Title
Deciphering transcriptional and epigenomic regulation of early cardiogenesis

Author
Krup, Alexis Leigh

Publication Date
2022

Peer reviewed|Thesis/dissertation

Deciphering transcriptional and epigenomic regulation of early cardiogenesis

by
Alexis Leigh Krup

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Biomedical Sciences

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
Jeremy Reiter
F07C889D1B164B3...

Jeremy Reiter
Chair

DocuSigned by:
Benoit Bruneau
DocuSigned by:
46F...

Benoit Bruneau

DocuSigned by:
Yin Shen
DocuSigned by:
42B...

Yin Shen

DocuSigned by:
Michael Ryan Corces
8CA3022ABDCE448...

Michael Ryan Corces

Committee Members

Dedication

This dissertation is dedicated to my parents,

who are generous in their love and assiduous in their support.

Acknowledgements

I am grateful to the many individuals who have provided support, advice, kindness, and encouragement throughout my graduate and scientific career.

Firstly, I am grateful to my advisor, Dr. Benoit Bruneau, for his mentorship and the opportunity to perform my graduate research in his lab. Generous in his support to his colleagues and trainees, he ensured I always had access to the resources and collaborations I needed to succeed in my endeavors, including the perhaps-overly-ambitious ones. His passion for studying gene regulation and cardiac development coupled with his enthusiasm for applying new technologies to complex problems creates a broad horizon for the types of questions he supports his trainees in pursuing. The resulting climate of scientific autonomy and opportunities enabled me to continuously push my thesis project through all manner of unexpected challenges, which in turn empowered me to develop clarity in my direction and desires for my scientific career. Additionally, his impressive depth of cardiovascular and developmental biology knowledge is perhaps only matched by his nuanced knowledge of his many favorite musicians' extensive discographies. I admire Benoit's ability to coordinate teams of experts and resources in large-scale collaborative studies, the pride he takes in supporting his trainees' ambitions and celebrating their achievements, and his special knack for identifying the "extra something" experiment or phrase likely needed to elevate a story or a grant.

My appreciation to my committee members, Dr. Yin Shen and Dr. Ryan Corces, who readily availed themselves for 1:1 meetings and provided mentorship in everything ranging from experimental design, analysis, and strategy. They sat patiently, posing

constructively direct questions and offering pointed comments, so especially necessary in some of those early committee meetings where I was still grasping for narrative focus in my data and my efforts. Their questions, advice, and actionable feedback helped me develop my scientific thinking, my project, and also myself as a scientist.

Special gratitude to Dr. Jeremy Reiter, whose mentorship as my committee chair has gone above and beyond all obligation through his kind encouragement, thoughtful strategy, and adeptly deployed humor in reminding me to maintain perspective. He would deftly illuminate any blind spot through a single question, and readily offer a tractable suggestion for how to address perceived issues. Before serving as my doctoral committee chair, he gave me my first full-time scientific opportunity as a research associate in his lab. He entrusted me with the responsibility and the resources to drive and contribute to research efforts initially beyond my skills. The confidence he placed in my abilities (and the abilities of all his lab members) coupled with his direct feedback, often while working side by side at the bench, reinforced and inspired me to be active in my learning and rise to the occasion with the best of my efforts every time. The experience and confidence I gained through working in his lab on such a wide range of topics catalyzed my scientific development and desire to pursue a PhD. Furthermore, Jeremy continuously inspires me for his boundless curiosity and creativity, work ethic of integrity, the breadth and depth of what he reads, and commitment to always asking the next level deeper question.

I am grateful to my undergraduate thesis advisor, Dr. Nipam Patel, who gave an enthusiastic UC Berkeley undergraduate her first chance to perform basic science research. After a few initial weeks of me mounting butterflies for the collection and doing

immunostains on *Drosophila* embryos, he agreed to let me restart a butterfly colony in the lab; getting permits and ordering the specimens that would become my thesis, teaching me how to use a confocal, and then turning me loose. The opportunity to perform evo-devo research on structural color in his lab enabled me to realize a childhood dream of researching butterfly iridescence and patterning. The experience also marked the moment I realized my creativity and penchant for “how does that happen?” questions had an important place in the research laboratory.

Thank you also to the many talented graduate students, postdocs, and professional scientific staff of both Patel Lab and Reiter Lab whose mentorship, friendship, thoughtfulness, and encouragement helped me develop my scientific confidence and led to me pursuing this doctoral degree. In particular; Crystal, Erin, Heather, and Ryan in Patel Lab; Elle, Colin, Andy K., David, and Monika in Reiter Lab.

I’d also like to thank the Bruneau Lab members I had the opportunity to work alongside during my graduate years; Sarah, Emily Bu., Kevin H., Rui, Kevin S., Swetansu, Kelly, Abby, Piyush, Martin, Patrick, Emily Br., Jingshing, Jon, Zoe, Irfan, Elphege, Kavitha, Vasu, and Austin. Sometimes there were late nights/early mornings, there was also mischief abound, IKIGM, The VK Method of temporal task organization and responsibility compartmentalization, and more experiments and analysis sessions than anyone should count. An additional note of gratitude for the friendships I’ve developed among members of this lab, which I look forward to continuing in both life and career beyond this dissertation.

During graduate school I’ve had the privilege to work with some inspiring scientific colleagues and mentors, many of whom have also become friends. In

particular: From Gladstone: Jeeves, Ariel, Vaishaali, Ana, Yvanka, Gokul, Hannah, Arun, Angelo, Byron, David, Ashley, Ivana, Kate, Casey, Joke; From UCSF: Tanvi, Trisha, Ryan, Bryan, Rebecca, Stephanie, Aydan, Michelle, Steffen, Melissa; From my cohort: Austin, Anthony, Parinaz, Katya, Johnny, Jerika, Sara, Camille, Bree, Ramiro, David, Kamir, Didi, Jeanmarie, Elma, Joe. My gratitude for all the constructive feedback and advice, good-natured hijinks, expertly-executed puns and memes, things you taught me at the bench or in the terminal, and camaraderie.

I benefited from the mentorship and administrative support of many UCSF and Gladstone professors and professional staff throughout my graduate career. Their open-door policies were sometimes taken so far as to manifest in me semi-gate-crashing their holiday parties as the tag-along with their actual, invited students (thanks, Dr. Rushika Perera). A special thank you to Demian Sainz and the BMS GMU office, for masterfully solving all problems and creating practical roadmaps so us students could put all our energy into our research and studies. Thank you to Dr. Sudha Krishnamurthy and Dr. Katy Claiborn at Gladstone for your coaching and strategic advice on fellowships, resource management, and manuscript writing. Thank you to Dr. Anita Sil and Dr. Mark Ansel for directing the BMS program and upholding standards of integrity for our training. An additional thank you to my qualifying exam committee; Dr. Licia Selleri, Dr. Alex Pollen, Dr. Jeremy Reiter, and my chair Dr. Takashi Mikawa, for kickstarting my thesis research. The intensity of preparation required for some of those 1:1 meetings with you all overprepared me to the point that the actual examination ended up being the easy part. Further gratitude to my rotation and early UCSF research advisors, Dr. Arturo Alvarez-Buylla, Dr. Michael Oldham, and Dr. Miguel Ramalho-Santos.

My eternal gratitude to the team of compassionate and thorough medical professionals who took care of me with such dignity and determination when I accidentally happened into the role of “patient” during my time at UCSF.

Thank you to the talented dancers and artists, avid outdoorsy folks, and down-for-any-adventure crew of delightful friends who have been supportive cheerleaders of my efforts in graduate training, and have kept me grounded in the aspects of my life outside the lab be they by land, sea, clinging to a rock wall up in the air, or ballet class.

Finally, my immense love and gratitude to my incredible family. To Max, for being my partner in life and adventure, and for climbing all sorts of figurative and literal mountains with me. To my younger brother, Michael, for being the Alfred to my Batman with every “endure, Master Wayne” reminder to stay the course of what I believe in, and for bringing unparalleled style to every dance floor. To Lillian, for all your words of encouragement. To my elder sister, Nicole, for being my role model and showing me what following one’s dreams can look like. To Johannes, for being the best kind of big brother(-in-law) anyone could ever want. To my late beloved grandparents, for their strict compass of integrity, and their unconditional love. To my father, Michael, for teaching me everything from falconry, fly fishing, butterfly catching, dark room photography, martial arts, and even how to ride a bike, and for encouraging me to approach challenges as opportunities for learning and evolution. To my mother, Michele, for demonstrating how to gracefully subvert systems with a principled mind towards bettering them, and for showing me how dedicated effort in the details can translate into beautiful stories.

Contributions

This dissertation encompasses my graduate research and training. The bodies of work presented in this dissertation were developed, led, executed, and written by me. The hypotheses, interpretations, and conclusions were developed through extensive consult with published literature, careful experimentation and analysis, thoughtful interpretations of analyses, and valuable discussions with colleagues and mentors.

Chapter 2 represents yet un-published work done with contributions from Sanjeev Ranade, Kavitha Rao, and Sarah Winchester. Aspects of this work are included in the publication that comprises the majority of the work presented in Chapter 4.

Chapter 3 represents an in-progress work performed in collaboration with Marco Osterwalder, Fabrice Darbellay, Diane Dickel, Axel Visel, and Len Pennacchio at Lawrence Berkeley National Laboratory, and with contributions from Junli Zhang at the Gladstone Institutes.

Chapter 4 is adapted with adjustments from portions of a submitted manuscript: A *Mesp1*-dependent developmental breakpoint in transcriptional and epigenomic specification of early cardiac precursors. Krup AL, Winchester SAB, Ranade SS, Agrawal A, Devine WP, Sinha T, Choudhary K, Dominguez MH, Thomas R, Black BL, Srivastava D, Bruneau BG. (*Under review*, August 2022).

Deciphering transcriptional and epigenomic regulation of early cardiogenesis

Alexis Leigh Krup

Abstract

Transcriptional networks governing early cardiac precursor cell (CPC) specification are incompletely understood due in part to the difficulty of distinguishing CPCs from their mesoderm germ layer of origin in early gastrulation. Cardiogenesis in the gastrulating embryo begins when mesoderm progenitor cells emerge from the primitive streak and migrate anterior-laterally to coalesce at the anterior midline. Errors during CPC specification and patterning can cause devastating Congenital Heart Defects (CHDs). Occurring in 1-2% of live births, CHDs often require surgical interventions and can result in secondary heart disease. The genetic etiology of CHDs indicates that genes encoding transcription factors (TFs) are overrepresented as causative and are predominantly haploinsufficient, indicating that fine dysregulation of gene expression is a driving mechanism for disease. Thus, understanding the transcriptional regulatory networks governing early cardiac specification is paramount for understanding CHDs and necessary to develop novel therapeutic strategies.

Our comprehension of transcriptional regulation at the initiation of cardiogenesis is hindered in part by the paucity of molecular tools capable of distinguishing the emerging cardiac lineage from the surrounding mesoderm. Prior studies leveraged lineage tracing from the basic-helix-loop-helix (bHLH) TF *Mesp1*, however as this lineage contributes to other mesodermal derivatives beyond the heart the method is insufficient for isolation of early CPCs. To overcome this challenge and investigate the cardiac lineage distinctly

from the surrounding mesoderm, we leveraged a pan-cardiac enhancer transgene reporter, *Smarcd3*-F6, that restrictively marks emerging, early CPC populations within the mesoderm. We utilized bioinformatic detection of fluorescent reporter transgenes tracking both the *Mesp1* lineage and *Smarcd3*-F6 expression in whole embryo single cell transcriptomic data to interrogate the heterogeneity of CPC transcriptional profiles in an *in vivo* mouse gastrulation time course. The dataset we generated towards this goal represents a valuable resource for investigations of the early cardiac mesoderm and for broader questions of cell fate allocation from germ layers during gastrulation.

We further leveraged the *Smarcd3*-F6 enhancer sequence as an experimental discovery platform for the identification of regulatory network logic during early cardiogenesis. We identified specific GATA and T-box motif sites necessary for a minimal *Smarcd3*-F6 sub-region's enhancer activity. This *in vivo* enhancer study provides a framework for functional characterization of transcriptional regulatory networks during development.

Lastly, we utilized single cell transcriptomic and chromatin accessibility sequencing to define the resilience and vulnerability of cardiac specification in embryos deficient for *Mesp1*, the early-expressed and often-posited 'cardiac master regulator'. Our results distinguish *Mesp1*-independent and dependent processes in early cardiogenesis, showing that *Mesp1* deficient CPCs progress through cardiogenesis until lateral plate mesoderm stages, at which point their disrupted regulatory landscape prohibits maturation further into patterned cardiac progenitor and cardiomyocyte fates. Collectively, these results illustrate the complex transcriptional and epigenomic

interdependence of regulatory networks during lineage specification and further advance our fundamental understanding of the processes governing cardiac specification *in vivo* at single cell resolution.

The investigative frameworks and the interpretations of findings described in this dissertation illuminate generalizable principles for the regulatory logic guiding the allocation and subsequent differentiation of precursor cells towards distinct, functional cell types during gastrulation.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

antEndoderm = anterior endoderm

antExEMeso = anterior extraembryonic mesoderm

antME= anterior mesendoderm

antPS = anterior primitive streak

CardiacMeso = cardiac mesoderm

CFMeso = craniofacial mesoderm

CM = cardiomyocyte

CP = cardiac progenitor

CPC = cardiac precursor cell

CRE = *cis*-regulatory element

DE = definitive endoderm

EarlyMeso = early mesoderm

EomesME = *Eomes*+ mesendoderm

EomesPSMeso = *Eomes*+ primitive streak mesoderm

Epi = epiblast

ExE = extraembryonic

ExEM-LPM = extraembryonic mesoderm / lateral plate mesoderm

ExEMeso = extraembryonic mesoderm

HPCs = hematopoietic precursors

IntMeso = intermediate mesoderm

LPM – ExEMeso = lateral plate mesoderm / extraembryonic mesoderm

LPM = lateral plate mesoderm

LPM-ExEMeso = lateral plate mesoderm / extraembryonic mesoderm

LSMeso = late streak mesoderm

MesEndo = mesendoderm

Meso = mesoderm

Mesp1M = *Mesp1*+ mesoderm

Mesp1ME = *Mesp1*+ mesendoderm

PGCs = primordial germ cells

postLPM = posterior lateral plate mesoderm

postMeso = posterior mesoderm

postPrxM = posterior paraxial mesoderm

preCardiacMeso = precardiac mesoderm

PrSoM = presomitic mesoderm

PrSoM-NPS = presomitic mesoderm / primitive streak bordering node

PrxM = paraxial mesoderm

PS = primitive streak

PSMeso = primitive streak mesoderm

RA signaling cells = retinoic acid signaling cells

Reichert's = Reichert's membrane

scATAC-seq = single cell Assay for Transposase-Accessible Chromatin sequencing

scRNA-seq = single cell RNA sequencing

TF = transcription factor

Chapter 1: Introduction

A single cell containing a genome is the opening act for all of vertebrate development to follow. Through the course of embryonic development, that single cell will divide and replicate its genome multiple times to create a mass of identical cells with identical genes expressed from their genomes. In mammalian embryos, this cellular uniformity is disrupted upon implantation and initiation of gastrulation, wherein cells are distributed into three embryonic germ layers. These germ layers are established through a combination of symmetry breaking events, subsequent reorganization of genome landscapes coupled with induction of germ-layer specific gene expression profiles, and morphological changes redistributing germ layer cells into stereotyped positions in relation to one another within the once uniform cell mass. Through the course of development, the cell lineages established from these three embryonic germ layers - the ectoderm, the mesoderm, and the endoderm - will differentiate and diversify further to build and pattern the many organs within vertebrate organisms.

Cardiac development and congenital heart defects

The heart is derived from the mesoderm germ layer during early gastrulation and represents the first organ lineage to form within the embryo proper. Mesoderm progenitors emerge from the primitive streak, migrate anterior-laterally to coalesce at the cardiac crescent, and subsequently undergo dramatic morphogenic events to create a multichambered heart (Saga et al., 1999). Cardiac precursor cells (CPCs) are prefigured towards distinct fate potentials in the heart at the time of gastrulation (Devine et al., 2014; Tyser et al., 2021; Wu, Chien & Mummery, 2008; Ivanovitch et al., 2021; Lescroart et al., 2014), indicating that very early regulatory processes are responsible

for instructing CPCs to heterogenous fates for ventricles, atria, greater arteries, and the other sub-anatomies within the developed heart. Moreover, the apparent lack of robust intrinsic regenerative properties in mammalian hearts (Doppler et al., 2017) indicates that cardiac cells are terminally differentiated. Consequently, and somewhat poetically, the post-mitotic nature of cardiac cells signifies that the cells within our hearts at birth are the same cells our hearts will use throughout the entirety of our lives. Thus, the early allocation of CPCs from the mesoderm primary germ layer coupled with the terminally differentiated status of the many heterogeneous derived cell types qualifies cardiac development as a unique biological paradigm for investigations of transcriptional regulation of cell fates. In this manner, interpretations of the transcriptional mechanisms regulating the cardiac lineage serve to enhance our knowledge of this vital organ's development, and also potentially elucidate generalizable principles for the establishment, allocation, and differentiation of precursor cell types in other organ lineages.

Furthermore, insights gained through investigations of cardiogenesis are highly clinically-relevant to our understanding of congenital heart defects (CHDs). Any errors during the specification or patterning of CPCs can manifest in the heart malformations underlying CHDs, with patients often requiring surgical intervention in the treatment of associated pathophysiology (Houyel & Meilhac, 2021). Furthermore, CHDs affect 1-2% of live births and are associated with increased susceptibly to secondary heart disease (Bruneau, 2008; Nees & Chung, 2019), underscoring significant prognostic burden to both medical systems and individual patients. The genetic etiology of CHDs indicates that genes encoding transcription factors (TFs) are overrepresented as causative and

are predominantly haploinsufficient (Nees & Chung, 2019), indicating that fine dysregulation of gene expression is a driving mechanism for disease.

Deeper knowledge of transcriptional regulation during cardiac development is broadly informative of principles for cell lineage specification during gastrulation and integral to clinical efforts for mitigation and correction of CHDs. Accordingly, the studies within this dissertation chiefly aim to decipher the regulation of CPC specification at the levels of the transcriptome and the epigenome.

Regulatory networks during early cardiogenesis

A gene regulatory network represents the collective interactions amongst molecular regulators, such as TFs, chromatin remodelers, RNA, DNA, and signaling molecules. These molecular regulators govern gene expression patterns imparting functional cellular outcomes. Molecular regulators can be activating, repressing, or exhibit different activities based on interactions with specific co-factors. Similarly, the same regulators might exhibit temporal or cell-type specific roles during development.

During gastrulation, *Pou5f1* (*Oct3/4*) is among the first TFs to be expressed in the embryo proper and exhibits regulatory roles in both pluripotency and embryonic germ layer cell fate differentiation (Zeineddine et al., 2006; Li et al., 2013). Pou5f1 interactions with Sox2 and Nanog serve gatekeeping roles in pluripotency, while Pou5f1 and canonical Wnt signaling regulate mesoderm germ layer activation via synergistic activity on the *Tcf/Lef-Oct4* element of the *Mesp1* promoter (Li et al., 2013).

Eomes and Brachyury (T) have roles within mesendoderm regulatory networks to specify the endoderm and mesoderm germ layers from the pluripotent mouse epiblast while repressing neurectoderm programs (Tosic et al., 2019; Probst et al., 2020; Costello et al., 2011). Together, Eomes and T also activate *Mesp1*, *Gata4/6*, *Tbx5, Tgfb1*, and structural myocyte genes such as *Myl7* within the emerging cardiac mesoderm (Tosic et al., 2019).

Mesp1 is a basic-helix-loop-helix (bHLH) TF expressed in the earliest mesoderm progenitors emerging from the primitive streak, including emerging CPCs (Saga et al., 1996, 1999). *Mesp1*+ CPCs subsequently migrate anterior-laterally to form the cardiac crescent. While Mesp1 is often positioned atop a cardiac regulatory hierarchy as a master transcriptional regulator, disruption of *Mesp1* variably affects mesoderm migration and cardiac specification (Ajima et al., 2021; Chiapparo et al., 2016; Bondue et al., 2008; Bondue & Blanpain, 2010; Lescroart et al., 2018; Lin et al., 2022; Lindsley et al., 2008), suggesting that these cardiogenic processes are likely regulated by networks involving but not solely controlled by Mesp1.

Gata/Mef/Tbx/Hand/Nk2 factors are often ascribed as the ‘core regulatory network’ for the induction of muscle genes in the four chambered heart (Olson, 2006), however other works show mesendoderm TFs are also capable of activating muscle genes (Tosic et al., 2019). Gata/Mef/Tbx/Hand/Nk2 factors regulate themselves, each other, and downstream growth and patterning genes, as well as the signaling molecules necessary for organogenesis of the heart following CPC specification (Kelly, Buckingham & Moorman, 2014; Miquerol & Kelly, 2013; Harvey, 2002).

Gata/Mef/Tbx/Hand/Nk2 factors are induced by upstream activators such as Isl1, Nkx2-5, Gata4, and Fox factors which themselves are induced by various upstream signaling molecules (Olson, 2006). Notably, Gata factors have multiple regulatory roles within both the mesoderm and endoderm, and coordinate their different functions through participation in multiple networks involving distinct regulatory co-factors and cell-type specific enhancer occupancy (Molkentin et al., 2000; Haworth et al., 2008; Song et al., 2022; Searcy et al., 1998; Rojas et al., 2009; Pikkarainen et al., 2004; Rojas et al., 2005; Reiter et al., 1999; Lou et al., 2011; Heslop et al., 2021; Luna-Zurita et al., 2016).

Numerous signaling molecules including Bmp4, Wnt, Fgf, Shh, Notch, Tgfβ and Nodal exert regulatory effects during specification and patterning of CPCs (Ladd, Yatskievych & Antin, 1998; Schultheiss, Burch & Lassar, 1997; Pater et al., 2012; Guzzetta et al., 2020; Schans, Smits & Blankesteijn, 2008; Cohen, Tian & Morrisey, 2008; Mandal et al., 2017; Costello et al., 2011; Takeuchi et al., 2007; Rojas et al., 2005). Gradients of these signaling molecules create concentration-dependent responses within embryonic domains, and interact with regulatory networks to produce spatially restricted patterns of gene expression (Cotterell & Sharpe, 2010).

Chromatin accessibility at *cis*-regulatory elements (CREs) such as enhancers and promoters further refine regulation of dynamic gene expression patterns within the developing heart. Mutations in chromatin remodelers and an enrichment of risk-associated genetic variants for heart disease within annotated candidate CREs underscore the functional importance of these non-coding factors within regulatory networks (Smemo et al., 2012; Zaidi & Brueckner, 2017; Hocker et al., 2021). Candidate

CREs and enhancers tend to be highly restricted to specific cell-types, thus contributing to the facilitation of cell-type specific activities for other regulatory factors such as TFs that are expressed broadly within many cell types of the gastrulating embryo (Hocker et al., 2021). Thus, chromatin accessibility remodeling contributes to regulation of dynamic developmental and pathophysiological gene expression changes during cardiogenesis and disease (Han et al., 2011).

The coordinated interactions amongst regulatory factors creates complex networks of numerous interdependent mechanisms governing specification and differentiation of the cardiac lineage during gastrulation.

Evolutionary perspectives for core principles of transcriptional regulation

Evolutionarily conserved gene expression patterns are observed during early cardiogenesis. Under the developmental hourglass model for conserved ontogenesis, conserved gene expression patterns within early- to mid-gastrulation embryos across various metazoan species is called the phylotypic stage (Kalinka et al., 2010; Domazet-Lošo & Tautz, 2010; Yuan et al., 2018). Many early cellular specification events appear highly conserved across the heart organs of various metazoan organisms, with similar regulatory networks governed by orthologous TFs (Yuan, Scott & Wilson, 2021).

Certainly not every regulatory factor performs identically in all species. Cross-species comparisons reveal that certain TFs are functional homologs to each other, and likewise orthologous cardiac TFs may perform different functional roles in different species. Notable examples include the functional homologs zebrafish *gata5* and mammalian

Gata4, and by contrast the divergent phenotypes observed with loss of *Drosophila tinman* and mouse ortholog *Nkx2-5* (Kathiriya, Nora & Bruneau, 2015; Molkentin et al., 2000; Song et al., 2022; Reiter et al., 1999; Sam et al., 2020; Bodmer, 1995; Ranganayakulu et al., 1998).

Conversely, sequence conservation of distal regulatory elements during the phylotypic periods of embryonic development and tissue lineage specification generally indicates conserved players operating within gene regulatory networks governing essential cardiogenic roles, such as conserved sequence motifs across species indicating regulatory networks comprised of similar factors (Nord et al., 2013; Bogdanović et al., 2012, 2016; Levine & Davidson, 2005). We can leverage these principles in the identification of regulatory activity within the epigenomic landscapes of emerging CPCs, as measured by transposase- or DNase1-accessible chromatin, chromatin immunoprecipitation, or enhancer activity transgene reporter assays.

Focus of dissertation

The work described in this dissertation investigates early development of the cardiac lineage in mouse embryos, with a focus on the role of transcriptional and epigenomic regulation during CPC specification.

In Chapter 2, I leverage single cell RNA sequencing (scRNA-seq) with computational detection of fluorescent reporter cell lineage transgenes to describe the emerging heterogeneity of the cardiac mesoderm during gastrulation in mouse embryos. Towards this goal, I generated an E6.0 – E7.75 *in vivo* time course for gene expression in

singularized cells from whole embryos, a rich resource dataset for studies of cardiac mesoderm specification and transcriptional events of early gastrulation broadly. Furthermore, I adapted a method for bioinformatic cell sorting to computationally label fluorescent reporter gene expression within cells in order to co-opt classical developmental biology transgenic reporter expression within high dimensional scRNA-seq data. This innovation enables tracking of cellular lineage relationships within high resolution, a-spatial single cell gene expression data.

In Chapter 3, I describe the presence of two distinct GATA and TBOX sequence regions interacting with candidate key regulatory factors in CPCs at the initiation of early cardiac specification. Using the sequence of the restricted, pan-cardiac *Smarcd3*-F6 enhancer as an experimental platform for discovery, I characterize discrete regions of high evolutionary conservation and neighborhoods of TF binding motifs within the enhancer sequence. Through reporter-mediated dissection of the *Smarcd3*-F6 enhancer's sequence activity *in vivo*, I discover a 217 bp minimally sufficient enhancer region that recapitulates the full enhancer's pan-cardiac activity in emerging CPCs. I further demonstrate that the minimized enhancer's activity depends, at least in part, on regulatory factors binding two separate GATA- and TBOX-containing motif regions in emerging CPCs during gastrulation.

In Chapter 4, I question the concept of a master transcriptional regulator for cardiogenesis. I found that embryos deficient in *Mesp1*, a TF that exhibits variable roles in migration and cardiac induction and is expressed early in cells of the developing cardiac lineage, maintain aspects of cardiogenesis despite disrupted mesodermal

migration and embryonic morphology. *Mesp1* deficient CPCs progress through initial cardiac specification up until lateral plate mesoderm cell stages, at which point mutant CPCs fail to activate the transcriptional programs necessary for patterning of cardiac progenitors and functional maturation of cardiomyocytes. Using integrated analyses of complementary single cell Assay for Transposase-Accessible-Chromatin with sequencing (scATAC-seq) and scRNA-seq *in vivo* datasets, I highlight the role of mesendoderm transcriptional programs in initiation of cardiac identity within the early mesoderm, independently of Mesp1 activity. I illustrate that mesendoderm TFs like Eomes likely activate early cardiac TFs and structural myocyte genes, and promote induction of active chromatin states near key cardiac genes in emerging CPCs. I show that de-repression of these same TFs and mesendoderm programs is likely also implicated in ectopic retention of early CPC transcriptional profiles in more mature *Mesp1*-deficient CPCs. I describe how the dysregulated epigenomic landscape of the *Mesp1* deficient lateral plate mesoderm CPCs coupled with their migratory defects leads to an ultimate halt in mutant cardiac lineage progression towards *Nkx2-5*+ cardiac progenitors. These results suggest that while the cardiac lineage is initiated in absence of *Mesp1*, lineage maturation and differentiation ultimately halts due to a *Mesp1*-dependent regulatory barrier for the transition from lateral plate mesoderm to cardiac progenitors. Together, these findings underscore the complex interdependence of transcriptional and epigenomic regulatory processes in the gastrulating embryo during cardiac lineage specification.

References

Ajima, R., Sakakibara, Y., Sakurai-Yamatani, N., Muraoka, M. & Saga, Y. (2021) Formal proof of the requirement of MESP1 and MESP2 in mesoderm specification and their transcriptional control via specific enhancers in mice. *Development*. 148 (20). doi:10.1242/dev.194613.

Bodmer, R. (1995) Heart development in Drosophila and its relationship to vertebrates. *Trends in Cardiovascular Medicine*. 5 (1), 21–28. doi:10.1016/1050-1738(94)00032-q.

Bogdanović, O., Fernandez-Miñán, A., Tena, J.J., Calle-Mustienes, E. de la, Hidalgo, C., Kruysbergen, I. van, Heeringen, S.J. van, Veenstra, G.J.C. & Gómez-Skarmeta, J.L. (2012) Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis. *Genome Research*. 22 (10), 2043–2053. doi:10.1101/gr.134833.111.

Bogdanović, O., Smits, A.H., Mustienes, E. de la C., Tena, J.J., Ford, E., et al. (2016) Active DNA demethylation at enhancers during the vertebrate phylotypic period. *Nature Genetics*. 48 (4), 417–426. doi:10.1038/ng.3522.

Bondue, A. & Blanpain, C. (2010) Mesp1. *Circulation Research*. 107 (12), 1414–1427. doi:10.1161/circresaha.110.227058.

Bondue, A., Lapouge, G., Paulissen, C., Semeraro, C., Iacovino, M., Kyba, M. & Blanpain, C. (2008) Mesp1 Acts as a Master Regulator of Multipotent

Cardiovascular Progenitor Specification. *Cell Stem Cell*. 3 (1), 69–84. doi:10.1016/j.stem.2008.06.009.

Bruneau, B.G. (2008) The developmental genetics of congenital heart disease. *Nature*. 451 (7181), 943–948. doi:10.1038/nature06801.

Chiapparo, G., Lin, X., Lescroart, F., Chabab, S., Paulissen, C., Pitisci, L., Bondue, A. & Blanpain, C. (2016) Mesp1 controls the speed, polarity, and directionality of cardiovascular progenitor migration. *The Journal of Cell Biology*. 213 (4), 463–477. doi:10.1083/jcb.201505082.

Cohen, E.D., Tian, Y. & Morrisey, E.E. (2008) Wnt signaling: an essential regulator of cardiovascular differentiation,morphogenesis and progenitor self-renewal. *Development*. 135 (5), 789–798. doi:10.1242/dev.016865.

Costello, I., Pimeisl, I.-M., Dräger, S., Bikoff, E.K., Robertson, E.J. & Arnold, S.J. (2011) The T-box transcription factor Eomesodermin acts upstream of Mesp1 to specify cardiac mesoderm during mouse gastrulation. *Nature Cell Biology*. 13 (9), 1084–1091. doi:10.1038/ncb2304.

Cotterell, J. & Sharpe, J. (2010) An atlas of gene regulatory networks reveals multiple three-gene mechanisms for interpreting morphogen gradients. *Molecular Systems Biology*. 6 (1), 425–425. doi:10.1038/msb.2010.74.

Devine, W.P., Wythe, J.D., George, M., Koshiba-Takeuchi, K. & Bruneau, B.G. (2014) Early patterning and specification of cardiac progenitors in gastrulating mesoderm. *eLife*. 3, e03848. doi:10.7554/elife.03848.

Domazet-Lošo, T. & Tautz, D. (2010) A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature*. 468 (7325), 815–818. doi:10.1038/nature09632.

Doppler, S.A., Deutsch, M.-A., Serpooshan, V., Li, G., Dzilic, E., Lange, R., Krane, M. & Wu, S.M. (2017) Mammalian Heart Regeneration. *Circulation Research*. 120 (4), 630–632. doi:10.1161/circresaha.116.310051.

Guzzetta, A., Koska, M., Rowton, M., Sullivan, K.R., Jacobs-Li, J., Kweon, J., Hidalgo, H., Eckart, H., Hoffmann, A.D., Back, R., Lozano, S., Moon, A.M., Basu, A., Bressan, M., Pott, S. & Moskowitz, I.P. (2020) Hedgehog–FGF signaling axis patterns anterior mesoderm during gastrulation. *Proceedings of the National Academy of Sciences*. 117 (27), 15712–15723. doi:10.1073/pnas.1914167117.

Han, P., Hang, C.T., Yang, J. & Chang, C.-P. (2011) Chromatin Remodeling in Cardiovascular Development and Physiology. *Circulation Research*. 108 (3), 378–396. doi:10.1161/circresaha.110.224287.

Harvey, R.P. (2002) Patterning the vertebrate heart. *Nature Reviews Genetics*. 3 (7), 544–556. doi:10.1038/nrg843.

Haworth, K.E., Kotecha, S., Mohun, T.J. & Latinkic, B.V. (2008) GATA4 and GATA5 are essential for heart and liver development in Xenopusembryos. *BMC Developmental Biology*. 8 (1), 74. doi:10.1186/1471-213x-8-74.

Heslop, J.A., Pournasr, B., Liu, J.-T. & Duncan, S.A. (2021) GATA6 defines endoderm fate by controlling chromatin accessibility during differentiation of human-induced pluripotent stem cells. *Cell Reports*. 35 (7), 109145. doi:10.1016/j.celrep.2021.109145.

Hocker, J.D., Poirion, O.B., Zhu, F., Buchanan, J., Zhang, K., Chiou, J., Wang, T.-M., Zhang, Q., Hou, X., Li, Y.E., Zhang, Y., Farah, E.N., Wang, A., McCulloch, A.D., Gaulton, K.J., Ren, B., Chi, N.C. & Preissl, S. (2021) Cardiac cell type–specific gene regulatory programs and disease risk association. *Science Advances*. 7 (20), eabf1444. doi:10.1126/sciadv.abf1444.

Houyel, L. & Meilhac, S.M. (2021) Heart Development and Congenital Structural Heart Defects. *Annual Review of Genomics and Human Genetics*. 22 (1), 1–28. doi:10.1146/annurev-genom-083118-015012.

Ivanovitch, K., Soro-Barrio, P., Chakravarty, P., Jones, R.A., Bell, D.M., Gharavy, S.N.M., Stamataki, D., Delile, J., Smith, J.C. & Briscoe, J. (2021) Ventricular, atrial, and outflow tract heart progenitors arise from spatially and molecularly distinct regions of the primitive streak. *PLoS Biology*. 19 (5), e3001200. doi:10.1371/journal.pbio.3001200.

Kalinka, A.T., Varga, K.M., Gerrard, D.T., Preibisch, S., Corcoran, D.L., Jarrells, J., Ohler, U., Bergman, C.M. & Tomancak, P. (2010) Gene expression divergence recapitulates the developmental hourglass model. *Nature*. 468 (7325), 811–814. doi:10.1038/nature09634.

Kathiriya, I.S., Nora, E.P. & Bruneau, B.G. (2015) Investigating the Transcriptional Control of Cardiovascular Development. *Circulation Research*. 116 (4), 700–714. doi:10.1161/circresaha.116.302832.

Kelly, R.G., Buckingham, M.E. & Moorman, A.F. (2014) Heart Fields and Cardiac Morphogenesis. *Cold Spring Harbor Perspectives in Medicine*. 4 (10), a015750. doi:10.1101/cshperspect.a015750.

Ladd, A.N., Yatskievych, T.A. & Antin, P.B. (1998) Regulation of Avian Cardiac Myogenesis by Activin/TGFβ and Bone Morphogenetic Proteins. *Developmental Biology*. 204 (2), 407–419. doi:10.1006/dbio.1998.9094.

Lescroart, F., Chabab, S., Lin, X., Rulands, S., Paulissen, C., Rodolosse, A., Auer, H., Achouri, Y., Dubois, C., Bondue, A., Simons, B.D. & Blanpain, C. (2014) Early lineage restriction in temporally distinct populations of Mesp1 progenitors during mammalian heart development. *Nature Cell Biology*. 16 (9), 829–840. doi:10.1038/ncb3024.

Lescroart, F., Wang, X., Lin, X., Swedlund, B., Gargouri, S., Sànchez-Dànes, A., Moignard, V., Dubois, C., Paulissen, C., Kinston, S., Göttgens, B. & Blanpain, C.

(2018) Defining the earliest step of cardiovascular lineage segregation by single-cell RNA-seq. *Science*. 359 (6380), eaao4174. doi:10.1126/science.aao4174.

Levine, M. & Davidson, E.H. (2005) Gene regulatory networks for development. *Proceedings of the National Academy of Sciences*. 102 (14), 4936–4942. doi:10.1073/pnas.0408031102.

Li, Y., Yu, W., Cooney, A.J., Schwartz, R.J. & Liu, Y. (2013) Brief Report: Oct4 and Canonical Wnt Signaling Regulate the Cardiac Lineage Factor Mesp1 Through a Tcf/Lef-Oct4 Composite Element. *STEM CELLS*. 31 (6), 1213–1217. doi:10.1002/stem.1362.

Lin, X., Swedlund, B., Ton, M.-L.N., Ghazanfar, S., Guibentif, C., Paulissen, C., Baudelet, E., Plaindoux, E., Achouri, Y., Calonne, E., Dubois, C., Mansfield, W., Zaffran, S., Marioni, J.C., Fuks, F., Göttgens, B., Lescroart, F. & Blanpain, C. (2022) Mesp1 controls the chromatin and enhancer landscapes essential for spatiotemporal patterning of early cardiovascular progenitors. *Nature Cell Biology*. 24 (7), 1114–1128. doi:10.1038/s41556-022-00947-3.

Lindsley, R.C., Gill, J.G., Murphy, T.L., Langer, E.M., Cai, M., Mashayekhi, M., Wang, W., Niwa, N., Nerbonne, J.M., Kyba, M. & Murphy, K.M. (2008) Mesp1 Coordinately Regulates Cardiovascular Fate Restriction and Epithelial-Mesenchymal Transition in Differentiating ESCs. *Cell Stem Cell*. 3 (1), 55–68. doi:10.1016/j.stem.2008.04.004.

Lou, X., Deshwar, A.R., Crump, J.G. & Scott, I.C. (2011) Smarcd3b and Gata5 promote a cardiac progenitor fate in the zebrafish embryo. *Development*. 138 (15), 3113–3123. doi:10.1242/dev.064279.

Luna-Zurita, L., Stirnimann, C.U., Glatt, S., Kaynak, B.L., Thomas, S., Baudin, F., Samee, M.A.H., He, D., Small, E.M., Mileikovsky, M., Nagy, A., Holloway, A.K., Pollard, K.S., Müller, C.W. & Bruneau, B.G. (2016) Complex Interdependence Regulates Heterotypic Transcription Factor Distribution and Coordinates Cardiogenesis. *Cell*. 164 (5), 999–1014. doi:10.1016/j.cell.2016.01.004.

Mandal, A., Holowiecki, A., Song, Y.C. & Waxman, J.S. (2017) Wnt signaling balances specification of the cardiac and pharyngeal muscle fields. *Mechanisms of Development*. 143, 32–41. doi:10.1016/j.mod.2017.01.003.

Miquerol, L. & Kelly, R.G. (2013) Organogenesis of the vertebrate heart. *Wiley Interdisciplinary Reviews: Developmental Biology*. 2 (1), 17–29. doi:10.1002/wdev.68.

Molkentin, J.D., Antos, C., Mercer, B., Taigen, T., Miano, J.M. & Olson, E.N. (2000) Direct Activation of a GATA6 Cardiac Enhancer by Nkx2.5: Evidence for a Reinforcing Regulatory Network of Nkx2.5 and GATA Transcription Factors in the Developing Heart. *Developmental Biology*. 217 (2), 301–309. doi:10.1006/dbio.1999.9544.

Nees, S.N. & Chung, W.K. (2019) Genetic Basis of Human Congenital Heart Disease. *Cold Spring Harbor Perspectives in Biology*. 12 (9), a036749. doi:10.1101/cshperspect.a036749.

Nord, A.S., Blow, M.J., Attanasio, C., Akiyama, J.A., Holt, A., Hosseini, R., Phouanenavong, S., Plajzer-Frick, I., Shoukry, M., Afzal, V., Rubenstein, J.L.R., Rubin, E.M., Pennacchio, L.A. & Visel, A. (2013) Rapid and Pervasive Changes in Genome-wide Enhancer Usage during Mammalian Development. *Cell*. 155 (7), 1521–1531. doi:10.1016/j.cell.2013.11.033.

Olson, E.N. (2006) Gene Regulatory Networks in the Evolution and Development of the Heart. *Science*. 313 (5795), 1922–1927. doi:10.1126/science.1132292.

Pater, E. de, Ciampricotti, M., Priller, F., Veerkamp, J., Strate, I., Smith, K., Lagendijk, A.K., Schilling, T.F., Herzog, W., Abdelilah-Seyfried, S., Hammerschmidt, M. & Bakkers, J. (2012) Bmp Signaling Exerts Opposite Effects on Cardiac Differentiation. *Circulation Research*. 110 (4), 578–587. doi:10.1161/circresaha.111.261172.

Pikkarainen, S., Tokola, H., Kerkelä, R. & Ruskoaho, H. (2004) GATA transcription factors in the developing and adult heart. *Cardiovascular Research*. 63 (2), 196–207. doi:10.1016/j.cardiores.2004.03.025.

Probst, S., Sagar, S., Tosic, J., Schwan, C., Grün, D. & Arnold, S.J. (2020) Spatiotemporal sequence of mesoderm and endoderm lineage segregation

during mouse gastrulation. *Development*. 148 (1), dev193789. doi:10.1242/dev.193789.

Ranganayakulu, G., Elliott, D.A., Harvey, R.P. & Olson, E.N. (1998) Divergent roles for NK-2 class homeobox genes in cardiogenesis in flies and mice. *Development*. 125 (16), 3037–3048. doi:10.1242/dev.125.16.3037.

Reiter, J.F., Alexander, J., Rodaway, A., Yelon, D., Patient, R., Holder, N. & Stainier, D.Y.R. (1999) Gata5 is required for the development of the heart and endoderm in zebrafish. *Genes & Development*. 13 (22), 2983–2995. doi:10.1101/gad.13.22.2983.

Rojas, A., Schachterle, W., Xu, S. & Black, B.L. (2009) An endoderm-specific transcriptional enhancer from the mouse Gata4 gene requires GATA and homeodomain protein–binding sites for function in vivo. *Developmental Dynamics*. 238 (10), 2588–2598. doi:10.1002/dvdy.22091.

Rojas, A., Val, S.D., Heidt, A.B., Xu, S.-M., Bristow, J. & Black, B.L. (2005) Gata4 expression in lateral mesoderm is downstream of BMP4 and is activated directly by Forkhead and GATA transcription factors through a distal enhancer element. *Development*. 132 (15), 3405–3417. doi:10.1242/dev.01913.

Saga, Y., Hata, N., Kobayashi, S., Magnuson, T., Seldin, M.F. & Taketo, M.M. (1996) MesP1: a novel basic helix-loop-helix protein expressed in the nascent mesodermal cells during mouse gastrulation. *Development*. 122 (9), 2769–2778. doi:10.1242/dev.122.9.2769.

Saga, Y., Miyagawa-Tomita, S., Takagi, A., Kitajima, S., Miyazaki, J. i & Inoue, T. (1999) MesP1 is expressed in the heart precursor cells and required for the formation of a single heart tube. *Development*. 126 (15), 3437–3447. doi:10.1242/dev.126.15.3437.

Sam, J., Mercer, E.J., Torregroza, I., Banks, K.M. & Evans, T. (2020) Specificity, redundancy and dosage thresholds among gata4/5/6 genes during zebrafish cardiogenesis. *Biology Open*. 9 (6), bio053611. doi:10.1242/bio.053611.

Schans, V.A.M. van de, Smits, J.F.M. & Blankesteijn, W.M. (2008) The Wnt/frizzled pathway in cardiovascular development and disease: Friend or foe? *European Journal of Pharmacology*. 585 (2–3), 338–345. doi:10.1016/j.ejphar.2008.02.093.

Schultheiss, T.M., Burch, J.B. & Lassar, A.B. (1997) A role for bone morphogenetic proteins in the induction of cardiac myogenesis. *Genes & Development*. 11 (4), 451–462. doi:10.1101/gad.11.4.451.

Searcy, R.D., Vincent, E.B., Liberatore, C.M. & Yutzey, K.E. (1998) A GATA-dependent nkx-2.5 regulatory element activates early cardiac gene expression in transgenic mice. *Development*. 125 (22), 4461–4470. doi:10.1242/dev.125.22.4461.

Smemo, S., Campos, L.C., Moskowitz, I.P., Krieger, J.E., Pereira, A.C. & Nobrega, M.A. (2012) Regulatory variation in a TBX5 enhancer leads to isolated congenital heart disease. *Human Molecular Genetics*. 21 (14), 3255–3263. doi:10.1093/hmg/dds165.

Song, M., Yuan, X., Racioppi, C., Leslie, M., Stutt, N., Aleksandrova, A., Christiaen, L., Wilson, M.D. & Scott, I.C. (2022) GATA4/5/6 family transcription factors are conserved determinants of cardiac versus pharyngeal mesoderm fate. *Science Advances*. 8 (10), eabg0834. doi:10.1126/sciadv.abg0834.

Takeuchi, J.K., Lickert, H., Bisgrove, B.W., Sun, X., Yamamoto, M., Chawengsaksophak, K., Hamada, H., Yost, H.J., Rossant, J. & Bruneau, B.G. (2007) Baf60c is a nuclear Notch signaling component required for the establishment of left–right asymmetry. *Proceedings of the National Academy of Sciences*. 104 (3), 846–851. doi:10.1073/pnas.0608118104.

Tosic, J., Kim, G.-J., Pavlovic, M., Schröder, C.M., Mersiowsky, S.-L., Barg, M., Hofherr, A., Probst, S., Köttgen, M., Hein, L. & Arnold, S.J. (2019) Eomes and Brachyury control pluripotency exit and germ-layer segregation by changing the chromatin state. *Nature Cell Biology*. 21 (12), 1518–1531. doi:10.1038/s41556-019-0423-1.

Tyser, R.C.V., Ibarra-Soria, X., McDole, K., Jayaram, S.A., Godwin, J., Brand, T.A.H. van den, Miranda, A.M.A., Scialdone, A., Keller, P.J., Marioni, J.C. & Srinivas, S. (2021) Characterization of a common progenitor pool of the epicardium and myocardium. *Science*. eabb2986. doi:10.1126/science.abb2986.

Wu, S.M., Chien, K.R. & Mummery, C. (2008) Origins and Fates of Cardiovascular Progenitor Cells. *Cell*. 132 (4), 537–543. doi:10.1016/j.cell.2008.02.002.

Yuan, X., Scott, I.C. & Wilson, M.D. (2021) Heart Enhancers: Development and Disease Control at a Distance. *Frontiers in Genetics*. 12, 642975. doi:10.3389/fgene.2021.642975.

Yuan, X., Song, M., Devine, P., Bruneau, B.G., Scott, I.C. & Wilson, M.D. (2018) Heart enhancers with deeply conserved regulatory activity are established early in zebrafish development. *Nature Communications*. 9 (1), 4977. doi:10.1038/s41467-018-07451-z.

Zaidi, S. & Brueckner, M. (2017) Genetics and Genomics of Congenital Heart Disease. *Circulation Research*. 120 (6), 923–940. doi:10.1161/circresaha.116.309140.

