

Genomics and gene regulation in allotetraploid *Xenopus* embryos

by

Wesley A. Phelps

Bachelor of Science, The Ohio State University, 2015

Submitted to the Graduate Faculty of the
Dietrich School of Arts and Sciences in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

University of Pittsburgh

2023

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Wesley A. Phelps

It was defended on

November 29, 2022

and approved by

Dr. Anne Carlson, PhD, Associate Professor, Dept. of Biological Sciences, Committee Member

Dr. Karen Arndt, PhD, Professor, Dept. of Biological Sciences, Committee Member

Dr. Andrea Berman, PhD, Associate Professor, Dept. of Biological Sciences, Committee
Member

Dr. Anne-Ruxandra Carvunis, PhD, Associate Professor, Dept. of Computational and Systems
Biology, Committee Member

Thesis Advisor: Dr. Miler T. Lee, PhD, Assistant Professor, Dept. of Biological Sciences,
Committee Chair

Copyright © by Wesley A. Phelps

2023

Genomics and gene regulation in allotetraploid *Xenopus* embryos

Wesley A. Phelps, Ph.D.

University of Pittsburgh, 2023

During early embryogenesis, key regulatory factors 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 pluripotency transition across animals, suggesting that pervasive regulatory remodeling over evolution has shaped the earliest stages of development; however it is unclear when and how such major modifications arose. For my dissertation, I have dissected the regulatory mechanisms underlying embryonic genome activation and pluripotency in the allotetraploid African-clawed frog, *Xenopus laevis*, and have additionally revealed key similarities in the pluripotency program across vertebrates despite many regulatory differences. First, I adapted two high-throughput sequencing methods for *X. laevis* embryos, which have traditionally been difficult to perform genomics on. These methods allow for genome-wide interrogation of gene expression and gene regulation using the limited number of cells present in early embryos. Second, I have demonstrated that the two distinct *X. laevis* “subgenomes” are activated independently to coordinate pluripotency induction, via subgenome-specific regulatory regions that drive differential gene expression. Maternal homologs of the mammalian pluripotency reprogramming factors OCT4 and SOX2 divergently activate the two subgenomes of *Xenopus laevis*. However, comparisons with other taxa reveal broad gene expression conservation despite high regulatory sequence turnover. Thus, my dissertation project both elucidates the previously uncharacterized mechanisms underlying genome activation for a long-established research model system and supports the existence of a core pluripotency program regulated by factors conserved across vertebrates.

Table of Contents

Preface.....	xii
1.0 Introduction.....	1
1.1 Pluripotency Induction	2
1.2 Genome Activation Reprograms a Differentiated Zygote to a Pluripotent Embryo	5
1.3 Maternal Clearance Facilitates Embryonic Reprogramming.....	7
1.4 <i>X. laevis</i> Offers Unique Insight into Genome Activation	9
1.5 Overall Goal	12
2.0 Genomic Techniques Adapted to Early Embryos.....	13
2.1 Optimized Design of Antisense Oligomers for Targeted rRNA Depletion	14
2.1.1 Materials and Methods	16
2.1.1.1 Animal Husbandry	16
2.1.1.2 Sample Collection	16
2.1.1.3 Total RNA Extraction	17
2.1.1.4 Antisense Oligo Design	18
2.1.1.5 RNaseH-mediated Depletion.....	20
2.1.1.6 Poly(A) Selection	21
2.1.1.7 Quantitative Reverse Transcription PCR (qRT-PCR)	22
2.1.1.8 RNA sequencing.....	23
2.1.1.9 RNA-seq Data Analysis	24
2.1.2 Method and Optimizations.....	25
2.1.3 Oligo Pool Efficiently Depletes the 28S rRNA.....	28

2.1.4 Oligo Pool Specifically Targets rRNA.....	29
2.1.5 rRNA Depletion and Poly(A)+ Selection Quantify Gene Expression Equivalently	31
2.1.6 Oligo Spacing and Melting Temperature Affect Depletion Efficiency	34
2.1.7 Oligo Pool Can Be Augmented to Target Multiple, Closely Related rRNAs	36
2.2 Cleavage Under Target and Release Using Nuclease (CUT&RUN) Facilitates Genome-wide DNA-binding Protein Mapping During Early Embryogenesis	42
2.2.1 Embryo Dissociation Increases Nuclei Binding to ConA Beads.....	43
2.2.2 CUT&RUN Maps Protein Binding Using Cell Numbers Amenable to Embryology.....	44
2.2.3 H3K4me3 CUT&RUN Coverage Correlates with Published ChIP-seq	44
2.3 Assay for Transposon Accessible Chromatin (ATAC) to Quantify Chromatin Accessibility During Genome Activation.....	45
2.3.1 Animal Cap Dissection Bypasses Yolk Inhibition of Tn5.....	46
2.3.2 Optimized Number of Animal Caps and Tn5 Concentration Facilitates Transposition.....	48
2.3.3 ATAC-seq at Early Developmental Stages Shows Open Chromatin at TSS's	48
2.4 Discussion	49
3.0 Hybridization Led to a Rewired Pluripotency Network in the Allotetraploid <i>Xenopus laevis</i>	52
3.1 Introduction	52

3.2 Materials and Methods	54
3.2.1 <i>Xenopus</i> Husbandry	54
3.2.2 Embryo Collection	54
3.2.3 RNA-seq Libraries	55
3.2.4 CUT&RUN	57
3.2.5 ATAC-seq	59
3.2.6 Transcriptomic Analysis.....	60
3.2.7 Chromatin Profiling Analysis	62
3.2.8 Motif Finding.....	64
3.2.9 Homeologous Enhancer Identification	64
3.2.10 Comparison with <i>X. tropicalis</i> and Zebrafish.....	65
3.3 Results.....	67
3.3.1 Homeologs are Activated Independently of Each Other	67
3.3.2 The microRNA <i>mir-427</i> is encoded on only one subgenome.....	69
3.3.3 Subgenomes Differ in Their Regulatory Architecture	73
3.3.4 Maternal Pluripotency Factors Differentially Engage the Subgenomes.....	75
3.3.5 The Ancestral Pluripotency Program is Maintained, Despite Enhancer Turnover	78
3.4 Discussion	86
4.0 Conclusions and Future Directions	90
4.1 Conclusions	90
4.2 Future Directions.....	91
Appendix A : XL rRNA oligo pool.....	96

Appendix B : XL test oligo pool.....	100
Appendix C : Zebrafish oligo pool	101
Appendix D : rRNA Depletion Protocol	107
Appendix E : <i>X. laevis</i> embryo CUT&RUN protocol	111
Appendix F : <i>X. laevis</i> embryo ATAC-seq Transposition Protocol.....	117
Appendix G : <i>X. borealis</i> and <i>X. muelleri</i> Husbandry	119
Appendix G.1 : General Guidelines	119
Appendix G.2 : Feeding	120
Appendix G.3 : Tank Cleaning.....	120
Appendix G.4 : hCG Preparation	121
Appendix G.5 : Amplexus.....	122
Appendix G.6 : Testis Dissection.....	123
Appendix G.7 : IVF and Dejellying	126
Bibliography	129

List of Tables

Table 1: Glossary	2
--------------------------------	----------

List of Figures

Figure 1-1: <i>Xenopus</i> natural history.	11
Figure 2-1: Oligo-ASST user interface	26
Figure 2-2: Oligo tiling strategy.....	27
Figure 2-3: Oligo pool specifically and efficiently depletes rRNA	29
Figure 2-4: rRNA depletion yields high quality libraries.....	30
Figure 2-5: Oligo pool can be easily augmented for addition targeting.....	33
Figure 2-6: Depletion efficiency is dependent on oligo spacing and melting temperature ..	36
Figure 2-7: Oligo pools can be augmented to deplete highly similar RNA species	39
Figure 2-8: RNA-seq methods are highly correlated in zebrafish.....	40
Figure 2-9: CUT&RUN and ATAC-seq optimizations	45
Figure 3-1: Identifying the first wave of genome activation across the two subgenomes. ...	68
Figure 3-2: Measuring genome activation	71
Figure 3-3: Homeologous genes are differentially activated in the early embryo.	72
Figure 3-4: The mir-427 locus.....	75
Figure 3-5: Differential homeolog activation is regulated by subgenome-specific enhancers	78
Figure 3-6: Profiling homeologous regulatory elements	80
Figure 3-7: Pou5f3.3 and Sox3 binding drives genome activation	82
Figure 3-8: Assessing Pou5f3 and Sox3 roles in genome activation	84
Figure 3-9: Regulatory divergence underlies dosage maintainence.....	85
Figure 3-10: Shared patterns of activation with other taxa	88

Figure 3-11: Model for Pluripotency Network Evolution	89
Figure 4-1: Additional <i>Xenopus</i> species	92
Figure 4-2 : <i>Xenopus</i> surgery diagrams	125

Preface

This dissertation is the product of six years of both professional and personal effort, but would not have been possible without the support of numerous people along the way. I have received advice and criticism from family, friends, and mentors alike and I am grateful to everyone who encouraged me onward even when I, myself, did not think it possible or desirable.

First, I'd like to thank my graduate mentor, Dr. Miler Lee. Despite his experience being primarily in fish, he accepted the challenge with me to take on a project in a system for which neither of us were initially prepared. However, he saw the potential this project had and fully supported and encouraged following this road wherever it led. Additionally, Miler put in a tireless and herculean effort into my personal development as a young scientist by persevering through both the expected challenges and unexpected conflicts we came across. At every turn of events, Miler always factored in my best interest, even when I acted contrarily.

Next, I would like to thank my dissertation committee: Anne Carlson, Karen Arndt, Andrea Berman, and Anne-Ruxandra Carvunis. They provided invaluable advice and critiques that shaped this project at every step and helped me grow as a scientist. I would also like to thank Sarah Hainer for her technical help with the CUT&RUN protocol. I would like to thank the Hainer, Arndt, and Kaplan labs for their valuable feedback.

Lastly, I would like to thank all of my friends and family, especially my incredible wife, Megan. Her unending support, belief in me, and love is a constant motivation and support, without which I never would have been able to complete this project. While I will never be able to repay the sacrifices she has made in my support, I look forward to trying for the rest of our lives together.

1.0 Introduction

Embryonic development relies on the ability of a single celled fertilized egg to give rise to a multicellular organism composed of many different cell types. As embryonic development progresses, a population of pluripotent cells emerges that can differentiate into all subsequent cell types. The potential cell types pluripotent cells can give rise to become progressively restricted as they reach their terminal cell type, a phenomenon Conrad Waddington famously depicted in 1957 as an initially undifferentiated ball rolling down a hill to its final differentiated state (1). This so-called Waddington's landscape model postulated that the terminal cell fate was permanent. However, in 1958, Sir John Gurdon demonstrated that differentiated cells can be artificially reprogrammed back to a pluripotent cell by implanting a differentiated nuclei into an enucleated egg (2), thus, demonstrating that the nucleus retains all the genetic instructions required for achieving pluripotency and that the cytoplasmic components of the egg can execute those instructions. Following Gurdon's successful somatic cell nuclear transfer (SCNT) experiments, Sir Martin Evans, Matthew Kaufman, and Gail Martin successfully established the first embryo-derived, pluripotent cell lines capable of generating all subsequent cell types (3, 4).

The establishment of embryonic stem (ES) cell lines was a major innovation in developmental biology, and it became evident that pluripotency depended on the specific gene expression profile of ES cells. Indeed, by fusing ES cells and somatic cells, the resulting heterokaryons were able to express pluripotency-associated genes (5). Thus, these studies suggested that pluripotent stem cells, such as ES cells, have the potential to reprogram somatic cells to a pluripotent state and imply the existence of one or more factors capable of reprogramming somatic cells. Evidence supporting the existence of reprogramming factors came from lineage

conversion studies in mammals by introducing a single transcription factor. The ectopic expression of MYOD alone is sufficient to transform mouse fibroblasts into myoblasts (6). Subsequent studies revealed ectopic expression of GATA1 converts myeloblasts to erythrocytes (7) and of CEBP α or CEBP β converts B lymphocytes to macrophages (8). Taken together, these studies demonstrate cell fates can be converted to different lineages via expression of only a few key transcription factors.

Table 1: Glossary

Diploid	A state in which the genome is comprised of two full copies.
Polyploid	A state in which the genome is comprised of more than two full copies.
Hybrid	An organism resulting from the cross fertilization of two species.
Autopolyploid	Ploidy in which every copy of the genome is derived from a single parental species.
Allopolyploid	Ploidy in which copies of the genome are derived from different parental species, usually two.
Homolog	Gene copies that are related to one another through ancestry from a common DNA sequence.
Paralog	Gene copies created by a local duplication within a single species.
Ohnolog	Gene copies created by a whole genome duplication within a single species.
Ortholog	Gene copies within <u>separate</u> species that resulted from speciation.
Homeolog	Gene copies within the <u>same</u> species that resulted from speciation.

1.1 Pluripotency Induction

During mammalian embryonic development, a single-celled zygote undergoes reductive cleavages for the first few days post fertilization, ultimately giving rise to a hollow ball of cells.

This blastocyst is composed of two distinct cell types: the inner cell mass (ICM) and the trophoblast. While the trophoblast will develop primarily into placenta, a mammalian-specific tissue, the ICM is composed of cells that each have the capacity to generate all the subsequent cell types of the adult organism, thus making them *pluripotent* (9). As development progresses, the ICM gives rise to a bilaminar embryonic disc composed of a further two cell types: the dorsal epiblast and the ventral hypoblast. The dorsal epiblast cells are thought to have a more restricted potency as they are unable to contribute to the resulting chimera after transplantation to a host blastocyte (10). Once the embryo begins gastrulation, the embryonic disc will give rise to the three primary germ layers, endoderm, ectoderm, and mesoderm, thus beginning the differentiation of all subsequent cell types.

Pluripotent stem cells can also be derived *in vitro* from embryonic and adult tissues upon explantation to culture, such as epiblast-derived stem cells (11, 12), embryonic germ cells (13, 14), and multipotent germline stem cells (15) in addition to ES cells. These cell lines all share their origins in early embryos or germ cell lineages, both of which harbor cells capable of spontaneously converting to a pluripotent state. Interestingly, another commonality among these cell types is the expression of the transcription factor OCT4. OCT4 along with SOX2 function as core transcription factors in mammalian pluripotency networks by regulating pluripotency-associated genes such as *FBX15*, as well as autoregulating their own expression (16–18). Soon after this discovery, various studies identified additional transcription factors that were also required for maintaining pluripotency in ES cells, such as *NANOG* (19, 20), *KLF4* (21), and *MYC* (22). However, the question remained as to whether any of these factors could reprogram somatic cells and induce pluripotency.

In 2003, Yamanaka and Takahashi performed an elegant screen to identify factors within a pool of 24 pluripotency-associated candidate genes based on the candidates' ability to activate a dormant drug resistance allele integrated into the *Fbx15* locus (16). Candidates that successfully induced the expression of this pluripotency gene in mouse fibroblasts could be reasonably expected to reprogram somatic cells into a pluripotent state. After successive rounds of elimination, four factors, Oct4, Sox2, Klf4, and c-Myc, were identified as the minimal combination required to induce pluripotency. Indeed, when these OSKM factors are ectopically expressed, they are able to induce pluripotency from both mouse (23) and human (24) fibroblasts. Thus, these landmark studies established the generation of induced pluripotent stem cells (iPSCs), a powerful tool in for biomedical studies and clinical use.

Around the same time, James Thomson's lab also reported generating iPSCs from human cell lines via ectopic expression of a specific transcription factor cocktail (25). Interestingly, Thomson's cocktail also included Oct4 and Sox2, but substituted Klf4 and c-Myc for Nanog and Lin28, demonstrating the central role that OCT4 and SOX2 have during pluripotency induction. Further studies revealed that of the OSKM factors, SKM were all redundant. SOX2 could be replaced by SOX1 or 3 and KLF4 could be replaced by KLF2 or 5, all of which were shown to have roles in mouse ES cell self-renewal (26, 27). In addition, it was demonstrated that the oncogenic transformation activity of L-MYC was much lower than that of other MYC family members, such as C-MYC (28). Therefore, the substitution of L-MYC allowed for the safer generation of human iPSCs for clinical application. Taken together, these subsequent studies would suggest the existence of an endogenous pluripotency network that is centered on a core set of transcription factors, such as OCT4, that is augmented by additional redundant mechanisms.

The culmination of this historic work is the establishment of iPSCs for use in biomedical research and clinical applications (29, 30). Many diseases are difficult to treat simply due to a lack of understanding of the mechanisms underlying the disease progression. Therefore, diseases need to be accurately modeled to develop an effective treatment aimed at the primary cause of the disease. This role has been traditionally fulfilled by animal models whose underlying biological functions closely resemble that of humans; however, no animal model perfectly replicates all aspects of a human disease. In addition to being a human-based model system, the primary benefit of using iPSCs is their inherent capability to self-renew and differentiate into all cell types in the human body, which can be harnessed to model a wide variety of diseases for study and drug discovery. In addition to disease modeling, iPSCs represent a powerful tool in regenerative medicine. With the help of iPSCs cultured in a lab, injured or degraded tissue could be replaced by transplanting artificially grown tissue into the injury or degeneration site (30). This bypasses potential immunorejection by the host as the new tissue can be grown from iPSCs reprogrammed from the host's own cells. Additionally, patients often die due to the lack of available donor tissue or organs, however, the use of iPSCs to artificially grow tissue could alleviate the limited availability of donor tissue and has the potential to obsolete the use of donor tissue entirely.

1.2 Genome Activation Reprograms a Differentiated Zygote to a Pluripotent Embryo

Following fertilization, the newly formed zygote is transcriptionally quiescent and must rely on a store of maternally provided factors to continue development. This maternal program carries out all cellular functions and ultimately acts upon the embryonic genome to drive *de novo* transcription, a phenomenon known as zygotic genome activation (ZGA) (31, 32). As development

progresses, these maternal factors are targeted for regulated decay as the new zygotic transcripts replace them, thus reprogramming the embryo from a maternally derived program to a zygotic program (31, 32). This maternal-to-zygotic transition (MZT) changes the embryo's fate from that of a terminally differentiated maternal state to a zygotic state that will eventually give rise to pluripotent stem cells, capable of developing into every cell type necessary for the adult.

The newly expressed transcripts from ZGA are necessary for the induction of pluripotent stem cells in animals. In mammals, ZGA occurs during the initial, slow embryonic cleavages and is 19 hours – 2 days post fertilization, for mouse and human, respectively (31), – the induction of pluripotent stem cells occurs much later, in the mammalian blastocyst driven by factors including OCT4, SOX2, and NANOG (29, 33). In contrast, more rapidly dividing species such as zebrafish, *Xenopus*, and *Drosophila*, activate their genomes within a few hours post fertilization, immediately leading to pluripotency induction (32, 34–36). Interestingly, maternally provided homologs of OCT4, SOX2, and NANOG are required for ZGA in zebrafish (37–39). Additionally, early gene expression in *Xenopus tropicalis* also relies on maternal homologs of OCT4 and SOX2 (40); however, the *Xenopus* clade does not have a direct homolog for NANOG (41, 42). Thus, vertebrate embryos rely on conserved pluripotency induction mechanisms at different times during early embryogenesis.

Outside of vertebrates, unrelated maternal factors drive genome activation and PSC induction, such as the factors Zelda (43), CLAMP (44, 45), and GAF (46) in *Drosophila*. These factors share many of the functional aspects of vertebrate pluripotency factors, including pioneering activity and establishing activating histone modifications (46–48). However, neither Zelda nor CLAMP have vertebrate homologs and the homolog of GAF does not have any known

role in pluripotency induction. This diversity of strategies implies extensive modifications to the gene network regulating pluripotency induction have occurred over evolutionary time (49, 50).

1.3 Maternal Clearance Facilitates Embryonic Reprogramming

During MZT, a subset of the initial maternal transcripts is programmatically decayed to facilitate embryonic reprogramming. The size of this maternal clearance can range from one-quarter of maternal transcripts, as in the case of zebrafish (51–54), to one-third in the case of mouse (55) and *C. elegans* (56), up to two-thirds in *Drosophila* (57). All animals undergo phasic maternal clearance, with the first phase being directed by maternally provided gene products and subsequent phases being directed by newly synthesized embryonic gene products. In *Drosophila*, a maternally directed wave of decay is triggered upon egg activation and does not require fertilization, followed by one or more waves that require genome activation (58–61). In total, 25% of the cleared transcripts are degraded strictly by the maternal machinery, 35% strictly through the zygotic machinery, and 40% show mixed decay effected by both maternal and zygotic mechanisms in *Drosophila melanogaster* (57, 61, 62).

In contrast, vertebrate maternal clearance is not as well dissected but still exhibit a phasic decay pattern. For example, during maternal decay in zebrafish, 60% of cleared transcripts are degraded between one-cell and sixteen-cell stage, prior to genome activation, whereas the remaining 40% of cleared transcripts are degraded coincident with genome activation (51, 54, 63). Similarly, in *Xenopus tropicalis*, 33% of maternal transcripts are cleared prior to ZGA (64), with another 15% are cleared after (65). Interestingly, some mammals, such as mice, exhibit an early wave of maternal clearance prior to fertilization after completing meiosis. After fertilization, a

second wave of clearance is triggered and a further third wave is triggered post ZGA (55, 66). The variable dependence on ZGA for different wave of clearance demonstrates different regulatory mechanisms underlying each wave of decay.

RNA-binding proteins (RBPs) have been shown to regulate the stabilization of stored maternal RNAs both during oogenesis and during the MZT. The specific RBPs that mediate stabilization vary across species, however the role of RBPs in maternal clearance has been well studied in *Drosophila*. Two RBPs, Smaug (61, 67) and Brain tumor (BRAT) (68) regulate RNA decay during the maternal wave, with BRAT and Pumilio (PUM) also regulating decay during subsequent decay waves post ZGA (57, 68–70). In vertebrates, however, certain classes of RBPs are shared among a wide array of species. For example, the Y-box RBP FRGY2 in *Xenopus laevis* (71, 72), Ybx1 in zebrafish (73), and MSY2 in mouse (74, 75) all bind maternal RNAs during oogenesis and/or early embryogenesis to translationally repress and stabilize the RNA. Y-box RBPs are also expressed during oogenesis in *Drosophila* and *C. elegans*, but no role in RNA stability has been identified (76–78). Another class of RBPs with a role in maternal clearance is AU-rich-elements binding proteins (ARE-BPs), which specifically bind sequence motifs comprised exclusively of adenosines and uracils. ARE-BPs have been demonstrated to have a role in destabilizing maternal RNAs during MZT in *C. elegans* (79–81), zebrafish (82), and *X. laevis* (65, 83–85), possibly in mouse as well (86). ARE motifs have also been shown to be enriched in degraded RNAs in *Drosophila*, however a role for ARE-BPs has not been identified yet (62). Interestingly, in *Xenopus*, ARE-BPs interact with another RBP, EDEN binding protein (EDEN-BP) to further regulate RNA stability prior to ZGA (65, 84). Thus, RBPs represent a major maternally derived regulatory mechanism in RNA stability during MZT.

In contrast, embryonically dependent maternal clearance waves frequently utilize micro RNAs (miRNAs) to target specific transcripts for deadenylation and, ultimately, degradation. Perhaps the most well studied example of this is the *miR-430* family of miRNAs in zebrafish. MiR-430 is embryonically expressed during ZGA and is required for the translational repression and deadenylation of up to 40% of the maternal transcripts that are cleared during MZT (52, 87). The ortholog, *miR-427*, is similarly required for deadenylation and transcript decay during *Xenopus* MZT (88). An analogous pathway exists in *Drosophila* utilizing the *miR-309* cluster, which is required for the degradation of 14% of late wave decay (89). However, *miR-309* and *miR-430/miR-427* do not share sequence homology, suggesting miRNA mediated maternal RNA clearance arose independently in vertebrates and invertebrates. Thus, miRNAs represent a major embryonic mode of maternal clearance during MZT.

1.4 *X. laevis* Offers Unique Insight into Genome Activation

Xenopus has long been used as a model to study the mechanisms of genome activation and pluripotency induction more broadly. As mentioned above, John Gurdon successfully performed the first SCNT experiments in *X. laevis*, not only demonstrating the ability of cells to be reprogrammed but also suggesting that the reprogramming factors were maternally contributed in the cytoplasm of the egg (2). Subsequently, in 1982, Newport and Kirschner published a pair of studies characterizing the timing and cellular changes during early *Xenopus* development (36, 90). Surprisingly, they found the onset of transcription could be induced prematurely by injecting non-specific DNA into the embryo (90), suggesting a model that utilized a titratable factor, likely histone proteins (91, 92), to prevent genome activation until a threshold DNA amount was reached.

The nuclear-to-cytoplasmic ratio model also suggests the existence of maternal activators that compete with the titratable factor for DNA occupancy.

However, those maternal ZGA activators have not been well established in the *Xenopus* clade, especially in *X. laevis*. Several studies have identified different maternal factors that have roles at the time of ZGA, but not a central role specifically to initiate broad transcription genome-wide. For example, the transcription factors Foxh1 and VegT are both maternally provided and have been shown to be active during MBT, however their pivotal functions are limited to mesendoderm specification in *X. tropicalis* (93–95). Another such example is that of the factors Otx2, Gsc, and Lhx1 which are expressed in the organizer of early *Xenopus* embryos and regulate a gene network required for mesendoderm specification that ultimately gives rise to head formation (96, 97). Finally, maternal Pou5 factors and Sox3 have been shown to act during ZGA to remodel embryonic chromatin and facilitate signaling competency in *X. tropicalis* (40, 98). These studies have identified numerous early embryonic transcription factors that each play a critical role in their respective regulatory networks, however these factors have not been shown to play a broad, central role in activating the embryonic genome or inducing pluripotency.

Additionally, much of the recent work in *Xenopus* ZGA studies has been conducted on sister species *X. tropicalis*, owing in part to the more complex genome of *X. laevis*. The ancestor of the *X. laevis* clade diverged from *X. tropicalis* ~48 million years ago (mya) (99). The *X. laevis* clade subsequently split ~36 mya into two, diploid *X. laevis* progenitors referred to as L and S (99). Subsequently, ~18 mya, the L and S progenitors hybridized, such that all the genomic material was retained in the hybrid, likely to maintain robust meiotic pairing (99–101) (Figure 1-1). The result of this hybridization gave rise to an allotetraploid organism in which the four sets of chromosomes are organized into two, diploid subgenomes – a long (L) and short (S) subgenome

due to the disproportionate pseudogenization and chromosomal deletions in the extant S subgenome (102, 103). It is this allotetraploidy that complicates genetic and genomic studies in *X. laevis*, but also has the potential to provide unique insights into gene expression regulation.

Allopolyploidy frequently induces acute effects on gene expression (103, 104) as a result of regulatory upheaval in the aftermath of such a dramatic genomic change. Over time, gene dosage imbalances between homeologous copies are reconciled due to both immediate and gradual regulatory shifts (105–107). This phenomenon has primarily been explored in plants due to their greater resilience to inter-species hybridization compared to animals (108–110). However, the

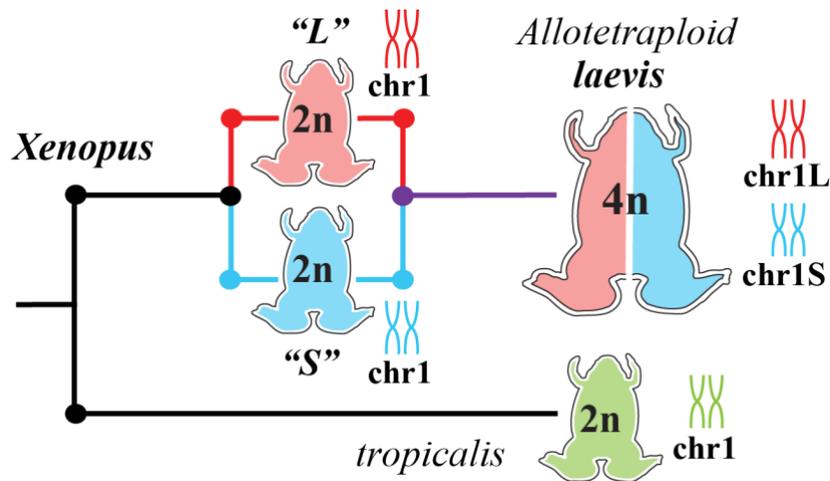


Figure 1-1: *Xenopus* natural history.

The allotetraploid *X. laevis* genome contains two distinct subgenomes “L” and “S” due to interspecific hybridization of ancestral diploids.

extent to which this has occurred in the few characterized allopolyploid vertebrates is unclear (111–113). Allotetraploid animals, such as *X. laevis*, allow for the study of gene expression regulation and transcriptome reprogramming using intra-specific comparative genomics; in the case of *X. laevis*, by comparing the highly similar L and S subgenomes.

1.5 Overall Goal

The goal of my dissertation is to characterize how the pluripotency regulatory network responds to dramatic, and often catastrophic, genomic upheavals to maintain function. To this end, I have conducted my dissertation in the model organism, *X. laevis*, which is uniquely positioned to provide valuable insight into pluripotency regulatory evolution. The allotetraploid genome allows for intra-specific transcriptomic and genomic comparisons between the L and S “species” (*i.e.* subgenomes) in the presence of a shared maternal contribution, thus negating any maternal variability in the comparison. In addition, the genome duplication post-hybridization provides a common regulatory upheaval event to both subgenomes, removing any variability in timing that would exist in two independent species. These comparisons will also elucidate the ZGA regulatory factors for a well-established, yet uncharacterized model organism. In chapter 2, I will describe three high-throughput genomic techniques I have developed and adapted for use in *X. laevis* embryos. These methods represent crucial tools necessary for analyzing gene expression regulation genome-wide in traditionally under-served taxa as well as early embryos more broadly, which typically are limited by low input material. In chapter 3, I will demonstrate that shared pluripotency factors drive differential genome activation through the use of subgenome-specific enhancers in the aftermath of allotetraploid hybridization. Thus, illustrating how gene regulatory networks accommodate regulatory disruptions. In chapter 4, I will summarize my conclusions and discuss further questions that have arisen from this project. Together, my project provides valuable genomic resources that I will use to elucidate when and how regulatory differences arise in the pluripotency regulatory network.

2.0 Genomic Techniques Adapted to Early Embryos

Portions of this chapter have been published in Phelps et. al. 2020 in *Nucleic Acids Research* (114). The Oligo-ASST Web tool is available at <https://mtleelab.pitt.edu/oligo>. Source code for the Web application and a command-line version of the program are available at <https://github.com/MTLeeLab/oligo-asst>. All raw sequencing reads are deposited in the NCBI Gene Expression Omnibus under accession number GSE152902.

In recent years, the advent of next-generation sequencing technology has facilitated a tremendous increase in the amount of DNA generated in a single study (115). Further developments and optimizations have dramatically reduced the cost of large data generation and rendered genome-wide studies feasible while increasing the statistical power provided by large data sets (115). As such, various methods have been developed in recent years to analyze numerous aspects of genome and transcriptome biology. By far, the most common technique to interrogate gene expression changes throughout the entire transcriptome is high-throughput RNA sequencing (RNA-seq), which is frequently used to create gene expression profiles transcriptome-wide in a wide variety of contexts (116–118). Other high-throughput techniques allow for the interrogation of chromatin biology by assessing protein binding, such as chromatin immunoprecipitation sequencing (ChIP-seq), or chromatin accessibility at a global, genome-wide level (119, 120). Such methods are frequently used in embryological studies to characterize organisms at different stages of development. However, early embryogenesis represents a unique obstacle to many high-throughput techniques. One of the more common limitations of high-throughput sequencing is the amount of input material required for a successful experiment (119, 120). The number of cells

required often translates to hundreds, even thousands, of embryos required for a given developmental stage. Thus, there is a need to develop low-input alternatives to many techniques. Another limitation is chemistry of certain methods is inadequate for many developmental situations or could be highly specific to a given taxa, limiting the propagation of that method despite its usefulness (40, 121). To that end, we sought to adapt three high-throughput sequencing techniques for use in *Xenopus* embryos: rRNA depletion RNA-seq, CUT&RUN as an alternative to ChIP-seq, and ATAC-seq.

2.1 Optimized Design of Antisense Oligomers for Targeted rRNA Depletion

High-throughput RNA sequencing (RNA-seq) has become a widespread method for measuring gene expression transcriptome-wide (116). Most RNA-seq studies focus on messenger RNA (mRNA); however, the vast majority of total RNA (>80%) (122, 123) is ribosomal RNA (rRNA). Therefore, RNA-seq is commonly paired with methods to reduce the amount of rRNA included in sequencing libraries, to maximize the proportion of sequencing reads derived from genes of interest.

An effective, widely used strategy for enriching mRNA is polyadenylate (poly(A)) selection (poly(A)+) (116). In eukaryotes, most mRNAs encode 3' poly(A) tails, which are used to select for and enrich mRNA pools using oligo(dT)-based methods (116, 124). However, many applications cannot take advantage of this approach, notably transcriptomics in prokaryotes, whose mRNA largely lack poly(A) tails (123), but also many eukaryotic contexts as well. Methods that aim to quantify message fragments separated from poly(A) tails, such as RNA-seq on degraded RNAs (125, 126), cap analysis gene expression (CAGE) (127), and ribosome profiling (128),

require alternate rRNA depletion strategies. Some RNAs of interest, such as nascent pre-mRNA, some histone mRNA and many non-coding RNAs (ncRNAs), do not encode poly(A) tails (129), thus their expression levels are underrepresented in poly(A)⁺ RNA-seq libraries. Finally, since poly(A) tail length is variable, it is a challenge to distinguish changes in poly(A) status, e.g. due to the activity of deadenylases, from changes in RNA molecule number using poly(A)⁺ RNA-seq (130). Indeed, in animals such as *Xenopus* and zebrafish, the maternal mRNA contribution to the egg is largely deadenylated (131), thus poly(A) selection is not well suited to accurately measure the transcriptome in the early embryo (132–134).

Antisense oligos can be used in conjunction with RNaseH to digest DNA–rRNA hybrids (125, 135). Several studies in both mammals and bacteria have shown that RNaseH-mediated rRNA depletion is efficient, resulting in sequencing libraries with minimal rRNA derived reads (125, 135–139). Commercial solutions have emerged for select taxa; however, the ease of this method allows it in principle to be readily adapted to any taxon, with the primary challenge being the design and acquisition of the 50 nt oligos that tile the specific rRNA sequences encoded in its transcriptome. Although rRNA sequences are generally well conserved between species relative to other genes, nucleotide differences and variable regions at even modest evolutionary distances (140) pose a challenge for reusing oligos designed for one taxon to effectively perform rRNA depletion in another.

To that end, we have developed an optimized strategy for RNaseH mediated rRNA depletion suitable for RNA-seq library construction that reduces up-front oligo costs by as much as 81%. Using *Xenopus laevis* and zebrafish (*Danio rerio*) embryos as test cases, we have demonstrated that depletion using short (39–40 nt) antisense DNA oligos sparsely tiled along rRNA, coupled with a 5-min digestion, effectively produces RNA-seq libraries with <5% rRNA-

derived reads, on par with poly(A) selection. We have shown that divergent rRNAs can be simultaneously digested with partially overlapping oligo pools that target regions of high sequence similarity, facilitating the design of flexible, cross-taxon reagents for rRNA depletion. Finally, we have developed a web tool, Oligo-ASST, that simplifies oligo design, allowing this approach to be easily adapted to any taxon or to target any other abundant RNAs for depletion.

2.1.1 Materials and Methods

2.1.1.1 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. *X. laevis* adults (NASCO NXR 0.0031) were housed in a recirculating aquatic system (Aquaneering) at 18°C with a 12/12 h light/dark cycle. Frogs were fed twice weekly with Frog Brittle (NASCO #SA05960(LM)M). *Danio rerio* (zebrafish) were housed in a recirculating aquatic system (Aquaneering) at 27°C with a 14/10 hour light/dark cycle and fed freshly hatched *Artemia* spp. nauplii twice daily, supplemented with TetraMin Tropical Flakes and dried krill.

2.1.1.2 Sample Collection

To obtain *Xenopus laevis* embryos, sexually mature females were injected with 1000 IU human chorionic gonadotropin into their dorsal lymph sac and incubated overnight at 16°C. In the morning, females were moved to room temperature where they laid eggs within an hour of being moved. Sexually mature males were euthanized by 30-minute submersion in 3.6 g/l tricaine-S (MS-222), pH 7.4, and testes were dissected. Cleaned testes were stored up to a week in L-15 medium at 4°C. Eggs were collected and artificially inseminated 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) (141). Zygotes were de-jellied (142) 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.

Zebrafish embryos were obtained from natural mating of TUAB strain fish 6–12 months old. Mating pairs were selected randomly from a pool of 24 males and 24 females ≥1 month since last breeding. Zebrafish were isolated in mating pairs overnight at room temperature in divided tanks. Dividers were removed the following morning, and eggs were collected in egg water (60 g/ml Ocean salt in RO water) and incubated at 28.5°C until the desired developmental stage.

To obtain fin clips, adult zebrafish were anesthetized in 500 mg/l MS-222 in system water for 2–5 min until gills stopped moving, then one lobe of the caudal fin was clipped. Fish were transferred to fresh system water for recovery.

2.1.1.3 Total RNA Extraction

For *X. laevis*, two embryos were pooled for RNA extraction; for *D. rerio*, 20 embryos or 10 fin clips were pooled. Samples were snap frozen in a 1.5 ml tube and homogenized with a pestle in 500 μl of TRIzol Reagent (Invitrogen #15596026) followed by 100 μl of chloroform. Tubes were centrifuged 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 –80°C for 1 h. Precipitated RNA was washed with cold 75% ethanol and resuspended in 50 μl of nuclease-free water. Concentration was determined by NanoDrop. RNA was stored at –80°C until use.

2.1.1.4 Antisense Oligo Design

The oligo tiling program is written in Python3. Each tiled oligo is defined by the start and end position of the complementary region in the target sequence (*e.g.* an rRNA). The algorithm assigns oligo positions left to right in a greedy fashion, such that each oligo is the maximum distance from the previous placed oligo while satisfying the parameter constraints – by default, melting temperature (T_m) between 70 and 80°C, length between 39–40 nucleotides (nts), and maximum untiled region of 30 nts. If no oligo exists that satisfies these constraints, the oligo with closest T_m to the ideal range is retained. The maximum untiled region is iteratively adjusted to take into account the remaining sequence length. Melting temperature is calculated using the nearest-neighbor method (143, 144) with RNA-DNA parameters (145): T_m (°C) = (H – helix initiation energy)/(S + R•ln(1/[oligo]) + 16.6 log₁₀([Na⁺]) – 273.15. This assumes oligo concentration is in excess of template, and is set to a conservatively low 50 mM compared to the oligo concentrations we use (100–400 mM), which will tend to slightly underestimate T_m by <2°C. Na⁺ concentration is set to 200 mM, and helix initiation energy for RNA-DNA hybrids is estimated as –3.1 kcal/°K•mol (145). At the time of writing, Banerjee et al. published improved RNA-DNA parameters (146), which may yield slightly different predicted T_m s compared to the old parameters used here; future software updates will incorporate these parameters. Once the entire target sequence is tiled, a second refinement phase adjusts each oligo position within the window defined by the upstream and downstream gaps, to yield maximized distances from upstream and downstream oligos within the ideal T_m range.

