Article

TRANSPARENT PROCESS OPEN ACCESS

THE EMBO JOURNAL

Combinatorial transcription factor activities on open chromatin induce embryonic heterogeneity in vertebrates

Ann Rose Bright[1], Siebe van Genesen[1], Qingqing Li[2], Alexia Grasso[1], Siebren Frölich[1], Maarten van der Sande[1], Simon J van Heeringen[1] & Gert Jan C Veenstra[1,*]

Abstract

During vertebrate gastrulation, mesoderm is induced in pluripotent cells, concomitant with dorsal-ventral patterning and establishing of the dorsal axis. We applied single-cell chromatin accessibility and transcriptome analyses to explore the emergence of cellular heterogeneity during gastrulation in *Xenopus tropicalis*. Transcriptionally inactive lineage-restricted genes exhibit relatively open chromatin in animal caps, whereas chromatin accessibility in dorsal marginal zone cells more closely reflects transcriptional activity. We characterized single-cell trajectories and identified head and trunk organizer cell clusters in early gastrulae. By integrating chromatin accessibility and transcriptome data, we inferred the activity of transcription factors in single-cell clusters and tested the activity of organizer-expressed transcription factors in animal caps, alone or in combination. The expression profile induced by a combination of Foxb1 and Eomes most closely resembles that observed in the head organizer. Genes induced by Eomes, Otx2, or the Irx3-Otx2 combination are enriched for maternally regulated H3K4me3 modifications, whereas Lhx8-induced genes are marked more frequently by zygotically controlled H3K4me3. Taken together, our results show that transcription factors cooperate in a combinatorial fashion in generally open chromatin to orchestrate zygotic gene expression.

Keywords cell trajectories; chromatin accessibility; mesendoderm; organizer; transcription factors

Subject Categories Chromatin, Transcription & Genomics; Development

DOI 10.15252/embj.2020104913 | Received 5 March 2020 | Revised 21 December 2020 | Accepted 24 December 2020 | Published online 8 February 2021

The EMBO Journal (2021) 40: e104913

Introduction

Cellular heterogeneity increases dramatically during early embryonic development in association with regional specification of the ectoderm, mesoderm, and endoderm lineages. During this process, cells respond to extracellular signals as dictated by cell-autonomous constraints such as chromatin state and the presence of factors mediating the response. The embryo is transcriptionally quiescent until zygotic genome activation (ZGA; Paranjpe & Veenstra, 2015; Vastenhouw *et al*, 2019), which gradually occurs during the mid-blastula stage in *Xenopus*. This is accompanied by slowing down of cell divisions and introduction of cell cycle gap phases. During the blastula stage, the cells at the animal pole are pluripotent. They are fated to become ectoderm but are competent to respond to mesoderm-inducing signals emanating from vegetal pole cells. During early gastrulation, there is a major cellular diversification, concomitant with germ layer formation and morphogenesis.

Within the animal cap at early gastrula stages, ectoderm primarily consists of a superficial (epithelial) layer, and an attached deep (sensorial) layer (Chalmers *et al*, 2002), which diverge to goblet cells, multiciliated cells, ionocytes, and small secretary cells in the epidermal ectoderm by larval stages (Angerilli *et al*, 2018). In addition, the animal cap cells are competent to form mesoderm, but can also be induced to neural ectoderm by underlying mesoderm during gastrulation. Mesoderm, located at the equatorial region (marginal zone) during early gastrula stages, is induced by nodal-related TGFβ family ligands, produced by vegetal cells. Dorsal mesoderm develops into a signaling center, the Spemann-Mangold Organizer, involved in dorsal-ventral patterning of mesoderm and formation of the dorsal axis during gastrulation (Agius *et al*, 2000). Using transplantation experiments, it has been established that the dorsal blastopore lip of early gastrula stages induces anterior dorsal structures (head organizer), whereas the blastopore lip at late gastrula stages specifies posterior structures (trunk organizer). Little is known how these activities relate to each other, but it has been shown that inhibition of Wnt signaling is important for the formation of head organizer structures such as anterior endoderm, prechordal plate, and anterior chordamesoderm (Niehrs, 1999). The dorsal blastopore lip is also home to superficial mesoderm. Under the influence of Wnt11b and FGF signaling, as well as the Foxj1

1 Department of Molecular Developmental Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Radboud University, Nijmegen, The Netherlands
2 Biomedical Pioneering Innovation Center, Peking University, Beijing, China
*Corresponding author. Tel: +31 24 3610541; E-mail: g.veenstra@science.ru.nl

transcription factor, involuting superficial mesoderm cells will constitute the ciliated gastrocoel roof that forms the left-right organizer (Walentek *et al*, 2013; Schneider *et al*, 2019).

During early development, secreted factors signal to the nucleus of exposed cells, impacting gene expression in conjunction with transcription factors. Chromatin plays a major role in this process, providing the gene-specific permissive or restrictive context for transcription (Perino & Veenstra, 2016; Jambhekar *et al*, 2019). This involves histone modifications such as the permissive promoter mark H3K4me3 and the repressive Polycomb mark H3K27me3, both of which increase dramatically during early *Xenopus* development (Bogdanović *et al*, 2012; Paranjpe & Veenstra, 2015; Hontelez *et al*, 2015). A large majority of genomic loci is decorated with these histone modifications by maternal factors; these loci are collectively referred to as the maternal regulatory space, whereas a relatively small number of promoters requires new embryonic transcription for the acquisition of H3K4me3 or H3K27me3 (Hontelez *et al*, 2015).

Very little is known how cellular heterogeneity and the cellular responses to extracellular signaling relate to chromatin accessibility and how this is regulated regionally during early development. Here we report on the regulation of chromatin accessibility during early development and in early gastrula animal cap and dorsal marginal zone using ATAC-sequencing. The regional differences in gene expression and chromatin accessibility were related to cellular heterogeneity observed in whole embryo single-cell RNA-sequencing (scRNA-seq) and scRNA-seq data of dissected animal cap and dorsal marginal zone tissue. We inferred transcription factor activities in specific cell clusters, which were tested in animal caps by microinjection and RNA-sequencing. We assessed the extent to which organizer-expressed transcription factors activate gene expression within open chromatin, and how this relates to H3K4me3 promoter marking by maternal factors. The data support the early emergence of head and trunk organizer as well as superficial mesoderm cells. Moreover, the data show how early cellular heterogeneity emerges in response to inductive events by action of zygotic transcription factors in the context of maternally marked, accessible promoters.

Results

Dynamics of chromatin accessibility during early development

To define the chromatin regulatory landscape during early development, we performed ATAC-seq of biological replicates for blastula, early and late gastrula, and neurula stages (respectively stage 9, 10½, 12, and 16; Fig 1A). We clustered the open chromatin ATAC-seq peaks (Dataset EV1) together with H3K4me3 and p300 ChIP-seq data (Hontelez *et al*, 2015) of the same developmental stages. Open chromatin as observed by ATAC-seq is found in regions with H3K4me3 (promoters), the co-activator p300 (enhancers) or both (promoter-proximal regulatory elements; Fig 1A). H3K4me3-decorated promoter regions displayed higher levels of chromatin accessibility than enhancer elements that recruit p300 (Fig EV1A). Generally, ATAC-seq enrichment increased from stage 9 onwards, both for H3K4me3-positive promoters and p300-bound enhancers (Fig EV1A). Before stage 9, we have not been able to obtain high-quality ATAC-seq tracks because of a lack of enrichment, which may suggest that regulatory elements become only accessible after the mid-blastula transition.

Pair-wise comparison of sequential stages yielded over 7000 differentially accessible regions (see Materials and Methods; Fig 1B). Clustering of these regions with p300 and H3K4me3 data revealed that the majority of these regions with dynamic open chromatin are enriched for p300 but not H3K4me3 (Fig EV1B). Moreover, the ATAC-seq and p300 signal intensities correlate for these dynamic open chromatin regions, suggesting they represent developmental stage-specific accessible enhancers. Clusters 2, 4, 5, and 6 showed accessibility signals increasing from stage 9 to stage 16, whereas clusters 1 and 3 consisted of regions showing a reduction in the signal after gastrulation. To assess the extent to which the open chromatin dynamics are linked to gene regulation, we analyzed the transcript levels of nearby genes. To associate each of these genomic regions to genes, we used GREAT regions (McLean *et al*, 2010). We found for each of the sequential-stage comparisons that genomic elements with increased chromatin accessibility are associated with an increased expression of the associated genes (Fig 1C; left side of each panel; Owens *et al*, 2016). Surprisingly, genes associated with genomic elements with decreased accessibility are also upregulated in many cases. It should be noted that many of these genes are regulated by multiple enhancers, sometimes with different dynamics. In addition, transcript stability may cause transcript dynamics to lag behind chromatin accessibility dynamics. Indeed, clusters with strongly increasing chromatin accessibility over multiple time points (clusters 2, 5, 6; Fig EV1B) show higher expression toward later time points (Fig EV1C), whereas clusters with reduced accessibility showed peak expression at early gastrulation (clusters 1, 3; Fig EV1B and C). The results raise the question how changes in chromatin accessibility relate to regional specification, germ layer formation, and the heterogeneity in gene expression programs associated with the onset of gastrulation.

Mesoderm-induced genes exhibit open chromatin in animal caps

To determine the extent to which chromatin accessibility is spatially acquired in early gastrula stage embryos (stage 10½), we first preformed ATAC-seq on ectodermal and organizer explants (Fig 2A; respectively, animal cap, AC; dorsal marginal zone, DMZ). Ectoderm-expressed genes such as *tfap2a* and *grhl3* showed high accessibility at regulatory regions in the AC and relatively low signals in DMZ (Fig 2A). By contrast, the organizer-expressed genes *gsc* and *chrd* were equally accessible in AC and DMZ explants. To assess how general these observations are, we performed differential gene expression analysis (fold change ≥ 2 and FDR < 0.05) of AC and DMZ RNA-seq samples (Blitz *et al*, 2016). Differential genes were then linked to ATAC-seq peaks (Materials and Methods) to see how well spatial expression differences match with differences in chromatin accessibility (Fig 2B). We observed that genes with higher expression in AC compared to DMZ, also exhibit higher chromatin accessibility in AC. However, for genes with higher expression in DMZ, the associated regulatory regions showed a similar accessibility signal in both explants, similar to what was observed at *gsc* and *chrd*. AC cells are considered pluripotent at the blastula stages and lose competence for mesoderm induction during gastrulation (Jones & Woodland, 1987; Borchers & Pieler, 2010). Consistent with the

competence for mesoderm induction, mesoderm-expressed genes appear to exhibit accessible chromatin in AC.

