Defining the interaction of Zelda with chromatin

By

Katharine N. Schulz

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Biochemistry)

at the

UNIVERSITY OF WISCONSIN-MADISON

2018

Date of final oral examination: 7/19/2018

The dissertation is approved by the following members of the Final Oral Committee:
Melissa M. Harrison, Assistant Professor, Biomolecular Chemistry
John M. Denu, Professor, Biomolecular Chemistry
Jill C. Wildonger, Assistant Professor, Biochemistry
Judith Kimble, Vilas Professor, Biochemistry
David Wassarman, Professor, Genetics

Dedicated to the Harrison Lab: past, present, and future.

Acknowledgements

When I rotated in the Harrison lab back in 2012, it was still in its infancy. The benches were spotless, the pipettes were fresh from the box, and there was a list of exciting, unclaimed projects waiting to be adopted. But while I was intrigued by this mysterious protein Zelda and wooed by the immaculate lab space, Melissa herself was the reason I decided to join. I may have been naive about all other aspects of grad school, but my sense that Melissa would make a great mentor was dead on. From training me at the bench to coaching me through my meager first attempts at writing, she made all of this happen. Outside the whole science thing, she has taught me how to be a more effective human. Thank you, Melissa, for being so understanding (even those times I totally lost it) and for being an excellent role model.

I also owe an enormous thanks to my labmate Danielle Hamm. Together, we were Melissa's first students and made up the entirety of the lab for a time. I am serious when I say that this work couldn't have been done without Danielle. She is what keeps the Harrison lab operating, and I can only wonder how much she has secretly helped me with all my experiments. But even more importantly, Danielle has kept me sane through many moments of disproportionate hysteria. Thank you, Danielle, for so very many things. I hope that a smidgen of your empathy and conscientiousness have rubbed off on me.

Over the years, we have watched our budding lab grow. When Markus Nevil joined, he brought with him a great sense of curiosity and much-needed calm. In recent years, our lab has been enlivened by the youthful energy of Liz Larson, the unyielding optimism of Marissa Gaskill, the levity of Stephen McDaniel, and the thoughtfulness of Tyler Gibson. Thank you all for making the lab a fun and interesting place to be. I am excited to see what discoveries await you.

Thank you to my committee members, who each brought something unique to the table. Jill Wildonger was our source for all things Drosophila, including the food that fed my countless flies. John Denu was our chromatin and protein chemistry expert. David Wassarman, the best lecturer I've ever

had, taught me all the basics of eukaryotic molecular biology. And Judith Kimble showed me how fun speculating about science can be in her lively seminar. Thank you all for your guidance and support.

Thank you to my IPiB cohort. Carousing with you made grad school fun, commiserating with you made it endurable. I'll be rooting for you. Thank you to my friends, near and far. And a special thanks to Mark DeCanio for being there with me through it all.

Finally, thank you to my family. To my sister Isa for helping me to see the world differently, and to my brother Zach for bringing the party wherever he goes. To my parents and grandparents, I truly appreciate how much you've done to pave the way for my privileged existence. Mom, you taught me to be dedicated. Mapz, you taught me to be critical. I see more and more of you in me every day, and I'm (mostly) thankful for that.

Abstract

Cell fate transitions are initiated by key transcription factors with a potent ability to reprogram the transcriptome. These master regulator transcription factors have been identified in numerous cellular contexts, yet, the mechanisms by which they produce global transcriptional change remain unclear. To further our understanding of how these factors function, I have studied a factor that initiates a massive transcriptional shift during embryogenesis.

In animals, the genome remains transcriptionally dormant following fertilization and development is launched by maternally supplied RNAs and proteins. Yet, within a matter of hours, the genome must undergo widespread transcriptional activation for development to proceed. In *Drosophila melanogaster* the transcription factor Zelda (ZLD) directs zygotic genome activation (ZGA). ZLD is required for the expression of hundreds of genes during ZGA, but the mechanisms by which ZLD initiates this widespread activation were initially unknown.

In this thesis, I present evidence that ZLD shapes the chromatin landscape to poise the zygotic genome for activation. I begin by introducing ZGA in Chapter 1, drawing on work from a range of model organisms. In Chapter 2, I include published work that demonstrates a role for ZLD in chromatin regulation. Here, we found that ZLD is required to maintain chromatin accessibility at a subset of its binding sites, and that this accessibility facilitates transcription factor binding and gene expression. Notably, we identified thousands of ZLD-binding sites that do not require ZLD for accessibility, and we determined that the chromatin regulator GAGA Factor (GAF) is enriched at these loci. Based on this finding, we proposed that GAF may serve as a co-regulator of ZGA. In Chapter 3, I describe my plan to assess this potential role for GAF using MS2-based reporters. Finally, to better understand how ZLD might regulate chromatin, I recently tested the ability of ZLD to bind to nucleosomes *in vitro*. In Chapter 4, I include evidence that ZLD has a unique ability to target nucleosomal DNA.

Table of Contents

Acknowledgements ii
Abstract iv
Table of Contents v
List of Tables and Figures vii
List of Abbreviations ix
Chapter 1 - Mechanisms of zygotic genome activation 1
Abstract 2
Introduction 3
Dynamics of genome activation 4
Mechanisms that control ZGA timing 5
Chromatin reprogramming during ZGA 8
Transcription factors direct ZGA 12
Conclusions and perspective 16
Figures 19
References 22
Chapter 2 - Zelda regulates chromatin accessibility in the early Drosophila embryo 34
Abstract 35
Introduction 36
Results 39
Discussion 51
Materials and Methods 57
Acknowledgments 61
Figures 62
References 78
Chapter 3 - GAGA Factor may serve as a co-activator of the Drosophila genome 83
Abstract 84
Introduction 85
Results 87
Discussion 90
Materials and Methods 93
Figures 95
References 98

Chapter 4 - Zelda binds nucleosomal DNA 101
Abstract 102
Introduction 103
Results 105
Discussion 108
Materials and Methods 111
Acknowledgements 114
Figures 115
References 120
Chapter 5 - Discussion and Future Directions 123
References 130

List of Tables and Figures

Table 2.1 76

Sup. Table 2.1 77

Figure 1.1: Zygotic genome activation is conserved across animals. 19

Figure 1.2: Several mechanisms contribute to the timing of zygotic genome activation 20

Figure 1.3. Chromatin undergoes reorganization during zygotic genome activation. 21

Figure 2.1: ZLD is required to maintain or establish discrete regions of open chromatin. 62

Figure 2.2: Many loci bound by ZLD in wild-type embryos remain accessible in zldM- embryos. 63

Figure 2.3: ZLD-mediated chromatin accessibility is correlated with early ZLD-dependent gene expression. 64

Figure 2.4: Early histone acetylation is enriched around loci that require ZLD for chromatin accessibility....................65

Figure 2.5: In the absence of ZLD, BCD binding is preferentially lost at those regions that lose chromatin accessibility. 66

Figure 2.6: GAF likely functions with ZLD to define chromatin accessibility in the early embryo. 67

Figure 2.7: Model for the role of ZLD in defining chromatin accessibility at the MZT. 68

Sup. Figure 2.1: ZLD protein is not detected in stage 5 embryos depleted of maternal zld....................69

Sup. Figure 2.2: FAIRE-seq data from 2-3 hour yw embryos are similar to previously published FAIRE-seq data from 2-4 hour OreR embryos. 70

Sup. Figure 2.3: Levels of ZLD binding, as assayed by ChIP-seq signal, are correlated with a dependence on ZLD for chromatin accessibility....................71

Sup. Figure 2.4: FAIRE-seq data from 2-3 hour yw embryos are similar to previously published FAIRE-seq data from 2-4 hour OreR embryos. 72

Sup. Figure 2.5: Early RNA Polymerase II binding is associated with promoters that require ZLD for chromatin accessibility. 73

Sup. Figure 2.6: Histone modifications associated with regions of open chromatin. 74

Sup. Figure 2.7: ZLD binding, and not a dependence on ZLD for chromatin accessibility, correlates with the mean number of transcription factors bound to the regions, regardless of genomic location. 75

Figure 3.1: MS2 reporters enable detection of nascent transcription in live embryos. 95

Figure 3.2: Expression patterns of tll reporters vary within each reporter class. 96

Sup. Figure 3.1: tll reporter expression is detected in nuclear cycle 12 (nc12) and expanded in nuclear cycle 14 (nc14). 97

Figure 4.1: ZLD binds nucleosomes containing its DNA-binding motif. 115

Figure 4.2: The interaction of ZLD with nucleosomes is largely sequence specific. 116

Sup. Figure 4.1: Nucleosomes are bound by multiple polypeptides at high protein concentrations. 117

Sup. Figure 4.2: ZLD binds nucleosomes under a second set of reaction conditions. 118

Sup. Figure 4.3: ZLD binding to the scute probe is sequence specific. 119

List of Abbreviations

3C, chromatin conformation capture

BCD, Bicoid

ChIP, chromatin immunoprecipitation

DL, Dorsal

EMSA, electrophoretic mobility shift assay

ESC, embryonic stem cell

FAIRE, formaldehyde-assisted isolation of regulatory elements

FISH, fluorescent *in situ* hybridization

GAF, GAGA Factor

HOT regions, highly occupied target regions

iPSC, induced pluripotent stem cell

MZT, maternal-to-zygotic transition

N/C ratio, nuclear-cytoplasmic ratio

Pol II, RNA polymerase II

TADs, topologically associating domains

TBP, TATA-binding protein

TE, transposable element

TF, transcription factor

TWI, Twist

ZGA, zygotic genome activation

ZLD, Zelda

ZLD^{DBD}, Zelda DNA-binding domain

Chapter 1 - Mechanisms of zygotic genome activation

The work presented in this chapter has been submitted for publication at *Nature Reviews Genetics*.

Statement of contribution: This review was co-written by Melissa Harrison and me.

Abstract

Following fertilization, the zygotic genome remains transcriptionally quiescent while the zygote is reprogrammed to a totipotent state. Through a process known as the maternal-to-zygotic transition, transcription of the zygotic genome is gradually activated, allowing zygotic gene products to replace the maternal supply that initiated development. This essential transition has been broadly characterized through decades of research in several model organisms. Yet, we still lack a full, mechanistic understanding of how genome activation is executed. Recent work highlights the central role that transcription factors play in this process and suggests that these genome-activating factors may coordinate transcriptional activation with additional developmental changes.

Introduction

Over half a century ago, John Gurdon generated the first cloned frog by transplanting the nucleus of a somatic cell into an enucleated egg (Gurdon 1962). This ground-breaking experiment highlighted the reprogramming power of the egg, whose cytoplasm contains all the factors needed to revert a differentiated genome to a totipotent state. In the first hours of animal development, these factors reset the chromatin contributed by the egg and sperm to ensure that the newly formed zygotic genome can give rise to the diverse cell types that form the adult animal.

The genome remains transcriptionally silent while this reprogramming takes place (Newport and Kirschner 1982b). For development to proceed in the absence of transcription, the embryo relies on maternally supplied RNAs and proteins that have been stockpiled in the oocyte. Ultimately, for the embryo to continue developing beyond this initial phase, the zygotic genome must be expressed. Transcriptional control is passed to the zygote through a process known as the maternal-to-zygotic transition (MZT), during which the degradation of maternal products is coordinated with zygotic genome activation (ZGA) (Tadros and Lipshitz 2009). There are many parallels between this rapid and efficient developmental transition and experimental reprogramming in culture (Yartseva and Giraldez 2015). Thus, understanding the mechanisms that underly genome activation will inform our efforts to manipulate cell identity.

The MZT is conserved across the kingdom Animalia. In many animals, this massive transcriptional shift coincides with changes to the cell cycle, referred to as the mid-blastula transition (Tadros and Lipshitz 2009). Prior to ZGA, these embryos undergo rapid cellular divisions, switching between DNA replication (S phase) and division (mitosis) without pausing in gap phase (Yuan et al. 2016). As the MZT nears completion, the cell cycle slows, and gap phase is introduced. Collectively, these changes prepare the embryo for gastrulation, during which cells begin to migrate and differentiate into the major germ layers of the animal (Tadros and Lipshitz 2009).

In recent years, technological innovations have dramatically improved our ability to interrogate the processes that govern ZGA. New live-imaging methods allow the expression of individual genes to be tracked with unprecedented spatial and temporal precision. Likewise, the increased accessibility of high-throughput sequencing has fostered numerous genome-profiling assays. With the advent of single-cell and low-input sequencing methods, we continue to improve the resolution at which we examine both ZGA and the chromatin remodelling that accompanies it.

In this review, we survey the mechanisms that regulate genome activation, highlighting the transcription factors at the center of this process. We will also discuss the dynamics of ZGA, models of ZGA timing, and the interplay between ZGA and chromatin remodelling. For a more comprehensive discussion of maternal RNA degradation (Yartseva and Giraldez 2015) or cell-cycle remodelling (Yuan et al. 2016) during the MZT, we direct you to these current reviews.

Dynamics of genome activation

ZGA is not a single event, but rather a period over which transcription is gradually activated (Collart et al. 2014; Harvey et al. 2013; Lott et al. 2011). It is characterized by two transcriptional bursts: a minor wave that occurs during the early cleavage divisions and a major wave that coincides with the first cell-cycle pause in many species (Tadros and Lipshitz 2009). However, the timing of these waves varies widely across animals. Rapidly developing species like worms (*C. elegans*), frogs (*Xenopis laevis*), fish (*Danio rerio*), and flies (*Drosophila melanogaster*) complete the MZT and enter gastrulation only a few hours after fertilization, while slower developing mammals, like mice (*Mus musculus*) and humans, take days (Figure 1.1) (Tadros and Lipshitz 2009).

This division between rapid and slow developing animals is thought to stem from the nature of the egg (Yuan et al. 2016). An egg that is abandoned in a predator-laden environment has different needs than one that is protected in a uterus and externally provided with nutrients. Of course,

important differences exist within these simplistic categories. In the fly embryo, for instance, cytokinesis is deferred in favor of speed. Here, nuclei divide in a common cytoplasm until the end of the MZT, when cellularization takes place (Tadros and Lipshitz 2009). While the particular needs of an egg dictate different modes of embryogenesis, many fundamental processes are conserved. Despite these differences, in all animals the accurate timing of ZGA onset depends on several intricately coordinated mechanisms.

Mechanisms that control ZGA timing

The first major model to explain the timing of ZGA was based on the idea that the early cell divisions served as a timer. Since the volume of the embryo remains constant through the MZT, the cell volume is cut in half with each division and the ratio of nuclear material to cytoplasm (the nucleocytoplasmic ratio, N/C) increases exponentially. In a landmark study, Newport and Kirschner discovered that ZGA takes place two cell cycles earlier in polyspermic frog embryos as compared to embryos fertilized by a single sperm (Newport and Kirschner 1982a). They proposed that the embryo's increasing supply of nuclear material could titrate a maternally supplied repressor to gradually relieve transcriptional repression (Figure 1.2A). This model was supported by a series of additional experiments in frogs. Here, experimentally increasing the DNA content of the embryo by injecting either plasmid DNA (Prioleau et al. 1994) or factors that increase nuclear size (Jevtic and Levy 2015) results in premature ZGA, while increasing the cytoplasmic volume using cleavage inhibitors or physical constriction results in delayed ZGA (Newport and Kirschner 1982a). Similarly, transcription is activated early in mutant zebrafish containing dense patches of DNA due to defective chromosome segregation (Dekens et al. 2003). However, the universality of this model was challenged by the discovery that haploid fly embryos execute ZGA with proper timing (Edgar, Kiehle, and Schubiger 1986). A detailed dissection using compound chromosomes revealed that only a subset of fly genes responds to the N/C ratio (X. Lu et al.

2009). Similarly, while this ratio influences morphological changes in mice, it does not affect transcription on a global scale (Lee et al. 2001).

The counterpart to this model, known as the maternal-clock model, posits that fertilization or egg activation initiates a biochemical cascade that serves as a molecular timer. The embryo receives many factors as maternally deposited mRNAs that are often held in a dormant state by inhibitory RNA-binding proteins. Even after this repression is released, it takes time for these transcripts to be polyadenylated, translated, and transported to the nucleus. Thus, in theory, accumulation of any essential factor required to either activate transcription or alleviate repression could contribute to ZGA timing (Figure 1.2B). While few such factors have been identified to date, canonical examples include components of the basal transcriptional machinery (discussed below) (Guven-Ozkan et al. 2008; Veenstra et al. 1999) and the fly maternal clearance factor Smaug (Benoit et al. 2009; Tadros et al. 2007).

These models are not mutually exclusive, and it is increasingly clear that multiple processes coordinate to regulate ZGA timing. In the following subsections, we discuss how titration of maternal repressors, cell-cycle lengthening, and accumulation of transcriptional activators collectively create a permissive environment for ZGA.

Titration of maternal repressors

The N/C ratio model relies of the existence of a titratable, maternally supplied repressor. While gene-specific repressors have long been known (Stancheva and Meehan 2000; Ruzov et al. 2004; Pritchard and Schubiger 1996), the idea that histones may serve as a global repressor of ZGA recently gained direct support. Histones bind ubiquitously across the genome and inhibit transcription by limiting the ability of regulators to access DNA, making them well suited for global repression. While the density of histones bound to DNA remains steady through the early cleavage cycles, a drop in the concentration

of unbound histones coincides with ZGA in zebrafish (Joseph et al. 2017). Thus, a model in which free histones compete with activators for DNA binding was proposed. In this model, excess histones initially buffer against premature transcription, but as activators accumulate and histone levels subside, the balance shifts to favor activation. Supporting this model, manipulation of histone levels has been shown to affect the timing of ZGA in both frogs and zebrafish (Joseph et al. 2017; Amodeo et al. 2015)

Cell-cycle lengthening

Since transcription is largely shut off when nuclei enter mitosis (Gottesfeld and Forbes 1997), the rapid cleavage cycles of the early embryo leave only brief windows of time for transcription (Rothe et al. 1992; Shermoen and O'Farrell 1991). Fittingly, genes expressed very early in development tend to be short and lack introns (Heyn et al. 2014; De Renzis et al. 2007; Rothe et al. 1992; Swinburne and Silver 2008). The transcriptional output of the genome gradually increases as the cell cycle slows, culminating in the major wave of ZGA as cells enter the first G2 pause.

These observations suggest that cell-cycle slowing could set the pace of genome activation (Figure 1.2C). Work in frogs lends credence to this model. Here, elongating the early cell cycle results in premature transcription (Collart et al. 2013; Kimelman et al. 1987). However, in zebrafish and flies, blocking cell-cycle progression does not affect ZGA timing (Farrell and O'Farrell 2013; Zhang et al. 2014; McCleland and O'Farrell 2008). In flies, the relationship between these processes is more complicated as zygotically expressed inhibitors are required to slow the cycle (Sung et al. 2017). Thus, while a rapid cell cycle places limits on transcription, cell-cycle slowing is interconnected with other mechanisms of ZGA regulation in at least some species.

Accumulation of transcriptional activators

In a third model of transcriptional silencing, the early embryo lacks an essential factor that must accumulate to enable transcription. A classic example of a factor that is rate-limiting for ZGA is TATA binding protein (TBP), a general transcription factor that promotes preinitiation complex formation as part of the TFIID complex (Hahn 2004). In frogs, translation of TBP is upregulated immediately before the major onset of transcription and precocious activation of TBP results in early ZGA (G. J. Veenstra, Destree, and Wolffe 1999). In worms, another component of TFIID, TAF4, is sequestered in the cytoplasm by repressor binding until a phosphorylation cascade triggered by fertilization results in its timely release and onset of ZGA (Guven-Ozkan et al. 2008a). However, a molecular timer based on basal transcription factors does not explain how the appropriate subset of genes is selected for activation. For this purpose, the embryo requires sequence-specific genome activators, which we will return to in the final section of this review.

Chromatin reprogramming during ZGA

In eukaryotes, the genome requires considerable compaction to fit inside the nucleus. DNA is spooled around octamers of histones to form nucleosomes, which are coiled into fibers and looped into higher-order structures (C. Jiang and Pugh 2009). This organization restricts the ability of the transcriptional machinery to access the DNA, and thus, plays a central role in gene regulation.

During the first hours of development, the chromatin is reprogrammed to a naive, globally accessible state. In this section, we discuss the multiple levels of chromatin reorganization that underly this developmental transition (Figure 1.3).

DNA Methylation

Methylation of cytosine to 5-methylcytosine promotes transcriptional silencing during processes such as genomic imprinting (E. Li, Beard, and Jaenisch 1993) and X-chromosome inactivation (Panning 2008). Evidence that human and mouse genomes undergo global demethylation prior to ZGA suggests that this mark could also contribute to the transcriptional silence of the early embryo (Guo et al. 2014; Shen et al. 2014). By contrast, both frog and zebrafish genomes remains heavily methylated through early development (G. J. C. Veenstra and Wolffe 2001; Bogdanovic et al. 2011; Potok et al. 2013; L. Jiang et al. 2013). Recent work in zebrafish suggests that DNA methylation may be involved in recruitment of the genome-activating transcription factors Nanog and Pou5f3 (G. Liu et al. 2018; Yin et al. 2017). However, while methylated promoters are robustly transcribed during ZGA in frogs (Bogdanovic et al. 2011), work in zebrafish suggests that DNA methylation serves to prevent precocious expression during ZGA (Potok et al. 2013). Further, the genomes of worms and flies possess limited amounts of DNA methylation (Takayama et al. 2014; Kelly 2014). Thus, while the role of this mark in ZGA remains unclear and appears to be species dependent, DNA methylation could play a part in priming genes for activation.

Histone modifications

Post-translational modifications chemically alter histone tails in ways that impact nucleosome stability and the recruitment of transcriptional regulators (Bannister and Kouzarides 2011). Chromatin immunoprecipitation coupled with high-throughput sequencing (ChIP-seq) time course data have revealed widespread changes in the abundance of these marks over the MZT. While work in flies suggests that histone acetylation may play a role early gene expression (X.-Y. Li et al. 2014), histone methylation has been more widely investigated across species. Trimethylation of histone H3 at lysine 4 (H3K4me3) is a canonical mark of activation that is often found at the transcription start site (TSS) of genes (Eissenberg and Shilatifard 2010). In zebrafish, H3K4me3 is detected at many promoters prior to

ZGA and appears to poise genes for activation (Vastenhouw et al. 2010). However, in frogs and flies, few promoters are marked with methylation prior to ZGA, indicating that early transcription can occur in the absence of this chromatin signature (X.-Y. Li et al. 2014; Akkers et al. 2009; Blythe et al. 2010). Despite some differences, a dramatic increase in histone methylation accompanies the major onset of transcription in all three of these species (Lindeman et al. 2011; Vastenhouw et al. 2010; X.-Y. Li et al. 2014; Akkers et al. 2009). This trend appears to be reversed in mice, where early embryos possess unusual, broad (5- to 10-kb) domains of H3K4me3. By the major wave of ZGA, these domains are largely restricted to conventional, transcription start site (TSS)-associated peaks (Dahl et al. 2016; X. Liu et al. 2016; B. Zhang et al. 2016). Notably, knockdown of the histone demethylases responsible for this pruning causes developmental arrest and downregulation of many ZGA genes (X. Liu et al. 2016; Dahl et al. 2016), suggesting that these broad domains could play a role in preventing premature transcription prior to ZGA. While there are some interspecies differences, all examined genomes exhibit global shifts in their histone-modification profile as the embryo proceeds through the MZT.

Histone variants

The constitution of a nucleosome can also be altered by replacing canonical histone proteins with histone variants (Weber and Henikoff 2014). Histone substitutions serve at least two purposes during ZGA: incorporation of permissive histone variants, such as H2A.Z or H3.3, marks specific genes for activation (Whittle et al. 2008; Lin, Conti, and Ramalho-Santos 2013), while removal of oocyte- or embryo-specific variants appears to more globally license the genome for activation. In flies, frogs, and mice, embryonic variants of linker histone H1 are replaced with their somatic counterparts at the major onset of ZGA (Perez-Montero et al. 2013; Smith, Dworkin-Rastl, and Dworkin 1988; Fu et al. 2003). These embryonic H1 variants are predicted to form less stable nucleosomes and may, therefore, contribute to the naive chromatin environment of the early embryo. The importance of this variant has been

demonstrated in flies, where a lack of embryonic H1 results in premature transcription and developmental arrest (Perez-Montero et al. 2013).

Nucleosome positioning

In zebrafish, reorganization of nucleosome positioning coincides with widespread genome activation. Here, well-positioned nucleosome arrays appear at the promoters of genes, marking many for future activation (Y. Zhang et al. 2014). Recently, chromatin accessibility profiling has revealed that defined regulatory regions are established concomitantly with transcriptional activation in flies (Blythe and Wieschaus 2016), zebrafish (G. Liu et al. 2018), mice (Wu et al. 2016; F. Lu et al. 2017), and humans (Gao et al. 2018; L. Li et al. 2018; Wu et al. 2018). Early mouse embryos possess broad regions of open chromatin that are narrowed down to mark promoters by the major wave of ZGA (Wu et al. 2016). Intriguingly, these broad regions often encompass transposable elements that are transiently expressed during ZGA in mice (Svoboda et al. 2004; Wu et al. 2016). While this burst of transposable element expression was initially assumed to be a side effect of global chromatin accessibility, recent work suggests that it may contribute to both chromatin opening (Jachowicz et al. 2017) and early gene expression (Macfarlan et al. 2012; Ishiuchi et al. 2015).

Chromatin domains

Condensed chromatin fibers form loops that are clustered together into topologically associated domains (TADs) (Dixon, Gorkin, and Ren 2016). By keeping specific promoters in the proximity of enhancers or silencers, TADs play an important role in transcriptional regulation (Dekker and Mirny 2016). Recently, chromosome conformation capture (3C) techniques, such as HiC, have been used to assess changes in these three-dimensional chromatin contacts over the course of the MZT. In fly embryos, TAD boundaries are gradually established in concert with gene activation (Hug et al. 2017).

Likewise, in mice, the genome lacks tightly defined TADs until well after ZGA (Ke et al. 2017; Du et al. 2017). Together, these studies support the notion that metazoan genomes are largely unstructured during the initial phase of embryogenesis.

Transcription factors direct ZGA

By binding to specific DNA sequences, transcription factors direct the transcriptional machinery to particular genes. This specificity is of critical importance during ZGA, when it is estimated that 12-15% of the genome is transcriptionally activated (Tadros and Lipshitz 2009).

The first identified master regulator of ZGA was discovered in flies. This transcription factor, Zelda (ZLD), is maternally deposited as mRNA and is translationally upregulated in the hour leading up to ZGA (Harrison et al. 2011; Nien et al. 2011; Harrison, Botchan, and Cline 2010). ZLD is required for the expression of hundreds of genes during both the major and minor waves of ZGA, and without this essential factor, fly embryos die before completing the MZT (Liang et al. 2008; Nien et al. 2011). Since ZLD orthologs are limited to the insect clade, it was initially unclear whether there were factors that function analogously in vertebrates. However, two independent studies identified Nanog, SoxB1, and Pou5f3 (Oct4) as activators of zebrafish ZGA (Leichsenring et al. 2013; M. T. Lee et al. 2013). Interest in these factors was incited by their identification as the most highly translated transcription factors in zebrafish embryos immediately after fertilization (M. T. Lee et al. 2013). These factors are homologues of the mammalian "pluripotency factors" which are known for their ability to reprogram differentiated cells to a stem cell-like state (Takahashi and Yamanaka 2016). Importantly, while they are not phylogenetically related to ZLD, these transcription factors share several functional characteristics. Like ZLD, the zebrafish activators are translated early in development allowing them to activate the earliest expressed genes and poise hundreds of additional genes for activation during the major onset of ZGA (Leichsenring et al. 2013; Liang et al. 2008; Harrison et al. 2011; Nien et al. 2011).

Advances in low-input sequencing methods have recently led to the first discoveries of mammalian genome activators. One study in human embryos found that Oct4-binding sites are enriched in accessible regulatory regions during ZGA, and knockdown of this factor results in downregulation of hundreds of ZGA genes (Gao et al. 2018). However, Oct4 motifs do not appear to be enriched in open regions identified by other chromatin profiling methods, so the role of Oct4 in human ZGA remains contested (Wu et al. 2018; L. Li et al. 2018). Importantly, this role for Oct4 is not shared in mice, where it is not required until much later in embryogenesis (Le Bin et al. 2014; Gao et al. 2018). In mice, a lesser known pluripotency factor, Nfy, is similarly enriched in open chromatin during ZGA and is required for expression of many ZGA genes (F. Lu et al. 2016). Humans and mice do share at least one family of genome activators known as the DUX transcription factors. DUX (mouse) and DUX4 (human) activate hundreds of ZGA genes in these species, including many MERVL transposable elements (De Iaco et al. 2017; Hendrickson et al. 2017; Whiddon et al. 2017). DUX genes are zygotically expressed as part of initial wave of ZGA and the mechanisms by which they are activated remain unknown (Hendrickson et al. 2017). Thus, while some genome activators are maternally deposited and translationally upregulated, others are regulated at the level of transcription. In all cases, genome activators are not present at fertilization, suggesting their activity must be carefully controlled to prevent premature transcriptional activation.

Potential mechanisms of genome activators

The widespread impact of genome activators may stem from their ability to regulate chromatin structure. ZLD-binding sites are strongly correlated with regions of chromatin accessibility during the major wave of ZGA in flies (Harrison et al. 2011), and embryos that lack ZLD lose accessibility at many of these sites (Schulz et al. 2015; Sun et al. 2015). ZLD binding reduces the local nucleosome occupancy and may promote histone acetylation at the regulatory regions of early expressed genes (Sun et al. 2015;

Schulz et al. 2015; Li et al. 2014). Likewise, Nanog and Pou5f3 are required for chromatin accessibility at enhancers of developmental genes during ZGA (G. Liu et al. 2018; Veil et al. 2018). Notably, recent work suggests that the binding of all three zebrafish activators (Nanog, Pou5f3, and SoxB1) is required to maintain nucleosome-free sites post-ZGA (Veil et al. 2018). While knockdown experiments demonstrate that Nfy is needed for open chromatin during mouse ZGA (F. Lu et al. 2016), DUX/DUX4 have only been shown to maintain sites of open chromatin in mouse embryonic stem cells (ESCs) (Hendrickson et al. 2017; Oldfield et al. 2014). Fittingly, the Nfy complex contains two histone-fold domains that interact with DNA in a histone-like manner (Nardini et al. 2013), allowing the complex to displace nucleosomes from DNA *in vitro* (Coustry et al. 2001). DUX4 can similarly displace histones in myoblasts and recruits the histone acetyltransferases p300/CBP to establish activating histone modifications (Choi et al. 2016). Finally, while there is evidence that many of these factors function as "pioneer factors", the case is arguably strongest for mammalian Oct4 which opens chromatin to promote induced pluripotent stem cell (iPSC) reprogramming (Soufi et al. 2012; Soufi et al. 2015). Oct4 binds nucleosomal DNA both *in vitro* and in fibroblasts (Zaret and Carroll 2011), and recent work suggests that both Oct4 and Pou5f3 may function, in part, by recruiting the chromatin remodeler BRG1 (Smarca4a in zebrafish) to stabilize nucleosome positioning (King and Klose 2017; G. Liu et al. 2018). Thus, defining regions of accessible chromatin is a shared function of genome activators.

Genome activators may also influence the establishment of higher-order chromatin structure. While TADs remain uncharted in human and zebrafish embryos, Oct4 and Sox2 have been shown to bind chromatin reorganization hotspots during iPSC generation and influence insulation strength at TAD borders (Stadhouders et al. 2018). In zebrafish embryos, these activators are enriched at sites bound by the architectural protein cohesin (Meier et al. 2018), a key player in both ZGA and TAD formation (Dixon, Gorkin, and Ren 2016). Recent work suggests that ZLD may also contribute to TAD boundary insulation at a subset of sites during ZGA (Hug et al. 2017). These observations raise the possibility that genome

activators help shape 3D genome reorganization by directing the binding of architectural proteins during ZGA.

Recently, single-molecule imaging has revealed that ZLD promotes formation of transient "hubs" of another transcription factor called Bicoid. Confinement of Bicoid to these hubs creates sites of high local concentration that potentiate its binding to DNA (Mir et al. 2017). Similar methods have revealed that Sox2 forms clusters of enhancers in ESCs (Z. Liu et al. 2014). Intriguingly, the transcriptional activation domains of both ZLD and Sox2 are predicted to be largely unstructured, which may contribute to this clustering function (Hamm, Bondra, and Harrison 2015; Z. Liu et al. 2014). It is tempting to speculate that these transcription factors could broadly influence gene expression by forming hubs at chromatin boundaries. However, the relationship between transcriptional hubs and TADs remains to be defined.

Misexpression of genome activators in development and disease

Given the potent effects of genome activators on gene expression, it might come as no surprise that misexpression of these factors has serious consequences for the embryo. In some cases, these factors are essential for development. For instance, both lack of ZLD and excessive ZLD activity are lethal to the fly embryo (Liang et al. 2008; Staudt et al. 2006; Hamm et al. 2017), demonstrating the need for precise regulation of this protein. In zebrafish, mutants that lack both maternal and zygotic *pou5f3* display numerous developmental defects including delayed gastrulation and an inability to form endoderm (Onichtchouk and Driever 2016; Lunde, Belting, and Driever 2004). However, the maternal and zygotic functions of this protein overlap such that maternal-only *pou5f3* mutants develop normally (Onichtchouk and Driever 2016). Further, zebrafish ZGA is only severely disrupted when at least two of the known genome activators are depleted in combination (M. T. Lee et al. 2013). Thus, some genome activators function redundantly with other transcription factors or with their zygotic counterparts.

Long before the DUX proteins were implicated in ZGA, they were known for their role in facioscapulohumeral dystrophy (FSHD), an untreatable form of muscular dystrophy that progresses from the face to the lower limbs (Vanderplanck et al. 2011). This disease is caused by misexpression of DUX4 in skeletal muscle cells, where it activates aberrant expression of germline- and stem cell-associated genes (Lemmers et al. 2010; Geng et al. 2012; Young et al. 2013). Activation of the pluripotency gene network is also a hallmark of cancer, where it may facilitate the proliferative potential of these cells (Iglesias et al. 2017). Hence, the ability of genome activators to reprogram cells towards pluripotency also makes them potent drivers of tumor development. Fittingly, aberrant expression of Oct4, Nanog, and Sox2 have been associated with numerous forms of cancer (A. Liu et al. 2013).

Conclusions and perspective

In the transcriptional silence that follows fertilization, the genome is reprogrammed to prepare the embryo to give rise to a new animal. The transition from silence to widespread gene expression requires precise regulation. This is accomplished through several coordinated mechanisms, in which transcription factors play a central role. After fertilization, translational upregulation promotes accumulation of genome-activating transcription factors. In one model of ZGA timing, these factors begin to successfully compete with histones for DNA binding after reaching a critical threshold. Competition of this nature would integrate a readout of the N/C-ratio (histone levels) with a readout of the embryo's molecular clock (activator levels). However, this tidy model is likely oversimplified. Given that histones are newly synthesized in the early embryo (Adamson and Woodland 1974), their levels should also be regulated by molecular clock-based mechanisms. Regardless, competition for access to regulatory regions of individual genes would help explain differences in the timing of gene activation.

The same mechanisms that drive this transcriptional shift in the embryo function in other cellular reprogramming contexts, including the creation of iPSCs. Transcriptional profiling has revealed a

remarkable overlap in the gene networks activated during iPSC generation and ZGA (Wang et al. 2018; Macfarlan et al. 2012). This overlap can be explained, in part, by the discovery that known master regulators of pluripotency also serve as genome activators in the embryo. These transcription factors direct chromatin remodeling in both of these reprogramming contexts, helping to erase the previous cell identity while creating a new one. These parallels demonstrate the ability of work in stem cell models to inform our understanding of the embryogenesis, and vice versa. The application of new technologies to these important biological questions will lead to more precise understanding of the mechanisms that drive these essential transitions.

The advent of single-cell-omics has unmasked cell-to-cell heterogeneity and led to the identification of hundreds of uncharacterized ZGA genes (White et al. 2017). Recently, these methods have been combined with advanced computational pipelines to generate transcriptome atlases of frog and zebrafish embryogenesis (Briggs et al. 2018; Farrell et al. 2018; Wagner et al. 2018), allowing the fate of individual cells to be traced from pluripotency to specification. Examining the branch points in these differentiation pathways will lead to the identification of novel lineage-defining transcription factors and expand our understanding of how such factors function. Additionally, new low-input sequencing techniques have allowed us to profile the transcriptome (Yan et al. 2013; Hendrickson et al. 2017), methylome (Guo et al. 2014), and chromatin-accessibility landscape (Gao et al. 2018) of embryos despite the limited material available for these studies. The first maps of early human embryos have identified fundamental differences in the regulation of human and mouse ZGA (Gao et al. 2018). These findings will allow decades of research in animal models to be more aptly applied to our own development.

