The role of maternal pioneer factors in predefining first zygotic responses to inductive signals

George E. Gentsch[1], Thomas Spruce[2], Nick D. L. Owens[3] & James C. Smith[1]

[1]Developmental Biology Laboratory, The Francis Crick Institute, London NW1 1AT, UK
[2]Centre for Genomic Regulation, Barcelona Institute for Science and Technology, 08003 Barcelona, Spain
[3]Department of Stem Cell and Developmental Biology, Pasteur Institute, 75015 Paris, France
Correspondence should be addressed to G.E.G (george.gentsch@crick.ac.uk) or J.C.S (jim.smith@crick.ac.uk)

Embryonic development yields many different cell types in response to just a few families of inductive signals. The property of a signal-receiving cell that determines how it responds to such signals, including the activation of cell type-specific genes, is known as its competence. Here, we show how maternal factors modify chromatin to specify initial competence in the frog *Xenopus tropicalis*. We identified the earliest active regulatory DNA sequences, and inferred from them critical activators of the zygotic genome. Of these, we showed that the pioneering activity of the maternal pluripotency factors Pou5f3 and Sox3 predefine competence for germ layer formation by extensively remodeling compact chromatin before the onset of signaling. The remodeling includes the opening and marking of thousands of regulatory elements, extensive chromatin looping, and the co-recruitment of signal-mediating transcription factors. Our work identifies significant developmental principles that inform our understanding of tissue engineering and tumorigenesis.

The specification of different cell types during embryonic development is achieved through the repeated use of a small number of highly conserved intercellular signals. The property of a cell that defines the way it responds to such signals (if it responds at all) is known as its competence[1]. Classical experiments with amphibian embryos show that competence is regulated in both space and time. For example, head ectoderm of the tailbud embryo responds to the underlying optic vesicle by forming a lens, while other surface ectoderm is unable to do so[2]. Similarly, on a temporal scale, naïve pluripotent (animal cap) tissue of the early *Xenopus* embryo loses the ability to form muscle in response to Nodal signaling by the mid-gastrula stage[3].

How is competence regulated at the molecular level? A simple mechanism would involve the loss of components of a signal transduction pathway, but this fails to explain differences in the context-dependent response to the same signals. In addition, although there are some tissue-specific signal receptor isoforms, transduction to the nucleus is limited to just a few signal mediators with transcriptional activity such as β-catenin (canonical Wnt signaling), Smad2 (Nodal signaling) and Smad1 (BMP signaling). Thus, a more plausible way in which tissue-specific competence might arise is through the recruitment of these signal mediators to different genomic sites.

A series of experiments indicates that cell lineage determinants such as sequence-specific transcription factors (TFs) may play a role in this. For lens induction, the responding cells of the head ectoderm depend on the presence of the eye-regulating TF Pax6[4]. However, although we know that certain TFs can act as competence factors, these TFs have not been identified in a systematic fashion, and their modes of action remain largely unknown. Our understanding of how chromatin interprets inductive signals is especially important because the generation of therapeutically relevant cells like insulin-producing pancreatic β-cells frequently relies on the deployment of signal modulators at different stages of cell differentiation *in vitro*[5]. Moreover, the initiation and progression of tumors is often associated with a change in their competence which, in some instances, is correlated with the erroneous re-activation of embryonic TFs stimulating excessive cell proliferation[6].

In an effort to analyze systematically the molecular basis of competence, we chose, for the following reasons, to analyze the first inductive signaling events in the embryo of *Xenopus tropicalis*. First, like most multicellular organisms, *X. tropicalis* begins development with a transcriptionally silent genome (**Fig. 1a**). After fertilization twelve rapid cleavages convert the egg into a mid-blastula embryo with little transcription and little diversification among its 4,096 cells[7]. Cellular homogeneity and low levels of transcription both assist in the interpretation of whole-embryo and loss-of-function chromatin profiling. Second, ENCODE studies highlight the high gene regulatory capacity of maternal or top-level TFs, which makes them more amenable to study[8]. Third, based on the nuclear accumulation of signal mediators[9] (**Fig. 1a,h**), we know a great deal about where and when inductive signaling occurs during zygotic genome activation (ZGA). Finally, *in vitro* fertilization yields thousands of synchronously developing embryos, which greatly facilitates temporal chromatin profiling.

In this way, using transcriptional, translational and multi-level chromatin profiling, we first identified the earliest active regulatory DNA sequences (*cis*-regulatory modules, CRMs) in *X. tropicalis*, and inferred from them the critical maternal activators of the zygotic genome as potential competence factors. Next, we compared the DNA binding of several maternal/zygotic TFs and signal mediators across the mid-blastula transition (MBT) to reveal the effect of TF co-expression on chromatin recruitment *in vivo*. Finally, we have demonstrated how maternal TFs of the pluripotency network restructure the chromatin landscape, which in turn determines the signal-mediated regionalization of gene activity and the

specification of the three germ layers: ectoderm, mesoderm and endoderm.

RESULTS

DNA motifs at active CRMs correlate with the affinities of frequently translated TFs during ZGA

In an effort to understand the early chromatin dynamics that may influence how signal mediators are recruited to the genome (**Fig. 1b**), we first identified transcriptionally engaged CRMs from the 32-cell to the late gastrula stage by mapping focal RNA polymerase II (RNAPII) depositions on a genome-wide scale (**Fig. 1c** and **Supplementary Table 1**). RNAPII is devoid of any DNA sequence preference and its chromatin engagement is a reliable and objective indicator of CRM usage[10]. RNAPII recruitment to CRMs proved to be dynamic, with the largest changes occurring between the 1,024-cell stage and MBT (**Fig. 1c** and **Supplementary Fig. 1a**). The analysis of enriched DNA motifs among RNAPII$^+$ CRMs suggests that pre-MBT recruitment is predominantly directed by members of the FOXH, POU, SOX and T domain TF families (**Fig. 1d** and **Supplementary Fig. 1b**).

However, our RNAPII-based approach of finding active CRMs discriminates against intronic or gene proximal CRMs because some of these are masked by RNAPII-mediated gene transcription. Thus, MBT-staged CRMs were further characterized for chromatin accessibility and for the enhancer-associated histone mark H3K4me1. In our hands, the high yolk content in early embryos made it impossible to employ transposition[11] for probing chromatin accessibility. Instead, DNase I mediated digestion followed by deep sequencing (DNase-Seq) was adapted to early *Xenopus* embryos with a novel approach to select small accessible DNA (see **Methods** and **Supplementary Fig. 1c** for exemplar comparison with other chromatin features). We detected ~20,000 accessible (DNase hypersensitive) CRMs occupied by RNAPII and flanked by H3K4me1$^+$ nucleosomes (**Fig. 1e**). The enriched DNA motifs among these CRMs were then correlated with the maternally inherited and translated sequence-specific factors identified by egg-staged mass spectrometry[12] and pre-MBT ribosome footprinting[13] to find the ZGA-critical members of the TF families (**Fig. 1f**, **Supplementary Fig. 1d** and **Supplementary Table 2**). It proved that the binding preferences of the most frequently translated maternal TFs and signal mediators matched the most significantly enriched DNA motifs (**Fig. 1g**). These were POU-SOX (Pou5f3-Sox3 heterodimer), Krüppel-like zinc finger (ZF; Sp1 and several Klf), POU (Pou5f3), SOX (Sox3), bZIP (Max), FOXH (Foxh1), ETS (Ets2), NFY (NFYa/b/c), SMAD (Smad1/2), T (mVegT, a maternal VegT isoform), TCF (Tcf/β-catenin), basic helix-span-helix domain (bHSH; Tfap2), and OTX (Otx1). Notably, the enrichment level of these motifs was independent of CRM sequence conservation (based

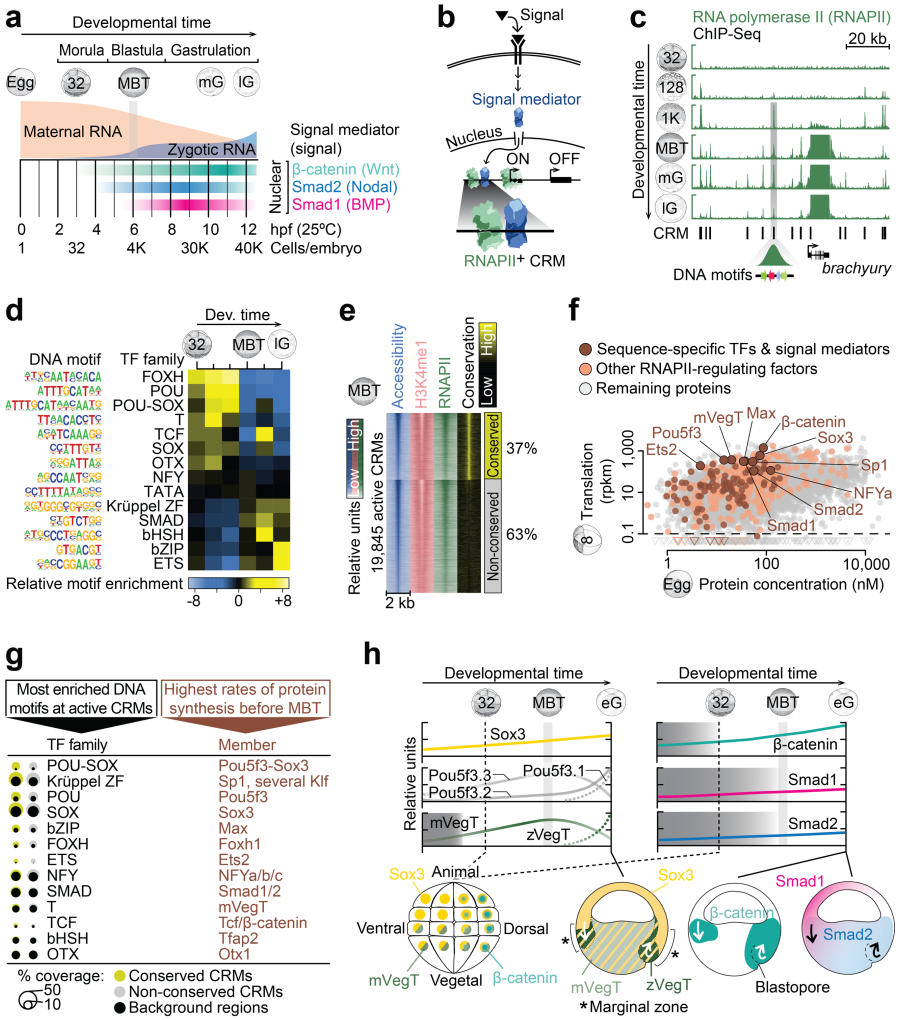


Figure 1 Finding critical sequence-specific TFs based on their significantly enriched DNA motifs at active CRMs and their high protein synthesis before MBT. (**a**) Timeline of the maternal-to-zygotic transition and earliest signaling events (nuclear accumulation of signal mediators β-catenin, Smad2 and Smad1) during early *Xenopus* embryogenesis. (**b**) Signal transduction pathway causing signal mediators to enter the nucleus and engage with RNAPII$^+$ CRMs. (**c**) Snapshot of RNAPII recruitment to CRMs from 32-cell to late gastrula stage at the *brachyury* locus. *De novo* discovery of enriched DNA motifs (illustrated as colored arrows for one RNAPII$^+$ CRM). (**d**) Temporal enrichment (Z-score) of consensus DNA motifs recognized by indicated TF families. (**e**) MBT-staged map of DNase-probed chromatin accessibility, RNAPII binding and H3K4me1 marking grouped by sequence conservation levels (phastCons) and sorted by the statistical significance of CRM accessibility. (**f**) Maternal protein concentrations[12] versus ribosome footprint levels at 8-cell stage[13]. Most frequently translated representatives of various TF families are labeled. (**g**) Matching canonical CRM-enriched DNA motifs (sorted by statistical significance) with frequently translated sequence-specific TFs and signal mediators. (**h**) Graphical illustration of protein levels and nuclear localizations of selected TFs and signal mediators based on our and previously published result[9,15-17,39,40]. Shaded boxes indicate periods of non-nuclear protein distribution. Arrows indicate tissue movements. Abbreviations for developmental timeline: 32, 128 and 1K, 32-, 128- and 1,024-cell stage; MBT, mid-blastula transition; eG, mG and lG, early, mid- and late gastrula stage.

on phastCons scoring) detected among vertebrates (**Fig. 1e,g**).

Based on this orthogonal correlation, we selected Pou5f3, Sox3 and mVegT as potential competence factors of canonical Wnt, Nodal and BMP signaling. Sox3 and presumably also Pou5f3 (based on the spatial distribution of its maternal transcripts *Pou5f3.2* and *Pou5f3.3*[14]) are detected ubiquitously, while mVegT is restricted to the vegetal hemisphere[15]. The zygotic isoform of VegT (zVegT) is induced within the marginal zone (**Fig. 1h** and **Supplementary Fig. 1h**). With respect to signal mediators, nuclear β-catenin begins to accumulate dorsally (first detected at the 32-cell stage[16]) before spreading more distinctly around the upper lip of the forming blastopore after MBT[9] (**Fig. 1a,h**). Signal-induced nuclear accumulation of Smad1 and Smad2 is first detected around MBT with Smad1 being preferentially ventral and Smad2 being vegetal and within the marginal zone[9,17] (**Fig. 1a,h** and **Supplementary Fig. 1e**).

Context-dependent TF co-expressions affect the dynamics of chromatin recruitment

Next, informed by the onset of their nuclear accumulation (**Fig. 1a,h**), we generated genome-wide chromatin profiles across MBT (**Fig. 2a** and **Supplementary Table 1**) to ask whether the recruitment of β-catenin, Smad1 and Smad2 might be influenced by co-expressed maternal TFs like Sox3, mVegT and Foxh1[18] or vice versa (**Supplementary Figs. 1j-n** and **2a**). As an outgroup control, we selected the binding profiles of several zygotic T-box TFs Eomesodermin (Eomes), zVegT, Brachyury[19] and Tbx6 (**Supplementary Figs. 1o-r** and **2b**), which collectively regulate the neuro-mesodermal cell lineage during and beyond gastrulation[20].

Maternal TFs and signal mediators shared many chromatin characteristics: (1) DNA occupancy levels followed a log-normal distribution with only a few hundred RNAPII-transcribed zygotic genes receiving high and super-enhancer-like[21] input (**Supplementary Fig. 2a-c**). (2) From the 1,024-cell to the late gastrula stage, frequent chromatin recruitment occurred on average to ~800 super-enhancers which had zygotic genes nearby (<5kb) that are primarily involved in early embryonic processes including germ layer and body axis formation (**Supplementary Fig. 2d**). (3) Among all zygotic genes, promoter-proximal regions were most consistently bound (**Supplementary Fig. 2e**). All these characteristics were also observed for chromatin accessibility and RNAPII engagement at CRMs (**Supplementary Fig. 2c,e**).

The comparison of DNA occupancy levels among and between signal mediators and TFs suggests that the recruitment of both depends on factor-intrinsic and developmental context-dependent elements (**Supplementary Fig. 3a,b**). The elements associated with developmental context were largely revealed by the binding pattern of sequence-nonspecific RNAPII at CRMs (**Supplementary Fig. 3c**). The discovery and co-clustering of DNA motifs at occupied CRMs indicate that factor-intrinsic elements include the sequence-specificity of DNA binding domains, oligomerization tendencies (e.g. Brachyury binding palindromic motifs due to its propensity to form homodimers[22]) and protein-protein interactions, while the developmental context reflects co-expressed TFs (**Fig. 2b** and **Supplementary Fig. 3d,e**). Notably, the comparison of TF- and signal-related DNA motifs suggested that maternal TFs affected the recruitment of signal mediators (see orange frame in **Fig. 2b**), but (with the exception of Foxh1, which interacts directly with Smad2[23])

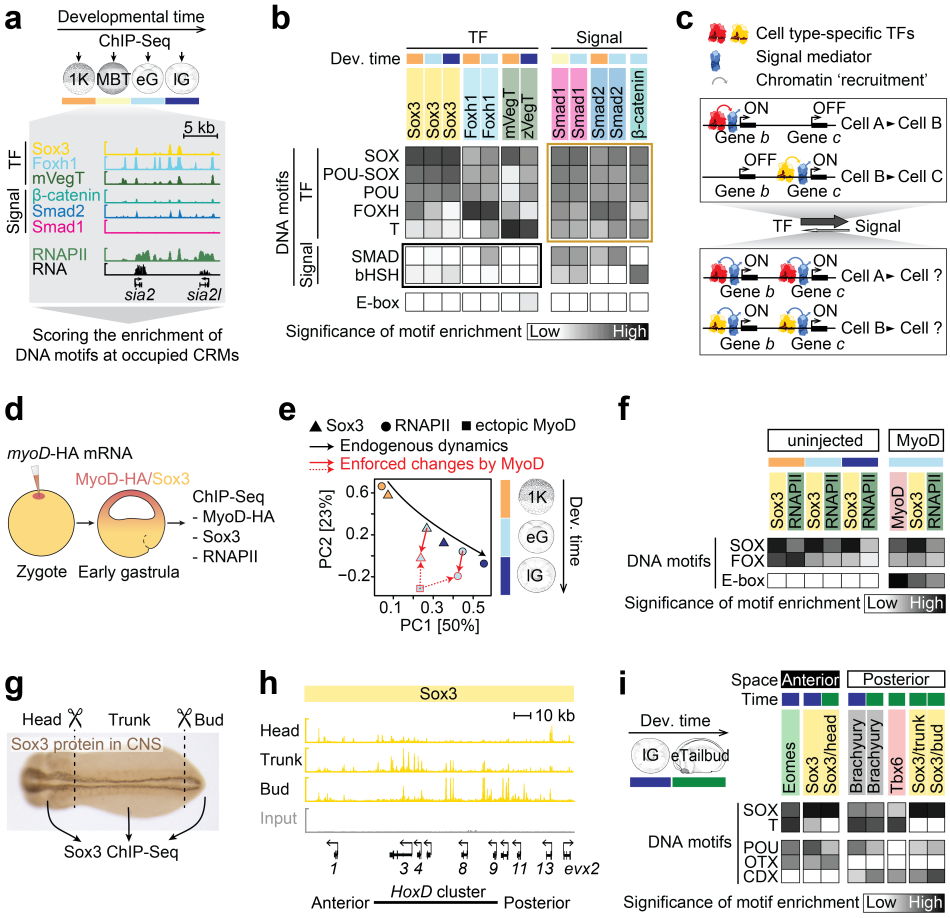


Figure 2 Spatio-temporal ChIP-Seq profiling reveals the influence of co-expressed TFs and signal mediators on chromatin engagement. (**a**) Chromatin profiling (ChIP-Seq) of selected TFs and signal mediators across MBT followed by scoring the enrichment of predetermined DNA motifs at occupied CRMs. (**b**) Statistical significance of finding TF- and signal-related DNA motifs at CRMs occupied by indicated proteins. Orange and black frames highlight the potential of TF and signal mediator-related DNA motifs to recruit signal mediators and TFs, respectively. (**c**) Illustration of TF- or signal-initiated chromatin engagements and their consequences on signal-induced gene activity. (**d-f**) MyoD-enforced recruitment of Sox3 and RNAPII. (**d**) Experimental design. (**e**) PCA of chromatin engagement. (**f**) Significance of DNA motif enrichments at occupied CRMs of indicated proteins in uninjected and MyoD-HA injected embryos. (**g-i**) CNS-specific Sox3 binding along the anterior-posterior axis (head, trunk and bud). (**g**) Experimental design. (**h**) Snapshot of Sox3 binding to *HoxD* cluster. (**i**) Significance of DNA motif enrichments at CRMs occupied by Sox3 and various T-box TFs (Eomes, Brachyury and Tbx6). (**b,e,f,i**) Graphs are separated by space (**i**, anterior and posterior) and time (color-coded as shown in **a**, **e** and **i**).

the reverse was much less the case (see black frame in **Fig. 2b**). This observation supports context-dependent signal interpretation (top panel in **Fig. 2c**): cell type-specific TFs predefine signal interpretation (of cell A or B) by shifting the binding of signal mediators and inducing different genes (gene *b* and *c*), which, ultimately, results in promoting different cell fate transitions (A to B or B to C, respectively). Conversely, signal-initiated recruitment fails to generate cell type-specific responses (bottom panel in **Fig. 2c**).

