THE UNIVERSITY OF CHICAGO

BREAKING ANTERIOR-POSTERIOR SYMMETRY IN THE MOTH FLY CLOGMIA ALBIPUNCTATA

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

COMMITTEE ON DEVELOPMENT, REGENERATION, AND STEM CELL BIOLOGY

BY

EZRA AMIRI

CHICAGO, ILLINOIS

AUGUST 2025

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to first thank my advisor, Urs Schmidt-Ott for his continued support and enthusiasm throughout my long PhD journey. Urs, thank you for your guidance, support and patience. I would like to sincerely thank my committee members Ilaria Rebay, Edwin "Chip" Ferguson and Shelby Blythe. Thank you Ilaria for your candor and pragmatism. Thank you Chip for your curiosity and uncanny ability to ask good questions. Shelby, thank you for all that you have taught me and for your mentorship.

To my colleagues in the Schmidt-Ott lab I am extremely grateful, you have all contributed your time and talents to help me. Thank you Ayse Tenger-Trolander, Koray Kasan, Muzi Li, and Natasha Megherea. I would like to thank my collaborators, Alex Julian and Misha Ludwig for their dedication and hard work.

I would like to thank my family chosen and blood. Thank you for pouring into me, supporting and believing in me. I specifically thank my mother, Angela, for instilling her values, intelligence, and creativity. Thank you, Chris, for understanding me and checking in on me. Thank you to my partner and best friend Jenny. Jenny, you have always had my back, and you will forever have my heart; I love you and thank you for everything. To the rest of my people, I know you know who you are, and I'll tell you when I see you.

ABSTRACT

Forming the anterior-posterior (AP) body axis is a fundamental process during embryogenesis. In flies (Diptera), symmetry breaking along the AP axis requires anterior determinants (ADs) that differ between species. In *Drosophila melanogaster*, localized mRNA of *bicoid* serves as the AD. Once translated, the Bicoid protein competes with nucleosomes and initiates symmetry-breaking along the AP axis by promoting chromatin accessibility at the loci of segmentation genes that are required to pattern the anterior embryo. The ADs of other dipteran insects are encoded by transcription factor (TF) genes that are unrelated to Bicoid, and little is known about they function and if they act through a shared mechanism. This study uses a *de novo* assembly and annotation of the genome of the moth fly *Clogmia albipunctata* to investigate whether the AD of this species initiates symmetry breaking along the AP axis during zygotic genome activation at the level of chromatin accessibility. In Clogmia, a maternally expressed transcript isoform of the pair-rule segmentation gene *odd-paired* is localized in the anterior egg and has been co-opted as the AD. I first describe the dynamics of chromatin accessibility during the last nuclear cycles before gastrulation. I find that in Clogmia, similar to Drosophila, the chromatin landscape progressively opens throughout development encompassing widespread zygotic transcription. However, following their initial opening, many regions close at earlier nuclear cycles in Clogmia than when such widespread changes occur in Drosophila. I investigate how knockdown of the pioneer factor gene *zelda* affects the chromatin landscape in Clogmia and find that Clogmia's *zelda* homolog is important for both opening and closing chromatin. I then show that maternal *odd-paired* activity promotes chromatin accessibility and anterior gene expression during the early phase of zygotic genome activation at Clogmia's *homeobrain* and *sloppy-paired* loci. These loci differ from the key targets of Bicoid but given that homologs

homeobrain and *sloppy-paired* have early anterior expression domains and function in head development in Drosophila and other insects, these genes may serve a more widely conserved role in the initiation of anterior pattern formation. I conclude that unrelated dipteran ADs initiate anterior-posterior axis-specification at the level of enhancer accessibility through distinct sets of target genes. My study motivates further investigation into how the network downstream of the primary AD targets converges in different dipteran species to yield the conserved segmented body plan.

INTRODUCTION

1.1 Prelude

Developmental gene networks drive all aspects of how a single fertilize egg grows and transforms into a complete organism with specialized cell types, while the evolution of such networks underpin biological robustness and diversity. Many regulatory networks are complex and difficult to study. However, the developmental system that organizes the body plan of flies (Diptera) is an experimental system that is amenable to genetic manipulation not only in the genetic model organism *Drosophila melanogaster* but also in many other fly species, due to the syncytial nature of early embryogenesis in these species. Flies have a segmented body plan that is organized from anterior to posterior by a cascade of transcription factor (TF) genes known as the segmentation cascade. The segmentation cascade initiates during the syncytial phase of embryogenesis and initially only requires gene products provided by the mother. One of those, called the anterior determinant (AD), is necessary and sufficient to establish anterior-posterior (AP) polarity of the segmentation cascade. The segmentation cascade of flies has been studied extensively in *Drosophila melanogaster,* but it is not known to what extent this gene regulatory network is conserved in other fly species. This question gained prominence when it was found that AD genes are not conserved across the dipteran clade, even though these species share many cellular and molecular mechanisms of early development. This discovery provides an excellent entry point to examine how developmental gene networks evolve in the context of early embryogenesis, and how their outcome, the largely conserved body plan, is preserved over substantial evolutionary distances. This study uses the moth fly, *Clogmia albipunctata*

(Psychodidae), to investigate the evolutionary diversification of the segmentation cascade between fly species that evolved independently for about 250 million years and specifically addresses the mechanisms by which the AD of *Clogmia albipunctata* initiates symmetry breaking symmetry and segmental patterning in comparison with the unrelated AD of Drosophila melanogaster.

1.2 Syncytial development and zygotic genome activation in the fruit fly *Drosophila melanogaster*

1.2.1 Overview of syncytial development

Early embryo development in flies has long been a fruitful system to address questions relating to morphogenesis, transcriptional dynamics, fate specification and sex determination (Bothma et al., 2014; Calderon et al., 2022; Cusanovich et al., 2018; Fierling et al., 2022; Garcia et al., 2013; Hammonds et al., 2013; Keyes et al., 1992; M. Ray et al., 2023; Young et al., 1993). In the fly, early embryogenesis is unique as the embryo develops as a syncytium. After two flies mate and the fertilized egg is laid, the pronuclei fuse forming the first genetically competent nucleus. This nucleus kicks off a period of intense proliferation as it rapidly replicates its DNA and divides without cytokinesis leading to two nuclei sharing the egg cytoplasm. The nuclei of the embryo repeatedly undergo nearly synchronous cycles of DNA replication and mitotic division without cytoplasmic cleavage termed nuclear cycles (abbreviated NC). There are a total of 13 nuclear divisions (14 nuclear cycles; NC1-NC14) before the embryo cellularizes and enters gastrulation. The first seven divisions occur deep within the egg cytoplasm and the duration of the nuclear cycles are short (Blumenthal et al., 1974; Edgar et al., 1994; Shermoen et al., 2010).

Due to the short duration of the initial nuclear cycles zygotic transcription is largely suppressed and instead early syncytial development is reliant on maternal factors supplied within the egg. At the start of the NC8, the nuclei begin to migrate outward. By NC10, they have settled in the periphery of the yolk sac, forming a single layer of nuclei, referred to as the syncytial blastoderm. The next four mitotic cycles of the embryo's blastoderm nuclei progressively lengthen due to the addition of a gap phase and an increase in the duration of the DNA replication phase (Blythe & Wieschaus, 2015a; Foe et al., 1993). At NC14, the cell cycle slows substantially while widespread zygotic transcription initiates and individual nuclei are enveloped in a plasma membrane forming cells (cellularization). Shortly after the cellular blastoderm is formed, morphogenic movements of gastrulation start.

1.2.2. Maternal-to-zygotic transition (MZT)

During oogenesis the fly egg is filled with mRNAs from nearly two-thirds of the protein-coding genome (Tadros et al., 2007; Thomsen et al., 2010). For the embryo to faithfully initiate transcription, cellularize and gastrulate it must transition from maternal to zygotic control. This phase of development is known as the Maternal-to-Zygotic transition (MZT) (Tadros & Lipshitz, 2009). The maternal transcriptome is transformed through selective degradation of maternally supplied factors and the initiation of zygotic transcription (Tadros & Lipshitz, 2009). Additionally, the maternal program of expedited cell cycling is dramatically slowed by zygotically introduced cell cycle remodeling (Blythe & Wieschaus, 2015a). Egg activation also triggers the translation of RNA-binding proteins such as Smaug, whose activity leads to the degradation of many maternal transcripts (Tadros et al., 2003, 2007). As Smaug protein

accumulates, it binds to the 3'UTR of target mRNAs and promotes RNA degradation by recruiting the CCR4-NOT-deadenylase complex (Chen et al., 2014; Semotok et al., 2005; Smibert et al., 1996; Tadros et al., 2007). In addition to the early degradation pathways, maternal RNAs are targeted for degradation by zygotically expressed microRNAs (Bushati et al., 2008). Simultaneously, ubiquitin-ligase complexes mediate clearance of a subset of the maternal proteome (Cao et al., 2020; Zavortink et al., 2020).

Zygotic transcription occurs gradually and proceeds in two waves, a minor wave starting ~NC8 and a major wave during NC14. Collectively the minor and major waves make up a process known as Zygotic Genome Activation (ZGA). At the minor wave only a few hundred genes are expressed starting at NC8 (Chen et al., 2013; De Renzis et al., 2007; Kwasnieski et al., 2019; Lott et al., 2011). Genes expressed in the minor wave are primarily involved in processes that unfold early in embryogenesis, including sex determination, dosage compensation, cell cycle regulation, and early embryonic patterning (Chen et al., 2013; Liang et al., 2008; ten Bosch et al., 2006). The major wave by contrast encompasses global zygotic transcription of thousands of genes (Blythe & Wieschaus, 2015b; Chen et al., 2013). Without ZGA, the transition from the maternal to the zygotic transcriptome would be incomplete and development would arrest.

During this developmental period, the zygote also implements regulatory mechanisms that slow the maternally expedited cell cycle, a process known as the Midblastula Transition (MBT) (**Figure 1.1**). Additionally, origins of DNA replication change, and the genome is no longer simultaneously replicated (reviewed in (Blythe & Wieschaus, 2015a)). At the initiation of the major wave of zygotic activation poised RNA polymerase II (RNAP II) is established *de novo* at promoters. Widespread RNAP II loading triggers a DNA replication checkpoint that slows cell cycle progression by increasing the length of S phase (Blythe & Wieschaus, 2015b).

Introduction of this checkpoint mitigates any conflicts between the RNAP II and the replication machinery and ensures that the DNA is fully replicated before entering mitosis. Major ZGA also leads to the expression of zygotic genes necessary to further regulate the cell cycle. The rapid nuclear cycles that occur before the MBT are driven by high levels of cyclin/Cdk1 (CDK) activity conferred in part by Cdc25 phosphatase (Blumenthal et al., 1974; Di Talia et al., 2013; Edgar et al., 1994; Edgar & Datar, 1996; Farrell et al., 2012; Farrell & O'Farrell, 2013). The Cdc25 phosphatase Twine is maternally supplied and acts to inhibit Wee1 from inhibiting the activity of CDK through phosphorylation (Di Talia et al., 2013; Farrell & O'Farrell, 2013; Price et al., 2000; Stumpff et al., 2004). Zygotic expression of a gene called *tribbles* leads to destabilization of Twine protein and its subsequent degradation allowing for inhibitory phosphorylation Cdk1 (Di Talia et al., 2013; Edgar & Datar, 1996; Farrell & O'Farrell, 2013). Another zygotic gene, *fruhstart*, is highly expressed at NC14 and functions to inhibit mitotic entry through cyclin binding and downregulation of CDK activity (Gawliński et al., 2007; Großhans et al., 2003).

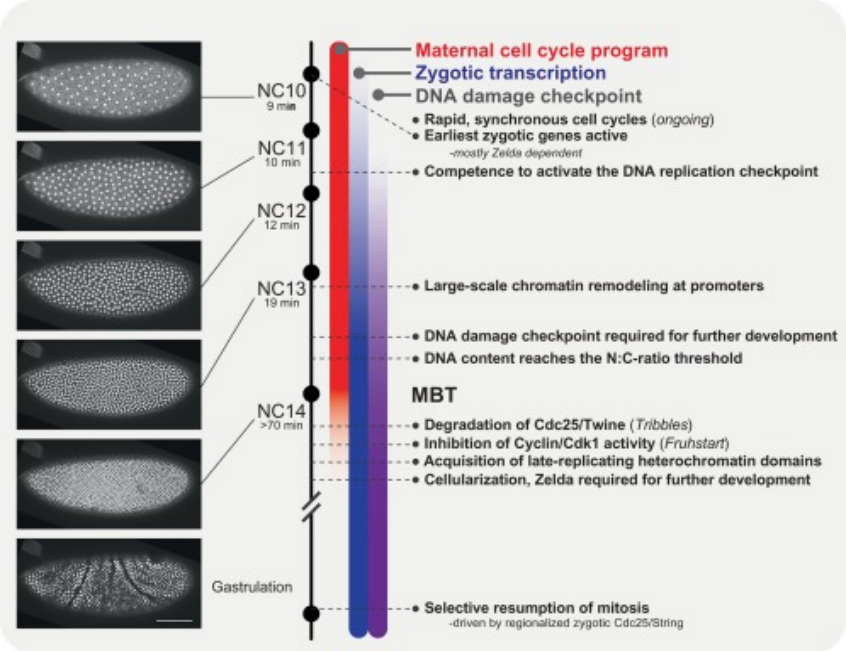

Figure 1. 1 Overview of key events during syncytial development.

Maternal control is transferred progressively to the zygote. Maternally supplied factors are degraded, and the maternal cell cycle program decreases (red). At minor ZGA the zygote initiates transcription and transcriptional activity increases (blue). DNA replication checkpoints (blue) are implemented, and major cell cycle remodeling occurs at the **Midblastula Transition (MBT)**. Images on the left are frames from a time-lapse of a Histone H2Av-GFP transgenic embryo spanning formation of the syncytial blastoderm at NC10, cellularization at NC14 and early gastrulation. Anterior, left. Dorsal, up. Figure from (Blythe & Wieschaus, 2015a).

1.2.3 Regulation of ZGA

The mechanisms regulating ZGA are complex and interconnected. Here I will briefly summarize the current understanding of how genomic activation is controlled and then focus on the importance of chromatin accessibility and the role of pioneer TFs in the proceeding section. Many of the earliest expressed embryonic genes contain promoters with the "TAGteam" motif CAGGTAG that is bound and activated by the TF Zelda (Zld) (De Renzis et al., 2007; Liang et al., 2008; Staudt et al., 2006; ten Bosch et al., 2006). *Zelda* is maternally deposited as an mRNA and is rapidly translated within the first hour of development (Nien et al., 2011). Zld binds thousands of promoters leading to the first wave of zygotic transcription at NC8 and initiates major ZGA by recruiting RNAP II to promoters at NC13 (Blythe & Wieschaus, 2015b; Harrison et al., 2011; Liang et al., 2008; Nien et al., 2011). Zld's function as a pioneer factor in driving chromatin accessibility will be further discussed in the subsequent section. As illustrated with Zld, accumulation of transcriptional activators serves as an internal timer of how soon zygotic transcription can start (Schulz & Harrison, 2019). However, timing of ZGA is not regulated by a simple activation mechanism alone, instead the activity of ZGA activators is constrained by factors that suppress transcription. These proposed mechanisms fall into three areas: 1) activity of global repressor, 2) kinetic limitations on transcription due to the cell cycle, and 3) modifications of histones that promote a repressive chromatin state.

Central to models of ZGA is the concept of a global repressor that is titrated away as the ratio of nuclear content to cytoplasm (termed nuclear-to-cytoplasm ratio, abbreviated N:C ratio) increases. Initially the volume of cytoplasm vastly outweighs the single nucleus of the embryo. However, as development progresses the number of nuclei increases while the volume of cytoplasm stays constant. As a result, the N:C ratio increases. This observation led to a model in which a global repressor is maternally supplied in excess to the embryo. As DNA content increases the repressor is titrated so that its concentration is no longer sufficient to suppress transcription (Newport & Kirschner, 1982a, 1982b). This mode was initially supported by experiments in Xenopus where manipulation of the N:C ratio by injecting exogenous DNA into the embryo spurred precocious transcription (Newport & Kirschner, 1982b). Subsequent studies in frog embryos demonstrated that titrating the concentration of maternally supplied histones either by increasing DNA content or nuclear volume can overcome histone-mediated transcriptional repression and induce zygotic transcription (Amodeo et al., 2015; Jevtić & Levy, 2015). Likewise, in zebrafish, the onset of zygotic transcription was linked to a reduction in unbound histones leading to a model where TFs must outcompete histone forming nucleosomes to initiate transcription (Joseph et al., 2017).

There is little evidence of a maternally supplied TF that acts to repress zygotic activation globally in Drosophila. Early studies in Drosophila initially proposed that the maternally supplied gene *tramtrack (ttk),* encoding the zinc-finger TF Tramtrack (Ttk), plays a broad role in suppressing ZGA (Harrison & Travers, 1990; Pritchard & Schubiger, 1996). These studies found that depletion of *ttk* lead to precocious transcription of a zygotic gene, *fushi tarazu*, and conversely, increasing the dose of *ttk* resulted in strong repression of *fushi tarazu* and other zygotically expressed genes (Brown et al., 1991; Brown & Wu, 1993; Pritchard & Schubiger,

1996). Ttk repression may be mediated by the repressive NuRD (nucleosome remodeling deacetylase) complex, as a more recent study demonstrated, Ttk interacts with NuRD subunits in turn recruiting the complex to target loci (Murawsky et al., 2001; Reddy et al., 2010). However, support for a global repressive role of Ttk remains limited, as its activity has been documented at only a few zygotically expressed genes, with some showing no evidence of Ttk-mediated repression (Brown & Wu, 1993; Pritchard & Schubiger, 1996). Moreover, Ttk is largely absent from chromatin during the early nuclear cycles and is therefore unable to function as a repressor at those stages (S. Blythe, personal communication). Instead, as observed with other species, evidence in Drosophila suggests that histones play a role in global repression of zygotic transcription. Modulation of maternal histone concentration by depletion or overexpression results in early or delayed ZGA, respectively (Chari et al., 2019). A subsequent study led to a model suggesting that histone concentration could mediate the timing of zygotic transcription by regulating the activity of checkpoint kinase Chk1. When bound to the tail of histone H3, Chk1 is inactive and cannot induce a DNA replication checkpoint (Shindo & Amodeo, 2021). As nuclear cycles progress and H3 levels wane, Chk1 activity increases causing the length of the cell cycle to increase (Harrison et al., 2023; Shindo & Amodeo, 2021).

Indeed, nuclear cycle duration is another mechanism that significantly limits zygotic transcription. Early nuclear division cycles are short because the gap phases of the mitotic cycles (G1 and G2) are largely suppressed, and DNA synthesis (S) is followed immediately by mitosis (M). Consequently, the window for RNAP II dependent transcription is reduced. Genes that are transcribed within the minor wave of ZGA are therefore usually short in length (~1-2 KB) and lack intron sequences (Chen et al., 2013; De Renzis et al., 2007; Heyn et al., 2014). Additionally, these genes are often regulated by strong promoters that contain TATA boxes, which are

associated with rapid RNAP II initiation and lack of polymerase stalling (Chen et al., 2013; Pimmett et al., 2021). Studies have shown that Drosophila embryos arrested in interphase at NC12 induce widespread transcription consistent with ZGA, suggesting that nuclear cycle duration influences transcriptional output (Strong et al., 2020). Moreover, the start of major ZGA is hallmarked by broad loading of RNAP II to promoters at NC13, triggering a DNA replication checkpoint that dramatically slows S phase (Blythe & Wieschaus, 2015b; Seller & O'Farrell, 2018). The latter study underscores the necessity of cell cycle remodeling during ZGA to avoid conflict between RNAP II and replication machinery and proposes the replication checkpoint as an indirect N:C sensor.

Transformations in chromatin structure, organization and posttranslational modification of histones occur throughout the MZT. However, there is little evidence that repressive marks are deposited globally after fertilization. To date only one study observed maternal inheritance of the histone modification Histone H3 lysine-27 trimethylation (H3K27me3) associated Polycomb-mediated repression. However, maternally inherited H3K27me3 was not observed globally and was limited to roughly 30 genomic regions (Zenk et al., 2017). Instead, global acquisition of histone modifications occurs gradually once the syncytial blastoderm is established. Starting at NC12, active histone mark Histone H3 lysine-4 trimethylation (H3K4me3) and repressive mark Histone H3 lysine-9 trimethylation (H3K9me3) emerge, demarcating euchromatin and heterochromatin respectively (Rudolph et al., 2007). Additional histone modifications including those marking active transcription (H3K27ac and H3K36me3) as well as the forementioned repressive mark H3K27me3, accumulate globally when major ZGA is well underway (Chen et al., 2013; Li et al., 2014). Likewise, higher order genome organization such as topologically associating domains (TADs) are also established with major zygotic transcription (Hug et al.,

2017). The physical access to chromatinized DNA, known as chromatin accessibility, is another important level of chromatin organization and will be discussed in the following section.

1.2.4 Zygotic genome activation; transitions in the chromatin landscape and the role of pioneer transcription factors

Nucleosome occupancy at promoters and enhancers hinder the ability of TFs to bind to these cis-regulatory elements and initiate transcription. Thus, acquisition of chromatin accessibility globally is essential to drive widespread transcription. Chromatin accessibility can be measured by the Assay for Transposase-Accessible Chromatin Sequencing (ATAC-seq) (Buenrostro et al., 2015). ATAC-seq on single embryos spanning NC11-N13 demonstrate the sequential opening of cis-regulatory elements leading up to ZGA and illustrate that garnering chromatin accessibility is a limiting factor of genome activation (Blythe & Wieschaus, 2016). Regions that gain accessibility ("open") starting at NC11 are predominately enhancers. Following, promoters and insulators start to open at NC12 and dynamically gain accessibility coinciding with global recruitment of RNAP II at NC13 (Blythe & Wieschaus, 2015b, 2016). Moreover, transient losses in accessibility caused by DNA replication and mitosis are rapidly re-established at the beginning of each nuclear cycle and subsequently increase corresponding to increased transcriptional output. These observations emphasize the importance of regulatory factors that work to initiate and maintain chromatin accessibility with each nuclear cycle to facilitate ZGA.

Pioneer factors are a distinct set of TFs that bind to nucleosome occupied DNA and promote nucleosome removal to make cis-regulatory regions accessible. Additionally, pioneer

factors facilitate downstream binding of other TFs (Iwafuchi-Doi & Zaret, 2014; Zaret, 2020). In Drosophila, the pioneer factor Zld acts early in the embryo during the minor wave of ZGA (Harrison et al., 2011; Liang et al., 2008; Sun et al., 2015). Zld acts by overcoming a high nucleosome barrier to bind to nucleosome bound DNA, leading to nucleosome removal and an open chromatin state (**Figure 1.2 A**) (Schulz et al., 2015; Sun et al., 2015). Zld's activity is required throughout genome activation and Zld binding enables the binding of other TFs including those involved in embryonic patterning (Brennan et al., 2023; Foo et al., 2014; McDaniel et al., 2019; Mir et al., 2017, 2018; Nien et al., 2011). Additional pioneer factors, GAGA Factor (GAF) and CLAMP, act independently and cooperatively with Zld to establish open chromatin and drive zygotic gene expression once major ZGA is underway (**Figure 1.2 B**) (Colonnetta et al., 2021; J. Duan et al., 2021; Gaskill et al., 2021; Moshe & Kaplan, 2017).

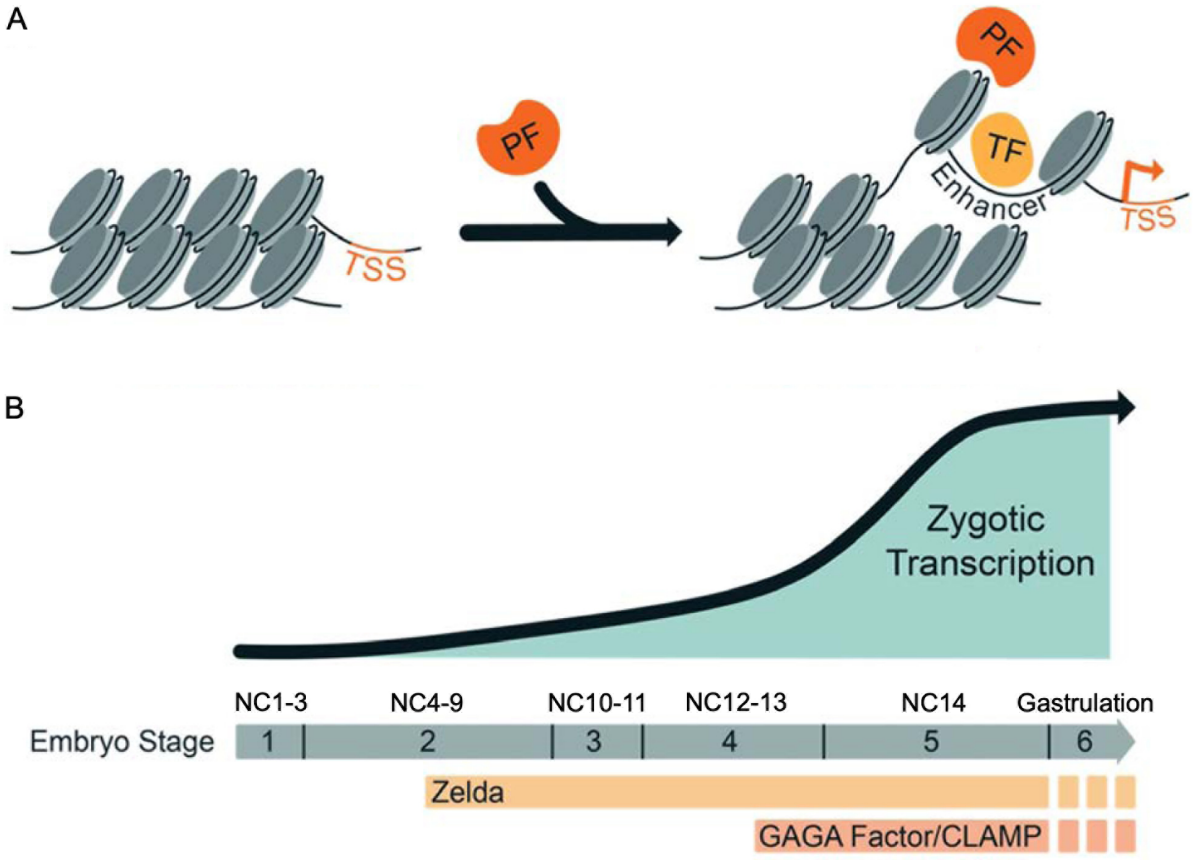

Figure 1. 2 Pioneer factors establish accessible chromatin to facilitate ZGA.
A) Pioneer factors (PF) bind DNA with high nucleosome occupancy and promote chromatin remolding leading to accessible region of chromatin. Pioneer factors recruit or facilitate binding of other factors including transcription factors (TF, shown), chromatin remodelers and RNAP II to regions of open chromatin. B) In Drosophila multiple pioneer factors (Zelda, GAGA factor, CLAMP) act temporally to regulate ZGA. Figure modified from figures in (Larson et al., 2021).

1.3 Anterior-Posterior axis specification and segmentation in *Drosophila melanogaster*

The adult fly has a segmented body consisting of three head (mandibular, maxillary, and labial), three thoracic (T1-T3) and eight abdominal (A1-A8) segments. The segmental organization is established during blastoderm stages of embryogenesis by a cascade of TFs and signaling pathways, known as the segmentation cascade. There are maternal inputs that establish AP polarity and zygotic inputs, commonly subdivided into three sequential tiers: the gap, pair-

rule and segment-polarity phases. The gap and pair-rule genes encode TFs. They function upstream of the segment polarity genes, which encode TFs as well as components of signaling pathways. The general logic of the cascade is to progressively restrict cell fates through successive refinement of gene expression. Ultimately, each segment contains a signaling center, known as parasegmental boundary, which establishes polarity within each segment (Lawrence, 1992).

1.3.1 Maternal inputs

The symmetry breaking maternal genes, also called maternal coordinate genes, provide the initial cues to initiate patterning. There are three coordinate systems that are relevant to patterning along the AP axis in *Drosophila melanogaster*: 1) The anterior determinant (AD), 2) the posterior germ plasm and 3) the terminal system. mRNA of the AD and a component of the germ plasm are localized to the anterior and posterior poles, respectively (Driever & Nüsslein-Volhard, 1988a; Ferrandon et al., 1997; Macdonald & Kerr, 1998; C. Wang & Lehmann, 1991; Wharton & Struhl, 1991). Once translated, these mRNAs form respective anterior-posterior and posterior-anterior protein gradients (Driever & Nüsslein-Volhard, 1988a; C. Wang & Lehmann, 1991). In *Drosophila melanogaster*, both gradients contribute independently to the embryo's AP polarity and function in distinct ways, but the posterior system is dispensable under certain conditions in Drosophila (Hülskamp et al., 1989; Irish et al., 1989; Struhl, 1989) and missing in some fly species, including the model organism of this study (see below). The terminal system involves localized signaling at the poles of the embryo by the receptor tyrosine kinase Torso but functions symmetrically in the absence of the other coordinate systems (Casanova & Struhl, 1989; Jiménez et al., 2000; Smits & Shvartsman, 2020).

1.3.2 The essential role of the anterior determinant gene, *bicoid*

In Drosophila, the AD gene, called *bicoid (bcd)*, encodes a homoedomain TF. Bicoid (Bcd) protein functions along the AP axis as a concentration dependent transcriptional activator of the early segmentation network (Driever et al., 1989; Driever & Nüsslein-Volhard, 1988a, 1988b; Jaeger, 2011; Nasiadka et al., 2002; Struhl et al., 1989). Multiple gap and pair-rule genes receive input from Bcd at different concentrations (Jaeger, 2011; Nasiadka et al., 2002), but Bcd's critical role in establishing AP-polarity is largely limited to the anterior embryo. Bcd deficient embryos do not develop a head and thorax and partially duplicate posterior structures (Driever et al., 1990). The asymmetric "double abdomen" phenotype is observed because maternally supplied *hunchback* (*hb*) is active in an AP gradient that forms independent of Bcd. *Hb* mRNA is deposit ubiquitously in the embryo, but the posteriorly enriched Nanos (Nos) protein represses its translation in the posterior, generating a shallow AP gradient of Hunchback (Hb) protein (Irish et al., 1989). The Hb gradient provides limited AP polarity to the gap-gene network in the absence of Bcd (Struhl, 1989; Struhl et al., 1992; Tautz, 1988). The maternal Hunchback (Hb) protein gradient obscures the essential role of Bcd in axis specification. However, it is not essential for embryonic development in *Drosophila melanogaster* (Hülskamp et al., 1989; Struhl, 1989). Indeed, in other flies that lack maternal *hb* or where maternal *hb* plays a lesser role in activating anterior genes, knockdown of *bcd* results in a complete duplication of posterior segments and a symmetric double abdomen phenotype (Lemke et al., 2010; Stauber et al., 2000). In the posterior of the embryo, where the concentration of Bcd is low, gap genes also receive input from maternal Hb and/or Caudal (Cad) (Hülskamp et al., 1990; Rivera-Pomar et al., 1995; Schroeder

et al., 2004a). In Drosophila, the early Cad gradient is complementary to the Bcd gradient because Bcd represses the translation of the ubiquitous maternal *caudal* (*cad*) mRNA (Dubnau et al., 1997; Rivera-Pomar et al., 1996). This function however is not conserved in other flies (Stauber et al., 2008). Loss of AP polarity is not observed in embryos depleted of germ plasm. Instead, embryos depleted of the germ plasm determinant Oskar, or the germ plasm component responsible for the maternal Hb gradient, Nos, lack abdominal segments and do not form, or lose pole cells (Lehmann & Nüsslein-Volhard, 1986, 1991; Schüpbach & Wieschaus, 1986).

1.3.3 Overview of the zygotic segmentation cascade

The zygotic segmentation cascade is initiated by gap and pair-rule genes (Nüsslein-Volhard & Wieschaus, 1980) , all of which encode TFs (**Figure 1.3**). The first to be expressed are the gap genes (for review see (Jaeger, 2011)). Bcd is required for establishing anterior expression domains of gap genes, including *hb*, *giant (gt)*, *Krüppel (Kr)*, *knrips (kni)*, and the "head gap genes" *buttonhead (btd)*, *empty-spiracles (ems)*, *orthodenticle (otd)* (Dalton et al., 1989; Driever & Nüsslein-Volhard, 1988b, 1989; Finkelstein & Perrimon, 1990; Gao & Finkelstein, 1998; Hoch et al., 1991a; Ochoa-Espinosa et al., 2005; Rivera-Pomar et al., 1995; Schroeder et al., 2004a; Struhl et al., 1989; Tautz, 1988; Walldorf & Gehring, 1992). Additionally, it modulates the expression of the "terminal gap genes", *tailless (tll)* and *huckebein (hkb)*, which are expressed at both poles in response to Tor signaling (Liaw & Lengyel, 1993; Pignoni et al., 1992; Reuter & Leptin, 1994). Maternally supplied Hb provides complementary input to specify anterior and posterior gap domains (Hoch et al., 1991a; Porcher et al., 2010; Schroeder et al., 2004a; Simpson-Brose et al., 1994; Struhl et al., 1992; Wimmer et al., 2000). The maternal gradient of Cad is essential for the expression of *kni* and *gt* in their posterior domains, and the

terminal system is critical for the activation *tll* and *hkb* (Klingler et al., 1988; Paroush et al., 1997; Pignoni et al., 1992; Rivera-Pomar et al., 1995; Rivera-Pomar & Jäckle, 1996; C. Schulz & Tautz, 1995; Weigel et al., 1990). The initial gap gene domains are further refined by mutual repression between neighboring gap genes rusting in sharp expression boundaries (Jäckle et al., 1986; Surkova et al., 2008). Genetic depletion of gap genes results in broad deletions of sequential segments (Nüsslein-Volhard et al., 1984b, 1984a; Nüsslein-Volhard & Wieschaus, 1980; Wieschaus et al., 1984).

The gap genes converge with maternal inputs to activate the pair-rule genes, named after loss of function phenotypes described as a complete or partial loss of alternating segments (Nüsslein-Volhard et al., 1984b, 1984a; Nüsslein-volhard et al., 1980; Wieschaus et al., 1984). The pair-rule genes have been subdivided into primary (*hairy* (*h*)*, runt* (*rnt*), *even-skipped* (*eve*)*, fushi tarazu* (*ftz*)*, odd-skipped* (*odd*)) and secondary genes (*paired* (*prd*) and *sloppy-paired* (*slp*)). The distinction is based on the observation that *prd* and *slp* are expressed later than the other pair-rule genes (Clark, 2017; Schroeder et al., 2011). The pair-rule genes are initially expressed in 7 stripes before cellularization and their activation is dictated by different combinations of gap input (Arnosti et al., 1996; Goto et al., 1989; Harding et al., 1989; Howard et al., 1988; Pankratz et al., 1990; Pankratz & Jäckle, 1990; Schroeder et al., 2011; Small et al., 1991, 1992; Stanojevic et al., 1991). The pair-rule periodicity is further refined by cross-regulatory interactions between pair-rule genes (Harding et al., 1986; Ingham & Gergen, 1988; Pankratz & Jäckle, 1990). At gastrulation the interactions between pair-rule genes become more complex causing the initial seven pair-rule stripes to double to 14 stripes corresponding to one stripe per segment (Clark, 2017; Schroeder et al., 2011). The exception to the traditional pair-rule periodicity is the non-canonical *odd-paired* (*opa*) which is expressed broadly in the trunk of the embryo (Benedyk et

al., 1994). Opa plays a key role in the late pair-rule network as it is required for the regulatory changes necessary for the transition from 7 stripes to 14 stripes. For example, Odd-skipped (Odd) repression of *prd* is Opa dependent, in *opa* mutant embryos the 7 stripes of *prd* fail to split into 14 stripes. (Clark & Akam, 2016). Opa is also required to activate the segment-polarity genes at the correct time (Benedyk et al., 1994).

The key outcomes of the pair-rule patterning network is the activation of engrailed (*en*) and *wingless* (*wg*) and the eventual establishment of segmental *wg* and *hedgehog* (*hh*) signaling centers within each future body segment (DiNardo et al., 1985; González et al., 1991; Ingham et al., 1985; Kornberg et al., 1985; Nüsslein-Volhard & Wieschaus, 1980; van den Heuvel et al., 1989). It is the reiterated pattern of these signaling centers that provides polarity to each body segment and constitutes the core of the segmented body plan. This segmental scaffold further differentiates under the influence of the homeotic (Hox) genes whose expression dictates the fate of each segment in the adult fly (Hubert & Wellik, 2023; Lewis, 1978).

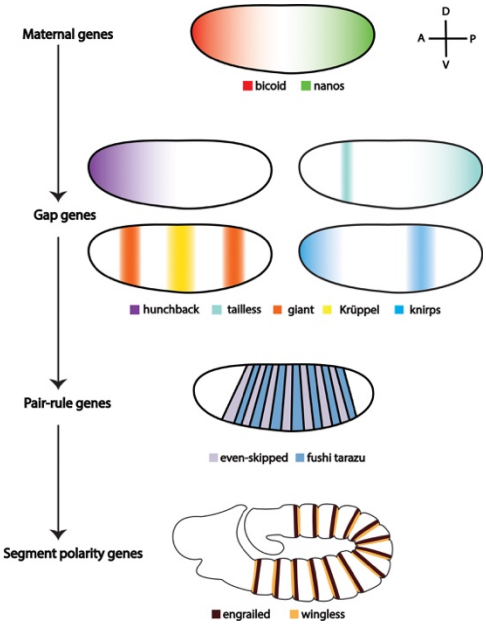

Figure 1. 3 The segmentation cascade of *Drosophila melanogaster*.
The segmentation cascade functions hierarchically. Maternal gradients initiate the cascade and activate the gaps genes. Gap genes encode TFs that cross regulate one another and with the maternal inputs activate the pair-rule genes. Pair-rule and gap genes activate the segment-

polarity genes. Segment-polarity genes solidify parasegmental organization. Schematic of the initial patterning of select gap genes (*hb*, *tll*, *gt*, *Kr*, *kni*) and the 7 pair-rule stripes (*eve* and *ftz*) are shown during the blastoderm stage. Parasegmental stripes of segment-polarity genes (*w*g and *e*n) are shown during germband extension. Figure from (Gheisari et al., 2020).

1.4 Symmetry breaking at the level of chromatin accessibility by the anterior determinant gene *bcd*

1.4.1 Segmentation gene enhancers display regional differences in chromatin accessibility

Studies in *Drosophila melanogaster*, performing ATAC-seq in single embryos, bisected embryos, regional populations of embryonic nuclei, and throughout embryogenesis at single-cells resolution, have revealed dynamic spatiotemporal transitions in chromatin accessibility that shape early patterning and lineage specification events (Bozek et al., 2019; Calderon et al., 2022; Cusanovich et al., 2018; Haines & Eisen, 2018). During the last 3 nuclear cycles before cellular blastoderm formation, chromatin accessibility is established incrementally (Blythe & Wieschaus, 2016) and by NC14, the chromatin state is generally homogenous in the degree of accessibility regionally in the embryo, particularly within gene bodies and promoters bound by pioneer factor Zld (Bozek et al., 2019; Haines & Eisen, 2018). As development progresses the chromatin accessibility state diversifies substantially, reflecting lineage commitment and differentiation (Calderon et al., 2022; Cardamone et al., 2025; Cusanovich et al., 2018).

Although the most dramatic changes in chromatin accessibility appear later in embryogenesis, regional heterogeneity in the degree of accessibility emerge early among enhancers of the AP and dorso-ventral (DV) patterning networks (Bozek et al., 2019; Haines &

Eisen, 2018; Hannon et al., 2017). AP enhancers are critical for regulating the spatiotemporal expression patterns of segmentation genes, including the early acting gap and pair-rule genes (El-Sherif & Levine, 2016; Fujioka et al., 1999; Perry et al., 2011; Small et al., 1996; Struhl et al., 1989). The chromatin state of gap and pair-rule enhancers neatly resolve along the AP axis; enhancers are accessible in regions where they are active and driving domains of segmentation gene expression and closed in regions where they are inactive (**Figure 1.4)** (Bozek et al., 2019). These observations suggest that the expression dynamics of segmentation genes is achieved through regulating enhancer activity by means of priming its accessibility (Bozek & Gompel, 2020; Eck et al., 2020).

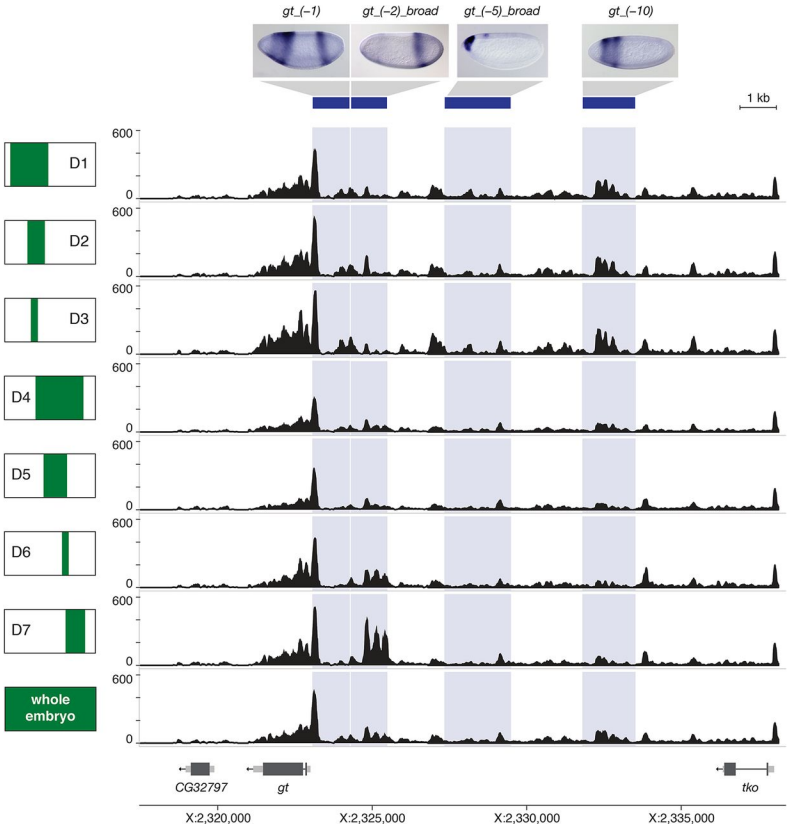

Figure 1. 4 Regional accessibility of the enhancers that drive *gt* expression.
Regional patterns of accessibility of *gt* enhancers correspond to *gt* expression domains. *Gt* enhancers are accessible in regions where they are actively driving *gt* expression. ATAC-seq accessibility data from purified nuclei along the AP axis (D1-D7) or from the whole embryo. Top: Images demonstrating enhancer activity in blastoderm embryos assessed by *in situ*

hybridization of a reporter gene (original study Schroeder et al., 2004). Genomic coordinates of known *gt* enhancers indicate by blue bar. Figure from (Bozek et al., 2019).

1.4.2 Bcd drives enhancer accessibility via concentration-dependent nucleosome competition

Activity of ubiquitous pioneer factors, particularly activity of Zld, lays the foundation of chromatin accessibility in the embryo (Colonnetta et al., 2021; J. Duan et al., 2021; Foo et al., 2014; Gaskill et al., 2021; Moshe & Kaplan, 2017; Sun et al., 2015). Once established some of these accessible regions are bound by Bcd and indeed it has been demonstrated that Zld facilities Bcd binding (Bishop et al., 2023; Fernandes et al., 2022; Harden et al., 2023; Harrison et al., 2011; Mir et al., 2017, 2018; Nien et al., 2011; Xu et al., 2014). However, a small cohort of early segmentation gene enhancers are largely insensitive to Zld and exhibit anteriorly restricted accessibility in response to high concentrations of Bcd. These enhancers drive anterior expression of gap and pair-rule genes (e.g. enhancer of head gap genes, anterior enhancers of *gt*, *kni*, and *eve*) (Bozek et al., 2019; Hannon et al., 2017; Liu et al., 2018). Bcd engages in a concentration-dependent competition with nucleosomes to open these otherwise inaccessible regulatory regions and spatially regulate their activity (Degen et al., 2024; Hannon et al., 2017). These observations suggest that Bcd breaks symmetry along the AP axis at the level of chromatin accessibility well before the bulk of zygotic genome activation and pattern formation occurs (Blythe & Wieschaus, 2016). Moreover, Bcd appears to have "key" targets who depend on Bcd to establish the accessibility of their enhancers and whose expression is lost entirely in Bcd deficient embryos, thereby causing the suppression of head and thorax formation (Datta et al., 2018; Hannon et al., 2017; Liu et al., 2018).

1.5 Evolution of anterior determinant genes in flies (Diptera)

1.5.1 Evolution of *bcd* in the Cyclorrhaphan lineage

The *bcd* gene evolved in a lineage of higher flies, called Cyclorrhapha, and is absent in non-cyclorrhaphan species (Stauber et al., 2000). Bcd arose from a duplication of the Hox3 gene *zerknüllt (zen),* which has a role in the formation of extra-embryonic tissues in flies and other insects (Rafiqi et al., 2008; Stauber et al., 2002; Van Der Zee et al., 2005). How *bcd* acquired its function in axis specification remains unknown (Klomp et al., 2015; Liu et al., 2018). Within cyclorrhaphan flies, Bcd is functionally conserved in the scuttle fly *Megaselia abdita* (Stauber et al., 2000), one of the most distant relatives of Drosophila within this group and where the gap gene network is highly conserved (Wotton, Jimenez-Guri, et al., 2015). In Megaselia, knockdown of *Mab-bcd* results in the loss of all anterior expression domains, including the central expression domain of the gap gene *Mab-Kr*, and mirror-image double abdomens (Stauber et al., 2000; Wotton, Jiménez-Guri, et al., 2015). These findings suggest that maternal *hb* expression plays a lesser role in the activation of gap genes in Megaselia than in Drosophila in which Kr is activated in the absence of Bcd by the maternal Hb gradient. Bcd is also functionally conserved in the hover fly *Episyrphus balteatus* (Lemke et al., 2010), and in blow flies (*Lucilia sericata* and *Calliphora vicina)* but Bcd protein from these species is insufficient to rescue a bcd null phenotype in Drosophila, suggesting significant functional changes to the protein throughout its evolution (Gregor et al., 2008).

1.5.2 Diversity of anterior determinant genes in non-cyclorrhaphan flies

Non-cyclorrhaphan flies have evolved an array of distinct AD genes (Klomp et al., 2015; Yoon et al., 2019). Their phylogenetic distribution suggests that the ancestral dipteran anterior determinant may have been encoded by *pangolin* (**Figure 1.6**) (Yoon et al., 2019). *Pangolin* also known as *T-cell factor* (*Tcf*), encodes a TF that is targeted by the Wnt signaling pathway (Brunner et al., 1997; Van de Wetering et al., 1997). However, the ADs of many other lower dipterans are unrelated to *Tcf* and *Bcd* and instead encode different zinc finger genes; *cucoid* in culicine mosquitoes, *panish* in certain harlequin flies (Chironomini), and *opa* (ortholog of *Zic*) in moth flies (Psychodidae) (Klomp et al., 2015; Yoon et al., 2019). The AD-encoded proteins (also abbreviated as AD) vary widely in their structure and DNA-binding domains. Bcd binds DNA via its homeodomain, Pangolin (Pan) through HMG and C-clamp domains, Panish through a C-clamp domain, and Opa and Cucoid through distinct C2H2 zinc finger domains (B. Duan et al., 2021; Hursh & Stultz, 2018; Li et al., 2023; Liu et al., 2018; Nitta et al., 2015; Ravindranath & Cadigan, 2014). Nonetheless, lower dipterans maternally deposit the mRNA of the AD at the anterior of the egg and depleting the embryo of the AD leads to a duplication of the abdomen at the expense of anterior fates (double abdomen phenotype).

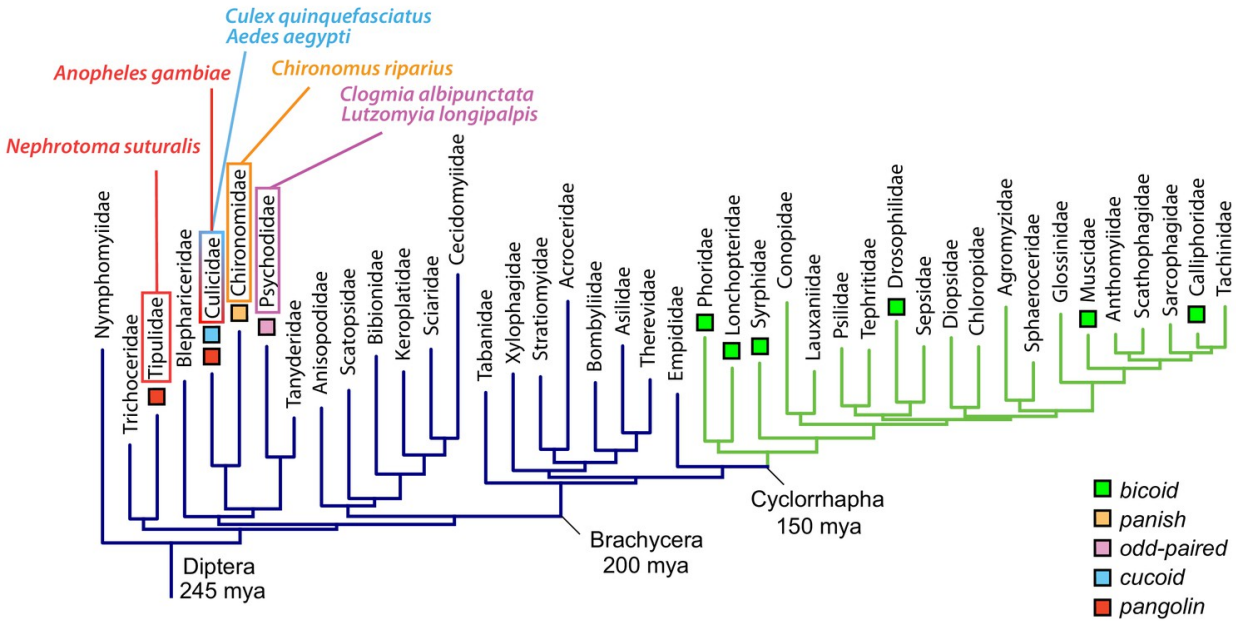

Figure 1. 5 Dipteran AD genes.
Phylogenetic tree of Diptera constructed from data in (Wiegmann et al., 2011), including the ADs described in (Yoon et al., 2019). Cyclorrhapha lineages are in green. Non-cyclorrhaphan lineages are in blue. Million years ago, (mya). Figure from (Yoon et al., 2019).

1.6 Axis specification in the moth fly *Clogmia albipunctata*

1.6.1 The maternal transcript isoform of the segmentation gene *opa* is the anterior determinant in Clogmia

In *Clogmia albipunctata*, a maternally deposited alternative transcript isoform of *opa* (*Cal-opa*) is localized at the anterior tip of the egg (**Figure 1.6 A**). Injection of double stranded RNA (dsRNA) targeting specifically this maternally expressed *Cal-opa* transcript in preblastoderm embryos causes ectopic expression of the posterior gene *caudal* (*Cal-cad*) and a mirror-image double abdomen (**Figure 1.6 B**). Conversely, injection of *Cal-opa* mRNA at the posterior pole of preblastoderm embryos is sufficient to ectopically induce expression of the head marker, *orthodenticle* (*Cal-otd*) and formation of a second head.

In *Cal-opa* RNAi embryos, a second set of *Cal-nos* expressing cells are observed in the duplicated posterior region, suggesting that Cal-Opa may act directly or indirectly to repress the induction of germ cells in the anterior of the embryo. Indeed, Clogmia embryos seem to lack maternal germ plasm at the posterior pole and Clogmia homologs of *nos* (*Cal-nos1*, *Cal-nos2*, *Cal-nos3*, and *Cal-nos4*) are not expressed until the cellular blastoderm stage (Yoon et al., 2019). This observation indicates that unlike in Drosophila, maternal germ plasm components do not play a role in axis specification in Clogmia.

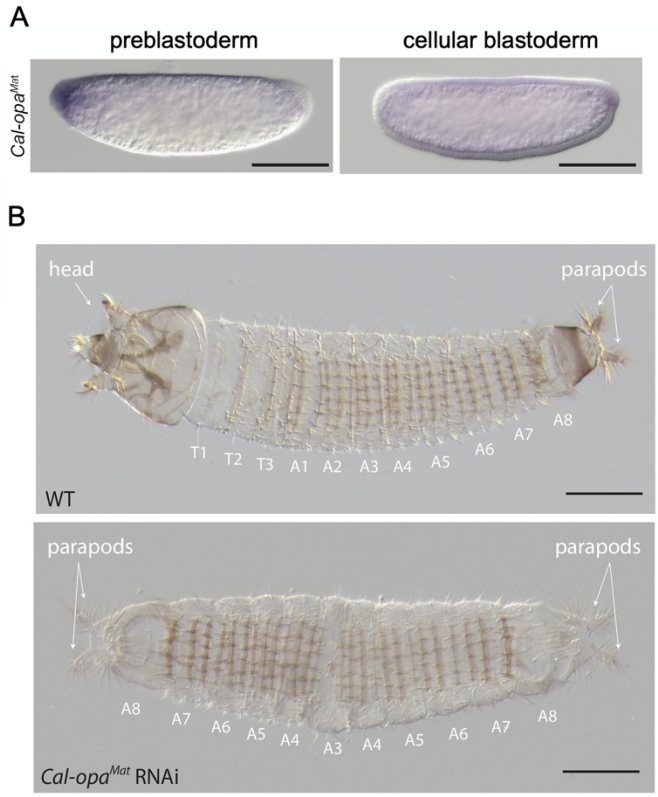

Figure 1. 6 Expression and function of the maternal transcript isoform of *Cal-opa*.
A) RNA *in situ* hybridization of the maternal transcript of *Cal-opa* (*Cal-opa*^{Mat}) in preblastoderm (1 hr-old) and cellular blastoderm (7 hr-old) embryos. B) First instar larval wild-type (WT) cuticle and cuticle phenotype following knockdown of maternal *Cal-opa* (*Cal-opa*^{Mat} RNAi). Anterior, left. Dorsal, up. Scale bar: 100µm. T, thoracic segment. A, abdominal segment. Figure modified from figures in (Yoon et al., 2019).

1.6.2 Functionally distinct roles of maternal and zygotic *Cal-opa* transcript isoforms

In Drosophila, *opa* is a late acting pair-rule gene expressed uniformly in the trunk of the embryo (Benedyk et al., 1994; Jürgens et al., 1984). Opa is required for the transition of pair-rule gene expression from double to single segment periodicity (i.e. 7 stripes to 14) and for timely expression of segment-polarity genes (Baumgartner & Noll, 1990; Benedyk et al., 1994; Clark & Akam, 2016; Klinger & Gergen, 1993; Swantek & Gergen, 2004).

This late segmentation function appears to be conserved in Clogmia but involves a strictly zygotic transcript isoform (*Cal-opa*^{Zyg}) (Yoon et al., 2019). *Cal-opa*^{Zyg} is expressed along the trunk of the embryo starting at cellular blastoderm stage when the maternal *Cal-opa* transcript is no longer present (**Figure 1.7 A**). RNAi knockdown of *Cal-opa*^{Zyg} results in half the number of *Cal-slp* positive segments and other defects in segmentation and head development, similar to the mutant phenotype of *opa* in Drosophila (**Figure 1.7 B**) (Jürgens et al., 1984; Yoon et al., 2019). Injection of dsRNA targeting both the maternal and zygotic *Cal-opa* transcripts resulted in double abdomens with missing segments (Yoon, 2019). Therefore, *Cal-opa* and *Cal-opa*^{Zyg} play distinct roles in segmentation; the maternal Cal-opa transcript functions as the AD and *Cal-opa*^{Zyg} as a component of the late pair-rule network.

Protein alignments indicate that *Cal-opa*^{Zyg} encodes the full-length Opa protein, while the protein encoded by the maternal transcript lacks 20 N-terminal amino acids. However, both isoforms can induce a second head when mRNA is provided before blastoderm formation at the posterior pole. Moreover, mRNA of *opa* homologs from Drosophila and other dipterans (Lutzomyia, Chironomus) also induce a second head when delivered to the posterior pole of 1-hour old Clogmia embryos (Yoon et al., 2019). This suggests that Cal-Opa evolved its role as the

AD without essential changes in its amino acid sequence, via a change in expression and co-option of downstream genes.

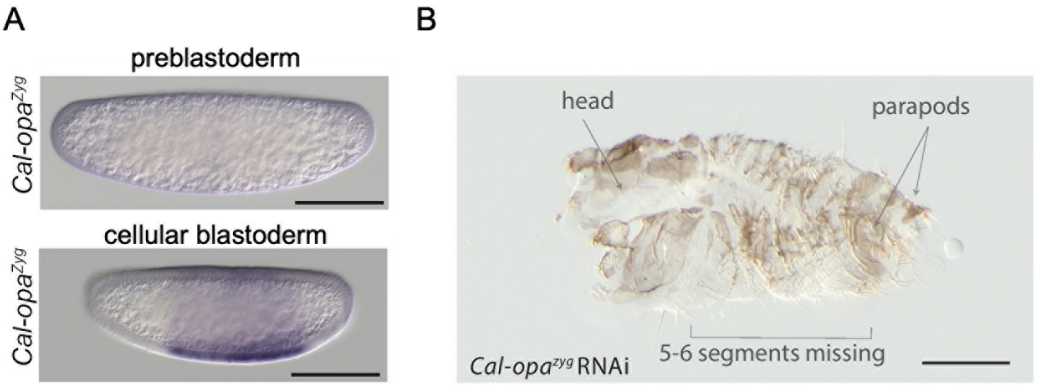

Figure 1. 7 Expression and function of the zygotic transcript isoform of *Cal-opa.*
A) RNA *in situ* hybridization of the zygotic transcript of *Cal-opa* (*Cal-opa*^{Zyg}) in preblastoderm (1 hr-old) and cellular blastoderm (7 hr-old) embryos. B) RNA *in situ* hybridization of *Cal-slp* in a wild-type (WT) embryo during germband extension and a stage matched embryo where the zygotic transcript of *Cal-opa* has been knocked down (*Cal-opa*^{Zyg} RNAi). C) First instar larval cuticle phenotype following *Cal-opa*^{Zyg} RNAi. Anterior, left. Dorsal, up. Scale bar: 100µm. Md, mandibular segment. Mx, maxillary segment. Lb, labial segment. T, thoracic segment. A, abdominal segment. Figure modified from figures in (Yoon et al., 2019).