Zeineddine, D., Papadimou, E., Chebli, K., Gineste, M., Liu, J., Grey, C., Thurig, S., Behfar, A., Wallace, V.A., Skerjanc, I.S. & Pucéat, M. (2006) Oct-3/4 Dose Dependently Regulates Specification of Embryonic Stem Cells toward a Cardiac Lineage and Early Heart Development. *Developmental Cell*. 11 (4), 535–546. doi:10.1016/j.devcel.2006.07.013.

Chapter 2: Characterizing the cardiac mesoderm during *in vivo* gastrulation with single cell transcriptomic sequencing

Abstract

Mammalian cardiogenesis begins at gastrulation (mouse ~E6.5), when *Mesp1*+ mesodermal progenitor cells migrate anterior-laterally from the primitive streak, already fated for distinct mesodermal substructures including the heart. The mechanisms governing cardiac precursor cell (CPC) specification from the mesoderm remain poorly understood, due in part to limitations of CPC-specific tools and markers for distinguishing the cardiac lineage from the surrounding mesoderm germ layer. We define and show utility for a computational method of "cell-sorting" that leverages classical lineage tracing in mouse embryos through bioinformatic analysis methods in single cell RNA sequencing (scRNA-seq) data. Using this method, we generate a scRNA-seq dataset of mouse embryos expressing fluorescent reporters for the *Mesp1* lineage and a pan-cardiac transgene driven by the *Smarcd3* "F6" enhancer. We characterize transcriptional identities of early CPCs to define the emerging heterogeneity of the cardiac mesoderm distinctly from the surrounding mesoderm. We show that the mesoderm developing between E6.0 – E7.75 generates a diversity of transcriptional profiles consistent with emerging cardiac lineage precursor cell subpopulations and contemporaneously specified precursor cell types of the non-cardiac mesoderm. Thus, we generate a high resolution scRNA-seq dataset containing a tight time course of mouse embryos across gastrulation up until cardiac crescent stages, with tracked expression of fluorescent lineage reporters for identification the emerging cardiac mesoderm.

Background

The heart is the first organ to develop in the embryo proper, and requires precise specification and patterning of cardiac progenitors during gastrulation. The mammalian heart is an anatomically complex organ containing specialized substructures arising from a pool of embryonic progenitors in the gastrulating mesoderm (Tyser et al., 2021; Kelly, Buckingham & Moorman, 2014; Tam et al., 1997; Evans et al., 2010; Bruneau, 2013; Zhang et al., 2021). This process of cardiogenesis relies on highly coordinated and regulated gene expression programs to specify cardiac precursor cells (CPCs) into diverse, terminally differentiated cell types within the four-chambered heart (Bruneau, 2008, 2013; Kelly, Buckingham & Moorman, 2014; Harvey, 2002). Understanding the molecular mechanisms specifying the cardiac lineage during embryonic development is integral to our efforts to design novel therapeutic strategies for congenital heart defects (CHDs) and reprogramming efforts to enhance or correct cardiac function through regenerative medicine. CHDs affect 1-2% of live births, while heart disease is the number one cause of fatality in adults in the Western World (Srivastava & Olson, 2000; Bruneau, 2008; Nees & Chung, 2019). Our current therapeutic options are limited in part because we require more mechanistic insight into how early CPCs are specified, patterned, and ultimately differentiated to build the heart. Thus, developing a transcriptional blueprint for CPC specification during embryonic development is a crucial prerequisite to designing strategies to combat CHDs and pathology of the heart.

Cardiogenic precursor cells are rapidly allocated from the gastrulating mesoderm at a developmental time point where we have a limited understanding of the extent of

cellular heterogeneity within either the cardiac lineage or the mesoderm germ layer broadly. Our understanding is hindered, in part, due to the lack of molecular markers available to further studies characterizing transcriptional diversity within relevant precursor cell pools at these early stages of cardiac mesoderm development. Dissecting mesoderm heterogeneity and the transcriptional programs specifying CPCs are necessary to advance our knowledge of the molecular mechanisms of cardiogenesis. Here, we leverage computational detection of endogenous fluorescent reporter transgenes for the cardiac mesoderm to create a scRNA-seq dataset of gastrulating mouse embryos in a tight timecourse of E6.0 to E7.75. In this manner, I couple use of classical developmental biology mouse models with high dimensional single cell transcriptomics to investigate a historically challenging question with a level of cardiac-specificity previously intractable in the developing embryo.

Early mesoderm progenitors emerging from the primitive streak express *Mesp1* and migrate towards the anterior-lateral aspects of the developing embryo where they will coalesce to form the cardiac crescent (Saga et al., 1999, 1996). Clonal lineage tracing studies of *Mesp1*+ mesodermal cells at gastrulation show that these cells are not cardiac-restricted, but contribute broadly to the heart cells, the endothelium of the aorta and brain, head and neck muscles, and somitic derivatives (Saga et al., 1999; Lescroart et al., 2010, 2014; Devine et al., 2014). Additionally, retrospective clonal lineage tracing at E6.5 using the *Mesp1*^{Cre}-MADM system illustrates that cardiac-fated *Mesp1* cells are prefigured at gastrulation to occupy distinct anatomic substructures of the heart, implying that regional identities are specified very early in cardiogenesis (Devine et al., 2014). Recent findings illustrate that CPC birthplace along the proximal-distal axis of the

primitive streak is correlated to eventual cell fate within specific anatomic regions within the heart (Zhang et al., 2021; Ivanovitch et al., 2021). Thus, while traditional mesoderm lineage tracing and longitudinal time course imaging studies have described rapid diversification of emerging CPCs and the heterogeneity of cardiac cell fates in the embryo, we still lack detailed comprehension of how gene expression changes occur during cell fate allocation. Thus, in order to understand the transcriptional profiles of this emerging CPC heterogeneity in higher resolution, we sought to leverage these lineage tracing techniques in evaluation of single cell transcriptomic profiles during cardiac development in the embryo.

Towards this effort, we first sought to improve our reporter system marking emerging CPCs within the gastrulating mesoderm. A characterized enhancer of *Smarcd3*, “F6”, shares expression with the *Mesp1* lineage within just CPCs. *Smarcd3*-F6::eGFP reporter transgene expression precedes expression of early cardiac TFs such as *Tbx5* and *Isl1*, early markers of the first and second heart fields, respectively (Devine et al., 2014; Kelly, Buckingham & Moorman, 2014; Zhang et al., 2021). The *Smarcd3* gene itself encodes Baf60c, a protein subunit of the Baf complex involved in chromatin remodeling during cardiac development (Takeuchi et al., 2007; Lickert et al., 2004; Sun et al., 2017; Hota et al., 2019). While constitutive genetic deletion of *Smarcd3* in mouse leads to cardiac hypoplasia with altered gene expression for contractile and cardiac metabolism programs between E12.5 - E14.5, the heart still forms, indicating no evidence for a *Smarcd3* -mediated effect on earlier CPC specification (Sun et al., 2017). Temporal gene expression analysis during *in vitro* differentiation of cardiomyocytes (CMs) from mouse embryonic stem cells (mESCs) shows that *Smarcd3* and *Smarcd3*-

F6 expression are rapidly up-regulated shortly after day 4, around the same time *Mesp1* expression is rapidly down-regulated *in vitro*. Bulk RNA-sequencing processed from *in vitro* cardiac differentiated *Smarcd3*-F6+ cells, isolated via cell-sorting of their fluorescent GFP reporter, revealed that *Smarcd3*-F6+ cells express genes associated with cardiac progenitors, such as *Hand2*, *Gata4*, and *Meis1* (Devine et al., 2014). *In vivo* studies reveal that *Smarcd3*-F6+ cells are present in all regions of the heart, and the *Smarcd3*-F6+ lineage is restricted to the heart (Devine et al., 2014; Yuan et al., 2018). Thus *Smarcd3*-F6 represents a novel molecular tool for the study of CPCs during early specification of the cardiac lineage from the developing mesoderm, uniquely suited to the goals of this study.

Results

Computational detection of fluorescent lineage transgenes in single cell transcriptomic data

To identify the emerging cardiogenic mesoderm cells at their earliest stages we employed a reporter transgene strategy in combination with scRNA-seq on a whole embryo time course spanning early gastrulation (E6.0) until cardiac crescent stages (E7.75) (Fig. 2.1). Embryos contained a fluorescent transgene reporter for the *Mesp1* lineage via $Mesp1^{Cre};Rosa26R^{Ai14}$ (Saga et al., 1999; Madisen et al., 2010), and the *Smarcd3*-F6::eGFP enhancer transgene that constitutively marks CPCs (Devine et al., 2014; Yuan et al., 2018). Due to the small size of mouse embryos during early

gastrulation stages, we sought to avoid processing dissociated embryos through physical fluorescent cell sorting. Physical sorting increases the risk of capturing altered gene expression profiles reflecting the sort-induced biomechanical stress of processing and increases the risk that prohibitive levels of cell loss from such small samples lead to inadequate material for assessment of cellular heterogeneity. Thus, we investigated the possibility to incorporate sequence detection of the fluorescent transgene reporters into our bioinformatic workflow following standard generation of single cell transcriptomic libraries from whole embryos. We determined that adding fluorescent reporter sequences to a pseudo-chromosome on the reference genome would theoretically enable us to detect transcripts during sequence read alignment, similar to how we detect the native transcripts within cells.

This effort was constrained by the chemistry of our single cell library preparation and sequencing paradigms; 10X Genomics, a highly reproducible and scalable commercial scRNA-seq method, utilizes a polydT sequence to capture 3' polyA tails of transcripts on beads, and then performs reverse transcription on the bound transcript so that libraries contain 3' biased capture of transcripts. Furthermore, 500 bp products are required for high throughput bridge amplification Illumina sequencing, limiting the depth of gene sequence that will ultimately be aligned to the genome for annotation. The sequences of our endogenous mouse fluorescent reporters were cloned with similar 3' ends, which prohibited dual reporter sequence detection during alignment. On account of the nearly identical 500 bp 3' region in each reporter, reads would get discarded for mapping to both reporter sequences in the reference genome. To circumvent the cross-alignment issue and maintain dual-reporter detection in our assay, we turned to the 10X

V(D)J library preparation system, which uses a switch oligo to capture 5' biased reads after unbound reverse transcription in the gel bead. Thus, since our fluorescent reporters had distinct 5' sequences, we were able to capture their expression in a manner that would render the colors distinguishable from each other during alignment to the reporter-appended reference genome. In this manner, we devised a system to leverage detection of endogenous mouse fluorescent reporters to enhance resolution of lineage relationships within single cell transcriptomic data. This innovation served to enhance our understanding of transcriptional profiles of early gastrulation cell types, when embryos aren't yet expressing many of the organ-restricted transcription factors typically used as markers for various cell lineages.

Cardiac mesoderm diversity during early gastrulation

Next, we deployed our method to process whole embryos for scRNA-seq, leveraging computational detection of the fluorescent transgene reporters during sequence alignment to identify the emerging cardiogenic mesoderm. Initially, we generated an atlas representing a fine time course of mouse gastrulation (Fig. 2.2A-B) and coarsely annotated clusters (Fig. 2.2D) to identify relevant mesodermal cell types. We then examined where expression for the fluorescent transgenes was detected within these clusters, finding *Mesp1* lineage reporter Ai14 and CPC-specific *Smarcd3*-F6-eGFP were restricted to mesodermal and cardiac mesodermal cell types within this dataset. *Smarcd3*-F6 expression is restricted within the *Mesp1* lineage, and thus more non-cardiac mesodermal derivatives expectedly express Ai14 while eGFP was restricted to early cardiac mesoderm cell types (Fig. 2.2A,C).

Following annotation of cell types in the whole embryo atlas (Fig. 2.2D), we subsetted cell clusters expressing the Ai14 and eGFP fluorescent transgenes and demonstrated that these cells resemble the emerging cardiac mesoderm through annotation of known cell type specific marker genes (Fig. 2.3). *Smarcd3*-F6+ cell clusters (Fig. 2.3A,C, Fig. S2.1) co-expressed early cardiac and cardiac mesoderm genes in the mesoderm exiting the primitive streak (meso exiting PS), anterior mesendoderm (antME), and cells of the lateral plate mesoderm (LPM) such as *T, Eomes*, *Mesp1*, *Mixl1,* and *Smarcd3* (Fig. 2.3A,C, Fig. S2.1A). *Smarcd3*-F6+ cardiomyocytes (CMs) co-expressed cardiac structural genes such as *Myl7*, *Tnnt2*, *Actc1* (Fig. 2.3A,C, Fig. S2.1B) and cardiac TFs such as *Tbx5*, *Hand1*, *Nkx2-5, Gata5* (Fig. 2.3A,C, Fig. S2.1C). Thus, *Smarcd3*-F6 enhancer transgene expressing cells have transcriptional signatures of early emerging CPCs, extending the initial description of this transgene (Devine et al., 2014) by validating high-fidelity demarcation of emerging early CPCs within the mesoderm prior to expression of cardiac-specific TFs in scRNA-seq data.

In summary, we generated a resource dataset and interrogated dynamic gene expression programs in the cardiogenic mesoderm using reporter transgenes, which will facilitate investigations and characterization of emerging heterogeneity within the cardiac lineage during gastrulation.

Trajectory analysis of emerging cardiac mesoderm reveals fate heterogeneity in progenitor populations

We next investigated how initiation of cardiac specification establishes heterogeneity of the cardiac mesoderm. Utilizing pseudotemporal trajectory ordering with URD (Farrell et

al., 2018) on the mesoderm dataset (Fig. 2.3A). We defined epiblast2 cells (C14), containing the youngest embryos, as the root and clusters containing the most differentiated mesodermal cells from the oldest stage embryos for the tips (Fig. 2.3A,B, Fig. 4.4A). Embryos ordering according to biological age in pseudotime (Fig. 2.4A). We layered expression of *Smarcd3*-F6-eGFP to identify the main cardiogenic path paths within the tree space (Fig. 2.4B), which also co-expressed CM genes such as *Tbx5* and *Hand2*, representing early TFs for the first and second heart fields, respectively (Fig. 2.4D) (Kelly, Buckingham & Moorman, 2014). Within the CMs1 and CMs2 fate branches, we see enriched expression for structural myocyte genes *Tnnt2* and *Myl7*, more so than cardiac transcription factors and initiating expression in earlier pseudotime fate branches (Fig. 2.4D), indicating that the structural myocyte identity of CPCs is defined either prior or independently of TFs for CPC patterning. We noted that the pseudotemporally-preceding fate branch events denoted C10-preCardiacMeso as the most closely related to CPCs, and the C13-postLPM and C16-PrxM fates being the next closest neighbors (Fig. 2.3A). We showed that broad expression of *Pou5f1* expectedly occupies earlier fate branches of the tree, consistent with its role in differentiating early germ layers during gastrulation, and continuing expression occurs in cardiac mesoderm cells where *Pou5f1* functions to promoted cell differentiation (Niwa, Miyazaki & Smith, 2000; Wu & Schöler, 2014; Plachta et al., 2011) (Fig. 2.4C). We observed transient, early *Mesp1* expression that rapidly turned off, as expected (Fig. 2.4C). We highlighted restricted expression of *Ramp2* and *Hba-x* in *Mesp1*-lineage derived fates of HPCs and endothelial cells (Fig. 2.4E). While this trajectory tree maps fate potentials of the total mesoderm, we can utilize expression of cardiac relevant transgene *Smarcd3*-F6 to

focus specifically on the far-left side of the tree to evaluate the key genetic differences between the related cardiac fates. In this manner, this dataset is a critical resource for investigating gene expression alterations during cardiac development and the mechanisms for allocation of cell fate heterogeneity within the cardiac mesoderm.

Discussion

By engineering a method for computational transgene detection using commercial scRNA-seq technology from 10X Genomics, we were able to avoid introducing the deleterious effects of a physical sort on small embryos into our dataset as confounding caveats. Thus, our dataset is uniquely suited to offer a complete picture of *in vivo* cellular heterogeneity in whole embryos, with distinct labeling of emerging CPCs in the mesoderm. Preliminary analyses reveal dynamic gene expression patterns during cardiac mesoderm development, highlighting the emerging heterogeneity during embryonic development in gastrulation.

Following annotation of cell types within the whole embryo scRNA-seq dataset, we evaluated presence of transgenes in clusters for mesoderm, cardiac mesoderm, and cardiac cell types, finding that reporters overlayed in expected cells and would enable us to bioinformatically subset our cardiac mesoderm clusters of interest, even in cases such as epiblast where heterogeneity of the emerging cardiac mesoderm is beginning but not yet known.

We finely annotated our wildtype mesoderm atlas, identifying multiple cardiac cell types, including two kinds of mature cardiomyocytes. We next defined relationships of cells along cardiac mesoderm and mesodermal fate paths in pseudotime, showing distinct branch points. Further differential gene expression analysis will enable distinction of cardiac regulatory candidates occupying different branch points.

Through these analyses we have generated a resource dataset and method for interrogation of dynamic gene expression programs in the cardiogenic mesoderm using fluorescent transgenes in whole embryo single cell transcriptomic data. We can further deploy this system towards describing of the emerging heterogeneity of the cardiac lineage during gastrulation. Our transcriptional profiling further cements that *Smarcd3*-F6 expression is a valid and restricted reporter of early CPCs within the mesoderm.

Figures

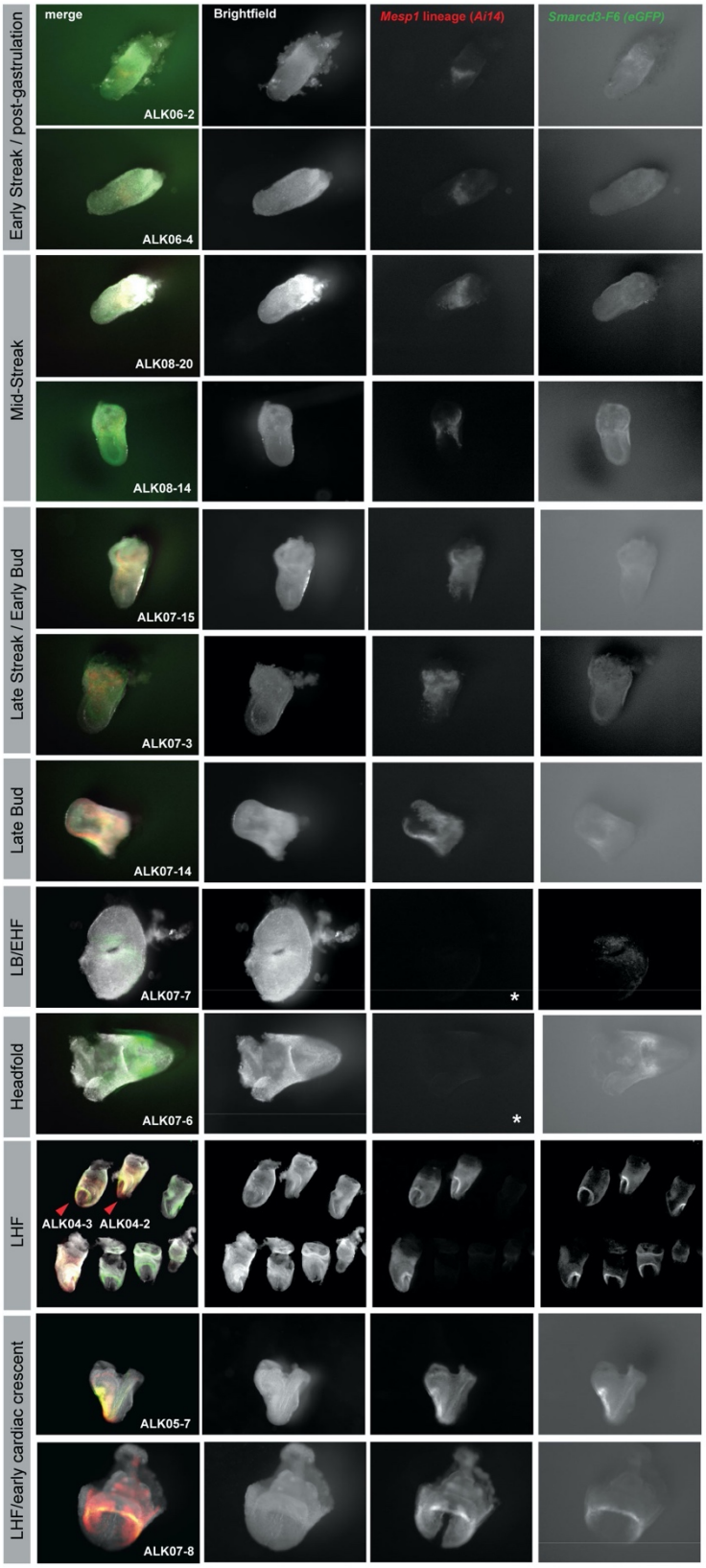

merge
Brightfield
Mesp1 lineage (Ai14)
Smarcd3-F6 (eGFP)
Early Streak / post-gastrulation
ALK06-2
ALK06-4
Mid-Streak
ALK08-20
ALK08-14
Late Streak / Early Bud
ALK07-15
ALK07-3
Late Bud
ALK07-14
LB/EHF
ALK07-7
*
Headfold
ALK07-6
*
LHF
ALK04-3
ALK04-2
LHF/early cardiac crescent
ALK05-7
ALK07-8

Figure 2.1. Fluorescent lineage transgenes in whole embryos. Images of all embryos utilized in generation of wildtype gastrulation atlas. *Mesp1* lineage visualized by Ai14 fluorescent reporter transgene. *Smarcd3*-F6 visualized by eGFP fluorescent reporter transgene. Images not acquired and processed identically. Embryos denoted with * lacked *Mesp1* lineage tracing by Ai14 transgene.

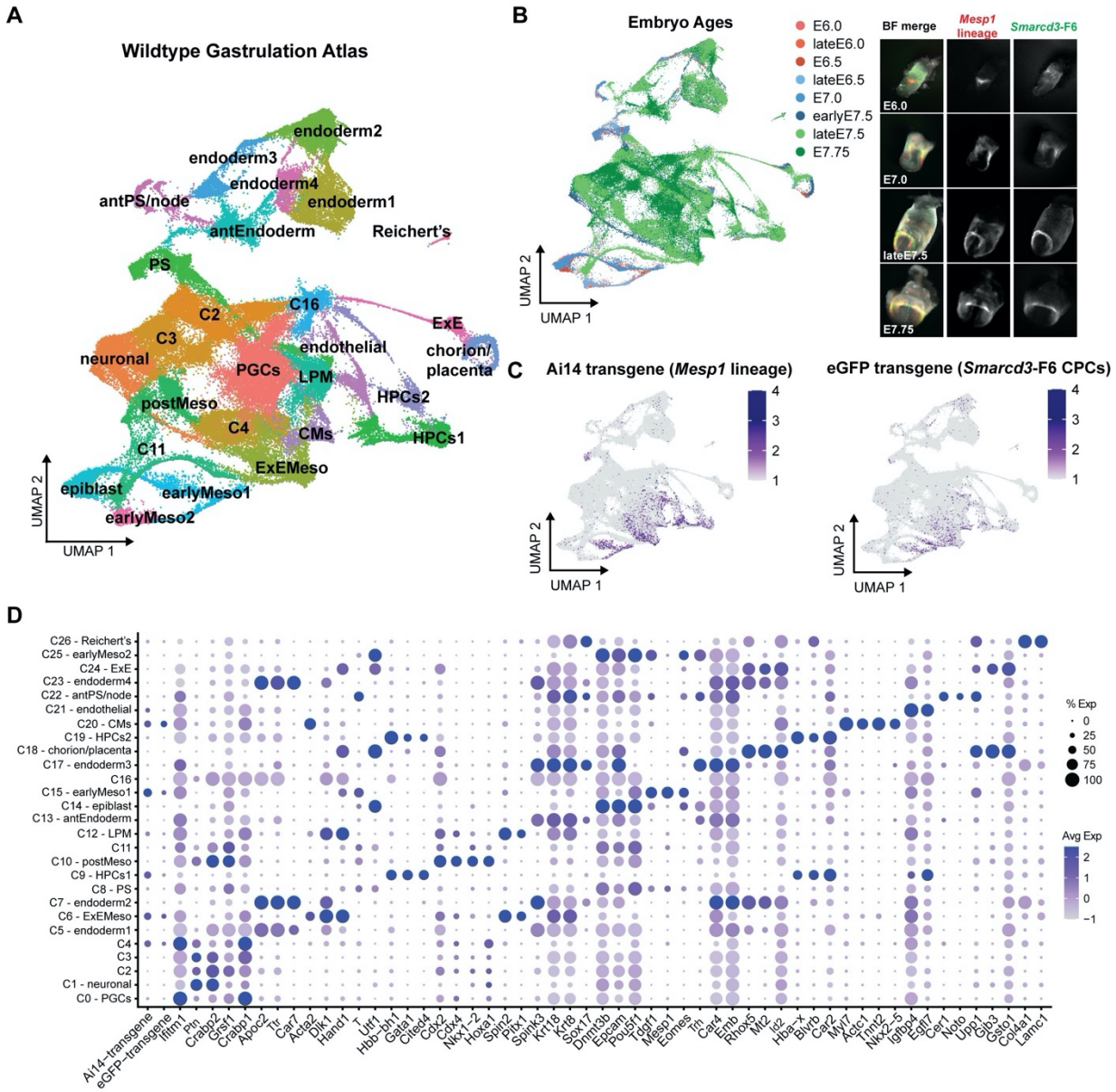


Figure 2.2. Utilizing fluorescent transgenes to identify the mesoderm in scRNA-seq. (A) Uniform manifold approximation and projection (UMAP) of 94,824 cells representing 27 cell types from gastrulating embryos. (B) UMAP labeled with embryo ages included in atlas and representative embryo images showing domains of fluorescent Ai14 (*Mesp1* lineage) and eGFP (*Smarcd3*-F6) transgenes. Images not scaled. (C) UMAP feature plots showing expression of fluorescent transgenes isolated to mesodermal cell types. (D) Dotplot denoting marker genes and cell type annotations by cluster in full embryo wildtype gastrulation atlas. Size of dot represents percent of cells expressing gene and color represents average expression level. Cluster number used to denote cell types when annotation was not possible.

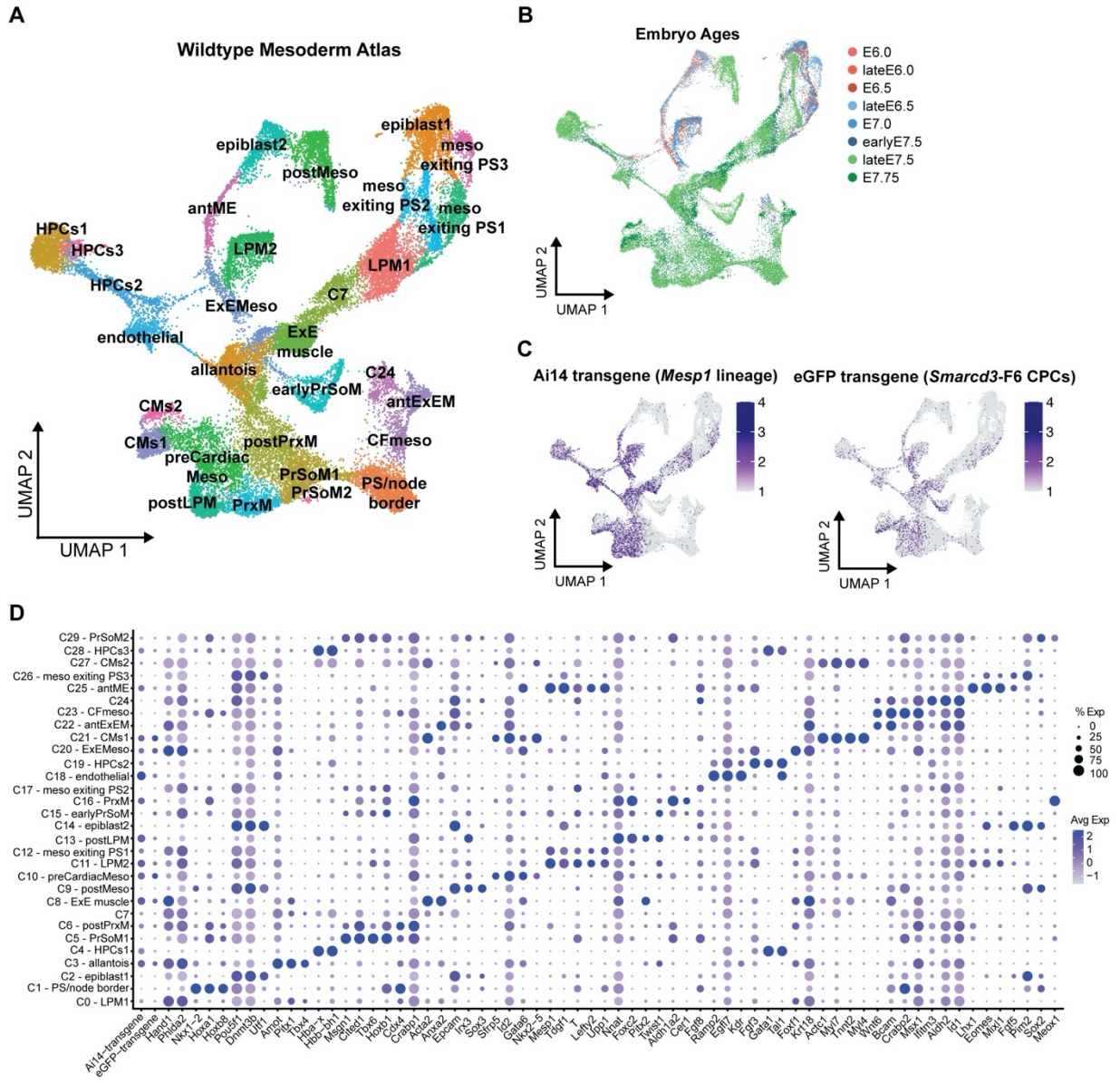


Figure 2.3. Cardiac mesoderm heterogeneity. (A) UMAP of 34,724 mesodermal cells subsetted from full atlas, representing 30 cell types. (B) UMAP labeled with embryo ages and (C) UMAP feature plots showing expression of fluorescent transgenes, Ai14 for *Mesp1* lineage and eGFP for *Smarcd3*-F6+ CPCs. (D) Doplot denoting marker genes and cell type annotations by cluster in mesoderm wildtype atlas. Size of dot represents percent of cells expressing gene and color represents average expression level. Cluster number used to denote cell types when annotation was not possible.

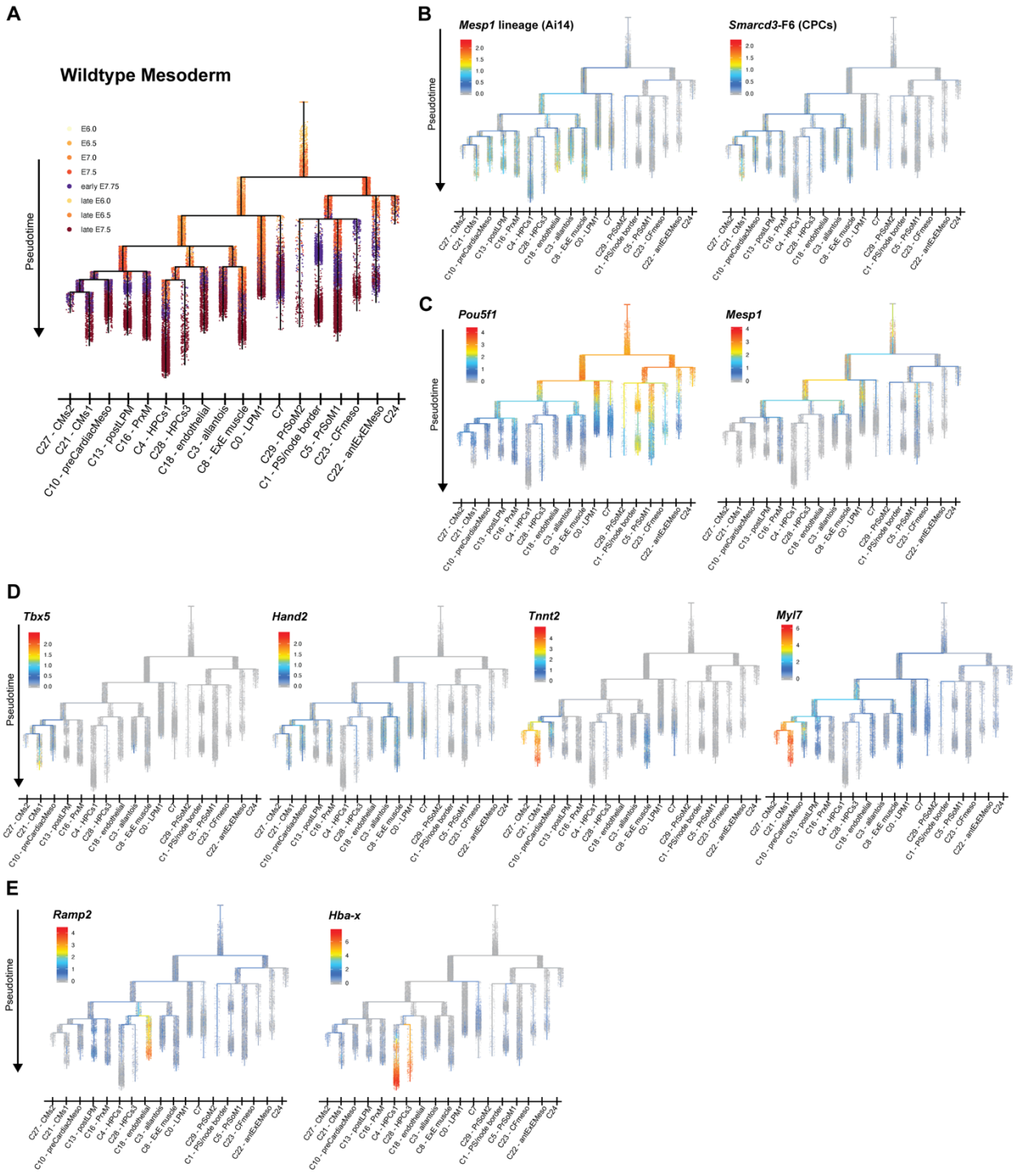
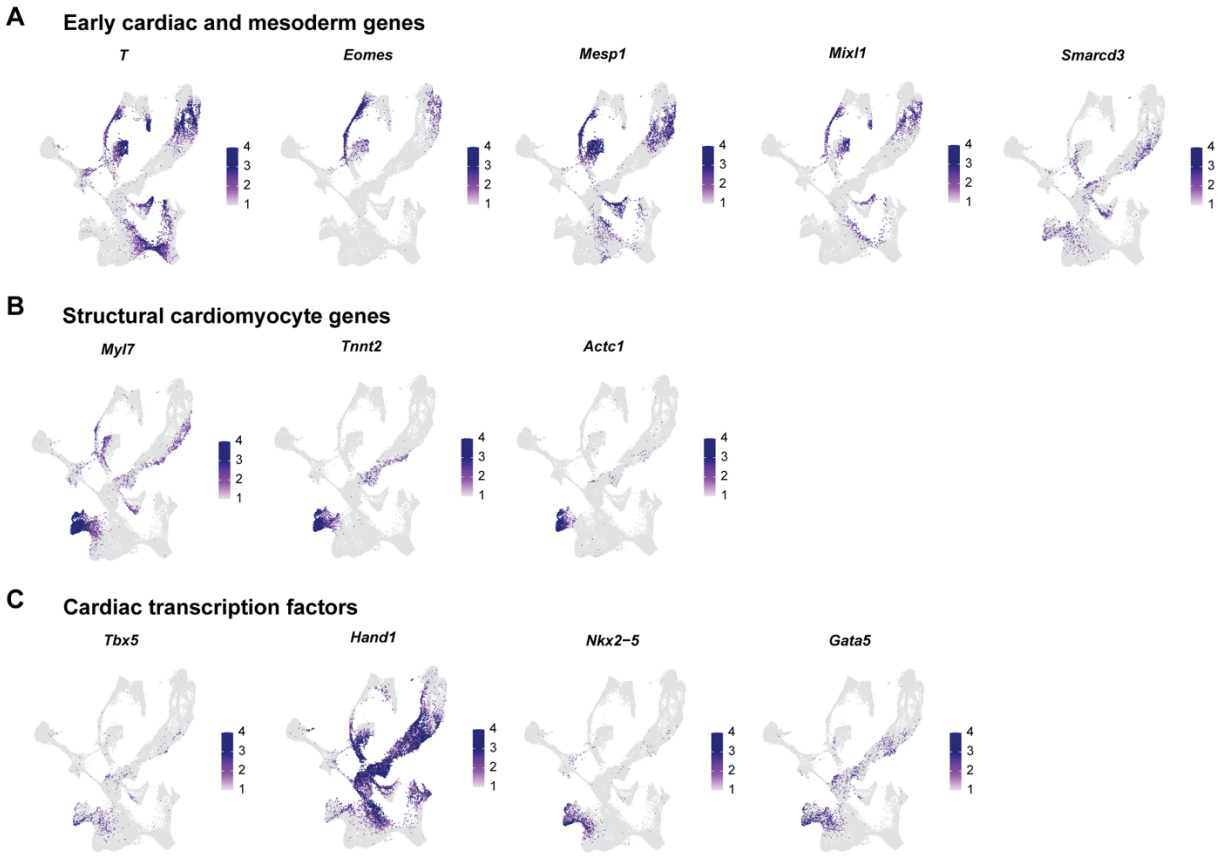


Figure 2.4. Heterogeneous fates for cardiac progenitors. (A) URD pseudotime trajectory tree for fate progression towards mesoderm fates, colored by embryonic age. (B) Overlay of fluorescent *Mesp1* lineage reporter transgerne and *Smarcd3*-F6 CPC transgene. (C) Overlay of early mesoderm genes *Pou5f1* and *Mesp1*. (D) Overlay of early cardiac TFs *Tbx5*, *Hand2*; and structural myocyte genes *Tnnt2, Myl7*. (E) Overlay of non-cardiac mesoderm genes *Ramp2* and *Hba-x*.



Supplementary Figure 2.1. Co-expression of cardiac mesoderm genes in *Smarcd3*-F6+ clusters in mesoderm. UMAP feature plots of wildtype mesoderm atlas showing gene expression (A) for early cardiac and mesoderm genes *T, Eomes, Mesp1, Mixl1, Smarcd3*, (B) structural cardiomyocyte genes *Myl7, Tnnt2, Actc1*, and (C) cardiac transcription factors *Tbx5, Hand1, Nkx2-5, Gata5*.

Materials and Methods

Mouse models

Animal studies were performed in strict compliance with the UCSF Institutional Animal Care and Use Committee. Mice were housed in a standard 12 hour light/dark animal husbandry barrier facility at the Gladstone Institutes. The *Mesp1*$^{Cre/+}$ knock-in mice were obtained from Yumiko Saga (Ajima et al., 2021; Saga et al., 1999). *Rosa26R*^{Ai14} mice were from Jackson Laboratory (strain #007914, (Madisen et al., 2010).

Embryos were generated from crosses of *Mesp1*$^{Cre/+}$*;Rosa26R*^{Ai14}*;Hipp11*$^{Smarcd3\text{-}F6::eGFP}$ males to C57BL/6J wildtype, *Mesp1*$^{Cre/+}$, or *Mesp1*$^{Cre/+}$*;Rosa26R*^{Ai14}*;Hipp11*$^{Smarcd3\text{-}F6::eGFP}$ females. “Control” denotes embryos with at least one wildtype allele in the *Mesp1* locus and includes genotypes *Mesp1*$^{Cre/+}$*;Rosa26R*^{Ai14}*;Hipp11*$^{Smarcd3\text{-}F6::eGFP}$, *Mesp1*$^{Cre/+}$*;Rosa26R*$^{Ai14/+}$ *;Hipp11*$^{Smarcd3\text{-}F6::eGFP/+}$, *Mesp1*$^{+/+}$*;Rosa26R*^{Ai14}*;Hipp11*$^{Smarcd3\text{-}F6::eGFP}$, or *Mesp1*$^{+/+}$*;Rosa26R*$^{Ai14/+}$ *Hipp11*$^{Smarcd3\text{-}F6::eGFP/+}$. Heterozygosity of *Mesp1*$^{Cre/+}$ or straight wildtype allele combination *Mesp1*$^{+/+}$ is noted when embryos were utilized in scRNA-seq library generation (Fig. 2.1).

Cloning and generation of TARGATT transgenic knock-in mice

The *Smarcd3*-F6 fragment was isolated and cloned with inclusion of an *nlsEGFP* under control of an *Hsp68* minimal promoter for TARGATT (Applied Stem Cells) insertion to the *Hipp11* locus as previously described (Devine et al., 2014) to create the *Hipp11*$^{Smarcd3\text{-}F6::eGFP}$ mouse. Purified construct DNA was injected into embryo pronuclei along with mRNA for the *Phi31o* transposase according to manufacturer’s protocols.

Timed matings and whole embryo dissections

To achieve timed matings, male and female mice were housed together in the evening and pregnancy was assessed by vaginal plug the following morning. Gestational stage was determined starting as day E0.5 at noon of plug detection. Females were confirmed pregnant by abdominal ultrasound (Vevo 3100, Visual Sonics) the afternoon of day 6 or else the morning of day 7, and sacrificed according to IACUC standard procedure at noon on day 7, or the early morning of day 8, as dictated by the desired age of embryo for the particular experiment. The embryonic ages captured in individual litters ranged from E6.0 to E7.5 on day 7, and E7.5 to E7.75 on day 8. The diversity of ages in litters aided in the construction of a fine timecourse for both mutant and control timelines.

Embryos were dissected and in later stages when yolk is present, also de-yolked, in ice-cold PBS (Life Technologies, 14190250) with 1% FBS (Thermo Fisher Scientific, 10439016) on ice. Embryos were screened for reporter expression using an upright epifluorescent dissecting microscope (Leica MZFLIII microscope, Lumen Dynamics XCite 120LED light source, Leica DFC 3000G camera) for presence of both red and green fluorescence, indicative of *Mesp1* lineage tracing from $Mesp1^{Cre}$;$Rosa26R^{Ai14}$ alleles and expression of the *Smarcd3*-F6::eGFP transgene reporter from the $Hipp11^{Smarcd3\text{-}F6::eGFP}$ allele, respectively. Embryos were staged according to standard schema (Downs & Davies, 1993). For difficult-to-capture control stages used in construction of the wildtype scRNA-seq timeline, absence of *Mesp1* lineage (Ai14) reporter was permitted and noted for those embryos (Fig. 2.1).

Embryo preparation for single-cell library generation

Due to the small size and lack of morphological distinction between tissue types of embryos at these early stages, which would otherwise enable microdissection of tissue, whole embryos were dissected and harvested for single cell library generation.

Whole embryos were incubated in 200 µL 0.25% TrypLE (ThermoFisher Scientific, 12563029) solution for 5 min at 37°C and triturated gently. Dissociated cell suspension was quenched with 600 µL of PBS with 1% FBS, singularized via passage through a 70 µm cell strainer (BD Falcon, 352235), pelleted by centrifugation at 150xg for 3 min, and resuspended in 34 µL of PBS with 1% FBS. At least 2 embryos were collected per genotype per embryonic stage in the datasets.

Single-cell transcriptome library preparation and sequencing

Libraries for scRNA-seq were prepared according to manufacturer's instructions using the 10X Genomics Chromium controller, Chromium Single Cell 5' Library and Gel Bead Kit v1 (10X Genomics, 1000006) and Chromium Single Cell A Chip Kit (10X Genomics, 1000151). A maximum of 10,000 cells per sample were loaded onto the 10X Genomics Chromium instrument, and each sample was indexed with a unique sample identifier (10X Genomics Chromium i7 Multiplex Kit, 120262). Final libraries were pooled and sequenced shallowly according to 10X protocol parameters on a NextSeq500 (Illumina), and then re-pooled for deeper sequencing on HighSeq4000 (Illumina) and/or NovaSeq using an S4 lane (Illumina). Littermate, stage-matched replicates were always

sequenced together in the same library pool. All scRNA-seq libraries were sequenced to a mean read depth of at least 50,000 total aligned reads per cell.

Processing raw scRNA-seq

Raw sequencing reads were processed using the 10X Genomics Cellranger v3.0.2 pipeline. Reads were demultiplexed using cellranger mkfastq and aligned with cellranger count to the Mm10 reference genome containing additional sequences for the Ai14 and eGFP. Cellranger “aggr” was used to aggregate and read depth normalize multiple GEM libraries to generate the dataset.

Seurat analysis of scRNA-seq data

Outputs from the Cellranger pipeline were analyzed using the Seurat Package v3.0.2 in R (Butler et al., 2018; Stuart et al., 2019; Satija et al., 2015). A single aggregated counts matrix was used as input for Read10X and CreateSeuratObject functions. Quality control steps were performed to remove dead cells or doublets.

Wildtype Atlas:

For the wildtype atlas, cells with <10% mitochondrial reads, UMI counts less than 50,000, and detected genes between 200 and 6,300 were retained. SCTransform (Hafemeister & Satija, 2019) was used to normalize and scale data with regressions performed with respect to mitochondrial percent, number of genes, and number of UMI counts detected. PCA analysis and batch correction were performed using FastMNN (Haghverdi et al., 2018) split by experimental group (experiment number denoted with

library prefixes ALK06, ALK08, ALK07, ALK05, ALK04). 94,824 cells were clustered based on the top 50 principal components and visualized using RunUMAP, FindNeighbors, and FindClusters and outputs were visualized as Uniform Manifold Approximation and Projection (UMAP) embeddings generated with DimPlot. Cell types were annotated at clustering resolution 0.4 using the FindAllMarkers function with Wilcoxon rank-sum test (min.pct = 0.1, logfc threshold = 0.25) to identify cluster specific marker genes. Relevant mesoderm cell types were subsetted based on cluster-wise detection of Smarcd3-F6::eGFP and Ai14 transgenes for CPCs and the Mesp1 lineage, respectively. The resulting 34,724 were re-clustered and re-annotated at resolution 1.2 to create the cardiac mesoderm wildtype atlas.

Single cell transcriptomic cell trajectories and pseudotime analysis

Pseudotime analysis was performed using the URD package (version 1.0.2 and 1.1.1) (Farrell et al., 2018). The Seurat object was first converted to an URD object using the seuratToURD function. Cell-to-cell transition probabilities were constructed by setting the number of near neighbors (knn) to 189 and sigma to 10. Pseudotime was then calculated by running 80 flood simulations with *Pou5f1*+ epiblast cells containing earliest staged staged embryos (cluster 14 of WT mesoderm Seurat Object resolution 1.2) as the “root” cells. Clusters containing the most defined mesodermal derivative cell types and containing the latest staged embryos were set as the “tip” cells (C4-HPCs1, C28-HPCs3, C18-endothelial, C3-allantois, C27-CMs2, C21-CMs1, C10-preCardiacMeso, C13-postLPM, C16-PrxM, C29-PrSoM2, C5-PrSoM1, C1-PS/node border, C23-CFmeso, C22-antExEM, C24, C8-ExE muscle, C7, C0-LPM1). The resulting URD tree

was subsequently built by simulated random walks from each tip. Overlay of relative developmental ages from embryo data was used to show consensus in pseudotime estimations of cell trajectories. Overlay of *Smarcd3*-F6::eGFP and various cardiac marker genes such as *Tbx5*, *Myl7*, *Hand2,* and *Tnnt2*, were used to identify the relevant cardiac-fated branching segments of the URD tree.

References

Ajima, R., Sakakibara, Y., Sakurai-Yamatani, N., Muraoka, M. & Saga, Y. (2021) Formal proof of the requirement of MESP1 and MESP2 in mesoderm specification and their transcriptional control via specific enhancers in mice. *Development*. 148 (20). doi:10.1242/dev.194613.

Bruneau, B.G. (2013) Signaling and Transcriptional Networks in Heart Development and Regeneration. *Cold Spring Harbor Perspectives in Biology*. 5 (3), a008292. doi:10.1101/cshperspect.a008292.

