ARTICLE

doi:10.1038/nature12632

Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition

Miler T. Lee[1]*, Ashley R. Bonneau[1]*, Carter M. Takacs[1], Ariel A. Bazzini[1], Kate R. DiVito[1], Elizabeth S. Fleming[1] & Antonio J. Giraldez[1,2]

After fertilization, maternal factors direct development and trigger zygotic genome activation (ZGA) at the maternal-to-zygotic transition (MZT). In zebrafish, ZGA is required for gastrulation and clearance of maternal messenger RNAs, which is in part regulated by the conserved microRNA miR-430. However, the factors that activate the zygotic program in vertebrates are unknown. Here we show that Nanog, Pou5f1 (also called Oct4) and SoxB1 regulate zygotic gene activation in zebrafish. We identified several hundred genes directly activated by maternal factors, constituting the first wave of zygotic transcription. Ribosome profiling revealed that *nanog*, *sox19b* and *pou5f1* are the most highly translated transcription factors pre-MZT. Combined loss of these factors resulted in developmental arrest before gastrulation and a failure to activate >75% of zygotic genes, including *miR-430*. Our results demonstrate that maternal Nanog, Pou5f1 and SoxB1 are required to initiate the zygotic developmental program and induce clearance of the maternal program by activating miR-430 expression.

In animals, maternal gene products drive early development in a transcriptionally silent embryo, and are responsible for ZGA. ZGA occurs during the MZT, when developmental control transfers to the embryonic nucleus. This universal transition represents a major reprogramming event that requires (1) chromatin remodelling to provide transcriptional competency; (2) specific activation of a new transcriptional program; and (3) clearance of the previous transcriptional program. In *Drosophila*, maternal Zelda is required for activating the first zygotic genes through binding of TAGteam cis elements[1,2]. However, the maternal factors that mediate ZGA in vertebrates remain largely unknown[3,4]. In zebrafish, ZGA coincides with the midblastula transition (MBT) ~3 h post-fertilization (h.p.f.), during which genome competency is established through widespread changes in chromatin[5,6] and DNA methylation[7,8]. Bivalent chromatin marks are associated with zygotic genes thought to be 'poised' for activation[5]. Yet, many loci with active marks seem to be transcriptionally inactive[5], indicating that competent genes require induction by additional factors. ZGA is required for epiboly[9] and the clearance of maternal mRNAs, a process regulated in part by the conserved microRNA (miRNA) miR-430 (refs 10–12). Although significant advances have taken place in understanding how vertebrate embryos acquire transcriptional competency and orchestrate the clearance of the maternal program, the factors that control activation of the specific genes during ZGA remain unknown. Here we combine loss-of-function analyses, high-throughput sequencing and ribosome footprinting to identify factors that activate the first wave of zygotic transcription to initiate nuclear control of embryonic development.

Identifying the first zygotic transcripts

To define factors that mediate transcriptional activation, we first sought to identify the earliest genes transcribed from the zygotic genome. Accurate characterization of the early transcriptome faces two main challenges: (1) zygotic transcription of a gene can be masked by a large maternal contribution; and (2) poly(A)$^+$ selection of mRNAs can lead to apparent increases in gene expression, reflecting delayed polyadenylation of maternal mRNAs rather than transcription. We reasoned that maternal mRNAs are spliced during oogenesis, so examining introns from total RNA would allow us to quantify *de novo* transcription independent of polyadenylation or maternal contribution. We performed Illumina total RNA sequencing on wild-type embryos after the onset of zygotic transcription (4 h.p.f., sphere, and 6 h.p.f., shield) (Fig. 1a) compared to embryos before the MZT (2 h.p.f., 64-cell stage) and α-amanitin-treated embryos (assayed at sphere and shield), which lack zygotic transcription. This analysis identified 608 genes with significant increases in exon or intron expression levels >5 RPKM (reads per kilobase per million reads) at sphere stage ($P < 0.1$, Benjamini–Hochberg multiple test correction) (Fig. 1b, c and Extended Data Fig. 1a–h). Intron signal identifies an additional 6,602 genes with low levels of transcription by 4 h.p.f., and 9,330 transcribed genes by 6 h.p.f., expanding the number of zygotically expressed genes previously identified[13,14] (Extended Data Fig. 1i–o and Supplementary Data 1). Over 74% of these are genes with maternal contributions (maternal and zygotic genes), most of which are only identified by elevated intron signal (Fig. 1b and Extended Data Fig. 1g), reflecting the sensitivity of this method to detect *de novo* transcription.

Next, we examined which genes are directly triggered by the maternal program in the 'first wave' of transcription by 4 h.p.f. versus those activated by zygotic factors. We reasoned that blocking zygotic gene function while leaving maternal factors unaffected would uncouple the first from subsequent waves of zygotic transcription. To this end, we inhibited splicing of zygotic mRNAs using morpholinos complementary to U1 and U2 spliceosomal RNAs (U1U2 MO) (Fig. 1d and Extended Data Fig. 1a–d)[15]. U1U2 MO embryos arrest before epiboly (Fig. 1a), despite remaining transcriptionally active. Illumina sequencing revealed an enrichment in intron–exon boundary reads (Fig. 1e) and activation of a subset of zygotic transcripts to levels >5 RPKM (Methods); these genes constitute the first wave of zygotic transcription (Fig. 1f). To test that these first-wave genes are indeed independent

[1]Department of Genetics, Yale University School of Medicine, New Haven, Connecticut 06510, USA. [2]Yale Stem Cell Center, Yale University School of Medicine, New Haven, Connecticut 06520, USA.
*These authors contributed equally to this work.

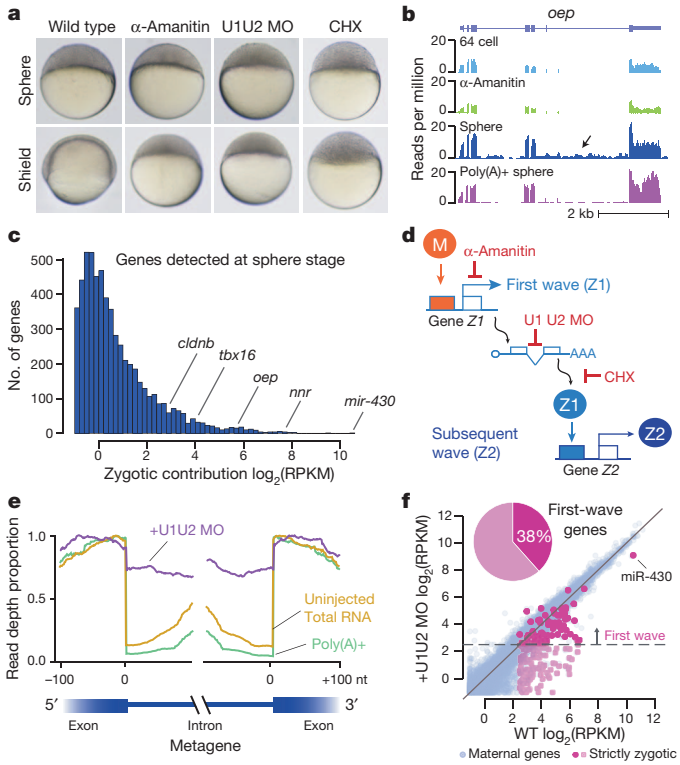


Figure 1 | Characterization of the zygotic transcriptome. **a**, Embryos showing the effects of α-amanitin, U1U2 morpholino (U1U2 MO) and cycloheximide (CHX). **b**, Sequencing read density across *oep*. Intronic signal increases with zygotic expression in total RNA. **c**, Expression histogram of zygotic genes. **d**, Maternal (M) but not zygotic factors (Z1) can activate transcription on splice or translation inhibition. **e**, Metagene of read density across exon–intron boundaries in first-wave genes. U1U2 MO shows enriched intron signal (purple). **f**, Biplot comparing expression in wild type and U1U2 MO. Points above 5 RPKM in U1U2 MO are considered first-wave genes.

of zygotic factors, we treated embryos with cycloheximide (CHX) before MBT (32-cell stage) to block translation of zygotic mRNAs selectively, while allowing translation of maternal mRNAs. CHX-treated embryos also fail to reach epiboly (Fig. 1a) and have a highly correlated transcriptome profile with U1U2 MO (Pearson's $R = 0.97$, Extended Data Fig. 2), confirming first-wave transcription in the absence of zygotic proteins. First-wave genes comprise both embryonic-specific and housekeeping genes ubiquitously expressed in adult tissues (Extended Data Fig. 3a) and are enriched in pattern specification, gastrulation and chromatin modifying functions (Extended Data Fig. 3b). We validated a subset of these genes by RT–PCR, including *klf4b*, *nnr* and *isg15* (Extended Data Fig. 3c–k). Notably, the pri-miR-430 polycistron is highly expressed as part of this first wave (>1,000 RPKM) (Fig. 1c, f). Together, these results identify 269 first-wave genes expressed by sphere stage for which maternal factors are sufficient for activation.

Nanog, SoxB1 and Pou5f1 activate the first wave

Considering the specific, widespread and steep pattern of zygotic gene activation, we proposed that the factors that trigger the first wave may include sequence-specific transcriptional regulators highly translated before ZGA. We analysed the translation levels of all maternal mRNAs using ribosome profiling data (Fig. 2a)[16]. We found that Nanog, Sox19b and Pou5f1 are the most highly translated sequence-specific transcription factors in the pre-MZT transcriptome (Fig. 2b). Pou5f1, the SoxB1 family (which includes Sox2 and Sox19b) and Nanog are key transcription factors involved in maintaining pluripotency in embryonic stem (ES) cells (reviewed in refs 17, 18). In zebrafish, Pou5f1 provides temporal control of gene expression[19] and together with SoxB1 regulates dorsal–ventral patterning and neuronal development[18,20–23], whereas Nanog is essential for endoderm formation through regulation of zygotic *mxtx2* (ref. 24).

To examine the roles of Nanog, Sox19b and Pou5f1 in activating zygotic gene expression, we combined a maternal–zygotic loss-of-function (LOF) Pou5f1 (MZ*pou5f1*)[21] with previously published translation-blocking morpholinos for Nanog (ref. 24) and SoxB1 (ref. 20) (Methods). Because Sox2, Sox3 and Sox19a have been shown to compensate for

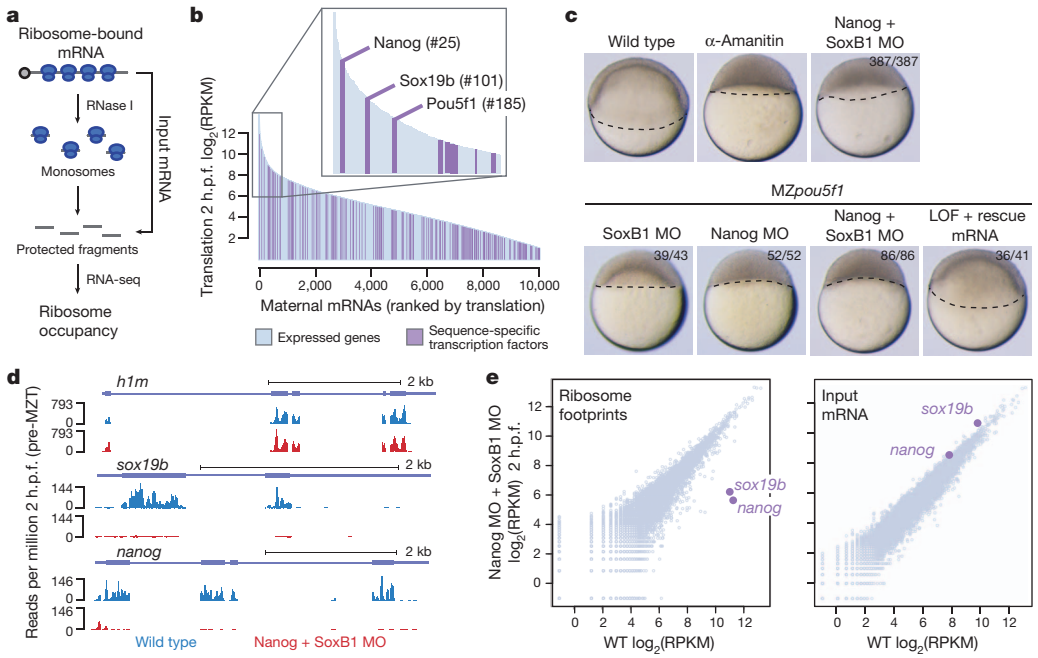


Figure 2 | Identification of Nanog, SoxB1 and Pou5f1 as zygotic gene regulators. **a**, Schematic diagram illustrating ribosome profiling. **b**, Rank plot showing translation levels pre-MZT. Sequence-specific transcription factors are highlighted. **c**, Embryos with combined loss of Nanog plus SoxB1, Nanog plus Pou5f1 or triple LOF arrest similar to α-amanitin and are rescued with *nanog*, *soxB1* and *pou5f1* mRNA injection. **d**, Ribosome footprints for *h1m*, *sox19b* and *nanog* in wild type and Nanog MO plus SoxB1 MO. *sox19b* and *nanog* are highly depleted in the morpholino condition. **e**, Biplots comparing wild-type and morpholino ribosome footprints and input mRNA.

Sox19b loss, we used a combination of morpholinos targeting all four *sox* genes[20] (Extended Data Fig. 4a). Simultaneous Nanog LOF in combination with SoxB1 or Pou5f1 resulted in complete developmental arrest before gastrulation, with >95% of the treated embryos failing to initiate epiboly ($n = 387$ and $n = 52$, respectively) (Fig. 2c and Extended Data Fig. 4b–e). This phenotype resembles that of α-amanitin-injected embryos, indicating that these factors have a role in activating zygotic genes. We used two different approaches to analyse the activity and specificity of these morpholinos. First, we performed ribosome profiling on wild-type and Nanog plus SoxB1 morpholino-injected embryos pre-MBT[16,25]. Translation efficiency for both Nanog and Sox19b was reduced >97% in the morpholino-injected embryos compared to wild type (Fig. 2d and Extended Data Fig. 4f), but was largely unaffected for the rest of the transcriptome (Fig. 2e). Second, we co-injected mRNAs encoding Nanog and SoxB1 with the morpholinos and were able to rescue gastrulation (Fig. 2c and Extended Data Fig. 4c–e). Together, these results show that Nanog, Sox19b and Pou5f1 regulate progression through zygotic development and gastrulation.

Illumina sequencing revealed that combined loss of Nanog, SoxB1 and Pou5f1 results in widespread reduction in first-wave gene expression by 4 h.p.f.: 77% for strictly zygotic genes, 50% for maternal and zygotic genes (Fig. 3a, b and Extended Data Fig. 5). By 6 h.p.f., expression loss is systemic, with 86% of strictly zygotic and 79% of maternal and zygotic genes failing to be expressed to wild-type levels (Fig. 3a, b and Extended Data Fig. 5), an effect that was rescued by injection of the cognate mRNAs (Fig. 3c and Extended Data Figs 5 and 6). Comparing the single and double LOF transcriptomes to the triple, we found that regulation is often combinatorial and redundant, with Nanog LOF having the strongest effect and SoxB1 the weakest (Fig. 3d and Extended Data Fig. 7a–c). By 6 h.p.f., affected genes include housekeeping genes, general transcription factors (for example, *gata6*, *otx1*, *irx1b*, *ntla*) and major signalling components in gastrulation, anterior–posterior axis and dorsal–ventral axis specification (for example, *oep*, *fgf3*, *wnt11*, *chd*, *nog1*, *ndr2*, *bmp2b*) (Extended Data Fig. 7d, e). Together, these results show that Nanog, Pou5f1 and SoxB1 have a fundamental role in activating the first wave, an effect that propagates to subsequent waves resulting in a global impact on zygotic gene expression.

