Title

CDX2 dose-dependently influences the gene regulatory network underlying human extraembryonic mesoderm development.

Permalink

https://escholarship.org/uc/item/5dq2w98v

Journal

Biology Open, 13(3)

Authors

Bulger, Emily

McDevitt, Todd

Bruneau, Benoit

Publication Date

2024-03-01

DOI

10.1242/bio.060323

Copyright Information

Peer reviewed

The Company of Biologists

RESEARCH ARTICLE

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

Emily A. Bulger[1,2], Todd C. McDevitt[1,3,*] and Benoit G. Bruneau[1,4,5,6,7,‡]

ABSTRACT

Loss of *Cdx2 in vivo* leads to stunted development of the allantois, an extraembryonic mesoderm-derived structure critical for nutrient delivery and waste removal in the early embryo. Here, we investigate how CDX2 dose-dependently influences the gene regulatory network underlying extraembryonic mesoderm development. By engineering human induced pluripotent stem cells (hiPSCs) consisting of wild-type (WT), heterozygous (CDX2-Het), and homozygous null *CDX2* (CDX2-KO) genotypes, differentiating these cells in a 2D gastruloid model, and subjecting these cells to single-nucleus RNA and ATAC sequencing, we identify several pathways that are dose-dependently regulated by CDX2 including VEGF and non-canonical WNT. snATAC-seq reveals that CDX2-Het cells retain a WT-like chromatin accessibility profile, suggesting accessibility alone is not sufficient to drive this variability in gene expression. Because the loss of CDX2 or TBXT phenocopy one another *in vivo*, we compared differentially expressed genes in our CDX2-KO to those from TBXT-KO hiPSCs differentiated in an analogous experiment. This comparison identifies several communally misregulated genes that are critical for cytoskeletal integrity and tissue permeability. Together, these results clarify how CDX2 dose-dependently regulates gene expression in the extraembryonic mesoderm and reveal pathways that may underlie the defects in vascular development and allantoic elongation seen *in vivo*.

KEY WORDS: Extraembryonic mesoderm, Gastrulation, CDX2, Allantois, Gastruloid

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

[1]Gladstone Institute of Cardiovascular Disease, Gladstone Institutes, San Francisco, CA 94158, USA. [2]Developmental and Stem Cell Biology Graduate Program, University of California, San Francisco, CA, 94158, USA. [3]Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA, 94158, USA. [4]Roddenberry Center for Stem Cell Biology and Medicine at Gladstone, San Francisco, CA, 94158, USA. [5]Department of Pediatrics, University of California, San Francisco, CA, 94158, USA. [6]Institute for Human Genetics, University of California, San Francisco, CA, 94158, USA. [7]Eli and Edythe Broad Center for Regeneration Medicine and Stem Cell Research, University of California, San Francisco, CA, 94158, USA.
*Present address: Genentech, South San Francisco, CA, USA.

‡Author for correspondence (benoit.bruneau@gladstone.ucsf.edu)

E.A.B., 0000-0003-2967-1994; T.C.M., 0000-0002-8905-4931; B.G.B., 0000-0002-0804-7597

Biology Open

primitive streak because of a network consisting of BMP, WNT, and NODAL signaling (Arnold and Robertson, 2009). *In vitro*, 48 h 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 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 subsequent 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 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 an analogous study looking 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 associated with impaired allantois formation and vasculogenesis in a dose-dependent manner, including but not limited to genes involved in cell adhesions, motility, and membrane permeability.

RESULTS
Generation of hiPSCs with varying CDX2 dosage and 2D gastruloids
To explore the effect of CDX2 dose on the development of the extraembryonic mesoderm population, we engineered CDX2$^{+/+}$ (WT), CDX2$^{+/-}$ (CDX2-Het), and CDX2$^{-/-}$ (CDX2-KO) human induced pluripotent stem cell lines (hiPSCs) derived from WTC11-LMNB1-GFP parental cell line. 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. 1A; Fig. S1, Table S1, Materials and Methods). All lines were shown to be karyotypically normal (Fig. S1). We conducted a western blot for CDX2 in a sparsely seeded monolayer of cells exposed to BMP4 for 48 h and this revealed the expected stepwise decrease in CDX2 expression across the genotypes, with the highest protein levels in the WT, intermediate levels in CDX2-Het, and no detectable protein in CDX2-KO cells (Fig. 1B). The absence of CDX2 protein in CDX2-KO cells was further confirmed through immunofluorescence (IF) (Fig. 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 each CDX2 genotype into 2D gastruloids via 48 h of BMP4 exposure (Warmflash et al., 2014; Minn et al., 2020) (Fig. 1C,

Materials and Methods). We then conducted two biological replicates of multiomic snRNA-seq and snATAC-seq (Buenrostro et al., 2015) on gastruloids of each genotype after 48 h 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 Bulger et al. (2024), which shares a WT dataset with the current analysis (Fig. 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. 2A-C; Figs S2-S3). 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. 2A′).

The extraembryonic cell type that emerges in 2D gastruloids has 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. S5). 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 (Materials and Methods), are enriched for motifs for the TEAD, TFAP, and HAND families of transcription factors, which are critical regulators of extraembryonic mesoderm fate (Fig. 2D; Figs S3,4). 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-
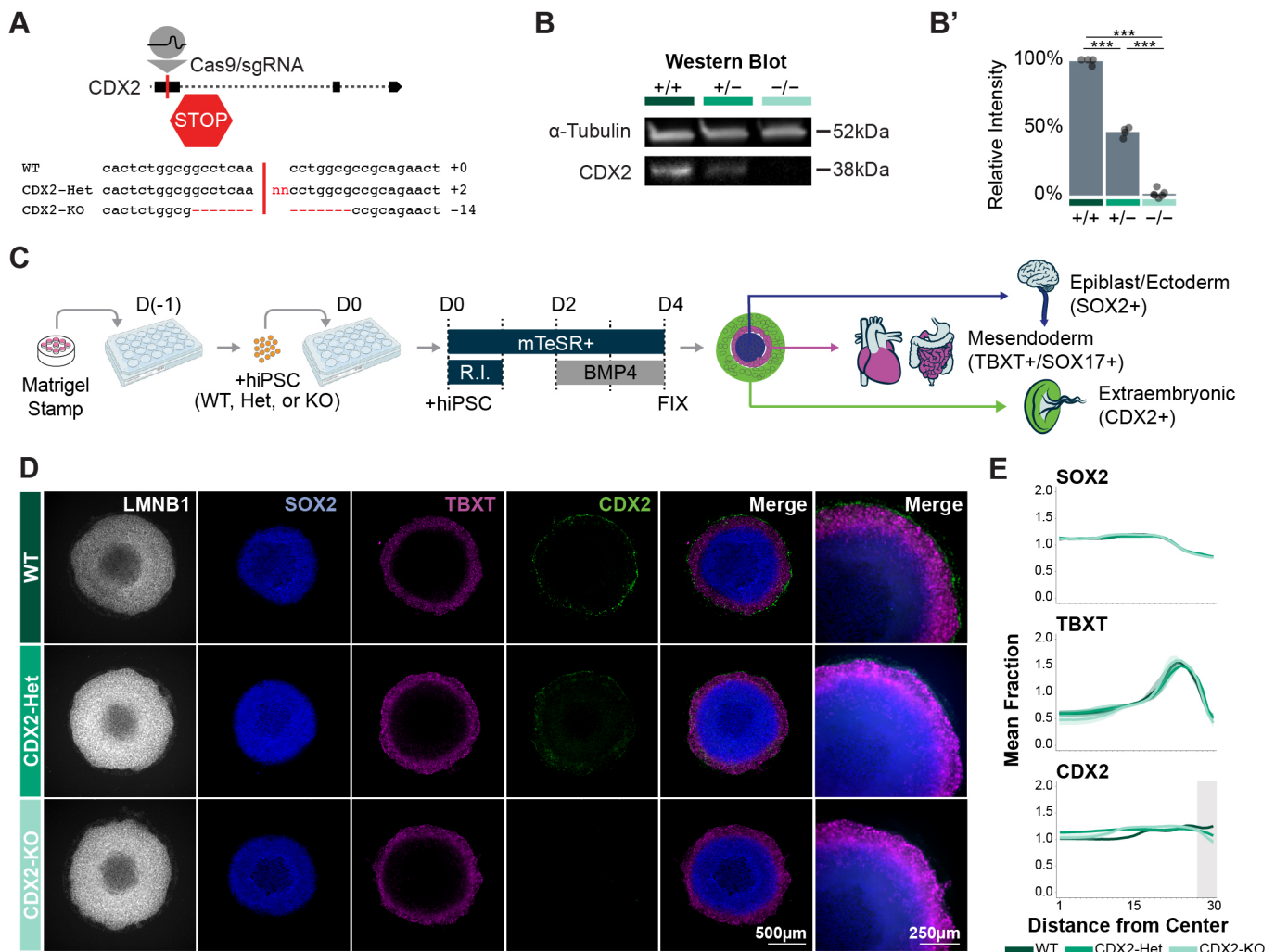
Fig. 1. Generation and validation of the CDX2 genotypes. (A) Schematic of CDX2 locus and sgRNA target in the first exon. Indel generated in CDX2-Het (+2 bp) or CDX2-KO (−14 bp). (B) Western blot and (B′) quantification CDX2 western blot signal intensity in WT, CDX2-Het, and CDX2-KO after 48 h BMP4 exposure. 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.