Advances in imaging methods continue to improve the resolution with which we visualize ZGA. Single-molecule fluorescence *in situ* hybridization (FISH) allowed quantitative measurement of endogenous transcription, while live-cell imaging using fluorescently tagged RNA enabled simultaneous

detection of temporal and spatial expression patterns. These methods have been used to track the activation of individual genes during ZGA (Campbell et al. 2015), and have led to the discovery of subtle transcriptional phenomena, such as mitotic memory (Ferraro et al. 2016) and transcriptional bursting (Bothma et al. 2014; Senecal et al. 2014). CRISPR/Cas9 technology offers new ways to image specific transcripts in living embryos (Nelles et al. 2016) and can be used in combination with FISH (Deng et al. 2015) or to introduce RNA reporters at endogenous loci to help avoid potential artifacts of these methods. Increasingly sensitive, fast, and precise imaging technologies, such as super-resolution microscopy and light-sheet microscopy, have enabled researchers to identify new structures and processes by imaging living embryos as they develop (Strom et al. 2017; Mir et al. 2017). These imaging strategies can be combined with optogenetic approaches to allow researchers to manipulate protein activity with unprecedented precision (Huang et al. 2017).

Together, new and forthcoming tools will enable researchers to address many outstanding questions about the mechanisms controlling ZGA. The causal relationships between the major processes that accompany ZGA remain unclear and the mechanisms by which genome activators function have yet to be defined. For instance, while TADs form independently of transcription (Ke et al. 2017; Du et al. 2017; Hug et al. 2017) and the artificial creation of chromatin loops drives gene expression in at least some contexts (Deng et al. 2014), it remains to be determined whether TAD formation is required for ZGA. Likewise, it is uncertain whether other chromatin changes are required for transcription or whether they are simply the biproduct of transcriptional activity (Henikoff and Shilatifard 2011). Defining how transcription factors function during these initial stages of development will uncover the connections between chromatin remodelling ,the mechanisms that govern ZGA, and the fundamental features required for developmental reprogramming.

Figures

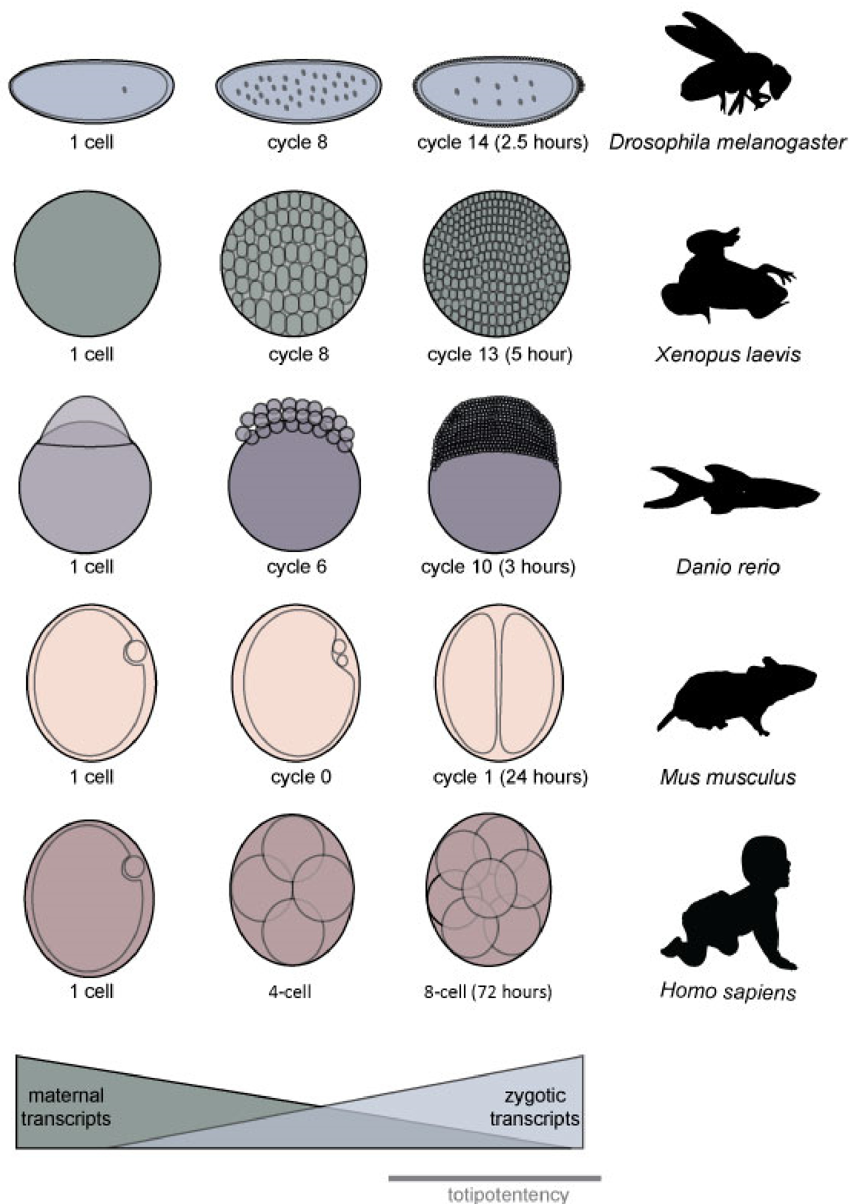


Figure 1.1: Zygotic genome activation is conserved across animals.

Key stages of zygotic genome activation are outlined for five model species, indicated on the right. Starting from the left, the initial "1 cell" stage, an intermediate stage, and a final stage representing the major wave of ZGA are depicted. The absolute time (in hours post fertilization) is indicated under the final stage. Below, the clearance of maternal products and activation of zygotic transcription are illustrated as triangles, and the period of totipotency is indicated with a bar.

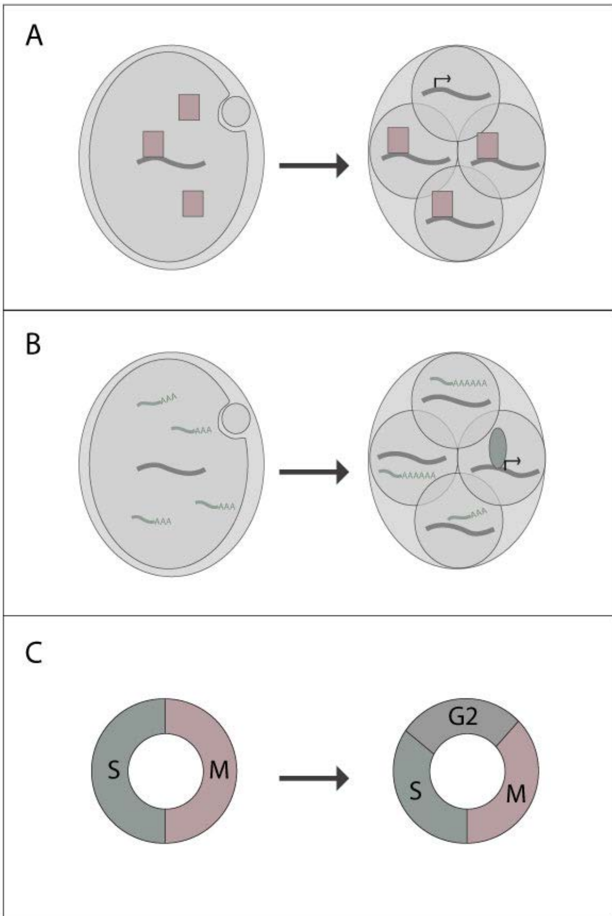


Figure 1.2: Several mechanisms contribute to the timing of zygotic genome activation

(A) A maternally supplied repressor (red square) prevents transcription in the early embryo. As the ratio of nuclei to cytoplasm increases with each cell division, the repressor is titrated by the growing supply of nuclear material and transcription initiates. **(B)** The early embryo lacks a key transcriptional activator (green oval). Polyadenylation and translation of maternally supplied mRNA leads to its accumulation, enabling the expression of target genes. **(C)** The rapid, early cell cycles consist of only a DNA replication (S) phase and mitosis (M). By the major onset of ZGA, the cell cycle has slowed and a gap phase (G2) is introduced, reducing the time restraint initially placed on transcription.

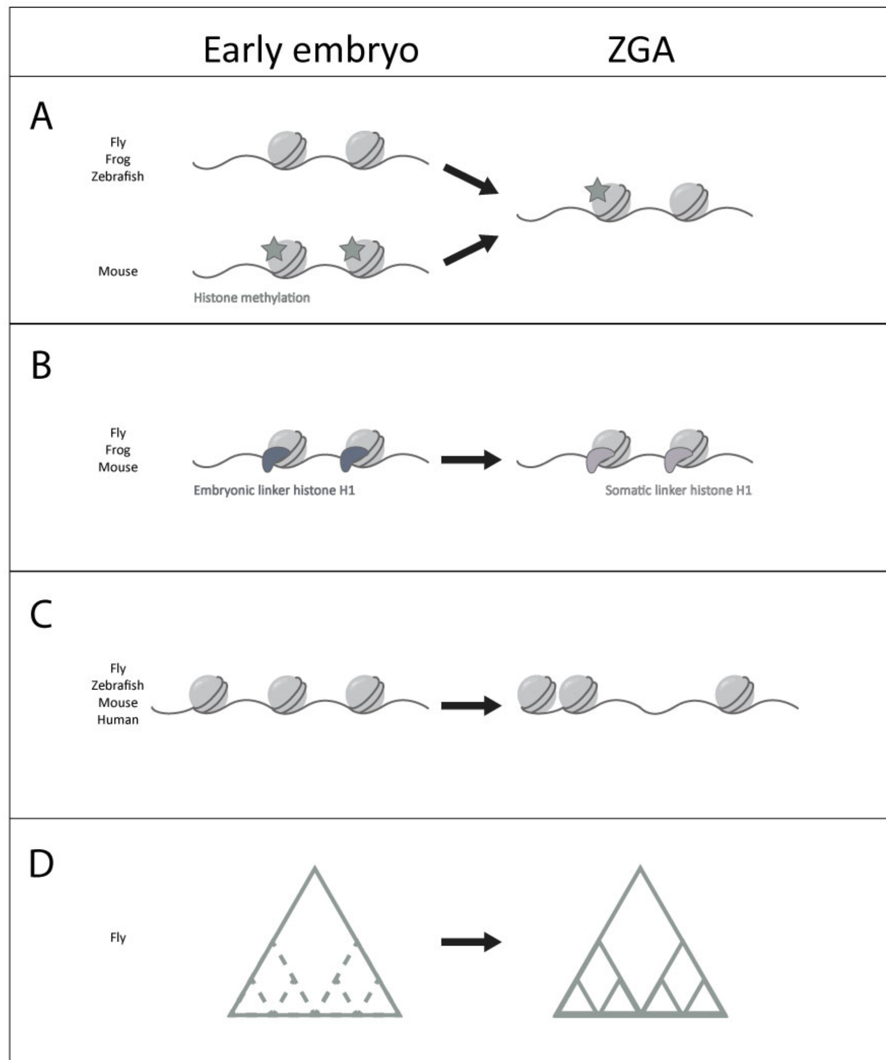


Figure 1.3. Chromatin undergoes reorganization during zygotic genome activation.

(A) In zebrafish, frogs, and flies, a dramatic increase in histone methylation accompanies the major onset of ZGA (top). In mice, unusual, broad methylation (H3K4me3) domains are pruned down into TSS-associated peaks for ZGA (bottom). **(B)** In flies, frogs, and mice, embryonic variants of linker histone H1 are replaced with their somatic counterparts at the major onset of ZGA. **(C)** Defined regulatory elements, characterized by open chromatin, are established during ZGA in flies, zebrafish, mice, and humans. **(D)** In flies, the boundaries of topologically associated domains (TADs) are established concurrently with ZGA.

References

Adamson, E D, and H R Woodland. 1974. "Histone Synthesis in Early Amphibian Development: Histone and DNA Syntheses Are Not Co-Ordinated." Journal of Molecular Biology 88 (2): 263–85. http://www.ncbi.nlm.nih.gov/pubmed/4476799.

Akkers, Robert C., Simon J. van Heeringen, Ulrike G. Jacobi, Eva M. Janssen-Megens, Kees-Jan Françoijs, Hendrik G. Stunnenberg, and Gert Jan C. Veenstra. 2009. "A Hierarchy of H3K4me3 and H3K27me3 Acquisition in Spatial Gene Regulation in Xenopus Embryos." Developmental Cell 17 (3): 425–34. https://doi.org/10.1016/j.devcel.2009.08.005.

Amodeo, Amanda A, David Jukam, Aaron F Straight, and Jan M Skotheim. 2015. "Histone Titration against the Genome Sets the DNA-to-Cytoplasm Threshold for the Xenopus Midblastula Transition." Proceedings of the National Academy of Sciences 112 (10): E1086–95. https://doi.org/10.1073/pnas.1413990112.

Bannister, A J, and T Kouzarides. 2011. "Regulation of Chromatin by Histone Modifications." Cell Res 21. https://doi.org/10.1038/cr.2011.22.

Benoit, B., C. H. He, F. Zhang, S. M. Votruba, W. Tadros, J. T. Westwood, C. A. Smibert, H. D. Lipshitz, and W. E. Theurkauf. 2009. "An Essential Role for the RNA-Binding Protein Smaug during the Drosophila Maternal-to-Zygotic Transition." Development 136 (6): 923–32. https://doi.org/10.1242/dev.031815.

Bin, G. C. Le, S. Munoz-Descalzo, A. Kurowski, H. Leitch, X. Lou, W. Mansfield, C. Etienne-Dumeau, et al. 2014. "Oct4 Is Required for Lineage Priming in the Developing Inner Cell Mass of the Mouse Blastocyst." Development 141 (5): 1001–10. https://doi.org/10.1242/dev.096875.

Blythe, Shelby A, Sang-Wook Cha, Emmanuel Tadjuidje, Janet Heasman, and Peter S Klein. 2010. "β-Catenin Primes Organizer Gene Expression by Recruiting a Histone H3 Arginine 8 Methyltransferase, Prmt2." Dev Cell 19 (2): 220–31. https://doi.org/http://dx.doi.org/10.1016/j.devcel.2010.07.007.

Blythe, Shelby A, and Eric F Wieschaus. 2016. "Establishment and Maintenance of Heritable Chromatin Structure during Early Drosophila Embryogenesis." Edited by Allan C Spradling. ELife 5. eLife Sciences Publications, Ltd: e20148. https://doi.org/10.7554/eLife.20148.

Bogdanovic, Ozren, Steven W Long, Simon J van Heeringen, Arie B Brinkman, Jose Luis Gómez-Skarmeta, Hendrik G Stunnenberg, Peter L Jones, and Gert Jan C Veenstra. 2011. "Temporal Uncoupling of the DNA Methylome and Transcriptional Repression during Embryogenesis." Genome Research 21 (8). Cold Spring Harbor Laboratory Press: 1313–27. https://doi.org/10.1101/gr.114843.110.

Bothma, Jacques P, Hernan G Garcia, Emilia Esposito, Gavin Schlissel, Thomas Gregor, and Michael Levine. 2014. "Dynamic Regulation of Eve Stripe 2 Expression Reveals Transcriptional Bursts in Living Drosophila Embryos." Proceedings of the National Academy of Sciences of the United States of America 111 (29). National Academy of Sciences: 10598–603. https://doi.org/10.1073/pnas.1410022111.

Briggs, James A, Caleb Weinreb, Daniel E Wagner, Sean Megason, Leonid Peshkin, Marc W Kirschner, and Allon M Klein. 2018. "The Dynamics of Gene Expression in Vertebrate Embryogenesis at Single-Cell Resolution." Science (New York, N.Y.), April. American Association for the Advancement of Science, eaar5780. https://doi.org/10.1126/science.aar5780.

Campbell, P. D., J. A. Chao, R. H. Singer, and F. L. Marlow. 2015. “Dynamic Visualization of Transcription and RNA Subcellular Localization in Zebrafish.” Development 142 (7): 1368–74. https://doi.org/10.1242/dev.118968.

Choi, Si Ho, Micah D. Gearhart, Ziyou Cui, Darko Bosnakovski, Minjee Kim, Natalie Schennum, and Michael Kyba. 2016. “DUX4 Recruits P300/CBP through Its C-Terminus and Induces Global H3K27 Acetylation Changes.” Nucleic Acids Research 44 (11): 5161–73. https://doi.org/10.1093/nar/gkw141.

Collart, Clara, George E Allen, Charles R Bradshaw, James C Smith, and Philip Zegerman. 2013. “Titration of Four Replication Factors Is Essential for the Xenopus Laevis Midblastula Transition.” Science (New York, N.Y.) 341 (6148). United States: 893–96. https://doi.org/10.1126/science.1241530.

Collart, Clara, Nick D. L. Owens, Leena Bhaw-Rosun, Brook Cooper, Elena De Domenico, Ilya Patrushev, Abdul K. Sesay, James N. Smith, James C. Smith, and Michael J. Gilchrist. 2014. “High-Resolution Analysis of Gene Activity during the Xenopus Mid-Blastula Transition.” Development 141 (9): 1927–39. https://doi.org/10.1242/dev.102012.

Coustry, Françoise, Qianghua Hu, Benoit de Crombrugghe, and Sankar N. Maity. 2001. “CBF/NF-Y Functions Both in Nucleosomal Disruption and Transcription Activation of the Chromatin-Assembled Topoisomerase IIα Promoter.” Journal of Biological Chemistry 276 (44): 40621–30. https://doi.org/10.1074/jbc.M106918200.

Dahl, John Arne, Inkyung Jung, Havard Aanes, Gareth D Greggains, Adeel Manaf, Mads Lerdrup, Guoqiang Li, et al. 2016. “Broad Histone H3K4me3 Domains in Mouse Oocytes Modulate Maternal-to-Zygotic Transition.” Nature 537 (7621). England: 548–52. https://doi.org/10.1038/nature19360.

Dekens, Marcus P S, Francisco J Pelegri, Hans-Martin Maischein, and Christiane Nusslein-Volhard. 2003. “The Maternal-Effect Gene Futile Cycle Is Essential for Pronuclear Congression and Mitotic Spindle Assembly in the Zebrafish Zygote.” Development (Cambridge, England) 130 (17). England: 3907–16.

Dekker, Job, and Leonid Mirny. 2016. “The 3D Genome as Moderator of Chromosomal Communication.” Cell 164 (6). Cell Press: 1110–21. https://doi.org/10.1016/J.CELL.2016.02.007.

Deng, Wulan, Jeremy W. Rupon, Ivan Krivega, Laura Breda, Irene Motta, Kristen S. Jahn, Andreas Reik, et al. 2014. “Reactivation of Developmentally Silenced Globin Genes by Forced Chromatin Looping.” Cell 158 (4): 849–60. https://doi.org/10.1016/j.cell.2014.05.050.

Deng, Wulan, Xinghua Shi, Robert Tjian, Timothée Lionnet, and Robert H Singer. 2015. “CASFISH: CRISPR/Cas9-Mediated in Situ Labeling of Genomic Loci in Fixed Cells.” Proceedings of the National Academy of Sciences of the United States of America 112 (38). National Academy of Sciences: 11870–75. https://doi.org/10.1073/pnas.1515692112.

Dixon, Jesse R., David U. Gorkin, and Bing Ren. 2016. “Chromatin Domains: The Unit of Chromosome Organization.” Molecular Cell 62 (5). Elsevier: 668–80. https://doi.org/10.1016/j.molcel.2016.05.018.

Du, Zhenhai, Hui Zheng, Bo Huang, Rui Ma, Jingyi Wu, Xianglin Zhang, Jing He, et al. 2017. “Allelic Reprogramming of 3D Chromatin Architecture during Early Mammalian Development.” Nature 2017 547:7662 547 (7662). Nature Publishing Group: 232. https://doi.org/10.1038/nature23263.

Edgar, B A, C P Kiehle, and G Schubiger. 1986. “Cell Cycle Control by the Nucleo-Cytoplasmic Ratio in Early Drosophila Development.” Cell 44 (2): 365–72.

http://www.ncbi.nlm.nih.gov/pubmed/3080248.

Eissenberg, Joel C, and Ali Shilatifard. 2010. "Histone H3 Lysine 4 (H3K4) Methylation in Development and Differentiation." Developmental Biology 339 (2). NIH Public Access: 240–49. https://doi.org/10.1016/j.ydbio.2009.08.017.

Farrell, Jeffrey A, and Patrick H O'Farrell. 2013. "Mechanism and Regulation of Cdc25/Twine Protein Destruction in Embryonic Cell-Cycle Remodeling." Current Biology : CB 23 (2). England: 118–26. https://doi.org/10.1016/j.cub.2012.11.036.

Farrell, Jeffrey A, Yiqun Wang, Samantha J Riesenfeld, Karthik Shekhar, Aviv Regev, and Alexander F Schier. 2018. "Single-Cell Reconstruction of Developmental Trajectories during Zebrafish Embryogenesis." Science (New York, N.Y.) 360 (6392). American Association for the Advancement of Science: eaar3131. https://doi.org/10.1126/science.aar3131.

Ferraro, Teresa, Emilia Esposito, Laure Mancini, Sam Ng, Tanguy Lucas, Mathieu Coppey, Nathalie Dostatni, Aleksandra M Walczak, Michael Levine, and Mounia Lagha. 2016. "Transcriptional Memory in the Drosophila Embryo." Current Biology : CB 26 (2). NIH Public Access: 212–18. https://doi.org/10.1016/j.cub.2015.11.058.

Fu, Germaine, Parinaz Ghadam, Allen Sirotkin, Saadi Khochbin, Arthur I. Skoultchi, and Hugh J. Clarke. 2003. "Mouse Oocytes and Early Embryos Express Multiple Histone H1 Subtypes1." Biology of Reproduction 68 (5): 1569–76. https://doi.org/10.1095/biolreprod.102.012336.

Gao, Lei, Keliang Wu, Zhenbo Liu, Xuelong Yao, Shenli Yuan, Wenrong Tao, Lizhi Yi, et al. 2018. "Chromatin Accessibility Landscape in Human Early Embryos and Its Association with Evolution." Cell 173 (1). Elsevier: 248–259.e15. https://doi.org/10.1016/j.cell.2018.02.028.

Geng, Linda N, Zizhen Yao, Lauren Snider, Abraham P Fong, Jennifer N Cech, Janet M Young, Silvere M van der Maarel, et al. 2012. "DUX4 Activates Germline Genes, Retroelements, and Immune Mediators: Implications for Facioscapulohumeral Dystrophy." Developmental Cell 22 (1). NIH Public Access: 38–51. https://doi.org/10.1016/j.devcel.2011.11.013.

Gottesfeld, J M, and D J Forbes. 1997. "Mitotic Repression of the Transcriptional Machinery." Trends in Biochemical Sciences 22 (6): 197–202. http://www.ncbi.nlm.nih.gov/pubmed/9204705.

Guo, Hongshan, Ping Zhu, Liying Yan, Rong Li, Boqiang Hu, Ying Lian, Jie Yan, et al. 2014. "The DNA Methylation Landscape of Human Early Embryos." Nature 511 (7511): 606–10. https://doi.org/10.1038/nature13544.

Gurdon, J. B. 1962. "The Developmental Capacity of Nuclei Taken from Intestinal Epithelium Cells of Feeding Tadpoles." Development 10 (4). http://dev.biologists.org/content/10/4/622.

Guven-Ozkan, Tugba, Yuichi Nishi, Scott M Robertson, and Rueyling Lin. 2008. "Global Transcriptional Repression in C. Elegans Germline Precursors by Regulated Sequestration of TAF-4." Cell 135 (1). NIH Public Access: 149–60. https://doi.org/10.1016/j.cell.2008.07.040.

Hahn, Steven. 2004. "Structure and Mechanism of the RNA Polymerase II Transcription Machinery." Nature Structural & Molecular Biology 11 (5). NIH Public Access: 394–403. https://doi.org/10.1038/nsmb763.

Hamm, Danielle C., Elizabeth D. Larson, Markus Nevil, Kelsey E. Marshall, Eliana R. Bondra, and Melissa M. Harrison. 2017. "A Conserved Maternal-Specific Repressive Domain in Zelda Revealed by Cas9-

Mediated Mutagenesis in Drosophila Melanogaster.” Edited by Michael B. Eisen. PLOS Genetics 13 (12). Public Library of Science: e1007120. https://doi.org/10.1371/journal.pgen.1007120.

Hamm, Danielle C, Eliana R Bondra, and Melissa M Harrison. 2015. “Transcriptional Activation Is a Conserved Feature of the Early Embryonic Factor Zelda That Requires a Cluster of Four Zinc Fingers for DNA Binding and a Low-Complexity Activation Domain.” The Journal of Biological Chemistry 290 (6). American Society for Biochemistry and Molecular Biology: 3508–18. https://doi.org/10.1074/jbc.M114.602292.

Harrison, Melissa M, Michael R Botchan, and Thomas W Cline. 2010. “Grainyhead and Zelda Compete for Binding to the Promoters of the Earliest-Expressed Drosophila Genes.” Developmental Biology 345 (2). United States: 248–55. https://doi.org/10.1016/j.ydbio.2010.06.026.

Harrison, Melissa M, Xiao-Yong Li, Tommy Kaplan, Michael R Botchan, and Michael B Eisen. 2011. “Zelda Binding in the Early Drosophila Melanogaster Embryo Marks Regions Subsequently Activated at the Maternal-to-Zygotic Transition.” PLoS Genetics 7 (10). United States: e1002266. https://doi.org/10.1371/journal.pgen.1002266.

Harvey, Steven A., Ian Sealy, Ross Kettleborough, Fruzsina Fenyes, Richard White, Derek Stemple, and James C. Smith. 2013. “Identification of the Zebrafish Maternal and Paternal Transcriptomes.” Development 140 (13): 2703–10. https://doi.org/10.1242/dev.095091.

Hendrickson, Peter G, Jessie A Dorais, Edward J Grow, Jennifer L Whiddon, Jong-Won Lim, Candice L Wike, Bradley D Weaver, et al. 2017. “Conserved Roles of Mouse DUX and Human DUX4 in Activating Cleavage-Stage Genes and MERVL/HERVL Retrotransposons.” Nat Genet 49 (6). Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.: 925–34. http://dx.doi.org/10.1038/ng.3844.

Henikoff, Steven, and Ali Shilatifard. 2011. “Histone Modification: Cause or Cog?” Trends in Genetics 27 (10). Elsevier Current Trends: 389–96. https://doi.org/10.1016/J.TIG.2011.06.006.

Heyn, Patricia, Martin Kircher, Andreas Dahl, Janet Kelso, Pavel Tomancak, Alex T Kalinka, and Karla M Neugebauer. 2014. “The Earliest Transcribed Zygotic Genes Are Short, Newly Evolved, and Different across Species.” Cell Reports 6 (2). United States: 285–92. https://doi.org/10.1016/j.celrep.2013.12.030.

Huang, Anqi, Christopher Amourda, Shaobo Zhang, Nicholas S Tolwinski, and Timothy E Saunders. 2017. “Decoding Temporal Interpretation of the Morphogen Bicoid in the Early Drosophila Embryo.” ELife 6. eLife Sciences Publications, Ltd. https://doi.org/10.7554/eLife.26258.

Hug, Clemens B, Alexis G Grimaldi, Kai Kruse, and Juan M Vaquerizas. 2017. “Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription.” Cell 169 (2). United States: 216–228.e19. https://doi.org/10.1016/j.cell.2017.03.024.

Iaco, Alberto De, Evarist Planet, Andrea Coluccio, Sonia Verp, Julien Duc, and Didier Trono. 2017. “DUX-Family Transcription Factors Regulate Zygotic Genome Activation in Placental Mammals.” Nat Genet 49 (6). Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.: 941–45. http://dx.doi.org/10.1038/ng.3858.

Iglesias, Juan Manuel, Juan Gumuzio, and Angel G Martin. 2017. “Linking Pluripotency Reprogramming and Cancer.” Stem Cells Translational Medicine 6 (2). Wiley-Blackwell: 335–39. https://doi.org/10.5966/sctm.2015-0225.

Ishiuchi, Takashi, Rocio Enriquez-Gasca, Eiji Mizutani, Ana Bošković, Celine Ziegler-Birling, Diego Rodriguez-Terrones, Teruhiko Wakayama, Juan M Vaquerizas, and Maria-Elena Torres-Padilla. 2015. “Early Embryonic-like Cells Are Induced by Downregulating Replication-Dependent Chromatin Assembly.” Nature Structural & Molecular Biology 22 (9): 662–71. https://doi.org/10.1038/nsmb.3066.

Jachowicz, Joanna W, Xinyang Bing, Julien Pontabry, Ana Bošković, Oliver J Rando, and Maria-Elena Torres-Padilla. 2017. “LINE-1 Activation after Fertilization Regulates Global Chromatin Accessibility in the Early Mouse Embryo.” Nature Genetics 49 (10). Nature Publishing Group: 1502–10. https://doi.org/10.1038/ng.3945.

Jevtic, Predrag, and Daniel L Levy. 2015. “Nuclear Size Scaling during Xenopus Early Development Contributes to Midblastula Transition Timing.” Current Biology : CB 25 (1). England: 45–52. https://doi.org/10.1016/j.cub.2014.10.051.

Jiang, Cizhong, and B. Franklin Pugh. 2009. “Nucleosome Positioning and Gene Regulation: Advances through Genomics.” Nature Reviews Genetics 10 (3): 161–72. https://doi.org/10.1038/nrg2522.

Jiang, Lan, Jing Zhang, Jing-Jing Wang, Lu Wang, Li Zhang, Guoqiang Li, Xiaodan Yang, et al. 2013. “Sperm, but Not Oocyte, DNA Methylome Is Inherited by Zebrafish Early Embryos.” Cell 153 (4). Cell Press: 773–84. https://doi.org/10.1016/J.CELL.2013.04.041.

Joseph, Shai R, Mate Palfy, Lennart Hilbert, Mukesh Kumar, Jens Karschau, Vasily Zaburdaev, Andrej Shevchenko, and Nadine L Vastenhouw. 2017. “Competition between Histone and Transcription Factor Binding Regulates the Onset of Transcription in Zebrafish Embryos.” ELife 6 (April). England. https://doi.org/10.7554/eLife.23326.

Ke, Yuwen, Yanan Xu, Xuepeng Chen, Songjie Feng, Zhenbo Liu, Yaoyu Sun, Xuelong Yao, et al. 2017. “3D Chromatin Structures of Mature Gametes and Structural Reprogramming during Mammalian Embryogenesis.” Cell 170 (2). Cell Press: 367–381.e20. https://doi.org/10.1016/J.CELL.2017.06.029.

Kelly, William G. 2014. “Transgenerational Epigenetics in the Germline Cycle of Caenorhabditis Elegans.” Epigenetics & Chromatin 7 (1): 6. https://doi.org/10.1186/1756-8935-7-6.

Kimelman, David, Marc Kirschner, and Talma Scherson. 1987. “The Events of the Midblastula Transition in Xenopus Are Regulated by Changes in the Cell Cycle.” Cell 48 (3): 399–407. https://doi.org/10.1016/0092-8674(87)90191-7.

King, Hamish W, and Robert J Klose. 2017. “The Pioneer Factor OCT4 Requires the Chromatin Remodeller BRG1 to Support Gene Regulatory Element Function in Mouse Embryonic Stem Cells.” ELife 6 (March). England. https://doi.org/10.7554/eLife.22631.

Lee, Dong Ryul, Jeoung Eun Lee, Hyun Soo Yoon, Sung Il Roh, and Moon Kyoo Kim. 2001. “Compaction in Preimplantation Mouse Embryos Is Regulated by a Cytoplasmic Regulatory Factor That Alters between 1- and 2-Cell Stages in a Concentration-Dependent Manner.” Journal of Experimental Zoology 290 (1). Wiley-Blackwell: 61–71. https://doi.org/10.1002/jez.1036.

Lee, Miler T, Ashley R Bonneau, Carter M Takacs, Ariel A Bazzini, Kate R DiVito, Elizabeth S Fleming, and Antonio J Giraldez. 2013. “Nanog, Pou5f1 and SoxB1 Activate Zygotic Gene Expression during the Maternal-to-Zygotic Transition.” Nature 503 (7476). England: 360–64. https://doi.org/10.1038/nature12632.

Leichsenring, Manuel, Julia Maes, Rebecca Mossner, Wolfgang Driever, and Daria Onichtchouk. 2013.

"Pou5f1 Transcription Factor Controls Zygotic Gene Activation in Vertebrates." Science (New York, N.Y.) 341 (6149). United States: 1005–9. https://doi.org/10.1126/science.1242527.

Lemmers, R. J. L. F., P. J. van der Vliet, R. Klooster, S. Sacconi, P. Camano, J. G. Dauwerse, L. Snider, et al. 2010. "A Unifying Genetic Model for Facioscapulohumeral Muscular Dystrophy." Science 329 (5999): 1650–53. https://doi.org/10.1126/science.1189044.

Li, En, Caroline Beard, and Rudolf Jaenisch. 1993. "Role for DNA Methylation in Genomic Imprinting." Nature 366 (6453). Nature Publishing Group: 362–65. https://doi.org/10.1038/366362a0.

Li, M M Harrison, J E Villalta, T Kaplan, and M B Eisen. 2014. "Establishment of Regions of Genomic Activity during the Maternal to Zygotic Transition." Elife 3. https://doi.org/10.7554/eLife.03737.

Li, Lin, Fan Guo, Yun Gao, Yixin Ren, Peng Yuan, Liying Yan, Rong Li, et al. 2018. "Single-Cell Multi-Omics Sequencing of Human Early Embryos." Nature Cell Biology 20 (7). Nature Publishing Group: 847–58. https://doi.org/10.1038/s41556-018-0123-2.

Li, Xiao-Yong, Melissa M Harrison, Jacqueline E Villalta, Tommy Kaplan, and Michael B Eisen. 2014. "Establishment of Regions of Genomic Activity during the Drosophila Maternal to Zygotic Transition." ELife 3 (October). England. https://doi.org/10.7554/eLife.03737.

Liang, Hsiao-Lan, Chung-Yi Nien, Hsiao-Yun Liu, Mark M Metzstein, Nikolai Kirov, and Christine Rushlow. 2008. "The Zinc-Finger Protein Zelda Is a Key Activator of the Early Zygotic Genome in Drosophila." Nature 456 (7220). Macmillan Publishers Limited. All rights reserved: 400–403. http://dx.doi.org/10.1038/nature07388.

Lin, C.-J., M. Conti, and M. Ramalho-Santos. 2013. "Histone Variant H3.3 Maintains a Decondensed Chromatin State Essential for Mouse Preimplantation Development." Development 140 (17): 3624–34. https://doi.org/10.1242/dev.095513.

Lindeman, Leif C, Ingrid S Andersen, Andrew H Reiner, Nan Li, Havard Aanes, Olga Ostrup, Cecilia Winata, et al. 2011. "Prepatterning of Developmental Gene Expression by Modified Histones before Zygotic Genome Activation." Developmental Cell 21 (6). United States: 993–1004. https://doi.org/10.1016/j.devcel.2011.10.008.

Liu, Anfei, Xiya Yu, and Shanrong Liu. 2013. "Pluripotency Transcription Factors and Cancer Stem Cells: Small Genes Make a Big Difference." Chinese Journal of Cancer 32 (9). BioMed Central: 483–87. https://doi.org/10.5732/cjc.012.10282.

Liu, Guifen, Wen Wang, Shengen Hu, Xiangxiu Wang, and Yong Zhang. 2018. "Inherited DNA Methylation Primes the Establishment of Accessible Chromatin during Genome Activation." Genome Research, May, gr.228833.117. https://doi.org/10.1101/gr.228833.117.

Liu, Xiaoyu, Chenfei Wang, Wenqiang Liu, Jingyi Li, Chong Li, Xiaochen Kou, Jiayu Chen, et al. 2016. "Distinct Features of H3K4me3 and H3K27me3 Chromatin Domains in Pre-Implantation Embryos." Nature 537 (7621). England: 558–62. https://doi.org/10.1038/nature19362.

Liu, Zhe, Wesley R Legant, Bi-Chang Chen, Li Li, Jonathan B Grimm, Luke D Lavis, Eric Betzig, and Robert Tjian. 2014. "3D Imaging of Sox2 Enhancer Clusters in Embryonic Stem Cells." ELife 3 (December). eLife Sciences Publications Limited: e04236. https://doi.org/10.7554/eLife.04236.