To demonstrate the importance of developmental context on chromatin engagement, we ectopically expressed HA-tagged MyoD (MyoD-HA) in animal cap cells and profiled the chromatin for MyoD-HA at the early gastrula stage, as well as for endogenous Sox3 and RNAPII (**Fig. 2d** and **Supplementary Table 1**). MyoD was chosen because its canonical E-box recognition motif is normally not enriched at the early gastrula stage (**Fig. 2b** and **Supplementary Fig. 3d**), while Sox3 and RNAPII were selected because they are ubiquitously expressed and represent sequence-specific and nonspecific DNA binding factors, respectively. As predicted, ectopic MyoD-HA altered the binding of both Sox3 and RNAPII by co-recruiting them to MyoD gene targets like *actc1* and *myl1* (**Fig. 2e** and **Supplementary Fig. 4a-c**). The E-box motif of

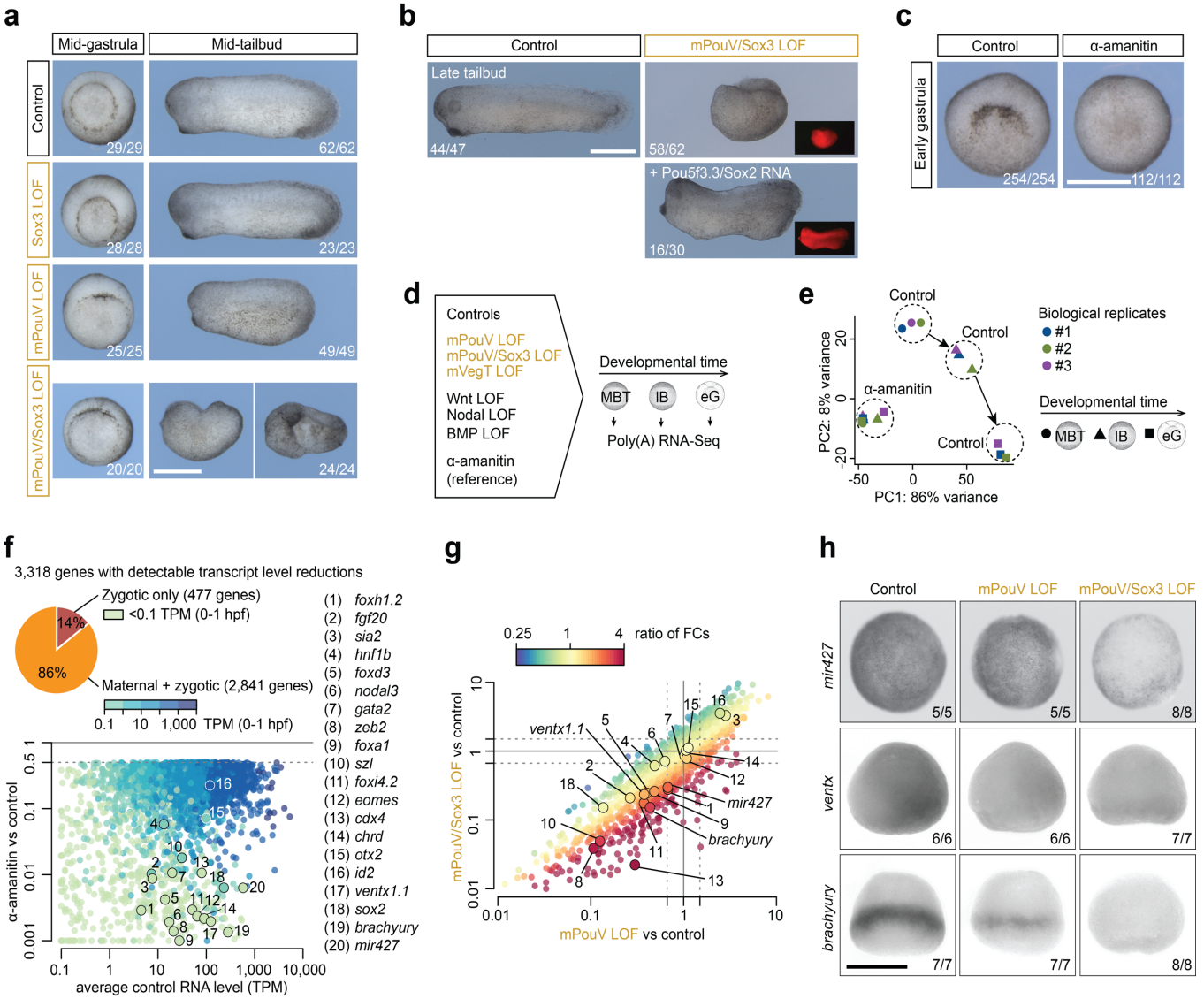


Figure 3 Synergistic relationship between maternal Pou5f3 (mPouV) and Sox3. (**a**) Morphological phenotypes caused by single and combined LOFs of Sox3 and mPouV when control embryos reached mid-gastrula and mid-tailbud stage. See also **Supplementary Movie 1**. (**b**) Phenotypical rescue of mPouV/Sox3 LOF embryos by the co-injection of both *X. laevis Pou5f3.3* and *Sox2* mRNA alongside *mCherry* mRNA as a tracer. (**c**) Morphological phenotype of embryos injected with α-amanitin. (**d**) Experimental design to profile the poly(A) RNA transcriptome over three consecutive developmental stages under indicated conditions. Abbrevations: MBT, midblastula transition (stage 8^{+}); lB, late blastula (stage 9^{+}); and eG, early gastrula (stage 10^{+}). (**e**) PCA of the poly(A) RNA transcriptome of control and α-amanitin-injected embryos at indicated developmental stages in biological triplicates (#1-3). (**f**) Detection of 3,318 zygotic genes with reduced transcript levels (≥50%) in α-amanitin-injected embryos. These genes were used as reference for all other LOFs. Dots in scatterplot are colored according to the maternal contribution[41] to the transcript level of each of these zygotic genes. List of some developmentally relevant genes whose numbers are used in (**g**), **Fig. 4a** and **Supplementary Fig. 5d**. (**g**) Scatterplot of transcript fold changes (FCs) caused by mPouV and mPouV/Sox3 LOFs with dots colored according to the ratio of FCs. (**h**) Early gastrula-staged WMISH: *mir427*, animal view; *ventx*, lateral view, ventral side facing right; and *brachyury*, dorsal view. Numbers in the right bottom corner of each image refer to the count of embryos detected with the displayed WMISH staining among all embryos analyzed per condition and *in situ* probe. Scale bars, 0.5 mm.

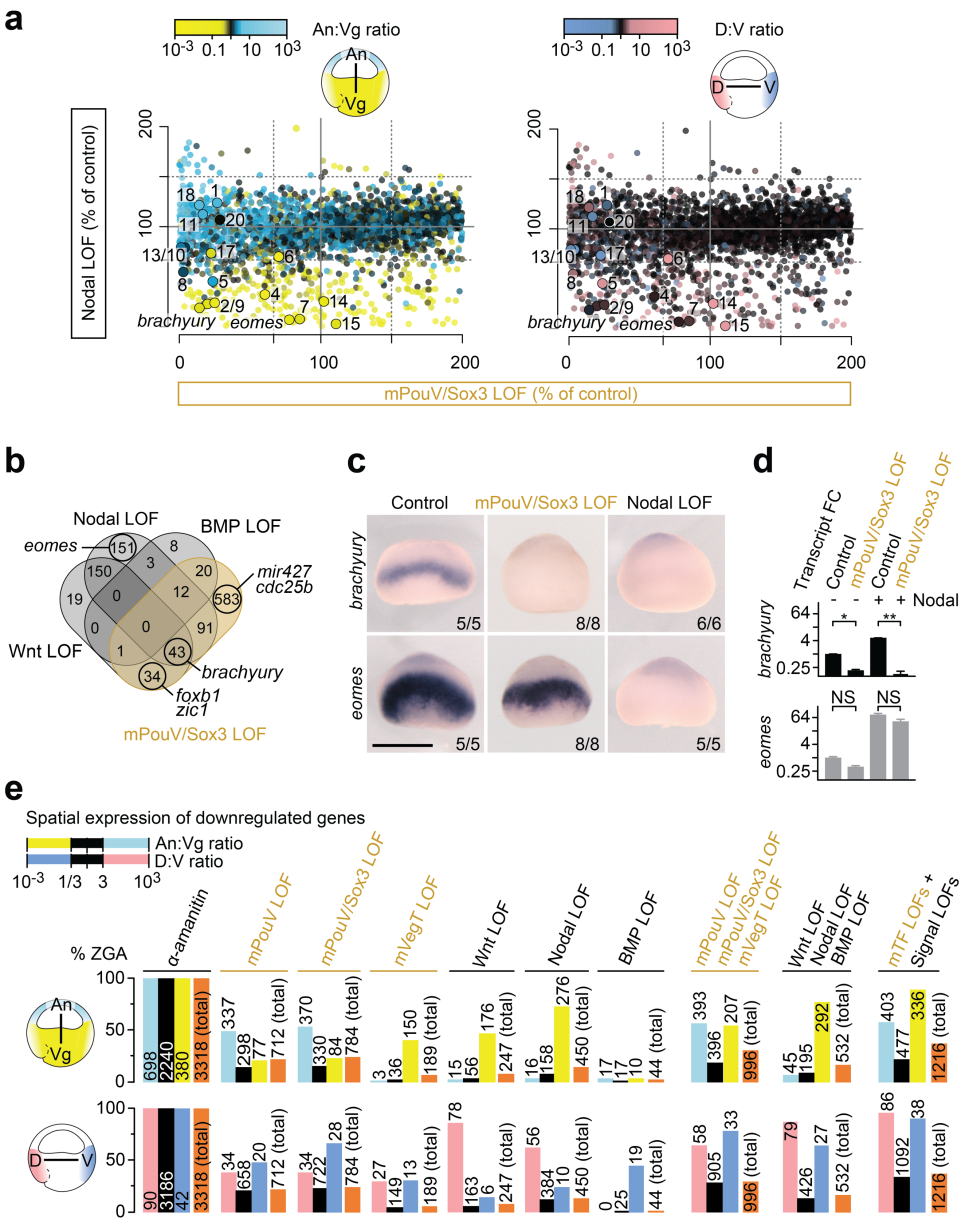


Figure 4 Signal-induced regionalization of ZGA depends on maternal TFs. (**a**) Transcriptional comparison of zygotic genes between the LOFs of maternal PouV/Sox3 and Nodal signaling. Dots are colored according to the normal ratio of transcript levels (regional expression) across the animal-vegetal (An:Vg) or dorso-ventral (D:V) axis[42]. Numbered dots refer to genes listed in **Fig. 3f**. (**b**) Venn diagram of genes downregulated by the mPouV/Sox3 LOF (orange) or the LOF of single signal transduction pathways (black). (**c**) Early gastrula-staged WMISH for *brachyury* and *eomes* under indicated LOFs. Scale bar, 0.5 mm. (**d**) Quantification of *brachyury* and *eomes* transcript levels in control and mPouV/Sox3 LOF animal caps with or without stimulated Nodal signaling. Error bars, mean+SD (n=2). Two-tailed Student's t-test: *, p=0.014; **, p=0.005; and NS, not significant (p≥0.02). (**e**) Bar graphs show the percentage (and number) of downregulated zygotic genes (% ZGA under indicated LOFs) grouped by the normal ratio of transcript levels (regional expression) across the animal-vegetal (An:Vg) and dorso-ventral (D:V) axis[42].

MyoD-HA emerged as a context-dependent motif of Sox3 and RNAPII binding, while MyoD-HA binding itself was influenced by the endogenous TF expression profile (**Fig. 2f** and **Supplementary Fig. 4b**). However, this so-called opportunistic recruitment to non-canonical binding sites, such as MyoD-HA to Sox3 gene targets (e.g. *otx2*, *sox2*), did not affect transcription in animal caps (**Supplementary Fig. 4d-f**).

The influence of TF co-expression on DNA occupancy was further substantiated by profiling chromatin for Sox3 in different anterior-posterior regions of the central nervous system (CNS) (**Fig. 2g** and **Supplementary Table 1**). The analysis of enriched DNA motifs suggested that Sox3 binding was affected by differentially expressed homeodomain TFs such as orthodenticle homeobox (OTX) in the brain (head) and caudal homeobox (CDX) in the spinal cord (trunk, tailbud) (**Fig. 2i** and **Supplementary Fig. 4g-j**). This was particularly apparent in Sox3 binding to the colinear *HoxD* cluster defining anterior-posterior cell identity (**Fig. 2h**). Similar influences on chromatin engagement were observed within anterior and posterior mesoderm marked by Eomes in gastrula embryos and Brachyury/Tbx6 in gastrula and early tailbud embryos, respectively (**Fig. 2i** and **Supplementary Fig. 4h**). On a temporal scale, the influence of FOXH and POU motif-recognizing TFs in recruiting Sox3 and T-box TF was more pronounced early than late in development (**Fig. 2i** and **Supplementary Figs. 3d** and **4h**).

Signal-induced regionalization of ZGA depends on maternal TFs

To ask whether signal-mediated ZGA requires maternal TFs, as suggested by the observed chromatin dynamics (**Fig. 2**), we next compared the effects of loss of maternal Sox3/PouV (Pou5f3.2 and Pou5f3.3; mPouV) or VegT (mVegT), and canonical Wnt, Nodal or BMP signal transduction, on zygotic transcription from MBT to late blastula (stage 9^{+}) and early gastrula stage (stage 10^{+}) (**Fig. 3d** and **Supplementary Table 1**). The high translation frequencies (**Fig. 1f** and **Supplementary Fig. 1d**) coupled with steady or even transient protein levels around MBT (**Fig. 1h**), suggest that maternal TFs and signal mediators have short half-lives. Consistent with this idea, the injection of antisense morpholino oligonucleotides (MOs) blocking

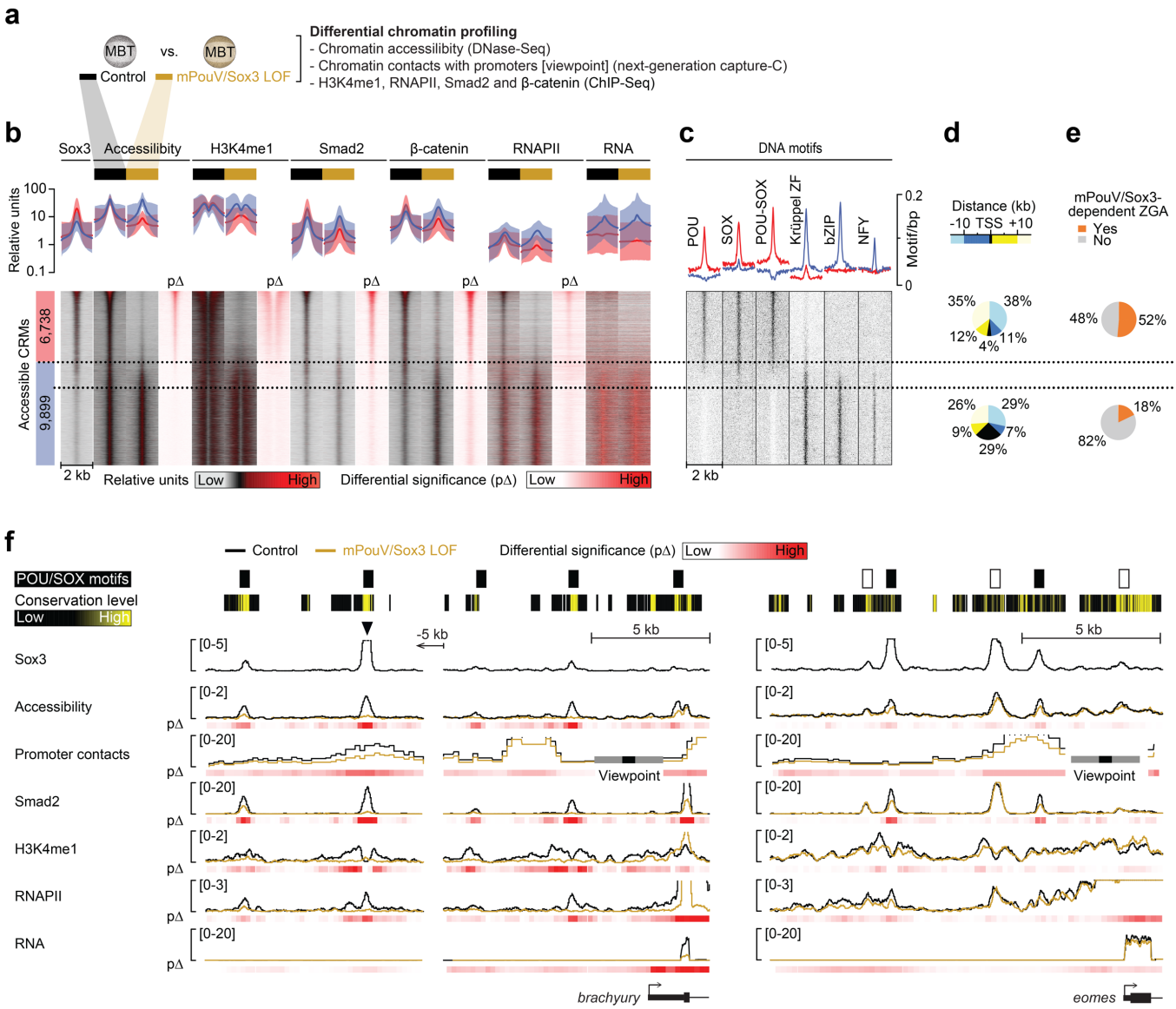


Figure 5 Maternal pluripotency TFs remodel the zygotic chromatin landscape to establish competence before MBT. (**a**) Approach to reveal the effect of mPouV/Sox3 on chromatin at MBT. (**b**) Heat map of chromatin (n=2) and RNA (n=3) profiles at accessible CRMs sorted by significantly reduced DNase cleavages (n=2) under mPouV/Sox3 LOF. β-catenin binding profile was generated at late blastula stage rather than early blastula stage (Sox3) or MBT (all others). (**b,c**) Characterization of various chromatin features, expressed RNA (**b**) and enriched DNA motifs (**c**) at CRMs of control and mPouV/Sox3 LOF embryos sorted by the significance of reduced accessibility (DNase cleavages) under mPouV/Sox3 LOF. The CRMs are grouped according to the differential FDR as shown in **Supplementary Fig. 7b**: affected (red group) and unaffected (blue group) CRMs. The β-catenin profiles were generated at late blastula stage rather than early blastula stage (Sox3) or MBT (all others). Top row, log-scaled meta-profiles (mean±SD) for chromatin features and RNA followed by the meta-profile of DNA motif instances for both affected (red) and unaffected (blue) CRMs. Bottom row, heat map of normalized read coverage and differential p-values (pΔ) (**b**) followed by co-aligned map of DNA motifs (**c**). (**d**) Distance of CRMs to nearest TSSs. (**e**) CRM association with mPouV/Sox3-dependent ZGA. (**f**) Superimposed line plots showing the LOF-induced changes to various chromatin features including chromatin contacts (n=3) with promoters (viewpoints) of *brachyury* and *eomes*. Heat maps (pΔ) below each superimposed line plot show statistical significance of mPouV/Sox3-induced changes. Header highlights the occurrence of POU/SOX motifs (black filled rectangles) at CRMs, the conservation level (phastCons) of the genomic DNA sequence among vertebrates and a strongly affected chromatin site with an arrowhead.

the translation of maternal transcripts such as *Sox3* or *mVegT* was effective in reducing protein levels (**Supplementary Fig. 1f-i**). We used previously-validated MOs to knock down mPouV[24] and β-catenin[25]. The nuclear accumulation of Smad1 and Smad2 in response to BMP and Nodal signaling was inhibited with the small molecules LDN193189[26,27] and SB431542[28,29], respectively. The morphological defects of these loss-of-function (LOF) treatments ranged from undetectable (Sox3), to weak (mVegT, BMP), to moderate (mPouV), to severe (mPouV/Sox3, Nodal, β-catenin) (**Fig. 3a** and **Supplementary Fig. 5a,c**). Moderate and severe defects affected gastrulation, while weak ones were only obvious later. All phenotypes were either in line with previous publications (β-catenin[25], Nodal[30] and BMP[31]) or could be rescued at the morphological (mPouV/Sox3; **Fig. 3b**) or transcriptional level (mVegT; **Supplementary Fig. 5b**) by co-injecting cognate mRNA. Interestingly, the role of maternal Sox3 could only be detected by

knocking it down together with mPouV. In contrast to single Sox3 or mPouV LOF, double LOF embryos failed to close the blastopore (**Fig. 3a** and **Supplementary Movie 1**).

Transcriptome analysis of TF and signal LOF embryos was confined to the 3,318 zygotic genes for which ≥two-fold reductions in exonic and/or intronic transcript level could be detected following injection of the RNAPII inhibitor α-amanitin (**Supplementary Table 3**). About 86% of these genes had a maternal contribution of ≥1 per 10 million transcripts (**Fig. 3f**). α-Amanitin prevented any transcriptional changes and thus blocked the gastrulation movements that are normally initiated by ZGA (**Fig. 3c,e**). The LOF-mediated reduction of ZGA ranged from ~1% for BMP to ~24% for mPouV/Sox3 (**Supplementary Fig. 6a**). We note that the additional loss of Sox3 in mPouV LOF embryos further reduced ZGA of developmentally relevant genes, including those that are expressed ubiquitously, like *mir427*, or those that are activated in a subset of mPouV/Sox3^{+} cells, like *ventx* and *brachyury* (**Fig. 3g,h**).

Spatial analysis of reduced gene activation confirmed that many of the 784 mPouV/Sox3-dependent genes showed enriched expression along the animal-vegetal or dorso-ventral axes (**Fig. 4a,e**). More specifically, comparison with signal LOFs revealed that 201 of the genes tissue-specifically induced by Wnt, Nodal or BMP also depended on mPouV/Sox3 (**Fig. 4b** and **Supplementary Fig. 5d,f**). Likewise, 142 of the 189 mVegT-dependent genes were activated in response to signaling within the vegetal hemisphere, such as *foxa1*, *nodal5/6* and *sox17a/b* (**Supplementary Fig. 5d,e**). Loss of mVegT and mPouV/Sox3 jointly affected 49 genes, 42 of which also depended on Wnt, Nodal or BMP signaling (**Supplementary Fig. 5f**).

Remarkably, the requirement for particular maternal TFs varied even among related genes with similar expression patterns that are activated by the same signals. For instance, *brachyury* and *eomes* are both activated by Nodal signaling in the marginal zone between the animal and vegetal hemispheres, but only *brachyury* requires mPouV/Sox3 for its signal responsiveness (**Fig. 4a-c**). Significantly, DNA occupancy levels did not explain the difference in gene regulation between *brachyury* and *eomes*, because both gene loci showed high levels of Sox3 binding (see Sox3 track in **Fig. 5f**). Similar dependencies were found for Wnt-responsive genes like *foxb1* and *zic1* (**Supplementary Fig. 5g**). We confirmed these mPouV/Sox3-dependent signal inductions by treating control and mPouV/Sox3-depleted animal cap tissue with Nodal (Activin) and Wnt (CHR99021) agonists (**Fig. 4d** and **Supplementary Fig. 5h**). Interestingly, mPouV/Sox3 also facilitated basal low-level expression of (especially) *brachyury* and *foxb1* without Nodal or Wnt stimulation.

Together, our selected maternal TFs and signals activated 1,216 of the 3,318 (~37%) zygotic genes. These included substantial percentages of genes with preferential expression in the animal (~58%), the vegetal (~88%), the dorsal (~96%) and the ventral (~90%) part of the embryo, while many ubiquitously expressed genes remained unaffected (**Fig. 4e**). Thus, mPouV and Sox3 tend to affect biological functions for which signal-induced regionalization of ZGA is essential, such as the formation of the main body axes and the segregation of the three germ layers (**Supplementary Fig. 6b**). We also note that LOF of mPouV/Sox3 caused significant increases in transcript levels of active genes encoding gamete-specific biological processes, suggesting that reprogramming towards embryonic pluripotency was compromised (**Supplementary Fig. 6a,b**).