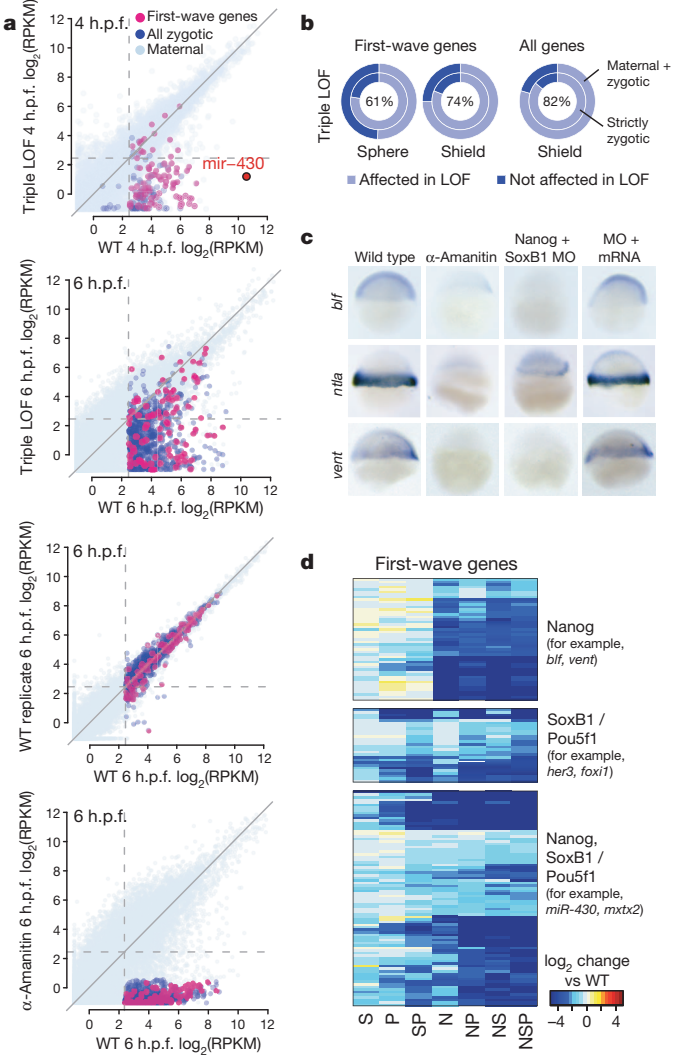


Figure 3 | Transcriptome-wide effects of loss of Nanog, SoxB1 and Pou5f1. **a**, Biplots showing widespread gene expression loss in the triple LOF at 4 and 6 h.p.f. **b**, Plots showing global effects of LOF. Percentages show the combined effect for strictly zygotic and maternal plus zygotic gene groups. **c**, *In situ* hybridization shows expression defects in LOF embryos, which are rescued by mRNA injection. **d**, Heat-map showing first-wave zygotic genes in single and combined LOF conditions. N, Nanog MO; P, MZ*pou5f1*; S, SoxB1 MO. Patterns shown are regulation by Nanog predominantly (top); SoxB1 and Pou5f1 (middle); or Nanog in combination with SoxB1 and Pou5f1 (bottom).

miR-430 is strongly activated by Nanog

Notably, among the first-wave genes co-regulated by Nanog, Pou5f1 and SoxB1 was miR-430, a miRNA family that functions in the clearance of maternal mRNAs in zebrafish and *Xenopus*[10–12]. Northern blot analysis revealed a strong reduction of mature miR-430 levels in Nanog LOF embryos (Fig. 4a). Although individual loss of SoxB1 or Pou5f1 had no detectable effect on miR-430 expression, when combined with Nanog LOF they reduced miR-430 levels even further, a phenotype that was rescued by co-injecting the respective mRNAs (Fig. 4a–c). Nanog morpholino embryos failed to repress a GFP reporter of endogenous miR-430 activity[26], consistent with Nanog's role in activating miR-430 (Extended Data Fig. 8a, b).

To determine whether Nanog specifically binds the *miR-430* genomic locus, we analysed Nanog chromatin immunoprecipitation sequencing (ChIP-seq) data at high (3.3 h.p.f.) and dome stage (4.3 h.p.f.)[24]. Consistent with widespread Nanog regulation, 74% of first-wave genes are bound by Nanog, a significant enrichment compared to subsequent-wave genes (Fig. 4d and Extended Data Fig. 9a). miR-430 is expressed from a 17-kilobase (kb) genomic region on chromosome 4, which includes 55 repeated miR-430 hairpin sequences. Because this locus is repetitive, it had been excluded from previous analyses; however, the sequences are largely unique relative to the rest of the genome. Reads aligning the *miR-430* locus were enriched >16-fold in the Nanog immunoprecipitation compared to whole-cell extract (Fig. 4e), indicating that strong Nanog binding throughout the locus correlates with strong miR-430 expression at ZGA. When the reads were aligned to the presumptive 5′ end of the polycistron, we observed a strong peak of binding in a ~600-nucleotide region between two miR-430 precursors, which contains three canonical Nanog binding sites (CATT[T/G][T/G]CA)[24,27].

To determine whether Nanog induces clearance of maternal mRNAs through activation of miR-430, we analysed the expression of an endogenous miR-430 target, *cd82b* (ref. 10). *cd82b* mRNA is maternally deposited and cleared in wild type by 6 h.p.f. (Fig. 5a). In contrast, *cd82b* mRNA is stabilized in MZ*dicer* mutants or α-amanitin-treated embryos, which lack miR-430 processing and expression, respectively. Similar loss of regulation is observed in Nanog plus SoxB1 MO, as well as triple LOF embryos, a defect that is rescued by providing the cognate mRNAs (Fig. 5b and Extended Data Fig. 8c). To determine the global effect of this regulation, we examined RNA-seq levels of maternal mRNAs containing miR-430 target sites. Loss of Nanog alone or in combination with loss of SoxB1 and MZ*pou5f1* resulted in miR-430 target stabilization, similar to MZ*dicer*[10,16,26] (Fig. 5c and Extended Data Fig. 8d–f) ($P < 1 \times 10^{-51}$, two-sided Wilcoxon rank-sum test). A significant, but weaker, effect was observed in Pou5f1 plus SoxB1 LOF embryos

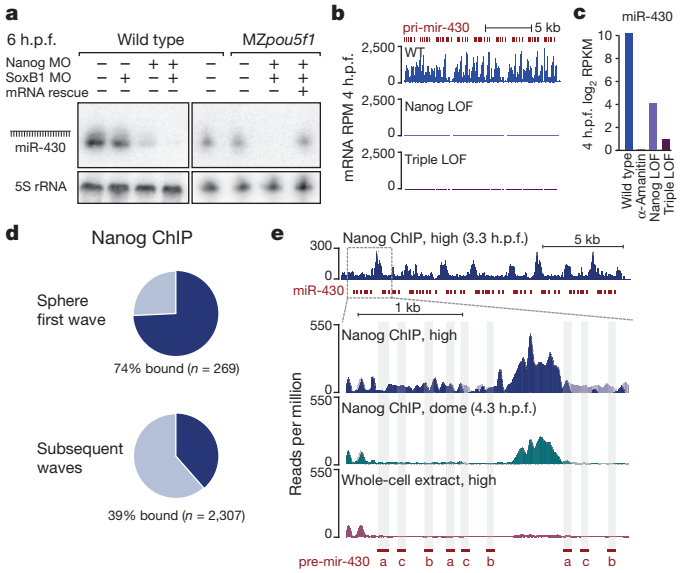


Figure 4 | miR-430 expression is regulated by Nanog. a, Northern blot shows that miR-430 is severely reduced in Nanog LOF and nearly undetectable in the triple LOF embryos. **b**. RNA-seq read levels of the pri-mir-430 polycistron in wild-type and LOF embryos. **c**, Bar plot of total miR-430 aligning reads. **d**, First-wave genes are highly bound by Nanog. **e**, Nanog binding across the miR-430 region (top panel) and a zoomed region where reads are preferentially aligned to the 5′ end (bottom). Binding profiles show a strong peak between two precursors. pre-mir-430a, pre-mir-430b and pre-mir-430c are marked in red.

($P < 1 \times 10^{-25}$) (Extended Data Fig. 8d). These results show that Nanog together with Pou5f1 and SoxB1 activate miR-430 expression, thus revealing a genetic network that links maternal regulation of zygotic gene expression to zygotic clearance of maternal mRNAs.

Discussion

Our transcriptome analysis during the maternal-to-zygotic transition provides three major insights. First, maternal factors directly regulate hundreds of mRNAs that constitute the first wave of zygotic transcription. These targets are activated in the absence of zygotic gene function and are enriched for genes that guide early embryonic development. Transcriptional competence coincides with changes in the chromatin and DNA methylation states of the genome[4–8]. Modifications to the epigenetic landscape during the MZT may be sufficient to allow basal levels of transcription; however, we show here that maternal transcription factors have a vital role in shaping transcriptional output.

Second, we observe that Nanog, SoxB1 and Pou5f1, previously implicated in the maintenance of pluripotency, contribute to widespread activation of zygotic genes during the MZT. These maternal factors enhance transcriptional activation of more than 74% of first-wave zygotic genes, and by 6 h.p.f. influence expression of >80% genes overall. Simultaneous removal of Nanog with SoxB1 and/or Pou5f1 results in complete block of gastrulation and developmental arrest, similar to global inhibition of zygotic gene expression (Fig. 2c and Extended Data Fig. 9c). Nanog binds 74% of first-wave genes during the early stages of ZGA (Fig. 4d). Additionally, while this manuscript was under review, Pou5f1 and Sox2 were also shown to associate with ~40% of early zygotic genes[28]. However, SoxB1 plus Pou5f1 LOF is insufficient to block gastrulation and zygotic development[28] (Fig. 2c). This highlights the central role of Nanog, which together with Pou5f1 and SoxB1 initiates the zygotic program of development, although it is likely that additional factors cooperate with them to provide genome competency and regulate the timing of ZGA[4]. These factors' role in vertebrates may be comparable to Zelda in *Drosophila*, in activating a large cohort of zygotic genes[2]. In mouse, Oct4 and Nanog have been proposed to regulate gene expression at the 2-cell stage[29,30] and along with Sox2 are required for specification of the blastocyst lineages[31–33]. In fact, when we analyse early zygotic genes in mouse, we find that they are enriched for Nanog, Oct4 and Sox2 binding in embryonic stem cells (Extended Data Fig. 9b). Conceptually and mechanistically, many parallels exist between the MZT and the cellular reprogramming that occurs in induced pluripotent stem cells (iPS cells)[3,12]. Indeed, reprogramming of terminally differentiated cells was first shown in the context of the early embryo through nuclear transfer[34,35]. The onset of zygotic development can be viewed as a major reprogramming event that occurs on fusion of two terminally differentiated cells (sperm and oocyte). As shown in ES cells and iPS cells, Pou5f1, Nanog and Sox2 are central players in the induction[36–40] and maintenance[41–43] of pluripotency *in vivo* and *in vitro*[17,35]. In these contexts, part of their role is to serve as 'pioneering' factors, binding to silent chromatin to facilitate *de novo* gene expression[44]. We propose that this pioneering activity is recapitulated during the MZT, where an endogenous function of Nanog, SoxB1 and Pou5f1 is to mediate activation of the first wave of zygotic genes, establishing a transient pluripotent state.

Third, we show that Nanog together with SoxB1 and Pou5f1 directly regulate miR-430, which is responsible for clearance of maternal mRNAs during the MZT[10–12], facilitating the transfer of developmental control to the zygotic program (Extended Data Fig. 9c). Members of the conserved miR-430/295/302/372 family of miRNAs stabilize self-renewal fate in ES cells and enhance reprogramming efficiency[45,46]. We propose that in both cases these miRNAs are 'clearing the slate' by accelerating the removal of mRNAs from the previous program, thus facilitating the establishment of new states by reprogramming factors[12]. The marked upregulation of miR-430 expression by Nanog, SoxB1 and Pou5f1 provides a central link between the mechanisms that drive zygotic gene activation and the clearance of the previous maternal history.

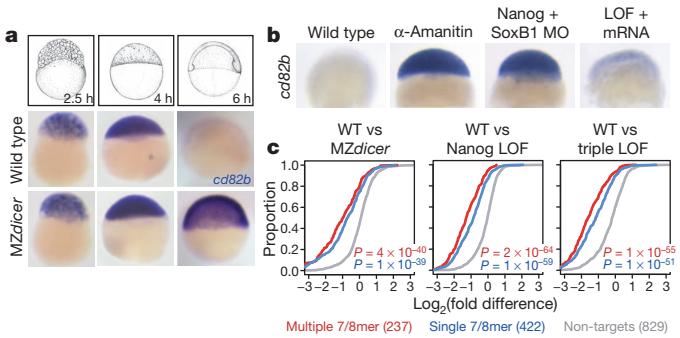


Figure 5 | miR-430 activity is abrogated by Nanog LOF. a, *In situ* hybridization showing degradation of miR-430 target *cd82b* at 6 h.p.f. in wild type, compared to stabilization in MZ*dicer* (lacking miR-430 activity). **b**, *cd82b* is stabilized in the Nanog-SoxB1 LOF embryo, indicating loss of miR-430 activity. The effect is rescued with injection of *nanog* and *soxB1* mRNA. **c**, Cumulative plots showing stabilized expression of miR-430 targets in MZ*dicer* and LOF embryos, compared to wild type. *P* values are for two-sided Wilcoxon rank-sum tests comparing each miR-430 target group to non-targets.

METHODS SUMMARY

MZ*pou5f1*$^{hi349Tg/hi349Tg}$ and MZ*dicer*$^{hu896/hu896}$ were generated as previously described[21,26]. All injections were performed at the one-cell stage. For translation inhibition, 32-cell stage embryos were incubated in media with 50 μg ml^{-1} cycloheximide (Sigma Aldrich) at 28 °C until collection. Total RNA libraries were constructed using the TruSeq Stranded and Ribo-Zero Gold kits (Epicentre). Aligned reads were intersected with Ensembl r70 and RefSeq gene exon and intron annotations. Differential expression was performed using DESeq[47]. ChIP-seq data were analysed as described previously[24], except for the *miR-430* locus, for which unique alignments were not required. Ribosome profiling was performed as described in ref. 16, using the Epicentre ARTseq kit. Sequencing samples are summarized in Extended Data Table 1.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 May; accepted 4 September 2013.
Published online 22 September 2013.

