Maternal pluripotency factors prime the zygotic genome to respond to intercellular signals

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

[1] Developmental Biology Laboratory, The Francis Crick Institute, 1 Midland Road, London, NW1 1AT, United Kingdom

[2] Present address: Centre for Genomic Regulation (CRG), The Barcelona Institute for Science and Technology, Dr Aiguader 88, 08003 Barcelona, Spain

[3] Present address: Epigenetics of Stem Cells, Department of Stem Cell and Developmental Biology, Pasteur Institute, CNRS UMR3738, 25 Rue du Dr Roux, 75015 Paris, France

*** Corresponding Authors**

George E. Gentsch and James C. Smith

The ability of cells to respond to intercellular signals, and the repertoire of possible responses to those signals, is defined by their 'competence', a long-standing developmental concept[1], that is key to tissue formation and tissue regeneration. Apart from a few cell type-specific transcription factors[2], little is known about the role of the chromatin landscape in defining competence. Here, we investigate the emergence of competence in *Xenopus tropicalis,* which, like most multicellular organisms, begins development with a transcriptionally silent genome. We first identified the earliest functional regulatory DNA sequences, and inferred from them the critical sequence-specific activators of the zygotic genome. Of these, we showed that the maternal pluripotency factors Pou5f3 and Sox3 specifically open thousands of compacted chromatin sites by recognising their canonical DNA binding sequences. This pioneering activity initiates local chromatin remodelling to facilitate poised or active transcription, including the post-translational marking of displaced nucleosomes and chromatin looping with promoters. These remodelled chromatin sites are then capable of responding to intercellular signalling by recruiting the signal mediators that convert signals into tissue-specific expression of developmentally critical genes. As well as shedding light on genome activation and competence, our findings suggest that transcriptional responses to signals can be re-awakened by overexpressing the appropriate pioneer factors—a much sought-after objective in cell replacement therapies.

36

37 Zygotic genome activation (ZGA) in the *Xenopus* embryo occurs in a defined spatio-temporal
38 manner around the mid-blastula transition (MBT) and depends on maternal gene products
39 and inductive signalling between cells[3] (Extended Data Fig. 1a). Thus, spatially restricted Wnt,
40 Nodal and BMP signals trigger the nuclear accumulation of their signal mediators β-catenin,
41 Smad2 and Smad1, respectively, which in turn induce tissue-specific transcriptional
42 responses that control the formation of the body axes and the formation of the three germ
43 layers (ectoderm, mesoderm and endoderm) from pluripotent tissue[4]. The same signals are
44 used subsequently in later development to elicit different responses, and we wondered
45 whether the chromatin landscape might underlie these differences in competence.

46

47 To identify chromatin factors likely to confer competence during ZGA, we first examined the
48 DNA motifs present at transcriptionally engaged *cis*-regulatory modules (CRMs) from the 32-
49 cell to the late gastrula stage, by mapping focal RNA polymerase II (RNAPII) depositions on
50 a genome-wide scale (Fig. 1a). RNAPII recruitment to CRMs proved to be dynamic, with the
51 largest changes occurring between the 1,000-cell stage and MBT (Fig. 1a). In particular, pre-
52 MBT recruitment coincided with the enrichment of canonical recognition motifs of the FOXH,
53 POU, SOX and T domain transcription factor (TF) families (Fig. 1b; Extended Data Fig. 1b).
54 Using this approach some of the RNAPII-engaged intronic or gene proximal CRMs were not
55 detectable because they were masked by RNAPII-mediated gene transcription. Thus, MBT-
56 specific CRMs were further characterised for chromatin accessibility (using a novel DNase-
57 Seq method; see exemplar comparison with other chromatin features in Extended Data Fig.
58 1c) and for the enhancer-associated histone mark H3K4me1. We detected ~20,000
59 accessible CRMs occupied by RNAPII and flanked by H3K4me1^{+} nucleosomes (Fig. 1c). The
60 enriched DNA motifs among accessible H3K4me1^{+}/RNAPII^{+} CRMs were then correlated with
61 the maternally inherited and translated sequence-specific factors identified by egg-staged
62 mass spectrometry[5] and pre-MBT ribosome footprinting[6] (Fig. 1d; Extended Data Fig. 1d;
63 Supplementary Table 1). It proved that the binding preference of the most frequently translated
64 maternal TFs and signal mediators matched the most significantly enriched DNA motifs (Fig.
65 1e). These were POU-SOX (Pou5f3-Sox3 heterodimer), Krüppel ZF (Sp/Klf), POU (Pou5f3),
66 SOX (Sox3), bZIP (Max), FOXH (Foxh1), ETS (Ets2), NFY (NFYa/b/c), SMAD (Smad1/2), T
67 (mVegT, a maternal VegT isoform), TCF (Tcf/β-catenin), basic helix-span-helix domain
68 (bHSH; Tfap2), and OTX (Otx1). Notably, the enrichment level of these motifs were
69 independent of CRM sequence conservation (based on phastCons and phyloP scoring)
70 detected among vertebrates (Fig. 1c,e).

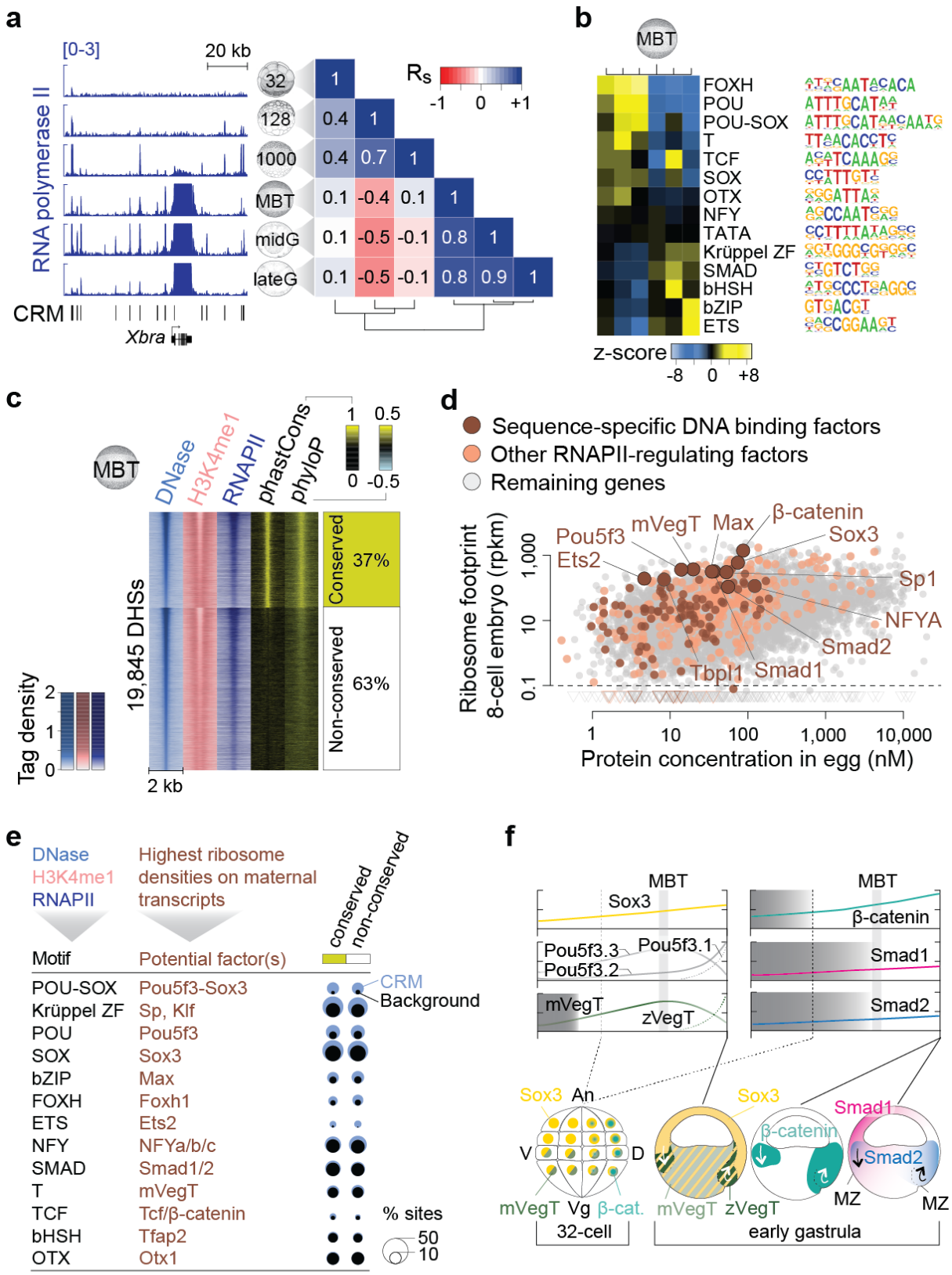


Figure 1. CRM-enriched DNA motifs match binding preferences of maternal TFs most frequently translated before MBT {dimension of figure: 1 column wide}

a,b, RNAPII recruitment to CRMs from 32-cell to late gastrula stage. **a**, Snapshots and Spearman correlations R_s. **b**, Z-score of enriched DNA motifs. **c**, MBT-staged map of DNase-probed chromatin accessibility, RNAPII binding and H3K4me1 marking grouped by sequence

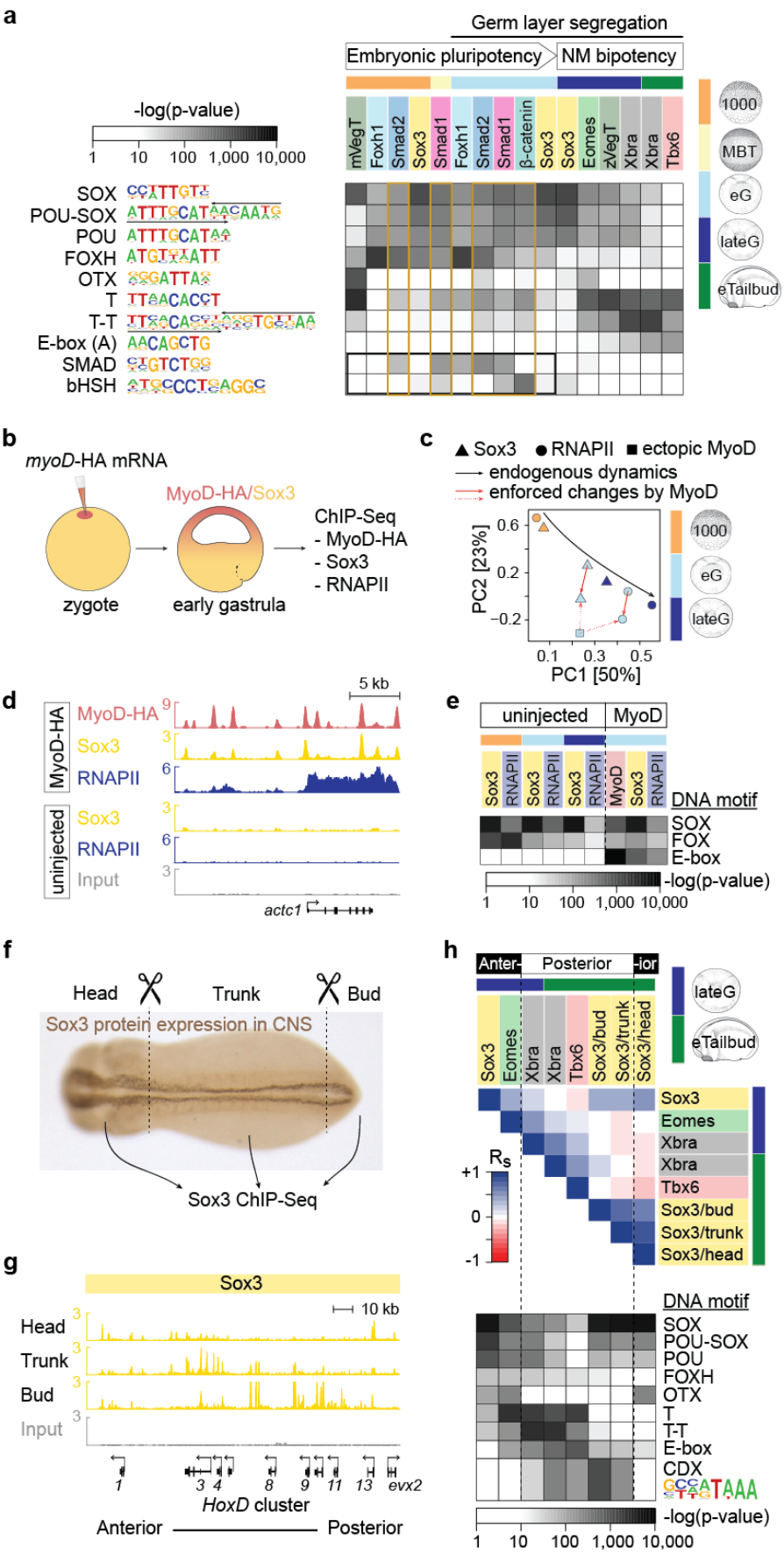
conservation levels (phastCons, phyloP) and sorted by statistical significance of DHSs. **d**, Maternal protein concentrations[5] versus ribosome footprint levels at 8-cell stage[6]. Most frequently translated representatives of various TF families are labelled. **e**, Matching canonical CRM-enriched DNA motifs (sorted by statistical significance) with frequently translated sequence-specific factors. **f**, Graphical illustration of protein levels and nuclear localisations of selected proteins based on our and previously published result[7-12]. Shaded boxes indicate periods of non-nuclear protein distribution. Arrows indicate tissue movements. Abbreviations: An, animal; D, dorsal; MZ, marginal zone; V, ventral; Vg, vegetal.

Based on this orthogonal correlation, we selected Pou5f3, Sox3 and mVegT as factors that might specify responsiveness to Wnt, Nodal and BMP signalling. Sox3 and presumably also Pou5f3 (based on the spatial distribution of its transcripts[13]) are detected ubiquitously, while mVegT is restricted to the vegetal hemisphere[9] (Fig. 1f; Extended Data Fig. 1g). With respect to signal mediators, nuclear β-catenin accumulates dorsally (first detected at the 32-cell stage[8]) before spreading around the forming blastopore[11] (Fig. 1f). 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[11,12] (Fig. 1f; Extended Data Fig. 1e).

Next, informed by the onset of their nuclear accumulation (Extended Data Fig. 1a), we generated genome-wide chromatin profiles across MBT to ask whether the recruitment of β-catenin, Smad1 and Smad2 might be influenced by co-expressed maternal TFs like Sox3, mVegT and Foxh1[14] (Extended Data Fig. 2 and 3a). As an outgroup control, we selected the binding profiles of several zygotic T-box TFs Eomes[15], zVegT[15], Xbra[15] and Tbx6 (Extended Data Fig. 2 and 3b), which collectively regulate the neuro-mesodermal cell lineage during and beyond gastrulation[16].

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[17] input (Extended Data Fig. 3a-c). (2) From the 1,000-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 (Extended Data Fig. 3d). (3) Among all zygotic genes, promoter-proximal regions were most consistently bound (Extended Data Fig. 3e). All these characteristics were also observed for chromatin accessibility and RNAPII engagement at CRMs (Extended Data Fig. 3c,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 (Extended Data Fig. 4a,b). The elements associated with developmental context were largely revealed by the binding

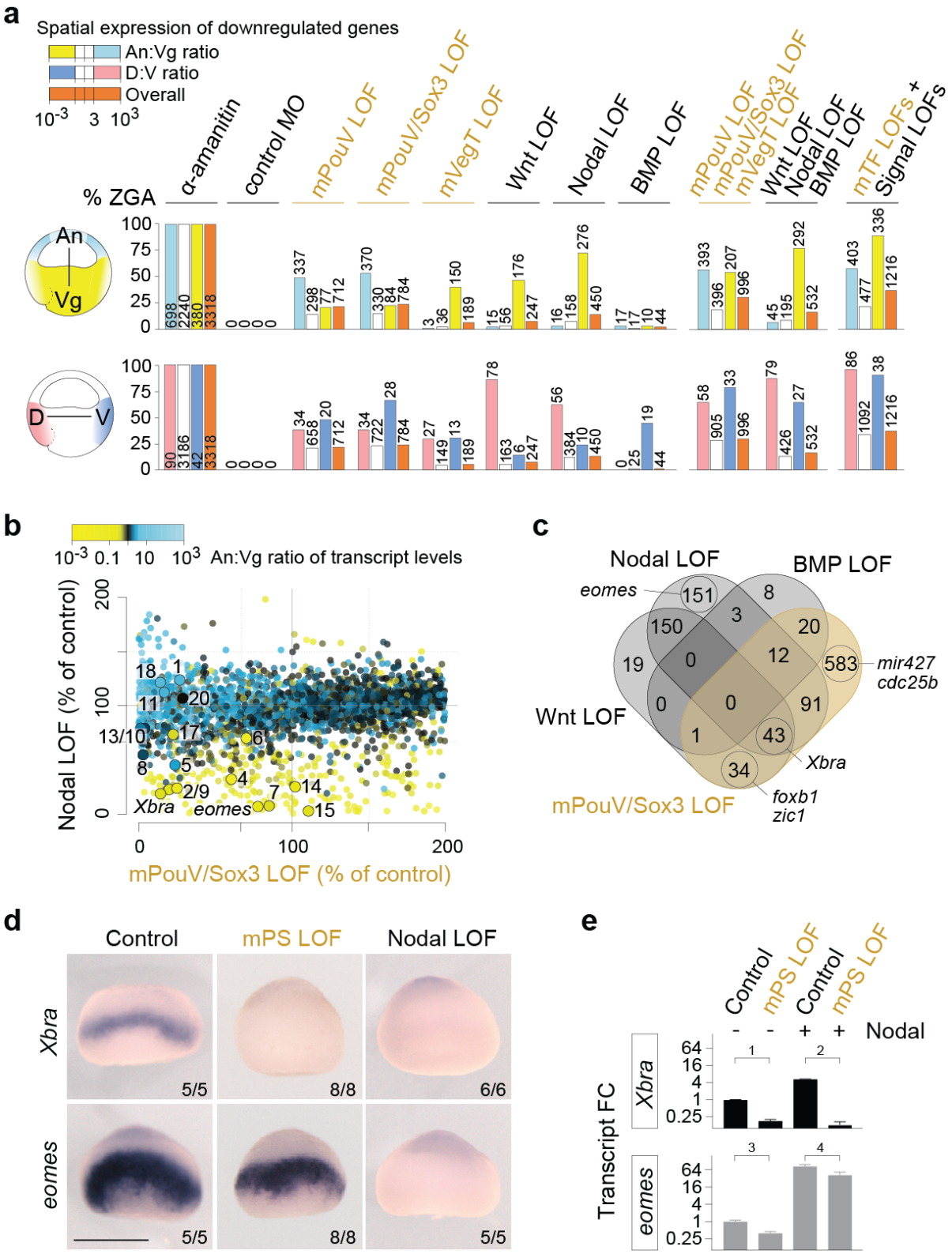
116 pattern of sequence-nonspecific RNAPII at CRMs (Extended Data Fig. 4c). The discovery and
117 co-clustering of enriched DNA motifs at occupied CRMs indicate that factor-intrinsic elements
118 include the sequence-specificity of DNA binding domains, oligomerisation tendencies (e.g.
119 Xbra binding palindromic motifs due to its propensity to form homodimers[18]) and protein-
120 protein interactions, while the developmental context reflects co-expressed TFs (Fig. 2a;
121 Extended Data Fig. 4d,e).
122
123 To demonstrate the importance of developmental context on chromatin engagement, we
124 ectopically expressed HA-tagged MyoD (MyoD-HA) in animal cap cells and profiled the
125 chromatin for MyoD-HA at the early gastrula stage, as well as for endogenous Sox3 and
126 RNAPII (Fig. 2b). MyoD was chosen because its canonical E-box recognition motif is normally
127 not enriched at the early gastrula stage (Fig. 2a), while Sox3 and RNAPII were selected
128 because they are ubiquitously expressed and represent sequence-specific and nonspecific
129 DNA binding factors, respectively. As predicted, the E-box motif of MyoD-HA emerged as a
130 context-dependent motif of Sox3 and RNAPII binding, while MyoD-HA binding itself was
131 influenced by the endogenous TF expression profile (Fig. 2c,e; Extended Data Fig. 5a,b).
132 MyoD-HA co-recruited Sox3 to MyoD gene targets like *actc1* and *myl1* (Fig. 2d; Extended
133 Data Fig. 5c,d). However, this so-called opportunistic recruitment to non-canonical binding
134 sites, such as MyoD-HA to Sox3 gene targets (e.g. *otx2*, *sox2*), did not affect transcription in
135 animal caps (Extended Data Fig. 5c-e).
136
137 The influence of TF co-expression on DNA occupancy was further substantiated by profiling
138 chromatin for Sox3 in different anterior-posterior regions of the central nervous system (CNS;
139 Fig. 2f). The analysis of enriched DNA motifs suggested that Sox3 binding was influenced by
140 differentially expressed homeodomain TFs such as orthodenticle homeobox (OTX) in the brain
141 (head) and caudal homeobox (CDX) in the spinal cord (trunk, tailbud) (Fig. 2h; Extended Data
142 Fig. 5f-h). This was particularly apparent in Sox3 binding to the colinear *HoxD* cluster defining
143 anterior-posterior cell identity (Fig. 2g). Similar influences on chromatin engagement were
144 observed within anterior and posterior mesoderm marked by Eomes in gastrula embryos and
145 Xbra/Tbx6 in gastrula and early tailbud embryos, respectively (Fig. 2h; Extended Data Fig. 5f).
146 On a temporal scale, the influence of FOXH and POU motif-recognising TFs in recruiting Sox3
147 and T-box TF was more pronounced early than late in development (Fig. 2h; Extended Data
148 Fig. 5f). Interestingly, the analysis of enriched DNA motifs suggests that maternal TFs affect
149 the recruitment of signal mediators, but (with the exception of Foxh1, which interacts directly
150 with Smad2[19]) the reverse is much less the case (compare highlighted orange and black boxes
151 in Fig. 2a; Extended Data Fig. 4d,e).



