

RESEARCH ARTICLE

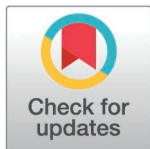
Asymmetric chromatin accessibility underlies anterior-posterior axis specification in moth fly embryos

Ezra E. Amiri¹*, Muzi Li¹*, Ayse Tenger-Trolander¹, Maxwell Devine¹, Alexander T. Julian², Koray Kasan¹, Sheri A. Sanders^{3,4}, Shelby A. Blythe^{5,6*}, Urs Schmidt-Ott^{1*}

1 Department of Organismal Biology and Anatomy, The University of Chicago, Chicago, Illinois, United States of America, **2** Department of Biology, Illinois Institute of Technology, Chicago, Illinois, United States of America, **3** University of Notre Dame, Department of Biological Sciences, Notre Dame, Indiana, United States of America, **4** Kallista Consulting, Niles, Michigan, United States of America, **5** Department of Molecular Biosciences, Northwestern University, Evanston, Illinois, United States of America, **6** National Institute for Theory and Mathematics in Biology, Northwestern University and The University of Chicago, Chicago, Illinois, United States of America

* These authors contributed equally to this work.

* shelby.blythe@northwestern.edu (SAB); uschmid@uchicago.edu (US)



OPEN ACCESS

Citation: Amiri EE, Li M, Tenger-Trolander A, Devine M, Julian AT, Kasan K, et al. (2026) Asymmetric chromatin accessibility underlies anterior-posterior axis specification in moth fly embryos. PLoS Biol 24(8): e3003896. <https://doi.org/10.1371/journal.pbio.3003896>

Academic Editor: François Schweisguth, Institut Pasteur, FRANCE

Received: August 14, 2025

Accepted: June 24, 2026

Published: August 6, 2026

Copyright: © 2026 Amiri et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: 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

Abstract

Establishing the anterior-posterior (AP) body axis is a fundamental process during embryogenesis, and the fruit fly, *Drosophila melanogaster*, provides one of the best-known case studies. But for unknown reasons, different species of flies (Diptera) establish the AP axis through unrelated, structurally distinct anterior determinants. The anterior determinant of *Drosophila*, Bicoid (Bcd), initiates symmetry-breaking during nuclear cleavage cycles (NCs) when ubiquitous pioneer factors, such as Zelda (Zld), drive zygotic genome activation (ZGA) at the level of chromatin accessibility by nucleosome depletion. While Bcd engages in a concentration-dependent competition with nucleosomes at the loci of a small set of transcription factor (TF) genes that are expressed in the anterior embryo, it remains unknown whether unrelated anterior determinants of other fly species function in the same way and target homologous genes. We have examined the symmetry-breaking mechanism of a moth fly, *Clogmia albipunctata*, in which a maternally expressed transcript isoform of the pair-rule segmentation gene *odd-paired* serves as an anterior determinant. We provide a de novo assembly and annotation of the *Clogmia* genome, report changes in chromatin accessibility during the nuclear cleavage cycles (NCs) of consecutive blastoderm stages, and describe how *Clogmia*'s orthologs of *zelda* (*Cal-zld*) and *odd-paired* (*Cal-opa*) affect chromatin accessibility and gene expression. We document extensive opening and closing of chromatin regions during cleavage cycles of blastoderm, important roles of *Cal-zld* in opening chromatin and driving zygotic gene expression, and show that maternal *Cal-opa* activity initiates zygotic symmetry-breaking along the AP axis by driving chromatin accessibility and expression at *Clogmia*'s *homeo-brain* and *sloppy-paired* loci. These genes are not known as key targets of Bcd but

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, including those samples that were not generated in this study but used as evidence for the annotation of the genome, are available as supplemental information (S1 Table). Scripts: Input files and custom R scripts used to parse MCSanX collinearity output and to generate circular synteny plots (Fig 1) with the R package 'circlize' have been deposited at the Zenodo repository <https://doi.org/10.5281/zenodo.20089250>. Input files and custom R scripts used to import and analyze all ATAC and RNA-seq datasets (Figs 2–9) have been deposited at the Zenodo repository <https://doi.org/10.5281/zenodo.20140158>. Genome Browser: The genomic data produced in this study can be further interrogated using a genome browser ecosystem centered on JBrowse2 [179] and delivered as a cloud image designed for individual use, currently available on NSF's Jetstream2 cloud platform, with a portable Docker image with documentation and BSGenome package created for R available at Clogmia - github: https://github.com/kallistaconsulting/genomic_resources_clogmia.

Funding: Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health (R01GM127366 to U.S.), by the Pew Scholars Program in the Biomedical Sciences of the Pew Charitable Trusts (to S.A.B.), predoctoral fellowships from the National Institute of General Medical Sciences (T32 GM139782 and T32 GM007183 to E.E.A.), and used Jetstream2 at Indiana University through allocation from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program (BIO220075 to U.S.), which is supported by National Science Foundation grants nos. #2138259, #2138286, #2138307, #2137603, and #2138296. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: S.A.S. is self-employed at Kallista Consulting.

Abbreviations: AP, anterior-posterior; BUSCO, Benchmarking Universal Single-Copy

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 insects.

Introduction

The insect order of flies (Diptera) offers excellent opportunities for comparative study of developmental mechanisms, as many dipteran species can be reared cost-effectively and because the fruit fly, *Drosophila melanogaster*, provides a well-established model for functional comparisons at the molecular and cellular level. Dipteran embryos are morphologically similar across the clade [1,2]. However, the mechanisms of axis specification [3,4], primordial germ cell specification [3], extra-embryonic tissue specification [5,6], morphogenesis [5,7–11], and sex determination [12–15] differ between dipteran species. Explaining such unexpected plasticity in developmental gene networks requires functional comparisons, which can reveal underlying general principles and mechanisms of evolutionary change. Anterior-posterior (AP) axis specification, a long-standing model for pattern formation and gene regulation in *Drosophila* [16–24], is particularly suitable for comparative functional studies because the syncytial nature of early dipteran embryos facilitates visualization and perturbation of gene activity in nontraditional model organisms. Many dipterans establish the embryo's head-to-tail polarity through transcription factors (TFs) that diffuse from maternally localized mRNA and form AP concentration gradients that regulate gene expression in a dose-dependent manner [25,26]. These anterior determinants prevent the formation of bicaudal embryos that lack head and thorax (double abdomen). Irrespective of this fundamental role, unrelated anterior determinants have been found in different fly species. While the anterior determinant of *Drosophila* and many other brachyceran fly species is encoded by a homeobox gene, *bicoid* [27–32], diverse zinc finger genes take on this role in lower dipterans, including *pangolin* (dipteran ortholog *Tcf* gene family) in anopheline mosquitoes and crane flies, *cucoid* in culicine mosquitoes, *panish* in certain harlequin flies (Chironomina), and *odd-paired* (dipteran ortholog of *Zic* gene family) in moth flies (Psychodidae) [3,4]. The phylogenetic occurrence of these anterior determinants indicates that they have been frequently replaced during the dipteran radiation. Moreover, the anterior determinant proteins identified so far 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 Odd-paired (Opa) and Cucoid through distinct C2H2 zinc finger domains [33–38]. This molecular diversity prompted us to ask whether the mechanisms of dipteran anterior determinants and their target genes differ between species or whether they are conserved due to contextual or downstream network constraints. In this study, we address these questions by analyzing the mechanism of action of the anterior determinant of the moth fly *Clogmia albipunctata*, a C2H2 zinc finger protein encoded by the *zic* gene family member *odd-paired* [3] and comparing it to the anterior determinant of *Drosophila melanogaster*, a homeodomain protein encoded by *bicoid* [27], which is missing in *Clogmia* and other midges.

Orthologs; COG, Cluster-of-Orthologous-Gene; CPM, counts per million; DGE, differential gene expression; FC, fold change; GO, Gene Ontology; HCR, hybridization chain reaction; NC, nuclear cycle; TEs, transposable elements; TF, transcription factor; TSS, transcription start site; UDI, Unique Dual Indexes; ZGA, zygotic genome activation.

Fly embryos begin embryonic development with a series of synchronized intravitel-line nuclear divisions. The nuclei are then arranged in a monolayer between the egg membrane and the central yolk sac. This ‘syncytial blastoderm’ undergoes additional nuclear cleavage cycles before it cellularizes and initiates gastrulation. The embryos of *Drosophila* and *Clogmia* undergo 13 nuclear division cycles, form the syncytial blastoderm at the end of nuclear cycle (NC) 9 (*Clogmia*) or the beginning of NC10 (*Drosophila*), and cellularize it during NC14, thereby ending the syncytial phase of embryonic development and setting the stage for gastrulation.

In *Drosophila*, only a few hundred genes are expressed before the midblastula transition in NC14 [39–41], when gap phases are added to the now desynchronized mitotic cycle and the embryo transitions from maternal to zygotic control of development [42]. Genes expressed during NC8–NC13 include regulators of the cell cycle, sex determination, dosage compensation, and early pattern formation [43–45]. However, widespread zygotic genome activation (ZGA) is delayed until NC14 when the prolongation of the mitotic cycle enables the expression of longer transcripts and provides more time for the regulation of global changes in chromatin accessibility [45–47]. After each DNA replication cycle, nucleosome formation reduces chromatin accessibility, creating a barrier to TF binding. This barrier is overcome by a special class of sequence-specific TFs capable of binding nucleosome occupied DNA and destabilizing nucleosomes [48,49]. The major driver of chromatin accessibility during *Drosophila*’s nuclear cleavage cycles is Zelda (Zld), a ubiquitously expressed zinc finger protein [44,50–53]. Zld’s activity is required throughout ZGA and enables the binding of other TFs, including those involved in embryonic patterning [51,54–58]. Two additional pioneer TFs, 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 [59–62].

At the beginning of the blastoderm stage, chromatin accessibility measured by ATAC-seq (assay for transposase-accessible chromatin with sequencing) is largely homogeneous across the embryo. Indeed, Zld establishes chromatin accessibility ubiquitously in the embryo at many sites that are targeted by Bcd [55,57,63–69]. However, Bcd’s role in AP patterning becomes indispensable after blastoderm formation at the beginning of NC10 [70], and a small cohort of early segmentation gene enhancers that drive Bcd-dependent anterior expression exhibits anteriorly restricted accessibility [35,71–77]. The underlying mechanism has been inferred by measuring accessibility and activity of the Bcd-dependent proximal P2 enhancer element of *hunchback*, an early target of Bcd, under experimentally modulated Bcd input and nucleosome stability at the enhancer [78]. Bcd engages in a concentration-dependent competition with nucleosomes to ‘open’ otherwise inaccessible regulatory regions. Bcd activity is thereby effectively breaking symmetry along the AP axis at the level of chromatin accessibility, well before the bulk of ZGA and pattern formation occurs [79]. However, Bcd remains active during the major wave of ZGA, when the embryo establishes a segmented body plan through a regulatory network of TF encoding genes, known as gap and pair-rule genes [20,22,80,81].

In *Clogmia*, the cleavage cycles prior to NC14 are ~3 times slower than in *Drosophila* [82,83] but its segmented body plan appears, like in *Drosophila*, during NC14 [83,84]. However, the activity window of its anterior determinant, a maternal transcript isoform of *Clogmia*'s *odd-paired* gene *Cal-opa*, is constrained by its zygotic function in the pair-rule gene network, which appears to be conserved [3,85]. Knockdown of the maternal *Cal-opa* transcript by isoform-specific RNA interference (RNAi) results in double abdomens that are normally segmented. Conversely, ectopic posterior expression of *Cal-opa* by mRNA injection can induce a bicephalic (double head) phenotype, but this phenotype is only observed when embryos are injected prior to blastoderm formation; delayed injection of the same mRNA during the syncytial blastoderm stage (4 hours after egg activation) fails to induce ectopic head structures [3]. Therefore, early Cal-Opa activity is both necessary and sufficient for establishing head-to-tail polarity of the *Clogmia* embryo and does not overlap with the function of the zygotic *Cal-opa* transcript isoform, which is expressed at NC14 and encodes a nearly identical protein.

Since zygotic Opa pioneers chromatin accessibility in late NC14 *Drosophila* embryos to advance the progression of segmentation and dorsoventral patterning [76,86], we asked whether the maternal *Cal-opa* gradient of *Clogmia* breaks axial symmetry by driving chromatin accessibility at gene loci that are transcribed in the anterior blastoderm before the beginning of NC14 and whether these genes are homologous to Bcd target genes. To set the stage for addressing these questions, we first assembled and annotated the *Clogmia* genome and identified homologs of *Drosophila*'s segmentation genes. We then characterized chromatin accessibility in single wild-type embryos at NC11, NC12, NC13, and both early and late NC14 to predict putative enhancer and promoter regions and to capture their stage-dependent accessibility. These experiments revealed opening and—unlike in *Drosophila* [79]—a lesser but substantial amount of chromatin closing during the roughly 3 times longer cleavage cycles of the syncytial blastoderm of *Clogmia*. We then assessed how chromatin accessibility and zygotic transcription are affected in embryos with reduced *Cal-zld* or reduced maternal *Cal-opa* activity. These experiments revealed (in conjunction with motif enrichment analyses) an important, likely direct, role of Cal-Zld in opening chromatin and activating zygotic gene expression in *Clogmia*, and identified homologs of *sloppy-paired* and *homeobrain* as target genes of maternal Cal-Opa, both in terms of chromatin accessibility and zygotic gene expression. Finally, we show that *Clogmia*'s *homeobrain* (*Cal-hbn*) and *sloppy-paired* homologs (*Cal-slp1*, *Cal-slp2*, *Cal-slp3*) function in complementary domains across the anterior half of the blastoderm, suggesting that these genes initiate zygotic symmetry-breaking along the AP axis downstream of *Clogmia*'s anterior determinant.

Results

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 (Fig 1, Table 1, S1 Appendix), consistent with the known chromosome number in this species [87,88]. 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) [89]. 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 (S1 Table), 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), in line with recently published dipteran genomes [90–92]. While synteny between the chromosomes within the Cyclorrhapha clade can be highly conserved (time of divergence ~145 million years ago) [91], synteny between the chromosomes of *D. melanogaster* and *C. albipunctata* (divergence time ~250 million years ago) was essentially lost (Fig 1).

Next, we interrogated the genome for the presence of predicted early segmentation genes (S2 Appendix). As expected, no *bicoid* ortholog was found but we recovered two putative Hox class 3 genes to which *bicoid* belongs [95], including the previously reported homolog of *zerknüllt* (*Cal-zen*) [96] 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). Our annotation also included homologs of *Drosophila*

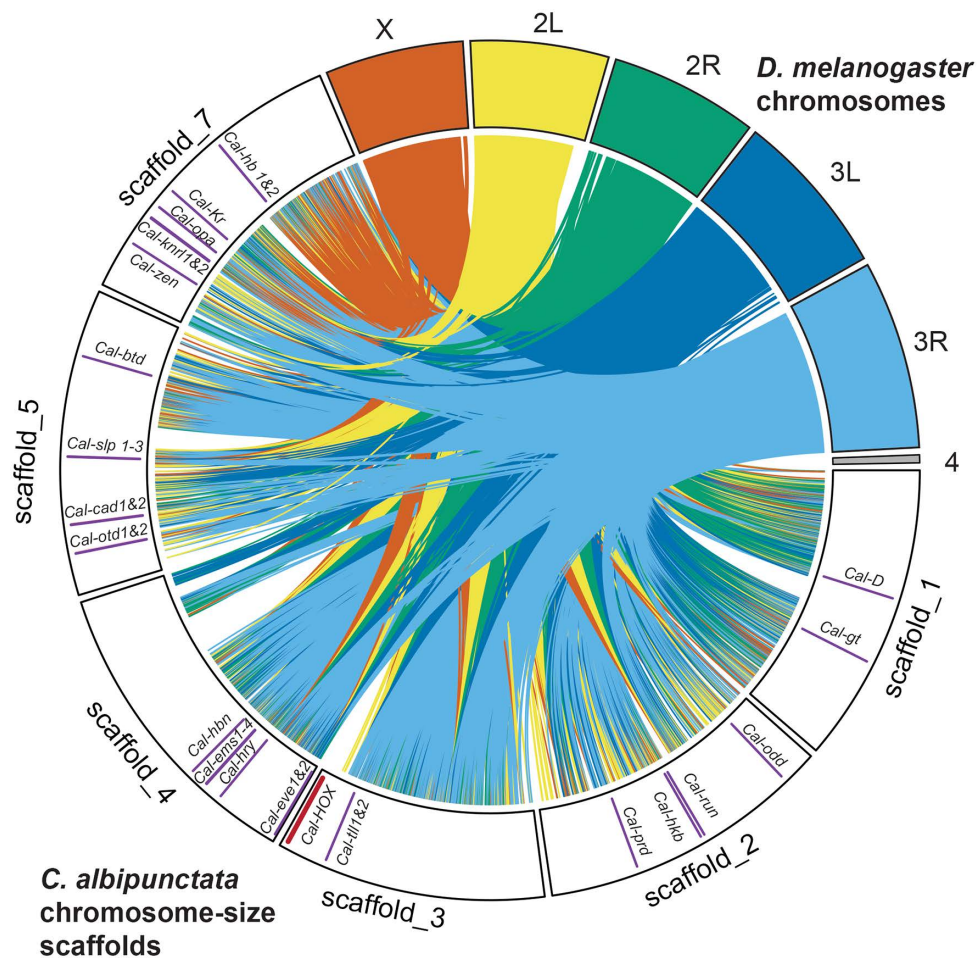


Fig 1. Synteny analysis of *Drosophila melanogaster* chromosomes and *Clogmia albipunctata* scaffolds. Synteny analysis between *Clogmia albipunctata* chromosome-sized scaffolds and *Drosophila melanogaster* chromosomes. Groups of collinear genes are represented by contact lines connecting the positions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

<https://doi.org/10.1371/journal.pbio.3003896.g001>

melanogaster's early segmentation genes, including gap genes (*tailless*, *huckebein*, *orthodenticle/ocelliless*, *empty spiracles*, *buttonhead*, *hunchback*, *Krüppel*, *giant*, and *knirps*), pair-rule genes (*even-skipped*, *fushi tarazu*, *odd-skipped*, *runt*, *sloppy-paired*, *paired*, *hairy*, and *odd-paired*), and other important regulators of the segmentation network (*caudal*, *Dichaete*) [20,22]. 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-knir1*, *Cal-knir2*), *orthodenticle* (*Cal-otd1*, *Cal-otd2*), *even-skipped* (*Cal-eve1*, *Cal-eve2*), *tailless* (*Cal-tll1*, *Cal-tll2*) and *caudal* (*Cal-cad1*, *Cal-cad2*), 3 copies of *sloppy-paired1/2/fd19B* (*Cal-slp1*, *Cal-slp2*, and *Cal-slp3*), and 4 copies of *empty spiracles* (*Cal-ems1*, *Cal-ems2*, *Cal-ems3*, and *Cal-ems4*) (Fig 1, S2 Table). Such gene duplications are a potential source of redundancies in *Clogmia*'s segmentation gene network (see below).

Patterns of chromatin opening and closing during nuclear cleavage cycles in *Clogmia* suggest coordinated regulation of groups of genes

To identify accessible chromatin before and during the main wave of ZGA, we performed single-embryo ATAC-seq at NC11, NC12, NC13, and NC14 (before cellularization), and Late NC14~20 min before gastrulation. The embryos were

Table 1. *Clogmia albipunctata* reference assembly statistics and quality metrics.

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%

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 [93]. L50: the minimum number of scaffolds whose collective length accounts for 50% of the total assembly length [94].

<https://doi.org/10.1371/journal.pbio.3003896.t001>

Table 2. *Clogmia albipunctata* annotation statistics and BUSCO scores.

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

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.