1.6.3 Pioneer activity of Opa/Zic in *Drosophila melanogaster*

Opa is a member of the Zic protein family that function as transcriptional regulators of development in vertebrates (Aruga & Millen, 2018; Houtmeyers et al., 2013; Hursh & Stultz, 2018). In Drosophila, Opa drives the accessibility of enhancers that advance the late segmentation network (pair-rule and segment-polarity) at NC14 (Koromila et al., 2020; Soluri et al., 2020). Opa functions both independently and cooperatively with Zld to establish enhancer accessibility. The absolute ATAC-seq signal (corresponding to the degree of accessibility) of enhancers regulated by both factors is greater than the accessibility of enhancers regulated by

one factor alone (Koromila et al., 2020). Beyond the segmentation cascade, Opa pioneers chromatin accessibility within the DV patterning network (Koromila et al., 2020). The vertebrate homolog, Zic3, has been shown to bind nucleosome DNA *in vivo* suggesting that Opa opens regions directly by binding nucleosome occupied DNA and stimulating nucleosome removal (Grand et al., 2024).

1.7 Research goal

1.7.1 Using the diversity of anterior determinant genes to understand how TF regulatory networks evolve in the context of early embryogenesis

The evolutionary instability of dipteran ADs presents an opportunity to investigate how developmental gene networks change and diversify over time. Historically, a question of this magnitude is difficult to address because one must ascertain the detailed knowledge of a developmental network as well as develop amenable and closely related species to compare said network. The segmentation network of dipteran insects fulfils these requirements. The simple anatomy of early fly embryos, the function of ADs during the syncytial phase of development when direct genetic perturbations are possible in a broad range of fly species, and detailed knowledge of the segmentation cascade in *Drosophila melanogaster* provide excellent conditions for comparative functional studies.

Although it is clear that dipteran ADs share the essential function of polarizing the embryo and specifying anterior fates, several aspects of axis specification and segmentation in these species remain unresolved. We do not know the mechanism of action of ADs other than Bcd, specifically if those ADs break symmetry early in the embryo by priming chromatin accessibility. Moreover, outside of Bcd the key zygotic targets of ADs are unknown. The

molecular diversity of ADs could correspondingly lead to major rewiring of the segmentation cascade. This would imply that the early zygotic gene network driving segmentation is itself unstable even though the segmented germband is morphologically very similar in these insects. Alternatively, the downstream network could impose constrains on the ADs resulting in a conserved network. For example, the complex cross-regulation among gap and pair-rule genes could impede the integration of new genes into the segmentation cascade.

This study explores the mechanism of symmetry breaking at the level of chromatin accessibility and gene expression initiated by unrelated ADs. Through this investigation early zygotic targets of the AD in *Clogmia albipunctata* are identified, providing the first evidence of how the dipteran segmentation network responds to dramatic changes in AD input.

1.7.2 The model organism *Clogmia albipunctata*

Clogmia albipunctata is a suitable model for combining genomic approaches and functional studies in embryos. Clogmia can be reared at a wide range of temperatures and does not require special lighting, or tight humidity control. At 25°C, the life cycle from egg to egg is approximately 27 days and embryogenesis lasts approximately 2.5 days. Like Drosophila, Clogmia embryos begin development as a syncytium. Clogmia forms a syncytial blastoderm at NC9 and cellularizes the blastoderm at NC14 shortly before entering gastrulation. Clogmia's nuclear cycles last longer than Drosophila: Pre-syncytial blastoderm cycles (i.e. NC1-9) ~3 times longer, NC10 ~2 times longer, NC11-13 ~3 times longer, and NC14 ~1.6 times longer (Jiménez-Guri et al., 2014). The blastoderm embryo is tractable for genetic manipulation including RNA interference (RNAi) and morphogenesis can be monitored by live imaging because the eggshell is transparent (Jiménez-Guri et al., 2014; Yoon et al., 2019). Hundreds of

fertilized eggs can be obtained at any time by dissecting female ovaries and activating the eggs by osmotic shock in water (Sander, 1985). This procedure enables the collection of precisely staged and synchronously developing embryos.

CONSERVATION OF SYMMETRY BREAKING AT THE LEVEL OF CHROMATIN ACCESSIBILITY BETWEEN FLY SPECIES WITH UNRELATED ANTERIOR DETERMINANTS

This chapter is a full reprint of a study by Ezra E. Amiri, Ayse Tenger-Trolander, Muzi Li, Alexander Thomas Julian, Koray Kasan, Sheri A. Sanders, Shelby Blythe, Urs Schmidt-Ott, in which I am the primary author. The work is included with permission from all authors. See Section 2.6 for author contributions. To be noted, the manuscript presented here will be formally submitted to a peer-reviewed journal and may undergo revision.

2.1 Abstract

Establishing the anterior-posterior body axis is a fundamental process during embryogenesis, and the fruit fly, *Drosophila melanogaster*, provides one of the best-known case studies of this process. In Drosophila, localized mRNA of *bicoid* serves as anterior determinant (AD). Bicoid engages in a concentration-dependent competition with nucleosomes and initiates symmetry-breaking along the AP axis by promoting chromatin accessibility at the loci of transcription factor (TF) genes that are expressed in the anterior of the embryo. However, ADs of other fly species are unrelated and structurally distinct, and little is known about how they function. We addressed this question in the moth fly, *Clogmia albipunctata*, in which a maternally expressed transcript isoform of the pair-rule segmentation gene *odd-paired* is localized in the anterior egg and has been co-opted as AD. We provide a *de novo* assembly and annotation of the Clogmia genome and describe how knockdown of *zelda* and maternal *odd-paired* transcript affect chromatin accessibility and expression of TF-encoding loci. The results of these experiments suggest direct roles of Cal-Zld in opening and closing chromatin during

nuclear cleavage cycles and show that Clogmia's maternal *odd-paired* activity promotes chromatin accessibility and anterior expression during the early phase of zygotic genome activation at Clogmia's *homeobrain* and *sloppy-paired* loci. We conclude that unrelated dipteran ADs initiate anterior-posterior axis-specification at the level of enhancer accessibility and that *homeobrain* and *sloppy-paired* homologs may serve a more widely conserved role in the initiation of anterior pattern formation given their early anterior expression and function in head development in other insects.

2.2 Introduction

Early embryos face the challenge of breaking symmetry and establishing spatially restricted domains of gene expression that define the body plan. Dipteran insects (flies, midges, mosquitoes) can achieve this through concentration gradients of transcription factors, referred to as morphogens, because their early embryos develop as a syncytium in which protein products from maternally localized mRNAs can diffuse (Driever & Nüsslein-Volhard, 1988a, 1988b). Dipteran embryos undergo multiple rounds of synchronous nuclear division cycles (**NC**) before the blastoderm, a monolayer of nuclei between the cell membrane of the egg and the yolk, cellularizes and enters gastrulation to form the segmented germband. Because of its anatomical simplicity and experimental accessibility, the syncytial and cellular blastoderm of Drosophila has been widely used as a model for axis specification, pattern formation, and gene regulation and is among the best-known developmental transcription factor (**TF**) / gene networks in animals (Clark et al., 2019; Crombach & Jaeger, 2021; Jaeger, 2011; Ma et al., 2016; St Johnston & Nüsslein-Volhard, 1992; Surkova et al., 2018; Tkačik & Gregor, 2021; Wieschaus, 2016). Many aspects of embryogenesis and pattern formation are conserved between distantly related

dipterans and provide a unifying framework for comparative studies (Anderson, 1966; Crombach & Jaeger, 2021; Lemke et al., 2020; Schmidt-Ott & Lynch, 2016). However, the maternally provided anterior determinants that establish anterior-specific zygotic gene expression and the embryo's overall head-to-tail polarity differ widely between species across the dipteran clade (Klomp et al., 2015; Yoon et al., 2019). This finding raises the question of which other aspects of Drosophila's classic model of axial pattern formation (Schmidt-Ott & Yoon, 2022) have diverged within the dipteran clade.

While the anterior determinant (**AD**) in Drosophila and many other higher flies (Cyclorrhapha) is encoded by a homeobox gene, *bicoid* (Berleth et al., 1988; Lemke et al., 2010; Shaw et al., 2001; Stauber et al., 2000; Wotton, Jiménez-Guri, et al., 2015), diverse zinc finger genes take on this role in lower dipterans, including *pangolin* (*Tcf* ortholog) in anopheline mosquitoes and crane flies, *cucoid* in culicine mosquitoes, *panish* in certain harlequin flies (Chironomini), and *odd-paired* (*Zic* ortholog) in moth flies (Psychodidae) (Klomp et al., 2015; Yoon et al., 2019). All these species deposit a maternally expressed transcript isoform of the respective AD gene in the anterior tip of the egg that, once translated, specifies anterior cell fates. Without the AD, embryos develop a bicaudal phenotype without head (double abdomen). The AD-encoded proteins (for simplicity we also refer to them as AD) differ widely in their structure and DNA-binding domains. Bicoid (Bcd) binds DNA via its homeodomain, Pangolin (Pan) through HMG and C-clamp domains, Panish through a C-clamp domain, and Opa and Cucoid through distinct C2H2 zinc finger domains (B. Duan et al., 2021; Hursh & Stultz, 2018; M. Li et al., 2023; Liu et al., 2018; Nitta et al., 2015; Ravindranath & Cadigan, 2014). This molecular diversity prompted us to ask whether targets and mechanisms of action of dipteran

ADs differ between species or whether they are conserved due to downstream network constraints.

Bcd binds to enhancers that promote the transcription of early segmentation genes, such as *hunchback* (Driever & Nüsslein-Volhard, 1989; Struhl et al., 1989), *Krüppel* (Hoch et al., 1991b), *knirps* (Rivera-Pomar et al., 1995; Rivera-Pomar & Jäckle, 1996), *giant* (Berman et al., 2002; Ochoa-Espinosa et al., 2005; Schroeder et al., 2004b), *orthodenticle* (Gao & Finkelstein, 1998), or *even-skipped* (Arnosti et al., 1996; Small et al., 1991, 1992; Stanojevic et al., 1991). This Bcd activity is complemented by the maternal TF gradients of Hunchback (Hb) (Manu et al., 2009; Schroeder et al., 2004b; Simpson-Brose et al., 1994; Wimmer et al., 2000), which peaks in the anterior embryo and Caudal (Cad) (Rivera-Pomar et al., 1995; Rivera-Pomar & Jäckle, 1996), which is repressed in the anterior embryo (Macdonald, 2005). Additionally, ubiquitous TFs, such as the pioneer factor Zelda (Zld) (Liang et al., 2008; Staudt et al., 2006), establish or promote chromatin accessibility at many sites that are targeted by Bcd (Bishop et al., 2023a; Crocker & Stern, 2017; Eck et al., 2020; Fernandes et al., 2022; Foo et al., 2014; Harden et al., 2023; Ling et al., 2019; Mir et al., 2018; Xu et al., 2014). During early blastoderm stages, chromatin accessibility measured by ATAC-seq (assay for transposase-accessible chromatin with sequencing) is largely homogeneous across the embryo; however, a small cohort of early segmentation gene enhancers exhibits anteriorly restricted accessibility in response to high concentrations of Bcd (e.g., enhancers of "head gap genes", *knirps*, and *giant*, and *even-skipped* that drive expression in anterior domains) (Calderon et al., 2022; Cusanovich et al., 2018; Haines & Eisen, 2018; Hannon et al., 2017; Soluri et al., 2020). Bcd engages in a concentration-dependent competition with nucleosomes to 'open' these otherwise inaccessible regulatory regions, effectively breaking symmetry along the AP axis at the level of chromatin accessibility

(Degen et al., 2024), well before the bulk of zygotic genome activation and pattern formation occurs (Blythe & Wieschaus, 2016). While Bcd exerts a concentration-dependent function in gene regulation through its impact on chromatin states (Hannon et al., 2017), it is not known whether this mechanism of action can be generalized for other dipteran ADs.

We have addressed this question by examining how the AD of *Clogmia albipunctata* affects chromatin accessibility and gene expression during the early phase of zygotic genome activation. Clogmia lacks *bicoid* and uses a maternal transcript isoform of *odd-paired* (*Cal-opa*) as anterior determinant (Yoon et al., 2019). Knockdown of the maternal *Cal-opa* transcript by isoform-specific RNA interference (RNAi) results in the double abdomen phenotype, and posterior mis-expression of *Cal-opa* by mRNA injection causes a bicephalic (double head) phenotype. An alternative zygotic transcript isoform of *Cal-opa* appears during blastoderm cellularization at NC14 and exerts a conserved function in the late segmentation gene network but has no role in establishing head-to-tail polarity of the embryo (Clark & Akam, 2016; Yoon et al., 2019). However, prior to activation of zygotic *Cal-opa* expression, ectopic posterior *odd-paired* activity, whether from the maternal or the zygotic isoform or from Drosophila *odd-paired*, results in double heads in Clogmia (Yoon et al., 2019). Early Cal-Opa activity is therefore both necessary and sufficient for establishing head-to-tail polarity of the Clogmia embryo. The symmetric double abdomen phenotype of Clogmia embryos with reduced maternal *Cal-opa* transcript is very similar to the phenotype of Drosophila embryos lacking the maternal transcripts of *bicoid* and *hunchback* (Hülskamp et al., 1990). We note here that in Drosophila, maternal *hunchback* provides residual polarity to *bicoid*-deficient embryos (Tautz, 1988). This is not the case in Clogmia, in which a gradient of Hb protein appears at NC12, which is later than in Drosophila but still before transcription initiation of several other “gap genes”, including

homologs of *Krüppel*, *giant*, *knirps*, and *tailless* (García-Solache et al., 2010; Janssens et al., 2014).

In Drosophila, zygotic Opa pioneers chromatin accessibility in late NC14 embryos to advance the progression of segmentation and dorso-ventral patterning (Koromila et al., 2020; Soluri et al., 2020). We therefore asked whether the maternal *Cal-opa* gradient of Clogmia breaks axial symmetry and initiates anterior pattern formation during zygotic genome activation by driving chromatin accessibility at the loci of select target genes. Like Drosophila, Clogmia forms a syncytial blastoderm at NC9 and cellularizes the blastoderm at NC14 shortly before entering gastrulation, although Clogmia's nuclear cycles last longer: NC1-9 ~3 times longer, NC10 ~2 times longer, NC11-13 ~3 times longer, and NC14 ~1.6 times longer (Jiménez-Guri et al., 2014). Here, we have used a *de novo* assembled and annotated genome sequence of *Clogmia albipunctata* to characterize chromatin accessibility in single wild-type embryos at NC11, NC12, NC13, and early and late NC14, using ATAC-seq. We then established a protocol for studying chromatin accessibility and gene expression changes in single blastoderm embryos that were subjected to RNA interference (RNAi), using Clogmia's ortholog of the pioneer factor gene *zelda* (*Cal-zld*) as target and confirmed that it has a pronounced impact on chromatin accessibility and gene expression. Finally, we examined how knockdown of maternal *odd-paired* transcript (*Cal-opa*) affects chromatin accessibility and gene expression during early stages of zygotic genome activation. Our results reveal groups of genes with similar dynamic changes in chromatin accessibility, point to direct roles of Cal-Zld in opening and closing chromatin, and support the hypothesis that the conserved feature of anterior determination between Clogmia and Drosophila is symmetry breaking at the level of chromatin accessibility, while key target genes may differ between these species.

2.3 Results

2.3.1 De novo assembly and annotation of the genome of *Clogmia albipunctata* reveals multiple lineage-specific duplications of early segmentation gene homologs

We generated a high quality *de novo* genome assembly of *Clogmia albipunctata* with 6 chromosome-size scaffolds (**Figure 2.1, Table 2.1, Appendix S2.1**), consistent with the known chromosome number in this species (Amabis, 1977; Amabis & Simoes, 1972). The total length of the chromosome-size scaffolds is 304.3 Mb, close to genome size estimates based on flow cytometry data (316.6 Mb) (Schmidt-Ott et al., 2009). An additional 268 small scaffolds (< 1 MB) contained mostly repetitive sequences. We annotated this new reference genome using RNA-seq transcripts from embryonic, larval, pupal, and adult stages (**Table S2.1**), protein sequences, and *ab initio* gene prediction software, and identified 15,047 protein coding genes. The annotated genome recovered 88%-94% of Complete Benchmarking Universal Single Copy Orthologs (BUSCO) (**Table 2.2**), in line with recently published dipteran genomes (Generalovic et al., 2021; Ji et al., 2024; Tenger-Trolander et al., 2025). While synteny between the chromosomes within the Cyclorrhapha clade can be highly conserved (time of divergence ~145 million years ago) (Tenger-Trolander et al., 2025), synteny between the chromosomes of *D. melanogaster* and *C. albipunctata* (divergence time ~250 million years ago) was essentially lost.

Next, we interrogated the genome for the presence of early developmental genes including predicted segmentation genes. As expected, no *bicoid* ortholog was found but we recovered two putative Hox class 3 genes to which *bicoid* belongs (Stauber et al., 1999), including the previously reported homolog of *zerknüllt* (*Cal-zen*) (Stauber et al., 2002) on scaffold 7 and a more diverged potential *Cal-zen* paralog in position 3 of the Hox gene complex

on scaffold 3 (evm.TU.scaffold_3.2943.1.65be5b92). Our annotation also included homologs of *Drosophila melanogaster*'s early segmentation genes, including gap genes (*tailless*, *huckebein*, *orthodenticle/ocelliless*, *empty spiracles*, *buttonhead*, *hunchback*, *Krüppel*, *giant, knirps*), pair-rule genes (*even-skipped*, *fushi tarazu*, *odd-skipped*, *runt*, *sloppy-paired*, *paired, hairy, odd-paired*), and other important regulators of the segmentation network (*caudal*, *Dichaete*) (Clark et al., 2019; Jaeger, 2011). Some of these genes have more than one copy in the Clogmia genome, including 2 copies of *hunchback* (*Cal-hb1, Cal-hb2*), *knirps/knirps-like* (*Cal-kni/knrl1, Cal-kni/knrl2*), *orthodenticle* (*Cal-otd1, Cal-otd2*), *even-skipped* (*Cal-eve1, Cal-eve2*), *tailless* (*Cal-tll1, Cal-tll2*) and *caudal* (*Cal-cad 1, Cal-cad2*), 3 copies of *sloppy-paired1/2/fd19B* (*Cal-slp1, Cal-slp2, Cal-slp3*), and 4 copies of *empty spiracles* (*Cal-ems1, Cal-ems2, Cal-ems3, Cal-ems4*) (**Figure 2.1, Table S2.2**). Such gene duplications are a potential source of redundancies in Clogmia's segmentation gene network (see below).

Table 2. 1 *Clogmia albipunctata* reference assembly statistics and quality metrics. Completeness and continuity of the assembly are assessed by Benchmarking Universal Single-Copy Orthologs (BUSCO) and scaffold N50 and L50, respectively. Eukaryotic BUSCOs recovered: the percentage of complete universal orthologs from the eukaryotic BUSCO dataset identified in the assembly. Scaffold N50: The length (in bp) of the shortest scaffold such that all scaffolds of equal or greater length account for at least 50% of the total assembly length (Lander et al., 2001). L50: The minimum number of scaffolds whose collective length accounts for 50% of the total assembly length (Gurevich et al., 2013).

Reference Assembly Statistics	
Total length (bp)	320,011,659
Number of scaffolds	274
Scaffold N50 (bp)	51,218,107
Scaffold L50	4
# of Ns per 100kb	0.19
GC (%)	32.34%
Heterozygosity (%)	0.018 - 0.038%
Masked (%)	48.55%
Eukaryotic BUSCOs recovered (%)	98.04%

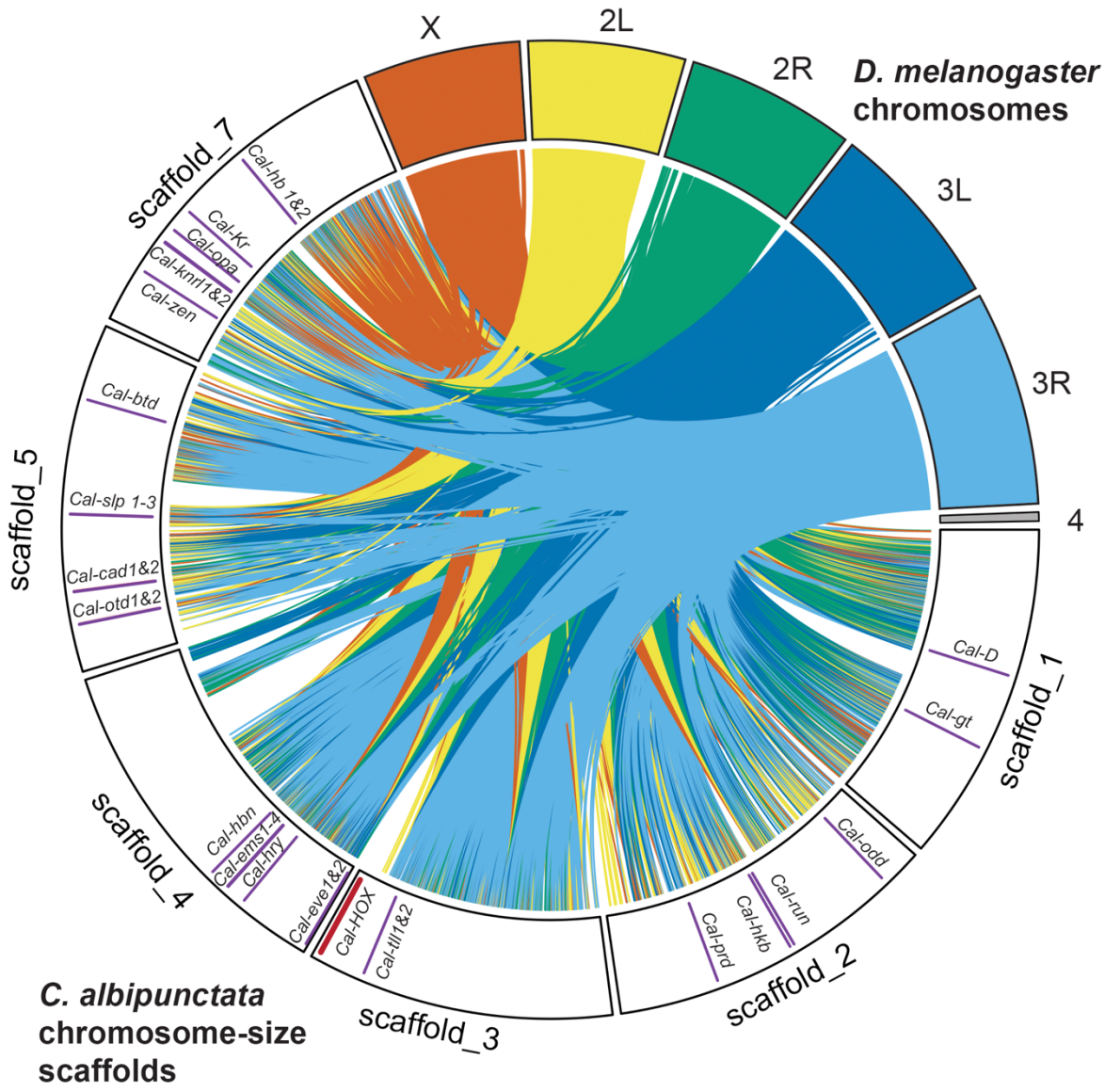

Figure 2. 1 Synteny analysis between *Clogmia albipunctata* chromosomes-sized scaffolds *and Drosophila melanogaster* chromosomes.
Groups of collinear genes are represented by contact lines connecting the positions shared between *D. melanogaster* chromosomes and *C. albipunctata* scaffolds. Colors correspond to *D. melanogaster* chromosomes. Positions of developmental genes of interest indicated by solid purple bars. Position of the HOX gene cluster indicated by a solid brown bar.

Table 2. 2 *Clogmia albipunctata* annotation statistics and BUSCO scores.
The quality metrics include the percentage of complete universal orthologs identified in the annotation across five BUSCO datasets. The number of genes in each lineage-specific BUSCO database is shown in parentheses.

Annotation Statistics	
Number of coding genes	15,047
Number of mRNAs	31,113
Number of mRNAs with 3' & 5' UTR	25,724
Mean mRNAs/gene	2.1
Mean gene length (bp)	9,163
Complete Single Copy BUSCOs	
Eukaryota (n = 255)	93.70%
Metazoa (n = 954)	93.30%
Insecta (n = 1,367)	94.00%
Endopterygota (n = 2,124)	92.90%
Diptera (n = 3,285)	88.80%

2.3.2 Distinct patterns of dynamic changes of ATAC-seq peaks during nuclear cycles 11 through 14 suggest coordinated regulation of associated genes

To identify accessible chromatin before and during the main wave of zygotic genome activation, we performed single embryo ATAC-seq at NC11, NC12, NC13, NC14 (before cellularization), and Late NC14 approximately 20 minutes before gastrulation. The embryos were staged *in vivo* using developmental time after egg activation at 25ºC and morphological criteria as previously described (Jiménez-Guri et al., 2014) and validated under our laboratory conditions. At least three replicates were performed for NC11 though NC14, and two replicates were performed for Late NC14. We pooled the ATAC-seq data from all stages to create a master peak list of 32157 open chromatin regions (peaks), using MACS2 (Zhang et al., 2008). We also computed the number of peaks for each developmental stage excluding stage-specific peaks that were not recovered as significant in the master peak list. In this analysis, we found that the overall number of peaks increases with each nuclear cycle, suggesting that the chromatin

landscape is continuously changing (**Table 2.3**). Since promoter accessibility does not necessarily indicate gene expression (Reddington et al., 2020), we divided the accessible chromatin regions into transcription start site (TSS) proximal regions (<500 bp of TSS) and TSS distal regions (>500 bp) to facilitate the distinction of open promoters and other open cis-regulatory elements. ATAC-seq peaks that only mapped to intergenic or intronic sequence were treated as candidate enhancers or silencers (henceforth collectively referred to as enhancers). The relative proportion of candidate enhancers increased by about 10% from NC11 (46.8%) to NC14 (58%) and then decreased from NC14 to Late NC14 (52.7%) (**Figure S2.1**). These observations suggest similar dynamic changes in chromatin accessibility during zygotic genome activation in Clogmia and Drosophila (Hamm & Harrison, 2018; Schulz & Harrison, 2019).

Table 2. 3 Comparison of chromatin accessibility peaks, nuclear cycle length in minutes, and haploid genome size of *Clogmia albipunctata* and *Drosophila melanogaster*.
Peaks for *Drosophila melanogaster* are based on (Blythe & Wieschaus, 2016; Hannon et al., 2017) and nuclear cycle lengths are based on (Jiménez-Guri et al., 2014).

	Clogmia albipunctata				
	NC11	NC12	NC13	NC14	Late NC14
Peaks	5013	5926	7586	16147	11628
Duration	30 m	37 m	1 h 13 m	1 h 47 m	
Genome size	~304 Mb				

	Drosophila melanogaster			
	NC11	NC12	NC13	NC14
Peaks	3084	6740	9824	13226
Duration	10 m	12 m	21 m	1 h 5 m
Genome size	~137 Mb			

To identify peaks with dynamic patterns of chromatin accessibility, we compared the degree of accessibility of each peak between successive timepoints (**Figure** 2.**2A-D**) using DESeq2 (Love et al., 2014). This analysis identified a total of 6930 dynamically regulated peaks (adjusted p-value ≤ 0.05 and $|\log_2 \text{Fold Change}| \geq 1$; 21.6% of all 32157 peaks) and enabled us to quantitatively determine the behavior of each peak (i.e., maintaining, gaining or losing accessibility) between consecutive developmental stages. We then clustered dynamic peaks into groups based on changes in accessibility from stage to stage and delineated 21 dynamic groups (**Figure S2.2**). Nearly 80% (5403 / 6930) of the dynamic peaks fall into one of 8 groups, which reveal the major accessibility trends we observe in the data (**Figure 2.2E)**. Many of the dynamic peaks closed (14.6%; Groups 5,13) or opened (~20.4%; Groups 8,12) gradually over time. The remaining dynamic peak groups exhibited more complex dynamics between NC11 and Late NC14, commonly with an inflection at NC13 (e.g., Groups 3,1,15,14; see also **Figure S2.2**). These observations suggest that remodeling of chromatin accessibility between NC11 and Late NC14 at individual peaks often follows patterns that are shared by many peaks and may reflect co-regulation of the associated genes.

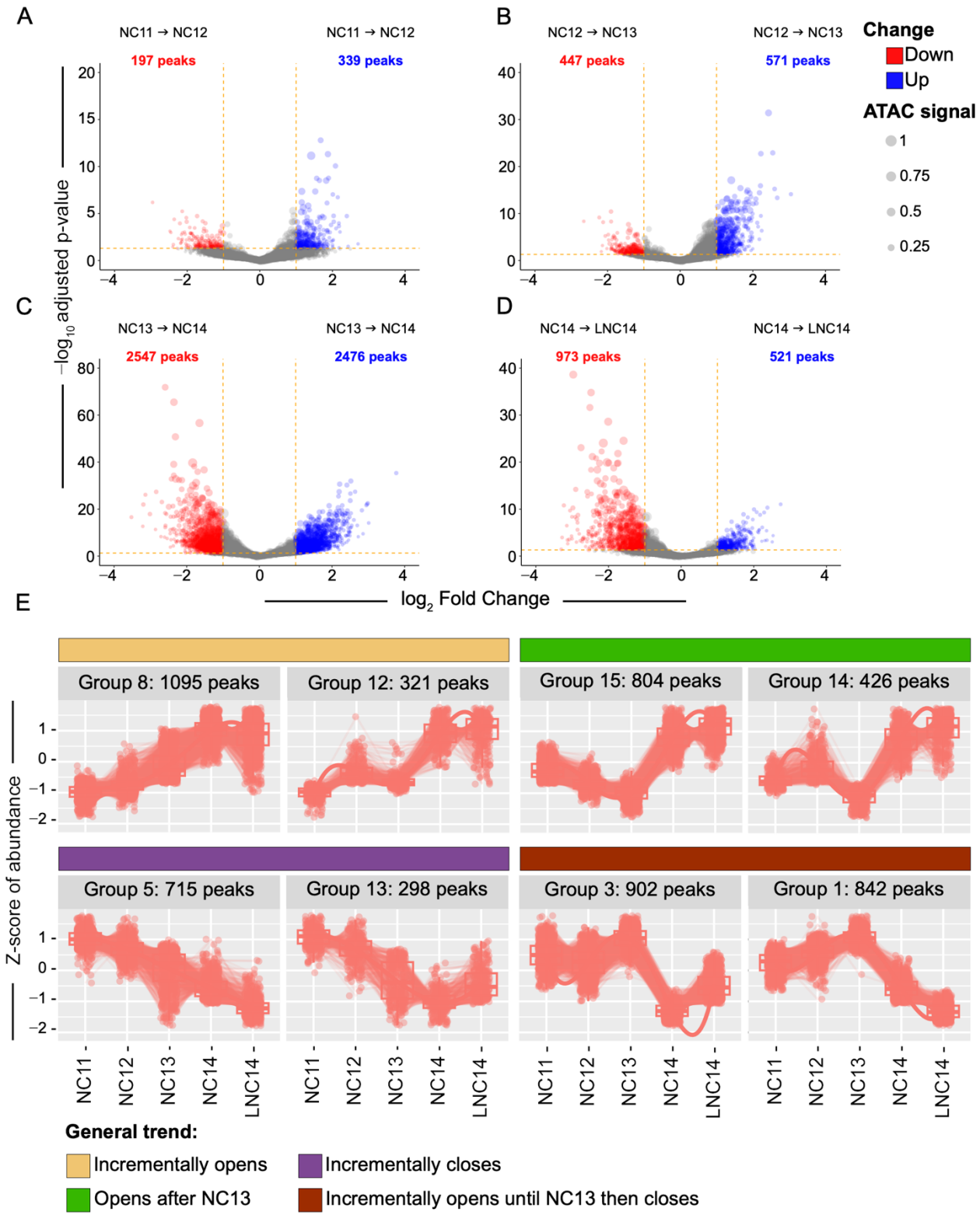

Figure 2. 2 Changes in chromatin accessibility during NC11 to Late NC14.
A-D) Volcano plots of differentially accessible ATAC-seq peaks between consecutive developmental stages. Significant change of adjusted p-value ≤ 0.05 and |$\log_2$ Fold Change| ≥ 1

highlighted; red (log_2 Fold Change ≤ -1) and blue (log_2 Fold Change ≥ 1). E) Dynamic peak groups produced using DEGreport (Pantano, 2017) (R package version 1.38.5). Y-axis, Z-score of abundance. X-axis, stage.

2.3.3 Cal-zld is required for both opening and closing chromatin at the blastoderm stage in Clogmia

Previous studies in Drosophila have shown that temporal changes in chromatin accessibility are regulated by the interplay of transcription factors that can interact with nucleosome-bound DNA (pioneer factors) and transcription factors that have limited or no ability to initiate chromatin accessibility but affect the accessibility of open chromatin (Brennan et al., 2023; Larson et al., 2021). The earliest changes in chromatin accessibility are primarily driven by the pioneer factor Zld (Harrison et al., 2011; Liang et al., 2008; McDaniel et al., 2019; Schulz & Harrison, 2019; Sun et al., 2015). As zygotic genome activation accelerates, additional pioneer factors, such as GAF (Gaskill et al., 2021), CLAMP (Colonnetta et al., 2021; J. Duan et al., 2021), and Opa (Koromila et al., 2020; Soluri et al., 2020) are required for maintaining, modulating, or gaining chromatin accessibility. In Clogmia, a motif enrichment analysis of peaks with the MEME software package (Bailey, 2021; Bailey et al., 2015) revealed an enrichment of a Zelda-binding motif (5'-CAGGTA), GA-rich motifs recognized by GAF and CLAMP (Gaskill et al., 2021; Kaye et al., 2018; Kuzu et al., 2016; Soruco et al., 2013), in addition to $(CA)_n$ motifs recognized by Combgap (Cg) (P. Ray et al., 2016), and others (**Figure S2.3**). Opa-binding motifs were not recovered in this analysis. These results are comparable to findings in Drosophila (Blythe & Wieschaus, 2016; Koromila et al., 2020; Soluri et al., 2020) and suggest that key pioneer factors acting during zygotic genome activation are likely conserved between Drosophila and Clogmia.

Homologs of *zelda* have been identified in many insects and some crustaceans (Ribeiro et al., 2017), and several transcriptomic studies suggest that Zld's function as pioneer transcription

factor during zygotic genome activation is conserved across insects (Arsala & Lynch, 2017; Biedler et al., 2012; Gao et al., 2023; Pires et al., 2016; Ventos-Alfonso et al., 2019). However, these comparative studies did not directly test for Zld-dependent effects on chromatin accessibility. We identified Clogmia *zld* (*Cal-zld*) (**Figure S2.4**) and confirmed high levels of *Cal-zld* expression at preblastoderm and blastoderm stages (**Figure S2.5**). Loss of Zld in *Drosophila* results in a failure to activate a large cohort of early zygotic genes, including those associated with the process of cellularization (Liang et al., 2008). To examine phenotypic effects of *Cal-zld* depletion, batches of ~30 embryos were injected with *Cal-zld* dsRNA (double-stranded RNA) within the first hour of development (NC2/3). A subset of embryos from each batch (3 replicates) were monitored *in vivo* and allowed to develop until the end of NC14 (roughly 7-7.5 hours after injection). Many of the embryos in each batch failed to develop a blastoderm (possibly due to injection artifacts) while some developed normally, presumably because of insufficient *Cal-zld* knockdown. A third class of embryos reached NC14 but failed to cellularize and developed aberrantly (**Movie S2.1**). This phenotype is comparable to the phenotype of *in vivo* imaged *zld* mutant Drosophila embryos (**Movie S2.2**) (McDaniel et al., 2019) and suggests that the role of Zld for the regulation of early zygotic genes is conserved in *Clogmia.* We also quantified knockdown efficiencies of *Cal-zld* transcript in individual embryos by RT-qPCR (**Table S2.3**). In one quarter of the *Cal-zld* RNAi embryos (5/20), *Cal-zld* mRNA levels were reduced by ~30-90%.

To assess *Cal-zld*'s role in shaping chromatin accessibility and zygotic genome activation, we adapted our ATAC-seq protocol to isolate RNA from single-embryo ATAC-seq library preparations of Late NC14 *Cal-zld* RNAi embryos (8 replicates). In parallel, we processed stage-matched untreated embryos that were allowed to develop under water in the dish

used for egg activation (wild-type controls; 5 replicates), and embryos that were prepared for injection under oil but not injected (alignment controls; 5 replicates). The isolated RNA was used to generate RNA-seq libraries to measure the efficiency of the *Cal-zld* knockdown as well as transcriptional effects of the knockdown in individual embryos. To identify individual injected embryos with strong *Cal-zld* knockdowns, we quantified *Cal-zld* expression levels using the RNA-seq data and chose embryos with knockdown efficiencies greater than 75% (n= 4, **S6 Figure** and see *Materials and Methods*). We then compared gene expression and ATAC-seq peaks of these *Cal-zld* RNAi embryos to those of the alignment control group using DESeq2. Loss of *Cal-zld* affected 2230 genes of which 865 (38.8%) were downregulated and 1365 (61.2%) were upregulated. We then compared genes differentially expressed in *Cal-zld* RNAi embryos to genes we found to be expressed maternally, zygotically, or both (**Figure 3.3A, Figure S2.7A** and *Materials and Methods*). 691 genes were classified as zygotically expressed of which 209 (30.2%) were downregulated in *Cal-zld* RNAi embryos and 63 (9.1%) were upregulated (corresponding to 24.2% of all 865 downregulated and 4.6% of all 1365 upregulated genes, respectively). 6658 genes were classified as expressed maternally and zygotically of which 593 (8.9%) were downregulated and 1232 (18.5%) were upregulated. Finally, of the 227 genes classified as maternally expressed, only 5 (2.2%) were found to be downregulated and 14 (6.2%) upregulated in *Cal-zld* RNAi embryos. Since the strictly zygotic genes were ~3 times more often downregulated than upregulated (209 versus 63), our results suggest that *Cal-zld* plays an important role in activating zygotic genes (**Figure 2.3A),** consistent with the role of Zld during zygotic genome activation in Drosophila embryos (Harrison et al., 2011; Liang et al., 2008). Changes in the levels of maternal transcripts in response to *Cal-zld* RNAi could reflect indirect posttranscriptional roles of *Cal-zld* in their regulation.

At the level of chromatin accessibility, *Cal-zld* depletion affected 4102 of all 32157 peaks (12.8%). 2399 of them lost accessibility (58.4%) and 1703 gained accessibility (41.5%). The majority (67%) of the peaks mapped to intergenic or intronic sequences (**Figure 2.3B)**. Depletion of *Cal-zld* mRNA affected both dynamic peaks and non-dynamic peaks (**Figure S2.7B**). Dynamic peaks were affected by *Cal-zld* knockdown in different ways (**Figure 2.3C**). Peaks in groups that progressively gained accessibility in wild-type embryos (Groups 8, 12, 15 and 14) were more likely to be *Cal-zld* dependent and to close in *Cal-zld* RNAi background (**Figure 2.3C, *left*** and **Figure S2.8A**). This observation is consistent with Cal-Zld functioning to open new chromatin regions during zygotic genome activation. Additionally, our data suggest that Cal-Zld also plays an important role in closing chromatin (**Figure S2.8B**). For example, peaks in dynamic groups whose accessibility decreased in wild-type with each nuclear cycle or that closed after NC13 (Groups 5 and 1) were more likely to require Cal-Zld to close (**Figure 2.3C, *right*** and **Figure S2.8B**).

To assess if the changes in accessibility we observed were a direct consequence of *Cal-zld* knockdown, we performed two types of motif analyses. We first performed a motif enrichment analysis with MEME on *Cal-zld* sensitive peaks to identify motifs enriched centrally within 50 bp of the peak summit (**Figure S2.9**). The Zld motif was enriched in peaks that lost accessibility following *Cal-zld* knockdown, consistent with a direct role of Cal-Zld in nucleosome depletion. A Zld motif was not enriched in peaks that gained accessibility following *Cal-zld* knockdown (**Figure 2.3D** and **Figure S2.9)**. The enrichment of Zld motifs in only one class of *Cal-zld* sensitive peaks may reflect the prevalence of Cal-Zld binding in those regions. In support of this, motif affinity and occupancy positively correlates with Zld's pioneering activity in Drosophila (Brennan et al., 2023; Gibson et al., 2024) . Next, we consider whether Cal-Zld

could bind at other sites within *Cal-zld* sensitive peaks by performing a motif search spanning the full width of each peak, using Zld motifs derived from our MEME analysis in Clogmia (5-CAGGTA; cf. **Figure S2.3)** and experimental data in *Drosophila melanogaster* (Fly Factor Survey) (Zhu et al., 2011). Overall, 2857 of the 4102 *Cal-zld* sensitive peaks contained at least one Zld motif within the full peak width. Consistent with the motif enrichment analysis, a Zld motif was prevalent in peaks whose accessibility was lost following *Cal-zld* depletion (81%, 1952/2399 peaks). Interestingly, we also observe Zld motifs in peaks whose accessibility increased following *Cal-zld* knockdown (53%, 905/1703 peaks), therefore a subset of these regions may close due to Cal-Zld activity directly.

Taken together, our findings suggest that Cal-Zld plays a central role in both opening and closing chromatin accessibility in pre-gastrulation embryos. This role appears to be unique to Cal-Zld as Zld activity in Drosophila has been documented as exclusively pioneering chromatin (Hannon et al., 2017; Harrison et al., 2011; Schulz et al., 2015; Soluri et al., 2020; Sun et al., 2015). Recent evidence suggests, however, that Zld can lead to both activating and repressive histone modifications (Galouzis et al., 2025; Gonzaga-Saavedra et al., 2025). Moreover, Zld isoforms have been described including a maternal specific isoform with repressive activity (Giannios & Tsitilou, 2013; Hamm et al., 2015, 2017; Pearson et al., 2012). We hypothesize that Cal-Zld can open or close chromatin depending on the context, and this activity is mediated through potential Cal-Zld isoforms and interactions with histone modifying complexes or cofactors.

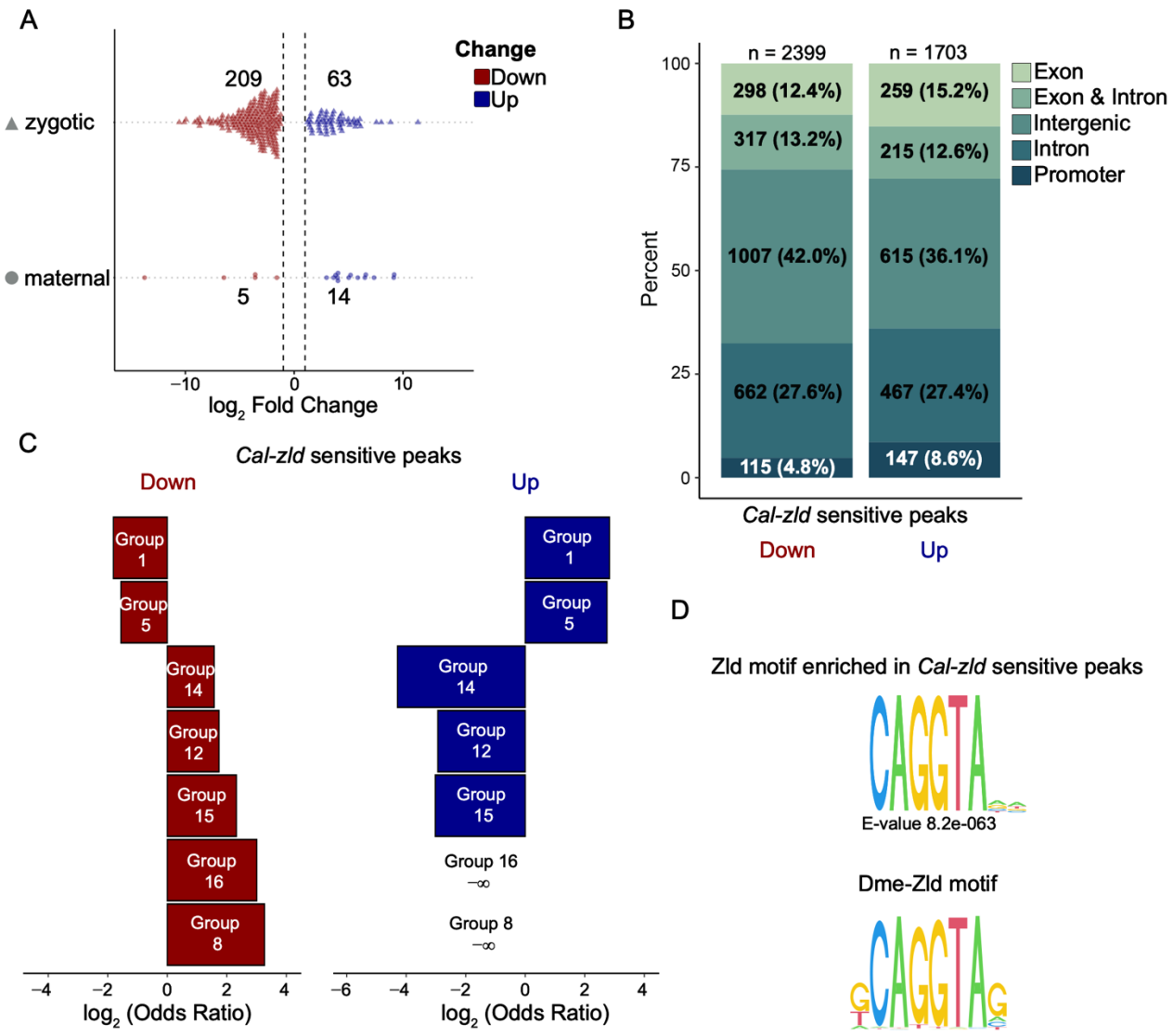

Figure 2. 3 Cal-Zld regulates chromatin accessibility and zygotic transcription positively and negatively.

A) Observed expression change of maternal or zygotic genes that are differentially expressed in *Cal-zld* RNAi embryos. On the y-axis, genes are grouped by maternal or zygotic classification (see *Materials and Methods*). Log_2 Fold Change on the x-axis. Only genes with an adjusted p-value ≤ 0.05 and $|\log_2 \text{Fold Change}| \geq 1$ are shown. B) Genomic distribution of *Cal-zld* sensitive peaks that were up-regulated (blue) and down-regulated (red). C) Dynamic groups whose accessibility increases over time are more likely to be *Cal-zld* sensitive peaks that lose accessibility (Group 8 p-value, 3.4e-176; Group 16 p-value, 8.7e-05; Group 15 p-value, 1.3e-55; Group 12 p-value, 4.6e-12; Group 14 p-value, 7.2e-12) unlike dynamic groups whose accessibility decrease over time or after NC13 (Group 5 p-value, 3.5e-08; Group 1 p-value, 2.7e-12). Dynamic groups whose accessibility decreases over time are more likely to be *Cal-zld* sensitive peaks that gain accessibility (Group 5 p-value, 1.7e-70; Group 1 p-value, 1.4e-95), unlike dynamic groups whose accessibility increases over time (Group 8 p-value, 3.4e-25; Group 16 p-value, 6.3e-01; Group 15 p-value, 1.8e-11; Group 12 p-value, 9e-05; Group 14 p-value,

2.6e-07). Odds ratios and p-values were calculated by a two-sided Fisher's exact test on contingency tables constructed by the presence/absence of dynamic groups in *Cal-zld* sensitive peaks up or down. Only groups with a $\log_2$(Odds Ratio) > 1.5 in at least one comparison are shown. D) Position weight matrix (PWM) logo of the Zld motif enriched in *Cal-zld* sensitive peaks as identified by MEME (top) and the canonical Zld motif in *Drosophila melanogaster*, shown for reference (bottom).

2.3.4 Chromatin accessibility and gene expression at NC12 and NC13 following knockdown of maternal *Cal-opa*

Having established that *Cal-zld* regulates chromatin accessibility in Clogmia embryos, we asked whether maternal *Cal-opa* affects chromatin accessibility during early stages of axial pattern formation. Since *Cal-opa* exerts its symmetry-breaking function in axis specification prior to the onset of zygotic *Cal-opa* expression during blastoderm cellularization at mid NC14 (Yoon et al., 2019), we were able to examine the function of *Cal-opa* without confounding effects of zygotic *Cal-opa* activity.

To assess the impact of maternal Cal-Opa activity on chromatin accessibility, we performed ATAC-seq and RNA-seq on single *Cal-opa* RNAi embryos at NC12 (6 replicates) and NC13 (8 replicates). In parallel, we processed stage-matched untreated controls that were allowed to develop under water in the dish used for egg activation (wild-type controls; 2 replicates for NC12 and 3 replicates for NC13), as well as controls that were prepared for injection under oil but not injected (alignment controls; 4 replicates for each developmental stage), and embryos that were injected with dsRNA of the extraneous gene *DsRed* (injection controls; 3 replicates for each stage). To identify RNAi embryos with maximally reduced *Cal-opa* mRNA levels, we used the RNA-seq data to measure *Cal-opa* expression and selected embryos with knockdown efficiencies greater than 75% (**Figure S2.10** and see *Materials and*

Methods). We identified 4 embryos at NC12 and 5 embryos at NC13 with strong *Cal-opa* knockdowns and assessed changes in chromatin accessibility in these embryos by performing DESeq2 analysis with the alignment control group as reference. At NC12, very few changes in chromatin accessibility were observed in response to *Cal-opa* RNAi. Out of the 86 differentially accessible regions, 79 (91.9%) lost accessibility and 7 (8.1%) gained accessibility (**Figure 2.4A**). At NC13, depletion of *Cal-opa* had a more pronounced effect on chromatin accessibility, with 250/292 (85.6%) peaks losing accessibility and 42/292 (14.4%) gaining accessibility **(Figure 2.4B**). Most of the differentially accessible regions mapped to putative enhancer regions (**Figure S2.11**). We also performed a motif enrichment analysis using the MEME suite on all peaks as well as the peaks that either lost or gained accessibility in *Cal-opa* RNAi embryos at either NC12 or NC13 (**Figure 2.4C**). In regions that lost accessibility in *Cal-opa* RNAi embryos, the most enriched motifs were GAF/CLAMP and Opa binding motifs. In addition, we recovered motifs that have not been shown to be bound by a TF in Drosophila. Regions that gained accessibility were enriched in DNA binding motifs of GAF/CLAMP, Abdominal-B (Abd-B), Twist (Twi), and two motifs with no associated transcription factor.

Next, we determined the number of *Cal-opa* sensitive peaks that contained putative Cal-Opa binding sites, including Opa motifs derived from our MEME analysis and a MEME analysis of ATAC-seq data from Drosophila (Soluri et al., 2020). At NC12, 63 of the 79 peaks losing accessibility and none of the 7 peaks gaining accessibility in response to reduced *Cal-opa* activity contained at least one Opa motif (**Appendix S2.2**). At NC13, 112 of the 250 peaks losing accessibility and 6 of the 42 peaks gaining accessibility in response to reduced *Cal-opa* activity contained at least one Opa motif (**Appendix S2.3**). These results support the hypothesis

that Opa directly promotes chromatin accessibility during the early phase of zygotic genome activation.

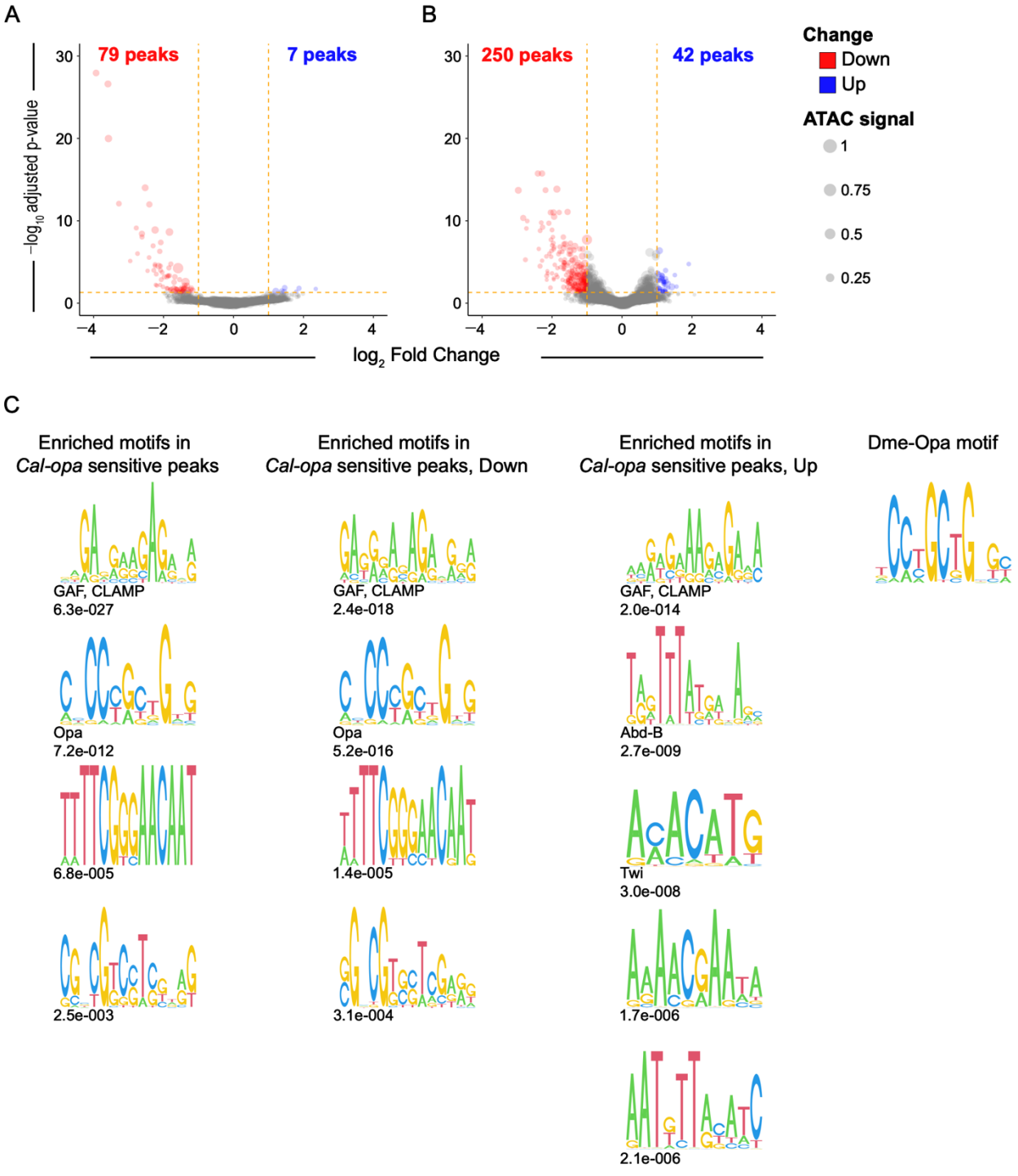

Figure 2. 4 Cal-Opa is required for establishing a selective set of chromatin accessibility peaks.

Volcano plots of differentially accessible peaks between the *Cal-opa* RNAi group and the alignment control group at A) NC12 and B) NC13. Significant change of adjusted p-value ≤ 0.05 and $|\log_2$ Fold Change$| \geq 1$ highlighted; red ($\log_2$ Fold Change ≤ -1) and blue ($\log_2$ Fold Change ≥ 1). C) PWM logos of motifs enriched in all *Cal-opa* sensitive peaks (n = 326), in *Cal-opa* sensitive peaks that lose accessibility (n = 277), and in *Cal-opa* sensitive peaks that gain accessibility (n = 49), as identified by the MEME suite. Data combined for both NC12 and NC13 stages. DNA binding proteins of *Drosophila melanogaster* that bind to these or similar motifs as identified by MEME are indicated. We include CLAMP in this list as it has been shown to bind GA-rich motifs (see text). E-value denotes the significance of recovered motifs. The canonical Opa motif in *Drosophila melanogaster* is shown for reference.

2.3.5 Maternal Cal-Opa drives chromatin accessibility and expression at *homeobrain* and *sloppy-paired* loci in complementary anterior domains

The early zygotic segmentation gene network of dipterans is dominated by TF-encoding genes, such as the gap genes and pair-rule genes (Bullock et al., 2004; Clark et al., 2019; García-Solache et al., 2010; Goltsev et al., 2004a; Jaeger, 2011; Lemke et al., 2010; Nüsslein-Volhard & Wieschaus, 1980; Wotton, Jimenez-Guri, et al., 2015). Since Bcd breaks axial symmetry by targeting these genes in a concentration-dependent manner, we focused our search of Cal-Opa targets on TF genes that are expressed during blastoderm formation. Using the RNA-seq data from the embryos that were used to identify regions of differential chromatin accessibility, we identified 84 differentially expressed genes at NC12, including 11 transcription factor genes, all of which were downregulated upon *Cal-opa* knockdown (**Figure 2.5A** and **Appendix S2.2**). At NC13, we identified 1047 differentially expressed genes, including 116 putative transcription factor genes of which 67 were downregulated and 49 were upregulated in *Cal-opa* RNAi embryos (**Figure 2.5B** and **Appendix S2.3**). We then asked which of the differentially expressed transcription factor genes are located near open chromatin regions (ATAC-seq peaks) with an Opa motif. This analysis was restricted to peaks within a 40 kb region centered on each gene body of interest and further narrowed the list of potential key target genes of Cal-Opa to 10

genes at NC12 and 72 genes at NC13 (**Appendix S2.2** and **Appendix S2.3)**. At NC12, 4 of the 10 candidate genes were associated with at least one Cal-Opa dependent ATAC-seq peak, including *Cal-opa,* and homologs of *homeobrain* (*Cal-hbn*) and *sloppy-paired* (*Cal-slp1*, *Cal-slp2*) (**Figure 2.6, Table S2.4,** and **Figure S2.12**). At NC13, we thereby identified *Cal-opa, Cal-slp1*, *Cal-slp2* and *Cal-hbn* in addition to Clogmia orthologs of *Dichaete* (*Cal-D*), *Krüppel* (*Cal-Kr*), and an unidentified C2H2 zinc finger gene among the 72 candidates (**Figure 2.7, Table S2.4, Figure S2.12** and **Figure S2.13**).