Bruneau, B.G. (2008) The developmental genetics of congenital heart disease. *Nature*. 451 (7181), 943–948. doi:10.1038/nature06801.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. (2018) Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology*. 36 (5), 411–420. doi:10.1038/nbt.4096.

Devine, W.P., Wythe, J.D., George, M., Koshiba-Takeuchi, K. & Bruneau, B.G. (2014) Early patterning and specification of cardiac progenitors in gastrulating mesoderm. *eLife*. 3, e03848. doi:10.7554/elife.03848.

Downs, K.M. & Davies, T. (1993) Staging of gastrulating mouse embryos by morphological landmarks in the dissecting microscope. *Development*. 118 (4), 1255–1266. doi:10.1242/dev.118.4.1255.

Evans, S.M., Yelon, D., Conlon, F.L. & Kirby, M.L. (2010) Myocardial Lineage Development. *Circulation Research*. 107 (12), 1428–1444. doi:10.1161/circresaha.110.227405.

Farrell, J.A., Wang, Y., Riesenfeld, S.J., Shekhar, K., Regev, A. & Schier, A.F. (2018) Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science*. 360 (6392), eaar3131. doi:10.1126/science.aar3131.

Hafemeister, C. & Satija, R. (2019) Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biology*. 20 (1), 296. doi:10.1186/s13059-019-1874-1.

Haghverdi, L., Lun, A.T.L., Morgan, M.D. & Marioni, J.C. (2018) Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nature Biotechnology*. 36 (5), 421–427. doi:10.1038/nbt.4091.

Harvey, R.P. (2002) Patterning the vertebrate heart. *Nature Reviews Genetics*. 3 (7), 544–556. doi:10.1038/nrg843.

Hota, S.K., Johnson, J.R., Verschueren, E., Thomas, R., Blotnick, A.M., Zhu, Y., Sun, X., Pennacchio, L.A., Krogan, N.J. & Bruneau, B.G. (2019) Dynamic BAF chromatin remodeling complex subunit inclusion promotes temporally distinct gene expression programs in cardiogenesis. *Development*. 146 (19), dev.174086. doi:10.1242/dev.174086.

Ivanovitch, K., Soro-Barrio, P., Chakravarty, P., Jones, R.A., Bell, D.M., Gharavy, S.N.M., Stamataki, D., Delile, J., Smith, J.C. & Briscoe, J. (2021) Ventricular, atrial, and outflow tract heart progenitors arise from spatially and molecularly distinct regions of the primitive streak. *PLoS Biology*. 19 (5), e3001200. doi:10.1371/journal.pbio.3001200.

Kelly, R.G., Buckingham, M.E. & Moorman, A.F. (2014) Heart Fields and Cardiac Morphogenesis. *Cold Spring Harbor Perspectives in Medicine*. 4 (10), a015750. doi:10.1101/cshperspect.a015750.

Lescroart, F., Chabab, S., Lin, X., Rulands, S., Paulissen, C., Rodolosse, A., Auer, H., Achouri, Y., Dubois, C., Bondue, A., Simons, B.D. & Blanpain, C. (2014) Early lineage restriction in temporally distinct populations of Mesp1 progenitors during mammalian heart development. *Nature Cell Biology*. 16 (9), 829–840. doi:10.1038/ncb3024.

Lescroart, F., Kelly, R.G., Garrec, J.-F.L., Nicolas, J.-F., Meilhac, S.M. & Buckingham, M. (2010) Clonal analysis reveals common lineage relationships between head muscles and second heart field derivatives in the mouse embryo. *Development*. 137 (19), 3269–3279. doi:10.1242/dev.050674.

Lickert, H., Takeuchi, J.K., Both, I. von, Walls, J.R., McAuliffe, F., Adamson, S.L., Henkelman, R.M., Wrana, J.L., Rossant, J. & Bruneau, B.G. (2004) Baf60c is essential for function of BAF chromatin remodelling complexes in heart development. *Nature*. 432 (7013), 107–112. doi:10.1038/nature03071.

Madisen, L., Zwingman, T.A., Sunkin, S.M., Oh, S.W., Zariwala, H.A., Gu, H., Ng, L.L., Palmiter, R.D., Hawrylycz, M.J., Jones, A.R., Lein, E.S. & Zeng, H. (2010) A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nature Neuroscience*. 13 (1), 133–140. doi:10.1038/nn.2467.

Nees, S.N. & Chung, W.K. (2019) Genetic Basis of Human Congenital Heart Disease. *Cold Spring Harbor Perspectives in Biology*. 12 (9), a036749. doi:10.1101/cshperspect.a036749.

Niwa, H., Miyazaki, J. & Smith, A.G. (2000) Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nature Genetics*. 24 (4), 372–376. doi:10.1038/74199.

Plachta, N., Bollenbach, T., Pease, S., Fraser, S.E. & Pantazis, P. (2011) Oct4 kinetics predict cell lineage patterning in the early mammalian embryo. *Nature Cell Biology*. 13 (2), 117–123. doi:10.1038/ncb2154.

Saga, Y., Hata, N., Kobayashi, S., Magnuson, T., Seldin, M.F. & Taketo, M.M. (1996) MesP1: a novel basic helix-loop-helix protein expressed in the nascent mesodermal cells during mouse gastrulation. *Development (Cambridge, England)*. 122 (9), 2769–2778.

Saga, Y., Miyagawa-Tomita, S., Takagi, A., Kitajima, S., Miyazaki, J. i & Inoue, T. (1999) MesP1 is expressed in the heart precursor cells and required for the formation of a single heart tube. *Development*. 126 (15), 3437–3447. doi:10.1242/dev.126.15.3437.

Satija, R., Farrell, J.A., Gennert, D., Schier, A.F. & Regev, A. (2015) Spatial reconstruction of single-cell gene expression data. *Nature Biotechnology*. 33 (5), 495–502. doi:10.1038/nbt.3192.

Srivastava, D. & Olson, E.N. (2000) A genetic blueprint for cardiac development. *Nature*. 407 (6801), 221–226. doi:10.1038/35025190.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P. & Satija, R. (2019) Comprehensive Integration of Single-Cell Data. *Cell*. 177 (7), 1888-1902.e21. doi:10.1016/j.cell.2019.05.031.

Sun, X., Hota, S.K., Zhou, Y.-Q., Novak, S., Miguel-Perez, D., Christodoulou, D., Seidman, C.E., Seidman, J.G., Gregorio, C.C., Henkelman, R.M., Rossant, J. & Bruneau, B.G. (2017) Cardiac-enriched BAF chromatin-remodeling complex subunit Baf60c regulates gene expression programs essential for heart development and function. *Biology Open*. 7 (1), bio029512. doi:10.1242/bio.029512.

Takeuchi, J.K., Lickert, H., Bisgrove, B.W., Sun, X., Yamamoto, M., Chawengsaksophak, K., Hamada, H., Yost, H.J., Rossant, J. & Bruneau, B.G. (2007) Baf60c is a nuclear Notch signaling component required for the establishment of left–right asymmetry. *Proceedings of the National Academy of Sciences*. 104 (3), 846–851. doi:10.1073/pnas.0608118104.

Tam, P.P., Parameswaran, M., Kinder, S.J. & Weinberger, R.P. (1997) The allocation of epiblast cells to the embryonic heart and other mesodermal lineages: the role of

ingression and tissue movement during gastrulation. *Development (Cambridge, England)*. 124 (9), 1631–1642.

Tyser, R.C.V., Ibarra-Soria, X., McDole, K., Jayaram, S.A., Godwin, J., Brand, T.A.H. van den, Miranda, A.M.A., Scialdone, A., Keller, P.J., Marioni, J.C. & Srinivas, S. (2021) Characterization of a common progenitor pool of the epicardium and myocardium. *Science*. eabb2986. doi:10.1126/science.abb2986.

Wu, G. & Schöler, H.R. (2014) Role of Oct4 in the early embryo development. *Cell Regeneration*. 3 (1), 7. doi:10.1186/2045-9769-3-7.

Yuan, X., Song, M., Devine, P., Bruneau, B.G., Scott, I.C. & Wilson, M.D. (2018) Heart enhancers with deeply conserved regulatory activity are established early in zebrafish development. *Nature Communications*. 9 (1), 4977. doi:10.1038/s41467-018-07451-z.

Zhang, Q., Carlin, D., Zhu, F., Cattaneo, P., Ideker, T., Evans, S.M., Bloomekatz, J. & Chi, N.C. (2021) Unveiling Complexity and Multipotentiality of Early Heart Fields. *Circulation Research*. 129 (4), 474–487. doi:10.1161/circresaha.121.318943.

Chapter 3: Dissecting a cardiac-specific enhancer sequence reveals motifs of candidate early cardiac regulators

Abstract

Evolutionarily conserved paradigms for phylogenesis and ontogenesis during development are supported by highly conserved gene expression patterns amongst diverse metazoan species, particularly during phylotypic stages of embryonic development. Many of these genes act in instructive roles for organ lineage specification and patterning, and are represented as orthologs across multiple species. Thus, the genetic mechanisms governing cell fate and function in organisms during phylotypic stages of development potentially represent critical core regulatory networks in lineage specification and organogenesis across metazoan species. Here, we leverage the 2.5 kb sequence of the cardiac-restricted enhancer, *Smarcd3*-F6, as a discovery platform for the characterization of candidate transcriptional regulators in early mouse cardiogenesis. We found that the *Smarcd3*-F6 enhancer sequence contains discrete regions of high conservation, as well transcription factor (TF) footprints and binding motifs for multiple cardiac mesoderm regulatory factors, such as T, Eomes, Nkx2-5, Tbx5, Hand2, Gata4/5/6. From the basis of CPC-restricted early enhancer activity, sequence conservation, and diverse TF motif hubs, we questioned if the *Smarcd3*-F6 enhancer may contain discrete enhancer sequences either with restricted activity to sub-anatomy within the heart reflective, or else representative of the minimally sufficient enhancer fragment for pan-cardiac *Smarcd3*-F6 activity. We hypothesized that regulatory signals acting within such a minimized enhancer sequences would represent key early cardiogenic transcriptional regulators. To characterize the factors activating *Smarcd3*-F6 in CPCs, we systematically dissected the sequence and screened for enhancer activity using the enSERT system for high-throughput, *in vivo* enhancer

activity screening. We characterized a minimally-sufficient 217 bp sub-fragment, mAK7, and two separate GATA and T-BOX motifs which appear critical for enhancer activity within developing CPCs. The results of this study provide the foundation for functional characterization of mechanisms governing transcriptional regulation, and resolving the temporal and spatial relationships of TFs and distal regulatory elements operating within complex transcriptional networks during cardiogenesis.

Background

The heart is a vital organ for life. Primarily acting as a pump to push blood into all organs, the heart is producing the pressure necessary for circulation even in mouse embryos without fully formed hearts (<E9.0). The importance of this central pump is supported by phylogenetic observations of ancient heart-like structures appearing in a diverse array of metazoan organisms, including arthropods, mollusks, and chordates (Yuan, Scott & Wilson, 2021). While final cardiac anatomy is diverse among metazoans, these differences arise later in development during morphogenesis and are likely reflective of the differing physiological demands for the cardiovascular systems of vastly different organisms. Despite these differences, many early cellular specification events appear highly conserved across vastly disparate metazoan species (Yuan, Scott & Wilson, 2021). The similarity in initial cardiac ontogeny is likewise correlated to shared molecular regulators, such as transcription factors (TFs), across different species. These observations point towards the existence of conserved, core regulatory networks with essential cardiogenic roles.

The challenge for studying these early regulatory networks is how early cardiogenesis begins during gastrulation, when embryos are quite small with morphologies lacking distinct sub-anatomic characteristics. Moreover, varied event orders within gastrulation, such as the species-specific sequences of uterine implantation and initiation of gastrulation events in mammals, renders capturing early molecular events of CPCs additionally difficult. Circumventing these problems towards this goal of characterizing early regulatory events requires a tractable tool, such as some sort of genomic sentinel beacon or marker within the emerging cardiac precursor cells (CPCs), influenced and activated by the same regulatory machinery in the same cells. We were curious if the characterized CPC-specific *Smarcd3*-F6 cardiac enhancer reporter (Devine et al., 2014) would enable such an investigation aiming to characterizeof the regulatory logic within very early cardiogenesis.

Within the *Mesp1* lineage, activity from the 2.5kb enhancer region "F6", located upstream of *Smarcd3*, labels the totality of CPCs prior to induction of other cardiac TFs, such as *Nkx2-5, Tbx5,* and *Isl1* (Devine et al., 2014). Enhancer activity begins shortly after *Mesp1* expression, and continues through gastrulation until looped heart stages, thus encompassing both initiation and early morphogenesis phases of cardiogenesis. Furthermore, *Smarcd3*-F6 activity and lineage contributions are confined heart cells (Devine et al., 2014), enabling clear distinction of early cardiac-fated cells from the surrounding mesoderm. Additionally, early cardiac progenitors within zebrafish show activity from mouse *Smarcd3*-F6 when injected as an EGFP reporter transgene (Yuan et al., 2018), indicating inter-species enhancer activity conservation. Thus, *Smarcd3*-F6's pan-cardiac activity and conservation across different species with vastly different

ultimate cardiac anatomies qualifies this enhancer sequence as a unique molecular tool towards understanding regulatory signals occurring within emerging CPCs.

Given that expression of *Smarcd3*-F6 expression precedes other early cardiac-specific TFs and *Smarcd3*-F6 activity is restricted to cardiac progenitors, we hypothesized that expression is initiated and restricted by some combination of early transcriptional regulators with yet uncharacterized roles in regulating early cardiac specification. Furthermore, we questioned if these early cardiac regulators might play instructive roles in cardiac specification. Additionally, since the *Smarcd3*-F6 expression pattern and lineage spans all regions of the developing heart, which are known to have diverse programs of transcriptional regulation, we hypothesized that the relatively large 2.5 kb *Smarcd3*-F6 enhancer region could contain discrete enhancer regions with potentially separate roles. These discrete sites within the “F6” region might potentially a) recapitulate anatomically restricted, partial *Smarcd3*-F6-expression, such as the FHF or SHF alone; or b) represent the minimally sufficient sequence(s) capable of recapitulating total *Smarcd3*-F6 expression, and these results would then represent a potentially previously uncharacterized, instructive early CPC regulatory program.

Results

Sequence analysis of *Smarcd3*-F6 illustrates evolutionary conservation and cardiac transcription factor binding motifs

Enhancer sequences are often conserved across multiple species, due in part to stereotyped transcription factor binding preferences to motifs, regardless of the genomic address (Chen, Fish & Capra, 2018). Thus, while genomic regions for enhancers and genes might be present on different chromosomes within different species, the sequences of those enhancers will potentially be quite conserved across organisms. In this manner, sequence conservation for an enhancer with known activity in one organism can be predictive of similar enhancer activity amongst species sharing that sequence. We extended this framework to hypothesize that multiple regions of conservation within the *Smarcd3*-F6 enhancer could indicate discrete enhancers with potentially distinct regulatory networks.

We interrogated these hypotheses by examining sequence conservation of the *Smarcd3*-F6 enhancer. We placed the *Smarcd3*-F6 sequence into UCSC Genome Browser (Fig. 3.1A) (Kent et al., 2002; Kent, 2002) and evaluated mammalian conservation with PhyloP base-wise conservation tracks (Fig. 3.1C,E) (Karolchik et al., 2004). Phylogenetically-spaced vertebrate species rat, human, orangutan, dog, horse, opossum, chicken, stickleback, zebrafish, frog, lizard, rhesus, chimp, guinea pig, pig, and turkey were evaluated with Multiz Alignment tracks and Non-mouse RefSeq gene tables turned on (Fig. 3.1E) (Karolchik et al., 2004). We further mapped the *Smarcd3*-F6 sequence into the ECR Browser (Ovcharenko et al., 2004), and noted similar blocks of

overlap as to those described in Fig. 3.1E. In total, we characterized 5 regions of high conservation, mapping mostly to human sequences (Fig. 3.1E, Fig. 3.2A).

Next, we investigated DNA-protein interactions to predict potential sites of TF binding within *Smarcd3*-F6. We saw no striking signal within *Smarcd3*-F6 for DNaseI foot-printing from ENCODE (Fig. 3.1D) (Thomas et al., 2007; Davis et al., 2017; Luo et al., 2019; Dunham et al., 2012), perhaps because the available data were poorly matched to our question; heart samples were from post-natal mice, and the embryonic sequence was from E14.5 brain samples. Thus, we turned to ChIP-exo data from *in vitro* cardiac differentiation of mouse embryonic stem cells (mESCs), finding that Tbx5, Nkx2-5, and Gata4 TF footprints within relevant cardiac progenitors (CPs) and cardiomyocytes (CMs) data stacked near the center of *Smarcd3*-F6 (Fig. 3.1B). Strikingly, footprints for all 3 TFs in both cell types stacked up in relatively the same location as a region of high sequence conservation with humans (Fig. 3.1B,C,E), indicating a potential, smaller area of regulatory activity within the middle of *Smarcd3*-F6.

Following this coarse annotation of species conservation and cardiac TF binding footprints, we more finely and unbiasedly queried motif presence within the total *Smarcd3*-F6 enhancer sequence. We applied the JASPAR Core motif set (Khan et al., 2017; Castro-Mondragon et al., 2021) to the full *Smarcd3*-F6 sequence, and compared the motif outputs to those generated with the MEME suite for motif discovery (Bailey et al., 2015), finding largely the same motifs annotated in the same sequence locations. To visualize potential patterns such as motif neighborhoods of known binding co-factors, we annotated these motifs within the *Smarcd3*-F6 sequence (Fig. S3.1). We found an

abundance of cardiac and mesoderm-relevant GATA, CDX, TBOX, FOX, HAND, and MEIS factors, as well a few non-cardiac motifs such as Neurod (Fig. S3.1). However, the main pattern resolved showed that regions of high sequence conservation were also 'hot spots' for detection of multiple cardiac and cardiac mesoderm motifs (Fig. S3.1). This observation may be biologically accurate, however could also potentially or partially represent a slight human bias of the datasets we used to annotate motif sequences; both the JASPAR and MEME databases contain enriched representation of human motifs annotations. Importantly, we detected an abundance of early gastrulation, mesoderm, and cardiac TF binding motifs within the *Smarcd3*-F6 sequence. While no motifs seemed to signify novel candidate cardiogenic regulatory factors, we were curious to 1) illuminate potentially uncharacterized roles in regulatory networks and 2) interrogate the potential for anatomically-restricted enhancer activity within the enhancer or potential cellular heterogeneity from fragmented enhancer activity.

From these collective analyses, we mapped a strategy to subdivide *Smarcd3*-F6 (mAK1) into sub-fragments for *in vivo* evaluation of enhancer activity (Fig. 3.1C, Fig. S3.1). The initial subdivision strategy was to test halves mAK2 and mAK3 first, and if those yielded positive enhancer activity we would devise anoter strategy to further subdivide them. For the sake of presentation, we have mapped all tested subdivisions in this same figure (Fig. 3.1C), and will explain the rationale behind mAK4-9 in the subsequent results section.

In vivo dissection of *Smarcd3*-F6 describes minimal cardiac-specific enhancer element

Following annotation of the enhancer sequence (Fig. 3.1, S3.1), we began our screen as previously by halving the enhancer sequence, taking care to avoid bifurcating any annotated motifs. To circumvent the low efficacy of positional uncertainty and random transgenic insertion events that often require large numbers of embryos and careful genotyping strategies, we turned to the enSERT system (Kvon et al., 2020) to screen our enhancer sub-fragments. The enSERT system poses several advantages for rapid F_0 *in vivo* screens of enhancer activity. 1) EnSERT makes use of CRISPR/Cas9-mediated targeted integration to insert enhancer-reporter transgenes into the safe harbor *Hipp11* locus, promoting both certainty in the number of genomic integration events as well as minimizing the effects of confounding regulatory topographies on levels of enhancer activity reporter read out (Kvon et al., 2020, 2016). 2) The enSERT transgene construct reports enhancer activity via LacZ driven by a minimal *Shh* promoter, which has been to shown to exhibit lower non-specific background activity relative to *Hsp68*::LacZ reporter constructs genome-wide (Kvon et al., 2020). 3) The enSERT donor plasmid can be delivered within a standard CRISPR/Cas9 targeting mastermix via pronuclear injection to standard FVB embryos, circumventing the specialized tools and recipient mice that are necessary in other site-specific enhancer transgene knock-in systems like TARGATT (Applied Stem Cell).

As we were utilizing the new enSERT system, we also generated a control construct, mAK1, which included the totality of the original *Smarcd3*-F6 enhancer sequence, only

now within the new enSERT-*Shh-lacZ* reporter plasmid. Since one of our initial questions was to query the possibility of anatomically-restricted enhancer activity within specific regions of the heart, we performed our initial screen in E9.5 embryos because embryonic hearts in this stage have a distinct looped patterning with discernable heart fields and structures. We found the enSERT-system read out for *Smarcd3*-F6 enhancer activity, via mAK1 (Fig. 3.2A-B), was identical to our previous accounts (Devine et al., 2014) and in Ch. 2 of this dissertation. Thus, we validated expected *Smarcd3*-F6 activity within the enSERT-system and proceeded with our screen. We performed our initial test screen to evaluate the 2 fragments halving the original enhancer, 1267 bp mAK2, and 1517 bp mAK3 (Fig. 3.2A). We found positive activity mirroring the restricted pan-cardiac activity of mAK1/*Smarcd3*-F6 with our results for mAK2 (Fig. 3.2C), and zero activity for mAK3 (Fig. 3.2E) in E9.5 embryos. To rule out the possibility for temporally distinct activity reflecting potential developmental requirements for regulation from distinct networks, we tested both mAK2 and mAK3 at E7.5, when the cardiac crescent is forming but morphogenesis into heart fields hasn't yet occurred. We found activity in the lateral plate mesoderm (LPM) of mAK2 E7.0 embryos (Fig. 3.2D), which appears similar to patterns of *Smarcd3*-F6 (Fig. 2.1) (Devine et al., 2014), and we detected no activity from mAK3 in E7.0 embryos.

We returned to our enhancer sequence annotations for motif, footprint, and conservation (Fig. 3.1, Fig. S3.1), noting that the spaced divide of 2 conservation regions (Fig. 3.2A) overlapping with the majority of detected motif sites, including GATA, SOX, FOX, T-BOX, SMAD, and MEIS motifs (Fig. S3.1), appeared to naturally divide mAK2 into thirds; 417 bp mAK4, 458 bp mAK5, 400 bp mAK6 (Fig. 3.2A, Fig. S3.1).

While mAK5 only contained a few motif annotations, aspects of the sequence overlapped with a highly conserved region (Fig. S3.1) so we still saw value in evaluating the fragment's activity. We evaluated enhancer activity for all 3 fragments at E9.5, however only detected activity from mAK4, which was again pan-cardiac (Fig. 3.2F, Table S3.1). Notably, the mAK4 fragment's endogenous genomic sequence position overlaps with the Tbx5, Nkx2-5, and Gata4 ChIP-exo footprints in CPs and CMs (Fig. 3.1B-C). Collectively, these initial results revealed that the 417 bp mAK4 fragment recapitulated all domains for *Smarcd3*-F6 activity, and no other fragments showed activity. We detected activity for mAK4 in the LPM of E7.5 embryos (Fig. 3.2G), which expectedly mirrors the domains of full *Smarcd3*-F6 activity (Ch.2) (Devine et al., 2014). In conclusion, while these findings provided no strong basis for multiple discrete *Smarcd3*-F6 sub-enhancers with anatomically restricted activity in the heart, there could still be a minimal enhancer sequence, either the 417 bp mAK4 or potentially even smaller. Thus, we focused on efforts on our remaining model that if present, a minimal enhancer sequence within *Smarcd3*-F6 would contain motifs for regulatory factors with potentially uncharacterized roles or uncharacterized participation in early regulatory networks during very early cardiogenesis.

We further subdivided mAK4 into the 217 bp mAK7, 200 bp mAK8, and a 157 bp mAK9 which contains a concentration of GATA binding motifs and overlaps both mAK7 and mAK8 (Fig. 3.2A, Fig. S3.1). Having forgone the discrete-anatomy enhancers hypothesis, we harvested embryos only at the younger pre-crescent stages prior to organ patterning. We characterized activity for mAK7 in the LPM of late E7.5 embryos (Fig. 3.2H), and no activity from either mAK8 (Fig. 3.2I) nor mAK9 (Fig. 3.2K, Table

S3.1). Since *Smarcd3*-F6 showed activity in mESCs that became depleted between days 0-4 during *in vitro* cardiac differentiation, only to reappear rapidly on day 4.5 (Devine et al., 2014), we asked if mAK8 or mAK9 might represent temporally restricted pluripotent enhancers. We evaluated both enhancer fragments at younger stages of E6.5 and E7.0 (Fig. 3.2J,L) but again found no activity. These results suggest that the 217 bp mAK7 fragment represents a minimally sufficient enhancer sequence containing motifs for candidate, critical early cardiac regulators.

Mutagenesis of distinct T-Box and GATA transcription factor binding motifs abolishes enhancer activity

Given the highly minimized nature of the mAK7 enhancer sequence, we gauged that site directed mutagenesis (SDM) of specific motif regions would provide higher resolution insights into key regulatory factors than further fragmentation of the sequence would. Thus, we focused in on the region of the mAK7 sequence which did not overlap with the negative mAK9 fragment (Fig. 3.3A). We noted an abundance of GATA and TBOX motifs (Fig. 3.3A), which are bound by TFs such as Gata4/5/6 and T, Tbx5, Eomes, respectively, in the early cardiac mesoderm, LPM, and CP stages of cardiogenesis (Kokkinopoulos et al., 2015; Costello et al., 2011; Probst et al., 2020; Tosic et al., 2019; Song et al., 2022; Irie & Kuratani, 2011). The known regulatory roles of these TFs are consistent with both the temporal and spatial activity for the mAK7 enhancer, as well as the parent fragment mAK4, within the LPM of E7.5 embryos and pan-cardiac activity within E9.5 embryos (Fig. 3.2F-H). We chose to perform SDM on one GATA motif near the border of mAK9, within the Gata4 ChIP-exo footprint (yellow

box, Fig. 3.3A) and one other SDM on the TBOX motif within the Nkx2-5 ChIP-exo footprint (yellow box, Fig. 3.3A). For each region, we deleted the totality of motif nucleotides. We evaluated transgenic mAK7-SDM embryos at ~E7.0, and observed no enhancer activity within embryos containing the mutated TBOX site (mAK7-SDM-TBOX, Fig. 3.3B, Table S3.1) and no activity within the embryos containing the mutated GATA site (mAK7-SDM-GATA, Fig. 3.3C, Table S3.1). We speculated that these sequences were bound by critical GATA and TBOX positive regulators, and thus ablation of binding motifs resulted in the observed loss of activity from the reporter signal. However, equally or additionally likely is the scenario where these deletions disrupted combinatorial binding of multiple TF regulators and their co-factors to act on this enhancer. Further analysis via point mutations or enhancer binding experiments like luciferase or EMSA would be required to differentiate these possibilities. Thus, we conclude that the 217 bp mAK7 represents our minimally sufficient enhancer sequence, and recapitulates *Smarcd3*-F6 activity in the emerging cardiac lineage. Furthermore, regulation around two defined GATA and T-box motif regions appears to be critical for enhancer activity.

Discussion

The developmental hourglass model suggests a phylotypic stage when species exhibit high morphological similarity and high conservation of gene expression patterns for the phylogenesis and ontogenesis promoting allocation of diverse cell lineages during mid-embryogenesis (Hanken & Carl, 1996; Domazet-Lošo & Tautz, 2010; Irie & Kuratani, 2014; Kalinka et al., 2010; Panigrahi & O'Malley, 2021). Initiation of mouse

cardiogenesis coincides with gastrulation during the phylotypic stage (Irie & Sehara-Fujisawa, 2007; Irie & Kuratani, 2011), and is coordinated through the regulatory activity of highly conserved cardiac TFs binding to *cis*-regulatory elements including promoters and distal enhancers (Spitz & Furlong, 2012; Levine & Davidson, 2005). The high conservation among putative mesoderm enhancers in early cardiogenic stages of development (Nord et al., 2013), coupled with the conserved functions of early, critical cardiac TFs prior to onset of *Nkx2-5* expression (Devine et al., 2014; Lescroart et al., 2018, 2014; Ivanovitch et al., 2021) implies that evaluation of regulatory relationships between TFs and *cis*-regulatory elements during early cardiogenesis will reveal critical or indispensable regulatory networks governing cardiac specification.

To interrogate this hypothesis, we utilized the early CPC-specific, pan-cardiac *Smarcd3*-F6 enhancer sequence as an experimental platform for discovery of the early regulatory signals acting within these cells as they emerge from the mesoderm. Comparative genomic analyses of the enhancer sequence revealed multiple discrete regions of high conservation to multiple metazoan species, with regions for the highest conservation matching to humans (Fig. 3.1). This characterization suggests that 1) *Smarcd3*-F6 activity is governed by highly conserved and potentially critical regulatory networks acting within CPCs, and 2) discrete sub-enhancers with activity restricted to sub-anatomy of the heterogeneous heart could potentially be present.

Annotation of the *Smarcd3*-F6 enhancer sequence revealed overlap with ChIP-exo footprints for critical cardiac TFs (Tbx5, Gata4, Nkx2-5), and binding motifs for additional, even earlier cardiac mesoderm-specific gastrulation TFs within these

conserved regions and throughout the full enhancer sequence (Fig. S3.1). The presence of these gastrulation, mesoderm, and cardiac TFs is consistent with the periods of embryonic development wherein *Smarcd3*-F6 is active from E6.5 through looped heart stages (E9.5).

Successive rounds of enhancer sequence dissection utilizing the enSERT *in vivo* high throughput mouse reporter assay revealed that the 217 bp mAK7 fragment represents a minimally sufficient sequence within *Smarcd3*-F6 for enhancer activity in early CPCs (Fig. 3.2 H). Separate instances of site directed mutagenesis to distinct GATA and TBOX binding motifs ablated mAK7 enhancer activity equally, suggesting disruption of critical regulatory networks containing factors or TFs normally bound to those motifs. These results implicate mesendoderm TFs such as Gata4/5/6, T, and Eomes as candidate regulators with potentially instructive roles during CPC allocation from the mesoderm at the initiation of cardiac lineage specification. Further investigation will be necessary to understand TF binding dynamics for these candidates, and the associated potential recruitment of binding co-factors, facilitation of signaling molecules or gradient interpretation, and participation within specific regulatory networks.

The results revealed over the course of this study led to us re-evaluating our initial hypotheses. When we found no evidence supporting the idea of sub-anatomic enhancers revealing early heterogeneity for the patterning of regionalized regulatory networks, we re-focused our investigation to specifically evaluate early regulatory signals in pre-crescent embryos. However, we didn't validate expression of mAK7 at looped heart stages and such an assessment would be edifying towards achieving

temporal resolution of regulatory network activities during cardiogenesis. We haven't yet investigated potential interdependent roles for candidate regulators and co-factors binding emerging cis-regulatory elements in a dynamically developing genomic landscape, and the mAK7 fragment and/or the larger mAK4 parent fragment could reveal cooperative effects within the *Smarcd3*-F6 enhancer at organogenesis stages succeeding specification beyond the cardiac crescent. We can further assess the lineage contributions of the minimal mAK7 enhancer, resources permitting, however finding an identical lineage renders mAK7 no more useful than our current *Smarcd3-F6-Cre* lineage driving allele and would only be informative if a different lineage contribution was defined by mAK7.

Additionally, the functional utility of *Smarcd3*-adjacent distal elements has not yet been evaluated via deletion or disruption of the endogenous locus. Stemming largely from genomic proximity and overlapping domains of gene expression and enhancer activity (Devine et al., 2014), the *Smarcd3*-F6 enhancer is presumed to have a distal regulatory role on the nearby *Smarcd3* gene, but the function of this enhancer has not been formally tested and such a study could yield insights into evolved redundancy or compensatory control over critical and finely-regulated transcriptional processes. Indeed, while shRNA-mediated knockdown of *Smarcd3* in mice and zebrafish produces organ patterning defects during heart formation (Takeuchi et al., 2007; Lickert et al., 2004), the constitutive genetic deletion of *Smarcd3* exhibits a milder phenotype for hypoplastic embryonic hearts and disrupted CM function in mice, (Sun et al., 2017). The discrepancy between the phenotypes could stem from loss of *Smarcd3*/Baf60c function (a chromatin remodeler), caveats stemming from high levels of shRNA exposure

disrupting transcriptional processing machinery, or failed induction of compensatory mechanisms with the rapid repression kinetics of shRNA treatment (Sun et al., 2017). Thus, while loss of Baf60c via knockout of *Smarcd3* appears to be minimally disruptive to cardiac development and physiology (Sun et al., 2017), the unknown regulatory role the nearby *Smarcd3*-F6 enhancer may exert on this locus or other potential targets remains to be investigated through knockout of the endogenous enhancer locus, and may serve to generate further clarity on compensatory regulatory mechanisms involving *Smarcd3*/Baf60c. Towards this goal, future studies would include endogenous deletions of the full *Smarcd3*-F6 enhancer, the mAK7 minimally sufficient enhancer, and disruption to the identified GATA and T-box motif regions, or yet to-be-defined point mutations within these regions.

In conclusion, we leveraged dissection of the *Smarcd3*-F6 enhancer sequence to interrogate candidate regulatory signals potentially acting in temporally and spatially restricted manners within CPCs of the developing heart. We found a minimal sequence with TF motif regions critical for enhancer activity in CPCs of the LPM at pre-crescent embryo stages. This study coordinates hypotheses of evolutionarily conserved regulatory logic during phylotypic stages of embryonic development to inform the design of a high-throughput, *in vivo* enhancer activity screen towards the goal of characterizing minimally-sufficient networks for lineage specification and organogenesis of the heart. The framework laid forth in this study can be extended to functional evaluation of other enhancer variants' phenotypic impacts, potentially enabling functional descriptions of regulatory roles for non-coding elements in development and disease.

Figures

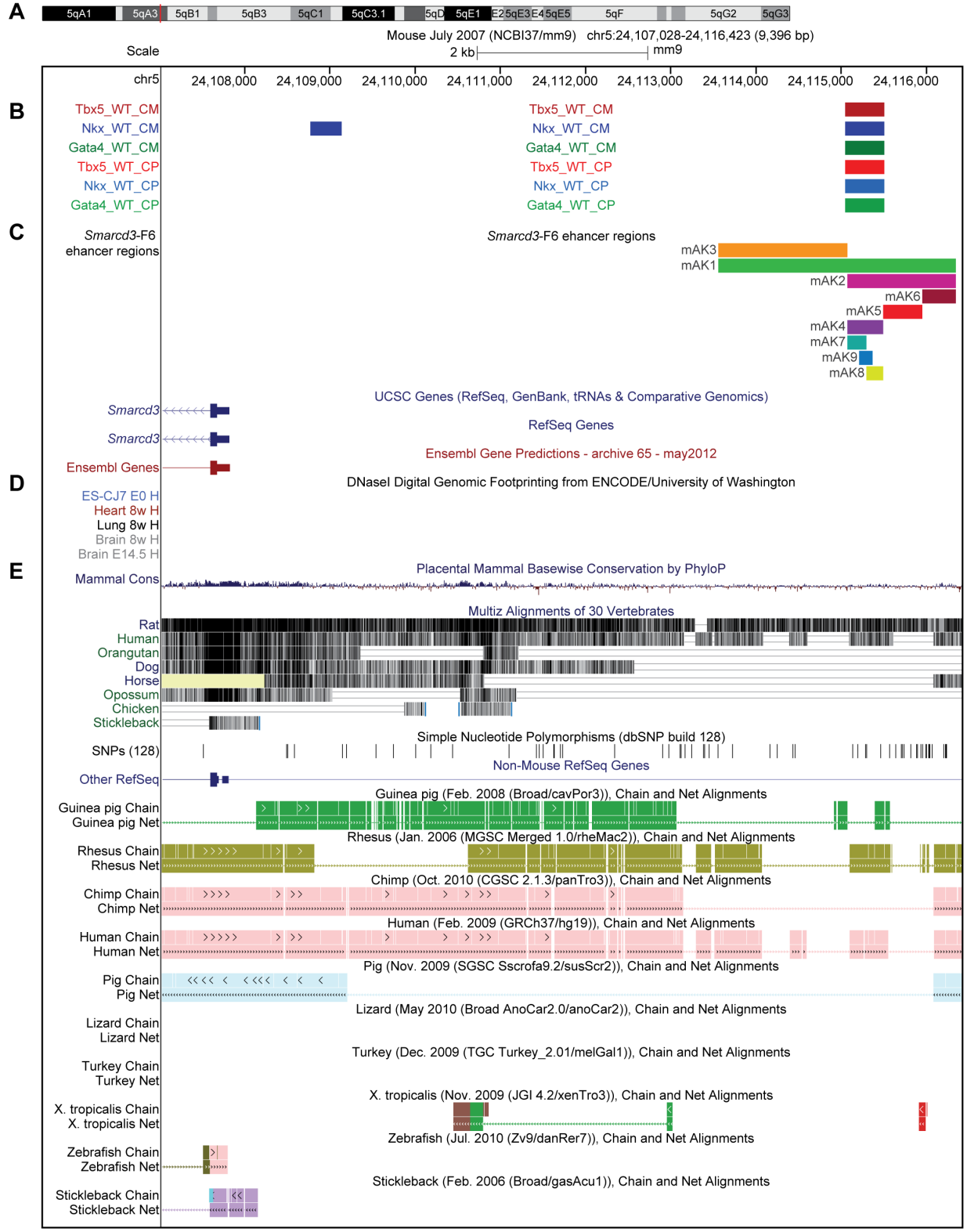

A
5qA1
5qA3
5qB1
5qB3
5qC1
5qC3.1
5qD
5qE1
E2
5qE3
E4
5qE5
5qF
5qG2
5qG3
Mouse July 2007 (NCBI37/mm9) chr5:24,107,028-24,116,423 (9,396 bp)
Scale
2 kb
mm9
chr5
24,108,000
24,109,000
24,110,000
24,111,000
24,112,000
24,113,000
24,114,000
24,115,000
24,116,000
B
Tbx5_WT_CM
Nkx_WT_CM
Gata4_WT_CM
Tbx5_WT_CP
Nkx_WT_CP
Gata4_WT_CP
C
Smarcd3-F6 ehancer regions
Smarcd3-F6 ehancer regions
mAK3
mAK1
mAK2
mAK6
mAK5
mAK4
mAK7
mAK9
mAK8
UCSC Genes (RefSeq, GenBank, tRNAs & Comparative Genomics)
Smarcd3
RefSeq Genes
Smarcd3
Ensembl Gene Predictions - archive 65 - may2012
Ensembl Genes
D
DNaseI Digital Genomic Footprinting from ENCODE/University of Washington
ES-CJ7 E0 H
Heart 8w H
Lung 8w H
Brain 8w H
Brain E14.5 H
E
Placental Mammal Basewise Conservation by PhyloP
Mammal Cons
Multiz Alignments of 30 Vertebrates
Rat
Human
Orangutan
Dog
Horse
Opossum
Chicken
Stickleback
Simple Nucleotide Polymorphisms (dbSNP build 128)
SNPs (128)
Non-Mouse RefSeq Genes
Other RefSeq
Guinea pig (Feb. 2008 (Broad/cavPor3)), Chain and Net Alignments
Guinea pig Chain
Guinea pig Net
Rhesus (Jan. 2006 (MGSC Merged 1.0/rheMac2)), Chain and Net Alignments
Rhesus Chain
Rhesus Net
Chimp (Oct. 2010 (CGSC 2.1.3/panTro3)), Chain and Net Alignments
Chimp Chain
Chimp Net
Human (Feb. 2009 (GRCh37/hg19)), Chain and Net Alignments
Human Chain
Human Net
Pig (Nov. 2009 (SGSC Sscrofa9.2/susScr2)), Chain and Net Alignments
Pig Chain
Pig Net
Lizard (May 2010 (Broad AnoCar2.0/anoCar2)), Chain and Net Alignments
Lizard Chain
Lizard Net
Turkey (Dec. 2009 (TGC Turkey_2.01/melGal1)), Chain and Net Alignments
Turkey Chain
Turkey Net
X. tropicalis (Nov. 2009 (JGI 4.2/xenTro3)), Chain and Net Alignments
X. tropicalis Chain
X. tropicalis Net
Zebrafish (Jul. 2010 (Zv9/danRer7)), Chain and Net Alignments
Zebrafish Chain
Zebrafish Net
Stickleback (Feb. 2006 (Broad/gasAcu1)), Chain and Net Alignments
Stickleback Chain
Stickleback Net

Figure 3.1. Genomic sequence analysis of *Smarcd3*-F6. (A) UCSC genome browser window chromosome 5 area around *Smarcd3* gene locus. (B) ChIP-exo genomic occupancy of TBX5, NKX2-5, and GATA4 in mouse *in vitro* differentiation CP and CM cells. (C) Alignment of *Smarcd3*-F6 enhancer (mAK1) and sub-fragments mAK2-9. (D) DNaseI digital genomic foot-printing from ENCODE. (E) Evolutionary conservation of Smarcd3 upstream genome sequences.

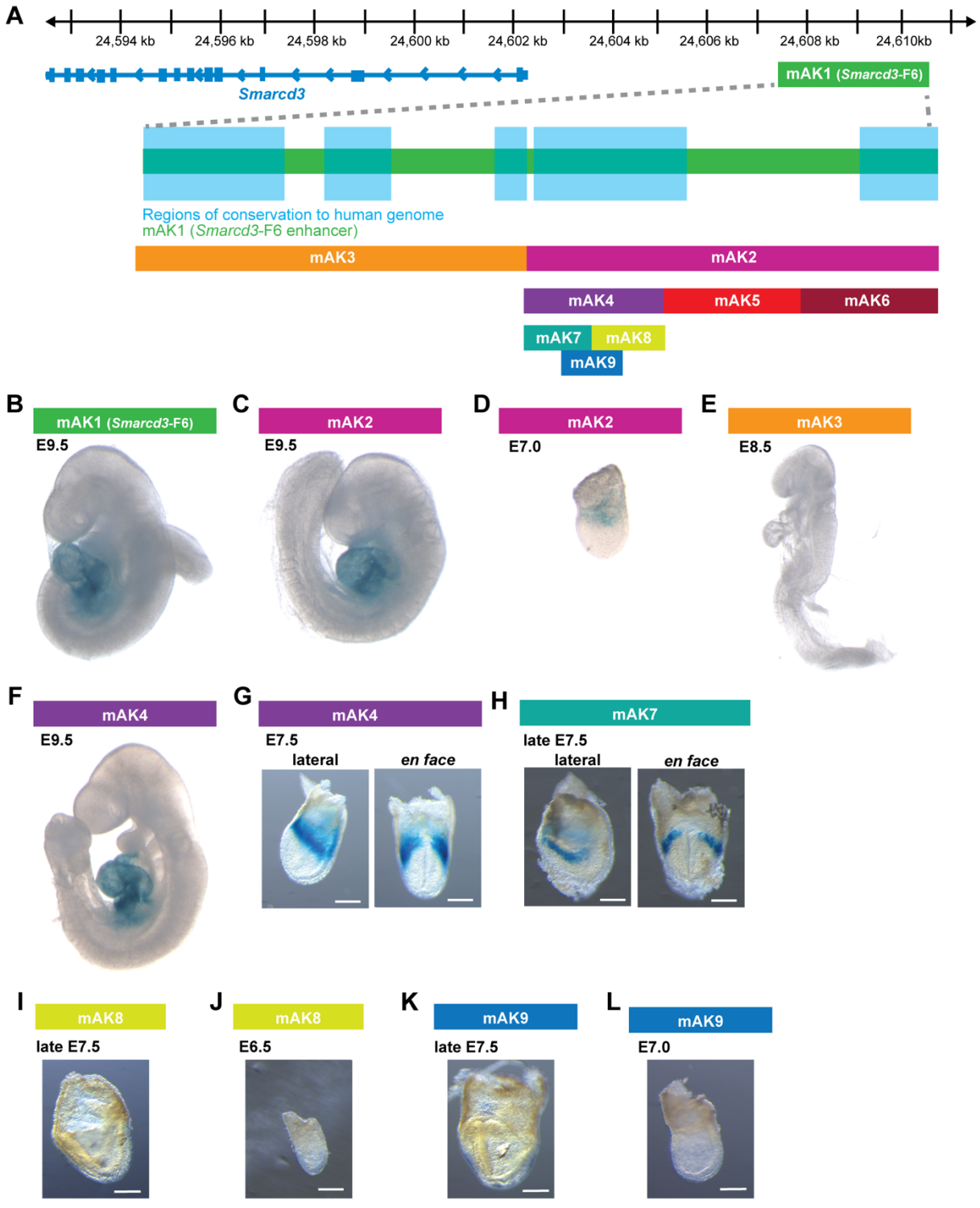

A
24,594 kb
24,596 kb
24,598 kb
24,600 kb
24,602 kb
24,604 kb
24,606 kb
24,608 kb
24,610kb
Smarcd3
mAK1 (Smarcd3-F6)
Regions of conservation to human genome
mAK1 (Smarcd3-F6 enhancer)
mAK3
mAK2
mAK4
mAK5
mAK6
mAK7
mAK8
mAK9
B
mAK1 (Smarcd3-F6)
E9.5
C
mAK2
E9.5
D
mAK2
E7.0
E
mAK3
E8.5
F
mAK4
E9.5
G
mAK4
E7.5
lateral
en face
H
mAK7
late E7.5
lateral
en face
I
mAK8
late E7.5
J
mAK8
E6.5
K
mAK9
late E7.5
L
mAK9
E7.0

Figure 3.2. A minimal cardiac specific enhancer within *Smarcd3*-F6. (A) Cartoon of conserved regions and *Smarcd3*-F6 enSERT fragments within the original enhancer sequence upstream of *Smarcd3*. (B-L) X-gal stains of enhancer activity within F_0 transgenic embryos. (B) Pan-cardiac mAK1 activity in E9.5 embryo (C) Pan-cardiac mAK2 activity in E9.5 embryos. (D) mAK2 activity within LPM in E7.0 embryos. (E) No activity from mAK3 in E9.5 embryos. (F) Pan-cardiac mAK4 activity in E9.5 embryos. (G) mAK4 activity within the LPM in E7.5 embryos. (H) mAK7 activity within LPM in E7.5 embryos. (I) No activity for mAK8 in E7.5 nor (J) E6.5 embryos. (K) No activity for mAK9 in E7.5 nor (L) E7.0 embryos. Scale bars are 200 μM.

