The Spatio-Temporal Control of Zygotic Genome Activation

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

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

[2]Department of Stem Cell and Developmental Biology, Pasteur Institute, 75015 Paris, France

[3]Lead contact: jim.smith@crick.ac.uk (J.C.S.)

*Correspondence: george.gentsch@crick.ac.uk (G.E.G.) and jim.smith@crick.ac.uk (J.C.S.)

SUMMARY

One of the earliest events in embryonic development is zygotic genome activation (ZGA). According to nascent transcript profiling most zygotic genes are inactive until the mid-blastula transition (MBT), and it has been suggested that only at this stage is the cell cycle slow enough, and the nuclear-to-cytoplamic (N/C) ratio of maternal repressors low enough, for bulk transcription to occur. Here we resolve the ZGA of the frog *Xenopus tropicalis* in time and space. We detect a gradual increase in quantity and length of RNA polymerase II-mediated elongation before the MBT, indicating that ZGA does not depend on a specific N/C ratio, and we observe that the size of newly transcribed genes is not necessarily constrained by cell cycle duration. We also reveal that canonical Wnt, Nodal and BMP signaling jointly generate most of the spatio-temporal dynamics of regional ZGA, converting the egg into a multi-axial embryo with proportionate germ layers.

INTRODUCTION

At the time of fertilization, the genomes of multicellular organisms are transcriptionally silent, and maternally deposited mRNA and protein direct the events of early development including zygotic genome activation (ZGA) (De Iaco et al., 2017; Gentsch et al., 2018b; Lee et al., 2013; Liang et al., 2008). The number of cell cycles after which ZGA becomes essential for development (at which embryos arrest if transcription is inhibited) is highly reproducible within each species. In zebrafish, *Xenopus* and *Drosophila*, this occurs after 10, 12 and 13 cell cycles, respectively, at the so-called mid-blastula transition (MBT) (Blythe and Wieschaus, 2015; Kane and Kimmel, 1993; Newport and Kirschner, 1982a). Early development in these species occurs with no gain in cytoplasmic volume, and studies in *Xenopus* suggested that the nuclear-to-cytoplasmic (N/C) ratio triggers ZGA when the increasing amount of nuclear DNA titrates out maternally deposited repressors (Newport and Kirschner, 1982b). Slower-developing mammalian embryos show major waves of transcription as early as the 2-cell

(mouse) (Hamatani et al., 2004) or the 4- to 8-cell stage (human) (Braude et al., 1988). This occurs days before the formation of the blastocyst, which, like the blastula, contains the pluripotent cells to form the embryo proper.

In *Xenopus*, ZGA triggers significant changes in cell behaviour after the MBT. First, rapid and nearly synchronous cell cleavages give way to longer and asynchronous cell divisions (Anderson et al., 2017; Newport and Kirschner, 1982a). Second, embryonic cells acquire the ability to respond to inductive signaling (Gentsch et al., 2018b), causing them to become motile, to establish dorso-ventral patterning, and to contribute to one or two of the three germ layers (endoderm, mesoderm and ectoderm). These germ layers emerge first during gastrulation and are the primordia of all organs. Third, embryos show accelerated degradation of maternal RNA, and fourth, cells gain apoptotic (Stack and Newport, 1997) and immunogenic (Gentsch et al., 2018a) capacities.

Conventionally, while large-scale ZGA occurs at the MBT, some genes escape the early transcriptionally repressive environment, and nascent transcripts can be detected in *Xenopus* embryos during rapid cleavage stages. For example, primary transcripts of the polycistronic MIR-427 gene are detectable in *X. tropicalis* after just three cell divisions (Owens et al., 2016). MIR-427, like its zebrafish equivalent MIR-430 (Lee et al., 2013), is strongly activated by the synergistic and pioneering activities of maternal members of the SoxB1 and Pou5F (Oct4) transcription factor (TF) families (Gentsch et al., 2018b). These core pluripotency TFs, represented by Sox3 and Pou5f3 in *Xenopus*, are characterized by ubiquitous and high translation frequencies in pre-MBT embryos (Gentsch et al., 2018b; Lee et al., 2013). Zygotic transcription of the Nodal ligand encoding genes *nodal3/5/6* and homeobox genes *siamois1/2* is initiated by nuclear β-catenin as early as the 32-cell stage (Owens et al., 2016; Skirkanich et al., 2011; Yang et al., 2002).

While miR-427 contributes to the clearance of maternal RNA, *nodal* and *siamois* genes establish the three germ layers and the Spemann organiser. All these genes, and other early-activated genes in *Drosophila*, *Xenopus* and zebrafish, are characterized by a coding sequence of less than 1 kb that either lacks introns or has just a few (Heyn et al., 2014). It has been suggested that the early rapid cell cycles cause DNA replication machinery to interfere with the transcription of large genes (Shermoen and O'Farrell, 1991), a suggestion supported, to date, by the profiling of nascent transcripts. We note, however, that the detection and temporal resolution of de novo transcription can be particularly challenging for genes showing both maternal and zygotic transcripts.

Here we use the continuous occupancy of RNAPII along gene bodies as a method to record ZGA. In contrast to transcript profiling techniques, this method (1) directly determines the activity of every gene; (2) is independent of metabolic labeling (Heyn et al., 2014) or any gene

feature such as introns (Lee et al., 2013), single nucleotide polymorphisms (Harvey et al., 2013) and transcript half-lives; and (3) circumvents difficulties in detecting nascent transcripts in a large pool of maternal transcripts. Combined with the profiling of the transcriptome along the primary body axes (Blitz et al., 2017), we resolve ZGA in time and space for wild-type and various loss-of-function embryos. We provide evidence that runs counter to our original understanding of the cell cycle or of the N/C ratio in constraining gene expression before MBT. And finally, we show how signaling initiates and coordinates spatio-temporal ZGA in the *Xenopus* embryo.

RESULTS

RNAPII Profiling Reveals Exponential ZGA before MBT

In an effort to resolve the progression of ZGA, we profiled chromatin for RNAPII engagement on hand-sorted *X. tropicalis* embryos over six developmental stages from the 32-cell to the late gastrula stage (Figure 1A,B). RNAPII was localized on the genome by chromatin immunoprecipitation followed by deep sequencing (ChIP-Seq) (Gentsch et al., 2018b). We complemented RNAPII profiling with high time-resolution transcriptomics (Owens et al., 2016) counting both exonic and intronic RNA at 30-min intervals from fertilization to the late gastrula stage (Figures 1A,B and S1A,B). For both maternal and zygotic genes, the detection threshold was set to ≥3 transcripts per million (TPM) averaged over any 1-h window during this developmental period. This restricted the analysis to 13,042 genes (Figure 1B). Genes were considered active when both RNAPII enrichment along their full length (see Methods) and corresponding transcripts (≥0.1 TPM) were simultaneously detected. RNAPII-guided ZGA profiling was verified in part by active post-translational histone marks (Hontelez et al., 2015) and by differential expression methods aiming at detecting nascent transcripts; zygotic transcript depletion (by blocking RNAPII elongation with α-amanitin) (Gentsch et al., 2018b) or enrichment (by selecting 4-thiouridine [4sU] tagged transcripts at the MBT and the mid-gastrula stage) showed substantial overlaps and positive correlations with RNAPII-covered genes (Figures 1A,B and S1C,D and Tables S1 and S2).

This analysis revealed an exponential ZGA before the MBT with 27, 144 and 1,044 active genes after 5 (32-cell, ~2.5 hpf), 7 (128-cell, ~3 hpf) and 10 (1,024-cell, ~4 hpf) cell cycles, respectively. Gene activation reached its peak around the MBT (~4.5 hpf), with 1,885 newly-activated genes, before dropping to 724 gene at the early-to-mid gastrula stage (~7.5 hpf) and increasing again to 1,214 genes towards the end of gastrulation (~10 hpf) (Figures 1B,C and S1E and Table S2). While most zygotic genes remain active beyond the mid-gastrula stage, 197 (including *siamois2* [*sia2*], *nodal5* and *znf470*) of the 4,836 zygotic genes (~4%) are deactivated within ~6 h of development (Figures 1C and S1F,G). Slightly less than one third of the activated genes had a regional expression pattern across either or both the

animal-vegetal (future antero-posterior) and dorso-ventral body axes (Figures 1D and S1F). The temporal order of enriched biological processes supported by ZGA matched the regulatory flow of gene expression starting with nucleosome assembly, nucleic acid synthesis, mRNA metabolism and production, post-translational modification and degradation of proteins (Figure 1E). Earliest transcriptional engagement was detected in gene clusters of tens to hundreds of kilobases (Figures 1A,C and S1H-J). These clusters featured close relatives of the same gene, some of which are critical to Nodal signaling (Nodal ligands), the formation of the Spemann organiser (Siamois homeobox transcription factors), nucleosome assembly (histones), mRNA decay (MIR-427), and further gene regulation (zinc finger [ZF] transcription factors with on average 10 Cys2-His2 [C2H2] domains; Figure S1J). These earliest activated genes were shorter and encoded smaller proteins than those within the maternal pool or those activated post-MBT (Figures 1F and S1K). The non-coding features that contributed most to the difference in length were the 3' UTRs and introns (Figure S1K).

We noted that the shorter zygotic genes observed before the MBT did not strictly correlate with the time constraints imposed by short cell cycles. We detected increasing and wider spread of de novo recruitment of RNAPII before the MBT when cleavages occur at fast and near-constant pace (Figures 1F and S1K). During this period, the median length of activated genes (and their coding sequences) increases from ~0.9 kb (~0.4 kb) to ~5.9 kb (~0.9 kb). It was not until after the MBT that the overall architecture of zygotic and maternal genes became indistinguishable (Figures 1F and S1K). Temporal comparison of RNAPII engagement and total RNA profiling suggested that the 'nominal' zygotic contribution to the transcriptome rose within seven cell cycles from ~0.2% at the 32-cell stage to ~22% at the MBT (Figure 1G). Further maternal degradation and more moderate transcriptional engagement extended the zygotic contribution to about one third of the transcriptome by the late gastrula stage. Apart from the pre-MBT ZGA, we detected maternal transcripts (≥0.1 TPM) for >90% of the activated genes (Figure 1G).

Wnt, Nodal and BMP Signals Are Key Drivers of Regional ZGA

Next, we sought to investigate the single and combined effects of different inductive signals on the spatio-temporal dynamics of ZGA. An early vertebrate embryo employs canonical Wnt, Nodal and BMP signals and their key transcriptional effectors β-catenin, Smad2 and Smad1, respectively, to establish the primary body axes and the three germ layers (reviewed by Arnold and Robertson, 2009; Kimelman, 2006). In *Xenopus*, β-catenin first translocates to the nuclei of dorsal blastomeres at the 32-cell stage (Larabell et al., 1997; Schneider et al., 1996) (Figure 2A). After the MBT, zygotic Wnt8a causes more nuclear β-catenin to accumulate around the upper lip of the forming blastopore (Christian and Moon, 1993; Schohl and Fagotto, 2002). The nuclear translocation of Smad1 and Smad2 is triggered around the MBT by various BMP and Nodal ligands. Nuclear Smad1 is primarily detected on the ventral side and the blastocoel roof of the embryo while nuclear Smad2 is detected within the vegetal

4

hemisphere and the marginal zone (Faure et al., 2000; Schohl and Fagotto, 2002) (Figure 2A). In an effort to inhibit canonical Wnt signaling, we injected into the *X. tropicalis* zygote a previously validated antisense morpholino oligonucleotide (MO) which interferes with β-catenin protein synthesis by annealing to the translation start codon (Gentsch et al., 2018b; Heasman et al., 2000). Nodal and BMP signals were selectively blocked by incubating dejellied embryos from the 8-cell stage onward with verified cell-permeable inhibitors SB431542 (Ho et al., 2006; Inman et al., 2002) and LDN193189 (Cuny et al., 2008; Young et al., 2017), respectively. The morphological phenotypes of these single loss-of-function (LOF) treatments were consistent with previous observations and ranged from impaired axial elongation causing the loss of tail structures (BMP, Reversade et al., 2005) to severe gastrulation defects (Wnt, Heasman et al., 2000, and Nodal, Ho et al., 2006) as shown in Figure 2B. Briefly, Nodal LOF impaired blastopore lip formation and bulk tissue movements of gastrulation (bullet points in Figure 2B). However, it did not preclude subsequent elongation of the antero-posterior axis. By contrast, Wnt LOF embryos underwent gastrulation (albeit initiated later and more circumferentially than in a dorsal to ventral wave), but failed to form an antero-posterior axis including head and tail. With respect to the joint effects of Wnt, Nodal and BMP signaling, most dual and triple LOFs added up their morphological defects such that, for example, Wnt/Nodal LOF resulted in the complete loss of gastrulation and axial elongation. Only Wnt/BMP LOF produced a seemingly non-additive phenotype including non-fusing neural folds (arrowheads in Figure 2B).

Next, changes to ZGA caused by the single or combined LOF of Wnt, Nodal and/or BMP were determined at the late blastula stage on a transcriptome-wide scale using deep RNA sequencing. The analysis was limited to the 3,315 zygotic genes for which spatio-temporal expression data was available (Blitz et al., 2017; Owens et al., 2016) and reduced expressions (≥50% loss of exonic and/or intronic transcript counts, FDR ≤10%) could be detected in α-amanitin-injected embryos (Figure 2C and Table S3) (Gentsch et al., 2018b). α-Amanitin-mediated inhibition of RNAPII elongation impedes the morphogenetic tissue movements of gastrulation and ultimately leads to early embryonic death (Gentsch et al., 2018b). The spatial expression was previously measured between dissected parts of an early gastrula embryo representing two of the three Cartesian body coordinates, that are the animal-vegetal and the dorso-ventral axes (Blitz et al., 2017). We did not include the left-right axis as no significant transcript level differences across this body axis were detected by gastrula stage (Blitz et al., 2017). The signal-mediated transcriptional effects (1.5-fold change from control RNA level) on zygotic genes, 86% (2840/3315) of which have ≥0.1 TPM maternal contribution, ranged from ~1.5% (~1.3% down and ~0.2% up) to ~26% (~19% down and 7% up) for single BMP LOF and triple Wnt/Nodal/BMP LOFs, respectively (Figure 2C). As expected the transcript levels of entirely zygotic genes were stronger affected than that of zygotic genes with maternally contributed transcripts (Figures 2C and 4A). They fairly

reflected the severity of the resulting morphological phenotypes at the late gastrula and the mid-tailbud stage (Figure 2B).