Lott, Susan E, Jacqueline E Villalta, Gary P Schroth, Shujun Luo, Leath A Tonkin, and Michael B Eisen. 2011. "Noncanonical Compensation of the Zygotic X Transcription in Early Drosophila

Melanogaster Development Revealed through Single-Embryo RNA-Seq." PLoS Biology 9 (2): e1000590. https://doi.org/10.1371/journal.pbio.1000590.

Lu, Falong, Yuting Liu, Azusa Inoue, Tsukasa Suzuki, Keji Zhao, and Yi Zhang. 2016. "Establishing Chromatin Regulatory Landscape during Mouse Preimplantation Development." Cell 165 (6): 1375–88. https://doi.org/10.1016/j.cell.2016.05.050.

———. 2017. "Establishing Chromatin Regulatory Landscape during Mouse Preimplantation Development." Cell 165 (6). Elsevier: 1375–88. https://doi.org/10.1016/j.cell.2016.05.050.

Lu, Xuemin, Jennifer M Li, Olivier Elemento, Saeed Tavazoie, and Eric F Wieschaus. 2009. "Coupling of Zygotic Transcription to Mitotic Control at the Drosophila Mid-Blastula Transition." Development (Cambridge, England) 136 (12). England: 2101–10. https://doi.org/10.1242/dev.034421.

Lunde, Karen, Heinz-Georg Belting, and Wolfgang Driever. 2004. "Zebrafish Pou5f1/Pou2, Homolog of Mammalian Oct4, Functions in the Endoderm Specification Cascade." Current Biology : CB 14 (1): 48–55. http://www.ncbi.nlm.nih.gov/pubmed/14711414.

Macfarlan, Todd S., Wesley D. Gifford, Shawn Driscoll, Karen Lettieri, Helen M. Rowe, Dario Bonanomi, Amy Firth, Oded Singer, Didier Trono, and Samuel L. Pfaff. 2012. "Embryonic Stem Cell Potency Fluctuates with Endogenous Retrovirus Activity." Nature 487 (7405): 57–63. https://doi.org/10.1038/nature11244.

McCleland, Mark L., and Patrick H. O'Farrell. 2008. "RNAi of Mitotic Cyclins in Drosophila Uncouples the Nuclear and Centrosome Cycle." Current Biology 18 (4). Cell Press: 245–54. https://doi.org/10.1016/J.CUB.2008.01.041.

Meier, Michael, Jenny Grant, Amy Dowdle, Amarni Thomas, Jennifer Gerton, Philippe Collas, Justin M. O'Sullivan, and Julia A. Horsfield. 2018. "Cohesin Facilitates Zygotic Genome Activation in Zebrafish." Development 145 (1): dev156521. https://doi.org/10.1242/dev.156521.

Mir, Mustafa, Armando Reimer, Jenna E Haines, Xiao-Yong Li, Michael Stadler, Hernan Garcia, Michael B Eisen, and Xavier Darzacq. 2017. "Dense Bicoid Hubs Accentuate Binding along the Morphogen Gradient." Genes & Development 31 (17). Cold Spring Harbor Laboratory Press: 1784–94. https://doi.org/10.1101/gad.305078.117.

Nardini, Marco, Nerina Gnesutta, Giacomo Donati, Raffaella Gatta, Claudia Forni, Andrea Fossati, Clemens Vonrhein, et al. 2013. "Sequence-Specific Transcription Factor NF-Y Displays Histone-like DNA Binding and H2B-like Ubiquitination." Cell 152 (1–2): 132–43. https://doi.org/10.1016/j.cell.2012.11.047.

Nelles, David A., Mark Y. Fang, Mitchell R. O'Connell, Jia L. Xu, Sebastian J. Markmiller, Jennifer A. Doudna, and Gene W. Yeo. 2016. "Programmable RNA Tracking in Live Cells with CRISPR/Cas9." Cell 165 (2): 488–96. https://doi.org/10.1016/j.cell.2016.02.054.

Newport, J, and M Kirschner. 1982a. "A Major Developmental Transition in Early Xenopus Embryos: I. Characterization and Timing of Cellular Changes at the Midblastula Stage." Cell 30 (3): 675–86. http://www.ncbi.nlm.nih.gov/pubmed/6183003.

———. 1982b. "A Major Developmental Transition in Early Xenopus Embryos: II. Control of the Onset of Transcription." Cell 30 (3): 687–96. http://www.ncbi.nlm.nih.gov/pubmed/7139712.

Nien, C Y, H L Liang, S Butcher, Y Sun, S Fu, T Gocha, N Kirov, J R Manak, and C Rushlow. 2011. "Temporal

Coordination of Gene Networks by Zelda in the Early Drosophila Embryo.” PLoS Genet 7 (10): e1002339. https://doi.org/10.1371/journal.pgen.1002339.

Oldfield, Andrew J, Pengyi Yang, Amanda E Conway, Senthilkumar Cinghu, Johannes M Freudenberg, Sailu Yellaboina, and Raja Jothi. 2014. “Histone-Fold Domain Protein NF-Y Promotes Chromatin Accessibility for Cell Type-Specific Master Transcription Factors.” Molecular Cell 55 (5). NIH Public Access: 708–22. https://doi.org/10.1016/j.molcel.2014.07.005.

Onichtchouk, Daria, and Wolfgang Driever. 2016. “Zygotic Genome Activators, Developmental Timing, and Pluripotency.” In Current Topics in Developmental Biology, 116:273–97. https://doi.org/10.1016/bs.ctdb.2015.12.004.

Panning, Barbara. 2008. “X-Chromosome Inactivation: The Molecular Basis of Silencing.” Journal of Biology 7 (8). BioMed Central: 30. https://doi.org/10.1186/jbiol95.

Perez-Montero, Salvador, Albert Carbonell, Tomas Moran, Alejandro Vaquero, and Fernando Azorin. 2013. “The Embryonic Linker Histone H1 Variant of Drosophila, DBigH1, Regulates Zygotic Genome Activation.” Developmental Cell 26 (6). United States: 578–90. https://doi.org/10.1016/j.devcel.2013.08.011.

Potok, Magdalena E., David A. Nix, Timothy J. Parnell, and Bradley R. Cairns. 2013. “Reprogramming the Maternal Zebrafish Genome after Fertilization to Match the Paternal Methylation Pattern.” Cell 153 (4). Cell Press: 759–72. https://doi.org/10.1016/J.CELL.2013.04.030.

Prioleau, M N, J Huet, A Sentenac, and M Mechali. 1994. “Competition between Chromatin and Transcription Complex Assembly Regulates Gene Expression during Early Development.” Cell 77 (3). United States: 439–49.

Pritchard, D K, and G Schubiger. 1996. “Activation of Transcription in Drosophila Embryos Is a Gradual Process Mediated Chardby the Nucleocytoplasmic Ratio.” Genes & Development 10 (9). United States: 1131–42.

Renzis, S De, O Elemento, S Tavazoie, and E F Wieschaus. 2007. “Unmasking Activation of the Zygotic Genome Using Chromosomal Deletions in the Drosophila Embryo.” PLoS Biol 5 (5): e117. https://doi.org/10.1371/journal.pbio.0050117.

Rothe, Mike, Michael Pehl, Heike Taubert, and Herbert Jackle. 1992. “Loss of Gene Function through Rapid Mitotic Cycles in the Drosophila Embryo.” Nature 359 (6391): 156–59. http://dx.doi.org/10.1038/359156a0.

Ruzov, Alexey, Donncha S. Dunican, Anna Prokhortchouk, Sari Pennings, Irina Stancheva, Egor Prokhortchouk, and Richard R. Meehan. 2004. “Kaiso Is a Genome-Wide Repressor of Transcription That Is Essential for Amphibian Development.” Development 131 (24). http://dev.biologists.org/content/131/24/6185.

Schulz, K N, E R Bondra, A Moshe, J E Villalta, J D Lieb, T Kaplan, D J McKay, and M M Harrison. 2015. “Zelda Is Differentially Required for Chromatin Accessibility, Transcription Factor Binding, and Gene Expression in the Early Drosophila Embryo.” Genome Res 25 (11): 1715–26. https://doi.org/10.1101/gr.192682.115.

Senecal, Adrien, Brian Munsky, Florence Proux, Nathalie Ly, Floriane E. Braye, Christophe Zimmer, Florian Mueller, and Xavier Darzacq. 2014. “Transcription Factors Modulate C-Fos Transcriptional Bursts.” Cell Reports 8 (1): 75–83. https://doi.org/10.1016/j.celrep.2014.05.053.

Shen, Li, Azusa Inoue, Jin He, Yuting Liu, Falong Lu, and Yi Zhang. 2014. "Tet3 and DNA Replication Mediate Demethylation of Both the Maternal and Paternal Genomes in Mouse Zygotes." Cell Stem Cell 15 (4): 459–71. https://doi.org/10.1016/j.stem.2014.09.002.

Shermoen, A W, and P H O'Farrell. 1991. "Progression of the Cell Cycle through Mitosis Leads to Abortion of Nascent Transcripts." Cell 67 (2). United States: 303–10.

Smith, R C, E Dworkin-Rastl, and M B Dworkin. 1988. "Expression of a Histone H1-like Protein Is Restricted to Early Xenopus Development." Genes & Development 2 (10): 1284–95. http://www.ncbi.nlm.nih.gov/pubmed/3060404.

Soufi, Abdenour, Greg Donahue, and Kenneth S. Zaret. 2012. "Facilitators and Impediments of the Pluripotency Reprogramming Factors' Initial Engagement with the Genome." Cell 151 (5): 994–1004. https://doi.org/10.1016/j.cell.2012.09.045.

Soufi, Abdenour, Meilin Fernandez Garcia, Artur Jaroszewicz, Nebiyu Osman, Matteo Pellegrini, and Kenneth S. Zaret. 2015. "Pioneer Transcription Factors Target Partial DNA Motifs on Nucleosomes to Initiate Reprogramming." Cell 161 (3): 555–68. https://doi.org/10.1016/j.cell.2015.03.017.

Stadhouders, Ralph, Enrique Vidal, François Serra, Bruno Di Stefano, François Le Dily, Javier Quilez, Antonio Gomez, et al. 2018. "Transcription Factors Orchestrate Dynamic Interplay between Genome Topology and Gene Regulation during Cell Reprogramming." Nature Genetics 50 (2). Nature Publishing Group: 238–49. https://doi.org/10.1038/s41588-017-0030-7.

Stancheva, I, and R R Meehan. 2000. "Transient Depletion of XDnmt1 Leads to Premature Gene Activation in Xenopus Embryos." Genes & Development 14 (3). Cold Spring Harbor Laboratory Press: 313–27. https://doi.org/10.1101/GAD.14.3.313.

Staudt, Nicole, Sonja Fellert, Ho-Ryun Chung, Herbert Jäckle, and Gerd Vorbrüggen. 2006. "Mutations of the Drosophila Zinc Finger-Encoding Gene Vielfältig Impair Mitotic Cell Divisions and Cause Improper Chromosome Segregation." Molecular Biology of the Cell 17 (5): 2356–65. https://doi.org/10.1091/mbc.E05-11-1056.

Strom, Amy R., Alexander V. Emelyanov, Mustafa Mir, Dmitry V. Fyodorov, Xavier Darzacq, and Gary H. Karpen. 2017. "Phase Separation Drives Heterochromatin Domain Formation." Nature 547 (7662). Nature Publishing Group: 241–45. https://doi.org/10.1038/nature22989.

Sun, Yujia, Chung-Yi Nien, Kai Chen, Hsiao-Yun Liu, Jeff Johnston, Julia Zeitlinger, and Christine Rushlow. 2015. "Zelda Overcomes the High Intrinsic Nucleosome Barrier at Enhancers during Drosophila Zygotic Genome Activation." Genome Research 25 (11). United States: 1703–14. https://doi.org/10.1101/gr.192542.115.

Sung, Hung-wei, Saskia Spangenberg, Nina Vogt, and Jörg Großhans. 2017. "Number of Nuclear Divisions in the <em>Drosophila</Em> Blastoderm Controlled by Onset of Zygotic Transcription." Current Biology 23 (2). Elsevier: 133–38. https://doi.org/10.1016/j.cub.2012.12.013.

Svoboda, Petr, Paula Stein, Martin Anger, Emily Bernstein, Gregory J Hannon, and Richard M Schultz. 2004. "RNAi and Expression of Retrotransposons MuERV-L and IAP in Preimplantation Mouse Embryos." Developmental Biology 269 (1): 276–85. https://doi.org/10.1016/j.ydbio.2004.01.028.

Swinburne, Ian A, and Pamela A Silver. 2008. "Intron Delays and Transcriptional Timing during Development." Developmental Cell 14 (3). NIH Public Access: 324–30. https://doi.org/10.1016/j.devcel.2008.02.002.

Tadros, and H D Lipshitz. 2009. "The Maternal-to-Zygotic Transition: A Play in Two Acts." Development 136 (18): 3033–42. https://doi.org/10.1242/dev.033183.

Tadros, Wael, Aaron L Goldman, Tomas Babak, Fiona Menzies, Leah Vardy, Terry Orr-Weaver, Timothy R Hughes, J Timothy Westwood, Craig A Smibert, and Howard D Lipshitz. 2007. "SMAUG Is a Major Regulator of Maternal MRNA Destabilization in Drosophila and Its Translation Is Activated by the PAN GU Kinase." Developmental Cell 12 (1). Elsevier: 143–55. https://doi.org/10.1016/j.devcel.2006.10.005.

Takahashi, Kazutoshi, and Shinya Yamanaka. 2016. "A Decade of Transcription Factor-Mediated Reprogramming to Pluripotency." Nature Reviews Molecular Cell Biology 17 (3). Nature Publishing Group: 183–93. https://doi.org/10.1038/nrm.2016.8.

Takayama, S., J. Dhahbi, A. Roberts, G. Mao, S.-J. Heo, L. Pachter, D. I. K. Martin, and D. Boffelli. 2014. "Genome Methylation in D. Melanogaster Is Found at Specific Short Motifs and Is Independent of DNMT2 Activity." Genome Research 24 (5): 821–30. https://doi.org/10.1101/gr.162412.113.

Vanderplanck, Céline, Eugénie Ansseau, Sébastien Charron, Nadia Stricwant, Alexandra Tassin, Dalila Laoudj-Chenivesse, Steve D Wilton, Frédérique Coppée, and Alexandra Belayew. 2011. "The FSHD Atrophic Myotube Phenotype Is Caused by DUX4 Expression." PloS One 6 (10). Public Library of Science: e26820. https://doi.org/10.1371/journal.pone.0026820.

Vastenhouw, Nadine L, Yong Zhang, Ian G Woods, Farhad Imam, Aviv Regev, X Shirley Liu, John Rinn, and Alexander F Schier. 2010. "Chromatin Signature of Embryonic Pluripotency Is Established during Genome Activation." Nature 464 (7290). Macmillan Publishers Limited. All rights reserved: 922–26. http://dx.doi.org/10.1038/nature08866.

Veenstra, G J, O H Destree, and A P Wolffe. 1999. "Translation of Maternal TATA-Binding Protein MRNA Potentiates Basal but Not Activated Transcription in Xenopus Embryos at the Midblastula Transition." Molecular and Cellular Biology 19 (12). United States: 7972–82.

Veenstra, Gert Jan C., and Alan P. Wolffe. 2001. "Constitutive Genomic Methylation during Embryonic Development of Xenopus." Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression 1521 (1–3). Elsevier: 39–44. https://doi.org/10.1016/S0167-4781(01)00280-9.

Veil, Marina, Lev Yampolsky, Bjoern Gruening, and Daria Onichtchouk. 2018. "Pou5f3, SoxB1 and Nanog Remodel Chromatin on High Nucleosome Affinity Regions at Zygotic Genome Activation." BioRxiv, June. Cold Spring Harbor Laboratory, 344168. https://doi.org/10.1101/344168.

Wagner, Daniel E, Caleb Weinreb, Zach M Collins, James A Briggs, Sean G Megason, and Allon M Klein. 2018. "Single-Cell Mapping of Gene Expression Landscapes and Lineage in the Zebrafish Embryo." Science (New York, N.Y.) 360 (6392). American Association for the Advancement of Science: 981–87. https://doi.org/10.1126/science.aar4362.

Wang, Yixuan, Chengchen Zhao, Zhenzhen Hou, Yuanyuan Yang, Yan Bi, Hong Wang, Yong Zhang, and Shaorong Gao. 2018. "Unique Molecular Events during Reprogramming of Human Somatic Cells to Induced Pluripotent Stem Cells (IPSCs) at Naïve State." ELife 7 (January). eLife Sciences Publications, Ltd. https://doi.org/10.7554/eLife.29518.

Weber, Christopher M, and Steven Henikoff. 2014. "Histone Variants: Dynamic Punctuation in Transcription." Genes & Development 28 (7). Cold Spring Harbor Laboratory Press: 672–82. https://doi.org/10.1101/gad.238873.114.

Whiddon, Jennifer L, Ashlee T Langford, Chao-Jen Wong, Jun Wen Zhong, and Stephen J Tapscott. 2017. “Conservation and Innovation in the DUX4-Family Gene Network.” Nat Genet 49 (6). Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.: 935–40. http://dx.doi.org/10.1038/ng.3846.

White, Richard J, John E Collins, Ian M Sealy, Neha Wali, Christopher M Dooley, Zsofia Digby, Derek L Stemple, et al. 2017. “A High-Resolution MRNA Expression Time Course of Embryonic Development in Zebrafish.” ELife 6 (November). eLife Sciences Publications, Ltd. https://doi.org/10.7554/eLife.30860.

Whittle, Christina M., Karissa N. McClinic, Sevinc Ercan, Xinmin Zhang, Roland D. Green, William G. Kelly, and Jason D. Lieb. 2008. “The Genomic Distribution and Function of Histone Variant HTZ-1 during C. Elegans Embryogenesis.” Edited by Asifa Akhtar. PLoS Genetics 4 (9). Public Library of Science: e1000187. https://doi.org/10.1371/journal.pgen.1000187.

Wu, Jingyi, Bo Huang, He Chen, Qiangzong Yin, Yang Liu, Yunlong Xiang, Bingjie Zhang, et al. 2016. “The Landscape of Accessible Chromatin in Mammalian Preimplantation Embryos.” Nature 534 (7609). Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.: 652–57. http://dx.doi.org/10.1038/nature18606.

Wu, Jingyi, Jiawei Xu, Bofeng Liu, Guidong Yao, Peizhe Wang, Zili Lin, Bo Huang, et al. 2018. “Chromatin Analysis in Human Early Development Reveals Epigenetic Transition during ZGA.” Nature, May. Nature Publishing Group, 1. https://doi.org/10.1038/s41586-018-0080-8.

Yan, Liying, Mingyu Yang, Hongshan Guo, Lu Yang, Jun Wu, Rong Li, Ping Liu, et al. 2013. “Single-Cell RNA-Seq Profiling of Human Preimplantation Embryos and Embryonic Stem Cells.” Nature Structural & Molecular Biology 20 (9). Nature Publishing Group: 1131–39. https://doi.org/10.1038/nsmb.2660.

Yartseva, Valeria, and Antonio J Giraldez. 2015. “The Maternal-to-Zygotic Transition During Vertebrate Development: A Model for Reprogramming.” Current Topics in Developmental Biology 113. NIH Public Access: 191–232. https://doi.org/10.1016/bs.ctdb.2015.07.020.

Yin, Yimeng, Ekaterina Morgunova, Arttu Jolma, Eevi Kaasinen, Biswajyoti Sahu, Syed Khund-Sayeed, Pratyush K. Das, et al. 2017. “Impact of Cytosine Methylation on DNA Binding Specificities of Human Transcription Factors.” Science 356 (6337): eaaj2239. https://doi.org/10.1126/science.aaj2239.

Young, Janet M., Jennifer L. Whiddon, Zizhen Yao, Bhavatharini Kasinathan, Lauren Snider, Linda N. Geng, Judit Balog, Rabi Tawil, Silvère M. van der Maarel, and Stephen J. Tapscott. 2013. “DUX4 Binding to Retroelements Creates Promoters That Are Active in FSHD Muscle and Testis.” Edited by Cédric Feschotte. PLoS Genetics 9 (11): e1003947. https://doi.org/10.1371/journal.pgen.1003947.

Yuan, Kai, Charles A Seller, Antony W Shermoen, and Patrick H O’Farrell. 2016. “Timing the Drosophila Mid-Blastula Transition: A Cell Cycle-Centered View.” Trends in Genetics : TIG 32 (8). England: 496–507. https://doi.org/10.1016/j.tig.2016.05.006.

Zaret, Kenneth S, and Jason S Carroll. 2011. “Pioneer Transcription Factors: Establishing Competence for Gene Expression.” Genes Dev 25 (21). Cold Spring Harbor Laboratory Press: 2227–41. https://doi.org/10.1101/gad.176826.111.

Zhang, Bingjie, Hui Zheng, Bo Huang, Wenzhi Li, Yunlong Xiang, Xu Peng, Jia Ming, et al. 2016. “Allelic

Reprogramming of the Histone Modification H3K4me3 in Early Mammalian Development." Nature 537 (7621). Macmillan Publishers Limited, part of Springer Nature. All rights reserved.: 553–57. http://dx.doi.org/10.1038/nature19361.

Zhang, Maomao, Priyanka Kothari, Mary Mullins, and Michael A Lampson. 2014. "Regulation of Zygotic Genome Activation and DNA Damage Checkpoint Acquisition at the Mid-Blastula Transition." Cell Cycle (Georgetown, Tex.) 13 (24). United States: 3828–38. https://doi.org/10.4161/15384101.2014.967066.

Zhang, Y, N L Vastenhouw, J Feng, K Fu, C Wang, Y Ge, A Pauli, P van Hummelen, A F Schier, and X S Liu. 2014. "Canonical Nucleosome Organization at Promoters Forms during Genome Activation." Genome Res 24 (2): 260–66. https://doi.org/10.1101/gr.157750.113.

Chapter 2 - Zelda regulates chromatin accessibility in the early Drosophila embryo

The work presented in this chapter has been published:

Schulz, K. N., Bondra, E. R., Moshe, A., Villalta, J. E., Lieb, J. D., Kaplan, T., McKay, D. J. and Harrison, M.M. (2015). Zelda is differentially required for chromatin accessibility, transcription factor binding, and gene expression in the early Drosophila embryo. *Genome Res.* 25(11): 1715–1726.

Statement of contribution: I performed the FAIRE-seq experiments described in this paper and assisted with the analysis of the results.

Abstract

The transition from a specified germ cell to a population of pluripotent cells occurs rapidly following fertilization. During this developmental transition, the zygotic genome is largely transcriptionally quiescent and undergoes significant chromatin remodeling. In *Drosophila*, the DNA-binding protein Zelda (also known as Vielfaltig) is required for this transition and for transcriptional activation of the zygotic genome. Open chromatin is associated with Zelda-bound loci as well as more generally with regions of active transcription. Nonetheless, the extent to which Zelda influences chromatin accessibility across the genome is largely unknown. Here we used Formaldehyde Assisted Isolation of Regulatory Elements to determine the role of Zelda in regulating regions of open chromatin in the early embryo. We demonstrate that Zelda is essential for hundreds of regions of open chromatin. This Zelda-mediated chromatin accessibility facilitates transcription-factor recruitment and early gene expression. Thus, Zelda possesses some key characteristics of a pioneer factor. Unexpectedly, chromatin at a large subset of Zelda-bound regions remains open even in the absence of Zelda. The GAGA factor-binding motif and embryonic GAGA factor binding are specifically enriched in these regions. We propose that both Zelda and GAGA factor function to specify sites of open chromatin and together facilitate the remodeling of the early embryonic genome.

Introduction

Transcription factors drive cell-fate specification by binding to cis-regulatory regions and controlling gene expression. Despite the broad impact of these factors, our understanding of how they access and function at the correct sites in the genome remains incomplete. In particular, while transcription factors recognize and bind specific DNA motifs, only a small fraction of their potential binding sites are occupied at a given time point (Carr and Biggin 1999; Iyer et al. 2001; Lieb et al. 2001; Liu et al. 2006; Yang et al. 2006; Li et al. 2008). Thus, other features in addition to DNA sequence must influence where transcription factors bind in the genome. *In vivo*, DNA is wrapped around histone proteins to form nucleosomes, which compete with transcription factors for access to DNA-encoded information. Reduced accessibility of nucleosomal DNA can largely explain the patterns of transcription factor binding *in vivo* (Kaplan et al. 2011; Li et al. 2011). However, the mechanisms that create sites of accessibility are poorly understood. It has been proposed that a special class of transcription factors, termed "pioneer factors", binds nucleosomal DNA and increases chromatin accessibility for other transcription factors (Zaret and Carroll 2011). This model is based on the activity of the Forkhead box (FOX) family of proteins that bind nucleosomal DNA early in hepatic development and facilitate open chromatin, making these regions accessible for subsequent transcription factor binding (Gualdi et al. 1996; Cirillo et al. 1998; Cirillo and Zaret 1999; Cirillo et al. 2002).

Although it is widely accepted that chromatin structure is reconfigured during early embryonic development, little is known about the factors that direct this process. These changes in chromatin structure accompany the transition from a specified germ cell to a population of pluripotent cells, which occurs rapidly following fertilization. At this time the zygotic genome is not transcribed, and maternally contributed mRNAs and proteins control embryonic development. Only at later cell cycles is widespread transcription initiated (Newport and Kirschner 1982; Tadros and Lipshitz 2009). The degradation of maternally deposited mRNAs is coordinated with the transcriptional activation of the zygotic genome

during the maternal-to-zygotic transition (MZT). This fundamental transition is conserved among metazoans and is essential for development.

In *Drosophila melanogaster*, early development is characterized by a series of 13 rapid synchronous nuclear divisions. Approximately 2-3 hours after fertilization, at the 14th nuclear cycle, the division cycle slows and widespread zygotic transcription initiates. This marks the end of the *D. melanogaster* MZT. Maternally deposited *zelda (zld)* (also known as *vielfaltig*) mRNA is necessary for zygotic genome activation and development beyond the MZT (Staudt et al. 2006; Liang et al. 2008). Previous work has demonstrated that loci bound by Zelda (ZLD) as early as nuclear cycle 8 are associated with regions of open chromatin and transcription-factor binding at cycle 14 (Li et al. 2008; MacArthur et al. 2009; Harrison et al. 2011). Indeed, ZLD binding at early time points is a robust predictor of subsequent transcription factor binding (Harrison et al. 2011), and it has recently been shown that ZLD-binding sites are instrumental in regulating DNA binding by the transcription factors Dorsal (DL), Twist (TWI) and Bicoid (BCD) (Yanez-Cuna et al. 2012; Foo et al. 2014; Xu et al. 2014). Furthermore, it has been proposed that ZLD acts to potentiate transcription factor binding by determining regions of open chromatin (Harrison et al. 2011; Satija and Bradley 2012). At regions bound by ZLD in wild-type embryos, histone H3 density increases when ZLD is depleted (Li et al. 2014), supporting this hypothesis. Additionally, chromatin accessibility as assayed by DNase I accessibility is correlated with levels of ZLD binding to the *brinker* and *short gastrulation (sog)* enhancers (Foo et al. 2014). Nonetheless, it is unknown whether ZLD is required for open chromatin at all of the thousands of loci to which it binds. Furthermore, whether ZLD-mediated chromatin accessibility facilitates transcription factor binding remains to be determined.

Here, we directly test the role of ZLD in shaping the chromatin environment during the MZT. We used Formaldehyde Assisted Isolation of Regulatory Elements (FAIRE) (Giresi et al. 2007; McKay and Lieb 2013) to perform genome-wide profiling of open chromatin in embryos lacking maternal ZLD (*zld*$^{M-}$). We

demonstrate that ZLD is required to establish or maintain hundreds of regions of open chromatin as was previously predicted. Unexpectedly, we also identified a large number of ZLD-bound loci that remain accessible in *zld*-depleted embryos. Together, our findings indicate that ZLD and another maternally deposited factor, GAGA-factor, coordinately function to define the regulatory landscape of the early embryo.

Results

ZLD establishes or maintains regions of open chromatin

We have previously shown that ZLD-bound regions are highly correlated with DNA accessibility (Harrison et al. 2011). To test the function of ZLD in establishing or maintaining these regions of open chromatin, we performed FAIRE on 2-3 hour embryos with wild-type levels of ZLD (*yw*) and embryos depleted for maternally contributed *zld* (zld^{M-}) (Liang et al. 2008). Obtaining embryos depleted for maternal *zld* requires generating germline clones using heat shock induced mitotic recombination (Chou et al. 1993; Liang et al. 2008). Thus, to control for any influence of this protocol on chromatin accessibility, we generated germline clones in our *yw* control strain in parallel with our *zld* mutant strain. Immunostaining confirmed the successful generation of embryos lacking ZLD (Sup. Figure 2.1). We sequenced and analyzed three biological replicates of the zld^{M-} embryos and two replicates of the *yw* control embryos. Regions enriched by FAIRE in our 2-3 hour *yw* control embryos correlated well with the regions enriched in 2-4 hour wild-type (OreR) embryos ((McKay and Lieb 2013), Figure 2.1A (compare *yw* control FAIRE to OreR FAIRE), Sup. Figure 2.2, Spearman's rho = 0.75). Differences between the two samples are likely the result of the different developmental time points at which the samples were collected as well as any effects from the induction of germline clones.

Because active cis-regulatory regions are correlated with chromatin accessibility (Fisher et al. 2012), we focused our analysis on the high-confidence open regions. Specifically, we took the top 5000 peaks in our zld^{M-} dataset (MACS adjusted -log10 p-value = 39) and controlled for differences in data quality between the experimental and control by similarly selecting the top 5000 peaks from the *yw* control (MACS adjusted -log10 p-value = 18). To determine how chromatin accessibility changed when maternal *zld* was removed, we took the union of the top 5000 peaks from both of our datasets (resulting in 6042 peak regions along the fly genome) and used edgeR to identify differentially accessible regions. That is, FAIRE peaks that vary between the two samples ($p < 0.05$) (Robinson et al. 2010). We found 540

regions that had enriched FAIRE signals in our control *yw* embryos as compared to zld^{M-} embryos (*yw* differential) (Figure 2.1A, B). These regions are exemplified in the cis-regulatory regions for *deadpan* (*dpn)* and *sog* by distinct FAIRE peaks that are present in both *yw* and OreR FAIRE, but absent in the zld^{M-} FAIRE (Figure 2.1A; blue boxes). Conversely, there were 145 regions that were enriched in the zld^{M-} embryos as compared to the control (zld^{M-} differential).

To determine whether ZLD directly contributes to the changes in chromatin accessibility we observed between our *yw* control and zld^{M-} embryos, we quantified the enrichment of the canonical ZLD-binding site in our constitutive FAIRE peaks along with those regions that become more accessible (zld^{M-} differential) and those that become less accessible (*yw* differential) in the absence of ZLD. Specifically, we searched for occurrences of CAGGTAG – the optimal sequence recognition element for ZLD – within those regions that were uniquely enriched in either *yw* or zld^{M-} embryos. While only 12% of the top 5000 FAIRE peaks in our *yw* control contain the CAGGTAG motif, 45% of the ZLD-dependent peaks (*yw* differential) contain CAGGTAG motifs (see Table 1 for full data, including the enrichment for CAGGTA). By contrast, of the 145 peaks that become more accessible in zld^{M-} embryos only 0.7% contain CAGGTAG (Table 1, zld^{M-} differential). Thus, the ZLD-binding site is selectively enriched in regions that depend on ZLD for accessibility.

When we compared our control (*yw*) FAIRE peaks to regions bound by ZLD, as determined by chromatin immunoprecipitation coupled with high throughput sequencing (ChIP-seq) (Harrison et al. 2011), we similarly identified binding enrichment specifically at those regions that depend on ZLD for accessibility (*yw* differential). At cycle 14, 54% (2681/5000) of the control FAIRE peaks overlap with a region bound by ZLD *in vivo*, while 77% (418/540) of the ZLD-dependent peaks (*yw* differential) overlap with a ZLD-bound region. This correlation between ZLD binding and regions that require ZLD for open chromatin is also evident at individual loci, such as *dpn* and *sog* (Figure 2.1A; blue boxes). Furthermore, only 15% of those peaks that increase in accessibility in the absence of maternal ZLD (zld^{M-} differential)

overlap with *in vivo* ZLD-binding sites. Thus, both the ZLD-binding sequence and *in vivo* ZLD binding are enriched specifically in those regions that are more open in the control as compared to *zld*$^{M-}$ embryos (*yw* differential). Together these data indicate that ZLD is required to maintain or establish distinct regions of open chromatin.

ZLD binds to promoters and enhancers of genes expressed through the MZT (Harrison et al. 2011; Nien et al. 2011) (Figure 2.1C). To assess which genomic regions might be particularly dependent on ZLD for chromatin accessibility, we determined the distribution of FAIRE peaks that were specific to the *yw* control (*yw* differential), FAIRE peaks that were accessible in both *yw* and *zld*$^{M-}$ embryos (constitutive) and FAIRE peaks specific to the *zld*$^{M-}$ embryos (*zld*$^{M-}$ differential)(Figure 2.1C). The observed distribution of FAIRE peaks that require ZLD (*yw* differential) was similar to the distribution of regions bound by ZLD *in vivo* and of CAGGTAG motifs (Figure 2.1C), suggesting that ZLD functions to establish or maintain chromatin accessibility at both promoters and enhancers. By contrast, the peaks specific to the *zld*$^{M-}$ (*zld*$^{M-}$ differential) showed a genomic distribution substantially different from ZLD binding and CAGGTAG motifs, with a strong enrichment in intergenic regions.

ZLD is not essential for chromatin accessibility at many ZLD-bound loci

During the initial analysis of our FAIRE data, we identified three classes of FAIRE peaks that we had expected to identify. First, fitting with the predicted role of ZLD as a pioneer-like factor, we identified ZLD-bound loci that required ZLD for chromatin accessibility. Second, we identified a large class of FAIRE peaks that did not bind ZLD or require ZLD for accessibility. Chromatin accessibility in these regions likely resulted from the activities of other proteins or from inherent DNA-sequence structure. Third, we identified a small number of regions that depend on ZLD for accessibility but that do not have ZLD bound, which are likely due to indirect effects of ZLD depletion. However, in addition to these three classes our analysis unexpectedly revealed a large class of FAIRE peaks that overlapped ZLD-

bound regions, but remained accessible in *zld*$^{M-}$ embryos. Thus, while ZLD was required for chromatin accessibility at 540 loci, there were nearly three times as many loci (1537) that were bound by ZLD, but did not require ZLD occupancy for accessibility.

For further analysis, we therefore divided the 5000 top FAIRE peaks from our *yw* embryos into four categories: differential, ZLD-bound (402 peaks); constitutive, ZLD-bound (1537 peaks); differential, not ZLD-bound (138 peaks); and constitutive, not ZLD-bound (2923 peaks) (Figure 2.2A). Examples of these different classes are highlighted in the regulatory regions surrounding both *dpn* and *sog* (Figure 2.1A). The average FAIRE signals from the differential and constitutive peaks were generally similar. However, the differential peaks showed a slightly lower level of FAIRE signal, especially those differential peaks that did not overlap ZLD-bound regions (differential, not ZLD-bound) (Figure 2.2B). As defined, the FAIRE peaks for the differential class are substantially decreased in the *zld*$^{M-}$ embryos (Figure 2.2B). Additionally, there was no detectable difference between the average FAIRE signals from constitutive regions either bound or not bound by ZLD (Figure 2.2B). Thus, the strength of the FAIRE signal is unlikely to explain the differences between these four classes.

When we used sequence surrounding the FAIRE peaks to predict nucleosome occupancy (Kaplan et al. 2009), we did not identify a predicted region of nucleosome depletion at the FAIRE peaks in any of the four classes (Figure 2.2C). On the contrary, we found that sequences underlying FAIRE peaks are predicted to have higher nucleosome occupancy, consistent with previous observations (McKay and Lieb 2013). This analysis suggests that the underlying sequence alone does not explain the open chromatin identified by FAIRE at any of these regions and that active mechanisms are likely working to create or maintain chromatin accessibility. To better define the features that determine chromatin accessibility in the early embryo, we focused our analysis on characterizing these classes of FAIRE peaks.