Pioneering activity of pluripotency TFs predefines signal-mediated gene induction

To discover how maternal TFs allow cell type-specific genes to be induced by intercellular signals, we compared various chromatin features between control and mPouV/Sox3 LOF embryos at MBT (**Fig. 5a** and **Supplementary Table 1**). Genome-wide analysis showed that mPouV/Sox3 LOF strongly reduced chromatin accessibility as measured by the significant loss of DNase cleavages in 6,738 of the 16,637 (~41%) CRMs (**Fig. 5b** and **Supplementary Fig. 7a-c**). Sorting of these CRMs according to the significance of lost accessibility reveals that chromatin opening depends on the pioneering activity of mPouV and Sox3 to recognize their canonical motifs in compacted chromatin (**Fig. 5c**). By comparison, unaffected CRMs rarely contained canonical POU/SOX motifs, and were strongly enriched for promoter-centric motifs of the Krüppel ZF, bZIP and NFY protein family (**Fig. 5c,d**). At the extreme end of

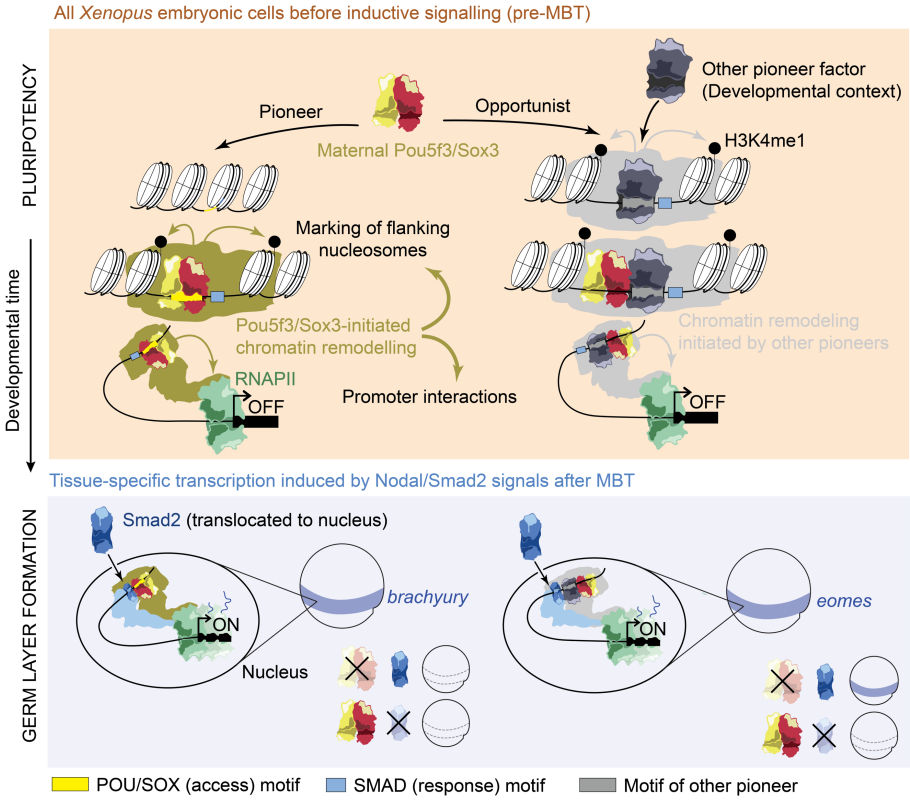


Figure 6 Model of pioneer-initiated chromatin remodeling to prime the first transcriptional response to inductive signals during ZGA.

affected loci (e.g. *brachyury*, *foxb1*, *cdc25b* and *zic1*) entire super-enhancers became accessible through mPouV/Sox3 activity (**Fig. 5f** and **Supplementary Fig. 7g**).

The loss of accessibility triggered further profound changes to chromatin *in situ* (see arrowheads in **Fig. 5f** and **Supplementary Fig. 7g** for strongly affected CRMs). Chromatin looping from affected CRMs to promoters (**Fig. 5f** and **Supplementary Fig. 7d-g**) was significantly reduced and so was the deposition of CRM-flanking H3K4me1 and co-recruitment of RNAPII and signal mediators Smad2 and β-catenin (**Fig. 5b,f** and **Supplementary Fig. 7g**). At the transcriptional level, lower usage of CRMs coincides with the reduction of coding as well as local non-coding RNA (**Fig. 5b,e** and **Supplementary Fig. 7g**). Importantly, the differential profiling explained how chromatin predetermines signal induction and why, for instance, Nodal-induced transcription of *brachyury* was strongly affected by the loss of mPouV/Sox3 and that of *Eomes* less so. In contrast to those of *eomes*, all promoter-tied CRMs of *brachyury* contain canonical POU/SOX motifs, so they were not accessible to Smad2 in mPouV/Sox3 LOF embryos (**Fig. 5f**). Thus, Smad2 interactions with critical Nodal responsive regulatory elements of *eomes* remained intact, but those of *brachyury* were impeded by compacted chromatin. Pioneer-initiated competence also applied to Wnt signaling with β-catenin failing to engage with Wnt responsive regulatory elements of *foxb1* in the absence of mPouV/Sox3 (**Supplementary Fig. 7g**).

DISCUSSION

Our results allow us to propose a model (**Fig. 6**) of pioneer-initiated chromatin remodeling, or priming, that unlocks context-specific CRMs, some of which contain signal responsive elements to enhance and regionalize transcription. The context is imprinted on chromatin by co-expressed sequence-specific TFs recognizing their canonical DNA binding motifs in compacted chromatin. Based on *in vitro* experiments[32], partial DNA motifs exposed on nucleosome surfaces can be enough to initiate chromatin access. Subsequent displacement of nucleosomes may be driven by cooperation among several TFs and the higher affinity of TFs for free genomic DNA. In the context of establishing pluripotency before MBT, we show that the pioneering activity of maternal Pou5f3 (mPouV) and Sox3 recognizing their POU/SOX motifs (**Fig. 6**, left branch in top panel) triggers extensive chromatin remodeling *in situ*. This includes the H3K4me1 marking of CRM-flanking nucleosomes and chromatin looping with promoters. Newly primed CRMs can initiate further canonical as well as opportunistic (sequence-nonspecific) binding. Thus, irrespective of their preference for certain DNA sequences, chromatin factors (e.g. TFs, signal mediators, RNAPII) can appear on accessible CRMs without any canonical DNA binding motifs (**Fig. 6**, right branch in top panel). However, at least the opportunistic recruitment of mPouV and Sox3 seems to have little to no effect on changing the chromatin landscape and its transcriptional readout. Such recruitment behaviors can be difficult to mask without the analysis of multi-level chromatin under LOF conditions. Overall, the canonical binding of mPouV/Sox3 strongly contributes to the pioneering of two fifth of all accessible CRMs at MBT which permits or instructs one quarter of the ZGA. We often observe that the usage of CRMs also coincides with the low-level activation of proximal non-coding RNA.

Though we have not performed chromatin profiling on a single-cell level, we suggest for three reasons that this priming occurs in every embryonic cell before the onset of regional Wnt, Nodal or BMP signaling that spatio-temporally restricts the recruitment of signal mediators to specific tissue domains. First, the nuclear accumulation of mPouV and Sox3 occurs ubiquitously and precedes that of signal mediators. Second, mPouV/Sox3 LOF affects chromatin similarly at zygotic genes with (e.g. *foxb1*) or without (e.g. *cdc25b*) tissue-specific expression. And finally, dissected animal caps replicate the *in vivo* response of the marginal zone tissue to Nodal and Wnt signaling. Importantly, consistent with context-dependent signal interpretation, our findings suggest that signal mediators have inferior pioneering activity and thus rely on cooperating pioneer factors like mPouV and Sox3 to make their signal response elements accessible. Analysis of the related genes *eomes* and *brachyury*, both of which are induced by Nodal signaling, but with only *brachyury* depending on mPouV/Sox3, reveals that the same context-dependent competence can be differentially encoded in the genome (e.g. presence or absence of POU/SOX motifs in signal-responsive CRMs) (**Fig. 6**, lower panel). This regulation of competence also applies to Wnt signaling and probably, based on transcriptional dependencies shown here and in the spinal cord[33], to other signaling pathways such as BMP and Sonic hedgehog.

Our work has focused on the pioneering activity of mPouV and Sox3, and the enrichment of DNA recognition motifs at active CRMs before MBT suggests that only a few more maternal TF families, including those of T-box, FoxH and Krüppel-like ZF families, are involved in the bulk acquisition of competence at the MBT. Indeed, many of these TF family members are relatively abundant and rapidly turned over, suggesting strong and dynamic TF activity[34]. Most of these TFs are already present in the *Xenopus* egg, which confers pluripotency to somatic cells when they are injected into its cytoplasm[35]. Their activity generates a signature of accessible DNA motifs of pluripotency or early cell lineage determination that is conserved among vertebrates and is most reminiscent of the chromatin footprints in embryonic stem cells *in vitro*[36]. Remarkably, as in the human genome[36], POU/SOX motifs populate distal CRMs, while Krüppel ZF motifs are frequently found in promoter-proximal CRMs. This creates a functional separation among the pluripotency factors where PouV/SoxB1 and Krüppel-like factors remodel enhancers and promoters, respectively. Collectively, we show that mPouV/Sox3 predetermine cell fate decisions by initiating access to signal-responsive CRMs via POU/SOX motifs. The usage of these permissive CRMs is most prominent in pluripotent cells[37] and proves to be critical *in vivo* to the induction of zygotic genes with functions in germ layer formation and primary axis determination.

The findings presented here reveal close connections between ZGA and pluripotency, and we anticipate that competence is controlled in human embryonic stem cells in a manner resembling that in the early *Xenopus* embryo. The approach described here can be applied to any other cell type to discover the molecular basis of its competence. Ultimately, this will generate a lexicon of competence that outlines which pioneer factors unlock which (signal-responsive) CRMs and gene loci. For example, the motif compositions of CRMs later engaged in the *Xenopus* tailbud embryo points at the potential of OTX and CDX TFs in conveying competence in anterior and posterior compartments, respectively. Interestingly, various tumors are associated with the mis-expression

of pioneer factors[38] and thus may display different competencies from the surrounding host tissue. Conversely, knowing which pioneers are required to trigger a specific signal response will increase the success of engineering patient-specific tissue for transplantation therapy. On a broader level, our profiling of chromatin states under LOF conditions discriminates functional from non-functional binding and provides a promising avenue of deciphering the non-coding part of the genome for basic and therapeutic research.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession code. Sequencing data are deposited in the Gene Expression Omnibus (GEO) database with accession number GSE113186.

ACKNOWLEDGMENTS

We thank Abdul Sesay, Leena Bhaw, Harsha Jani, Deborah Jackson and Meena Anissi for deep sequencing; Mike Klymkowsky, Manolis Gialitakis and John Gurdon for supplying antibodies; Stefan Hoppler and Yukio Nakamara for discussing β-catenin ChIP protocol; Jim Hughes and Damien Downes for sharing pre-publication results of an improved next-generation capture-C method; Ying Wang for providing sequence conservation tracks; Mareike Thompson, Rita Monteiro and Clara Collart for critical reading of the manuscript; and the Smith lab for discussions and advice. G.E.G and J.C.S. were supported by the Medical Research Council (program number U117597140) and are now supported by the Francis Crick Institute, which receives its core funding from Cancer Research UK (FC001-157), the UK Medical Research Council (FC001-157), and the Wellcome Trust (FC001-157).

AUTHOR CONTRIBUTIONS

Conceptualization, G.E.G.; Methodology, G.E.G; Software, G.E.G.; Formal Analysis, G.E.G. and N.D.L.O.; Investigation, G.E.G. and T.S.; Writing – Original Draft, G.E.G. and J.C.S.; Writing – Review & Editing, G.E.G and J.C.S.; Funding Acquisition, G.E.G. and J.C.S.; Supervision, G.E.G. and J.C.S.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Waddington, C. H. *Organisers and Genes*. Cambridge University Press (1940).
2. Grainger, R. M. Embryonic lens induction: shedding light on vertebrate tissue determination. *Trends Genet.* **8,** 349–355 (1992).
3. Tada, M., O'Reilly, M. & Smith, J. Analysis of competence and of Brachyury autoinduction by use of hormone-inducible Xbra. *Development* **124,** 2225–2234 (1997).
4. Fujiwara, M., Uchida, T., Osumi-Yamashita, N. & Eto, K. Uchida rat (rSey): a new mutant rat with craniofacial abnormalities resembling those of the mouse Sey mutant. *Differentiation* **57,** 31–38 (1994).
5. Pagliuca, F. W. *et al.* Generation of functional human pancreatic β cells in vitro. *Cell* **159,** 428–439 (2014).
6. Torres, J. *et al.* A switch in transcription and cell fate governs the onset of an epigenetically-deregulated tumor in Drosophila. *Elife* **7,** e32697 (2018).
7. Briggs, J. A. *et al.* The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science* eaar5780 (2018).
8. Gerstein, M. B. *et al.* Architecture of the human regulatory network derived from ENCODE data. *Nature* **489,** 91–100 (2012).
9. Schohl, A. & Fagotto, F. Beta-catenin, MAPK and Smad signaling during early Xenopus development. *Development* **129,** 37–52 (2002).
10. Buecker, C. & Wysocka, J. Enhancers as information integration hubs in development: lessons from genomics. *Trends Genet.* **28,** 276–284 (2012).
11. 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 (2013).
12. Wühr, M. *et al.* Deep proteomics of the Xenopus laevis egg using an mRNA-derived reference database. *Curr. Biol.* **24,** 1467–1475 (2014).
13. Subtelny, A. O., Eichhorn, S. W., Chen, G. R., Sive, H. & Bartel, D. P. Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature* **508,** 66–71 (2014).
14. De Domenico, E., Owens, N. D. L., Grant, I. M., Gomes-Faria, R. & Gilchrist, M. J. Molecular asymmetry in the 8-cell stage Xenopus tropicalis embryo described by single blastomere transcript sequencing. *Dev. Biol.* **408,** 252–268 (2015).
15. Stennard, F., Zorn, A. M., Ryan, K., Garrett, N. & Gurdon, J. B. Differential expression of VegT and Antipodean protein isoforms in Xenopus. *Mech. Dev.* **86,** 87–98 (1999).
16. Larabell, C. A. *et al.* Establishment of the dorso-ventral axis in Xenopus embryos is presaged by early asymmetries in beta-catenin that are modulated by the Wnt signaling pathway. *J. Cell. Biol.* **136,** 1123–1136 (1997).
17. Faure, S., Lee, M. A., Keller, T., Dijke, ten, P. & Whitman, M. Endogenous patterns of TGFbeta superfamily signaling during early Xenopus development. *Development* **127,** 2917–2931 (2000).
18. Charney, R. M. *et al.* Foxh1 Occupies cis-Regulatory Modules Prior to Dynamic Transcription Factor Interactions Controlling the Mesendoderm Gene Program. *Dev. Cell* **40,** 595–607.e4 (2017).
19. Gentsch, G. E. *et al.* In vivo T-box transcription factor profiling reveals joint regulation of embryonic neuromesodermal bipotency. *Cell Rep.* **4,** 1185–1196 (2013).
20. Gentsch, G. E., Monteiro, R. S. & Smith, J. C. Cooperation Between T-Box Factors Regulates the Continuous Segregation of Germ Layers During Vertebrate Embryogenesis. *Curr. Top. Dev. Biol.* **122,** 117–159 (2017).
21. Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153,** 307–319 (2013).
22. Müller, C. W. & Herrmann, B. G. Crystallographic structure of the T domain-DNA complex of the Brachyury transcription factor. *Nature* **389,** 884–888 (1997).
23. Chen, X., Rubock, M. J. & Whitman, M. A transcriptional partner for MAD proteins in TGF-beta signalling. *Nature* **383,** 691–696 (1996).
24. Chiu, W. T. *et al.* Genome-wide view of TGF /Foxh1 regulation of the early mesendoderm program. *Development* **141,** 1–114 (2014).
25. Heasman, J., Kofron, M. & Wylie, C. βCatenin Signaling Activity Dissected in the Early Xenopus Embryo: A Novel Antisense Approach. *Dev. Biol.* **222,** 124–134 (2000).
26. Cuny, G. D. *et al.* Structure-activity relationship study of bone morphogenetic protein (BMP) signaling inhibitors. *Bioorg. Med. Chem. Lett.* **18,** 4388–4392 (2008).
27. Young, J. J. *et al.* Noggin is required for first pharyngeal arch differentiation in the frog Xenopus tropicalis. *Dev. Biol.* **426,** 245–254 (2017).
28. Ho, D. M., Chan, J., Bayliss, P. & Whitman, M. Inhibitor-resistant type I receptors reveal specific requirements for TGF-beta signaling in vivo. *Dev. Biol.* **295,** 730–742 (2006).
29. Inman, G. J. *et al.* SB-431542 is a potent and specific inhibitor of transforming growth factor-beta superfamily type I activin receptor-like kinase (ALK) receptors ALK4, ALK5, and ALK7. *Mol. Pharmacol.* **62,** 65–74 (2002).
30. Skirkanich, J., Luxardi, G., Yang, J., Kodjabachian, L. & Klein, P. S. An essential role for transcription before the MBT in Xenopus laevis. *Dev. Biol.* **357,** 478–491 (2011).
31. Reversade, B., Kuroda, H., Lee, H., Mays, A. & De Robertis, E. M. Depletion of Bmp2, Bmp4, Bmp7 and Spemann organizer signals induces massive brain formation in Xenopus embryos. *Development* **132,** 3381–3392 (2005).
32. Soufi, A. *et al.* Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* **161,** 555–568 (2015).
33. Oosterveen, T. *et al.* Mechanistic differences in the transcriptional interpretation of local and long-range Shh morphogen signaling. *Dev. Cell* **23,** 1006–1019 (2012).
34. Gurdon, J. B. Cell Fate Determination by Transcription Factors. *Curr. Top. Dev. Biol.* **116,** 445–454 (2016).
35. Pasque, V., Jullien, J., Miyamoto, K., Halley-Stott, R. P. & Gurdon, J. B. Epigenetic factors influencing resistance to nuclear reprogramming. *Trends Genet.* **27,** 516–525 (2011).
36. Neph, S. *et al.* An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* **489,** 83–90 (2012).

37. Tapia, N. *et al.* Dissecting the role of distinct OCT4-SOX2 heterodimer configurations in pluripotency. *Sci. Rep.* **5,** 13533 (2015).
38. Iwafuchi-Doi, M. & Zaret, K. S. Cell fate control by pioneer transcription factors. *Development* **143,** 1833–1837 (2016).
39. Peshkin, L. *et al.* On the Relationship of Protein and mRNA Dynamics in Vertebrate Embryonic Development. *Dev. Cell* **35,** 383–394 (2015).
40. Zhang, C., Basta, T., Jensen, E.D. & Klymkowsky, M.W. The beta-catenin/VegT-regulated early zygotic gene Xnr5 is a direct target of SOX3 regulation. *Development* **130,** 5609–5624 (2003).
41. Owens, N. D. L. *et al.* Measuring Absolute RNA Copy Numbers at High Temporal Resolution Reveals Transcriptome Kinetics in Development. *Cell Rep.* **14,** 632–647 (2016).
42. Blitz, I. L. *et al.* A catalog of Xenopus tropicalis transcription factors and their regional expression in the early gastrula stage embryo. *Dev. Biol.* **426,** 409–417 (2017).
43. Hontelez, S. *et al.* Embryonic transcription is controlled by maternally defined chromatin state. *Nat. Commun.* **6,** 10148 (2015).
44. Beck, C. W. & Slack, J. M. Analysis of the developing Xenopus tail bud reveals separate phases of gene expression during determination and outgrowth. *Mech. Dev.* **72,** 41–52 (1998).

ONLINE METHODS

***Xenopus* manipulation.** Standard procedures were used for ovulation, fertilization, and manipulation and incubation of embryos[45,46]. Briefly, frogs were obtained from Nasco (Wisconsin, USA). Ovulation was induced by injecting serum gonadotropin (Intervet) and chorionic gonadotropin (Intervet) into the dorsal lymph sac of mature female frogs. Eggs were fertilized *in vitro* with sperm solution consisting of 90% Leibovitz's L-15 medium (Gibco) and 10% fetal calf serum (Gibco). After 10 min, fertilized eggs were de-jellied with 2.2% (w/v) L-cysteine (Sigma) equilibrated to pH 8.0. *X. tropicalis* embryos were cultured in 5% Marc's Modified Ringer's solution (MMR)[46] at 21ºC-28ºC. *X. laevis* embryos were cultured in 10% Normal Amphibian Medium (NAM)[46] at 14ºC-22ºC. Embryos were staged as previously described[47]. All *Xenopus* work fully complied with the UK Animals (Scientific Procedures) Act 1986 as implemented by the Francis Crick Institute.

Nucleic acid injections and treatments of embryos. Microinjections were performed using calibrated needles and embryos equilibrated in 4% (w/v) Ficoll PM-400 (Sigma) in 5% MMR or 10% NAM. Microinjection needles were generated from borosilicate glass capillaries (Harvard Apparatus, GC120-15) using the micropipette puller Sutter p97. Maximally two nanolitres of morpholino (MO) and/or mRNA were injected into de-jellied embryos at the 1-cell, 2-cell or 4-cell stage using the microinjector Narishige IM-300. Injections for the mVegT loss-of-function (LOF) and rescue were targeted to the vegetal hemisphere. All other injections were targeted to the animal hemisphere. Embryos were transferred to fresh 5% MMR or 10% NAM (without Ficoll) once they reached about blastula stage.

The following mRNA amounts were used for ectopic protein expression: 80 pg and 400 pg *MyoD-HA* mRNA into *X. tropicalis* and *X. laevis* embryos, respectively; and 400 pg *Sox3-HA* into *X. laevis* embryos.