In comparison, the LOFs of critical maternal TFs like Pou5f3/Sox3 or VegT (Gentsch et al., 2018b) misregulated 61% (~24% down and ~37% up) and 13% (~6% down and ~7% up) of the zygotic genes, respectively. The quadruple LOFs of zygotic T-box TFs (zVegT, eomes, t and t2), whose expressions depend on Nodal signaling, mildly misregulated 19% (~9% down and ~10% up) of the zygotic genes as detected later over three consecutive developmental time points during gastrulation (Table S3). Among the ZGA-enriched biological functions (Figure 1D), Wnt, Nodal and BMP signals, like the maternal Pou5f3/Sox3 and VegT, strongly affected the ZGA of animal organogenesis including genes associated with cell migration, gastrulation (segregation of the germ layers ectoderm, mesoderm and endoderm), dorso-ventral and antero-posterior body axis formation and regionalization (Figure 2D). Impaired tissue movements during gastrulation as observed in various LOFs (Figure 2B and Movie S1) was best indicated by the strong enrichment for cell migration-associated genes. The genes suppressed or unaffected by the selected signals and maternal TFs were enriched for the ZGA-critical biological processes of mRNA metabolism and translation. For instance, the transient activation of the entire zinc finger cluster (Figure S2A) was not affected by any tested LOF. Because family members are frequently cross-regulated and the MBT-staged chromatin contains many Krüppel-like zinc finger 'footprints' (Gentsch et al., 2018b) it is conceivable that the unaffected, tissue-nonspecific part of ZGA is regulated by maternal TFs like KLF4. This vertebrate gene regulatory branch may be more ancient as zinc finger TFs like Zelda are also key to ZGA of the invertebrate *Drosophila* (Liang et al., 2008).

Next, signal-dependent ZGA was resolved in time and space based on the profiling of RNAPII-engaged chromatin from the 32-cell stage to the MBT and transcript enrichments (Blitz et al., 2017) across the animal-vegetal and dorso-ventral axes (Figures 2E and S2B-F). In line with the nuclear translocation of their signal mediators (Figure 2A), Wnt, Nodal and BMP induced tissue-specific genes in different spatio-temporal domains of the early embryo: β-catenin induced ~87% and ~46% of genes preferentially expressed on the dorsal side and in the vegetal hemisphere (VH) / marginal zone (MZ), respectively. Some of its target genes like *nodal3.1* and *sia2* were already active by the 32-cell stage (Figures 1C and 2E). The early Wnt LOF-mediated transcription loss was followed by the misregulation of opposing cell fate specifiers: the upregulation of ventral genes (e.g., *id2*, *szl*) versus the downregulation of dorsal genes (e.g., *chrd*, *otx2*) suggesting that β-catenin protects the dorsal fate from ventralization (Figures 2E and S2C). Similarly, Nodal predominantly activated dorsal (~63%) and VH/MZ-specific (~73%) genes although with no effect on earliest activated genes at the 32-cell stage and opposing cell fate regulators (Figure S2A,B). The MZ-specific FGF ligand *fgf20* activated by the 128-cell stage was among the first Nodal-responsive genes (Figures 1C and S2A,B). By contrast, BMP on its own only contributed to ventrally enriched transcription

(~45%) from the 1,024-cell stage onwards (Figure S2A,C). As a comparison, the ubiquitous expression of maternal Pou5f3/Sox3 facilitated transcription within all spatio-temporal domains including the uniform expression of miR-427 (Figure S2A,D). Overall, however, Pou5f3/Sox3 preferentially acts on tissue-specific genes, in particular, those expressed within animal- (~55%) and ventral-specific (~67%) domains (Figure S2A,D). Likewise, the maternal TF VegT only upregulated genes (~40%) within its expression domain, the vegetal hemisphere, while suppressing opposite, animal-specific fate specifiers (e.g., *foxi4.2*). VegT showed equal preference for ventral- and dorsal-specific (~31% and ~30%) genes.

Wnt/BMP Synergy Enables Uniform ZGA Across the Dorso-Ventral Axis

Next, the relationships between inductive signals to regionalize ZGA was explored by comparing the zygotic transcriptome between single and dual inhibitions. Interestingly, while Nodal/BMP and Wnt/Nodal shared more additive relationships (Figures 3E and S3A-N), BMP can acted synergistically with Wnt (Figure 3A,F). This reflected the various morphological anomalies amassed from single to double LOFs (Figure 2B). Single LOF hardly revealed any common gene targets between Wnt and BMP signaling (Figure 3B), which at first was not surprising given their initial activity at opposite ends of the dorso-ventral axis. However, dual Wnt/BMP inhibition dramatically increased the number of downregulated genes by 292, which is a rise of ~118% and ~664% with respect to Wnt and BMP-dependent genes (Figure 3B). Interestingly, this synergy affected 166 Nodal-dependent genes, most of which had uniform expression levels across the dorso-ventral axis (Figure 3C,D,F-H). Thus, spatially-restricted Wnt, BMP and Nodal signals act together to establish dorso-ventral expression uniformity of genes such as *t* and *eomes* (Figure 3I). It is not entirely clear yet whether Wnt/BMP synergy arises from the joint chromatin engagement, mutual or post-translational interactions. For instance, Wnt8a signal transduction can enhance BMP transcriptional readouts by inhibiting GSK3 phosphorylations that target Smad1 for degradation (Fuentealba et al., 2007). Here, however, we were less likely to interfere with this biochemical pathway as Wnt signaling was precluded by the depletion of β-catenin, which did not diminish *wnt8a* transcription. Previous reports (Harvey et al., 2010), and the higher enrichment of canonical DNA recognition motifs for the Wnt-associated basic helix-span-helix TF AP-2 at Smad1 rather than Smad2 *cis*-regulatory binding sites, suggest that Wnt preferentially cooperates with BMP on the chromatin level (Figure S3O).

Overall, the loss of canonical Wnt, Nodal and/or BMP signaling misregulated ~39% (~22.1% down, ~2.1% down/up and ~14.4% up) of ZGA (Figure 4A). These signals induced most regional ZGA on the dorsal side (~89%) and in the VH/MZ (~82%) reaching virtually all tissue-specific expression (~98%, 56/57 genes) enriched in the dorso-vegetal/MZ quadrant (Figure 4B,C). The cascade of Wnt, Nodal and BMP signaling fully covered regional ZGA in most anatomical domains of the early gastrula embryo with the exception of animally enriched transcription (~19%). One of the most potent hierarchical gene regulatory network is that of β-

catenin initiating the Nodal pathway and the Spemann organiser, a source of BMP antagonists (e.g., chordin [chrd]) and head organizers (e.g., otx2) on the dorsal side of the embryo, by activating several Nodal ligands and the *siamois* gene cluster by the 32-cell stage, respectively. We note that animal- and ventral-specific expressions were stronger affected by ubiquitous maternal TFs like Pou5f3/Sox3 and depended rather on Wnt, Nodal and/or BMP signal repression (Figure 4B,D) and other maternal TFs like VegT protecting opposite cell fates (Figure S2F). Further research is required to estimate the regional influence of other early signals such as retinoic acid or FGF on ZGA.

DISCUSSION

Our study provides two major insights on how ZGA is initiated in time and space. First, RNAPII spreads increasingly across longer gene bodies before the MBT, when reductive cell cleavages at intervals of ~20 min (28ºC) subdivide the *Xenopus tropicalis* zygote into 4,096 blastomeres (Figure 4E). This indicates that rapid cell divisions do not necessarily restrict RNAPII elongation to short runs of ~1 kb, the average gene length among earliest activated genes. In fact, recent long-read sequencing of the early zebrafish transcriptome identified a zygotic 8-kb transcript spanning multiple pri-miR-430 elements (Nudelman et al., 2018). Thus, it remains to be determined why RNAPII, despite its abundance, is initially prevented from more wide-spread elongation. Multiple studies show that the temporal progression of ZGA is preceded by or coincides with gradual chromatin remodeling at multiple levels, from the accessibility of *cis*-regulatory elements (Gentsch et al., 2018b) to the spatial organisation of an initially unstructured chromatin landscape (Hug et al., 2017; Kaaij et al., 2018; Ke et al., 2017). The gradual increase of elongated RNAPII chromatin engagement before the MBT also suggests that ZGA occurs over time and does not depend on a MBT-specific N/C ratio. Similar conclusions were drawn from profiling the zygotic transcriptome of haploid *Drosophila* (Lu et al., 2009) and cell cycle-arrested zebrafish (Chan et al., 2018). However, the highest number of newly engaged gene bodies is indeed reached at the MBT, confirming it as the developmental stage at which bulk ZGA occurs (Langley et al., 2014). Furthermore, the N/C ratio remains more critical for other aspects of the MBT. For instance, the DNA-based titration of four replication factors slows the cell cycle (Collart et al., 2013).

Nevertheless, it is possible that cell cycles contribute to the temporal progression of ZGA. We observed that the number of activated genes increases exponentially with the number of cell cell cycles before the MBT. Cell cycles are considered important to accelerate chromatin remodeling by displacing suppressors in mitotic chromatin and providing unique access to TFs (Halley-Stott et al., 2014) and structural proteins of high-order chromatin (Ke et al., 2017). For example, maternal core histones were shown to prevent premature ZGA by competing with specific TFs (Joseph et al., 2017).

In addition to the short lengths of the earliest activated genes, we observed that most of these genes code for groups of related factors like histones or zinc finger TFs, and that they appear as clusters spanning up to several hundred kilobases. The number and spatial proximity of clustered genes enhances the transcriptional output by sharing multiple cis-regulatory elements (arranged as super-enhancers) (Whyte et al., 2013) and fortifying transcriptional condensates of Mediator coactivator and RNAPII (Cho et al., 2018). Overall, based on enriched gene functions, we discovered that ZGA progressively initiates steps of gene expression control from nucleosome remodeling before the MBT to protein degradation after the MBT.

The second insight is that we can assign a large proportion of spatio-temporal ZGA to key signaling pathways. Canonical Wnt, Nodal and BMP largely govern regional ZGA in line with the nuclear translocation of their signal mediators. In terms of space, Nodal signaling mainly affects transcription within the vegetal hemisphere, including the animally-derived marginal zone, while Wnt and BMP initiate dorsally and ventrally enriched transcription, respectively. In summary, Wnt, Nodal and BMP are key drivers of regional ZGA (Figure 4E). ZGA timings also depend on the sequential translocation of signal mediators such that first, nuclear β-catenin directs earliest regional ZGA at the 32-cell stage, followed by nuclear Smad2 and Smad1 at the 128-cell and 1,024-cell stage, respectively, both of which require ZGA to express their ligands in previous developmental stages.

We also show that the synergy of opposing signals of the Wnt and BMP pathway affects many Nodal-dependent genes with uniform expression along the dorso-ventral axis such as *eomes* and *Brachyury* (*t*). Analysis of a zebrafish *Brachyury* gene orthologue suggests that *cis*-regulatory elements can integrate this triangle of signals to produce dorso-ventral expression uniformity (Harvey et al., 2010). We therefore propose that signal mediators are critical to regional ZGA and that they balance initially opposing cell fate commitments. However, their transcriptional roles rely on maternal pioneer TFs like Pou5f3 and Sox3 to make their signal-responsive *cis*-regulatory elements accessible for binding (Gentsch et al., 2018b). Critically, signal mediators are potent gene regulators, but they lack pioneering activity, which ensures tissue-specific differences in transcriptionally responding to the same signals. For example, signal-induced *Brachyury* expression requires maternal Pou5f3 and Sox3 (Gentsch et al., 2018b).

In summary, our results favour a ZGA model in which pioneer factors translated from maternal RNA remodel a transcriptionally repressive chromatin landscape. The elapsed time to achieve transcriptional competency may be shortened by the number of DNA replication, but importantly depends neither on a specific N/C ratio and nor on a lengthening of the cell cycle. Nuclear signal mediators and the transcription machinery then act upon the unlocked

9

cis-regulatory elements to orchestrate regional ZGA and generate a multi-axial embryo with proportionate germ layers.

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures, three tables and one movie and can be found with this article online.

