2001). Early expression of endomucin on endothelium of the mouse embryo and on putative hematopoietic clusters in the dorsal aorta. *Dev Dyn.*, 410-9.

Cai Z, d. B. (2000). Haploinsufficiency of AML1 affects the temporal and spatial generation of hematopoietic stem cells in the mouse embryo. *Immunity*, 423-31.

Carson CT, K. E. (2000). Tbx12, a novel T-box gene, is expressed during early stages of heart and retinal development. *Mech Dev.*, 137-40.

Catron KM, W. H.-S. (1996). Comparison of MSX-1 and MSX-2 suggests a molecular basis for functional redundancy. *Mech Dev.*, 185-99.

Chan SS, S. X. (2013). Mesp1 patterns mesoderm into cardiac, hematopoietic, or skeletal myogenic progenitors in a context-dependent manner. *Cell Stem Cell.*, 587-601.

Chan SS., H. H. (2016). Development of Bipotent Cardiac/Skeletal Myogenic Progenitors from MESP1+ Mesoderm. *Stem Cell Reports.*, 26-34.

Chapman DL, A. I. (1996). Tbx6, a mouse T-Box gene implicated in paraxial mesoderm formation at gastrulation. *Dev Biol.*, 534-42.

Chen MJ, L. Y.-I. (2011). Erythroid/myeloid progenitors and hematopoietic stem cells originate from distinct populations of endothelial cells. *Cell Stem Cell*, 541-52.

Chen MJ, Y. T. (2009). Runx1 is required for the endothelial to haematopoietic cell transition but not thereafter. *Nature*, 889–891.

Chiapparo G, L. X. (2016). Mesp1 controls the speed, polarity, and directionality of cardiovascular progenitor migration. *J Cell Biol.*, 463-77.

Clevers H, L. K. (2014). Stem cell signaling. An integral program for tissue renewal and regeneration: Wnt signaling and stem cell control. *Science*.

David R, B. C.-H. (2008). Nat Cell Biol. 338-45.

Davidson B, L. M. (2003). Evolutionary origins of the vertebrate heart: specification of the cardiac lineage in Ciona intestinalis. *Proc Natl Acad Sci USA*, 11469–11473.

de Lau W, P. W. (2014). The R-spondin/Lgr5/Rnf43 module: regulator of Wnt signal strength. *Genes Dev.*, 305-16.

Devine WP, W. J.-T. (2014). Early patterning and specification of cardiac progenitors in gastrulating mesoderm. *Elife*.

Diogo R, K. R. (2015). A new heart for a new head in vertebrate cardiopharyngeal evolution. *Nature*, 466-73.

Dupé V, D. M. (1997). In vivo functional analysis of the Hoxa-1 3' retinoic acid response element (3'RARE). *Development*, 399-410.

Espinoza-Lewis RA, W. D. (2014). Generation of a Cre knock-in into the Myocardin locus to mark early cardiac and smooth muscle cell lineages. *Genesis*, 879-87.

Fan J, S. N. (2016). Characterizing transcriptional heterogeneity through pathway and gene set overdispersion analysis. . *Nat Methods*, 241-4.

Ferkowicz MJ, S. M. (2003). CD41 expression defines the onset of primitive and definitive hematopoiesis in the murine embryo. *Development*, 4393-403.

Fletcher RB, D. D. (2017). Deconstructing Olfactory Stem Cell Trajectories at Single-Cell Resolution. *Cell Stem Cell*, 817-830.

Ginhoux, F. G. (2010). Fate mapping analysis reveals that adult microglia derive from primitive macrophages. *Science*, 841–845.

Glinka A, D. C. (2011). LGR4 and LGR5 are R-spondin receptors mediating Wnt/β-catenin and Wnt/PCP signalling. *EMBO Rep.*, 1055-61.

Gomez Perdiguero, E. K. (2015). Tissue-resident macrophages originate from yolk-sac-derived erythro-myeloid progenitors. *Nature*, 547–551.

Grün D, L. A. (2015). Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature*, 251-5.

Hatada, Y. a. (1994). A fate map of the epiblast of early chick embryo. . *Development*, 2879–2989 .

Herbomel, P. T. (1999). Ontogeny and behaviour of early macrophages in the zebrafish embryo. *Development*, 3735–3745.

Hoffmeyer K, R. A. (2012). Wnt/β-catenin signaling regulates telomerase in stem cells and cancer cells. *Science*, 1549-54.

Janda CY, W. D. (2012). Structural basis of Wnt recognition by Frizzled. *Science*, 59-64.

Kaestner KH, B. S. (1996). Clustered arrangement of winged helix genes fkh-6 and MFH-1: possible implications for mesoderm development. *Development*, 1751-8.

Kharchenko PV, S. L. (2014). Bayesian approach to single-cell differential expression analysis. *Nat Methods*, 740-2.