To find shared oligo pools between two or more unaligned target sequences, oligo tiling proceeds as above for the first sequence. For each subsequent sequence, oligos from the first set with exact complementary matches are selected, then the remaining untiled regions are subjected

to the tiling procedure as above. To find shared oligo pools between aligned target sequences, a consensus sequence from the alignment is used for the first round of oligo tiling to generate the candidate common oligos for subsequent rounds of tiling for each individual sequence. If wildcard bases are allowed, the consensus sequence will incorporate IUPAC wildcard bases. Wildcard-containing oligos are retained if the number of possible target sequences does not exceed the threshold specified by the user (*e.g.*, an oligo with two wildcard positions, R and Y, would target four different sequences encoding all combinations of C/T and A/G at the complementary positions, respectively). We ordered wildcard-containing oligos directly from the manufacturer as oligo mixtures, but they can also be ordered as individual unambiguous oligos and subsequently mixed.

Oligos for *X. laevis* rRNA were designed individually for 28S (X02995.1:3836–7917), 18S (X02995.1:1030–2854), 5.8S (X02995.1:3412–3573), 16S (M10217.1:3093–4723), and 12S (M10217.1:2205–3023). Aligned consensus oligos were designed for the maternal and somatic 5S (maternal: M10635:352–471, somatic: J01009.1:607–726) (*147*). The alternate mitochondrial rRNA sequences encoded in our samples are HM991335.1:1086–2726 for 16S and HM991335.1:69–1016 for 12S. COX2 and COX3 sequences were obtained from the *X. laevis* v9.2 genome assembly, chrM:9109–9796(+) and chrM:10711–11491(+) respectively. For zebrafish, aligned consensus oligos were designed for maternal and somatic 28S (chr4:77556054–77560323(-) and chr5:820029–824137(-) respectively), maternal and somatic 18S (chr4:77561203–77563141(-) and chr5:824921–826807(-) respectively), maternal and somatic 5.8S (chr4:77560653–77560810(-) and chr5:824488–824644(-) respectively), and maternal and somatic 5S (chr4:41890222–41890340(-) and chr18:30048558–30048676(-) respectively), according to previous annotations (*148*, *149*). Individual oligos sets were designed for 16S

(chrM:1020–1971(+)) and 12S (chrM:2043–3725(+)). All coordinates are from the GRCz11 genome build.

Oligos were ordered from Thermo Fisher as individual dry, desalted tubes at 25 nmol scale. At the time of writing, value oligo pricing (≥ 25 oligos with length ≤ 40 nts) was US\$4.64 per oligo, thus a full *X. laevis* set (137 oligos) would cost ~US\$636. In contrast, standard 50mer oligos are US\$19 each (without institutional discount), thus the 176 oligos required for full tiling would total US\$3344. With an institutional discount, this would likely still be >US\$1400.

2.1.1.5 RNaseH-mediated Depletion

Individual dry oligos were resuspended to 1000 μM . For *X. laevis*, a 10X working stock for nuclear rRNA (28S, 18S, 5.8S, maternal and somatic 5S) was created by pooling 1 μl of each of the 96 oligos and diluting to 4 μM per individual oligo (250 μl total volume, 384 μM total oligo concentration). At 1x concentration in 10 μl , each oligo is at 400 nM, which we estimate to be 10-fold in excess of its target in 1 μg of total RNA: assuming 80% of total RNA is derived from 40S rRNA (28S, 18S, 5.8S in equimolar amounts) and 28S rRNA is 2x the length of 18S+5.8S, this corresponds to ~530 ng of the ~4000-nt 28S rRNA, or ~41 nM in 10 μl . A similar stock of 41 oligos targeting the less abundant mitochondrial rRNA (16S, 12S) was prepared at 1 μM per individual oligo. For the parameter evaluation libraries, 15 of the original mitochondrial rRNA oligos, seven new mitochondrial rRNA oligos, seven COX2 oligos and nine COX3 oligos were pooled at 1 μM per oligo, and this stock was used with the original nuclear rRNA oligo stock for depletion. For zebrafish, separate working stocks for maternal nuclear (112 oligos at 4 μM per oligo), somatic nuclear (109 oligos at 4 μM per oligo), and mitochondrial rRNA (42 oligos at 1 μM per oligo) were similarly constructed. Maternal and somatic nuclear pools were then proportionally mixed

according to developmental stage (1:0 for two-cell, 1:1 for 28hpf, 0:1 for adult) (149). Hybridization procedure was based on *Adiconis et al.* (135) with slight modifications: 1 μ l of the nuclear pool (final concentration 0.4 μ M per oligo) and 1 μ l of the mitochondrial pool (final concentration 0.1 μ M per oligo) were combined with 1 μ g of total RNA (and optionally 150 ng of in vitro transcribed mCherry mRNA) in a 10 μ l buffered reaction volume (100 mM Tris-HCl pH 7.4, 200mMNaCl, 10 mM DTT), heated at 95°C for 2 min and cooled to 22°C at a rate of 0.1°C /s in a thermal cycler. Next, 10 U of thermostable RNaseH (NEB #M0523S) and 2 μ l of provided 10x RNaseH buffer were added and volume brought to 20 μ l with nuclease-free water. We achieved the best results with NEB thermostable RNaseH compared to other commercial RNaseH products. The reaction was incubated at either 45°C or 65°C for 5 or 30 min, then 5U of TURBO DNase (Invitrogen #AM2238) and 5 μ l of provided 10x buffer was added, volume brought to 50 μ l with nuclease-free water and incubated at 37°C for 30 min. Oligos were omitted from input control samples prior to heating and enzyme addition. For visualization, 12.5 μ l of each reaction was run on a 1% formaldehyde 1.2% agarose gel in MOPS buffer (10x stock: 200 mM MOPS, 50 mM NaAc, 10 mM Na₂EDTA, pH 7.0) at 80V. Gels were stained with SYBR Gold (Invitrogen #S11494) for 30 minutes. For qRT-PCR and RNA-seq, the reaction was purified and size selected to >200 nts using Zymo Clean and Concentrator-5 (Zymo #R1013) according to manufacturer's protocol, eluting in 10 μ l of nuclease-free water. RNA was stored at -80°C.

2.1.1.6 Poly(A) Selection

Polyadenylated mRNA was selected using the NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB #E7490L) according to the manufacturer's protocol: 1 μ g of total RNA was denatured at 65°C for 5 min then hybridized to buffered dT magnetic beads at room

temperature for 2 min. Selected RNA was eluted in 50 μ l of Tris buffer at 80°C for 2 min and rehybridized to the same beads for a second round of selection at room temperature for 2 min. Re-selected RNA was eluted in a final volume of 17 μ l of Tris buffer and stored at -80°C until further use.

2.1.1.7 Quantitative Reverse Transcription PCR (qRT-PCR)

For first strand synthesis, a 20 μ l reaction consisting of Zymo-cleaned RNA (~50 ng), dNTP (1 μ M), random primer (NEB #S1330S) (3 μ M), and DTT (10 mM) was incubated at 65°C for 5 min then transferred to ice for 2 min. 1 μ l of SuperScript III reverse transcriptase enzyme and 8 μ l of 5x buffer (Invitrogen #18080085) were added to a final reaction volume of 40 μ l, incubated at 42°C for 90 min, then heat inactivated at 70°C for 15 min. Initial samples for the *X. laevis* 28S qRT-PCR were column purified (Qiagen #28704) and used at full concentration for qRT-PCR; subsequent samples were used directly at 1:10 dilution for qRT-PCR based on the results of a 4-sample, 1:5 dilution calibration curve analysis. qRT-PCR was performed in triplicate using 10 μ l reactions (2.5 μ l of cDNA, 5 μ M of each forward and reverse primers, and 2x SYGreen mix (Genesee #17-505B)). qPCR was performed on QuantStudio 3 (Applied Biosystems) with an initial heat activation at 50°C for 2 min and then 95°C for 10 min. The reactions were cycled at 95°C for 15 s and 60°C for 1 min for 40 cycles. Specificity was determined via a 3-stage melt curve analysis conducted at 95°C for 15 s, dropped to 60°C for 1 min, and then raising the temperature from 60°C to 95°C at 0.1°C/s. No-template negative controls were run for each primer pair. Data analysis was conducted in Design and Analysis Application v1.5.1 (Thermo Fisher) and C_t values were calculated automatically from that application. Each NTC sample resulted in a $C_t > 34$. Experimental samples resulted in C_t values ranging between 14 and 33. ΔC_t values were

calculated from the average of three technical replicates for each sample using mCherry as the reference gene and plotted $\Delta\Delta C_t$ values represent depletion conditions ΔC_t over input RNA control ΔC_t . Statistical comparisons were done using two-tailed paired *t* tests on ΔC_t values (each treated sample is paired with the input RNA that was used for treatment). Primers were: **28S** (F-TGTGATTTCTGCCCAGTGCT; R-GACGAGGCATTTGGCTACCT, amplicon: 107 bp), **16S** (F-TCCAAAAACCTAGCATTCCAATTAT; R-TTTCATCTTTCCTTACGGTACTTTTTTC, amplicon: 140 bp), **mCherry** (F-GCCCCGTAATGCAGAAGAAG; R-TCAGCTTCAGCCTCTGCTTG, amplicon: 105 bp), **sub1.L – XM_018266533.1** (F-AGCAGGAGAAATGAAGCCAGG-exon 4; R-CCGACATCTGCTCCTTCAGT-exon 5, amplicon: 80 bp) (*150*); **helb.L – XM_018252426.1** (F-TTCCAGGGTTCAGAAGAGGAG-exon12/13 junction; R-TGCTATGGCTTCACCCAACCT-exon 13, amplicon: 148 bp); **nudt15.L – XM_018245539.1** (F-CCTGAGAAAAACGAAGGTTGGAA-exon3/4 junction; R-TGGATTGTAGCCTTGCTGCT-exon 4, amplicon: 105 bp). Primer specificity was verified using NCBI Primer-BLAST.

2.1.1.8 RNA sequencing

Strand-specific RNA-seq libraries were constructed using the 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 1.8x AmpureXP beads (Beckman #A63880), end repaired, then ligated to sequencing adaptors diluted 1:5. Ligated DNA was purified with 0.9x AmpureXP beads and PCR amplified for 8 cycles, then purified again with 0.9x AmpureXP beads. Libraries were verified by Qubit dsDNA high sensitivity (Invitrogen #Q32851) 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.

2.1.1.9 RNA-seq Data Analysis

RNA-seq reads were mapped to the *X. laevis* v9.2 or GRCz11 (zebrafish) genomes using HISAT2 v2.0.5 (151) (`-no-mixed -no-discordant`) and assigned to genes (Xenbase v9.2 models for *X. laevis* and Ensembl r99 for zebrafish) using featureCounts v1.5.1 (152) in reversely-stranded paired end mode with default parameters. To more accurately quantify rRNA levels in the *X. laevis* genome, due to poor assembly at the 40S rDNA locus, we additionally aligned to a separate HISAT2 index consisting of only the 40S (X02995.1) and 5S (J01009.1) sequences. Coverage plots were generated using BEDTools v2.25.0 genome-CoverageBed (153) and visualized on the UCSC Genome Browser (154). To annotate histone mRNA, *X. laevis* and zebrafish protein sequences were curated from HistoneDB 2.0 (155) and used to construct NCBI BLAST blastx databases (156). Xenbase and Ensembl zebrafish mRNA hits with $E\text{-value} < 1e-40$ were annotated as histones. To correlate oligo features with depletion efficiency, oligo positions or gap positions were converted to a genomic coordinate BED file and used to calculate RNA-seq coverage per feature using the BEDTools multicov command. To estimate oligo off targeting, we constructed an NCBI BLAST database of oligos and performed a gapless blastn (word size 5, reverse strand only, E-value threshold 100) querying every transcript in the transcriptome, retaining the best oligo hit per transcript. Bit score and number of identities were each plotted against log₂ TPM fold difference in rRNA depleted versus poly(A)⁺ samples, for all genes with TPM ≥ 1 in either sample. All plots and analyses were generated using R-3.4.4.

2.1.2 Method and Optimizations

The lack of an appropriate computational method to implement a gapped tiling strategy prompted us to build a Webtool called Oligo-ASST (which stands for Antisense Spaced Tiling) using the Python Dash v1.0 framework, available at <https://mtleelab.pitt.edu/oligo>. Oligo-ASST iteratively positions antisense oligos along a target sequence to maximize distance between consecutive oligos up to a threshold (e.g. 30 nts) while attempting to maintain a predicted T_m as close to the user defined target (e.g. 70°–80°C) as possible according to RNA/DNA duplex thermodynamic parameters (145). To design oligos, the user uploads one or more sequences in FASTA format (Figure 2-1A), selects oligo and gap length parameters according to their needs, then the resulting oligo sequences, coordinates and properties are displayed in the Web interface, where they can be downloaded in text format (Figure 2-1B, right). Tiled positions are also highlighted on a Dash Bio Sequence Viewer (Figure 2-1B, left). Pre-calculated oligo sets, including the ones presented here, are also provided for download. When multiple sequences are input, users can choose to design independent oligos per sequence or a shared set with common oligos targeting either identical subsequences or subsequences with one or more mismatches using wildcard bases. Sequences can be aligned beforehand using a tool such as MUSCLE (157) to improve identification of identical subsequences and yield a maximally compact oligo pool to target heterogeneous RNA species.

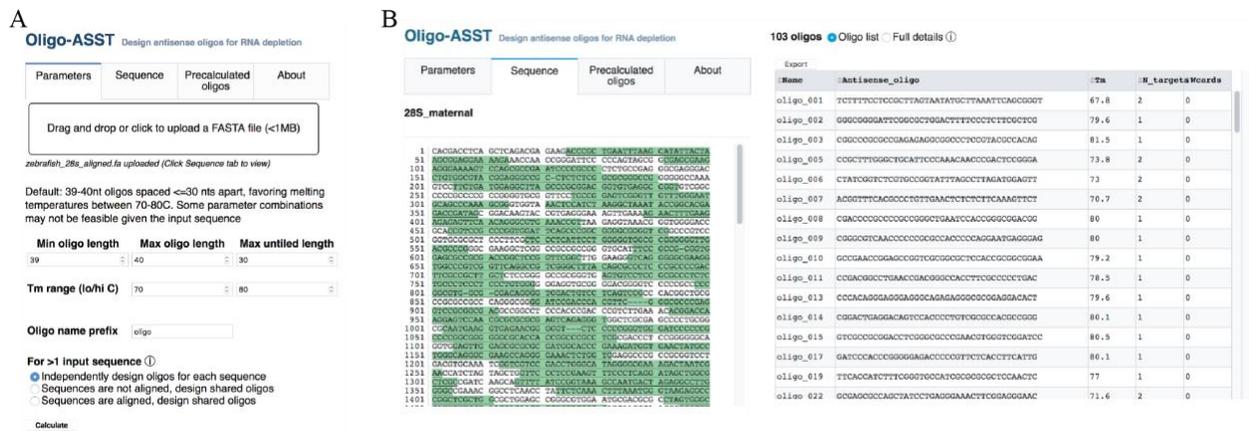


Figure 2-1: Oligo-ASST user interface

(A) The Oligo-ASST Web interface allows users to upload a FASTA file for target sequences and select parameters for oligo design. (B) Designed oligos are highlighted in a sequence viewer (left) and listed in the right pane in abbreviated form or with full details (not shown), which can be downloaded in text format.

Previous RNaseH-based depletion methods used 50-nt DNA antisense oligomers that completely tile target RNA species (125, 135, 137) (Figure 2-2A). We reasoned that for many applications, *e.g.* RNA-seq for non-degraded samples, tiling with gaps should be effective if the resulting fragments are short enough to be filtered out by size selection prior to cDNA generation. To test this strategy, we designed 39–40 nt oligos spaced ≤ 30 nt apart (Figure 2-3A) to tile the *X. laevis* nuclear (28S, 18S, 5.8S, 5S) and mitochondrial (16S, 12S) rRNA. The gaps would yield fragments ≤ 30 nts if the oligos induce digestion to completion, or ~ 70 nts with partial digestion if the flanking oligos each induce cleavage in the center (Figure 2-2B), comparable to the size of tRNAs. By including gaps, we were able to select oligos with high predicted melting temperatures (T_m), to ensure that they are hybridized with their targets at the digestion reaction temperature. With the aid of a computational tool we developed (see below), we designed most oligos with T_m between 65 and 87°C, with the exception of seven oligos targeting 16S rRNA with T_m between 58

and 64°C due to sequence constraints. By comparison, an end-to-end 50-nt tiling strategy would produce oligos with T_m between 52 and 94°C. Shortening the oligo lengths allowed us to take advantage of value oligo pricing, and the overall strategy used 137 individual oligos tiling 5434 total bases (Appendix A), compared to 176 oligos and 8639 bases (a 37% reduction) for the end-to-end 50-nt strategy. In our study, this led to an 81% reduction in total oligo cost according to list prices.

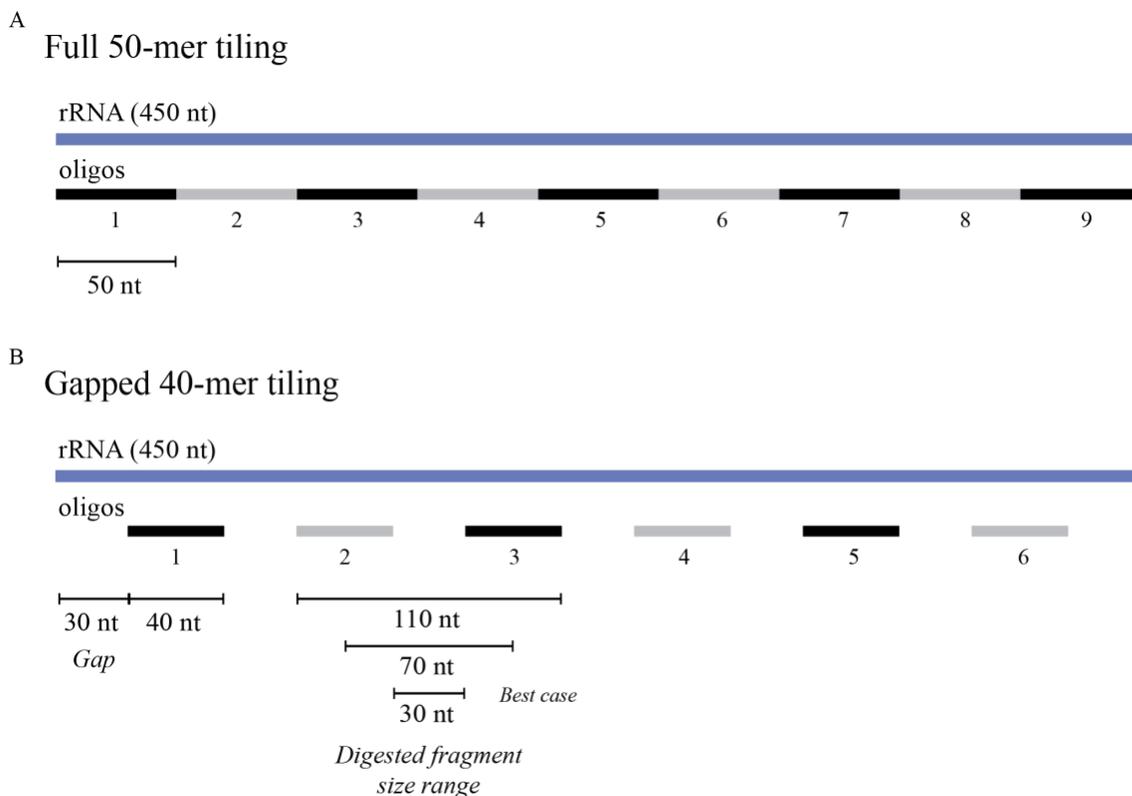


Figure 2-2: Oligo tiling strategy

A) Traditional design for rRNA depletion using 50mer antisense oligos that fully tile the target rRNA. B) A gapped tiling design using 40mer antisense oligos with 30 nt gaps uses fewer oligos to digest the target rRNA. The undigested fragments, optimally 30 nt in length, can be subsequently depleted from the sample using column-based size selection.

2.1.3 Oligo Pool Efficiently Depletes the 28S rRNA

At 1X, the oligo pools target ~1 µg of total RNA, such that each oligo is in ~10-fold excess of its rRNA target (Methods). To test the efficacy of the oligo pools, we subjected Nieuwkoop and Faber (NF) embryonic stage 0 *X. laevis* total RNA to RNaseH treatment and visualized the digested RNA, without any cleanup, on a 1% formaldehyde-agarose gel. We tested previously published reaction parameters (45°C for 30 min) using thermostable RNaseH along with two other conditions that reduced reaction time (45°C for 5 min) and additionally increased reaction temperature (65°C for 5 min) (Figure 2-3B). All three reaction conditions were effective, eliminating the upper bands corresponding to the 28S (4082 nts) and 18S rRNA (1825 nts) (Figure 2-3B). A large mass that is likely digested RNA and DNA oligos is visible at the bottom of each lane at <50 nts (Figure 2-3B), which is largely excluded due to size selection that is performed after digestion.

To precisely quantify the rRNA depletion, we subjected samples in triplicate to quantitative reverse transcription PCR (qRT-PCR) probing for 28S rRNA. All three depletion conditions significantly reduce the level of 28S rRNA compared to untreated RNA, with the 45°C/30 min and 65°C/5 min reactions reducing 28S rRNA levels by 99.99% ($P < 0.05$, two-tailed paired t test) (Figure 2-3C) – the optimal reaction temperature for the thermostable RNaseH is 65°C, and these results demonstrate that digestion is rapid at this temperature.

2.1.4 Oligo Pool Specifically Targets rRNA

To test specificity of the treatment for rRNA, we spiked 150 ng of *in vitro* transcribed mCherry mRNA into each reaction. All three reaction conditions eliminated the upper rRNA bands while leaving the mCherry (1037 nts) band intact (Figure 2-3B). A diffuse band migrating at ~500

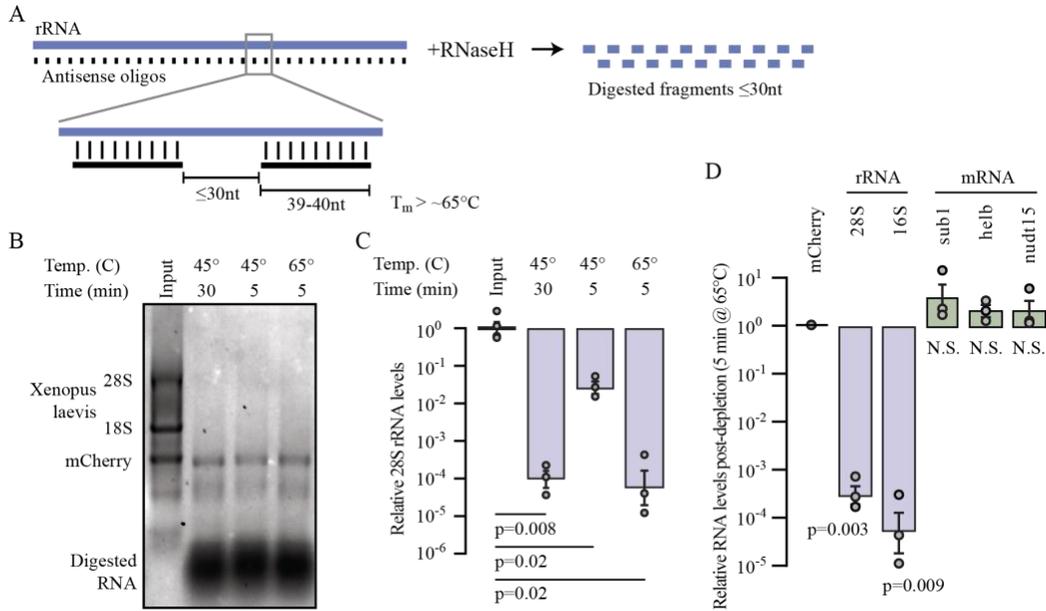
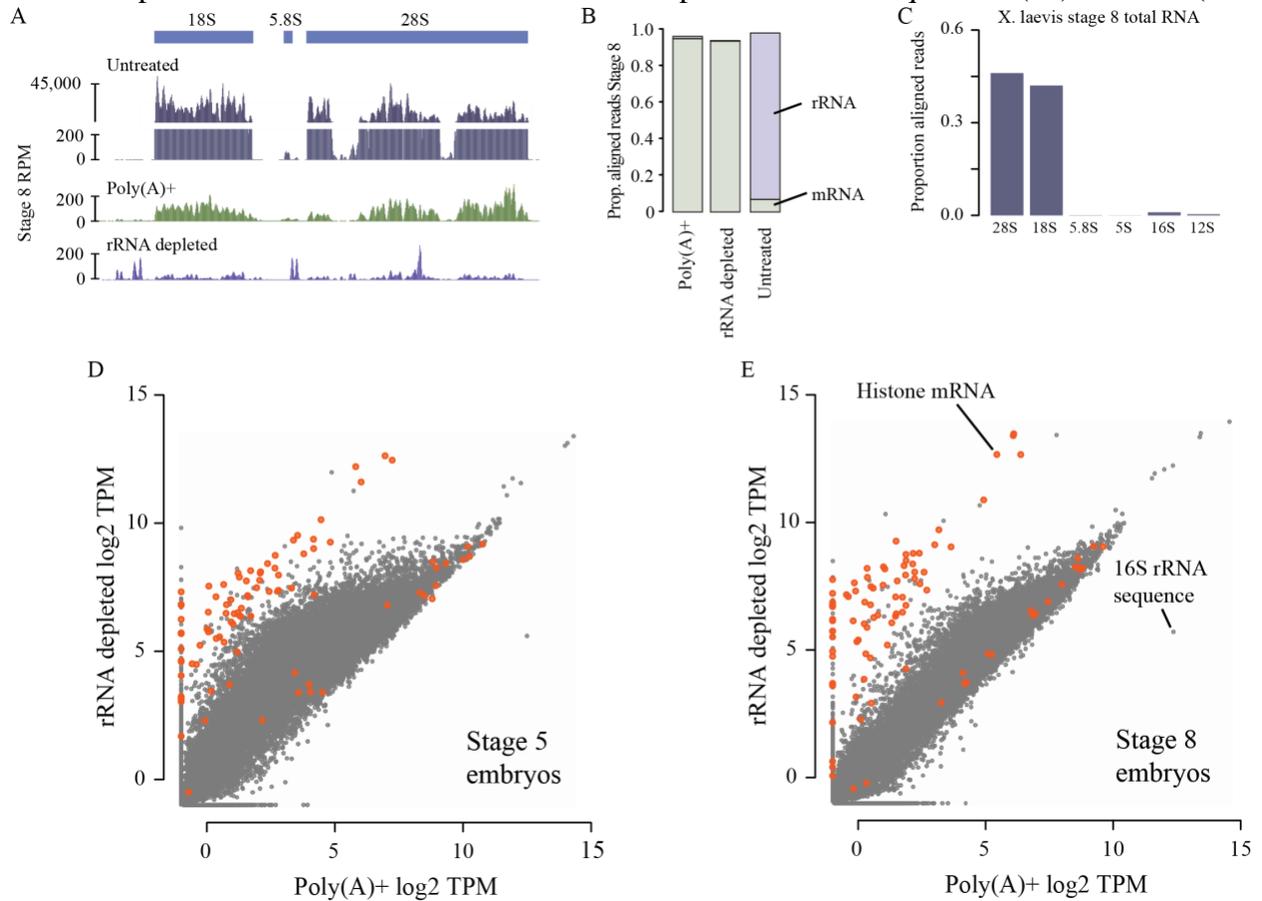


Figure 2-3: Oligo pool specifically and efficiently depletes rRNA

(A) Schematic of rRNA depletion strategy using 39–40 nt antisense oligos spaced ≤ 30 -nt apart. (B) *X. laevis* stage 0 total RNA (input, lane 1) and with rRNA depletion using different reaction conditions visualized on a 1% formaldehyde 1.2% agarose gel. *In vitro* transcribed mCherry mRNA was spiked into the input RNA prior to digestion. (C) qRT-PCR comparing 28S rRNA levels in *X. laevis* stage 0 total RNA (input, left) versus depletion conditions normalized to mCherry. *P* values are from two-tailed paired t tests comparing depleted samples to their corresponding total RNA input. (D) qRT-PCR measuring mCherry-normalized rRNA and mRNA levels in *X. laevis* stage 0 rRNA-depleted samples divided by levels in untreated samples. *P*-values are from two-tailed paired t tests for each gene comparing depleted samples to their corresponding total RNA input. N.S. = not significant.

nts is also intact in the digested samples, which likely corresponds to highly abundant histone mRNA species, based on inspection of RNA-seq datasets.

To assess the effects of rRNA depletion on mRNA as compared to rRNA, we performed qRT-PCR on treated (RNaseH 65°C/5 minutes) versus untreated total RNA, probing for embryonic mRNA expressed at low to moderate levels based on previous RNA-seq studies (99) – sub1.L (133



(A) Genome browser tracks comparing read coverage at the *X. laevis* 40S rDNA locus in untreated total RNA, poly(A)+ and rRNA depleted RNA-seq libraries from stage 8 embryos. Y-axis is discontinuous for the total RNA sample. (B) Stacked barplots showing proportion of aligned reads deriving from mRNA or lncRNA (green) versus rRNA (purple) in untreated, poly(A)+ and rRNA depleted RNA-seq libraries. (C, D) Biplots comparing log₂ TPM expression levels and poly(A)+ and rRNA-depleted libraries at stage 5 and 8, respectively. Histone genes are highlighted in orange. RPM = reads per million, TPM = transcripts per million.

transcripts per million (TPM)), *helb.L* (5 TPM), and *nudt15.L* (1 TPM) – along with 28S and mitochondrial encoded 16S rRNA, normalizing to mCherry spike in. Both rRNA species were significantly depleted in treated versus untreated samples ($P < 0.001$, two-tailed paired t test) (Figure 2-3D), while the mRNA levels were not significantly different ($P > 0.1$, two-tailed paired t test) (Figure 2-3D). Taken together, we find that the optimized oligo design effectively and specifically degraded targeted rRNA, using streamlined reaction times.

2.1.5 rRNA Depletion and Poly(A)+ Selection Quantify Gene Expression Equivalently

Next, we sought to determine whether our depletion strategy could be used to construct high quality RNA-seq libraries. We collected total RNA from two different *X. laevis* embryonic stages (NF 5 and 8) and performed either rRNA depletion (65°C/5 min) or poly(A)+ selection. We column size-selected the rRNA depleted samples to enrich for RNAs >200 nts, thus excluding digested rRNA fragments, then built Illumina strand-specific libraries and sequenced each sample to 5–10 million read pairs. Both the poly(A)+ and rRNA depleted samples show a >100-fold reduction in reads aligning the 40S rDNA locus compared to unselected total RNA (Figure 2-4A). Indeed, overall <0.3% of reads derive from rRNA while >90% of reads align to annotated mRNA or long non-coding RNA (lncRNA) in both the poly(A)+ and rRNA depleted samples, compared to <7% mRNA reads for unselected total RNA (Figure 2-4B,C). This efficiency is comparable to previous rRNA depletion strategies using full 50-nt end-to-end oligo tiling (125, 135, 137), demonstrating that gapped oligo tiling can achieve highly efficient rRNA depletion.

Transcriptome wide, expression levels correlate well between poly(A)+ and rRNA depletion for most genes (Figure 2-4D,E). However, at stage 5, a population of transcripts shows elevated apparent levels with rRNA depletion compared to poly(A)+ (Figure 2-4D). Indeed, the

maternal RNA contribution to the egg is largely deadenylated, with poly(A) tails lengthening during early embryonic stages through cytoplasmic polyadenylation (131). Thus, rRNA depletion avoids the depressed expression levels arising from inefficient capture of mRNA with short poly(A) tails, typical of poly(A)+ RNA-seq (53, 132–134). By the mid-blastula transition (NF stage 8), poly(A) tails are longer, so poly(A)+ and rRNA depletion yield comparable expression

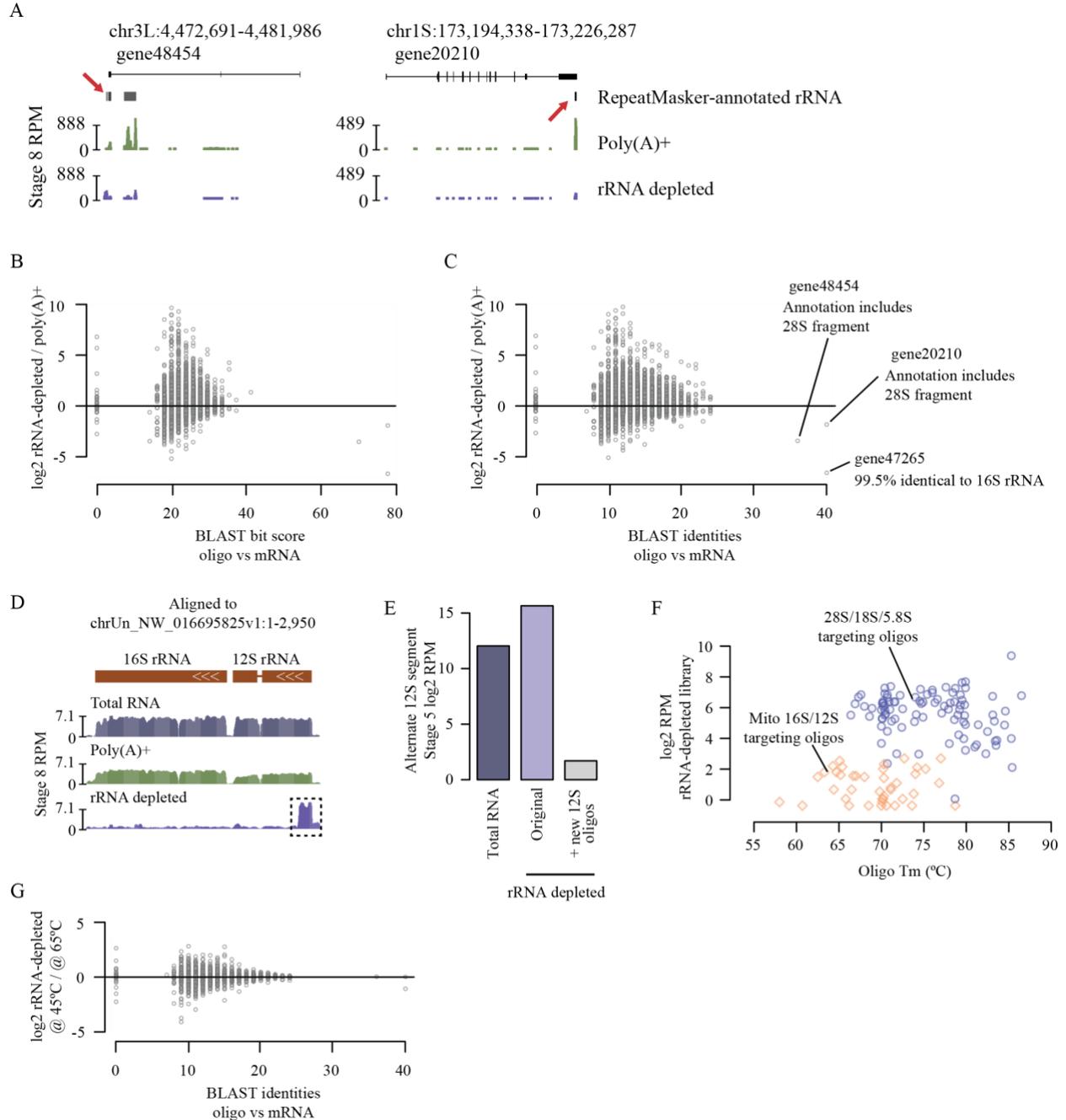


Figure 2-5: Oligo pool can be easily augmented for addition targeting

A) Bar plots showing proportion of aligned reads from untreated total RNA mapping to rRNA species in *X. laevis*. B) Genome browser tracks illustrating rRNA sequences as annotated by RepeatMasker (red arrows) falling within annotated exons of two genes. C,D) Biplots showing the BLAST bit score (C) and number of identities (D) of each *X. laevis* mRNA to the most similar oligo sequence in the depletion pool, versus the log₂ fold expression difference between the stage 8 rRNA-depleted and poly(A)+ samples. Three mRNA have very high similarity to oligos, which all contain sequences identical to rRNA, suggesting spurious annotations. E) Genome browser track of an unassembled *X. laevis* scaffold that harbors the sequences of alternate mitochondrial rDNA locus. The alternate 12S gene encodes a divergent 5' end compared to the reference 12S gene, which fails to be depleted by the original oligo pool. F) Augmenting the original oligo pool with two oligos that target the alternate 12S 5' divergent end effectively depletes this sequence. G) Biplot comparing oligo melting temperature (T_m) to log₂ RPM coverage of the region targeted by each oligo in the *X. laevis* stage 8 rRNA depleted sample. Blue circles are nuclear rRNA-targeting oligos, orange diamonds are mitochondrial rRNA-targeting oligos. H) Same as (D) except comparing depletion at 45°C to depletion at 65°C. RPM = reads per million.

values for these mRNA (Figure 2-4E). However, some RNA species are still better represented in the rRNA depletion libraries, suggesting these transcripts lack poly(A) tails. Indeed, replication-dependent histone mRNA encode 3' stem loops instead of poly(A) tails (158), and these transcripts are much more efficiently sequenced with rRNA depletion (Figure 2-4D,E). Thus, our optimized rRNA depletion strategy effectively quantifies expression levels of both the adenylated and non-adenylated transcriptome.

To determine if the antisense oligos were inducing off-target depletion of mRNA, we performed BLAST alignment for each oligo against the annotated transcriptome and found that the majority of transcripts have low sequence similarity to any rRNA-targeting oligo: 99.4% of

transcripts have <50% sequence identity to any oligo. However, we did identify three predicted transcripts that have high similarity to rRNA sequences, which are likely falsely annotated mRNA (Figure 2-5A). We additionally found no correlation between sequence similarity to an oligo and RNA-seq fold difference between the rRNA-depleted and poly(A)+ samples, whereas if significant off-target digestion were occurring, we would expect higher BLAST similarity to be correlated with a lower sequenced expression level with rRNA depletion compared to poly(A)+ (Figure 2-5B,C). Thus, antisense-oligo depletion using 39–40mers can achieve high efficiency and on-target specificity.