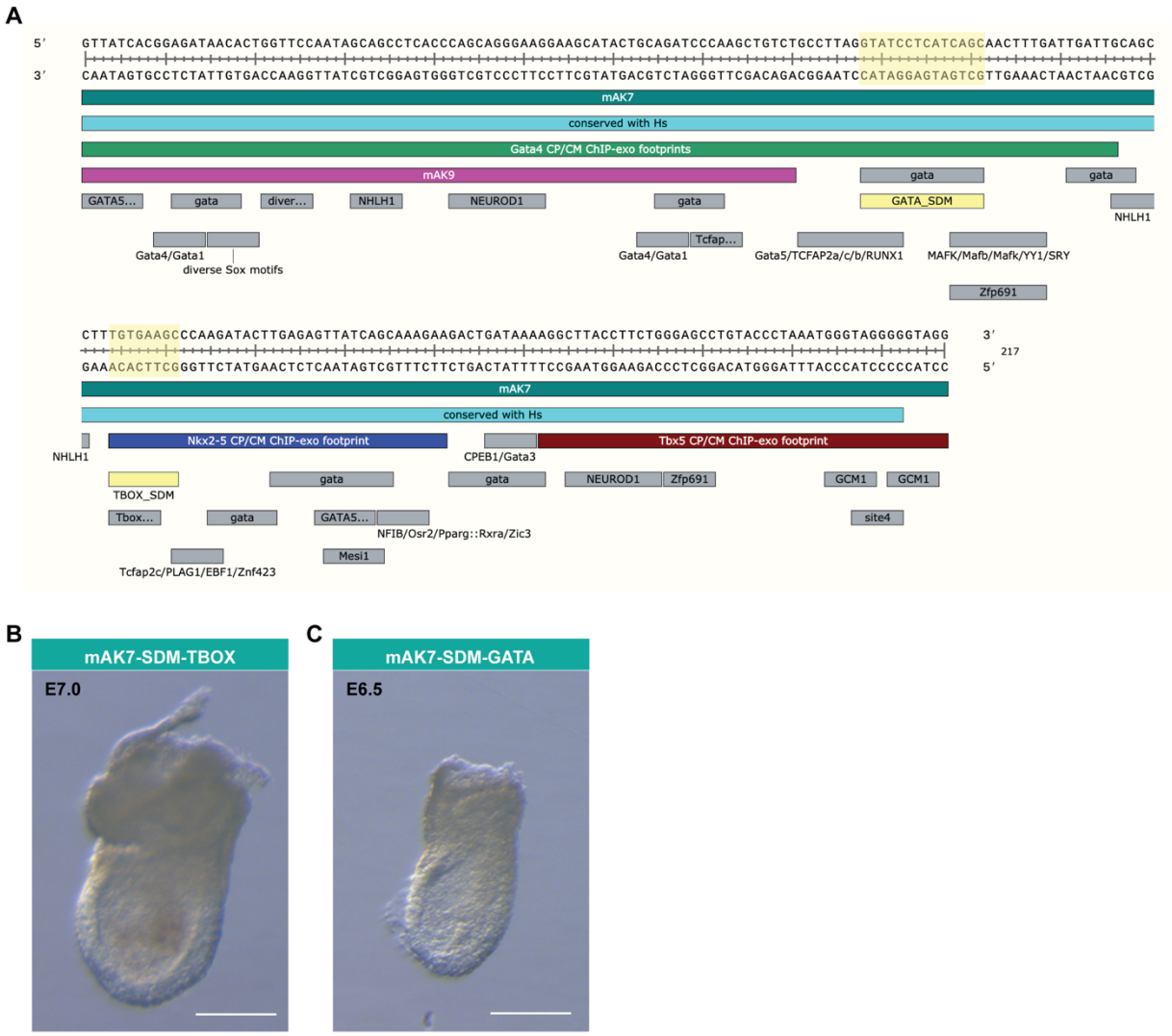
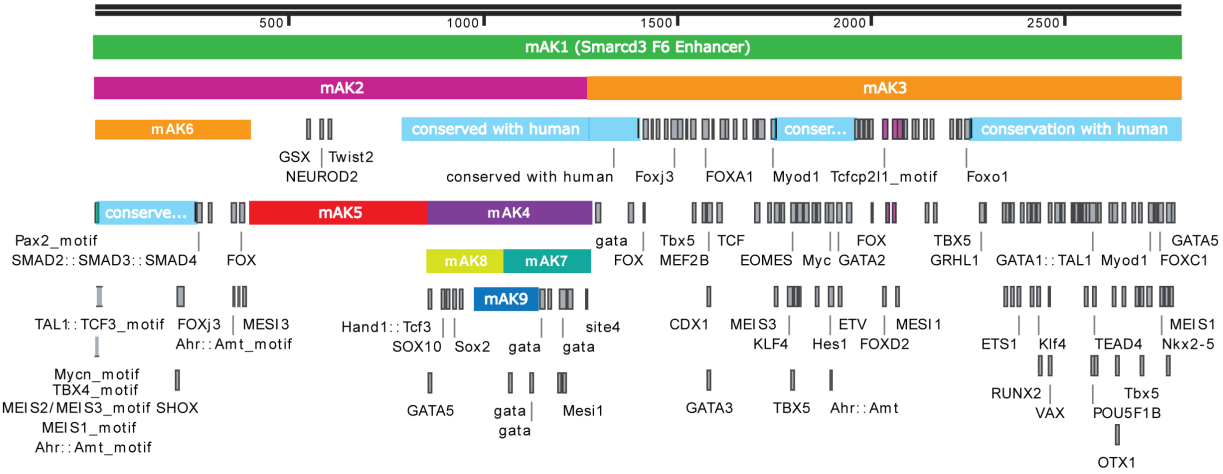


Figure 3.3. Mutagenesis of T-box and GATA motifs in minimal enhancer sequence. (A) Nucleotide sequence for the mAK7 enhancer, a sub-fragment of *Smarcd3*-F6. Human-conserved regions denoted in blue. CP and CM ChIP-exo footprints for Gata4 (green), Nkx2-5 (blue), and Tbx5 (red). mAK7 fragment (teal) contains a portion of the mAK9 fragment (violet). Motif annotations from JASPAR and MEME denoted with grey boxes and labeled for motif recognized by databases. Yellow boxes denoting motif regions ablated during SDM. (B) Negative activity in E7.0 embryo for the mAK7-SDM-TBOX enhancer. (C) Negative activity in E6.5 embryo for the mAK7-SDM-GATA enhancer. Scale bars are 200 μM.



Supplementary Figure 3.1. Motif annotation of *Smarcd3*-F6 sequence. Motif annotations within full the *Smarcd3*-F6 enhancer sequence (mAK1) and sub-fragments mAK2-9. Sequence conservation to *Homo sapiens* overlayed in light blue, and motifs annotated (grey) throughout.

Supplementary Table 3.1. Summary of transgenic embryo results. Table of assayed enhancer fragments, detection of enhancer activity through LacZ, embryonic stages, and genotypes for transgenic embryos.

Enhancer	ACTIVITY	Stage	LacZ activity	Total Transgenics	Tandem Knock-in	Single Knock-in	Random Knock-in
mAK1	**ACTIVE**	**E9.5**	5	6	4	0	2
mAK2	**ACTIVE**	**E9.5**	6	7	6	0	1
mAK3	NEGATIVE	E9.5	0	6	5	1	0
mAK4	**ACTIVE**	**E9.5**	8	8	7	0	1
mAK5	NEGATIVE	E9.5	0	5	2	2	1
mAK6	NEGATIVE	E9.5	0	4	1	2	1
mAK2	**ACTIVE**	**E7.5**	6	9	3	5	1
mAK3	NEGATIVE	E7.5	0	6	not known	not known	not known
mAK4	**ACTIVE**	**E7.5**	4	6	4	1	1
mAK7	**ACTIVE**	**E7.5**	3	5	2	0	3
mAK8	NEGATIVE	E7.5	0	5	1	1	3
mAK9	NEGATIVE	E7.5	0	12	3	4	5
mAK7_SDM_TBOX	NEGATIVE	E7.5	0	3	1	0	2
mAK7_SDM_GATA	NEGATIVE	E7.5	0	3	1	0	1

Materials and Methods

Mouse models

Animal studies were performed both at Lawrence Berkeley National Labs (LBNL) and within the Gladstone Institutes. At LBNL, all animal work was approved by the Lawrence Berkeley National Laboratory Animal Welfare and Research Committee. Mice were housed at the Animal Care Facility (ACF) of LBNL, and all procedures were performed in strict compliance with the LBNL Institutional Animal Care and Use Committee. At the Gladstone Institutes, all animal studies were performed in strict compliance with the UCSF Institutional Animal Care and Use Committee. Mice were housed in a standard 12 hour light/dark animal husbandry barrier facility at the Gladstone Institutes. Transgenic mouse assays for enhancer knock-in were performed in FVB strain mice (Jackson Laboratory strain #001800). CD-1 strain females were used as surrogate mothers (Jackson Laboratory strain #003814).

Conservation annotation of *Smarcd3*-F6 enhancer sequence

The *Smarcd3*-F6 fragment (mAK1) was aligned to the mm10 genome within UCSC Genome Browser (Kent et al., 2002; Kent, 2002). Mammalian conservation was determine using PhyloP conservation tracks (Karolchik et al., 2004). Phylogenetically-spaced vertebrate species rat, human, orangutan, dog, horse, opossum, chicken, stickleback, zebrafish, frog, lizard, rhesus, chimp, guinea pig, pig, and turkey were evaluated using Multiz Alignment tracks and Non-mouse RefSeq gene tables (Karolchik et al., 2004). The ECR Browser (Ovcharenko et al., 2004) was utilized as a secondary

check of conservation. Regions of high conservation were annotated within the *Smarcd3*-F6 fragment file using SnapGene.

Annotating protein-binding within the *Smarcd3*-F6 enhancer sequence

We aligned *Smarcd3*-F6 (mAK1) to ENCODE DNaseI footprinting tracks in the UCSC Genome Browser (Thomas et al., 2007; Davis et al., 2017; Luo et al., 2019; Dunham et al., 2012; Kent et al., 2002). We aligned *Smarcd3*-F6 (mAK1) to ChIP-exo sequencing data for TF binding footprints of Gata4, Tbx5, and Nkx2-5 in wildtype cardiac progenitor and cardiomyocyte data from *in vitro* differentiation mouse embryonic stem cell data (Luna-Zurita et al., 2016).

Motif annotation of *Smarcd3*-F6 enhancer sequence

The full *Smarcd3*-F6 enhancer sequence (mAK1) was aligned against the JASPAR Core database (Khan et al., 2017; Castro-Mondragon et al., 2021) and the MEME suite for motif discovery (Bailey et al., 2015) and resulting motifs were annotated on the enhancer sequence in SnapGene. The mAK4 and mAK7 sub-fragments were further annotated again by the same methods in order to determine further sequence dissection and where to direct site directed mutagenesis, respectively.

Generation of ENSERT constructs

Transgenic donor plasmids were generated using the enSERT knock-in system (Kvon et al., 2020). Donor plasmid backbone vector PCR4-*Shh*::lacZ-H11 (Addgene 139098) for site-specific integration of enhancer reporter transgenes was linearized and purified

following Not1 digestion (NEB R0189S). Enhancer fragments were PCR isolated or geneblocks (IDT), and assembled into enSERT backbone using Gibson assembly (NEB E2611S) to create the transgenic donor knock-in plasmids. Final purified constructs were resuspended in water for microinjection.

Fragments mAK1-3 of the *Smarcd3*-F6 enhancer were PCR isolated from mouse genomic DNA using the following primers at LBNL:

mAK1-F1: cacgtgtccagctgtaatccag

mAK1-R1: atgcccctcctatctgtcctgc

mAK2-F1: cacgtgtccagctgtaatccag

mAK2-R1: cctacccccctacccatttaggg

mAK3-F1: gggagtggtggagagagagtgt

mAK3-R1: atgcccctcctatctgtcctgc

mAK4-F1: agatctggcatgcagccctggtc

mAK4-R1: cctacccccctacccatttaggg

mAK5-F1: atgtatgtttgtggatttggtc

mAK5-R1: cctgatatggatggaggtccttg

mAK6-F1: cacgtgtccagctgtaatccag

mAK6-R1: gtacaccatacatagacatgca

Fragments mAK4, mAK7-9 of the *Smarcd3*-F6 enhancer were PCR isolated from mouse genomic DNA using the following primers at Gladstone:

mAK4-F1: ctacttactcttcaggctgaagctgatggaacagcggccgcagatctggcatgcagccc

mAK4-R1: aagagagaaagaaaggctgctcagtttggatgttcctggccctacccccctacccatt

mAK7-F1: tactcttcaggctgaagctgatggaacagcgttatcacggagataacactggttccaata

mAK7-R1: aagagagaaagaaaggctgctcagtttggatgttcctgcctacccccctacccatttagg

mAK8-F1: gagaactacttactcttcaggctgaagctgatggaacaagatctggcatgcagccct

mAK8-R1: aagaaaggctgctcagtttggatgttcctggcacacagcttcatttgcacaggactcc

mAK9-F1: gaactacttactcttcaggctgaagctgatggaacagctgtaaggagaaggcaggtcc

mAK9-R1: aagaaaggctgctcagtttggatgttcctggcagacagcttgggatctgcagtatgc

Site-directed mutagenesis geneblock (IDT) sequences:

mAK7-SDM-TBOX:

tactcttcaggctgaagctgatggaacagcggccgcgttatcacggagataacactggttccaatagcagcctc

acccagcagggaaggaagcatactgcagatcccaagctgtctgccttaggtatcctcatcagcaactttgattga

ttgcagccttccaagatacttgagagttatcagcaaagaagactgataaaaggcttaccttctgggagcctgtac

cctaaatgggtaggggtagggcggccgccaggaacatccaaactgagcagcctttctttct

mAK7-SDM-GATA:

tactcttcaggctgaagctgatggaacagcggccgcgttatcacggagataacactggttccaatagcagcctc

acccagcagggaaggaagcatactgcagatcccaagctgtctgccttagaactttgattgattgcagcctttgtga

agcccaagatacttgagagttatcagcaaagaagactgataaaaggcttaccttctgggagcctgtaccctaaa

tgggtaggggtagggcggccgccaggaacatccaaactgagcagcctttctttctctctt

Preparation of embryo microinjection transgenic mastermix

All transgenic knock-in embryos were generated using a modified CRISPR/Cas9 protocol as previously described in (Kvon et al., 2016, 2020). Briefly, a microinjection mastermix was generated with final concentrations of 20 ng/µl Cas9 protein (IDT 1081058), 50 ng/µl sgRNA (IDT) and 12.5 ng/µl donor plasmid mixed with injection

buffer (IDTE pH 7.5 11-01-02-0). To generate the *Hipp11* locus targeting sgRNA, IDT Alt-R crRNA (5'-gctgatggaacaggtaacaa-3') and Alt-R tracrRNA were first pre-assembled by mixing gently with IDTE pH 7.5 to generate a 50 µM stock solution, incubated for 5 min at 95°C, spun quickly, then cooled to room temperature. Diluted Cas9 protein stock solution 1000 ng/µl was added to the sgRNA mix and incubated for 15 min at room temperature. Donor plasmids were then mixed separately with IDTE pH 7.5 in a DNA LoBind tube (Eppendorf 022431021) and combined with the Cas9 + sgRNA mixture for a final concentration of 12.5 ng/µl donor plasmid in mastermix. The mastermix was filtered through a 0.1 µm MilliporeSigma filter (Fisher UFC30VV00) for 10 min at 14,000 xg. Tubes were parafilmed and stored at 4°C for up to a week prior to microinjection.

Embryo microinjection

Super-ovulated FVB female mice were mated to FVB stud males and fertilized embryos were collected from oviducts. Transgenic mastermix was injected into the pronucleus of the harvested FVB embryos. Injected zygotes were cultured in M16 solution with amino acids at 37°C and 5% CO_2 for ~2 hours, then transferred to pseudopregnant CD-1 females. The procedures for generating transgenic and genetically engineered were approved protocols under LBNL and Gladstone. Initial screens mAK1-6 were performed at LBNL (mAK1-6 at ~E9.5, mAK2,3 at ~E7.5), subsequent screens for mAK4 (~E7.5), and mAK7-9, mAK7-SDM-TBOX, and mAK7-SDM-GATA (all ~E7.5) were performed at the Gladstone Institutes.

Embryo harvest

Transgenic F_0 embryos for screened knock-in constructs mAK1-6 were collected at ~E9.5. Transgenic F_0 embryos for screened knock-in constructs mAK2, mAK3, mAK4, mAK7-9, mAK7-SDM-TBOX, and mAK7-SDM-GATA were collected at ~E7.5. Embryos were dissected in ice-cold PBS (Life Technologies, 14190250) with 1% FBS (Thermo Fisher Scientific, 10439016) on ice. Yolk sac (E9.5 embryos) or a small knick of anterior extraembryonic region (E7.5 embryos) were harvested for genotyping. Embryos were processed in standard 24-well tissue culture plates to track embryo identifier.

X-gal staining and imaging

Freshly dissected embryos were fixed on ice with freshly-made fixative solution (2% formaldehyde, 0.2% glutaraldehyde, 0.02% sodium deoxycholate (Sigma D6750), 0.01% NP-40, 1X PBS), shaking gently for 8 min if E9.5 and 5 min if E7.5. Embryos were washed in wash buffer (2 mM $MgCl_2$ (Ambion AM9530), 0.02% Nonidet P 40 substitute (Fluka 74385), 0.01% Sodium deoxycholate, phosphate buffer pH 7.3 to volume), rocking at room temperature for 10-30 minutes each wash for a total of 3 washes, assuring embryos were submerged the whole time. X-gal stain was prepared as a 40 mg/mL stock from 5 mL N-N-Dimethylformamide (Sigma D4551) and 200 mg 5-Bromo-4-chloro-3-inolyl β-D-galactopyranoside (MilliporeSigma B4252). X-gal staining solution (toxic) was prepared with final constitution 4 nM K-Ferricyanide (Sigma P3667, red), 4nM K-Ferrocyanide (Sigma P9387, yellow), 20 nM Tris pH 7.5 (Invitrogen 15567027), and 0.8 mg/mL X-gal stain, and equilibrated to room temperature, protected from light with foil. Embryos were stained in X-gal staining solution, protected from light

with foil, gently rocking at room temperature, and then placed in cold room at 4°C to continue rocking overnight if signal on batch needed to develop longer. The following morning X-gal was pipetted off and disposed appropriately. Embryos were washed 3 time, 30 min each, rocking at room temperature. Embryos were transferred to tubes containing 4% paraformaldehyde (Electron Microscopy Sciences 15710) for storage overnight prior to imaging, and protected from light. Embryos were imaged in brightfield using a Leica MX165 FC stereomicroscope with DFC450 camera. Imaging was done blind, with genotyping performed afterwards.

Genotyping knock-in events

DNA for genotyping was extracted from yolk sacs (~E9.5) or microdissection of anterior extraembryonic regions (~E7.5) using QuickExtract DNA Extraction Solution (Lucigen, QE09050). Genotyping was performed using Phusion polymerase (NEB M0491S). Random knock-in events were positive for PCR1 only. Tandem knock-in events were positive for PCR1+PCR2. Single knock-in events were positive for PCR2 only.

PCR1-F: AGACTGGGCGGTTTTATGGACA

PCR1-R: TCTGACGCTCAGTGGAACGAAA

PCR2-F: TGCATTCTAGTTGTGGTTTGTCCA

PCR2-R: ACCTTTGCTCTTGGGGCTTAGA

References

Bailey, T.L., Johnson, J., Grant, C.E. & Noble, W.S. (2015) The MEME Suite. *Nucleic Acids Research*. 43 (W1), W39–W49. doi:10.1093/nar/gkv416.

Castro-Mondragon, J.A., Riudavets-Puig, R., Rauluseviciute, I., Berhanu Lemma, R., Turchi, L., et al. (2021) JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Research*. 50 (D1), gkab1113-. doi:10.1093/nar/gkab1113.

Chen, L., Fish, A.E. & Capra, J.A. (2018) Prediction of gene regulatory enhancers across species reveals evolutionarily conserved sequence properties. *PLoS Computational Biology*. 14 (10), e1006484. doi:10.1371/journal.pcbi.1006484.

Costello, I., Pimeisl, I.-M., Dräger, S., Bikoff, E.K., Robertson, E.J. & Arnold, S.J. (2011) The T-box transcription factor Eomesodermin acts upstream of Mesp1 to specify cardiac mesoderm during mouse gastrulation. *Nature Cell Biology*. 13 (9), 1084–1091. doi:10.1038/ncb2304.

Davis, C.A., Hitz, B.C., Sloan, C.A., Chan, E.T., Davidson, J.M., Gabdank, I., Hilton, J.A., Jain, K., Baymuradov, U.K., Narayanan, A.K., Onate, K.C., Graham, K., Miyasato, S.R., Dreszer, T.R., Strattan, J.S., Jolanki, O., Tanaka, F.Y. & Cherry, J.M. (2017) The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Research*. 46 (Database issue), gkx1081-. doi:10.1093/nar/gkx1081.

Devine, W.P., Wythe, J.D., George, M., Koshiba-Takeuchi, K. & Bruneau, B.G. (2014) Early patterning and specification of cardiac progenitors in gastrulating mesoderm. *eLife*. 3, e03848. doi:10.7554/elife.03848.

Domazet-Lošo, T. & Tautz, D. (2010) A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature*. 468 (7325), 815–818. doi:10.1038/nature09632.

Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., et al. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*. 489 (7414), 57–74. doi:10.1038/nature11247.

Hanken, J. & Carl, T.F. (1996) The shape of life: Genes, development, and the evolution of animal form by Rudolf A. Raff University of Chicago Press, 1996. $55.00 hbk, $29.95 pbk (520 pages) ISBN 0 226 70265 0. *Trends in Ecology & Evolution*. 11 (10), 441–442. doi:10.1016/0169-5347(96)81153-0.

Irie, N. & Kuratani, S. (2011) Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis. *Nature Communications*. 2 (1), 248. doi:10.1038/ncomms1248.

Irie, N. & Kuratani, S. (2014) Erratum: Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis. *Nature Communications*. 5 (1), 3366. doi:10.1038/ncomms4366.

Irie, N. & Sehara-Fujisawa, A. (2007) The vertebrate phylotypic stage and an early bilaterian-related stage in mouse embryogenesis defined by genomic information. *BMC Biology*. 5 (1), 1. doi:10.1186/1741-7007-5-1.

Ivanovitch, K., Soro-Barrio, P., Chakravarty, P., Jones, R.A., Bell, D.M., Gharavy, S.N.M., Stamataki, D., Delile, J., Smith, J.C. & Briscoe, J. (2021) Ventricular, atrial, and outflow tract heart progenitors arise from spatially and molecularly distinct regions of the primitive streak. *PLoS Biology*. 19 (5), e3001200. doi:10.1371/journal.pbio.3001200.

Kalinka, A.T., Varga, K.M., Gerrard, D.T., Preibisch, S., Corcoran, D.L., Jarrells, J., Ohler, U., Bergman, C.M. & Tomancak, P. (2010) Gene expression divergence recapitulates the developmental hourglass model. *Nature*. 468 (7325), 811–814. doi:10.1038/nature09634.

Karolchik, D., Hinrichs, A.S., Furey, T.S., Roskin, K.M., Sugnet, C.W., Haussler, D. & Kent, W.J. (2004) The UCSC Table Browser data retrieval tool. *Nucleic Acids Research*. 32 (suppl_1), D493–D496. doi:10.1093/nar/gkh103.

Kent, W.J. (2002) BLAT—The BLAST-Like Alignment Tool. *Genome Research*. 12 (4), 656–664. doi:10.1101/gr.229202.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M. & Haussler, and D. (2002) The Human Genome Browser at UCSC. *Genome Research*. 12 (6), 996–1006. doi:10.1101/gr.229102.

Khan, A., Fornes, O., Stigliani, A., Gheorghe, M., Castro-Mondragon, J.A., et al. (2017) JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Research*. 46 (Database issue), gkx1126-. doi:10.1093/nar/gkx1126.

Kokkinopoulos, I., Ishida, H., Saba, R., Ruchaya, P., Cabrera, C., Struebig, M., Barnes, M., Terry, A., Kaneko, M., Shintani, Y., Coppen, S., Shiratori, H., Ameen, T., Mein, C., Hamada, H., Suzuki, K. & Yashiro, K. (2015) Single-Cell Expression Profiling Reveals a Dynamic State of Cardiac Precursor Cells in the Early Mouse Embryo. *PLoS ONE*. 10 (10), e0140831. doi:10.1371/journal.pone.0140831.

Kvon, E.Z., Kamneva, O.K., Melo, U.S., Barozzi, I., Osterwalder, M., et al. (2016) Progressive Loss of Function in a Limb Enhancer during Snake Evolution. *Cell*. 167 (3), 633-642.e11. doi:10.1016/j.cell.2016.09.028.

Kvon, E.Z., Zhu, Y., Kelman, G., Novak, C.S., Plajzer-Frick, I., et al. (2020) Comprehensive In Vivo Interrogation Reveals Phenotypic Impact of Human Enhancer Variants. *Cell*. 180 (6), 1262-1271.e15. doi:10.1016/j.cell.2020.02.031.

Lescroart, F., Chabab, S., Lin, X., Rulands, S., Paulissen, C., Rodolosse, A., Auer, H., Achouri, Y., Dubois, C., Bondue, A., Simons, B.D. & Blanpain, C. (2014) Early lineage restriction in temporally distinct populations of Mesp1 progenitors during mammalian heart development. *Nature Cell Biology*. 16 (9), 829–840. doi:10.1038/ncb3024.

Lescroart, F., Wang, X., Lin, X., Swedlund, B., Gargouri, S., Sànchez-Dànes, A., Moignard, V., Dubois, C., Paulissen, C., Kinston, S., Göttgens, B. & Blanpain, C. (2018) Defining the earliest step of cardiovascular lineage segregation by single-cell RNA-seq. *Science*. 359 (6380), eaao4174. doi:10.1126/science.aao4174.

Levine, M. & Davidson, E.H. (2005) Gene regulatory networks for development. *Proceedings of the National Academy of Sciences*. 102 (14), 4936–4942. doi:10.1073/pnas.0408031102.

Lickert, H., Takeuchi, J.K., Both, I. von, Walls, J.R., McAuliffe, F., Adamson, S.L., Henkelman, R.M., Wrana, J.L., Rossant, J. & Bruneau, B.G. (2004) Baf60c is essential for function of BAF chromatin remodelling complexes in heart development. *Nature*. 432 (7013), 107–112. doi:10.1038/nature03071.

Luna-Zurita, L., Stirnimann, C.U., Glatt, S., Kaynak, B.L., Thomas, S., Baudin, F., Samee, M.A.H., He, D., Small, E.M., Mileikovsky, M., Nagy, A., Holloway, A.K., Pollard, K.S., Müller, C.W. & Bruneau, B.G. (2016) Complex Interdependence Regulates Heterotypic Transcription Factor Distribution and Coordinates Cardiogenesis. *Cell*. 164 (5), 999–1014. doi:10.1016/j.cell.2016.01.004.

Luo, Y., Hitz, B.C., Gabdank, I., Hilton, J.A., Kagda, M.S., et al. (2019) New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic Acids Research*. 48 (D1), D882–D889. doi:10.1093/nar/gkz1062.

Nord, A.S., Blow, M.J., Attanasio, C., Akiyama, J.A., Holt, A., Hosseini, R., Phouanenavong, S., Plajzer-Frick, I., Shoukry, M., Afzal, V., Rubenstein, J.L.R.,

Rubin, E.M., Pennacchio, L.A. & Visel, A. (2013) Rapid and Pervasive Changes in Genome-wide Enhancer Usage during Mammalian Development. *Cell*. 155 (7), 1521–1531. doi:10.1016/j.cell.2013.11.033.

Ovcharenko, I., Nobrega, M.A., Loots, G.G. & Stubbs, L. (2004) ECR Browser: a tool for visualizing and accessing data from comparisons of multiple vertebrate genomes. *Nucleic Acids Research*. 32 (suppl_2), W280–W286. doi:10.1093/nar/gkh355.

Panigrahi, A. & O'Malley, B.W. (2021) Mechanisms of enhancer action: the known and the unknown. *Genome Biology*. 22 (1), 108. doi:10.1186/s13059-021-02322-1.

Probst, S., Sagar, S., Tosic, J., Schwan, C., Grün, D. & Arnold, S.J. (2020) Spatiotemporal sequence of mesoderm and endoderm lineage segregation during mouse gastrulation. *Development*. 148 (1), dev193789. doi:10.1242/dev.193789.

Song, M., Yuan, X., Racioppi, C., Leslie, M., Stutt, N., Aleksandrova, A., Christiaen, L., Wilson, M.D. & Scott, I.C. (2022) GATA4/5/6 family transcription factors are conserved determinants of cardiac versus pharyngeal mesoderm fate. *Science Advances*. 8 (10), eabg0834. doi:10.1126/sciadv.abg0834.

Spitz, F. & Furlong, E.E.M. (2012) Transcription factors: from enhancer binding to developmental control. *Nature Reviews Genetics*. 13 (9), 613–626. doi:10.1038/nrg3207.

Sun, X., Hota, S.K., Zhou, Y.-Q., Novak, S., Miguel-Perez, D., Christodoulou, D., Seidman, C.E., Seidman, J.G., Gregorio, C.C., Henkelman, R.M., Rossant, J. & Bruneau, B.G. (2017) Cardiac-enriched BAF chromatin-remodeling complex subunit Baf60c regulates gene expression programs essential for heart development and function. *Biology Open*. 7 (1), bio029512. doi:10.1242/bio.029512.

Takeuchi, J.K., Lickert, H., Bisgrove, B.W., Sun, X., Yamamoto, M., Chawengsaksophak, K., Hamada, H., Yost, H.J., Rossant, J. & Bruneau, B.G. (2007) Baf60c is a nuclear Notch signaling component required for the establishment of left–right asymmetry. *Proceedings of the National Academy of Sciences*. 104 (3), 846–851. doi:10.1073/pnas.0608118104.

Thomas, D.J., Rosenbloom, K.R., Clawson, H., Hinrichs, A.S., Trumbower, H., Raney, B.J., Karolchik, D., Barber, G.P., Harte, R.A., Hillman-Jackson, J., Kuhn, R.M., Rhead, B.L., Smith, K.E., Thakkapallayil, A., Zweig, A.S., Consortium, T.E.P., Haussler, D. & Kent, W.J. (2007) The ENCODE Project at UC Santa Cruz. *Nucleic Acids Research*. 35 (suppl_1), D663–D667. doi:10.1093/nar/gkl1017.

Tosic, J., Kim, G.-J., Pavlovic, M., Schröder, C.M., Mersiowsky, S.-L., Barg, M., Hofherr, A., Probst, S., Köttgen, M., Hein, L. & Arnold, S.J. (2019) Eomes and Brachyury control pluripotency exit and germ-layer segregation by changing the chromatin state. *Nature Cell Biology*. 21 (12), 1518–1531. doi:10.1038/s41556-019-0423-1.

Yuan, X., Scott, I.C. & Wilson, M.D. (2021) Heart Enhancers: Development and Disease Control at a Distance. *Frontiers in Genetics*. 12, 642975. doi:10.3389/fgene.2021.642975.

Yuan, X., Song, M., Devine, P., Bruneau, B.G., Scott, I.C. & Wilson, M.D. (2018) Heart enhancers with deeply conserved regulatory activity are established early in zebrafish development. *Nature Communications*. 9 (1), 4977. doi:10.1038/s41467-018-07451-z.

Chapter 4: A *Mesp1*-dependent developmental breakpoint in transcriptional and epigenomic specification of early cardiac precursors

Abstract

Transcriptional networks governing cardiac precursor cell (CPC) specification are incompletely understood due in part to limitations in distinguishing CPCs from non-cardiac mesoderm in early gastrulation. We leveraged detection of early cardiac lineage transgenes within a granular single cell transcriptomic time course of mouse embryos to identify emerging CPCs and describe their transcriptional profiles. *Mesp1*, a transiently-expressed mesodermal transcription factor (TF), is canonically described as an early regulator of cardiac specification. However, we observed perdurance of CPC transgene-expressing cells in *Mesp1* mutants, albeit mis-localized, prompting us to investigate the scope of *Mesp1*'s role in CPC emergence and differentiation. *Mesp1* mutant CPCs failed to robustly activate markers of cardiomyocyte maturity and critical cardiac TFs, yet they exhibited transcriptional profiles resembling cardiac mesoderm progressing towards cardiomyocyte fates. Single cell chromatin accessibility analysis defined a *Mesp1*-dependent developmental breakpoint in cardiac lineage progression at a shift from mesendoderm transcriptional networks to those necessary for cardiac patterning and morphogenesis. These results reveal *Mesp1*-independent aspects of early CPC specification and underscore a *Mesp1*-dependent regulatory landscape required for progression through cardiogenesis.

Background

Cardiogenesis requires precise specification and patterning of the cardiac precursor cells (CPCs) as they emerge from the gastrulating mesoderm in very early stages of embryogenesis. Errors in this process lead to congenital heart defects (CHDs), which affect 1-2% of live births (Bruneau, 2008). The genetic etiology of CHDs indicates that genes encoding transcriptional regulators are overrepresented as causative and are predominantly haploinsufficient, indicating that fine dysregulation of gene expression is a critical mechanism for disease (Zug, 2022; Nees & Chung, 2019). A thorough delineation of the transcriptional networks governing cardiogenesis is foundational to understanding how defects in this process manifest as CHDs, and may inform the design of strategies to treat CHDs and heart disease broadly.

Cardiogenesis begins when mesoderm progenitors emerge from the primitive streak and migrate towards the anterior-lateral aspects of the developing embryo (Saga, Kitajima & Miyagawa-Tomita, 2000; Saga et al., 1999). Interrogating the earliest cardiac progenitors distinctly from the developing mesoderm has historically been challenging due to a paucity of molecular markers available to distinguish a CPC from the rest of the developing mesoderm. Prior studies used lineage tracing of mesoderm progenitors expressing the basic-helix-loop-helix (bHLH) transcription factor (TF) *Mesp1,* which is transiently expressed in cells that go on to contribute to the heart, somitic mesoderm derivatives, and craniofacial mesoderm (Devine et al., 2014; Lescroart et al., 2014; Saga et al., 1999). Clonal lineage tracing studies have shown that a subset of *Mesp1+* cells at early gastrulation are fated for distinct cardiac substructures well before

anatomy is patterned, highlighting extensive diversification among early mesodermal progenitors (Devine et al., 2014; Lescroart et al., 2014; Liu, 2017).

Deletion of *Mesp1* in mice variably disrupts specification and migration of cardiac progenitors (Ajima et al., 2021; Saga, Kitajima & Miyagawa-Tomita, 2000; Saga et al., 1999; Kitajima et al., 2000; Lescroart et al., 2018). During in vitro cardiac differentiation, overexpression of *Mesp1* induces expression of subsequent cardiac TFs, indicating a potentially instructive role in cardiogenesis (Chiapparo et al., 2016; Bondue et al., 2008; Lindsley et al., 2008; Wu, 2008; Kelly, 2016; Bondue & Blanpain, 2010; Lin et al., 2022; Soibam et al., 2015). Gain of function experiments suggest a broad and important function for *Mesp1* in mesoderm differentiation, but the *in vivo* gene regulatory landscape controlled by *Mesp1* remains unclear (Costello et al., 2011; Saga et al., 1999; Liu, 2017; Saga, Kitajima & Miyagawa-Tomita, 2000; Ajima et al., 2021; Saga et al., 1996; Kitajima et al., 2000; Lin et al., 2022).

Previous studies identified an enhancer of *Smarcd3,* "F6", which is specifically active in CPCs fated to become the totality of heart cells, and is active shortly after *Mesp1* expression and before other early cardiac-specific TFs are expressed (Devine et al., 2014; Yuan et al., 2018). Thus, *Smarcd3*-F6 activity enables distinct identification of CPCs as they emerge from the developing mesoderm. We found that the *Smarcd3*-F6 enhancer remains active in posterior regions of *Mesp1* KOs, indicating perdurance of cardiogenesis in some capacity. Here, we utilized the *Smarcd3*-F6 transgene to comprehensively delineate the dynamic transcriptional and epigenomic consequences of *Mesp1* loss during early cardiogenesis and reveal *Mesp1*-independent aspects of

cardiac specification. This study challenges the concept of a master regulator for cardiac specification by defining transcriptional phases during cardiogenesis with different vulnerabilities to *Mesp1* loss.

Results

Transcriptional profiling of *Smarcd3*-F6+ cells shows enduring expression of cardiac genes in *Mesp1* knockout embryos

To determine the requirement for *Mesp1* in establishing CPC identity, we investigated the transcriptional identities of *Smarcd3*-F6+ cells upon loss of *Mesp1*. We detected *Smarcd3*-F6 expressing cells in *Mesp1*$^{Cre/Cre}$ (*Mesp1* KO) embryos, although positive cells are localized posteriorly relative to control embryos at early cardiac crescent stages (Fig. 4.1A). The persistence of *Smarcd3*-F6+ cells led us to hypothesize that these cells represent retained CPCs, suggesting that as previously described (Saga, Kitajima & Miyagawa-Tomita, 2000; Ajima et al., 2021; Saga et al., 1999), aspects of early cardiac specification may be *Mesp1*-independent. Thus, the transcriptional and epigenomic programs regulated by and independent of *Mesp1* remain to be understood during *in vivo* cardiogenesis.

We performed scRNA-seq on whole *Mesp1* KO embryos and littermate controls along a timeline of developmental stages for early cardiogenesis spanning early gastrulation (E6.0) to cardiac crescent formation (E7.75) (Fig. S4.3A). We bioinformatically identified *Smarcd3*-F6-eGFP-expressing cells from the whole embryo time course (Fig. S4.3A-D)

to generate an atlas of 4,868 *Smarcd3*-F6+ cells representing 24 cell types (Fig. 4.1B, Fig. S4.1A). The majority of *Smarcd3*-F6+ cells represented early cardiac mesodermal derivatives such as the late streak mesoderm (LSMeso), *Mesp1*+ mesoderm (Mesp1M), posterior mesoderm (postMeso), LPM, precardiac mesoderm (preCardiacMeso), and early CMs (Fig. 4.1B-C, Fig. S4.1A). We detected cells of the allantois, lateral plate mesoderm/extraembryonic mesoderm (LPM-ExEM), and the node/notochord, consistent with *Smarcd3* expression in these domains (Fig 4.1B, Fig. S4.1A) (Takeuchi et al., 2007; Devine et al., 2014). Additionally, we found populations of blood, endothelial cells, Reichert's membrane, posterior paraxial mesoderm (postPrxM) and cells appearing endoderm-like, potentially representing early mesendoderm cells (Fig. 4.1B, Fig. S4.1A). Inducible lineage labeling of *Smarcd3*-F6+ cells at E6.5 excluded lineage contributions to non-cardiac cell types (Devine et al., 2014), suggesting detection here is the result of genotype-agnostic, weak, or transient transgene expression (Fig. S4.1A).

To examine overall trends in gene expression differences between *Mesp1* KO and control, we performed a comparison of all *Smarcd3*-F6+ cells between genotypes irrespective of cell type or embryonic stage in the developmental timeline (Fig. S4.1B). We found that *Mesp1* KO *Smarcd3*-F6+ cells express mesodermal genes of the emerging cardiac lineage such as *Tdgf1*, *Lhx1*, *Eomes*, and *Myl7*, however mostly lacked expression of more mature cardiac progenitor markers such as *Nkx2-5* (Fig. 2E, Fig. S4A). When we divided the "all cells" genotype analysis into relative developmental stages separating "Early" embryos (E6.0-E6.5), "Middle" embryos (late E6.5-E7.5), and "Late" embryos (late E7.5 to early E7.75), we found that genotype discrepancies in

cardiac-related gene expression were minor at Early stages and diverged with increasing embryonic age (Fig. 4.1E).

Relatedly, the distribution of genotypes across *Smarcd3*-F6+ cell types shows that *Mesp1* KO cells are not fully represented in every cell type (Fig. 4.1B-D, Fig. S4.1C). Both genotypes were present in mesoderm clusters (C1, C3), the preCardiacMeso (C4), the postMeso (C5), retinoic acid signaling cells (C6), LSMeso (C7), allantois (C8), endothelial (C9), postPrxM (C10), the endoderm-like clusters (C11, C14, C19), the LPM-ExEM cluster (C15), the primitive streak (PS) (C17), postMeso (C20), blood (C21), and Reichert's (C23) (Fig. 2B-C, Fig. S4C). Only control cells were present in LPMs (C0, C16), CMs (C2, C12), postLPM (C13), Mesp1M (C18), node/notochord (C22). Many of the cell types comprised only of control were Late-stage embryo cells (Fig. 4.1B-D, Fig. S4.1C), indicating that cell type heterogeneity was affected with loss of *Mesp1* in *Smarcd3*-F6+ cells with increasing severity as development progresses. Furthermore, while the preCardiacMeso and LSMeso cell types were represented by both genotypes in Early- and Middle-staged embryos, the Late-stage embryo cells represented in the preCardiacMeso were exclusively *Mesp1* KO (Fig. 4.1B-D, Fig. S4.1C), indicating retention of precursor transcriptional profiles.

To understand *Mesp1*-correlated differences in emerging *Smarcd3*-F6+ CPCs in individual cell types, we performed differential expression testing within cell types present in both genotypes. Within preCardiacMeso and LSMeso cells, we found similar expression of *Tdgf1*, *Eomes, Fgf8*, genes involved in early mesoderm specification (Fig. S4.2A-C) (Probst et al., 2020; Reifers et al., 2000). These results were confirmed by

multiplexed RNA *in situ* hybridization, which showed co-expression of *Smarcd3*-F6 with these markers in cardiogenic regions of E6.0-E6.5 (Fig. 4.1F, Fig. 4.1I) and E7.0 (Fig. 4.1H) embryos. Notably, *Mesp1* KO embryos showed decreased levels of *Smarcd3*-F6 and broad posterior expansion of *Tdgf1* (Fig. 4.1F) and *Fgf8* (Fig. 4.1H) expression beyond *Smarcd3*-F6+ cardiogenic regions. Additionally, *Tdgf1* and *Eomes* expression aberrantly perdured through late E7.5 (Fig. 4.1F,G) and E7.0 (Fig. 4.1I-J), respectively. Other genes involved in early mesoderm specification (*Fgf10),* lineage specification and pluripotency exit (*Chchd2* and *Nme2*), and non-cardiac mesoderm genes (*Amot*) were upregulated in *Mesp1* KO cells relative to controls, while genes involved in migration and patterning (*Lefty2*, *Rac1*, *Foxf1*) were downregulated (Fig. S4.2B,C) (Zhu et al., 2009, 2016; Migeotte, Grego-Bessa & Anderson, 2011; Sang et al., 2021).

Within the Late-stage-dominated LPM-ExEM and endoderm-like1 cell types, we found similar expression levels of *Myl7* between genotypes (Fig. 4.1K, Fig. S4.2A, Fig. S4.2D,E). *Mesp1* KO cells displayed relative upregulation of early mesoderm specification genes (*Tdgf1, Eomes, Fgf8, S100a10, Ifitm2*, *Fn1)* and downregulation of morphogenesis and migration genes (*Dlk1, Elavl1*) (Fig. S4.2D,E) (Probst et al., 2020; Cheng et al., 2013; Klymiuk et al., 2012; Saykali et al., 2019; Katsanou et al., 2009)

Collectively, these analyses indicate that the *Mesp1* KO transcriptional phenotype of *Smarcd3*-F6+ cells becomes increasingly disrupted as embryonic development progresses, consistent with the divergent morphology of *Mesp1* KO embryos at cardiac crescent stages (Fig. 4.1A-D).

Alterations to cardiac mesoderm in *Mesp1* knockout embryos become increasingly severe as gastrulation progresses

Following characterization of *Mesp1* KO effects in *Smarcd3*-F6+ cells specifically, we sought to understand alterations to the mesoderm, inclusive of *Smarcd3*-F6+ cells and the cardiac mesoderm, more broadly. We applied our method of dual-reporter transgene identification (Fig. 2.1) to generate an atlas of 35,792 mesodermal cells from both control and *Mesp1* KO embryos (Fig. S4.3A-F). The relative Early- and Middle-stage embryos showed a similar census of mesodermal cell types between genotypes, including preCardiacMeso (Fig. S4.3E-I). However, *Mesp1* KO Late-stage embryo mesoderm lacked many of the cell types present in control, such as mature CMs, PrxM, and PrSoM cells (Fig. S4.3E-I). To interrogate how these changes occur in developmental time, we divided the mesoderm dataset into the Early, Middle, and Late developmental stages as defined in Fig. 4.1E. Mesodermal cells for each stage were re-clustered, and differential gene expression was assessed between genotypes (Fig. 4.2).

Within the Early mesoderm dataset (Fig. 4.2A, Fig. S4.4A), we identified the LSMeso2 and *Eomes*+ primitive streak mesoderm (EomesPSMeso) as clusters of interest for cardiac specification based on enriched *Smarcd3*-F6+ expression (Fig. 4.2C). Both genotypes were present in each cell type (Fig. 4.2B), indicating that *Mesp1* KO cells are able to engage with transcriptional programs to exit pluripotency and initiate cardiac mesoderm specification. Differential gene expression analysis revealed *Mesp1* KO cells showed upregulation of mesendoderm and PS markers (*Fgf5*, *Mixl1*, *Upp1*, *Fgf5*, *Sox2*, *Tdgf1*), downregulation of LPM differentiation genes (*Foxf1*, *Taf10*),

downregulation of migration and patterning genes (*Rac1*, *Elavl1*), and persistent but decreased expression of cardiac *Myl7* (Fig. 4.2D-E).

Within the Middle mesoderm dataset (Fig. 4.2F-H, Fig. S4.4B) we focused on the *Smarcd3*-F6 enriched *Mesp1*+ mesendoderm cluster (Mesp1ME) and its developmental predecessors, LSMeso2 cells. Middle-stage LSMeso2 cells (Fig. 4.2I) showed similar expression patterns between genotypes to Early-stage LSMeso cells (Fig. 4.2D). *Mesp1* KO cells of the Mesp1ME upregulated posterior mesoderm organization genes (*Fgf10* and *Gsc*) (Probst et al., 2020; Meijer et al., 2000; Branney et al., 2009) and the non-cardiac mesoderm gene *Anxa2* (Schwartz et al., 2014; Wang et al., 2015), and downregulated *Lefty2, Rac1*, and myogenesis differentiation gene *Pcbp1* (Shi & Grifone, 2021) (Fig. 4.2J). Notably, there was an absence of *Mesp1* KO cells in *Smarcd3*-F6 and *Mesp1* enriched clusters representing *Foxc2*+ mesoderm cells (Fig. 4.2F-H, Fig. S4.5A). *Foxc2* operates in cardiac field diversification and morphogenesis (Seo & Kume, 2006; Lescroart et al., 2018). Examination in E6.75 embryos by immunohistochemistry and light sheet imaging showed that anterior-proximal marker domains were misaligned in *Mesp1* KO embryos, and *Foxc2* was absent (Fig. S4.5A-B). Together, these results indicate dysregulation of networks controlling cellular movements and domain boundaries, as well as reduced cellular diversification in *Mesp1* KO embryos of pre-crescent stages.

Analysis of Late mesoderm *Mesp1* KO cells revealed restricted diversity of both cardiac and other mesodermal cell types (Fig. 4.2K-L, Fig. S4.4C). Furthermore, while both genotypes were found in *Smarcd3*-F6 enriched clusters (Meso1, Meso_2) and the

postLPM, there were no *Mesp1* KO cells in the CM clusters (Fig. 4.2K,M). *Mesp1* KO cells from Meso1 and Meso2 clusters had highly disrupted transcriptional profiles characterized by upregulation of several mesodermal genes (*Cited2*, *Ifitm2*, *Mif*, *Ahnak, Ankrd11, Myl6*) (Weninger et al., 2005; Lange et al., 2003; Huang et al., 2022) and downregulation of cardiac maturation genes (*Dlk1*, *Acta2, Ifitm1*) (Pursani et al., 2017; Klymiuk et al., 2012) (Fig. 4.2N,O). Additionally, the few *Mesp1* KO cells present in the postLPM cluster upregulated genes involved in mesendoderm specification and organization (*Lhx1, Eomes, Asb4*) (Fernandez-Guerrero et al., 2021) and downregulated or else lacked patterning, morphogenesis, and maturation genes (*Crabp1*, *Foxc2*, *Meis2*) (Fig. 4.2P). From these results we conclude that Late stage *Mesp1* KO embryos fail to produce mature CMs and various mesoderm cell types, and display highly disrupted transcriptional profiles in retained cardiac mesoderm cells.

Thus, similar to the patterns described specifically in *Mesp1* KO *Smarcd3*-F6+ CPCs, *Mesp1* KO cardiac mesoderm cells show transcriptional dysregulation that becomes increasingly divergent as embryonic development progresses. Additionally, we observed gross disruption of mesoderm diversification beyond purely cardiogenic cell types in Middle- and Late-stage *Mesp1* KO embryos (Fig. 4.2F-H,K-M, Fig. S4.5), consistent with their altered morphology.

