The transcripts of a gastrula-premarked enhancer prime posterior tissue development through cross-talk with morphogen effector

Authors: Yingying Chen[1]†, Fengxiang Tan[1]†, Xianfa Yang[1]†*, Qing Fang [2]†, Lin Zhang[3]†, Jiaoyang Liao[4], Penglei Shen[1], Yun Qian[3], Mingzhu Wen[1], Rui Song[1], Yonggao Fu[1], He Jax Xu[3], Ran Wang[5], Cheng Li[2,6]*, Zhen Shao[7]*, Jinsong Li[3]*,Naihe Jing[1]*.

Affiliations:

[1]Guangzhou National Laboratory, No. 9 XingDaoHuanBei Road, Guangzhou International Bio Island, Guangzhou 510005, Guangdong Province, China.

[2]School of Life Sciences, Peking University; Beijing, China.

[3]State Key Laboratory of Cell Biology, Center for Excellence in Molecular Cell Science, Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences, University of Chinese Academy of Sciences, Shanghai, China.

[4]Shanghai Frontiers Science Center of Genome Editing and Cell Therapy, Shanghai Key Laboratory of Regulatory Biology, Institute of Biomedical Sciences and School of Life Sciences, East China Normal University, Shanghai, China

[5]Key Laboratory of Biomedical Information Engineering of Ministry of Education, Biomedical Informatics & Genomics Center, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China

[6]Center for Statistical Science, Peking University; Beijing, China.

[7]CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, Chinese Academy of Sciences, Shanghai, 200031, China.

*Corresponding author. Email: yang_xianfa@gzlab.ac.cn (X.Y.), cheng_li@pku.edu.cn (C.L.), shaozhen@sinh.ac.cn (Z.S.), jsli@sibcb.ac.cn (J.L.), jing_naihe@gzlab.ac.cn (N.J.)

†These authors contributed equally to this work.

Abstract

The regulatory mechanisms governing cell fate determination, particularly lineage diversification during mammalian embryonic development, remain poorly understood with in-depth regulatory paradigms yet to be fully elucidated. Here, leveraging the epigenetic landscape of mouse gastrula, we identified p-Enh, a pre-marked enhancer in primitive streak region, as pivotal regulator for posterior tissue development in mouse embryos. Morphological and single-cell transcriptomic analyses confirmed embryonic lethality phenotype with disrupted posterior tissue development trajectories in p-Enh-KO embryos. Molecularly, apart from regulating the neighboring coding-gene *Cdx2 in cis*, we found that p-Enh can also modulate the global transcriptome and epigenomic landscape through the transient production of chromatin-binding eRNA *in trans*. Further investigation revealed p-Enh-eRNA participate in the regulatory cascades of TGF-β signaling by colocalizing with TFs such as SMAD4. Chemical modulation of TGF-β signaling or over-expression of nuclei-resident eRNAs can morphologically rescue the posterior development in *in vitro* gastruloids. Thus, we propose that the broadly distributed p-Enh transcripts within the nucleus serve as essential coordinators to prime the posterior development of mouse embryo.

Main Text:

Gastrulation stands as a pivotal developmental process to set up the embryonic body plan, characterized by the differentiation of the epiblast into three germ layers (*1*, *2*). Studies reported that the posteriorly positioned primitive streak and adjacent posterior epiblast (PS) region can integrate the intricate developmental signals, and then serve as a progenitor reservoir with the competence to give rise to a diverse range of tissues, spanning from mesodermal tissues such as the anlage of somitic, limb and cardiac mesoderm to the posterior neural lineage of the spinal cord (*3*–*5*). Nonetheless, how cells reside in the PS region coordinate the extrinsic signals and intrinsic molecular features and sustain the competence for prospective cell fate differentiation remains largely elusive.

Vital developmental morphogens lay the foundation for tissue patterning and homeostasis across various organs (*6*). The cross-talk among different pathways forms a "signal cocktail" with an optimal concentration niche, which facilitates diverse cell differentiation programs. For instance, during gastrulation, Wnt and TGF-β signaling pathways are activated in the PS and mesodermal cells, then diffuse from the organized sources and form morphogen gradients, thus instructing the cell fate of surrounding tissues (*7*, *8*). However, β-catenin and SMAD4, the downstream effectors of Wnt and TGF-β signaling, are broadly expressed throughout the entire embryo during gastrulation (*9*). Distinct sets of downstream developmental genes are then activated to promote cell fate diversification in different regions of the embryo (*1*). The specific underlying mechanisms for how these signals anchor precisely to a specific set of developmental genes remain as mysteries.

The selective chromatin responsiveness, manifested as various chromatin accessibilities, chromatin modifications, or chromatin topologies at regulatory loci, to signal effectors and co-factors, have been reported as a vital dimension in orchestrating the regulative forces of extrinsic signals (*10*). Notably, the regulatory elements are frequently docking platforms for transcription factors (TFs) and are usually annotated as enhancers or promoters (*11*). Dynamics of regulatory elements contribute to the regulation of lineage-specific transcriptional programs under distinct developmental conditions (*12*). Mechanistic studies indicate that regulatory elements can not only regulate the expression of neighboring genes but also hold the proper cell fate plasticity through direct enhancer-promoter chromatin interactions or self-transcribed enhancer RNAs (eRNAs) (*13*, *14*). In particular, the discovery of eRNAs has expanded the understanding of enhancer functions from modulating local gene targets *in cis* to even

genome-wide chromatin remodeling and cross-modular interactions with the transcriptional machinery *in trans* (*15–17*). Recently, we and others have reported that enhancers crucial for prospective tissues development are frequently pre-marked by epigenetic modifications or TFs' binding prior to the corresponding gene expression (*13*, *18–22*) , a phenomena termed "chromatin priming". However, limited knowledge has been acquired concerning the coordination between chromatin regulation and the extrinsic signals in the establishment of cellular developmental competence and the subsequent lineage diversification of the PS region.

In this study, through combined bioinformatic analysis and experimental screening of the pre-marked regulatory elements in the PS regions of the mouse gastrula, we identified a regulatory element, p-Enh, which is located in the first intron of *Cdx2.* Transgenic enhancer reporter assay revealed that p-Enh is pre-activated in the PS region at the late gastrulation stage and consistently activated in the tailbud region during early organogenesis. Phenotypic analyses of the p-Enh knockout embryo mutants revealed embryo lethality at around the E10.5 stage with severe disruptions of posterior tissue development. Single-cell RNA-seq confirmed that the cellular abundances of posterior tissues, such as neuromesodermal progenitors (NMP), limb mesenchyme, and somatic mesoderm, were heavily decreased. Mechanistically, besides regulating its neighboring coding-gene *Cdx2*, p-Enh could transiently express abundant eRNAs, in the nuclei of the PS cells at E7.5 but not in later-stage embryos. The genetic removal of p-Enh in *in vitro* differentiated NMP leads to dramatic remodeling of epigenomic landscape specifically at SMAD binding hotspots independent of CDX2 dependency. Furthermore, p-Enh-eRNA might participate in the regulatory cascades of TGF-β signaling by colocalizing with TF repertoires containing SMAD4. Functional assays demonstrate both the re-expression of p-Enh-eRNA and adjusting TGF-β signaling can rescue the p-Enh-KO defects. Thus, we propose that the transient expressed p-Enh-eRNA can coordinate with signal effectors, therefore contributing to the priming of posterior tissue development.

Screening and identification of pre-marked distal regulatory elements within PS cells in the mouse gastrula

During mouse gastrulation, the PS cells firstly emerge at the E6.5 stage, then gradually form a morphologically distinct structure at the posterior embryonic region, and finally establish a diverse cell fate-specified progenitor pool at the late gastrulation stage (E7.0 to E7.5) (*4*, *23*). PS cells possess a remarkable capacity to give rise to a diverse range of tissues, encompassing mesodermal cells such as somites (*24*), limb mesenchyme (*26*),

and cardiac mesoderm (*27*) (Fig. 1A). Moreover, it has been recognized that the PS region of the mouse gastrula also harbors a newly identified bi-potent NMP, which can contribute to both the posterior spinal cord and mesoderm (Fig. 1A) (*28–30*). Thus, the mechanisms governing PS cell development are fundamental for subsequent patterning of diverse tissues and organs.

Previously, we identified the epigenetic pre-mark at regulatory elements associated with developmental genes during gastrulation. This mechanism may serve as a widely available mechanism that primes prospective cell fate for following organogenesis (*18*). Here, to pinpoint the specific regulatory elements responsible for the developmental competence of the PS cells, we first employed MAnorm2 (*31*) to systematically analyze the dynamics of H3K27ac enrichment during the late gastrulation stage (from the E7.0 to the E7.5 stage) (Fig. 1B). In total, 1828 distal regulatory elements (DREs) were identified with significant upregulation (2-fold, p-value < 0.01) of H3K27ac enrichment within the posterior tissue (fig. S1A). Genomic Regions Enrichment of Annotations Tool (GREAT) analysis of the top-100 DREs, ranked by p-value, revealed that these DREs are related to genes involved in anterior-posterior (A-P) patterning, animal organ morphogenesis, neural tube development, as well as mesenchymal cell differentiation, which further indicate that the pre-mark of H3K27ac at DREs are functionally related to the posterior tissue development (Fig. S1B). To determine the enrichment of H3K27ac during the following organogenesis, we checked the H3K27ac distribution at the DREs in E9.5 tailbud tissues by incorporating the published dataset (*32*). As shown in Fig. 1B, among the top-30 ranked DREs, 10 DREs (marked with asterisk) maintain abundant H3K27ac enrichment in E9.5 tailbud.

Enhancer RNAs (eRNAs), which are transcribed from corresponding enhancers and often pre-loaded with RNA polymerase II and H3K27ac, could be used to represent the enhancer activity and also be involved in the regulation of gene expression and cell fate determination (*33–37*). To further refine the scope of potential functional DREs, we determined the abundance of eRNA transcripts for the top-30 DREs across various tissues of embryos at different stages (E7.0, E7.5, E9.5) using RT-qPCR (Fig. 1C). Based on the expression pattern of these eRNAs, we clustered the eRNAs into three groups: posterior early-activated group with relatively high and specific eRNA expression in the PS cells of the E7.5 stage; posterior late-activated group with obvious eRNA expression in the E9.5 trunk and tailbud; and posterior no-enriched group (Fig. 1C). Next, to determine the innate regulatory activity during mouse embryogenesis, we

used *in vivo* transgenic enhancer reporter system to validate the DREs from the above three groups (fig. S1C) (*18*, *38*). We performed transgenic reporter assays with the top-10 DREs, spanning the three eRNA groups (highlighted in bold in Fig. 1C) and found that 6 out of 10 DREs could drive LacZ expression in the time window tested (E7.0 to E9.5). Notably, the majority (5 out of 6) of DREs with LacZ signals belonged to the early-activated group (DRE-1, 2, 3, 5, and 8). Only one DRE (DRE-6) from the late-activated group showed LacZ signal (1 out of 3), and no positive results were observed in the posterior no-enriched group (Fig. 1D, fig. S1D). Within these six DREs, DRE-1 exhibits the most pronounced caudal tissue specificity of LacZ signals. DRE-1 is *de novo* activated in the E7.5 PS cells with extraembryonic LacZ expression in E7.0 embryos, and its posterior region-specific enhancer activity is sustained in the tailbud region till later organogenetic stages (E8.5, E9.5 and E11.5) (Fig. 1D; fig. S1E). Further exploration of the genomic features of DRE-1 revealed that the evolutionarily conserved DRE-1 element is located in the first intron of the coding-gene, *Cdx2*, and harbors pervasive binding sites for epigenetic activators (MED1 and p300) (*39*, *40*), signal effectors (β-catenin, TCF-1, and SMAD4) (*41–43*), and master regulators (T and SOX2) (*44*, *45*) (Fig. 1E; fig. S1F). Published gene run-on sequencing (GRO-seq) (*46*) using *in vitro* PS counterpart also demonstrated the existence of nascent transcripts from the DRE-1 locus (Fig. 1E). Given the highly tissue-specificity of DRE-1 element in the posterior embryonic tissue, we named the element as posterior enhancer (p-Enh) in the following studies. Consistent with the GRO-seq and eRNA quantification results (Fig. 1C, 1E), RNA *in situ* visualization using RNAscope revealed that only sense transcripts (oriented identically to the nearby gene) could be specifically detected in the PS region of the E7.5 embryos (Fig. 1F, fig. S1G). There was no eRNA signals could be detected in ExE cells, PS cells of earlier embryonic stages (E6.5 and E7.0), or in the tailbud of the organogenetic embryos (E9.5) (Fig. 1F; fig. S1H, S1I).

Taken together, we identified a DRE, p-Enh, which is pre-marked by H3K27ac specifically in PS cells at E7.5, possessing the posterior tissue-specific enhancer activity from late gastrulation stage to organogenesis stage. Notably, p-Enh shows transient sense eRNA transcription with strict spatial-temporal specificity, occurring only in E7.5 PS cells.

Genetic removal of p-Enh causes severe posterior tissue developmental defects and embryonic lethality at the organogenetic stage

To investigate the potential functional role of p-Enh during mouse embryo development, we genetically knockout the p-Enh (~1.3 kb) in the mouse embryo using the CRISPR/Cas9 system (fig. S2A). No live mouse individual with homozygous p-Enh removal (p-Enh-KO) could be acquired (fig. S2B). Further analyses identified that most p-Enh-KO embryos are deceased around the E10.5-E11.5 stage with noticeable posterior tailbud developmental abnormalities and hindlimb absence (Fig. 2A; fig. S2C, S2D). TUNEL assay further confirmed that the apoptotic signals become prominent in the E10.5 p-Enh-KO embryos, predominantly enriched at the tailbud and the brain regions (fig. S2E).

To systematically characterize the molecular phenotypes of p-Enh-KO embryos, we performed single-cell transcriptomic analyses in p-Enh-KO and wildtype counterparts from the same litter covering successive early organogenetic stages from E8.5 to E10.5 (Fig. 2B). In total, 180,930 cells across three sequential stages were sequenced with a median coverage of 23,783 unique molecular identifiers (UMIs) per cell. 127,069 cells were retained for further analysis after rigorous quality filters (fig. S2F, Method). By referencing the published embryo atlas dataset (*47*, *48*), 21 cell clusters of the WT embryos with precise annotation were identified (Fig. 2B; fig. S2G). Generally, our annotation unveiled diverse clusters with unique set of markers spanning all three germ layer derivatives. For example, the neural tube cells were defined by the expression of *Sox2, Pax6, and Fabp7*, while NMPs were distinguished by *Epha5*, *T, and Fgf8* (fig. S2G). Additionally, the dynamics of cell type proportions across developmental stages also reflect the sequential cell fate differentiation during organogenesis (Fig. 2B). Thus, the single-cell transcriptome profile of early organogenesis ranging from E8.5 to E10.5 faithfully captures the *in vivo* developmental events.

Next, we projected the single-cell transcriptome of p-Enh-KO embryos onto the WT reference, and found that the major cell type compositions between the WT and p-Enh-KO embryos were comparable (Fig. 2B). The prominent apoptotic signals identified in E10.5 p-Enh-KO embryos were further confirmed by the assessment of apoptosis and hypoxia activity hallmarks expression across all samples (fig. S2H). To mitigate the potential bias caused by apoptotic cells, we focused our analyses on the E8.5 and E9.5 embryos in the following sections. Differential cellular abundance analysis highlighted a substantial decrease in posterior derived tissues of p-Enh-KO embryos, such as NMP,

somatic mesoderm, and limb mesenchyme (Fig. 2C; fig. S2I). Following this, we calculated the enrichment scores of marker gene sets within each cluster. Remarkably, while the most significant cell density changes were observed at E9.5 (Fig. 2C), the marker gene expression patterns, especially for NMP and somatic cells, showed severe abnormalities as early as E8.5 in p-Enh-KO embryos (Fig. 2D, fig. S3A, S3B). The alterations of gene expression prior to those of cellular phenotypes indicates that p-Enh may function at an earlier stage when the posterior tissues were differentiated from the PS region in the gastrula.