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

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

DECLARATION OF INTERESTS

The authors declare no competing interests.
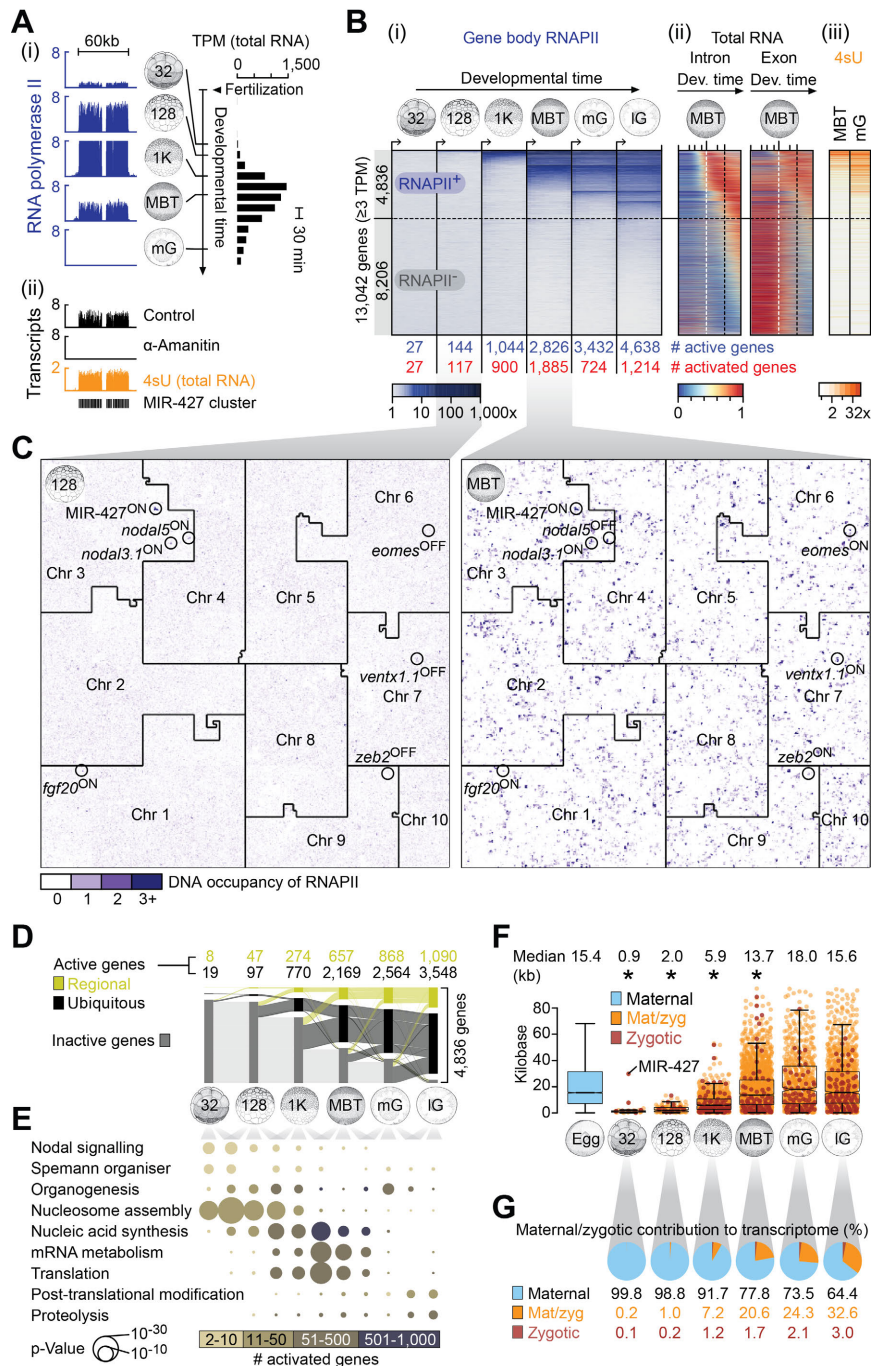
10

MAIN FIGURES

Figure 1. Dynamics and Architecture of ZGA in *X. tropicalis*

(A) (i) Strategy to record ZGA by profiling RNAPII and total RNA (Owens et al., 2016) using deep sequencing technologies. (ii) Alternative detection of active genes by α-amanitin-induced loss and 4sU-enrichment of nascent (zygotic) transcripts.

(B) Progression of zygotically activated genes (≥3 TPM) from the 32-cell to the late gastrula stage based on (i) full-length RNAPII occupancy (i.e., RNAPII was enriched across entire gene bodies; see Methods). Co-aligned: (ii) High time-resolution of total RNA, separated by intron- and exon-derived signals, from fertilization to the late gastrula stage, and (iii)

enrichments of 4sU-tagged RNA at the MBT and the mid-gastrula stage. Numbers below RNAPII heatmap represent counts of active (blue) and activated (red) genes at the indicated developmental stages. The horizontal dotted line separates RNAPII-engaged (RNAPII$^+$) from non-engaged (RNAPII$^-$) genes as detected until the late gastrula stage. The vertical dotted lines indicate the developmental time points of the MBT (white) and the late gastrula stage (black).

(C) 2D space-filling (Hilbert) curves showing RNAPII recruitment to chromosomes (Chr) at the 128-cell stage and the MBT. A few zygotic genes are highlighted as being active (ON) or not (OFF) based on their engagement with RNAPII.

(D) Alluvial diagram of spatio-temporal ZGA. Tissue-specificity inferred from regional transcript enrichments along the animal-vegetal or the dorso-ventral axes or both (Blitz et al., 2017).

(E) ZGA-associated enrichment of biological processes.

(F) Length of maternal and/or zygotic genes. *, p <1.9e-7 (Wilcoxon rank-sum test against maternal and post-MBT activated genes); r_{effect} 0.06 ('MBT' vs 'Egg') - 0.48 ('128' vs 'mG').

(G) Nominal contribution to the transcriptome deduced from full-length RNAPII occupancy and maternally inherited RNA.

Abbreviations: 32, 32-cell; 128, 128-cell; 1K, 1,024-cell; MBT, mid-blastula transition; mG, mid-gastrula; lG, late gastrula; 4sU, 4-thiouridine; Mdn, median; TPM, transcripts per million.

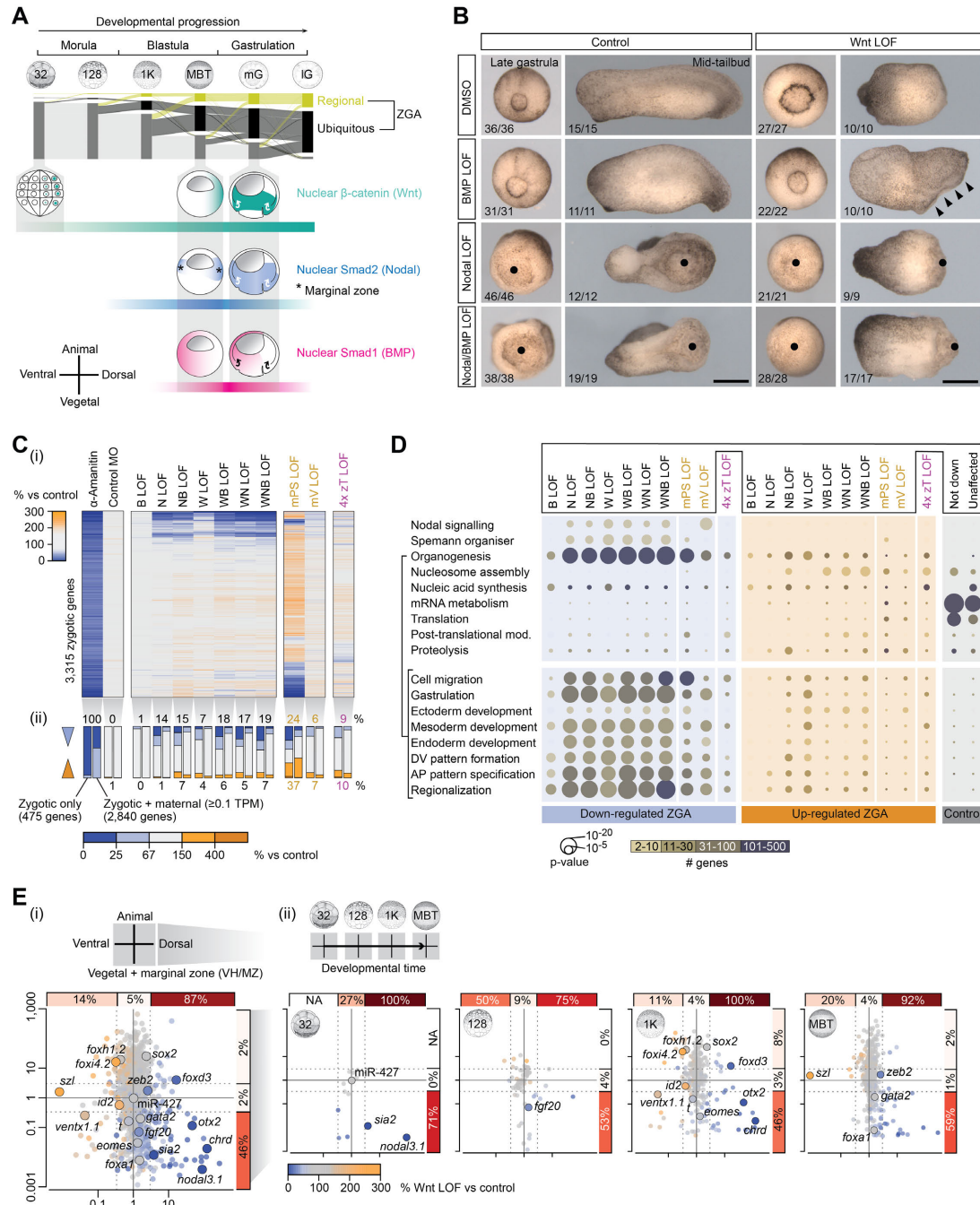See also Figure S1 and Tables S1 and S2.

Figure 2. Spatio-temporal ZGA Regulated by Canonical Wnt, Nodal and BMP Signals

(A) Spatio-temporal ZGA and nuclear localization of signal mediators β-catenin (canonical Wnt), Smad2 (Nodal) and Smad1 (BMP) (Faure et al., 2000; Larabell et al., 1997; Schohl and Fagotto, 2002; Stennard et al., 1999) from the 32-cell to the late gastrula stage.

(B) Morphological phenotypes of single and combined signal LOFs at the late gastrula and the mid-tailbud stage. Left ('control') pictures are taken from Gentsch et al. (2018b). Bullet points, failed blastopore formation. Arrowheads, excessive neural fold formation. Scale bar, 0.5 mm.

(C) Heatmap (i) and bar graph summary (ii) of ZGA misregulated in various LOF embryos: α-amanitin, positive control; control MO, negative control. Abbreviations: B, BMP; N, Nodal; W,

canonical Wnt; mPS, maternal Pou5f3/Sox3; mV, maternal VegT; 4x zT, four zygotic T-box TFs (zygotic VegT, Eomes, T and T2).

(D) Biological processes enriched with mis-regulated and control (not down-regulated or unaffected by the loss of maternal TFs or signaling) sets of zygotic genes under indicated LOFs.

(E) Summary (i) and temporal resolution (ii) of Wnt LOF effects on regional ZGA. Percentages only refer to the down-regulated genes among all zygotic genes with similar expression ratios along the animal-vegetal or the dorso-ventral axes.

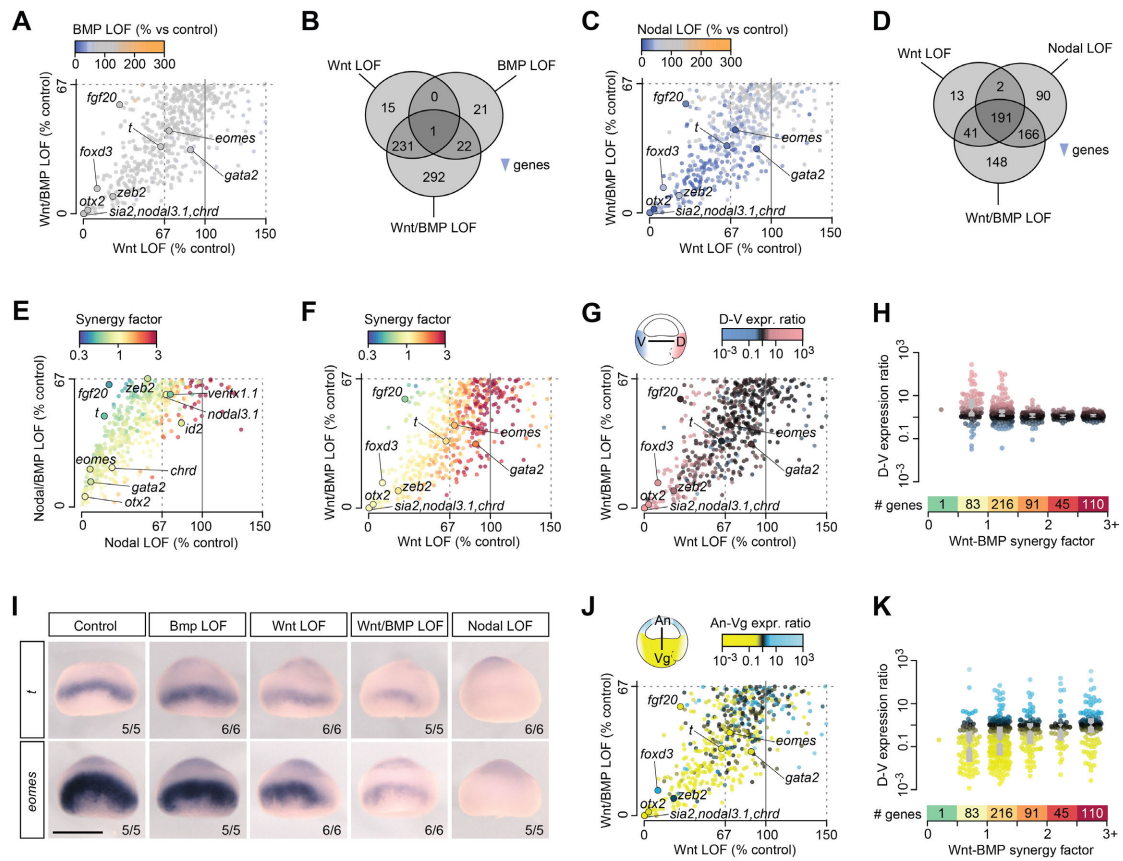See also Figure S2, Tables S1 and S3, and Movie S1.

Figure 3. Uniform Gene Expression Across Dorso-Ventral Axis Achieved by Wnt/BMP Synergy.

(A,C,E,F,G,J) Scatter plots of relative (% to control) transcript levels between indicated LOFs with each dot (gene) color-coded according to a third attribute: (A,B) relative (% to control) transcript levels, (E,F) synergy factors between single inductive signals, and (G,J) regional expression ratios between opposite ends of the indicated axis.

(B,D) Venn diagram of down-regulated genes by indicated LOFs.

(H,K) Beeswarm plot of regional expression (as measured along the indicated axes) depending on increased Wnt-BMP synergy.