The following MO amounts and treatments were used for LOF experiments: 5 ng *Pou5f3.2* MO, 5 ng *Pou5f3.3* MO and 5 ng standard control MO (mPouV LOF); 5 ng *Pou5f3.2* MO, 5 ng *Pou5f3.3* MO and 5 ng *Sox3* MO (mPouV/Sox3 LOF); 5 ng *β-catenin* MO (Wnt LOF); 10 ng m*VegT* MO (mVegT LOF); standard control MO (5-10 ng according to the dose used for the β-catenin or mVegT LOF experiment); and 30 pg α-amanitin (Sigma). To block Nodal (Nodal LOF) or BMP (BMP LOF) signaling, embryos were treated with 100 µM SB431542 (Tocris, #1614) or 10 µM LDN193189 (Selleckchem, #S2618) from the 8-cell stage onwards. Control embryos were treated accordingly with DMSO, in which these antagonists were dissolved. MOs were designed and produced by Genetools (Oregon, USA) to block translation. β-catenin MO (5'-TTTCAAC<u>A</u>GTTTCCAAAGAACCAGG-3', underlined base was changed from the original MO[25]), Pou5f3.2 MO (5'-GCTGTTGGCTGTACAT<u>A</u>GTGTC-3', underlined base was changed from the original MO[24]), Pou5f3.3 MO[24] (5'-GTACAGAACGGGTTGGTCCATGTTC-3'), Sox3 MO (5'-GTCTGTGTCCAACATGCTATACATC-3'), Tbx6 MO (5'-TACATTGGGTGCAGGGACCCTCTCA-3') and mVegT MO[48] (5'-TGTGTTCCTGACAGCAGTTTCTCAT-3'; see **Supplementary Fig. 1h**). For the rescue of mVegT LOF, 25-50 pg z*VegT* mRNA were injected at the 2- or 4-cell stage. For the rescue of mPouV/Sox3 LOF, 75 pg *Pou5f3.3* mRNA and 75 pg *Sox2* mRNA were injected at the 2- or 4-cell stage. For both rescue experiment, 150-300 pg *mCherry* mRNA was co-injected to trace mRNA expression in embryonic cells.

RNA constructs. All constructs are products of Gateway (Invitrogen) cloning. First, the coding sequences of interest were amplified from either plasmid using *Pfu*Turbo DNA polymerase (Agilent, #600250) or cDNA using Phusion HF DNA polymerase (NEB, #M0530), and unidirectionally cloned into the pENTR/D-TOPO vector (Thermo Fisher Scientific, #K2400). Next, these clones were LR-recombined with Gateway-compatible pCS2+ vectors containing no tag or a C-terminal 3xHA tag (see **Supplementary Fig. 4d**). The following forward and reverse primers were used for PCR amplification: *X. laevis myoD1a* from *MyoD*2-24 pBluescript plasmid[49], 5'-CACCATGGAGCTGTTGCCCCCAC-3', 5'-TAAGACGTGATAGATGGTGCTG-3'; *X. laevis sox3* from the IMAGE:443920 clone (European *Xenopus* Resource Centre), 5'-CACCATGTATAGCATGTTGGACAC-3', 5'-TATGTGAGTGAGCGGTACAGTG-3'; *X. laevis sox2* from stage 11 cDNA, 5'-CACCATGTACAGCATGATGGAGACCGA-3', 5'-TCACATGTGCGACAGAGGCAGC-3'; *X. laevis pou5f3.3* from stage 11 cDNA, 5'-CACCATGGACCAGCCCATATTGTACA-3', 5'-TCAGCCGGTCAGGACCCCCA-3'; *X. tropicalis* z*VegT* from stage 11 cDNA, 5'-CACCATGCACTCTCTGCCGGATGTA-3', 5'-TTACCAACAGCTGTATGGAAAGA-3'. pCS2+8N*mCherry*[50] (#34936) was obtained from Addgene.

***In vitro* transcription of capped mRNA.** About 10 µg plasmid was linearized by restriction digestion and purified using the QIAquick PCR purification kit (Qiagen, #28104). Capped mRNA was produced from ~1 µg linearized plasmid using mMessage mMachine SP6 (Thermo Fisher Scientific, #AM1340) according to the manufacturer's instructions. The following restrictions enzymes (NEB) were used for linearizing the pCS2+ constructs: *Asp*781 (*myoD1a*, *sox3*), *Pvu*II (*sox2*), *Psp*OMI (*pou5f3.3*, z*VegT*) and *Not*I (*mCherry*).

Generation of digoxigenin-labeled RNA probes. DNA template for generating the *foxb1 in situ* hybridization probe was amplified from *X. tropicalis* stage 19-20 cDNA. KAPA HiFi HotStart polymerase (Kapa Biosystems, #KK2602) and the following PCR cycling conditions were used: 98ºC for 45 sec followed by 40 cycles (98ºC for 10 sec, 63ºC for 10 sec, 72ºC for 15 sec) and final elongation step of 20 sec at 72ºC. Primers were 5'-GTCAGCGCCTATGGAGTACC-3' and 5'-AACACTGGAGATGCCATGC-3'. Fresh PCR product was zero-blunt cloned into the pCRII-TOPO vector (Thermo Fisher Scientific, #K2800). The identity and direction of insert was verified by restriction digest and Sanger sequencing. Plasmids were linearized by restriction digestion and purified using the QIAquick PCR purification kit (Qiagen, #28104). All *in situ* hybridization probes were transcribed from ~1 µg linearized plasmid using 1x digoxigenin-11-UTP (Roche, #11277065910), 40 U RiboLock RNase inhibitor (Thermo Fisher Scientific, #EO0381), 1x transcription buffer (Roche) and T7 RNA polymerase (Roche, #10881767001) at 37ºC for 2 h. The probe was treated with 2 U Turbo DNase (Thermo Fisher Scientific, #AM2238) to remove the DNA template and purified by spin-column chromatography (Clontech). RNA was diluted to 10 ng/µl (10x stock) with hybridization buffer (50% formamide, 5x SSC, 1x Denhardt's, 10 mM EDTA, 1 mg/ml torula RNA, 100 µg/ml heparin, 0.1% Tween-20 and 0.1% CHAPS) and stored at -20ºC. The following plasmids and restriction enzymes (NEB) were used for plasmid linearization: *X. laevis eomes* pCRII-TOPO[19], *Bam*HI; *X. tropicalis foxb1* pCRII-TOPO, *Bam*HI; *X. tropicalis mir427* pCS108 (IMAGE:7545411)[41], *Sal*I; *X. laevis ventx2.1* pBluescript SK-[51], *Eco*RI; *X. laevis brachyury* pSP73[52], *Bgl*II; and *X. tropicalis zic1* pCMV-SPORT6 (IMAGE:7668846), *Sal*I.

Animal cap assays. These assays were carried out for the ectopic expression of mRNA constructs in *X. laevis* and for the MO-mediated LOF in *X. tropicalis*. All animal caps were dissected at the blastula stage (stage 8 to 9) and cultured in 75% NAM containing 0.1% bovine serum albumin (Sigma). Dissections were carried out with 13 µm loop electrodes (Protech International, 13-Y1) connected to a MC-2010 micro cautery instrument (Protech International) operating at power level 2. As illustrated in **Supplementary Fig. 4d**, control (uninjected) and MyoD-HA or Sox3-HA expressing animal caps were cultured at 20ºC until sibling embryos reached early neurula stage (stage 13). For the experiment shown in **Fig. 4d** and **Supplementary Fig. 5h**, animal caps of control and mPouV/Sox3 LOF embryos were cultured with or without 10 ng/ml recombinant human activin A (Nodal agonist; R&D Systems, #338-AC) or 50 µM CHIR99021 (canonical Wnt agonist; Tocris, #4423) at 22ºC for ~2 h until sibling embryos reached early gastrula stage (stage 10^+).

Extraction of total RNA. 10-15 embryos (or 15-20 animal caps) were homogenized in 800 (400) µl TRIzol reagent (Thermo Fisher Scientific, #15596018) by vortexing. The homogenate was either snap-frozen in liquid nitrogen and stored at -80ºC or processed immediately. For phase separation, the homogenate together with 0.2x volume of chloroform was transferred to pre-spun 1.5-ml Phase Lock Gel Heavy microcentrifuge tubes (VWR), shaken vigorously for 15 sec, left on the bench for 2 min and spun at ~16,000 g (4ºC) for 5 min. The upper phase was mixed well with one volume of 95-100% ethanol and spun through the columns of Zymo RNA Clean and Concentrator 5 or 25 (Zymo Research) at ~12,000 g for 30 sec. Next, the manufacturer's instructions were followed for the recovery of total RNA (>17 nt) with minor modifications. First, the flow-through of the first spin was re-applied to the column. Second, the RNA was treated in-column with 3 U Turbo DNase (Thermo Fisher Scientific, #AM2238). Third, the RNA was eluted twice with 25 µl molecular-grade water. The concentration was determined on the NanoDrop 1000 spectrophotometer.

Reverse transcription (RT). About 0.2-1 µg total RNA was denatured at 75ºC for 5 min before setting up the RT reaction including 40 U M-MLV (RNase H minus, point mutant) RT (Promega, #M3681), 500 µM of each dNTP (NEB), and 10 µM random hexamers (Sigma) in a final volume of 10 µl. The following incubation settings were used: 25ºC for 15 min, 37ºC for 15 min, 55ºC for 45 min, 85ºC for 15 min and 4ºC for indefinite. The reaction was diluted with 90-190 µl molecular-grade water before using 1-2.5 µl for each qPCR reaction.

qPCR primers. The PCR primers were designed to hybridize at ~60ºC (T_m) and to generate 75-125 bp amplicons. Where possible, primers were checked for their specificity *in silico* using Jim Kent's PCR tool (http://genomes.crick.ac.uk/cgi-bin/hgPcr). All primer pairs were tested to produce one amplicon of the correct size through diagnostic PCR using RedTaq DNA polymerase (Sigma, #D4309). Exceptions were previously designed primer pairs for *X. laevis actc1*, *odc1* and *otx2*, all of which generate amplicons of 200-250 bp.

Primers (forward, reverse) for *X. laevis* (L) or *X. tropicalis* (T) cDNA: *actc1* (L, 5'-TCCTGTACGCTTCTGGTCGTA-3', 5'-TCTCAAAGTCCAAAGCCACATA-3'); *eomes* (T, 5'-ATTGACCACAACCCATTTGC-3', 5'-TCTAGGAGAATCCGCAGGAG-3'); *foxb1* (T, 5'-CCGGATGCAGAGTGACAAC-3', 5'-GAGGCTTCTGGTCGCTGTAG-3'); *gata4* (T, 5'-TCCTCTGATCAAGCCTCAAAG-3', 5'-ACGGCGCCAGAGTGTAGTAG-3'); *myl1* (L, 5'-GCACCTGCATTTGATCTGTC-3', 5'-TTGCTGTCTCCTGTCCTGTC-3'); *nodal5* (T, 5'-TTCAAGCCAACAAACCACGC-3', 5'-CGCATCTTCACCGGTATGCA-3'); *nodal6* (T, 5'-CGCATTTCGTTGTGAGGGGT-3', 5'-TCGTACAGCTTGACCAAACTCT-3'); *odc1* (L, 5'-GCCATTGTGAAGACTCTCTCC-3', 5'-TTCGGGTGATTCCTTGCCAC-3'); *odc1* (T, 5'-GGGCAAAAGAGCTTAATGTGG-3', 5'-CATCGTGCATCTGAGACAGC-3'); *otx2* (L, 5'-TTGAACCAGACCTGGACT-3', 5'-CGGGATGGATTTGTTGCA-3'); *rpl8* (T, 5'-CTGGCTCCAAGAAGGTCATC-3', 5'-CAGGATGGGTTTGTCAATACG-3'); *sox2* (L, 5'-ACATCACATGGACGGTTGTC-3', 5'-TCTCTCGGTTTCTGCTCTCG-3'); *sox17b* (T, 5'- GTTTATGGTGTGGGCAAAGG-3', 5'-TCAGCGATTTCCATGATTTG-3'); *tbx6* (L, 5'-GGGGCACCACGACTCTTAC-3', 5'-CCCAGAATCCTCTGTGAAGG-3'); and *brachyury* (T, 5'-CCTGTGGATGAGGTTCAAGG-3', 5'-CACGCTCACCTTTAGAACTGG-3').

Primers (forward, reverse) for *X. tropicalis* genomic DNA: *β-actin* -0.5kb (5'-TTGTACGCAACACAATGCAG-3', 5'-AAATGTCAGGACCAAATGCAG-3'); *dlc* -3.2kb (5'-CATCCTGCAATTAACTGTCTGC-3', 5'-AGGTGTGACAAATGGCACTG-3'); *gdf3* -0.1kb (5'-TGGATGAGCACAGAGAGGTG-3', 5'-GTCAGGAGGGAGGTGTCAAC-3'); *hoxd1* +3.2kb (5'-TGTTGTAGATGCTGATGCTTATCG-3', 5'-AACAGAAAATCAAAGGCTTGCA-3'); *mespa* -1.3kb (5'-CTCCTTTGCCCTCTGAAATG-3', 5'-CTTCAGGGAATTGGCTTGTG-3'); *msgn1* -0.1kb (5'-CCAGTCCATTTTCCATGTTG-3', 5'-GGGGAGGCCCTTTTATACAG-3'); *myf5* -1.1kb (5'-AAAGAGTTCCTGCACCTTGG-3', 5'-TCCCAAATCATCAGCACAAC-3'); *nodal6* -0.1kb (5'-ACATGAGTTTGGCCTTGGTC-3', 5'-AGGTAACCCCTGAGCTGTCA-3'); *pax3* 5'UTR (5'-CAGATGCCAGAGGAGAGAGC-3', 5'-GGAACTCTCTGGGTCCCAAC-3'); *pax3* +2.7kb (5'-ATGAGAATGTGAGCGACACG-3', 5'-AAACCCAAGGTGAGCAAGTG-3'); *sia1* -0.2kb (5'-TGGCTTTTAAGTTCGTCTTGG-3', 5'-CAAAGGGCTTCTCTTTGTGC-3'); *sox3* TSS (5'-GATGGGGCATTAATGGACAG-3', 5'-CACGGGAGAAAATGAAGCTG-3'); *sox17b* -0.1kb (5'-GTTAGACTCCGCTCCAGACG-3', 5'-TCTCCACCTAAGGGGAAACC-3'); *wnt8a* -5.7kb (5'-GGAAACTGCCTGTGGAAATG-3', 5'-ACAAAAGCCCACAGGAGATG-3'); *brachyury* -4kb (5'-CAATGGATGCCTCTCTGGAC-3', 5'-TGGGTTTAGAATGGCAAAGG-3'); and *brachyury* TSS (5'-CACACCTTGGGTTTTGTTCC-3', 5'-TTCCTCAAGAGAGGGGCTTC-3').

Whole-mount *in situ* hybridization (WMISH). WMISH was conducted using digoxigenin-labeled RNA probes. It was based on previously published protocols[46,53]. *X. tropicalis* embryos were fixed in MEMFA (1x MEM and 3.7% formaldehyde) at room temperature for 1 h. The embryos were then washed once in 1x PBS and two to three times in ethanol. Fixed and dehydrated embryos were kept at -20ºC for ≥24 h to ensure proper dehydration before starting hybridization. Dehydrated embryos were washed once more in ethanol before rehydrating them in two steps to PBT (1x PBS and 0.1% Tween-20). Embryos were treated with 5 µg/ml proteinase K (Thermo Fisher Scientific, #AM2548) in PBT for 6-8 min, washed briefly in PBT, fixed again in MEMFA for 20 min and washed three times in PBT. Embryos were transferred into baskets, which were kept in an 8x8 microcentrifuge tube holder sitting inside a 10x10 slot plastic box filled with PBT. Baskets were built by replacing the round bottom of 2-ml microcentrifuge tubes with a Sefar Nitex mesh. This container system was used to readily process several batches of embryos at once. These baskets were maximally loaded with 40 to 50 *X. tropicalis* embryos. The microcentrifuge tube holder was used to transfer all baskets at once and to submerge embryos into subsequent buffers of the WMISH protocol. Next, the embryos were incubated in 500 µl hybridization buffer (50% formamide, 5x SSC, 1x Denhardt's, 10 mM EDTA, 1 mg/ml torula RNA, 100 µg/ml heparin, 0.1% Tween-20 and 0.1% CHAPS) for 2 h in a hybridization oven set to 60ºC. After this pre-hybridization step, the embryos were transferred into 500 µl digoxigenin-labeled probe (1 ng/µl) preheated to 60ºC and further incubated overnight at 60ºC. The pre-hybridization buffer was kept at 60ºC. The next day embryos were transferred back into the pre-hybridization buffer and incubated at 60ºC for 10 min. Subsequently, they were washed three times in 2x SSC/0.1% Tween-20 at 60ºC for 10 min, twice in 0.2x SSC/0.1%

Tween-20 at 60ºC for 20 min and twice in 1x maleic acid buffer (MAB) at room temperature for 5 min. Next, the embryos were treated with blocking solution (2% Boehringer Mannheim blocking reagent in 1x MAB) at room temperature for 30 min, and incubated in antibody solution (10% lamb or goat serum, 2% Boehringer Mannheim blocking reagent, 1x MAB and 1:2,000 Fab fragments from polyclonal anti-digoxigenin antibodies conjugated to alkaline phosphatase) at room temperature for 4 h. The embryos were then washed four times in 1x MAB for 10 min before leaving them in 1x MAB overnight at 4ºC. On the final day of the WMISH protocol, the embryos were washed another three times in 1x MAB for 20 min and equilibrated to working conditions of alkaline phosphatase (AP) for a total of 10 min by submerging embryos twice into AP buffer (50 mM $MgCl_2$, 100 mM NaCl, 100 mM Tris-HCl pH 9.5 and 1% Tween-20). At this stage, the embryos were transferred to 5-ml glass vials for monitoring the progression of the AP-catalyzed colorimetric reaction. Any residual AP buffer was discarded before adding 700 µl staining solution (AP buffer, 340 µg/ml nitro-blue tetrazolium chloride, 175 µg/ml 5-bromo-4-chloro-3'-indolyphosphate). The colorimetric reaction was developed at room temperature in the dark. Once the staining was clear and intense enough, the color reaction was stopped by two washes in 1x MAB. To stabilize and preserve morphological features, the embryos were fixed with Bouin's fixative without picric acid (9% formaldehyde and 5% glacial acetic acid) at room temperature for 30 min. Next, the embryos were washed twice in 70% ethanol/PBT to remove the fixative and residual chromogens. After rehydration to PBT in two steps, the embryos were treated with weak Curis solution (1% H_2O_2, 0.5x SSC and 5% formamide) at 4ºC in the dark overnight. Finally, the embryos were washed twice in PBS before imaging them in PBS on a thick agarose dish by light microscopy.

Antibody generation. Affinity-purified rabbit polyclonal antibodies were generated by sdix (Delaware, USA) against different amino acid (aa) sequences of *X. tropicalis* Tbx6 avoiding its highly conserved DNA binding domain (aa 95-280) among T-box proteins: aa 1-94 (#4596), aa 304-403 (#5061) and aa 418-517 (#5107). Antibody #5107 did not work (data not shown). Antibodies #4596 and #5061 were suitable for immunoprecipitation and Western blotting to detect endogenous Tbx6 in both *X. tropicalis* and *X. laevis* embryos (**Supplementary Fig. 1o-q**). Antibody #4596 complied with ENCODE guidelines[54] and thus was suitable for ChIP (**Supplementary Fig. 1r**).

Immunoprecipitation (IP). For each IP, 40 de-jellied embryos were collected at the developmental stage of interest. They were homogenized in 400 µl ice-cold IP buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 1 mM EDTA and 0.25 % Igepal CA-630) supplemented with an EDTA-free protease inhibitor cocktail (Roche, #11873580001). The extract was kept on ice for 5 min. Subsequently, the extract was mixed with 400 µl 1,1,2-trichloro-1,2,2-trifluoroethane (Sigma) by vigorously inverting the tube and cleared by centrifugation at 16,000 g (4ºC) for 5 min. An aliquot of 20 µl supernatant was saved as input. The remaining supernatant was incubated with 0.5 µl antibody (anti-Sox3[40], anti-VegT[15] or anti-Tbx6; all rabbit polyclonal antibodies) at 4ºC for 1 h. Subsequently, each embryonic extract was supplemented with 20 µl washed magnetic beads conjugated to sheep anti-rabbit IgG (Dynabeads M280, Invitrogen) and kept on a tube rotator at 4ºC for another 2 h. Next, the beads were washed three times with IP buffer. The immunoprecipitate was eluted off the beads with 20 µl 1x SDS loading buffer (50 mM Tris-HCl pH 6.8, 2% SDS, 6% glycerol and 5% β-mercaptoethanol). Western blotting was carried out with TrueBlot (Rockland Immunochemicals) and normal horseradish peroxidase (HRP)-conjugated secondary antibodies for IP and input samples, respectively.

Western blotting. Protein samples were denatured in 1x (final concentration) SDS loading buffer at 70ºC for 5 min. Denatured samples were run alongside a standard protein ladder into pre-cast gradient gels (Any kD Mini-PROTEAN TGX, BioRad) at constant 200 V for 1 h. Size-separated proteins were immediately transferred onto hydrophobic Immobilon-P PVDF transfer membranes (Millipore) at constant 100 V for 30 min using standard protein electrophoresis equipment. The membranes were blocked with 5% milk powder in PBS/0.1% Tween-20 (MPBTw) at room temperature for 30 min. Next, the membranes were incubated at room temperature for 1 h with the primary antibodies diluted in MPBTw: 1:5,000 of mouse monoclonal anti-α-tubulin (Sigma, #T5168), 1:2,000 rabbit polyclonal anti-Sox3[40], 1:2,000 rabbit polyclonal anti-VegT[15] or 1:2,000 rabbit polyclonal anti-Tbx6 (#5061). The membranes were washed three times with PBST for 10 min before applying 1:1,000 TrueBlot HRP-conjugated anti-rabbit IgG (Rockland Immunochemicals, #18-8816-31) or 1:2,000 normal goat HRP-conjugated anti-mouse IgG (Thermo Fisher Scientific, #31430) in MPBTw at room temperature. After 1 h, the membranes were washed three times in PBTw for 10 min and once in PBS for 5 min. Finally, the membranes were treated with UptiLight US HRP WB reagent (interchim, #58372) for 1 min in the dark. HRP-catalyzed chemiluminescence was detected on the ChemiDoc XRS+ system (BioRad).