152 **Figure 2. Developmental context influences chromatin engagement** {dimension of figure:

153 1 column wide}
154 **a**, Significantly enriched DNA motifs at CRMs occupied by indicated proteins covering
155 developmental process of embryonic pluripotency and neuromesodermal (NM) bipotency. **b-**
156 **e**, MyoD-enforced recruitment of Sox3 and RNAPII. **b**, Experimental design. **c**, PCA of
157 chromatin engagement. **d**, Snapshot of chromatin co-recruitment. **e**, Significantly enriched
158 motifs. **f-h**, CNS-specific Sox3 binding along the anterior-posterior axis (head, trunk and bud).
159 **f**, Experimental design. **g**, Snapshot of Sox3 binding to *HoxD* cluster. **h**, Spearman correlation
160 R_s of TF binding. Below, heat map of significantly enriched DNA motifs.
161
162 To ask whether signal-mediated ZGA requires maternal TFs, as suggested by the observed
163 chromatin dynamics, we next compared the effects of loss of maternal Sox3/PouV (Pou5f3.2
164 and Pou5f3.3; mPouV) or VegT (mVegT), and canonical Wnt, Nodal or BMP signal
165 transduction, on zygotic transcription from MBT to late blastula and early gastrula stage. The
166 high translation frequencies (Fig. 1d; Extended Data Fig. 1d) coupled with steady or even
167 transient protein levels around MBT (Fig. 1f), suggest that maternal TFs and signal mediators
168 have short half-lives. Consistent with this idea, the injection of morpholinos (MOs) blocking
169 the translation of maternal transcripts such as *Sox3* or *mVegT* was effective in reducing protein
170 levels (Extended Data Fig. 1f,i). We used previously validated MOs to knock down mPouV[20]
171 and β-catenin[21]. The nuclear accumulation of Smad1 and Smad2 in response to BMP and
172 Nodal signalling was inhibited with the small molecules LDN193189 and SB431542,
173 respectively. The morphological defects of these loss-of-function (LOF) treatments ranged
174 from undetectable (Sox3), to weak (mVegT, BMP), to moderate (mPouV), to severe
175 (mPouV/Sox3, Nodal, β-catenin) (Extended Data Fig. 6a,c,e). Moderate and severe defects
176 affected gastrulation, while weak ones were only obvious later. All phenotypes were either in
177 line with previous publications (β-catenin[21], Nodal[22], BMP[23]) or could be rescued on a
178 morphological (mPouV/Sox3; Extended Data Fig. 6b) or transcriptional level (mVegT,
179 Extended Data Fig. 6d) by co-injecting cognate mRNA. Interestingly, the role of maternal Sox3
180 could only be detected by knocking it down together with mPouV. In contrast to single Sox3
181 or mPouV LOF, double LOF embryos failed to close the blastopore (Extended Data Fig. 6a;
182 Supplementary Video 1).
183
184 Transcriptome analysis of TF and signal LOF embryos was confined to the 3,318 zygotically
185 expressed genes for which ≥two-fold reductions in exonic and/or intronic transcript level could
186 be detected following injection of the RNAPII inhibitor α-amanitin (Supplementary Table 2).
187 About 86% of these genes had a maternal contribution of ≥1 per 10 million transcripts
188 (Extended Data Fig. 7a). α-amanitin prevented any transcriptional changes and thus blocked
189 the gastrulation movements that are normally initiated by ZGA (Extended Data Fig. 6f,g). The
190 LOF-mediated reduction of gene activation ranged from ~1% for BMP to ~24% for

191 mPouV/Sox3 (Fig. 3a). We note that the additional loss of Sox3 in mPouV LOF embryos
192 further reduced transcription of developmentally relevant genes, including those that are
193 expressed ubiquitously, like *mir427*, or those that are activated in a subset of mPouV/Sox3^{+}
194 cells, like *ventx* and *Xbra* (Extended Data Fig. 7b,c).
195
196 Spatial analysis of reduced gene activation confirmed that many of the 784 mPouV/Sox3-
197 dependent genes showed enriched expression along the animal-vegetal or dorso-ventral axis
198 (Fig. 3a). More specifically, comparison with signal LOFs revealed that 201 of the genes
199 tissue-specifically induced by Wnt, Nodal or BMP also depended on mPouV/Sox3 (Fig. 3b,c;
200 Extended Data Fig. 7d,f). Likewise, 142 of the 189 mVegT-dependent genes were activated
201 in response to signalling within the vegetal hemisphere, such as *foxa1*, *nodal5/6* and *sox17a/b*
202 (Extended Data Fig. 7d,e). Loss of mVegT and mPouV/Sox3 jointly affected 49 genes, 42 of
203 which also depended on Wnt, Nodal or BMP signalling (Extended Data Fig. 7f).
204
205 Remarkably, the requirement for particular maternal TFs varied even among related genes
206 with similar expression patterns that are activated by the same signals. For instance, *Xbra*
207 and *eomes* are both activated by Nodal signalling in the marginal zone between the animal
208 and vegetal hemispheres, but only *Xbra* requires mPouV/Sox3 for its signal responsiveness
209 (Fig. 3c). Significantly, DNA occupancy levels did not explain the difference in gene regulation
210 between *Xbra* and *eomes*, because both gene loci showed high levels of Sox3 binding (see
211 Fig. 4c). Identical dependencies were found for Wnt-responsive genes like *foxb1* and *zic1*
212 (Extended Data Fig. 7g).
213
214 We confirmed these mPouV/Sox3-dependent signal inductions by treating control and
215 mPouV/Sox3-depleted naïve (animal cap) tissue with Nodal (Activin) and Wnt (CHR99021)
216 agonists (Fig. 3c; Extended Data Fig. 7h). Interestingly, mPouV/Sox3 could trigger basal low-
217 level expression of (especially) *Xbra* and *foxb1* in the absence of signalling. Together, our
218 selected maternal TFs and signals activated 1,213 of the 3,318 (~37%) zygotic genes. These
219 included ~58% of those expressed animally, ~88% of the vegetal genes, ~96% of the dorsal
220 genes and 90% of those expressed ventrally, while many ubiquitously expressed genes
221 remained unaffected (Fig. 3a). Thus, mPouV and Sox3 tend to affect biological functions for
222 which tissue-specific expression is essential, such as main body axis formation,
223 regionalisation and germ layer segregation (Extended Data Fig. 8b). We also note that LOF
224 of mPouV/Sox3 also caused significant increases in transcript levels of active genes encoding
225 gamete-specific biological processes, suggesting that reprogramming towards embryonic
226 pluripotency was compromised (Extended Data Fig. 8a,b).



227

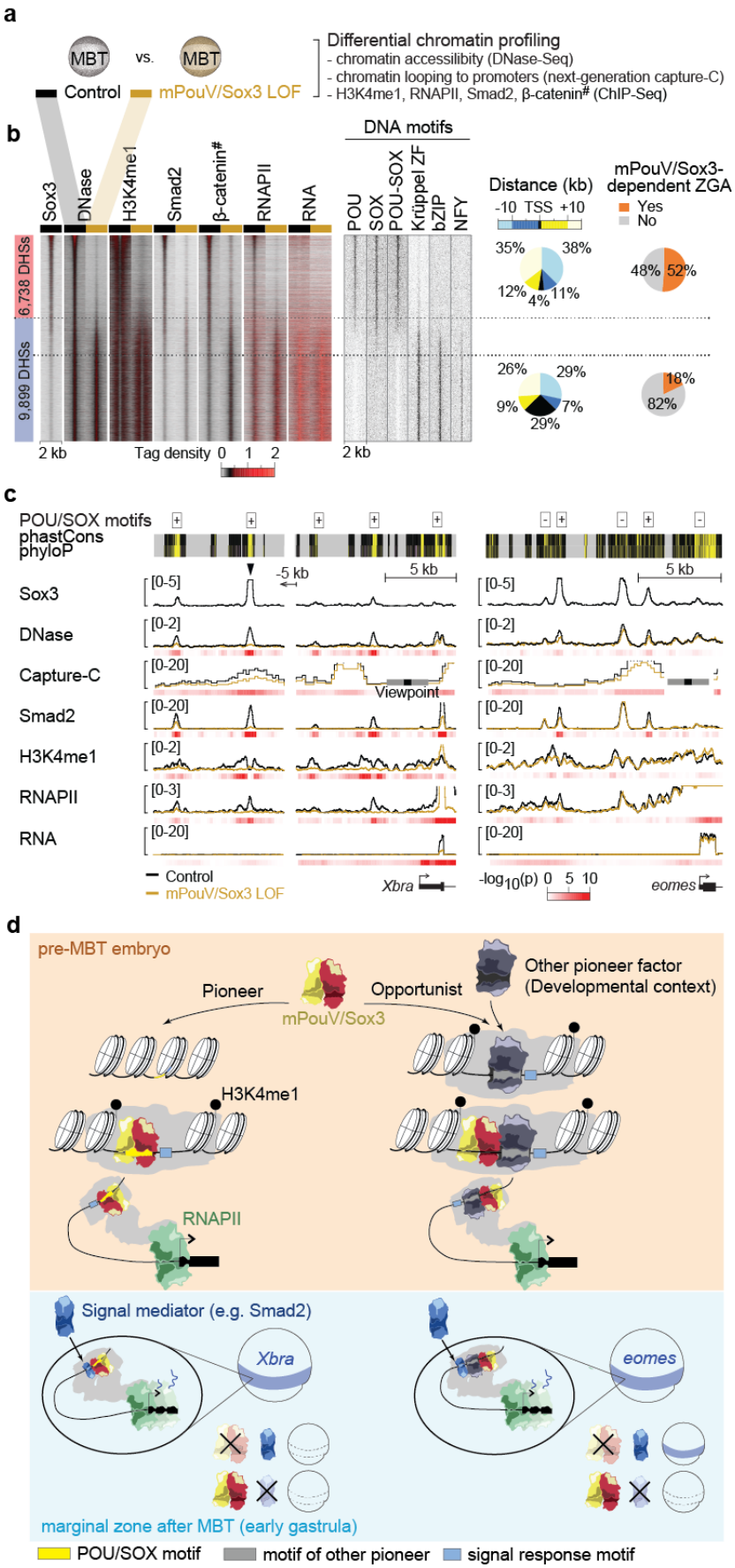
228 **Figure 3. Signal-induced ZGA depends on maternal TFs** {dimension of figure: 1 column
229 wide}

230 **a**, Regional expression of genes downregulated by indicated LOFs. Regional expression
231 refers to transcript level ratios along the animal-vegetal (An:Vg) or dorso-ventral (D:V) axis in
232 unperturbed early gastrula embryos[24]. Numbers and percentages of downregulated genes are

either specific to regional expression or reflect overall statistics. **b**, Scatterplot of transcript level changes caused by LOF of mPouV/Sox3 or Nodal signalling. Dots are coloured according to An:Vg transcript level ratios. Numbers refer to some more developmentally relevant genes (see Extended Data Fig. 5h). **c**, Venn diagram of downregulated genes. **d**, Early gastrula-staged WMISH for *Xbra* and *eomes* under indicated LOFs. Scale bar, 0.5 mm. **e**, Quantification of *Xbra* and *eomes* transcript levels in control and mPouV/Sox3 (mPS) LOF animal caps with or without Nodal signalling. Error bars, mean+SD (n=2). Two-tailed Student's t-test: 1, p=0.014; 2, p=0.005; 3, p=0.126; and 4, p=0.019.

To discover how maternal TFs identify genes to become competent to respond to signals, we compared various chromatin features between control and mPouV/Sox3 LOF embryos at MBT (Fig. 4a). Genome-wide analysis showed that mPouV/Sox3 LOF strongly reduced chromatin accessibility as measured by the significant loss of cleavages in 6,738 of the 16,637 (~41%) DNase hypersensitive sites (DHSs) (Fig. 4b; Extended Data Fig. 9a-c). Sorting of DHSs according to the significance of lost accessibility indicates that accessibility depends on the ability of mPouV and Sox3 to recognise their canonical motifs (Fig. 4b; Extended Data Fig. 9d). By comparison, unaffected DHSs rarely contained canonical mPouV/Sox3 motifs, and were strongly enriched for promoter-centric motifs[25] of the Krüppel ZF, bZIP and NFY protein family (Fig. 4b). Indeed, promoters (+/-0.5 kb from a transcription start site, TSS) were less frequently located within mPouV/Sox3-affected (~4%) than unaffected DHSs (~29%). By contrast, at the extreme end of affected loci (e.g. *Xbra*, *foxb1*, *cdc25b* and *zic1*) entire super-enhancers became accessible through the activity of mPouV/Sox3 (Fig. 4c, Extended Data Fig. 10d).

The loss of accessibility at distal DHSs triggered further profound changes to chromatin *in situ* (see arrowheads in Fig. 4c and Extended Data Fig. 10d for strongly affected CRMs). Chromatin looping from affected DHSs to promoters (Fig. 4c; Extended Data Fig. 10a-d) was significantly reduced and so was the deposition of DHS-flanking H3K4me1 and co-recruitment of RNAPII and signal mediators Smad2 and β-catenin (Fig. 4b,c; Extended Data Fig. 9d). At the transcriptional level, a reduction of non-coding and coding RNA also coincided with the affected DHSs (Fig. 4b; Extended Data Figs. 9d and 10d). Importantly, the differential chromatin profiling explained how tissue competence can be established and why, for instance, Nodal-induced transcription of *Xbra* was strongly affected by the loss of mPouV/Sox3 and that of *Eomes* less so. Thus, in contrast to those of *eomes*, all promoter-tied CRMs of *Xbra* contain canonical POU/SOX motifs, so they were not accessible to Smad2 in mPouV/Sox3 LOF embryos (Fig. 4c,d). Thus, Smad2 interactions with critical Nodal responsive regulatory elements of *eomes* remained intact, but those of *Xbra* were



270 **Figure 4. Maternal pluripotency factors pioneer silent zygotic genome to control its**

response to signalling {dimension of figure: 1 column wide}
a, Approach to reveal effect of mPouV/Sox3 on chromatin landscape at MBT. **b**, Heat map of chromatin and RNA profiles at DHSs sorted by significantly reduced DNase cleavages under mPouV/Sox3 LOF. #, β-catenin binding profile generated at late blastula stage rather than early blastula stage (Sox3) or MBT (all others). To the right, co-aligned map of DNA motifs, distance of DHSs to nearest TSSs and correlation with mPouV/Sox3-dependent ZGA. **c**, Superimposed line plots showing the LOF-induced changes to various chromatin features including chromatin contacts (capture-C) with promoters (viewpoints) of *Xbra* and *eomes*. For statistical significance check heat map below each profile. Header, POU/SOX motif occurrences at DHSs and sequence conservation heat maps (phastCons, phyloP). Arrowhead denotes strongly affected chromatin site. **c**, Model of pioneer-initiated chromatin remodelling to facilitate the transcriptional response to signals during ZGA.

impeded by closed chromatin. mPouV/Sox3-dependent tissue competence also applied to Wnt signalling with β-catenin failing to access Wnt responsive regulatory elements of *foxb1* in the absence of mPouV/Sox3 (Extended Data Fig. 10d).

Our results allow us to propose a model (Fig. 4d) of pioneer-initiated chromatin remodelling that facilitates the transcriptional response to inductive signals and thus defines competence during ZGA. First, the maternal TFs mPouV and Sox3 are recruited to silent CRMs either by recognising canonical DNA motifs or in a more opportunistic manner, presumably through the activity of other pioneers. Based on *in vitro* experiments[26], pioneering activity may require only partial motifs to be displayed on nucleosome surfaces to initiate access to compacted chromatin. Subsequent displacement of nucleosomes may be driven by the higher affinity of TFs for free DNA. In contrast, opportunistic binding occurs independent of sequence preference and is influenced by developmental context such as the co-expression of sequence-specific TFs as shown for the recruitment of Sox3 in various tissues. However, while pioneering activity triggers extensive local chromatin remodelling (e.g. H3K4me1 marking of displaced nucleosomes and the chromatin looping with promoters), opportunistic binding seems to have much less of an effect on chromatin.

We suggest that pioneer-initiated chromatin remodelling, or priming, occurs in every embryonic cell before the onset of Wnt, Nodal or BMP signalling to allow the recruitment of signal mediators. This is for three reasons. 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 signalling.

310 Our work has focussed on the pioneer activity of mPouV and Sox3, and, based on the
311 recognition motifs enriched at active CRMs before MBT, we suggest that only a few more
312 maternal TF families, including those of T-box, FoxH and Krüppel-like ZF families, are involved
313 in the acquisition of competence at the ZGA. Indeed, many of these TF family members are
314 relatively abundant and rapidly turned over, suggesting strong and dynamic TF activity[27]. Most
315 of these TFs are already present in the *Xenopus* egg, which confers pluripotency to somatic
316 cells when they are injected into its cytoplasm[28]. Their activity generates a signature of
317 accessible DNA motifs at the MBT that is conserved among vertebrates and most reminiscent
318 of the chromatin footprints in embryonic stem cells[25].
319
320 The findings presented here and elsewhere[29] reveal close connections between ZGA and
321 pluripotency, and we anticipate that competence is controlled in human embryonic stem cells
322 in a manner resembling that in the early *Xenopus* embryo. The approach described here can
323 be applied to any other cell type to discover the molecular basis of its competence.
324 Interestingly, various tumours are associated with the mis-expression of pioneer factors[30] and
325 thus may display different competencies from the surrounding host tissue. Conversely,
326 knowing which pioneers are required to trigger a specific signal response will increase the
327 success of engineering patient-specific tissue for transplantation therapy.
328
329
330 **Methods**
331
332 ***Xenopus* manipulation**. Standard procedures were used for ovulation, fertilisation, and
333 manipulation and incubation of embryos[31,32]. Briefly, frogs were obtained from Nasco
334 (Wisconsin, USA). Ovulation was induced by injecting serum gonadotropin (Intervet) and
335 chorionic gonadotropin (Intervet) into the dorsal lymph sac of mature female frogs. Eggs were
336 fertilised *in vitro* with sperm solution including 90% Leibovitz's L-15 medium (Gibco) and 10%
337 fetal calf serum (Gibco). After 10 min, fertilised eggs were de-jellied with 2.2% L-cysteine
338 (Sigma) equilibrated to pH 8.0. *X. tropicalis* embryos were cultured in 5% Marc's Modified
339 Ringer's solution[32] (MMR) at 21ºC-28ºC. *X. laevis* embryos were cultured in 10% Normal
340 Amphibian Medium[32] (NAM) at 14ºC-22ºC. Embryos were staged as previously described[33].
341 All *Xenopus* work complied fully with the UK Animals (Scientific Procedures) Act 1986 as
342 implemented by the Francis Crick Institute.
343
344 **Nucleic acid injections and treatments of embryos.** Microinjections were performed using
345 calibrated needles and embryos equilibrated in 4% Ficoll PM-400 (Sigma) in 5% MMR or 10%
346 NAM. Microinjection needles were generated from borosilicate glass capillaries (Harvard