2.1.6 Oligo Spacing and Melting Temperature Affect Depletion Efficiency

To further evaluate antisense oligo design parameters, we sequenced additional libraries using variations to our original depletion strategy. First, we observed that our *X. laevis* seemed to be encoding a mitochondrial variant that differs from the xenLae 9.2 reference genome. However, the alternate sequence spanning the 12S and 16S rRNA genes was included in the genome sequence as a scaffold (chrUn_NW_016695825v1). Although the 12S and 16S sequences are each >97% identical between the variants, the alternate 12S gene has a divergent 5' portion that was not targeted by our original oligo pool (Figure 2-5D). By adding two new oligos to the depletion reaction, we effectively depleted the alternate 12S rRNA sequence in a new library (Figure 2-5E).

To measure the effect of oligo spacing on depletion efficiency, we selectively omitted oligos from the 12S and 16S rRNA targeting pool to leave gaps of up to 215 nts untiled by oligos. We additionally tiled the 3' section of 16S rRNA end-to-end with 50mer oligos to resemble a traditional design strategy (Appendix B). We found that increasing the untiled gap size is correlated with poorer depletion efficiency (*i.e.*, higher read depth), demonstrating that larger

digested RNA fragments are less efficiently excluded during column clean up (Figure 2-6A,B). Conversely, there was a negligible difference in efficiency when gaps were eliminated, suggesting that a tiling strategy with moderate gaps (≤ 30 nts) is comparable to end-to-end tiling for rRNA-depleted sequencing libraries when paired with size selection (Figure 2-6A,B).

Finally, in our original depletion we used oligos with melting temperatures near or above the 65°C reaction temperature, and we observed no correlation between oligo T_m and depletion efficiency (Figure 2-5F). To further explore the relationship between oligo T_m and reaction temperature on depletion, we targeted two highly abundant mRNA, COX2 and COX3 with high- T_m ($67.2^\circ\text{C} \leq T_m \leq 71^\circ\text{C}$) and low- T_m ($47.4^\circ\text{C} \leq T_m \leq 55.2^\circ\text{C}$) oligos respectively (Appendix B). Performing digestion at 65°C for 5 min yielded strong depletion of COX2 but less efficient depletion of COX3, suggesting that the low- T_m oligos are poorly hybridized to their targets at 65°C (Figure 2-6C). However, lowering the reaction temperature (45°C for 30 min) improved COX3 depletion (Figure 2-6C,D) with no other effects on the rest of the transcriptome (Figure 2-5G). Thus, better depletion is achieved using oligos with melting temperatures at or above the digestion reaction temperature.

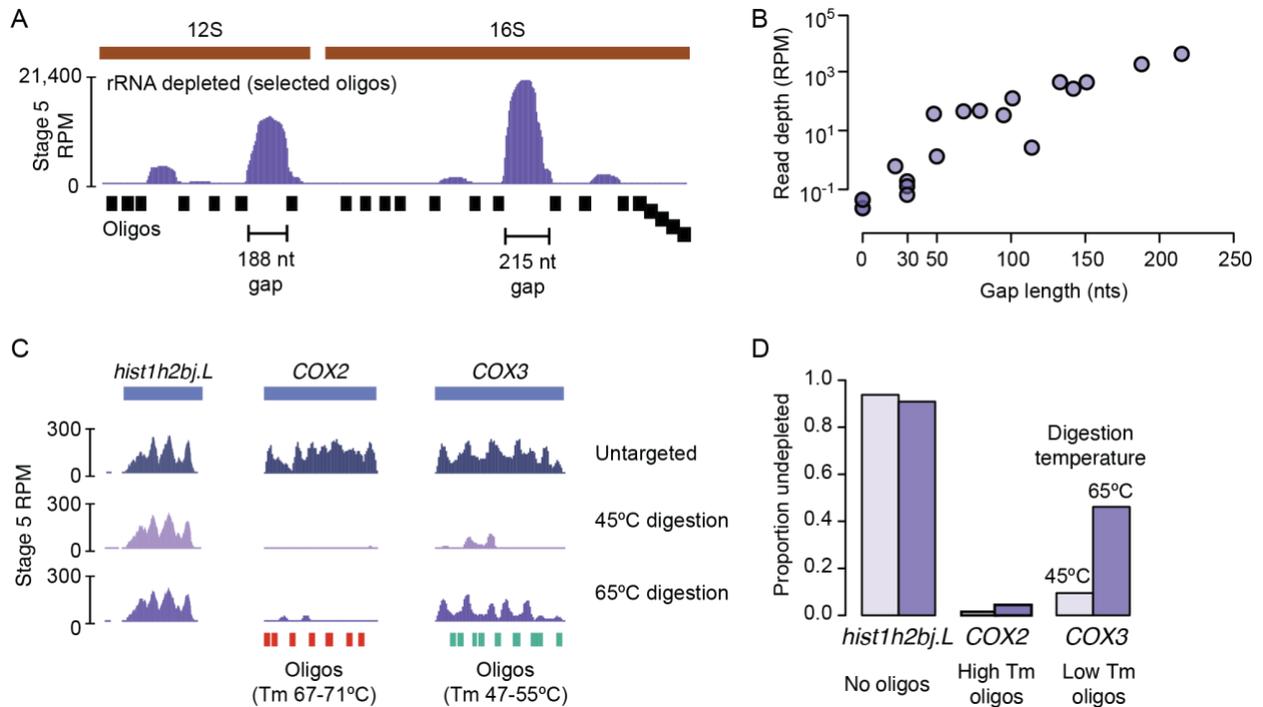


Figure 2-6: Depletion efficiency is dependent on oligo spacing and melting temperature

(A) Genome browser track showing read coverage over *X. laevis* 12S and 16S rDNA, illustrating the effect of rRNA depletion using variably spaced oligos (bottom). The two largest untiled regions are highlighted. (B) Biplot showing sequencing read depth at the center of each untiled region (gap) over 12S and 16S as a function of length of the region. (C) Genome browser tracks showing read coverage for *hist1h2bj.L*, not targeted for depletion; *COX2*, targeted using oligos with high melting temperature (T_m); and *COX3*, targeted using oligos with low T_m . Depletion reactions at 45°C for 30 min and 65°C for 5 min are compared to a reaction lacking the gene-targeting oligos. (D) Barplot for the three genes as in (C) showing the ratio of RNA-seq transcripts per million in depletion reactions (45°C left bars, 65°C right bars) over the non depleted condition. RPM = reads per million.

2.1.7 Oligo Pool Can Be Augmented to Target Multiple, Closely Related rRNAs

Given the gapped design strategy, it is likely that some sequence differences in target RNAs would be tolerated, allowing oligo pools designed for the rRNAs of one taxon to be used for

another closely related taxon. At greater sequence dissimilarity, we reasoned that shared oligos could be designed to target common subsequences between two or more RNAs, with gaps positioned over variable regions, avoiding the need to design completely separate reagents for rRNA depletion.

To test this, we designed a combined oligo pool to target the two versions of the zebrafish nuclear rRNAs, which are 86% similar. Zebrafish encode maternal-specific 28S, 18S, 5.8S and 5S rRNAs that are deposited into eggs during oogenesis (148, 149). After zygotic genome activation, distinct somatic rRNAs begin to be transcribed and slowly replace the maternal versions as the embryo develops (Figure 2-7A). Thus, to effectively deplete rRNAs in zebrafish embryos, both versions would need to be targeted. We aligned each rRNA sequence pair and designed 46 oligos that target identical regions between the maternal and somatic versions. To target regions that differed at only one position, we additionally designed 22 oligos containing a wildcard base (*e.g.* R to represent either A or G), which we ordered as mixtures of two oligos (Figure 2-7B). Finally, 44 and 41 additional oligos were required to target divergent maternal and somatic regions, respectively. In all, the combined design required 153 total oligos to together target both sets of nuclear rRNAs (Appendix C), compared to 201 total oligos for two independent sets.

We combined the common and unique oligos to create separate maternal and somatic pools each at 4 μ M per individual oligo. We also created a mitochondrial rRNA targeting pool at 1 μ M per oligo (there is only one known version of the 16S and 12S rRNAs). In the 2-cell stage embryo, the maternal+mitochondrial pools effectively and specifically induce rRNA depletion from total RNA, which is entirely maternally derived (Figure 2-7C, left), while in adult fins, the somatic+mitochondrial pools are effective (Figure 2-7C, right). We additionally tested depletion in 28 h post fertilization (h.p.f.) embryos, which express roughly equal amounts of maternal and

somatic rRNA (*I49*). Neither the maternal pool nor the somatic pool alone was as effective as a 1:1 mixture of both pools: using only the maternal or somatic pools produced several RNA species between 300 and 800 nts, suggesting incomplete digestion (Figure 2-7C, middle).

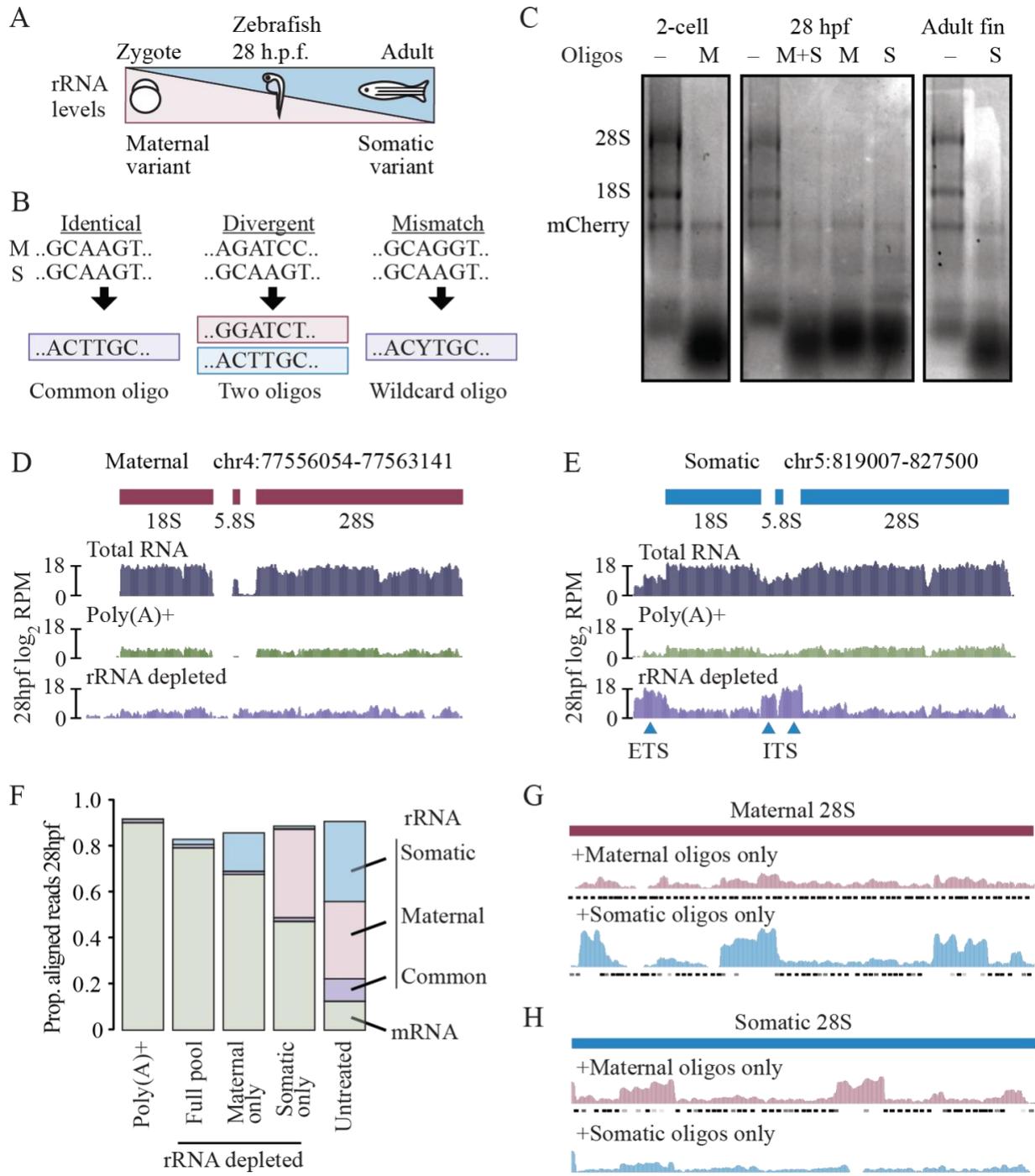


Figure 2-7: Oligo pools can be augmented to deplete highly similar RNA species

(A) Diagram illustrating the relative expression of the maternal and somatic nuclear rRNA variants over development. h.p.f. = hours post fertilization. (B) Schematic showing how oligos (bottom) can target similar sequences (top) between two RNAs. (C) Gels showing total RNA samples after rRNA depletion across three zebrafish developmental timepoints using only the maternal pool (M), only the somatic pool (S), or a mixture of the two pools (M+S), as compared to untreated input (-). In vitro transcribed mCherry mRNA was spiked into the input RNA prior to digestion. (D, E) Genome browser tracks comparing read coverage at the maternal (D) and somatic (E) 45S rDNA loci in untreated, poly(A)+ and rRNA depleted libraries from zebrafish 28 h.p.f. (F) Stacked barplots showing proportion of aligned reads deriving from mRNA (green) versus rRNA (blue, uniquely somatic; pink, uniquely maternal; purple, common) in untreated, poly(A)+ and rRNA depleted RNA-seq libraries. (G, H) Genome browser tracks comparing read coverage at the maternal (G) and somatic (H) 28S rDNA loci in rRNA depletion libraries depleted using only maternal (top row) or only somatic (bottom row) oligo pools from zebrafish 28 h.p.f. Targeted regions by each oligo pool are shown beneath each track.

To quantify this difference in efficiency, we constructed RNA-seq libraries at 28 h.p.f. rRNA depletion with the combined oligo pool effectively reduced the number of sequencing reads mapping to either the maternal or somatic rRNA loci compared to untreated total RNA, comparable to poly(A)+ (Figure 2-7D,E, Figure 2-8A). We did observe elevated levels of reads mapping to the external and internal transcribed spacers of the full somatic 45S transcript (5' ETS and two ITS regions; Figure 2-7E), which were omitted from the oligo design; as well as a small region of 16S rRNA where targeting was less efficient (Figure 2-8B-D). Nonetheless, 79% of reads mapped to mRNA or lncRNA, compared to 90% for poly(A)+ (Figure 2-7F), and expression quantification was highly correlated between the two methods (Figure 2-8E). rRNA depletion additionally recovered highly expressed non-coding RNAs such as the signal recognition particle and 7SK RNAs, which are not efficiently sequenced with poly(A)+ (Figure 2-8E).

In contrast, rRNA depletion using only the maternal or somatic pools was less efficient. By targeting only the maternal rRNA, 15% of the library is still rRNA, mapping to the somatic 45S locus; and by targeting only the somatic rRNA, 38% of reads derive from rRNA, corresponding to the maternal 45S locus. This leaves only 68% and 47% of the library mapping to mRNA+lncRNA,

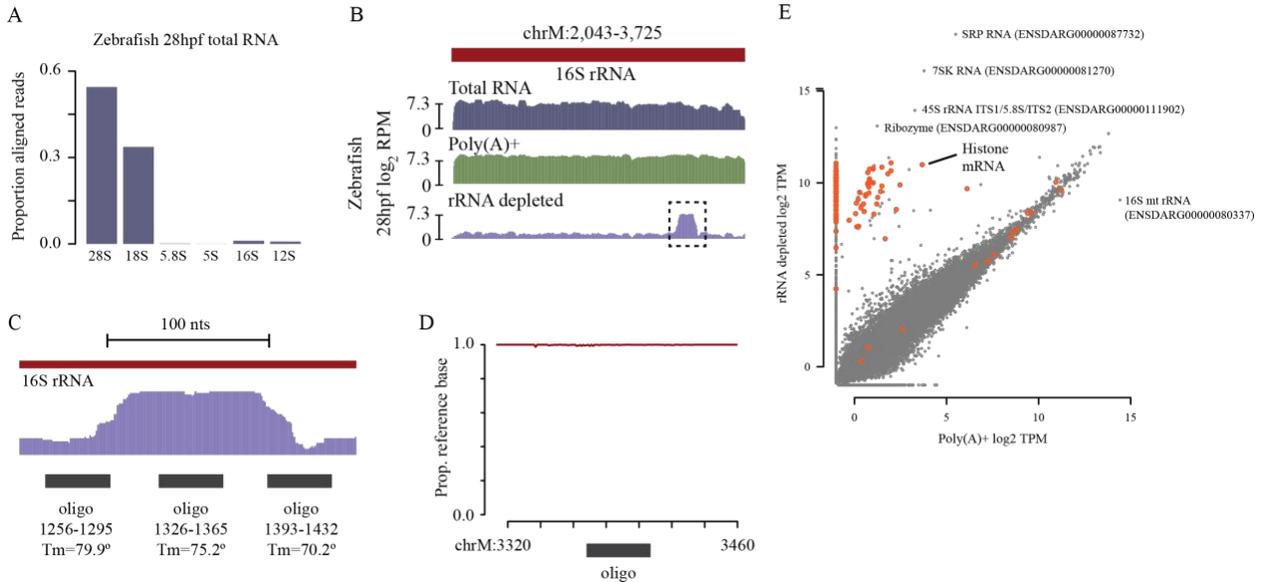


Figure 2-8: RNA-seq methods are highly correlated in zebrafish

A) Barplots showing proportion of aligned reads from untreated total RNA mapping to rRNA species in zebrafish. B) Genome browser tracks comparing read coverage at the 16S rDNA loci in untreated, poly(A)+ and rRNA depleted libraries from zebrafish 28 h.p.f. embryos. One region with less efficient rRNA depletion is boxed. C) Zoomed browser track for the boxed region in (B) showing that the inefficient digestion occurred over the region targeted by one oligo (positions 1326-1365 relative to the 16S sequence). This likely resulted in rRNA fragments that were slightly too large to be efficiently excluded during cleanup and library build. D) Variant analysis of sequencing reads mapping to chrM:3320-2460 showing that nearly all read sequences match the GRCz11 reference sequence, suggesting that there is no defect in the oligo's ability to target; rather, it is likely that this oligo was omitted from the pool in error. E) Biplot comparing log₂ TPM expression levels from poly(A)+ and rRNA depleted libraries at 28 h.p.f. Histone genes are highlighted in orange. Several highly expressed noncoding RNAs are labeled. h.p.f. = hours post fertilization, RPM = reads per million, T_m = melting temperature, TPM = transcripts per million.

respectively (Figure 2-7F). Read coverage over the maternal 28S rRNA gene indeed shows a failure to digest sequence regions where the somatic oligos lack complementarity (Figure 2-7G), while a similar pattern is observed over the somatic 28S gene when only maternal oligos are used (Figure 2-7H). These results show that a full maternal+somatic targeting strategy is required to achieve a maximally effective rRNA depletion and demonstrate that a shared, compact oligo pool can efficiently target these divergent sequences simultaneously. Thus, oligo pools can be augmented to target highly similar rRNA species and avoids the need for separate oligo compliments, thereby expanding the range of depletion targets with a minimum increase in cost and complexity.

In conclusion, we have developed and optimized antisense oligo-based rRNA depletion for *X. laevis* and zebrafish RNA-seq libraries and provide a tool Oligo-ASST to design similar reagents for any other species. For many taxa, designing new oligo sets should be straightforward with Oligo-ASST, given the availability of rRNA sequences in databases such as GenBank. Although there may be use cases where magnetic bead-based methods would be more appropriate, *e.g.* highly degraded RNA (126) or libraries where precise ends are required such as for ribosome profiling (159), for many RNA-seq applications RNaseH digestion should yield excellent results. We anticipate this will be of benefit to researchers who need alternatives to poly(A) selection for RNA-seq, particularly those working with taxa that were inadequately served by previous rRNA depletion methods.

2.2 Cleavage Under Target and Release Using Nuclease (CUT&RUN) Facilitates Genome-wide DNA-binding Protein Mapping During Early Embryogenesis

Detailed methods and materials for this section can be found below in section 3.2.4 and in Appendix E.

Chromatin immunoprecipitation followed by high throughput sequencing (ChIP-seq) is a common technique for accessing protein binding location genome-wide (119). One main limitation to traditional ChIP-seq methods is the amount of input material required to conduct a successful experiment. A typical ChIP-seq experiment requires $10^6 - 10^7$ cells and will yield between 10 – 100 ng of DNA for library building, depending on the species (119, 120). In the case of early embryos, this cell number can translate to 100s, sometimes 1000s, of embryos required per ChIP-seq experiment. For example, in *X. laevis*, conducting a single ChIP-seq experiment on NF stage 8 embryos (the traditional onset of ZGA) would require 250 – 2500 embryos to obtain typical amounts of input material. While protocols that utilize lower amounts of input do exist (160), they require abundant proteins of interest and a high-quality antibody; requirements that are not always achievable for certain targets or species (such *X. laevis*). Therefore, a viable, low-input alternative for ChIP-seq is needed to investigate chromatin binding locations genome-wide during genome activation in *Xenopus*.

One such alternative that has been developed in recent years is Cleavage Under Target and Release Using Nuclease (CUT&RUN). Originally developed for use in cell culture (161), CUT&RUN utilizes protein A and G to localized the conjugated MNase to a particular protein of interest. After cutting the DNA, the protein protected fragments can be isolated for library construction while the remaining chromatin is washed away. This technique dramatically

concentrates the signal-to-noise ratio, thus requiring much less input to be able to robustly identify binding locations throughout the genome. Indeed, CUT&RUN has even been shown to accurately profile protein binding genome-wide in a single cell (162). Therefore, we sought to adapt CUT&RUN for use in *Xenopus* embryos to assay protein binding genome-wide during genome activation.

2.2.1 Embryo Dissociation Increases Nuclei Binding to ConA Beads

Initially, my attempts at CUT&RUN in *X. laevis* embryos yield little DNA suitable for library construction. We reasoned that the prolonged nuclear extraction needed to free all the nuclei in a fully structured embryo was damaging or lysing the nuclei from the outer cells before the inner cell nuclei could be freed; an issue that did not arise in the original cell culture protocol as every cell is able to have the buffer applied equally throughout the culture. To address this, we dissociated the embryo into the component cells prior to nuclear extraction using Newport 2.0 buffer, a buffer designed to prepare *Xenopus* embryos for single cell RNA-seq (163). We compared dissociated and undissociated DNA recovery by extracting nuclei from twelve stage 8 embryos and quantifying the DNA via phenol:chloroform:isoamyl alcohol extraction followed by Nanodrop quantification. To ensure maximum nuclei recovery, we also tested 3 different lysing methods: gentle washing (pipetting the buffer up and down to make a gentle current), pipet mixing (pipetting the entire slurry up and down with a P1000), and vortexing at 1500 rpm. Dissociating the embryos prior to nuclear extraction facilitated the recovery of over 80% of input DNA (Figure 2-9B). Additionally, DNA recovery was four-fold higher with dissociated embryos compared to undissociated embryos (Figure 2-9B). Regardless of dissociation, gently mixing the embryos facilitated a higher nuclei recovery than either pipet mixing or vortexing (Figure 2-9B). Therefore,

embryo dissociation not only increases nuclei binding to conA beads but also is necessary for robust CUT&RUN in *X. laevis*.

2.2.2 CUT&RUN Maps Protein Binding Using Cell Numbers Amenable to Embryology

The input material required for a successful CUT&RUN experiment can range from 500,000 cells to just a single cell (161, 162, 164). To determine if CUT&RUN could be used in *X. laevis* embryos, we conducted CUT&RUN using 12 embryos at NF stage 9, corresponding to ~50,000 cells, for the active promoter associated histone modification, histone 3 lysine 4 trimethylation (H3K4me3), and plotted the coverage over every annotated promoter in the XenLae v9.2 genome. The resulting promoter coverage depicted a typical H3K4me3 pattern with coverage extending into the gene body of the most highly coverage genes, indicating active transcription, and tapering off further down the plot as transcriptional activity is reduced (Figure 2-9A left and middle). Thus, demonstrating CUT&RUN can be used in *X. laevis* with cell numbers amenable for use in early embryos.

2.2.3 H3K4me3 CUT&RUN Coverage Correlates with Published ChIP-seq

To determine how CUT&RUN results compared with that of traditional ChIP-seq, we aligned publicly available ChIP-seq for H3K4me3 at stage 9 published by the Veenstra lab (102) to the XenLae v9.2 genome and quantified the coverage over annotated promoters, similar to my previously generated stage 9 CUT&RUN. Similar to CUT&RUN, ChIP-seq also produces a typical H3K4me3 promoter pattern with gene body extension exhibited in the most active genes which tapers off with less gene activity (Figure 2-9A). One noticeable difference between both

techniques is the additional sequencing depth of the ChIP-seq dataset (approximately 3.5X deeper) may yield greater resolution. Indeed, both the +1 and -1 nucleosomes are clearly visible in the ChIP-seq sample for the most activated genes (Figure 2-9A). Comparatively, the +1 nucleosome is also clearly visible in the CUT&RUN sample, but the -1 nucleosome is less intense compared

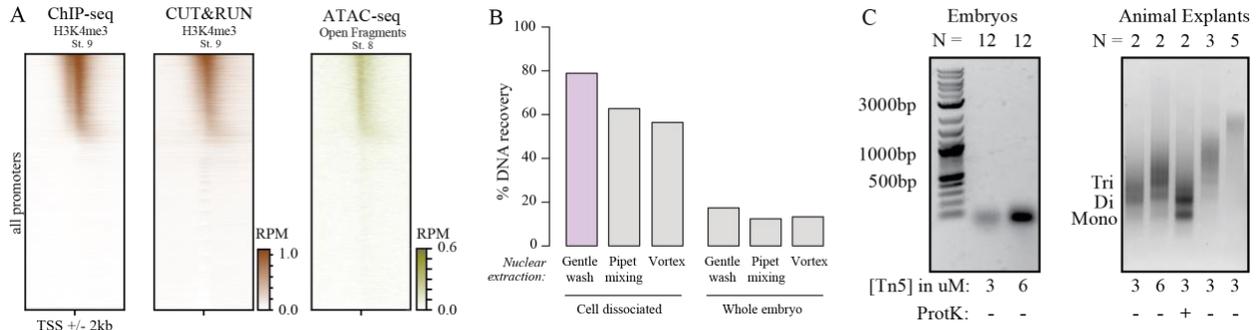


Figure 2-9: CUT&RUN and ATAC-seq optimizations

A) Heatmaps comparing H3K4me3 CUT&RUN with traditional ChIP-seq (left and middle) and with early stage ATAC-seq (right). B) CUT&RUN nuclear extraction optimizations to increase DNA recovery. C) ATAC-seq optimizations to increase transposition yield in accessible fragments. to ChIP-seq (likely due to a greater H3K4me3 concentration at the +1 than the -1 nucleosome). However, we would expect both techniques to yield comparable resolution at equivalent sequencing depths. Thus, CUT&RUN is a viable ChIP-seq alternative that yields similar results as traditional ChIP-seq for assaying protein binding genome-wide.

2.3 Assay for Transposon Accessible Chromatin (ATAC) to Quantify Chromatin

Accessibility During Genome Activation

Detailed method and materials for this section can be found below in section 3.2.5 and Appendix F.

Eukaryotic genomic DNA is tightly packaged into an array of nucleosomes, each composed of a histone octamer core wrapped around 147 bp of DNA (165–167). Transcriptional activation is tightly linked with disrupted nucleosome organization at regulatory regions (i.e. promoters, enhancers, etc.) due to the need for transcription factors to bind the DNA. Therefore, regulatory DNA is highly correlated with accessible genomic locations. Chromatin accessibility can be assayed either chemically (168) or enzymatically (169–171) to isolate accessible or inaccessible regions. Subsequently, isolated DNA can be used as the basis for library construction and sequencing to identify regions of interest genome-wide.

A commonly used chemical assay for chromatin accessibility is FAIRE-seq (168), which uses formaldehyde fixation to separate protein bound fragments from accessible fragments via phenol:chloroform extraction. While this method is widely considered one of the easiest methods to use (172), it has a high background, and this combined with the high number of required input cells limits its use in early embryonic studies (168, 172). A similar high input requirement also limits the usefulness of both DNase-seq and MNase-seq, with both techniques requiring a minimum of 1 million input cells (169, 170, 172). However, Assay for Transposon Accessible Chromatin (ATAC-seq) can be used with as few as 500 cells for input (171–173) and can directly probe chromatin accessibility genome-wide, as opposed to methods like MNase-seq in which low signal, not high signal, indicates open chromatin. Therefore, we sought to develop an ATAC-seq method to use in early *X. laevis* embryos.

2.3.1 Animal Cap Dissection Bypasses Yolk Inhibition of Tn5

ATAC-seq relies on the Tn5 transposase to cut and insert sequencing adaptors at accessible chromatin. Genomic locations that are inaccessible to the transposase are inferred to be bound by

another protein, and therefore represent closed chromatin. However, it is thought that Tn5 is highly inhibited by yolk protein. Indeed, previous accessibility assays in *Xenopus* resorted to alternative methods due to the observed lack of transposition (98). ATAC-seq protocols in other egg-laying vertebrates, such as zebrafish (174), remove the yolk prior to transposition to avoid poor yield, however this is not possible in *Xenopus* since the yolk is internalized in the cells. To bypass yolk inhibition, Peter Klein's lab first dissected animal explants for transposition due to relatively light yolk content compared to the vegetal cells (175) and was able to successfully conduct ATAC-seq in gastrulation stage embryos. We sought to use the same innovation to conduct ATAC-seq in earlier, blastulae stage embryos.

We conducted ATAC-seq on stage 10 animal caps to compare with ATAC-seq conducted on whole embryos at the same stage and at amenable numbers that are comparable to CUT&RUN. After library construction, we ran each library on a 1% agarose gel to assay transposition. In whole embryos, no observable transposition has occurred as evidenced by only being able to recover sequencing adaptors. (Figure 2-9C). However, equivalent treatment on only 2 animal explants yielded visible banding corresponding to mono-, di-, and tri-nucleosomal fragments, with the higher Tn5 concentration yielding more material (Figure 2-9C). Additionally, we treated one sample with proteinase K to immediately stop transposition and facilitate fragment release (175). This optimization from the Klein lab further increased yield, primarily in fragments less than tri-nucleosomal in size (*i.e.* accessible and various nucleosomal fragments) (Figure 2-9C). Thus, animal explant dissection bypasses yolk inhibition and facilitates ATAC-seq in *X. laevis* embryos.

2.3.2 Optimized Number of Animal Caps and Tn5 Concentration Facilitates Transposition

To further increase transposition yield, we increased both the effective concentration of Tn5 transposome and increased the number of caps transposed. Compared to the previous transposition, doubling the transposome concentration increased the amount of transposed DNA as evidence by the increased amount of library. However, most of this gain are in fragments larger than tri-nucleosomal fragments and therefore are of no use for assaying accessible chromatin (Figure 2-9C). It is likely the excess transposome targeted less accessible sites among larger nucleosome arrays, leading to larger fragment generation. Additionally, the increased number of animal caps resulted in poor transposition, with little to no transposition having occurred using 5 animal caps (Figure 2-9C). This seems to indicate that 2 is the optimized number of animal caps as this provides enough cells to conduct ATAC-seq without adding an inhibitory amount of yolk to the reaction. Thus, the optimized ATAC-seq protocol for *X. laevis* embryos utilizes 2 animals caps and 3 μ M transposome.

2.3.3 ATAC-seq at Early Developmental Stages Shows Open Chromatin at TSS's

Previous ATAC-seq studies in *Xenopus* focus on later (gastrulae) stage embryos as early embryos did not have good transposition (175), likely due to a poor DNA-to-yolk ratio. We reason that this poor transposition was likely due to a high signal-to-noise ratio that was rendering early embryo ATAC-seq ineffective. Therefore, we sought to boost the signal-to-noise ratio in early embryos by enriching for accessible fragments via size-selection. We transposed stage 8 embryos, earlier than what previous studies had successfully transposed, using the aforementioned optimizations and constructed libraries for high-throughput sequencing. Once constructed, each

library was run on a 2% TBE gel and size-selected to separate 150-250 (accessible) and 250-600 (nucleosomal) bp fragments. Following sequencing, we plotted the accessibility coverage over TSS's.

The resulting coverage showed enriched accessibility over the TSS's of active genes (Figure 2-9A, right), where we would expect to observe chromatin accessibility due to active regulation. Interestingly, comparison with the H3K4me3 ChIP-seq signal indicates that this enrichment is in between the +1 and -1 nucleosome, exactly where transcription machinery would be binding (Figure 2-9A, left and right). This accessibility extends into gene body and overlaps with H3K4me3 binding, likely a result of nucleosomal eviction and replacement as the transcriptional elongation over these genes. Taken together, these results demonstrate that the optimized ATAC-seq protocol we have developed facilitates the interrogation of chromatin accessibility genome-wide at previously unprobed developmental stages.

2.4 Discussion

We demonstrate that a streamlined RNaseH digestion protocol using easily obtained reagents efficiently and cost-effectively achieves ribosomal RNA depletion, which we estimate to be ~US\$8 per reaction. Our Web tool Oligo-ASST improves oligo design to use shorter antisense DNA oligos (39-40 nts) that tile rRNA target sequences with gaps, thereby reducing reagent cost compared to previous methods (*125*, *135*) while still producing high-quality RNA-seq libraries comparable to those constructed with poly(A) selection. In addition, since Oligo-ASST can also design compact oligo sets for treating multiple different rRNAs by targeting shared sequences, this protocol is especially advantageous for researchers to achieve rRNA depletion in diverse taxa.

Our digestion reaction proceeds for only 5 minutes at 65°C using thermostable RNaseH, enabled by our use of oligos with T_m near or above 65°C. Reactions at lower temperatures produce comparable results, which would be beneficial when using oligos with lower T_m , e.g. to target AT-rich RNAs. Thus, we have demonstrated that the previously standard full tiling strategy is unnecessary for a typical RNA-seq use case. Increasing allowable gap lengths leaves larger digested fragments and reduces the efficiency of the depletion, though there may be applications for which this may be acceptable.

Additionally, we have also adapted the low-input ChIP-seq alternative CUT&RUN and ATAC-seq for use in *X. laevis* embryos. These innovations bypass a major limitation to characterizing the chromatin landscape during pluripotency induction, namely the amount of input material. Using only 12 blastula stage embryos, we have generated H3K4me3 genome-wide profiles that highly resemble those generated with traditional ChIP-seq using less than 10% of the input material. Dissociating embryos prior to nuclear extraction makes CUT&RUN feasible in *X. laevis* embryos by ensuring a minimal number of nuclei are lysed from prolonged exposure to the extraction buffer. These profiles were also generated using 3.5X less sequencing depth, demonstrating a higher signal-to-noise ratio compared to traditional ChIP-seq. For ATAC-seq, size selection the libraries prior to sequencing boosts the signal-to-noise ratio to sufficient levels as to facilitate assessing chromatin accessibility at developmental stages previously described as ineffective (175). However, by enriching for accessible fragments, we have demonstrated that ATAC-seq on stage 8 embryos characterizes chromatin accessibility at sufficient resolution to identify accessible loci flanked by nucleosomes at actively regulated genes. These adapted protocols will facilitate future pluripotency and genome activation studies by providing high-throughput tools to interrogate transcriptional regulation genome-wide. We anticipate these

methods will be of benefit to researchers who need low-input alternatives to traditional ChIP-seq and chromatin accessibility techniques, particularly those working with taxa that were inadequately served by previous methods.

3.0 Hybridization Led to a Rewired Pluripotency Network in the Allotetraploid *Xenopus laevis*

Portions of this chapter have been published in Phelps et. al. 2022 on *BioRxiv* (176).

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, github.com/MTLeeLab/xl-zga. Additional data files including chromosome alignments are available at OSF, osf.io/ct6g8/

3.1 Introduction

In mammals, zygotic genome activation (ZGA) is triggered after an initial period of transcriptional quiescence, during the slow first cleavages post fertilization (31). This is 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 (29, 33). In contrast, faster-dividing taxa including zebrafish, *Xenopus*, and *Drosophila* activate their genomes in the blastula hours after fertilization during the maternal-to-zygotic transition (MZT) (32, 34, 35, 90), which leads immediately to pluripotency. In zebrafish, maternally provided homologs of NANOG, OCT4 and SOX2 are required for genome activation (37–39); 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, e.g. Zelda (43), CLAMP (44, 45) and GAF (46) in *Drosophila*, though they seem to

share many functional aspects with vertebrate pluripotency factors, including pioneering roles in opening repressed embryonic chromatin and establishing activating histone modifications (46–48). This diversity of strategies implies that the gene network regulating pluripotency has been extensively modified over evolutionary time (49, 50), 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 (99) (Figure 3-1A). 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. 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 (99, 102).

Allopolyploidy often provokes acute effects on gene expression (103, 104), leading to regulatory shifts over time to reconcile dosage imbalances and incompatibilities between gene copies (105–107). This phenomenon has been explored primarily in plants (108–110), but the extent to which this has occurred in the few characterized allopolyploid vertebrates is unclear (111–113). For *X. laevis*, there is a broad trend toward balanced homeolog expression across development and adult tissues (99) and an overall ontogenetic and transcriptomic trajectory similar to 48-million-years diverged diploid *X. tropicalis* (177, 178). However, initial observations suggest a divergent cis-regulatory landscape between the two *X. laevis* subgenomes (102, 179).

Although *Xenopus* embryos have long been a model for understanding the MZT, e.g. (2, 36, 40, 90, 92, 93, 177, 180–185), ZGA regulators have not previously been identified in *X. laevis*. Here, we elucidate the top-level regulators of *X. laevis* pluripotency and ZGA, and the 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.

3.2 Materials and Methods

3.2.1 *Xenopus* 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 (Aquanearing) at 18°C with a 12/12 h light/dark cycle. Frogs were fed 3x weekly with Frog Brittle (NASCO #SA05960 (LM)M).

3.2.2 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) (142). Dissected testes were stored up to one week in L-15 medium at 4°C prior to use. Zygotes were de-jellied (142) 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.