<https://doi.org/10.1371/journal.pbio.3003896.t002>

staged in vivo using developmental time after egg activation at 25 °C and morphological criteria as previously described [83] 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 32,157 open chromatin regions (peaks), using MACS2 [97]. 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 NC, suggesting that the chromatin landscape is continuously changing (Table 3, S3 Appendix). Since promoter accessibility does not necessarily indicate gene expression [98], we divided the accessible chromatin regions into transcription start site (TSS) proximal regions (within 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

Table 3. Comparison of chromatin accessibility peaks, nuclear cycle length in minutes, and haploid genome size of *Clogmia albipunctata* and *Drosophila melanogaster*.

	<i>Clogmia albipunctata</i>				
	NC11	NC12	NC13	NC14	Late NC14
Peaks	5,013	5,926	7,586	16,147	11,628
Duration	30 m	37 m	1 h 13 m	1 h 47 m	
Genome size	~304 Mb				
	<i>Drosophila melanogaster</i>				
	NC11	NC12	NC13	NC14	
Peaks	3,084	6,740	9,824	13,226	
Duration	10 m	12 m	21 m	1 h 5 m	
Genome size	~137 Mb				

Peaks for *Drosophila melanogaster* are based on [71,79] and nuclear cycle lengths are based on [83].

<https://doi.org/10.1371/journal.pbio.3003896.t003>

(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%) (S1 Fig). These observations suggest similar dynamic changes in chromatin accessibility during ZGA in *Clogmia* and *Drosophila* [99,100].

To identify peaks with dynamic patterns of chromatin accessibility, we compared the degree of accessibility of each peak between successive timepoints (Fig 2A) using DESeq2 [101]. This analysis identified a total of 6,930 dynamically regulated peaks (adjusted p -value ≤ 0.05 and $|\log_2$ fold change| ≥ 1 ; 21.6% of all 32,157 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 shared changes in accessibility from stage to stage and delineated 21 dynamic groups (S2 Fig). Nearly 80% (5,403/ 6,930) of the dynamic peaks fell into one of 8 groups, which reveal the major accessibility trends we observe in the data (Fig 2B). 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 S2 Fig). These observations suggest that remodeling of chromatin accessibility (both opening and closing) 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.

***Cal-zld* dependent gene expression and chromatin accessibility**

Previous studies in *Drosophila* have shown that temporal changes in chromatin accessibility are regulated by the interplay of TFs that can interact with nucleosome-bound DNA (pioneer factors) and TFs that have limited or no ability to initiate chromatin accessibility [47,54]. The earliest changes in chromatin accessibility are primarily driven by the pioneer factor Zld [44,52,56,99,102]. As ZGA accelerates, additional pioneer factors, such as GAF [61], CLAMP [59,60], and Opa [76,86] are required for maintaining, modulating, or gaining chromatin accessibility. In *Clogmia*, a motif enrichment analysis of peaks with the MEME software package [103,104] revealed an enrichment of a Zelda-binding motif (5'-CAGGTA), GA-rich motifs recognized by GAF and CLAMP [61,105–107], in addition to (CA)_n motifs recognized by Combgap (Cg) [108], and others (S3 Fig). Opa-binding motifs were not recovered in this analysis. These results are comparable to findings in *Drosophila* [76,79,86] and suggest that motif enrichment within chromatin accessible at ZGA is very similar between *Drosophila* and *Clogmia*.

In early *Drosophila* embryos, loss of Zld causes widespread reduction of chromatin accessibility and failure to activate a large cohort of early zygotic genes, including genes required for pattern formation and cellularization [44,52]. Both motif

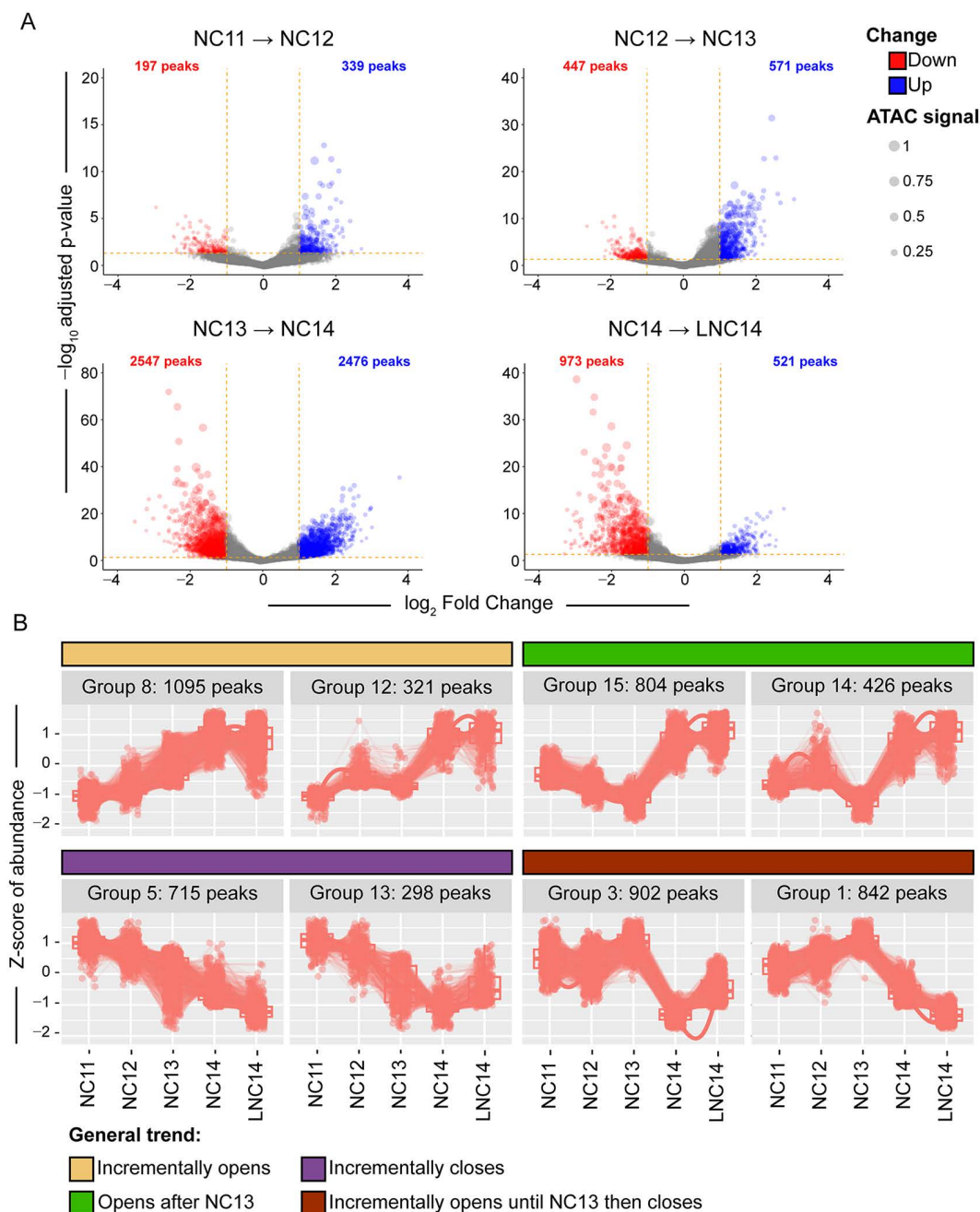


Fig 2. Changes in chromatin accessibility during NC11 to Late NC14. **A)** Volcano plots of differentially accessible ATAC-seq peaks between consecutive developmental stages. Significant change of adjusted p -value ≤ 0.05 and $|\log_2$ fold change ≥ 1 are highlighted in red ($\log_2$ fold change ≤ -1) or blue ($\log_2$ fold change ≥ 1). **B)** Dynamic peak groups were identified using DESeq2 (Pantano, 2017) (R package version 1.38.5). Y-axis, Z-score of abundance. X-axis, developmental stage. Underlying data is available in [S1 Data](https://doi.org/10.1371/journal.pbio.3003896.g002) and NCBI (bioproject: PRJNA1198932).

<https://doi.org/10.1371/journal.pbio.3003896.g002>

affinity and occupancy positively correlate with Zld's pioneering activity [44,52,54,109]. Homologs of *zelda* have been identified in many insects and some crustaceans [110], and several transcriptomic studies suggest that Zld's function as pioneer TF during ZGA is conserved across insects [111–115]. However, these comparative studies did not directly test

for Zld-dependent effects on chromatin accessibility. We wanted to know how *zelda* might affect chromatin accessibility in regions that open or close as the *Clogmia* embryo reaches NC14. We identified a single ortholog of *zelda* in the *Clogmia* genome (*Cal-zld*; [S4 Fig](#)) and observed high levels of *Cal-zld* expression at preblastoderm and blastoderm stages ([S5 Fig](#)). To examine phenotypic effects of *Cal-zld* depletion, we injected batches of ~30 embryos with *Cal-zld* dsRNA (double-stranded RNA) within the first hour of development (NC2/3) and monitored their development in vivo until the end of NC14 (roughly 7–7.5 hours after injection). Preblastoderm mortality of injected embryos (33/59; 56%) was higher than in the uninjected controls (6/22; 27%), but this difference may have been caused by unspecific side effects of the injection procedure. Across three independent replicates, a subset of the injected embryos reached NC14 but failed to cellularize and developed abnormally ([S1 Movie](#)). This phenotype is comparable to the phenotype of *zld* mutant *Drosophila* embryos ([S2 Movie](#)) [56] and suggests that Zld performs similar developmental roles in both species.

To assess *Cal-zld*'s role in shaping chromatin accessibility and zygotic transcription, we adapted our ATAC-seq protocol to isolate RNA from single-embryo ATAC-seq library preparations and used the RNA-seq data to quantify *Cal-zld* expression levels in Late NC14 *Cal-zld* RNAi embryos and uninjected control embryos ([S6 Fig](#) and see [Materials and methods](#)). We then compared gene expression (RNA-seq data) and chromatin accessibility (ATAC-seq data) between RNAi embryos with knockdown efficiencies greater than 75% ($n=4$; zldKD group), RNAi embryos in which no knockdown of *Cal-zld* was detected ($n=3$; zldnoKD group), and uninjected embryos that were otherwise treated like the RNAi embryos ($n=5$; alignment control group, AC) using DESeq2. To better discern effects of *Cal-zld* knockdown on the initiation of zygotic transcription, we distinguished genes with maternal expression (227), genes with maternal and zygotic expression (6658), and genes with only zygotic expression (691) ([Fig 3](#), [S4 Appendix](#)). As expected, only very few maternal genes were affected by knockdown of *Cal-zld*. In contrast, genes with maternal and zygotic expression were predominantly upregulated (1.7-fold in zldKD-vs-AC; 2.1-fold in zldKD-vs-zldnoKD) while genes with strictly zygotic expression were predominantly downregulated (3.3-fold in zldKD-vs-AC; 13.7-fold in zldKD-vs-zldnoKD). While the predominant upregulation of many genes with maternal and zygotic expression could reflect indirect posttranscriptional roles of *Cal-zld* in their regulation, the predominant downregulation of strictly zygotic genes in response to reduced *Cal-zld* levels is consistent with an important role of Cal-Zld in their transcriptional activation.

At the level of chromatin accessibility, ([Figs 4](#), [S7](#), [S5 Appendix](#)), 4,102 regions (12.8% of all 32,157 ATAC-seq peaks) were affected in the zldKD-vs-AC comparison, and 2,145 peaks (8.6% of all peaks) were affected in the zldKD-vs-zldnoKD comparison. In both cases, the majority of the peaks lost accessibility (1.4-fold difference in zldKD-vs-AC; 3.2-fold difference in zldKD-vs-zldnoKD; 5.7-fold differences between consistently downregulated and upregulated peaks). These observations should be compared to our negative control (zldnoKD-vs-AC), in which only 2.3% of all peaks exhibited differential accessibility and no enrichment of downregulated peaks was observed ([Fig 4](#)). Depletion of *Cal-zld* mRNA affected both dynamic peaks and nondynamic peaks. Dynamic peaks that lost accessibility in response to *Cal-zld* knockdown mostly belonged to groups that opened before late NC14 (e.g., groups 8, 15, 14, 12, 6, 10; [S8 Fig](#)). Taken together, these findings suggest that Cal-Zld plays an important role in opening chromatin while a limited additional role in closing select chromatin regions cannot be ruled out.

To assess if the changes in accessibility were a direct or indirect consequence of *Cal-zld* knockdown, we performed two types of motif analyses. First, we performed a differential motif enrichment analysis with MEME on *Cal-zld* sensitive peaks to identify motifs enriched centrally within 50 bp of the peak summit ([S9 Fig](#)). Relative to all nonsensitive peaks, the Zld motif was enriched in peaks that lost accessibility following *Cal-zld* knockdown (zldKD versus AC: 5-CAGGTA, $E = 1.9e^{-76}$; zldKD versus zldnoKD: 5-CAGGTA, $E = 1.9e^{-89}$), consistent with a direct role of Cal-Zld in nucleosome depletion and chromatin opening. Conversely, the Zld motif was not enriched in peaks that gained accessibility. We then extended the search for Cal-Zld motifs to the full width of each peak, using Zld motifs derived from our MEME analysis in *Clogmia* and experimental data in *Drosophila melanogaster* [116]. 81.4% of the 2,399 downregulated peaks in zldKD-vs-AC and 92.1% of the 1,632 downregulated peaks

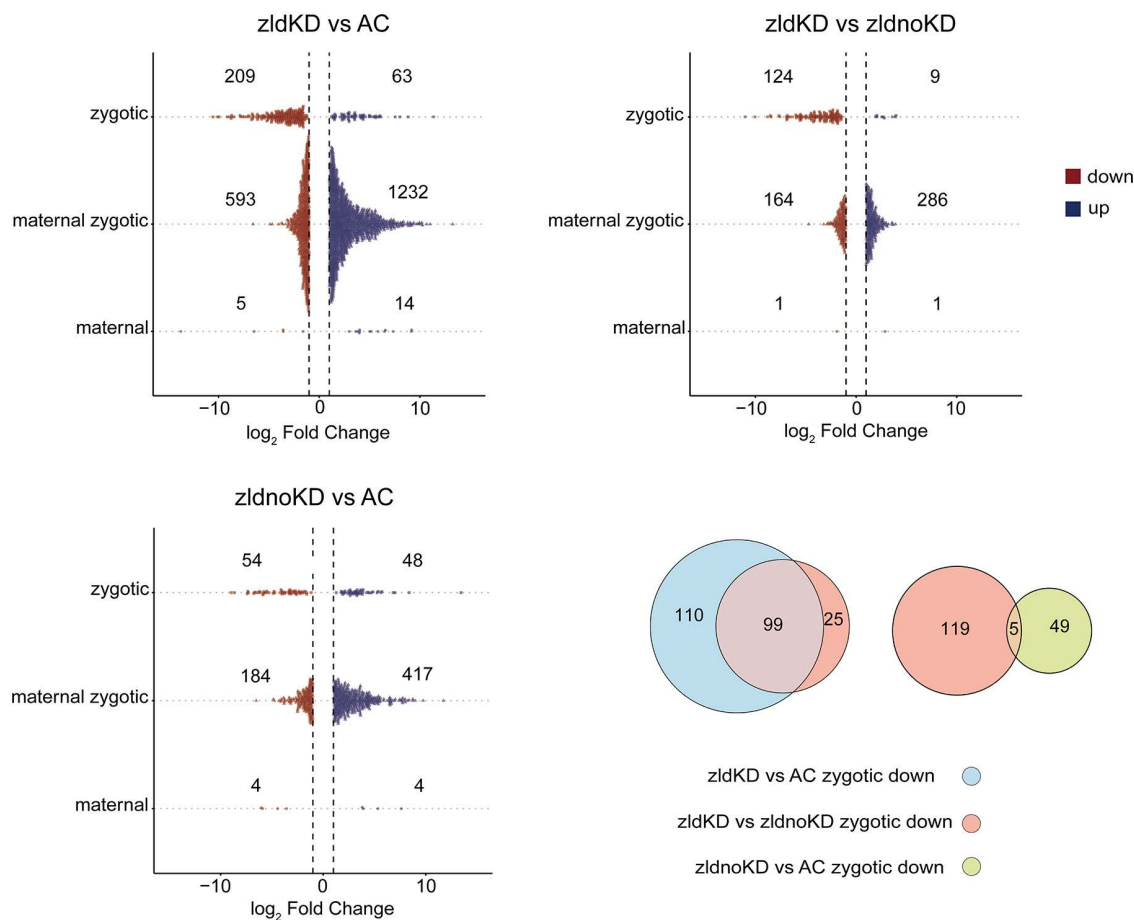


Fig 3. Cal-Zld promotes zygotic transcription. Individual panels show the number of downregulated and upregulated genes when comparing *Cal-zld* RNAi embryos with efficient *Cal-zld* knockdown to alignment controls (zldKD vs. AC) or to *Cal-zld* RNAi embryos without *Cal-zld* knockdown (zldKD vs. zldnoKD). The comparison of *Cal-zld* RNAi embryos without *Cal-zld* knockdown to alignment controls (zldnoKD vs. AC) serves as a negative control. On the y-axis, genes are grouped based on maternal, maternal and zygotic, or zygotic expression (see [Materials and methods](#)). Log₂ fold change is shown on the x-axis. Only genes with an adjusted *p*-value ≤ 0.05 and $|\log_2 \text{fold change}| \geq 1$ are shown. Venn diagrams show overlap of downregulated zygotic genes between zldKD vs. zldnoKD (pink) and zldKD vs. AC (blue) or zldnoKD vs. AC (green). Underlying data is available in [S2 Data](#) and NCBI (bioproject: PRJNA1198932).

<https://doi.org/10.1371/journal.pbio.3003896.g003>

in zldKD-vs-zldnoKD contained at least one Zld motif compared to ~60% of the respective upregulated peaks or the Cal-Zld-insensitive peaks. This difference was slightly more pronounced when restricting the comparison to the 1,091 consistently downregulated peaks (93.3% with at least one Zld motif; [S7A Fig](#)) and the 191 consistently upregulated peaks (47.1% with at least one Zld motif; [S7B Fig](#)). These findings suggest that Cal-Zld plays a direct role in opening chromatin. In support of this hypothesis, we recall that both motif affinity and occupancy positively correlate with Zld's pioneering activity in *Drosophila* [54,109]. Conversely, the closure of many chromatin regions that we observed during the final cleavage cycles in wild-type *Clogmia* embryos may generally occur without direct Cal-Zld input, although a direct role in closing some of these regions cannot be ruled out since 90 of the 191 consistently upregulated peaks in the zldKD-vs-AC and zldKD-vs-zldnoKD comparisons contain at least one Zld motif.