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. S5B).

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. 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. 2D).

CDX2 primarily influences the GRN underlying extraembryonic mesoderm identity

To understand how CDX2 influences early embryonic morphogenesis, we next isolated each cluster and compared the number of differentially expressed genes (DEGs) between WT versus CDX2-Het, WT versus CDX2-KO, or CDX2-Het versus CDX2-KO. As expected, the ExeM-Late cluster yielded both the highest CDX2 expression in WT cells (Fig. 3A) and the largest number of DEGs between WT and CDX2-KO relative to the other clusters (Table S2). 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. 3B).

Comparing CDX2-Het to WT, we detected 170 genes downregulated and 77 genes upregulated in the CDX2-Het (Fig. 3C; Table S2). Downregulated genes included canonical
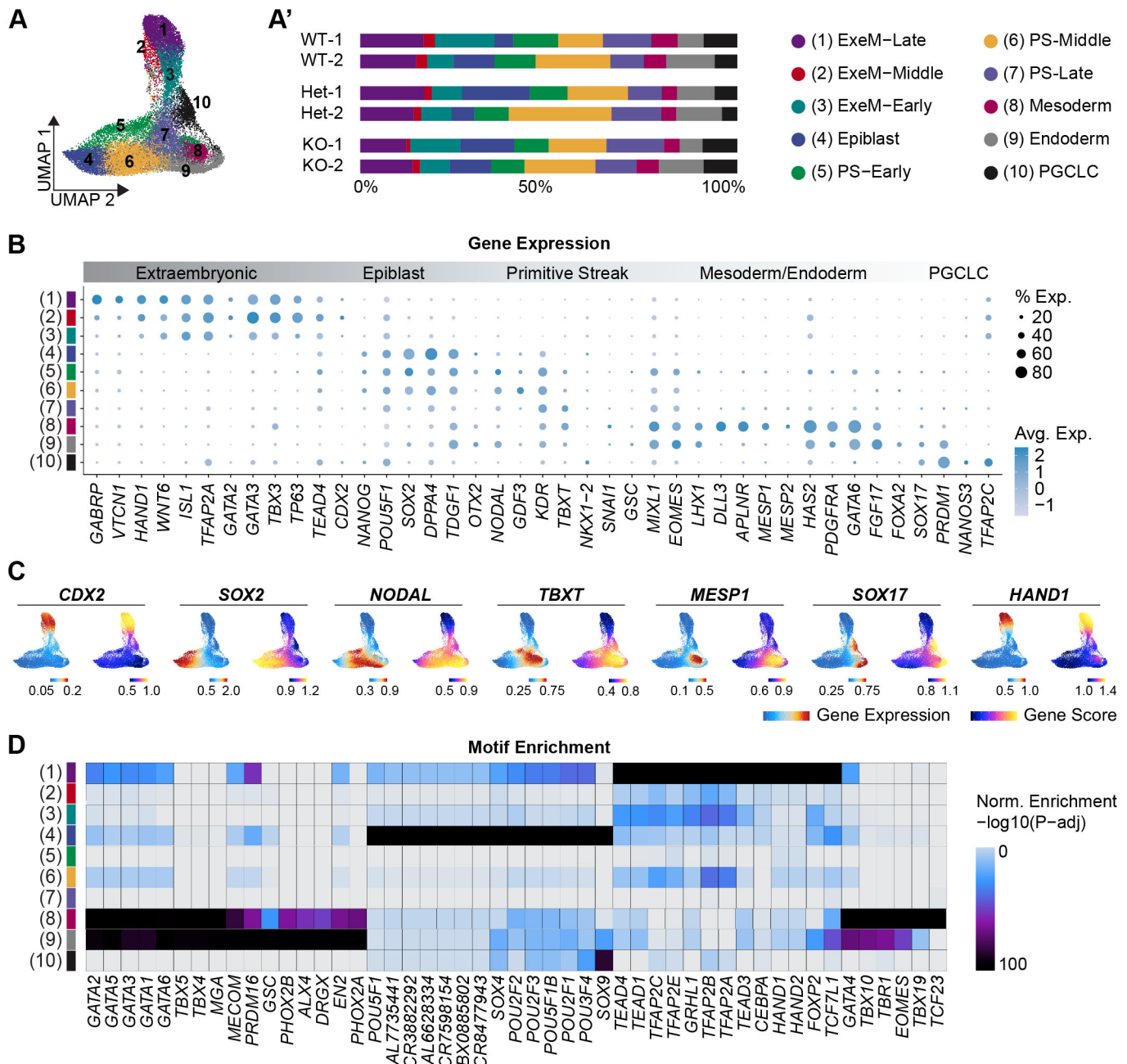
Fig. 2. Lineage emergence is minimally altered in 2D gastruloids of varying CDX2 dose. (A) UMAP reflecting ten clusters comprised of WT, CDX2-Het, and CDX2-KO cells from 2D gastruloids. (A′) Proportion of cells of each sample in each cluster. (B) Dot plot depicting key lineage markers across all ten 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 in each cluster [FDR≤0.1 & abs(Log2FC)≥0.5, maximum motifs per cluster plotted=20].

