Understanding dose-dependent gene regulation using in vitro models of early human development

by
Emily Anne Bulger

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

in

Developmental and Stem Cell Biology

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
Jeremy Reiter
F07C889D1B164B3...

Jeremy Reiter
Chair

DocuSigned by:
Benoit Bruneau
[illegible]

Benoit Bruneau

DocuSigned by:
Jeffrey Bush
[illegible]

Jeffrey Bush

DocuSigned by:
Todd McDevitt
6E4BA4FF7EA4430...

Todd McDevitt

Committee Members

The work is dedicated to my family: Mom, Dad, and Elf.

Acknowledgments

When I first started graduate school, my first advisor, Todd McDevitt, told me that scientific significance would come from the body of work I would produce, but that the most important output of my Ph.D. would have far more to do with the type of scientist I would grow to become. The last six years have shaped me as both a scientist and a person in more ways than I can count, and I'm incredibly fortunate to have had so many role models, both scientific and non-scientific, who have helped me learn how to craft and answer questions about the world while keeping curiosity and compassion at the forefront. I am proud to present this dissertation and would like to acknowledge the endless gratitude I have for the community that made it possible.

First, I would like to thank my advisors, Todd McDevitt and Benoit Bruneau, who have enabled me to a scientific home in two labs that both have some of the most inspiring scientists I've had the opportunity to work with. To Todd, thank you for providing the type of mentorship that gave me the freedom to follow my curiosities and approaching my questions and confusions with support and kindness. I frequently walked into our meetings feeling chaotic in my project and walked out feeling like the chaos was still there but a little more manageable, and perhaps more importantly like I had the tools within myself to figure out the solution. To Benoit, thank you for providing the mentorship and resources to see my projects through and being a strong foundation in times of uncertainty. You taught me how to persevere in my experiments when a discovery was just around the corner and perhaps more importantly when to be willing to walk away from a project in pursuit of something better, which is a skill I will continue to carry into so many facets of my life.

I would like to thank the old guard of the McDevitt lab: Ariel Kauss, Ana Silva, Serah Kang, Sarah Rockwood, Marty Krakora, Ron Manlapaz, Vaishaali Natarajan, Oriane Mathys, Olaia Fernandez-Vila, and particularly Ashley Libby and David Joy for teaching me the vast majority of the scientific tools I know, lovingly editing my half-baked grant proposals into a sea of red, and providing the best pie manifesto a person could ask for. From lab meetings to Zoom-based

qualifying exams to the caves of Portugal, your constructive criticism, patience, and silliness helped me grow so much as a scientist and I am incredibly grateful for your continued mentorship. To Lana Zholudeva, thank you for pushing me to see all of the possibilities in front of me, and providing an endless stream of much-needed kitten pictures. To Jasmine King, Ivana Muncie-Vasic, Will Flanigan, and Nick Elder, the NE to the NEEB, thank you for making lab such a fun place to be and critiquing my science in a way only great friends can. There's no one else I'd rather annoy with Evo-Devo facts relentlessly, and I'm so grateful have gotten to do this grad school journey together.

I am fortunate to have not just one supportive lab community but two, and I'd like to thank the members of Bruneau Lab for their endless support: Kelly Hayes, Swetansu Hota, Irfan Kathiriya, Kavitha Rao, Kevin Hu, Megan Matthews, Emily Brower, Jon Muncie-Vasic, Alexis Krup, Austin Hsu, Jingshing Wu, Zoe Grant, Abe Horrillo, Vasu Kameswaran, Carine Joubran, Nicole Greenberg, and Nicole Velasquez. You all infuse the lab with the perfect combination of scientific rigor, silliness, Taylor Swift fandom, and genuine compassion when life takes turns we don't expect, and I feel so fortunate to have been welcomed into and grown as a part of this community.

I would like to thank the Developmental and Stem Cell Biology graduate program and Gladstone Institutes for providing me with the opportunity and resources to make this Ph.D. a reality. To the members of my Thesis Committee, Jeremy Reiter and Jeff Bush, thank you for on so many occasions helping me see the forest through the trees. I am a far better scientist for your guidance. To Demian Sainz and Sudha Krishnamurthy, thank you for being incredible advocates for your students. And to Nicole Koutsodendris, Matt Keefe, Dylan Maghini, Antara Rao, Maxine Nelson, and so many other close friends I've met through these communities, thank you for standing in my corner these last six years. I am so incredibly grateful for you.

I'd next like to thank the mentors who made graduate school a real possibility for me. To Rosemary Gillespie and Katrina Lohan, thank you for taking a chance on me as an undergraduate and showing me what it looks like to be curious as a career. To Ethan Bier, Valentino Gantz, and

the rest of the Bier Lab at UCSD, thank you for teaching me how to think like a scientist and showing me what I am capable of. My experiences within the lab solidified my decision to go to graduate school and I am continuously thankful for your mentorship.

I have a special place in my heart for the friends who knew me before I knew myself, as much as that is ever a finished process. To Tania Carrillo, oh BABY how much has happened since we sang the cups song in Ms. Earp's class. Thank you for being my forever cheerleader; here's to one more story for the Rocket. To Annie Sturges and Lauren Roger, thank you for listening to my hours of triumphs and anxieties, and for coming up with some of the most genuinely creative ways to explain my job to people. Ender's game has nothing on us.

Finally, I am so deeply thankful for my family and their unrelenting love and support. To my mom, Beth Bohlman, my dad, Steve Bulger, and my step-dad Rick Bohlman, thank you for giving me every tool to let me explore who I want to become and make mistakes along the way. Thank you for being a source of strength and amusement and for asking the best questions when I explain my science, and, Mom, for seeing me through the best and the worst days with a morning song and unrelenting optimism. Words can't express how important you are to me and all the ways you've shaped me, and I am so, so grateful for each of you.

Contributions

The majority of the material in this dissertation is accessible via the references below:

Bulger, E.A., McDevitt, T.C. and Bruneau, B.G. (2024) “CDX2 dose-dependently influences the gene regulatory network underlying human extraembryonic mesoderm development.,” *Biology open* [Preprint]. doi:10.1242/bio.060323.

Bulger, E.A. *et al.* (2024) “TBXT dose sensitivity and the decoupling of nascent mesoderm specification from EMT progression in 2D human gastruloids.,” *Development* [Preprint]. doi:10.1242/dev.202516.

When the light turns green, you go. When the light turns red, you stop. But what do you do when the light turns blue with orange and lavender spots?

– Shel Silverstein

Understanding dose-dependent gene regulation using i*n vitro* models of early human development

Emily A. Bulger

Abstract

Precise regulation of both the timing and dosage of gene expression is critical for the development of the early embryo and the extraembryonic tissues that support it. These signals enable molecular communication within and between tissues, thereby allowing cells to differentiate to the correct identity and migrate to the correct location for patterned organ systems to emerge. Some of the genes active during early development orchestrate morphogenesis in a dose-dependent manner, whereby reduced expression fulfills some of the gene's functions, but is insufficient to drive normal patterning. How gene dosage informs morphogenesis is not well understood. This dissertation focuses on how varying expression levels of two transcription factors with dose-dependent phenotypes, TBXT and CDX2, influence the patterning and kinetics of the embryonic and extraembryonic mesoderm populations during gastrulation. By utilizing *in vitro* models of early human gastrulation and mesoderm development, I demonstrate that TBXT dosage directly influences the temporal progression of the epithelial-to-mesenchymal transition (EMT) in the nascent mesoderm and hypothesize that this influences the temporal and spatial migration kinetics surrounding primitive streak morphogenesis. I additionally show that CDX2 dosage influences the gene regulatory network (GRN) underlying extraembryonic mesoderm development despite heterozygous expression being sufficient to maintain a wild-type-like chromatin accessibility profile. This work suggests that the regulation of downstream gene expression is not solely dependent on chromatin remodeling and implies that proper regulation of this GRN is potentially critical for the development of extraembryonic structures such as the allantois. These findings clarify how varying dosages of specific transcription factors can influence the gene regulatory networks underlying early gastrulation, thereby contributing to our understanding of both dose-dependent gene regulation and early human development.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AU	*arbitrary units*
BMP	*bone morphogenic protein*
CADM	*cell adhesion molecule*
CDX2	*caudal type homeobox 2*
CDX2-Het	*caudal type homeobox 2 heterozygous*
CDX2-KO	*caudal type homeobox 2 knock-out*
ChIP-seq	*chromatin immunoprecipitation with sequencing*
CRISPR	*clustered regularly interspaced short palindromic repeats*
cWNT	*canonical wingless-related integration site*
DARs	*differentially accessible regions*
DEGs	*differentially expressed genes*
DNA	*deoxyribonucleic acid*
ECM	*extracellular matrix*
EEM Progenitors	*extraembryonic progenitors*
EMT	*epithelial to mesenchymal transition*
Epi	*epiblast*
ExeM	*extraembryonic mesoderm*
GFP	*green fluorescence protein*
GO	*gene ontogeny*
hESC/mESC	*human/mouse embryonic stem cell*
hiPSC	*human induced pluripotent stem cell*
HOX	*homeobox*
ICM	*inner cell mass*
IF	*immunofluorescence*
Indel	*insertion deletion*

JAM	*junctional adhesion molecule*
L-R	*ligand-receptor*
Meso	*mesoderm*
MIM	*mesoderm induction media*
ncWNT	*noncanonical wingless-related integration site*
NMPs	*neuromesodermal progenitors*
PCR	*polymerase chain reaction*
PGCLC	*primordial germ cell-like cells*
PS	*primitive streak*
RI	*rho-associated protein kinase inhibitor (see also ROCKi)*
RNA	*ribonucleic acid*
ROCKi	*rho-associated protein kinase inhibitor (see also RI)*
sgRNA	*single guide ribonucleic acid*
snATAC-seq	*single nucleus assay for transposase-accessible chromatin with sequencing*
snRNA-seq	*single nucleus assay RNA sequencing*
SWI-SNF	*switch-sucrose non-fermentable*
TBXT	*T-box transcription factor T, also Brachury/T/Bra/XBra*
TBXT-Het	*T-box transcription factor T heterozygous*
TBXT-KO	*T-box transcription factor T knock-out*
TE	*trophectoderm*
TF	*transcription factor*
VEGF	*vascular endothelial growth factor*
WNT	*wingless-related integration site*
WT	*wildtype*
WTC11	*Wildtype-C11 (human induced pluripotent stem cell line)*

Chapter 1: Introduction

Gastrulation and Early Development

Across phyla, cell movements and cell fate specification must be precisely orchestrated for a fertilized egg to develop into a multifaceted organism with complex organ systems. Initially, the morphogenesis of cells within the early embryo largely preserves radial symmetry as the single-celled zygote undergoes several rounds of cleavage to form the 16-celled morula (Muhr, Arbor and Ackerman, 2024). At this point, the morula "breaks symmetry", undergoing compaction and cavitation to form the blastocyst. The blastocyst consists of two layers – the outer shell of the trophoblast, which forms the placenta and enables the embryo to implant into the uterine lining, and the inner cell mass (ICM), which goes on to form the embryo proper (Wolpert *et al.*, 2019).

Cells of the ICM are considered pluripotent, meaning they are able to give rise to all the cell types of the adult organism. During week two of human development, ICM cells begin to flatten into a two-layered ("bilaminar") disc consisting of epiblast cells and hypoblast cells. This bilaminar disc is located between the amniotic cavity and the primitive yolk sac and is the first demarcation of the dorsal/ventral axis. The hypoblast will go on to surround and contribute to the yolk sac, while the epiblast will go on to for the three primary germ layers of the embryo proper in a process known as gastrulation (Palmer and Kaldis, 2016).

Gastrulation is initiated around the start of week 3 of human development and is denoted by the appearance of the primitive streak (PS), which is a groove that forms at the caudal end of the epiblast. This groove becomes visible as cells in this region thicken and adjacent cells proliferate and migrate toward the embryonic midline. Next, following a highly coordinated waterfall-like motion, epithelial cells near the primitive streak undergo an epithelial-to-mesenchymal transition (EMT) primarily regulated by the WNT, BMP, and FGF signaling pathways. This transition allows the cells to detach from the neighboring sheet of cells and ingress into the PS itself, settling between the epiblast layer and the underlying hypoblast layer (Winnier *et al.*, 1995; Sun *et al.*, 1999; Huelsken *et al.*, 2000; Ciruna and Rossant, 2001; Mohamed, Clarke and Dufort, 2004; Arnold and Robertson, 2009; Bardot and Hadjantonakis, 2020). The

first cells to ingress form the first primary germ layer, the endoderm, which will ultimately give rise to the gastrointestinal tract and lungs. Subsequently ingressing cells migrate between the epiblast layer and the newly established endoderm layer, forming the second germ layer, the mesoderm, which contributes to the musculoskeletal and circulatory systems. The remaining epiblast cells form the third and final germ layer, the ectoderm, which largely contributes to the nervous system and epidermis (Muhr, Arbor and Ackerman, 2024). Cells continue to proliferate and ingress as the embryo grows until the PS has receded around the end of the fourth week of development. At this point, the emergent structure, the gastrula, is primed for organ formation and will continue to develop based on interactions between and within the three primary germ layers.

Embryonic and Extraembryonic Mesoderm Development

Cells undergoing gastrulation rely on precise paracrine, juxtracrine, and mechanical cues to migrate to the correct location at the correct time. In human mesoderm specifically, the time and position at which a cell ingresses influences the signals it is exposed to, and the combination of these signals informs which mesodermal subtype the cell is fated to become (Kinder *et al.*, 1999). More posterior, earlier ingressing cells form the lateral plate mesoderm, which gives rise to the circulatory system, body cavity, pelvis, limb bones, and part of the extraembryonic mesoderm (Prummel, Nieuwenhuize and Mosimann, 2020). Extraembryonic mesoderm goes on to contribute to the amnion, allantois, chorion, and visceral yolk sac, making it critical for maternal-fetal communication and primitive erythropoiesis (Watson and Cross, 2005; Saykali *et al.*, 2019). This ingression is followed by the intermediate mesoderm, which gives rise to the kidneys and gonads, and then the paraxial mesoderm, which gives rise to the somites including tendons and muscles. The last, most anterior mesodermal cells to ingress give rise to the notochord, an important signaling center during embryonic development (de Bree, de Bakker and Oostra, 2018). Throughout and after gastrulation, a progenitor population remains near the tailbud that gives rise to both the spinal cord and adjacent somites. These bipotent progenitors, known as

neuromesodermal progenitors (NMPs), continue to contribute to the trunk as the nascent embryo undergoes axial elongation (Tzouanacou *et al.*, 2009; Henrique *et al.*, 2015).

TBXT and CDX2 in early development

Successful early embryonic development relies on gene expression at the correct time, location, and dosage. In the nascent mesoderm, levels of the transcription factor (TF) Brachyury (TBXT) must be precisely regulated to ensure cells exit the primitive streak and pattern the anterior-posterior axis. Loss of *Tbxt* in the mouse leads to an accumulation of cells along the embryonic midline and disrupts both embryonic and extraembryonic mesoderm development in a dose-dependent manner (Yanagisawa, Fujimoto and Urushihara, 1981; Hashimoto, Fujimoto and Nakatsuji, 1987; Rashbass *et al.*, 1991; Wilson *et al.*, 1995) (Fig. 1.1). In many aspects the loss of *Cdx2* phenocopies the loss of *Tbxt*, most significantly leading to axial truncations and stunted development of the allantois, an extraembryonic mesoderm-derived structure critical for nutrient delivery and waste removal in the early embryo (Chawengsaksophak *et al.*, 2004; Brooke-Bisschop *et al.*, 2017; Foley and Lohnes, 2022). Understanding how these genes influence early mesodermal development, and specifically how their varying dosage informs morphogenesis, is foundational for disentangling the signals driving early embryonic development.

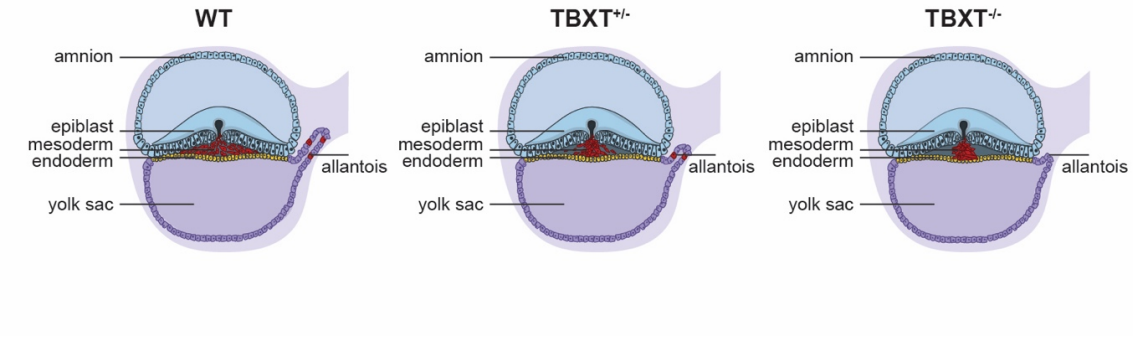


Figure 1.1: TBXT and CDX2 are required for migration out of the primitive streak.
Schematic of a developing embryo displaying normal (WT) and impaired ($TBXT^{+/-}$ and $TBXT^{-/-}$) cell migration out of the primitive streak and allantois outgrowth.

Focus of Dissertation

In this dissertation, I investigate how two genes critical for gastrulation, TBXT and CDX2, dose-dependently regulate mesoderm and extraembryonic mesoderm development, respectively.

In Chapter 2, I define the transcriptional consequences of TBXT dose reduction during early human gastrulation using human induced pluripotent stem cell (hiPSC)-based models of gastrulation and mesoderm differentiation. Multiomic single-nucleus RNA and single-nucleus ATAC sequencing of 2D gastruloids comprised of WT, TBXT heterozygous (TBXT-Het), or TBXT null (TBXT-KO) hiPSCs reveal that varying TBXT dosage does not compromise a cell's ability to differentiate into nascent mesoderm, but that the loss of TBXT significantly delays the temporal progression of the epithelial to mesenchymal transition (EMT). This delay is dependent on TBXT dosage, as cells heterozygous for TBXT proceed with EMT at an intermediate pace relative to WT or TBXT-KO. By differentiating iPSCs of the allelic series into nascent mesoderm in a monolayer format, I further illustrate that TBXT dose directly impacts the persistence of junctional proteins and cell-cell adhesions. These results demonstrate that EMT progression can be decoupled from the acquisition of mesodermal identity in the early gastrula and shed light on the mechanisms underlying human embryogenesis.

In Chapter 3, I investigate how CDX2 dose-dependently influences the gene regulatory network underlying extraembryonic mesoderm development. As with TBXT, I generate an allelic series for CDX2 in human induced pluripotent stem cells consisting of WT, heterozygous (CDX2-Het), and homozygous null (CDX2-KO) CDX2 genotypes, differentiate these cells in a 2D gastruloid model, and subject these cells to multiomic single-nucleus RNA and ATAC sequencing. By isolating the extraembryonic mesoderm population, I identify several genes dose-dependently regulated by CDX2 that are important for cytoskeletal integrity, adhesiveness, and cell permeability, including regulators of VEGF, canonical WNT, and non-canonical WNT signaling pathways. Single nucleus ATAC-seq reveals that heterozygous CDX2 expression is capable of maintaining a WT-like chromatin accessibility profile despite these dose-dependent gene

expression patterns, suggesting chromatin remodeling is not sufficient to drive the observed variability in gene expression. Finally, because the loss of CDX2 or TBXT phenocopy one another *in vivo*, we compare differentially expressed genes in our CDX2 knock-out model to those from TBXT knock-out hiPSCs differentiated in the analogous experiment. This comparison identifies several communally misregulated genes critical for cytoskeletal integrity and vasculogenesis *in vivo*, including *ANK3* and *ANGPT1*. This work suggests that the regulation of downstream gene expression is not solely dependent on chromatin accessibility and reveals components of the gene regulatory network that are potentially critical for the development of extraembryonic mesoderm-derived structures such as the allantois.

Together, these results inform how TBXT and CDX2 dose-dependently regulate gene expression in the embryonic and extraembryonic mesoderm and provide new insight into the gene regulatory networks underpinning the defects in somitogenesis, vascular development, and allantois development observed when TBXT or CDX2 expression is disrupted *in vivo.*

Chapter 2: TBXT dose sensitivity and the decoupling of nascent mesoderm specification from EMT progression in 2D human gastruloids

INTRODUCTION

In the early embryo, cells must precisely regulate gene expression to ensure the organism progresses through standard hallmarks of development. One key developmental timepoint in vertebrate embryogenesis is the establishment and morphogenesis of the primitive streak (PS), a transient structure of the posterior embryo that initiates germ layer formation and establishes bilateral symmetry (Mikawa *et al.*, 2004). Cells follow precisely orchestrated migration patterns as they ingress into the streak, undergo an epithelial-to-mesenchymal transition (EMT) to delaminate from the streak, and expand outward as individual mesenchymal cells to form endodermal and mesodermal germ lineages (Schoenwolf and Smith, 2000; Ciruna and Rossant, 2001; Dale *et al.*, 2006). Timing of epithelial cell ingression into the PS informs cell fate, as cells ingressing early during gastrulation give rise to cranial mesoderm (Lawson, Meneses and Pedersen, 1991), while subsequently ingressing cells form axial, paraxial, and lateral mesoderm of the anterior trunk (Wilson and Beddington, 1996). Posterior trunk mesoderm, including caudal somites, emerges later from a separate progenitor population in the tailbud known as the neuromesodermal progenitors (NMPs) (Tzouanacou *et al.*, 2009; Henrique *et al.*, 2015).

The transcription factor Brachyury (TBXT (human), also T/Bra (mouse)) has a conserved role in mesoderm differentiation across vertebrates (Technau, 2001) and is widely utilized as one of the first markers of nascent mesoderm. *T/Bra* is initially expressed in the posterior embryo just before the emergence of the PS, and as gastrulation progresses its expression domain expands to include the PS, notochord, and later the tailbud NMP population (Wilkinson, Bhatt and Herrmann, 1990; Rivera-Pérez and Magnuson, 2005). Still, how precisely *T/Bra* orchestrates the interplay between germ layer specification and PS morphogenesis is unclear. Human stem-cell-based models show TBXT is required for mesoderm induction (Bernardo *et al.*, 2011; Faial *et al.*, 2015) and studies in Xenopus show that the TBXT homolog, XBra, can induce different mesodermal cell types in a dose-dependent manner (O'Reilly, Smith and Cunliffe, 1995; Faial *et al.*, 2015). In contrast, mice with loss-of-function *T/Bra* mutations appear to generate an early

mesoderm population (Yanagisawa, Fujimoto and Urushihara, 1981; Hashimoto, Fujimoto and Nakatsuji, 1987; Rashbass *et al.*, 1991; Wilson *et al.*, 1995). Morphogenesis, however, is dramatically altered in these mutants, as *T/Bra*$^{-/-}$ mesoderm-like cells lose their ability to properly migrate away from the embryonic midline and accumulate in the node, PS, and regions immediately ventral to the streak. This aberrant cellular distribution, coupled with later-onset defects in tailbud mesoderm specification, ultimately causes mutant embryos to not develop a notochord, have significant body axis truncations rostral to somite seven, and die from incomplete allantois development. How *T/Bra* loss drives this phenotype, including if and how its mode of misregulation is conserved in humans, has not been fully defined.

Interestingly, this mutant axial truncation phenotype appears to be correlated to *T/Bra* dose. Mice heterozygous for *T/Bra* show a small but notable accumulation of cells in the same domains that see cell accumulation in *T/Bra* homozygous knock-out mice (Wilson, Rashbass and Beddington, 1993; Wilson *et al.*, 1995; Wilson and Beddington, 1997). Heterozygous mice are viable but later generate short tails and can display notochord and sacral malformations (Dobrovolskaia-Zavadskaia, 1927; Chesley, 1935; Stott, Kispert and Herrmann, 1993). In humans, hypomorphic *TBXT* expression manifests as partial absences or abnormal fusions of the tailbone, pelvis, or lower vertebrae, and, like in mouse, *TBXT*-loss-of-function mutations are embryonic lethal (Papapetrou *et al.*, 1999; Ghebranious *et al.*, 2008; Postma *et al.*, 2014; Chen *et al.*, 2023). The intermediate axial truncation phenotype seen in both human and mouse when TBXT expression is reduced suggests that the mechanism governing the spatial patterning of mesoderm once it exits the vertebrate PS is dependent on precisely calibrated expression levels of *TBXT*.

For TBXT+ cells to become motile and exit the PS, they must undergo EMT, including restructuring cell-cell adhesions, the cytoskeleton, and the composition of the extracellular matrix (ECM). In early amniote gastrulation, EMT is tightly associated with the acquisition of mesoderm fate, and it has been suggested that EMT initiates mesoderm commitment in hESCs (Evseenko

et al., 2010). TBXT also promotes EMT in the context of cancer (Fernando *et al.*, 2010; Roselli *et al.*, 2012), however, it has been proposed to be dispensable for EMT during notochord development (Zhu, Kwan and Mackem, 2016). How TBXT dosage modulates EMT in the context of mesoderm commitment to ensure cells acquire increased motility is therefore not fully defined.

Acquisition of cell fate preceding and throughout the establishment of the primitive streak is primarily controlled by a network consisting of BMP, WNT, and Nodal signaling (Arnold and Robertson, 2009). This network can be manipulated *in vitro* to yield hESC colonies, termed 2D gastruloids, that reproducibly generate concentric rings of epiblast, mesoderm, endoderm, and extraembryonic cells from the center outwards (Warmflash *et al.*, 2014; Minn *et al.*, 2020, 2021). 2D gastruloids have been applied to understanding the minimal inputs driving multicellular patterning and migration kinetics throughout the course of gastrulation, including PS morphogenesis (Libby *et al.*, 2019; Martyn, Siggia and Brivanlou, 2019; Joy, Libby and McDevitt, 2021).

Here, we adapt this 2D micropatterned gastruloid culture to investigate how TBXT dose-dependently controls mesodermal cell identity, EMT, and subsequent migratory behavior during early human gastrulation. We demonstrate that varying levels of TBXT expression modulate human PS morphogenesis by controlling the timing of EMT, including the persistence of junctional proteins and cell-cell adhesions, without compromising a cell's ability to differentiate into nascent mesoderm. We conclude that initial mesoderm specification can be decoupled from the temporal progression of EMT during early gastrulation, thus significantly improving our understanding of cell fate acquisition and morphogenesis during early human development.

RESULTS

Generation of hiPSC TBXT allelic series and 2D Gastruloids

To investigate the effect of TBXT dose on PS morphogenesis, we engineered a WTC11-LMNB1-GFP-derived hiPSC allelic series: *TBXT*$^{+/+}$ (WT), *TBXT*$^{+/-}$ (TBXT-Het), and *TBXT*$^{-/-}$ (TBXT-

KO), by targeting the first exon of TBXT with CRISPR/Cas9 (Fig. 1A, S1). The resulting indel created a premature stop codon in one or two alleles, respectively. The WT line used in subsequent experiments was derived from a subclone that was exposed to the TBXT sgRNA but remained unedited. Conducting a western blot for TBXT in a monolayer of cells exposed to a proprietary mesoderm induction media for 48 hours revealed the expected decrease in TBXT expression across the allelic series (Figs. 2.1B-B', methods). Of note, the TBXT protein level in the TBXT-Het was approximately 75-80% of the WT protein level, suggesting intrinsic dose compensation mechanisms are active in the TBXT gene regulatory network (Fig. 2.1B', S2.1C). This result was replicated in an independent heterozygous clone, which expressed TBXT at 91% of the WT level (Fig. S2.1). Knock-out efficiency was further confirmed through immunofluorescence (IF), which demonstrated the complete absence of TBXT protein in the TBXT-KO (Figs. 2.1D).

TBXT dose does not dramatically impact lineage emergence in 2D gastruloids

To begin to dissect how TBXT dose shapes the earliest stages of nascent mesoderm morphogenesis, we subjected the allelic series to 2D gastruloid differentiation, which reproducibly generates concentric rings of radially patterned primary germ layers, primordial germ cell-like cells (PGCLCs), and extraembryonic-like cells after 48 hours of BMP4 exposure (Warmflash *et al.*, 2014; Minn *et al.*, 2020) (Fig.2.1C, methods). We then conducted IF in colonies of each genotype to identify if and how TBXT expression broadly influences germ layers' patterning, proportion, and identity during early human gastrulation.

Using antibodies targeting epiblast/ectoderm (SOX2+), mesoderm (EOMES+/SOX17-), endoderm/PGCLCs (SOX17+), and extraembryonic cells (CDX2+), we observed a conserved spatial distribution of the germ layers between WT, TBXT-Het, and TBXT-KO colonies (Figs. 2.1D, F, S2.2). The consistent expression pattern of these broad germ layer markers suggests that the intrinsic BMP → WNT→ NODAL feedback loop is sustained in the absence of TBXT as predicted

in silico (Kaul *et al.*, 2023), and reflects the ability of nascent mesoderm to develop in murine *T/Bra*$^{-/-}$ models during the initial stages of gastrulation (Beddington, Rashbass and Wilson, 1992; Rashbass *et al.*, 1994; Wilson *et al.*, 1995).

To validate the conservation of this signaling network, we conducted IF for pSMAD1/5, WNT3a, NODAL, and the NODAL inhibitor LEFTY. The consistent localization of pSMAD1/5 along the colony periphery revealed that all genotypes maintain BMP4 signaling in this region (Fig. 2.1E, F). Additionally, NODAL was detected broadly and upregulated just interior toward the nascent mesoderm domain in gastruloids of all genotypes, and LEFTY was seen adjacent to this domain near the outer mesendoderm boundary. WNT3a protein expression was also maintained in the mesendoderm domain across all genotypes but was slightly decreased in this region of TBXT-KO gastruloids, suggesting the positive feedback loop between TBXT and canonical WNT signaling is disrupted in the TBXT-KO (Fig. 2.1E, F).

The uniformity in germ layer identity and distribution across genotypes suggests that cell fate specification is not dependent on TBXT dose during the early stages of human PS morphogenesis, and that cell fate specification and induction of morphogenetic movements are likely governed independently during *in vitro* gastrulation.

Defining a genomic regulatory network underlying TBXT dosage

To gain a more comprehensive understanding of similarities to the malformations observed in *T/Bra* mutants observed *in vivo*, we sought to precisely define disparities in cell identity and the expression of migration regulators, including drivers of EMT, within gastruloids of each genotype. TBXT directly binds to the regulatory regions of key genes in mesoderm development and can influence chromatin accessibility (Faial *et al.*, 2015; Koch *et al.*, 2017), so we further hypothesized that TBXT expression level may influence chromatin accessibility at genes that are required for mesoderm maturation, and that altered accessibility may precede changes in protein level at the initial stages of mesoderm development (48-hour BMP4 exposure).

To explore these possibilities, we conducted multiomic single nucleus RNA sequencing (snRNA-seq) and single nucleus Assay for Transposase-Accessible Chromatin (snATAC-seq) on gastruloids of each genotype after 48 hours of BMP4 treatment.

Our analysis of all 3 pooled genotypes yielded 11 clusters consisting of three extraembryonic cell populations (Clusters 1–3; “Extraembryonic-1–3”, extraembryonic progenitor cells (Cluster 4; “EEM Progenitors”), epiblast-like cells (Cluster 5; “Epiblast”), three primitive streak-like cell populations (Clusters 6–8; “PS-1–3“), nascent mesoderm-like cells (Cluster 9; “Mesoderm”), nascent endoderm-like cells (Cluster 10; “Endoderm”), and primordial germ cell-like cells (Cluster 11; “PGCLC”) (Figs. 2.2A-D, relationships between clusters at various resolutions illustrated in Figs. S2.3, S2.7). In agreement with immunofluorescence data, we found there was not a significant difference in the proportion of cells from each genotype assigned to each cluster, supporting the notion that cell identity during early gastrulation was not significantly affected by the loss of TBXT.

Looking at these cluster identities in more depth, we observed that clusters 1–3 share an extraembryonic gene expression signature reflecting broad markers of both trophectoderm (TE) and amnion (*CDX2, GATA3, TFAP2A, HAND1, WNT6, GATA2*) (Fig. 2.2C-D). To distinguish these possibilities, we evaluated these clusters for key markers of trophectoderm, late-amnion, and early-amnion lineages. This analysis revealed a bias in all three clusters toward late-amnion (*GAPBRP, HEY1, HAND1, VTCN1, TPM1, IGFBP3, ANKS1A*) relative to embryonic clusters (Fig. S2.4), in agreement with published findings that primed hiPSCs are biased toward a late-amnion fate in the presence of BMP4 (Rostovskaya *et al.*, 2022). This late-amnion gene expression signature was highest in cluster 1 and lowest in cluster 3, suggesting clusters 1–3 are likely distinguished by subtle variations in developmental timing with cluster 1 being a relatively more differentiated amnion and cluster 3 being nascent amnion. The companion snATAC-seq dataset revealed that clusters 1–3 have very similar chromatin accessibility and inferred peaks to both one another relative to all other clusters and to published H3K27AC accessibility data from

primary amnion tissue (Bernstein *et al.*, 2010), reinforcing their developmental similarity and amnion identity (Figs. 2.2E, S2.5D). These peaks are also enriched for GATA motifs, which are key regulators of extraembryonic cell fate (Fig. S2.5E). With these observations in mind, we designated “Extraembryonic-1” as “Extraembryonic-Late”, “Extraembryonic-2” as “Extraembryonic-Middle” and “Extraembryonic-3” as “Extraembryonic-Early”.

Following this developmental trajectory, cluster 4 shows a slight upregulation of extraembryonic markers (*GATA3, TFAP2A, HAND1, ISL1, TBX3*) relative to cluster 5, and cluster 5 exhibits canonical hallmarks of epiblast fate (*SOX2, POU5F1, NANOG, DPPA4*), including simultaneous motif enrichment of *SOX2*, *POU5F1*, and *NANOG* (Fig. 2.2C-D, S2.5E). Even with an apparent extraembryonic gene signature, cluster 4 maintains an epiblast-like gene signature overlapping with that of cluster 5. This gene expression pattern suggests that cluster 4 likely contains cells analogous to amnion progenitors leaving the PS, which we refer to as “Extraembryonic progenitors” (“EEM Progenitors”), while cluster 5 reflects epiblast-like cells (“Epiblast”).

The three PS clusters, clusters 6–8, share many elements of a PS gene signature, including *TBXT, MIXL1, and EOMES*, and are differentiated from one another by the progressive downregulation of epiblast markers such as *SOX2, TDGF1*, and *NODAL* (Fig. 22.C-D*)*. This pattern suggests that cluster 6 represents the least differentiated state between epiblast and mesoderm-like cells, followed by clusters 7 and then 8. Therefore, we designated “PS-1” (cluster 6) as “PS-Early”, “PS-2” (cluster 7) as “PS-Middle”, and “PS-3” (cluster 8) as “PS-Late.” Notably, clusters 4–8 share similar gene expression, chromatin accessibility, and peak distributions to one another, which reflects their recently shared epiblast-like origin and closely related identities (Fig. S2.5D). Analysis of differentially accessible regions (DARs) between clusters revealed a prospective SOX2 enhancer and the POU5F1 promoter that was uniquely accessible in the epiblast, extraembryonic progenitor, and early PS-like cell types (clusters 4–7) (Fig. 2.2E). These DARs correspond to H3K27ac ChIP-seq peaks in pluripotent HUES cells (Tsankov *et al.*, 2015),

reflective of these clusters' pluripotency-related expression pattern and subsequent downregulation as differentiation progresses away from pluripotency.