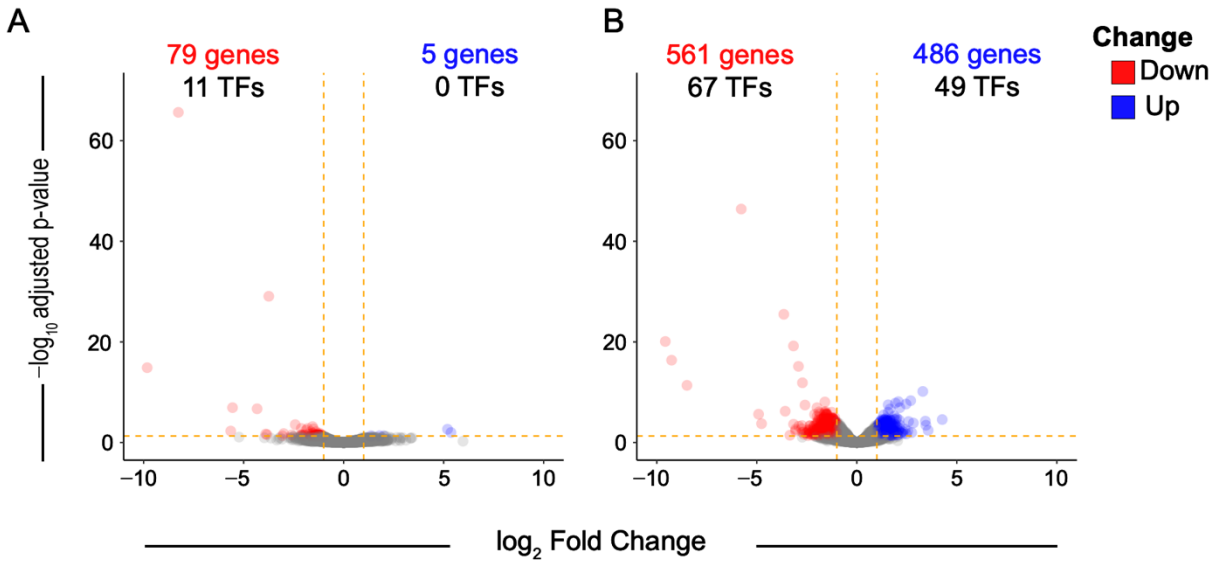

Figure 2. 5 *Cal-opa* knockdown leads to the miss-regulation of a small set of transcription factor genes.
Volcano plots of differentially expressed genes between the *Cal-opa* RNAi group and the alignment control group at A) NC12 and B) NC13. Significant change of adjusted p-value ≤ 0.05 and $|\log_2 \text{Fold Change}| \geq 1$ highlighted; red ($\log_2 \text{Fold Change} \leq -1$) and blue ($\log_2 \text{Fold Change} \geq 1$). The number of transcription factors (TFs) is indicated in the text.

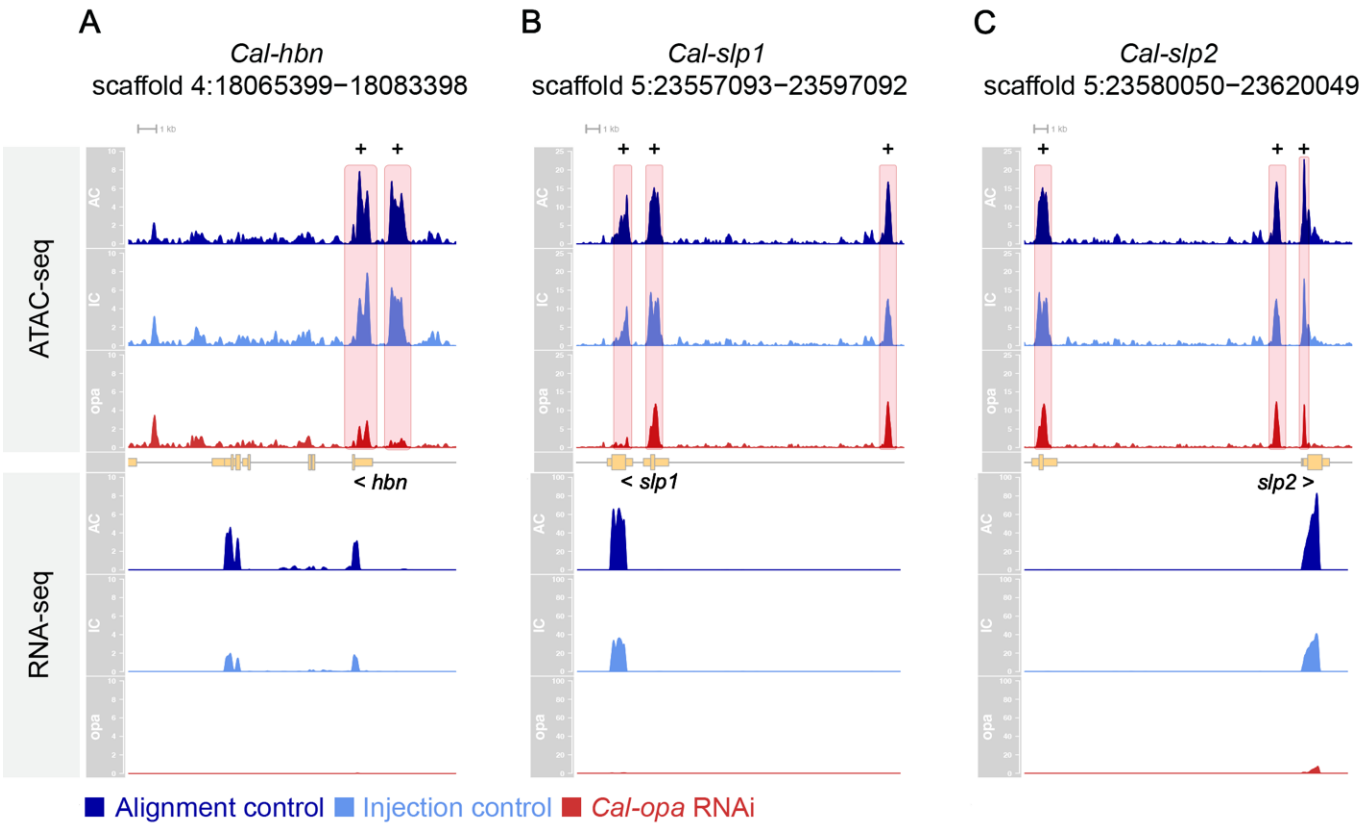

Figure 2. 6 Key Cal-Opa targets at NC12.
Gene loci of A) *Cal-hbn*, B) *Cal-slp1*, and C) *Cal-slp2* at NC12. Top tracks: CPM of ATAC-seq data (Alignment control in blue, Injection control in red, *Cal-opa* RNAi in yellow). Middle tracks show gene models in pale yellow. Bottom tracks: CPM of RNA-seq data. Peaks with significant changes in accessibility are highlighted in pink. Peaks containing Opa motifs are marked with a plus sign. The y-axis is the same for all tracks of the same data type at each locus. *Cal-hbn* ATAC-seq: 0-10 CPM. *Cal-hbn* RNA-seq: 0-10 CPM. *Cal-slp1* ATAC-seq: 0-25 CPM. *Cal-slp1* RNA-seq: 0-100 CPM. *Cal-slp2* ATAC-seq: 0-25 CPM. *Cal*-slp2 RNA-seq: 0-100 CPM. Blue: Alignment control. Light blue: Injection control. Red: *Cal-opa* RNAi.

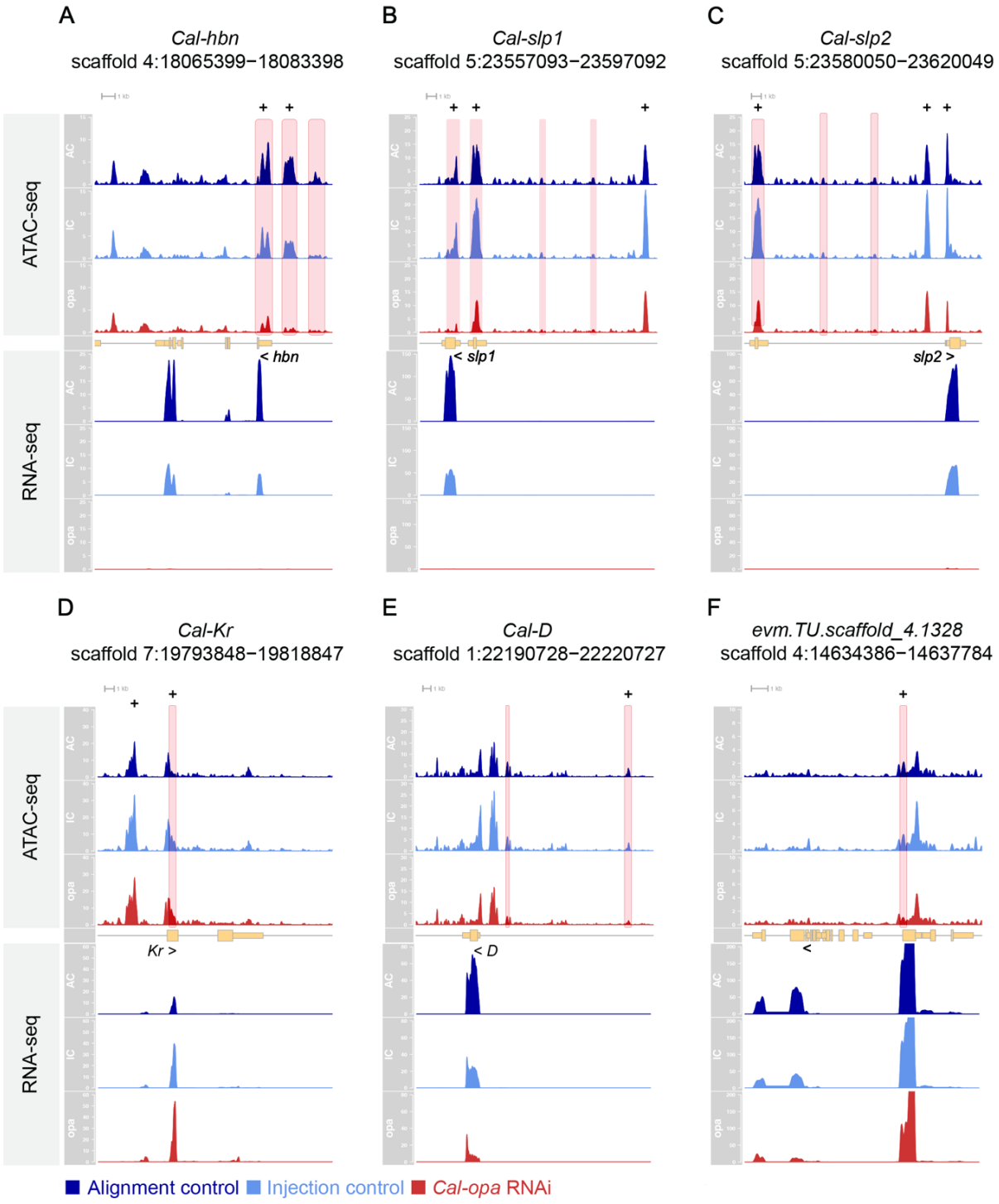

Figure 2. 7 Key Cal-Opa targets at NC13.
Gene loci of A) *Cal-hbn*, B) *Cal-slp1*, C) *Cal-slp2*, D) *Cal-Kr*, E) *Cal-D* and F) *evm.TU.scaffold_4.1328* at NC13. Top tracks: CPM of ATAC-seq data. Middle track gene model in pale yellow. Bottom tracks: CPM of RNA-seq data. Peaks with significant changes in accessibility highlighted in pink. Peaks containing Opa motifs are marked with a plus sign. The

y-axis is the same for all tracks of the same data type at each locus. *Cal-hbn* ATAC-seq: 0-15 CPM. *Cal-hbn* RNA-seq: 0-25 CPM. *Cal-slp1* ATAC-seq: 0-25 CPM. *Cal-slp1* RNA-seq: 0-150 CPM. *Cal-slp2*: ATAC-seq: 0-25 CPM. *Cal-slp2* RNA-seq: 0-100 CPM. *Cal-Kr* ATAC-seq: 0-40 CPM. *Cal-Kr* RNA-seq: 0-60 CPM. *Cal-D* ATAC-seq: 0-30 CPM. *Cal-D* RNA-seq: 0-80 CPM. *evm.TU.scaffold_4.1328* ATAC-seq: 0-10 CPM. *evm.TU.scaffold_4.1328* RNA-seq: 0-200 CPM. Blue: Alignment control. Light blue: Injection control. Red: *Cal-opa* RNAi.

We tested the robustness of these results by repeating the analysis using the injection control group as reference. At NC12, we recovered *Cal-opa*, *Cal-hbn, Cal-slp1,* and *Cal-slp2* as candidate targets, but only three of them (*Cal-opa*, *Cal-hbn, Cal-slp1*) were associated with Cal-Opa dependent ATAC-seq peaks with Cal-Opa binding sites. In the case of *Cal-slp2*, the reduction of these peaks was insignificant (**Table S2.4** and **Appendix S2.2**). At NC13, we recovered *Cal-hbn, Cal-slp1 and Cal-slp2* (**Table S2.4** and **Appendix S2.3**). *Cal-opa*, *Cal-Kr*, *Cal-D* and the unidentified C2H2 zinc finger gene were not recovered in this analysis at NC13 because they did not meet the threshold for statistical significance (settings identical for NC12 and NC13). We therefore focused on Clogmia's *homeobrain* and *sloppy-paired* homologs.

Cal-hbn and *Cal-slp1* expression was strictly zygotic and detected during and after NC11 when both genes were expressed in non-overlapping anterior domains (**Figure 2.8A, B**). Maternal *Cal-slp2* transcript was previously found to be distributed in a shallow anterior-to-posterior gradient and may contribute to the embryo's head-to-tail polarity in subtle ways (Yoon et al., 2019). Zygotic *Cal-slp2* expression during and after NC11 was detected in an anterior stripe (**Figure 2.8C**). A third *sloppy-paired* homolog that we identified in the Clogmia genome (evm.TU.scaffold_5.987; **Figure S2.14 and Figure S2.15**) was expressed at low levels at NC12 and NC13 and robustly at NC14. The expression of this gene was reduced in *Cal-opa* RNAi embryos but associated ATAC-seq peaks did not contain Opa-binding motifs. Taken together,

our findings suggest that *Cal-hbn*, and *Cal-slp1* and *Cal-slp2* are direct early targets of Clogmia's AD and that Cal-Opa serves as their transcriptional activator.

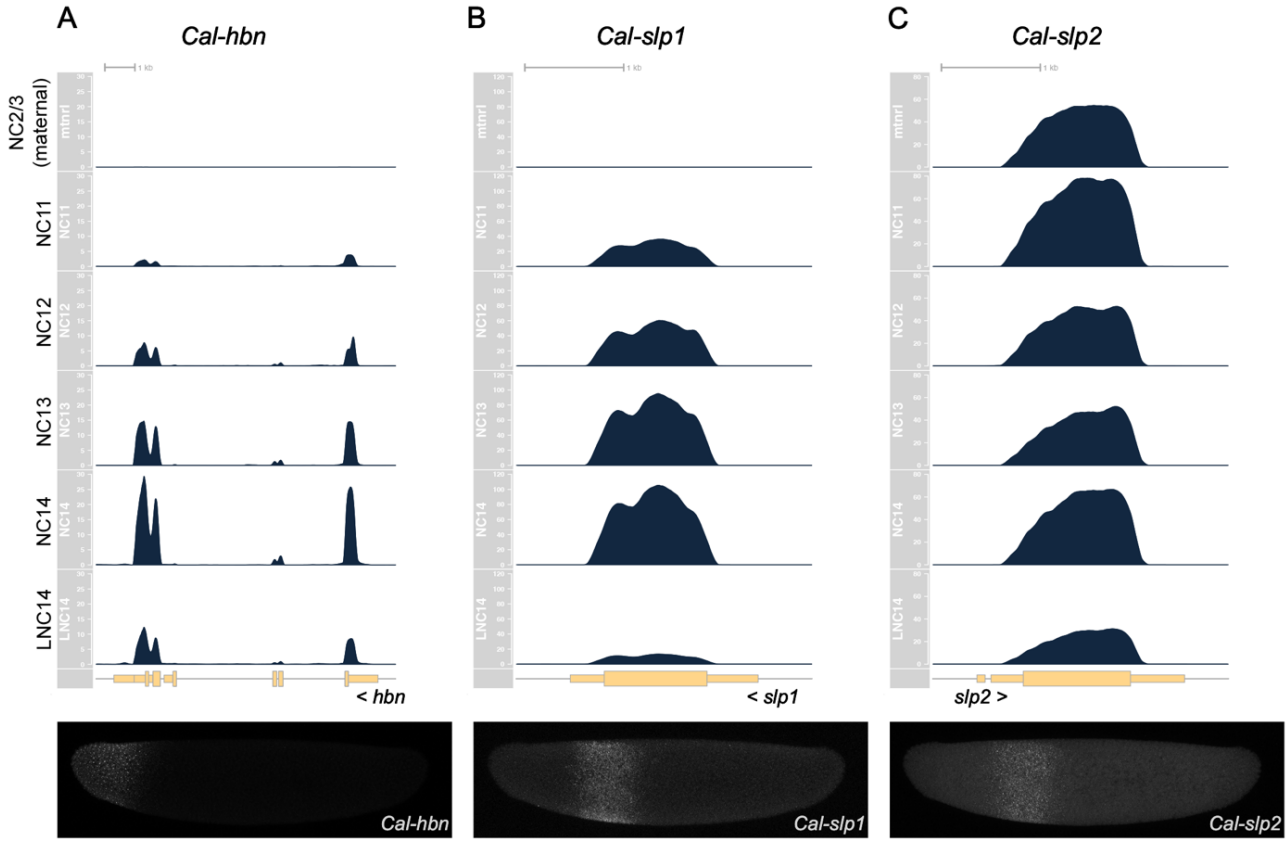

Figure 2. 8 Time-course RNA-seq data for *Cal-hbn*, *Cal-slp1*, and *Cal-slp2*.
Expression levels (CPM) of A) *Cal-hbn*, B) *Cal-slp1*, C) *Cal-slp2* at 6 consecutive stages, including NC2/3 (maternal, ~45-minute old embryos), NC11, NC12, NC13, NC14 and Late NC14 are shown above fluorescent *in situ* hybridization chain reaction (HCR) staining of each gene at NC13. Anterior, left. Dorsal, up. Gene models are shown in pale yellow. The y-axis is the same for all tracks at each locus. *Cal-hbn* RNA-seq: 0-30 CPM. *Cal-slp1* RNA-seq: 0-120 CPM. *Cal-slp2* RNA-seq: 0-80 CPM.

2.4 Discussion

2.4.1 Dynamic chromatin states in the embryo of *Clogmia albipunctata*

Dipteran embryos are morphologically similar (Anderson, 1966; Campos-Ortega & Hartenstein, 1997), but their mechanisms of axis specification (Klomp et al., 2015; Yoon et al., 2019), primordial germ cell specification (Yoon, 2019; Yoon et al., 2019), extraembryonic tissue specification (Kwan et al., 2016; Rafiqi et al., 2008, 2012), morphogenesis (Caroti et al., 2018; Rafiqi et al., 2008; Urbansky et al., 2016), and sex determination (Hall et al., 2015; Krzywinska et al., 2016; Saccone, 2022; Vicoso & Bachtrog, 2015) are not conserved across the dipteran clade. This clade is therefore an attractive model taxon to study how molecular mechanisms of embryogenesis evolve. The segmentation process of flies is particularly suitable for studying the evolution of developmental TF gene networks because it unfolds before the completion of cellularization of the multinucleated embryo when gene expression can be readily visualized and perturbed in non-traditional model organisms.

In this study, we have established the resources for dissecting the developmental gene network of early Clogmia embryos by providing an annotated chromosome-level genome sequence for our inbred laboratory strain and by describing chromatin accessibility of precisely staged consecutive blastoderm stages. We then have used these resources to examine how *zelda* and maternal *odd-paired* activities shape chromatin accessibility and gene expression in Clogmia. We found that, as in Drosophila, the period spanning zygotic genome activation is accompanied by large-scale changes in chromatin accessibility in Clogmia. However, unlike in Drosophila, where only gains in accessibility are observed during the cleavage divisions (Blythe & Wieschaus, 2016), in Clogmia we observed a substantial cohort of regions that lose accessibility over the final cleavage divisions. These regions represent sites that gain accessibility prior to NC11 and thereafter lose accessibility. In Drosophila, losses in chromatin accessibility are only observed following widespread ZGA and lengthening of the cell cycle at

NC14, which coincides with the initial formation of germ layer identities (Calderon et al., 2022; Cardamone et al., 2025; Cusanovich et al., 2018; Soluri et al., 2020). These programmatic losses in accessibility are presumed to be driven by the activity of repressive factors, either operating globally in all cells of the embryo, or conditionally as part of developmental programs for cell fate decisions (Cardamone et al., 2025).

It is important to consider that the accessibility dynamics that we observe in Clogmia may reflect the consequences of development with longer cell cycle timing. In Clogmia, NC11 through NC13 take ~3x longer than in Drosophila, in which NC11 only lasts ~10 minutes, NC12 ~12 minutes, and NC13 ~21 minutes (Jiménez-Guri et al., 2014). The much shorter cell cycle times of Drosophila present significant limitations to establish and maintain chromatin states necessary for transcriptional activity (Blythe & Wieschaus, 2016). Assuming that kinetic rates for nuclear import, DNA replication, and DNA binding of transcription factors is similar between Clogmia and Drosophila, the significantly longer syncytial interphase times in Clogmia could alleviate the impact of cell cycle dependent constraints on transcriptional regulation. It is therefore conceivable that events initiating ZGA occur at earlier nuclear cycles in Clogmia than in Drosophila, leading to earlier formation of germ layer specific chromatin accessibility patterns. Alternatively, ZGA occurs in Clogmia as it does in Drosophila, but distinct germ layer chromatin states could be primed prior to widespread ZGA. Investigating how ZGA unfolds in Clogmia and how distinct chromatin states emerge temporally at single-cell resolution will help distinguish between these possibilities.

2.4.2 Roles of Zld and Opa in shaping chromatin accessibility in Clogmia.

We demonstrate that depletion of *Cal-zld* results in widespread losses of chromatin accessibility, consistent with a conserved role for Zld in pioneering chromatin accessibility within the context of zygotic genome activation. This study also describes a novel function of Cal-Zld consistent with the unique chromatin dynamics of Clogmia. Specifically, many regions that lose accessibility during the syncytial cleavage divisions do so in a Zld-dependent manner. Based on the frequent occurrence of Zld motifs within this class of regions (53%), many of these effects are likely direct. We therefore propose that Cal-Zld functions not only to open and maintain chromatin accessibility but also to regulate the closure of many regions during nuclear cleavage cycles. This contrasts with Drosophila, in which Zld almost exclusively drives gains in chromatin accessibility during nuclear cleavage cycles (Brennan et al., 2023; Hannon et al., 2017; Li et al., 2014; Schulz et al., 2015; Soluri et al., 2020; Sun et al., 2015) . This difference remains unexplained, although we suspect that the longer nuclear cleavage cycles in Clogmia allow for additional levels of regulation of chromatin states. In this context, Cal-Zld may act upstream to activate a set of genes that function to close chromatin at later nuclear cycles. Differences in the activities of Zld and Cal-Zld could also reflect distinct protein structures or isoforms. For example, Cal-Zld contains an additional zinc-finger that has been eroded in Drosophila (**Figure S2.4**) (Ribeiro et al., 2017). Additionally, stage-specific isoforms of Zld have been described in Drosophila, including a variant lacking the C-terminal three zinc-fingers (Giannios & Tsitilou, 2013; Hamm et al., 2015, 2017; Pearson et al., 2012). This isoform functions as a dominant negative form in cell culture but flies in which the expression of this isoform is suppressed are viable and fertile (Hamm et al., 2017). Another feature to consider is

how Cal-Zld may interact with other trans-acting factors such as histone remodeling complexes. In Drosophila, Zld enables histone acetylation at promoters and enhancers through its interaction with the Drosophila homolog of LIM-domain-binding protein 1 (Ldb1, called Chip in Drosophila) which in turn recruits histone acetyltransferase CBP (Galouzis et al., 2025). In addition, Zld activity has also been shown to lead to changes in the deposition of repressive histone modification H3K27me3 within specific Polycomb Group domains (Gonzaga-Saavedra et al., 2025). Therefore, the molecular underpinnings of the distinct functional readouts of Zld and Cal-Zld are complex and likely involve both direct and indirect mechanisms.

Clogmia's anterior determinant, Cal-Opa, also plays a role in shaping chromatin accessibility states; loss of Cal-Opa results in reduced chromatin accessibility, albeit at a smaller cohort of regions compared with Cal-Zld. These Opa-dependent regions include zygotic loci whose functions we further hypothesize are necessary for mediating anterior specification downstream of Opa. Collectively, these results are consistent with a model in which the anterior determinants of Drosophila and Clogmia, while unrelated transcription factors, share a common feature of symmetry breaking at the level of cis-regulatory sequence accessibility to establish anterior fates. Mechanistically, this could occur through a canonical pioneering mechanism which requires binding of nucleosome occupied DNA and promoting nucleosome instability (Colonnetta et al., 2021; Iwafuchi-Doi & Zaret, 2014; Judd et al., 2021; Larson et al., 2021; McDaniel et al., 2019; Tsukiyama et al., 1994). A vertebrate homolog of Opa, Zic3, was found to bind nucleosomal DNA *in vivo*, suggesting that Cal-Opa may pioneer chromatin directly (Grand et al., 2024). Alternatively, Cal-Opa may function in a concentration-dependent manner to outcompete nucleosomes for DNA occupancy (Polach & Widom, 1995), similar to Bcd (Brennan et al., 2023; Degen et al., 2024; Hannon et al., 2017).

2.4.3 Functional comparison of the ADs in Clogmia and Drosophila

We find that homologs of *homeobrain* and *sloppy-paired* are early targets of Cal-Opa. These genes may play key roles in initiating anterior pattern formation not just in Clogmia but also in other insects. In Drosophila, *homeobrain* is required for labrum and brain development (Hildebrandt et al., 2022; Kolb et al., 2021; Walldorf et al., 2000) and is associated with enhancers that open in a Bcd-dependent manner (Hannon et al., 2017), and in the beetle *Tribolium castaneum*, *homeobrain* was found to play an important zygotic role in establishing the embryo's anterior-posterior polarity (Ansari et al., 2018). The *sloppy-paired* gene, a member of the *forkhead* domain gene family (Schomburg et al., 2022), is best known for its role as a pair-rule segmentation gene (Clark, 2017; Clark et al., 2019), but it is first expressed in a head domain of the syncytial Drosophila embryo, like its closest paralogs, *sloppy-paired 2* and *fd19B* (Cadigan et al., 1994a; Fujioka & Jaynes, 2012; Grossniklaus et al., 1992; Lee & Frasch, 2004). Partially redundant functions of sloppy-paired paralogs, which are common across insects (**Figure S2.14** and **Figure S2.15**), could have obscured a fundamental role of *sloppy-paired* activity in head development in Drosophila. However, in Tribolium, there is stronger evidence suggesting that *sloppy-paired* genes may have played a more critical role in anterior patterning. In this species, the *slp* paralogs may have both gap and pair-rule functions; the mandibular, maxillary, and labial segments are lost in addition to even-numbered trunk segments when *Tca-slp1* is knocked down (Choe et al., 2006; Choe & Brown, 2007).

An important limitation of our study is that we cannot rule out the possibility that maternal Cal-Opa targets additional genes via constitutively accessible cis-regulatory DNA. Testing this possibility will require new genetic tools or reagents to assess where maternal Cal-Opa binds to

genomic DNA *in vivo*. In the meantime, we provide a preliminary discussion of how targets of Cal-Opa and Bcd compare. Bcd contributes to the activation of many early segmentation genes but its role in the posterior embryo is much more subtle than in the anterior embryo, where all expression domains directly or indirectly depend on Bcd. Target enhancers of Bcd have been classified based on how their accessibility depends on Bcd and Zld (Hannon et al., 2017). The accessibility of enhancers targeted by Bcd that drive expression in the posterior region of the embryo is independent of Bcd (e.g., *Krüppel* CD1, *knirps* posterior, *giant* posterior). However, multiple enhancers that open in a Bcd-dependent manner were found to be associated with genes that are expressed in the anterior region of the embryo (e.g., *orthodenticle*, *empty spiracles*, *buttonhead*, *giant*, *knirps*, *hunchback*, *sloppy-paired 1*, *fd19B*, *even-skipped*, *paired*, *cap'n'collar*, *huckebein, homeobrain*), and many of them correspond to enhancers that are known to drive expression of a reporter gene in the anterior embryo (e.g., *giant* anterior (-6), *knirps* anterior, *orthodenticle*, *buttonhead, even-skipped* stripe 1/5, *hunchback* P2, *hunchback* shadow enhancers, *eve-skipped* stripe 2*,* an early *paired* enhancer, *giant* anterior (-10), *cap'n'collar*, *huckebein*. reviewed in (Hannon et al., 2017)).

In Clogmia, the expression of both *hunchback* orthologs (*Cal-hb1*, *Cal-hb2*) was reduced in *Cal-opa* RNAi embryos along with associated ATAC-seq peaks but these regions did not include Opa motifs (**Figure S2.16A** and **Figure S2.16B**). Likewise, the expression of one of Clogmia's two orthologs of *orthodenticle* (*Cal-otd1*) was reduced at NC13 in *Cal-opa* RNAi embryos but was not associated with *Cal-opa* sensitive peaks containing Opa motifs (**S2.16C Figure**). The second ortholog (*Cal-otd2*) was not expressed at NC12 and NC13. In the case of Clogmia *buttonhead* (*Cal-btd*), we identified a *Cal-opa* sensitive peak with an Opa motif but the observed reduction of *Cal-btd* expression was not significant (**Figure S2.16D**). *Cal-btd*

expression could be under the control of AD-dependent and AD-independent enhancers, as previously reported for *orthodenticle* in Drosophila (Gao & Finkelstein, 1998) . Clogmia orthologs of *giant* (*Cal-gt*), *knirps* (*Cal-kni/knrl1, Cal-kni/knrl2*), and *empty spiracles* (*Cal-ems1-4*) were not expressed at NC12 or NC13 and no significant changes in chromatin accessibility were detected at these loci at NC12 or NC13 following knockdown of *Cal-opa*. The ortholog of *huckebein* (*Cal-hkb*) was not expressed at NC12 or NC13, either; however, its locus contained one Opa motif positive ATAC-seq peak that lost accessibility following *Cal-opa* knockdown at NC13. The expression of Clogmia's *cap'n'collar* (*Cal-cnc*) and *paired* (*Cal-prd*) was unaffected by *Cal-opa* RNAi, and the ATAC-seq peaks at their loci were largely *Cal-opa* insensitive. Only one small intronic peak within the Cal-*cnc* locus displayed reduced accessibility in *Cal-opa* RNAi embryos at NC13. The loci of *even-skipped* orthologs (*Cal-eve1, Cal-eve2*) contain ATAC-seq peaks with Opa motifs that were insensitive to *Cal-opa* RNAi. We note that the expression of *Cal-eve2* was elevated in *Cal-opa* RNAi embryos at NC13. Taken together, these observations point to significant differences in the key targets of Cal-Opa in Clogmia and Bcd in Drosophila and raise the question of how the downstream segmentation gene networks of these species converge.

2.5 Materials and Methods

2.5.1 Clogmia culture and embryo collection

The Clogmia culture was maintained in 100mm x15mm Petri Dishes (Fisher Scientific FB0875712) on moist, compressed cotton (Jorvet SKU: J0197) sprinkled with parsley leaf

powder (Starwest Botanicals SKU: 209895-5). For embryo collections, ~2-day old females were transferred to round 250 mL glass bottles with a ~1 inch bottom layer of moistened and compressed cotton and mated at room temperature for 2-3 days. Female ovaries were dissected with a pair of fine forceps (Dumont 5) and eggs released into water for egg activation and embryogenesis. For staged egg collections, embryos were released into flat bottom glass 100mm x15mm Petri Dishes (Corning CM: 101425255) containing 15ml of water. Care was taken to disperse the eggs on the bottom of the dish as crowding can delay development. Petri Dishes were covered with a lid and moved to an incubator set at 25° C until the desired stage. Egg harvesting post activation was on average 47 minutes for NC2/3, 261 minutes for NC11, 292 minutes for NC12, 332 minutes for NC13, 401 minutes for NC14, and 465 minutes for Late NC14.

2.5.2 Embryo fixation

Clogmia Embryos were dechorionated using a 10% dilution of commercial bleach (8.25% sodium hypochloride) for 90 seconds. Dechorionated embryos were fixed in a 50 mL falcon tube, using 20 mL of boiling salt/detergent-solution (100 µL 10% triton-X; 500 µL 28% NaCl; up to 20 mL of water). The mixture was gently swirled for 10 seconds. To stop the heat fixation, 30 mL of deionized water was added. Embryos were transferred to 1.5 mL reaction vials and devitellinized in a 1:1 mixture of n-heptane and methanol by vigorous shaking. This step was followed by 3 washes with cold methanol (stored at -80C). Embryos were inspected under a stereomicroscope and manual devitellinizations were performed as needed using sharp tungsten needles in a 3% agar plate covered with methanol. Devitellinized embryos were stored in 100% methanol at −20°C.

2.5.3 Microinjections of embryos and *Cal-opa* RNAi efficiency

Clogmia albipunctata embryo injection was done as previously described (Yoon et al., 2019). We note here that RNAi efficiency in Clogmia embryos varies (Yoon et al., 2019). This variability could be due to technical variables (e.g., injection needle, amount injected) or biological variables (e.g., differences in the RNAi response between individual embryos, or the general health or genetic background of the embryos or their mothers) or a combination of such effects. Therefore, we took care to control the age of the females and selected all experimental embryos from females with healthy looking ovaries and eggs. To assess maternal *Cal-opa* RNAi efficiency, we injected embryos in the first hour of development with dsRNA of the isoform-specific first exon of the maternal *Cal-opa* isoform. In three independently conducted experiments we reproduced the double abdomen phenotype, confirming that our knockdown reagent was satisfactory. We also quantified knockdown efficiencies of *Cal-opa* transcript in individual embryos by RT-qPCR and performed FC calculations of *Cal-opa* mRNA levels with the average level in alignment controls as reference (**Table S2.5**). This analysis suggests that *Cal-opa* RNAi can nearly abolish the *Cal-opa* transcript.

2.5.4 dsRNA synthesis

A PCR template was made using primers specific for the target gene with T7 overhangs and either cDNA, gDNA or a plasmid depending on the target sequence. The PCR product was run on a gel and purified using the Zymo Gel DNA Recovery Kit (D4007/D4008) following the manufactures instructions. The PCR template was used as the input for the transcription reaction

producing dsRNA. dsRNA synthesis and purification were done using the Invitrogen MEGAscript RNAi Kit (AM1626) following the manufacturer's instructions. Purified dsRNA was eluted in injection buffer (0.1 mM NaH_2PO_4 Ph 6.8; 5 mM KCl).

Forward and reverse primer sequences for dsRNA (lengths of dsRNAs in brackets; gene specific sequence of primers underlined):

Cal-opa, maternal transcript (201 bp):

5'-CAGAGATGCATAATACGACTCACTATAGGGAGAAAACAATTGTGAAGTGCGACA

5'-

CAGAGATGCATAATACGACTCACTATAGGGAGACAAATTTCCAAACGATGACAGA

Cal-zld (a combination of two dsRNAs were used 579 bp and 668 bp):

5'-TAATACGACTCACTATAGGGAGAAGTCCCGCAATTGATACAGC

5'-TAATACGACTCACTATAGGGAGAAGGATGTTGGTGGACACTCC

5'-TAATACGACTCACTATAGGGAGACGGAATGGTGTGTGAAACAG

5'-TAATACGACTCACTATAGGGAGATTCGAGGGGTTATTGTCCTG

dsRed (273 bp):

5'-CAGAGATGCATAATACGACTCACTATAGATGCAGAAGAAGACTATGG

5'- CAGAGATGCATAATACGACTCACTATAGCTACAGGAACAGGTGGTG

2.5.5 RT-qPCR

RT-qPCR reactions were performed using the Luna® Universal One-Step RT-qPCR Kit (New England Biolabs Ipswich, MA).

Forward and reverse primer sequences for targets (lengths of amplicon in brackets):

Cal-opa, maternal transcript (112 bp):

5'- TTGCGCTTCATTCTGTCATCG

5'- CACTGAGGGCTGATTCTCGG

Cal-zld (138 bp):

5'-AAAGATGAACCACCGTGCGA

5'-CATTGATGGCAGTGGACCCT

Cal-rpl35 (119 bp):

5'- CTGTCCAAAATCCGAGTCGT

5'- AGGTCGAGGGGCTTGTATTT

2.5.6 Fluorescent HCR *in situ* hybridizations

Gene specific probes and amplifier-fluorophore sets were designed and synthesized by Molecular Instruments Inc (Los Angeles, CA). The amplifier-fluorophore set used was B1-647 for *Cal-slp1*, B2-546 for *Cal-hbn*, and B3-488 for *Cal-slp2*. Buffers (probe hybridization, probe wash and amplification buffers) throughout the protocol were purchased from Molecular Instruments Inc. The HCR *in situ* hybridizations were performed as described (Bruce et al., 2021; Choi et al., 2018). Briefly, fixed embryos were rehydrated and washed 3 times in PBS-Tween20 (1x PBS, 0.1% Tween-20). Embryos were permeabilized for 30 minutes at room temperature (100 ml Permeabilization buffer: 100 µL 100% Triton X-100, 50 µL Igepal CA-630, 50 mg Sodium Deoxycholate (Sigma D6750-25G), 50 mg Saponin (Sigma 47036-50G-F), 200

mg BSA Fraction V (Sigma A5611-5G). Following embryos were pre-hybridized in probe hybridization buffer for 30 minutes in a 37º C water bath. Once pre-hybridization was complete, fresh hybridization solution containing gene specific probes were added and the embryos were left to incubate overnight in a 37º C water bath. On the second day the hybridization solution was removed, and the embryos were washed 4 times with pre-warmed probe wash buffer followed by two washes in 5x SSCT (5x SSC, 0.1% Tween-20). An amplifier hairpin solution was prepared following the manufacturer's instructions. Embryos were incubated with pre-warmed amplification buffer for 30 minutes at room temperature before adding amplifier hairpin solution. Embryos were left to incubate overnight at room temperature in the dark. The following day the embryos were washed 5 times in 5x SSCT at room temperature. The embryos were then counter stained with DAPI and mounted in Aqua-Polymount for imaging. Confocal imaging was performed on the Zeiss LSM 900.

2.5.7 Image Acquisition and Processing

Confocal imaging was performed on the Zeiss LSM 900 using a Plan-Apochromat 20x/0.8 M27 objective. Pixel-density of 1024 x 1024. Excitation/Emission wavelengths (in nanometers) were: 493/517 for Alexa Flour 488, 653/668 for Alexa Flour 647, 401/422 for DAPI, and 548/561 for Alexa Flour 546. For each image a Z-stack of images was captured from the embryo's surface to approximately midway through the embryo. Then a max projection was created from this Z-stack using FIJI/Image J. Brightness was increased in Photoshop. No further adjustments to the images were made.

2.5.8 ATAC-seq library preparation

Single embryo ATAC-seq library preparations were done essential as described previously in *Drosophila melanogaster* (Blythe & Wieschaus, 2016; Soluri et al., 2020). Embryos developed in an incubator at 25ºC until the desired nuclear cycle. Nuclear cycle was confirmed by time after egg activation and the morphology was checked under water using a transmission light microscope (Zeiss Stemi 2000-C). Once at the desired nuclear cycle individual embryos were placed in the lid of a 1.5 ml low-retention Eppendorf tube and homogenized in 10 µl of cold lysis buffer (10 mM Tris-HCl pH 7.4; 10 mM NaCl; 3 mM MgCl2; 1% Igepal CA-630) (Buenrostro et al., 2013), using a fire-sealed microcapillary tube (Drummond Microcap 25). Once homogenized, 40 µl of lysis buffer was added to the lid, and the body of the tube was carefully closed over the lid. Samples were place on ice until all embryos were processed. The nuclei were pelleted by centrifugation at 800 RCF at 4ºC for 10 minutes. Supernatant removal was observed under a dissection scope to ensure that the nuclei pellet was not dislodged. The nuclear pellet was resuspended in 7.5 µl of Tagmentation buffer (Illumina Tagment DNA TDE1 Enzyme and Buffer Kit) and placed on ice until all samples were resuspended. Then 2.5 µl of Tn5 Transposase was added and mixed by pipetting up and down. Tagmentation and amplification of ATAC-seq libraries were performed as described previously (Buenrostro et al., 2015). The tagmentation reaction was performed at 37ºC and 800 RPM for 30 minutes using an Eppendorf Thermomixer. Immediately after this step, the reactions were purified using the Qiagen Minelute kit following the manufacturer's instructions. Purified tagmented DNA was eluted in 10 µl of buffer EB. Library amplification was performed as described previously (Buenrostro et al., 2015). Typically, 12-14 total PCR were performed. Amplified libraries were purified using 1.8x Ampure SpRI beads following the manufacturer's instructions. Purified ATAC-seq libraries were eluted in 15 ul. The ATAC-seq library profile was evaluated using

Agilent high sensitivity bio-analyzer. Concentrations were estimated using a Qubit fluorometer. Libraries from the same experiment were pooled and sequenced together. Sequencing was done at the University of Chicago Genomics Core (Chicago, IL, USA) using the Illumina NextSeq (PE 75 bp) and at Admera Health (South Plainfield, NJ, USA) using the Illumina HiSeq (PE 150 bp) and the Illumina NovaSeq X Plus (PE 150 bp).

ATAC-seq libraries were amplified using a set of modified Buenrostro primers that introduced Unique Dual Indexes (UDI) (Soluri et al., 2020). The generalized UDI ATAC primer sequences are:

ATAC UDI Index Read 2 (i5):

5'- AATGATACGGCGACCACCGAGATCTACACnnnnnnnnTCGTCGGCAGCGTCAGATG T*G -3'

ATAC UDI Index Read 1 (i7):

5'- CAAGCAGAAGACGGCATACGAGATnnnnnnnnGTCTCGTGGGCTCGGAGATG*T -3'

Primers were synthesized with a terminal phosphorothioate bond (*) by IDT (Integrated DNA Technologies, Coralville, Iowa).

2.5.9 RNA isolation for RT-qPCR and RNA-seq experiments

Embryos developed in an incubator at 25ºC until the desired nuclear cycle. Nuclear cycle was confirmed by time after egg activation and the morphology was checked under water using a transmission light microscope (Zeiss Stemi 2000-C). RNA isolation was performed using the Zymo Quick-RNA Tissue/Insect Microprep Kit following the manufacturer's instructions except for the tissue homogenization step. Instead, homogenization was done as follows. Once at the

desired nuclear cycle individual embryos were placed in the lid of a 1.5 ml low-retention Eppendorf tube and homogenized in 10 µl of cold lysis buffer (10 mM Tris-HCl pH 7.4; 10 mM NaCl; 3 mM MgCl2; 1% Igepal CA-630) (Buenrostro et al., 2013), using a fire-sealed microcapillary tube (Drummond Microcap 25). Once homogenized the lysate was added to 450 µl of Zymo RNA Lysis Buffer and thoroughly mixed. RNA isolation resumed following the manufacturer's instructions including DNAse treatment. Purified RNA was eluted in DNase/RNase-Free Water.

2.5.10 RNA-seq library preparation

RNA quality and quantity was assessed using the Agilent bio-analyzer. Non-Strand-specific RNA-seq libraries were prepared using the Smarter v4 Ultra-Low Input protocol from Takara for cDNA synthesis and Nextera XT from Illumina for Library prep (protocols provided by Takara and Illumina). Library quality and quantity was assessed using the Agilent bio-analyzer. Libraries from the same experiment were pooled and sequenced together. Sequencing was done at the University of Chicago Genomics Core (Chicago, IL, USA) using the Illumina NovaSeq X (PE 150 bp) and at Admera Health (South Plainfield, NJ, USA) using the Illumina NovaSeq X Plus (PE 150bp).

2.5.11 ATAC-seq library preparation and RNA isolation

The first steps are the same as the standard ATAC-seq protocol above. Individual embryos are homogenized in cold lysis buffer, and a centrifugation step is performed to pellet the nuclei. Following the supernatant (~47 ul) is removed and mixed thoroughly with 450 µl Zymo

RNA lysis buffer in a fresh low-bind tube. The nuclear pellet is resuspended in 7.5 µl Tag mentation buffer and the ATAC-seq protocol continues as above with no deviations. RNA isolation proceeded immediately, or the RNA lysate was stored at temperately at -20ºC. RNA isolation was performed using the Zymo Quick-RNA Tissue/Insect Microprep Kit following the manufacturer's instructions. Isolated RNA was for RT-qPCR and RNA-seq experiments.

2.5.12 Preparation of injected embryos for ATAC-seq/ATAC-seq and RNA isolation

Following the injection embryos were left to develop on the injection slide at 25ºC. Once the embryos reached the desired nuclear stage any excess injection oil was removed using a pipette. The embryos were then washed 6 times using DI water. Embryos were removed from the injection slide using a blunted tungsten needle and transferred to the lid of a low bind tube. 10 µl of cold lysis buffer was added and the embryos were homogenized using a fire-sealed microcapillary tube. The ATAC-seq protocol proceeds as above.

2.5.13 ATAC-seq data processing

Demultiplexed reads were trimmed of adapters using TrimGalore! (Krueger, 2012) (version 0.6.10, cutadapt version 4.4) and mapped to the *Clogmia albipunctata* assembly (calbi-uni4000) using Bowtie2 (Langmead & Salzberg, 2012) (version 2.4.1) with option -X 2000. Suspected optical and PCR duplicates were marked by Picard MarkDuplicates (version 2.21.4) using default parameters (https://broadinstitute.github.io/picard/). Mapped, trimmed, duplicate marked reads were imported into R using the GenomicAlignments (Lawrence et al., 2013) (version 1.38.2) and Rsamtools (Morgan et al., 2013) (version 2.18.0). Reads that mapped, were

properly paired, non-secondary and had a map quality scores ≥ 10 were imported. Only reads mapping to scaffolds 1-5 and scaffold 7 were used for downstream analysis. The start and end coordinates of the reads were adjusted to account for the Tn5 interaction site. Watson strand start sites had four base pairs subtracted, and Crick-strand start sites had five base pairs subtracted (Buenrostro et al., 2013). Reads with mapped length ≤ 120 bp were considered to have originated from 'open' chromatin.

2.5.13.1 MACS2

Accessible chromatin peaks from NC11 through NC14 were determined using MACS2 (Zhang et al., 2008) (version 2.2.7.1) with options -f BEDPE -q 1e-25 on a merged dataset comprising all ATAC reads corresponding to open chromatin (length ≤ 120 bp). Peaks were assigned to a genomic feature on the basis of overlapping with one or more features of a gene (e.g. exon, intron, promoter). Peaks that overlapped with no gene feature were classified as intergenic.

2.5.13.2 DESeq2

A count matrix for differential enrichment analyses using DESeq2 (Love et al., 2014) (version 1.42.1) was generated by counting the number of open ATAC reads (length ≤ 120 bp) that overlapped with a MAC2 identified peak. Before running DESeq2, peaks with low count values were removed. DESeq2 design parameters were specific to the experimental data set. For the WT stage-wise comparisons (e.g. NC12 vs NC11, NC13 vs NC12, NC14 vs NC13 and LNC14 vs NC14) the design parameters passed to DESeq2 were stage and sequencing platform (~instrument.atac.data + stage). For the comparison between RNAi embryos and alignment

control embryos (e.g., *Cal-zld* RNAi vs alignment control and *Cal-opa* RNAi vs alignment control) the design components were embryo batch and genotype (~Experiment.Date + genotype). All control groups were given their own genotype to account for differences in embryo manipulation, and the genotype for the RNAi embryos was determined by the strength of knockdown (see *Fold change calculations*). The criteria for a statistical significance change were a $\log_2$ Fold Change ≥ 1 and an adjusted p-value ≤ 0.05 or a $\log_2$ Fold Change ≤ -1 and an adjusted p-value ≤ 0.05.

Dynamic peaks were defined as peaks that underwent a statistically significant change in accessibility in at least one of the WT stage wise comparisons. Different dynamic peak classes were found using the R package DEGreport (Pantano, 2017) (R package version 1.38.5) with the parameter of the minimum group size set to 10.

2.5.13.3 MEME

MEME (Bailey et al., 2015) (version 5.4.1) analysis was performed using the following options -order 1 -meme-mod zoops -maxw 14 -meme-nmotifs 15 -meme-searchsize 100000 -centrimo-score 5.0 -centrimo-ethresh 10.0. Motifs were reported from all enrichment or discovery programs within the MEME suite (i.e., MEME, STREME (Bailey, 2021), CentriMo), unless specified otherwise. MEME was given a set of sequences containing the 100 bp sequence centered on the summit of the peak. A set of background sequences were also given to MEME. To determine motifs enriched in dynamic peaks a set of non-dynamic peak sequences were provided as background sequences. To determine motifs enriched in *Cal-zld* sensitive peaks a set of sequences from peaks unaffected by *Cal-zld* were provided as background sequences. To determine motifs enriched in *Cal-opa* sensitive peaks a set of sequences from peaks unaffected

by *Cal-opa* were provided as background sequences. Identified motifs were compared to known *Drosophila melanogaster* motifs from the following databases: OnTheFly_2014, Fly Factor Survey, FLYREG, iDMMPMM, and DMMPMM databases (Bergman et al., 2005; Kulakovskii & Makeev, 2009; Kulakovskiy et al., 2009; Shazman et al., 2014; Zhu et al., 2011).

2.5.14 RNA-seq data processing

Demultiplexed reads were trimmed of adapters using TrimGalore! (Krueger, 2012) (version 0.6.10, cutadapt version 4.4; https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) and mapped to the *Clogmia albipunctata* assembly (calbi-uni4000) using Bowtie2 (Langmead & Salzberg, 2012) (version 2.4.1) with default parameters. Suspected optical and PCR duplicates were marked by Picard MarkDuplicates (version 2.21.4) using default parameters. Mapped, trimmed, duplicate marked reads were imported into R using the GenomicAlignments(Lawrence et al., 2013) (version 1.38.2) and Rsamtools (Morgan et al., 2013) (version 2.18.0). Reads that mapped, were properly paired, non-secondary and had a map quality scores ≥ 10 were imported.

2.5.14.1 DESeq2

A gene count matrix for differential enrichment analyses was generated by counting the number of RNA-seq reads that overlapped with a transcript of a gene. Before running DESeq2, genes with low count values were removed. DESeq2 design parameters were specific to the experimental data set. For the comparison between RNAi embryos and alignment control embryos (e.g., *Cal-zld* RNAi vs alignment control and *Cal-opa* RNAi vs alignment control) the design components were embryo batch and genotype (~Experiment.Date + genotype). All

control groups were given their own genotype to account for differences in embryo manipulation, and the genotype for the RNAi embryos was determined by the strength of knockdown (see *Fold change calculations*). The criteria for a statistically significant change were a $\log_2$ Fold Change ≥ 1 and an adjusted p-value ≤ 0.05 or a $\log_2$ Fold Change ≤ -1 and an adjusted p-value ≤ 0.05.

2.5.14.2 Maternal, zygotic, maternal and zygotic classification

Gene expression was compared in NC2/3 (maternal, ~45-minute old) and syncytial stage embryos (NC11- Late NC14). We assume here that NC2/3 embryos have little to no active transcription and instead are enriched in maternally deposited mRNAs. The distribution of gene expression was plotted for both stages, and a gene was determined to be expressed if its expression level was above 2^5 counts. Then genes were classified based on their expression profile in NC2/3 and syncytial stage embryos. We found 227 maternal genes, 691 zygotic genes, 6658 maternal and zygotic genes, and 7473 genes not expressed during either stage.

2.5.14.3 Fold change calculations

Fold change was calculated as the sample's counts per million divided by the control's counts per million for the gene of interest. First a counts per million (CPM) matrix was made by counting the number of RNA-seq reads that overlapped with a gene's transcript and then normalizing to counts per million. Fold change calculations were done with embryos grouped by experimental date (i.e. embryos from the same egg packet). For each embryo the CPM was divided by the average CPM of wild-type and alignment control embryos. If a date did not have any wild-type or alignment control embryos, fold change was calculated using the average CPM

of all wild-type and alignment control embryos in the data set. For *Cal-zld* RNAi experiments a fold change (FC) calculation of *Cal-zld* mRNA levels was conducted with the average of control embryos (untreated wild-type and alignment controls) from the same experimental date as reference. *Cal-zld* RNAi embryos with a FC < 0.25 (knockdown efficiency greater than 75%) were used for further analysis. Similarly, for *Cal-opa* RNAi experiments FC calculations were done separately for each stage. The fold change (FC) calculation of *Cal-opa* mRNA levels was conducted with the average of control embryos (untreated wild-type and alignment controls) from the same experimental date as reference. *Cal-opa* RNAi embryos with a FC < 0.25 (knockdown efficiency greater than 75%) were used for further analysis.

2.5.15 Genome Assembly

We used Cantata Bio's genome assembly services to generate a *de novo* reference genome for *Clogmia albipunctata* from ~70 freshly eclosed males of a 12-generation inbred line of an established laboratory culture (K. B. Rohr et al., 1999). High-fidelity (HiFi) PacBio reads were used to generate an initial draft assembly subsequently refined with Omni-C reads (**Appendix S2.1**). Details on sequencing and software used to assemble the genome have been described elsewhere (Tenger-Trolander et al., 2025). We estimated genome heterozygosity and sequencing error rates using jellyfish to count 21-mers and GenomeScope2.0 to analyze the k-mer frequencies.

2.5.16 Genome Annotation

We used *RepeatModeler* (Smit & Hubley, 2008) (version v2.0.4) to identify repeat elements in the *Clogmia albipunctata* genome and *RepeatMasker*'s (Smit et al., 2013) (version v4.1.5) *fambd.py* script to extract *Arthropoda* records from the *Dfam* database. These sequences were combined to generate a custom repeat library, which was then used to soft-mask the genome. To predict gene structures, we used RNA-seq and protein evidence. mRNA data from 26 individual embryos of various stages, pooled samples of first and fourth-instar larvae, one female and one male pupa generated in this study were mapped to the genome (**Table S2.1**). Additionally, 11 RNA samples (9 embryonic and two adult) from other studies were downloaded from NCBI (**S1 Table**). For protein evidence, we downloaded all available Dipteran sequences from NCBI's RefSeq (January 2024) and UniProt's complete protein sequences (Release 2023_04). A detailed description of the annotation pipeline is available elsewhere (Tenger-Trolander et al., 2025). Briefly, we assembled and mapped the RNA transcripts and protein sequences to the reference and used these data as evidence of gene structures and to predict gene models. *Evidence Modeler* (Haas et al., 2008) (EVM, version 2.1.0) was used to create a consensus gene set, which was further refined with Program to Assemble Spliced Alignments' (PASA version 2.5.3) to add UTRs and identify alternative transcripts. Gene structures overlapping with repeats and transposable elements (TEs) were filtered out. Functional annotation was performed with *EggNOG-mapper* (Cantalapiedra et al., 2021; Huerta-Cepas et al., 2019) (version 2.1.12), assigning gene names based on orthology. We assessed the final annotation quality using BUSCO (Benchmarking Universal Single-Copy Orthologs) (Manni et al., 2021; Simao et al., 2015) on the predicted protein sequences. Finally, we analyzed synteny between the *C. albipunctata* and *D. melanogaster* genomes using *MCScanX* (Y. Wang et al., 2012) (primary release) with parameters: match_score = 50 (default), match size = 20 (minimum

genes per syntenic block), gap_penalty = -1 (default), and max_gaps = 100. Results were visualized using the *circlize* R package (Gu et al., 2014) (version 4.1.2). Annotations of *Cal-eve1, Cal-eve2, Cal-tll1,* and *Cal-tll2* were manually corrected in the final output file (.gff3) based on RNA-seq data and transcript data available on NCBI.

2.5.17 Transcription factor identification

Any gene in our annotation of the Clogmia genome that contained the Cluster-of-Orthologous-Gene (COG) symbol "K" for transcription or the Gene Ontology (GO) term "0003700" for DNA-binding transcription factor activity was treated as a putative transcription factor gene.

2.6 Author Contributions

Ezra E. Amiri: Conceptualization, Methodology, Data collection and curation, Investigation, Formal analysis (including all differential enrichment analysis), Visualization, Writing – original draft, and Writing – review and editing. Figures 2.2-2.8, Supplemental Figures S2.1-S2.13 and S2.16, Table 2.3, Supplemental Tables -S2.1-S2.5.

Ayse Tenger-Trolander: Developed and implemented the genome annotation pipeline used to annotate the genome of *Clogmia albipunctata*, Visualization, Writing – review and editing. Figure 2.1, Tables 2.1-2.2, Supplemental Table S2.1.

Muzi Li: Performed all injections of Clogmia embryos.

Alexander Thomas Julian: *De novo* transcriptome assembly.

Koray Kasan: Phylogenic analysis, including the analysis *Sloppy-paired* homologs. Supplemental Figures S2.14 and S2.15.

Sheri A. Sanders: Preparation of genomic resources, Writing – review and editing.

Shelby Blythe: Conceptualization, Resources, Funding acquisition, Project administration, Supervision, Writing – original draft, Writing – review and editing.

Urs Schmidt-Ott: Conceptualization, Resources, Funding acquisition, Project administration, Supervision, Writing – original draft, Writing – review and editing.

2.7 Acknowledgments

We thank Lily Shiue and Qianyu Jin at Cantata Bio for assembling the genome, Melissa Harrison for the Cry2-zld mutant Drosophila strain, Natasha Megherea for managing the *Clogmia albipunctata* culture, and Suryadi Mcqueen for the illustrations of *Clogmia albipunctata* and *Drosophila melanogaster*. This work used Jetstream2 at Indiana University through allocation BIO220075 from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program, which is supported by National Science Foundation grants #2138259, #2138286, #2138307, #2137603, and #2138296. Bioinformatic work was supported in part by the Notre Dame University Genomics and Bioinformatics Core Facility.

2.8 Competing interests

The authors declare no competing interests.

2.9 Funding

E. A. was the recipient of predoctoral fellowships from the National Institute of General Medical Sciences (T32 GM139782, T32 GM007183). S.A.B. is a Pew Scholar in the Biomedical Sciences. Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R01 GM127366. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

2.10 Data and Resource Availability

Genome: The annotated *Clogmia albipunctata* genome can be found in NCBI's Genome database under BioProject Accession PRJNA1165226.

RNA-seq and ATAC-seq data: All fastq files containing the raw sequencing reads for each sample have been uploaded to NCBI's Sequence Read Archive (SRA) and can be found under the BioProject Accession PRJNA1198932. Individual BioSample and SRA accession numbers for the annotation of the genome are listed in S1 Table, including those samples that were not generated in this study but used as evidence for the annotation of the genome.