3.2.3 RNA-seq Libraries

All stage 9 embryos were collected halfway through the stage, at 8 hours post fertilization. Triptolide samples were bathed in 20 μM triptolide in DMSO (200X 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 *pou5f3.3* (GTACAATATGGGCTGGTCCATCTCC) (186) and *sox3* (AACATGCTATACATTTGGAGCTTCA) (187) 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: 40 ng *pou5f3.3* + 40 ng GFP, 40 ng *sox3* + 40 ng GFP, 40 ng *pou5f3.3* + 40 ng *sox3*, or 80 ng GFP. High concentration morpholino injections were 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.

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 minutes, 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 (114): 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, 10U of thermostable RNaseH (NEB #M0523S) and 2 µl of provided 10X 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 minutes, then 5U of TURBO DNase (Invitrogen #AM2238) and 5 µl of provided 10x buffer was added, volume brought to 50 µl with nuclease-free water and incubated at 37°C for 30 minutes. 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 minutes. Following first and second strand synthesis, DNA was purified with 1.8X AmpureXP beads (Beckman #A63880), end repaired, then ligated to sequencing adaptors diluted 1:5. Ligated DNA was purified with 0.9X AmpureXP beads and PCR amplified for 8 cycles, then purified again with 0.9X 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, d, e).

3.2.4 CUT&RUN

CUT&RUN procedure was adapted from Hainer et al (*164*) optimizations of the method of Skene and Henikoff (*161*). 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 (*163*) 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 (20mM HEPES-KOH, pH 7.9, 10mM 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 mins. 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 CaCl₂ and allowed to digest for 30 mins, then stopped using 2x 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 µL of water and verified by Qubit dsDNA high sensitivity and Fragment Analyzer. Antibodies used were: H3K4me3, Invitrogen #711958, RRID:AB_2848246, Lot #2253580; H3K27ac, ActiveMotif #39135, RRID:AB_2614979, Lot #06419002; V5, Invitrogen #R960-25, RRID:AB_2556564, Lot #2148086. Biological duplicate libraries from different embryo collection days were constructed for the key samples (St. 8 & 9 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:GGACAGCACGGGAGGCGGGGGATCCGACCAGCCCATATTGTACAGCCAAAC; R:TATCATGTCTGGATCTACGTCTAGATCAGCCGGTCAGGACCCC) and *sox3.S* - NM_001090679.1 (F:TATAGCATGTTGGACACCGACATCA; R:TTATATGTGAGTGAGCGGTACCGTG) 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 2hrs at 37°C, then treated with 5U 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 5 nL of 40 ng/µL of mRNA on opposite sides, for a total of 10 nL per embryo. 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 0.9x AmpureXP beads and PCR

amplified for 15 cycles, then purified again with 0.9x AmpureXP beads. Libraries were size selected to 175 – 650 bp via 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.

3.2.5 ATAC-seq

ATAC procedure was from Esmaeili et al (175) Embryos were grown in MR/3 until desired NF stage and devitellinized individually with fine watch-maker forceps. Ectodermal explants (animal caps) were dissected using watch-maker forceps in 0.7x 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 (10mM 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: CAAGCAGAAGACGGCATAACGAGAT[i7]GTCTCGTGGGCTCGG with i7 indices 707 – gtagagag; 714 – tcatgagc; 716 – tagcgagt; and AATGATACGGCGACCACCGAGATCTACAC[i5]TCGTCGGCAGCGTC with i5 indices 505 – gtaaggag; 510 – cgtctaata; 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. 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. Biological duplicate libraries from different embryo collection days were performed for each stage.

Transposomes were constructed according to Picelli et. al. (188) Adapter duplexes for Tn5ME-A (TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG) + Tn5MErev ([phos]CTGTCTCTTATACACATCT) and Tn5ME-B (GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG) + Tn5MErev were each annealed in 2 µl of 10X 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.

3.2.6 Transcriptomic Analysis

RNA-seq reads were mapped to the *X. laevis* v9.2 genome using HISAT2 v2.0.5 (151) (--no-mixed --no-discordant). Mapped reads were assigned to gene exons (Xenbase v9.2 models) using featureCounts v2.0.1 (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 (189) 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 $P < 0.05$ and \log_2 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. 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 (190) hairpin sequences MI0001449 and MI0038331 were aligned to the v9.2 and v10.1 reference genomes using UCSC BLAT (191) 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 (156).

dN/dS ratios were calculating using PAML v4.9f (192) with L-S pairwise CDS alignments produced by pal2nal v14 (193) on amino-acid alignments by EMBOSS needle v6.6.0.0 (-gapopen 10 -gapextend 0.5) (194).

All other statistical tests were performed using R v4.0.4 (195).

3.2.7 Chromatin Profiling Analysis

CUT&RUN and ATAC-seq paired-end reads were mapped to the *X. laevis* v9.2 genome using bowtie2 v2.4.2 (196) (--no-mixed --no-discordant) and only high-quality alignments (MAPQ \geq 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 (197) in paired-end mode (--illumina --trim-n). Downstream analyses were performed using custom scripts with the aid of BEDtools v2.30.0 (153), Samtools v1.12 (198), and deepTools v3.5.1 (199).

For promoter-centered analyses, one transcript isoform per gene was selected from Xenbase 9.2 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.

To identify open chromatin regions, aligned ATAC-seq fragments pooled between replicates were filtered to <130 bp, then peaks called using MACS2 v2.2.7.1 (200) with an effective genome size of 2.4e9 (number of non-N bases in the 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 RepeatMasker and BLASTed 45S, 16S, 12S, and 5S sequences. Peaks on unassembled scaffolds were also excluded.

Putative enhancers had 2-fold enriched pooled stage 8 H3K27ac CUT&RUN coverage over no antibody, ≥ 1 RPM H3K27ac coverage, and < 0.5 RPM no antibody coverage, in a 500-bp window centered on ATAC-seq peak summits. Enhancers were classified as distal if they were > 1 kb from any Xenbase 9.2 annotated TSS, proximal otherwise.

For transcription factor peak calling, the Sox3 sample was down-sampled to match Pou5f3 read depth (~12 M read pairs) using samtools view -s. No-antibody samples were pooled as a uniform background. MACS2 was run as above, and SEACR v1.3 (201) was run in norm relaxed mode. Peak calls were not used for enhancer analyses; rather, enhancers or homeologous regions with ≥ 1 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, enhancer pairs were annotated as differential or conserved based on one or both partners, respectively, mapping to a putative enhancer, as described above. The total region of each putative enhancer corresponding to ≥ 2 -fold H3K27ac enrichment was calculated and converted to a bigWigs representing the genomic location of each enriched region. 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.

3.2.8 Motif Finding

Enriched sequence motifs in enhancers were identified using Homer v4.11.1 (202) in scanning mode against the vertebrate database, using 200 bp of sequence centered on the ATAC-seq peak for enhancers and 500 bp of sequences 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 used in de novo mode on the top 500 peaks. The top motif was extracted for each of Pou5f3 and Sox3, then scanned against the entire set of peaks.

3.2.9 Homeologous Enhancer Identification

Each chromosome pair (*e.g.*, chr1L and chr1S) was aligned using lastZ-1.04.00 (203) and UCSC Genome Browser utilities (154) 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, gappedthreshold=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.

3.2.10 Comparison with *X. tropicalis* and Zebrafish

X. tropicalis wild-type RNA-seq reads from Owens et al (204), 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 (40), 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.

Zebrafish annotations for activated and Pou5f3 / Nanog / SoxB1 affected genes were obtained from Lee & Bonneau et al (37) 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 were lifted over as above 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 (205) 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 (206) (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.

3.3 Results

3.3.1 Homeologs are Activated Independently of Each Other

At genome activation, the *X. laevis* pluripotency network consists of maternal regulators acting directly on the first embryonic genes (Figure 3-1B). To identify these genes, we performed a total RNA-seq early embryonic time course using our *X. laevis*-specific ribosomal RNA depletion protocol (114) (Figure 3-1A/B). We identified 4772 genes with significant activation by the middle of Nieuwkoop and Faber (N.F.) stage 9 (8 hours post fertilization [h.p.f.] at 23°C) (Figure 3-1C), through a combination of exon- and intron-overlapping sequencing reads deriving from nascent pre-mRNA (37). Indeed, two-thirds of these genes had substantial maternal contributions that masked their activation when quantifying exon-overlapping reads alone (Figure 3-1C). These genes fail to be activated in embryos treated at 1-cell stage with the transcription inhibitor triptolide (181) when compared to DMSO vehicle control embryos (Figure 3-1B,C, Figure 3-2A-C).

To distinguish direct targets of maternal factors (primary activation) (Figure 3-1B), 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 (37, 53). 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 3-1C, Figure 3-2A).

We analyzed subgenome of origin for activated genes and found that they are preferentially encoded as two homeologous copies in the genome ($P = 2.9 \times 10^{-181}$, χ -squared test, 6 d.o.f.) (Figure 3-3A). 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 3-3B-C, Figure 3-2D). This degree of divergent activation suggests large differences in the cis-regulatory architecture between gene homeologs in the two subgenomes. Genes activated from both subgenomes are enriched in transcriptional regulators ($P < 0.01$, Fisher's exact test, two-sided) (Figure 3-2E), suggesting that gene function may have influenced homeolog expression patterns.

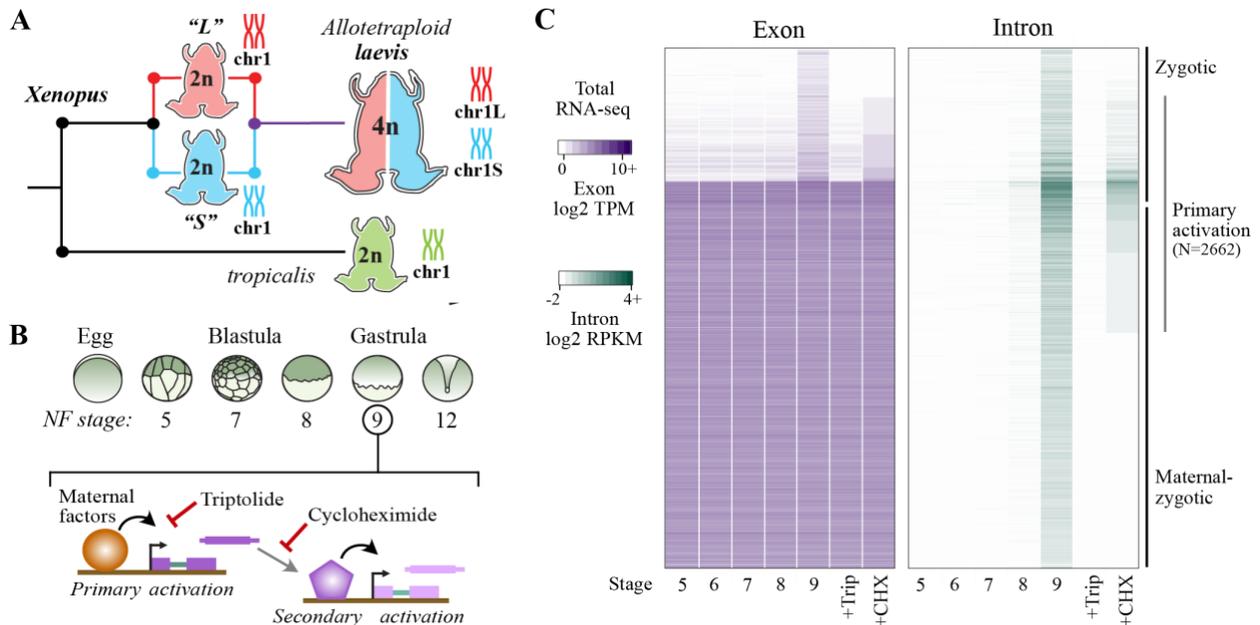


Figure 3-1: Identifying the first wave of genome activation across the two subgenomes.

(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.

However, there is no evidence for strong 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 3-2F-G).

3.3.2 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 (88). Similar to *X. tropicalis mir-427* (204) and the related zebrafish *mir-430* (37), *mir-427* is one of the most strongly activated genes in the *X. laevis* embryonic genome (Figure 3-3D, Figure 3-2C, Figure 3-4A). 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 3-4B). This is in stark contrast to the 171 tandemly arrayed precursors in the two *X. tropicalis mir-427* loci on Chr03, which is thought to accelerate mature *miR-427* accumulation during the MZT to facilitate rapid maternal clearance (204). Zebrafish similarly encodes a large array of 55 *mir-430* precursors, which begin to target maternal mRNA for clearance shortly after ZGA (37, 52, 87).

To better capture the genomic configuration of the *mir-427* primary transcript, we aligned the miRBase-annotated precursor sequence (190) to the recently released 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 3-3D, Figure 3-4A). This is reminiscent of the *X. tropicalis* configuration (204), though smaller in scale and on a non-homologous chromosome. The corresponding region on Chr1S is unalignable (Figure 3-4C), 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 3-4D). 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 between the *tropicalis* and *laevis* lineages.

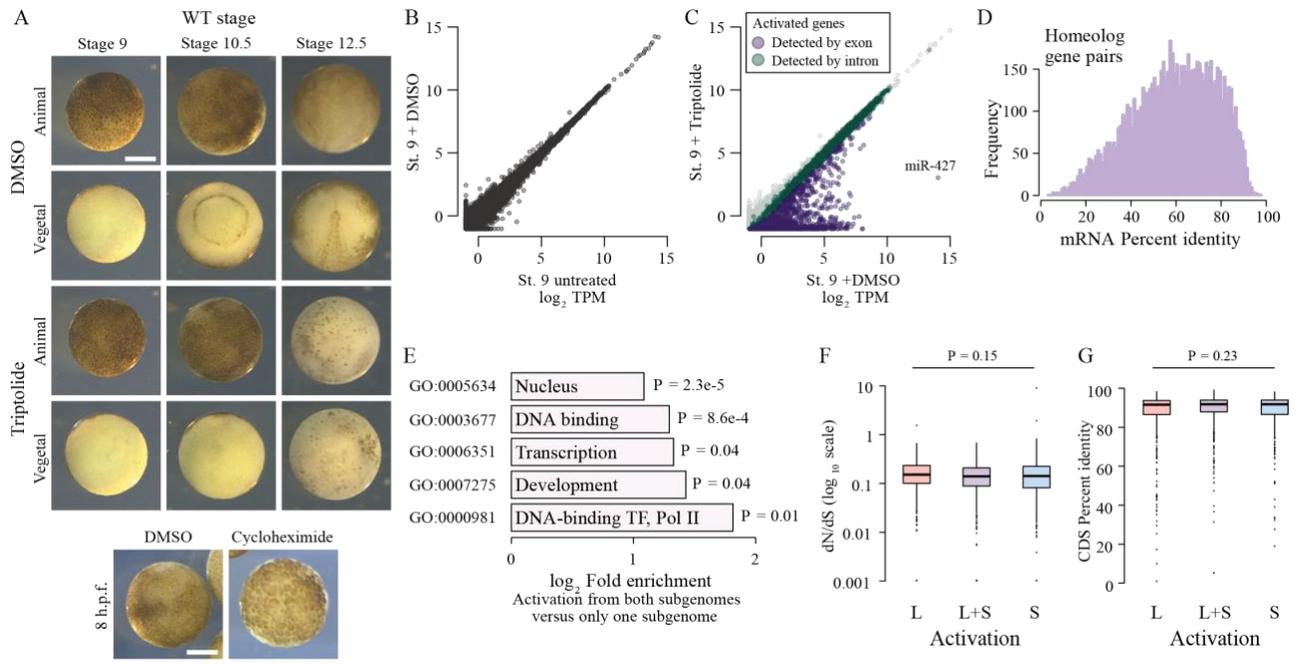


Figure 3-2: Measuring genome activation

(A) (Top) Animal and vegetal views of embryos treated with DMSO (vehicle) versus triptolide. Triptolide-treated embryos fail to gastrulate. (Bottom) Comparison of DMSO versus cycloheximide treated embryo. Treatment was at stage 8, which inhibits progression to stage 9. Scale bar = 0.5 mm. (B) Biplot of RNA-seq for untreated versus DMSO-treated embryos at stage 9, showing no effect on the transcriptome. (C) Biplot of RNA-seq for DMSO versus triptolide treated embryos, showing inhibited activation as detected by exonic (purple) and intronic (green) signal. The predicted mir-427 primary transcript is labeled and exhibits >95% expression inhibition. (D) Histogram of mRNA percent identity between homeolog pairs, as measured by Needleman-Wunsch alignment. Maximum is 0.975 (5 in every 200 bases, which should be generally distinguishable by RNA-seq using 2x100 sequencing reads) (E) Significantly (FDR < 0.05, Fisher's exact test, two-sided) enriched Gene Ontology terms in genes activated from both homeologs, as compared to genes activated from only one subgenome. (F) Boxplots of non-synonymous to synonymous substitution rate ratio (dN/dS) shown on a log₁₀ scale, for genes activated from both subgenomes or only one subgenome ($P = 0.15$, Kruskal-Wallis test; median L = 0.15, LS = 0.14, S = 0.14). All gene groups trend toward stabilizing selection. (G) Boxplots of CDS percent similarity for activation groups ($P = 0.23$, Kruskal-Wallis test). For all boxplots: center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers.

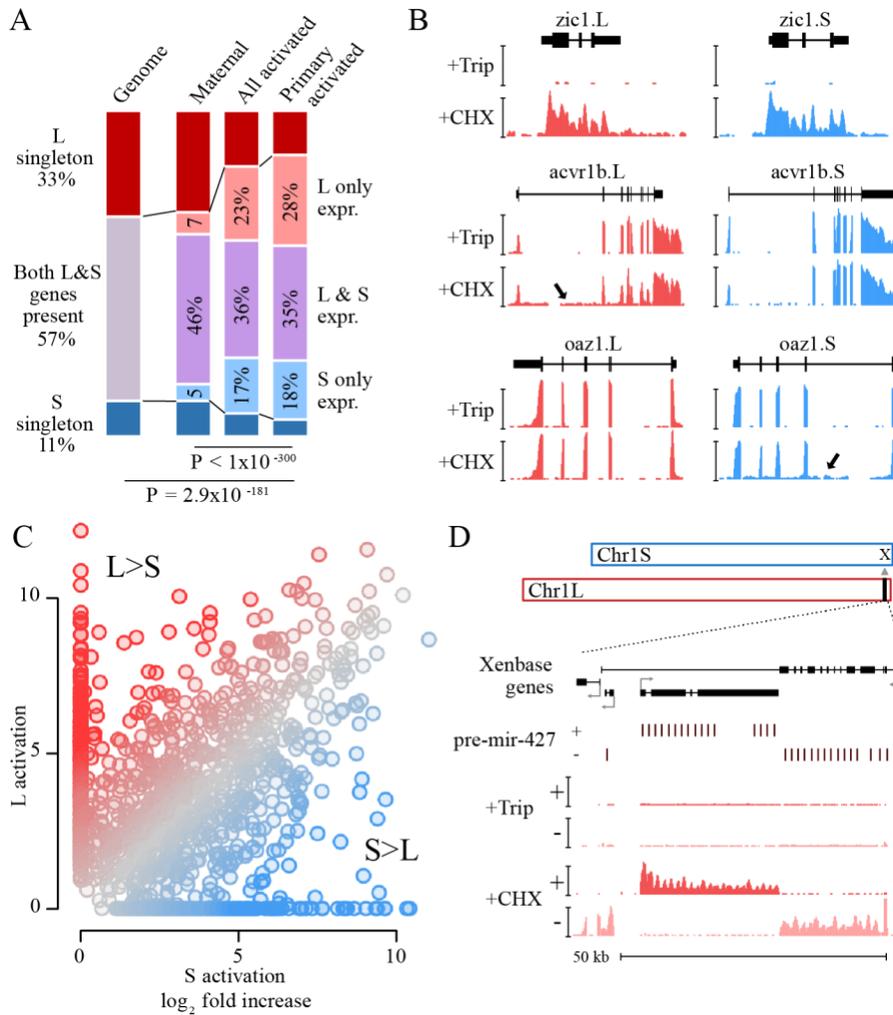


Figure 3-3: Homeologous genes are differentially activated in the early embryo.

(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. $P = 2.9 \times 10^{-181}$, c-squared test, 6 d.o.f., comparing genomic to expressed proportions; $P < 1 \times 10^{-300}$, c-squared test, 8 d.o.f., comparing proportions within expressed genes. (B) Browser tracks showing \log_2 reads-per-million RNA-seq coverage of equivalently activated homeologs (top) and differentially activated homeologs (L-specific, middle; S-specific, bottom). (C) Biplot comparing \log_2 fold activation of homeologs in cycloheximide versus triptolide treated embryos. (D) 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). Trip = triptolide, CHX = cycloheximide.

3.3.3 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) (161, 164), 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 3-5A, Figure 3-6A-C). At stages 8 and 9, the active marks H3 lysine 4 trimethylation (H3K4me3) and H3 lysine 27 acetylation (H3K27ac) were enriched in activated genes compared to their unactivated homeologs (stage 8 H3K27ac, $P < 3 \times 10^{-8}$; stage 8 H3K4me3, $P < 0.01$; stage 9 H3K4me3, $P < 2 \times 10^{-10}$, paired t-tests, two-sided) (Figure 3-5B-C, Figure 3-6D). Differential promoter engagement by transcriptional machinery likely underlies the differential active histone levels; however, we found no promoter sequence differences between homeologs that would implicate differential recruitment of specific transcription factors.

Instead, we searched for differences in gene-distal regulatory elements – i.e., 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 from stage 8 and 9 embryos; the high concentration of yolk in vegetal cells inhibits the Tn5 transposase (175). We called peaks of elevated sub-nucleosome sized fragment coverage, then intersected the open regions with our H3K27ac CUT&RUN. This yielded 7562 putative open and acetylated enhancers at genome activation (Figure 3-6E).

To identify homeologous L and S enhancer regions, we constructed a subgenome chromosome-chromosome alignment using LASTZ (203). This yielded a syntenic structure consistent with genetic maps (Figure 3-5D) (99), recapitulating the large inversions between chr3L/chr3S and chr8L/chr8S. 79% of enhancer regions successfully lifted over to homeologous

chromosomes, and of these, >90% of these are flanked by the same homeologous genes (Figure 3-6F), confirming local synteny.

Among the paired regions, only 23% had conserved enhancer activity in both homeologs, with the remaining pairs exhibiting differential H3K27ac and chromatin accessibility (Figure 3-5E, Figure 3-6G). Differential enhancer density around genes significantly correlated with differential activation ($P = 1.3 \times 10^{-16}$, Pearson's correlation test) (Figure 3-5C, middle), 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 ($P = 0.20$, Pearson's correlation test) (Figure 3-5C, right). Thus, differences in enhancer activity likely underlie divergent gene homeolog transcription at genome activation.

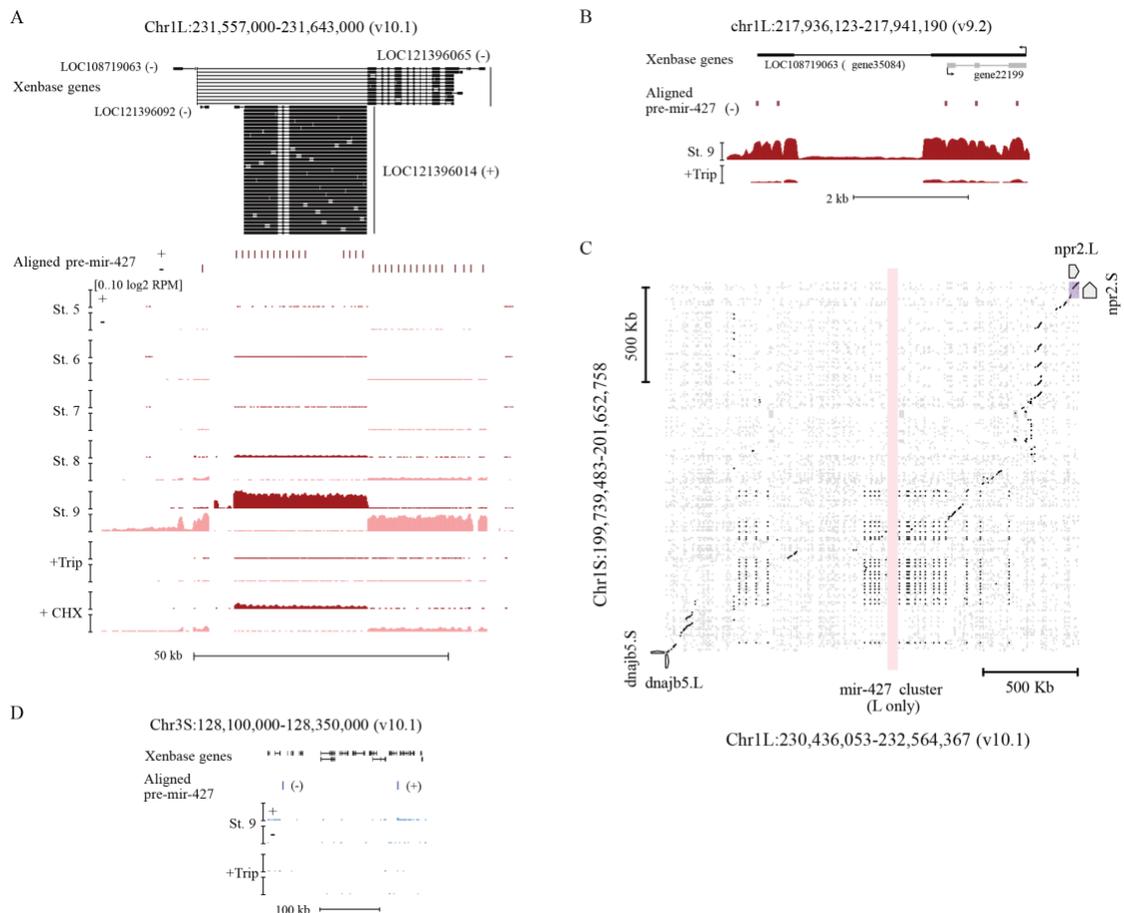


Figure 3-4: The mir-427 locus

(A) Browser tracks showing strand-separated log₂ reads-per-million RNA-seq coverage over the predicted mir-427 primary transcript near the telomere of Chr1L on the v10.1 genome assembly. Xenbase gene isoforms are annotated at the top, aligned precursor mir-427 sequences are annotated in the middle according to strand orientation. (B) Browser track showing log₂ reads-per-million RNA-seq coverage over the presumed mir-427 encoding region on the v9.2 genome assembly. The overlapping antisense transcript is not transcribed (all coverage shown is sense to the mir-427 transcript). (C) Dot matrix alignment plot showing BLAST local alignments between v10.1 Chr1L and Chr1S in the region flanking the mir-427 locus. Repetitive sequence alignments (Xenbase soft-masked genomic sequence) are shown in gray, non-repetitive alignments in black. Upstream (dnajb5) and downstream (npr2) homeologous genes are labeled. The L-specific mir-427 locus is highlighted in light red, showing no alignments to Chr1S. (D) Region of v10.1 Chr3S where two additional sequence matches to the mir-427 hairpin are found by BLAT. However, there is minimal RNA-seq coverage, suggesting the Chr1L locus is the only bona fide mir-427 encoding region in the v10.1 assembly. Log₂ reads-per-million coverage is shown on the same scale as panel (A).

3.3.4 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 3-5F). Since mammalian OCT4 and SOX2 are master regulators of pluripotent stem cell induction (29), and zebrafish homologs of these factors are maternally provided and required for embryonic genome activation (37–39), 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 revealed high maternal levels of *pou5f3.3* (OCT4 homolog) and *sox3* mRNA, each deriving from both subgenomes (Figure 3-6H). To assess their roles in genome activation, we inhibited their translation using previously validated antisense morpholinos (*I86*, *I87*) injected into stage 1 embryos. Each of the two morpholinos was complementary to both the L and S homeologs of *pou5f3.3* and *sox3*, respectively, but not to their paralogs that are primarily expressed zygotically (i.e., *pou5f3.1* and weakly maternal *pou5f3.2*). To again focus specifically on maternal regulation of primary genome activation, we treated the injected embryos with cycloheximide at stage 8 and collected them at stage 9 for RNA-seq. When both Pou5f3.3 and Sox3 were inhibited, we observed significant downregulation of 62% of activated genes compared to embryos injected with a control morpholino, including the *mir-427* transcript (Figure 3-7A, Figure 3-8A). Targeting *pou5f3.3* or *sox3* mRNA individually had minimal impact on genome activation (Figure 3-8A), suggesting these two maternal transcription factors together coordinate early gene expression.

To interrogate Pou5f3.3 and Sox3 chromatin binding across the subgenomes, we performed CUT&RUN on stage 8 embryos injected at stage 1 with mRNA encoding V5 epitope-tagged *pou5f3.3.L* and *sox3.S*. Peak calling revealed thousands of binding sites for each factor (Figure 3-8B-F), and Homer *de novo* motif analysis recovered the OCT4 and SOX3 binding sequences as top hits ($P = 10^{-184}$ and $P = 10^{-98}$, respectively) (Figure 3-7B, Figure 3-8G-H). CUT&RUN signal for both factors is enriched in the vicinity of activated genes, with stronger association to genes affected by morpholino treatment ($P < 1 \times 10^{-300}$, Kruskal-Wallis test) (Figure 3-7B-C), and indeed comparison between differentially affected homeolog pairs showed preferential binding in enhancers near the Pou5f3.3/Sox3-dependent homeolog (Pou5f3.3: $P = 1.5 \times 10^{-5}$; Sox3: $P = 1.8 \times 10^{-5}$, Kruskal-Wallis tests) (Figure 3-8J-K). Together, these results implicate Pou5f3.3 and Sox3 in regulating ZGA differentially between the two subgenomes.

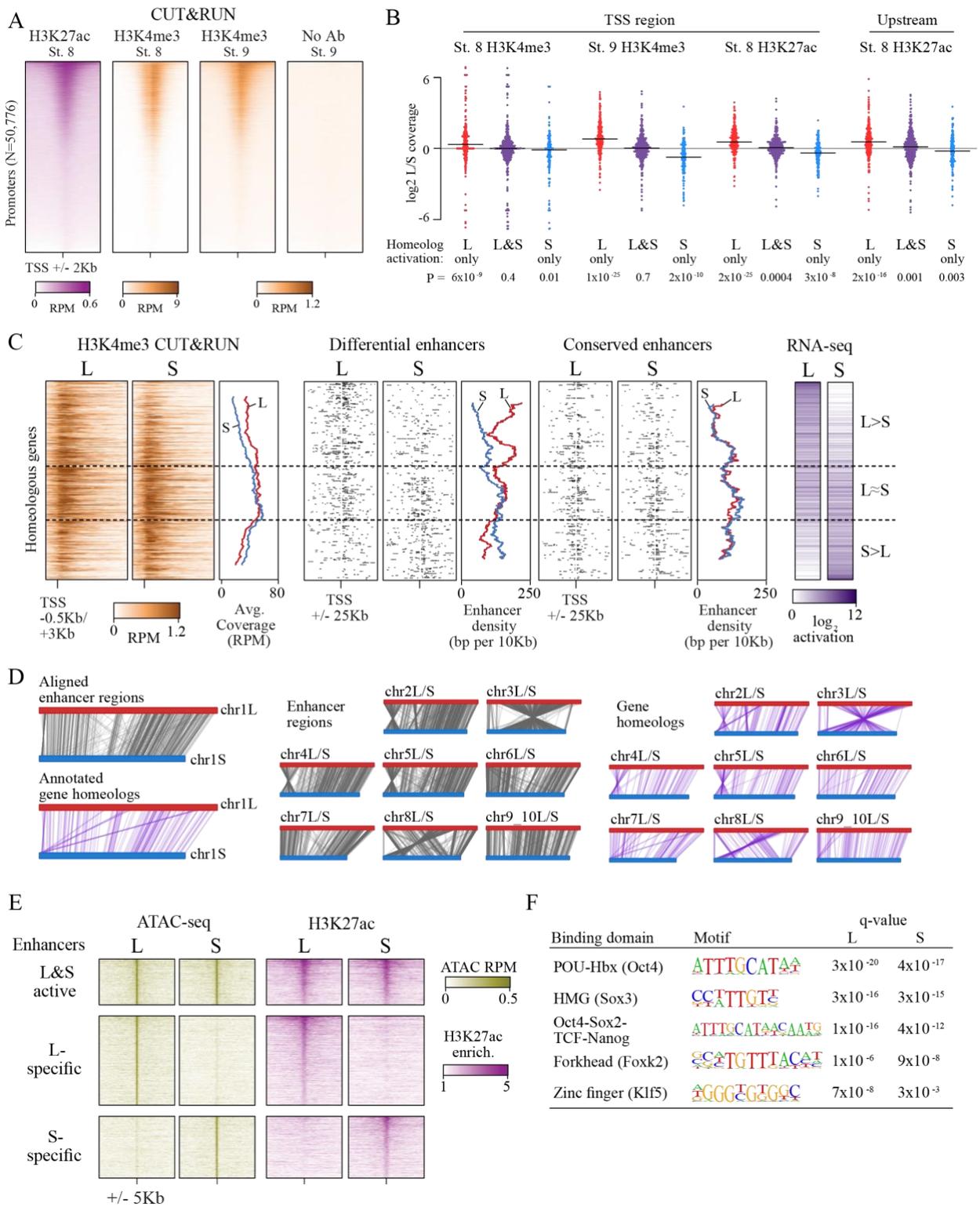


Figure 3-5: Differential homeolog activation is regulated by subgenome-specific enhancers

(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. *P* values are from two-sided paired t-tests of log₂ L homeolog coverage vs log₂ S homeolog coverage. (C) Stage 9 H3K4me3 CUT&RUN coverage over paired homeologous gene regions around the TSS (left) and maps comparing enhancer density near homeologous TSSs (middle). Differential enhancers are active in only one subgenome, conserved 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 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 enhancers (top) and subgenome-specific enhancers. (F) Top enriched transcription factor motif families in L-specific and S-specific active enhancers compared to inactive homeologous regions. FDR-corrected *P*-values from Homer are shown. RPM = reads per million.

3.3.5 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 (99). For three-way homeologs/orthologs with minimal maternal contribution in *X. laevis*, there is broad 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 (Pearson's $r = 0.72$) (Figure 3-9A left). 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, Pearson's $r = 0.60$; S, Pearson's $r = 0.55$) (Figure 3-9A right). 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 remodeling post allotetraploidization that maintained relative gene expression dosage for embryonic genome activation.

However, most differentially activated genes also have a maternal contribution, which could offset asymmetries in homeolog activation levels. Indeed, overall when both *X. laevis* gene homeologs are activated, the *X. tropicalis* ortholog is more likely also to be activated, compared to genes where only one homeolog is activated ($P = 5.0 \times 10^{-20}$, χ -squared test, 4 d.o.f.) (Figure 3-

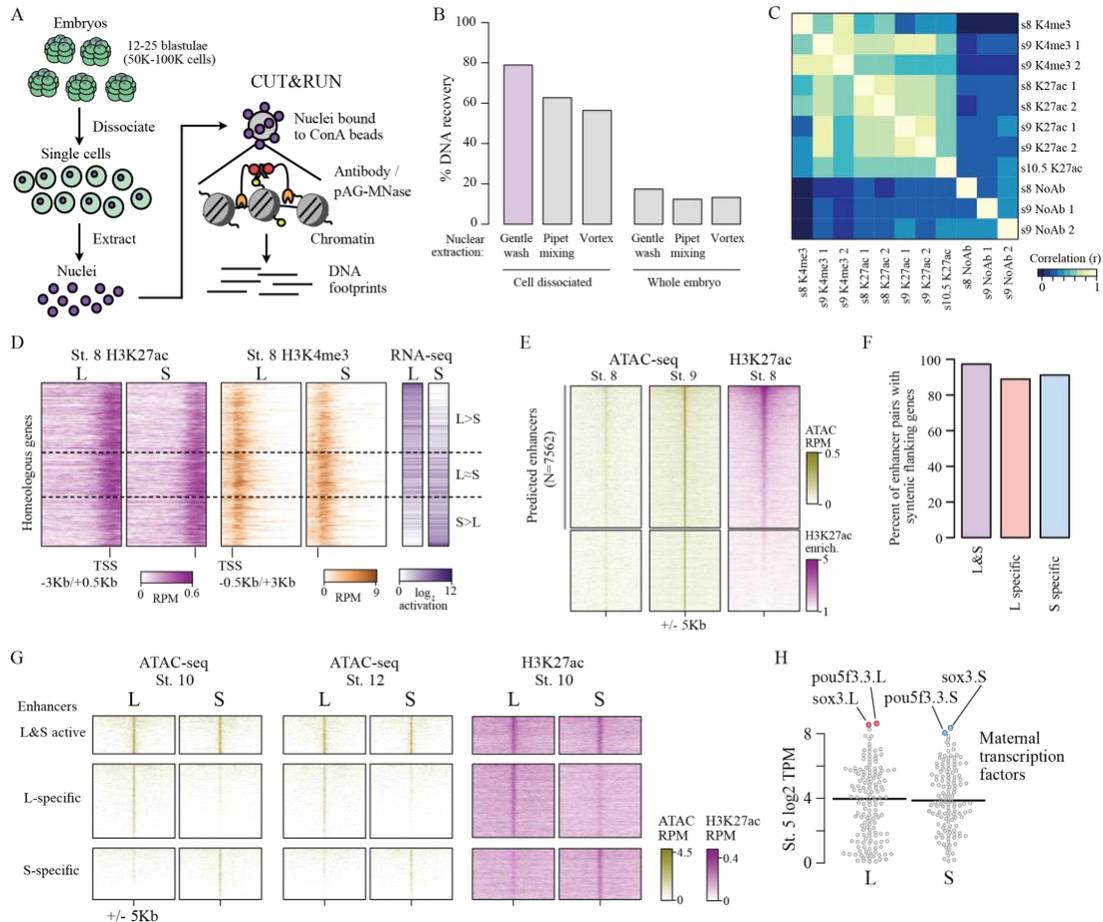


Figure 3-6: Profiling homeologous regulatory elements

(A) CUT&RUN for *X. laevis* blastulae requires cell dissociation prior to nuclear extraction. (B) Comparison of different nuclear extraction techniques. Percent DNA recovered was estimated by NanoDrop quantification of phenol-chloroform extracted DNA after nuclear extraction, as a percentage of theoretical total nuclear DNA mass based on the length of the reference genome sequence. Three nuclear extraction methods were tested with and without cell dissociation: gentle washing by pipeting buffer on the surface of the cells, pipet mixing, vortexing at 1500 rpm. (C) Heatmap of pairwise sample correlation between CUT&RUN samples, as measured by log₂ coverage in a 1 Kb window around the center of ATAC-seq open regions (N = 41083). (D) CUT&RUN coverage over paired homeologous gene regions around the TSS. Gene pairs are sorted according to L versus S RNA-seq activation ratio (right). (E) Heatmaps showing ATAC-seq peaks divided into putative enhancers with H3K27ac CUT&RUN enrichment, versus non-enhancers lacking H3K27ac. (F) Proportion of predicted homeologous enhancers that are flanked upstream and downstream by homeologous genes (at least 1 of the 5 nearest genes up/downstream). (G) Heatmaps of later-stage ATAC-seq (Esmaili et al 2020) and H3K27ac CUT&RUN coverage (this study) plotted over homeologous enhancers. (H) Maternal (stage 5) RNA-seq levels for L and S sequence-specific transcription factors, as annotated by Gene Ontology. *pou5f3.3* and *sox3* are the top expressed transcription factors for both L and S.