Clusters 9 and 10, the nascent mesoderm and nascent endoderm-like clusters, respectively, share a mesendoderm-like gene expression signature (*GSC, MIXL1, EOMES, LHX1, HAS2, GATA6,* and *PDGFRA*). They are distinguished by the increased expression of *MESP1, MESP2, and APLNR* in mesoderm-like cells and *SOX17* in endoderm-like cells (Fig. 2.2C-D). In accordance, these two clusters are enriched for several motifs regulating mesendoderm specification, including GATA factors and FOXA2 (Fig. S2.5E). Chromatin accessibility near the *TBXT* promoter is increased in the PS and mesendoderm-like subclusters (clusters 6–10) and, interestingly, there appears to be a potential *TBXT* regulatory domain uniquely accessible in extraembryonic cell types (Fig. 2.2E). This peak is not linked to *TBXT* gene expression by Peak2Gene analysis, but it does correlate with the H3K27ac ChIP-seq profile of amnion tissue (Bernstein *et al.*, 2010), suggesting that *TBXT* expression may be uniquely regulated in extraembryonic tissue relative to embryonic, a possibility that warrants further study. Finally, cluster 11 co-expresses *SOX17, PRDM1, NANOS3*, and *TFAP2C*, consistent with PGCLC identity (Fig. 2.2D).

TBXT influences downstream gene expression in a dose-dependent manner

To understand why the loss of *T/Bra in vivo* most dramatically affects the morphogenesis of the mesoderm population, we assessed the nuanced differences that exist between the mesoderm populations of each genotype and how these may contribute to abnormal migration patterns observed later in development. We first compared differentially expressed (DE) genes between TBXT-Het vs. WT or TBXT-KO vs. WT within the mesoderm cluster (Fig. 2.3A-A'). This analysis identified 14 genes downregulated in the TBXT-Het compared with WT and 44 genes downregulated in the TBXT-KO compared with WT (Figs. 2.3B-D, S2.6A-B) (adj. $p < 0.05$, Log2FC < -0.25). Two general classes of DE genes emerged: genes that required a relatively binary

threshold level of TBXT expression to fully activate or repress downstream expression and genes whose expression levels scaled with TBXT expression levels. Nine genes were statistically significantly downregulated in both the TBXT-Het and TBXT-KO compared to WT, putting them in the former category, including the non-canonical WNT pathway component *WNT5A* and BMP inhibitor *BMPER* (Fig. 2.3C-D, starred). In contrast, several genes including *WDPCP* and *TCF4* were significantly downregulated in the TBXT-KO compared to WT but had intermediate expression in the TBXT-Het (Fig. 2.3D, unstarred). GO Biological Pathway Enrichment for the 44 genes significantly downregulated in the TBXT-KO vs. WT, the majority of which also had reduced expression in the TBXT-Het, denoted “regulation of cell motility”, ”locomotion”, and “PCP pathway involved in axis elongation,” reflecting the attenuated migratory phenotype and increased adhesiveness seen when TBXT expression is reduced *in vivo* (Wilson and Beddington, 1997)(Wilson *et al.*, 1995) (Fig. 2.3E). As anticipated, *in situ* hybridization confirmed decreased expression of the mesoderm marker *MESP1* and the WNT regulators *WNT5A* and *RSPO3* in TBXT-KO gastruloids relative to WT (Fig. 2.3F).

On the other hand, we identified 54 genes upregulated in the TBXT-Het compared to WT and 45 genes upregulated in the TBXT-KO compared to WT (Fig. 2.3B-D, S2.6A) (adj. $P < 0.05$, Log2FC > 0.25). Of these, 15 genes overlapped, including the posterior morphogens *CYP26A1* and *FGF17* and the broad mesendoderm marker *PRDM1* (Fig. 2.3C-D, starred). We conducted *in situ* hybridization for *FGF17* and *CYP26A1* to validate these results and observed increased expression of both transcripts in TBXT-KO gastruloids relative to the WT (Fig. 2.3F).Genes that demonstrated expression patterns that scaled with TBXT dose included migration and adhesion ligand *EFNA5*, cell adhesion molecules *CADM1*, *CADM2,* and *CDH12*, and endoderm marker *LHX1* (Fig. 2.3D). GO enrichment of the 45 genes significantly upregulated in the TBXT-KO denoted “cell-cell adhesion” and “adherens junction organization,” reflective of a persistent epithelial character and impaired migration in the absence or reduction of TBXT (Fig. 2.3E). In addition, several terms reflective of neurogenesis emerged. These terms are almost exclusively

driven by genes involved in axon guidance or migration such as *NTN4*, *SEMA6A*, and *EFNA5*, rather than markers of neural identity such as *SOX2*, *OLIG3*, or *PAX6*, so while it is possible that these changes in gene expression reflect a neural bias we believe it more accurately represents the misregulation of genes involved in the migratory mesenchymal phenotype. Furthermore, it is also possible that this neural bias in GO terms reflects a bias of a neuromesodermal progenitor-like (NMP) population toward neural identities at the expense of mesodermal, a decision point which TBXT has been previously shown to directly modulate (Koch *et al.*, 2017). However, the NMP marker NKX1-2 is very minimally detected across all clusters, and no cluster shows robust co-expression of SOX2 and TBXT even at higher clustering resolutions (Fig. 2.2D, S2.7). This prevents us from conclusively defining the presence of an NMP subpopulation within the mesodermal population or in the dataset more broadly.

Overall, the level of expression of differentially expressed genes in the TBXT-Het was always similar to the TBXT-KO, similar to the WT, or an intermediate between the two. This pattern suggests that while some genes respond to TBXT dose in a binary manner where WT expression levels are needed for downstream expression, more often the expression level of TBXT is closely linked to the expression level of its downstream targets.

TBXT dose subtly influences the expression profile within the mesodermal population

While the presence and spatial distribution of broad mesoderm markers were not affected by varying TBXT expression, closer examination revealed that the gene regulatory network controlling early mesoderm identity itself is only subtly influenced by TBXT dose. The broad mesendoderm and PGCLC marker, *PRDM1*, was significantly upregulated in both the TBXT-Het and the TBXT-KO relative to WT, suggesting its expression is sensitive to TBXT dose (Fig. 2.3D, G, S2.6A). *LHX1* and *PDGFRA* had a similar expression pattern as *PRDM1*, with intermediate expression levels in the TBXT-Het and significant upregulation in the TBXT-KO compared to WT. To understand if this gene expression pattern reflected a larger bias toward endodermal cell

identity, we manually isolated several canonical markers of mesodermal and endodermal identity from our dataset and looked at their expression patterns. *SOX17* and *EOMES* followed a similar pattern to PRDM1, with slightly elevated but not significantly differential expression in the TBXT-KO relative to the WT. In contrast, mesoderm markers *MESP1*, *TBX6,* and *BMP4* were slightly higher in the WT compared to the KO, but again these changes in expression were not statistically significant (Fig. 2.3G, S2.6C). Canonical WNT pathway components *RSPO3*, *WDPCP*, and *TCF4* did display significantly graded expression, with the highest expression in WT and the lowest in the TBXT-KO (Fig. 2.3D, H, S2.6D). *WNT3A* itself was very sparsely detected in the nascent mesoderm population via snRNA-seq, however, the significant upregulation of several canonical WNT pathway components in WT agrees with the WNT3A expression pattern observed via IF (Fig. 2.1D-F, 2.3H). We then reanalyzed our dataset at a higher resolution (0.6) to understand if these subtle trends reflected larger biases toward endoderm fate that were being masked by our clustering resolution(Fig. S2.7). While this new resolution did segregate our mesoderm cluster into a mesendoderm-like population and a relatively more mature mesoderm-like population, there is not a clear genotype-specific bias in the distribution of cells assigned to each of these new clusters, reinforcing the consistency in lineage allocation regardless of TBXT dosage.

Notably, several direct targets of TBXT that are implicated in osteogenesis were significantly upregulated in WT, including *BMPER*, *GNAI2, ENPP1, THSD7A,* and *ADAMTS3* (Fig. S2.6C). This is relevant because *Tbxt*$^{+/-}$ mutant mice frequently display skeletal malformations later in development including rib fusions, osteochondrodysplasia, and brachydactyly (Grüneberg, 1958; Herrmann *et al.*, 1990; Wilson, Rashbass and Beddington, 1993). In addition, *MAML3*, which amplifies transcription of *HES1* to drive oscillations in somitogenesis, was increased in the TBXT-KO gastruloids. This expression change is potentially related to somitic fusions and subsequent rib fusions seen in *Tbxt*$^{+/-}$ and *Tbxt*$^{-/-}$ animal models later in development (Wu *et al.*, 2002; William *et al.*, 2007). Finally, the BMP inhibitor *BMPER* was significantly downregulated in both TBXT-Het and TBXT-KO compared to WT. Because BMP4 signaling activates *TBXT*

expression and TBXT directly activates *BMPER* expression, decreased *BMPER* expression in the TBXT-KO could reflect a compensatory pathway to rescue *TBXT* expression levels in TBXT-Het or TBXT-KO.

These gene expression trends suggest that while reduced TBXT expression does not preclude cells from forming mesoderm, it does influence the gene regulatory networks that may have a later-onset role in downstream axial elongation and patterning.

TBXT dose influences the expression of genes that modulate EMT

In addition to genes related to endoderm identity, we observed several gene expression patterns reflective of impaired EMT in our mutant gastruloids. For example, cell adhesion molecules *CADM1, CADM2*, *CDH12*, and *EFNA5* were significantly upregulated in the TBXT-KO mesoderm, intermediate in the TBXT-Het mesoderm, and downregulated in the WT mesoderm (Fig. 2.3I). *CDH1* shared this expression pattern albeit with reduced statistical significance, suggesting that the mutant cell lines retain an epithelial character, unlike their WT counterparts. Indeed, *SNAI1* and *SNAI2*, canonical regulators of EMT, both had the highest expression in the WT mesoderm, although they were detected in very few cells (Fig. 2.3I). The upregulation of adhesion molecules and downregulation of SNAI family proteins observed in TBXT-Het and TBXT-KO suggests that while the WT population is actively undergoing EMT, this process is impaired in the mutant gastruloids.

The majority of differentially expressed genes are direct targets of TBXT

To better isolate the key components of the gene regulatory network underlying TBXT dose-responsive gene expression, we sought to identify which differentially expressed genes are also likely direct targets of TBXT based on promoter binding proximity. To do this, we leveraged four existing TBXT chromatin immunoprecipitation with sequencing (ChIP-seq) datasets from human and mouse embryonic stem cells (hESC/mESCs) grown *in vitro*. The first two datasets

were derived from hESCs differentiated into monolayer TBXT+ cell populations using either activin (endoderm-biased) or BMP4 (mesoderm-biased) protocols (Faial *et al.*, 2015). The third dataset utilized a mesendoderm-biased hESC population driven by TGF-β and WNT signaling (Tsankov *et al.*, 2015). The final study used an Activin-A mediated protocol to drive mESC embryoid bodies to a Tbxt+ PS fate (Lolas *et al.*, 2014).

This comparison revealed that the vast majority of the genes that are differentially expressed between the different genotypes within the mesoderm cluster have TBXT binding sites adjacent to their promoters and are therefore likely direct targets of TBXT. Specifically, 33 of the 44 differentially expressed genes downregulated in TBXT-KO compared to WT had proximal TBXT binding sites in at least one ChIP-seq dataset (Fig. 2.3D (bold), 2.3J, S2.8). Six of these 33 genes were detected in both human and mouse ChIP-seq datasets (*BMPER, DTWD2, EXT1, MDFIC, ENPP1,* and *RSPO3*), suggesting they play an evolutionarily conserved function in early development. The remaining 26 genes were specifically identified in the human ChIP-seq datasets. One gene, *PCED1B*, was detected as a potential direct target of TBXT in the mouse datasets but not in any of the human datasets. Seven of the 33 differentially expressed genes identified in the ChIP-seq datasets were significantly downregulated in both the TBXT-Het and TBXT-KO colonies compared to WT, including *BMPER*, *WNT5A*, and *UNC5C,* suggesting these genes are particularly sensitive to reduced TBXT expression. The 11 differentially expressed genes that were not identified as likely direct targets of TBXT included 4 protein-coding genes (*LDHA, AFDN, NLGN4Y,* and *SLC9C1*) and 7 long non-coding RNAs.

Thirty-five out of the 45 genes upregulated in the TBXT-KO compared to WT were found to have proximal TBXT binding sites in at least one ChIP-seq dataset (log2FC > 0.25, adj. $p < 0.05$) (Fig. 2.3D (bold), 2.3J, S2.8). Of these, 9 were identified as potential direct targets of TBXT in both human and mouse datasets (*FGF17, LHX1, SAMD3, UBL3, CADM1, EFNA5, MAML3, SEMA6A, and TCF7L1).* The remaining 26 genes were identified specifically in the human ChIP-seq datasets. Nine of the 35 potential direct targets, including *FGF17, CYP26A1,* and *PRDM1,*

were significantly upregulated in both TBXT-Het and TBXT-KO colonies compared to WT, reflecting a more pronounced dosage sensitivity (Fig. S2.8 (bold)). Ten differentially expressed genes were not identified in any of the ChIP-seq datasets, of which 3 were detected in coding regions (*CST1*, *PKIB*, and *RPS21*), 1 was mitochondrial, and the remaining 6 were long non-coding RNAs.

Taken together, these results indicate that the majority of genes that display TBXT dosage sensitivity have proximal binding sites for TBXT, and therefore their transcription is likely directly modulated by TBXT.

TBXT does not significantly impact chromatin accessibility in nascent mesoderm

It has previously been shown that TBXT plays a role in the deposition of H3K27ac at target genes during hematopoietic and endothelial development to alter transcription (Beisaw *et al.*, 2018; Chen *et al.*, 2023) and that TBXT is essential for remodeling chromatin during NMP development (Koch *et al.*, 2017). We were interested in determining whether TBXT similarly modulates chromatin accessibility during early PS morphogenesis. However, in our gastruloid mesoderm population, very few genomic regions were differentially accessible between either the TBXT-Het and WT or TBXT-KO and WT conditions (26 and 6 genes, respectively; FDR < 0.1, log2FC > 0.5) (Fig. S2.9A-C). While it is true that TBXT-Het does reflect a higher number of differentially accessible regions (DARs) than TBXT-KO when compared to WT, the Log2FC values of these DARs are very close to the significance cutoff and several are microRNAs. The 6 DARs identified in the TBXT-KO included adhesion protein *MDGA2*, vitamin D metabolizing enzyme *CYP2R1*, pluripotency-related gene *EYS*, histone components *HIST1H4L*, *HIST1H1B*, and *HIST1H3I*, and several long non-coding RNAs. Likewise, very few peaks were differentially accessible between the two conditions (18 peaks between WT and TBXT-Het and 0 peaks between WT and TBXT-KO, respectively; FDR < 0.1, log2FC > 0.5), and no motifs were significantly enriched within these peaks, as was true for all clusters. We observed slight variation

in peak height at genes demonstrating differential expression such as *WNT5A, RSPO3*, and *LHX1*, although these variations did not reach statistical significance (Fig. S2.9D). These results suggest that at this early stage of mesoderm specification in gastruloids, TBXT expression does not contribute to mesodermal patterning by influencing chromatin accessibility.

CellChat reveals TBXT-driven regulation of cell-cell adhesions and ncWNT signaling across and within clusters

Morphogenesis frequently involves paracrine and juxtracrine signaling between different cell populations. For example, mesodermal cells in contact with the epiblast and visceral endoderm have distinct protrusions whereas mesoderm cells in contact with other mesoderm cells appear smoother, reflective of distinct cell responses to specific migratory guidance cues (Saykali *et al.*, 2019). Therefore, we questioned whether TBXT-dose-dependent changes in cell behavior are restricted to the mesoderm population or if they may be influenced by adjacent cell types. To begin to address this question, we turned to the software CellChat, which analyzes the expression of ligand-receptor pairs within and across clusters to predict patterns in cell-cell communication.

First, we investigated how TBXT expression influences broad patterns in pathway activation by assessing which pathways have the largest changes in signals between or within different cell types, termed 'information flow,' when comparing WT, TBXT-Het, and TBXT-KO snRNA-seq data across all 11 clusters. This analysis revealed several pathways that have varying levels of information flow, and we focused on the three in which the WT and TBXT-KO have distinctly different patterns from one another (Fig. 2.4A). The first two pathways, cell adhesions (CADM) and junctional adhesions (JAM), were predicted to have higher information flow in the TBXT-Het and TBXT-KO gastruloids relative to WT, while the third pathway, non-canonical WNT (ncWNT) signaling, was predicted to have higher information flow in WT and TBXT-Het gastruloids relative to TBXT-KO.

To understand how this expression pattern affects interactions across clusters, we looked at predicted ligand-receptor interactions of each of these three pathways within and between each cluster for all three genotypes. Analysis of CADM pathway and JAM pathway communication across and with clusters revealed these hits were primarily driven by varying expression levels of *CADM1* and *F11R*, respectively, although the JAM pathway was also influenced by *JAM3* expression (Fig. S2.10A). We then visualized ligand-receptor interactions between and within clusters using circle plots, where the presence and thickness of a line correlate with the degree of predicted communication within or between clusters. An overall increase in CADM signaling was evident in TBXT-Het and TBXT-KO gastruloids relative to WT, and this seemed to be largely due to increased communication within the PS-Late and mesoderm clusters (Fig. 2.4B). Similarly, in TBXT-Het and TBXT-KO gastruloids, JAM signaling appeared exaggerated in PS and mesoderm clusters whereas these signaling dynamics were largely lost in WT. These trends suggest that unlike the WT, the TBXT-Het and TBXT-KO colonies are maintaining their cell-cell adhesions and junctional adhesions, and these differences are uniquely apparent in cell types expected to undergo EMT such as PS-Late and Mesoderm.

Within the ncWNT pathway, the TBXT-Het and TBXT-KO gastruloids had reduced ligand-receptor interactions compared to WT both between clusters but also within clusters, with the degree of self-regulation changing most notably within the three PS clusters and the mesoderm cluster (Fig. 2.4B). This variation in ncWNT signaling across genotypes was driven by the expression patterns of *WNT5A*, *WNT5B*, and several *FZD* receptors (Fig. 2.4B, S2.10A-C). WNT5A is a direct target of TBXT and a crucial component of both the non-canonical WNT/Planar Cell Polarity (PCP) pathway, which controls many aspects of directed cell migration and convergent extension, and the canonical WNT pathway, which is critical for sustained mesoderm development (Yamaguchi *et al.*, 1999; Dunty *et al.*, 2008; Kikuchi *et al.*, 2012).

To better understand how CADM, JAM, and ncWNT pathways are influenced by TBXT dose, we next explored the extent to which different clusters operate as senders, receivers,

mediators, and influencers of ligands and receptors in these pathways. Mediators serve as gatekeepers to control cell communication between any two groups, while influencers are predicted to control information flow more generally (Jin *et al.*, 2021). For both the CADM and JAM pathways, the TBXT-Het and TBXT-KO PS-late and mesoderm clusters increased their ability to serve as senders, receivers, mediators, or influencers relative to WT (Fig. 2.4C). These trends likely reflected genotype-dependent patterns in *CADM1* and *F11R* expression within these specific clusters. In contrast, the PS-late, mesoderm, and PGCLC clusters largely lost their ability to operate as senders, mediators, and influencers of the ncWNT pathway, but they maintained their ability to serve as receivers (Fig. 2.4C). These patterns suggest that variability in ncWNT pathway information flow is likely more highly dependent on varying levels of *WNT5A* or *WNT5B* expression (senders) rather than *FZD* expression (receivers).

To clarify how specific ligands or receivers modulated predicted information flow in the ncWNT pathway, we looked at how these gastruloid communication patterns were affected by the expression of specific ligand-receptor pairs sent or received by the mesoderm cluster (Fig. 2.4D). We found that WNT5A signals sent from the mesoderm to FZD receptors across all other clusters, including the mesoderm itself, were largely lost in the TBXT-KO but maintained in WT. WNT5B, however, was not detected as a major ligand sent from the mesoderm. Reciprocally, ncWNT signals such as WNT5A and WNT5B sent from extraembryonic-early, PS-late, and PGCLC clusters to the mesoderm cluster showed increased signal in the WT, likely reflecting an increase in *WNT5A* expression in all TBXT-expressing clusters. This analysis also revealed that WNT5B is most impactful in extraembryonic and PGCLC clusters, but relatively negligible in PS and mesendodermal clusters.

Taken together, our analyses directly reflect the persistence of cell adhesions and junctional adhesions in the TBXT-Het and TBXT-KO PS-late and mesoderm clusters, suggesting that these cells do not readily acquire a mesenchymal identity, a key requirement for motility and subsequent PS morphogenesis. The results additionally reflect the downregulation of *WNT5A* in

the TBXT-Het and TBXT-KO compared to WT and identify the non-canonical WNT signaling pathway as a key TBXT-dependent regulator of nascent mesoderm development.

TBXT dose influences the persistence of cell-cell adhesions

To understand how cell-cell adhesions and EMT are affected by TBXT dose, we conducted IF for several markers of EMT progression including SNAI1, the tight junction regulator ZO1, and basement membrane protein FN1 in 48hr 2D gastruloid colonies (Fig. S2.11). However, gastruloid colonies develop a very dense mesoderm layer that grows into the z-axis, and we found that the colony density and multilayered growth, in conjunction with the presence of multiple cell types, prohibited the clear visualization and quantification of cell morphology and junctions. In addition, the timepoint recapitulated by the gastruloids is likely too early to visualize the full process of EMT, and the 48-hour limitation of colony growth limits our ability to monitor EMT progression over time. Therefore, we turned to an alternate protocol that induces early mesoderm in a monolayer using StemCell Technology's Mesoderm Induction Media (MIM). In WT cells, this media induces TBXT expression in >90% of cells within 48 hours, thereby recapitulating the initial stages of nascent mesoderm specification (Fig. 2.5A-B, 2.12A). This early mesoderm population maintains consistent EOMES expression regardless of TBXT genotype with expression peaking at 48hrs, reinforcing the shared nascent mesoderm identity across genotypes (Fig. S2.12B). At 48 hours we observe a TBX6+ mesoderm population in WT and TBXT-Het and a reduced but present TBX6+ population in TBXT-KO. This is in agreement with our snRNA-seq data and the established role of TBX6 as a direct target of TBXT in the PS and later in the paraxial mesoderm. (Fig. 2.3G, S2.12C). We differentiated cells of each genotype using MIM for either 0hr, 24hr, 48hr, or 72hr, and assessed cell morphology and protein expression via IF.

Varying the timing of MIM exposure revealed a robust and stepwise delay in the timing of the downregulation of CDH1, a canonical marker of the epithelial state, across genotypes (Fig. 2.5C-D). At 24 hours of MIM treatment, all genotypes displayed robust junctional CDH1

expression. However, by 48 hours, CDH1 was largely eliminated in WT cells, TBXT-Het cells had variable expression, and TBXT-KO cells maintained robust junctional expression, in agreement with our snRNA-seq data. By 72 hours, TBXT-Het cells had lost CDH1 expression to mirror WT cells, whereas CDH1 was reduced but still detectable in TBXT-KO cells. This expression pattern demonstrates that TBXT dose is directly correlated to the temporal downregulation of CDH1 in the nascent mesoderm, which in turn correlates with stunted EMT progression. We also looked at CDH2 expression, as CDH1 downregulation is generally associated with CDH2 upregulation as EMT progresses. However, at 48 hours CDH2 appeared to be consistently expressed regardless of CDH1 expression level or genotype (Fig. S2.13B).

To better understand how TBXT influences EMT progression more broadly, we looked at the distribution of SNAI1, an established marker of EMT and inhibitor of CDH1 (Muqbil *et al.*, 2014; Xu *et al.*, 2019), across the same 72hr timespan. As previously indicated, we detected higher *SNAI1* RNA expression in WT compared to TBXT-Het and TBXT-KO via snRNA-seq (Fig. 2.3I). IF revealed that WT and TBXT-Het had a significantly higher level of nuclear SNAI1 relative to TBXT-KO at the 48hr timepoint (Fig. 2.5E-F). At 72hr this expression was beginning to drop off in the WT but was maintained in the TBXT-Het and TBXT-KO, revealing an inversed correlation to CDH1 expression. Furthermore, we also observed increased protein expression of other epithelial markers including junctional β-catenin and ZO1 in the TBXT-KO cells relative to the WT after 48hr MIM, in accordance with TBXT dose-dependently augmenting the progression of EMT (Fig. S2.13A).

EMT progression and mesenchymal cell motility are linked to alterations in extracellular matrix composition. In particular, the deposition of the basement membrane protein and SNAI1-target fibronectin (FN1) reflects the acquisition of a mesenchymal phenotype. We observed decreased deposition of basement membrane component FN1 in our TBXT-KO relative to TBXT-Het and WT, reflecting maintained cell adhesions and decreased motility in the absence of TBXT (Fig. S2.13A).

Next, we utilized the scratch wound assay to visualize the migration patterns of cells of each genotype and quantify how TBXT-dose-dependent changes in EMT-related protein expression correlate with cell migration kinetics. Comparing the difference in the percent of occupied tissue area between the starting time point (24hr MIM) and ending time point (48hr MIM) across genotypes revealed that the WT population migrated significantly farther into the wound space than the TBXT-Het or TBXT-KO populations (Fig. 2.5G-H). WT cells were noticeably larger and much more motile both as a group and independently. Staining with F-Actin revealed long protrusions in WT cells, reflective of their motile mesenchymal character. TBXT-KO cells, in contrast, remained tightly packed and epithelial, and cell movement appeared to be driven by overconfluence more so than directed cell movement. A fraction of TBXT-Het cells acquired a similar mesenchymal phenotype to WT, but others appeared epithelial and non-migratory. Therefore, increased SNAI1 and decreased CDH1 expression seen in WT cells at the 48-hour MIM timepoint reflect the acquisition of a TBXT-dependent migratory mesenchymal phenotype.

Through these studies, we observe that TBXT-Het is able to increase SNAI1, FN1, and β-catenin expression in tandem with WT. However, at the 48-hour timepoint, there are clusters of TBXT-Het cells that maintain their junctional protein expression, including CDH1 and ZO1, in a way that more closely mirrors the TBXT-KO phenotype. We hypothesize that both the upregulation of SNAI1 and the downregulation of CDH1 are critical for migration. and that specifically the downregulation of CDH1 is required for delamination and increased cell motility. Therefore, perduring junctional CDH1 likely explains why we observe a less migratory phenotype in TBXT-Het.

The expression patterns of CDH1, β-catenin, ZO1, SNAI1, and FN1 in addition to the distinct migratory dynamics seen across genotypes indicate that TBXT plays a crucial role in regulating the temporal component of EMT, where TBXT dose-dependently promotes the reduction of CDH1, nuclear localization of SNAI, and subsequent acquisition of a mesenchymal phenotype.

DISCUSSION

Our study demonstrates that the acquisition of mesodermal identity can be decoupled from the acquisition of a mesenchymal character, and TBXT is required for the latter in a dose-dependent manner. After 48 hours of BMP4 exposure, the mesoderm cluster of our gastruloids exhibit TBXT-dose dependent changes in gene expression related to mesendoderm identity, both non-canonical and canonical WNT signaling, and EMT, and the majority of these differentially expressed genes are likely directly regulated by TBXT as evidenced by comparisons to existing ChIP-seq datasets. CellChat analysis highlighted the JAM, CADM, and ncWNT pathways as uniquely regulated between WT and TBXT-KO colonies and showed that the mesoderm and PS-late clusters play key roles in the regulation of these pathways. We then utilized a 2D monolayer culture system to demonstrate that TBXT dose directly correlates with the temporal downregulation of genes related to cell adhesions, the upregulation of drivers of EMT, and the subsequent acquisition of a mesenchymal phenotype.

These findings have interesting implications for our understanding of how TBXT influences embryonic patterning at the earliest stages of gastrulation, even before the commencement of posterior trunk development. As opposed to NMP populations which require TBXT for mesoderm differentiation (Gouti *et al.*, 2017; Koch *et al.*, 2017), we demonstrated that early PS populations do not require TBXT for the initial establishment of a relatively anterior mesoderm identity. This result is in agreement with studies showing that another T-box factor, EOMES, primarily drives mesoderm differentiation at this stage of development (Schüle *et al.*, 2023). However, we also discovered that TBXT is not dispensable at this early stage, as its expression directly influences the timing of EMT and therefore the ability of the nascent mesoderm cells to properly acquire a migratory mesenchymal character. We believe that this impaired EMT observed *in vitro* is reflected in the “pile-up” phenotype observed *in vivo,* as cells accumulate in the PS when *T/Bra* expression is lost.

The timepoint reflected in our 2D gastruloid culture reflects the earliest stages of mesoderm commitment. However, the question remains how TBXT dosage influences the differentiation of mesodermal subtypes after EMT has been initiated. We observe that the early mesoderm-like cells emerge before EMT regardless of TBXT dosage, however this does not negate the very likely possibility that both TBXT expression and changes in gene regulation driven by the mechanical forces of EMT itself are required for more mature mesodermal subtypes to correctly differentiate as development progresses. In our late stage (72hr) MIM experiments, we observe a reduction in TBX6 expression in TBXT-KO conditions. TBX6 is a direct target of TBXT in the primitive streak and it is also a canonical marker of the paraxial mesoderm, and so this reduction possibly indicates a reduction or delay in the differentiation of paraxial mesoderm after EMT commencement. Additionally, TBXT is known to augment cell fate in populations not clearly defined in our 2D gastruloid or MIM systems, including the notochord and NMPs (Zhu, Kwan and Mackem, 2016; Koch *et al.*, 2017). Additional experiments, possibly using *in vivo* models, will likely be required to further disentangle how these later mesodermal cell identities are influenced by the reduction of TBXT expression and the corresponding delayed onset of EMT.

In our study, we observe that several of the differentially expressed genes identified between WT and TBXT-KO have been previously identified as probable direct targets of TBXT through ChiP-seq in both human and mouse datasets. Based on their evolutionary preservation, the genes that we identified as shared between this study and ChIP datasets from both species are likely uniquely important for gastrulation and subsequent survival of the embryo. Genes identified as specific to human or mouse models may reflect species-specific adaptations to TBXT's function, or, alternately, they may reflect the unique binding capabilities of TBXT in different tissue types. To this end, the mouse and human datasets generated mesoderm and/or endoderm using different differentiation protocols, and tissue-specific gene expression patterns for TBXT have been previously documented in (Faial *et al.*, 2015). Further experiments, such as those conducted by (Xia *et al.*, 2024) looking at the role of TBXT in human and ape axial

patterning, will be useful for deciphering how these species-specific differences influence gastrulation and subsequent developmental patterning.

Somewhat surprisingly, we did not observe large-scale changes in chromatin accessibility across TBXT genotypes, despite detecting variation in gene expression. While this may be specific to our *in vitro* model system, it likely better reflects the relatively early time point modeled by the gastruloids. At 48 hours mesoderm is just beginning to emerge, and it is possible that with continued differentiation over an additional 24-48 hours, the mesoderm identity would advance, and chromatin remodeling would be more readily apparent. It is also possible that TBXT does not directly remodel chromatin during early gastrulation and might instead be interacting with histone-modifying complexes.

Throughout this study we focused on cell populations analogous to those leaving the PS, however, extraembryonic tissues were also affected by the loss of TBXT. Like in the PS and mesoderm, CellChat specifically identified alterations in signaling within the JAM, CADM, and ncWNT pathways in extraembryonic populations. TBXT-KO embryos die from asphyxiation due to impaired allantois development, and so it will be interesting to explore the extent to which this is driven by analogous defects in EMT that limit extraembryonic cell migration.

Overall, this study clarifies the role of TBXT during early PS development and sheds light on its ability to promote the temporal progression of EMT in a dose-dependent manner. While TBXT-KO cells do ultimately downregulate CDH1 and acquire a motile phenotype, this transition occurs considerably later than in WT cells, demonstrating that the correct timing of EMT is critical for proper morphogenesis. In addition, this study decouples EMT progression from initial mesoderm specification in the PS and complements *in vivo* studies to both improve our core understanding of vertebrate mesoderm development and identify nuances of human development. This understanding will help us to effectively design directed differentiation strategies that incorporate both EMT and cell fate acquisition, in addition to improving our foundational understanding of human embryogenesis.

MAIN FIGURES

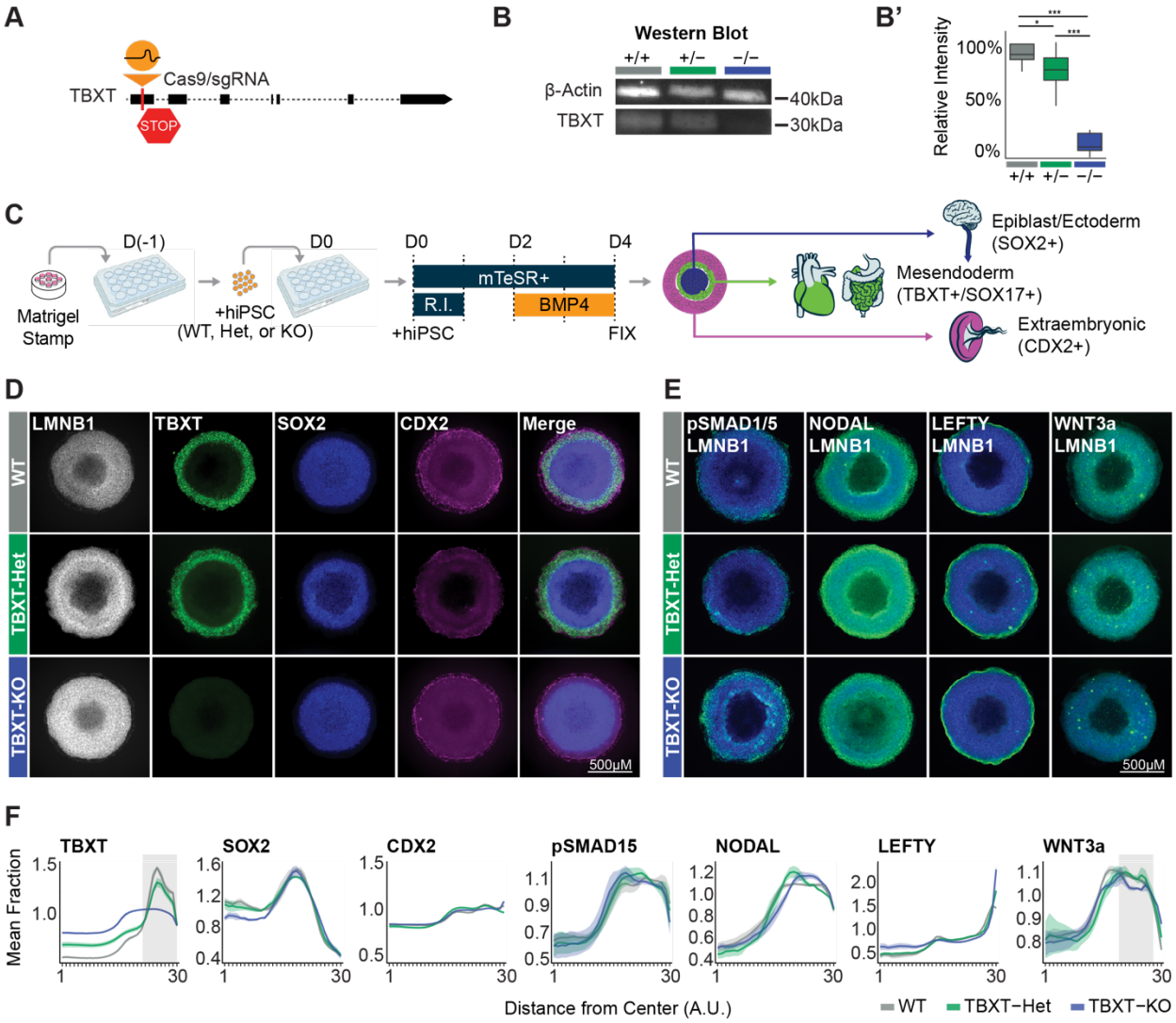


Figure 2.1: TBXT dose does not dramatically impact lineage emergence in 2D gastruloids (**A**) Schematic of the location within the TBXT locus targeted by sgRNA to generate TBXT-Het and TBXT-KO. (**B**) Western blot and (**B'**) quantification of fluorescent signal showing the knock-out efficiency of TBXT sgRNA across the allelic series (*p<0.5, **p<0.1, ***p<0.01 by unpaired t-test, n = 5). (**C**) Schematic depicting the 2D gastruloid differentiation strategy. (**D-E**) Immunofluorescent images of gastruloids for canonical gastrulation markers. "R.I." indicates ROCK inhibitor. (**E**) Green staining indicates protein of interest, while blue staining indicates LMNB1 (**F**) Quantification of fluorescent intensity of canonical gastrulation markers (n = TBXT: 5 WT, 15 TBXT-Het, 10 TBXT-KO; SOX2: 4 WT, 3 TBXT-Het, 3 TBXT-KO; CDX2: 5 WT, 15 TBXT-Het, 10 TBXT-KO; Lefty: 6 WT, 7 TBXT-Het, 8 TBXT-KO; Nodal: 8 WT, 3 TBXT-Het, 6 TBXT-KO; pSMAD1/5: 8 WT, 7 TBXT-Het, 4 TBXT-KO; WNT3A: 8 WT, 7 TBXT-Het, 4 TBXT-KO). Shadows around data indicate SEM. Gray vertical highlights indicate regions of interest based on variability in expression.