Genome Browser: The genomic data produced in this study can be further interrogated using a genome browser ecosystem centered on JBrowse2 (Diesh et al., 2023) and delivered as a cloud image designed for individual use, currently available on NSF's Jetstream2 cloud platform, with a portable Docker image in development. Jetstream2's user-friendly interface (image: *Clogmia albipunctata* Genome Resources) (Boerner et al., 2023; Hancock et al., 2021) integrates tools for genomic analysis, including BLAST search (SequenceServer2.0) (Priyam et al., 2019), CRISPR guide RNA designer (Beeber & Chain, 2020), R shiny-driven differential gene expression (DGE)

analysis (freecount) (Brooks et al., 2024), and synteny mapping for comparative genomics (ShinySyn) (Xiao & Lam, 2022). Each tool is connected back to the browser, leveraging JBrowse2's ability to dynamically visualize and contextualize the genomic data. BLAST results link directly to genomic regions, enabling users to explore expression tracks, variants, and synteny relationships. CRISPR tools provide sgRNA design with metrics for selection and visualization of target regions alongside expression data, off-target effects, and variants. DGE analysis visualizations are linked back to the genome browser for further exploration. All gene model annotations in the browser connect to common external databases, such as NCBI, Uniprot, FlyBase, and Google Scholar, for further exploring and interpreting the data. Future enhancements will include expanded sgRNA profiling, support for Docker to enable deployment on commercial cloud platforms, and further optimization of workflows, ensuring this ecosystem remains a powerful and accessible solution for the use of our genomic data presented here and elsewhere (Tenger-Trolander et al., 2025) for research and education. To access the *Clogmia albipunctata* genome browser and related tools, create an ACCESS ID and use this ID to create an ACCESS account to log in to Jetstream2 to apply for an 'allocation' of credits that can be used on Jetstream2. Detailed instructions on the use of Jetstream2 can be found at https://jetstream-cloud.org/get-started/index.html. This on-demand model reduces resource costs and security concerns compared to hosting a centralized site. Administrative tools, (e.g. automated reboot scripts, system service management, and load balancing for multi-user groups or workshops) enhance usability while minimizing operational overhead.

2.11 Supplemental Data

2.11.1 Supplemental Tables

Table S2. 1 RNA-seq samples used for annotation.

SRA file ID	Experiment Title	Instr.	Study Title	Sample Accession	Stage	Sex
SRR7132659	RNA-Seq of Clogmia albipunctata: cellular blastoderm stage	Illumina HiSeq 1000	Evolution of an Embryonic Axis Determinant via Alternative Transcription	SRS3270620	Embryo, cellular blastoderm	Unk.
SRR7132660	RNA-Seq of Clogmia albipunctata: early gastrula stage	Illumina HiSeq 1000	Evolution of an Embryonic Axis Determinant via Alternative Transcription	SRS3270619	Embryo, early gastrulation	Unk.
SRR7132661	RNA-Seq of Clogmia albipunctata: maternal stage	Illumina HiSeq 1000	Evolution of an Embryonic Axis Determinant via Alternative Transcription	SRS3270618	Embryo, maternal	Unk.

Table S2.1 RNA-seq samples used for annotation, continued

SRR7 1326 62	RNA-Seq of Clogmia albipunctata: syncytial stage	Illumina HiSeq 1000	Evolution of an Embryonic Axis Determinant via Alternative Transcription	SRS32706 16	Embryo, syncytial blastoderm	Unk.
SRR7 1326 63	RNA-Seq of Clogmia albipunctata: posterior embryo maternal stage replicate 2	Illumina HiSeq 1000	Evolution of an Embryonic Axis Determinant via Alternative Transcription	SRS32706 15	Embryo, maternal	Unk.
SRR7 1326 64	RNA-Seq of Clogmia albipunctata: anterior embryo maternal stage replicate 2	Illumina HiSeq 1000	Evolution of an Embryonic Axis Determinant via Alternative Transcription	SRS32706 17	Embryo, maternal	Unk.
SRR7 1326 65	RNA-Seq of Clogmia albipunctata: posterior embryo maternal stage replicate 1	Illumina HiSeq 1000	Evolution of an Embryonic Axis Determinant via Alternative Transcription	SRS32706 14	Embryo, maternal	Unk.

Table S2.1 RNA-seq samples used for annotation, continued

SRR7132666	RNA-Seq of Clogmia albipunctata: anterior embryo maternal stage replicate 1	Illumina HiSeq 1000	Evolution of an Embryonic Axis Determinant via Alternative Transcription	SRS3270613	Embryo, maternal	Unk.
SRR5559337	C_albipunctata_M_RNAseq	Illumina Genome Analyzer	Illumina sequencing of male and female transcriptomes of B.oleae, C.riparius, C.albipunctata, C.trivittatus, T.oleracea, C.fuscipes, C.patibulatus, M.abdita and S.bullata	SRS2199522	Adult	Male
SRR5559338	C_albipunctata_F_RNAseq	Illumina Genome Analyzer	Illumina sequencing of male and female transcriptomes of B.oleae, C.riparius, C.albipunctata, C.trivittatus, T.oleracea, C.fuscipes, C.patibulatus, M.abdita and S.bullata	SRS2199521	Adult	Fem.

Table S2.1 RNA-seq samples used for annotation, continued

ERR16007 1	RNA-Seq of Clogmia albipunctata embryos (8hrs, 10hrs, 12hrs at 25C)	Illumina HiSeq 2000	Comparative transcriptomics of early dipteran development	ERS158633	Unknown	Unk.
23003FL-09-01-01	Initiation of anterior pattern formation in moth fly embryos by Opa/Zic	Illumina NovaSeq X Plus	This study	SAMN45858185	Embryo, maternal	Unk.
23003FL-09-01-02	Initiation of anterior pattern formation in moth fly embryos by Opa/Zic	Illumina NovaSeq X Plus	This study	SAMN45858186	Embryo, maternal	Unk.
23003FL-09-01-03	Initiation of anterior pattern formation in moth fly embryos by Opa/Zic	Illumina NovaSeq X Plus	This study	SAMN45858187	Embryo, maternal	Unk.

Table S2.1 RNA-seq samples used for annotation, continued

2300 3FL-09-01-04	Initiation of anterior pattern formation in moth fly embryos by Opa/Zic	Illumina NovaSeq X Plus	This study	SAMN45858188	Embryo, maternal	Unk.
2300 3FL-09-01-05	Initiation of anterior pattern formation in moth fly embryos by Opa/Zic	Illumina NovaSeq X Plus	This study	SAMN45858189	Embryo, NC11	Unk.
2300 3FL-09-01-06	Initiation of anterior pattern formation in moth fly embryos by Opa/Zic	Illumina NovaSeq X Plus	This study	SAMN45858190	Embryo, NC11	Unk.
2300 3FL-09-01-07	Initiation of anterior pattern formation in moth fly embryos by Opa/Zic	Illumina NovaSeq X Plus	This study	SAMN45858191	Embryo, NC11	Unk.

Table S2.1 RNA-seq samples used for annotation, continued

2300 3FL-09-01-08	Initiation of anterior pattern formation in moth fly embryos by Opa/Zic	Illumina NovaSeq X Plus	This study	SAMN45858192	Embryo, NC11	Unk.
2300 3FL-09-01-09	Initiation of anterior pattern formation in moth fly embryos by Opa/Zic	Illumina NovaSeq X Plus	This study	SAMN45858196	Embryo, NC12	Unk.
2300 3FL-09-01-10	Initiation of anterior pattern formation in moth fly embryos by Opa/Zic	Illumina NovaSeq X Plus	This study	SAMN45858193	Embryo, NC12	Unk.
2300 3FL-09-01-11	Initiation of anterior pattern formation in moth fly embryos by Opa/Zic	Illumina NovaSeq X Plus	This study	SAMN45858194	Embryo, NC12	Unk.

Table S2.1 RNA-seq samples used for annotation, continued

2300 3FL-09-01-12	Initiation of anterior pattern formation in moth fly embryos by Opa/Zic	Illumina NovaSeq X Plus	This study	SAMN45858195	Embryo, NC12	Unk.
2300 3FL-09-01-13	Initiation of anterior pattern formation in moth fly embryos by Opa/Zic	Illumina NovaSeq X Plus	This study	SAMN45858197	Embryo, NC13	Unk.
2300 3FL-09-01-14	Initiation of anterior pattern formation in moth fly embryos by Opa/Zic	Illumina NovaSeq X Plus	This study	SAMN45858198	Embryo, NC13	Unk.
2300 3FL-09-01-15	Initiation of anterior pattern formation in moth fly embryos by Opa/Zic	Illumina NovaSeq X Plus	This study	SAMN45858199	Embryo, NC13	Unk.

Table S2.1 RNA-seq samples used for annotation, continued

2300 3FL-09-01-16	Initiation of anterior pattern formation in moth fly embryos by Opa/Zic	Illumina NovaSeq X Plus	This study	SAMN45858200	Embryo, NC13	Unk.
2300 3FL-09-01-17	Initiation of anterior pattern formation in moth fly embryos by Opa/Zic	Illumina NovaSeq X Plus	This study	SAMN45858201	Embryo, NC14	Unk.
2300 3FL-09-01-18	Initiation of anterior pattern formation in moth fly embryos by Opa/Zic	Illumina NovaSeq X Plus	This study	SAMN45858202	Embryo, NC14	Unk.
2300 3FL-09-01-19	Initiation of anterior pattern formation in moth fly embryos by Opa/Zic	Illumina NovaSeq X Plus	This study	SAMN45858203	Embryo, NC14	Unk.

Table S2.1 RNA-seq samples used for annotation, continued

2300 3FL-09-01-20	Initiation of anterior pattern formation in moth fly embryos by Opa/Zic	Illumina NovaSeq X Plus	This study	SAMN45858204	Embryo, NC14	Unk.
2300 3FL-09-01-21	Initiation of anterior pattern formation in moth fly embryos by Opa/Zic	Illumina NovaSeq X Plus	This study	SAMN45858205	Embryo, Late NC14	Unk.
2300 3FL-09-01-22	Initiation of anterior pattern formation in moth fly embryos by Opa/Zic	Illumina NovaSeq X Plus	This study	SAMN45858206	Embryo, Late NC14	Unk.
2300 3FL-09-01-23	Initiation of anterior pattern formation in moth fly embryos by Opa/Zic	Illumina NovaSeq X Plus	This study	SAMN45858207	Embryo, Late NC14	Unk.

Table S2.1 RNA-seq samples used for annotation, continued

2300 3FL-09-01-24	Initiation of anterior pattern formation in moth fly embryos by Opa/Zic	Illumina NovaSeq X Plus	This study	SAMN45858208	Embryo, Late NC14	Unk.
URS-EA-01	Initiation of anterior pattern formation in moth fly embryos by Opa/Zic	Illumina NovaSeq 6000	This study	SAMN45858209	Embryos, Stage 13	Fem. & Male
URS-EA-02	Initiation of anterior pattern formation in moth fly embryos by Opa/Zic	Illumina NovaSeq 6000	This study	SAMN45858210	Embryos, Stage 15	Fem. & Male
URS-EA-03	Initiation of anterior pattern formation in moth fly embryos by Opa/Zic	Illumina NovaSeq 6000	This study	SAMN45858211	Larvae, L1	Fem. & Male

Table S2.1 RNA-seq samples used for annotation, continued

URS-EA-04	Initiation of anterior pattern formation in moth fly embryos by Opa/Zic	Illumina NovaSeq 6000	This study	SAMN45858212	Larvae, L4	Fem. & Male
URS-EA-05	Initiation of anterior pattern formation in moth fly embryos by Opa/Zic	Illumina NovaSeq 6000	This study	SAMN45858213	Pupa	Fem.
URS-EA-06	Initiation of anterior pattern formation in moth fly embryos by Opa/Zic	Illumina NovaSeq 6000	This study	SAMN45858214	Pupa	Male

Table S2. 2 Homologs of early segmentation genes in *Clogmia albipunctata*.

Gene name	Gene sym.	FlybaseID	Clogmia homolog	Cal. gene symb.	Ref.
hunchback	*hb*	FBgn0001180	evm.TU.scaffold_7.1783	*Cal-hb1*	AJ131041 (Rohr et al. 1999)
			evm.TU.scaffold_7.1785	*Cal-hb2*	This study
giant	*gt*	FBgn0001150	evm.TU.scaffold_1.1986	*Cal-gt*	GU137318 (Garcia Solache et al. 2010)
knirps	*kni*	FBgn0001320	None	N. A.	NA
knirps-like	*knrl*	FBgn0001323	evm.TU.scaffold_7.729	*Cal-kni/knrl 1*	GU137321 (Garcia Solache et al. 2010)
			evm.TU.scaffold_7.724	*Cal-kni/knrl 2*	This study
Krüppel	*Kr*	FBgn0001325	evm.TU.scaffold_7.1074	*Cal-Kr*	GU137323 (Garcia Solache et al. 2010) and AJ131042 (Rohr et al. 1999)

Table S2.2 Homologs of early segmentation genes in *Clogmia albipunctata*, continued

empty spiracles	*ems*	FBgn0000576	evm.TU.scaffold_4.1375	*Cal-ems1*	This study
			evm.TU.scaffold_4.1371	*Cal-ems2*	This study
			evm.TU.scaffold_4.1372	*Cal-ems3*	This study
			evm.TU.scaffold_4.1374	*Cal-ems4*	This study
buttonhead	*btd*	FBgn0000233	evm.TU.scaffold_5.1944	*Cal-btd*	This study
orthodenticle/ ocelliless	*otd/ oc*	FBgn0004102	evm.TU.scaffold_5.235	*Cal-otd1*	MN122108 (Yoon et al. 2019)
			evm.TU.scaffold_5.236	*Cal-otd2*	This study
huckebein	*hkb*	FBgn0261434	evm.TU.scaffold_2.1090	*Cal-hkb*	GU137311 (Garcia Solache et al. 2010)
tailless	*tll*	FBgn0003720	evm.TU.scaffold_3.2548a	*Cal-tll1*	GU137320 (Garcia Solache et al. 2010)

Table S2.2 Homologs of early segmentation genes in *Clogmia albipunctata*, continued

			evm.TU.scaffold_3.2548b	*Cal-tll2*	This study
hairy	*hry*	FBgn0001168	evm.TU.scaffold_4.1097	*Cal-hry*	AY645026 (Bullock et al. 2004)
even-skipped	*eve*	FBgn0000606	evm.TU.scaffold_4.40a	*Cal-eve1*	AY645024 (Bullock et al. 2004) and GU137324 (Garcia Solache et al. 2010)
			evm.TU.scaffold_4.40b	*Cal-eve2*	AY645025 (Bullock et al. 2004)
runt	*run*	FBgn0003300	evm.TU.scaffold_2.1058	*Cal-run*	This study
fushi tarazu	*ftz*	FBgn0001077	evm.TU.scaffold_3.2936	*Cal-ftz*	This study
odd-skipped	*odd*	FBgn0002985	evm.TU.scaffold_2.281	*Cal-odd*	This study
paired	*prd*	FBgn0003145	evm.TU.scaffold_2.1785	*Cal-prd*	This study

Table S2.2 Homologs of early segmentation genes in *Clogmia albipunctata*, continued

sloppy-paired 1	*slp1*	FBgn0003430	None	N. A.	N.A.
sloppy-paired 2	*slp2*	FBgn0004567	evm.TU.scaffold_5.983	*Cal-slp1*	This study
			evm.TU.scaffold_5.985	*Cal-slp2*	MN122106 (Yoon et al. 2019)
			evm.TU.scaffold_5.987	*Cal-slp3*	This study
forkhead domain 19B	*fd19B*	FBgn0031086	None	N. A.	N. A.
caudal	*cad*	FBgn0000251	evm.TU.scaffold_5.401	*Cal-cad1*	Rivera-Pomar et al.1996
caudal	*cad*	FBgn0000251	evm.TU.scaffold_5.403	*Cal-cad2*	This study
odd-paired	*opa*	FBgn0003002	evm.TU.scaffold_7.878	*Cal-opa*	MN122104 (Yoon et al. 2019)
Dichaete	*D*	FBgn0000411	evm.TU.scaffold_1.1497	*Cal-D*	This study

Table S2.2 Homologs of early segmentation genes in *Clogmia albipunctata*, continued

zerknüllt	*zen*	FBgn0004053	evm.TU.scaffold_7.614	*Cal-zen*	AJ419659 (Stauber et al. 2002)
homeobrain	*hbn*	FBgn0008636	evm.TU.scaffold_4.1647	*Cal-hbn*	This study

Table S2. 3 *Cal-zld* RT-qPCR data.
Fold change calculated using the ΔΔCt method. Reference gene *Cal-rpl35*. Target gene *Cal-zld*. Percent knockdown was calculated as 100-(FC*100).

Treatment	Embryo No	Fold change	Percent knockdown
Alignment control	1	1.04427378	-4.4273782
Alignment control	2	0.95760328	4.23967193
Cal-zld RNAi	3	1.92919637	-92.919637
Cal-zld RNAi	4	1.1011416	-10.11416
Cal-zld RNAi	5	1.61440215	-61.440215
Cal-zld RNAi	6	0.88300897	11.6991029
Cal-zld RNAi	7	1.30450276	-30.450276
Cal-zld RNAi	8	0.82188019	17.8119813
Alignment control	9	0.8630404	13.69596
Alignment control	10	1.15869431	-15.869431
Cal-zld RNAi	11	2.17723933	-117.72393
Cal-zld RNAi	12	0.4735209	52.6479103
Cal-zld RNAi	13	0.06091748	93.9082523
Cal-zld RNAi	14	1.63580412	-63.580412
Cal-zld RNAi	15	2.08132174	-108.13217
Cal-zld RNAi	16	0.68349373	31.6506274

Table S2.3 *Cal-zld* RT-qPCR data, continued

Alignment control	17	0.98043915	1.95608492
Alignment control	18	1.01995111	-1.995111
Cal-zld RNAi	19	1.11457987	-11.457987
Cal-zld RNAi	20	0.51245598	48.7544016
Cal-zld RNAi	21	1.1045816	-10.45816
Cal-zld RNAi	22	1.27059126	-27.059126
Cal-zld RNAi	23	5.13014641	-413.01464
Cal-zld RNAi	24	0.5913155	40.86845
Cal-zld RNAi	25	1.80562713	-80.562713
Cal-zld RNAi	26	0.96392981	3.60701923

Table S2. 4 Lists of transcription factor genes at NC12 and NC13 that are sensitive to *Cal-opa* knockdown.
1) Differentially expressed transcription factor genes (DE TFs) containing Opa motif positive ATAC-seq peaks (potential direct Cal-Opa targets). 2) Differentially expressed transcription factor genes with *Cal-opa* sensitive ATAC-seq peaks. 3) Differentially expressed transcription factor genes with *Cal-opa* sensitive ATAC-seq peaks containing an Opa motif. Gene model name is listed in addition to a flybaseID if a Drosophila homolog was identified from a BLASTp search. Genes with an asterisk (*) are also identified in the *Cal-opa* RNAi versus injection control DESeq2 analysis.

NC12				
DE TFs with Opa+ ATAC-seq peaks	evm.TU.scaffold_2.1802	evm.TU.scaffold_3.1882(FBgn0039270)	evm.TU.scaffold_4.1647(Cal-hbn) *	evm.TU.scaffold_5.1243(FBgn0031573)
	evm.TU.scaffold_5.150(FBgn0035160)	evm.TU.scaffold_5.184(FBgn0013753)	evm.TU.scaffold_5.983(Cal-slp1) *	evm.TU.scaffold_5.985(Cal-slp2) *
	evm.TU.scaffold_7.2949(FBgn0037491)	evm.TU.scaffold_7.878(Cal-opa) *		
DE TFs with *Cal-opa* sensitive ATAC-seq peaks	evm.TU.scaffold_4.1647(Cal-hbn) *	evm.TU.scaffold_5.983(Cal-slp1) *	evm.TU.scaffold_5.985(Cal-slp2)	evm.TU.scaffold_7.878(Cal-opa) *
DE TFs with Opa+ *Cal-opa* sensitive ATAC-seq peaks	evm.TU.scaffold_4.1647(Cal-hbn) *	evm.TU.scaffold_5.983(Cal-slp1) *	evm.TU.scaffold_5.985(Cal-slp2)	evm.TU.scaffold_7.878(Cal-opa) *
NC13				

Table S2.4, continued

DE TFs with Opa+ ATAC-seq peaks	evm.TU.scaffold_1.1194(FBgn0002735)	evm.TU.scaffold_1.1262	evm.TU.scaffold_1.134	evm.TU.scaffold_1.1422
	evm.TU.scaffold_1.1497(Cal-D)	evm.TU.scaffold_1.1676(FBgn0010417)	evm.TU.scaffold_1.182(FBgn0043002)	evm.TU.scaffold_1.2409(FBgn0011715)
	evm.TU.scaffold_1.2511(FBgn0005427)	evm.TU.scaffold_1.2651	evm.TU.scaffold_1.2674(FBgn0263757)	evm.TU.scaffold_1.2960(FBgn0086655)
	evm.TU.scaffold_1.348	evm.TU.scaffold_1.824(FBgn0002573)	evm.TU.scaffold_2.1031(FBgn0010422)	evm.TU.scaffold_2.1246(FBgn0003687)
	evm.TU.scaffold_2.1413(FBgn0031776)	evm.TU.scaffold_2.1535(FBgn0037363)	evm.TU.scaffold_2.1830	evm.TU.scaffold_2.2083
	evm.TU.scaffold_2.250(FBgn0266557)	evm.TU.scaffold_2.275(FBgn0028398)	evm.TU.scaffold_2.751(FBgn0004606)	evm.TU.scaffold_2.825
	evm.TU.scaffold_2.935(FBgn0032016)	evm.TU.scaffold_3.1271(FBgn0038499)	evm.TU.scaffold_3.1502(FBgn0004897)	evm.TU.scaffold_3.1619(FBgn0003612)
	evm.TU.scaffold_3.1882(FBgn0039270)	evm.TU.scaffold_3.2548b(Cal-tll2)	evm.TU.scaffold_3.2623(FBgn0001235)	evm.TU.scaffold_3.2809(FBgn0262127)
	evm.TU.scaffold_3.514(FBgn0038903)	evm.TU.scaffold_3.913	evm.TU.scaffold_4.1248	evm.TU.scaffold_4.1328

Table S2.4, continued

	evm.TU.scaffold_4.1461(FBgn0033540)	evm.TU.scaffold_4.1647(Cal-hbn) *	evm.TU.scaffold_4.1673(FBgn0042085)	evm.TU.scaffold_4.1700(FBgn0032329)
	evm.TU.scaffold_4.1897	evm.TU.scaffold_4.2102(FBgn0031822)	evm.TU.scaffold_4.2428	evm.TU.scaffold_4.40b(Cal-eve2)
	evm.TU.scaffold_4.667(FBgn0261934)	evm.TU.scaffold_4.874(FBgn0027549)	evm.TU.scaffold_4.881(FBgn0014143)	evm.TU.scaffold_4.918(FBgn0040273)
	evm.TU.scaffold_4.958(FBgn0003415)	evm.TU.scaffold_5.1243(FBgn0031573)	evm.TU.scaffold_5.1294(FBgn0261617)	evm.TU.scaffold_5.1512(FBgn0035145)
	evm.TU.scaffold_5.175(FBgn0004859)	evm.TU.scaffold_5.1857(FBgn0036581)	evm.TU.scaffold_5.2222(FBgn0002931)	evm.TU.scaffold_5.235(Cal-otd1)
	evm.TU.scaffold_5.842(FBgn0086447)	evm.TU.scaffold_5.983(Cal-slp1) *	evm.TU.scaffold_5.985(Cal-slp2) *	evm.TU.scaffold_7.1036(FBgn0024841)
	evm.TU.scaffold_7.1074(Cal-Kr)	evm.TU.scaffold_7.1127(FBgn0014037)	evm.TU.scaffold_7.1185(FBgn0037085)	evm.TU.scaffold_7.1252(FBgn0023518)
	evm.TU.scaffold_7.1663(FBgn0038306)	evm.TU.scaffold_7.1925(FBgn0027529)	evm.TU.scaffold_7.228	evm.TU.scaffold_7.2949(FBgn0037491)
	evm.TU.scaffold_7.53(FBgn0004396) *	evm.TU.scaffold_7.614(Cal-zen)	evm.TU.scaffold_7.878(Cal-opa)	evm.TU.scaffold_7.993(FBgn0051950)

Table S2.4, continued

DE TFs with *Cal-opa* sensitive ATAC-seq peaks	evm.TU.scaffold_1.1497(Cal-D)	evm.TU.scaffold_1.2511(FBgn0005427)	evm.TU.scaffold_3.2623(FBgn0001235)	evm.TU.scaffold_4.1328
	evm.TU.scaffold_4.1449(FBgn0037834)	evm.TU.scaffold_4.1647(Cal-hbn) *	evm.TU.scaffold_4.2497	evm.TU.scaffold_5.235(Cal-otd1)
	evm.TU.scaffold_5.983(Cal-slp1) *	evm.TU.scaffold_5.985(Cal-slp2) *	evm.TU.scaffold_5.987(Cal-slp3)	evm.TU.scaffold_7.1074(Cal-Kr)
	evm.TU.scaffold_7.1783(Cal-hb1)	evm.TU.scaffold_7.1785(Cal-hb2) *	evm.TU.scaffold_7.878(Cal-opa)	
DE TFs with Opa+ *Cal-opa* sensitive ATAC-seq peaks	evm.TU.scaffold_1.1497(Cal-D)	evm.TU.scaffold_4.1328	evm.TU.scaffold_4.1647(Cal-hbn) *	evm.TU.scaffold_5.983(Cal-slp1) *
	evm.TU.scaffold_5.985(Cal-slp2) *	evm.TU.scaffold_7.1074(Cal-Kr)	evm.TU.scaffold_7.878(Cal-opa)	

Table S2. 5 *Cal-opa* RT-qPCR data.
Fold change calculated using the ΔΔCt method. Reference gene *Cal-rpl35*. Target gene *Cal-opa*. Percent knockdown was calculated as 100-(FC*100). Indicated is whether the RNA was isolated immediately or in parallel with an ATAC-seq library preparation.

Treatment	**Embryo No**	**Fold change**	**Percent knockdown**	**RNA isolation Method**
Alignment control	1	0.90093817	9.90618329	Standard RNA isolation
Cal-opa RNAi	2	0.73739012	26.2609876	Standard RNA isolation
Injection control	3	0.90469287	9.53071341	RNA isolated from ATAC-seq library prep
Alignment control	4	1.10995409	-10.995409	Standard RNA isolation
Cal-opa RNAi	5	0.54961756	45.0382441	Standard RNA isolation
Cal-opa RNAi	6	0.97857206	2.14279379	Standard RNA isolation
Cal-opa RNAi	7	1.08729996	-8.729996	Standard RNA isolation
Injection control	8	0.704416	29.55838	RNA isolated from ATAC-seq library prep

Table S2.5 *Cal-opa* RT-qPCR data, continued

Injection control	9	0.661356	33.86435	RNA isolated from ATAC-seq library prep
Alignment control	10	1.0085272	-0.8527204	Standard RNA isolation
Alignment control	11	0.99154489	0.84551058	Standard RNA isolation
Cal-opa RNAi	12	5.13815353	-413.81535	Standard RNA isolation
Cal-opa RNAi	13	0.57345413	42.6545869	Standard RNA isolation
Cal-opa RNAi	14	0.18884179	81.1158206	Standard RNA isolation
Cal-opa RNAi	15	2.14094836	-114.09484	Standard RNA isolation
Injection control	16	1.09619161	-9.6191612	RNA isolated from ATAC-seq library prep
Injection control	17	0.56487361	43.5126393	RNA isolated from ATAC-seq library prep
Alignment control	18	0.96845036	3.1549645	Standard RNA isolation

Table S2.5 *Cal-opa* RT-qPCR data, continued

Cal-opa RNAi	19	2.07663836	-107.66384	Standard RNA isolation
Cal-opa RNAi	20	0.63650776	36.3492244	Standard RNA isolation
Cal-opa RNAi	21	1.95782254	-95.782254	Standard RNA isolation
Cal-opa RNAi	22	1.23306598	-23.306598	Standard RNA isolation
Injection control	23	1.30359886	-30.359886	RNA isolated from ATAC-seq library prep
Injection control	24	0.8411879	15.8812102	RNA isolated from ATAC-seq library prep
Alignment control	25	1.03257745	-3.2577452	Standard RNA isolation
Alignment control	26	0.73866903	26.1330968	Standard RNA isolation
Cal-opa RNAi	27	0.07282108	92.7178917	Standard RNA isolation
Cal-opa RNAi	28	0.1320357	86.7964297	Standard RNA isolation
Cal-opa RNAi	29	0.25578437	74.4215633	Standard RNA isolation

Table S2.5 *Cal-opa* RT-qPCR data, continued

Cal-opa RNAi	30	0.03349292	96.6507079	Standard RNA isolation
Cal-opa RNAi	31	0.19895306	80.1046939	Standard RNA isolation
Cal-opa RNAi	32	0.63683874	36.3161262	Standard RNA isolation
Injection control	33	2.2137618	-121.37618	Standard RNA isolation
Injection control	34	1.93790789	-93.790789	Standard RNA isolation
Injection control	35	2.34323232	-134.32323	Standard RNA isolation
Alignment control	36	1.35378628	-35.378628	Standard RNA isolation
Alignment control	37	0.12904982	87.0950183	Standard RNA isolation
Alignment control	38	0.84177117	15.8228833	Standard RNA isolation
Cal-opa RNAi	39	0.23882073	76.1179266	Standard RNA isolation
Cal-opa RNAi	40	0.26370558	73.6294421	Standard RNA isolation
Cal-opa RNAi	41	0.18313761	81.6862391	Standard RNA isolation

Table S2.5 *Cal-opa* RT-qPCR data, continued

Cal-opa RNAi	42	0.79802177	20.1978235	Standard RNA isolation
Cal-opa RNAi	43	0.30418059	69.581941	Standard RNA isolation
Injection control	44	1.77727468	-77.727468	Standard RNA isolation
Injection control	45	1.90087896	-90.087896	Standard RNA isolation
Injection control	46	1.22235782	-22.235782	Standard RNA isolation
Cal-opa RNAi	47	0.99343673	0.65632653	Standard RNA isolation
Cal-opa RNAi	48	0.32522273	67.4777269	Standard RNA isolation
Alignment control	49	0.99584975	0.41502469	Standard RNA isolation
Alignment control	50	1.19292224	-19.292224	Standard RNA isolation
Cal-opa RNAi	51	0.118517	88.14829	Standard RNA isolation
Cal-opa RNAi	52	0.88403	11.59703	Standard RNA isolation
Alignment control	53	0.952418	4.758208	Standard RNA isolation

Table S2.5 *Cal-opa* RT-qPCR data, continued

Alignment control	54	0.989428	1.057198	Standard RNA isolation
Cal-opa RNAi	55	0.007323	99.26769	Standard RNA isolation
Cal-opa RNAi	56	0.028206	97.17936	Standard RNA isolation
Cal-opa RNAi	57	5.637934	-463.793	Standard RNA isolation
Cal-opa RNAi	58	0.20537512	79.4624879	Standard RNA isolation
Cal-opa RNAi	59	0.13395621	86.6043793	Standard RNA isolation
Cal-opa RNAi	60	0.52443408	47.5565925	Standard RNA isolation
Cal-opa RNAi	61	0.43719031	56.2809689	Standard RNA isolation
Cal-opa RNAi	62	0.6726059	32.7394103	Standard RNA isolation
Alignment control	63	1.061178	-6.1178	Standard RNA isolation
Alignment control	64	0.81356715	18.643285	Standard RNA isolation

Table S2.5 *Cal-opa* RT-qPCR data, continued

Cal-opa RNAi	65	0.03331926	96.6680744	RNA isolated from ATAC-seq library prep
Cal-opa RNAi	66	0.001725	99.82754	RNA isolated from ATAC-seq library prep
Cal-opa RNAi	67	0.002495	99.75054	RNA isolated from ATAC-seq library prep
Cal-opa RNAi	68	0.758384	24.16162	RNA isolated from ATAC-seq library prep
Cal-opa RNAi	69	0.0045529	99.5447105	RNA isolated from ATAC-seq library prep
Cal-opa RNAi	70	0.83393104	16.6068956	RNA isolated from ATAC-seq library prep
Cal-opa RNAi	71	0.85916075	14.0839245	RNA isolated from ATAC-seq library prep

Table S2.5 *Cal-opa* RT-qPCR data, continued

Cal-opa RNAi	72	0.00488306	99.5116944	RNA isolated from ATAC-seq library prep
Cal-opa RNAi	73	1.63240609	-63.240609	RNA isolated from ATAC-seq library prep
Alignment control	74	0.90114635	9.88536479	Standard RNA isolation
Alignment control	75	1.40915795	-40.915795	Standard RNA isolation
Alignment control	76	1	0	RNA isolated from ATAC-seq library prep
Alignment control	77	1	0	RNA isolated from ATAC-seq library prep
Alignment control	78	0.9862327	1.37672955	RNA isolated from ATAC-seq library prep
Alignment control	79	1.01395948	-1.395948	RNA isolated from ATAC-seq library prep

Table S2.5 *Cal-opa* RT-qPCR data, continued

Cal-opa RNAi	80	0.78024548	21.975452	RNA isolated from ATAC-seq library prep
Cal-opa RNAi	81	0.00900848	99.0991516	RNA isolated from ATAC-seq library prep
Cal-opa RNAi	82	0.00221036	99.7789636	RNA isolated from ATAC-seq library prep
Cal-opa RNAi	83	11.6924066	-1069.2407	Standard RNA isolation
Injection control	84	0.882703	11.7297004	Standard RNA isolation
Alignment control	85	1	0	RNA isolated from ATAC-seq library prep
Alignment control	86	0.67618973	32.3810275	Standard RNA isolation
Alignment control	87	1.47887488	-47.887488	Standard RNA isolation
Cal-opa RNAi	88	1.40785622	-40.785622	Standard RNA isolation

Table S2.5 *Cal-opa* RT-qPCR data, continued

Cal-opa RNAi	89	0.3023937	69.7606298	Standard RNA isolation
Cal-opa RNAi	90	0.70686176	29.3138241	Standard RNA isolation
Cal-opa RNAi	91	0.25836812	74.1631875	Standard RNA isolation
Cal-opa RNAi	92	0.64104609	35.8953907	Standard RNA isolation
Injection control	93	1.490883	-49.0883	Standard RNA isolation
Injection control	94	1.85339	-85.339	Standard RNA isolation
Injection control	95	2.033784	-103.378	Standard RNA isolation
Alignment control	96	0.86314011	13.6859892	Standard RNA isolation
Alignment control	97	1.58190906	-58.190906	Standard RNA isolation
Alignment control	98	0.7323812	26.7618798	Standard RNA isolation
Cal-opa RNAi	99	1.43561278	-43.561278	Standard RNA isolation
Cal-opa RNAi	100	0.24496825	75.5031752	Standard RNA isolation

Table S2.5 *Cal-opa* RT-qPCR data, continued

Cal-opa RNAi	101	1.04367076	-4.367076	Standard RNA isolation
Cal-opa RNAi	102	1.46984769	-46.984769	Standard RNA isolation
Cal-opa RNAi	103	1.28847689	-28.847689	Standard RNA isolation
Injection control	104	1.5613926	-56.13926	Standard RNA isolation
Injection control	105	1.639588	-63.9588	Standard RNA isolation
Injection control	106	2.34458622	-134.45862	Standard RNA isolation
Injection control	107	0.99711605	0.28839467	Standard RNA isolation

2.11.2 Supplemental Figures

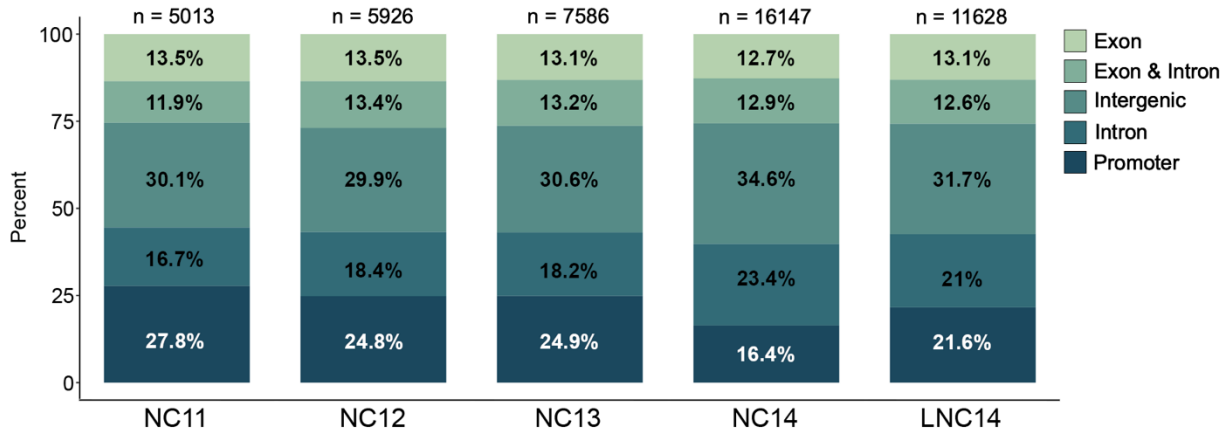

Figure S2. 1 Genomic distribution of ATAC-seq peaks at nuclear cycles NC11, NC12, NC13, NC14 and Late NC14.

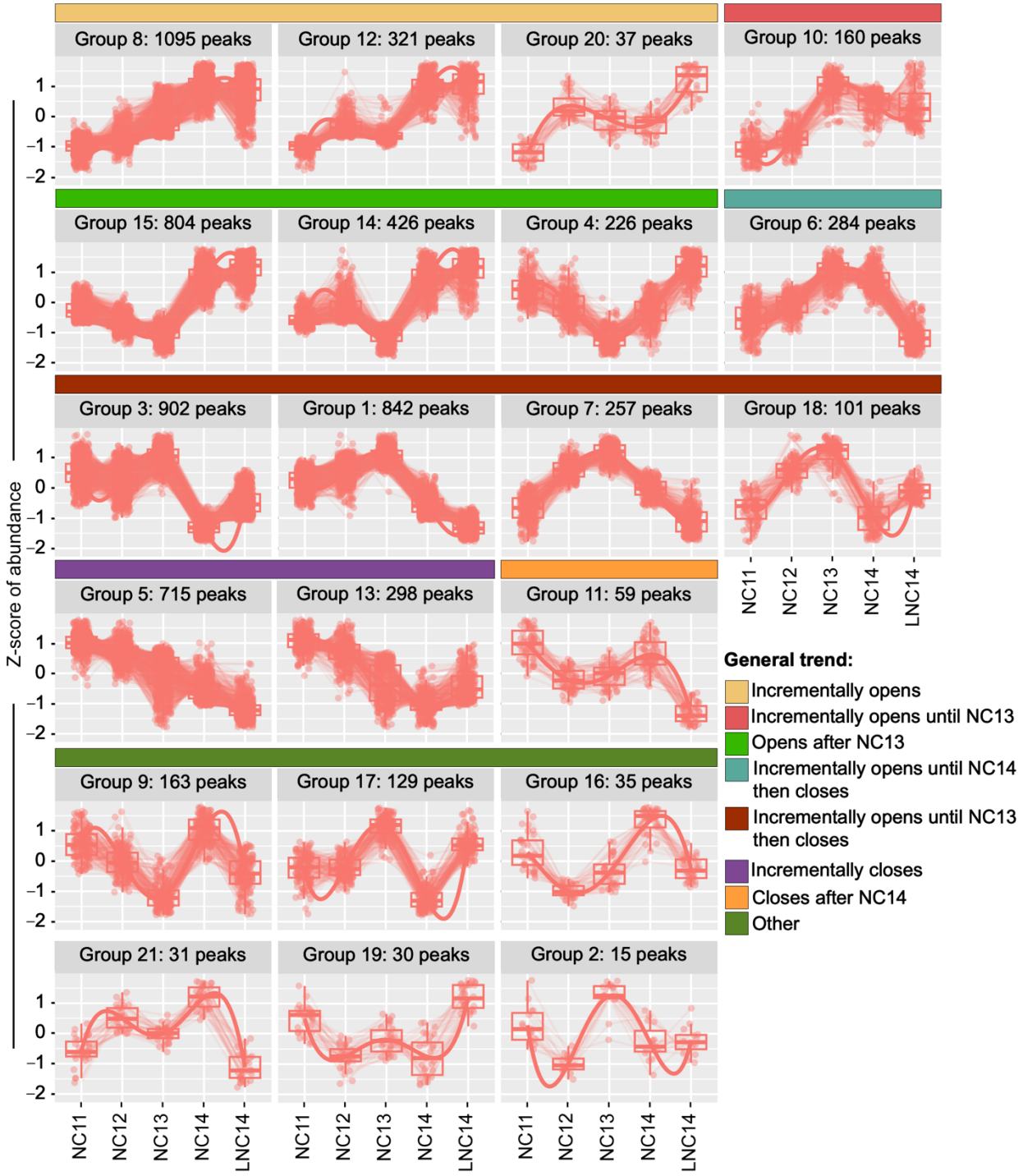

Figure S2. 2 Dynamic ATAC-seq peak groups.
Dynamic peak groups were generated using DEG reports (Pantano, 2017) (R package version 1.38.5). Y-axis, Z-score of abundance. X-axis, developmental stage. Colored bars mark manually established super-groups with similar features described as follows: Incrementally opens (yellow bar; Groups 8, 12, 20); Incrementally opens until NC13 (red bar; Group 10); Opens after NC13

(green; Groups 15, 14, 4); Incrementally opens until NC14, then closes (blue; Group 6); Incrementally opens until NC13, then closes (red-brown; Groups 3, 1, 7, 18); Incrementally closes (violet; Groups 5, 13); Closes after NC14 (orange; Group 11); Other (olive; Groups 9, 17, 16, 21, 19, 15).

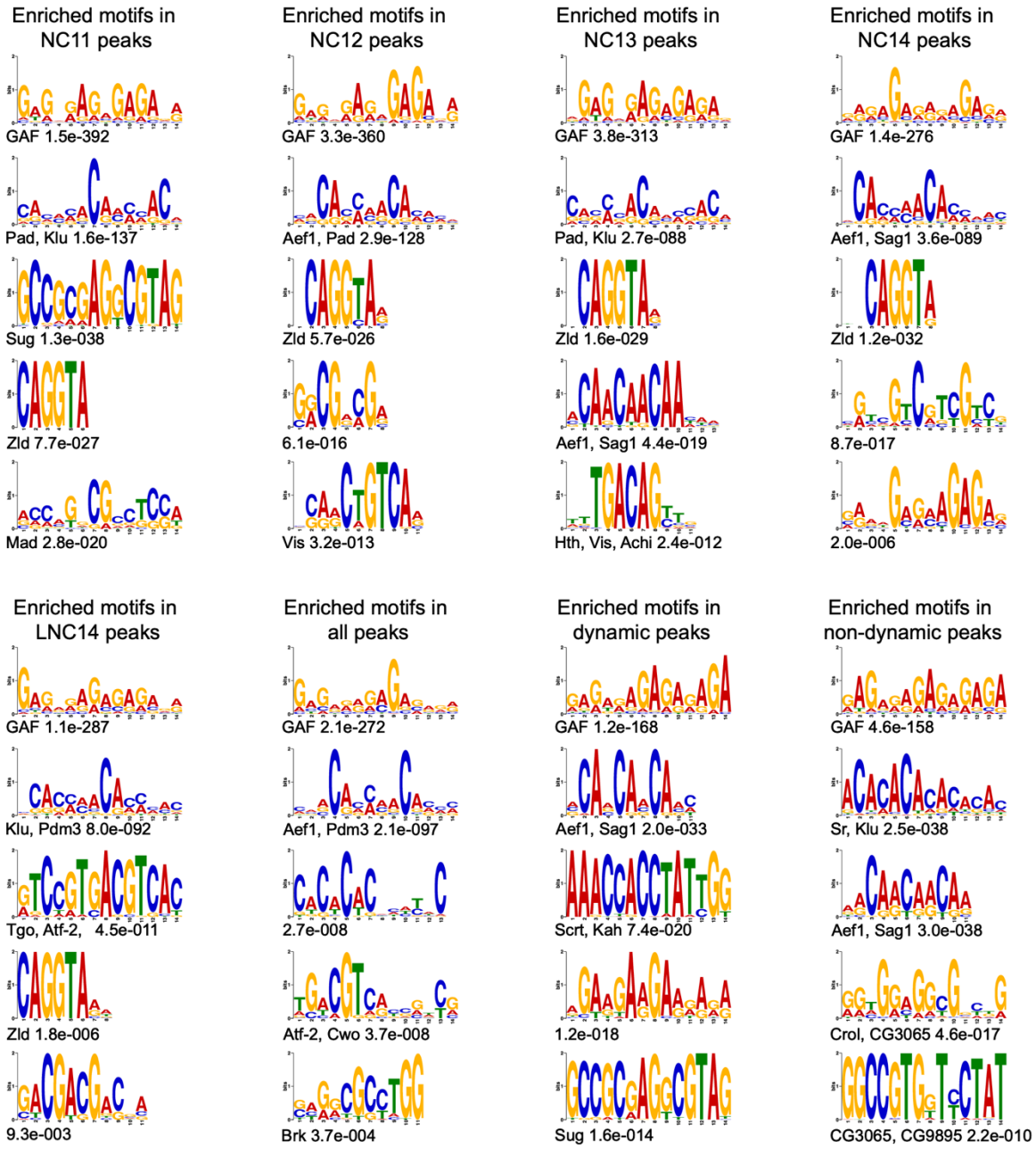

Figure S2. 3 Enriched motifs in ATAC-seq peaks.
Enriched motifs identified by MEME for 8 sets of peaks: NC11 peaks (n = 5013), NC12 peaks (n = 5926), NC13 peaks (n = 7586), NC14 peaks (n = 16147), Late NC14 peaks (n = 11628), all peaks (n = 32157), dynamic peaks (n = 6930), and non-dynamic peaks (right column; n =

25227). The top five motifs discovered by MEME are reported for each group. Motifs are shown as PWM logos with significance (E-value). DNA binding proteins of *Drosophila melanogaster* that bind to these or similar motifs as identified by MEME are indicated. We note that although not identified by MEME, CLAMP has been shown to bind GA-rich motifs and Cg has been shown to bind (CA)n like the motifs above (Kaye et al., 2018; Kuzu et al., 2016; P. Ray et al., 2016; Soruco et al., 2013).

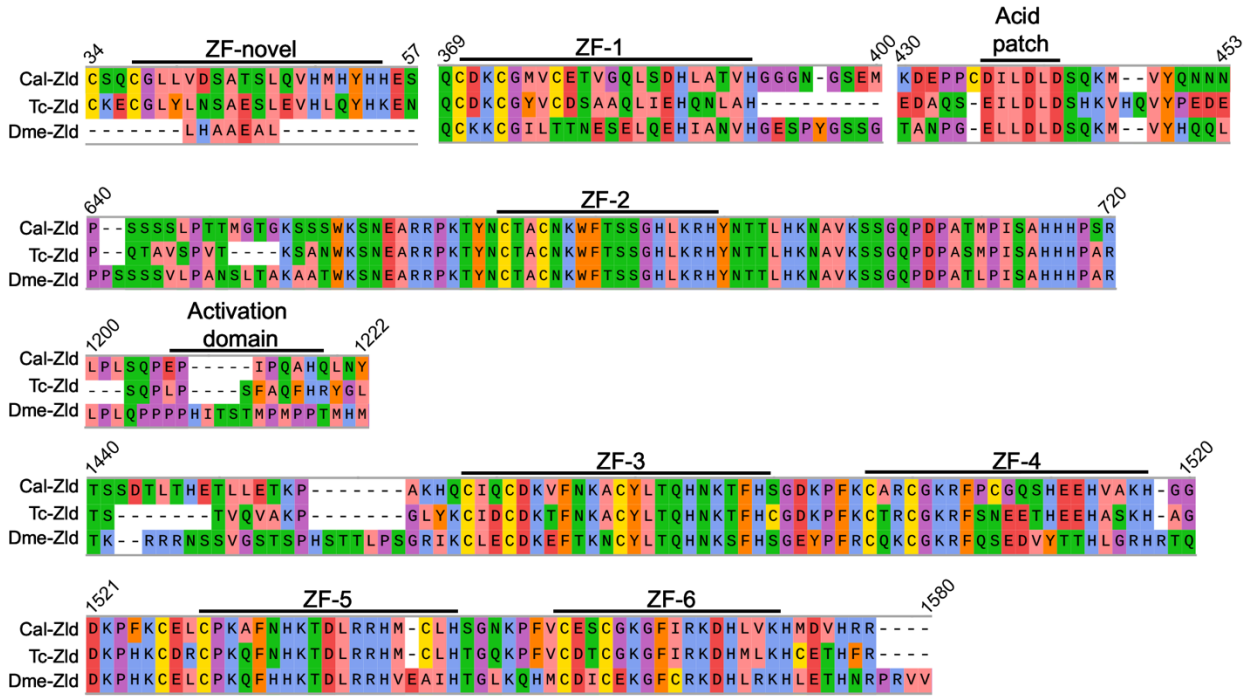

Figure S2. 4 Protein alignment of conserved Zelda domains.
Cal-Zld sequence is shown above Zld sequences of *Tribolium castaneum* (Tc-Zld, XP_001812268.1) and *Drosophila melanogaster* (Dme-Zld, NP_608356.1). Alignment was performed using MAFFT (Katoh et al., 2019) (v7) with the default settings. Amino acid color indicates physio-chemical property (based on Zappo Color Scheme). Important protein domains are indicated (Ribeiro et al., 2017).

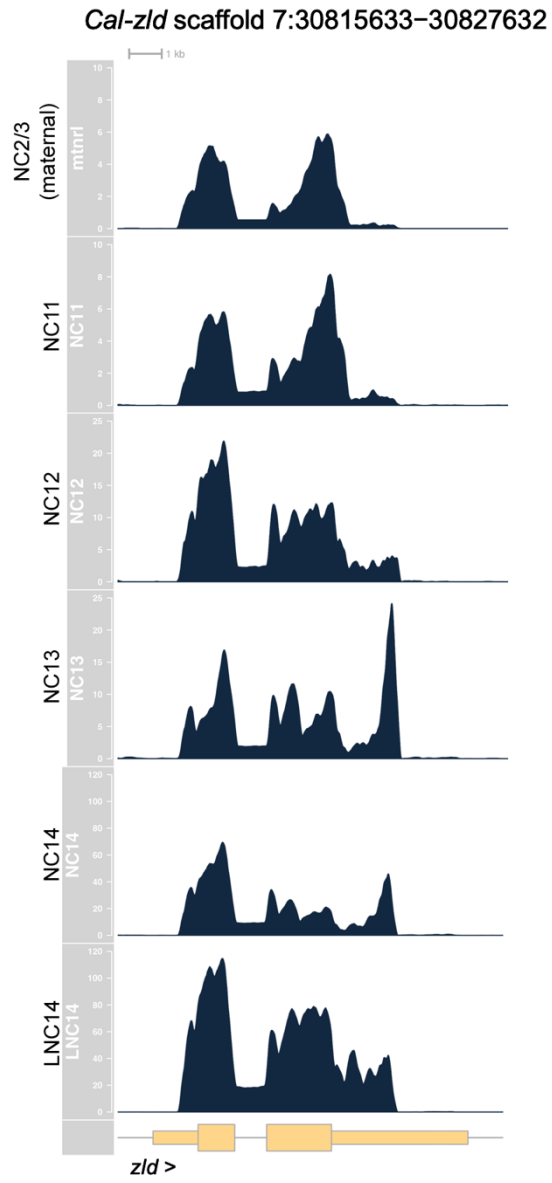

Figure S2. 5 Time-course RNA-seq data of *Cal-zld* at 6 developmental stages.
Gene model in pale yellow. RNA-seq data CPM. Y-axis for NC2/3 (maternal, ~45-minute old embryos) and NC11: 0-10 CPM. Y-axis for NC12 and NC13: 0-25 CPM. Y-axis for NC14 and Late NC14 (LCN14): 0-120 CPM.

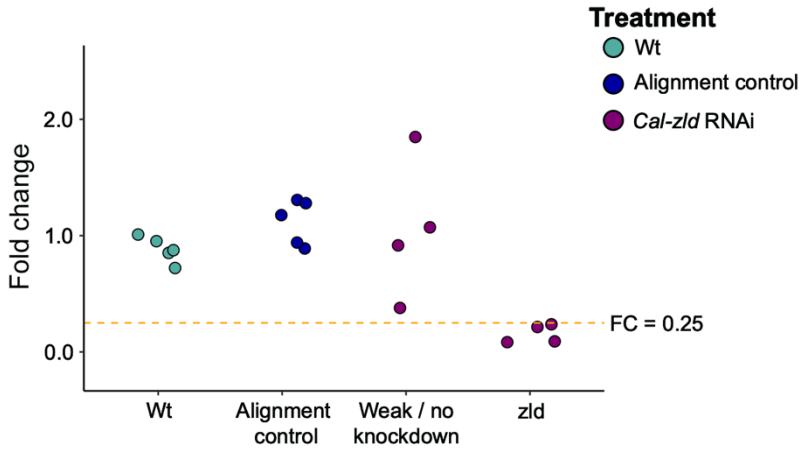

Figure S2. 6 Identification of *Cal-zld* RNAi embryos with strong knockdowns.
Plot of *Cal-zld* fold change for individual embryos. Each embryo is represented by a circle. Embryos are grouped along the x-axis as indicated, spacing within a group was done arbitrarily for visualization. Treatment group of each embryo is indicated by color. Fold change (FC) for each embryo was calculated by dividing the embryo's CPM of *Cal-zld* by the average of wild-type and alignment control *Cal-zld* CPMs. Fold change was calculated using embryos from a single female (same activation batch). If controls from the same batch were lost, fold change was calculated using the average *Cal-zld* CPM of all wild-type and alignment control embryos in the data set. *Cal- zld* RNAi embryos with FC > 0.25 (weak / no knockdown of *Cal-zld*) were excluded from further analysis.

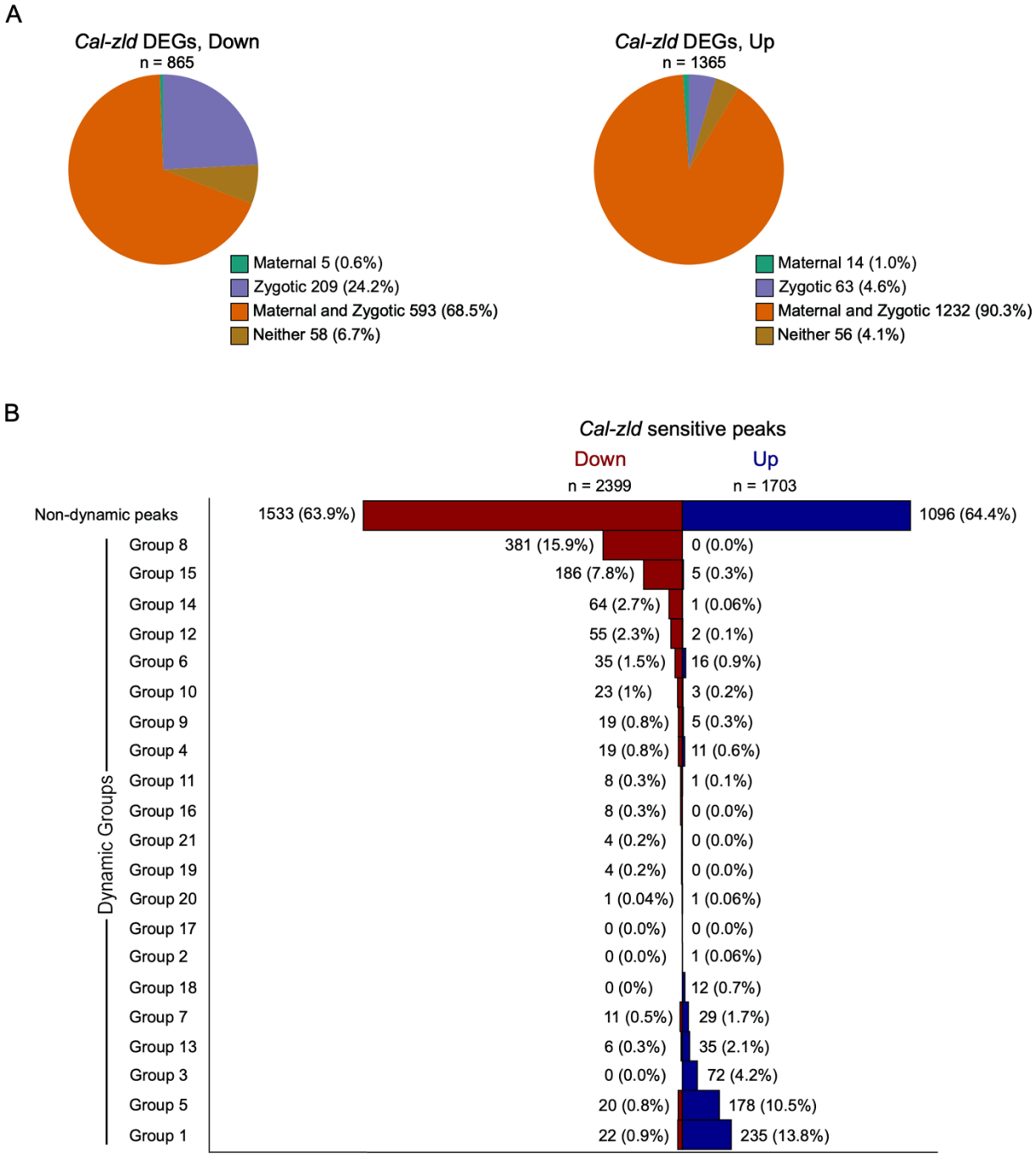

Figure S2. 7 Classifications of differentially expressed genes in *Cal-zld* RNAi embryos and Cal-zld sensitive ATAC-seq peaks.

A) Percentage of down-regulated (left) and up-regulated (right) genes in *Cal-zld* RNAi embryos classified as maternal, zygotic, maternal and zygotic or with little to no expression (neither) in wild type. Classifications were determined by expression analysis of genes at NC2/3 (~45-minute old embryos) or at syncytial blastoderm stages (NC11 through Late NC14, see *Materials*

and Methods). B) *Cal-zld* sensitive ATAC-seq peaks in dynamic and non-dynamic groups. Peaks in different groups are ordered by *Cal-zld* sensitivity. Peaks that gained accessibility are shown in blue and peaks that lost accessibility are shown in red. Numbers of peaks and percentage relative to all down- or upregulated peaks are indicated.