1. ten Bosch, J. R., Benavides, J. A. & Cline, T. W. The TAGteam DNA motif controls the timing of *Drosophila* pre-blastoderm transcription. *Development* **133,** 1967–1977 (2006).
2. Liang, H. L. *et al.* The zinc-finger protein Zelda is a key activator of the early zygotic genome in *Drosophila. Nature* **456,** 400–403 (2008).
3. Tadros, W. & Lipshitz, H. D. The maternal-to-zygotic transition: a play in two acts. *Development* **136,** 3033–3042 (2009).
4. Andersen, I. S. *et al.* Epigenetic marking of the zebrafish developmental program. *Curr. Top. Dev. Biol.* **104,** 85–112 (2013).
5. Vastenhouw, N. L. *et al.* Chromatin signature of embryonic pluripotency is established during genome activation. *Nature* **464,** 922–926 (2010).
6. Lindeman, L. C. *et al.* Prepatterning of developmental gene expression by modified histones before zygotic genome activation. *Dev. Cell* **21,** 993–1004 (2011).
7. Potok, M. E., Nix, D. A., Parnell, T. J. & Cairns, B. R. Reprogramming the maternal zebrafish genome after fertilization to match the paternal methylation pattern. *Cell* **153,** 759–772 (2013).
8. Jiang, L. *et al.* Sperm, but not oocyte, DNA methylome is inherited by zebrafish early embryos. *Cell* **153,** 773–784 (2013).
9. Kane, D. A. & Kimmel, C. B. The zebrafish midblastula transition. *Development* **119,** 447–456 (1993).
10. Giraldez, A. J. *et al.* Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* **312,** 75–79 (2006).
11. Lund, E., Liu, M., Hartley, R. S., Sheets, M. D. & Dahlberg, J. E. Deadenylation of maternal mRNAs mediated by miR-427 in *Xenopus laevis* embryos. *RNA* **15,** 2351–2363 (2009).
12. Giraldez, A. J. microRNAs, the cell's Nepenthe: clearing the past during the maternal-to-zygotic transition and cellular reprogramming. *Curr. Opin. Genet. Dev.* **20,** 369–375 (2010).
13. Harvey, S. A. *et al.* Identification of the zebrafish maternal and paternal transcriptomes. *Development* **140,** 2703–2710 (2013).
14. Aanes, H. *et al.* Zebrafish mRNA sequencing deciphers novelties in transcriptome dynamics during maternal to zygotic transition. *Genome Res.* **21,** 1328–1338 (2011).
15. Kaida, D. *et al.* U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* **468,** 664–668 (2010).
16. Bazzini, A. A., Lee, M. T. & Giraldez, A. J. Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. *Science* **336,** 233–237 (2012).
17. Chambers, I. & Tomlinson, S. R. The transcriptional foundation of pluripotency. *Development* **136,** 2311–2322 (2009).
18. Onichtchouk, D. Pou5f1/oct4 in pluripotency control: insights from zebrafish. *Genesis* **50,** 75–85 (2012).
19. Onichtchouk, D. *et al.* Zebrafish Pou5f1-dependent transcriptional networks in temporal control of early development. *Mol. Syst. Biol.* **6,** 354 (2010).
20. Okuda, Y., Ogura, E., Kondoh, H. & Kamachi, Y. B1 SOX coordinate cell specification with patterning and morphogenesis in the early zebrafish embryo. *PLoS Genet.* **6,** e1000936 (2010).
21. Lunde, K., Belting, H. G. & Driever, W. Zebrafish pou5f1/pou2, homolog of mammalian Oct4, functions in the endoderm specification cascade. *Curr. Biol.* **14,** 48–55 (2004).
22. Hauptmann, G. *et al.* spiel ohne grenzen/pou2 is required for zebrafish hindbrain segmentation. *Development* **129,** 1645–1655 (2002).
23. Belting, H. G. *et al.* Pou5f1 contributes to dorsoventral patterning by positive regulation of *vox* and modulation of *fgf8a* expression. *Dev. Biol.* **356,** 323–336 (2011).
24. Xu, C. *et al.* Nanog-like regulates endoderm formation through the Mxtx2-Nodal pathway. *Dev. Cell* **22,** 625–638 (2012).
25. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. & Weissman, J. S. Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science* **324,** 218–223 (2009).
26. Giraldez, A. J. *et al.* MicroRNAs regulate brain morphogenesis in zebrafish. *Science* **308,** 833–838 (2005).
27. Loh, Y. H. *et al.* The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nature Genet.* **38,** 431–440 (2006).
28. Leichsenring, M., Maes, J., Mossner, R., Driever, W. & Onichtchouk, D. Pou5f1 transcription factor controls zygotic gene activation in vertebrates. *Science* **341,** 1005–1009 (2013).
29. Foygel, K. *et al.* A novel and critical role for Oct4 as a regulator of the maternal-embryonic transition. *PLoS ONE* **3,** e4109 (2008).
30. Tan, M. H. *et al.* An Oct4-Sall4-Nanog network controls developmental progression in the pre-implantation mouse embryo. *Mol. Syst. Biol.* **9,** 632 (2013).
31. Keramari, M. *et al.* Sox2 is essential for formation of trophectoderm in the preimplantation embryo. *PLoS ONE* **5,** e13952 (2010).
32. Frum, T. *et al.* Oct4 cell-autonomously promotes primitive endoderm development in the mouse blastocyst. *Dev. Cell* **25,** 610–622 (2013).
33. Messerschmidt, D. M. & Kemler, R. Nanog is required for primitive endoderm formation through a non-cell autonomous mechanism. *Dev. Biol.* **344,** 129–137 (2010).
34. Gurdon, J. B., Elsdale, T. R. & Fischberg, M. Sexually mature individuals of *Xenopus laevis* from the transplantation of single somatic nuclei. *Nature* **182,** 64–65 (1958).
35. Gurdon, J. B. & Melton, D. A. Nuclear reprogramming in cells. *Science* **322,** 1811–1815 (2008).
36. Mitsui, K. *et al.* The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113,** 631–642 (2003).
37. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126,** 663–676 (2006).
38. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131,** 861–872 (2007).
39. Wernig, M. *et al. In vitro* reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* **448,** 318–324 (2007).
40. Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318,** 1917–1920 (2007).
41. Chambers, I. *et al.* Nanog safeguards pluripotency and mediates germline development. *Nature* **450,** 1230–1234 (2007).
42. Masui, S. *et al.* Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nature Cell Biol.* **9,** 625–635 (2007).
43. Niwa, H., Miyazaki, J. & Smith, A. G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nature Genet.* **24,** 372–376 (2000).
44. Soufi, A., Donahue, G. & Zaret, K. S. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell* **151,** 994–1004 (2012).
45. Judson, R. L., Babiarz, J. E., Venere, M. & Blelloch, R. Embryonic stem cell-specific microRNAs promote induced pluripotency. *Nature Biotechnol.* **27,** 459–461 (2009).
46. Subramanyam, D. *et al.* Multiple targets of miR-302 and miR-372 promote reprogramming of human fibroblasts to induced pluripotent stem cells. *Nature Biotechnol.* **29,** 443–448 (2011).
47. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11,** R106 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank Y. Kamachi, W. Driever, A. F. Schier, N. Ivanova and I.-H. Park for reagents and fish lines; S. Mane and J. Overton for outstanding sequencing support; E. Zdobnov for hosting genomics data; A. Hubaud, N. Darricarrere and M. Koziol for contribution in the early stages of this project; and all the members of the Giraldez laboratory for discussions. This work was supported by NIH grants F32HD071697-02 (M.T.L.), T32GM007499 (A.R.B), F32HD061194-03 (C.M.T.), Pew Fellows Program in Biomedical Sciences (A.A.B), R01GM081602-06, R01GM103789-01, R01HD074078-02, the Pew Scholars Program in the Biomedical Sciences and the Yale Scholars Program (A.J.G.).

Author Contributions M.T.L., A.R.B. and A.J.G. designed the project, performed experiments and data analysis. M.T.L., A.R.B., C.M.T. and A.J.G. wrote the manuscript. C.M.T. designed and performed the cycloheximide experiment and contributed to *in situ* hybridizations. A.A.B. designed and performed ribosome profiling and U1U2 experiments. K.R.D. and E.S.F. assisted with gene validation.

Author Information Sequencing data are deposited in the Gene Expression Omnibus (GEO) database with accession number GSE47558. 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 A.J.G. (antonio.giraldez@yale.edu).

METHODS

Zebrafish maintenance. MZ*pou5f1*$^{hi349Tg/hi349Tg}$ (ref. 48) were generated as previously described[21]. Embryos obtained from natural crosses between homozygous MZ*pou5f1*$^{hi349Tg/hi349Tg}$ mutants were injected with 30 pg of *pou5f1* mRNA at the one-cell stage. MZ*dicer*$^{hu896/hu896}$ fish were generated as described previously[26]. Zebrafish wild-type embryos were obtained from natural crosses of TU-AB and TLF strains of mixed ages (5–17 months). Selection of mating pairs was random from a pool of 60 males and 60 females allocated for a given day of the month. Fish lines were maintained in accordance with AAALAC research guidelines, under a protocol approved by Yale University IACUC.

Treatments and mRNA injection. Embryos from all wild-type crosses were pooled following collection and distributed equally between experimental conditions. Unless otherwise stated, a minimum of 30 wild-type embryos were subjected to each treatment in each experimental replicate. Morpholinos were obtained from Gene Tools and re-suspended in nuclease-free water. Unless otherwise stated, 1 nl of morpholino solution was injected into dechorionated embryos at the one-cell stage. A combination of two morpholinos was used to target each gene in a 1:1 ratio as described in ref. 20, with one SoxB1 morpholino targeting a conserved region of both *sox2* and *sox3*. Nanog and SoxB1 morpholinos were previously described in refs 20, 24, respectively. For individual and combinatorial loss of function, wild-type and MZ*pou5f1* embryos were injected with 1 ng of each SoxB1 morpholino (0.125 mM each) and 5 ng of Nanog morpholino (0.6 mM each). For inhibition of splicing, one morpholino (1.25 mM each) complementary to U1 and two morpholinos (0.6 mM each) complementary to isoforms of U2 spliceosomal RNAs (U1U2) were used[15,49,50]. Divergence of the U2 genes in zebrafish requires the use of two different morpholinos to block activity.

Zebrafish Nanog and SoxB1 capped mRNA was generated by *in vitro* transcription using mMessage mMachine Sp6 kit (Ambion) in accordance with the manufacturer's instructions. For Nanog morpholino rescue, zebrafish *nanog* was cloned into a pCS2 vector and sense mutations introduced during PCR amplification (indicated in lowercase): 5′-ATGGCaGAtTGGAAaATGCCgGTGAG TTAC-3′. SoxB1 rescue constructs were provided by Y. Kamachi[20]. To rescue the loss-of-function phenotype, 50 pg of Nanog and 20 pg of SoxB1 (5 pg each) mRNAs were injected either individually or together into morpholino-injected embryos at the one-cell stage. Triple loss-of-function embryos were additionally injected with 30 pg of *pou5f1* mRNA[51].

For polymerase II inhibition, α-amanitin was obtained from Sigma Aldrich and re-suspended in nuclease-free water. Dechorionated embryos were injected with 0.2 ng of α-amanitin at the one-cell stage[52].

For translation inhibition, wild-type embryos were collected and dechorionated at the one-cell stage. To allow for translation of maternal mRNAs, at 32-cell stage, embryos were transferred to media containing cycloheximide (50 μg ml^{-1}) (Sigma Aldrich) and incubated at 28 °C. Embryos were collected and frozen in liquid nitrogen at sphere and shield stage. Total RNA was extracted from ten embryos using Trizol (Invitrogen) and re-suspended in 10 μl RNase-free water.

To assay miR-430 activity, a GFP reporter was used as previously described[26]. GFP and dsRed mRNAs were *in vitro* transcribed using mMessage mMachine Sp6 kit (Ambion) in accordance to the manufacturer's instructions. Embryos were injected with 150 pg of GFP reporter and 100 pg of dsRed loading control at the one-cell stage.

All phenotypes were initially assayed by one experimenter and blindly confirmed and/or imaged by another. Distribution-free statistics were used to determine significance, except for calculating RNA-seq differential expression (see below).

***In situ* hybridization.** Template for *in situ* probes was amplified from shield stage cDNA and a T7-promoter sequence added for *in vitro* transcription. Primers are listed below. Antisense digoxigenin (DIG) RNA probes were generated by *in vitro* transcription in 20 μl reactions consisting of 100 ng purified PCR product (8 μl), 2 μl DIG RNA labelling mix (Roche), 2 μl ×10 transcription buffer (Roche), and 2 μl T7 RNA polymerase (Roche) in RNase-free water and purified using a Qiagen RNEasy kit. *In situ* protocol was followed as detailed previously[26]. To reduce variability, the following conditions were combined in the same tube during *in situ* hybridization and recognized based on their morphology: (1) wild-type and α-amanitin-injected embryos and (2) Nanog plus SoxB1 MO with and without rescue mRNA. Before photo documentation, embryos were cleared using a 2:1 benzyl benzoate:benzyl alcohol solution. Images were obtained using a Zeiss stereo Discovery V12.

Northern analysis. To detect endogenous miR-430, ten wild-type and MZ*pou5f1* embryos injected with Nanog morpholino and SoxB1 morpholino mix were collected at 6 h.p.f. and flash frozen in liquid nitrogen. Total RNA was extracted using Trizol (Invitrogen) and re-suspended in 5 μl RNase-free water and 5 μl ×2 loading buffer (8 M urea, 50 mM EDTA, 0.2 mg ml^{-1} xylene cyanol, and 0.2 mg ml^{-1} bromophenol blue). Northern protocol was followed as detailed previously[16].

Ribosome profiling. Fifty wild-type embryos injected with 1 nl of Nanog morpholino (0.6 mM each) and SoxB1 morpholino (0.125 mM each) mix and fifty non-injected embryos were collected at the 64-cell stage. Embryos were lysed using 800 μl of a mammalian cell lysis buffer containing 100 μg ml^{-1} cycloheximide as per the manufacturer's instruction (ARTseq Ribosome Profiling kit, RPHMR12126, Epicentre). For nuclease treatment, 3 μl of ARTseq nuclease was used. Ribosome protected fragments were run and 28–29-nt fragments were gel purified as previously described[16] and cloned according to the manufacturer's protocol (ARTseq kit). Illumina libraries were constructed and sequence reads analysed as in ref. 16. Subsequent to sequencing, traces of exogenous RNA corresponding to a *nanog* antisense probe, and *ntla* sense and antisense, were detected outside the expected size range. Only 28- and 29-nt sense sequences were used in the analysis matching the size of the ribosome footprint.

Reverse transcription PCR (RT–PCR). Total RNA from ten embryos was extracted using Trizol (Invitrogen) at sphere and shield stage for each experimental condition. RNA was treated with TURBO DNase (Ambion) for 30 min at 37 °C and extracted using phenol chloroform. cDNA was generated by reverse transcription with random hexamers using SuperscriptII (Invitrogen). RT–PCR reactions were carried out at an annealing temperature of 60 °C for 35 cycles. Primers are listed below.

Illumina sequencing. Total RNA was extracted as above, and strand-specific TruSeq Illumina RNA sequencing libraries were constructed by the Yale Center for Genome Analysis. Before sequencing, samples were treated with Epicentre Ribo-Zero Gold kits according to the published protocol, to deplete ribosomal RNA. Samples were multiplexed on Illumina HiSeq 2000/2500 machines to produce single-end 76-nt reads. Sequencing samples are summarized in Extended Data Table 1.

Raw reads were initially filtered by aligning permissively to a ribosomal DNA index using Bowtie v0.12.9[53] with switches -seedlen 25 -n 3 -k 1 -y -e 10000. Unaligned reads were then aligned to the zebrafish Zv9 (UCSC danRer7) genome sequence using Tophat v2.0.7[54] with default parameters.

Hybrid gene models were constructed from the union of zebrafish Ensembl r70, RefSeq annotations (downloaded from http://www.genome.ucsc.edu on 8 February 2013) and Ensembl RNA-seq gene models[55]. All overlapping transcript isoforms were merged to produce maximal exonic annotations. To quantify exonic expression levels per gene, genome-uniquely aligning reads overlapping ≥10 nt to the exonic region of a given gene were summed. To quantify intronic expression levels per gene, an annotation mask was first created consisting of repetitive sequences as annotated by RepeatMasker in addition to any region aligned by ≥2 reads in the α-amanitin samples; this is to minimize false-positive introns due to annotation inconsistencies, under the assumption that the transcriptionally inhibited α-amanitin transcriptome should contain no intron-containing transcripts. Valid intron-overlapping reads aligned the intronic region uniquely and overlapped no more than 50% to the masked regions. For the purposes of RPKM normalization, we considered intron length to be the number of unmasked nucleotides. We additionally identified reads that mapped to at most two different genic loci (for example, two closely related paralogues) and from these calculated 'meta gene' expression values. Meta genes were treated as conventional genes for differential expression, but counted as two different genes in subsequent analyses.

The *miR-430* locus is internally repetitive; therefore, reads were aligned to *miR-430* in a separate step using Bowtie with switches -n 2 -k 1 on the genomic region chr4:27999472-28021845, which spans the presumed mir-430 polycistron. Reads overlapping any of the Ensembl annotated miR-430 hairpins in this region were counted as mir-430 cluster reads. Reads are counted only once, regardless of the number of times they overlap.

Differential gene expression analysis. Differential expression analysis was performed using the R package DESeq[47] with the parameters fit-type = local and sharingMode = fit-only. For exonic expression comparisons, raw exon-overlapping read counts were assembled for all genes with a raw read count of at least 10 in one or more of the samples. Genes annotated as Ensembl biotypes 'IG_C_pseudogene', 'IG_pseudogene', 'IG_V_pseudogene', 'misc_RNA', 'Mt_rRNA', 'Mt_tRNA', 'non_coding', 'nonsense_mediated_decay', 'retained_intron', 'rRNA', 'sense_intronic', 'sense_overlapping', 'snoRNA', 'snRNA' were excluded. Additionally, all Ensembl miR-430 annotations were excluded, and a meta 'miR-430 hairpin' gene added in, based on the quantification described in the previous section. For intronic expression comparisons, because overall counts are lower, variance models for DESeq were calculated using both intronic counts and exonic counts as separate gene entries (that is, at most 1 intronic count entry and 1 exonic count entry per gene). Differential expression proceeded as normal, except multiple test correction of *P* values was applied relative only to the intronic counts.

Six sets of differential expression analyses were performed separately: exons and introns for each of (group 1) wild-type 64 cell, wild-type sphere, wild-type shield, U1U2 MO 4 h.p.f., α-amanitin 4 h.p.f. and α-amanitin 6 h.p.f., with the two α-amanitin conditions serving as pseudo replicates for DESeq for variance

estimation; (group 2) sphere stage wild type, Nanog MO, SoxB1 MO, Nanog MO plus SoxB1 MO, MZ*pou5f1*, Nanog MO plus MZ*pou5f1*, SoxB1 MO plus MZ*pou5f1*,Nanog MO plus SoxB1 MO plus MZ*pou5f1*, and two biological replicate shield stage wild-type samples for variance estimation; (group 3) shield stage wild-type, Nanog MO, two Nanog MO plus SoxB1 MO conditions treated as non-replicates, MZ*pou5f1*, SoxB1 MO plus MZ*pou5f1*, Nanog MO plus SoxB1 MO plus MZ*pou5f1*, and two additional biological replicate shield stage wild-type samples to parallel group 2. For groups 2 and 3, we applied an exonic RPKM $\geq$1 and intronic RPKM $\geq$0.5 threshold in one or more of the samples.

