

Hybridization led to a rewired pluripotency network in the allotetraploid *Xenopus laevis*

Wesley A Phelps, Matthew D Hurton, Taylor N Ayers, Anne E Carlson, Joel C Rosenbaum, Miler T Lee*

Department of Biological Sciences, University of Pittsburgh, Pittsburgh, United States

Abstract After fertilization, maternally contributed factors to the egg initiate the transition to pluripotency to give rise to embryonic stem cells, in large part by activating de novo transcription from the embryonic genome. Diverse mechanisms coordinate this transition across animals, suggesting that pervasive regulatory remodeling has shaped the earliest stages of development. Here, we show that maternal homologs of mammalian pluripotency reprogramming factors OCT4 and SOX2 divergently activate the two subgenomes of *Xenopus laevis*, an allotetraploid that arose from hybridization of two diploid species ~18 million years ago. Although most genes have been retained as two homeologous copies, we find that a majority of them undergo asymmetric activation in the early embryo. Chromatin accessibility profiling and CUT&RUN for modified histones and transcription factor binding reveal extensive differences in predicted enhancer architecture between the subgenomes, which likely arose through genomic disruptions as a consequence of allotetraploidy. However, comparison with diploid *X. tropicalis* and zebrafish shows broad conservation of embryonic gene expression levels when divergent homeolog contributions are combined, implying strong selection to maintain dosage in the core vertebrate pluripotency transcriptional program, amid genomic instability following hybridization.

*For correspondence:
miller@pitt.edu

Competing interest: The authors declare that no competing interests exist.

Funding: See page 18

Preprinted: 17 September 2022

Received: 04 October 2022

Accepted: 02 October 2023

Published: 03 October 2023

Reviewing Editor: Carole LaBonne, Northwestern University, United States

© Copyright Phelps et al. This article is distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.

Editor's evaluation

This paper reports fundamental findings that substantially advance our understanding of a major research question – how hybridization events influence gene regulatory programs and how evolutionary pressures have shaped these programs in response to such events. This convincing work uses appropriate and validated methodology in line with the current state-of-the-art.

Introduction

In animals, zygotic genome activation (ZGA) is triggered after an initial period of transcriptional quiescence following fertilization of the egg, during the maternal-to-zygotic transition (MZT; [Lee et al., 2014](#); [Vastenhouw et al., 2019](#)). In mammals, this occurs during the slow first cleavages ([Svoboda, 2018](#)), a few days removed from the subsequent induction of pluripotent stem cells in the blastocyst by a core network of factors including NANOG, OCT4, and SOX2 ([Li and Belmonte, 2017](#); [Takahashi and Yamanaka, 2016](#)). In contrast, faster-dividing taxa including zebrafish, *Xenopus*, and *Drosophila* activate their genomes in the blastula hours after fertilization ([Foe and Alberts, 1983](#); [Kane and Kimmel, 1993](#); [Newport and Kirschner, 1982a](#); [Vastenhouw et al., 2019](#)), which leads immediately to pluripotency. In zebrafish, maternally provided homologs of NANOG, OCT4, and SOX2 are required for a large share of genome activation ([Lee et al., 2013](#); [Leichsenring et al., 2013](#); [Miao](#)

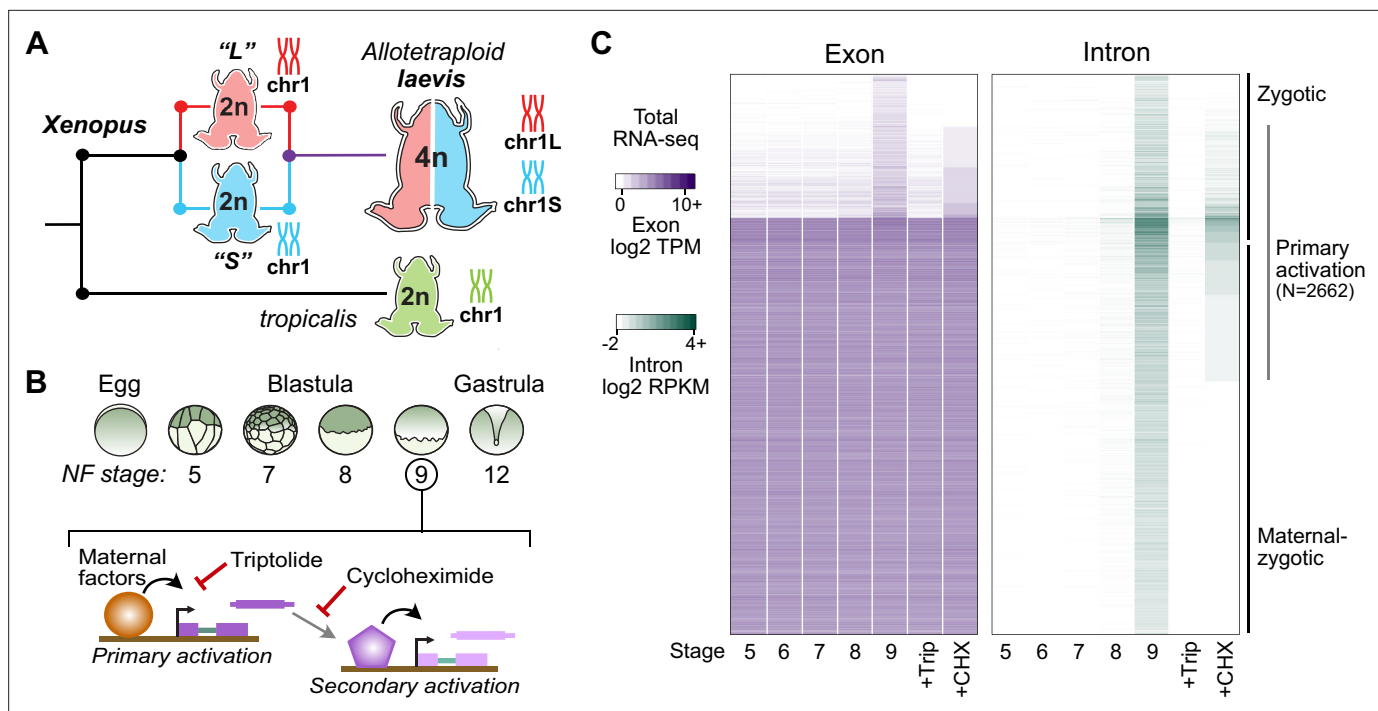


Figure 1. Identifying the first wave of genome activation across the two subgenomes. **(A)** The allotetraploid *X. laevis* genome contains two distinct subgenomes "L" and "S" due to interspecific hybridization of ancestral diploids. **(B)** Triptolide inhibits genome activation, as measured in the late blastula, while cycloheximide inhibits only secondary activation, distinguishing genes directly activated by maternal factors. NF = Nieuwkoop and Faber. **(C)** Heatmap of RNA-seq coverage over exons (left) and introns (right) of activated genes.

The online version of this article includes the following figure supplement(s) for figure 1:

Figure supplement 1. Measuring genome activation.

et al., 2022); thus, vertebrate embryos deploy conserved pluripotency induction mechanisms at different times during early development.

Beyond vertebrates, unrelated maternal factors direct genome activation and the induction of stem cells, for example Zelda (Liang et al., 2008), CLAMP (Colonna et al., 2021; Duan et al., 2021), and GAF (Gaskill et al., 2021) in *Drosophila*, although they seem to share many functional aspects with vertebrate pluripotency factors, including pioneering roles in opening repressed embryonic chromatin and establishing activating histone modifications (Blythe and Wieschaus, 2016; Gaskill et al., 2021; Hug et al., 2017). This diversity of strategies implies that the gene network regulating pluripotency has been extensively modified over evolutionary time (Endo et al., 2020; Fernandez-Tresguerres et al., 2010), though it is unknown when and under what circumstances major modifications arose.

We sought to understand how recent genome upheaval has affected the pluripotency regulatory network in the allotetraploid *Xenopus laevis*, by deciphering how embryonic genome activation is coordinated between its two subgenomes. *X. laevis*'s L (long) and S (short) subgenomes are inherited from each of two distinct species separated by ~34 million years that hybridized ~18 million years ago (Session et al., 2016; Figure 1A). A subsequent whole-genome duplication restored meiotic pairing. Despite extensive rearrangements and deletions, most genes are still encoded as two copies (homeologs) on parallel, non-inter-recombining chromosomes (Session et al., 2016). Previously, homeologs had been challenging to distinguish due to high functional and sequence similarity; however, the recent high-quality *X. laevis* genome assembly has made it feasible to resolve differential expression and regulation genome-wide between the two subgenomes (Elurbe et al., 2017; Session et al., 2016).

Allopolyploidy often provokes acute effects on gene expression (Hu and Wendel, 2019; Moran et al., 2021), leading to regulatory shifts over time to reconcile dosage imbalances and incompatibilities between gene copies (Grover et al., 2012; Song et al., 2020; Swamy et al., 2021). This phenomenon has been explored primarily in plants (Adams and Wendel, 2005; Husband et al.,

2013; Mable, 2004), but the extent to which this has occurred in the few characterized allopolyploid vertebrates is unclear (Chen et al., 2019a; Li et al., 2021; Luo et al., 2020). For *X. laevis*, there is a broad trend toward balanced homeolog expression across development and adult tissues, with a subtle bias in favor of the L homeolog that emerges after genome activation (Session et al., 2016), and an overall ontogenetic and transcriptomic trajectory similar to 48 million-years diverged diploid *X. tropicalis* (Harland and Grainger, 2011; Yanai et al., 2011). Initial observations in *X. laevis* have demonstrated differential homeologous enhancer activity in the eye (Ochi et al., 2017) and a divergent cis-regulatory landscape of histone modifications and recruitment of transcriptional machinery in the early gastrula (Elurbe et al., 2017), suggesting that embryonic genome activation is likely also asymmetric between the two subgenomes.

Although *Xenopus* embryos have long been a model for understanding the MZT, for example (Amodeo et al., 2015; Charney et al., 2017; Chen et al., 2019b; Gentsch et al., 2019; Gibeaux et al., 2018; Gurdon et al., 1958; Kimelman et al., 1987; Newport and Kirschner, 1982a; Newport and Kirschner, 1982b; Paraiso et al., 2019; Skirkanich et al., 2011; Veenstra et al., 1999; Yanai et al., 2011), ZGA regulators have not previously been identified in *X. laevis*. Here, we identify maternal Pou5f3 and Sox3 as top-level regulators of *X. laevis* pluripotency and ZGA and elucidate the predicted enhancer architecture that differentially recruits them to homeologous gene copies between the two subgenomes. Despite differential subgenome activation, combined transcriptional output converges to proportionally resemble the diploid state, maintaining gene dosage for the embryonic pluripotency program.

Results

Identifying divergently activated homeologous genes

At genome activation, the *X. laevis* pluripotency network consists of maternal regulators acting directly on the first embryonic genes (Figure 1B). To identify these genes, we performed a total RNA-seq early embryonic time course using our *X. laevis*-specific ribosomal RNA depletion protocol (Phelps et al., 2021; Figure 1A and B, Supplementary file 1). Subtle gene activation is observed in the blastula at Nieuwkoop and Faber (N.F.) stage 8, culminating in 4772 genes with significant activation by the middle of stage 9 (8 hours post fertilization [h.p.f.] at 23 °C) (Figure 1C, Supplementary file 2). Gene activation was detected through a combination of exon- and intron-overlapping sequencing reads deriving from nascent pre-mRNA (Lee et al., 2013) – indeed, two-thirds of these genes had substantial maternal contributions that masked their activation when quantifying exon-overlapping reads alone (Figure 1C). These genes fail to be activated in embryos treated at 1 cell (stage 1) with the transcription inhibitor triptolide (Gibeaux et al., 2018) when compared to DMSO vehicle control embryos (Figure 1B and C, Figure 1—figure supplement 1).

To distinguish direct targets of maternal factors (primary activation) (Figure 1B), we then performed RNA-seq on stage 9 embryos treated with cycloheximide at stage 8, to inhibit translation of newly synthesized embryonic transcription factors that could regulate secondary activation (Harvey et al., 2013; Lee et al., 2013). A total of 2662 genes (56% of all activated genes) were still significantly activated in cycloheximide-treated embryos compared to triptolide-treated embryos, representing the first wave of genome activation in the embryo (Figure 1C, Supplementary file 1A).

We analyzed subgenome of origin for activated genes and found that they are preferentially encoded as two homeologous copies in the genome ($p=2.3 \times 10^{-225}$, χ -squared test, 10 d.o.f.; Figure 2A). However, a majority of these genes have asymmetric expression between the two homeologs, often with transcription deriving from only the L or S copy alone (Figure 2B–C). This asymmetry is more pronounced at stage 8, but balances somewhat as genome activation progresses, suggesting timing differences for homeolog activation that could result from subtle regulatory divergence (Figure 2A, Figure 2—figure supplement 1A–C); and slightly less pronounced for strictly zygotic genes compared to maternal-zygotic genes (maternal contribution ≥ 1 TPM; Figure 2D, Figure 2—figure supplement 1D), which are often reactivated with different homeolog expression patterns compared to their maternal contribution (Figure 2E).

After genome activation, a heightened imbalance in favor of the L homeolog emerges, as measured by activation patterns in four differentiated cell lineages (Johnson et al., 2022; Figure 2D, Figure 2—figure supplement 1E–I), that appears to indicate a shift toward more divergent homeolog regulation