347 Apparatus, GC120-15) using the micropipette puller Sutter p97. Maximally two nanolitres of
348 morpholino (MO) and/or mRNA were injected into de-jellied embryos at the 1-cell, 2-cell or 4-
349 cell stage using the microinjector Narishige IM-300. Injections for the mVegT LOF and rescue
350 were targeted to the vegetal hemisphere. All other injections were targeted to the animal
351 hemisphere. Embryos were transferred to fresh 5% MMR or 10% NAM (without Ficoll) once
352 they reached about stage 8.
353
354 The following mRNA amounts were used for ectopic expression: 80 pg and 400 pg *MyoD-HA*
355 mRNA into *X. tropicalis* and *X. laevis* embryos, respectively; and 400 pg *Sox3-HA* into *X.*
356 *laevis* embryos.
357
358 The following MO amounts and treatments were used for LOF experiments: 5 ng *Pou5f3.2*
359 MO, 5 ng *Pou5f3.3* MO and 5 ng standard control MO (mPouV LOF); 5 ng *Pou5f3.2* MO, 5 ng
360 *Pou5f3.3* MO and 5 ng *Sox3* MO (mPouV/Sox3 LOF); 5 ng *β-catenin* MO (Wnt LOF); 10 ng
361 m*VegT* MO (mVegT LOF); standard control MO (5-10 ng according to the dose used for the
362 β-catenin or mVegT LOF experiment); and 30 pg α-amanitin (Sigma). To block Nodal (Nodal
363 LOF) or BMP (BMP LOF) signalling, embryos were treated with 100 μM SB431542 or 10 μM
364 LDN19318 from the 8-cell stage onwards. Control embryos were treated accordingly with
365 DMSO, in which these antagonists were dissolved.
366
367 For the rescue of mVegT LOF, 25-50 pg z*VegT* mRNA were injected at the 2- or 4-cell stage.
368 For the rescue of mPouV/Sox3 LOF, 75 pg *Pou5f3.3* mRNA and 75 pg *Sox2* mRNA were
369 injected at the 2- or 4-cell stage. For both rescue experiment, 150-300 pg *mCherry* mRNA
370 was co-injected to trace mRNA expression in embryonic cells.
371
372 **Morpholino (MO) oligonucleotide sequences.** MOs were designed and produced by
373 Genetools (Oregon, USA) to block the translation of the following gene transcripts:
374 *β-catenin* MO (5'-TTTCAACAGTTTCCAAAGAACCAGG-3', underlined base was changed
375 from the original MO[21]), *Pou5f3.2* MO (5'-GCTGTTGGCTGTACATAGTGTC-3', underlined
376 base was changed from the original MO[20]), *Pou5f3.3* MO[20] (5'-
377 GTACAGAACGGGTTGGTCCATGTTC-3'), *Sox3* MO (5'-
378 GTCTGTGTCCAACATGCTATACATC-3'), *Tbx6* MO (5'-
379 TACATTGGGTGCAGGGACCCTCTCA-3') and m*VegT* MO[34] (5'-
380 TGTGTTCCTGACAGCAGTTTCTCAT-3'; see Extended Data Fig. 1h).
381
382 **RNA constructs.** All constructs are products of Gateway (Invitrogen) cloning. First, the coding
383 sequences of interest were amplified from either plasmid using *Pfu*Turbo DNA polymerase

384 (Agilent) or cDNA using Phusion HF DNA polymerase (NEB), and unidirectionally cloned into
385 the pENTR/D-TOPO vector (Invitrogen). Next, these clones were LR-recombined with
386 Gateway-compatible pCS2+ vectors containing no tag or a C-terminal 3xHA tag (see
387 Extended Data Fig. 5c). The following forward and reverse primers were used for PCR
388 amplification: *X. laevis myoD1a* from *MyoD*2-24 pBluescript plasmid[35], 5'-
389 CACCATGGAGCTGTTGCCCCCAC-3', 5'-TAAGACGTGATAGATGGTGCTG-3'; *X. laevis*
390 *sox3* from IMAGE:443920 clone, 5'-CACCATGTATAGCATGTTGGACAC-3', 5'-
391 TATGTGAGTGAGCGGTACAGTG-3'; *X. laevis sox2* from stage 11 cDNA, 5'-
392 CACCATGTACAGCATGATGGAGACCGA-3', 5'-TCACATGTGCGACAGAGGCAGC-3'; *X.*
393 *laevis pou5f3.3* from stage 11 cDNA, 5'-CACCATGGACCAGCCCATATTGTACA-3', 5'-
394 TCAGCCGGTCAGGACCCCCA-3'; *X. tropicalis* z*VegT* from stage 11 cDNA, 5'-
395 CACCATGCACTCTCTGCCGGATGTA-3', 5'-TTACCAACAGCTGTATGGAAAGA-3'.
396 pCS2+8N*mCherry*[36] (#34936) was obtained from Addgene.
397
398 ***In vitro* transcription of capped mRNA.** About 10 µg plasmid was linearised by restriction
399 digestion and purified using the QIAquick PCR purification kit (Qiagen). Capped mRNA was
400 produced from ~1 µg linearised plasmid using mMessage mMachine SP6 (Ambion) according
401 to the manufacturer's instructions. The following restrictions enzymes (NEB) were used for
402 linearising the pCS2+ constructs: *Asp*781 (*myoD1a*, *sox3*), *Pvu*II (*sox2*), *Psp*OMI (*pou5f3.3*,
403 z*VegT*) and *Not*I (*mCherry*).
404
405 **Generation of digoxigenin-labelled RNA probes.** DNA template for generating the *foxb1 in*
406 *situ* hybridisation probe was amplified from *X. tropicalis* stage 19-20 cDNA. KAPA HiFi
407 HotStart polymerase and the following PCR cycling conditions were used: 98ºC for 45 sec
408 followed by 40 cycles (98ºC for 10 sec, 63ºC for 10 sec, 72ºC for 15 sec) and final elongation
409 step of 20 sec at 72ºC. Primers were 5'-GTCAGCGCCTATGGAGTACC-3' and 5'-
410 AACACTGGAGATGCCATGC-3'. Fresh PCR product was zero-blunt cloned into the pCRII-
411 TOPO vector (Invitrogen). The identity and direction of insert was verified by restriction digest
412 and Sanger sequencing. Plasmids were linearised by restriction digestion and purified using
413 the QIAquick PCR purification kit (Qiagen). All *in situ* hybridisation probes were transcribed
414 from ~1 µg linearised plasmid using 1x digoxigenin-11-UTP (Roche), 40 U RiboLock RNase
415 inhibitor (Thermo Fisher Scientific), 1x transcription buffer (Roche) and T7 RNA polymerase
416 (Roche) at 37ºC for 2 h. The probe was treated with 2 U Turbo DNase (Ambion) to remove
417 the DNA template and purified by spin-column chromatography (Clontech). RNA was diluted
418 to 10 ng/µl (10x stock) with hybridisation buffer (50% formamide, 5x SSC, 1x Denhardt's, 10
419 mM EDTA, 1 mg/ml torula RNA, 100 µg/ml heparin, 0.1% Tween-20 and 0.1% CHAPS) and
420 stored at -20ºC. The following plasmids and restriction enzymes (NEB) were used for plasmid

421 linearisation: *X. laevis eomes* pCRII-TOPO[15], *Bam*HI; *X. tropicalis foxb1* pCRII-TOPO, *Bam*HI;
422 *X. tropicalis mir427* pCS108 (IMAGE:7545411)[37], *Sal*I; *X. laevis ventx2.1* pBluescript SK-[38],
423 *Eco*RI; *X. laevis Xbra* pSP73[39], *Bgl*II; and *X. tropicalis zic1* pCMV-SPORT6 (IMAGE:7668846),
424 *Sal*I.
425
426 **Animal cap assay.** These assays were carried out for the ectopic expression of mRNA
427 constructs in *X. laevis* and for the MO-mediated LOF in *X. tropicalis*. All animal caps were
428 dissected at the blastula stage (stage 8 to 9) and cultured in 75% NAM containing 0.1% bovine
429 serum albumin (Sigma). Dissections were carried out with 13 µm loop electrodes (Protech
430 International, 13-Y1) connected to a MC-2010 micro cautery instrument (Protech International)
431 operating at power level 2. As illustrated in Extended Data Fig. 5c, control (uninjected) and
432 MyoD-HA or Sox3-HA expressing animal caps were cultured at 20ºC until sibling embryos
433 reached early neurula stage (stage 13). For the experiment shown in Fig. 3e and Extended
434 Data Fig. 7h, animal caps of control and mPouV/Sox3 LOF embryos were cultured with or
435 without 10 ng/ml recombinant human activin A (R&D Systems, 338-AC) (Nodal agonist) or 50
436 µM CHIR99021 (Tocris, 4423) (canonical Wnt agonist) at 22ºC for ~2 h until sibling embryos
437 reached early gastrula stage (stage 10^+).
438
439 **Extraction of total RNA.** 10-15 embryos (or 15-20 animal caps) were homogenised in 800
440 (400) µl TRIzol reagent (Ambion) by vortexing. The homogenate was either snap-frozen in
441 liquid nitrogen and stored at -80ºC or processed immediately. For phase separation, the
442 homogenate together with 0.2x volume of chloroform was transferred to pre-spun 1.5-ml
443 Phase Lock Gel Heavy (PLGH) microcentrifuge tubes, shaken vigorously for 15 sec, left on
444 the bench for 2 min and spun at ~16,000 g (4ºC) for 5 min. The upper phase was mixed well
445 with one volume of 95-100% ethanol and spun through the columns of Zymo RNA Clean and
446 Concentrator 5 or 25 (Zymo Research) at ~12,000 g for 30 sec. Next, the manufacturer's
447 instructions were followed for the recovery of total RNA (>17 nt) with minor modifications. First,
448 the flow-through of the first spin was re-applied to the column. Second, the RNA was treated
449 in-column with 3 U Turbo DNase (Ambion). Third, the RNA was eluted twice with 25 µl
450 molecular-grade water. The concentration was determined on the NanoDrop 1000
451 spectrophotometer.
452
453 **Reverse transcription (RT).** About 0.2-1 µg total RNA was denatured at 75ºC for 5 min before
454 setting up the RT reaction including 40 U M-MLV RT, RNase H minus, point mutant (Promega)
455 500 µM of each dNTP (NEB), and 10 µM random hexamers (Sigma) in a final volume of 10
456 µl. The following incubation settings were used: 25ºC for 15 min, 37ºC for 15 min, 55ºC for 45

457 min, 85ºC for 15 min and 4ºC for indefinite. The reaction was diluted with 90-190 µl molecular-
458 grade water before using 1-2.5 µl for each qPCR reaction.
459
460 **qPCR primer sequences.** The PCR primers were designed to hybridize at ~60ºC (T_m) and to
461 generate 75-125 bp amplicons. Where possible, primers were checked for their specificity *in*
462 *silico* via the UCSC Genome Browser. All primer pairs were tested to produce one amplicon
463 of the correct size through diagnostic PCR using RedTaq DNA polymerase (Sigma).
464 Exceptions were previously designed primer pairs for *X. laevis actc1*, *odc1* and *otx2*, all of
465 which generate amplicons of 200-250 bp.
466
467 Primers (forward, reverse) for *X. laevis* (L) or *X. tropicalis* (T) cDNA: *actc1* (L, 5'-
468 TCCTGTACGCTTCTGGTCGTA-3', 5'-TCTCAAAGTCCAAAGCCACATA-3'); *eomes* (T, 5'-
469 ATTGACCACAACCCATTTGC-3', 5'-TCTAGGAGAATCCGCAGGAG-3'); *foxb1* (T, 5'-
470 CCGGATGCAGAGTGACAAC-3', 5'-GAGGCTTCTGGTCGCTGTAG-3'); *gata4* (T, 5'-
471 TCCTCTGATCAAGCCTCAAAG-3', 5'-ACGGCGCCAGAGTGTAGTAG-3'); *myl1* (L, 5'-
472 GCACCTGCATTTGATCTGTC-3', 5'-TTGCTGTCTCCTGTCCTGTC-3'); *nodal5* (T, 5'-
473 TTCAAGCCAACAAACCACGC-3', 5'-CGCATCTTCACCGGTATGCA-3'); *nodal6* (T, 5'-
474 CGCATTTCGTTGTGAGGGGT-3', 5'-TCGTACAGCTTGACCAAACTCT-3'); *odc1* (L, 5'-
475 GCCATTGTGAAGACTCTCTCC-3', 5'-TTCGGGTGATTCCTTGCCAC-3'); *odc1* (T, 5'-
476 GGGCAAAAGAGCTTAATGTGG-3', 5'-CATCGTGCATCTGAGACAGC-3'); *otx2* (L, 5'-
477 TTGAACCAGACCTGGACT-3', 5'-CGGGATGGATTTGTTGCA-3'); *rpl8* (T, 5'-
478 CTGGCTCCAAGAAGGTCATC-3', 5'-CAGGATGGGTTTGTCAATACG-3'); *sox2* (L, 5'-
479 ACATCACATGGACGGTTGTC-3', 5'-TCTCTCGGTTTCTGCTCTCG-3'); *sox17b* (T, 5'-
480 GTTTATGGTGTGGGCAAAGG-3', 5'-TCAGCGATTTCCATGATTTG-3'); *tbx6* (L, 5'-
481 GGGGCACCACGACTCTTAC-3', 5'-CCCAGAATCCTCTGTGAAGG-3'); and *Xbra* (T, 5'-
482 CCTGTGGATGAGGTTCAAGG-3', 5'-CACGCTCACCTTTAGAACTGG-3').
483
484 Primers (forward, reverse) for *X. tropicalis* genomic DNA: *β-actin* -0.5kb (5'-
485 TTGTACGCAACACAATGCAG-3', 5'- AAATGTCAGGACCAAATGCAG-3'); *dlc* -3.2kb (5'-
486 CATCCTGCAATTAACTGTCTGC-3', 5'-AGGTGTGACAAATGGCACTG-3'); *gdf3* -0.1kb (5'-
487 TGGATGAGCACAGAGAGGTG-3', 5'-GTCAGGAGGGAGGTGTCAAC-3'); *hoxd1* +3.2kb (5'-
488 TGTTGTAGATGCTGATGCTTATCG-3', 5'-AACAGAAAATCAAAGGCTTGCA-3'); *mespa* -
489 1.3kb (5'-CTCCTTTGCCCTCTGAAATG-3', 5'-CTTCAGGGAATTGGCTTGTG-3'); *msgn1* -
490 0.1kb (5'-CCAGTCCATTTTCCATGTTG-3', 5'-GGGGAGGCCCTTTTATACAG-3'); *myf5* -
491 1.1kb (5'-AAAGAGTTCCTGCACCTTGG-3', 5'-TCCCAAATCATCAGCACAAC-3'); *nodal6* -
492 0.1kb (5'-ACATGAGTTTGGCCTTGGTC-3', 5'-AGGTAACCCCTGAGCTGTCA-3'); *pax3*
493 5'UTR (5'-CAGATGCCAGAGGAGAGAGC-3', 5'-GGAACTCTCTGGGTCCCAAC-3'); *pax3*

494 +2.7kb (5'-ATGAGAATGTGAGCGACACG-3', 5'-AAACCCAAGGTGAGCAAGTG-3'); *sia1* -
495 0.2kb (5'-TGGCTTTTAAGTTCGTCTTGG-3', 5'-CAAAGGGCTTCTCTTTGTGC-3'); *sox3*
496 TSS (5'-GATGGGGCATTAATGGACAG-3', 5'-CACGGGAGAAAATGAAGCTG-3'); *sox17b* -
497 0.1kb (5'-GTTAGACTCCGCTCCAGACG-3', 5'-TCTCCACCTAAGGGGAAACC-3'); *wnt8a* -
498 5.7kb (5'-GGAAACTGCCTGTGGAAATG-3', 5'- ACAAAAGCCCACAGGAGATG-3'); *Xbra* -
499 4kb (5'-CAATGGATGCCTCTCTGGAC-3', 5'-TGGGTTTAGAATGGCAAAGG-3'); and *Xbra*
500 TSS (5'-CACACCTTGGGTTTTGTTCC-3', 5'-TTCCTCAAGAGAGGGGCTTC-3').
501
502 **Whole mount *in situ* hybridisation (WMISH)**. WMISH was conducted using digoxigenin-
503 labelled RNA probes. It was based on previously published protocols[32,40]. *X. tropicalis*
504 embryos were fixed in MEMFA (1x MEM and 3.7% formaldehyde) at room temperature for 1
505 h. The embryos were then washed once in 1x PBS and two to three times in ethanol. Fixed
506 and dehydrated embryos were kept at -20ºC for ≥24 h to ensure proper dehydration before
507 starting hybridisation. Dehydrated embryos were washed once more in ethanol before
508 rehydrating them in two steps to PBT (1x PBS and 0.1% Tween-20). Embryos were treated
509 with 5 µg/ml proteinase K (Ambion) in PBT for 6 to 8 min, washed briefly in PBT, fixed again
510 in MEMFA for 20 min and washed three times in PBT. Embryos were transferred into baskets,
511 which are kept in an 8x8 microcentrifuge tube holder sitting inside a 10x10 slot plastic box
512 filled with PBT. Baskets were built by replacing the round bottom of 2-ml microcentrifuge tubes
513 with a Sefar Nitex mesh. This container system was used to readily process several batches
514 of embryos at once. These baskets were maximally loaded with 40 to 50 *X. tropicalis* embryos.
515 The microcentrifuge tube holder was used to transfer all baskets at once and to submerge
516 embryos into subsequent buffers of the WMISH protocol. Next, the embryos were incubated
517 in 500 µl hybridisation buffer (50% formamide, 5x SSC, 1x Denhardt's, 10 mM EDTA, 1 mg/ml
518 torula RNA, 100 µg/ml heparin, 0.1% Tween-20 and 0.1% CHAPS) for 2 h in a hybridisation
519 oven set to 60ºC. After this pre-hybridisation step, the embryos were transferred into 500 µl
520 digoxigenin-labelled probe (1 ng/µl) preheated to 60ºC and further incubated overnight at
521 60ºC. The pre-hybridisation buffer was kept at 60ºC. The next day embryos were transferred
522 back into the pre-hybridisation buffer and incubated at 60ºC for 10 min. Subsequently, they
523 were washed three times in 2x SSC/0.1% Tween-20 at 60ºC for 10 min, twice in 0.2x
524 SSC/0.1% Tween-20 at 60ºC for 20 min and twice in 1x maleic acid buffer (MAB) at room
525 temperature for 5 min. Next, the embryos were treated with blocking solution (2% Boehringer
526 Mannheim blocking reagent in 1x MAB) at room temperature for 30 min, and incubated in
527 antibody solution (10% lamb or goat serum, 2% Boehringer Mannheim blocking reagent, 1x
528 MAB and 1:2,000 Fab fragments from polyclonal anti-digoxigenin antibodies conjugated to
529 alkaline phosphatase) at room temperature for 4 h. The embryos were then washed four times
530 in 1x MAB for 10 min before leaving them in 1x MAB overnight at 4ºC. On the final day of the

531 WMISH protocol, the embryos were washed another three times in 1x MAB for 20 min and
532 equilibrated to working conditions of alkaline phosphatase (AP) for a total of 10 min by
533 submerging embryos twice into AP buffer (50 mM $MgCl_2$, 100 mM NaCl, 100 mM Tris-HCl pH
534 9.5 and 1% Tween-20). At this stage, the embryos were transferred to 5-ml glass vials for
535 monitoring the progression of the AP-driven colorimetric reaction. Any residual AP buffer was
536 discarded before adding 700 µl staining solution (AP buffer, 340 µg/ml nitro-blue tetrazolium
537 chloride, 175 µg/ml 5-bromo-4-chloro-3'-indolyphosphate). The colorimetric reaction was
538 developed at room temperature in the dark. Once the staining was clear and intense enough,
539 the colour reaction was stopped by two washes in 1x MAB. To stabilise and preserve
540 morphological features, the embryos were fixed with Bouin's fixative without picric acid (9%
541 formaldehyde and 5% glacial acetic acid) at room temperature for 30 min. Next, the embryos
542 were washed twice in 70% ethanol/PBT to remove the fixative and residual chromogens. After
543 rehydration to PBT in two steps, the embryos were treated with weak Curis solution (1% H_2O_2,
544 0.5x SSC and 5% formamide) at 4ºC in the dark overnight. Finally, the embryos were washed
545 twice in PBS before imaging them in PBS on a thick agarose dish by light microscopy.
546
547 **Antibody generation.** Affinity-purified rabbit polyclonal antibodies were generated by sdix
548 (Delaware, USA) against different amino acid (aa) sequences of *X. tropicalis* Tbx6 avoiding
549 its highly conserved DNA binding domain (aa 95-280) among T-box proteins: aa 1-94 (#4596),
550 aa 304-403 (#5061) and aa 418-517 (#5107). Antibody #5107 did not work (data not shown).
551 Antibodies #4596 and #5061 were suitable for immunoprecipitation and Western blotting to
552 detect endogenous Tbx6 in both *X. tropicalis* and *X. laevis* embryos (Extended Data Fig. 2f-
553 h). Antibody #4596 complied with ENCODE guidelines[41] and thus was suitable for ChIP
554 (Extended Data Fig. 2i).

555 **Immunoprecipitation (IP).** For each IP, 40 de-jellied embryos were collected at the
556 developmental stage of interest. They were homogenised in 400 µl ice-cold IP buffer (50 mM
557 Tris-HCl pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 1 mM EDTA and 0.25 % Igepal CA-630)
558 supplemented with an EDTA-free protease inhibitor cocktail (Roche). The extract was kept on
559 ice for 5 min before adding 400 µl 1,1,2-trichloro-1,2,2-trifluoroethane (Sigma). The extract
560 was mixed with Freon by vigorously inverting the tube and cleared by centrifugation at 16,000
561 g (4ºC) for 5 min. An aliquot of 20 µl supernatant was saved as input. The remaining
562 supernatant was incubated with 0.5 µl antibody (anti-Sox3[42], anti-VegT[9] or anti-Tbx6; all rabbit
563 polyclonal antibodies) at 4ºC for 1 h. Subsequently, each embryonic extract was
564 supplemented with 20 µl washed magnetic beads conjugated to sheep anti-rabbit IgG
565 (Dynabeads M280, Invitrogen) and kept on a tube rotator at 4ºC for another 2 h. Next, the
566 beads were washed three times with IP buffer. The immunoprecipitate was eluted off the

567 beads with 20 µl 1x SDS loading buffer (50 mM Tris-HCl pH 6.8, 2% SDS, 6% glycerol and
568 5% β-mercaptoethanol). Western blotting was carried out with TrueBlot (Rockland
569 Immunochemicals) and normal horseradish peroxidase (HRP)-conjugated secondary
570 antibodies for IP and input samples, respectively.