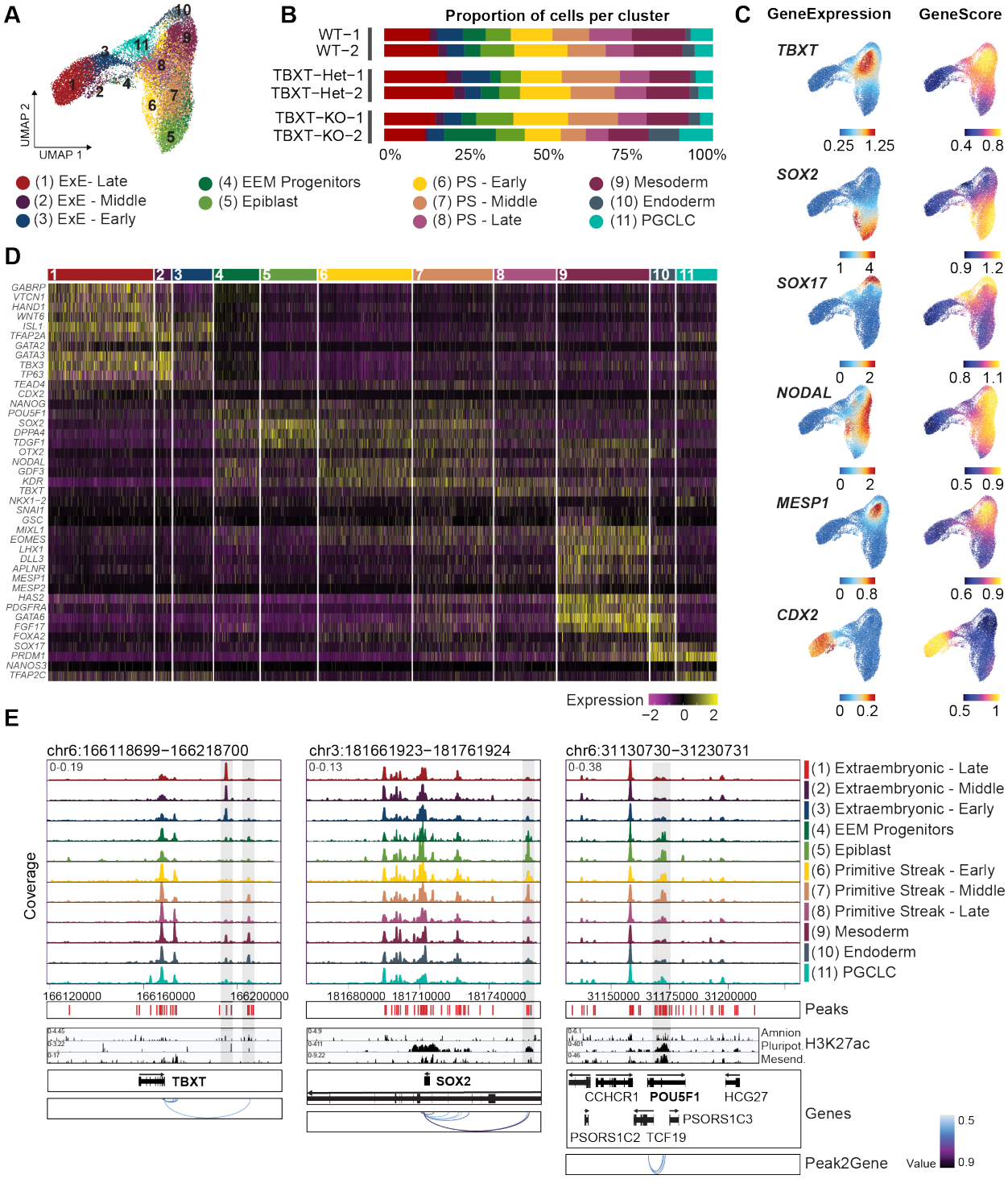


Figure 2.2: snRNA-seq reveals conserved lineage emergence in the absence of TBXT (**A**) UMAP derived from snRNA-seq and snATAC-seq data depicting 11 clusters identified through Seurat analysis of snRNA-seq data after 48hr BMP exposure (nCells = 23,838, nReplicates = 2). (**B**) Bar Plot depicting the proportion of cells from each sample assigned to each cluster. (Figure caption continued on the next page.)

(Figure caption continued from the previous page.)
(**C**) UMAP feature plots depicting GeneExpressionMatrix plots (derived from snRNA-seq) or GeneScoreMatrix plots for chromatin accessibility loci (derived from snATAC-seq) for key markers of gastrulation lineages. (**D**) Heatmap of snRNA-seq expression patterns for key markers of gastrulation lineages across clusters. (**E**) snATAC-seq accessibility tracks across clusters centered around TBXT, SOX2, and POU5F1 and aligned to detected peaks, H3K27ac tracks from published datasets, and Peak2Gene linkage predictions of regulatory connections between distal accessible regions (peaks) and nearby genes. Grey bars indicate peaks of interest.

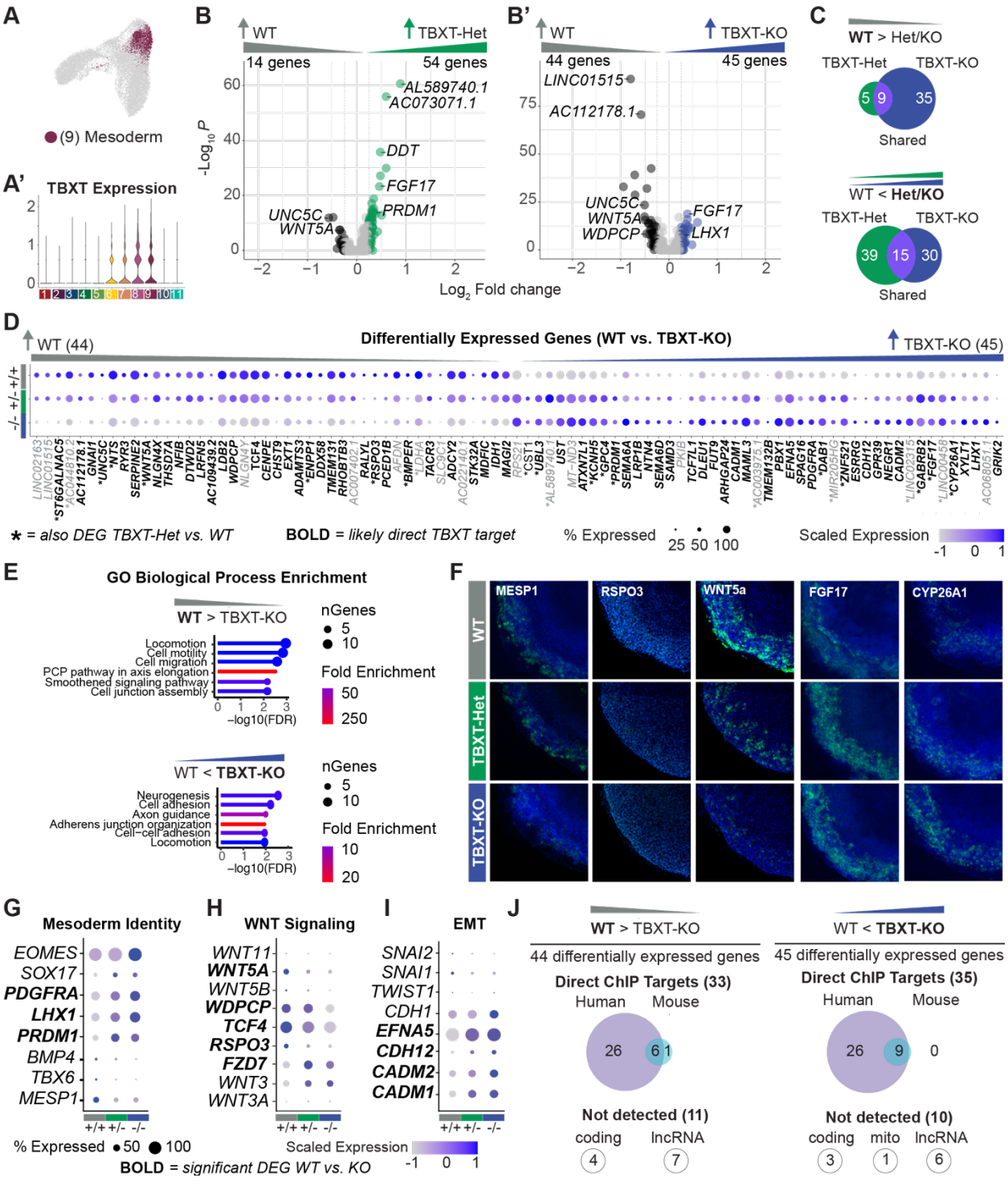


Figure 2.3: TBXT dose influences downstream gene expression in a dose-dependent manner (**A**) UMAP highlighting the mesoderm cluster (Cluster "9") and (**A'**) Violin plot showing the relative expression levels of TBXT in each cluster.
(Figure caption continued on the next page.)

(Figure caption continued from the previous page.)
(**B**) Volcano plot of differentially expressed genes between WT and TBXT-Het or (**B'**) WT and TBXT-KO after subsetting the mesoderm cluster (n = 3,212 cells, adj. p < 0.05, log2FC > 0.25) (**C**) Venn diagram of the number of DE genes identified in the TBXT-Het (green), TBXT-KO (Blue), or both TBXT-Het and TBXT-KO (purple) compared to WT. The top diagram includes genes downregulated in TBXT-Het and/or TBXT-KO compared to WT, while the bottom diagram includes genes upregulated in TBXT-Het and/or TBXT-KO compared to WT. (**D**) Dotplot of 89 DE genes between WT and TBXT-KO in order of log2FC. '*' indicates genes that are also differentially expressed between WT to TBXT-Het, while the bolded black text indicates genes identified as likely direct targets of TBXT by comparison with ChIP-seq datasets. (**E**) Key results from ShinyGO GO Biological Process Analysis (FDR < 0.05, Pathway Size 2-2000 genes) (**F**) Multiplexed fluorescence *in situ* hybridization for *MESP1*, *RSPO3, WNT5A, FGF17*, and *CYP26A1* in gastruloids at 48hr. *In situ* hybridizations were repeated for >=3 gastruloids of each genotype. (**G-I**) Dotplots showing gene expression trends for key genes related to mesoderm identity. Bold genes are significantly differentially expressed (adj. p < 0.05, log2FC > 0.25). (**G**), WNT signaling (**H**), and EMT (**I**). (**J**) Venn diagrams summarizing DE genes detected in corresponding ChIP-seq datasets.

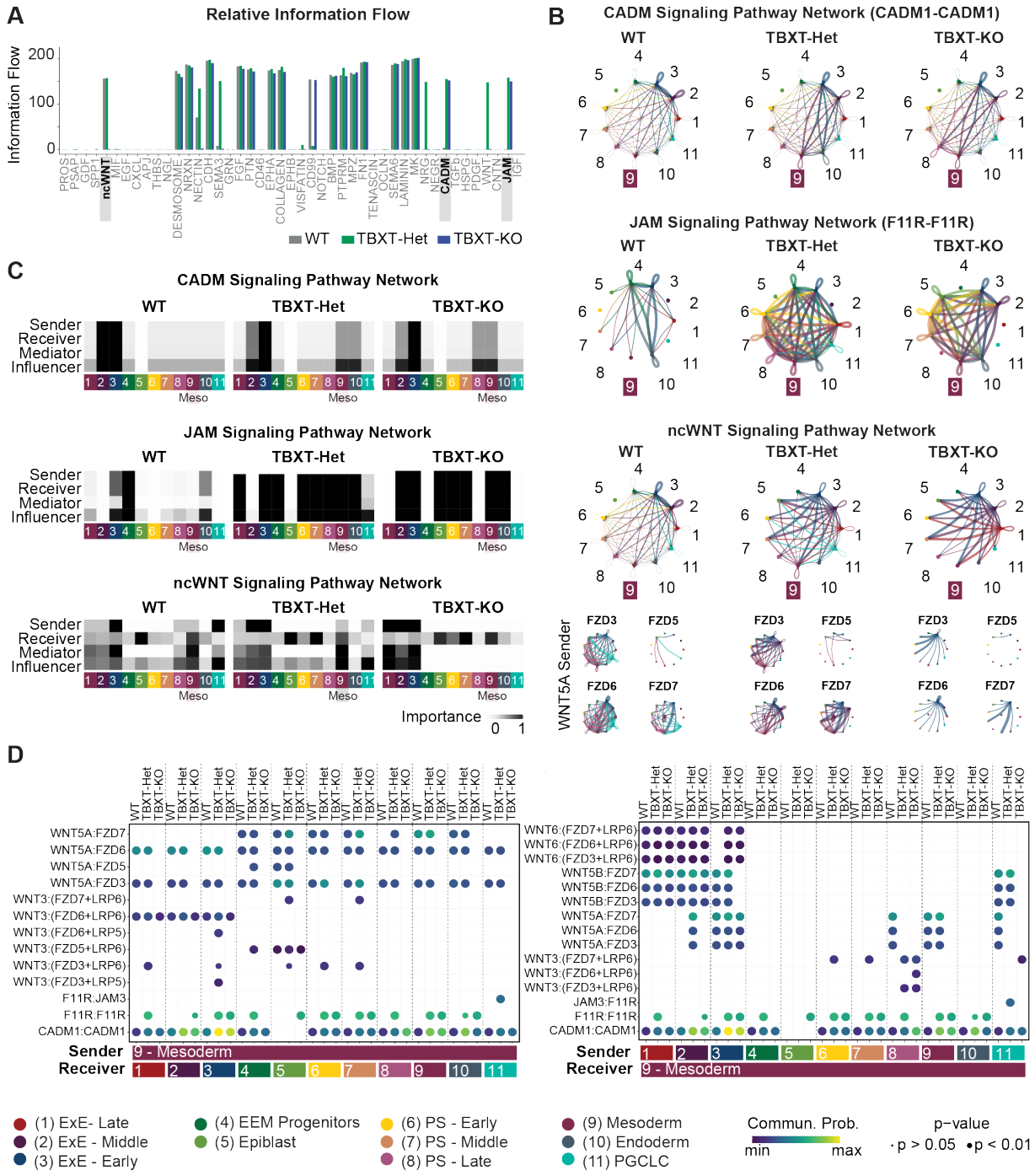


Figure 2.4: CellChat reveals TBXT-driven regulation of cell-cell adhesions and ncWNT signaling across and within clusters (**A**) Bar plot comparing information flow (A. U.) of key pathways between WT, TBXT-Het, and TBXT-KO. (**B**) Circle plots visualizing communication between or within clusters for the CADM, JAM, and ncWNT signaling pathways. Numbers correlate with cluster identity, while line thickness corresponds to the strength of predicted communication. “9” indicates the mesoderm cluster. Small circle plots under the “ncWNT Signaling Pathway” header depict interactions between WNT5A and designated FZD receptors. (Figure caption continued on the next page.)

(Figure caption continued from the previous page.)
(**C**) Heatmap showing the communication dynamics for each genotype for the CADM, JAM, and ncWNT signaling pathways. (**D**) Bubble plot comparing predicted ligand-receptor interactions for CADM, JAM, WNT, and ncWNT pathways. Only ligand-receptor interactions with variability in communication probability are shown.

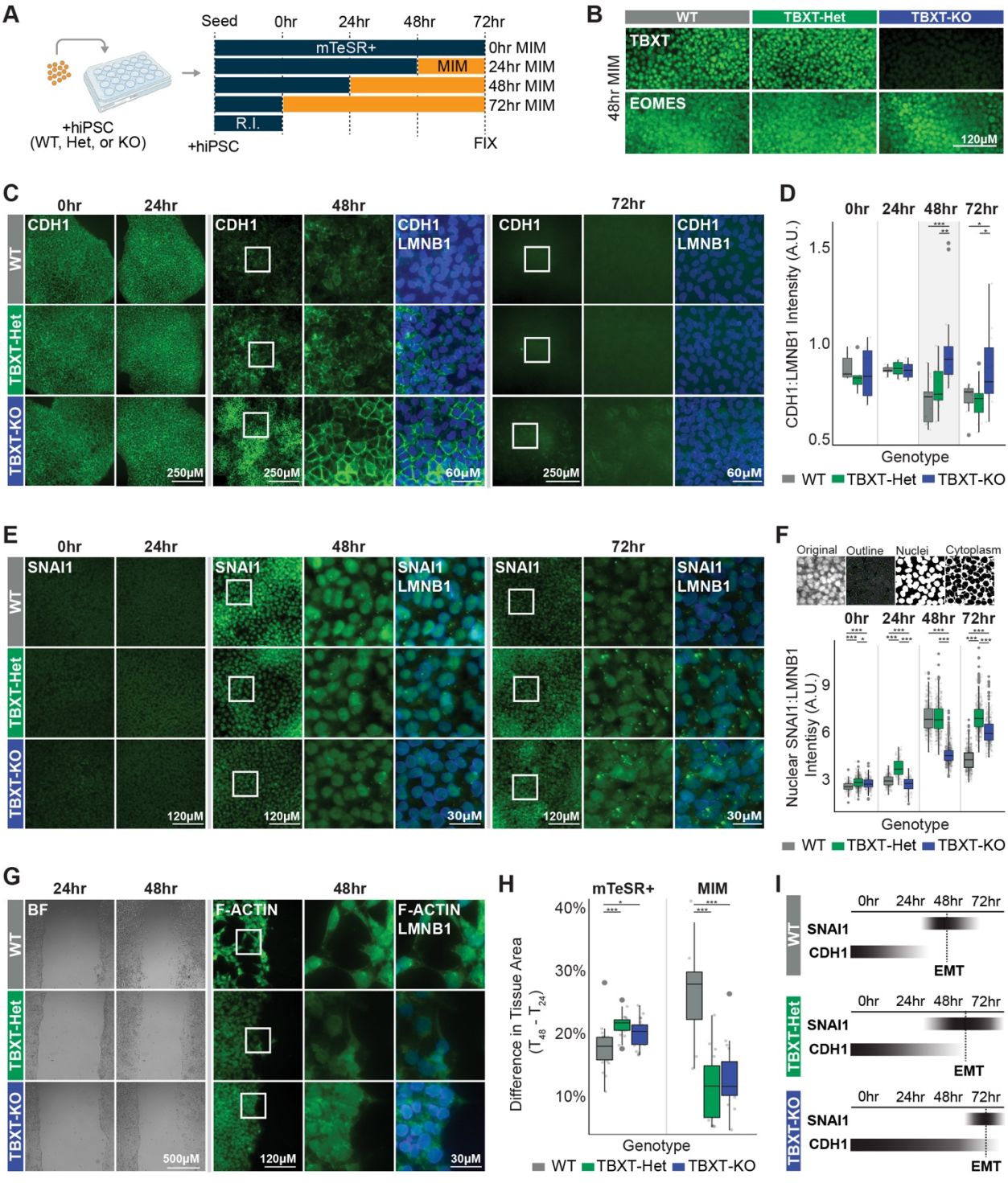


Figure 2.5: TBXT dose directly impacts the timing of EMT to permit migration (**A**) Schematic of the experimental setup. MIM = Mesoderm Induction Media. R.I. = ROCK inhibitor. (**B**) Immunofluorescence for TBXT or EOMES expression in each genotype after 48hr MIM exposure. (**C**) Immunofluorescence for CDH1 or (**E**) SNAI at 0hr, 24hr, 48, or 72hr MIM exposure. (**D**) Quantification of the ratio between CDH1 and LMNB1 intensity. n >= 6 wells/genotype. (Figure caption continued on the next page.)

(Figure caption continued from the previous page.)
(**F**) Schematic showing segmentation of nuclei and cytoplasm and the quantification of the ratio between SNAI and LMNB1 specifically within cell nuclei. WT = 1,228 cells; TBXT-Het = 1,423 cells TBXT-KO = 1,373 cells. n = 3 images/genotype/timepoint. (**G**) Brightfield images of the scratch assay at 24hrs or 48hrs MIM exposure (0hr or 24hr relative to scratch). Immunofluorescence for F-ACTIN at the edge of a scratch wound at 48hrs (24hr relative to scratch). (**H**) Quantification of the difference in tissue area between 48hr and 24hr MIM exposure (n = mTeSR+: WT= 23, TBXT-Het = 19, TBXT-KO = 21. MIM: WT = 8, TBXT-Het = 15, TBXT-KO = 16). **(I)** Overview of temporal onset of EMT in each genotype.

SUPPLEMENTARY FIGURES

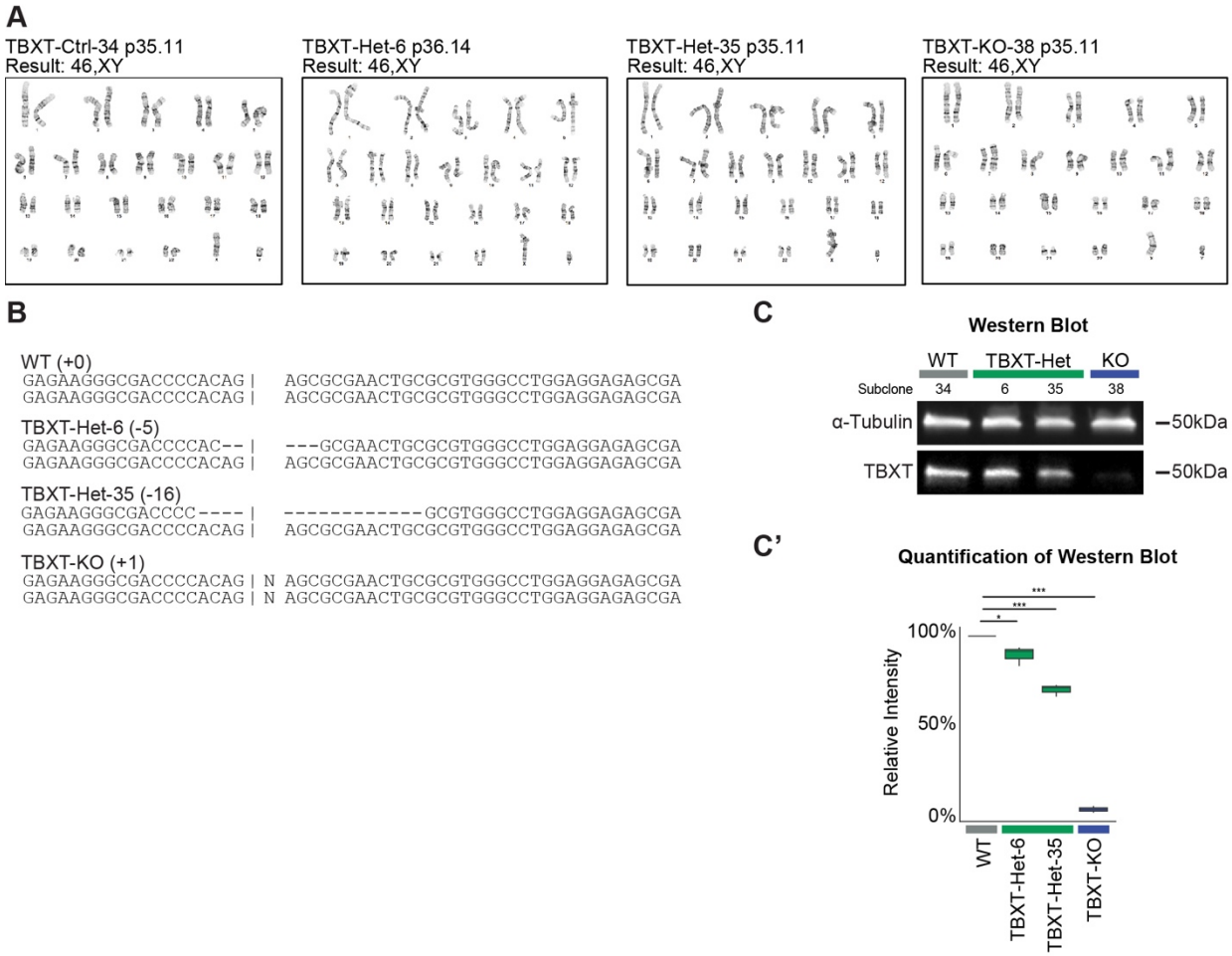


Figure S21: Quality Control and Karyotyping for TBXT allelic series. (**A**) Karyotyping results from WT (TBXT-Ctrl-34), TBXT-Het (TBXT-Het-35 and TBXT-Het-6), and TBXT-KO (TBXT-KO-38) cell lines. (**B**) Sequences of WT, TBXT-Het, and TBXT-KO cells illustrating corresponding indels. (**C**) Western blot and (**C'**) quantification of CDX2 in western blot in WT, TBXT-Het-6, TBXT-Het-35, and TBXT-KO cells after being differentiated in MIM for 48 hours in a monolayer. n = 3 replicates/genotype.

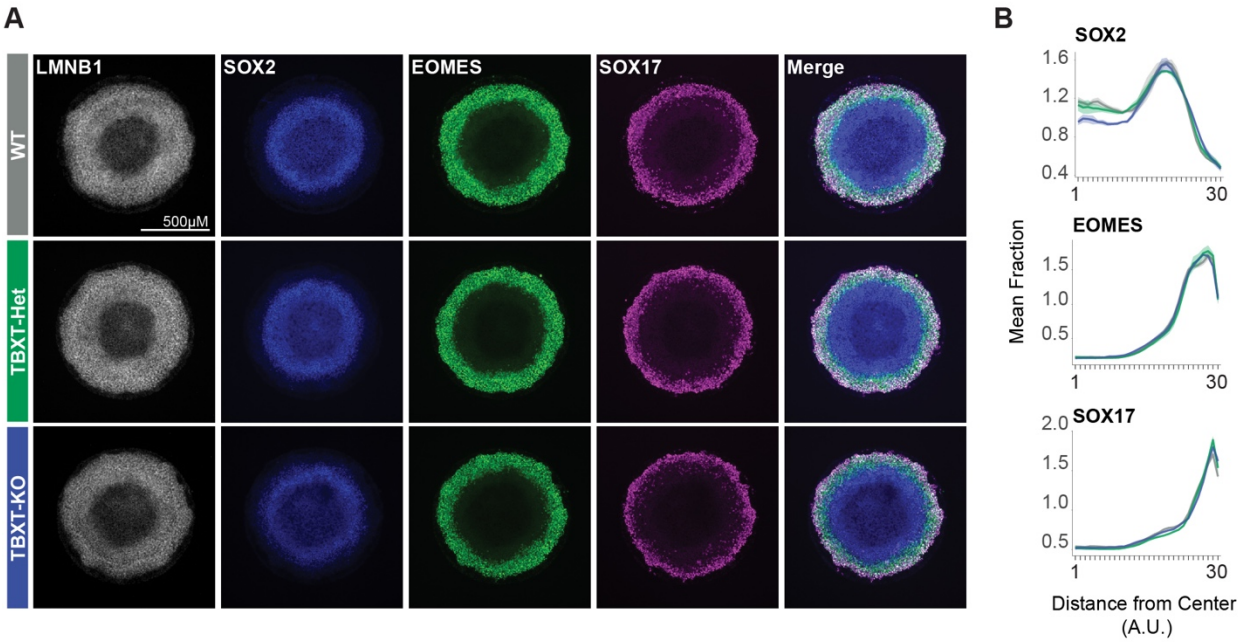


Figure S2.2: Immunofluorescent images of 2D gastruloids for canonical gastrulation markers. (**A**) Immunofluorescent images of 2D gastruloids for canonical gastrulation markers and (**B**) quantification of fluorescent intensity (n = 4 WT, 3 TBXT-Het, 3 TBXT-KO). Shadows around quantification data points indicate SEM.

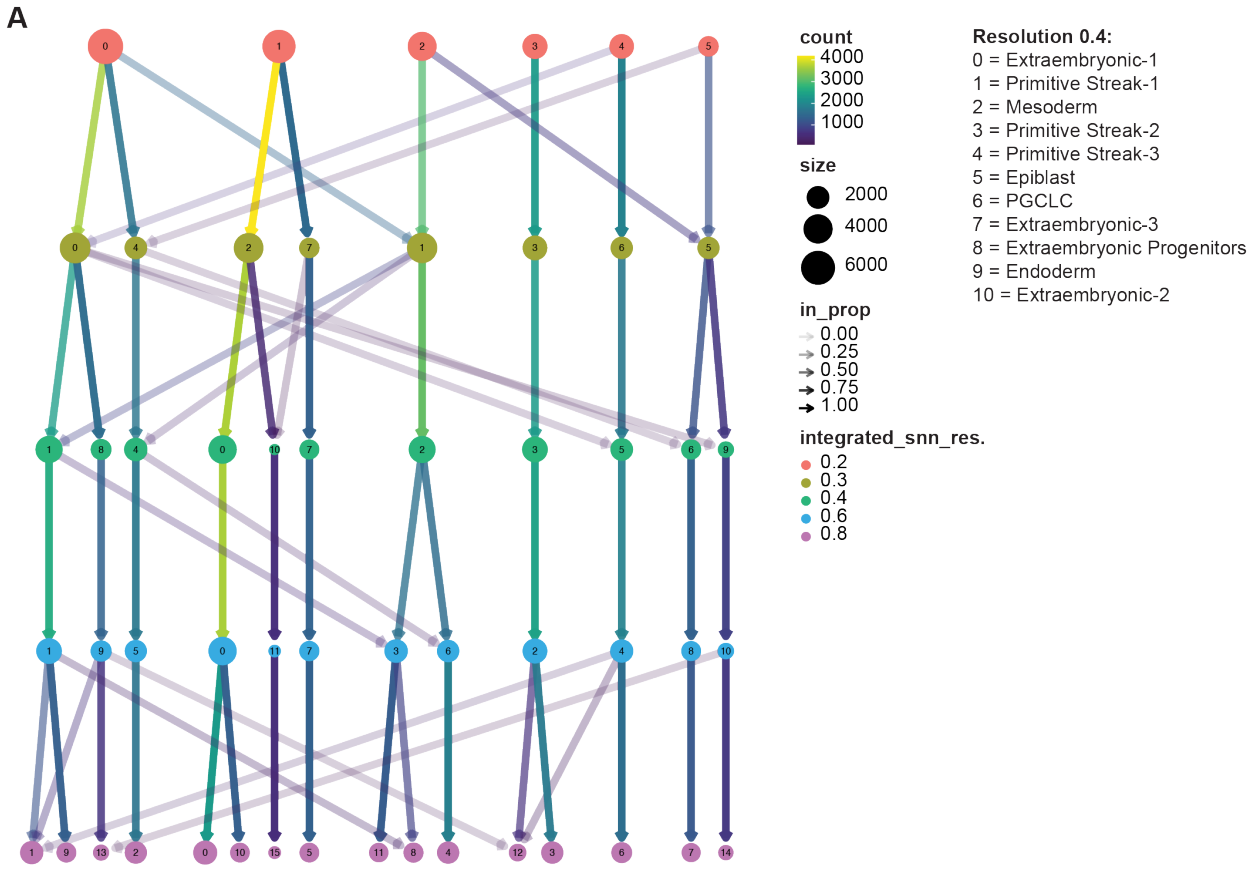


Figure S2.3: ClusTree analysis of clusters determined at various resolutions. (**A**) ClusTree analysis of clusters determined at 0.2, 0.4, 0.6, and 0.8 resolution.

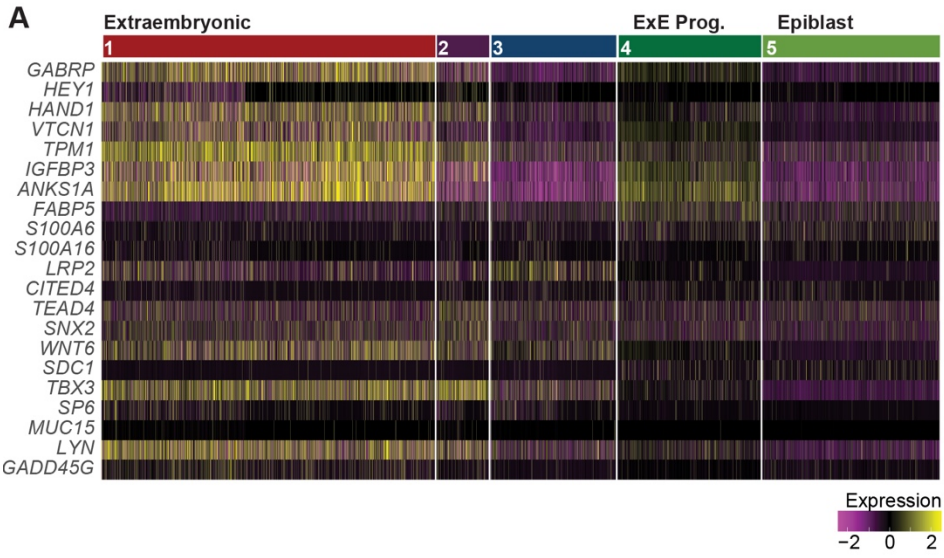


Figure S2.4: Extraembryonic mesoderm and trophectoderm marker expression across clusters. (**A**) Heatmap of extraembryonic/amnion markers for snRNA-seq data for clusters 1-5.