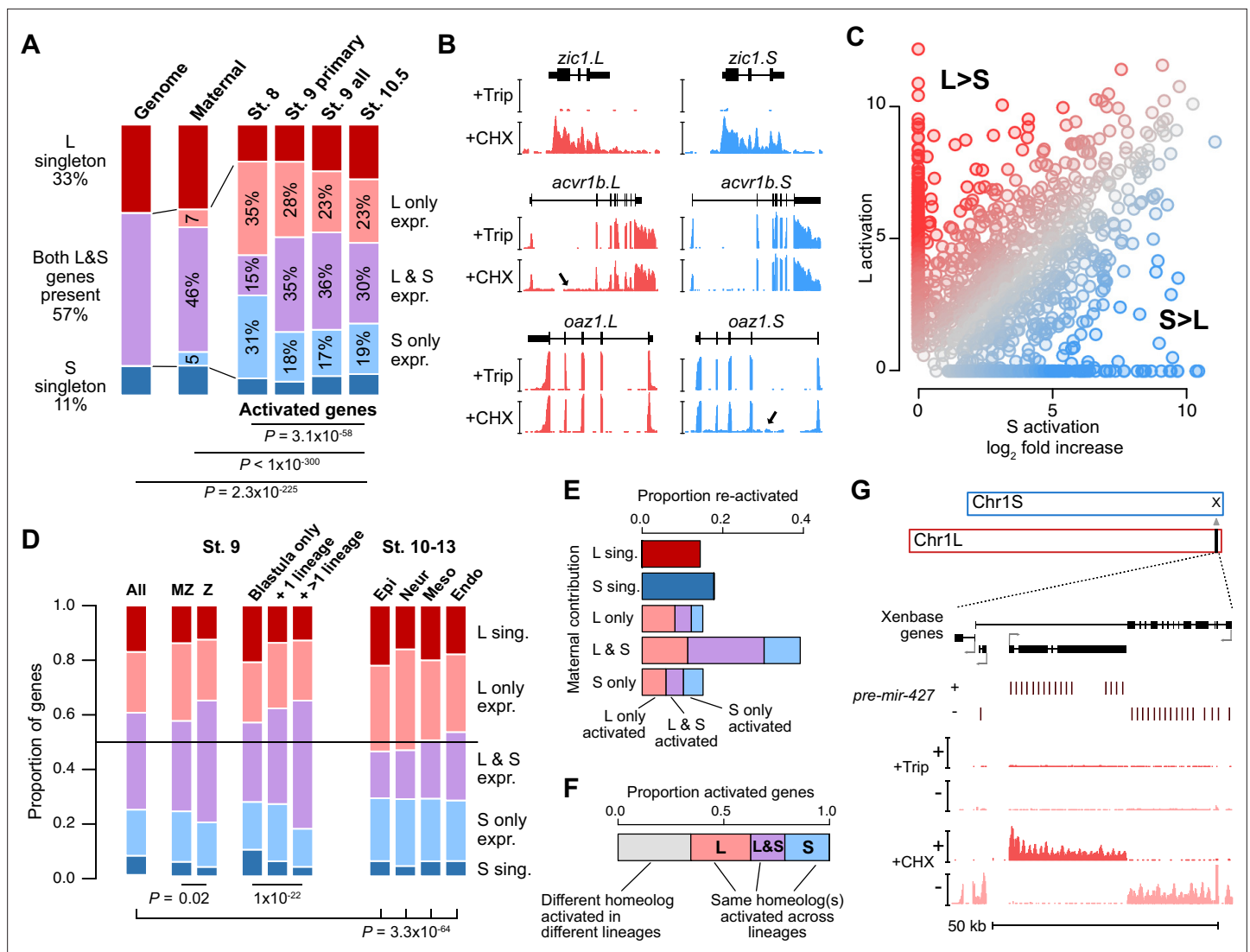


Figure 2. Homeologous genes are differentially activated in the early embryo. **(A)** Proportion of genes encoded as homeologs on both subgenomes versus only one subgenome (singleton) (left), as compared to expression patterns in the early embryo. p Values are from χ -squared tests, 10 d.o.f., comparing genomic to expressed proportions, 16 d.o.f., comparing proportions between activated genes and the maternal contribution, 12 d.o.f., comparing proportions at subsequent stages of activation. **(B)** Browser tracks showing log₂ reads-per-million RNA-seq coverage of equivalently activated homeologs (top) and differentially activated homeologs (L-specific, middle; S-specific, bottom). Trip = triptolide, CHX = cycloheximide. **(C)** Biplot comparing log₂ fold primary activation over triptolide treated embryos for the S homeolog (x axis) versus the L homeolog. **(D)** Left, proportion of genes activated symmetrically or asymmetrically from the L or S subgenomes, stratified into whether there is a maternal contribution for either homeolog (MZ) or not (Z) ($p=0.02$, χ -squared test, 4 d.o.f.); and whether a gene is activated only in the stage 9 blastula or is additionally increased in only one or more than one differentiated lineage from stages 10–13 ($p=1.3 \times 10^{-22}$, χ -squared test, 8 d.o.f.). Right, homeolog proportions of later gene activation in epidermal (Epi), neural progenitor (Neur), ventral mesodermal (Meso), and endodermal (Endo) lineages from stages 10–13 ($p=3.3 \times 10^{-64}$, χ -squared test comparing stage 9 and the four lineages, 16 d.o.f.). Lineage-specific gene expression data are from **Johnson et al., 2022**. **(E)** Homeolog-specific stage 9 activation proportions, versus maternal contribution homeolog expression patterns, for maternal-zygotic genes. **(F)** Concordance of homeolog activation patterns across the differentiated lineages at stages 10–13, for genes initially activated at stage 9 and also increased in at least two differentiated lineages. **(G)** Browser track showing strand-separated reads-per-million RNA-seq coverage over the *mir-427* encoding locus on the distal end of Chr1L (v10.1).

The online version of this article includes the following figure supplement(s) for figure 2:

Figure supplement 1. Differential homeolog activation over early development.

Figure supplement 2. The mir-427 locus.

as development proceeds, as was observed previously (Session et al., 2016). However, it is likely that some of the shared homeolog activation as measured in the whole blastula is actually composed of homeolog-specific regional activation (Chen and Good, 2022). Indeed, for one-third of genes activated in more than one lineage, different homeologs are activated in different lineages (Figure 2F, Figure 2—figure supplement 1F and G), and for those genes that are already activated at stage 9, this seems to result in a higher proportion of both-homeolog activation, as compared to genes with single-lineage or blastula-specific activation ($p=1.3 \times 10^{-22}$, χ -squared test, 8 d.o.f.). Overall, this indicates a high degree of divergent cis-regulatory architecture between gene homeologs throughout early development.

Genes activated from both subgenomes are enriched in transcriptional regulators ($p<0.01$, Fisher's exact test, two-sided) (Figure 2—figure supplement 1J), suggesting that gene function may have influenced homeolog expression patterns. However, there is no evidence for stronger functional divergence between homeologs expressed asymmetrically between the subgenomes, as estimated by non-synonymous versus synonymous mutation rate in coding regions (dN/dS ratio; Figure 2—figure supplement 1K and L).

The microRNA *mir-427* is encoded on only one subgenome

Among the first-wave genes is the microRNA *mir-427*, which plays a major role in clearance of maternally contributed mRNA (Lund et al., 2009). Similar to *X. tropicalis mir-427* (Owens et al., 2016) and the related zebrafish *mir-430* (Lee et al., 2013), *mir-427* is one of the most strongly activated genes in the *X. laevis* embryonic genome (Figure 2D, Figure 2—figure supplements 1C and 2A). In version 9.2 of the *X. laevis* genome assembly, the *miR-427* precursor hairpin sequence is found in only five copies overlapping a Xenbase-annotated long non-coding RNA on chr1L (Figure 2—figure supplement 2B). To better capture the genomic configuration of the *mir-427* primary transcript, we aligned the miRBase-annotated precursor sequence (Kozomara et al., 2019) to the version 10.1 *X. laevis* genome assembly. This revealed an expanded *mir-427* locus at the distal end of Chr1L composed of 33 precursor copies, encoded in both strand orientations over 55 kilobases (Figure 2D, Figure 2—figure supplement 2A and B). The corresponding region on Chr1S is unalignable (Figure 2—figure supplement 2C), suggesting that *mir-427* is encoded on only the L subgenome. We additionally found two *mir-427* hairpin sequence matches to the distal end of Chr3S, but these loci were not supported by substantial RNA-seq coverage (Figure 2—figure supplement 2D).

This is reminiscent of the *X. tropicalis mir-427* genomic configuration (Owens et al., 2016), although smaller in scale and on a non-homologous chromosome. In *X. tropicalis*, 171 tandemly arrayed *mir-427* precursors are found on the distal end of Chr03, which is thought to accelerate mature *miR-427* accumulation during the MZT to facilitate rapid maternal clearance (Owens et al., 2016). Zebrafish similarly encodes a large array of more than 2000 *mir-430* precursors, which begin to target maternal mRNA for clearance shortly after ZGA (Bazzini et al., 2012; Giraldez et al., 2006; Hadzhiev et al., 2023; Lee et al., 2013). These results strongly suggest that the *mir-427* locus has undergone genomic remodeling, resulting in absence from the S subgenome, but possibly also translocation between chromosomes in the *tropicalis* or *laevis* lineages.

Subgenomes differ in their regulatory architecture

To discover the maternal regulators of differential homeolog activation, we first profiled embryonic chromatin using Cleavage Under Target & Release Using Nuclease (CUT&RUN) (Hainer and Fazzio, 2019; Skene and Henikoff, 2017), which we adapted for blastulae. We found that cell dissociation was necessary for efficient nuclear isolation to carry out the on-bead CUT&RUN chemistry (Figure 3A, Figure 3—figure supplement 1A–D). At stages 8 and 9, the active marks H3 lysine 4 trimethylation (H3K4me3) and H3 lysine 27 acetylation (H3K27ac) were enriched in the transcription start site (TSS) regions of activated genes, and differential homeolog activation measured by RNA-seq significantly correlates with differential histone modification profiles, with a slight overall bias toward stronger L homeolog chromatin activity (Figure 3B and C, Figure 3—figure supplement 1E–G, Supplementary file 3). Differential promoter engagement by transcriptional machinery likely underlies the differential histone modification levels; however, we found no promoter sequence differences between homeologs that would implicate differential recruitment of specific transcription factors (Supplementary file 4).

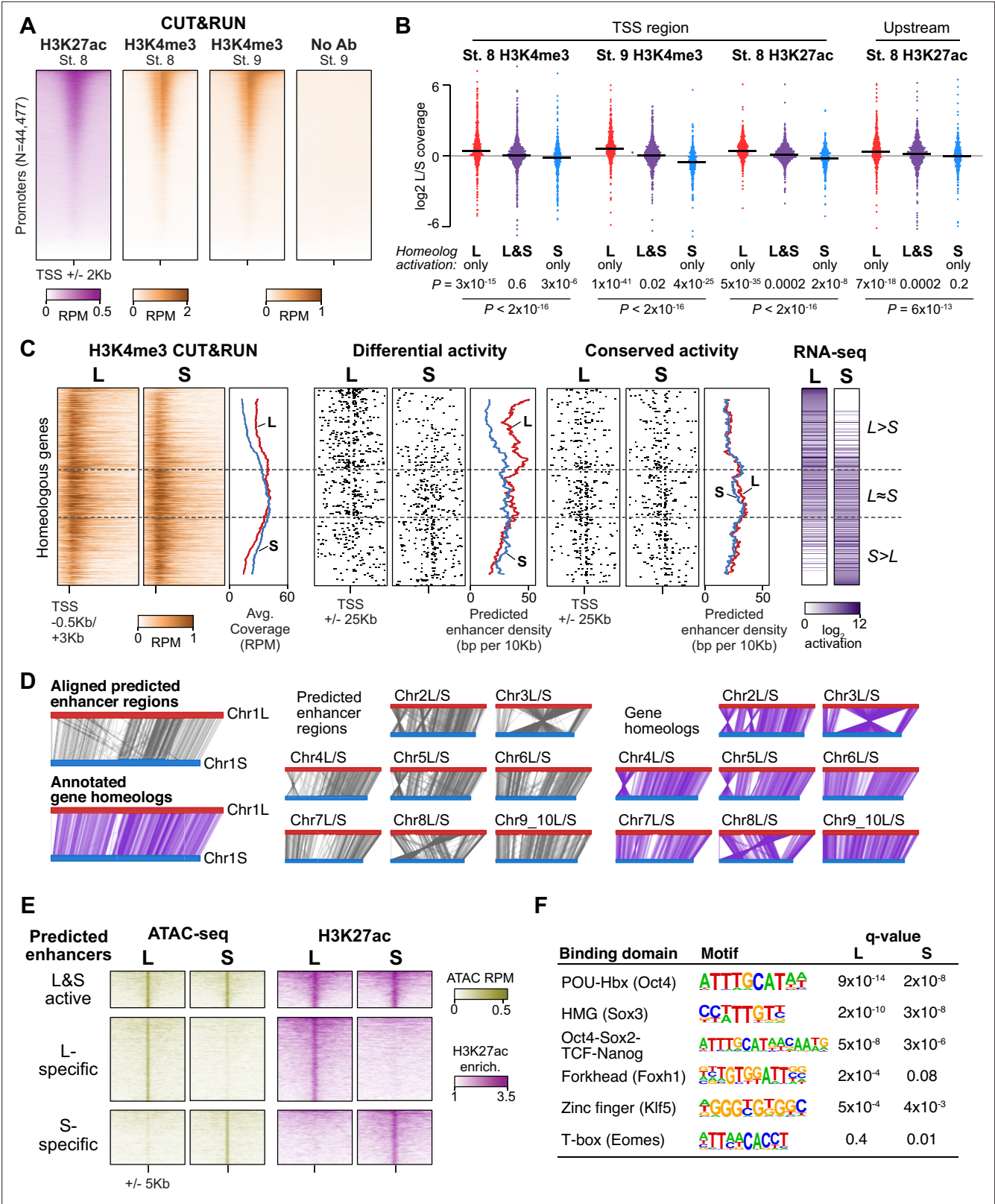


Figure 3. Differential homeolog activation is regulated by subgenome-specific enhancers. **(A)** CUT&RUN coverage over all annotated transcription-start site (TSS) regions, sorted by descending stage 8 H3K27ac signal. **(B)** Bee-swarm plots showing the log₂ ratio of L versus S homeolog coverage among genes where only one homeolog is activated (L only, S only), or both homeologs are activated. TSS region is 1 kb centered on the TSS; upstream region is 500 bp to 3 kb upstream of the TSS. Horizontal bars show medians. Individual category p values are from two-sided paired t-tests of log₂ L

Figure 3 continued on next page

Figure 3 continued

homeolog coverage vs log2 S homeolog coverage, p values across the three categories are from a one-way ANOVA on the log2 ratios. **(C)** Stage 9 H3K4me3 CUT&RUN coverage over paired homeologous gene regions around the TSS (left) and maps comparing high-confidence predicted enhancer density near homeologous TSSs (middle). Differential predicted enhancers are active in only one subgenome, conserved predicted enhancers are active in both. Average densities are plotted to the right of each paired map. Gene pairs are sorted according to L versus S subgenome RNA-seq activation ratio (right). **(D)** Schematics showing aligned predicted enhancers and their homeologous regions (gray) mapped onto L (red, top lines) and S (blue, bottom lines) chromosomes. Comparable schematics show Xenbase annotated homeologous gene pairs (lavender). **(E)** Heatmap of stage 9 ATAC-seq and stage 8 H3K27ac CUT&RUN over L & S homeologous regions for equivalently active high-confidence predicted enhancers (top) and subgenome-specific predicted enhancers. **(F)** Top enriched transcription factor motif families in L-specific and S-specific active high-confidence predicted enhancers compared to inactive homeologous regions. FDR-corrected p-values from Homer are shown. RPM = reads per million.

The online version of this article includes the following figure supplement(s) for figure 3:

Figure supplement 1. Profiling homeologous regulatory elements.

Figure supplement 2. Enhancer prediction and homeologous region comparison.

Instead, we searched for differences in gene-distal regulatory elements – that is enhancers – between the two subgenomes. To identify regions of open chromatin characteristic of enhancers, we performed Assays for Transposase-Accessible Chromatin with sequencing (ATAC-seq) on dissected animal cap explants; the high concentration of yolk in vegetal cells inhibits the Tn5 transposase (Esmaili et al., 2020). Accessible chromatin is already evident at stage 8 in putative enhancer regions, though the overall signal is weak, and by stage 9, these regions exhibit robust accessibility (Figure 3—figure supplement 2A). We called peaks of elevated sub-nucleosome sized fragment coverage at stage 9, then intersected the open regions with our H3K27ac CUT&RUN. This yielded 15,654 putative open and acetylated gene-distal regulatory regions at genome activation, of which we classified 5047 as high confidence predicted enhancers that had ≥ 2 fold signal enrichment in each of at least three H3K27ac replicates and three ATAC-seq replicates individually (Figure 3—figure supplement 2A, Supplementary file 5).

To identify homeologous L and S enhancer regions, we constructed a subgenome chromosome-chromosome alignment using LASTZ (Harris, 2007). This yielded a syntenic structure consistent with genetic maps (Figure 3D; Session et al., 2016), recapitulating the large inversions between chr3L/chr3S and chr8L/chr8S. Seventy-nine percent of predicted enhancer regions successfully lifted over to homeologous chromosomes, and of these, >92% of these are flanked by the same homeologous genes (Figure 3—figure supplement 2B), confirming local synteny.

Among the paired regions involving high-confidence predicted enhancers, only 21% had conserved activity in both homeologs, with the remaining pairs exhibiting differential H3K27ac and chromatin accessibility (Figure 3E, Figure 3—figure supplement 2C). Differential predicted enhancer density around genes significantly correlated with differential activation ($p=1.3 \times 10^{-16}$, Pearson's correlation test; Figure 3C, middle, Figure 3—figure supplement 2D), with greater L enhancer density around differentially activated L genes, and similarly for S enhancers and S genes. In contrast, conserved enhancers had equivalent density near both homeologs regardless of activation status ($p=0.67$, Pearson's correlation test; Figure 3C, right). Thus, differences in enhancer activity likely underlie divergent gene homeolog transcription at genome activation.

Maternal pluripotency factors differentially engage the subgenomes

Given that these paired enhancer regions are differentially active despite having similar base sequences, we searched for transcription factor binding motifs that distinguished active enhancers from their inactive homeolog. Two motifs were strongly enriched in both active L enhancers and active S enhancers, corresponding to the binding sequences of the pluripotency factors OCT4 and SOX2/3 (SOXB1 family; Figure 3F, Supplementary file 4). Since mammalian OCT4 and SOX2 are master regulators of pluripotent stem cell induction (Li and Belmonte, 2017; Takahashi and Yamanaka, 2016), and zebrafish homologs of these factors are maternally provided and required for embryonic genome activation (Lee et al., 2013; Leichsenring et al., 2013; Miao et al., 2022), we hypothesized that differential enhancer binding by maternal *X. laevis* OCT4 and SOXB1 homologs underlies asymmetric activation of the L and S subgenomes.