Given the phenotypes of p-Enh-KO were mainly associated with the posterior tissue development, we focused on cell types comprising the embryonic A-P axis from the single-cell reference atlas, encompassing forebrain/midbrain/hindbrain, neural tube, NMP, somitic mesoderm, and paraxial mesoderm. Sub-clustering of these single cells by UMAP further divided the cells into 7 clusters (Fig. 2E). RNA velocity and pseudo-time transcriptional ordering analyses showed a clear bi-directional trajectory originated from NMP towards the neural and mesodermal lineages, respectively (Fig. 2E). This result aligns with the model that two distinct developmental trajectories exist in neural tube cells: one originating from the anterior brain cells and the other from the bipotential NMP in the posterior region (*28*, *29*). The developmental signal gradients (RA, Wnt, and FGF signaling) and A-P axis-related marker genes alignment (*49*, *50*) signify the robust correlation between the plotted single-cell transcriptome and the A-P axis patterning process *in vivo* (fig. S3C-S3F), suggesting that the reconstructed UMAP plot could effectively represent cellular distribution along the *in vivo* A-P axis, and highlight NMP as the origin for posterior neural tube cells (Fig. 2E).

Subsequently, we compared the cellular distribution between wildtype and p-Enh-KO embryos along with the reconstructed digital A-P map. By linearly plotting the single cells along with the reconstructed A-P axis, the p-Enh-KO embryo exhibited a noticeable reduction in cell density in NMP and their immediate developmental progenies (Fig. 2F). Trajectory-based gene expression analyses revealed significant disruptions in the developmental programs during both posterior neural tube and mesodermal lineage development in p-Enh-KO embryos (Fig. 2G). Experimental validation showed that the expression territory and abundance of *Hoxb9,* a marker for posterior neural tube and mesodermal cells with the M-shaped distribution, was restricted in the tailbud tip of the p-Enh-KO embryos (Fig. 2H). Additionally, *Tcf15*, marker for anterior presomitic mesoderm (PSM) and somite (*51*), exhibited a

substantial decrease and shrinkage in the undifferentiated progenitor pool for PSM (*Tcf15*-negative region) in E9.5 p-Enh-KO embryos (Fig. 2I). Besides, the highly organized stripe-like pattern of *Uncx4.1* in WT embryos (*52*) was severely disrupted in p-Enh-KO embryos, especially for embryos at E9.5 (Fig. 2J). Interestingly, the anterior somite structure appears to be normal in the E8.5 p-Enh-KO embryos (Fig. 2J). Given the different trajectory routes for the early-anterior (early PS-derived) and late-posterior somitogenesis (NMP-derived) during mouse embryonic development (*24*), the disrupted *Uncx4.1* distribution indicated that p-Enh may specifically regulate the NMP-derived late-posterior somitogenesis process in the mouse embryo.

Additionally, we also observed a notable decrease in the posterior-derived limb mesenchyme cell population in E9.5 p-Enh-KO embryos (Fig. 2C). To delve into these abnormalities, we specifically analyzed cells associated with limb mesenchyme and further categorized them into three sub-clusters representing forelimb progenitors, hindlimb progenitors, and other mesenchyme based on marker gene expression. RNA velocity analysis confirmed that both forelimb and hindlimb progenitors seem to share one common developmental origin (fig. S3G). Cell density analyses reveal comparable patterns between WT and p-Enh-KO embryos at E8.5, whereas there is a pronounced reduction in hindlimb progenitors in E9.5 p-Enh-KO embryos (fig. S3H). Molecular analyses identified that genes related to hindlimb development, such as *Hoxc10*, *Hoxd11*, *Pitx1*, and *Tmem119*, were significantly down-regulated in p-Enh-KO embryos, while the markers of forelimb showed limited expression alterations (fig. S3I). Further experimental validation of the typical hindlimb marker *Pitx1* confirmed the absence of hindlimb structure in the p-Enh-KO embryos with the normal distribution of forelimb cells as marked by *Tbx5* (fig. S3J).

Together, these results indicate that the deletion of p-Enh had a profound impact on the developmental programs of multiple organs, including retardation in posterior tissue growth and eventual embryonic lethality. Single-cell transcriptomics analysis reveals that p-Enh can regulate the gene expression of NMP-derived clusters as early as E8.5, p-Enh-KO results in a pronounced decrease in cell density in posterior cell clusters at E9.5, and ultimately contributing to the lethality phenotype at E10.5.

Two distinct mechanisms with differential CDX2 dependencies underlying p-Enh function

A substantial body of work on various enhancers have proposed mechanisms of enhancer function by regulating adjacent genes (*53–57*). Considering that p-Enh is located in the first intron of *Cdx2* and crucial role of *Cdx2* in body axis development (*32*, *58–60*), we examined whether the expression of *Cdx2* was affected in p-Enh-KO embryos. There was no change in *Cdx2* expression in the extraembryonic cells, with partial down-regulation in the posterior regions at E7.5 (Fig. 3A). Consistently, single-cell transcriptome data also showed reduced *Cdx2* expression in posterior tissues such as NMP and somitic mesoderm (fig. S4A). To probe the downstream molecular mechanisms for p-Enh function and dissect the potential relevance between p-Enh and its neighboring coding gene-*Cdx2*, we prepared the p-Enh-KO and Cdx2-KO mouse embryonic stem cell lines with CRISPR/Cas9 system, respectively (fig. S4B) (*61*). It is noteworthy that the Cdx2-KO cell line features a specific 35 bp deletion in the first exon, causing a frameshift mutation and eliminating functional CDX2 protein without interrupting the p-Enh element genetically (fig. S4C, S4D). Generally, the morphologies of stem cell clones and distributions of pluripotent markers NANOG and OCT4 were comparable among WT, Cdx2-KO, and p-Enh-KO ESCs (fig. S4E), indicating that the removal of CDX2 protein or p-Enh element did not affect stem cell pluripotency. Next, we applied the WT, Cdx2-KO and p-Enh-KO cell lines to the newly developed *in vitro* gastruloids (*62*) and Trunk-like structures (TLSs) (*63*) differentiation system (Fig. 3B), which can be used to model the anterior-posterior axis formation and later somitogenesis in the embryo. We found that the WT embryoids could generate tailbud-like structures that resemble the caudal region of embryos in both gastruloids and TLSs system (Fig. 3C, fig. S4F). However, the elongation of tailbud regions in both the gastruloids and TLSs for Cdx2-KO and the p-Enh-KO cells was largely compromised in comparison to the WT group. Moreover, in contrast to the WT and Cdx2-KO gastruloids and TLSs, the embryoids derived from p-Enh-KO ESCs remain as large spherical aggregates with no morphological evidence of A-P axis elongation (Fig. 3C, fig. S4F). Geometric analysis of the acquired gastruloids and TLSs further confirmed the consistent and significant defects ($p<0.0001$) in A-P axis organization and elongation in both Cdx2-KO and p-Enh-KO embryoids. Moreover, the phenotypic abnormalities were more pronounced in p-Enh-KO embryoids compared to Cdx2-KO counterparts (Fig. 3D, fig. S4G).

To comprehensively determine the molecular architectures of gastruloids derived from Cdx2-KO and p-Enh-KO cells, we performed transcriptomic profiling of the

gastruloids with three biological replicates for each group (fig. S5A). Our analysis revealed a significant decrease in *Cdx2* expression in both Cdx2-KO and p-Enh-KO samples, which was further experimentally confirmed (Fig. S5B, S5C). Principal component analyses (PCA) revealed distinct transcriptomic features for Cdx2-KO and p-Enh-KO samples (Fig. 3E). Additionally, p-Enh-KO exhibited a greater number of differentially expressed genes (DEGs) compared to *Cdx2*-KO (p-Enh-KO: 608 up and 1451 down, vs. Cdx2-KO: 819 up and 455 down) (fig. S5D). Remarkably, only a small portion of DEGs were overlapped between Cdx2-KO and p-Enh-KO group (330/1274, 25.9% of Cdx2-KO DEGs; 330/2059, 16% of p-Enh-KO DEGs) (Fig. 3F), suggesting both shared and distinct regulatory roles of *Cdx2* and p-Enh during A-P axis formation and elongation in embryoids.

Next, we categorized these DEGs (Fig. 3F) into 5 clusters based on gene expression patterns, including the shared DEGs (Group 1), Cdx2-KO specific up (Group 2), Cdx2-KO specific down (Group 3), p-Enh-KO specific up (Group 4), and p-Enh-KO specific down (Group 5) (Fig. 3G). Gene ontology (GO) analysis revealed that genes in Group 3 were enriched in the biological process related to anterior/posterior patterning and somitogenesis process, consistent with the well-known functions of CDX2 protein (*58*). As for p-Enh-KO specific down (Group 5) genes, the pathways related to a broad spectrum of mesodermal development, such as gland development, mesenchyme development, and heart morphogenesis, were enriched (Fig. 3G, fig. S5E). Following this, we conducted an enrichment analysis of the DEGs of Group 1, 3, 5 with the constructed *in vivo* single-cell A-P atlas (Fig. 2E), and found that Group 1 and Group 3 exhibited a clear posteriorized enrichment, displaying high enrichment scores with NMP, somitic mesoderm, and posterior neural tube clusters. However, genes in the p-Enh-KO specific down group (Group 5) showed a strong correlation with NMP and the mesodermal clusters (Fig. 3H), indicating p-Enh may participate into posterior mesoderm development through a CDX2-independent mechanism.

To further validate this notion, we established inducible *Cdx2* overexpression cell line on the background of p-Enh-KO (p-Enh-KO/iCdx2) and subject this cell line into gastruloid differentiation system (Fig. 3I, fig. S5F). We observed that the expression of the somitic mesoderm marker *Tbx6*, a downstream target of *Cdx2*, was reactivated in the p-Enh-KO/iCdx2 gastruloids. However, the A-P elongated morphology as well as the stripe-like structure marked by *Uncx4.1* transcripts remained compromised in p-Enh-KO/iCdx2 (Fig. 3J). Transcriptome analyses showed that more than 50% of genes

(651/1224, 53%) failed to be fully rescued by the inducible *Cdx2* overexpression in comparison with the WT counterpart (Fig. 3K, fig. S5G), which enriched organ formation-related genes. In particular, genes involved in the response to the transforming growth factor beta (TGF-β) pathway remains abnormal even with the restoration of CDX2 (Fig. 3K). Regulon analyses summarizing the upstream regulators for the un-rescuable genes revealed that major effectors involved in TGF-β signaling, such as Smad1 and Smad9, were specifically enriched (Fig. S5H).

In summary, *in vitro* embryoid data showed that p-Enh can regulate posterior tissue development partially through the neighboring gene *Cdx2*. Moreover, these results also suggest that a previously unrecognized CDX2-independent mechanism may contribute to p-Enh function, possibly involving cellular responsiveness to TGF-β signaling pathway.

Genome-wide epigenomic remodeling suggests a potential functional relevance between p-Enh and TGF-β signaling effectors

To systematically determine the downstream molecular cascades of p-Enh, we profiled the transcriptome, H3K27ac modification enrichment from both WT and p-Enh-KO NMP cells acquired from directed NMP differentiation system (Fig. 4A; fig. S6A-S6C) (*64*). Besides, we also performed proximity ligation-assisted ChIP-seq (PLAC-seq) (*65*) as well as circular chromatin conformation capture (4C-seq) (*66*) to specifically capture the genome-wide chromatin interactome of p-Enh in WT NMP (Fig. 4A; fig. S6D-S6H). The comparative analysis of chromatin activities marked by H3K27ac revealed the occurrence of pervasive epigenomic remodeling in p-Enh-KO NMP cells, especially at distal intergenic chromatin regions (Fig. 4B; fig. S6C). GREAT analyses identified that peaks with weakened enrichment of H3K27ac in p-Enh-KO cells are strongly enriched with SMAD binding capabilities (Fig. 4B), which are crucial to mediate the responsiveness to TGF-β signaling (*8*). Following this, we also incorporated published dataset of SMADs ChIP-seq data (*67*, *68*), as well as CDX2 ChIP-seq (*32*) data from *in vitro* sample counterpart, and checked the distribution of these signaling effectors around peak sets with changed H3K27ac level in p-Enh-KO cells (Fig. 4C). We observed enrichment of TGF-β signal effectors (SMADs) binding with these peaks, but with minimal CDX2 binding (Fig. 4C) This result was further supported by the examination of developmental gene loci related to A-P axis patterning such as *Tbx3* and *Tcf15*, as well as genes involved in developmental signaling pathway, like *Tgfbr2* (Fig. 4D). These results, together with previous transcriptome data (Fig. 3F, 3K), indicate

that the CDX2-independent mechanism underlying p-Enh function may participate in coordinating with TGF-β signaling.

To explore the potential direct interacting regions of p-Enh, H3K27ac PLAC-seq revealed 38 interacting fragments around the p-Enh locus on chromosome 5 (within 3 megabases), exhibiting frequent chromatin interactions with the p-Enh locus (Fig. 4E). The 4C-seq analyses revealed the consistent frequency of chromatin interactions between p-Enh and neighboring genomic bins (fig. S6H). However, when integrated with transcriptome data from both WT and p-Enh-KO NMP cells, we found that only three genes (*Rasl11a*, *Cdx2*, and *Slc46a3*) exhibited significant transcriptomic changes between WT and p-Enh-KO cells (Fig. 4E). After referencing the published studies for both *Rasl11a* and *Slc46a3* (*69–71*), we chose to determine whether the aberrant up-regulation of *Slc46a3* in p-Enh-KO cells is responsible for the posterior tissue defects. However, the inhibition of *Slc46a3* expression level in p-Enh-KO cells through shRNA-mediated knocking-down (fig. S6I) failed to rescue the developmental abnormalities (Fig. 4F). These results suggest that the *Cdx2*-independent functions of p-Enh are not likely due to its *cis* chromatin interactions with distal genes, such as *Slc46a3*.

The nuclear co-localization of p-Enh transcripts and SMAD4 protein at the Hox cluster regulates the posterior tissue development

Recent studies have found that a broad spectrum of TFs bind with RNA molecules, highlighting the conserved and essential nature of interactions between TFs and RNA in vertebrate development (*72*). We found that p-Enh can transcribe into RNAs, which are largely distributed in the posterior regions at the late-gastrulation stage (E7.5) (Fig. 1F, 5A). Co-detection of p-Enh transcripts and *Cdx2* mRNA in the E7.5 mouse gastrula revealed that both transcripts were specifically enriched in the posterior region (Fig. 5A). Moreover, sub-cellular distribution analyses showed that the p-Enh-eRNAs were prominently enriched in the nuclei, forming puncta-like structure at multiple loci within a single nucleus. In contrast, the transcripts of the neighboring coding-gene, *Cdx2,* were universally located in the cytoplasm (Fig. 5A). Meanwhile, the signaling activity of TGF-β was also enriched in the posterior embryonic region at E7.5, represented by the spatial gene expression enrichment of typical TGF-β signaling target genes (Fig. 5B). The spatial co-occurrence of p-Enh transcripts and TGF-β signaling activity suggests an intriguing involvement of p-Enh-eRNA in TGF-β signaling transduction, possibly through direct interaction with nuclear SMAD4 protein. Following this hypothesis, we

performed co-staining of p-Enh-eRNA and SMAD4 protein in the E7.5 mouse embryo. As shown in Fig. 5C, more than 40% nuclei-enriched p-Enh-eRNA puncta were specifically co-localized with SMAD4 protein. *In vitro* assay further confirmed that synthesized p-Enh-eRNAs were prone to enriched at the central region of SMAD4 droplets (fig. S7A, S7B). Since nuclear SMAD4 can directly bind to the chromatin regions of TGF-β target genes (*8*), we further asked whether the SMAD4 co-localized p-Enh transcripts could bind to the SMAD4-bound chromatin regions. ChIRP-seq analysis confirmed that p-Enh-eRNA can specifically bind to the flank region of SMAD4 targeted chromatin, such as the *Hox* cluster and developmental gene loci (Fig. 5D, fig. S7C). These results suggest that p-Enh may function through direct interaction with nuclear SMAD4 in the form of RNA puncta.