571 **Western blotting.** Protein samples were denatured in 1x SDS loading buffer (final
572 concentration) at 70ºC for 5 min. Denatured samples were run alongside a standard protein
573 ladder into pre-cast gradient gels (Any kD Mini-PROTEAN TGX, BioRad) at constant 200 V
574 for 1 h. Size separated proteins were immediately transferred onto hydrophobic Immobilon-P
575 PVDF transfer membranes (Millipore) at constant 100 V for 30 min using standard protein
576 electrophoresis equipment. The membranes were blocked with 5% milk powder in PBS/0.1%
577 Tween-20 (MPBTw) at room temperature for 30 min. Next, the membranes were incubated at
578 room temperature for 1 h with the primary antibodies diluted in MPBTw: 1:5,000 of mouse
579 monoclonal anti-α-tubulin (Sigma, T5168), 1:2,000 rabbit polyclonal anti-Sox3[42], 1:2,000
580 rabbit polyclonal anti-VegT[9] or 1:2,000 rabbit polyclonal anti-Tbx6 (#5061). The membranes
581 were washed three times with PBST for 10 min before applying 1:1,000 TrueBlot HRP-
582 conjugated anti-rabbit IgG (Rockland Immunochemicals, 18-8816-31) or 1:2,000 normal goat
583 HRP-conjugated anti-mouse IgG (Thermo Fisher Scientific, 31430) in MPBTw at room
584 temperature. After 1 h, the membranes were washed three times in PBTw for 10 min and once
585 in PBS for 5 min. Finally, the membranes were treated with UptiLight US HRP WB reagent
586 (interchim) for 1 min in the dark. HRP-linked chemiluminescence was detected on the
587 ChemiDoc XRS+ system (BioRad).

588 **Whole-mount immunohistochemistry (WMIHC).** Embryos were fixed with MEMFA (0.1 M
589 MOPS, 2 mM EGTA, 1 mM $MgSO_4$ and 3.7% formaldehyde) in capped 5-ml glass vials at
590 room temperature for 2 h, dehydrated and stored in absolute ethanol at -20ºC for ≥24 h. For
591 sagittal sections, fixation was stopped after 30 min. Embryos were placed on a piece of
592 flattened Bostik Blu-Tack in a large Petri dish filled with PBS to bisect them with a scalpel.
593 Following a three-step rehydration (50%, 75% and 100% PBS) the embryos were bleached
594 with weak Curis solution (1% H_2O_2, 0.5x SSC and 5% formamide) on a light box at room
595 temperature for 2 h or in the dark at 4ºC overnight. Bleached embryos were washed three
596 times with PBS/0.3% Triton X-100 (PBT) and pre-incubated in blocking solution (PBS, 20%
597 goat or donkey serum, 2% BSA and 0.1% Triton X-100) at 4ºC for 6 h. The embryos were
598 then transferred to 2-ml round-bottom microcentrifuge tubes before discarding all remaining
599 blocking solution. The embryos were incubated at 4ºC for 1-3 days in 50 µl blocking solution
600 containing the primary antibody at the following dilutions: 1:1,000 rabbit polyclonal anti-Sox3[42],
601 1:500 rabbit polyclonal anti phospho-Smad1/5/9 (Cell Signalling, #9511) or 1:500 goat

602 polyclonal Smad2/3 (R&D Systems, AF3797). Afterwards, the embryos were transferred back
603 to capped 5-ml glass vials, washed three times with RIPA buffer (50 mM HEPES pH 7.5, 0.5
604 M LiCl, 1 mM EDTA, 1% Igepal CA-630 and 0.7% sodium deoxycholate) for 1 h and once with
605 PBT for 5 min. Next, the embryos were incubated with the secondary HRP-conjugated
606 antibody diluted in blocking solution at 4ºC overnight: 1:400 goat anti-rabbit IgG-HRP
607 (ThermoFisher Scientific, G-21234) or 1:200 donkey anti-goat IgG-HRP (Santa Cruz
608 Biotechnology, sc-2020). The washes of the previous day were repeated before pre-
609 incubating the embryos in 400 µl inactive 3,3'-diaminobenzidine tetrahydrochloride with cobalt
610 (DAB with metal enhancer; Sigma). After 1-2 min, the inactive solution was replaced with the
611 H_2O_2-activated DAB solution. The HRP reaction was stopped after 40 sec by washing the
612 embryos several times with PBT. In Extended Data Fig. 1g, embryos were dehydrated in three
613 steps to absolute methanol and cleared with Murray's clear (2:1 benzyl benzoate/benzyl
614 alcohol) on a glass depression slide.

615 **Deep sequencing and quality filter.** All deep sequencing libraries were quality controlled:
616 The DNA yield and fragment size distribution were determined by fluorometry and Chip-based
617 capillary or polyacrylamide gel electrophoresis, respectively. Libraries were sequenced on the
618 Illumina platforms GAIIx or HiSeq by the Advanced Sequencing Facility of the Francis Crick
619 Institute to produce single or paired-end reads of at least 40 bases. Next-generation capture-
620 C libraries were sequenced on MiSeq with a read length of 150 bases. Sequencing samples
621 and read alignment results are summarised in Supplementary Table 3. The metrics of paired-
622 end alignments such as insert size mean and standard deviation were determined by Picard
623 (CollectInsertSizeMetrics) from the Broad Institute (USA).

624 **Chromatin immunoprecipitation (ChIP).** ChIP was carried out as detailed previously[43].
625 Briefly, de-jellied *X. tropicalis* embryos were treated with 1% formaldehyde (Sigma, F8775) in
626 1% MMR at room temperature for 15-45 min to cross-link chromatin proteins to nearby
627 genomic DNA. Duration of fixation was determined empirically and depended mainly on the
628 developmental stage and antibody epitopes[43]. Fixation was terminated by rinsing embryos
629 three times with ice-cold 1% MMR. Where required, post-fixation embryos were dissected to
630 select specific anatomical regions in ice-cold 1% MMR. Fixed embryos were homogenised in
631 CEWB1 (10 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 1% Igepal CA-630, 0.25%
632 sodium deoxycholate and 0.1% SDS) supplemented with 0.5 mM DTT, protease inhibitors
633 and, if using phospho-specific antibodies, phosphatase blockers (0.5 mM orthovanadate and
634 2.5 mM NaF). To solubilise yolk platelets and separate them from the nuclei, the homogenate
635 was left on ice for 5 min and then centrifuged at 1,000 g (4ºC) for 5 min. Homogenisation and
636 centrifugation was repeated once before resuspending the nuclei containing pellet in 1-3 ml

637 CEWB1. Nuclear chromatin was solubilised and fragmented by isothermal focused or
638 microtip-mediated sonication. The solution of fragmented chromatin was cleared by
639 centrifuging at 16,000 g (4ºC) for 5 min. Where required, ~1% of the cleared chromatin extract
640 was set aside for the input sample (negative control for ChIP). ChIP-grade antibodies were
641 used to recognise specific chromatin features and to enrich these by coupling the antibody-
642 chromatin complex to protein G magnetic beads and extensive washing. These steps were
643 carried out at 4ºC. The beads were washed twice in CEWB1, twice in WB2 (10 mM Tris-HCl
644 pH 8.0, 500 mM NaCl, 1 mM EDTA, 1% Igepal CA-630, 0.25% sodium deoxycholate and 0.1%
645 SDS), twice in WB3 (10 mM Tris-HCl pH 8.0, 250 mM LiCl, 1 mM EDTA, 1% Igepal CA-630
646 and 1% sodium deoxycholate) and once in TEN (10 mM Tris-HCl pH 8.0, 150 mM NaCl and
647 1 mM EDTA). ChIP was eluted off the beads twice with 100 µl SDS elution buffer (50 mM Tris-
648 HCl pH 8.0, 1 mM EDTA and 1% SDS) at 65ºC. ChIP eluates were pooled before reversing
649 DNA-protein cross-links. Input (filled up to 200 µl with SDS elution buffer) and ChIP samples
650 were supplemented with 10 µl 5 M NaCl and incubated at 65ºC for 6-16 h. Samples were
651 treated with proteinase K and RNase to remove any proteins and RNA from the co-
652 immunoprecipitated DNA fragments. The DNA was purified with phenol:chloroform:isoamyl
653 alcohol (25:24:1, pH 7.9) using 1.5-ml Phase Lock Gel Heavy (PLGH) microcentrifuge tubes
654 for phase separation and precipitated with 1/70 volume of 5 M NaCl, 2 volumes of absolute
655 ethanol and 15 µg GlycoBlue (Ambion). After centrifugation, the DNA pellet was air-dried and
656 dissolved in 11 µl elution buffer (10 mM Tris-HCl pH 8.5). The DNA concentration was
657 determined on a fluorometer using high-sensitivity reagents for double-stranded DNA (10
658 pg/µl to 100 ng/µl).

659 The following antibodies were used for ChIP: rabbit polyclonal anti-β-catenin (Santa-Cruz, H-
660 102), rabbit polyclonal anti-H3K4me1 (Abcam, ab8895), rabbit polyclonal anti-HA (Abcam,
661 ab9110), mouse monoclonal anti-RNAPII (Covance, MMS-126R), rabbit polyclonal anti-
662 phospho-Smad1/5/9 (Cell Signalling, #9511), goat polyclonal anti-Smad2/3 (R&D Systems,
663 AF3797), rabbit polyclonal anti-Sox3[42], rabbit polyclonal anti-Tbx6 (#4596) and rabbit
664 polyclonal anti-VegT[9].

665 **Quantitative PCR (qPCR).** DNA was amplified in technical duplicates with SYBR Green I
666 master mix (Roche) on a Light Cycler 480 II (Roche) cycling 55-times between 94ºC, 60ºC
667 and 72ºC with each temperature step running for 10 sec and switching at +4.8ºC/sec and -
668 2.5ºC/sec. At the end qPCR reactions were heated from 65ºC to 97ºC with a gradual increase
669 of 0.11ºC/sec (melting curve) to ensure only fluorescence was collected from one specific
670 amplicon. ChIP-qPCR results were based on absolute quantification using an 8-point standard
671 curve of three-fold dilutions of ~1% input DNA (Extended Data Fig. 2b-e,i). RT-qPCR results

672 were normalised to the housekeeping gene *odc1* (Fig. 3e; Extended Data Fig. 5d and 7h) or
673 *rpl8* (Extended Data Fig. 6d), and scaled relative to control embryos using the $2^{-\Delta\Delta Ct}$ method[44].
674 The threshold cycle (C_t) was derived from the maximum acceleration of SYBR fluorescence
675 (second derivative maximum method).

676 **ChIP-Seq library preparation.** 0.5 to 10 ng ChIP DNA or 5 ng input DNA were used to
677 prepare single (only Tbx6 ChIP-Seq) or indexed paired-end libraries as previously
678 described[43,45,46].
679
680 **Post-sequencing analysis of ChIP-Seq.** Single reads of maximal 50 bases were processed
681 using trim_galore v0.4.2 (Babraham Institute, UK) to trim off low-quality bases (default Phred
682 score of 20, i.e. error probability was 0.01) and adapter contamination from the 3' end.
683 Processed reads were aligned to the *X. tropicalis* genome assembly v7.1 (and v9.1 for some
684 ChIP-Seq data) running Bowtie2 v2.2.9[47] with default settings (Supplementary Table 3).
685 Alignments were converted to the HOMER's tag density format[48] with redundant reads being
686 removed (makeTagDirectory -single -tbp 1 -unique -mapq 10 -fragLength 175 -totalReads all).
687 Only uniquely aligned reads (i.e. MAPQ ≥10) were processed. We pooled all input alignments
688 from various developmental stages. This created a comprehensive mappability profile that
689 covered ~400 million unique base pair positions. All chromatin profiles were position-adjusted
690 and normalised to the effective total of 1 million aligned reads including multimappers (counts
691 per million aligned reads, CPM). Agreeing biological replicates according to ENCODE
692 guidelines[41] were subsequently merged. For stage 10^+ β-catenin and Smad2 ChIP-Seq,
693 external datasets[20,49,50] were used as biological replicates. HOMER's peak caller was used to
694 identify transcription factor binding sites by virtue of ChIP-mediated read enrichments
695 (hereafter called peaks): findpeaks -style factor -minDist 175 -fragLength 175 -
696 inputFragLength 175 -fdr 0.001 -gsize 1.435e9 -F 3 -L 1 -C 0.97. This means that both ChIP
697 and input alignments were extended 3' to 175 bp for the detection of significant (0.1% FDR)
698 peaks being separated by ≥175 bp. The effective size of the *X. tropicalis* genome assembly
699 v7.1 was set to 1.435 billion bp, an estimate obtained from the mappability profile. These
700 peaks showed equal or higher tag density than the surrounding 10 kb, ≥3-fold more tags than
701 the input and ≥0.97 unique tag positions relative to the expected number of tags. To detect
702 focal RNAPII recruitment to putative CRMs and avoid calling peaks within broad regions of
703 RNAPII reflecting transcript elongation, the threshold of focal ratio and local enrichment within
704 10 kb was elevated to 0.6 and 3 (-L 3), respectively. To further eliminate any false positive
705 peaks, we removed any peaks with <0.5 (TFs including signal mediators) or <1 (RNAPII) CPM
706 and those falling into blacklisted regions showing equivocal mappability due to genome
707 assembly errors, gaps or simple/tandem repeats. Regions of equivocal mappability were

708 identified by a two-fold lower (poor) or three-fold higher (excessive) read coverage than the
709 average detected in 400-bp windows sliding at 200-bp intervals through normalised ChIP input
710 and DNase-digested naked genomic DNA. All identified regions ≤800 bp apart were
711 subsequently merged. Gap coordinates were obtained from the Francis Crick UCSC genome
712 browser (http://genomes.crick.ac.uk). Simple repeats were masked with RepeatMasker v4.0.6
713 using the crossmatch search engine v1.090518 and the following settings: RepeatMasker -
714 species "xenopus silurana tropicalis" -s -xsmall. Tandem repeats were masked with Kent's
715 trfBig wrapper script of the Tandem Repeat Finder v4.09[51] using the following settings: weight
716 for match, 2; weight for mismatch, 7; delta, 7; matching probability, 80; indel probability, 10;
717 minimal alignment score, 50; maximum period size, 2,000; and longest tandem repeat array
718 (-l), 2 [million bp]. Multi-genome sequence conservation tracks (phastCons and phyloP) for *X.*
719 *tropicalis* genome assembly v9.1 were obtained from Xenbase[52]. The following 11 vertebrate
720 species were used to evaluate sequence similarity: *X. tropicalis*, *X. laevis*, *Nanorana parkeri*
721 (High Himalaya Frog), *Fugu rubripes* (Japanese Pufferfish), *Chrysemys picta* (Painted Turtle),
722 *Gallus gallus* (Chicken), *Anolis carolinensis* (Green Anole lizard), *Monodelphis domestica*
723 (Gray Short-tailed Opossum), *Canis lupus familiaris* (dog), mouse and human. CRMs with a
724 phastCons average ≥0.4 were considered 'conserved'.
725
726 **Detecting zygotic genes by means of RNAPII chromatin profiles.** RNAPII and input tag
727 densities were calculated across the gene body in 10 bins of equal size. Gene annotations
728 v7.1 were altered based on a few known zygotic isoforms[37] and some corrections obtained
729 from assembling total and poly(A) RNA[37] from stage 6 to stage 12.5 *de novo* as outlined
730 elsewhere[53]. A few genes had previously been annotated as gene clusters due to assembly
731 uncertainties. We reduced the annotation of polycistronic *mir427* to the minus arm
732 (scaffold_3b:3516900-3523400) and only monitored *nodal3.5* and *nodal5.3* within their
733 respective gene clusters. Gene bodies with <40% mappability were removed. Here, the
734 threshold of mappability per bin was set at 10% of the input read density averaged across all
735 gene bodies in use. Subsequently, enrichment values were only obtained for all mappable
736 bins by dividing read densities of RNAPII and input. Further, we restricted the analysis to
737 genes for which ≥3 transcripts per million (TPM) could be detected on average over 3
738 consecutive time points (i.e. over the developmental time of 1 h) of a high-resolution profile of
739 total RNA[37] from fertilisation to after gastrulation (stage 13). Genes were considered activated
740 at a certain developmental stage if ≥0.1 TPM were detected over 3 consecutive time points
741 within a 2 h window of the developmental stage in question and either of the following RNAPII
742 enrichments covering ≥80% of the mappable gene body were reached: (1) 2.6-fold, (2) 1.8-
743 fold and 1.4-fold at the next or previous stage, (3) 1.4-fold and 1.8-fold at the next or previous
744 stage, or (4) 1.4-fold over 3 consecutive stages.

745

746 **Poly(A) RNA-Seq.** Libraries were made from ~1 µg total RNA by following the low sample
747 protocol of the TruSeq RNA sample preparation guide version 2 with a few modifications. First,
748 1 µl cDNA purified after second strand synthesis was quantified on a fluorometer using high-
749 sensitivity reagents for double-stranded DNA (10 pg/µl to 100 ng/µl). By this stage, the yield
750 was ~10 ng. Second, the numbers of PCR cycles were adjusted to the detected yield of cDNA
751 to avoid products of over-amplification such as chimera fragments: 7 (~20 ng), 8 (~10 ng), 9
752 (~5 ng) and 12 (~1 ng).
753

754 **RNA-Seq read alignment.** Paired-end reads were aligned to the *X. tropicalis* genome
755 assembly v7.1 using STAR v2.5.3a[54] with default settings and a revised version of gene
756 models v7.2[37] to improve mapping accuracy across splice junctions. The alignments were
757 sorted by read name using the sort function of samtools v1.3.1. Exon and intron counts (-t
758 'exon;intron') were extracted from unstranded (-s 0) alignment files using VERSE v0.1.5[55] in
759 featureCounts (default) mode (-z 0). Intron coordinates were adjusted to exclude any overlap
760 with exon annotation. For visualisation, genomic BAM files of biological replicates were
761 merged using samtools v1.3.1[56] and converted to the bigWig format. These genome tracks
762 were normalised to the wigsum of 1 billion excluding any reads with mapping quality <10 using
763 the python script bam2wig.py from RSeQC v2.6.4[57].
764

765 **Differential gene expression analysis.** Differential expression analysis was performed with
766 both raw exon and intron counts excluding those belonging to ribosomal and mitochondrial
767 RNA using the Bioconductor/R package DESeq2 v1.14.1[58]. In an effort to find genes with
768 consistent fold changes over time, p-values were generated according to a likelihood ratio test
769 reflecting the probability of rejecting the reduced (~ developmental stage) over the full (~
770 developmental stage + condition) model. Resulting p-values were adjusted to obtain false
771 discovery rates (FDR) according to the Benjamini-Hochburg procedure with thresholds on
772 Cook's distances and independent filtering being switched off. Equally, regional expression
773 datasets[24] without time series were subjected to likelihood ratio tests with reduced (~1) and
774 full (~ condition) models for statistical analysis. Fold changes of intronic and exonic transcript
775 levels were calculated for each developmental stage and condition using the mean of
776 DESeq2-normalised read counts from biological replicates. Both intronic and exonic datasets
777 were filtered for ≥10 DESeq2-normalised read counts detected at least at one developmental
778 stage in all uninjected or DMSO-treated samples. Gene-specific fold changes were removed
779 at developmental stages that yielded <10 normalised read counts in corresponding control
780 samples. Next, the means of intronic and exonic fold changes were calculated across
781 developmental stages. The whole dataset was cut down to 3,318 genes for which at least 50%