Mesp1 knockout cardiac mesoderm cells progress incompletely and imperfectly towards cardiomyocyte fates

We next investigated the steps of cardiac fate progression to understand how *Mesp1* KO embryos initiate cardiogenesis but fail to produce matures CMs. Utilizing pseudotemporal trajectory ordering with URD (Farrell et al., 2018) on the full mesoderm dataset, we defined the epiblast cells, the cluster also containing the earliest staged embryos (C1-Epiblast in Fig. S4.3E,G), as the root, and clusters containing the most differentiated mesodermal cells from the oldest stage embryos as the tips (Fig. 4.3A, Fig. S4.3E,G, Fig. S4.6A,B). We layered expression of *Smarcd3*-F6-eGFP to identify the main cardiogenic fate paths within the tree space, which also co-expressed CM genes such as *Nkx2-5*, *Myl7*, and *Smarcd3* (Fig. 4.3C, Fig. S4.6C). Within CM and CardiacMeso fate branches, *Mesp1* KO cells occupied the youngest pseudotemporal positions near the top of the branch segment, and were more represented in younger pseudotime segment branches of the tree, including their own earlier-pseudotime branch fate “C22” which was defined by multiple mesodermal genes not representative of any particular wildtype cell type (Fig. 4.3A-B, Fig. S4.3E,I).

Focusing on the cardiogenic fate tree section beginning at segment 34, we performed differential gene expression analysis to compare cell types of similar fate potentials within branch segments or pseudotemporal levels of the trajectory (Fig. 4.3D-H). Among CM-fated cells, *Mesp1* KO cells were enriched for expression of *Anxa2*, *Hand1*, *Krt8* and other genes reminiscent of extraembryonic mesoderm, expressed lower levels of structural myocyte genes such as *Myl7* and *Tnnt2* relative to control, and lacked *Nkx2-5*

transcripts. (Fig. 4.3D). *Hand1* was similarly enriched in *Mesp1* KO CardiacMeso-fated cells, along with *Vim*, a fibroblast gene, and *Tagln2*, a gene involved in cell transformation and cell morphology (Han et al., 2017) (Fig. 4.3E). In the CardiacMeso-fated segment, *Myl7* and *Id2* were reduced relative to controls, as was *Ankrd1*, a gene implicated in sarcomere-binding and dilated cardiomyopathy that is known to be upregulated with overexpression of *Mesp1* (Bondue & Blanpain, 2010; Moulik et al., 2009) (Fig. 4.3E). In the branch that gave rise to CM and CardiacMeso fates, segment 30, *Amot*, *Hand1*, and *Ifitm2*, genes expressed in the posterior proximal extraembryonic border of the murine embryo and ExEMeso, were increased in *Mesp1* KO cells (Fig. 4.3F). By contrast, myocyte and cardiac progenitor genes *Myl7*, *Gata5*, and *Gata4* were decreased in *Mesp1* KO cells relative to control (Fig. 4.3F). In *Mesp1* KO cells in segment 34, the predecessors to LPM mesodermal derivatives, pronephros gene *Cox6b1*, spongiotrophoblast and extraembryonic energy storage gene *Phlda2*, and ESC self-renewal gene *Nme2* (Zhu et al., 2009) were enriched, while retinoic acid gene *Crabp2* and early gastrulation genes *Dnmt3b*, *Pou5f1* were downregulated (Fig. 4.3G). Finally, we compared the *Mesp1* KO cell-dominated segment 22 to its pseudotime-branching contemporary segment 30, and found enriched expression of mesoderm-fate promoting gastrulation TFs *Cdx2* and *T*, along with mesendoderm allocation gene *Tdgf1* (Fig. 4.3H). Conversely, cardiac progenitor morphogenesis TFs *Mef2c*, *Gata6*, and *Gata4* were downregulated (Fig. 4.3H).

We summarize these analyses of *Mesp1* KO cardiac mesoderm fates into two categories; 1) retained expression of some cardiac progenitor genes (*Myl7, Gata4/5/6, Id2, Tnnt2*), albeit at decreased levels relative to control, and absence of others (*Nkx2-*

5, Ankrd1), and 2) ectopic enrichment of ExEMeso and other mesoderm associated genes (*Hand1, Anxa2, Amot, Vim, Tagln2*). We used multiplexed fluorescent RNA *in situ* hybridization to validate the spatial domains of differentially expressed genes in Late-stages, and confirmed presence of *Myl7*+ cells co-expressing *Smarcd3*-F6 in the posterior distal compartment of *Mesp1* KO embryos (Fig. 4.3I) along with absence of *Nkx2-5* expression in *Mesp1* KO embryos (Fig. 4.3J). We also showed ectopic *Anxa2* expression into the embryo proper, overlapping with *Smarcd3*-F6+ cells in their posterior position in *Mesp1* KO embryos, in contrast to the anterior extraembryonic-restricted expression pattern of controls (Fig. 4.3K). These results further highlight that *Mesp1* KO CPCs ectopically express non-cardiac mesodermal genes, and reveals that *Mesp1* KO CPCs progress towards CM fates incompletely in part through a failure to express requisite TFs. Thus, *Mesp1* KO CPCs reach a cardiogenic breakpoint during gastrulation prior to cardiac crescent formation.

scATAC-seq analysis reveals regulatory barrier in *Mesp1* knockout mesoderm progression towards cardiomyocyte fates

To characterize the regulatory landscape prohibiting *Mesp1* KO cells from progressing fully towards CM fates, we turned to single cell Assay for Transposase Accessible Chromatin (scATAC-seq) (Buenrostro et al., 2015) of Middle- and Late-stage embryos ages E7.5 - E7.75 (Fig. S4.7). We processed whole embryos and performed preliminary atlasing analysis in ArchR (Granja et al., 2021). We utilized integration with the complementary whole embryo scRNA-seq dataset along with chromatin accessibility profiles near marker genes (gene scores) to subset mesodermal cell type clusters (Fig.

S4.8A-D) in order to generate a subset scATAC-seq atlas of 16 mesodermal cell types (Fig. 4.4A). *Mesp1* KO and controls had strikingly divergent regulatory landscapes (Fig. 4.4B). *Mesp1* KO cells were confined to scATAC-seq clusters representing epiblast (Epi), mesendoderm, and LPM cell types, while control cells were represented in the LPM cell types, the more mature cardiac progenitor (CP) and CM cluster, and mesodermal derivative cell types (Fig. 4.4A-C). Integration with the complementary mesoderm scRNA-seq dataset, visualization of key marker gene scores and integrated expression (Fig. 4.4D-E), and Jaccard indexing (Fig. S4.9) were used to assign relative cell identities to each mesoderm scATAC-seq cluster (Fig. 4.4C). While some cardiac TFs such as *Nkx2-5* were not active in *Mesp1* KO cells, others such as *Tbx5* had chromatin accessibility in *Mesp1* KO cells, but integrated expression only in control CM/CP cells (Fig. 4.4B,D-E). Other cardiac TFs *Hand1* and *Gata4* had similar activity between *Mesp1* KO and control cells (Fig. 4.4B,D-E), and while *Mesp1* KO cells downregulated *Smarcd3* and *Myl7* expression, chromatin accessibility for these genes was similar between genotypes (Fig. 4.4B,D-E). These results indicate a perdurance of active chromatin states in the steps preceding cardiogenic differentiation.

To interrogate the developmental relationship between *Mesp1* KO cells failing to mature and control CMs, we performed an ArchR trajectory inference analysis assessing pseudotime along the cardiac fate path. We defined a trajectory backbone in the *Mesp1* KO cells traversing the expected differentiation path of Epi, *Eomes*+ mesendoderm (EomesME), *Mesp1*+ mesoendoderm (Mesp1ME), lateral plate mesoderm (LPM2, LPM1), to cardiac progenitors (CP) and cardiomyocytes (CM) clusters. This trajectory analysis revealed that while *Mesp1* KO cells traversed the normal path from epiblast to

LPM, they abruptly failed to progress further towards CPs and CMs (Fig. 4.4F). Notably, the most mature cell identity *Mesp1* KO cells achieved (LPM1) also contained control cells capable of progressing to CPs past this point where *Mesp1* KO cells halted, indicating the LPM1-to-CP transition represents the breakpoint in cardiogenesis for *Mesp1* KO cells (Fig. 4.4F).

From this trajectory analysis, we assessed dynamic shifts in the correlation of TF gene scores and gene expression with corresponding TF motifs in accessible chromatin peaks across pseudotime (Fig. S4.10A-B) to reveal a biologically-sensical order of TF regulators involved in cardiogenesis. Notably, TFs represented in early pseudotime and *Mesp1* KO cells (*Lhx1, T, Eomes, Zic2/3, Pitx2, Isl1*, Fig. S13A-B) were consistent with early gastrulation mesodermal regulatory networks, indicating that aspects of these networks are either *Mesp1*-independent or resilient to *Mesp1* loss. TFs represented in later pseudotime (*Hand2, Gata4/5/6, Hoxb1*, Fig. S4.10A-B) were concordant with downregulated gene expression in *Mesp1* KO CPCs and mesoderm by scRNA-seq (Fig. 4.1-4.3), suggesting that failed induction of these TFs and their programs is either *Mesp1*-dependent or vulnerable to secondary effects of *Mesp1* loss.

To ascertain which gene regulatory networks were present in which cell types, and thus which genotypes, along the cardiogenic trajectory, we performed an orthogonal analysis to identify putative positive transcriptional drivers (Fig. 4.4G) and visualized resulting TFs' motif enrichments in UMAP space (Fig. 4.4H, Fig. S14.10A-B). In particular, the *Mesp1* KO Epi cluster is driven in part by pluripotency TFs Pou5f1 and Mesp1-cofactor Zic3 (Lin et al., 2022) (Fig. 4.4D-E, 4.4H). Mesendoderm TFs Eomes and Zic3 were

drivers of EomesME and Mesp1ME (Fig. 4.4D-E, 4.4H). ExEM and first heart field TF Hand1 appeared in the "last-stop" LPM1 cell types where *Mesp1* KO cells failed to progress towards more mature cardiac fates (Fig. 4.4D-E, 4.4H), consistent with the upregulated expression observed in *Mesp1* KO CM-fated cells (Fig. 4.3D-F). While Gata motifs were present in LPM2 and LPM1, the latter of which contains both genotypes, Gata4 was most enriched in the later cardiac fate destinations of CPs and CMs (Fig. 4.4H). This result coupled with the *Gata4*'s representation in late trajectory pseudotime (Fig. S4.10A) and downregulated expression in cardiac-fated *Mesp1* KO mesoderm cells (Fig. 4.3F) likely signifies *Mesp1*-dependent induction and/or influence of Gata factor-associated networks within emerging CMs. Indeed, *Gata4* was shown to be activated by Mesp1 during *in vitro* differentiation (Soibam et al., 2015), and while Gata4 binds the minority of Mesp1-bound enhancers, Gata4 binds nearly half of enhancers opened following *in vitro* induction of Mesp1 (Lin et al., 2022). Separately, Mesp1 target gene *Hoxb1*'s motif was distinctly expressed in PrSoM cell types, coincident with the "late phase" role for *Mesp1* (Lin et al., 2022; Haraguchi et al., 2001) in mesoderm diversification beyond the cardiac lineage (Fig. 4.4H). The preponderance and accordance of these results supports that early cardiogenic phases proceed resilient to *Mesp1*-loss, however *Mesp1* KO cells cannot proceed to later phases.

Given enrichment of Eomes motifs, gene score, and gene expression in mesendoderm clusters (Fig. 4.4D-E, 4.4I), its apparent role as a positive TF driver (Fig. 4.4G,I) and its direct involvement in Mesp1 induction (Tosic et al., 2019; Costello et al., 2011; Alexanian et al., 2017; Guo et al., 2018; Probst et al., 2020), we investigated Eomes as a potential driver of *Mesp1*-independent early phases of cardiogenesis. Eomes directly

binds *Myl7* regulatory regions (Tosic et al., 2019), and *Eomes* loss disrupts induction of *Myl7* (Costello et al., 2011), supporting that expression of *Myl7* in *Mesp1* KO CPCs (Fig. 4.1E,K, Fig. S4.2A,C-E, Fig. 4.2D, Fig. 4.3C-F,J) is regulated by Eomes at least partially independently of Mesp1. Furthermore, domains of *Eomes* expression anomalously endured in cardiogenic *Smarcd3*-F6+ regions and are ectopically expanded in lateral aspects of the embryo proper in cardiac-crescent staged *Mesp1* KOs (Fig. 4.4I), indicating improper repression of *Eomes* in cardiogenic regions.

Taken together, these results describe a shift between mesendoderm and cardiac patterning regulatory programs during pre-crescent phases of cardiogenesis. *Mesp1* KO cells are unable to traverse beyond LPM cell types to initiate cardiac patterning programs and instead retain gene expression indicative of earlier cardiac mesoderm regulatory programs. The ectopically retained expression of *Eomes* may be a driving mechanism for this halt in cardiogenesis.

The disrupted regulatory landscape of *Mesp1* KO embryos is characterized by ectopic endurance of mesendoderm gene programs

To understand how the *Mesp1* KO disrupted regulatory landscape underlies the transcriptional barriers to progression towards more mature cardiac fates, we characterized cell type peak accessibility profiles and the motif enrichment within these peaks (Fig. S4.11A-B). We performed differential accessibility testing of peaks between cell types along the cardiac trajectory (Fig. S4.11C-G). Focusing specifically on the “last stop” for *Mesp1* KO cells, we compared motif enrichment within differential peaks of CMs/CPs containing only control and LPM1 containing both control and *Mesp1* KO cells

(Fig. 4.5A). In agreement with motif enrichment scores for positive TF regulators (Fig. 4.4G, 4.4H), Gata and Mef2c motifs were among those enriched in the more mature cardiac fates, while motifs for cardiac differentiation and myogenesis-promoting Tead factors were relatively enriched within LPM1, offering further explanation for retention of some myocyte identity within *Mesp1* KO CPCs (Fig. 4.5A, Fig. 4.1E, 4.1K, Fig. S4.2D, Fig. 4.3J) (Han et al., 2020; Akerberg et al., 2019). To measure correlations between the differential accessibility profiles behind these motifs and the complementary gene expression profiles of these cells, we performed an association analysis measuring the probability that peak accessibility near genes corresponds to gene expression. We applied this analysis to find an odds ratio of 16.7 for the probability that significantly differentially open peaks corresponded to upregulated gene expression (Q3, Fig. 4.5B) while significantly differentially closed peaks corresponded to downregulated gene expression (Q1, Fig. 4.5B) in CMs/CPs relative to LPM1. Thus, gene expression profiles enriched in control-only CMs/CPs (Q3: *Mef2c, Tbx5, Gata5, Nkx2-5, Tnnt2*, Fig. 4.5B) and transcriptional profiles of LPM1 cells (Q1: *Hand1*, *Anxa2, Cdx2, Krt8/18*, Fig. 4.5B*)* are associated with these cells' differing chromatin landscapes.

We next compared control-only CMs/CPs to *Mesp1* KO-only Mesp1ME and LPM2 (Fig. 4.5C-D) because these cells had similar gene scores for the *Smarcd3* locus (Fig. 4.4D), a proxy for *Smarcd3*-F6 enhancer activity. Motifs including those for Gata and Hox factors were relatively enriched in CMs/CPs, and T-box motifs including Eomes and T were enriched in Mesp1ME and LPM2 (Fig. 4.5C). The correlation odds ratio of 8.08 highlighted corresponding peak accessibility and gene expression enrichment for CP patterning and CM genes in control CMs/CPs (Q3: *Nkx2-5, Tbx5, Wnt2, Mef2c, Meis1,*

Ttn, Tnnt2) and relative enriched peak accessibility near upregulated genes for earlier cardiac mesoderm and mesendoderm programs in Mesp1ME and LPM2 *Mesp1* KO-only cells (Q1: *Tdgf1, Fgf3, Eomes, Mixl1, T, Krt8, Hand1, Pou5f1*) (Fig. 4.5D).

Applying this analysis paradigm to multiple pairwise comparisons along the cardiogenic trajectory (Fig. S4.11H-M) showed that the predominant regulatory signature of control CMs/CPs, is characterized by TFs such as *Gata4/5/6, Hoxb1, Mef2c, Foxf1, and Tbx5*, which are required for initiation of cardiac patterning and morphogenesis programs upon formation of the cardiac crescent, subsequent heart fields, and higher level organogenesis (Pikkarainen et al., 2004; Kokkinopoulos et al., 2015; Bruneau, 2013; Stefanovic et al., 2020; Harvey, 2002; Kelly, Buckingham & Moorman, 2014). *Mesp1* KO cells were unable to activate these same regulatory programs, instead retaining TFs for mesendoderm and other mesoderm networks (*T, Eomes, Hand1)* (Fig. S4.11H-M).

To visualize regulatory interactions between chromatin accessibility and integrated gene expression agnostic of differential accessibility and expression testing between specific cell types, we utilized the ArchR pipeline's orthogonal "peak2gene" linkage approach (Granja et al., 2021). This linkage prediction method identified both known and uncharacterized distal regulatory elements (Fig. 4.5E-J). The *Smarcd3*-F6 enhancer (Devine et al., 2014) expectedly showed linkage to *Smarcd3* and similar accessibility across cardiogenesis, including the *Mesp1* KO cells Mesp1ME, LPM2 (Fig. 4.5E), consistent with our detection of the transgene by scRNA-seq.

Accordant with the modular enhancer landscape of *Nkx2-5*, multiple peak linkages were defined for the *Nkx2-5* locus, including two distal uncharacterized regions (Fig. 4.5F).

Two linkages were appropriately mapped to the characterized Gata4-, Nfat-, Mesp1/Mzf1-, and Isl1-regulated 9 kb-upstream *Nkx2-5* cardiac enhancer sequence (*Nkx2-5*-AR1) (Lien et al., 1999; Chen & Cao, 2009; Clark et al., 2013; Doppler et al., 2014; Bondue et al., 2008) and the distal-linked AR1 peak was increased in control CM/CP cells only (Fig. 4.5F). Similarly, the Gata-, Smad4-, Nfat-, Isl1-regulated *Nkx2-5*-AR2 enhancer (Searcy et al., 1998; Liberatore et al., 2002; Lien et al., 2002) and the two uncharacterized linked regions ~25 kb and ~30 kb-upstream of the TSS showed enriched accessibility in control CM/CP cells (Fig. 4.5F) while Gata4- and Smad1/4-responsive 6 kb-upstream enhancer (*Nkx2-5*-GS) (Brown et al., 2004) didn't show accessibility in any cells (Fig. 4.5F). These results are consistent with absence of *Nkx2-5* in *Mesp1* KO embryos (Fig. 4.3), underscore the complexity of regulation on this critical cardiac TF, and provide further evidence for the regulatory shift between LPM1 and CM/CP cells (Fig. 4.4F) that *Mesp1* KO cells are unable to progress through.

Examination of the *Gata5* locus revealed a linkage to the characterized cardiac crescent and mesodermal derivatives enhancer (*Gata5*-CC-meso) (MacNeill et al., 2000) as well as several uncharacterized linked distal elements with accessibility in Mesp1ME, LPM2, LPM1, and CM/CP cells (Fig. 4.5G). Several characterized *Gata4* enhancer regions were linked, including lateral mesoderm enhancer *Gata4*-G2 (Rojas et al., 2005) and cardiac crescent enhancer *Gata4*-G9 (Schachterle et al., 2012). Foxf1 and Gata4-bound enhancer *Gata4*-C2 showed enriched accessibility in mesendoderm *Mesp1* KO cells, while ETS-activated *Gata4*-G9 was similarly accessible between between Mesp1ME, LPM2, LPM1, and CM/CPs (Fig. 4.5H), highlighting retention of active chromatin states preceding cardiac patterning and differentiation despite loss of *Mesp1*.

Evaluation of loci for mesendoderm genes *Eomes* and *Tdgf1,* which ectopically perdure in *Mesp1* KO embryos, showed a corresponding pattern of enriched linked peaks in *Mesp1* KO cell types (Fig. 4.5I). The characterized distal element *Meteor*, a lncRNA (Alexanian et al., 2017), was linked to *Eomes* with greatest accessibility enrichment in Epi, EomesME, and Mesp1ME *Mesp1* KO cells (Fig. 4.5I). Similarly, characterized PSEa, PSEb, and VME regulatory regions (Simon et al., 2017) were linked (Fig. 4.5I), supporting that the early cardiac mesoderm transcriptional landscape is intact despite *Mesp1* absence, however retained later in development than it should be for the age of these embryos. Upstream of *Tdgf1*, a previously characterized enhancer sequence and direct transcriptional target of Mef2c (Barnes et al., 2016) displayed enrichment of proximal peaks in *Mesp1* KO cells (Fig. 4.5J), which we confirmed by increased *Tdgf1* enhancer transgene activity in posterior domains of E7.5 *Mesp1* KO embryos (Fig. 4.5K). Increased *Tdgf1* enhancer activity mimicked the enriched *Tdgf1* gene expression in *Mesp1* KO embryos (Fig. 4.1F-G), further supporting the hypothesis that early programs are de-repressed in absence of *Mesp1*.

We examined additional linked peak profiles around differentially expressed genes (Fig. S4.12). We detected linkages between *Mesp1* and the characterized “EME” enhancer (Haraguchi et al., 2001; Ajima et al., 2021; Costello et al., 2011; Guo et al., 2018) with enrichment in *Mesp1* KO cell types likely indicative of retained early chromatin landscape or de-repression of the locus without appropriate regulation from downstream targets (Fig. S4.12A). We detected linkages to 3 uncharacterized distal regions near the *Gata6* locus, as well as the Nkx2-5-targeted enhancer regions (Molkentin et al., 2000) which had similar accessibility profiles across *Mesp1* KO Mesp1ME, LPM2 cells, and

the LPM1 cells containing both genotypes (Fig. S4.12B). We noted linkages to multiple characterized *Hand1* enhancer regions (Vincentz et al., 2021, 2019; George & Firulli, 2021) across both genotypes and multiple cell types, with accessibility for some enhancers decreasing in CM/CP cells (Fig. S4.12C), consistent *Hand1*'s more robust activity in LPM1 cells (Fig. 4.4D,E,H). Peaks with similar accessibility across Mesp1ME and LPM cell types containing both genotypes were detected in linkages near *Tbx5,* including the Tbx5-CRE16 (Smemo et al., 2012), and downregulated but retained structural myocyte genes *Tnnt2* (Fig. S4.12E) and *Myl7* (Fig. S4.12F). Increased accessibility for *Anxa2*-linked peaks in *Mesp1* KO LPM2 cells and control/*Mesp1* KO LPM1 cells contrasted near-inaccessibility in control CM/CPs (Fig. S4.12G), consistent with the upregulated *Anxa2* expression in Late-stage embryo cardiogenic regions (Fig. 4.3K). Downstream distal peaks were linked to Mesp1-induced EMT-gene *Snai1* (Fig. S4.12H) (Lin et al., 2022), including the Mesp1-binding site. These peak2gene linkage analyses further illustrate the correlation between differentially expressed genes and the altered chromatin landscape in *Mesp1* KO mesoderm cells that prevents progression towards more mature cardiac fates, and also highlights that despite this disrupted regulatory landscape, some distal-regulatory elements relevant to early cardiogenesis are still retained in their active states.

Discussion

We generated scRNA-seq and scATAC-seq datasets from whole mouse embryos in a timeline of gastrulation, creating a valuable *in vivo* resource for high-resolution studies of gene regulatory networks in early embryonic development. We utilized computational detection of the CPC-labeling transgenes to focus on early cardiac specification, showing that while *Mesp1* KO embryos are capable of initiating and progressing through early cardiac mesoderm specification, a *Mesp1*-dependent regulatory barrier prevents *Mesp1* KO CPCs from progressing completely towards CM fates. We characterized improper repression of early mesendoderm programs at this breakpoint, such as how absence of *Mesp1* leads to enduring *Eomes* activity, which in turn promotes ectopic perdurance of mesendoderm transcriptional networks when cardiac crescent-staged embryos should instead be upregulating cardiac patterning programs. Additionally, this disrupted regulatory landscape likely contributes to *Mesp1* KO cardiac mesoderm and CPCs ectopically expressing non-cardiac mesoderm genes. Despite this ectopic expression, CPCs do not appear to deviate from a cardiac-directed mesodermal fate path. Ultimately, while *Mesp1* KO embryos specify early cardiac lineage cell types, the deficient regulatory landscapes within *Mesp1* KO CPCs prove prohibitive against further lineage development (Fig. 4.6).

Positing *Mesp1* as a master transcriptional regulator of early cardiac fate is largely informed by overexpression studies (Chiapparo et al., 2016; Bondue et al., 2008; Lindsley et al., 2008; Wu, 2008; Kelly, 2016; Bondue & Blanpain, 2010; Lin et al., 2022) in contrast to earlier *in vivo* studies which suggested a *Mesp1*-dependent role for

cardiac mesoderm migration (Saga, Kitajima & Miyagawa-Tomita, 2000). Indeed, in this work we note downregulation of migratory genes in *Mesp1* KO cells, and a companion work demonstrates that *Mesp1*-dependent migration patterns are critical for spatial organization of CPCs during cardiogenesis (Dominguez et al., 2022). Additional interpretations in *Mesp1/Mesp2* double knockouts underscore the potential for more complex networks of TF dependency in cardiac specification not fully explained by regulatory hierarchies (Ajima et al., 2021; Kitajima et al., 2000; Saga, 1998). While the concept of a "master transcription factor" is a broadly applied hierarchical framework for interrogation of gene regulatory networks (Cai et al., 2020; Davis & Rebay, 2017; Yin & Wang, 2014), and *Mesp1*'s coincident expression in emerging CPCs supports an instructive role for *Mesp1* in cardiogenesis, this model likely oversimplifies cardiogenesis. Indeed, our high resolution, single cell transcriptional and epigenomic analyses reveal both transcriptional resilience and vulnerability of early cardiogenesis in a regulatory landscape lacking *Mesp1*.

Transcriptional profiling of *Smarcd3*-F6+ cells highlighted that *Mesp1* KO cells were mostly represented in cell types of early cardiogenesis and in Early- and Middle-stage embryos, indicating that *Mesp1* KO CPCs not only initiate but also progress through early stages of cardiac specification. This finding contrasts with the *Mesp1*-dependent failure to exit pluripotency previously highlighted (Lescroart et al., 2018). We interpret the failed induction of *Nkx2-5*, which is critical for patterning of the first and second heart fields (Harvey, 2002), and the inappropriate levels of *Gata* factors in *Mesp1* KO CPCs as representing a breakpoint between phases of the cardiogenic process.

To characterize this breakpoint, we utilized complementary scATAC-seq and scRNA-seq mesoderm datasets to conclude that mesendoderm regulatory programs, instructed at least partially by *Eomes,* are responsible for the initiation and progression through Middle phases of cardiac specification prior to cardiac crescent formation. However, the perdurance of these programs coupled with the failure of LPM to properly migrate anterior-laterally in *Mesp1* KO embryos leads to aberrant upregulation and ectopic expression of early cardiac mesoderm, non-cardiac mesoderm, and mesendoderm genes and TFs. Additionally, we hypothesize that the posterior positioning of CPCs in *Mesp1* KO embryos further compounds cardiac maturation and CPC transcriptional profiles via improper exposure to signaling gradients and growth factors. The dysregulated identity of *Mesp1* KO cardiac mesoderm in this phase between Middle and Late embryonic stages stalls *Mesp1*-deficient cardiogenesis due to failed induction of cardiac progenitor patterning, morphogenesis, and CM maturation regulatory programs.

Although this developmental breakpoint is observed between E7.5-E7.75, well after transient *Mesp1* expression has declined, these processes appear to be *Mesp1*-dependent. Possible explanations for this phenomenon are 1) improper repression of earlier regulators, such as *Eomes*; 2) compounded, *Mesp1*-dependent secondary effects influencing de-repression or ectopic activation; or 3) *Mesp1* KO CPCs are exposed to improper embryonic signaling cues as a result of their aberrant posterior localization within the embryo. Future studies with additional genetic models and perturbations along with assays of embryos representing earlier developmental timepoints are needed to disentangle these possibilities.

Overall, our work shows that complex transcriptional networks and interdependent hierarchies govern CPC emergence and differentiation. We characterize an initial, transcriptionally resilient, phase of CPC specification and identify that the epigenomic landscape necessary for CPCs to transition from LPM to CPs and CMs is dependent on upstream Mesp1 activity. Our results point to generalizable transcriptional regulatory principles during gastrulation for the allocation of precursor cells from embryonic germ layers towards restricted fates, and their differentiation to distinct functional cell types.

Figures

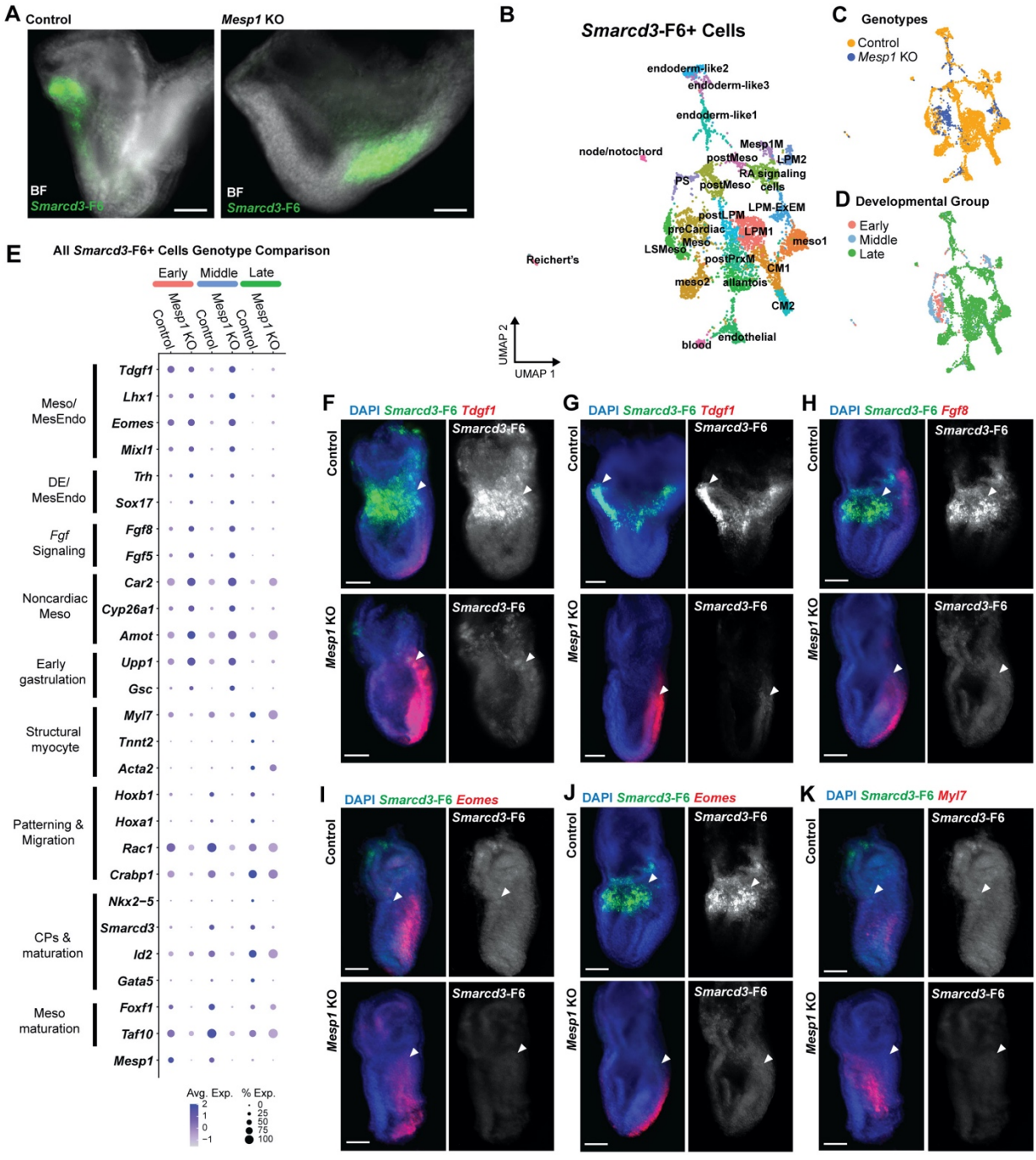

A
Control
Mesp1 KO
BF
Smarcd3-F6
B
Smarcd3-F6+ Cells
endoderm-like2
endoderm-like3
endoderm-like1
node/notochord
Mesp1M
postMeso
LPM2
RA signaling
PS
postMeso
cells
LPM-ExEM
postLPM
preCardiac
LPM1
meso1
LSMeso
Meso
Reichert's
postPrxM
CM1
meso2
allantois
CM2
endothelial
blood
UMAP 2
UMAP 1
C
Genotypes
Control
Mesp1 KO
D
Developmental Group
Early
Middle
Late
E
All Smarcd3-F6+ Cells Genotype Comparison
Early
Middle
Late
Control
Mesp1 KO
Meso/ MesEndo
Tdgf1
Lhx1
Eomes
Mixl1
DE/ MesEndo
Trh
Sox17
Fgf Signaling
Fgf8
Fgf5
Noncardiac Meso
Car2
Cyp26a1
Amot
Early gastrulation
Upp1
Gsc
Structural myocyte
Myl7
Tnnt2
Acta2
Patterning & Migration
Hoxb1
Hoxa1
Rac1
Crabp1
CPs & maturation
Nkx2-5
Smarcd3
Id2
Gata5
Meso maturation
Foxf1
Taf10
Mesp1
Avg. Exp.
% Exp.
F
DAPI Smarcd3-F6 Tdgf1
G
DAPI Smarcd3-F6 Tdgf1
H
DAPI Smarcd3-F6 Fgf8
I
DAPI Smarcd3-F6 Eomes
J
DAPI Smarcd3-F6 Eomes
K
DAPI Smarcd3-F6 Myl7
Smarcd3-F6

Figure 4.1 Transcriptional profiles of *Smarcd3*-F6+ cells in *Mesp1* KO embryos. (A) Fluorescence *in situ* hybridization for *Smarcd3*-F6 expression (green) in cardiac crescent stage (E7.75) *Mesp1* KO and control littermate embryos. (B) UMAP atlas of 4,868 *Smarcd3*-F6+ cells representing 24 cell types. (C-D) UMAPs colored by (C) genotype and (D) relative developmental stages, Early (E6.0-E6.5), Middle (late E6.5 – E7.5), Late (late E7.5- early E7.75). (E) Dotplot representation of gene expression across genotypes at relative developmental stages. Size of dot denotes percent of cells expressing gene, color of dot represents average gene expression. (F-K) Multiplexed fluorescence *in situ* hybridization for *Smarcd3*-F6 (green) and (F-G) *Tdgf1* (red) in representative (F) Early and (G) Middle stages, (H) *Fgf8* (red) in Middle stages, (I-J) *Eomes* (red) in (I) Early and (J) Middle stages, (K) *Myl7* (red) in Early stages. Arrowheads denote *Smarcd3*-F6+ cardiogenic regions. Scale bars are 100 μm.

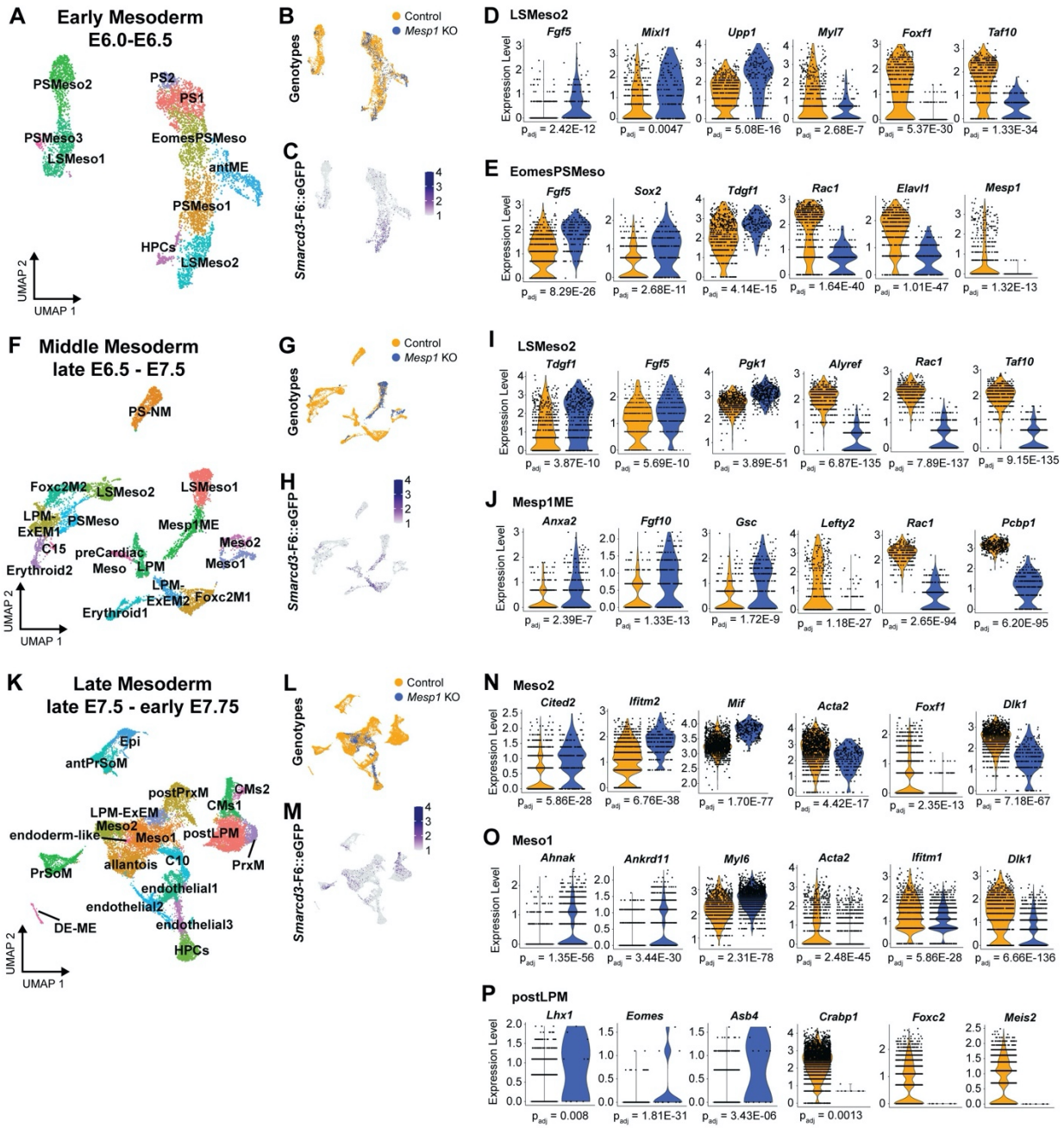


Figure 4.2 Transcriptional profiles of cardiac mesoderm in *Mesp1* KO embryos. Mesoderm scRNA-seq UMAP atlases for (A) Early (5,504 cells), (F) Middle (7,666 cells), and (K) Late (22,622 cells) developmental stages. Associated UMAPs for each stage atlas colored by (B, G, L) genotype and (C, H, M) *Smarcd3*-F6-eGFP expression. Differentially expressed genes in Early mesoderm in (D) LSMeso2 and (E) EomesPSMeso. Differentially expressed genes in Middle mesoderm in (I) LSMeso2 and (J) Mesp1ME. Differentially expressed genes in Late mesoderm in (N) Meso2, (O) Meso1, and (P) postLPM. Significant changes denoted with adjusted p values < 0.05.

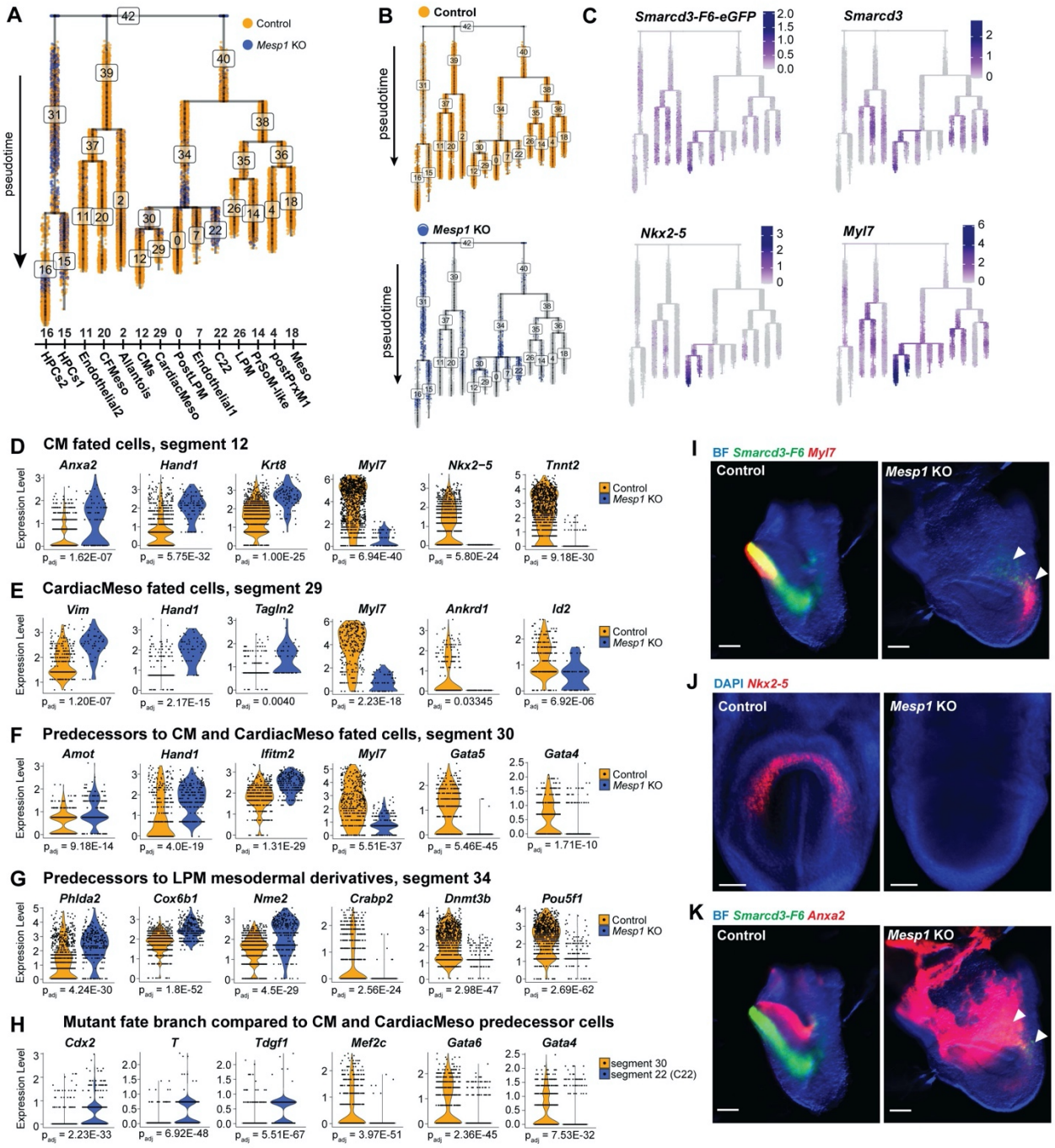

A
Control
Mesp1 KO
pseudotime
16 15 11 20 2 12 29 0 7 22 26 14 4 18
HPCs2
HPCs1
Endothelial2
CFMeso
Allantois
CMs
CardiacMeso
postLPM
Endothelial1
C22
LPM
PrSoM-like
postPrxM1
Meso
B
Control
Mesp1 KO
pseudotime
C
Smarcd3-F6-eGFP
Smarcd3
Nkx2-5
Myl7
D
CM fated cells, segment 12
Anxa2
Hand1
Krt8
Myl7
Nkx2−5
Tnnt2
Expression Level
padj = 1.62E-07
padj = 5.75E-32
padj = 1.00E-25
padj = 6.94E-40
padj = 5.80E-24
padj = 9.18E-30
E
CardiacMeso fated cells, segment 29
Vim
Hand1
Tagln2
Myl7
Ankrd1
Id2
padj = 1.20E-07
padj = 2.17E-15
padj = 0.0040
padj = 2.23E-18
padj = 0.03345
padj = 6.92E-06
F
Predecessors to CM and CardiacMeso fated cells, segment 30
Amot
Hand1
Ifitm2
Myl7
Gata5
Gata4
padj = 9.18E-14
padj = 4.0E-19
padj = 1.31E-29
padj = 5.51E-37
padj = 5.46E-45
padj = 1.71E-10
G
Predecessors to LPM mesodermal derivatives, segment 34
Phlda2
Cox6b1
Nme2
Crabp2
Dnmt3b
Pou5f1
padj = 4.24E-30
padj = 1.8E-52
padj = 4.5E-29
padj = 2.56E-24
padj = 2.98E-47
padj = 2.69E-62
H
Mutant fate branch compared to CM and CardiacMeso predecessor cells
Cdx2
T
Tdgf1
Mef2c
Gata6
Gata4
padj = 2.23E-33
padj = 6.92E-48
padj = 5.51E-67
padj = 3.97E-51
padj = 2.36E-45
padj = 7.53E-32
segment 30
segment 22 (C22)
I
BF Smarcd3-F6 Myl7
Control
Mesp1 KO
J
DAPI Nkx2-5
Control
Mesp1 KO
K
BF Smarcd3-F6 Anxa2
Control
Mesp1 KO

Figure 4.3. Pseudotime trajectory for mesoderm fates in *Mesp1* KO embryos. (A) URD pseudotime tree for fate progression towards mature mesoderm fates colored by genotypes together and (B) separately. (C) Overlay of cardiac marker gene expression. (D-H) Differentially expressed genes in cells of shared fates and pseudotime identities; (D) CM fated cells, (E) CardiacMeso fated cells, (F) predecessors to CM and CardiacMeso fates, (G) predecessors to LPM derivate fates, (H) comparison of mutant fate branch C22 to predecessors to CardiacMeso fates. (I-J) Multiplexed fluorescence *in situ* hybridization for *Smarcd3*-F6 (green) and (I) *Myl7* (red) and (J) *Nkx2-5* (red), and (K) *Anxa2* in cardiac crescent stage embryos. Arrowheads denote *Smarcd3*-F6+ cardiogenic regions in *Mesp1* KO embryos. Scale bars are 100 μm.