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. 3D; Table S3). 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. 3C′; Table S2). 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. 3C′; Table S2). Genes involved in cell migration and axon guidance such as *GLI3*, *UNC5B*, *SEMA3E*, and *ISL1* were also identified. GO biological process enrichment revealed a fourfold enrichment of genes involved in 'canonical WNT signaling', as well as enrichment for 'vasculature development', 'circulatory system development',
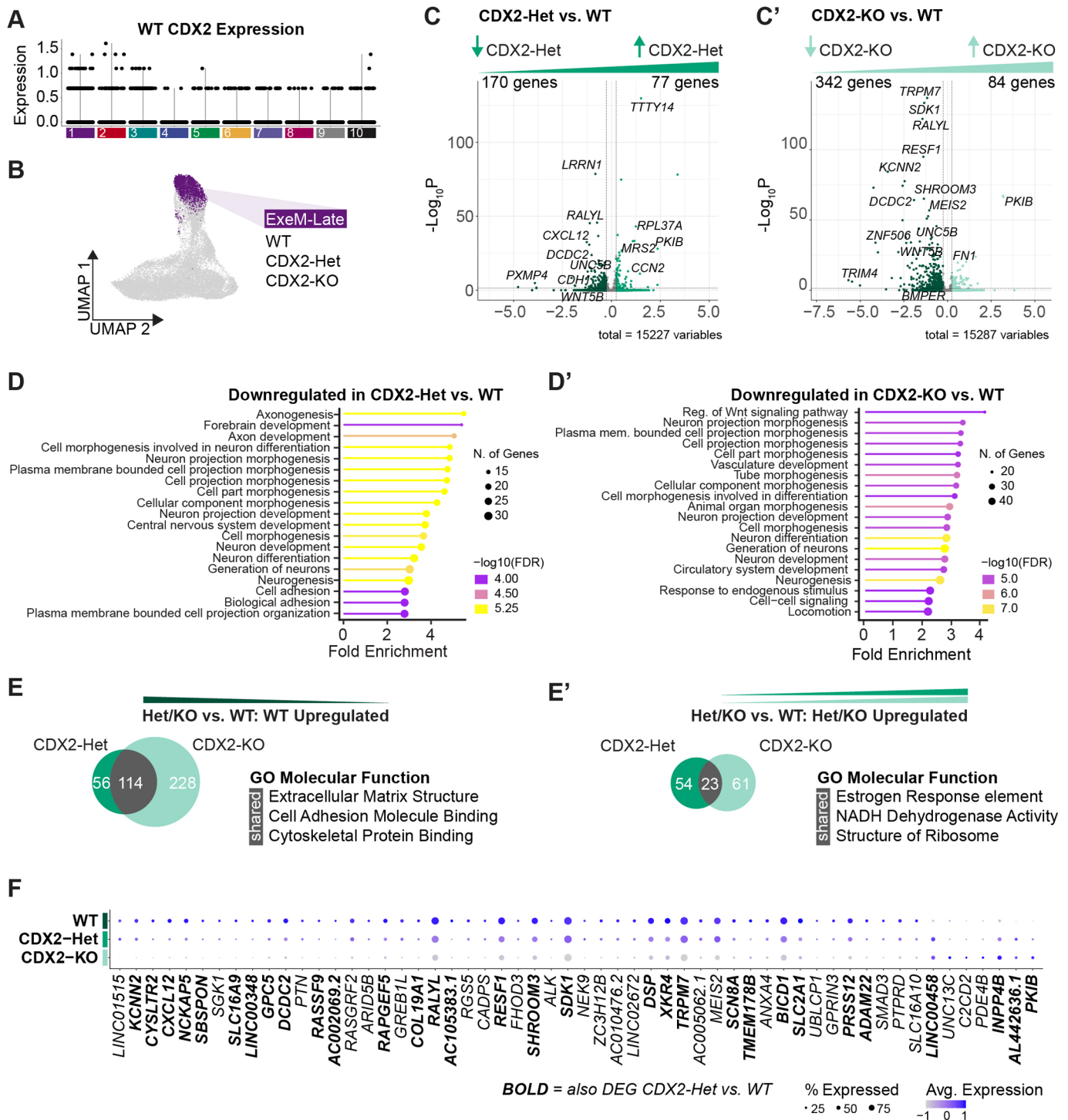
Fig. 3. CDX2 dose-dependently influences extraembryonic mesoderm gene expression. (A) Violin plot of CDX2 expression across clusters in WT cells. (B) ExeM-Late cluster isolated for genotype-specific analyses. (C) Volcano plot of differentially expressed genes comparing WT and CDX2-Het or (C′) WT and CDX2-KO [abs(Log2FC)>0.25, P-adj<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, P-adj<0.05, ShinyGO FDR<0.05] within the ExeM-Late cluster. (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 versus WT and CDX2-KO versus WT.

'tube morphogenesis', 'cell-cell signaling', and 'locomotion' (Fig. 3D′; Fig. S7, Table S3). Taken together, the differentially expressed genes that are seen in both CDX2-Het and CDX2-KO versus WT correspond to several 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 versus WT and the CDX2-KO versus 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. 3E,F; Table S2). GO molecular function analysis of these 114 genes revealed significant enrichment for 'extracellular matrix structure', 'cell adhesion molecule binding', and 'cytoskeletal protein binding' (Fig. 3E; Fig. S7, Table S3). GO biological process enrichment for 'vasculature development' remained enriched by approximately threefold (FDR=0.03), and of the 29 genes driving 'vasculature development' enrichment in CDX2-KO, ten remained differentially expressed in CDX2-Het (*ISL1*, *ATP2B4*, *GLI3*, *ENG*, *UNC5B*, *GPLD1*, *RTN4*, *COL4A2*, *HEY1*, *SEMA3E*) (Table S2). 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. 3E′; Table S3). 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. 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 both physical cell structure and 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) (Table S2). 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. 4A; Table S2). 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 maintaining accessibility in the extraembryonic mesoderm population. Additionally, the lack of DARs detected in the CDX2-Het versus WT comparison indicates that an intermediate level of CDX2 is sufficient to induce a WT-like chromatin accessibility profile (Fig. 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(*P*-adj) ~3] and the low number of DARs driving these results makes their biological relevance uncertain (Fig. 4C; Fig. S8). 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(*P*-adj) ~60] (Fig. 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 et al., 2009; Bulajić et al., 2020). Because of the limited detection of posterior HOX gene expression in our datasets (Fig. S6A), we believe that these hits 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. 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 versus 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 expression within the HOXB locus

Because CDX and posterior HOX motifs share a common binding motif that was identified as 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 the allantois specifically, posterior HOXA genes have been shown to be crucial for the expansion of fetal vasculature across amniotes and are expressed both before and after chorioallantoic fusion (Scotti and Kmita, 2012).
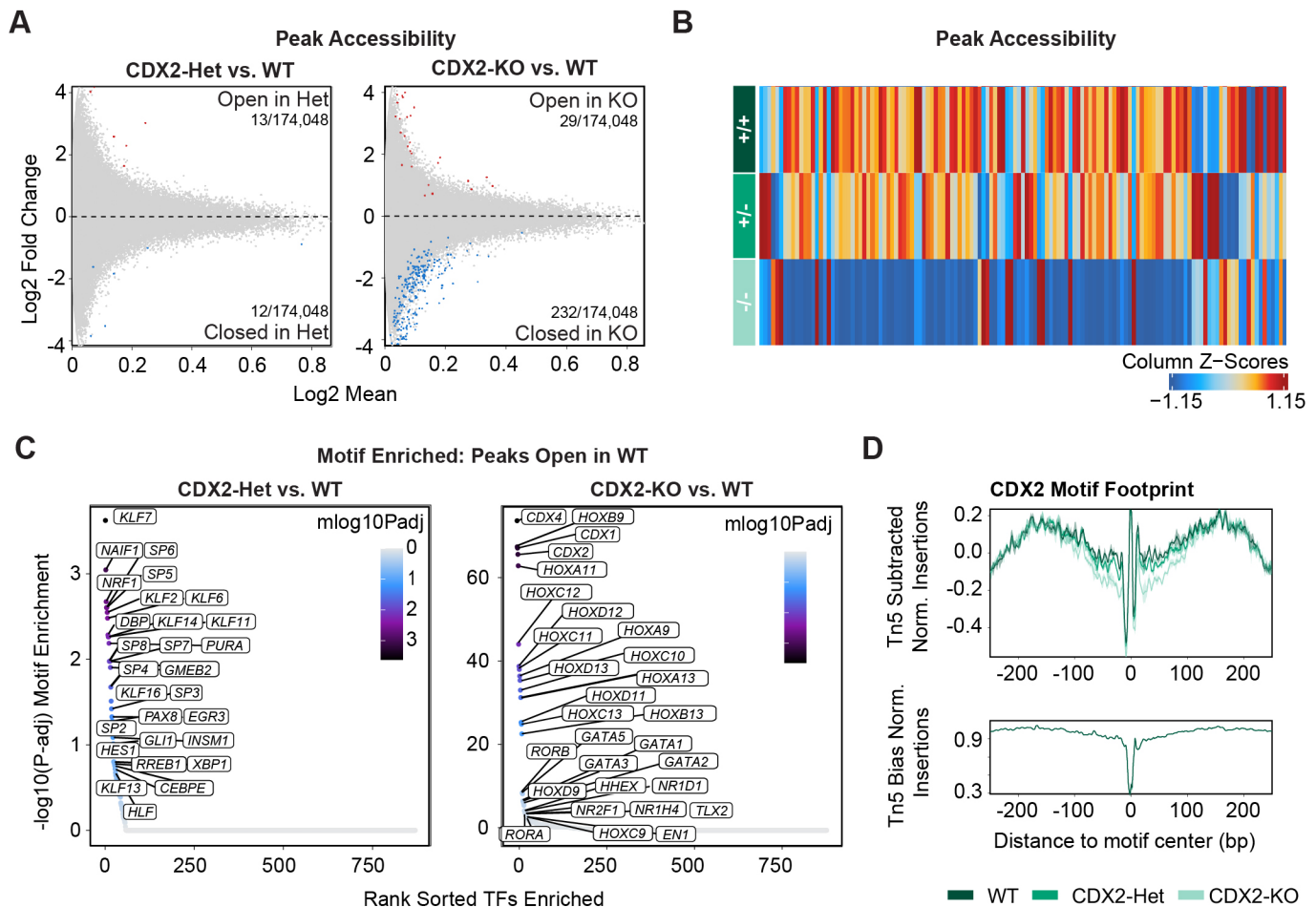
Fig. 4. CDX2 expression influences chromatin organization at regions containing CDX motifs. (A) Differential peak accessibility comparing CDX2-Het and WT (left) or CDX2-KO and WT (right) within the ExeM-Late cluster [FDR<0.1 & abs(Log2FC)>0.5]. Red, regions more accessible in mutant; blue, regions more 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).