9B), suggesting a greater degree of regulatory innovation among differentially activated homeologs. Indeed, enhancers conserved between the *X. laevis* subgenomes exhibit significantly higher conservation with *X. tropicalis*, versus subgenome-specific enhancers ($P = 1.0 \times 10^{-300}$, χ -squared test, 4 d.o.f.) (Figure 3-9C, Figure 3-10A). However, total embryonic expression (i.e., maternal + zygotic) appears to be broadly maintained between *X. laevis* and *X. tropicalis* (Figure 3-9B), suggesting that much of the divergent subgenome activation is buffered by the maternal contribution, maintaining the stoichiometry of mRNA in the embryonic transcriptome.

This trend is also apparent at greater evolutionary distances. We found that genes activated in *X. laevis* are largely also expressed in zebrafish embryos (~450 million years separated) (Figure 3-10B). 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 ($P = 8.0 \times 10^{-12}$, χ -

squared test, 4 d.o.f.) and targeted by maternal homologs of OCT4 and SOX2, but also NANOG ($P = 1.5 \times 10^{-138}$, χ -squared test, 6 d.o.f.) (Figure 3-9D, Figure 3-10C-E). Interestingly, *Xenopus* and possibly all Anuran amphibians lack a NANOG ortholog, likely due to a chromosomal deletion (42). 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 (41, 42). 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.

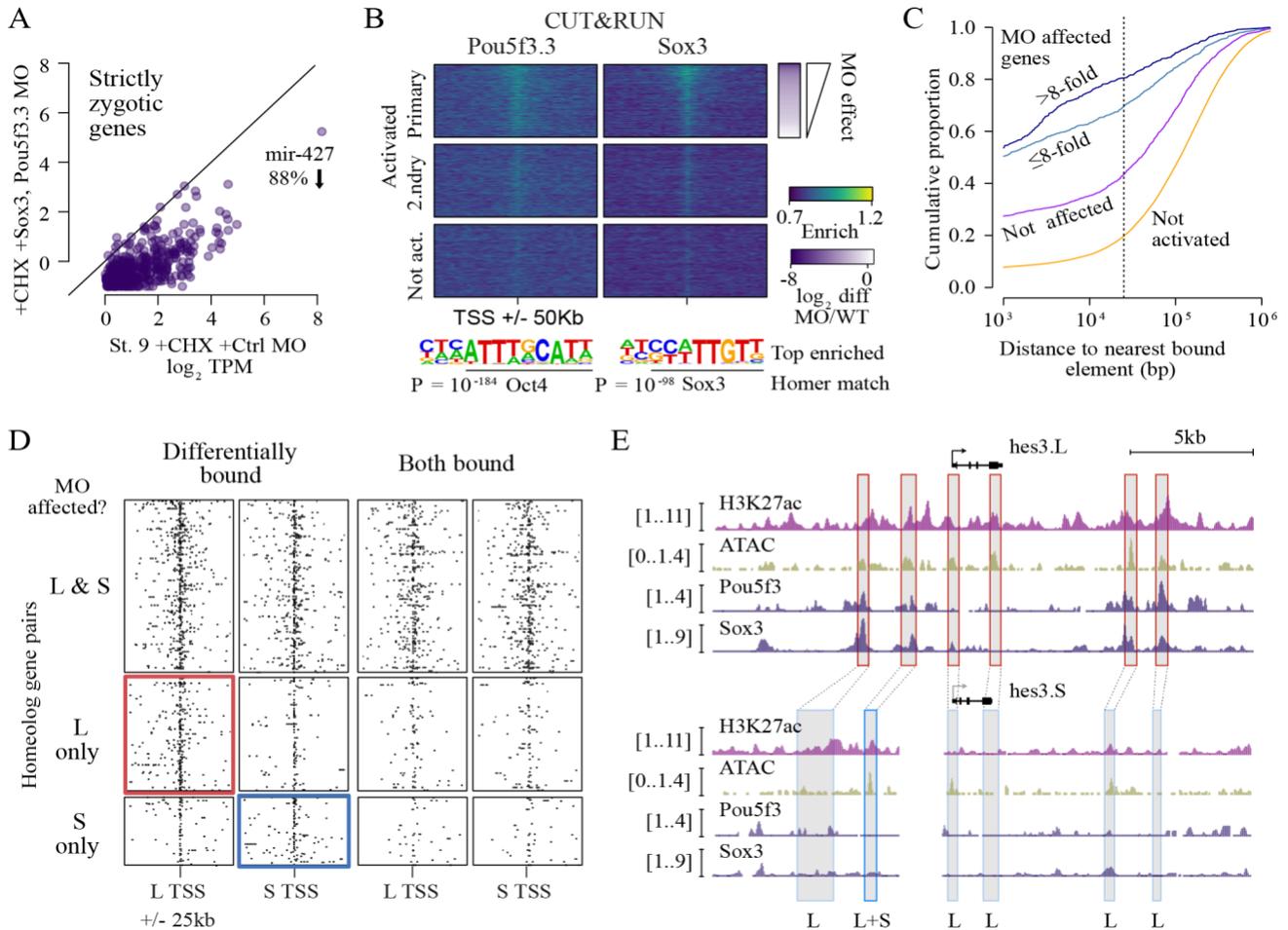


Figure 3-7: Pou5f3.3 and Sox3 binding drives genome activation

(A) Biplot showing inhibited gene activation in pou5f3.3/sox3 morpholino-treated embryos compared to controls. (B) Stage 8 Pou5f3.3 (left) and Sox3 (right) CUT&RUN coverage near TSSs for genes activated in the primary and secondary waves (top, middle) and unactivated genes (bottom). Primary activated genes are sorted by RNA-seq sensitivity to pou5f3.3/sox3 morpholino. Top enriched motifs for each factor are shown below. (C) Cumulative distributions of distance from a Pou5f3/Sox3-bound regulatory element for genes strongly (>8-fold) and less strongly affected by pou5f3.3/sox3 morpholino compared to unaffected and unactivated genes. (D) 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 pou5f3.3/sox3 morpholino treatment. (E) Browser tracks showing CUT&RUN enrichment and ATAC-seq coverage near active homeolog

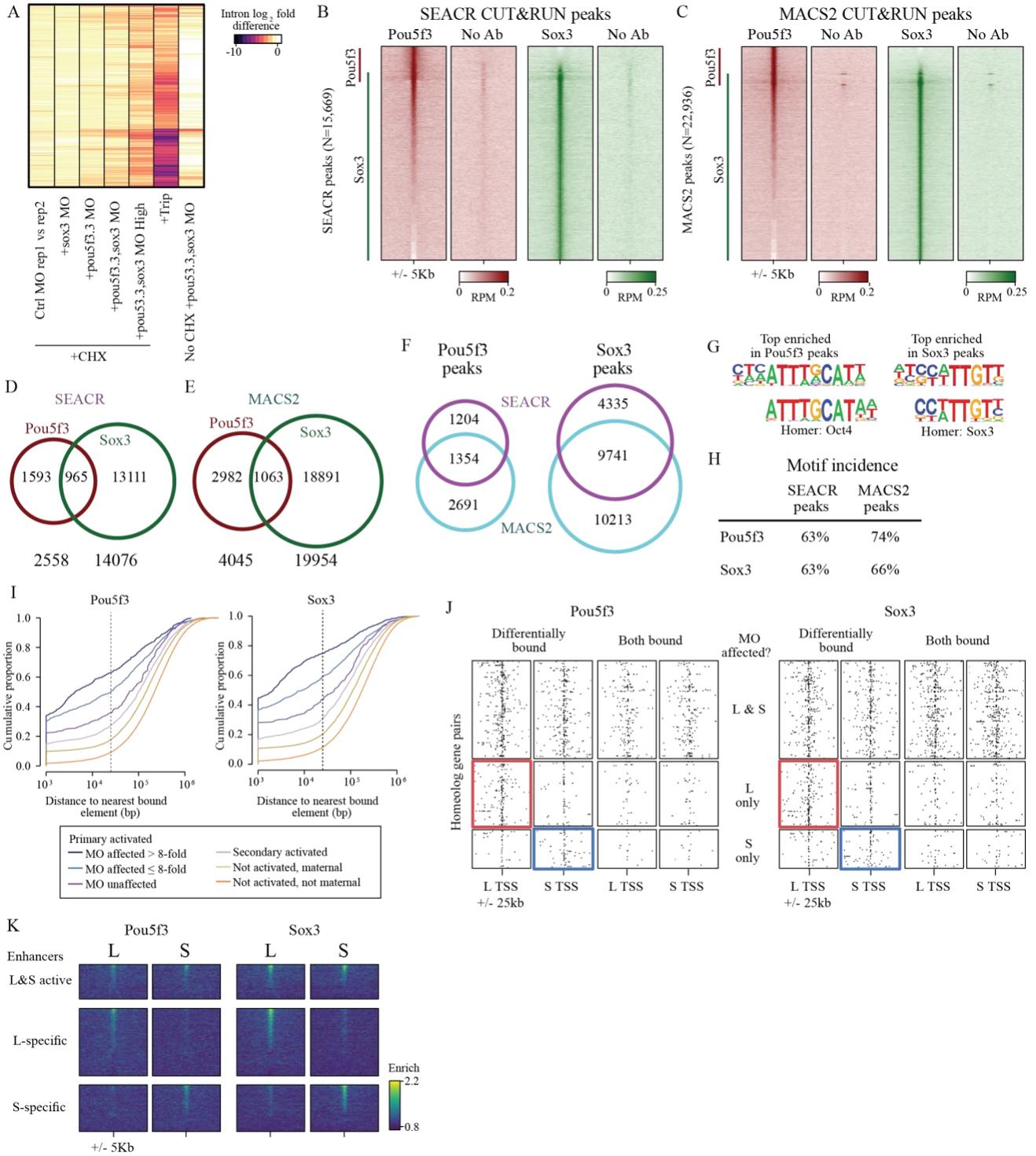


Figure 3-8: Assessing Pou5f3 and Sox3 roles in genome activation

(A) Heatmap showing RNA-seq intron log₂ fold difference compared to control for intron-containing activated genes. Column 1 compares replicates of cycloheximide (CHX) treated embryos injected with GFP control morpholino (MO); columns 2-5 show cycloheximide-treated embryos injected with individual sox3 and pou5f3.3 morpholino and both sox3 + pou5f3.3 morpholino at lower and higher concentrations compared to their respective control GFP morpholino injected embryos; column 6 shows triptolide-treated embryos compared to DMSO control; and column 7 shows embryos treated with sox3 + pou5f3.3 morpholino without cycloheximide compared to GFP morpholino. All samples are from embryos collected when untreated wild-type controls were at stage 9. (B, C) Heatmaps of Pou5f3.3 and Sox3 CUT&RUN coverage over SEACR (C) and MACS2 (D) predicted peaks. The union 491 of peaks per method is shown. Pou5f3 and Sox3 signal overlaps across most of the peaks, though it appears that both methods under-call Pou5f3 peaks. No Ab = no antibody. (D, E) Venn diagrams showing peak overlap between factors for SEACR (E) and MACS2 (F). (F) Venn diagrams showing peak overlap between methods. (G) Top enriched motif for Homer de novo motif finding for each factor compared to the closest database match. (H) Table of motif occurrence of the de novo identified Pou5f3 and Sox3 motifs among the peaks. (I) Cumulative distributions of distance from a Pou5f3-bound (left) or Sox3-bound (right) regulatory element. Curves represent gene groups according to the degree that they are affected by pou5f3.3/sox3 morpholino treatment. (J) Maps showing density of Pou5f3-bound (left) and Sox3-bound (right) regulatory elements around paired homeologous TSSs, divided into elements with differential homeologous L&S binding (i.e., one bound, the other not) versus both L & S homeologous region bound. TSSs are grouped according to whether both L & S homeologs are affected by pou5f3.3/sox3 morpholino treatment, or only one or the other homeolog is affected. For Pou5f3, $P = 1.5 \times 10^{-5}$, Kruskal-Wallis test for L-S differential bound enhancer count difference among the three gene groups; for Sox3, $P = 1.8 \times 10^{-5}$. For both-bound enhancers, $P = 0.13$ for Pou5f3, $P = 0.067$ for Sox3. (K) Heatmaps showing Pou5f3 and Sox3 CUT&RUN binding enrichment over no-antibody control, plotted over homeologous enhancers as previously defined by ATAC-seq and H3K27ac.

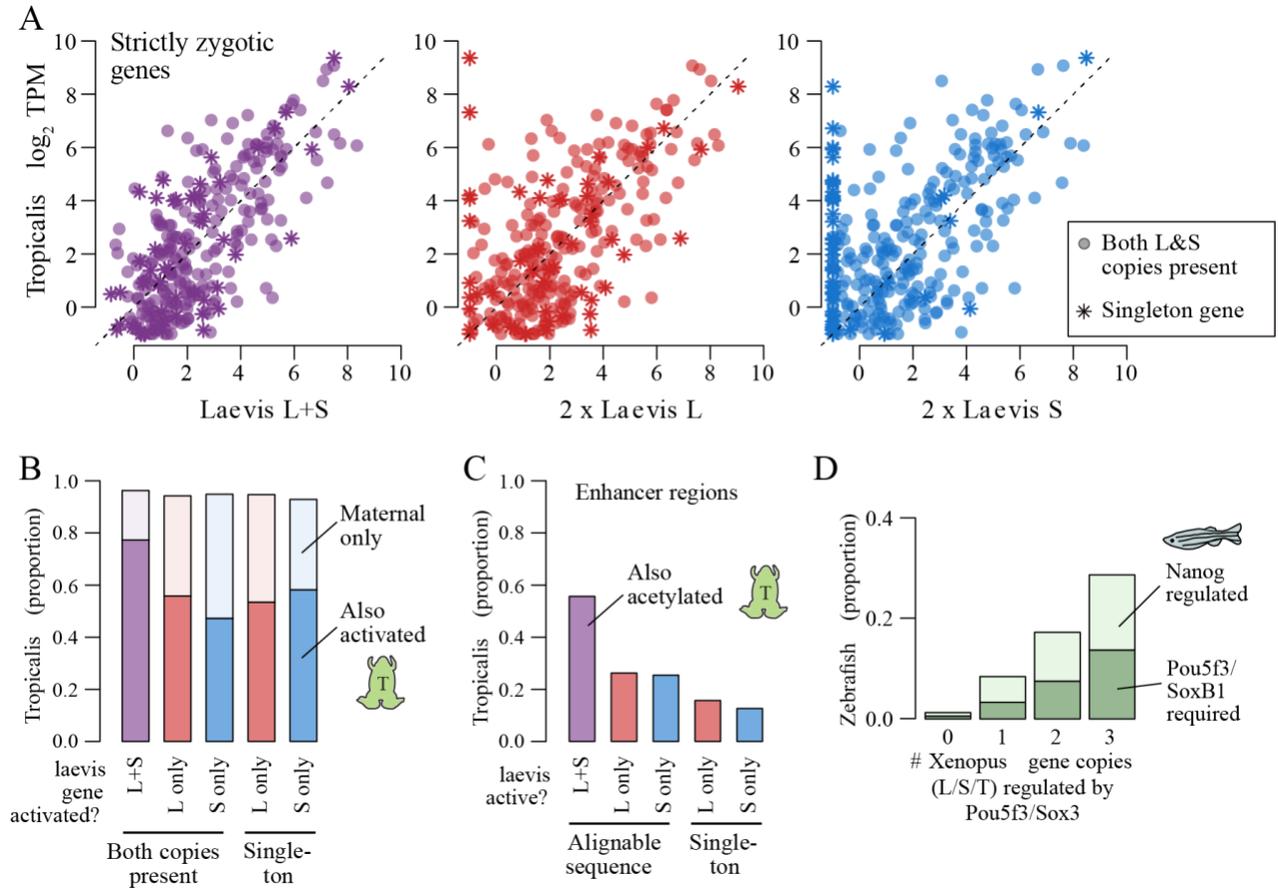


Figure 3-9: Regulatory divergence underlies dosage maintainence

(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 2x, since transcript per million (TPM) normalization is calculated relative to the entire *X. laevis* transcriptome. (B) Barplots showing the proportion of *X. laevis* genes across 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 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/SoxB1. 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.

3.4 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 3-11). Of the thousands of genes activated during the MZT, a majority of annotated homeolog pairs experience differential activation, which appears to be 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 (40, 93, 185), and binding motifs for some of these are found in differentially active *X. laevis* enhancers (Figure 3-5F). 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 (105, 207), we propose that regulatory upheaval in *X. laevis* post-hybridization (i.e., “genome shock” (208)) led to expression level gain or loss in one homeolog, which was subsequently corrected by compensatory changes to the other homeolog, possibly repeatedly (209, 210). 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.

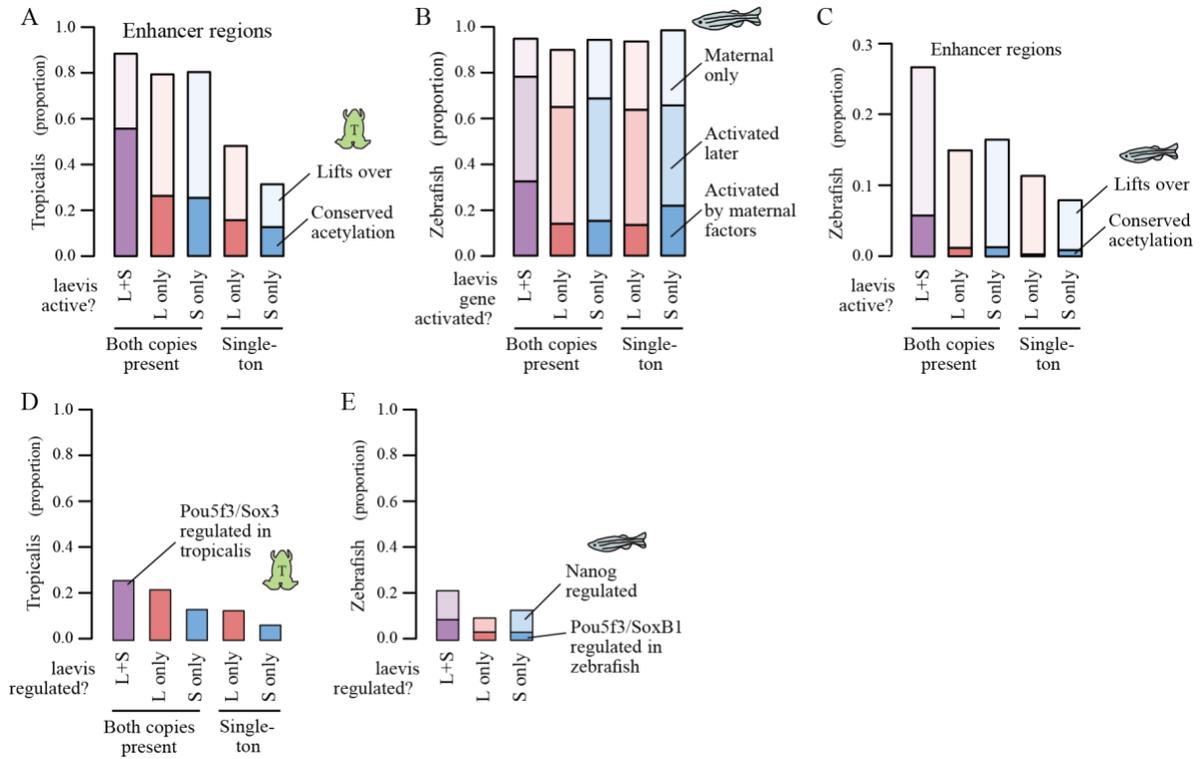


Figure 3-10: Shared patterns of activation with other taxa

(A) Similar to Figure 3-9: barplots showing the proportion of *X. laevis* enhancers across activity categories that are acetylated in *X. tropicalis*, additionally showing the proportion of enhancers that lift over but are not acetylated in *X. tropicalis*. (B) Barplots showing the proportion of *X. laevis* genes in different homeolog activation categories whose orthologs are also activated in zebrafish as part of the first wave by maternal factors, activated by 6 h.p.f., or part of the maternal contribution. Both activated homeologs are more likely to also be activated in zebrafish in the first wave ($P = 8.0 \times 10^{-12}$, c-squared test, 4 d.o.f.). (C) Barplots showing the proportion of enhancers that lift over and are acetylated in zebrafish according to Bogdanovich et al. 2012. L+S conserved enhancers have low conservation with zebrafish, but significantly higher proportion than L- or S-only enhancers ($P = 9.4 \times 10^{-80}$, c-squared test, 4 d.o.f.). (D) Barplots showing the proportion of Pou5f3/Sox3-regulated *X. laevis* genes also regulated by Pou5f3/Sox3 in *X. tropicalis* according to Gentsch et al. Both-regulated homeologs are more likely to also be regulated in *X. tropicalis* ($P = 4.2 \times 10^{-4}$, c-squared test, 4 d.o.f.). (E) Barplots showing the proportion of Pou5f3/Sox3-regulated *X. laevis* genes also regulated by Nanog/Pou5f3/Sox3 in zebrafish according to Lee & Bonneau et al. 2013. Both-regulated homeologs are more likely to also be regulated by Pou5f3/SoxB1 in zebrafish ($P = 0.0096$, c-squared test, 2 d.o.f.), but also more likely to be regulated by Nanog in zebrafish ($P = 3.4 \times 10^{-4}$, c-squared test, 2 d.o.f.).

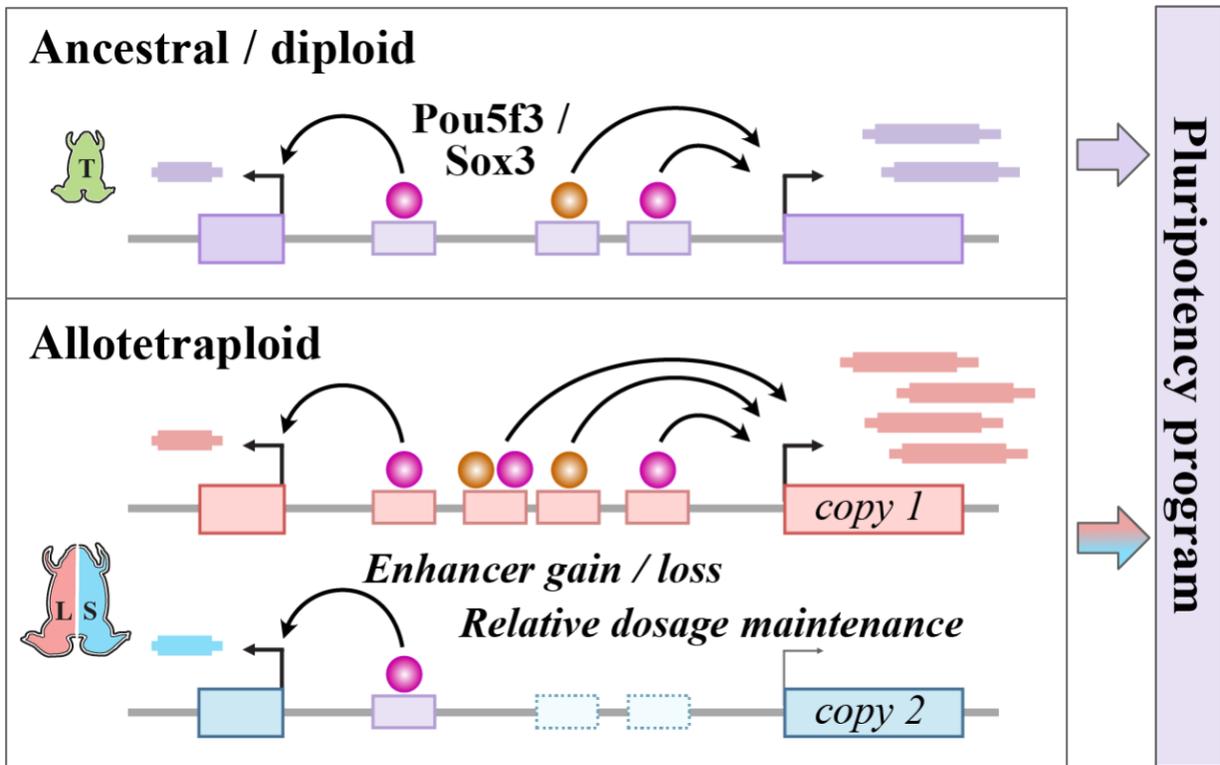


Figure 3-11: Model for Pluripotency Network Evolution

X. laevis likely underwent extensive enhancer turnover between its two subgenomes, which nonetheless maintained stoichiometry of pluripotency reprogramming in the early embryo.

4.0 Conclusions and Future Directions

4.1 Conclusions

In conclusion, we have demonstrated pluripotency factor homologs, Pou5f3.3 and Sox3, drive differential homeolog expression during genome activation through engaging subgenome-specific enhancers in such a way as to maintain homeolog dosage comparable to that of the ancestral *X. tropicalis* and zebrafish. In chapter 2, we developed a gapped-tiling oligo rRNA depletion strategy that yields high quality RNA-seq libraries comparable to those constructed using poly(A) selection. Additionally, we adapted the low-input ChIP-seq alternative, CUT&RUN for use in *X. laevis* embryos by dissociating the embryos prior to nuclear extraction and we adapted ATAC-seq for use in early *X. laevis* embryos by enriching our libraries for accessible fragments and reducing yolk inhibition. In chapter 3, we used the aforementioned techniques to characterize genome activation in the previously uncharacterized species, *X. laevis*, and find that differential homeolog expression compensates for the overall gene dosage despite the high regulatory turnover that gave rise to the subgenome-specific enhancers regulating this expression. Taken together, my dissertation uses *X. laevis* embryos to show how the pluripotency program adapts to regulatory disruptions and genomic instability in vertebrates, and potentially across animals.

4.2 Future Directions

My dissertation project expanded the repertoire of high-throughput sequencing tools available to embryological studies and has shed light on how the pluripotency regulatory program adapts to genomic disruptions. However, the project raises a number of further questions. One yet uncharacterized aspect for *X. laevis* MZT concerns how maternal mRNA clearance may differ between the subgenomes, and what the underlying regulatory mechanisms are that mediate these differences. My dissertation has demonstrated subgenome-specific regulation underlying differential homeolog activation – a similar paradigm could underlie maternal clearance with asymmetrical RNA degradation occurring to equivalently provided maternal homeologs. Such a phenomenon could be regulated by RNA binding proteins, such as Y-box, ARE-BPs, and/or EDEN-BPs that bind differentially encoded sequence motifs contained within L and S homeologs. Alternatively, miRNAs, such *mir-427*, could regulate L and S maternal homeologs in a subgenome-specific manner via the differential presence of seed sequence motifs in the 3'UTRs of homeologs. Thus, the selective depletion of only one maternally provided homeolog could be regulated by RBPs, miRNAs, or some combination thereof. Another interesting possibility is the utilization of entirely subgenome-specific mechanisms selectively degrade the homolog from their corresponding subgenome, *i.e.* an L mechanism that selectively degrade L homeologs and vice versa. Such mechanisms are likely to be ancestral regulatory differences that are perpetuated in the hybrid species and could be invoking maternally derived dosage compensation during oogenesis. This question is most easily addressed in *X. laevis* and the data generated from this project could easily be repurposed to look at RNA decay over time, with any additional time point being generated using the described rRNA depletion strategy from Chapter 2.

Another question that arises is when did these regulatory disruptions arise. My analysis has compared *X. tropicalis* and *X. laevis*, which are separated by ~48 million years, and the L and S subgenomes, diverged by ~36 million years (99). A comparative analysis of the regulatory landscapes of two more closely related genomes would further our understanding of regulatory

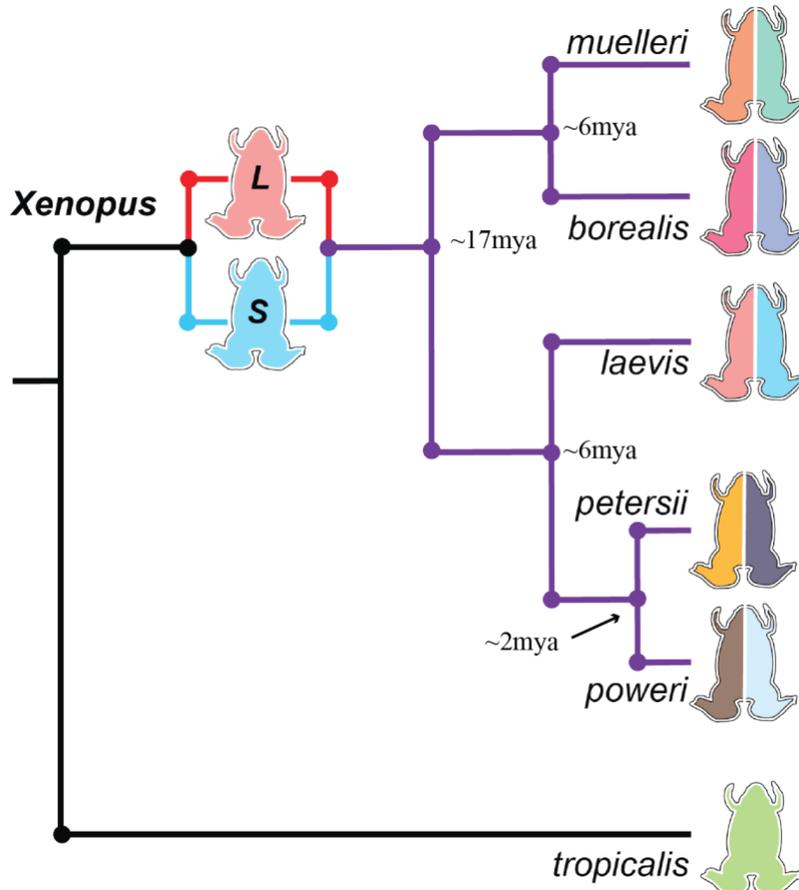


Figure 4-1: Additional *Xenopus* species

Comparisons with other allotetraploid *Xenopus* species could reveal novel regulatory paradigms underlying pluripotency induction.

adaptation at finer resolution and facilitate the study of more short-term responses. This question could be addressed by expanding my analysis to encompass more recently diverged *Xenopus* species. The sister species *X. borealis* and *X. muelleri* are diverged from one another by ~6 million years (Figure 4-1) (211). Additionally, they are diverged from *X. laevis* by ~17 million years, roughly a third of the evolutionary distance between *X. laevis* and *X. tropicalis* and half the distance

between the L and S subgenomes (211). Therefore, any syntenic regulatory regions that are active in both *X. borealis* and *X. muelleri* but not active in *X. laevis* represent novel regulation that evolved within 17 million years (since *X. laevis* split from *X. borealis* and *X. muelleri*). In addition, syntenic regulatory regions that are active in either *X. borealis* or *X. muelleri* but not active in *X. laevis* is also an example of novel gene regulation but is likely to have evolved within the last 6 million years (since *X. borealis* and *X. muelleri* diverged from each other.) Furthermore, two additional sister species, *X. petersii* and *X. poweri*, are only 2 million years diverged from each other and 6 million years diverged from *X. laevis* (Figure 4-1) (211). The addition comparison of these two species is highly likely to elucidate more immediate regulatory and expression changes following speciation (within the last 2 million years) and, thus are critical to understand the progression of regulatory divergence. Finally, any active regulatory regions that are shared among all 5 species would represent ancestral mechanisms within the clade. Thus, comparative genomics using four additional species, *X. borealis*, *X. muelleri*, *X. petersii*, and *X. poweri*, would represent a fine resolution expansion to the regulatory adaption of the *Xenopus* clade.

Another question that is raised is how much flexibility the regulatory response has. *X. laevis* has a specific homeolog expression configuration that facilitates pluripotency induction despite a massive genomic upheaval. But it remains unclear if the regulatory network had to respond in the way that it did or if there are other expression configurations that also would maintain pluripotency induction. In addition to the inter-species comparisons described previously, these four additional species can provide much insight via intra-species comparisons. All four of these species are allotetraploids that arose from the same hybridization event that gave rise to *X. laevis* (211). Therefore, each species has the potential to have vastly different L and S expression configurations and could each invoke a similar pluripotent regulatory network using different L

and S expression. For example, an L-activated homeolog pair in *X. laevis* that is also L-activated in each of the other 4 species would represent a highly conserved expression pattern and suggests either this pattern was established early on in their evolutionary trajectory and/or that the singular expression of the L homeolog confers some advantage over activation from only the S subgenome or equivalent subgenome expression during pluripotency induction. Conversely, an L-activated expression pattern that exists in *X. petersii* and/or *X. poweri* but not in *X. laevis*, *X. borealis*, or *X. muelleri* (Figure 4-1) likely represents a more recently evolved expression pattern unique to the progenitor of those two species and would demonstrate flexibility in the expression pattern of these genes when inducing pluripotency. A similar evolutionary trajectory can be inferred for a homeolog expression pattern that is conserved in *X. borealis* and *X. muelleri* but not in *X. laevis*, *X. poweri*, and *X. petersii* (Figure 4-1). Finally, homeolog expression patterns with no discernable evolutionary trajectories would represent genes with no observable evolutionary constraint during pluripotency induction. A comparison of genome activation between these five frogs has the potential to identify homeologs that are required to be expressed from one subgenome versus the other or which subgenome's enhancers (if they have to be subgenome-specific) are under evolutionary pressure to maintain their ancestral function. Thus, these comparisons could establish new evolutionary paradigms in the vertebrate lineage.

To address these questions, it is necessary to develop well characterized genomic and transcriptomic infrastructure as all four of these new species are completely uncharacterized and no assembled genome or transcriptome exists, let alone annotations. In addition, characterizing their husbandry and gamete acquisition for the purposes of creating such resources for each species will also be necessary. One of the more straightforward resources to develop would be an embryonic transcriptome whose genes can be further annotated and aligned to. By collecting RNA

over a developmental time span that encompass purely maternal transcripts through gastrulation, it is possible to conduct long read sequencing via Nanopore sequencing to obtain long reads that can be assembled into intact transcripts. These transcripts can then be compared with those of *X. laevis* and *X. tropicalis* to begin establishing their identity as well as their subgenome of origin. The Nanopore sequencing can also provide the sequence of each species' rRNA and oligos can be designed for rRNA depletion probes to enrich for mRNA using the Oligo-ASST tool described in Chapter 2. Once assembled and adequately annotated, rRNA depletion RNA-seq libraries can be constructed for sequencing and aligning against the new transcriptomes. Additionally, it is likely both CUT&RUN and ATAC-seq can be applied to each species with minimal adaptation due to the high relatedness of each species. However, to fully utilize these techniques, it would be necessary to develop a genome to which to align in order to identify active regulatory regions. Thus, expanding the intra-species comparative genomics to encompass additional *Xenopus* species could reveal novel regulatory paradigms underlying pluripotency induction.