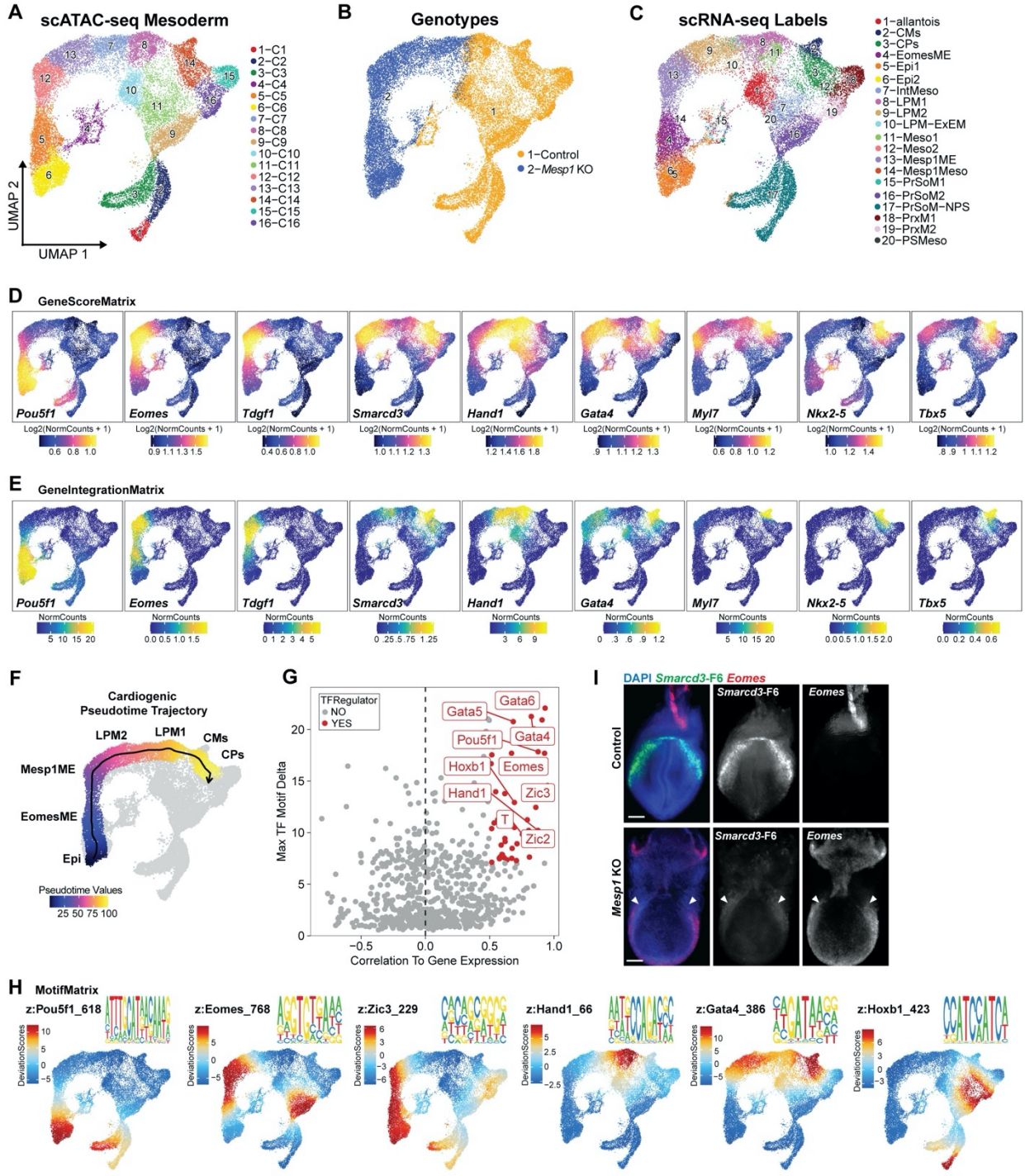

A
scATAC-seq Mesoderm
1−C1
2−C2
3−C3
4−C4
5−C5
6−C6
7−C7
8−C8
9−C9
10−C10
11−C11
12−C12
13−C13
14−C14
15−C15
16−C16
UMAP 2
UMAP 1
B
Genotypes
1−Control
2−Mesp1 KO
C
scRNA-seq Labels
1−allantois
2−CMs
3−CPs
4−EomesME
5−Epi1
6−Epi2
7−IntMeso
8−LPM1
9−LPM2
10−LPM−ExEM
11−Meso1
12−Meso2
13−Mesp1ME
14−Mesp1Meso
15−PrSoM1
16−PrSoM2
17−PrSoM−NPS
18−PrxM1
19−PrxM2
20−PSMeso
D
GeneScoreMatrix
Pou5f1
Eomes
Tdgf1
Smarcd3
Hand1
Gata4
Myl7
Nkx2-5
Tbx5
Log2(NormCounts + 1)
E
GeneIntegrationMatrix
NormCounts
F
Cardiogenic
Pseudotime Trajectory
LPM2
LPM1
CMs
CPs
Mesp1ME
EomesME
Epi
Pseudotime Values
25 50 75 100
G
TFRegulator
NO
YES
Gata5
Gata6
Pou5f1
Gata4
Hoxb1
Eomes
Hand1
Zic3
T
Zic2
Max TF Motif Delta
Correlation To Gene Expression
I
DAPI Smarcd3-F6 Eomes
Control
Mesp1 KO
Smarcd3-F6
Eomes
H
MotifMatrix
z:Pou5f1_618
z:Eomes_768
z:Zic3_229
z:Hand1_66
z:Gata4_386
z:Hoxb1_423
DeviationScores

Figure 4.4. Transcriptional drivers in *Mesp1* KO mesoderm. (A) Mesoderm scATAC-seq atlas of 16 cell types with overlays for (B) genotype and (C) relative cell types from integration of a complementary scRNA-seq dataset. (D) GeneScoreMatrix plots for chromatin accessibility around gene loci and (E) GeneIntegrationMatrix plots for scRNA-seq integrated gene expression for cardiac mesoderm marker genes and TFs. (F) Pseudotime values for cells along the *Mesp1* KO cardiac-fate trajectory path. (G) Maximum z-score delta for TF motif variance between clusters correlated to gene expression within clusters to identify positive TF drivers (red). (G-H) Highlighted positive regulator TFs' motif z-scores mapped in UMAP space, with associated position weight matrix plots. (I) Multiplexed fluorescence *in situ* hybridization for *Smarcd3*-F6 (green) and *Eomes* (red) in cardiac crescent stage embryos. Arrowheads denote cardiogenic regions in *Mesp1* KO embryo. Scale bars are 100 μm.

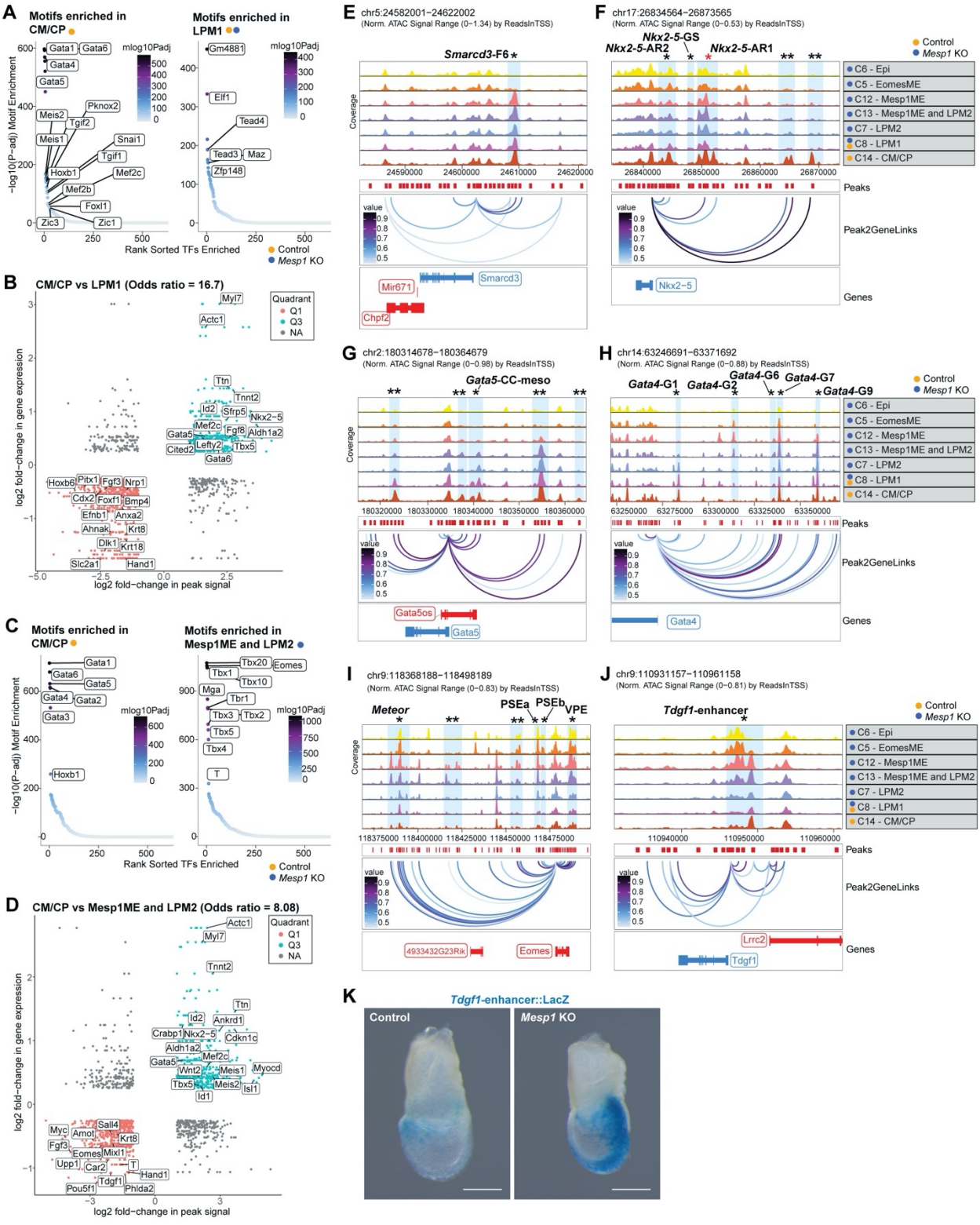

A
Motifs enriched in CM/CP
Motifs enriched in LPM1
−log10(P−adj) Motif Enrichment
mlog10Padj
Rank Sorted TFs Enriched
Control
Mesp1 KO
B
CM/CP vs LPM1 (Odds ratio = 16.7)
Quadrant
Q1
Q3
NA
log2 fold−change in gene expression
log2 fold−change in peak signal
C
Motifs enriched in CM/CP
Motifs enriched in Mesp1ME and LPM2
D
CM/CP vs Mesp1ME and LPM2 (Odds ratio = 8.08)
E
chr5:24582001−24622002
(Norm. ATAC Signal Range (0−1.34) by ReadsInTSS)
Smarcd3-F6
Coverage
Peaks
Peak2GeneLinks
Genes
F
chr17:26834564−26873565
(Norm. ATAC Signal Range (0−0.53) by ReadsInTSS)
Nkx2-5-GS
Nkx2-5-AR2
Nkx2-5-AR1
C6 - Epi
C5 - EomesME
C12 - Mesp1ME
C13 - Mesp1ME and LPM2
C7 - LPM2
C8 - LPM1
C14 - CM/CP
G
chr2:180314678−180364679
(Norm. ATAC Signal Range (0−0.98) by ReadsInTSS)
Gata5-CC-meso
H
chr14:63246691−63371692
(Norm. ATAC Signal Range (0−0.88) by ReadsInTSS)
Gata4-G1
Gata4-G2
Gata4-G6
Gata4-G7
Gata4-G9
I
chr9:118368188−118498189
(Norm. ATAC Signal Range (0−0.83) by ReadsInTSS)
Meteor
PSEa
PSEb
VPE
J
chr9:110931157−110961158
(Norm. ATAC Signal Range (0−0.81) by ReadsInTSS)
Tdgf1-enhancer
K
Tdgf1-enhancer::LacZ
Control
Mesp1 KO

Figure 4.5. Disrupted regulatory landscape of *Mesp1* KO mesoderm. (A,C) Motifs enriched in differentially accessible peaks between (A) CM/CP vs. LPM1 and (C) CM/CP vs Mesp1ME and LPM2. (B,D) Plots for peak,gene associations showing correlations between differential peak accessibility and gene expression in comparisons between cells type1 vs type2. Q3 peak,gene pairs represent significantly more accessible peaks paired with upregulated gene expression in type1 cells. Q1 peak,gene pairs represent significantly more accessible peaks paired with upregulated gene expression in type2 cells. Odds ratio denotes probability for observed peak,gene relationships. (B) Peak,gene association plot for cells in CM/CP vs LPM1 comparison and (D) in CM/CP vs Mesp1ME and LPM2 comparison. (E-J) Peak2Gene linkage browser tracks for cell types showing predicted regulatory connections between distal accessible regions (Peaks) and nearby genes. Shaded bars denote predicted distal regulatory regions; *denotes characterized elements; red* denotes regions with Mesp1-binding; **denotes uncharacterized elements. Characterized elements named when available. (E) *Smarcd3* linkage to the “F6” enhancer. Peak linkages to genes (F) *Nkx2-5*, (G) *Gata5*, (H) *Gata4*, (I) *Eomes*, and (J) *Tdgf1.* (K) X-gal stain for activity of characterized *Tdgf1* enhancer, scale bars are 200 μm.

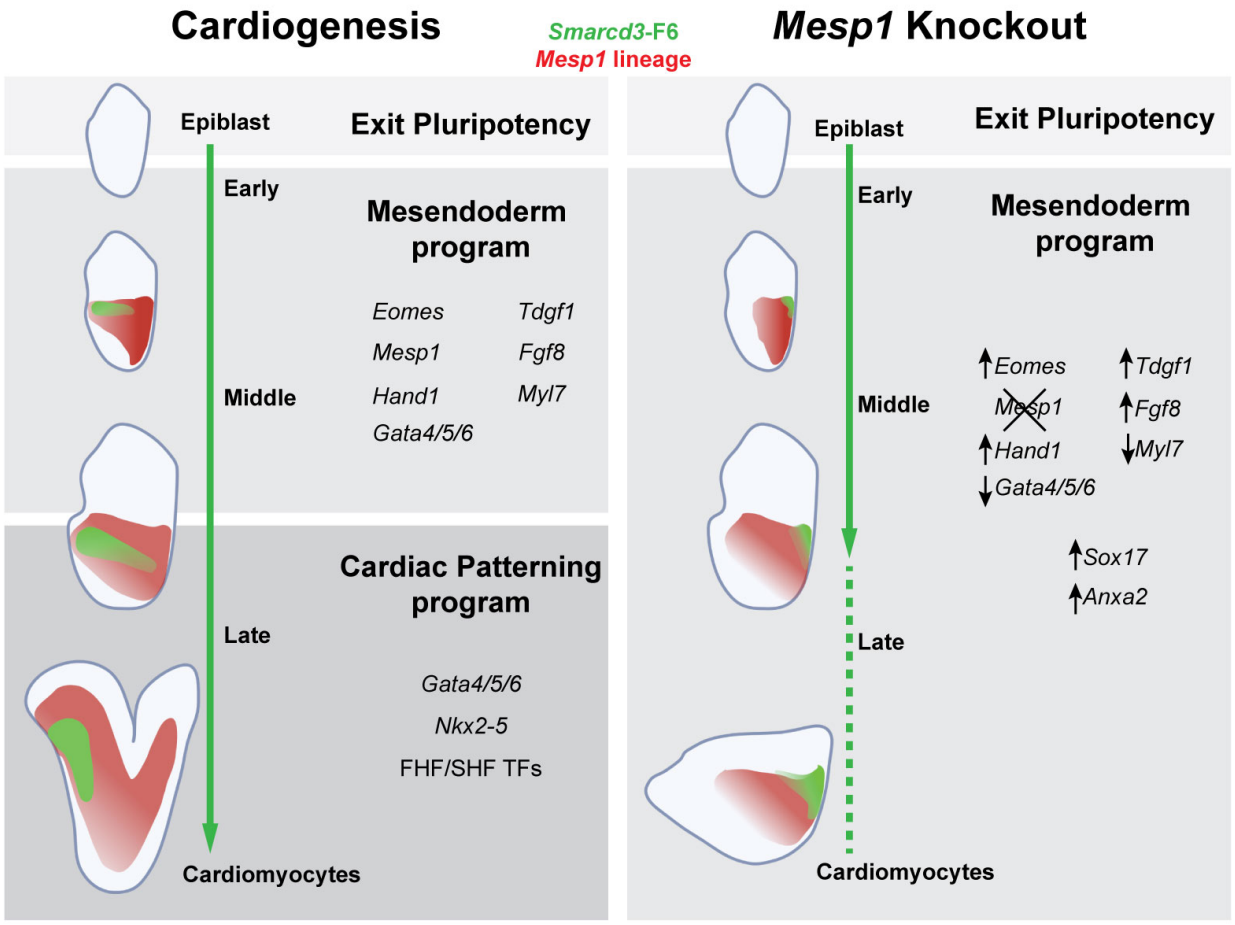
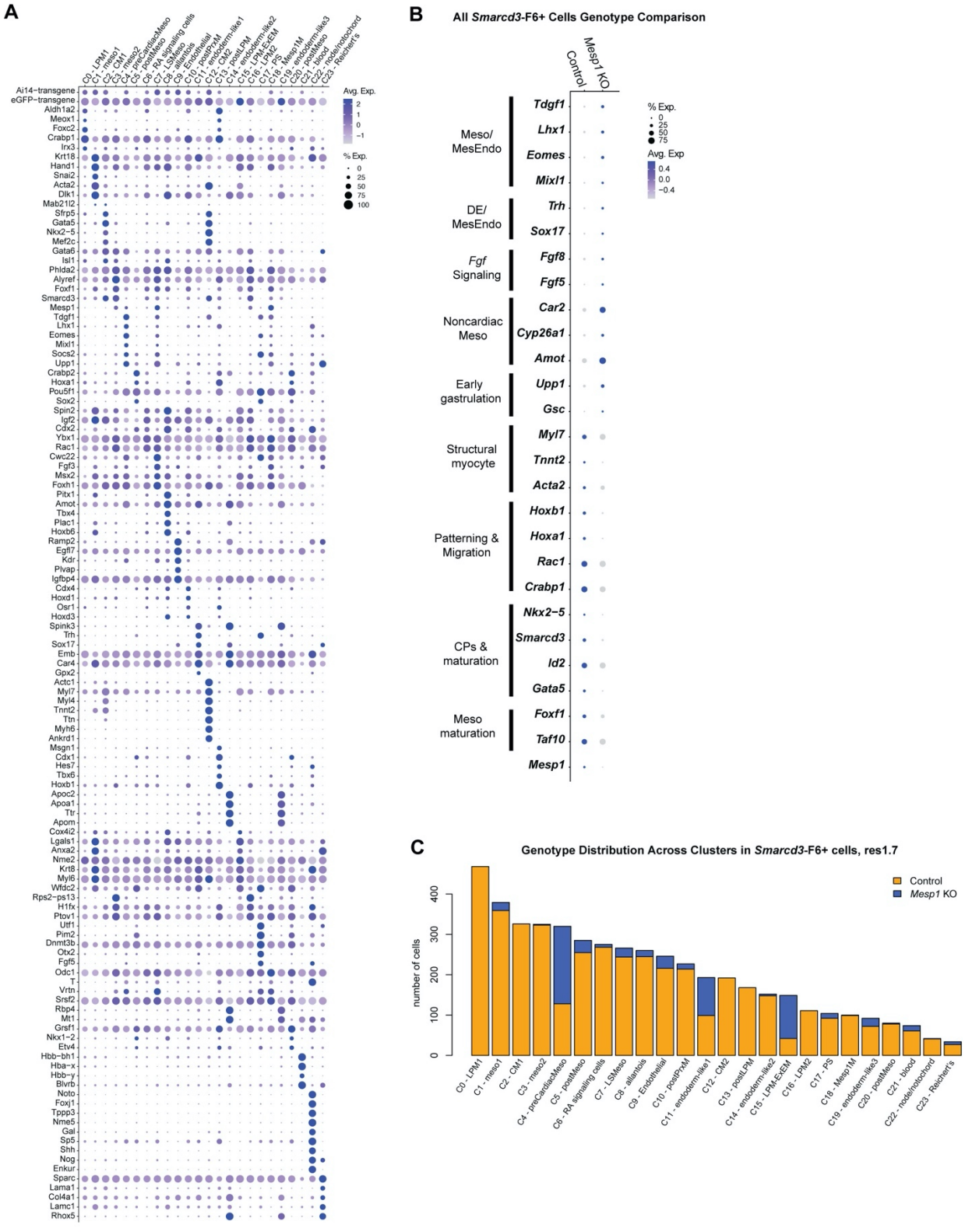
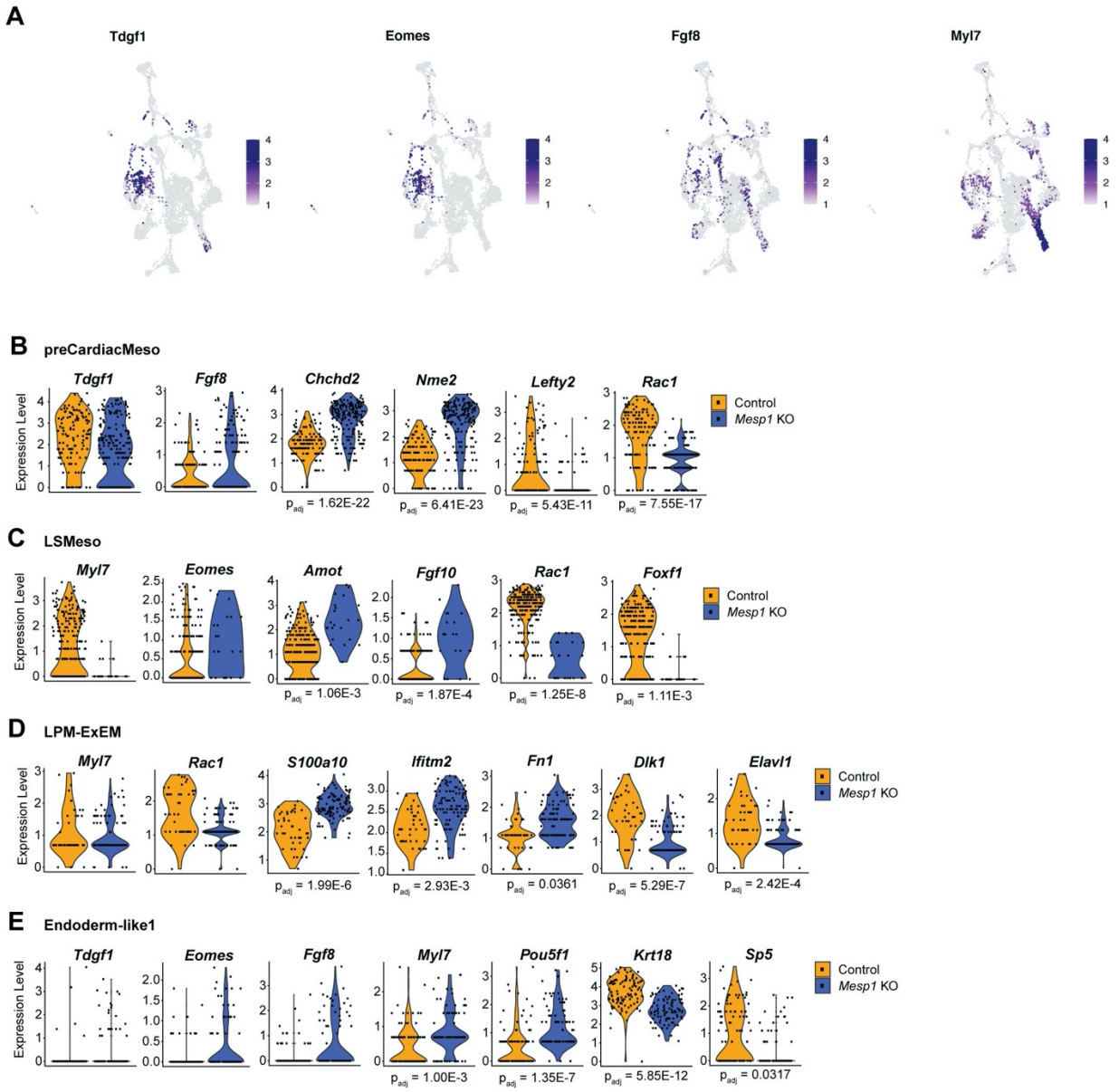


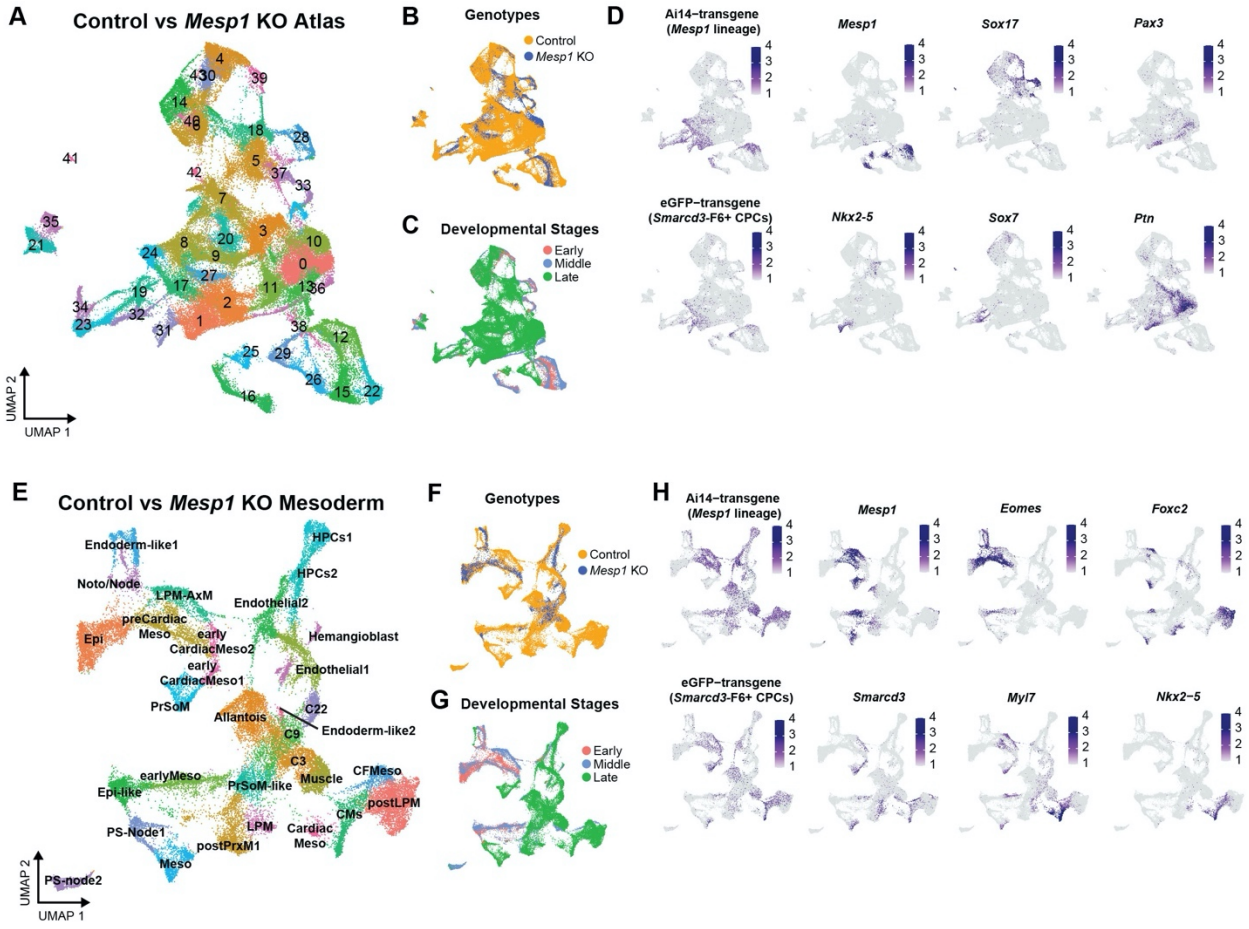
Figure 4.6. Model for transcriptional regulatory landscape of cardiogenesis and loss of *Mesp1*. Schematic model of gene regulatory program phases during cardiac mesoderm specification and differentiation. *Mesp1* KO cardiac mesoderm cells exit pluripotency, induce early cardiac specification genes under control of mesendoderm programs, yet fail to activate critical cardiac TFs at cardiac crescent stages to initiate cardiac patterning programs.


A
B
All Smarcd3-F6+ Cells Genotype Comparison
Control
Mesp1 KO
Meso/ MesEndo
DE/ MesEndo
Fgf Signaling
Noncardiac Meso
Early gastrulation
Structural myocyte
Patterning & Migration
CPs & maturation
Meso maturation
Tdgf1
Lhx1
Eomes
Mixl1
Trh
Sox17
Fgf8
Fgf5
Car2
Cyp26a1
Amot
Upp1
Gsc
Myl7
Tnnt2
Acta2
Hoxb1
Hoxa1
Rac1
Crabp1
Nkx2−5
Smarcd3
Id2
Gata5
Foxf1
Taf10
Mesp1
% Exp.
Avg. Exp
C
Genotype Distribution Across Clusters in Smarcd3-F6+ cells, res1.7
Control
Mesp1 KO
number of cells

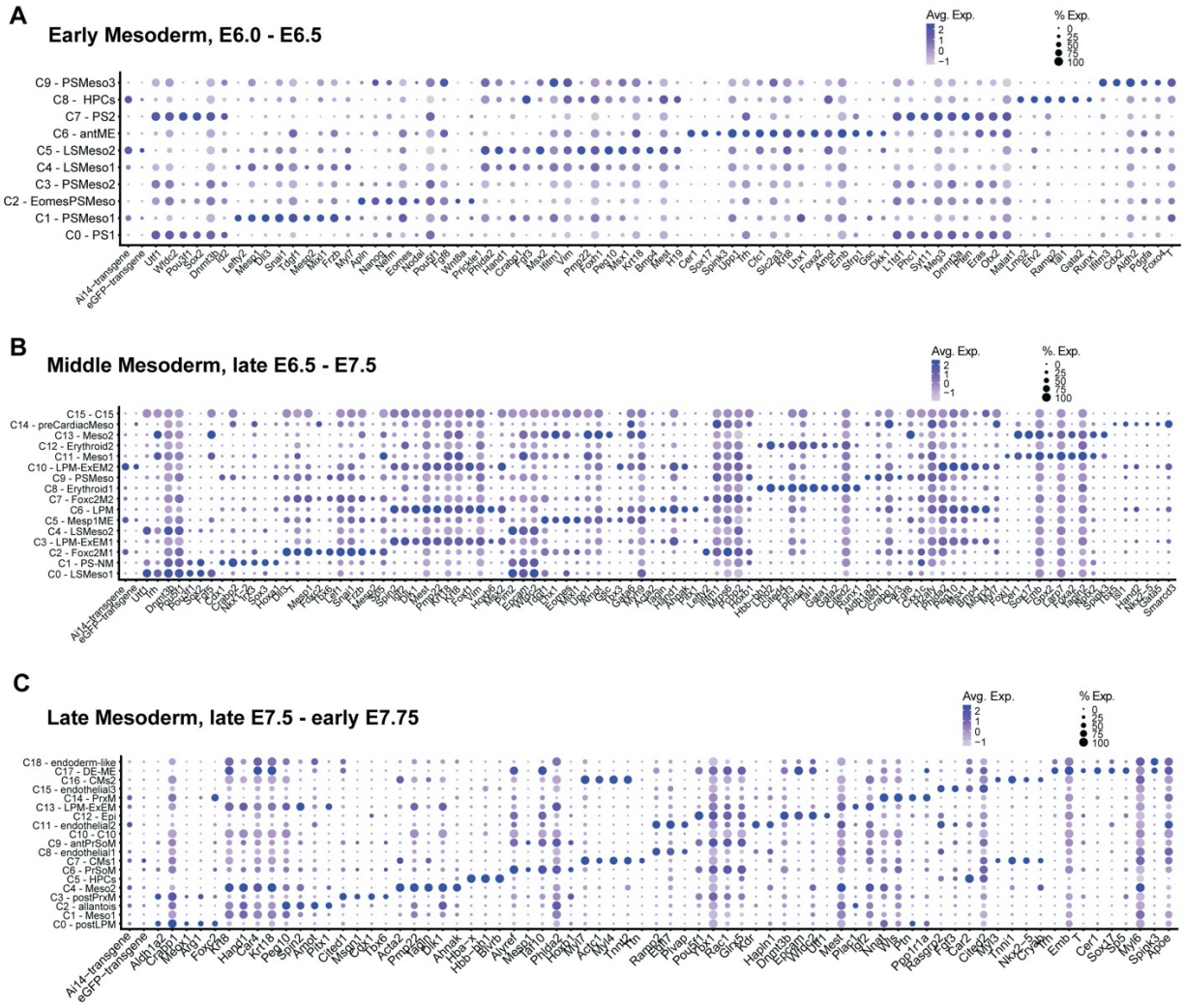
Supplementary Figure 4.1. *Smarcd3*-F6+ cell type cluster annotations. (A) Dotplot denoting marker genes and cell type annotations by cluster in *Smarcd3*-F6+ cells atlas. (B) Dotplot representation of differential gene expression between genotypes across all cells. Size of dot denotes percent of cells expressing gene, color of dot represents average gene expression. (C) Barplot denoting distribution of number of cells from genotypes across cluster identities for *Smarcd3*-F6+ atlas.



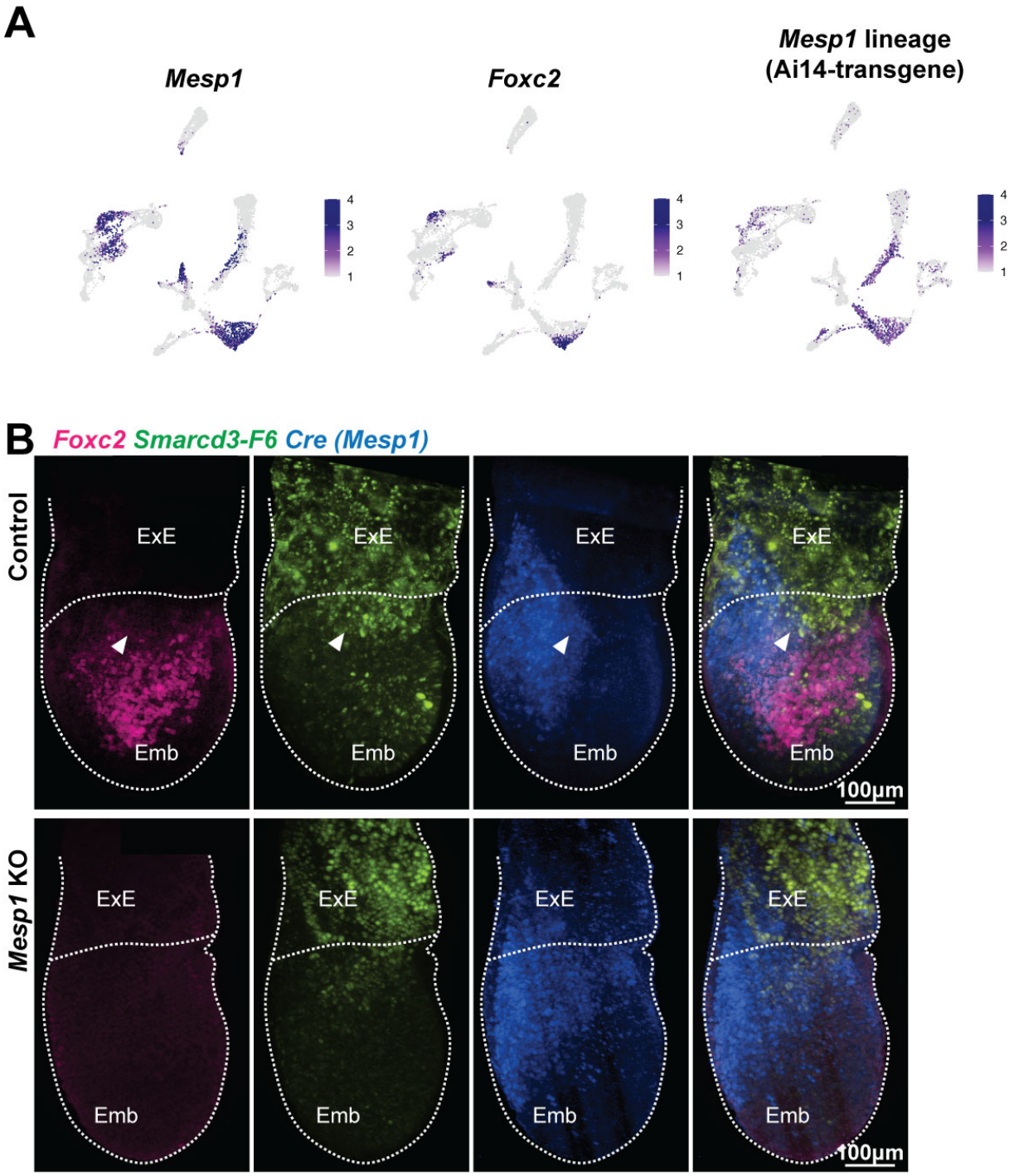
Supplementary Figure 4.2. Differentially expressed genes in *Smarcd3*-F6+ cells from *Mesp1* KO embryos. (A) Overlay of gene expression in UMAP space for early cardiac marker genes *Tdgf1*, *Eomes*, *Fgf8*, *Myl7*. (B-E) Differential gene expression profiling highlights similar cardiac marker gene expression between genotypes in (B) preCardiacMeso, (C) LSMeso, (D)LPM-ExEM, and (E) endoderm-like1 cells. (B-E) Differentially expressed genes plotted with adj p values < 0.05.


A
Control vs Mesp1 KO Atlas
UMAP 2
UMAP 1
B
Genotypes
Control
Mesp1 KO
C
Developmental Stages
Early
Middle
Late
D
Ai14−transgene (Mesp1 lineage)
Mesp1
Sox17
Pax3
eGFP−transgene (Smarcd3-F6+ CPCs)
Nkx2-5
Sox7
Ptn
E
Control vs Mesp1 KO Mesoderm
Endoderm-like1
HPCs1
HPCs2
Noto/Node
LPM-AxM
Endothelial2
preCardiac Meso
Epi
early CardiacMeso2
Hemangioblast
early CardiacMeso1
Endothelial1
PrSoM
Allantois
C22
Endoderm-like2
C9
C3
earlyMeso
Muscle
CFMeso
Epi-like
PrSoM-like
postLPM
CMs
PS-Node1
LPM
Cardiac Meso
postPrxM1
Meso
PS-node2
F
Genotypes
Control
Mesp1 KO
G
Developmental Stages
Early
Middle
Late
H
Ai14−transgene (Mesp1 lineage)
Mesp1
Eomes
Foxc2
eGFP−transgene (Smarcd3-F6+ CPCs)
Smarcd3
Myl7
Nkx2−5
I
C30 - Endoderm-like2
C29 - CardiacMeso
C28 - earlyCardiacMeso2
C27 - earlyCardiacMeso1
C26 - LPM
C25 - Hemangioblast
C24 - Noto/Node
C23 - PS-node2
C22 - C22
C21 - PS-node1
C20 - CFMeso
C19 - Endoderm-like1
C18 - Meso
C17 - PrSoM
C16 - HPCs2
C15 - HPCs1
C14 - PrSoM-like
C13 - LPM-AxM
C12 - CMs
C11 - Endothelial2
C10 - Epi-like
C9 - C9
C8 - earlyMeso
C7 - Endothelial1
C6 - Muscle
C5 - preCardiacMeso
C4 - postPrxM1
C3 - C3
C2 - allantois
C1 - Epi
C0 - postLPM
% Exp.
Avg. Exp.

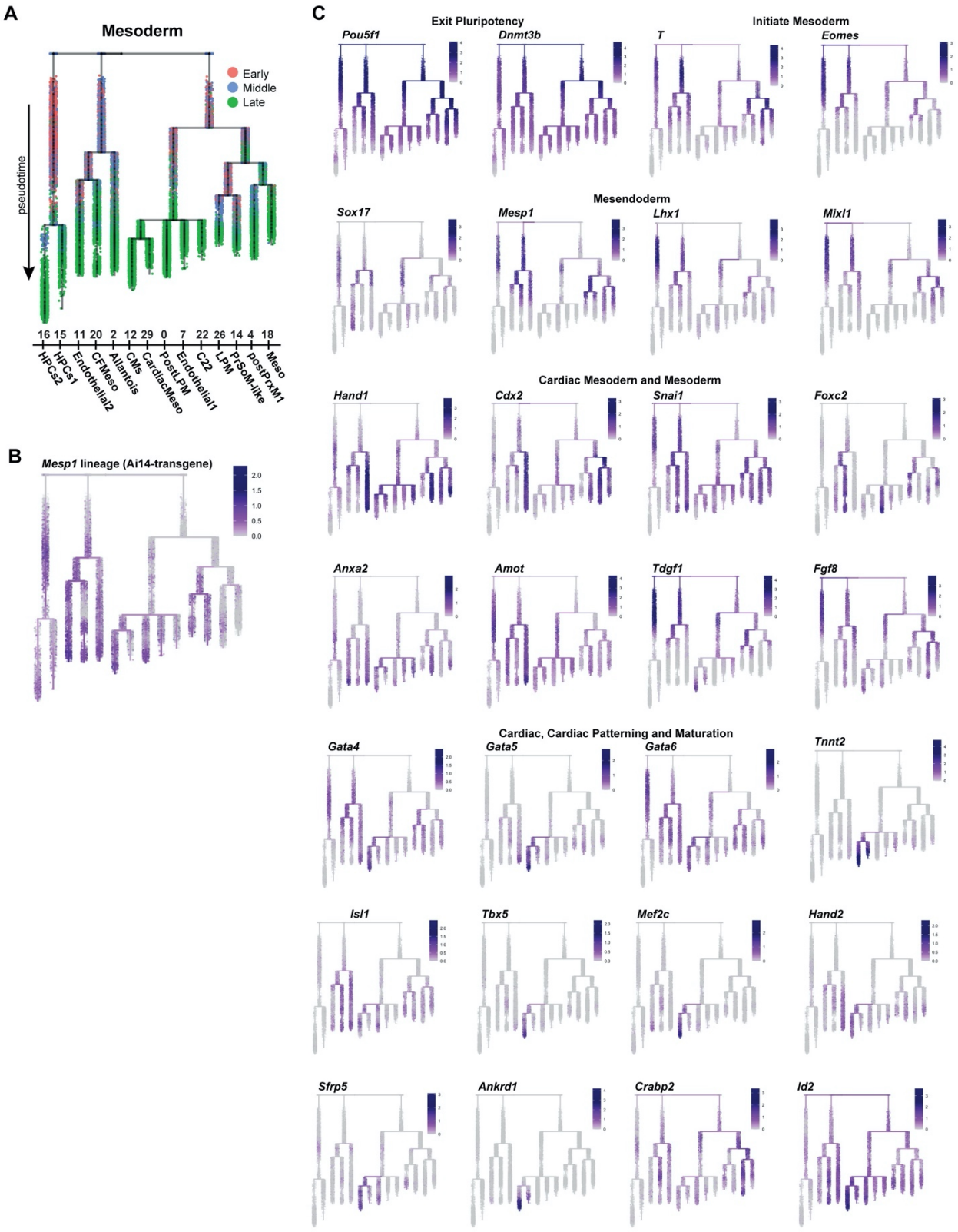
Supplementary Figure 4.3. Identification of emerging cardiac mesoderm in control and *Mesp1* KO scRNA-seq data. (A) Atlas UMAP of 96,027 cells representing whole embryos with overlay of (B) genotypes (C) relative developmental stages Early, Middle, and Late. (D) UMAPs showing expression of *Mesp1* lineage transgene Ai14, CPC-specific *Smarcd3*-F6 transgene eGFP, cardiac mesoderm markers *Mesp1*, *Nkx2-5,* endoderm markers *Sox17*, S*ox7*, neural markers *Pax3*, *Ptn*. (E) Atlas UMAP of 35,792 mesoderm cells with overlay of (F) genotypes and (E) relative developmental stages. (H) UMAPs showing gene expression of cardiac and mesoderm genes. (I) Doplot denoting marker genes and cell type annotations by cluster in mesoderm atlas. Size of dot represents percent of cells expressing gene and color represents average expression level. Cluster numbers denote cell types when annotation was not possible.



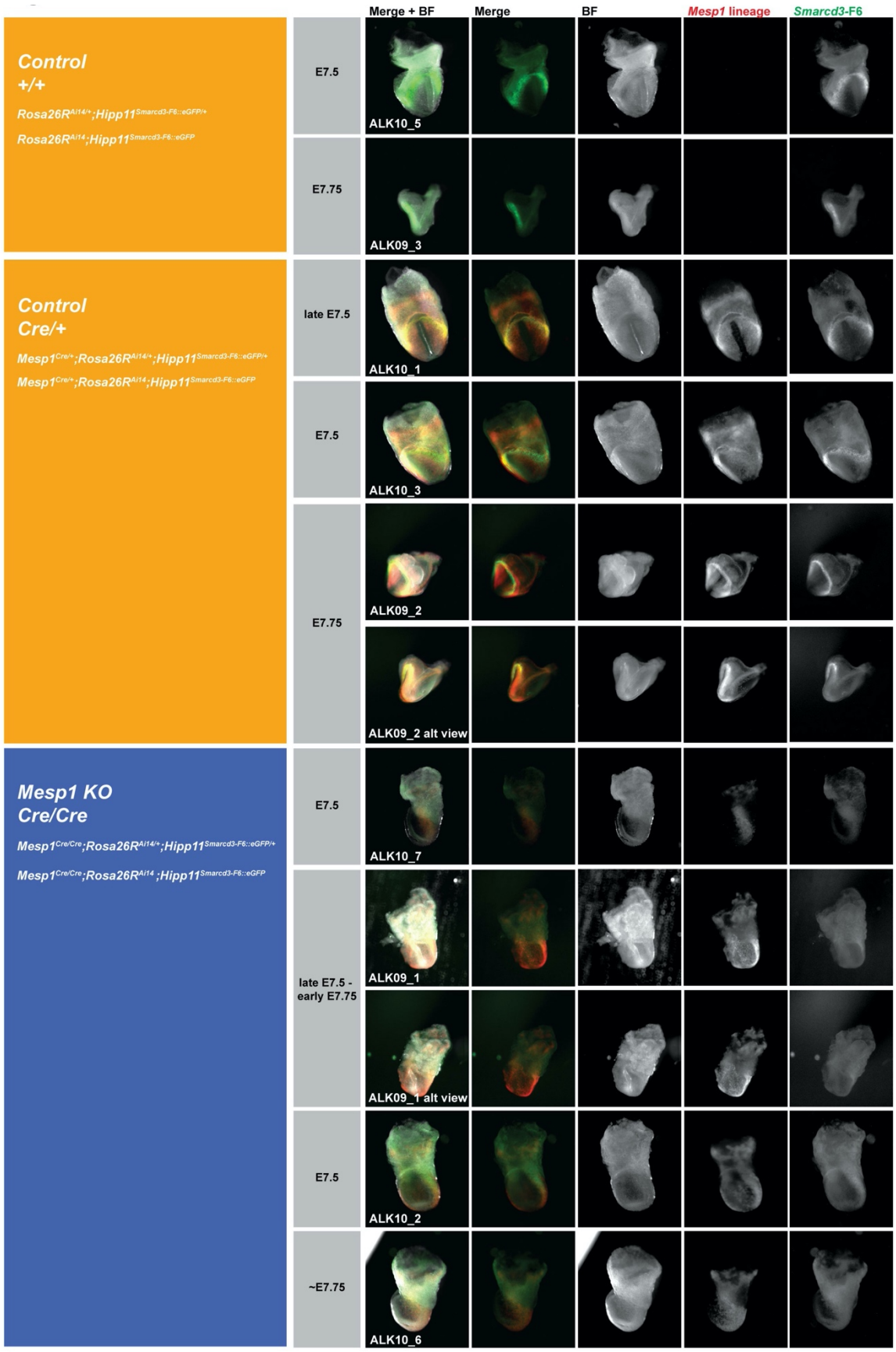
Supplementary Figure 4.4. Cell type labels of developmental stage mesoderm atlases. (A) Doplot denoting marker genes and cell type annotations by cluster in Early mesoderm atlas. (B) Doplot denoting marker genes and cell type annotations by cluster in Middle mesoderm atlas. (C) Doplot denoting marker genes and cell type annotations by cluster in Late mesoderm atlas. Size of dot represents percent of cells expressing gene and color represents average expression level.