To further characterize regional chromatin accessibility at promoters and enhancers, we divided AC and DMZ peaks into p300 and H3K4me3-positive regions. Hierarchical clustering showed somewhat lower ATAC signals in DMZ for both H3K4me3 and p300 positive regions (Fig 2C), but the significance of this result is not clear. Most of the p300-associated regulatory elements of

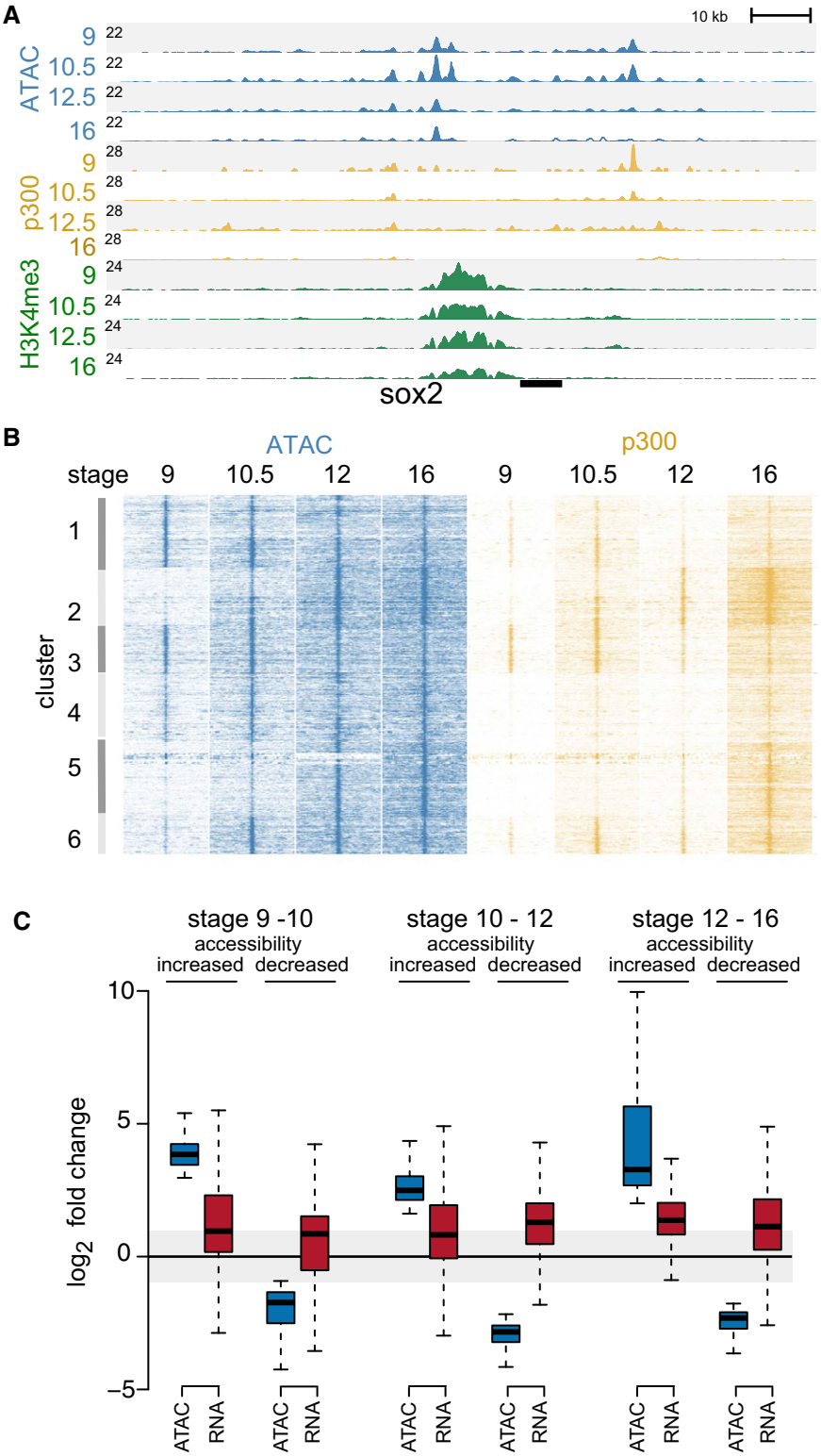

Figure 1.

Figure 1. Chromatin accessibility during early development.

A Genome browser view showing chromatin accessibility (ATAC-seq) and ChIP-seq (p300 and H3K4me3) profiles at stage 9, 10½, 12, and 16 at *sox2* gene locus. The number at the left (black line) indicates the Y-axis scale of the profile. ATAC-seq peaks were found at promoter (H3K4me3) and enhancer (p300-bound) regulatory regions.

B Chromatin accessibility and p300 binding at differential ATAC-seq peaks visualized using K-means clustering.

C Boxplots showing pair-wise sequential-stage comparisons of fold change in accessibility (ATAC-seq) and corresponding changes in gene expression (RNA-seq data set; Owens *et al*, 2016). The data represents two biological replicates. The central band within the boxplot represents the median (50th percentile), the box represents the range between the first and third quartile (25th–75th percentile), and the whiskers show 1.5× the interquartile range (IQR).

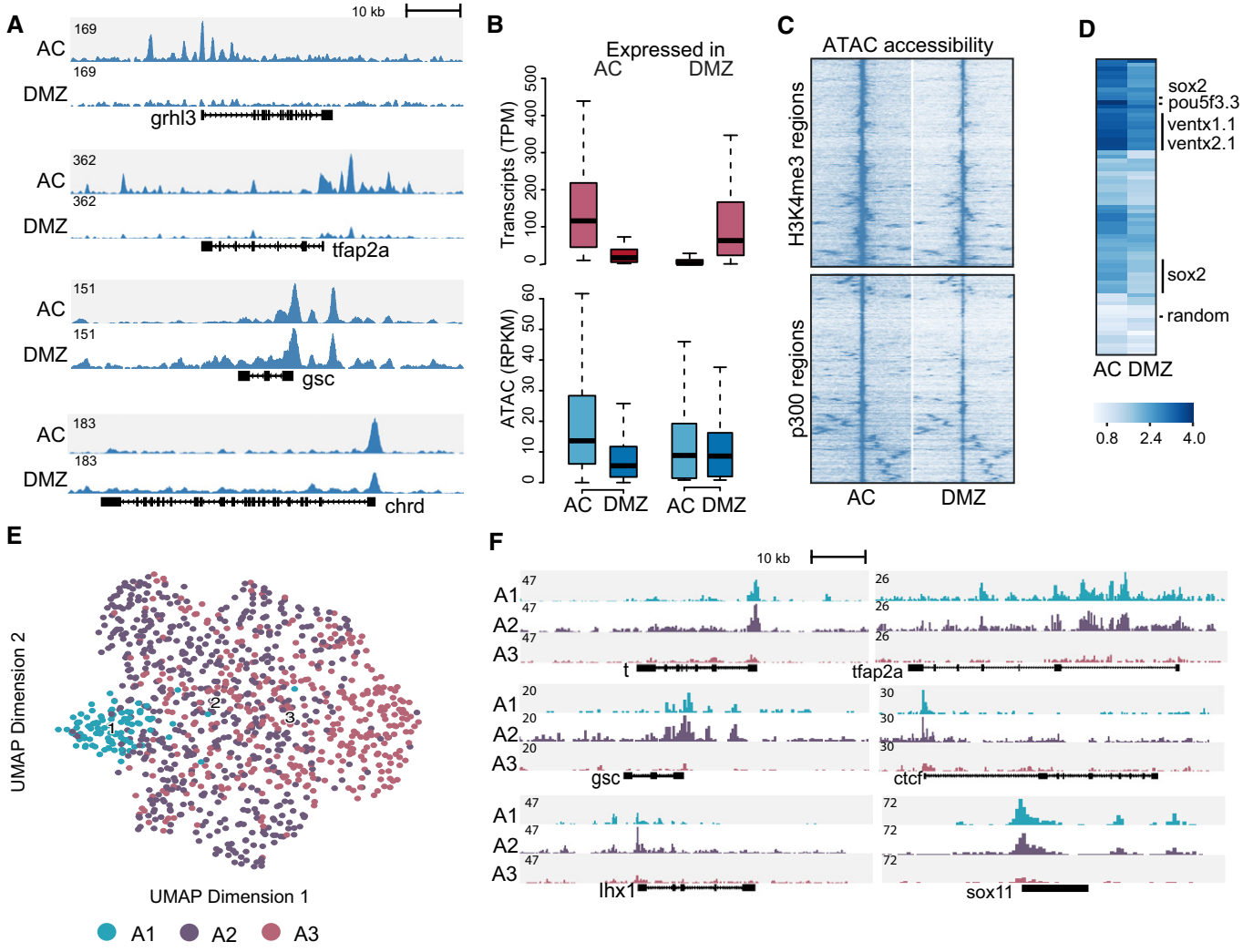

Figure 2. Chromatin accessibility in animal cap (AC) and dorsal marginal zone (DMZ).

A Genome browser view of AC and DMZ accessibility profiles for ectoderm-expressed (*tfap2a* and *grhl3*) and organizer-expressed (*gsc* and *chrd*) marker genes.

B Boxplots showing differential gene expression (AC versus DMZ) and associated ATAC-seq signals (Two biological replicates). The central band within the boxplot represents the median and the whiskers show 1.5 times the range between the first and third quartile (IQR).

C Hierarchical clustering of AC and DMZ ATAC-seq data on H3K4me3-positive (top) and p300-positive (bottom) ATAC-seq peaks.

D Heatmap showing accessibility signal (log1p of fold over background) at p300-positive ATAC-seq peaks surrounding pluripotency genes (*ventx1/ventx2, pou5f3* and *sox2*). The row labeled "random" shows accessibility signals at random genomic loci.