782 reductions (FDR <10%) in exonic (default) or intronic counts could be detected in α-amanitin-
783 injected embryos. Regional expression was based on exonic read counts by default unless
784 the intronic fold changes were significantly (FDR <10%) larger than the exonic fold changes
785 (Supplementary Table 2).
786
787 **Analysing ribosome footprinting and mass-spectrometry data.** The ribosome footprinting
788 reads[6] were trimmed 5' by 8 bases and 3' by as many bases overlapping with the adapter
789 sequence 5'-TCGTATGCCGTCTTCTGCTTG-3' from the 5' end. All trimmed reads between
790 27 to 32 bases in length were aligned first to ribosomal RNA as listed in the SILVA rRNA
791 database[59] using Bowtie v1.0.1[60] with the following parameters: --seedlen 25 (seed length) --
792 seedmms 1 (number of mismatches allowed in the seed) --un (unaligned reads were
793 reported). All non-aligned (rRNA-depleted) reads were mapped to the gene model 6.0 of *X.*
794 *laevis* using Tophat2 v2.0.10[61] --no-novel-juncs (spliced reads must match splice junctions of
795 gene model 6.0) --no-novel-indels (indel detetion disabled) --segment-length 15 (minimal
796 length of read fragment to be aligned) –GTF v6.0.gff3 --prefilter-multihits (reads aligned first
797 to whole genome to exclude reads aligning >10 times) --max-multihits 10. Alignment files were
798 converted to the HOMER[48] tag density format before retrieving reads for each CDS per
799 kilobase and one million mapped reads (rpkm) using HOMER's perl script analyzeRNA.pl.
800 This read count table was merged with a published list of estimated protein concentrations
801 (nM) in the *X. laevis* egg using mass-spectrometry[5] (Supplementary Table 1). Chromatin-
802 associated proteins regulating RNAPII-mediated transcription were filtered based on human
803 (version 2.0, 09/2014) and Xenbase-released gene ontology (GO) associations. The genes
804 associated with any of the following GO terms were verified with UniProt[62] whether chromatin
805 binding is supported by functional evidence: chromatin (GO:0000785), DNA binding
806 (GO:0003677), DNA-templated regulation of transcription (GO:0006355), sequence-specific
807 DNA binding and TF activity (GO:0003700) and DNA replication (GO:0006260). We separated
808 specific TFs from all other chromatin binding proteins to form three categories: (1) sequence-
809 specific DNA binding factors, (2) other RNAPII-regulating factors and (3) remaining genes,
810 whereby genes associated with DNA repair (GO:0006281) and RNAPIII (GO:0006383) were
811 moved to the third category.
812
813 **SPRI-based enrichment of small DNase-digested fragments (DNase-Seq).** In our hands,
814 the high yolk content in vegetal blastomeres made it impossible to employ ATAC-Seq on whole
815 *Xenopus* embryos before the onset of gastrulation. Therefore, DNase-Seq was adapted to
816 early *Xenopus* embryos with a novel approach to select small DNase-digested DNA
817 fragments. Ultracentrifugation- or gel electrophoresis-mediated size selection[63,64] was
818 replaced by two rounds of high molecular DNA depletion by solid phase reverse immobilisation

819 (SPRI). Wide-bore pipette tips were used for the resuspensions and the transfers of samples
820 from the second homogenisation step until after SPRI to minimise fragmentation of high
821 molecular DNA. Approximately 250 de-jellied mid-blastula embryos were collected in 2-ml
822 round-bottom microcentrifuge tubes and homogenised in 2 ml ice-cold LB-DNase buffer (15
823 mM Tris-HCl pH 8.0, 15 mM NaCl, 60 mM KCl, 1 mM EDTA, 0.5 mM EGTA and 0.5 mM
824 spermidine) supplemented with 0.05% Igepal CA-630. This lysate was left on ice for 3 min
825 before centrifuging at 1,000 g (4ºC) for 2 min. The supernatant was discarded without
826 disrupting the pellet. The pellet was gently resuspended in 2 ml ice-cold LB-DNase buffer
827 (without Igepal CA-630) before centrifuging again at 1,000 g (4ºC) for 2 min. After discarding
828 the supernatant, the pellet was resuspended at room temperature in 600 µl LB-DNase buffer
829 (without Igepal CA-630) with 6 mM $CaCl_2$. The sample was distributed equally to two 1.5-ml
830 microcentrifuge tubes, one for probing chromatin accessibility with DNase and the other one
831 for creating a reference profile of DNase-digested naked genomic DNA. Approximately 0.1 U
832 recombinant DNase I (Roche) was added to one aliquot. Both samples were incubated at
833 37ºC for 8 min before adding 300 µl STOP buffer (0.1% SDS, 100 mM NaCl, 100 mM EDTA
834 and 50 mM Tris-HCl pH 8.0) including 80 µg RNase A (Thermo Fisher Scientific), 333 nM
835 spermine and 1 µM spermidine. The tubes were inverted gently to mix samples before
836 incubating them at 55ºC for 15 min. Next, 200 µg proteinase K (Ambion) were added and the
837 tubes were inverted gently to mix the samples again. After 2 h at 55ºC, the digests were
838 transferred to pre-spun 1.5-ml Phase Lock Gel Heavy (PLGH) microcentrifuge tubes. 600 µl
839 phenol:chloroform:isoamylalcohol (25:24:1, pH 7.9) were added to the digests. The tubes
840 were shaken gently and then centrifuged at 1,500 g (room temperature) for 4 min. The top
841 phase was transferred to a fresh 2-ml microcentrifuge tube and mixed with 60 µl 3 M sodium
842 acetate (pH 5.2) and 1.2 ml of absolute ethanol to precipitate the genomic DNA. The
843 precipitation was stored at -20ºC overnight at least and centrifuged at 16,000 g (4ºC) for 30
844 min. The supernatant was discarded and the DNA pellet washed by adding 500 µl ice-cold
845 80% ethanol and centrifuging at 16,000 g (4ºC) for 3 min. The ethanol was discarded and the
846 DNA pellet was air-dried at room temperature for 10 min. After that the DNA pellet was left on
847 ice to dissolve in 27 µl elution buffer (10 mM Tris-HCl pH 8.5) for 20 min. To remove any
848 residual RNA, 10 µg RNase A were added to the DNA. 1 µl was used to determine the DNA
849 fragment size distribution by gel electrophoresis. On a 0.6% TAE agarose gel, a smear of very
850 high molecular DNA was visible as expected from previous DNase experiments[64,65]. Genomic
851 DNA to the amount of 20 untreated mid-blastula embryos was digested with 0.3 U DNase I at
852 37ºC for 5 min following the same steps as described above, which generated a low molecular
853 smear of DNA fragments. Next, 70 µl SPRI beads (Agencourt AMPure XP, Beckman Coulter)
854 per DNase sample were equilibrated to room temperature for 10 min. 22.5 µl SPRI beads
855 (0.9x of the volume) were added to the DNA sample without pipetting up and down achieving

856 a final polyethylene glycol (PEG) concentration of ~9.5%. After 3 min, by which time high-
857 molecular DNA causes beads to coalesce, the tubes were clipped into a magnetic stand for
858 microcentrifuge tubes. After 3 min or until the beads were separated from the supernatant, the
859 latter was transferred to a 96-well microplate (350-µl round wells with V-shaped bases). 47.5
860 µl elution buffer and 43 µl SPRI beads were added sequentially and mixed gently by slowly
861 pipetting up and down. After 3 min, the plate was transferred to a magnetic stand for 96-well
862 plates. Once the beads have completely separated from the suspension, the supernatant was
863 transferred to a pre-spun PLGH tube. 162 µl elution buffer and 300 µl
864 phenol:chloroform:isoamylalcohol (25:24:1, pH 7.9) were added. The tubes were shaken
865 gently and then centrifuged at 1,500 g (room temperature) for 4 min. The top phase was
866 transferred to a 1.5-ml low-retention microcentrifuge tube and mixed with 30 µl 3 M sodium
867 acetate (pH 5.2) and 900 µl absolute ethanol to precipitate the DNA as outlined above. The
868 DNA pellet was dissolved in 12 µl elution buffer. The concentrations of the DNA samples were
869 determined on a fluorometer using high sensitivity reagents for dsDNA (10 pg/µl to 100 ng/µl).
870 Libraries were generated as previously outlined[43] except that all cleaning steps were executed
871 with 0.2x more SPRI beads.
872
873 **Post-sequencing analysis of DNase-Seq.** Single and paired-end reads of maximal 50 bases
874 were processed using trim_galore v0.4.2 from the Babraham Institute (UK) to trim off low-
875 quality bases (default Phred score of 20, i.e. error probability was 0.01) and adapter
876 contamination from the 3' end. Processed reads were aligned to the *X. tropicalis* genome v7.1
877 and v9.1 using Bowtie2 v2.2.9[47] with default settings apart from -X (fragment length), which
878 was reduced to 250 bp for paired-end reads. Alignments were sorted by genomic coordinates
879 and only those with a quality score of ≥10 were retained using samtools v1.3.1[56]. Duplicates
880 were removed using Picard (MarkDuplicates) from the Broad Institute (USA). Paired-end
881 alignments were dissociated using hex flags (-f 0x40 or 0x80) of samtools view. Single
882 alignments were converted to HOMER's tag density format[48] (makeTagDirectory -single -
883 unique -fragLength 100 -totalReads all). DNase hypersensitive sites (DHSs) were identified
884 using the following or otherwise default parameters of HOMER's peak calling: findpeaks -style
885 factor -minDist 100 -fragLength 100 -inputFragLength 100 -fdr 0.001 -gsize 1.435e9 -F 3 -L 1
886 -C 0.97. This means that alignments of DNase-digested chromatin fragments (or naked
887 genomic DNA fragments considered here as 'input') were extended 3' to 100 bp from the
888 DNase cleavage site to detect significant (0.1% FDR) DNase-mediated read enrichments
889 (hereafter called peaks) being separated by ≥100 bp. The effective size of the *X. tropicalis*
890 genome assembly v7.1 was set to 1.435 billion bp, an estimate obtained from the mappability
891 of ChIP input reads. These peaks showed equal or higher tag density than the surrounding 10
892 kb, at least three-fold more tags than the input and ≥0.97 unique tag positions relative to the

893 expected number of tags. Peaks falling into blacklisted regions (see Post-sequencing analysis
894 of ChIP-Seq) were removed. DNase-probed chromatin accessibility profiles were position-
895 adjusted and normalised to the effective total of 1 million aligned reads including
896 multimappers.
897
898 **Next-generation capture-C.** About 500 mid-blastula embryos per condition were collected
899 in 9-ml capped glass vials and briefly washed once with 1% MMR. The embryos were then
900 fixed with 1% formaldehyde (Sigma, F8775) in 9 ml 1% MMR at room temperature for 40 min.
901 The fixation reaction was terminated by briefly rinsing the embryos three times with ice-cold
902 1% MMR. The embryos were aliquoted into round-bottom 2-ml microcentrifuge tubes in
903 batches of ~250 embryos (filling the volume of ~250 µl water). The supernatant was removed
904 before equilibrating embryos in 250 µl ice-cold HEG buffer (50 mM HEPES pH 7.5, 1 mM
905 EDTA and 20% glycerol). Once the embryos settled to the bottom of the tube as much liquid
906 as possible was discarded. The aliquots were snap-frozen in liquid nitrogen and stored at -
907 80ºC.
908
909 For each 3C experiment, 10 ml chromatin extraction buffer CEB-3C (10 mM Tris-HCl pH 8.0,
910 150 mM NaCl, 1 mM EDTA, 1% Igepal CA-630, 0.25% sodium deoxycholate and 0.2% SDS)
911 were supplemented with a mini protease inhibitor tablet (Roche) and 0.05 mM DTT (CEB-
912 3C*). Both aliquots of ~250 embryos of each condition were homogenised in 2 ml ice-cold
913 CEB-3C*. The tubes were kept on ice for 5 min before centrifuging at 1,000 g (4ºC) for 2 min.
914 The supernatants were discarded and the pellet resuspended in 0.5 ml ice-cold CEB-3C*. The
915 resuspensions of each condition were pooled and then equally divided to two 50-ml conical
916 tubes and each filled up with CEB-3C (without protease inhibitors and DTT) to 50 ml. The
917 embryonic extracts were incubated at 37ºC for 1 h in a hybridisation oven with the tubes
918 rotating inside a hybridisation bottle. The tubes were then centrifuged at 1,000 g (room
919 temperature) for 5 min. The supernatants were discarded and the pellets resuspended in 50
920 ml double distilled water. The tubes were centrifuged again at 1,000 g (room temperature) for
921 5 min. The supernatants were discarded and each pellet was resuspended in 300 µl double
922 distilled water. The resuspensions containing cross-linked nuclei of each condition were
923 pooled in a 1.5-ml microcentrifuge tube. The samples were then digested with 400 U *Dpn*II
924 (NEB) in a total volume of 800 µl containing 1x *Dpn*II reaction buffer. The digest was incubated
925 overnight in a thermomixer set to 37ºC and 1,400 rpm. Of note, any residual yolk platelets in
926 the solution did not interfere with the restriction digest. The next day, the digest was
927 supplemented with 200 U *Dpn*II and incubated at 37ºC and 1,400 rpm for another 6-8 h. After
928 that, *Dpn*II was heat-inactivated at 65ºC for 20 min.
929

930 To verify the degree of chromatin digestion, an aliquot of 50 µl from each condition was
931 transferred into separate 1.5-ml microcentrifuge tubes and processed as follows: First, any
932 remaining RNA was degraded by incubating the aliquot with 20 µg RNase A (Thermo Fisher
933 Scientific) at 37ºC for 30 min. Second, proteins were digested by incubating the aliquot mixed
934 with 50 µl SDS elution buffer (50 mM Tris pH 8.0, 1 mM EDTA and 1% SDS) and 20 µg
935 proteinase K (Ambion) at 65ºC in a hybridisation oven overnight. Third, the DNA was purified
936 and ethanol precipitated: 300 µl TE pH 8.0 were added to the digest before transferring it to
937 pre-spun 1.5-ml Phase Lock Gel Heavy (PLGH) microcentrifuge tubes. The diluted digests
938 were mixed vigorously with 400 µl (1 volume) phenol:chloroform:isoamylalcohol (25:24:1, pH
939 7.9) before centrifuging at 16,000 g (room temperature) for 1 min. The top (aqueous) phases
940 were transferred into 1.5-ml microcentrifuge tubes and mixed with 40 µl sodium acetate (pH
941 5.2) and 800 µl absolute ethanol and 15 µg GlycoBlue (Ambion). The DNA precipitations were
942 kept at -80ºC for 30 min before centrifuging at 16,000 g (4ºC) for 30 min. The supernatants
943 were discarded without disturbing the DNA pellets which were subsequently washed once
944 with 400 µl ice-cold 80% ethanol. The tubes were centrifuged at 16,000 g (4ºC) for 30 min
945 before discarding the supernatants. The DNA pellets were air-dried for 10 min and
946 resuspended in 10 µl elution buffer (10 mM Tris-HCl pH 8.5). The DNA concentrations were
947 determined with a fluorometer and the DNA fragment distributions were visualised on a 0.6%
948 TAE agarose gel. The digestion of cross-linked chromatin with *Dpn*II caused extensive
949 fragmentation of the genomic DNA such that no high molecular DNA fragment bands (>10 kb)
950 were visible.
951
952 The chromatin digests were processed for proximity ligation by adding 240 U T4 DNA ligase
953 (Thermo Fisher Scientific) in a total volume of 1.2 ml containing 1x T4 DNA ligase buffer
954 including 5 mM ATP. The ligation reactions were incubated at 16ºC and 1,400 rpm for ≥16 h
955 before centrifuging at 16,000 g (4ºC) for 1 min. The supernatants were discarded. 200 µl
956 elution buffer and 40 µg RNase A were added to the pellets to degrade any residual RNA at
957 37ºC and 1,400 rpm for 30 min. Proteins were digested by adding 200 µl SDS elution buffer
958 and 160 µg proteinase K and incubating at 65ºC in a hybridisation oven overnight. The digest
959 was transferred to pre-spun PLGH tubes and processed as described in the previous
960 paragraph to purify and precipitate ligated genomic DNA. The air-dried DNA pellet was
961 dissolved in 20 µl elution buffer and quality controlled as above. Proximity ligation of *Dpn*II-
962 digested chromatin fragments massively increased the size of DNA fragments to the range of
963 high molecular weight (>10 kb) representing the chromatin conformation capture (3C) library.
964 The concentration was measured on a fluorometer with broad range concentration (5-500
965 ng/µl) reagents.
966

967 About 10-15 µg of the 3C libraries were diluted with elution buffer to the total volume of 130 µl
968 and transferred to a designated glass vial (microTUBE) for isothermal focused sonication
969 (Covaris). The following settings of the focused ultrasonicator Covaris S220 were used to
970 shear the libraries to an average DNA fragment length of ~200 bp: duty cycle, 10%; intensity,
971 5; cycles per burst, 200; duration, 60 sec in frequency sweeping mode. Sonication was run
972 with 6 cycles in batch mode. The degree of DNA fragmentation was verified by gel
973 electrophoresis and the DNA concentration was measured on a fluorometer with broad range
974 concentration reagents.
975
976 Approximately 2-4 µg of the DNA fragments were converted into next-generation paired-end
977 libraries using the enzymes of the KAPA Hyper Prep Kit (Kapa Biosystems) and TruSeq
978 adaptors (Illumina). The end-repair and A-tailing reactions were set up according to the
979 manufacturer's instructions and incubated at 20ºC for 1 h followed by 30 min at 65ºC. End-
980 repaired and A-tailed DNA fragments were ligated to 150 pmol TruSeq adaptors of index 6
981 and 12, respectively, at 20ºC for 1 h. The DNA was purified using 0.8x volume of SPRI beads
982 and amplified in a total volume of 100 µl using 50 pmol of each TruSeq PCR primer, KAPA
983 HiFi HotStart ReadyMix and the following PCR conditions: 98ºC for 45 sec followed by 3 cycles
984 (98ºC for 15 sec, 98ºC for 30 sec and 72ºC for 30 sec) and a final elongation step of 1 min at
985 72ºC. The lid temperature was set to 105ºC. The PCR reactions were cleaned up with 100 µl
986 SPRI beads and eluted in 14 µl elution buffer. The integrity of the library DNA was verified on
987 a Chip-based capillary electrophoresis system, an 8% TBE polyacrylamide gel or a 2% TAE
988 agarose gel. DNA concentration was measured on a fluorometer with broad range
989 concentration reagents. The DNA yield was 2-4 µg.
990
991 To capture the genomic regions of interest (viewpoints), 5'-biotinylated oligonucleotides of 120
992 bases (hereafter called probe; Supplementary Table 4) were designed for each viewpoint as
993 follows: The viewpoints were *Dpn*II fragments of 300-3,300 bp which were overlapping gene
994 promoters or located <1 kb from them. The probe matched the terminal sequence of the same
995 DNA fragment strand including the *Dpn*II restriction site GATC. The probe sequence was
996 examined using BLAT to check whether it is unique and did not partially match any other
997 genomic regions. Furthermore, CENSOR v4.2.29[66] and RepeatMasker v4.0.6[67] programs
998 were run to discard any probes that contained repeats. Because of these design restrictions,
999 six of the selected 30 gene promoters were captured with only one probe. All the probes were
1000 purchased as desalted oligomers (4 nmol) from IDT (ultramer technology) and re-constituted
1001 in molecular-grade water to 10 µM. The probes were mixed in equimolar quantities and diluted
1002 to 2.9 nM such that 4.5 µl contained 13 fmol of oligomers.
1003

1004 The capture was performed with the SeqCap Hybridization and Wash Kit (Roche) and SeqCap
1005 HE-Oligo Kit A (Roche) as follows. Exactly 1 µg of each TruSeq library with index 6 and 12
1006 was mixed with 1 nmol TruSeq universal blocking oligonucleotide and 0.5 nmol blocking
1007 oligonucleotides specific to TruSeq index 6 and 12 in a 1.5-ml low-retention microcentrifuge
1008 tube. Of note, Cot-1 DNA commonly used to mask repetitive DNA in hybridisation reactions
1009 proved to be not required for capturing the genomic DNA of interest. The mixture of library
1010 and blocking oligonucleotides was dried in a vacuum centrifuge before adding 7.5 µl 2x
1011 hybridisation buffer (vial 5) and 3 µl hybridisation component A (vial 6). This mixture was
1012 vortexed for 10 sec, centrifuged at 16,000 g for 10 sec and incubated at 95ºC for 10 min. In
1013 the meantime, 4.5 µl (13 fmol) of the equimolar probe mixture were transferred to a PCR tube
1014 and incubated at 47ºC in a PCR machine with the volume and the lid temperature set to 15 µl
1015 and 57ºC, respectively. Upon denaturation, the libraries and blocking oligonucleotides were
1016 added to the probes without removing either tubes from the heating block or the PCR machine.
1017 The hybridisation mixture was incubated at 47ºC for 64-72 h. The wash buffers were prepared
1018 according to the manufacturer's instructions to make 1x working solutions for one single
1019 capture experiment. The stringent wash buffer and wash buffer I were pre-warmed to 47ºC
1020 and 50 µl M270 streptavidin beads (Invitrogen) were transferred into a 1.5-ml low-retention
1021 microcentrifuge tube to let them equilibrate to room temperature for 10 min. The magnetic
1022 beads were washed twice with 200 µl bead wash buffer. Immediately after the final wash, the
1023 hybridisation mixture was directly added to the beads and vortexed for 10 sec. The sample
1024 was incubated for 45 min in a thermomixer set to 47ºC and 1,100 rpm. The beads were
1025 washed by adding 100 µl pre-warmed wash buffer I and vortexing for 10 sec. The tube was
1026 placed into a magnetic rack. Once the liquid was clear, the supernatant was discarded, and
1027 200 µl pre-warmed stringent wash buffer were added to the beads. In an effort to avoid any
1028 temperature drop, it is important to work quickly according to the manufacturer's instructions.
1029 The beads in stringent wash buffer were incubated in a thermomixer set to 47ºC and 1,100
1030 rpm for 5 min. The wash with pre-warmed stringent wash buffer was repeated once. Next, 200
1031 µl pre-warmed wash buffer I was added to the beads and vortexed at 1,400 rpm (room
1032 temperature) for 2 min. After removing the respective wash buffer, beads were vortexed at
1033 1,400 rpm in 200 µl wash buffer II for 1 min and 200 µl wash buffer III for 30 sec. After the final
1034 the final wash, as much liquid as possible was removed and the beads were resuspended in
1035 40 µl molecular-grade water.
1036
1037 The captured DNA was directly amplified from the beads using KAPA HiFi HotStart master
1038 mix and 25 pmol of each TruSeq PCR primer in two separate 50-µl PCR reactions. The PCR
1039 conditions were the same as outlined above except that 14 cycles were used for amplification.
1040 The PCR reactions were pooled and the DNA was purified using 100 µl SPRI beads. The DNA