To validate the notion that p-Enh transcripts and its molecular partner SMAD4 could coordinately function in the A-P patterning, we knocked down p-Enh-eRNA in WT ESCs and overexpressed p-Enh-eRNA in p-Enh-KO cells and subjected these cells into gastruloid differentiation (Fig. 5E; fig. S7D, S7E). We found that knocking down p-Enh transcripts could largely mimic the developmental defects in p-Enh-KO cells (Fig. 5F, 5G), and the nuclei-specific overexpression of p-Enh transcripts (*73*) or specific activation of TGF-β signaling using Activin A can obviously alleviate the differentiation defects caused by p-Enh knockout. However, inhibiting TGF-β signaling by SB431542 induced extensive cell death, indicating the essential roles of TGF-β signaling in A-P axis patterning (Fig. 5F, 5G). Together, these results demonstrate that p-Enh transcripts and the molecular interactions between p-Enh transcripts and SMAD4 protein are of great biological significances underlying posterior tissue development.

Discussion

In this study, we have identified a pre-marked regulatory element, p-Enh, exhibiting highly restricted spatial-temporal activity and significant biological importance during mouse early embryonic development. The pivotal role of p-Enh in posterior tissue development is underscored by the embryonic lethality observed in p-Enh-KO embryos, accompanied by severe posterior tissue growth failure, disruption of A-P axis gene expression pattern, and dysfunctions across multiple cell types (Fig. 2). Mechanism analyses through transcriptomic dissection pinpointed the co-occurrence of two distinct mechanisms underlying the p-Enh function: (i) regulating the expression level of its neighboring gene *Cdx2 in cis*, thereby affecting the downstream cascades of CDX2 (Fig. 3); (ii) transient production of chromatin-binding eRNAs, potentially involved in coordinating the nuclear responsiveness to the extrinsic developmental signals, such as TGF-β signaling pathway (Fig. 4-5). Thus, we propose a molecular model for pre-marked developmental enhancers, wherein genetic sequence-based *in cis* local regulation and RNA transcripts-based *in trans* genome-wide regulation co-exists to facilitate tight spatiotemporal control of global molecular architecture and thus set up the molecular blueprint for future tissue development (Fig. 5H).

Recently, there has been a growing interest in dissecting the molecular hierarchy of gene regulatory networks (GRNs) that govern rapid cell fate commitment (*9*, *18*, *74*, *75*). Previous studies have primarily focused on identifying specific pioneer transcription factors capable of steering the developmental competence towards specific developmental lineages (*76–78*). In line with the pioneering factors, emerging evidences also indicated that the epigenetic priming or pre-mark at regulatory elements could offer a notable feature prior to cell fate decision (*13*, *79*). However, it remains unclear whether the "chromatin priming" of regulatory elements holds potential functional role in biological process beyond merely serving as a marker. Here, we determined one pre-marked regulatory element, p-Enh, which is specifically activated in the PS cells at late gastrulation stage, and sustains the epigenetic activity till late organogenesis. Phenotypic and molecular analysis demonstrated that genetic loss of p-Enh leads to dramatic developmental defects when the posterior lineage diversifies. Moreover, the precedence of transcriptomic disruptions over phenotypic anomalies strengthens the evidence for the importance of the priming effects. Thus, future evaluation of pre-marked regulatory elements and further phenotyping through incorporating multi-dimensional data may better elucidate the function of novel crucial regulatory elements.

Generally, chromatin loops formed between distal regulatory elements and target genes play crucial roles in transducing the regulatory force from the upstream TFs to the downstream local target genes (*57*, *80–82*). Here, through in-depth transcriptomic dissection, genetic analyses, and intricate chromatin conformation capture techniques, we discovered that only a small subset of p-Enh physically linked genes (*Cdx2*, *Rasl11a*, *Slc46a3*) exhibit significant gene expression change in p-Enh-KO cells (Fig. 4E). For the local gene-*Cdx2*, as reflected by the transcriptomic comparison among multiple genetic edited cell lines (WT, p-Enh-KO, Cdx2-KO, p-Enh-KO/iCdx2), we found that only 16% genes were regulated through the p-Enh-*Cdx2* axis (Fig. 3F). Meanwhile, two intriguing observations suggest potential roles for p-Enh beyond direct physical interactions. First, TGF-β signaling responsive genes were highly enriched in the cluster of p-Enh specifically regulated genes (Fig. 3K, fig. S5H). Second, through profiling H3K27ac pattern in both WT and p-Enh-KO cells, we found that the global distribution of H3K27ac was pervasively remodeled in p-Enh-KO cells (Fig. 4B). Interestingly, the genomic regions with changed H3K27ac level are enriched with TGF-β signaling effectors but not CDX2 binding (Fig. 4C). Thus, the regulatory relay underlying p-Enh function should be involved with TGF-β signaling but not simply viewed as the linear molecular cascades from the upstream TF to p-Enh and its subsequent locally linked targets.

The discovery of enhancer-derived transcripts largely broaden the current scope of the molecular mechanisms governing chromatin structure organization and gene expression regulation (*83*, *84*). The abundant expression and nuclear localization of p-Enh-eRNAs at multiple loci within individual nuclei in the PS region suggest their potential roles in global GRNs instead of localized functions. Their broad distribution within nuclei may contribute to the observed pervasive chromatin activity changes in p-Enh-KO cells (Fig. 5A, 4B). Furthermore, recent study has underscored the extensive interaction between TFs and RNA molecules (*72*). In our study, we observed the cross-modular cellular location of eRNA transcripts and SMAD4 protein. Chromatin distribution analysis further supported this observation, demonstrating that p-Enh-eRNA can interact with SMAD4 protein at the *Hox* gene loci (Fig. 5C, 5D). Importantly, functional analyses confirmed the importance of both eRNA transcripts and TGF-β signaling pathway in A-P axis patterning (Fig. 5E-5G). This finding prompts us to hypothesize that, the widespread eRNA-TFs condensate complexes within individual nuclei may act as “regulatory machinery” to modulate pivotal biological events, such as the long-range regulatory functions of p-Enh (Fig. 5H). This regulatory machinery

could potentially facilitate the efficient guidance of interacting TF repertoires to their target chromatin regions, bridging the enhancer element and signaling pathway targets through eRNA-TFs molecules, thereby influencing the long-range modulation of global GRNs for proper developmental programming. Future in-depth characterization of p-Enh-eRNA structure as well as identification of the molecular partners of this transcript will clarify the complex regulatory network and promote a comprehensive understanding of mouse embryogenesis.

In summary, to delve into the molecular mechanisms underlying epigenetic pre-marked enhancers, we systematically dissect both the *in cis* and *in trans* paths for p-Enh, proposing this as a novel strategy of developmental enhancer to coordinate multiple downstream modules genome-widely, and then establish a responsive cellular state for forthcoming lineage diversification.

References and Notes

1. P. P. L. Tam, D. A. F. Loebel, Gene function in mouse embryogenesis: get set for gastrulation. *Nat. Rev. Genet.* **8**, 368–381 (2007).
2. N. Fossat, V. Jones, M. J. Garcia-Garcia, P. P. L. Tam, Modulation of WNT signaling activity is key to the formation of the embryonic head. *Cell Cycle* **11**, 26–32 (2012).
3. C. Qiu, J. Cao, B. K. Martin, T. Li, I. C. Welsh, S. Srivatsan, X. Huang, D. Calderon, W. S. Noble, C. M. Disteche, S. A. Murray, M. Spielmann, C. B. Moens, C. Trapnell, J. Shendure, Systematic reconstruction of cellular trajectories across mouse embryogenesis. *Nat. Genet.* **54**, 328–341 (2022).
4. R. Wang, X. Yang, J. Chen, L. Zhang, J. A. Griffiths, G. Cui, Y. Chen, Y. Qian, G. Peng, J. Li, L. Wang, J. C. Marioni, P. P. L. Tam, N. Jing, Time space and single-cell resolved tissue lineage trajectories and laterality of body plan at gastrulation. *Nat. Commun.* **14**, 5675 (2023).
5. G. Sheng, A. Martinez Arias, A. Sutherland, The primitive streak and cellular principles of building an amniote body through gastrulation. *Science* **374**, abg1727 (2021).
6. A. Kicheva, J. Briscoe, Control of Tissue Development by Morphogens. *Annu. Rev. Cell Dev. Biol.* **39**, annurev-cellbio-020823-011522 (2023).
7. P. P. Tam, D. A. Loebel, S. S. Tanaka, Building the mouse gastrula: signals, asymmetry and lineages. *Curr. Opin. Genet. Amp Dev.* **16**, 419–425 (2006).
8. C. Liu, G. Peng, N. Jing, TGF-β signaling pathway in early mouse development and embryonic stem cells. *Acta Biochim. Biophys. Sin.* **50**, 68–73 (2018).
9. G. Peng, S. Suo, G. Cui, F. Yu, R. Wang, J. Chen, S. Chen, Z. Liu, G. Chen, Y. Qian, P. P. L. Tam, J.-D. J. Han, N. Jing, Molecular architecture of lineage allocation and tissue organization in early mouse embryo. *Nature* **572**, 528–532 (2019).
10. A. I. Badeaux, Y. Shi, Emerging roles for chromatin as a signal integration and storage platform. *Nat. Rev. Mol. Cell Biol.* **14**, 211–224 (2013).
11. G. A. Maston, S. K. Evans, M. R. Green, Transcriptional regulatory elements in the human genome. *Annu. Rev. Genomics Hum. Genet.* **7**, 29–59 (2006).
12. J. Huang, X. Liu, D. Li, Z. Shao, H. Cao, Y. Zhang, E. Trompouki, T. V. Bowman, L. I. Zon, G.-C. Yuan, S. H. Orkin, J. Xu, Dynamic Control of Enhancer Repertoires Drives Lineage and Stage-Specific Transcription during Hematopoiesis. *Dev. Cell* **36**, 9–23 (2016).

13. A. Wang, F. Yue, Y. Li, R. Xie, T. Harper, N. A. Patel, K. Muth, J. Palmer, Y. Qiu, J. Wang, D. K. Lam, J. C. Raum, D. A. Stoffers, B. Ren, M. Sander, Epigenetic Priming of Enhancers Predicts Developmental Competence of hESC-Derived Endodermal Lineage Intermediates. *Cell Stem Cell* **16**, 386–399 (2015).

14. P. R. Arnold, A. D. Wells, X. C. Li, Diversity and Emerging Roles of Enhancer RNA in Regulation of Gene Expression and Cell Fate. *Front. Cell Dev. Biol.* **7** (2020).

15. H. K. Long, S. L. Prescott, J. Wysocka, Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell* **167**, 1170–1187 (2016).

16. M. T. Y. Lam, W. Li, M. G. Rosenfeld, C. K. Glass, Enhancer RNAs and regulated transcriptional programs. *Trends Biochem. Sci.* **39**, 170–182 (2014).

17. Z. Han, W. Li, Enhancer RNA: What we know and what we can achieve. *Cell Prolif.* **55**, e13202 (2022).

18. X. Yang, B. Hu, J. Liao, Y. Qiao, Y. Chen, Y. Qian, S. Feng, F. Yu, J. Dong, Y. Hou, H. Xu, R. Wang, G. Peng, J. Li, F. Tang, N. Jing, Distinct enhancer signatures in the mouse gastrula delineate progressive cell fate continuum during embryo development. *Cell Res.* **29**, 911–926 (2019).

19. A. J. Rubin, B. C. Barajas, M. Furlan-Magaril, V. Lopez-Pajares, M. R. Mumbach, I. Howard, D. S. Kim, L. D. Boxer, J. Cairns, M. Spivakov, S. W. Wingett, M. Shi, Z. Zhao, W. J. Greenleaf, A. Kundaje, M. Snyder, H. Y. Chang, P. Fraser, P. A. Khavari, Lineage-specific dynamic and pre-established enhancer-promoter contacts cooperate in terminal differentiation. *Nat. Genet.* **49**, 1522–1528 (2017).

20. H. S. Kim, Y. Tan, W. Ma, D. Merkurjev, E. Destici, Q. Ma, T. Suter, K. Ohgi, M. Friedman, D. Skowronska-Krawczyk, M. G. Rosenfeld, Pluripotency factors functionally premark cell-type-restricted enhancers in ES cells. *Nature* **556**, 510–514 (2018).

21. S. Ai, H. Xiong, C. C. Li, Y. Luo, Q. Shi, Y. Liu, X. Yu, C. Li, A. He, Profiling chromatin states using single-cell itChIP-seq. *Nat. Cell Biol.* **21**, 1164–1172 (2019).

22. M. Sendra, K. McDole, D. Jimenez-Carretero, J. de D. Hourcade, S. Temiño, M. Raiola, L. Guignard, P. J. Keller, F. Sánchez-Cabo, J. N. Domínguez, M. Torres, Epigenetic priming of embryonic lineages in the mammalian epiblast. bioRxiv [Preprint] (2024). https://doi.org/10.1101/2024.01.11.575188.

23. E. S. Bardot, A.-K. Hadjantonakis, Mouse gastrulation: Coordination of tissue patterning, specification and diversification of cell fate. *Mech. Dev.* **163**, 103617 (2020).