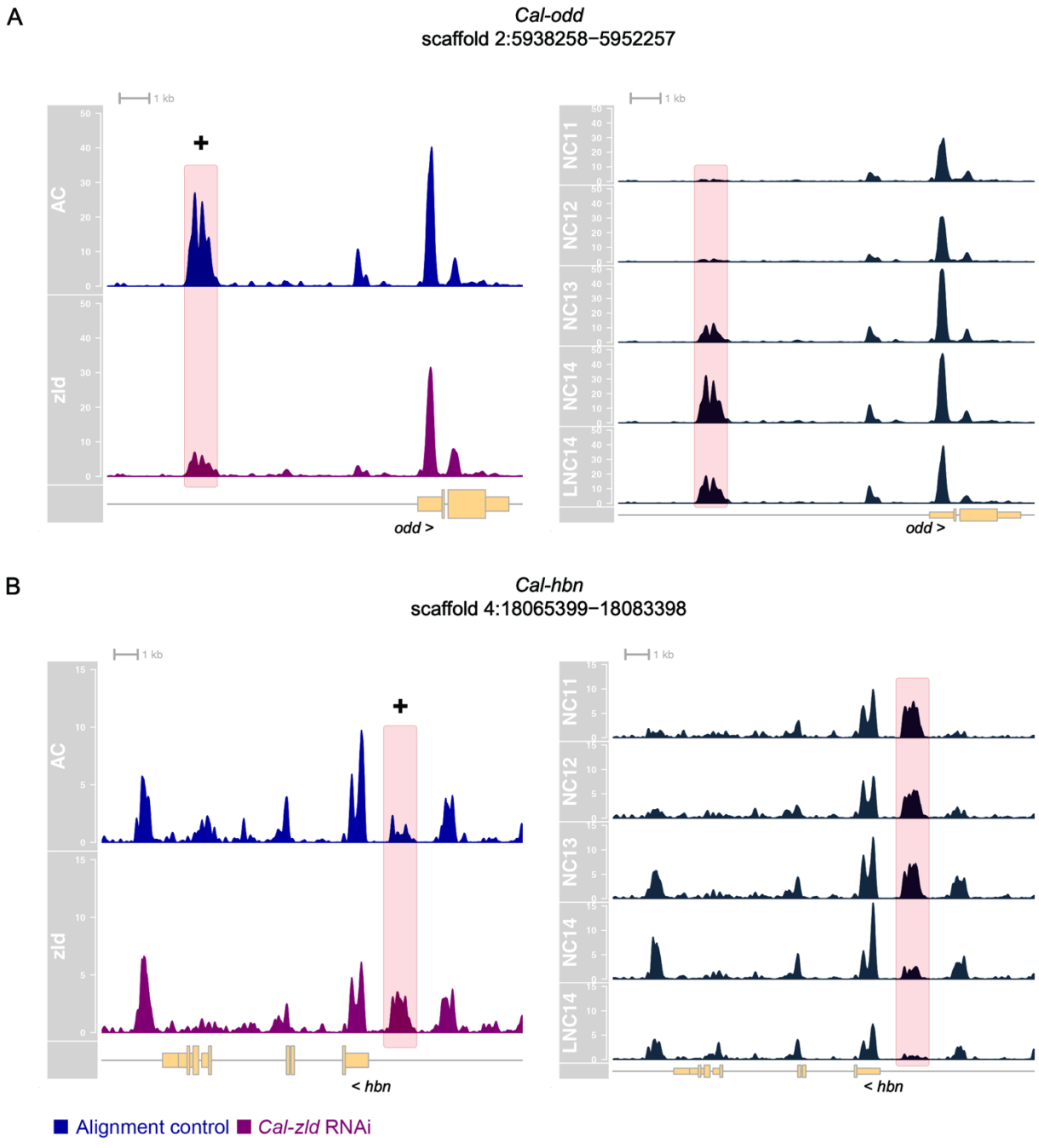

Figure S2. 8 Examples from the loci of two developmental genes that display *Cal-zld* sensitive peaks.

A) The locus of *Cal-odd* contains one chromatin accessibility peak whose accessibility is dependent on *Cal-zld* to open and contains a Zld motif (highlighted). On the right is the same view of the locus in wild-type embryos at NC11 through Late NC14. The peak is in dynamic group 8 and progressively gains accessibility as nuclear cycles progress. ATAC-seq tracks: 0-50 CPM. B) The locus of *Cal-hbn* contains one chromatin accessibility peak whose accessibility is dependent on *Cal-zld* to close and contains a Zld motif (highlighted). On the right is the same view of the locus in WT embryos at NC11 through Late NC14. The peak is in dynamic group 7 and progressively closes as nuclear cycles progress. ATAC-seq tracks: 0-15 CPM.

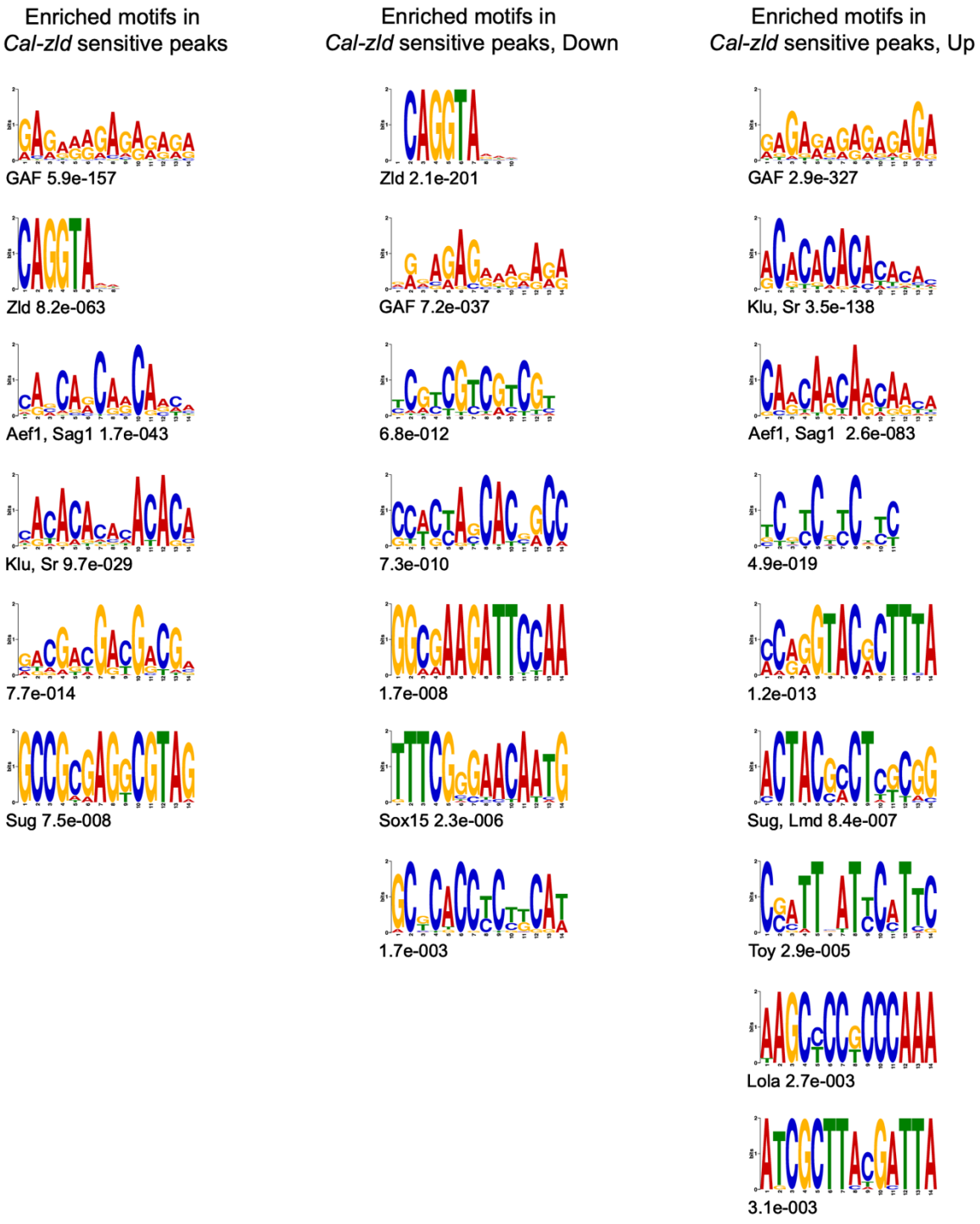

Figure S2. 9 Enriched motifs in *Cal-zld* sensitive peaks.
Enriched motifs identified by the MEME suite in all *Cal-zld* sensitive peaks (n = 4102, left), peaks whose accessibility decreases in *Cal-zld* RNAi embryos (n = 2399, center) and peaks whose accessibility increases in *Cal-zld* RNAi embryos (n = 1703, right). Motifs shown as a PWM logo with significance (E-value). DNA binding proteins of *Drosophila melanogaster* that bind to these or similar motifs as identified by MEME are indicated. We note that although not

identified by MEME, CLAMP has been shown to bind GA-rich motifs and Cg has been shown to bind (CA)n like the motifs above (Kaye et al., 2018; Kuzu et al., 2016; P. Ray et al., 2016; Soruco et al., 2013).

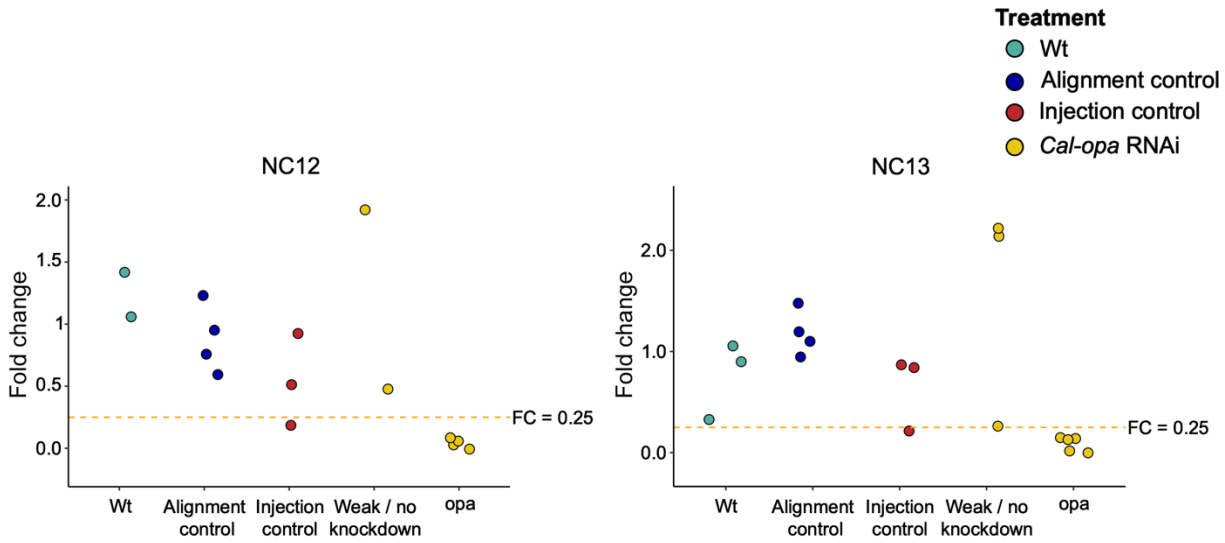

Figure S2. 10 Identification of *Cal-opa* RNAi embryos with strong knockdowns.
Plots of *Cal-opa* fold change (FC) for individual embryos at NC12 and NC13. Each embryo is represented by a circle. Embryos are grouped along the x-axis as indicated, spacing within a group was done arbitrarily for visualization. Treatment group of each embryo is indicated by color. FC for each embryo was calculated by dividing the embryo's CPM of *Cal-opa* by the average of wild-type and alignment control *Cal-opa* CPMs. FC was calculated using embryos from a single female (same activation batch). *Cal-opa* RNAi embryos with FC > 0.25 (weak / no knockdown of *Cal-opa*) were excluded from further analysis. WT, wild type.

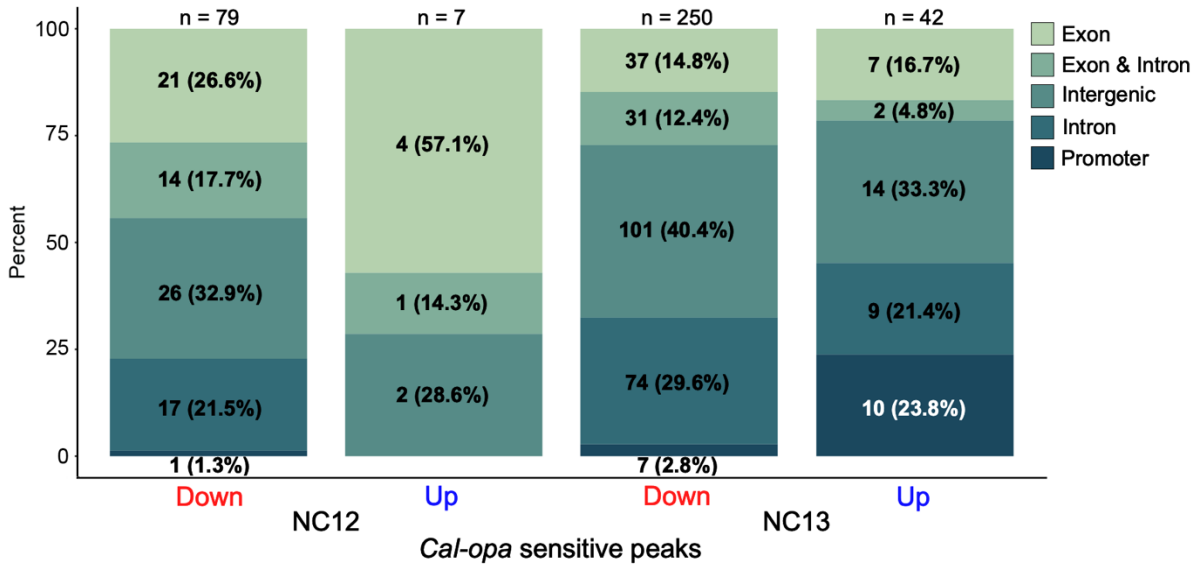

Figure S2. 11 Genomic distribution of *Cal-opa* sensitive peaks at NC12 and NC13.

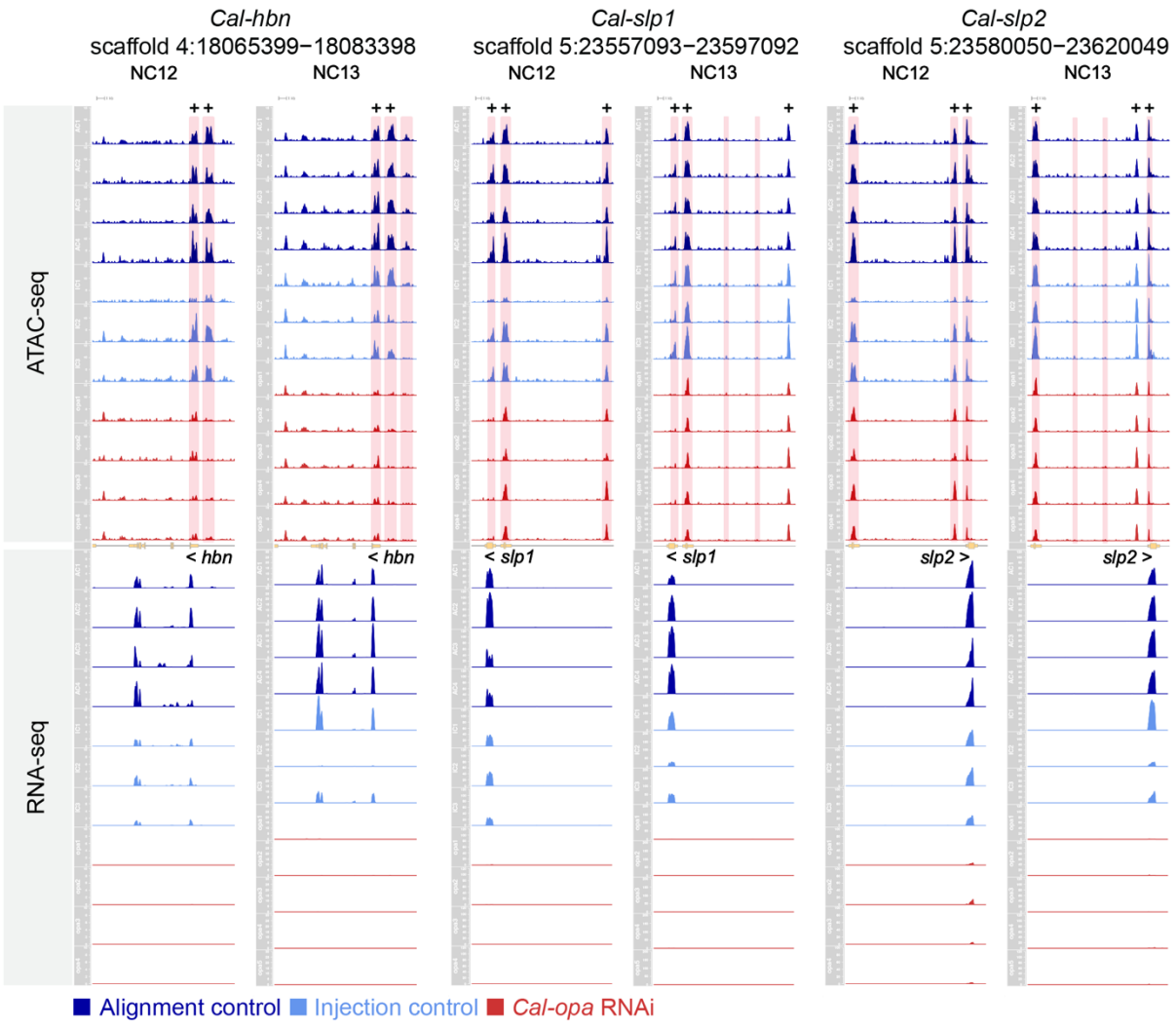

Figure S2. 12 ATAC-seq and RNA-seq tracks of *Cal-hbn*, *Cal-slp1 and Cal-slp2* plotted for individual replicates.

Alignment controls (blue), injection controls (light blue), and *Cal-opa* RNAi embryos (red) are shown at NC12 and NC13. Top tracks: CPM of ATAC-seq data. Middle track: gene model in pale yellow. Bottom tracks: CPM of RNA-seq data. Peaks with significant changes in accessibility highlighted in pink. Peaks containing Opa motifs are marked with a plus sign. The y-axis is the same for all tracks of the same data type at each locus. *Cal-hbn* ATAC-seq at NC12 and NC13: 0-15 CPM. *Cal-hbn* RNA-seq at NC12: 0-10 CPM. *Cal-hbn* RNA-seq at NC13: 0-30 CPM. *Cal-slp1* ATAC-seq at NC12 and NC13: 0-30 CPM. *Cal-slp1* RNA-seq at NC12: 0-120 CPM. *Cal-slp1* RNA-seq at NC13: 0-200 CPM. *Cal- slp2* ATAC-seq at NC12 and NC13: 0-30 CPM. *Cal-slp2* RNA-seq at NC12: 0-100 CPM. *Cal-slp2* RNA-seq at NC13: 0-120 CPM.

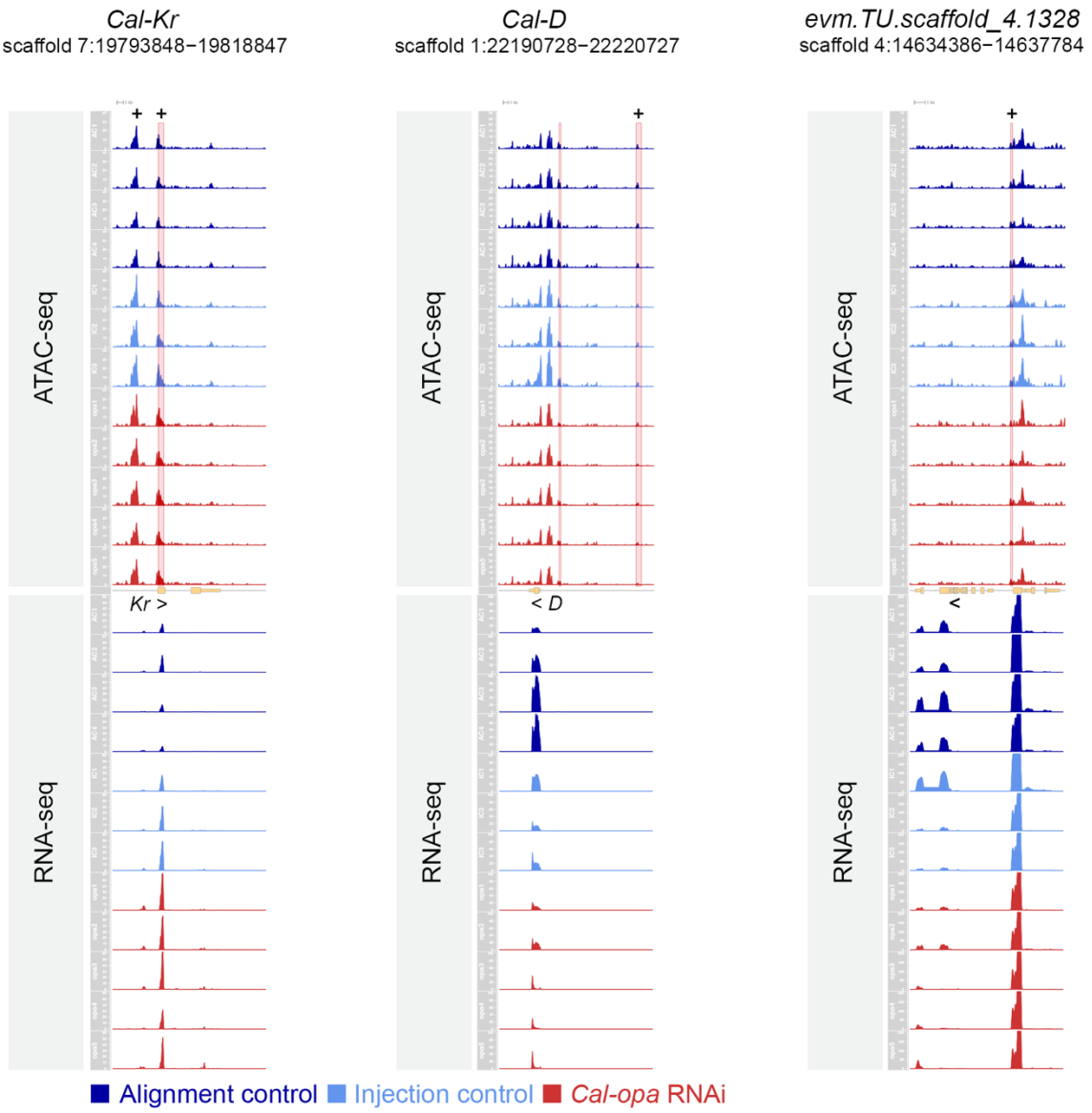

Figure S2. 13 ATAC-seq and RNA-seq tracks of *Cal-Kr*, *Cal-D* and *evm.TU.scaffold_4.1328* plotted for individual replicates.
Alignment controls (blue), injection controls (light blue), and *Cal-opa* RNAi embryos (red) at NC13. Top tracks: CPM of ATAC-seq data. Middle track gene model in pale yellow. Bottom tracks: CPM of RNA-seq data. Peaks with significant changes in accessibility highlighted in pink. Peaks containing Opa motifs are marked with a plus sign. The y-axis is the same for all tracks of the same data type at each locus. *Cal-Kr* ATAC-seq: 0-40 CPM. *Cal-Kr* RNA-seq: 0-60 CPM. *Cal-D* ATAC-seq: 0-30 CPM. *Cal-D* RNA-seq: 0-100 CPM. *evm.TU.scaffold_4.1328* ATAC-seq: 0-10 CPM. *evm.TU.scaffold_4.1328* RNA-seq: 0-200 CPM.

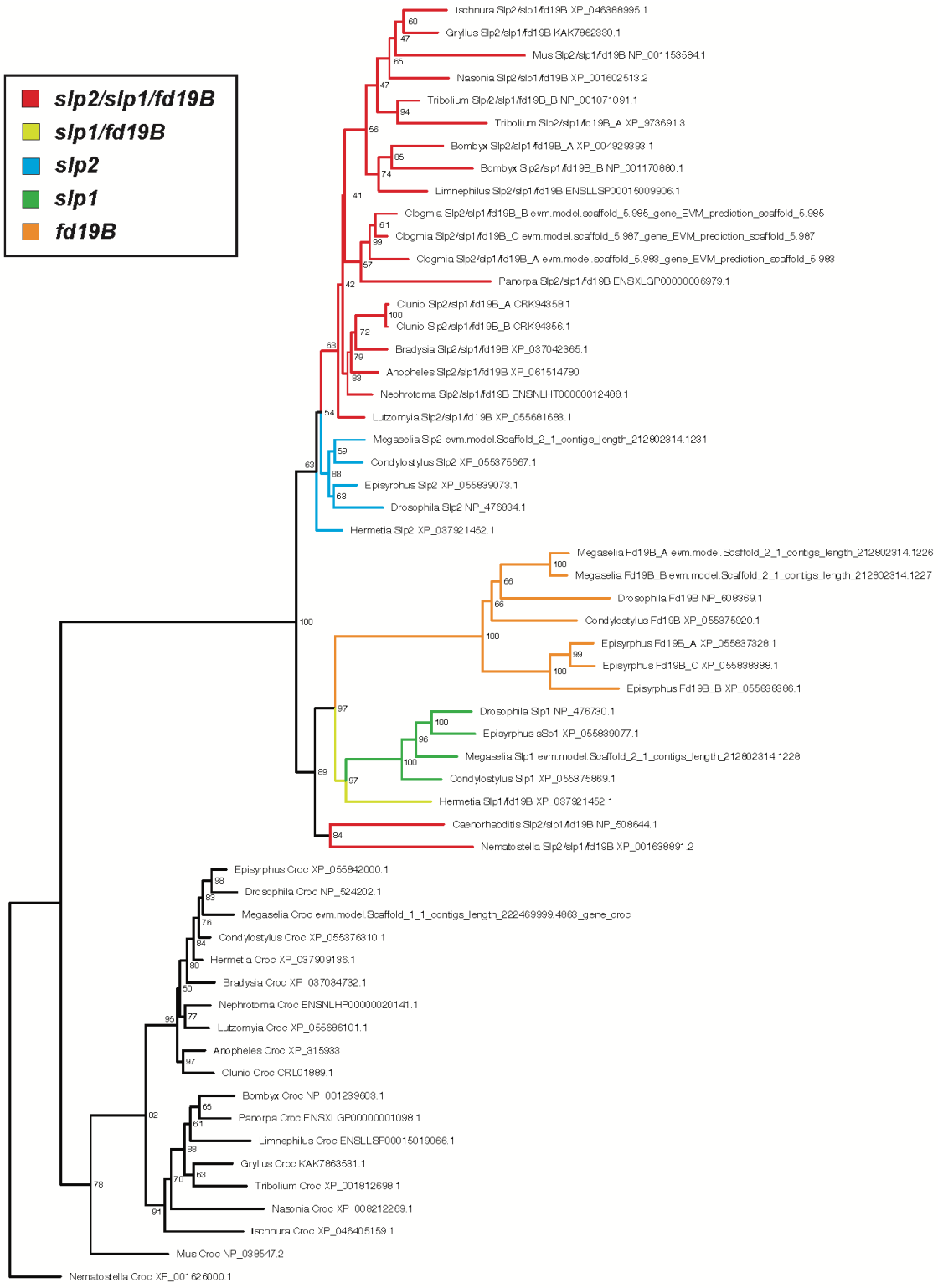

Figure S2. 14 Maximum-likelihood tree of Sloppy-paired homologs.
Reciprocal protein BLAST was performed using Slp1 (NP_476730.1), Slp2 (NP_476834.1), Fd19B (NP_608369.1), and Crocodile (Croc; outgroup; NP_524202.1) as queries. Homologous protein sequences in *Nematostella vectensis*, *Mus musculus*, *Caenorhapditis elegans*, *Ischnura*

elegans, *Gryllus longicernus*, *Nasonia vitripennis*, *Tribolium castaneum*, *Bombyx mori*, *Clunio marinus*, *Anopheles gambiae*, *Lutzomyia longipalpis*, *Bradysia coprophila*, *Hermetia illucens*, *Condylostylus longicornis*, and *Episyrphus balteatus* were identified by reciprocal Protein BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) (Camacho et al., 2009). Homologous sequences in *Limnephilus lunatus* (Austin et al., 2023) *Panorpa germanica* (Sivell et al., 2024), and *Nephrotoma appendiculata* (Crowley et al., 2024) were identified by reciprocal BLAST inGeneious Prime 2024.0.7 (https://www.geneious.com/), using respective transcript and protein fasta files available at ensemble-Darwin Tree of life (https://projects.ensembl.org/darwin-tree-of-life/). E value cut-off threshold was set to 0.05 and maximum number of target sequences set to 100. Identified sequences (for accession numbers see tree) were aligned using MAFFT v.7.526 with the L-INS-i strategy (https://mafft.cbrc.jp/alignment/software/) (Katoh et al., 2019). Maximum likelihood trees were constructed using IQ-TREE v.2.3.6. (Minh et al., 2020). The best molecular substitution model for each partition under the Bayesian information criterion was selected by partition merging strategy with ModelFinder (Kalyaanamoorthy et al., 2017). Maximum likelihood trees were constructed under the selected substitution models (VT+F+I+G4), with branch support values estimated by the ultrafast bootstrap approximation (Minh et al., 2020) using 1000 replicates. A majority-rule consensus tree was then generated based on the bootstrap results. The consensus tree was visualized on FigTree v.1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/) (Rambaut, 2018). Nematostella Croc sequence was manually chosen as root.

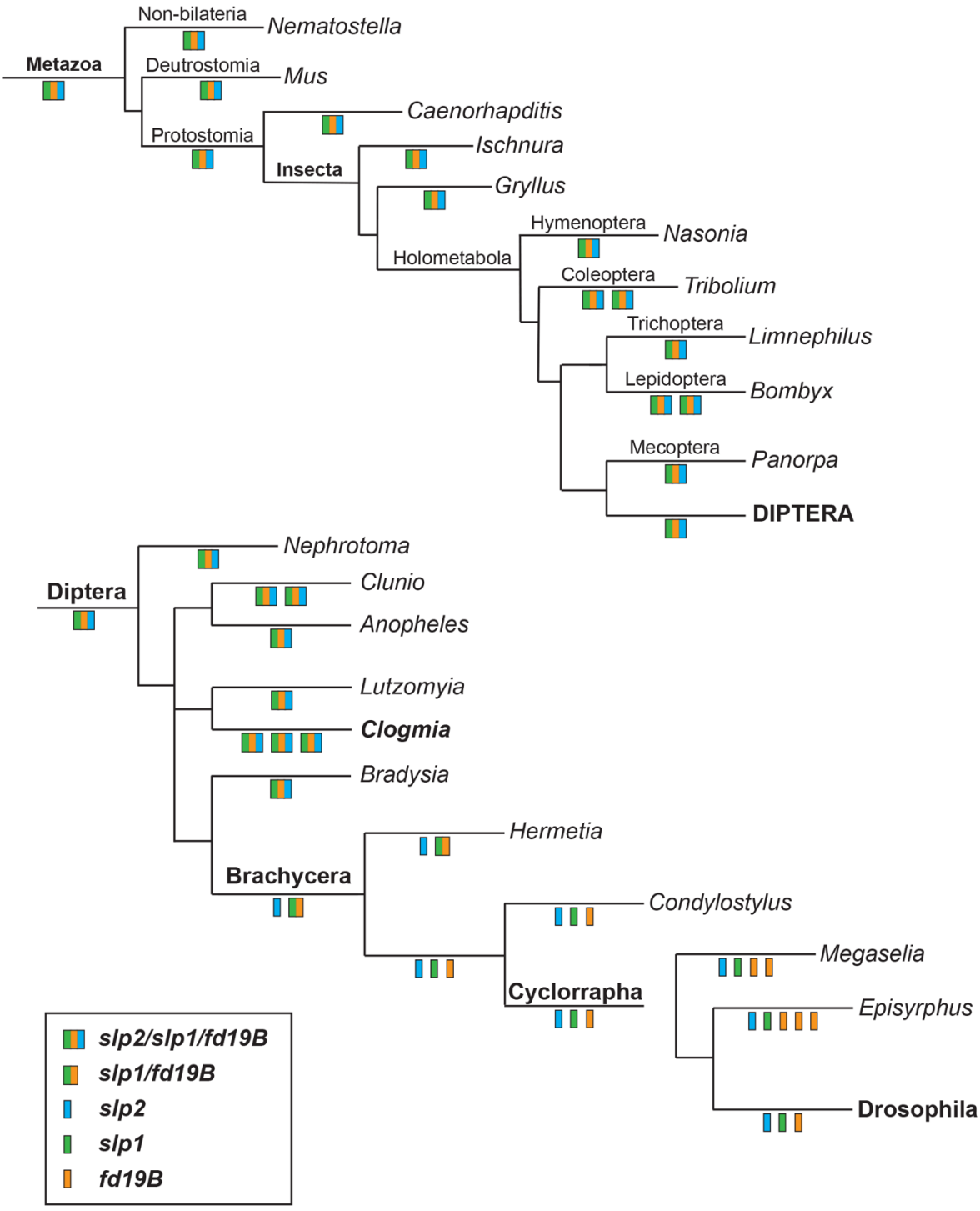

Figure S2. 15 The evolution of *sloppy-paired*.
Homologs of *sloppy-paired* are mapped onto simplified trees of Metazoa (top) and Diptera (bottom). Orthologs of *slp1* (green), *Fd19B* (orange), *slp2* (blue), and the inferred *slp1/fd19B/slp2* and *Slp1/fd19B* progenitors are color coded.

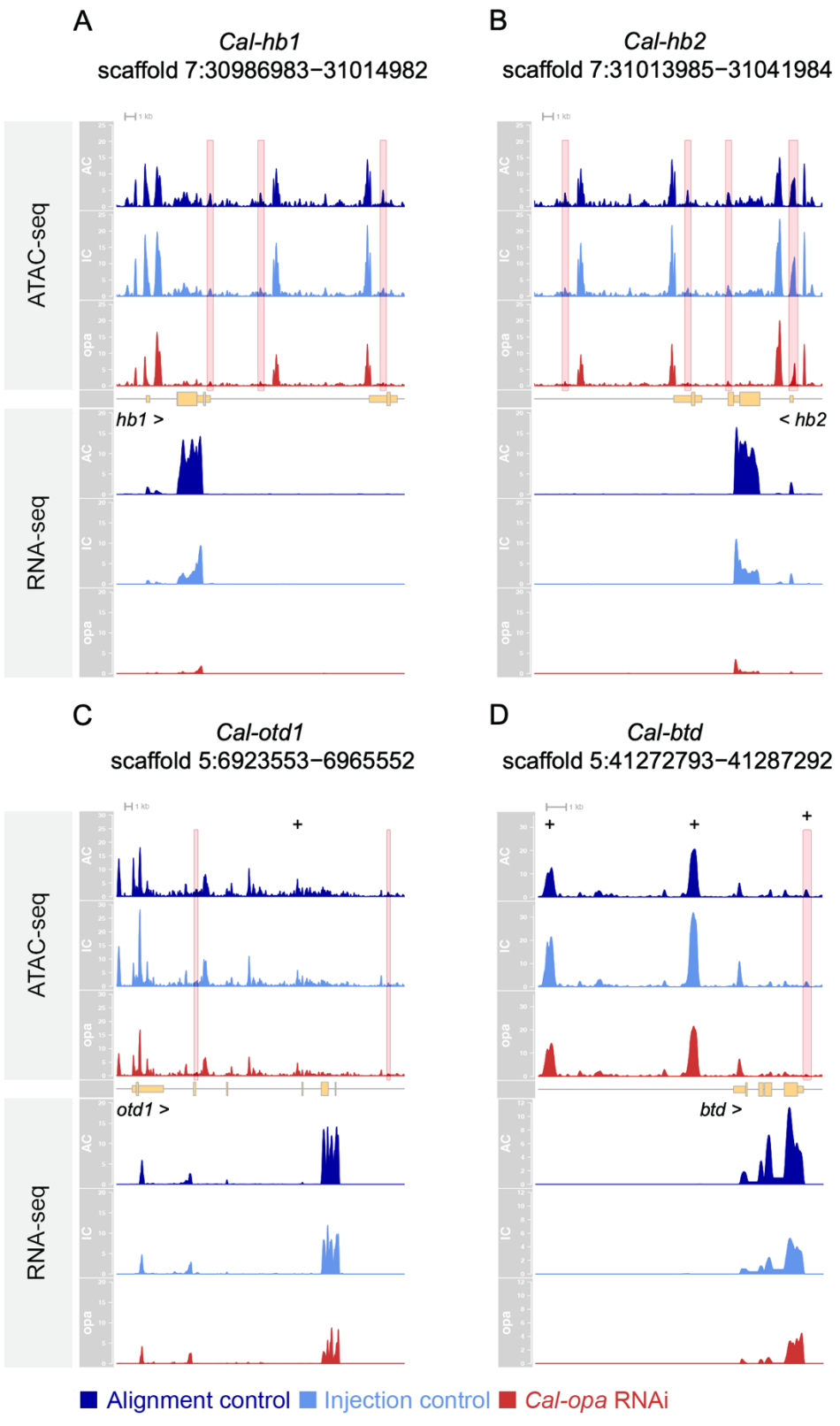
A
Cal-hb1
scaffold 7:30986983−31014982
B
Cal-hb2
scaffold 7:31013985−31041984
ATAC-seq
RNA-seq
AC
IC
opa
hb1 >
< hb2
C
Cal-otd1
scaffold 5:6923553−6965552
D
Cal-btd
scaffold 5:41272793−41287292
otd1 >
btd >
Alignment control
Injection control
Cal-opa RNAi

Figure S2. 16 ATAC-seq and RNA-seq tracks of *Cal-hb1*, *Cal-hb2*, *Cal-otd1* and *Cal-btd* at NC13.
Top tracks: CPM of ATAC-seq data. Middle track gene model in pale yellow. Bottom tracks: CPM of RNA-seq data. Peaks with significant changes in accessibility highlighted in pink. Peaks containing Opa motifs are marked with a plus sign. The y-axis is the same for all tracks of the same data type at each locus. *Cal-hb1*ATAC-seq 0-25 CPM. *Cal-hb1* RNA-seq 0-20 CPM. *Cal-hb2* ATAC-seq 0-25 CPM. *Cal-hb2* RNA-seq 0-20 CPM. *Cal-otd1* ATAC-seq 0-30 CPM. *Cal-otd1* RNA-seq 0-20 CPM. *Cal-btd* ATAC-seq 0-35 CPM. *Cal-btd* RNA-seq 0-12 CPM. Blue: Alignment control. Light blue: Injection control. Red: *Cal*-opa RNAi.

2.11.3 Movies and Appendix

Available at https://schmidtottlab.uchicago.edu/data/

Movie S2.1: Time-lapse movie from ~NC11 to gastrulation. Embryos from left to right; 1) *Cal-zld* RNAi 2) *Cal-zld* RNAi, 3) *Cal-zld* RNAi, 4) alignment control, and 5) *Cal-zld* RNAi. Transmitted light imaging was performed on a Leica DM5000B at 1-minute intervals. Imaging was stopped and needed to be restarted twice, this accrues at the ~30 second and ~65 second marks. Scale bar 100 μM.

Movie S2.2: Loss of Drosophila melanogaster *zelda* impairs cellularization and cytoplasmic clearing during zygotic genome activation. This movie is a composite of two independently recorded movies aligned to coincide with the beginning of nuclear cycle 14, both of Cry2-zelda embryos. Cry2-zelda embryos express a chimeric Zld protein fused to the light-inducible Cry2 module, allowing *zelda* loss of function upon exposure to blue light (McDaniel et al., 2019). The top embryo was imaged at a permissive wavelength (zld+, 594 nm) and the bottom one was imaged at the restrictive wavelength (zld-, 488 nm) on a confocal microscope with a 30-second frame rate. The timestamp indicates minutes relative to the beginning of NC14. Note that while the zld+ embryo develops clear cortical cytoplasm and cellularizes during the ~1h period of NC14, the zld- embryo retains dark cytoplasm at the basal surface of the nuclei and fails to

undergo cellularization, prior to loss of epithelial integrity (~ +45 minutes, lower right). Scale bar, 50 μm.

Appendix S2.1: Dovetail/Cantata Bio Hifiasm and HiRise Scaffolding Reports

Appendix S2.2: Summary *Cal-opa* RNAi analysis at NC12. 1) DESeq2 analysis of *Cal-opa* RNAi verses alignment control RNA-seq data. 2) DESeq2 analysis of *Cal-opa* RNAi verses alignment control ATAC-seq data. 3) Differentially expressed genes and ATAC-seq peaks within their loci. 4) Differentially expressed transcription factors and ATAC-seq peaks within their loci. 5) Differentially expressed transcription factors containing Opa motif positive ATAC-seq peaks (putative direct Cal-Opa targets). 6) Differentially expressed transcription factors with *Cal-opa* sensitive ATAC-seq peaks. 7) Differentially expressed transcription factors with *Cal-opa* sensitive ATAC-seq peaks containing an Opa motif. 8) DESeq2 analysis of *Cal-opa* RNAi verses injection control RNA-seq data. 9) DESeq2 analysis of Cal-*opa* RNAi verses injection control ATAC-seq data.

Appendix S2.3: Summary *Cal-opa* RNAi analysis at NC13. 1) DESeq2 analysis of *Cal-opa* RNAi verses alignment control RNA-seq data. 2) DESeq2 analysis of *Cal-opa* RNAi verses alignment control ATAC-seq data. 3) Differentially expressed genes and ATAC-seq peaks within their loci. 4) Differentially expressed transcription factors and ATAC-seq peaks within their loci. 5) Differentially expressed transcription factors containing Opa motif positive ATAC-seq peaks (putative direct Cal-Opa targets). 6) Differentially expressed transcription factors with *Cal-opa* sensitive ATAC-seq peaks. 7) Differentially expressed transcription factors with *Cal-opa* sensitive ATAC-seq peaks containing an Opa motif. 8) DESeq2 analysis of *Cal-opa* RNAi verses injection control RNA-seq data. 9) DESeq2 analysis of Cal-*opa* RNAi verses injection control ATAC-seq data.

DISCUSSION

3.1 The chromatin landscape during syncytial development in *Clogmia albipunctata*

This study shows that the chromatin landscape changes dramatically throughout the final four nuclear cycles of the blastoderm stage in *Clogmia albipunctata*. As in *Drosophila melanogaster*, widespread changes in the chromatin accessibility state are imposed upon the initial maternal ground state to guide zygotic genome activation. There are, however, notable differences between Clogmia and Drosophila in the early patterns of establishment and maintenance of accessible states. In Clogmia, the chromatin landscape is highly dynamic in that gains as well as losses occur throughout syncytial development. Conversely, in Drosophila, only gains in accessibility are observed during the early cleavage divisions and losses occur following the initiation of ZGA and lengthening of the cell cycle at NC14 (Blythe & Wieschaus, 2016; Soluri et al., 2020).

Clogmia and Drosophila differ in the length of their nuclear cycles during syncytial development. Although both species cellularize and gastrulate at the end of NC14, the duration of nuclear cycles is significantly longer in Clogmia than Drosophila. For example, once the syncytial blastoderm is formed, NC10 through NC14 are ~2-3 times longer (Jiménez-Guri et al., 2014). In Drosophila, short nuclear cycles impose significant limitations on establishing and stabilizing chromatin states and on transcriptional output (Blythe & Wieschaus, 2016). As such, for ZGA to occur faithfully, RNAP II loading at promoters *de novo* triggers a DNA replication checkpoint that dramatically slows the nuclear cycle at NC14 (Blythe & Wieschaus, 2015b). Once ZGA is underway in Drosophila, distinct chromatin states emerge corresponding to the major germ layers and their subtypes driven by both regional activity of activators, and by

repressors, either operating globally in all cells, or conditionally as part of developmental programs for cell fate decisions (Calderon et al., 2022; Cardamone et al., 2025; Cusanovich et al., 2018).

The significantly longer syncytial interphase times in Clogmia could alleviate the impact of cell cycle dependent constraints on maintaining chromatin states and transcription. This raises the possibility that the Clogmia embryo exhibits increased transcriptional activity and potentially initiates major ZGA before NC14. Correspondingly, as a result of ZGA occurring at an earlier nuclear cycle, distinct germ layer chromatin accessibility patterns emerge sooner than they would in Drosophila.

Likely, the dynamics we observe in Clogmia reflect both temporal changes in the activity of cis-regulatory elements, as well as the underlying heterogeneity in chromatin accessibility that arise as nuclei both activate and restrict the accessibility of cis-regulatory elements in time and space. Performing temporal single-cell ATAC-seq experiments in Clogmia will help distinguish between these possibilities. Single-cell ATAC-seq experiments provide the necessary spatial resolution of chromatin accessibility that cannot be addressed by bulk ATAC-seq. When measured by bulk ATAC-seq in a single embryo, the chromatin accessibility profile represents a weighted average of the degree of accessibility of elements across all nuclei, which does not capture the spatial differences in accessibility of these elements between individual nuclei (Bozek et al., 2019). Single-cell ATAC-seq can be performed across nuclear cycles to determine when global cell-type specific chromatin accessibility patterns develop and how the activity of cis-regulatory elements may change over time. Distinct chromatin profiles can be assigned to specific germ layers and their subtypes using known markers from similar studies in Drosophila (Calderon et al., 2022; Cardamone et al., 2025). Determining how and when ZGA

unfolds is another key component of this discussion (see future directions). It is important to determine whether germ layer–specific chromatin initiates once ZGA is underway, as observed in *Drosophila*, or whether chromatin signatures are instead primed before widespread ZGA. Additionally, ZGA can serve as a benchmark to facilitate more detailed comparisons of chromatin dynamics between Clogmia and Drosophila.

3.2 A unique role of Cal-Zld in regulating the transitions of the chromatin landscape

I find that mechanistically, Cal-Zld primes the accessibility of cis-regulatory elements at the loci of zygotic genes and is required to activate zygotic gene transcription. Depletion of *Cal-zld* results in widespread losses of chromatin accessibility and the loss of expression of target genes, consistent with a conserved role as a pioneer TF. This study also describes a novel function of Cal-Zld, consistent with the unique chromatin dynamics of Clogmia. Early open regions that lose accessibility during the syncytial cleavage divisions do so in a Zld-dependent manner. Based on the frequent occurrence (53%) of Zld motifs within this class of regions, these effects could be direct. This is in contrast to Drosophila, where Zld almost exclusively drives gains in chromatin accessibility through direct binding to DNA motifs independent of nucleosome association (Brennan et al., 2023; Hannon et al., 2017; Li et al., 2014; Schulz et al., 2015; Soluri et al., 2020; Sun et al., 2015). This observation suggests that Cal-Zld plays a broad role in regulating the chromatin landscape by both opening and closing chromatin. The proposed context-dependent activity of Cal-Zld may be mediated through a combination of direct and indirect mechanisms (considered below). Additionally, Cal-Zld's function may evolve depending on whether the embryo is approaching, undergoing, or has completed ZGA. Investigating how

Cal-Zld functions temporally in conjunction with the mechanisms discussed below will be critical to understanding its context-dependent roles in chromatin regulation.

Cal-Zld may function by recruiting or permitting the binding of activating or repressive complexes. Recently in Drosophila, Zld has been shown to interact with the Drosophila homolog of LIM-domain-binding protein 1 (Ldb1, called Chip in Drosophila) (Galouzis et al., 2025). In turn, Chip recruits histone acetyltransferase (HAT) CBP, leading to H3K27 acetylation of promoters and enhancers and subsequent transcriptional activation. Here, Chip acts downstream of Zld: Zld binds target promoters and enhancers, priming their initial accessibility independent of Chip. Once recruited by Zld, Chip is essential for the H3K27 acetylation of Zld targets (Galouzis et al., 2025). Zld may also play a role in depositing repressive H3K27me3 modifications; loss of *zld* results in both losses and gains of H3K27me3 within a subset of Polycomb Group domains. In some cases where H3K27me3 is reduced, binding of the catalytic subunit of Polycomb repressive complexes 2 (PRC2), E(z) is also lost (Gonzaga-Saavedra et al., 2025). Although the mechanism by which Zld distinguishes between different sites is unknown, these observations suggest that Zld binding also directs the activity of repressive complexes. Taken together, it is plausible that Cal-Zld interacts directly or through a cofactor with both activating and repressive histone modifying complexes. The context of Cal-Zld binding, in relation to cofactor or complex availability, nuclear concentration, or post-translational regulation of these factors could lead to different outcomes. In some contexts, Cal-Zld binding recruits a HAT complex leading to activating histone modifications and transcriptional activation. In another context, Cal-Zld binds target regions leading to their initial opening, however repressive complexes (e.g., histone deacetylases (HDAC) or H3K27 methyltransferases) are recruited to such regions leading to repressive histone modifications,

closing of chromatin and transcriptional repression. Similarly, the outcome of Cal-Zld activity could be influenced by the presence of transcriptional repressors or activators. In Drosophila, Zld facilitates binding of TFs leading to both activation and repression of the gene target (Bishop et al., 2023; Foo et al., 2014; Mir et al., 2017, 2018; Nien et al., 2011). Several experiments can be used to identify the complexes or cofactors that Cal-Zld may interact with. Co-immunoprecipitation will identify Cal-Zld interacting proteins, and binding comparisons of Cal-Zld and various chromatin remodeling complexes, cofactors and TFs will identify regions where these factors co-occupy. Measuring occupancy of these factors at putative target loci in *Cal-zld* mutant embryos will reveal if their binding is Cal-Zld dependent. Comparing loci either repressed or activated by Cal-Zld and performing these assays temporally will help determine whether Cal-Zld activity is ultimately governed by other trans-acting factors.

Context-specific activity of Cal-Zld may be mediated by the presence of stage-specific isoforms, which have been described in Drosophila (Giannios & Tsitilou, 2013; Hamm et al., 2015, 2017; Pearson et al., 2012). A short Zld isoform lacking the terminal four zinc-fingers has self-inhibitory activity, suggesting that some Zld domains have repressive functions (Hamm et al., 2017). Notably, the zinc-finger domains of Cal-Zld have not been functionally characterized, and Cal-Zld contains a novel zinc-finger that has been eroded in Drosophila (Ribeiro et al., 2017). Differences in protein structure and domains may help explain both the context-dependent Cal-Zld activity and divergent functions of Zld in Clogmia and Drosophila. With respect to context-dependent activity in Clogmia, the interplay of Cal-Zld isoforms at different nuclear cycles could mediate the outcome of Cal-Zld activity. Changes to the overall protein structure could result in interactions with different cofactors as discussed above and changes in the number of DNA-binding domains could alter the prevalence of Cal-Zld binding at different

targets. Functional studies of Cal-Zld protein domains, identification of putative *Cal-zld* isoforms and interacting partners, and temporal analysis of Cal-Zld activity across nuclear cycles are necessary to test this hypothesis.

Cal-Zld may act temporally with other pioneer factors, as documented in Drosophila (Colonnetta et al., 2021; J. Duan et al., 2021; Gaskill et al., 2021; Moshe & Kaplan, 2017). GAF and CLAMP are good candidates, as the GA-rich motif bound by both factors is consistently enriched in Clogmia's ATAC-seq peaks including those whose accessibility is affected by *Cal-zld* depletion. Moreover, regions that remain open in *zld* mutants in Drosophila are enriched in GAF motifs (Schulz et al., 2015). In Clogmia, Cal-GAF could prime the initial accessibility of regions that are later closed through Cal-Zld activity. Alternatively, Cal-Zld activity may be the result of competition with other factors. In Drosophila, TAGteam sites are bound by both Zelda and Grainyhead, a TF that has a pioneering role in establishing epithelial enhancers during larval stages of development (Harrison et al., 2010; Jacobs et al., 2018; Nevil et al., 2017, 2020). While both factors have different affinities for individual elements, Grh competes for binding to TAGteam sites bound by Zld, and overexpression of *grh* results in defects that resemble *zld* mutants. This observation led the authors to hypothesize that competition between Zld and Grh fine-tunes early zygotic transcription in pre-cellular blastoderm embryos (Harrison et al., 2010). Applying this to Clogmia, Cal-Zld could compete against other TFs that bind similar motifs or whose motifs overlap Cal-Zld binding sites. In this case, the strength of Cal-Zld binding, as well as its concentration compared to the competitive TF, dictate the observed effect of Cal-Zld activity.

Motif position, number, and motif strength could dictate how Cal-Zld switches from opening regions to closing regions. For example, TFs can exhibit so called "incoherent" activity,

activating or repressing transcription depending on the strength, number, and position of their binding motifs (Martinez-Corral et al., 2024). Interestingly, I observed that both classes of peaks, either opened or closed by Cal-Zld, contain Zld motifs, however the canonical Zld motif (5'-CAGGTA, (Nien et al., 2011)) was enriched within the peak summit of regions that require Cal-Zld to open and not in regions that require Cal-Zld to close. It is plausible that the enrichment of Cal-Zld motifs in one class over the other reflects the strength of Cal-Zld binding in those regions. In Drosophila, Zld's pioneering activity has been shown to increase with motif strength and increased binding (Brennan et al., 2023; Gibson et al., 2024). To fully consider the hypothesis that the context of the motif (i.e. position, strength, number) influences Cal-Zld activity, one would first need to perform a binding assay to determine motifs bound by Cal-Zld and rank them by strength of binding. Using these motifs, one could do several types of analyses. For each peak, the overall Cal-Zld binding could be scored, based on the composition of strong and weak Zld motifs and overall number of motifs within the peak. The position of the motifs within the peak as well as the position of the entire peak in relation to its putative gene target can also be measured. Once delineated, one could then probe whether any factor or set of factors correlate with Cal-Zld-dependent activation or repression.

3.3 Symmetry breaking at the level of chromatin accessibility is a shared feature of dipteran anterior determinants

Clogmia's AD, Cal-Opa, influences the chromatin landscape at the level of accessibility. Loss of Cal-Opa results in reduced chromatin accessibility at a smaller cohort of regions compared to Cal-Zld. These Opa-dependent regions include zygotic genes expressed in the anterior of the embryo. I hypothesize that these zygotic genes are necessary for mediating

anterior specification downstream of Cal-Opa. These observations are consistent with studies in Drosophila, where Bcd acts at a small number of segmentation genes expressed in the anterior of the embryo that are required for anterior patterning (Hannon et al., 2017). Collectively, my study suggests that AD-dependent symmetry breaking along the AP axis in Clogmia and Drosophila is initiated at the level of chromatin accessibility at a small number of critical cis-regulatory elements.

ADs could function in a classical sense of pioneering, which requires binding of nucleosome occupied DNA and facilitation of nucleosome removal via recruitment of nucleosome remodelers or cofactors (Colonnetta et al., 2021; Iwafuchi-Doi & Zaret, 2014; Judd et al., 2021; Larson et al., 2021; McDaniel et al., 2019; Tsukiyama et al., 1994). Alternatively, ADs could function by competing with nucleosomes to occupy naked DNA. TF-nucleosome competition occurs when nucleosomes transiently leave DNA or when they are displaced during replication (Polach & Widom, 1995; Ramachandran & Henikoff, 2016). This appears to be the case for Bcd; recent studies demonstrated that Bcd does not pioneer chromatin as observed with Zld and instead appears to function through nucleosome competition (Brennan et al., 2023; Degen et al., 2024). Degen et al. used the MS2-MCP system to measure the dynamics of zygotic *hb* transcription driven solely by Bcd target enhancer *hb* P2. The accessibility of this enhancer is Bcd dependent, as Bcd concentration increases nucleosome occupancy at *hb* P2 decreases resulting in the enhancer's open chromatin state (Degen et al., 2024; Hannon et al., 2017). Using this system, the authors monitored when transcription of *hb* initiates (onset time) and measured the *hb* expression domain along the AP axis at NC13. These observations were further investigated by mathematical modeling. Measurements of *hb* transcription were successfully recapitulated by a model in which Bcd competes with nucleosomes to gain access to its binding

sites within the P2 element. While this model takes the replication state of the P2 enhancer into account, Bcd binding to P2 is not dependent on replication. Taken together these data support a mechanism where Bcd outcompetes nucleosomes above a concentration threshold and in turn opens regions in the anterior of the embryo (Degen et al., 2024; Hannon et al., 2017).

In contrast to Bcd, Cal-Opa may pioneer chromatin directly. Zic3, one of the vertebrate homologs of Opa, was found to bind nucleosomal DNA *in vivo*, and another homolog, Zic2, interacts with chromatin remodeling complex Mbd3/NuRD in mouse embryonic stem cells (Grand et al., 2024; Luo et al., 2015). To address the hypothesis that Cal-Opa pioneers chromatin through the canonical mechanism one could perform a binding assay such as the electrophoretic mobility shift assay (EMSA) to determine whether Cal-Opa binds nucleosome occupied DNA. Following, interactions with suspected cofactors or nucleosome remodeling complexes can be addressed using co-immunoprecipitation experiments. Testing the nucleosome competition mechanism is more difficult to study directly and requires one to first rule out canonical pioneering. Then one could measure chromatin accessibility in embryos where the concentration of Cal-Opa has been manipulated via injection of mRNA or by creating a genetic strain. In such experiments, one would expect that as the concentration of Cal-Opa increases, the accessibility of its target cis-regulatory elements would also increase and the modeled nucleosome occupancy at these elements would decrease.

3.4 Zygotic targets of the anterior determinants in Clogmia and Drosophila

Although the ADs of Clogmia and Drosophila share the function of breaking symmetry at their target loci, the targets themselves appear to have diverged. Cal-Opa targets the loci of *Cal-*

hbn and *Cal-slp* homologs to prime accessibility of their cis-regulatory elements and initiate their expression, while Bcd-dependent enhancers drive anterior expression of gap and pair-rule genes, including enhancers of *otd*, *ems*, *btd*, *gt*, *kni*, *hb*, *slp1*, *eve*, *prd*, *cap'n'collar* (*cnc*), and *hkb*. In Clogmia, only a few Bcd targets were affected by the loss of Cal-Opa including homologs of *hb* (*Cal-hb1*, *Cal-hb2*), *otd* (*Cal-otd1*) and *btd* (*Cal-btd*). However, only *Cal-btd* showed evidence for being directly targeted by Cal-Opa; the locus of *Cal-btd* contained one Cal-Opa dependent peak containing an Opa motif. However, despite evidence that Cal-Opa pioneers chromatin at the *Cal-btd* locus, the expression of *Cal-btd* was not significantly affected by loss of *Cal-opa*. In the case of Cal-*hb1*, *Cal-hb2* and *Cal-otd1* the accessibility at their loci was *Cal-opa* insensitive except for a few small regions that did not contain Opa motifs. The expression of both *hb* homologs was reduced in *Cal-opa* RNAi embryos; however, it is unclear if this is a direct or indirect effect. Other Bcd target orthologs (e.g. *Cal-gt*, *Cal-kni/knrl1*, *Cal-kni/knrl2*, *Cal-ems1-4*, *Cal-otd2*, *Cal-hkb*) are not expressed until NC14 or later when the maternal Cal-Opa protein has cleared from the embryo (Misha Ludwig, personal communication). Additional studies characterizing Cal-Opa binding *in vivo*, are required to determine if additional Cal-Opa targets remain to be discovered including those discussed above.