1041 was eluted with 11 µl elution buffer. The eluted DNA was subjected to another round of probe-
1042 mediated capture with the hybridisation timespan reduced to 24 h. Furthermore, after washing
1043 the beads, any NG capture-C fragments were amplified using only 10 cycles of PCR.
1044
1045 **Post-sequencing analysis of next-generation capture-C reads.** The analysis was carried
1046 out in accordance with the original description of the method[68] with some modifications.
1047 Paired-end reads were processed using trim_galore v0.4.2 from the Babraham Institute (UK)
1048 to trim off low-quality bases (default Phred score of 20, i.e. error probability was 0.01) and
1049 adapter contamination from the 3' end. Only complete mate pairs were processed further to
1050 reconstruct single reads from overlapping paired-end sequences using FLASH v1.2.11[69] with
1051 interleaved output settings for non-extended reads (flash --interleaved-output --max-overlap
1052 150). FLASH reads were split in *silico* at *Dpn*II restriction sites using a designated perl script
1053 (https://github.com/telenius/captureC/releases) before aligning them to the *X. tropicalis*
1054 genome assembly 7.1 using Bowtie2 v2.2.9[47]. The alignment was run with default settings and
1055 one thread only to keep the order of the reads. The view function of samtools v1.3.1[56] was
1056 used to retain alignments with a quality score of ≥10. The alignments were analysed further
1057 using a suite of perl scripts (https://github.com/Hughes-Genome-Group/captureC/releases)
1058 modified to process both chromosome and scaffold coordinates. Viewpoint coordinates
1059 included a 1-kb proximity exclusion range. Restriction fragments were classified as capture,
1060 proximity-exclusion or reporter. PCR duplicates were removed. The interaction map was
1061 based on the number of unique paired-end reads per restriction fragment. Windows of 2 kb
1062 incrementing by 200 bp were used to consolidate interactions, which were normalised to
1063 10,000 interactions per viewpoint.
1064
1065 **Analysis of genomic profiles, DNA motifs and super-enhancers.** Unless otherwise stated,
1066 Spearman correlations (R_s), relative tag densities, PCAs and enrichment of DNA motifs were
1067 determined using a limited number of detected peaks per chromatin feature and
1068 developmental stage: top 2,000 peaks (Fig. 2a; Extended Data Fig. 4a-e), top 5,000 peaks
1069 (Extended Data Fig. 1b) and top 10,000 peaks (Fig. 2c,e,h; Extended Data Fig. 5a,b,f,g). R_s
1070 and PCA were calculated using DiffBind v2.4.8[70] in R (Fig. 1a and 2c,h; Extended Data Fig.
1071 4a-c and 5a). Z-scores of motif enrichments at RNAPII peaks were calculated using ISMARA
1072 v1.2.2[71] (https://ismara.unibas.ch/fcgi/mara) in expert mode (Fig. 1b). Heat maps of relative
1073 tag densities were hierarchically clustered (Ward's method) except for those in Extended Data
1074 Fig. 5b, Fig. 4b and 9d, where genomic coordinates were sorted upon significant changes in
1075 DNA occupancy or chromatin accessibility (see figure legends). In Extended Data Fig. 4d,
1076 cluster leafs were additionally ordered with the Gruvaeus-Wainer's algorithm. The matrices of
1077 normalised tag densities, p-values and motif instances were generated with HOMER as

1078 follows: annotatePeaks.pl -size 2000 (or 200 for Extended Data Fig. 4d) -hist 25 (bin size) -
1079 ghist. In heat maps all genomic profiles were scaled to the means of the highest bin values of
1080 the top 100 (or 1,000; Fig. 4b and Extended Data Fig. 9d) peaks. In addition, to visualise low
1081 RNAPII binding and RNA expression, their top means were divided by 5 or 500, respectively.
1082 Super-enhancers were assembled by stitching together peaks that are ≤25 kb apart:
1083 findPeaks -style super -fragLength 175 -inputFragLength 175 -fdr 0.001 -gsize 1.435e9 -F 3 -
1084 L 1 -localSize 10000 -C 0.97 -superSlope 1 -minDist 25000. Subsequently, super-enhancers
1085 ≤5 kb from TSS apart were associated with zygotic genes defined by RNAPII coverage (see
1086 Detecting zygotic genes by means of RNAPII chromatin profiles).
1087
1088 **Analysis of enriched gene ontology (GO) terms.** Over-represented GO terms were found
1089 by applying hypergeometric tests of the Bioconductor/R package GOstats v2.40.0[72] on gene
1090 lists. The gene universe was associated with GO terms by means of BLAST2GO[73] as
1091 previously outlined[46,74].
1092
1093 **Differential analysis of chromatin features.** The significance of differential chromatin
1094 accessibility between control and mPouV/Sox3 LOF embryos was assessed at DHSs using
1095 the Bioconductor/R package DiffBind v2.4.8[70] and the DESeq2[58] algorithm. The sizes of all
1096 DHSs were fixed to 500 bp. Only the very 5' end of each aligned DNase-Seq read, which
1097 represents the DNase cleavage site, was considered for this analysis. Read counts from
1098 mapping DNase-digested naked genomic DNA were subtracted from chromatin accessibility
1099 readings at each DHS. Any changes to DNase-mediated cleavages at DHSs with FDR ≤10%
1100 were determined as statistically significant (Extended Data Fig. 9b). For visualisation, the
1101 orientations of DHSs were aligned to the direction of transcription of the closest zygotic gene.
1102 Matrices of normalised and raw tag counts were generated at DHSs for various chromatin
1103 features and RNA (Fig. 4b and Extended Data Fig. 9d). Their rows were sorted according to
1104 the statistical significance of differential chromatin accessibility. Raw tag counts were
1105 processed to estimate p-values of chromatin changes across DHSs using DESeq2 v1.14.1[58].
1106 Read count dispersions were locally fitted. The shrinkage of $\log_2$ fold changes, threshold on
1107 Cook's distances and independent filtering were all switched off. For the comparison of NG
1108 capture-C interactions with promoters at affected and unaffected DHSs, the size of the DHSs
1109 was widened to 1.2 kb and filtered for ≥5 reporter fragments (Extended Data Fig. 10c).
1110 Normalised line and bar graph tracks of control and mPouV/Sox3 LOF embryos were
1111 superimposed in the IGV genome browser v2.3.92[75] (Fig. 4c and Extended Data Fig. 10d).
1112

1113 **Software for computational analysis and graphical illustration.** Post-alignment analysis
1114 and graphical illustration were performed using samtools v1.3.1[56], bedtools v2.25.0[76], HOMER
1115 v4.8.3[48] and tools in R v3.4.1 / Bioconductor v3.5, Perl v5.18.2 or Python v2.7.12.
1116
1117 **Data availability.** Raw and processed sequencing data for this study is available at the Gene
1118 Expression Omnibus (GSE113186).
1119
1120 **External next-generation datasets.** The following external datasets were used for this study:
1121 5mC at stage 9, and ChIP for H3K4me3 and H3K27me3 at stage 9 (GSE67974), β-catenin
1122 ChIP at stage 10^+ (GSE72657), Smad2 ChIP at stage 10^+ (GSE30146 and GSE53654), Foxh1
1123 ChIP at stage 8 and 10^+ (GSE85273 and GSE53654), total RNA recorded at 30-min intervals
1124 from unfertilised eggs to 23.5 hpf (GSE65785), ribosome footprinting from stage 3, 9 and 12^+
1125 (GSE52809), and regional expression along the animal-vegetal and dorso-ventral axis at
1126 stage 10^+ (GSE81458).
1127
1128 **Statistical analysis.** No statistical method was used for determining sample size. We followed
1129 the literature to select the appropriate sample size. The experiments were not randomized.
1130 Due to the nature of experiments, the authors were not blinded to allocation during
1131 experiments and results analysis. Only viable embryos and embryonic tissues were included
1132 for analysis. Due to space constraints, data from the trials of establishing and optimising
1133 protocols (ChIP-Seq, DNase-Seq and next-generation capture-C) were excluded.
1134 Frequencies of shown morphological phenotypes and WMISH patterns are included in every
1135 image. Significances of normally or non-normally distributed data points were calculated using
1136 a paired Student's t-test (one-tailed or two-tailed) or paired Wilcoxon test (two-tailed),
1137 respectively. The effect size r was estimated from the standard normal deviate of the Wilcoxon
1138 p-value (p) as previously described[80], r_{effect}=Z/sqrt(N), where Z=qnorm(1-p/2) is the
1139 standardised Z-score and N is the number of observations. Significance of motif enrichments
1140 or GO terms are based on hypergeometric tests. The significance of transcriptome-wide
1141 differential expression was adjusted for multiple comparisons using the Benjamini-Hochburg
1142 procedure.
1143
1144 **Code availability.** Code for the analyses described in this paper is available from the authors
1145 upon request.
1146
1147 **References**
1148 1. Waddington, C. H. Organisers and Genes. Cambridge University Press (1940).
1149 2. Sagner, A. & Briscoe, J. Morphogen interpretation: concentration, time, competence, and

1150 signaling dynamics. *Wiley Interdiscip Rev Dev Biol* **6**, e271 (2017).
1151 3. Langley, A. R., Smith, J. C., Stemple, D. L. & Harvey, S. A. New insights into the maternal to
1152 zygotic transition. *Development* **141,** 3834–3841 (2014).
1153 4. Kimelman, D. Mesoderm induction: from caps to chips. *Nat Rev Genet* **7,** 360–372 (2006).
1154 5. Wühr, M. *et al.* Deep proteomics of the Xenopus laevis egg using an mRNA-derived reference
1155 database. *Curr Biol* **24,** 1467–1475 (2014).
1156 6. Subtelny, A. O., Eichhorn, S. W., Chen, G. R., Sive, H. & Bartel, D. P. Poly(A)-tail profiling
1157 reveals an embryonic switch in translational control. *Nature* **508,** 66–71 (2014).
1158 7. Zhang, C., Basta, T., Fawcett, S. R. & Klymkowsky, M. W. SOX7 is an immediate-early target
1159 of VegT and regulates Nodal-related gene expression in Xenopus. *Dev Biol* **278,** 526–541
1160 (2005).
1161 8. Larabell, C. A. *et al.* Establishment of the dorso-ventral axis in Xenopus embryos is presaged
1162 by early asymmetries in beta-catenin that are modulated by the Wnt signaling pathway. *J Cell*
1163 *Biol* **136,** 1123–1136 (1997).
1164 9. Stennard, F., Zorn, A. M., Ryan, K., Garrett, N. & Gurdon, J. B. Differential expression of VegT
1165 and Antipodean protein isoforms in Xenopus. *Mech Dev* **86,** 87–98 (1999).
1166 10. Peshkin, L. *et al.* On the Relationship of Protein and mRNA Dynamics in Vertebrate Embryonic
1167 Development. *Dev Cell* **35,** 383–394 (2015).
1168 11. Schohl, A. & Fagotto, F. Beta-catenin, MAPK and Smad signaling during early Xenopus
1169 development. *Development* **129,** 37–52 (2002).
1170 12. Faure, S., Lee, M. A., Keller, T., Dijke, ten, P. & Whitman, M. Endogenous patterns of TGFbeta
1171 superfamily signaling during early Xenopus development. *Development* **127,** 2917–2931 (2000).
1172 13. De Domenico, E., Owens, N. D. L., Grant, I. M., Gomes-Faria, R. & Gilchrist, M. J. Molecular
1173 asymmetry in the 8-cell stage Xenopus tropicalis embryo described by single blastomere
1174 transcript sequencing. *Dev Biol* **408,** 252–268 (2015).
1175 14. Charney, R. M. *et al.* Foxh1 Occupies cis-Regulatory Modules Prior to Dynamic Transcription
1176 Factor Interactions Controlling the Mesendoderm Gene Program. *Dev Cell* **40,** 595–607.e4
1177 (2017).
1178 15. Gentsch, G. E. *et al.* In vivo T-box transcription factor profiling reveals joint regulation of
1179 embryonic neuromesodermal bipotency. *Cell Rep* **4,** 1185–1196 (2013).
1180 16. Gentsch, G. E., Monteiro, R. S. & Smith, J. C. Cooperation Between T-Box Factors Regulates
1181 the Continuous Segregation of Germ Layers During Vertebrate Embryogenesis. *Curr Top Dev*
1182 *Biol* **122,** 117–159 (2017).
1183 17. Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key
1184 cell identity genes. *Cell* **153,** 307–319 (2013).
1185 18. Müller, C. W. & Herrmann, B. G. Crystallographic structure of the T domain-DNA complex of the
1186 Brachyury transcription factor. *Nature* **389,** 884–888 (1997).
1187 19. Chen, X., Rubock, M. J. & Whitman, M. A transcriptional partner for MAD proteins in TGF-beta
1188 signalling. *Nature* **383,** 691–696 (1996).
1189 20. Chiu, W. T. *et al.* Genome-wide view of TGFβ/Foxh1 regulation of the early mesendoderm
1190 program. *Development* **141**, 4537–4547 (2014).
1191 21. Heasman, J., Kofron, M. & Wylie, C. βCatenin Signaling Activity Dissected in the Early Xenopus
1192 Embryo: A Novel Antisense Approach. *Dev Biol* **222,** 124–134 (2000).
1193 22. Skirkanich, J., Luxardi, G., Yang, J., Kodjabachian, L. & Klein, P. S. An essential role for
1194 transcription before the MBT in Xenopus laevis. *Dev Biol* **357,** 478–491 (2011).
1195 23. Reversade, B., Kuroda, H., Lee, H., Mays, A. & De Robertis, E. M. Depletion of Bmp2, Bmp4,
1196 Bmp7 and Spemann organizer signals induces massive brain formation in Xenopus embryos.
1197 *Development* **132,** 3381–3392 (2005).
1198 24. Blitz, I. L. *et al.* A catalog of Xenopus tropicalis transcription factors and their regional expression
1199 in the early gastrula stage embryo. *Dev Biol* **426,** 409–417 (2017).
1200 25. Neph, S. *et al.* An expansive human regulatory lexicon encoded in transcription factor footprints.
1201 *Nature* **489,** 83–90 (2012).
1202 26. Soufi, A. *et al.* Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate
1203 reprogramming. *Cell* **161,** 555–568 (2015).
1204 27. Gurdon, J. B. Cell Fate Determination by Transcription Factors. *Curr Top Dev Biol* **116,** 445–
1205 454 (2016).
1206 28. Pasque, V., Jullien, J., Miyamoto, K., Halley-Stott, R. P. & Gurdon, J. B. Epigenetic factors
1207 influencing resistance to nuclear reprogramming. *Trends Genet* **27,** 516–525 (2011).
1208 29. Lee, M. T. *et al.* Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the
1209 maternal-to-zygotic transition. *Nature* **503,** 360–364 (2013).

30. Iwafuchi-Doi, M. & Zaret, K. S. Cell fate control by pioneer transcription factors. *Development* **143,** 1833–1837 (2016).
31. Khokha, M. *et al.* Techniques and probes for the study of Xenopus tropicalis development. *Dev Dyn* **225,** 499–510 (2002).
32. Sive, H., Grainger, R. & Harland, R. Early development of Xenopus laevis: A laboratory manual. Cold Spring Harbor Press (2000).
33. 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).
34. 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).
35. 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).
36. Gökirmak, T. *et al.* Localization and substrate selectivity of sea urchin multidrug (MDR) efflux transporters. *J Biol Chem* **287,** 43876–43883 (2012).
37. 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).
38. 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).
39. 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).
40. Monsoro-Burq, A. H. A Rapid Protocol for Whole-Mount In Situ Hybridization on Xenopus Embryos. *Cold Spring Harb Protoc***,** pdb.prot4809 (2007).
41. Landt, S. G. *et al.* ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res* **22,** 1813–1831 (2012).
42. 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).
43. 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).
44. 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).
45. Gentsch, G. E. & Smith, J. C. Investigating physical chromatin associations across the Xenopus genome by chromatin immunoprecipitation. *Cold Spring Harb Protoc* **2014,** (2014).
46. 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* **96**, e52535 (2015).
47. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9,** 357–359 (2012).
48. 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).
49. 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).
50. 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).
51. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* **27,** 573–580 (1999).
52. 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).
53. 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).
54. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).
55. Zhu, Q., Fisher, S. A., Shallcross, J. & Kim, J. VERSE: a versatile and efficient RNA-Seq read counting tool. *bioRxiv*, 053306 (2016).
56. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
57. Wang, L., Wang, S. & Li, W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28,** 2184–2185 (2012).
58. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for

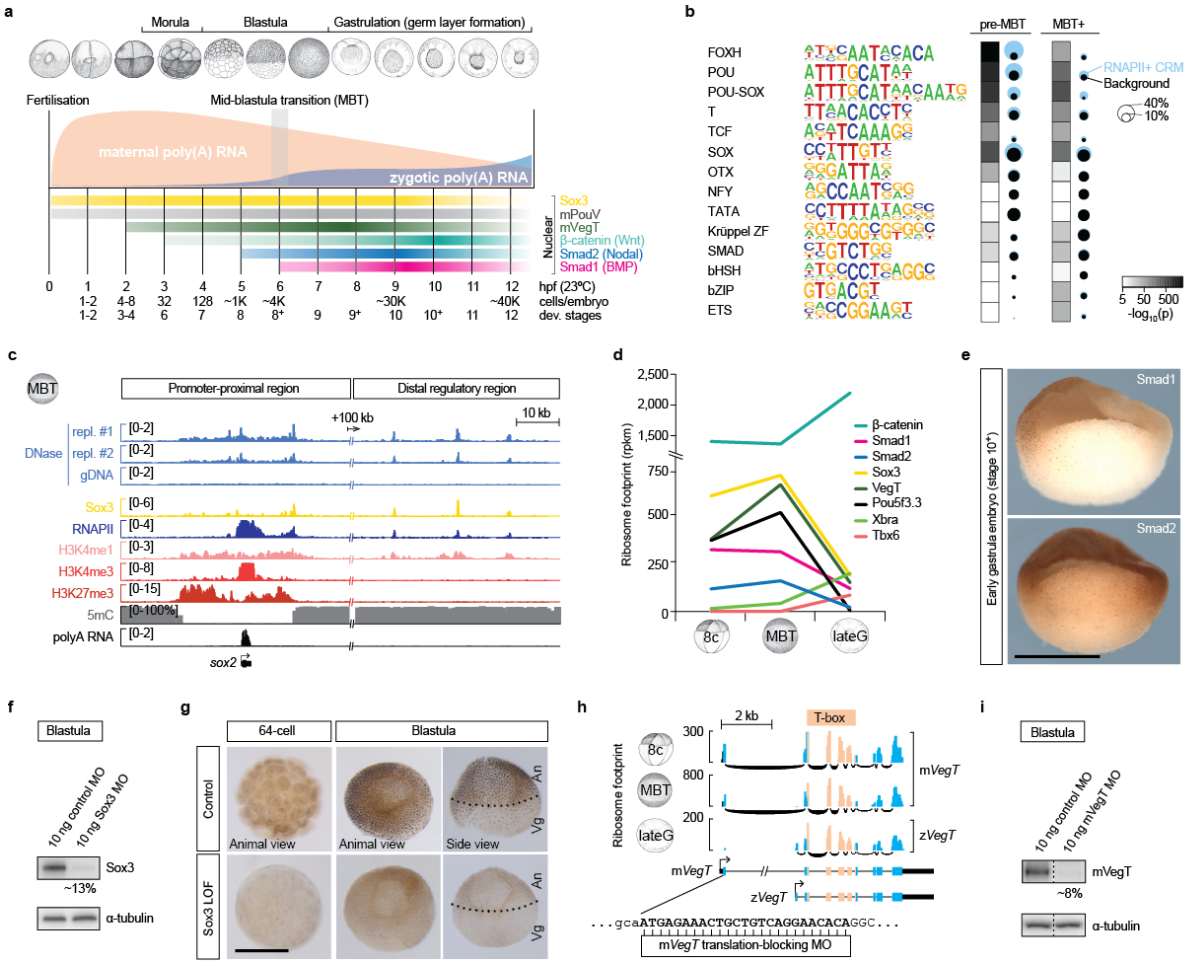
RNA-seq data with DESeq2. *Genome Biol* **15,** 550 (2014).
59. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41,** D590–6 (2013).
60. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10,** R25 (2009).
61. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14,** R36 (2013).
62. UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res* **43,** D204–12 (2015).
63. 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).
64. 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).
65. Song, L. & Crawford, G. E. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harb Protoc*, pdb.prot5384 (2010).
66. 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).
67. Smit, A., Hubley, R. & Green, P. RepeatMasker Open-4.0. 2013-2015.
68. Davies, J. O. J. *et al.* Multiplexed analysis of chromosome conformation at vastly improved sensitivity. *Nat Methods* **13,** 74–80 (2016).
69. Magoč, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27,** 2957–2963 (2011).
70. Stark, R. & Brown, G. DiffBind: Differential binding analysis of ChIP-Seq peak data. 1–33 (2018).
71. Balwierz, P. J. *et al.* ISMARA: automated modeling of genomic signals as a democracy of regulatory motifs. *Genome Res* **24,** 869–884 (2014).
72. Falcon, S. & Gentleman, R. Using GOstats to test gene lists for GO term association. *Bioinformatics* **23,** 257–258 (2007).
73. Conesa, A. *et al.* Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21,** 3674–3676 (2005).
74. Collart, C. *et al.* High-resolution analysis of gene activity during the Xenopus mid-blastula transition. *Development* **141,** 1927–1939 (2014).
75. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29,** 24–26 (2011).
76. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).
77. Rosenthal, R. Meta-Analytic Procedures for Social Research. SAGE (1991).
78. Hontelez, S. *et al.* Embryonic transcription is controlled by maternally defined chromatin state. *Nat Commun* **6,** 10148 (2015).
79. 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).