RNA-seq confirms high maternal levels of *sox3* and *pou5f3.3* (OCT4 homolog) mRNA, as well as lower levels of paralog *pou5f3.2*, each deriving from both subgenomes (Figure 3—figure supplement

2E). To assess their roles in genome activation, we inhibited their translation using previously validated antisense morpholinos (Morrison and Brickman, 2006; Takebayashi-Suzuki et al., 2007; Zhang et al., 2003) injected into stage 1 embryos. Combinations of *pou5f3.3+sox3* morpholinos and *pou5f3.2+pou5f3.3* morpholinos led to mild and severe gastrulation defects, respectively, while combining all three morpholinos led to developmental arrest with a complete failure to close the blastopore (Figure 4—figure supplement 1A), consistent with what has been reported in *X. tropicalis* (Gentsch et al., 2019).

RNA-seq of morpholino-treated embryos at stage 9 revealed extensive misregulation of genome activation, though only 15% of genes exhibited deficient activation, while 43% of genes actually exhibited slightly higher levels in the morphants (Figure 4A, Figure 4—figure supplement 1B and G–K), which could be due to direct or indirect transcriptional repression mediated by Pou5f3 and Sox3. Increases were predominantly detected from intron signal (Figure 4—figure supplement 1J), which would largely rule out post-transcriptional effects. A larger proportion of strictly zygotic genes were down-regulated in the morphants compared to maternal-to-zygotic genes ($p=6.6 \times 10^{-18}$, χ^2 -squared test, 3 d.o.f.), perhaps reflecting a more complex regulatory network that regulates maternal gene reactivation (Figure 4—figure supplement 1L).

To further clarify the regulatory network, we also performed morpholino treatments followed by cycloheximide treatment at stage 8, collecting at stage 9 for RNA-seq, to focus the loss of function on primary activation. In these embryos, nearly 70% of first-wave genes were down regulated, including the *mir-427* transcript (Figure 4A–C, Figure 4—figure supplement 1B–F, I), suggesting that maternal Pou5f3 and Sox3 directly activate a large proportion of first-wave genome activation, but newly synthesized zygotic factors rapidly mobilize to refine target gene expression levels (Figure 4H).

In the absence of wild-type Pou5f3 and Sox3 activity, divergent homeolog activation is reduced for a subset of genes, indicating that these factors at least partially underlie differential subgenome activation (Figure 4D–G). Among primary-activated genes, there does not seem to be a strong bias toward greater regulation of either homeolog; however, a significantly larger proportion of strictly zygotic genes encoded on both subgenomes is activated by Pou5f3 and Sox3 compared to singleton genes ($p=0.0020$, χ^2 -squared test, 5 d.o.f.; Figure 4—figure supplement 1M, N), which may reflect a reliance on these factors to mediate homeolog-specific expression when two copies exist.

To interrogate Pou5f3 and Sox3 chromatin binding across the subgenomes, we performed CUT&RUN on stage 8 embryos injected with mRNA encoding V5 epitope-tagged *pou5f3.3.L* and *sox3.S*. Peak calling revealed thousands of binding sites for each factor (Figure 4—figure supplement 2A–D), and Homer de novo motif analysis recovered the OCT4 and SOX3 binding sequences as top hits ($p=10^{-252}$ and $p=10^{-104}$, respectively; Figure 4I, Figure 4—figure supplement 2E and F). A subset of peaks have CUT&RUN enrichment for both factors, and at least 10% of peaks contain matches to the Oct4-Sox2 heterodimer motif (Figure 4—figure supplement 2F), suggesting Pou5f3 and Sox3 may form a complex in the blastula, similar to mammalian OCT4 and SOX2 (Boyer et al., 2005; Dailey and Basilico, 2001). CUT&RUN signal for both factors is enriched in the vicinity of genes down-regulated in morphants with or without cycloheximide treatment, but notably not for genes up-regulated ($p<1 \times 10^{-43}$, Kruskal-Wallis test; Figure 4I and J), confirming that up regulation is likely an indirect effect of Pou5f3/Sox3 loss of function.

Down-regulated genes are highly significantly nearer to predicted regulatory elements with enriched Pou5f3 and Sox3 binding ($p<1 \times 10^{-300}$, Kruskal-Wallis test; Figure 4K, Figure 4—figure supplement 2G). Differential Pou5f3 and Sox3 binding mirrors differential predicted enhancer activity (Figure 4—figure supplement 2I), and subgenome-specific Pou5f3 and Sox3 binding is enriched in the vicinity of the homeolog affected by Pou5f3/Sox3 loss of function ($p=7.9 \times 10^{-13}$, Kruskal-Wallis test; Figure 4L, Figure 4—figure supplement 2H). Together, these results implicate Pou5f3.3 and Sox3 in regulating ZGA differentially between the two subgenomes.

The ancestral pluripotency program is maintained, despite enhancer turnover

Finally, to understand differential activation given the natural history of *X. laevis* allotetraploidy, we compared *X. laevis* subgenome activation patterns to diploid *X. tropicalis* as a proxy for the ancestral *Xenopus*, since there are no known extant diploid descendants of either *X. laevis* progenitor (Session et al., 2016). For three-way homeologs/orthologs that are strictly zygotic in *X. laevis*, there is broad

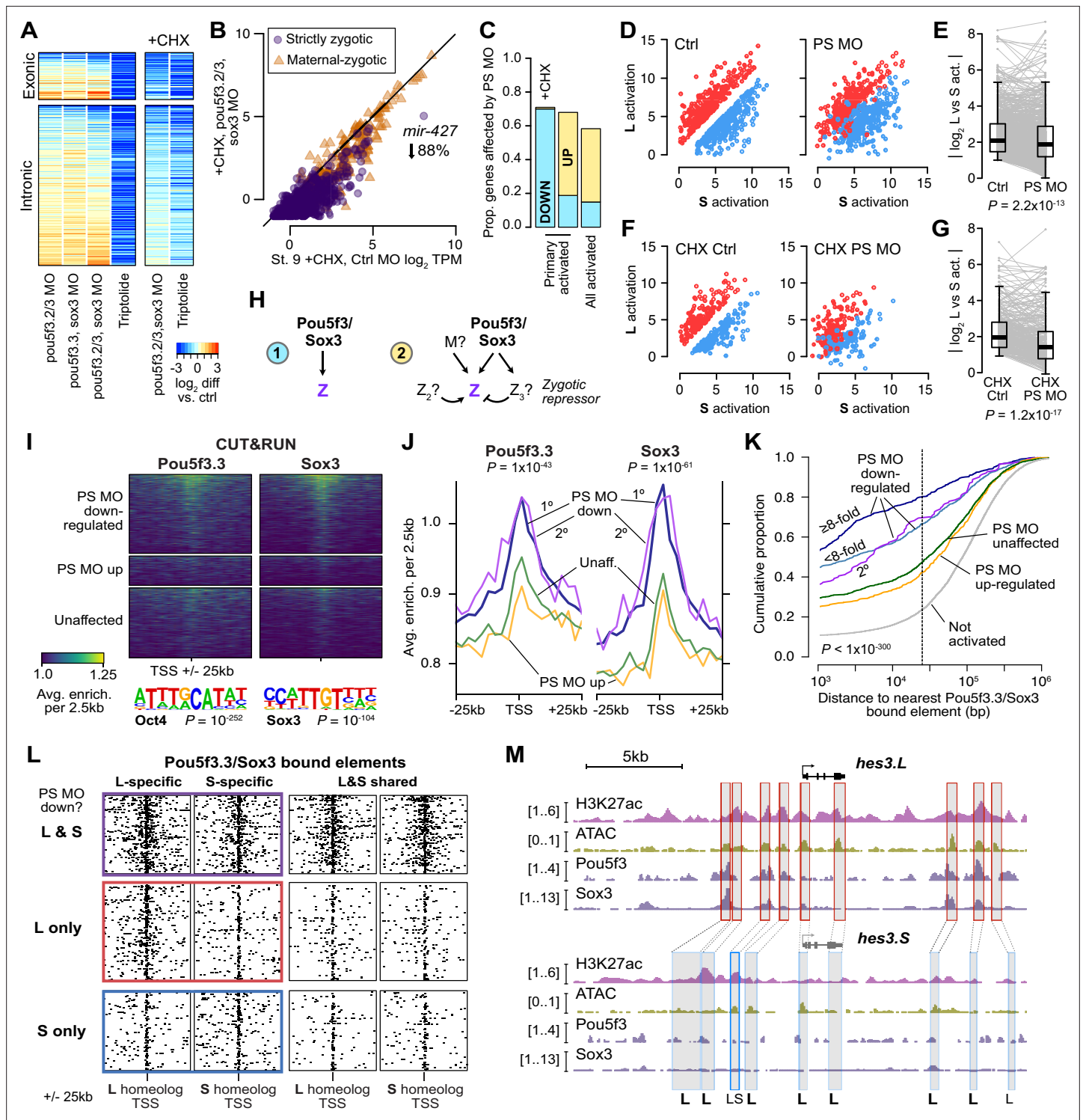


Figure 4. Pou5f3.3 and Sox3 binding drives genome activation. (A) Heatmap showing log₂ fold activation differences for exonic and intronic regions of primary-activated genes for combinations of *pou5f3.2*, *pou5f3.3*, and *sox3* morpholino-treatments, or Triptolide treatment, compared to controls. Right panel is in the presence of cycloheximide (CHX). (B) Biplot comparing exonic expression levels in cycloheximide-treated control embryos versus embryos also injected with *pou5f3.2*, *pou5f3.3*, and *sox3* morpholinos. Primary-activated genes with maternal contribution <1 TPM (strictly zygotically) are purple circles, maternal-zygotically genes detected by exonic increases are orange triangles. TPM = transcripts per million. (C) Barplot summarizing the proportion of genes affected by morpholino treatment with cycloheximide on primary-activated genes (left bar), without cycloheximide (middle bar), and all stage 9 activated genes without cycloheximide (right bar). Down = significantly decreased in one of the morpholino treatments, up = significantly increased. (D, F) Biplots showing genes with >2 fold L or S biased activation (upper red and lower blue points, respectively) in control

Figure 4 continued on next page

Figure 4 continued

embryos (left panel) versus their activation in *pou5f3.2*, *pou5f3.3*, and *sox3* morpholino-treated embryos (right panel, maintaining the same color per gene). (E, G) Quantification of the biplots in (D, F) in before-and-after plots. Y-axis is the absolute value of the log₂ L vs S activation difference. p Values are from Wilcoxon signed-rank tests (paired). Overlaid boxplots show median, upper and lower quartiles, and 1.5 x interquartile range. (H) Regulatory networks consistent with direct regulation of embryonic gene activation by Pou5f3 and Sox3 (1) versus additional regulation by zygotic factors (2), which likely accounts for genes up-regulated in MO treatments. (I) Stage 8 Pou5f3.3 (left) and Sox3 (right) CUT&RUN coverage near TSSs for genes down-regulated in morpholino-treated embryos with or without cycloheximide (top), genes up-regulated (middle), and genes not significantly affected in any morpholino treatment (bottom). Top enriched motifs for each factor are shown below with p-values from Homer de novo discovery. (J) Aggregate plots of the binding signal in (I), with down-regulated genes further separated into genes down-regulated with morpholino treatment and cycloheximide (1°) or only down-regulated without cycloheximide (2°). p Values are from Kruskal-Wallis tests on summed signal per TSS. (K) Cumulative distributions of distance from a Pou5f3/Sox3-bound regulatory element for genes strongly (≥ 8 fold) and less strongly (< 8 fold) down-regulated in morpholino-treated embryos with or without cycloheximide, compared to up-regulated, unaffected and unactivated genes. p Value is from a Kruskal-Wallis test. (L) Maps showing density of Pou5f3/Sox3-bound regulatory elements around paired homeologous TSSs, divided into elements with differential homeologous L & S binding (left panels) versus both bound (right panels). TSSs are grouped according to L versus S homeolog sensitivity to morpholino treatment. (M) Browser tracks showing CUT&RUN enrichment and ATAC-seq coverage near active homeolog *hes3.L* and inactive homeolog *hes3.S*. Seven L-specific high-confidence regulatory regions are highlighted with their homeologous S regions (bold 'L'), as well as two lower-confidence enhancers, one of which also has weak activity in S, but minimal Pou5f3 or Sox3 binding (labeled 'LS').

The online version of this article includes the following figure supplement(s) for figure 4:

Figure supplement 1. Assessing Pou5f3 and Sox3 roles in genome activation.

Figure supplement 2. Pou5f3 and Sox3 CUT&RUN.

conservation of relative expression levels between the *X. tropicalis* and *X. laevis* embryonic transcriptomes after genome activation, when *X. laevis* homeolog levels are summed gene-wise (Spearman's $\rho=0.67$) (Figure 5A, left, Figure 5—figure supplement 1A–C, Supplementary file 6). However, the correlation weakens when the *X. laevis* subgenomes are considered independently: relative activation levels in one subgenome alone are depressed relative to *X. tropicalis*, with expression of some genes completely restricted to one subgenome or the other (L, Spearman's $\rho=0.56$; S, $\rho=0.47$; Figure 5A, middle, right, Figure 5—figure supplement 1A–C). Indeed, the *X. tropicalis* embryonic transcriptome is a better estimator for the composite *X. laevis* transcriptome than for either subtranscriptome individually ($p < 4.3 \times 10^{-13}$ for strictly zygotic genes, $p < 1.4 \times 10^{-83}$ for maternal-zygotic genes, Wilcoxon signed-rank test on residuals; Figure 5—figure supplement 1B). If the diploid L and S progenitor embryos each exhibited the inferred ancestral activation levels, then these trends strongly suggest that *X. laevis* underwent regulatory changes post allotetraploidization that maintained relative gene expression dosage for embryonic genome activation.

Most activated genes also have a maternal contribution, which can offset asymmetries in homeolog activation levels (Figure 5—figure supplement 1C); and indeed, most *X. laevis* genes without conserved activation in *X. tropicalis* nonetheless have conserved embryonic expression due to the maternal contribution (Figure 5B). When we specifically compare zygotic activation between the two species, genes activated from both *X. laevis* homeologs are more likely to have orthologous *X. tropicalis* activation ($p = 1.7 \times 10^{-44}$, χ^2 -squared test, 4 d.o.f.; Figure 5B), as well as conserved *X. tropicalis* Pou5f3/Sox3 regulation ($p < 1.7 \times 10^{-6}$, χ^2 -squared test, 8 d.o.f.; Figure 5—figure supplement 1D; Gentsch et al., 2019). This suggests that many differentially activated homeologs have acquired novel ZGA regulation by Pou5f3 and Sox3 compared to the inferred ancestral state. Predicted enhancers also follow this trend: subgenome-specific predicted enhancers are less likely to be conserved with *X. tropicalis* ($p = 1.5 \times 10^{-76}$, χ^2 -squared test for high confidence enhancers, 4 d.o.f.; Figure 5C, Figure 5—figure supplement 1E), consistent with a greater degree of regulatory innovation underlying differentially activated homeologs.

