Spatial patterning of the epigenome during vertebrate gastrulation

Authors: Ana Paula Azambuja[1,2,3]†, Megan Rothstein[3]†, Tatiane Y. Kanno[1,2,3], Marcos Simoes-Costa[1,2,3]*

Affiliations:

[1]Department of Systems Biology, Harvard Medical School; Boston, MA, USA.

[2]Department of Pathology, Boston Children's Hospital; Boston, MA, USA.

[3]Department of Molecular Biology and Genetics, Cornell University; Ithaca, NY, USA.

*Corresponding author. Email: marcos@hms.harvard.edu

† These authors contributed equally to this work.

Summary

Embryonic patterning relies on the ability of cells to interpret the spatial information they receive from the environment. In vertebrates, this process begins with the localized activation of signaling systems that mediate axial specification. How embryonic cells record these spatial coordinates within their regulatory states has remained unclear. Here, we report that positional information is encoded within the chromatin landscape of embryonic cells. By utilizing spatially resolved genomic analysis in the avian gastrula, we found that the epigenome is organized in gradients of accessibility along the embryonic axes. Positional gradients of chromatin activation are established at the *loci* of developmental genes at the onset of gastrulation and can be used to predict the location of cells in amniote embryos. Our results show that axial specification involves the spatial patterning of the epigenome and support the existence of a Cartesian system of chromatin activation in the developing embryo.

Key Words: axial specification; gastrulation; embryonic patterning; positional information; epigenome; morphogen gradients.

Introduction

Embryonic cells must differentiate at precise locations to give rise to the adult body plan. This requires the generation of positional information, a set of spatial coordinates that reflect cellular location within a developing system[1-3]. Positional information has been a fundamental concept in developmental biology, offering a framework to explain how progenitor cells are patterned in space. It was outlined by Lewis Wolpert[1], based on his work in *Hydra* regeneration and on classical experiments performed by Hans Driesch, Thomas Hunt Morgan, Hildegard Stumpf and others[4]. Wolpert postulated that pattern formation is initiated by the establishment of gradients of signaling molecules along the embryonic axes, known as morphogens[5]. Cells that are closer to the source of the morphogen receive a higher concentration of the molecule, while cells further away receive a lower concentration[6]. Morphogen gradients would provide cells with spatial context, allowing them to differentiate into the appropriate cell type[1,3]. The concept of positional information predicted a number of subsequent discoveries, including the identification of the first morphogen molecule in *Drosophila*[7], and the demonstrations that extracellular signals lead to differential activation of the gene regulatory networks that determine cell fate[8,9].

Much of the positional information required for the establishment of the body plan emerges during axial specification[10]. This refers to the emergence of molecular asymmetries along the anterior-posterior (AP), dorsal-ventral (DV), and medial-lateral (ML) axes in the developing embryo. In *Drosophila melanogaster*, gradients of maternally derived transcription factors establish the embryonic axes, resulting in a system of spatial coordinates across the syncytial embryo[7,11,12]. In the cellularized vertebrate embryo, this process is mediated by signal transduction pathways. WNTs, BMPs, FGFs and retinoic acid signaling have all been implicated in generating spatial complexity during early development[13-15]. Ligands are produced by specific cells and generate concentration gradients via passive or active diffusion, and cells respond to these gradients by activating different sets of genes[8]. Varying concentrations of signals and patterning genes are thus thought to encode positional information[2]. Yet, further studies have questioned whether diffusible molecules are sufficient to drive positional specification and suggested that additional regulatory mechanisms could also mediate embryonic patterning[16].

A cellular process that may play a pivotal role in embryonic patterning is the regulation of the chromatin landscape. As early as 1969, long before the characterization of mechanisms of epigenetic regulation, Wolpert proposed that the establishment of positional information would require a coordinate system of spatial genomic responses in the embryo[1]. Nevertheless, our understanding of how the epigenome is remodeled during pattern formation is still limited. This is due to the technical challenges of profiling chromatin organization of embryonic cells while preserving their spatial context within the embryo. To address this, we took advantage of the large size of the avian embryo and performed spatially resolved genomic analyses during gastrulation. This approach allowed us to map gene expression and chromatin organization along the embryonic axes. Our results show that axial specification involves the spatial patterning of chromatin itself, and that the organization of the epigenome changes according to cellular position. We report that approximately 10% of the accessible cis-regulatory elements in the gastrula encode measurable positional values, which can predict where individual cells reside within the embryo, reconstruct gene expression patterns, and quantify cellular composition along the embryonic axes. These findings demonstrate that embryonic cells record

75 their spatial coordinates within the epigenome via the differential activation of specialized cis-regulatory elements.

Results

Reconstruction of gene expression patterns in the avian gastrula

To map the transcriptional signatures that underlie axial specification, we conducted a genome-wide reconstruction of gene expression patterns in the avian gastrula. Similar to the human embryo, the avian gastrula is a trilaminar structure shaped like a flat disc, making it suitable for this approach. We performed Tomography-Seq (tomo-seq)[17] in Hamburger Hamilton 5 (HH5) gastrulas[18], which were cryosectioned along the AP and ML axes (Fig. 1A). Individual 100 um tissue sections were then subjected to 3' RNA-seq (Fig. 1B). Tomo-seq allowed for genome-wide quantification of gene expression levels along embryonic axes, with normalized mRNA levels successfully recapitulating the axial domains of genes like *OTX2*, *MEOX2* and *TBXT* (Fig. 1C). Integration of AP and ML tomo-seq datasets allowed us to reconstruct of gene expression patterns in the avian embryo (Fig. 1D) and to simulate *in situ* hybridization experiments *in silico* for hundreds of genes (Fig. S1). Inferred gene expression patterns were consistent with those of known axial-restricted genes (Fig. 1C-E, Fig. S2A), including *OTX2*, *CHRD* and *WNT5A*. Identification of genes with similar expression patterns also allowed us to delineate gene signatures for morphological landmarks like the anterior neural plate, primitive streak, and the notochord (Table S1).

Next, we used the tomo-seq datasets to define genes involved in the AP patterning of the avian gastrula. Hierarchical clustering of gene expression profiles from two tomo-seq biological replicates (Fig. S2) identified distinct clusters of genes along the AP axis (Fig. 1F, Fig. S2B). We detected expression of approximately 13,500 genes (>50 counts across all slices), of which 1000 displayed a statistically significant association between gene expression and position along the AP axis. Of these, 706 genes were found to be highly enriched at the anterior pole of the embryo (Table S2), including previously characterized anterior factors such as *OTX2*, *SIX3* and *CYP26A* (Fig. 1G, Table S2)[19-21]. In contrast, only 294 genes were found to be posteriorly enriched. This cohort of genes contained *HOX* family members, and other posteriorizing factors like *CDX4* and *WNT5A* (Fig. 1G, Table S2)[22,23]. These results underscored an unexpected regulatory complexity of the anterior gastrula. Gene ontology (GO) analysis of anterior-enriched gene sets showed an enrichment of terms associated with neurogenesis, head development, and sensory structure morphogenesis, while those associated with posterior-enriched genes included mesenchymal tissue formation, including heart and skeletal morphogenesis (Fig. S2C). Averaging *in silico* expression patterns of all anterior genes showed that their expression was the strongest around the Hensen's node, whereas posterior genes were enriched around the primitive streak (Fig. 1H). Together, these results allowed for a high-throughput reconstruction of expression patterns along the AP and ML axes and led to the identification of AP gene signatures in the avian gastrula.

Spatial patterning of the epigenome during axial specification

Next, we examine the organization of epigenome along the AP axis of the gastrula. We divided HH5 avian embryos into seven transverse sections along the AP axis and subjected them to ATAC-seq analysis[24] (Fig. 2A, Fig. S3A,B). This tomoATAC-seq strategy led to the identification

of a set of cis-regulatory elements (CREs) distributed throughout the genome that display a graded pattern of chromatin accessibility along the AP axis (Fig. 2B,C). These genomic regions, which we termed positional cis-regulatory elements (pCREs), were also present in the *loci* of genes with regionalized expression patterns. For instance, the *OTX2* locus contained pCREs that were highly accessible at the anterior-most sections but displayed a gradually weaker ATAC signal in more posterior sections (Fig. 2B). The converse was observed in the *loci* of *WNT5A*, which contained many posterior pCREs (Fig. 2C). Quantification of normalized chromatin accessibility confirmed that, rather than recapitulating transcriptional domains, pCREs formed stereotypical gradients of accessibility along the AP axis (Fig. 2D, Fig. S4-S5). Chromatin gradients were not uniform within a given locus, as pCREs were often interspersed by elements that were not patterned or that were patterned along the opposing direction (arrow, Fig. 2B,C). We next compiled anterior and posterior sets of pCREs to assemble AP epigenomic signatures (Fig. 2E). Hierarchical clustering of the ATAC-seq datasets revealed patterns of chromatin accessibility along the AP axis (Fig 2E). Differential accessibility analysis identified two sets of 10913 and 12532 elements (out of 250,772 total ATAC peaks) that were significantly enriched at the anterior and posterior poles of the embryo, respectively, which we deemed anterior and posterior pCREs (Table S3).

To identify the main regulators of AP identity in the embryo, we conducted transcription factor motif enrichment analysis with HOMER[25] on anterior and posterior pCRE sets. Peaks active at the anterior end of the embryo were enriched for motifs for *OTX2*, *GSC*, *SOX3*, and *FOXA2*, while those active at the posterior end of the embryo contained *HOX*, *CDX2*, and *TBXT* motifs (Fig. S3C,D). Analysis of transcription factor motif activity across the seven ATAC-seq sections via chromvar[26] confirmed high variability of anterior and posterior transcription factor binding motifs along the AP axis (Fig. 2F). Furthermore, we found a strong correlation between chromatin accessibility and deposition of H3K27AC at anterior and posterior pCREs, confirming that these genomic regions were not only open, but also displayed polarized chromatin activation in the embryo (Fig. S4, Fig. S5). Similar to chromatin accessibility, H3K27AC deposition in pCREs was also graded along the AP axis (Fig. 2G). Together, these experiments show the existence of spatial gradients of chromatin activity along the AP embryonic axis. Such gradients are restricted to specialized elements in the genome that are present in the *loci* of genes with regionalized gene expression.

pCRE gradients are shaped by signaling pathways at the onset of gastrulation

To determine when pCRE gradients first appear during development, we analyzed chromatin patterning during at blastula and gastrula stages (Fig. 3A). HH1, HH2, and HH3 embryos were divided in five sections along the AP axis and subjected to CUT&RUN for H3K27AC. When visualizing trends of chromatin activity at anterior and posterior pCREs, we observed little to no chromatin polarization at the early (HH1) and late (HH2) blastula stages. Graded activation of pCREs became conspicuous at HH3, coinciding with onset of gastrulation (Fig. 3A, Fig. S6). To verify how pCRE polarization correlates with transcription, we also performed RNA-seq along the AP axis in these early stages. The results showed that expression of the tomo-seq AP gene sets (Fig. 1) became regionalized following the establishment of pCRE gradients (Fig. S6). Analysis of *loci* individual genes support that

chromatin patterning precedes gene expression, as formation of pCRE gradients can already be detected in the *loci* of genes of *DMBX1* and *HOX11A* at HH3, before transcription begins (Fig. S7A-B). These results indicate pCRE patterning is not a consequence of regionalized gene expression.