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. S6A). 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 et al., 2015). Alternatively, HOX paralogs have been shown to exert different roles in different tissues (Kachgal et al., 2012), and *HOXB3*, *HOXB5*, and *HOXB7* specifically have been shown to impact vasculogenesis (Miano et al., 1996; Myers et al., 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. S6A). 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 and CDX2-Het than 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 et al., 2015) (Fig. S6A). The posterior HOX genes, specifically *HOXB9*, *HOXA13*, and *HOXC13*, were detected in very few cells. Taken together, these results, while subtle, illustrate how reduced CDX2 dose-dependently limits the expression of downstream central HOX genes.

Overall, these results reflect the ability of CDX2 to augment HOX expression, including the downregulation of central HOX genes and slight 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 juxtacrine 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. However, the non-canonical WNT (ncWNT) signaling pathway has reduced information flow in both CDX2-Het and CDX2-KO (Fig. 5A). CellChat defines the ncWNT signaling pathway broadly based on the expression of WNT5A and WNT5B as ligands and FZD family members as receptors. These ligands and receptors coincide with several non-canonical WNT pathways important during embryogenesis, including both the WNT/calcium pathway and the WNT/PCP signaling pathway, the latter of which has frequently been shown to be crucial for in directional cell migration, cell polarity, and convergent extension. We visualized the precise changes in predicted communication between and within clusters in circle plots, where line thickness correlates with the degree of predicted communication. This analysis revealed that 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. 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 near eliminated signaling between WNT5B and those same receptors, suggesting a critical role for WNT5B in the maintenance of the ncWNT pathway (Fig. 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 et al., 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 recently conducted a parallel analysis of a TBXT dosage series in the 2D gastruloid model that uses the same WT cell population as this current study (Bulger et al., 2024). We isolated the analogous extraembryonic mesoderm population from this TBXT dataset ('Extraembryonic-Late'), conducted TBXT-Het versus WT and TBXT-KO versus WT comparisons within this cluster, and looked at how the resulting lists of differentially expressed genes intersected with those of the CDX2 dosage series ExeM-Late population.

From this comparative analysis, we uncovered 11 genes downregulated and 32 genes upregulated in TBXT-Het versus WT and 35 downregulated and 49 upregulated genes in TBXT-KO relative to WT (Fig. 6A,C; Table S4). Of the 35 genes downregulated in TBXT-KO, eight were also downregulated in CDX2-KO relative to WT. These eight 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. 6B). Five of these eight genes were also significantly downregulated in CDX2-Het and zero were significantly downregulated in TBXT-Het, although trends toward intermediate expression are evident (*P*-adj<0.05, Log2FC>0.25) (Fig. 6B; Table S4). Notably, both *ANK3* and *ANGPT1* have been shown to regulate angiogenesis *in vivo* via the VEGF signaling pathway (Gavard et al., 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. 6C). These nine genes included the cytoskeleton modulator *TMSB10*, hedgehog pathway and myogenesis effector *CDON*, and placental adhesion regulators *FBLN1* and *IGFBP7* (Fig. 6D; Table S4). 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 nine genes were also significantly upregulated in CDX2-Het and zero were upregulated in TBXT-Het (Table S4). While speculative based on bioinformatics analysis, these results suggest that one method by which CDX2 and TBXT regulate the development of extraembryonic mesoderm structures could be 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 potentially 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.

A

B

ncWNT Signaling Pathway

C

D

ncWNT Signaling Pathway
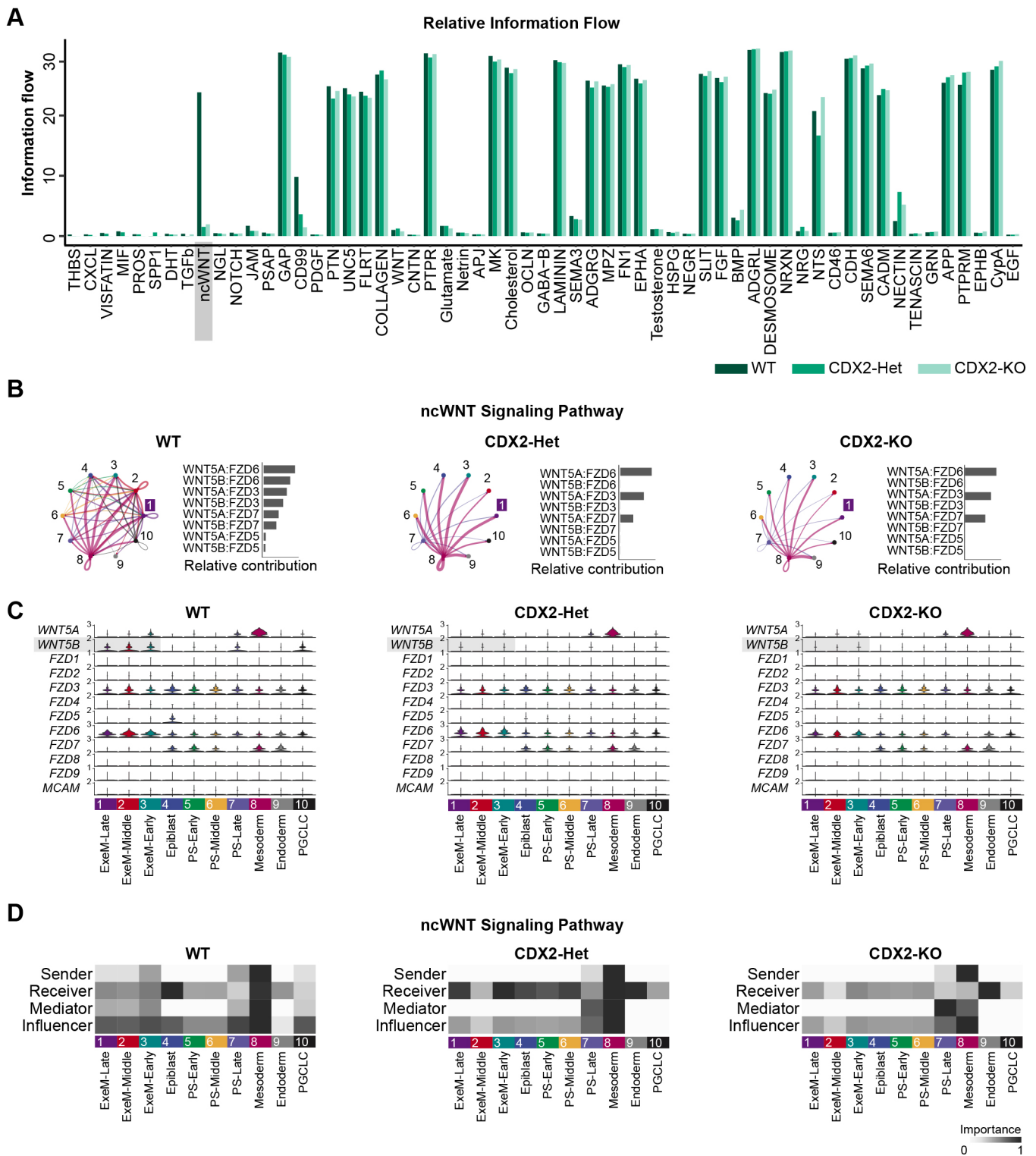
Fig. 5. Cellchat reveals the CDX2-dose-dependent regulation of the ncWNT signaling pathway. (A) Information flow (arbitrary units; A.U.) of various pathways across all clusters within WT, CDX2-Het, or CDX2-KO populations. (B) Circle plots (left) and bar plots (right) visualizing ligand-receptor communication 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.