Zygotic transcription was determined on the basis of significant exon and intron increases in sphere and shield stages relative to α-amanitin. 64 cell (pre-MZT) was used as further confirmation when no significant changes in intron level were detected or the gene was intronless (genes with $<$10 nt of unmasked intron sequence were considered effectively intronless). Increases in either exon signal, intron signal, or both determined positive zygotic transcription. For genes with a maternal contribution, increases in intronic signal due to zygotic transcription can be accompanied by no change or decreases in exonic signal. For genes significantly expressed, zygotic expression contribution is estimated using either intronic RPKM level or the RPKM difference between the post-MZT condition and the maximum of 64-cell and α-amanitin expression levels. Expression calls are provided in Supplementary Data 1.

To define first-wave genes, genes that were detected as transcribed in the U1U2 MO treated embryos above an expression level of 5 RPKM were considered to be first wave, using an estimate for zygotic transcription based on intronic signal for multi-exon genes, or comparison to α-amanitin and 64 cell for single-exon genes as described above. Although a cutoff of 5 RPKM was used for the main analyses, lower levels of transcription were observed for many genes, indicating weaker degrees of activation. Genes that were not called as transcribed in wild-type sphere were removed from the analysis.

Classification of loss-of-function expression categories. Significant changes in LOF conditions relative to wild type were determined using either intron or exon signal, depending on the pattern of signal originally used to call the gene as zygotically expressed. For genes with no maternal contribution, decreases in either exon or intron levels relative to wild type are considered to be loss of zygotic expression, whereas increases in either exon or intron levels are considered to be ectopic increases in zygotic expression. For genes with maternal contribution, we distinguish between two cases: (1) if zygotic transcription was originally detected in wild type only using intronic signal, then loss of zygotic transcription in the loss-of-function conditions is called only when intronic signal is lost; (2) if zygotic transcription was originally detected in wild type with both exonic and/or intronic signal, then decreases in either intronic levels or exonic levels indicate loss of zygotic expression, with intronic signal taking precedence when the directions of change disagree. For LOF embryos with the MZ*pou5f1* genotype, differential expression was additionally performed between uninjected and injected MZ*pou5f1* conditions, and expression differences between the injected conditions and wild type were required to be transitively consistent—for example, if a gene is called significantly lower in uninjected MZ*pou5f1* than wild type, and a gene is significantly lower in injected MZ*pou5f1* than uninjected MZ*pou5f1*, then the gene must also be considered lower in the injected compared to wild type. To ensure that expression level differences in the MZ*pou5f1* background are due to zygotic contributions, in addition to relying on intron signal, we filtered out any genes that were previously reported to be differentially maternally provided in MZ*pou5f1* (ref. 19).

ChIP-seq analysis. Re-analysis of previously published Nanog ChIP-seq data (GSE34683) was performed as described[24], except using the current version of the zebrafish genome, Zv9. For *miR-430* locus alignment, reads were aligned exhaustively to the region chr4:27994413–28019085 (2 kb ± the miR-430 polycistron) using Bowtie with parameters -v 1 -best -strata -all. To estimate read depth and enrichment, reads were normalized by the number of times the read aligned the genome. To focus on the maximally non-redundant region in the locus, reads were preferentially aligned closest to the presumptive 5′ boundary of the polycistron (chr4:28000732, corresponding to the 5′ end of ENSDARG00000082539).

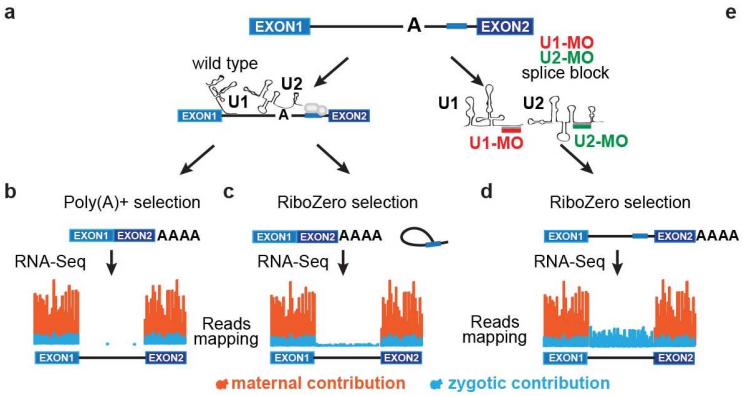
Morpholino oligonucleotide sequence. Sox2 MO1 5′- GAGAGGCTGCTGAA GTTACCTTAGC-3′; Sox2 MO2 5′-CTCGGTTTCCATCATGTTATACATT-3′; Sox3 MO1 5′-TACATTCTTAAAAGTGGTGCCAAGC-3′; Sox3 MO2 5′-GAAG TCAGTCAAAAGTTCAGAGAGC-3′; Sox19a MO1 5′-GTACATGGCTGCCA ACAGAAGTTAG-3′; Sox19a MO2 5′-AAAACGAGAGCGAGCCGTCTGTAA C-3′; Sox19b MO1 5′-GTACATCATGCCACTTCTCGCTTTG-3′; Sox19b MO2 5′-ACGAGCGAGCCTAATCAGGTCAAAC-3′; Nanog MO1 5′-CTGGCATCT TCCAGTCCGCCATTTC-3′; Nanog MO2 5′-AGTCCGCCATTTCGCCGTTA GATAA-3′; U1 MO1 5′-GGTATCTCCCCTGCCAGGTAAGTAT-3′; U2 MO1 5′-TGATAAGAACAGATACTACACTTGA-3′; U2 MO2 5′-TATCAGATATT AAACTGATAAGAAC-3′.

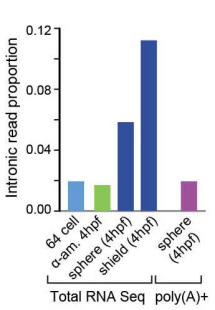
***In situ* primers.** *ntla* forward 5′-TGGAAATACGTGAACGGTGA-3′, reverse 5′-*GTACGAACCCGAGGAGTGAA-3′; *isg15* forward 5′-AGAAGGGCCAGG TCAAAACT-3′, reverse 5′-*CATCACGGCATTGAAAACAC-3′; *cebpb* forward 5′-GTATGCAAGCAGCCAGTCAA-3′, reverse 5′-*TGTACTCGTCGCTGTCC TTG-3′; *cldne* forward 5′-TGGTGTCTATGTGCCGAGAG-3′, reverse 5′-*CGG CTGGGAGTATTTCATGT-3′; *krt18* forward 5′-ATCACCGGCCTAAGAAAG GT-3′, reverse 5′-*TCGTACTCCTGCGTCTGATG-3′; *foxa3* forward 5′-CTTC AACGATTGCTTCGTCA-3′ reverse 5′-*CATCTTCTGCTCGTTGGAC-3′; *vent* forward 5′-ACCCAGCAAGTTCTCAGTGG-3′, reverse 5′-*TAGCAGCGTGTG AACAGCAT-3′; *nnr* forward 5′-CAGAGATGGACAGCGATTCA-3′, reverse 5′-*TTCGTTTCCTTCTGGGAGTTT-3′; *blf* forward 5′-GTCTCACAAGCGAATC CACA-3′, reverse 5′-*GTGTGGGTCTTCTCGTGGTT-3′. Asterisks indicate where a T7 promoter sequence gactTAATACGACTCACTATAGGG was added for *in vitro* transcription

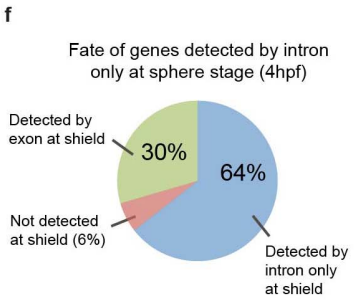
RT–PCR primers. *nnr* forward 5′-AGCGTTTACAGCGGATCTCA-3′, reverse 5′-*AGTGGACGGGGAAATAAACC-3′; *isg15* forward 5′-CGAAAGCCTCA TTCAGCAAC-3′, reverse 5′-*GTGCAACTTCATGCCAGACTC-3′; *cldne* forward 5′-TGGTGTCTATGTGCCGAGAG-3′, reverse 5′-*CGGCTGGGAGTATTTC ATGT-3′; *sox11a* forward 5′-CGAAACGGACAGCATGTCTA-3′, reverse 5′-GG AGTCGTCATCGTCGTCTT-3′; *grhl3* (1/2) forward 5′-GAGGAGACCGGATA CCAAACT-3′, reverse 5′-CCAAGCTCCACTGTGTTTGT-3′; *grhl3* (1/3) forward 5′-GAGGAGACCGGATACCAAACT-3′, reverse 5′-TTGTAAATGCTGCTCT CACG-3′; *cldnb* forward 5′-ACTCCCCATGTGGAAAGTCA-3′, reverse 5′-GG GGTTGCGTTGTATTTAGC-3′; *krt4* forward 5′-GCAACCTCCTCCACTCAC TC-3′, reverse 5′-AATTGTGGGGTCAATTTCCA-3′; *hist1h2aa* forward 5′-CA AAGGCTAAGACTCGCTCCT-3′, reverse 5′-TCTGTCTTCTTGGGCAGCAG-3′; *tubb4b* forward 5′-AGGTCTGGTCCATTTGGTCA-3′, reverse 5′-CATCCA GAACGGAATCAACC-3′; *klf4b* forward 5′-ACAGTTGTGAATTCCCTGGAT G-3′, reverse 5′-GTTTACATGTGCCTCTTCATGTG-3′; *vox* forward 5′-GAC TGGCTTGCTCAGAGCTT-3′, reverse 5′-GGCCGCTTCACTCTCATAAC-3′; *tbx16* forward 5′-AACCTTTACCTTCCCCGAGA-3′, reverse 5′-CAAGACTCG GGACTCAAAGC-3′.

qRT–PCR primers. *blf* forward 5′-CCCTGCTGAGCTTGCATAGT-3′, reverse 5′-CCCACACTGAGGACACTTGA-3′; *cldne* forward 5′-GGCTTCTTGGGAG CCATTAT-3′, reverse 5′-GCGAAAAAGCTGACGATGAT-3′; *ctcf* forward 5′-GTTAGCAGAGGCTTGCTTTACTG-3′, reverse 5′-GCAGTGAAATTTCGCC ACA-3′; *dact1* forward 5′-AGCCTCGGTTCTTCTTCACA-3′, reverse 5′-GGA GGATTTGTGCAAGTGGT-3′; *dusp1* forward 5′-CTCCAGTAATGTGCGCTT CA-3′, reverse 5′-TGGTCGAACTTTTGACCTTCA-3′; *ef1a* forward 5′-TGAT CTACAAATGCGGTGGA-3′, reverse 5′-CAATGGTGATACCACGCTCA-3′; *her5* forward 5′-CCAAGCCTCTCATGGAGAAA-3′, reverse 5′-TAGCTCTGA CGTTTGCATGG-3′; *mtATP6* forward 5′-CTTTAGCGGCCACAAATGAG-3′, reverse 5′-ATGGGGGTTCCTTCTGGTAA-3′; *mtND5* forward 5′-TTCTTAT GCTCAGGGGCAAT-3′, reverse 5′-TTAGGGCTCAGGCGTTAAGA-3′; *mxtx1* forward 5′-GAAATGCAAGGGTGGAAAAA-3′, reverse 5′-ACCCCAGTTAGG AGGCATCT-3′; *oep* forward 5′-TTCTGGAAAGCCAAAGCAAT-3′, reverse 5′-TCATGTCAGTGTGCAGCTTG-3′; *pcf11* forward 5′-CCTCGCTGGAAGATC TGACT-3′, reverse 5′-CATGTTACAGGCCTCATGTCA-3′; *tdp2b* forward 5′-GG AGCCCACCTGCTCTATTA-3′, reverse 5′-ACCCTGCCAATTGTGAAGATA-3′.

48. Amsterdam, A. *et al.* Identification of 315 genes essential for early zebrafish development. *Proc. Natl Acad. Sci. USA* **101,** 12792–12797 (2004).
49. Matter, N. & Konig, H. Targeted 'knockdown' of spliceosome function in mammalian cells. *Nucleic Acids Res.* **33,** e41 (2005).
50. Rösel, T. D. *et al.* RNA-Seq analysis in mutant zebrafish reveals role of U1C protein in alternative splicing regulation. *EMBO J.* **30,** 1965–1976 (2011).
51. Iwafuchi-Doi, M. *et al.* The Pou5f1/Pou3f-dependent but SoxB-independent regulation of conserved enhancer N2 initiates Sox2 expression during epiblast to neural plate stages in vertebrates. *Dev. Biol.* **352,** 354–366 (2011).
52. Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B. & Schilling, T. F. Stages of embryonic development of the zebrafish. *Dev. Dyn.* **203,** 253–310 (1995).
53. 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).
54. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25,** 1105–1111 (2009).
55. Collins, J. E., White, S., Searle, S. M. & Stemple, D. L. Incorporating RNA-seq data into the zebrafish Ensembl genebuild. *Genome Res.* **22,** 2067–2078 (2012).
56. Berg, M. G. *et al.* U1 snRNP determines mRNA length and regulates isoform expression. *Cell* **150,** 53–64 (2012).
57. Chen, X. *et al.* Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133,** 1106–1117 (2008).
58. Marson, A. *et al.* Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* **134,** 521–533 (2008).
59. Hamatani, T., Carter, M. G., Sharov, A. A. & Ko, M. S. Dynamics of global gene expression changes during mouse preimplantation development. *Dev. Cell* **6,** 117–131 (2004).



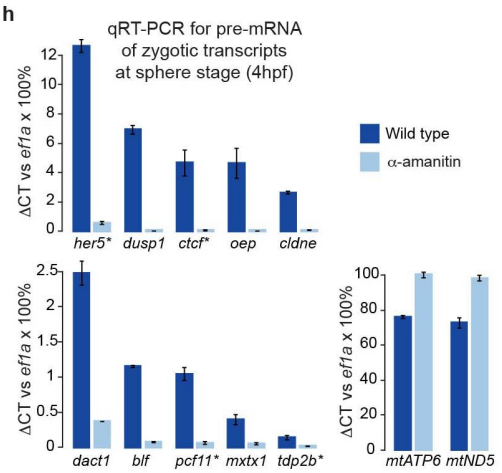


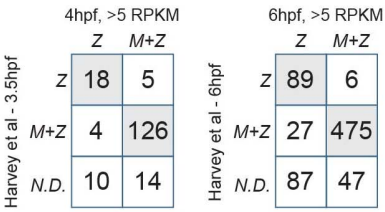


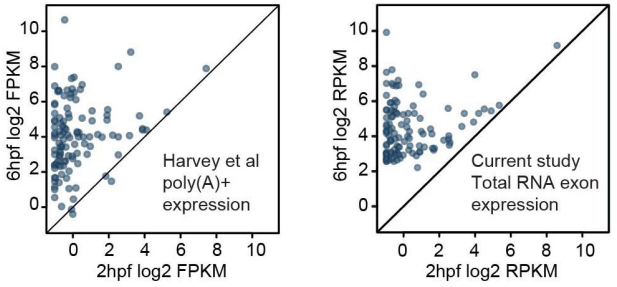
g

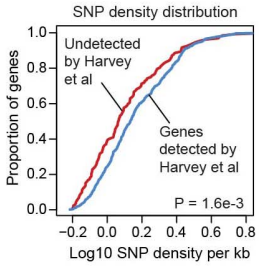
	All detected (Ensembl + RefSeq)			>5 RPKM zygotic gain		
	Sphere	Sphere + U1U2 MO	Shield	Sphere	Sphere + U1U2 MO	Shield
Strictly zygotic *	1760	770	2422	197	104	791
Maternal + zygotic †						
Detected through exon signal	723	648	2346	159	91	1454
Detected through intron signal only	4727	2976	4562	252	79	193
Total	7210	4394	9330	608	274	2438

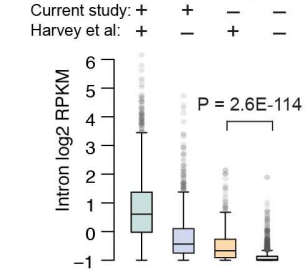
* Maternal contribution <= 1 RPKM, as observed in α-amanitin treated or 64-cell embryos
† Maternal contribution > 1 RPKM