We next examined the mechanisms that shape chromatin gradients during gastrulation. A well-known positional cue involved in the establishment of the anteroposterior axis is WNT signaling[14]. This highly conserved signaling system acts as a posteriorizing agent across metazoans[27]. We thus tested the hypothesis that pCREs are directly regulated by nuclear effectors of WNT signaling. Several lines of evidence from the analysis of our tomo-seq datasets point to a role for WNT signaling in the establishment of pCREs. First, a number of WNT ligands and WNT pathway-associated genes are restricted in expression along the AP axis (Table S2). Moreover, motif enrichment analysis of posterior pCREs revealed robust enrichment of motifs for nuclear effectors of WNT signaling, including LEF1 and TCF-like factors (Fig. 2F). To investigate the relationship between pCREs and WNT signaling, we performed CUT&RUN for LEF1/CTNNB1 in three sections along the AP axis of the HH5 embryo (Fig. 3B). Analysis of our CUT&RUN datasets revealed 1558 peaks, representing genomic regions bound by LEF1/CTNNB1. Motif enrichment analysis of LEF1/CTNNB1 peaks showed TCF/LEF as the highest ranked motifs, confirming our ability to isolate regions bound by these factors (Fig. 3C). Furthermore, consistent with the posteriorizing role of WNTs, we found that LEF1/CTNNB1 were predominantly associated with posterior pCREs (Fig. 3D), and that CUT&RUN signal displayed an anterior to posterior gradient (Fig. 3D-E). These results indicate that posterior pCREs are directly regulated by components of the WNT signaling pathway.

To test if WNT signaling is consequential for chromatin patterning, we next assessed the effect of modulation of WNT signaling on pCREs. We isolated cells of the anterior epiblast at stage HH2, prior to the activation of pCREs, and cultured the cells for 24 hours under WNT gain-of-function (GOF) and loss-of-function (LOF) conditions (Fig. 3F). The media was supplemented with either a WNT pathway agonist, CHIR99021[28], a WNT antagonist, XAV939[29], or DMSO as a control. Analysis of anterior- and posterior-enriched pCREs revealed a significant change in accessibility upon WNT modulation. We observed a decrease in the accessibility of anterior CREs upon treatment with CHIR99021, and a concomitant increase in accessibility following XAV939 treatment (Fig. 3G-H). The opposite trend was observed for posterior CREs, which displayed an increase in accessibility following CHIR treatment and a decrease in accessibility upon treatment with XAV939. Thus, posteriorizing signals increase the accessibility of posterior pCREs, while inhibiting their anterior counterparts.

We next examined how peaks bound by LEF1/CTNN1B responded to manipulations in Wnt signaling. We projected LEF1/CTNN1B-bound peaks onto our WNT modulation epigenomic dataset (Fig. 3I). As expected, genomic regions associated with LEF1/CTNN1B binding were positively regulated by WNT signaling (i.e., their accessibility is promoted by CHIR99021 and repressed by XAV939). The integration of these datasets allowed us to compile a list of genomic targets of canonical WNT signaling that respond to the manipulation of the pathway (Table S4). Finally, we examined the spatial patterns of accessibility in WNT-responsive genomic regions to examine if they display features of pCREs. Projection of the accessibility of WNT-responsive peaks (Fig. 3J) in our tomoATAC-seq datasets showed that they display AP polarization and

graded accessibility along the AP axis (Fig. 3K). As expected, peaks activated by WNTs formed posterior to anterior gradients, while peaks inhibited by the signaling system are anteriorly polarized (Fig. 3J,K). These results show how signaling systems modulate spatial gradients of chromatin activity to endow cells with spatial information.

Spatial chromatin gradients predict axial position of embryonic cells

The finding that pCREs are established during gastrulation led us to postulate that these elements encode positional information in embryonic cells. To test this, we first examined if pCRE signatures (Table S3) could predict the position of embryonic cells along the AP axis. We thus performed single cell multiome analysis[30] of the avian gastrula using the 10X genomics platform to measure both chromatin accessibility and gene expression in individual cells (Fig. 4A). This analysis allowed us to estimate cellular AP position based on pCRE accessibility, and to subsequently test the accuracy of this prediction based on the gene expression signatures from the tomo-seq analysis (Fig. 1). Upon filtering for high quality cells (n=7750) and performing dimensionality reduction, we identified 16 distinct clusters within the multiome dataset (Fig. 4B). Clusters were distributed similarly within the UMAP projection regardless of which dataset was used for dimensionality reduction (RNA-seq, ATAC-seq, or both) (Fig. S8).

Cellular identities were assigned to individual clusters using a combination of cluster-specific marker gene expression from scRNA-seq, along with motif enrichment analysis from scATAC-seq (Fig. S9). Cluster annotation showed separation of cell identity primarily based on their germ layer of origin. Of the 16 identified clusters, six clusters were mesodermal, four clusters were ectodermal, and one cluster was endodermal (Fig. 4B). Extraembryonic tissues, including blood islands, extraembryonic endoderm, and the area opaca were also identified in the dataset. After cluster assignment, we attempted to predict the position of individual cells using our anterior and posterior epigenomic signatures. First, we examined whether anterior and posterior pCREs were associated with regionalized gene expression. We employed peak-gene co-accessibility analysis[31] in the multiome dataset to assign pCREs to their cognate target genes (Fig. 4C) and visualized the expression profiles of these genes along the AP axis. As expected, genes found to be co-accessible with AP pCREs displayed spatial enrichment of anterior or posterior tomo-seq gene signatures (Fig. 4D).

Next, we attempted to order cells along the AP axis based on epigenomic information. We utilized a quadratic programming approach[32,33] to calculate the fractional identity of each cell based on its similarity to the anterior vs. posterior pCRE signatures (Fig. 4E). This strategy allowed us to assign the approximate position of each cell based on pCRE score (Fig. 4F). To test the accuracy of pCRE-based cell ordering, we projected transcriptomic information from the multiome on the ordered cells. We first assessed the expression of the anterior- and posterior-enriched gene sets identified in the tomo-seq analysis (Fig. 1G, Table S2). We found that these gene sets were enriched in the anterior and posterior positioned cells, validating our cell ordering analysis (Fig. 4G). Reconstruction of expression patterns of individual AP genes, including *OTX2*, *WNT5A*, and *TCF7*, were consistent with their expression domains (Fig. 4H). This was also true for genes expressed in intermediate AP positions within the embryo, such as *CHRD,* indicating that pCRE gradients contain positional information along the entire embryonic

axis. Cell ordering was also able to reconstruct the early *HOX* code in the gastrula (Fig. 4H). Expression patterns were consistent with chromvar analysis, which identified enrichment of motifs for regulators at their predicted expression domains (Fig. 4I). These results indicate that pCREs can be used to order cells *in silico* along the AP axis and support that a cell's positional information is encoded within its chromatin landscape.

pCREs encode positional information within embryonic germ layers

Our multiome analysis demonstrated that the avian gastrula is composed of 16 different cell types (Fig. 4B). We next tested whether we could use the combination of pCRE signatures and quadratic programing to define how these cell types are organized along the AP axis. We first examined the cluster identity of cells enriched for anterior vs. posterior pCREs. We found that the anterior cells are largely ectodermal and endodermal, while the posterior cells consist mostly of mesoderm (Fig. 5A). Indeed, projection of the anterior pCREs on the single-cell multiome UMAP shows robust signal in the ectodermal and endodermal clusters, whereas the posterior signal is strongest in the mesodermal cluster (Fig. 5B). To examine AP composition of the embryo in more detail, we generated 50-cell bins ordered from the anterior to the posterior poles of the embryo. We then assigned the identity of the cells in each bin to determine the regional abundance of distinct cell types (Fig. 5C). This confirmed that a larger percentage of ectodermal and endodermal cells were positioned in the anterior half of the embryo, while mesodermal clusters were posteriorly enriched. To verify the accuracy of this prediction, we performed immunostaining in sagittal sections of the avian gastrula with antibodies for markers for the three germ layers (Fig. 5D)[34-36]. Consistent with the cellular composition analysis, endodermal (SOX17+) and ectodermal (SOX2+) progenitors were enriched in the anterior region of the embryo, whereas the posterior gastrula is predominantly mesodermal (TBXT+). Furthermore, we also observed that cells assigned to distinct embryonic structures, including the primitive streak and the notochord, occupied the expected position along the AP axis (Fig. 5C).

The asymmetric distribution of cell types along the AP axis led us to test if pCREs could be used to predict the spatial distribution of cells within individual germ layers. We subdivided our multiome dataset into ectodermal, mesodermal, and endodermal clusters (2481, 2936, and 440 cells, respectively) (Fig. S10). Projection of the anterior- and posterior-enriched pCREs onto germ layer-specific subclusters revealed distinct subpopulations of cells enriched for axial-restricted peaks in each of the three germ layers (Fig. 5E, Fig. S10). Thus, although unsupervised clustering largely groups cells based on cellular origin, a closer look at intra-cluster organization suggests additional separation based on positional identity. To test this, we examined the axial identity of cells from neural ectoderm, which undergoes early regionalization along the AP axis during the formation of the central nervous system[37]. We employed pCRE-based cell ordering in the neural ectoderm subcluster (708 cells), and assessed the expression of the forebrain, midbrain, and hindbrain markers (*OTX2*, *GBX2*, and *HOXA2*, respectively) (Fig. 5E)[38,39]. Ordered cells displayed the expected expression profiles of these marker genes along the AP axis (Fig. 5E). Thus, pCREs encode cellular positional information within a single tissue or germ layer. These results support a decoupling of the molecular mechanisms controlling cell identity and axial identity.