24. C. Guibentif, J. A. Griffiths, I. Imaz-Rosshandler, S. Ghazanfar, J. Nichols, V. Wilson, B. Göttgens, J. C. Marioni, Diverse Routes toward Early Somites in the Mouse Embryo. *Dev. Cell* **56**, 141-153.e6 (2020).
25. S. Markman, M. Zada, E. David, A. Giladi, I. Amit, E. Zelzer, A single-cell census of mouse limb development identifies complex spatiotemporal dynamics of skeleton formation. *Dev. Cell* **0** (2023).
26. C. McQueen, M. Towers, Establishing the pattern of the vertebrate limb. *Development* **147**, dev177956 (2020).
27. R. C. V. Tyser, X. Ibarra-Soria, K. McDole, S. A. Jayaram, J. Godwin, T. A. H. van den Brand, A. M. A. Miranda, A. Scialdone, P. J. Keller, J. C. Marioni, S. Srinivas, Characterization of a common progenitor pool of the epicardium and myocardium. *Science*, doi: 10.1126/science.abb2986 (2021).
28. D. Henrique, E. Abranches, L. Verrier, K. G. Storey, Neuromesodermal progenitors and the making of the spinal cord. *Development* **142**, 2864–2875 (2015).
29. M. Gouti, J. Delile, D. Stamataki, F. J. Wymeersch, Y. Huang, J. Kleinjung, V. Wilson, J. Briscoe, A Gene Regulatory Network Balances Neural and Mesoderm Specification during Vertebrate Trunk Development. *Dev. Cell* **41**, 243-261.e7 (2017).
30. V. Metzis, S. Steinhauser, E. Pakanavicius, M. Gouti, D. Stamataki, K. Ivanovitch, T. Watson, T. Rayon, S. N. Mousavy Gharavy, R. Lovell-Badge, N. M. Luscombe, J. Briscoe, Nervous System Regionalization Entails Axial Allocation before Neural Differentiation. *Cell* **1751**, 1105-1118.e17 (2018).
31. S. Tu, M. Li, H. Chen, F. Tan, J. Xu, D. J. Waxman, Y. Zhang, Z. Shao, MAnorm2 for quantitatively comparing groups of ChIP-seq samples. *Genome Res.* **31**, 131–145 (2021).
32. S. Amin, R. Neijts, S. Simmini, C. van Rooijen, S. C. Tan, L. Kester, A. van Oudenaarden, M. P. Creyghton, J. Deschamps, Cdx and T Brachyury Co-activate Growth Signaling in the Embryonic Axial Progenitor Niche. *Cell Rep.* **17**, 3165–3177 (2016).
33. R. Ye, C. Cao, Y. Xue, Enhancer RNA: biogenesis, function, and regulation. *Essays Biochem.* **64**, 883–894 (2020).
34. V. Sartorelli, S. M. Lauberth, Enhancer RNAs are an important regulatory layer of the epigenome. *Nat. Struct. Mol. Biol.* **27**, 521–528 (2020).
35. D. A. Bose, G. Donahue, D. Reinberg, R. Shiekhattar, R. Bonasio, S. L. Berger, RNA Binding to CBP Stimulates Histone Acetylation and Transcription. *Cell* **168**, 135-149.e22 (2017).
36. L. J. Harrison, D. Bose, Enhancer RNAs step forward: new insights into enhancer function. *Development* **149**, dev200398 (2022).

37. A. M. Costa, D. M. Bear, D. A. Harmin, D. Kuhl, E. Markenscoff-Papadimitriou, G. Kreiman, H. Bito, J. M. Gray, J. Wu, K. Barbara-Haley, M. Hemberg, M. E. Greenberg, M. Laptewicz, P. F. Worley, S. Kuersten, T.-K. Kim, Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182 (2010).
38. L. A. Pennacchio, N. Ahituv, A. M. Moses, S. Prabhakar, M. A. Nobrega, M. Shoukry, S. Minovitsky, I. Dubchak, A. Holt, K. D. Lewis, I. Plajzer-Frick, J. Akiyama, S. De Val, V. Afzal, B. L. Black, O. Couronne, M. B. Eisen, A. Visel, E. M. Rubin, In vivo enhancer analysis of human conserved non-coding sequences. *Nature* **444**, 499–502 (2006).
39. M. H. Kagey, J. J. Newman, S. Bilodeau, Y. Zhan, D. A. Orlando, N. L. van Berkum, C. C. Ebmeier, J. Goossens, P. B. Rahl, S. S. Levine, D. J. Taatjes, J. Dekker, R. A. Young, Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430–435 (2010).
40. C.-X. Song, K. E. Szulwach, Q. Dai, Y. Fu, S.-Q. Mao, L. Lin, C. Street, Y. Li, M. Poidevin, H. Wu, J. Gao, P. Liu, L. Li, G.-L. Xu, P. Jin, C. He, Genome-wide Profiling of 5-Formylcytosine Reveals Its Roles in Epigenetic Priming. *Cell* **153**, 678–691 (2013).
41. X. Zhang, K. Peterson, X. Liu, M. Andrew, S. Ohba, Gene Regulatory Networks Mediating Canonical Wnt Signal-Directed Control of Pluripotency and Differentiation in Embryo Stem Cells. *Stem Cells* **31**, 2667–2679 (2013).
42. J. L. Johnson, G. Georgakilas, J. Petrovic, M. Kurachi, S. Cai, C. Harly, W. S. Pear, A. Bhandoola, E. J. Wherry, G. Vahedi, Lineage-Determining Transcription Factor TCF-1 Initiates the Epigenetic Identity of T Cells. *Immunity* **48**, 243-257.e10 (2018).
43. P. Martin-Malpartida, M. Batet, Z. Kaczmarska, R. Freier, T. Gomes, E. Aragón, Y. Zou, Q. Wang, Q. Xi, L. Ruiz, A. Vea, J. A. Márquez, J. Massagué, M. J. Macias, Structural basis for genome wide recognition of 5-bp GC motifs by SMAD transcription factors. *Nat. Commun.* **8**, 2070 (2017).
44. A. Beisaw, P. Tsaytler, F. Koch, S. U. Schmitz, M. Melissari, A. D. Senft, L. Wittler, T. Pennimpede, K. Macura, B. G. Herrmann, P. Grote, BRACHYURY directs histone acetylation to target loci during mesoderm development. *EMBO Rep.* **19**, 118–134 (2018).
45. Z. Liu, W. L. Kraus, Catalytic-Independent Functions of PARP-1 Determine Sox2 Pioneer Activity at Intractable Genomic Loci. *Mol. Cell* **65**, 589-603.e9 (2017).
46. A. Strikoudis, C. Lazaris, T. Trimarchi, A. L. Galvao Neto, Y. Yang, P. Ntziachristos, S. Rothbart, S. Buckley, I. Dolgalev, M. Stadtfeld, B. D. Strahl, B. D. Dynlacht, A. Tsirigos, I. Aifantis, Regulation of transcriptional elongation in

pluripotency and cell differentiation by the PHD-finger protein Phf5a. *Nat. Cell Biol.* **18**, 1127–1138 (2016).

47. J. Cao, M. Spielmann, X. Qiu, X. Huang, D. M. Ibrahim, A. J. Hill, F. Zhang, S. Mundlos, L. Christiansen, F. J. Steemers, C. Trapnell, J. Shendure, The single-cell transcriptional landscape of mammalian organogenesis. *Nature*, doi: 10.1038/s41586-019-0969-x (2019).

48. B. Pijuan-Sala, J. A. Griffiths, C. Guibentif, T. W. Hiscock, W. Jawaid, F. J. Calero-Nieto, C. Mulas, X. Ibarra-Soria, R. C. V. Tyser, D. L. L. Ho, W. Reik, S. Srinivas, B. D. Simons, J. Nichols, J. C. Marioni, B. Göttgens, A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature*, doi: 10.1038/s41586-019-0933-9 (2019).

49. K. Dorey, E. Amaya, FGF signalling: diverse roles during early vertebrate embryogenesis. *Development* **137**, 3731–3742 (2010).

50. S. Bernheim, S. M. Meilhac, Mesoderm patterning by a dynamic gradient of retinoic acid signalling. *Philos. Trans. R. Soc. B Biol. Sci.* **375**, 20190556 (2020).

51. R. Burgess, A. Rawls, D. Brown, A. Bradley, E. N. Olson, Requirement of the paraxis gene for somite formation and musculoskeletal patterning. *Nature* **384**, 570–573 (1996).

52. A. Mansouri, A. K. Voss, T. Thomas, Y. Yokota, P. Gruss, Uncx4.1 is required for the formation of the pedicles and proximal ribs and acts upstream of Pax9. *Dev. Camb. Engl.* **127**, 2251–2258 (2000).

53. H. Shiratori, R. Sakuma, M. Watanabe, H. Hashiguchi, K. Mochida, Y. Sakai, J. Nishino, Y. Saijoh, M. Whitman, H. Hamada, Two-Step Regulation of Left–Right Asymmetric Expression of Pitx2: Initiation by Nodal Signaling and Maintenance by Nkx2. *Mol. Cell* **7**, 137–149 (2001).

54. H. Yamagishi, J. Maeda, T. Hu, J. McAnally, S. J. Conway, T. Kume, E. N. Meyers, C. Yamagishi, D. Srivastava, Tbx1 is regulated by tissue-specific forkhead proteins through a common Sonic hedgehog-responsive enhancer. *Genes Dev.* **17**, 269–281 (2003).

55. G. F. Mok, L. Folkes, S. A. Weldon, E. Maniou, V. Martinez-Heredia, A. M. Godden, R. M. Williams, T. Sauka-Spengler, G. N. Wheeler, S. Moxon, A. E. Münsterberg, Characterising open chromatin in chick embryos identifies cis-regulatory elements important for paraxial mesoderm formation and axis extension. *Nat. Commun.* **12**, 1157 (2021).

56. T. Takemoto, M. Uchikawa, M. Yoshida, D. M. Bell, R. Lovell-Badge, V. E. Papaioannou, H. Kondoh, Tbx6-dependent Sox2 regulation determines neural or mesodermal fate in axial stem cells. *Nature* **470**, 394–398 (2011).

57. T.-T. Liu, S. Kim, P. Desai, D.-H. Kim, X. Huang, S. T. Ferris, R. Wu, F. Ou, T. Egawa, S. J. Van Dyken, M. S. Diamond, P. F. Johnson, M. Kubo, T. L. Murphy, K. M. Murphy, Ablation of cDC2 development by triple mutations within the Zeb2 enhancer. *Nature* **607**, 142–148 (2022).
58. K. Chawengsaksophak, W. de Graaff, J. Rossant, J. Deschamps, F. Beck, Cdx2 is essential for axial elongation in mouse development. *Proc. Natl. Acad. Sci.* **101**, 7641–7645 (2004).
59. E. van den Akker, S. Forlani, K. Chawengsaksophak, W. de Graaff, F. Beck, B. I. Meyer, J. Deschamps, Cdx1 and Cdx2 have overlapping functions in anteroposterior patterning and posterior axis elongation. *Development* **129**, 2181–2193 (2002).
60. J. G. A. Savory, N. Bouchard, V. Pierre, F. M. Rijli, Y. D. Repentigny, R. Kothary, D. Lohnes, Cdx2 regulation of posterior development through non-Hox targets. *Development* **136**, 4099–4110 (2009).
61. L. Cong, F. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. Hsu, X. Wu, W. Jiang, L. Marraffini, F. Zhang, Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339**, 819–823 (2013).
62. S. C. van den Brink, A. Alemany, V. van Batenburg, N. Moris, M. Blotenburg, J. Vivié, P. Baillie-Johnson, J. Nichols, K. F. Sonnen, A. Martinez Arias, A. van Oudenaarden, Single-cell and spatial transcriptomics reveal somitogenesis in gastruloids. *Nature*, 1–5 (2020).
63. J. V. Veenvliet, A. Bolondi, H. Kretzmer, L. Haut, M. Scholze-Wittler, D. Schifferl, F. Koch, L. Guignard, A. S. Kumar, M. Pustet, S. Heimann, R. Buschow, L. Wittler, B. Timmermann, A. Meissner, B. G. Herrmann, Mouse embryonic stem cells self-organize into trunk-like structures with neural tube and somites. *Science* **370** (2020).
64. M. Gouti, A. Tsakiridis, F. J. Wymeersch, Y. Huang, J. Kleinjung, V. Wilson, J. Briscoe, In Vitro Generation of Neuromesodermal Progenitors Reveals Distinct Roles for Wnt Signalling in the Specification of Spinal Cord and Paraxial Mesoderm Identity. *PLOS Biol.* **12**, e1001937 (2014).
65. R. Fang, M. Yu, G. Li, S. Chee, T. Liu, A. D. Schmitt, B. Ren, Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq. *Cell Res.* **26**, 1345–1348 (2016).
66. P. H. L. Krijger, G. Geeven, V. Bianchi, C. R. E. Hilvering, W. de Laat, 4C-seq from beginning to end: A detailed protocol for sample preparation and data analysis. *Methods* **170**, 17–32 (2020).
67. E. Aragón, Q. Wang, Y. Zou, S. M. Morgani, L. Ruiz, Z. Kaczmarska, J. Su, C. Torner, L. Tian, J. Hu, W. Shu, S. Agrawal, T. Gomes, J. A. Márquez, A.-K.

Hadjantonakis, M. J. Macias, J. Massagué, Structural basis for distinct roles of SMAD2 and SMAD3 in FOXH1 pioneer-directed TGF-β signaling. *Genes Dev.*, genesdev;gad.330837.119v1 (2019).

68. Q. Wang, Y. Zou, S. Nowotschin, S. Y. Kim, Q. V. Li, C.-L. Soh, J. Su, C. Zhang, W. Shu, Q. Xi, D. Huangfu, A.-K. Hadjantonakis, J. Massagué, The p53 Family Coordinates Wnt and Nodal Inputs in Mesendodermal Differentiation of Embryonic Stem Cells. *Cell Stem Cell* **20**, 70–86 (2017).

69. M. Pistoni, A. Verrecchia, M. Doni, E. Guccione, B. Amati, Chromatin association and regulation of rDNA transcription by the Ras-family protein RasL11a. *EMBO J.* **29**, 1215–1224 (2010).

70. J.-H. Kim, T. Matsubara, J. Lee, C. Fenollar-Ferrer, K. Han, D. Kim, S. Jia, C. J. Chang, H. Yang, T. Nagano, K. W. Krausz, S.-H. Yim, F. J. Gonzalez, Lysosomal SLC46A3 modulates hepatic cytosolic copper homeostasis. *Nat. Commun.* **12**, 290 (2021).

71. K. J. Hamblett, A. P. Jacob, J. L. Gurgel, M. E. Tometsko, B. M. Rock, S. K. Patel, R. R. Milburn, S. Siu, S. P. Ragan, D. A. Rock, C. J. Borths, J. W. O'Neill, W. S. Chang, M. F. Weidner, M. M. Bio, K. C. Quon, W. C. Fanslow, SLC46A3 Is Required to Transport Catabolites of Noncleavable Antibody Maytansine Conjugates from the Lysosome to the Cytoplasm. *Cancer Res.* **75**, 5329–5340 (2015).

72. O. Oksuz, J. E. Henninger, R. Warneford-Thomson, M. M. Zheng, H. Erb, A. Vancura, K. J. Overholt, S. W. Hawken, S. F. Banani, R. Lauman, L. N. Reich, A. L. Robertson, N. M. Hannett, T. I. Lee, L. I. Zon, R. Bonasio, R. A. Young, Transcription factors interact with RNA to regulate genes. *Mol. Cell* **83**, P2449-2463 (2023).

73. Q.-F. Yin, S.-B. Hu, Y.-F. Xu, L. Yang, G. G. Carmichael, L.-L. Chen, SnoVectors for nuclear expression of RNA. *Nucleic Acids Res.* **43**, e5–e5 (2015).

74. H. Yang, Y. Luan, T. Liu, H. J. Lee, L. Fang, Y. Wang, X. Wang, B. Zhang, Q. Jin, K. C. Ang, X. Xing, J. Wang, J. Xu, F. Song, I. Sriranga, C. Khunsriraksakul, T. Salameh, D. Li, M. N. K. Choudhary, J. Topczewski, K. Wang, G. S. Gerhard, R. C. Hardison, T. Wang, K. C. Cheng, F. Yue, A map of cis -regulatory elements and 3D genome structures in zebrafish. *Nature*, 1–7 (2020).

75. M. Shu, D. Hong, H. Lin, J. Zhang, Z. Luo, Y. Du, Z. Sun, M. Yin, Y. Yin, L. Liu, S. Bao, Z. Liu, F. Lu, J. Huang, J. Dai, Single-cell chromatin accessibility identifies enhancer networks driving gene expression during spinal cord development in mouse. *Dev. Cell*, doi: 10.1016/j.devcel.2022.11.011 (2022).