E Single-cell ATAC-seq UMAP projection of cells derived from gastrula stage embryos (stage 10½), colored by cluster.

F Genomic tracks showing aggregated accessibility of single-cell ATAC-seq clusters at the *t* (*tbxt*), *tfap2a*, *gsc*, *ctcf*, *lhx1* and *sox11* loci.

pluripotency genes (*pou5f3.3, pou5f3.1, sox2,* and the *Nanog*-related genes *ventx2.1,* and *ventx1.1* (Scerbo *et al*, 2012) were accessible in both AC and DMZ, although to variable degrees, whereas random genomic regions were not accessible in either tissue (Fig 2D). The transcription start sites (TSS) of these pluripotency genes also showed relatively strong chromatin accessibility in AC compared to DMZ (Fig EV2A). Overall, these observations confirmed that regulatory regions in AC cells have relatively open chromatin, irrespective of the transcriptional activity of the associated genes, in line with the pluripotent nature of the AC and its competence for mesoderm induction.

As a next step, we used Chromium single-cell ATAC-seq to uncover differences in open chromatin with single-cell resolution. We isolated nuclei from stage 10½ embryos, transposed them in bulk, then encapsulated them in barcoded gel beads, each barcode representing a unique cell. After library preparation, sequencing and filtering, we recovered 1,072 barcodes (cells), with a median 5.56× enrichment of transposition events at transcription starts sites (TSSs) and a median of 7,053 transposition events per cell. Dimensionality reduction and clustering uncovered three closely associated clusters with only marginal differences in locus-specific enrichment patterns (Fig 2E and F). Cluster A1 likely corresponds to AC cells, with elevated levels of accessibility at 1,935 genes (FDR < 0.05, $\log_2$ fold change > 1; Dataset EV2), including genes like *sox11* and *tfap2a* (Fig 2F). Clusters A2 and A3 are less distinct, with, respectively, only 34 and 30 genes with cluster-enriched accessibility (Dataset EV2). Cluster A3 consists of cells with relatively few transposition events compared to clusters A1 and A2. Cluster A1 almost exclusively shows transposition events at promoters, whereas A2 and A3 show more modest promoter enrichment (Figs EV2B and C, and 2F). The covariates associated with these cell populations appear to rise to prominence in the absence of large intercellular differences in chromatin accessibility within the embryo. Together, these results highlight the relatively subtle regional differences in the context of broadly accessible chromatin in the early gastrula embryo.

Single-cell analysis of spatiotemporal trajectories of ectodermal and mesendodermal cell states

Our results indicated that chromatin accessibility differs in relatively subtle ways in the early embryo for some genes, notably those expressed at the animal pole, but not for many others. During gastrulation, cellular heterogeneity is known to increase dramatically at the transcriptional level. For example, single-cell profiles sampled from blastula to tailbud stage embryos (stages 8–22) has documented the emergence of an increasing number of cell states during early development (Briggs *et al*, 2018). We analyzed stage 8, 10, and 12 whole embryo single-cell data as a first step to assess the emerging heterogeneity during gastrulation and the associated developmental cell trajectories. We filtered, normalized, and visualized the data with UMAP, using available stage and cell type annotations (Figs 3A and EV3A). We called Louvain clusters (cell clusters L0-L20) and hypervariable genes (differential between cell clusters) (Dataset EV3). We labeled cell clusters based on predominant cell annotations in these clusters (Fig EV3A and B). The cells annotated as stage 10 neural ectoderm express both *sox2* and *tfap2a*, whereas *tfap2a* is not expressed in neural ectoderm at later stages (Appendix Fig S1). This suggests that these cells are closely related to non-neural ectoderm at early gastrula stages, consistent with Louvain clusters of cells with mixed neural and non-neural annotations at this stage (clusters L3, L16, L18, L19, Fig EV3A and B). To assist the interpretation, we color-coded clusters based on similarity in cell type annotations and gene expression (Figs 3BandEV3B, Appendix Fig S1). Stage 8 blastomeres are relatively homogeneous, represented by a single cluster (L4) that is most related to clusters comprised of non-neural/neural ectoderm at stage 10 (L3, L8, L18; Figs 3BandEV3B). From these basal stage 10 clusters, there is a continuous trajectory to stage 10 clusters with neural ectoderm annotation (L5), mixed neural ectoderm and marginal zone annotations (L10), marginal zone (L12), and organizer mesendoderm (L2). Cells in stage 10 cluster L10, with mixed marginal zone and neural ectoderm annotations, express both *t (tbxt)* and *sox2* in the same cells, but low levels of *tfap2a*, in line with a potential bi-potent neuro-mesodermal cell state (Fig EV3B, Appendix Fig S1). Endoderm cells are rather sparse in this data set. An ectodermal trajectory, however, can be observed from basal ectoderm clusters to stage 10 non-neural ectoderm. Clusters comprised mainly of stage 12 cells are more located to the periphery relative to the stage 10 clusters, with a distinct stage 12 neural ectoderm cluster that is juxtaposed to stage 10 ectoderm (L18) and neural ectoderm (L5). Stage 12 involuted dorsal mesoderm (L9) is juxtaposed to stage 10 organizer (L2) and marginal zone (L12), as well as a stage 12 cluster with tailbud cell annotations (L12; Figs 3B and EV3B).

These whole embryo single-cell profiles lack spatial information, although the localization of cell clusters can be tentatively inferred from highly expressed genes. To define the early events associated with the specification of mesodermal and epidermal lineage during gastrulation, and to link single-cell transcriptomic profiles to spatially localized gene regulation, we performed single-cell RNA-sequencing (scRNA-seq) on dissected animal cap (AC) and dorsal marginal zone (DMZ) explants. We analyzed hand-picked cells from dissociated stage 10½ AC and DMZ explants, collected from two different experiments (Materials and Methods; Dataset EV4). We obtained seven single-cell clusters (C0-C6), with the cells of the same region of the embryo generally clustering together (Fig 4A and B). Based on known marker gene expression of *sox11* (ectoderm inner layer) and *grhl3*, *krt*, and *upk3b* (ectoderm outer layer; Chalmers *et al*, 2006), we tentatively assigned AC clusters C0 and C1 to these ectodermal layers (Figs 4C and D, Appendix Fig S2A and B, Dataset EV5). Both clusters express relatively high levels of the pluripotency factor-encoding transcripts *pou5f3.3* and *sox2*. These clusters also express *ventx1.1* and *ventx1.2*, the closest amphibian homologs of the mammalian Nanog protein (Scerbo *et al*, 2012), which are abundant in the ventral and animal cap regions (Appendix Figs S2 and S3). Clusters C2, C3, and C4 all express high levels of mesendodermal markers such as *t* (*tbxt*), *vegt*, and *mix1*. C2 shows the highest levels of *wnt11b* and early expression of *foxj1* (Walentek *et al*, 2013), which mark, respectively, involuting mesoderm and superficial mesoderm, the epithelial layer of involuting mesoderm. C3 expresses the highest levels of well-known organizer genes such as *gsc*, *otx2* and *chrd*, in addition to endodermal markers such as *gata4*, *sox17a*, and *sox17b* (Fig 4C, Appendix Figs S2 and S3). In addition, C3 cells express the head organizer genes *cer1*, *dkk1*, *frzb*, and *fst*, suggesting these cells comprise the precursors of anterior endoderm, prechordal plate mesendoderm and anterior chordamesoderm. C4 expresses

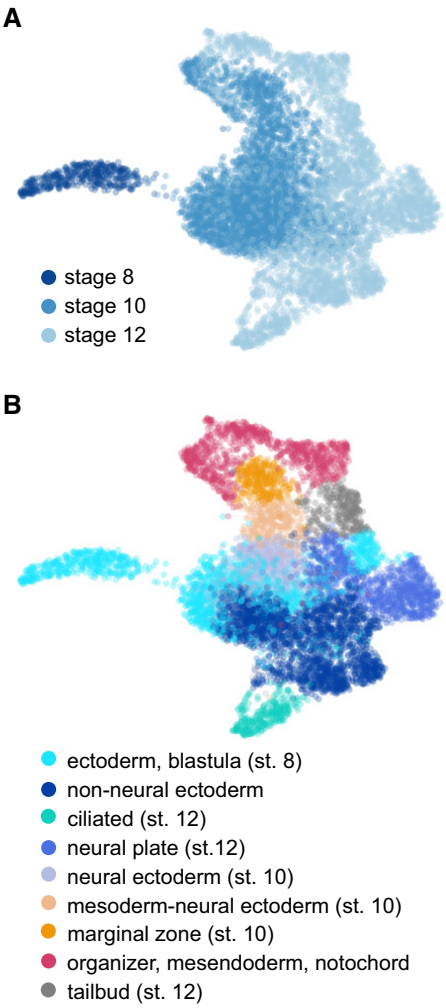

Figure 3. Cellular heterogeneity and developmental trajectories in blastula and gastrula stages.

A, B UMAP visualization of whole embryo scRNA-seq for stages 8, 10, and 12 colored by stage (A) and cell type annotations (B).

also organizer genes such as *gsc*, *otx2*, and *chrd*, but compared to C3, C4 cells express lower levels of head organizer genes and relatively high levels of *cdx4*, *t* (*tbxt*) and *irx3* (Figs 4C and D, Appendix Figs S2 and S3, Fig EV4). This cluster likely constitutes non-involuted mesoderm and prospective trunk organizer cells. C5 is derived from the AC and shows both inner layer characteristics (*sox11*) and expression of *foxj1* and *klf5* (Appendix Fig S2); therefore, most likely constituting progenitors of ciliated cells. C6 exhibits a mosaic of ectodermal (*sox11, sox2*) and mesodermal expression (*t, chrd, eomes, vegt*) that is present at the level of individual single cells (Fig 4D, Appendix Fig S2B). Genes specifically expressed in this cluster, such as *fgf8* and *foxb1*, are expressed in the upper blastopore lip (Appendix Fig S3), in the inner layer of dorsal ectoderm and in non-involuted mesoderm (Gamse & Sive, 2001; Takebayashi-Suzuki *et al*, 2011; Gentsch *et al*, 2013; Pera *et al*, 2014).