Appendix A : XL rRNA oligo pool

Target	Tm	Antisense_oligo
28S	70.1	TTCCTCCGCTTAGTAATATGCTTAAATTCAGCGGGTCGCC
28S	79.8	ATTCGGCGCTGGGCTCTCCCTCTTCACTCGCCGTTACTG
28S	82.5	CGGTCTCCCGTACGCCACACGTCCC CGCGCCCGGGCG
28S	77.9	GGCCTAACACCGTCCGCGGGCTGGGCTCGATCAGAAGGA
28S	71	CTTTGGGCTGCATTCCCAAACAACCCGACTCCGAGGAGAC
28S	76.6	CTTTCCTTACGGTACTTGTCCGCTATCGGTCTCGTGCCGG
28S	77.1	CACGGCCCCACCCGTTTACCTCTTAACGGTTTCACGCCCT
28S	79.9	CGGCGGCGCTGACCCGCCGGTTGAATCCTCCGGGCGGA
28S	84.5	CGCGCGACCGCCCTCCGGCGAGGGGAGGGGGCCGGCGG
28S	78.7	GCGGCCCGGAGCCGGCGCGCACCGCCGCGGAGGAAACG
28S	83.1	TAGCCGCGCGCGGGGGCGGCCGGCCACCTTCCCCTGAAC
28S	83.5	CGGAGGCCGCGTCTCCCTCGGCCCGGGCGACGGCGAG
28S	81.2	CGGGGGGAGCGGCGGGACGCGCTCCGGGGAGGAGGACCG
28S	86.5	CGCCCCGGTTGACAGCCGCGCCGGGAGCGGGGGGCCCTT
28S	83.1	CTCCCGCCTCGGCGGCGGGCGGACGGGGCGCACTGG
28S	74.4	TGGTCCGTGTTTCAAGACGGGTCCGGTGGGACACCGACAT
28S	79.7	CCCTCACCTTCAATTGCGCCACAGGGTTTCGCGCAGAGTC
28S	83.4	CGGAGGGAGGGGCGGCGGGATCCACCTCAGCCGGGGCG
28S	82.3	CGACGGGGCGGGCGAGACGGGCCGGTGGTGGCCCCGCCGG
28S	79.7	TGGCTTCGCTGCCAGGCATAGTTCACCATCTTTCGGG
28S	74.1	CGCCCCTATACCAGGTCGGACGACCGATTTGCACGTCAGG
28S	72.5	ACGGACAGCGCCAGCTATCCTGAGGGAACTTCGGAGGGA
28S	70.2	GATTCGGCCCCAAGACCTCTAATCATTGCTTTACCGG
28S	79.8	CCCGGCTCCAAGCCAGCGAGCCGGGCTTCTTACCCATTTA
28S	74	GTTTCGGTTCATCCCGCAGCGCCAGTTCTGCTTACCAAAAG
28S	70.4	CTGCTGTCTATATCAACCAACACCTTTTCTGGGGTCTGAT
28S	75.4	AGGGCTAGTTGATTCGGCAGGTGAGTTGTTACACTCCT
28S	79.9	CCCCGCGGACTGACCCAGCGCCGGCGACGGCCGGGTATG
28S	84.6	CGCTTCCGCGCCACCGCCGGGGCCCTCCTACTCAGTCG
28S	70	GAATATTTGCTACTACCACCAAGATCTGCACCCGCGGCGG
28S	71.6	CCGACTGACCCATGTTCAACTGCTGTTACATGGAACCCT
28S	79.9	TTTCGATCGGCCGAGGGCGACGGAGGCCATCGCCCGTCCC
28S	79.5	CCCGTCTCCGCCACTCCGGGTTCCGGGGATCTGAACCCGA

28S	85.3	CACTGGACGCCCCGCCCCCGCCGCCGCGAGGGCTCG
28S	70.9	GCCCTTCACAAAGAAAAGAGAACTCTTCTCCCGGGCTCCC
28S	78.6	CCGCCGGAACCGCGACGCTTCCAACGGCGCGGGCCCCTC
28S	78.1	TACGGCCCCGGCGCAGAGATTTACACCCTCTCCCCGGATT
28S	71.6	GACTTCCCTTACCTACATTGTTCTAACATGCCAGAGGGCTG
28S	79.9	CCCCAGCCCGACCGACCCAGCCCTTAGAGCCAATCCTTA
28S	78	CGGGAGGAGGGAAAGAGCGCCACGGGGGCGCCTTCGTCC
28S	85.3	CGCGGCCCCCGCGAGGGGGGGCCGGAGAGAGAGGGGGG
28S	83.6	AAGGGCCCCGGCGCGCGTCCAGAGTCGCCGCCGGGGCG
28S	81	ACGGGGGGGAGGCGCAGGAGGGGGACGGCGCGGGGGAGAG
28S	85.4	TAAGTCAGCTGCTAGGCGCCGGCCGAGGCCGGGCCCGCCC
28S	71.2	AACACCCGCCTCGGGCCTTCGCGATGCTTTGTTTTAATTA
28S	68.9	CCC GCGTTCAATTGAATTTCTTCACTTTGACATTCAGAGC
28S	70.7	CGAGGCATTTGGCTACCTTAAGAGAGTCATAGTTACTCCC
28S	70.4	TTTCGCTAGATAGTAGGTAGGGACAGTGGGAATCTCGTTC
28S	74.7	AGACTAGAGTCAAGCTCAACAGGGTCTTCTTTCCCCGCTG
28S	79.2	CCCTTTGCGACGAGCGCGGGGGCCTCCCACTTATCCTAC
28S	71.5	TCGGGGCTCGCCCCGCCTCACCGGGTAAGTGAAAAACG
28S	81.4	CCCGGCGCGGGGGCCGGCGCTTGGGTCCAGAAGCGAGAG
28S	74.5	CGTTACGGTTTGACAGGTGTACCGCCCCAGTCAAACCTCCC
28S	78.6	AAGCGAGCTTTTGCCCTTCTGCTCCACGGGAGGTTTCTGT
28S	70.2	AAAAGTCAGAAGGATCGTGAGGCCCGCGTTTACGGTC
28S	76.3	TGAACGCTTGCCGGCCACAAGCCAGTTATCCCTGTGGTA
28S	66.4	GCTTGGTGAATTCTGCTTCACAATGATAGGAAGAGCCGAC
28S	70.2	ACTAACCTGTCTCACGACGGTCTAAACCCAGCTCACGTT
28S	74.6	GTCTGAACCTGCGGTTCTCTCGTACTGAGCAGGATTACT
28S	76	GCGTTCAGTCATAATCCCACAGATGGTAGCTTCGCCCAT
28S	77.4	TAATCCGAGGCCGACCGAGGCTCCGCGCGCTGCGGTATC
28S	83.3	CTTTCGTCCGCGCTCCGGTCCCGAGGCCGAGCGGCTCTGC
28S	70	GTCTACGAATGATTTAGCACCAGGTTCCCCACGAACGTGC
28S	71.8	GATGACTTTCAATAGATCGCAGCGAGGTAGCTGCTCTGCT
18S	68.6	CTTTGAGACAAGCATATGCTACTGGCAGGATCAACCAGGT
18S	70.5	ACTGATTTAATGAGCCATTCGCAGTTTCACTGTACCGGCC
18S	66.9	GTCGGCATGTATTAGCTCTAGAATTACCACAGTTATCCAA
18S	78.4	CCGGGGCGCGGGGGCCCCGGATTGGTTTTGGTCTGATAA
18S	70.7	CATCCGAATGTATCGTCGCCGTACGGGGACGTGCGATCG
18S	79.5	CCTGATTCCTCGTTACCCGTGGTCACCATGGTAGGGCA
18S	79	GTAATTTGCGCGCTGCTGCCTTCCTTGGATGTGGTAGCC
18S	70.1	GAAAGAGTCCTGTATTGTTATTTTCGTCCTACTACCTCCCC

18S	68	CACCAGACTTGCCCTCCAATAGATCCTCGTTAAAGGATTT
18S	67.3	AATATACGCTATTGGAGCTGGAATTACCGCGGCTGCTGG
18S	79.3	CCGCTCGCGGGGACCGCCAGCTCGATCCCAAGATCCA
18S	79.3	CGGGCCCCGGGACACTCAGTCAAGAGCATCGGGGAGGCG
18S	74.9	TATTCCTAGCTGAAGTATCCAGGCGACGCGGCCTGCTTG
18S	74.9	CCCGGCCGTCCCTCTTAATCATGGCCCCAGTCCGAAAAC
18S	70.4	AAATGCTTTCGCTTTGGTTCTGCTTGCGCCGGTCCAAGAA
18S	70.4	GACGGTATCTGATCGTCTTTCGAACCTCCGACTTTCGTTCT
18S	72.4	AAGCTGCTCGGCGGGTCATGGGAATAACGCCCGGGATCG
18S	70.7	GTCAATTCCTTTAAGTTTCAGCTTTGCAACCATACTCCCC
18S	73.9	CCGGGCCGGGTGAGGTTTCCCGTGTGAGTCAAATTAAGC
18S	70.1	GGCCATGCACCACCACCCACAGAATCGAGAAAGAGCTATC
18S	67.6	GTTAGCATGGAGGAGTCTCGTTCGTTATCGGAATTAACCA
18S	72.2	TCGTGTGGCTGAACGCCACTTGTCCCTCTAAGAAGTTGGA
18S	77.6	CCGTTCAAGTGTAGCGCGCGTGCAGCCCCGGACATCTAAGG
18S	76	ATCCCTATCACGAACGGGGTTCAGCGGGTTACCCGCACCT
18S	71.1	ACGCGAGCTTATGACCCGCACTTACTGGGAATTCCTCGT
18S	70.4	ACCTCACTAAACCATCCAATCGGTAGTAGCGACGGGCGGT
18S	79.1	CAAGTTTGATCGTCTTCTCGGCGCTCCGCCAGGGCCGTGG
18S	70	GATCCTTCCGCAGGTTACCTACGGAAACCTTGTACGA
5.8S	70.7	ACTAATTCTCGCAGCTAGCTGCGTTCATCGACGCACG
5.8S	78.7	GGCCCCGGGAGGAACCCGGGGCCGCAAGGTGCGTTCGAAG
5S_mat_som	73.1	GCTTCYGAGATCAGACGAGATCRGGCACTTTCAGGGTGG
5S_mat_som	77.7	AAGCTACGACACCTGGTATCCAGGCGGTCTCCCATC
16S	60.7	GAGGTTATTGTTATTATAATTGGAATGCTAGGTTTTTGG
16S	58	GCTATTATGATTGTTCTATCGCCTATACTAAAATTTTAGA
16S	65.8	TTTCATTTCTATTTTCATCTTTCCTTACGGTACTTTTTCT
16S	69.8	AGGTACGAGGTAAGTTCTCTGCTTTTTGTTGCTTAGTTA
16S	66.5	TCGTTTTGCTTGATTATGACTGGCTAGACCATTATGCA
16S	71.1	AAAGCTGTCTCGGAGTAGATCGCTTAGTTTCGGGTAGTC
16S	76.8	CTCGTTTGGTCTGTACACCCTACTCGGAGATCTTCCCAC
16S	67	GGGTAGAACTTATATTCATTTCTGAGCAACCAGCTATCA
16S	63.3	TCAGGCTGTACCCTGATTGAATAAATCCTAAGTAGACTTT
16S	65.4	ATAATCTTTACCCAGTATTATAGGTTGTATCCTGTTTCA
16S	74.2	GAGTGAGCTTTGACGCTGTCTTTACAGGTGGCTGCTTTTA
16S	65.1	AGCTCAGTATTGTTTGGGGTTTAGAATTAGTTATACTA
16S	66.7	ATCGTGTATCACACTAGTTCTAGCATAAGTGCTTCT
16S	64.3	GGATCTCAGGGAGGACGTTAATTATCAGTGATTTATTCGA
16S	64.4	ACGGTAATAAGTGCATGGTTTTCTGTTTTGTTATTGCA

16S	73.8	GGGGTTCATAGTTTGCCGAGTTCCTTCTGCGTCTTTAAT
16S	68.5	CTTGGGCAGGCTGGACCTTTATAACAATGTTTTTAAGCA
16S	70	AGACAGTGATTACGCTACCTTTGCACGGTCAGATACCGC
16S	70.1	ATGGATTGGATGCAGGAGACAGTTGAACCTTCGTGGCCG
16S	70.6	TTAGTTTAAAGCTCCATAGGGCTTCTCGTTCTATGGTCT
16S	70.2	GGTCAGTTTCTGCTTGTTAATTGTTATTCCTTATGGGT
16S	71.8	GGTGGTAGGCCCTATTCTTCAAGGAGGATTTTTTATTCTC
16S	62.5	AGTCATAAATTTTGTACTTAGAATGGTGGTCTTGGTGA
16S	70.3	AATGGATTGCGCTGTTATCCCTAGGGTAACTTGGTTCGT
16S	77	GGCTGCACCACTGGGATGCCCTGATTCCAACATCGAGGTC
16S	72.7	CCTGGATTACTCCGGTCTGAACTCAGATCACGTAGGGCTT
16S	65.2	ACATTGGCCTCATTTTTTCGGTCTTTTCGTA TAGAAAA
16S	64.7	CTTGGGCAGAGTATTAGTTCTATTTTCAATTTAGTTGACT
12S	70.1	TCGAGCTTTGTGTCGCTCATGTTCAATGTTTACTACTGC
12S	70.3	ACTTGAGTTTCTCGTATAACCGCGGCGGCTGCACGAGAT
12S	72.5	TGCGACAGCTTGGTTGGAGTTTGACTCTAGTTTGGGTTAC
12S	68	GTGTAATATGGGTAGAGTTACTTTCGTGAGTGTCTTCA
12S	70.2	GTCAAAGTTTATGGCTAGGCATAGTGGGGTATCTAATCCC
12S	74	TTTGGAGCACCGCCAAGTCCTTTGGGTTTTAAGCTTAGGC
12S	75.2	GGCGGGTTTGGCAAGAAGTGGTGAGGTTTAGCGAGGGGTA
12S	70.3	AAGCCTACTAAGAATCTCTCACGAGGTGGGCTGGCGACGG
12S	71.1	AATGTAGCCCATTTCTTCCCCTTCATATGCTACACCTT
12S	71.6	ATCCGCCTTTTCTCGATCTGGTTTCATAGAGATCTTTCGT
12S	73.6	CGGTGTGTGCGCGCTCCAGGGCCGTTTTAAAGAGGAACTC
12S	66.2	GTGTGTTTATAGACGTTGGTTGATTTTTGTAGAAGAGGGT
12S	78.7	CCAAGCACACCTCCGGCTGGTGATCCCGATCTCAGTGGT
12S_alt	72.6	GGCATTTTACGGGAGTGGGATGCTTGCATGTGTAAGTT
12S_alt	72.5	AGAGCAAGGTGTCATGGGCTATTAGTTGTCCTGATACCG

Appendix B : XL test oligo pool

Target	Tm	Antisense_oligo
16S	78.2	CTTTAGTAGCGGCTGCACCACTGGGATGCCCTGATTCCAACATCGAGGTC
16S	65.2	TCCGGTCTGAACTCAGATCACGTAGGGCTTTAATCGTTGAACAAACGAAC
16S	58.4	CCTTTCGTACTAGAAAAATACTTCATAGATAGAACTGACCTGGATTAC
16S	64.7	ACATTGATATAGAGAGAGGCTTATTAACATTGGCCTCATTTTTTCGGT
16S	69.3	TAGCCCTAATCTTGGGCAGAGTATTAGTTCTATTTCAATTTAGTTGACT
COX2	69.5	GAGGCTGCGTCTTGAAAACCTAATTGTGATGGGTGTGCCA
COX2	68.3	GGCTATGAGGGTATGGTCGTGGAAGTAAGTAATTCTTCT
COX2	70	TCACACTATTTTCGATCTTGTGCGTCCATTAGGTTTGTA
COX2	67.2	TCATTGGTGGCCGATTGCTTTAATTGTTAAGTGTGGATCA
COX2	70.7	TTATCAACTTCTAGCAGCCGGAATTGTCCAGGGGTAAGGT
COX2	71	TTGATGAAGTCGTCCTGGGATTGCATCTGTTTTGACACCC
COX2	70.2	AGCTGTGGTTTGCTCCGCAAATTTCTGAACATTGTCCGT
COX3	47.4	GTAAAAGAATTATTGATCCAAAGTGAAATCATATAG
COX3	54.2	TCATCATTGAATTATAGTTAGTACTATAGTAATTAGG
COX3	53.1	TAATAAATAGGATTATCCATATCGTAATCCTTTTTG
COX3	47.4	TAAAATGCTCAGAAGAATCCAATAAAGAAGAATACTT
COX3	55.2	CTGTGTTTTAAAAGTGGAACTTCAAATGGGTTTAATGG
COX3	51.5	AATAAAGTCCAAGAAGAATGGTTAAAGTTAGTGATTGAAT
COX3	55.2	ATGAAGACCGTGGAAACCAGTTGCTACAAAAAATGTT
COX3	48.8	AAGAAGACAAACAGATAGGAATAATGAGCCAATAATG
COX3	53.1	ATCAATAGATCGATACGTAAAGGAATAGTCATACTAC

Appendix C : Zebrafish oligo pool

Target	Tm	Antisense_oligo
28S_mat_som	68	TCTTTTCTCCGCTTAGTAATATGCTTAAATTCAGCGGGT
28S_mat	80	GGGCGGGGATTCGGCGCTGGACTTTTCCCTCTTCGCTCG
28S_mat	82	CGGCCCGCGCCGAGAGAGGCGGCCCTCCGTACGCCACAG
28S_mat_som	78	ACCGGCCTCACACCGTCCRCGGGCTAAGCCTCCATCAGAA
28S_mat_som	74	CCGCTTTGGGCTGCATTCCCAAACAACCCGACTCCGGGA
28S_mat_som	73	CTATCGGTCTCGTGCCGGTATTTAGCCTTAGATGGAGTT
28S_mat_som	71	ACGGTTTCACGCCCTGTTGAACTCTCTCTTCAAAGTTCT
28S_mat	80	CGACCCCGCCCCGCCGGGCTGAATCCACCGGGCGGACGG
28S_mat	80	CGGGCGTCAACCCCCGCGCCACCCCCAGGAATGAGGGAG
28S_mat	79	GCCGAACCGGAGCCGGTTCGCGGCGCTCCACCGCGGCGGAA
28S_mat	79	CCGACGGCCTGAACCGACGGGCCACCTTCGCCCCCTGAC
28S_mat_som	75	AGCAAGCGGCGAAGTCGGGGCGGGRRGGGCGCTGTAAAGC
28S_mat	80	CCCACAGGGAGGGAGGGCAGAGAGGGCGCGGAGGACACT
28S_mat	80	CGGACTGAGGACAGTCCACCCCTGTCGCGCCACGCCGGG
28S_mat	81	GTCGCCGCGGACCTCGGGCGCCGAACGTGGGTTCGGATCC
28S_mat_som	80	ACCCTCTGACTCGCGCGCGGTTRGACTCCTTGGTCCGT
28S_mat	80	GATCCCACCCGGGGGAGACCCCCGTTCTCACCTTCATTG
28S_mat_som	89	CGAGRCGGGCCGGTGGTGCGCCCCCGCCGGGGCGGGGG
28S_mat	77	TTCACCATCTTTCGGGTGCCATCGCGCGCGCTCCAATC
28S_mat_som	83	CACCAGAGTTTCCYCTGGCTTCGCCCTGCCAGGCATAG
28S_mat_som	76	TTCGATTAGTCTTTCGCCCTATRCCCAGGTTCGGACGACC
28S_mat_som	72	GCGAGCGCCAGCTATCCTGAGGGAAACTTCGGAGGGAAC
28S_mat_som	70	GCCCCAAGGCCTCTAGTCATTWGTCTTACCGGATAAAAC
28S_mat_som	70	CCAGCGAGCCGGGCTTACCCATTTAAAGTTTGAGAA
28S_mat	72	CCCGCAGCACCAGTTCTGCTTACCAAAAATGGCCCACTAG
28S_mat_som	77	TGGGGTCTGATGAGCGTCGGCATCGGGCGCCTTAACCCG
28S_mat_som	76	CCGACTTCCATGGCCACCGTCCTGCTGTCTATATCAACC
28S_mat_som	79	GCTCCAGCGCCATCCATTTTCAGGGCTAGTTGATTCGGCA
28S_mat	80	AGCGCGCGGACCGCTCGCGTGTCCCTTGTGCCGTCGACG
28S_mat_som	83	TTCWGCGCCACCGCGGCGGCCCTCTACTCGTCGAGGCTT
28S_mat_som	71	AATATTTGCTACTACCACCAAGATCTGCRCCCGCGGCGG
28S_mat_som	73	ACCATGTTCAACTGCTGTTACATGGAACCTTCTCCA
28S_mat	80	GCGACAGAGGCCATCGCCCTCCCCTCCGAACGGCGTAG

28S_mat	80	TCCGGGCTCGGGGATCTGGACCCGACTCCCTTTCGGTCCG
28S_mat_som	80	CGGGTTCGTTTGCGTACCCGACTGGGCGCCTCGCGGCG
28S_mat_som	69	TTCCAGGGCGCCCTGCCCTTACAAAGAAAAGAGAACTCT
28S_mat_som	84	CGGCGCTTTCCAGGGCGCGGGCCCCTCTCTCGGGGCGAAC
28S_mat_som	73	TTACACCTTCTCCCCGGATTTTCAAGGGCCGACGAGAG
28S_mat_som	77	TGTTCCAACACGCCAGAGGCTGTTACCTTGGAGACCTGC
28S_mat_som	71	CAGCCCTTAGAGCCAATCCTTATCCCGAAGTTACGGATC
28S_mat	85	CGGCTCGGGCCAGGCCCGCTTCGCACCTCAGCCCGACCG
28S_mat	81	CACCGCGCGCGCACCCCCGCGGCTCGGGAACGGGGGTGAG
28S_mat	73	GAGTCGGGGGGAGTCGAGACCGCAAAAAGGCCCGAGGCGG
28S_mat	80	CTCCGAGACCCCCCGCCCGACCCCGCGGACGGGGAAA
28S_mat	80	CGACGCCACGAACCACCCTTTCCTCCCCACGGCCTCTC
28S_mat	81	CAGCGGAGGCATCGCGCGGCAGGCGAGGACCCTCCCGTTG
28S_mat	82	AAGGGCGCGGCGCGGCCAGGGTCGCCGCCGAACCG
28S_mat	82	CGGACGCGGGCCCCGGCGCGGGGCCGTAGCTGCGGAGAT
28S_mat	84	CGGCCGAGGCGCCCGCACCGGGGGGAGCCGAGGCCGCGCG
28S_mat_som	70	GCTTTGTTTTAATTAACAGTCGGATTCCCCTGGTCCGC
28S_mat_som	71	TCAGAGCACTGGGCAGAAATCACATCGCGTCAACACCCGC
28S_mat_som	74	GCTACCTTAAGAGAGTCATAGTTACTCCCGCCGTTTACCC
28S_mat_som	72	AGCAGGTAGGGACAGTGGGAATCTCGTTCATCCATTCATG
28S_mat_som	74	AAGCTCAACAGGGTCTTCTTTCCCCGCTGATTCTGCCAAG
28S_mat_som	80	GGGCCTCCACTTATTCTACACCCCTCATGTCTCTTCAC
28S_mat_som	67	AGTGAGGAAACGATAAGAGTAGTGGTATTTACCGGCGGC
28S_mat	81	CTTGACGCCAGAAGCAGGGGCGGCCCCCGCCGGGGGACG
28S_mat	84	TCGCGCCCCGGTTCGGCGGCCGGGAGGGGAAGGGGGTTG
28S_mat	79	AGGTGTACCGCCCCAGTCAAACCTCCCCACCTGACGCTGT
28S_mat_som	76	CCCTGAGCTCGCCTTAGGACACCTGCGTTACCGTTTGAC
28S_mat_som	66	CTGAAAATCAAGATCAAGCGAGCTTTTGCCCTTCTGCTC
28S_mat_som	77	GAAGGATCGTGAGGCCCGCTTTCRCGGTCCGTAATCAT
28S_mat_som	75	TTGGCCGCCACAAGCCAGTTATCCCTGTGGTAACTTTTCT
28S_mat_som	64	TCTGCTTACAATGATAGGAAGAGCCGACATCGAAGGATC
28S_mat_som	71	AAACCCAGCTCACGTTCCCTRRTTAGTGGGTGAACAATCC
28S_mat_som	66	ACATCAGTAGGGTAAACTAACCTGTCTCACGACGGTCT
28S_mat_som	70	GCGGTTCTCTCGTACTGAGCRGGATTACTATTGCAACAA
28S_mat	80	GTGGCTTCGCGCCAGTGGCTCCTCAGCCAAGCGCACGAA
28S_mat_som	70	TTAGGCGGGATTCTGACTTAGAGGCGTTCAGTCATAATCC
28S_mat	80	CGGAGAGGGGCTTCCCGGTTACCCGGGGCCTATCGGAGA
28S_mat	80	GCGCTCGTTCTAGCCGCGCCCCGGTTCCTGTCACGAATG
28S_mat	73	AGGTCGTCTACGAGTCATTTAAGCACCGGGTTCACCAA

28S_mat	71	CGAGGGCTGACTTTCAATGGATCGCAGCAACAGAGCTGCT
28S_som	79	CTGGGCTCTTCCCTCTTCGCTCGCCGCTACTGAGGGAAT
28S_som	75	CGCGCCGAGAGAGAGCGATCTCCGTACGCCACATTTCCC
28S_som	80	CCGACCAGCCCCGCCGGTTGAATCCTCCGGGCGGACCG
28S_som	81	CTCCCCGAGAGGGGGGGGATCCGAGCGCACCGGACGGG
28S_som	79	GTCGCGGCGCACCGCCGACGGAGGAAGTGCGCCCGGCGG
28S_som	80	CCGGAGCCGCGAGCCACCTTCGCCCCGAACCTTCCAA
28S_som	79	CCCCCCGAGAGAGGGGGCCGGAGAGAAGGCGCGGAGGAC
28S_som	70	GAAACGGACTGAGGACAGTCCGTTCCGGTCGACGCACGCG
28S_som	80	ATCACCGCGGACCTCGGGCGCCTTTTGGACGTGGGCCTG
28S_som	82	CACCTGGGCCGGCGCGCGCCGGCCTTACCTTATTGCG
28S_som	76	TTCACCATCTTTCGGGTACCATCGCACGCGCTCTAGCTC
28S_som	73	TCCCGCAGCGCCAGTTCTGCTTACCAAAAAGGCCCACTG
28S_som	82	CGGAGGGGGGGGCTCCTTCTGCCGTGACGGCCGGGTATG
28S_som	79	CCGGGGGCGACGGAGGCCATCGCTCCTCCCTCCGAACG
28S_som	80	CCCGTCTCCGCACTCCGGGTTCCGGGGATCTGAACCCGA
28S_som	85	CGGCGCGGGCCAGCCCCGTTTCGCACCTCAGCCCGACCG
28S_som	75	AGGGGCCGCACGACGCGCTCCGGAGACGGGAACGAGAGGG
28S_som	79	CCTCCCCACCCGAAGGCGGGGGCGACCGGACGGACGCG
28S_som	80	CTACCCGCAACACCCACCCCGAGCCCACGAGCCGCCCGG
28S_som	85	CCCCGCAGGCCCCGACCTCGGCGCGCCGCGGCGGACAC
28S_som	79	AGCTGGGGAGATCCGCGAGAAGGGCCCGGCGCACGTCCAG
28S_som	80	CCGCCCCTGCCCTCCCCCGAGAGGAGAAGACGCCAGACG
28S_som	81	CAGTTCTAAGCCAGCTGCTAGGCGCCGGCCGAGGCGCCG
28S_som	80	CCGAGGCGCTTGACGCCAGAAGCAGGAGCCCGCCGGGGG
28S_som	81	CAAACCTCCCACCTGCCACTGTCCCCGGGGCAGGTCGCG
28S_som	77	CACAGATGGTAGCTTCGCACCAGTGGCTCCTCAGCCAAG
28S_som	78	GACCGAGGGCGAACCCCGGCCGGCTATCCAGGCCAATCG
28S_som	80	TGCGTACCGGAAGAGGGGCGGCGCTCGTTCCGCCGCGCC
28S_som	70	GCACGAAACCCTGACCCAGAATCAGGTCGTCTACGAGTCA
28S_som	71	GGATCGAGGGCTGACTTTCAATGGATCGCAGCGATGGAGC
18S_mat_som	67	GTGCACTTAGACYTGCATGGCTTAATCTTTGAGACAAGCA
18S_mat_som	71	ACTGATTTAATGAGCCATTCGCAGTTTCACTGTACCGGCC
18S_mat_som	70	GGCATGTATTAGCTCTGGAATTGCCACAGTTATCCAAGT
18S_mat	83	CCGCCCCGAGGGGAGAAGGGCGGCCAGGTCGGCGCTCG
18S_mat_som	70	GCACCCGWATGGGTTTTGGATCTGATAAATGCACGCGTC
18S_mat	82	AGCGAGGCCGGGGGCGGACGGCGGCGGGTAGGCGCGAGG
18S_mat_som	75	AGGGCGCGCGATCGGCCCGAGGTTATCTAGAGTCACCAA
18S_mat_som	70	AGTTGATAGGGCAGACATTCGAATGAAYCGTCGCCGCCGC

18S_mat_som	80	ACCCTGATTCCCCGYACCCGTGGTCACCATGGTAGGCG
18S_mat_som	78	GGGTAATTTGCGCGCCTGCTGCCTTCCTGGATGTGGTAG
18S_mat_som	69	AGGGCCTCGAAAGAGACCTGYATTGTTATTTTTTCGTCCT
18S_mat_som	80	CTGGCACCAGACTTGCCCTCCAATGGGTCCTCGCCCATG
18S_mat	69	ACGTTAGCATACGCTATTGGAGCTGGAATTACCGCGGCT
18S_mat	80	CTCGCGGCGGACCGCGCGGCCCGGTCCCCGAGATCCAAC
18S_mat	81	AGGACACCCAGTCAAGGGCATCCGGGGGGCGCCGGGAGG
18S_mat_som	64	GAACACTCTAATTTTTTCAAAGTAAACGCTCCGGGCCCC
18S_mat	73	ATAGAACCAGGGTCTATTCCATTATTCCTAGCTGGGGTA
18S_mat_som	87	ATGCCCCCGGCCGTCCCTCTCRATCATGGCCCCGGGTTCC
18S_mat_som	76	GTCCTCTTTCGCGCCGGTCCAAGAATTTACCTCTAGCGG
18S_mat_som	65	GATCGTCTTCGAACCTCCGACTTTCGTTCTTGATTAATGA
18S_mat_som	75	CGCCGCCGGATCGCGGGTCGGCATCGTTTACGGTCGGAA
18S_mat	80	ATACTCCCCCGGAGCCAGAGACTCGTGGTTTTCCCGCAA
18S_mat_som	80	GGCTCCACTCCTGGTGGTGCCCTTCCGTCAATTCCTTAA
18S_mat_som	77	TCAATCTGTCAATCCTTTCCGTGTCCGGGCCGGGTGAGG
18S_mat_som	69	ACGGCCATGCACCACCACCACAGAATCGAGAAAGAGC
18S_mat_som	67	GTCTCGTTCGTTATCGGAATWAACCAGACAAATCGCTCCA
18S_mat_som	76	GACGCCGACCGCGCGGGGCCGTAATAKTTAGCATGC
18S_mat_som	73	TCTCGCGTGGCTGAACGCCACTTGTCCCTCTAAGAAGTT
18S_mat_som	79	ATCCGCCATTGTGGCGCGCGTGCAGCCCCGGACATCTAA
18S_mat_som	80	CGGGGTTCAACGGGTTACCCGCGCCTWTCGGCGCAGGGT
18S_mat	71	TGATGACCTGCGCTTACTGGGAATTCCTCGTTCTCGGGAA
18S_mat_som	74	TCGGTAGTAGCGACGGGCGGTGTGTACAAAGGGCAGGGAC
18S_mat	80	TAAGGGAGCCCCGGGCGGGGCCGATCCGAGGACCTCACT
18S_mat	71	ACTTCTCTAAACGACCGAGTTCGATCGTCTTCTCGGCCG
18S_mat_som	70	GATCCTTCCGAGGTTACCTACGGAAACCTTGTTACGA
18S_som	83	CCGGGGACGAGGCCCGACCCCCGAGGGGGAGCCACG
18S_som	65	ACTTTAATATACGCTATTGGAGCTGGAATTACCGCGGCT
18S_som	80	CCTCGCGGCGGACCGCCAGCCCACTCCCGAGATCCAAC
18S_som	80	TACCGGACACCCAGCTAAGGGCATCCGGGGGGCGCCGGG
18S_som	71	ATAGAACCAGGATCCTATTCCATTATTCCTAGCTGCGGT
18S_som	72	CAACCATACTCCCCCGGAACCCAAAGACTCGTGGTTTCC
18S_som	69	ACTGGGAATTCCTCGTTGATGGGAAACAGTTTCAAGCCCC
18S_som	79	GCGAGGAGCCCCGGCGGGGCCGATCCGAGGACCTACTAA
18S_som	71	TACTTCTCTAGAGGATCAAGTTTGATCGTCTTCTCGGC
5.8S_mat_som	74	CACATTAGTTCTCGCAGCTAGCTGCGTTCCTCATCGACGC
5.8S_mat	78	TGGCCCCGGGATGGACCCGGGGCCGCAATGTGCGTTCGAA
5.8S_som	78	CGTGGCCCCGGGAGGAACCCGGGGCCGCAAAGTGCGTTCC

5S_mat	75	CAGGCCCGACCCTGCTTTACTTCCGAGTTCAGATGAGATC
5S_mat_som	78	CTTACAGCACCTGGTATTCCCAGGCGGTCTCCCATCCAA
5S_som	76	CAGGCCCGACCCTGCTTAGCTTCCGAGATCGGACGAGATC
16S	64	GTTTTAAGTTGATCTTGATGGTTAAGCTAGCTATTTAGC
16S	66	GTCTCCCATACTAGGATAATATAATGGTTTAGTTTGTGGT
16S	76	GGGTTATTTCAATTTCTCTTTCAGCGTTCCTTGCGGTA
16S	64	GCCTAGTGGTGTGGCTAAATCATGATGCAAAGGTACAA
16S	75	GGCCCTAAGTTAAATAGGCTGTCTCGGGGTAGCTCACCTG
16S	79	GGGTTTCGATAGGTCTGTCACTTCTACCCGGAGCTCTTCC
16S	70	GCTGAACTTTTATTCATCTCCAGACAACCAGCTATTGCC
16S	55	ACTAACTTTCGTATTTTTCTCTTAATTTTGTGTTAAATTA
16S	70	AAGGTTGTATCCTTTGTTAAAGGGGCTGTACCCTCTTTA
16S	75	TTTTTAGGTGGCTGCTTCCAGGCCCACTACGACAACGTGT
16S	63	GTCAGAATTATGAATTTCTTCTTACCTGAGCTTTAACGCT
16S	66	TCTCCATATTTGTATGGATTAGTCTGGTATTTTTAGGGA
16S	72	GTCCAACCTGGCTTACACTTGTGCTTGGAGAAGTACAGGT
16S	71	GTTGTGGATTTTTCTCAGTTTTTAACTCACATTGCCCTCT
16S	76	GTTTGCCGAGTTCCTTCCCTTTCTTTTAGTCTTTCCTTAT
16S	77	ACTGGGCAGGCTGGACCTCCTATATTGGGATGTCAGGAGG
16S	70	AGTGATTGCGCTACCTTGCACGGTCAAATACCGCGGCC
16S	72	TCACTGACTTGATGGGGGAGACAGTTAAGCCCTCGTTTTGG
16S	70	AGCTCAAAGGGTCTTCTCGTCTTGTATTATTATATCCGC
16S	63	GGTTTATCTTTTACTAGTTTAATCACTTTTATTGGTTTGC
16S	80	GGTTTTAGGGCTGTGCCCGTCCGCTTGGAGGTTGTTTTT
16S	75	GGTTCGTTGATCGGCTTATAGGCCGGATCGTCTTTGGTCA
16S	70	AGGTCGTAAACCCCTCGTCGATATGGACTCTTAGAGAGG
16S	71	GAACCCTTAATAGCGGCTGCACCATTAGGATGCCTGATC
16S	70	GACCTGGATTACTCCGGTCTGAACTCAGATCACGTAGGAC
16S	76	TGCTTTTAGTATGGGCCCTTTTTCCGATCCTTTCGTAC
16S	67	TCGGGAAGAGGTTTTAGCTCTCCTTTATTTTTTTATCT
12S	65	AGGGGTGCGGAGACTTGCATGTGTAATTAGGTAAGGCT
12S	70	TTGGGCTAAAAATTTGTGCCTGATACCTGCTCCTCGTCC
12S	70	GTTTTCACTTATAGCTTAATGTTTGTCACTGCTGGATCCC
12S	74	CTAGAGCCTCTCGTTAACC
12S	71	AGGGCCGTTTGGCTTTATTTTACATTGTTCTTAACCACC
12S	71	AGCTACTTTCGTGTTTGGGCTCTAGCATCTAAAAGCGT
12S	72	GGCTAAGCATAGTGGGGTATCTAATCCCAGTTTGTCTT
12S	74	TCCTTTGGGTTTTAAGCTTATGCTTGTAGTGTCCGGGCGA
12S	71	AACGGGGATAATCGGTTCTAGAACAGGCTCCTCTAGGGGG

12S	71	TGTTATTCCTTCACAGGGTAAGCTGGCGACGGTGGTATAT
12S	76	TAGCCCATTTCTTCCCATTTTCATGTGCTACACCTCGACCT
12S	66	ATCCTCCTTCAAGTATTGTTTCATGTTACATTTTCGTAT
12S	72	TGTGTACGCGCCTTAGAGCCGGGTTCAAAGAGACACTCT
12S	72	TTTCCCCCTTGTCTGTGCTTCTGTATTAATTATTTCTT
12S	70	GATAATGTCCAAGTGCACCTCCGGTACACTTACCTTGT

Appendix D : rRNA Depletion Protocol

Oligo design (once per target RNA)

1. Curate the target RNA sequences using Web databases such as Genbank, the UCSC Genome Browser, or Ensembl. For eukaryotic rRNA depletion, you would typically include the three nuclear rDNA sequences (e.g., 28S, 18S and 5.8S) and the two mitochondrial rDNA sequences (16S and 12S).
2. Design antisense oligos using the Oligo-ASST Web tool, <https://mtleelab.pitt.edu/oligo>. Input your sequences one at a time or all at once in a single FASTA file and click the “Calculate” button to use the default settings (39-40 nt oligos spaced ≤ 30 nts apart, with melting temperature as close to 70-80°C as possible). Download the oligo sequences to your computer using the “Export” button and open the file using Microsoft Excel or similar program.
3. Order the oligos at the desired scale and standard desalting – 25 nmol scale per oligo should be sufficient for >1000 nuclear rRNA depletion reactions from 1 μ g of total RNA. Some vendors offer bulk/value pricing when ordering above a threshold number of oligos. (Warning: if you downloaded the Details file from Oligo-ASST, be sure to order the antisense oligos and *NOT* the target sequences).

All remaining steps should be done in conditions that minimize contamination with unwanted RNases. Wipe all surfaces and tools with RNase-Away and use filter tips and clean consumables.

Constructing oligo pools

4. Resuspend each dry oligo tube to 1000 μ M with nuclease-free water.

5. Create a 10x working stock of oligos by combining 1 μ L of each oligo into a tube, then diluting with nuclease-free water to the desired concentration per oligo. For eukaryotes, a nuclear rRNA 10x stock should be 4 μ M per individual oligo, to be about 10-fold in excess of the target RNA species in 1 μ g of total RNA in a 10 μ L hybridization reaction. A 10x stock at 1 μ M per oligo may be sufficient for a mitochondrial rRNA pool, depending on taxon and cellular context.

Oligo hybridization to RNA

6. Combine 1 μ L of each 10x oligo pool (e.g., nuclear rRNA pool + mitochondrial rRNA pool) with 1 μ g of total RNA in a PCR tube.
7. Add 2 μ L of 5x hybridization buffer and bring the reaction to 10 μ L with nuclease-free water.
8. Mix well by pipetting and place the capped tube in the thermocycler with the following hybridization program:
 - a. 95°C for 2 minutes (this facilitates RNA denaturation)
 - b. Cool the reaction to 22°C at a rate of 0.1°C/s and hold at 22°C for 5 minutes
9. Place the reaction on ice. RNaseH-mediated depletion
10. Add 10U of thermostable RNaseH and 2 μ L of 10x RNaseH buffer to the reaction. Bring the reaction volume to 20 μ L with nuclease-free water.
11. Mix well by pipetting and heat the depletion reaction at 65°C for 5 minutes or 45°C for 30 minutes. Reaction temperature should be near or above the predicted melting temperatures of the oligos.
12. Place the reaction on ice.

13. Add 5U of TURBO DNase and 5 μ L of 10x DNase buffer. Bring the reaction volume to 50 μ L with nuclease-free water.
14. Mix well by pipetting and heat the reaction at 37°C for 30 minutes.
15. Place the reaction on ice.

Clean-up and size selection

Follow the manufacturer's protocol for the Zymo Clean and Concentrator-5 kit to enrich for RNA molecules >200 nts:

16. Mix 50 μ L of Zymo RNA Binding Buffer and 50 μ L of 100% ethanol to create the Adjusted Binding Buffer.
17. Add 100 μ L (2 volumes) of Adjusted Binding Buffer to depleted sample, which should be in a volume of 50 μ L after Step 15. Mix well by pipetting.
18. Transfer the sample to a Zymo spin column with collection tube. Centrifuge for 30 seconds at 15,000 x g.
19. Discard the flow through, which will contain RNAs < 200 nts
20. Add 400 μ L of RNA Prep Buffer to the column. Centrifuge for 30 seconds at 15,000 x g. Discard flow through.
21. Add 700 μ L of RNA Wash Buffer to the column. Centrifuge for 30 seconds at 15,000 x g. Discard flow through.
22. Add 400 μ L of RNA Wash Buffer to the column. Centrifuge for 1 minute at 15,000 x g. Discard flow through.
23. Place the column into a nuclease-free 1.5 mL microcentrifuge tube for elution.
24. Add 6 μ L of nuclease-free water directly to the column matrix and centrifuge for 30 seconds at 15,000 x g.