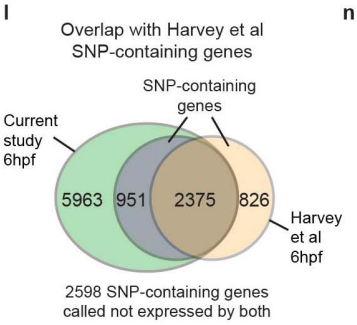


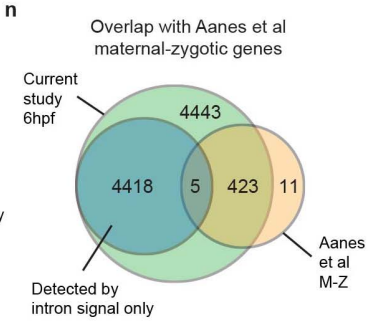


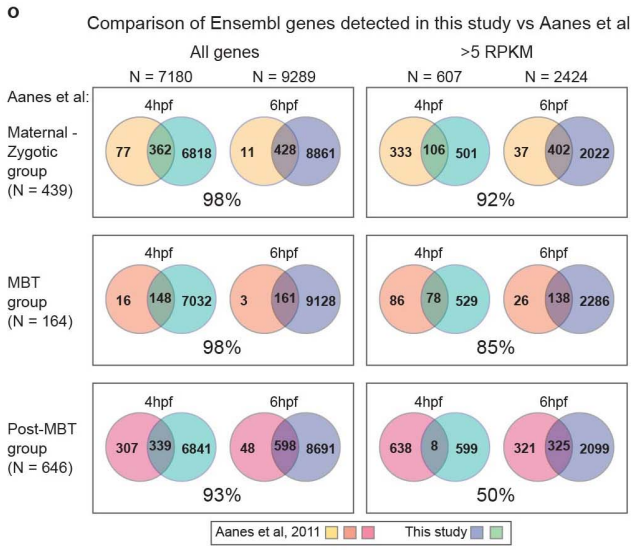




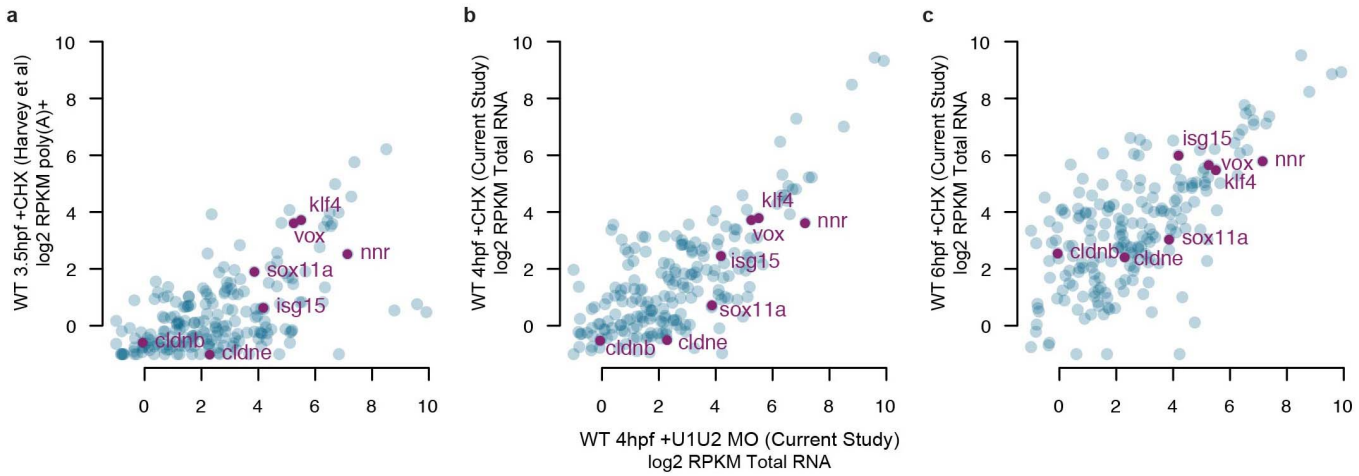


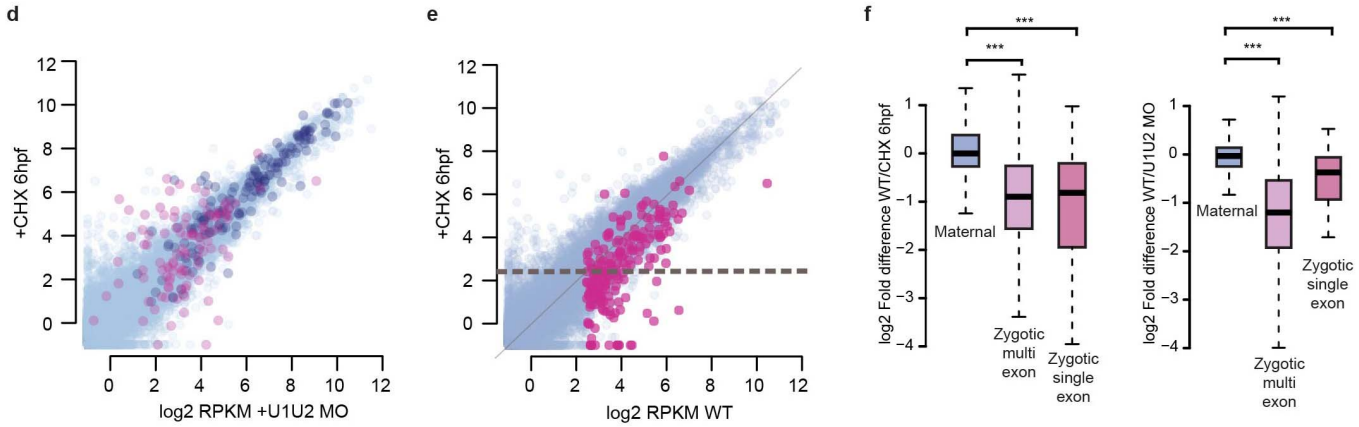


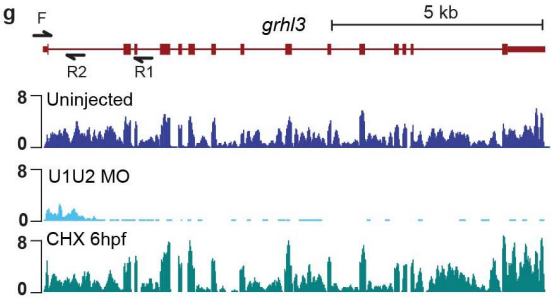


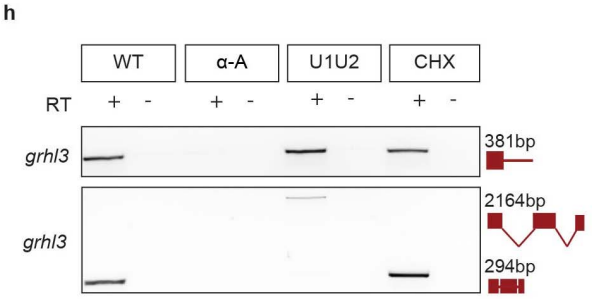


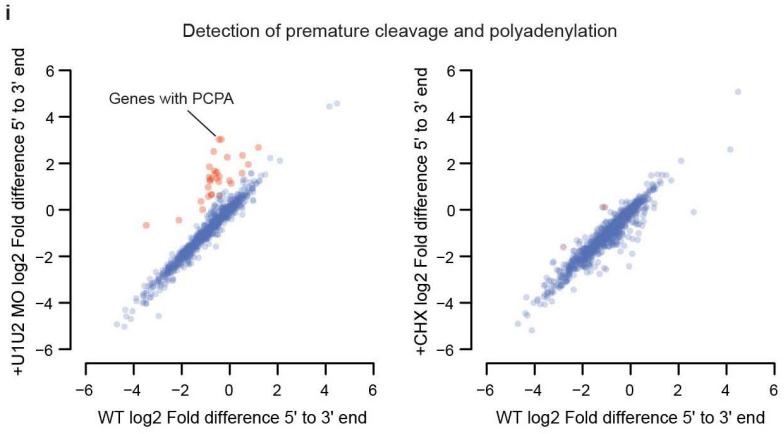
Extended Data Figure 1 | Identifying *de novo* zygotic transcription. **a**, Schematic of the sequencing strategy used in this study. Most zebrafish protein-coding genes (>95%) contain introns. *De novo* transcription produces intronic RNA sequences, which are spliced out of pre-mRNAs by the spliceosome, consisting of several ncRNA species including U1 and U2. **b**, Typical mRNA-seq applications use poly(A)$^+$ selection to enrich for the mature mRNA population. Sequence reads map predominantly to exonic regions, with very few reads mapping to introns. During embryogenesis, many zygotic transcribed genes are expected to have a maternal contribution in the cytoplasm from the oocyte. The resulting signal will be a mixture of maternal-derived (orange) and zygotic-derived (blue) mRNA molecules, which cannot be deconvoluted without comparing to a reference sample to look for exon expression level change. **c**, mRNA-seq applications that skip poly(A)$^+$ selection and instead use a rRNA depletion protocol (RiboZero) will not enrich for the mature mRNA population. Thus, transcripts in all stages of biogenesis (pre-mRNA, partially spliced mRNA, spliced introns) will be sequenced, and reads are expected to map both to exons and introns. Because maternally contributed mRNAs are mature, any intron signal detected must derive from *de novo* zygotic transcription. To determine the background signal for each intron, α-amanitin is used as a negative control for transcription. **d**, Morpholinos complementary to U1 and U2 injected into one-cell embryos inhibit zygotic splicing. Thus, pre-mRNAs fail to be processed, and the entire population of zygotic mRNAs will be unspliced. There are two benefits: (1) intron signal is amplified, as introns are stabilized in the pre-mRNA compared to spliced out introns; (2) protein production from zygotic mRNAs is effectively halted, as pre-mRNAs are generally not competent for normal translation. Only the first wave of transcription, resulting from activation by maternal factors, is observed. Transcription that requires zygotic proteins (subsequent waves) will be largely absent. **e**, The proportion of sequencing reads aligning to gene introns. Total RNA sequencing reveals elevated intronic sequence reads, corresponding to *de novo* zygotic transcription. **f**, The fate of the 5,318 sphere-stage (4 h.p.f.) zygotic genes that are only detectable through significant changes in intron sequence. At shield stage (6 h.p.f.), 64% of the genes are still detected as zygotically transcribed based only on intron signal. These include genes that have simultaneous zygotic transcription with decay of the maternal contribution. 30% of the genes are detected using both exon and intron signal by shield stage, indicating that transcription levels at sphere stage were too low to detect differences in exons, but were apparent in the introns. **g**, Number of genes detected in wild-type sphere-stage embryos, sphere embryos injected with U1U2 MO and wild-type shield-stage embryos, at different thresholds of detection. For both groups, a multiple test-corrected $P < 0.1$ threshold (Benjamin–Hochberg) was used for differential expression of exonic signal. For intronic signal, an uncorrected $P < 0.1$ was used for the 'All detected' group, whereas a multiple test-corrected $P < 0.1$ was used for the >5 RPKM gain group. **h**, Quantitative RT–PCR was performed for select genes to confirm zygotic transcription in wild-type sphere-stage embryos (dark blue bars) compared to α-amanitin-treated embryos (light blue bars). Primers were designed to amplify pre-mRNAs across exon–intron boundaries, except for *cldne*. Expression levels are reported as percentage of CT value compared to a maternally provided housekeeping gene (*ef1a*) ($\Delta CT \times 100\%$). Error bars show s.e.m. for three technical replicates. Increased pre-mRNA levels were observed for all zygotic genes tested between wild type and α-amanitin. Maternally provided genes *mtATP6* and *mtND5* show no increase in wild type. Genes marked with an asterisk represent the bottom 10% of significant differential intron expression based on the RNA-seq data (which quantifies both pre-mRNA and spliced introns). This shows that using intron signal is a reliable indication of zygotic transcription. **i**, Genes detected in this study were compared to previous annotations of zygotic transcripts[13], which used SNPs to identify transcripts derived from paternal alleles, to distinguish zygotic transcription from the maternal contribution. From their genomic sequencing results, we extracted 6,750 genes with informative exonic SNPs, which were consistently called between the two sets of matings. 178 of the genes we call zygotically transcribed at sphere stage at levels >5 RPKM are among the 6,750 informative genes. 87% of these are also found to be transcribed by ref. 13, with agreement between both strictly zygotic genes (Z) and maternal+zygotic genes (M+Z). 24 genes were not detected by ref. 13 (N.D.). At shield stage, 82% of the zygotic genes are also found by ref. 13, with 134 genes not detected. **j**, These undetected genes nevertheless have highly increased expression pre-64-cell to post-MZT (shield) using the RNA-seq data generated by ref. 13 (left) and in the current study (right). **k**, Cumulative plots show that SNP density is significantly lower among ref. 13 undetected genes at shield compared to detected genes ($P = 1.6 \times 10^{-3}$, two-sided Wilcoxon rank sum test), suggesting that low SNP density may account for the missed genes. **l**, Overall, ref. 13 and the current study distinguish a similar number of zygotic versus maternal transcripts at 6 h.p.f., among Ensembl genes with informative SNPs, with 74% agreement. However, 64% of zygotic transcripts identified in the current study do not have informative SNPs, and are thus not called transcribed by ref. 13. **m**, Genes called transcribed by ref. 13 but not in the current study have significantly higher intron signal than maternal genes ($P = 1.4 \times 10^{-95}$, two-sided Wilcoxon rank sum test), indicating that our significance threshold to detect zygotic transcription is conservative. **n**, Reference 14 used a time course poly(A)$^+$ RNA-seq strategy to define zygotic transcripts. The comparable r70 Ensembl genes in the ref. 14 maternal+zygotic gene category are largely found in our study; however, we find thousands more transcribed genes based on intron signal—these genes represent transcription that is masked by the maternal contribution. **o**, Overall, our study captures most of the zygotic genes in the three categories described by ref. 14: maternal–zygotic genes (zygotic genes with maternal contribution, yellow), MBT genes (strictly zygotic genes detected at MBT, 3.5 h.p.f., orange), and post-MBT genes (strictly zygotic genes detected at 5.3 h.p.f., pink). Venn diagrams show the number of comparable r70 Ensembl genes that overlap between the two studies. Left panels include all zygotic genes detected in this study; right panels impose a zygotic expression threshold of >5 RPKM. Percentages within each box are calculated as the number of genes detected in this study (at either time point) that overlap the respective ref. 14 group, divided by the size of the ref. 14 group. The overlap percentages are generally high, indicating that our study recovered genes previously annotated as zygotically transcribed as well as many additional zygotic genes based on the use of intronic reads.