76. K. Lee, H. Cho, R. W. Rickert, Q. V. Li, J. Pulecio, C. S. Leslie, D. Huangfu, FOXA2 Is Required for Enhancer Priming during Pancreatic Differentiation. *Cell Rep.* **28**, 382-393.e7 (2019).

77. R. C. Adam, H. Yang, S. Rockowitz, S. B. Larsen, M. Nikolova, D. S. Oristian, L. Polak, M. Kadaja, A. Asare, D. Zheng, E. Fuchs, Pioneer factors govern super-enhancer dynamics in stem cell plasticity and lineage choice. *Nature* **521**, 366–370 (2015).
78. S. Bessonnard, L. D. Mot, D. Gonze, M. Barriol, C. Dennis, A. Goldbeter, G. Dupont, C. Chazaud, Gata6, Nanog and Erk signaling control cell fate in the inner cell mass through a tristable regulatory network. *Development* **141**, 3637–3648 (2014).
79. R. Xu, S. Li, Q. Wu, C. Li, M. Jiang, L. Guo, M. Chen, L. Yang, X. Dong, H. Wang, C. Wang, X. Liu, X. Ou, S. Gao, Stage-specific H3K9me3 occupancy ensures retrotransposon silencing in human pre-implantation embryos. *Cell Stem Cell* **29**, 1051-1066.e8 (2022).
80. N. Koyano-Nakagawa, W. Gong, S. Das, J. W. M. Theisen, T. B. Swanholm, D. Van Ly, N. Dsouza, B. N. Singh, H. Kawakami, S. Young, K. Q. Chen, Y. Kawakami, D. J. Garry, Etv2 regulates enhancer chromatin status to initiate Shh expression in the limb bud. *Nat. Commun.* **13**, 4221 (2022).
81. A. Balsalobre, J. Drouin, Pioneer factors as master regulators of the epigenome and cell fate. *Nat. Rev. Mol. Cell Biol.*, doi: 10.1038/s41580-022-00464-z (2022).
82. W. Wang, A. Chandra, N. Goldman, S. Yoon, E. K. Ferrari, S. C. Nguyen, E. F. Joyce, G. Vahedi, TCF-1 promotes chromatin interactions across topologically associating domains in T cell progenitors. *Nat. Immunol.* **23**, 1052–1062 (2022).
83. C. Anania, Alu elements confer enhancer–promoter specificity. *Nat. Genet.* **55**, 1421–1421 (2023).
84. L. Liang, C. Cao, L. Ji, Z. Cai, D. Wang, R. Ye, J. Chen, X. Yu, J. Zhou, Z. Bai, R. Wang, X. Yang, P. Zhu, Y. Xue, Complementary Alu sequences mediate enhancer–promoter selectivity. *Nature* **619**, 868–875 (2023).
85. T. Chen, X. Chen, S. Zhang, J. Zhu, B. Tang, A. Wang, L. Dong, Z. Zhang, C. Yu, Y. Sun, L. Chi, H. Chen, S. Zhai, Y. Sun, L. Lan, X. Zhang, J. Xiao, Y. Bao, Y. Wang, Z. Zhang, W. Zhao, The Genome Sequence Archive Family: Toward Explosive Data Growth and Diverse Data Types. *Genomics Proteomics Bioinformatics* **19**, 578–583 (2021).
86. CNCB-NGDC Members and Partners, Database Resources of the National Genomics Data Center, China National Center for Bioinformation in 2022. *Nucleic Acids Res.* **50**, D27–D38 (2022).
87. X. Yang, Y. Chen, L. Song, T. Zhang, N. Jing, Wholemount in situ Hybridization for Spatial-temporal Visualization of Gene Expression in Early Post-implantation Mouse Embryos. *Bio-Protoc.* **11**, e4229–e4229 (2021).

88. Q.-L. Ying, J. Wray, J. Nichols, B.-M. Laura, B. Doble, J. Woodgett, P. Cohen, A. Smith, The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519–523 (2008).
89. L. Beccari, N. Moris, M. Girgin, D. A. Turner, P. Baillie-Johnson, A.-C. Cossy, M. P. Lutolf, D. Duboule, A. M. Arias, Multi-axial self-organization properties of mouse embryonic stem cells into gastruloids. *Nature*, 1 (2018).
90. Q.-F. Yin, L. Yang, Y. Zhang, J.-F. Xiang, Y.-W. Wu, G. G. Carmichael, L.-L. Chen, Long Noncoding RNAs with snoRNA Ends. *Mol. Cell* **48**, 219–230 (2012).
91. K. Zhang, F. Yu, J. Zhu, S. Han, J. Chen, X. Wu, Y. Chen, T. Shen, J. Liao, W. Guo, X. Yang, R. Wang, Y. Qian, J. Yang, L. Cheng, Y. Zhao, C.-C. Hui, J. Li, G. Peng, S. He, N. Jing, K. Tang, Imbalance of Excitatory/Inhibitory Neuron Differentiation in Neurodevelopmental Disorders with an NR2F1 Point Mutation. *Cell Rep.* **31**, 107521 (2020).
92. N. Renier, Z. Wu, D. J. Simon, J. Yang, P. Ariel, M. Tessier-Lavigne, iDISCO: A Simple, Rapid Method to Immunolabel Large Tissue Samples for Volume Imaging. *Cell* **159**, 896–910 (2014).
93. C. Chu, J. Quinn, H. Y. Chang, Chromatin Isolation by RNA Purification (ChIRP). *JoVE J. Vis. Exp.*, e3912 (2012).
94. C. Chu, K. Qu, F. L. Zhong, S. E. Artandi, H. Y. Chang, Genomic Maps of Long Noncoding RNA Occupancy Reveal Principles of RNA-Chromatin Interactions. *Mol. Cell* **44**, 667–678 (2011).
95. S. Alberti, S. Saha, J. B. Woodruff, T. M. Franzmann, J. Wang, A. A. Hyman, A User's Guide for Phase Separation Assays with Purified Proteins. *J. Mol. Biol.* **430**, 4806–4820 (2018).
96. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10 (2011).
97. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
98. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9** (2008).
99. H. M. Amemiya, A. Kundaje, A. P. Boyle, The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci. Rep.* **9**, 9354 (2019).
100. A. Frankish, S. Carbonell-Sala, M. Diekhans, I. Jungreis, J. E. Loveland, J. M. Mudge, C. Sisu, J. C. Wright, C. Arnan, I. Barnes, A. Banerjee, R. Bennett, A. Berry, A. Bignell, C. Boix, F. Calvet, D. Cerdán-Vélez, F. Cunningham, C. Davidson, S. Donaldson, C. Dursun, R. Fatima, S. Giorgetti, C. G. Giron, J. M. Gonzalez, M. Hardy, P. W. Harrison, T. Hourlier, Z. Hollis, T. Hunt, B. James, Y. Jiang, R.

Johnson, M. Kay, J. Lagarde, F. J. Martin, L. M. Gómez, S. Nair, P. Ni, F. Pozo, V. Ramalingam, M. Ruffier, B. M. Schmitt, J. M. Schreiber, E. Steed, M.-M. Suner, D. Sumathipala, I. Sycheva, B. Uszczynska-Ratajczak, E. Wass, Y. T. Yang, A. Yates, Z. Zafrulla, J. S. Choudhary, M. Gerstein, R. Guigo, T. J. P. Hubbard, M. Kellis, A. Kundaje, B. Paten, M. L. Tress, P. Flicek, GENCODE: reference annotation for the human and mouse genomes in 2023. *Nucleic Acids Res.* **51**, D942–D949 (2023).

101. G. X. Y. Zheng, J. M. Terry, P. Belgrader, P. Ryvkin, Z. W. Bent, R. Wilson, S. B. Ziraldo, T. D. Wheeler, G. P. McDermott, J. Zhu, M. T. Gregory, J. Shuga, L. Montesclaros, J. G. Underwood, D. A. Masquelier, S. Y. Nishimura, M. Schnall-Levin, P. W. Wyatt, C. M. Hindson, R. Bharadwaj, A. Wong, K. D. Ness, L. W. Beppu, H. J. Deeg, C. McFarland, K. R. Loeb, W. J. Valente, N. G. Ericson, E. A. Stevens, J. P. Radich, T. S. Mikkelsen, B. J. Hindson, J. H. Bielas, Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).

102. S. L. Wolock, R. Lopez, A. M. Klein, Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst.* **8**, 281-291.e9 (2019).

103. Y. Hao, S. Hao, E. Andersen-Nissen, W. M. Mauck, S. Zheng, A. Butler, M. J. Lee, A. J. Wilk, C. Darby, M. Zager, P. Hoffman, M. Stoeckius, E. Papalexi, E. P. Mimitou, J. Jain, A. Srivastava, T. Stuart, L. M. Fleming, B. Yeung, A. J. Rogers, J. M. McElrath, C. A. Blish, R. Gottardo, P. Smibert, R. Satija, Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).

104. T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck, Y. Hao, M. Stoeckius, P. Smibert, R. Satija, Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).

105. L. Waltman, N. J. van Eck, A smart local moving algorithm for large-scale modularity-based community detection. *Eur. Phys. J. B* **86**, 471 (2013).

106. L. McInnes, J. Healy, N. Saul, L. Großberger, UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* **3**, 861 (2018).

107. I. Tirosh, B. Izar, S. M. Prakadan, M. H. Wadsworth, D. Treacy, J. J. Trombetta, A. Rotem, C. Rodman, C. Lian, G. Murphy, M. Fallahi-Sichani, K. Dutton-Regester, J.-R. Lin, O. Cohen, P. Shah, D. Lu, A. S. Genshaft, T. K. Hughes, C. G. K. Ziegler, S. W. Kazer, A. Gaillard, K. E. Kolb, A.-C. Villani, C. M. Johannessen, A. Y. Andreev, E. M. Van Allen, M. Bertagnolli, P. K. Sorger, R. J. Sullivan, K. T. Flaherty, D. T. Frederick, J. Jané-Valbuena, C. H. Yoon, O. Rozenblatt-Rosen, A. K. Shalek, A. Regev, L. A. Garraway, Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).

108. G. La Manno, R. Soldatov, A. Zeisel, E. Braun, H. Hochgerner, V. Petukhov, K. Lidschreiber, M. E. Kastriti, P. Lönnerberg, A. Furlan, J. Fan, L. E. Borm, Z. Liu,

D. van Bruggen, J. Guo, X. He, R. Barker, E. Sundström, G. Castelo-Branco, P. Cramer, I. Adameyko, S. Linnarsson, P. V. Kharchenko, RNA velocity of single cells. *Nature* **560**, 494–498 (2018).

109. V. Bergen, M. Lange, S. Peidli, F. A. Wolf, F. J. Theis, Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* **38**, 1408–1414 (2020).

110. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

111. Y. Liao, G. K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

112. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

113. T. Wu, E. Hu, S. Xu, M. Chen, P. Guo, Z. Dai, T. Feng, L. Zhou, W. Tang, L. Zhan, X. Fu, S. Liu, X. Bo, G. Yu, clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* **2**, 100141 (2021).

114. G. Geeven, H. Teunissen, W. de Laat, E. de Wit, peakC: a flexible, non-parametric peak calling package for 4C and Capture-C data. *Nucleic Acids Res.* **46**, e91 (2018).

115. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

116. P. Danecek, J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane, S. A. McCarthy, R. M. Davies, H. Li, Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008 (2021).

117. R. Stark, G. Brown, DiffBind: Differential binding analysis of ChIP-Seq peak data. doi: 10.18129/B9.bioc.DiffBind.

118. S. Heinz, C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh, C. K. Glass, Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol. Cell* **38**, 576–589 (2010).

119. F. Ramírez, D. P. Ryan, B. Grüning, V. Bhardwaj, F. Kilpert, A. S. Richter, S. Heyne, F. Dündar, T. Manke, deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).

120. G. Yu, L.-G. Wang, Q.-Y. He, ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383 (2015).

121. N. Servant, N. Varoquaux, B. R. Lajoie, E. Viara, C.-J. Chen, J.-P. Vert, E. Heard, J. Dekker, E. Barillot, HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015).
122. C. A. Lareau, M. J. Aryee, hichipper: a preprocessing pipeline for calling DNA loops from HiChIP data. *Nat. Methods* **15**, 155–156 (2018).
123. D. Kim, J. M. Paggi, C. Park, C. Bennett, S. L. Salzberg, Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
124. S. Aibar, C. B. González-Blas, T. Moerman, V. A. Huynh-Thu, H. Imrichova, G. Hulselmans, F. Rambow, J.-C. Marine, P. Geurts, J. Aerts, J. van den Oord, Z. K. Atak, J. Wouters, S. Aerts, SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
125. R. Neijts, S. Amin, C. van Rooijen, S. Tan, M. P. Creyghton, W. de Laat, J. Deschamps, Polarized regulatory landscape and Wnt responsiveness underlie Hox activation in embryos. *Genes Dev.* **30**, 1937–1942 (2016).
126. A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, J. P. Mesirov, Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* **102**, 15545–15550 (2005).

Acknowledgments:

We are grateful to Prof. Chi-Chung Hui for his constructive suggestions. We thank the core facilities at Guangzhou National Laboratory and Center for Excellence in Molecular Cell Science, CAS for their excellent support, especially in animal care and imaging support.

Funding: This work was supported in part by

the National Key Basic Research and Development Program of China (2019YFA0801402, 2018YFA0800100, 2018YFA0107200, 2018YFA0801402),

the Major Project of Guangzhou National Laboratory (GZNL2023A02005),

the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA16020501, XDA16020404),

the National Natural Science Foundation of China (32130030, 31900454).

Author contributions:

Conceptualization: N.J., Y.C., X.Y.

Investigation: Y.C., F.T., X.Y., L.Z., Q.F., J.L., P.S., M.W., Y.Q., R.S., Y.F., H.J.X., R.W.

Visualization: Y.C., X.Y., F.T., Q.F., R.S., Y.F.

Funding acquisition: N.J., X.Y.

Supervision: N.J., J.L., Z.S., C.L.

Writing – original draft: Y.C., X.Y., F.T., Q.F., N.J.

Writing – review & editing: N.J., Y.C., X.Y., F.T., Q.F., C.L.

Competing interests: The authors declare that they have no competing interests.

Data and materials availability: All raw and processed sequence data reported in this paper have been deposited in the Genome Sequence Archive (*85*) in National Genomics Data Center (*86*), China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences (GSA: CRA014616) that are publicly accessible at https://ngdc.cncb.ac.cn/gsa.

Fig. 1. Screening and identification of pre-marked distal regulatory elements within PS cells in the mouse gastrula.