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 ChIP protocol for β-catenin; 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).

Authors Contributions G.E.G. designed the project, performed experiments and data analysis. G.E.G. and J.C.S wrote the manuscript. T.S. helped with some embryo collections, immunostainings and library preparations. N.D.L.O. assisted with the mapping of some sequencing data.

Author Information Sequencing data are deposited in the Gene Expression Omnibus (GEO) database with accession number GSE113186. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.E.G. and J.C.S. (george.gentsch@crick.ac.uk; jim.smith@crick.ac.uk).

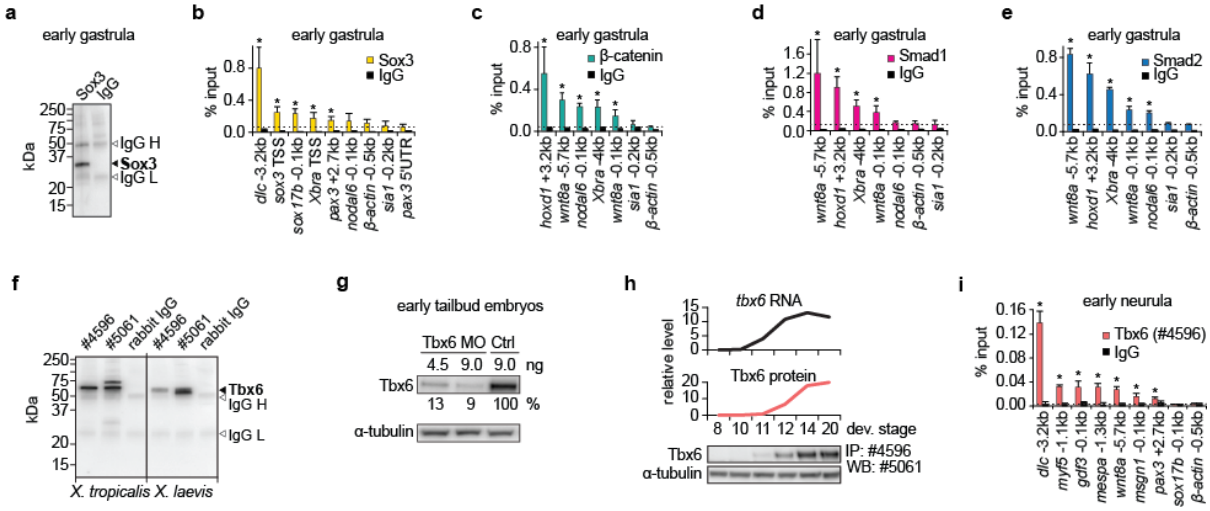
Extended Data



Extended Data Figure 1. Identification of most influential TFs on ZGA

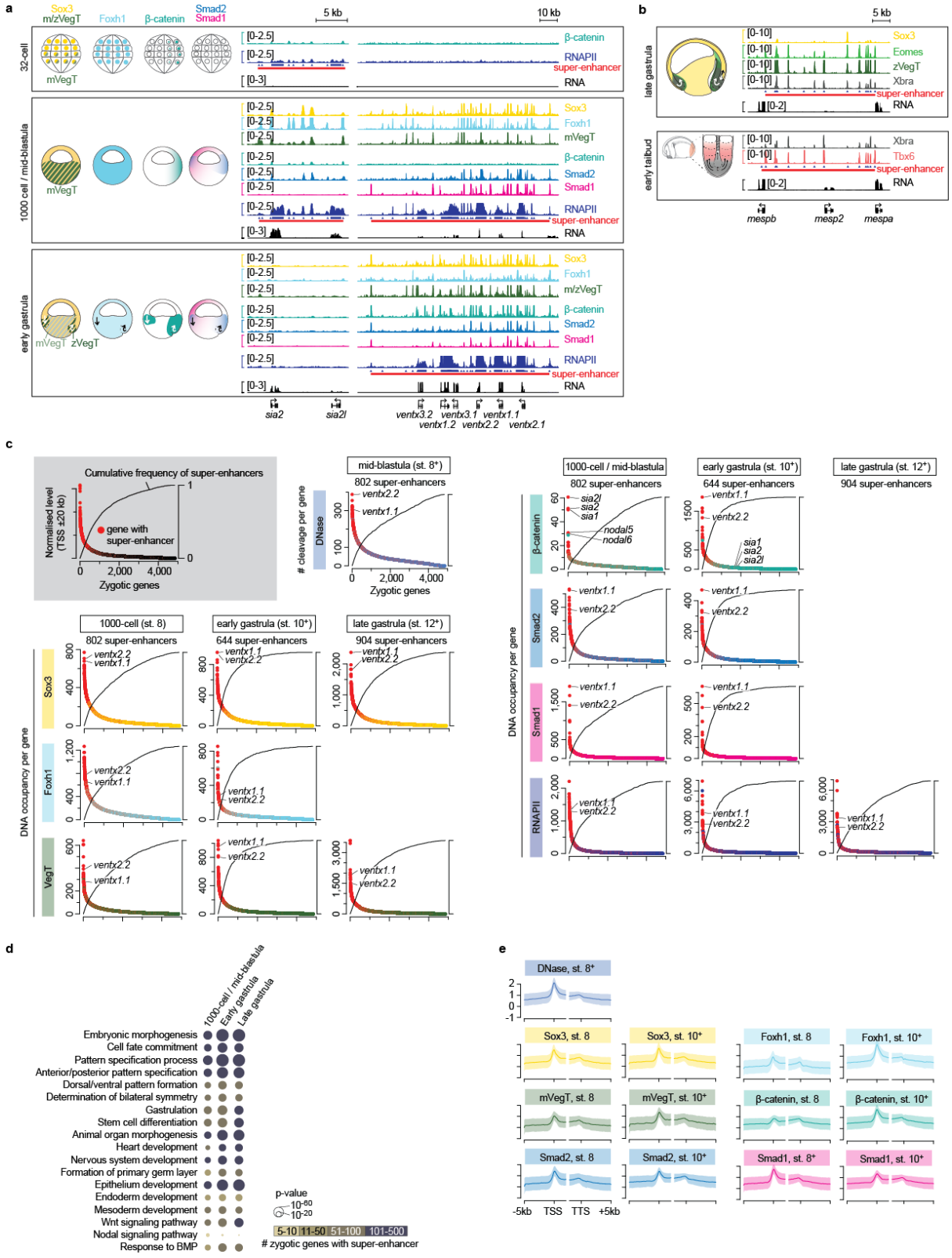
a, Timeline of the maternal-to-zygotic transition during early *Xenopus* embryogenesis. Examples of maternal TFs Sox3, mPouV and mVegT being nuclear before the signal mediators β-catenin, Smad2 and Smad1 are based on our and previously published result[7-12] (also see Fig. 1f). **b**, Enrichment (hypergeometric p-value 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**, 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[78]) at the *sox2* locus (promoter-proximal and downstream distal regulatory region) around MBT to validate our novel DNase-Seq method. Negative control of DNase-Seq, DNase-treated naked genomic DNA (gDNA). **d**, Ribosome footprinting[6] of selected maternal and zygotic TFs across MBT. **e**, WMIHC of Smad1 and Smad2 on bisected early gastrula embryos. **f**, Western blotting of Sox3 in standard control and Sox3 morphants (Sox3 LOF) at blastula stage. α-tubulin, loading control. **g**, WMIHC of Sox3 in control and Sox3 LOF embryos at 64-cell and blastula stage. Nuclear accumulation of Sox3 detected in both the animal (an) and vegetal (veg) 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 in standard control and mVegT morphants (mVegT LOF). α-tubulin, loading control.



Extended Data Figure 2. Validation of ChIP-grade antibodies

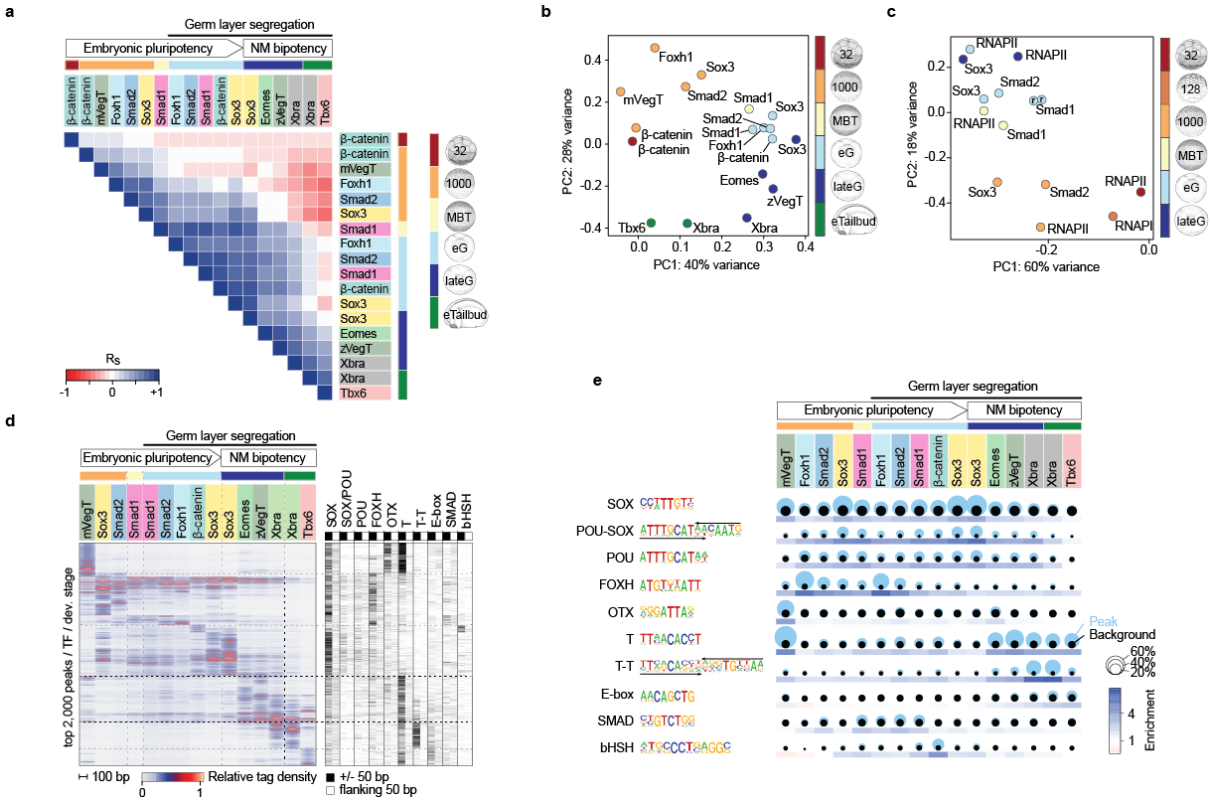
a, Sox3 immunoprecipitation (IP). IgG H and L, detected IgG heavy and light chains of the IP antibody. **b-e,i**, ChIP-qPCR for Sox3, β-catenin, Smad1, Smad2 and Tbx6 at indicated developmental stages. Mean+SD (n=2-3). One-tailed Student's t-test (comparing to IgG control): *, p ≤0.1 and ≥2-fold enrichment relative to lowest DNA recovery (% input) with ChIP antibody (dotted line). **f**, 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. **g**, Western blotting (WB) of Tbx6 in standard control and Tbx6 morphants at early tailbud stage. α-tubulin, loading control. **h**, Relative expression levels of the transcript (RT-qPCR) and its protein (IP/WB) from early blastula to early tailbud stage. α-tubulin, loading control.



Extended Data Figure 3. Super-enhancers of ZGA are associated with embryonic and tissue-specific gene functions

a, Dynamic chromatin engagement of endogenous Sox3, Foxh1[14], 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 localisation of the selected TFs with arrows pointing to any tissue movements. Arrowheads,

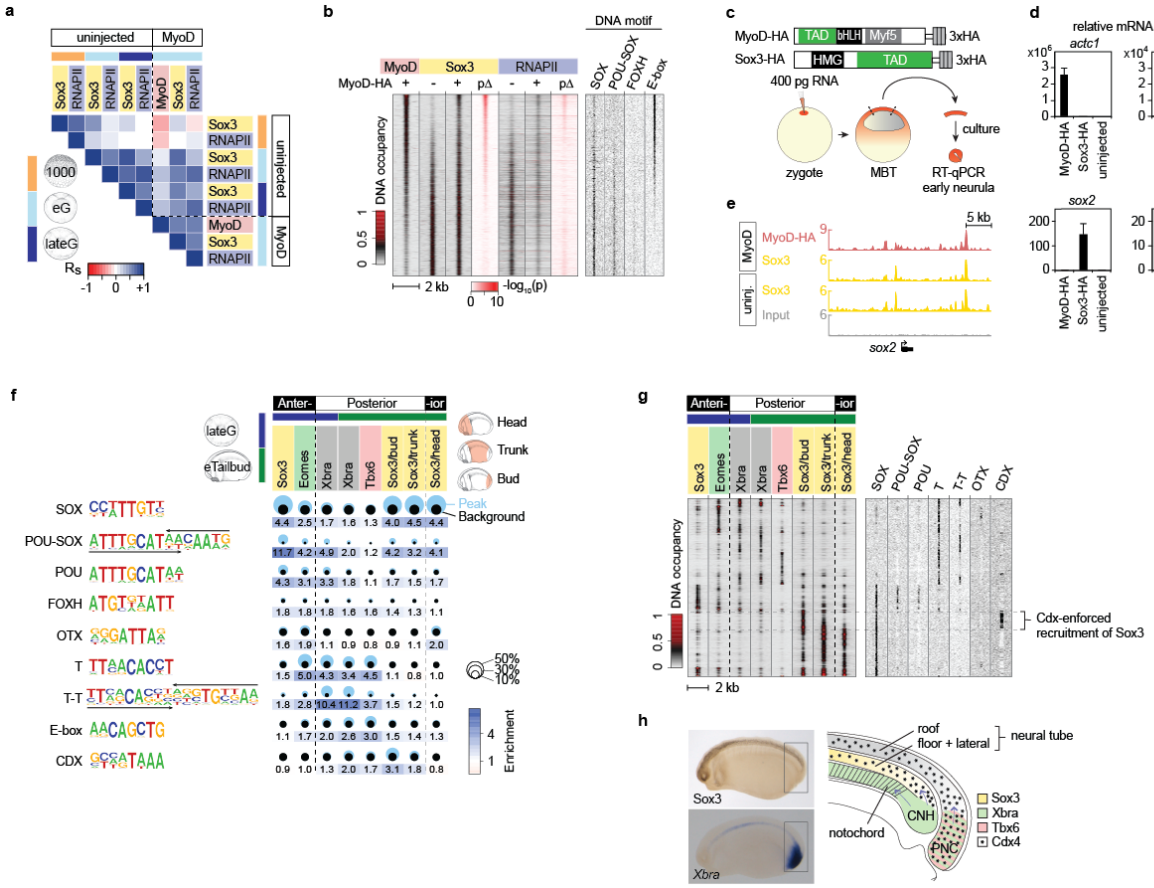
promoter and enhancer activity based on the peak-like recruitment of RNAPII and/or TFs. Rectangles, transcriptional engagement based on the broad distribution of RNAPII and the occurrence of intronic RNA. **b**, Chromatin landscape of endogenous Sox3, Eomes, zVegT, Xbra and Tbx6 to super-enhancer of the *mesp* gene cluster at late gastrula and/or early tailbud stage. Illustrated sagittal sections (dorsal side is right) show the nuclear localisation of selected TFs in late gastrula embryo and at the caudal end of early tailbud embryo. **c**, Super-enhancer-like binding correlates with high levels of chromatin accessibility and DNA occupancy of TFs, signal mediators and RNAPII. Grey box explains following plots. They show normalised 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 the cumulative frequency of detected super-enhancers. Genes with super-enhancer-like 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).



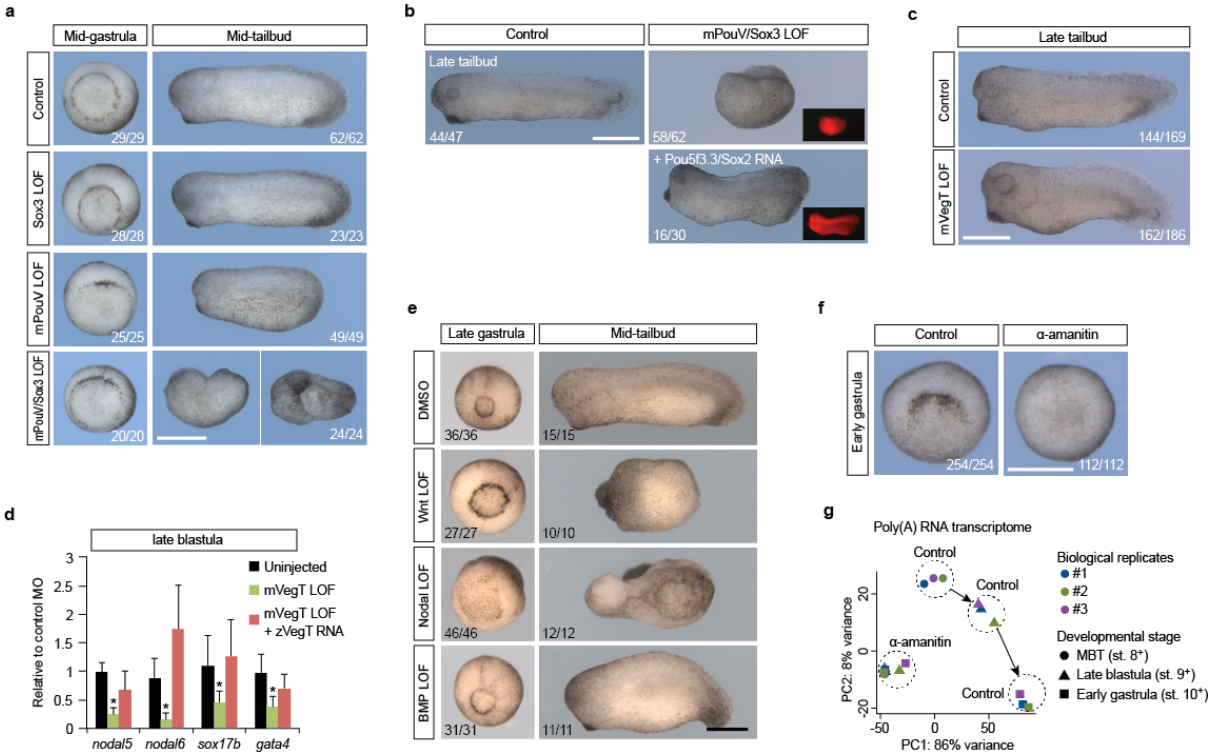
Extended Data Figure 4. Endogenous chromatin dynamics in cells of embryonic pluripotency and neuro-mesodermal bipotency

a-e, Endogenous dynamics of chromatin engagement between TFs (Sox3, mVegT, Foxh1[14], Eomes[15], zVegT[15], Xbra[15] and Tbx6) and signal mediators (β-catenin, Smad1, Smad2) covering developmental process of embryonic pluripotency and neuromesodermal (NM) bipotency. **a**, Spearman correlation 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**, DNA occupancy-assisted clusters of CRMs co-aligned

with enriched DNA motifs detected in peak centres (±50 bp) and flanking sequences (each 50 bp). **e**, Coverage and enrichment of DNA motifs.