To place our spatially localized AC and DMZ single-cell clusters in the whole embryo stage 8–12 data set, we determined the correlations of the AC and DMZ clusters (C0–C6) with all clusters in the whole embryo data set (L0–L20). As expected, AC cluster C0, C1 and C5 correlated mostly with cells annotated as stage 10 (neural/non-neural) ectoderm and stage 8 blastomeres (Figs 4E and EV4). Involuting mesoderm (C2) and head organizer mesendoderm (C3) correlated best with cells annotated as organizer (L2), with C3 matching L2 best. Organizer mesoderm cluster C4 correlated with both organizer and marginal zone (L12, L2), whereas upper blastopore lip cluster C6 correlated only weakly with stage 10 organizer (L2). Overall, these combined analyses chart trajectories of the cellular heterogeneity that arises during gastrulation in the AC and DMZ. Interestingly, even though the head and trunk organizer historically have been defined based on their location at the dorsal blastopore lip at early and late gastrula stages, respectively, the early gastrula is home to (cellular precursors of) both. The data also highlight the progressive nature of germ layer specification, with partially overlapping cell states across developmental stages.

Integration of single-cell transcriptome clusters and chromatin accessibility

We next sought to identify the drivers of cellular heterogeneity using motif analysis of regulatory elements active in the early gastrula. Peak calling of chromatin accessibility did not reveal subsets of genomic regions that were specifically accessible in a subset of cells, there are only subtle differences in relative peak strength. We therefore used motif activity analysis, a regression approach that has been shown to robustly identify the contribution of individual motifs to differential chromatin accessibility or gene expression (Suzuki *et al*, 2009; Balwierz *et al*, 2014; Madsen *et al*, 2018). A positive motif activity indicates a positive correlation between presence of the motif and associated accessibility signal, while a negative motif activity means the motif is associated with lower accessibility. As a first step, we determined the motifs that were associated with differences in chromatin accessibility between explants (AC, DMZ) and whole embryos (stage10½). In order to select the motifs with highest differential motif activity we ranked the motifs by the maximum z-score difference (Fig 5A). We plotted expressed transcription factors capable of binding to these motifs. Interestingly, we observed various ectodermal and mesodermal factors to display motif activity patterns that correlated with their known gene expression profile. Chromatin accessibility-associated motifs in AC included factors involved in ectodermal and epidermal development (Tafp2a, Tfap2c, Grhl1; Luo *et al*, 2005; Tao *et al*, 2005). In addition, motifs for Klf factors (Klf2, Klf5) were identified, which are highly expressed in the animal cap cells of *Xenopus* and are involved in pluripotency and self-renewal in mammalian cells (Gao *et al*, 2015). For DMZ, the motifs included those bound by Eomes, T/Tbxt, Otx2 and Forkhead factors such as Foxb1 and Foxc1, which are known to play roles in mesendoderm specification and axial mesoderm (Steiner *et al*, 2006; Charney *et al*, 2017). Factors like Grhl1, Eomes, and T/Tbxt showed overlapping motif activity in AC/DMZ with the whole embryo, but GATA factor-motif activity was enriched in whole embryo relative to both AC and DMZ (Fig 5A). Likewise, single-cell ATAC cluster A1 shows relatively high motif activities for Sox2 and Tfap2a, whereas A2 shows relatively high motif activities for CTCF and T/Tbxt, and Forkhead motifs show up in both A1 and A2 clusters (Fig EV5A). A3 shows

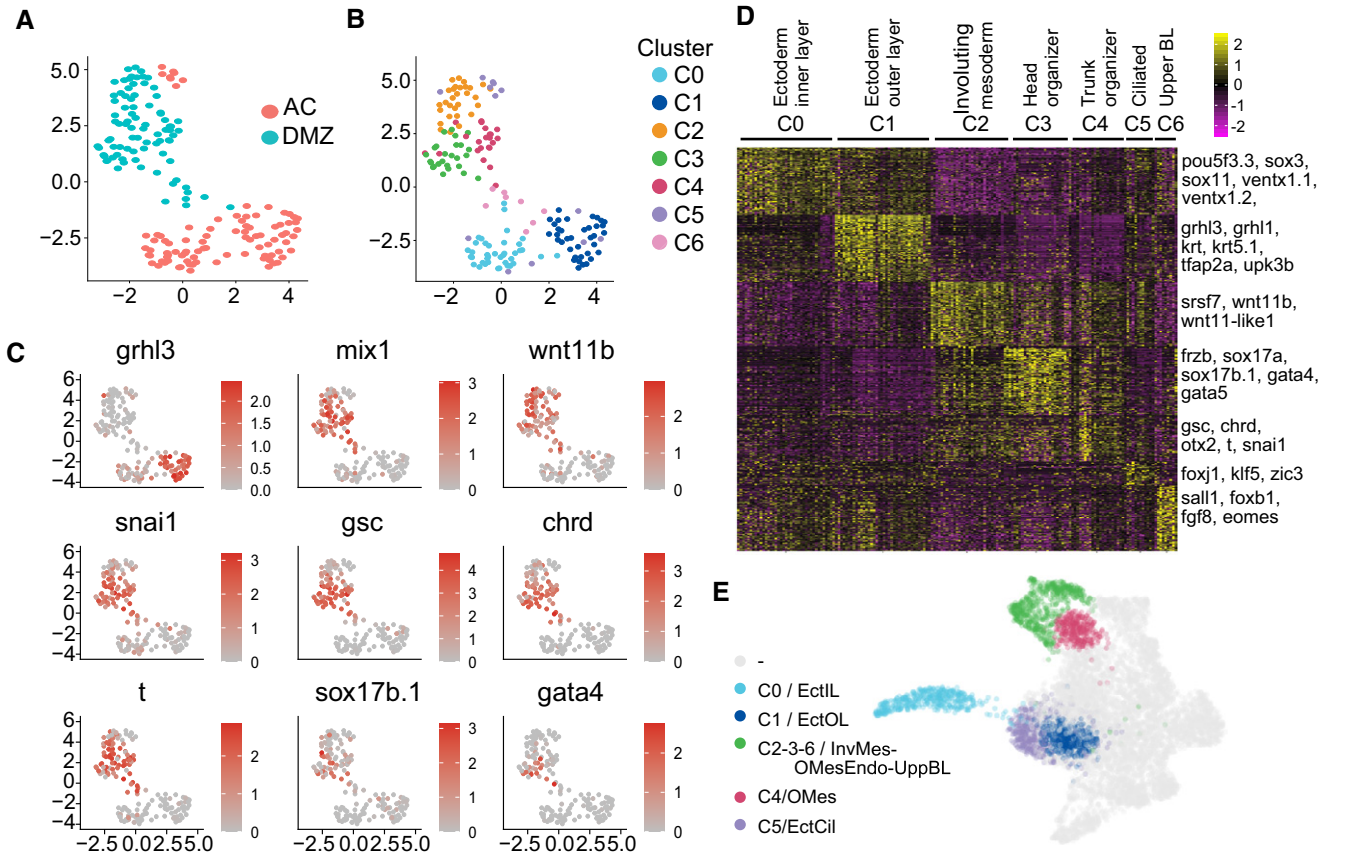

Figure 4. scRNA-seq of hand-picked cells from stage 10½ dissected AC and DMZ explants.

A Dimensionality reduction using UMAP. Each dot represents a cell. Colors indicate AC and DMZ.

B As panel (A), colors indicate clusters (C0-C6).

C Feature maps, showing the expression of selected genes in single cells (cf. Appendix Fig S2). Color scale represents the gene expression value (log1p-transformed counts per 10,000 unique reads) for each cell for a given gene, from low (gray) to high (orange).

D Heatmap depicting top 100 hypervariable genes in each cluster (cf. Dataset EV5). Color scale represents scaled gene expression (z-score values). The z-score values are ranging from −2 to 2.

E Spatially restricted AC and DMZ cell clusters embedded in UMAP of whole embryo scRNA-seq data (cf. Fig 3).

activity for Crx and Gata3. These analyses show that the relatively subtle differences in chromatin accessibility that are observed in the embryo are correlated with specific motifs that can be bound by transcription factors with localized expression.

Next, we wondered whether we could extend this analysis by identifying potential regulators of cluster-specific gene expression. We analyzed the regulatory elements associated with hypervariable genes of the AC and DMZ scRNA-seq clusters, i.e., transcription factor motifs in regulatory elements near genes with cluster-enriched gene expression. Similar to the analysis of chromatin accessibility, we used a regression approach (Materials and Methods), now regressing motifs in open chromatin to the variance in gene expression between single-cell clusters. The motif activity in this context represents the relation between the motif in accessible regions and the variance in gene expression associated with these elements. This recovers many of the same motifs identified in the previous analyses, but now linking them to cluster-specific gene expression (Fig 5B). Some of these motifs can be bound by multiple factors, often do to similarities in their DNA-binding domain (for example, Forkhead, Sox, T-box factors, Iroquois factors, etc.). Therefore, we visualized the expression of transcription factors capable of binding specific motifs together with the cluster-specific motif activity, so as to infer which transcription factor of a family may be most relevant (Figs 5C and EV5B). This analysis is agnostic to whether DNA-binding proteins activate or repress transcription, as both positive and negative contributions to cluster gene expression are included. A positive motif activity in a particular cluster means that genes with the motif tend to be more abundantly expressed in that particular cluster. The single-cell resolution of the motif analysis comes from the data type to which the presence of motifs is regressed, in this case single-cell gene expression (clusters C0-6). To test whether it matters which ATAC peak set is used for the analysis, we ran the analysis with AC peaks, DMZ peaks and whole embryo stage 10½ peaks. The results show similar results with some differences (Figs 5C and EV5C and D).