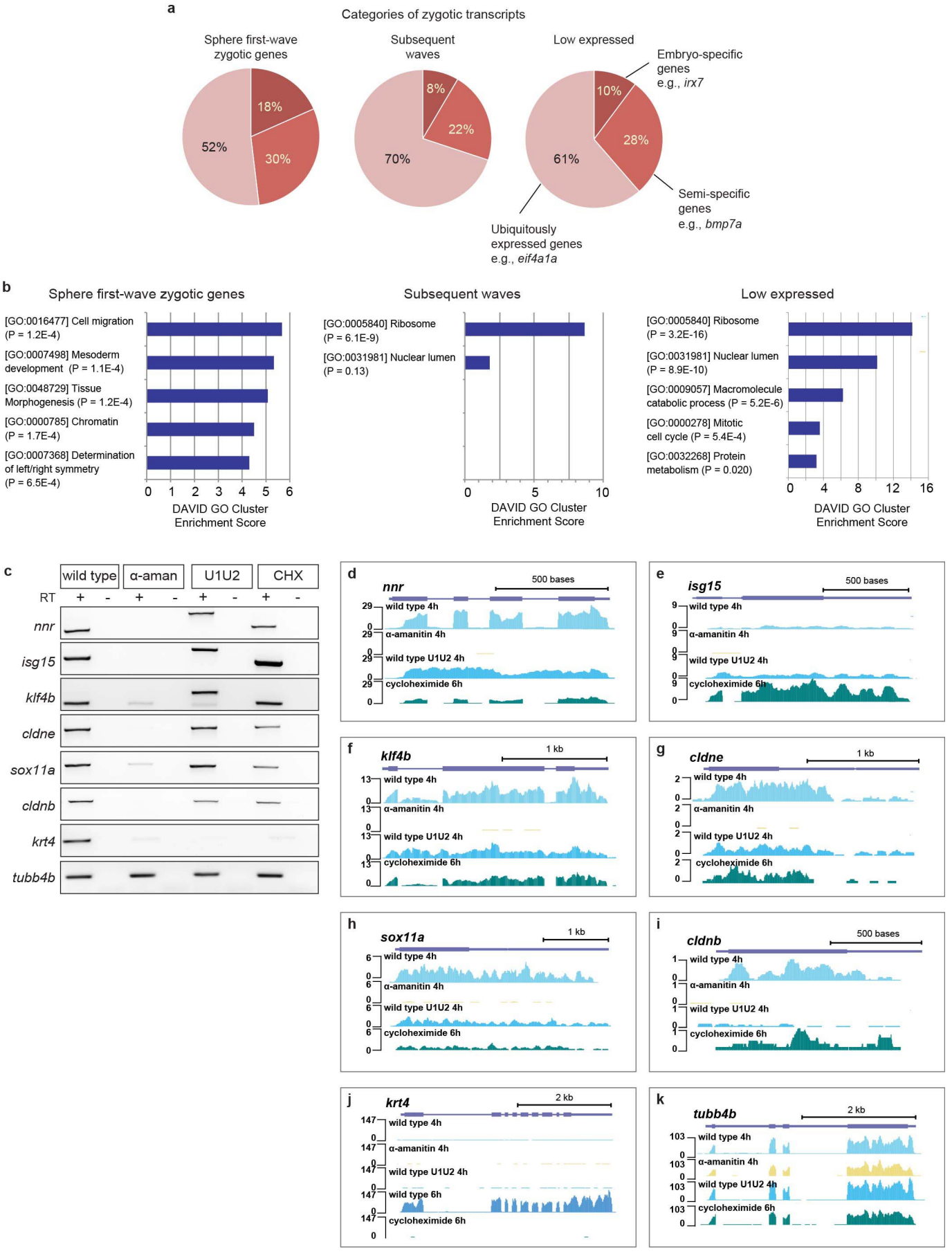




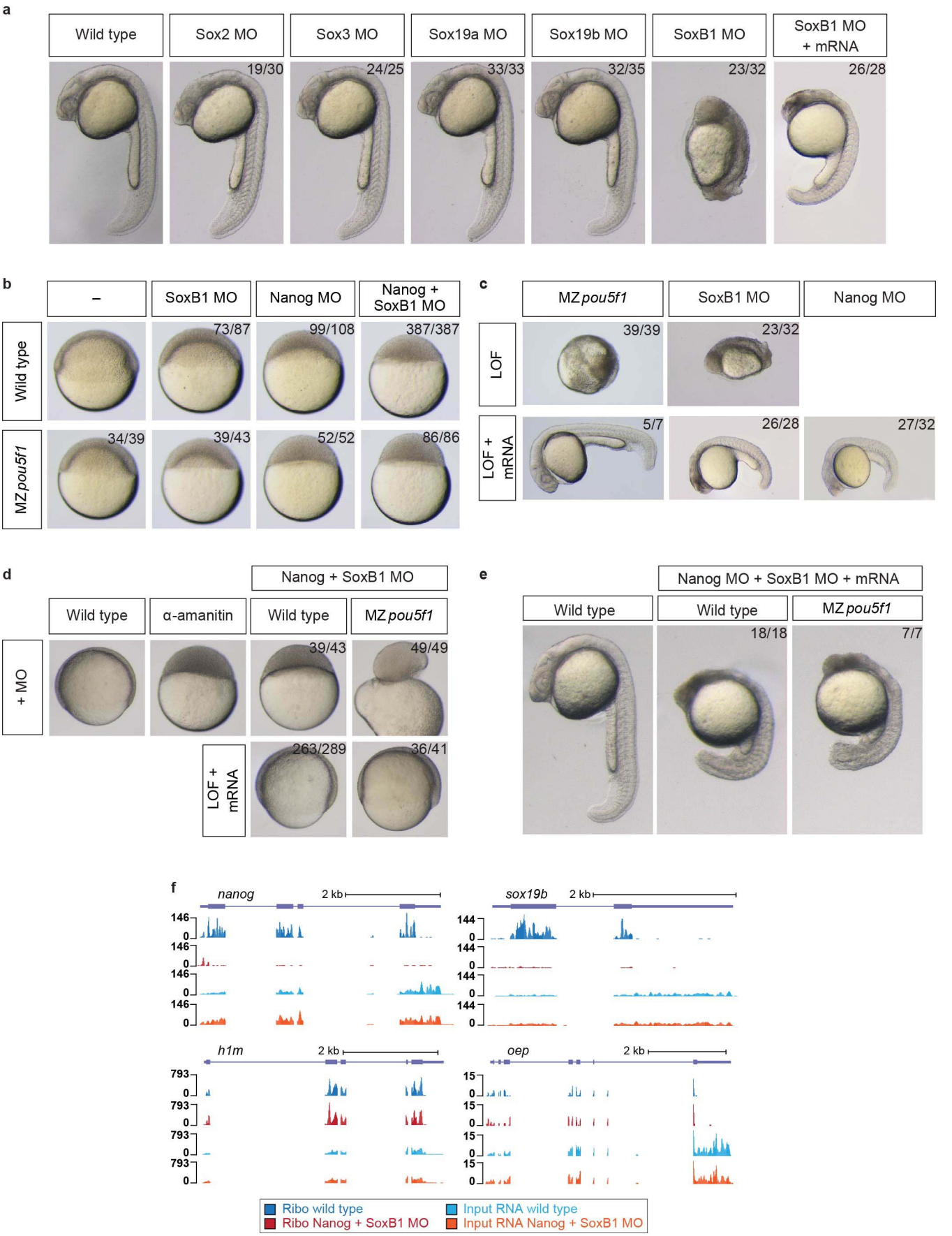




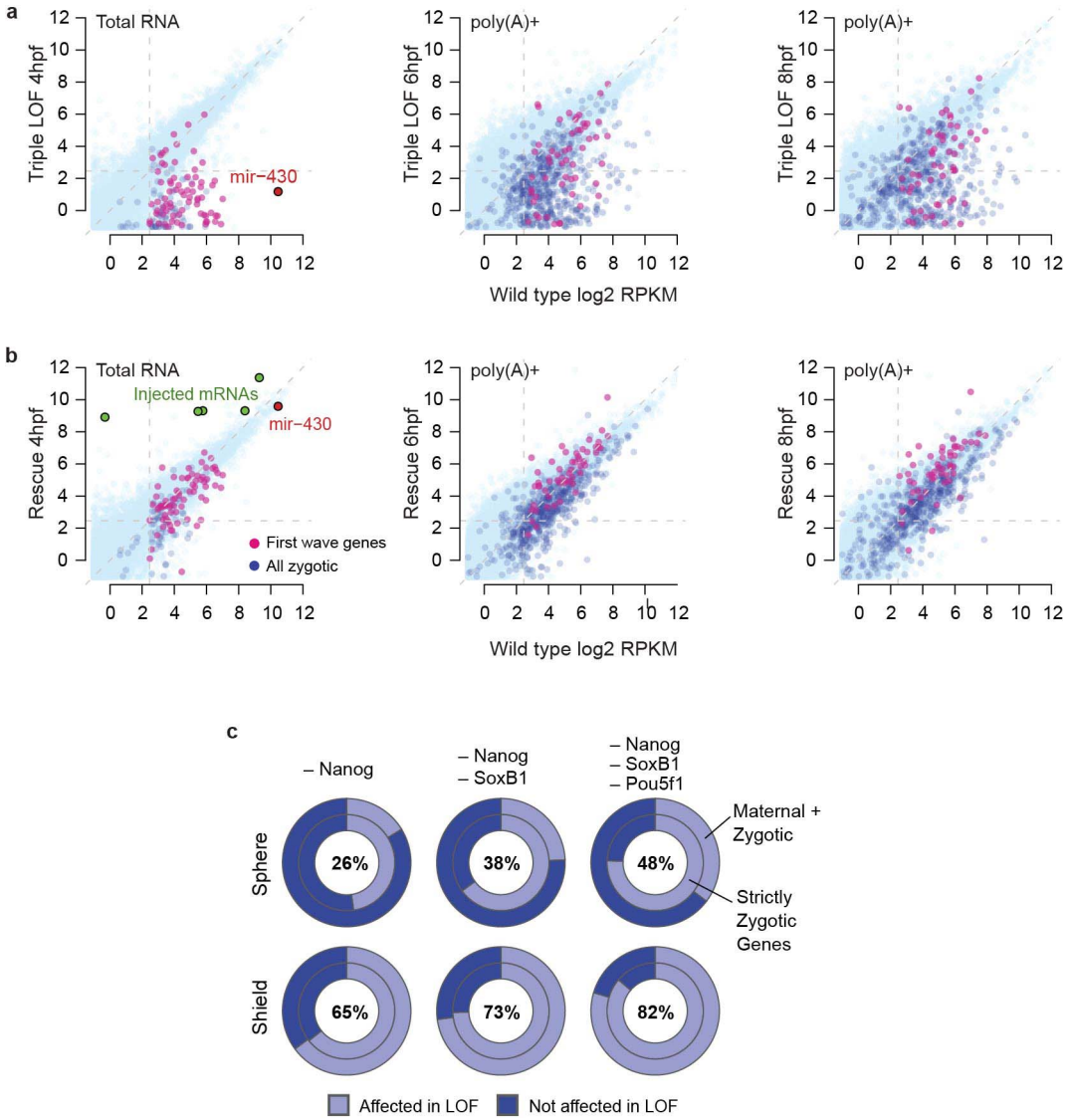
Extended Data Figure 2 | Cycloheximide and U1U2 MO transcriptomes show first-wave genes. **a**–**c**, Biplots comparing strictly zygotic genes found by either the current study or ref. 13 at >5 RPKM ($N = 202$). Zygotic expressed genes of ref. 13 were identified by comparing their raw RNA-seq data at 128-cell (pre-MZT) versus 3.5 h.p.f. In **a**, zygotic expression in U1U2 MO treated embryos (Total RNA, 4hpf) is compared to ref. 13 embryos treated with cycloheximide (CHX) (poly(A)$^+$, assayed at 3.5 h.p.f.), which shows lagging expression of many first-wave genes (defined as having >5 RPKM in U1U2 MO). Genes verified by RT–PCR as first wave (*klf4*, *nnr*, *sox11a*, *isg15*, *cldne*) are highlighted, in addition to *cldnb*, which misses the threshold for first wave in the U1U2 MO transcriptome, and *vox*, which was highlighted by ref. 13. In **b**, **c**, Embryos treated with CHX and assayed in the current study at 4 h.p.f. and 6 h.p.f. (Total RNA) show gradual increases in expression of zygotic genes. Together these results suggest that expression of first-wave genes is independent of *de novo* zygotic factors, and that transcription overall is slower in CHX-treated embryos compared to wild type or U1U2 MO. **d**, Biplot showing gene expression levels (exonic) for all genes in U1U2 MO embryos at 4 h.p.f. compared to CHX-treated embryos assayed at 6 h.p.f. Magenta points, strictly zygotic genes; dark-blue points, maternal+zygotic genes. 97% of the first-wave genes called in U1U2 MO were expressed >1 RPKM in the CHX condition. **e**, Biplot comparing exonic expression levels between wild-type (4 h.p.f.) and CHX-treated embryos. Magenta points are strictly zygotic genes expressed >5 RPKM in wild type. The dotted line indicates 5 RPKM expression in CHX. **f**, Box-and-whisker plots comparing exonic expression level differences between wild-type and treated embryos in maternal genes, strictly zygotic multi-exon genes, and strictly zygotic single-exon genes. Both U1U2 MO and CHX-treated embryos show loss of expression in zygotic genes compared to wild type (U1U2 MO: $P = 9.4 \times 10^{-207}$ for multi-exonic, $P = 4.2 \times 10^{-4}$ for single exon, Wilcoxon rank-sum test comparing to maternal; CHX: $P = 4.3 \times 10^{-137}$ multi-exon, $P = 1.5 \times 10^{-6}$ single exon). The box defines the first and third quartiles, with the median indicated with a thick black line. The systemic decreases in expression in the U1U2 MO or CHX conditions compared to wild type indicate that although maternal factors can activate to a large extent expression of the first-wave genes, additional zygotic contribution of transcription factors (Nanog, SoxB1 and Pou5f1, but possibly others as well) might be required to reach wild-type levels of expression for many genes. This was also observed in ref. 13 for the gene *vox*. Alternatively, lower expression of first-wave zygotic genes might be caused by reduced level of maternal encoded proteins, as incubation with CHX at 32-cell stage might also decrease translation of the maternally deposited mRNAs. We consistently observe that CHX-treated embryos show lower/delayed expression compared with U1U2-MO-treated embryos, indicating that premature inhibition of maternal mRNA translation has an effect on the rate of activation of the first-wave genes. **g**, UCSC Genome Browser track showing an example of premature cleavage and polyadenylation (PCPA) for *grhl3*. Arrows indicate primer sites for RT–PCR. Previously, it was shown that U1 snRNA also serves to protect nascent mRNAs from PCPA, and that U1 inhibition results in 3′-truncation that may affect transcript level quantification[56]. **h**, RT–PCR for *grhl3* on shield-stage embryos ($N = 5$). Wild-type (WT), U1U2 MO and CHX-treated embryos all amplify a 381-bp fragment from exon 1 to the beginning of intron 1. U1U2-MO-injected embryos amplify an unspliced 2,164-bp gene product spanning exon 1 to 3, whereas wild-type and CHX-treated embryos have a 294-bp spliced product, with α-amanitin as a negative control. **i**, Biplots comparing expression levels at the 5′ end of a transcript compared to the 3′ end, to detect PCPA at 4 h.p.f. Read density was assayed in up to 1,000 nucleotides of 5′ and 3′ sequence per transcript. The range of asymmetry values in wild type reflects sequencing biases or transcript annotation irregularities. Several genes in U1U2 MO embryos show elevated asymmetry compared to wild-type (orange dots, >twofold), reflecting a drop-off of read density moving 5′–3′ in the transcript, indicative of PCPA. These genes are included in our annotations of the zygotic first wave of expressed genes. The minor extent of PCPA during embryogenesis may reflect the short length of many of the zygotic genes, as PCPA is associated with longer genes that are likely to harbour cryptic polyadenylation sites. Transcripts in CHX-treated embryos generally do not show this trend.