This trend is also apparent at greater evolutionary distances. We find that genes activated in *X. laevis* are largely also expressed in zebrafish embryos (~450 million years separated) (Figure 5—figure supplement 1F). Despite considerable divergence in activation timing, co-activated *X. laevis* homeologs are still more likely to be part of the first wave of zebrafish genome activation ($p = 4.3 \times 10^{-12}$, χ^2 -squared test, 4 d.o.f.) and targeted by zebrafish maternal homologs of OCT4 and SOX2, but also NANOG ($p = 1.8 \times 10^{-45}$, χ^2 -squared test, 6 d.o.f.) (Figure 5D, Figure 5—figure supplement 1G). Subgenome-shared predicted enhancers are also more likely to have evidence for conservation in zebrafish ($p = 2.8 \times 10^{-58}$ for all enhancers, $p = 5.0 \times 10^{-4}$ for high-confidence enhancers, χ^2 -squared tests, 4 d.o.f.) (Figure 5—figure supplement 1H; Bogdanovic et al., 2012). Taken together, this

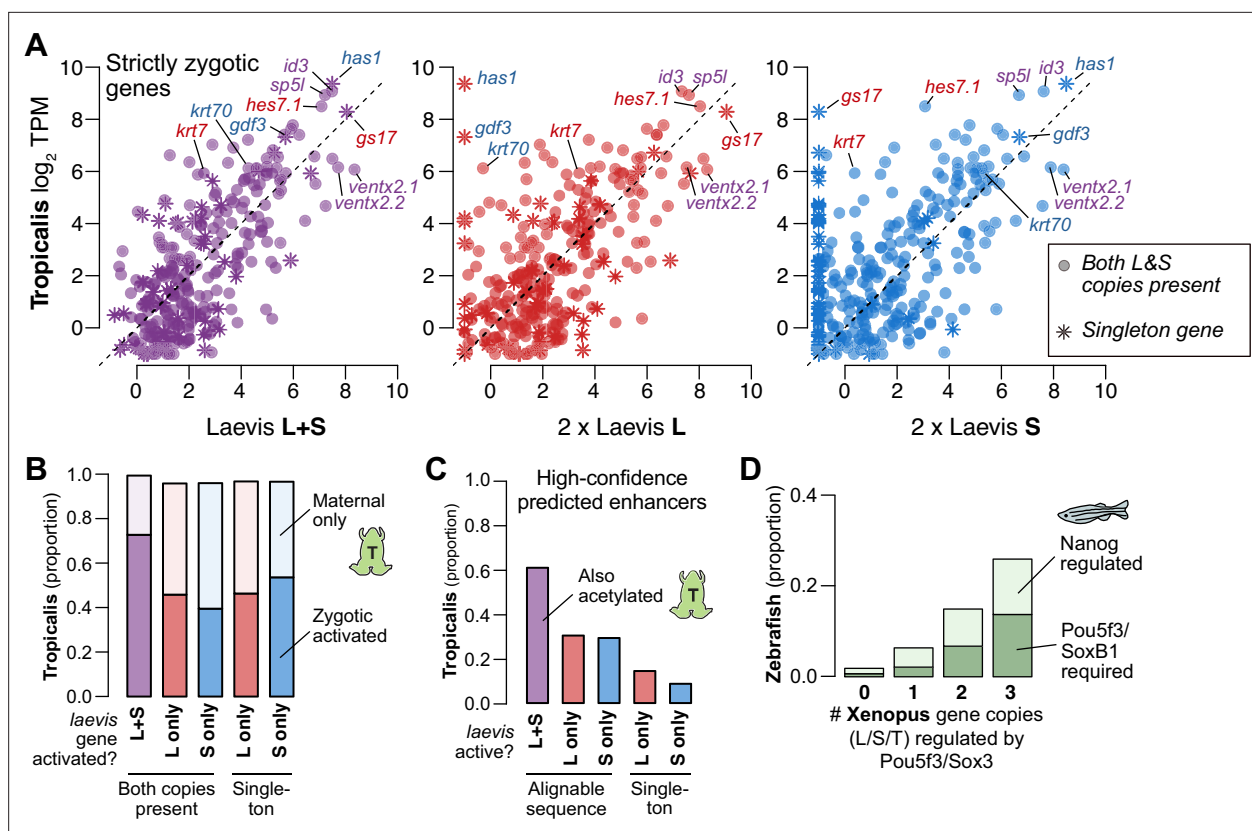


Figure 5. Regulatory divergence underlies dosage maintenance. **(A)** Biplots comparing relative expression levels of activated genes in *X. laevis* and *X. tropicalis*, treating L and S homeolog contributions separately (middle, right) or summed (left). Individual subgenome expression is scaled 2 x, since transcript per million (TPM) normalization is calculated relative to the entire *X. laevis* transcriptome. Individual labeled genes are color coded according to the dominant expressed homeolog (red = L, blue = S, purple = equivalent). **(B)** Barplots showing the proportion of *X. laevis* genes across homeolog activation categories whose orthologs are also activated in *X. tropicalis* or part of the maternal contribution. **(C)** Barplots showing the proportion of *X. laevis* enhancers across homeolog activity categories that are acetylated in *X. tropicalis*. **(D)** Barplots showing the proportion of *Xenopus* genes whose orthologs are regulated by Pou5f3/SoxB1 and Nanog in zebrafish. *Xenopus* genes are classified according to how many homeo/orthologs are regulated by Pou5f3/Sox3. Genes with conserved regulation in both *X. laevis* homeologs and *X. tropicalis* are more likely to be regulated by Pou5f3/SoxB1 in zebrafish, but also more likely to be regulated by Nanog.

The online version of this article includes the following figure supplement(s) for figure 5:

Figure supplement 1. Shared patterns of activation with other taxa.

suggests that the regulatory architecture underlying differential homeolog activation in *X. laevis* is more likely to be derived, in contrast to the deeply conserved networks that regulate many co-activated homeologs.

Interestingly, *Xenopus* and possibly all Anuran amphibians lack a NANOG ortholog, likely due to a chromosomal deletion (Schuff et al., 2012). In the absence of a Nanog homolog in the maternal contribution, we find that maternal Pou5f3.3 and Sox3 seem to have subsumed NANOG's roles in *X. laevis* genome activation, while zygotic factors such as Ventx help promote cell potency in the early gastrula (Scerbo et al., 2012; Schuff et al., 2012). This demonstrates core-vertebrate mechanistic conservation in genome activation amid both cis- and trans-regulatory shuffling, which converge to support pluripotent stem cell induction and embryonic development.

Discussion

Together, our findings establish the pluripotency factors Pou5f3.3 and Sox3 as maternal activators of embryonic genome activation, which are differentially recruited to the two homeologous subgenomes of *X. laevis* by a rewired enhancer network (Figure 6). Of the thousands of genes activated during the MZT, a majority of annotated homeolog pairs experience differential activation, which appears to be

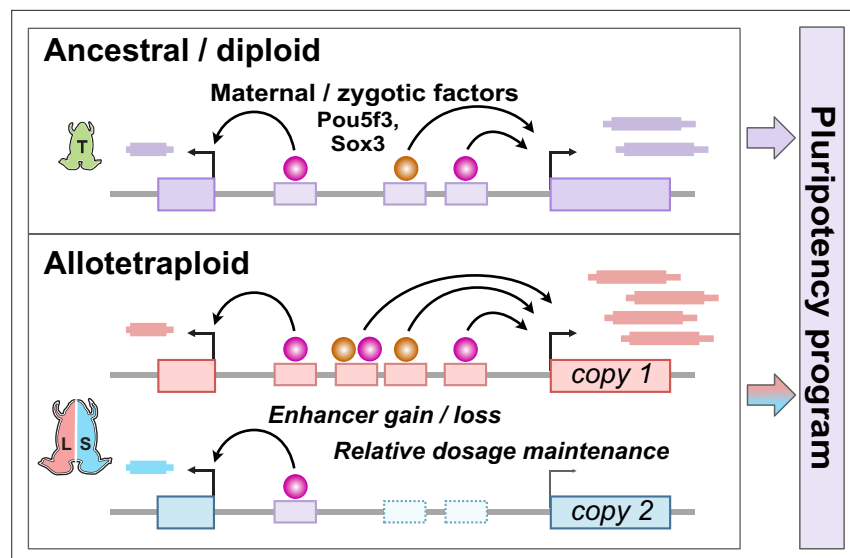


Figure 6. Model for pluripotency network evolution. *X. laevis* likely underwent extensive enhancer turnover between its two subgenomes, which nonetheless maintained stoichiometry of pluripotency reprogramming in the early embryo.

driven by subgenome-specific enhancer gain and/or loss correlated with differential Pou5f3.3/Sox3 binding and regulation. However, this magnitude of regulatory divergence seems to have had a net neutral effect, as combined subgenome activation produces a composite reprogrammed embryonic transcriptome akin to diploid *X. tropicalis*.

As embryogenesis proceeds, regulatory divergence between the subgenomes is likely even broader. In *X. tropicalis*, signal transducers and transcription factors including Pou5f3.2/3, Sox3, Smad1/2, β -catenin, Vegt, Otx1, and Foxh1 regulate embryo-wide and regional gene activation (Charney et al., 2017; Gentsch et al., 2019; Paraiso et al., 2019), and binding motifs for some of these are found in differentially active *X. laevis* enhancers (Figure 3F, Supplementary file 4). Additionally, by focusing on accessible chromatin in animal caps, we may have underestimated the magnitude of homeologous enhancer divergence regulating endodermal fate in the vegetal cells. But based on the close morphological similarity of *X. tropicalis* and *X. laevis* embryos, we would predict that these subgenome regulatory differences also converge to producing ancestral dosages in the transcriptome.

Although homeolog expression bias can derive from gene regulatory differences evolved in the parental species prior to hybridization (Buggs et al., 2014; Grover et al., 2012), we propose that regulatory upheaval in *X. laevis* post-hybridization (i.e. ‘genome shock’ McClintock, 1984) led to expression level gain or loss in one homeolog, which was subsequently corrected by compensatory changes to the other homeolog, possibly repeatedly (Shi et al., 2012; Tirosh et al., 2009). This implies that early development exerts constraint on the reprogrammed embryonic transcriptome while tolerating (or facilitating) regulatory turnover. The apparent reconfiguration of the *mir-427* cluster after the *X. laevis* and *tropicalis* lineages split similarly highlights how essential MZT regulatory mechanisms can evolve, ostensibly neutrally given that *mir-427*-directed maternal clearance is conserved in *Xenopus*. Thus, *X. laevis* embryos illustrate how the pluripotency program may have accommodated regulatory network disruptions, genomic instability, and aneuploidy across the animal tree.

Methods

Animal husbandry

All animal procedures were conducted under the supervision and approval of the Institutional Animal Care and Use Committee at the University of Pittsburgh under protocol #21120500. *Xenopus laevis* adults (Research Resource Identifier NXR_0.0031; NASCO) were housed in a recirculating aquatic system (Aquaneering) at 18 °C with a 12/12 hr light/dark cycle. Frogs were fed 3 x weekly with Frog Brittle (NASCO #SA05960 (LM)M).

Embryo collection

Sexually mature females were injected with 1000 IU human chorionic gonadotropin into their dorsal lymph sac and incubated overnight at 16 °C. Females were moved to room temperature to lay. Eggs from two mothers per collection were pooled and artificially inseminated using dissected testes in MR/3 (33 mM NaCl, 0.6 mM KCl, 0.67 mM CaCl₂, 0.33 mM MgCl₂, 1.67 mM HEPES, pH 7.8; **Sive and Richard, 2000**). Dissected testes were stored up to one week in L-15 medium at 4 °C prior to use. Zygotes were de-jellied (**Sive and Richard, 2000**) in MR/3 pH 8.5, with 0.3% β-mercaptoethanol with gentle manual agitation, neutralized with MR/3 pH 6.5, washed twice with MR/3 and incubated in MR/3 at 23 °C until desired developmental stage based on morphology, for genomics experiments.

RNA-seq libraries

All stage 9 embryos were collected halfway through the stage, at 8 hours post fertilization ('stage 9.5'). Stage 10.5 embryo libraries were spiked with GFP, mCherry, and luciferase in vitro transcribed RNA for an unrelated purpose. Triptolide samples were bathed in 20 μM triptolide in DMSO (200 X stock added to MR/3) at stage 1 and cycloheximide samples were bathed in 500 μg/mL cycloheximide in DMSO at the beginning of stage 8; both were collected when batch-matched, untreated embryos were halfway through stage 9. Equivalent volumes of DMSO were used to treat control samples. Previously validated morpholinos targeting *pou5f3.2* (AGGGCTGTTGGCTGTACATGGTGTC) (**Takebayashi-Suzuki et al., 2007**) *pou5f3.3* (GTACAATATGGGCTGGTCCATCTCC) (**Morrison and Brickman, 2006**) and *sox3* (AACATGCTATACATTGGAGCTTCA) (**Zhang et al., 2003**), along with control GFP morpholino (ACAGCTCCTCGCCCTTGCTCACCAT) were ordered from GeneTools. Morpholino treated embryos were injected at stage 1 with *pou5f3.3*, *sox3*, and/or GFP control morpholino. Non-cycloheximide treated embryos all received 120 ng total morpholino, consisting of 40 ng of each target morpholino augmented with 40 ng of GFP morpholino for two-morpholino conditions. Cycloheximide-treated embryos received 40 ng of each target morpholino for the triple condition, 40 ng *pou5f3.3*+40 ng GFP, 40 ng *sox3* +40 ng GFP, 40 ng *pou5f3.3*+40 ng *sox3*, or 80 ng GFP. An additional cycloheximide-treated high concentration morpholino condition used 55 ng *pou5f3.3*+75 ng *sox3*. Each embryo was injected twice with 5 nl of MO on opposite sides. Embryos were allowed to recover to stage 5 before moving to MR/3 to develop, and collected when batch-matched, untreated embryos were halfway through stage 9. Samples from the 'H' batch likely had an issue with the cycloheximide treatment, based on greater similarity of gene expression to untreated samples than other cycloheximide-treated samples, and were removed from subsequent analyses.

For phenotype observation, embryos were incubated at 23 °C or 18 °C after injection and photographed when control embryos reached stage 10.5 for 23 °C and stage 12 for 18 °C. For RNA extraction, two embryos per sample were snap frozen and homogenized in 500 μl of TRIzol Reagent (Invitrogen #15596026) followed by 100 μl of chloroform. Tubes were spun at 18,000 x g at 4 °C for 15 min, the aqueous phase was transferred to a fresh tube with 340 μl of isopropanol and 1 μl of GlycoBlue (Invitrogen #AM9515), then precipitated at -20 °C overnight. Precipitated RNA was washed with cold 75% ethanol and resuspended in 50 μl of nuclease-free water. Concentration was determined by NanoDrop.

For library construction, rRNA depletion was performed as per **Phelps et al., 2021** with *X. laevis* specific oligos reported previously: 1 μl of antisense nuclear rRNA oligos and 1 μl of antisense mitochondrial rRNA oligos (final concentration 0.1 μM per oligo) were combined with 1 μg of total RNA in a 10 μl buffered reaction volume (100 mM Tris-HCl pH 7.4, 200 mM NaCl, 10 mM DTT), heated at 95 °C for 2 minutes and cooled to 22 °C at a rate of 0.1 °C/s in a thermocycler. Next, 10 U of thermostable RNaseH (NEB #M0523S) and 2 μl of provided 10 X RNaseH buffer were added and volume brought to 20 μl with nuclease-free water. The reaction was incubated at 65 °C for 5 or 30 min, then 5 U of TURBO DNase (Invitrogen #AM2238) and 5 μl of provided 10 x buffer was added, volume brought to 50 μl with nuclease-free water and incubated at 37 °C for 30 min. The reaction was purified and size selected to >200 nts using Zymo RNA Clean and Concentrator-5 (Zymo #R1013) according to manufacturer's protocol, eluting in 10 μl of nuclease-free water. The WT Stage 5 sample was also depleted of mitochondrial COX2 and COX3 mRNA as part of the **Phelps et al., 2021** study. Strand-specific RNA-seq libraries were constructed using NEB Ultra II RNA-seq library kit (NEB #E7765) according to manufacturer's protocol with fragmentation in first-strand buffer at 94 °C for 15 min. Following first and second strand synthesis, DNA was purified with 1.8 X AmpureXP beads

(Beckman #A63880), end repaired, then ligated to sequencing adaptors diluted 1:5. Ligated DNA was purified with 0.9 X AmpureXP beads and PCR amplified for 8 cycles, then purified again with 0.9 X AmpureXP beads. Libraries were verified by Qubit dsDNA high sensitivity (Invitrogen #Q32851) and Fragment Analyzer prior to multiplexed sequencing at the Health Sciences Sequencing Core at Children's Hospital of Pittsburgh.

For samples used for differential expression analysis, separate libraries were constructed for each of two replicate sets of embryos from each experimental day, which were considered biological replicates for DESeq2. All libraries from the same experimental day are labeled with the same batch designation (e.g. a, b, c,...).