The transcription factor-motif combinations uncovered in this way include many known well-known regulators, for example,

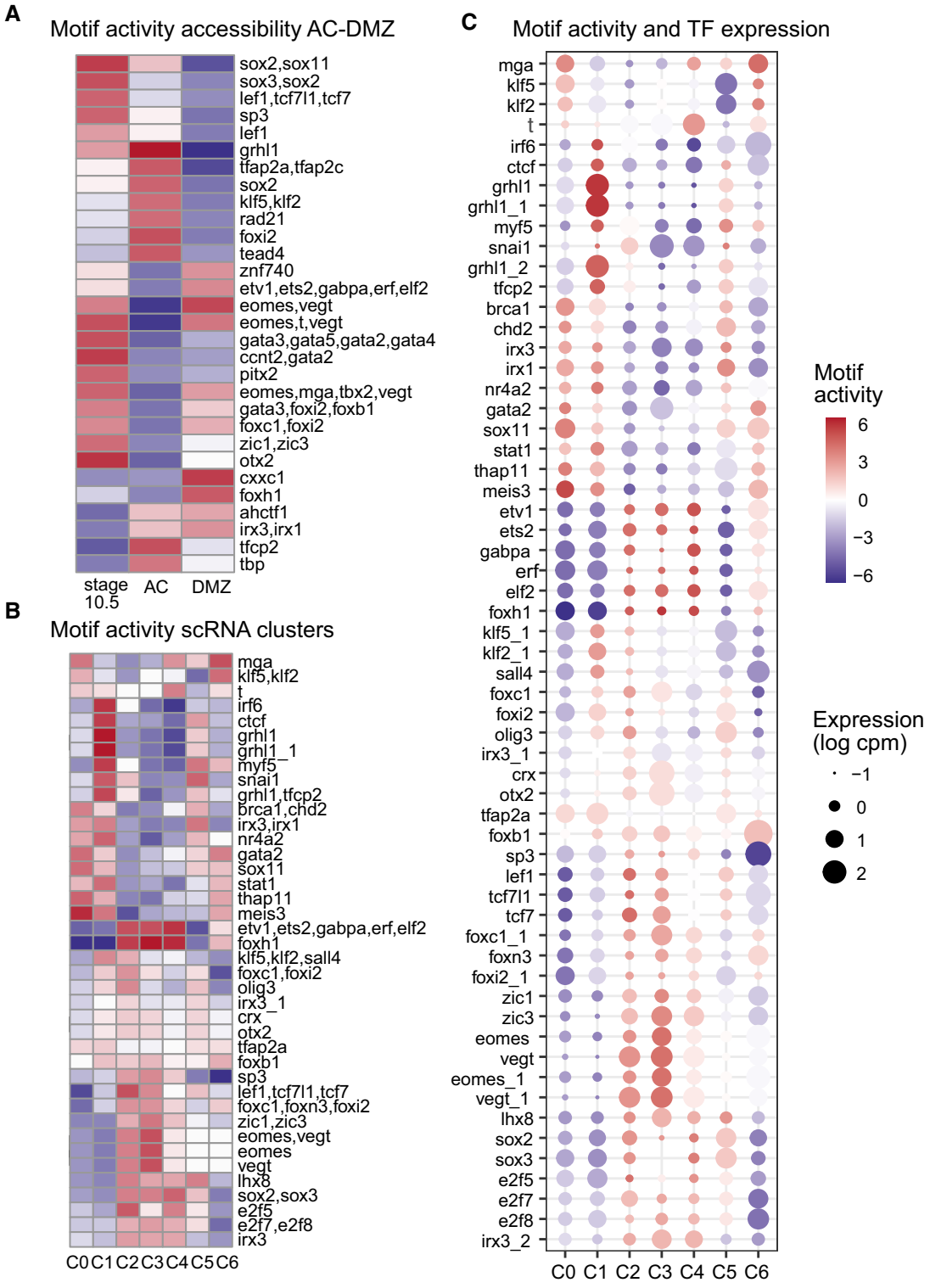

Figure 5. Integration of single-cell transcriptomics and chromatin accessibility, for identifying regulators driving cluster-specific gene expression.

A Heatmap showing motif activity inferred from differential chromatin accessibility of stage 10½-AC-DMZ.

B Heatmap showing motifs identified based on regulatory regions closest to cluster-specific genes, and regression to cluster gene expression.

C Heatmap of transcription factor (TF)-motif combinations showing cluster-specific motif activity (z-score, color) and gene expression (size of dot). Motifs and the Motif-TF combinations were hierarchically clustered.

Grhl1 (AC outer layer cluster C1), Vegt, T and Eomes (DMZ involuted mesoderm and head organizer clusters C2-3), and Zic1 and Zic3 (DMZ clusters C2-C3) (Figs 5C and EV5C and D). Interestingly, additional motif activity for the T-box factors is picked up with DMZ peaks in upper blastopore cluster C6, in which Eomes is more abundant compared to T/Tbxt and Vegt. Less well-known is Foxb1; both motif activity and cluster-enriched expression support a potential role in upper blastopore cluster C6, similar to Eomes. Several Iroquois motifs are found with opposing motif activities (line 15,16, bottom; Fig 5C). Irx3 is expressed in organizer clusters in C3 and C4, whereas Irx1 is more abundant in AC clusters C0 and C5. Lhx8 expression is relatively low at stage 10½ and but its motif activity was mostly restricted to C2 cells, with some activity in C3 and C4 where Lhx8 is also expressed. These results identify potential regulators of the gene regulatory network in early gastrula embryos. This raises the question how these transcription factors contribute to gene regulation, and whether some of these factors can act in a combinatorial fashion in promoting cluster-specific or regional gene expression.

Combinatorial action of transcription factors induces organizer gene expression in animal cap explants

To address the contributions of the transcription factors identified by integrated analysis of single-cell transcriptomes and chromatin accessibility, we selected candidates for functional characterization based on cluster-specific motif activity and gene expression in DMZ clusters. We selected Lhx8 (DMZ clusters C2-C4, resp. putative involuted mesoderm, head organizer, trunk organizer), Irx3 (C4), Otx2 (C4), Foxb1 (C6, upper blastopore), and Eomes (C6) to test their ability to induce DMZ gene expression in AC. We injected synthetic mRNA encoding these factors in the animal pole of one-cell stage embryos, cut animal caps at the blastula stage (stage 8), and collected the animal caps for RNA-sequencing when control embryos reached stage 10½. We also injected combinations of *foxb1* and *eomes*, as well as *irx3* and *otx2* mRNA. Differential gene expression analysis identified a total of around 600 genes (Fig 6A, Dataset EV6) that were differentially expressed (DE) in the overexpressing AC explants compared to water-injected AC explants (adjusted *P*-value < 0.05, fold change ≥ 2). Among individually overexpressed transcription factors, *eomes* and *otx2* overexpression had a larger effect on the transcriptome compared to *foxb1, irx3* or *lhx8* overexpression (Fig 6A, Dataset EV6). The combination of *foxb1* and *eomes* resulted in a marked difference in transcriptional response, with some genes activated more strongly and other genes less strongly compared to *eomes* alone. For example, whereas *fst* was robustly upregulated in *eomes* but not in *eomes-foxb1*-injected caps, *frzb* was induced only in *eomes-foxb1* caps. The combination of *irx3* and *otx2* resulted in a pattern of up- and downregulation that was similar to that caused by *otx2* alone, but a relatively small number of genes was upregulated more strongly when the two factors were combined. *Lhx8*-injected caps showed less profound activation of DMZ-expressed genes compared to *eomes* or *otx2* injections. Nonetheless, several of the most strongly induced genes in *foxb1-eomes* caps were also expressed in *lhx8* caps. In all these AC explants with overexpressed DMZ factors, genes with mesendodermal or organizer expression such as *t (tbxt), mixer, vegt, foxa1, nog, foxd4, chrd, gsc,* and *sox17* were activated, whereas the expression of AC genes like *tfap2a* and *foxi4.2* was reduced (Fig 6A, Dataset EV6).

To define the induced cell states, we compared the profiles of overexpression RNA-seq samples and the single-cell clusters. We performed a correlation analysis using the expression values of genes common between differentially expressed genes in the bulk data set, and hypervariable genes in the single-cell data set (Fig 6B). Eomes by itself induced transcriptomic changes that are related to the gene expression profiles of C3-C4 cells (organizer), and Foxb1-induced transcripts correlated with C4 (presumptive trunk organizer) and C6 (upper dorsal blastopore lip) expression. In combination, however, these factors increase the transcriptome similarity to C2-C3-C4 cells, whereas the correlation with C6 gene expression is similar to that caused by Foxb1 alone. Irx3 and Otx2 individually cause some similarity to, respectively, C3 and C4 cells, whereas in combination the correlation with both C3 and C4 cells is increased. Lhx8 overexpression caused expression of organizer and mesodermal genes (*gsc, chrd, t, mespb, wnt11-like*) and the induced transcriptome correlated most strongly with C4, with lower correlations with C3 and C2 cells. These data show that all these factors induce transcriptional changes related to mesodermal and organizer cells, and moreover that Foxb1-Eomes and Irx3-Otx2 cause qualitatively and quantitatively different effects when expressed in combination. Notably these effects differ for specific spatial gene expression programs, as Foxb1-expressing AC cells partially recapitulate the transcriptome observed in upper blastopore lip (C6) cells, but the resemblance to C6 expression is unaffected by Eomes co-expression.