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 six peaks with decreased accessibility and five peaks with increased accessibility in WT relative to TBXT-KO within the equivalent cluster. Of these 11
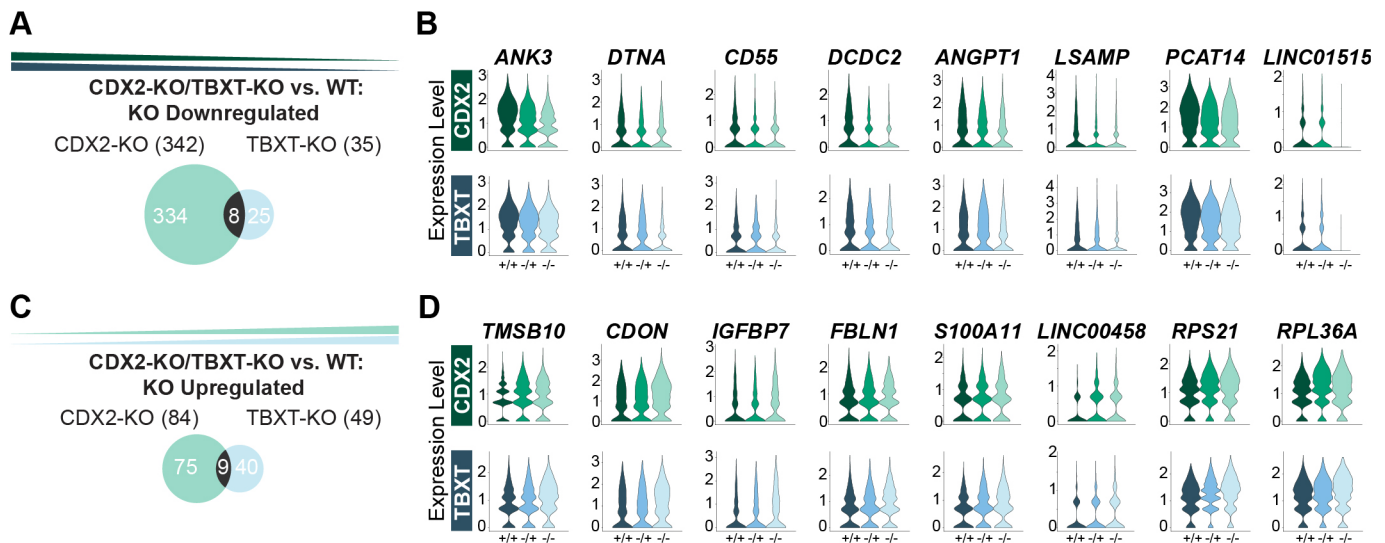
Fig. 6. TBXT and CDX2 mutants both misregulate genes associated with the VEGF signaling pathway. (A) Venn diagram illustrating overlapping genes downregulated or (C) upregulated in CDX2-KO (left) or TBXT-KO (right) relative to WT in the ExeM-Late cluster. (B) Violin plots of gene expression for overlapping genes identified in A. MT-ND3 shown in Fig. S9. (D) Violin plots of gene expression for overlapping genes identified in C.

regions, five 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 specifically 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 understand the gene regulatory network underlying the development of these structures and how it is dysregulated in the absence of CDX2. We additionally asked the extent to which this network is affected in a dose-dependent manner, motivated by studies showing both a correlation between CDX2 expression levels and the development of embryonic mesoderm, including elongation of the tailbud and

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 two-thirds 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, even if 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 et al., 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 the R package, 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 differentially expressed genes 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 clarify 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.

MATERIALS AND METHODS
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 *CDX2* [*CDX2⁺/⁻ (CDX2-Het) and CDX2⁻/⁻ (CDX2-KO)*] were generated for this study and derived from the Allen Institute WTC11-LaminB parental cell line (AICS-0013 cl.210). The WT line was derived from a WTC11-LaminB subclone that was exposed to a TBXT sgRNA [CAGAGCGCGAACTGCGCGTG; a gift from Jacob Hanna (Addgene plasmid #5972)] but remained unedited. 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 cm². After 24 h, cells were maintained in mTeSR-Plus media until 80% confluent.

Generation of WT, CDX2-Het, and CDX2-KO hiPSC lines
To generate the CDX2 dosage series we first lipofected WTC11-LaminB cells (AICS-0013 cl.210) with 125 ng sgRNA and 500 ng 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 62.5 ng each: (CDX2-A: CCUAGCUCC-GUGCGCCACUC and CDX2-B: AGUUCUGCGGCGCCAGGUUG). After recovery for 48 h in mTeSR-Plus supplemented with ROCK inhibitor, lipofected cells were dissociated using Accutase and passaged into a GFR-Matrigel coated 10 cm dish, where they were expanded for 24 h in mTeSR-Plus with 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 *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 (For1: cctcgacgtctccaaccattg and Rev1: gcctctgcttaccttggctg) and sequenced using Rev1. 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 S1). To eliminate the possibility of a heterozygous line being a mixed population of wild-type and homozygous null alleles, eight subclones of the prospective heterozygous line were isolated and Sanger sequenced as before, and all subclones were confirmed to contain identical genotypes (Table S1). After sequencing confirmation of respective genotypes, karyotypically normal cells from each hiPSC line (Fig. S1) were expanded for subsequent studies.

PDMS stamp fabrication
Stamps were fabricated as described in Bulger et al. (2024). 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 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, USA) 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
Matrigel stamps were applied as described in Bulger et al. (2024). PDMS stamps (each containing 12×1000 µM 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 h. 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 1 h 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 h before being stored in DPBS−/− solution at 4°C for up to 2 weeks.

Confined 2D gastruloid differentiation

2D gastruloids were differentiated as described in Bulger et al. (2024). hiPSCs were dissociated with Accutase and resuspended in mTeSR-Plus supplemented with ROCK inhibitor. 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 h before the well was rinsed 1x with DPBS and given fresh mTeSR-Plus supplemented with ROCK inhibitor. Approximately 24 h post-seeding, media was exchanged for mTeSR-Plus. After another 24 h or upon confluency of the stamped colony, media was exchanged for mTeSR-Plus supplemented with BMP4 (50 ng/ml). Colonies were allowed to differentiate in the presence of BMP4 for 48 h before being processed for downstream analyses.

Western blot

Cells of each genotype were induced to form CDX2+ extraembryonic mesoderm in a sparsely seeded monolayer exposed to 50 ng/μl BMP4 in mTeSR+ for 48 h before 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 (Bio-Rad; 1704157). The membrane was then blocked overnight at 4°C using Intercept TBS Blocking Buffer (Li-COR; 927-70001). Primary antibodies CDX2 (12306; 1:1000; Rb), ß-Actin (ab8226; 1:1000; Ms), or α-Tubulin (T5168; 1:1000; 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 h at RT in the dark with species-specific secondary antibodies (Rb-680; 926-68071; Ms-800 926-32212) (VWR) at 1:10,000. Membranes were subsequently washed and developed using the Bio-Rad 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

Immunofluorescence was conducted as described in Bulger et al. (2024). hiPSCs were rinsed with PBS once, fixed in 4% paraformaldehyde (VWR) for 15-20 min, and subsequently washed three times 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 SOX2 (3579s; 1:200; Rb) CDX2 (ab157524; 1:500; Ms), or TBXT (AF2085; 1:400; Gt) diluted in antibody dilution buffer (0.3% Triton, 1% BSA in PBS) overnight. The following day, samples were washed three times with PBS and incubated with secondary antibodies in antibody dilution buffer at room temperature for 2 h. 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).

Cell harvesting for single nuclei multiome ATAC+RNA sequencing

Each of the *CDX2* genotypes was differentiated, harvested, and prepared at the same time for each of the two biological replicates, as described in Bulger et al. (2024). Therefore, each set of biological replicates represents an experimental batch. For each genotype (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 ~10,000-19,000 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). WT-1 (eb01)=12,072 nuclei, 20,556.09 reads/nucleus. CDX2-KO (eb04)=18,239 nuclei, 15,288.63 reads/nucleus. CDX2-Het-1 (eb05)=18,806 nuclei, 19,435.86 reads/nuclei. WT-2 (eb06)=10,903 nuclei, 26,003.63 reads/nucleus. CDX2-KO-2 (eb09)=12,314 nuclei, 20,150.10 reads/nucleus. CDX2-Het-2 (eb10)=11,602 nuclei, 18,929.24 reads/nucleus. 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.