25. Proceed straight to downstream application (e.g., RNA-seq library building) or store purified RNA at -80°C until needed.

Appendix E : *X. laevis* embryo CUT&RUN protocol

Protocol is adapted from Hainer et. al. 2019 (162) and Skene and Henikoff 2017 (161)

Buffers

1. Newport 2.0 (163) (100mM sodium isethionate, 20mM sodium pyrophosphate, 10mM CAPS buffer) pH to 10.5 with NaOH
2. Nuclear Extraction (20mM HEPES-KOH, pH 7.9, 10mM KCl, 0.5mM Spermidine, 0.1% Triton X-100, 20% Glycerol) 1 tablet of Pierce Protease Inhibitor Tablets / 50mL
3. Binding (20mM HEPES-KOH, pH 7.9, 10mM KCl, 1mM CaCl₂, 1mM MnCl₂)
4. Wash (20mM HEPES-KOH, pH 7.5, 150mM NaCl, 0.5mM Spermidine, 0.1% BSA) 1 tablet of Pierce Protease Inhibitor Tablets / 50mL
5. Blocking (20mM HEPES-KOH, pH 7.5, 150mM NaCl, 0.5mM Spermidine, 0.1% BSA, 2mM EDTA) 1 tablet of Pierce Protease Inhibitor Tablets / 50mL
6. 2XSTOP+ (200mM NaCl, 20mM EDTA, 4mM EGTA, 50µg/mL RNase A, 40µg/mL glycogen) Add spike-in DNA as desired.

Dejelly Embryos

1. After fertilization (~20 minutes after sperm addition or upon observation of contraction), withdraw supernatant with a clipped pipet and cover the embryos with 0.3% BME in MR/3 (pH 8.5). Gently shake the dish for 2 minutes. Check the dish under a dissecting microscopy to see the jelly has completely dissolved.
2. Transfer the embryos to a new, dry MR/3 agarose dish. Take care to transfer as little of the BME as possible.

3. To neutralize any residual BME, add MR/3 (pH 6.5) to the embryos and gently shake the dish for 2 minutes.
4. Transfer the embryos to a new, dry MR/3 agarose dish. Take care to transfer as little of the MR/3 as possible.
5. Wash the embryos with MR/3 twice (2x) by adding MR/3, gently shaking the dish, and then removing the MR/3.
6. Add fresh MR/3 and incubate dejellied embryos to desired stage.

Devitellinize Embryos

7. Devitellinize embryos with pronase (1mg/mL). To maximize survival, you can devitellinize just before the desired stage, but accurate staging can be more difficult with this approach.
8. Collect dejellied embryos with a green glass pipette. Place the pipette tip into an MR/3 agarose dish with 1-3 mL of pronase dissolved in MR/3. Ensure the embryos sink to the bottom of the pipette and gently release into the dish, allowing for minimal dilution of the pronase.
9. Periodically shake the dish by hand to encourage the membrane to release. Continue to pronase until the vitelline envelope is visibly removed from the embryos. Gently pipetting individual embryos up and down with a flame-tipped glass pipet can help manually remove difficult ones. Be careful to not over pronase, as this will cause cells to lysis early and result in reduced nuclear recovery.

Nuclear Extraction

10. Use a flame-tipped glass pipet to transfer devitellinized embryos to a 1.5mL Eppendorf tube containing 1mL of Newport 2.0 buffer. Allow the embryos sink to the bottom of

- the pipette and gently release into the tube. Periodically, pipet the buffer up and down with a P1000 to facilitate cell dissociation. Be as gentle as possible as to not lyse the cells!
11. Allow the dissociated cells to sink to the bottom of the tube. Withdraw as much of the Newport 2.0 buffer as possible without disturbing the cells.
 12. Add 1mL of ice-cold NE buffer. Gently pipet up and down with a P1000 to lyse the cells.
 13. Spin down at 600g 3' 4°C and remove supernatant with a P1000. Do not disturb the pellet at the bottom.
 14. Resuspend in 600 μ L NE buffer.

Bind to Magnetic Beads

15. Gently resuspend Bio-Mag Plus Concanavalin A coated beads (Polysciences, Inc. #86057).
16. Withdraw 150 μ L bead slurry, and transfer to 850 μ L Binding buffer in a 1.7 ml Eppendorf tube.
17. Place on a magnet stand and wash twice in 1 ml Binding buffer.
18. Resuspend in 300 μ L Binding buffer.
19. While gently vortexing the nuclei (~1500 rpm on the mini vortexer), slowly add the bead slurry.
20. Rotate 10 min at room temperature.

Block

21. Place on the magnet stand, allow to clear (~2 min) and pull off the liquid.
22. Add 1 ml Blocking buffer and mix with gentle pipetting.

23. Incubate 5' at RT.

Bind Primary Antibody

24. Place on the magnet stand and pull off the liquid.

25. (Repeat 2x) Add 1 ml Wash buffer, mix with gentle pipetting. Place on the magnet stand, wait for solution to clear, and pull off the liquid.

26. Resuspend in 250 μ L Wash buffer with gentle pipetting.

27. While gently vortexing (~1500 rpm) add 250 μ L primary antibody in Wash buffer (typically 1:100, final. So 5 μ L primary antibody).

28. Incubate on rotator ~2 hr at 4°C.

29. (Repeat 2x) Quick spin in microfuge (1 sec) and wash in 1 ml Wash buffer. For wash steps: remove from magnet stand and pipette up and down ~ten times; beads should come off the side)

Bind pAG-MNase

30. Magnetize, then pull off the liquid and resuspend each sample in 250 μ L Wash buffer.

31. Add 5 μ L pAG-MNase enzyme into 245 μ L Wash buffer, then add to tube while gently vortexing (~1500 rpm) for a final pAG-MNase concentration of 1:100 (WARNING: each new batch of pA/pAG-MNase must be titrated to determine working concentration).

32. Incubate 1 hr on rotator at 4°C.

33. Quick spin on microfuge (1 sec) and wash twice in 1 ml Wash buffer (same as with primary)

Digestion

34. Pull off the liquid and resuspend in 150 μ L Wash buffer.

35. Equilibrate to 0°C in wet ice (~5 minutes) (it is imperative the digestion is performed at 0°C – you can use an aluminum block to maintain the temperature).
36. Remove a tube from 0°C, add 3 µL 100 mM CaCl₂ per 150 µL while vortexing (1500 rpm), flick tube 2-3x, then return to 0°C. Mixing well is essential.
37. Stop each time point at the designated time (15-30 min) with 150 µL 2XSTOP+. Pipette to mix and place on ice. Usually, 30 min for *X. laevis*.

Fractionation

38. Incubate 20 min 37°C to RNase and release CUT&RUN fragments from the insoluble nuclear chromatin.
39. Spin 5' 16,000 x g 4°C (with hinge pointed outward)
40. Carefully pull off the liquid and add to new tube – use a P200 and stick your pipette on the opposite side of the pellet

Extraction

41. To each sample add 3 µL 10% SDS (to 0.1%), and 2.5 µL Proteinase K (20 mg/ml).
42. Mix by inversion and incubate 10 min 70°C (Shift the 65° block).
43. If you crosslinked: Incubate overnight at 55°C (use the mini incubator)
44. Add 300 µL buffered phenol-chloroform-isoamyl solution (25:24:1) and vortex.
45. Transfer to a phase-lock tube, and spin 5 min full speed.
46. Add 300 µL chloroform and invert ~10x to mix and spin 2 min full speed.
47. If you crosslinked: Remove aqueous phase to a new tube with 150ul AMPure XP beads, let sit for 15 minutes. Magnetize 5 min.
48. Remove aqueous phase to a fresh tube.

49. Add 5 μL of 20 mg/ml glycogen and 1 mL 100% ethanol and mix by vortexing or tube inversion.
50. To increase yield, chill at -20°C for 2 hours
51. Chill on ice then spin 10-30 min full speed 4°C .
52. Wash the pellet (hardly visible) in 1 ml 80% ethanol, spin briefly full speed 4°C .
53. Discard S/N. Air dry.
54. When the pellet is dry, dissolve in 36.5 μL 0.1xTE. Can store at -20°C .

Appendix F : *X. laevis* embryo ATAC-seq Transposition Protocol

Protocol is adapted from Esmaeili et. al. 2020 (175).

Animal Cap Dissection

1. Place devitellinized embryos in a 1% agarose dish filled with 0.7x MMR
2. Use fine watch-makers forceps to make cuts just above the marginal zone of the embryo(s).
3. Cut around the perimeter of each embryo and remove the animal cap.
4. Places caps in a fresh dish filled with 0.7x MMR to wash them.

Lysing

5. Transfer caps (2/rxn) to a 1.5mL tube containing 1mL of ice-cold PBS
6. Centrifuge at 500xg for 5 min at 4°C.
7. Draw off as much PBS as possible without exposing the caps to air.
8. Add 1mL of ice-cold PBS and centrifuge again.
9. Remove the PBS completely
10. Resuspend the caps in 50 μ L of lysis buffer (10mM Tris pH 7.4, 10mM NaCl, 3mM MgCl₂, 0.1% IGEPAL CA-630).
11. Pipet up and down with a clipped P200 at 4°C.
12. Centrifuge at 500xg for 10min at 4°C.

Transposition

13. Remove lysis buffer and resuspend the pellet in 47.5 μ L of TD buffer (10mM Tris pH 7.6, 5mM MgCl₂, 10% dimethylformamide).
14. Add 2.5 μ L of 3 μ M Tn5 transposome and mix.

15. Incubate for 1hr at 37°C with pipet mixing every 15 min.
16. Add 2.5µL of proteinase K and incubate at 37°C overnight.
17. MinElute transposed DNA into 20µL of water.

PCR and Size Selection

18. Add 2.5µL of desired i7 and 2.5µL of desired i5 primer to the DNA.
19. Add 25µL of Phusion Flash 2X Master Mix (Fischer Product Code: 10043967) and mix well
20. Amply the library according to the following:
 - a. 72°C – 5 min
 - b. 98°C – 30s
 - c. 98°C – 10s
 - d. 63°C – 30s
 - e. 72°C – 1min
 - i. Repeat c – e for a total of 12 cycles
 - f. 72°C – 1min
 - g. Hold at 4°C
21. MinElute the library into 10µL of water.
22. Run the library on a 2% TBE gel as long as possible to allow for clear separation.
23. Excise the fragments between 150 – 250 bp (these are the accessible fragments).
24. Excise the fragments between 250 – 600 bp (these are the nucleosomal fragments.)
25. Gel extract (NEB: T1020S) each library separately and elute in 6µL of water to acquire as high of a concentration as possible. Store at -20°C.

Appendix G : *X. borealis* and *X. muelleri* Husbandry

Appendix G.1 : General Guidelines

Any personnel handling the frogs should wash their hands *without soap* prior to contact, as soap has the potential to irritate the animals' skin. Frogs should be housed in tanks that allow a minimum of 2L of water per individual with males and females in separate tanks. Water should have a pH of 7.4 ± 0.1 , a salinity of $1200 \mu\text{S} \pm 100$, and be 22°C . The pH can be adjusted with dilute (1:60 dilution) hydrochloric acid and solid sodium bicarbonate. Salinity can be adjusted using a saturated sodium chloride solution or the addition of de-ionized water. Tanks should be checked routinely (twice a week, minimally) for signs of aggression (*i.e.* claw marks, broken claws, bite marks) and signs of stress (*i.e.* skin sloughing and sluggish/shy behavior). In the case of observed aggression, remove the aggressive individual (the one with little to no injuries) and place it in a tank with individuals of a similar size; swap in another individual from the new tank into the old tank ensuring size balancing as well. To reduce stress, the tanks should provide some sort of enrichment tube large enough to allow all individuals to hide in and around the tube without provoking aggressive behavior. The most common cause of stress is chemical stress and may be an indication of improper water conditions or the need for more frequently cleaning.

Appendix G.2 : Feeding

Frogs should be fed between two to three times a week and should be fed two hours prior to any manual cleaning, minimally. For standing water tanks, it is advisable to clean the tanks on the same day as feeding to efficiently remove waste products and maintain minimal ammonium levels. Frogs should be fed “floating/sinking” food (FFF, Xenopus Express) at a density of 10 pellets per frog. After two hours, excess food and visible waste should be removed from the tank if the tank is not being cleaned at that time.

Appendix G.3 : Tank Cleaning

Tanks can be cleaned either by a complete water exchange (draining and replacing dirty water) or by a full tank exchange (remove frogs into a fresh tank with clean water). For a water exchange, the tank should be drained as low as possible and then flushed with fresh system water while still draining. Enrichment tubes should be removed and rinsed with de-ionized water before replacing into the tank. The frogs can stay within the tank during this process so long as there is no risk to them being trapped in the drain (*i.e.* the drain is small enough to not allow entry). For a tank exchange, a fresh tank should be sanitized beforehand using a sanitizer that will not irritate the frogs’ skin (*i.e.* ReptiSan) and rinsed with de-ionized water. The sanitized tank can then be filled with fresh system water according to the above parameters and rinsed enrichment tube placed within the tank. Frogs should be moved individually by net and placed into the fresh tank. The net should be large enough to completely contain the frog and allow the technician to pinch off the opening of the net with their free hand; thereby completely encasing the frog in the net.

In the event a frog is dropped, *do not try to catch the frog in mid-air*. It is far more likely that the frog will be injured trying to catch them in mid-air rather than impacting the floor. Instead, cease all movement until the frog is on the floor. *Xenopus* are easily able to see movement. Therefore, any fast and/or sudden movements are likely to drive the loose frog away from the source of those movements (*i.e.* the pursuing technician) leading to additional animal stress and the potential of losing the frog. However, *Xenopus* are entirely aquatic species and struggle to effectively move outside of water. Therefore, slowly, calmly, and gently catch the loose frog with an available net or gloved hands. *Do not try to block the frog with feet or large, heavy items*. Once recaptured, replace the frog into the dirty tank to allow cleaning of debris as well as acclimation to a familiar environment. Once the panic has subsided (fast, jerky movements in search of an avenue of escape and/or safety), transfer the frog to the clean tank.

Appendix G.4 : hCG Preparation

Human chorionic gonadotropin (hCG; Merck Animal Health) is prepared by dissolving 10,000 units in 2mL of sterile water (provided with freeze-dried hCG). This stock should be aliquoted into 8-250 μ L shots at 5U/ μ L. These are *X. borealis* and *X. muelleri* boosting shots. Further diluting 2 of these shots with sterile water to 650 μ L each will yield 1.3mL of 2U/ μ L and can be aliquoted into 5-250 μ L shots. These are *X. tropicalis* boosting shots. These boosting shots are at a higher concentration than priming shots and are intended to induce ovulation in the female while encouraging latching behavior in the male, in the case of amplexus.

To prepare priming shots, a single *X. borealis/X. muelleri* boosting shot should be split and diluted to 1250 μ L each. This should yield 10-250 μ L shots of 0.5U/ μ L and are *X. borealis/X.*

muelleri priming shots. Further diluting 2 of these priming shots to 1250 μ L each will yield 10-250 μ L shots at 0.1U/ μ L. These are *X. tropicalis* priming shots. Priming shots are 1/10th to 1/20th the concentration of boosting shots and are intended to prepare the animals for mating. Priming is especially important for couple in way one or both partners have long (4 – 6 months) stretching between mating sessions or have never been mated before. If the frogs are kept on a regular mating schedule, it is possible to forego priming as they should already be acclimated to the sudden hormonal spike from the boosting shot. All shots should be stored at -20°C.

Appendix G.5 : Amplexus

An individual male and female should be selected for amplexus based on past breeding record. It is advisable to allow at least 4 weeks of recovery between amplexus sessions. A 27G hypodermic needle should be loaded with 200 μ L of 0.5U/ μ L human chorionic gonadotropin (priming shot) to prepare *both* individuals, as necessary (see previous section). After netting an individual, place the frog on the benchtop in hopping position and gently hold down the frog within the net with a free hand. The positioning should be such that the head is entirely contained within the net, but the hind leg is exposed from the net, preferably the hind leg that corresponds to the technician's dominant hand. While in hopping position, the dorsal lymph sac is pressed up against the skin behind the hind leg, underneath the lateral lines, and does not require deep insertion to inject into. Therefore, insert the needle through the lateral lines, following the path of the lines, behind the hind leg (Figure 4-2A). After insertion, inject 100 μ L of priming dose hCG to prime the frog for amplexus. The male and female should be separated and placed in individual containment with at least 2L of water for each. In the late afternoon on the day prior to embryo collection, the

female frog should be boosted with 100 μ L of 5U/ μ L hCG via injection into the dorsal lymph sac as described above. The male should be similarly injected with only 50 μ L of boosting dose hCG. After boosting both individuals, place both frogs together in containment with 2L of system water and place the containment overnight in isolation with dim light where they won't be disturbed. The following morning, the male should be latched behind the female and hundreds to a thousand embryos should be laid in the container. If the couple is still latched, prepare a clean container with 2L of water and gently transfer the couple, preferably without detaching the male, to the new container and collect the embryos. In our lab's hands, couples can latch and lay eggs as soon as 2 hours after the boosting injection, leading to embryos as old as NF stage 12 the next day. This suggests that couples could be boosting during the morning to obtain embryos by noon on the same day, however this has not been achieved successfully yet. Note: if performing IVF, the female will be housed overnight alone after hCG injection.

Appendix G.6 : Testis Dissection

While amplexus is technically simpler and more likely to yield viable embryos, the primary cost is the vast differences in stages due to multiple fertilization times while latched. Alternatively, *Xenopus* eggs can be artificially inseminated to produce a clutch of embryos with a synchronized fertilization point. This requires the initial euthanasia of the male and dissecting out the testes. A sexually mature male is placed in 1L of 3.6g/L MS-222 for at least 30 minutes. Then the male is removed from the MS-222 and placed on his back with his legs outstretched on sterile, absorbent benchtop surface. To ensure the male has been anesthetized, pinch one of his toes on each foot with a pair of fine precision medium tipped tweezers (Fisher: 12-000-157). If this provokes any

movement, place the male back in the MS-222 solution for another 30 minutes. If there is no observable movement, turn the male onto his stomach and sever his spinal cord with a pair of 6.5” sharp/sharp straight blade operating scissors (Fisher: 13-810-4).

After sacrificing the male, return him to his back and puncture the skin with a 16G 1” hypodermic needle (Fisher: 14-826-18A) such that the needle creates a small impalement just above the hip and in line with the shoulder (Figure 4-2B). Place the blade of a 4.5” sharp-pointed dissecting scissor (Fisher: 08-940) in line with the needle to guide the blade under the skin and create an initial incision. This incision should be expanded laterally across the frog to the other hip and anteriorly up to the shoulder (Figure 4-2B). At the shoulder, the incision can be extended laterally again to the other shoulder, creating a flap covering the entirety of the torso (Figure 4-2B). At this point, the muscle should be exposed but not cut in any way. Similar to the skin incision, an initial incision should be created in the muscle with a 16G 1” needle and 4.5” sharp-pointed dissecting scissors (Figure 4-2C). Take care not to penetrate too deeply as this risks excessive bleeding due to organ and/or artery puncture. Extend the incision laterally to the other hip, like with the skin incision (Figure 4-2C). Unlike the skin incision, *do not extend anteriorly from hip to shoulder* as the subsequent lateral extension from shoulder to shoulder is blocked by the sternum. Instead, extend two incisions anteriorly on both sides of the sternum as far as possible (eventually the hard, cartilaginous clavicle area will prevent further dissection) (Figure 4-2C). These two incisions can then be extended laterally to their respective shoulders (Figure 4-2C). Unlike the skin dissection, which will create a “single door” opening covering the whole torso, the muscle dissection should result in a “double door” opening exposing the vital organs on both sides with the sternum covering the center, with minimal observed bleeding. As a warning, *the heart will still be beating after proper euthanasia.*

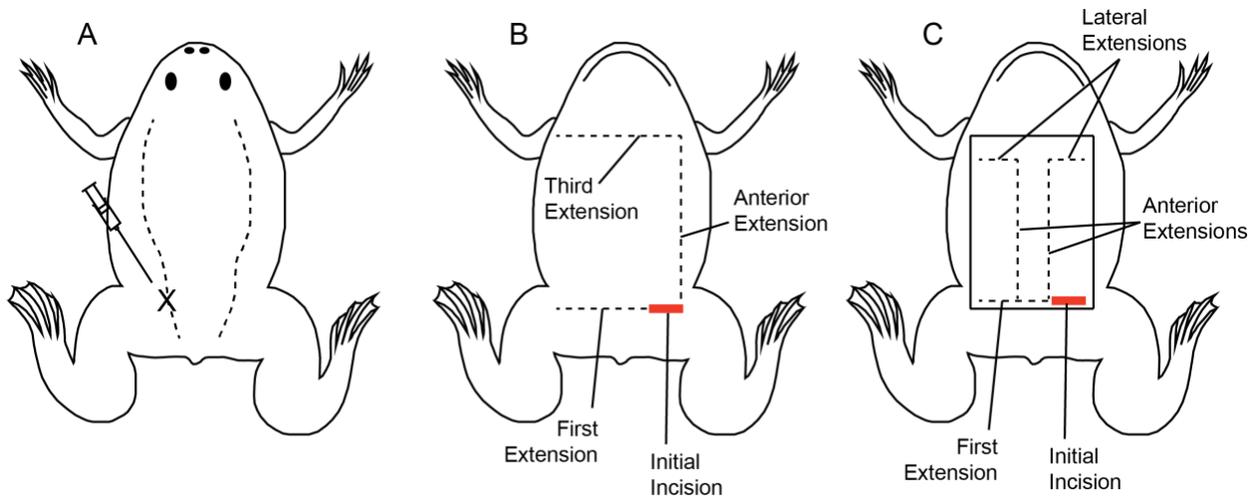


Figure 4-2 : *Xenopus* surgery diagrams

(A) Injection into the dorsal lymph sac behind the hindleg. (B) Skin incisions to flap over the skin and expose the muscle. (C) Muscle incisions to reveal vital organs and testes.

The testes are white, jellybean-like organs with one on each side of the frog. They will be beneath the lungs, the pair of large, dark brown/black organs, and are likely underneath the bright orange/yellow fat lobes near the posterior of the frog. Using the 4.5" sharp-pointed dissecting scissors, each teste should be gently cut away from the fat by severing the fibrous tissue attaching the teste to the fat lobes without puncturing the testes. Dissected testes can be placed in a Kimwipe soaked with 1x MMR. It is advisable to remove as much excess tissue as possible post-dissection using Vannas-Tübingen spring scissors (Fine Science Tools: 15008-08) under a dissection microscope. For longer-term storage, the testes from be placed in L-15 media (Fisher: SH3052501) and can be stored at 4°C for up to 3 days.

Appendix G.7 : IVF and Dejelling

Once testes have been dissected, the females can be squeezed to obtain unfertilized eggs. If the hCG injection was successful, it is likely there are dozens to hundreds of eggs already in her container. Her cloaca should, also, be swollen and deep red in color. After coming into contact with water, the eggs become activated and the jelly coat will begin to thicken. After 10 – 15 minutes, the jelly coat will become impermeable to sperm and, therefore, cannot be fertilized. As such, it is advisable to dissect the testes prior to egg acquisition, but only once it is certain eggs can be acquired. To prepare the sperm slurry, testes should be kept on ice and ~ one-sixth of a testis can be cut away using a razor blade. This testis piece should be placed in 1mL of 1x MMR and shredded with fine forceps and dissecting scissors/razor blade. The slurry can be kept on ice until used.

To extract fertilization competent eggs, the female can be “squeezed” over a petri dish. The female should be grabbed with both hands such that the palms are over her back and she is facing the technician. The pinky fingers should be positioned in front of her forelegs, the middle and ring fingers positioned between the forelegs and hindlegs, and the pointer fingers behind the hindlegs. The pinky fingers should wrap around and be positioned on either side of her head in order to control her movement. Using the pointer fingers of both hands, gently pin her hind legs forward so her cloaca is unobstructed. Once in this position, the female should be lifted out of the water and held such that her cloaca is over an empty petri dish. This grip should be tight enough as to control her movements but not so tight to completely restrict her movements. This is partial due to the risk of injuring the female, but primarily due to the fact that the intermittent resistance to being held is what will release the eggs (*i.e.* “squeeze” them out of the female). As such, *there will never be a need to physically squeeze the frog.*

Once in the dish, the eggs should be covered with MR/3 and 500 μ L of the sperm slurry can be applied. Apply the slurry slowly by targeting and spreading the slurry to as many eggs as possible. Gently shake the dish once applied and note the time; this is the time of fertilization of all subsequent downstream time points. The zygotes will take 90 minutes, on average, to divide to NF stage 2, however, the animal should show contraction 20 – 30 minutes after fertilization, indicating successful fertilization. If the contraction is not obvious, fertilized eggs will roll animal pole up ~30 minutes post fertilization.

Once fertilization is confirmed, the clutch should be dejellied immediately to facilitate downstream manipulation, including injection. To dejelly eggs, the MR/3 should be removed as much as possible (the eggs are not yet sensitive to air) and replaced with a dejellying solution, either 2.5% L-cysteine or 0.3% β -mercaptoethanol. Of the two treatments, the β -mercaptoethanol is the harsher of the two and, thus, more likely to remove the jelly completely in a shorter amount of time. However, the L-cysteine treatment is gentler on the eggs but may require a longer dejelly treatment. *It is worth noting that in our lab's hands, the L-cysteine treatment appears to more efficiently remove the jelly coat for X. muelleri, however this would require further examination.* Once submerged in the dejelly solution, the eggs should be gently swirled on the benchtop for 2 minutes. After 2 minutes, it is advisable to check the jelly coat under a dissecting microscope, however the eggs should be visibly freer as the dejellying proceeds. Once the jelly coat is removed, the eggs should be washed twice with MR/3 and can then be incubated/manipulated as desired. Typical incubation is at 23°C but both frogs can be incubated at 18°C for slower development. Unlike *X. laevis*, it is not advisable to incubate at 16°C as the embryos do not survive well, if at all, at this temperature. MR/3 should be changed daily if not visible buildup of debris is observed;

if a buildup is observed, it is advisable to change the MR/3 immediately or to transfer the embryos to a new dish, depending on the severity.

Bibliography

1. C. H. Waddington, A discussion of some aspects of theoretical biology. *The Strategy of the Genes*, ed. CH Waddington (London: London George Allen and Unwin) (1957).
2. J. B. Gurdon, T. R. Elsdale, M. Fischberg, Sexually mature individuals of *Xenopus laevis* from the transplantation of single somatic nuclei. *Nature*. **182**, 64–5 (1958).
3. M. J. Evans, M. H. Kaufman, Establishment in culture of pluripotential cells from mouse embryos. *Nature*. **292**, 154–156 (1981).
4. G. R. Martin, Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proceedings of the National Academy of Sciences*. **78**, 7634–7638 (1981).
5. H. M. Blau, C.-P. Chiu, C. Webster, Cytoplasmic activation of human nuclear genes in stable heterocaryons. *Cell*. **32**, 1171–1180 (1983).
6. R. L. Davis, H. Weintraub, A. B. Lassar, Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell*. **51**, 987–1000 (1987).
7. H. Kulesa, J. Frampton, T. Graf, GATA-1 reprograms avian myelomonocytic cell lines into eosinophils, thromboplasts, and erythroblasts. *Genes & development*. **9**, 1250–1262 (1995).
8. H. Xie, M. Ye, R. Feng, T. Graf, Stepwise Reprogramming of B Cells into Macrophages. *Cell*. **117**, 663–676 (2004).
9. J. Nichols, A. Smith, Naive and Primed Pluripotent States. *Cell Stem Cell*. **4**, 487–492 (2009).
10. J. Nichols, A. Smith, Pluripotency in the embryo and in culture. *Cold Spring Harbor perspectives in biology*. **4**, a008128 (2012).
11. I. G. M. Brons, L. E. Smithers, M. W. B. Trotter, P. Rugg-Gunn, B. Sun, S. M. Chuva de Sousa Lopes, S. K. Howlett, A. Clarkson, L. Ahrlund-Richter, R. A. Pedersen, L. Vallier, Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature*. **448**, 191–195 (2007).
12. P. J. Tesar, J. G. Chenoweth, F. A. Brook, T. J. Davies, E. P. Evans, D. L. Mack, R. L. Gardner, R. D. G. McKay, New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature*. **448**, 196–199 (2007).
13. Y. Matsui, K. Zsebo, B. L. M. Hogan, Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. *Cell*. **70**, 841–847 (1992).

14. J. L. Resnick, L. S. Bixler, L. Cheng, P. J. Donovan, Long-term proliferation of mouse primordial germ cells in culture. *Nature*. **359**, 550–551 (1992).
15. M. Kanatsu-Shinohara, K. Inoue, J. Lee, M. Yoshimoto, N. Ogonuki, H. Miki, S. Baba, T. Kato, Y. Kazuki, S. Toyokuni, M. Toyoshima, O. Niwa, M. Oshimura, T. Heike, T. Nakahata, F. Ishino, A. Ogura, T. Shinohara, Generation of Pluripotent Stem Cells from Neonatal Mouse Testis. *Cell*. **119**, 1001–1012 (2004).
16. Tokuzawa Yoshimi, Kaiho Eiko, Maruyama Masayoshi, Takahashi Kazutoshi, Mitsui Kaoru, Maeda Mitsuyo, Niwa Hitoshi, Yamanaka Shinya, Fbx15 Is a Novel Target of Oct3/4 but Is Dispensable for Embryonic Stem Cell Self-Renewal and Mouse Development. *Molecular and Cellular Biology*. **23**, 2699–2708 (2003).
17. M. Tomioka, M. Nishimoto, S. Miyagi, T. Katayanagi, N. Fukui, H. Niwa, M. Muramatsu, A. Okuda, Identification of Sox-2 regulatory region which is under the control of Oct-3/4–Sox-2 complex. *Nucleic Acids Research*. **30**, 3202–3213 (2002).
18. S. Okumura-Nakanishi, M. Saito, H. Niwa, F. Ishikawa, Oct-3/4 and Sox2 regulate Oct-3/4 gene in embryonic stem cells. *Journal of Biological Chemistry*. **280**, 5307–5317 (2005).
19. K. Mitsui, Y. Tokuzawa, H. Itoh, K. Segawa, M. Murakami, K. Takahashi, M. Maruyama, M. Maeda, S. Yamanaka, The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell*. **113**, 631–42 (2003).
20. I. Chambers, D. Colby, M. Robertson, J. Nichols, S. Lee, S. Tweedie, A. Smith, Functional Expression Cloning of Nanog, a Pluripotency Sustaining Factor in Embryonic Stem Cells. *Cell*. **113**, 643–655 (2003).
21. H. Niwa, K. Ogawa, D. Shimosato, K. Adachi, A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. *Nature*. **460**, 118–122 (2009).
22. P. Cartwright, C. McLean, A. Sheppard, D. Rivett, K. Jones, S. Dalton, LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism. *Development*. **132**, 885–896 (2005).
23. K. Takahashi, S. Yamanaka, Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*. **126**, 663–76 (2006).
24. K. Takahashi, K. Tanabe, M. Ohnuki, M. Narita, T. Ichisaka, K. Tomoda, S. Yamanaka, Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*. **131**, 861–72 (2007).
25. J. Yu, M. A. Vodyanik, K. Smuga-Otto, J. Antosiewicz-Bourget, J. L. Frane, S. Tian, J. Nie, G. A. Jonsdottir, V. Ruotti, R. Stewart, I. I. Slukvin, J. A. Thomson, Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. *Science*. **318**, 1917–1920 (2007).
26. M. Nakagawa, M. Koyanagi, K. Tanabe, K. Takahashi, T. Ichisaka, T. Aoi, K. Okita, Y. Mochiduki, N. Takizawa, S. Yamanaka, Generation of induced pluripotent stem cells

- without Myc from mouse and human fibroblasts. *Nature Biotechnology*. **26**, 101–106 (2008).
27. J. Jiang, Y.-S. Chan, Y.-H. Loh, J. Cai, G.-Q. Tong, C.-A. Lim, P. Robson, S. Zhong, H.-H. Ng, A core Klf circuitry regulates self-renewal of embryonic stem cells. *Nature Cell Biology*. **10**, 353–360 (2008).
 28. M. Nakagawa, N. Takizawa, M. Narita, T. Ichisaka, S. Yamanaka, Promotion of direct reprogramming by transformation-deficient Myc. *Proceedings of the National Academy of Sciences*. **107**, 14152–14157 (2010).
 29. K. Takahashi, S. Yamanaka, A decade of transcription factor-mediated reprogramming to pluripotency. *Nature Reviews Molecular Cell Biology*. **17**, 183–193 (2016).
 30. A. Trounson, N. D. DeWitt, Pluripotent stem cells progressing to the clinic. *Nature Reviews Molecular Cell Biology*. **17**, 194–200 (2016).
 31. P. Svoboda, Mammalian zygotic genome activation. *Seminars in Cell & Developmental Biology*. **84**, 118–126 (2018).
 32. N. L. Vastenhouw, W. X. Cao, H. D. Lipshitz, The maternal-to-zygotic transition revisited. *Development (Cambridge, England)*. **146** (2019), doi:10.1242/dev.161471.
 33. M. Li, J. C. Belmonte, Ground rules of the pluripotency gene regulatory network. *Nature reviews. Genetics*. **18**, 180–191 (2017).
 34. V. E. Foe, B. M. Alberts, Studies of nuclear and cytoplasmic behaviour during the five mitotic cycles that precede gastrulation in *Drosophila* embryogenesis. *Journal of Cell Science*. **61**, 31–70 (1983).
 35. D. A. Kane, C. B. Kimmel, The zebrafish midblastula transition. *Development (Cambridge, England)*. **119**, 447–56 (1993).
 36. J. Newport, M. Kirschner, A major developmental transition in early *Xenopus* embryos: I. characterization and timing of cellular changes at the midblastula stage. *Cell*. **30**, 675–86 (1982).
 37. M. T. Lee, A. R. Bonneau, C. M. Takacs, A. A. Bazzini, K. R. DiVito, E. S. Fleming, A. J. Giraldez, Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. *Nature*. **503**, 360–4 (2013).
 38. M. Leichsenring, J. Maes, R. Mossner, W. Driever, D. Onichtchouk, Pou5f1 transcription factor controls zygotic gene activation in vertebrates. *Science (New York, N.Y.)*. **341**, 1005–9 (2013).
 39. L. Miao, Y. Tang, A. R. Bonneau, S. H. Chan, M. L. Kojima, M. E. Pownall, C. E. Vejnar, F. Gao, S. Krishnaswamy, C. E. Hendry, A. J. Giraldez, The landscape of pioneer factor

- activity reveals the mechanisms of chromatin reprogramming and genome activation. *Molecular Cell*. **82**, 986-1002.e9 (2022).
40. G. E. Gentsch, T. Spruce, N. D. L. Owens, J. C. Smith, Maternal pluripotency factors initiate extensive chromatin remodelling to predefine first response to inductive signals. *Nature communications*. **10**, 4269 (2019).
 41. P. Scerbo, F. Girardot, C. Vivien, G. V. Markov, G. Luxardi, B. Demeneix, L. Kodjabachian, L. Coen, Ventx Factors Function as Nanog-Like Guardians of Developmental Potential in *Xenopus*. *PLOS ONE*. **7**, e36855 (2012).
 42. M. Schuff, D. Siegel, M. Philipp, K. Bundschu, N. Heymann, C. Donow, W. Knochel, Characterization of *Danio rerio* Nanog and functional comparison to *Xenopus* Vents. *Stem cells and development*. **21**, 1225–38 (2012).
 43. H. L. Liang, C. Y. Nien, H. Y. Liu, M. M. Metzstein, N. Kirov, C. Rushlow, The zinc-finger protein Zelda is a key activator of the early zygotic genome in *Drosophila*. *Nature*. **456**, 400–3 (2008).
 44. M. M. Colonna, J. E. Abrahante, P. Schedl, D. M. Gohl, G. Deshpande, CLAMP regulates zygotic genome activation in *Drosophila* embryos. *Genetics*. **219**, iyab107 (2021).
 45. J. Duan, L. Rieder, M. M. Colonna, A. Huang, M. McKenney, S. Watters, G. Deshpande, W. Jordan, N. Fawzi, E. Larschan, CLAMP and Zelda function together to promote *Drosophila* zygotic genome activation. *eLife*. **10** (2021), doi:10.7554/eLife.69937.
 46. M. M. Gaskill, T. J. Gibson, E. D. Larson, M. M. Harrison, GAF is essential for zygotic genome activation and chromatin accessibility in the early *Drosophila* embryo. *eLife*. **10**, e66668 (2021).
 47. S. A. Blythe, E. F. Wieschaus, Establishment and maintenance of heritable chromatin structure during early *Drosophila* embryogenesis. *eLife*. **5**, e20148 (2016).
 48. C. B. Hug, A. G. Grimaldi, K. Kruse, J. M. Vaquerizas, Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. *Cell*. **169**, 216-228.e19 (2017).
 49. Y. Endo, K. Kamei, M. Inoue-Murayama, Genetic Signatures of Evolution of the Pluripotency Gene Regulating Network across Mammals. *Genome Biology and Evolution*. **12**, 1806–1818 (2020).
 50. B. Fernandez-Tresguerres, S. Cañon, T. Rayon, B. Pernaute, M. Crespo, C. Torroja, M. Manzanares, Evolution of the mammalian embryonic pluripotency gene regulatory network. *Proceedings of the National Academy of Sciences*. **107**, 19955–19960 (2010).
 51. H. Aanes, C. L. Winata, C. H. Lin, J. P. Chen, K. G. Srinivasan, S. G. Lee, A. Y. Lim, H. S. Hajan, P. Collas, G. Bourque, Zebrafish mRNA sequencing deciphers novelties in