Conservation of pCRE signatures across vertebrates

Spatial epigenomic analysis of the avian gastrula allowed for the identification of anterior and posterior pCREs, which provide a molecular signature of cellular position in the embryo. To determine if this mechanism is present on other vertebrate embryos, we analyzed multiomic data from the E7.5 mouse embryo produced by Argelaguet and colleagues[40]. Liftover of anterior and posterior pCREs[41] identified 1560 anterior and 1724 posterior pCREs conserved in mouse. We then used quadratic programming to assign the relative position of individual cells from the E7.5 mouse dataset along the AP axis (Fig. 5F). Upon ordering cells based on their chromatin accessibility at conserved pCREs, we assessed the expression patterns of anterior- and posterior-restricted genes as 200-cell bins along the AP axis. This analysis revealed spatial restriction of anterior genes *OTX2* and *POU3F1*, as well as posterior genes *TBXT* and *WNT3* in the developing mouse gastrula (Fig. 5G)[42]. Therefore, even a small subset (~15%) of pCREs can predict anterior/posterior gene expression across species. These results support a general role for pCREs in recording positional information in vertebrate embryonic cells.

Discussion

The amniote gastrula consists of a trilaminar disc-like structure with a few morphological landmarks, such as the organizer and the primitive streak[43]. Our mapping of spatial patterns of gene expression and chromatin activation in avian embryos underscored that this apparent simplicity belies a large degree of spatial regulatory complexity. This analysis also revealed that gradients of chromatin activity are formed during axial specification. According to the concept of positional information, embryonic patterning involves the establishment of a graded set of values that varies continuously along a field of progenitor cells[1,3]. These values have been generally thought to be generated by morphogenetic gradients[44], but our results show that they are also recorded within the epigenomic landscape. We found that positional gradients of chromatin activity occur at specialized genomic elements, which we dubbed positional cis-regulatory elements (pCREs). pCREs comprise a significant percentage of accessible regions in the genome (~10%), and are often interspersed with other open chromatin regions that are not spatially patterned. These elements are generally associated with developmental genes with axially-restricted expression domains, with some sections of the genome (e.g., the *HOX* clusters) displaying higher density of pCREs. We also observed that chromatin patterning occurs within the germ layers, suggesting a shared system of positional information is used by distinct cell types.

Chromatin organization has been shown to be dynamic and cell type dependent, implying a sophisticated level of regulation that responds to internal and external cues[45,46]. Here we show that the chromatin landscape also varies depending on cellular position. While it is often intuited that chromatin accessibility and activity recapitulate gene expression, our results show that these epigenomic patterns are initially set up as slopes that run across the embryonic axes. There are thus significant differences between the domains of pCRE accessibility and transcription of their cognate genes. For instance, cis-regulatory elements at the *loci* of an anterior gene like *OTX2* are still accessible at posterior cell populations, although at a lower level than what is observed in the anterior region of the embryo where the gene is expressed. This suggests that the interpretation of ATAC-seq data from developmental systems must consider the spatial context of the cells being analyzed. We also observed that the establishment of chromatin gradients precedes regionalized transcription, indicating that patterning of the epigenome may play a role in differential gene expression. Understanding how a Cartesian system of chromatin activation supports the emergence of complex tridimensional expression patterns will require a deeper understanding of both physical and functional interactions between cis-regulatory elements. Identification of the principles that allow processing of spatial information at the chromatin level will also shed light on how gradients of transcription factors [47] and signaling systems[48] regulate cell fate commitment.

Gradients, whether of secreted ligands, inhibitors, or transcription factors, have been an important concept in developmental biology, often described as necessary for the generation of spatial complexity[9,49,50]. The cellularized vertebrate embryo requires the transfer of spatial information across cell membranes, thus relying on the action of extracellular ligands[9]. Our results indicate that spatial information from signaling systems gets recorded within the epigenome, with chromatin organization retaining the patterns of activity of morphogen gradients. Indeed, we showed that canonical WNT signaling directly regulates posterior pCREs,

with WNT nuclear effectors binding to pCREs in a graded fashion along the AP axis. Thus, chromatin patterning may depend upon two important factors: the establishment of gradients of activity of nuclear effectors and the ability of CREs to interact with nuclear effectors in a concentration-dependent manner[51,52]. We propose that epigenomic patterning depends on the integration of the spatial information provided by the multiple signaling systems involved in axial specification. This is supported by the abundance of nuclear effector motifs of WNTs, TGFB, HIPPO, and retinoic acid signaling in pCREs (Fig. 2). Gradients of chromatin accessibility may also be downstream of numerous mechanisms that confer robustness to signaling responses[53,54], and thus constitute a more reliable readout of positional information than morphogen levels.

Our results also inform upon how embryonic cells are allocated to anterior or posterior fates during gastrulation. Existing models for axial specification in vertebrates have focused on the development of the central nervous system (CNS)[55-57]. Based on grafting experiments performed in *Xenopus*, Nieuwkoop proposed a two-signal model that states the CNS is first induced with an anterior 'default' identity that is subsequently posteriorized by signals from the trunk[58]. Thus, the two-signal model states neural induction precedes the allocation of embryonic cells into anterior or posterior fates[55]. In contrast, classical work from Otto Mangold supported an earlier acquisition of axial identity by embryonic cells. After grafting different cell populations derived from the dorsal blastopore lip into host embryos, Mangold reported that different parts of the organizer induce CNS cells of distinct AP identities[59]. His results suggest that different signals drive induction of the anterior or posterior regions of the CNS. Our observation that anterior and posterior epigenomic signatures arise simultaneously at the onset of gastrulation is consistent with Mangold's results and align with recent findings that indicate axial identity of CNS progenitors arises sooner than previously thought. In a study utilizing cell reprogramming experiments, Metzis and colleagues observed that spinal cord cells acquire posterior identity before undergoing neural induction[56]. Our results indicate that axial identity is established at the onset of gastrulation and that it can be measured by a positional signature that is recorded within the chromatin landscape of embryonic cells.

380 **Acknowledgments:** The authors thank Peter Schweitzer from Cornell University BRC Genomics Core Facility for all help generating the multiome libraries and next-generation sequencing. We also acknowledge the Nikon Imaging Center at Harvard Medical School for microscopy and image analysis resources.

Funding: National Institutes of Health grant DP2 HD102043.

385 **Author contributions:** Conceptualization: MSC, APA, MR; Methodology: MSC, APA, MR, TK; Investigation MSC, APA, MR, TK; Visualization: MSC, MR, APA, TK; Funding acquisition: MSC; Project administration: MSC; Supervision: MSC; Writing: MSC, MR.

Declaration of interests: Authors declare that they have no competing interests.

Data and materials availability: All datasets generated in this article have been deposited to
390 the Gene Expression Omnibus: GSE224724

Supplementary Materials:

Figs. S1 to S10

Tables S1 to S4

395

Materials and Methods

Data availability

All datasets have been deposited to the Expression Omnibus: SuperSeries GSE224724.

Tissue collection

Fertilized Leghorn White chicken eggs were obtained from the Department of Animal Science, University of Connecticut. Eggs were incubated at 37º C and 50% humidity until embryos reached the desired developmental stage. Embryos were collected in Ringer's solution, staged according to Hamburger and Hamilton[18], and prepared for genomic protocols as described below.

Tomo-seq sample and library preparation

HH5 embryos were collected, staged, and cleaned in RNase-free solution. Next, under a ZEISS Stemi 508 stereo microscope, the anterior-posterior and lateral axes of each embryo were measured. Similar sized embryos were then transferred to freezing molds, positioned depending on the axis to be sectioned and embedded in frozen section medium (Thermo Scientific™ Richard-Allan Scientific Neg 50™, #6502). Before freezing, each embryo had its Hensen's node, anterior-posterior and lateral ends marked with dye to serve as landmarks for the sectioning process. In a RNase-free environment, frozen samples were trimmed to avoid extra-embryonic tissues and frozen medium excess, and cryosectioned into 100 um. Each anterior-posterior or lateral section was then collected for total RNA extraction with the RNAqueous-Micro Total RNA Isolation Kit (Thermo Fisher Scientific, # AM1931) according to the kit's protocol and processed for 3'RNA-seq library preparation. Libraries were prepared with the kit QuantSeq 3'mRNA-seq Library Prep Kit (FWD) for Illumina (Lexogen, #K01596), following the manufacturer's recommendations for low RNA input samples. Final libraries were quantified, pooled and sequenced with single-end 75bp reads on an Illumina NextSeq500 instrument.

Tomo-seq data analysis

Trimming of adapter sequences and poly(A) tails was performed using BBDuk (version 38.93) using the following parameters: "ref=polyA.fa.gz,truseq_rna.fa.gz k=13 ktrim=r useshortkmers=t mink=5 qtrim=r trimq=10 minlength=20". Next, STAR (version 2.7.3a)[60] was used to align trimmed reads to the galGal6 chicken reference genome using the following arguments: "--outFilterType BySJout --outFilterMultimapNmax 20 --alignSJoverhangMin 8 --alignSJDBoverhangMin 1 --outFilterMismatchNmax 999 --outFilterMismatchNoverLmax 0.1 --alignIntronMin 20 --alignIntronMax 1000000 --alignMatesGapMax 1000000 --outSAMattributes NH HI NM MD --outSAMtype BAM SortedByCoordinate". Bam alignment files were indexed using samtools (version 1.10)[61] and reads were counted with htseq-count (version 0.13.5)[62] using a GTF file containing chicken gene annotations and using the following parameters: "-m intersection-nonempty -s yes -f bam -r pos". Differential gene expression analysis was performed on two embryo replicates using TradeSeq (version 1.8)[63]. Anterior- and posterior-enriched genes were identified using TradeSeq's startVsEndTest function and genes with a p-value <0.01 and a log fold change >1 or <-1 were considered significantly enriched. For plotting of AP and 2D gene expression profiles, raw counts were normalized to the total number of

spike-in reads and the moving average of the normalized counts were plotted along each axis. Gene ontology analysis was performed using PANTHER's[64,65] Fisher's Exact test with Bonferroni correction.