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

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 and nFeature_RNA<4500, *n*Count_RNA>200 and *n*Count_RNA<12,000, 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", "G2 M. 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 ten 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

Indexed Fragment files generated by the CellRanger-Arc counts function served as input for the generation of sample-specific ArrowFiles (minTSS=4 and minFrags=1000) using the R package ArchR v1.0.2 (Granja et al., 2021). ArrowFile creation also generates a genome wide TileMatrix using 500 bp 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 500 bp

TileMatrix with default settings (iterations=2, sampleCells=10,000, n.start=10, resolution=2, maxClusters=6). This was repeated using the Gene Expression Matrix based on 2500 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 3792 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 versus CDX2-KO or WT versus 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].

Gene Ontology analysis
Gene Ontology (GO) analysis for downregulated or upregulated CDX2-dependent genes was performed with ShinyGO V0.77 (Ge et al., 2020) using GO Biological Process or Molecular Function terms (FDR<0.05, pathways size 2-2000). Downregulated or upregulated CDX2-dependent gene lists from the ExeM-Late subcluster were assembled from differential tests between CDX2-Het versus WT or CDX2-KO versus WT in Seurat. Gene sets were filtered with a significance threshold set at an adjusted P-value>0.05 and the abs(Log2FC)>0.25. The first 20 hits for the CDX2-KO or CDX2-Het versus WT comparison were visualized with lollipop plots. The process was repeated for genes identified as overlapping in both TBXT-KO and CDX2-KO versus WT comparisons.

CellChat
Cell signaling analysis was performed using the R package CellChat (Jin et al., 2021) as described in Bulger et al. (2024). The Seurat object containing all samples was subset by genotype, yielding a separate Seurat object for WT, CDX2-Het, or CDX2-KO. These three 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, CDX2-Het, and CDX2-KO datasets independently and differential signaling analysis was then run by merging the WT, CDX2-Het, and CDX2-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 and heatmaps of pathways of interest were then generated for each genotype separately using the standard settings for netVisual_aggregate(cellchat) or netVisual_heatmap(cellchat), respectively. Violin plots of differential gene expression were generated using plotGeneExpression(cellchat) with the standard settings.

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 post hoc 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 P-adj <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.

Data and reagent availability
snATAC-seq and snRNA-seq data have been deposited in GEO under the accession number GSE245998 (WT) and GSE251813 (CDX2-Het and CDX2-KO). Analysis scripts used to generate figure panels and relevant cell lines are available from the authors upon request. R scripts for sequencing analyses can be found at https://github.com/ebulger3/TBXT-CDX2-Analysis.

Competing interests
B.G.B. is a founder, shareholder, and advisor of Tenaya Therapeutics and is an advisor for Silver Creek Pharmaceuticals. T.C.M. is an employee at Genentech and is an advisor for Vitra Labs. The work presented here is not related to the interests of these commercial entities.

Author contributions
Conceptualization: E.A.B. , T.C.M., and B.G.B.; Methodology: E.A.B.; Software: E.A.B.; Validation: E.A.B.; Formal analysis: E.A.B.; Investigation: E.A.B.; Resources: E.A.B., T.C.M., B.G.B.; Data curation: E.A.B.; Writing - original draft: E.A.B.; Writing - review & editing: E.A.B., T.C.M., B.G.B.; Visualization: E.A.B.; Supervision: T.C.M., B.G.B.; Project administration: E.A.B., T.C.M., B.G.B.; Funding acquisition: E.A.B., T.C.M., B.G.B.

Data availability

All relevant data can be found within the article – see 'Data and reagent availability – and its supplementary information.

First Person

This article has an associated First Person interview with the first author of the paper.

References

Akoumianakis, I., Polkinghorne, M. and Antoniades, C. (2022). Non-canonical WNT signalling in cardiovascular disease: mechanisms and therapeutic implications. *Nat. Rev. Cardiol.* **19**, 783-797. doi:10.1038/s41569-022-00718-5

Alom Ruiz, S. and Chen, C. S. (2007). Microcontact printing: a tool to pattern. *Soft Mat.* **3**, 168-177. doi:10.1039/B613349E

Amin, S., Neijts, R., Simmini, S., van Rooijen, C., Tan, S. C., Kester, L., van Oudenaarden, A., Creyghton, M. P. and Deschamps, J. (2016). Cdx and T brachyury co-activate growth signaling in the embryonic axial progenitor niche. *Cell Rep.* **17**, 3165-3177. doi:10.1016/j.celrep.2016.11.069

Anand, G. M., Megale, H. C., Murphy, S. H., Weis, T., Lin, Z., He, Y., Wang, X., Liu, J. and Ramanathan, S. (2023). Controlling organoid symmetry breaking uncovers an excitable system underlying human axial elongation. *Cell* **186**, 497-512.e23. doi:10.1016/j.cell.2022.12.043

Arnold, S. J. and Robertson, E. J. (2009). Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. *Nat. Rev. Mol. Cell Biol.* **10**, 91-103. doi:10.1038/nrm2618

Babcock, K., Rosner, I. L., Srivastava, S., Dobi, A. and Tan, S.-H. (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 Res.* **78** 13_Suppl., 4486-4486. doi:10.1158/1538-7445.AM2018-4486

Beck, F., Erler, T., Russell, A. and James, R. (1995). Expression of Cdx-2 in the mouse embryo and placenta: possible role in patterning of the extra-embryonic membranes. *Dev. Dyn.* **204**, 219-227. doi:10.1002/aja.1002040302

Berger, M. F., Badis, G., Gehrke, A. R., Talukder, S., Philippakis, A. A., Peña-Castillo, L., Alleyne, T. M., Mnaimneh, S., Botvinnik, O. B., Chan, E. T. et al. (2008). Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell* **133**, 1266-1276. doi:10.1016/j.cell.2008.05.024

Bernardo, A. S., Faial, T., Gardner, L., Niakan, K. K., Ortmann, D., Senner, C. E., Callery, E. M., Trotter, M. W., Hemberger, M., Smith, J. C. 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**, 144-155. doi:10.1016/j.stem.2011.06.015

Brooke-Bisschop, T., Savory, J. G. A., Foley, T., Ringuette, R. and Lohnes, D. (2017). Essential roles for Cdx in murine primitive hematopoiesis. *Dev. Biol.* **422**, 115-124. doi:10.1016/j.ydbio.2017.01.002

Buenrostro, J. D., Wu, B., Chang, H. Y. and Greenleaf, W. J. (2015). ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* **109**, 21.29.1-21.29.9. doi:10.1002/0471142727.mb2129s109

Bulajić, M., Srivastava, D., Dasen, J. S., Wichterle, H., Mahony, S. and Mazzoni, E. O. (2020). Differential abilities to engage inaccessible chromatin diversify vertebrate Hox binding patterns. *Development* **147**, dev194761. doi:10.1242/dev.194761

Bulger, E. A., Muncie-Vasic, I., Libby, A. R. G., McDevitt, T. C. and Bruneau, B. G. (2024). TBXT dose sensitivity and the decoupling of nascent mesoderm specification from EMT progression in 2D human gastruloids. *Development*. doi:10.1242/dev.202516

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

Cao, Y., Linden, P., Farnebo, J., Cao, R., Eriksson, A., Kumar, V., Qi, J.-H., Claesson-Welsh, L. and Alitalo, K. (1998). Vascular endothelial growth factor C induces angiogenesis in vivo. *Proc. Natl. Acad. Sci. USA* **95**, 14389-14394. doi:10.1073/pnas.95.24.14389

Cao, J., Ehling, M., März, S., Seebach, J., Tarbashevich, K., Sixta, T., Pitulescu, M. E., Werner, A.-C., Flach, B., Montanez, E. et al. (2017). Polarized actin and VE-cadherin dynamics regulate junctional remodelling and cell migration during sprouting angiogenesis. *Nat. Commun.* **8**, 2210. doi:10.1038/s41467-017-02373-8

Chawengsaksophak, K., James, R., Hammond, V. E., Köntgen, F. and Beck, F. (1997). Homeosis and intestinal tumours in Cdx2 mutant mice. *Nature* **386**, 84-87. doi:10.1038/386084a0