Limitations aside, this study suggests that Cal-Opa targets fewer genes that Bcd in Drosophila. The differences between the two species in nuclear cycle length and timing of segmentation gene expression may help explain this difference. Bcd protein is present in the Drosophila embryo through the start of NC14, spanning the period when gap and many pair-rule genes are activated (~2.5 hours). Within the same period (absolute time), the Clogmia embryo would not have even formed the syncytial blastoderm (Jiménez-Guri et al., 2014). Moreover, Cal-Opa is cleared from the embryo by NC14 before many segmentation genes are activated.

Therefore, while Bcd has a *longer* opportunity to regulate segmentation gene expression, Cal-Opa has a *shorter* window. Consequently, Cal-Opa targets may be reduced to the fewest zygotic genes required to define the AP axis through repression of posterior genes and initiate anterior development by activating anterior specific gene expression. These genes, in turn, as Cal-Opa activity subsides, could take on the role of activating other anterior gap and pair-rule domains and repressing the expression of posterior segmentation genes. Interestingly, others have suggested that a limited number of target genes may facilitate AD substitution in lower dipterans (Yoon, 2019). Future studies are required to test this hypothesis, including identifying the downstream targets of *Cal-hbn*, *Cal-slp1* and *Cal-slp2*, detailing the exact expression patterns of other segmentation genes, and knockdown studies addressing the genetic relationships between segmentation genes.

I acknowledge that much of this discussion centers on whether Cal-Opa targets the same genes as Bcd, those genes being the canonical segmentation genes. Now I consider if there is evidence that Bcd targets the Cal-Opa targets identified in this study. From reviewing the available data from Hannon et al., I find that Bcd binds enhancers of *hbn* and *slp1* and the paralog of *slp1*, *fd19B*. In the case of *hbn,* there are two Bcd bound enhancers that close when the concentration of Bcd is reduced significantly. In contrast, *slp1* and *fd19B* each have one enhancer that is highly sensitive to Bcd concentration and dramatically loses accessibility when Bcd concentration is slightly reduced (Hannon et al., 2017). The *slp1* enhancer drives the anterior stripe of its pair-rule expression pattern, which is lost in *bcd* mutants (Datta et al., 2018; Liu et al., 2018). Both *hbn* and *fd19B* have anterior expression in the syncytial blastoderm, but it has not been determined if this expression is Bcd dependent. Although limited in scope this data suggests that Bcd may have once regulated Cal-Opa targets.

I consider the "hand-over" mechanism described by Li & Johnson (**Figure 3.1**). Here an original regulatory network composed of TF1 and its targets x1 and x2 is gradually taken over by TF2 through genetic interactions established *de novo* (H. Li & Johnson, 2010). Notice, regulation of y1 and y2, the original targets of TF2, is maintained. Hypothetically, the functions of y1 and y2 could gradually evolve leading to redeployment of these genes in the original network (Peter & Davidson, 2011; Voordeckers et al., 2015). This framework could be applied to explain how AD changeover throughout Diptera and how new genes are introduced into an existing network. First, the novel AD gains the correct expression pattern (i.e. maternal expression) through changes in its regulation including mutations in cis-regulatory sequence and change in transcription factor regulation (Peter & Davidson, 2011). Then the new AD gradually usurps regulation of the original network through a hand-over mechanism. This transition may also introduce new genes that were regulated by the new AD in a different context. Returning to the discussion of Bcd targets in Drosophila, evidence of genetic interactions of Bcd at the loci of *hbn* and *slp1* suggests that at one point these genes were important Bcd targets that functioned as canonical anterior gap genes. As the lineage of Drosophila evolved, these genes were gradually replaced by others giving rise to the extant network observed in *Drosophila melanogaster* today.

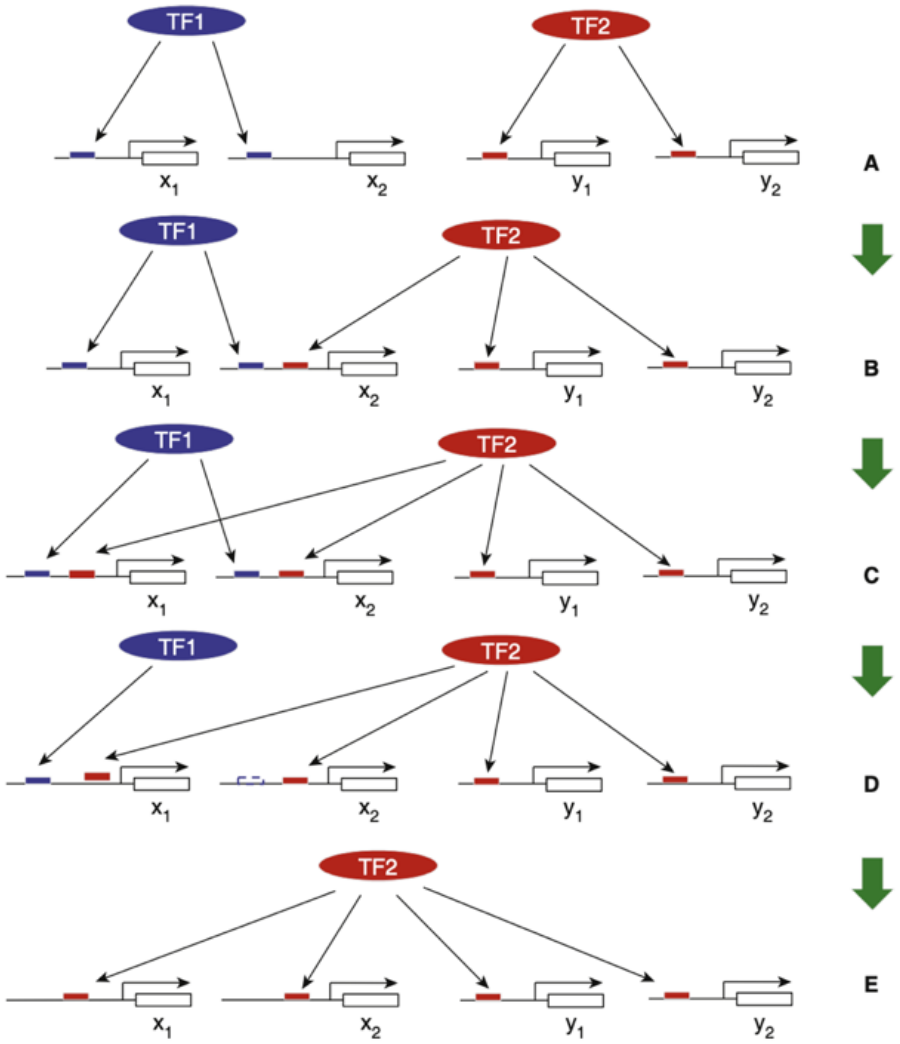

Figure 3. 1 Handover of a regulatory network from one TF to another.
The ancestral network is composed of x1 and x2 regulated by TF1 (A). The network goes through a series of intermediate states (B-D), where redundant regulation of the original network (x1 and x2) shifts to being completely under the control of TF2. Figure from (H. Li & Johnson, 2010).

3.5 Roles of *hbn* and *slp* homologs in the initiation of anterior patterning

This study finds that *Cal-hbn*, *Cal-slp1, and Cal-slp2* are targets of Cal-Opa. These genes may play key roles in anterior patterning not just in Clogmia, but also in other insects. In

Drosophila, *hbn* is required for structures within head segments, and at later stages during brain development (Hildebrandt et al., 2022; Kolb et al., 2021; Walldorf et al., 2000). In the beetle *Tribolium castaneum, Tca-hbn* plays an important role in establishing the embryo's AP polarity zygotically (Ansari et al., 2018). In Tribolium, the anterior of the embryo is specified by maternally supplied Axin, which represses Wnt signaling by promoting the degradation of ß-catenin. In the absence of Wnt signaling, *Tca-hbn* and *Tca-zen1* are zygotically expressed and appear to form a positive feedback loop that leads to the repression of *Tc-cad* (Ansari et al., 2018). Although the direct targets of Tca-Hbn are unknown, these data suggests that *Tca-hbn* functions to promote anterior identities in part by repressing genes that promote posterior fates and does not rule out that Tca-Hbn also initiates anterior patterning through activation of anterior specific gene expression. In conclusion, evidence from Tribolium and in this study suggests that *hbn* may have once held a broader role in specifying and patterning anterior fates.

The *slp* gene is a member of the *forkhead* domain gene family (Schomburg et al., 2022). *Slp1* is best known for its role as a pair-rule segmentation gene (Clark, 2017; Clark et al., 2019). However, it is first expressed in an anterior domain of the syncytial Drosophila embryo, like its closest paralogs, *slp2* and *fd19B* (Cadigan et al., 1994a; Fujioka & Jaynes, 2012; Grossniklaus et al., 1992, 1994; Lee & Frasch, 2004). Deletions at the *slp* locus cause defects in formation of the anterior mandibular segment, as well as the antennal and ocular regions of the head (Grossniklaus et al., 1992, 1994). In the beetle Tribolium, *slp* paralogs *Tca-slp1* and *Tca-slp2* are expressed in anterior segments and in a pair-rule pattern along the elongating germband (Choe et al., 2006; Choe & Brown, 2007; Janssen, 2020). Here, *slp* paralogs may have both gap and pair-rule functions; the mandibular, maxillary, and labial segments, as well as even-numbered trunk segments are lost when *Tca-slp1* is knocked down (Choe et al., 2006; Choe & Brown, 2007).

How *Cal-hbn* and *Cal-slp1* and *Cal-slp2* promotes anterior patterning through downstream targets remains unknown. The Homeobrain (Hbn) protein of Drosophila contains an octapeptide/GEH motif (YSIDQILG) associated with transcriptional repression in addition to its homeodomain (Walldorf et al., 2000). This domain is also conserved in Cal-Hbn. Therefore, in Clogmia, Cal-Hbn could function to regulate downstream segmentation genes as both an activator of genes that promote anterior identities and a repressor of posterior genes. Likewise, Slp1 and Slp2 may function similarly, as Slp1 can function as both a transcriptional repressor and activator in Tribolium and Drosophila (Andrioli et al., 2002; Buescher et al., 2004; Cadigan et al., 1994b; Choe & Brown, 2009; Clark, 2017; Konstantinides et al., 2022).

As a preliminary observation I find that the motifs of Hbn, Slp1, Slp2 are prevalent in the ATAC-seq peaks of segmentation genes (**Figure 3.2**). These observations suggest that *Cal-hbn* and the *slp* paralogs may converge to regulate the expression of canonical segmentation genes. It is important to note the functional significance of the ATAC-seq peaks at each locus has not been addressed and binding assays are necessary to confirm these observations. Moreover, non-canonical segmentation genes may also be downstream targets. Future studies combining binding assays of Cal-Hbn, Cal-Slp1, and Cal-Slp2 with functional analyses will help elucidate their downstream targets and their genetic relationships.

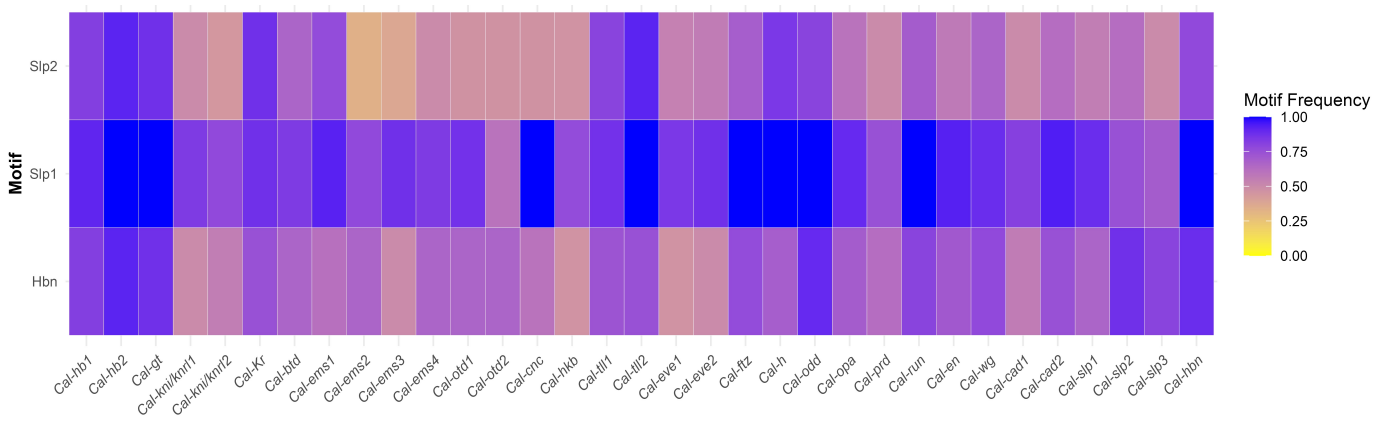

Figure 3. 2 Motif Frequency of Hbn, Slp1 and Slp2 at the loci of segmentation genes. Heatmap of motif frequencies of Hbn, Slp1 and Slp2 across segmentation gene loci. For each gene, all peaks within a 40 kb region centered on the gene were considered. Presence of the TF motif (Hbn, Slp1 or Slp2) was assessed for all peaks within a gene locus. Motif frequency was then quantified as the percentage of peaks within the locus that contained at least one binding motif for the indicated TF. Frequencies are represented on a yellow-to-blue scale from 0 to 100%.

3.6 Evolution of the dipteran segmentation network

Dipteran ADs are diverse, and this study suggests that their initial targets differ (Klomp et al., 2015; Yoon et al., 2019). Below, I summarize what else is known about the segmentation cascade in other dipteran insects starting with the first tier of gap genes. Collectively these studies highlight variations in the plasticity of individual network components and suggest that the segment-polarity genes represent a convergence point for different configurations of the cascade.

This study and others have identified homologs of core gap genes (e.g. *hb*, *gt*, *Kr*, *kni/knirps-like, tll*) and found differences in some of their expression patterns in Clogmia (García-Solache et al., 2010; Janssens et al., 2014; K. Rohr et al., 1999). For example, in Drosophila *gt* and *hb* have posterior domains (in addition to their anterior domains), that are largely absent in Clogmia.

The posterior domain of *Cal-gt* appears only briefly before disappearing at NC14 and the posterior domain of *Cal-hb1* (in the study *Calb-hb*) appears after gastrulation (García-Solache et al., 2010; Janssens et al., 2014). These expression patterns imply that at least some of the genetic relationships between gap genes, particularly those in the posterior of the embryo have changed. In Drosophila, *Kr* is expressed roughly in the center of the embryo, and its posterior boundary is repressed by the posterior domain of *gt*. In Clogmia, the expression domain of *Cal-Kr* is comparable to Drosophila, however it is unclear how its posterior boundary is set. The authors of the study under discussion, Garcia-Solache et al., proposed that *Cal-kni*/*knrl1 (in the study Calb-knl)* represses *Cal-Kr* in the posterior of the embryo, setting the posterior boundary. Moreover, a study that used expression data to predict gap gene interactions in Clogmia predicted multiple potential gap gene network structures (Crombach et al., 2014). Therefore, the network structure of the gap gene network in Clogmia remains speculative.

Functional evidence for gap genes in Clogmia is very limited. While some genes that function as gap genes in Drosophila or the flour beetle *Tribolium castaneum* are expressed in broad domains, they do not result in a gap phenotype when knocked down by RNAi in Clogmia (Jiménez-Guri et al., 2018). Only a thoracic defect was observed in *Cal-hb1* (in the study *Calb-hb*) RNAi experiments and *Cal-gt* RNAi did not cause obvious segmentation defects. Likewise, RNAi targeting another gap gene candidate, a Clogmia homolog of the *tarsal-less* gene of Tribolium with a posterior expression domain in Clogmia reminiscent to the posterior *gt* domain in Drosophila, did not cause segmentation defects. I note here that the authors did not publish data on their knockdown efficiencies, so it is possible that lack of a phenotype was due to insufficient knockdown. I also note that at the time the authors were unaware of a second copy of *hb* in Clogmia. In fact, my study found that several other gap genes are duplicated in Clogmia. It

is possible that some gap genes (e.g., *Cal-hb1* and *Cal-hb2*; *Cal-otd1 and Cal-otd2; Cal-tll1 and Cal-tll2*) may function redundantly with a paralog. Additionally, it seems likely (given the reduced posterior domain of *Cal-gt*) that not all Clogmia gap genes are currently known, and it remains unclear whether all Clogmia orthologs of Drosophila gap genes function as gap genes in Clogmia.

Overall, gap genes have not been comprehensively investigated in any lower fly species (Suborder: Nematocera, does not include Cyclorrhapha). In *Anopheles gambiae*, gap genes exhibit spatiotemporal differences from the expression patterns observed in both Drosophila and Clogmia (García-Solache et al., 2010; Goltsev et al., 2004b). This observation is consistent with the theory that there have been multiple iterations of the gap gene circuit within dipteran insects. Within higher flies (Suborder: Brachycera, includes Cyclorrhapha) the expression patterns of gap genes appear to be comparable to Drosophila, although in most cases not all gap genes have been evaluated (Bonneton et al., 1997; Lemke et al., 2010; Lukowitz et al., 1994; McGregor et al., 2001; Schetelig et al., 2008; Sommer & Tautz, 1991; Stauber et al., 2000; Treier et al., 1989; Wotton, Jiménez-Guri, et al., 2015; Wratten et al., 2006).

What about the pair-rule and segment-polarity genes? The Clogmia genome contains the canonical pair-rule genes described in Drosophila (*eve*, *h*, *rnt*, *odd*, *opa*, *ftz*, *slp*, *prd*). As others have noted, *eve* has duplicated in Clogmia (*Cal-eve1* and *Cal-eve2* (Bullock et al., 2004)), and my study found three lineage-specific copies of *slp* paralogs in Clogmia (*Cal-slp1*, *Cal-slp2*, Cal-*slp3*). To date little is known about the expression and function of pair-rule genes in Clogmia. *Cal-eve1* expression exhibits noteworthy spatiotemporal differences in Clogmia and Drosophila. The stripes of *Cal-eve1* are posteriorly shifted compared to Drosophila and form later into NC14 and the 7th stripe does not appear until well after the initiation of gastrulation (Bullock et al.,

2004; García-Solache et al., 2010; Jiménez-Guri et al., 2014; K. Rohr et al., 1999). Note, the expression patterns of *Cal-eve1* and *Cal-eve2* are identical in the blastoderm and early in gastrulation and the localization of Cal-eve1 protein is consistent with the expression data (García-Solache et al., 2010; Janssens et al., 2014). *Opa's* expression and pair-rule function is conserved in Clogmia and conferred by a strictly zygotic transcript isoform (Yoon et al., 2019). One *slp* homolog, *Cal-slp2,* is maternally deposited and localized in the anterior of the embryo forming an AP gradient (Yoon, 2019). *Cal-slp2* is zygotically expressed first in an anterior domain and forms seven stripes at the cellular blastoderm stage then parasegmental stripes during germband extension (Amiri et al., 2025; Yoon, 2019). These expression dynamics are consistent with the expression patterns of *slp1* and *slp2* in Drosophila (Cadigan et al., 1994a; Fujioka & Jaynes, 2012; Grossniklaus et al., 1992; Lee & Frasch, 2004). Finally, and most notably, the expression of the primary pair-rule gene *hairy* (*h*) is not conserved in Clogmia, suggesting that its function is not conserved (Bullock et al., 2004). Due to the sparse data available it is difficult to speculate further on pair-rule organization in Clogmia However, it is important to note that the suspected reorganization of the gap gene network may have altered how the initial pair-rule network is established. For example, because of the missing gap gene domain of *gt* in the posterior of the embryo it is unclear how *Cal-eve* stripes 3-7 are specified (García-Solache et al., 2010). Moreover, as delineated by Goltsev et al. from their evaluation of the gap network in *Anopholes gambiae*, changes in gap expression led to different combinations of gap gene activity to specify the primary *eve* stripes (Goltsev et al., 2004b). Regarding the segment-polarity genes, *wg* expression appears similar to Drosophila but has not been documented in the same level of detail as *Cal-en* (Bullock et al., 2004). *Cal-en* patterning along

the germband is highly comparable to Drosophila (Patel et al., 1989; Schmidt-Ott et al., 1994; Schmidt-Ott & Technau, 1992).

There is limited data in other dipteran insects on the pair-rule and segment-polarity genes, however, studies evaluating their expression patterns find them to be generally comparable to patterns described in Drosophila (Lemke & Schmidt-Ott, 2009; Schmidt-Ott et al., 1994; Sommer & Tautz, 1991). Data from the mosquito *Anopheles stephensi* demonstrate how functionally equivalent substitutions within the pair-rule network can ensure proper segmental patterning. Pair-rule and segment-polarity expression patterns in Anopheles are highly comparable to Drosophila except for the pair-rule gene *Ast-rnt* which is expressed in a broad domain like *opa* in Drosophila. Anopheles, however, lacks the pair-rule gene *prd* and instead the Pax family member *gooseberry* (*gsb*) functionally replaces the gene in the pair-rule network of this species. *Ast-gsb* mutant embryos form half the number of *Ast-en* stripes and the larva form cuticles missing alternating segments (Cheatle Jarvela et al., 2020). However, it is unclear how *Ast-gsb* recovered *prd* function. One possibility is that there were two functionally redundant genes (*gsb* and *prd*) in the last common ancestor before the split between Drosophila and Anopheles. Alternatively, in Drosophila *gsb* could have duplicated and undergone sub-functionalization giving rise to *prd*.

In conclusion, the current evidence suggests that the early components of the dipteran segmentation network (the AD, gap and early pair-rule genes) are evolutionarily labile. Iterations of the network include changes in the genes involved, changes in expression patterns and altered genetic interactions within the cascade. Despite early flexibility, segmental boundaries established by segment-polarity genes expression is conserved. Much more work is required,

particularly at later points of the cascade, to determine how the network resolves to establish the parasegmental patterning that gives rise to the segment-polarity genes.

FUTURE DIRECTIONS

4.1 Investigating the regulatory landscape of segmentation downstream of symmetry breaking in Clogmia

This study identified *Cal-slp1*, *Cal-slp2,* and *Cal-hbn* as targets immediately downstream of the AD Cal-Opa. However, additional targets may remain to be discovered. An assay that profiles Cal-Opa binding *in vivo* such as Chromatin immunoprecipitation sequencing (ChIP-seq) or in vitro using the high-throughput technique DNA affinity purification sequencing (DAP-seq) could identify candidates that were missed due to limitations of this study. DAP-seq identifies binding sites of *in vitro* expressed tagged full-length TFs in endogenous DNA sequence from the same species. This technique has been demonstrated to recapitulate TF binding *in vivo* and circumvents the need to generate an antibody for each TF of interest (Bartlett et al., 2017). Candidate Cal-Opa targets identified using a binding assay would be expected to be expressed in gap or pair-rule patterns, which can be confirmed by *in situ* hybridization. Likewise, RNAi KD of the candidate genes would confirm their function in the segmentation cascade. Characterizing the change in expression of these genes in *Cal-opa* RNAi embryos would then elucidate the epistatic relationship between the identified gene and maternal Cal-Opa. Gathering this data could result in the identification of additional Cal-Opa targets.

Putative enhancers of the segmentation gene network have been identified in this study; however, the regulatory input to these regions remain unidentified. DAP-seq could be applied here to characterize binding of TFs that are suspected to be involved in Clogmia's segmentation cascade (e.g. *Cal-hbn*, *Cal-slp1 Cal-slp2* and canonical gap genes *Cal-hb1, Cal-hb2*, *Cal-gt*, *Cal-Kr*, etc). Coupling DAP-seq and the time-course ATAC-seq data from this study will lay the foundation to assemble the regulatory landscape of segmentation in Clogmia. The ATAC-seq

data provides information on the temporal activity of a putative enhancer while specific regulatory input can be deduced from the DAP-seq data. Functional studies including RNAi knockdowns, *in situ* hybridizations, genome editing, and enhancer reporter assays would be useful to formally validate the inferred gene regulatory networks. Dissection of how Clogmia's segmentation gene network is wired will enable detailed comparisons with Drosophila and advance our understanding of how the segmentation cascades of these species converge to establish the segmented body plan.

4.2 Symmetry breaking at the level of chromatin accessibility in other dipteran insects

The model proposed in this study, that unrelated anterior determinants break AP symmetry at the level of chromatin accessibility, is based on data from only two dipteran species. Data from additional dipteran species are necessary to test how generalizable the proposed mechanism of AP symmetry breaking is across the dipteran clade. Firsts this mechanism can be explored in *Nephrotoma suturalis,* the most basal dipteran in which the AD has been identified as TCF family member Pangolin (Pan) (Yoon et al., 2019). Pioneer activity by Pan has not been described in dipteran insects; however, the murine homolog, TCF-1 acts as a pioneer during T-cell development. TCF-1 is required to open enhancers that drive early phases of T-cell maturation, and when expressed ectopically in fibroblasts, TCF-1 can establish the accessibility of these regions by binding nucleosome occupied DNA (Johnson et al., 2018). Investigating pioneer activity by Nsu-Pan will determine if breaking symmetry by pioneering chromatin is an ancestral function of dipteran ADs. Additionally, these experiments will identify Nsu-Pan target genes, providing insight into the ancestral AD target genes.

Other notable species are *Megaselia abdita* and *Lutzomyia longipalpis,* whose ADs are Bcd and Opa, respectively (Stauber et al., 2000; Yoon et al., 2019). In Megaselia, Bcd is functionally conserved, and depleting the embryo of *bcd* transcripts results in mirror-image double abdomens, indicating that although maternal *hb* transcript is present in the early embryo, it does not substantially contribute to embryo polarity (Stauber et al., 2000; Wotton, Jiménez-Guri, et al., 2015). However, the gap gene network is believed to be highly conserved in Megaselia (Wotton, Jimenez-Guri, et al., 2015). Studying Bcd in Megaselia would not only test if Bcd breaks symmetry in the same way, but would also test how well conserved Bcd targets are in two Cyclorrhapha flies that diverged ~ 145 million years ago. Likewise, investigating the pioneer activity of Llo-Opa will address if ADs encoded by the same gene in different lineages function through a common mechanism and converge on the same targets. Injection of *llo-opa*, mRNA into the posterior pole of the Clogmia embryo can induce a second head, suggesting Opa may target the same genes in both species (Yoon et al., 2019) .

4.3 Timing of zygotic genome activation in *Clogmia albipunctata*

The longer syncytial interphase times in Clogmia compared to Drosophila could alleviate the impact of cell cycle dependent constraints on transcriptional regulation. In turn, widespread zygotic activation could start at an earlier nuclear cycle. In line with this hypothesis, the chromatin state of Clogmia is highly dynamic, in contrast with the stepwise gains in accessibility that occur leading up to major ZGA in Drosophila (Blythe & Wieschaus, 2016; Soluri et al., 2020). Time course binding assays of RNA polymerase II would elucidate the temporal sequence of zygotic gene activation and allow one to identify the earliest transcribed genes and probe their

function. Coupling ChIP-seq of histone modifications with the chromatin accessibility data will provide definitive data on the activity state of cis-regulatory elements. Used together one could distinguish between genes required only early in development that are then turned off, corresponding to the loss in accessibility of the gene's cis-regulatory elements, versus turnover in enhancer use for a gene whose transcription is sustained but becomes for example restricted in space. The proposed study would provide useful insight to how increased nuclear cycle duration affects enhancer usage, turnover and timing of large-scale zygotic genome activation.

APPENDIX A: PRELIMINARY DATA ON ZYGOTIC CAL-OPA'S ROLE IN PROMOTING CHROMATIN ACCESSIBILITY DURING THE BLASTODERM STAGE

A.1 Introduction

In Drosophila, Opa is required for the transition of pair-rule gene expression from 7 to 14 stripes and is also required to activate the segment-polarity genes at the correct time (Benedyk et al., 1994; Clark & Akam, 2016). In addition, two studies found that Opa drives the accessibility of enhancers of the pair-rule, segment-polarity and dorso-ventral (DV) gene networks in late NC14 (Koromila et al., 2020; Soluri et al., 2020). Soluri et al. identified hundreds of regions bound by Opa that require Opa to gain accessibility. Many of these regions, gain accessibility late in NC14 including late pair-rule and segment-polarity enhancers (e.g. enhancers of *odd*, *eve*, *en, wg*, *patched* (*ptc*), *slp1*, *gsb*). Ectopic expression of *opa* caused many of the Opa-dependent regions to precociously gain accessibility at NC13. This study also showed that Opa-binding was sufficient to reduce nucleosome binding. Similarly, Koromila et al., identified Opa bound regions dependent on Opa to gain accessibility late in NC14. These regions included segmentation network enhancers (e.g. enhancers of *otd*, *hb*, *en*) and DV network enhancers (e.g. *short gastrulation* (*sog*) distal enhancer). Additionally, *opa* depletion resulted in the miss-regulation of hundreds of genes, including downregulation of the DV gene *sog*. The authors also investigated regions targeted by both Opa and Zld. Nearly 40% of the identified Opa bound regions were also bound by Zld, and moreover, regions bound by both Zld and Opa exhibited increased chromatin accessibility compared to regions bound by either factor alone. This result suggests, that Opa also contributes to the overall accessibility of those regions.

In Clogmia the late segmentation function of Opa appears to be conserved but involves a strictly zygotic transcript isoform (*Cal-opa*^{Zyg}) (Yoon et al., 2019). In my study of the maternal function of Opa in Clogmia, I found that Cal-Opa drives chromatin accessibility of select enhancers within the early segmentation network. These observations motivated investigation into the zygotic function of Cal-Opa (Cal-Opa^{Zyg}) with respect to its role in promoting chromatin accessibility in the final nuclear cycle of the blastoderm stage.

A.2 Results and Discussion

To assess the impact of Cal-Opa^{Zyg} activity on chromatin accessibility, I performed ATAC-seq and RNA-seq on single *Cal-opa*^{Zyg} RNAi embryos late in NC14. *Cal-opa*^{Zyg} was knocked down by injecting dsRNA targeting the zygotic transcript 4-5 hours into development (Yoon et al., 2019). In parallel, I processed stage-matched untreated controls that were allowed to develop under water in the dish used for egg activation (wild-type controls), as well as controls that were prepared for injection under oil but not injected (alignment controls). The RNA-seq data was used to measure *Cal-opa*^{Zyg} expression and select embryos with strong knockdown efficiencies and maximally reduced *Cal-opa*^{Zyg} mRNA levels. In analogous experiments, where either maternal *Cal-opa* or *Cal-zld* was knocked down, a knockdown efficiency greater than 75% was treated as a successful knockdown. However, the strongest knockdowns of *Cal-opa*^{Zyg} ranged from 72%-61% (**Figure A.1**). In addition, two alignment control embryos had particularly low *Cal-opa*^{Zyg} expression. These embryos may have been mislabeled or were inadvertently damaged during the preparation process. These two samples were excluded from further analysis. Using DESeq2 I compared the changes in chromatin

accessibility and expression of *Cal-opa*^{Zyg} knockdown embryos (n=4) to alignment controls (n=2).

Depletion of *Cal-opa*^{Zyg} lead to a modest effect on chromatin accessibility, with 44.9% (105/234) peaks losing accessibility and 55.1% (129/234) gaining accessibility (**Figure A.2 A**). Peaks whose accessibility decreased in *Cal-opa*^{Zyg} RNAi embryos had a higher proportion that map to intergenic or intronic sequences (~78%), compared to peaks whose accessibility increased upon *Cal-opa*^{Zyg} depletion (~52%, **Figure A.2 B)**. Further work is required to determine if Cal-Opa^{Zyg} functions directly to repress chromatin regions, or if the observed gains in accessibility following *Cal-opa*^{Zyg} depletion are caused by indirect compounding effects. In Drosophila, Opa functions as both a transcriptional activator and repressor (Hursh & Stultz, 2018) raising the possibility that Cal-Opa^{Zyg} can open or close chromatin regions directly. To better distinguish between direct and indirect targets, I performed an Opa motif search within all *Cal-opa*^{Zyg} sensitive peaks. This search used the Opa motif previously derived from MEME analysis in Clogmia as well as an Opa motif derived from a MEME analysis of ATAC-seq data in Drosophila (Soluri et al., 2020). An Opa motif was not among the motifs identified by MEME analysis as enriched in *Cal-opa*^{Zyg} sensitive peaks (**Figure A.2 C**). However, I found that 19.0% (20/105) of the peaks that lost accessibility and 20.2% (26/129) of the peaks that gained accessibility in response to *Cal-opa*^{Zyg} knockdown contain at least one Opa motif. I note here that the cohort of putative direct Cal-Opa^{Zyg} targets could be larger given that the knockdown of the *Cal-opa*^{Zyg} transcript was moderate.

Next, I determined the overlap between *Cal-opa*^{Zyg} and *Cal-zld* sensitive peaks. I found that 27% (64/234) of the *Cal-opa*^{Zyg} sensitive regions are also sensitive to *Cal-zld* depletion (**Figure A.2 D**). Among the regions sensitive to both *Cal-opa*^{Zyg} and Cal-*zld* depletion,12.5%

(8/64) contained both Opa and Zld binding motifs. These results raise the possibility that Cal-Opa^{Zyg} and Cal-Zld cooperatively regulate chromatin accessibility as shown in Drosophila, or alternatively, that Cal-Opa^{Zyg} is dependent on Cal-Zld for binding. Experiments measuring Cal-Opa^{Zyg} binding in *Cal-zld* RNAi embryos and chromatin accessibility in embryos depleted for both *Cal-opa^{Zyg}* and *Cal-zld* could help distinguish between these possibilities.

Finally, I assembled a list of developmental genes including segmentation genes shown to be regulated by Opa either at the level of chromatin accessibility or transcription in Drosophila (Koromila et al., 2020; Soluri et al., 2020). Ultimately, I sought to identify how *Cal-opa^{Zyg}* depletion affected the expression of these genes and the chromatin state at the loci of these genes. In addition, I included TF genes that were either differentially expressed upon *Cal-opa^{Zyg}* or whose loci contained *Cal-opa^{Zyg}* sensitive peaks. Table A.1 summarizes these results. I find that 46 genes have at least one *Cal-opa^{Zyg}* sensitive ATAC-seq peak, including the locus of *Cal-en* (**Figure A.3 A**). 13 of these loci, including the locus of *Cal-btd,* contain at least one Opa motif positive ATAC-seq peak that was sensitive to *Cal-opa^{Zyg}* depletion (**Figure A.3 B**). In all cases, the expression of these genes was not significantly affected by *Cal-opa^{Zyg}* depletion. Only three gene loci exhibited changes in accessibility that led to differential gene expression following *Cal-opa^{Zyg}* depletion. These loci, however, do not contain peaks that are both Opa motif positive and *Cal-opa^{Zyg}* sensitive. One example of this class is the homolog of *sog* (*Cal-sog*); depletion of *Cal-opa^{Zyg}* led to increased expression of *Cal-sog* and increased accessibility of one peak at the *Cal-sog* locus (**Figure A.3 C**).

Overall, most of the pair-rule and segment-polarity genes whose accessibility was Opa-dependent in Drosophila (e.g. *odd*, *eve*, *en, wg*, *ptc*, *slp1*, *gsb, hb, otd*) did not display *Cal-opa^{Zyg}* dependent ATAC-seq peaks. *Cal-en* and *Cal-hb2* appear to be the only shared targets of Cal-

Opa^{Zyg} and Opa in this regard. However, it is not clear if *Cal-en* and *Cal-hb2* are directly targeted by Cal-Opa^{Zyg}, as the *Cal-opa^{Zyg}* sensitive peak at their loci do not contain an Opa motif. *Cal-opa^{Zyg}* sensitive peaks were observed at the loci of segmentation genes *Cal-btd*, *Cal-cad1,* and *Cal-cad2*, however, only the peak at the locus of *Cal-btd* was positive for an Opa motif. I note that the expression of *Cal-prd* and *Cal-h* were reduced in *Cal-opa^{Zyg}* RNAi embryos and their loci contain at least one Opa positive ATAC-seq peak. It is possible that both genes are activated by Cal-Opa^{Zyg} directly, although the ATAC-seq peaks at their loci were insensitive to Cal-*opa^{Zyg}* depletion. In the case of *Cal-otd1*, the expression was increased in *Cal-opa^{Zyg}* RNAi embryos and one peak contained an Opa motif. Likewise, the expression of *Cal-slp1* and *Cal-tll2* increased upon Cal-*opa^{Zyg}* depletion and both loci contained at least one Opa positive ATAC-seq peak. Cal-Opa^{Zyg} may serve to repress these genes directly late in NC14 and further work is required to determine if this repression is spatial, temporal, or both. I also note that the expression of *Cal-slp1* and *Cal-otd1* is reduced in embryos depleted for maternal *Cal-opa* at NC12 and NC13 for *Cal-slp1* and at NC13 for *Cal-otd1*. This result suggests that the maternal Cal-Opa and zygotic Cal-Opa^{Zyg} may target some of the same genes and further suggests that the context dictates if the target gene is repressed or activated. Cal-Opa^{Zyg} repression for example, may depend on co-repressors. My analysis identified *Cal-sog* as a potential Cal-Opa^{Zyg} within the DV network. However, depletion of *Cal-opa^{Zyg}* causes increased expression of *Cal-sog* and increased accessibility of one *Cal-sog* ATAC-seq peak. In Drosophila *sog* depends on Opa for activation and to drive the accessibility of its distal enhancer. This result may be biological and reflect differing regulation of *Cal-sog* by Cal-Opa^{Zyg} in Clogmia or these results could reflect an indirect effect. Moreover, technical effects could be exacerbated given that the knockdown of the *Cal-opa^{Zyg}* transcript was moderate.

In conclusion, more work is required to fully appreciate the zygotic function of Cal-Opa^{Zyg} and to determine how conserved this function is with Opa in Drosophila. Evidence in Yoon et al. suggest that Cal-Opa^{Zyg}'s role in the late segmentation network is conserved, and this study identifies potential Cal-Opa^{Zyg} targets. However, many of the pair-rule and segment-polarity genes targeted by Opa in Drosophila are not significantly impacted in this study. I suspect that a significant reason for this is that the knockdown of *Cal-opa^{Zyg}* was moderate, unlike analogous experiments in Drosophila where Opa protein is abolished (Koromila et al., 2020; Soluri et al., 2020). Repeating these experiments in *Cal-opa^{Zyg}* RNAi with stronger knockdowns and identifying where Cal-Opa binds in the genome, will help to elucidate the full zygotic function of Cal-Opa^{Zyg}.

A.2 Figures and Tables

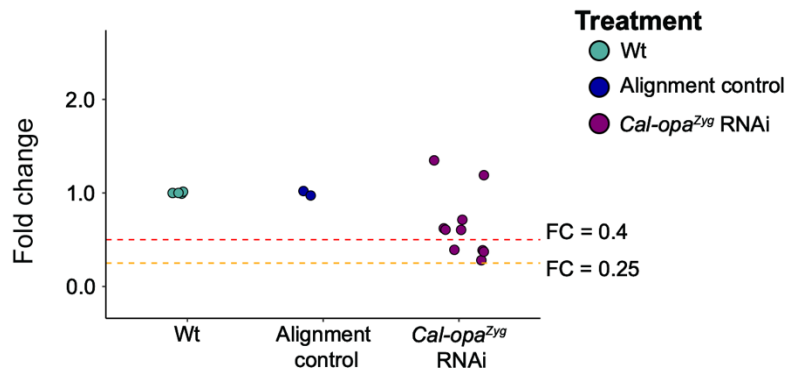

Figure A. 1 Identification of *Cal-opa^{Zyg}* RNAi embryos with strong knockdowns.
Plots of opa fold change for individual embryos at LNC14. Each embryo is represented by a circle. Embryos are grouped along the x-axis as indicated, spacing within a group was done arbitrarily for visualization. Treatment group of each embryo is indicated by color. Fold change (FC) for each embryo was calculated by dividing the embryo's CPM of Cal-opa^{Zyg} by the average of wild-type and alignment control Cal-opa CPMs. FC was calculated using embryos from a single female (same activation batch). Cal-opa^{Zyg} RNAi embryos with a FC < 0.4 (knockdown efficiency greater than 60%) were used for further analysis (n=4). Cal-opa^{Zyg} RNAi

embryos FC > 0.4 were excluded from further analysis. Orange dashed line, FC = 0.25. Red dashed line, FC = 0.4.

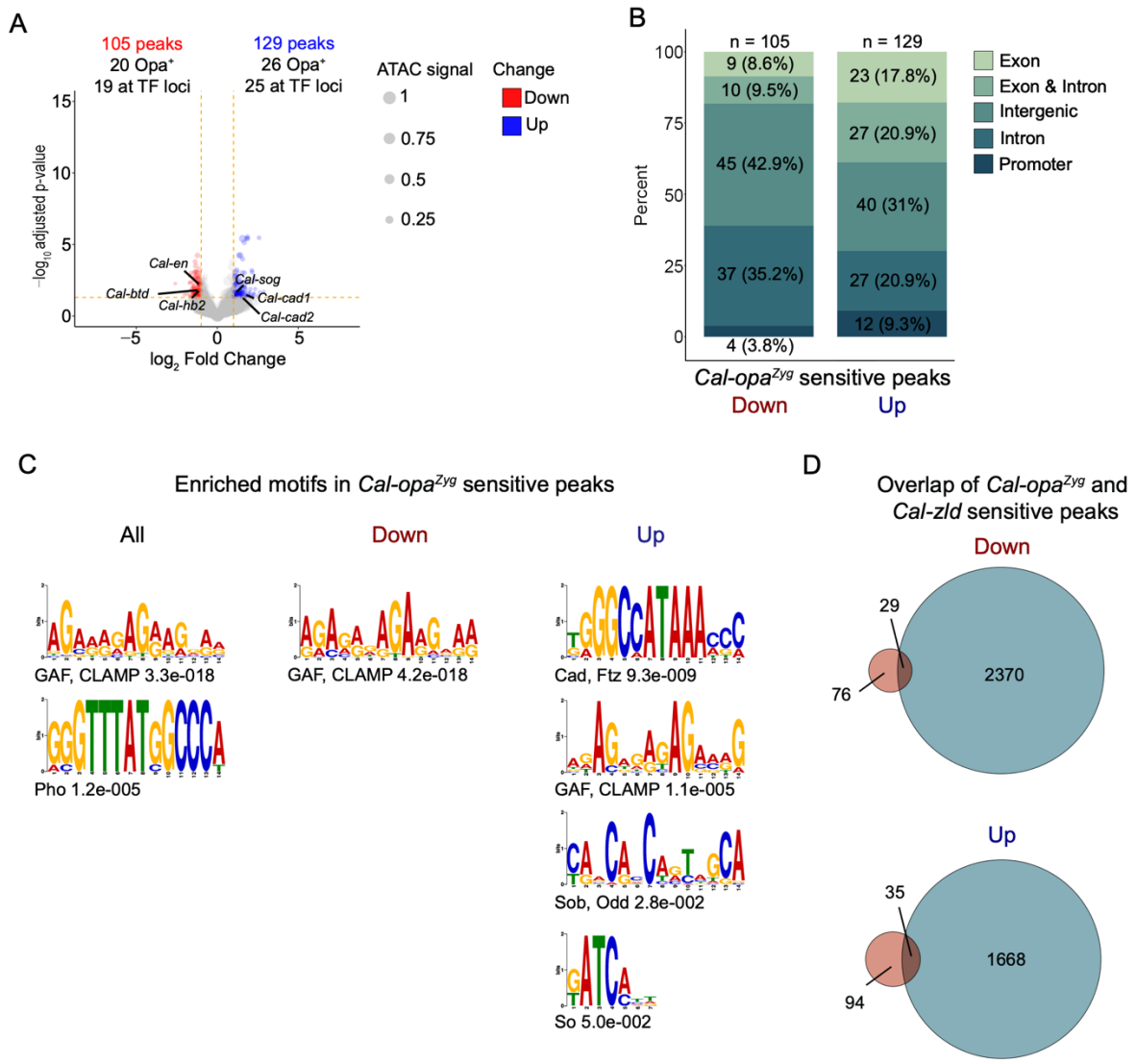

Figure A. 2 Cal-Opa^{Zyg} regulates chromatin accessibility positively and negatively.
A) Volcano plot of differentially accessible peaks between the *Cal*-opa^{Zyg} RNAi group and the alignment control group. Significant change of adjusted p-value ≤ 0.05 and |$\log_2$ Fold Change| ≥ 1 highlighted; red ($\log_2$ Fold Change ≤ -1) and blue ($\log_2$ Fold Change ≥ 1). Select developmental genes of interest whose loci contain *Cal*-opa^{Zyg} sensitive peaks are labeled. The number of peaks that contain an Opa motif (Opa+) and the number of peaks that fall within the locus of a TF are indicated by text. B) Genomic distribution of *Cal*-opa^{Zyg} sensitive peaks that were up-regulated (blue) and down-regulated (red). C) PWM logos of motifs enriched in all *Cal*-opa^{Zyg} sensitive peaks (n = 234), in *Cal*-opa^{Zyg} sensitive peaks that lose accessibility (n = 105), and in *Cal*-opa^{Zyg} sensitive peaks that gain accessibility (n = 129), as identified by the MEME suite. When available factors that bind to the motif or a similar motif are listed. E-value, denoting the significance of a motif is provided. D) Overlap of *Cal*-opa^{Zyg} (n = 105) and Cal-zld

(n = 2399) sensitive peaks whose accessibility decreases (Down, top). Overlap of *Cal*-opa^{Zyg} (n = 129) and Cal-zld (n = 1703) sensitive peaks whose accessibility increases (Up, bottom).

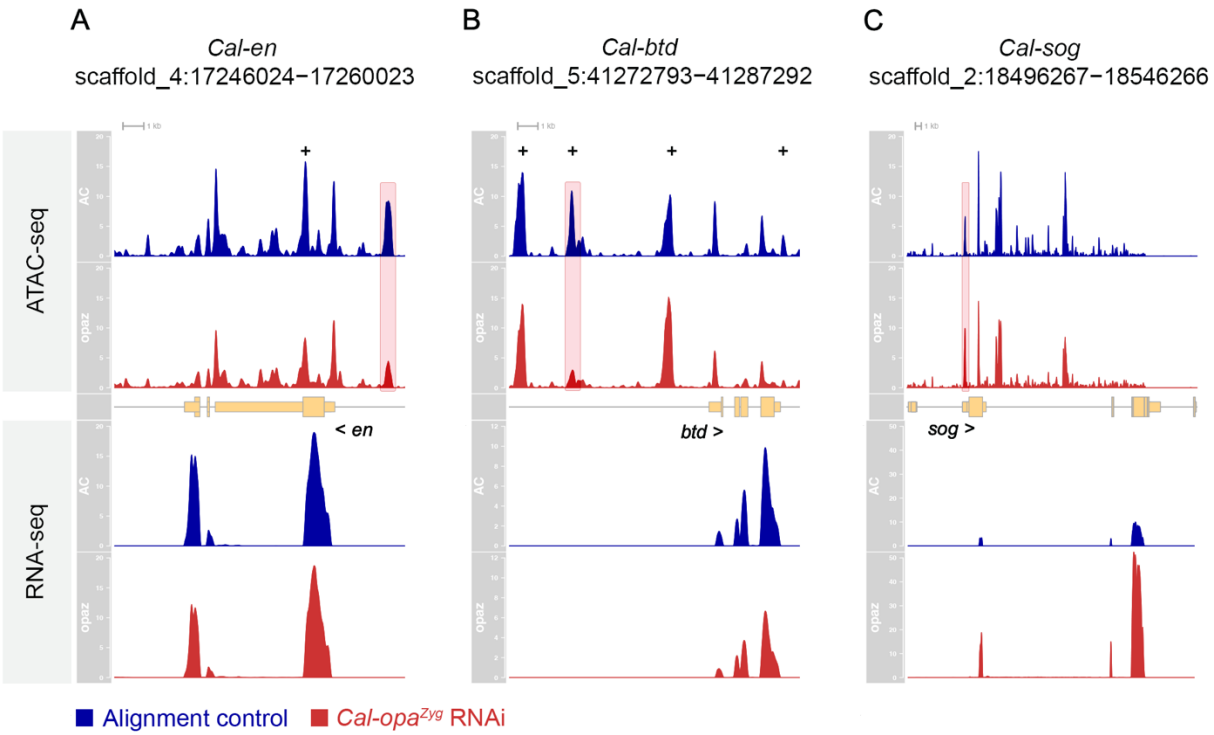

Figure A.3 Examples of the loci of three developmental genes that display *Cal*-opa^{Zyg} sensitive peaks.
Gene loci of A) *Cal-en*, B) *Cal-btd*, and C) *Cal-sog* at LNC14. Top tracks: CPM of ATAC-seq data) Middle tracks show gene models in pale yellow. Bottom tracks: CPM of RNA-seq data. Peaks with significant changes in accessibility are highlighted in pink. Peaks containing Opa motifs are marked with a plus sign. The y-axis is the same for all tracks of the same data type at each locus. *Cal-en* ATAC-seq: 0-20 CPM. *Cal-en* RNA-seq: 0-20 CPM. *Cal-btd* ATAC-seq: 0-20 CPM. *Cal-btd* RNA-seq: 0-12 CPM. *Cal-sog* ATAC-seq: 0-20 CPM. *Cal-sog* RNA-seq: 0-50 CPM. Blue: Alignment control. Red: *Cal*-opa^{Zyg} RNAi.

Table A. 1 Summary of Cal-Opa^{Zyg}'s effect on the transcription and chromatin accessibility of developmental genes of interest and other TF genes.
Shown are developmental genes of interests (red bar) and TF genes (blue bar) that become differentially expressed (DE) in *Cal*-opa^{Zyg} RNAi embryos or whose loci contain *Cal*-opa^{Zyg} sensitive peaks.

	Gene locus	Gene name or annotation name	Cal gene symb.	Flybase ID	DE in *Cal-opa^{Zyg}*	Change	Opa+ ATAC-seq peaks	*Cal-opa^{Zyg}* sensitive ATAC-seq peaks
	evm.TU.scaffold_1.1497	*Dichaete*	*Cal-D*	FBgn0000411			TRUE	
	evm.TU.scaffold_1.1986	*giant*	*Cal-gt*	FBgn0001150			TRUE	
	evm.TU.scaffold_1.788	*patched*	*Cal-ptc*	FBgn0003892			TRUE	
	evm.TU.scaffold_2.1058	*runt*	*Cal-run*	FBgn0003300			TRUE	
	evm.TU.scaffold_2.1090	*huckebein*	*Cal-hkb*	FBgn0261434			TRUE	
	evm.TU.scaffold_2.1785	*paired*	*Cal-prd*	FBgn0003145	TRUE	DOWN	TRUE	
	evm.TU.scaffold_2.281	*odd-skipped*	*Cal-odd*	FBgn0002985			TRUE	
	evm.TU.scaffold_2.871	*short gastrulation*	*Cal-sog*	FBgn0003463	TRUE	UP		TRUE
	evm.TU.scaffold_3.1360	*hedgehog*	*Cal-hh*	FBgn0004644			TRUE	
	evm.TU.scaffold_3.2548a	*tailless*	*Cal-tll1*	FBgn0003720			TRUE	
	evm.TU.scaffold_3.2548b	*tailless*	*Cal-tll2*	FBgn0003720	TRUE	UP	TRUE	

Table A.1, continued

	evm.TU.scaffold_3.2936	*fushi tarazu*	*Cal-ftz*	FBgn0001077			TRUE	
	evm.TU.scaffold_4.1097	*hairy*	*Cal-h*	FBgn0001168	TRUE	DOWN	TRUE	
	evm.TU.scaffold_4.1371	*empty spiracles*	*Cal-ems2*	FBgn0000576				
	evm.TU.scaffold_4.1372	*empty spiracles*	*Cal-ems3*	FBgn0000576				
	evm.TU.scaffold_4.1374	*empty spiracles*	*Cal-ems4*	FBgn0000576				
	evm.TU.scaffold_4.1375	*empty spiracles*	*Cal-ems1*	FBgn0000576			TRUE	
	evm.TU.scaffold_4.1575	*engrailed*	*Cal-en*	FBgn0000577			TRUE	TRUE
	evm.TU.scaffold_4.1647	*homeobrain*	*Cal-hbn*	FBgn0008636			TRUE	
	evm.TU.scaffold_4.40a	*even-skipped*	*Cal-eve1*	FBgn0000606			TRUE	
	evm.TU.scaffold_4.40b	*even-skipped*	*Cal-eve2*	FBgn0000606			TRUE	
	evm.TU.scaffold_5.1433	*wingless*	*Cal-wg*	FBgn0284084			TRUE	
	evm.TU.scaffold_5.1944	*buttonhead*	*Cal-btd*	FBgn0000233			TRUE	TRUE
	evm.TU.scaffold_5.235	*orthodenticle/ocelliless*	*Cal-otd1*	FBgn0004102	TRUE	UP	TRUE	
	evm.TU.scaffold_5.236	*orthodenticle/ocelliless*	*Cal-otd2*	FBgn0004102				
	evm.TU.scaffold_5.401	*caudal*	*Cal-cad1*	FBgn0000251			TRUE	TRUE

Table A.1, continued

	evm.TU.scaffold_5.403	*caudal*	*Cal-cad2*	FBgn0000251			TRUE	TRUE
	evm.TU.scaffold_5.983	*sloppy-paired*	*Cal-slp1*	FBgn0004567	TRUE	UP	TRUE	
	evm.TU.scaffold_5.985	*sloppy-paired*	*Cal-slp2*	FBgn0004567			TRUE	
	evm.TU.scaffold_5.987	*sloppy-paired*	*Cal-slp3*	FBgn0004567				
	evm.TU.scaffold_7.1074	*Krüppel*	*Cal-Kr*	FBgn0001325			TRUE	
	evm.TU.scaffold_7.154	*gooseberry*	*Cal-gsb*	FBgn0001148			TRUE	
	evm.TU.scaffold_7.155	*gooseberry-neuro*	*Cal-gsb-n*	FBgn0001147				
	evm.TU.scaffold_7.1783	*hunchback*	*Cal-hb1*	FBgn0001180				
	evm.TU.scaffold_7.1785	*hunchback*	*Cal-hb2*	FBgn0001180				TRUE
	evm.TU.scaffold_7.614	*zerknüllt*	*Cal-zen*	FBgn0004054			TRUE	
	evm.TU.scaffold_7.724	*knirps-like*	*Cal-kni/knrl2*	FBgn0001323			TRUE	
	evm.TU.scaffold_7.729	*knirps-like*	*Cal-kni/knrl1*	FBgn0001323			TRUE	
	evm.TU.scaffold_7.878	*odd-paired*	*Cal-opa*	FBgn0003002	TRUE	DOWN	TRUE	
	evm.TU.scaffold_1.1459	-	-	FBgn0039937			TRUE	TRUE
	evm.TU.scaffold_1.1761	*SCMH1*	*Cal-scmh1*	FBgn0003334				TRUE

Table A.1, continued

	evm.TU.scaffold_1.1933	*HLX*	*Cal-hlx*	FBgn0001170			TRUE	TRUE
	evm.TU.scaffold_1.1934	-	-	-			TRUE	TRUE
	evm.TU.scaffold_1.2123	*CIC*	*Cal-cic*	FBgn0262582			TRUE	TRUE
	evm.TU.scaffold_1.2253	*SSRP1*	*Cal-ssrp1*	FBgn0010278	TRUE	UP	TRUE	TRUE
	evm.TU.scaffold_1.2724	*ZNF384.2*	*Cal-znf384*	-	TRUE	UP	TRUE	
	evm.TU.scaffold_1.561	*CNOT8*	*Cal-cnot8*	FBgn0036239			TRUE	TRUE
	evm.TU.scaffold_1.562	-	-	-			TRUE	TRUE
	evm.TU.scaffold_2.1526	-	-	-			TRUE	TRUE
	evm.TU.scaffold_2.1788	*POU2F1*	*Cal-pou2f1*	FBgn0085424				TRUE
	evm.TU.scaffold_2.2347	*dl.2*	*Cal-dl.2*	FBgn0260632	TRUE	DOWN		
	evm.TU.scaffold_2.872	*SIM2*	*Cal-sim2*	FBgn0004666	TRUE	DOWN	TRUE	
	evm.TU.scaffold_2.878	*SUPT6H*	*Cal-supt6h*	FBgn0028982				TRUE
	evm.TU.scaffold_2.881	-	-	-				TRUE
	evm.TU.scaffold_3.1502	*FOXB1*	*Cal-foxb1*	FBgn0004897			TRUE	TRUE
	evm.TU.scaffold_3.1509	-	-	FBgn0032130				TRUE

Table A.1, continued

	evm.TU.scaffold_3.194	*SMAD7*	*Cal-smad7*	FBgn0020493	TRUE	UP	TRUE	
	evm.TU.scaffold_3.2255	-	-	-	TRUE	UP	TRUE	TRUE
	evm.TU.scaffold_3.2280	*topi*	*Cal-topi*	FBgn0037751				TRUE
	evm.TU.scaffold_3.2281	*topi.2*	*Cal-topi2*	-				TRUE
	evm.TU.scaffold_3.2330	*CEBPG*	*Cal-cebpg*	-	TRUE	UP		
	evm.TU.scaffold_3.2809	*WWC1*	*Cal-wwc1*	FBgn0262127	TRUE	UP	TRUE	
	evm.TU.scaffold_3.317	*ey*	*Cal-ey*	FBgn0005558			TRUE	TRUE
	evm.TU.scaffold_3.477	*Sox11*	*Cal-sox11*	FBgn0005612			TRUE	TRUE
	evm.TU.scaffold_3.696	*ATXN7L3*	*Cal-atxn7l3*	FBgn0036804				TRUE
	evm.TU.scaffold_4.1224	*ONECUT2*	*Cal-onecut2*	FBgn0028996			TRUE	TRUE
	evm.TU.scaffold_4.1225	*SOX6*	*Cal-sox6*	FBgn0039938			TRUE	TRUE
	evm.TU.scaffold_4.1441	*ETS1*	*Cal-ets1*	FBgn0003118			TRUE	TRUE
	evm.TU.scaffold_4.1449	*PRMT8*	*Cal-prmt8*	FBgn0037834				TRUE
	evm.TU.scaffold_4.1644	*Rx*	*Cal-rx*	FBgn0020617				TRUE
	evm.TU.scaffold_4.1842	-	-	-			TRUE	TRUE

Table A.1, continued

	evm.TU.scaffold_4.2089	*DPF2*	*Cal-dpf2*	FBgn0033015				TRUE
	evm.TU.scaffold_4.2097	-	-	FBgn0086680	TRUE	UP	TRUE	
	evm.TU.scaffold_4.2411	*dl.2*	*Cal-dl.2*	FBgn0260632			TRUE	TRUE
	evm.TU.scaffold_4.336	*BTF3L4*	*Cal-btf3l4*	FBgn0000181				TRUE
	evm.TU.scaffold_4.402	-	-	-			TRUE	TRUE
	evm.TU.scaffold_4.406	*SND1*	*Cal-snd1*	FBgn0035121			TRUE	TRUE
	evm.TU.scaffold_4.696	-	-	FBgn0052105	TRUE	DOWN	TRUE	
	evm.TU.scaffold_5.1145	-	-	-	TRUE	UP	TRUE	
	evm.TU.scaffold_5.1146	-	-	-	TRUE	UP	TRUE	
	evm.TU.scaffold_5.1204	*MEN1*	*Cal-men1*	FBgn0031885				TRUE
	evm.TU.scaffold_5.1536	*TBPL1*	*Cal-tbpl1*	FBgn0261793			TRUE	TRUE
	evm.TU.scaffold_5.265	-	-	FBgn0004915			TRUE	TRUE
	evm.TU.scaffold_5.480	-	-	-				TRUE
	evm.TU.scaffold_7.1127	*ELL*	*Cal-ell*	FBgn0014037	TRUE	UP	TRUE	
	evm.TU.scaffold_7.1276	*E75*	*Cal-e75*	FBgn0000568	TRUE	UP	TRUE	

Table A.1, continued

	evm.TU.scaffold_7.1522	*Eyg*	*Cal-eyg*	FBgn0000625	TRUE	DOWN	TRUE	
	evm.TU.scaffold_7.1551	*LHX1*	*Cal-lhx1*	FBgn0026411			TRUE	TRUE
	evm.TU.scaffold_7.1664	-	-	-	TRUE	DOWN	TRUE	
	evm.TU.scaffold_7.167	-	-	-				TRUE
	evm.TU.scaffold_7.2015	*SLC30A9*	*Cal-slc30a9*	FBgn0033762				TRUE
	evm.TU.scaffold_7.2227	*E2F4*	*Cal-e2f4*	FBgn0024371				TRUE
	evm.TU.scaffold_7.2237	*MED4*	*Cal-med4*	FBgn0035754				TRUE
	evm.TU.scaffold_7.2594	-	-	FBgn0034643			TRUE	TRUE
	evm.TU.scaffold_7.2612	*PHF20*	*Cal-phf20*	FBgn0038016			TRUE	TRUE
	evm.TU.scaffold_7.844	*BBX*	*Cal-bbx*	FBgn0024251	TRUE	DOWN	TRUE	
	evm.TU.scaffold_7.913	*MSL3*	*Cal-msl3*	FBgn0002775			TRUE	TRUE

APPENDIX B: DETAILED ATAC-SEQ LIBRARY PREPARATION PROTOCOL

B.1 Standard ATAC-seq protocol

This protocol is based on the Blythe protocol (Blythe & Wieschaus, 2016). The Blythe protocol is very similar to the Buenrostro protocol (Buenrostro et al., 2015) except that it has been optimized for the *Drosophila melanogaster* embryo.