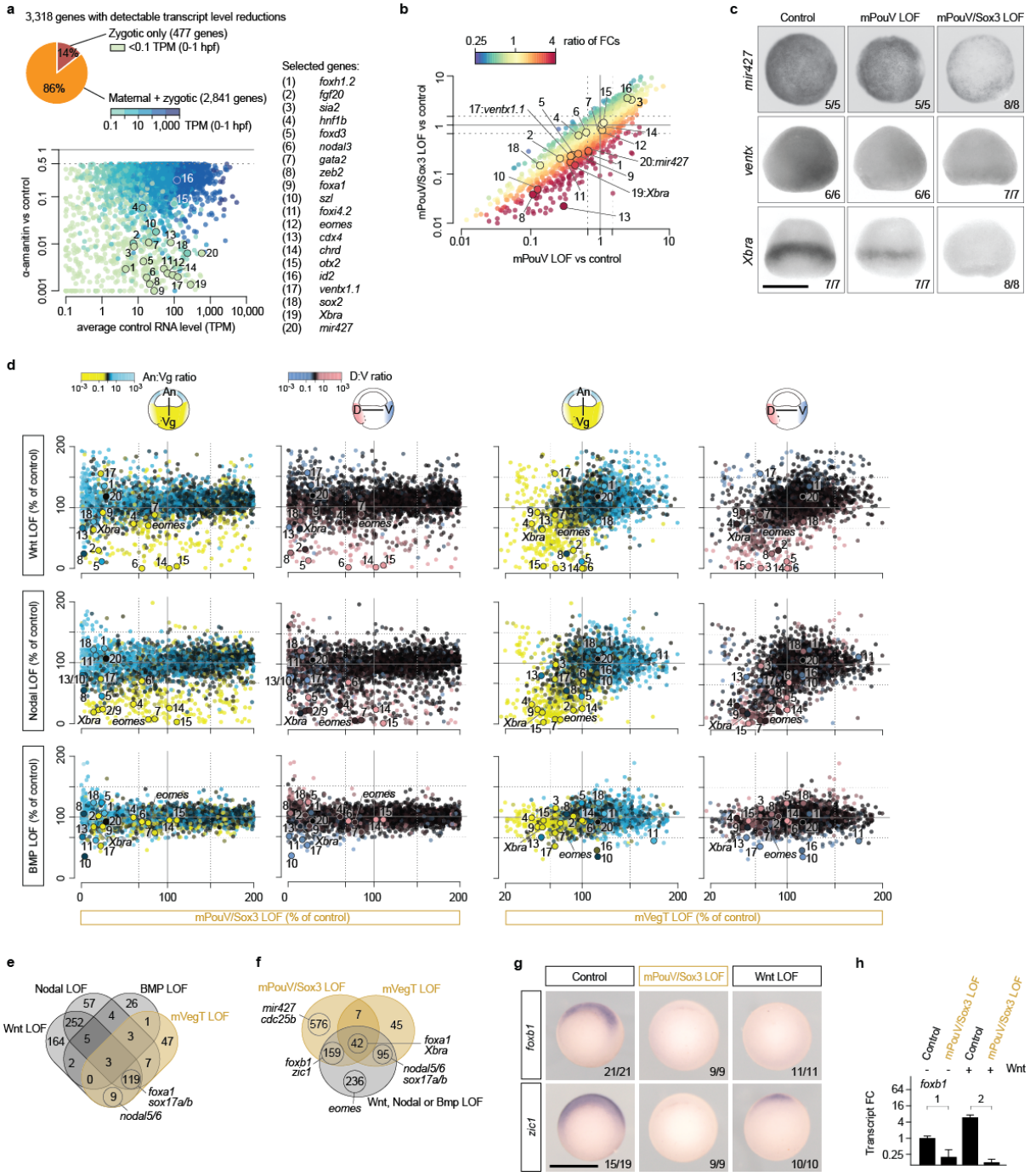


Extended Data Figure 5. Developmental context influences chromatin engagement
a-e, MyoD-enforced change of the development context in early gastrula embryos and its influence on chromatin recruitment of Sox3 and RNAPII. **a**, Spearman correlation R_s of DNA occupancy levels. **b**, Map of DNA occupancies and enriched DNA motifs sorted by significantly differential (pΔ) Sox3 binding. **c**, Animal cap assay to measure MyoD-enforced transcription. **d**, RT-qPCR results from the animal cap assay. Error bars, mean+SD (n=2). **e**, Opportunistic recruitment of MyoD-HA to Sox3 target gene *sox2*. **f-h**, Spatio-temporal co-enrichment of DNA motifs at CRMs occupied by Sox3, Eomes[15], Xbra[15] and/or Tbx6 at indicated developmental stages reflect developmental context. **f**, Coverage and enrichment of DNA motifs. **g**, Hierarchical clustering of DNA occupancies separated by time and space. **h**, Sox3-expressing cells of the posterior neural tube originate from the chordoneural hinge (CNH) and posterior wall of the neurenteric canal (PNC) expressing Xbra and Xbra/Tbx6, respectively. Both neural tube and PNC are also exposed to the expression of Cdx such as Cdx4[79].



Extended Data Figure 6. Maternal TFs and signals are critical to ZGA-dependent tissue movements of gastrulation

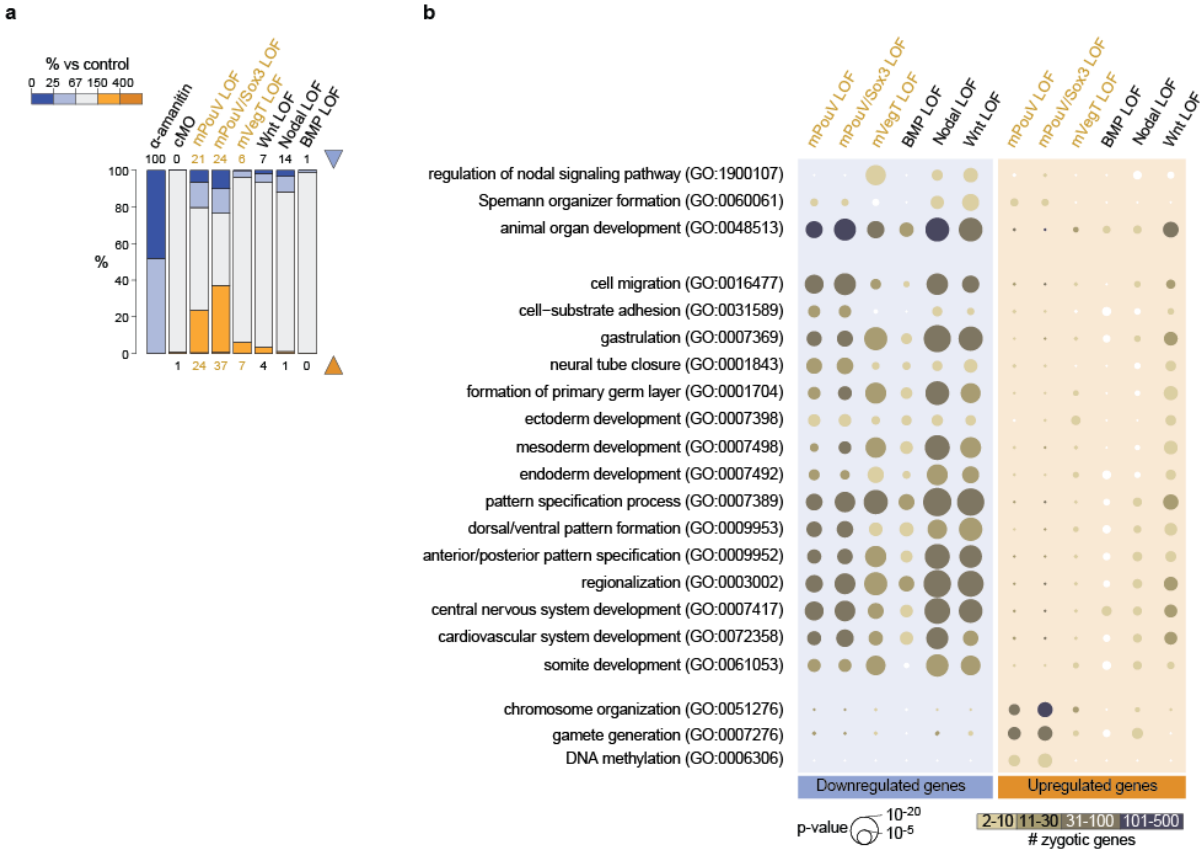
Numbers attached to any embryo images refer to the count of embryos detected with the displayed phenotype among all embryos analysed. **a**, Morphological phenotypes caused by single and combined LOFs of Sox3 and mPouV when control embryos reached mid-gastrula and mid-tailbud stage. **b**, Phenotypical 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 caused by the LOF of mVegT at late tailbud stage. **d**, 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: *, $p < 0.02$. **e**, Morphological phenotypes caused by the single LOF of canonical Wnt, Nodal or BMP signalling when control embryos reached late gastrula and mid-tailbud stage. **f**, Morphological phenotype and transcriptome analysis of embryos injected with α-amanitin. PCA of the poly(A) RNA transcriptome of control and α-amanitin-injected embryos at indicated developmental stages in biological triplicates (#1-3).



Extended Data Figure 7. Signal-induced tissue-specific gene expression depends on maternal TFs

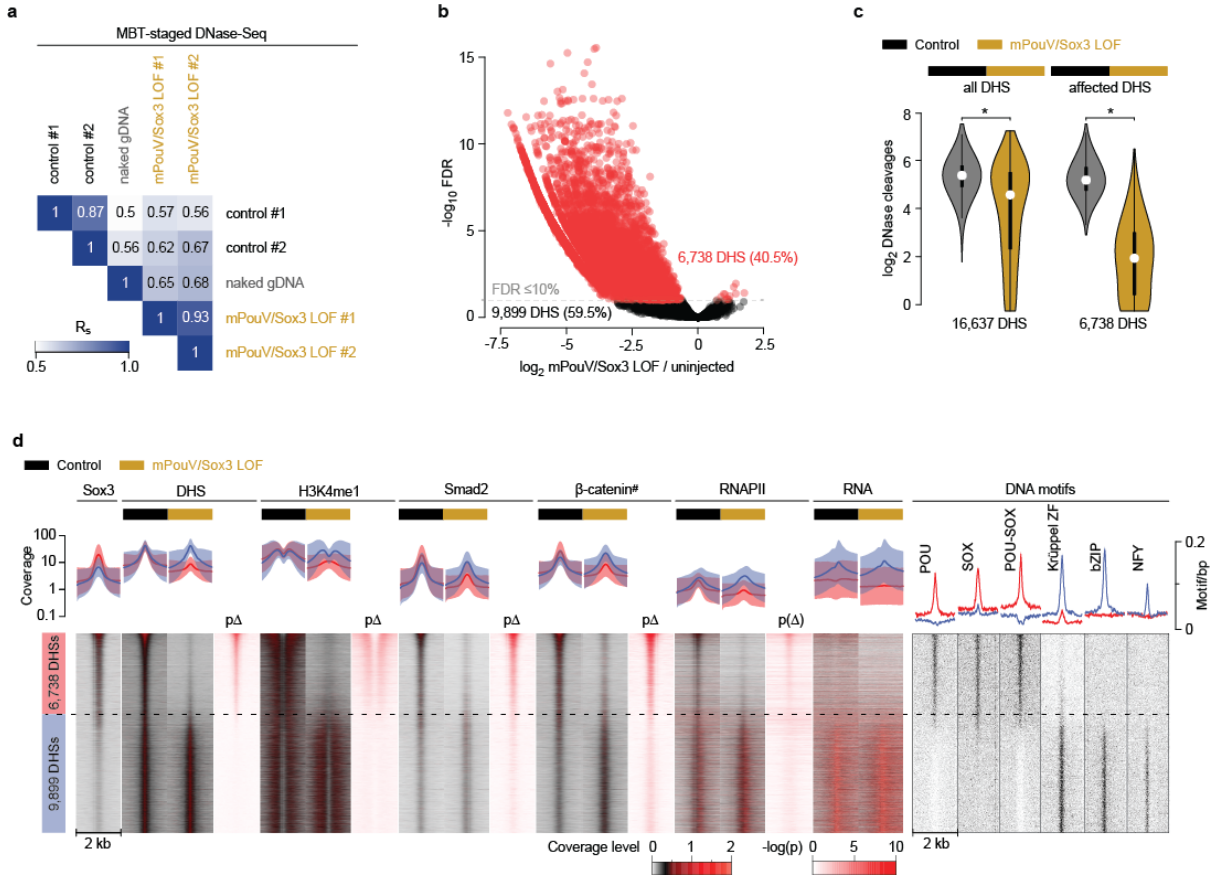
a, Detection of 3,318 zygotic genes with reduced transcript levels (≥50%) in α-amanitin-injected embryos of stage 8^+, 9^+ and 10^+ (Extended Data Fig. 6f). These genes were used as reference for all other LOFs. Dots in scatterplot are coloured according to the maternal contribution[37] to the transcript level of each of these zygotic genes. List of some developmentally relevant genes whose numbers are used in **b** and **d** and Fig. 3b. **b**, Scatterplot of transcript fold changes (FCs) caused by mPouV and mPouV/Sox3 LOFs with dots coloured according to the ratio of FCs. **c**,**g**, Early gastrula-staged WMISH: *mir427*, animal view; *ventx*, lateral view, ventral side facing right; *Xbra*, dorsal view; *foxb1* and *zic1*, vegetal view, dorsal side facing top. Numbers in the right bottom corner of each image refer to the count of embryos detected with the displayed WMISH staining among all embryos analysed

per condition and *in situ* probe. Scale bars, 0.5 mm. **d**, Transcriptional comparison of zygotic genes between indicated LOFs of maternal TFs or signals. Dots are coloured according to regional expression along the animal-vegetal (An:Vg) or dorso-ventral (D:V) axis. **e**,**f**, Venn diagrams of genes downregulated by indicated LOFs. **h**, Quantification of *foxb1* transcript levels in control and mPouV/Sox3 LOF animal caps with or without canonical Wnt signalling. Error bars, mean+SD (n=2). Two-tailed Student's t-test: 1, p=0.050; and 2, p=0.047.



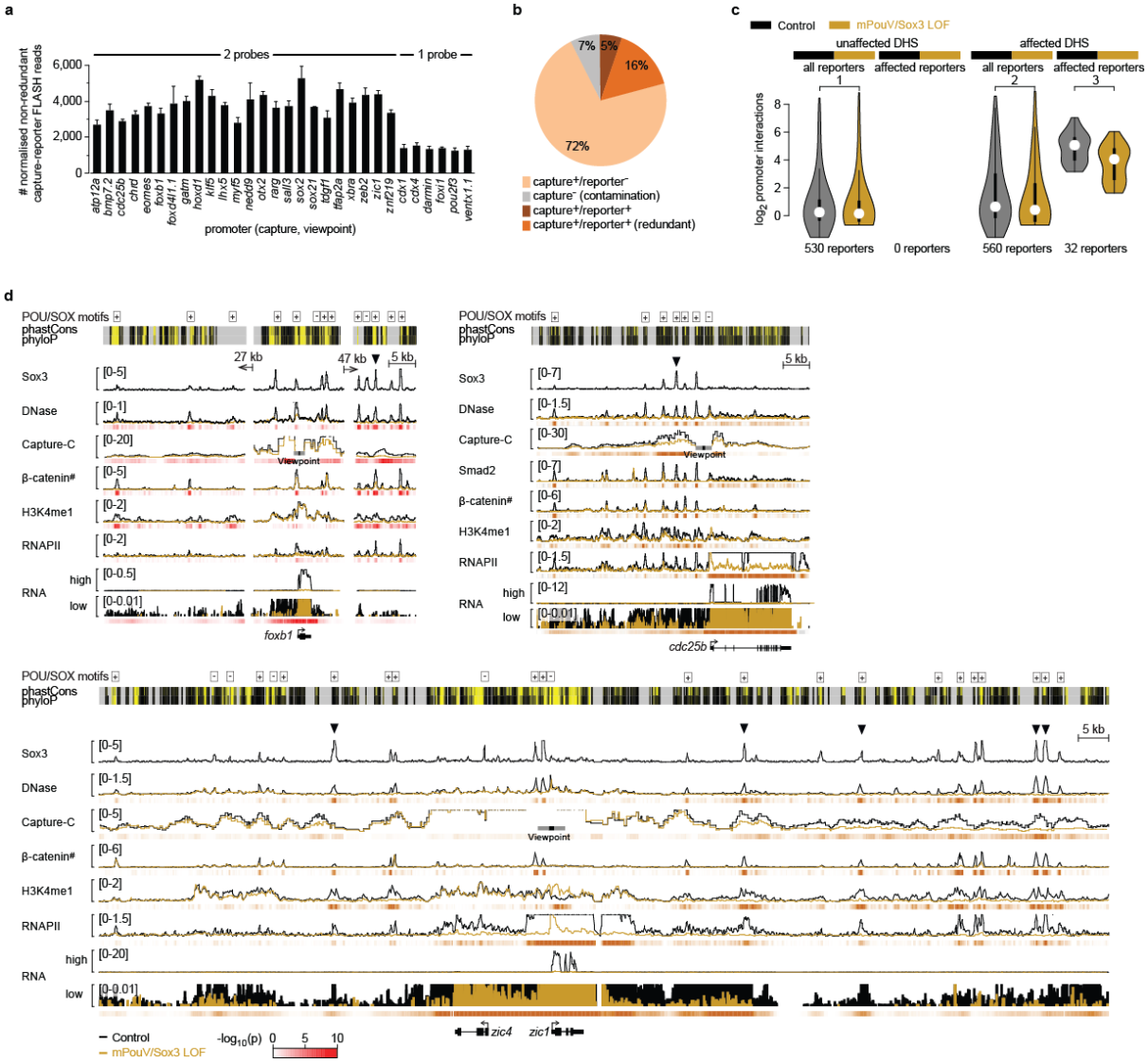
Extended Data Figure 8. Selected maternal TFs and signals regulate gene signature of earliest tissue regionalisation and animal organogenesis

a, Percentage of genes misregulated 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 colour indicates the number of affected genes.



Extended Data Figure 9. Pioneering activity of mPouV and Sox3 initiate extensive chromatin remodelling.

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 DNase cleavage changes at DHSs between mPouV/Sox3 LOF and uninjected embryos. **c**, Violin plot of DNase cleavages comparing uninjected and mPouV/Sox3 LOF embryos at all and affected (FDR <10%) DHSs. Wilcoxon test: *, $p < 2.2 \times 10^{-16}$. **d**, Map of various chromatin features, expressed RNA and enriched DNA motifs at DHSs sorted by significantly reduced DNase cleavages under mPouV/Sox3 LOF. #, β-catenin binding profile 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 meta-profile of DNA motif instances. Bottom row, normalised read coverage levels (and differential p-values, pΔ) followed by co-aligned map of DNA motifs.



Extended Data Figure 10. mPouV and Sox3 stabilise chromatin contacts with promoters **a**,**b**, Performance of next-generation capture-C using novel 3C (chromatin conformation capture) protocol adapted to early *Xenopus* embryos. **a**, Bar graph of normalised counts (mean+SD; n=6) of non-redundant capture-reporter FLASH[69] reads for each viewpoint captured with 1 or 2 probes. **b**, Summary of FLASH reads from sequencing 3C libraries after two rounds of hybridisation-based capture. **c**, Violin plot of promoter-proximal interactions with unaffected and affected DHS for both all and affected (FDR <10%) chromatin contacts. Wilcoxon test and effect size estimates: 1, $p=5x10^{-5}$ and $r_{effect}=0.12$ (small effect); 2, $p=5x10^{-30}$ and $r_{effect}=0.34$ (medium effect); and 3, $p=5x10^{-10}$ and $r_{effect}=0.78$ (large effect). **d**, Superimposed line plots showing the LOF-induced changes to various chromatin features at MBT including chromatin contacts with promoters (viewpoints) of *foxb1*, *cdc25b* and *zic1*. #, β-catenin binding profile generated at late blastula stage rather than early blastula stage (Sox3) or MBT (all others). RNA profile split into low and high expression windows. Low expression profile shows pervasively transcribed non-coding RNA. For statistical significance check heat map below each profile. Header, POU/SOX motif occurrences at DHSs and sequence conservation heat maps (phastCons, phyloP). Arrowheads denote strongly affected chromatin sites.

1483 **Supplementary Table 1**
1484 Maternal protein concentrations versus ribosome densities (translation) across the mid-
1485 blastula transition
1486
1487 **Supplementary Table 2**
1488 Transcriptional effect of maternal transcription factors and intercellular signals on zygotic
1489 genome activation
1490
1491 **Supplementary Table 3**
1492 Summary of deep sequencing and read alignments
1493
1494 **Supplementary Table 4**
1495 Genomic coordinates of selected capture-C viewpoints and probe DNA sequences used for
1496 hybridisation-based capture
1497
1498 **Supplementary Video 1**
1499 Gastrulation defects caused by the loss-of-function of maternal pluripotency factors