A large number of promoters do not require ZLD for accessibility

Many factors are known to shape chromatin accessibility so we were not surprised to identify a relatively large class of FAIRE peaks that were unchanged in the absence of ZLD and which did not possess ZLD-binding sites (constitutive, not ZLD-bound; n = 2923). When we analyzed the distribution of these regions throughout the genome we noted they were highly enriched for promoter regions with nearly 70% of these FAIRE peaks being localized to promoters (Figure 2.2D). Promoters, especially of transcriptionally active genes, are known to be relatively nucleosome free in a wide variety of species (Lee et al. 2007; Schones et al. 2008; Weiner et al. 2010; Thomas et al. 2011; Valouev et al. 2011). In zebrafish, well-positioned nucleosome arrays appear at promoters during zygotic genome activation in a transcription-independent fashion, suggesting that active mechanisms may shape nucleosome occupancy at promoters during early development (Zhang et al. 2014). When we searched for motifs enriched among these constitutively open peaks, all k-mers identified were highly AT rich (Supplemental Table 1). AT-rich sequences were also enriched when we analyzed all FAIRE peaks lacking ZLD as compared to all FAIRE peaks with ZLD bound. AT-rich sequences are associated with decreased nucleosome occupancy and have also been shown to be enriched at promoters (Sekinger et al. 2005; Hughes et al. 2012). Because the underlying sequence does not predict any inherent decrease in nucleosome occupancy at FAIRE peaks that are independent of ZLD binding (constitutive, not ZLD-bound) (Figure 2.2C), we suggest the identification of the AT-rich sequence motif in these regions likely results from the fact that this sets of peaks is enriched for promoters and does not by itself explain the open chromatin at these regions.

ZLD-mediated chromatin accessibility is associated with early gene expression

Our analysis of the top FAIRE peaks demonstrated that there were two classes of accessible chromatin regions bound by ZLD: those that required ZLD for accessibility (differential, ZLD-bound; n =

402) and those that did not (constitutive, ZLD-bound; n = 1537). To begin to differentiate between these two classes of ZLD-bound loci, we used ChIP-seq data to determine the relative strength of ZLD binding at nuclear cycle 14 (Harrison et al. 2011). Constitutive peaks with ZLD bound (constitutive, ZLD-bound) had on average 3-fold lower ZLD ChIP-signals than those that were dependent on ZLD for accessibility (differential, ZLD-bound) (Figure 2.3A). This shows that regions highly occupied by ZLD *in vivo* are more dependent on ZLD for establishing or maintaining chromatin accessibility than those less-occupied sites (see also Sup. Figure 2.3). We have previously shown that levels of ZLD binding are correlated with the timing of gene expression; regions with higher ZLD ChIP-seq signal are expressed earlier in embryonic development (Harrison et al. 2011). This suggests that regions that depend on ZLD for accessibility may be enriched near genes that require ZLD for expression.

To test the functional impact of ZLD-mediated chromatin accessibility on gene expression, we performed RNA-seq from single embryos and compared stage 5 *zld*$^{M-}$ to *yw* controls. We then associated FAIRE peaks with the nearest gene and identified what percent of these FAIRE-associated genes require ZLD for robust expression. Because ZLD is known to act as a transcriptional activator (Liang et al. 2008), we focused on genes whose expression decreased by greater than two fold in the *zld*$^{M-}$ embryos. The analysis shows that 48% of genes associated with ZLD-dependent FAIRE peaks (differential, ZLD-bound) had a greater than two-fold decrease in expression in the absence of ZLD as compared to only 18% for all FAIRE peaks (Figure 2.3B). Those regions that do not depend on ZLD for accessibility (constitutive, ZLD-bound) were also enriched near genes that require ZLD for expression, but to a lesser degree than those genes near regions that require ZLD for open chromatin (differential, ZLD-bound). However, the large amount of maternally supplied RNA in the early embryo masked any changes in expression of genes transcribed both maternally and zygotically. Thus, our RNA-seq only identified changes in gene expression for those genes that are exclusively zygotically expressed. For this reason, the actual numbers of down-regulated genes in the *zld*$^{M-}$ embryos may be larger than reported. To independently

verify these results, we repeated our gene expression analysis using published microarray data (Liang et al. 2008), and similarly demonstrated that those regions that require ZLD for accessibility were enriched near genes that require ZLD for expression (Sup. Figure 2.4).

ZLD influences expression of both a small number of genes expressed early in the MZT as well as a large number of genes expressed at nuclear cycle 14. To determine if ZLD-mediated accessibility might specifically drive expression of either one of these sets of genes, we used a high-temporal resolution dataset to examine the relationship between ZLD-mediated chromatin accessibility and the timing of gene expression (Lott et al. 2011). Genes were grouped based on their timing of initial expression (Li et al. 2014), and we analyzed the correlation between ZLD-mediated changes in accessibility and ZLD-mediated changes in expression. Specifically, we determined the change in FAIRE signal between control and zld^{M-} embryos and plotted this against the change in RNA levels, as determined by RNA-seq, for the gene with the nearest associated transcription start site. There was a significant correlation between ZLD-mediated chromatin accessibility and ZLD-mediated gene expression for genes that initiate expression during nuclear cycles 10-11 (Spearman rho = 0.33, p = 0.0005), but not for genes that are expressed during late cycle 14 (Spearman rho = 0.11, p = 0.25) (Figure 2.3C,D). Furthermore, during nuclear cycles 8-12 RNA Polymerase II (Pol II) is bound only to the promoters of the early expressed genes (Chen et al. 2013; Li et al. 2014), and we demonstrated that this promoter localized Pol II is preferentially found at promoters that require ZLD for chromatin accessibility (Sup. Figure 2.5, differential, ZLD-bound). Only later, at cycles 13-14, was Pol II associated with promoters from every FAIRE class (Sup. Figure 2.5) (Chen et al. 2013). Thus ZLD-mediated chromatin accessibility likely drives the very earliest expressed zygotic genes. Later in embryonic development, ZLD may activate gene expression by a different mechanism.

H3K18ac and H4K8ac are enriched around ZLD-bound regions that require ZLD for accessibility

We recently mapped the location of nine histone modifications at four different developmental times spanning the MZT using ChIP-seq (Li et al. 2014) and demonstrated that acetylation of three different lysines on two different histones (H3K18, H3K27 and H4K8) correlated with ZLD binding, especially early in development (Li et al. 2014). Thus we speculated that ZLD-mediated chromatin accessibility might correlate with specific histone modifications. To further investigate the connection between chromatin accessibility and early gene expression, we examined the relationship between chromatin accessibility and histone modifications during the MZT. We analyzed the enrichment of acetylation and methylation marks around our four classes of FAIRE peaks (Figure 2.4, Sup. Figure 2.6). Prior to widespread zygotic genome activation (at cycles 8 and 12), H3K18ac and H4K8ac were enriched on nucleosomes surrounding the ZLD-bound regions that require ZLD for accessibility (Figure 2.4; differential, ZLD-bound). Similar enrichment was not evident at constitutively open regions or those that lack ZLD binding (Figure 2.4; constitutive, ZLD-bound; constitutive, not ZLD-bound, differential, not ZLD-bound). Thus, ZLD-mediated chromatin accessibility, and not ZLD-binding in general, is correlated with early histone acetylation. We previously demonstrated that these early acetylation marks are associated with early expressed genes (Li et al. 2014), strengthening the connection between regions that require ZLD for accessibility and early gene expression. Furthermore, H3K18ac levels are decreased in *zld*$^{M-}$ embryos (Li et al. 2014) demonstrating that ZLD is required for H3K18 acetylation at specific regions. Together with our FAIRE data, this suggests that ZLD may potentiate chromatin accessibility and early gene expression by increasing histone acetylation.

Regions that require ZLD for chromatin accessibility also require ZLD for BCD binding

One mechanism by which ZLD may activate gene expression is by facilitating transcription factor binding. We have previously demonstrated that early ZLD binding is predictive of where transcription

factors will later bind (Harrison et al. 2011), and it has been recently reported that ZLD promotes binding of the transcription factors TWI, DL and BCD to DNA (Yanez-Cuna et al. 2012; Foo et al. 2014; Xu et al. 2014). To determine if ZLD enabled binding of these proteins by creating open regions of chromatin, we reanalyzed a published ChIP-seq dataset identifying BCD-binding sites in both wild-type embryos and those lacking maternal *zld* (Xu et al. 2014). Previous analysis had shown that BCD binding is partially redistributed when ZLD is absent. Specifically, BCD binding in the mutant was decreased at a large number of sites that overlap with ZLD-bound regions (Xu et al. 2014). Comparing our four classes of peaks with the BCD ChIP-seq dataset demonstrated that BCD binds preferentially to ZLD-bound loci, both those that require ZLD for accessibility and those that do not (Figure 2.5A), and the average BCD ChIP-signal at these two classes of ZLD-bound regions is approximately equivalent (Figure 2.5A). We identified a preferential reduction in BCD binding at the differential peaks with ZLD bound (differential, ZLD-bound) as compared to the constitutively open peaks with ZLD bound (constitutive, ZLD-bound) upon ZLD depletion (Figure 2.5B). This is demonstrated by the regulatory regions of *dpn* and *sog* (Figure 2.1A). In these examples, promoter proximal FAIRE peaks do not require ZLD for accessibility or BCD binding (dark orange box), while neighboring enhancers lose both chromatin accessibility and BCD binding in the absence of ZLD (blue box). This shows that ZLD-dependent BCD binding is correlated with ZLD-mediated chromatin accessibility and not just with ZLD binding. Together this suggests that ZLD facilitates transcription factor binding through the establishment or maintenance of open chromatin.

Collective binding by a large number of transcription factors does not distinguish between ZLD-bound FAIRE classes

In addition to being associated with regions of open chromatin, another feature correlated with ZLD-bound loci is occupancy by a large number of transcription factors (Li et al. 2008; MacArthur et al. 2009; Roy et al. 2010; Kvon et al. 2012). These regions have been termed highly occupied target (HOT)

regions and have been identified in a wide-variety of organisms (Moorman et al. 2006; The ENCODE Project Consortium 2007; Li et al. 2008; MacArthur et al. 2009; Gerstein et al. 2010; Roy et al. 2010; Negre et al. 2011). Thus it was possible that ZLD-bound regions that do not require ZLD for accessibility are maintained in an open chromatin state by binding of a large number of additional factors. To determine if this was the case, we analyzed the number of transcription factors bound *in vivo* to each of the four classes of FAIRE peaks. As expected, we showed that FAIRE peaks overlapping ZLD-bound regions were enriched for binding by more than ten transcription factors (Figure 2.6A; constitutive, ZLD-bound and differential, ZLD-bound). On average there were 2.5 times as many transcription factors present at ZLD-bound FAIRE peaks than at FAIRE peaks without ZLD bound (Figure 2.6A). Thus, among regions of chromatin accessibility, ZLD binding is correlated with those loci that are bound by a large number of transcription factors. Nonetheless, we found no obvious difference in the number of transcription factors bound when we compared ZLD-bound peaks that depend on ZLD for accessibility (differential, ZLD-bound) to those that do not require ZLD (constitutive, ZLD-bound). This was also the case when we controlled for the fact that constitutive, ZLD-bound peaks are enriched for promoters as compared to differential, ZLD-bound peaks by further subdividing each class by genome annotation (Sup. Figure 2.7). Thus, these data demonstrate that binding by a large number of transcription factors is not what functionally distinguishes these classes of ZLD-bound regions and that some other feature must allow the chromatin to remain accessible in the absence of ZLD at those regions that remain constitutively open.

Constitutively open ZLD-bound loci contain GAGA motifs and are preferentially bound by GAGA factor

Because binding by a large number of transcription factors did not functionally distinguish between the two ZLD-bound FAIRE classes, we proposed that binding by a specific transcription factor might maintain chromatin accessibility at constitutive, ZLD-bound peaks in the absence of ZLD. To

determine potential factors, we identified motifs enriched in the DNA sequences underlying the FAIRE peaks. For both classes of peaks with ZLD bound (differential and constitutive), we showed, as expected, that the most highly enriched motif was the canonical ZLD-binding motif, CAGGTA (Figure 2.6B, Supplemental Table 1). For the differential ZLD-bound peaks, the next most highly enriched k-mers were overlapping or related to this canonical ZLD-binding site (Supplemental Table 1) (ten Bosch et al. 2006; Liang et al. 2008; Harrison et al. 2011; Nien et al. 2011; Struffi et al. 2011). However, for the constitutively open ZLD-bound peaks, there was a strong enrichment for GAGA motifs (Figure 2.6B, Supplemental Table 1), suggesting that these regions may be occupied by GAGA factor (GAF; encoded by the gene *Trithorax-like* (*Trl*)) in addition to ZLD (Wilkins and Lis 1997). Importantly, GAF binding is strongly correlated with chromatin accessibility, and GAGA motifs are enriched in HOT regions (Lu et al. 1993; Kvon et al. 2012; Slattery et al. 2014). Because GAF-binding motifs are known to be enriched at promoters, it was possible that the identification of GAF-binding motifs in constitutive, ZLD-bound peaks was a function of the fact that 39% of these peaks were in promoter regions as compared to only 18% for the differential, ZLD-bound peaks (Figure 2.2D). Thus, we directly determined the enrichment of the GAF-binding motif in the constitutive, ZLD-bound versus the differential, ZLD-bound peaks for promoters, intergenic regions, intronic regions, and coding regions. We found that in all cases there was a significant enrichment of the GAF-binding motif in the constitutive, ZLD-bound peaks. This enrichment was most significant at intronic and intergenic peaks. Comparing GAGA motif enrichment in constitutive versus differential ZLD-bound FAIRE peaks, we demonstrated that at intronic regions 62% of constitutively accessible ZLD-bound regions contained GAGA motifs while only 25% of the introns that depended on ZLD for accessibility (differential, ZLD-bound) contained the motif ($p = 10^{-15}$). In intergenic regions, 20% of constitutively accessible ZLD-bound regions had GAGA motifs in comparison to none of the differential, ZLD-bound intergenic peaks ($p < 10^{-8}$). These analyses demonstrate that the GAF-binding

motif is specifically enriched in the constitutive, ZLD-bound regions independent of where in the genome these peaks are located.

Consistent with this sequence enrichment, GAF binding, as assayed by ChIP-seq in 0-8 hour embryos (Negre et al. 2011), is enriched in those regions that are constitutively accessible and bound by ZLD (constitutive, ZLD-bound). Specifically, 52% (796/1537) of constitutively open ZLD-bound regions overlap GAF peaks, compared to only 15% (59/402) of regions that depend on ZLD for open chromatin (Figure 2.6C, $p < 0.0001$). Thus both the GAF-binding site and embryonic GAF binding are specifically enriched at ZLD-bound loci that remain accessible in the absence of ZLD. Furthermore, only 28% (829/2923) of constitutively open regions that do not bind ZLD overlap GAF-binding sites, indicating that, in addition to GAF, other transcription factors may be required for open chromatin. Together, these data suggest that a subset of ZLD-bound regions that do not change in accessibility in the zld^{M-} embryos remain open due to co-occupancy by GAF.

Discussion

We have used FAIRE to identify regions of open chromatin in the early embryo and determine the role of ZLD in establishing or maintaining chromatin accessibility. We demonstrate on a genome-wide level that ZLD is instrumental in defining specific regions of open chromatin. Furthermore, this ZLD-mediated chromatin accessibility dictates both transcription factor binding and early gene expression. Unexpectedly, we discovered that most open chromatin regions to which ZLD is bound do not absolutely require ZLD for chromatin accessibility. At these regions ZLD may function redundantly with GAF to determine the chromatin state. We suggest that ZLD directly mediates the very earliest gene expression by facilitating chromatin accessibility. At cycle 14, when thousands of genes are transcribed, ZLD and GAF may coordinate to determine both regions of open chromatin and levels of gene expression.

ZLD-mediated chromatin accessibility facilitates transcription-factor binding and early gene expression

ZLD is known to be instrumental in regulating expression of both the very first set of zygotic genes transcribed after fertilization as well as a large set of genes transcribed at cycle 14. ZLD is already bound to thousands of loci at cycle 10, including those that will not be activated until four nuclear cycles later during the major wave of genome activation. (Harrison et al. 2011). This suggests that early ZLD-binding is poising genes for later activation. Nonetheless, it remains unclear what differentiates the small subset of ZLD-bound loci that are transcribed early from the hundreds of ZLD-bound genes activated at cycle 14. Here we demonstrate regions that require ZLD for chromatin accessibility are correlated with the subset of genes transcribed prior to cycle 14 and with histone acetylation (Figure 2.3,4) (Chen et al. 2013; Li et al. 2014). However, not all ZLD-bound regions are equally dependent on ZLD for chromatin accessibility. We therefore propose that ZLD is essential for creating regions of open chromatin that drive expression of the subset of earliest expressed genes. This may be mediated, in

part, by local histone acetylation. At cycle 14, other factors likely function with ZLD to determine chromatin accessibility.

ZLD-mediated transcriptional activation may be potentiated by the subsequent binding of additional transcription factors. Early ZLD binding is a robust predictor of where multiple additional transcription factors will later bind (Harrison et al. 2011). More recently it has been shown that ZLD is required for the DNA binding of three different transcription factors: TWI, DL, and BCD (Yanez-Cuna et al. 2012; Foo et al. 2014; Xu et al. 2014). Additionally, transgenic versions of the *brinker (brk)* and *sog* enhancers show a correlation between the number of ZLD-binding sites and both DL binding and DNase I accessibility (Foo et al. 2014). Thus, prior work has clearly demonstrated a role for ZLD in mediating transcription factor binding, but the mechanism by which ZLD served this function was unclear. Here we demonstrate that BCD binding is lost in zld^{M-} embryos preferentially at those regions that depend on ZLD for chromatin accessibility. These data show that ZLD potentiates transcription-factor binding through the establishment or maintenance of open chromatin, and this is likely to be important for ZLD-mediated transcriptional activation (Figure 2.7).

The mechanism by which ZLD establishes or maintains chromatin accessibility remains unknown. Unlike the pioneer factor FoxA1, which can open chromatin by binding chromatin through a winged-helix domain (Cirillo et al. 1998; Cirillo et al. 2002), the ZLD DNA-binding domain does not resemble that of a linker histone. Instead, ZLD binds DNA through a cluster of four zinc fingers in the C-terminus (Struffi et al. 2011; Hamm et al. 2015). In addition, ZLD is a large protein with no recognizable enzymatic domains that activates transcription through a low-complexity protein domain (Hamm et al. 2015). Thus, ZLD likely facilitates open chromatin through interactions with cofactors, and it is possible that recruitment of different cofactors to distinct ZLD-bound loci could partially explain the differential requirement on ZLD for chromatin accessibility in the early embryo.

ZLD binding but not ZLD-mediated chromatin accessibility is a defining feature of HOT regions

HOT regions, loci that are bound by a large number of different transcription factors, have been identified in multiple organisms, including worms, flies and humans (Moorman et al. 2006; The ENCODE Project Consortium 2007; Li et al. 2008; MacArthur et al. 2009; Gerstein et al. 2010; Roy et al. 2010; Negre et al. 2011). Unexpectedly, these HOT regions are not strongly enriched for the DNA-sequence motifs bound by the transcription factors that define them (Moorman et al. 2006; MacArthur et al. 2009; Gerstein et al. 2010; Roy et al. 2010). Instead, HOT regions are associated with open chromatin, suggesting that chromatin accessibility along with sequence motif enrichment drives the high transcription factor occupancy (Roy et al. 2010). In *Drosophila*, HOT regions are enriched for developmental enhancers that contain the canonical ZLD-binding site, CAGGTAG, as well as for *in vivo* ZLD binding (Li et al. 2008; MacArthur et al. 2009; Harrison et al. 2011; Nien et al. 2011; Kvon et al. 2012). By analyzing the 5000 regions with the highest FAIRE signal, we demonstrate that high transcription factor occupancy is correlated with ZLD-bound regions of accessible chromatin and not with open chromatin more generally (Figure 2.6A). Furthermore, this association was not specific for those regions that require ZLD for accessibility. Thus, HOT regions overlap with ZLD-bound regions of open chromatin regardless of whether these loci require ZLD for accessibility. Our data suggest that, while ZLD-mediated chromatin accessibility may facilitate gene expression, it is not this function of ZLD alone that defines HOT regions.

The open chromatin landscape and transcriptional profile of the embryo are likely defined by the functions of both GAGA factor and ZLD

Our FAIRE data showed that more than 400 regions are bound by ZLD and require ZLD for chromatin accessibility. However, at least three times as many regions are bound by ZLD, but remain open even in its absence. Our data predict GAF functions at many of these constitutively open

chromatin regions to maintain chromatin accessibility, even in the absence of ZLD (Figure 2.7). Along with the CAGGTAG element, GAF-binding motifs are enriched in HOT regions (Kvon et al. 2012; Slattery et al. 2014). Like ZLD, GAF is maternally deposited into embryos. Furthermore, GAF is known to facilitate nuclease-hypersensitive regions and interact with members of the NURF ATP-dependent chromatin-remodeling complex (Lu et al. 1993; Tsukiyama and Wu 1995).

Our data show that at early expressed genes there is a correlation between regions that require ZLD for chromatin accessibility (differential, ZLD-bound) and ZLD-dependent gene expression (Figure 2.3C; cycles 10-11). However, this association is not found for genes expressed during cycle 14 (Figure 2.3D). Instead, our data suggest that at loci associated with this later gene expression, GAF is functioning together with ZLD to regulate chromatin accessibility and gene expression. Maternally deposited GAF is required for robust transcription and nuclear divisions during the MZT (Bhat et al. 1996; Lagha et al. 2013). GAF is thought to mediate transcription, at least in part, through a role in the establishment of poised polymerase (Shopland et al. 1995; Bhat et al. 1996; Lee et al. 2008). The fact that poised polymerase is not established until cycle 13 (Chen et al. 2013; Blythe and Wieschaus 2015), supports our model that GAF is required specifically for gene expression at cycles 13-14. Thus, we suggest that ZLD-dependent early embryonic enhancers may be unique in that they rely only on ZLD for chromatin accessibility. Although there are likely additional factors involved, our data demonstrate that later in development ZLD and GAF likely function together to define the chromatin landscape of the early embryo.

Pioneer factors as drivers of zygotic genome activation

Pioneer factors are a specialized class of transcription factors that bind nucleosomal DNA and initiate chromatin remodeling, allowing the recruitment of additional transcription factors (Iwafuchi-Doi and Zaret 2014). ZLD-binding is strongly driven by DNA sequence, much more so than the binding of

other transcription factors (Harrison et al. 2011). This observation combined with our FAIRE data and analyses demonstrates that ZLD exhibits many of the characteristics of a pioneer factor as defined by Iwafuchi-Doi and Zaret (2014): 1) engaging chromatin prior to gene activity; 2) establishing or maintaining chromatin accessibility to facilitate transcription factor binding; and 3) playing a primary role in cell reprogramming. Additional properties have been shown for classical pioneer factors, including remaining bound to the mitotic chromosomes (i.e. bookmarking) and binding to nucleosomal DNA (Cirillo et al. 2002; Sekiya et al. 2009; Kadauke et al. 2012; Caravaca et al. 2013). It will be important to determine whether ZLD shares these characteristics with other pioneer factors.

Pioneer factors, such as FoxA1, can bind to closed chromatin and subsequently increase accessibility of the target site (Cirillo et al. 2002). However the chromatin of the early embryo may provide a unique environment with little compacted chromatin. Heterochromatin formation is not observed until the 14th nuclear cycle (Foe and Alberts 1983). Chromatin bound H3 levels increase through the MZT, and histone modifications indicative of silent genes, such as H3K27 trimethylation, are not evident until there is widespread activation of the zygotic genome (Li et al. 2014). Thus, while ZLD binds to genes prior to zygotic genome activation this activity may not require binding to compacted chromatin. It may be that ZLD is distinctive in the timing of its expression rather than in its chromatin-binding properties and that the sequence-driven binding of ZLD is a property of the open chromatin and rapid nuclear divisions that characterize the earliest stages of embryonic development.

Despite the fact that we have demonstrated a critical role for ZLD in determining chromatin accessibility at hundreds of genomic regions, our data show that this role is limited to specific regions associated with the earliest-expressed embryonic genes. Other factors, such as GAF, likely work redundantly with ZLD to define chromatin accessibility during the MZT. The coordinated function of multiple factors in determining chromatin structure and genome activation is not without precedent. It has recently been demonstrated homologs of the core pluripotency factors, Nanog, Pou5f3 (also known

as Pou5f1 and Oct 4), and Sox19B (a member of the SoxB1 family), act analogously to ZLD during the zebrafish MZT to drive genome activation (Lee et al. 2013; Leichsenring et al. 2013). Furthermore, Oct 4 and Sox2 are known to be pioneer factors instrumental in reprogramming differentiated cells to a pluripotent state (Soufi et al. 2012; Soufi et al. 2015). Together, these data suggest that chromatin remodeling in the early embryo requires the function of multiple factors, and this activity facilitates the transition from the specified germ cells to the pluripotent cells of the early embryo.

Materials and Methods

Fly Strains

Embryos lacking maternally supplied ZLD were generated as described by Liang et al. (2008). *zld*^{294} *FRT19A/FM7* females were crossed with *y w sn P{mini w+, ovo*$^{D1\text{-}26}$*}*^{25} *FRT19A, hsFLP122/Y* males and allowed to lay for 24 hours. The resulting offspring were heat-shocked twice for 30 minutes at 37°C with a 24 hour interval between (at ~24-48 hours and 48-72 hours after laying) to generate germ-line clones. In parallel, *y w FRT19A* females were crossed to *y w sn P{mini w+, ovo*$^{D1\text{-}26}$*}*^{25} *FRT19A, hsFLP122/Y* males, and the offspring were heat-shocked to serve as paired controls. Since only females with the desired germ-line clones should produce embryos, non-heat-shocked offspring from each cross were preserved and allowed to mature to check for escapers. Any crosses that produced embryos without heat-shock (escapers) were discarded.

Immunostaining

Embryos depleted for maternal *zld* and paired *yw* controls were dechorionated and fixed in 3 ml fixation buffer (1.3 X PBS, 67 mM EGTA pH 8), 1 ml 37% formaldehyde and 5 ml heptanes for 25 minutes. Vitelline membranes were removed with methanol. Embryos were rehydrated and stained using anti-ZLD antibodies (Harrison et al. 2011) along with DM1A anti-tubulin (Sigma, St. Louis, MO) as a positive control. Goat anti-rabbit Alexa-488 and goat anti-mouse Alexa-688 were used as secondary antibodies (Life Technologies, Grand Island, NY).

FAIRE

Embryos ranging in age from 2 to 3 hours were collected, dechorionated, and crosslinked for 10 min with 3:1 heptane:fix solution at room temperature. Following washing, embryos were flash frozen in liquid nitrogen and stored at -80°C. FAIRE was performed on approximately 0.1 g of embryos, as

previously described in McKay and Lieb (2013) with the following exceptions. The homogenized embryos were filtered through Miracloth (EMD Millipore, Billerica, MA). After resuspension in Lysis Buffer (10 mM Tris-Cl pH 8, 100 mM NaCl, 1 mM EDTA, 2% Triton X-100, 1% SDS), the nuclei were subjected to four rounds of bead beating with 0.5 mm glass beads. The chromatin was sonicated three times for 30 seconds using a Branson Sonifier 250 at a power setting of 1.5 to achieve a size range of 200bp-1kb.

Single-embryo RNA preparation

Embryos depleted for maternal *zld* and paired *yw* controls were dechorionated and analyzed under halocarbon oil to determine stage. Individual paired stage 5 embryos were selected and oil was removed. Embryos were lysed in TRIzol (Life Technologies, Grand Island, NY) supplemented with 150 µg/ml glycogen, and RNA was prepared according to the manufacturer's instruction.

Sequencing and Data Analysis

High-throughput sequencing libraries were prepared by the University of Wisconsin Biotechnology Center DNA Sequencing Facility (FAIRE) and by Jacqueline Villalta (RNA). Libraries were prepared from both FAIRE-enriched DNA and from sonicated genomic DNA to serve as input material. The libraries were sequenced using the Illumina HiSeq 2000 platform. The total reads for *yw* were 42524780 and for were *zld*$^{M-}$ 50462591. The sequence data was processed essentially as previously described (McKay and Lieb 2013). Briefly, reads were aligned to the genome using Bowtie 2 (Langmead and Salzberg 2012) with the --sensitive and --end-to-end flags. Aligned reads with a MAPQ score less than 10 were removed using SAMtools (Li et al. 2009). Finally, reads were extended to 110bp. Since replicate data were highly correlated, reads from each replicate were pooled for further analysis, resulting in 23646263 reads for *yw* and 13371214 reads for *zld*$^{M-}$. This corresponds to 92% of the euchromatic region covered by greater than one read for *yw* and 87% for *zld*$^{M-}$. FAIRE peaks were called

with MACS2 (Feng et al. 2012) using a shift size of 125bp. Differential FAIRE peaks between *yw* and *zld*$^{M-}$ were identified with edgeR (Robinson et al. 2010), using the "classic analysis". The input file consisted of a table of FAIRE read counts for each *yw* and *zld*$^{M-}$ replicate intersecting the union set of the top 5000 FAIRE peaks. FAIRE peaks with a p-value lower than 0.05 were called as differential.

Data for BCD binding in wild-type and *zld*$^{M-}$ embryos were from GSE55256 (Xu et al. 2014). GAF binding was from GSE23537 (Negre et al. 2011). Histone modification data was obtained from GSE58935 (Li et al. 2014). Microarray expression data for wild-type and *zld*$^{M-}$ embryos were from GSE11231 (Liang et al. 2008). ZLD binding sites were from GSE30757 (Harrison et al. 2011). Data for RNA Polymerase II binding during the MZT were obtained from GSE41703 (Chen et al. 2013).

Peak classification

FAIRE peaks were classified into four functional groups based on their location. Peaks located -500 to +150 bases from a known transcription start site were classified as “Promoter” peaks. Peaks overlapping coding regions were classified as “Genic” or “Intronic”, and the remaining peaks classified as “Intergenic”.

Motif Analysis

The DNA sequence in FAIRE peaks was analyzed using SeedSearcher software by Yoseph Barash and colleagues (Barash 2005). Specifically, we enumerated over all possible k-mers (ranging from length 3 to 8, while allowing one or two wildcards in specific positions). Each k-mer was then scored using a differential hypergeometric score, comparing the number of motif occurrences within a set of peaks to the overall number of occurrences within a background set of control sequences. Here we compared each group of FAIRE peaks (constitutive or differential, with or without in vivo ZLD or GAF binding) to the

remaining of the 5000 high-accessibility *yw* FAIRE peaks. All reported k-mers are highly significant after corrections for multiple hypothesis testing.

Data access

The FAIRE-seq and RNA-seq data from this study have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE65837.

Acknowledgments

We thank Michael Eisen for assistance with the single-embryo RNA-seq and helpful discussions. T.K. is a member of the Israeli Center of Excellence (I-CORE) for Gene Regulation in Complex Human Disease (no. 41/11) and the Israeli Center of Excellence (I-CORE) for Chromatin and RNA in Gene Regulation (no. 1796/12). K.S. was supported by the NIH National Research Service Award T32 GM07215. This work was supported in part by a Basil O'Connor Starter Scholar Research Award (#5-FY14-29) to M.M.H. and grant 1R01GM111694 from the National Institute of General Medical Sciences to M.M.H.

Figures

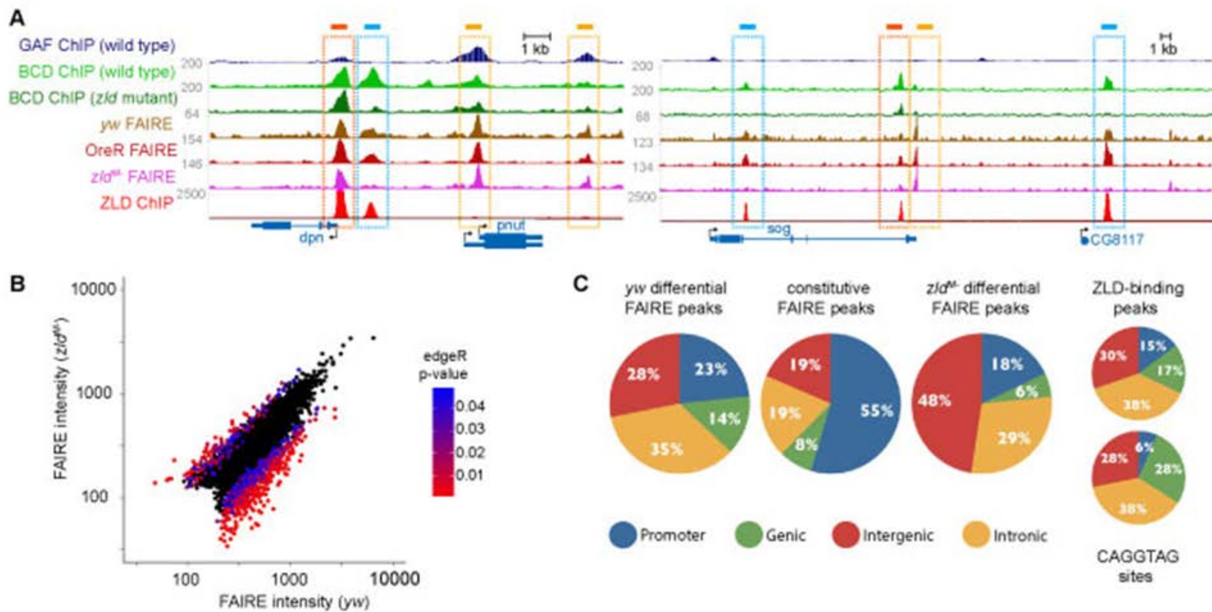


Figure 2.1: ZLD is required to maintain or establish discrete regions of open chromatin.

(A) Normalized FAIRE-seq and ChIP-seq read profiles as labeled on the left for two genomic regions. FAIRE data for *yw* and zld^{M-} embryos (this study). FAIRE data for 2-4 hour OreR (wild-type) embryos from McKay and Lieb (2013). ChIP data for ZLD from Harrison et al. (2011), for BCD from Xu et al. (2014) and for GAF from Negre et al. (2011). Genes are shown at the bottom with arrows to indicate direction of transcription. Boxes indicate different classes of FAIRE peaks: blue box, differential, ZLD-bound; dark orange box, constitutive, ZLD-bound; light orange box, constitutive, not ZLD-bound. **(B)** Scatter plot of the FAIRE signal from *yw* embryos versus the FAIRE signal from zld^{M-} embryos. Black dots indicate the union set of regions identified by FAIRE in both *yw* and zld^{M-} embryos. Colored dots indicate differential peaks identified by edgeR. Colors represent p-vales (< 0.05) as indicated by the scale. **(C)** Pie charts showing the genomic distribution of FAIRE peaks, ZLD-binding sites and CAGGTAG motifs to promoters, genes, introns and intergenic regions as shown.

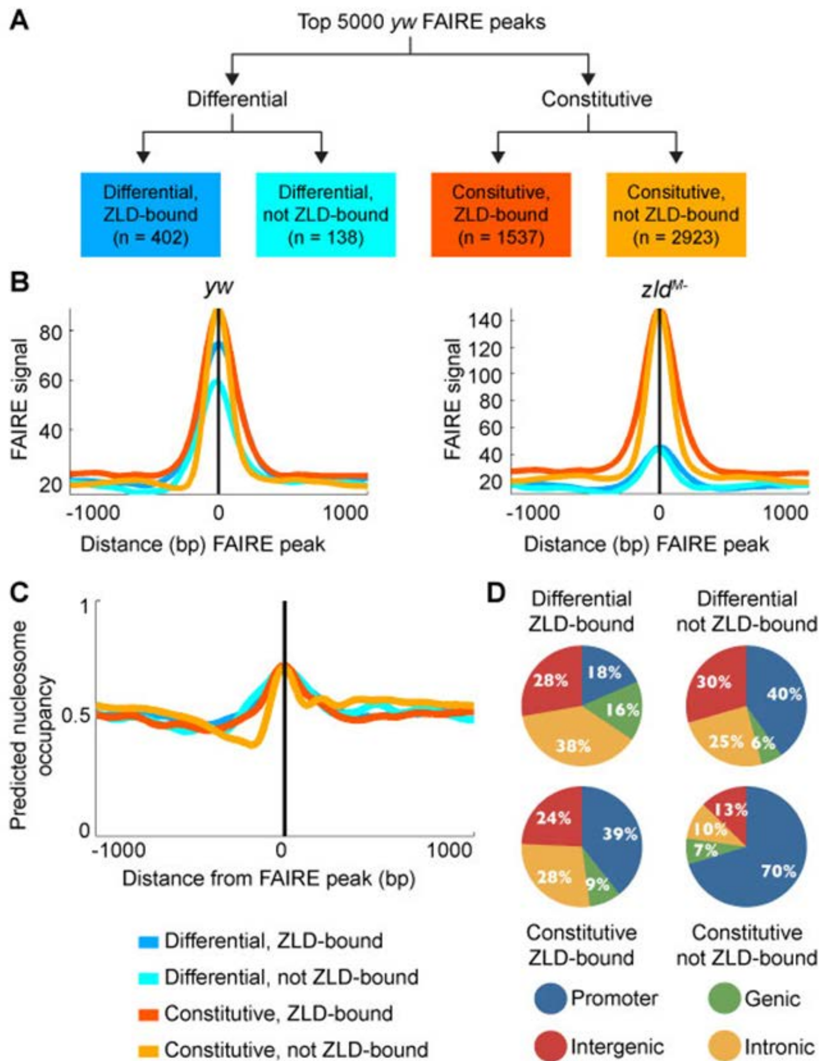


Figure 2.2: Many loci bound by ZLD in wild-type embryos remain accessible in *zld^{M}*- embryos.