(I) WMISH of *t* and *eomes* transcripts under various LOFs. Control and Nodal LOF pictures are from (Gentsch et al., 2018b).
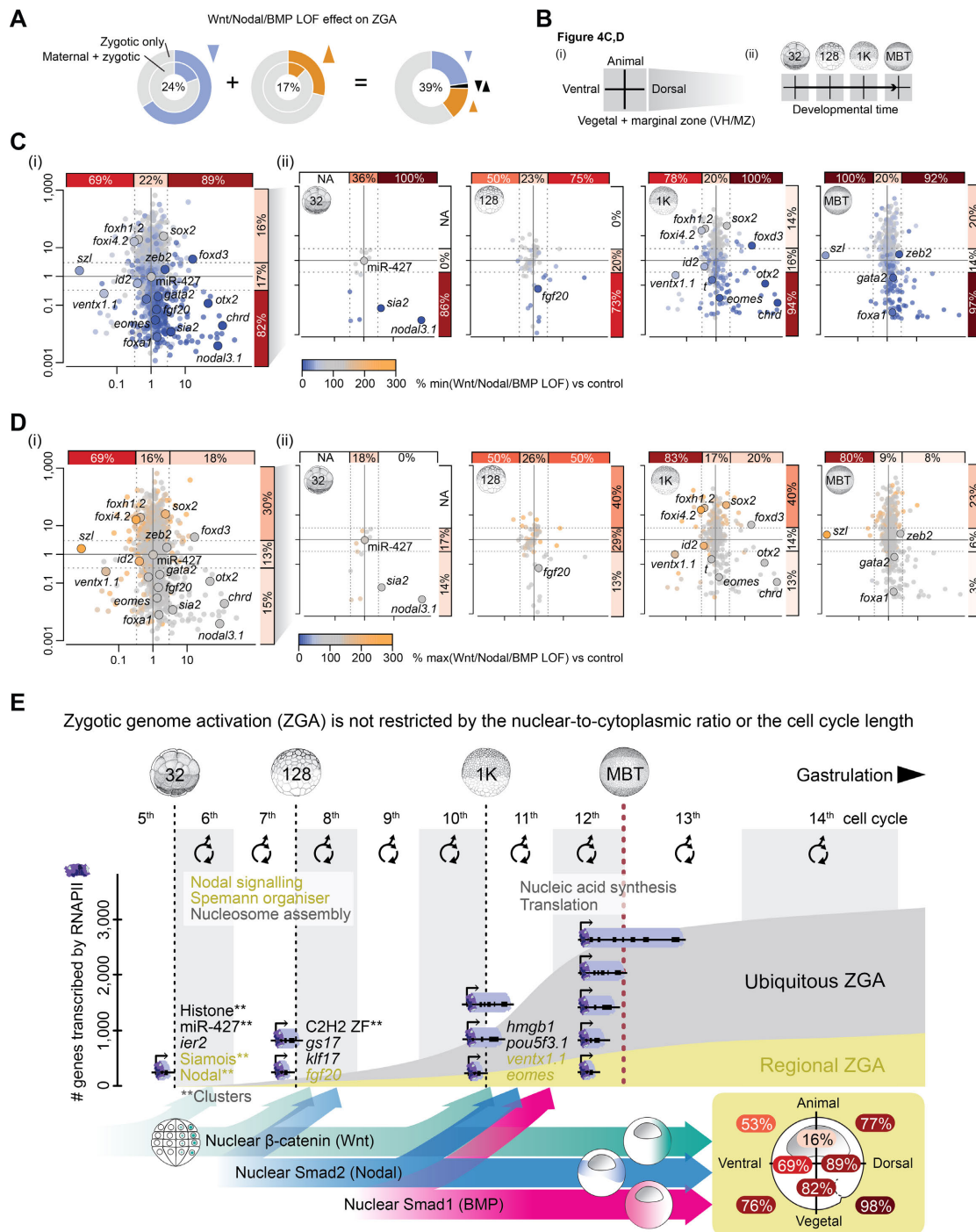
See also Figure S3 and Table S3.

15

Figure 4. Canonical Wnt, Nodal and BMP Signals Induce the Majority of Regional ZGA.

(A) Total percentage of active (zygotic only and maternal-zygotic) genes mis-regulated by Wnt, Nodal and/or BMP LOF.

(B) Graphical explanations of figure panels (C) and (D).

(C,D) Summary (i) and temporal resolution (ii) of minimal (C) and maximal (D) transcript levels of active genes (separated by the regional expression along the primary body axes) detected among Wnt, Nodal and/or BMP LOFs. Percentages only refer to the down-regulated (C) or up-regulated (D) genes among all zygotic genes with the same range of expression ratios along the animal-vegetal or dorso-ventral axes.

16

(E) Exponential activation of gradually longer genes before the MBT (bulk ZGA) when cell divisions occur at rapid and nearly constant intervals (~20 min at 28ºC). Sequential induction of the canonical Wnt, Nodal and BMP pathway is critical to high percentages of regional ZGA (as measured along the two primary body axes within the indicated halves and quadrants of an early gastrula embryo): e.g. ~89% or ~98% of gene expression enriched in the dorsal half or vegetal-dorsal quadrant, respectively.

STAR METHODS

KEY RESOURCES TABLE

CONTACT FOR REAGENTS AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, James C. Smith (jim.smith@crick.ac.uk).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Xenopus tropicalis Manipulation

Standard procedures were used for ovulation, fertilization, and manipulation and incubation of embryos (Khokha et al., 2002; Sive et al., 2000). Briefly, frogs were obtained from Nasco (Wisconsin, USA). Ovulation was induced by injecting serum gonadotropin (Intervet) and chorionic gonadotropin (Intervet) into the dorsal lymph sac of mature female frogs. Eggs were fertilized *in vitro* with sperm solution consisting of 90% Leibovitz's L-15 medium (Thermo Fisher Scientific) and 10% fetal bovine serum (Thermo Fisher Scientific). After 10 min, fertilized eggs were de-jellied with 2.2% (w/v) L-cysteine (Merck) equilibrated to pH 8.0. Embryos were cultured in 5% Marc's Modified Ringer's solution (MMR) (5 mM NaCl, 0.1 mM KCl, 0.1 mM $CaCl_2$, 0.05 mM $MgSO_4$ and 0.25 mM HEPES pH7.5) at 21ºC-28ºC. Embryos were staged according to (Nieuwkoop and Faber, 1994). All *Xenopus* work fully complied with the UK Animals (Scientific Procedures) Act 1986 as implemented by the Francis Crick Institute.

Chromatin immunoprecipitation (ChIP)

ChIP was carried out as detailed previously (Gentsch and Smith, 2017). Briefly, de-jellied *X. tropicalis* embryos were fixed at room temperature with 1% formaldehyde (Merck) in 1% MMR for 25 min. The fixation time was extended to 45 min for pre-gastrula stages. The following number of embryos were used for ChIP-Seq: 1,400 at the 32-cell stage, 1,000 at the 128-cell stage, 700 at the 1,024-cell stage, 450 at the MBT and 350 for the post-MBT stages. Fixation was terminated by rinsing embryos three times with ice-cold 1% MMR. Fixed embryos were homogenized in CEWB1 (150 mM NaCl, 1 mM EDTA, 1% (v/v) Igepal CA-630, 0.25% (w/v) sodium deoxycholate, 0.1% (w/v) sodium dodecyl sulfate and 10 mM Tris-HCl pH 8.0) supplemented with 0.5 mM DL-Dithiothreitol (Fluorochem) and protease inhibitors (Roche). The homogenate was left on ice for 5 min and then centrifuged at 1,000 g (4ºC) for 5 min. Homogenization and centrifugation was repeated once before resuspending the pellet in 1-3 ml CEWB1. Chromatin was solubilized and fragmented by microtip-mediated ultra-sonication (Misonix 3000 sonicator with a tapered 1/16-inch microtip). The solution of fragmented chromatin was cleared by centrifuging aat 16,000 g (4ºC) for 5 min. About 1% of the cleared chromatin extract was set aside for the input sample (negative control). The remaining chromatin was incubated overnight at 4ºC on a vertical rotor (10 rpm) with 20 µl of the mouse

monoclonal anti-RNAPII (8WG16, Covance) antibody. After adding 100 µl of washed protein G magnetic beads (Thermo Fisher Scientific) the solution was incubated for another 4 h at 4ºC on a vertical rotor (10 rpm). The beads were washed eight times in CEWB1 and once in TEN (10 mM Tris-HCl pH 8.0, 150 mM NaCl and 1 mM EDTA) at 4ºC. ChIP was eluted off the beads twice with 100 µl SDS elution buffer (50 mM Tris-HCl pH 8.0, 1 mM EDTA and 1% (w/v) sodium dodecyl sulfate) at 65ºC. ChIP eluates were pooled before reversing DNA-protein cross-links. Input (filled up to 200 µl with SDS elution buffer) and ChIP samples were supplemented with 10 µl 5 M NaCl and incubated at 65ºC for 6-16 h. Samples were treated with proteinase K (Thermo Fisher Scientific) and RNase A (Thermo Fisher Scientific) to remove any proteins and RNA from the co-immunoprecipitated DNA fragments. The DNA was purified with phenol:chloroform:isoamyl alcohol (25:24:1, pH 7.9) (Thermo Fisher Scientific) using 2.0-ml Phase Lock Gel Heavy microcentrifuge tubes (VWR) for phase separation and precipitated with 1/70 volume of 5 M NaCl, 2 volumes of absolute ethanol and 15 µg GlycoBlue (Thermo Fisher Scientific). After centrifugation, the DNA pellet was air-dried and dissolved in 11 µl elution buffer (10 mM Tris-HCl pH 8.5). The DNA concentration was determined on a Qubit fluorometer using high-sensitivity reagents for detecting double-stranded DNA (10 pg/µl to 100 ng/µl) (Thermo Fisher Scientific).

ChIP-Seq Library Preparation

Using the KAPA Hyper Prep Kit (Roche), 2.5-5 ng ChIP DNA or 5 ng input DNA were converted into indexed paired-end libraries as previously described (Gentsch and Smith, 2017). Briefly, DNA fragments were end-repaired and A-tailed for 30 min at 20ºC followed by 30 min at 65ºC before cooling to 4ºC. 7.5 pmol TruSeq (single index) Y-adapters (IDT) were ligated to the DNA fragment ends for 20 min at 20ºC. The DNA ligation product was extracted with 0.8x SPRI (solid phase reversible immobilisation) beads (Beckman Coulter) and amplified in five PCR cycles (15 sec at 98ºC, 30 sec at 60ºC and 30 sec at 72ºC) using the KAPA high-fidelity polymerase master mix (Roche) and 25 pmol Illumina P5 (forward) and P7 (reverse) primers (IDT). After cleaning up the PCR reaction with 1x SPRI beads, the DNA library was size-separated by electrophoresis using E-gel EX agarose gels (Thermo Fisher Scientific). A gel slice containing DNA ranging from 250 to 450 bp in size was dissolved shaking in 350 µl QG buffer (Qiagen) using a thermomixer (1,000 rpm) at room temperature. The DNA was purified with MinElute columns (Qiagen) and eluted off these columns twice using 11 µl elution buffer (10 mM Tris-HCl pH 8.5). The library was re-amplified using another 6-8 PCR cycles yielding 100-200 ng DNA without adapter dimer contamination. The DNA library was cleaned up with 1x SPRI beads.

Illumina Sequencing

All sequencing libraries were quality controlled: The DNA yield and fragment size distribution were determined by fluorometry and chip-based capillary electrophoresis, respectively. ChIP-Seq and RNA-Seq libraries were sequenced on the Illumina HiSeq 2500 and 4000, respectively, by

the Advanced Sequencing Facility of the Francis Crick Institute. Sequencing samples and read alignment results are summarized in Table S1.

Post-Sequencing Analysis of ChIP-Seq

Single reads of maximal 50 bases were processed using trim_galore v0.4.2 (Babraham Institute, UK) to trim off low-quality bases (default Phred score of 20, i.e. error probability was 0.01) and adapter contamination from the 3' end. Processed reads were aligned to the *X. tropicalis* genome assembly v7.1 and v9.1 (for Hilbert curves) running Bowtie2 v2.2.9 (Langmead and Salzberg, 2012) with default settings (Table S1). Alignments were converted to the HOMER's tag density format (Heinz et al., 2010) with redundant reads being removed (makeTagDirectory -single -tbp 1 -unique -mapq 10 -fragLength 175 -totalReads all). Only uniquely aligned reads (i.e. MAPQ ≥10) were processed. We pooled all input alignments from various developmental stages (Gentsch et al., 2018b). This created a comprehensive mappability profile that covered ~400 million unique base pair positions. For Hilbert curves, tag densities were generated across the genome v9.1 using sliding (200-bp increments) 400-bp window. Background signals (<0.3 reads per 1 million mapped reads) were removed. Blacklisted (Gentsch et al., 2018b) regions (except for MIR-427) were excluded using intersectBed (-v -f 0.5) from BEDtools (Quinlan and Hall, 2010).

Detecting Zygotic Genes Using RNAPII Chromatin Profiling

Normalized RNAPII and input tag densities were calculated across the gene body in 10 bins of equal size. Gene annotations v7.1 were altered based on a few known zygotic isoforms and some corrections obtained from assembling total and poly(A) RNA (Owens et al., 2016) from stage 6 to stage 12.5 *de novo* (Pertea et al., 2016). A few genes had previously been annotated as gene clusters due to assembly uncertainties. We reduced the annotation of polycistronic MIR-427 to the minus arm (scaffold_3b:3516900-3523400) and only monitored *nodal3.5* and *nodal5.3* within their respective gene clusters. Gene bodies with <40% mappability were removed. Here, the threshold of mappability per bin was set at 10% of the input read density averaged across all gene bodies in use. Subsequently, enrichment values were only obtained for all mappable bins by dividing read densities of RNAPII and input. Further, we restricted the analysis to genes for which ≥3 transcripts per million (TPM) could be detected on average over three consecutive time points (i.e. over the developmental time of 1 h) of a high-resolution profile of total RNA (Owens et al., 2016) from fertilization to after gastrulation (stage 13). Genes were considered active when RNAPII enrichments along their full length (see thresholds below) and corresponding transcripts (≥0.1 TPM) were simultaneously detected. Transcript levels were calculated over three consecutive time points +/- 1 h from the developmental stage of RNAPII profiling. RNAPII enrichment covered ≥80% of the mappable gene body and reached at least one of the following thresholds: (1) 2.6-fold, (2) 1.8-fold and 1.4-fold at the next or previous stage, (3) 1.4-fold and 1.8-fold at the next or previous stage, or (4) 1.4-fold over three consecutive stages. The heatmap (Figure 1B and S1A) was sorted by the developmental stage (1st) and the overall fold (2nd) of

RNAPII enrichment. Zygotic and maternal contributions to transcriptome (Figure 1G) were based on RNAPII enrichment (see above) and mean transcript levels (≥0.1 TPM) detected between 0 and 1 hpf, respectively.