B.1.1 Key points before starting

- Spend time with your system and the embryos you are working with. Develop a way to reproducibly stage your system. Develop tactile skills to handle your embryos as there are steps early in the protocol that will test your fine motor coordination.
- The amount of Tn5 enzyme needs to be optimized. 2.5 µl will likely be sufficient but it is best practice to confirm this before running an actual experiment. Factors such as the number of nuclei, genome size, effectiveness of lysing the embryo and collection of the nuclear pellet, etc., all affect to the success of the tagmentation reaction.
- Any 1.5 ml tube used in this protocol should be an Eppendorf DNA Lobind tube (Catalog No. 0030108310)

B.1.2 To do before starting

- Prepare glass pestles. Glass pestles are used to homogenize the embryo in step 1. Fire one end of a Drummond microcapillary using a Bunsen burner. It will only take a few moments for the end to start to melt and close forming a tear dropped shaped bulb. This is

the desired end point; the bulbed end will be used to crush the embryo. Make sure the knobbed end is completely sealed prior to use. Glass pestles can be prepared in mass and saved. They can also be baked to make them RNase free.

- Pre-cool the centrifuge to 4°C
- Set the thermomixer to 37°C
- Get an ice bucket

B.1.3 Materials

- Drummond microcapillary (1-000-0200 or 1-000-0250)
- Eppendorf DNA Lobind tube (Catalog No. 0030108310)
- Filtered low retention tips (any brand you like)
- Eppendorf Thermomixer 5350 Mixer
- Magnet stand Invitrogen DynaMag™-2 Magnet (12321D) or comparable available on ebay
- NEBNext High-Fidelity 2x PCR Master Mix (M0541)
- Qiagen MinElute Reaction Cleanup Kit (28204)
- SYBR Green solution (Invitrogen S7563)
- Molecular biology grade ethanol

B.1.4 Protocol

1. Lyse the embryo and collect the nuclear pellet
    a. If you are working with embryos that have been injected, see Section B.2.
    b. It is advisable to perform this step under a dissection scope.

c. Add 40 µl of ice-cold lysis buffer to bottom of the tube of a 1.5 µl tube.
d. To the same tube place an embryo in the lid of the 1.5 µl tube. You can use a fine paint brush, a loop made from tungsten wire, or you can pipette the embryo using a 20 µl pipette. If you are pipetting the embryo you need to take care that you do not damage the embryo and ensure the embryo does not get stuck in the pipette tip. You also need to carefully remove any of the liquid that transferred with the embryo before adding the lysis buffer.
e. Add 10 µl of lysis buffer over the embryo and thoroughly homogenize the embryo using the glass pestle.
f. Carefully and quickly close the lid. Make sure NO liquid spills.
g. Alternatively, if this is too difficult. remove the cap of a 1.5 µl tube and lyse the embryo in the cap without the tube. Carefully, transfer the cap to a new 1.5 µl tube containing the remaining 40 µl of lysis buffer. Make sure to label both lids. You can get rid of the lid that is not attached to the tube after the centrifugation step and once you have confirmed that you have a nuclear pellet (in step 2).
h. Once closed you can gently invert the tube up and down to wash off any residual liquid on the lid. Do not introduce bubbles.
i. While you are processing multiple samples you want to keep the tubes on ice. I like to keep the tube inverted in the ice as the embryo was lysed in the cap and I want to make sure that everything goes into solution.

Table B. 1 Lysis buffer recipe

Lysis buffer (5 ml) 10 mM Tris-HCl pH 7.4; 10 mM NaCl; 3 mM MgCl2; 0.1% Igepal CA-630			
Component	Desired []	Stock []	Volume of stock to add
10mM Tris HCl ph 7.4	0.01M	1M	50 ul
10mM NaCl	0.01M	2M	25 ul
3mM MgCl2	0.003M	2M	7.5 µl
0.1% (v/v) Igepal CA-630	0.1%	100%	5 µl
Molecular biology grade H2O			To 5 ml

2. Pellet the nuclei by centrifugation at 4°C for 10 minutes.
    a. I have tried both 500 and 800 RCF either works for Clogmia embryos.
    b. Once complete place the tubes back on ice carefully as not to dislodge the nuclear pellet.

3. Tagmentation reaction
    a. Remove the supernatant. Remove as much as possible without losing the nucellar pellet.
    b. Remove the supernatant under a dissection scope to make sure that the nuclear pellet is not lost.
    c. Keep the supernatant if you want to isolate RNA from the prep as well, see Section B.3.
    d. Resuspend the nuclear pellet in 7.5 µl of tagmentation master mix.
    e. Resuspending the pellet is key. You want to pipette up and down while running the tip circularly along the bottom. However, you do not want to make the solution foamy or introduce bubbles.
    f. Once all samples have been resuspended add the 2.5 µl of Tn5 enzyme.
    g. Carefully pipette up and down.

h. Add all tubes to the thermomixer. Incubate at 37°C with 800 RPM shaking for 30 minutes.

Table B. 2 Tagmentation reaction mix

Reaction mix	
2x TD reaction buffer	5 µl
TDE1 Tn5 transposase enzyme	2.5 µl
Nuclease- free H2O	2.5 µl

4. Purify Tagmentation reaction with the Qiagen Minelute purification
   a. Overall, you follow the manufactures instructions including the additional points here. All centrifugation steps are done at RT at the speed recommended by the manufacturer. Alternatively, you can use a vacuum manifold.
   b. First add 300 µl ERC to the 10 µl of tagmented DNA and transfer to the minelute column as instructed.
   c. Wash 2x with 750 µl of buffer PE. On the second wash incubate for 3 minutes. When adding buffers make sure to wash the walls of the column.
   d. Following the last wash make sure you do an additional spin with an unused collection column. This is key to get rid of residual ethanol.
   e. Elute with 10 µl of EB (elution buffer). Incubate for 3 min.
   f. If necessary, the purified tagmented DNA can be stored overnight at -20°C.

5. PCR Pre-Amplification
   a. This is a critical step **do not skip** or **modify** the PCR cycle.
   b. The initial 5-minute extension covalently links the inserted adapter to the genomic DNA and ensures that there is a continuous DNA fragment that represents the transposition

site. After the tagmentation reaction, before the extension, only one end of the inserted adapter is covalently linked to the genomic DNA. If one was to denature the DNA *before the extension*, the priming site for the PCR would float away. Extending the transposed DNA ensures that there is a continuous DNA fragment that represents the transposition site.

c. Stick with the NEBNext High-Fidelity 2x PCR master mix. Do not use a hot-start polymerase

d. There are 5 PCR cycles. Following the preamplification proceed to the qPCR step.

Table B. 3 PCR Pre-Amplification reaction mix and cycle

Reaction mix		
Transposed DNA sample	10 µl	
Nuclease-free H2O	13.5 µl	
25 µM Unique Dual-Index Primer Mix	1.5 µl	
NEBNext Q5 2x PCR master mix	25 µl	
Total volume **one** reaction	50 µl	
PCR cycle		
1 cycle	5 min	72°C
5 cycles	30 sec	98°C
	10 sec	98°C
	30 sec	63°C
	1 min	72°C

6. qPCR side reaction

a. The qPCR will determine the additional numbers of cycles required to amplify the tagmented DNA that will be used for the PCR amplification.

b. You will use 5 µl from the pre-amplified reaction for the qPCR. Keep the remaining 45 at 4°C while the qPCR is running.

c. To determine the number of cycles, plot the linear fluorescence intensity (Rn) versus the cycle number. Find the maximum fluorescence intensity value (the plateau) then find the

cycle number that corresponds to 1/3 the maximum intensity value. An additional 7-9 cycles are typical.

Table B. 4 qPCR reaction mix and cycle

Reaction mix		
PCR Pre-amplification product	5 µl	
Nuclease- free H2O	3.91 µl	
6.25 µM Unique Dual-Index Primer Mix *	1 µl	
NEBNext Q5 2x PCR master mix	5 µl	
SYBR Green 100x	0.09 µl	
Total volume **one** reaction	15 µl	
*Dilute the 25 µM primer mix for the qPCR reaction		
PCR cycle		
1 cycle	30 sec	98°C
20 cycles	10 sec	98°C
	30 sec	63°C
	1 min	72°C

7. Complete PCR Amplification
    a. Remove the remaining 45 µl of the reaction and place it into the PCR machine.
    b. Perform the additional number of cycles determined in step 6.
    c. Depending on the additional cycles required for each sample you may have to run multiple rounds of PCR. If so, keep the other samples at 4°C until all samples are done.

Table B. 5 Final PCR cycle

PCR cycle		
1 cycle	30 sec	98°C
N cycles	10 sec	98°C
	30 sec	63°C
	1 min	72°C

8. Purify the PCR Reaction
    a. Clean up with Ampure following the standard protocol.
    b. Before starting make sure the beads have been warmed to RT.

c. Transfer the 45 µl of the reaction to a lobind tube.

d. Add 1.8X of Ampure beads to each reaction. Mix thoroughly. Incubate for 5 minutes at RT.

e. Add the tubes to a magnet stand. Let the beads separate from the solution for 1 minute. Remove the supernatant

f. Wash 2X with **fresh** 80% EtOH without removing from the magnetic stand.

g. Dry the beads. This will take ~5 minutes. Do not over dry the beads so that they start to crack.

h. Resuspend beads in 15 µl Buffer EB. Incubate for 3 minutes

i. Place the tubes back on the stand. Let the beads separate from the solution for 1 minute. Keep the elutant, transfer to a new labeled tube.

9. Library QC

a. Measure the library concentration with a Qubit fluorometer.

b. Check the size distribution using a Bioanalyzer High Sensitivity Chip. Depending on what you have access to a fragment analyzer will also do. Depending on the library concentration prepare it neat or a dilute it 1:5.

c. You are expecting to see a sharp low molecular weight peak ~200 bp and 2 or more undulations of higher molecular weight fragments corresponding to nucleosomal DNA. Ideally the lower molecular weight peak is at a higher concentration than the higher molecular weight fragments. Additionally, a good library does not have a high concentration of primer peaks. If so, you may need to repeat the Ampure purification and may need to optimize the primer concentration in the PCR pre-amplification step.

d. The image below is from a bioanalyzer trace of a sequence-able ATAC-seq library. The ladder peaks are at 35 bp and 10380. The numbers above the peaks (in green or purple respectively) indicate the migration time. The low molecular weight peak is at ~200 bp with a time of 57.83. Between the ~200 bp peak and the lower ladder peak are primer peaks that can be removed by a second round of Ampure purification. The high molecular weight undulations start at ~350 bp.

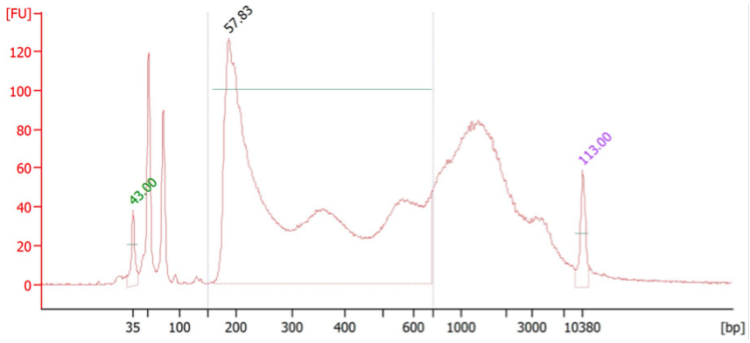

Figure B. 1 Example bioanalyzer trace of an ATAC-seq library.

B.2 Preparing injected embryos for ATAC-seq library prep

This section is based on working with Clogmia embryos. You may have to find some work arounds depending on the injection setup for your system. For injection, Clogmia embryos are affixed to a coverslip with double sided tape and covered with halocarbon oil. It is important to wash off as much of the injection oil before starting the ATAC-seq library prep.

1. First use a vacuum to gently suction away as much oil as possible. Then wash the injection coverslip with embryos by submerging it in a petri dish with H2O, rock the dish back and forth gently. Remove the coverslip from the dish, replace the water and place

the cover slip back into the dish. Repeat 4-5 times until you see that the oil has been removed. A well-placed embryo does dislodge during the washes. It is important to check under the dissection scope that the embryos are staying put while washing.

    - You can use filter paper to try to wick away oil. You can also briefly place the embryo on filter paper before transferring it to a tube for lysing. If you choose to do so you must work quickly so the embryo does not dry out.

2. After washing, remove the embryos from the tape. To remove the embryos from the tape you want to use a blunted tungsten wire. Push the wire under the embryo to separate it from the tape. You should avoid pushing on the embryo directly, rather you are trying to put space between the embryo at the tape using the tungsten wire. Do not let the embryos dry out.

3. Transfer the embryos. You want to avoid using a paint brush to transfer embryos if they have come in contact with oil. In this case, the brush can transfer residual oil along with the embryo. Instead use a tungsten loop or wire to transfer the embryo. The embryos generally like to stick to the wire and will come off when you delicately place them on the inside of tube. This may take some practice. Alternatively, transfer the embryos to a dish with water and carefully pipette them to individual tubes.

B.3 Recovering RNA from an ATAC-seq library prep

After lysing the embryo and pelleting the nuclei (step 1 and 2) RNA can be purified from the supernatant. This section uses a Zymo RNA isolation kit and modifies steps in the protocol to isolate RNA.

B.3.1 Additional Materials

- Zymo Quick-RNA Tissue/Insect Microprep (R2030)
- This protocol should also be compatible with the Zymo Quick-RNA Microprep Kit (R1050)
- Zymo DNase I set (E1010)

B.3.2 Protocol

Perform all centrifugations at 10,000-16,000 x g for 1 minute at room temperature.

1. Collect as much supernatant as possible (45-50 µl) and mix with 450 µl of Zymo RNA lysis buffer.
    a. Pipette up and down several times to ensure complete mixing.
    b. This is equivalent to step 3 in the manufacture's protocol
    c. At this point the RNA can be stored in lysis buffer at -20°C. The longest I have stored samples is a couple days.

2. Transfer the cleared supernatant into a Zymo-Spin™ IIICG Column1 (green) in a collection tube and centrifuge. **Save the flow-through**.
3. Add an equal volume (1:1) of molecular biology grade ethanol (100%) to the flow-through and mix by pipetting up and down.
4. Transfer the mixture into a Zymo-Spin™ IC Column2 in a collection tube and centrifuge. Discard the flow-through.
5. Perform the DNase treatment.
    a. Add 400 µl RNA Wash Buffer to the column, centrifuge and discard the flow-through.
    b. Prepare the DNase reaction mix and add directly to the column.
    c. Incubate for 15 min at RT.

Table B. 6 DNAse reaction mix

DNase reaction mix	
DNase I (reconstituted 1 U/µl)	5 µl
DNA digestion buffer	35 µl

6. Add 400 µl RNA Prep Buffer to the column and centrifuge. Discard the flow-through.
7. Add 700 µl RNA Wash Buffer to the column and centrifuge. Discard the flow-through.
8. Add 400 µl RNA Wash Buffer and centrifuge. Discard the flow-through.
9. Ensure complete removal of the wash buffer. Place the column in a new collection tube and centrifuge.
10. Elute the RNA. Place the collection tube into a labeled 1.5 ml tube. Add 6-15 µl of DNase/RNase free molecular biology grade water. Incubate for 3 minutes at RT. Centrifuge.
    a. Purified RNA can be used immediately or stored at -20°C/-80°C.

REFERENCES

Amabis, J. M. (1977). Cytological evidence of male sex heterogamety in sex determination of Telmatoscopus albipunctatus (Diptera: Psychodidae). *Chromosoma*, *62*, 133–138.

Amabis, J. M., & Simoes, L. C. G. (1972). Chromosome studies in a species of Telmatoscopus sp. *Caryologia*, *25*(2).

Amiri, E. E., Tenger-Trolander, A., Li, M., Julian, A. T., Kasan, K., Sanders, S. A., Blythe, S., & Schmidt-Ott, U. (2025). Conservation of symmetry breaking at the level of chromatin accessibility between fly species with unrelated anterior determinants. *BioRxiv*, 2025.01.13.632851. https://doi.org/10.1101/2025.01.13.632851

Amodeo, A. A., Jukam, D., Straight, A. F., & Skotheim, J. M. (2015). Histone titration against the genome sets the DNA-to-cytoplasm threshold for the Xenopus midblastula transition. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(10), E1086–E1095. https://doi.org/10.1073/PNAS.1413990112/SUPPL_FILE/PNAS.1413990112.SM03.AVI

Anderson, D. T. (1966). Comparative Embryology of Diptera. *Annual Review of Entomology*, *11*, 23-+. https://doi.org/DOI 10.1146/annurev.en.11.010166.000323

Andrioli, L. P. M., Visisht, V., Theodosopoulou, E., Oberstein, A., & Small, S. (2002). Anterior repression of a Drosophila stripe enhancer requires three position-specific mechanisms. *Development (Cambridge, England)*, *129*(21), 4931–4940. https://doi.org/10.1242/DEV.129.21.4931

Ansari, S., Troelenberg, N., Dao, V. A., Richter, T., Bucher, G., & Klingler, M. (2018). Double abdomen in a short-germ insect: Zygotic control of axis formation revealed in the beetle

Tribolium castaneum. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(8), 1819–1824. https://doi.org/10.1073/pnas.1716512115

Arnosti, D. N., Barolo, S., Levine, M., & Small, S. (1996). The eve stripe 2 enhancer employs multiple modes of transcriptional synergy. *Development (Cambridge, England)*, *122*(1), 205–214. https://doi.org/10.1242/DEV.122.1.205

Arsala, D., & Lynch, J. A. (2017). Ploidy has little effect on timing early embryonic events in the haplo-diploid wasp Nasonia. *Genesis*, *55*(5). https://doi.org/10.1002/dvg.23029

Aruga, J., & Millen, K. J. (2018). ZIC1 Function in Normal Cerebellar Development and Human Developmental Pathology. *Advances in Experimental Medicine and Biology*, *1046*, 249–268. https://doi.org/10.1007/978-981-10-7311-3_13

Austin, M., Clifford, C., Rutt, G., Price, B. W., Natural History Museum Genome Acquisition, L., Wellcome Sanger Institute Tree of Life, programme, Wellcome Sanger Institute Scientific Operations, D. N. A. P. collective, Tree of Life Core Informatics, collective, Wallace, I., & Darwin Tree of Life, C. (2023). The genome sequence of a caddisfly, Limnephilus lunatus (Curtis, 1834). *Wellcome Open Res*, *8*, 25. https://doi.org/10.12688/wellcomeopenres.18752.1

Bailey, T. L. (2021). STREME: accurate and versatile sequence motif discovery. *Bioinformatics*, *37*(18), 2834–2840. https://doi.org/10.1093/bioinformatics/btab203

Bailey, T. L., Johnson, J., Grant, C. E., & Noble, W. S. (2015). The MEME Suite. *Nucleic Acids Res*, *43*(W1), W39-49. https://doi.org/10.1093/nar/gkv416

Bartlett, A., O'Malley, R. C., Huang, S. S. C., Galli, M., Nery, J. R., Gallavotti, A., & Ecker, J. R. (2017). Mapping genome-wide transcription-factor binding sites using DAP-seq. *Nature Protocols 2017 12:8*, *12*(8), 1659–1672. https://doi.org/10.1038/nprot.2017.055

Baumgartner, S., & Noll, M. (1990). Network of interactions among pair-rule genes regulating paired expression during primordial segmentation of Drosophila. *Mechanisms of Development*, *33*(1), 1–18. https://doi.org/10.1016/0925-4773(90)90130-E

Beeber, D., & Chain, F. J. (2020). crispRdesignR: A Versatile Guide RNA Design Package in R for CRISPR/Cas9 Applications. *J Genomics*, *8*, 62–70. https://doi.org/10.7150/jgen.41196

Benedyk, M. J., Mullen, J. R., & DiNardo, S. (1994). odd-paired: a zinc finger pair-rule protein required for the timely activation of engrailed and wingless in Drosophila embryos. *Genes & Development*, *8*(1), 105–117. https://doi.org/10.1101/GAD.8.1.105

Bergman, C. M., Carlson, J. W., & Celniker, S. E. (2005). Drosophila DNase I footprint database: a systematic genome annotation of transcription factor binding sites in the fruitfly, Drosophila melanogaster. *Bioinformatics*, *21*(8), 1747–1749. https://doi.org/10.1093/bioinformatics/bti173

Berleth, T., Burri, M., Thoma, G., Bopp, D., Richstein, S., Frigerio, G., Noll, M., & Nusslein-Volhard, C. (1988). The role of localization of bicoid RNA in organizing the anterior pattern of the Drosophila embryo. *EMBO J*, *7*(6), 1749–1756. https://www.ncbi.nlm.nih.gov/pubmed/2901954

Berman, B. P., Nibu, Y., Pfeiffer, B. D., Tomancak, P., Celniker, S. E., Levine, M., Rubin, G. M., & Eisen, M. B. (2002). Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the Drosophila genome. *Proc Natl Acad Sci U S A*, *99*(2), 757–762. https://doi.org/10.1073/pnas.231608898

Biedler, J. K., Hu, W., Tae, H., & Tu, Z. (2012). Identification of early zygotic genes in the yellow fever mosquito Aedes aegypti and discovery of a motif involved in early zygotic genome activation. *PLoS One*, *7*(3), e33933. https://doi.org/10.1371/journal.pone.0033933

Bishop, T. R., Onal, P., Xu, Z., Zheng, M., Gunasinghe, H., Nien, C. Y., Small, S., & Datta, R. R. (2023). Multi-level regulation of even-skipped stripes by the ubiquitous factor Zelda. *Development (Cambridge, England)*, *150*(23). https://doi.org/10.1242/DEV.201860

Blumenthal, A. B., Kriegstein, H. J., & Hogness, D. S. (1974). The units of DNA replication in Drosophila melanogaster chromosomes. *Cold Spring Harbor Symposia on Quantitative Biology*, *38*, 205–223. https://doi.org/10.1101/SQB.1974.038.01.024

Blythe, S. A., & Wieschaus, E. F. (2015a). Coordinating Cell Cycle Remodeling with Transcriptional Activation at the Drosophila MBT. In *Current Topics in Developmental Biology* (1st ed., Vol. 113). Elsevier Inc. https://doi.org/10.1016/bs.ctdb.2015.06.002

Blythe, S. A., & Wieschaus, E. F. (2015b). Zygotic genome activation triggers the DNA replication checkpoint at the midblastula transition. *Cell*, *160*(6), 1169–1181. https://doi.org/10.1016/j.cell.2015.01.050

Blythe, S. A., & Wieschaus, E. F. (2016). Establishment and maintenance of heritable chromatin structure during early drosophila embryogenesis. *ELife*, *5*(NOVEMBER2016), 1–21. https://doi.org/10.7554/eLife.20148

Boerner, T. J., Deems, S., Furlani, T. R., Knuth, S. L., & Towns, J. (2023). ACCESS: Advancing Innovation: NSF's Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support. In *Practice and Experience in Advanced Research Computing (PEARC '23)*. Association for Computing Machinery. https://doi.org/10.1145/3569951.3597559

Bonneton, F., Shaw, P. J., Fazakerley, C., Min, S., & Dover, G. A. (1997). Comparison of bicoid-dependent regulation of hunchback between Musca domestica and Drosophila melanogaster. *Mechanisms of Development*, *66*(1–2), 143–156. https://doi.org/10.1016/S0925-4773(97)00100-7

Bothma, J. P., Garcia, H. G., Esposito, E., Schlissel, G., Gregor, T., & Levine, M. (2014). *Dynamic regulation of eve stripe 2 expression reveals transcriptional bursts in living Drosophila embryos*. *111*(29). https://doi.org/10.1073/pnas.1410022111

Bozek, M., Cortini, R., Storti, A. E., Unnerstall, U., Gaul, U., & Gompel, N. (2019). ATAC-seq reveals regional differences in enhancer accessibility during the establishment of spatial coordinates in the Drosophila blastoderm. *Genome Research*, *29*(5), 771–783. https://doi.org/10.1101/gr.242362.118

Bozek, M., & Gompel, N. (2020). Developmental Transcriptional Enhancers: A Subtle Interplay between Accessibility and Activity: Considering Quantitative Accessibility Changes between Different Regulatory States of an Enhancer Deconvolutes the Complex Relationship between Accessibility a. *BioEssays*, *42*(4), 1–11. https://doi.org/10.1002/bies.201900188

Brennan, K. J., Weilert, M., Krueger, S., Pampari, A., Liu, H. yun, Yang, A. W. H., Morrison, J. A., Hughes, T. R., Rushlow, C. A., Kundaje, A., & Zeitlinger, J. (2023). Chromatin accessibility in the Drosophila embryo is determined by transcription factor pioneering and enhancer activation. *Developmental Cell*, *58*(19), 1898-1916.e9. https://doi.org/10.1016/J.DEVCEL.2023.07.007

Brooks, E. M., Sanders, S. A., & Pfrender, M. E. (2024). FreeCount: A coding-free framework for guided count data visualization and analysis. In *Practice and Experience in Advanced Research Computing 2024: Human Powered Computing (PEARC '24)*. Association for Computing Machinery. https://doi.org/10.1145/3626203.3670605

Brown, J. L., Sonoda, S., Ueda, H., Scott, M. P., & Wu, C. (1991). Repression of the Drosophila fushi tarazu (ftz) segmentation gene. *The EMBO Journal*, *10*(3), 665–674. https://doi.org/10.1002/J.1460-2075.1991.TB07995.X

Brown, J. L., & Wu, C. (1993). Repression of Drosophila pair-rule segmentation genes by ectopic expression of tramtrack. *Development*, *117*(1), 45–58. https://doi.org/10.1242/DEV.117.1.45

Bruce, H. S., Jerz, G., Kelly, S. R., McCarthy, J., Pomerantz, A., Senevirathne, G., Sherrard, A., Sun, D. A., Wolff, C., & Patel, N. H. (2021). Hybridization chain reaction (HCR) in situ protocol V.1. In *Protocols.io*. https://doi.org/doi.org/10.17504/protocols.io.bunznvf6

Brunner, E., Peter, O., Schweizer, L., & Basler, K. (1997). pangolin encodes a Lef-1 homologue that acts downstream of Armadillo to transduce the Wingless signal in Drosophila. *Nature*, *385*(6619), 829–833. https://doi.org/10.1038/385829A0

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., & Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods*, *10*(12), 1213–1218. https://doi.org/10.1038/nmeth.2688

Buenrostro, J. D., Wu, B., Chang, H. Y., & Greenleaf, W. J. (2015). ATAC-seq: A method for assaying chromatin accessibility genome-wide. *Current Protocols in Molecular Biology*, *2015*, 21.29.1-21.29.9. https://doi.org/10.1002/0471142727.mb2129s109

Buescher, M., Svendsen, P. C., Tio, M., Miskolczi-McCallum, C., Tear, G., Brook, W. J., & Chia, W. (2004). Drosophila T box proteins break the symmetry of hedgehog-dependent activation of wingless. *Current Biology*, *14*(19), 1694–1702. https://doi.org/10.1016/j.cub.2004.09.048

Bullock, S. L., Stauber, M., Prell, A., Hughes, J. R., Ish-Horowicz, D., & Schmidt-Ott, U. (2004). Differential cytoplasmic mRNA localization adjusts pair-rule transcription factor activity to cytoarchitecture in dipteran evolution. *Development*, *131*, 4251–4261.

Bushati, N., Stark, A., Brennecke, J., & Cohen, S. M. (2008). Temporal reciprocity of miRNAs and their targets during the maternal-to-zygotic transition in Drosophila. *Current Biology : CB*, *18*(7), 501–506. https://doi.org/10.1016/J.CUB.2008.02.081

Cadigan, K. M., Grossniklaus, U., & Gehring, W. J. (1994a). Functional redundancy: the respective roles of the two sloppy paired genes in Drosophila segmentation. *Proc Natl Acad Sci U S A*, *91*(14), 6324–6328. https://doi.org/10.1073/pnas.91.14.6324

Cadigan, K. M., Grossniklaus, U., & Gehring, W. J. (1994b). Localized expression of sloppy paired protein maintains the polarity of Drosophila parasegments. *Genes & Development*, *8*(8), 899–913. https://doi.org/10.1101/GAD.8.8.899

Calderon, D., Blecher-Gonen, R., Huang, X., Secchia, S., Kentro, J., Daza, R. M., Martin, B., Dulja, A., Schaub, C., Trapnell, C., Larschan, E., O'Connor-Giles, K. M., Furlong, E. E. M., & Shendure, J. (2022). The continuum of Drosophila embryonic development at single-cell resolution. *Science*, *377*(6606). https://doi.org/10.1126/SCIENCE.ABN5800/SUPPL_FILE/SCIENCE.ABN5800_MDAR_REPRODUCIBILITY_CHECKLIST.PDF

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, *10*, 421. https://doi.org/10.1186/1471-2105-10-421

Campos-Ortega, J. A., & Hartenstein, V. (1997). *The embryonic development of Drosophila melanogaster* (2nd ed.). Springer-Verlag.

Cantalapiedra, C. P., Hernandez-Plaza, A., Letunic, I., Bork, P., & Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol Biol Evol*, *38*(12), 5825–5829. https://doi.org/10.1093/molbev/msab293

Cao, W. X., Kabelitz, S., Gupta, M., Yeung, E., Lin, S., Rammelt, C., Ihling, C., Pekovic, F., Low, T. C. H., Siddiqui, N. U., Cheng, M. H. K., Angers, S., Smibert, C. A., Wühr, M., Wahle, E., & Lipshitz, H. D. (2020). Precise Temporal Regulation of Post-transcriptional Repressors Is Required for an Orderly Drosophila Maternal-to-Zygotic Transition. *Cell Reports*, *31*(12). https://doi.org/10.1016/J.CELREP.2020.107783

Cardamone, F., Piva, A., Löser, E., Eichenberger, B., Romero-Mulero, M. C., Zenk, F., Shields, E. J., Cabezas-Wallscheid, N., Bonasio, R., Tiana, G., Zhan, Y., & Iovino, N. (2025). Chromatin landscape at cis-regulatory elements orchestrates cell fate decisions in early embryogenesis. *Nature Communications 2025 16:1*, *16*(1), 1–22. https://doi.org/10.1038/s41467-025-57719-4

Caroti, F., Gonzalez Avalos, E., Noeske, V., Gonzalez Avalos, P., Kromm, D., Wosch, M., Schutz, L., Hufnagel, L., & Lemke, S. (2018). Decoupling from yolk sac is required for extraembryonic tissue spreading in the scuttle fly Megaselia abdita. *Elife*, *7*. https://doi.org/10.7554/eLife.34616

Casanova, J., & Struhl, G. (1989). Localized surface activity of torso, a receptor tyrosine kinase, specifies terminal body pattern in Drosophila. *Genes & Development*, *3*(12B), 2025–2038. https://doi.org/10.1101/GAD.3.12B.2025

Chari, S., Wilky, H., Govindan, J., & Amodeo, A. A. (2019). Histone concentration regulates the cell cycle and transcription in early development. *Development (Cambridge, England)*, *146*(19). https://doi.org/10.1242/DEV.177402

Cheatle Jarvela, A. M., Trelstad, C. S., & Pick, L. (2020). Regulatory gene function handoff allows essential gene loss in mosquitoes. *Communications Biology 2020 3:1*, *3*(1), 1–10. https://doi.org/10.1038/s42003-020-01203-w

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*, *2013*(2). https://doi.org/10.7554/ELIFE.00861

Chen, L., Dumelie, J. G., Li, X., Cheng, M. H. K., Yang, Z., Laver, J. D., Siddiqui, N. U., Westwood, J. T., Morris, Q., Lipshitz, H. D., & Smibert, C. A. (2014). Global regulation of mRNA translation and stability in the early Drosophila embryo by the Smaug RNA-binding protein. *Genome Biology*, *15*(1), 1–21. https://doi.org/10.1186/GB-2014-15-1-R4/FIGURES/10

Choe, C. P., & Brown, S. J. (2007). Evolutionary flexibility of pair-rule patterning revealed by functional analysis of secondary pair-rule genes, paired and sloppy-paired in the short-germ insect, Tribolium castaneum. *Developmental Biology*, *302*(1), 281–294. https://doi.org/10.1016/J.YDBIO.2006.09.037,

Choe, C. P., & Brown, S. J. (2009). Genetic regulation of engrailed and wingless in Tribolium segmentation and the evolution of pair-rule segmentation. *Developmental Biology*, *325*(2), 482–491. https://doi.org/10.1016/J.YDBIO.2008.10.037

Choe, C. P., Miller, S. C., & Brown, S. J. (2006). A pair-rule gene circuit defines segments sequentially in the short-term insect Tribolium castaneum. *Proceedings of the National*

Academy of Sciences of the United States of America, *103*(17), 6560–6564. https://doi.org/10.1073/PNAS.0510440103/SUPPL_FILE/10440FIG7.JPG

Choi, H. M. T., Schwarzkopf, M., Fornace, M. E., Acharya, A., Artavanis, G., Stegmaier, J., Cunha, A., & Pierce, N. A. (2018). Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. *Development*, *145*(12). https://doi.org/10.1242/dev.165753

Clark, E. (2017). Dynamic patterning by the Drosophila pair-rule network reconciles long-germ and short-germ segmentation. In *PLoS Biology* (Vol. 15, Issue 9). https://doi.org/10.1371/journal.pbio.2002439

Clark, E., & Akam, M. (2016). Odd-paired controls frequency doubling in Drosophila segmentation by altering the pair-rule gene regulatory network. *ELife*, *5*(AUGUST), 1–42. https://doi.org/10.7554/eLife.18215

Clark, E., Peel, A. D., & Akam, M. (2019). Arthropod segmentation. *Development (Cambridge)*, *146*(18). https://doi.org/10.1242/dev.170480

Colonnetta, M. M., Abrahante, J. E., Schedl, P., Gohl, D. M., & Deshpande, G. (2021). CLAMP regulates zygotic genome activation in Drosophila embryos. *Genetics*, *219*(2). https://doi.org/10.1093/genetics/iyab107

Crocker, J., & Stern, D. L. (2017). Functional regulatory evolution outside of the minimal even-skipped stripe 2 enhancer. *Development*, *144*(17), 3095–3101. https://doi.org/10.1242/dev.149427

Crombach, A., García-Solache, M. A., & Jaeger, J. (2014). Evolution of early development in dipterans: Reverse-engineering the gap gene network in the moth midge Clogmia

albipunctata (Psychodidae). *BioSystems*, *123*, 74–85. https://doi.org/10.1016/J.BIOSYSTEMS.2014.06.003

Crombach, A., & Jaeger, J. (2021). Life's attractors continued: progress in understanding developmental systems through reverse engineering and in silico evolution. In A. Crombach (Ed.), *Evolutionary Systems Biology*. Springer Nature Switzerland AG. https://doi.org/10.1007/978-3-030-71737-7_4

Crowley, L. M., Wawman, D. C., University of, O., Wytham Woods Genome Acquisition, L., Darwin Tree of Life Barcoding, collective, Wellcome Sanger Institute Tree of Life Management, S., Laboratory, team, Wellcome Sanger Institute Scientific Operations: Sequencing, O., Wellcome Sanger Institute Tree of Life Core Informatics, team, Tree of Life Core Informatics, collective, & Darwin Tree of Life, C. (2024). The genome sequence of the spotted cranefly, Nephrotoma appendiculata (Pierre, 1919). *Wellcome Open Res*, *9*, 38. https://doi.org/10.12688/wellcomeopenres.20886.1

Cusanovich, D. A., Reddington, J. P., Garfield, D. A., Daza, R. M., Aghamirzaie, D., Marco-Ferreres, R., Pliner, H. A., Christiansen, L., Qiu, X., Steemers, F. J., Trapnell, C., Shendure, J., & Furlong, E. E. M. (2018). The cis-regulatory dynamics of embryonic development at single-cell resolution. *Nature*, *555*(7697), 538–542. https://doi.org/10.1038/nature25981

Dalton, D., Chadwick, R., & McGinnis, W. (1989). Expression and embryonic function of empty spiracles: a Drosophila homeo box gene with two patterning functions on the anterior-posterior axis of the embryo. *Genes & Development*, *3*(12A), 1940–1956. https://doi.org/10.1101/GAD.3.12A.1940

Datta, R. R., Ling, J., Kurland, J., Ren, X., Xu, Z., Yucel, G., Moore, J., Shokri, L., Baker, I., Bishop, T., Struffi, P., Levina, R., Bulyk, M. L., Johnston, R. J., & Small, S. (2018). A feed-

forward relay integrates the regulatory activities of bicoid and orthodenticle via sequential binding to suboptimal sites. *Genes and Development*, *32*(9–10), 723–736. https://doi.org/10.1101/gad.311985.118

De Renzis, S., Elemento, O., Tavazoie, S., & Wieschaus, E. F. (2007). Unmasking activation of the zygotic genome using chromosomal deletions in the Drosophila embryo. *PLoS Biology*, *5*(5), 1036–1051. https://doi.org/10.1371/journal.pbio.0050117

Degen, E. A., Croslyn, C., Mangan, N. M., & Blythe, S. A. (2024). *Bicoid-nucleosome competition sets a concentration threshold for transcription constrained by genome replication*. https://doi.org/10.1101/2024.12.10.627802

Di Talia, S., She, R., Blythe, S. A., Lu, X., Zhang, Q. F., & Wieschaus, E. F. (2013). Posttranslational control of Cdc25 degradation terminates Drosophila's early cell-cycle program. *Current Biology : CB*, *23*(2), 127–132. https://doi.org/10.1016/J.CUB.2012.11.029

Diesh, C., Stevens, G. J., Xie, P., De Jesus Martinez, T., Hershberg, E. A., Leung, A., Guo, E., Dider, S., Zhang, J., Bridge, C., Hogue, G., Duncan, A., Morgan, M., Flores, T., Bimber, B. N., Haw, R., Cain, S., Buels, R. M., Stein, L. D., & Holmes, I. H. (2023). JBrowse 2: a modular genome browser with views of synteny and structural variation. *Genome Biol*, *24*(1), 74. https://doi.org/10.1186/s13059-023-02914-z

DiNardo, S., Kuner, J. M., Theis, J., & O'Farrell, P. H. (1985). Development of embryonic pattern in D. melanogaster as revealed by accumulation of the nuclear engrailed protein. *Cell*, *43*(1), 59–69. https://doi.org/10.1016/0092-8674(85)90012-1

Driever, W., & Nüsslein-Volhard, C. (1988a). A gradient of bicoid protein in Drosophila embryos. *Cell*, *54*(1), 83–93. https://doi.org/10.1016/0092-8674(88)90182-1

Driever, W., & Nüsslein-Volhard, C. (1988b). The bicoid protein determines position in the Drosophila embryo in a concentration-dependent manner. *Cell*, *54*(1), 95–104. https://doi.org/10.1016/0092-8674(88)90183-3

Driever, W., & Nüsslein-Volhard, C. (1989). The bicoid protein is a positive regulator of hunchback transcription in the early Drosophila embryo. *Nature*, *337*(6203), 138–143. https://doi.org/10.1038/337138a0

Driever, W., Siegel, V., & Nüsseein-Volhard, C. (1990). Autonomous determination of anterior structures in the early Drosophila embryo by the bicoid morphogen. *Development (Cambridge, England)*, *109*(4), 811–820. https://doi.org/10.1242/DEV.109.4.811

Driever, W., Thoma, G., & Nüsslein-Volhard, C. (1989). Determination of spatial domains of zygotic gene expression in the Drosophila embryo by the affinity of binding sites for the bicoid morphogen. *Nature*, *340*(6232), 363–367. https://doi.org/10.1038/340363A0

Duan, B., Fu, D., Zhang, C., Ding, P., Dong, X., & Xia, B. (2021). Selective Nonmethylated CpG DNA Recognition Mechanism of Cysteine Clamp Domains. *Journal of the American Chemical Society*, *143*(20), 7688–7697. https://doi.org/10.1021/JACS.1C00599

Duan, J., Rieder, L., Colonnetta, M. M., Huang, A., McKenney, M., Watters, S., Deshpande, G., Jordan, W., Fawzi, N., & Larschan, E. (2021). CLAMP and Zelda function together to promote Drosophila zygotic genome activation. *Elife*, *10*. https://doi.org/10.7554/eLife.69937

Dubnau, J., Struhl, G., Rivera-Pomar, R., Niessing, D., Schmidt-Ott, U., Gehring, W. J., Jäckle, H., Treisman, J., Gönczy, P., Vashishtha, M., Harris, E., Desplan, C., Hanes, S. D., Riddihough, G., Ish-Horowicz, D., & Brent, R. (1997). Sequence-specific RNA binding by Bicoid. *Nature 1997 388:6643*, *388*(6643), 634–634. https://doi.org/10.1038/41692

Eck, E., Liu, J., Kazemzadeh-Atoufi, M., Ghoreishi, S., Blythe, S., & Garcia, H. G. (2020). Quantitative dissection of transcription in development yields evidence for transcription factor-driven chromatin accessibility. *ELife*, *9*, 1–99. https://doi.org/10.7554/ELIFE.56429

Edgar, B. A., & Datar, S. A. (1996). Zygotic degradation of two maternal Cdc25 mRNAs terminates Drosophila's early cell cycle program. *Genes & Development*, *10*(15), 1966–1977. https://doi.org/10.1101/GAD.10.15.1966

Edgar, B. A., Sprenger, F., Duronio, R. J., Leopold, P., & O'Farrell, P. H. (1994). Distinct molecular mechanism regulate cell cycle timing at successive stages of Drosophila embryogenesis. *Genes & Development*, *8*(4), 440–452. https://doi.org/10.1101/GAD.8.4.440

El-Sherif, E., & Levine, M. (2016). Shadow Enhancers Mediate Dynamic Shifts of Gap Gene Expression in the Drosophila Embryo. *Current Biology*, *26*(9), 1164–1169. https://doi.org/10.1016/j.cub.2016.02.054

Farrell, J. A., & O'Farrell, P. H. (2013). Mechanism and regulation of Cdc25/Twine protein destruction in embryonic cell-cycle remodeling. *Current Biology : CB*, *23*(2), 118–126. https://doi.org/10.1016/J.CUB.2012.11.036

Farrell, J. A., Shermoen, A. W., Yuan, K., & O'Farrell, P. H. (2012). Embryonic onset of late replication requires Cdc25 down-regulation. *Genes & Development*, *26*(7), 714. https://doi.org/10.1101/GAD.186429.111

Fernandes, G., Tran, H., Andrieu, M., Diaw, Y., Romero, C. P., Fradin, C., Coppey, M., Walczak, A. M., & Dostatni, N. (2022). Synthetic reconstruction of the hunchback promoter specifies the role of Bicoid, Zelda and Hunchback in the dynamics of its transcription. *ELife*, *11*. https://doi.org/10.7554/ELIFE.74509

Ferrandon, D., Koch, I., Westhof, E., & Nüsslein-Volhard, C. (1997). RNA-RNA interaction is required for the formation of specific bicoid mRNA 3' UTR-STAUFEN ribonucleoprotein particles. *The EMBO Journal*, *16*(7), 1751–1758. https://doi.org/10.1093/EMBOJ/16.7.1751

Fierling, J., John, A., Delorme, B., Torzynski, A., Blanchard, G. B., Lye, C. M., Popkova, A., Malandain, G., Sanson, B., Étienne, J., Marmottant, P., Quilliet, C., & Rauzi, M. (2022). Embryo-scale epithelial buckling forms a propagating furrow that initiates gastrulation. *Nature Communications 2022 13:1*, *13*(1), 1–14. https://doi.org/10.1038/s41467-022-30493-3

Finkelstein, R., & Perrimon, N. (1990). The orthodenticle gene is regulated by bicoid and torso and specifies Drosophila head development. *Nature*, *346*(6283), 485–488. https://doi.org/10.1038/346485A0

Foe, V. E., Odell, G. M., & Bruce, E. A. (1993). Mitosis and morphogenesis in the Drosophila embryo, Point and counterpoint. In M. Bate & A. Martinez-Arias (Eds.), *The Development of Drosophila melanogaster.* (Vol. 1, pp. 149–300). Cold Spring Harbor Laboratory Press.

Foo, S. M., Sun, Y., Lim, B., Ziukaite, R., O'Brien, K., Nien, C. Y., Kirov, N., Shvartsman, S. Y., & Rushlow, C. A. (2014). Zelda potentiates morphogen activity by increasing chromatin accessibility. *Current Biology : CB*, *24*(12), 1341–1346. https://doi.org/10.1016/J.CUB.2014.04.032

Fujioka, M., Emi-Sarker, Y., Yusibova, G. L., Goto, T., & Jaynes, J. B. (1999). Analysis of an even-skipped rescue transgene reveals both composite and discrete neuronal and early blastoderm enhancers, and multi-stripe positioning by gap gene repressor gradients. *Development (Cambridge, England)*, *126*(11), 2527–2538. https://doi.org/10.1242/DEV.126.11.2527

Fujioka, M., & Jaynes, J. B. (2012). Regulation of a duplicated locus: Drosophila sloppy paired is replete with functionally overlapping enhancers. *Developmental Biology*, *362*(2), 309–319. https://doi.org/10.1016/J.YDBIO.2011.12.001

Galouzis, C. C., Kherdjemil, Y., Forneris, M., Viales, R. R., Marco-Ferreres, R., & Furlong, E. E. M. (2025). Chip (Ldb1) is a putative cofactor of Zelda forming a functional bridge to CBP during zygotic genome activation. *Molecular Cell*, *85*(12), 2425-2441.e9. https://doi.org/10.1016/J.MOLCEL.2025.05.018

Gao, Q., & Finkelstein, R. (1998). Targeting gene expression to the head: the Drosophila orthodenticle gene is a direct target of the Bicoid morphogen. *Development (Cambridge, England)*, *125*(21), 4185–4193. https://doi.org/10.1242/DEV.125.21.4185

Gao, S., Xue, S., Gao, T., Lu, R., Zhang, X., Zhang, Y., Zhang, K., & Li, R. (2023). Transcriptome analysis reveals the role of Zelda in the regulation of embryonic and wing development of Tribolium castaneum. *Bull Entomol Res*, *113*(5), 587–597. https://doi.org/10.1017/S0007485323000263

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

García-Solache, M., Jaeger, J., & Akam, M. (2010). A systematic analysis of the gap gene system in the moth midge Clogmia albipunctata. *Developmental Biology*, *344*(1), 306–318. https://doi.org/10.1016/J.YDBIO.2010.04.019

Gaskill, M. M., Gibson, T. J., Larson, E. D., & Harrison, M. M. (2021). GAF is essential for zygotic genome activation and chromatin accessibility in the early Drosophila embryo. *ELife*, *10*, 1–30. https://doi.org/10.7554/eLife.66668

Gawliński, P., Nikolay, R., Goursot, C., Lawo, S., Chaurasia, B., Herz, H. M., Kußler-Schneider, Y., Ruppert, T., Mayer, M., & Großhans, J. (2007). The Drosophila mitotic inhibitor Frühstart specifically binds to the hydrophobic patch of cyclins. *EMBO Reports*, *8*(5), 490–496. https://doi.org/10.1038/SJ.EMBOR.7400948

Generalovic, T. N., McCarthy, S. A., Warren, I. A., Wood, J. M. D., Torrance, J., Sims, Y., Quail, M., Howe, K., Pipan, M., Durbin, R., & Jiggins, C. D. (2021). A high-quality, chromosome-level genome assembly of the Black Soldier Fly (Hermetia illucens L.). *G3 (Bethesda)*, *11*(5). https://doi.org/10.1093/g3journal/jkab085

Gheisari, E., Aakhte, M., & Müller, H. A. J. (2020). Gastrulation in Drosophila melanogaster: Genetic control, cellular basis and biomechanics. *Mechanisms of Development*, *163*, 103629. https://doi.org/10.1016/J.MOD.2020.103629

Giannios, P., & Tsitilou, S. G. (2013). The embryonic transcription factor Zelda of Drosophila melanogaster is also expressed in larvae and may regulate developmentally important genes. *Biochemical and Biophysical Research Communications*, *438*(2), 329–333. https://doi.org/10.1016/J.BBRC.2013.07.071

Gibson, T. J., Larson, E. D., & Harrison, M. M. (2024). Protein-intrinsic properties and context-dependent effects regulate pioneer factor binding and function. *Nat Struct Mol Biol*, *31*(3), 548–558. https://doi.org/10.1038/s41594-024-01231-8

Goltsev, Y., Hsiong, W., Lanzaro, G., & Levine, M. (2004a). Different combinations of gap repressors for common stripes in Anopheles and Drosophila embryos. *Dev. Biol.*, *275*, 435–446.

Goltsev, Y., Hsiong, W., Lanzaro, G., & Levine, M. (2004b). Different combinations of gap repressors for common stripes in Anopheles and Drosophila embryos. *Developmental Biology*, *275*(2), 435–446. https://doi.org/10.1016/J.YDBIO.2004.08.021

Gonzaga-Saavedra, N., Degen, E. A., Soluri, I. V., Croslyn, C., & Blythe, S. A. (2025). Nucleation-dependent propagation of Polycomb modifications emerges during the Drosophila maternal to zygotic transition. *BioRxiv*, 2025.07.02.662854. https://doi.org/10.1101/2025.07.02.662854

González, F., Swales, L., Bejsovec, A., Skaer, H., & Arias, A. M. (1991). Secretion and movement of wingless protein in the epidermis of the Drosophila embryo. *Mechanisms of Development*, *35*(1), 43–54. https://doi.org/10.1016/0925-4773(91)90040-D

Goto, T., Macdonald, P., & Maniatis, T. (1989). Early and late periodic patterns of even skipped expression are controlled by distinct regulatory elements that respond to different spatial cues. *Cell*, *57*(3), 413–422. https://doi.org/10.1016/0092-8674(89)90916-1

Grand, R. S., Pregnolato, M., Baumgartner, L., Hoerner, L., Burger, L., & Schübeler, D. (2024). Genome access is transcription factor-specific and defined by nucleosome position. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2024.08.009

Gregor, T., McGregor, A. P., & Wieschaus, E. F. (2008). Shape and function of the Bicoid morphogen gradient in dipteran species with different sized embryos. *Developmental Biology*, *316*(2), 350–358. https://doi.org/10.1016/J.YDBIO.2008.01.039

Großhans, J., Müller, H. A. J., & Wieschaus, E. (2003). Control of cleavage cycles in Drosophila embryos by frühstart. *Developmental Cell*, *5*(2), 285–294. https://doi.org/10.1016/S1534-5807(03)00208-9

Grossniklaus, U., Cadigan, K. M., & Gehring, W. J. (1994). Three maternal coordinate systems cooperate in the patterning of the Drosophila head. *Development*, *120*(11), 3155–3171. https://doi.org/10.1242/DEV.120.11.3155

Grossniklaus, U., Pearson, R. K., & Gehring, W. J. (1992). The Drosophila sloppy paired locus encodes two proteins involved in segmentation that show homology to mammalian transcription factors. *Genes Dev*, *6*(6), 1030–1051. https://doi.org/10.1101/gad.6.6.1030

Gu, Z., Gu, L., Eils, R., Schlesner, M., & Brors, B. (2014). circlize Implements and enhances circular visualization in R. *Bioinformatics*, *30*(19), 2811–2812. https://doi.org/10.1093/bioinformatics/btu393

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics*, *29*(8), 1072–1075. https://doi.org/10.1093/BIOINFORMATICS/BTT086,

Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., White, O., Buell, C. R., & Wortman, J. R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol*, *9*(1), R7. https://doi.org/10.1186/gb-2008-9-1-r7

Haines, J. E., & Eisen, M. B. (2018). Patterns of chromatin accessibility along the anterior-posterior axis in the early Drosophila embryo. *PLoS Genetics*, *14*(5). https://doi.org/10.1371/journal.pgen.1007367

Hall, A. B., Basu, S., Jiang, X., Qi, Y., Timoshevskiy, V. A., Biedler, J. K., Sharakhova, M. V, Elahi, R., Anderson, M. A., Chen, X. G., Sharakhov, I. V, Adelman, Z. N., & Tu, Z. (2015). SEX DETERMINATION. A male-determining factor in the mosquito Aedes aegypti. *Science*, *348*(6240), 1268–1270. https://doi.org/10.1126/science.aaa2850

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. *Journal of Biological Chemistry*, *290*(6), 3508–3518. https://doi.org/10.1074/jbc.M114.602292

Hamm, D. C., & Harrison, M. M. (2018). Regulatory principles governing the maternal-to-zygotic transition: insights from Drosophila melanogaster. *Open Biol*, *8*(12), 180183. https://doi.org/10.1098/rsob.180183

Hamm, D. C., Larson, E. D., Nevil, M., Marshall, K. E., Bondra, E. R., & Harrison, M. M. (2017). A conserved maternal-specific repressive domain in Zelda revealed by Cas9-mediated mutagenesis in Drosophila melanogaster. *PLOS Genetics*, *13*(12), e1007120. https://doi.org/10.1371/JOURNAL.PGEN.1007120

Hammonds, A. S., Bristow, C. A., Fisher, W. W., Weiszmann, R., Wu, S., Hartenstein, V., Kellis, M., Yu, B., Frise, E., & Celniker, S. E. (2013). Spatial expression of transcription factors in Drosophila embryonic organ development. *Genome Biology*, *14*(12). https://doi.org/10.1186/GB-2013-14-12-R140

Hancock, D. Y., Fischer, J., Lowe, J. M., Fischer, J., Lowe, J. M., Snapp-Childs, W., Pierce, M., Marru, S., Coulter, J. E., Vaughn, M., Beck, B., Merchant, N., Skidmore, E., & Jacobs, G. (2021). Jetstream2: Accelerating cloud computing via Jetstream. In *Practice and Experience in Advanced Research Computing (PEARC '21)*. Association for Computing Machinery. https://doi.org/10.1145/3437359.3465565

Hannon, C. E., Blythe, S. A., & Wieschaus, E. F. (2017). Concentration dependent chromatin states induced by the bicoid morphogen gradient. *ELife*, *6*, 1–29. https://doi.org/10.7554/eLife.28275

Harden, T. T., Vincent, B. J., & DePace, A. H. (2023). Transcriptional activators in the early Drosophila embryo perform different kinetic roles. *Cell Systems*, *14*(4), 258-272.e4. https://doi.org/10.1016/J.CELS.2023.03.006

Harding, K., Hoey, T., Warrior, R., & Levine, M. (1989). Autoregulatory and gap gene response elements of the even-skipped promoter of Drosophila. *The EMBO Journal*, *8*(4), 1205–1212. https://doi.org/10.1002/J.1460-2075.1989.TB03493.X

Harding, K., Rushlow, C., Doyle, H. J., Hoey, T., & Levine, M. (1986). Cross-regulatory interactions among pair-rule genes in Drosophila. *Science (New York, N.Y.)*, *233*(4767), 953–959. https://doi.org/10.1126/SCIENCE.3755551

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. 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). https://doi.org/10.1371/journal.pgen.1002266

Harrison, M. M., Marsh, A. J., & Rushlow, C. A. (2023). Setting the stage for development: the maternal-to-zygotic transition in Drosophila. *Genetics*, *225*(2). https://doi.org/10.1093/GENETICS/IYAD142

Harrison, S. D., & Travers, A. A. (1990). The tramtrack gene encodes a Drosophila finger protein that interacts with the ftz transcriptional regulatory region and shows a novel embryonic expression pattern. *The EMBO Journal*, *9*(1), 207. https://doi.org/10.1002/J.1460-2075.1990.TB08097.X

Heyn, P., Kircher, M., Dahl, A., Kelso, J., Tomancak, P., Kalinka, A. T., & Neugebauer, K. M. (2014). The earliest transcribed zygotic genes are short, newly evolved, and different across species. *Cell Reports*, *6*(2), 285–292. https://doi.org/10.1016/J.CELREP.2013.12.030

Hildebrandt, K., Kolb, D., Kloppel, C., Kaspar, P., Wittling, F., Hartwig, O., Federspiel, J., Findji, I., & Walldorf, U. (2022). Regulatory modules mediating the complex neural expression patterns of the homeobrain gene during Drosophila brain development. *Hereditas*, *159*(1), 2. https://doi.org/10.1186/s41065-021-00218-5

Hoch, M., Seifert, E., & Jäckle, H. (1991a). Gene expression mediated by cis-acting sequences of the Krüppel gene in response to the Drosophila morphogens bicoid and hunchback. *The EMBO Journal*, *10*(8), 2267–2278. https://doi.org/10.1002/J.1460-2075.1991.TB07763.X

Hoch, M., Seifert, E., & Jäckle, H. (1991b). Gene expression mediated by cis-acting sequences of the Krüppel gene in response to the Drosophila morphogens bicoid and hunchback. *Embo j*, *10*(8), 2267–2278.