(A) Flow chart illustrating the four classes of *yw* FAIRE peaks used for future analysis. **(B)** Average FAIRE signal for each class of FAIRE peaks in both *yw* and *zld^{M-}* embryos. **(C)** Predicted nucleosome occupancy based on DNA sequence surrounding FAIRE peaks in each of the indicated classes. Colors indicate the different FAIRE classes. Plots are centered on the FAIRE peak midpoint and oriented relative to the transcription start site of the nearest gene. **(D)** Pie charts showing the genomic distribution to promoters, genes, introns and intergenic regions for each class of FAIRE peaks.

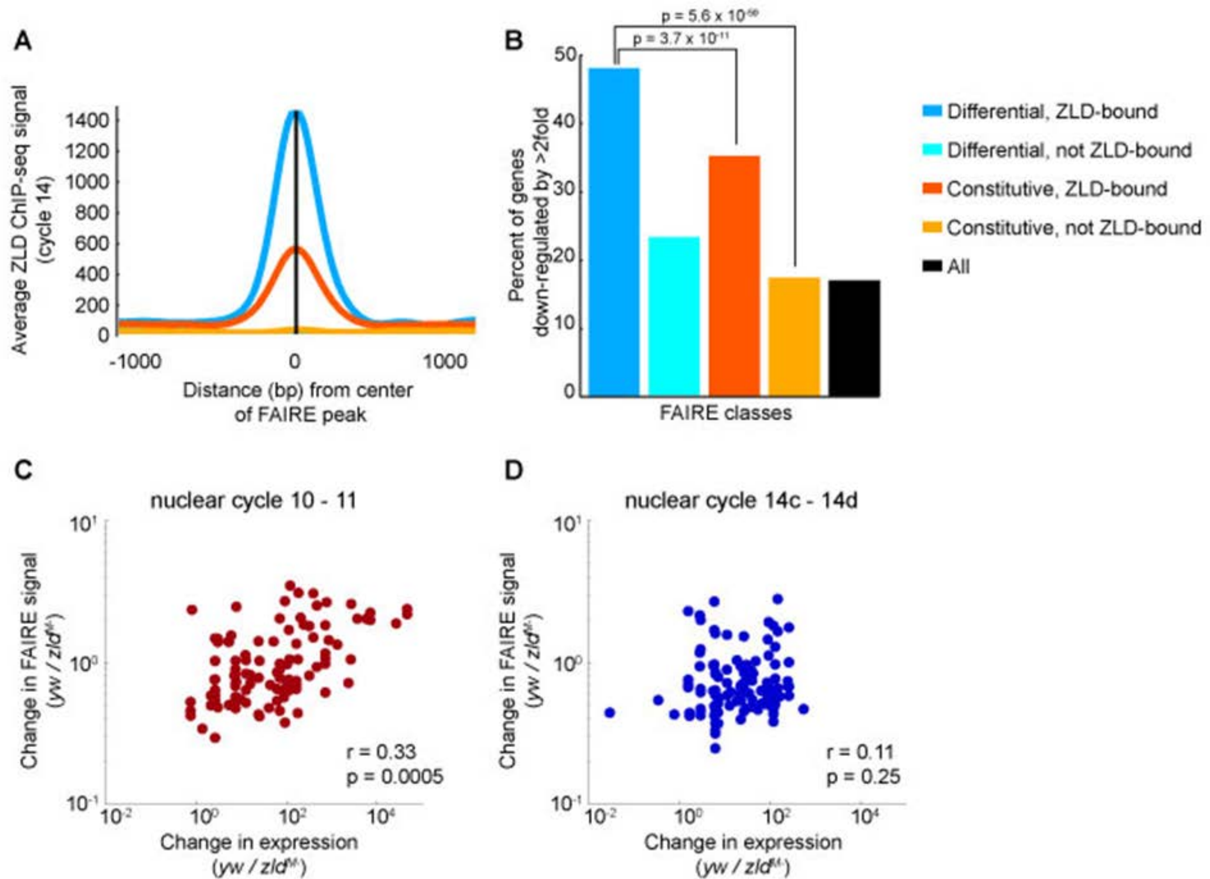


Figure 2.3: ZLD-mediated chromatin accessibility is correlated with early ZLD-dependent gene expression.

(A) Average ChIP-seq signal for ZLD at cycle 14 for each class of FAIRE peaks. **(B)** FAIRE-enriched sites were associated with the nearest gene. For each associated gene the expression change when maternal *zld* was depleted was identified by RNA-seq. The percent of genes that show a greater than two fold decrease in expression are plotted for each class of FAIRE peaks. Hypergeometric p-values are shown. **(C, D)** Correlation between ZLD-mediated changes in gene expression and FAIRE signal for 83 genes expressed at nuclear cycles 10-11 **(C)** or 91 genes expressed at late nuclear cycle 14 **(D)**. r = Spearman's correlation coefficient.

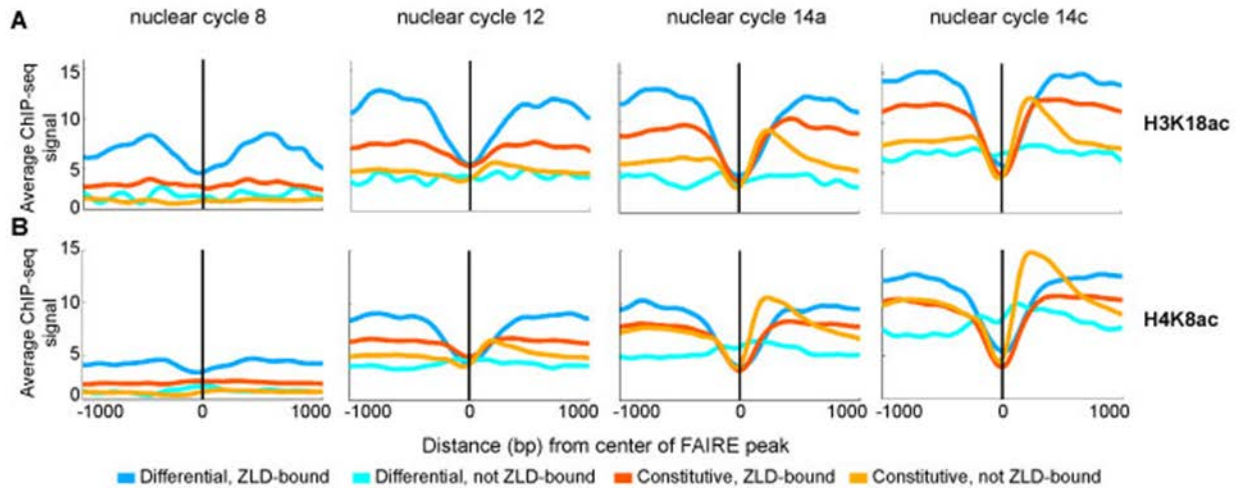


Figure 2.4: Early histone acetylation is enriched around loci that require ZLD for chromatin accessibility.

(A-B) Distribution of ChIP-seq signals for H3K18ac **(A)** and H4K8ac **(B)** surrounding individual classes of FAIRE peaks at specific nuclear cycles during the early stages of embryonic development. Colors indicate the different FAIRE classes. Plots are centered on the FAIRE peak and oriented relative to the transcription start site of the nearest gene.

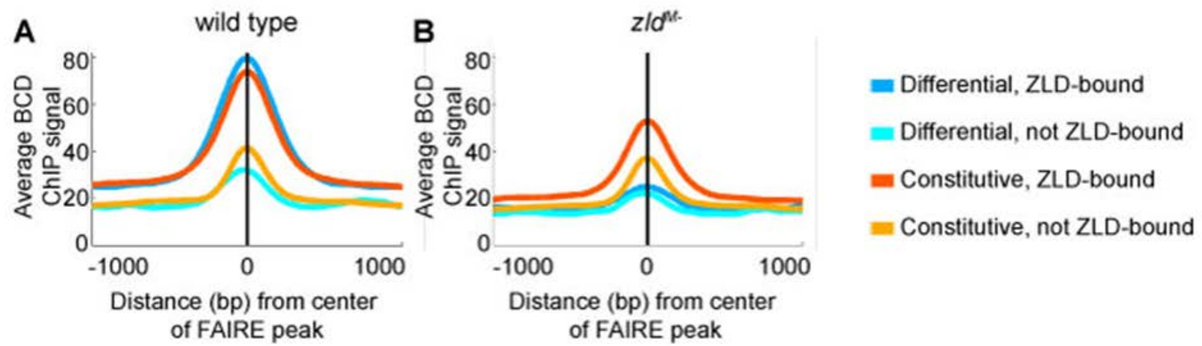


Figure 2.5: In the absence of ZLD, BCD binding is preferentially lost at those regions that lose chromatin accessibility.

(A-B) Average BCD binding in wild-type embryos **(A)** and embryos lacking maternal ZLD **(B)** for each class of FAIRE peaks. Colors indicate the different FAIRE classes.

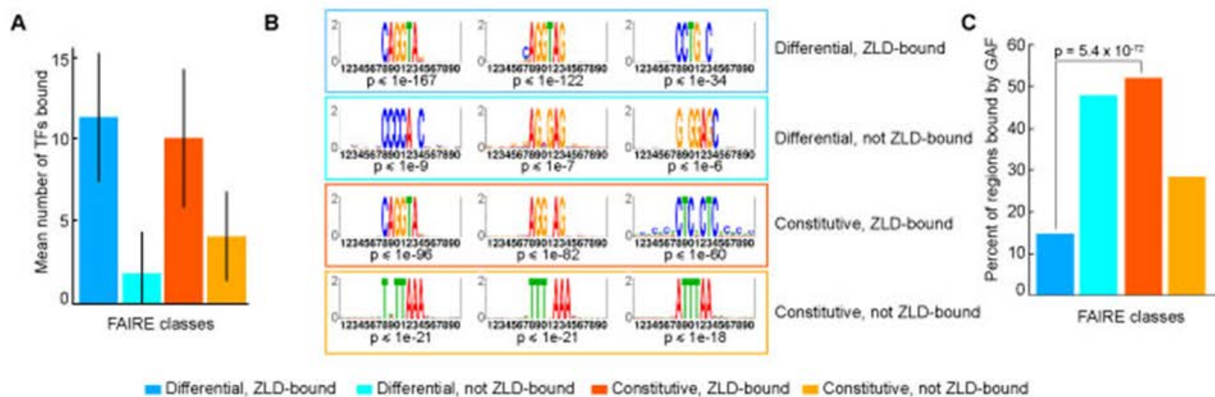


Figure 2.6: GAF likely functions with ZLD to define chromatin accessibility in the early embryo.

(A) Mean number of transcription factors (TFs) bound to each class of FAIRE peaks. Colors indicate the different FAIRE classes. Error bars indicate standard deviation. **(B)** Position weight matrices for the motifs enriched in the FAIRE peaks for each class as compared to all of the top 5000 FAIRE peaks identified in *yw* embryos. Hypergeometric p-values are shown below each motif. The top three motifs for each class are shown. **(C)** Percent of FAIRE peaks in each class that overlap a GAF-binding site as identified by ChIP-seq in 0-8hr embryos (Negre et al. 2011). Hypergeometric p-values are shown.

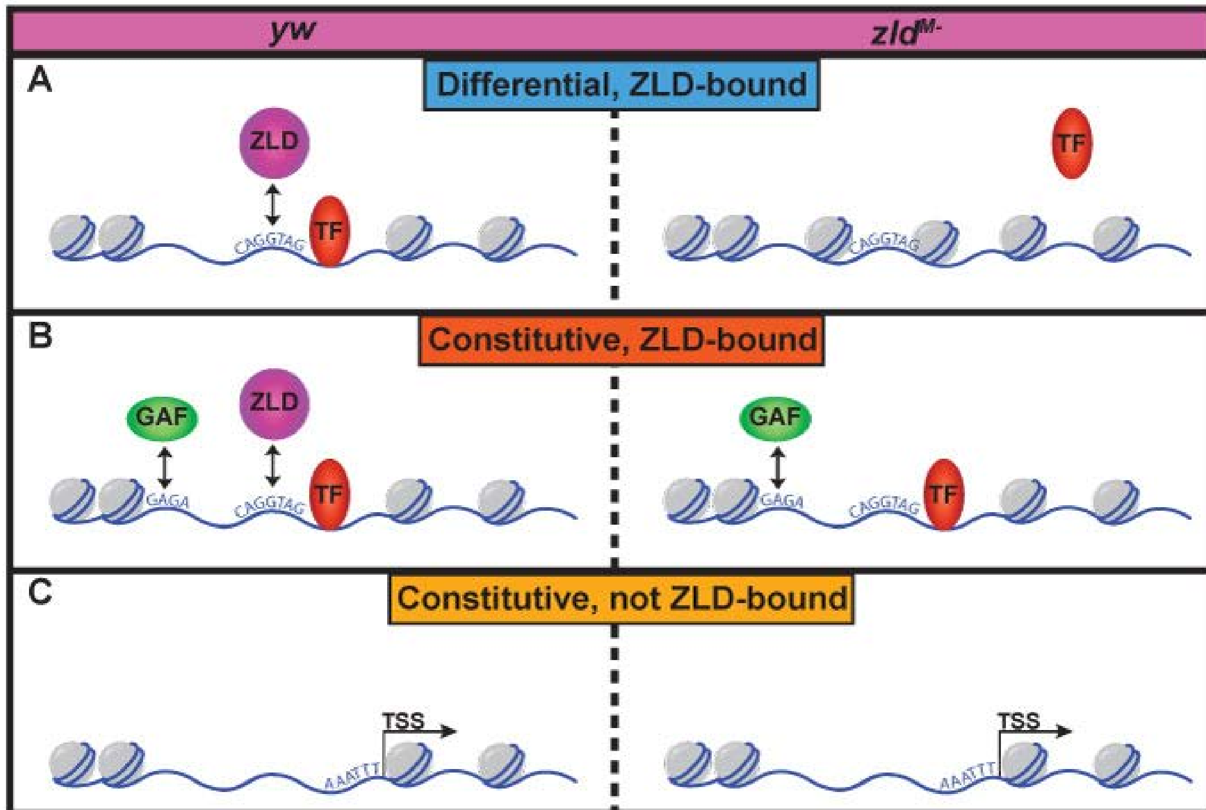
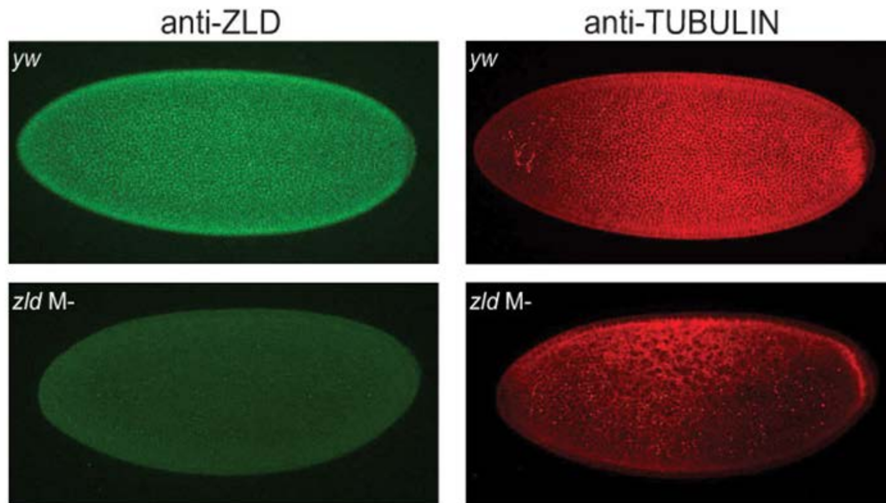


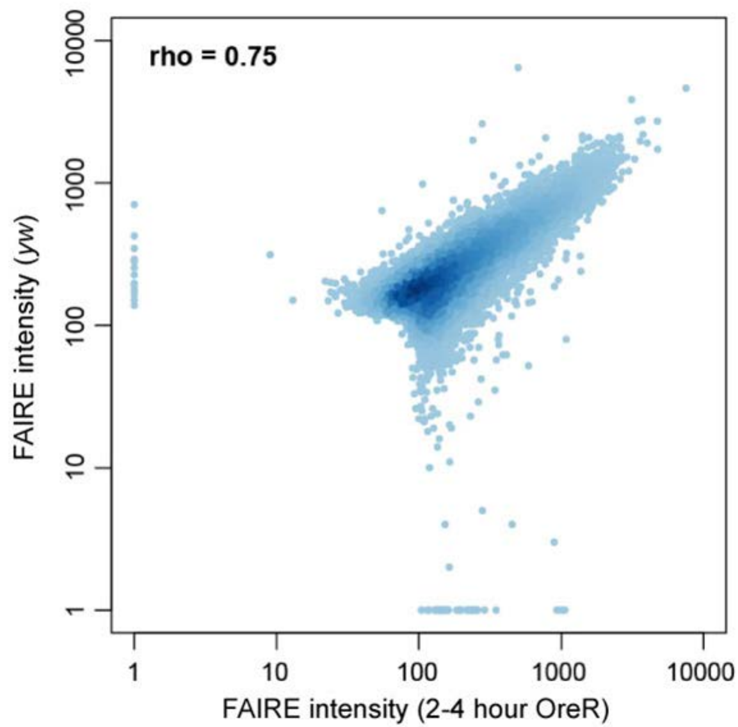
Figure 2.7: Model for the role of ZLD in defining chromatin accessibility at the MZT.

(A) ZLD determines chromatin accessibility at early embryonic enhancers, allowing other transcription factors (TF) access to their binding sites. **(B)** At a large subset of regions, ZLD binds chromatin, but is not required for accessibility. GAF facilitates open chromatin at regions that remain accessible in *zld*$^{M-}$ embryos, allowing transcription factors to bind to these loci. **(C)** Promoter regions are enriched among those regions that do not require ZLD for accessibility.



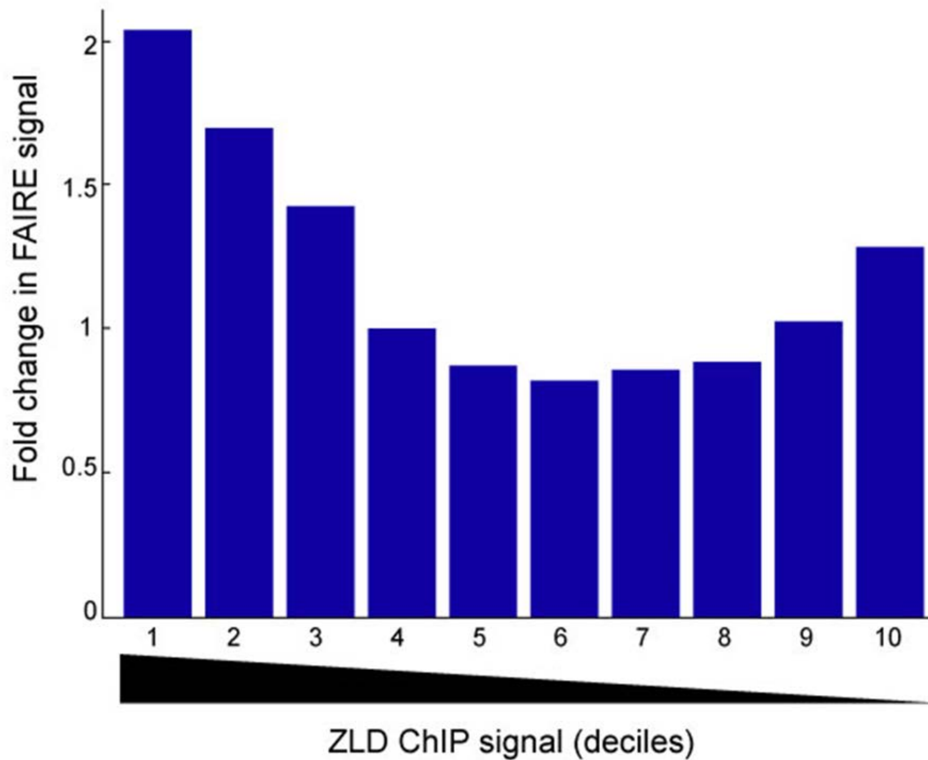
Sup. Figure 2.1: ZLD protein is not detected in stage 5 embryos depleted of maternal *zld*.

A stage 5 *yw* control embryo and an embryo depleted of maternal *zld* (*zld*$^{M-}$) stained with anti-ZLD (green) and anti-tubulin antibodies (red) as a staining control.



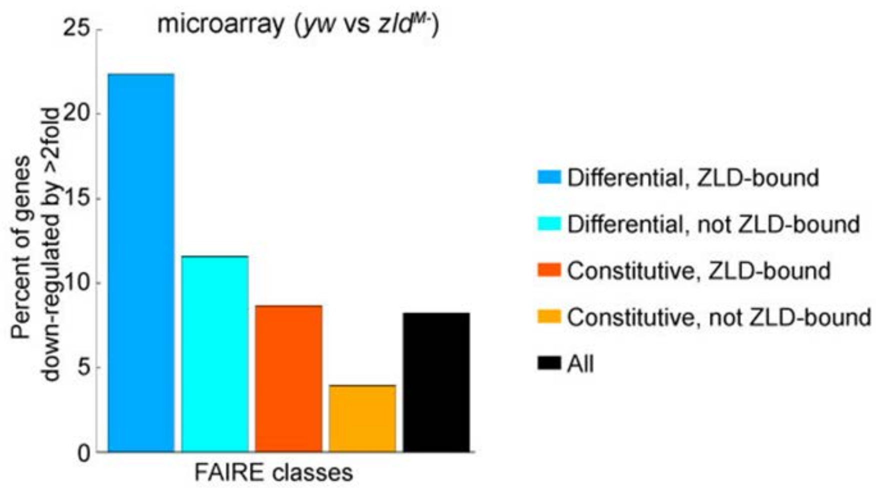
Sup. Figure 2.2: FAIRE-seq data from 2-3 hour *yw* embryos are similar to previously published FAIRE-seq data from 2-4 hour OreR embryos.

Scatter plot of the FAIRE signal from 2-3 hour *yw* embryos versus the FAIRE signal from 2-4 hour OreR embryos (McKay and Lieb 2013). Spearman's rho = 0.75.



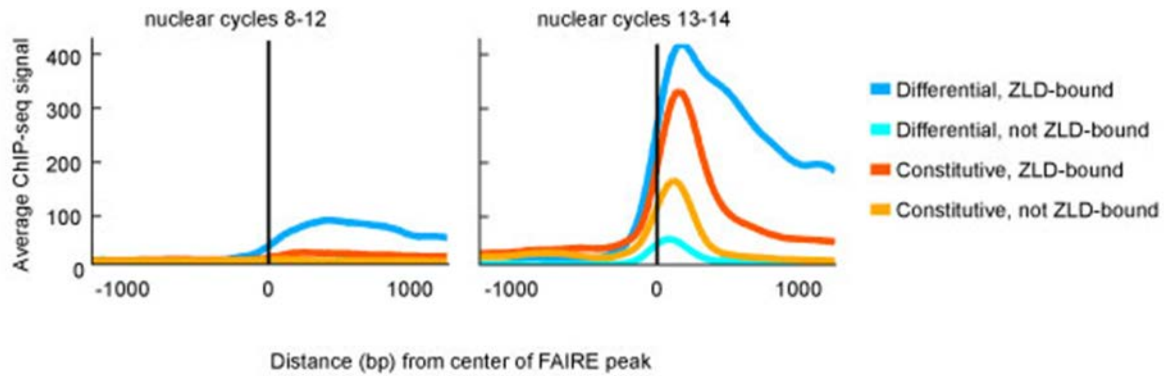
Sup. Figure 2.3: Levels of ZLD binding, as assayed by ChIP-seq signal, are correlated with a dependence on ZLD for chromatin accessibility.

ZLD ChIP-seq peaks were binned into ten equally sized groups based on the peak height. For each decile, the fold change in FAIRE signal was calculated by dividing the average FAIRE z-score from the *yw* control embryos by the average FAIRE z-score in the zld^{M-} embryos. z-scores were used to allow quantitative comparisons between the two data sets. Similar results were obtained when the raw FAIRE signal was used.



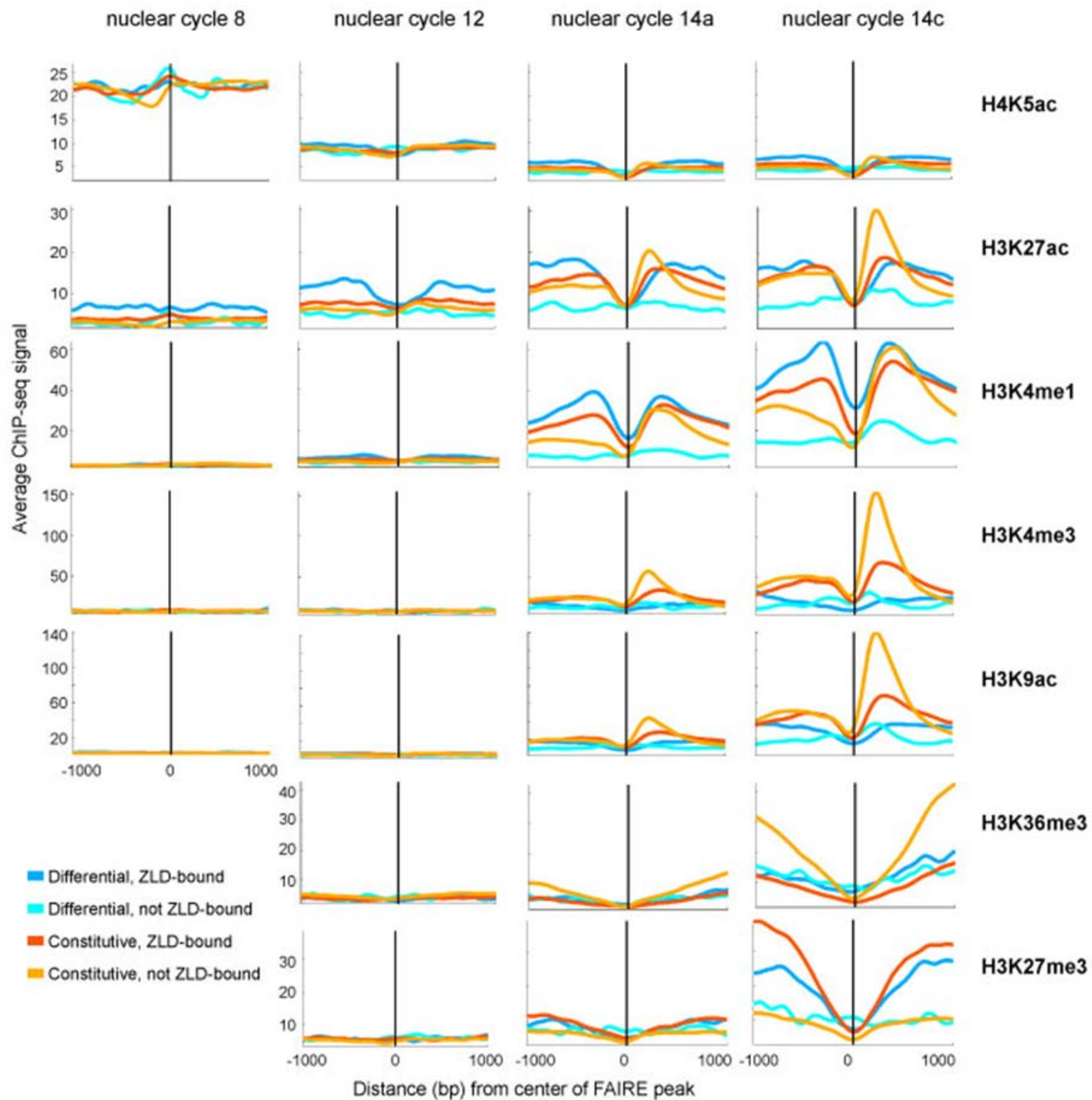
Sup. Figure 2.4: FAIRE-seq data from 2-3 hour *yw* embryos are similar to previously published FAIRE-seq data from 2-4 hour OreR embryos.

Published microarray data (Liang et al., 2008) was used to determine ZLD-dependent gene expression.



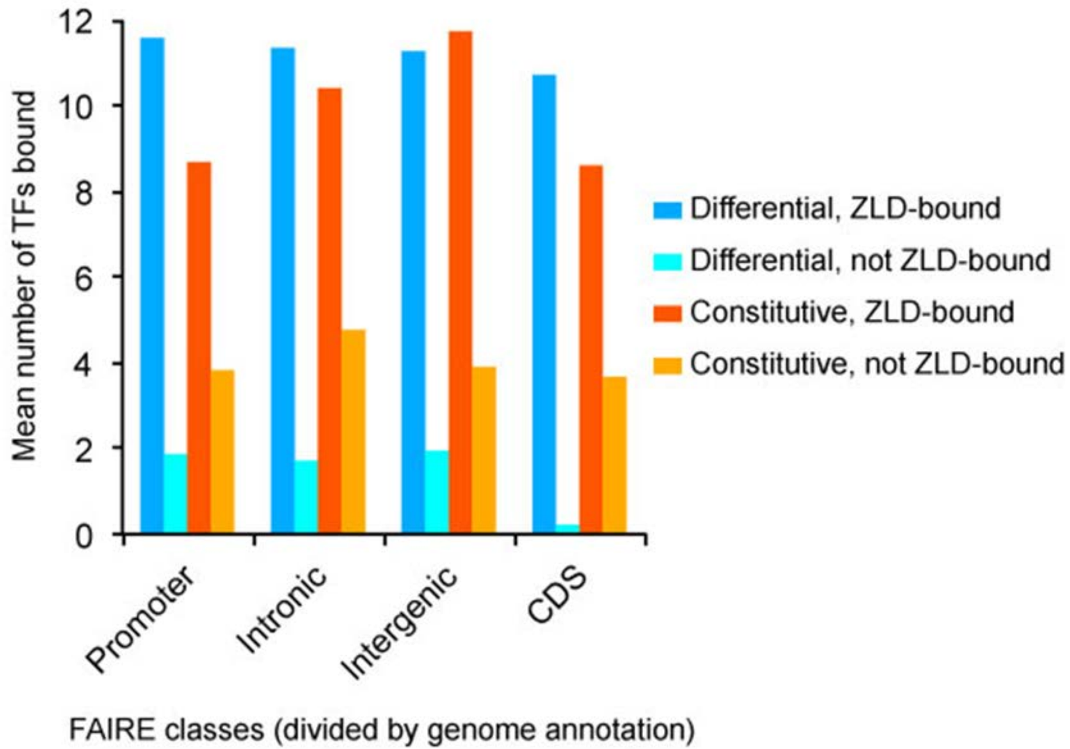
Sup. Figure 2.5: Early RNA Polymerase II binding is associated with promoters that require ZLD for chromatin accessibility.

Plot of average ChIP-signal for RNA Pol II at promoters at two different developmental time points spanning the MZT. Colors indicate the different FAIRE classes. Plots are centered on the FAIRE peak and oriented relative to the transcription start site of the nearest gene.



Sup. Figure 2.6: Histone modifications associated with regions of open chromatin.

Distribution of ChIP-seq signals for histone marks indicated on the right surrounding individual classes of FAIRE peaks at specific nuclear cycles during the early stages of embryonic development. Colors indicate the different FAIRE classes. Plots are centered on the FAIRE peak and oriented relative to the transcription start site of the nearest gene.



Sup. Figure 2.7: ZLD binding, and not a dependence on ZLD for chromatin accessibility, correlates with the mean number of transcription factors bound to the regions, regardless of genomic location.

Mean number of transcription factors (TFs) bound to each class of FAIRE peaks subdivided by genome annotation. Colors indicate the different FAIRE classes.

Table 2.1

Enrichment of ZLD-binding motifs and *in vivo* ZLD-binding sites (with ZLD) in FAIRE peaks. Without ZLD indicates no *in vivo* bound ZLD was identified by ChIP-seq (Harrison et al. 2011).

		CAGGTAG		CAGGTA		Total
yw differential	Total	45%	244	67%	361	540
	with ZLD	61%	244	87%	351	402
	without ZLD	0%	0	7%	10	138
constitutive	Total	8.3%	370	20%	900	4460
	with ZLD	21%	321	45%	692	1537
	without ZLD	1.7%	49	7%	208	2923
zld^{M-} differential	Total	0.7%	1	6%	9	145
	with ZLD	0%	0	23%	5	22
	without ZLD	1%	1	3%	4	123

Sup. Table 2.1

Motifs enriched in individual classes of FAIRE peaks with p-values and descriptions of the motif.

differential, ZLD-bound	Motif	p-value	Description
	CAGGTA	1 e-167	ZLD binding
	AGGTAG	1 e-122	
	CCTGNC	1 e-34	
	CCTGNA	1 e-29	
	CCTNCC	1 e-28	
	GNTACC	1 e-25	
	GATCC	1 e-23	
differential, not ZLD-bound			
	CCCCANC	1 e-9	
	GNGGAGC	1 e-8	
	AGNGAG	1 e-6	
constitutive, ZLD-bound			
	CAGGTA	1 e-96	ZLD binding
	AGGNAG	1 e-82	
	CTCNCTC	1 e-61	GAGA binding
	CTCTNTC	1 e-54	
	CNCTCTC	1 e-53	
	AGNGAGA	1 e-51	
	GAGANA	1 e-48	
constitutive, not ZLD-bound			
	TNTTAAA	1 e-22	AT-rich
	TTTXAAA	1 e-21	
	ATTTAA	1 e-18	
	TNAATAA	1 e-14	
	TANTTTA	1 e-12	
	ANTATT	1 e-10	
differential vs. constitutive			
	CAGGTA	1 e-105	ZLD binding
	AGGTAG	1 e-84	
	ACCTGC	1 e-34	
	AGANCC	1 e-24	
	CCTGNCC	1 e-22	
	CCCAG	1 e-20	
constitutive vs. differential			
	TNTTAAA	1 e-20	AT-rich
	ANAAAAA	1 e-17	
	AANTAAA	1 e-17	
	AAAAAT	1 e-16	
	TAAATA	1 e-15	
	AAAATA	1 e-15	
	TNTAAAA	1 e-15	
	TTNTAAA	1 e-15	
	ATTTAA	1 e-13	

References

Barash Y. 2005. Unified Models for Regulatory Mechanisms. PhD Thesis. Jerusalem: Hebrew University.

Bhat KM, Farkas G, Karch F, Gyurkovics H, Gausz J, Schedl P. 1996. The GAGA factor is required in the early Drosophila embryo not only for transcriptional regulation but also for nuclear division. Development 122: 1113-1124.

Blythe SA, Wieschaus EF. 2015. Zygotic genome activation triggers the DNA replication checkpoint at the midblastula transition. Cell 160: 1169-1181.

Caravaca JM, Donahue G, Becker JS, He X, Vinson C, Zaret KS. 2013. Bookmarking by specific and nonspecific binding of FoxA1 pioneer factor to mitotic chromosomes. Genes Dev 27: 251-260.

Carr A, Biggin MD. 1999. A comparison of in vivo and in vitro DNA-binding specificities suggests a new model for homeoprotein DNA binding in Drosophila embryos. EMBO J 18: 1598-1608.

Chen K, Johnston J, Shao W, Meier S, Staber C, Zeitlinger J. 2013. A global change in RNA polymerase II pausing during the Drosophila midblastula transition. eLife 2: e00861.

Chou TB, Noll E, Perrimon N. 1993. Autosomal P[ovoD1] dominant female-sterile insertions in Drosophila and their use in generating germ-line chimeras. Development 119: 1359-1369.

Cirillo LA, Lin FR, Cuesta I, Friedman D, Jarnik M, Zaret KS. 2002. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. Mol Cell 9: 279-289.

Cirillo LA, McPherson CE, Bossard P, Stevens K, Cherian S, Shim EY, Clark KL, Burley SK, Zaret KS. 1998. Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. EMBO J 17: 244-254.