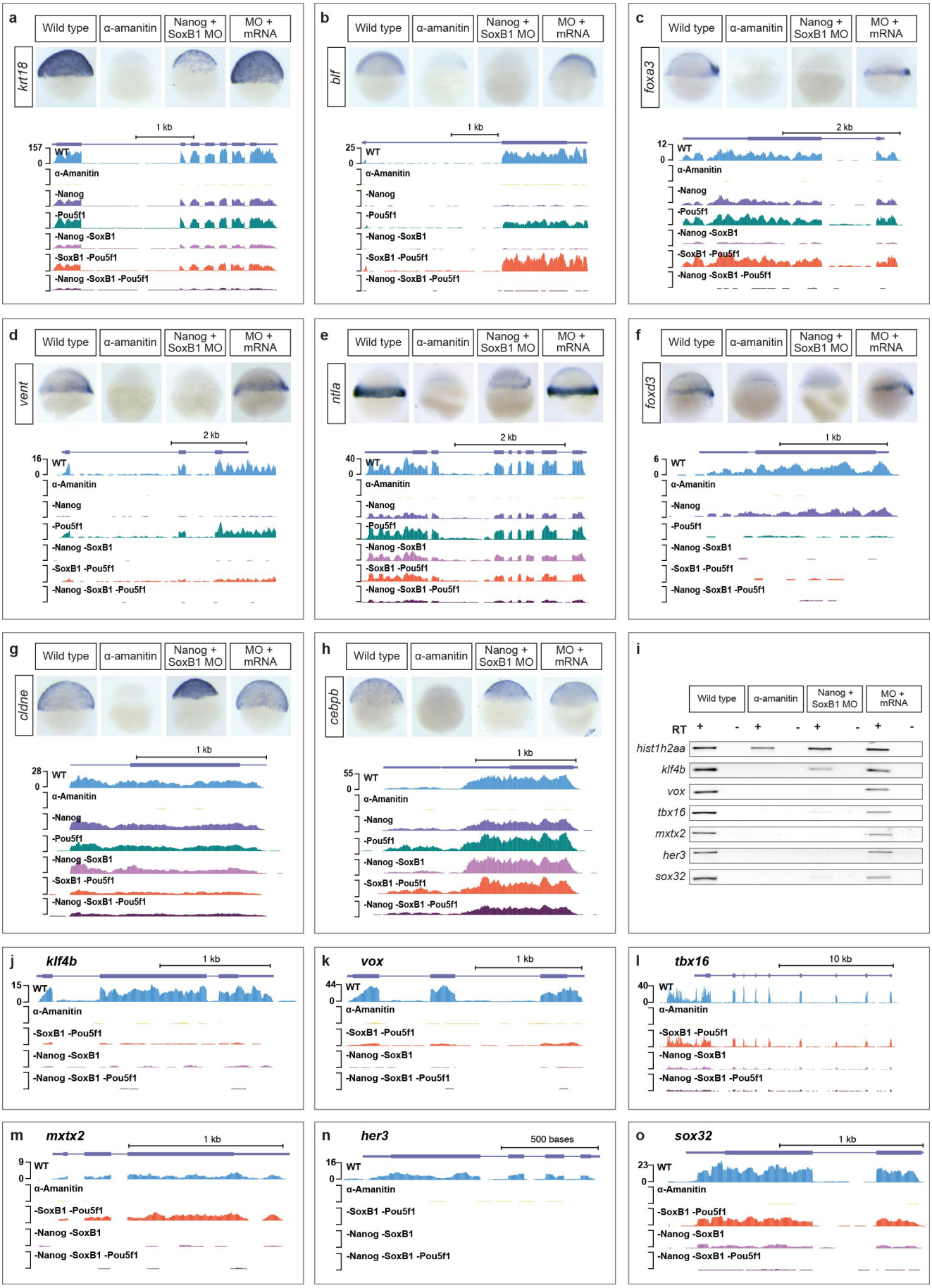
Extended Data Figure 3 | Verification of first-wave gene expression and functional categories. **a**, To assay the embryonic specificity of the first-wave genes, we used publicly available microarray data from NCBI GEO across eight normal adult tissue types (brain, GSE11107; liver, GSE11107; heart, GSE17993; skin, GSE24528; kidney, GSE32363; digestive tract, GSE35889; ovary, GSE14979; testis, GSE14979) to classify genes as expressed specifically in the embryo (called 'present' by the MAS5 algorithm in 0–2 different adult tissues), genes expressed semi-specifically (present in 3–5 different adult tissues), and genes expressed ubiquitously (present in 6–8 different adult tissues); this latter group would correspond to 'housekeeping' genes. Sphere-stage first-wave genes consist of a mixture of specifically expressed and housekeeping genes. Subsequent-wave genes and genes expressed at levels <5 RPKM consist of a larger proportion of genes typically expressed ubiquitously in adult fish, suggesting a widespread activation of genes encoding general cellular processes in addition to developmentally specific ones. **b**, Gene Ontology enrichment analysis for first-wave, subsequent-wave and the low expressed genes with intronic RPKM >0.5. Top 5 scoring clusters are shown for each gene set. Clusters were defined using DAVID (http://david.abcc.ncifcrf.gov) Gene Functional Annotation Clustering on GO 'FAT' annotations and 'high' stringency. Clusters are annotated with representative GO terms and corresponding Benjamini–Hochberg FDR corrected *P* values. **c**, To validate genes activated in the first wave versus subsequent waves, RT–PCR was performed on shield stage (6 h.p.f.) in wild-type, α-amanitin, U1U2 MO and cycloheximide (CHX)-treated embryos. The unspliced products for *nnr*, *isg15* and *klf4* are detected only in U1U2 morphants, confirming that U1U2 is indeed blocking splicing. CHX treatment indicates the single-exon genes *cldne* and *sox11a* are activated in the first wave. *cldnb* is detected at low levels in wild type, as well as both U1U2 MO and CHX-treated embryos; however, based on RNA-seq levels at sphere stage, this gene does not pass the expression threshold to be called first wave. *krt4* is significantly reduced in U1U2 MO and CHX-treated embryos, indicating that zygotic factors are required for its activation. Maternal *tubb4b* is present in all conditions. **d**–**h**, UCSC Genome Browser tracks for first-wave genes *nnr*, *isg15*, *klf4*, *cldne* and *sox11a*. **i**, UCSC Genome Browser track for *cldnb*, which shows low expression levels at sphere stage. **j**, **k**, UCSC Genome Browser track for a gene activated in subsequent waves (*krt4*) and for a maternally provided gene (*tubb4b*).



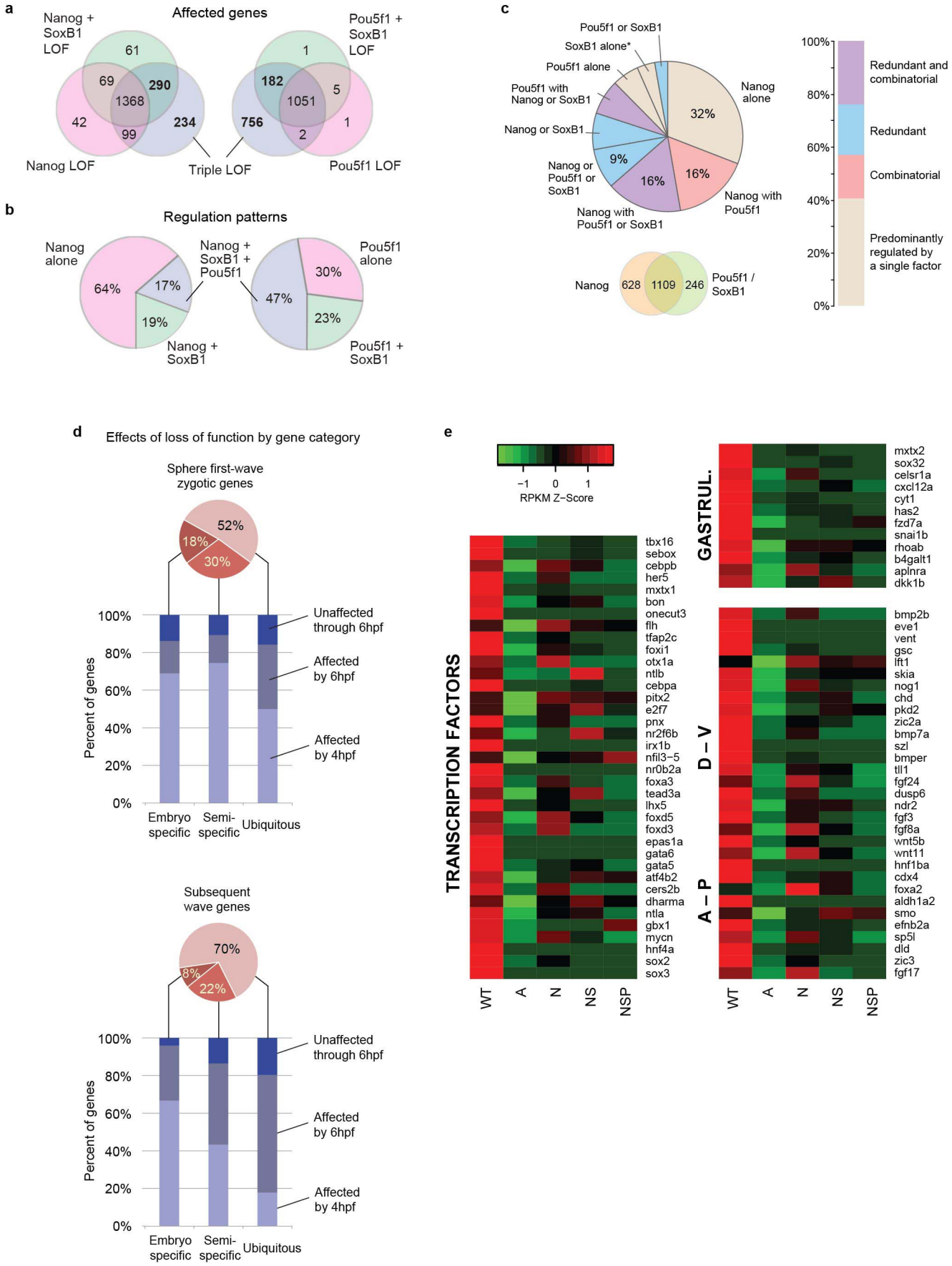
Extended Data Figure 4 | Loss-of-function and rescue for Nanog, SoxB1 and Pou5f1. **a**, Wild-type embryos were injected with Sox2, Sox3, Sox19a and Sox19b morpholinos individually and in combination (0.125 mM). Consistent with other reports, only quadruple LOF results in severe developmental defects (27 h.p.f.)[20]. LOF phenotype is rescued by injecting *soxb1* mRNA (imaged at 24 h.p.f.). **b**, Wild-type and MZ*pou5f1* embryos were injected with *SoxB1* MO (0.125mM each) and *Nanog* MO (0.6mM each) individually and in combination (Nanog + SoxB1). Loss of Nanog results in severe gastrulation defects and failure to progress past 80% epiboly, as previously reported[24]. Loss of SoxB1 in both wild-type and MZ*pou5f1* embryos showed developmental delay, whereas combined LOF for Nanog/SoxB1 or Pou5f1/Nanog completely arrested development before epiboly. Triple LOF embryos also arrested and failed to undergo gastrulation. **c**, Individual LOF for Nanog, SoxB1 and Pou5f1 resulted in developmental abnormalities (top panel). Embryos with Nanog LOF did not progress past 80% epiboly. The LOF phenotypes were rescued by injecting the respective mRNAs (LOF + mRNA) (bottom panel). Embryos imaged at 23 h.p.f. **d**, **e**, Wild-type and MZ*pou5f1* embryos were co-injected with Nanog + SoxB1 MO. LOF embryos arrest at sphere stage and resemble α-amanitin-injected embryos (+MO). Combinatorial LOF is rescued with co-injection of the respective mRNAs (MO + mRNA). Embryos were imaged when wild-type siblings reached 80% epiboly (**d**) and 24 h.p.f. (**e**). **f**, Ribosome profiling was performed at 2 h.p.f. on wild-type embryos and embryos injected with Nanog and SoxB1 morpholino at one-cell stage, to determine the specificity of the morpholinos to repress translation of *nanog* and *soxB1* mRNA. Sequenced ribosome protected fragments (RPFs) were predominantly 28–29 nucleotides long, indicative of the width of the ribosome footprint. UCSC Genome Browser tracks (sense strand) showing ribosome profiling (top 2 tracks per gene) and input mRNA (bottom 2 tracks per gene). *nanog* and *sox19b* show significant reduction in RPFs in the Nanog MO + SoxB1 MO injected embryos compared to wild type. Input mRNA is unaffected. Neither *h1m*, a highly expressed gene, nor *oep*, a low expressed gene, has any change in either RPFs or input mRNA between wild-type and injected embryos.



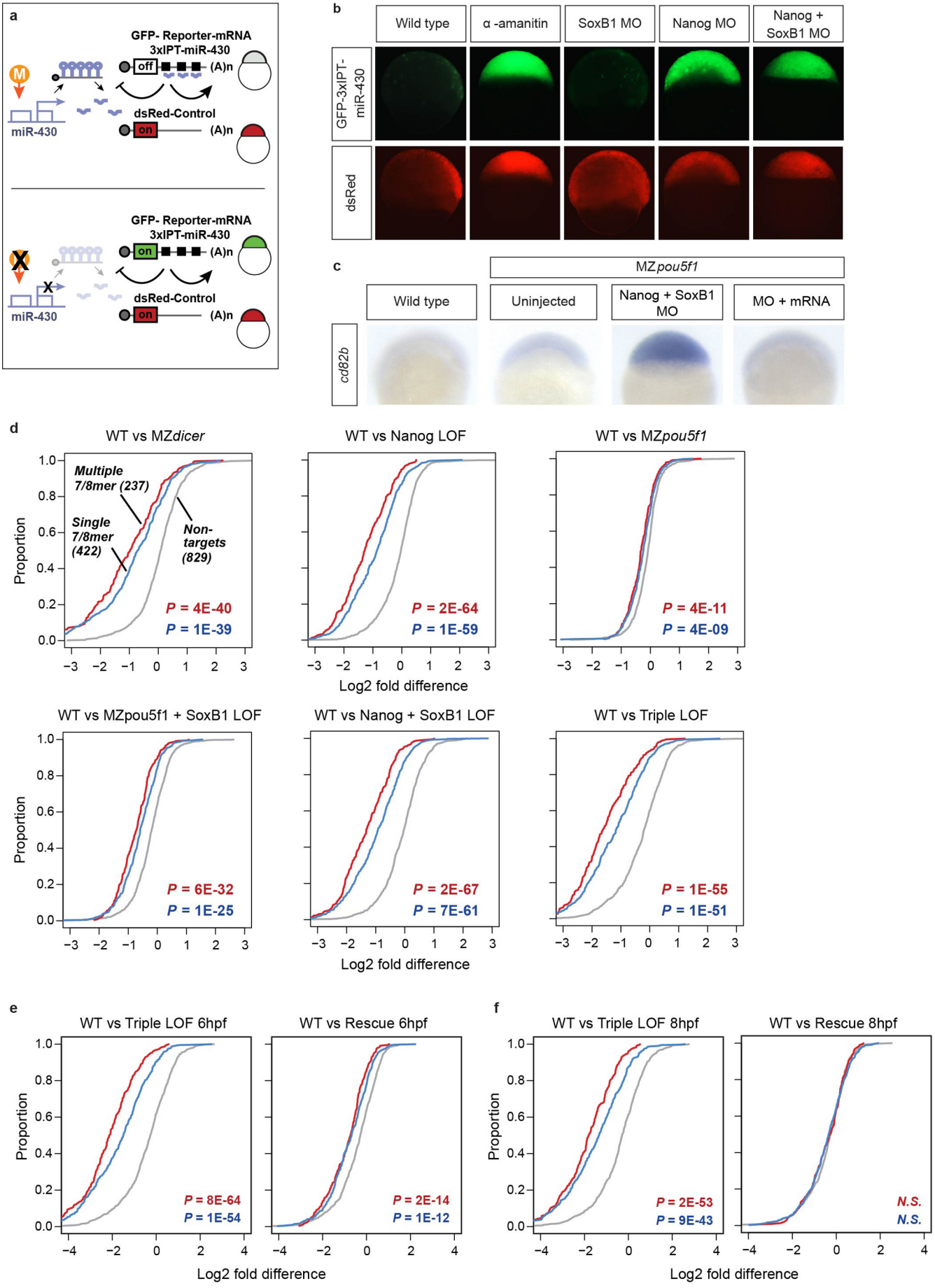
Extended Data Figure 5 | A transcriptome-wide effect is observed in LOF embryos. a, **b**, Biplots comparing $\log_2$ RPKM exonic expression levels between time-matched wild-type and Nanog + SoxB1 + Pou5f1 LOF embryos (**a**); and between wild-type and triple LOF embryos co-injected with mRNA for *nanog*, *soxB1* and *pou5f1* (**b**) at 4 h.p.f., 6 h.p.f. and 8 h.p.f. Dark blue points highlight all strictly zygotic genes, whereas magenta points highlight the first-wave zygotic genes. miR-430 is highlighted at 4 h.p.f. in red, whereas green points indicate expression levels of (left to right) *sox2*, *sox3*, *sox19a*, *sox19b* and *nanog*. **c**, Plots showing proportion of the zygotic transcriptome affected (including first and subsequent waves). For sphere and shield stages and each LOF (Nanog MO, Nanog MO + SoxB1 MO, MZ*pou5f1* + Nanog MO + SoxB1 MO), dark blue regions represent genes with normal expression compared to wild type; light blue regions represent genes with significant loss of expression. Inner ring comprises zygotic genes with <1 RPKM of maternal contribution; outer ring comprises zygotic genes with maternal contribution. Percentages represent total affected genes in that condition over both gene categories. At sphere stage (4 h.p.f.) the effect for maternal and zygotic (M+Z) genes is weaker than for strictly zygotic genes, which may reflect a reduced power to detect changes due to the maternal contribution (see also Fig. 3b).