Peak Calling and Motif Enrichment Analysis

Peak calling and motif enrichment analysis were carried out as previously reported (Gentsch et al., 2018b). Briefly, HOMER v4.8.3 (Heinz et al., 2010) was used to identify the binding sites of Smad1 (Gentsch et al., 2018b), Smad2 (Chiu et al., 2014; Gentsch et al., 2018b; Yoon et al., 2011) and β-catenin (Gentsch et al., 2018b; Nakamura et al., 2016) by virtue of ChIP-enriched read alignments (hereafter called peaks): findpeaks -style factor -minDist 175 -fragLength 175 -inputFragLength 175 -fdr 0.001 -gsize 1.435e9 -F 3 -L 1 -C 0.97. This means that both ChIP and input alignments were extended 3' to 175 bp for the detection of significant (FDR ≤0.1%) peaks being separated by ≥175 bp. The effective size of the *X. tropicalis* genome assembly v7.1 was set to 1.435 billion bp, an estimate obtained from the mappability profile (Gentsch et al., 2018b). These peaks showed equal or higher tag density than the surrounding 10 kb, ≥3-fold more tags than the input and ≥0.97 unique tag positions relative to the expected number of tags. To further eliminate any false positive peaks, we removed any peaks with <0.5 CPM and those falling into blacklisted regions showing equivocal mappability due to genome assembly errors, gaps or simple/tandem repeats. Regions of equivocal mappability were identified by a two-fold lower (poor) or three-fold higher (excessive) read coverage than the average detected in 400-bp windows sliding at 200-bp intervals through normalized ChIP input and DNase-digested naked genomic DNA (Gentsch et al., 2018b). All identified regions ≤800 bp apart were subsequently merged. Gap coordinates were obtained from the Francis Crick mirror site of the UCSC genome browser (http://genomes.crick.ac.uk). Simple repeats were masked with RepeatMasker v4.0.6 (Smit et al.) using the crossmatch search engine v1.090518 and the following settings: RepeatMasker -species "xenopus silurana tropicalis" -s -xsmall. Tandem repeats were masked with Jim Kent's trfBig wrapper script of the Tandem Repeat Finder v4.09 (Benson, 1999) using the following settings: weight for match, 2; weight for mismatch, 7; delta, 7; matching probability, 80; indel probability, 10; minimal alignment score, 50; maximum period size, 2,000; and longest tandem repeat array (-l), 2 [million bp]. The enrichment and occurrence of predetermined DNA binding motifs was calculated using 100 bp centred across the top 2,000 peaks per chromatin feature and developmental stage: findMotifsGenome.pl -size 100 -mknown -nomotif.

Injections and Treatments of Embryos

Microinjections were performed using calibrated needles and embryos equilibrated in 4% (w/v) Ficoll PM-400 (Merck) in 5% MMR. Microinjection needles were generated from borosilicate glass capillaries (Harvard Apparatus, GC120-15) using the micropipette puller Sutter p97. Maximally three nanolitres were injected into the animal hemisphere of de-jellied zygotes using the microinjector Narishige IM-300. Embryos were transferred to fresh 5% MMR (without Ficoll PM-400) once they reached about the mid-blastula stage.

For profiling the nascent transcriptome, embryos were injected with 75 ng 4-thiouridine-5'-triphosphate (4sU) (TriLink BioTechnologies), which is incorporated into newly synthesized transcripts.

Loss-of-functions (LOFs) were generated by treating embryos with small molecule inhibitors and/or injecting them with morpholinos (MOs) or α-amanitin. MOs were designed and produced by Gene Tools (Oregon, USA) to block splicing (MO_{splice}) or translation (MO_{transl}). MO sequences are listed in the Key Resources Table: maternal Pou5f3/Sox3 (mPou5f3/Sox3) LOF, 5 ng *Pou5f3.2* MO_{transl} (Chiu et al., 2014; Gentsch et al., 2018b), 5 ng *Pou5f3.3* MO_{transl} (Chiu et al., 2014; Gentsch et al., 2018b) and 5 ng *Sox3* MO_{transl} (Gentsch et al., 2018b); maternal VegT (mVegT) LOF, 10 ng m*VegT* MO_{transl} (Gentsch et al., 2018b; Rana et al., 2006); canonical Wnt LOF, 5 ng *β-catenin* MO_{transl} (Gentsch et al., 2018b; Heasman et al., 2000); LOF of four zygotic T-box TFs (4x zT LOF), 2.5 ng *t* MO_{splice} (Gentsch et al., 2013), 2.5 ng *t* MO_{transl} (Gentsch et al., 2013), 2.5 ng *t2* MO_{splice} (Gentsch et al., 2013), 2.5 ng *t2* MO_{transl} (Gentsch et al., 2013), 5 ng z*VegT* MO_{transl} (Fukuda et al., 2010; Gentsch et al., 2013) and 5 ng *eomes* MO_{splice} (Fukuda et al., 2010; Gentsch et al., 2013); control MO, 5-20 ng standard control MO according to the dose used for the β-catenin, mVegT and 4x zT LOF experiment; and 30 pg α-amanitin (BioChemica). To block Nodal (Nodal LOF) and BMP (BMP LOF) signaling, embryos were treated with 100 µM SB431542 (Tocris) and/or 10 µM LDN193189 (Selleckchem) from the 8-cell stage onwards. Control embryos were treated accordingly with DMSO, in which these antagonists were dissolved. Transcriptional effects of combinatorial signal LOF were determined at late blastula stage (stage 9^+), while those of all other maternal LOFs were determined over three consecutive time points: the MBT (stage 8+), the late blastula (stage 9^+) and the early gastrula (stage 10^+) stage. The 4x zT LOF was transcriptionally profiled at early, mid and late gastrula stage (stage 10^+, 11^+ and 12^+). The 4x zT LOF comparison has four biological replicates (n=4). All other comparisons entail three biological replicates (n=3).

Extraction of Total RNA

Embryos were homogenized in 800 µl TRIzol reagent (Thermo Fisher Scientific) by vortexing. The homogenate was either snap-frozen in liquid nitrogen and stored at -80ºC or processed immediately. For phase separation, the homogenate together with 0.2x volume of chloroform was transferred to pre-spun 2.0-ml Phase Lock Gel Heavy microcentrifuge tubes (VWR), shaken vigorously for 15 sec, left on the bench for 2 min and spun at ~16,000 g (4ºC) for 5 min. The upper phase was mixed well with one volume of 95-100% ethanol and spun through the columns of the RNA Clean & Concentrator 25 Kit (Zymo Research) at ~12,000 g for 30 sec. Next, the manufacturer's instructions were followed for the recovery of total RNA (>17 nt) with minor modifications. First, the flow-through of the first spin was re-applied to the column. Second, the RNA was treated in-column with 3 U Turbo DNase (Thermo Fisher Scientific). Third, the RNA was eluted twice with 25 µl molecular-grade water. The concentration was determined on the NanoDrop 1000 spectrophotometer or by fluorometry before depleting

ribosomal RNA from total RNA (Profiling the Nascent Transcriptome).

Tagging the Nascent Transcriptome

Thirty 4sU-injected embryos were collected at the MBT and the early-to-mid gastrula stage. Total RNA was extracted as outlined above. The 4sU-tagging was performed according to (Gay et al., 2014) with few minor modifications. The RNA Clean & Concentrator 5 Kit (Zymo Research) was used to purify RNA. Briefly, the Ribo-Zero Gold rRNA Removal Kit (Illumina) was used according to the manufacturer's instructions to deplete ribosomal RNA from ~10 μg total RNA. The RNA was purified and fragmented for 4 min at 95ºC using the NEBNext Magnesium RNA Fragmentation Module (NEB). The RNA was purified again before conjugating HDPD-Biotin (Thermo Fisher Scientific) to 4sU via disulfide bonds for 3 h in the dark. Purified RNA was mixed with Streptavidin beads (Thermo Fisher Scientific) to pull down biotin-tagged RNA. The RNA was eluted off the beads by treating them twice with 100 μl pre-heated (80ºC) 100 mM β-mercaptoethanol (Merck), which breaks the disulfide bond between Biotin and 4sU. Subsequently, the RNA was converted into a deep sequencing library by following the manual instructions (Rev. C, 8/2014) of the ScriptSeq v2 RNA-Seq Library Preparation Kit (Illumina) starting with 4.1.A. (Anneal the cDNA Synthesis Primer) and 4.1.B. (Synthesize cDNA), RNA and ending with part 3.C (Synthesize 3'-Tagged DNA) to 3.G. (Assess Library Quantity and Quality). cDNA was purified using 1.8x SPRI beads. Input and 4sU-enriched cDNA were PCR-amplified with 11 and 15 cycles, respectively. The 4sU RNA-Seq library was purified with 1x SPRI beads.

Post-Sequencing Analysis of 4sU Tagging

Paired-end reads were aligned to the *X. tropicalis* transcriptome assembly v7.1 running Bowtie2 (Langmead and Salzberg, 2012) with the following constraints: -k 200 (maximal allowed number of alignments per fragment) -X 800 (maximum fragment length in bp) --rdg 6,5 (penalty for read gaps of length N, 6+N*5) --rfg 6,5 (penalty for reference gaps of length N, 6+N*5) --score-main L,-.6,-.4 (minimal alignment score as a linear function of the read length x, f(x) = -0.6 - 0.4*x) --no-discordant (no paired-end read alignments breaching maximum fragment length X) --no-mixed (only concordant alignment of paired-end reads). Only read pairs that uniquely align to one gene were counted. Raw read counts were normalized with DESeq2 v1.14.1 (Love et al., 2014) and then scaled to the input.

Poly(A) RNA-Seq Profiling

10-15 embryos were collected per stage and condition. Total RNA was extracted as outlined above. Libraries were made from ~1 μg total RNA by following the low-sample protocol of the TruSeq RNA Library Prep Kit v2 with a few modifications. First, 1 μl cDNA purified after second strand synthesis was quantified on a Qubit fluorometer using high-sensitivity reagents for detecting double-stranded DNA (10 pg/μl to 100 ng/μl). By this stage, the yield was ~10 ng. Second, the number of PCR cycles was reduced to eight to avoid products of over-amplification such as chimera fragments.

Poly(A) RNA-Seq Read Alignment

Paired-end reads were aligned to the *X. tropicalis* genome assembly v7.1 using STAR v2.5.3a (Dobin et al., 2013) with default settings. The alignment was guided by a revised version of the gene models v7.2 (Collart et al., 2014) to improve mapping accuracy across splice junctions. The alignments were sorted by read name using the sort function of Samtools v1.3.1 (Li et al., 2009). Exon and intron counts (-t 'exon;intron') were extracted from unstranded (-s 0) alignment files using VERSE v0.1.5 (Zhu et al., 2016) in featureCounts (default) mode (-z 0). Intron coordinates were adjusted to exclude any overlap with exon annotation. For visualization, genomic BAM files of biological replicates were merged using Samtools and converted to the bigWig format. These genome tracks were normalized to the wigsum of 1 billion excluding any reads with mapping quality <10 using the python script bam2wig.py from RSeQC v2.6.4 (Wang et al., 2012).

Differential Gene Expression Analysis

Differential expression analysis was performed with both raw exon and intron counts excluding those belonging to ribosomal and mitochondrial RNA using the Bioconductor/R package DESeq2 v1.14.1 (Love et al., 2014). In an effort to find genes with consistent fold changes over time, p-values were generated according to a likelihood ratio test reflecting the probability of rejecting the reduced (~ developmental stage) over the full (~ developmental stage + condition) model. Resulting p-values were adjusted to obtain false discovery rates (FDR) according to the Benjamini-Hochburg procedure with thresholds on Cook's distances and independent filtering being switched off. Equally, combinatorial LOF profiling and regional expression datasets (Blitz et al., 2017) without time series were subjected to likelihood ratio tests with reduced (~ 1) and full (~ condition) models for statistical analysis. Fold changes of intronic and exonic transcript levels were calculated for each developmental stage and condition using the mean of DESeq2-normalized read counts from biological replicates. Both intronic and exonic datasets were filtered for ≥10 DESeq2-normalized read counts that were detected at least at one developmental stage in all uninjected or DMSO-treated samples. Gene-specific fold changes were removed at developmental stages that yielded <10 normalized read counts in corresponding control samples. Next, the means of intronic and exonic fold changes were calculated across developmental stages. The whole dataset was confined to 3,318 genes for which at least 50% reductions (FDR ≤10%) in exonic (default) or intronic counts could be detected in α-amanitin-injected embryos. Regional expression was based on exonic read counts by default unless the intronic fold changes were significantly (FDR ≤10%) larger than the exonic fold changes (Table S3). For the hierarchical clustering of relative gene expression (Figure 2C), increased transcript levels were masked and only data points from signal LOFs, mPou5f3/Sox3 LOF and 4x zT LOF embryos were used. Euclidean distance-derived clusters were linked according to Ward's criterion and sorted using the optimal leaf ordering (OLO) algorithm. The synergy factor (SF) between signals x and y (Figures 3 and S3) were calculated as follows: $SF_{xy} = \Delta_{xy} / (\Delta_x + \Delta_y)$. Δ is the relative loss of

gene expression caused by signal depletion. For these calculations, any gene upregulations were neutralised (i.e. set to 1).

Analysis of Enriched Gene Ontology (GO) Terms

Over-represented GO terms were found by applying hypergeometric tests of the Bioconductor/R package GOstats v2.42.0 (Falcon and Gentleman, 2007) on gene lists. The process was also supported by the Bioconductor/R packages GSEABase v1.38.1 (Morgan et al., 2017) and GO.db v3.4.1 (Carlson et al., 2007). The gene universe was associated with GO terms by means of BLAST2GO (Conesa et al., 2005) as previously outlined (Collart et al., 2014; Gentsch et al., 2015).