Cirillo LA, Zaret KS. 1999. An early developmental transcription factor complex that is more stable on nucleosome core particles than on free DNA. Mol Cell 4: 961-969.

The ENCODE Project Consortium. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447: 799-816.

Feng J, Liu T, Qin B, Zhang Y, Liu XS. 2012. Identifying ChIP-seq enrichment using MACS. Nature protocols 7: 1728-1740.

Fisher WW, Li JJ, Hammonds AS, Brown JB, Pfeiffer BD, Weiszmann R, MacArthur S, Thomas S, Stamatoyannopoulos JA, Eisen MB et al. 2012. DNA regions bound at low occupancy by transcription factors do not drive patterned reporter gene expression in Drosophila. Proceedings of the National Academy of Sciences of the United States of America 109: 21330-21335.

Foe VE, Alberts BM. 1983. Studies of nuclear and cytoplasmic behaviour during the five mitotic cycles that precede gastrulation in Drosophila embryogenesis. J Cell Sci 61: 31-70.

Foo SM, Sun Y, Lim B, Ziukaite R, O'Brien K, Nien CY, Kirov N, Shvartsman SY, Rushlow CA. 2014. Zelda potentiates morphogen activity by increasing chromatin accessibility. Current biology : CB 24: 1341-1346.

Gerstein MB Lu ZJ Van Nostrand EL Cheng C Arshinoff BI Liu T Yip KY Robilotto R Rechtsteiner A Ikegami K et al. 2010. Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. Science 330: 1775-1787.

Giresi PG, Kim J, McDaniell RM, Iyer VR, Lieb JD. 2007. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. Genome Res 17: 877-885.

Gualdi R, Bossard P, Zheng M, Hamada Y, Coleman JR, Zaret KS. 1996. Hepatic specification of the gut endoderm in vitro: cell signaling and transcriptional control. Genes Dev 10: 1670-1682.

Hamm DC, Bondra ER, Harrison MM. 2015. Transcriptional activation is a conserved feature of the early embryonic factor Zelda that requires a cluster of four zinc fingers for DNA binding and a low-complexity activation domain. The Journal of biological chemistry 290: 3508-3518.

Harrison MM, Li XY, Kaplan T, Botchan MR, Eisen MB. 2011. Zelda binding in the early Drosophila melanogaster embryo marks regions subsequently activated at the maternal-to-zygotic transition. PLoS Genet 7: e1002266.

Hughes AL, Jin Y, Rando OJ, Struhl K. 2012. A functional evolutionary approach to identify determinants of nucleosome positioning: a unifying model for establishing the genome-wide pattern. Mol Cell 48: 5-15.

Iwafuchi-Doi M, Zaret KS. 2014. Pioneer transcription factors in cell reprogramming. Genes Dev 28: 2679-2692.

Iyer VR, Horak CE, Scafe CS, Botstein D, Snyder M, Brown PO. 2001. Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. Nature 409: 533-538.

Kadauke S, Udugama MI, Pawlicki JM, Achtman JC, Jain DP, Cheng Y, Hardison RC, Blobel GA. 2012. Tissue-specific mitotic bookmarking by hematopoietic transcription factor GATA1. Cell 150: 725-737.

Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Tillo D, Field Y, LeProust EM, Hughes TR, Lieb JD, Widom J et al. 2009. The DNA-encoded nucleosome organization of a eukaryotic genome. Nature 458: 362-366.

Kaplan T, Li X-y, Sabo PJ, Thomas S, Stamatoyannopoulos JA, Biggin MD, Eisen MB. 2011. Quantitative Models of the Mechanisms That Control Genome-Wide Patterns of Transcription Factor Binding during Early Drosophila Development. PLoS Genetics 7: e1001290.

Kvon EZ, Stampfel G, Yanez-Cuna JO, Dickson BJ, Stark A. 2012. HOT regions function as patterned developmental enhancers and have a distinct cis-regulatory signature. Genes Dev 26: 908-913.

Lagha M, Bothma JP, Esposito E, Ng S, Stefanik L, Tsui C, Johnston J, Chen K, Gilmour DS, Zeitlinger J et al. 2013. Paused Pol II coordinates tissue morphogenesis in the Drosophila embryo. Cell 153: 976-987.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nature methods 9: 357-359.

Lee C, Li X, Hechmer A, Eisen M, Biggin MD, Venters BJ, Jiang C, Li J, Pugh BF, Gilmour DS. 2008. NELF and GAGA factor are linked to promoter-proximal pausing at many genes in Drosophila. Molecular and cellular biology 28: 3290-3300.

Lee MT, Bonneau AR, Takacs CM, Bazzini AA, Divito KR, Fleming ES, Giraldez AJ. 2013. Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. Nature. 503: 360-364.

Lee W, Tillo D, Bray N, Morse RH, Davis RW, Hughes TR, Nislow C. 2007. A high-resolution atlas of

nucleosome occupancy in yeast. Nat Genet 39: 1235-1244.

Leichsenring M, Maes J, Mossner R, Driever W, Onichtchouk D. 2013. Pou5f1 Transcription Factor Controls Zygotic Gene Activation In Vertebrates. Science. 341: 1005-1009.

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

Li XY, Harrison MM, Villalta JE, Kaplan T, Eisen MB. 2014. Establishment of regions of genomic activity during the maternal to zygotic transition. eLife 3.

Li XY, MacArthur S, Bourgon R, Nix D, Pollard DA, Iyer VN, Hechmer A, Simirenko L, Stapleton M, Luengo Hendriks CL et al. 2008. Transcription factors bind thousands of active and inactive regions in the Drosophila blastoderm. PLoS Biol 6: e27.

Li XY, Thomas S, Sabo PJ, Eisen MB, Stamatoyannopoulos JA, Biggin MD. 2011. The role of chromatin accessibility in directing the widespread, overlapping patterns of Drosophila transcription factor binding. Genome Biol 12: R34.

Liang HL, Nien CY, Liu HY, Metzstein MM, Kirov N, Rushlow C. 2008. The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. Nature 456: 400-403.

Lieb JD, Liu X, Botstein D, Brown PO. 2001. Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. Nat Genet 28: 327-334.

Liu X, Lee CK, Granek JA, Clarke ND, Lieb JD. 2006. Whole-genome comparison of Leu3 binding in vitro and in vivo reveals the importance of nucleosome occupancy in target site selection. Genome Res 16: 1517-1528.

Lott SE, Villalta JE, Schroth GP, Luo S, Tonkin LA, Eisen MB. 2011. Noncanonical Compensation of the Zygotic X Transcription in Early Drosophila melanogaster Development Revealed through Single-Embryo RNA-Seq. PLoS Biology 9: e1000590.

Lu Q, Wallrath LL, Granok H, Elgin SC. 1993. (CT)n (GA)n repeats and heat shock elements have distinct roles in chromatin structure and transcriptional activation of the Drosophila hsp26 gene. Molecular and cellular biology 13: 2802-2814.

MacArthur S, Li XY, Li J, Brown JB, Chu HC, Zeng L, Grondona BP, Hechmer A, Simirenko L, Keranen SV et al. 2009. Developmental roles of 21 Drosophila transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions. Genome Biol 10: R80.

McKay DJ, Lieb JD. 2013. A common set of DNA regulatory elements shapes Drosophila appendages. Developmental cell 27: 306-318.

Moorman C, Sun LV, Wang J, de Wit E, Talhout W, Ward LD, Greil F, Lu XJ, White KP, Bussemaker HJ et al. 2006. Hotspots of transcription factor colocalization in the genome of Drosophila melanogaster. Proceedings of the National Academy of Sciences of the United States of America 103: 12027-12032.

Negre N, Brown CD, Ma L, Bristow CA, Miller SW, Wagner U, Kheradpour P, Eaton ML, Loriaux P, Sealfon R et al. 2011. A cis-regulatory map of the Drosophila genome. Nature 471: 527-531.

Newport J, Kirschner M. 1982. A major developmental transition in early Xenopus embryos: II. Control of the onset of transcription. Cell 30: 687-696.

Nien CY, Liang HL, Butcher S, Sun Y, Fu S, Gocha T, Kirov N, Manak JR, Rushlow C. 2011. Temporal coordination of gene networks by Zelda in the early Drosophila embryo. PLoS Genet 7: e1002339.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26: 139-140.

Roy S Ernst J Kharchenko PV Kheradpour P Negre N Eaton ML Landolin JM Bristow CA Ma L Lin MF et al. 2010. Identification of functional elements and regulatory circuits by Drosophila modENCODE. Science 330: 1787-1797.

Satija R, Bradley RK. 2012. The TAGteam motif facilitates binding of 21 sequence-specific transcription factors in the Drosophila embryo. Genome Res 22: 656-665.

Schones DE, Cui K, Cuddapah S, Roh TY, Barski A, Wang Z, Wei G, Zhao K. 2008. Dynamic regulation of nucleosome positioning in the human genome. Cell 132: 887-898.

Sekinger EA, Moqtaderi Z, Struhl K. 2005. Intrinsic histone-DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast. Mol Cell 18: 735-748.

Sekiya T, Muthurajan UM, Luger K, Tulin AV, Zaret KS. 2009. Nucleosome-binding affinity as a primary determinant of the nuclear mobility of the pioneer transcription factor FoxA. Genes Dev 23: 804-809.

Shopland LS, Hirayoshi K, Fernandes M, Lis JT. 1995. HSF access to heat shock elements in vivo depends critically on promoter architecture defined by GAGA factor, TFIID, and RNA polymerase II binding sites. Genes Dev 9: 2756-2769.

Slattery M, Ma L, Spokony RF, Arthur RK, Kheradpour P, Kundaje A, Negre N, Crofts A, Ptashkin R, Zieba J et al. 2014. Diverse patterns of genomic targeting by transcriptional regulators in Drosophila melanogaster. Genome Res 24: 1224-1235.

Soufi A, Donahue G, Zaret KS. 2012. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. Cell 151: 994-1004.

Soufi A, Garcia MF, Jaroszewicz A, Osman N, Pellegrini M, Zaret KS. 2015. Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. Cell 161: 555-568.

Staudt N, Fellert S, Chung HR, Jackle H, Vorbruggen G. 2006. Mutations of the Drosophila zinc finger-encoding gene vielfaltig impair mitotic cell divisions and cause improper chromosome segregation. Mol Biol Cell 17: 2356-2365.

Struffi P, Corado M, Kaplan L, Yu D, Rushlow C, Small S. 2011. Combinatorial activation and concentration-dependent repression of the Drosophila even skipped stripe 3+7 enhancer. Development 138: 4291-4299.

Tadros W, Lipshitz HD. 2009. The maternal-to-zygotic transition: a play in two acts. Development 136: 3033-3042.

ten Bosch JR, Benavides JA, Cline TW. 2006. The TAGteam DNA motif controls the timing of Drosophila pre-blastoderm transcription. Development 133: 1967-1977.

Thomas S, Li XY, Sabo PJ, Sandstrom R, Thurman RE, Canfield TK, Giste E, Fisher W, Hammonds A, Celniker SE et al. 2011. Dynamic reprogramming of chromatin accessibility during Drosophila embryo development. Genome Biol 12: R43.

Tsukiyama T, Wu C. 1995. Purification and properties of an ATP-dependent nucleosome remodeling factor. Cell 83: 1011-1020.

Valouev A, Johnson SM, Boyd SD, Smith CL, Fire AZ, Sidow A. 2011. Determinants of nucleosome organization in primary human cells. Nature 474: 516-520.

Weiner A, Hughes A, Yassour M, Rando OJ, Friedman N. 2010. High-resolution nucleosome mapping reveals transcription-dependent promoter packaging. Genome Res 20: 90-100.

Wilkins RC, Lis JT. 1997. Dynamics of potentiation and activation: GAGA factor and its role in heat shock gene regulation. Nucleic Acids Res 25: 3963-3968.

Xu Z, Chen H, Ling J, Yu D, Struffi P, Small S. 2014. Impacts of the ubiquitous factor Zelda on Bicoid-dependent DNA binding and transcription in Drosophila. Genes Dev 28: 608-621.

Yanez-Cuna JO, Dinh HQ, Kvon EZ, Shlyueva D, Stark A. 2012. Uncovering cis-regulatory sequence requirements for context-specific transcription factor binding. Genome Res 22: 2018-2030.

Yang A, Zhu Z, Kapranov P, McKeon F, Church GM, Gingeras TR, Struhl K. 2006. Relationships between p63 binding, DNA sequence, transcription activity, and biological function in human cells. Mol Cell 24: 593-602.

Zaret KS, Carroll JS. 2011. Pioneer transcription factors: establishing competence for gene expression. Genes Dev 25: 2227-2241.

Zhang Y, Vastenhouw NL, Feng J, Fu K, Wang C, Ge Y, Pauli A, van Hummelen P, Schier AF, Liu XS. 2014. Canonical nucleosome organization at promoters forms during genome activation. Genome Res 24: 260-266.

Chapter 3 - GAGA Factor may serve as a co-activator of the Drosophila genome

Statement of contribution: I performed the work described in this chapter with guidance from our collaborators in the Garcia lab (see Acknowledgements).

Abstract

Master regulator transcription factors initiate changes in cell fate by reprogramming the transcriptome. In *Drosophila melanogaster,* the transcription factor Zelda (ZLD) activates hundreds of genes in the early embryo during zygotic genome activation (ZGA). ZLD facilitates transcription by establishing or maintaining sites of open chromatin, but ZLD alone cannot account for the thousands of accessible loci that are present during the major onset of activation. One candidate co-regulator of ZGA is GAGA factor (GAF), a transcription factor with a known role in regulating chromatin accessibility. Here, we describe a set of MS2-based reporters that we have created to tease apart the roles of ZLD and GAF in the activation of shared target genes.

Introduction

Studies in a variety of developmental and cellular models have found that changes in cell identity are initiated by key transcription factors with a potent ability to reprogram the transcriptome. Yet, the mechanisms by which these factors produce widespread changes remain unclear. One popular model posits that “pioneer” transcription factors independently initiate reprogramming through regulation of chromatin accessibility (Zaret and Carroll 2011). However, recent work has challenged the notion of a lone pioneer, favoring a model in which several transcription factors reprogram closed chromatin cooperatively (Swinstead et al. 2016; Chronis et al. 2017). While these models may each accurately describe individual factors within specific chromatin contexts, this debate highlights a significant gap in our understanding of how master regulator transcription factors function.

A monumental transcriptional shift takes place during the first hours of animal life. Following fertilization, the zygotic genome is initially dormant while maternally supplied products direct development. Gradually, the zygotic genome is activated while maternal products are degraded in a hand-off known as the maternal-to-zygotic transition (MZT) (Tadros and Lipshitz 2009). In many animals, this transition occurs as the embryo is undergoing rapid nuclear divisions, switching between DNA replication and mitosis without pausing in gap phase (Yuan et al. 2016). Zygotic genome activation (ZGA) occurs in two transcriptional bursts. First, a small number of genes are expressed in a minor wave while the genome is still rapidly replicating. Then, as the MZT nears completion, the cell cycle slows and gap phase is introduced in concert with the major wave of genome activation (Tadros and Lipshitz 2009).

In *Drosophila melanogaster,* a single transcription factor, Zelda (ZLD), is responsible for expression of hundreds of genes during zygotic genome activation (ZGA) (Liang et al. 2008; Nien et al. 2011). In flies, the major wave of ZGA occurs two hours into development, after the embryo has gone through 13 rapid nuclear cycles to amplify the genome (Farrell and O’Farrell 2014). While ZLD is required for expression of many genes during the major wave of activation at nuclear cycle 14 (nc14), the genes

expressed earliest in ZGA (in nc7-12) are particularly reliant on ZLD for activation (Harrison et al. 2011). Given that ZLD-binding is strongly correlated with accessible chromatin (Harrison et al. 2011), we wanted to test whether ZLD has a role in chromatin regulation. Using Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE)-seq we determined that ZLD is required for open chromatin at hundreds of ZLD- binding sites during ZGA (Schulz et al. 2015). However, we also identified over three times as many ZLD-binding sites that do not require ZLD to remain accessible. Our analysis revealed that these ZLD-independent sites are enriched for the binding motif of another transcription factor called GAGA factor (GAF) (Schulz et al. 2015).

GAF is a promising candidate co-regulator of Drosophila ZGA for several reasons. Like ZLD, GAF is maternally deposited in the embryo and is highly expressed throughout early embryogenesis (Weiszmann, Hammonds, and Celniker 2009; Lott et al. 2011). Importantly, the GAF-binding motif is correlated with chromatin accessibility and gene expression during the major onset of ZGA (Blythe and Wieschaus 2016; Chen et al. 2013). Further, GAF has been shown to regulate chromatin accessibility in cell culture and is known to interact with chromatin remodelers (Okada and Hirose 1998; Lomaev et al. 2017; Fuda et al. 2015). Thus, GAF could be responsible for maintaining open chromatin at ZLD-binding sites that remain open in the absence of ZLD. We showed that the sites that depend on ZLD for accessibility are enriched for early expressed genes (Schulz et al. 2015). Thus, ZLD may function relatively independently during the early phase of the MZT, but may rely on additional factors, such as GAF, to activate genes during the major onset of transcription. Here, we describe our plan to use transgenic reporters to test the potential contributions of GAF to ZGA.

Results

We have designed a series of reporter constructs to assess the impact of ZLD and GAF on activation of shared target genes using live embryo imaging. To build these reporters we selected two genes, *snail* and *tailless* (*tll*), that have been shown to be mis-expressed in the absence of ZLD by *in situ* (Liang et al. 2008; Nien et al. 2011). The upstream regulatory regions of these genes contain sites that are bound by ZLD and GAF *in vivo* (Harrison et al. 2011; Negre et al. 2011) and that do not require ZLD to remain accessible during ZGA (Schulz et al. 2015). The enhancer/promoter regions of these genes were cloned upstream of a LacZ reporter gene containing 24 tandem repeats of the MS2 motif at its 5' end (Garcia et al. 2013). As these MS2 motifs are transcribed, they form mRNA stem loops that are bound by a maternally provided MS2 coat protein fused to GFP (MCP-GFP) to collectively produce a fluorescent spot (Figure 3.1A). Confocal microscopy is used to image live transgenic embryos possessing these reporters. In addition to a wild-type version of each reporter, we have built versions in which the binding sites of ZLD, GAF, or both factors have been mutated in the enhancer/promoter region to ablate the binding of these factors. Comparing the wild-type and mutant "ZLD/GAF" reporters will allow us to identify any changes in onset or pattern of expression that occur in the absence of regulation by ZLD and GAF.

As a control, we also generated reporters containing promoter regions from the ZLD target genes *bottleneck (bnk)* and *serendipity alpha (Sry-α)*. Unlike the regions chosen for our ZLD/GAF reporters, these regions depend only on ZLD to remain accessible in the early embryo and the associated genes fail to be expressed in the absence of ZLD (Schulz et al. 2015; Liang et al. 2008). For each of these genes, we created both a wild-type reporter and a version in which the ZLD-binding motifs had been mutated. By imaging embryos containing our wild-type *Sry-α* reporter, we have demonstrated that this reporter is activated early, starting at nc11, and is expressed ubiquitously (Figure 3.1B), as expected based on the literature (Liang et al. 2008). However, the mutant version of this reporter

produces very few transcriptional foci (data not shown). Thus, as expected, ZLD-binding motifs are necessary for expression of our *Sry-α* reporter gene. In addition to validating our experimental design, analysis of the *bnk* and *Sry-α* reporters will provide a benchmark to which the ZLD/GAF reporters can be compared.

To survey the effects of binding site mutations in our *tll* reporters, we first performed *in situ* hybridization on embryos containing reporters with wild-type binding sites (*tll* WT), mutated ZLD-binding sites (*tll* ZLD), mutated GAF-binding sites (*tll* GAF), or a combination of mutated ZLD- and GAF-binding sites (*tll* BOTH) using a probe that recognizes *lacZ* mRNA. Additionally, we probed a fifth set of embryos for endogenous *tll* expression and showed that our wild-type reporter recapitulates the endogenous pattern at nc14 (Figure 3.2A). In bulk, the embryos with mutant reporters appeared to have reduced staining as compared to those with the wild-type reporter. However, a systematic analysis revealed that there was substantial staining variability across embryos with any given reporter (Figure 3.2A). Without a reliable way to stage these embryos, it was impossible to determine if the reduced staining we observed was simply a result of a higher proportion of young embryos in the mutant samples.

To assess *tll* reporter expression at higher resolution, we next analyzed transgenic embryos using confocal microscopy. By this method, we could detect transcriptional foci at the anterior and posterior termini of the embryos as early as nc12, with a more robust pattern that mimicked endogenous expression appearing by nc14 (Sup. Figure 3.1). However, *tll* expression has been observed as early as nc10 (Nien et al. 2011), so either we did not fully recapitulate the endogenous expression pattern or we have failed to detect early expression at the termini of these embryos. Again, mutant reporter expression appeared to be qualitatively weaker than wild-type reporter expression, but variability within each reporter class has prevented us from drawing any conclusions at this time.

Further, expression of these reporters appears to weaken as nc14 progresses (Figure 3.2B). Therefore, precise staging will be critical for identifying subtle differences in expression of these reporters.

Discussion

With the ability to simultaneously reveal temporal and spatial expression patterns, our MS2-based reporters are a useful tool for determining the contributions of ZLD and GAF to gene activation. Using reporters based on the ZLD-dependent gene *Sry-α*, we have recapitulated *in situ* results using live embryo imaging (Liang et al. 2008). However, our survey of reporters based on the gene *tll,* which may be co-regulated by GAF, revealed that further analysis will be required to detect the effects of mutating ZLD- or GAF- binding motifs in the enhancer/promoter region of this gene.

While the expression patterns produced by our *tll* reporters mimic the endogenous pattern at nc14, we have been unable to detect expression as early (nc10-11) as has been described in the literature (Nien et al. 2011). Previous work suggests that ~6 kb of upstream regulatory sequence may be required to drive reporter expression in the full, embryonic *tll* pattern (Rudolph et al. 1997). Thus, the fact that our *tll* reporters include only 3.3kb of this sequence could account for differences in expression. Accordingly, it will be important to solidly establish the timing and patterning of “wild-type” reporter expression to serve as the baseline to which the mutant reporters are compared.

For genes like *tll,* for which expression is aberrant and expanded but not abolished in the absence of ZLD, the effects of ZLD-motif mutation may prove to be especially subtle. In the absence of maternally supplied ZLD, *tll* is expressed 1-2 cleavage cycles later in an abnormal, expanded pattern (Nien et al. 2011). ZLD has been shown to regulate transcription by potentiating the binding of other transcription factors, including the factors Bicoid (BCD) and Dorsal, which are both regulators of *tll* (Yanez-Cuna et al. 2012; Xu et al. 2014; Foo et al. 2014; Liaw and Lengyel 1993). BCD is expressed in an anterior-to-posterior gradient (Driever and Nüsslein-Volhard 1988), and ZLD is particularly important for activation of BCD-target genes in areas of the embryo where its concentration is low (Xu et al. 2014). This can partially explain why the requirement for ZLD for *tll* expression is spatially variable and may, thus, prove to be challenging to detect.

If differences in expression of our reporters are not detected by a qualitative survey, a more intensive analysis can be performed. Our collaborators in the Garcia lab have developed a computational analysis pipeline in which reporter expression can be tracked in individual nuclei across multiple nuclear cycles (Garcia et al. 2013). It has been shown that adding ZLD-binding motifs to an enhancer will cause it to drive earlier and more robust reporter expression in the embryo, with activation occurring in a greater proportion of nuclei (Foo et al. 2014; Crocker, Tsai, and Stern 2017). Thus, the effect of mutating ZLD motifs could be as subtle as a reduction in the number of nuclei that are activated at a given point in development. Using the Garcia lab pipeline, we could record the number and localization of transcriptionally active nuclei to capture differences of this nature with temporal precision.

We may, nonetheless, be unable to detect measurable differences between our wild-type and mutant *tll* reporters. In this case, one possibility is that we failed to mutate all the relevant binding motifs in our enhancer/promoter sequences. GAF can recognize very short DNA sequences, making complete ablation of GAF motifs challenging. While the GAF motif was originally thought to be comprised of ~3.5 (GA) repeats (Granok et al. 1995), GAF has been shown to bind sequences as short as three bases (GAG) in certain contexts (Wilkins and Lis 1998). We mutated all such motifs within the region marked by GAF ChIP-seq signal (Negre et al. 2011). Yet, several short GA-rich motifs that lie outside this region remain intact in our GAF mutant reporters. Thus, it is possible that GAF binding at these sites could compensate for the absence of its binding elsewhere, obscuring the effects of our mutations.

Another possibility is that GAF does not contribute to *tll* expression during ZGA. The ChIP-seq data used to select potential GAF-regulated genes was produced from 0-8 hour embryos, so GAF may only bind the *tll* promoter later in development. Notably, GAF is known to promote RNA Polymerase II (Pol II) promoter pausing (Boija et al. 2017; Duarte et al. 2016; C. Lee et al. 2008; Tsai et al. 2016), and

appears to be involved in the recruitment of paused polII during the major wave of ZGA (Chen et al., 2013). While the *tll* promoter shows no evidence of pausing, our second ZLD/GAF gene, *snail*, has been shown to gradually accumulate Pol II at its pause site during ZGA (Chen et al. 2013), making it a promising candidate for GAF regulation. In future work, we will continue our examination of the *tll* reporters and begin to assess the roles of ZLD and GAF at the *snail* promoter.

Materials and Methods

Cloning and Transgenesis

The Garcia lab provided us with the reporter plasmid pIB-hbP2 p2pWT-MS2v7-LacZ-TUb3'UTR. This plasmid contains the hunchback P2 enhancer and promoter driving the expression of LacZ fused to an αTubulin 3’UTR with a series of 24 MS2 motifs in the 5’UTR. This plasmid is nearly identical to the previously described plasmid piB-hbP2-P2P-MS2-24x-lacZ-αTub3’UTR (Garcia et al. 2013), apart from a new MS2 sequence in which potential ZLD-binding motifs have been mutated. The reporter sequence is flanked by attB sites in this plasmid, allowing it to be inserted into the fly genome at specific landing sites by recombinase-mediated cassette exchange (RMCE) (Bateman, Lee, and Wu 2006).

To build our reporters, the enhancer/promoter regions associated with genes of interest were cloned into a pBluescript vector using restriction digest. For *bottleneck* (exact region: chr3R:27,018,085-27,019,109) and *Sry-α* (chr3R:25,866,080-25,867,061), these regions consisted of a ~1kb sequence found upstream of the transcription start site, while the *snail* (chr2L:15,478,105-15,480,409) and *tll* (chr3R:26,674,967-26,678,272) regions were 2.3-3.3kb in length. Gibson Assembly was used to incorporate gBlocks (IDT) containing the desired mutations in ZLD and/or GAF binding sites into the enhancer/promoter regions. To mutate potential ZLD motifs, any AGGTA sequences (the central portion of the canonical motif cAGGTAg) were substituted with the sequence CTTGA. To mutate potential GAF motifs, any (GA)-repeats (GAGA or longer) were disrupted by substitutions with T/C. Finally, the hunchback p2 enhancer was replaced with each of our wild-type and mutant enhancer/promoter regions in the pIB-hbP2 p2pWT-MS2v7-LacZ-TUb3'UTR plasmid using Gibson Assembly. These constructs were inserted into the fly genome (site 38F1, chrII) using RMCE to create transgenic fly lines (BestGene, injected stock: BDSC#27388).

In Situ Hybridization

In situ hybridization was performed using the standard method with complementary digoxigenin-labelled mRNA probes for *tll* and *lacZ*.

Live Imaging Sample Preparation

Virgin female flies expressing MCP-GFP and Histone-RFP (*yw; Histone-RFP;MCP-NoNLS-GFP*) (Garcia et al. 2013) were crossed with males from the MS2-LacZ reporter lines. Embryos were collected, dechorinated with bleach, and mounted in Halocarbon 27 oil between a semipermeable membrane (Biofolie, In Vitro Systems & Services) and a coverslip (Fisherbrand 12-541A). Embryos were imaged on a Nikon A1R + Confocal Microscope using a 60x objective. MCP-GFP and Histone-RFP were excited with laser wavelengths of 488 and 561 nm, respectively.

Figures

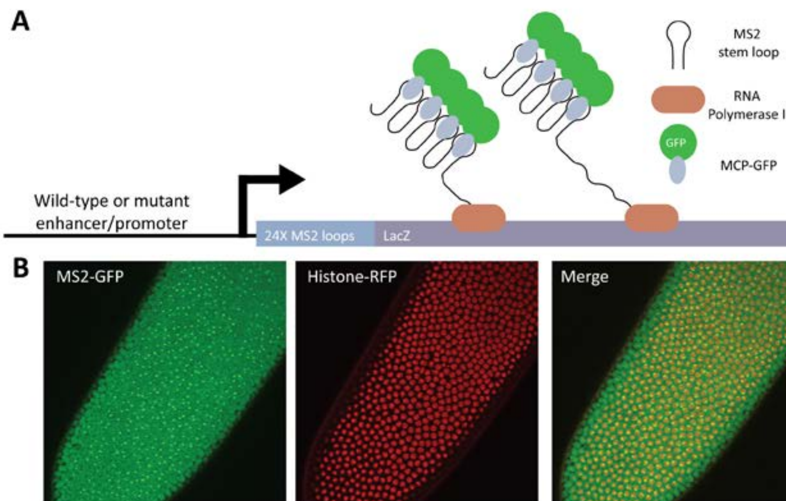


Figure 3.1: MS2 reporters enable detection of nascent transcription in live embryos.

(A) Schematic depicting our enhancer/promoter regions upstream of a lacZ gene containing 24 MS2 stem loops at its 5’-end. Maternally provided MCP-GFP protein binds to the stem loops as they are transcribed, creating fluorescent foci. **(B)** Snapshot of the anterior region of an embryo expressing the wild-type *Sry-α* MS2-MCP system at nuclear cycle 14.

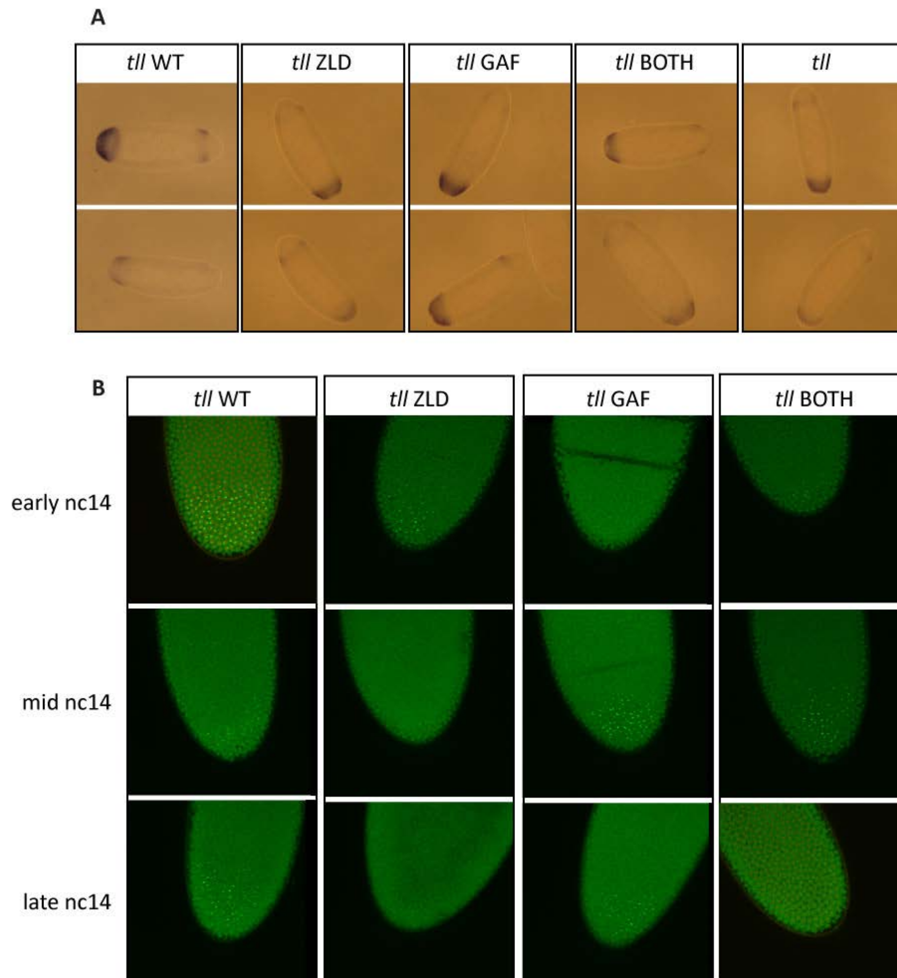
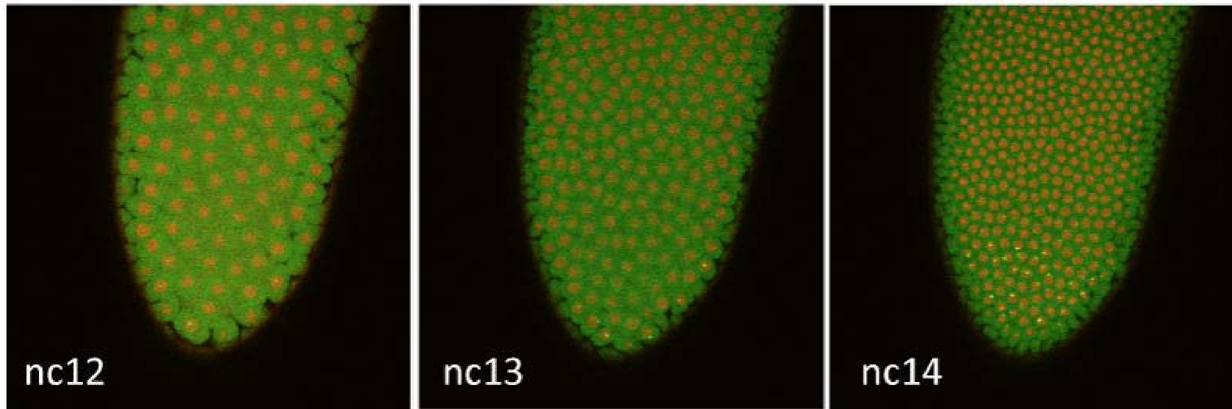


Figure 3.2: Expression patterns of *tll* reporters vary within each reporter class.

(A) Representative embryos on which *in situ* hybridization was performed to label the *lacZ* RNA produced from *tll* reporters with wild-type binding sites (*tll* WT), mutated ZLD-binding sites (*tll* ZLD), mutated GAF-binding sites (*tll* GAF), or both mutated ZLD- and mutated GAF- binding sites (*tll* BOTH). A fifth set of embryos was treated with probe for *tll* RNA to label the expression of the endogenous gene (*tll).* **(B)** Snapshots of the posterior region of live embryos expressing the *tll* reporters (*tll* WT, *tll* ZLD, *tll* GAF, *tll* BOTH) at various stages of nuclear cycle 14, indicated on the left.



Sup. Figure 3.1: *tll* reporter expression is detected in nuclear cycle 12 (nc12) and expanded in nuclear cycle 14 (nc14).

Snapshots of the posterior region of live embryos expressing the *tll* reporter with mutated GAF-binding sites (*tll* GAF) during nc12-14, as indicated.