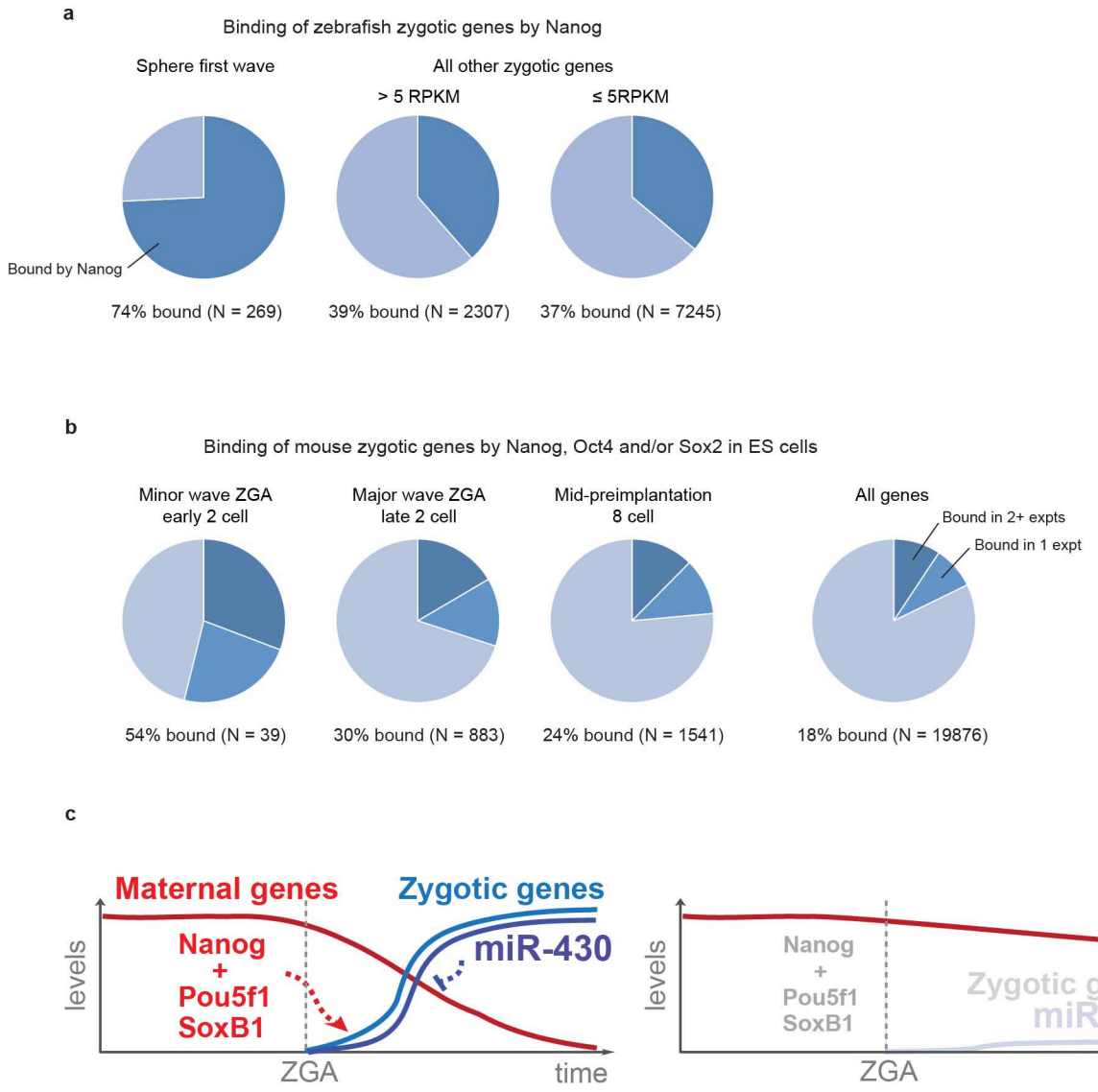
Extended Data Figure 6 | Zygotic genes fail to be activated with Nanog, SoxB1 and Pou5f1 LOF. **a**–**f**, *In situ* images showing that loss of Nanog and SoxB1 function results in a significant reduction in zygotic *foxa3*, *blf*, *vent*, *foxd3*, *krt18* and *ntla* expression. LOF embryos (Nanog + SoxB1 MO) resemble α-amanitin-injected embryos by *in situ*, as well as in their transcriptome profiles. Loss of Nanog and SoxB1 is rescued by *nanog* and *soxb1* mRNA (MO + mRNA), which is sufficient to restore wild-type expression profiles. **g**, **h**, *In situ* hybridization for zygotically transcribed *cldne* and *cebpb* shows that loss of Nanog and SoxB1 (Nanog + SoxB1 MO) has minimal effect on activation of *cldne* and *cebpb*. However, triple LOF shows a decrease in expression for both genes, as shown in the UCSC tracks. **i**–**o**, RT–PCR analysis (**i**) and UCSC Genome Browser tracks (**j**–**o**) for zygotic genes *klf4b*, *vox*, *tbx16*, *mxtx2*, *her3* and *sox32*, showing differential expression of zygotic genes in LOF conditions. Expression levels were rescued by injecting *nanog* and *soxb1* mRNA (MO + mRNA). Maternal *hist1h2aa* was present in the α-amanitin control. RT (−) indicates the absence of reverse transcriptase, to control for genomic DNA contamination. In UCSC tracks, loss of Nanog, SoxB1 and Pou5f1 in each sequenced condition is indicated by (−).



Extended Data Figure 7 | Loss of function affects genes across functional categories in a combinatorial manner. **a**, Comparisons of the single and double LOF transcriptomes to the triple LOF reveal that regulation is often combinatorial and redundant. Although all three factors seem to exert some influence on most of the transcribed genes, the effects observed in the combined LOF are not usually additive. Nanog seems to have the strongest individual effect of the three factors, but Pou5f1/SoxB1 can often act redundantly, or amplify the effect of Nanog alone. Venn diagrams show overlap between genes significantly downregulated at shield stage in single (pink), double (green) and triple (blue) LOF embryos. $n = 2{,}172$, left; $n = 2{,}027$, right. **b**, Pie charts showing the relative influence of each factor in the triple LOF. For each pie chart, genes downregulated in the triple LOF were compared in the single and double LOF transcriptomes. If the downregulation of a gene observed in the single LOF was less than twofold different from that observed in the triple LOF, the gene was considered to be regulated by the single factor alone. Otherwise, if the downregulation in the double was less than twofold different than the triple LOF, the gene was considered regulated by the combination of two factors. All remaining genes display the strongest downregulation in the triple LOF. Note that genes in each category may be affected by other combinations of LOF; however, the effect there is weaker. **c**, Breakdown of effects showing the redundancy of regulation in genes downregulated in the triple LOF. The largest category of genes seems to be regulated exclusively by Nanog (31%), as loss of Nanog function is equivalent to the triple LOF. 16% of genes seem to be regulated by both Nanog and Pou5f1 together, as loss of either Nanog alone or loss of Pou5f1 alone is sufficient to achieve the loss of function observed in the triple LOF. 16% of genes have equivalent effects with either Nanog LOF or Pou5f1 + SoxB1 double LOF, suggesting that Pou5f1 and SoxB1 act redundantly for these genes to co-regulate with Nanog. 9% of genes show the strongest effect only in the triple LOF. This suggests that there is redundancy between all three factors, as these genes can still be activated when one or two factors are lost. In all, 76% of the affected genes are subject to some form of redundant or combinatorial regulation. Asterisk indicates that for genes where the effect in the triple LOF was equivalent to both the double loss of SoxB1 and Nanog, and the double loss of SoxB1 and Pou5f1, we inferred that the effect was conferred by SoxB1 alone. **d**, Most genes are affected in the double or triple LOF conditions, across the gene categories defined in Extended Data Fig. 3a, including both embryo-specific genes and housekeeping (ubiquitously expressed) genes. **e**, Heat map showing specific embryonic functional categories of genes downregulated in LOF embryos. Three GO categories of genes expressed in wild type at shield stage are shown: general transcription factors, gastrulation and cell movement genes, and patterning genes (anterior–posterior axis and dorsal–ventral axis). Expression levels are represented as row-normalized values on a red–green colour scale for wild type (WT), α-amanitin treated (A), Nanog LOF (N), Nanog + SoxB1 LOF (NS), and Nanog + SoxB1 + Pou5f1 triple LOF (NSP). Widespread loss of expression is observed across these functional categories, with the triple LOF exhibiting the greatest similarity to α-amanitin.



Extended Data Figure 8 | miR-430 activity requires Nanog function. **a**, Schematic representation of miR-430 activity reporter GFP-3×IPT-miR-430 containing three complementary target sites to miR-430 (ref. 26). If maternal factor (M) is present, miR-430 is expressed and represses translation of the target mRNAs (no GFP expressed). Conversely, loss (X) of the maternal factor required for miR-430 activation would lead to a failure to repress miR-430 targets and GFP expression. dsRed is a control mRNA that is not subject to regulation by miR-430 and is co-injected with the target mRNA. **b**, GFP-reporter and dsRed (injection control) mRNAs were co-injected into embryos at one-cell stage and fluorescence assayed 7–8 h.p.f. GFP-reporter is repressed in wild-type and SoxB1 morphants by endogenous miR-430 (ref. 26), as shown by a decrease in GFP expression. The GFP-reporter fails to be repressed in α-amanitin (that fail to activate zygotic transcription and do not express miR-430) and Nanog-MO-injected embryos, indicating a loss of miR-430 activity. **c**, *In situ* hybridization for maternal miR-430 target gene *cd82b*. At shield stage, *cd82b* is cleared from wild-type and MZ*pou5f1* embryos. Combined Nanog, SoxB1 and Pou5f1 LOF causes a failure in clearance (MZ*pou5f1* + Nanog + SoxB1 MO). Injection of *nanog*, *soxb1* and *pou5f1* mRNA rescues the phenotype (MO + mRNA). **d**, Cumulative plots showing the effect of each LOF condition on miR-430 target repression, as in ref. 16, using Total RNA-seq. Plots show the distribution of $\log_2$ fold expression level difference for each condition relative to wild type in three groups of genes defined in ref. 16: miR-430 targets with multiple 7mer or 8mer seed target sites in their 3′ UTR; miR-430 targets with a single 7mer or 8mer seed in the 3′ UTR; and genes lacking miR-430 seed sites in their 3′ UTRs. *P* values are for two-sided Wilcoxon rank-sum tests comparing each of the two miR-430 target groups to the non-targets. MZ*dicer* expression data are from ref. 16. Displacement of the curve to the left (−) from the grey control line indicates a larger fraction of genes are accumulated (fail to be degraded) in the indicated condition compared to wild type. Nanog has the strongest effect, although there is also an effect from the combined loss of Pou5f1 and SoxB1. **e**, Cumulative plots showing the effect of triple LOF with and without mRNA rescue on miR-430 target repression, using poly(A)$^+$ selection RNA-seq. At 6 h.p.f., miR-430 targets fail to be degraded in the LOF condition compared to wild type, with expression levels of targets high in the LOF relative to wild type. Co-injection of *nanog*, *soxB1* and *pou5f1* mRNAs restores miR-430 activity, and the targets' expression levels are restored to near wild-type levels. **f**, At 8 h.p.f., miR-430 targets are still undegraded in the LOF, but are degraded to wild-type levels in the rescue. *P* values are for two-sided Wilcoxon rank-sum tests comparing each of the two miR-430 target groups to the non-targets.



Extended Data Figure 9 | Nanog, Pou5f1 and SoxB1 bind to and regulate embryonic genes. **a**, Nanog chromatin immunoprecipitation sequencing binding data in zebrafish at 3.3 h.p.f. (ref. 24) was re-analysed to determine Nanog-bound regions genome wide. Pie charts show percentage of genes in each category that are associated with Nanog bound regions (±5 kb). 74% of first-wave genes detected at sphere were associated with Nanog binding, twofold higher than subsequent-wave genes ($P = 3.7 \times 10^{-29}$, two-sided Fisher's exact test). Low expressed zygotic genes are also less associated with Nanog-bound regions. For those genes that are nonetheless affected by Nanog LOF, this suggests that they are influenced by Nanog indirectly, rather than through Nanog binding at the gene locus. The enrichment of Nanog binding on the first-wave genes versus subsequent waves supports a model where Nanog has a central role in the regulation of the activation of the first wave of zygotic transcription. **b**, ChIP-seq data for Nanog, Oct4 and Sox2 in mouse embryonic stem cells[57,58] were used to examine the binding profiles of genes transcribed during pre-implantation mouse embryogenesis[59], as ChIP data do not exist for early mouse embryos. Three gene groups were analysed: α-amanitin-sensitive genes expressed at early 2-cell stage (minor wave ZGA), α-amanitin sensitive genes expressed at late 2-cell stage (major wave ZGA), and genes expressed during the 4–8-cell stages (mid-preimplantation). Gene promoters (defined to be 5 kb upstream to 50 bp downstream the annotated transcription start site of a gene) are highly enriched in binding sites among the genes comprising ZGA, as compared to the genome as a whole ($P = 4.03 \times 10^{-7}$ for the minor wave, $P = 6.05 \times 10^{-18}$ major wave, two-sided Fisher's exact test). Genomic coordinates (mm8) for genes were defined by NIA/NIH U-cluster annotations for the microarray probes in ref. 59. Note that not all of the genes expressed during ZGA are necessarily expressed in ES cells; thus, the binding proportions are likely to be underestimates. Although these represent two different states of development, these results are consistent with a role for these factors in activating the earliest waves of zygotic gene expression also in mammals. **c**, Model showing maternal gene expression in red and zygotic gene expression in blue during the maternal to zygotic transition. Gene expression is depicted on the *y* axis and time on the *x* axis. During the MZT, Nanog, SoxB1 and Pou5f1 are required to activate a large fraction of zygotic genes, including miR-430, which in turn is responsible for the clearance of a significant portion of maternal mRNAs. In the loss of function of Nanog, SoxB1 and Pou5f1, there is a reduction in zygotic gene activation, causing a failure in the establishment of the zygotic developmental program, including loss of miR-430 expression and maternal mRNA clearance.

Extended Data Table 1 | Summary of Illumina sequencing data generated in this study

Sample*	Preparation	Age†	Genotype	Treatment	Total reads	rRNA	Aligned‡
1	input mRNA	2	WT	none	11,701,690	7,122,193	3,601,785
2	ribosome profiling	2	WT	none	35,324,638	28,557,085	4,782,034
3	input mRNA	2	WT	Nanog MO, SoxB1 MO	10,054,885	5,882,165	3,376,709
4	ribosome profiling	2	WT	Nanog MO, SoxB1 MO	37,708,163	28,354,953	5,946,384
5	RiboZero	2	WT	none	13,290,599	6,830,823	4,757,404
6	poly(A)+	4	WT	poly(A)+	21,504,328	NA	17,269,920
7	RiboZero	4	WT	none	49,104,024	29,072,153	16,633,109
8	RiboZero	4	WT	a-amanitin	43,280,984	17,159,279	22,541,771
9	RiboZero	4	WT	cycloheximide	60,496,090	13,980,195	40,960,186
10	RiboZero	4	WT	U1U2 MO	57,668,297	37,115,620	16,937,564
11	RiboZero	4	WT	Nanog MO	15,630,076	6,248,360	7,983,685
12	RiboZero	4	WT	SoxB1 MO	17,468,157	8,655,861	7,315,193
13	RiboZero	4	WT	Nanog MO, SoxB1 MO	13,583,155	6,597,214	5,853,123
14	RiboZero	4	MZpou5f1	none	116,396,185	90,173,314	20,274,383
15	RiboZero	4	MZpou5f1	Nanog MO	91,577,210	45,269,682	39,254,068
16	RiboZero	4	MZpou5f1	SoxB1 MO	47,420,118	32,192,741	12,137,699
17	RiboZero	4	MZpou5f1	Nanog MO, SoxB1 MO	42,220,676	28,214,894	11,452,962
18	RiboZero	4	MZpou5f1	Nanog MO, SoxB1 MO, rescue mRNA	63,785,933	22,119,249	36,078,935
19a	RiboZero	6	WT	none	14,503,666	5,448,147	7,487,251
19b	RiboZero	6	WT	none	15,074,846	8,338,535	5,303,876
19c	RiboZero	6	WT	none	17,682,683	8,153,773	7,806,485
20	poly(A)+	6	WT	none	22,626,103	NA	20,010,462
21	RiboZero	6	WT	a-amanitin	35,748,801	3,075,825	9,151,010
22	RiboZero	6	WT	cycloheximide	11,123,998	3,691,903	6,742,954
23	RiboZero	6	WT	Nanog MO	16,430,596	7,745,144	7,096,717
24a	RiboZero	6	WT	Nanog MO, SoxB1 MO	14,084,576	8,615,769	4,263,310
24b	RiboZero	6	WT	Nanog MO, SoxB1 MO	14,567,957	7,517,631	5,664,836
25	RiboZero	6	MZpou5f1	none	101,366,092	81,625,522	13,520,349
26	RiboZero	6	MZpou5f1	SoxB1 MO	13,616,658	5,839,148	6,383,511
27	RiboZero	6	MZpou5f1	Nanog MO, SoxB1 MO	28,543,110	13,273,679	12,670,402
28	poly(A)+	6	MZpou5f1	Nanog MO, SoxB1 MO	25,148,861	NA	22,263,359
29	poly(A)+	6	MZpou5f1	Nanog MO, SoxB1 MO, rescue mRNA	23,785,791	NA	21,033,046
30	poly(A)+	8	WT	none	23,504,890	NA	20,790,090
31	poly(A)+	8	MZpou5f1	Nanog MO, SoxB1 MO	25,758,851	NA	22,615,585
32	poly(A)+	8	MZpou5f1	Nanog MO, SoxB1 MO, rescue mRNA	23,475,791	NA	20,291,649

* All rows represent separately collected biological samples; that is, 19a, 19b and 19c, and 24a and 24b are biological replicates.
† Age in hours post fertilization.
‡ Reads aligning to the genome, minus rRNA-aligning reads where applicable.