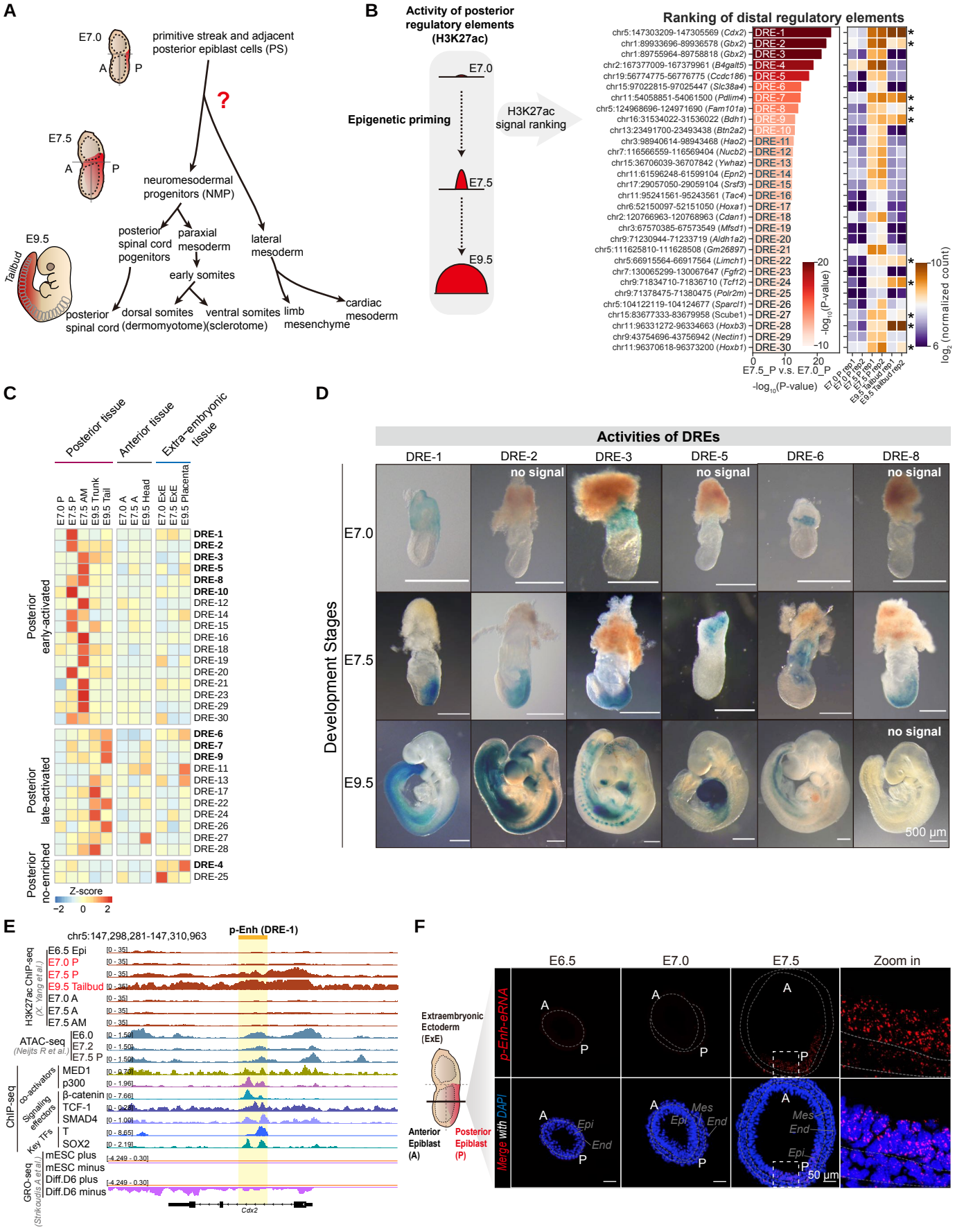

Fig. 2. Genetic removal of p-Enh causes severe posterior tissue developmental defects and embryonic lethality at the organogenetic stage.

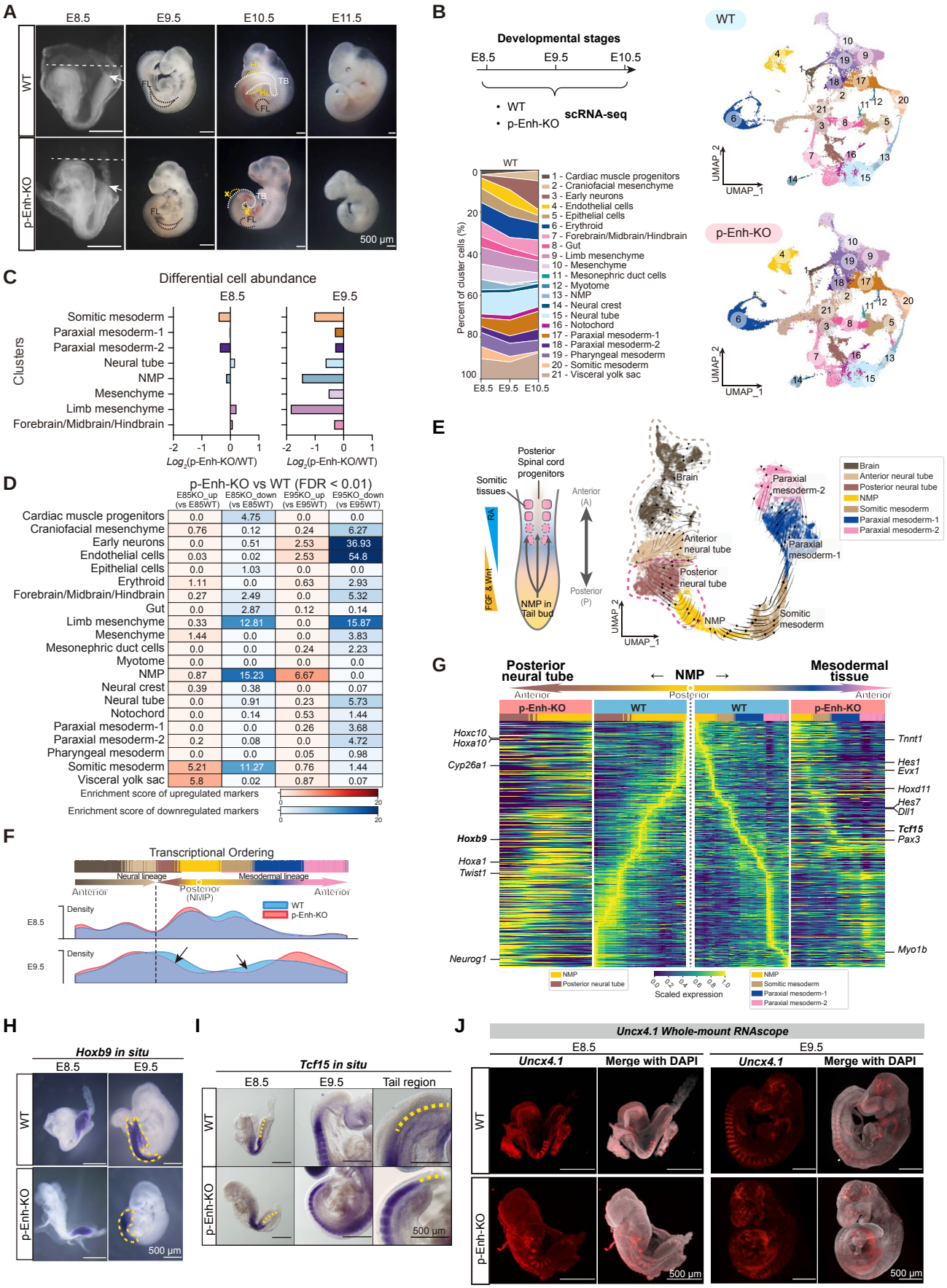

Fig. 3. Two distinct mechanisms with differential CDX2 dependencies underlying p-Enh function.

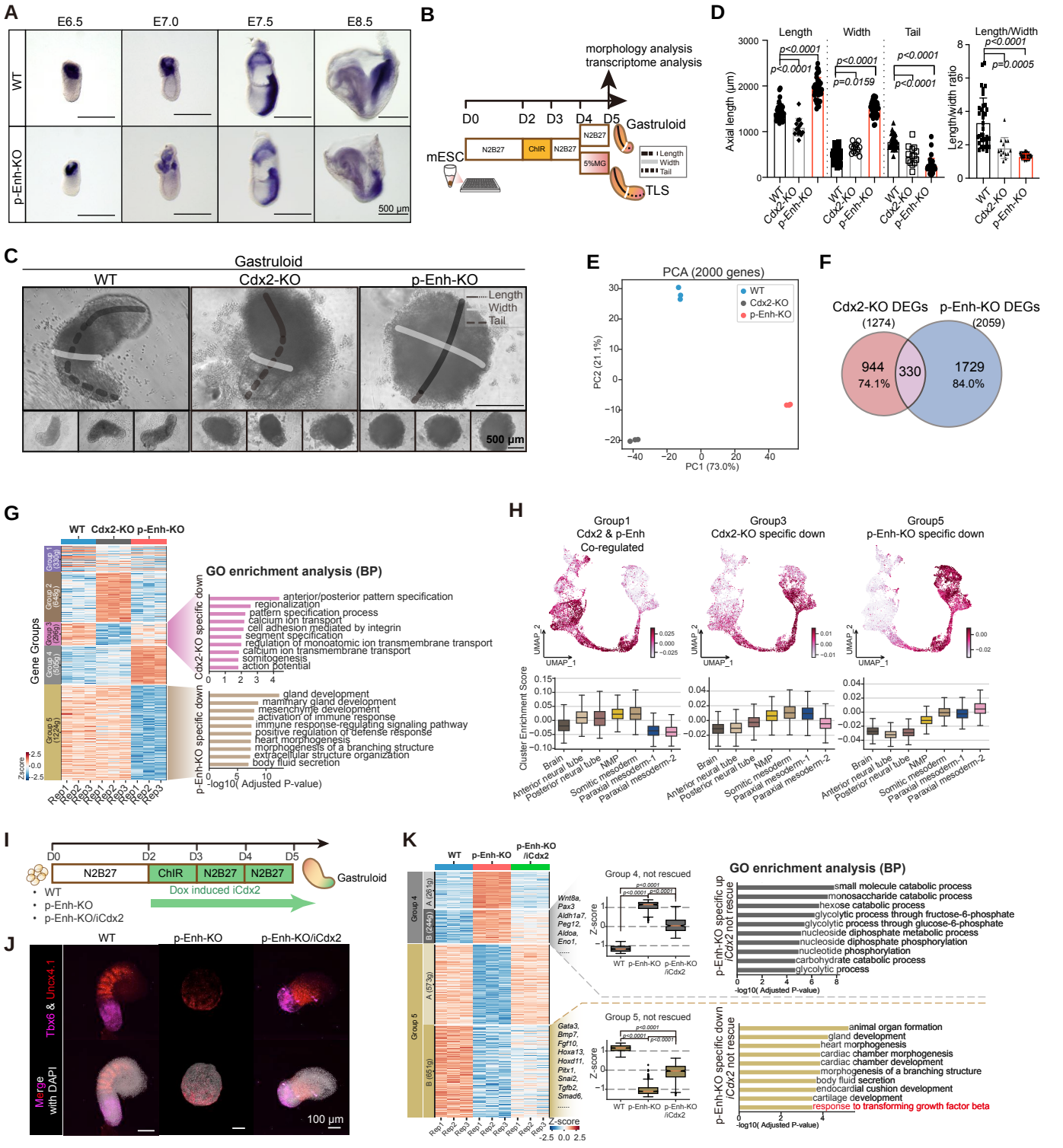

Fig. 4. Genome-wide epigenomic remodeling suggests a potential functional relevance between p-Enh and TGF-β signaling effectors.

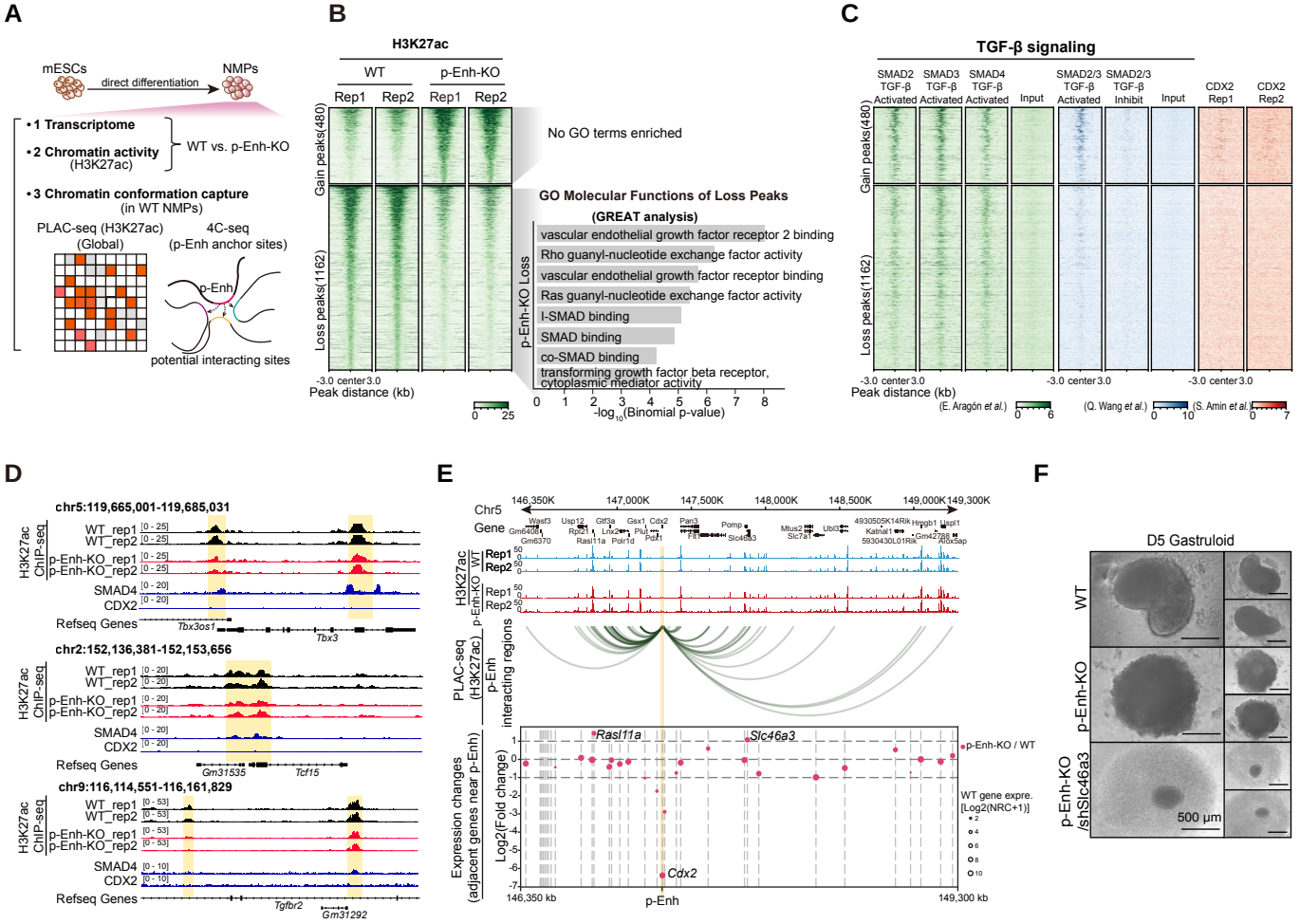

Fig. 5. The nuclear co-localization of p-Enh transcripts and SMAD4 protein at the *Hox* cluster regulates the posterior tissue development