- transcriptome dynamics during maternal to zygotic transition. *Genome research*. **21**, 1328–1338 (2011).
52. A. A. Bazzini, M. T. Lee, A. J. Giraldez, Ribosome Profiling Shows That miR-430 Reduces Translation Before Causing mRNA Decay in Zebrafish. *Science*. **336**, 233–237 (2012).
 53. S. A. Harvey, I. Sealy, R. Kettleborough, F. Fenyés, R. White, D. Stemple, J. C. Smith, Identification of the zebrafish maternal and paternal transcriptomes. *Development (Cambridge, England)*. **140**, 2703–10 (2013).
 54. Y. Mishima, Y. Tomari, Codon Usage and 3' UTR Length Determine Maternal mRNA Stability in Zebrafish. *Molecular Cell*. **61**, 874–885 (2016).
 55. T. Hamatani, M. G. Carter, A. A. Sharov, M. S. Ko, Dynamics of global gene expression changes during mouse preimplantation development. *Developmental cell*. **6**, 117–31 (2004).
 56. L. R. Baugh, A. A. Hill, D. K. Slonim, E. L. Brown, C. P. Hunter, Composition and dynamics of the *Caenorhabditis elegans* early embryonic transcriptome. *Development (Cambridge, England)*. **130**, 889–900 (2003).
 57. S. Thomsen, S. Anders, S. C. Janga, W. Huber, C. R. Alonso, Genome-wide analysis of mRNA decay patterns during early *Drosophila* development. *Genome Biology*. **11**, R93 (2010).
 58. A. Bashirullah, S. R. Halsell, R. L. Cooperstock, M. Kloc, A. Karaiskakis, W. W. Fisher, W. Fu, J. K. Hamilton, L. D. Etkin, H. D. Lipshitz, Joint action of two RNA degradation pathways controls the timing of maternal transcript elimination at the midblastula transition in *Drosophila melanogaster*. *The EMBO journal*. **18**, 2610–20 (1999).
 59. A. Bashirullah, R. L. Cooperstock, H. D. Lipshitz, Spatial and temporal control of RNA stability. *Proceedings of the National Academy of Sciences*. **98**, 7025–7028 (2001).
 60. W. Tadros, S. A. Houston, A. Bashirullah, R. L. Cooperstock, J. L. Semotok, B. H. Reed, H. D. Lipshitz, Regulation of Maternal Transcript Destabilization During Egg Activation in *Drosophila*. *Genetics*. **164**, 989–1001 (2003).
 61. W. Tadros, A. L. Goldman, T. Babak, F. Menzies, L. Vardy, T. Orr-Weaver, T. R. Hughes, J. T. Westwood, C. A. Smibert, H. D. Lipshitz, SMAUG is a major regulator of maternal mRNA destabilization in *Drosophila* and its translation is activated by the PAN GU kinase. *Developmental cell*. **12**, 143–55 (2007).
 62. S. De Renzis, O. Elemento, S. Tavazoie, E. F. Wieschaus, Unmasking activation of the zygotic genome using chromosomal deletions in the *Drosophila* embryo. *PLoS biology*. **5**, e117 (2007).
 63. S. Mathavan, S. G. Lee, A. Mak, L. D. Miller, K. R. Murthy, K. R. Govindarajan, Y. Tong, Y. L. Wu, S. H. Lam, H. Yang, Y. Ruan, V. Korzh, Z. Gong, E. T. Liu, T. Lufkin,

- Transcriptome analysis of zebrafish embryogenesis using microarrays. *PLoS genetics*. **1**, 260–76 (2005).
64. M. H. Tan, K. F. Au, A. L. Yablonovitch, A. E. Wills, J. Chuang, J. C. Baker, W. H. Wong, J. B. Li, RNA sequencing reveals a diverse and dynamic repertoire of the *Xenopus tropicalis* transcriptome over development. *Genome research*. **23**, 201–16 (2013).
 65. A. Graindorge, R. Thuret, N. Pollet, H. B. Osborne, Y. Audic, Identification of post-transcriptionally regulated *Xenopus tropicalis* maternal mRNAs by microarray. *Nucleic Acids Research*. **34**, 986–995 (2006).
 66. P. Svoboda, V. Franke, R. M. Schultz, "Chapter Nine - Sculpting the Transcriptome During the Oocyte-to-Embryo Transition in Mouse" in *Current Topics in Developmental Biology*, H. D. Lipshitz, Ed. (Academic Press, 2015; <https://www.sciencedirect.com/science/article/pii/S0070215315000320>), vol. 113, pp. 305–349.
 67. L. Chen, J. G. Dumelie, X. Li, M. H. Cheng, Z. Yang, J. D. Laver, N. U. Siddiqui, J. T. Westwood, Q. Morris, H. D. Lipshitz, C. A. Smibert, Global regulation of mRNA translation and stability in the early *Drosophila* embryo by the Smaug RNA-binding protein. *Genome Biology*. **15**, R4 (2014).
 68. J. D. Laver, X. Li, D. Ray, K. B. Cook, N. A. Hahn, S. Nabeel-Shah, M. Kekis, H. Luo, A. J. Marsolais, K. Y. Fung, T. R. Hughes, J. T. Westwood, S. S. Sidhu, Q. Morris, H. D. Lipshitz, C. A. Smibert, Brain tumor is a sequence-specific RNA-binding protein that directs maternal mRNA clearance during the *Drosophila* maternal-to-zygotic transition. *Genome Biology*. **16**, 94 (2015).
 69. A. P. Gerber, S. Luschnig, M. A. Krasnow, P. O. Brown, D. Herschlag, Genome-wide identification of mRNAs associated with the translational regulator PUMILIO in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences*. **103**, 4487–4492 (2006).
 70. J. D. Laver, A. J. Marsolais, C. A. Smibert, H. D. Lipshitz, "Chapter Two - Regulation and Function of Maternal Gene Products During the Maternal-to-Zygotic Transition in *Drosophila*" in *Current Topics in Developmental Biology*, H. D. Lipshitz, Ed. (Academic Press, 2015; <https://www.sciencedirect.com/science/article/pii/S0070215315000356>), vol. 113, pp. 43–84.
 71. P. Bouvet, A. P. Wolffe, A role for transcription and FRGY2 in masking maternal mRNA within *Xenopus* oocytes. *Cell*. **77**, 931–941 (1994).
 72. K. Matsumoto, F. Meric, A. P. Wolffe, Translational repression dependent on the interaction of the *Xenopus* Y-box protein FRGY2 with mRNA: role of the cold shock domain, tail domain, and selective RNA sequence recognition. *Journal of biological chemistry*. **271**, 22706–22712 (1996).

73. J. Sun, L. Yan, W. Shen, A. Meng, Maternal Ybx1 safeguards zebrafish oocyte maturation and maternal-to-zygotic transition by repressing global translation. *Development*. **145**, dev166587 (2018).
74. S. Medvedev, H. Pan, R. M. Schultz, Absence of MSY2 in Mouse Oocytes Perturbs Oocyte Growth and Maturation, RNA Stability, and the Transcriptome1. *Biology of Reproduction*. **85**, 575–583 (2011).
75. J. Yu, N. B. Hecht, R. M. Schultz, Expression of MSY2 in Mouse Oocytes and Preimplantation Embryos1. *Biology of Reproduction*. **65**, 1260–1270 (2001).
76. A. Arnold, M. M. Rahman, M. C. Lee, S. Muehlhaeuser, I. Katic, D. Gaidatzis, D. Hess, C. Scheckel, J. E. Wright, A. Stetak, P. R. Boag, R. Ciosk, Functional characterization of C. elegans Y-box-binding proteins reveals tissue-specific functions and a critical role in the formation of polysomes. *Nucleic Acids Research*. **42**, 13353–13369 (2014).
77. P. R. Boag, A. Nakamura, T. K. Blackwell, A conserved RNA-protein complex component involved in physiological germline apoptosis regulation in C. elegans. *Development*. **132**, 4975–4986 (2005).
78. J. H. Mansfield, J. E. Wilhelm, T. Hazelrigg, Ypsilon Schachtel, a Drosophila Y-box protein, acts antagonistically to Orb in the oskar mRNA localization and translation pathway. *Development*. **129**, 197–209 (2002).
79. I. D'Agostino, C. Merritt, P.-L. Chen, G. Seydoux, K. Subramaniam, Translational repression restricts expression of the C. elegans Nanos homolog NOS-2 to the embryonic germline. *Developmental Biology*. **292**, 244–252 (2006).
80. C. M. Gallo, E. Munro, D. Rasoloson, C. Merritt, G. Seydoux, Processing bodies and germ granules are distinct RNA granules that interact in C. elegans embryos. *Developmental Biology*. **323**, 76–87 (2008).
81. C. M. Schubert, R. Lin, C. J. de Vries, R. H. A. Plasterk, J. R. Priess, MEX-5 and MEX-6 Function to Establish Soma/Germline Asymmetry in Early C. elegans Embryos. *Molecular Cell*. **5**, 671–682 (2000).
82. M. Rabani, L. Pieper, G.-L. Chew, A. F. Schier, A Massively Parallel Reporter Assay of 3' UTR Sequences Identifies In Vivo Rules for mRNA Degradation. *Molecular Cell*. **68**, 1083-1094.e5 (2017).
83. L. Detivaud, G. Pascreau, A. Karaïskou, H. B. Osborne, J. Z. Kubiak, Regulation of EDEN-dependent deadenylation of Aurora A/Eg2-derived mRNA via phosphorylation and dephosphorylation in Xenopus laevis egg extracts. *Journal of Cell Science*. **116**, 2697–2705 (2003).
84. L. Paillard, F. Omilli, V. Legagneux, T. Bassez, D. Maniey, H. B. Osborne, EDEN and EDEN-BP, a cis element and an associated factor that mediate sequence-specific mRNA deadenylation in Xenopus embryos. *The EMBO journal*. **17**, 278–87 (1998).

85. G. K. Voeltz, J. A. Steitz, AUUUA sequences direct mRNA deadenylation uncoupled from decay during *Xenopus* early development. *Molecular and cellular biology*. **18**, 7537–45 (1998).
86. S. B. V. Ramos, D. J. Stumpo, E. A. Kennington, R. S. Phillips, C. B. Bock, F. Ribeiro-Neto, P. J. Blackshear, The CCCH tandem zinc-finger protein Zfp3612 is crucial for female fertility and early embryonic development. *Development*. **131**, 4883–4893 (2004).
87. A. J. Giraldez, Y. Mishima, J. Rihel, R. J. Grocock, S. Van Dongen, K. Inoue, A. J. Enright, A. F. Schier, Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science (New York, N.Y.)*. **312**, 75–9 (2006).
88. E. Lund, M. Liu, R. S. Hartley, M. D. Sheets, J. E. Dahlberg, Deadenylation of maternal mRNAs mediated by miR-427 in *Xenopus laevis* embryos. *RNA (New York, N.Y.)*. **15**, 2351–63 (2009).
89. N. Bushati, A. Stark, J. Brennecke, S. M. Cohen, Temporal reciprocity of miRNAs and their targets during the maternal-to-zygotic transition in *Drosophila*. *Current biology : CB*. **18**, 501–6 (2008).
90. J. Newport, M. Kirschner, A major developmental transition in early *Xenopus* embryos: II. Control of the onset of transcription. *Cell*. **30**, 687–96 (1982).
91. G. Almouzni, M. Méchali, A. P. Wolffe, Competition between transcription complex assembly and chromatin assembly on replicating DNA. *The EMBO Journal*. **9**, 573–582 (1990).
92. A. A. Amodeo, D. Jukam, A. F. Straight, J. M. Skotheim, Histone titration against the genome sets the DNA-to-cytoplasm threshold for the *Xenopus* midblastula transition. *Proceedings of the National Academy of Sciences of the United States of America*. **112**, E1086-95 (2015).
93. R. M. Charney, E. Forouzmand, J. S. Cho, J. Cheung, K. D. Paraiso, Y. Yasuoka, S. Takahashi, M. Taira, I. L. Blitz, X. Xie, K. W. Cho, Foxh1 Occupies cis-Regulatory Modules Prior to Dynamic Transcription Factor Interactions Controlling the Mesendoderm Gene Program. *Developmental cell*. **40**, 595-607.e4 (2017).
94. J. Zhang, D. W. Houston, M. L. King, C. Payne, C. Wylie, J. Heasman, The Role of Maternal VegT in Establishing the Primary Germ Layers in *Xenopus* Embryos. *Cell*. **94**, 515–524 (1998).
95. M. Kofron, T. Demel, J. Xanthos, J. Lohr, B. Sun, H. Sive, S. Osada, C. Wright, C. Wylie, J. Heasman, Mesoderm induction in *Xenopus* is a zygotic event regulated by maternal VegT via TGFbeta growth factors. *Development*. **126**, 5759–5770 (1999).
96. Y. Yasuoka, Y. Suzuki, S. Takahashi, H. Someya, N. Sudou, Y. Haramoto, K. W. Cho, M. Asashima, S. Sugano, M. Taira, Occupancy of tissue-specific cis-regulatory modules by Otx2 and TLE/Groucho for embryonic head specification. *Nature Communications*. **5**, 4322 (2014).

97. E. Gazdag, U. G. Jacobi, I. van Kruijsbergen, D. L. Weeks, G. J. C. Veenstra, Activation of a T-box-Otx2-Gsc gene network independent of TBP and TBP-related factors. *Development*. **143**, 1340–1350 (2016).
98. G. E. Gentsch, N. D. L. Owens, J. C. Smith, The Spatiotemporal Control of Zygotic Genome Activation. *iScience*. **16**, 485–498 (2019).
99. A. M. Session, Y. Uno, T. Kwon, J. A. Chapman, A. Toyoda, S. Takahashi, A. Fukui, A. Hikosaka, A. Suzuki, M. Kondo, S. J. van Heeringen, I. Quigley, S. Heinz, H. Ogino, H. Ochi, U. Hellsten, J. B. Lyons, O. Simakov, N. Putnam, J. Stites, Y. Kuroki, T. Tanaka, T. Michiue, M. Watanabe, O. Bogdanovic, R. Lister, G. Georgiou, S. S. Paranjpe, I. van Kruijsbergen, S. Shu, J. Carlson, T. Kinoshita, Y. Ohta, S. Mawaribuchi, J. Jenkins, J. Grimwood, J. Schmutz, T. Mitros, S. V. Mozaffari, Y. Suzuki, Y. Haramoto, T. S. Yamamoto, C. Takagi, R. Heald, K. Miller, C. Haudenschild, J. Kitzman, T. Nakayama, Y. Izutsu, J. Robert, J. Fortriede, K. Burns, V. Lotay, K. Karimi, Y. Yasuoka, D. S. Dichmann, M. F. Flajnik, D. W. Houston, J. Shendure, L. DuPasquier, P. D. Vize, A. M. Zorn, M. Ito, E. M. Marcotte, J. B. Wallingford, Y. Ito, M. Asashima, N. Ueno, Y. Matsuda, G. J. Veenstra, A. Fujiyama, R. M. Harland, M. Taira, D. S. Rokhsar, Genome evolution in the allotetraploid frog *Xenopus laevis*. *Nature*. **538**, 336–343 (2016).
100. H. R. Kobel, L. Du Pasquier, Genetics of polyploid *Xenopus*. *Trends in Genetics*. **2**, 310–315 (1986).
101. C. A. Bisbee, M. A. Baker, A. C. Wilson, I. Haji-Azimi, M. Fischberg, Albumin phylogeny for clawed frogs (*Xenopus*). *Science (New York, N.Y.)*. **195**, 785–7 (1977).
102. D. M. Elurbe, S. S. Paranjpe, G. Georgiou, I. van Kruijsbergen, O. Bogdanovic, R. Gibeaux, R. Heald, R. Lister, M. A. Huynen, S. J. van Heeringen, G. J. C. Veenstra, Regulatory remodeling in the allo-tetraploid frog *Xenopus laevis*. *Genome biology*. **18**, 198 (2017).
103. G. Hu, J. F. Wendel, Cis–trans controls and regulatory novelty accompanying allopolyploidization. *New Phytologist*. **221**, 1691–1700 (2019).
104. B. M. Moran, C. Payne, Q. Langdon, D. L. Powell, Y. Brandvain, M. Schumer, The genomic consequences of hybridization. *eLife*. **10**, e69016 (2021).
105. C. E. Grover, J. P. Gallagher, E. P. Szadkowski, M. J. Yoo, L. E. Flagel, J. F. Wendel, Homoeolog expression bias and expression level dominance in allopolyploids. *The New Phytologist*. **196**, 966–971 (2012).
106. M. J. Song, B. I. Potter, J. J. Doyle, J. E. Coate, Gene Balance Predicts Transcriptional Responses Immediately Following Ploidy Change in *Arabidopsis thaliana*. *The Plant Cell*. **32**, 1434–1448 (2020).
107. K. B. S. Swamy, S. C. Schuyler, J.-Y. Leu, Protein Complexes Form a Basis for Complex Hybrid Incompatibility. *Frontiers in Genetics*. **12** (2021) (available at <https://www.frontiersin.org/articles/10.3389/fgene.2021.609766>).

108. K. L. Adams, J. F. Wendel, Polyploidy and genome evolution in plants. *Curr Opin Plant Biol.* **8**, 135–141 (2005).
109. B. C. Husband, S. J. Baldwin, J. Suda, "The Incidence of Polyploidy in Natural Plant Populations: Major Patterns and Evolutionary Processes" in *Plant Genome Diversity Volume 2: Physical Structure, Behaviour and Evolution of Plant Genomes*, J. Greilhuber, J. Dolezel, J. F. Wendel, Eds. (Springer Vienna, Vienna, 2013; https://doi.org/10.1007/978-3-7091-1160-4_16), pp. 255–276.
110. B. K. Mable, 'Why polyploidy is rarer in animals than in plants': myths and mechanisms. *Biological Journal of the Linnean Society.* **82**, 453–466 (2004).
111. Z. Chen, Y. Omori, S. Koren, T. Shirokiya, T. Kuroda, A. Miyamoto, H. Wada, A. Fujiyama, A. Toyoda, S. Zhang, De novo assembly of the goldfish (*Carassius auratus*) genome and the evolution of genes after whole-genome duplication. *Science advances.* **5**, eaav0547 (2019).
112. J.-T. Li, Q. Wang, M.-D. Huang Yang, Q.-S. Li, M.-S. Cui, Z.-J. Dong, H.-W. Wang, J.-H. Yu, Y.-J. Zhao, C.-R. Yang, Y.-X. Wang, X.-Q. Sun, Y. Zhang, R. Zhao, Z.-Y. Jia, X.-Y. Wang, Parallel subgenome structure and divergent expression evolution of allo-tetraploid common carp and goldfish. *Nature Genetics.* **53**, 1493–1503 (2021).
113. Luo Jing, Chai Jing, Wen Yanling, Tao Min, Lin Guoliang, Liu Xiaochuan, Ren Li, Chen Zeyu, Wu Shigang, Li Shengnan, Wang Yude, Qin Qinbo, Wang Shi, Gao Yun, Huang Feng, Wang Lu, Ai Cheng, Wang Xiaobo, Li Lianwei, Ye Chengxi, Yang Huimin, Luo Mi, Chen Jie, Hu Hong, Yuan Liujiào, Zhong Li, Wang Jing, Xu Jian, Du Zhenglin, Ma Zhanshan (Sam), Murphy Robert W., Meyer Axel, Gui Jianfang, Xu Peng, Ruan Jue, Chen Z. Jeffrey, Liu Shaojun, Lu Xuemei, Zhang Ya-ping, From asymmetrical to balanced genomic diversification during rediploidization: Subgenomic evolution in allotetraploid fish. *Science Advances.* **6**, eaaz7677.
114. W. A. Phelps, A. E. Carlson, M. T. Lee, Optimized design of antisense oligomers for targeted rRNA depletion. *Nucleic acids research.* **49**, e5 (2021).
115. B. Schmidt, A. Hildebrandt, Next-generation sequencing: big data meets high performance computing. *Drug Discovery Today.* **22**, 712–717 (2017).
116. Z. Wang, M. Gerstein, M. Snyder, RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews. Genetics.* **10**, 57–63 (2009).
117. P. H. Sudmant, M. S. Alexis, C. B. Burge, Meta-analysis of RNA-seq expression data across species, tissues and studies. *Genome biology.* **16**, 287 (2015).
118. S. Schuierer, W. Carbone, J. Knehr, V. Petitjean, A. Fernandez, M. Sultan, G. Roma, A comprehensive assessment of RNA-seq protocols for degraded and low-quantity samples. *BMC Genomics.* **18**, 442 (2017).
119. T. S. Furey, ChIP-seq and beyond: new and improved methodologies to detect and characterize protein–DNA interactions. *Nature Reviews Genetics.* **13**, 840–852 (2012).

120. P. J. Park, ChIP-seq: advantages and challenges of a maturing technology. *Nature Reviews Genetics*. **10**, 669–680 (2009).
121. S. Zhao, Y. Zhang, R. Gamini, B. Zhang, D. von Schack, Evaluation of two main RNA-seq approaches for gene quantification in clinical RNA sequencing: polyA⁺ selection versus rRNA depletion. *Scientific reports*. **8**, 4781 (2018).
122. D. O’Neil, H. Glowatz, M. Schlumpberger, *Current protocols in molecular biology*, in press, doi:10.1002/0471142727.mb0419s103.
123. A. H. van Vliet, Next generation sequencing of microbial transcriptomes: challenges and opportunities. *FEMS microbiology letters*. **302**, 1–7 (2010).
124. F. Cheung, B. J. Haas, S. M. Goldberg, G. D. May, Y. Xiao, C. D. Town, Sequencing *Medicago truncatula* expressed sequenced tags using 454 Life Sciences technology. *BMC genomics*. **7**, 272 (2006).
125. J. D. Morlan, K. Qu, D. V. Sinicropi, Selective depletion of rRNA enables whole transcriptome profiling of archival fixed tissue. *PloS one*. **7**, e42882 (2012).
126. A. J. Kraus, B. G. Brink, T. N. Siegel, Efficient and specific oligo-based depletion of rRNA. *Scientific reports*. **9**, 12281 (2019).
127. X. Adiconis, A. L. Haber, S. K. Simmons, A. Levy Moonshine, Z. Ji, M. A. Busby, X. Shi, J. Jacques, M. A. Lancaster, J. Q. Pan, A. Regev, J. Z. Levin, Comprehensive comparative analysis of 5’-end RNA-sequencing methods. *Nature methods*. **15**, 505–511 (2018).
128. N. T. Ingolia, S. Ghaemmaghami, J. R. Newman, J. S. Weissman, Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science (New York, N.Y.)*. **324**, 218–23 (2009).
129. L. Yang, M. O. Duff, B. R. Graveley, G. G. Carmichael, L. L. Chen, Genomewide characterization of non-polyadenylated RNAs. *Genome biology*. **12**, R16 (2011).
130. N. L. Garneau, J. Wilusz, C. J. Wilusz, The highways and byways of mRNA decay. *Nature reviews. Molecular cell biology*. **8**, 113–26 (2007).
131. J. D. Richter, Cytoplasmic polyadenylation in development and beyond. *Microbiology and molecular biology reviews : MMBR*. **63**, 446–56 (1999).
132. M. T. Lee, A. R. Bonneau, A. J. Giraldez, Zygotic genome activation during the maternal-to-zygotic transition. *Annu Rev Cell Dev Biol*. **30**, 581–613 (2014).
133. C. Collart, N. D. Owens, L. Bhaw-Rosun, B. Cooper, E. De Domenico, I. Patrushev, A. K. Sesay, J. N. Smith, J. C. Smith, M. J. Gilchrist, High-resolution analysis of gene activity during the *Xenopus* mid-blastula transition. *Development (Cambridge, England)*. **141**, 1927–39 (2014).

134. S. S. Paranjpe, U. G. Jacobi, S. J. van Heeringen, G. J. Veenstra, A genome-wide survey of maternal and embryonic transcripts during *Xenopus tropicalis* development. *BMC genomics*. **14**, 762 (2013).
135. X. Adiconis, D. Borges-Rivera, R. Satija, D. S. DeLuca, M. A. Busby, A. M. Berlin, A. Sivachenko, D. A. Thompson, A. Wysoker, T. Fennell, A. Gnirke, N. Pochet, A. Regev, J. Z. Levin, Comparative analysis of RNA sequencing methods for degraded or low-input samples. *Nature methods*. **10**, 623–9 (2013).
136. P. H. Culviner, C. K. Guegler, M. T. Laub, A Simple, Cost-Effective, and Robust Method for rRNA Depletion in RNA-Sequencing Studies. *mBio*. **11** (2020), doi:10.1128/mBio.00010-20.
137. Y. Huang, R. U. Sheth, A. Kaufman, H. H. Wang, Scalable and cost-effective ribonuclease-based rRNA depletion for transcriptomics. *Nucleic acids research*. **48**, e20 (2020).
138. A. Baldwin, A. R. Morris, N. Mukherjee, *bioRxiv*, in press, doi:10.1101/2020.05.21.109033.
139. J. R. Fauver, S. Akter, A. I. O. Morales, W. C. th Black, A. D. Rodriguez, M. D. Stenglein, G. D. Ebel, J. Weger-Lucarelli, A reverse-transcription/RNase H based protocol for depletion of mosquito ribosomal RNA facilitates viral intrahost evolution analysis, transcriptomics and pathogen discovery. *Virology*. **528**, 181–197 (2019).
140. and D P Mindell, R. L. Honeycutt, Ribosomal RNA in Vertebrates: Evolution and Phylogenetic Applications. *Annual Review of Ecology and Systematics*. **21**, 541–566 (1990).
141. S. H. J. Heasman C. C. Wylie, "Fertilization of Cultured *Xenopus* Oocytes and Use in Studies of Maternally Inherited Molecules" in *Methods in Cell Biology* (Elsevier, Cambridge, England, 1991; [https://doi.org/10.1016/S0091-679X\(08\)60279-4](https://doi.org/10.1016/S0091-679X(08)60279-4)), vol. 36, pp. 213–218.
142. R. M. G. Hazel L Sive Richard M. Harland Hazel L., in *Early development of *Xenopus laevis*: a laboratory manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 2000).
143. W. A. Kibbe, OligoCalc: an online oligonucleotide properties calculator. *Nucleic acids research*. **35**, W43-6 (2007).
144. K. J. Breslauer, R. Frank, H. Blöcker, L. A. Marky, Predicting DNA duplex stability from the base sequence. *Proceedings of the National Academy of Sciences of the United States of America*. **83**, 3746–50 (1986).
145. N. Sugimoto, S. Nakano, M. Katoh, A. Matsumura, H. Nakamuta, T. Ohmichi, M. Yoneyama, M. Sasaki, Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes. *Biochemistry*. **34**, 11211–6 (1995).
146. D. Banerjee, H. Tateishi-Karimata, T. Ohyama, S. Ghosh, T. Endoh, S. Takahashi, N. Sugimoto, Improved nearest-neighbor parameters for the stability of RNA/DNA hybrids under a physiological condition. *Nucleic Acids Research* (2020), doi:10.1093/nar/gkaa572.

147. A. P. Wolffe, D. D. Brown, Developmental regulation of two 5S ribosomal RNA genes. *Science (New York, N.Y.)*. **241**, 1626–32 (1988).
148. M. D. Locati, J. F. Pagano, W. A. Ensink, M. van Olst, S. van Leeuwen, U. Nehrdich, K. Zhu, H. P. Spaink, G. Girard, H. Rauwerda, M. J. Jonker, R. J. Dekker, T. M. Breit, Linking maternal and somatic 5S rRNA types with different sequence-specific non-LTR retrotransposons. *RNA (New York, N.Y.)*. **23**, 446–456 (2017).
149. M. D. Locati, J. F. B. Pagano, G. Girard, W. A. Ensink, M. van Olst, S. van Leeuwen, U. Nehrdich, H. P. Spaink, H. Rauwerda, M. J. Jonker, R. J. Dekker, T. M. Breit, Expression of distinct maternal and somatic 5.8S, 18S, and 28S rRNA types during zebrafish development. *RNA (New York, N.Y.)*. **23**, 1188–1199 (2017).
150. B. B. Mughal, M. Leemans, P. Spirhanzlova, B. Demeneix, J. B. Fini, Reference gene identification and validation for quantitative real-time PCR studies in developing *Xenopus laevis*. *Scientific reports*. **8**, 496 (2018).
151. D. Kim, B. Langmead, S. L. Salzberg, HISAT: a fast spliced aligner with low memory requirements. *Nature methods*. **12**, 357–60 (2015).
152. Y. Liao, G. K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics (Oxford, England)*. **30**, 923–30 (2014).
153. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)*. **26**, 841–2 (2010).
154. W. J. Kent, C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler, D. Haussler, The human genome browser at UCSC. *Genome research*. **12**, 996–1006 (2002).
155. E. J. Draizen, A. K. Shaytan, L. Mariño-Ramírez, P. B. Talbert, D. Landsman, A. R. Panchenko, HistoneDB 2.0: a histone database with variants--an integrated resource to explore histones and their variants. *Database : the journal of biological databases and curation*. **2016** (2016), doi:10.1093/database/baw014.
156. C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T. L. Madden, BLAST+: architecture and applications. *BMC Bioinformatics*. **10**, 421 (2009).
157. R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*. **32**, 1792–7 (2004).
158. W. F. Marzluff, E. J. Wagner, R. J. Duronio, Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. *Nature reviews. Genetics*. **9**, 843–54 (2008).
159. B. Zinshteyn, J. R. Wangen, B. Hua, R. Green, Nuclease-mediated depletion biases in ribosome footprint profiling libraries. *RNA (New York, N.Y.)* (2020), doi:10.1261/rna.075523.120.

160. L. G. Acevedo, A. Leonardo Iniguez, H. L. Holster, X. Zhang, R. Green, P. J. Farnham, Genome-scale ChIP-chip analysis using 10,000 human cells. *BioTechniques*. **43**, 791–797 (2007).
161. P. J. Skene, S. Henikoff, An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife*. **6** (2017), doi:10.7554/eLife.21856.
162. S. J. Hainer, A. Boskovic, K. N. McCannell, O. J. Rando, T. G. Fazio, Profiling of Pluripotency Factors in Single Cells and Early Embryos. *Cell*. **177**, 1319-1329.e11 (2019).
163. J. A. Briggs, C. Weinreb, D. E. Wagner, S. Megason, L. Peshkin, M. W. Kirschner, A. M. Klein, The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science (New York, N.Y.)*. **360** (2018), doi:10.1126/science.aar5780.
164. S. J. Hainer, T. G. Fazio, High-Resolution Chromatin Profiling Using CUT&RUN. *Current protocols in molecular biology*. **126**, e85 (2019).
165. K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, T. J. Richmond, Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*. **389**, 251–260 (1997).
166. T. J. Richmond, C. A. Davey, The structure of DNA in the nucleosome core. *Nature*. **423**, 145–150 (2003).
167. R. D. Kornberg, Chromatin structure: a repeating unit of histones and DNA. *Science (New York, N.Y.)*. **184**, 868–71 (1974).
168. P. G. Giresi, J. Kim, R. M. McDaniell, V. R. Iyer, J. D. Lieb, FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome research*. **17**, 877–885 (2007).
169. J. M. Rizzo, S. Sinha, "Analyzing the Global Chromatin Structure of Keratinocytes by MNase-Seq" in *Epidermal Cells: Methods and Protocols*, K. Turksen, Ed. (Springer New York, New York, NY, 2014; https://doi.org/10.1007/7651_2014_77), pp. 49–59.
170. L. Song, G. E. Crawford, *Cold Spring Harbor Protocols*, in press.
171. J. D. Buenrostro, B. Wu, H. Y. Chang, W. J. Greenleaf, *Current protocols in molecular biology*, in press, doi:10.1002/0471142727.mb2129s109.
172. M. Tsompana, M. J. Buck, Chromatin accessibility: a window into the genome. *Epigenetics & Chromatin*. **7**, 33 (2014).
173. J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, W. J. Greenleaf, Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature methods*. **10**, 1213–8 (2013).

174. M. Palfy, G. Schulze, E. Valen, N. L. Vastenhouw, Chromatin accessibility established by Pou5f3, Sox19b and Nanog primes genes for activity during zebrafish genome activation. *PLoS genetics*. **16**, e1008546 (2020).
175. M. Esmaeili, S. A. Blythe, J. W. Tobias, K. Zhang, J. Yang, P. S. Klein, Chromatin accessibility and histone acetylation in the regulation of competence in early development. *Developmental biology*. **462**, 20–35 (2020).
176. W. A. Phelps, M. D. Hurton, T. N. Ayers, A. E. Carlson, J. C. Rosenbaum, M. T. Lee, *bioRxiv*, in press, doi:10.1101/2022.09.14.507817.
177. I. Yanai, L. Peshkin, P. Jorgensen, M. W. Kirschner, Mapping Gene Expression in Two Xenopus Species: Evolutionary Constraints and Developmental Flexibility. *Developmental Cell*. **20**, 483–496 (2011).
178. R. M. Harland, R. M. Grainger, Xenopus research: metamorphosed by genetics and genomics. *Trends in Genetics*. **27**, 507–515 (2011).
179. H. Ochi, A. Kawaguchi, M. Tanouchi, N. Suzuki, T. Kumada, Y. Iwata, H. Ogino, 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 (2017).
180. H. Chen, L. C. Einstein, S. C. Little, M. C. Good, Spatiotemporal Patterning of Zygotic Genome Activation in a Model Vertebrate Embryo. *Developmental Cell*. **49**, 852-866.e7 (2019).
181. R. Gibeaux, R. Acker, M. Kitaoka, G. Georgiou, I. van Kruijsbergen, B. Ford, E. M. Marcotte, D. K. Nomura, T. Kwon, G. J. C. Veenstra, R. Heald, Paternal chromosome loss and metabolic crisis contribute to hybrid inviability in Xenopus. *Nature*. **553**, 337–341 (2018).
182. D. Kimelman, M. Kirschner, T. Scherson, The events of the midblastula transition in Xenopus are regulated by changes in the cell cycle. *Cell*. **48**, 399–407 (1987).
183. G. J. Veenstra, O. H. Destree, A. P. Wolffe, 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**, 7972–82 (1999).
184. J. Skirkanich, G. Luxardi, J. Yang, L. Kodjabachian, P. S. Klein, An essential role for transcription before the MBT in Xenopus laevis. *Developmental Biology*. **357**, 478–491 (2011).
185. K. D. Paraiso, I. L. Blitz, M. Coley, J. Cheung, N. Sudou, M. Taira, K. W. Y. Cho, Endodermal Maternal Transcription Factors Establish Super-Enhancers during Zygotic Genome Activation. *Cell Reports*. **27**, 2962-2977.e5 (2019).

186. G. M. Morrison, J. M. Brickman, Conserved roles for Oct4 homologues in maintaining multipotency during early vertebrate development. *Development (Cambridge, England)*. **133**, 2011–22 (2006).
187. C. Zhang, T. Basta, E. D. Jensen, M. W. Klymkowsky, The beta-catenin/VegT-regulated early zygotic gene *Xnr5* is a direct target of SOX3 regulation. *Development (Cambridge, England)*. **130**, 5609–24 (2003).
188. S. Picelli, A. K. Björklund, B. Reinius, S. Sagasser, G. Winberg, R. Sandberg, Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome research*. **24**, 2033–40 (2014).
189. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*. **15**, 550 (2014).
190. A. Kozomara, M. Birgaoanu, S. Griffiths-Jones, miRBase: from microRNA sequences to function. *Nucleic Acids Research*. **47**, D155–D162 (2019).
191. W. J. Kent, BLAT—the BLAST-like alignment tool. *Genome research*. **12**, 656–664 (2002).
192. Z. Yang, PAML: a program package for phylogenetic analysis by maximum likelihood. *Bioinformatics*. **13**, 555–556 (1997).
193. M. Suyama, D. Torrents, P. Bork, PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research*. **34**, W609–W612 (2006).
194. P. Rice, I. Longden, A. Bleasby, EMBOSS: the European molecular biology open software suite. *Trends in genetics*. **16**, 276–277 (2000).
195. R. C. R. Core Team, R: A language and environment for statistical computing (2013).
196. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nature methods*. **9**, 357–9 (2012).
197. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*. **17** (2011), p. 3.
198. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. **25**, 2078–2079 (2009).
199. F. Ramírez, F. Dünder, S. Diehl, B. A. Grüning, T. Manke, deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Research*. **42**, W187–W191 (2014).
200. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, X. S. Liu, Model-based analysis of ChIP-Seq (MACS). *Genome biology*. **9**, R137 (2008).

201. M. P. Meers, D. Tenenbaum, S. Henikoff, Peak calling by Sparse Enrichment Analysis for CUT&RUN chromatin profiling. *Epigenetics & chromatin*. **12**, 42 (2019).
202. S. Heinz, C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh, C. K. Glass, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell*. **38**, 576–89 (2010).
203. R. S. Harris, *Improved pairwise alignment of genomic DNA* (The Pennsylvania State University, 2007).
204. N. D. Owens, I. L. Blitz, M. A. Lane, I. Patrushev, J. D. Overton, M. J. Gilchrist, K. W. Cho, M. K. Khokha, Measuring absolute RNA copy numbers at high temporal resolution reveals transcriptome kinetics in development. *Cell reports*. **14**, 632–647 (2016).
205. R. Gupta, A. Wills, D. Ucar, J. Baker, Developmental enhancers are marked independently of zygotic Nodal signals in *Xenopus*. *Developmental Biology*. **395**, 38–49 (2014).
206. O. Bogdanovic, A. Fernandez-Minan, J. J. Tena, E. de la Calle-Mustienes, C. Hidalgo, I. van Kruysbergen, S. J. van Heeringen, G. J. Veenstra, J. L. Gomez-Skarmeta, Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis. *Genome research*. **22**, 2043–53 (2012).
207. R. J. A. Buggs, J. F. Wendel, J. J. Doyle, D. E. Soltis, P. S. Soltis, J. E. Coate, The legacy of diploid progenitors in allopolyploid gene expression patterns. *Philosophical Transactions of the Royal Society B: Biological Sciences*. **369**, 20130354 (2014).
208. McClintock Barbara, The Significance of Responses of the Genome to Challenge. *Science*. **226**, 792–801 (1984).
209. Tirosh Itay, Reikhav Sharon, Levy Avraham A., Barkai Naama, A Yeast Hybrid Provides Insight into the Evolution of Gene Expression Regulation. *Science*. **324**, 659–662 (2009).
210. X. Shi, D. W.-K. Ng, C. Zhang, L. Comai, W. Ye, Z. Jeffrey Chen, Cis- and trans-regulatory divergence between progenitor species determines gene-expression novelty in Arabidopsis allopolyploids. *Nature Communications*. **3**, 950 (2012).
211. B. J. Evans, T. F. Carter, E. Greenbaum, V. Gvoždík, D. B. Kelley, P. J. McLaughlin, O. S. G. Pauwels, D. M. Portik, E. L. Stanley, R. C. Tinsley, M. L. Tobias, D. C. Blackburn, Genetics, Morphology, Advertisement Calls, and Historical Records Distinguish Six New Polyploid Species of African Clawed Frog (*Xenopus*, Pipidae) from West and Central Africa. *PLOS ONE*. **10**, e0142823 (2015).