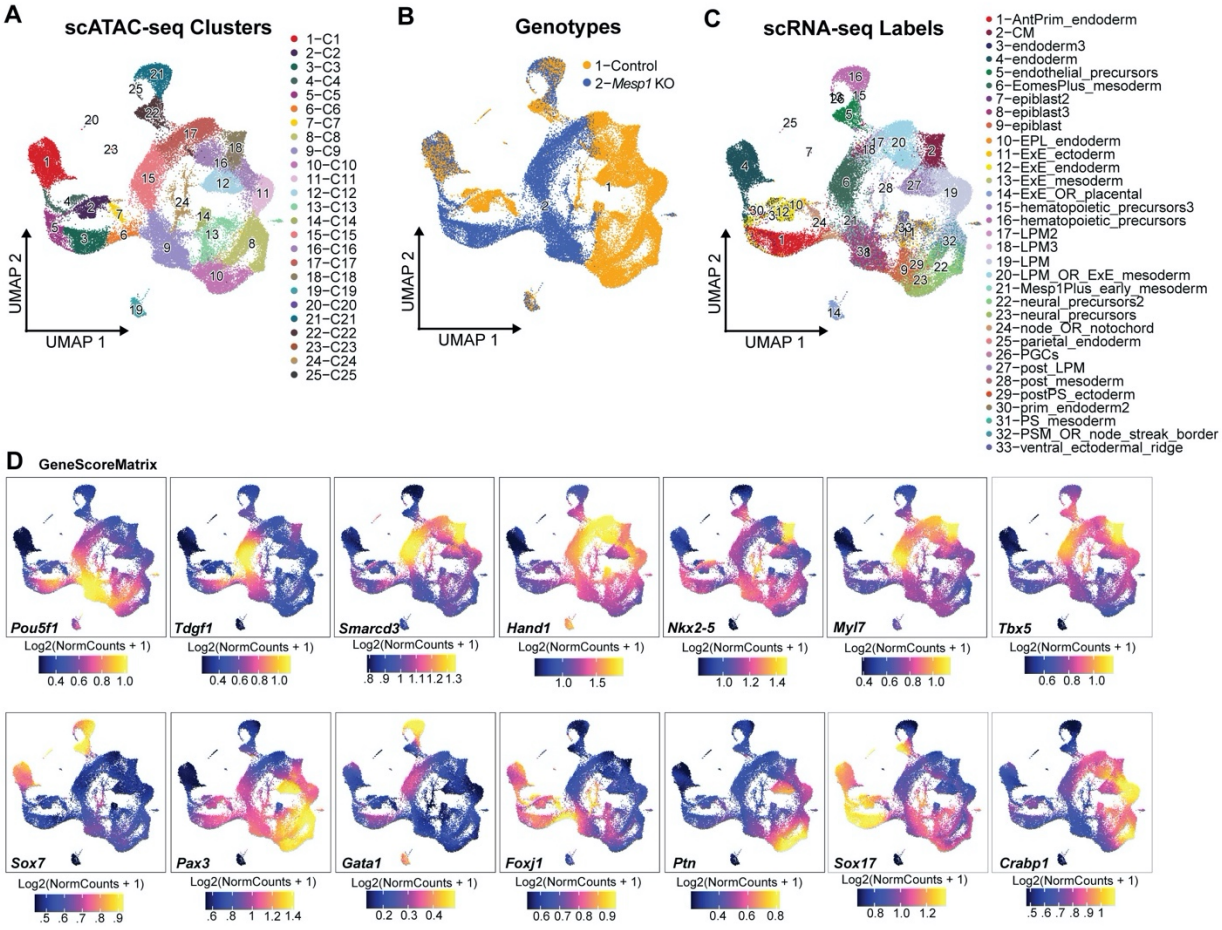
Supplementary Figure 4.5. Disrupted organization of mesoderm in Middle stage *Mesp1* KO embryos. (A) Overlay of *Mesp1*, *Foxc2*, and Ai14 *Mesp1*-lineage gene expression in cell types of Middle mesoderm atlas UMAP. (B) Immunostaining and Light Sheet Confocal microscopy for *Foxc2* (magenta), *Smarcd3*-F6 (green) and *Mesp1* via Cre detection (blue) in Middle stage embryos (~E6.75). Arrowheads denote domain boundaries in control and disruption in *Mesp1* KO embryo. Scale bars are 100 μm.


A
Mesoderm
Early
Middle
Late
pseudotime
16 15 11 20 2 12 29 0 7 22 26 14 4 18
HPCs2
HPCs1
Endothelial2
CFMeso
Allantois
CMs
CardiacMeso
PostLPM
Endothelial1
C22
LPM
pSoM-like
posPrxM1
Meso
B
Mesp1 lineage (Ai14-transgene)
2.0
1.5
1.0
0.5
0.0
C
Exit Pluripotency
Pou5f1
Dnmt3b
Initiate Mesoderm
T
Eomes
Mesendoderm
Sox17
Mesp1
Lhx1
Mixl1
Cardiac Mesodern and Mesoderm
Hand1
Cdx2
Snai1
Foxc2
Anxa2
Amot
Tdgf1
Fgf8
Cardiac, Cardiac Patterning and Maturation
Gata4
Gata5
Gata6
Tnnt2
Isl1
Tbx5
Mef2c
Hand2
Sfrp5
Ankrd1
Crabp2
Id2

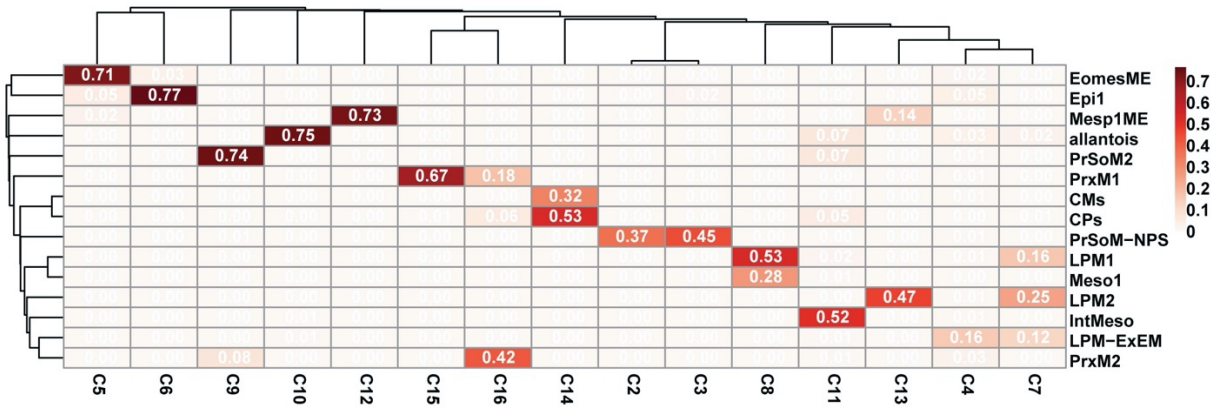
Supplementary Figure 4.6. URD trajectory for pseudotime ordering of control and *Mesp1* KO mesoderm. (A) URD tree labeled with relative developmental stages of embryos. (B) URD tree labeled *with Mesp1* lineage transgene reporter Ai14. (C) URD trees labeled with gene expression of various mesodermal genes and TFs involved in regulation, differentiation, and progression of cardiogenesis.


Merge + BF
Merge
BF
Mesp1 lineage
Smarcd3-F6
Control
+/+
Rosa26R^Ai14/+;Hipp11^Smarcd3-F6::eGFP/+
Rosa26R^Ai14;Hipp11^Smarcd3-F6::eGFP
E7.5
ALK10_5
E7.75
ALK09_3
Control
Cre/+
Mesp1^Cre/+;Rosa26R^Ai14/+;Hipp11^Smarcd3-F6::eGFP/+
Mesp1^Cre/+;Rosa26R^Ai14;Hipp11^Smarcd3-F6::eGFP
late E7.5
ALK10_1
E7.5
ALK10_3
E7.75
ALK09_2
ALK09_2 alt view
Mesp1 KO
Cre/Cre
Mesp1^Cre/Cre;Rosa26R^Ai14/+;Hipp11^Smarcd3-F6::eGFP/+
Mesp1^Cre/Cre;Rosa26R^Ai14;Hipp11^Smarcd3-F6::eGFP
E7.5
ALK10_7
late E7.5 -
early E7.75
ALK09_1
ALK09_1 alt view
E7.5
ALK10_2
~E7.75
ALK10_6

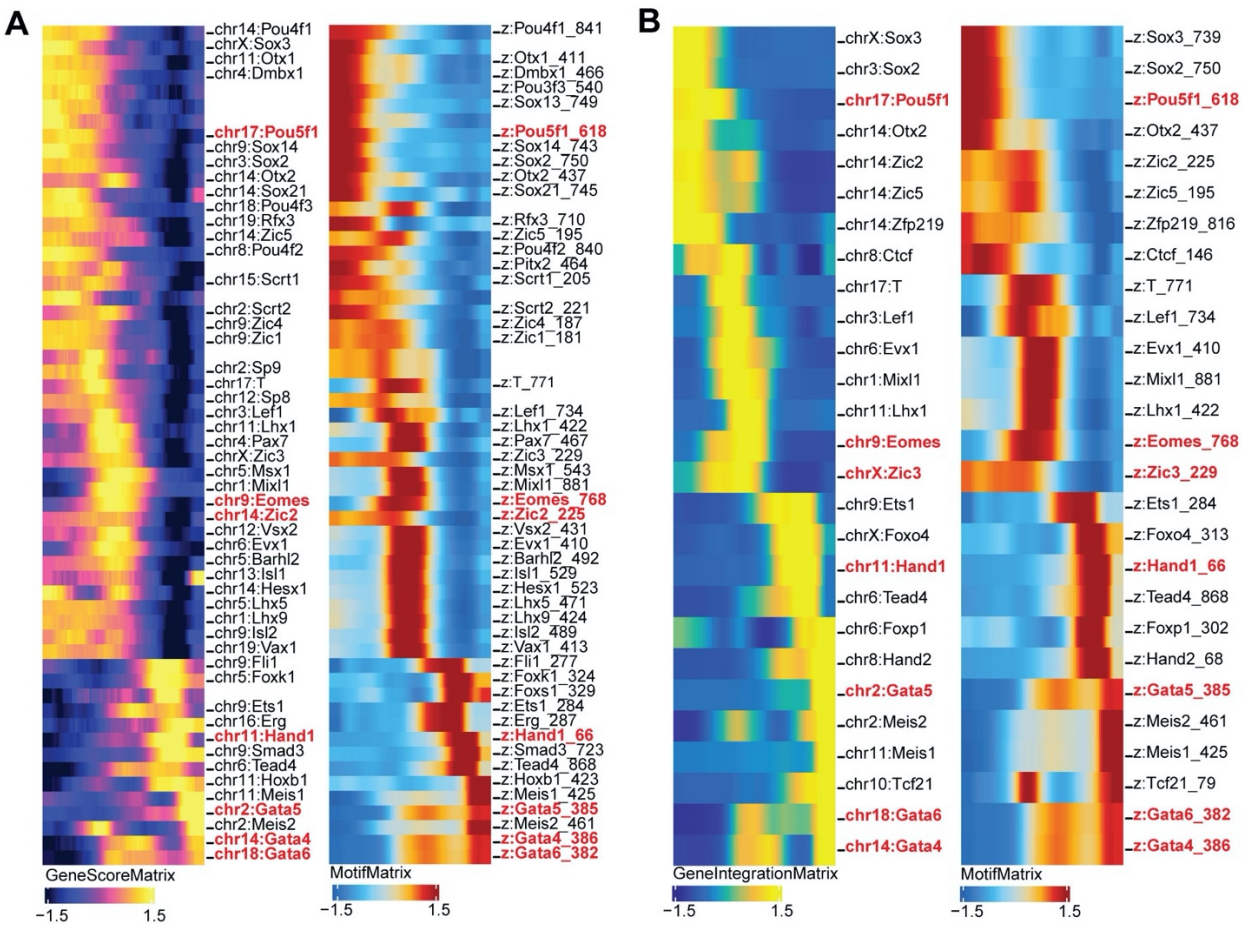
Supplementary Figure 4.7. Middle and Late stage embryos assayed in scATAC-seq. Images of embryos utilized in generation of control and *Mesp1* KO scATAC-seq dataset. *Mesp1* lineage visualized by endogenous Ai14 fluorescent reporter transgene. *Smarcd3*-F6 visualized by endogenous eGFP fluorescent reporter transgene. Images not acquired and processed identically. Genotypes and ages denoted.



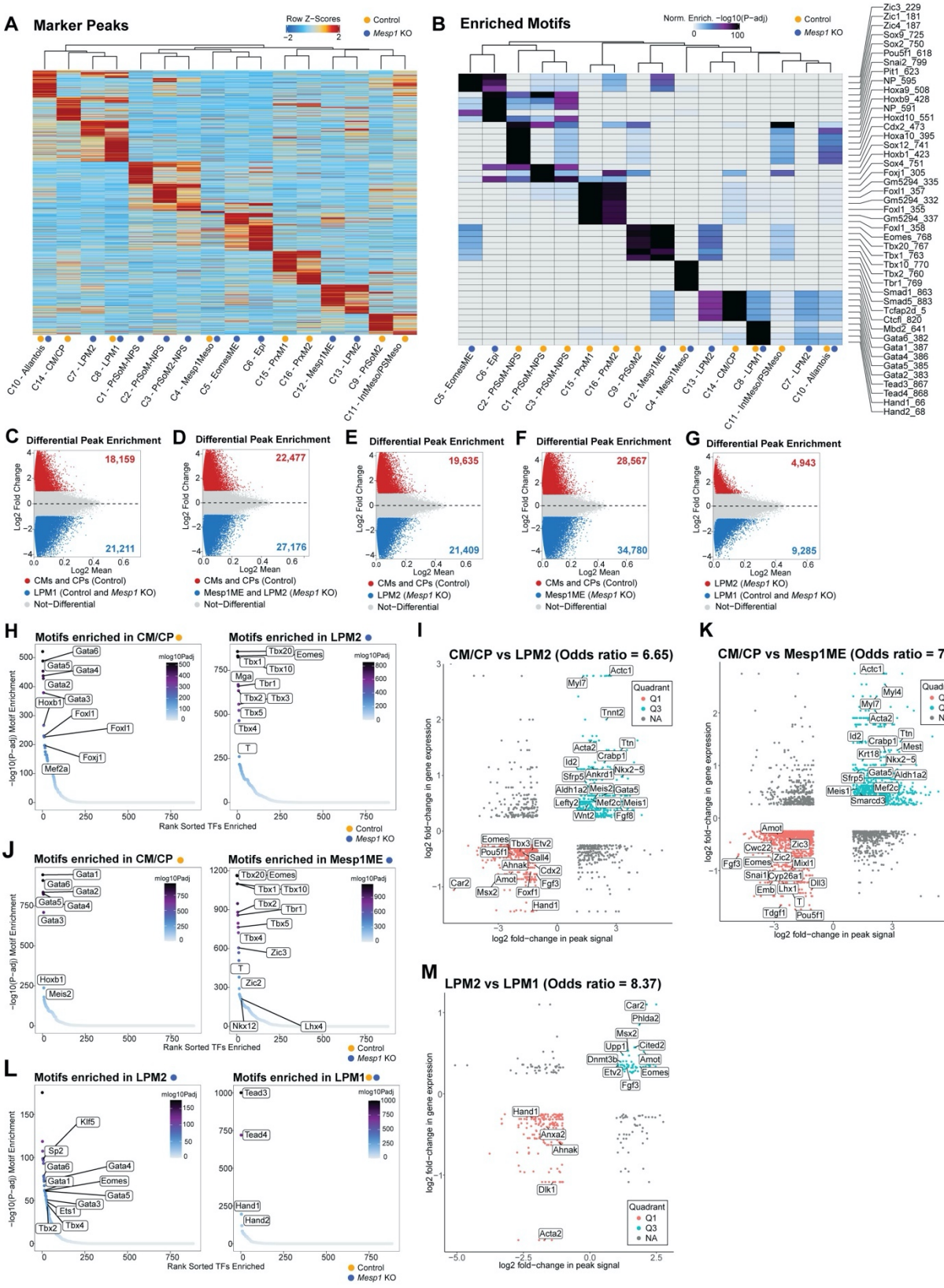
Supplementary Figure 4.8. Identification of mesoderm in scATAC-seq whole embryo data. (A) Whole embryo scATAC-seq atlas with overlays for (B) genotype and (C) relative cell type identities from integration of the complementary scRNA-seq dataset. (D) GeneScoreMatrix plots for chromatin accessibility around gene loci of various mesoderm, cardiac, endoderm, ectoderm, neuronal marker genes.



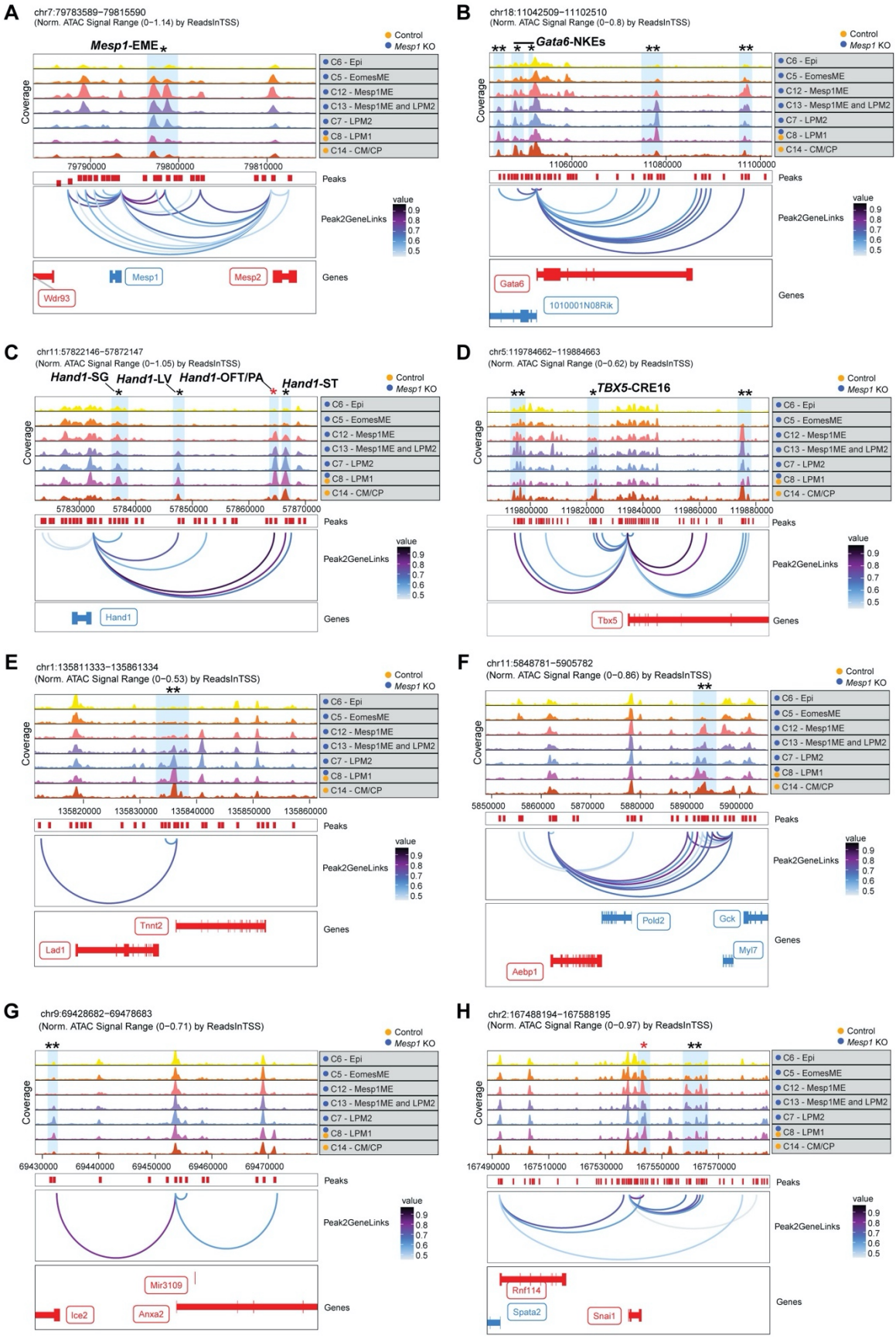
Supplementary Figure 4.9. Jaccard Similarity Index for scATAC-seq mesoderm cluster annotation. Scaled strength of similarity match for scRNA-seq complementary dataset label transfer (rows) onto scATAC-seq clusters (columns). Values 0-1 indicate strength of similarity match for relative cell type annotations. Greater values indicate stronger label matching.



Supplementary Figure 4.10. Integrative scATAC-seq trajectory pseudotime correlation analysis. Heatmap visualizations of dynamic shifts along pseudotime progress for correlation matrices (A) between accessibility near TF loci, GeneScoreMatrix, with associated TF motifs, MotifMatrix and (B) between TF gene expression, GeneIntegrationMatrix, with associated TF motifs, MotifMatrix. Motifs in red represent selected putative positive regulators.


A Marker Peaks
Row Z-Scores
−2
2
Control
Mesp1 KO
C10 - Allantois
C14 - CM/CP
C7 - LPM2
C8 - LPM1
C1 - PSoM-NPS
C2 - PSoM-NPS
C3 - PSoM2-NPS
C4 - Mesp1Meso
C5 - EomesME
C6 - Ep
C15 - PrxM1
C16 - PrxM2
C12 - Mesp1ME
C13 - LPM2
C9 - PSoM2
C11 - IntMesoPSMeso
B Enriched Motifs
Norm. Enrich. −log10(P−adj)
0
100
Control
Mesp1 KO
Zic3_229
Zic1_181
Zic4_187
Sox9_725
Sox2_750
Pou5f1_618
Snai2_799
Pit1_623
NP_595
Hoxa9_508
Hoxb9_428
NP_591
Hoxd10_551
Cdx2_473
Hoxa10_395
Sox12_741
Hoxb1_423
Sox4_751
Foxj1_305
Gm5294_335
Foxl1_357
Gm5294_332
Foxl1_355
Gm5294_337
Foxl1_358
Eomes_768
Tbx20_767
Tbx1_763
Tbx10_770
Tbx2_760
Tbr1_769
Smad1_863
Smad5_883
Tcfap2d_5
Ctcfl_820
Mbd2_641
Gata6_382
Gata1_387
Gata4_386
Gata5_385
Gata2_383
Tead3_867
Tead4_868
Hand1_66
Hand2_68
C5 - EomesME
C6 - Ep
C2 - PSoM-NPS
C1 - PSoM-NPS
C3 - PSoM-NPS
C15 - PrxM1
C16 - PrxM2
C9 - PSoM2
C12 - Mesp1ME
C4 - Mesp1Meso
C13 - LPM2
C14 - CM/CP
C8 - LPM1
C11 - IntMesoPSMeso
C7 - LPM2
C10 - Allantois
C Differential Peak Enrichment
18,159
21,211
CMs and CPs (Control)
LPM1 (Control and Mesp1 KO)
Not−Differential
D Differential Peak Enrichment
22,477
27,176
CMs and CPs (Control)
Mesp1ME and LPM2 (Mesp1 KO)
Not−Differential
E Differential Peak Enrichment
19,635
21,409
CMs and CPs (Control)
LPM2 (Mesp1 KO)
Not−Differential
F Differential Peak Enrichment
28,567
34,780
CMs and CPs (Control)
Mesp1ME (Mesp1 KO)
Not−Differential
G Differential Peak Enrichment
4,943
9,285
LPM2 (Mesp1 KO)
LPM1 (Control and Mesp1 KO)
Not−Differential
Log2 Fold Change
Log2 Mean
H Motifs enriched in CM/CP
Motifs enriched in LPM2
−log10(P−adj) Motif Enrichment
Rank Sorted TFs Enriched
Gata6
Gata5
Gata4
Gata2
Hoxb1
Gata3
Foxl1
Foxj1
Mef2a
Tbx20
Eomes
Tbx1
Tbx10
Mga
Tbr1
Tbx2
Tbx3
Tbx5
Tbx4
T
I CM/CP vs LPM2 (Odds ratio = 6.65)
log2 fold−change in gene expression
log2 fold−change in peak signal
Quadrant
Q1
Q3
NA
Actc1
Myl7
Tnnt2
Ttn
Acta2
Crabp1
Id2
Sfrp5
Ankrd1
Nkx2−5
Aldh1a2
Meis2
Gata5
Lefty2
Mef2c
Meis1
Wnt2
Fgf8
Eomes
Tbx3
Etv2
Pou5f1
Sall4
Ahnak
Cdx2
Car2
Amot
Fgf3
Msx2
Foxf1
Hand1
K CM/CP vs Mesp1ME (Odds ratio = 7.7)
Myl4
Id2
Krt18
Smarcd3
Amot
Cwc22
Zic3
Zic2
Mixl1
Fgf3
Snai1
Cyp26a1
Dll3
Emb
Lhx1
Tdgf1
Pou5f1
J Motifs enriched in CM/CP
Motifs enriched in Mesp1ME
Gata1
Gata6
Gata2
Gata5
Gata4
Gata3
Hoxb1
Meis2
Tbx20
Eomes
Tbx1
Tbx10
Tbx2
Tbr1
Tbx5
Tbx4
Zic3
T
Zic2
Nkx12
Lhx4
M LPM2 vs LPM1 (Odds ratio = 8.37)
Car2
Phlda2
Msx2
Upp1
Cited2
Dnmt3b
Amot
Etv2
Eomes
Fgf3
Hand1
Anxa2
Ahnak
Dlk1
Acta2
L Motifs enriched in LPM2
Motifs enriched in LPM1
Klf5
Sp2
Gata6
Gata4
Gata1
Eomes
Gata5
Gata3
Ets1
Tbx2
Tbx4
Tead3
Tead4
Hand1
Hand2
Control
Mesp1 KO

Supplementary Figure 4.11. Differential peak and motif enrichment in cardiogenic cell types comprised of control and *Mesp1* KO mesoderm cells. (A) Heatmap for Marker Peak (FDR <= 0.05, Log2FC >=1) accessibility profiles of mesoderm cell types comprised of control, *Mesp1* KO, or both genotypes. (B) Heatmap for enriched motifs (FDR <=0.05, Log2FC >=1) in cluster Marker Peaks. (C-G) MA plots for pairwise comparisons of differential peak enrichment between cardiogenic cell types with genotypes noted. (H, J, L) Motifs enriched in differentially accessible peaks between noted cell types and cluster genotypes. (I, K, M) Plots for peak,gene associations showing correlations between differential peak accessibility and gene expression in comparisons between cells type1 vs type2. Q3 peak,gene pairs represent significantly more accessible peaks paired with upregulated gene expression in type1 cells. Q1 peak,gene pairs represent significantly more accessible peaks paired with upregulated gene expression in type2 cells. Odds ratio denotes probability for observed peak,gene relationships. (I) Peak,gene association plot for cells in CMCP vs LPM2 comparison, (K) CMCP vs Mesp1ME comparison, (M) LPM2 vs LPM1 comparison.


A
chr7:79783589−79815590
(Norm. ATAC Signal Range (0−1.14) by ReadsInTSS)
Mesp1-EME
Control
Mesp1 KO
C6 - Epi
C5 - EomesME
C12 - Mesp1ME
C13 - Mesp1ME and LPM2
C7 - LPM2
C8 - LPM1
C14 - CM/CP
Coverage
79790000
79800000
79810000
Peaks
Peak2GeneLinks
value
0.9
0.8
0.7
0.6
0.5
Genes
Mesp1
Mesp2
Wdr93
B
chr18:11042509−11102510
(Norm. ATAC Signal Range (0−0.8) by ReadsInTSS)
Gata6-NKEs
11060000
11080000
11100000
Gata6
1010001N08Rik
C
chr11:57822146−57872147
(Norm. ATAC Signal Range (0−1.05) by ReadsInTSS)
Hand1-SG
Hand1-LV
Hand1-OFT/PA
Hand1-ST
57830000
57840000
57850000
57860000
57870000
Hand1
D
chr5:119784662−119884663
(Norm. ATAC Signal Range (0−0.62) by ReadsInTSS)
TBX5-CRE16
119800000
119820000
119840000
119860000
119880000
Tbx5
E
chr1:135811333−135861334
(Norm. ATAC Signal Range (0−0.53) by ReadsInTSS)
135820000
135830000
135840000
135850000
135860000
Tnnt2
Lad1
F
chr11:5848781−5905782
(Norm. ATAC Signal Range (0−0.86) by ReadsInTSS)
5850000
5860000
5870000
5880000
5890000
5900000
Pold2
Gck
Myl7
Aebp1
G
chr9:69428682−69478683
(Norm. ATAC Signal Range (0−0.71) by ReadsInTSS)
69430000
69440000
69450000
69460000
69470000
Mir3109
Ice2
Anxa2
H
chr2:167488194−167588195
(Norm. ATAC Signal Range (0−0.97) by ReadsInTSS)
167490000
167510000
167530000
167550000
167570000
Rnf114
Spata2
Snai1

Supplementary Figure 4.12. Peak2Gene linkage plots for dysregulated genes in *Mesp1* KO embryos. Peak2Gene linkage browser tracks for cell types showing predicted regulatory connections between distal accessible regions (Peaks) and nearby genes. Shaded bars denote predicted distal regulatory regions; *denotes characterized elements; red* denotes regions with Mesp1-binding; **denotes uncharacterized elements. Characterized elements named when available. Peak linkages to genes (A) *Mesp1* (B) *Gata6,* (C) *Hand1,* (D) *Tbx5,* (E) *Tnnt2*, (F) *Myl7*, (G) *Anxa2*, (H) *Snai1*.

Materials and Methods

Mouse models

Animal studies were performed in strict compliance with the UCSF Institutional Animal Care and Use Committee. Mice were housed in a standard 12 hour light/dark animal husbandry barrier facility at the Gladstone Institutes. The *Mesp1*$^{Cre/+}$ knock-in mice were obtained from Yumiko Saga (Ajima et al., 2021; Saga et al., 1999). *Rosa26R*^{Ai14} mice were from Jackson Laboratory (strain #007914, (Madisen et al., 2010). Tdgf1::*LacZ* mice containing a transgene for *Tdgf1* enhancer with a LacZ reporter were obtained from Brian Black (Barnes et al., 2016).

Control embryos were generated from crosses of *Mesp1*$^{Cre/+}$*;Rosa26R*^{Ai14}*;Hipp11*$^{Smarcd3\text{-}F6::eGFP}$ males to C57BL/6J wildtype, *Mesp1*$^{Cre/+}$, or *Mesp1*$^{Cre/+}$*;Rosa26R*^{Ai14}*;Hipp11*$^{Smarcd3\text{-}F6::eGFP}$ females. *Mesp1* KO embryos were generated from crosses of *Mesp1*$^{Cre/+}$*;Rosa26R*^{Ai14}*;Hipp11*$^{Smarcd3\text{-}F6::eGFP}$ males to *Mesp1*$^{Cre/+}$, or *Mesp1*$^{Cre/+}$*;Rosa26R*^{Ai14}*;Hipp11* $^{Smarcd3\text{-}F6::eGFP}$ females. Transgenic embryos for single cell transcriptomic and epigenomic sequencing experiments were all on a C57BL/6J background. Transgenic embryos for whole mount *in situ* hybridizations and immunohistochemistry validations were on C57BL/6J backgrounds or a mixed CD1 / C57BL/6J background, in order to facilitate better littermate stage matching via larger litters, with litters born to *Mesp1*$^{Cre/+}$ CD1 / C57BL/6J hybrid females mated to *Mesp1*$^{Cre/+}$*;Rosa26R*^{Ai14}*;Hipp11*$^{Smarcd3\text{-}F6::eGFP}$ C57BL/6J males. “Control” denotes embryos with at least one wildtype allele in the *Mesp1* locus and includes genotypes *Mesp1*$^{Cre/+}$*;Rosa26R*^{Ai14}*;Hipp11*$^{Smarcd3\text{-}F6::eGFP}$, *Mesp1*$^{Cre/+}$*;Rosa26R*$^{Ai14/+}$ *;Hipp11*$^{Smarcd3\text{-}}$

$^{F6::eGFP/+}$, *Mesp1*$^{+/+}$;*Rosa26R*^{Ai14};*Hipp11*$^{Smarcd3\text{-}F6::eGFP}$, or *Mesp1*$^{+/+}$;*Rosa26R*$^{Ai14/+}$ *Hipp11*$^{Smarcd3\text{-}F6::eGFP/+}$. Heterozygosity of *Mesp1*$^{Cre/+}$ or *Mesp1*$^{+/+}$ is noted when control embryos were utilized in scRNA-seq (Fig. S1) or scATAC-seq (Fig. S10) library generation. "*Mesp1* KO" denotes embryos with homozygosity of the Cre insertion disrupting the *Mesp1* locus and includes genotypes *Mesp1*$^{Cre/Cre}$;*Rosa26R*^{Ai14};*Hipp11*$^{Smarcd3\text{-}F6::eGFP}$ or *Mesp1*$^{Cre/Cre}$;*Rosa26R*$^{Ai14/+}$;*Hipp11*$^{Smarcd3\text{-}F6::eGFP/+}$.

Control embryos for activity assessment of the *Tdgf1* enhancer had genotypes *Mesp1*$^{Cre/+}$;*Tdgf1*::LacZ or *Tdgf1*::LacZ, and *Mesp1* KO embryos had genotypes *Mesp1*$^{Cre/Cre}$:*Tdgf1*::LacZ.

Cloning and generation of TARGATT transgenic knock-in mice

The *Smarcd3*-F6 fragment was isolated and cloned with inclusion of an *nlsEGFP* under control of an *Hsp68* minimal promoter for TARGATT (Applied Stem Cells) insertion to the *Hipp11* locus as previously described (Devine et al., 2014) to create the *Hipp11*$^{Smarcd3\text{-}F6::eGFP}$ mouse. Purified construct DNA was injected into embryo pronuclei along with mRNA for the *Phi31o* transposase according to manufacturer's protocols.

Timed matings and whole embryo dissections

To achieve timed matings, male and female mice were housed together in the evening and pregnancy was assessed by vaginal plug the following morning. Gestational stage was determined starting as day E0.5 at noon of plug detection. Females were confirmed pregnant by abdominal ultrasound (Vevo 3100, Visual Sonics) the afternoon of day 6 or

morning of day 7 and sacrificed according to IACUC standard procedure at noon on day 7, or the early morning of day 8. The embryonic ages captured in individual litters ranged from E6.0 to E7.5 on day 7, and E7.5 to E7.75 on day 8. The diversity of ages in litters aided in the construction of a fine timecourse for embryonic timelines.

Embryos were dissected and in later stages when yolk is present, also de-yolked, in ice-cold PBS (Life Technologies, 14190250) with 1% FBS (Thermo Fisher Scientific, 10439016) on ice. Embryos were screened using an upright epifluorescent dissecting microscope (Leica MZFLIII microscope, Lumen Dynamics XCite 120LED light source, Leica DFC 3000G camera) for presence of both red and green fluorescent reporters, indicative of *Mesp1* lineage tracing from $Mesp1^{Cre}$;$Rosa26R^{Ai14}$ alleles and expression of the *Smarcd3*-F6::eGFP transgene reporter from the $Hipp11^{Smarcd3\text{-}F6::eGFP}$ allele, respectively. Embryos were staged according to (Downs & Davies, 1993). For difficult-to-capture control stages used in construction of the wildtype scRNA-seq timeline, absence of *Mesp1* lineage (Ai14) reporter was permitted and noted for those embryos (Fig. S1). Additionally, $Mesp1^{+/+}$ alleles were specifically included in addition to $Mesp1^{Cre/+}$ *as* controls for scATAC-seq library generation in the event locus-specific effects of Cre insertion required additional consideration, which we didn't find to be the case as both control genotypes appeared identically in the dataset (Fig. S10). DNA for genotyping was extracted using QuickExtract DNA Extraction Solution (Lucigen, QE09050) from harvested yolk sac tissue if available or else from a micro-dissected nick of the extraembryonic anterior proximal region. Genotyping was performed to distinguish *Mesp1* KO embryos from control embryos using Phire Green Hot Start II DNA Polymerase (Thermo Fisher Scientific, F124L) according to manufacturer's

protocols using primers to detect wildtype bands (control, P1+P3) and Cre alleles (*Mesp1* KO, P1+P2):

Mesp1 FWD, P1: GGC CAT AGG TGC CTG ACT TA

Cre2 REV, P2: CCT GTT TTG CAC GTT CAC GG

Mesp1 REV, P3: ACC AGC GGG ACT CAG GAT

Embryo preparation for single-cell library generation

Due to the small size and lack of morphological distinction between tissue types of embryos at these early stages, whole embryos were dissected and harvested for single cell library generation.

Whole embryos were incubated in 200 µL 0.25% TrypLE (ThermoFisher Scientific, 12563029) solution for 5 min at 37°C and triturated gently. Dissociated cell suspension was quenched with 600 µL of PBS with 1% FBS, singularized via passage through a 70 µm cell strainer (BD Falcon, 352235), pelleted by centrifugation at 150xg for 3 min, and resuspended in 34 µL of PBS with 1% FBS. At least 2 embryos were collected per genotype per embryonic stage in all datasets except for the *Mesp1* KO embryos in the scRNA-seq dataset where this was not possible, and the use of relative developmental stages was employed in analysis along with replicate validations via *in-situ* hybridization for differentially expressed genes.

Single-cell transcriptome library preparation and sequencing

Libraries for scRNA-seq were prepared according to manufacturer's instructions using the 10X Genomics Chromium controller, Chromium Single Cell 5' Library and Gel Bead Kit v1 (10X Genomics, 1000006) and Chromium Single Cell A Chip Kit (10X Genomics, 1000151). A maximum of 10,000 cells per sample were loaded onto the 10X Genomics Chromium instrument, and each sample was indexed with a unique sample identifier (10X Genomics Chromium i7 Multiplex Kit, 120262). Final libraries were pooled and sequenced shallowly according to 10X protocol parameters on a NextSeq500 (Illumina), and then re-pooled for deeper sequencing on HighSeq4000 (Illumina) and/or NovaSeq using an S4 lane (Illumina). Littermate, stage-matched comparisons of control and *Mesp1* KO libraries were always sequenced together in the same library pool. All scRNA-seq libraries were sequenced to a mean read depth of at least 50,000 total aligned reads per cell.

Processing raw scRNA-seq

Raw sequencing reads were processed using the 10X Genomics Cellranger v3.0.2 pipeline. Reads were demultiplexed using cellranger mkfastq and aligned with cellranger count to the Mm10 reference genome containing additional sequences for the Ai14 and eGFP. Cellranger "aggr" was used to aggregate and read depth normalize multiple GEM libraries for the atlas dataset of control and *Mesp1* KO embryo libraries.

Seurat analysis of scRNA-seq data

Outputs from the Cellranger pipeline were analyzed using the Seurat Package v3.0.2 in R (Butler et al., 2018; Stuart et al., 2019; Satija et al., 2015). The “WTvsMut” dataset as a single aggregated counts matrix was input to Read10X and CreateSeuratObject functions. Quality control steps were performed to remove dead cells or doublets.

Whole Embryo Control vs. Mesp1 KO Atlas:

For the WTvsMut atlas, cells with <10% mitochondrial reads, UMI counts less than 50,000, and detected genes between 200 and 7,000 were retained. SCTransform was used to normalize and scale data with regressions performed with respect to mitochondrial percent, number of genes, and number of UMI counts detected. PCA analysis and batch correction were performed using FastMNN (Haghverdi et al., 2018) split by experimental group (experiment number denoted with library prefixes ALK06, ALK08, ALK07, ALK05, ALK04). 96,027 cells were clustered based on the top 50 principal components and visualized using RunUMAP, FindNeighbors, and FindClusters and outputs were visualized as Uniform Manifold Approximation and Projection (UMAP) embeddings generated with DimPlot. Cells were clustered using the FindAllMarkers function with Wilcoxon rank-sum test (min.pct = 0.1, logfc threshold = 0.25) at resolution 1.0 to identify cluster specific marker genes. Iterative clustering performed following removal of low-quality clusters. This WTvsMut dataset represents 96,027 cells containing 79,725 control and 16,302 *Mesp1* KO cells.

The relevant developmental stages were annotated within Seurat meta data. Cells from 6 embryos staged E6.0 - E6.5 (ALK06_2_E60_con_rep1, ALK06_4_E60_con_rep2, ALK08_20_E60_con_rep3, ALK08_14_lateE60_con_rep1, ALK07_15_E65_con_rep1, ALK08_6_E65_Mesp1KO_rep1) were denoted as “Early” stages. Cells from 4 embryos staged late E6.5 – early E7.5 (ALK07_3_lateE65_con_rep1, ALK07_14_E70_con_rep1, ALK08_11_E70_Mesp1KO_rep1, ALK07_7_earlyE75_con_rep1) were denoted as “Middle” stages. Cells from 5 embryos staged late E7.5 to early E7.75 when cardiac crescent is formed (ALK07_6_lateE75_con_rep1, ALK04_3_lateE75_con_rep2, ALK05_7_E775_con_rep1, ALK05_2_lateE75_Mesp1KO_rep1, ALK07_8_E775_con_rep2) were denoted as “Late” stages. While we set out to acquire replicates of both genotypes per each stage as the most optimal statistical scenario, the 25% yield of *Mesp1* KO embryos within C57BL/6J litter sizes at these early gastrulation stages proved prohibitive. Thus we relied on validations of key scRNA-seq findings via the orthogonal approach of multiplexed whole mount *in situ* hybridizations.

Smarcd3-F6+ Control vs Mesp1 KO Atlas:

To analyze putative CPCs, all cells expressing the *Smarcd3*-F6-eGFP transgene were subsetted from the full WTvsMut atlas and re-clustered into their own Seurat object containing 4,868 cells (4,276 control and 592 *Mesp1* KO cells). FindAllMarkers function was used to identify cluster marker genes of represented cell types at resolution 1.7. The analysis between control and *Mesp1* KO genotypes irrespective of cell type was performed using FindMarkers function between genotypes with Wilcoxon rank-sum test (min.pct = 0.1, logfc threshold = 0.25). Cluster-wise differential gene expression testing

was performed using FindMarkers function and Wilcoxon rank-sum test (min.pct = 0.1, logfc threshold = 0.25) between genotypes within specific cell type clusters, and visualized with the VlnPlot function. Differential gene expression results irrespective of cell type were visualized by DotPlot function separated by genotypes and also genotypes separated by developmental stages.

Mesoderm Control vs Mesp1 KO Atlas:

Relevant mesoderm cells were subsetted from the full WTvsMut atlas based on cluster-wise detection via FeaturePlot and VlnPlot at cluster resolution 1.0 of *Smarcd3*-F6::eGFP and Ai14 transgenes for CPCs and the *Mesp1* lineage, respectively. The resulting 35,792 cells (29,924 control and 5,868 *Mesp1* KO cells) of the WTvsMut mesoderm dataset was re-clustered and annotated at resolution 1.5 using FindAllMarkers function as above to identify cell type marker genes as described above.

Embryos representing the relative developmental stages of “Early” (5,504 cells; 4,472 control and 1,032 *Mesp1* KO), “Middle” (7,666 cells; 6,734 control and 932 *Mesp1* KO), and “Late” (22,622 cells; 18,718 control and 3,904 *Mesp1* KO) as described above were subsetted into respective individual Seurat objects, re-clustered as described, and cell type clusters were further re-annotated (at resolutions 0.7, 0.7, 0.7 for Early, Middle, and Late objects, respectively). Clusters representing cell types relevant for cardiac development were identified through cluster-wise enrichment of *Smarcd3*-F6::eGFP transgene expression overlayed in UMAP space via FeaturePlot. Differential gene expression testing between genotypes within cardiogenic cell type clusters was performed using FindMarkers function with Wilcoxon rank-sum test (min.pct = 0.1, logfc

threshold = 0.25). Differentially expressed genes with adjusted p-values < 0.05 were plotted as violin plots in Seurat except in cases to highlight total absence of transcript in one genotype condition.

Whole Embryo Control vs Mesp1 KO Atlas for scATAC-seq integration:

For the scRNA-seq WTvsMut atlas for integration with scATAC-seq data, libraries from Middle stage embryos (ALK07_3_lateE65_con_rep1, ALK07_14_E70_con_rep1, ALK08_11_E70_Mesp1KO_rep1, ALK07_7_earlyE75_con_rep1) and Late stage embryos (ALK07_6_lateE75_con_rep1, ALK04_3_lateE75_con_rep2, ALK05_7_E775_con_rep1, ALK05_2_lateE75_Mesp1KO_rep1, ALK07_8_E775_con_rep2) were subsetted from the aggregated WTvsMut counts matrix. Cells with <7.5% mitochondrial reads, UMI counts less than 50,000, and detected genes between 200 and 7,000 were retained. SCTransform was used to normalize and scale data with regressions performed with respect to mitochondrial percent, number of genes, and number of UMI counts detected. PCA analysis and batch correction were performed using FastMNN split by experimental group. After initial clustering as previously described, cell clusters representing low quality cells were removed and clustering was iterated again. The resulting dataset represents 82,536 cells containing 68,717 control and 13,819 *Mesp1* KO cells. Cluster cell types were annotated at resolution 1.2 using FindAllMarkers as described above.

Mesoderm Control vs Mesp1 KO Atlas for scATAC-seq integration:

Relevant mesoderm cells were subsetted from the whole embryo matched scATAC-seq WTvsMut atlas based on cluster-wise detection via FeaturePlot and VlnPlot of *Smarcd3*-F6::eGFP and Ai14 transgenes for CPCs and the *Mesp1* lineage, respectively. The resulting 30,427 cells (26,054 control and 4,373 *Mesp1* KO cells) of the scATAC-seq matched mesoderm WTvsMut dataset were re-processed from RNA assay slot with the standard Seurat workflow NormalizeData, FindVariableFeatures and ScaleData. SCTransform was not used in this mesoderm scRNA-seq dataset because we found that while cell type label-transfer with scATACseq was successful as previously described for the whole embryo integration, downstream scATAC-seq analyses leveraging the scRNA-seq gene integration matrix performed in the mesoderm scATAC-seq dataset were incompatible with SCT-normalized values. PCA analysis and batch correction were performed using FastMNN split by experimental group. From here clustering was performed as previously described and cell types were annotated at resolution 1.2 using FindAllMarkers function as above to identify cell type marker genes as described above.

Differential gene expression testing between genotypes within cell type clusters and between cell type clusters was performed using FindMarkers function with Wilcoxon rank-sum test (min.pct = 0.1, logfc threshold = 0.25). These lists of differentially expressed genes served as inputs to the (peak, gene) association analyses with scATAC-seq differential peaks using rGreat (below in methods).

Single cell transcriptomic cell trajectories and pseudotime analysis

Pseudotime analysis was performed using the URD package (version 1.0.2 and 1.1.1) (Farrell et al., 2018). The WTvsMut mesoderm Seurat object containing all three relative developmental stages, processed as previously described, was converted to an URD object using the seuratToURD function. Cell-to-cell transition probabilities were constructed by setting the number of near neighbors (knn) to 189 and sigma to 10. Pseudotime was then calculated by running 80 flood simulations with *Pou5f1*+ epiblast cells containing "Early" staged embryos (cluster 1 of WTvsMut mesoderm Seurat object at resolution 1.5) as the "root" cells. Clusters containing the most defined mesodermal derivative cell types and containing the "Late" staged embryos were set as the "tip" cells (C15-,C16-HPCs, C11-, C7-Endothelial, C20-CFMeso, C2-Allantois, C12-CMs, C29-CardiacMeso, C0-postLPM, C22, C26-LPM, C14-PrSoM-like, C4-postPrxM1,C18-Meso). The resulting URD tree was subsequently built by simulated random walks from each tip. Overlay of relative developmental stages from embryo data was used to show consensus in pseudotime estimations of cell trajectories. Overlay of *Smarcd3*-F6::eGFP and various cardiac marker genes such as *Nkx2-5*, *Myl7*, *Smarcd3*, *Tnnt2*, and various *Gata* transcription factors were used to identify the relevant cardiac-fated branching segments of the URD tree.