Kinder SJ, L. D. (2001). Allocation and early differentiation of cardiovascular progenitors in the mouse embryo. . *Trends Cardiovasc Med.*, 177-84.

Kinder SJ, T. T. (1999). The orderly allocation of mesodermal cells to the extraembryonic structures and the anteroposterior axis during gastrulation of the mouse embryo. *Development*, 4691-4701.

Kingsley, P. D. (2004). Yolk sac-derived primitive erythroblasts enucleate during mammalian embryogenesis. *Blood*, 19–25.

Kinzel B, P. M. (2014). Functional roles of Lgr4 and Lgr5 in embryonic gut, kidney and skin development in mice. *Dev Biol.*, 181-90.

Kitajima S, T. A. (2000). MesP1 and MesP2 are essential for the development of cardiac. *Development*, 3215-26.

Kouskoff V, L. G. (2005). Sequential development of hematopoietic and cardiac mesoderm during embryonic stem cell differentiation. *Proc Natl Acad Sci USA*, 13170–13175.

Kraus F, H. B. (2001). Cloning and expression analysis of the mouse T-box gene tbx20. *Mech Dev.* , 87-91.

Lawson KA, M. J. (1991). Clonal analysis of epiblast fate during germ layer formation in the mouse embryo. *Development*, 891–911.

Lawson, K. A. (1991). Clonal analysis of epiblast fate during germ layer formation in the mouse embryo. *Development*, 891–911.

Leblond CP, S. C. (1948). The constant renewal of the intestinal epithelium in the albino rat. . *Anat Rec.*, 357-77.

Leblond CP, W. B. (1956). Renewal of cell populations. *Physiol Rev.*, 255-76.

Lescroart F, C. S. (2014). Early lineage restriction in temporally distinct populations of Mesp1 progenitors during mammalian heart development. *Nat Cell Biol.*, 829-40.

Leussink B, B. A.-d. (1995). Expression patterns of the paired-related homeobox genes MHox/Prx1 and S8/Prx2 suggest roles in development of the heart and the forebrain. *Mech Dev.*, 51-64.

Li Y, W. J. (2013). Mesothelial cells give rise to hepatic stellate cells and myofibroblasts via mesothelial-mesenchymal transition in liver injury. *Proc Natl Acad Sci USA*, 2324–2329.

Liu D, H. X. (2014). Leucine-rich repeat-containing G-protein-coupled Receptor 5 marks short-term hematopoietic stem and progenitor cells during mouse embryonic development. *J Biol Chem.*, 23809-16.

Lua I, L. Y. (2015). Myofibroblastic conversion and regeneration of mesothelial cells in peritoneal and liver fibrosis. *Am J Pathol* , 3258–3273.

Lua I., J. D. (2014). Mesodermal mesenchymal cells give rise to myofibroblasts, but not epithelial cells, in mouse liver injury. *Hepatology*, 311–322.

Morita H, M. S. (2004). Neonatal lethality of LGR5 null mice is associated with ankyloglossia and gastrointestinal distension. *Mol Cell Biol.*, 9736-43.

Muñoz J, S. D. (2012). The Lgr5 intestinal stem cell signature: robust expression of proposed quiescent '+4' cell markers. *EMBO J.*, 3079-91.

Najdi R, P. K. (2012). A uniform human Wnt expression library reveals a shared secretory pathway and unique signaling activities. *Differentiation*, 203-13.

Nakano H, L. X. (2013). Haemogenic endocardium contributes to transient definitive haematopoiesis. *Nature Communications*.

Narumiya H, H. K. (2007). Endocardiogenesis in embryoid bodies: novel markers identified by gene expression profiling. *Biochem Biophys Res Commun.*, 896-902.

Palis, J. (2014). Primitive and definitive erythropoiesis in mammals. *Frontiers in Physiology*.

Palis, J. R. (1999). Development of erythroid and myeloid progenitors in the yolk sac and embryo proper of the mouse. *Development*, 5073–5084.

Parameswaran M, T. P. (1995). Regionalisation of cell fate and morphogenetic movement of the mesoderm during mouse gastrulation. *Dev Genet.*, 16-28.

Potts, K. S. (2014). A lineage of diploid platelet-forming cells precedes polyploid megakaryocyte formation in the mouse embryo. *Blood*, 2725–2729.

Robb L, M. L. (1998). Epicardin: A novel basic helix-loop-helix transcription factor gene expressed in epicardium, branchial arch myoblasts, and mesenchyme of developing lung, gut, kidney, and gonads. *Dev Dyn.*, 105-13.

Saga Y, H. N. (1996). MesP1: a novel basic helix–loop–helix protein expressed in the nascent mesodermal cells during mouse gastrulation. *Development*, 2769–2778.