Whole-mount immunohistochemistry (WMIHC). Embryos were fixed with MEMFA (0.1 M MOPS, 2 mM EGTA, 1 mM $MgSO_4$ and 3.7% formaldehyde) in capped 5-ml glass vials at room temperature for 2 h, dehydrated and stored in absolute ethanol at -20ºC for ≥24 h. For sagittal sections, fixation was stopped after 30 min. Embryos were placed on a piece of flattened Blu-Tack (Bostik) in a large Petri dish filled with PBS to bisect them with a scalpel. Following a three-step rehydration (50%, 75% and 100% PBS) the embryos were bleached with weak Curis solution (1% H_2O_2, 0.5x SSC and 5% formamide) on a light box at room temperature for 2 h or in the dark at 4ºC overnight. Bleached embryos were washed three times with PBS/0.3% Triton X-100 (PBT) and pre-incubated in blocking solution (PBS, 20% goat or donkey serum, 2% BSA and 0.1% Triton X-100) at 4ºC for 6 h. The embryos were then transferred to 2-ml round-bottom microcentrifuge tubes before discarding all remaining blocking solution. The embryos were incubated at 4ºC for 1-3 days in 50 µl blocking solution containing the primary antibody at the following dilutions: 1:1,000 rabbit polyclonal anti-Sox3[40], 1:500 rabbit polyclonal anti phospho-Smad1/5/8 (Cell Signaling, #9511) or 1:500 goat polyclonal Smad2/3 (R&D Systems, #AF3797). Afterwards, the embryos were transferred back to capped 5-ml glass vials, washed three times with RIPA buffer (50 mM HEPES pH 7.5, 0.5 M LiCl, 1 mM EDTA, 1% Igepal CA-630 and 0.7% sodium deoxycholate) for 1 h and once with PBT for 5 min. Next, the embryos were incubated with the secondary HRP-conjugated antibody diluted in blocking solution at 4ºC overnight: 1:400 goat anti-rabbit IgG-HRP (Thermo Fisher Scientific, #G-21234) or 1:200 donkey anti-goat IgG-HRP (Santa Cruz Biotechnology, #sc-2020). The washes of the previous day were repeated before pre-incubating the embryos in 400 µl inactive 3,3'-diaminobenzidine tetrahydrochloride with cobalt (Sigma, #D0426). After 1-2 min, the inactive solution was replaced with the H_2O_2-activated DAB solution. The HRP reaction was stopped after 40 sec by washing the embryos several times with PBT. In **Supplementary Fig. 1g**, embryos were dehydrated in three steps to absolute methanol and cleared with Murray's clear (2:1 benzyl benzoate/benzyl alcohol) on a glass depression slide.

Deep sequencing and quality filter. All deep sequencing libraries were quality controlled: The DNA yield and fragment size distribution were determined by fluorometry and chip-based capillary or polyacrylamide gel electrophoresis, respectively. Libraries were sequenced on the Illumina platforms GAIIx or HiSeq by the Advanced Sequencing Facility of the Francis Crick Institute to produce single or paired-end reads of at least 40 bases. Next-generation capture-C libraries were sequenced on MiSeq with a read length of 150 bases. Sequencing samples and read alignment results are summarized in **Supplementary Table 1**. The metrics of paired-end alignments such as insert size mean and standard deviation were determined by Picard (CollectInsertSizeMetrics) from the Broad Institute (USA).

Chromatin immunoprecipitation (ChIP). ChIP was carried out as detailed previously[55]. Briefly, de-jellied *X. tropicalis* embryos were treated with 1% formaldehyde (Sigma, #F8775) in 1% MMR at room temperature for 15-45 min to cross-link chromatin proteins to nearby genomic DNA. Duration of

fixation was determined empirically and depended mainly on the developmental stage and antibody epitopes[55]. Fixation was terminated by rinsing embryos three times with ice-cold 1% MMR. Where required, post-fixation embryos were dissected to select specific anatomical regions in ice-cold 1% MMR. Fixed embryos were homogenized in CEWB1 (10 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 1% Igepal CA-630, 0.25% sodium deoxycholate and 0.1% SDS) supplemented with 0.5 mM DTT, protease inhibitors and, if using phospho-specific antibodies, phosphatase blockers (0.5 mM orthovanadate and 2.5 mM NaF). To solubilize yolk platelets and separate them from the nuclei, the homogenate was left on ice for 5 min and then centrifuged at 1,000 g (4ºC) for 5 min. Homogenization and centrifugation was repeated once before resuspending the nuclei containing pellet in 1-3 ml CEWB1. Nuclear chromatin was solubilized and fragmented by isothermal focused or microtip-mediated sonication. The solution of fragmented chromatin was cleared by centrifuging at 16,000 g (4ºC) for 5 min. Where required, ~1% of the cleared chromatin extract was set aside for the input sample (negative control). ChIP-grade antibodies were used to recognize specific chromatin features and to enrich these by coupling the antibody-chromatin complex to protein G magnetic beads (Thermo Fisher Scientific, #10003D) and extensive washing. These steps were carried out at 4ºC. The beads were washed twice in CEWB1, twice in WB2 (10 mM Tris-HCl pH 8.0, 500 mM NaCl, 1 mM EDTA, 1% Igepal CA-630, 0.25% sodium deoxycholate and 0.1% SDS), twice in WB3 (10 mM Tris-HCl pH 8.0, 250 mM LiCl, 1 mM EDTA, 1% Igepal CA-630 and 1% sodium deoxycholate) and once in TEN (10 mM Tris-HCl pH 8.0, 150 mM NaCl and 1 mM EDTA). ChIP was eluted off the beads twice with 100 µl SDS elution buffer (50 mM Tris-HCl pH 8.0, 1 mM EDTA and 1% SDS) at 65ºC. ChIP eluates were pooled before reversing DNA-protein cross-links. Input (filled up to 200 µl with SDS elution buffer) and ChIP samples were supplemented with 10 µl 5 M NaCl and incubated at 65ºC for 6-16 h. Samples were treated with proteinase K (Thermo Fisher Scientific, #AM2548) and RNase A (Thermo Fisher Scientific, #12091021) to remove any proteins and RNA from the co-immunoprecipitated DNA fragments. The DNA was purified with phenol:chloroform:isoamyl alcohol (25:24:1, pH 7.9) using 1.5-ml Phase Lock Gel Heavy microcentrifuge tubes (VWR) for phase separation and precipitated with 1/70 volume of 5 M NaCl, 2 volumes of absolute ethanol and 15 µg GlycoBlue (Thermo Fisher Scientific, #AM9516). After centrifugation, the DNA pellet was air-dried and dissolved in 11 µl elution buffer (10 mM Tris-HCl pH 8.5). The DNA concentration was determined on a fluorometer using high-sensitivity reagents for double-stranded DNA (10 pg/µl to 100 ng/µl).

The following antibodies were used for ChIP: rabbit polyclonal anti-β-catenin (Santa Cruz Biotechnology, #H-102), rabbit polyclonal anti-H3K4me1 (Abcam, #ab8895), rabbit polyclonal anti-HA (Abcam, #ab9110), mouse monoclonal anti-RNAPII (Covance, #MMS-126R), rabbit polyclonal anti-phospho-Smad1/5/8 (Cell Signaling, #9511), goat polyclonal anti-Smad2/3 (R&D Systems, #AF3797), rabbit polyclonal anti-Sox3[40], rabbit polyclonal anti-Tbx6 (#4596) and rabbit polyclonal anti-VegT[15].

Quantitative PCR (qPCR). DNA was amplified in technical duplicates with SYBR Green I master mix (Roche, #04707516001) on a Light Cycler 480 II (Roche) cycling 55-times between 94ºC, 60ºC and 72ºC with each temperature step running for 10 sec and switching at +4.8ºC/sec and -2.5ºC/sec. At the end qPCR reactions were heated from 65ºC to 97ºC with a gradual increase of 0.11ºC/sec (melting curve) to ensure only fluorescence was collected from one specific amplicon. ChIP-qPCR results were based on absolute quantification using an eight-point standard curve of three-fold dilutions of ~1% input DNA (**Supplementary Fig. 1k-n,r**). RT-qPCR results were normalized to the housekeeping gene *odc1* (**Fig. 4d** and **Supplementary Figs. 4e** and **5h**) or *rpl8* (**Supplementary Fig. 5b**), and scaled relative to control embryos using the $2^{-\Delta\Delta Ct}$ method[56]. The threshold cycle (C_t) was derived from the maximum acceleration of SYBR fluorescence (second derivative maximum method).

ChIP-Seq library preparation. 0.5 to 10 ng ChIP DNA or 5 ng input DNA were used to prepare single (only Tbx6 ChIP-Seq) or indexed paired-end libraries as previously described[55,57,58].

Post-sequencing analysis of ChIP-Seq. Single reads of maximal 50 bases were processed using trim_galore v0.4.2 (Babraham Institute, UK) to trim off low-quality bases (default Phred score of 20, i.e. error probability was 0.01) and adapter contamination from the 3' end. Processed reads were aligned to the *X. tropicalis* genome assembly v7.1 (and v9.1 for some ChIP-Seq data) running Bowtie2 v2.2.9[59] with default settings (**Supplementary Table 1**). Alignments were converted to the HOMER's tag density format[60] with redundant reads being removed (makeTagDirectory -single -tbp 1 -unique -mapq 10 -fragLength 175 -totalReads all). Only uniquely aligned reads (i.e. MAPQ ≥10) were processed. We pooled all input alignments from various developmental stages. This created a comprehensive mappability profile that covered ~400 million unique base pair positions. All chromatin profiles were position-adjusted and normalized to the effective total of 1 million aligned reads including multimappers (counts per million aligned reads, CPM). Agreeing biological replicates according to ENCODE guidelines{Landt:2012cl} were subsequently merged. For stage 10^+ β-catenin and Smad2 ChIP-Seq, external datasets[24,61,62] were used as biological replicates. HOMER's peak caller was used to identify transcription factor binding sites by virtue of ChIP-enriched read alignments (hereafter called peaks): findpeaks -style factor -minDist 175 -fragLength 175 -inputFragLength 175 -fdr 0.001 -gsize 1.435e9 -F 3 -L 1 -C 0.97. This means that both ChIP and input alignments were extended 3' to 175 bp for the detection of significant (0.1% FDR) peaks being separated by ≥175 bp. The effective size of the *X. tropicalis* genome assembly v7.1 was set to 1.435 billion bp, an estimate obtained from the mappability profile. These peaks showed equal or higher tag density than the surrounding 10 kb, ≥3-fold more tags than the input and ≥0.97 unique tag positions relative to the expected number of tags. To detect focal RNAPII recruitment to putative CRMs and avoid calling peaks within broad regions of RNAPII reflecting transcript elongation, the threshold of focal ratio and local enrichment within 10 kb was elevated to 0.6 and 3 (-L 3), respectively. To further eliminate any false positive peaks, we removed any peaks with <0.5 (TFs including signal mediators) or <1 (RNAPII) CPM and those falling into blacklisted regions showing equivocal mappability due to genome assembly errors, gaps or simple/tandem repeats. Regions of equivocal mappability were identified by a two-fold lower (poor) or three-fold higher (excessive) read coverage than the average detected in 400-bp windows sliding at 200-bp intervals through normalized ChIP input and DNase-digested naked genomic DNA. All identified regions ≤800 bp apart were subsequently merged. Gap coordinates were obtained from the Francis Crick mirror site of the UCSC genome browser (http://genomes.crick.ac.uk). Simple repeats were masked with RepeatMasker v4.0.6[63] using the crossmatch search engine v1.090518 and the following settings: RepeatMasker -species "xenopus silurana tropicalis" -s -xsmall. Tandem repeats were masked with Jim Kent's trfBig wrapper script of the Tandem Repeat Finder v4.09[64] using the following settings: weight for match, 2; weight for mismatch, 7; delta, 7; matching probability, 80; indel probability, 10; minimal alignment score, 50; maximum period size, 2,000; and longest tandem repeat array (-l), 2 [million bp]. The multi-genome sequence conservation track (phastCons) for *X. tropicalis* genome assembly v9.1 was obtained from Xenbase[65]. The following eleven vertebrate species were used to evaluate sequence similarity: *X. tropicalis*, *X. laevis*, *Nanorana parkeri* (High Himalaya Frog), *Fugu rubripes* (Japanese Pufferfish), *Chrysemys picta* (Painted Turtle), *Gallus gallus* (Chicken), *Anolis carolinensis* (Green Anole lizard), *Monodelphis domestica* (Gray Short-tailed Opossum), *Canis lupus familiaris* (dog), mouse and human. CRMs with a phastCons average ≥0.4 were considered 'conserved'.

Detecting zygotic genes by means of RNAPII chromatin profiles. RNAPII and input tag densities were calculated across the gene body in 10 bins of equal size. Gene annotations v7.1 were altered based on a few known zygotic isoforms and some corrections obtained from assembling total and poly(A) RNA[41] from stage 6 to stage 12.5 *de novo* as outlined elsewhere[66]. A few

genes had previously been annotated as gene clusters due to assembly uncertainties. We reduced the annotation of polycistronic *mir427* to the minus arm (scaffold_3b:3516900-3523400) and only monitored *nodal3.5* and *nodal5.3* within their respective gene clusters. Gene bodies with <40% mappability were removed. Here, the threshold of mappability per bin was set at 10% of the input read density averaged across all gene bodies in use. Subsequently, enrichment values were only obtained for all mappable bins by dividing read densities of RNAPII and input. Further, we restricted the analysis to genes for which ≥3 transcripts per million (TPM) could be detected on average over three consecutive time points (i.e. over the developmental time of 1 h) of a high-resolution profile of total RNA[41] from fertilization to after gastrulation (stage 13). Genes were considered activated at a certain developmental stage if ≥0.1 TPM were detected over three consecutive time points within a 2 h window of the developmental stage in question and either of the following RNAPII enrichments covering ≥80% of the mappable gene body were reached: (1) 2.6-fold, (2) 1.8-fold and 1.4-fold at the next or previous stage, (3) 1.4-fold and 1.8-fold at the next or previous stage, or (4) 1.4-fold over three consecutive stages.

Poly(A) RNA-Seq. Libraries were made from ~1 µg total RNA by following the low-sample protocol of the TruSeq RNA sample preparation guide version 2 with a few modifications. First, 1 µl cDNA purified after second strand synthesis was quantified on a fluorometer using high-sensitivity reagents for double-stranded DNA (10 pg/µl to 100 ng/µl). By this stage, the yield was ~10 ng. Second, the numbers of PCR cycles were adjusted to the detected yield of cDNA to avoid products of over-amplification such as chimera fragments: 7 (~20 ng), 8 (~10 ng), 9 (~5 ng) and 12 (~1 ng).

RNA-Seq read alignment. Paired-end reads were aligned to the *X. tropicalis* genome assembly v7.1 using STAR v2.5.3a[67] with default settings (**Supplementary Table 1**) and a revised version of gene models v7.2[41] to improve mapping accuracy across splice junctions. The alignments were sorted by read name using the sort function of samtools v1.3.1[68]. Exon and intron counts (-t 'exon;intron') were extracted from unstranded (-s 0) alignment files using VERSE v0.1.5[69] in featureCounts (default) mode (-z 0). Intron coordinates were adjusted to exclude any overlap with exon annotation. For visualization, genomic BAM files of biological replicates were merged using samtools v1.3.1 and converted to the bigWig format. These genome tracks were normalized to the wigsum of 1 billion excluding any reads with mapping quality <10 using the python script bam2wig.py from RSeQC v2.6.4[70].

Differential gene expression analysis. Differential expression analysis was performed with both raw exon and intron counts excluding those belonging to ribosomal and mitochondrial RNA using the Bioconductor/R package DESeq2 v1.14.1[71]. In an effort to find genes with consistent fold changes over time, p-values were generated according to a likelihood ratio test reflecting the probability of rejecting the reduced (~ developmental stage) over the full (~ developmental stage + condition) model. Resulting p-values were adjusted to obtain false discovery rates (FDR) according to the Benjamini-Hochburg procedure with thresholds on Cook's distances and independent filtering being switched off. Equally, regional expression datasets[42] without time series were subjected to likelihood ratio tests with reduced (~ 1) and full (~ condition) models for statistical analysis. Fold changes of intronic and exonic transcript levels were calculated for each developmental stage and condition using the mean of DESeq2-normalized read counts from biological replicates. Both intronic and exonic datasets were filtered for ≥10 DESeq2-normalized read counts detected at least at one developmental stage in all uninjected or DMSO-treated samples. Gene-specific fold changes were removed at developmental stages that yielded <10 normalized read counts in corresponding control samples. Next, the means of intronic and exonic fold changes were calculated across developmental stages. The whole dataset was confined to 3,318 genes for which at least 50% reductions (FDR <10%) in exonic (default) or intronic counts could be detected in α-amanitin-injected embryos. Regional expression was based on exonic read counts by default unless the intronic fold changes were significantly (FDR <10%) larger than the exonic fold changes (**Supplementary Table 3**).

Analyzing ribosome footprinting and mass-spectrometry data. The ribosome footprinting reads[13] were trimmed 5' by 8 bases and 3' by as many bases overlapping with the adapter sequence 5'-TCGTATGCCGTCTTCTGCTTG-3' from the 5' end. All trimmed reads between 27 to 32 bases in length were aligned first to ribosomal RNA as listed in the SILVA rRNA database[72] using Bowtie v1.0.1[73] with the following parameters: --seedlen 25 (seed length) --seedmms 1 (number of mismatches allowed in the seed) --un (unaligned reads were reported). All non-aligned (rRNA-depleted) reads were mapped to the gene model 6.0 of *X. laevis* using Tophat v2.0.10[74] --no-novel-juncs (spliced reads must match splice junctions of gene model 6.0) --no-novel-indels (indel detetion disabled) --segment-length 15 (minimal length of read fragment to be aligned) –GTF v6.0.gff3 --prefilter-multihits (reads aligned first to whole genome to exclude reads aligning >10 times) --max-multihits 10. Alignment files were converted to the HOMER's tag density format[60] before retrieving reads for each CDS per kilobase and one million mapped reads (rpkm) using HOMER's perl script analyzeRNA.pl. This read count table was merged with a published list of estimated protein concentrations (nM) in the *X. laevis* egg using mass-spectrometry[12] (**Supplementary Table 2**). Chromatin-associated proteins regulating RNAPII-mediated transcription were filtered based on human (version 2.0, 09/2014) and Xenbase-released gene ontology (GO) associations. The genes associated with any of the following GO terms were verified with UniProt (UniProt Consortium, 2015) whether chromatin binding is supported by functional evidence: chromatin (GO:0000785), DNA binding (GO:0003677), DNA-templated regulation of transcription (GO:0006355), sequence-specific DNA binding and TF activity (GO:0003700) and DNA replication (GO:0006260). We separated specific TFs from all other chromatin binding proteins to form three categories: (1) sequence-specific DNA binding factors, (2) other RNAPII-regulating factors and (3) remaining genes, whereby genes associated with DNA repair (GO:0006281) and RNAPIII (GO:0006383) were moved to the third category.

SPI-Based enrichment of small DNase-digested fragments (DNase-Seq). In our hands, the high yolk content in vegetal blastomeres made it impossible to employ ATAC-Seq[11] on whole *Xenopus* embryos before the onset of gastrulation. Therefore, DNase-Seq was adapted to early *Xenopus* embryos with a novel approach to select small DNase-digested DNA fragments. Ultracentrifugation- or gel electrophoresis-mediated size selection[75,76] was replaced by two rounds of solid phase immobilization (SPI) to remove long inaccessible DNA from short accessible DNA. Wide-bore pipette tips were used for the resuspensions and the transfers of samples from the second homogenization step until after SPI to minimize fragmentation of high-molecular DNA. Approximately 250 de-jellied mid-blastula embryos were collected in 2-ml round-bottom microcentrifuge tubes and homogenized in 2 ml ice-cold LB-DNase buffer (15 mM Tris-HCl pH 8.0, 15 mM NaCl, 60 mM KCl, 1 mM EDTA, 0.5 mM EGTA and 0.5 mM spermidine) supplemented with 0.05% Igepal CA-630. This lysate was left on ice for 3 min before centrifuging at 1,000 g (4ºC) for 2 min. The supernatant was discarded without disrupting the pellet. The pellet was gently resuspended in 2 ml ice-cold LB-DNase buffer (without Igepal CA-630) before centrifuging again at 1,000 g (4ºC) for 2 min. After discarding the supernatant, the pellet was resuspended at room temperature in 600 µl LB-DNase buffer (without Igepal CA-630) with 6 mM $CaCl_2$. The sample was distributed equally to two 1.5-ml microcentrifuge tubes, one for probing chromatin accessibility with DNase and the other one for creating a reference profile of DNase-digested naked genomic DNA. Approximately 0.1 U recombinant DNase I (Roche, #04716728001) was added to one aliquot. Both samples were incubated at 37ºC for 8 min before adding 300 µl STOP buffer (0.1% SDS, 100 mM NaCl, 100 mM EDTA and 50 mM Tris-HCl pH 8.0) including 80 µg RNase A (Thermo Fisher Scientific, #12091021), 333 nM spermine and 1 µM spermidine. The tubes were inverted gently to mix samples before incubating them at 55ºC for 15 min. Next, 200 µg proteinase K (Thermo Fisher

Scientific, #AM2548) were added and the tubes were inverted gently to mix the samples again. After 2 h at 55ºC, the digests were transferred to pre-spun 1.5-ml Phase Lock Gel Heavy microcentrifuge tubes (VWR). 600 µl phenol:chloroform:isoamylalcohol (25:24:1, pH 7.9) were added to the digests. The tubes were shaken gently and then centrifuged at 1,500 g (room temperature) for 4 min. The top phase was transferred to a fresh 2-ml microcentrifuge tube and mixed with 60 µl 3 M sodium acetate (pH 5.2) and 1.2 ml absolute ethanol to precipitate the genomic DNA. The precipitation was stored at -20ºC overnight and centrifuged at 16,000 g (4ºC) for 30 min. The supernatant was discarded and the DNA pellet washed by adding 500 µl ice-cold 80% ethanol and centrifuging at 16,000 g (4ºC) for 3 min. The ethanol was discarded and the DNA pellet was air-dried at room temperature for 10 min. After that, the DNA pellet was left on ice to dissolve in 27 µl elution buffer (10 mM Tris-HCl pH 8.5) for 20 min. To remove any residual RNA, 10 µg RNase A were added to the DNA. 1 µl was used to determine the DNA fragment size distribution by gel electrophoresis. On a 0.6% TAE agarose gel, a smear of very high molecular DNA was visible as expected from previous DNase experiments{He:2014cl, Song:2010ib}. Genomic DNA to the amount of 20 untreated mid-blastula embryos was digested with 0.3 U DNase I at 37ºC for 5 min following the same steps as described above, which generated a low-molecular smear of DNA fragments. Next, 70 µl AMPure XP beads (Beckman Coulter, #A63880) per DNase sample were equilibrated to room temperature for 10 min. 22.5 µl AMPure XP beads (0.9x of the volume) were added to the DNA sample without pipetting up and down achieving a final polyethylene glycol concentration of ~9.5%. After 3 min, by which time high-molecular DNA causes beads to coalesce, the tubes were clipped into a magnetic stand for microcentrifuge tubes. After 3 min or until the beads were separated from the supernatant, the latter was transferred to a 96-well microplate (350-µl round wells with V-shaped bases). 47.5 µl elution buffer and 43 µl AMPure XP beads were added sequentially and mixed gently by slowly pipetting up and down. After 3 min, the plate was transferred to a magnetic stand for 96-well plates. Once the beads have completely separated from the suspension, the supernatant was transferred to a pre-spun 1.5-ml Phase Lock Gel Heavy tube. 162 µl elution buffer and 300 µl phenol:chloroform:isoamylalcohol (25:24:1, pH 7.9) were added. The tubes were shaken gently and then centrifuged at 1,500 g (room temperature) for 4 min. The top phase was transferred to a 1.5-ml low-retention microcentrifuge tube and mixed with 30 µl 3 M sodium acetate (pH 5.2) and 900 µl absolute ethanol to precipitate the DNA as outlined above. The DNA pellet was dissolved in 12 µl elution buffer. The concentrations of the DNA samples were determined on a fluorometer using high sensitivity reagents for dsDNA (10 pg/µl to 100 ng/µl). Libraries were generated as previously outlined[55] except that all cleaning steps were executed with 0.2x more AMPure XP beads.