Previously, we defined sets of promoters based on whether they required embryonic transcription for gaining the permissive promoter mark H3K4me3 (Hontelez *et al*, 2015). Maternally and zygotically defined (MaD, ZyD) H3K4me3 on promoters is related to DNA methylation; unmethylated CpG island promoters acquire H3K4me3 independent of embryonic transcription, revealing mechanistically different modes of transcriptional activation. We wondered to what extent the transcription factors tested in this study, could activate genes in zygotic regulatory space, alone or in combination. We tested if genes more than two-fold upregulated with these transcription factors were enriched for ZyD H3K4me3 genes. We found insignificant or no enrichment of ZyD genes among genes activated by individual transcription factors in AC, with the exception of Lhx8 (hypergeometric *P*-value 0.001; Fig 6C). Rather, ZyD H3K4me3 genes were depleted among genes which expression was increased in Eomes (*P*-value 0.006), Otx2 (*P*-value 0.0009), and Irx3 plus Otx2-injected (*P*-value 0.00004) animal caps. We also wondered if the Foxb1-Eomes and Irx3-Otx2 combinations could activate genes synergistically and identified just over 40 genes with more than two-fold cooperativity for each group (fold change ≥ 2 and fold change ≥ 2x the product of individual fold changes). *Frzb* and *chrd* were among the top genes induced synergistically by Foxb1-Eomes, whereas and *dmbx1 (otx3), pcdh8, sp5, nog* and *lhx8* were among the genes induced synergistically by Irx3-Otx2 (Dataset EV6). *Lhx1, chrd, gsc,* and 7 other genes were synergistically induced by both combinations of transcription factors. The two groups of synergistically activated genes included ZyD H3K4me3 genes; however, they were neither enriched nor depleted significantly in these groups. These results indicated that the Foxb1-Eomes and Irx3-Otx2 combinations had cooperative roles in regulating

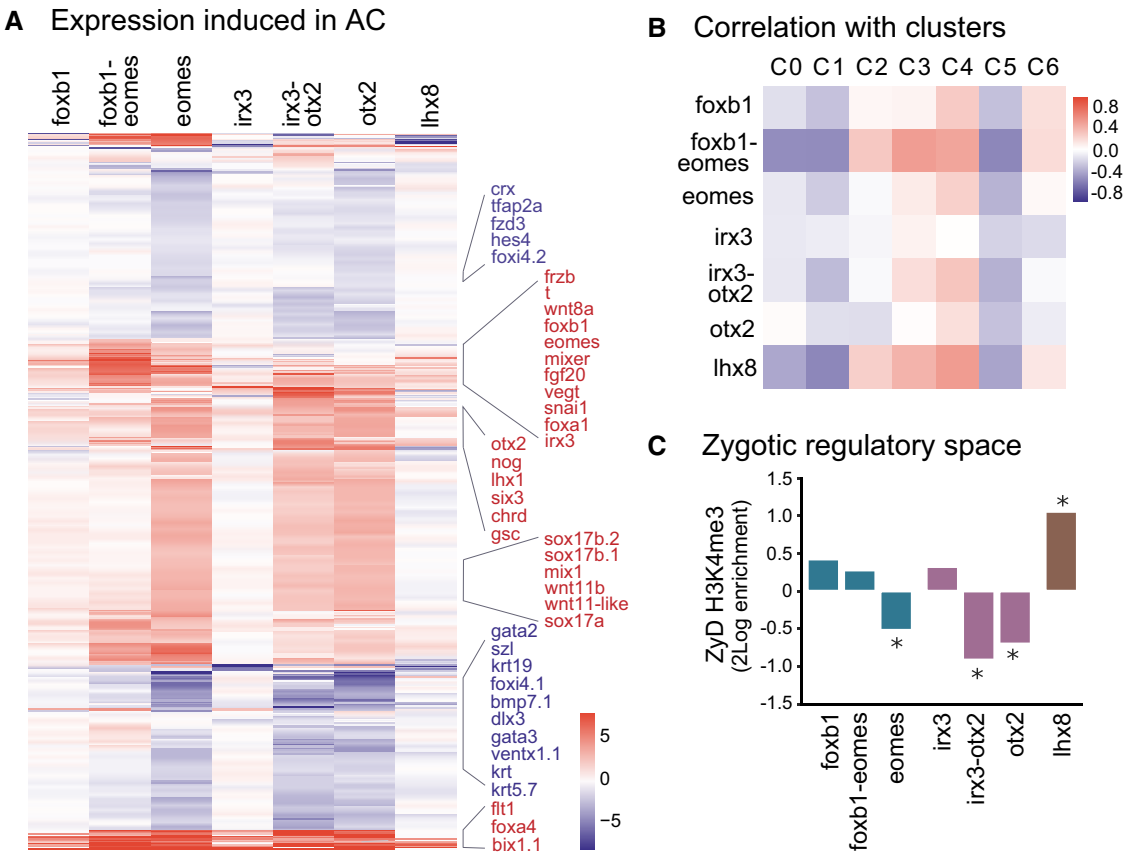

Figure 6. Induction of organizer gene expression in AC cells.

A Heatmap showing $\log_2$ fold expression changes of differentially expressed genes in AC tissues overexpressing Foxb1, Foxb1-Eomes, Eomes, Irx3, Irx3-Otx2, Otx2, and Lhx8.

B Correlation heatmap of overexpression RNA-seq samples and single-cell clusters.

C Fold enrichment of genes with zygotically defined (ZyD) H3K4me3 at their promoter in AC with transcription factor overexpression. Asterisk indicates hypergeometric *P*-value ≤ 0.01.

organizer gene expression in animal caps. Together, these observations indicate that Eomes, Otx2, and the combination of Irx3 and Otx2 tend cause expression of genes that have maternally controlled H3K4me3-decoration of their promoters, whereas H3K4me3-marking of a relatively high number of genes induced in Lhx8-caps requires zygotic transcription.

Discussion

This study explores the relationships between chromatin state, regulatory elements and spatial regulation of gene expression during early development. Genome-wide analysis of chromatin accessibility demonstrated pluripotent animal cap (AC) cells to have open chromatin for both ectoderm-expressed and mesoderm-expressed genes, whereas dorsal marginal zone (DMZ) cells exhibit a more restricted pattern of chromatin accessibility. This is concordant with studies that have shown that embryonic stem cells cultured *in vitro* have a more open, accessible chromatin compared to differentiated cells (Schlesinger & Meshorer, 2019). Lineage commitment involves changes in accessibility of genes with lineage-restricted expression, many of which are accessible in pluripotent cells but inaccessible in lineages where they are not expressed. Similarly, we found that DMZ cells exhibit reduced chromatin accessibility for ectodermal genes. The earliest accessibility detected with the ATAC-seq method roughly coincides with the mid-blastula transition. This suggests that early development involves a major transition from a generally closed chromatin state to an open state that accommodates developmental competence. This is in line with studies showing that maternal transcription factors, such as Pou5f3, Sox3, Foxh1, Otx1, and Vegt, not only bind to DNA before zygotic genome activation, but also have a role in opening up chromatin (Paraiso *et al*, 2019; Gentsch *et al*, 2019). Chromatin characteristics such as DNA methylation (Hontelez *et al*, 2015) and chromatin accessibility (Esmaeili *et al*, 2020) of Wnt targets has been suggested to represent an early-late switch in Wnt signaling (Hontelez *et al*, 2015; Nakamura *et al*, 2016). Apart from a few early but important Wnt signaling targets, our data suggest that signaling-induced transcriptional programs are broadly facilitated by open, accessible chromatin during gastrulation.

During gastrulation, cellular heterogeneity rapidly increases beyond regional differences due to induction and morphogenesis. Analysis of blastula, early gastrula, and mid-to-late gastrula single-cell data showed well-resolved cellular trajectories for non-neural ectoderm, neural ectoderm, mesoderm and organizer mesendoderm. Sequencing hand-picked single cells from dissected AC and DMZ tissue achieved both a known spatial localization and a deeper transcriptome of the cells, allowing a more detailed characterization of cell clusters. We identified organizer mesendodermal cells that express relatively high levels of Wnt inhibitors, constituting the head organizer, in addition to organizer mesodermal cells that did not abundantly express endoderm genes or Wnt inhibitors, the prospective trunk organizer. In addition, we identified cells with expression patterns consistent with superficial mesoderm and upper blastopore lip. Our analysis of whole embryo data supports the competence of *sox2/sox3*-expressing ectodermal cells for both neural as well mesodermal induction at an early gastrula stage. At the bifurcation of the neural and mesodermal trajectories, some of the cells express both *sox2* and *t (tbxt)*, which have been shown to promote, respectively, neural and mesodermal fates in an antagonistic fashion (Gentsch *et al*, 2013; Koch *et al*, 2017). There is a continuum between non-organizer and organizer mesoderm in the two-dimensional representation of the cells, with head and prospective trunk organizer cells exhibiting a more distinct identity.

By integrating single-cell transcriptomics and chromatin accessibility landscapes, we identified active motifs and associated transcription factors, which, when expressed in animal caps, induced organizer gene expression. *Eomes* induced *fst*, an activin and BMP antagonist, whereas *frzb*, a Wnt antagonist, was highly enriched in *foxb1-eomes* injected caps. Eomes and other T-box transcription factors are expressed in partially overlapping expression domains, with Eomes expressed in prospective head mesoderm (Gentsch *et al*, 2013). Foxb1 can be induced by FGF signaling (Chung *et al*, 2004) and is expressed in non-involuted mesoderm and upper blastopore lip during early gastrulation (Gamse & Sive, 2001). In our data, delta-like *dll1* expression is moderately enhanced by *foxb1-eomes* overexpression in animal caps. Its expression has been reported in dorsal marginal zone and prospective mesoderm, as well as neuroectoderm (Kinoshita *et al*, 2011), where it is involved in lateral inhibition of neurogenesis. Although *foxb1* by itself induces a gene expression pattern similar to that observed in the upper blastopore lip cells, it also induced head organizer gene expression, especially in combination with *eomes*. Notably *frzb, chrd* and *gsc* are strongly induced by the Foxb1-Eomes combination. Ectopic expression of *iroquois3* in zebrafish (*iro3, irx3* in frogs) induced organizer gene expression, including that of *lhx1* and *chrd* (Kudoh & Dawid, 2001). In our experiments, *irx3* by itself induced some organizer gene expression in animal caps, but in combination with *otx2* expression this was strongly enhanced. Previously, we showed that the embryonic chromatin state of promoters is largely established by maternal factors (Hontelez *et al*, 2015). This observation extends to the promoters of genes that are activated by zygotic transcription factors. In this study, we find that zygotic factors such as Foxb1, Eomes, Irx3, and Otx2 predominantly act within open chromatin, using promoters that have gained the H3K4me3 permissive promoter mark by the activity of maternal factors. Interestingly, the combinatorial action of the transcription factors indicates that their contributions to the early developmental program are not simply additive. We observed that the transcription factors, while strongly inducing the expression of many organizer-expressed genes, did not induce highly cluster-specific patterns of gene expression (*foxb1-eomes* C6; *irx3-otx2* C3). This indicates that their role is more selective or that their effects are further modified by other transcription factors and signaling. Single-cell technologies hold great promise to analyze developmental gene regulation, especially when combined with spatial techniques. Future single-cell analyses and perturbation studies will not only build on the current approaches, but will also provide the data and analytical power to reconstruct the gene regulatory networks in a spatio-temporally resolved manner.