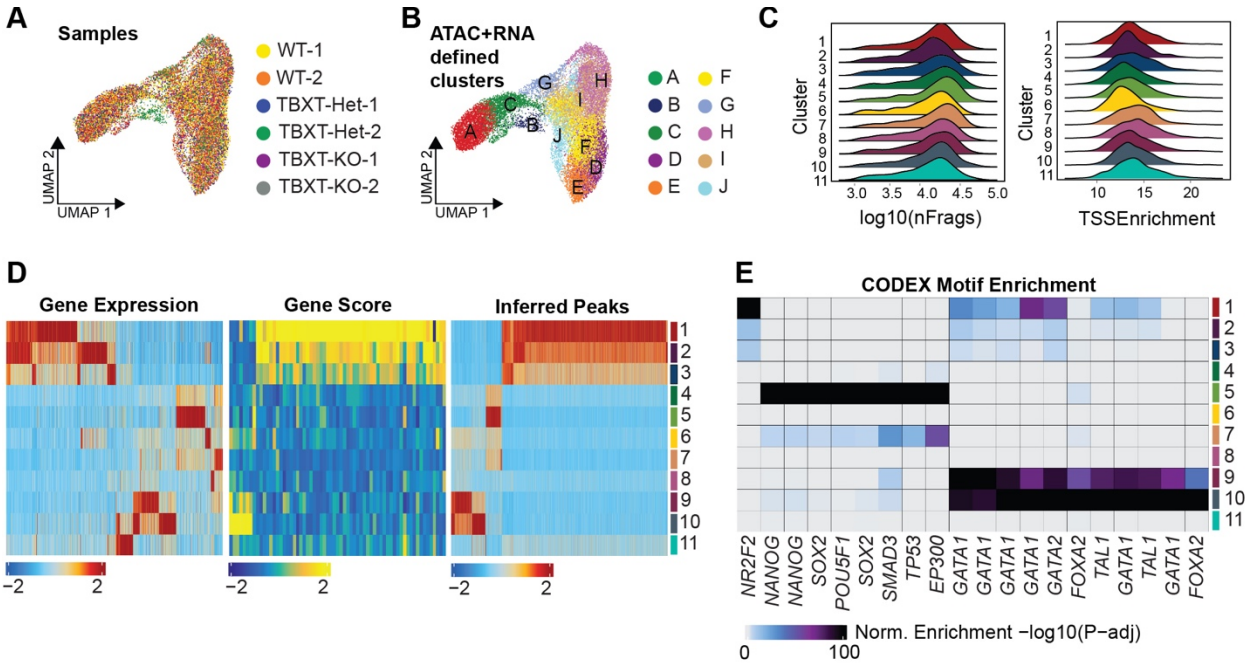


Figure S2.5: ArchR analysis of TBXT snATAC-seq data across clusters. (**A**) UMAP colored by sample and (**B**) ArchR-defined clusters. ArchR clustering integrates both snRNA-seq and snATAC-seq. (**C**) Ridgeplot showing the number of fragments (log10nFrags) and TSS enrichment score across snRNA-seq defined clusters from Fig. 2.2. (**D**) Comparison between gene expression (GeneExpressionMatrix), predicted expression based on accessibility (GeneScoreMatrix), and inferred peaks (PeakMatrix) across snRNA-seq defined clusters from Fig. 2.2. (**E**) CODEX motif enrichment across snRNA-seq defined clusters from Fig. 2.2.

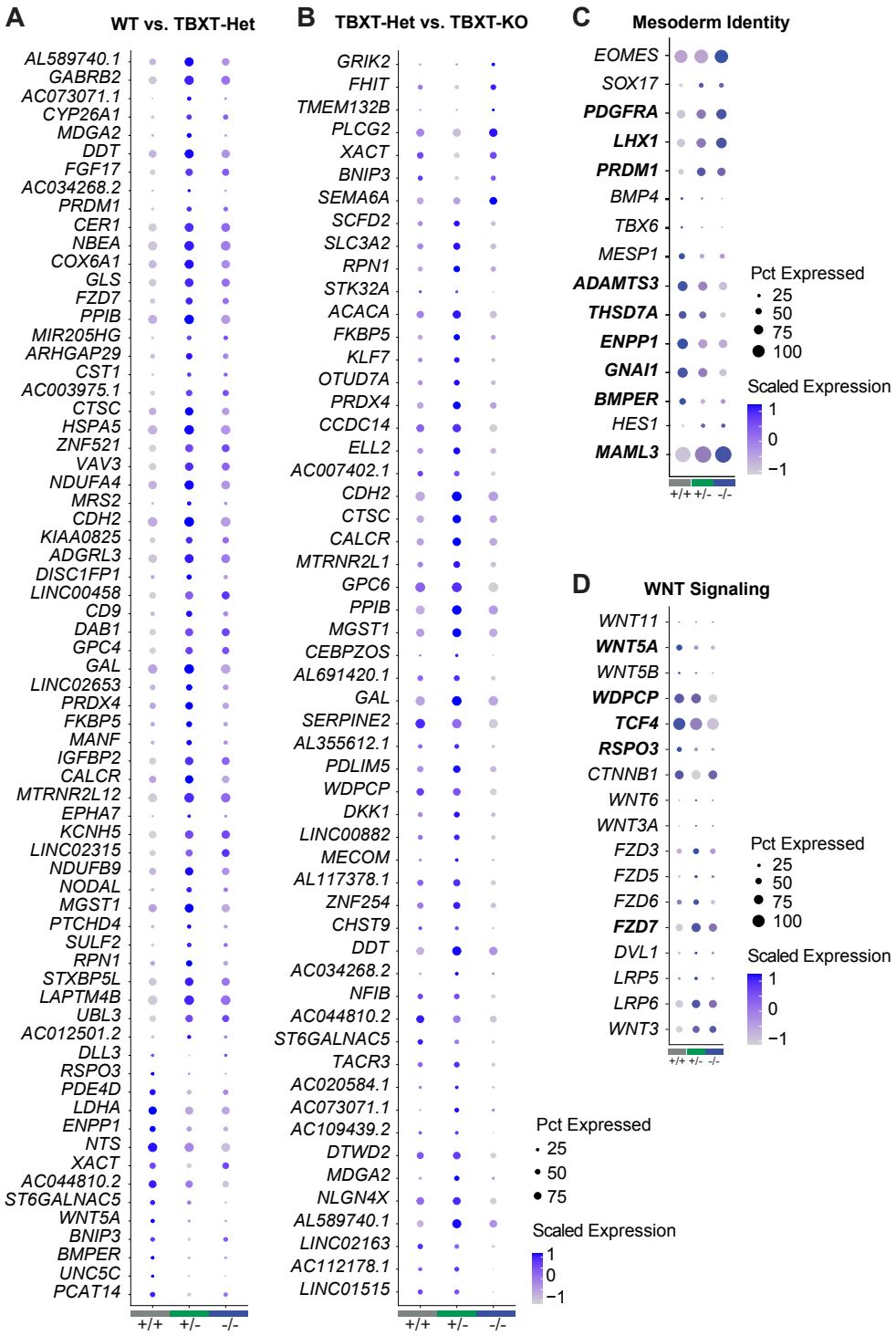


Figure S2.6: Differentially expressed genes in the mesoderm cluster. (**A**) Dot plot depicting all DE genes between WT and TBXT-Het or (**B**) TBXT-Het and TBXT-KO. (**C**) Dot plot depicting additional genes related to mesoderm identity and (**D**) WNT signaling. Bold genes in (**C-D)** indicate statistically significant DE genes (p adj. < 0.05, abs(Log2FC) > 0.25).

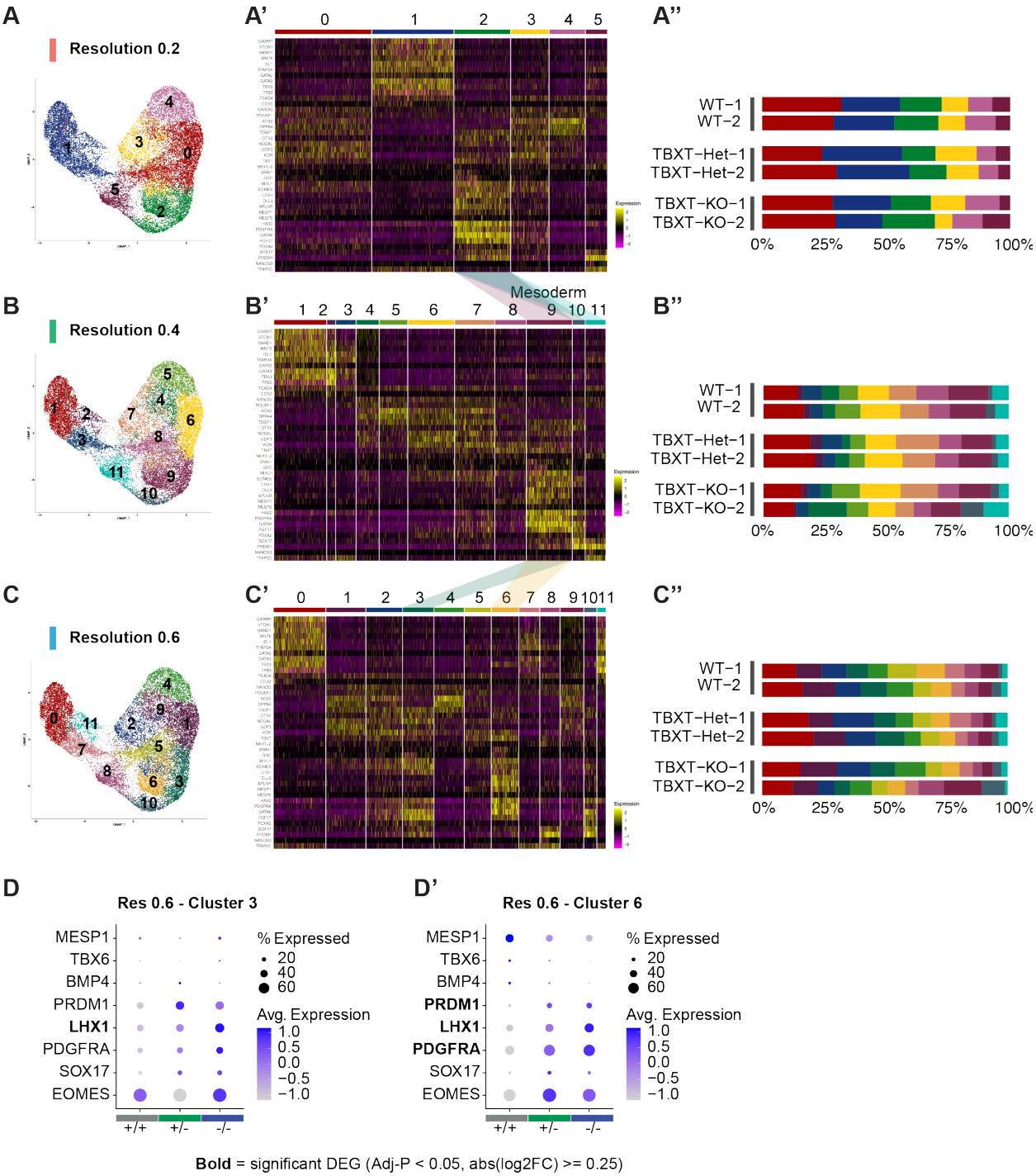
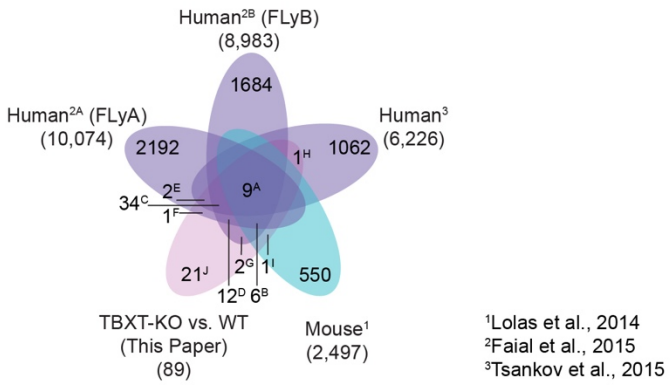


Figure S2.7: Characterization of snRNA-seq at additional clustering resolutions. UMAPs labeled by cluster at (**A**) resolution 0.2, (**B**) resolution 0.4, or (**C**) resolution 0.6. (**A'-C'**) Heatmap of key lineage markers of each cluster. (**A"-C"**) Stacked bar graph of the proportion of cells per sample assigned to each cluster. (**D**) Mesendoderm markers at resolution 0.6 for (**D**) cluster 3 or (**D'**) cluster 6.

A



B

A	**BMPER,** CADM1, DTWD2, EFNA5, EXT1, MAML3, MDFIC, SEMA6A, TCF7L1	M^{1} + $H^{2A, 2B, 3}$
B	**ENPP1, FGF17, UBL3,** LHX1, **RSPO3**, SAMD3	M^{1} + $H^{2A, 2B}$
C	**CYP26A1, DAB1, GABRB2, GPC4, ZNF521,** DLGAP1, LDB2, RYR3, UST, ADAMTS3, ATXN7L1, CADM2, CHST11, CHST9, DLEU1, ESRRG, GNAI1, GPR39, LRFN5, LRP1B, NEGR1, NFIB, NTN4, PBX1, PDGFRA, RALYL, SEMA6D, SERPINE2, SPAG16, TCF4, THSD7A, TMEM131, TMEM132B, WDPCP	$H^{2A, 2B, 3}$
D	**PRDM1, ST6GALNAC5, UNC5C, WNT5A,** ARHGAP24,CENPE, IDH1, STK32A, DDX58, GRIK2, NLGN4X, XYLT1	$H^{2A, 2B}$
E	CDH12, FUT9	$H^{2A, 3}$
F	TACR3	H^{2A}
G	**NTS,** RHOBTB3	H^{2B}
H	**KCNH5**	H^{3}
I	PCED1B	M^{1}
J	**LINC00458**, **LINC02315**, **MIR205HG**, **AC003975.1**, **AL589740.1**, **AC044810.2**, AFDN, AC068051.1, PKIB, MT-ND3, CST1, RPS2, AC022140.1, SLC9C1, LDHA, AC007402.1, NLGN4Y, AC109439.2, AC112178.1, LINC01515, LINC02163	Undetected

Figure S2.8: Intersection of differentially expressed genes in mesoderm and TBXT ChIP-seq datasets. (**A-B**) Venn Diagram depicting the intersection between DE genes identified between WT and TBXT-KO and each of the 4 TBXT ChIP-seq datasets. Bold indicates significant DE genes comparing both TBXT-Het and TBXT-KO to WT ($p<0.05$, Log2FC >= 0.25).

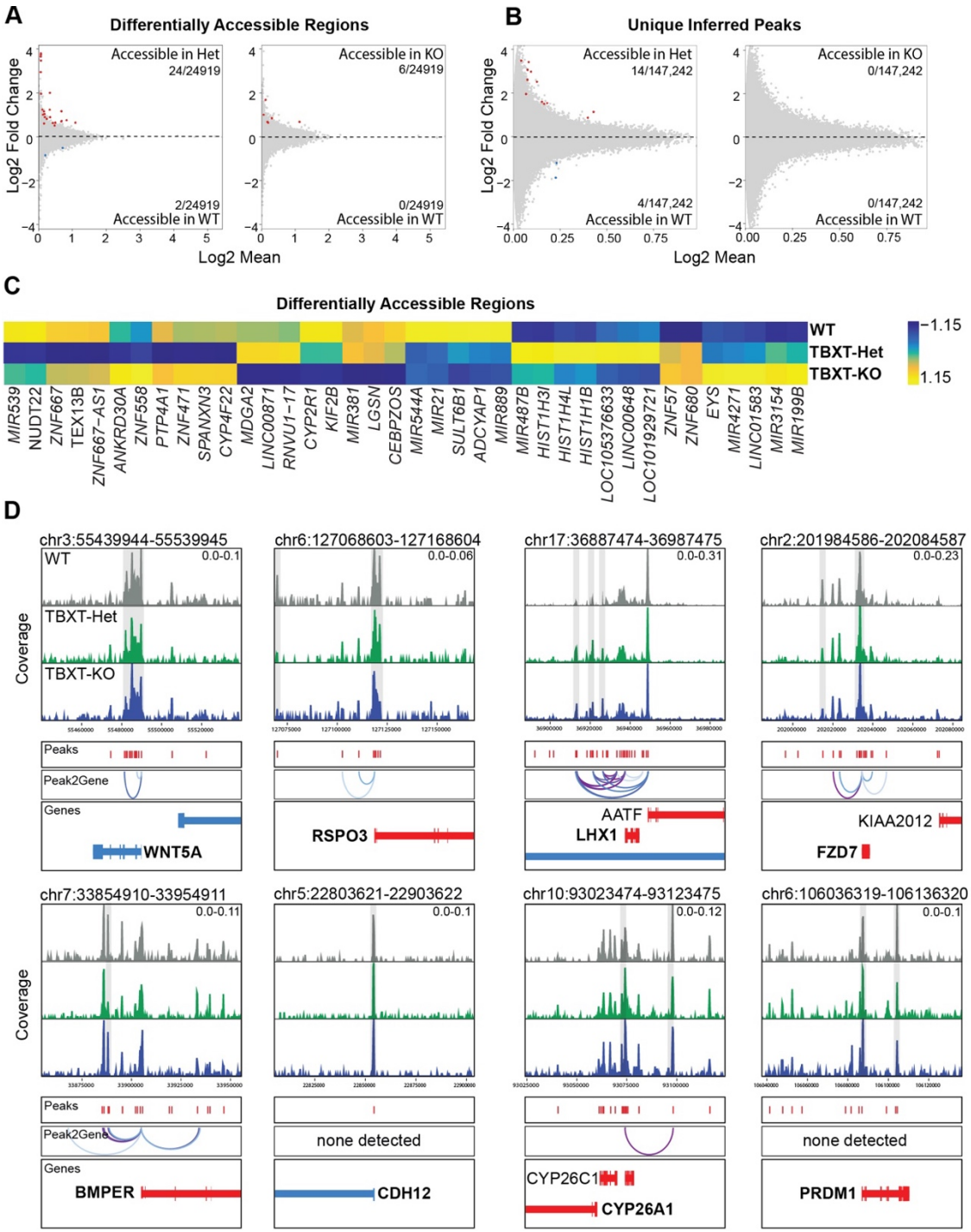


Figure S2.9: Differential Chromatin Accessibility in mesoderm across TBXT allelic series. (**A**) MA plots depicting pairwise DARs and unique peaks between WT and TBXT-Het or (**B**) WT and TBXT-KO identified in the mesoderm subset (FDR < 0.1, abs(Log2FC) > 0.5). (**C**) Heatmap of markerFeatures (DARs) detected across all 3 genotypes (FDR < 0.01 & abs(Log2FC) > 1.25) (**D**) Accessibility tracks comparing genotypes within the mesoderm subset centered around DE genes identified by snRNA-seq, including Peak2Gene linkage predictions of regulatory connections between distal accessible regions (peaks) and nearby genes. Gray bars indicate peaks of interest.

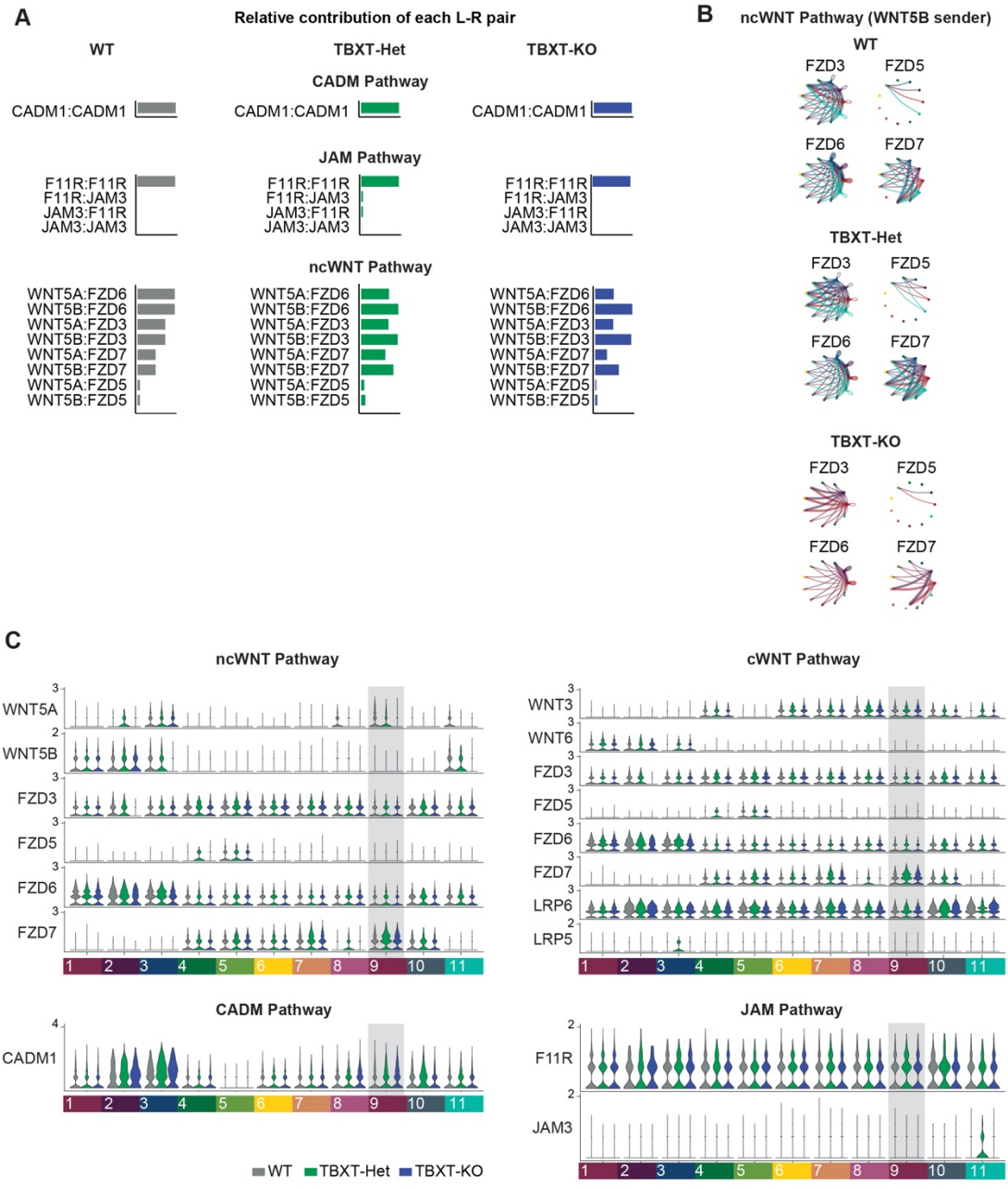


Figure S2.10: Relative contribution of ligand-receptor pairs in key pathways identified by Cellchat. (A) Bar plot of the relative contribution of each ligand-receptor (L-R) pair within the CADM, JAM, and ncWNT pathways. (**B**) Circle plots visualizing predicted patterns between WNT5B and designated FZD receptors across clusters as indicated in Fig. 2.4. (**C**) Violin plots comparing snRNA-seq expression for key ligands or receptors between genotypes for each cluster.

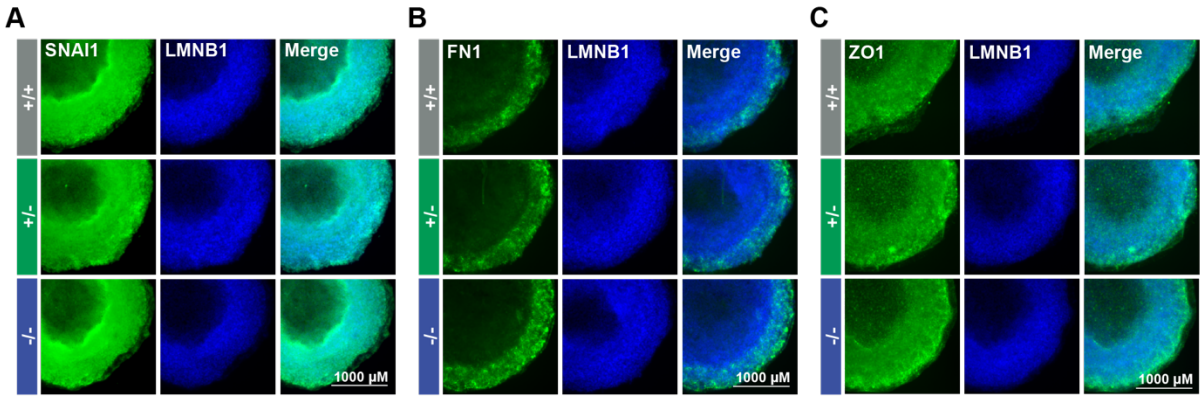


Figure S2.11: Characterization of EMT regulators in 2D gastruloids. IF for (**A**) SNAI1, (**B**) FN1, or (**C**) ZO1 in 2D gastruloids of each genotype at 48hrs. LMNB1 detects nuclei.

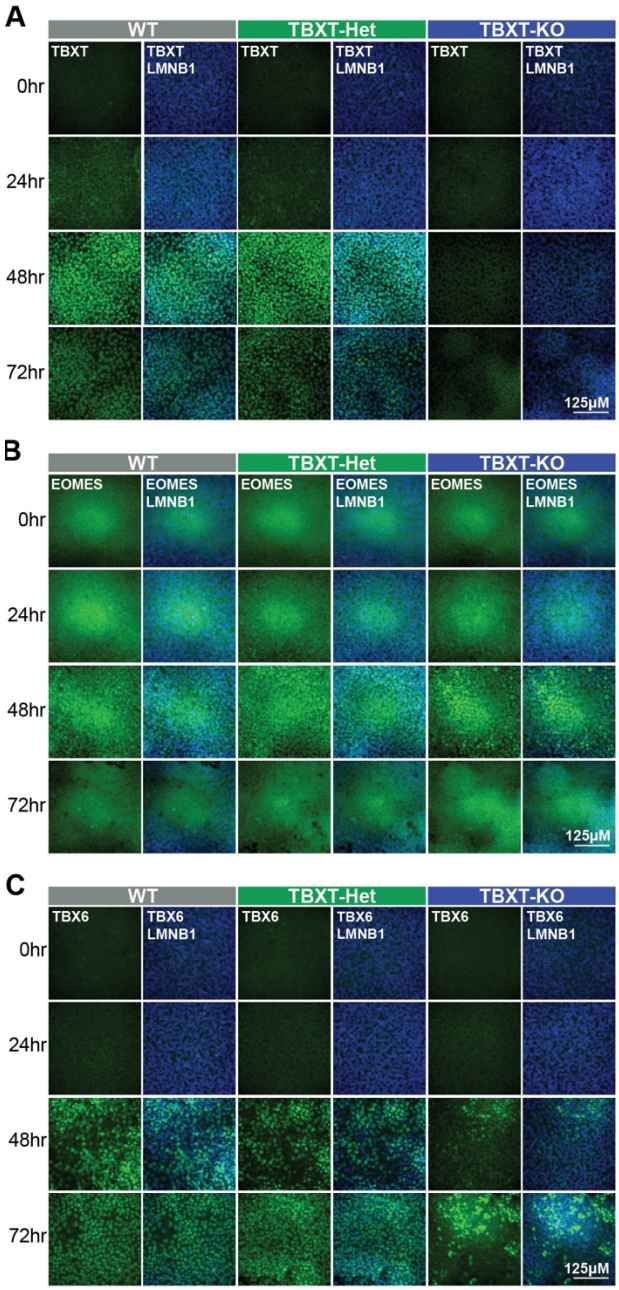


Figure S2.12: Characterization of cell types induced by Mesoderm Induction Media across TBXT allelic series . (**A-C**) Immunofluorescence for (**A**) TBXT, (**B**) EOMES, or (**C**) TBX6 at 0hr, 24hr, 48hr, and 72hr MIM exposure for each genotype. Protein of interest = green. LMNB1 = blue.

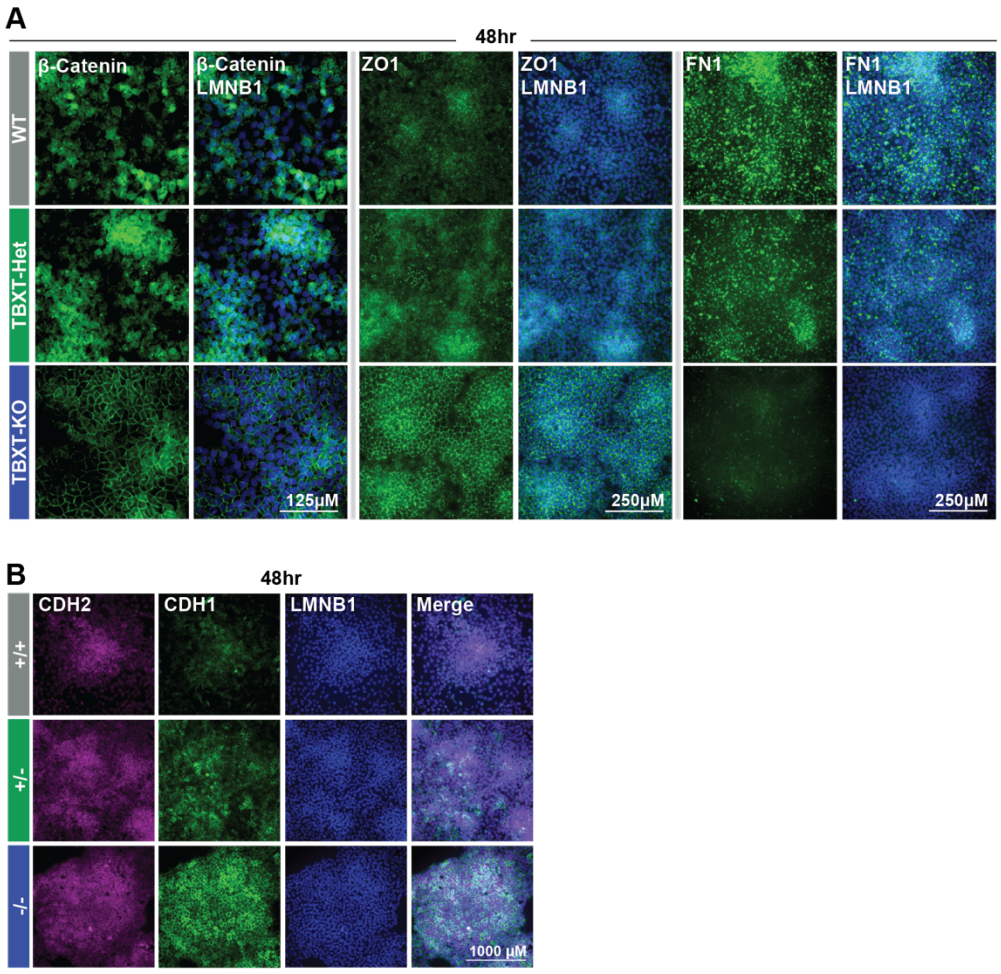


Figure S2.13: Characterization of junctional proteins induced by Mesoderm Induction Media across TBXT allelic series. (**A**) Immunofluorescence for β -catenin, ZO1, and FN1 at 48hr MIM exposure for each genotype. Protein of interest = green. LMNB1 = blue. (**B**) Immunofluorescence for CDH2 (magenta), CDH1 (green), or LMNB1 (blue) at 48 hours MIM exposure for each genotype.

Chapter 3: CDX2 dose-dependently influences the gene regulatory network underlying human extraembryonic mesoderm development

INTRODUCTION

Throughout embryonic development, proper regulation of gene dosage is necessary for the morphogenesis of the embryo and the tissues that support it. One transcription factor with critical roles in both extraembryonic and embryonic tissue development is CDX2, a homeodomain-containing protein that is critical for placental development and later the axial elongation of the tailbud (Beck *et al.*, 1995; van den Akker *et al.*, 2002; Chawengsaksophak *et al.*, 2004; Strumpf *et al.*, 2005; Savory *et al.*, 2009; Foley *et al.*, 2019). CDX2 is first expressed during the blastocyst stage of embryogenesis during the specification of the trophectoderm, where it is required for proper implantation (Chawengsaksophak *et al.*, 1997; Niwa *et al.*, 2005). As development progresses toward gastrulation, CDX2 expression becomes localized to the posterior primitive streak where CDX2+ cells give rise to the extraembryonic mesoderm. The extraembryonic mesoderm contributes to the formation of the heavily vascularized allantois, enables the fusion of the allantois with the chorion in a process known as chorioallantoic fusion, and contributes to vasculogenesis of the yolk sac mesoderm. These structures are critical for nutrition, gas exchange, and waste removal in the early embryo, and their failure to develop properly leads to asphyxiation, nutrient deprivation, and premature death. As such, epiblast-specific *Cdx2* mutant mice have severely underdeveloped allantoic buds and these buds do not fuse with the chorion, preventing the formation of a functional chorioallantoic placenta (Chawengsaksophak *et al.*, 2004; Brooke-Bisschop *et al.*, 2017; Foley and Lohnes, 2022). Yolk sac vasculogenesis is also impaired in these mutants, preventing normal circulation. Despite thorough documentation of the physical manifestations induced by impaired *Cdx2* expression, the specific molecular role of CDX2 in regulating the morphogenesis of these extraembryonic structures is not fully understood, nor has it been explored in a human system.

Interestingly, the loss of *Cdx2* in mice phenocopies the loss of several genes involved in the canonical WNT signaling pathway, including *Tcf/Lef*, *Wnt3*, and *Tbxt* (Rashbass *et al.*, 1991; Galceran *et al.*, 1999; Rossant and Cross, 2001; Inman and Downs, 2006). Much like *Cdx2*

mutant embryos, *Tbxt* mutant embryos are unable to survive past approximately E10 due to impaired allantois formation. Cells in the allantoic core where *Tbxt* would normally be expressed are significantly reduced in *Tbxt* mutants, however, cells of the outer mesothelium appear intact and these cells remain seemingly competent to adhere to the chorion (Inman and Downs, 2006). These observations suggest that *Tbxt* is necessary for the allantois to extend long enough to reach the chorion but is not necessarily required for chorioallantoic fusion itself. In addition, mice that are heterozygous for a functional *Tbxt* allele have an intermediate phenotype to WT or *Tbxt*-null mice, displaying variable allantois and blood island development and delayed vasculogenesis. Even so, allantoic growth in these heterozygotes is usually sufficient to allow for eventual chorioallantoic fusion, enabling pups to survive to term. Whether *Cdx2* heterozygosity also influences allantoic growth and chorioallantoic fusion is not yet known.

Acquisition of extraembryonic mesoderm identity is intimately tied to gastrulation, as this tissue emerges from the posterior primitive streak because of a network consisting of BMP, WNT, and NODAL signaling (Arnold and Robertson, 2009). *In vitro,* 48 hours of BMP4 exposure can induce this network in 2D cell colonies ("2D gastruloids") in a way that reproducibly generates concentric rings of epiblast-like cells, embryonic mesoderm-like cells, endoderm-like cells, and extraembryonic mesoderm-like cells (Warmflash *et al.*, 2014; Minn *et al.*, 2020, 2021). This model allows us to investigate both how specific genes active during early gastrulation augment cell identity and how changes in adjacent tissues influence cell-cell communication and the gene regulatory networks underlying lineage emergence.

In this study, we employ this 2D gastruloid model and multiomic single nucleus RNA sequencing (snRNA-seq) and Assay for Transposase Accessible Chromatin sequencing (snATAC-seq) to identify how CDX2 regulates the proper morphogenesis of extraembryonic mesoderm and the extent to which this regulation is controlled in a dose-dependent manner during early gastrulation. We demonstrate that varying CDX2 dose at this stage of early development directly influences genes involved in cell-cell adhesions, extracellular matrix

integrity, cytoskeletal architecture, and HOX gene expression, in addition to influencing key regulators of extraembryonic mesoderm fate such as the WNT and GATA families of factors. We then compare this dataset to our data generated in Chapter 2 which looks at the impact of TBXT dose on extraembryonic mesoderm development (Bulger *et al.*, 2024). This comparison allows us to further isolate several genes with shared misregulation that are involved in cytoskeletal integrity and tissue permeability, including *ANK3* and *ANGPT1,* which are both involved in VEGF-directed blood vessel maturation. Taken together, this study suggests that CDX2 activates gene regulatory networks that regulate allantois formation and vasculogenesis in a dose-dependent manner, including but not limited to genes involved in cell adhesion, motility, and membrane permeability.