In situ hybridization

For *in situ* hybridization, embryos were fixed in phosphate buffer saline (PBS) containing 4% paraformaldehyde (PFA) for 2 hours at room temperature or overnight at 4 ºC. Following fixation, embryos were dissected, washed with PBS containing 0.1% Tween (PBST), dehydrated and stored in methanol at -20º C. Whole-mount in situ hybridization was performed as previously described[66].

HCR RNA-FISH

Hybridization chain reaction RNA fluorescent in situ hybridization (Molecular Instruments HCR v3.0) was performed in HH5 (19-22h) whole mount chick embryos[67]. Briefly, embryos were collected and fixed in 4% paraformaldehyde for 2 hours at room temperature. Embryos were dissected to remove the vitelline membrane, washed with PBST on ice and then stored in methanol at -20° C until use. Next, embryos were rehydrated with a series of graded Methanol/PBST washes on ice. For the detection step, embryos were pre-hybridized with probe hybridization buffer for 30 minutes at 37° C, then incubated with 4 pmol of each probe set overnight. The excess of probe was removed by washing the embryos with probe wash buffer at 37° C. The embryos were then pre-amplified with amplification buffer and incubated with hairpin solution overnight in the dark at room temperature. Excess hairpins were removed by washing the embryos with 5X SSCT at room temperature. Embryos were incubated with DAPI for 1 hour followed by sample mounting for microscopy. Embryos were imaged at the Nikon Imaging Center at Harvard Medical School using a Nikon AX R point scanning confocal. The target mRNA probe sets, amplifiers and buffers were purchased from Molecular Instruments (https://www.molecularinstruments.com/).

Immunohistochemistry

For whole-mount immunohistochemistry, embryos were collected and fixed in 4% PFA-PB for 20 minutes at room temperature. Post fixation, embryos were dissected from the filter paper and washed in TBS containing 0.1% Triton and 1% DMSO (TBTD). Embryos were blocked at room temperature for 2 hours in TBTD supplemented with 10% donkey serum and incubated in anti-SOX2 (AbCam, #ab97959) and anti-SOX17 (R&D, #AF1924) primary antibodies diluted in blocking solution, overnight at 4º C. Following the primary antibody incubation, embryos were washed, blocked for 30 minutes at room temperature, and stained with appropriate secondary antibodies and anti-Brachyury Alexa Fluor 488 Conjugated (R&D, #IC2085G) for 2 hours at room temperature. Secondary antibodies used included donkey anti-rabbit/goat IgG conjugated with Alexa Fluor594/647 (Molecular Probes). Following the secondary antibody step, the embryos were washed, stained with DAPI and post-fixed with 4% PFA for 1 hour, prior to imaging. Whole-mount images were collected using an upright Zeiss Axio Imager fluorescent microscope.

TomoATAC-seq

HH5 chicken embryos were collected in Ringer's solution and stage- and size-matched embryos were dissected out of any extra embryonic tissue. Then, using the Hensen's node as a landmark, seven sections were obtained along the antero-posterior axis, three anterior to the node and four posterior to it. Five position-matching sections were pooled and dissociated in Accumax (Innovative Cell Technologies, #AM105). Cell suspensions were washed and processed using the OMNI-ATAC-Seq protocol[68].

TomoATAC-seq data analysis

Trimming of Nextera adapter sequences was performed using cutadapt (version 3.4)[69] using the following parameters: "-a CTGTCTCTTATACACATCT -A AGATGTGTATAAGAGACAG --minimum-length=25". Trimmed reads were aligned to the galGal6 chicken reference genome using bowtie2 (version 2.4.4)[70] with the following arguments: "--local --very-sensitive-local --no-unal --no-mixed --no-discordant -X 2000". Bam alignment files were indexed using samtools (version 1.10), followed by marking of duplicate reads using Picard's (version 2.26.2) MarkDuplicates function. Duplicate reads were then removed using samtools view with the following arguments: "-F 1804 -f 2". Peaks were called with macs2 (version 2.2.7.1)[71] using the parameters: "-f BAMPE -g 1e9 -q 0.05 --call-summits" and bigwigs were generated from Bam files using DeepTools' bamCoverage function (version 3.5.1)[72] with the following arguments: "--binSize 5 --normalizeUsing RPKM --extendReads". Reads in peaks were counted using DiffBind (version 3.4.11)[73]. Peaks with anterior- and posterior-enriched accessibility were identified using TradeSeq's startVsEndTest function and peaks with a p-value <0.01 and a log fold change >1 or <-1 were considered significantly enriched. For plotting of accessibility profiles, raw counts were library size normalized and the moving average of the normalized counts were plotted along the AP axis. Motif enrichment analysis was performed by calculating motif deviations using ChromVar (version 1.16), as well as using HOMER's[25] findMotifsGenome.pl function.

Anterior epiblast ATAC-seq

As described previously, chick embryos were obtained from fertile eggs, collected, and staged in Ringer's solution. HH2 embryos were dissected off any extra-embryonic tissue and had the anterior hypoblast removed. Then, the anterior, naive, epiblasts from three embryos were microdissected, pooled, and transferred to 96-well plates previously coated with fibronectin and containing 100 uL of 10% FBS DMEM culture media containing the WNT agonist CHIR99021 (Tocris, #4423), or antagonist XAV939 (Sigma, #X3004-5MG) at the final concentrations of 10 uM (the same volume of vehicle, DMSO, was used for control explants). The explants were incubated for 24 hours at 37º C in a CO2 incubator. After this period, explants were washed in PBS and collected with Accumax. Cell suspensions were, then, washed and processed using the OMNI-ATAC-Seq protocol[68].

Epiblast ATAC-seq data analysis

Initial processing of ATAC-seq libraries, including trimming, alignment, removal of duplicate reads, peak calling, and generation of bigwigs, was performed in the same manner as the tomoATAC-seq analysis. After initial processing, counting of reads at peaks and differential accessibility analysis was performed with DiffBind (version 3.4.11) using three biological

replicates.

530

CUT&RUN

As described above for the TomoATAC-seq experiments, the embryos processed for CUT&RUN were first collected and stage- and size-matched. For each experiment, position matched sections from 3 embryos were pooled together. H3K27AC CUT&RUN experiments
535 were performed with HH1, HH2, HH3, and HH5 embryos sectioned into 5, 5, 6 and 7 sections, respectively. Embryonic structures like Hensen's node and primitive streak were used as landmarks. The sectioning was performed in a way to preserve the amount of tissue per fragment. For the CTNNB1/LEF1 CUT&RUN, embryos were dissected into 3 anterior-posterior sections. The rostral segment was cut right after the regressing Hensen's node and the
540 remaining embryo was cut in half, dividing the middle and caudal segments. CUT&RUN experiments were conducted as previously described[74]. First, the tissue was dissociated into a single cell suspension using Accumax at room temperature for 20 minutes under mild agitation. Next, cells were bound to BioMag Plus Concavalin A magnetic beads (Bangs Laboratories, BP531) and incubated with rabbit anti H3K27AC (AbCam, #ab177178), or rabbit anti-Beta
545 catenin (Abcam, ab32572) plus anti-LEF1 (Abcam, #ab137872) antibodies overnight at 4° C, rocking. After washing away the unbound antibody, protein A-MNase was added to a final concentration of 700 ng/mL and incubated on the tube rotator for 1 hour at 4° C. Cells were chilled down to 0° C and CaCl2 was added to a final concentration of 2 mM to activate the MNase enzyme. MNase digestion was conducted for 45 minutes and terminated by adding the
550 2XSTOP buffer containing heterologous *Saccharomyces cerevisiae* spike-in DNA at a concentration of 2 pg/mL. The protein-DNA complexes were released by centrifugation and digested with proteinase K for 10 minutes at 70° C. DNA fragments were extracted using phenol-chloroform and ethanol precipitation.

555 *CUT&RUN data analysis*

Trimming of TruSeq adapter sequences was performed using cutadapt (version 3.4) using the following parameters: "-a AGATCGGAAGAGCACACGTCTGAACTCCAGTCA -A AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT --minimum-length=25". Trimmed reads were aligned to the galGal6 chicken reference genome using bowtie2 (version 2.4.4) with the
560 following arguments: "--local --very-sensitive-local --no-unal --no-mixed --no-discordant -I 10 -X 1000". Bam alignment files were indexed using samtools (version 1.10), followed by marking of duplicate reads using Picard's (version 2.26.2) MarkDuplicates function. Duplicate reads were then removed using samtools view with the following arguments: "-F 1804 -f 2". Peaks were called with macs2 (version 2.2.7.1) using the parameters: "-f BAMPE -g 1e9 -q 0.05 --call-
565 summits" and bigwigs were generated from Bam files using DeepTools' bamCoverage function (version 3.5.1) with the following arguments: "--binSize 5 --normalizeUsing RPKM --extendReads". Reads in peaks were counted using DiffBind (version 3.4.11). For plotting of chromatin activity profiles, raw counts were library size normalized and the moving average of the normalized counts were plotted along the AP axis. Motif enrichment analysis and
570 identification of the top enriched transcription factor binding motifs was performed using HOMER's findMotifsGenome.pl function. Plotting of LEF1/bCAT binding scores at anterior-/posterior-enriched peaks was accomplished using DeepTools' (version 3.5.1) computeMatrix

and plotHeatmap functions.

575 *Tissue isolation and 10X multiome library preparation*

The nuclei of three whole chick embryos at stage HH5 were isolated following the Chromium Next GEM Single Cell Multiome ATAC + Gene Expression (GEX) protocol. Briefly, the embryos were dissected and dissociated using Accumax for 15-20 minutes at room temperature to generate a single cell suspension. Next, the cell suspension was filtered using a Flowmi cell
580 strainer (Millipore-Sigma), porosity 40uM to remove cell debris and clumps. To isolate the nuclei, cells were centrifuged at 500 rcf for 5 minutes at 4° C and then lysed using a 0.1X lysis buffer for 5 minutes on ice. The isolated nuclei were washed and resuspended in Diluted nuclei buffer. Finally, the nuclei concentration was determined using an automated cell counter and 15,000 nuclei were used to target a 10,000 nuclei recovery. The ATAC and gene expression
585 library preparation and sequencing were performed at Cornell University BRC Genomics Core Facility according to the manufacturer's instructions (10X genomics).