Saga Y, M.-T. S. (1999). MesP1 is expressed in the heart precursor cells and required for the formation of a single heart tube. *Development*, 3437-47.

Sasaki H, H. B. (1993). Differential expression of multiple fork head related genes during gastrulation and axial pattern formation in the mouse embryo. *Development*, 47-59.

Schepers AG, V. R. (2011). Lgr5 intestinal stem cells have high telomerase activity and randomly segregate their chromosomes. *EMBO J*, 1104-9.

Shen MM, W. H. (1997). A differential display strategy identifies Cryptic, a novel EGF-related gene expressed in the axial and lateral mesoderm during mouse gastrulation. *Development*, 429-42.

Simrick S, S. D. (2012). Biallelic expression of Tbx1 protects the embryo from developmental defects caused by increased receptor tyrosine kinase signaling. *Dev Dyn.*, 1310-24.

Slack, J. (1999). From Egg to Embryo: Determinative Events in Early Development. Ch 1.

Smart N, H. A. (2002). A differential screen for putative targets of the bHLH transcription factor Hand1 in cardiac morphogenesis. *Mech Dev*, 65-71.

Stamos JL, C. M. (2014). Structural basis of GSK-3 inhibition by N-terminal phosphorylation and by the Wnt receptor LRP6. *Elife*.

Sturgeon CM, D. A. (2014). Wnt signaling controls the specification of definitive and primitive hematopoiesis from human pluripotent stem cells. *Nat Biotechnol.*, 554-61.

Tam PP, G. D. (2000). Early events of somitogenesis in higher vertebrates: allocation of precursor cells during gastrulation and the organization of a meristic pattern in the paraxial mesoderm. *Curr Top Dev Biol.* , 1-32.

Tam PP, P. M. (1997). The allocation of epiblast cells to the embryonic heart and other mesodermal lineages: the role of ingression and tissue movement during gastrulation. *Development*, 1631-42.

Tam PP., B. R. (1997). Mouse gastrulation: the formation of a mammalian body plan. *Mech Dev.*, 3-25.

Tam, P. (1989). Regionalization of the mouse embryonic ectoderm: allocation of prospective ectodermal tissues during gastrulation. *Development*, 55-67.

Tam, P. P. (1987). The formation of mesodermal tissues in the mouse embryo during gastrulation and early organogenesis. *Development*, 109–126.

Tamplin OJ, K. D. (2008). Microarray analysis of Foxa2 mutant mouse embryos reveals novel gene expression and inductive roles for the gastrula organizer and its derivatives. *BMC Genomics*, 511.

Tanaka Y, H. M. (2012). Early ontogenic origin of the hematopoietic stem cell lineage. . *Proc Natl Acad Sci U S A*, 4515-20.

Tanaka Y, S. V. (2014). Circulation-independent differentiation pathway from extraembryonic mesoderm toward hematopoietic stem cells via hemogenic angioblasts. . *Cell Rep.*, 31-9.

Tober J, M. M. (2016). Taking the Leap: Runx1 in the Formation of Blood from Endothelium. *Current Topics in Developmental Biology*, 113-162.

Trapnell C, C. D. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol.*, 381-6.

Trapnell, C. (2015). Defining cell types and states with single-cell genomics. *Genome Res.*, 1491-8.

Vlaeminck-Guillem V, C. S.-C. (2000). The Ets family member Erg gene is expressed in mesodermal tissues and neural crests at fundamental steps during mouse embryogenesis. *Mech Dev.*, 331-5.

Wei TC, L. H. (2011). Expression of Crip2, a LIM-domain-only protein, in the mouse cardiovascular system under physiological and pathological conditions. *Gene Expr Patterns*, 384-94.

Yamaguchi TP, D. D. (1993). flk-1, an flt-related receptor tyrosine kinase is an early marker for endothelial cell precursors. . *Development*, 489-98.

Yamaguchi, T. P. (1994). Fgfr- 1 is required for embryonic growth and mesodermal patterning during mouse gastrulation. *Genes Dev.*, 3032-3044.

Yoder MC. (2014). Inducing definitive hematopoiesis in a dish. *Nature Biotechnology*, 539–541.

Yoshida T, V. P.-K. (2008). Cell lineage in mammalian craniofacial mesenchyme. *Mech Dev*, 797–808.

Yule, G. U. (1903). Notes on the Theory of Associationof Attributes in Statistics. *Biometrika*, 121-134.

Zamir L, S. R.-R. (2017). Nkx2.5 marks angioblasts that contribute to hemogenic endothelium of the endocardium and dorsal aorta. *Elife*.

Zeng YA, N. R. (2010). Wnt proteins are self-renewal factors for mammary stem cells and promote their long-term expansion in culture. *Cell Stem Cell.*, 568-77.