To identify differentially expressed genes in fate-related cells of cardiac-relevant branches, cell barcodes from relevant branch segments were extracted from the URD object and assigned their relevant segment branch identities in the corresponding Seurat object. Differential gene testing using the Wilcoxon rank sum test (min.pct = 0.1,

logfc threshold = 0.25) was then performed between genotypes within a segment or between noted segments related in their pseudotemporal progression. Differentially expressed genes with adjusted p-values less than 0.05 were plotted as violin plots in Seurat and representative genes were overlayed on the URD tree to visualize expression patterns in pseudotime space.

Single cell Assay for Transposase Accessible Chromatin (scATAC-seq) library generation

For scATACseq library generation we used the 10X Genomics Chromium, scATACseq library kit v1 (10X Genomics, 1000110) and Chromium Chip E (10X Genomics, 1000156) according to manufacturer's protocols. Embryos were dissected and dissociated into single cells as described above and cells were resuspended in pre-chilled Lysis buffer for isolation of single nuclei. A maximum of 10,000 nuclei per sample were subjected to transposition and loaded into the 10X Genomics Chromium instrument. Final libraries were pooled and sequenced shallowly according to 10X protocol parameters on a NextSeq500 (Illumina). Littermate, stage-matched comparisons comprising a total of 5 control and 4 *Mesp1* KO embryos were ultimately re-pooled and sequenced together for deep sequencing on a NovaSeq6000 S4 lane (Illumina). All libraries were sequenced to depths of at least 24,000 median fragments per cell, and at most 35,000 median fragments per cell.

Processing raw scATAC-seq

Raw sequencing reads were processed using the 10X Cellranger ATAC v1.2.0 software pipeline. Reads were demultiplexed using cellranger-atac mkfastq. Cell barcodes were filtered and aligned to the Mm10 reference genome using cellranger-atac count. The resulting output indexed fragment files from each library were not aggregated and served as inputs for downstream computational analysis in ArchR (Granja et al., 2021).

ArchR analysis of scATAC-seq

Downstream computational analysis of scATAC-seq data was done with the ArchR software package v1.0.1 in R (Granja et al., 2021). Initial Arrow files were generated for all samples from inputs of respective indexed fragment files and sample meta-data. Samples from embryos aged E7.5 were called “Middle” stage (libraries ALK10_5_E75_con_rep1, ALK10_3_E75_con_rep2, ALK10_1_lateE75_con_rep1, ALK10_7_E75_Mesp1KO_rep1, ALK10_2_E75_Mesp1KO_rep2). Samples from embryos aged E7.75 were called “Late” stage (libraries ALK09_3_E775_con_rep1, ALK09_2_E775_con_rep2, ALK09_1_E775_Mesp1KO_rep1, ALK10_6_E775_Mesp1KO_rep2). The function createArrowFiles was run on each sample, removing cells with a transcription start site (TSS) enrichment score less than 4, and fragments less than 5000. This initialization also creates a genome-wide TileMatrix of 500 base pair bins and a weighted calculation of accessibility within and surrounding gene loci annotated from the Mm10 genome, called a GeneScoreMatrix. While CellRanger v1.2.0 implements removal of multi-cell capture, ArchR recommends an additional round of cell doublet removal using functions addDoubletScores and

filterDoublets. Individual ArrowFiles for each sample were aggregated into a single WTvsMut whole embryo ArchRProject containing 46,819 cells (26,295 control, 20,524 *Mesp1* KO) with a median TSS enrichment score of 10.675 and median of 30,703 fragments per cell. Dimensionality reduction was performed with addIterativeLSI (2 iterations, resolution 0.2, 30 dimensions). Clustering was performed using addClusters with “Seurat” method (resolution 0.8) and addUMAP was used to embed values for dimensionality reduced visualizations with the function plotEmbedding. Relative cell-type annotation of clusters was performed with consideration of combined information from GeneScore plots and label transfer from the complementary annotated whole embryo WTvsMut scRNA-seq Seurat analysis object of stage-matched control and *Mesp1* KO embryos for the relative Middle (embryos ALK07_3_lateE65_con_rep1, ALK07_14_E70_con_rep1, ALK08_11_E70_Mesp1KO_rep1, ALK07_7_earlyE75_con_rep1) and Late (embryos ALK07_6_lateE75_con_rep1, ALK04_3_lateE75_con_rep2, ALK05_7_E775_con_rep1, ALK05_2_lateE75_Mesp1KO_rep1, ALK07_8_E775_con_rep2) stages. For scRNA-seq integration, the addGeneIntegrationMatrix function utilizes Seurat’s FindTransferAnchors to perform Canonical Correlation Analysis. Relevant mesoderm clusters ("C15", "C9", "C24", "C17", "C16", "C18", "C12", "C11", "C8") were identified based on relative overlay of scRNA-seq cell type labels onto scATAC-seq clusters and GeneScoreMatrix for key marker genes, and subsetted into a WTvsMut mesoderm ArchRProject containing 25,848 cells (14,212 control and 11,636 *Mesp1* KO).

Dimensionality reduction was performed on the subsetted WTvsMut mesoderm ArchRProject with addIterativeLSI (4 iterations, resolution 0.2, 30 dimensions), which

was then batch corrected using addHarmony. Harmonized clustering was then performed using addClusters with “Seurat” method (resolution 0.8) and addUMAP was performed. Clusters were visualized using plotEmbedding. Relative cell-type annotation of clusters was again performed following integration with the mesoderm WTvsMut complementary, annotated, Seurat analysis scRNA-seq object from stage-matched control and *Mesp1* KO embryos for the relative Middle and Late stages. The addGeneIntegrationMatrix function was used to generate GeneIntegration plots, which were compared to GeneScore plots for understanding of cluster markers. A Jaccard Similarity Analysis from the predicted scRNA-seq integration for scATAC-seq clusters annotation was performed similarly to as described (Sarropoulos et al., 2021) to assess the strength of predictive labels, and the resulting proportions were visualized with the pheatmap function from the ComplexHeatmap R package (Gu, Eils & Schlesner, 2016). Cluster identities from the mesoderm subset scATAC-seq dimensionality reduction were utilized for downstream cluster-wise analyses.

Peak calling and motif enrichment:

Peaks were called using pseudo-bulkification and MACS2. Cell replicates for pseudobulks were created using addGroupCoverages on scATAC-seq clusters (40 minimum and 500 maximum cells in a replicate, minimum 2 replicates per cluster, 0.8 sampling ratio, kmerlength for Tn5 bias correction of 6). Peaks were called using addReproduciblePeakSet (500 peaks per cell, 1.5E5 maximum peaks per cluster) with MACS2 (-75 base pair shift per Tn5 insertion, 150 basepair extension after shift, excluding mitochondrial chromosome genes and chromosome Y genes, with a q-value

significance cutoff 0.1). Peaks were then merged using ArchR's iterative overlap method. Cluster enriched marker peaks were identified with getMarkerFeatures (FDR <= 0.05, Log2FC >=1) and visualized with plotMarkerHeatmap. Cluster motif enrichment was ascertained with addMotifAnnotations using the CIS-BP database motif set. Cluster enriched motifs were visualized with peakAnnoEnrichment (FDR <=0.05, Log2FC >=1) and then the top 7 motifs per cluster were plotted with plotEnrichHeatmap and ComplexHeatmap. Single cell resolution motif enrichment was computed using the chromVAR package (Schep et al., 2017) by adding background peaks (addBgdPeaks) and then motif z-score deviations were computed per cell with addDeviationsMatrix. Motif enrichments were visualized in UMAP embeddings with plotEmbedding.

Pseudotime ordering of cardiogenic trajectory:

A pseudotime trajectory approximating the differentiation of progenitor cell types to mature cell types was curated using the addTrajectory function (preFilterQuantile = 0.9, postFilterQuantile = 0.9) to order cells along the trajectory backbone C6, C5, C12, C13, C7, C8, C14. This backbone represents the biologically relevant cardiogenic differentiation path; epiblast, EomesME, Mesp1ME, LPM2, LPM1, CMs/CPs. We leveraged ArchR's series of pseudotime vector calculations to fit and align individual cells based on their Euclidean distances to the defined backbone's cell type clusters' mean coordinates in order to fit a continuous trajectory path in batch corrected LSI dimensional space. This resulting path with scaled, per-cell pseduotime values was then visualized in UMAP space using the plotTrajectory function. We then performed an integrative analysis to identify positive TF regulators along trajectory pseudotime. We

integrated gene accessibility scores and gene expression data with motif accessibility across pseudotime using correlateTrajectories function and visualized correlated matrices in trajectory space with plotTrajectoryHeatmap function.

Assessment of positive transcription factor regulators:

A putative positive regulator represents a TF whose gene expression is positively correlated to changes in accessibility of its corresponding motifs. Using the previously calculated motif z-score deviations, we stratified motif z-scores variation between all clusters to identify the maximum motif z-score delta. We next used the correlateMatrices function to correlate motifs to gene expression in batch-corrected LSI dimensional space, then used these correlations to identify motifs with maximized deviance from expected accessibility averages in other cells, and ranked TFs accordingly. We required positive TF regulators to have correlations greater than 0.5 (and adjusted p value < 0.01) between their gene expression and corresponding motifs, and deviation z-scores with maximum inter-cluster variation difference in the top quartile (quantile 0.75). Correlations were plotted for visualization using the ggplot function. While the ranking association with analysis might be vulnerable to generating false-negatives, wherein potential TF drivers aren't recognized, we found overlay of motifs with TF gene expression and gene score values along the cardiogenic trajectory and in UMAP cluster space served to sufficiently identify the highest confidence drivers.

Differential peaks and differential motif enrichment comparisons between cell types:

Pairwise comparisons between cell types of accessible peak differences was performed using the getMarkerFeatures function (Wilcoxon test, TSS enrichment and log10(nFrags) bias, 100 nearby cells for biased-matched background, 0.8 buffer ratio, 500 maximum cells) by setting one cell type as the lead comparison (useGroup) and one cell type as the relative comparison (bgdGroup). These pairwise comparisons of differential peaks were saved as .RDS objects and served as inputs to the (peak,gene) association analyses with rGreat (below in Methods). Differentially enriched peaks (FDR<= 0.05, abs(Log2FC) >=1) were visualized as MA plots. Motif enrichment of differential peaks was determined using the peakAnnoEnrichment function (FDR<= 0.05 and Log2FC >=1 for useGroup enrichment or else Log2FC<= -1 for bgdGroup enrichment) to determine motifs enriched in differential peaks between cell type groups. Enriched motifs were rank-sorted and colored by significance of enrichment, then plotted using the ggplot function.

Assessment of peak-to-gene linkages:

Peak-to-gene linkage analysis to assess correlations between chromatin accessibility and gene expression was performed using the addPeak2GeneLinks function on batch corrected LSI dimensions (correlation cut off > 0.45, FDR < 1E-4, resolution 1000 bp for optimized browser track visualization). Peak-to-gene linkages for differentially expressed genes (identified in scRNA-seq analyses) were visualized with cell type cluster browser tracks using plotBrowserTrack.

Association between scATAC-seq differential peaks and scRNA-seq differentially expressed genes

The rGREAT1(v1.26.0) bioconductor R package (Gu, 2022) was used to generate gene lists linked to scATAC-seq differential peaks based on gene regulatory domains defined as 5 kb upstream, 1 kb downstream of the Transcription Start Site (TSS) and up to 100 kb to the nearest gene. The log Fold Change (logFC) for the (peak,gene) pairs where the peak was differentially accessible (FDR <= 0.05, Log2FC >=1) were plotted to show how the log fold change of the gene expression is associated with the log fold change of the accessibility of peaks. The (peak,gene) pairs in the top-right quadrant (Q3) of the plot correspond to differentially open peaks linked with genes whose expressions are up-regulated. Similarly, the (peak,gene) pairs in the bottom-left quadrant (Q1) correspond to differentially closed peaks linked with genes whose expressions are down-regulated. Fisher's test (Pearce, 1992) was performed on the counts of (peak, gene) pairs in each of the four quadrants; up-regulated genes:differentially open peak regions, down-regulated genes:differentially closed peak regions, up-regulated genes:differentially closed peak regions and down-regulated genes:differentially open peak regions. This provided an estimate of the ratio of the odds of upregulated genes linked to differentially open peak regions versus the odds of up-regulated genes linked to differentially closed peak regions.

Whole mount fluorescent *in situ* hybridization experiments

Validation of spatial gene expression and differentially expressed genes was conducted in stage-matched, littermate whole-mount embryos. The assay for whole-mount embryo

in situ was adapted from the optimized whole-mount zebrafish embryo protocol using the RNAscope Multiplex Fluorescent Reagent Kit v2 and ProteasePlus (ACDBio) for embryo permeabilization as previously described (Gross-Thebing, Paksa & Raz, 2014; Soysa et al., 2019). De-yolked whole embryos were fixed in 4% paraformaldehyde solution (Electron Microscopy Sciences 15710) overnight at 4°C. Embryos were then washed 2x in PBST and processed through 10 min incubations in a dehydration series of 25%, 50%, 75%, 100% methanol on ice. Embyros were stored in 100% methanol at -20°C short term until initiation of the *in situ* hybridization protocol. Yolk sac DNA or anterior proximal extraembryonic regions prior to fixation were used for genotyping. Catalogue numbers for ACDBio RNAscope probes used in this study: eGFP (400281-C1, -C2, -C4), Tdgf1 (506411-C1), Fgf8 (313411-C1), Eomes (429641-C2), Myl7 (584271-C3), Anxa2 (501011-C2), Nkx2-5 (428241-C2). Whole-mount embryos were imaged in cold PBS using an upright epifluorescent microscope (Leica MZFLIII, Leica DFC 3000G, Lumen Dynamics XCite 120LED) and acquisition software LASX (Leica). Control and *Mesp1* KO embryo comparisons were processed with identical parameters.

Whole-mount embryo X-gal staining and imaging

X-gal staining for LacZ enhancer activity was performed according to standard protocols (Anderson et al., 2004; Materna et al., 2018; Sinha et al., 2015; Wilkinson & Nieto, 1993). Briefly, embryos were fixed in 4% paraformaldehyde at 4°C and stored in PBS until initiation of standard X-gal staining protocol. Littermate embryos were processed and imaged identically and simultaneously in brightfield using a Leica MZ165 FC stereomicroscope with DFC450 camera. Genotyping done after blind processing.

Whole-mount embryo immunostaining and light sheet imaging

Dissected embryos were fixed in 4% paraformaldehyde for 1 hour at room temperature with gentle agitation, washed in PBS, and stored in PBS + 0.2% sodium azide short-term at 4°C until initiation of immunostaining. Immunostaining was performed in PCR strip tubes. Embryos were incubated in blocking solution; PBS + 5% normal donkey serum, 0.2% sodium azide, 0.5% Triton X-100 (Sigma, X100-500 mL) with 100 µg/mL unconjugated Fab fragment donkey anti-mouse (Jackson Immunoresearch, 715-007-003) for 2 hours at 37°C with gentle rocking agitation. Following PBS washes, primary staining was done in blocking solution overnight and subsequently washed with PBS. Secondary staining incubation was done in blocking solution for 2-3 hours protected from light, and embryos were subjected to final PBS washes. All steps of immunostaining protocol were done at 37°C with gentle rocking and rotation. Antibodies used in this study: sheep polyclonal Foxc2 (R&D, AF6989), chicken polyclonal GFP (Aves, GFP-1020), rabbit polyclonal Cre (Millipore, 69050). Light sheet embryo images were acquired using Z1 Light Sheet Microscope (Zeiss) and processed as described (Dominguez et al., 2022).

References

Ajima, R., Sakakibara, Y., Sakurai-Yamatani, N., Muraoka, M. & Saga, Y. (2021) Formal proof of the requirement of MESP1 and MESP2 in mesoderm specification and their transcriptional control via specific enhancers in mice. *Development*. 148 (20). doi:10.1242/dev.194613.

Akerberg, B.N., Gu, F., VanDusen, N.J., Zhang, X., Dong, R., et al. (2019) A reference map of murine cardiac transcription factor chromatin occupancy identifies dynamic and conserved enhancers. *Nature Communications*. 10 (1), 4907. doi:10.1038/s41467-019-12812-3.

Alexanian, M., Maric, D., Jenkinson, S.P., Mina, M., Friedman, C.E., et al. (2017) A transcribed enhancer dictates mesendoderm specification in pluripotency. *Nature Communications*. 8 (1), 1806. doi:10.1038/s41467-017-01804-w.

Anderson, J.P., Dodou, E., Heidt, A.B., Val, S.J.D., Jaehnig, E.J., Greene, S.B., Olson, E.N. & Black, B.L. (2004) HRC Is a Direct Transcriptional Target of MEF2 during Cardiac, Skeletal, and Arterial Smooth Muscle Development In Vivo. *Molecular and Cellular Biology*. 24 (9), 3757–3768. doi:10.1128/mcb.24.9.3757-3768.2004.

Barnes, R.M., Harris, I.S., Jaehnig, E.J., Sauls, K., Sinha, T., Rojas, A., Schachterle, W., McCulley, D.J., Norris, R.A. & Black, B.L. (2016) MEF2C regulates outflow tract alignment and transcriptional control of Tdgf1. *Development*. 143 (5), 774–779. doi:10.1242/dev.126383.

Bondue, A. & Blanpain, C. (2010) Mesp1. *Circulation Research*. 107 (12), 1414–1427. doi:10.1161/circresaha.110.227058.

Bondue, A., Lapouge, G., Paulissen, C., Semeraro, C., Iacovino, M., Kyba, M. & Blanpain, C. (2008) Mesp1 Acts as a Master Regulator of Multipotent Cardiovascular Progenitor Specification. *Cell Stem Cell*. 3 (1), 69–84. doi:10.1016/j.stem.2008.06.009.

Branney, P.A., Faas, L., Steane, S.E., Pownall, M.E. & Isaacs, H.V. (2009) Characterisation of the Fibroblast Growth Factor Dependent Transcriptome in Early Development. *PLoS ONE*. 4 (3), e4951. doi:10.1371/journal.pone.0004951.

Brown, C.O., Chi, X., Garcia-Gras, E., Shirai, M., Feng, X.-H. & Schwartz, R.J. (2004) The Cardiac Determination Factor, Nkx2-5, Is Activated by Mutual Cofactors GATA-4 and Smad1/4 via a Novel Upstream Enhancer*. *Journal of Biological Chemistry*. 279 (11), 10659–10669. doi:10.1074/jbc.m301648200.

Bruneau, B.G. (2013) Signaling and Transcriptional Networks in Heart Development and Regeneration. *Cold Spring Harbor Perspectives in Biology*. 5 (3), a008292. doi:10.1101/cshperspect.a008292.

Bruneau, B.G. (2008) The developmental genetics of congenital heart disease. *Nature*. 451 (7181), 943–948. doi:10.1038/nature06801.

Buenrostro, J.D., Wu, B., Litzenburger, U.M., Ruff, D., Gonzales, M.L., Snyder, M.P., Chang, H.Y. & Greenleaf, W.J. (2015) Single-cell chromatin accessibility reveals

principles of regulatory variation. *Nature*. 523 (7561), 486–490. doi:10.1038/nature14590.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. (2018) Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology*. 36 (5), 411–420. doi:10.1038/nbt.4096.

Cai, W., Zhou, W., Han, Z., Lei, J., Zhuang, J., Zhu, P., Wu, X. & Yuan, W. (2020) Master regulator genes and their impact on major diseases. *PeerJ*. 8, e9952. doi:10.7717/peerj.9952.

Chen, Y. & Cao, X. (2009) NFAT directly regulates Nkx2-5 transcription during cardiac cell differentiation. *Biology of the Cell*. 101 (6), 335–350. doi:10.1042/bc20080108.

Cheng, P., Andersen, P., Hassel, D., Kaynak, B.L., Limphong, P., Juergensen, L., Kwon, C. & Srivastava, D. (2013) Fibronectin mediates mesendodermal cell fate decisions. *Development*. 140 (12), 2587–2596. doi:10.1242/dev.089052.

Chiapparo, G., Lin, X., Lescroart, F., Chabab, S., Paulissen, C., Pitisci, L., Bondue, A. & Blanpain, C. (2016) Mesp1 controls the speed, polarity, and directionality of cardiovascular progenitor migration. *The Journal of Cell Biology*. 213 (4), 463–477. doi:10.1083/jcb.201505082.

Clark, C.D., Zhang, B., Lee, B., Evans, S.I., Lassar, A.B. & Lee, K.-H. (2013) Evolutionary conservation of Nkx2.5 autoregulation in the second heart field. *Developmental Biology*. 374 (1), 198–209. doi:10.1016/j.ydbio.2012.11.007.

Costello, I., Pimeisl, I.-M., Dräger, S., Bikoff, E.K., Robertson, E.J. & Arnold, S.J. (2011) The T-box transcription factor Eomesodermin acts upstream of Mesp1 to specify cardiac mesoderm during mouse gastrulation. *Nature Cell Biology*. 13 (9), 1084–1091. doi:10.1038/ncb2304.

Davis, T.L. & Rebay, I. (2017) Master regulators in development: Views from the Drosophila retinal determination and mammalian pluripotency gene networks. *Developmental Biology*. 421 (2), 93–107. doi:10.1016/j.ydbio.2016.12.005.

Devine, W.P., Wythe, J.D., George, M., Koshiba-Takeuchi, K. & Bruneau, B.G. (2014) Early patterning and specification of cardiac progenitors in gastrulating mesoderm. *eLife*. 3, e03848. doi:10.7554/elife.03848.

Dominguez, M.H., Krup, A.L., Muncie, J.M. & Bruneau, B.G. (2022) A spatiotemporal gradient of mesoderm assembly governs cell fate and morphogenesis of the early mammalian heart. *bioRxiv*. 2022.08.01.502159. doi:10.1101/2022.08.01.502159.

Doppler, S.A., Werner, A., Barz, M., Lahm, H., Deutsch, M.-A., Dreßen, M., Schiemann, M., Voss, B., Gregoire, S., Kuppusamy, R., Wu, S.M., Lange, R. & Krane, M. (2014) Myeloid Zinc Finger 1 (Mzf1) Differentially Modulates Murine

Cardiogenesis by Interacting with an Nkx2.5 Cardiac Enhancer. *PLoS ONE*. 9 (12), e113775. doi:10.1371/journal.pone.0113775.

Downs, K.M. & Davies, T. (1993) Staging of gastrulating mouse embryos by morphological landmarks in the dissecting microscope. *Development*. 118 (4), 1255–1266. doi:10.1242/dev.118.4.1255.

Farrell, J.A., Wang, Y., Riesenfeld, S.J., Shekhar, K., Regev, A. & Schier, A.F. (2018) Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science*. 360 (6392), eaar3131. doi:10.1126/science.aar3131.

Fernandez-Guerrero, M., Zdral, S., Castilla-Ibeas, A., Lopez-Delisle, L., Duboule, D. & Ros, M.A. (2021) Time-sequenced transcriptomes of developing distal mouse limb buds: A comparative tissue layer analysis. *Developmental Dynamics*. doi:10.1002/dvdy.394.

George, R.M. & Firulli, A.B. (2021) Deletion of a Hand1 lncRNA-Containing Septum Transversum Enhancer Alters lncRNA Expression but Is Not Required for Hand1 Expression. *Journal of Cardiovascular Development and Disease*. 8 (5), 50. doi:10.3390/jcdd8050050.

Granja, J.M., Corces, M.R., Pierce, S.E., Bagdatli, S.T., Choudhry, H., Chang, H.Y. & Greenleaf, W.J. (2021) ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nature Genetics*. 53 (3), 403–411. doi:10.1038/s41588-021-00790-6.

Gross-Thebing, T., Paksa, A. & Raz, E. (2014) Simultaneous high-resolution detection of multiple transcripts combined with localization of proteins in whole-mount embryos. *BMC Biology*. 12 (1), 55. doi:10.1186/s12915-014-0055-7.

Gu, Z., Eils, R. & Schlesner, M. (2016) Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*. 32 (18), 2847–2849. doi:10.1093/bioinformatics/btw313.

Gu, Z., 2022. *rGreat: Client for GREAT Analysis.* https://github.com/jokergoo/rGREAT, http://great.stanford.edu/public/html/

Guo, X., Xu, Y., Wang, Z., Wu, Y., Chen, J., Wang, G., Lu, C., Jia, W., Xi, J., Zhu, S., Jiapaer, Z., Wan, X., Liu, Z., Gao, S. & Kang, J. (2018) A Linc1405/Eomes Complex Promotes Cardiac Mesoderm Specification and Cardiogenesis. *Cell Stem Cell*. 22 (6), 893-908.e6. doi:10.1016/j.stem.2018.04.013.

Hafemeister, C. & Satija, R. (2019) Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biology*. 20 (1), 296. doi:10.1186/s13059-019-1874-1.

Haghverdi, L., Lun, A.T.L., Morgan, M.D. & Marioni, J.C. (2018) Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nature Biotechnology*. 36 (5), 421–427. doi:10.1038/nbt.4091.

Han, M.-Z., Xu, R., Xu, Y.-Y., Zhang, X., Ni, S.-L., Huang, B., Chen, A.-J., Wei, Y.-Z., Wang, S., Li, W.-J., Zhang, Q., Li, G., Li, X.-G. & Wang, J. (2017) TAGLN2 is a

candidate prognostic biomarker promoting tumorigenesis in human gliomas. *Journal of Experimental & Clinical Cancer Research*. 36 (1), 155. doi:10.1186/s13046-017-0619-9.

Han, Z., Yu, Y., Cai, B., Xu, Z., Bao, Z., et al. (2020) YAP/TEAD3 signal mediates cardiac lineage commitment of human-induced pluripotent stem cells. *Journal of Cellular Physiology*. 235 (3), 2753–2760. doi:10.1002/jcp.29179.

Haraguchi, S., Kitajima, S., Takagi, A., Takeda, H., Inoue, T. & Saga, Y. (2001) Transcriptional regulation of Mesp1 and Mesp2 genes: differential usage of enhancers during development. *Mechanisms of Development*. 108 (1–2), 59–69. doi:10.1016/s0925-4773(01)00478-6.

Harvey, R.P. (2002) Patterning the vertebrate heart. *Nature Reviews Genetics*. 3 (7), 544–556. doi:10.1038/nrg843.

Huang, G., Ma, L., Shen, L., Lei, Y., Guo, L., Deng, Y. & Ding, Y. (2022) MIF/SCL3A2 depletion inhibits the proliferation and metastasis of colorectal cancer cells via the AKT/GSK-3β pathway and cell iron death. *Journal of Cellular and Molecular Medicine*. 26 (12), 3410–3422. doi:10.1111/jcmm.17352.

Katsanou, V., Milatos, S., Yiakouvaki, A., Sgantzis, N., Kotsoni, A., Alexiou, M., Harokopos, V., Aidinis, V., Hemberger, M. & Kontoyiannis, D.L. (2009) The RNA-Binding Protein Elavl1/HuR Is Essential for Placental Branching Morphogenesis and Embryonic Development. *Molecular and Cellular Biology*. 29 (10), 2762–2776. doi:10.1128/mcb.01393-08.

Kelly, R.G. (2016) How Mesp1 makes a move. *The Journal of Cell Biology*. 213 (4), 411–413. doi:10.1083/jcb.201604121.

Kelly, R.G., Buckingham, M.E. & Moorman, A.F. (2014) Heart Fields and Cardiac Morphogenesis. *Cold Spring Harbor Perspectives in Medicine*. 4 (10), a015750. doi:10.1101/cshperspect.a015750.

Kitajima, S., Takagi, A., Inoue, T. & Saga, Y. (2000) MesP1 and MesP2 are essential for the development of cardiac mesoderm. *Development*. 127 (15), 3215–3226. doi:10.1242/dev.127.15.3215.

Klymiuk, I., Kenner, L., Adler, T., Busch, D.H., Boersma, A., Irmler, M., Fridrich, B., Gailus-Durner, V., Fuchs, H., Leitner, N., Müller, M., Kühn, R., Schlederer, M., Treise, I., Angelis, M.H. de & Beckers, J. (2012) In Vivo Functional Requirement of the Mouse Ifitm1 Gene for Germ Cell Development, Interferon Mediated Immune Response and Somitogenesis. *PLoS ONE*. 7 (10), e44609. doi:10.1371/journal.pone.0044609.

Kokkinopoulos, I., Ishida, H., Saba, R., Ruchaya, P., Cabrera, C., Struebig, M., Barnes, M., Terry, A., Kaneko, M., Shintani, Y., Coppen, S., Shiratori, H., Ameen, T., Mein, C., Hamada, H., Suzuki, K. & Yashiro, K. (2015) Single-Cell Expression Profiling Reveals a Dynamic State of Cardiac Precursor Cells in the Early Mouse Embryo. *PLoS ONE*. 10 (10), e0140831. doi:10.1371/journal.pone.0140831.

Lange, U., Saitou, M., Western, P., Barton, S. & Surani, M. (2003) The Fragilis interferon-inducible gene family of transmembrane proteins is associated with

germ cell specification in mice. *BMC Developmental Biology*. 3 (1), 1–1. doi:10.1186/1471-213x-3-1.

Lescroart, F., Chabab, S., Lin, X., Rulands, S., Paulissen, C., Rodolosse, A., Auer, H., Achouri, Y., Dubois, C., Bondue, A., Simons, B.D. & Blanpain, C. (2014) Early lineage restriction in temporally distinct populations of Mesp1 progenitors during mammalian heart development. *Nature Cell Biology*. 16 (9), 829–840. doi:10.1038/ncb3024.

Lescroart, F., Wang, X., Lin, X., Swedlund, B., Gargouri, S., Sànchez-Dànes, A., Moignard, V., Dubois, C., Paulissen, C., Kinston, S., Göttgens, B. & Blanpain, C. (2018) Defining the earliest step of cardiovascular lineage segregation by single-cell RNA-seq. *Science*. 359 (6380), eaao4174. doi:10.1126/science.aao4174.

Liberatore, C.M., Searcy-Schrick, R.D., Vincent, E.B. & Yutzey, K.E. (2002) Nkx-2.5 Gene Induction in Mice Is Mediated by a Smad Consensus Regulatory Region. *Developmental Biology*. 244 (2), 243–256. doi:10.1006/dbio.2002.0604.

Lien, C.-L., McAnally, J., Richardson, J.A. & Olson, E.N. (2002) Cardiac-Specific Activity of an Nkx2–5 Enhancer Requires an Evolutionarily Conserved Smad Binding Site. *Developmental Biology*. 244 (2), 257–266. doi:10.1006/dbio.2002.0603.

Lien, C.L., Wu, C., Mercer, B., Webb, R., Richardson, J.A. & Olson, E.N. (1999) Control of early cardiac-specific transcription of Nkx2-5 by a GATA-dependent enhancer. *Development*. 126 (1), 75–84. doi:10.1242/dev.126.1.75.

Lin, X., Swedlund, B., Ton, M.-L.N., Ghazanfar, S., Guibentif, C., Paulissen, C., Baudelet, E., Plaindoux, E., Achouri, Y., Calonne, E., Dubois, C., Mansfield, W., Zaffran, S., Marioni, J.C., Fuks, F., Göttgens, B., Lescroart, F. & Blanpain, C. (2022) Mesp1 controls the chromatin and enhancer landscapes essential for spatiotemporal patterning of early cardiovascular progenitors. *Nature Cell Biology*. 24 (7), 1114–1128. doi:10.1038/s41556-022-00947-3.

Lindsley, R.C., Gill, J.G., Murphy, T.L., Langer, E.M., Cai, M., Mashayekhi, M., Wang, W., Niwa, N., Nerbonne, J.M., Kyba, M. & Murphy, K.M. (2008) Mesp1 Coordinately Regulates Cardiovascular Fate Restriction and Epithelial-Mesenchymal Transition in Differentiating ESCs. *Cell Stem Cell*. 3 (1), 55–68. doi:10.1016/j.stem.2008.04.004.

Liu, Y. (2017) Earlier and broader roles of Mesp1 in cardiovascular development. *Cellular and Molecular Life Sciences*. 74 (11), 1969–1983. doi:10.1007/s00018-016-2448-y.

MacNeill, C., French, R., Evans, T., Wessels, A. & Burch, J.B.E. (2000) Modular Regulation of cGATA-5 Gene Expression in the Developing Heart and Gut. *Developmental Biology*. 217 (1), 62–76. doi:10.1006/dbio.1999.9539.

Madisen, L., Zwingman, T.A., Sunkin, S.M., Oh, S.W., Zariwala, H.A., Gu, H., Ng, L.L., Palmiter, R.D., Hawrylycz, M.J., Jones, A.R., Lein, E.S. & Zeng, H. (2010) A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nature Neuroscience*. 13 (1), 133–140. doi:10.1038/nn.2467.

Materna, S.C., Sinha, T., Barnes, R.M., Bueren, K.L. van & Black, B.L. (2018) Cardiovascular development and survival require Mef2c function in the myocardial but not the endothelial lineage. *Developmental Biology*. 445 (2), 170–177. doi:10.1016/j.ydbio.2018.12.002.

Meijer, H.A., Pavert, S.A.V.D., Stroband, H.W.J. & Boerjan, M.L. (2000) Expression of the organizer specific homeobox gene Goosecoid (gsc) in porcine embryos. *Molecular Reproduction and Development*. 55 (1), 1–7. doi:10.1002/(sici)1098-2795(200001)55:1<1::aid-mrd1>3.0.co;2-a.

Migeotte, I., Grego-Bessa, J. & Anderson, K.V. (2011) Rac1 mediates morphogenetic responses to intercellular signals in the gastrulating mouse embryo. *Development*. 138 (14), 3011–3020. doi:10.1242/dev.059766.

Molkentin, J.D., Antos, C., Mercer, B., Taigen, T., Miano, J.M. & Olson, E.N. (2000) Direct Activation of a GATA6 Cardiac Enhancer by Nkx2.5: Evidence for a Reinforcing Regulatory Network of Nkx2.5 and GATA Transcription Factors in the Developing Heart. *Developmental Biology*. 217 (2), 301–309. doi:10.1006/dbio.1999.9544.

Moulik, M., Vatta, M., Witt, S.H., Arola, A.M., Murphy, R.T., McKenna, W.J., Boriek, A.M., Oka, K., Labeit, S., Bowles, N.E., Arimura, T., Kimura, A. & Towbin, J.A. (2009) ANKRD1, the Gene Encoding Cardiac Ankyrin Repeat Protein, Is a Novel Dilated Cardiomyopathy Gene. *Journal of the American College of Cardiology*. 54 (4), 325–333. doi:10.1016/j.jacc.2009.02.076.

Nees, S.N. & Chung, W.K. (2019) Genetic Basis of Human Congenital Heart Disease. *Cold Spring Harbor Perspectives in Biology*. 12 (9), a036749. doi:10.1101/cshperspect.a036749.

Pearce, S.C. (1992) Breakthroughs in Statistics, Methodology and Distribution. *Springer Series in Statistics*. 59–65. doi:10.1007/978-1-4612-4380-9_5.

Pikkarainen, S., Tokola, H., Kerkelä, R. & Ruskoaho, H. (2004) GATA transcription factors in the developing and adult heart. *Cardiovascular Research*. 63 (2), 196–207. doi:10.1016/j.cardiores.2004.03.025.

Probst, S., Sagar, S., Tosic, J., Schwan, C., Grün, D. & Arnold, S.J. (2020) Spatiotemporal sequence of mesoderm and endoderm lineage segregation during mouse gastrulation. *Development*. 148 (1), dev193789. doi:10.1242/dev.193789.

Pursani, V., Pethe, P., Bashir, M., Sampath, P., Tanavde, V. & Bhartiya, D. (2017) Genetic and Epigenetic Profiling Reveals EZH2-mediated Down Regulation of OCT-4 Involves NR2F2 during Cardiac Differentiation of Human Embryonic Stem Cells. *Scientific Reports*. 7 (1), 13051. doi:10.1038/s41598-017-13442-9.

Reifers, F., Walsh, E.C., Leger, S., Stainier, D.Y. & Brand, M. (2000) Induction and differentiation of the zebrafish heart requires fibroblast growth factor 8 (fgf8/acerebellar). *Development*. 127 (2), 225–235. doi:10.1242/dev.127.2.225.

Rojas, A., Val, S.D., Heidt, A.B., Xu, S.-M., Bristow, J. & Black, B.L. (2005) Gata4 expression in lateral mesoderm is downstream of BMP4 and is activated directly by Forkhead and GATA transcription factors through a distal enhancer element. *Development*. 132 (15), 3405–3417. doi:10.1242/dev.01913.

Saga, Y. (1998) Genetic rescue of segmentation defect in MesP2-deficient mice by MesP1 gene replacement. *Mechanisms of Development*. 75 (1–2), 53–66. doi:10.1016/s0925-4773(98)00077-x.

Saga, Y., Hata, N., Kobayashi, S., Magnuson, T., Seldin, M.F. & Taketo, M.M. (1996) MesP1: a novel basic helix-loop-helix protein expressed in the nascent mesodermal cells during mouse gastrulation. *Development*. 122 (9), 2769–2778. doi:10.1242/dev.122.9.2769.

Saga, Y., Kitajima, S. & Miyagawa-Tomita, S. (2000) Mesp1 Expression Is the Earliest Sign of Cardiovascular Development. *Trends in Cardiovascular Medicine*. 10 (8), 345–352. doi:10.1016/s1050-1738(01)00069-x.

Saga, Y., Miyagawa-Tomita, S., Takagi, A., Kitajima, S., Miyazaki, J. i & Inoue, T. (1999) MesP1 is expressed in the heart precursor cells and required for the formation of a single heart tube. *Development*. 126 (15), 3437–3447. doi:10.1242/dev.126.15.3437.

Sang, T., Yang, J., Liu, J., Han, Y., Li, Y., Zhou, X. & Wang, X. (2021) AMOT suppresses tumor progression via regulating DNA damage response signaling in

diffuse large B-cell lymphoma. *Cancer Gene Therapy*. 28 (10–11), 1125–1135. doi:10.1038/s41417-020-00258-5.

Sarropoulos, I., Sepp, M., Frömel, R., Leiss, K., Trost, N., Leushkin, E., Okonechnikov, K., Joshi, P., Giere, P., Kutscher, L.M., Cardoso-Moreira, M., Pfister, S.M. & Kaessmann, H. (2021) Developmental and evolutionary dynamics of cis-regulatory elements in mouse cerebellar cells. *Science*. 373 (6558). doi:10.1126/science.abg4696.

Satija, R., Farrell, J.A., Gennert, D., Schier, A.F. & Regev, A. (2015) Spatial reconstruction of single-cell gene expression data. *Nature Biotechnology*. 33 (5), 495–502. doi:10.1038/nbt.3192.

Saykali, B., Mathiah, N., Nahaboo, W., Racu, M.-L., Hammou, L., Defrance, M. & Migeotte, I. (2019) Distinct mesoderm migration phenotypes in extra-embryonic and embryonic regions of the early mouse embryo. *eLife*. 8, e42434. doi:10.7554/elife.42434.

Schachterle, W., Rojas, A., Xu, S.-M. & Black, B.L. (2012) ETS-dependent regulation of a distal Gata4 cardiac enhancer. *Developmental Biology*. 361 (2), 439–449. doi:10.1016/j.ydbio.2011.10.023.

Schep, A.N., Wu, B., Buenrostro, J.D. & Greenleaf, W.J. (2017) chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nature Methods*. 14 (10), 975–978. doi:10.1038/nmeth.4401.

Schwartz, B., Marks, M., Wittler, L., Werber, M., Währisch, S., Nordheim, A., Herrmann, B.G. & Grote, P. (2014) SRF is essential for mesodermal cell migration during elongation of the embryonic body axis. *Mechanisms of Development*. 133, 23–35. doi:10.1016/j.mod.2014.07.001.

Searcy, R.D., Vincent, E.B., Liberatore, C.M. & Yutzey, K.E. (1998) A GATA-dependent nkx-2.5 regulatory element activates early cardiac gene expression in transgenic mice. *Development*. 125 (22), 4461–4470. doi:10.1242/dev.125.22.4461.

Seo, S. & Kume, T. (2006) Forkhead transcription factors, Foxc1 and Foxc2, are required for the morphogenesis of the cardiac outflow tract. *Developmental Biology*. 296 (2), 421–436. doi:10.1016/j.ydbio.2006.06.012.

Shi, D.-L. & Grifone, R. (2021) RNA-Binding Proteins in the Post-transcriptional Control of Skeletal Muscle Development, Regeneration and Disease. *Frontiers in Cell and Developmental Biology*. 9, 738978. doi:10.3389/fcell.2021.738978.

Simon, C.S., Downes, D.J., Gosden, M.E., Telenius, J., Higgs, D.R., Hughes, J.R., Costello, I., Bikoff, E.K. & Robertson, E.J. (2017) Functional characterisation of cis-regulatory elements governing dynamic Eomes expression in the early mouse embryo. *Development (Cambridge, England)*. 144 (7), 1249–1260. doi:10.1242/dev.147322.

Sinha, T., Lin, L., Li, D., Davis, J., Evans, S., Wynshaw-Boris, A. & Wang, J. (2015) Mapping the dynamic expression of Wnt11 and the lineage contribution of

Wnt11-expressing cells during early mouse development. *Developmental Biology*. 398 (2), 177–192. doi:10.1016/j.ydbio.2014.11.005.

Smemo, S., Campos, L.C., Moskowitz, I.P., Krieger, J.E., Pereira, A.C. & Nobrega, M.A. (2012) Regulatory variation in a TBX5 enhancer leads to isolated congenital heart disease. *Human Molecular Genetics*. 21 (14), 3255–3263. doi:10.1093/hmg/dds165.

Soibam, B., Benham, A., Kim, J., Weng, K., Yang, L., Xu, X., Robertson, M., Azares, A., Cooney, A.J., Schwartz, R.J. & Liu, Y. (2015) Genome-Wide Identification of MESP1 Targets Demonstrates Primary Regulation Over Mesendoderm Gene Activity. *STEM CELLS*. 33 (11), 3254–3265. doi:10.1002/stem.2111.

Soysa, T.Y. de, Ranade, S.S., Okawa, S., Ravichandran, S., Huang, Y., Salunga, H.T., Schricker, A., Sol, A. del, Gifford, C.A. & Srivastava, D. (2019) Single-cell analysis of cardiogenesis reveals basis for organ-level developmental defects. *Nature*. 572 (7767), 120–124. doi:10.1038/s41586-019-1414-x.

Stefanovic, S., Laforest, B., Desvignes, J.-P., Lescroart, F., Argiro, L., Maurel-Zaffran, C., Salgado, D., Plaindoux, E., Bono, C.D., Pazur, K., Théveniau-Ruissy, M., Béroud, C., Puceat, M., Gavalas, A., Kelly, R.G. & Zaffran, S. (2020) Hox-dependent coordination of mouse cardiac progenitor cell patterning and differentiation. *eLife*. 9, e55124. doi:10.7554/elife.55124.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P. & Satija, R. (2019) Comprehensive Integration of Single-Cell Data. *Cell*. 177 (7), 1888-1902.e21. doi:10.1016/j.cell.2019.05.031.

Takeuchi, J.K., Lickert, H., Bisgrove, B.W., Sun, X., Yamamoto, M., Chawengsaksophak, K., Hamada, H., Yost, H.J., Rossant, J. & Bruneau, B.G. (2007) Baf60c is a nuclear Notch signaling component required for the establishment of left–right asymmetry. *Proceedings of the National Academy of Sciences*. 104 (3), 846–851. doi:10.1073/pnas.0608118104.

Tosic, J., Kim, G.-J., Pavlovic, M., Schröder, C.M., Mersiowsky, S.-L., Barg, M., Hofherr, A., Probst, S., Köttgen, M., Hein, L. & Arnold, S.J. (2019) Eomes and Brachyury control pluripotency exit and germ-layer segregation by changing the chromatin state. *Nature Cell Biology*. 21 (12), 1518–1531. doi:10.1038/s41556-019-0423-1.

Vincentz, J.W., Firulli, B.A., Toolan, K.P., Arking, D.E., Sotoodehnia, N., Wan, J., Chen, P.-S., Vries, C. de G., Christoffels, V.M., Lohe, M.R. der & Firulli, A.B. (2019) Variation in a Left Ventricle–Specific Hand1 Enhancer Impairs GATA Transcription Factor Binding and Disrupts Conduction System Development and Function. *Circulation Research*. 125 (6), 575–589. doi:10.1161/circresaha.119.315313.

Vincentz, J.W., Firulli, B.A., Toolan, K.P., Osterwalder, M., Pennacchio, L.A. & Firulli, A.B. (2021) HAND transcription factors cooperatively specify the aorta and pulmonary trunk. *Developmental Biology*. doi:10.1016/j.ydbio.2021.03.011.

Wang, T., Yuan, J., Zhang, J., Tian, R., Ji, W., Zhou, Y., Yang, Y., Song, W., Zhang, F. & Niu, R. (2015) Anxa2 binds to STAT3 and promotes epithelial to mesenchymal transition in breast cancer cells. *Oncotarget*. 6 (31), 30975–30992. doi:10.18632/oncotarget.5199.

Weninger, W.J., Floro, K.L., Bennett, M.B., Withington, S.L., Preis, J.I., Barbera, J.P.M., Mohun, T.J. & Dunwoodie, S.L. (2005) Cited2 is required both for heart morphogenesis and establishment of the left-right axis in mouse development. *Development*. 132 (6), 1337–1348. doi:10.1242/dev.01696.

Wilkinson, D.G. & Nieto, M.A. (1993) Detection of messenger RNA by in Situ hybridization to tissue sections and whole mounts. *Methods in Enzymology*. 225, 361–373. doi:10.1016/0076-6879(93)25025-w.

Wu, S.M. (2008) Mesp1 at the Heart of Mesoderm Lineage Specification. *Cell Stem Cell*. 3 (1), 1–2. doi:10.1016/j.stem.2008.06.017.

Yin, J. & Wang, G. (2014) The Mediator complex: a master coordinator of transcription and cell lineage development. *Development*. 141 (5), 977–987. doi:10.1242/dev.098392.

Yuan, X., Scott, I.C. & Wilson, M.D. (2021) Heart Enhancers: Development and Disease Control at a Distance. *Frontiers in Genetics*. 12, 642975. doi:10.3389/fgene.2021.642975.

Yuan, X., Song, M., Devine, P., Bruneau, B.G., Scott, I.C. & Wilson, M.D. (2018) Heart enhancers with deeply conserved regulatory activity are established early in zebrafish development. *Nature Communications*. 9 (1), 4977. doi:10.1038/s41467-018-07451-z.

Zhu, L., Gomez-Duran, A., Saretzki, G., Jin, S., Tilgner, K., Melguizo-Sanchis, D., Anyfantis, G., Al-Aama, J., Vallier, L., Chinnery, P., Lako, M. & Armstrong, L. (2016) The mitochondrial protein CHCHD2 primes the differentiation potential of human induced pluripotent stem cells to neuroectodermal lineages. *The Journal of Cell Biology*. 215 (2), 187–202. doi:10.1083/jcb.201601061.

Zhu, S., Wurdak, H., Wang, J., Lyssiotis, C.A., Peters, E.C., Cho, C.Y., Wu, X. & Schultz, P.G. (2009) A Small Molecule Primes Embryonic Stem Cells for Differentiation. *Cell Stem Cell*. 4 (5), 416–426. doi:10.1016/j.stem.2009.04.001.

Zug, R. (2022) Developmental disorders caused by haploinsufficiency of transcriptional regulators: a perspective based on cell fate determination. *Biology Open*. 11 (1), bio058896. doi:10.1242/bio.058896.

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:
Alexis Leigh Krup
5E968A103DB846F...
Author Signature

8/27/2022
Date