Post-sequencing analysis of DNase-Seq. Single and paired-end reads of maximal 50 bases were processed using trim_galore v0.4.2 from the Babraham Institute (UK) to trim off low-quality bases (default Phred score of 20, i.e. error probability was 0.01) and adapter contamination from the 3' end. Processed reads were aligned to the *X. tropicalis* genome v7.1 and v9.1 using Bowtie2 v2.2.9[59] with default settings apart from -X (fragment length), which was reduced to 250 bp for paired-end reads. Alignments were sorted by genomic coordinates and only those with a quality score of ≥10 were retained using samtools v1.3.1[68]. Duplicates were removed using Picard (MarkDuplicates) from the Broad Institute (USA). Paired-end alignments were dissociated using hex flags (-f 0x40 or 0x80) of samtools view. Single alignments were converted to HOMER's tag density format[60] (makeTagDirectory -single -unique -fragLength 100 -totalReads all). DNase hypersensitive sites were identified using the following or otherwise default parameters of HOMER's peak calling: findpeaks -style factor -minDist 100 -fragLength 100 -inputFragLength 100 -fdr 0.001 -gsize 1.435e9 -F 3 -L 1 -C 0.97. This means that alignments of DNase-digested chromatin fragments (or naked genomic DNA fragments considered here as 'input') were extended 3' to 100 bp from the DNase cleavage site to detect significant (0.1% FDR) DNase-enriched read alignments (hereafter called peaks) being separated by ≥100 bp. The effective size of the *X. tropicalis* genome assembly v7.1 was set to 1.435 billion bp, an estimate obtained from the mappability of ChIP input reads. These peaks showed equal or higher tag density than the surrounding 10 kb, at least three-fold more tags than the input and ≥0.97 unique tag positions relative to the expected number of tags. Peaks falling into blacklisted regions (see Post-sequencing analysis of ChIP-Seq) were removed. The profiles of DNase-probed chromatin accessibility were position-adjusted and normalized to the effective total of 1 million aligned reads including multimappers.

Next-generation capture-C. About 500 mid-blastula embryos per condition were collected in 9-ml capped glass vials and briefly washed once with 1% MMR. The embryos were then fixed with 1% formaldehyde (Sigma, #F8775) in 9 ml 1% MMR at room temperature for 40 min. The fixation reaction was terminated by briefly rinsing the embryos three times with ice-cold 1% MMR. The embryos were aliquoted into 2-ml round-bottom microcentrifuge tubes in batches of ~250 embryos (filling the volume of ~250 µl water). The supernatant was removed before equilibrating embryos in 250 µl ice-cold HEG buffer (50 mM HEPES pH 7.5, 1 mM EDTA and 20% glycerol). Once the embryos settled to the bottom of the tube as much liquid as possible was discarded. The aliquots were snap-frozen in liquid nitrogen and stored at -80ºC.

For each chromatin conformation capture (3C) experiment, 10 ml chromatin extraction buffer CEB-3C (10 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 1% Igepal CA-630, 0.25% sodium deoxycholate and 0.2% SDS) were supplemented with a mini protease inhibitor tablet (Roche, #11873580001) and 0.05 mM DTT (CEB-3C*). Both aliquots of ~250 embryos of each condition were homogenized in 2 ml ice-cold CEB-3C*. The tubes were kept on ice for 5 min before centrifuging at 1,000 g (4ºC) for 2 min. The supernatants were discarded and the pellet resuspended in 0.5 ml ice-cold CEB-3C*. The resuspensions of each condition were pooled and then equally divided to two 50-ml conical tubes and each filled up with CEB-3C (without protease inhibitors and DTT) to 50 ml. The embryonic extracts were incubated at 37ºC for 1 h in a hybridization oven with the tubes rotating inside a hybridization bottle. The tubes were then centrifuged at 1,000 g (room temperature) for 5 min. The supernatants were discarded and the pellets resuspended in 50 ml double-distilled water. The tubes were centrifuged again at 1,000 g (room temperature) for 5 min. The supernatants were discarded and each pellet was resuspended in 300 µl double-distilled water. The resuspensions containing cross-linked nuclei of each condition were pooled in a 1.5-ml microcentrifuge tube. The samples were then digested with 400 U *Dpn*II (NEB, #R0543) in a total volume of 800 µl containing 1x *Dpn*II reaction buffer. The digest was incubated overnight in a thermomixer set to 37ºC and 1,400 rpm. Of note, any residual yolk platelets in the solution did not interfere with the restriction digest. The next day, the digest was supplemented with 200 U *Dpn*II and incubated at 37ºC and 1,400 rpm for another 6-8 h. After that, *Dpn*II was heat-inactivated at 65ºC for 20 min.

To verify the degree of chromatin digestion, an aliquot of 50 µl from each condition was transferred into separate 1.5-ml microcentrifuge tubes and processed as follows: First, any remaining RNA was degraded by incubating the aliquot with 20 µg RNase A (Thermo Fisher Scientific, #12091021) at 37ºC for 30 min. Second, proteins were digested by incubating the aliquot mixed with 50 µl SDS elution buffer (50 mM Tris pH 8.0, 1 mM EDTA and 1% SDS) and 20 µg proteinase K (Thermo Fisher Scientific, #AM2548) at 65ºC in a hybridization oven overnight. Third, the DNA was purified and ethanol precipitated: 300 µl TE pH 8.0 were added to the digest before transferring it to pre-spun 1.5-ml Phase Lock Gel Heavy microcentrifuge tubes (VWR). The diluted digests were mixed vigorously with 400 µl (1 volume) phenol:chloroform:isoamylalcohol (25:24:1, pH 7.9) before centrifuging at 16,000 g (room temperature) for 1 min. The top (aqueous) phases were transferred into 1.5-ml microcentrifuge tubes and mixed with 40 µl sodium acetate (pH 5.2) and 800 µl absolute ethanol and 15 µg GlycoBlue (Thermo Fisher Scientific, #AM9516). The DNA precipitations were kept at -80ºC for 30 min before centrifuging at 16,000 g (4ºC) for 30 min. The supernatants were discarded without disturbing the DNA pellets which were subsequently washed once with 400 µl ice-cold 80% ethanol. The tubes were

centrifuged at 16,000 g (4ºC) for 30 min before discarding the supernatants. The DNA pellets were air-dried for 10 min and resuspended in 10 µl elution buffer (10 mM Tris-HCl pH 8.5). The DNA concentrations were determined with a fluorometer and the DNA fragment distributions were visualized on a 0.6% TAE agarose gel. The digestion of cross-linked chromatin with *Dpn*II caused extensive fragmentation of the genomic DNA such that no high-molecular DNA fragment bands (>10 kb) were visible.

The chromatin digests were processed for proximity ligation by adding 240 U T4 DNA ligase (Thermo Fisher Scientific, #EL0013) in a total volume of 1.2 ml containing 1x T4 DNA ligase buffer including 5 mM ATP. The ligation reactions were incubated at 16ºC and 1,400 rpm for ≥16 h before centrifuging at 16,000 g (4ºC) for 1 min. The supernatants were discarded. 200 µl elution buffer and 40 µg RNase A were added to the pellets to degrade any residual RNA at 37ºC and 1,400 rpm for 30 min. Proteins were digested by adding 200 µl SDS elution buffer and 160 µg proteinase K and incubating at 65ºC in a hybridization oven overnight. The digest was transferred to pre-spun 1.5-ml Phase Lock Gel Heavy tubes and processed as described in the previous paragraph to purify and precipitate ligated genomic DNA. The air-dried DNA pellet was dissolved in 20 µl elution buffer and quality controlled as above. Proximity ligation of *Dpn*II-digested chromatin fragments massively increased the size of DNA fragments to the range of high molecular weight (>10 kb) representing the 3C library. The concentration was measured on a fluorometer with broad range concentration (5-500 ng/µl) reagents.

About 10-15 µg of the 3C libraries were diluted with elution buffer to the total volume of 130 µl and transferred to a designated glass vial (microTUBE) for isothermal focused sonication (Covaris). The following settings of the focused ultrasonicator Covaris S220 were used to shear the libraries to an average DNA fragment length of ~200 bp: duty cycle, 10%; intensity, 5; cycles per burst, 200; duration, 60 sec in frequency sweeping mode. Sonication was run with 6 cycles in batch mode. The degree of DNA fragmentation was verified by gel electrophoresis and the DNA concentration was measured on a fluorometer with broad range concentration reagents.

Approximately 2-4 µg of the DNA fragments were converted into next-generation paired-end libraries using the enzymes of the KAPA Hyper Prep Kit (Kapa Biosystems, #KR0961) and TruSeq adaptors (Illumina). The end-repair and A-tailing reactions were set up according to the manufacturer's instructions and incubated at 20ºC for 1 h followed by 30 min at 65ºC. End-repaired and A-tailed DNA fragments were ligated to 150 pmol TruSeq adaptors of index 6 and 12, respectively, at 20ºC for 1 h. The DNA was purified using 0.8x volume of AMPure XP beads (Beckman Coulter, #A63880) and amplified in a total volume of 100 µl using 50 pmol of each TruSeq PCR primer, KAPA HiFi HotStart ReadyMix (Kapa Biosystems, #KK2602) and the following PCR conditions: 98ºC for 45 sec followed by 3 cycles (98ºC for 15 sec, 98ºC for 30 sec and 72ºC for 30 sec) and a final elongation step of 1 min at 72ºC. The lid temperature was set to 105ºC. The PCR reactions were cleaned up with 100 µl AMPure XP beads and eluted in 14 µl elution buffer. The integrity of the library DNA was verified on a chip-based capillary electrophoresis system, an 8% TBE polyacrylamide gel or a 2% TAE agarose gel. DNA concentration was measured on a fluorometer with broad range concentration reagents. The DNA yield was 2-4 µg.

To capture the genomic regions of interest (viewpoints), 5'-biotinylated oligonucleotides of 120 bases (hereafter called probes) (**Supplementary Table 4**) were designed for each viewpoint as follows: The viewpoints were *Dpn*II fragments of 300-3,300 bp which were overlapping gene promoters or located <1 kb from them. Each probe matched the terminal sequence of the same DNA fragment strand including the *Dpn*II restriction site GATC. The probe sequence was examined using BLAT to check whether it is unique and did not partially match any other genomic regions. Furthermore, CENSOR v4.2.29[77] and RepeatMasker v4.0.6[63] programs were run to discard any probes that contained repeats. Because of these design restrictions, six of the selected 30 gene promoters were captured with only one probe. All the probes were purchased as desalted oligomers (4 nmol) from IDT (ultramer technology) and reconstituted in molecular-grade water to 10 µM. The probes were mixed in equimolar quantities and diluted to 2.9 nM such that 4.5 µl contained 13 fmol of oligomers.

The capture was performed with the SeqCap Hybridization and Wash Kit (Roche, #05634253001) and SeqCap HE-Oligo Kit A (Roche, #06777287001) as follows. Exactly 1 µg of each TruSeq library with index 6 and 12 was mixed with 1 nmol TruSeq universal blocking oligonucleotide and 0.5 nmol blocking oligonucleotides specific to TruSeq index 6 and 12 in a 1.5-ml low-retention microcentrifuge tube. Of note, Cot-1 DNA commonly used to mask repetitive DNA proved to be unnecessary here in reducing non-specific hybridization. The mixture of library and blocking oligonucleotides was dried in a vacuum centrifuge before adding 7.5 µl 2x hybridization buffer (vial 5) and 3 µl hybridization component A (vial 6). This mixture was vortexed for 10 sec, centrifuged at 16,000 g for 10 sec and incubated at 95ºC for 10 min. In the meantime, 4.5 µl (13 fmol) of the equimolar probe mixture were transferred to a PCR tube and incubated at 47ºC in a PCR machine with the volume and the lid temperature set to 15 µl and 57ºC, respectively. Upon denaturation, the libraries and blocking oligonucleotides were added to the probes without removing either tubes from the heating block or the PCR machine. The hybridization mixture was incubated at 47ºC for 64-72 h. The wash buffers were prepared according to the manufacturer's instructions to make 1x working solutions for one single capture experiment. The stringent wash buffer and wash buffer I were pre-warmed to 47ºC and 50 µl M270 streptavidin-conjugated magnetic beads (Thermo Fisher Scientific, #65305) were transferred into a 1.5-ml low-retention microcentrifuge tube to let them equilibrate to room temperature for 10 min. The beads were washed twice with 200 µl bead wash buffer. Immediately after the final wash, the hybridization mixture was directly added to the beads and vortexed for 10 sec. The sample was incubated for 45 min in a thermomixer set to 47ºC and 1,100 rpm. The beads were washed by adding 100 µl pre-warmed wash buffer I and vortexing for 10 sec. The tube was placed into a magnetic rack. Once the liquid was clear, the supernatant was discarded, and 200 µl pre-warmed stringent wash buffer were added to the beads. In an effort to avoid any temperature drop, it is important to work quickly according to the manufacturer's instructions. The beads in stringent wash buffer were incubated in a thermomixer set to 47ºC and 1,100 rpm for 5 min. The wash with pre-warmed stringent wash buffer was repeated once. Next, 200 µl pre-warmed wash buffer I was added to the beads and vortexed at 1,400 rpm (room temperature) for 2 min. After removing the respective wash buffer, beads were vortexed at 1,400 rpm in 200 µl wash buffer II for 1 min and 200 µl wash buffer III for 30 sec. After the final the final wash, as much liquid as possible was removed and the beads were resuspended in 40 µl molecular-grade water.

The captured DNA was directly amplified from the beads using KAPA HiFi HotStart master mix and 25 pmol of each TruSeq PCR primer in two separate 50-µl PCR reactions. The PCR conditions were the same as outlined above except that 14 cycles were used for amplification. The PCR reactions were pooled and the DNA was purified using 100 µl AMPure XP beads. The DNA was eluted with 11 µl elution buffer. The eluted DNA was subjected to another round of probe-mediated capture with the hybridization timespan reduced to 24 h. Furthermore, after washing the beads, the captured fragments were amplified using only 10 cycles of PCR.

Post-sequencing analysis of next-generation capture-C data. The analysis was carried out in accordance with the original description of the method[78] with some modifications. Paired-end reads were processed using trim_galore v0.4.2 from the Babraham Institute (UK) to trim off low-quality bases (default Phred score of 20, i.e. error probability was 0.01) and adapter contamination from the 3' end. Only complete mate pairs were processed further to reconstruct single reads from overlapping paired-end sequences using FLASH v1.2.11[79] with interleaved output settings for non-extended reads (flash --interleaved-output --max-overlap 150). FLASH reads were split in *silico* at *Dpn*II restriction sites using a designated perl script (https://github.com/telenius/captureC/releases) before aligning them to the *X. tropicalis* genome assembly 7.1 using Bowtie2 v2.2.9[59]. The alignment was run with default settings and one thread only to keep the order of the reads. The view function of samtools v1.3.1[68] was used to retain alignments with a quality score of ≥10. The alignments were analyzed further using a suite of perl scripts (https://github.com/Hughes-Genome-Group/captureC/releases)

modified to process both chromosome and scaffold coordinates. Viewpoint coordinates included a 1-kb proximity exclusion range. Restriction fragments were classified as capture, proximity-exclusion or reporter. PCR duplicates were removed. The interaction map was based on the number of unique paired-end reads per restriction fragment. Windows of 2 kb incrementing by 200 bp were used to consolidate interactions, which were normalized to 10,000 interactions per viewpoint.

Analysis of genomic profiles, DNA motifs and super-enhancers. Unless otherwise stated, Spearman correlation (R_s), relative tag density, principal component analysis (PCAs) and enrichment of DNA motifs were determined using a limited number of detected peaks per chromatin feature and developmental stage: top 2,000 peaks (**Fig. 2b** and **Supplementary Fig. 3a-e**), top 5,000 peaks (**Fig. 1d**) and top 10,000 peaks (**Fig. 2e,f,i** and **Supplementary Fig. 4a,b,g-i**). R_s and PCA were calculated using DiffBind v2.4.8[80] in R (**Fig. 2e** and **Supplementary Figs. 1a**, **3a-c** and **4a,g**). Z-scores of motif enrichments at RNAPII peaks were calculated using ISMARA v1.2.2[81] (https://ismara.unibas.ch/fcgi/mara) in expert mode (**Fig. 1d**). Heat maps of relative tag densities were hierarchically clustered (Ward's method) using the R package seriation[82] except for those in **Fig. 5b** and **Supplementary Fig. 4b**, where genomic coordinates were sorted upon significant changes in DNA occupancy or chromatin accessibility (see Figure legends). In **Supplementary Fig. 3e**, cluster leaves were additionally ordered with the Gruvaeus-Wainer's algorithm. The matrices of normalized tag densities, p-values and motif instances were generated with HOMER as follows: annotatePeaks.pl -size 2000 (or 200 for **Supplementary Fig. 4e**) -hist 25 (bin size) -ghist. In heat maps, all genomic profiles were scaled to the means of the highest bin values of the top 100 (or 1,000 in **Fig. 5b** and **Supplementary Fig. 4b**) peaks. In addition, to visualize low RNAPII binding and RNA expression, their top means were divided by 5 or 500, respectively (**Fig. 5b**). Super-enhancers were assembled by stitching together peaks that are ≤25 kb apart: findPeaks -style super -fragLength 175 -inputFragLength 175 -fdr 0.001 -gsize 1.435e9 -F 3 -L 1 -localSize 10000 -C 0.97 -superSlope 1 -minDist 25000. Subsequently, super-enhancers ≤5 kb from TSS apart were associated with zygotic genes defined by RNAPII coverage.

Analysis of enriched gene ontology (GO) terms. Over-represented GO terms were found by applying hypergeometric tests of the Bioconductor/R package GOstats v2.40.0[83] on gene lists. The process was also supported by the Bioconductor/R packages GSEABase v1.36.0[84] and GO.db v3.6.0[85]. The gene universe was associated with GO terms by means of BLAST2GO[86] as previously outlined[58,87].

Differential analysis of chromatin features. The significance of differential chromatin accessibility between control and mPouV/Sox3 LOF embryos was assessed at CRMs using the Bioconductor/R package DiffBind v2.4.8[80] and the DESeq2[71] algorithm. The size of all CRMs was fixed to 500 bp. Only the very 5' end of each aligned DNase-Seq read, which represents the DNase cleavage site, was considered for this analysis. Read counts from mapping DNase-digested naked genomic DNA were subtracted from chromatin accessibility readings at each CRMs. Any changes to DNase-mediated cleavages at CRMs with FDR ≤10% were determined as statistically significant (**Supplementary Fig. 7b**). For visualization, the orientations of CRMs were aligned to the direction of transcription of the closest zygotic gene. Matrices of normalized and raw tag counts were generated at CRMs for various chromatin features and RNA (**Fig. 5b**). Their rows were sorted according to the statistical significance of differential chromatin accessibility. Raw tag counts were processed to estimate p-values of chromatin changes across CRMs using DESeq2 v1.14.1[71]. Read count dispersions were locally fitted. The shrinkage of $\log_2$ fold changes, threshold on Cook's distances and independent filtering were all switched off. For the comparison of local interactions with promoters at affected and unaffected CRMs, the size of the CRMs was widened to 1.2 kb and filtered for ≥5 reporter fragments (**Supplementary Fig. 7f**). Normalized line and bar graph tracks of control and mPouV/Sox3 LOF embryos were superimposed in the IGV genome browser v2.3.92[88] (**Fig. 5f** and **Supplementary Fig. 7g**).

Software for computational analysis and graphical illustration. Post-alignment analysis and graphical illustration were performed using samtools v1.3.1[68], bedtools v2.25.0[89], HOMER v4.8.3[60] and tools in R v3.4.1 / Bioconductor v3.5, Perl v5.18.2 or Python v2.7.12.

Quantification and statistical analysis. No statistical method was used for determining sample size. We followed the literature to select the appropriate sample size. The experiments were not randomized. Due to the nature of experiments, the authors were not blinded to group allocation during data collection and analysis. Only viable embryos and embryonic tissues were included in the analysis. Due to space constraints, data from the trials of establishing and optimizing protocols (ChIP-Seq, DNase-Seq and next-generation capture-C) were excluded. Frequencies of shown morphological phenotypes and WMISH patterns are included in every image. Significances of normally or non-normally distributed data points were calculated using a paired Student's t-test (one-tailed or two-tailed) or paired Wilcoxon test (two-tailed), respectively. The effect size r was estimated from the standard normal deviate of the Wilcoxon p-value (p) as previously described[90], r_{effect}=Z/sqrt(N), where Z=qnorm(1-p/2) is the standardized Z-score and N is the number of observations. Significance of motif enrichments or GO terms are based on hypergeometric tests. The significance of transcriptome-wide differential expression was adjusted for multiple comparisons using the Benjamini-Hochburg procedure.

External datasets. The following external datasets were used for this study: 5mC at stage 9, and ChIP for H3K4me3 and H3K27me3 at stage 9 (GSE67974), β-catenin ChIP at stage 10^+ (GSE72657), Smad2 ChIP at stage 10^+ (GSE30146 and GSE53654), Foxh1 ChIP at stage 8 and 10^+ (GSE85273 and GSE53654), total RNA recorded at 30-min intervals from unfertilized eggs to 23.5 hpf (GSE65785), ribosome footprinting from stage 3, 9 and 12^+ (GSE52809), and regional expression along the animal-vegetal and dorso-ventral axis at stage 10^+ (GSE81458).