References

Blythe, S. A., & Wieschaus, E. F. (2016). Establishment and maintenance of heritable chromatin structure during early Drosophila embryogenesis. ELife, 5, e20148. https://doi.org/10.7554/eLife.20148

Bateman, J. R., Lee, A. M., & Wu, C. (2006). Site-Specific Transformation of Drosophila via C31 Integrase-Mediated Cassette Exchange. Genetics, 173(2), 769–777. https://doi.org/10.1534/genetics.106.056945

Blythe, S. A., & Wieschaus, E. F. (2016). Establishment and maintenance of heritable chromatin structure during early Drosophila embryogenesis. ELife, 5, e20148. https://doi.org/10.7554/eLife.20148

Boija, A., Mahat, D. B., Zare, A., Holmqvist, P.-H., Philip, P., Meyers, D. J., … Mannervik, M. (2017). CBP Regulates Recruitment and Release of Promoter-Proximal RNA Polymerase II. Molecular Cell, 68(3), 491–503.e5. https://doi.org/10.1016/j.molcel.2017.09.031

Chen, K., Johnston, J., Shao, W., Meier, S., Staber, C., & Zeitlinger, J. (2013). A global change in RNA polymerase II pausing during the Drosophila midblastula transition. ELife, 2, e00861. https://doi.org/10.7554/eLife.00861

Chronis, C., Fiziev, P., Papp, B., Butz, S., Bonora, G., Sabri, S., … Plath, K. (2017). Cooperative Binding of Transcription Factors Orchestrates Reprogramming. Cell, 168(3), 442–459.e20. https://doi.org/10.1016/j.cell.2016.12.016

Crocker, J., Tsai, A., & Stern, D. L. (2017). A Fully Synthetic Transcriptional Platform for a Multicellular Eukaryote. Cell Reports, 18(1), 287–296. https://doi.org/10.1016/j.celrep.2016.12.025

Driever, W., & Nüsslein-Volhard, C. (1988). A gradient of bicoid protein in Drosophila embryos. Cell, 54(1), 83–93. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3383244

Duarte, F. M., Fuda, N. J., Mahat, D. B., Core, L. J., Guertin, M. J., & Lis, J. T. (2016). Transcription factors GAF and HSF act at distinct regulatory steps to modulate stress-induced gene activation. Genes & Development, 30(15), 1731–1746. https://doi.org/10.1101/gad.284430.116

Farrell, J. A., & O'Farrell, P. H. (2014). From egg to gastrula: how the cell cycle is remodeled during the Drosophila mid-blastula transition. Annual Review of Genetics, 48, 269–294. https://doi.org/10.1146/annurev-genet-111212-133531

Foo, S. M., Sun, Y., Lim, B., Ziukaite, R., O'Brien, K., Nien, C. Y., … Rushlow, C. A. (2014). Zelda potentiates morphogen activity by increasing chromatin accessibility. Curr Biol, 24(12), 1341–1346. https://doi.org/10.1016/j.cub.2014.04.032

Fuda, N. J., Guertin, M. J., Sharma, S., Danko, C. G., Martins, A. L., Siepel, A., & Lis, J. T. (2015). GAGA Factor Maintains Nucleosome-Free Regions and Has a Role in RNA Polymerase II Recruitment to Promoters. PLOS Genetics, 11(3), e1005108. Retrieved from http://dx.doi.org/10.1371%2Fjournal.pgen.1005108

Garcia, H. G., Tikhonov, M., Lin, A., & Gregor, T. (2013). Quantitative Imaging of Transcription in Living Drosophila Embryos Links Polymerase Activity to Patterning. Current Biology, 23(21), 2140–2145. https://doi.org/10.1016/j.cub.2013.08.054

Granok, H., Leibovitch, B. A., Shaffer, C. D., & Elgin, S. C. (1995). Chromatin. Ga-ga over GAGA factor. Current Biology : CB, 5(3), 238–241. https://doi.org/10.1016/S0960-9822(95)00048-0

Harrison, M. M., Li, X.-Y., Kaplan, T., Botchan, M. R., & Eisen, M. B. (2011). Zelda binding in the early Drosophila melanogaster embryo marks regions subsequently activated at the maternal-to-zygotic transition. PLoS Genetics, 7(10), e1002266. https://doi.org/10.1371/journal.pgen.1002266

Lee, C., Li, X., Hechmer, A., Eisen, M., Biggin, M. D., Venters, B. J., … Gilmour, D. S. (2008). NELF and GAGA Factor Are Linked to Promoter-Proximal Pausing at Many Genes in Drosophila. Molecular and Cellular Biology, 28(10), 3290–3300. https://doi.org/10.1128/MCB.02224-07

Liang, H.-L., Nien, C.-Y., Liu, H.-Y., Metzstein, M. M., Kirov, N., & Rushlow, C. (2008). The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. Nature, 456(7220), 400–403. Retrieved from http://dx.doi.org/10.1038/nature07388

Liaw, G. J., & Lengyel, J. A. (1993). Control of tailless expression by bicoid, dorsal and synergistically interacting terminal system regulatory elements. Mechanisms of Development, 40(1–2), 47–61. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8443106

Lomaev, D., Mikhailova, A., Erokhin, M., Shaposhnikov, A. V, Moresco, J. J., Blokhina, T., … Chetverina, D. (2017). The GAGA factor regulatory network: Identification of GAGA factor associated proteins. PloS One, 12(3), e0173602. https://doi.org/10.1371/journal.pone.0173602

Lott, S. E., Villalta, J. E., Schroth, G. P., Luo, S., Tonkin, L. A., & Eisen, M. B. (2011). Noncanonical Compensation of the Zygotic X Transcription in Early Drosophila melanogaster Development Revealed through Single-Embryo RNA-Seq. PLoS Biology, 9(2), e1000590. https://doi.org/10.1371/journal.pbio.1000590

Negre, N., Brown, C. D., Ma, L., Bristow, C. A., Miller, S. W., Wagner, U., … White, K. P. (2011). A cis-regulatory map of the Drosophila genome. Nature, 471(7339), 527–531. https://doi.org/10.1038/nature09990

Nien, C. Y., Liang, H. L., Butcher, S., Sun, Y., Fu, S., Gocha, T., … Rushlow, C. (2011). Temporal coordination of gene networks by Zelda in the early Drosophila embryo. PLoS Genet, 7(10), e1002339. https://doi.org/10.1371/journal.pgen.1002339

Okada, M., & Hirose, S. (1998). Chromatin Remodeling Mediated by Drosophila GAGA Factor and ISWI Activates fushi tarazu Gene Transcription In Vitro. Mol Cell Biol, 18(5), 2455–2461.

Rudolph, K. M., Liaw, G. J., Daniel, A., Green, P., Courey, A. J., Hartenstein, V., & Lengyel, J. A. (1997). Complex regulatory region mediating tailless expression in early embryonic patterning and brain development. Development (Cambridge, England), 124(21), 4297–4308. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9334278

Schulz, K. N., Bondra, E. R., Moshe, A., Villalta, J. E., Lieb, J. D., Kaplan, T., … Harrison, M. M. (2015). Zelda is differentially required for chromatin accessibility, transcription factor binding, and gene expression in the early Drosophila embryo. Genome Res, 25(11), 1715–1726. https://doi.org/10.1101/gr.192682.115

Swinstead, E. E., Miranda, T. B., Paakinaho, V., Baek, S., Goldstein, I., Hawkins, M., … Hager, G. L. (2016). Steroid Receptors Reprogram FoxA1 Occupancy through Dynamic Chromatin Transitions. Cell, 165(3), 593–605. https://doi.org/10.1016/j.cell.2016.02.067

Tadros, W., & Lipshitz, H. D. (2009). The maternal-to-zygotic transition: a play in two acts. Development, 136(18), 3033–3042. https://doi.org/10.1242/dev.033183

Tsai, S.-Y., Chang, Y.-L., Swamy, K. B. S., Chiang, R.-L., & Huang, D.-H. (2016). GAGA factor, a positive regulator of global gene expression, modulates transcriptional pausing and organization of upstream nucleosomes. Epigenetics & Chromatin, 9(1), 32. https://doi.org/10.1186/s13072-016-0082-4

Weiszmann, R., Hammonds, A. S., & Celniker, S. E. (2009). Determination of gene expression patterns using high-throughput RNA in situ hybridization to whole-mount Drosophila embryos. Nature Protocols, 4(5), 605–618. https://doi.org/10.1038/nprot.2009.55

Wilkins, R. C., & Lis, J. T. (1998). GAGA factor binding to DNA via a single trinucleotide sequence element. Nucleic Acids Research, 26(11), 2672–2678. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9592153

Xu, Z., Chen, H., Ling, J., Yu, D., Struffi, P., & Small, S. (2014). Impacts of the ubiquitous factor Zelda on Bicoid-dependent DNA binding and transcription in Drosophila. Genes Dev, 28(6), 608–621. https://doi.org/10.1101/gad.234534.113

Yanez-Cuna, J. O., Dinh, H. Q., Kvon, E. Z., Shlyueva, D., & Stark, A. (2012). Uncovering cis-regulatory sequence requirements for context-specific transcription factor binding. Genome Res, 22(10), 2018–2030. https://doi.org/10.1101/gr.132811.111

Yuan, K., Seller, C. A., Shermoen, A. W., & O'Farrell, P. H. (2016). Timing the Drosophila Mid-Blastula Transition: A Cell Cycle-Centered View. Trends in Genetics : TIG, 32(8), 496–507. https://doi.org/10.1016/j.tig.2016.05.006

Zaret, K. S., & Carroll, J. S. (2011). Pioneer transcription factors: establishing competence for gene expression. Genes Dev, 25(21), 2227–2241. https://doi.org/10.1101/gad.176826.111

Chapter 4 - Zelda binds nucleosomal DNA

Statement of contribution: I performed the bulk of the work described in this chapter with guidance from our collaborators in the Zaret and Lewis labs. Nucleosome assembly was performed by Siddhant Jain and a second set of EMSAs was performed by Meilin Fernandez Garcia (see Acknowledgements).

Abstract

While the power of master regulator transcription factors to shape cell identity has been widely demonstrated, their capacity to elicit massive transcriptional change remains largely unexplained. In Drosophila, the transcription factor Zelda (ZLD) is responsible for global activation of the zygotic genome during the first hours of embryogenesis. ZLD serves this role, at least in part, by regulating chromatin accessibility, but the mechanisms by which ZLD influences chromatin are unknown. A special class of transcription factors, known as pioneer factors, initiate chromatin remodeling using their unique ability to bind nucleosomal DNA. Here, we present evidence that ZLD shares this ability. ZLD binds nucleosomes *in vitro* through an interaction that is driven by the sequence-specific activity of its DNA-binding domain.

Introduction

Transcription factors shape cell identity by binding to specific DNA sequences and regulating gene expression. Efforts to experimentally reprogram cells have revealed the remarkable potency of key transcription factors that reside at the top of regulatory hierarchies. Some of these factors have the ability to confine a cell to a particular fate or even revert it to pluripotency (Vierbuchen et al. 2010; Takahashi et al. 2006). Yet, while "master regulator" transcription factors have been identified in many cellular contexts, the mechanisms by which they produce global transcriptional change remain unclear.

In *Drosophila melanogaster,* a single transcription factor called Zelda (ZLD) initiates a global transcriptional shift during early embryogenesis. In animals, the genome remains transcriptionally silent following fertilization while development is driven by maternally deposited transcripts and proteins. But within hours, these maternal products are selectively degraded and zygotic genome activation (ZGA) occurs (Tadros and Lipshitz 2009). ZLD is required for the expression of hundreds of genes during this transition (Liang et al. 2008; Nien et al. 2011). ZLD binds to thousands of genomic sites at the beginning of ZGA, and is required for the activation of both early expressed genes and genes that are expressed during the major burst of transcription that concludes ZGA (Harrison et al. 2011). Thus, early ZLD binding poises the genome for activation. Further, ZLD-binding is strongly correlated with regions of open chromatin (Harrison et al. 2011), and embryos that lack ZLD lose accessibility at many of these sites (Schulz et al. 2015; Sun et al. 2015). Both this ability to regulate chromatin accessibility and the ability to poise genes for later expression suggest that ZLD might function as a pioneer factor.

Pioneer factors are specialized transcription factors that bind to regions of silent chromatin to initiate cell reprogramming (Zaret and Carroll 2011). Nucleosomes present a barrier to most transcription factors by restricting their ability to access the underlying DNA (Kornberg and Lorch 1999). Pioneer factors overcome this barrier, allowing additional transcription factors to bind and influence gene expression. The truly defining feature of a pioneer factor is the ability to access binding sites in

nucleosomal DNA (Zaret and Carroll 2011). Yet, while numerous transcription factors have been dubbed pioneers based on their ability to promote chromatin accessibility, few factors have been directly shown to interact with nucleosomes. Here, we present evidence that both full-length ZLD and its DNA-binding domain alone bind to nucleosomes *in vitro* and that this interaction is largely driven by DNA-sequence specificity.

Results

ZLD binds nucleosomes in vitro

To test the ability of ZLD to bind nucleosomes, we purified recombinant ZLD protein to use in electrophoretic mobility shift assays (EMSA). Full-length ZLD was purified from insect cells using baculovirus infection and affinity purification, and a truncated ZLD protein consisting of only the DNA binding domain (ZLD^{DBD}) was purified from *E. coli*. The concentration of each protein was determined from SYPRO-Red stained SDS-PAGE gels using BSA standards for quantification.

To construct nucleosomes for these assays, we chose a 160bp sequence in the promoter of the ZLD-target gene *bottleneck* (*bnk*) that contains four ZLD-binding motifs. This locus is bound by ZLD throughout ZGA and requires ZLD for accessibility (Schulz et al. 2015; Sun et al. 2015; Harrison et al. 2011). Further, this site is enriched for nucleosome occupancy in the absence of ZLD, as measured by MNase-seq (Sun et al. 2015), suggesting that this sequence would likely form a nucleosome *in vitro*. We performed PCR on Drosophila genomic DNA to generate Cy5-labelled *bnk*-DNA, which was assembled into nucleosomes (*bnk*-nuc) by salt gradient dilution.

In our assays, full-length ZLD was able to bind *bnk*-nuc, though with roughly six-fold lower affinity than for *bnk*-DNA (Figure 4.1A), indicating that ZLD has a lower affinity for nucleosomes than for free DNA. Likewise, ZLD^{DBD} was also able to bind *bnk*-nuc, though at roughly three-fold lower affinity than the full-length protein (Figure 4.1B), indicating that the DNA-binding domain is sufficient for nucleosome binding. As we increase the concentration of ZLD in reactions with free DNA probe, we detect multiple proteins bound to *bnk*-DNA, as evidenced by a slower mobility band. However, at equivalent protein levels, *bnk*-nuc appears to be bound by a single polypeptide (Figure 4.1A). Using high concentrations of ZLD^{DBD} we are able to detect multiple polypeptides binding to nucleosomes, potentially at the four canonical ZLD motifs present in the *bnk* sequence (Sup. Figure 4.1). Thus, it is possible that the full-length protein was simply not concentrated enough to saturate binding to *bnk*-nuc

in these reactions. Alternatively, steric hinderance may prevent the binding of multiple full-length polypeptides, which are much larger than ZLD^{DBD} (180kDa vs. 85kDa). To confirm that histones were present in the shifted *bnk*-nuc complexes, we transferred protein from EMSA gels to PVDF membrane and blotted for histone H3. The shifted complexes resulting from binding of both full-length ZLD and ZLD^{DBD} contained H3, indicating that ZLD binding does not evict the histones from these nucleosomes (Figure 4.1C,D).

To verify these results, our collaborators in the Zaret lab performed a set of analogous assays using their preferred conditions and nucleosomes containing two suboptimal ZLD-binding motifs. In their hands, full-length ZLD similarly binds to nucleosomes with about three-fold lower affinity than to naked DNA, and ZLD^{DBD} binds nucleosomes with about three-fold lower affinity that the full-length protein (Sup. Figure 4.2A,B). However, under these conditions ZLD binds all probes with roughly 10-fold greater affinity than in our assays. This is likely due to the lack of poly[d-(IC)] in these reactions, which we use in our assays to decrease nonspecific interactions. Nonetheless, these results independently confirm our conclusion that ZLD has the capacity to interact with nucleosomes.

The interaction of ZLD with nucleosomes is predominantly sequence specific

Pioneer factors have been shown to interact with nucleosomes through a combination of specific and non-specific interactions (Sekiya et al. 2009; Soufi et al. 2015). To determine whether ZLD can bind to nucleosomes in the absence of its canonical binding motif, we repeated our assays using nucleosomes containing the Widom 601 sequence (601-nuc), which was identified based on its ability to form highly stable nucleosomes (Lowary and Widom 1998). Both full-length ZLD and ZLD^{DBD} were able to bind free 601-DNA with similar affinity as was observed with *bnk*-DNA (Figure 4.2A,B), suggesting that this protein has a strong, nonspecific affinity for long (147-160bp) DNA probes. However, neither protein was able to bind to 601-nuc (Figure 4.2A,B). This supports the idea that ZLD requires its DNA-binding

motif to target nucleosomes. However, a recent analysis of the 601 nucleosome revealed that it substantially more rigid than endogenous nucleosomes, calling its biological relevance into question (Takizawa et al. 2018). Nonetheless, these data demonstrate that ZLD must engage with the nucleosomal DNA to bind nucleosomes, as interactions with other protein and structural elements of the nucleosome are not sufficient for binding.

To further assess the degree to which sequence specificity contributes to the interactions of ZLD with nucleosomes, we added varying amounts of competitor probe to our binding reactions. Our competitor probe consisted of unlabeled DNA based on the *scute* promoter. We confirmed that both full-length ZLD and ZLD^{DBD} could bind to *scute* probes containing a canonical binding motif (*sc* WT) but not to probes in which this motif has been mutated (*sc* MUT, Sup. Figure 4.3A-B), as we have previously demonstrated under different conditions (Harrison, Botchan, and Cline 2010; Hamm, Bondra, and Harrison 2015). Surprisingly, the *scute* competitor probe was unable to compete with full-length ZLD binding to *bnk*-DNA, even when added at 320-fold molar excess (data not shown). In contrast, the binding of ZLD^{DBD} to *bnk*-DNA was abrogated by only 32-fold specific competitor (Figure 4.2D), suggesting that regions outside of the DNA-binding domain may be involved in nonspecific interactions with DNA. The increased affinity of ZLD for *bnk*-DNA as compared to the short competitor probe could be a result of its longer length (*bnk*:160bp, *scute*:24bp) or additional ZLD-binding sites (*bnk* contains 4, *sc* contains 1), which would both effectively increase the concentration of this probe in the binding reactions. In contrast, an 80-fold molar excess of specific probe (s) competed with full-length ZLD binding to *bnk*-nuc, while the same amount of nonspecific probe (ns) did not appreciably reduce binding (Figure 4.2C). Similarly, at 32-fold molar excess, only the specific competitor competed with ZLD^{DBD} binding to *bnk*-nuc (Figure 4.2D). Together these results suggest the interaction of ZLD with nucleosomes is largely sequence driven.

Discussion

In this work, we demonstrate that ZLD can bind nucleosomes containing its canonical binding motif *in vitro*. In doing so, we add ZLD to the short list of transcription factors for which this ability has been directly tested. These factors include the canonical pioneer FoxA1 (Sekiya et al. 2009), a master regulator of liver cell fate in mammals, its *C. elegans* homolog PHA-4 (Hsu et al. 2015), and the mammalian pluripotency factors Oct4, Sox2, and Klf4 (Soufi et al. 2015). Like ZLD, several of these factors (FoxA1, PHA-4, Klf4) bind nucleosomes with lower affinity than free DNA (Sekiya et al. 2009; Hsu et al. 2015; Soufi et al. 2015). While most of these factors bind nucleosomes using a combination of sequence-specific and -nonspecific interactions, competition assays indicate that the binding of both Oct4 and ZLD are predominantly sequence specific (Soufi et al. 2015). Our assays also revealed that the DNA-binding domain of ZLD is sufficient for nucleosome binding, which has also been demonstrated for FoxA1. The DNA-binding domain of FoxA1 binds to free DNA with nearly (70%) the affinity of the full-length protein, but binds with a relatively lower affinity (30%) to nucleosomes (L A Cirillo et al. 1998), similarly to what we observed with ZLD. Thus, ZLD is similar to this limited set of pioneer factors in both its ability to bind nucleosomes and in the characteristics of this interaction.

While we have confirmed that ZLD can interact with nucleosomes, the mechanism by which it impacts chromatin accessibility remains unresolved. For FoxA1, simply binding to a nucleosome facilitates chromatin opening (Lisa Ann Cirillo et al. 2002). FoxA1 disrupts interactions between adjacent nucleosomes and displaces H1 using a protein domain that structurally resembles this linker histone (Clark et al. 1993; Taube et al. 2010). Other pioneers that lack inherently disruptive domains can act as beacons for chromatin regulators. For example, Oct4 has been shown to recruit the chromatin remodeler BRG1 to stabilize nucleosome positioning (King and Klose 2017). For ZLD, no protein interactors have been identified and its structural domains offer few clues to its function. However, an association of ZLD binding with histone acetylation may offer a mechanistic clue (Li et al. 2014). Loci that

require ZLD for accessibility in the early embryo are enriched for histone acetylation marks (Schulz et al. 2015), and in the absence of ZLD, acetylation is specifically lost at its binding sites (Li et al. 2014). While it is possible that ZLD directly recruits a histone acetyltransferase, ZLD could also facilitate acetylation by maintaining open chromatin at specific sites to direct the binding of other factors that serve this function.

The pioneer factor model offers a tidy explanation of how silent chromatin is converted into functional cis-regulatory elements and provides a convenient shorthand to discuss how transcription factors function. However, recent work has called this hierarchical model into question, favoring a more egalitarian model in which several transcription factors bind closed chromatin cooperatively (Swinstead et al. 2016; Chronis et al. 2017). Further, many pioneers possess only a subset of pioneer-like traits and may only bind to nucleosomes within a particular chromatin context. While we've shown that ZLD binds nucleosomes *in vitro*, it does not share certain characteristics with the canonical pioneer factors, such as slow nuclear diffusion and mitotic bookmarking (unpublished work). Thus, by characterizing each factor individually, we will gain a more nuanced understanding of transcription factor function.

While our findings suggest that ZLD may have the ability to bind closed chromatin, ZLD may not require pioneer activity to access binding sites in the early embryo. Due to experimental limitations, we have been unable to profile chromatin accessibility in the embryo prior to ZLD binding. Thus, it remains to be determined if ZLD initially binds to closed chromatin or if it binds to open regions and simply maintains the local accessibility. While the very early embryo remains uncharted, a growing body of evidence suggests that this chromatin may exist in a naïve, globally accessible state (Li et al. 2014; Hug et al. 2017; Blythe and Wieschaus 2016). Thus, it is possible that ZLD simply binds to open regions and serves as a placeholder that blocks the establishment of repressive chromatin structures at key sites. In this case, early upregulation of ZLD would be sufficient to facilitate its widespread binding, regardless of its chromatin-binding capacity. The fact that ZLD is one of the earliest factors to be translated to

detectable levels in the embryo lends plausibility to this model (Nien et al. 2011). Even in this scenario, the ability of ZLD to interact with nucleosomes could be critical to its function, allowing it to track its target loci across several nuclear cycles. In the early embryo, chromatin accessibility patterns are disrupted with each round of DNA replication, and transcription factors may play an important role in resetting the chromatin behind the replication fork (Blythe and Wieschaus 2016; Ramachandran and Henikoff 2016). ZLD could be one such factor, binding to the freshly deposited nucleosomes to repeatedly re-establish the cis-regulatory elements required for ZGA. Ultimately, advances in chromatin profiling methods will reveal the chromatin landscape in which ZLD first accesses its binding sites, providing the context needed to more fully understand how this transcription factor functions.

Materials and Methods

Protein expression and purification

MBP-ZLD$^{1117\text{-}1487}$ (ZLD^{DBD}) was purified from *E. coli* as described previously (Hamm, Bondra, and Harrison 2015), with the exclusion of the final dialysis step. Briefly, protein was bound to amylose resin (New England Biolabs) and eluted with 20 mM maltose. Full-length ZLD protein was prepared using baculovirus infection and affinity purification. Baculovirus expression constructs containing the full-length open reading frames for *zld* (DGRC clone #LD47819) with an N-terminal FLAG affinity tag were created for use in Harrison *et al.* (2010) (Harrison, Botchan, and Cline 2010). Infection and protein purification methods were adapted from Ilves *et al.* (2010). Here, baculovirus stocks were amplified for four days in Sf9 cells grown in 150mm tissue culture dishes with Grace's media (Himedia) supplemented with 10% fetal bovine serum. Hi5 cells were grown in ESF 921 media (Expression Systems) and were infected with 1 ml of freshly amplified virus. After three days, the cells were collected on ice, centrifuged, washed with PBS containing 5mM $MgCl_2$, and centrifuged once more. The cells were resuspended in hypotonic buffer (15mM HEPES, 15mM KCl, 2mM $MgCl_2$, 0.02% Tween, 10% glycerol, 2mM BME, 1mM EGTA, 0.4mM PMSF, Roche complete protease inhibitor cocktail), flash frozen in liquid nitrogen, and kept at -80°C.

FLAG-ZLD was purified from 102 dishes of cells. Cell suspensions were thawed and dounced. KCl was then added to bring the concentration to 300mM and the lysate was cleared by centrifugation. The extract was then cleared twice more using the Beckman Avanti J-26 centrifuge with the JA 25.50 rotor (10,000 rpm for 10 min at 4°C, 16,000rpm for 10 min at 4°C). 20µM PMSF and 750µL pre-washed anti-flag M2 affinity beads (Sigma A2220) were added to the cytoplasmic extract and allowed to incubate with end-over-end mixing at 4°C for at least 2 hours. The beads were washed and applied to a disposable chromatography column (BioRad) for additional washing. Protein was eluted twice (total volume: 3ml) in buffer containing 150 mM KCl and 200µg /mL flag peptide following a 30 min incubation

with end-over end mixing at 4°C. The fractions were pooled and further concentrated via injection onto a Mono Q PC 1.6/5 anion exchange chromatography column (GE Healthcare Life Sciences) connected to an ÄKTAmicro system (GE Healthcare Life Sciences) equilibrated in 150mM KCl buffer. Using gradient elution, ZLD was eluted ~400 mM KCl. Protein concentration was determined using SYPRO-Red (Invitrogen) stained SDS-PAGE gels using BSA protein standards as a reference. Gels were visualized with a Typhoon FLA9000 using the SYPRO-Red fluorescence setting (excitation at 532nm).

Nucleosome construction

The 160bp *bnk* DNA fragment (dm3_chr3R:27018900-27019059) was amplified from Drosophila genomic DNA using one Cy5-labeled primer and one unlabeled primer (see *bnk* and primer sequences below). The DNA was ethanol precipitated and purified with AxyPrep Mag PCR Clean-up (Axygen) using a 1.8x ratio of beads to sample. The nucleosomes were reconstituted using a standard salt-urea gradient.

bnk DNA*:*

TCCTTTTTTACTTTCATAGCTTAGGTAGTGATCTCAGGTAGTTTCCCGGAATTAAGTTAGGCTGGGTATCGCCTATC
GGGAGCGGCTACCTGAACTTTTGGCACCAGCTGTCGGGGGTGAAAACTGGACCAGGTAGTCTTTAGAAGTGCACC
TATATAAG

Primers:

KS111+	Cy5-TCCTTTTTTACTTTCATAGCTT
KS112	CTTATATAGGTGCACTTCTAAA

Binding reactions

50fmol Cy5-labelled probe (DNA or nucleosome) was incubated with recombinant ZLD protein in buffer containing: 5ng poly[d-(IC)], 12.5 mM HEPES, 0.5mM EDTA, 0.5mM EGTA, 5% glycerol, 93.75µM ZnCl2, 0.375mM DTT, 75µM PMSF, 0.075 mg/ml BSA, 5 mM MgCl2, 0.00625% NP-40, 50mM KCl at room

temperature for 60 min. For competition assays, a 32 to 320-fold molar excess of unlabeled *scute* probe (*sc* WT or *sc* MUT) was added to the binding reaction (see sequences below). Reactions were run in a 4% non-denaturing polyacrylamide gel (recipe below) in 0.5X Tris–borate–EDTA. Gels were visualized with a Typhoon FLA9000 using Cy5 fluorescence setting (excitation at 635nm).

Our collaborators performed similar reactions using 40fmol Cy5-labelled probe (DNA or nucleosome) in buffer containing: 10 mM Tris-HCl (pH7.5), 1 mM MgCl2, 10 uM ZnCl2, 1 mM DTT, 50 mM KCl, 0.5 mg/ml BSA, and 5% glycerol.

Specific competitor (*sc* WT)	Cy5-GAGAGAGA**CTACCTG**TGGCTCACT
Nonspecific competitor (*sc* MUT)	Cy5-GAGAGAGA**GTAGTTC**TGGCTCACT

4% Tris Glycine Gel Recipe (50mL): 6.7 mL 30% acrylamide (4%), 2.5 mL 10x Tris–borate–EDTA (0.5x), 40.4 mL water, 350 ul 10% APS, 50 ul TEMED.

Western blotting after EMSA

After the EMSA was run as described above, the gel was cropped to fit into a BioRad Mini Trans-Blot Cell. Proteins were transferred from the gel to a 0.45µm Immobilon-P PVDF membrane (Millipore) in transfer buffer (25mM Tris, 200mM Glycine, 20% methanol) for 75 min at 500mA at 4°C. The membranes were blocked with 5% non-fat dry milk in TBST for 20 min at room temperature and then incubated with anti-histone H3 antibody overnight at 4°C (1mg/ml, 1:5000 dilution, Abcam #ab1791). The secondary incubation was performed with goat anti-rabbit IgG-HRP conjugate (1:10,000 dilution, Bio-Rad #1706515) for 1 hr at room temperature. Blots were treated with SuperSignal West Pico PLUS chemiluminescent substrate (Thermo-Scientific # 34577) and visualized using the Azure Biosystems c600.

Acknowledgements

This work was completed with the help of several collaborators, all of whom guided the project through helpful discussions. Ken Zaret and Meilin Fernandez Garcia contributed a wealth of EMSA technical knowledge and performed the series of experiments shown in Sup. Figure 4.2. Peter Lewis and Siddhant Jain assisted with our protein purifications, supplied us with histone antibodies, and generated the nucleosomes (*bnk*-nuc and 601-nuc) used in this work. Additional thanks to my labmate Stephen McDaniel, who has been a great purification assistant and will adopt this work in my absence.

Figures

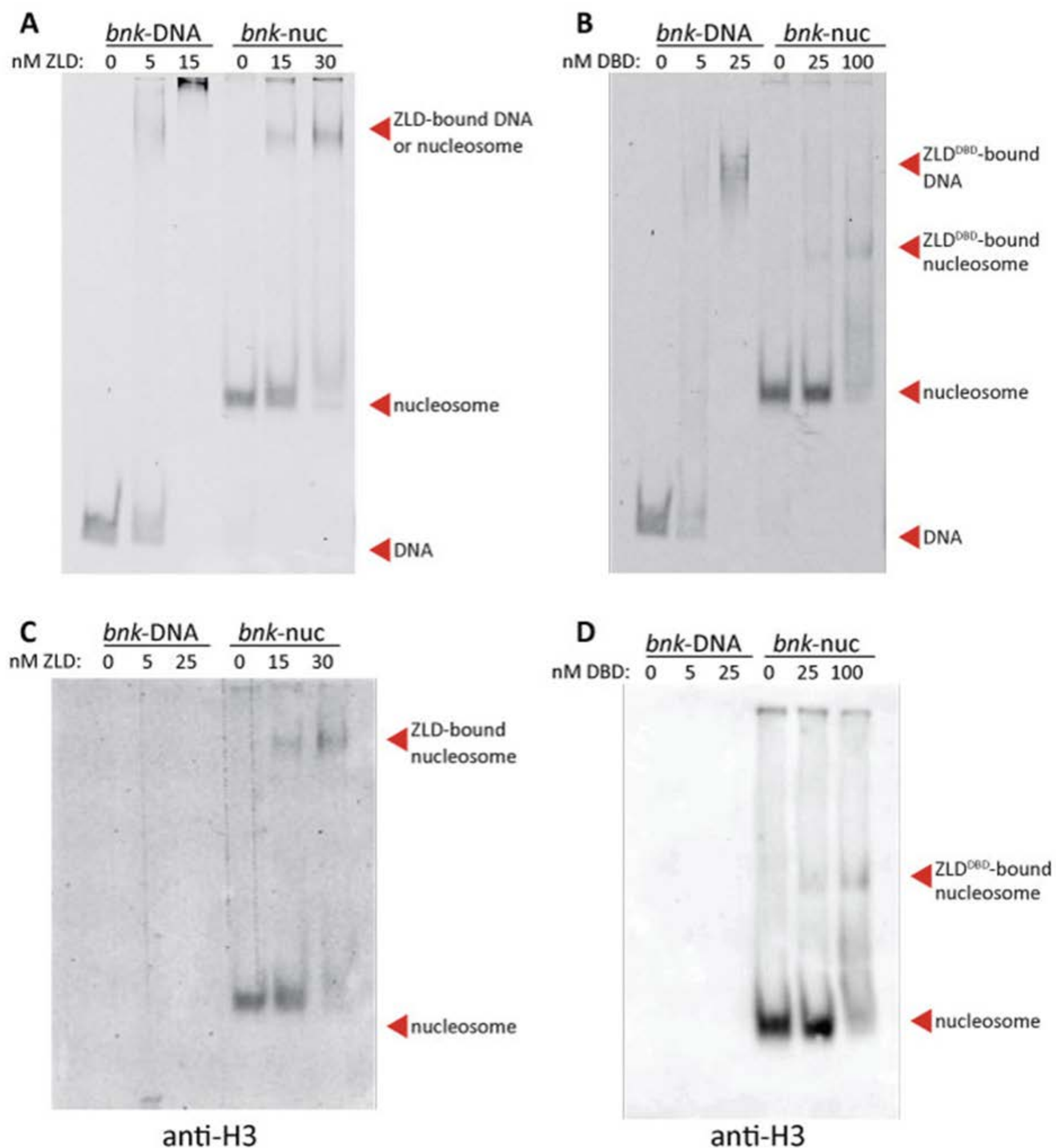


Figure 4.1: ZLD binds nucleosomes containing its DNA-binding motif.

(A-B) EMSA showing the binding of increasing amounts of ZLD to Cy5-labeled free DNA (*bnk*-DNA) and nucleosomes (*bnk*-nuc) containing four canonical ZLD-binding motifs, using full-length protein **(A)** and ZLD DNA-binding domain (ZLD^{DBD}) **(B)**. **(C)** Protein transferred onto PVDF from the EMSA shown in A and blotted for histone H3. **(D)** Protein transferred onto PVDF from the EMSA shown in B and blotted for histone H3.