Houtmeyers, R., Souopgui, J., Tejpar, S., & Arkell, R. (2013). The ZIC gene family encodes multi-functional proteins essential for patterning and morphogenesis. *Cellular and Molecular Life Sciences : CMLS*, *70*(20), 3791–3811. https://doi.org/10.1007/S00018-013-1285-5

Howard, K., Ingham, P., & Rushlow, C. (1988). Region-specific alleles of the Drosophila segmentation gene hairy. *Genes & Development*, *2*(8), 1037–1046. https://doi.org/10.1101/GAD.2.8.1037

Hubert, K. A., & Wellik, D. M. (2023). Hox genes in development and beyond. *Development (Cambridge)*, *150*(1). https://doi.org/10.1242/DEV.192476/286593

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernandez-Plaza, A., Forslund, S. K., Cook, H., Mende, D. R., Letunic, I., Rattei, T., Jensen, L. J., von Mering, C., & Bork, P. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res*, *47*(D1), D309–D314. https://doi.org/10.1093/nar/gky1085

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

Hülskamp, M., Pfeifle, C., & Tautz, D. (1990). A morphogenetic gradient of hunchback protein organizes the expression of the gap genes Krüppel and knirps in the early Drosophila embryo. *Nature 1990 346:6284*, *346*(6284), 577–580. https://doi.org/10.1038/346577A0

Hülskamp, M., Schröder, C., Pfeifle, C., Jäckle, H., & Tautz, D. (1989). Posterior segmentation of the Drosophila embryo in the absence of a maternal posterior organizer gene. *Nature 1989 338:6217*, *338*(6217), 629–632. https://doi.org/10.1038/338629a0

Hursh, D. A., & Stultz, B. G. (2018). Odd-Paired: The Drosophila Zic Gene. *Advances in Experimental Medicine and Biology*, *1046*, 41–58. https://doi.org/10.1007/978-981-10-7311-3_3

Ingham, P., & Gergen, P. (1988). Interactions between the pair-rule genes runt, hairy, even-skipped and fushi tarazu and the establishment of periodic pattern in the Drosophila embryo. *Development*, *104*(Supplement), 51–60. https://doi.org/10.1242/DEV.104.SUPPLEMENT.51

Ingham, P., Martinez-Arias, A., Lawrence, P. A., & Howard, K. (1985). Expression of engrailed in the parasegment of Drosophila. *Nature 1985 317:6038*, *317*(6038), 634–636. https://doi.org/10.1038/317634a0

Irish, V., Lehmann, R., & Akam, M. (1989). The Drosophila posterior-group gene nanos functions by repressing hunchback activity. *Nature*, *338*(6217), 646–648. https://doi.org/10.1038/338646a0

Iwafuchi-Doi, M., & Zaret, K. S. (2014). Pioneer transcription factors in cell reprogramming. *Genes and Development*, *28*(24), 2679–2692. https://doi.org/10.1101/gad.253443.114

Jäckle, H., Tautz, D., Schuh, R., Seifert, E., & Lehmann, R. (1986). Cross-regulatory interactions among the gap genes of Drosophila. *Nature 1986 324:6098*, *324*(6098), 668–670. https://doi.org/10.1038/324668a0

Jacobs, J., Atkins, M., Davie, K., Imrichova, H., Romanelli, L., Christiaens, V., Hulselmans, G., Potier, D., Wouters, J., Taskiran, I. I., Paciello, G., González-Blas, C. B., Koldere, D., Aibar, S., Halder, G., & Aerts, S. (2018). The transcription factor Grainy head primes epithelial enhancers for spatiotemporal activation by displacing nucleosomes. *Nature Genetics 2018 50:7*, *50*(7), 1011–1020. https://doi.org/10.1038/s41588-018-0140-x

Jaeger, J. (2011). The gap gene network. *Cellular and Molecular Life Sciences*, *68*(2), 243–274. https://doi.org/10.1007/s00018-010-0536-y

Janssen, R. (2020). The embryonic expression pattern of a second, hitherto unrecognized, paralog of the pair-rule gene sloppy-paired in the beetle Tribolium castaneum. *Development Genes and Evolution*, *230*(3), 247–256. https://doi.org/10.1007/S00427-020-00660-X,

Janssens, H., Siggens, K., Cicin-Sain, D., Jiménez-Guri, E., Musy, M., Akam, M., & Jaeger, J. (2014). A quantitative atlas of Even-skipped and Hunchback expression in Clogmia

albipunctata (Diptera: Psychodidae) blastoderm embryos. *EvoDevo*, *5*(1), 1–13. https://doi.org/10.1186/2041-9139-5-1/FIGURES/7

Jevtić, P., & Levy, D. L. (2015). Nuclear Size Scaling during Xenopus Early Development Contributes to Midblastula Transition Timing. *Current Biology*, *25*(1), 45–52. https://doi.org/10.1016/J.CUB.2014.10.051

Ji, J., Gao, Y., Xu, C., Zhang, K., Li, D., Li, B., Chen, L., Gao, M., Huangfu, N., Elumalai, P., Gao, X., Zhu, X., Wang, L., Luo, J., & Cui, J. (2024). Chromosome-level genome assembly of marmalade hoverfly Episyrphus balteatus (Diptera: Syrphidae). *Sci Data*, *11*(1), 844. https://doi.org/10.1038/s41597-024-03666-6

Jiménez, G., Guichet, A., Ephrussi, A., & Casanova, J. (2000). Relief of gene repression by Torso RTK signaling: role of capicua in Drosophila terminal and dorsoventral patterning. *Genes & Development*, *14*(2), 224–231. https://doi.org/10.1101/GAD.14.2.224

Jiménez-Guri, E., Wotton, K. R., Gavilán, B., & Jaeger, J. (2014). A staging scheme for the development of the moth midge Clogmia albipunctata. *PLoS ONE*, *9*(1). https://doi.org/10.1371/journal.pone.0084422

Jiménez-Guri, E., Wotton, K. R., & Jaeger, J. (2018). tarsal-less is expressed as a gap gene but has no gap gene phenotype in the moth midge Clogmia albipunctata. *Royal Society Open Science*, *5*(8). https://doi.org/10.1098/rsos.180458

Johnson, J. L., Georgakilas, G., Petrovic, J., Kurachi, M., Cai, S., Harly, C., Pear, W. S., Bhandoola, A., Wherry, E. J., & Vahedi, G. (2018). Lineage-Determining Transcription Factor TCF-1 Initiates the Epigenetic Identity of T Cells. *Immunity*, *48*(2), 243-257.e10. https://doi.org/10.1016/J.IMMUNI.2018.01.012/ATTACHMENT/CE01F91C-03D6-4C3B-90CF-0C17464EB4AA/MMC8.PDF

Joseph, S. R., Pálfy, M., Hilbert, L., Kumar, M., Karschau, J., Zaburdaev, V., Shevchenko, A., & Vastenhouw, N. L. (2017). Competition between histone and transcription factor binding regulates the onset of transcription in zebrafish embryos. *ELife*, *6*. https://doi.org/10.7554/ELIFE.23326

Judd, J., Duarte, F. M., & Lis, J. T. (2021). *Pioneer-like factor GAF cooperates with PBAP (SWI/SNF) and NURF (ISWI) to regulate transcription*. https://doi.org/10.1101/gad.341768.120

Jürgens, G., Wieschaus, E., Nüsslein-Volhard, C., & Klu. (1984). Mutations affecting the pattern of the larval cuticle in Drosophila melanogaster II. Zygotic loci on the third chromosome. *Roux's Archives of Developmental Biology*, *193*(3), 283–295.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*, *14*(6), 587–589. https://doi.org/10.1038/nmeth.4285

Katoh, K., Rozewicki, J., & Yamada, K. D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform*, *20*(4), 1160–1166. https://doi.org/10.1093/bib/bbx108

Kaye, E. G., Booker, M., Kurland, J. V, Conicella, A. E., Fawzi, N. L., Bulyk, M. L., Tolstorukov, M. Y., & Larschan, E. (2018). Differential Occupancy of Two GA-Binding Proteins Promotes Targeting of the Drosophila Dosage Compensation Complex to the Male X Chromosome. *Cell Rep*, *22*(12), 3227–3239. https://doi.org/10.1016/j.celrep.2018.02.098

Keyes, L. N., Cline, T. W., & Schedl, P. (1992). The primary sex determination signal of Drosophila acts at the level of transcription. *Cell*, *68*(5), 933–943. https://doi.org/10.1016/0092-8674(92)90036-C

Klinger, M., & Gergen, J. P. (1993). Regulation of runt transcription by Drosophila segmentation genes. *Mechanisms of Development*, *43*(1), 3–19. https://doi.org/10.1016/0925-4773(93)90019-T

Klingler, M., Erdélyi, M., Szabad, J., & Nüsslein-Volhard, C. (1988). Function of torso in determining the terminal anlagen of the Drosophila embryo. *Nature*, *335*(6187), 275–277. https://doi.org/10.1038/335275A0

Klomp, J., Athy, D., Kwan, C. W., Bloch, N. I., Sandmann, T., Lemke, S., & Schmidt-Ott, U. (2015). A cysteine-clamp gene drives embryo polarity in the midge Chironomus. *Science*, *348*(6238), 1040–1042. https://doi.org/10.1126/science.aaa7105

Kolb, D., Kaspar, P., Klöppel, C., & Walldorf, U. (2021). The Drosophila homeodomain transcription factor Homeobrain is involved in the formation of the embryonic protocerebrum and the supraesophageal brain commissure. *Cells & Development*, *165*. https://doi.org/10.1016/J.CDEV.2021.203657

Konstantinides, N., Holguera, I., Rossi, A. M., Escobar, A., Dudragne, L., Chen, Y. C., Tran, T. N., Martínez Jaimes, A. M., Özel, M. N., Simon, F., Shao, Z., Tsankova, N. M., Fullard, J. F., Walldorf, U., Roussos, P., & Desplan, C. (2022). A complete temporal transcription factor series in the fly visual system. *Nature 2022 604:7905*, *604*(7905), 316–322. https://doi.org/10.1038/s41586-022-04564-w

Kornberg, T., Sidén, I., O'Farrell, P., & Simon, M. (1985). The engrailed locus of Drosophila: in situ localization of transcripts reveals compartment-specific expression. *Cell*, *40*(1), 45–53. https://doi.org/10.1016/0092-8674(85)90307-1

Koromila, T., Gao, F., Iwasaki, Y., He, P., Pachter, L., Peter Gergen, J., & Stathopoulos, A. (2020). Odd-paired is a pioneer-like factor that coordinates with zelda to control gene expression in embryos. *ELife*, *9*, 1–71. https://doi.org/10.7554/eLife.59610

Krueger, F. (2012). *Trim Galore!* Babraham Bioinformatics.

Krzywinska, E., Dennison, N. J., Lycett, G. J., & Krzywinski, J. (2016). A maleness gene in the malaria mosquito Anopheles gambiae. *Science*, *353*(6294), 67–69. https://doi.org/10.1126/science.aaf5605

Kulakovskii, I. V, & Makeev, Vi. (2009). [Integration of data obtained by different experimental methods to determine the motifs in DNA sequences recognized by transcription-regulating factors]. *Biofizika*, *54*(6), 965–974. https://www.ncbi.nlm.nih.gov/pubmed/20067172

Kulakovskiy, I. V, Favorov, A. V, & Makeev, V. J. (2009). Motif discovery and motif finding from genome-mapped DNase footprint data. *Bioinformatics*, *25*(18), 2318–2325. https://doi.org/10.1093/bioinformatics/btp434

Kuzu, G., Kaye, E. G., Chery, J., Siggers, T., Yang, L., Dobson, J. R., Boor, S., Bliss, J., Liu, W., Jogl, G., Rohs, R., Singh, N. D., Bulyk, M. L., Tolstorukov, M. Y., & Larschan, E. (2016). Expansion of GA Dinucleotide Repeats Increases the Density of CLAMP Binding Sites on the X-Chromosome to Promote Drosophila Dosage Compensation. *PLoS Genet*, *12*(7), e1006120. https://doi.org/10.1371/journal.pgen.1006120

Kwan, C. W., Gavin-Smyth, J., Ferguson, E. L., & Schmidt-Ott, U. (2016). Functional evolution of a morphogenetic gradient. *Elife*, *5*. https://doi.org/10.7554/eLife.20894

Kwasnieski, J. C., Orr-Weaver, T. L., & Bartel, D. P. (2019). Early genome activation in Drosophila is extensive with an initial tendency for aborted transcripts and retained introns. *Genome Research*, *29*(7), 1188–1197. https://doi.org/10.1101/GR.242164.118

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., Fitzhugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., Levine, R., McEwan, P., … Morgan, M. J. (2001). Initial sequencing and analysis of the human genome. *Nature 2001 409:6822*, *409*(6822), 860–921. https://doi.org/10.1038/35057062

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Larson, E. D., Marsh, A. J., & Harrison, M. M. (2021). Pioneering the developmental frontier. *Molecular Cell*, *81*(8), 1640–1650. https://doi.org/10.1016/J.MOLCEL.2021.02.020

Lawrence, M., Huber, W., Pages, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M. T., & Carey, V. J. (2013). Software for computing and annotating genomic ranges. *PLoS Comput Biol*, *9*(8), e1003118. https://doi.org/10.1371/journal.pcbi.1003118

Lawrence, P. A. (1992). *The Making of a Fly: the Genetics of Animal Design*. Blackwell Scientific Publications.

Lee, H. H., & Frasch, M. (2004). Survey of forkhead domain encoding genes in the Drosophila genome: Classification and embryonic expression patterns. *Dev Dyn*, *229*(2), 357–366. https://doi.org/10.1002/dvdy.10443

Lehmann, R., & Nüsslein-Volhard, C. (1986). Abdominal segmentation, pole cell formation, and embryonic polarity require the localized activity of oskar, a maternal gene in Drosophila. *Cell*, *47*(1), 141–152. https://doi.org/10.1016/0092-8674(86)90375-2

Lehmann, R., & Nüsslein-Volhard, C. (1991). The maternal gene nanos has a central role in posterior pattern formation of the Drosophila embryo. *Development (Cambridge, England)*, *112*(3), 679–691. https://doi.org/10.1242/DEV.112.3.679

Lemke, S., Busch, S. E., Antonopoulos, D. A., Meyer, F., Domanus, M. H., & Schmidt-Ott, U. (2010). Maternal activation of gap genes in the hover fly Episyrphus. *Development*, *137*(10), 1709–1719. https://doi.org/10.1242/DEV.046649

Lemke, S., Kale, G., & Urbansky, S. (2020). Comparing gastrulation in flies: Links between cell biology and the evolution of embryonic morphogenesis. *Mechanisms of Development*, *164*. https://doi.org/10.1016/j.mod.2020.103648

Lemke, S., & Schmidt-Ott, U. (2009). Evidence for a composite anterior determinant in the hover fly Episyrphus balteatus (Syrphidae), a cyclorrhaphan fly with an anterodorsal serosa anlage. *Development (Cambridge, England)*, *136*(1), 117–127. https://doi.org/10.1242/DEV.030270

Lewis, E. B. (1978). A gene complex controlling segmentation in Drosophila. *Nature 1978 276:5688*, *276*(5688), 565–570. https://doi.org/10.1038/276565a0

Li, H., & Johnson, A. D. (2010). Evolution of Transcription Networks — Lessons from Yeasts. *Current Biology : CB*, *20*(17), R746. https://doi.org/10.1016/J.CUB.2010.06.056

Li, M., Kasan, K., Saha, Z., Yoon, Y., & Schmidt-Ott, U. (2023). Twenty-seven ZAD-ZNF genes of Drosophila melanogaster are orthologous to the embryo polarity determining mosquito gene cucoid. *PloS One*, *18*(1). https://doi.org/10.1371/JOURNAL.PONE.0274716

Li, X. Y., Harrison, M. M., Villalta, J. E., Kaplan, T., & Eisen, M. B. (2014a). 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. (2014b). 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. https://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. https://doi.org/10.1016/0925-4773(93)90087-E

Ling, J., Umezawa, K. Y., Scott, T., & Small, S. (2019). Bicoid-Dependent Activation of the Target Gene hunchback Requires a Two-Motif Sequence Code in a Specific Basal Promoter. *Mol Cell*, *75*(6), 1178-1187 e4. https://doi.org/10.1016/j.molcel.2019.06.038

Liu, Q., Onal, P., Datta, R. R., Rogers, J. M., Schmidt-Ott, U., Bulyk, M. L., Small, S., & Thornton, J. W. (2018). Ancient mechanisms for the evolution of the bicoid homeodomain's function in fly development. *ELife*, *7*, 1–28. https://doi.org/10.7554/eLife.34594

Lott, S. E., Villalta, J. E., Schroth, G. P., Luo, S., Tonkin, L. A., & Eisen, M. B. (2011). Noncanonical Compensation of 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

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Lukowitz, W., Schröder, C., Glaser, G., Hülskamp, M., & Tautz, D. (1994). Regulatory and coding regions of the segmentation gene hunchback are functionally conserved between Drosophila virilis and Drosophila melanogaster. *Mechanisms of Development*, *45*(2), 105–115. https://doi.org/10.1016/0925-4773(94)90024-8

Luo, Z., Gao, X., Lin, C., Smith, E. R., Marshall, S. A., Swanson, S. K., Florens, L., Washburn, M. P., & Shilatifard, A. (2015). Zic2 is an enhancer-binding factor required for embryonic stem cell specification. *Molecular Cell*, *57*(4), 685–694. https://doi.org/10.1016/J.MOLCEL.2015.01.007/ATTACHMENT/E024ADBC-D1D3-44E4-B403-24A80EA0772F/MMC3.PDF

Ma, J., He, F., Xie, G., & Deng, G. X. (2016). Maternal AP determination in the Drosophila oocyte and embryo. *Wiley Interdiscip Rev Dev Biol*, *5*, 562–581.

Macdonald, P. M. (2005). Translational repression by Bicoid: competition for the cap. *Cell*, *121*(3), 321–322. https://doi.org/10.1016/j.cell.2005.04.018

Macdonald, P. M., & Kerr, K. (1998). Mutational analysis of an RNA recognition element that mediates localization of bicoid mRNA. *Molecular and Cellular Biology*, *18*(7), 3788–3795. https://doi.org/10.1128/MCB.18.7.3788

Manni, M., Berkeley, M. R., Seppey, M., & Zdobnov, E. M. (2021). BUSCO: Assessing Genomic Data Quality and Beyond. *Curr Protoc*, *1*(12), e323. https://doi.org/10.1002/cpz1.323

Manu, Surkova, S., Spirov, A. V, Gursky, V. V, Janssens, H., Kim, A. R., Radulescu, O., Vanario-Alonso, C. E., Sharp, D. H., Samsonova, M., & Reinitz, J. (2009). Canalization of gene expression and domain shifts in the Drosophila blastoderm by dynamical attractors. *PLoS Comput Biol*, *5*(3), e1000303. https://doi.org/10.1371/journal.pcbi.1000303

Martinez-Corral, R., Friedrich, D., Frömel, R., Velten, L., Gunawardena, J., & DePace, A. H. (2024). Emergence of activation or repression in transcriptional control under a fixed molecular context. *BioRxiv*, 2024.05.29.596388. https://doi.org/10.1101/2024.05.29.596388

McDaniel, S. L., Gibson, T. J., Schulz, K. N., Fernandez Garcia, M., Nevil, M., Jain, S. U., Lewis, P. W., Zaret, K. S., & Harrison, M. M. (2019). Continued Activity of the Pioneer Factor Zelda Is Required to Drive Zygotic Genome Activation. *Molecular Cell*, *74*(1), 185-195.e4. https://doi.org/10.1016/j.molcel.2019.01.014

McGregor, A. P., Shaw, P. J., & Dover, G. A. (2001). Sequence and expression of the hunchback gene in Lucilia sericata: a comparison with other Dipterans. *Development Genes and Evolution*, *211*(6), 315–318. https://doi.org/10.1007/S004270100148

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol*, *37*(5), 1530–1534. https://doi.org/10.1093/molbev/msaa015

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

Mir, M., Stadler, M. R., Ortiz, S. A., Hannon, C. E., Harrison, M. M., Darzacq, X., & Eisen, M. B. (2018). Dynamic multifactor hubs interact transiently with sites of active transcription in drosophila embryos. *ELife*, *7*(Dv), 1–27. https://doi.org/10.7554/eLife.40497

Morgan, M., Pages, H., Obenchain, V., & Hayden, N. (2013). *Rsamtools: Binary alignment (BAM), FASTA, variant call (BCF), and tabix file import.* Bioconductor.

Moshe, A., & Kaplan, T. (2017). Genome-wide search for Zelda-like chromatin signatures identifies GAF as a pioneer factor in early fly development. *Epigenetics and Chromatin*, *10*(1), 1–14. https://doi.org/10.1186/S13072-017-0141-5/FIGURES/7

Murawsky, C. M., Brehm, A., Badenhorst, P., Lowe, N., Becker, P. B., & Travers, A. A. (2001). Tramtrack69 interacts with the dMi-2 subunit of the Drosophila NuRD chromatin remodelling complex. *EMBO Reports*, *2*(12), 1089. https://doi.org/10.1093/EMBO-REPORTS/KVE252

Nasiadka, A., Dietrich, B. H., & Krause, H. M. (2002). Anterior-posterior patterning in the Drosophila embryo. *Advances in Developmental Biology and Biochemistry*, *12*(C), 155–204. https://doi.org/10.1016/S1569-1799(02)12027-2

Nevil, M., Bondra, E. R., Schulz, K. N., Kaplan, T., & Harrison, M. M. (2017). Stable binding of the conserved transcription factor grainy head to its target genes throughout Drosophila melanogaster development. *Genetics*, *205*(2), 605–620. https://doi.org/10.1534/GENETICS.116.195685/-/DC1

Nevil, M., Gibson, T. J., Bartolutti, C., Iyengar, A., & Harrison, M. M. (2020). Establishment of chromatin accessibility by the conserved transcription factor Grainy head is developmentally regulated. *Development (Cambridge)*, *147*(5). https://doi.org/10.1242/DEV.185009/VIDEO-1

Newport, J., & Kirschner, M. (1982a). A major developmental transition in early Xenopus embryos: I. characterization and timing of cellular changes at the midblastula stage. *Cell*, *30*(3), 675–686. https://doi.org/10.1016/0092-8674(82)90272-0

Newport, J., & Kirschner, M. (1982b). A major developmental transition in early Xenopus embryos: II. Control of the onset of transcription. *Cell*, *30*(3), 687–696. https://doi.org/10.1016/0092-8674(82)90273-2

Nien, C. Y., Liang, H. L., Butcher, S., Sun, Y., Fu, S., Gocha, T., Kirov, N., Manak, J. R., & Rushlow, C. (2011). Temporal coordination of gene networks by Zelda in the early Drosophila embryo. *PLoS Genetics*, *7*(10). https://doi.org/10.1371/journal.pgen.1002339

Nitta, K. R., Jolma, A., Yin, Y., Morgunova, E., Kivioja, T., Akhtar, J., Hens, K., Toivonen, J., Deplancke, B., Furlong, E. E. M., & Taipale, J. (2015). Conservation of transcription factor binding specificities across 600 million years of bilateria evolution. *ELife*, *4*(4). https://doi.org/10.7554/ELIFE.04837

Nüsslein-Volhard, C., & Wieschaus, E. (1980). Mutations affecting segment number and polarity in Drosophila. *Nature*, *287*, 795–801.

Nüsslein-Volhard, C., Wieschaus, E., & Kluding, H. (1984a). Mutations affecting the pattern of the larval cuticle in Drosophila melanogaster - I. Zygotic loci on the second chromosome. *Wilhelm Roux's Archives of Developmental Biology*, *193*(5), 267–282. https://doi.org/10.1007/BF00848156/METRICS

Nüsslein-Volhard, C., Wieschaus, E., & Kluding, H. (1984b). Mutations affecting the pattern of the larval cuticle inDrosophila melanogaster : I. Zygotic loci on the second chromosome. *Wilhelm Roux's Archives of Developmental Biology*, *193*(5), 267–282. https://doi.org/10.1007/BF00848156

Ochoa-Espinosa, A., Yucel, G., Kaplan, L., Pare, A., Pura, N., Oberstein, A., Papatsenko, D., & Small, S. (2005). The role of binding site cluster strength in Bicoid-dependent patterning in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(14), 4960–4965. https://doi.org/10.1073/PNAS.0500373102

Pankratz, M. J., & Jäckle, H. (1990). Making stripes in the Drosophila embryo. *Trends in Genetics : TIG*, *6*(9), 287–292. https://doi.org/10.1016/0168-9525(90)90234-W

Pankratz, M. J., Seifert, E., Gerwin, N., Billi, B., Nauber, U., & Jäckle, H. (1990). Gradients of Krüppel and knirps gene products direct pair-rule gene stripe patterning in the posterior region of the Drosophila embryo. *Cell*, *61*(2), 309–317. https://doi.org/10.1016/0092-8674(90)90811-R

Pantano, L. (2017). *DEGreport: report of DEG analysis*. https://doi.org/10.18129/B9.bioc.DEGreport

Paroush, Z., Mark Wainwright, S., & Ish-Horowicz, D. (1997). Torso signalling regulates terminal patterning in Drosophila by antagonising Groucho-mediated repression. *Development*, *124*(19), 3827–3834. https://doi.org/10.1242/DEV.124.19.3827

Patel, N. H., Martin-Blanco, E., Coleman, K. G., Poole, S. J., Ellis, M. C., Kornberg, T. B., & Goodman, C. S. (1989). Expression of engrailed proteins in arthropods, annelids, and chordates. *Cell*, *58*(5), 955–968. https://doi.org/10.1016/0092-8674(89)90947-1

Pearson, J. C., Watson, J. D., & Crews, S. T. (2012). Drosophila melanogaster Zelda and Single-minded collaborate to regulate an evolutionarily dynamic CNS midline cell enhancer. *Developmental Biology*, *366*(2), 420–432. https://doi.org/10.1016/j.ydbio.2012.04.001

Perry, M. W., Boettiger, A. N., & Levine, M. (2011). Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(33), 13570–13575. https://doi.org/10.1073/PNAS.1109873108

Peter, I. S., & Davidson, E. H. (2011). Evolution of gene regulatory networks controlling body plan development. *Cell*, *144*(6), 970–985. https://doi.org/10.1016/J.CELL.2011.02.017/ASSET/1FFFFB6F-60B2-49B6-AC76-4B324505AEB5/MAIN.ASSETS/GR3.JPG

Pignoni, F., Steingrímsson, E., & Lengyel, J. A. (1992). bicoid and the terminal system activate tailless expression in the early Drosophila embryo. *Development (Cambridge, England)*, *115*(1), 239–251. https://doi.org/10.1242/DEV.115.1.239

Pimmett, V. L., Dejean, M., Fernandez, C., Trullo, A., Bertrand, E., Radulescu, O., & Lagha, M. (2021). Quantitative imaging of transcription in living Drosophila embryos reveals the impact of core promoter motifs on promoter state dynamics. *Nature Communications 2021 12:1*, *12*(1), 1–16. https://doi.org/10.1038/s41467-021-24461-6

Pires, C. V, Freitas, F. C., Cristino, A. S., Dearden, P. K., & Simoes, Z. L. (2016). Transcriptome Analysis of Honeybee (Apis Mellifera) Haploid and Diploid Embryos Reveals Early Zygotic Transcription during Cleavage. *PLoS One*, *11*(1), e0146447. https://doi.org/10.1371/journal.pone.0146447

Polach, K. J., & Widom, J. (1995). Mechanism of Protein Access to Specific DNA Sequences in Chromatin: A Dynamic Equilibrium Model for Gene Regulation. *Journal of Molecular Biology*, *254*(2), 130–149. https://doi.org/10.1006/JMBI.1995.0606

Porcher, A., Abu-Arish, A., Huart, S., Roelens, B., Fradin, C., & Dostatni, N. (2010). The time to measure positional information: maternal Hunchback is required for the synchrony of the Bicoid transcriptional response at the onset of zygotic transcription. *Development*, *137*(16), 2795–2804. https://doi.org/10.1242/DEV.051300

Price, D., Rabinovitch, S., O'Farrell, P. H., & Campbell, S. D. (2000). Drosophila wee1 has an essential role in the nuclear divisions of early embryogenesis. *Genetics*, *155*(1), 159. https://doi.org/10.1093/GENETICS/155.1.159

Pritchard, D. K., & Schubiger, G. (1996). Activation of transcription in Drosophila embryos is a gradual process mediated by the nucleocytoplasmic ratio. *Genes & Development*, *10*(9), 1131–1142. https://doi.org/10.1101/GAD.10.9.1131

Priyam, A., Woodcroft, B. J., Rai, V., Moghul, I., Munagala, A., Ter, F., Chowdhary, H., Pieniak, I., Maynard, L. J., Gibbins, M. A., Moon, H., Davis-Richardson, A., Uludag, M., Watson-Haigh, N. S., Challis, R., Nakamura, H., Favreau, E., Gomez, E. A., Pluskal, T., … Wurm, Y. (2019). Sequenceserver: A Modern Graphical User Interface for Custom BLAST Databases. *Mol Biol Evol*, *36*(12), 2922–2924. https://doi.org/10.1093/molbev/msz185

Rafiqi, A. M., Lemke, S., Ferguson, S., Stauber, M., & Schmidt-Ott, U. (2008). Evolutionary origin of the amnioserosa in cyclorrhaphan flies correlates with spatial and temporal expression changes of zen. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(1), 234–239. https://doi.org/10.1073/pnas.0709145105

Rafiqi, A. M., Park, C. H., Kwan, C. W., Lemke, S., & Schmidt-Ott, U. (2012). BMP-dependent serosa and amnion specification in the scuttle fly Megaselia abdita. *Development*, *139*(18), 3373–3382. https://doi.org/10.1242/dev.083873

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

Rambaut, A. (2018). *Figtree ver 1.4.4*. Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK. http://tree.bio.ed.ac.uk/software/figtree/

Ravindranath, A., & Cadigan, K. M. (2014). Structure-function analysis of the C-clamp of TCF/Pangolin in Wnt/ß-catenin signaling. *PloS One*, *9*(1). https://doi.org/10.1371/JOURNAL.PONE.0086180

Ray, M., Conard, A. M., Urban, J., Mahableshwarkar, P., Aguilera, J., Huang, A., Vaidyanathan, S., & Larschan, E. (2023). Sex-specific splicing occurs genome-wide during early Drosophila embryogenesis. *ELife*, *12*. https://doi.org/10.7554/ELIFE.87865

Ray, P., De, S., Mitra, A., Bezstarosti, K., Demmers, J. A., Pfeifer, K., & Kassis, J. A. (2016). Combgap contributes to recruitment of Polycomb group proteins in Drosophila. *Proc Natl Acad Sci U S A*, *113*(14), 3826–3831. https://doi.org/10.1073/pnas.1520926113

Reddington, J. P., Garfield, D. A., Sigalova, O. M., Karabacak Calviello, A., Marco-Ferreres, R., Girardot, C., Viales, R. R., Degner, J. F., Ohler, U., & Furlong, E. E. M. (2020). Lineage-Resolved Enhancer and Promoter Usage during a Time Course of Embryogenesis. *Dev Cell*, *55*(5), 648-664 e9. https://doi.org/10.1016/j.devcel.2020.10.009

Reddy, B. A., Bajpe, P. K., Bassett, A., Moshkin, Y. M., Kozhevnikova, E., Bezstarosti, K., Demmers, J. A. A., Travers, A. A., & Verrijzer, C. P. (2010). Drosophila Transcription Factor Tramtrack69 Binds MEP1 To Recruit the Chromatin Remodeler NuRD. *Molecular and Cellular Biology*, *30*(21), 5234. https://doi.org/10.1128/MCB.00266-10

Reuter, R., & Leptin, M. (1994). Interacting functions of snail, twist and huckebein during the early development of germ layers in Drosophila. *Development (Cambridge, England)*, *120*(5), 1137–1150. https://doi.org/10.1242/DEV.120.5.1137

Ribeiro, L., Tobias-Santos, V., Santos, D., Antunes, F., Feltran, G., de Souza Menezes, J., Aravind, L., Venancio, T. M., & Nunes da Fonseca, R. (2017). Evolution and multiple roles of the Pancrustacea specific transcription factor zelda in insects. *PLoS Genet*, *13*(7), e1006868. https://doi.org/10.1371/journal.pgen.1006868

Rivera-Pomar, R., & Jäckle, H. (1996). From gradients to stripes in Drosophila embryogenesis: filling in the gaps. *Trends in Genetics : TIG*, *12*(11), 478–483. https://doi.org/10.1016/0168-9525(96)10044-5

Rivera-Pomar, R., Lu, X., Perrimon, N., Taubert, H., & Jäckle, H. (1995). Activation of posterior gap gene expression in the Drosophila blastoderm. *Nature 1995 376:6537*, *376*(6537), 253–256. https://doi.org/10.1038/376253a0

Rivera-Pomar, R., Niessing, D., Schmidt-Ott, U., Gehring, W. J., & Jackle, H. (1996). RNA binding and translational suppression by bicoid. *Nature 1996 379:6567*, *379*(6567), 746–749. https://doi.org/10.1038/379746a0

Rohr, K. B., Tautz, D., & Sander, K. (1999). Segmentation gene expression in the mothmidge Clogmia albipunctata (Diptera, psychodidae) and other primitive dipterans. *Dev Genes Evol*, *209*(3), 145–154. https://doi.org/10.1007/s004270050238

Rohr, K., Tautz, D., & Sander, K. (1999). Segmentation gene expression in the mothmidge Clogmia albipunctata (Diptera, Psychodidae) and other primitive dipterans. *Dev Genes Evol*, *209*, 145–154.

Rudolph, T., Yonezawa, M., Lein, S., Heidrich, K., Kubicek, S., Schäfer, C., Phalke, S., Walther, M., Schmidt, A., Jenuwein, T., & Reuter, G. (2007). Heterochromatin Formation in Drosophila Is Initiated through Active Removal of H3K4 Methylation by the LSD1 Homolog SU(VAR)3-3. *Molecular Cell*, *26*(1), 103–115. https://doi.org/10.1016/J.MOLCEL.2007.02.025

Saccone, G. (2022). A history of the genetic and molecular identification of genes and their functions controlling insect sex determination. *Insect Biochem Mol Biol*, *151*, 103873. https://doi.org/10.1016/j.ibmb.2022.103873

Sander, K. (1985). Experimental egg activation in lower dipterans (Psychoda, Smittia) by low osmolarity. *International Journal of Invertebrate Reproduction and Development*, *8*, 175–183.

Schetelig, M. F., Schmid, B. G. M., Zimowska, G., & Wimmer, E. A. (2008). Plasticity in mRNA expression and localization of orthodenticle within higher Diptera. *Evolution & Development*, *10*(6), 700–704. https://doi.org/10.1111/J.1525-142X.2008.00283.X

Schmidt-Ott, U., & Lynch, J. A. (2016). Emerging developmental genetic model systems in holometabolous insects. *Curr Opin Genet Dev*, *39*, 116–128. https://doi.org/10.1016/j.gde.2016.06.004

Schmidt-Ott, U., Rafiqi, A. M., Sander, K., & Johnston, J. S. (2009). Extremely small genomes in two unrelated dipteran insects with shared early developmental traits. *Dev Genes Evol*, *219*(4), 207–210. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19308443

Schmidt-Ott, U., Sander, K., & Technau, G. M. (1994). Expression of engrailed in embryos of a beetle and five dipteran species with special reference to the terminal regions. *Roux's Archives of Developmental Biology*, *203*(6), 298–303. https://doi.org/10.1007/BF00457800/METRICS

Schmidt-Ott, U., & Technau, G. M. (1992). Expression of en and wg in the embryonic head and brain of Drosophila indicates a refolded band of seven segment remnants. *Development*, *116*(1), 111–125. https://doi.org/10.1242/DEV.116.1.111

Schmidt-Ott, U., & Yoon, Y. (2022). Evolution and loss of beta-catenin and TCF-dependent axis specification in insects. *Curr Opin Insect Sci*, *50*, 100877. https://doi.org/10.1016/j.cois.2022.100877

Schomburg, C., Janssen, R., & Prpic, N. M. (2022). Phylogenetic analysis of forkhead transcription factors in the Panarthropoda. *Development Genes and Evolution*, *232*(1), 39–48. https://doi.org/10.1007/S00427-022-00686-3/FIGURES/3

Schroeder, M. D., Greer, C., & Gaul, U. (2011). How to make stripes: deciphering the transition from non-periodic to periodic patterns in Drosophila segmentation. *Development (Cambridge, England)*, *138*(14), 3067–3078. https://doi.org/10.1242/DEV.062141

Schroeder, M. D., Pearce, M., Fak, J., Fan, H. Q., Unnerstall, U., Emberly, E., Rajewsky, N., Siggia, E. D., & Gaul, U. (2004a). Transcriptional Control in the Segmentation Gene Network of Drosophila. *PLOS Biology*, *2*(9), e271. https://doi.org/10.1371/JOURNAL.PBIO.0020271

Schroeder, M. D., Pearce, M., Fak, J., Fan, H.-Q., Unnerstall, U., Emberly, E., Rajewsky, N., Siggia, E. D., & Gaul, U. (2004b). Transcriptional control in the segmentation gene network of Drosophila. *PLOS*, *2*, e271.

Schulz, C., & Tautz, D. (1995). Zygotic caudal regulation by hunchback and its role in abdominal segment formation of the Drosophila embryo. *Development*, *121*(4), 1023–1028.

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

Schulz, K. N., & Harrison, M. M. (2019). Mechanisms regulating zygotic genome activation. *Nature Reviews Genetics*, *20*. https://doi.org/https://doi.org/10.1038/s41576-018-0087-x

Schüpbach, T., & Wieschaus, E. (1986). Maternal-effect mutations altering the anterior-posterior pattern of the Drosophila embryo. *Roux's Archives of Developmental Biology 1986 195:5*, *195*(5), 302–317. https://doi.org/10.1007/BF00376063

Seller, C. A., & O'Farrell, P. H. (2018). Rif1 prolongs the embryonic S phase at the Drosophila mid-blastula transition. *PLoS Biology*, *16*(5). https://doi.org/10.1371/JOURNAL.PBIO.2005687

Semotok, J. L., Cooperstock, R. L., Pinder, B. D., Vari, H. K., Lipshitz, H. D., & Smibert, C. A. (2005). Smaug recruits the CCR4/POP2/NOT deadenylase complex to trigger maternal transcript localization in the early Drosophila embryo. *Current Biology : CB*, *15*(4), 284–294. https://doi.org/10.1016/J.CUB.2005.01.048

Shaw, P. J., Salameh, A., McGregor, A. P., Bala, S., & Dover, G. A. (2001). Divergent structure and function of the bicoid gene in Muscoidea fly species. *Evol Dev*, *3*(4), 251–262. https://doi.org/10.1046/j.1525-142x.2001.003004251.x

Shazman, S., Lee, H., Socol, Y., Mann, R. S., & Honig, B. (2014). OnTheFly: a database of Drosophila melanogaster transcription factors and their binding sites. *Nucleic Acids Res*, *42*(Database issue), D167-71. https://doi.org/10.1093/nar/gkt1165

Shermoen, A. W., McCleland, M. L., & O'Farrell, P. H. (2010). Developmental Control of Late Replication and S Phase Length. *Current Biology*, *20*(23), 2067–2077. https://doi.org/10.1016/J.CUB.2010.10.021

Shindo, Y., & Amodeo, A. A. (2021). Excess histone H3 is a competitive Chk1 inhibitor that controls cell-cycle remodeling in the early Drosophila embryo. *Current Biology*, *31*(12), 2633-2642.e6. https://doi.org/10.1016/J.CUB.2021.03.035

Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V, & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Simpson-Brose, M., Treisman, J., & Desplan, C. (1994). Synergy between the hunchback and bicoid morphogens is required for anterior patterning in Drosophila. *Cell*, *78*(5), 855–865. https://doi.org/10.1016/S0092-8674(94)90622-X

Sivell, D., Mitchell, R., Crowley, L. M., Natural History Museum Genome Acquisition, L., University of, O., Wytham Woods Genome Acquisition, L., Darwin Tree of Life Barcoding, collective, Wellcome Sanger Institute Tree of Life Management, S., Laboratory, team, Wellcome Sanger Institute Scientific Operations: Sequencing, O., Wellcome Sanger Institute Tree of Life Core Informatics, team, Tree of Life Core Informatics, collective, & Darwin Tree of Life, C. (2024). The genome sequence of the German scorpionfly, Panorpa germanica Linnaeus, 1758. *Wellcome Open Res*, *9*, 285. https://doi.org/10.12688/wellcomeopenres.22259.1

Small, S., Blair, A., & Levine, M. (1992). Regulation of even-skipped stripe 2 in the Drosophila embryo. *The EMBO Journal*, *11*(11), 4047–4057. https://doi.org/10.1002/J.1460-2075.1992.TB05498.X

Small, S., Blair, A., & Levine, M. (1996). Regulation of two pair-rule stripes by a single enhancer in the Drosophila embryo. *Developmental Biology*, *175*(2), 314–324. https://doi.org/10.1006/DBIO.1996.0117

Small, S., Kraut, R., Hoey, T., Warrior, R., & Levine, M. (1991). Transcriptional regulation of a pair-rule stripe in Drosophila. *Genes & Development*, *5*(5), 827–839. https://doi.org/10.1101/GAD.5.5.827

Smibert, C. A., Wilson, J. E., Kerr, K., & Macdonald, P. M. (1996). smaug protein represses translation of unlocalized nanos mRNA in the Drosophila embryo. *Genes & Development*, *10*(20), 2600–2609. https://doi.org/10.1101/GAD.10.20.2600

Smit, A., & Hubley, R. (2008). *RepeatModeler*. https://www.repeatmasker.org/

Smit, A., Hubley, R., & Green, P. (2013). *RepeatMasker*. https://www.repeatmasker.org/

Smits, C. M., & Shvartsman, S. Y. (2020). The design and logic of terminal patterning in Drosophila. *Current Topics in Developmental Biology*, *137*, 193–217. https://doi.org/10.1016/BS.CTDB.2019.11.008

Soluri, I. V., Zumerling, L. M., Parra, O. A. P., Clark, E. G., & Blythe, S. A. (2020). Zygotic pioneer factor activity of Odd-paired/Zic is necessary for establishing the Drosophila Segmentation Network. *ELife*. https://doi.org/10.1101/852707

Sommer, R., & Tautz, D. (1991). Segmentation gene expression in the housefly Musca domestica. *Development*, *113*(2), 419–430. https://doi.org/10.1242/DEV.113.2.419

Soruco, M. M., Chery, J., Bishop, E. P., Siggers, T., Tolstorukov, M. Y., Leydon, A. R., Sugden, A. U., Goebel, K., Feng, J., Xia, P., Vedenko, A., Bulyk, M. L., Park, P. J., & Larschan, E. (2013). The CLAMP protein links the MSL complex to the X chromosome during

Drosophila dosage compensation. *Genes Dev*, *27*(14), 1551–1556. https://doi.org/10.1101/gad.214585.113

St Johnston, D., & Nüsslein-Volhard, C. (1992). The origin of pattern and polarity in the Drosophila embryo. *Cell*, *68*(2), 201–219. https://doi.org/10.1016/0092-8674(92)90466-p

Stanojevic, D., Small, S., & Levine, M. (1991). Regulation of a segmentation stripe by overlapping activators and repressors in the Drosophila embryo. *Science (New York, N.Y.)*, *254*(5036), 1385–1387. https://doi.org/10.1126/SCIENCE.1683715

Stauber, M., Jäckle, H., & Schmidt-Ott, U. (1999). The anterior determinant bicoid of Drosophila is a derived Hox class 3 gene. *Proc. Natl. Acad. Sci. USA*, *96*, 3786–3789.

Stauber, M., Lemke, S., & Schmidt-Ott, U. (2008). Expression and regulation of caudal in the lower cyclorrhaphan fly Megaselia. *Development Genes and Evolution*, *218*(2), 81. https://doi.org/10.1007/S00427-008-0204-5

Stauber, M., Prell, A., & Schmidt-Ott, U. (2002). A single Hox3 gene with composite bicoid and zerknüllt expression characteristics in non-Cyclorrhaphan flies. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(1), 274–279. https://doi.org/10.1073/pnas.012292899

Stauber, M., Taubert, H., & Schmidt-Ott, U. (2000). Function of bicoid and hunchback homologs in the basal cyclorrhaphan fly Megaselia (Phoridae). *Proceedings of the National Academy of Sciences of the United States of America*, *97*(20), 10844–10849. https://doi.org/10.1073/pnas.190095397

Staudt, N., Fellert, S., Chung, H. R., Jäckle, H., & Vorbrüggen, G. (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. https://doi.org/10.1091/MBC.E05-11-1056

Strong, I. J. T., Lei, X., Chen, F., Yuan, K., & O'Farrell, P. H. (2020). Interphase-arrested Drosophila embryos activate zygotic gene expression and initiate mid-blastula transition events at a low nuclear-cytoplasmic ratio. *PLoS Biology*, *18*(10). https://doi.org/10.1371/JOURNAL.PBIO.3000891

Struhl, G. (1989). Differing strategies for organizing anterior and posterior body pattern in Drosophila embryos. *Nature*, *338*(6218), 741–744. https://doi.org/10.1038/338741a0

Struhl, G., Johnston, P., & Lawrence, P. A. (1992). Control of Drosophila body pattern by the hunchback morphogen gradient. *Cell*, *69*(2), 237–249. https://doi.org/10.1016/0092-8674(92)90405-2

Struhl, G., Struhl, K., & Macdonald, P. M. (1989). The gradient morphogen bicoid is a concentration-dependent transcriptional activator. *Cell*, *57*(7), 1259–1273. https://doi.org/10.1016/0092-8674(89)90062-7

Stumpff, J., Duncan, T., Homola, E., Campbell, S. D., & Su, T. T. (2004). Drosophila Wee1 Kinase Regulates Cdk1 and Mitotic Entry during Embryogenesis. *Current Biology : CB*, *14*(23), 2143. https://doi.org/10.1016/J.CUB.2004.11.050

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 Res*, *25*(11), 1703–1714. https://doi.org/10.1101/gr.192542.115

Surkova, S., Golubkova, E., Mamon, L., & Samsonova, M. (2018). Dynamic maternal gradients and morphogenetic networks in Drosophila early embryo. *Biosystems*, *173*, 207–213. https://doi.org/10.1016/j.biosystems.2018.10.009

Surkova, S., Kosman, D., Kozlov, K., Manu, Myasnikova, E., Samsonova, A. A., Spirov, A., Vanario-Alonso, C. E., Samsonova, M., & Reinitz, J. (2008). Characterization of the Drosophila segment determination morphome. *Developmental Biology*, *313*(2), 844–862. https://doi.org/10.1016/J.YDBIO.2007.10.037

Swantek, D., & Gergen, J. P. (2004). Ftz modulates Runt-dependent activation and repression of segment-polarity gene transcription. *Development (Cambridge, England)*, *131*(10), 2281–2290. https://doi.org/10.1242/DEV.01109

Tadros, W., Goldman, A. L., Babak, T., Menzies, F., Vardy, L., Orr-Weaver, T., Hughes, T. R., Westwood, J. T., Smibert, C. A., & Lipshitz, H. D. (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), 143–155. https://doi.org/10.1016/J.DEVCEL.2006.10.005

Tadros, W., Houston, S. A., Bashirullah, A., Cooperstock, R. L., Semotok, J. L., Reed, B. H., & Lipshitz, H. D. (2003). Regulation of Maternal Transcript Destabilization During Egg Activation in Drosophila. *Genetics*, *164*(3), 989–1001. https://doi.org/10.1093/GENETICS/164.3.989

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

Tautz, D. (1988). Regulation of the Drosophila segmentation gene hunchback by two maternal morphogenetic centers. *Nature (London)*, *332*, 281–284.

ten Bosch, J. R., Benavides, J. A., & Cline, T. W. (2006). The TAGteam DNA motif controls the timing of Drosophila pre-blastoderm transcription. *Development*, *133*(10), 1967–1977. https://doi.org/10.1242/dev.02373

Tenger-Trolander, A., Amiri, E., Gantz, V., Kwan, C. W., Sanders, S. A., & Schmidt-Ott, U. (2025). Genomic resources for the scuttle fly Megaselia abdita: A Model Organism for Comparative Developmental Studies in Flies. *BioRxiv*.

Thomsen, S., Anders, S., Janga, S. C., Huber, W., & Alonso, C. R. (2010). Genome-wide analysis of mRNA decay patterns during early Drosophila development. *Genome Biology*, *11*(9), 1–27. https://doi.org/10.1186/GB-2010-11-9-R93/FIGURES/10

Tkačik, G., & Gregor, T. (2021). The many bits of positional information. *Development*, *148*(2). https://doi.org/10.1242/dev.176065

Treier, M., Pfeifle, C., & Tautz, D. (1989). Comparison of the gap segmentation gene hunchback between Drosophila melanogaster and Drosophila virilis reveals novel modes of evolutionary change. *The EMBO Journal*, *8*(5), 1517–1525. https://doi.org/10.1002/J.1460-2075.1989.TB03536.X

Tsukiyama, T., Becker, P. B., & Wu, C. (1994). ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor. *Nature*, *367*(6463), 525–532. https://doi.org/10.1038/367525A0;KWRD=SCIENCE

Urbansky, S., Gonzalez Avalos, P., Wosch, M., & Lemke, S. (2016). Folded gastrulation and T48 drive the evolution of coordinated mesoderm internalization in flies. *Elife*, *5*. https://doi.org/10.7554/eLife.18318

Van de Wetering, M., Cavallo, R., Dooijes, D., Van Beest, M., Van Es, J., Loureiro, J., Ypma, A., Hursh, D., Jones, T., Bejsovec, A., Peifer, M., Mortin, M., & Clevers, H. (1997).

Armadillo coactivates transcription driven by the product of the Drosophila segment polarity gene dTCF. *Cell*, *88*(6), 789–799. https://doi.org/10.1016/S0092-8674(00)81925-X

van den Heuvel, M., Nusse, R., Johnston, P., & Lawrence, P. A. (1989). Distribution of the wingless gene product in Drosophila embryos: a protein involved in cell-cell communication. *Cell*, *59*(4), 739–749. https://doi.org/10.1016/0092-8674(89)90020-2

Van Der Zee, M., Berns, N., & Roth, S. (2005). Distinct functions of the Tribolium zerknüllt genes in serosa specification and dorsal closure. *Current Biology*, *15*(7), 624–636. https://doi.org/10.1016/j.cub.2005.02.057

Ventos-Alfonso, A., Ylla, G., & Belles, X. (2019). Zelda and the maternal-to-zygotic transition in cockroaches. *FEBS J*, *286*(16), 3206–3221. https://doi.org/10.1111/febs.14856

Vicoso, B., & Bachtrog, D. (2015). Numerous transitions of sex chromosomes in Diptera. *PLoS Biol*, *13*(4), e1002078. https://doi.org/10.1371/journal.pbio.1002078

Voordeckers, K., Pougach, K., & Verstrepen, K. J. (2015). How do regulatory networks evolve and expand throughout evolution? *Current Opinion in Biotechnology*, *34*, 180–188. https://doi.org/10.1016/J.COPBIO.2015.02.001

Walldorf, U., & Gehring, W. J. (1992). Empty spiracles, a gap gene containing a homeobox involved in Drosophila head development. *The EMBO Journal*, *11*(6), 2247. https://doi.org/10.1002/J.1460-2075.1992.TB05284.X

Walldorf, U., Kiewe, A., Wickert, M., Ronshaugen, M., & McGinnis, W. (2000). Homeobrain, a novel paired-like homeobox gene is expressed in the Drosophila brain. *Mech Dev*, *96*(1), 141–144. https://doi.org/10.1016/s0925-4773(00)00380-4

Wang, C., & Lehmann, R. (1991). Nanos is the localized posterior determinant in Drosophila. *Cell*, *66*(4), 637–647. https://doi.org/10.1016/0092-8674(91)90110-K

Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., Lee, T. H., Jin, H., Marler, B., Guo, H., Kissinger, J. C., & Paterson, A. H. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res*, *40*(7), e49. https://doi.org/10.1093/nar/gkr1293

Weigel, D., Jürgens, G., Klingler, M., & Jäckle, H. (1990). Two gap genes mediate maternal terminal pattern information in Drosophila. *Science (New York, N.Y.)*, *248*(4954), 495–498. https://doi.org/10.1126/SCIENCE.2158673

Wharton, R. P., & Struhl, G. (1991). RNA regulatory elements mediate control of Drosophila body pattern by the posterior morphogen nanos. *Cell*, *67*(5), 955–967. https://doi.org/10.1016/0092-8674(91)90368-9

Wiegmann, B. M., Trautwein, M. D., Winkler, I. S., Barr, N. B., Kim, J. W., Lambkin, C., Bertone, M. A., Cassel, B. K., Bayless, K. M., Heimberg, A. M., Wheeler, B. M., Peterson, K. J., Pape, T., Sinclair, B. J., Skevington, J. H., Blagoderov, V., Caravas, J., Kutty, S. N., Schmidt-Ott, U., … Yeates, D. K. (2011). Episodic radiations in the fly tree of life. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(14), 5690–5695. https://doi.org/10.1073/pnas.1012675108

Wieschaus, E. (2016). Positional Information and Cell Fate Determination in the Early Drosophila Embryo. *Curr Top Dev Biol*, *117*, 567–579. https://doi.org/10.1016/bs.ctdb.2015.11.020

Wieschaus, E., Nüsslein-Volhard, C., & Jürgens, G. (1984). Mutations affecting the pattern of the larval cuticle inDrosophila melanogaster : III. Zygotic loci on the X-chromosome and fourth chromosome. *Wilhelm Roux's Archives of Developmental Biology*, *193*(5), 296–307. https://doi.org/10.1007/BF00848158

Wimmer, E. A., Carleton, A., Harjes, P., Turner, T., & Desplan, C. (2000). Bicoid-independent formation of thoracic segments in Drosophila. *Science (New York, N.Y.)*, *287*(5462), 2476–2479. https://doi.org/10.1126/SCIENCE.287.5462.2476

Wotton, K. R., Jimenez-Guri, E., Crombach, A., Janssens, H., Alcaine-Colet, A., Lemke, S., Schmidt-Ott, U., & Jaeger, J. (2015). Quantitative system drift compensates for altered maternal inputs to the gap gene network of the scuttle fly Megaselia abdita. *ELife*, *4*(4). https://doi.org/10.7554/ELIFE.04785

Wotton, K. R., Jiménez-Guri, E., & Jaeger, J. (2015). Maternal Co-ordinate Gene Regulation and Axis Polarity in the Scuttle Fly Megaselia abdita. *PLoS Genetics*, *11*(3), 1–20. https://doi.org/10.1371/journal.pgen.1005042

Wratten, N. S., McGregor, A. P., Shaw, P. J., & Dover, G. A. (2006). Evolutionary and functional analysis of the tailless enhancer in Musca domestica and Drosophila melanogaster. *Evolution & Development*, *8*(1), 6–15. https://doi.org/10.1111/J.1525-142X.2006.05070.X

Xiao, Z., & Lam, H. M. (2022). ShinySyn: a Shiny/R application for the interactive visualization and integration of macro- and micro-synteny data. *Bioinformatics*, *38*(18), 4406–4408. https://doi.org/10.1093/bioinformatics/btac503

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 and Development*, *28*(6), 608–621. https://doi.org/10.1101/gad.234534.113

Yoon, Y. (2019). *EVOLUTION OF EMBRYONIC AXIS DETERMINANTS VIA ALTERNATIVE TRANSCRIPTION* [The University of Chicago]. https://doi.org/.1037//0033-2909.I26.1.78

Yoon, Y., Klomp, J., Martin-Martin, I., Criscione, F., Calvo, E., Ribeiro, J., & Schmidt-Ott, U. (2019). Embryo polarity in moth flies and mosquitoes relies on distinct old genes with localized transcript isoforms. *ELife*, *8*, 1–30. https://doi.org/10.7554/eLife.46711

Young, P. E., Richman, A. M., Ketchum, A. S., & Kiehart, D. P. (1993). Morphogenesis in Drosophila requires nonmuscle myosin heavy chain function. *Genes & Development*, *7*(1). https://doi.org/10.1101/GAD.7.1.29

Zaret, K. S. (2020). Pioneer Transcription Factors Initiating Gene Network Changes. *Annual Review of Genetics*, *54*, 367–385. https://doi.org/10.1146/annurev-genet-030220-015007

Zavortink, M., Rutt, L. N., Dzitoyeva, S., Henriksen, J. C., Barrington, C., Bilodeau, D. Y., Wang, M., Chen, X. X. L., & Rissland, O. S. (2020). The E2 Marie Kondo and the CTLH E3 ligase clear deposited RNA binding proteins during the maternal-to-zygotic transition. *ELife*, *9*, 1–48. https://doi.org/10.7554/ELIFE.53889

Zenk, F., Loeser, E., Schiavo, R., Kilpert, F., Bogdanovic, O., & Iovino, N. (2017). Germ line–inherited H3K27me3 restricts enhancer function during maternal-to-zygotic transition. *Science*, *357*(6347), 212–216. https://doi.org/10.1126/SCIENCE.AAM5339/SUPPL_FILE/AAM5339_ZENK_SM.PDF

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., & Liu, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol*, *9*(9), R137. https://doi.org/10.1186/gb-2008-9-9-r137

Zhu, L. J., Christensen, R. G., Kazemian, M., Hull, C. J., Enuameh, M. S., Basciotta, M. D., Brasefield, J. A., Zhu, C., Asriyan, Y., Lapointe, D. S., Sinha, S., Wolfe, S. A., & Brodsky, M. H. (2011). FlyFactorSurvey: a database of Drosophila transcription factor binding

specificities determined using the bacterial one-hybrid system. *Nucleic Acids Res*, *39*(Database issue), D111-7. https://doi.org/10.1093/nar/gkq858