10X multiome data analysis

Initial processing and alignment of paired scRNA-seq and scATAC-seq datasets from the 10X
590 multiome kit was performed using the cellranger-arc pipeline (version 2.0.0)[75]. Upon generation of the UMI count matrix via cellranger-arc's count function, downstream processing was performed using Seurat (version 4.3.0)[76] and Signac (version 1.8.0)[77]. In addition to cellranger's initial cell quality filtering, cells were additionally filtered in Seurat using the following criteria: "nCount_ATAC<8e4 & nCount_ATAC>4e3 & nCount_RNA<5000 & nCount_RNA>300 &
595 percent.mt<10". In total, 7750 cells passed cell quality filtering. RNA- and ATAC-seq datasets underwent independent preprocessing and dimensionality reduction using the standard Seurat and Signac pipelines, respectively. The two datasets were then integrated by calculating a weighted-nearest neighbor (WNN) graph, which was used for final UMAP visualization and clustering. Clusters were empirically assigned using marker genes identified through gene
600 expression and motif enrichment across UMAP clusters. Ordering of cells along the AP axis was accomplished using a quadratic programming approach developed by Yu and colleagues[32]. Extraembryonic clusters were excluded from the cell ordering analysis, as they were not predicted to be restricted along the AP axis. Cells were ranked according to their accessibility at anterior-/posterior- enriched peaks derived from TradeSeq analysis of the bulk tomoATAC-seq
605 datasets. Cell ranks were assigned using custom R scripts from Yu and colleagues[32]. Peak-gene linkage analysis was performed by computing the correlation between gene expression and accessibility at nearby peaks using Signac's LinkPeaks function. Analysis of individual germ layers was performed by subsetting the Seurat object to include only the clusters of interest, followed by repeating dimensionality reduction and WNN integration on each germ layer-specific
610 data subset.

References

1 Wolpert, L. Positional information and the spatial pattern of cellular differentiation. *J Theor Biol* **25**, 1-47, doi:10.1016/s0022-5193(69)80016-0 (1969).

2 Tkacik, G. & Gregor, T. The many bits of positional information. *Development* **148**, doi:10.1242/dev.176065 (2021).

3 Wolpert, L. Positional Information and Pattern Formation. *Curr Top Dev Biol* **117**, 597-608, doi:10.1016/bs.ctdb.2015.11.008 (2016).

4 Wolpert, L. One hundred years of positional information. *Trends Genet* **12**, 359-364, doi:10.1016/s0168-9525(96)80019-9 (1996).

5 Turing, A. M. The chemical basis of morphogenesis. 1953. *Bull Math Biol* **52**, 153-197; discussion 119-152, doi:10.1007/BF02459572 (1990).

6 Crick, F. Diffusion in embryogenesis. *Nature* **225**, 420-422, doi:10.1038/225420a0 (1970).

7 Driever, W. & Nusslein-Volhard, C. A gradient of bicoid protein in Drosophila embryos. *Cell* **54**, 83-93, doi:10.1016/0092-8674(88)90182-1 (1988).

8 Ashe, H. L. & Briscoe, J. The interpretation of morphogen gradients. *Development* **133**, 385-394, doi:10.1242/dev.02238 (2006).

9 Rogers, K. W. & Schier, A. F. Morphogen gradients: from generation to interpretation. *Annu Rev Cell Dev Biol* **27**, 377-407, doi:10.1146/annurev-cellbio-092910-154148 (2011).

10 Goldstein, B. & Freeman, G. Axis specification in animal development. *Bioessays* **19**, 105-116, doi:10.1002/bies.950190205 (1997).

11 Driever, W. & Nusslein-Volhard, C. The bicoid protein determines position in the Drosophila embryo in a concentration-dependent manner. *Cell* **54**, 95-104, doi:10.1016/0092-8674(88)90183-3 (1988).

12 Jaeger, J. & Martinez-Arias, A. Getting the measure of positional information. *PLoS Biol* **7**, e81, doi:10.1371/journal.pbio.1000081 (2009).

13 Slack, J. M. & Tannahill, D. Mechanism of anteroposterior axis specification in vertebrates. Lessons from the amphibians. *Development* **114**, 285-302, doi:10.1242/dev.114.2.285 (1992).

14 Hikasa, H. & Sokol, S. Y. Wnt signaling in vertebrate axis specification. *Cold Spring Harb Perspect Biol* **5**, a007955, doi:10.1101/cshperspect.a007955 (2013).

15 Yamaguchi, T. P. Heads or tails: Wnts and anterior-posterior patterning. *Curr Biol* **11**, R713-724, doi:10.1016/s0960-9822(01)00417-1 (2001).

16 Kerszberg, M. & Wolpert, L. Specifying positional information in the embryo: looking beyond morphogens. *Cell* **130**, 205-209, doi:10.1016/j.cell.2007.06.038 (2007).

17 Junker, J. P. *et al.* Genome-wide RNA Tomography in the zebrafish embryo. *Cell* **159**, 662-675, doi:10.1016/j.cell.2014.09.038 (2014).

18 Hamburger, V. & Hamilton, H. L. A series of normal stages in the development of the chick embryo. *Journal of morphology* **88**, 49-92 (1951).

19 Simeone, A. & Acampora, D. The role of Otx2 in organizing the anterior patterning in mouse. *Int J Dev Biol* **45**, 337-345 (2001).

20 Lagutin, O. V. *et al.* Six3 repression of Wnt signaling in the anterior neuroectoderm is essential for vertebrate forebrain development. *Genes Dev* **17**, 368-379, doi:10.1101/gad.1059403 (2003).

21 Sakai, Y. *et al.* The retinoic acid-inactivating enzyme CYP26 is essential for establishing an uneven distribution of retinoic acid along the anterio-posterior axis within the mouse embryo. *Genes Dev* **15**, 213-225, doi:10.1101/gad.851501 (2001).

22 Pilon, N. *et al.* Cdx4 is a direct target of the canonical Wnt pathway. *Dev Biol* **289**, 55-63, doi:10.1016/j.ydbio.2005.10.005 (2006).
23 Andre, P., Song, H., Kim, W., Kispert, A. & Yang, Y. Wnt5a and Wnt11 regulate mammalian anterior-posterior axis elongation. *Development* **142**, 1516-1527, doi:10.1242/dev.119065 (2015).
24 Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**, 1213-1218, doi:10.1038/nmeth.2688 (2013).
25 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589, doi:10.1016/j.molcel.2010.05.004 (2010).
26 Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat Methods* **14**, 975-978, doi:10.1038/nmeth.4401 (2017).
27 Petersen, C. P. & Reddien, P. W. Wnt signaling and the polarity of the primary body axis. *Cell* **139**, 1056-1068, doi:10.1016/j.cell.2009.11.035 (2009).
28 Ladewig, J. *et al.* Small molecules enable highly efficient neuronal conversion of human fibroblasts. *Nat Methods* **9**, 575-578, doi:10.1038/nmeth.1972 (2012).
29 Huang, S. M. *et al.* Tankyrase inhibition stabilizes axin and antagonizes Wnt signalling. *Nature* **461**, 614-620, doi:10.1038/nature08356 (2009).
30 Montefiori, L. E. *et al.* Enhancer Hijacking Drives Oncogenic BCL11B Expression in Lineage-Ambiguous Stem Cell Leukemia. *Cancer Discov* **11**, 2846-2867, doi:10.1158/2159-8290.CD-21-0145 (2021).
31 Granja, J. M. *et al.* 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 (2021).
32 Yu, Q. *et al.* Charting human development using a multi-endodermal organ atlas and organoid models. *Cell* **184**, 3281-3298 e3222, doi:10.1016/j.cell.2021.04.028 (2021).
33 Wapinski, O. L. *et al.* Rapid Chromatin Switch in the Direct Reprogramming of Fibroblasts to Neurons. *Cell Rep* **20**, 3236-3247, doi:10.1016/j.celrep.2017.09.011 (2017).
34 Zhao, S., Nichols, J., Smith, A. G. & Li, M. SoxB transcription factors specify neuroectodermal lineage choice in ES cells. *Mol Cell Neurosci* **27**, 332-342, doi:10.1016/j.mcn.2004.08.002 (2004).
35 Kanai-Azuma, M. *et al.* Depletion of definitive gut endoderm in Sox17-null mutant mice. *Development* **129**, 2367-2379, doi:10.1242/dev.129.10.2367 (2002).
36 Marcellini, S., Technau, U., Smith, J. C. & Lemaire, P. Evolution of Brachyury proteins: identification of a novel regulatory domain conserved within Bilateria. *Dev Biol* **260**, 352-361, doi:10.1016/s0012-1606(03)00244-6 (2003).
37 Lumsden, A. & Krumlauf, R. Patterning the vertebrate neuraxis. *Science* **274**, 1109-1115, doi:10.1126/science.274.5290.1109 (1996).
38 Millet, S. *et al.* A role for Gbx2 in repression of Otx2 and positioning the mid/hindbrain organizer. *Nature* **401**, 161-164, doi:10.1038/43664 (1999).
39 Irving, C. & Mason, I. Signalling by FGF8 from the isthmus patterns anterior hindbrain and establishes the anterior limit of Hox gene expression. *Development* **127**, 177-186, doi:10.1242/dev.127.1.177 (2000).
40 Argelaguet, R. *et al.* Multi-omics profiling of mouse gastrulation at single-cell resolution. *Nature* **576**, 487-491, doi:10.1038/s41586-019-1825-8 (2019).
41 Hinrichs, A. S. *et al.* The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res* **34**, D590-598, doi:10.1093/nar/gkj144 (2006).