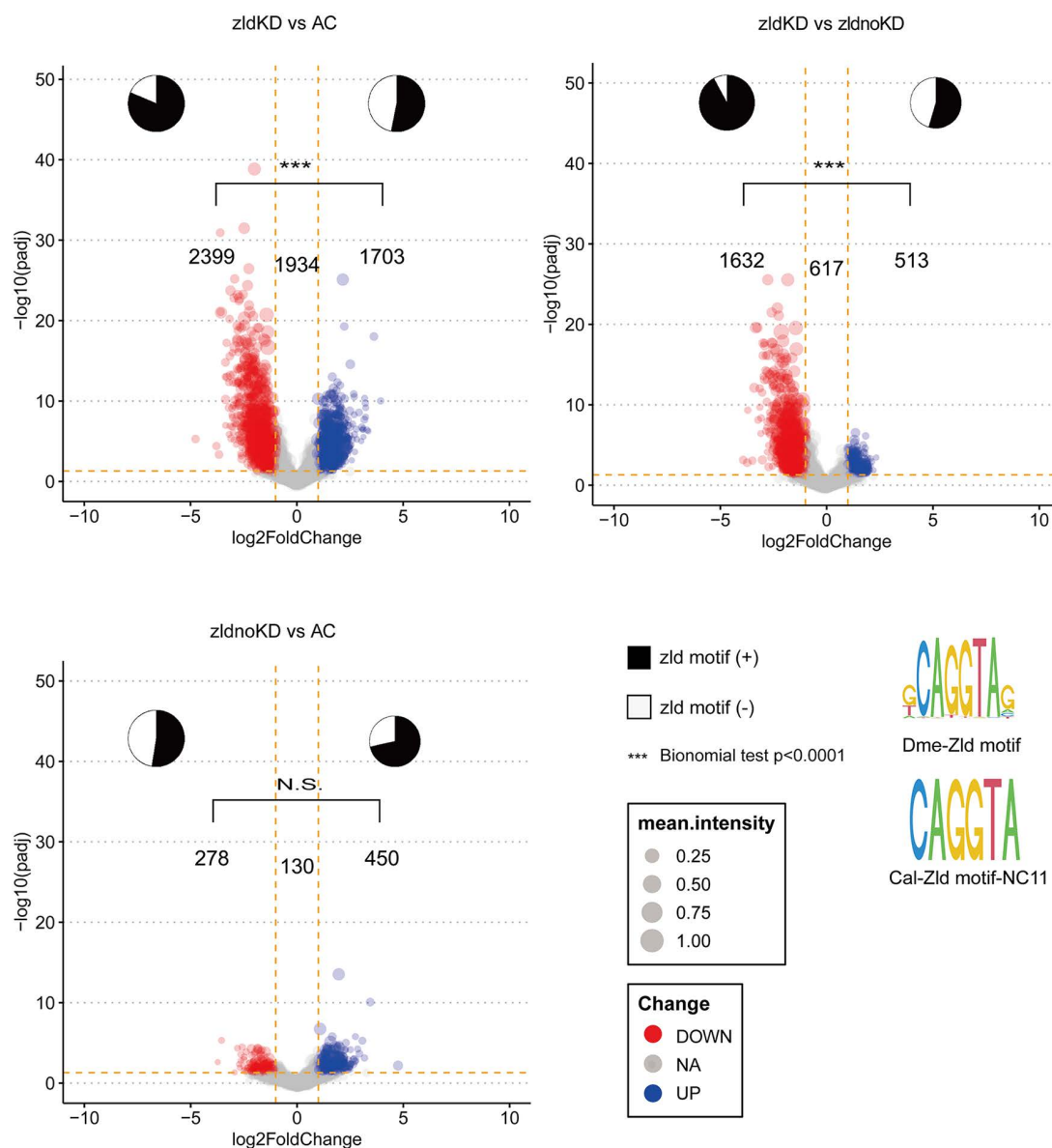


Fig 4. Cal-Zld promotes chromatin accessibility. Volcano plots of differentially accessible chromatin regions (ATAC-seq peaks) between *Cal-zld* RNAi embryos with efficient *Cal-zld* knockdown and alignment controls (zldKD vs. AC) or *Cal-zld* RNAi embryos without *Cal-zld* knockdown to alignment controls (zldnoKD vs. AC) serves as a negative control. The comparison of *Cal-zld* RNAi embryos without *Cal-zld* knockdown to alignment controls (zldnoKD vs. AC) serves as a negative control. Significant change of adjusted p -value ≤ 0.05 and $|\log_2$ fold change ≥ 1 are highlighted in red ($\log_2$ fold change ≤ -1) or blue ($\log_2$ fold change ≥ 1). The inserted pie charts show the proportion of peaks with at least one Zld motif (black) as reported in FlyFactorSurvey or inferred from the discriminative MEME analysis on NC11 peaks in wild-type *Clogmia* embryos (cf. position weight matrix logos). Underlying data is available in [S3 Data](#) and NCBI (bio-project: PRJNA1198932).

<https://doi.org/10.1371/journal.pbio.3003896.g004>

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 [20,22,30,80,117–120]. Bcd breaks axial symmetry by targeting these genes in a

concentration-dependent manner throughout the blastoderm stage, including NC14. In contrast, maternal *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 [3]. Therefore, we focused our search of Cal-Opa targets on TF genes that are expressed before NC14 and asked whether maternal *Cal-opa* activity affects their expression and chromatin accessibility during NC12 and NC13.

To assess the impact of maternal Cal-Opa activity on gene expression and chromatin accessibility, we performed RNA-seq and ATAC-seq on single *Cal-opa* RNAi embryos (6 replicates for NC12, 8 replicates for NC13) and selected four NC12 embryos and five NC13 embryos with knockdown efficiencies greater than 75% (NC12_opaKD, NC13_opaKD; [S10 Fig](#) and [Materials and methods](#)). At NC13, we also identified two RNAi embryos without *Cal-opa* knockdown (NC13_opanoKD) and used them as a negative control group. In parallel, we processed stage-matched embryos that were prepared for injection under oil but not injected (alignment controls, AC; 4 replicates for each developmental stage), and embryos that were injected with dsRNA of the extraneous gene *DsRed* (injection controls, IC; 3 replicates for each stage). We then used DESeq2 to compare gene expression (RNA-seq peaks; [Fig 5](#), [S6 Appendix](#)) and chromatin accessibility (ATAC-seq peaks; [Fig 6](#), [S7 Appendix](#)) at NC12 and NC13 between embryos of the opaKD groups and their respective control groups (AC and IC for NC12; AC, IC, and opanoKD for NC13). At NC12, opaKD-vs-AC identified 79 downregulated genes (including 11 TF genes) and 5 upregulated genes (no TF genes) while opaKD-vs-IC identified only 22 downregulated genes (including 5 TF genes) and no upregulated genes. An opanoKD group was not available for this stage. *Cal-opa* (positive control), *Clogmia*'s three *sloppy-paired* homologs (*Cal-slp1*, *Cal-slp 2*, *Cal-slp3*), and *homeobrain* (*Cal-hbn*) were the only consistently downregulated TF genes at NC12 ([Fig 5](#)). At NC13, the opaKD-vs-AC comparison yielded 561 downregulated genes (including 67 TF genes) and 486 upregulated genes (including 49 TF genes). However, far fewer down- and upregulated genes were identified when the opaKD group was compared to the injected control groups ([Fig 5](#)). In opaKD-vs-IC, 27 genes were downregulated (including 5 TF genes: *Cal-hbn*, *Cal-slp 1*, *Cal-slp2*, *Cal-hb1*, *Cal-hb2*) and 17 were upregulated (including 1 TF gene). Downregulation of *Cal-opa* in the IC control group was not statistically significant ($\log_{10}$ adjusted *p*-value of 0.1169), falling slightly above our significance cutoff: $\log_{10}$ adjusted *p*-value of < 0.05 . This observation likely reflects the low detection of *Cal-opa* transcript in one of the three embryos in the injection control group ([S10 Fig](#)). In opaKD-vs-opanoKD, only 13 genes were downregulated (including 6 TF genes: *Cal-hbn*, *Cal-slp1*, *Cal-slp2*, *Cal-hb1*, *Cal-hb2*, and *Cal-opa*) and 3 genes were upregulated (0 TF genes). These results should be compared to our negative control for NC13 (opanoKD-vs-AC), which only yielded a random set of 3 slightly downregulated genes (0 TF genes) and no upregulated genes ([S11A Fig](#)). Taken together, our results suggest that maternal *Cal-opa* promotes the activation of *Cal-slp1*, *Cal-slp 2*, *Cal-slp3*, and *Cal-hbn* at NC12 and of *Cal-slp1*, *Cal-slp2*, *Cal-hbn*, *Cal-hb1*, and *Cal-hb2* at NC13. Of these genes, only the loci of *Cal-slp1*, *Cal-slp 2*, and *Cal-hbn* were associated with ATAC-seq peaks that spanned Opa motifs ([S7 Appendix](#)).

To assess the effects of *Cal-opa* knockdown on chromatin accessibility we performed DESeq2 analyses on the ATAC-seq peaks ([Fig 6](#), [S7 Appendix](#)). At NC12, the opaKD-vs-AC comparison identified 79 downregulated and 7 upregulated peaks. Sixty-three of the 79 peaks downregulated peaks but none of the 7 upregulated peaks contained at least one Opa motif. The opaKD-vs-IC comparison only yielded 35 downregulated peaks of which 30 contained at least one Opa motif. The 32 consistently downregulated peaks were associated with the loci of *Cal-slp1*, *Cal-hbn*, *Cal-opa*, and various non-TF genes ([Figs 7](#), [8](#), [S7 Appendix](#)). One additional ATAC-seq peak associated with *Cal-slp2* narrowly missed our significance threshold in the opaKD-vs-IC comparison (adjusted *p*-value 0.062).

At NC13, opaKD-vs-AC yielded 250 downregulated peaks and 42 upregulated peaks, opaKD-vs-IC identified 11 downregulated peaks and 5 upregulated peaks, and opaKD-vs-opanoKD recovered only 5 downregulated peaks and 5 upregulated peaks without Opa motifs ([Fig 6](#), [S7 Appendix](#)). Of the consistently downregulated peaks, two were associated with *Cal-hbn* ([Fig 7](#)), and two were associated with *Cal-slp1* ([Fig 8](#)), all covering Opa motifs. The fifth consistently downregulated peak was associated with an uncharacterized gene without homology to known genes (evm.TU.scaffold_7.1426)

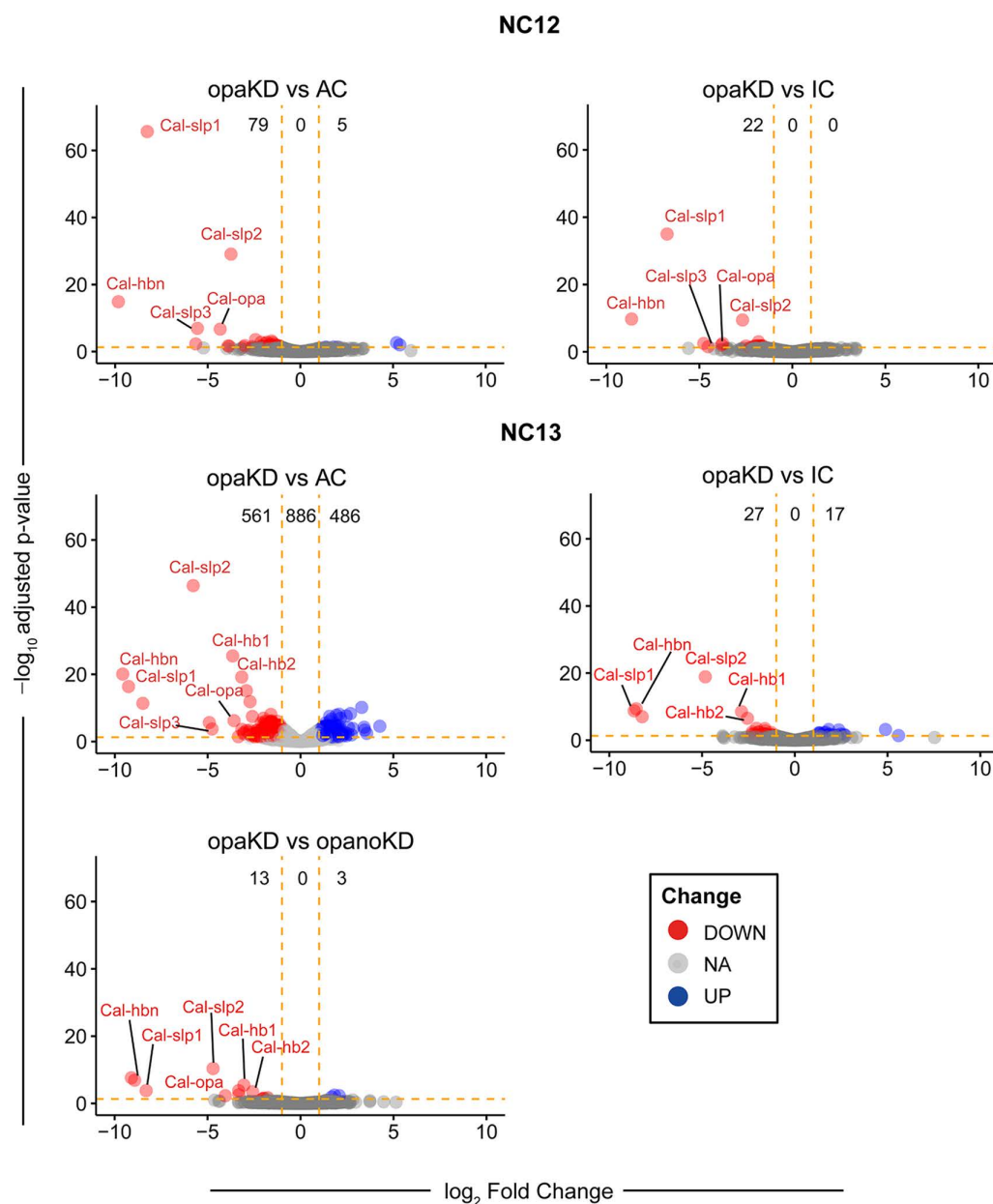


Fig 5. Maternal *Cal-opa* activates a small set of transcription factor genes at NC12 and NC13. Volcano plots of differentially expressed genes between embryo groups with reduced *Cal-opa* expression (opaKD) and control groups (AC and IC for NC12; AC, IC, and opoKD for NC13). See [S11A Fig](#) for a negative control (opoKD vs. AC at NC13). Significant changes of adjusted p -value ≤ 0.05 and $|\log_2$ fold change| ≥ 1 are highlighted in red ($\log_2$ fold change ≤ -1) and blue ($\log_2$ fold change ≥ 1). Consistently affected TF genes are labeled. Underlying data is available in [S4 Data](#) and NCBI (bioproject: PRJNA1198932).

<https://doi.org/10.1371/journal.pbio.3003896.g005>

and did not cover an Opa motif. Significant downregulation of peaks associated with *Cal-hb1* and *Cal-hb2* was only observed in comparisons with AC and IC controls at NC13, and no Opa motifs were detected in these peaks ([S12 Fig](#)). In our negative control (opoKD-vs-AC), none of these ATAC-seq peaks were recovered ([S11B Fig](#), [S7 Appendix](#)).

Given that maternal *Cal-opa* is required for zygotic transcription *homeobrain* and *sloppy-paired* homologs, we examined the expression of these genes in more detail. Maximum-likelihood trees with dipteran and nondipteran

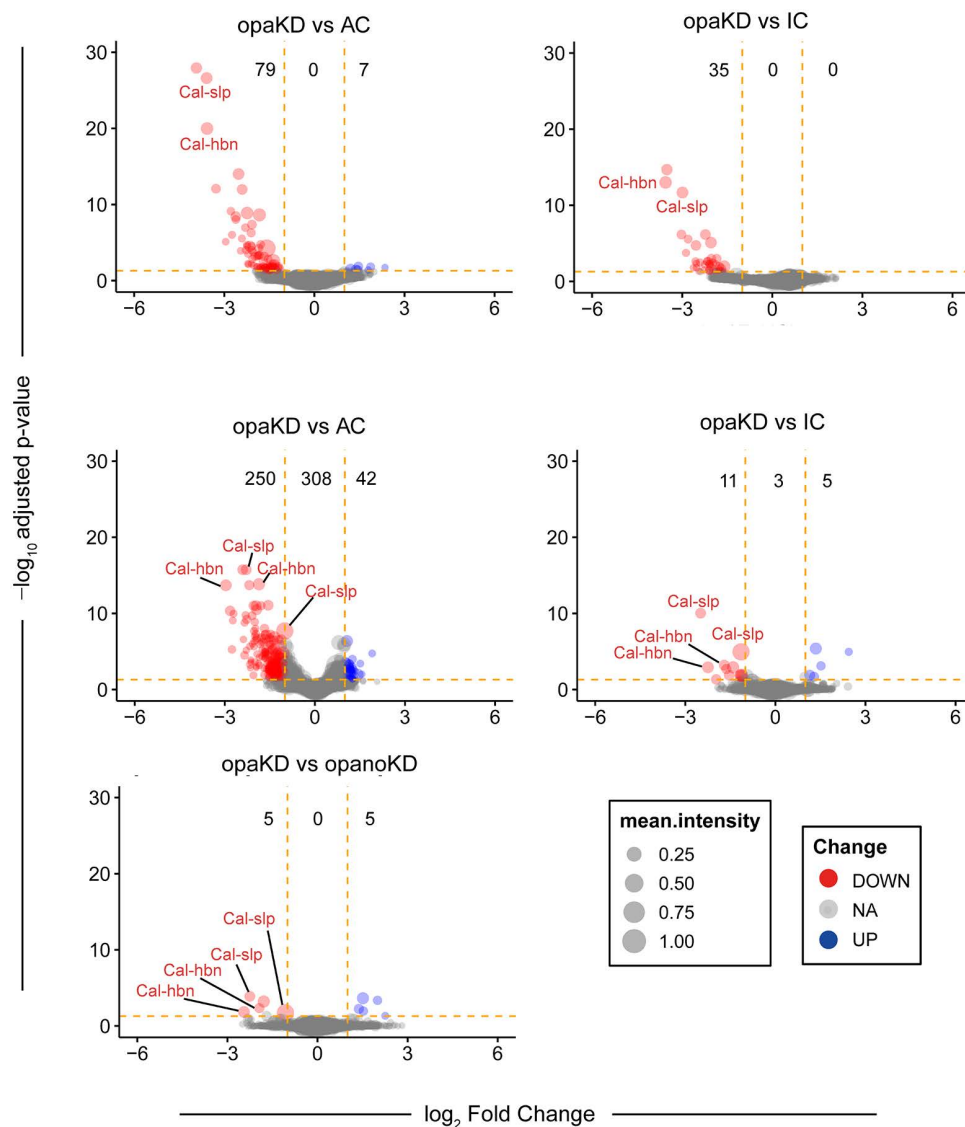


Fig 6. *Cal-Op* is required for establishing a selective set of chromatin accessibility peaks. Volcano plots of differentially accessible chromatin regions (ATAC-seq peaks) between *Cal-opa* RNAi embryos with reduced *Cal-opa* expression (*opaKD*) and control groups (AC and IC for NC12; AC, IC, and *opanoKD* for NC13). See S11B Fig for negative control (*opanoKD* vs. AC). Significant changes of adjusted *p*-value ≤ 0.05 and $|\log_2$ fold change ≥ 1 are highlighted in red ($\log_2$ fold change ≤ -1) and blue ($\log_2$ fold change ≥ 1). Underlying data is available in S5 Data and NCBI (bioproject: PRJNA1198932).