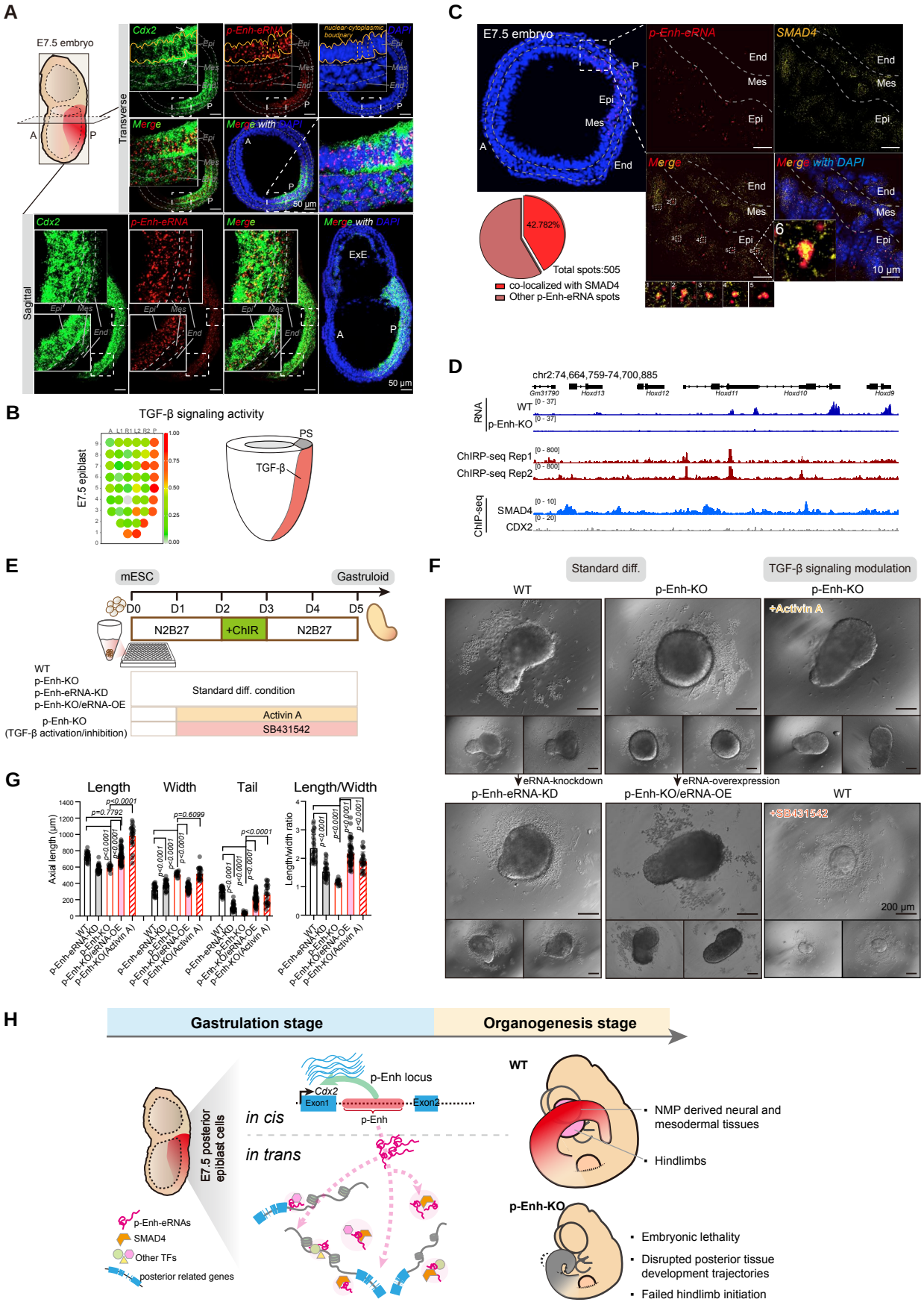

Fig. S1. screening and identification of posterior development related distal regulatory elements.

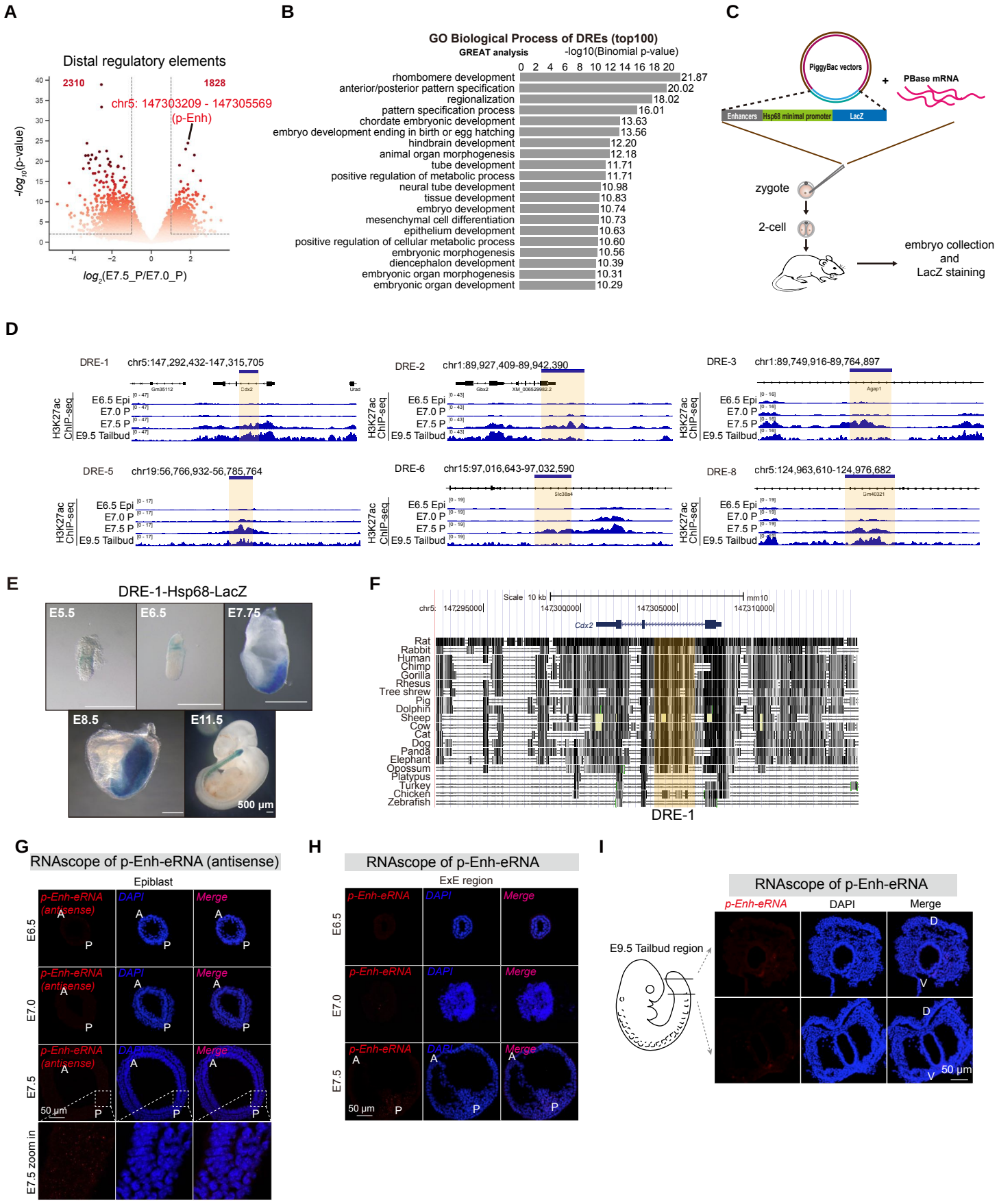

Fig. S2. p-Enh-KO mouse embryos exhibit embryonic lethality.

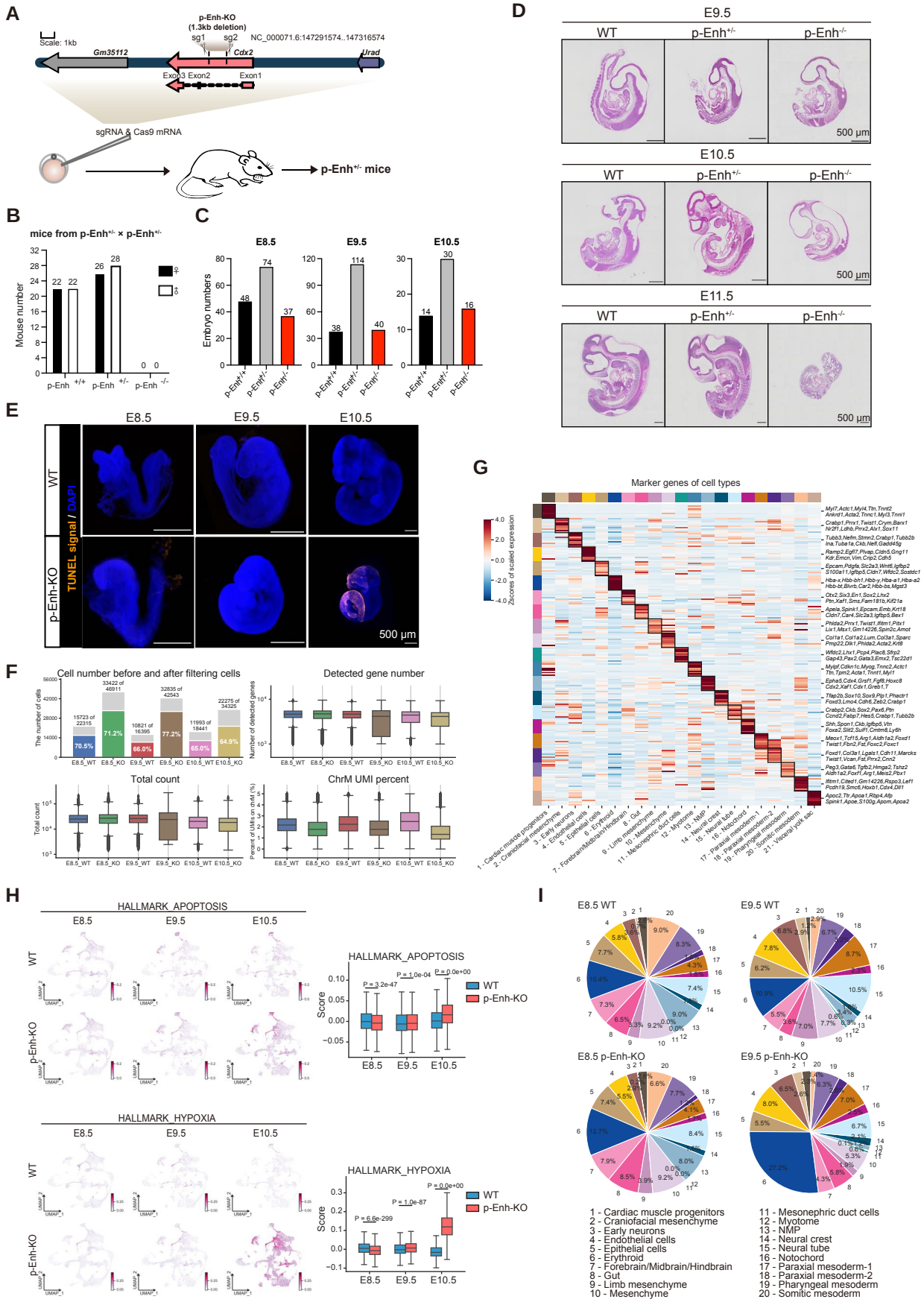

Fig. S3. A-P axis patterning and limb development exhibit deficiencies in p-Enh-KO embryos.

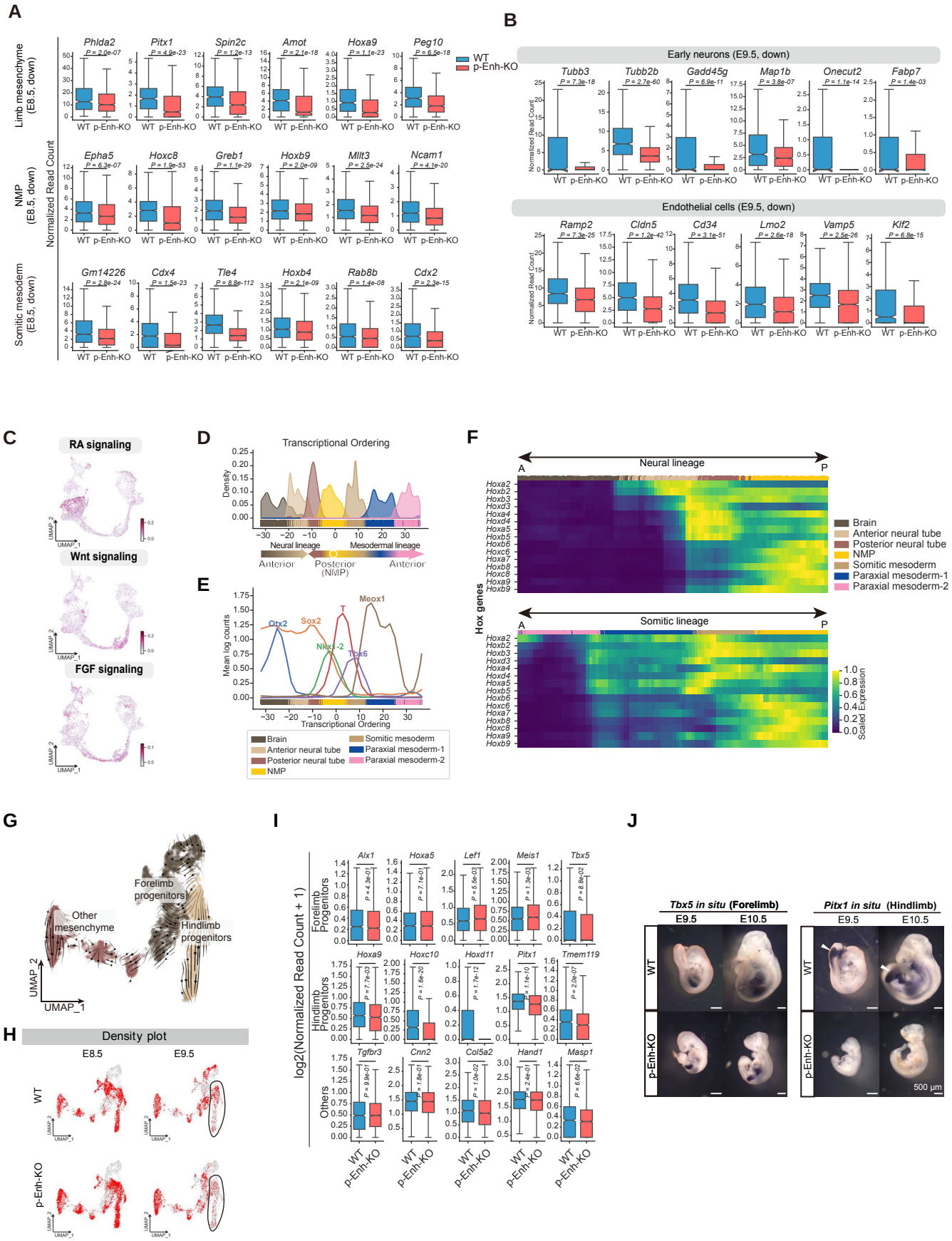

Fig. S4. Generation of p-Enh-KO and Cdx2-KO mESC and their differentiation deficiency.