42 Peng, G. *et al.* Spatial Transcriptome for the Molecular Annotation of Lineage Fates and Cell Identity in Mid-gastrula Mouse Embryo. *Dev Cell* **55**, 802-804, doi:10.1016/j.devcel.2020.11.018 (2020).
43 Stern, C. D. *Gastrulation : from cells to embryo*. (Cold Spring Harbor Laboratory Press, 2004).
44 Nusslein-Volhard, C. Determination of the embryonic axes of Drosophila. *Dev Suppl* **1**, 1-10 (1991).
45 Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nat Rev Genet* **20**, 207-220, doi:10.1038/s41576-018-0089-8 (2019).
46 Prieto, E. I. & Maeshima, K. Dynamic chromatin organization in the cell. *Essays Biochem* **63**, 133-145, doi:10.1042/EBC20180054 (2019).
47 Rossi, A. M., Jafari, S. & Desplan, C. Integrated Patterning Programs During Drosophila Development Generate the Diversity of Neurons and Control Their Mature Properties. *Annu Rev Neurosci* **44**, 153-172, doi:10.1146/annurev-neuro-102120-014813 (2021).
48 Niehrs, C. On growth and form: a Cartesian coordinate system of Wnt and BMP signaling specifies bilaterian body axes. *Development* **137**, 845-857, doi:10.1242/dev.039651 (2010).
49 Lewis, J. From signals to patterns: space, time, and mathematics in developmental biology. *Science* **322**, 399-403, doi:10.1126/science.1166154 (2008).
50 Green, J. B. & Sharpe, J. Positional information and reaction-diffusion: two big ideas in developmental biology combine. *Development* **142**, 1203-1211, doi:10.1242/dev.114991 (2015).
51 Azambuja, A. P. & Simoes-Costa, M. The connectome of neural crest enhancers reveals regulatory features of signaling systems. *Dev Cell* **56**, 1268-1282 e1266, doi:10.1016/j.devcel.2021.03.024 (2021).
52 Ramos, A. I. & Barolo, S. Low-affinity transcription factor binding sites shape morphogen responses and enhancer evolution. *Philos Trans R Soc Lond B Biol Sci* **368**, 20130018, doi:10.1098/rstb.2013.0018 (2013).
53 Barkai, N. & Shilo, B. Z. Robust generation and decoding of morphogen gradients. *Cold Spring Harb Perspect Biol* **1**, a001990, doi:10.1101/cshperspect.a001990 (2009).
54 Lo, W. C., Zhou, S., Wan, F. Y., Lander, A. D. & Nie, Q. Robust and precise morphogen-mediated patterning: trade-offs, constraints and mechanisms. *J R Soc Interface* **12**, 20141041, doi:10.1098/rsif.2014.1041 (2015).
55 Stern, C. D. Initial patterning of the central nervous system: how many organizers? *Nat Rev Neurosci* **2**, 92-98, doi:10.1038/35053563 (2001).
56 Metzis, V. *et al.* Nervous System Regionalization Entails Axial Allocation before Neural Differentiation. *Cell* **175**, 1105-1118 e1117, doi:10.1016/j.cell.2018.09.040 (2018).
57 Stern, C. D. *et al.* Head-tail patterning of the vertebrate embryo: one, two or many unresolved problems? *Int J Dev Biol* **50**, 3-15, doi:10.1387/ijdb.052095cs (2006).
58 Nieuwkoop, P. D. Activation and organization of the central nervous system in amphibians. Part III. Synthesis of a new working hypothesis. *J. Exp. Zool.*, 83-108 (1952).
59 Mangold, O. Über die Induktionsfähighkeit der verschiedenen Bezirke der Neurula von Urodelen. *Naturwissenshaften*, 761-766 (1933).
60 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).
61 Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, doi:10.1093/gigascience/giab008 (2021).
62 Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169, doi:10.1093/bioinformatics/btu638 (2015).

63 Van den Berge, K. *et al.* Trajectory-based differential expression analysis for single-cell sequencing data. *Nat Commun* **11**, 1201, doi:10.1038/s41467-020-14766-3 (2020).
64 Mi, H. *et al.* Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). *Nat Protoc* **14**, 703-721, doi:10.1038/s41596-019-0128-8 (2019).
65 Thomas, P. D. *et al.* PANTHER: Making genome-scale phylogenetics accessible to all. *Protein Sci* **31**, 8-22, doi:10.1002/pro.4218 (2022).
66 Wilkinson, D. G. & Nieto, M. A. Detection of messenger RNA by in situ hybridization to tissue sections and whole mounts. *Methods Enzymol* **225**, 361-373, doi:10.1016/0076-6879(93)25025-w (1993).
67 Choi, H. M. T. *et al.* Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. *Development* **145**, doi:10.1242/dev.165753 (2018).
68 Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods* **14**, 959-962, doi:10.1038/nmeth.4396 (2017).
69 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal* **17**, 10-12 (2011).
70 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).
71 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).
72 Ramirez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160-165, doi:10.1093/nar/gkw257 (2016).
73 Ross-Innes, C. S. *et al.* Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389-393, doi:10.1038/nature10730 (2012).
74 Rothstein, M. & Simoes-Costa, M. Heterodimerization of TFAP2 pioneer factors drives epigenomic remodeling during neural crest specification. *Genome Res* **30**, 35-48, doi:10.1101/gr.249680.119 (2020).
75 Zheng, G. X. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat Commun* **8**, 14049, doi:10.1038/ncomms14049 (2017).
76 Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e3529, doi:10.1016/j.cell.2021.04.048 (2021).
77 Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nat Methods* **18**, 1333-1341, doi:10.1038/s41592-021-01282-5 (2021).

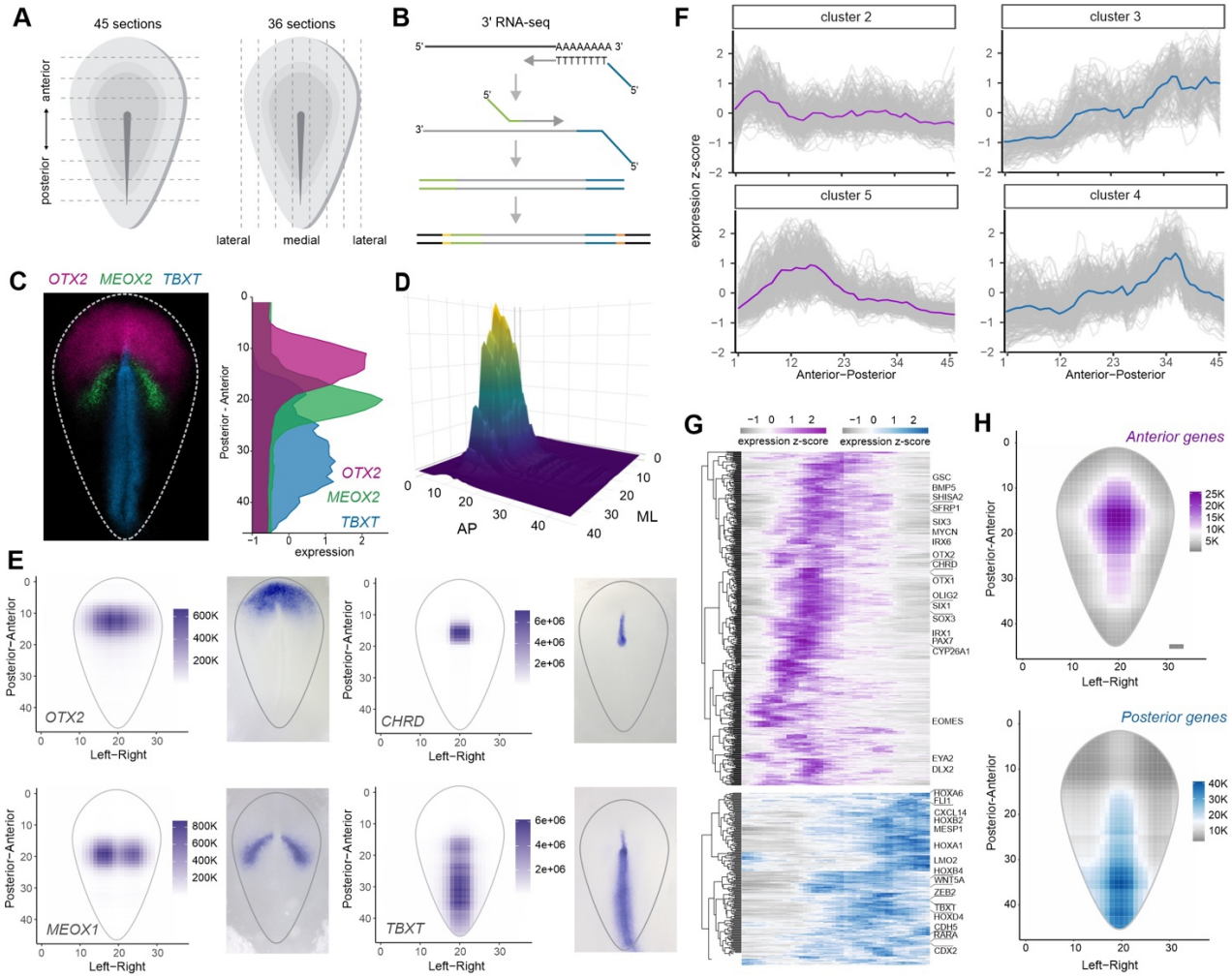

Fig. 1. Reconstruction of gene expression patterns in the avian gastrula. (**A**) Spatial transcriptomic analysis was performed via cryosectioning of gastrula stage (HH5) embryos along the anteroposterior (AP) and mediolateral axes (ML). (**B**) Individual tissue sections were subsequently processed for 3' RNA-seq. (**C**) Multiplexed HCR *in situ* hybridization showing the expression of *OTX2, MEOX2* and *TBXT* in the HH5 chicken gastrula. Density plots to the right display normalized gene expression profiles along the AP axis derived from analysis of the tomo-seq datasets. (**D**) Integration of AP and ML transcriptomic datasets allows for quantification of gene expression in two dimensions across the avian gastrula. The two-dimensional expression of *OTX2* is shown for reference (**E**) *In silico* (left) reconstruction of the *OTX2*, *CHRD*, *MEOX1*, and *TBXT* 2D gene expression patterns, along with the corresponding *in situ* hybridization images (right). *In silico* projections recapitulate the gene expression profiles as seen via *in situ* hybridization. (**F**) Hierarchical clustering analysis of AP expression profiles identifies anterior and posterior enriched gene clusters. Colored lines represent the mean of all profiles shown. (**G**) Heatmap showing 706 anterior- and 294 posterior-enriched genes expressed in the avian gastrula. (**H**) Spatial averages of anterior and posterior genes in the gastrula. AP: anteroposterior, ML: mediolateral