CUT&RUN

CUT&RUN procedure was adapted from *Hainer and Fazio, 2019* optimizations of the method of *Skene and Henikoff, 2017*. For nuclear extraction, embryos were de-vitellinized using 1 mg/mL pronase dissolved in MR/3. Once the vitelline envelope was removed, 12–24 embryos (50K – 100K cells) were carefully transferred into 1 mL of NP2.0 buffer (*Briggs et al., 2018*) in a 1.5 mL tube and gently agitated (pipetting buffer over the surface of the embryos) until cells have dissociated. The buffer was carefully drawn off to the level of the cells and 1 mL of Nuclear Extraction (NE) buffer (20 mM HEPES-KOH, pH 7.9, 10 mM KCl, 500 μ M spermidine, 0.1% Triton X-100, 20% glycerol) with gentle pipetting with a clipped P1000, and the lysate was centrifuged at 600xg in 4 °C for 3 min. The free nuclei were then bound to 300 μ L of activated concanavalin A beads (Polysciences #86057) at RT for 10 min. Nuclei were blocked for 5 min at RT then incubated in 1:100 dilution of primary antibody for 2 hr at 4 °C, washed, incubated in a 1:200 dilution of pAG MNase for 1 hr at 4 °C, and washed again. The bound MNase was activated with 2 mM CaCl_2 and allowed to digest for 30 min, then stopped using 2 x STOP buffer (200 mM NaCl, 20 mM EDTA, 4 mM EGTA, 50 μ g/mL RNase A, 40 μ g/mL glycogen). Nuclei were incubated at 37 °C for 20 min followed by centrifuging for 5 min at 16,000xg, drawing off the DNA fragments with the supernatant. The extracted fragments were treated with SDS and proteinase K at 70 °C for 10 min followed by phenol chloroform extraction. Purified DNA was resuspended in 50 μ L of water and verified by Qubit dsDNA high sensitivity and Fragment Analyzer. Antibodies used were: H3K4me3, Millipore #05–745 R, RRID:AB_1587134, Lot #3257057 (stage 8 rep 1 & stage 9 reps 1 & 2) and Invitrogen #711958, RRID:AB_2848246, Lot #2253580; H3K27ac, ActiveMotif #39135, RRID:AB_2614979, Lot #06419002; V5, Invitrogen #R960-25, RRID:AB_2556564, Lot #2148086. At least three biological replicate libraries from different embryo collection days were constructed for the key samples (St. 8 H3K27ac, St. 9 H3K4me3).

For transcription factor CUT&RUN, *pou5f3.3.L* and *sox3.S* IVT templates were cloned from cDNA using primers for *pou5f3.3.L* – NM_001088114.1 (F: GGACAGCACGGGAGGCGGGGATCCGAC CAGCCCATATTGTACAGCCAAAC; R: TATCATGTCTGGATCTACGTCTAGATCAGCCGGTCAGGAC CCC) and *sox3.S* - NM_001090679.1 (F: aaagatccTATAGCATGTTGGACACCGACATCA; R: aaatctag aTTATATGTGAGTGAGCGGTACCGTG) into N-terminal V5-pBS entry plasmids using HiFi assembly (NEB #E2621) for *pou5f3.3* and BamHI/XbaI for *sox3*. IVT was done using NEB HiScribe T7 ARCA kit (#E2065S) on NotI-linearized plasmid for 2 hr at 37 °C, then treated with 5 U of TURBO DNaseI (Invitrogen #AM2238) for 15 min. mRNA was purified using NEB Monarch RNA Cleanup Columns (#T2030) and stored at –80 °C until use. For injection, immediately after dejellying, stage 1 embryos were placed in 4% Ficoll-400 in MR/3. Each embryo was injected with 2.5 nL of 40 ng/ μ L of mRNA into each cell at stage 3 (4 cell), for a total of 10 nL per embryo. Three biological replicates from different days for each factor were generated. Factor-specific no-antibody CUT&RUN samples were made using the same injected embryos.

CUT&RUN libraries were constructed using the NEB Ultra II DNA library prep kit (NEB #E7645) according to manufacturer's protocol. DNA was end repaired and then ligated to sequencing adaptors diluted 1:10. Ligated DNA was purified with 0.9 x AmpureXP beads and PCR amplified for 15 cycles, then purified again with 0.9 x AmpureXP beads. Libraries were size selected to 175–650 bp for histone modifications and 150–650 bp for transcription factors on a 1.5% TBE agarose gel and gel purified using the NEB Monarch DNA gel extraction kit (#T1020) before being verified by Qubit dsDNA high sensitivity and Fragment Analyzer prior to multiplexed paired-end sequencing on an Illumina NextSeq 500 at the Health Sciences Sequencing Core at Children's Hospital of Pittsburgh.

ATAC-seq

ATAC procedure was from *Esmaeili et al., 2020*. Embryos were grown in MR/3 until desired NF stage and devitelinized individually with fine watch-maker forceps. Ectodermal explants (animal caps) were dissected using watch-maker forceps in 0.7 x MR. Two caps were transferred to 1 mL of ice-cold PBS and centrifuged at 500xg in 4 °C for 5 min twice. After washing with PBS, caps were lysed in 50 µl of RSB buffer (10 mM Tris pH 7.4, 10 mM NaCl, 3 mM MgCl₂, 0.1% Igepal CA-630) with a clipped P200 pipet. The lysate was centrifuged again for 10 min and the supernatant was drawn off. The pellet was resuspended in 47.5 µl TD buffer (10 mM Tris pH 7.6, 5 mM MgCl₂, 10% dimethylformamide) and 2.5 µl of 3 µM transposome (see below) was added. Nuclei were transposed with gentle shaking for 1 hr at 37 °C before adding 2.5 µl proteinase K and incubating overnight at 37 °C. Transposed DNA was purified using EconoSpin Micro columns (Epoch) and amplified using 25 µM indexed Nextera primers with Thermo Phusion Flash master mix for 12 cycles. Primers used were: CAAGCAGAAGACGGCATACGAGAT[i7]GTCTCGTGGGCTCGG with i7 indices 707 – gtagagag; 714 – tcatgagc; 716 – tagcgagt; and AATGATACGGCGACCACCGAGATCTACAC[i5]TCGTCGGCAGCGTC with i5 indices 505 – gtaaggag; 510 – cgtctaat; 517 – gcgtaaga; 520 – aaggctat. The amplified library was column cleaned and verified by Qubit dsDNA high sensitivity and Fragment Analyzer and sequenced multiplexed paired end at the Health Sciences Sequencing Core at Children's Hospital of Pittsburgh. For the first two replicates per stage, after initial sequencing, libraries were subsequently size selected on an agarose gel to enrich for 150–250 and 250–600 bp fragments and resequenced pooled. For a third stage 8 replicate and third and fourth stage 9 replicate, only the 150–250 bp fragments were sequenced. Biological replicate libraries are from different embryo collection days.

Transposomes were constructed according to *Picelli et al., 2014*. Adapter duplexes for Tn5ME-A (TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG) + Tn5MErev ([phos]CTGTCTCTTATACACATCT) and Tn5ME-B (GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG) + Tn5MErev were each annealed in 2 µl of 10 X annealing buffer (100 mM HEPES pH 7.2, 500 mM NaCl, 10 mM EDTA) using 9 µl of each oligo at 100 µM, heated to 95 °C for 1 min then ramped down to 25 °C at 0.1 °C/s in a thermocycler. The two duplexes were held at 25 °C for 5 min then mixed together. On ice, 35 µl of hot glycerol was cooled to 4 °C then 35 µl of the primer mixture and 25 µl of Tn5 (Addgene #112112) was added and mixed and held at 1 hr at RT with gentle pipet mixing every 15 min. Transposomes were stored at –20 °C.

Transcriptomic analysis

RNA-seq reads were mapped to the *X. laevis* v9.2 genome using HISAT2 v2.0.5 (*Kim et al., 2015*) (--no-mixed --no-discordant). Mapped reads were assigned to gene exons (Xenbase v9.2 models) using featureCounts v2.0.1 *Liao et al., 2014* in reversely-stranded paired-end mode with default parameters, and to introns with --minOverlap 10 on a custom intron annotation: starting with all introns from the v9.2 GFF file, subtract (a) all regions detected in stage 5 RNA-seq at >2 read coverage, strand specifically; (b) all regions that overlap an annotated exon from a different transcript form; (c) regions that overlap repetitive elements as defined by RepeatMasker (UCSC) and Xenbase-annotated transposons, not strand specifically; (d) regions that ambiguously map to more than one distinct gene's intron (i.e. transcript forms of the same gene are allowed to share an intron, but not between different genes).

DESeq2 v4.0.3 (*Love et al., 2014*) was used for statistical differential expression analysis. To build the DESeq2 model, exon and intron raw read counts were treated as separate rows per gene in the same counts matrix (intron gene IDs were preceded with a 'i_' prefix). Only genes annotated by Xenbase as 'protein_coding', 'lncRNA', or 'pseudogene' were retained. Low-expressed genes were removed (exon reads per million (RPM) <0.5 across all samples) and then low-depth intron features were removed (intron raw read count ≤10 or reads per kilobase per million (RPKM) <0.25 across all samples). Comparisons were made between batch-matched samples where possible, to account for variations in the maternal contribution between mothers. Significant differences with adjusted <0.05 and log2 difference ≥1.5 were used for downstream analysis. High-confidence activated genes had significant increases in DMSO vs Triptolide for both batches and stage 9 vs stage 5. High-confidence primary-activation 'first-wave' genes were high-confidence activated and had significant increase in DMSO vs Cycloheximide. High-confidence activated genes significantly changed in any Pou/Sox morpholino treatment were considered to be affected genes. For chromatin profiling, genes

were classified as Pou/Sox down-regulated if they were significantly decreased in morpholino treatment either with or without cycloheximide. Homeologous genes were paired according to Xenbase GENEPAGE annotations. Genes were considered maternal if they had average stage 5 TPM ≥ 1 . To calculate magnitude of effect for graphing and sorting, the maximal $|\log_2$ fold difference| of average exon TPM and average intron RPKM was chosen per gene.

For mir-427 gene identification and RNA-seq coverage visualization, miRBase (Kozomara et al., 2019) hairpin sequences MI0001449 and MI0038331 were aligned to the v9.2 and v10.1 reference genomes using UCSC BLAT (Kent, 2002) and maximal possible read coverage was graphed allowing all multimappers. To align the v10.1 Chr1L and Chr1S regions flanking the Chr1L mir-427 locus, genomic sequence was extracted between homeologous genes upstream and downstream mir-427. Local alignments with E-value $< 1e-10$ were retained from an NCBI BLAST 2.11.0+blastn alignment (Camacho et al., 2009).

dN/dS ratios were calculating using PAML v4.9f (Yang, 1997) with L-S pairwise CDS alignments produced by pal2nal v14 (Suyama et al., 2006) on amino-acid alignments by EMBOSS needle v6.6.0.0 (-gapopen 10 -gapextend 0.5) (Rice et al., 2000).

All other statistical tests were performed using R v4.0.4 (R Development Core Team, 2013).

Chromatin profiling analysis

CUT&RUN and ATAC-seq paired-end reads were mapped to the *X. laevis* v10.1 genome using bowtie2 v2.4.2 (Langmead and Salzberg, 2012) (--no-mixed --no-discordant) and only high-quality alignments (MAPQ ≥ 30) were retained for subsequent analysis. Read pairs were joined into contiguous fragments for coverage analyses. For transcription factor CUT&RUN, reads were trimmed using trim_galore v0.6.6 and Cutadapt v1.15 (Martin, 2011) in paired-end mode (--illumina --trim-n). Downstream analyses were performed using custom scripts with the aid of BEDtools v2.30.0 (Quinlan and Hall, 2010), Samtools v1.12 (Li et al., 2009), and deepTools v3.5.1 (Ramírez et al., 2014).

For promoter-centered analyses, one transcript isoform per gene was selected from Xenbase v9.2 annotations: the most upstream TSS with non-zero RNA-seq coverage at stage 9 was used, otherwise the most upstream TSS if no RNA-seq evidence. Then the corresponding v10.1 coordinates were obtained based on gene name match.

To identify open chromatin regions, aligned stage 9 ATAC-seq fragments pooled between replicates were filtered to < 130 bp, then peaks called using MACS2 v2.2.7.1 (Zhang et al., 2008) with an effective genome size of $2.74e9$ (number of non-N bases in the v10 reference sequence). CUT&RUN no-antibody samples were used as the control sample. To further exclude probable false-positive regions, peaks overlapping any of the following repetitive regions were removed: (a) scRNA, snRNA, snoRNA, or tRNA as annotated by Xenbase; (b) rRNA as determined by BLASTed 45 S, 16 S, 12 S, and 5 S sequences. Peaks on unassembled scaffolds were also excluded.

Putative enhancers had twofold enriched stage 8 H3K27ac CUT&RUN coverage over no antibody, with ≥ 1 RPKM pooled H3K27ac coverage and < 1 RPKM no-antibody coverage, in a 500 bp window centered on ATAC-seq peak summits. A subset of these were additionally annotated as high-confidence ('hi') if they had twofold enrichment in each of at least three individual H3K27ac CUT&RUN samples and three ATAC-seq samples, and lower confidence otherwise ('lo'). Stage 9 ATAC-seq replicates 3 and 4 were pooled to serve as a single sample for this purpose, due to lower read depth. Enhancers were classified as distal ('dist') if they were > 1 kb from any Xenbase v10.1 annotated TSS, proximal ('prox') otherwise.

For transcription factor peak calling, replicates were pooled per factor, then individual replicates were verified for enrichment at peaks. No-antibody samples were pooled as a uniform background. MACS2 was run as above, and SEACR v1.3 (Meers et al., 2019) was run in norm stringent mode. Peak calls were not used for enhancer analyses; rather, enhancers or homeologous regions with ≥ 0.5 RPM CUT&RUN coverage and ≥ 2 -fold enrichment over no antibody in a 200 bp window were considered bound.

Coverage heatmaps were generated using deepTools on reads-per-million normalized bigWigs; or enrichment over no-antibody bigWigs generated using deepTools bigwigCompare (--operation ratio --pseudocount 0.1 --binSize 50 --skipZeroOverZero).

For density heatmaps, L/S enhancer pairs were annotated as differential or shared based on one or both partners, respectively, mapping to a putative enhancer, as described above. Pairs were similarly

annotated as differentially or both TF bound based on ≥ 2 -fold enrichment over no antibody for either TF at one or both partners, respectively, and converted to bigWigs representing the genomic location of each bound putative enhancer. Density heatmaps were generated as above and plotted with respect to selected TSSs.

Motif finding

Enriched sequence motifs in enhancers were identified using Homer v4.11.1 (Heinz et al., 2010) in scanning mode against the vertebrate database, using 200 bp of sequence centered on the ATAC-seq peak for enhancers; and 500 bp of sequence centered on the TSS for promoters. Enrichment was calculated using one set of homeologous regions (L or S) as the foreground and the other as the background. The top representative motif per DNA binding domain was reported. For transcription factor peaks, Homer was first used in de novo mode on the top 1000 MACS peaks for Pou5f3 and Sox3 separately; the top motif matched mammalian Oct4 and Sox3 database motifs, respectively. To determine the sequence logo for the Pou5f3-Sox3 dimer motif, a subset of Sox3 peaks with adjacent Pou5f3 and Sox3 motifs was extracted and Homer motif finding was performed using `-len 15`. To calculate motif prevalence across all peaks, Homer database motif matrices for Oct4 (GSE11431), Sox3 (GSE33059) and OCT4-SOX2-TCF-NANOG (GSE11431) (representing the OCT4-SOX2 dimer motif) were scanned against the entire set of peaks. A set of ATAC-seq accessible regions with no H3K27ac enrichment (rejected regions from the above enhancer prediction analysis) and <0.5 fold enrichment of Pou5f or Sox3 CUT&RUN signal was also scanned to estimate background motif frequencies. Peaks with hits for the dimer motif were secondarily filtered to additionally require the presence of the Pou5f3 and Sox3 individual motifs.