<https://doi.org/10.1371/journal.pbio.3003896.g006>

sloppy-paired homologs generated with IQ-TREE v.2.3.6. [121] and ModelFinder [122] suggest that *Cal-slp1*, *Cal-slp2*, and *Cal-slp3* resulted from recent gene duplications in the *Clogmia* lineage (S13, S14 Figs). Zygotic transcription of *Clogmia*'s *sloppy-paired* and *homeobrain* homologs was detected during and after NC11 and was confined to the anterior half of the blastoderm (Fig 9). *Cal-hbn* was expressed in the anterior-most portion of the blastoderm, complementary to the overlapping expression domains of *Cal-slp1*, *Cal-slp2*, and *Cal-slp3* (which was only expressed at low levels prior to NC14). In the case of *Cal-slp2*, we also detected high levels of maternal transcript, previously found to be distributed in a shallow anterior-to-posterior gradient [3]. The early timing and spatial

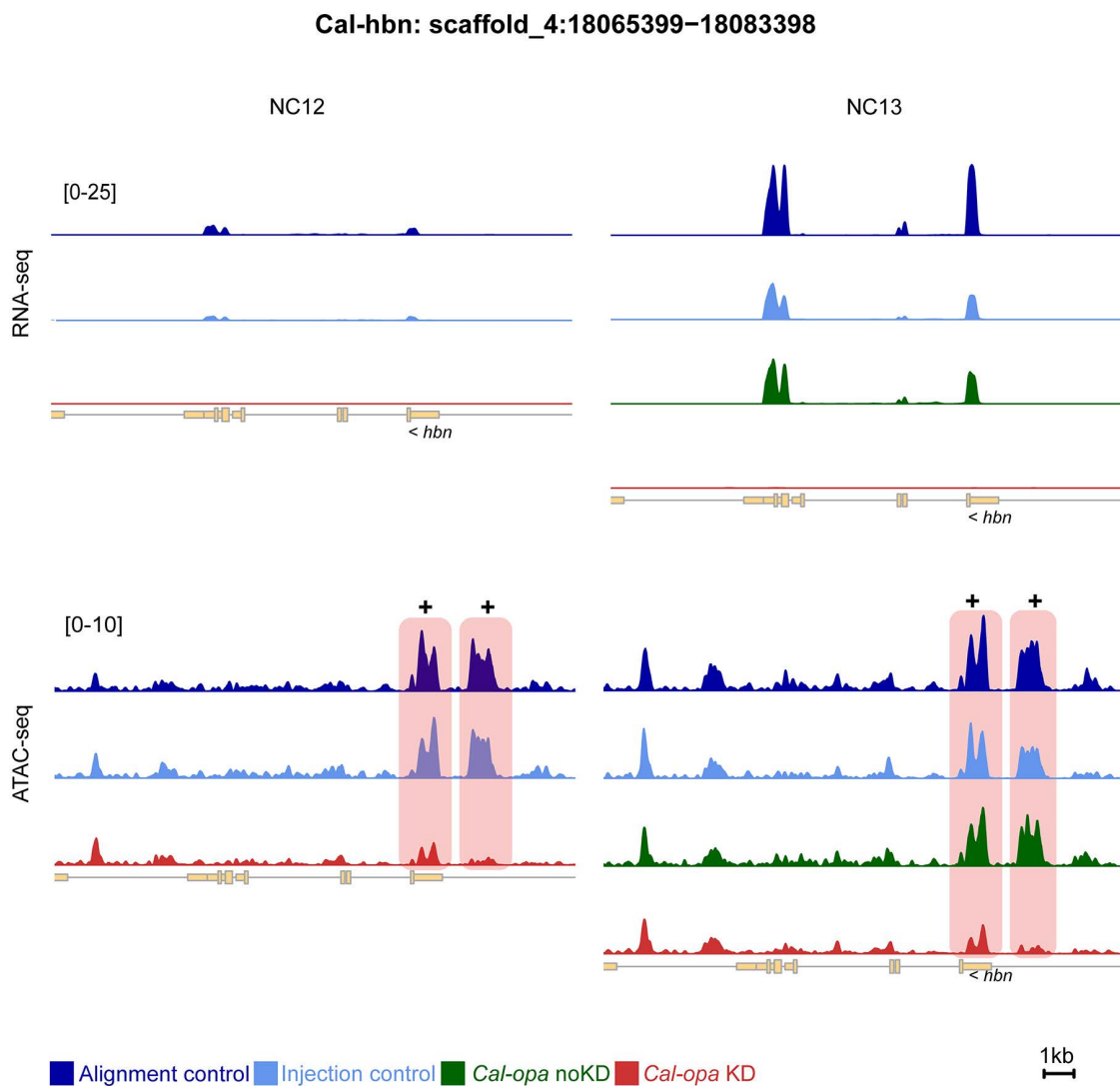


Fig 7. Expression and chromatin accessibility at the *Cal-hbn* locus at NC12 and NC13. RNA-seq and ATAC-seq tracks (CPM) are shown for NC12 and NC13 with alignment control in blue, injection control in light blue, *Cal-opa* RNAi without KD in green, and *Cal-opa* RNAi with KD in red. Gene models are shown in pale yellow. ATAC-seq peaks with significant changes in accessibility are highlighted in pink and are marked with a plus sign if they contain at least one Opa motif. The scale of the y-axis is adjusted for each data type but kept constant between stages and is indicated with numbers in square brackets: [0-25] for RNA-seq data; [0-10] for ATAC-seq data. Underlying data is available in NCBI (bioproject: PRJNA1198932).

<https://doi.org/10.1371/journal.pbio.3003896.g007>

expression of *Clogmia*'s *homeobrain* and *sloppy-paired* homologs suggest that these genes likely play an important zygotic role in the initiation of anterior pattern formation.

Discussion

Dynamic chromatin states in the embryo of *Clogmia albipunctata*

In this study, we have established the resources for dissecting developmental gene networks of early *Clogmia* embryos by providing an annotated chromosome-level genome sequence for our inbred laboratory strain and by describing chromatin

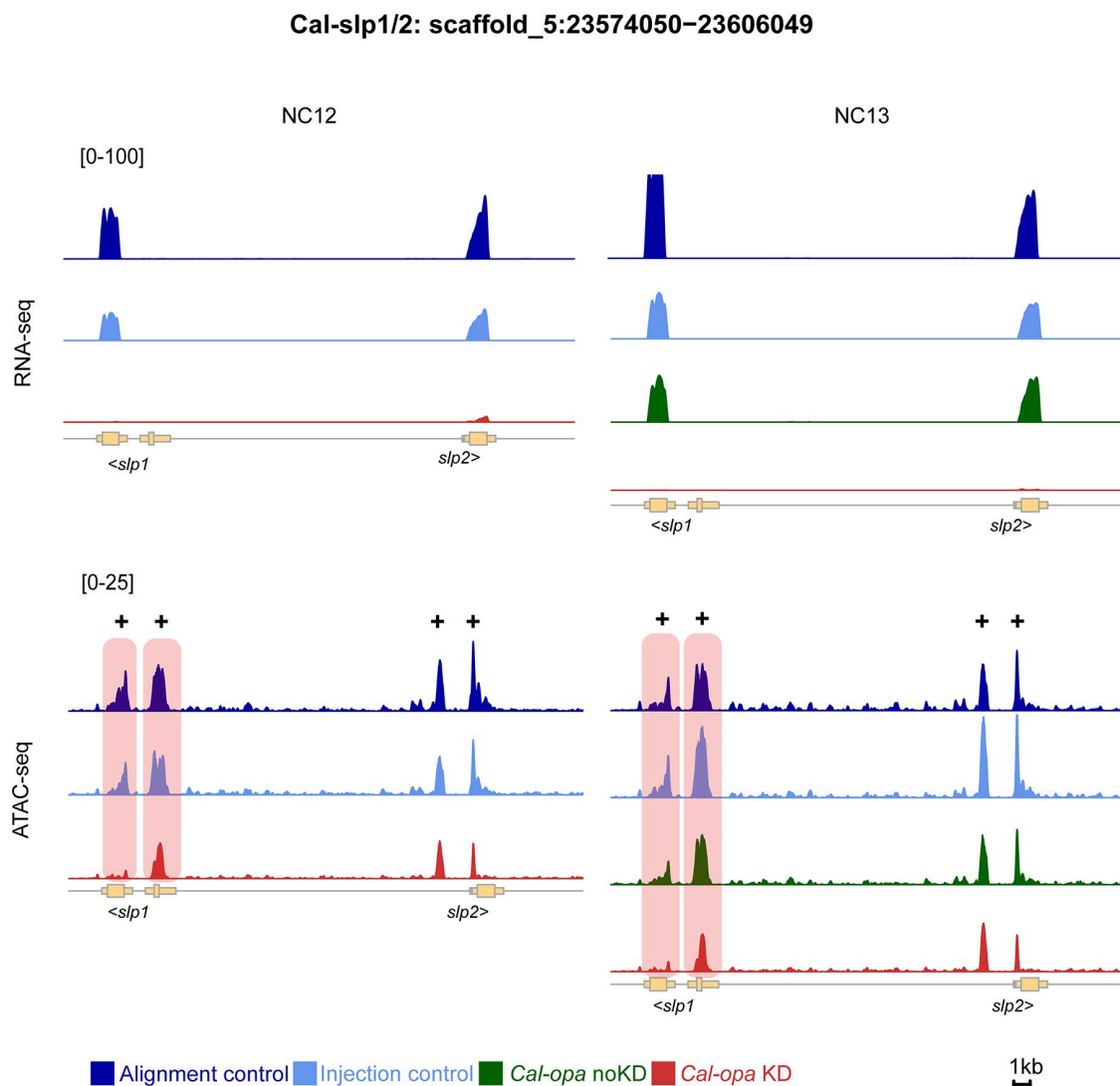


Fig 8. Expression and chromatin accessibility at the *Cal-slp1* and *Cal-slp2* locus at NC12 and NC13. RNA-seq and ATAC-seq tracks are shown for NC12 and NC13 with alignment control in blue, injection control in light blue, *Cal-opa* RNAi without KD in green, and *Cal-opa* RNAi with KD in red. Gene models are shown in pale yellow. ATAC-seq peaks with significant changes in accessibility are highlighted in pink and are marked with a plus sign if they contain at least one Opa motif. The scale of the y-axis (CPM) is adjusted for each data type but kept constant between stages and is indicated with numbers in square brackets: [0-100] for RNA-seq data; [0-25] for ATAC-seq data. Underlying data is available in NCBI (bioproject: PRJNA1198932).

<https://doi.org/10.1371/journal.pbio.3003896.g008>

accessibility of precisely staged consecutive blastoderm stages. We have used these resources to examine how the activities of *zelda* and maternal *odd-paired* shape chromatin accessibility and gene expression in *Clogmia*. We found that, as in *Drosophila*, the period spanning ZGA is accompanied by large-scale changes in chromatin accessibility. However, there are notable differences between *Clogmia* and *Drosophila* in the early patterns of establishment and maintenance of accessible states before NC14. In *Clogmia*, the chromatin landscape is highly dynamic in that gains as well as losses occur during the nuclear cleavage cycles of the blastoderm stage. In *Drosophila*, only gains in accessibility are observed until widespread ZGA and lengthening of the cell cycle at NC14 [74,76,77,79,123]. This difference could be a consequence of cell cycle timing. In *Clogmia*, NC11 through NC13 last ~3x longer than in *Drosophila* [83], in which NC11 only

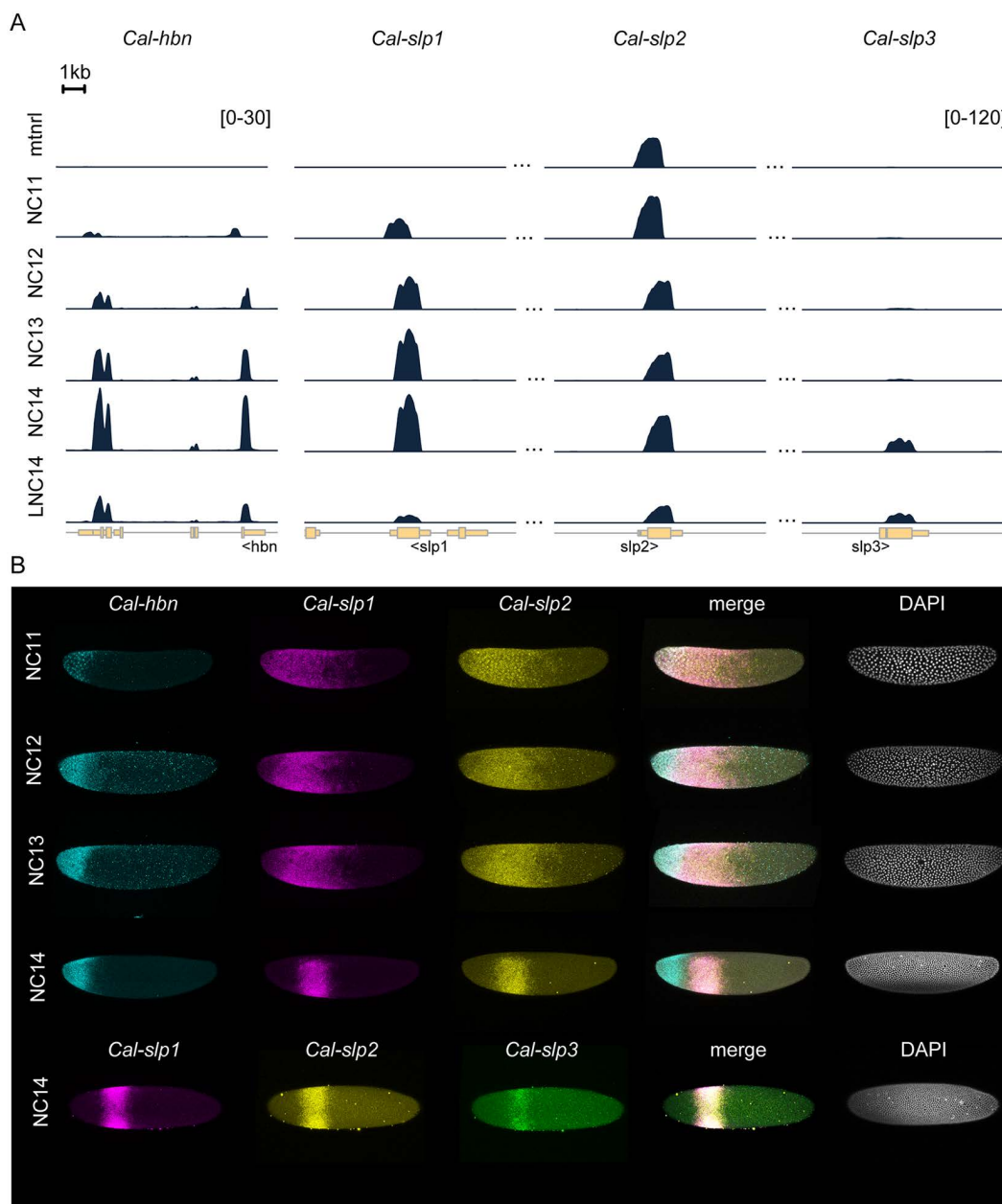


Fig 9. Expression of *homeobrain* and *sloppy-paired* homologs in *Clogmia*. **A)** Expression levels (CPM) of *Cal-hbn*, *Cal-slp1*, *Cal-slp2*, and *Cal-slp3* at 6 consecutive stages, including NC2/3 (maternal, ~45-min old embryos), NC11, NC12, NC13, NC14 and Late NC14. Gene models are shown in pale yellow. Synteny of the three sloppy-paired loci is indicated by three dots. The y-axis is the same for all tracks at each locus. *Cal-hbn* RNA-seq: 0-30 CPM. *Cal-slp1-3* RNA-seq: 0-120 CPM. **B)** Whole-mount fluorescent in situ hybridizations at consecutive blastoderm stages. Embryos are shown with anterior left and dorsal up. Underlying data is available in NCBI (bioproject: PRJNA1198932).

<https://doi.org/10.1371/journal.pbio.3003896.g009>

lasts ~10 min, NC12~12 min, and NC13~21 min [82]. The much shorter cell cycle times of *Drosophila* present significant limitations to establish and maintain chromatin states necessary for transcriptional activity [79]. Assuming that kinetic rates for nuclear import, DNA replication, and DNA binding of TFs is similar between *Clogmia* and *Drosophila*, the significantly longer syncytial interphase times in *Clogmia* could alleviate constraints on the initiation of zygotic transcription. However,

the timing of AP patterning between the two species appears comparable [84,117], suggesting that the timing of ZGA might also be conserved. Alternatively, the longer syncytial interphase times in *Clogmia* could alleviate constraints on maintaining or silencing active chromatin states. In *Drosophila*, Polycomb group (PcG) proteins begin to establish heritable, transcriptionally silenced chromatin states during NC14 [124], but in *Clogmia*, the longer nuclear cleavage cycles may provide time for the maturation of repressive histone marks at accessible chromatin regions before NC14.

Role of Zld in regulating chromatin accessibility during *Clogmia*'s nuclear cleavage cycles

We found that reduced *Cal-zld* expression results in widespread losses of chromatin accessibility, consistent with a conserved role for Zld to pioneer chromatin accessibility within the context of ZGA. In *Drosophila*, Zld almost exclusively drives gains in chromatin accessibility during nuclear cleavage cycles [71,76,102,125,126]. However, two recent studies found that during ZGA, Zld activity directs how repressive chromatin marks are deposited. Loss of *zld* results in context-specific losses and gains of H3K27me3 within specific PcG domains [124]. Zld also enables histone acetylation at promoters and enhancers through its interaction with Chip, *Drosophila*'s homolog of LIM-domain-binding protein 1 (Ldb1), which in turn recruits the histone acetyltransferase CBP [127]. The longer nuclear cleavage cycles in *Clogmia* might enable similar roles of Cal-Zld before NC14. In this case, *Clogmia* and *Drosophila* may differ in the timing of when specific histone marks are deposited and affected by Zld depletion. Differences in the activities of Zld and Cal-Zld could also reflect differences in their protein structures or isoforms. For example, Cal-Zld contains an additional zinc finger that has been eroded in *Drosophila* (S4 Fig) [110]. Additionally, stage-specific isoforms of Zld have been described in *Drosophila*, including a variant lacking the C-terminal three zinc fingers [128–131]. This isoform functions as a dominant negative form in cell culture although flies in which the expression of this isoform is suppressed are viable and fertile [130]. Therefore, the molecular underpinnings of the functional readouts of Zld and Cal-Zld could differ and involve both direct and indirect mechanisms.

Functional comparison of the anterior determinants in *Clogmia* and *Drosophila*

Bcd contributes to the activation of many early segmentation genes; however, 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. Bcd target enhancers have been classified based on how their accessibility depends on Bcd and Zld [71]. The accessibility of Bcd target enhancers that drive expression in the posterior of the embryo is established independent of Bcd (e.g., *Krüppel* CD1, *knirps* posterior, *giant* posterior). However, enhancers that open in a Bcd-dependent manner are associated with genes expressed in the anterior of the embryo (e.g., *orthodenticle*, *empty spiracles*, *buttonhead*, *giant*, *knirps*, *hunchback*, *sloppy-paired 1*, *fd19B*, *even-skipped*, *paired*, *cap'n'collar*, *huckebein*, and *homeobrain*). Moreover, many correspond to enhancers that drive reporter gene expression in the anterior embryo including *giant* anterior (–6), *giant* anterior (–10), *knirps* anterior, *orthodenticle*, *buttonhead*, *even-skipped* stripe 1/5, *hunchback* P2, *hunchback* shadow enhancers, *even-skipped* stripe 2, an early *paired* enhancer, *cap'n'collar*, and *huckebein* (reviewed in [71]). The majority of Bcd-dependent anterior expression domains first appear during NC13, with the noteworthy exception the P2 transcript of *hunchback*. The promoter of this transcript is in an open conformation as early NC10 [79] and drives Bcd-dependent expression throughout the anterior half starting at NC11 [45,66,132], thereby boosting the maternal Hb gradient which forms independent of Bcd [133].

In *Clogmia*, neither the expression nor chromatin accessibility at the *Cal-hb1* and *Cal-hb2* loci was reduced in NC12 opaKD embryos (S12 Fig) and maternal *hunchback* activity does not contribute to embryo polarity in this species, given that knockdown of maternal *Cal-opa* transcripts results in symmetrical double abdomens [3]. Additionally, homologs of several Bcd target genes were not expressed at NC12 and NC13 (e.g., *Cal-gt*, *Cal-kni/knr1-2*, *Cal-ems1-4*, and *Cal-hkb*) or their expression levels at these stages were not affected by the knockdown of maternal *Cal-opa* activity (e.g., *Cal-cnc*, *Cal-prd*, and *Cal-btd*). However, maternal Cal-Opa was required for the zygotic transcription of *sloppy-paired* and

homeobrain homologs in complementary, evolutionarily conserved domains across the anterior half of the embryo. In *Drosophila*, these genes are regulated by Bcd-bound enhancers whose accessibility is sensitive to the concentration of Bcd [134,135], even though neither of them has received much attention as Bcd target genes. Taken together, these findings suggest that the anterior determinant of *Clogmia* targets a small subset of the genes known as direct targets of Bcd in *Drosophila*.

Mechanistically, Cal-Opa could prime accessibility at target loci through a canonical pioneering mechanism which requires binding of nucleosome occupied DNA and promoting nucleosome instability [47,48,56,60,136,137]. A vertebrate homolog of Opa, Zic3, was found to bind nucleosomal DNA in vivo, suggesting that Cal-Opa may pioneer chromatin directly [138]. Alternatively, Cal-Opa binding may compete with nucleosomes to bind unoccupied DNA during each nuclear cleavage cycle [139], similar to Bcd [54,71,78].