RESULTS

Generation of a hiPSC CDX2 allelic series and 2D gastruloids

To explore the effect of CDX2 dose on the development of the extraembryonic mesoderm population, we engineered a WTC11-LMNB1-GFP-derived allelic series consisting of $CDX2^{+/+}$ (WT), $CDX2^{+/-}$ (CDX2-Het), and $CDX2^{-/-}$ (CDX2-KO) human induced pluripotent stem cell lines (hiPSCs). These mutants were generated by targeting the first exon of CDX2 with CRISPR/Cas9, generating a premature stop codon on one or two alleles, respectively, as confirmed by Sanger sequencing (Fig. 3.1A, S3.1, Table 4.1, methods). We conducted a western blot for CDX2 in a sparsely seeded monolayer of cells exposed to BMP4 for 48 hours and this revealed the expected stepwise decrease in CDX2 expression across the allelic series, with the highest protein levels in the WT, intermediate levels in CDX2-Het, and no detectable protein in CDX2-KO cells (Fig. 3.1B). The complete absence of CDX2 protein in CDX2-KO cells was further confirmed through immunofluorescence (IF) (Fig. 3.1D-E).

Lineage emergence is minimally altered in 2D gastruloids of varying CDX2 dose

To investigate how CDX2 dose influences extraembryonic mesoderm specification and morphogenesis, we differentiated the allelic series into 2D gastruloids via 48 hours of BMP4 exposure (Warmflash *et al.*, 2014; Minn *et al.*, 2020) (Fig. 3.1C, methods). We then conducted two biological replicates of multiomic single-nucleus RNA-sequencing (snRNA-seq) and single-nucleus Assay for Transposase-Accessible chromatin (snATAC-seq) (Buenrostro *et al.*, 2015) on gastruloids of each genotype after 48 hours of BMP4 treatment. Through this in-depth sequencing approach, we sought to precisely define how CDX2 dose influences cell identity and the expression of morphogenesis regulators during early gastrulation. In addition, CDX2 has been shown to sustain newly accessible chromatin regions in mature tissues (Verzi *et al.*, 2010, 2013; Neijts *et al.*, 2017; Saxena *et al.*, 2017; Kumar *et al.*, 2019) and also to interact with the Brg1 subunit of the switch-sucrose non-fermentable (SWI-SNF) chromatin remodeling complex (Yamamichi *et al.*, 2009; Nguyen *et al.*, 2017). We were interested in the extent to which CDX2 can dose-dependently influence chromatin accessibility in this nascent extraembryonic tissue.

After computationally pooling cells of all three genotypes and conducting dimensionality reduction and clustering of snRNA-seq data using the R package, Seurat (Satija *et al.*, 2015), our analysis yielded ten distinct clusters reflecting various populations expected in the gastrulating embryo. Clusters were identified as outlined in Chapter 2, which shares a WT dataset with the current analysis (Fig. 3.2B-C). These included three extraembryonic cell populations (C1–C3; “Extraembryonic-1–3”, epiblast-like cells (C4; “Epiblast”), three primitive streak-like cell populations (C5–C7; “PS-1–3“), nascent mesoderm-like cells (C8; “Mesoderm”), nascent endoderm-like cells (C9; “Endoderm”), and primordial germ cell-like cells (C10; “PGCLC”) (Fig. 3.2A-C, S3.2-3.3). The proportion of cells of each genotype assigned to each cluster was largely consistent, suggesting that the acquisition of cell identity during early gastrulation is not significantly affected by the loss of CDX2 (Fig. 3.2A’).

The extraembryonic cells that emerge in 2D gastruloids have previously been shown to express genes that are shared between extraembryonic mesoderm/amnion and trophectoderm, including *CDX2, GATA3, TFAP2A, HAND1,* and *WNT6* (Bernardo *et al.*, 2011; Warmflash *et al.*, 2014; Ma *et al.*, 2019; Niu *et al.*, 2019; Zheng *et al.*, 2019; Minn *et al.*, 2020). To distinguish between these possible cell types and identify differences between the three extraembryonic clusters, we evaluated these clusters for key markers of trophectoderm, late-amnion, and early-amnion lineages. This analysis revealed a bias in all three clusters toward late-amnion identity (*GAPBRP, HEY1, HAND1, VTCN1, TPM1, IGFBP3, ANKS1A*) relative to embryonic clusters, in agreement with published findings that BMP4 drives primed hiPSCs toward an amnion-like fate (Rostovskaya *et al.*, 2022) and reflecting an extraembryonic mesoderm rather than trophectoderm origin (Fig. S3.5). This late-amnion expression pattern was the highest in extraembryonic-1 and the lowest in extraembryonic-3, suggesting extraembryonic-1 is a relatively more differentiated extraembryonic mesoderm while extraembryonic-3 remains more nascent. Upon analysis using the R package ArchR (Granja *et al.*, 2021), the companion snATAC-seq dataset revealed that the accessible regions of chromatin in cells within these extraembryonic clusters, whose identity information was imported from our snRNA-seq analysis based on matching cell barcodes (methods), are enriched for motifs for the TEAD, TFAP, and HAND families of transcription factors, which are critical regulators of extraembryonic mesoderm fate (Fig. 3.2D, S3.3-3.4). Again, we observe a stepwise enrichment pattern where these motifs are most highly enriched in extraembryonic-1 and have variable enrichment in extraembryonic-2 and extraembryonic-3. We therefore designated clusters 1-3 as extraembryonic-mesoderm-like late (“ExeM-Late”), middle (“ExeM-Middle”), and early (“ExeM-Early”), respectively. This annotation coincided with the trajectory analysis which ordered cells in pseudo time from ExeM-Early to ExeM-Late (Fig. S3.5B)

Looking at clusters correlating to embryonic cell types, cluster 4 exhibits canonical hallmarks of epiblast cell fate including gene expression of *POU5F1*, *SOX2*, and *NANOG.* The

three primitive streak clusters share canonical elements of the PS gene signature including *TBXT, MIXL1,* and *EOMES,* and are differentiated from one another by the progressive downregulation of epiblast markers such as *SOX2, TDGF1*, and *NODAL*. Based on this observation, we designated “PS-1” (C5) as “PS-Early”, “PS-2” (C6) as “PS-Middle”, and “PS-3” (C7) as “PS-Late.” Clusters 8 and 9 share a mesendoderm-like signature including enriched GATA and TBX family motifs and are distinguished by a mesoderm-like gene expression signature in cluster 8 (*MESP1, MESP2, ALPNR, TBXT*) and an endoderm-like gene expression pattern in cluster 9 (*FOXA2, SOX17, PRDM1*) (Fig. 3.2B-D). Lastly, cluster 10 reflects PGCLC identity, as evidenced by the coexpression of *SOX17, PRDM1, NANOS3*, and *TFAP2C*. The SOX9 motif was also enriched in the accessible regions that distinguish cells of this cluster, in agreement with its role as a marker of PGCLC-derived Sertoli cells later in development (Hayashi *et al.*, 2018) (Fig. 3.2D).

CDX2 primarily influences the GRN underlying extraembryonic mesoderm identity

To understand how CDX2 influences gene expression during early embryonic morphogenesis, we next isolated each cluster and compared the number of differentially expressed genes (DEGs) between WT vs. CDX2-Het, WT vs. CDX2-KO, or CDX2-Het vs. CDX2-KO. As expected, the ExeM-Late cluster yielded both the highest CDX2 expression in WT cells (Fig. 3.3A) and the largest number of DEGs between WT and CDX2-KO relative to the other clusters. Based on these changes in expression and because of the biological relevance of this cluster to extraembryonic mesoderm development, we focused on this cluster for subsequent analyses (Fig. 3.3B).

Comparing CDX2-Het to WT, we detected 170 genes downregulated and 77 genes upregulated in the CDX2-Het (Fig. 3.3C). Downregulated genes included canonical WNT inhibitor *DCDC2*, cell-cell adhesion regulator *CDH1*, WNT ligand *FZD6*, and BMP inhibitor *BMPER,* while the upregulated genes largely consisted of ribosomal proteins. GO biological process enrichment for the genes downregulated in CDX2-Het revealed several pathways related to “cell projection

morphogenesis" and "cell adhesion", and also several neural-related pathways (Fig. 3.3D). These neural pathways are largely driven by genes involved in axon guidance such as *FYN*, *UN5B*, and *SEMA3E*, so while they may reflect a neural bias in cell identity it is also probable that they reflect broader patterns in misregulated cell migration.

Comparing CDX2-KO to WT, we detected 342 genes downregulated and 84 genes upregulated in the CDX2-KO (Fig. 3.3C'). Like in CDX2-Het, downregulated genes in CDX2-KO included several regulators of the WNT signaling pathway including *WNT5B*, *DCDC2*, and *FZD6*, as well as several genes related to cell-cell adhesion including *CDH1* (Fig. 3.3C'). Genes involved in cell migration and axon guidance such as *GLI3*, *UNC5B*, *SEMA3E*, and *ISL1* were also identified. GO biological process enrichment revealed a 4-fold enrichment of genes involved in "canonical WNT signaling," as well as enrichment for "vasculature development", "circulatory system development", "tube morphogenesis", "cell-cell signaling", and "locomotion" (Fig. 3.3D'. S3.7) Taken together, the differentially expressed genes that are seen in both CDX2-Het and CDX2-KO vs. WT correspond to properties of the developing allantois and other extraembryonic-mesoderm-derived structures including cell projection morphogenesis, adhesions, and WNT signaling, and potentially underly the impaired vasculogenesis phenotype observed in mouse models.

To clarify the extent to which these pathways are dependent on CDX2 dosage, we next determined the number of overlapping downregulated genes from both the CDX2-Het vs. WT and the CDX2-KO vs. WT comparisons. Of the 170 genes downregulated in CDX2-Het and the 342 genes downregulated in CDX2-KO, we found that 114 genes were conserved including *SMAD3*, *MEIS2*, and *DCDC2* (Fig. 3.3E-F) GO Molecular Function analysis of these 114 genes revealed significant enrichment for "extracellular matrix structure", "cell adhesion molecule binding", and "cytoskeletal protein binding" (Fig. 3.3E. S3.7). GO biological process enrichment for "vasculature development" remained enriched by approximately 3-fold (FDR = 0.03), and of the 29 genes driving "vasculature development" enrichment in CDX2-KO, 10 remained differentially expressed

in CDX2-Het (*ISL1, ATP2B4, GLI3, ENG, UNC5B, GPLD1, RTN4, COL4A2, HEY1, SEMA3E*). In contrast, there were 23 genes upregulated in both CDX2-Het and CDX2-KO, and the GO enrichment terms for these genes included "estrogen response element", "NADH dehydrogenase activity" and "structure of ribosome." However, it is difficult to evaluate the biological significance of these terms due to the low number of overlapping genes (Fig. 3.3E'). We additionally plotted the top genes that are differentially expressed between the WT and the CDX2-KO to clarify whether CDX2-Het often has a gene expression profile that is intermediate between the WT and CDX2-KO even when certain genes are not detected as differentially accessible between WT and CDX2-Het (Fig. 3.3F). This analysis suggested that genes downstream of CDX2 often have intermediate expression when WT and CDX2-Het do not have statistically different expression, reinforcing dose-sensitivity in downstream gene expression. These results demonstrate that CDX2 dose-dependently influences determinants of physical cell structure as well as the gene regulatory network involved in the development of extraembryonic tissue and largely acts as a transcriptional activator.

snATAC-seq reveals loss of accessibility in regions with CDX2 motifs

To better understand how CDX2 dosage influences the gene regulatory network underlying extraembryonic-mesoderm development, we next turned to the paired snATAC-seq dataset generated from the same nuclei as our snRNA-seq dataset. As for snRNA-seq, we subset the ExeM-Late cluster and identified regions with differential chromatin accessibility (differentially accessible regions; DARs) between either WT and CDX2-Het or WT and CDX2-KO populations.

Through this analysis, we identified very few DARs between CDX2-Het and WT in the Exe-Late cluster (13/174,048 DARs with increased accessibility in CDX2-Het and 12/174,048 DARs with increased accessibility in WT). In contrast, comparing CDX2-KO to WT in the same cluster revealed a slightly higher number of DARs, as we detected 29/174,048 DARs with increased accessibility in CDX2-KO and 232/174,048 DARs with increased accessibility in WT

(Fig. 3.4A). The relatively higher number of reduced-accessibility peaks detected in the CDX2-KO compared to the WT condition suggests that CDX2 is either directly or indirectly responsible for making specific regions of chromatin more accessible or maintain accessibility in the extraembryonic mesoderm population. Additionally, the lack of DARs detected in the CDX2-Het vs. WT comparison indicates that an intermediate level of CDX2 is sufficient to induce or maintain a WT-like chromatin accessibility profile (Fig. 3.4B).

To better understand the composition of DARs that are uniquely accessible in each CDX2 condition, we next conducted motif enrichment to look for motifs that are overrepresented in these regions. Motifs enriched in peaks that are differentially accessible between WT and CDX2-Het or WT and CDX2-KO include proliferation regulators *KLF7*, *SP6*, and *GATA1*, but these motifs are enriched at a relatively low level (-log10(Adj. P) ~ 3) and the low number of DARs driving these results makes their biological relevance uncertain (Fig. 3.4C, S3.8). In contrast, the top motifs in DARs that are more accessible in WT but less accessible in CDX2-KO exhibited significantly higher enrichment scores and consisted almost exclusively of homeobox genes, including *CDX2*, *CDX4*, and posterior *HOX* genes, and also *GATA* factors (-log10(Adj. P) ~60) (Fig. 3.4C). While this may reflect the established role of CDX2 in regulating HOX gene expression and associated chromatin dynamics (Amin *et al.*, 2016; Neijts *et al.*, 2017), CDX genes and posterior HOX genes share very similar binding motif (TTAT) that is distinct from anterior or central HOX genes (Ekker *et al.*, 1994; Berger *et al.*, 2008; Noyes *et al.*, 2008; Mann, Lelli and Joshi, 2009; Bulajić *et al.*, 2020). Because of the limited detection of posterior HOX gene expression in our datasets (Fig. S6A), we believe that these results likely capture the redundancy in these motif annotations rather than true HOX binding. These results highlight how CDX2 expression correlates with the accessibility of regions containing its binding motifs, suggesting it has the capability to remodel chromatin in regions where it is bound.

We next conducted transcription factor footprinting analysis to understand how TF binding at these CDX motifs is influenced by CDX2 dosage. As anticipated, we observe a lower footprint

at the CDX2 motif for CDX2-KO reflective of reduced occupancy relative to WT (Fig. 3.4D). The footprint in the CDX2-Het was intermediate between WT and CDX2-KO, suggesting that the reduced gene expression correlates with reduced TF binding at CDX motifs. Because TF footprinting denotes average motif occupancy, it is unclear whether this is because fewer CDX TFs bind to a specific locus in favor of another with higher affinity or due to a global reduction in CDX TF binding across all loci containing CDX motifs. However, the lack of differentially accessible peaks observed in CDX2-Het vs. WT indicates that the intermediate CDX2 expression level and subsequent reduced occupancy of CDX motifs remains sufficient to induce a WT-like chromatin accessibility profile, despite the lower dose influencing gene expression levels for markers related to cell architecture and adhesions.

CDX2 dose-dependently influences gene expression within the HOXB locus

Because CDX and posterior HOX motifs share a common binding motif that enriched in DARs, and because CDX2 influences HOX gene expression in the embryo proper(Neijts *et al.*, 2017), we next asked how CDX2 dosage influences HOX gene expression patterns in the extraembryonic mesoderm. HOX clusters can be subdivided into three subsections containing paralogs A-D; the 3' cluster containing HOX genes 1-4, the middle cluster containing HOX genes 5-9, and a 5' cluster consisting of HOX genes 10-13 (Neijts *et al.*, 2017), The 3' cluster is thought to be activated by WNT signaling, while all paralogs within the middle cluster depend on CDX-transcription factors to become accessible. The 5' cluster is activated much later in development by central HOX genes in a colinear fashion (Neijts *et al.*, 2017).

In agreement with previous work documenting an anterior homeotic shift in CDX2 mutant mice (van den Akker *et al.*, 2002), we observed slightly elevated expression of the central HOX genes, specifically *HOXB3*-*HOXB7*, in WT relative to CDX2-KO (Fig. S3.6A). Of these, only *HOXB6* was significantly differentially expressed between CDX2-KO and WT, and no HOX genes were significantly differentially expressed between CDX2-Het and WT. Most genes from *HOXA*,

HOXC, and *HOXD* paralogs were not detected in our dataset, and this may be because the HOXB cluster typically slightly precedes the expression of other paralogs (Denans, Iimura and Pourquié, 2015). Alternatively, HOX paralogs have been shown to exert different roles in different tissues (Kachgal, Mace and Boudreau, 2012), and *HOXB3*, *HOXB5*, and *HOXB7* specifically have been shown to impact vasculogenesis (Miano *et al.*, 1996; Myers, Charboneau and Boudreau, 2000; Wu *et al.*, 2003). It is, therefore, possible that HOXB paralogs serve a unique role in early extraembryonic mesoderm development and subsequent placental vasculogenesis.

Like CDX2-KO, central HOX expression in CDX2-Het was also slightly reduced relative to WT (Fig. S3.6A). This suggests that while intermediate CDX2 expression may be sufficient to induce changes in chromatin accessibility, a higher threshold of expression may be required to activate downstream HOX genes to WT levels. In contrast, the anterior HOX gene *HOXB2* was slightly higher in CDX2-KO than both CDX2-Het and WT, likely because of the established role of more posterior HOX genes in suppressing the function of more anterior HOX genes (Chisaka and Capecchi, 1991; Lufkin *et al.*, 1991, 1992; Duboule and Morata, 1994; Iimura and Pourquié, 2006; Denans, Iimura and Pourquié, 2015) (Fig. S3.6A). The posterior HOX genes, specifically *HOXB9*, *HOXA13*, and *HOXC13*, were detected in very few cells. Taken together, these results illustrate a dose-dependent role CDX2 in limiting the expression of downstream central HOX genes.

Overall, these results reflect the ability of CDX2 to modulate HOX expression, including the downregulation of central HOX genes and upregulation of anterior and posterior HOX genes, in a dose-dependent and possibly paralog-specific manner.

CellChat reveals a dose-dependent role for CDX2 in regulating the non-canonical WNT signaling pathway

In the early extraembryonic mesoderm, paracrine and juxtracrine signals from a variety of tissues are required to orchestrate morphogenesis (Stewart, 1996; Downs *et al.*, 2009). With this

in mind, we utilized the R package CellChat (Jin *et al.*, 2021) to identify patterns of ligand-receptor communication across clusters. We first investigated how CDX2 dosage influences broad patterns in pathway activation by isolating pathways with identifiable changes in “information flow”, which predicts patterns in cell-cell communication by quantifying the changes in signals between or within cell types.

This analysis revealed that the amount of information flow across most pathways is largely conserved across CDX2 genotypes, except the non-canonical WNT (ncWNT) signaling pathway which has reduced information flow in both CDX2-Het and CDX2-KO (Fig. 5A). We visualized the changes in predicted communication between and within clusters in circle plots, where line thickness correlates with the degree of predicted communication. The marked decrease in ncWNT communication within CDX2-Het and CDX2-KO was most clearly derived from changes in the three ExeM clusters (C1-C3), where both paracrine signaling (loops) and juxtracrine signaling (lines) are reduced (Fig. 3.5B). We quantified the relative contribution of the different ligands and receptor pairs that define the ncWNT signaling pathway in each genotype, which in CDX2-Het and CDX2-KO revealed conserved signaling between WNT5A and various frizzled receptors but nearly eliminated signaling between WNT5B and those same receptors, suggesting a critical role for WNT5B in the maintenance of the ncWNT pathway (Fig. 3.5B-C). We next examined trends in the clusters acting as senders, receivers, mediators, and influencers of the ncWNT pathway. In this analysis, mediators specifically control cell communication between any two groups, and influencers influence information flow more generally (Jin *et al.*, 2021). Most clearly, we identified a marked decrease in the ability of CDX2-Het and CDX2-KO to act as senders and mediators in the three ExeM clusters (C1-C3). This pattern was also observed in PGCLCs, which share an extraembryonic origin (Sasaki *et al.*, 2016).

This pattern, in conjunction with the snRNA-seq data, suggests that the changes in information flow across genotypes are largely due to the inability of CDX2-Het and CDX2-KO to express WT levels of *WNT5B* in their extraembryonic populations, and implicate the non-

canonical WNT signaling pathway in regulating early extraembryonic mesoderm development in a CDX2 dose-dependent manner. CDX2 has been previously been shown to be important for the expression of ncWNT ligands WNT5A and WNT5B, and loss of WNT5B phenocopies the axial truncations observed with the loss of CDX2 (Savory *et al.*, 2009; Anand *et al.*, 2023). WNT5B has also been shown to influence both canonical WNT/ β-catenin signaling and VEGF-C expression (Kanazawa *et al.*, 2005), both of which have been shown to regulate vasculogenesis *in vivo* (Cao *et al.*, 1998; Drake *et al.*, 2000; Kanazawa *et al.*, 2005; Shibuya, 2011). Additionally, components of the ncWNT pathway, including WNT5 and WNT11, have been shown to influence angiogenesis via regulation of the VEGF inhibitor FLT1 (Stefater *et al.*, 2011; Akoumianakis, Polkinghorne and Antoniades, 2022). Thus, CDX2-driven *WNT5B* expression may be required to properly regulate these two pathways and sustain development within the allantois. Additional experiments, possibly using *in vivo* models of allantois development, will be required to validate these hypotheses.

CDX2 and TBXT jointly regulate genes involved in extraembryonic mesoderm development

Because CDX2 and TBXT null animals share an embryonic lethal defect in allantois development and chorioallantoic fusion, we next asked whether we could identify common genes that are misregulated during extraembryonic mesoderm development both CDX2-KO and TBXT-KO conditions. We isolated the analogous extraembryonic mesoderm population from the TBXT dataset generated in Chapter 2 (“Extraembryonic-Late”), conducted TBXT-Het vs. WT and TBXT-KO vs. WT comparisons within this cluster, and looked at how the resulting lists of differentially expressed genes intersected with those of the CDX2 allelic series ExeM-Late population.

From this comparative analysis, we uncovered 11 genes downregulated and 32 genes upregulated in TBXT-Het vs. WT and 35 downregulated and 49 upregulated genes in TBXT-KO relative to WT (Fig. 3.6A). Of the 35 genes downregulated in TBXT-KO, 8 were also downregulated in CDX2-KO relative to WT. These 8 genes included *ANK3, LSAMP,* and *ANGPT1*,

which regulate adhesions and the organization of the actin cytoskeleton (Babcock *et al.*, 2018; Liu *et al.*, 2021), and the canonical WNT inhibitor *DCDC2* (Fig. 3.6B). 5 of these 8 genes were also significantly downregulated in CDX2-Het and 0 were significantly downregulated in TBXT-Het, although trends toward intermediate expression are evident (Adj. $P < 0.05$, $Log2FC > 0.25$) (Fig. 3.6B). Notably, both *ANK3* and *ANGPT1* have been shown to regulate angiogenesis *in vivo* via the VEGF signaling pathway (Gavard, Patel and Gutkind, 2008; Cao *et al.*, 2017; Liu *et al.*, 2021). VEGF increases vascular permeability while ANGPT1 and ANK3 reduce permeability (Senger *et al.*, 1983; Thurston *et al.*, 1999; Liu *et al.*, 2021), and the proper balance of these factors is likely required for proper placental vasculogenesis. In addition, both ANGPT1 and ANK3 reduce the cell surface localization of ß-catenin and endothelial barrier function. This function is impaired in ANGPT1 heterozygotes (Durak *et al.*, 2015; d'Apolito *et al.*, 2019), suggesting that proper regulation of their expression is required for the development of extraembryonic-mesoderm-derived structures.

Of the 49 upregulated genes in TBXT-KO, 9 were also upregulated in CDX2-KO (Fig. 3.6C). These 9 genes included the cytoskeleton modulator *TMSB10*, hedgehog pathway and myogenesis effector *CDON*, and placental adhesion regulators *FBLN1* and *IGFBP7* (Fig. 3.6D). Like *ANK3* and *ANGPT1*, *TMSB10* influences VEGF expression, specifically inhibiting VEGF-induced endothelial cell proliferation, migration, and invasion (Lee *et al.*, 2005; Pan *et al.*, 2020). Five of these 9 genes were also significantly upregulated in CDX2-Het and 0 were upregulated in TBXT-Het. While speculative based on bioinformatics analysis, these these results suggest that one method by which CDX2 and TBXT regulate the development of extraembryonic mesoderm structures is by promoting the expression of *ANGPT1* and *ANK3*, which in turn sequester β-catenin in the cytoskeleton, preventing its nuclear translocation and fine-tuning the expression of downstream canonical WNT effectors. These factors also likely influence the activity of the VEGF signaling pathway and the associated development of extraembryonic mesoderm.

In addition to comparing differentially expressed genes, we also utilized our snATAC-seq data to look at whether certain peaks were uniquely accessible in each mutant line relevant to WT. As previously indicated, there are 29 peaks with decreased accessibility and 232 peaks with increased accessibility in WT relative to CDX2-KO within the ExeM-Late cluster (Fig., 4A, Table S2). In contrast, there are 6 peaks with decreased accessibility and 5 peaks with increased accessibility in WT relative to TBXT-KO within the equivalent cluster. Of these 11 regions, 5 overlap between the two datasets (Fig. S9B, Table S5). The region adjacent to CEBPZOS was more accessible in WT relevant to both mutant lines, while regions adjacent to USP4, DZIP1, TPPGS1, and ENSG00000286456 were less accessible in WT relative to the mutant lines. These genes are generally involved with the mitochondrial membrane, ER function, hedgehog signaling, and microtubule binding. The small number of regions differentially accessible between TBXT-KO and WT suggests that TBXT may act on a small handful of regions made or kept accessible by CDX2, however, TBXT itself seems to remodel chromatin only minimally during the early development of the extraembryonic mesoderm.

Taken together, the changes in gene expression shared between the TBXT-KO and CDX2-KO compared to WT reveal effectors that regulate cytoskeletal architecture, cell adhesions, and cell permeability via the VEGF and WNT signaling pathways. The canonical WNT pathway regulates several pro-angiogenic molecules including VEGF family members, whose expression in turn have been shown to positively correlate with cytoplasmic β-catenin localization in the context of cancer (Kasprzak, 2020). While further validation is necessary, we hypothesize that proper regulation of these factors may be required for the development of the allantois and subsequent chorioallantoic fusion, and their misregulation contributes to the early embryonic lethality observed *in vivo* in the absence of TBXT or CDX2.

DISCUSSION

Proper morphogenesis of extraembryonic structures is critical for an embryo to develop to term. CDX2 is required for the development of several extraembryonic-mesoderm-derived structures, including the growth of the allantois, chorioallantoic fusion, and yolk sac vasculogenesis, and its absence leads to early embryonic lethality. In this study, we sought to identify the gene regulatory networks underlying the development of these structures and how they are dysregulated in the absence of CDX2. We additionally determined the extent to which this network is CDX2 dose-dependent, motivated by studies showing both a correlation between CDX2 expression levels and the development of embryonic mesoderm, including elongation of the tailbud as well as studies showing that genes in related pathways such as TBXT can dose-dependently regulate allantois development.

Through this work, we demonstrate that relative to WT, both CDX2-Het and CDX2-KO have altered expression of both canonical and non-canonical WNT and BMP signaling pathways, in addition to changes in cell adhesions and cytoskeletal regulators. Over 2/3 of the genes misregulated in CDX2-Het are also misregulated in CDX2-KO, revealing a striking dose dependence in gene expression downstream of CDX2. Is it therefore likely that the changes in gene expression observed in CDX2-Het correlate with the impaired or delayed development of the allantois and associated vasculature observed *in vivo*. However, this impairment is likely not sufficient to prevent chorioallantoic fusion and cause lethality, as evidenced by the viability of heterozygous CDX2 mice (Chawengsaksophak *et al.*, 2004). This prediction agrees with the phenotype observed in mice heterozygous for TBXT.

Even with these changes in gene expression, we observe fewer differentially accessible chromatin peaks in CDX2-Het than in CDX2-KO when compared to WT. The peaks identified in CDX2-KO are heavily enriched for CDX motifs, suggesting that CDX2 binding drives their accessibility. Additionally, we observe that CDX2-Het binds to CDX motifs at a reduced frequency compared to WT, but CDX2-Het and WT share very similar chromatin accessibility profiles. This

observation suggests that reduced CDX2 expression is sufficient to drive or sustain a WT-like chromatin accessibility profile, though downstream gene expression is compromised. This perhaps indicates that open chromatin is not sufficient to drive the expression of downstream genes and that a threshold amount of CDX2 must be bound to regulatory elements for expression to reach WT levels. It is also possible that a balance between CDX2 and its co-factors is required to activate downstream gene expression, and this balance is not achieved in CDX2-Het despite the chromatin being accessible.

We next demonstrate that CDX2 dose slightly influences HOX gene expression patterns and identify HOXB paralog as being uniquely regulated in our dataset. The detection of differently expressed genes within the HOXB cluster perhaps reflects a temporal delay in expression between different HOX paralogs, or it may be a reflection of a HOXB-specific role in the extraembryonic mesoderm. Regardless, we observe a reduction in central HOX gene expression in WT relative to both CDX2-Het and CDX2-KO, again emphasizing how CDX2 dose influences the expression of its downstream targets. Additionally, we observe the slight upregulation of *HOXB2* in CDX2-Het and CDX2-KO, reminiscent of the anterior homeotic shifts seen *in vivo* (van den Akker *et al.*, 2002). We believe this expression pattern is related to the idea of "posterior prevalence" where more posterior HOX genes suppress the expression of more anterior HOX genes (Krumlauf, 1993; Duboule and Morata, 1994; Yekta, Tabin and Bartel, 2008). Because the mutants have lower central HOX expression, more anterior HOX genes are not as severely suppressed, leading them to have higher expression relative to WT.

We next look at how communication between ligands and receptors of various clusters is influenced by CDX2 dose using CellChat. Through this analysis, we isolate the ncWNT pathway as uniquely misregulated in both CDX2-Het and CDX2-KO relative to WT. Changes in the ncWNT signaling pathway are largely restricted to signals from and within the three ExeM clusters and specifically reflect the reduction in *WNT5B* expression in the ExeM-late cluster of the mutant cell lines. *WNT5B* influences both canonical WNT signaling and VEGF signaling, both established

regulators of early embryonic vasculogenesis *in vivo*, and while further validation is required these results implicate *WNT5B* expression could be critical for early morphogenesis of extraembryonic mesoderm-derived structures.

Finally, we compare how the CDX2-KO and TBXT-KO influence extraembryonic mesoderm gene expression and isolate DEGs that are shared between the two datasets relative to WT. These genes are reflective of adhesions and cytoskeletal dynamics, and several converge on the VEGF signaling pathway, suggesting that CDX2 and TBXT both disrupt pathways that have been shown *in vivo* to be critical for early vasculogenesis. We hypothesize that these pathways may contribute to malformations in the development of the allantois, preventing chorioallantoic fusion and subsequent placental development.

Taken together, these results solidify the dose-dependent role of CDX2 in the formation of extraembryonic structures crucial for early embryogenesis. Understanding the genetic patterns underlying the development of these structures is critical for our foundational understanding of how gene dosage influences morphogenesis, chromatin conformation, and downstream gene expression, with the ultimate goal of better understanding extraembryonic mesoderm and subsequent placental development.

MAIN FIGURES

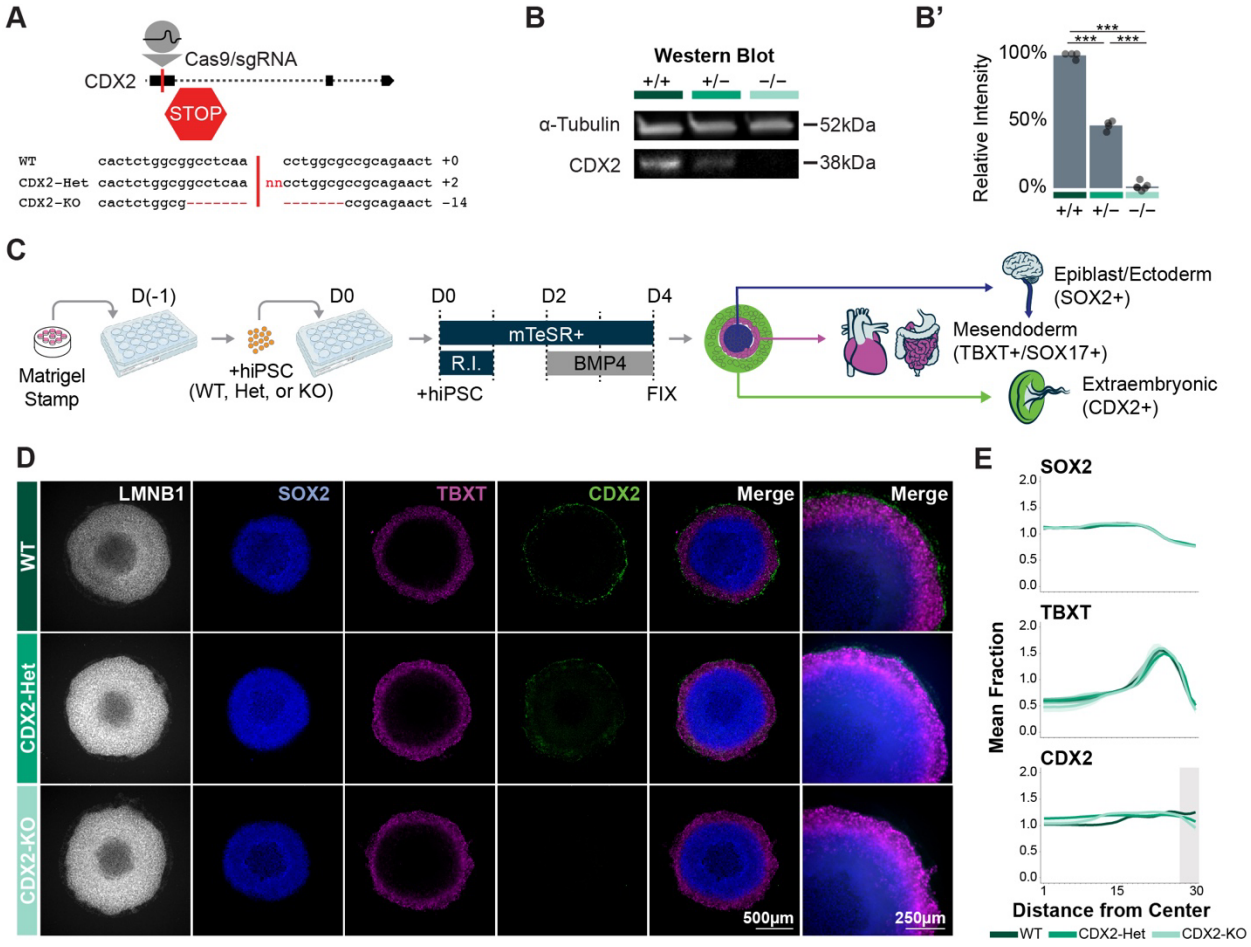


Figure 3.1: Generation and validation of the CDX2 allelic series. (**A**) Schematic of CDX2 locus and sgRNA target in the first exon. Indel generated in CDX2-Het (+2bp) or CDX2-KO (-14bp). (**B**) Western blot and (**B'**) quantification CDX2 western blot signal intensity in WT, CDX2-Het, and CDX2-KO after 48hr BMP4 exposure. n=4 replicates/genotype. (**C**) Schematic of differentiation protocol. (**D**) Immunofluorescence for SOX2, TBXT, and CDX2 in WT, CDX2-Het, and CDX2-KO 2D gastruloids. Nuclei labeled with LMNB1. (**E**) Quantification of the mean fraction of fluorescence intensity across gastruloids of each genotype. 1 = center of gastruloid, 30 = outer gastruloid. n = 3-7 gastruloids/genotype. Gray bar indicates region of interest.