Code availability. Programming code used in this paper is available from the authors upon request

45. Khokha, M. *et al.* Techniques and probes for the study of Xenopus tropicalis development. *Dev. Dyn.* **225,** 499–510 (2002).
46. Sive, H., Grainger, R. & Harland, R. Early development of Xenopus laevis: A laboratory manual. *Cold Spring Harbor Laboratory Press* (2000).
47. Nieuwkoop, P. & Faber, J. Normal table of Xenopus laevis (Daudin): a systematical and chronological survey of the development from the fertilized egg till the end of metamorphosis. Garland (1994).
48. Rana, A., Collart, C., Gilchrist, M. & Smith, J. Defining synphenotype groups in Xenopus tropicalis by use of antisense morpholino oligonucleotides. *PLoS Genet.* **2,** e193 (2006).
49. Hopwood, N., Pluck, A. & Gurdon, J. MyoD expression in the forming somites is an early response to mesoderm induction in Xenopus embryos. *Embo J.* **8,** 3409–3417 (1989).
50. Gökirmak, T. *et al.* Localization and substrate selectivity of sea urchin multidrug (MDR) efflux transporters. *J. Biol. Chem.* **287,** 43876–43883 (2012).
51. Ladher, R., Mohun, T. J., Smith, J. C. & Snape, A. M. Xom: a Xenopus homeobox gene that mediates the early effects of BMP-4. *Development* **122,** 2385–2394 (1996).
52. Smith, J., Price, B., Green, J., Weigel, D. & Herrmann, B. Expression of a Xenopus homolog of Brachyury (T) is an immediate-early response to mesoderm induction. *Cell* **67,** 79–87 (1991).
53. Monsoro-Burq, A. H. A Rapid Protocol for Whole-Mount In Situ Hybridization on Xenopus Embryos. *Cold Spring Harb. Protoc.* **2007,** pdb.prot4809 (2007).
54. Landt, S. G. *et al.* ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* **22,** 1813–1831 (2012).
55. Gentsch, G. E. & Smith, J. C. Efficient Preparation of High-Complexity ChIP-Seq Profiles from Early Xenopus Embryos. *Methods Mol. Biol.* **1507,** 23–42 (2017).
56. Livak, K. J. & Schmittgen, T. D. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2CT Method.

Methods **25,** 402–408 (2001).
57. Gentsch, G. E. & Smith, J. C. Investigating physical chromatin associations across the Xenopus genome by chromatin immunoprecipitation. *Cold Spring Harb Protoc* **2014,** pdb.prot080614 (2014).
58. Gentsch, G. E., Patrushev, I. & Smith, J. C. Genome-wide snapshot of chromatin regulators and states in Xenopus embryos by ChIP-Seq. *J. Vis. Exp.* e52535 (2015).
59. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9,** 357–359 (2012).
60. 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 (2010).
61. Nakamura, Y., de Paiva Alves, E., Veenstra, G. J. C. & Hoppler, S. Tissue- and stage-specific Wnt target gene expression is controlled subsequent to β-catenin recruitment to cis-regulatory modules. *Development* **143,** 1914–1925 (2016).
62. Yoon, S.-J., Wills, A. E., Chuong, E., Gupta, R. & Baker, J. C. HEB and E2A function as SMAD/FOXH1 cofactors. *Genes Dev.* **25,** 1654–1661 (2011).
63. Smit, A., Hubley, R. & Green, P. RepeatMasker Open-4.0. 2013-2015.
64. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27,** 573–580 (1999).
65. Karpinka, J. B. *et al.* Xenbase, the Xenopus model organism database; new virtualized system, data types and genomes. *Nucleic Acids Res.* **43,** D756–63 (2015).
66. Pertea, M., Kim, D., Pertea, G. M., Leek, J. T. & Salzberg, S. L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **11,** 1650–1667 (2016).
67. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).
68. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
69. Zhu, Q., Fisher, S. A., Shallcross, J. & Kim, J. VERSE: a versatile and efficient RNA-Seq read counting tool. *bioRxiv* 053306 (2016).
70. Wang, L., Wang, S. & Li, W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28,** 2184–2185 (2012).
71. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** 550 (2014).
72. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41,** D590–6 (2013).
73. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25 (2009).
74. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14,** R36 (2013).
75. He, H. H. *et al.* Refined DNase-seq protocol and data analysis reveals intrinsic bias in transcription factor footprint identification. *Nat. Methods* **11,** 73–78 (2014).
76. Vierstra, J., Wang, H., John, S., Sandstrom, R. & Stamatoyannopoulos, J. A. Coupling transcription factor occupancy to nucleosome architecture with DNase-FLASH. *Nat. Methods* **11,** 66–72 (2014).
77. Kohany, O., Gentles, A. J., Hankus, L. & Jurka, J. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics* **7,** 474 (2006).
78. Davies, J. O. J. *et al.* Multiplexed analysis of chromosome conformation at vastly improved sensitivity. *Nat. Methods* **13,** 74–80 (2016).
79. Magoč, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27,** 2957–2963 (2011).
80. Stark, R. & Brown, G. DiffBind: Differential binding analysis of ChIP-Seq peak data. 1–33 (2018).
81. Balwierz, P. J. *et al.* ISMARA: automated modeling of genomic signals as a democracy of regulatory motifs. *Genome Res.* **24,** 869–884 (2014).
82. Hahsler, M., Hornik, K. & Buchta, C. Getting Things in Order: An Introduction to the R Package seriation. *J. Stat. Softw.* **25** (2008).
83. Falcon, S. & Gentleman, R. Using GOstats to test gene lists for GO term association. *Bioinformatics* **23,** 257–258 (2007).
84. Morgan, M., Falcon, S. & Gentleman, R. *GSEABase: Gene set enrichment data structures and methods* (2017).
85. Carlson, M., Falcon, S., Pages, H. & Li, N. *Bioconductor - GO.db* (2007).
86. Conesa, A. *et al.* Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21,** 3674–3676 (2005).
87. Collart, C. *et al.* High-resolution analysis of gene activity during the Xenopus mid-blastula transition. *Development* **141,** 1927–1939 (2014).
88. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol* **29,** 24–26 (2011).
89. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).
90. Rosenthal, R. Meta-Analytic Procedures for Social Research. SAGE Publishing (1991).

Supplementary information

Supplementary Figure 1 Identification of potential ZGA regulators and competence factors. (**a**) Spearman correlations (R_s) between RNAPII ChIP-Seq profiles at indicated developmental stages. (**b**) Enrichment significance and coverage of DNA motifs at RNAPII$^+$ CRMs from 32-cell to 1000-cell stage (pre-MBT) and from MBT to late gastrula stage (MBT+). (**c**) Validation of our novel DNase-Seq method: Snapshot of DNase-probed chromatin accessibility (biological replicates #1 and #2) and various other chromatin features (Sox3, RNAPII, H3K4me1, poly(A) RNA from this study and H3K4me3, H3K27me3, 5-methylcytosine [5mC] from a previously published study[43]) at the *sox2* locus (gene-centric region and downstream distal regulatory region) around MBT. Negative control of DNase-Seq, DNase-treated naked genomic DNA (gDNA). (**d**) Ribosome footprinting[13] of selected maternal and zygotic TFs across MBT. (**e**) WMIHC of Smad1 and Smad2 protein on bisected early gastrula embryos. Scale bar, 0.5 mm. (**f**) Western blotting of Sox3 protein in control (control MO) and Sox3-depleted (Sox3 MO) embryos (Sox3 LOF) at blastula stage. α-tubulin, loading control. (**g**) WMIHC of Sox3 protein in control and Sox3 LOF embryos at 64-cell and blastula stage. Nuclear accumulation of Sox3 protein detected in both the animal (An) and vegetal (Vg) hemisphere of control embryos. Scale bar, 0.5 mm. (**h**) Ribosome footprinting mapped to the *VegT* locus showing the post-MBT switch of translation from the maternal (m) to the zygotic (z) *VegT* transcript. Design of MO to block translation of the maternal transcript only. (**i**) Western blotting of mVegT protein in control (control MO) and mVegT-depleted (mVegT MO) embryos (mVegT LOF). Dotted line indicates the elimination of irrelevant lanes from the western blot. α-tubulin, loading control. (**j-r**) Validation of ChIP-grade antibodies. (**j**) Sox3 immunoprecipitation (IP). IgG H and L, detected IgG heavy and light chains of the IP antibody. (**k-n,r**) ChIP-qPCR for Sox3, β-catenin, Smad1, Smad2 and Tbx6 at indicated developmental stages. Mean+SD (n=2-3). One-tailed Student's t-test (comparing to IgG control): *, $p \leq 0.1$ and ≥2-fold enrichment relative to lowest DNA recovery (% input) with ChIP antibody (dotted line). (**o**) Tbx6 IP using *X. tropicalis* and *X. laevis* embryonic extract from late gastrula embryos. Antibody #4596 was chosen for subsequent IP and ChIP experiments, and #5061 for Western blotting. (**p**) Western blotting (WB) of Tbx6 in standard control and Tbx6 morphants at early tailbud stage. α-tubulin, loading control. (**q**) Relative expression levels of the transcript (RT-qPCR) and its protein (IP/WB) from early blastula to early tailbud stage. α-tubulin, loading control.

Supplementary Figure 2 Super-enhancers of ZGA are associated with the formation of the three germ layers and the primary body axes. (**a**) Dynamic chromatin engagement of endogenous Sox3, Foxh1[24], VegT, β-catenin, Smad2, Smad1 and RNAPII to super-enhancers of the *siamois* and *ventx* gene cluster from 32-cell to early gastrula stage. Illustrated sagittal sections (dorsal side is right) show the nuclear localization of the selected TFs with arrows pointing to the tissue movements of gastrulation. Arrowheads, promoter and enhancer activity based on the peak-like recruitment of RNAPII and/or TFs. (**b**) Chromatin landscape of endogenous Sox3, Eomes, zVegT, Brachyury and Tbx6 to super-enhancer of the *mesp* gene cluster at late gastrula and/or early tailbud stage. Illustrated sagittal sections (dorsal side is right) show the nuclear localization of selected TFs in late gastrula embryo and at the caudal

end of early tailbud embryo. (**c**) Super-enhancer-associated binding correlates with high levels of chromatin accessibility and DNA occupancy of TFs, signal mediators and RNAPII. Grey box explains the plot composition: (1) the normalized levels of chromatin accessibility or DNA occupancy (±20 kb from TSSs) for each zygotic gene (detected by RNAPII profiling from 32-cell to late gastrula stage, see Methods) and (2) the cumulative frequency of detected super-enhancers. Genes with super-enhancer-associated binding <5kb from TSSs are highlighted in red. (**d**) Significantly enriched biological processes associated with the gene function of nearby super-enhancers detected at indicated developmental stages. (**e**) Meta-summary of chromatin accessibility and DNA occupancy levels of TFs and signal mediators at zygotic genes (mean±SD; $\log_{10}$ scale).

Supplementary Figure 3 Endogenous chromatin dynamics in embryonic cells of pluripotency and neuro-mesodermal bipotency. (**a-e**) Endogenous dynamics of chromatin engagement between TFs (Sox3, mVegT, Foxh1[24], Eomes[19], zVegT[19], Brachyury[19] and Tbx6) and signal mediators (β-catenin, Smad1, Smad2) at indicated developmental stages (color-coded). (**a**) Spearman correlations (R_s) of DNA occupancy levels. (**b**) PCA of chromatin engagement. (**c**) PCA of RNAPII recruitment levels at CRMs occupied by TFs and signal mediators. r, biological replicates. (**d**) Coverage and enrichment of DNA motifs. (**e**) DNA occupancy-assisted clusters of CRMs co-aligned with enriched DNA motifs detected in peak centres (±50 bp) and flanking sequences (each 50 bp).

Supplementary Figure 4 Developmental context influences chromatin engagement. (**a-f**) MyoD-enforced change of the development context in early gastrula embryos and its influence on chromatin recruitment of Sox3 and RNAPII. (**a**) Spearman correlations (R_s) of DNA occupancy levels. (**b**) Map of DNA occupancies and enriched DNA motifs sorted by significantly differential (pΔ) Sox3 binding. (**c**) Snapshot of chromatin co-recruitment at canonical MyoD target gene *actc1*. (**d**) Animal cap assay to measure MyoD-enforced transcription. (**e**) RT-qPCR results from the animal cap assay. Error bars, mean+SD (n=2). (**f**) Opportunistic recruitment of MyoD-HA to canonical Sox3 target gene *sox2*. (**g-j**) Spatio-temporal co-enrichment of DNA motifs at CRMs occupied by Sox3, Eomes[19], Brachyury[19] and/or Tbx6 at indicated developmental stages reflect developmental context. (**g**) Spearman correlations (R_s) of DNA occupancy levels. (**h**) Coverage and enrichment of DNA motifs. (**i**) Hierarchical clustering of DNA occupancies separated by time and space. (**j**) Sox3-expressing cells of the posterior neural tube originate from the chordoneural hinge (CNH) and posterior wall of the neurenteric canal (PNC) expressing Brachyury and Brachyury/Tbx6, respectively. Both neural tube and PNC are also exposed to the expression of Cdx such as Cdx4[44]. Abbreviations of the developmental timeline: 1K, 1,024-cell (early blastula); eG, early gastrula; lG, late gastrula; and eTailbud, early tailbud.

Supplementary Figure 5 Signal-induced regionalization of ZGA depends on maternal TFs. (**a**) Morphological phenotype caused by the LOF of mVegT at late tailbud stage. (**b**) Rescue of reduced *nodal5*, *nodal6*, *sox17b* and *gata4* transcript levels in mVegT LOF embryos by the injection of *X. tropicalis* z*VegT* mRNA as determined by RT-qPCR at blastula stage. Error bars, mean+SD (n=5). Two-tailed Student's t-test: *, $p < 0.02$. (**c**) Morphological phenotypes caused by the single LOF of canonical Wnt, Nodal or BMP signaling when control embryos

reached late gastrula and mid-tailbud stage. (**d**) Transcriptional comparison of zygotic genes between indicated LOFs of maternal TFs or signals. Dots are colored according to the normal ratio of transcript levels (regional expression) across the animal-vegetal (An:Vg) or dorso-ventral (D:V) axis{Blitz:2017fi}. (**e,f**) Venn diagrams of genes downregulated by indicated LOFs. (**g**) Early gastrula-staged WMISH: *foxb1* and *zic1*, vegetal view, dorsal side facing top. (**h**) Quantification of *foxb1* transcript levels in control and mPouV/Sox3 LOF animal caps with or without canonical Wnt signaling. Error bars, mean+SD (n=2). Two-tailed Student's t-test: *, p=0.05. Numbers in the right bottom corner of each image refer to the count of embryos detected with the displayed morphological phenotype (**a,c**) or WMISH staining (**g**) among all embryos analyzed per condition and *in situ* probe. Scale bars, 0.5 mm.

Supplementary Figure 6 Selected maternal TFs and signals regionalize zygotic genome activation (ZGA) to induce the formation of the three germ layers and the primary body axes. (**a**) Percentage of zygotic genes misregulated (% ZGA) by indicated LOFs. (**b**) Enrichment analysis of key GO term-related biological processes among zygotic genes misregulated by indicated LOFs. Size of bubble reflects the statistical significance (hypergeometric p-value) of enrichment while the color indicates the number of affected genes.

Supplementary Figure 7 Pioneering activity of mPouV and Sox3 initiates extensive chromatin remodeling such as the chromatin looping of distal CRMs with promoters. (**a**) Heat map of Spearman correlations (R_s) among indicated biological replicates of MBT-staged DNase-Seq on extracted chromatin and naked genomic DNA (gDNA). (**b**) Double-logarithmic volcano plot of chromatin accessibility (DNase cleavage) changes at CRMs between mPouV/Sox3 LOF and uninjected embryos. CRMs with statistical significant changes (n=2, FDR ≤10%) are marked in red. (**c**) Violin plot of chromatin accessibility (DNase cleavages) comparing uninjected and mPouV/Sox3 LOF embryos at all and affected (FDR ≤10%) CRMs. Wilcoxon test: *, $p < 2.2 \times 10^{-16}$. (**d,e**) Performance of next-generation capture-C using novel chromatin conformation capture (3C) protocol adapted to early *Xenopus* embryos. (d) Bar graph of normalized counts (mean+SD; n=6) of non-redundant capture-reporter FLASH (Magoč and Salzberg, 2011) reads for each viewpoint captured with 1 or 2 probes (**Supplementary Table 4**). (**e**) Summary of FLASH reads from sequencing 3C libraries after two rounds of hybridization-based capture. (**f**) Violin plot of promoter-proximal interactions with unaffected and affected CRMs for both all and affected (FDR <10%) chromatin contacts. Wilcoxon test and effect size estimates (from left to right): $p=5 \times 10^{-5}$ and $r_{effect}=0.12$ (small effect); $p=5 \times 10^{-30}$ and $r_{effect}=0.34$ (medium effect); and $p=5 \times 10^{-10}$ and $r_{effect}=0.78$ (large effect). (**g**) Superimposed line plots showing the LOF-induced changes to various chromatin features at MBT including chromatin contacts with the promoters (viewpoints) of *foxb1*, *cdc25b* and *zic1*. The β-catenin profiles were generated at late blastula stage rather than early blastula stage (Sox3) or MBT (all others). RNA profile is split into low and high expression windows. Low-expression profiles show that locally transcribed non-coding RNAs depend on mPouV/Sox3 like the associated genes. Heat maps (pΔ) below each superimposed line plot show the statistical significance of mPouV/Sox3-induced changes. Header highlights the occurrences of canonical POU/SOX motifs (black filled rectangles) at CRMs, the conservation level (phastCons) of the genomic DNA sequence among vertebrates and some strongly affected chromatin sites with an arrowhead.

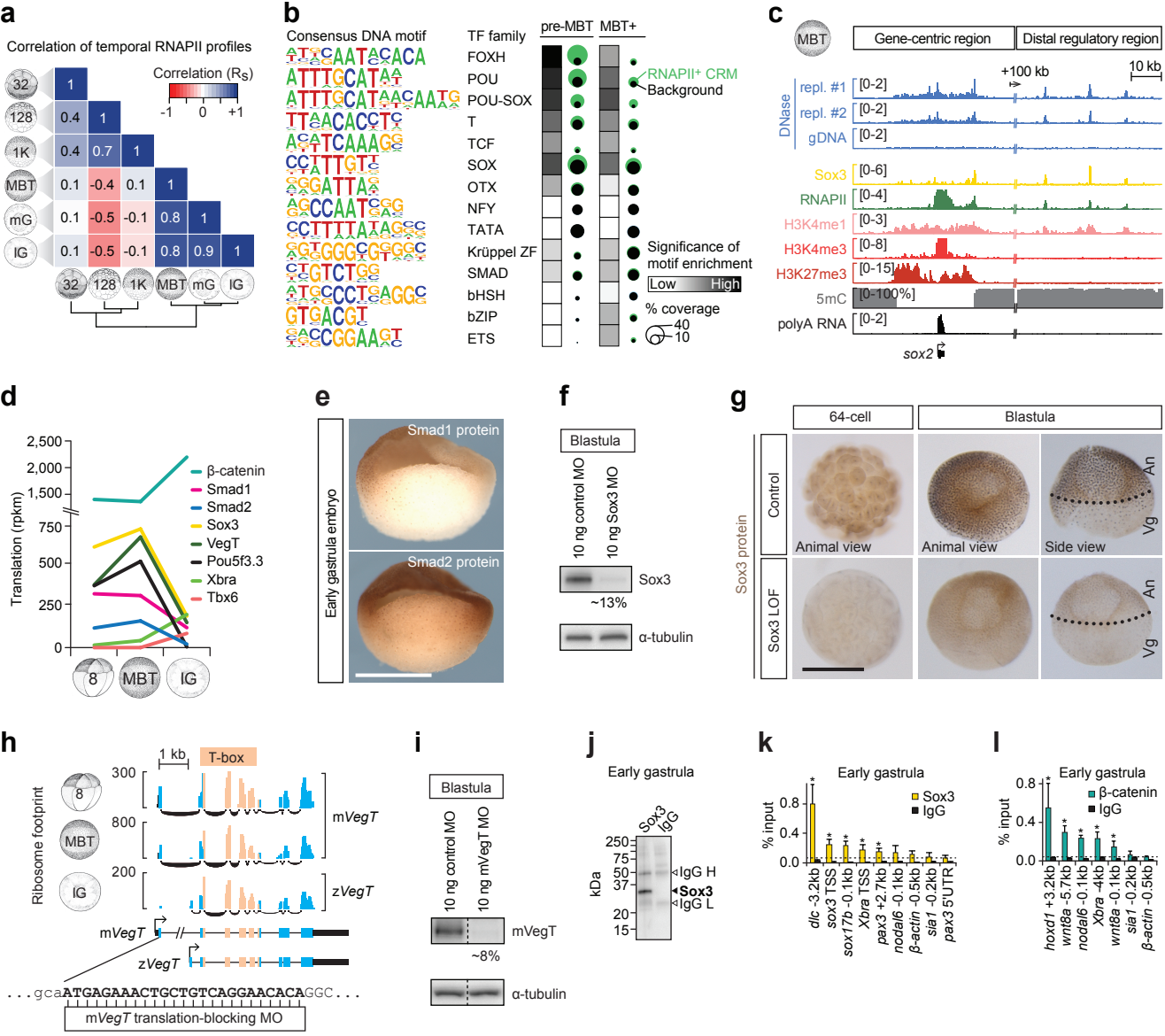
Supplementary Movie 1 Gastrulation defects caused by the loss-of-function of maternal pluripotency factors.

Supplementary Table 1 Summary of deep sequencing and read alignments.