Role of *homeobrain* and *sloppy-paired* in breaking axial symmetry in insects

It is conceivable that *homeobrain* nor *sloppy-paired* homologs serve a more broadly conserved role in the initiation of anterior pattern formation. In *Drosophila*, *homeobrain* mutant embryos lack the labrum and anterior portions of the brain [140–143], indicating that *homeobrain* is essential for specifying the most anterior portion of the embryo. A more extreme *homeobrain* loss-of-function phenotype has been observed in the beetle *Tribolium castaneum*, where this gene plays an important zygotic role in establishing the embryo's AP polarity [144]. In *Tribolium*, knockdown of *homeobrain* resulted in loss of the entire head and—in extreme cases—duplications of the tail end. The duplicated abdomen was defective, and thoracic structures persisted in these embryos, but complete double abdomens could be obtained by simultaneously suppressing the expression *Tc-zen1* (like *bicoid* a *Hox3* derivative [35,95], thereby preventing the anterior repression of *caudal*, an important gene for posterior patterning [145]. Additionally, the *Tribolium* ortholog of *odd-paired*, which is expressed maternally and required for head development (including the labrum, antennae, and head capsule) [146], could potentially complement the strictly zygotic roles of *Tc-zen* and *Tc-hbn* in breaking symmetry along the AP axis of the *Tribolium* embryo.

The *sloppy-paired* gene, a member of the *forkhead* domain gene family [147], is best-known for its role as a pair-rule segmentation gene [22,81], but it is initially expressed in a head domain of the syncytial *Drosophila* embryo, along with its paralogs *sloppy-paired 2* and *fd19B* [135,148–153]. The *sloppy-paired* paralogs, which are common across insects (S13, S14 Figs), may have obscured a fundamental role of *sloppy-paired* activity in initiating anterior pattern formation in *Drosophila*. This also applies to the two *sloppy-paired* homologs of *Tribolium*, in which only the gnathocephalon (mandibular, maxillary, and labial segment) and even-numbered trunk segments are lost when *Tca-slp1* is knocked down [154,155]. We therefore propose that along with *homeobrain*, *sloppy-paired* may play a more widely conserved role in establishing the AP axis in insect embryos.

Limitations of the study

It is conceivable that *homeobrain* and *sloppy-paired* activities are sufficient for breaking symmetry along the AP axis of the *Clogmia* embryo. To test this hypothesis, it might be necessary to completely suppress the activities of four genes, including three with early zygotic expression (*Cal-hbn*, *Cal-slp1*, and to a lesser degree *Cal-slp3*) and one with anterior-enriched maternal transcript and zygotic expression (*Cal-slp2*). This is particularly challenging in *Clogmia* because this species did not tolerate more than one microinjection per embryo in our hands. Ongoing research in our laboratory seeks to overcome this technical issue by developing more effective knockout and knockdown reagents. Alternatively, maternal Cal-Opa could control the zygotic expression of additional genes. We did not observe other TF genes that were consistently downregulated in embryos with reduced maternal *Cal-opa* transcript, but this could be due to incomplete knockdown of the maternal *Cal-opa* transcript or limited statistical power of our opaKD-vs-IC and/or opaKD-vs-opanoKD comparisons or a combination of both. To further explore this possibility and to formally prove a direct interaction of maternal Cal-Opa with

the identified target genes, we are developing protocols and reagents for Cal-Opa chromatin immunoprecipitation experiments. In the present study, the identification of Clogmia's anterior determinant target genes rests on Opa-dependent chromatin accessibility along with the presence of Opa-binding motifs, Opa-dependent gene expression in the anterior blastoderm, and the early timing of transcriptional activation.

Materials and methods

Clogmia culture and embryo collection

The Clogmia culture was maintained in 100 mm × 15 mm 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 100 mm × 15 mm Petri Dishes (Corning CM: 101425,55) containing 15 ml 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 min for NC2/3, 261 min for NC11, 292 min for NC12, 332 min for NC13, 401 min for NC14, and 465 min for Late NC14.

Embryo fixation

Embryos were dechorionated using a 50% dilution of commercial bleach (8.25% sodium hypochloride) for 60 s. 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 s. 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 –80 °C). 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.

Microinjections of embryos and RNAi efficiency

Embryo injection was done as previously described [3]. We note here that RNAi efficiency in Clogmia embryos varies [3]. 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. Embryos were injected within the first hour of development (NC2/3). To test *Cal-zld* dsRNA, we quantified knockdown of the *Cal-zld* transcript in individual embryos by RT-qPCR and performed fold change calculations of *Cal-zld* mRNA levels with the average level in uninjected control embryos as reference (S3 Table). In one quarter of these *Cal-zld* RNAi embryos (5/20), *Cal-zld* mRNA levels were reduced by ~30%–90%. For *Cal-opa* RNAi, we used dsRNA of the isoform-specific first exon of the maternal *Cal-opa* transcript and confirmed that it can induce the double abdomen phenotype. We also quantified knockdown efficiencies of the *Cal-opa* transcript in individual embryos by RT-qPCR and performed fold change calculations of *Cal-opa* mRNA levels with the average level in uninjected control embryos as reference (S4 Table). This analysis suggests that *Cal-opa* RNAi can nearly abolish the *Cal-opa* transcript.

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 NaH₂PO₄ 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'-TAATACGACTCACTATAGGGAGAAAGTCCCGCAATTGATACAGC

5'-TAATACGACTCACTATAGGGAGAAAGGATGTTGGTGGACACTCC

5'-TAATACGACTCACTATAGGGAGACGGAATGGTGTGTGAAACAG

5'-TAATACGACTCACTATAGGGAGATTCCGAGGGGTTATTGTCCTG

dsRed (273 bp):

5'-CAGAGATGCATAATACGACTCACTATAGATGCAGAAGAAGACTATGG

5'-CAGAGATGCATAATACGACTCACTATAGCTACAGGAACAGGTGGTG

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'- AGGTCGAGGGGCTTGATTT

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-fluorophores were used as follows: B1-647 for *Cal-slp1*, B2-546 for *Cal-hbn* and *Cal-slp3*, and B3-488 for *Cal-slp2*, and *Cal-otd1*. Buffers (probe hybridization, probe wash, and amplification buffers) throughout the protocol were purchased from Molecular Instruments. The HCR in situ hybridizations were performed as described [156,157]. Briefly, fixed embryos were rehydrated and washed 3 times in PBS-Tween20 (1x PBS, 0.1% Tween-20). Embryos were permeabilized for 30 min 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), and 200 mg BSA Fraction V (Sigma A5611-5G). Embryos were then pre-hybridized in probe hybridization buffer for 30 min in a 37 °C water bath. Once pre-hybridization was complete, fresh hybridization solution containing gene specific probes was added, and the embryos were left to incubate overnight in a 37 °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 min 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.

Image acquisition and processing

Confocal imaging was performed on the Zeiss LSM 900 using a Plan-Apochromat 20x/0.8 M27 objective. Pixel-density was 1,024 x 1,024 and physical pixel size was X = 0.46980808984799 μm and Y = 0.46980808984799 μm . Excitation wavelengths were 561 nm for Alexa Fluor 546 with B2 amplifier (*Cal-hbn/Cal-slp3*; collection window: 550–645 nm), 640 nm for Alexa Fluor 647 with B1 amplifier (*Cal-slp1*; collection window: 645–700 nm), 488 nm for Alexa Fluor 488 with B3 amplifier (*Cal-slp2*; collection window: 410–550), and 405 nm for DAPI (collection window: 410–550 nm). Z-slice spacing was 1 μm . For each image a Z-stack of images was captured from the embryo's surface to approximately halfway through the embryo. Then a maximal projection was created from this Z-stack using FIJI/Image J. To minimize background fluorescence, the display range minimum of each channel was slightly increased in all images (0–255 scale): 18 for 541 nm, 25 for 640 nm, and 40 for 488 nm. No further adjustments to the images were made.

ATAC-seq library preparation

Single-embryo ATAC-seq library preparations were done essentially as described previously in *Drosophila melanogaster* [76,79]. Embryos developed in an incubator at 25 °C until the desired NC. NC 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 NC 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) [158], 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 placed on ice until all embryos were processed. The nuclei were pelleted by centrifugation at 800 RCF at 4 °C for 10 min. Supernatant removal was observed under a stereo microscope 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 [159]. The tagmentation reaction was performed at 37 °C and 800 RPM for 30 min 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 [159]. 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 μl of Qiagen elution buffer (10 mM Tris-Cl, pH 8.5). 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) [76]. The generalized UDI ATAC primer sequences are:

ATAC UDI Index Read 2 (i5): 5'- AATGATACGGCGACCGAGATCTACACnnnnnnnnTCGTCGGCAGCGTCAGATGT*G -3'

ATAC UDI Index Read 1 (i7): 5'- CAAGCAGAAGACGGCATACGAGATnnnnnnnnGTCTCGTGGGCTCGGAGATG*-3'

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

RNA isolation for RT-qPCR and RNA-seq experiments

Embryos developed in an incubator at 25 °C until the desired NC. NC 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 NC 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; and 1% Igepal CA-630) [158], 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.

RNA-seq library preparation

RNA quality and quantity were assessed using the Agilent bio-analyzer. NonStrand-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 were 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 150 bp).

Experimental setup of ATAC/RNA-seq experiments in *Cal-zld* and *Cal-opa* RNAi embryos

RNAi embryos were processed in parallel with stage-matched untreated embryos that were allowed to develop under water in the dish used for egg activation (wild-type controls), embryos that were prepared for injection under oil but not injected (alignment controls, AC), and embryos that were injected with double-stranded RNA of the extraneous gene *DsRed* (injection controls, IC). *Cal-zld* and *Cal-opa* expression levels were quantified using the RNA-seq data (see *Fold change calculations*).

ATAC-seq library preparation and RNA isolation

The first steps were the same as in the standard ATAC-seq protocol above. Individual embryos were homogenized in cold lysis buffer, and a centrifugation step was performed to pellet the nuclei. The supernatant (~47 µl) was then removed and mixed thoroughly with 450 µl Zymo RNA lysis buffer in a fresh low-bind reaction vial. The nuclear pellet was resuspended in 7.5 µl Tagmentation buffer and the ATAC-seq protocol was continued as described above. RNA isolation proceeded immediately, or the RNA lysate was stored at -20 °C for up to ~1 week. 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.

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 25 °C. Once the embryos had reached the desired stage, any excess oil was removed from the slide using a pipette. The embryos were then washed 6 times using deionized water. Embryos were removed from the injection slide using a blunted tungsten needle and transferred to the lid of a low bind reaction vial. 10 µl of cold lysis buffer was added and the embryos were homogenized using a fire-sealed microcapillary tube. The ATAC-seq protocol was conducted as described above.

ATAC-seq data processing

Demultiplexed reads were trimmed of adapters using TrimGalore! [160] (version 0.6.10, cutadapt version 4.4) and mapped to the *Clogmia albipunctata* assembly (calbi-uni4000) using Bowtie2 [161] (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 [162] (version 1.38.2) and Rsamtools [163] (version 2.18.0). Reads that mapped, were properly paired, nonsecondary and had 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 [158]. Reads with mapped length ≤ 120 bp were considered to have originated from 'open' chromatin.

MACS2. Accessible chromatin peaks from NC11 through NC14 were determined using MACS2 [97] (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, and promoter). Peaks that overlapped with no gene feature were classified as intergenic.

DESeq2. A count matrix for differential enrichment analyses using DESeq2 [101] (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 versus NC11, NC13 versus NC12, NC14 versus NC13, and LNC14 versus NC14) the design parameters passed to DESeq2 were stage and sequencing platform ($\sim$ instrument.atac.data + stage). For the comparison between RNAi embryos and alignment control embryos (e.g., *Cal-zld* RNAi versus alignment control and *Cal-opa* RNAi versus alignment control) the design components were embryo batch and genotype ($\sim$ 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 DESeq2 [164] (R package version 1.38.5) with the parameter of the minimum group size set to 10.

MEME. For [S3 Fig](#), MEME [103] (version 5.4.1) analysis was performed using the following options -order 1 -meme-mod zoops -minw6 -psp-gen -maxw 14 -meme-nmotifs 15 -meme-searchsize 100,000 -centrimo-score 5.0 -centrimo-ethresh 10.0. For [S9 Fig](#), we used MEME (version 5.5.8) with -order 1 -meme-mod zoops -minw4 -maxw 14 -meme-nmotifs 15 -meme-searchsize 100,000 -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 [104], CentriMo), unless specified otherwise. MEME was given a set of sequences containing the 100 bp sequence centered on the summit of the peak. To determine motifs enriched in wild-type ATAC-seq peaks, a set of 1,000 background sequences in genomic regions without ATAC-seq peaks was used as background sequences. To determine motifs enriched in dynamic wild-type peaks, 1,000 randomly chosen nondynamic peak sequences were provided as background sequences. To determine motifs enriched in *Cal-zld* sensitive or *Cal-opa* sensitive peaks, all peaks unaffected by *Cal-zld* knockdown or *Cal-opa* knockdown 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 [116,165–168].

RNA-seq data processing

Demultiplexed reads were trimmed of adapters using TrimGalore [160] (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 [161] (version 2.4.1) with default parameters and maximum fragment length (-X) of 2000. 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 [162] (version 1.38.2) and Rsamtools [163] (version 2.18.0). Reads that mapped, were properly paired, nonsecondary and had a map quality scores ≥ 10 were imported.

DESeq2. A gene count matrix for differential enrichment analyses was generated by counting the number of RNA-seq reads that overlapped with the transcripts 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 versus alignment control and *Cal-opa* RNAi versus 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 .

Maternal, zygotic, maternal, and zygotic classification. Gene expression was compared in NC2/3 (maternal, ~45-min 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, 6,658 maternal and zygotic genes, and 7,473 genes not expressed during either stage.

Fold change calculations. Fold change was calculated as the sample's counts per million (CPM) divided by the control's CPM for the gene of interest. First, a CPM matrix was made by counting the number of RNA-seq reads that overlapped with a gene's transcripts and then normalizing to CPM. 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 from the same experimental date. 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 fold change < 0.25 (knockdown efficiency greater than 75%) were selected to assess the effect of *Cal-zld* knockdown; *Cal-zld* RNAi embryos with a fold change ~ 1 or higher (no *Cal-zld* knockdown detected) and uninjected embryos (alignment controls) were selected for comparison in further analyses. Similarly, for *Cal-opa* RNAi experiments fold change calculations were done separately for each stage. The fold change 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 a reference. *Cal-opa* RNAi embryos with a fold change < 0.25 (knockdown efficiency greater than 75%) and with a fold change > 1 (no knockdown observed, 2 embryos at NC13) were used for further analysis.

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 [169]. High-fidelity (HiFi) PacBio reads were used to generate an initial draft assembly, subsequently refined with Omni-C reads (S1 Appendix). Details on sequencing and software used to assemble the genome have been described elsewhere [91]. We estimated

genome heterozygosity and sequencing error rates using jellyfish to count 21-mers and GenomeScope2.0 to analyze the k-mer frequencies.

Genome annotation

We used *RepeatModeler* [170] (version v2.0.4) to identify repeat elements in the *Clogmia albipunctata* genome and *RepeatMasker*'s [171] (version v4.1.5) *famdb.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 (S1 Table). 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 [91]. 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* [172] (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* [173,174] (version 2.1.12), assigning gene names based on orthology. We assessed the final annotation quality using BUSCO (Benchmarking Universal Single-Copy Orthologs) [175,176] on the predicted protein sequences. Finally, we analyzed synteny between the *C. albipunctata* and *D. melanogaster* genomes using *MCSanX* [177] (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 *circize* R package [178] (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.

TF 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 TF activity was treated as a putative TF gene.

Genome browser

The genomic data produced in this study can be further interrogated using a genome browser ecosystem centered on JBrowse2 [179] and delivered as a cloud image designed for individual use, currently available on NSF's Jetstream2 cloud platform, with a portable Docker image with documentation and BSGenome package created for R. Jetstream2's user-friendly interface (image: *Clogmia albipunctata* Genome Resources) [180,181] integrates tools for genomic analysis, including BLAST search (SequenceServer2.0) [182], CRISPR guide RNA design (modified crisprDesigner) [183], R shiny-driven differential gene expression (DGE) analysis (freecount) [184], and synteny mapping for comparative genomics (ShinySyn) [185]. 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, 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 [91] for

research and education. Installation instructions for accessing the *Clogmia albipunctata* genome browser and related tools are available at https://github.com/kallistaconsulting/genomic_resources_clogmia. The container requires a Linux web server on any cloud platform. The docker was developed on Jetstream2, which can be accessed via instructions at <https://jetstream-cloud.org/get-started/index.html>. This on-demand container-based model reduces resource costs and security concerns compared to hosting a centralized site. Administrative tools (e.g., automated reboot scripts, system service management, load balancing for multi-user groups or workshops) enhance usability while minimizing operational overhead.

Supporting information

S1 Table. Relating to Fig 1 and Table 2: RNA-seq samples used for annotation.

(XLSX)

S2 Table. Relating to Fig 1: Homologs of early segmentation genes in *Clogmia albipunctata*. Homologs were identified by reciprocal BLASTp [186] searches between *Drosophila melanogaster* and *Clogmia albipunctata*. Note that *knirps* and *knirps-like* as well as *sloppy-paired 1*, *sloppy-paired 2*, and *forkhead domain 19B* resulted from gene duplications specific to the *Drosophila* lineage, while *Cal-knrl1* and *Cal-knrl2* as well as *Cal-slp1*, *Cal-slp2*, and *Cal-slp3* resulted from gene duplication specific to the *Clogmia* lineage.

(XLSX)

S3 Table. *Cal-zld* RT-qPCR data. Fold change calculated using the $\Delta\Delta C_t$ method. Reference gene *Cal-rpl35*. Target gene *Cal-zld*. Percent knockdown was calculated as $100 - (FC \times 100)$.

(XLSX)

S4 Table. *Cal-opa* RT-qPCR data. Fold change calculated using the $\Delta\Delta C_t$ method. Reference gene *Cal-rpl35*. Target gene *Cal-opa*. Percent knockdown was calculated as $100 - (FC \times 100)$. Indicated is whether the RNA was isolated immediately or in parallel with an ATAC-seq library preparation. RNA was either isolated using the Zymo Quick-RNA Tissue/Insect Microprep Kit following the manufacturer's instructions kit or following an ATAC-seq library preparation using the protocol by Li and colleagues [187]. Briefly following the tagmentation reaction, mRNA was isolated from the supernatant using Dynabeads Oligo (dT)25 beads. Subsequent washes were performed and mRNA was eluted directly from the beads.

(XLSX)

S1 Fig. Relating to Fig 2: genomic distribution of ATAC-seq peaks at nuclear cycles NC11, NC12, NC13, NC14, and Late NC14.

(TIF)

S2 Fig. Relating to Fig 2: dynamic ATAC-seq peak groups. Dynamic peak groups were generated using DEG reports [164] (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).

(TIF)

S3 Fig. Relating to Fig 2: enriched motifs in ATAC-seq peaks. Enriched motifs identified by MEME for 8 sets of peaks: NC11 peaks ($n=5,013$), NC12 peaks ($n=5,926$), NC13 peaks ($n=7,586$), NC14 peaks ($n=16,147$), Late NC14 peaks ($n=11,628$), all peaks ($n=32,157$), dynamic peaks ($n=6,930$), and nondynamic peaks (right column; $n=25,227$). 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 motifs [105–108].

(TIF)

S4 Fig. 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 [188] (v7) with the default settings. Amino acid color indicates physio-chemical property (based on Zappo Color Scheme). Important protein domains are indicated [110].