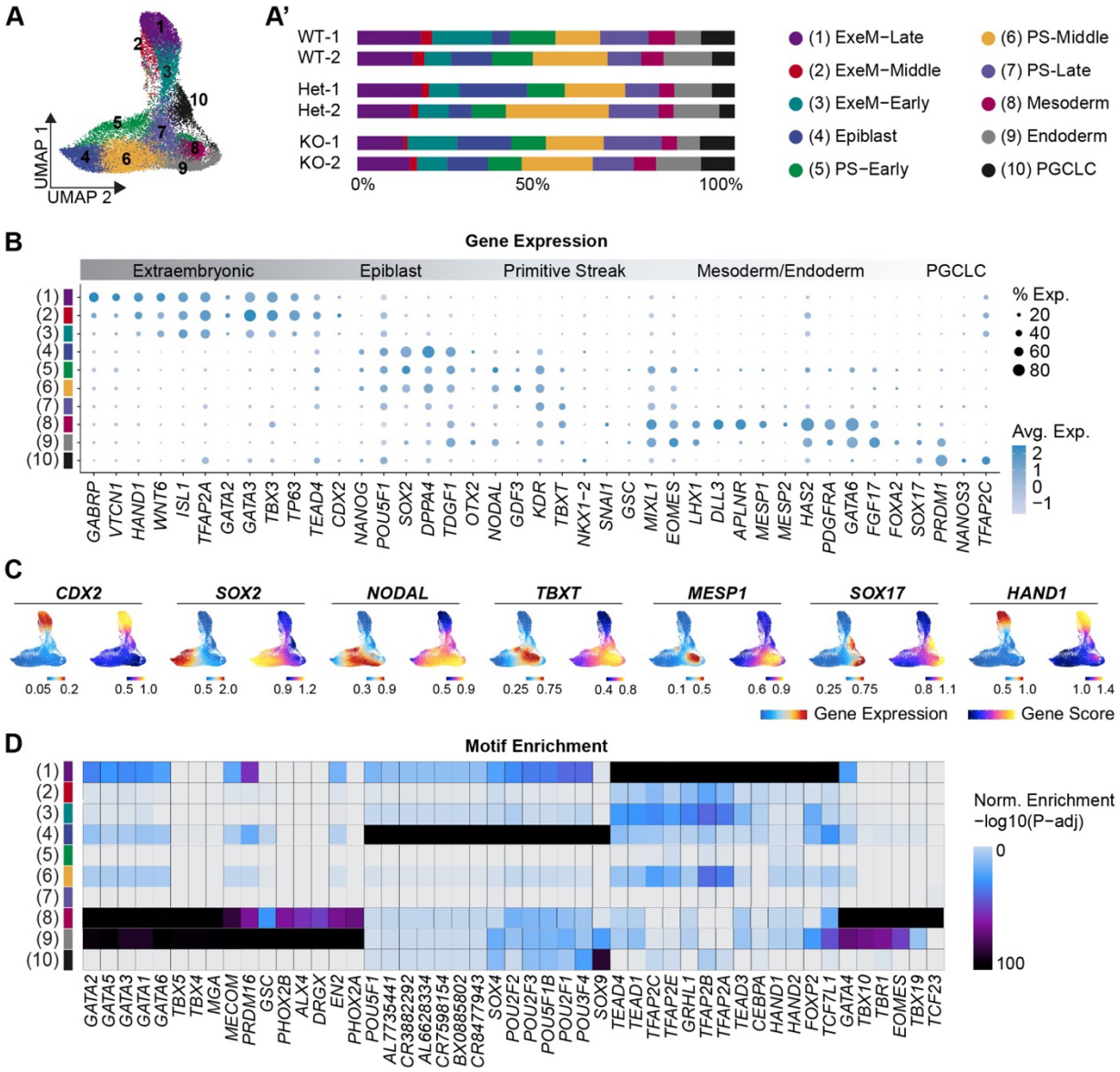


Figure 3.2: Lineage emergence is minimally altered in 2D gastruloids of varying CDX2 dose. (**A**) UMAP reflecting 10 clusters comprised of WT, CDX2-Het, and CDX2-KO cells from 2D gastruloids. (**A'**) Proportion of cells of each sample in each cluster. (**B**) Dot plot depicting key lineage markers across all 10 clusters. (**C**) Feature Plots reflecting the gene expression (snRNA-seq) or gene score (snATAC-seq) of key lineage markers. (**D**) Heatmap of the top motifs enriched

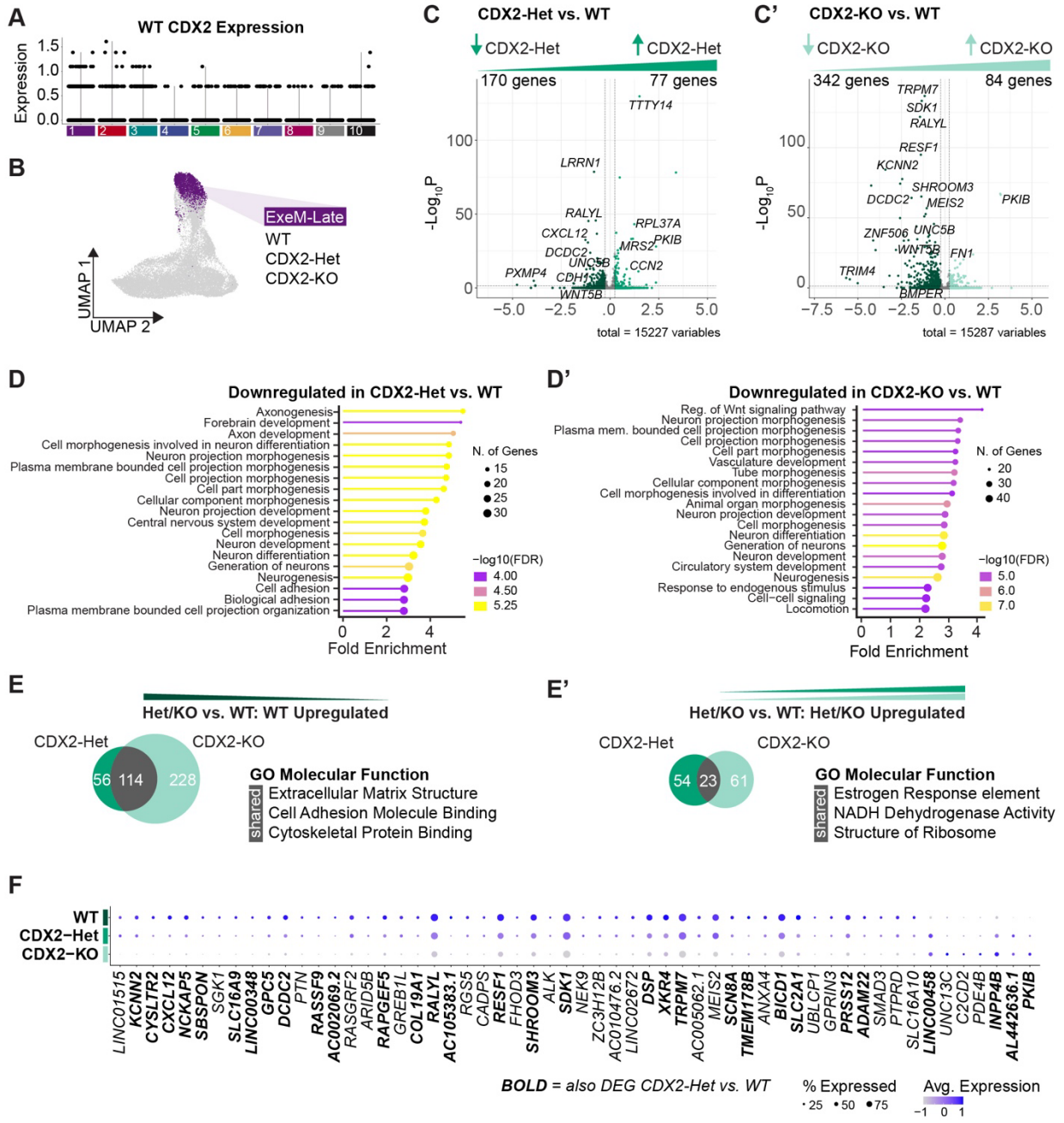


Figure 3.3: CDX2 dose-dependently influences extraembryonic mesoderm gene expression. (A) Violin plot of CDX2 expression across clusters in WT cells. **(B**) ExeM-Late cluster isolated for genotype-specific analyses. (**C**) Volcano plot of differentially expressed genes comparing WT and CDX2-Het or (**C'**) WT and CDX2-KO (abs(Log2FC) > 0.25, Adj. P < 0.05) within the ExeM-Late cluster. (**D**) GO biological process enrichment for genes downregulated in CDX2-Het relative to WT or (**D'**) CDX2-KO relative to WT (abs(Log2FC) > 0.25, Adj. P < 0.05, ShinyGO FDR < 0.05) within the ExeM-Late cluster.

(Figure caption continued on the next page.)

(Figure caption continued from the previous page.)

(**E**) Venn diagram of genes upregulated or (**E'**) downregulated in WT relative to CDX2-Het (left) or CDX2-KO (right). Select results from GO molecular function enrichment for overlapping genes. (**F**) Dotplot of the top differentially expressed genes between WT and CDX2-KO (p-adj < 0.05, abs(Log2FC >1, pct cells > 20%) within the ExeM-Late cluster. Bold text indicates genes with significantly differential expression comparing both CDX2-Het vs. WT and CDX2-KO vs. WT.

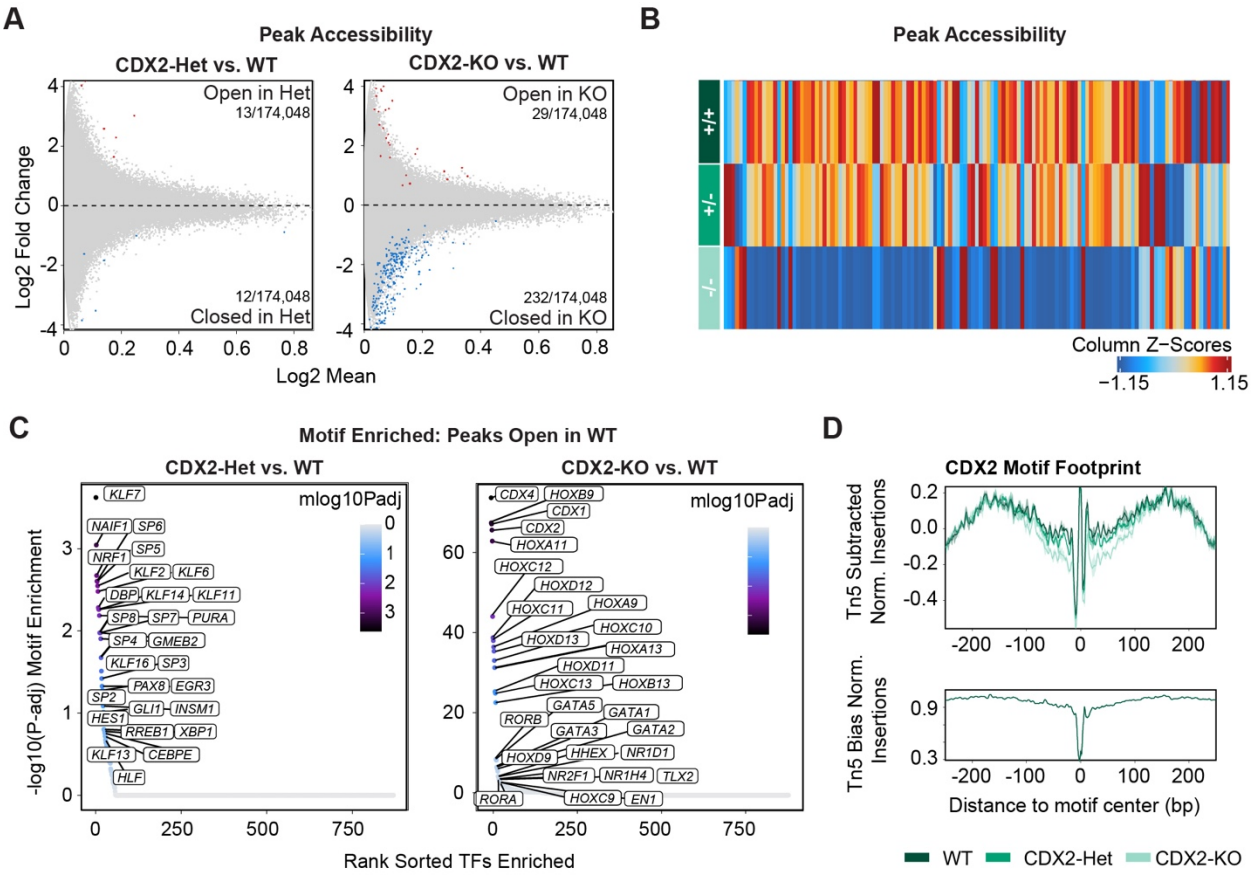


Figure 3.4: CDX2 expression influences chromatin organization at regions containing CDX motifs (A) Differential peak accessibility comparing CDX2-Het and WT (left) or CDX2-KO and WT (right) within the ExeM-Late cluster (FDR < 0.1 & abs(Log2FC) > 0.5). Red = regions more accessible in mutant, blue = regions more accessible in WT. (**B**) Heatmap of Differentially accessible peaks across WT, CDX2-Het, and CDX2-KO within the ExeM-Late cluster (**C**) Enriched motifs detected in peaks uniquely accessible in WT relative to CDX2-Het (left) and CDX2-KO (right). (**D**) Footprint for the CDX2 motif across WT (dark green), CDX2-Het (middle green), and CDX2-KO (light green).

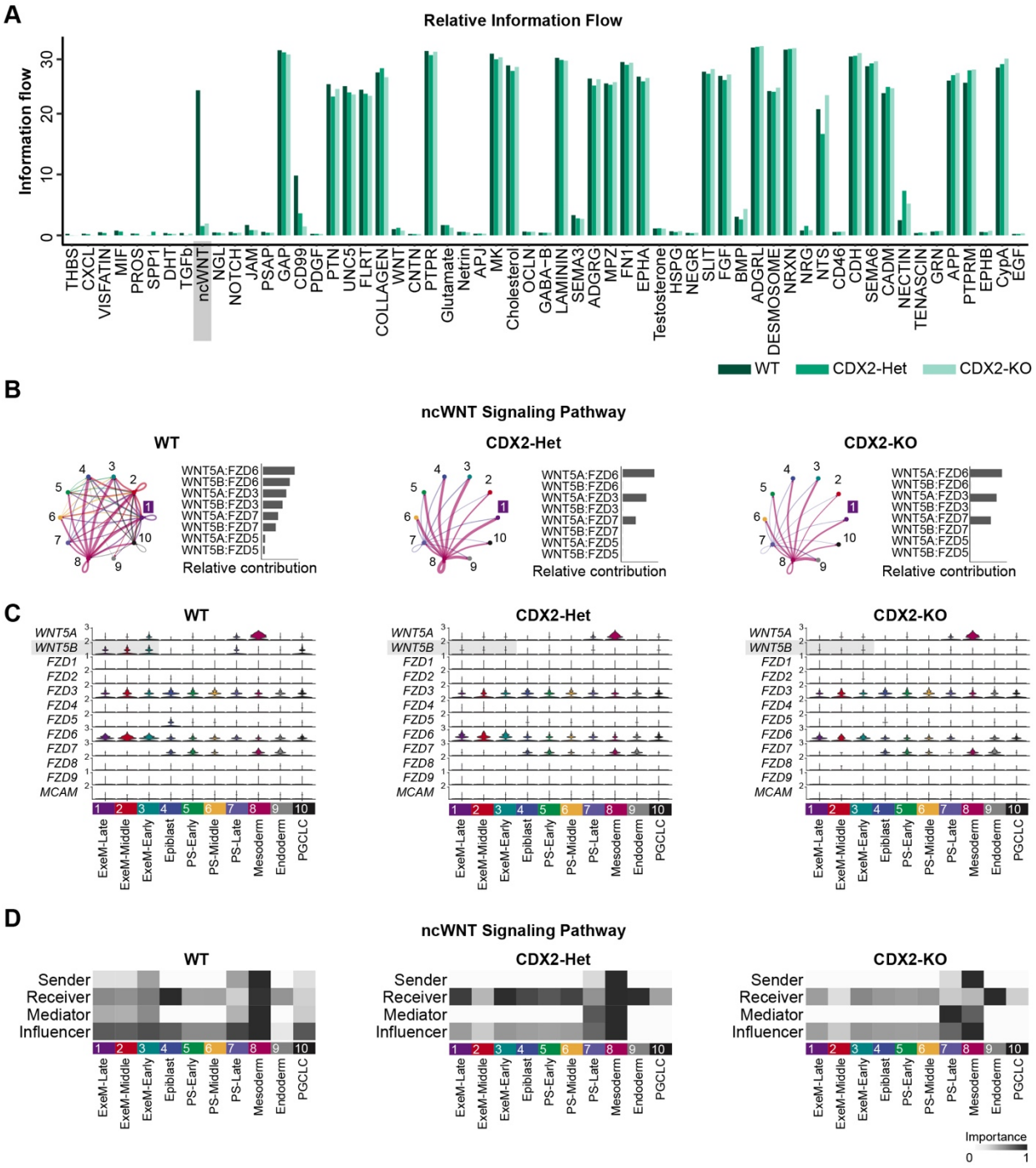


Figure 3.5: Cellchat reveals the CDX2-dose-dependent regulation of ncWNT signaling pathway. (**A**) Information flow (arbitrary units; A. U.) of various pathways across all clusters within WT, CDX2-Het, or CDX2-KO populations. (**B**) Circle plots (left) and bar plots (right) visualizing ligand-receptor communication across all clusters. The number correlates with cluster identity, and "1" indicates ExeM-Late. Line thickness corresponds with the strength of the predicted communication. (**C**) Violin plot of gene expression for each ligand or receptor comprising the ncWNT signaling pathway. (**D**) Heatmap showing the communication dynamics for the ncWNT signaling pathway in each cluster.

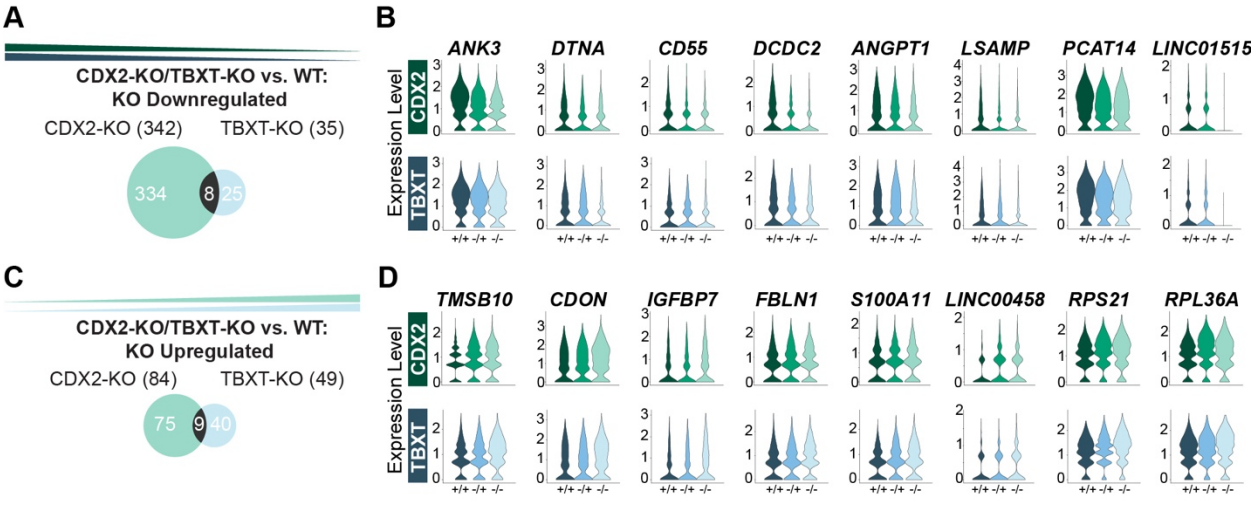


Figure 3.6. TBXT and CDX2 mutants both misregulate genes associated with the VEGF signaling pathway. (**A**) Venn diagram illustrating overlapping genes downregulated or (**C**) upregulated in CDX2-KO (left) or TBXT-KO (right) relative to WT in the ExeM-Late cluster. (**B**) Violin plots of gene expression for overlapping genes identified in (**A**). MT-ND3 shown in Fig. S9. (**D**) Violin plots of gene expression for overlapping genes identified in (**C**).

SUPPLEMENTAL FIGURES

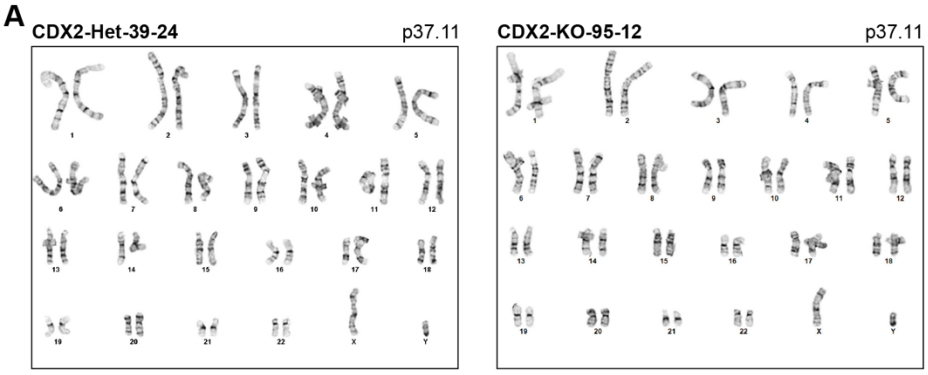


Figure S3.1: Karyotyping results from CDX2-Het and CDX2-KO isogenic lines. For WT, see Chapter 2.

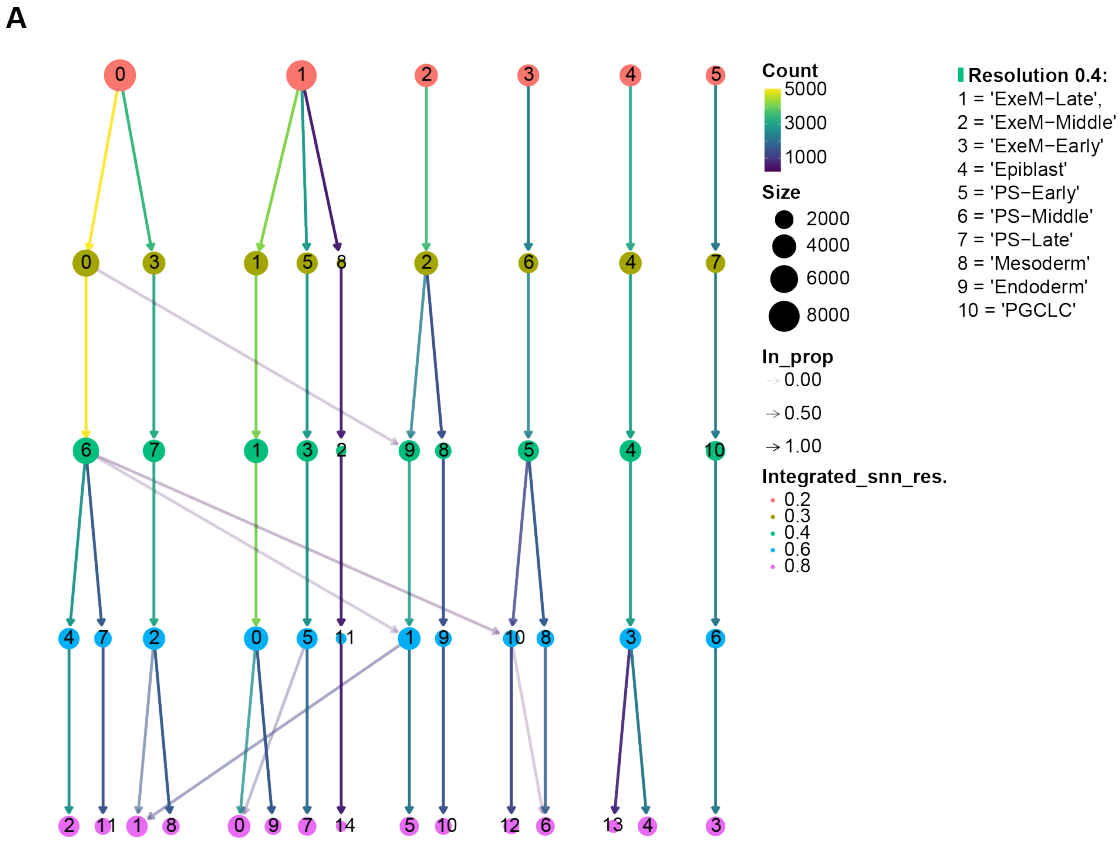


Figure S3.2: Clustree analysis of clusters at 0.2, 0.3, 0.4, 0.6, and 0.8 resolution. A resolution of 0.4 was used for subsequent analyses.

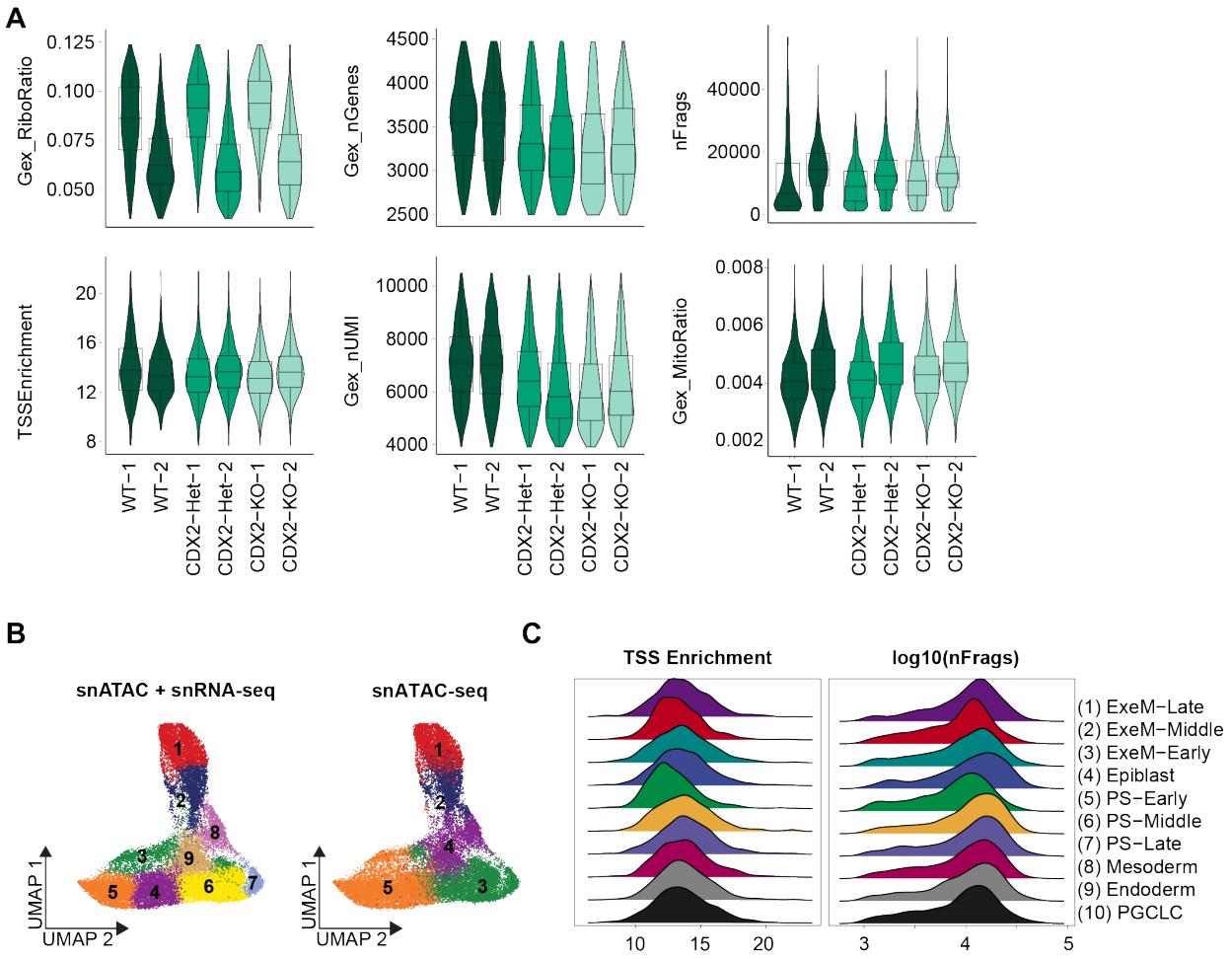


Figure S3.3: Quality control parameters for CDX2 snRNA-seq and snATAC-seq data A) Quality control parameters after filtration separated by sample. B) TSS enrichment and log10(nFrags) separated by cluster.

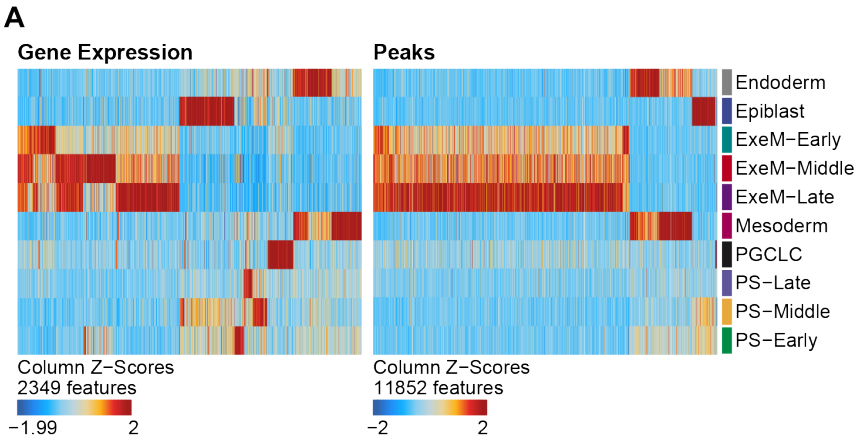


Figure S3.4: Gene expression and peak accessibility separated by cluster. The order of rows is based on hierarchical clustering.

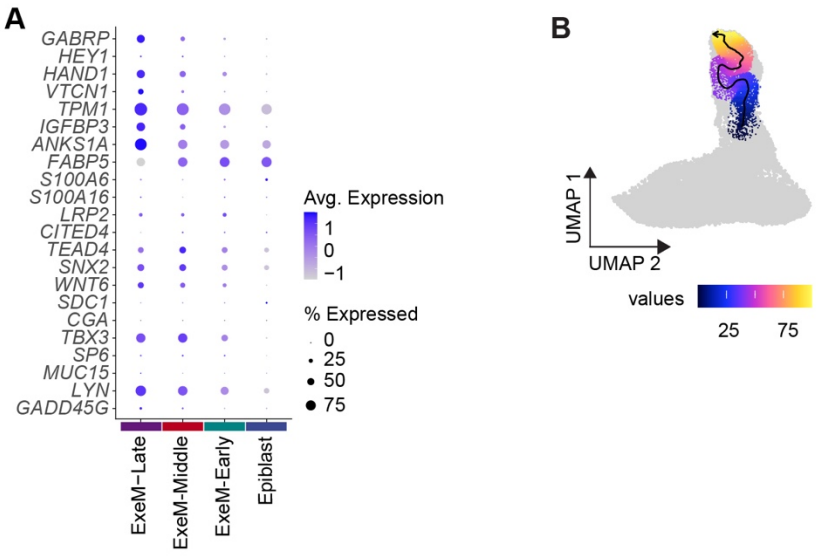


Figure S3.5: Extraembryonic mesoderm and trophectoderm marker expression across clusters A) Dotplot reflecting key markers of Amnion and trophectoderm for the extraembryonic and epiblast clusters (C1-C4). B) Trajectory analysis of ExeM-Early, ExeM-Middle, and ExeM-Late (C3-C1).

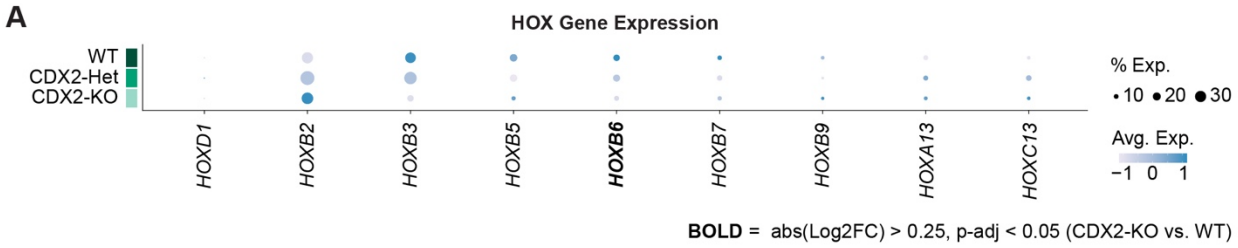


Figure S3.6: CDX2 dose-dependently influences downstream HOX expression. A) Gene expression of detectable HOX genes in the ExeM-Late cluster across WT, CDX2-Het, and CDX2-KO. BOLD text indicates Log2FC > 0.25 and p-adj < 0.05 (CDX2-KO vs. WT)

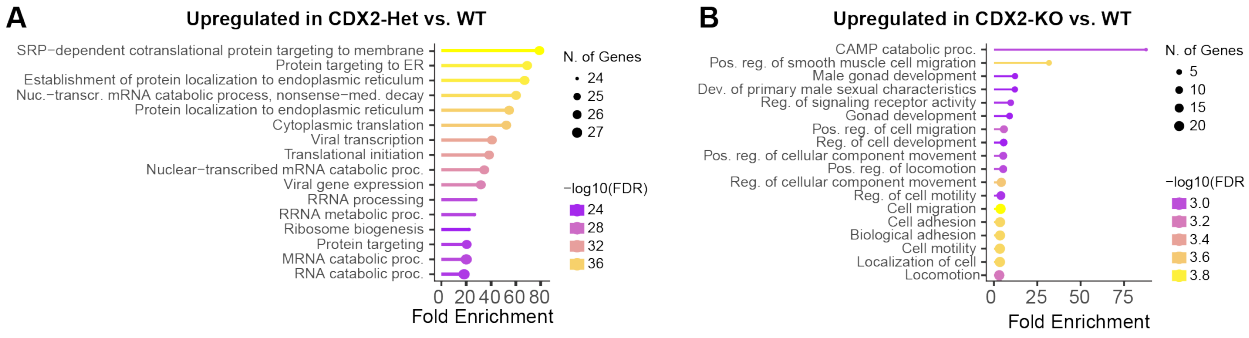


Figure S3.7: ShinyGO analysis of CDX2 allelic series A) ShinyGO analysis of genes upregulated in CDX2-Het vs. WT or B) genes upregulated in CDX2-KO vs. WT (right) (Log2FC >-0.25, Adj. P < 0.05).