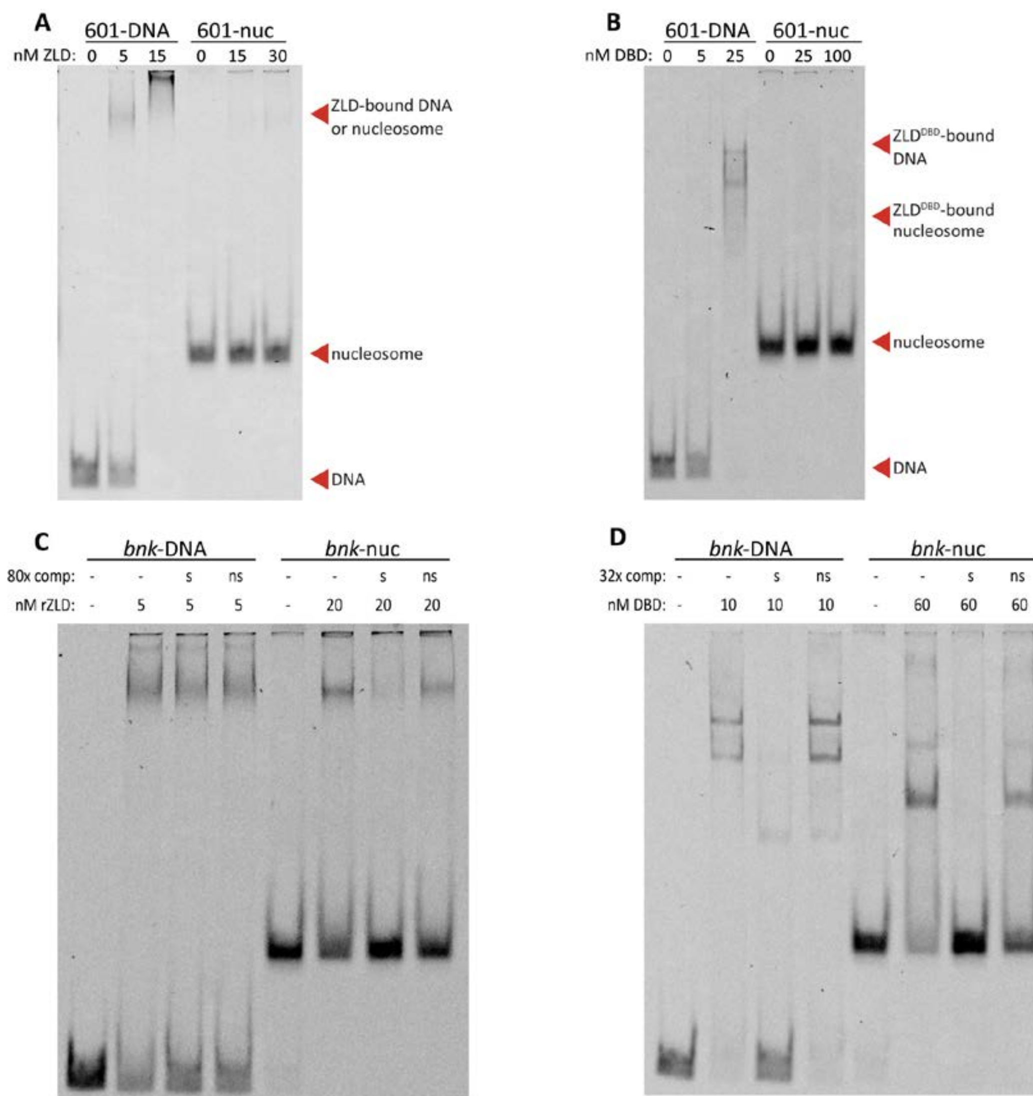
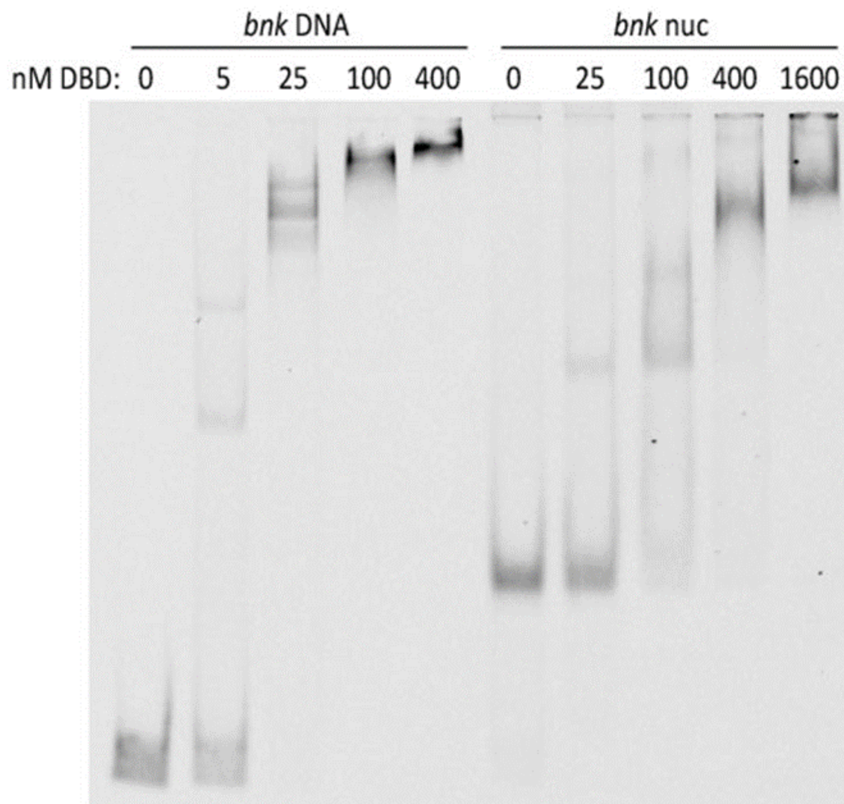


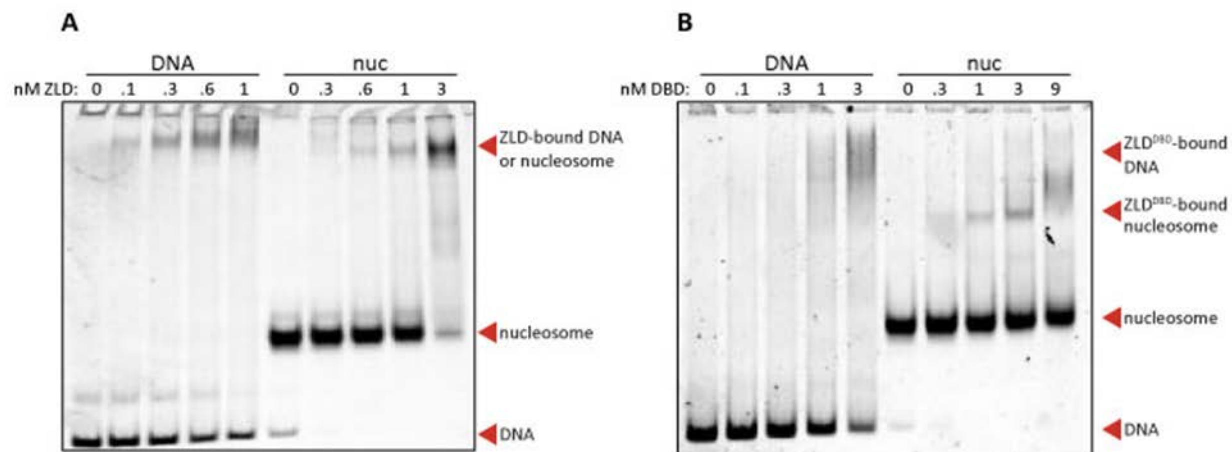
Figure 4.2: The interaction of ZLD with nucleosomes is largely sequence specific.

(A-B) EMSA showing the binding of increasing amounts of ZLD to Cy5-labeled free DNA (601-DNA) and nucleosomes (601-nuc) without ZLD-binding motifs, using full-length protein **(A)** and ZLD DNA-binding domain (ZLD^{DBD})**(B)**. **(C-D)** Competition assays showing the binding of ZLD to Cy5-labeled *bnk*-DNA and *bnk*-nuc in the presence of specific competitor (s), non-specific competitor (ns), or in the absence of competitor (-). Full-length protein was assayed using an 80-fold molar excess of competitor **(C)**, and ZLD^{DBD} was assayed using a 32-fold molar excess of competitor **(D)**.



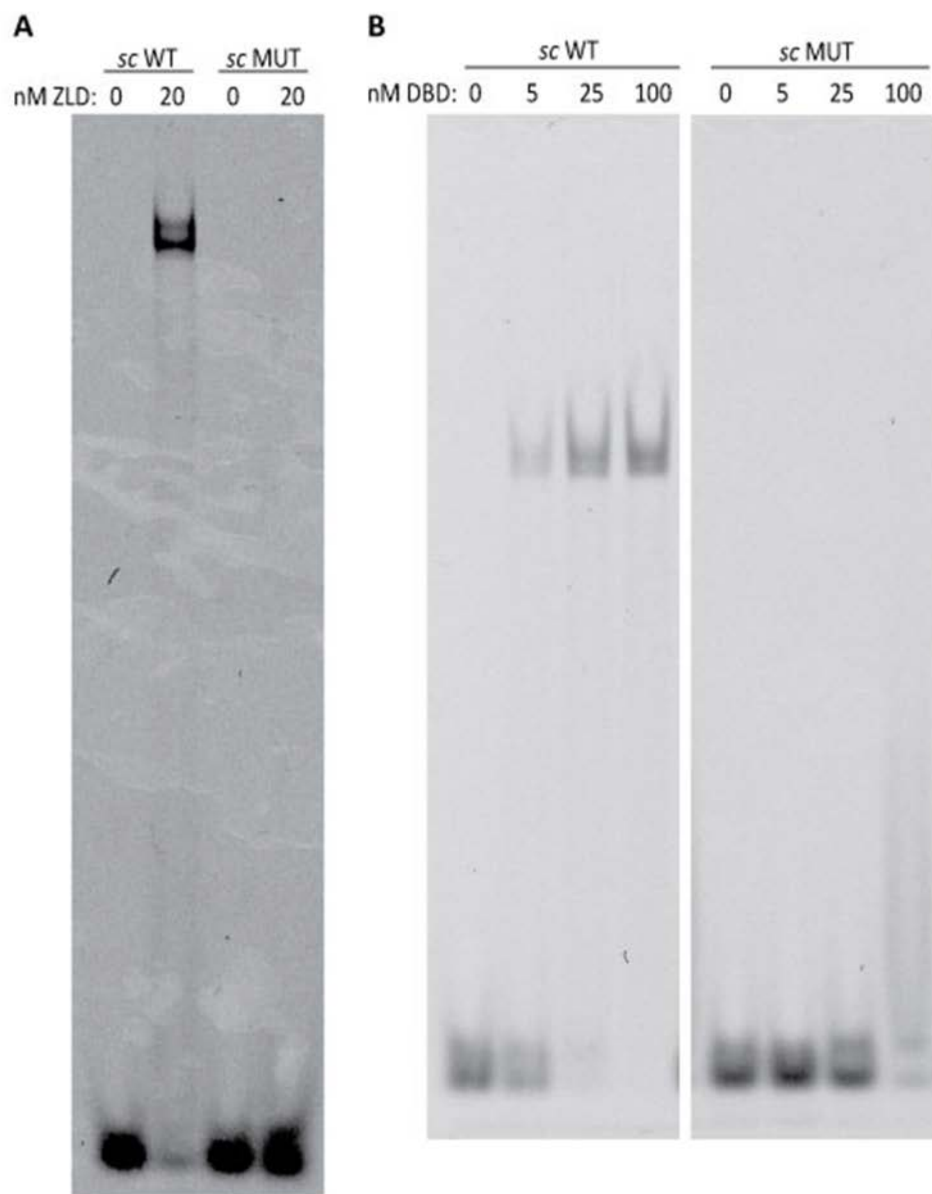
Sup. Figure 4.1: Nucleosomes are bound by multiple polypeptides at high protein concentrations.

EMSA showing the binding of increasing amounts of ZLD DNA binding domain (ZLD^{DBD}) to Cy5-labeled free DNA (*bnk*-DNA) and nucleosomes (*bnk*-nuc).



Sup. Figure 4.2: ZLD binds nucleosomes under a second set of reaction conditions.

(A-B) EMSA showing the binding of increasing amounts of ZLD to Cy5-labeled free DNA and nucleosomes containing two suboptimal ZLD-binding motifs in a reaction that did not include poly[d-(IC)], using full-length protein **(A)** and ZLD DNA-binding domain (ZLD^{DBD}) **(B).**



Sup. Figure 4.3: ZLD binding to the *scute* probe is sequence specific.

(A-B) EMSA showing the binding of increasing amounts of ZLD to Cy5-labeled *scute* probe containing a wildtype ZLD-binding motif (*sc* WT) and a probe in which this motif is mutated (*sc* MUT), using full-length protein **(A)** and ZLD DNA-binding domain (ZLD^{DBD}) **(B).**

References

Blythe, S. A., & Wieschaus, E. F. (2016). Establishment and maintenance of heritable chromatin structure during early Drosophila embryogenesis. ELife, 5, e20148. https://doi.org/10.7554/eLife.20148

Chronis, C., Fiziev, P., Papp, B., Butz, S., Bonora, G., Sabri, S., … Plath, K. (2017). Cooperative Binding of Transcription Factors Orchestrates Reprogramming. Cell, 168(3), 442–459.e20. https://doi.org/10.1016/j.cell.2016.12.016

Cirillo, L. A., Lin, F. R., Cuesta, I., Friedman, D., Jarnik, M., & Zaret, K. S. (2002). Opening of Compacted Chromatin by Early Developmental Transcription Factors HNF3 (FoxA) and GATA-4. Molecular Cell, 9(2), 279–289. https://doi.org/10.1016/S1097-2765(02)00459-8

Cirillo, L. A., McPherson, C. E., Bossard, P., Stevens, K., Cherian, S., Shim, E. Y., … Zaret, K. S. (1998). Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. Embo J, 17(1), 244–254. https://doi.org/10.1093/emboj/17.1.244

Clark, K. L., Halay, E. D., Lai, E., & Burley, S. K. (1993). Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. Nature, 364(6436), 412–420. https://doi.org/10.1038/364412a0

Hamm, D. C., Bondra, E. R., & Harrison, M. M. (2015). Transcriptional activation is a conserved feature of the early embryonic factor Zelda that requires a cluster of four zinc fingers for DNA binding and a low-complexity activation domain. The Journal of Biological Chemistry, 290(6), 3508–3518. https://doi.org/10.1074/jbc.M114.602292

Harrison, M. M., Botchan, M. R., & Cline, T. W. (2010). Grainyhead and Zelda compete for binding to the promoters of the earliest-expressed Drosophila genes. Developmental Biology, 345(2), 248–255. https://doi.org/10.1016/j.ydbio.2010.06.026

Harrison, M. M., Li, X.-Y., Kaplan, T., Botchan, M. R., & Eisen, M. B. (2011). Zelda binding in the early Drosophila melanogaster embryo marks regions subsequently activated at the maternal-to-zygotic transition. PLoS Genetics, 7(10), e1002266. https://doi.org/10.1371/journal.pgen.1002266

Hsu, H.-T., Chen, H.-M., Yang, Z., Wang, J., Lee, N. K., Burger, A., … Mango, S. E. (2015). Recruitment of RNA polymerase II by the pioneer transcription factor PHA-4. Science (New York, N.Y.), 348(6241), 1372–1376. https://doi.org/10.1126/science.aab1223

Hug, C. B., Grimaldi, A. G., Kruse, K., & Vaquerizas, J. M. (2017). Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. Cell, 169(2), 216–228.e19. https://doi.org/10.1016/j.cell.2017.03.024

Ilves, I., Petojevic, T., Pesavento, J. J., & Botchan, M. R. (2010). Activation of the MCM2-7 Helicase by Association with Cdc45 and GINS Proteins. Molecular Cell, 37(2), 247–258. https://doi.org/10.1016/J.MOLCEL.2009.12.030

King, H. W., & Klose, R. J. (2017). The pioneer factor OCT4 requires the chromatin remodeller BRG1 to support gene regulatory element function in mouse embryonic stem cells. ELife, 6. https://doi.org/10.7554/eLife.22631

Kornberg, R. D., & Lorch, Y. (1999). Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. Cell, 98(3), 285–294. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10458604

Li, X.-Y., Harrison, M. M., Villalta, J. E., Kaplan, T., & Eisen, M. B. (2014). Establishment of regions of genomic activity during the Drosophila maternal to zygotic transition. ELife, 3. https://doi.org/10.7554/eLife.03737

Liang, H.-L., Nien, C.-Y., Liu, H.-Y., Metzstein, M. M., Kirov, N., & Rushlow, C. (2008). The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. Nature, 456(7220), 400–403. Retrieved from http://dx.doi.org/10.1038/nature07388

Lowary, P. T., & Widom, J. (1998). New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. J Mol Biol, 276(1), 19–42. https://doi.org/10.1006/jmbi.1997.1494

Nien, C. Y., Liang, H. L., Butcher, S., Sun, Y., Fu, S., Gocha, T., … Rushlow, C. (2011). Temporal coordination of gene networks by Zelda in the early Drosophila embryo. PLoS Genet, 7(10), e1002339. https://doi.org/10.1371/journal.pgen.1002339

Ramachandran, S., & Henikoff, S. (2016). Transcriptional Regulators Compete with Nucleosomes Post-replication. Cell, 165(3), 580–592. https://doi.org/10.1016/j.cell.2016.02.062

Schulz, K. N., Bondra, E. R., Moshe, A., Villalta, J. E., Lieb, J. D., Kaplan, T., … Harrison, M. M. (2015). Zelda is differentially required for chromatin accessibility, transcription factor binding, and gene expression in the early Drosophila embryo. Genome Res, 25(11), 1715–1726. https://doi.org/10.1101/gr.192682.115

Sekiya, T., Muthurajan, U. M., Luger, K., Tulin, A. V, & Zaret, K. S. (2009). Nucleosome-binding affinity as a primary determinant of the nuclear mobility of the pioneer transcription factor FoxA. Genes & Development, 23(7), 804–809. https://doi.org/10.1101/gad.1775509

Soufi, A., Garcia, M. F., Jaroszewicz, A., Osman, N., Pellegrini, M., & Zaret, K. S. (2015). Pioneer Transcription Factors Target Partial DNA Motifs on Nucleosomes to Initiate Reprogramming. Cell, 161(3), 555–568. https://doi.org/10.1016/j.cell.2015.03.017

Sun, Y., Nien, C.-Y., Chen, K., Liu, H.-Y., Johnston, J., Zeitlinger, J., & Rushlow, C. (2015). Zelda overcomes the high intrinsic nucleosome barrier at enhancers during Drosophila zygotic genome activation. Genome Research, 25(11), 1703–1714. https://doi.org/10.1101/gr.192542.115

Swinstead, E. E., Miranda, T. B., Paakinaho, V., Baek, S., Goldstein, I., Hawkins, M., … Hager, G. L. (2016). Steroid Receptors Reprogram FoxA1 Occupancy through Dynamic Chromatin Transitions. Cell, 165(3), 593–605. https://doi.org/10.1016/j.cell.2016.02.067

Tadros, W., & Lipshitz, H. D. (2009). The maternal-to-zygotic transition: a play in two acts. Development, 136(18), 3033–3042. https://doi.org/10.1242/dev.033183

Takahashi, K., Yamanaka, S., Randle, D. H., Kamijo, T., Cleveland, J. L., Sherr, C. J., … al., et. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell, 126(4), 663–676. https://doi.org/10.1016/j.cell.2006.07.024

Takizawa, Y., Tanaka, H., Machida, S., Koyama, M., Maehara, K., Ohkawa, Y., … Kurumizaka, H. (2018). Cryo-EM structure of the nucleosome containing the ALB1 enhancer DNA sequence. Open Biology, 8(3). https://doi.org/10.1098/rsob.170255

Taube, J. H., Allton, K., Duncan, S. A., Shen, L., & Barton, M. C. (2010). Foxa1 Functions as a Pioneer Transcription Factor at Transposable Elements to Activate Afp during Differentiation of Embryonic

Stem Cells. Journal of Biological Chemistry, 285(21), 16135–16144. https://doi.org/10.1074/jbc.M109.088096

Vierbuchen, T., Ostermeier, A., Pang, Z. P., Kokubu, Y., Südhof, T. C., & Wernig, M. (2010). Direct conversion of fibroblasts to functional neurons by defined factors. Nature, 463(7284), 1035–1041. https://doi.org/http://www.nature.com/nature/journal/v463/n7284/suppinfo/nature08797_S1.html

Zaret, K. S., & Carroll, J. S. (2011). Pioneer transcription factors: establishing competence for gene expression. Genes Dev, 25(21), 2227–2241. https://doi.org/10.1101/gad.176826.111

Chapter 5 - Discussion and Future Directions

Transcription factors initiate zygotic genome activation

While recent studies have established that zygotic genome activation (ZGA) is broadly initiated by key transcription factors with a potent ability to reprogram the transcriptome, Zelda (ZLD) was the only known genome activator for several years. Since ZLD orthologs are only found in insects, it was initially unclear whether there would be factors that function analogously in vertebrates. Then, in 2013, two independent labs identified Nanog, SoxB1, Pou5f3 as activators of zebrafish ZGA (Leichsenring et al. 2013; M. T. Lee et al. 2013). Within the past two years, advanced low-input sequencing methods enabled the discovery of the first mammalian genome activators (Gao et al. 2018; De Iaco et al. 2017; Hendrickson et al. 2017; Whiddon et al. 2017). While these vertebrate activators are not phylogenetically related to ZLD, they appear to share key functional characteristics. Thus, this role of transcription factors in genome activation can be added to the list of phenomena discovered in flies and later recognized in higher eukaryotes

ZLD defines regions of open chromatin during ZGA

ZLD had been shown to activate hundreds of genes during ZGA (Liang et al. 2008; Nien et al. 2011), but the mechanisms by which ZLD regulates transcription were initially unclear. The structural domains of this protein offer few clues to its function, and, to this day, no protein interacting partners have been identified for ZLD. However, upon finding that ZLD binding is strongly correlated with chromatin accessibility, it was proposed that ZLD might function by establishing or maintaining regions of open chromatin (Harrison et al. 2011). To test this model, we profiled the chromatin accessibility of embryos that lack maternally supplied ZLD and determined that ZLD is required for open chromatin at hundreds of loci during ZGA (Chapter 2). Along with a co-published study (Sun et al. 2015), our data provided the first evidence that genome activators may broadly function by defining regions of open chromatin. Importantly, recent studies have shown that the newly identified vertebrate factors are

associated with chromatin accessibility during ZGA, supporting the notion that this function may be shared across species (G. Liu et al. 2018; Veil et al. 2018; Hendrickson et al. 2017; Oldfield et al. 2014; Choi et al. 2016).

ZLD had previously been shown to facilitate the binding of several transcription factors, though it was not clear how it served this function. In this study (Chapter 2), we demonstrated that ZLD potentiates the binding of other factors by establishing or maintaining sites of open chromatin. Further, we found that while ZLD remains bound to thousands of genes over the course of ZGA, ZLD-mediated accessibility is only essential for the activation of a small subset of early expressed zygotic genes. ZLD is bound but not required for chromatin accessibility at thousands of genes, including many that are not activated until the end of ZGA (Harrison et al. 2011). While our failure to identify interacting partners for ZLD suggests that this protein may not form stable interactions with other factors, it is unlikely that ZLD functions alone. Recent work in zebrafish has shown that the combinatorial binding of three genome activators (Pou5f3, SoxB1, and Nanog) is required for stable chromatin remodelling during ZGA (Veil et al. 2018). Fittingly, we found that ZLD-bound regions that do not require ZLD to remain open are enriched for a known chromatin regulator called GAGA factor (GAF). We propose that while ZLD is critical for activation of the earliest expressed zygotic genes, GAF may be important for activating genes in the later stages of ZGA. Notably, GAF binding appears to promote the recruitment of paused RNA Polymerase II and the establishment open chromatin immediately prior to the major onset of transcription (Blythe and Wieschaus 2016; Chen et al. 2013). To test this model, I have generated MS2-based reporters driven by regulatory regions that are bound by ZLD and GAF *in vivo* (Chapter 4). In addition to the wild-type version of each reporter, I have built versions in which the binding sites of ZLD, GAF, or both factors have been mutated. Comparing these wild-type and mutant reporters will allow us identify changes in the onset or pattern of expression that result from the absence of regulation by

these transcription factors. This will allow us to tease out the contributions of ZLD and GAF to activation of shared target genes and, thus, potentially identify a role for GAF in co-activation of ZGA.

ZLD may perform its role in ZGA using a unique ability to bind nucleosomes

After we established that ZLD has a role in chromatin regulation, we next wanted to determine how ZLD might facilitate chromatin accessibility. Around this time, several transcription factors that similarly poise genes for expression by maintaining regions of open chromatin were being described as "pioneer factors". This term was coined to describe a group of specialized transcription factors that can reprogram regions of silent chromatin to initiate cell-fate transitions (Zaret and Carroll 2011). The defining feature of a pioneer factor is a unique ability to bind nucleosomal DNA. By overcoming the barrier presented by nucleosomes, pioneers pave the way for additional transcription factors to bind and regulate gene expression (Zaret and Carroll 2011).

While we have shown that ZLD does not possess certain pioneer characteristics, such as slow nuclear diffusion and mitotic bookmarking, we recently demonstrated that ZLD shares the ability to bind nucleosomes *in vitro* (Chapter 4). ZLD binding is remarkably widespread in the early embryo, covering 64% of the canonical ZLD-binding motifs that are found across the genome (Harrison et al. 2011). ZLD binds nucleosomes in a sequence-specific manner. Thus, this ability could enable widespread binding by allowing ZLD to target motifs within regions of closed chromatin. However, chromatin accessibility has not been assayed in the embryo prior to ZLD binding, so it is unclear whether ZLD initially binds within closed chromatin. Several lines of evidence suggest that chromatin may exist in a naïve, globally accessible state prior to ZGA (Hug et al. 2017; Blythe and Wieschaus 2016; Strom et al. 2017; Li et al. 2014). Thus, it is possible that ZLD locates its binding motifs within a globally accessible genome and simply blocks the formation of heterochromatin at these sites. In this scenario, an ability to interact with nucleosomes could still be critical to its function. ZLD remains bound to a majority of its binding sites

across six cycles of nuclear replication. In the fly embryo, chromatin accessibility patterns appear to be disrupted with each round of DNA replication, and it has been proposed that transcription factors may be responsible for reconstructing chromatin organization after the replication fork has passed (Blythe and Wieschaus 2016; Ramachandran and Henikoff 2016). With its ability to bind nucleosomal DNA, ZLD could be uniquely adept at binding freshly deposited nucleosomes behind the fork. With each cycle of binding, ZLD could re-establish key cis-regulatory elements to enable the gradual activation of gene expression. Thus, while we have shown that ZLD can bind nucleosomes *in vitro*, to understand the biological significance of this activity, we must assess the chromatin state in which ZLD first locates its binding sites. Advanced low-input sequencing methods will ultimately allow us to overcome the pitfalls of our current chromatin profiling assays to uncover the features of the early chromatin landscape.

Towards a molecular mechanism

Despite the progress we have made towards understanding this transcription factor, it remains unclear how ZLD would induce chromatin opening upon binding a nucleosome. Unlike the canonical pioneer FoxA1, which has a DNA-binding domain that displaces histones (Clark et al. 1993; Taube et al. 2010), ZLD lacks an inherently disruptive structural domain. While ZLD appears to promote histone acetylation at its binding sites (Li et al. 2014), it may only do so indirectly, by maintaining open chromatin to allow additional factors to bind and serve this function. However, it is possible that the ability of ZLD to facilitate the binding of other factors is all that is required for its role in ZGA. ZLD binding is a robust predictor of where many additional transcription factors will later bind (Harrison et al. 2011). Thus, by simply potentiating the binding of a diverse set of factors, ZLD could direct the recruitment of chromatin regulators to produce the widespread reprogramming required for genome activation.

The ability of genome activators to elicit global transcriptional change could also stem from a role in higher-order chromatin organization. In fly embryos, the boundaries of discrete chromatin

regulatory domains are gradually established in concert with gene activation (Hug et al. 2017). These "topologically associated domains" (TADs) play an important role in transcriptional regulation by clustering specific promoters in the proximity of enhancers or silencers (Dekker and Mirny 2016). Recent work suggests that ZLD may contribute to TAD boundary insulation at a subset of sites during ZGA (Hug et al. 2017). The fact that GAF can function as an insulator protein suggests that it could also be involved in three-dimensional genome organization (Adkins, Hagerman, and Georgel 2006; Kaye et al. 2017). One way ZLD could promote TAD formation is by maintaining sites of open chromatin that facilitate the binding of architectural proteins. In support of this model, the structural protein cohesin was recently found to be enriched at sites bound by the genome activators Pou5f3 and Sox2 in zebrafish (Meier et al. 2018). Transcription factors could also impact genome organization by directly promoting enhancer clustering, a function that has been demonstrated for the factor Sox2 in embryonic stem cells (Z. Liu et al. 2014). Enhancer clusters facilitate transcription by creating a high local concentration of bound factors (Hnisz et al. 2017). Notably, ZLD was recently shown to potentiate the binding of the transcription factor Bicoid by promoting the formation of "hubs" containing high concentrations of this factor (Mir et al. 2017). Further, ZLD and GAF both contain intrinsically disordered, low-complexity domains (Adkins, Hagerman, and Georgel 2006; Hamm, Bondra, and Harrison 2015), a feature that allows certain proteins to undergo a phase transition to a hydrogel-like state (Kato et al. 2012). Thus, it is tempting to speculate that these factors could nucleate the formation of super enhancers using an inherent tendency to aggregate at high concentrations. These super enhancers could serve as transcriptional factories that enable widespread gene activation. In the future, lattice light-sheet microscopy will be used to determine whether ZLD and GAF form foci with liquid properties, testing the viability of this model.

Recent work has demonstrated that chromatin is remodeled at multiple organizational levels during ZGA, but the role of this remodeling remains poorly defined. In many cases, it is unclear whether

chromatin changes are required for transcription or whether they are simply the biproduct of transcriptional activity (Henikoff and Shilatifard 2011). Genome activators initiate widespread transcriptional change, at least in part, by shaping the chromatin landscape. Thus, these factors may serve as a link between the chromatin and transcriptional changes that constitute ZGA. A more comprehensive understanding of how these transcription factors function will help uncover the mechanisms by which these embryonic processes are coordinated.

References

Adkins, N. L., Hagerman, T. A., & Georgel, P. (2006). GAGA protein: a multi-faceted transcription factor. Biochemistry and Cell Biology, 84(4), 559–558. https://doi.org/10.1139/o06-062

Blythe, S. A., & Wieschaus, E. F. (2016). Establishment and maintenance of heritable chromatin structure during early Drosophila embryogenesis. ELife, 5, e20148. https://doi.org/10.7554/eLife.20148

Chen, K., Johnston, J., Shao, W., Meier, S., Staber, C., & Zeitlinger, J. (2013). A global change in RNA polymerase II pausing during the Drosophila midblastula transition. ELife, 2, e00861. https://doi.org/10.7554/eLife.00861

Choi, S. H., Gearhart, M. D., Cui, Z., Bosnakovski, D., Kim, M., Schennum, N., & Kyba, M. (2016). DUX4 recruits p300/CBP through its C-terminus and induces global H3K27 acetylation changes. Nucleic Acids Research, 44(11), 5161–5173. https://doi.org/10.1093/nar/gkw141

Clark, K. L., Halay, E. D., Lai, E., & Burley, S. K. (1993). Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. Nature, 364(6436), 412–420. https://doi.org/10.1038/364412a0

De Iaco, A., Planet, E., Coluccio, A., Verp, S., Duc, J., & Trono, D. (2017). DUX-family transcription factors regulate zygotic genome activation in placental mammals. Nat Genet, 49(6), 941–945. Retrieved from http://dx.doi.org/10.1038/ng.3858

Dekker, J., & Mirny, L. (2016). The 3D Genome as Moderator of Chromosomal Communication. Cell, 164(6), 1110–1121. https://doi.org/10.1016/J.CELL.2016.02.007

Gao, L., Wu, K., Liu, Z., Yao, X., Yuan, S., Tao, W., … Liu, J. (2018). Chromatin Accessibility Landscape in Human Early Embryos and Its Association with Evolution. Cell, 173(1), 248–259.e15. https://doi.org/10.1016/j.cell.2018.02.028

Hamm, D. C., Bondra, E. R., & Harrison, M. M. (2015). Transcriptional activation is a conserved feature of the early embryonic factor Zelda that requires a cluster of four zinc fingers for DNA binding and a low-complexity activation domain. The Journal of Biological Chemistry, 290(6), 3508–3518. https://doi.org/10.1074/jbc.M114.602292

Harrison, M. M., Li, X.-Y., Kaplan, T., Botchan, M. R., & Eisen, M. B. (2011). Zelda binding in the early Drosophila melanogaster embryo marks regions subsequently activated at the maternal-to-zygotic transition. PLoS Genetics, 7(10), e1002266. https://doi.org/10.1371/journal.pgen.1002266

Hendrickson, P. G., Dorais, J. A., Grow, E. J., Whiddon, J. L., Lim, J.-W., Wike, C. L., … Cairns, B. R. (2017). Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. Nat Genet, 49(6), 925–934. Retrieved from http://dx.doi.org/10.1038/ng.3844

Henikoff, S., & Shilatifard, A. (2011). Histone modification: cause or cog? Trends in Genetics, 27(10), 389–396. https://doi.org/10.1016/J.TIG.2011.06.006

Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K., & Sharp, P. A. (2017). A Phase Separation Model for Transcriptional Control. Cell, 169(1), 13–23. https://doi.org/10.1016/j.cell.2017.02.007

Hug, C. B., Grimaldi, A. G., Kruse, K., & Vaquerizas, J. M. (2017). Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. Cell, 169(2), 216–228.e19.

https://doi.org/10.1016/j.cell.2017.03.024

Kato, M., Han, T. W., Xie, S., Shi, K., Du, X., Wu, L. C., … McKnight, S. L. (2012). Cell-free Formation of RNA Granules: Low Complexity Sequence Domains Form Dynamic Fibers within Hydrogels. Cell, 149(4), 753–767. https://doi.org/10.1016/j.cell.2012.04.017

Kaye, E. G., Kurbidaeva, A., Wolle, D., Aoki, T., Schedl, P., & Larschan, E. (2017). Drosophila Dosage Compensation Loci Associate with a Boundary-Forming Insulator Complex. Molecular and Cellular Biology, 37(21), e00253-17. https://doi.org/10.1128/MCB.00253-17

Lee, M. T., Bonneau, A. R., Takacs, C. M., Bazzini, A. A., DiVito, K. R., Fleming, E. S., & Giraldez, A. J. (2013). Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. Nature, 503(7476), 360–364. https://doi.org/10.1038/nature12632

Leichsenring, M., Maes, J., Mossner, R., Driever, W., & Onichtchouk, D. (2013). Pou5f1 transcription factor controls zygotic gene activation in vertebrates. Science (New York, N.Y.), 341(6149), 1005–1009. https://doi.org/10.1126/science.1242527

Li, X.-Y., Harrison, M. M., Villalta, J. E., Kaplan, T., & Eisen, M. B. (2014). Establishment of regions of genomic activity during the Drosophila maternal to zygotic transition. ELife, 3. https://doi.org/10.7554/eLife.03737

Li, X. Y., Harrison, M. M., Villalta, J. E., Kaplan, T., & Eisen, M. B. (2014). Establishment of regions of genomic activity during the maternal to zygotic transition. Elife, 3. https://doi.org/10.7554/eLife.03737

Liang, H.-L., Nien, C.-Y., Liu, H.-Y., Metzstein, M. M., Kirov, N., & Rushlow, C. (2008). The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila. Nature, 456(7220), 400–403. Retrieved from http://dx.doi.org/10.1038/nature07388

Liu, G., Wang, W., Hu, S., Wang, X., & Zhang, Y. (2018). Inherited DNA methylation primes the establishment of accessible chromatin during genome activation. Genome Research, gr.228833.117. https://doi.org/10.1101/gr.228833.117

Liu, Z., Legant, W. R., Chen, B.-C., Li, L., Grimm, J. B., Lavis, L. D., … Tjian, R. (2014). 3D imaging of Sox2 enhancer clusters in embryonic stem cells. ELife, 3, e04236. https://doi.org/10.7554/eLife.04236

Meier, M., Grant, J., Dowdle, A., Thomas, A., Gerton, J., Collas, P., … Horsfield, J. A. (2018). Cohesin facilitates zygotic genome activation in zebrafish. Development, 145(1), dev156521. https://doi.org/10.1242/dev.156521

Mir, M., Reimer, A., Haines, J. E., Li, X.-Y., Stadler, M., Garcia, H., … Darzacq, X. (2017). Dense Bicoid hubs accentuate binding along the morphogen gradient. Genes & Development, 31(17), 1784–1794. https://doi.org/10.1101/gad.305078.117

Nien, C. Y., Liang, H. L., Butcher, S., Sun, Y., Fu, S., Gocha, T., … Rushlow, C. (2011). Temporal coordination of gene networks by Zelda in the early Drosophila embryo. PLoS Genet, 7(10), e1002339. https://doi.org/10.1371/journal.pgen.1002339

Oldfield, A. J., Yang, P., Conway, A. E., Cinghu, S., Freudenberg, J. M., Yellaboina, S., & Jothi, R. (2014). Histone-fold domain protein NF-Y promotes chromatin accessibility for cell type-specific master transcription factors. Molecular Cell, 55(5), 708–722. https://doi.org/10.1016/j.molcel.2014.07.005

Ramachandran, S., & Henikoff, S. (2016). Transcriptional Regulators Compete with Nucleosomes Post-replication. Cell, 165(3), 580–592. https://doi.org/10.1016/j.cell.2016.02.062

Strom, A. R., Emelyanov, A. V., Mir, M., Fyodorov, D. V., Darzacq, X., & Karpen, G. H. (2017). Phase separation drives heterochromatin domain formation. Nature, 547(7662), 241–245. https://doi.org/10.1038/nature22989

Sun, Y., Nien, C.-Y., Chen, K., Liu, H.-Y., Johnston, J., Zeitlinger, J., & Rushlow, C. (2015). Zelda overcomes the high intrinsic nucleosome barrier at enhancers during Drosophila zygotic genome activation. Genome Research, 25(11), 1703–1714. https://doi.org/10.1101/gr.192542.115

Taube, J. H., Allton, K., Duncan, S. A., Shen, L., & Barton, M. C. (2010). Foxa1 Functions as a Pioneer Transcription Factor at Transposable Elements to Activate Afp during Differentiation of Embryonic Stem Cells. Journal of Biological Chemistry, 285(21), 16135–16144. https://doi.org/10.1074/jbc.M109.088096

Veil, M., Yampolsky, L., Gruening, B., & Onichtchouk, D. (2018). Pou5f3, SoxB1 and Nanog remodel chromatin on High Nucleosome Affinity Regions at Zygotic Genome Activation. BioRxiv, 344168. https://doi.org/10.1101/344168

Whiddon, J. L., Langford, A. T., Wong, C.-J., Zhong, J. W., & Tapscott, S. J. (2017). Conservation and innovation in the DUX4-family gene network. Nat Genet, 49(6), 935–940. Retrieved from http://dx.doi.org/10.1038/ng.3846

Zaret, K. S., & Carroll, J. S. (2011). Pioneer transcription factors: establishing competence for gene expression. Genes Dev, 25(21), 2227–2241. https://doi.org/10.1101/gad.176826.111