(TIF)

S5 Fig. Time-course of *Cal-zld* expression. Gene model in pale yellow. RNA-seq data in counts per million (CPM). Y-axis for NC2/3 (maternal, ~45-min 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.

(TIF)

S6 Fig. Relating to Figs 3 and 4: 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. The treatment group of each embryo is indicated by color. Fold change 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 fold change < 0.25 were selected for the zldKD group (*n* = 4). Alignment control embryos (AC; *n* = 5) and *Cal-zld* RNAi embryos with fold change ~1 or higher (zldnoKD; *n* = 3) served as controls. A *Cal-zld* RNAi embryos with intermediate *Cal-zld* fold change (intKD) was excluded from further analysis. Wt, wild-type embryos.

(TIF)

S7 Fig. Relating to Fig 4: comparison of downregulated and upregulated ATAC-seq peaks in response to *Cal-zld* RNAi. **A)** Overlap of downregulated peaks between zldKD-vs-AC and zldKD-vs-zldnoKD, and between zldKD-vs-zldnoKD and zldnoKD -vs-AC. **B)** Overlap of upregulated peaks between zldKD-vs-AC and zldKD-vs-zldnoKD, and between zldKD-vs-zldnoKD and zldnoKD -vs-AC.

(TIF)

S8 Fig. Relating to Fig 4: *Cal-zld* sensitive ATAC-seq peaks in dynamic and nondynamic 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. The number of peaks and percentage relative to all down- or upregulated peaks are indicated.

(PDF)

S9 Fig. Relating to Fig 4: MEME analysis of *Cal-zld*-sensitive ATAC-seq peaks. Top three differentially enriched motifs (within 50 bp of the peak summit) of downregulated or upregulated ATAC-seq peaks relative to all consistently nonsensitive peaks are shown for zldKD-vs-AC, zldKD-vs-zldnoKD, and the negative control, zldnoKD-vs-AC. Motifs for each comparison are shown as a position weight matrix logo with decreasing significance (increasing *E*-value) from left to right. 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 [105–107].

(PDF)

S10 Fig. Relating to Figs 5, 6, 7, and 8: identification of *Cal-opa* RNAi embryos with strong knockdowns. Plots of *Cal-opa* fold change 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. The treatment group of each embryo is indicated by color. Fold change 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. Fold change was calculated using embryos from a single female (same activation batch). *Cal-opa* RNAi embryos with fold change <0.25 were selected for the opaKD group ($n=4$ for each stage). *Cal-opa* RNAi embryos with intermediate *Cal-opa* fold change (intKD) and one NC12 embryo with *Cal-opa* fold change > 1 (noKD) were excluded from further analysis. Wt, wild-type embryos.
(TIF)

S11 Fig. Relating to Figs 5 and 6: comparisons of opanoKD-vs-AC RNA-seq and ATAC-seq peaks at NC13. A) Downregulated (red) and upregulated (blue) RNA-seq peaks (no TF genes; cf. S6 Appendix). B) Downregulated (red) and upregulated (blue) ATAC-seq peaks (cf. S7 Appendix).
(TIF)

S12 Fig. Relating to Figs 5 and 6: RNA-seq and ATAC-seq tracks of *Cal-hb*. RNA-seq and ATAC-seq tracks are shown for NC12 and NC13 with alignment control in blue, injection control in light blue, *Cal-opa* RNAi without KD in green, and *Cal-opa* RNAi with KD in red. Gene models are shown in pale yellow. ATAC-seq peaks with significant changes in accessibility are highlighted in pink (cf. S7 Appendix). The scale of the y-axis (CPM) is adjusted for each data type but kept constant between stages and is indicated with numbers in square brackets: [0–20] for RNA-seq data; [0–25] for ATAC-seq data.
(TIF)

S13 Fig. 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*, *Caenorhabditis elegans*, *Ischnura elegans*, *Gryllus longicernus*, *Nasonia vitripennis*, *Tribolium castaneum*, *Bombyx mori*, *Clunio marinus*, *Anopheles gambiae*, *Lutzomyia longipalpis*, *Bradysia coprophila*, *Hermetia illucens*, *Condyllostylus longicornis*, and *Episyrphus balteatus* were identified by reciprocal Protein BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) [186]. Homologous sequences in *Limnephilus lunatus* [189], *Panorpa germanica* [190], and *Nephrotoma appendiculata* [191] were identified by reciprocal BLAST in oHoGeneious 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 cutoff 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/>) [188]. Maximum-likelihood trees were constructed using IQ-TREE v.2.3.6. [121]. The best molecular substitution model for each partition under the Bayesian information criterion was selected by partition merging strategy with ModelFinder [122]. Maximum-likelihood trees were constructed under the selected substitution models (VT+I+G4), with branch support values estimated by the ultrafast bootstrap approximation [121] using 1,000 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/>) [192]. *Nematostella* Croc sequence was manually chosen as root.
(PDF)

S14 Fig. 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 progenitors *slp1/fd19B/slp2* (joined rectangles in green, orange, and blue) and *Slp1/fd19B* (joined rectangles in green and orange) are color-coded. Inferred gene duplications are indicated by white spaces between colored boxes.
(TIF)

S1 Movie. 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-min intervals. Imaging was stopped and needed to be restarted twice, this accrues at the ~30 s and ~65 s marks. Scale bar 100 μ M.

(MOV)

S2 Movie. 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 [56]. 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-s 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 min, lower right). Scale bar, 50 μ m.

(MP4)

S1 Data. Relating to Fig 2: ATAC-seq count matrix (sheet 1) and metadata (sheet 2) for time course of wild-type embryos.

(XLSX)

S2 Data. Relating to Fig 3: RNA-seq count matrices and metadata for time course of wild-type embryos (sheet 1 and 2) and *Cal-zld* RNAi embryos (sheet 3 and 4).

(XLSX)

S3 Data. Relating to Fig 4: ATAC-seq count matrix (sheet 1) and metadata (sheet 2) for *Cal-zld* RNAi and control embryos.

(XLSX)

S4 Data. Relating to Fig 5: RNA-seq count matrices and metadata for *Cal-opa* RNAi and control embryos at NC12 (sheet 1 and 2) and NC13 (sheet 3 and 4).

(XLSX)

S5 Data. Relating to Fig 6: ATAC-seq count matrices and metadata for *Cal-opa* RNAi and control embryos at NC12 (sheet 1 and 2) and NC13 (sheet 3 and 4).

(XLSX)

S1 Appendix. Relating to Fig 1: Dovetail/Cantata Bio Hifiasm and HiRise scaffolding reports.

(PDF)

S2 Appendix. Relating to Fig 1: blast hits of annotated *Clogmia albipunctata* genes with flybaseID.

(XLSX)

S3 Appendix. Relating to Fig 2: chromatin accessibility at stages NC11, NC12, NC13, NC14, and Late NC14.

(XLSX)

S4 Appendix. Relating to Fig 3: *Cal-zld* RNAi RNA-seq analysis. 1) zldKD versus alignment control. 2) zldKD versus zldnoKD. 3) zldnoKD versus alignment control.

(XLSX)

S5 Appendix. Relating to Fig 4: *Cal-zld* RNAi ATAC-seq analysis. 1) zldKD versus alignment control. 2) zldKD versus zldnoKD. 3) zldnoKD versus alignment control.

(XLSX)

S6 Appendix. Relating to Fig 5: *Cal-opa* RNAi RNA-seq analysis. 1) NC12 opaKD versus alignment control. 2) NC12 opaKD versus injection control. 3) NC13 opaKD versus alignment control. 4) NC13 opaKD versus alignment control. 5) NC13 opaKD versus opanoKD. 6) NC13 opanoKD versus alignment control.

(XLSX)

S7 Appendix. Relating to Fig 6: *Cal-opa* RNAi ATAC-seq analysis. 1) NC12 opaKD versus alignment control. 2) NC12 opaKD versus injection control. 3) NC13 opaKD versus alignment control. 4) NC13 opaKD versus alignment control. 5) NC13 opaKD versus opanoKD. 6) NC13 opanoKD versus alignment control.

(XLSX)

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*. Edwin “Chip” Ferguson critically reviewed previous versions of this manuscript and provided valuable feedback. Bioinformatic work was supported in part by the Notre Dame University Genomics and Bioinformatics Core Facility.

Disclaimer

The content is solely the responsibility of the authors.

Author contributions

Conceptualization: Ezra E. Amiri, Shelby A. Blythe, Urs Schmidt-Ott.

Data curation: Ezra E. Amiri, Muzi Li.

Formal analysis: Ezra E. Amiri, Muzi Li, Ayse Tenger-Trolander, Maxwell Devine, Alexander T. Julian, Koray Kasan.

Funding acquisition: Shelby A. Blythe, Urs Schmidt-Ott.

Investigation: Ezra E. Amiri, Muzi Li, Maxwell Devine.

Methodology: Ezra E. Amiri, Muzi Li, Ayse Tenger-Trolander, Alexander T. Julian.

Project administration: Sheri A. Sanders, Shelby A. Blythe, Urs Schmidt-Ott.

Resources: Sheri A. Sanders, Shelby A. Blythe, Urs Schmidt-Ott.

Supervision: Shelby A. Blythe, Urs Schmidt-Ott.

Visualization: Ezra E. Amiri, Muzi Li, Ayse Tenger-Trolander, Maxwell Devine, Koray Kasan.

Writing – original draft: Ezra E. Amiri, Shelby A. Blythe, Urs Schmidt-Ott.

Writing – review & editing: Ezra E. Amiri, Muzi Li, Ayse Tenger-Trolander, Maxwell Devine, Sheri A. Sanders, Shelby A. Blythe, Urs Schmidt-Ott.

References

1. Anderson DT. The comparative embryology of the Diptera. *Annu Rev Entomol.* 1966;11:23–64.
2. Campos-Ortega JA, Hartenstein V. The embryonic development of *Drosophila melanogaster*. 2 ed. Berlin, Heidelberg, New York: Springer-Verlag; 1997.

3. Yoon Y, Klomp J, Martin-Martin I, Criscione F, Calvo E, Ribeiro J, et al. Embryo polarity in moth flies and mosquitoes relies on distinct old genes with localized transcript isoforms. *Elife*. 2019;8:e46711. <https://doi.org/10.7554/eLife.46711> PMID: 31591963
4. Klomp J, Athy D, Kwan CW, Bloch NI, Sandmann T, Lemke S, et al. Embryo development. A cysteine-clamp gene drives embryo polarity in the midge *Chironomus*. *Science*. 2015;348(6238):1040–2. <https://doi.org/10.1126/science.aaa7105> PMID: 25953821
5. Rafiqi AM, Lemke S, Ferguson S, Stauber M, Schmidt-Ott U. Evolutionary origin of the amnioserosa in cyclorrhaphan flies correlates with spatial and temporal expression changes of *zen*. *Proc Natl Acad Sci U S A*. 2008;105(1):234–9. <https://doi.org/10.1073/pnas.0709145105> PMID: 18172205
6. Kwan CW, Gavin-Smyth J, Ferguson EL, Schmidt-Ott U. Functional evolution of a morphogenetic gradient. *Elife*. 2016;5:e20894. <https://doi.org/10.7554/eLife.20894> PMID: 28005004
7. Caroti F, González Avalos E, Noeske V, González Avalos P, Kromm D, Wosch M, et al. Decoupling from yolk sac is required for extraembryonic tissue spreading in the scuttle fly *Megaselia abdita*. *eLife*. 2018;7. <https://doi.org/10.7554/elife.34616>
8. Urbansky S, González Avalos P, Wosch M, Lemke S. Folded gastrulation and T48 drive the evolution of coordinated mesoderm internalization in flies. *Elife*. 2016;5:e18318. <https://doi.org/10.7554/eLife.18318> PMID: 27685537
9. Fraire-Zamora JJ, Jaeger J, Solon J. Two consecutive microtubule-based epithelial seaming events mediate dorsal closure in the scuttle fly *Megaselia abdita*. *eLife*. 2018;7. <https://doi.org/10.7554/elife.33807>
10. Dey B, Kaul V, Kale G, Scorcelletti M, Takeda M, Wang Y-C, et al. Divergent evolutionary strategies pre-empt tissue collision in gastrulation. *Nature*. 2025;646(8085):637–46. <https://doi.org/10.1038/s41586-025-09447-4> PMID: 40903584
11. Vellutini BC, Cuenca MB, Krishna A, Szalapak A, Modes CD, Tomancak P. Patterned invagination prevents mechanical instability during gastrulation. *Nature*. 2025;646(8085):627–36. <https://doi.org/10.1038/s41586-025-09480-3> PMID: 40903575
12. Hall AB, Basu S, Jiang X, Qi Y, Timoshevskiy VA, Biedler JK, et al. A male-determining factor in the mosquito *Aedes aegypti*. *Science*. 2015;348(6240):1268–70. <https://doi.org/10.1126/science.aaa2850>
13. Krzywinska E, Dennison NJ, Lycett GJ, Krzywinski J. A maleness gene in the malaria mosquito *Anopheles gambiae*. *Science*. 2016;353(6294):67–9. <https://doi.org/10.1126/science.aaf5605> PMID: 27365445
14. Saccone G. A history of the genetic and molecular identification of genes and their functions controlling insect sex determination. *Insect Biochem Mol Biol*. 2022;151:103873. <https://doi.org/10.1016/j.ibmb.2022.103873> PMID: 36400424
15. Vicoso B, Bachtrog D. Numerous transitions of sex chromosomes in *Diptera*. *PLoS Biol*. 2015;13(4):e1002078. <https://doi.org/10.1371/journal.pbio.1002078> PMID: 25879221
16. St Johnston D, Nüsslein-Volhard C. The origin of pattern and polarity in the *Drosophila* embryo. *Cell*. 1992;68(2):201–19. [https://doi.org/10.1016/0092-8674\(92\)90466-p](https://doi.org/10.1016/0092-8674(92)90466-p) PMID: 1733499
17. Surkova S, Golubkova E, Mamon L, Samsonova M. Dynamic maternal gradients and morphogenetic networks in *Drosophila* early embryo. *Biosystems*. 2018;173:207–13. <https://doi.org/10.1016/j.biosystems.2018.10.009> PMID: 30315821
18. Ma J, He F, Xie G, Deng W-M. Maternal AP determinants in the *Drosophila* oocyte and embryo. *Wiley Interdiscip Rev Dev Biol*. 2016;5(5):562–81. <https://doi.org/10.1002/wdev.235> PMID: 27253156
19. Wieschaus E. Positional information and cell fate determination in the early *Drosophila* embryo. *Curr Top Dev Biol*. 2016;117:567–79. <https://doi.org/10.1016/bs.ctdb.2015.11.020> PMID: 26970001
20. Jaeger J. The gap gene network. *Cell Mol Life Sci*. 2011;68(2):243–74. <https://doi.org/10.1007/s00018-010-0536-y> PMID: 20927566
21. Tkačik G, Gregor T. The many bits of positional information. *Development*. 2021;148(2):dev176065. <https://doi.org/10.1242/dev.176065> PMID: 33526425
22. Clark E, Peel AD, Akam M. Arthropod segmentation. *Development*. 2019;146(18):dev170480. <https://doi.org/10.1242/dev.170480> PMID: 31554626
23. Crombach A, Jaeger J. Life's attractors continued: progress in understanding developmental systems through reverse engineering and in silico evolution. In: Crombach A, editor. *Evolutionary systems biology*. Springer International Publishing; 2021. p. 59–88. https://doi.org/10.1007/978-3-030-71737-7_4
24. Irizarry J, Stathopoulos A. Dynamic patterning by morphogens illuminated by cis-regulatory studies. *Development*. 2021;148(2):dev196113. <https://doi.org/10.1242/dev.196113> PMID: 33472851
25. Driever W, Nüsslein-Volhard C. A gradient of bicoid protein in *Drosophila* embryos. *Cell*. 1988;54(1):83–93. [https://doi.org/10.1016/0092-8674\(88\)90182-1](https://doi.org/10.1016/0092-8674(88)90182-1) PMID: 3383244
26. Driever W, Nüsslein-Volhard C. The bicoid protein determines position in the *Drosophila* embryo in a concentration-dependent manner. *Cell*. 1988;54(1):95–104. [https://doi.org/10.1016/0092-8674\(88\)90183-3](https://doi.org/10.1016/0092-8674(88)90183-3) PMID: 3383245
27. Berleth T, Burri M, Thoma G, Bopp D, Richstein S, Frigerio G, et al. The role of localization of bicoid RNA in organizing the anterior pattern of the *Drosophila* embryo. *EMBO J*. 1988;7(6):1749–56. <https://doi.org/10.1002/j.1460-2075.1988.tb03004.x> PMID: 2901954
28. Stauber M, Taubert H, Schmidt-Ott U. Function of *bicoid* and *hunchback* homologs in the basal cyclorrhaphan fly *Megaselia* (Phoridae). *Proc Natl Acad Sci U S A*. 2000;97(20):10844–9. <https://doi.org/10.1073/pnas.190095397> PMID: 10995461
29. Shaw PJ, Salameh A, McGregor AP, Bala S, Dover GA. Divergent structure and function of the *bicoid* gene in Muscoidea fly species. *Evol Dev*. 2001;3(4):251–62. <https://doi.org/10.1046/j.1525-142x.2001.003004251.x> PMID: 11478522