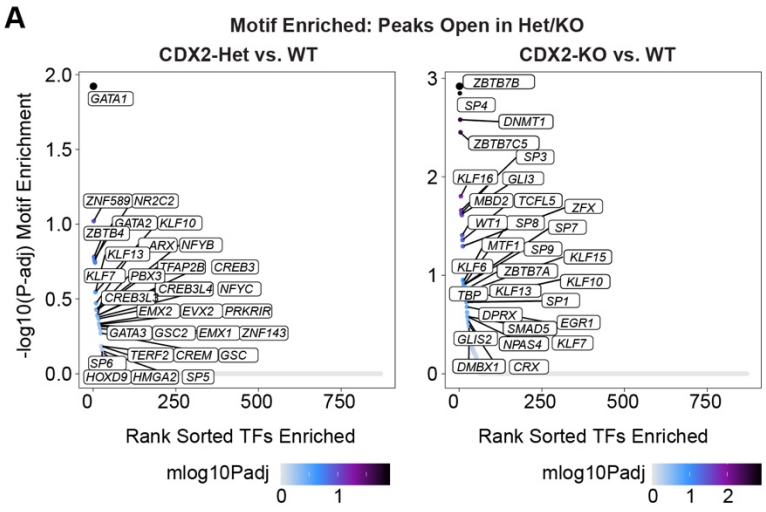


Figure S3.8: Motif enrichment in ExeM-Late across CDX2 allelic series A) Motifs enriched in DARs more accessible in CDX2-Het relative to WT or B) CDX2-KO relative to WT.

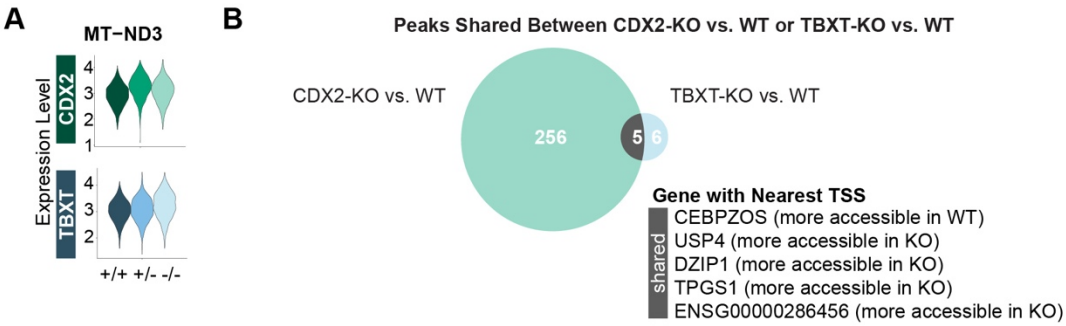


Figure S3.9: CDX2-KO and TBXT-KO DEG and Peak comparisons. A**)** Gene Expression within the extraembryonic mesoderm cluster for MT-ND3 across CDX2 dosage series (top) or TBXT dosage series (bottom). **B**) Venn diagram of DARs shared between CDX2-KO vs. WT and TBXT-KO vs. WT (abs(Log2FC) > 0.5, FDR <0.1)

Chapter 4: Methods

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to Benoit G. Bruneau (benoit.bruneau@gladstone.ucsf.edu).

Data availability

snATAC-seq and snRNA-seq data have been deposited in GEO under the accession number GSE245998 (TBXT) and GSE251813(CDX2). Analysis scripts used to generate figure panels are available from the authors upon request.

METHOD DETAILS

Cell Lines

All work with human induced pluripotent stem cells (hiPSCs) was approved by the University of California, San Francisco Human Gamete, Embryo, and Stem Cell Research (GESCR) Committee. Human iPS cells harboring genome-edited indel mutations for *TBXT* (*TBXT-Het, TBXT-KO*) or *CDX2* (*CDX2-Het, CDX2-KO*) were generated for this study and derived from the Allen Institute WTC11-LaminB parental cell line (AICS-0013 cl.210). TBXT-Het-6 was derived from the Allen Institute WTC11-LMNB1- mTagRFP-T parental cell line (AICS-0034 cl.62). The WT line used for all studies was derived from a WTC11-LaminB subclone that was exposed to the TBXT sgRNA (see "TBXT Allelic series generation," below) but remained unedited. Throughout this dissertation, "TBXT-Het" refers to "TBXT-Het-35" unless otherwise noted. All cell lines were karyotyped by Cell Line Genetics and reported to be karyotypically normal. Additionally, all cell lines tested negative for mycoplasma using a MycoAlert Mycoplasma Detection Kit (Lonza).

Maintenance of iPS Cells

Human iPSCs were cultured on growth factor-reduced (GFR) Matrigel (Corning Life Sciences) and fed at minimum every other day with mTeSR-Plus medium (STEMCELL Technologies) (Ludwig et al., 2006). Cells were passaged by dissociation with Accutase (STEMCELL Technologies) and re-seeded in mTeSR-Plus medium supplemented with the small molecule Rho-associated coiled-coil kinase (ROCK) inhibitor Y276932 (10 μM; Selleckchem) (Park et al., 2015) at a seeding density of 12,000 cells per cm^2. After 24 hours, cells were maintained in mTeSR-Plus media until 80% confluent.

Allelic Series Generation

To generate the TBXT allelic series we first lipofected parental cells with 125ng sgRNA and 500ng Cas9 protein according to the Lipofectamine Stem Transfection Reagent Protocol (Invitrogen). TBXT-Het-35 and TBXT-KO were derived from the Allen Institute WTC11-LMNB1-GFP parental cell line (AICS-0013 cl.210) while TBXT-Het-6 was derived from the Allen Institute WTC11-LMNB1-mTagRFP-T parental cell line (AICS-0034 cl.62). The human TBXT sgRNA (CAGAGCGCGAACUGCGCGUG) was a gift from Jacob Hanna (Addgene plasmid #59726; http://n2t.net/addgene:59726; RRID: Addgene_59726) and targeted the first exon of the TBXT gene.

To generate the CDX2 allelic series we lipofected WTC11-LaminB cells (AICS-0013 cl.210) with 125ng sgRNA and 500ng Cas9 protein according to the Lipofectamine Stem Transfection Reagent Protocol (Invitrogen). Two CDX2 sgRNAs that both target the first exon of CDX2 gene were co-lipofected using 62.5ng each: (CDX2-A: CCUAGCUCCGUGCGCCACUC and CDX2-B: AGUUCUGCGGCGCCAGGUUG).

After recovery for 48 hours in mTeSR-Plus supplemented with ROCK inhibitor, lipofected cells were dissociated using Accutase and passaged into a GFR-Matrigel coated 10cm dish, where they were expanded for 24hr in mTeSR-Plus with with the small molecule Rho-associated coiled-coil kinase (ROCK) inhibitor. Media was replaced with mTeSR-Plus without ROCK inhibitor and cells continued to grow another 2-4 days before the manual selection of 20-30 single colonies into individual wells of a 96-well plate. After the expansion of the clonal populations for 5-10 days, cells were passaged into a new 96-well plate at a 1:5 dilution ratio in mTeSR-Plus supplemented with ROCK inhibitor, and the remaining cells were used for genotyping.

For screening of *TBXT* exon 1 non-homologous end-joining (NHEJ) mutations, DNA was isolated using QuickExtract DNA lysis solution (Epicentre #QE0905T), and genomic DNA flanking the targeted sequence was amplified by PCR (For1: gaaggtggatctcaggtagcgagtctgg and Rev1: cagcaggaaggagtacatggcgttgg) and sequenced using Rev1. For screening of *CDX2* exon 1

non-homologous end-joining (NHEJ) mutations, DNA was isolated using QuickExtract DNA lysis solution (Epicentre #QE0905T), and genomic DNA flanking the targeted sequence was amplified by PCR (For2: cctcgacgtctccaaccattg and Rev2: gcctctgcttaccttggctg) and sequenced using Rev2.

Synthego ICE analysis was employed to quantify editing efficiency and identify clones with heterozygous (45-55% KO score) or homozygous null (>90% KO score) mutations (Table 4.1). To eliminate the possibility of a heterozygous line being a mixed population of wildtype and homozygous null alleles, a minimum of 8 subclones of the prospective heterozygous clones were isolated and Sanger sequenced as before, and all subclones were confirmed to contain identical genotypes. After sequencing confirmation of respective genotypes, karyotypically normal cells from each hiPSC line were expanded for subsequent studies.

Table 4.1: Indel frequency of clonal or subclonal cell populations

Genotype	Sequence	Indel	Indel %	R^2
WT (TBXT)	gagaagggcgaccccacag| agcgcgaactgcgcgtgggcctgga	+0bp	0%	0.99
TBXT-Het	gagaagggcgacccc----| ----------gcgtgggcctgga	-16bp	45-51%	0.84
TBXT-KO	gagaagggcgaccccacag|n agcgcgaactgcgcgtgggcctgga	+1bp	100%	0.97
WT (CDX2)	cactctggcggcctcaa| cctggcgccgcagaact	+0bp	0%	0.99
CDX2-Het	cactctggcggcctcaa| nn cctggcgccgcagaact	+2bp	47-50%	0.92
CDX2-KO	cactctggcg-------| -------ccgcagaact	-14bp	100%	1.0

PDMS stamp fabrication

Standard photolithography methods were used to fabricate a master template, which was provided as a gift from PengFei Zhang and the Abate lab at the University of California, San

Francisco (Alom Ruiz and Chen, 2007; Théry and Piel, 2009; Minn *et al.*, 2020). The photoresist master was then coated with a layer of chlorotrimethylsilane in the vacuum for 30 min. Polydimethylsiloxane (PDMS) and its curing agent, Sylgard 184, (Dow Corning, Midland, MI) were mixed in a 10:1 ratio, degassed, poured over the top of the master, and cured at 60°C overnight, after which the PDMS layer was peeled off to be used as a stamp in micro-contact printing.

Microcontact Printing

PDMS stamps (each containing 12 x 1000uM circles) were sterilized by washing in a 70% ethanol solution and dried in a laminar flow hood. Growth factor reduced Matrigel (Corning) was diluted in DMEM/F-12 (Gibco) at 1:100 dilution and incubated on the stamps to cover the entire surface of the feature side at 37°C for 1 hour. The Matrigel solution was then aspirated off the stamps, which were air-dried. Using tweezers, the Matrigel-coated surface of stamps was brought in contact with glass or plastic substrate, usually a glass 24-well plate or removable 3-chamber slide (Ibidi), and incubated on the substrate for 1hr at 37°C. The stamps were then removed and rinsed in ethanol for future use. Matrigel-printed substrates were incubated with 1% Bovine Serum Albumin (Sigma Aldrich) in DPBS-/- at room temperature for 1 hour before being stored in DPBS-/- solution at 4°C for up to 2 weeks.

Confined 2D Gastruloid Differentiation

hiPSCs were dissociated with Accutase and resuspended in mTeSR-Plus supplemented with ROCK inhibitor, as previously described. Cells were then seeded onto a stamped well at a concentration of approximately 750 cells/mm^2. Cells were incubated at 37°C for 3 hours before the well was rinsed 1x with DPBS and given fresh mTeSR-Plus supplemented with ROCK inhibitor. Approximately 24 hours post-seeding, media was exchanged for mTeSR-Plus. After another 24 hours or upon confluency of the stamped colony, media was exchanged for mTeSR-

Plus supplemented with BMP4 (50ng/mL). Colonies were allowed to differentiate in the presence of BMP4 for 48 hours before being processed for downstream analyses.

Mesoderm Induction Media Differentiation

hiPSCs were dissociated with Accutase and resuspended in mTeSR-Plus supplemented with ROCK inhibitor, as previously described. Cells were then seeded as a monolayer at a concentration of approximately 500 cells/mm^2. Cells were incubated at 37°C for 24 hours before being rinsed 1x with DPBS and given fresh mTeSR-Plus without ROCK inhibitor. Approximately 24 hours later, media was exchanged for Mesoderm Induction Media (MIM). Colonies were allowed to differentiate for 24-72 hours, with MIM being exchanged daily, before use in downstream analyses.

Western Blot

Cells of each genotype were induced to form TBXT+ mesoderm with either MIM or 4uM CHIR99021 in mTeSR+ for 48 hours prior to protein isolation. Cells were washed twice with ice-cold PBS and lysed in RIPA lysis buffer (Fisher Scientific; A32965). Three replicate wells were pooled for each genotype for each differentiation condition. The protein concentration was determined using the Pierce BCA Protein Assay Kit (Life Technologies, 23227) and quantified on a SpectraMax i3 Multi-Mode Platform (Molecular Devices). following the manufacturer's instructions. Protein (~20-40 µg) was transferred to the membrane using the Trans-Blot Turbo Transfer System (Biorad; 1704157). The membrane was then blocked overnight at 4°C using Intercept TBS Blocking Buffer (Li-COR; 927-7000ß1). Primary antibodies TBXT (AF2085; 1000x; Gt) or CDX2 (12306; 1000x; Rb) and either GAPDH (ab9485; 1000x; Rb), α-Tubulin (T5168; 1000x; Ms) or β-actin (ab8226, 1000x, Ms) were diluted in Intercept T20 (TBS) Antibody dilution buffer (Li-COR) at a 1:1000 ratio and incubated with the membrane overnight at 4°C.

The next morning, membranes were washed in 1x TBS-T and incubated for 1 hour at RT in the dark with species-specific secondary antibodies (Rb-680; 926-68071; Gt-680 925-68074, Ms-800 926-32212 Gt-800; 926-32214) (VWR) at 1:10,000. Membranes were subsequently washed and developed using the BioRad ChemiDoc MP. Protein levels were quantified using ImageJ by first subtracting the intensity of a blank ROI from the experimental ROI, and then calculating a normalization factor by dividing the observed housekeeping intensity by the highest observed housekeeping intensity. The observed experimental signal was then divided by the lane normalization factor to generate a normalized experimental signal. Each lane from the same blot was then converted to a percentage of the highest WT normalized experimental signal on that blot.

Immunofluorescence

hiPSCs were rinsed with PBS 1X, fixed in 4% paraformaldehyde (VWR) for 15-20 minutes, and subsequently washed 3X with PBS. The fixed cells were permeabilized and blocked in a buffer comprised of 0.3% Triton X-100 (Sigma Aldrich) and 5% normal donkey serum in PBS for one hour, and then incubated with primary antibodies diluted in antibody dilution buffer (0.3% Triton, 1% BSA in PBS) overnight (Table 4.2). The following day, samples were washed 3X with PBS and incubated with secondary antibodies in antibody dilution buffer at room temperature for 2 hours. Secondary antibodies used were conjugated with Alexa 405, Alexa 555, or Alexa 647 (Life Technologies) at a dilution of 1:400. Cells were imaged at 10x, 20x, or 40x magnification on an inverted AxioObserver Z1 (Zeiss) with an ORCA-Flash4.0 digital CMOS camera (Hamamatsu).

Table 4.2: Antibodies used in this study

Antibody	Catalog	Supplier	Dilution	Host
CDH1	ab1416	AbCam	1:400	Ms
CDH2	ab76057	AbCam	1:200	Rb
CDX2	ab157524	AbCam	1:500	Ms
EOMES	MAB6166	R&D Systems	1:400	Ms
F-ACTIN	BS-1571R	Bioss	1:100	Rb

Antibody	Catalog	Supplier	Dilution	Host
FN1	ab281574	AbCam	1:200	Ms
LEFTY	PA519507	Invitrogen	1:1000	Rb
NODAL	PA5-23084	ThermoFisher	1:100	Rb
pSMAD1/5	3579	Cell Signaling	1:200	Rb
SNAI1	ab180714	AbCam	1:400	Rb
SOX17	AF1924	R&D Systems	1:200	Gt
SOX2	3579s	Cell Signaling	1:200	Rb
TBX6	AF4744	R&D Systems	1:200	Gt
TBXT	AF2085	R&D Systems	1:400	Gt
WNT3A	ab169175	AbCam	1:100	Ms
ZO-1	33-9100	Invitrogen	1:200	Ms
ZO-1	HPA001636	Sigma	1:200	Rb
β-Catenin	610154	BD Transduction Labs	1:800	Ms

Scratch Assay

Approximately 50k hiPSCs of each genotype were seeded into each well of a 96-well plate following the MIM induction protocol previously described. 24 hours after the addition of MIM, a scratch was made in confluent wells manually by using a p200 pipette tip. Using ZenPro software, cells were then imaged every 12 minutes across 24 hours in Brightfield at 10x magnification on an inverted AxioObserver Z1 (Zeiss) with an ORCA-Flash4.0 digital CMOS camera (Hamamatsu) at 37°C with 5% CO^2. Images of cells were computationally segmented, and the area occupied by cells was calculated using CellProfiler at the 24hr and 48hr time points relative to initial MIM addition (0hr and 24hr of live imaging). The area occupied at the final time point was subtracted from the area at the initial time point to yield the change in area. Wells that were not confluent at the time of the scratch or wells in which the basement membrane had been removed by the scratch were omitted from the final dataset.

Fluorescent *In situ* hybridization

PDMS stamps coated in GFR-Matrigel were applied to 3-well chamber slides with removable silicone chamber walls (Ibidi, 80381) and gastruloids were generated as described

previously. Colonies were then fixed in 4% paraformaldehyde for 15-30 minutes, rinsed in PBS, and dehydrated according to the RNAscope cultured Adherent Cell Sample Preparation for RNA Multiplex Fluorescent v2 Assay (Advanced Cell Diagnostics). Slides were stored in 100% ethanol at −20°C short term until initiation of the *in situ* hybridization protocol. The *RNAscope* protocol was then performed as outlined in User Manual 323100-USM. Catalog numbers for ACDBio RNAscope probes used in this study include FGF17 (1148351-C1), RSPO3 (413701-C2), MESP1(849231), CYP26A1 (487741), and WNT5A (604921). Colonies were imaged using the Olympus Fluoview FV3000 Confocal Microscope or the Nikon C2 laser scanning confocal microscope equipped with a Prime 95B 25mm sCMOS camera in collaboration with the UCSF Nikon Imaging Core and Gladstone Microscopy Core.

Cell Harvesting for Single Nuclei Multiome ATAC + RNA Sequencing

Each of the WT, TBXT or CDX2 genotypes was differentiated, harvested, and prepared at the same time for each of the two biological replicates. Therefore, each biological replicate represents an experimental batch. For each sample within the batch, 12 micropatterns were differentiated within each well of a 24-well plate and cells from all wells on a plate were pooled, yielding a cell suspension comprising approximately 288 colonies per sample. Nuclei were isolated and ~9,000-19,000 nuclei/sample were transposed and loaded onto a 10x Chromium Chip J to generate gel bead-in emulsions (GEMs) following the 10x Chromium Next GEM Single Cell Multiome ATAC and Gene Expression Kit (10x Genomics, CG000338). GEMs were processed to produce ATAC and gene expression libraries in collaboration with the Gladstone Genomics Core. Deep sequencing was performed on the NovaSeq 6000 S4 200 cycle flow cell for a read depth of >15k reads per cell. WT-1 (eb01)= 12072 nuclei, 20,556.09 reads/nucleus. TBXT-KO-1 (eb02) = 10242 nuclei, 30,683.65 reads/nucleus.TBXT-Het-1 (eb03) = 8819 nuclei, 34,179.63 reads/nucleus. CDX2-KO (eb04) = 18239 nuclei, 15,288.63 reads/nucleus. CDX2-Het-1 (eb05) = 18806 nuclei, 19,435.86 reads/nuclei. WT-2 (eb06) = 10903 nuclei, 26,003.63

reads/nucleus. TBXT-KO-2 (eb07) = 19,325 nuclei, 12,453.80 reads/nucleus.TBXT-Het-2 (eb08) = 8,863 nuclei, 29,849.86 reads/nucleus. CDX2-KO-2 (eb09) = 12314 nuclei, 20,150.10 reads/nucleus. CDX2-Het-2 (eb10) = 11602 nuclei, 18,929.24 reads/nucleus.

Data Processing Using CellRanger-Arc

All ATAC and GEX datasets were processed using CellRanger-Arc 2.0.0. FASTQ files were generated using the mkfastq function, and reads were aligned to the hg38 reference genome (version 2.0.0).

Seurat Analysis - TBXT

Outputs from the CellRanger-Arc count pipeline were analyzed using the Seurat package (version 4.3.0)(Satija *et al.*, 2015; Butler *et al.*, 2018; Stuart *et al.*, 2019) in R (v4.2.0). Quality control filtering included the removal of outliers due to the number of features/genes (nFeature_RNA > 2500 & nFeature_RNA < 4500, nCount_RNA > 5000 & nCount_RNA < 12,500, mitochondrial percentage > 5% and mitochondrial percentage < 20%, and ribosomal percentage > 3% and ribosomal percentage < 15%). Cell cycle scores were added using the function CellCycleScoring. ScTransform v2 normalization was then performed to integrate samples based on batch with regression based on cell cycle scores and ribosomal content (vars.to.regress = c("S.Score", "G2M.Score", “percent_ribo”)). Principal component analysis (PCA) was performed using the most highly variable genes, and cells were clustered based on the top 15 principal components using the functions RunUMAP, FindNeighbors, and FindClusters, and the output UMAP graphs were generated by DimPlot. The resolution parameter of 0.4 was set so that cluster boundaries largely separated the likely major cell types. Cluster annotation was performed based on the expression of known marker genes, leading to 11 broadly assigned cell types. Cells filtered out of the ArchR dataset based on doublet identification (see” ArchR Analysis” below) were removed from the Seurat dataset (final n = 23,838 cells). Differential gene expression was then

performed with the function FindAllMarkers (logfc.threshold = 0.25 and min.pct = 0.1) to generate a list of top marker genes for each cluster. In pairwise comparisons of differential gene expression, positive values reflect upregulation in mutant lines, while negative values reflect upregulation in WT.

Seurat Analysis - CDX2

Outputs from the CellRanger-Arc count pipeline were analyzed using the Seurat package (version 4.3.0)(Satija *et al.*, 2015; Butler *et al.*, 2018; Stuart *et al.*, 2019) in R (v4.2.0). Quality control filtering included the removal of outliers due to the number of features/genes (nFeature_RNA > 2500 & nFeature_RNA < 4500, nCount_RNA > 200 & nCount_RNA < 12000, mitochondrial percentage > 5% and mitochondrial percentage < 20%, and ribosomal percentage > 3% and ribosomal percentage < 10%). Cell cycle scores were added using the function CellCycleScoring. ScTransform v2 normalization was then performed to integrate samples based on batch with regression based on cell cycle scores and ribosomal content (vars.to.regress = c("S.Score", "G2M.Score", “percent_ribo”)). Principal component analysis (PCA) was performed using the most highly variable genes, and cells were clustered based on the top 15 principal components using the functions RunUMAP, FindNeighbors, and FindClusters, and the output UMAP graphs were generated by DimPlot. The resolution parameter of 0.4 was set so that cluster boundaries largely separated the likely major cell types. Cluster annotation was performed based on the expression of known marker genes, leading to 10 broadly assigned cell types. Cells filtered out of the ArchR dataset based on doublet identification (see” ArchR Analysis” below) were removed from the Seurat dataset (final n = 25,557 cells). Differential gene expression was then performed with the function FindAllMarkers (logfc.threshold = 0.25 and min.pct = 0.1) to generate a list of top marker genes for each cluster. In pairwise comparisons of differential gene expression, positive values reflect upregulation in mutant lines, while negative values reflect upregulation in WT.

ArchR Analysis - TBXT

Indexed Fragment files generated by the CellRanger-Arc counts function served as input for the generation of sample-specific ArrowFiles (minTSS = 4 & minFrags = 1000) using the R package ArchR v1.0.2(Granja *et al.*, 2021). ArrowFile creation also generates a genome wide TileMatrix using 500bp bins and a GeneScoreMatrix, an estimated value of gene expression based on a weighted calculation of accessibility within a gene body and surrounding locus. Each Arrow file (n=6 total) was then aggregated into a single ArchRProject for downstream analysis. Corresponding Gene Expression Matrices were imported to the project based on the filtered feature barcode matrix h5 file generated by CellRanger-arc counts and descriptive cluster labels were imported from the corresponding Seurat object based on cell barcodes. Cells filtered out of the Seurat dataset based on QC metrics previously described were also removed from the ArchR dataset. Cell doublet removal was performed in ArchR using the functions addDoubletScores and filterDoublets, leaving 23,838 cells with a median TSS of 13.278 and a median value of 12,774.5 fragments per cell (Cells filtered = WT-1 767/8759, TBXT-Het-1 613/7831, TBXT-KO-1 826/9092, WT-2 916/9572, TBXT-Het-2 617/7857, TBXT-KO-2 1623/12740).

After generation of the aggregated ArchR project, dimensionality reduction was performed using ArchR's implementation of Iterative Latent Semantic Indexing (LSI) with the function addIterativeLSI based on the 500bp TileMatrix with 4 iterations, increasing resolution values (0.1, 0.2, and 0.4) each iteration. This was repeated using the Gene Expression Matrix based on 2,500 variable features, yielding "LSI-ATAC" and LSI-RNA" reduced dimensions, respectively. The two reduced dimension values were then combined using addCombinedDims to yield "LSI_Combined," which was used as input for batch correction using Harmony with the function addHarmony (groupby = "Sample, "Batch"). Clustering was then performed using Harmony-corrected values with addClusters with a resolution of 0.4 from the R package Seurat. Finally, clusters were visualized with function plot embedding, using batch-corrected single-cell embedding values from Uniform Manifold Approximation and Projection (UMAP) using the

function addUMAP. Clusters and their corresponding UMAP projection were very similar to those generated based on RNA data in Seurat, and unless otherwise stated. Cluster identities in figures are based on barcodes transferred from Seurat rather than ArchR's LSI implementation.

After cluster annotation, pseudobulk replicates of cells within similar groups were created to facilitate peak calling. Replicates were created using addGroupCoverages and peak calling was performed using addReproduciblePeakSet using standard settings by implementing MACS2. We then used ArchR's iterative overlap peak merging method to create a union peakset of 322,520 unique peaks.

Cluster-enriched marker peaks were identified with getMarkerFeatures, using a Wilcoxon test and normalizing for biases from TSS enrichment scores and sequencing depth, and visualized with plotMarkerHeatmap, filtering for FDR < 0.01 and abs(Log2FC) > 1.25. Motif enrichment of cluster-enriched peaks was done using addMotifAnnotations with the "CODEX" motif set. Enriched motifs per cluster were visualized by first running peakAnnoEnrichment, with FDR < 0.1 and Log2FC > 0.5. The top 7 motifs per cluster were visualized as a heatmap using plotEnrichHeatmap.

Peak-to-gene linkage analysis was performed in ArchR using the addPeak2GeneLinks command, using the batch-corrected Harmony embedding values. A total of 3,010,318 linkages were found using FDR 1e-04, corCutOff = 0.95 and a resolution of 1.

Differential accessibility within the mesoderm cluster was performed by using the command subsetArchrProject to subset the ArchR project based on the mesoderm-annotated cluster as determined from Seurat. This subsetting yielded 3,212 cells, with a median TSS of 3,085 and a median number of fragments of 12,423. Differentially expressed genes predicted pairwise across genotypes (WT vs. TBXT-KO or WT vs. TBXT-Het) were identified with getMarkerFeatures based on the GeneScoreMatrix, using a Wilcoxon Test and normalizing for biases from TSS enrichment scores and sequencing depth. GetMarkers was then run and visualized as a volcano plot using plotMarkers (FDR < 0.1 and abs(Log2FC) > 0.5). This process

was repeated for the PeakMatrix to determine uniquely accessible peaks. 18 Peaks were detected between WT and TBXT-Het and 0 peaks were detected between WT and TBXT-KO. No CODEX motif enrichments were detected between genotypes (FDR < 0.1 and abs(Log2FC) > 0.5).

ArchR Analysis - CDX2

Indexed Fragment files generated by the CellRanger-Arc counts function served as input for the generation of sample-specific ArrowFiles (minTSS = 4 & minFrags = 1000) using the R package ArchR v1.0.2(Granja *et al.*, 2021). ArrowFile creation also generates a genome wide TileMatrix using 500bp bins and a GeneScoreMatrix, an estimated value of gene expression based on a weighted calculation of accessibility within a gene body and surrounding locus. Each Arrow file (n=6 total) was then aggregated into a single ArchRProject for downstream analysis. Corresponding Gene Expression Matrices were imported to the project based on the filtered feature barcode matrix h5 file generated by CellRanger-arc counts and descriptive cluster labels were imported from the corresponding Seurat object based on cell barcodes. Cells filtered out of the Seurat dataset based on QC metrics previously described were also removed from the ArchR dataset. Cell doublet removal was performed in ArchR using the functions addDoubletScores and filterDoublets, leaving 25,557 cells with a median TSS of 13.431 and a median value of 11,458 fragments per cell (cells filtered = WT-1 0/2908, CDX2-Het-1 257/5072, CDX2-KO-1 0/ 2836, WT-2 288/5368, CDX2-Het-2 = 243/4935, CDX2-KO-2 = 306/5532).

After generation of the aggregated ArchR project, dimensionality reduction was performed using ArchR's implementation of Iterative Latent Semantic Indexing (LSI) with the function addIterativeLSI based on the 500bp TileMatrix with default settings (iterations = 2, sampleCells = 10000, n.start = 10, resolution = 2, maxClusters = 6). This was repeated using the Gene Expression Matrix based on 2,500 variable features, yielding "LSI-ATAC" and LSI-RNA" reduced dimensions, respectively. The two reduced dimension values were then combined using addCombinedDims to yield "LSI_Combined," which was used as input for batch correction using

Harmony with the function addHarmony (groupby = “Sample, “Batch”). Clustering was then performed using Harmony-corrected values with addClusters with a resolution of 0.4 from the R package Seurat. Finally, clusters were visualized with function plot embedding, using batch-corrected single-cell embedding values from Uniform Manifold Approximation and Projection (UMAP) using the function addUMAP. Clusters and their corresponding UMAP projection were very similar to those generated based on RNA data in Seurat, and unless otherwise stated cluster identities in figures are based on barcodes transferred from Seurat rather than ArchR’s LSI implementation.

After cluster annotation, pseudobulk replicates of cells within similar groups were created to facilitate peak calling. Replicates were created using addGroupCoverages and peak calling was performed using addReproduciblePeakSet using standard settings by implementing MACS2. We then used ArchR’s iterative overlap peak merging method to create a union peakset of 305,429 unique peaks.

Cluster-enriched marker peaks were identified with getMarkerFeatures, using a Wilcoxon test and normalizing for biases from TSS enrichment scores and sequencing depth, and visualized with plotMarkerHeatmap, filtering for FDR =< 0.01 and abs(Log2FC) => 1.25. Motif enrichment of cluster-enriched peaks was done using addMotifAnnotations with the “cisbp” motif set. Enriched motifs per cluster were visualized by first running peakAnnoEnrichment, with FDR <= 0.1 and Log2FC >= 0.5. The top 20 significantly enriched motifs per cluster were visualized as a heatmap using plotEnrichHeatmap.

Peak-to-gene linkage analysis was performed in ArchR using the addPeak2GeneLinks command, using the batch-corrected Harmony embedding values. A total of 24,487 linkages were found using FDR 1e-04, corCutOff = 0.45, and a resolution of 1.

Differential accessibility within the extraembryonic mesoderm cluster was performed by using the command subsetArchrProject to subset the ArchR project based on the ‘ExeM-Late’ annotated cluster as determined from Seurat. This subsetting yielded 3,792 cells, with a median

TSS of 13.396 and a median number of fragments of 11,419. Differentially expressed genes predicted pairwise across genotypes (WT vs. CDX2-KO or WT vs. CDX2-Het) were identified with getMarkerFeatures based on the GeneScoreMatrix, using a Wilcoxon Test and normalizing for biases from TSS enrichment scores and sequencing depth. GetMarkers was then run and visualized as a volcano plot using plotMarkers (FDR =< 0.1 and abs(Log2FC) >= 0.5). This process was repeated for the PeakMatrix to determine uniquely accessible peaks. 25 Peaks were detected between WT and CDX2-Het and 261 peaks were detected between WT and CDX2-KO. Significant 'cisbp' motif enrichments detected between genotypes within these peaks were calculated using peakAnnoEnrichment() (FDR <= 0.1 and abs(Log2FC) >= 0.5).

CellChat

Cell signaling analysis was performed using the R package CellChat (Jin *et al.*, 2021). The Seurat object containing all samples related to the TBXT dataset was subset by genotype, yielding a separate Seurat object for WT, TBXT-Het, or TBXT-KO. These 3 objects were then imported into CellChat using the function createCellChat. These 3 objects were then imported into CellChat using the function createCellChat. All ligand-receptor and signaling pathways within the CellChatDB.human were kept for analysis. Initial preprocessing to identify over-expressed ligands and receptors was performed using the functions identifyOverExpressedGenes and identifyOverExpressedInteractions with standard settings. Inference of cell communication was calculated with computeCommunProb(cellchat) and filtered by filterCommunication(cellchat, min.cells = 10). Pathway-level cell communication was calculated with computeCommunProbPathway, and aggregated networks were identified with aggregateNet, using standard settings. Network centrality scores were assigned with the function netAnalysis_computeCentrality. This workflow was run for WT, TBXT-Het, and TBXT-KO datasets independently and differential signaling analysis was then run by merging the WT, TBXT-Het, and TBXT-KO objects with mergeCellChat(). Information flow, which is defined by the sum

of communication probability among all pairs of cell groups in the inferred network (i.e., the total weights in the network), was compared across genotypes using rankNet(cellchat). The distance of signaling networks between WT and TBXT-KO datasets was calculated by performing joint manifold learning and classification of communication networks based on functional similarity using computeNetSimilarityPairwise(cellchat), netEmbedding(cellchat), and netClustering (cellchat). Circle diagrams, heatmaps, and bubble plots of pathways of interest were then generated for each genotype separately using the standard settings for netVisual_aggregate(cellchat), netVisual_heatmap(cellchat), or netVisual_bubble(cellchat), respectively. Violin plots of differential gene expression were generated using plotGeneExpression(cellchat) with the standard settings. Separately, this process was repeated using the Seurat object containing all samples related to the CDX2 dataset including WT, CDX2-Het, or CDX2-KO.

Gene Ontology Analysis

Gene Ontology (GO) analysis for downregulated or upregulated TBXT or CDX2 dependent genes was performed with ShinyGO V0.77 (Ge, Jung and Yao, 2020) using GO Biological Process terms. Downregulated or upregulated TBXT-dependent gene lists from the mesoderm subcluster were assembled from differential tests between *TBXT-Het vs. WT* or *TBXT-KO vs. WT* in Seurat. Downregulated or upregulated CDX2-dependent gene lists from the ExeM-Late subcluster were assembled from differential tests between *CDX2-Het vs. WT* or *CDX2-KO vs. WT in Seurat*. Gene sets were filtered with a significance threshold set at adj. $p > 0.05$ and abs(log2FC) > 0.25. Biologically relevant pathways within the first twenty hits for the TBXT-KO vs. WT comparison were visualized with bar plots. The first twenty hits for the CDX2-KO or CDX2-Het vs. WT comparison were visualized with lollipop plots. The process was repeated for genes identified as overlapping in both TBXT-KO and CDX2-KO vs. WT comparisons.