Generation of Hybridization Probes

Plasmids *X. laevis eomes* pCRII-TOPO (Gentsch et al., 2013) and *X. laevis brachyury* pSP73 (Smith et al., 1991) were linearized by restriction digestion (*Bam*HI and *Bgl*II, respectively) and purified using the QIAquick PCR Purification Kit (Qiagen). The hybridization probes were transcribed from ~1 µg linearized plasmid using 1x digoxigenin-11-UTP (Roche), 40 U RiboLock RNase inhibitor (Thermo Fisher Scientific), 1x transcription buffer (Roche) and T7 RNA polymerase (Roche) at 37ºC for 2 h. The probe was treated with 2 U Turbo DNase (Thermo Fisher Scientific) to remove the DNA template and purified by LiCl precipitation. RNA was diluted to 10 ng/µl (10x stock) with hybridization buffer. The hybridization buffer (stored at -20ºC) consists of 50% formamide (Fisher Scientific), 5x saline sodium citrate (SSC), 1x Denhardt's solution (Thermo Fisher Scientific) , 10 mM EDTA, 1 mg/ml torula RNA (Merck), 100 µg/ml heparin (Merck), 0.1% (v/v) Tween-20 (Merck) and 0.1% (w/v) CHAPS (Merck).

Whole-Mount In Situ Hybridization (WMISH)

WMISH was conducted using digoxigenin-labeled RNA probes (Monsoro-Burq, 2007; Sive et al., 2000). Briefly, *X. tropicalis* embryos were fixed in MEMFA (100 mM MOPS pH 7.4, 2 mM EDTA, 1 mM $MgSO_4$ and 3.7% formaldehyde) at room temperature for 1 h. The embryos were then washed once in 1x PBS and two to three times in ethanol. Fixed and dehydrated embryos were kept at -20ºC for ≥24 h to ensure proper dehydration before starting hybridization. Dehydrated embryos were washed once more in ethanol before rehydrating them in two steps to PBT (1x PBS and 0.1% (v/v) Tween-20). Embryos were treated with 5 µg/ml proteinase K (Thermo Fisher Scientific) in PBT for 6-8 min, washed briefly in PBT, fixed again in MEMFA for 20 min and washed three times in PBT. Embryos were transferred into baskets, which were kept in an 8x8 microcentrifuge tube holder sitting inside a 10x10 slot plastic box filled with PBT. Baskets were built by replacing the round bottom of 2-ml microcentrifuge tubes with a Sefar Nitex mesh. This container system was used to readily process several batches of embryos at once. These baskets were maximally loaded with 40 to 50 *X. tropicalis* embryos. The microcentrifuge tube holder was used to transfer all baskets at once and to submerge embryos into subsequent buffers of the WMISH protocol. Next, the embryos were incubated in 500 µl hybridization buffer (see recipe above) for 2 h in a

hybridization oven set to 60ºC. After this pre-hybridization step, the embryos were transferred into 500 µl digoxigenin-labeled probe (1 ng/µl) preheated to 60ºC and further incubated overnight at 60ºC. The pre-hybridization buffer was kept at 60ºC. The next day embryos were transferred back into the pre-hybridization buffer and incubated at 60ºC for 10 min. Subsequently, they were washed three times in 2x SSC/0.1% Tween-20 at 60ºC for 10 min, twice in 0.2x SSC/0.1% Tween-20 at 60ºC for 20 min and twice in 1x maleic acid buffer (MAB) at room temperature for 5 min. Next, the embryos were treated with blocking solution (2% Blocking Reagent (Merck) in 1x MAB) at room temperature for 30 min, and incubated in antibody solution (10% lamb serum (Thermo Fisher Scientific), 2% Blocking Reagent (Merck), 1x MAB and 1:2,000 Fab fragments from polyclonal anti-digoxigenin antibodies conjugated to alkaline phosphatase) at room temperature for 4 h. The embryos were then washed four times in 1x MAB for 10 min before leaving them in 1x MAB overnight at 4ºC.

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

Processing of External Datasets

High-time (30-min intervals) resolution of total and poly(A) RNA-Seq (GSE65785) was processed as reported in the original publication (Owens et al., 2016). In addition, intron read counts were corrected by spike RNA-derived normalization factors. For visualization normalized exon and intron counts were scaled to the maximal count detected across the time course and fitted using cubic smoothing splines from 0 to 23.5 hpf: smooth.spline(1:48, x, spar=0.6). Other ChIP-Seq (see Key Resources Table) were processed as described in detail above except for H3K4me3 and H3K36me3 whose enriched regions were detected as follows: findPeaks -style histone -fragLength 175 -inputFragLength 175 -fdr 0.001 -gsize 1.435e9 -F 2 -C 1 -region -size 350 -minDist 500. Thus, we detected significant regions of

histone modifications (-style histone) of at least the lengths of two DNA fragments (-size 350) and being separated by at least 500 bp from each other.

Generation of Plots and Heatmaps

Genomic snapshots were generated with the IGV genome browser v2.4-rc6 (Robinson et al., 2011). All plots and heatmaps were generated using R v3.5.1 (http://cran.r-project.org/). The following add-on R and Bioconductor packages were used for sorting and graphical visualization of data: alluvial v0.1-2, beeswarm v0.2.3, circlize v0.4.5 (Gu et al., 2014), complexHeatmap v1.20.0 (Gu *et al.*, 2016a), dplyr v0.7.8, ggplot2 v3.1.0 (Wickham, 2016), gplots v3.0.1, HilbertCurve v1.12.0 (Gu et al., 2016b), limma v3.38.2 (Ritchie et al., 2015) and seriation v1.2-3 (Hahsler et al., 2008).

QUANTIFICATION AND STATISTICAL ANALYSIS

No statistical method was used for determining sample size. We followed the literature to select the appropriate sample size. The experiments were not randomized. Due to the nature of experiments, the authors were not blinded to group allocation during data collection and analysis. Only viable embryos were included in the analysis. Frequencies of shown morphological phenotypes and WMISH patterns are included in every image. The significance of over-represented GO terms was based on hypergeometric tests. Significances of non-normally distributed data points (gene features) across ZGA were calculated using paired Wilcoxon rank-sum tests (alternative hypothesis 'less'). The effect size (r_{effect}) was estimated from the standard normal deviate of the Wilcoxon p-value (p) as previously described (Rosenthal, 1991), $r_{effect}=Z/sqrt(N)$, where $Z=qnorm(1-p/2)$ is the standardized Z-score and N is the number of observations. The statistical significance of differential RNA-Seq was corrected for multiple comparison according to the Benjamini-Hochberg procedure.

DATA AND SOFTWARE AVAILABILITY

Sequencing reads (FASTQ files) and raw RNA-Seq read counts reported in this paper are available in the GEO database (www.ncbi.nlm.nih.gov/geo) under the accession number GEO: GSE113186 and GSE122551. All analyses were performed in R v3.5.1 (Bioconductor v3.8), Perl v5.18.2 and Python v2.7.12 as detailed above. The R code, genome annotation, intermediate datasets and graphs are available at https://github.com/gegentsch/SpatioTemporalControlZGA.

Supplemental Information

Spatio-Temporal Dynamics of Zygotic Genome Activation in Early *Xenopus* Embryos

George E. Gentsch, Nick D. L. Owens, and James C. Smith
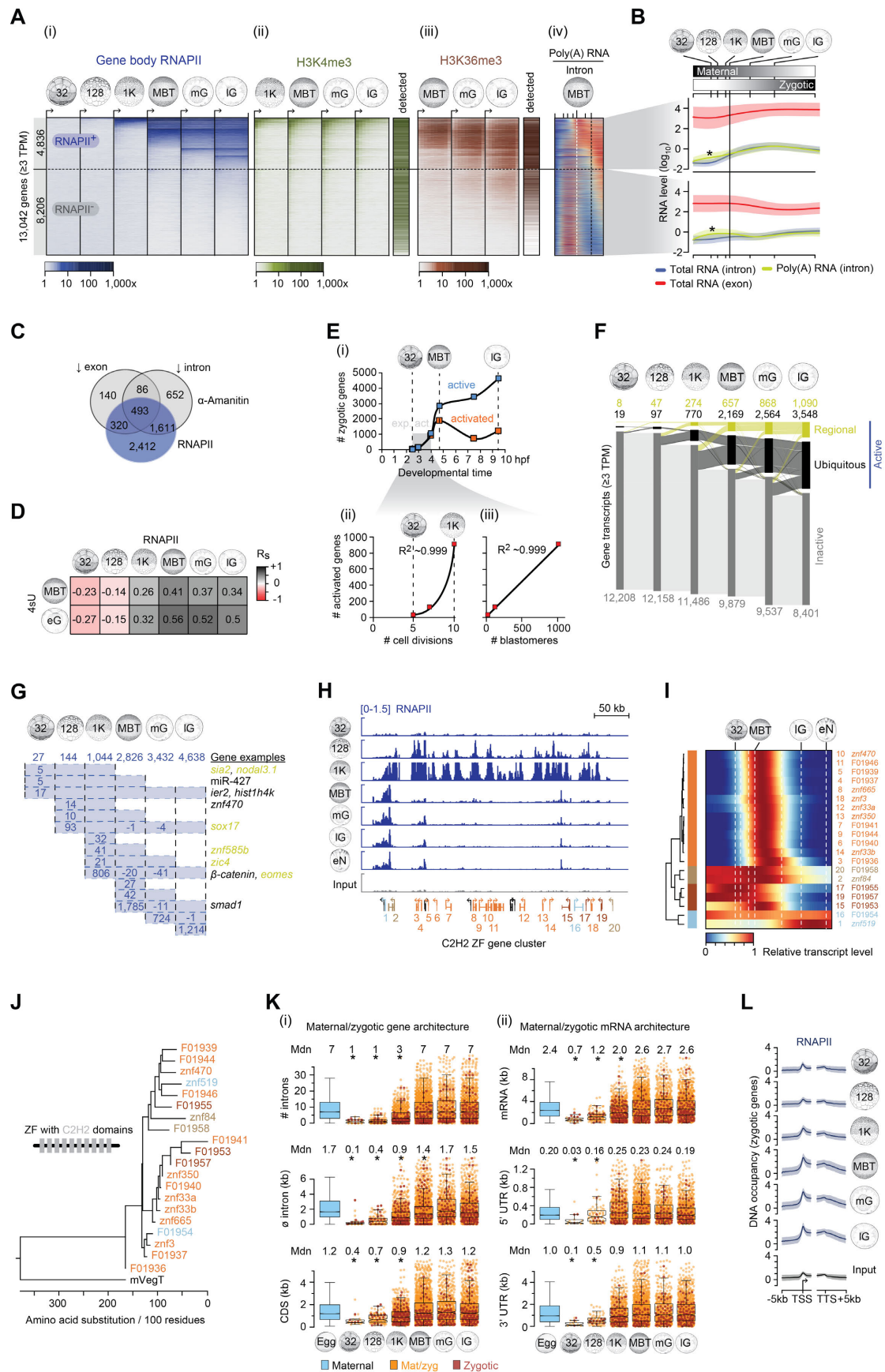
Supplemental Figures

Figure S1. Dynamics and Architecture of ZGA in *X. tropicalis*, Related to Figure 1.

(A) Progression of zygotically activated genes (≥3 TPM) from the 32-cell to the late gastrula stage based on (i) full-length RNAPII occupancy. Co-aligned: Active histone marks H3K4me1 (ii) and H3K36me3 (iii) (Hontelez et al., 2015) and intronic signal from high time-resolution of poly(A) RNA (iv) (Owens et al., 2016). Horizontal dotted line separates RNAPII-engaged (RNAPII$^+$) from non-engaged (RNAPII$^-$) genes as detected until the late gastrula stage. Vertical dotted line indicates the developmental time points of the MBT (white) and the late gastrula stage (black).

(B) Transcript feature levels (mean +/- SD) during the maternal-to-zygotic transition. Asterisk, polyadenylation immediately after fertilization (Collart et al., 2014) transiently increased the intron signal obtained from poly(A) RNA preparations.

(C) Venn diagram of zygotic genes detected by full-length RNAPII occupancy or reduced exonic or intronic transcript counts upon blocking RNAPII-mediated transcription with α-amanitin.

(D) Pairwise spearman correlations (R_s) of enrichment values resulting from RNAPII profiling and 4sU tagging to detect zygotic genes at indicated developmental stages.

(E) Plots of the number of zygotic (active and newly activated) genes (i) or newly activated genes versus the developmental time (i), the number of completed cell divisions (ii) or formed blastomeres (iii).

(F) Alluvial diagram of spatio-temporal ZGA including maternally inherited RNA (≥3 TPM) of genes not activated by the late gastrula stage. Tissue-specificity inferred from regional transcript enrichments along the animal-vegetal or the dorso-ventral or both axes (Blitz et al., 2017).

(G) Numbers of genes with full-length RNAPII occupancy and ≥3 TPM at indicated developmental stages. Examples of ubiquitously (black) and tissue-specifically (orange) expressed genes are listed to the right.

(H) RNAPII dynamics at the C2H2 zinc finger (ZF) cluster from the 32-cell to the early neurula stage.

(I) Expression dynamics of the C2H2 ZF cluster (normalized to maximal transcript level recorded between fertilization and 23.5 hpf, (Owens et al., 2016)).

(J) Phylogenetic tree of the clustered genes (H,I) all of which contain Cys2-His2 [C2H2] domains (10 on average). Maternal VegT (mVegT), 'outgroup' of this phylogenetic tree.

(K) Box and beeswarm plots showing various metrics of the zygotic/maternal genes (i) and mRNA (ii) during ZGA. Wilcoxon rank-sum test against maternal and post-MBT activated genes and corresponding effect sizes (r_{effect}). Bottom value of statistically significant (p) r_{effect} were consistently determined for the comparison between '32' and 'Egg'. Top r_{effect} related to '32' or '128' vs 'mG' as indicated: # introns, *, p <2.1e-13, r_{effect} 0.07-0.43 ('128' vs 'mG'); Ø intron (kb), *, p <7.9e-7, r_{effect} 0.07-0.42 ('128' vs 'mG'); CDS (kb), *, p <3e-7, r_{effect} 0.05-0.27 ('128' vs 'mG'); mRNA (kb), *, p <7e-11, r_{effect} 0.06-0.32 ('128' vs 'mG'); 5' UTR (kb), *, p <0.015, r_{effect} 0.03-0.15 ('32' vs 'mG'); and 3' UTR (kb), *, p <1.4e-6, r_{effect} 0.04-0.21 ('128' vs 'mG').