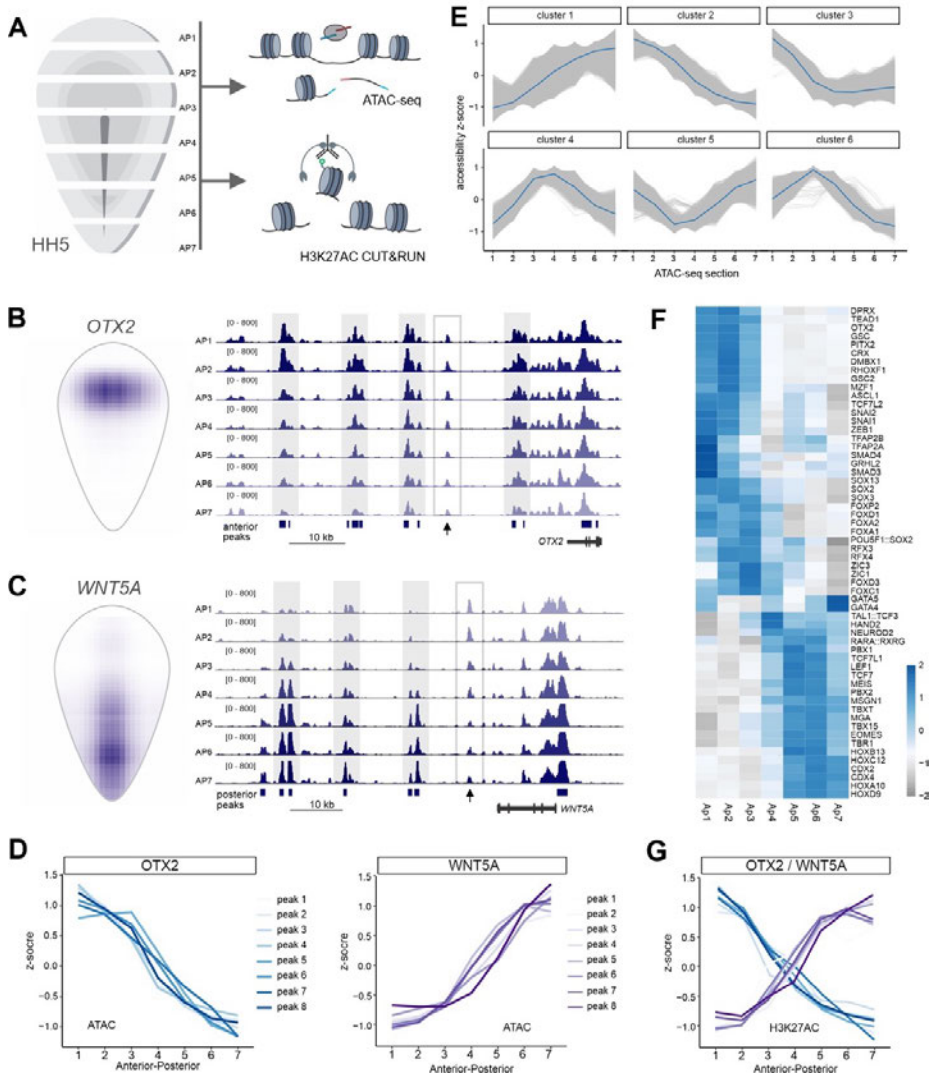

Fig. 2. Identification of gradients of chromatin activity in positional cis-regulatory elements (pCREs). (**A**) HH5 avian embryos were divided into seven sections along the AP axis and subjected to ATAC-seq and CUT&RUN for H3K27AC. (**B**) *In silico* reconstruction of *OTX2* gene expression alongside ATAC-seq profiles in the *OTX2* locus. Note the gradients of accessibility that are formed from the anterior to the posterior regions of the embryo (shaded regions). Arrow highlights a region of chromatin accessibility that is not patterned. (**C**) *In silico* reconstruction of *WNT5A* gene expression alongside ATAC-seq profiles in the *WNT5A* locus. Accessibility profiles display a posterior to anterior gradient (shaded regions). Arrow highlights an element with anterior-enriched accessibility. (**D**) Quantification of pCRE accessibility in 7 ATAC-seq datasets along the AP axis. Accessibility profiles are shown for eight peaks in the *OTX2* locus(Fig. S4), as well as eight peaks in the *WNT5A* locus (Fig. S4). (**E**) Hierarchical clustering of chromatin accessibility profiles along the AP axis. Blue lines represent the mean of all profiles shown. (**F**) Heatmap displaying variable motifs enriched along the AP axis as identified via Chromvar. (**G**) Quantification of H3K27AC signal in eight peaks in the *OTX2* locus, as well as eight peaks in the *WNT5A* locus. Peaks correspond to those shown in panel D. HH: Hamburger Hamilton

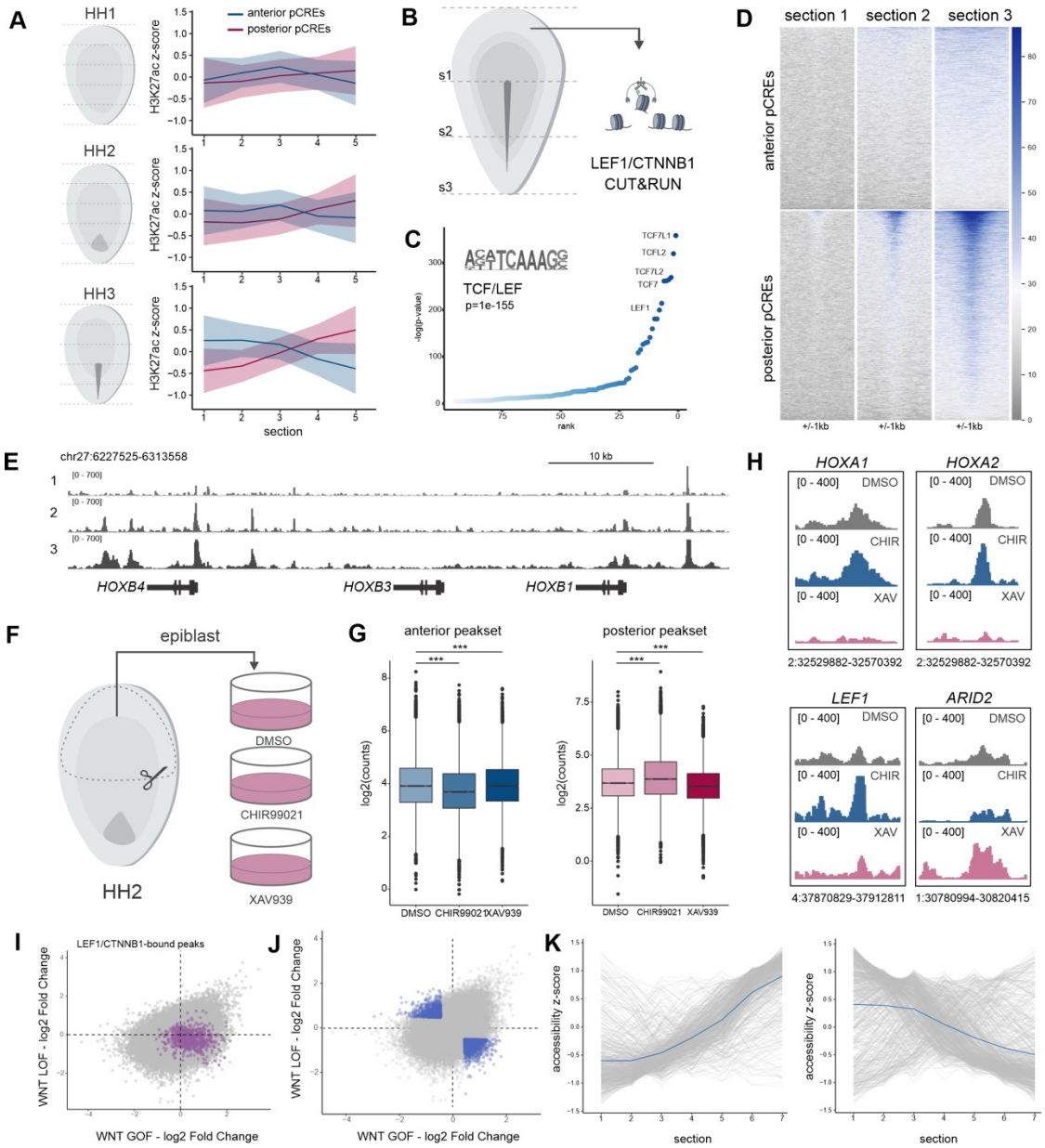

Fig. 3. Regulation of pCRE gradients by signaling systems. (**A**) Profiles of H3K27AC signal at anterior and posterior pCREs in 5 sections along the AP axis at three developmental stages in the avian embryo (HH1, HH2, and HH3). Lines represent the average H3K27AC signal, while ribbons represent standard deviation. AP gradients of H3K27AC activity are formed at the onset of gastrulation (HH3). (**B**) Strategy for genome-wide mapping of nuclear effectors of canonical WNT signaling, LEF1/CTNNB1, with CUT&RUN in anterior, medial, and posterior regions of the gastrula. (**C**) Motif enrichment analysis of LEF1/CTNNB1 CUT&RUN peaks via HOMER. Top enriched motifs include the WNT nuclear effectors TCF3, TCFL2, TCF4, TCF7, and LEF1. (**D**) Tornado plots displaying LEF1/CTNNB1 signal at anterior/posterior pCREs. Binding of WNT nuclear effectors is enriched in posterior pCREs in an anterior to posterior gradient. (**E**) Genome browser tracks showing abundant binding of WNT nuclear effectors in the *HOX* cluster. (**F**) Strategy for manipulation of WNT signaling in naïve cells of the avian blastoderm. (**G**) Box plots showing distribution of ATAC-seq signal in anterior/posterior pCREs upon WNT manipulation. (**H**) Examples of pCREs that are affected by manipulation of WNT signaling. Peaks in the

HOXA1, *HOXA2*, and *LEF1* loci are positively regulated by WNT signaling, while a peak in the *ARID2* locus is negatively regulated by WNT signaling. (**I**) Scatter plot of all accessible peaks across ATAC-seq datasets. Log2 fold change in accessibility upon WNT gain-of-function or WNT loss-of-function is shown on the x and y axes, respectively. LEF1/CTNNB1-associated peaks (highlighted) display increased accessibility under CHIR conditions and loss of ATAC signal following XAV treatment. (**J**) Scatter plot as shown in (I), with highly WNT-responsive peaks highlighted. (**K**) Profiles of chromatin accessibility of WNT-responsive peaks highlighted in (**J**) from the HH5 tomoATAC-seq dataset. Blue line represents the mean of all profiles shown. WNT-responsive elements display gradients in accessibility along the AP axis. HH: Hamburger Hamilton, ***P≤0.001