30. Lemke S, Busch SE, Antonopoulos DA, Meyer F, Domanus MH, Schmidt-Ott U. Maternal activation of gap genes in the hover fly *Episyrphus*. *Development*. 2010;137(10):1709–19. <https://doi.org/10.1242/dev.046649> PMID: 20430746
31. Wotton KR, Jiménez-Guri E, Jaeger J. Maternal co-ordinate gene regulation and axis polarity in the scuttle fly *Megaselia abdita*. *PLoS Genet*. 2015;11(3):e1005042. <https://doi.org/10.1371/journal.pgen.1005042> PMID: 25757102
32. Mulhair PO, Pennati A, Herrera-Ubeda C, Holland PWH. Revised evolutionary relationships within Brachycera and the early origin of bicoid in flies. *Curr Biol*. 2025;35(21):5308–5319.e3. <https://doi.org/10.1016/j.cub.2025.09.051> PMID: 41109215
33. Hursh DA, Stultz BG. Odd-Paired: The *Drosophila* Zic Gene. *Adv Exp Med Biol*. 2018;1046:41–58.
34. Ravindranath AJ, Cadigan KM. Structure-function analysis of the C-clamp of TCF/Pangolin in Wnt/ β -catenin signaling. *PLoS One*. 2014;9(1):e86180. <https://doi.org/10.1371/journal.pone.0086180> PMID: 24465946
35. Liu Q, Onal P, Datta RR, Rogers JM, Schmidt-Ott U, Bulyk ML, et al. Ancient mechanisms for the evolution of the bicoid homeodomain's function in fly development. *Elife*. 2018;7:e34594. <https://doi.org/10.7554/eLife.34594> PMID: 30298815
36. Nitta KR, Jolma A, Yin Y, Morgunova E, Kivioja T, Akhtar J, et al. Conservation of transcription factor binding specificities across 600 million years of bilateria evolution. *Elife*. 2015;4:e04837. <https://doi.org/10.7554/eLife.04837> PMID: 25779349
37. Duan B, Fu D, Zhang C, Ding P, Dong X, Xia B. Selective nonmethylated CpG DNA recognition mechanism of cysteine clamp domains. *J Am Chem Soc*. 2021;143(20):7688–97.
38. Li M, Kasan K, Saha Z, Yoon Y, Schmidt-Ott U. Twenty-seven ZAD-ZNF genes of *Drosophila melanogaster* are orthologous to the embryo polarity determining mosquito gene cucoid. *PLoS One*. 2023;18(1):e0274716. <https://doi.org/10.1371/journal.pone.0274716> PMID: 36595500
39. De Renzis S, Elemento O, Tavazoie S, Wieschaus EF. Unmasking activation of the zygotic genome using chromosomal deletions in the *Drosophila* embryo. *PLoS Biol*. 2007;5(5):e117. <https://doi.org/10.1371/journal.pbio.0050117> PMID: 17456005
40. Lott SE, Villalta JE, Schroth GP, Luo S, Tonkin LA, Eisen MB. Noncanonical compensation of zygotic X transcription in early *Drosophila melanogaster* development revealed through single-embryo RNA-seq. *PLoS Biol*. 2011;9(2):e1000590. <https://doi.org/10.1371/journal.pbio.1000590> PMID: 21346796
41. Kwasniewski JC, Orr-Weaver TL, Bartel DP. Early genome activation in *Drosophila* is extensive with an initial tendency for aborted transcripts and retained introns. *Genome Res*. 2019;29(7):1188–97. <https://doi.org/10.1101/gr.242164.118> PMID: 31235656
42. Tadros W, Lipshitz HD. The maternal-to-zygotic transition: a play in two acts. *Development*. 2009;136(18):3033–42. <https://doi.org/10.1242/dev.033183> PMID: 19700615
43. ten Bosch JR, Benavides JA, Cline TW. The TAGteam DNA motif controls the timing of *Drosophila* pre-blastoderm transcription. *Development*. 2006;133(10):1967–77. <https://doi.org/10.1242/dev.02373> PMID: 16624855
44. Liang H-L, Nien C-Y, Liu H-Y, Metzstein MM, Kirov N, Rushlow C. The zinc-finger protein Zelda is a key activator of the early zygotic genome in *Drosophila*. *Nature*. 2008;456(7220):400–3. <https://doi.org/10.1038/nature07388> PMID: 18931655
45. Chen K, Johnston J, Shao W, Meier S, Staber C, Zeitlinger J. A global change in RNA polymerase II pausing during the *Drosophila* midblastula transition. *Elife*. 2013;2:e00861. <https://doi.org/10.7554/eLife.00861> PMID: 23951546
46. Blythe SA, Wieschaus EF. Zygotic genome activation triggers the DNA replication checkpoint at the midblastula transition. *Cell*. 2015;160(6):1169–81. <https://doi.org/10.1016/j.cell.2015.01.050> PMID: 25748651
47. Larson ED, Marsh AJ, Harrison MM. Pioneering the developmental frontier. *Mol Cell*. 2021;81(8):1640–50. <https://doi.org/10.1016/j.molcel.2021.02.020> PMID: 33689750
48. Iwafuchi-Doi M, Zaret KS. Pioneer transcription factors in cell reprogramming. *Genes Dev*. 2014;28(24):2679–92. <https://doi.org/10.1101/gad.253443.114> PMID: 25512556
49. Zaret KS. Pioneer transcription factors initiating gene network changes. *Annu Rev Genet*. 2020;54:367–85. <https://doi.org/10.1146/annurev-genet-030220-015007> PMID: 32886547
50. Staudt N, Fellert S, Chung H-R, Jäckle H, Vorbrüggen G. Mutations of the *Drosophila* zinc finger-encoding gene vielfältig impair mitotic cell divisions and cause improper chromosome segregation. *Mol Biol Cell*. 2006;17(5):2356–65. <https://doi.org/10.1091/mbc.e05-11-1056> PMID: 16525017
51. Nien C-Y, Liang H-L, Butcher S, Sun Y, Fu S, Gocha T, et al. Temporal coordination of gene networks by Zelda in the early *Drosophila* embryo. *PLoS Genet*. 2011;7(10):e1002339. <https://doi.org/10.1371/journal.pgen.1002339> PMID: 22028675
52. Harrison MM, Li X-Y, Kaplan T, Botchan MR, Eisen MB. Zelda binding in the early *Drosophila melanogaster* embryo marks regions subsequently activated at the maternal-to-zygotic transition. *PLoS Genet*. 2011;7(10):e1002266. <https://doi.org/10.1371/journal.pgen.1002266> PMID: 22028662
53. Satija R, Bradley RK. The TAGteam motif facilitates binding of 21 sequence-specific transcription factors in the *Drosophila* embryo. *Genome Res*. 2012;22(4):656–65. <https://doi.org/10.1101/gr.130682.111> PMID: 22247430
54. Brennan KJ, Weilert M, Krueger S, Pampari A, Liu H-Y, Yang AWH, et al. Chromatin accessibility in the *Drosophila* embryo is determined by transcription factor pioneering and enhancer activation. *Dev Cell*. 2023;58(19):1898–1916.e9. <https://doi.org/10.1016/j.devcel.2023.07.007> PMID: 37557175
55. Foo SM, Sun Y, Lim B, Ziukaite R, O'Brien K, Nien C-Y, et al. Zelda potentiates morphogen activity by increasing chromatin accessibility. *Curr Biol*. 2014;24(12):1341–6. <https://doi.org/10.1016/j.cub.2014.04.032> PMID: 24909324

56. McDaniel SL, Gibson TJ, Schulz KN, Fernandez Garcia M, Nevil M, Jain SU, et al. Continued activity of the pioneer factor zelda is required to drive zygotic genome activation. *Mol Cell*. 2019;74(1):185–195.e4. <https://doi.org/10.1016/j.molcel.2019.01.014> PMID: [30797686](#)
57. Mir M, Stadler MR, Ortiz SA, Hannon CE, Harrison MM, Darzacq X, et al. Dynamic multifactor hubs interact transiently with sites of active transcription in *Drosophila* embryos. *Elife*. 2018;7:e40497. <https://doi.org/10.7554/eLife.40497> PMID: [30589412](#)
58. Mir M, Reimer A, Haines JE, Li X-Y, Stadler M, Garcia H, et al. Dense Bicoid hubs accentuate binding along the morphogen gradient. *Genes Dev*. 2017;31(17):1784–94. <https://doi.org/10.1101/gad.305078.117> PMID: [28982761](#)
59. Duan J, Rieder L, Colonna MM, Huang A, Mckenney M, Watters S, et al. CLAMP and Zelda function together to promote *Drosophila* zygotic genome activation. *Elife*. 2021;10:e69937. <https://doi.org/10.7554/eLife.69937> PMID: [34342574](#)
60. Colonna MM, Abrahante JE, Schedl P, Gohl DM, Deshpande G. CLAMP regulates zygotic genome activation in *Drosophila* embryos. *Genetics*. 2021;219(2):iyab107. <https://doi.org/10.1093/genetics/iyab107> PMID: [34849887](#)
61. Gaskill MM, Gibson TJ, Larson ED, Harrison MM. GAF is essential for zygotic genome activation and chromatin accessibility in the early *Drosophila* embryo. *eLife*. 2021;10. <https://doi.org/10.7554/eLife.66668>
62. Moshe A, Kaplan T. Genome-wide search for Zelda-like chromatin signatures identifies GAF as a pioneer factor in early fly development. *Epigenetics Chromatin*. 2017;10(1):33. <https://doi.org/10.1186/s13072-017-0141-5> PMID: [28676122](#)
63. Xu Z, Chen H, Ling J, Yu D, Struffi P, Small S. Impacts of the ubiquitous factor Zelda on Bicoid-dependent DNA binding and transcription in *Drosophila*. *Genes Dev*. 2014;28(6):608–21.
64. Bishop TR, Onal P, Xu Z, Zheng M, Gunasinghe H, Nien C-Y, et al. Multi-level regulation of even-skipped stripes by the ubiquitous factor Zelda. *Development*. 2023;150(23):dev201860. <https://doi.org/10.1242/dev.201860> PMID: [37934130](#)
65. Crocker J, Stern DL. Functional regulatory evolution outside of the minimal even-skipped stripe 2 enhancer. *Development*. 2017;144(17):3095–101. <https://doi.org/10.1242/dev.149427> PMID: [28760812](#)
66. Ling J, Umezawa KY, Scott T, Small S. Bicoid-dependent activation of the target gene hunchback requires a two-motif sequence code in a specific basal promoter. *Mol Cell*. 2019;75(6):1178–1187.e4. <https://doi.org/10.1016/j.molcel.2019.06.038> PMID: [31402096](#)
67. Eck E, Liu J, Kazemzadeh-Atoufi M, Ghoreishi S, Blythe SA, Garcia HG. Quantitative dissection of transcription in development yields evidence for transcription-factor-driven chromatin accessibility. *Elife*. 2020;9:e56429. <https://doi.org/10.7554/eLife.56429> PMID: [33074101](#)
68. Fernandes G, Tran H, Andrieu M, Diaw Y, Perez Romero C, Fradin C, et al. Synthetic reconstruction of the hunchback promoter specifies the role of Bicoid, Zelda and Hunchback in the dynamics of its transcription. *Elife*. 2022;11:e74509. <https://doi.org/10.7554/eLife.74509> PMID: [35363606](#)
69. Harden TT, Vincent BJ, DePace AH. Transcriptional activators in the early *Drosophila* embryo perform different kinetic roles. *Cell Syst*. 2023;14(4):258–272.e4. <https://doi.org/10.1016/j.cels.2023.03.006> PMID: [37080162](#)
70. Huang A, Amourda C, Zhang S, Tolwinski NS, Saunders TE. Decoding temporal interpretation of the morphogen Bicoid in the early *Drosophila* embryo. *Elife*. 2017;6:e26258. <https://doi.org/10.7554/eLife.26258> PMID: [28691901](#)
71. Hannon CE, Blythe SA, Wieschaus EF. Concentration dependent chromatin states induced by the bicoid morphogen gradient. *Elife*. 2017;6:e28275. <https://doi.org/10.7554/eLife.28275> PMID: [28891464](#)
72. Datta RR, Ling J, Kurland J, Ren X, Xu Z, Yucel G, et al. A feed-forward relay integrates the regulatory activities of Bicoid and Orthodenticle via sequential binding to suboptimal sites. *Genes Dev*. 2018;32(9–10):723–36. <https://doi.org/10.1101/gad.311985.118> PMID: [29764918](#)
73. Haines JE, Eisen MB. Patterns of chromatin accessibility along the anterior-posterior axis in the early *Drosophila* embryo. *PLoS Genet*. 2018;14(5):e1007367. <https://doi.org/10.1371/journal.pgen.1007367> PMID: [29727464](#)
74. Cusanovich DA, Reddington JP, Garfield DA, Daza RM, Aghamirzaie D, Marco-Ferreres R, et al. The *cis*-regulatory dynamics of embryonic development at single-cell resolution. *Nature*. 2018;555(7697):538–42. <https://doi.org/10.1038/nature25981> PMID: [29539636](#)
75. Bozek M, Cortini R, Storti AE, Unnerstall U, Gaul U, Gompel N. ATAC-seq reveals regional differences in enhancer accessibility during the establishment of spatial coordinates in the *Drosophila* blastoderm. *Genome Res*. 2019;29(5):771–83. <https://doi.org/10.1101/gr.242362.118> PMID: [30962180](#)
76. Soluri IV, Zumerling LM, Payan Parra OA, Clark EG, Blythe SA. Zygotic pioneer factor activity of Odd-paired/Zic is necessary for late function of the *Drosophila* segmentation network. *Elife*. 2020;9:e53916. <https://doi.org/10.7554/eLife.53916> PMID: [32347792](#)
77. Calderon D, Blecher-Gonen R, Huang X, Secchia S, Kentro J, Daza RM, et al. The continuum of *Drosophila* embryonic development at single-cell resolution. *Science*. 2022;377(6606):eabn5800. <https://doi.org/10.1126/science.abn5800> PMID: [35926038](#)
78. Degen EA, Croslyn C, Mangan NM, Blythe SA. Bicoid-nucleosome competition sets a concentration threshold for transcription constrained by genome replication. *Cell Rep*. 2025;44(8):116121. <https://doi.org/10.1016/j.celrep.2025.116121> PMID: [40779393](#)
79. Blythe SA, Wieschaus EF. Establishment and maintenance of heritable chromatin structure during early *Drosophila* embryogenesis. *Elife*. 2016;5:e20148. <https://doi.org/10.7554/eLife.20148> PMID: [27879204](#)
80. Nüsslein-Volhard C, Wieschaus E. Mutations affecting segment number and polarity in *Drosophila*. *Nature*. 1980;287(5785):795–801. <https://doi.org/10.1038/287795a0> PMID: [6776413](#)
81. Clark E. Dynamic patterning by the *Drosophila* pair-rule network reconciles long-germ and short-germ segmentation. *PLoS Biol*. 2017;15(9):e2002439. <https://doi.org/10.1371/journal.pbio.2002439> PMID: [28953896](#)

82. Foe VE, Odell GM, Edgar BA. Mitosis and morphogenesis in the *Drosophila* embryo: point and counterpoint. In: Bate M, Martinez-Arias A, editors. The development of *Drosophila melanogaster*. Cold Spring Harbor: Cold Spring Harbor Laboratory Press; 1993. p. 149–300.
83. Jiménez-Guri E, Wotton KR, Gavilán B, Jaeger J. A staging scheme for the development of the moth midge *Clogmia albipunctata*. PLoS One. 2014;9(1):e84422. <https://doi.org/10.1371/journal.pone.0084422> PMID: 24409296
84. Janssens H, Siggins K, Cicin-Sain D, Jiménez-Guri E, Musy M, Akam M, et al. A quantitative atlas of even-skipped and hunchback expression in *Clogmia albipunctata* (Diptera: Psychodidae) blastoderm embryos. Evodevo. 2014;5(1):1. <https://doi.org/10.1186/2041-9139-5-1> PMID: 24393251
85. Clark E, Akam M. Odd-paired controls frequency doubling in *Drosophila* segmentation by altering the pair-rule gene regulatory network. Elife. 2016;5:e18215. <https://doi.org/10.7554/eLife.18215> PMID: 27525481
86. Koromila T, Gao F, Iwasaki Y, He P, Pachter L, Gergen JP, et al. Odd-paired is a pioneer-like factor that coordinates with Zelda to control gene expression in embryos. Elife. 2020;9:e59610. <https://doi.org/10.7554/eLife.59610> PMID: 32701060
87. Amabis JM, Simões LCG. Chromosome studies in a species of *Telmatoscopus* Sp. Caryologia. 1972;25(2):199–210. <https://doi.org/10.1080/00087114.1972.10796477>
88. Amabis JM. Cytological evidence of male heterogamety in sex determination of *Telmatoscopus albipunctatus* (Diptera: Psychodidae). Chromosoma. 1977;62(2):133–8. <https://doi.org/10.1007/BF00292635> PMID: 880846
89. Schmidt-Ott U, Rafiqi AM, Sander K, Johnston JS. Extremely small genomes in two unrelated dipteran insects with shared early developmental traits. Dev Genes Evol. 2009;219(4):207–10. <https://doi.org/10.1007/s00427-009-0281-0> PMID: 19308443
90. Generalovic TN, McCarthy SA, Warren IA, Wood JMD, Torrance J, Sims Y, et al. A high-quality, chromosome-level genome assembly of the Black Soldier Fly (*Hermetia illucens* L.). G3 (Bethesda). 2021;11(5):jkab085. <https://doi.org/10.1093/g3journal/jkab085> PMID: 33734373
91. Tenger-Trolander A, Amiri E, Gantz V, Kwan CW, Sanders SA, Schmidt-Ott U. Genomic resources for the scuttle fly *Megaselia abdita*: a model organism for comparative developmental studies in flies. bioRxiv. 2025.
92. Ji J, Gao Y, Xu C, Zhang K, Li D, Li B, et al. Chromosome-level genome assembly of marmalade hoverfly *Episyrphus balteatus* (Diptera: Syrphidae). Sci Data. 2024;11(1):844. <https://doi.org/10.1038/s41597-024-03666-6> PMID: 39097648
93. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409(6822):860–921. <https://doi.org/10.1038/35057062> PMID: 11237011
94. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013;29(8):1072–5. <https://doi.org/10.1093/bioinformatics/btt086> PMID: 23422339
95. Stauber M, Jäckle H, Schmidt-Ott U. The anterior determinant bicoid of *Drosophila* is a derived Hox class 3 gene. Proc Natl Acad Sci U S A. 1999;96(7):3786–9. <https://doi.org/10.1073/pnas.96.7.3786> PMID: 10097115
96. Stauber M, Prell A, Schmidt-Ott U. A single Hox3 gene with composite bicoid and zerknullt expression characteristics in non-Cyclorrhaphan flies. Proc Natl Acad Sci U S A. 2002;99(1):274–9. <https://doi.org/10.1073/pnas.012292899> PMID: 11773616
97. Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9):R137. <https://doi.org/10.1186/gb-2008-9-9-r137> PMID: 18798982
98. Reddington JP, Garfield DA, Sigalova OM, Karabacak Calviello A, Marco-Ferreres R, Girardot C, et al. Lineage-resolved enhancer and promoter usage during a time course of embryogenesis. Dev Cell. 2020;55(5):648–664.e9. <https://doi.org/10.1016/j.devcel.2020.10.009> PMID: 33171098
99. Schulz KN, Harrison MM. Mechanisms regulating zygotic genome activation. Nat Rev Genet. 2019;20(4):221–34. <https://doi.org/10.1038/s41576-018-0087-x> PMID: 30573849
100. Hamm DC, Harrison MM. Regulatory principles governing the maternal-to-zygotic transition: insights from *Drosophila melanogaster*. Open Biol. 2018;8(12):180183. <https://doi.org/10.1098/rsob.180183> PMID: 30977698
101. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550. <https://doi.org/10.1186/s13059-014-0550-8> PMID: 25516281
102. Sun Y, Nien C-Y, Chen K, Liu H-Y, Johnston J, Zeitlinger J, et al. Zelda overcomes the high intrinsic nucleosome barrier at enhancers during *Drosophila* zygotic genome activation. Genome Res. 2015;25(11):1703–14. <https://doi.org/10.1101/gr.192542.115> PMID: 26335633
103. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME suite. Nucleic Acids Res. 2015;43(W1):W39–49.
104. Bailey TL. STREME: accurate and versatile sequence motif discovery. Bioinformatics. 2021;37(18):2834–40. <https://doi.org/10.1093/bioinformatics/btab203> PMID: 33760053
105. Soruco MML, Chery J, Bishop EP, Siggers T, Tolstorukov MY, Leydon AR, et al. The CLAMP protein links the MSL complex to the X chromosome during *Drosophila* dosage compensation. Genes Dev. 2013;27(14):1551–6. <https://doi.org/10.1101/gad.214585.113> PMID: 23873939
106. Kuzu G, Kaye EG, Chery J, Siggers T, Yang L, Dobson JR, et al. Expansion of GA dinucleotide repeats increases the density of clamp binding sites on the X-chromosome to promote *Drosophila* dosage compensation. PLoS Genet. 2016;12(7):e1006120. <https://doi.org/10.1371/journal.pgen.1006120> PMID: 27414415
107. Kaye EG, Booker M, Kurland JV, Conicella AE, Fawzi NL, Bulky ML, et al. Differential occupancy of two GA-binding proteins promotes targeting of the *Drosophila* dosage compensation complex to the male X chromosome. Cell Rep. 2018;22(12):3227–39. <https://doi.org/10.1016/j.celrep.2018.02.098> PMID: 29562179