Chawengsaksophak, K., de Graaff, W., Rossant, J., Deschamps, J. and Beck, F. (2004). Cdx2 is essential for axial elongation in mouse development. *Proc. Natl. Acad. Sci. USA* **101**, 7641-7645. doi:10.1073/pnas.0401654101

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**, 473-479. doi:10.1038/350473a0

d'Apolito, M., Santacroce, R., Colia, A. L., Cordisco, G., Maffione, A. B. and Margaglione, M. (2019). Angiopoietin-1 haploinsufficiency affects the endothelial barrier and causes hereditary angioedema. *Clin. Exp. Allergy* **49**, 626-635. doi:10.1111/cea.13349

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

Downs, K. M., Inman, K. E., Jin, D. X. and Enders, A. C. (2009). The Allantoic Core Domain: new insights into development of the murine allantois and its relation to the primitive streak. *Dev. Dyn.* **238**, 532-553. doi:10.1002/dvdy.21862

Drake, C. J., LaRue, A., Ferrara, N. and Little, C. D. (2000). VEGF regulates cell behavior during vasculogenesis. *Dev. Biol.* **224**, 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 Genet.* **10**, 358-364. doi:10.1016/0168-9525(94)90132-5

Durak, O., de Anda, F. C., Singh, K. K., Leussis, M. P., Petryshen, T. L., Sklar, P. and Tsai, L.-H. (2015). Ankyrin-G regulates neurogenesis and Wnt signaling by altering the subcellular localization of β-catenin. *Mol. Psychiatry* **20**, 388-397. doi:10.1038/mp.2014.42

Ekker, S. C., Jackson, D. G., von Kessler, D. P., Sun, B. I., Young, K. E. and Beachy, P. A. (1994). The degree of variation in DNA sequence recognition among four Drosophila homeotic proteins. *EMBO J.* **13**, 3551-3560. doi:10.1002/j.1460-2075.1994.tb06662.x

Foley, T. and Lohnes, D. (2022). Cdx regulates gene expression through PRC2-mediated epigenetic mechanisms. *Dev. Biol.* **483**, 22-33. doi:10.1016/j.ydbio.2021.12.014

Foley, T. E., Hess, B., Savory, J. G. A., Ringuette, R. and Lohnes, D. (2019). Role of Cdx factors in early mesodermal fate decisions. *Development* **146**, dev170498. doi:10.1242/dev.170498

Galceran, J., Farinas, I., Depew, M. J., Clevers, H. and Grosschedl, R. (1999). Wnt3a-/–like phenotype and limb deficiency in Lef1(−/−)Tcf1(−/−) mice. *Genes Dev.* **13**, 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. *Dev. Cell* **14**, 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**, 2628-2629. doi:10.1093/bioinformatics/btz931

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

Hayashi, M., Kawaguchi, T., Durcova-Hills, G. and Imai, H. (2018). Generation of germ cells from pluripotent stem cells in mammals. *Reprod. Med. Biol.* **17**, 107-114. doi:10.1002/rmb2.12077

Iimura, T. and Pourquié, O. (2006). Collinear activation of Hoxb genes during gastrulation is linked to mesoderm cell ingression. *Nature* **442**, 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**, 2947-2959. doi:10.1242/dev.02454

Jin, S., Guerrero-Juarez, C. F., Zhang, L., Chang, I., Ramos, R., Kuan, C.-H., Myung, P., Plikus, M. V. and Nie, Q. (2021). Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* **12**, 1088. doi:10.1038/s41467-021-21246-9

Kachgal, S., Mace, K. A. and Boudreau, N. J. (2012). The dual roles of homeobox genes in vascularization and wound healing. *Cell Adhes. Migr.* **6**, 457-470. doi:10.4161/cam.22164

Kanazawa, A., Tsukada, S., Kamiyama, M., Yanagimoto, T., Nakajima, M. and Maeda, S. (2005). Wnt5b partially inhibits canonical Wnt/β-catenin signaling pathway and promotes adipogenesis in 3T3-L1 preadipocytes. *Biochem. Biophys. Res. Commun.* **330**, 505-510. doi:10.1016/j.bbrc.2005.03.007

Kasprzak, A. (2020). Angiogenesis-related functions of Wnt signaling in colorectal carcinogenesis. *Cancers* **12**, 3601. doi:10.3390/cancers12123601

Krumlauf, R. (1993). Mouse Hox genetic functions. *Curr. Opin. Genet. Dev.* **3**, 621-625. doi:10.1016/0959-437x(93)90098-a

Kumar, N., Tsai, Y.-H., Chen, L., Zhou, A., Banerjee, K. K., Saxena, M., Huang, S., Toke, N. H., Xing, J., Shivdasani, R. A. et al. (2019). The lineage-specific transcription factor CDX2 navigates dynamic chromatin to control distinct stages of intestine development. *Development* **146**, dev172189. doi:10.1242/dev.172189

Lee, S.-H., Son, M. J., Oh, S.-H., Rho, S.-B., Park, K., Kim, Y.-J., Park, M.-S. and Lee, J.-H. (2005). Thymosin {beta}(10) inhibits angiogenesis and tumor growth by interfering with Ras function. Cancer Res. 65, 137-148. doi:10.1158/0008-5472.137.65.1

Liu, J., Liao, X., Zhou, J., Li, B., Xu, L., Liu, S., Li, Y., Yuan, D., Hu, C., Jiang, W. et al. (2021). A rare variant of ANK3 is associated with intracranial aneurysm. Front. Neurol. 12, 672570. doi:10.3389/fneur.2021.672570

Ludwig, T., Levenstein, M., Jones, J. et al. (2006). Derivation of human embryonic stem cells in defined conditions. Nat Biotechnol 24, 185-187. doi:10.1038/nbt1177

Lufkin, T., Dierich, A., LeMeur, M., Mark, M. and Chambon, P. (1991). Disruption of the Hox-1.6 homeobox gene results in defects in a region corresponding to its rostral domain of expression. Cell 66, 1105-1119. doi:10.1016/0092-8674(91)90034-v

Lufkin, T., Mark, M., Hart, C. P., Dollé, P., LeMeur, M. and Chambon, P. (1992). Homeotic transformation of the occipital bones of the skull by ectopic expression of a homeobox gene. Nature 359, 835-841. doi:10.1038/359835a0

Ma, H., Zhai, J., Wan, H., Jiang, X., Wang, X., Wang, L., Xiang, Y., He, X., Zhao, Z.-A., Zhao, B. et al. (2019). In vitro culture of cynomolgus monkey embryos beyond early gastrulation. Science 366, eaax7890. doi:10.1126/science.aax7890

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

Miano, J. M., Firulli, A. B., Olson, E. N., Hara, P., Giachelli, C. M. and Schwartz, S. M. (1996). Restricted expression of homeobox genes distinguishes fetal from adult human smooth muscle cells. Proc. Natl. Acad. Sci. U.S.A. 93, 900-905. doi:10.1073/pnas.93.2.900

Minn, K. T., Fu, Y. C., He, S., Dietmann, S., George, S. C., Anastasio, M. A., Morris, S. A. and Solnica-Krezel, L. (2020). High-resolution transcriptional and morphogenetic profiling of cells from micropatterned human ESC gastruloid cultures. eLife 9, e59445. doi:10.7554/eLife.59445

Minn, K. T., Dietmann, S., Waye, S. E., Morris, S. A. and Solnica-Krezel, L. (2021). Gene expression dynamics underlying cell fate emergence in 2D micropatterned human embryonic stem cell gastruloids. Stem Cell Rep. 16, 1210-1227. doi:10.1016/j.stemcr.2021.03.031

Myers, C., Charboneau, A. and Boudreau, N. (2000). Homeobox B3 promotes capillary morphogenesis and angiogenesis. J. Cell Biol. 148, 343-352. doi:10.1083/jcb.148.2.343

Neijts, R., Amin, S., van Rooijen, C. and Deschamps, J. (2017). Cdx is crucial for the timing mechanism driving colinear Hox activation and defines a trunk segment in the Hox cluster topology. Dev. Biol. 422, 146-154. doi:10.1016/j.ydbio.2016.12.024