Homeologous enhancer identification

Each v10.1 chromosome pair (e.g. Chr1L and Chr1S) was aligned using lastZ-1.04.00 (Harris, 2007) and UCSC Genome Browser utilities (Kent et al., 2002) with parameters adapted from the UCSC Genome Browser previously used to align *X. tropicalis* with *X. laevis* (http://www.bx.psu.edu/miller_lab/dist/README.lastz-1.02.00/README.lastz-1.02.00a.html; http://genomewiki.ucsc.edu/index.php/XenTro9_11-way_conservation_lastz_parameters; no automatic chaining; open = 400, extend = 30, masking = 0, seed = 1 {12of19}, hspthreshold = 3000, chain = 0, ydropoff = 9400, gapped-threshold = 3000, inner = 2000). Chaining and netting were done with axtChain linearGap set to medium and chainSplit lump = 50. Nets were generated using default chainNet and the highest scoring chains were selected from those nets using default netChainSubset. Reciprocal best chains were identified according to UCSC Genome Browser guidelines. The highest scoring chains were reverse referenced, sorted, and then converted to nets using default chainPreNet and chainNet (`-minSpace=1 -minScore=0`). Reciprocal best nets were selected with default netSyntenic. The new highest scoring best chains were extracted using netChainSubset, converted back to the original reference, and netted as described prior, resulting in reciprocal best, highest scoring chains for use with liftOver.

In the first pass, 500 bp enhancer regions centered on the ATAC-seq peak were lifted to the homeologous subgenome with a 10% minimum sequence match requirement. For enhancers that failed this liftOver, 5 kb enhancer regions were lifted over; as a stringency check, each 2.5 kb half was also individually lifted over, and only regions correctly flanked by both halves were retained. If an enhancer's homeologous region also overlaps an annotated enhancer, it was considered conserved, otherwise it was considered subgenome-specific. To test synteny, the 5 closest Xenbase-annotated genes up- and downstream of each region in a homeologous pair were compared.

Comparison with *X. tropicalis* and zebrafish

X. tropicalis wild-type RNA-seq reads from Owens et al., 2016, RiboZero stage 5 (SRA: SRR1795666) and stage 9 (SRA: SRR1795634), were aligned by HISAT2 as above and mapped to Xenbase v10 gene annotations using featureCounts. Pou5f3/Sox3 morpholino and alpha-amanitin-affected genes were obtained from published data tables from Gentsch et al., 2019, and the JGI gene accession numbers were mapped to Xenbase GenePage IDs (v7.1). Significantly affected genes were 1.5-fold decreased and adjusted $p < 0.05$. Genes with TPM > 1 at either stage 5 or stage 9 were considered embryonic expressed.

For transcriptome comparisons between *X. laevis* subtranscriptomes and *X. tropicalis*, log₂ TPM values for non-zero expressed genes per transcriptome (L homeologs only, S homeologs only, L+S homeologs summed, *tropicalis*) were Z-normalized to calculate correlations. To measure gene-wise deviation of the *laevis* transcriptome/sub-transcriptome from *tropicalis*, residuals were calculated using *tropicalis* Z-normalized expression as the predictor variable (i.e. *tropicalis* expression minus *laevis* expression per gene), with the null hypothesis that *tropicalis* is equally as good a predictor for the L+S composite transcriptome compared to L only or S only.

Zebrafish annotations for activated and Pou5f3/Nanog / SoxB1 affected genes were obtained from Lee et al., 2013 and associated to *Xenopus* genes using Ensembl ortholog annotations (Xenbase to Zfin). First-wave activated zebrafish genes are significantly increased in the U1/U2 spliceosomal RNA inhibited sample over alpha-amanitin (DESeq2 adjusted p<0.05), activated genes are significantly increased by 6 h.p.f. over alpha-amanitin. Pou5f3/SoxB1 affected genes were significantly decreased in the Pou5f3-SoxB1 double loss of function versus wild-type. Nanog-affected genes were significantly decreased in triple loss of function (NSP) but not Pou5f3-SoxB1 double loss of function. Genes with TPM >1 at 2, 4, or 6 h.p.f. were considered embryonic expressed.

To identify putative conserved enhancers in *X. tropicalis*, *X. laevis* enhancers on the v10.1 genome were BLATed (Kent, 2002) to the *X. laevis* v9.2 genome, then lifted over to the *X. tropicalis* v9.2 genome using liftOver chains from the UCSC Genome Browser (xenLae2ToXenTro9, 10% minimum sequence match). Successfully lifted over regions were intersected with published *X. tropicalis* H3K27ac stage 9 peaks from Gupta et al., 2014 that were lifted from the *X. tropicalis* v2 genome to the v9 genome, passing through v7 and requiring 90% minimum sequence match, using liftOver chains from UCSC Genome Browser (xenTro2ToXenTro7 and xenTro7ToXenTro9). *X. laevis* enhancers were lifted over to the zebrafish GRCz11 genome using liftOver chains from the UCSC Genome Browser, passing through *X. tropicalis* (xenLae2ToXenTro9, 10% minimum sequence match; then xenTro9ToXenTro7, 90% minimum sequence match, then xenTro7ToDanRer10, 10% minimum sequence match, then danRer10ToDanRer11 requiring 90% minimum sequence match). Acetylation at zebrafish dome stage was then assessed by intersecting with H3K27ac ChIP-seq peaks from Bogdanovic et al., 2012 (GEO: GSM915197): reads were aligned to the GRCz11 genome using bowtie2 as above, and peaks called using macs2 as above with an effective genome size of 4.59e8 and no control sample.

Acknowledgements

We thank S Hainer and lab for providing the pAG-MNase enzyme, assistance with the CUT&RUN protocol, and feedback. We thank M Rebeiz, T Levin, M Turcotte, K Arndt and lab, C Kaplan and lab, and the entire Lee lab for feedback. This project used the University of Pittsburgh Health Sciences Core at UPMC Children's Hospital Pittsburgh for sequencing. This work was supported by the March of Dimes #5-FY16-307, the National Institutes of Health R35GM137973, the Samuel and Emma Winters Foundation, and start-up funds from the University of Pittsburgh to MTL. This research was supported in part by the University of Pittsburgh Center for Research Computing, RRID:SCR_022735, through the resources provided. Specifically, this work used the H2P cluster, which is supported by NSF award number OAC-2117681.

Additional information

Funding

| Funder | Grant reference number | Author |
|------------------------------------|------------------------|-------------|
| National Institutes of Health | R35GM137973 | Miler T Lee |
| March of Dimes Foundation | 5-FY16-307 | Miler T Lee |
| Samuel and Emma Winters Foundation | | Miler T Lee |

| Funder | Grant reference number | Author |
|--|------------------------|--------|
| The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication. | | |

Author contributions

Wesley A Phelps, Conceptualization, Data curation, Software, Formal analysis, Validation, Investigation, Visualization, Methodology, Writing – original draft, Writing – review and editing; Matthew D Hurton, Investigation, Methodology; Taylor N Ayers, Investigation; Anne E Carlson, Joel C Rosenbaum, Resources; Miler T Lee, Conceptualization, Data curation, Software, Formal analysis, Supervision, Funding acquisition, Validation, Investigation, Visualization, Methodology, Writing – original draft, Project administration, Writing – review and editing

Author ORCIDs

Wesley A Phelps  <https://orcid.org/0000-0002-4056-2345>
Taylor N Ayers  <http://orcid.org/0000-0003-0680-0773>
Anne E Carlson  <http://orcid.org/0000-0003-2724-1325>
Miler T Lee  <https://orcid.org/0000-0003-0933-0551>

Ethics

All animal procedures were conducted under the supervision and approval of the Institutional Animal Care and Use Committee at the University of Pittsburgh under protocol #21120500.

Decision letter and Author response

Decision letter <https://doi.org/10.7554/eLife.83952.sa1>
Author response <https://doi.org/10.7554/eLife.83952.sa2>

Additional files

Supplementary files

- Supplementary file 1. RNA-seq expression values and DESeq2 comparisons.
- Supplementary file 2. Annotations and expression values for activated genes.
- Supplementary file 3. Transcription start site coordinates used for all genes.
- Supplementary file 4. Motif search results.
- Supplementary file 5. Enhancer annotations.
- Supplementary file 6. Comparative genomics with *X. tropicalis* and zebrafish.
- MDAR checklist

Data availability

All data and analysis files are available with no restrictions on access. Sequencing data are available in the Gene Expression Omnibus (GEO) under accession number GSE207027. Code and auxiliary data files are available on Github, <https://github.com/MTLeeLab/xl-zga> (copy archived at **Phelps and Lee, 2023**). Additional data files including chromosome alignments are available at OSF, <https://osf.io/ct6g8/>.

The following datasets were generated:

| Author(s) | Year | Dataset title | Dataset URL | Database and Identifier |
|-------------------|------|---|---|---|
| Phelps WA, Lee MT | 2022 | Hybridization led to a rewired pluripotency network in the allotetraploid <i>Xenopus laevis</i> | https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE207027 | NCBI Gene Expression Omnibus, GSE207027 |
| Phelps WA, Lee MT | 2022 | <i>Xenopus</i> MZT | https://doi.org/10.17605/OSF.IO/CT6G8 | Open Science Framework, 10.17605/OSF.IO/CT6G8 |

The following previously published datasets were used:

| Author(s) | Year | Dataset title | Dataset URL | Database and Identifier |
|--|------|---|---|---|
| Bogdanović O, Fernandez-Miñán A, Tena JJ, de la Calle-Mustienes E, Hidalgo C, van Heeringen SJ, Veenstra GJ, Gómez-Skarmeta JL | 2011 | Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis | https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE32483 | NCBI Gene Expression Omnibus, GSE32483 |
| Lee MT, Phelps WA | 2020 | Optimized design of antisense oligomers for targeted rRNA depletion | https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE152902 | NCBI Gene Expression Omnibus, GSE152902 |
| Owens ND, Blitz IL, Lane MA, Patrushev I, Overton JD, Gilchrist MJ, Cho KW, Khokha MK | 2016 | Measuring Absolute RNA Copy Numbers at High Temporal Resolution Reveals Transcriptome Kinetics in Development | https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE65785 | NCBI Gene Expression Omnibus, GSE65785 |
| Gentsch GE, Smith JC | 2019 | Maternal pluripotency factors initiate extensive chromatin remodelling to predefine first response to inductive signals | https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE113186 | NCBI Gene Expression Omnibus, GSE113186 |
| Lee MT, Bonneau AR, Giraldez AJ | 2013 | Nanog, SoxB1 and Pou5f1/Oct4 regulate widespread zygotic gene activation during the maternal-to-zygotic transition | https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE47558 | NCBI Gene Expression Omnibus, GSE47558 |
| Gupta R, Baker JC | 2014 | Enhancer chromatin signatures predict Smad2/3 binding in <i>Xenopus</i> | https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE56000 | NCBI Gene Expression Omnibus, GSE56000 |
| Johnson K, LaBonne C | 2022 | Transcriptome sequencing of <i>Xenopus laevis</i> animal caps at six time points in transit from pluripotency to 4 lineages: epidermal, neural, ventral mesoderm and endoderm | https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE198598 | NCBI Gene Expression Omnibus, GSE198598 |

References

- Adams KL, Wendel JF. 2005. Polyploidy and genome evolution in plants. *Current Opinion in Plant Biology* **8**:135–141. DOI: <https://doi.org/10.1016/j.pbi.2005.01.001>, PMID: 15752992
- Amodeo AA, Jukam D, Straight AF, Skotheim JM. 2015. Histone titration against the genome sets the DNA-to-cytoplasm threshold for the *Xenopus* midblastula transition. *PNAS* **112**:E1086–E1095. DOI: <https://doi.org/10.1073/pnas.1413990112>
- Bazzini AA, Lee MT, Giraldez AJ. 2012. Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. *Science* **336**:233–237. DOI: <https://doi.org/10.1126/science.1215704>, PMID: 22422859
- Blythe SA, Wieschaus EF. 2016. Establishment and maintenance of heritable chromatin structure during early *Drosophila* embryogenesis. *eLife* **5**:e20148. DOI: <https://doi.org/10.7554/eLife.20148>, PMID: 27879204
- Bogdanovic O, Fernandez-Miñán A, Tena JJ, de la Calle-Mustienes E, Hidalgo C, van Kruysbergen I, van Heeringen SJ, Veenstra GJC, Gómez-Skarmeta JL. 2012. Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis. *Genome Research* **22**:2043–2053. DOI: <https://doi.org/10.1101/gr.134833.111>, PMID: 22593555
- Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG, Gifford DK, Melton DA, Jaenisch R, Young RA. 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**:947–956. DOI: <https://doi.org/10.1016/j.cell.2005.08.020>
- Briggs JA, Weinreb C, Wagner DE, Megason S, Peshkin L, Kirschner MW, Klein AM. 2018. The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science* **360**:eaar5780. DOI: <https://doi.org/10.1126/science.aar5780>

- Buggs RJA**, Wendel JF, Doyle JJ, Soltis DE, Soltis PS, Coate JE. 2014. The legacy of diploid progenitors in allopolyploid gene expression patterns. *Philosophical Transactions of the Royal Society B* **369**:20130354. DOI: <https://doi.org/10.1098/rstb.2013.0354>
- Camacho C**, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* **10**:421. DOI: <https://doi.org/10.1186/1471-2105-10-421>, PMID: 20003500
- Charney RM**, Forouzmand E, Cho JS, Cheung J, Paraiso KD, Yasuoka Y, Takahashi S, Taira M, Blitz IL, Xie X, Cho KWW. 2017. Foxh1 occupies cis-regulatory modules prior to dynamic transcription factor interactions controlling the mesendoderm gene program. *Developmental Cell* **40**:595–607.. DOI: <https://doi.org/10.1016/j.devcel.2017.02.017>, PMID: 28325473
- Chen H**, Einstein LC, Little SC, Good MC. 2019a. Spatiotemporal patterning of zygotic genome activation in a model vertebrate embryo. *Developmental Cell* **49**:852–866.. DOI: <https://doi.org/10.1016/j.devcel.2019.05.036>
- Chen Z**, Omori Y, Koren S, Shirokiya T, Kuroda T, Miyamoto A, Wada H, Fujiyama A, Toyoda A, Zhang S, Wolfsberg TG, Kawakami K, Phillippy AM, NISC Comparative Sequencing Program, Mullikin JC, Burgess SM. 2019b. De novo assembly of the goldfish (*Carassius auratus*) genome and the evolution of genes after whole-genome duplication . *Science Advances* **5**:eaav0547. DOI: <https://doi.org/10.1126/sciadv.aav0547>
- Chen H**, Good MC. 2022. Nascent transcriptome reveals orchestration of zygotic genome activation in early embryogenesis. *Current Biology* **32**:4314–4324.. DOI: <https://doi.org/10.1016/j.cub.2022.07.078>
- Colonna MM**, Abrahante JE, Schedl P, Gohl DM, Deshpande G. 2021. CLAMP regulates zygotic genome activation in *Drosophila* embryos. *Genetics* **219**:iyab107. DOI: <https://doi.org/10.1093/genetics/iyab107>, PMID: 34849887
- Dailey L**, Basilico C. 2001. Coevolution of HMG domains and homeodomains and the generation of transcriptional regulation by Sox/POU complexes. *Journal of Cellular Physiology* **186**:315–328. DOI: [https://doi.org/10.1002/1097-4652\(2001\)9999:9999<000::AID-JCP1046>3.0.CO;2-Y](https://doi.org/10.1002/1097-4652(2001)9999:9999<000::AID-JCP1046>3.0.CO;2-Y), PMID: 11169970
- Duan J**, Rieder L, Colonna MM, Huang A, Mckenney M, Watters S, Deshpande G, Jordan W, Fawzi N, Larschan E. 2021. CLAMP and Zelda function together to promote *Drosophila* zygotic genome activation. *eLife* **10**:e69937. DOI: <https://doi.org/10.7554/eLife.69937>, PMID: 34342574
- Elurbe DM**, Paranjpe SS, Georgiou G, van Kruijsbergen I, Bogdanovic O, Gibeaux R, Heald R, Lister R, Huynen MA, van Heeringen SJ, Veenstra GJC. 2017. Regulatory remodeling in the allo-tetraploid frog *Xenopus laevis*. *Genome Biology* **18**:198. DOI: <https://doi.org/10.1186/s13059-017-1335-7>, PMID: 29065907
- Endo Y**, Kamei K-I, Inoue-Murayama M. 2020. Genetic signatures of evolution of the pluripotency gene regulating network across mammals. *Genome Biology and Evolution* **12**:1806–1818. DOI: <https://doi.org/10.1093/gbe/evaa169>, PMID: 32780791
- Esmaeili M**, Blythe SA, Tobias JW, Zhang K, Yang J, Klein PS. 2020. Chromatin accessibility and histone acetylation in the regulation of competence in early development. *Developmental Biology* **462**:20–35. DOI: <https://doi.org/10.1016/j.ydbio.2020.02.013>, PMID: 32119833
- Fernandez-Tresguerres B**, Cañon S, Rayon T, Pernaute B, Crespo M, Torroja C, Manzanares M. 2010. Evolution of the mammalian embryonic pluripotency gene regulatory network. *PNAS* **107**:19955–19960. DOI: <https://doi.org/10.1073/pnas.1010708107>
- Foe VE**, Alberts BM. 1983. Studies of nuclear and cytoplasmic behaviour during the five mitotic cycles that precede gastrulation in *Drosophila* embryogenesis. *Journal of Cell Science* **61**:31–70. DOI: <https://doi.org/10.1242/jcs.61.1.31>, PMID: 6411748
- Gaskill MM**, Gibson TJ, Larson ED, Harrison MM. 2021. GAF is essential for zygotic genome activation and chromatin accessibility in the early *Drosophila* embryo. *eLife* **10**:e66668. DOI: <https://doi.org/10.7554/eLife.66668>
- Gentsch GE**, Spruce T, Owens NDL, Smith JC. 2019. Maternal pluripotency factors initiate extensive chromatin remodelling to predefine first response to inductive signals. *Nature Communications* **10**:4269. DOI: <https://doi.org/10.1038/s41467-019-12263-w>, PMID: 31537794
- Gibeaux R**, Acker R, Kitaoka M, Georgiou G, van Kruijsbergen I, Ford B, Marcotte EM, Nomura DK, Kwon T, Veenstra GJC, Heald R. 2018. Paternal chromosome loss and metabolic crisis contribute to hybrid inviability in *Xenopus*. *Nature* **553**:337–341. DOI: <https://doi.org/10.1038/nature25188>, PMID: 29320479
- Giraldez AJ**, Mishima Y, Rihel J, Grocock RJ, Van Dongen S, Inoue K, Enright AJ, Schier AF. 2006. Zebrafish MiR-430 Promotes deadenylation and clearance of maternal mRNAs. *Science* **312**:75–79. DOI: <https://doi.org/10.1126/science.1122689>
- Grover CE**, Gallagher JP, Szadkowski EP, Yoo MJ, Flagel LE, Wendel JF. 2012. Homoeolog expression bias and expression level dominance in allopolyploids. *The New Phytologist* **196**:966–971. DOI: <https://doi.org/10.1111/j.1469-8137.2012.04365.x>, PMID: 23033870
- Gupta R**, Wills A, Ucar D, Baker J. 2014. Developmental enhancers are marked independently of zygotic Nodal signals in *Xenopus*. *Developmental Biology* **395**:38–49. DOI: <https://doi.org/10.1016/j.ydbio.2014.08.034>, PMID: 25205067
- Gurdon JB**, Elsdale TR, Fischberg M. 1958. Sexually mature individuals of *Xenopus laevis* from the transplantation of single somatic nuclei. *Nature* **182**:64–65. DOI: <https://doi.org/10.1038/182064a0>, PMID: 13566187
- Hadzhiev Y**, Wheatley L, Cooper L, Ansaloni F, Whalley C, Chen Z, Finaurini S, Gustincich S, Sanges R, Burgess S, Beggs A, Müller F. 2023. The miR-430 locus with extreme promoter density forms a transcription body during