Materials and Methods

Xenopus embryo manipulation

Xenopus tropicalis embryos were obtained by in vitro fertilization, dejellied in 3% cysteine and collected at the desired stages. Fertilized eggs were injected with 500 ng of synthetic mRNA at the 1-cell stage and cultured until control embryos reached stage 8. Animal caps were explanted at stage 8 and cultured until stage 10.5 in 0.1× MMR. Animal use licenses were provided by DEC permits RU-DEC 2012–116, 2014–122 and CCD approval AVD1030020171826.

Stage and tissue-specific ATAC-seq

ATAC-seq was performed as described (Bright & Veenstra, 2018). Library quality was assessed using the Agilent Bioanalyzer High Sensitivity DNA kit for checking the fragment size distribution and signal to noise ratio was checked by performing qPCR primers spanning open and closed regions. Fragments above 700 bps were removed using AMPure XP beads to reduce unamplified clusters during sequencing. The concentration of the prepared library was quantified using Qubit and KAPA library quantification kit. The libraries were sequenced on the Illumina HiSeq 2500 with 43 bps paired-end reads each (Dataset EV7). ATAC-seq reads were processed using the seq2science pipeline (v0.3.0; https://github.com/vanheeringen-lab/seq2science).

After demultiplexing, the reads were trimmed using fastp v0.20.1 (Chen *et al*, 2018) and aligned to the *X. tropicalis* genome (xt9.0 and xt10.0) with bwa-mem v0.7.17 using default settings. Reads mapping to mitochondrial DNA were excluded from the analysis together with low-quality reads including repeats and duplicates (MAPQ < 10). All mapped reads were offset by + 4 bp for the + strand and − 5 bp for the − strand. Peaks were called for each sample using MACS2 v2.2.7 (Zhang *et al*, 2008) with parameters "-q 0.05 --nomodel –shift −100 --extsize 200 –keep-dup 1".

Single-cell ATAC-seq

Nuclei were isolated with a method adapted from Mariano (Mariano, 1964). Briefly, 100 *Xenopus* embryos were suspended in 300 µl ice-cold E1-buffer (110 mM KCl, 50 mM Tris–HCl pH 7.4, 5 mM $MgCl_2$, 0.1 mM EDTA, 2 mM DTT) containing 0.25 M sucrose. To the sample, 2.4 ml ice-cold E1-buffer with 2.2 M sucrose was added and gently mixed. The embryo suspension was then layered on a 150 µl cushion of ice-cold 2.2 M sucrose in a SW60 Beckman

polyallomer tube. Nuclei were pelleted by centrifugation at 130,000 *g* at 4°C. The pellet was carefully resuspended in 250 µl nuclear buffer (25% glycerol, 25 mM Tris–HCl pH 7.4, 70 mM KCl, 5 mM $MgCl_2$, 0.2 mM EDTA, 2 mM DTT), and nuclei were washed twice by spinning them through a 20 µl 80% glycerol cushion. After the last wash, the nuclei were resuspended in 200 µl 1× nuclei buffer (10× Genomics). Libraries were generated using Chromium Single Cell ATAC (v1.1 Chemistry; 10× Genomics). Base calling, demultiplexing, and mapping were performed using cellranger-atac (v1.2.0, 10× Genomics). The reads were mapped to the *X. tropicalis* genome (xt10.0), downloaded from Xenbase (http://www.xenbase.org/, RRID:SCR_003280; Karimi *et al*, 2018). Sequencing and mapping statistics are presented in Dataset EV7. Barcodes with low sequence coverage (cells with < 1,000 transposition fragments), low enrichment (TSS enrichment < 4.5), or highly variable chromosome coverage (unique fragments per Mbp, calculated per chromosome; coefficient of variation of chromosome coverage > 1) were excluded from downstream analysis.

Single-cell RNA and total RNA library preparation and sequencing

Dissected embryo explants were dissociated into single cells using Ca^{2+} and Mg^{2+} free media (Sive *et al*, 2007). 200 cells were picked for performing single-cell RNA-seq using the modified version of STRT-seq protocol (Islam *et al*, 2012; Dong *et al*, 2018) ERCC spike-in RNA (Thermo Fisher scientific, 4456740) was added to the lysis buffer. After the reverse transcription reaction, we performed 4 + 16 cycles of PCR amplification for cDNA synthesis. Amplified cDNA was purified with the Zymo purification kit and Agencourt AMPure XP beads, respectively, and its concentration was measured with Qubit® 2.0 Fluorometer (Q32866, Life Technologies). The quality of the amplified cDNA and distribution of DNA fragment size were assessed by Agilent 2100 Bioanalyzer (Agilent Technologies) with the High Sensitivity DNA Kit (5067-4626, Agilent). The sequencing libraries were prepared using the KAPA Hyper Pre-Kit (KR0961 – v6.17, Kapa Biosystems). The concentrations of the fragments with the approximate indexed adapters were quantified by KAPA library quantification. The libraries were sequenced in the Illumina platform and the data were mapped using Bowtie (version: 1.2.2; Langmead & Salzberg, 2012) to the indexed files of the xt9 genome (cf. Dataset EV7).

Total RNA was extracted from animal caps using our in-house adapted Trizol-Zymo Hybrid protocol. The concentrations of all the RNA samples were measured using the DeNovix dsDNA High Sensitivity Assay (Catalog number: KIT-DSDNA-HIGH-1). The cDNA was constructed using KAPA RNA HyperPrep with RiboErase (Catalog number: 08278555702, Kapa Biosystems). We checked the quality of the samples by RT–qPCR using primers spanning coding regions of candidate and housekeeping genes.

ATAC-seq data analysis

Differential peaks were identified using DESeq2 (Love *et al*, 2014) and were used for further downstream analysis. Heatmaps were generated using fluff (Georgiou & van Heeringen, 2016). Differential peaks were then annotated to the closest gene using BEDtools (Quinlan & Hall, 2010) intersect with GREAT regions (McLean *et al*, 2010). Genome-wide boxplots of accessibility and transcriptional signal were plotted using ggplot2 (Wickham, 2016). Motif analysis on peak regions was performed using GimmeMotifs (preprint: Bruse & Heeringen, 2018). Analysis of single-cell ATAC-seq data was performed using ArchR (v0.9.5, https://www.archrproject.com/), with barcode filter settings (fragments > 1,000, TSS enrichment > 4.5) and clustering resolution 0.2. Pseudo-bulk profiles of the single-cell clusters were analyzed for motifs using GimmeMotifs.

Single-cell RNA-seq data analysis

The data set was filtered and quality-checked for cells and genes using the package Scater (McCarthy *et al*, 2017). The filtered data set was further loaded into R package Seurat (Satija *et al*, 2015). The hypervariable genes were used for principal component analysis, from which the statistically significant PCs were used for UMAP projection (dimensionality reduction). We identified seven distinct clusters of cells using the FindClusters function in Seurat. Based on the predicted clusters, the marker genes relevant to each cluster were taken for further analysis with other data sets. Processing and visualization of the whole embryo single-cell RNA-sequencing data (Briggs *et al*, 2018) were performed with scanpy (Wolf, Angerer & Theis, 2018). Stage 10½ AC and DMZ single-cell clusters were placed in the whole embryo data based on Spearman correlations between clusters of both data sets, based on cluster mean expression of the hypervariable genes common to the two data sets.

Integration of ATAC-seq and single-cell RNA-seq

Top hypervariable genes (HVGs) from the scRNA-seq analysis were associated with the closest AC-DMZ ATAC-seq peaks using GREAT analysis (Fig 5B and C). This peak-to-gene model comprised of a matrix, with rows as peak locations and columns as expression values of target genes across the seven clusters. This was used as input of motif prediction using the gimme maelstrom function of GimmeMotifs (preprint: Bruse & Heeringen, 2018). The transcription factors associated with the predicted motifs were then screened based on their correlation between motif activity and their gene expression across the clusters.

Data availability

The data sets produced in this study are available in the following databases:

- bulk ATAC-sequencing: GEO accession number GSE145619 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE145619).
- single-cell ATAC-sequencing: GEO accession number GSE145619 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE145619).
- single-cell RNA-sequencing: GEO accession number GSE145619 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE145619).
- bulk RNA-sequencing: GEO accession number GSE145619 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE145619).

Expanded View for this article is available online.

Acknowledgments

The authors are grateful to Dr. Huiqing (Jo) Zhou for helpful discussions, to Dr. Georgios Georgiou for bioinformatics support, and to Saskia Heffener and

Laura Wingens for help with the single-cell ATAC samples. The work of ARB and GJCV has been financially supported by the European Commission (EC), People Program (Marie Curie Actions) of the Seventh Framework Program FP7 under REA grant agreement number 607142 (DevCom). SJvH is supported by the Netherlands Organization for Scientific research (NWO), Aard-en Levenswetenschappen, grant 016.Vidi.189.081.

Author contributions

ARB: conceptualization, investigation (performing experiments), formal analysis, visualization, writing original draft; SG: investigation (performing experiments), visualization; QL: methodology, investigation (performing experiments); AG: investigation (performing experiments); SF: software, resources; MS: software, resources; SJH: formal analysis, software, resources, supervision; GJCV: Conceptualization, formal analysis, funding acquisition, resources, supervision, visualization, writing original draft. All authors were involved in writing (review and editing).

Conflict of Interest

The authors declare that they have no conflict of interest.

References

Agius E, Oelgeschläger M, Wessely O, Kemp C, De Robertis EM (2000) Endodermal Nodal-related signals and mesoderm induction in Xenopus. *Development* 127: 1173 – 1183

Angerilli A, Smialowski P, Rupp RA (2018) The Xenopus animal cap transcriptome: building a mucociliary epithelium. *Nucleic Acids Res* 46: 8772 – 8787

Balwierz PJ, Pachkov M, Arnold P, Gruber AJ, Zavolan M, Van Nimwegen E (2014) ISMARA: automated modeling of genomic signals as a democracy of regulatory motifs. *Genome Res* 24: 869 – 884

Blitz IL, Paraiso KD, Patrushev I, Chiu WTY, Cho KWY, Gilchrist MJ (2016) A catalog of Xenopus tropicalis transcription factors and their regional expression in the early gastrula stage embryo. *Dev Biol* 426: 409 – 417.