Quantification and Statistical Analysis

Each experiment was performed with at least three biological replicates except multiomic snATAC and snRNA-seq, which was performed with two biological replicates. Multiple comparisons were used to compare multiple groups followed by unpaired *t*-tests (two-tailed) between two groups subject to a posthoc Bonferroni correction. In gene expression analysis, two replicates were used for each condition, and all gene expression was normalized to control wild-type populations followed by unpaired *t*-tests (two-tailed). Significance was specified as Adj. $P < 0.05$ unless otherwise specified in figure legends. All error bars represent the standard error of the mean (s.e.m.) unless otherwise noted in the figure legend.

Scientific Acknowledgements

We would like to thank the Mylinh Bernardi and Horng-Ru Lin of the Gladstone Genomics Core, the Gladstone Histology and Light Microscopy Core, the Gladstone Stem Cell Core, and the Nikon Microscopy Core at the University of California, San Francisco for their experimental expertise. In addition, we thank Pengfei Zhang and the Abate Lab at UCSF for assistance with PDMS stamp fabrication, as well as Nicholas Elder, David Joy, Jonathon Muncie-Vasic, and Martin Dominguez, and the rest of the McDevitt and Bruneau labs for their experimental and computational expertise.

REFERENCES

van den Akker, E. *et al.* (2002) “Cdx1 and Cdx2 have overlapping functions in anteroposterior patterning and posterior axis elongation.,” *Development*, 129(9), pp. 2181–2193.

Akoumianakis, I., Polkinghorne, M. and Antoniades, C. (2022) “Non-canonical WNT signalling in cardiovascular disease: mechanisms and therapeutic implications.,” *Nature Reviews. Cardiology*, 19(12), pp. 783–797. doi:10.1038/s41569-022-00718-5.

Alom Ruiz, S. and Chen, C.S. (2007) “Microcontact printing: A tool to pattern,” *Soft matter*, 3(2), pp. 168–177. doi:10.1039/B613349E.

Amin, S. *et al.* (2016) “Cdx and T Brachyury Co-activate Growth Signaling in the Embryonic Axial Progenitor Niche.,” *Cell reports*, 17(12), pp. 3165–3177. doi:10.1016/j.celrep.2016.11.069.

Anand, G.M. *et al.* (2023) “Controlling organoid symmetry breaking uncovers an excitable system underlying human axial elongation.,” *Cell*, 186(3), pp. 497-512.e23. doi:10.1016/j.cell.2022.12.043.

d’Apolito, M. *et al.* (2019) “Angiopoietin-1 haploinsufficiency affects the endothelial barrier and causes hereditary angioedema.,” *Clinical and Experimental Allergy*, 49(5), pp. 626–635. doi:10.1111/cea.13349.

Arnold, S.J. and Robertson, E.J. (2009) “Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo.,” *Nature Reviews. Molecular Cell Biology*, 10(2), pp. 91–103. doi:10.1038/nrm2618.

Babcock, K. *et al.* (2018) “Abstract 4486: Reexpression of *LSAMP* , a gene frequently deleted in African American prostate cancers, alters adhesive qualities of prostate tumor cells and inhibits Akt, ERK1/2, and β-Catenin signaling axis,” *Cancer Research*, 78(13_Supplement), pp. 4486–4486. doi:10.1158/1538-7445.AM2018-4486.

Bardot, E.S. and Hadjantonakis, A.-K. (2020) “Mouse gastrulation: Coordination of tissue patterning, specification and diversification of cell fate.,” *Mechanisms of Development*, 163, p. 103617. doi:10.1016/j.mod.2020.103617.

Beck, F. *et al.* (1995) “Expression of Cdx-2 in the mouse embryo and placenta: possible role in patterning of the extra-embryonic membranes.,” *Developmental Dynamics*, 204(3), pp. 219–227. doi:10.1002/aja.1002040302.

Beddington, R.S., Rashbass, P. and Wilson, V. (1992) “Brachyury--a gene affecting mouse gastrulation and early organogenesis.,” *Development (Cambridge, England). Supplement*, pp. 157–165.

Beisaw, A. *et al.* (2018) “BRACHYURY directs histone acetylation to target loci during mesoderm development.,” *EMBO Reports*, 19(1), pp. 118–134. doi:10.15252/embr.201744201.

Berger, M.F. *et al.* (2008) “Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences.,” *Cell*, 133(7), pp. 1266–1276. doi:10.1016/j.cell.2008.05.024.

Bernardo, A.S. *et al.* (2011) “BRACHYURY and CDX2 mediate BMP-induced differentiation of human and mouse pluripotent stem cells into embryonic and extraembryonic lineages.,” *Cell Stem Cell*, 9(2), pp. 144–155. doi:10.1016/j.stem.2011.06.015.

Bernstein, B.E. *et al.* (2010) “The NIH roadmap epigenomics mapping consortium.,” *Nature Biotechnology*, 28(10), pp. 1045–1048. doi:10.1038/nbt1010-1045.

de Bree, K., de Bakker, B.S. and Oostra, R.-J. (2018) “The development of the human notochord.,” *Plos One*, 13(10), p. e0205752. doi:10.1371/journal.pone.0205752.

Brooke-Bisschop, T. *et al.* (2017) “Essential roles for Cdx in murine primitive hematopoiesis.,” *Developmental Biology*, 422(2), pp. 115–124. doi:10.1016/j.ydbio.2017.01.002.

Buenrostro, J.D. *et al.* (2015) “ATAC-seq: a method for assaying chromatin accessibility genome-wide.,” *Current Protocols in Molecular Biology*, 109, p. 21.29.1-21.29.9. doi:10.1002/0471142727.mb2129s109.

Bulajić, M. *et al.* (2020) “Differential abilities to engage inaccessible chromatin diversify vertebrate Hox binding patterns.,” *Development*, 147(22). doi:10.1242/dev.194761.

Bulger, E.A., McDevitt, T.C. and Bruneau, B.G. (2024) “CDX2 dose-dependently influences the gene regulatory network underlying human extraembryonic mesoderm development.,” *Biology open* [Preprint]. doi:10.1242/bio.060323.

Bulger, E.A. *et al.* (2024) “TBXT dose sensitivity and the decoupling of nascent mesoderm specification from EMT progression in 2D human gastruloids.,” *Development* [Preprint]. doi:10.1242/dev.202516.

Butler, A. *et al.* (2018) “Integrating single-cell transcriptomic data across different conditions, technologies, and species.,” *Nature Biotechnology*, 36(5), pp. 411–420. doi:10.1038/nbt.4096.

Cao, J. *et al.* (2017) “Polarized actin and VE-cadherin dynamics regulate junctional remodelling and cell migration during sprouting angiogenesis.,” *Nature Communications*, 8(1), p. 2210. doi:10.1038/s41467-017-02373-8.

Cao, Y. *et al.* (1998) “Vascular endothelial growth factor C induces angiogenesis in vivo.,” *Proceedings of the National Academy of Sciences of the United States of America*, 95(24), pp. 14389–14394. doi:10.1073/pnas.95.24.14389.

Chawengsaksophak, K. *et al.* (1997) “Homeosis and intestinal tumours in Cdx2 mutant mice.,” *Nature*, 386(6620), pp. 84–87. doi:10.1038/386084a0.

Chawengsaksophak, K. *et al.* (2004) “Cdx2 is essential for axial elongation in mouse development.,” *Proceedings of the National Academy of Sciences of the United States of America*, 101(20), pp. 7641–7645. doi:10.1073/pnas.0401654101.

Chen, S. *et al.* (2023) “A mutation in TBXT causes congenital vertebral malformations in humans and mice.,” *Journal of Genetics and Genomics = Yi Chuan Xue Bao* [Preprint]. doi:10.1016/j.jgg.2023.09.009.

Chesley, P. (1935) “Development of the short-tailed mutant in the house mouse,” *Journal of Experimental Zoology*, 70(3), pp. 429–459. doi:10.1002/jez.1400700306.

Chisaka, O. and Capecchi, M.R. (1991) “Regionally restricted developmental defects resulting from targeted disruption of the mouse homeobox gene hox-1.5.,” *Nature*, 350(6318), pp. 473–479. doi:10.1038/350473a0.

Ciruna, B. and Rossant, J. (2001) “FGF signaling regulates mesoderm cell fate specification and morphogenetic movement at the primitive streak.,” *Developmental Cell*, 1(1), pp. 37–49. doi:10.1016/s1534-5807(01)00017-x.

Dale, J.K. *et al.* (2006) “Oscillations of the snail genes in the presomitic mesoderm coordinate segmental patterning and morphogenesis in vertebrate somitogenesis.,” *Developmental Cell*, 10(3), pp. 355–366. doi:10.1016/j.devcel.2006.02.011.

Denans, N., Iimura, T. and Pourquié, O. (2015) “Hox genes control vertebrate body elongation by collinear Wnt repression,” *eLife*, 4(e04379). doi:10.7554/eLife.04379.

Dobrovolskaia-Zavadskaia, N. (1927) “Sur la mortification spontanee de la queuw chez la spuris nouveau et sur l’existence d’un caractere (facteur) hereditaire non viable,” *C. R. Seances Soc. Biol. Fil.*, (97), pp. 114–116.

Downs, K.M. *et al.* (2009) “The Allantoic Core Domain: new insights into development of the murine allantois and its relation to the primitive streak.,” *Developmental Dynamics*, 238(3), pp. 532–553. doi:10.1002/dvdy.21862.

Drake, C.J. *et al.* (2000) “VEGF regulates cell behavior during vasculogenesis.,” *Developmental Biology*, 224(2), pp. 178–188. doi:10.1006/dbio.2000.9744.

Duboule, D. and Morata, G. (1994) “Colinearity and functional hierarchy among genes of the homeotic complexes.,” *Trends in Genetics*, 10(10), pp. 358–364. doi:10.1016/0168-9525(94)90132-5.

Dunty, W.C. *et al.* (2008) “Wnt3a/beta-catenin signaling controls posterior body development by coordinating mesoderm formation and segmentation.,” *Development*, 135(1), pp. 85–94. doi:10.1242/dev.009266.

Durak, O. *et al.* (2015) “Ankyrin-G regulates neurogenesis and Wnt signaling by altering the subcellular localization of β-catenin.,” *Molecular Psychiatry*, 20(3), pp. 388–397. doi:10.1038/mp.2014.42.

Ekker, S.C. *et al.* (1994) “The degree of variation in DNA sequence recognition among four Drosophila homeotic proteins.,” *The EMBO Journal*, 13(15), pp. 3551–3560. doi:10.1002/j.1460-2075.1994.tb06662.x.

Evseenko, D. *et al.* (2010) “Mapping the first stages of mesoderm commitment during differentiation of human embryonic stem cells.,” *Proceedings of the National Academy of Sciences of the United States of America*, 107(31), pp. 13742–13747. doi:10.1073/pnas.1002077107.

Faial, T. *et al.* (2015) “Brachyury and SMAD signalling collaboratively orchestrate distinct mesoderm and endoderm gene regulatory networks in differentiating human embryonic stem cells.,” *Development*, 142(12), pp. 2121–2135. doi:10.1242/dev.117838.

Fernando, R.I. *et al.* (2010) “The T-box transcription factor Brachyury promotes epithelial-mesenchymal transition in human tumor cells.,” *The Journal of Clinical Investigation*, 120(2), pp. 533–544. doi:10.1172/JCI38379.

Foley, T. and Lohnes, D. (2022) “Cdx regulates gene expression through PRC2-mediated epigenetic mechanisms.,” *Developmental Biology*, 483, pp. 22–33. doi:10.1016/j.ydbio.2021.12.014.

Foley, T.E. *et al.* (2019) “Role of Cdx factors in early mesodermal fate decisions.,” *Development*, 146(7). doi:10.1242/dev.170498.

Galceran, J. *et al.* (1999) “Wnt3a-/--like phenotype and limb deficiency in Lef1(-/-)Tcf1(-/-) mice.,” *Genes & Development*, 13(6), pp. 709–717. doi:10.1101/gad.13.6.709.

Gavard, J., Patel, V. and Gutkind, J.S. (2008) “Angiopoietin-1 prevents VEGF-induced endothelial permeability by sequestering Src through mDia.,” *Developmental Cell*, 14(1), pp. 25–36. doi:10.1016/j.devcel.2007.10.019.

Ge, S.X., Jung, D. and Yao, R. (2020) “ShinyGO: a graphical gene-set enrichment tool for animals and plants.,” *Bioinformatics*, 36(8), pp. 2628–2629. doi:10.1093/bioinformatics/btz931.

Ghebranious, N. *et al.* (2008) “A missense T (Brachyury) mutation contributes to vertebral malformations.,” *Journal of Bone and Mineral Research*, 23(10), pp. 1576–1583. doi:10.1359/jbmr.080503.

Gouti, M. *et al.* (2017) “A Gene Regulatory Network Balances Neural and Mesoderm Specification during Vertebrate Trunk Development.,” *Developmental Cell*, 41(3), pp. 243-261.e7. doi:10.1016/j.devcel.2017.04.002.

Granja, J.M. *et al.* (2021) “ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis.,” *Nature Genetics*, 53(3), pp. 403–411. doi:10.1038/s41588-021-00790-6.

Grüneberg, H. (1958) “Genetical studies on the skeleton of the mouse,” *Development*, 6(3), pp. 424–443. doi:10.1242/dev.6.3.424.

Hashimoto, K., Fujimoto, H. and Nakatsuji, N. (1987) “An ECM substratum allows mouse mesodermal cells isolated from the primitive streak to exhibit motility similar to that inside the embryo and reveals a deficiency in the T/T mutant cells.,” *Development*, 100(4), pp. 587–598. doi:10.1242/dev.100.4.587.

Hayashi, M. *et al.* (2018) “Generation of germ cells from pluripotent stem cells in mammals.,” *Reproductive medicine and biology*, 17(2), pp. 107–114. doi:10.1002/rmb2.12077.

Henrique, D. *et al.* (2015) “Neuromesodermal progenitors and the making of the spinal cord.,” *Development*, 142(17), pp. 2864–2875. doi:10.1242/dev.119768.

Herrmann, B.G. *et al.* (1990) “Cloning of the T gene required in mesoderm formation in the mouse.,” *Nature*, 343(6259), pp. 617–622. doi:10.1038/343617a0.

Huelsken, J. *et al.* (2000) “Requirement for beta-catenin in anterior-posterior axis formation in mice.,” *The Journal of Cell Biology*, 148(3), pp. 567–578. doi:10.1083/jcb.148.3.567.

Iimura, T. and Pourquié, O. (2006) “Collinear activation of Hoxb genes during gastrulation is linked to mesoderm cell ingression.,” *Nature*, 442(7102), pp. 568–571. doi:10.1038/nature04838.

Inman, K.E. and Downs, K.M. (2006) “Brachyury is required for elongation and vasculogenesis in the murine allantois.,” *Development*, 133(15), pp. 2947–2959. doi:10.1242/dev.02454.

Jin, S. *et al.* (2021) “Inference and analysis of cell-cell communication using CellChat.,” *Nature Communications*, 12(1), p. 1088. doi:10.1038/s41467-021-21246-9.

Joy, D.A., Libby, A.R.G. and McDevitt, T.C. (2021) “Deep neural net tracking of human pluripotent stem cells reveals intrinsic behaviors directing morphogenesis.,” *Stem cell reports*, 16(5), pp. 1317–1330. doi:10.1016/j.stemcr.2021.04.008.

Kachgal, S., Mace, K.A. and Boudreau, N.J. (2012) “The dual roles of homeobox genes in vascularization and wound healing.,” *Cell Adhesion & Migration*, 6(6), pp. 457–470. doi:10.4161/cam.22164.

Kanazawa, A. *et al.* (2005) “Wnt5b partially inhibits canonical Wnt/beta-catenin signaling pathway and promotes adipogenesis in 3T3-L1 preadipocytes.,” *Biochemical and Biophysical Research Communications*, 330(2), pp. 505–510. doi:10.1016/j.bbrc.2005.03.007.

Kasprzak, A. (2020) “Angiogenesis-Related Functions of Wnt Signaling in Colorectal Carcinogenesis.,” *Cancers*, 12(12). doi:10.3390/cancers12123601.

Kaul, H. *et al.* (2023) “Virtual cells in a virtual microenvironment recapitulate early development-like patterns in human pluripotent stem cell colonies.,” *Stem cell reports*, 18(1), pp. 377–393. doi:10.1016/j.stemcr.2022.10.004.

Kikuchi, A. *et al.* (2012) “Wnt5a: its signalling, functions and implication in diseases.,” *Acta Physiologica*, 204(1), pp. 17–33. doi:10.1111/j.1748-1716.2011.02294.x.

Kinder, S.J. *et al.* (1999) “The orderly allocation of mesodermal cells to the extraembryonic structures and the anteroposterior axis during gastrulation of the mouse embryo.,” *Development*, 126(21), pp. 4691–4701. doi:10.1242/dev.126.21.4691.

Koch, F. *et al.* (2017) “Antagonistic Activities of Sox2 and Brachyury Control the Fate Choice of Neuro-Mesodermal Progenitors.,” *Developmental Cell*, 42(5), pp. 514-526.e7. doi:10.1016/j.devcel.2017.07.021.

Krumlauf, R. (1993) “Mouse Hox genetic functions.,” *Current Opinion in Genetics & Development*, 3(4), pp. 621–625. doi:10.1016/0959-437x(93)90098-a.

Kumar, N. *et al.* (2019) “The lineage-specific transcription factor CDX2 navigates dynamic chromatin to control distinct stages of intestine development.,” *Development*, 146(5). doi:10.1242/dev.172189.

Lawson, K.A., Meneses, J.J. and Pedersen, R.A. (1991) “Clonal analysis of epiblast fate during germ layer formation in the mouse embryo.,” *Development*, 113(3), pp. 891–911. doi:10.1242/dev.113.3.891.

Lee, S.-H. *et al.* (2005) “Thymosin {beta}(10) inhibits angiogenesis and tumor growth by interfering with Ras function.,” *Cancer Research*, 65(1), pp. 137–148.

Libby, A.R.G. *et al.* (2019) “Automated Design of Pluripotent Stem Cell Self-Organization.,” *Cell Systems*, 9(5), pp. 483-495.e10. doi:10.1016/j.cels.2019.10.008.

Liu, J. *et al.* (2021) “A rare variant of ANK3 is associated with intracranial aneurysm.,” *Frontiers in neurology*, 12, p. 672570. doi:10.3389/fneur.2021.672570.

Lolas, M. *et al.* (2014) “Charting Brachyury-mediated developmental pathways during early mouse embryogenesis.,” *Proceedings of the National Academy of Sciences of the United States of America*, 111(12), pp. 4478–4483. doi:10.1073/pnas.1402612111.

Lufkin, T. *et al.* (1991) “Disruption of the Hox-1.6 homeobox gene results in defects in a region corresponding to its rostral domain of expression.,” *Cell*, 66(6), pp. 1105–1119. doi:10.1016/0092-8674(91)90034-v.

Lufkin, T. *et al.* (1992) “Homeotic transformation of the occipital bones of the skull by ectopic expression of a homeobox gene.,” *Nature*, 359(6398), pp. 835–841. doi:10.1038/359835a0.

Mann, R.S., Lelli, K.M. and Joshi, R. (2009) “Chapter 3 hox specificity,” in *Hox Genes*. Elsevier (Current topics in developmental biology), pp. 63–101. doi:10.1016/S0070-2153(09)88003-4.

Martyn, I., Siggia, E.D. and Brivanlou, A.H. (2019) “Mapping cell migrations and fates in a gastruloid model to the human primitive streak.,” *Development*, 146(17). doi:10.1242/dev.179564.

Ma, H. *et al.* (2019) “In vitro culture of cynomolgus monkey embryos beyond early gastrulation.,” *Science*, 366(6467). doi:10.1126/science.aax7890.

Miano, J.M. *et al.* (1996) “Restricted expression of homeobox genes distinguishes fetal from adult human smooth muscle cells.,” *Proceedings of the National Academy of Sciences of the United States of America*, 93(2), pp. 900–905. doi:10.1073/pnas.93.2.900.

Mikawa, T. *et al.* (2004) “Induction and patterning of the primitive streak, an organizing center of gastrulation in the amniote.,” *Developmental Dynamics*, 229(3), pp. 422–432. doi:10.1002/dvdy.10458.

Minn, K.T. *et al.* (2020) “High-resolution transcriptional and morphogenetic profiling of cells from micropatterned human ESC gastruloid cultures.,” *eLife*, 9. doi:10.7554/eLife.59445.

Minn, K.T. *et al.* (2021) “Gene expression dynamics underlying cell fate emergence in 2D micropatterned human embryonic stem cell gastruloids.,” *Stem cell reports*, 16(5), pp. 1210–1227. doi:10.1016/j.stemcr.2021.03.031.

Mohamed, O.A., Clarke, H.J. and Dufort, D. (2004) “Beta-catenin signaling marks the prospective site of primitive streak formation in the mouse embryo.,” *Developmental Dynamics*, 231(2), pp. 416–424. doi:10.1002/dvdy.20135.

Muhr, J., Arbor, T.C. and Ackerman, K.M. (2024) “Embryology, Gastrulation,” in *StatPearls*. Treasure Island (FL): StatPearls Publishing.

Muqbil, I. *et al.* (2014) “Snail nuclear transport: the gateways regulating epithelial-to-mesenchymal transition?,” *Seminars in Cancer Biology*, 27, pp. 39–45. doi:10.1016/j.semcancer.2014.06.003.

Myers, C., Charboneau, A. and Boudreau, N. (2000) “Homeobox B3 promotes capillary morphogenesis and angiogenesis.,” *The Journal of Cell Biology*, 148(2), pp. 343–351. doi:10.1083/jcb.148.2.343.

Neijts, R. *et al.* (2017) “Cdx is crucial for the timing mechanism driving colinear Hox activation and defines a trunk segment in the Hox cluster topology.,” *Developmental Biology*, 422(2), pp. 146–154. doi:10.1016/j.ydbio.2016.12.024.

Nguyen, T.T. *et al.* (2017) “Cdx2 Regulates Gene Expression through Recruitment of Brg1-associated Switch-Sucrose Non-fermentable (SWI-SNF) Chromatin Remodeling Activity.,” *The Journal of Biological Chemistry*, 292(8), pp. 3389–3399. doi:10.1074/jbc.M116.752774.

Niu, Y. *et al.* (2019) “Dissecting primate early post-implantation development using long-term in vitro embryo culture.,” *Science*, 366(6467). doi:10.1126/science.aaw5754.

Niwa, H. *et al.* (2005) “Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation.,” *Cell*, 123(5), pp. 917–929. doi:10.1016/j.cell.2005.08.040.

Noyes, M.B. *et al.* (2008) “Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites.,” *Cell*, 133(7), pp. 1277–1289. doi:10.1016/j.cell.2008.05.023.

O'Reilly, M.A., Smith, J.C. and Cunliffe, V. (1995) "Patterning of the mesoderm in Xenopus: dose-dependent and synergistic effects of Brachyury and Pintallavis.," *Development*, 121(5), pp. 1351–1359. doi:10.1242/dev.121.5.1351.

Palmer, N. and Kaldis, P. (2016) "Regulation of the embryonic cell cycle during mammalian preimplantation development.," *Current Topics in Developmental Biology*, 120, pp. 1–53. doi:10.1016/bs.ctdb.2016.05.001.

Pan, Q. *et al.* (2020) "TMSB10 acts as a biomarker and promotes progression of clear cell renal cell carcinoma.," *International Journal of Oncology*, 56(5), pp. 1101–1114. doi:10.3892/ijo.2020.4991.

Papapetrou, C. *et al.* (1999) "A genetic study of the human T gene and its exclusion as a major candidate gene for sacral agenesis with anorectal atresia.," *Journal of Medical Genetics*, 36(3), pp. 208–213.

Postma, A.V. *et al.* (2014) "Mutations in the T (brachyury) gene cause a novel syndrome consisting of sacral agenesis, abnormal ossification of the vertebral bodies and a persistent notochordal canal.," *Journal of Medical Genetics*, 51(2), pp. 90–97. doi:10.1136/jmedgenet-2013-102001.

Prummel, K.D., Nieuwenhuize, S. and Mosimann, C. (2020) "The lateral plate mesoderm.," *Development*, 147(12). doi:10.1242/dev.175059.

Rashbass, P. *et al.* (1991) "A cell autonomous function of Brachyury in T/T embryonic stem cell chimaeras.," *Nature*, 353(6342), pp. 348–351. doi:10.1038/353348a0.

Rashbass, P. *et al.* (1994) “Alterations in gene expression during mesoderm formation and axial patterning in Brachyury (T) embryos.,” *The International Journal of Developmental Biology*, 38(1), pp. 35–44.

Rivera-Pérez, J.A. and Magnuson, T. (2005) “Primitive streak formation in mice is preceded by localized activation of Brachyury and Wnt3.,” *Developmental Biology*, 288(2), pp. 363–371. doi:10.1016/j.ydbio.2005.09.012.

Roselli, M. *et al.* (2012) “Brachyury, a driver of the epithelial-mesenchymal transition, is overexpressed in human lung tumors: an opportunity for novel interventions against lung cancer.,” *Clinical Cancer Research*, 18(14), pp. 3868–3879. doi:10.1158/1078-0432.CCR-11-3211.

Rossant, J. and Cross, J.C. (2001) “Placental development: lessons from mouse mutants.,” *Nature Reviews. Genetics*, 2(7), pp. 538–548. doi:10.1038/35080570.

Rostovskaya, M. *et al.* (2022) “Amniogenesis occurs in two independent waves in primates.,” *Cell Stem Cell*, 29(5), pp. 744-759.e6. doi:10.1016/j.stem.2022.03.014.

Sasaki, K. *et al.* (2016) “The germ cell fate of cynomolgus monkeys is specified in the nascent amnion.,” *Developmental Cell*, 39(2), pp. 169–185. doi:10.1016/j.devcel.2016.09.007.

Satija, R. *et al.* (2015) “Spatial reconstruction of single-cell gene expression data.,” *Nature Biotechnology*, 33(5), pp. 495–502. doi:10.1038/nbt.3192.

Savory, J.G.A. *et al.* (2009) “Cdx2 regulation of posterior development through non-Hox targets.,” *Development*, 136(24), pp. 4099–4110. doi:10.1242/dev.041582.

Saxena, M. *et al.* (2017) “Transcription factor-dependent ‘anti-repressive’ mammalian enhancers exclude H3K27me3 from extended genomic domains.,” *Genes & Development*, 31(23–24), pp. 2391–2404. doi:10.1101/gad.308536.117.

Saykali, B. *et al.* (2019) “Distinct mesoderm migration phenotypes in extra-embryonic and embryonic regions of the early mouse embryo.,” *eLife*, 8. doi:10.7554/eLife.42434.

Schoenwolf, G.C. and Smith, J.L. (2000) “Gastrulation and early mesodermal patterning in vertebrates.,” *Methods in Molecular Biology*, 135, pp. 113–125. doi:10.1385/1-59259-685-1:113.

Schüle, K.M. *et al.* (2023) “Eomes restricts Brachyury functions at the onset of mouse gastrulation.,” *Developmental Cell*, 58(18), pp. 1627-1642.e7. doi:10.1016/j.devcel.2023.07.023.

Senger, D.R. *et al.* (1983) “Tumor cells secrete a vascular permeability factor that promotes accumulation of ascites fluid.,” *Science*, 219(4587), pp. 983–985. doi:10.1126/science.6823562.

Shibuya, M. (2011) “Vascular Endothelial Growth Factor (VEGF) and Its Receptor (VEGFR) Signaling in Angiogenesis: A Crucial Target for Anti- and Pro-Angiogenic Therapies.,” *Genes & cancer*, 2(12), pp. 1097–1105. doi:10.1177/1947601911423031.

Stefater, J.A. *et al.* (2011) “Regulation of angiogenesis by a non-canonical Wnt-Flt1 pathway in myeloid cells.,” *Nature*, 474(7352), pp. 511–515. doi:10.1038/nature10085.

Stewart, F. (1996) “Roles of mesenchymal-epithelial interactions and hepatocyte growth factor-scatter factor (HGF-SF) in placental development.,” *Reviews of reproduction*, 1(3), pp. 144–148. doi:10.1530/ror.0.0010144.

Stott, D., Kispert, A. and Herrmann, B.G. (1993) “Rescue of the tail defect of Brachyury mice.,” *Genes & Development*, 7(2), pp. 197–203. doi:10.1101/gad.7.2.197.

Strumpf, D. *et al.* (2005) “Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst.,” *Development*, 132(9), pp. 2093–2102. doi:10.1242/dev.01801.

Stuart, T. *et al.* (2019) “Comprehensive Integration of Single-Cell Data.,” *Cell*, 177(7), pp. 1888-1902.e21. doi:10.1016/j.cell.2019.05.031.

Sun, X. *et al.* (1999) “Targeted disruption of Fgf8 causes failure of cell migration in the gastrulating mouse embryo.,” *Genes & Development*, 13(14), pp. 1834–1846. doi:10.1101/gad.13.14.1834.

Technau, U. (2001) “Brachyury, the blastopore and the evolution of the mesoderm.,” *Bioessays: News and Reviews in Molecular, Cellular and Developmental Biology*, 23(9), pp. 788–794. doi:10.1002/bies.1114.

Théry, M. and Piel, M. (2009) “Adhesive micropatterns for cells: a microcontact printing protocol.,” *Cold Spring Harbor Protocols*, 2009(7), p. pdb.prot5255. doi:10.1101/pdb.prot5255.

Thurston, G. *et al.* (1999) “Leakage-resistant blood vessels in mice transgenically overexpressing angiopoietin-1.,” *Science*, 286(5449), pp. 2511–2514. doi:10.1126/science.286.5449.2511.

Tsankov, A.M. *et al.* (2015) “Transcription factor binding dynamics during human ES cell differentiation.,” *Nature*, 518(7539), pp. 344–349. doi:10.1038/nature14233.

Tzouanacou, E. *et al.* (2009) “Redefining the progression of lineage segregations during mammalian embryogenesis by clonal analysis.,” *Developmental Cell*, 17(3), pp. 365–376. doi:10.1016/j.devcel.2009.08.002.

Verzi, M.P. *et al.* (2010) “Differentiation-specific histone modifications reveal dynamic chromatin interactions and partners for the intestinal transcription factor CDX2.,” *Developmental Cell*, 19(5), pp. 713–726. doi:10.1016/j.devcel.2010.10.006.

Verzi, M.P. *et al.* (2013) “Intestinal master transcription factor CDX2 controls chromatin access for partner transcription factor binding.,” *Molecular and Cellular Biology*, 33(2), pp. 281–292. doi:10.1128/MCB.01185-12.

Warmflash, A. *et al.* (2014) “A method to recapitulate early embryonic spatial patterning in human embryonic stem cells.,” *Nature Methods*, 11(8), pp. 847–854. doi:10.1038/nmeth.3016.

Watson, E.D. and Cross, J.C. (2005) “Development of structures and transport functions in the mouse placenta.,” *Physiology*, 20, pp. 180–193. doi:10.1152/physiol.00001.2005.

Wilkinson, D.G., Bhatt, S. and Herrmann, B.G. (1990) “Expression pattern of the mouse T gene and its role in mesoderm formation.,” *Nature*, 343(6259), pp. 657–659. doi:10.1038/343657a0.

William, D.A. *et al.* (2007) “Identification of oscillatory genes in somitogenesis from functional genomic analysis of a human mesenchymal stem cell model.,” *Developmental Biology*, 305(1), pp. 172–186. doi:10.1016/j.ydbio.2007.02.007.

Wilson, V. and Beddington, R.S. (1996) “Cell fate and morphogenetic movement in the late mouse primitive streak.,” *Mechanisms of Development*, 55(1), pp. 79–89. doi:10.1016/0925-4773(95)00493-9.

Wilson, V. and Beddington, R. (1997) “Expression of T protein in the primitive streak is necessary and sufficient for posterior mesoderm movement and somite differentiation.,” *Developmental Biology*, 192(1), pp. 45–58. doi:10.1006/dbio.1997.8701.

Wilson, V., Rashbass, P. and Beddington, R.S. (1993) “Chimeric analysis of T (Brachyury) gene function.,” *Development*, 117(4), pp. 1321–1331. doi:10.1242/dev.117.4.1321.

Wilson, V. *et al.* (1995) “The T gene is necessary for normal mesodermal morphogenetic cell movements during gastrulation.,” *Development*, 121(3), pp. 877–886. doi:10.1242/dev.121.3.877.

Winnier, G. *et al.* (1995) “Bone morphogenetic protein-4 is required for mesoderm formation and patterning in the mouse.,” *Genes & Development*, 9(17), pp. 2105–2116. doi:10.1101/gad.9.17.2105.

Wolpert, L. *et al.* (2019) *Principles of Development*. Oxford University Press. doi:10.1093/hesc/9780198800569.001.0001.

Wu, L. *et al.* (2002) “Identification of a family of mastermind-like transcriptional coactivators for mammalian notch receptors.,” *Molecular and Cellular Biology*, 22(21), pp. 7688–7700. doi:10.1128/MCB.22.21.7688-7700.2002.

Wu, Y. *et al.* (2003) “HoxB5 is an upstream transcriptional switch for differentiation of the vascular endothelium from precursor cells.,” *Molecular and Cellular Biology*, 23(16), pp. 5680–5691. doi:10.1128/MCB.23.16.5680-5691.2003.

Xia, B. *et al.* (2024) “On the genetic basis of tail-loss evolution in humans and apes.,” *Nature*, 626(8001), pp. 1042–1048. doi:10.1038/s41586-024-07095-8.

Xu, R. *et al.* (2019) “Roles of the Phosphorylation of Transcriptional Factors in Epithelial-Mesenchymal Transition.,” *Journal of Oncology*, 2019, p. 5810465. doi:10.1155/2019/5810465.

Yamaguchi, T.P. *et al.* (1999) “A Wnt5a pathway underlies outgrowth of multiple structures in the vertebrate embryo.,” *Development*, 126(6), pp. 1211–1223. doi:10.1242/dev.126.6.1211.

Yamamichi, N. *et al.* (2009) “Cdx2 and the Brm-type SWI/SNF complex cooperatively regulate villin expression in gastrointestinal cells.,” *Experimental Cell Research*, 315(10), pp. 1779–1789. doi:10.1016/j.yexcr.2009.01.006.

Yanagisawa, K.O., Fujimoto, H. and Urushihara, H. (1981) “Effects of the brachyury (T) mutation on morphogenetic movement in the mouse embryo.,” *Developmental Biology*, 87(2), pp. 242–248. doi:10.1016/0012-1606(81)90147-0.

Yekta, S., Tabin, C.J. and Bartel, D.P. (2008) “MicroRNAs in the Hox network: an apparent link to posterior prevalence.,” *Nature Reviews. Genetics*, 9(10), pp. 789–796. doi:10.1038/nrg2400.

Zheng, Y. *et al.* (2019) “Controlled modelling of human epiblast and amnion development using stem cells.,” *Nature*, 573(7774), pp. 421–425. doi:10.1038/s41586-019-1535-2.

Zhu, J., Kwan, K.M. and Mackem, S. (2016) “Putative oncogene Brachyury (T) is essential to specify cell fate but dispensable for notochord progenitor proliferation and EMT.,” *Proceedings of the National Academy of Sciences of the United States of America*, 113(14), pp. 3820–3825. doi:10.1073/pnas.1601252113.

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:
Emily Bulger
9A1633BEA61B432...
Author Signature

2/12/2024
Date