- the minor wave of zygotic genome activation. *Developmental Cell* **58**:155–170.. DOI: <https://doi.org/10.1016/j.devcel.2022.12.007>, PMID: 36693321
- Hainer SJ, Fazio TG. 2019. High-Resolution Chromatin profiling using CUT&RUN. *Current Protocols in Molecular Biology* **126**:e85. DOI: <https://doi.org/10.1002/cpmb.85>, PMID: 30688406
- Harland RM, Grainger RM. 2011. *Xenopus* research: metamorphosed by genetics and genomics. *Trends in Genetics* **27**:507–515. DOI: <https://doi.org/10.1016/j.tig.2011.08.003>, PMID: 21963197
- Harris RS. 2007. *Improved Pairwise Alignment of Genomic DNA* The Pennsylvania State University.
- Harvey SA, Sealy I, Kettleborough R, Fenyes F, White R, Stemple D, Smith JC. 2013. Identification of the zebrafish maternal and paternal transcriptomes. *Development* **140**:2703–2710. DOI: <https://doi.org/10.1242/dev.095091>
- Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular Cell* **38**:576–589. DOI: <https://doi.org/10.1016/j.molcel.2010.05.004>, PMID: 20513432
- Hu G, Wendel JF. 2019. Cis-trans controls and regulatory novelty accompanying allopolyploidization. *The New Phytologist* **221**:1691–1700. DOI: <https://doi.org/10.1111/nph.15515>, PMID: 30290011
- Hug CB, Grimaldi AG, Kruse K, Vaquerizas JM. 2017. Chromatin architecture emerges during zygotic genome activation independent of transcription. *Cell* **169**:216–228.. DOI: <https://doi.org/10.1016/j.cell.2017.03.024>, PMID: 28388407
- Husband BC, Baldwin SJ, Suda J. 2013. The incidence of Polyploidy in natural plant populations: major patterns and evolutionary processes. Greilhuber J, Dolezel J, Wendel JF (Eds). *Plant Genome Diversity Volume 2: Physical Structure, Behaviour and Evolution of Plant Genomes* Springer. p. 255–276. DOI: <https://doi.org/10.1007/978-3-7091-1160-4>
- Johnson K, Freedman S, Braun R, LaBonne C. 2022. Quantitative analysis of transcriptome dynamics provides novel insights into developmental state transitions. *BMC Genomics* **23**:723. DOI: <https://doi.org/10.1186/s12864-022-08953-3>, PMID: 36273135
- Kane DA, Kimmel CB. 1993. The zebrafish midblastula transition. *Development* **119**:447–456. DOI: <https://doi.org/10.1242/dev.119.2.447>
- Kent WJ. 2002. BLAT—the BLAST-like alignment tool. *Genome Research* **12**:656–664. DOI: <https://doi.org/10.1101/gr.229202>, PMID: 11932250
- Kent W, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The Human Genome Browser at UCSC. *Genome Research* **12**:996–1006. DOI: <https://doi.org/10.1101/gr.229102>
- Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods* **12**:357–360. DOI: <https://doi.org/10.1038/nmeth.3317>
- Kimelman D, Kirschner M, Scherson T. 1987. The events of the midblastula transition in *Xenopus* are regulated by changes in the cell cycle. *Cell* **48**:399–407. DOI: [https://doi.org/10.1016/0092-8674\(87\)90191-7](https://doi.org/10.1016/0092-8674(87)90191-7)
- Kozomara A, Birgaoanu M, Griffiths-Jones S. 2019. miRBase: from microRNA sequences to function. *Nucleic Acids Research* **47**:D155–D162. DOI: <https://doi.org/10.1093/nar/gky1141>, PMID: 30423142
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**:357–359. DOI: <https://doi.org/10.1038/nmeth.1923>, PMID: 22388286
- Lee MT, Bonneau AR, Takacs CM, Bazzini AA, DiVito KR, Fleming ES, Giraldez AJ. 2013. Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. *Nature* **503**:360–364. DOI: <https://doi.org/10.1038/nature12632>, PMID: 24056933
- Lee MT, Bonneau AR, Giraldez AJ. 2014. Zygotic genome activation during the maternal-to-zygotic transition. *Annual Review of Cell and Developmental Biology* **30**:581–613. DOI: <https://doi.org/10.1146/annurev-cellbio-100913-013027>, PMID: 25150012
- Leichsenring M, Maes J, Mössner R, Driever W, Onichtchouk D. 2013. Pou5f1 transcription factor controls zygotic gene activation in vertebrates. *Science* **341**:1005–1009. DOI: <https://doi.org/10.1126/science.1242527>, PMID: 23950494
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* **25**:2078–2079. DOI: <https://doi.org/10.1093/bioinformatics/btp352>
- Li M, Belmonte JCI. 2017. Ground rules of the pluripotency gene regulatory network. *Nature Reviews. Genetics* **18**:180–191. DOI: <https://doi.org/10.1038/nrg.2016.156>, PMID: 28045100
- Li JT, Wang Q, Huang Yang MD, Li QS, Cui MS, Dong ZJ, Wang HW, Yu JH, Zhao YJ, Yang CR, Wang YX, Sun XQ, Zhang Y, Zhao R, Jia ZY, Wang XY. 2021. Parallel subgenome structure and divergent expression evolution of allo-tetraploid common carp and goldfish. *Nature Genetics* **53**:1493–1503. DOI: <https://doi.org/10.1038/s41588-021-00933-9>, PMID: 34594040
- Liang HL, Nien CY, Liu HY, Metzstein MM, Kirov N, Rushlow C. 2008. The zinc-finger protein Zelda is a key activator of the early zygotic genome in *Drosophila*. *Nature* **456**:400–403. DOI: <https://doi.org/10.1038/nature07388>, PMID: 18931655
- Liao Y, Smyth GK, Shi W. 2014. Feature Counts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**:923–930. DOI: <https://doi.org/10.1093/bioinformatics/btt656>, PMID: 24227677
- Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**:550. DOI: <https://doi.org/10.1186/s13059-014-0550-8>

- Lund E, Liu M, Hartley RS, Sheets MD, Dahlberg JE. 2009. Deadenylation of maternal mRNAs mediated by miR-427 in *Xenopus laevis* embryos. *RNA* **15**:2351–2363. DOI: <https://doi.org/10.1261/rna.1882009>
- Luo J, Chai J, Wen Y, Tao M, Lin G, Liu X, Ren L, Chen Z, Wu S, Li S, Wang Y, Qin Q, Wang S, Gao Y, Huang F, Wang L, Ai C, Wang X, Li L, Ye C, et al. 2020. From asymmetrical to balanced genomic diversification during rediploidization: Subgenomic evolution in allotetraploid fish. *Science Advances* **6**:eaaz7677. DOI: <https://doi.org/10.1126/sciadv.aaz7677>, PMID: 32766441
- Mable BK. 2004. ‘Why polyploidy is rarer in animals than in plants’: myths and mechanisms. *Biological Journal of the Linnean Society* **82**:453–466. DOI: <https://doi.org/10.1111/j.1095-8312.2004.00332.x>
- Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal* **17**:10. DOI: <https://doi.org/10.14806/ej.17.1.200>
- McClintock B. 1984. The significance of responses of the genome to challenge. *Science* **226**:792–801. DOI: <https://doi.org/10.1126/science.15739260>, PMID: 15739260
- Meers MP, Tenenbaum D, Henikoff S. 2019. Peak calling by Sparse Enrichment Analysis for CUT&RUN chromatin profiling. *Epigenetics & Chromatin* **12**:42. DOI: <https://doi.org/10.1186/s13072-019-0287-4>, PMID: 31300027
- Miao L, Tang Y, Bonneau AR, Chan SH, Kojima ML, Pownall ME, Vejnar CE, Gao F, Krishnaswamy S, Hendry CE, Giraldez AJ. 2022. The landscape of pioneer factor activity reveals the mechanisms of chromatin reprogramming and genome activation. *Molecular Cell* **82**:986–1002. DOI: <https://doi.org/10.1016/j.molcel.2022.01.024>, PMID: 35182480
- Moran BM, Payne C, Langdon Q, Powell DL, Brandvain Y, Schumer M. 2021. The genomic consequences of hybridization. *eLife* **10**:e69016. DOI: <https://doi.org/10.7554/eLife.69016>, PMID: 34346866
- Morrison GM, Brickman JM. 2006. Conserved roles for Oct4 homologues in maintaining multipotency during early vertebrate development. *Development* **133**:2011–2022. DOI: <https://doi.org/10.1242/dev.02362>
- Newport J, Kirschner M. 1982a. A major developmental transition in early *Xenopus* embryos: I. characterization and timing of cellular changes at the midblastula stage. *Cell* **30**:675–686. DOI: [https://doi.org/10.1016/0092-8674\(82\)90272-0](https://doi.org/10.1016/0092-8674(82)90272-0)
- Newport J, Kirschner M. 1982b. A major developmental transition in early *Xenopus* embryos: II. control of the onset of transcription. *Cell* **30**:687–696. DOI: [https://doi.org/10.1016/0092-8674\(82\)90273-2](https://doi.org/10.1016/0092-8674(82)90273-2)
- Ochi H, Kawaguchi A, Tanouchi M, Suzuki N, Kumada T, Iwata Y, Ogino H. 2017. Co-accumulation of cis-regulatory and coding mutations during the pseudogenization of the *Xenopus laevis* homoeologs six6.L and six6.S. *Developmental Biology* **427**:84–92. DOI: <https://doi.org/10.1016/j.ydbio.2017.05.004>, PMID: 28501477
- Owens ND, Blitz IL, Lane MA, Patrushev I, Overton JD, Gilchrist MJ, Cho KKY, Khokha MK. 2016. Measuring absolute RNA Copy Numbers at High Temporal Resolution Reveals Transcriptome Kinetics in Development. *Cell Reports* **14**:632–647. DOI: <https://doi.org/10.1016/j.celrep.2015.12.050>, PMID: 26774488
- Paraiso KD, Blitz IL, Coley M, Cheung J, Sudou N, Taira M, Cho KKY. 2019. Endodermal maternal transcription factors establish super-enhancers during Zygotic Genome Activation. *Cell Reports* **27**:2962–2977. DOI: <https://doi.org/10.1016/j.celrep.2019.05.013>, PMID: 31167141
- Phelps WA, Carlson AE, Lee MT. 2021. Optimized design of antisense oligomers for targeted rRNA depletion. *Nucleic Acids Research* **49**:e5. DOI: <https://doi.org/10.1093/nar/gkaa1072>, PMID: 33221877
- Phelps WA, Lee MT. 2023. Xl-Zga. swb:1:rev:9787c36788368d2d7003c973e8c27176e3832cbc. Software Heritage. <https://archive.softwareheritage.org/swb/1:dir:1cbb7eafcd2d739365d6af5a574501e4d877ef7;origin=https://github.com/MTLeeLab/xl-zga;visit=swb:1:snp:26285a8a149d3a54c47b8e6a443856593bae0848;anchor=swb:1:rev:9787c36788368d2d7003c973e8c27176e3832cbc>
- Picelli S, Björklund AK, Reinius B, Sagasser S, Winberg G, Sandberg R. 2014. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Research* **24**:2033–2040. DOI: <https://doi.org/10.1101/gr.177881.114>, PMID: 25079858
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**:841–842. DOI: <https://doi.org/10.1093/bioinformatics/btq033>, PMID: 20110278
- Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. 2014. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Research* **42**:W187–W191. DOI: <https://doi.org/10.1093/nar/gku365>, PMID: 24799436
- R Development Core Team. 2013. R: A language and environment for statistical computing. Vienna, Austria. R Foundation for Statistical Computing. <http://www.r-project.org/>
- Rice P, Longden I, Bleasby A. 2000. EMBOSS: the european molecular biology open software suite. *Trends in Genetics* **16**:276–277. DOI: [https://doi.org/10.1016/s0168-9525\(00\)00204-2](https://doi.org/10.1016/s0168-9525(00)00204-2), PMID: 10827456
- Scerbo P, Girardot F, Vivien C, Markov GV, Luxardi G, Demeneix B, Kodjabachian L, Coen L. 2012. Ventx factors function as Nanog-like guardians of developmental potential in *Xenopus*. *PLOS ONE* **7**:e36855. DOI: <https://doi.org/10.1371/journal.pone.0036855>, PMID: 22606298
- Schuff M, Siegel D, Philipp M, Bundschu K, Heymann N, Donow C, Knöchel W. 2012. Characterization of *Danio rerio* Nanog and functional comparison to *Xenopus* Vents. *Stem Cells and Development* **21**:1225–1238. DOI: <https://doi.org/10.1089/scd.2011.0285>, PMID: 21967637
- Session AM, Uno Y, Kwon T, Chapman JA, Toyoda A, Takahashi S, Fukui A, Hikosaka A, Suzuki A, Kondo M, van Heeringen SJ, Quigley I, Heinz S, Ogino H, Ochi H, Hellsten U, Lyons JB, Simakov O, Putnam N, Stites J, et al. 2016. Genome evolution in the allotetraploid frog *Xenopus laevis*. *Nature* **538**:336–343. DOI: <https://doi.org/10.1038/nature19840>, PMID: 27762356
- Shi X, Ng DW-K, Zhang C, Comai L, Ye W, Chen ZJ. 2012. Cis- and trans-regulatory divergence between progenitor species determines gene-expression novelty in Arabidopsis allopolyploids. *Nature Communications* **3**:950. DOI: <https://doi.org/10.1038/ncomms1954>, PMID: 22805557