Bogdanović O, van Heeringen SJ, Veenstra GJC (2012) The epigenome in early vertebrate development. *Genesis* 50: 192 – 206

Borchers A, Pieler T (2010) Programming pluripotent precursor cells derived from Xenopus embryos to generate specific tissues and organs. *Genes* 1: 413 – 426

Briggs JA, Weinreb C, Wagner DE, Megason S, Peshkin L, Kirschner MW, Klein AM (2018) The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science* 360: eaar5780

Bright AR, Veenstra GJC (2018) Assay for transposase-accessible chromatin-sequencing using xenopus embryos. *Cold Spring Harb Protoc* 2019: 39 - 45

Bruse N, van Heeringen SJ (2018) GimmeMotifs: an analysis framework for transcription factor motif analysis. *bioRxiv* https://doi.org/10.1101/474403 [PREPRINT]

Chalmers AD, Lachani K, Shin Y, Sherwood V, Cho KWY, Papalopulu N (2006) Grainyhead-like 3, a transcription factor identified in a microarray screen, promotes the specification of the superficial layer of the embryonic epidermis. *Mech Dev* 123: 702 – 718

Chalmers AD, Welchman D, Papalopulu N (2002) Intrinsic differences between the superficial and deep layers of the Xenopus ectoderm control primary neuronal differentiation. *Dev Cell* 2: 171 – 182

Charney RM, Paraiso KD, Blitz IL, Cho KWY (2017) A gene regulatory program controlling early Xenopus mesendoderm formation: Network conservation and motifs. *Semin Cell Dev Biol* 66: 12 – 24

Chen S, Zhou Y, Chen Y, Gu J (2018) Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34: i884 – i890

Chung HA, Hyodo-Miura J, Kitayama A, Terasaka C, Nagamune T, Ueno N (2004) Screening of FGF target genes in *Xenopus* by microarray: temporal dissection of the signalling pathway using a chemical inhibitor. *Genes Cells* 9: 749 – 761

Dong J, Hu Y, Fan X, Wu X, Mao Y, Hu B, Guo H, Wen L, Tang F (2018) Single-cell RNA-seq analysis unveils a prevalent epithelial/mesenchymal hybrid state during mouse organogenesis. *Genome Biol* 19: 31

Esmaeili M, Blythe SA, Tobias JW, Zhang K, Yang J, Klein PS (2020) Chromatin accessibility and histone acetylation in the regulation of competence in early development. *Dev Biol* 462: 20 – 35

Gamse JT, Sive H (2001) Early anteroposterior division of the presumptive neurectoderm in Xenopus. *Mech Dev* 104: 21 – 36

Gao Y, Cao Q, Lu L, Zhang X, Zhang Z, Dong X, Jia W, Cao Y (2015) Kruppel-like factor family genes are expressed during *Xenopus* embryogenesis and involved in germ layer formation and body axis patterning. *Dev Dyn* 244: 1328 – 1346

Gentsch GE, Owens NDL, Martin SR, Piccinelli P, Faial T, Trotter MWB, Gilchrist MJ, Smith JC (2013) In Vivo T-Box transcription factor profiling reveals joint regulation of embryonic neuromesodermal bipotency. *Cell Rep* 4: 1185 – 1196

Gentsch GE, Spruce T, Owens NDL, Smith JC (2019) Maternal pluripotency factors initiate extensive chromatin remodelling to predefine first response to inductive signals. *Nat Commun* 10: 4269

Georgiou G, van Heeringen SJ (2016) fluff: exploratory analysis and visualization of high-throughput sequencing data. *PeerJ* 4: e2209

Hontelez S, van Kruijsbergen I, Georgiou G, van Heeringen SJ, Bogdanovic O, Lister R, Veenstra GJC (2015) Embryonic transcription is controlled by maternally defined chromatin state. *Nat Commun* 6: 10148

Islam S, Kjällquist U, Moliner A, Zajac P, Fan JB, Lönnerberg P, Linnarsson S (2012) Highly multiplexed and strand-specific single-cell RNA 5′ end sequencing. *Nat Protoc* 7: 813 – 828

Jambhekar A, Dhall A, Shi Y (2019) Roles and regulation of histone methylation in animal development. *Nat Rev Mol Cell Biol* 20: 625 – 641.

Jones EA, Woodland HR (1987) The development of animal cap cells in Xenopus: a measure of the start of animal cap competence to form mesoderm. *Development* 101: 557LP - 563

Karimi K, Fortriede JD, Lotay VS, Burns KA, Wang DZ, Fisher ME, Pells TJ, James-Zorn C, Wang Y, Ponferrada VG *et al* (2018) Xenbase: A genomic, epigenomic and transcriptomic model organism database. *Nucleic Acids Res* 46: D861 – D868

Kinoshita T, Haruta Y, Sakamoto C, Imaoka S (2011) Antagonistic role of XESR1 and XESR5 in mesoderm formation in Xenopus laevis. *Int J Dev Biol* 55: 25 – 31

Koch F, Scholze M, Wittler L, Schifferl D, Sudheer S, Grote P, Timmermann B, Macura K, Herrmann BG (2017) Antagonistic activities of Sox2 and brachyury control the fate choice of neuro-mesodermal progenitors. *Dev Cell* 42: 514 – 526

Kudoh T, Dawid IB (2001) Role of the iroquois3 homeobox gene in organizer formation. *Proc Natl Acad Sci USA* 98: 7852 – 7857

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

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

Luo T, Zhang Y, Khadka D, Rangarajan J, Cho KWY, Sargent TD (2005) Regulatory targets for transcription factor AP2 in Xenopus embryos. *Dev Growth Differ* 47: 403 – 413

Madsen JGS, Rauch A, Van Hauwaert EL, Schmidt SF, Winnefeld M, Mandrup S (2018) Integrated analysis of motif activity and gene expression changes of transcription factors. *Genome Res* 28: 243 – 255

Mariano EE (1964) The isolation of nuclei from Xenopus laevis embryonic cells. *Exp Cell Res* 34: 201 – 205

McCarthy DJ, Campbell KR, Lun ATL, Wills QF (2017) Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* 33: 1179 – 1186

McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G (2010) GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* 28: 495 – 501

Nakamura Y, De Paiva AE, Veenstra GJC, Hoppler S (2016) Tissue-and stage-specific Wnt target gene expression is controlled subsequent to β-catenin recruitment to cis-regulatory modules. *Dev* 143: 1914 – 1925

Niehrs C (1999) Head in the WNT: the molecular nature of Spemann's head organizer. *Trends Genet* 15: 314 – 319

Owens NDL, Blitz IL, Lane MA, Patrushev I, Overton JD, Gilchrist MJ, Cho KWY, Khokha MK (2016) Measuring absolute RNA copy numbers at high temporal resolution reveals transcriptome kinetics in development. *Cell Rep* 14: 632 – 647

Paraiso KD, Blitz IL, Coley M, Cheung J, Sudou N, Taira M, Cho KWY (2019) Endodermal maternal transcription factors establish super-enhancers during zygotic genome activation. *Cell Rep* 27: 2962 – 2977

Paranjpe SS, Veenstra GJC (2015) Establishing pluripotency in early development. *Biochim Biophys Acta* 1849: 626 – 636

Pera EM, Acosta H, Gouignard N, Climent M, Arregi I (2014) Active signals, gradient formation and regional specificity in neural induction. *Exp Cell Res* 321: 25 – 31

Perino M, Veenstra GJC (2016) Chromatin control of developmental dynamics and plasticity. *Dev Cell* 38: 610 – 620

Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26: 841 – 842

Satija R, Farrell JA, Gennert D, Schier AF, Regev A (2015) Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* 33: 495 – 502

Scerbo P, Girardot F, Vivien C, Markov GV, Luxardi G, Demeneix B, Kodjabachian L, Coen L (2012) Ventx factors function as nanog-like guardians of developmental potential in xenopus. *PLoS One* 7: e36855

Schlesinger S, Meshorer E (2019) Open chromatin, epigenetic plasticity, and nuclear organization in pluripotency. *Dev Cell* 48: 135 – 150.

Schneider I, Kreis J, Schweickert A, Blum M, Vick P (2019) A dual function of FGF signaling in Xenopus left-right axis formation. *Development* 146: dev173575

Sive HL, Grainger RM, Harland RM (2007) Dissociation and Reaggregation of Xenopus laevis Animal Caps. *CSH Protoc* 2007: pdb.prot4745

Steiner AB, Engleka MJ, Lu Q, Piwarzyk EC, Yaklichkin S, Lefebvre JL, Walters JW, Pineda-Salgado L, Labosky PA, Kessler DS (2006) FoxD3 regulation of Nodal in the Spemann organizer is essential for Xenopus dorsal mesoderm development. *Development* 133: 4827 – 4838

Suzuki H, Forrest ARR, Van Nimwegen E, Daub CO, Balwierz PJ, Irvine KM, Lassmann T, Ravasi T, Hasegawa Y, De Hoon MJL *et al* (2009) The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. *Nat Genet* 41: 553 – 562

Takebayashi-Suzuki K, Kitayama A, Terasaka-Iioka C, Ueno N, Suzuki A (2011) The forkhead transcription factor FoxB1 regulates the dorsal–ventral and anterior–posterior patterning of the ectoderm during early Xenopus embryogenesis. *Dev Biol* 360: 11 – 29

Tao J, Kuliyev E, Wang X, Li X, Wilanowski T, Jane SM, Mead PE, Cunningham JM (2005) BMP4-dependent expression of Xenopus Grainyhead-like 1 is essential for epidermal differentiation. *Development* 132: 1021 – 1034

Vastenhouw NL, Cao WX, Lipshitz HD (2019) The maternal-to-zygotic transition revisited. *Development* 146: dev161471

Walentek P, Schneider I, Schweickert A, Blum M (2013) Wnt11b is involved in cilia-mediated symmetry breakage during Xenopus left-right development. *PLoS One* 8: e73646

Wickham H (2016) *ggplot2: elegant graphics for data analysis*. New York, NY: Springer-Verlag

Wolf FA, Angerer P, Theis FJ (2018) SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* 19: 15

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W *et al* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9: R137