30

(L) Metaprofile (mean +/- SD) for RNAPII and input (negative control) at zygotic genes.

Abbreviations: 32, 32-cell; 128, 128-cell; 1K, 1,024-cell; MBT, mid-blastula transition; mG, mid-gastrula; lG, late gastrula; eN, early neurula; 4sU, 4-thiouridine; Mdn, median; TPM, transcripts per million.

31
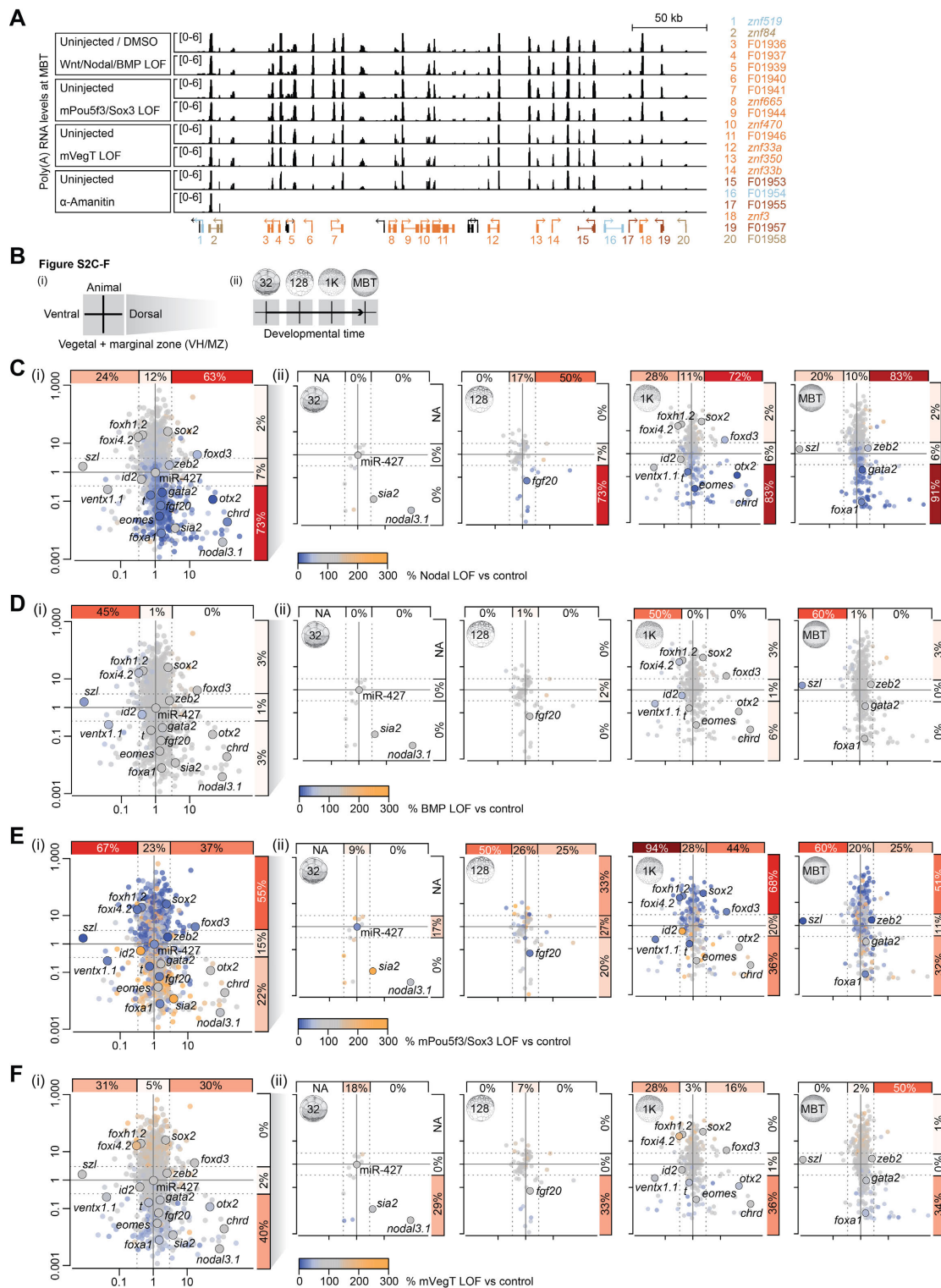
Figure S2. Effect of Canonical Wnt, Nodal and BMP Signals on ZGA, Related to Figure 2.

(A) Poly(A) RNA profiles of the C2H2 ZF cluster (Figure S1H) for indicated control and LOFs at the MBT.

(B) Graphical explanations of figure panels (C) to (F).

(C-F) Summary (i) and temporal resolution (ii) of gene misregulations upon the LOF of Nodal (C) or BMP (D) signaling or maternal Pou5f3/Sox3 (mPou5f3/Sox3) (E) or VegT (mVegT) (F).

32

Percentages only refer to the down-regulated genes (by ≥1/3 compared to control expression level) among all zygotic genes with the same range of expression ratios along the animal-vegetal or the dorso-ventral axes.
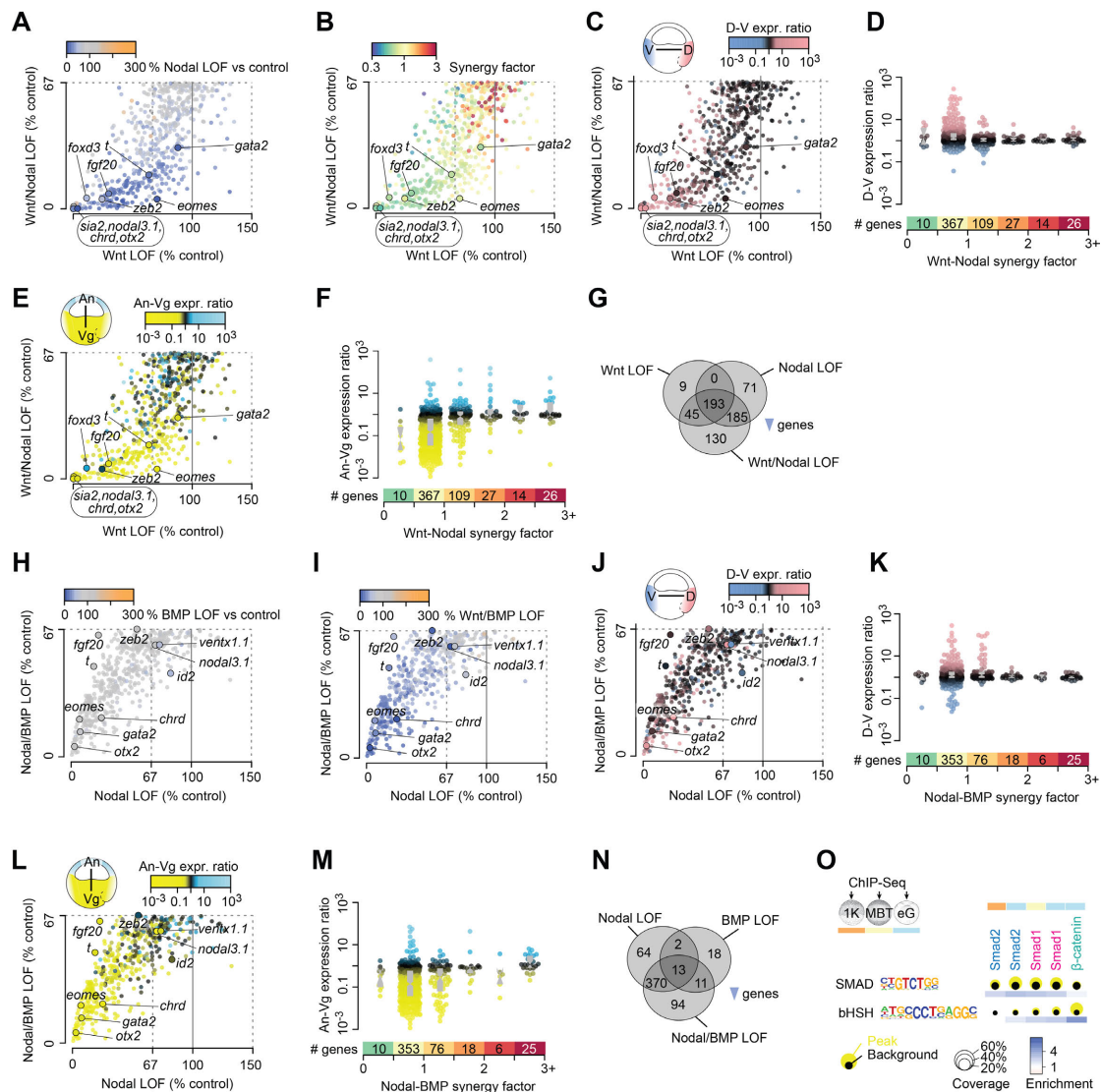
Figure S3. Relationship between Canonical Wnt, Nodal and BMP to Control Regional ZGA, Related to Figure 3.

(A-C,E,H-J,L) Scatter plots of relative (% to control) transcript levels between indicated LOFs with each dot (gene) color-coded according to a third attribute: (A,H,I) relative (% to control) transcript levels, (B) synergy factors between single inductive signals, and (C,E,J,L) regional expression ratios between opposite ends of the indicated axis.

(G,N) Venn diagram of down-regulated genes by indicated LOFs.

(D,F,K,M) Beeswarm plot of regional expression (as measured along the indicated axes) depending on increased Wnt-BMP synergy.

(O) Coverage and enrichment of Smad and β-catenin-associated DNA motifs (SMAD and bHSH motifs) at endogenous binding sites of β-catenin, Smad1, Smad2 (Gentsch et al., 2018b) at indicated developmental stages (color-coded).

34

Table S1. Summary of Deep sequencing and Read Alignments. Related to Figure 1 and 2.

Contains the metadata of ChIP-Seq, 4sU-Seq and RNA-Seq runs including the developmental stage, condition, read type, and total and genome-aligned read numbers.

Table S2. Temporal Progression of ZGA. Related to Figure 1.

Activated genes are listed according to their developmental stage of first detected RNAPII elongation. The lists also include the overall enrichment of RNAPII across the gene body, the maternal contribution (TPM) (Owens et al., 2016) and the enrichment of de novo synthesized transcripts determined by 4sU tagging at the MBT and the early gastrula stage.

Table S3. Differential ZGA Analysis. Related to Figure 2, 3 and 4.

Differential expression analysis of genes showing ≥50% (FDR ≤10%) reduced transcript levels in α-amanitin-injected embryos. Normalised transcript levels (inferred from exon or intron counts) are scaled (percentage, %) to the expression level in control embryos. The list also includes FDRs, expression ratios across the animal-vegetal or dorso-ventral axis (Blitz et al., 2017), the earliest developmental stage of RNAPII elongation across the full gene body, and average expression levels between 0 and 1 hpf (maternal) and 5 and 9 hpf (from the MBT to the mid-gastrula stage).

Movie S1. Quadruple LOF of Zygotic T-box TFs. Related to Figure 2.

Shows simultaneous filming of the vegetal (top row) and animal (bottom row) hemisphere of 4x zT LOF (labelled as T-box KD in movie) (left) and control (right) embryos from early gastrula to mid-tailbud stage (12 h 50 min).

REFERENCES

Anderson, G.A., Gelens, L., Baker, J.C., and Ferrell, J.E. (2017). Desynchronizing Embryonic Cell Division Waves Reveals the Robustness of Xenopus laevis Development. Cell. Rep. *21*, 37–46.

Arnold, S.J., and Robertson, E.J. (2009). Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. Nat. Rev. Mol. Cell Biol. *10*, 91–103.

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. *27*, 573–580.

Blitz, I.L., Paraiso, K.D., Patrushev, I., Chiu, W.T.Y., Cho, K.W.Y., and Gilchrist, M.J. (2017). A catalog of Xenopus tropicalis transcription factors and their regional expression in the early gastrula stage embryo. Dev. Biol. *426*, 409–417.

Blythe, S.A., and Wieschaus, E.F. (2015). Zygotic genome activation triggers the DNA replication checkpoint at the midblastula transition. Cell *160*, 1169–1181.

Braude, P., Bolton, V., and Moore, S. (1988). Human gene expression first occurs between the four- and eight-cell stages of preimplantation development. Nature *332*, 459–461.

Carlson, M., Falcon, S., Pages, H., and Li, N. (2007). Bioconductor - GO.db.

Chan, S.H., Tang, Y., Miao, L., Darwich-Codore, H., Vejnar, C.E., Beaudoin, J.-D., Musaev, D., Fernandez, J.P., Moreno-Mateos, M.A., and Giraldez, A.J. (2018). Brd4 and P300 regulate zygotic genome activation through histone acetylation. bioRxiv 369231.

Chiu, W.T., Le, R.C., Blitz, I.L., Fish, M.B., Li, Y., Biesinger, J., Xie, X., and Cho, K.W.Y. (2014). Genome-wide view of TGFβ/Foxh1 regulation of the early mesendoderm program. Development *141*, 1–114.

Cho, W.-K., Spille, J.-H., Hecht, M., Lee, C., Li, C., Grube, V., and Cisse, I.I. (2018). Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. Science *361*, 412–415.

Christian, J.L., and Moon, R.T. (1993). Interactions between Xwnt-8 and Spemann organizer signaling pathways generate dorsoventral pattern in the embryonic mesoderm of Xenopus. Genes Dev. *7*, 13–28.

Collart, C., Allen, G.E., Bradshaw, C.R., Smith, J.C., and Zegerman, P. (2013). Titration of four replication factors is essential for the Xenopus laevis midblastula transition. Science *341*, 893–896.

Collart, C., Owens, N.D.L., Bhaw-Rosun, L., Cooper, B., De Domenico, E., Patrushev, I., Sesay, A.K., Smith, J.N., Smith, J.C., and Gilchrist, M.J. (2014). High-resolution analysis of gene activity during the Xenopus mid-blastula transition. Development *141*, 1927–1939.

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics *21*, 3674–3676.