- Sive HL**, Richard M. 2000. *Early Development of Xenopus Laevis: A Laboratory Manual* Cold Spring Harbor Laboratory Press.
- Skene PJ**, Henikoff S. 2017. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* **6**:e21856. DOI: <https://doi.org/10.7554/eLife.21856>, PMID: 28079019
- Skirkanich J**, Luxardi G, Yang J, Kodjabachian L, Klein PS. 2011. An essential role for transcription before the MBT in *Xenopus laevis*. *Developmental Biology* **357**:478–491. DOI: <https://doi.org/10.1016/j.ydbio.2011.06.010>, PMID: 21741375
- Song MJ**, Potter BI, Doyle JJ, Coate JE. 2020. Gene balance predicts transcriptional responses immediately following ploidy change in *Arabidopsis thaliana* *The Plant Cell* **32**:1434–1448. DOI: <https://doi.org/10.1105/tpc.19.00832>, PMID: 32184347
- Suyama M**, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research* **34**:W609–W612. DOI: <https://doi.org/10.1093/nar/gkl315>, PMID: 16845082
- Svoboda P**. 2018. Mammalian zygotic genome activation. *Seminars in Cell & Developmental Biology* **84**:118–126. DOI: <https://doi.org/10.1016/j.semcdb.2017.12.006>
- Swamy KBS**, Schuyler SC, Leu JY. 2021. Protein complexes form a basis for complex hybrid incompatibility. *Frontiers in Genetics* **12**:609766. DOI: <https://doi.org/10.3389/fgene.2021.609766>, PMID: 33633780
- Takahashi K**, Yamanaka S. 2016. A decade of transcription factor-mediated reprogramming to pluripotency. *Nature Reviews. Molecular Cell Biology* **17**:183–193. DOI: <https://doi.org/10.1038/nrm.2016.8>, PMID: 26883003
- Takebayashi-Suzuki K**, Arita N, Murasaki E, Suzuki A. 2007. The *Xenopus* POU class V transcription factor XOct-25 inhibits ectodermal competence to respond to bone morphogenetic protein-mediated embryonic induction. *Mechanisms of Development* **124**:840–855. DOI: <https://doi.org/10.1016/j.mod.2007.09.005>, PMID: 17950579
- Tirosh I**, Reikhav S, Levy AA, Barkai N. 2009. A yeast hybrid provides insight into the evolution of gene expression regulation. *Science* **324**:659–662. DOI: <https://doi.org/10.1126/science.1169766>, PMID: 19407207
- Vastenhouw NL**, Cao WX, Lipshitz HD. 2019. The maternal-to-zygotic transition revisited. *Development* **146**:dev161471. DOI: <https://doi.org/10.1242/dev.161471>, PMID: 31189646
- Veenstra GJ**, Destree OH, Wolffe AP. 1999. Translation of maternal TATA-binding protein mRNA potentiates basal but not activated transcription in *Xenopus* embryos at the midblastula transition. *Molecular and Cellular Biology* **19**:7972–7982. DOI: <https://doi.org/10.1128/MCB.19.12.7972>, PMID: 10567523
- Yanai I**, Peshkin L, Jorgensen P, Kirschner MW. 2011. Mapping gene expression in two *Xenopus* species: evolutionary constraints and developmental flexibility. *Developmental Cell* **20**:483–496. DOI: <https://doi.org/10.1016/j.devcel.2011.03.015>, PMID: 21497761
- Yang Z**. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Bioinformatics* **13**:555–556. DOI: <https://doi.org/10.1093/bioinformatics/13.5.555>
- Zhang C**, Basta T, Jensen ED, Klymkowsky MW. 2003. The beta-catenin/VegT-regulated early zygotic gene Xnr5 is a direct target of SOX3 regulation. *Development* **130**:5609–5624. DOI: <https://doi.org/10.1242/dev.00798>, PMID: 14522872
- Zhang Y**, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biology* **9**:R137. DOI: <https://doi.org/10.1186/gb-2008-9-9-r137>, PMID: 18798982

Appendix 1

Appendix 1—key resources table

| Reagent type (species) or resource | Designation | Source or reference | Identifiers | Additional information |
|---|--|--|-----------------------------------|--|
| Strain, strain background (<i>Xenopus laevis</i>) | Nasco wildtype | eNasco | Research Resource #NXR_0.0031 | |
| Antibody | Anti-H3K4me3 recombinant (rabbit polyclonal) | Invitrogen | Cat #711958 RRID #AB_2848246 | CUT&RUN (1:100) |
| Antibody | Anti-H3K4me3 recombinant (rabbit monoclonal) | Millipore | Cat #05-745 R RRID #AB_1587134 | CUT&RUN (1:100) |
| Antibody | Anti-H3K27ac recombinant (rabbit polyclonal) | Active Motif | Cat #39135 RRID #AB_2614979 | CUT&RUN (1:100) |
| Antibody | Anti-V5 (mouse monoclonal) | Invitrogen | Cat #R960-25 RRID #2556564 | CUT&RUN (1:100) |
| Chemical compound, drug | Isethionic acid | Sigma Aldrich | Cat #220078 | |
| Chemical compound, drug | Triptolide | Apexbio | Cat #50-101-1030 | |
| Chemical compound, drug | Cycloheximide | Sigma Aldrich | Cat #01810 | |
| Gene (<i>Xenopus laevis</i>) | <i>pou5f3.3.L</i> | RefSeq | NM_001088114.1 | Homeolog used for TF CUT&RUN |
| Gene (<i>Xenopus laevis</i>) | <i>sox3.S</i> | RefSeq | NM_001090679.1 | Homeolog used for TF CUT&RUN |
| Sequence-based reagent | <i>X. laevis</i> rRNA depletion oligomers | Phelps et al., 2021 | DOI: 10.1093/nar/gkaa1072 | |
| Recombinant DNA reagent | Tn5 (plasmid) | Addgene | Cat #112112 | |
| Recombinant DNA reagent | pA/G-MNase (plasmid) | Addgene | Cat #123461 | Purified enzyme gift from S. Hainer |
| Sequence-based reagent | Tn5ME-A | Picelli et al., 2014 | DOI: 10.1101/gr.177881.114 | |
| Sequence-based reagent | Tn5ME-B | Picelli et al., 2014 | DOI: 10.1101/gr.177881.114 | |
| Sequence-based reagent | Tn5MErev | Picelli et al., 2014 | DOI: 10.1101/gr.177881.114 | |
| Sequence-based reagent | Pou5f3.3 morpholino | GeneTools/ Morrison and Brickman, 2006 | DOI: 10.1242/dev.02362 | Targets both homeologs; GTACAATATGGGCTGGTCCATCTCC |
| Sequence-based reagent | Sox3 morpholino | GeneTools/ Zhang et al., 2003 | DOI: 10.1242/dev.00798 | Targets both homeologs; AACATGCTATACATTGGAGCTTCA |
| Sequence-based reagent | Pou5f3.2 morpholino | Genetools/ Takebayashi-Suzuki et al., 2007 | DOI: 10.1016 /j.mod.2007.09.005 | Targets both homeologs; AGGGCTGTTGGCTGTACATGGTGTC |
| Sequence-based reagent | GFP control morpholino | Genetools | | ACAGCTCCTCGCCCTTGCTACCAT |
| Sequence-based reagent | Hi-Fi F primer for Pou5f3.3.L ORF | This paper | | GGACAGCACGGGAGGCGGGGATCC GACCAGCCCATATTGTACAGCCAAAC |
| Sequence-based reagent | Hi-Fi R primer for Pou5f3.3.L ORF | This paper | | TATCATGTCTGGATCTACGTCTAGAT CAGCCGGTCAGGACCCC |
| Sequence-based reagent | F primer for Sox3.S ORF | This paper | | aaaggatcc TATAGCATGTTGGACACCGACATCA |
| Sequence-based reagent | R primer for Sox3.S ORF | This paper | | aaatctaga TTATATGTGAGTGAGCGGTACCGTG |
| Commercial assay, kit | Ultra II RNA library build kit | NEB | Cat #E7760 | |
| Commercial assay, kit | Ultra II DNA library build kit | NEB | Cat #E7645 | |
| Commercial assay, kit | RNA Clean and Concentrator-5 | Zymo | Cat #R1013 | |
| Software, algorithm | Bowtie2 | Langmead and Salzberg, 2012 ; http://bowtie2-bio.sourceforge.net/bowtie2 | DOI: 10.1038/nmeth.1923 | v2.4.2 |

Appendix 1 Continued on next page

Appendix 1 Continued

| Reagent type (species) or resource | Designation | Source or reference | Identifiers | Additional information |
|------------------------------------|--|---|--|---|
| Software, algorithm | Hisat2 | Kim et al., 2015 ; http://daehwankimlab.github.io/hisat2/ | DOI: 10.1038/nmeth.3317 | v2.0.5 |
| Software, algorithm | featureCounts | Liao et al., 2014 ; https://subread.sourceforge.net/ | DOI: 10.1093/bioinformatics/btt656 | v2.0.1 |
| Software, algorithm | SEACR | Meers et al., 2019 ; https://github.com/FredHutch/SEACR | DOI: 10.1186/s13072-019-0287-4 | v1.3 |
| Software, algorithm | MACS2 | Zhang et al., 2008 ; https://github.com/taoliu/MACS | DOI: 10.1186/gb-2008-9-9-r137 | v2.2.7.1 |
| Software, algorithm | BEDtools | Quinlan and Hall, 2010 ; https://bedtools.readthedocs.io/en/latest/ | DOI: 10.1093/bioinformatics/btq033 | v2.30.0 |
| Software, algorithm | DESeq2 | Love et al., 2014 ; https://bioconductor.org/packages/release/bioc/html/DESeq2.html | DOI: 10.1186/s13059-014-0550-8 | v4.0.3 |
| Software, algorithm | LiftOver | Kent et al., 2002 ; https://hgdownload.soe.ucsc.edu/downloads.html#utilities_downloads | DOI: 10.1101/gr.229102 | |
| Software, algorithm | BLAT | Kent et al., 2002 ; https://hgdownload.soe.ucsc.edu/downloads.html#utilities_downloads | DOI: 10.1101/gr.229102 | |
| Software, algorithm | Blast | Camacho et al., 2009 ; https://blast.ncbi.nlm.nih.gov/doc/blast-help/downloadblastdata.html | DOI: 10.1186/1471-2105-10-421 | v2.11.0+ |
| Software, algorithm | Samtools | Li et al., 2009 ; http://www.htslib.org | DOI: 10.1093/bioinformatics/btp352 | v1.12 |
| Software, algorithm | deeptools | Ramírez et al., 2014 ; https://github.com/deeptools/deepTools | DOI: 10.1519/JSC.0b013e3182a1f44c | v3.5.1 |
| Software, algorithm | LastZ | Harris, 2007 ; https://github.com/lastz/lastz | | v1.04.00 |
| Software, algorithm | Homer | Heinz et al., 2010 ; http://homer.ucsd.edu/homer/download.html | DOI: 10.1016/j.molcel.2010.05.004 | v4.11.1 |
| Software, algorithm | Paml | Yang, 1997 ; https://github.com/abacus-gene/paml | DOI: 10.1093/bioinformatics/13.5.555 | v4.9f |
| Software, algorithm | pal2nal | Suyama et al., 2006 ; http://www.bork.embl.de/pal2nal/ | DOI: 10.1093/nar/gkl315 | v14 |
| Software, algorithm | EMBOSS needle | Rice et al., 2000 ; https://emboss.sourceforge.net/download/ | DOI: 10.1016/s0168-9525(00)02024-2 | v6.6.0 |
| Software, algorithm | R | R core team, 2013; https://www.r-project.org | | v4.0.4 |
| Software, algorithm | Trim_galore | Martin, 2011 ; https://github.com/FelixKrueger/TrimGalore | DOI: 10.14806/ej.17.1.200 | v0.6.6 |
| Software, algorithm | Cutadapt | Martin, 2011 ; https://github.com/marcelm/cutadapt | DOI: 10.14806/ej.17.1.200 | v1.15 |
| Other | <i>X. laevis</i> genome | Xenbase; https://www.xenbase.org/xenbase/static-xenbase/ftpDatafiles.jsp | <i>X. laevis</i> v9.2 genome assembly | Genomic resource. v9.2 |
| Other | <i>X. laevis</i> genome | Xenbase; https://www.xenbase.org/xenbase/static-xenbase/ftpDatafiles.jsp | <i>X. laevis</i> v10.1 genome assembly | Genomic resource. v10.1 |
| Other | <i>X. laevis</i> gene models | Xenbase; https://www.xenbase.org/xenbase/static-xenbase/ftpDatafiles.jsp | <i>X. laevis</i> v9.2 gene models | Genomic resource. v9.2 |
| Other | <i>X. laevis</i> gene models | Xenbase; https://www.xenbase.org/xenbase/static-xenbase/ftpDatafiles.jsp | <i>X. laevis</i> v10.1 gene models | Genomic resource. v10.1 |
| Other | <i>X. laevis</i> page IDs | Xenbase; https://www.xenbase.org/xenbase/static-xenbase/ftpDatafiles.jsp | <i>X. laevis</i> v7.1 page IDs | Genomic resource. v7.1 |
| Other | CUT&RUN for histone marks, ATAC-seq, RNA-seq in <i>X. laevis</i> | This study | GEO #GSE207027 | High-throughput sequencing data |
| Other | mir-427 gene model | miRBase; Kozomara et al., 2019 | MI0001449 and MI0038331 | High-throughput sequencing data |
| Other | <i>X. laevis</i> wildtype stage 5 RNA-seq | Phelps et al., 2021 | GEO #GSE152902 | High-throughput sequencing data. SRR12758941; SRR12758940 |
| Other | <i>X. tropicalis</i> wildtype RNA-seq | Owens et al., 2016 | GEO #GSE65785 | High-throughput sequencing data. SRR1795666; SRR1795634 |
| Other | <i>X. tropicalis</i> morpholino and amanitin affected genes | Gentsch et al., 2019 | GEO #GSE113186 | High-throughput sequencing data |

Appendix 1 Continued on next page

Appendix 1 Continued

| Reagent type (species) or resource | Designation | Source or reference | Identifiers | Additional information |
|------------------------------------|--|--------------------------------|--------------------|--|
| Other | Zebrafish Pou/Sox/Nanog affected genes | Lee et al., 2013 | GEO #GSE47558 | High-throughput sequencing data |
| Other | <i>X. tropicalis</i> acetylated enhancers | Gupta et al., 2014 | GEO #GSE56000 | High-throughput sequencing data |
| Other | Zebrafish acetylated enhancers | Bogdanovic et al., 2012 | GEO #GSM915197 | High-throughput sequencing data |
| Other | Chains from <i>X. laevis</i> v2 to <i>X. tropicalis</i> v9 | UCSC genome browser | xenLae2ToXenTro9 | Genomic resource for liftover. 10% minimum sequence matching |
| Other | Chains from <i>X. tropicalis</i> v2 to <i>X. tropicalis</i> v7 | UCSC genome browser | xenTro2ToXenTro7 | Genomic resource for liftover. 90% minimum sequence matching |
| Other | Chains from <i>X. tropicalis</i> v7 to <i>X. tropicalis</i> v9 | UCSC genome browser | xenTro7ToXenTro9 | Genomic resource for liftover. 90% minimum sequence matching |
| Other | Chains from <i>X. tropicalis</i> v7 to zebrafish v10 | UCSC genome browser | xenTro7ToDanRer10 | Genomic resource for liftover. 10% minimum sequence matching |
| Other | Chains from zebrafish v10 to zebrafish v11 | UCSC genome browser | danRer10ToDanRer11 | Genomic resource for liftover. 90% minimum sequence matching |