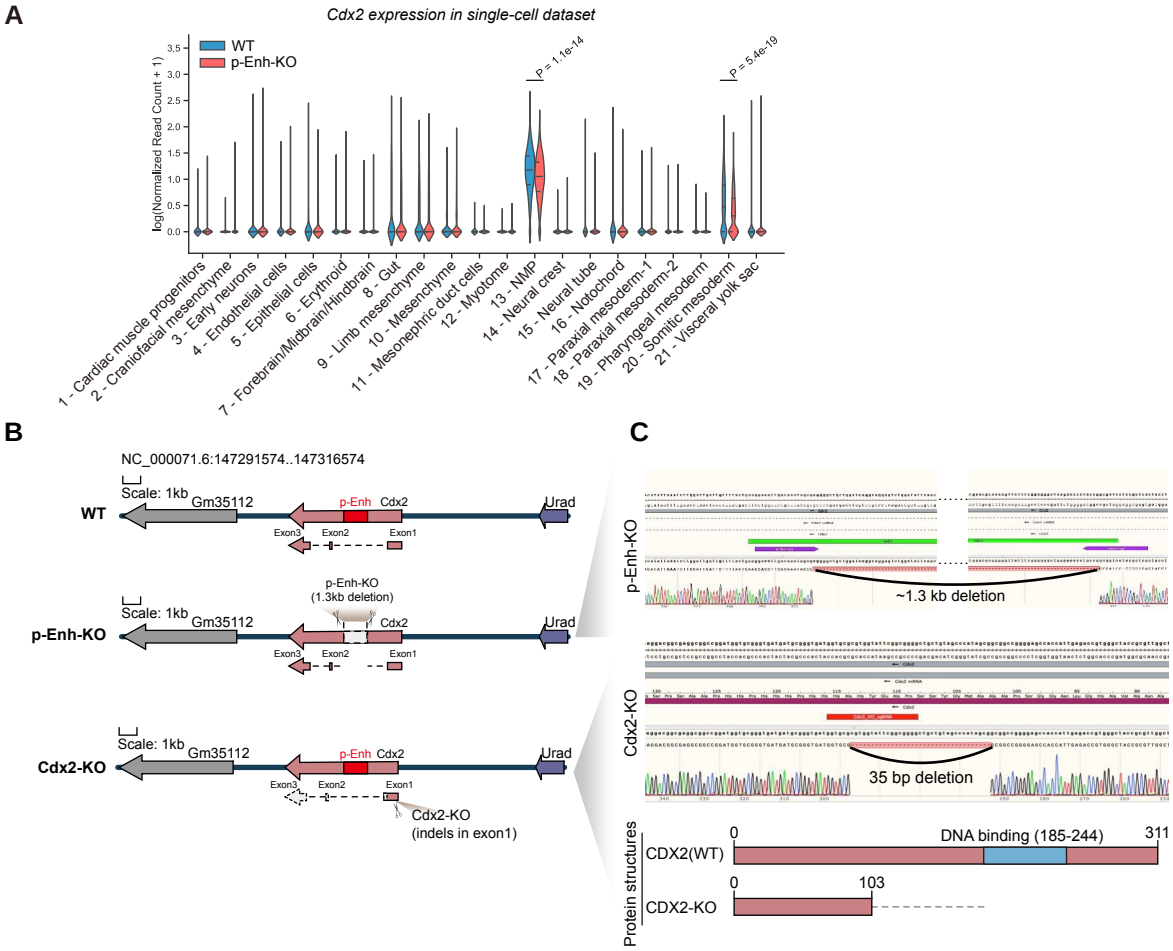

Fig. S5. Transcriptome analysis revealed distinct functions between p-Enh and *Cdx2*.

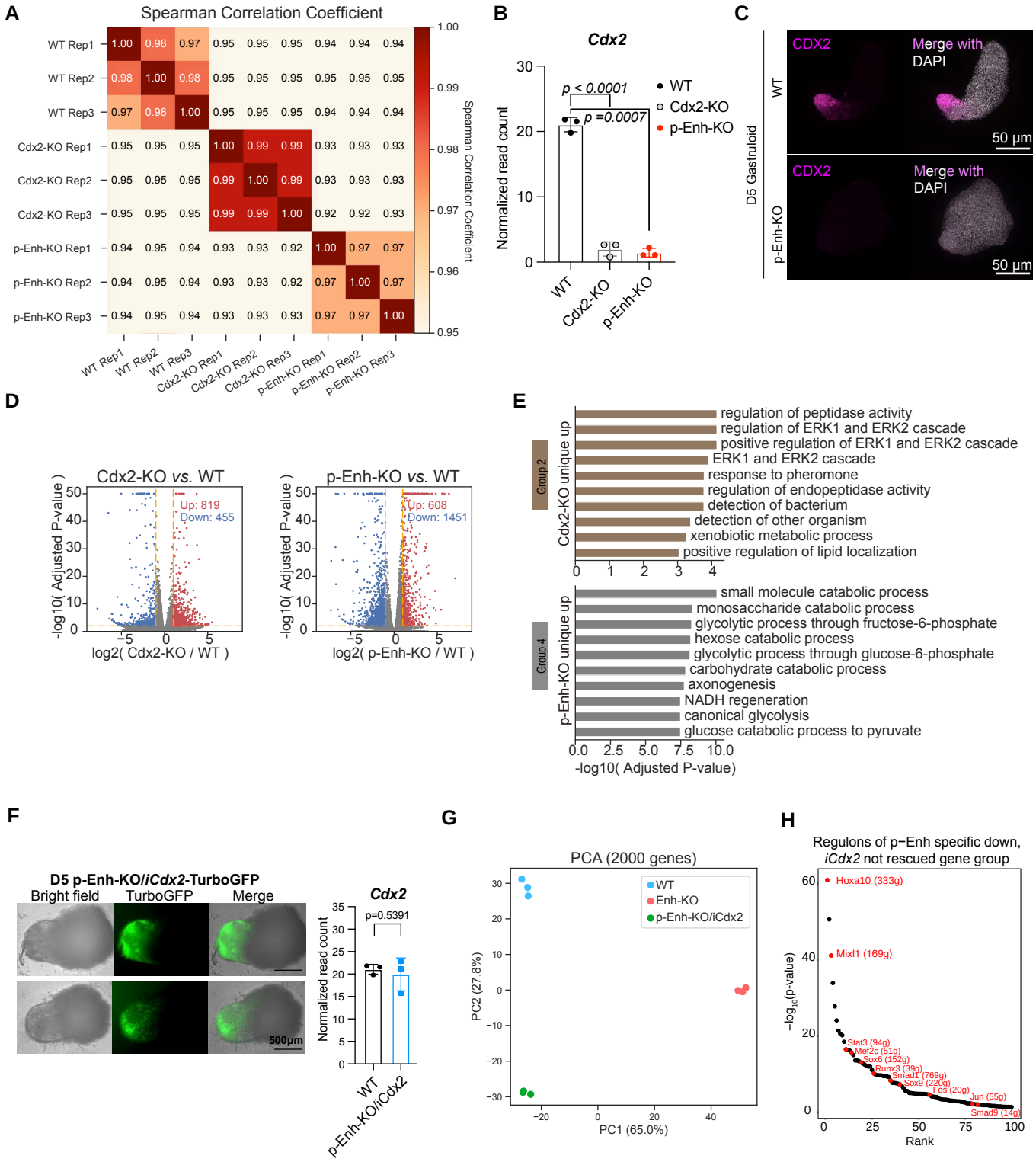

Fig. S6. Investigation of potential p-Enh interacting regions.

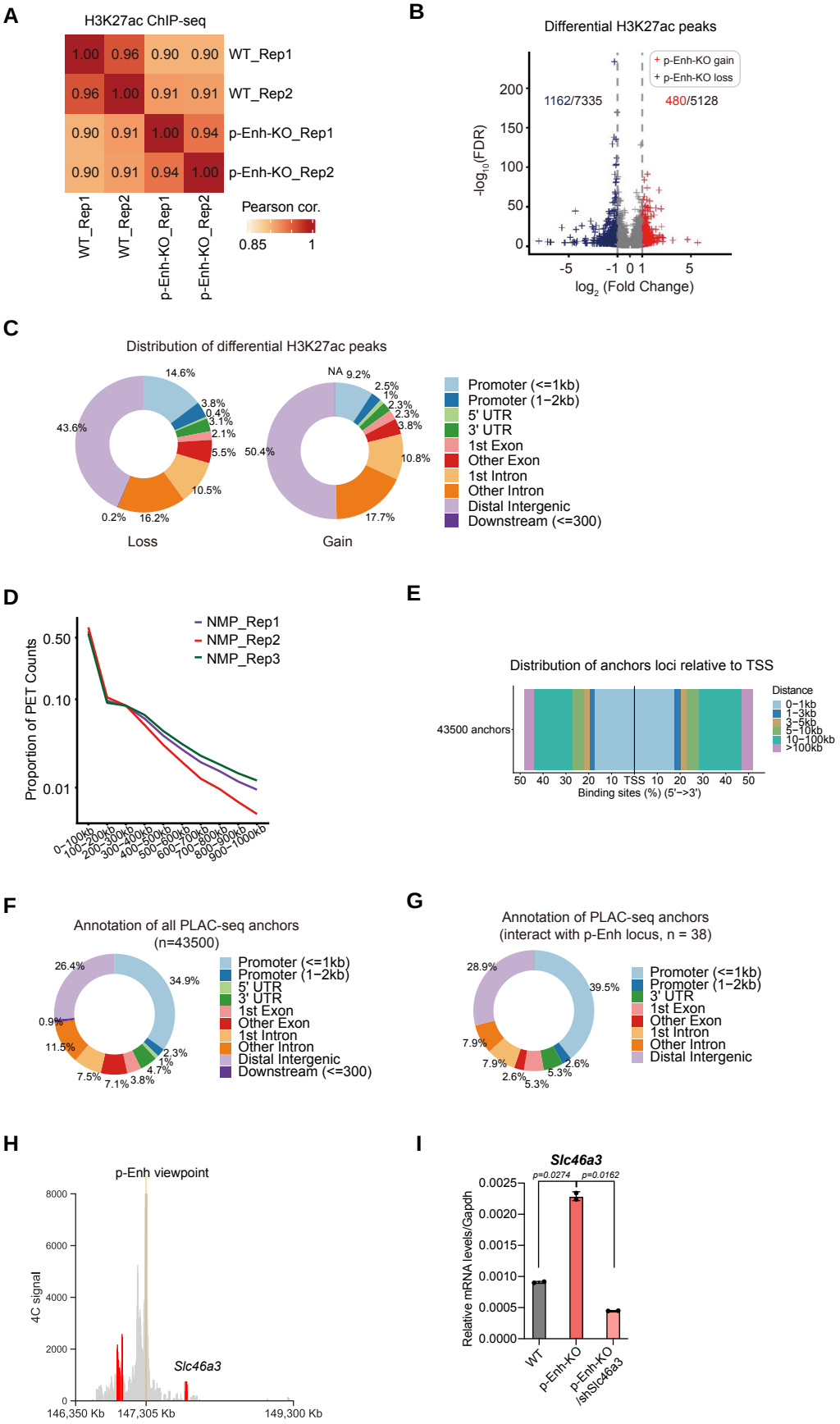
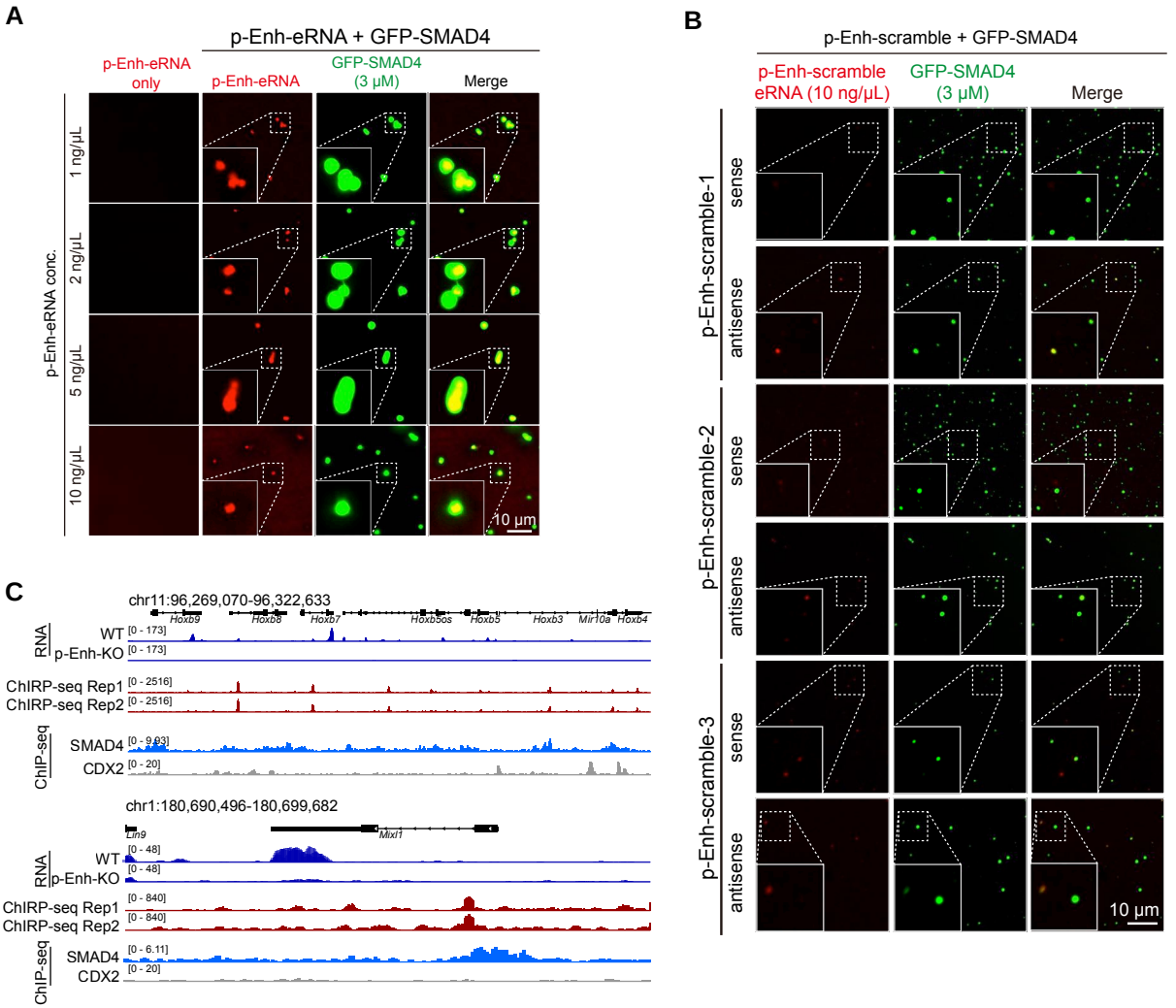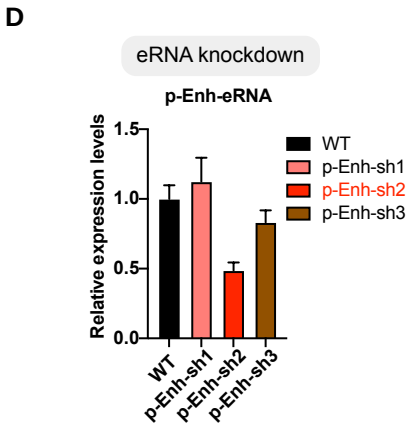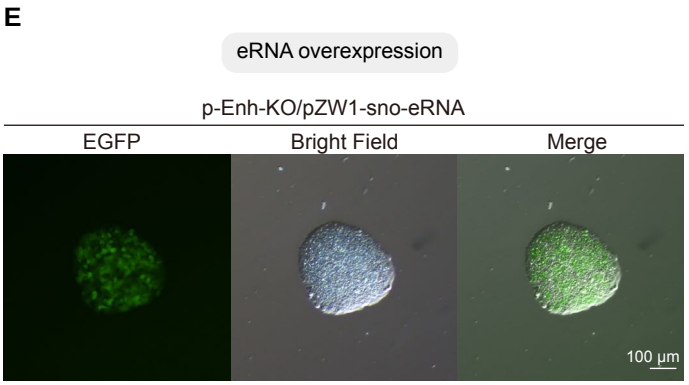
Fig. S7. Investigating the role of p-Enh-eRNA in posterior tissue patterning.