Nguyen, T. T., Savory, J. G. A., Brooke-Bisschop, T., Ringuette, R., Foley, T., Hess, B. L., Mulatz, K. J., Trinkle-Mulcahy, L. and Lohnes, D. (2017). Cdx2 regulates gene expression through recruitment of Brg1-associated Switch-Sucrose Non-fermentable (SWI-SNF) chromatin remodeling activity. J. Biol. Chem. 292, 3389-3399. doi:10.1074/jbc.M116.752774

Niu, Y., Sun, N., Li, C., Lei, Y., Huang, Z., Wu, J., Si, C., Dai, X., Liu, C., Wei, J. et al. (2019). Dissecting primate early post-implantation development using long-term in vitro embryo culture. Science 366, eaaw5754. doi:10.1126/science.aaw5754

Niwa, H., Toyooka, Y., Shimosato, D., Strumpf, D., Takahashi, K., Yagi, R. and Rossant, J. (2005). Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. Cell 123, 917-929. doi:10.1016/j.cell.2005.08.040

Noyes, M. B., Christensen, R. G., Wakabayashi, A., Stormo, G. D., Brodsky, M. H. and Wolfe, S. A. (2008). Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. Cell 133, 1277-1289. doi:10.1016/j.cell.2008.05.023

Pan, Q., Cheng, G., Liu, Y., Xu, T., Zhang, H. and Li, B. (2020). TMSB10 acts as a biomarker and promotes progression of clear cell renal cell carcinoma. Int. J. Oncol. 56, 1101-1114. doi:10.3892/ijo.2020.4991

Park, S., Kim, D., Jung, Y.-G. and Roh, S. (2015). Thiazovivin, a Rho kinase inhibitor, improves stemness maintenance of embryo-derived stem-like cells under chemically defined culture conditions in cattle. Anim. Reprod. Sci 161, 47-57. doi:10.1016/j.anireprosci.2015.08.003

Rashbass, P., Cooke, L. A., Herrmann, B. G. and Beddington, R. S. P. (1991). A cell autonomous function of Brachyury in T/T embryonic stem cell chimaeras. Nature 353, 348-351. doi:10.1038/353348a0

Richard S. Mann, Katherine M. Lelli, Rohit Joshi (2009). Chapter 3 Hox Specificity: Unique Roles for Cofactors and Collaborators, Current Topics in Developmental Biology, Academic Press, Volume 88, 63-101, ISSN 0070-2153, ISBN 9780123745293. doi:10.1016/S0070-2153(09)88003-4

Rossant, J. and Cross, J. C. (2001). Placental development: lessons from mouse mutants. Nat. Rev. Genet. 2, 538-548. doi:10.1038/35080570

Rostovskaya, M., Andrews, S., Reik, W. and Rugg-Gunn, P. J. (2022). Amniogenesis occurs in two independent waves in primates. Cell Stem Cell 29, 744-759.e6. doi:10.1016/j.stem.2022.03.014

Sasaki, K., Nakamura, T., Okamoto, I., Yabuta, Y., Iwatani, C., Tsuchiya, H., Seita, Y., Nakamura, S., Shiraki, N., Takakuwa, T. et al. (2016). The germ cell fate of cynomolgus monkeys is specified in the nascent amnion. Dev. Cell 39, 169-185. doi:10.1016/j.devcel.2016.09.007

Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol. 33, 495-502. doi:10.1038/nbt.3192

Savory, J. G. A., Bouchard, N., Pierre, V., Rijli, F. M., De Repentigny, Y., Kothary, R. and Lohnes, D. (2009). Cdx2 regulation of posterior development through non-Hox targets. Development 136, 4099-4110. doi:10.1242/dev.041582

Saxena, M., Roman, A. K. S., O'Neill, N. K., Sulahian, R., Jadhav, U. and Shivdasani, R. A. (2017). Transcription factor-dependent 'anti-repressive' mammalian enhancers exclude H3K27me3 from extended genomic domains. Genes Dev. 31, 2391-2404. doi:10.1101/gad.308536.117

Scotti, M. and Kmita, M. (2012). Recruitment of 5′ Hoxa genes in the allantois is essential for proper extra-embryonic function in placental mammals. Development 139, 731-739. doi:10.1242/dev.075408

Senger, D. R., Galli, S. J., Dvorak, A. M., Perruzzi, C. A., Harvey, V. S. and Dvorak, H. F. (1983). Tumor cells secrete a vascular permeability factor that promotes accumulation of ascites fluid. Science 219, 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, 1097-1105. doi:10.1177/1947601911423031

Stefater, J. A., III, Lewkowich, I., Rao, S., Mariggi, G., Carpenter, A. C., Burr, A. R., Fan, J., Ajima, R., Molkentin, J. D., Williams, B. O. et al. (2011). Regulation of angiogenesis by a non-canonical Wnt-Flt1 pathway in myeloid cells. Nature 474, 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. Rev. Reprod. 1, 144-148. doi:10.1530/ror.0.0010144

Strumpf, D., Mao, C.-A., Yamanaka, Y., Ralston, A., Chawengsaksophak, K., Beck, F. and Rossant, J. (2005). Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. Development 132, 2093-2102. doi:10.1242/dev.01801

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., Hao, Y., Stoeckius, M., Smibert, P. and Satija, R. (2019). Comprehensive integration of single-cell data. Cell 177, 1888-1902.e21. doi:10.1016/j.cell.2019.05.031

Théry, M. and Piel, M. (2009). Adhesive micropatterns for cells: a microcontact printing protocol. Cold Spring Harb. Protoc. 2009, pdb.prot5255. doi:10.1101/pdb.prot5255

Thurston, G., Suri, C., Smith, K., McClain, J., Sato, T. N., Yancopoulos, G. D. and McDonald, D. M. (1999). Leakage-resistant blood vessels in mice transgenically overexpressing angiopoietin-1. Science 286, 2511-2514. doi:10.1126/science.286.5449.2511

van den Akker, E., Forlani, S., Chawengsaksophak, K., de Graaff, W., Beck, F., Meyer, B. I. and Deschamps, J. (2002). Cdx1 and Cdx2 have overlapping functions in anteroposterior patterning and posterior axis elongation. Development 129, 2181-2193. doi:10.1242/dev.129.9.2181

Verzi, M. P., Shin, H., He, H. H., Sulahian, R., Meyer, C. A., Montgomery, R. K., Fleet, J. C., Brown, M., Liu, X. S. and Shivdasani, R. A. (2010). Differentiation-specific histone modifications reveal dynamic chromatin interactions and partners for the intestinal transcription factor CDX2. Dev. Cell 19, 713-726. doi:10.1016/j.devcel.2010.10.006

Verzi, M. P., Shin, H., San Roman, A. K., Liu, X. S. and Shivdasani, R. A. (2013). Intestinal master transcription factor CDX2 controls chromatin access for partner transcription factor binding. Mol. Cell. Biol. 33, 281-292. doi:10.1128/MCB.01185-12

Warmflash, A., Sorre, B., Etoc, F., Siggia, E. D. and Brivanlou, A. H. (2014). A method to recapitulate early embryonic spatial patterning in human embryonic stem cells. Nat. Methods 11, 847-854. doi:10.1038/nmeth.3016

Wu, Y., Moser, M., Bautch, V. L. and Patterson, C. (2003). HoxB5 is an upstream transcriptional switch for differentiation of the vascular endothelium from precursor cells. Mol. Cell. Biol. 23, 5680-5691. doi:10.1128/MCB.23.16.5680-5691.2003

Yamamichi, N., Inada, K., Furukawa, C., Sakurai, K., Tando, T., Ishizaka, A., Haraguchi, T., Mizutani, T., Fujishiro, M., Shimomura, R. et al. (2009). Cdx2 and the Brm-type SWI/SNF complex cooperatively regulate villin expression in gastrointestinal cells. Exp. Cell Res. 315, 1779-1789. doi:10.1016/j.yexcr.2009.01.006

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, 789-796. doi:10.1038/nrg2400

Zheng, Y., Xue, X., Shao, Y., Wang, S., Esfahani, S. N., Li, Z., Muncie, J. M., Lakins, J. N., Weaver, V. M., Gumucio, D. L. et al. (2019). Controlled modelling of human epiblast and amnion development using stem cells. Nature 573, 421-425. doi:10.1038/s41586-019-1535-2