Cuny, G.D., Yu, P.B., Laha, J.K., Xing, X., Liu, J.-F., Lai, C.S., Deng, D.Y., Sachidanandan, C., Bloch, K.D., and Peterson, R.T. (2008). Structure-activity relationship study of bone morphogenetic protein (BMP) signaling inhibitors. Bioorg. Med. Chem. Lett. *18*, 4388–4392.

De Iaco, A., Planet, E., Coluccio, A., Verp, S., Duc, J., and Trono, D. (2017). DUX-family transcription factors regulate zygotic genome activation in placental mammals. Nat. Genet. *49*, 941–945.

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

Falcon, S., and Gentleman, R. (2007). Using GOstats to test gene lists for GO term association. Bioinformatics *23*, 257–258.

Faure, S., Lee, M.A., Keller, T., Dijke, ten, P., and Whitman, M. (2000). Endogenous patterns of TGFbeta superfamily signaling during early Xenopus development. Development *127*, 2917–2931.

Fuentealba, L.C., Eivers, E., Ikeda, A., Hurtado, C., Kuroda, H., Pera, E.M., and De Robertis, E.M. (2007). Integrating patterning signals: Wnt/GSK3 regulates the duration of the BMP/Smad1 signal. Cell *131*, 980–993.

Fukuda, M., Takahashi, S., Haramoto, Y., Onuma, Y., Kim, Y.-J., Yeo, C.-Y., Ishiura, S., and Asashima, M. (2010). Zygotic VegT is required for Xenopus paraxial mesoderm formation and is regulated by Nodal signaling and Eomesodermin. Int. J. Dev. Biol. *54*, 81–92.

Gay, L., Karfilis, K.V., Miller, M.R., Doe, C.Q., and Stankunas, K. (2014). Applying thiouracil tagging to mouse transcriptome analysis. Nat. Protoc. *9*, 410–420.

Gentsch, G.E., and Smith, J.C. (2017). Efficient Preparation of High-Complexity ChIP-Seq Profiles from Early Xenopus Embryos. Methods Mol. Biol. *1507*, 23–42.

Gentsch, G.E., Owens, N.D.L., Martin, S.R., Piccinelli, P., Faial, T., Trotter, M.W.B., Gilchrist, M.J., and Smith, J.C. (2013). In vivo T-box transcription factor profiling reveals joint regulation of embryonic neuromesodermal bipotency. Cell. Rep. *4*, 1185–1196.

Gentsch, G.E., Patrushev, I., and Smith, J.C. (2015). Genome-wide snapshot of chromatin regulators and states in Xenopus embryos by ChIP-Seq. J. Vis. Exp. e52535–e52535.

Gentsch, G.E., Spruce, T., Monteiro, R.S., Owens, N.D.L., Martin, S.R., and Smith, J.C. (2018a). Innate Immune Response and Off-Target Mis-splicing Are Common Morpholino-Induced Side Effects in Xenopus. Dev. Cell *44*, 597–610.e10.

Gentsch, G.E., Spruce, T., Owens, N.D.L., and Smith, J.C. (2018b). The role of maternal pioneer factors in predefining first zygotic responses to inductive signals. bioRxiv 306803.

Gu, Z. Eils, R., and Schlesner, M. (2016a). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics *32*, 2847-2849.

Gu, Z., Eils, R., and Schlesner, M. (2016b). HilbertCurve: an R/Bioconductor package for high-resolution visualization of genomic data. Bioinformatics *32*, 2372–2374.

Gu, Z., Gu, L., Eils, R., Schlesner, M., and Brors, B. (2014). circlize Implements and enhances circular visualization in R. Bioinformatics *30*, 2811–2812.

Hahsler, M., Hornik, K., and Buchta, C. (2008). Getting Things in Order: An Introduction to the R Package seriation. J. Stat. Softw. *25*.

Halley-Stott, R.P., Jullien, J., Pasque, V., and Gurdon, J. (2014). Mitosis gives a brief window of opportunity for a change in gene transcription. PLoS Biol. *12*, e1001914.

Hamatani, T., Carter, M.G., Sharov, A.A., and Ko, M.S.H. (2004). Dynamics of global gene expression changes during mouse preimplantation development. Dev. Cell *6*, 117–131.

Harvey, S.A., Sealy, I., Kettleborough, R., Fenyes, F., White, R., Stemple, D., and Smith, J.C. (2013). Identification of the zebrafish maternal and paternal transcriptomes. Development

140, 2703–2710.

Harvey, S.A., Tümpel, S., Dubrulle, J., Schier, A.F., and Smith, J.C. (2010). no tail integrates two modes of mesoderm induction. Development *137*, 1127–1135.

Heasman, J., Kofron, M., and Wylie, C. (2000). βCatenin Signaling Activity Dissected in the Early Xenopus Embryo: A Novel Antisense Approach. Dev. Biol. *222*, 124–134.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Heyn, P., Kircher, M., Dahl, A., Kelso, J., Tomancak, P., Kalinka, A.T., and Neugebauer, K.M. (2014). The earliest transcribed zygotic genes are short, newly evolved, and different across species. Cell Rep. *6*, 285–292.

Ho, D.M., Chan, J., Bayliss, P., and Whitman, M. (2006). Inhibitor-resistant type I receptors reveal specific requirements for TGF-beta signaling in vivo. Dev. Biol. *295*, 730–742.

Hontelez, S., van Kruijsbergen, I., Georgiou, G., van Heeringen, S.J., Bogdanović, O., Lister, R., and Veenstra, G.J.C. (2015). Embryonic transcription is controlled by maternally defined chromatin state. Nat. Commun. *6*, 10148.

Hug, C.B., Grimaldi, A.G., Kruse, K., and Vaquerizas, J.M. (2017). Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. Cell *169*, 216–228.e219.

Inman, G.J., Nicolás, F.J., Callahan, J.F., Harling, J.D., Gaster, L.M., Reith, A.D., Laping, N.J., and Hill, C.S. (2002). SB-431542 is a potent and specific inhibitor of transforming growth factor-beta superfamily type I activin receptor-like kinase (ALK) receptors ALK4, ALK5, and ALK7. Mol. Pharmacol. *62*, 65–74.

Joseph, S.R., Pálfy, M., Hilbert, L., Kumar, M., Karschau, J., Zaburdaev, V., Shevchenko, A., and Vastenhouw, N.L. (2017). Competition between histone and transcription factor binding regulates the onset of transcription in zebrafish embryos. Elife *6*.

Kaaij, L.J.T., van der Weide, R.H., Ketting, R.F., and de Wit, E. (2018). Systemic Loss and Gain of Chromatin Architecture throughout Zebrafish Development. Cell Rep. *24*, 1–10.e4.

Kane, D.A., and Kimmel, C.B. (1993). The zebrafish midblastula transition. Development *119*, 447–456.

Ke, Y., Xu, Y., Chen, X., Feng, S., Liu, Z., Sun, Y., Yao, X., Li, F., Zhu, W., Gao, L., et al. (2017). 3D Chromatin Structures of Mature Gametes and Structural Reprogramming during Mammalian Embryogenesis. Cell *170*, 367–381.e20.

Khokha, M., Chung, C., Bustamante, E., Gaw, L., Trott, K., Yeh, J., Lim, N., Lin, J., Taverner, N., Amaya, E., et al. (2002). Techniques and probes for the study of Xenopus tropicalis development. Dev. Dyn. *225*, 499–510.

Kimelman, D. (2006). Mesoderm induction: from caps to chips. Nat. Rev. Genet. *7*, 360–372.

Langley, A.R., Smith, J.C., Stemple, D.L., and Harvey, S.A. (2014). New insights into the maternal to zygotic transition. Development *141*, 3834–3841.

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

Larabell, C.A., Torres, M., Rowning, B.A., Yost, C., Miller, J.R., Wu, M., Kimelman, D., and Moon, R.T. (1997). Establishment of the dorso-ventral axis in Xenopus embryos is presaged by early asymmetries in beta-catenin that are modulated by the Wnt signaling pathway. J. Cell Biol. *136*, 1123–1136.

Lawrence M., Gentleman R., and Carey V. (2009). rtracklayer: an R package for interfacing with genome browsers. Bioinformatics *25*, 1841-1842.

Lawrence M., Huber W., Pagès H., Aboyoun P., Carlson M., Gentleman R., Morgan M., and Carey V. (2013). Software for Computing and Annotating Genomic Ranges. PLoS Comput. Biol. *9*, e1003118.

Lee, M.T., Bonneau, A.R., Takacs, C.M., Bazzini, A.A., Divito, K.R., Fleming, E.S., and Giraldez, A.J. (2013). Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. Nature *503*, 360–364.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Liang, H.-L., Nien, C.-Y., Liu, H.-Y., Metzstein, M.M., Kirov, N., and Rushlow, C. (2008). The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. Nature *456*, 400–403.

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

Lu, X., Li, J.M., Elemento, O., Tavazoie, S., and Wieschaus, E.F. (2009). Coupling of zygotic transcription to mitotic control at the Drosophila mid-blastula transition. Development *136*, 2101–2110.

Monsoro-Burq, A.H. (2007). A Rapid Protocol for Whole-Mount In Situ Hybridization on Xenopus Embryos. Cold Spring Harb. Protoc. *2007*, pdb.prot4809.

Morgan, M., Falcon, S., and Gentleman, R. (2017). *GSEABase: Gene set enrichment data structures and methods*.

Nakamura, Y., de Paiva Alves, E., Veenstra, G.J.C., and Hoppler, S. (2016). Tissue- and stage-specific Wnt target gene expression is controlled subsequent to β-catenin recruitment to cis-regulatory modules. Development *143*, 1914–1925.

Newport, J., and Kirschner, M.W. (1982a). A major developmental transition in early xenopus embryos: I. characterization and timing of cellular changes at the midblastula stage. Cell *30*, 675–686.

Newport, J., and Kirschner, M.W. (1982b). A major developmental transition in early xenopus embryos: II. control of the onset of transcription. Cell *30*, 687–696.

Nieuwkoop, P., and Faber, J. (1994). Normal table of Xenopus laevis (Daudin): a systematical and chronological survey of the development from the fertilized egg till the end of metamorphosis. New York: Garland.

Nudelman, G., Frasca, A., Kent, B., Sadler, K.C., Sealfon, S.C., Walsh, M.J., and Zaslavsky, E. (2018). High resolution annotation of zebrafish transcriptome using long-read sequencing. Genome Res. *28*, 1415-1425.

Owens, N.D.L., Blitz, I.L., Lane, M.A., Patrushev, I., Overton, J.D., Gilchrist, M.J., Cho, K.W.Y., and Khokha, M.K. (2016). Measuring Absolute RNA Copy Numbers at High Temporal Resolution Reveals Transcriptome Kinetics in Development. Cell Rep. *14*, 632–647.

Pertea, M., Kim, D., Pertea, G.M., Leek, J.T., and Salzberg, S.L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat. Protoc. *11*, 1650–1667.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Rana, A., Collart, C., Gilchrist, M., and Smith, J. (2006). Defining synphenotype groups in Xenopus tropicalis by use of antisense morpholino oligonucleotides. PLoS Genet. *2*, e193.

Reversade, B., Kuroda, H., Lee, H., Mays, A., and De Robertis, E.M. (2005). Depletion of Bmp2, Bmp4, Bmp7 and Spemann organizer signals induces massive brain formation in Xenopus embryos. Development *132*, 3381–3392.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26.

Rosenthal, R. (1991). Meta-Analytic Procedures for Social Research. SAGE Publishing.

Schneider, S., Steinbeisser, H., Warga, R., and Hausen, P. (1996). Beta-catenin translocation into nuclei demarcates the dorsalizing centers in frog and fish embryos. Mech. Dev. *57*, 191–198.

Schohl, A., and Fagotto, F. (2002). Beta-catenin, MAPK and Smad signaling during early Xenopus development. Development *129*, 37–52.

Shermoen, A.W., and O'Farrell, P.H. (1991). Progression of the cell cycle through mitosis leads to abortion of nascent transcripts. Cell *67*, 303–310.

Sive, H., Grainger, R., and Harland, R. (2000). Early development of Xenopus laevis: A laboratory manual. Cold Spring Harbor Laboratory Press.

Skirkanich, J., Luxardi, G., Yang, J., Kodjabachian, L., and Klein, P.S. (2011). An essential role for transcription before the MBT in Xenopus laevis. Dev. Biol. *357*, 478–491.

Smit, A., Hubley, R., and Green, P. RepeatMasker Open-4.0. 2013-2015.

Smith, J., Price, B., Green, J., Weigel, D., and Herrmann, B. (1991). Expression of a Xenopus homolog of Brachyury (T) is an immediate-early response to mesoderm induction. Cell *67*, 79–87.

Stack, J.H., and Newport, J.W. (1997). Developmentally regulated activation of apoptosis early in Xenopus gastrulation results in cyclin A degradation during interphase of the cell cycle. Development *124*, 3185–3195.

Stennard, F., Zorn, A.M., Ryan, K., Garrett, N., and Gurdon, J.B. (1999). Differential expression of VegT and Antipodean protein isoforms in Xenopus. Mech. Dev. *86*, 87–98.

Wang, L., Wang, S., and Li, W. (2012). RSeQC: quality control of RNA-seq experiments. Bioinformatics *28*, 2184–2185.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307–319.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer.

Yang, J., Tan, C., Darken, R.S., Wilson, P.A., and Klein, P.S. (2002). Beta-catenin/Tcf-regulated transcription prior to the midblastula transition. Development *129*, 5743–5752.

Yoon, S.-J., Wills, A.E., Chuong, E., Gupta, R., and Baker, J.C. (2011). HEB and E2A function as SMAD/FOXH1 cofactors. Genes Dev. *25*, 1654–1661.

Young, J.J., Kjolby, R.A.S., Wu, G., Wong, D., Hsu, S.-W., and Harland, R.M. (2017). Noggin is required for first pharyngeal arch differentiation in the frog Xenopus tropicalis. Dev. Biol. *426*, 245–254.

Zhu, Q., Fisher, S.A., Shallcross, J., and Kim, J. (2016). VERSE: a versatile and efficient RNA-Seq read counting tool. bioRxiv 053306.