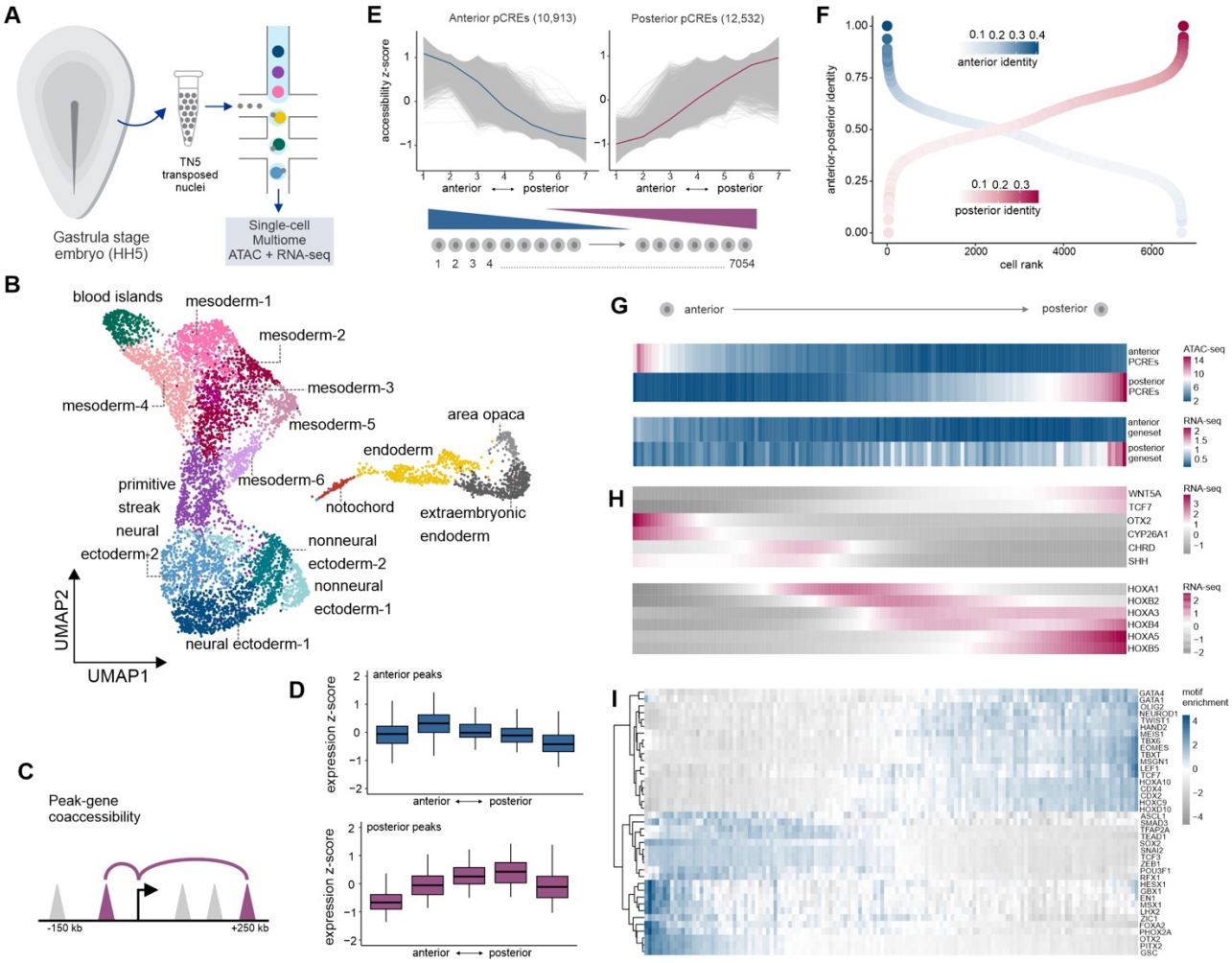

Fig. 4. pCREs predict the axial position of embryonic cells. (**A)** Single-cell multiomic analysis (RNA-seq & ATAC-seq) of the HH5 chicken embryo via 10X Genomics. **(B)** UMAP projection of 7750 cells in the avian gastrula, with sixteen distinct clusters labeled by cellular identity. (**C**) Assignment of pCREs to putative target genes using co-accessibility analysis. (**D**) Boxplot displaying aggregate AP gene expression profiles via tomo-seq of genes displaying strong co-accessibility with anterior and posterior pCREs. (**E**) Strategy to assign cellular position of cells from the multiome dataset based on anterior and posterior pCRE scores. (**F**) Dotplot displaying fractional identity of individual cells along the AP axis. Cells are colored by similarity to the anterior/posterior pCRE datasets and are ranked from anterior to posterior. (**G**) Heatmaps displaying accessibility of the anterior and posterior pCRE peaksets (Table S3), as well as average expression of anterior/posterior gene sets (Table S2) in 50-cell bins ordered along the AP axis. Aggregate scores of anterior and posterior genes are consistent with cell order, demonstrating pCRE score is sufficient to predict cell position. (**H**) Projection of normalized expression of individual genes from the multiome RNA-seq in 50-cell binds ordered along the AP axis. (**I**) Chromvar variability of transcription factor motifs in 50-cell binds along the AP axis. HH: Hamburger Hamilton

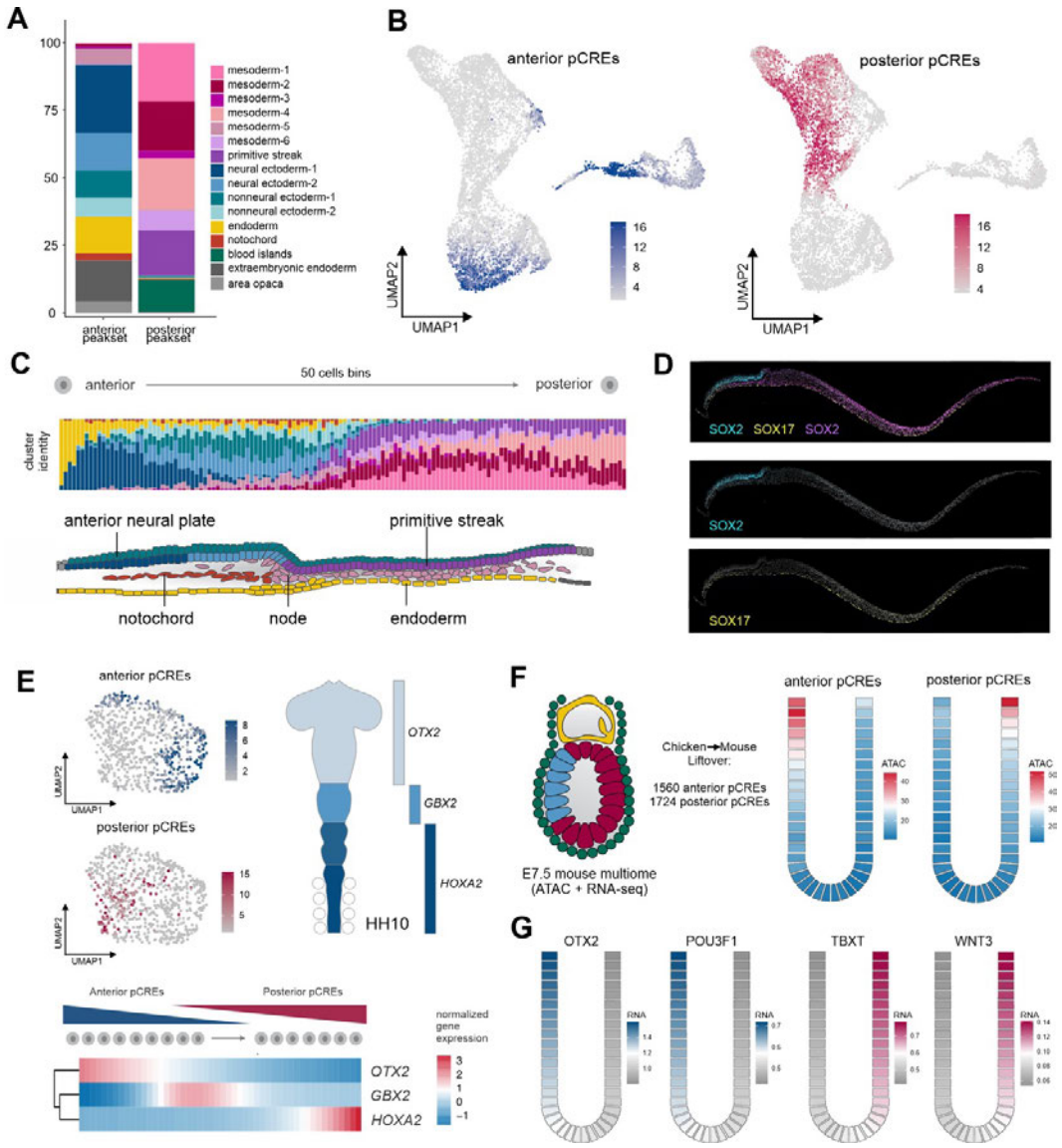

Fig. 5. pCREs contain positional information within individual germ layers. (**A**) Stacked bar plot displaying cellular composition of groups of cells with high anterior vs posterior pCRE scores from the multiome dataset. (**B**) Projection of pCRE scores in the multiome UMAP. Aggregate accessibility of anterior pCREs is the strongest in ectodermal and endodermal cells, whereas posterior pCREs are on average more accessible in the mesodermal. (**C**) Cellular composition of 50-cell bins ordered from anterior to posterior using quadratic programming. Cellular composition along the AP axis shows an uneven distribution of germ layers in the avian gastrula. (**D**) Immunohistochemistry for germ layer markers SOX2, SOX17, and TBXT show ectoderm and endoderm predominantly occupy anterior positions, while most of the posterior embryo is composed of mesoderm. (**E**) Projection of anterior and posterior pCRE signatures in 708 neuroectodermal cells isolated from the multiome dataset. Ordering of neuroectodermal cells with quadratic programming allows for the reconstruction of expected expression patterns for axial markers of the central nervous system, like *OTX2*, *GBX2,* and *HOXA2*. (**F**) 1560 anterior pCREs and 1724 posterior pCREs are conserved between chicken and mouse.

Conserved pCREs were used to order cells from the E7.5 mouse multiome dataset from Argelaguet et al. as 200-cell bins from anterior to posterior. (**G**) Heatmaps displaying expression of *OTX2*, *POU3F1*, *TBXT*, and *WNT3* in 200-cell bins ordered from anterior to posterior using pCRE accessibility. Conserved pCREs can be used for cell ordering and prediction of expression patterns of axial genes in the cup-shaped murine gastrula.