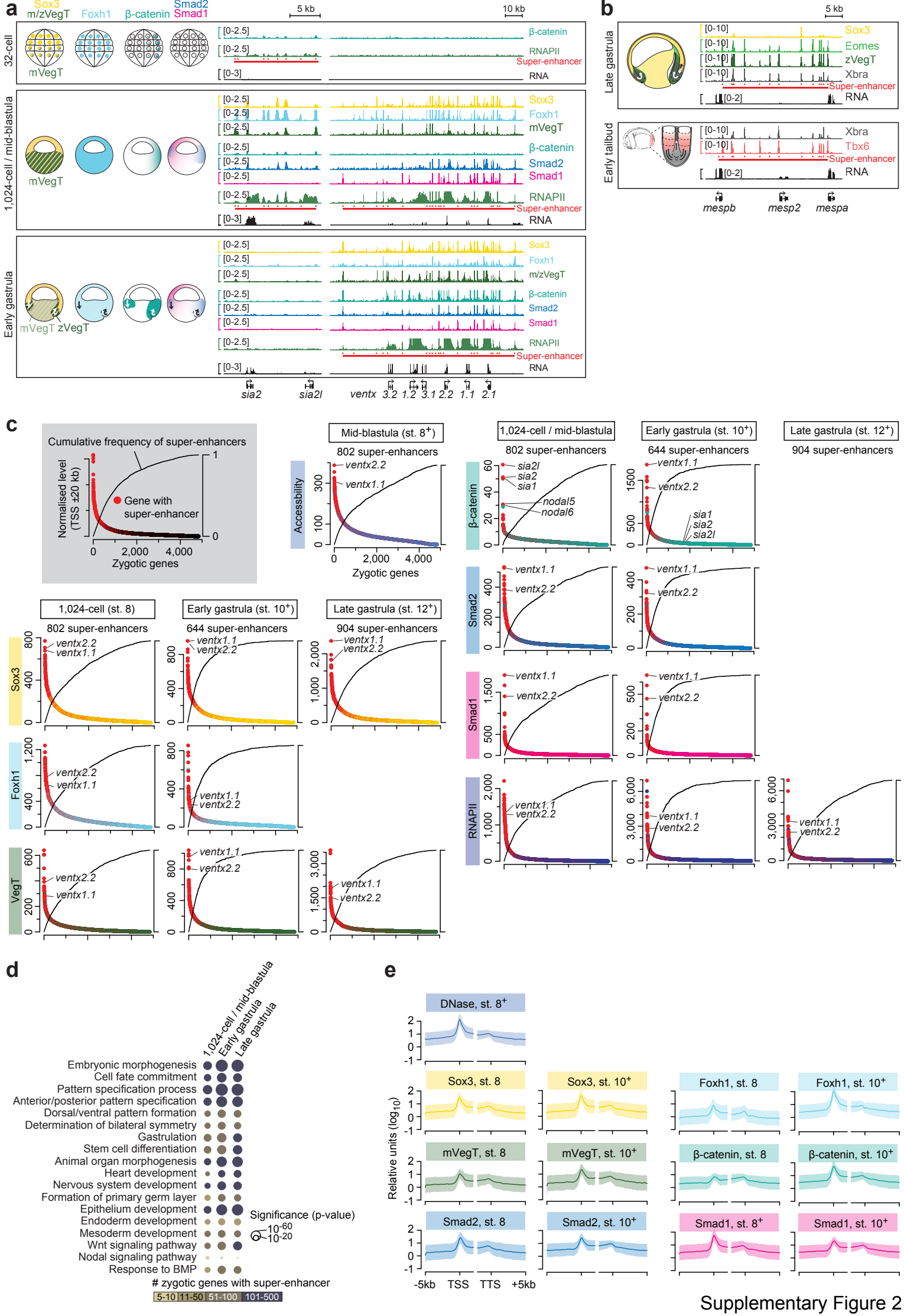
Supplementary Table 2 Maternal protein concentrations versus ribosome densities (translation) across the mid-blastula transition.

Supplementary Table 3 Transcriptional effect of maternal transcription factors and inductive signals on zygotic genome activation.

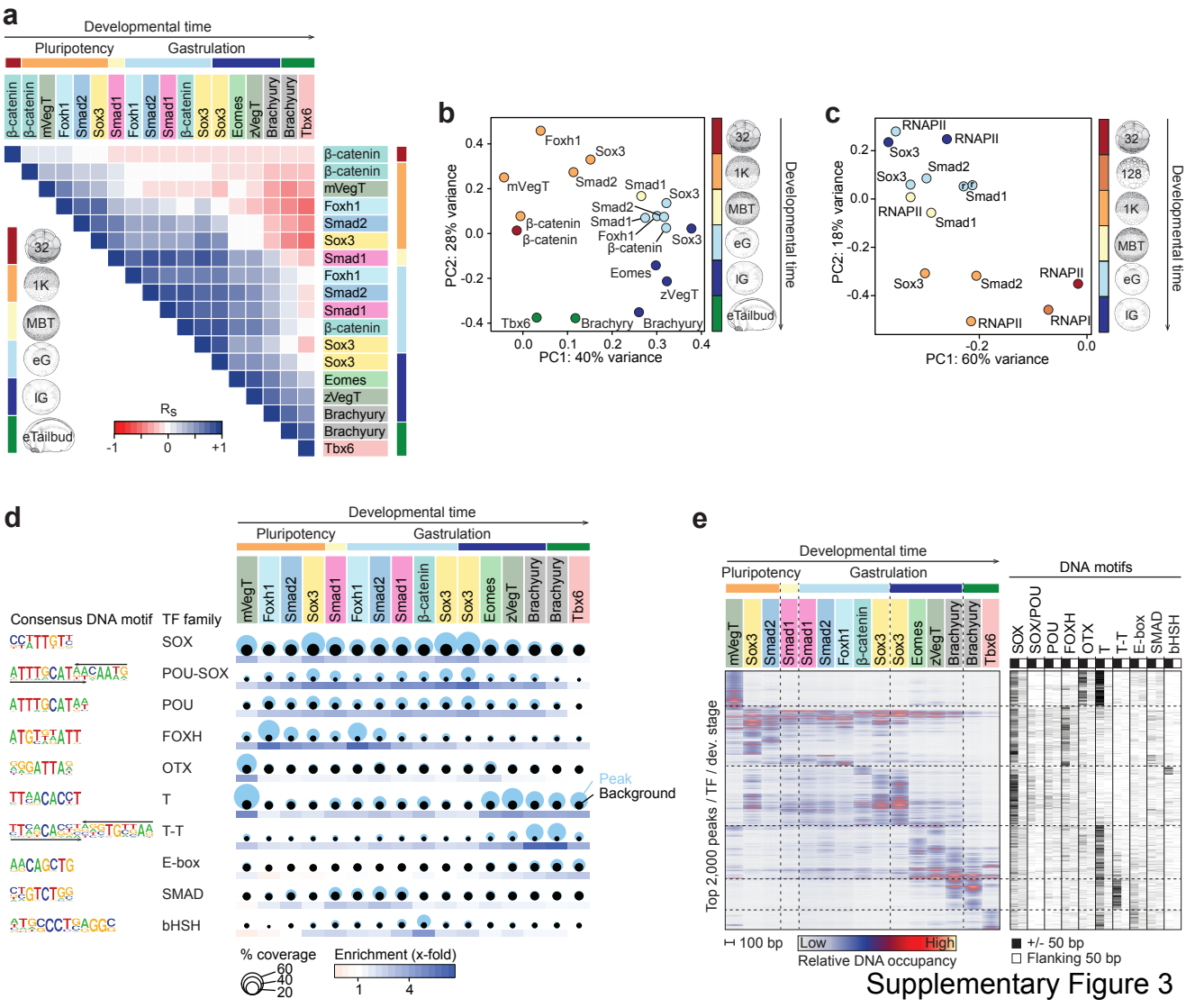
Supplementary Table 4 Genomic coordinates of selected capture-C viewpoints and DNA sequences used for capturing promoter-tied regulatory elements.



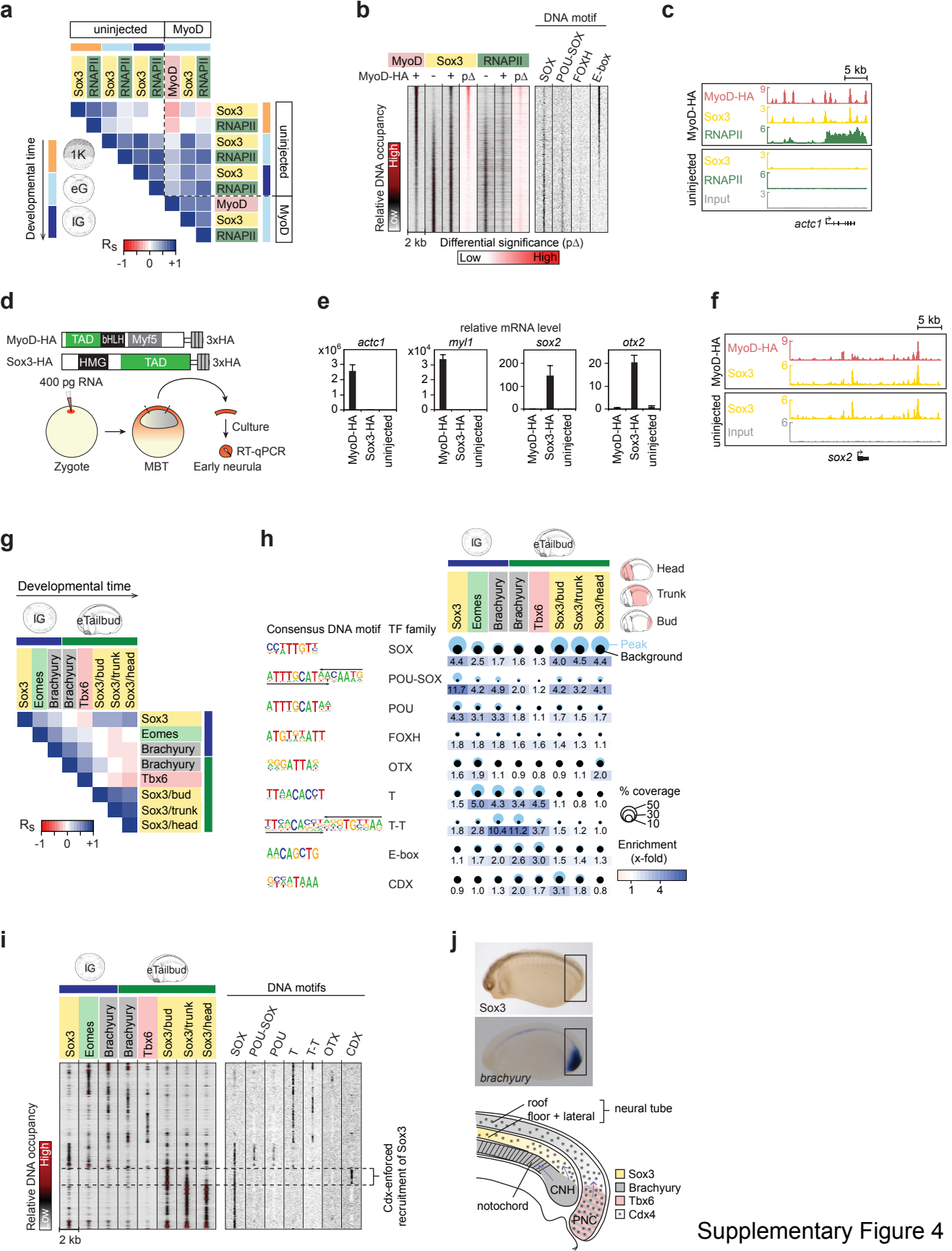
Supplementary Figure 1



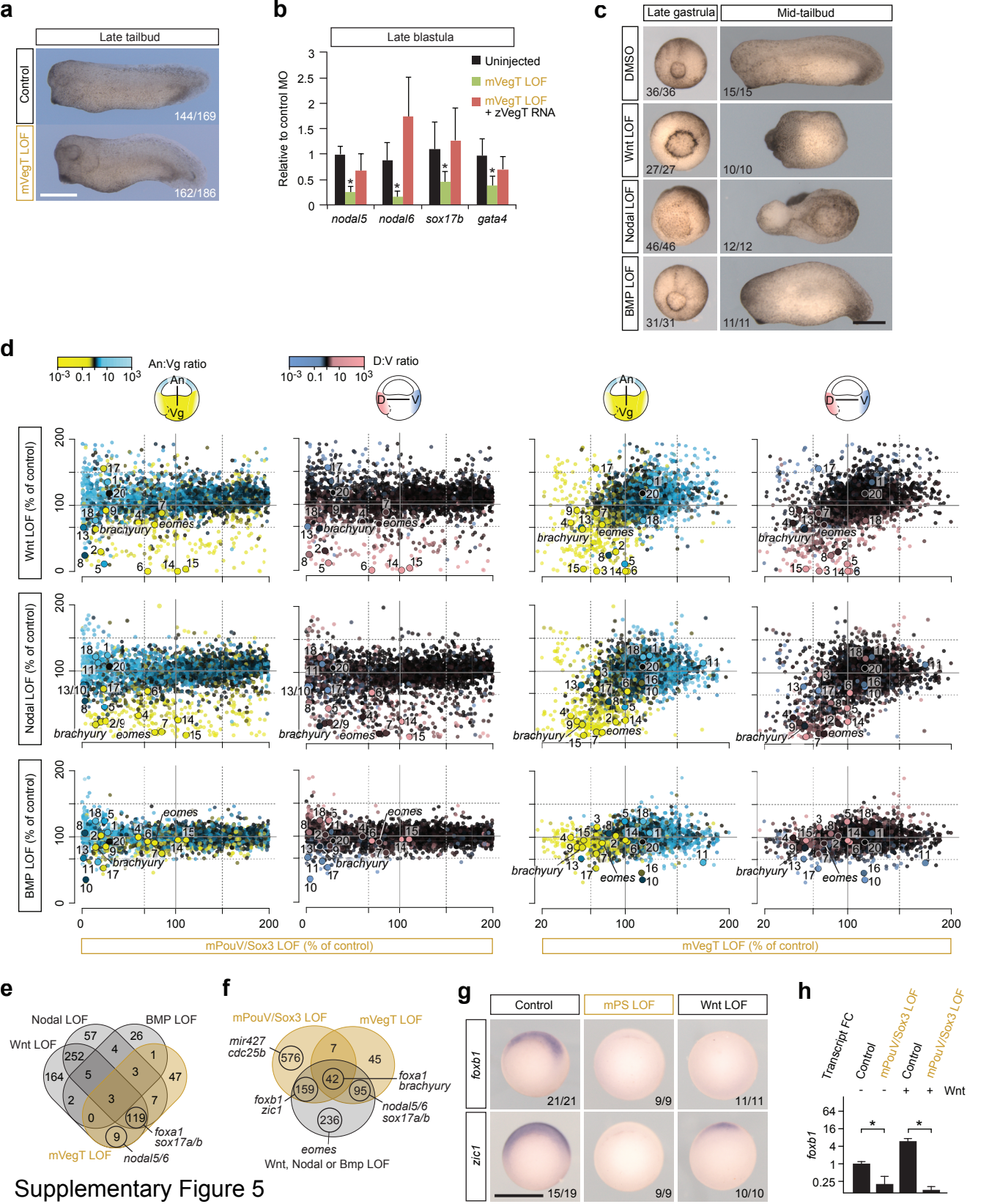
Supplementary Figure 2



Supplementary Figure 3

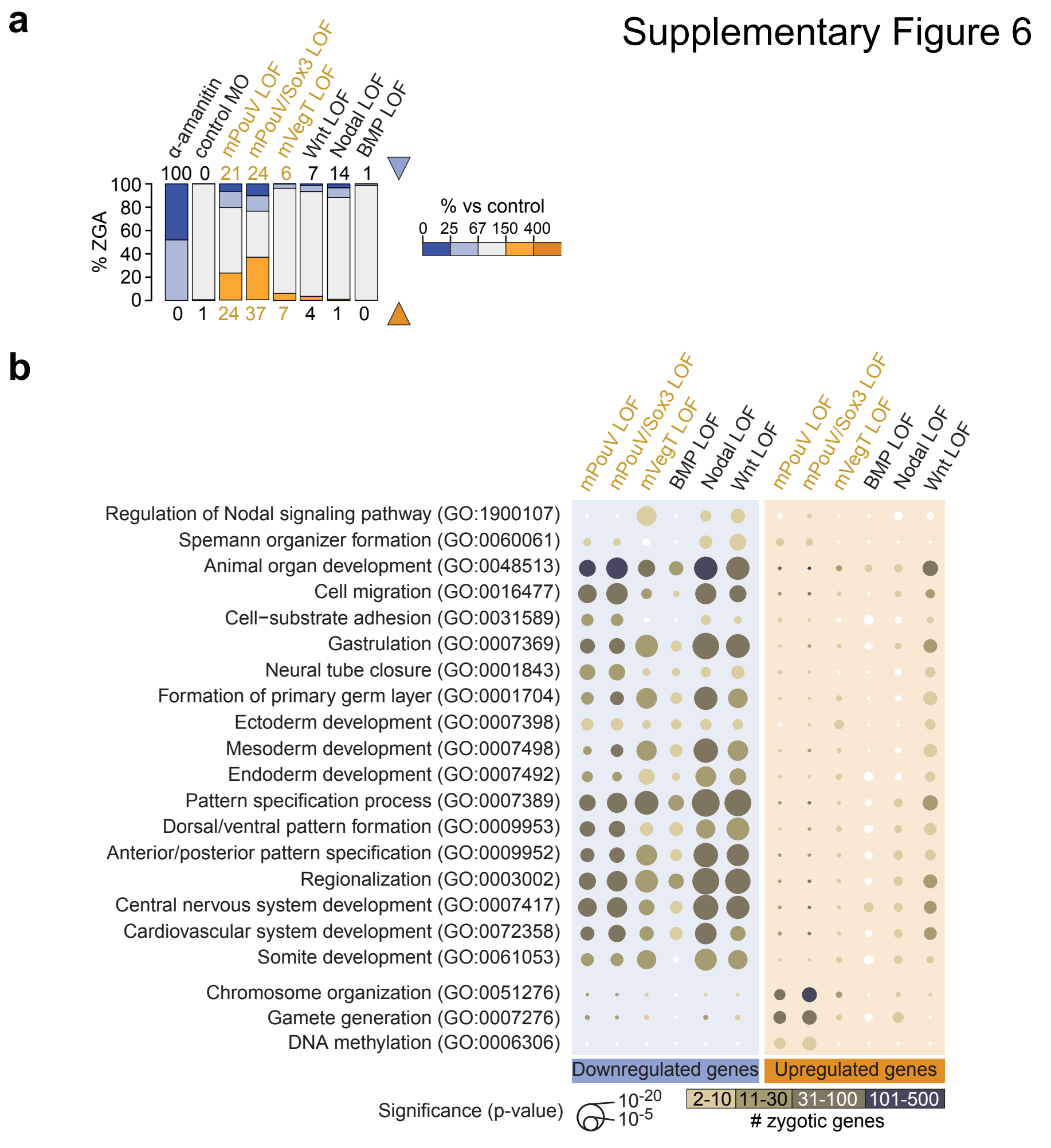


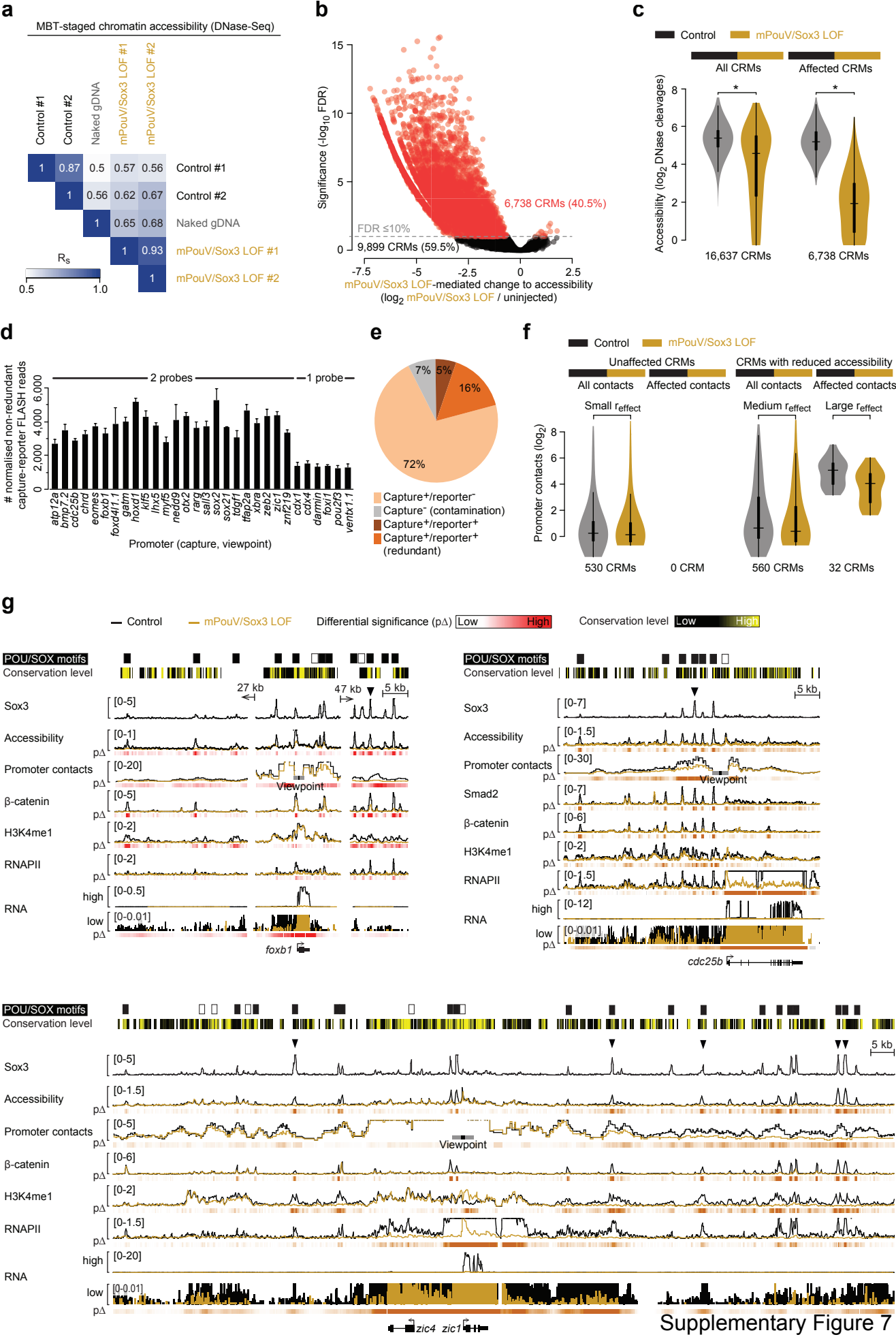
Supplementary Figure 4



Supplementary Figure 5

Supplementary Figure 6





Supplementary Figure 7