108. Ray P, De S, Mitra A, Bezstarosti K, Demmers JAA, Pfeifer K, et al. Combgap contributes to recruitment of polycomb group proteins in *Drosophila*. *Proc Natl Acad Sci U S A*. 2016;113(14):3826–31. <https://doi.org/10.1073/pnas.1520926113> PMID: 27001825
109. Gibson TJ, Larson ED, Harrison MM. Protein-intrinsic properties and context-dependent effects regulate pioneer factor binding and function. *Nat Struct Mol Biol*. 2024;31(3):548–58. <https://doi.org/10.1038/s41594-024-01231-8> PMID: 38365978
110. Ribeiro L, Tobias-Santos V, Santos D, Antunes F, Feltran G, de Souza Menezes J, et al. Evolution and multiple roles of the Pancrustacea specific transcription factor zelda in insects. *PLoS Genet*. 2017;13(7):e1006868. <https://doi.org/10.1371/journal.pgen.1006868> PMID: 28671979
111. Biedler JK, Hu W, Tae H, Tu Z. 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*. 2012;7(3):e33933. <https://doi.org/10.1371/journal.pone.0033933> PMID: 22457801
112. Arsala D, Lynch JA. Ploidy has little effect on timing early embryonic events in the haplo-diploid wasp *Nasonia*. *Genesis*. 2017;55(5):10.1002/dvg.23029. <https://doi.org/10.1002/dvg.23029> PMID: 28432826
113. Pires CV, Freitas FC de P, Cristino AS, Dearden PK, Simões ZLP. Transcriptome analysis of honeybee (*Apis Mellifera*) haploid and diploid embryos reveals early zygotic transcription during cleavage. *PLoS One*. 2016;11(1):e0146447. <https://doi.org/10.1371/journal.pone.0146447> PMID: 26751956
114. Gao S, Xue S, Gao T, Lu R, Zhang X, Zhang Y, et al. Transcriptome analysis reveals the role of Zelda in the regulation of embryonic and wing development of *Tribolium castaneum*. *Bull Entomol Res*. 2023;113(5):587–97. <https://doi.org/10.1017/S0007485323000263> PMID: 37476851
115. Ventos-Alfonso A, Ylla G, Belles X. Zelda and the maternal-to-zygotic transition in cockroaches. *FEBS J*. 2019;286(16):3206–21. <https://doi.org/10.1111/febs.14856> PMID: 30993896
116. Zhu LJ, Christensen RG, Kazemian M, Hull CJ, Enuameh MS, Basciotta MD, et al. FlyFactorSurvey: a database of *Drosophila* transcription factor binding specificities determined using the bacterial one-hybrid system. *Nucleic Acids Res*. 2011;39(Database issue):D111–7. <https://doi.org/10.1093/nar/gkq858> PMID: 21097781
117. García-Solache M, Jaeger J, Akam M. A systematic analysis of the gap gene system in the moth midge *Clogmia albipunctata*. *Dev Biol*. 2010;344(1):306–18. <https://doi.org/10.1016/j.ydbio.2010.04.019> PMID: 20433825
118. Goltsev Y, Hsiong W, Lanzaro G, Levine M. Different combinations of gap repressors for common stripes in *Anopheles* and *Drosophila* embryos. *Dev Biol*. 2004;275(2):435–46. <https://doi.org/10.1016/j.ydbio.2004.08.021> PMID: 15501229
119. Wotton KR, Jiménez-Guri E, Crombach A, Janssens H, Alcaine-Colet A, Lemke S, et al. Quantitative system drift compensates for altered maternal inputs to the gap gene network of the scuttle fly *Megaselia abdita*. *Elife*. 2015;4:e04785. <https://doi.org/10.7554/eLife.04785> PMID: 25560971
120. Bullock SL, Stauber M, Prell A, Hughes JR, Ish-Horowicz D, Schmidt-Ott U. Differential cytoplasmic mRNA localisation adjusts pair-rule transcription factor activity to cytoarchitecture in dipteran evolution. *Development*. 2004;131(17):4251–61. <https://doi.org/10.1242/dev.01289> PMID: 15280214
121. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol*. 2020;37(5):1530–4. <https://doi.org/10.1093/molbev/msaa015> PMID: 32011700
122. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermini LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 2017;14(6):587–9. <https://doi.org/10.1038/nmeth.4285> PMID: 28481363
123. Cardamone F, Piva A, Löser E, Eichenberger B, Romero-Mulero MC, Zenk F, et al. Chromatin landscape at cis-regulatory elements orchestrates cell fate decisions in early embryogenesis. *Nat Commun*. 2025;16(1):3007. <https://doi.org/10.1038/s41467-025-57719-4> PMID: 40148291
124. Gonzaga-Saavedra N, Degen EA, Soluri IV, Croslyn C, Blythe SA. Nucleation-dependent propagation of Polycomb modifications emerges during the *Drosophila* maternal to zygotic transition. *eLife*. 2025;14:RP108371.
125. Schulz KN, Bondra ER, Moshe A, Villalta JE, Lieb JD, Kaplan T, et al. Zelda is differentially required for chromatin accessibility, transcription factor binding, and gene expression in the early *Drosophila* embryo. *Genome Res*. 2015;25(11):1715–26. <https://doi.org/10.1101/gr.192682.115> PMID: 26335634
126. Li X-Y, Harrison MM, Villalta JE, Kaplan T, Eisen MB. Establishment of regions of genomic activity during the *Drosophila* maternal to zygotic transition. *Elife*. 2014;3:e03737. <https://doi.org/10.7554/eLife.03737> PMID: 25313869
127. Galouzis CC, Kherdjemil Y, Forneris M, Viales RR, Marco-Ferreres R, Furlong EEM. Chip (Ldb1) is a putative cofactor of Zelda forming a functional bridge to CBP during zygotic genome activation. *Mol Cell*. 2025;85(12):2425–2441.e9. <https://doi.org/10.1016/j.molcel.2025.05.018> PMID: 40494353
128. Giannini P, Tsitilou SG. The embryonic transcription factor Zelda of *Drosophila melanogaster* is also expressed in larvae and may regulate developmentally important genes. *Biochem Biophys Res Commun*. 2013;438(2):329–33. <https://doi.org/10.1016/j.bbrc.2013.07.071> PMID: 23891688
129. Hamm RL, Meisel RP, Scott JG. The evolving puzzle of autosomal versus Y-linked male determination in *Musca domestica*. *G3 (Bethesda)*. 2014;5(3):371–84. <https://doi.org/10.1534/g3.114.014795> PMID: 25552607
130. Hamm DC, Larson ED, Nevil M, Marshall KE, Bondra ER, Harrison MM. A conserved maternal-specific repressive domain in Zelda revealed by Cas9-mediated mutagenesis in *Drosophila melanogaster*. *PLoS Genet*. 2017;13(12):e1007120. <https://doi.org/10.1371/journal.pgen.1007120> PMID: 29261646
131. Pearson JC, Watson JD, Crews ST. *Drosophila melanogaster* Zelda and Single-minded collaborate to regulate an evolutionarily dynamic CNS midline cell enhancer. *Dev Biol*. 2012;366(2):420–32. <https://doi.org/10.1016/j.ydbio.2012.04.001> PMID: 22537497

132. Schröder C, Tautz D, Seifert E, Jäckle H. Differential regulation of the two transcripts from the *Drosophila* gap segmentation gene hunchback. *EMBO J*. 1988;7(9):2881–7. <https://doi.org/10.1002/j.1460-2075.1988.tb03145.x> PMID: 2846287
133. Tautz D. Regulation of the *Drosophila* segmentation gene hunchback by two maternal morphogenetic centres. *Nature*. 1988;332(6161):281–4. <https://doi.org/10.1038/332281a0> PMID: 2450283
134. Löhner U, Chung H-R, Beller M, Jäckle H. Antagonistic action of Bicoid and the repressor Capicua determines the spatial limits of *Drosophila* head gene expression domains. *Proc Natl Acad Sci U S A*. 2009;106(51):21695–700. <https://doi.org/10.1073/pnas.0910225106> PMID: 19959668
135. Chen H, Xu Z, Mei C, Yu D, Small S. A system of repressor gradients spatially organizes the boundaries of Bicoid-dependent target genes. *Cell*. 2012;149(3):618–29. <https://doi.org/10.1016/j.cell.2012.03.018> PMID: 22541432
136. Judd J, Duarte FM, Lis JT. Pioneer-like factor GAF cooperates with PBAP (SWI/SNF) and NURF (ISWI) to regulate transcription. *Genes Dev*. 2021;35(1–2):147–56. <https://doi.org/10.1101/gad.341768.120> PMID: 33303640
137. Tsukiyama T, Becker PB, Wu C. ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor. *Nature*. 1994;367(6463):525–32. <https://doi.org/10.1038/367525a0> PMID: 8107823
138. Grand RS, Pregniolato M, Baumgartner L, Hoerner L, Burger L, Schübeler D. Genome access is transcription factor-specific and defined by nucleosome position. *Mol Cell*. 2024;84(18):3455–3468.e6. <https://doi.org/10.1016/j.molcel.2024.08.009> PMID: 39208807
139. Polach KJ, Widom J. Mechanism of protein access to specific DNA sequences in chromatin: a dynamic equilibrium model for gene regulation. *J Mol Biol*. 1995;254(2):130–49. <https://doi.org/10.1006/jmbi.1995.0606> PMID: 7490738
140. Hildebrandt K, Kolb D, Klöppel C, Kaspar P, Wittling F, Hartwig O, et al. Regulatory modules mediating the complex neural expression patterns of the homeobrain gene during *Drosophila* brain development. *Hereditas*. 2022;159(1):2. <https://doi.org/10.1186/s41065-021-00218-5> PMID: 34983686
141. Kolb D, Kaspar P, Klöppel C, Walldorf U. The *Drosophila* homeodomain transcription factor Homeobrain is involved in the formation of the embryonic protocerebrum and the supraesophageal brain commissure. *Cells Dev*. 2021;165:203657. <https://doi.org/10.1016/j.cdev.2021.203657> PMID: 33993980
142. Walldorf U, Kiewe A, Wickert M, Ronshaugen M, McGinnis W. Homeobrain, a novel paired-like homeobox gene is expressed in the *Drosophila* brain. *Mech Dev*. 2000;96(1):141–4. [https://doi.org/10.1016/S0925-4773\(00\)00380-4](https://doi.org/10.1016/S0925-4773(00)00380-4) PMID: 10940637
143. Ransick A, Davidson EH. A complete second gut induced by transplanted micromeres in the sea urchin embryo. *Science*. 1993;259(5098):1134–8. <https://doi.org/10.1126/science.8438164> PMID: 8438164
144. Ansari S, Troelenberg N, Dao VA, Richter T, Bucher G, Klingler M. Double abdomen in a short-germ insect: Zygotic control of axis formation revealed in the beetle *Tribolium castaneum*. *Proc Natl Acad Sci U S A*. 2018;115(8):1819–24. <https://doi.org/10.1073/pnas.1716512115> PMID: 29432152
145. Schoppmeier M, Fischer S, Schmitt-Engel C, Löhner U, Klingler M. An ancient anterior patterning system promotes *caudal* repression and head formation in ecdysozoa. *Curr Biol*. 2009;19(21):1811–5. <https://doi.org/10.1016/j.cub.2009.09.026> PMID: 19818622
146. Clark E, Peel AD. Evidence for the temporal regulation of insect segmentation by a conserved sequence of transcription factors. *Development*. 2018;145(10):dev155580. <https://doi.org/10.1242/dev.155580> PMID: 29724758
147. Schomburg C, Janssen R, Prpic N-M. Phylogenetic analysis of forkhead transcription factors in the Panarthropoda. *Dev Genes Evol*. 2022;232(1):39–48. <https://doi.org/10.1007/s00427-022-00686-3> PMID: 35230523
148. Cadigan KM, Grossniklaus U, Gehring WJ. Functional redundancy: the respective roles of the two sloppy paired genes in *Drosophila* segmentation. *Proc Natl Acad Sci U S A*. 1994;91(14):6324–8. <https://doi.org/10.1073/pnas.91.14.6324> PMID: 8022780
149. Fujioka M, Jaynes JB. Regulation of a duplicated locus: *Drosophila* sloppy paired is replete with functionally overlapping enhancers. *Dev Biol*. 2012;362(2):309–19. <https://doi.org/10.1016/j.ydbio.2011.12.001> PMID: 22178246
150. Grossniklaus U, Pearson RK, Gehring WJ. The *Drosophila* sloppy paired locus encodes two proteins involved in segmentation that show homology to mammalian transcription factors. *Genes Dev*. 1992;6(6):1030–51. <https://doi.org/10.1101/gad.6.6.1030> PMID: 1317319
151. Lee H-H, Frasch M. Survey of forkhead domain encoding genes in the *Drosophila* genome: classification and embryonic expression patterns. *Dev Dyn*. 2004;229(2):357–66. <https://doi.org/10.1002/dvdy.10443> PMID: 14745961
152. Andrioli LP, Oberstein AL, Corado MSG, Yu D, Small S. Groucho-dependent repression by sloppy-paired 1 differentially positions anterior pair-rule stripes in the *Drosophila* embryo. *Dev Biol*. 2004;276(2):541–51. <https://doi.org/10.1016/j.ydbio.2004.09.025> PMID: 15581884
153. Schroeder MD, Greer C, Gaul U. How to make stripes: deciphering the transition from non-periodic to periodic patterns in *Drosophila* segmentation. *Development*. 2011;138(14):3067–78. <https://doi.org/10.1242/dev.062141> PMID: 21693522
154. Choe CP, Miller SC, Brown SJ. A pair-rule gene circuit defines segments sequentially in the short-germ insect *Tribolium castaneum*. *Proc Natl Acad Sci U S A*. 2006;103(17):6560–4. <https://doi.org/10.1073/pnas.0510440103> PMID: 16611732
155. Choe CP, Brown SJ. 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*. *Dev Biol*. 2007;302(1):281–94. <https://doi.org/10.1016/j.ydbio.2006.09.037> PMID: 17054935
156. Choi HMT, Schwarzkopf M, Fornace ME, Acharya A, Artavanis G, Stegmaier J, et al. Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. *Development*. 2018;145(12):dev165753. <https://doi.org/10.1242/dev.165753> PMID: 29945988

157. Bruce HS, Jerz G, Kelly SR, McCarthy J, Pomerantz A, Senevirathne G, et al. Hybridization chain reaction (HCR) in situ protocol V.1. 2021. <https://doi.org/10.17504/protocols.io.bunznvf6>
158. Buenroostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods*. 2013;10(12):1213–8. <https://doi.org/10.1038/nmeth.2688> PMID: [24097267](https://pubmed.ncbi.nlm.nih.gov/24097267/)
159. Buenroostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: A method for assaying chromatin accessibility genome-wide. In: Ausubel FM, editor. *Current protocols in molecular biology*. 2015. p. 21.9.1–9.
160. Krueger F. *Trim galore!*. Babraham bioinformatics. 2012.
161. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357–9. <https://doi.org/10.1038/nmeth.1923> PMID: [22388286](https://pubmed.ncbi.nlm.nih.gov/22388286/)
162. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, et al. Software for computing and annotating genomic ranges. *PLoS Comput Biol*. 2013;9(8):e1003118. <https://doi.org/10.1371/journal.pcbi.1003118> PMID: [23950696](https://pubmed.ncbi.nlm.nih.gov/23950696/)
163. Morgan M, Pages H, Obenchain V, Hayden N. Rsamtools: Binary alignment (BAM), FASTA, variant call (BCF), and tabix file import. *Bioconductor*; 2013.
164. Pantano L. DEGreport: report of DEG analysis 2017. Available from: <https://doi.org/10.18129/B9.bioc.DEGreport>
165. Bergman CM, Carlson JW, Celniker SE. *Drosophila* DNase I footprint database: a systematic genome annotation of transcription factor binding sites in the fruitfly, *Drosophila melanogaster*. *Bioinformatics*. 2005;21(8):1747–9. <https://doi.org/10.1093/bioinformatics/bti173> PMID: [15572468](https://pubmed.ncbi.nlm.nih.gov/15572468/)
166. Kulakovskii IV, Makeev VI. Integration of data obtained by different experimental methods to determine the motifs in DNA sequences recognized by transcription-regulating factors. *Biofizika*. 2009;54(6):965–74. PMID: [20067172](https://pubmed.ncbi.nlm.nih.gov/20067172/)
167. Kulakovskiy IV, Favorov AV, Makeev VJ. Motif discovery and motif finding from genome-mapped DNase footprint data. *Bioinformatics*. 2009;25(18):2318–25. <https://doi.org/10.1093/bioinformatics/btp434> PMID: [19605419](https://pubmed.ncbi.nlm.nih.gov/19605419/)
168. Shazman S, Lee H, Socol Y, Mann RS, Honig B. OnTheFly: a database of *Drosophila melanogaster* transcription factors and their binding sites. *Nucleic Acids Res*. 2014;42(Database issue):D167–71. <https://doi.org/10.1093/nar/gkt1165> PMID: [24271386](https://pubmed.ncbi.nlm.nih.gov/24271386/)
169. Rohr KB, Tautz D, Sander K. Segmentation gene expression in the mothmidge *Clogmia albipunctata* (Diptera, psychodidae) and other primitive dipterans. *Dev Genes Evol*. 1999;209(3):145–54. <https://doi.org/10.1007/s004270050238> PMID: [10079357](https://pubmed.ncbi.nlm.nih.gov/10079357/)
170. Smit A, Hubley R. RepeatModeler 2008. Available from: <https://www.repeatmasker.org/>
171. Smit A, Hubley R, Green P. RepeatMasker 2013. Available from: <https://www.repeatmasker.org/>
172. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol*. 2008;9(1):R7. <https://doi.org/10.1186/gb-2008-9-1-r7> PMID: [18190707](https://pubmed.ncbi.nlm.nih.gov/18190707/)
173. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res*. 2019;47(D1):D309–14. <https://doi.org/10.1093/nar/gky1085> PMID: [30418610](https://pubmed.ncbi.nlm.nih.gov/30418610/)
174. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol Evol*. 2021;38(12):5825–9. <https://doi.org/10.1093/molbev/msab293> PMID: [34597405](https://pubmed.ncbi.nlm.nih.gov/34597405/)
175. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015;31(19):3210–2. <https://doi.org/10.1093/bioinformatics/btv351> PMID: [26059717](https://pubmed.ncbi.nlm.nih.gov/26059717/)
176. Manni M, Berkeley MR, Seppely M, Zdobnov EM. BUSCO: assessing genomic data quality and beyond. *Curr Protoc*. 2021;1(12):e323. <https://doi.org/10.1002/cpz1.323> PMID: [34936221](https://pubmed.ncbi.nlm.nih.gov/34936221/)
177. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res*. 2012;40(7):e49. <https://doi.org/10.1093/nar/gkr1293> PMID: [22217600](https://pubmed.ncbi.nlm.nih.gov/22217600/)
178. Gu Z, Gu L, Eils R, Schlesner M, Brors B. circlize Implements and enhances circular visualization in R. *Bioinformatics*. 2014;30(19):2811–2. <https://doi.org/10.1093/bioinformatics/btu393> PMID: [24930139](https://pubmed.ncbi.nlm.nih.gov/24930139/)
179. Diesh C, Stevens GJ, Xie P, De Jesus Martinez T, Hershberg EA, Leung A, et al. JBrowse 2: a modular genome browser with views of synteny and structural variation. *Genome Biol*. 2023;24(1):74. <https://doi.org/10.1186/s13059-023-02914-z> PMID: [37069644](https://pubmed.ncbi.nlm.nih.gov/37069644/)
180. Hancock DY, Fischer J, Lowe JM, Snapp-Childs W. Jetstream2: accelerating cloud computing via jetstream. In: *Practice and experience in advanced research computing*; 2021.
181. Boerner TJ, Deems S, Furlani TR, Knuth SL, Towns J. ACCESS: advancing innovation: NSF's advanced cyberinfrastructure coordination ecosystem: services & support. In: *Practice and experience in advanced research computing (PEARC'23)*; 2023.
182. Priyam A, Woodcroft BJ, Rai V, Moghul I, Munagala A, Ter F, et al. Sequenceserver: a modern graphical user interface for custom BLAST databases. *Mol Biol Evol*. 2019;36(12):2922–4. <https://doi.org/10.1093/molbev/msz185> PMID: [31411700](https://pubmed.ncbi.nlm.nih.gov/31411700/)
183. Beeber D, Chain FJ. crisperDesignR: a versatile guide RNA design package in R for CRISPR/Cas9 applications. *J Genomics*. 2020;8:62–70. <https://doi.org/10.7150/jgen.41196> PMID: [32494309](https://pubmed.ncbi.nlm.nih.gov/32494309/)

184. Brooks EM, Sanders SA, Pfrender ME. freeCount: a coding free framework for guided count data visualization and analysis. In: Practice and experience in advanced research computing 2024: human powered computing, 2024. pp. 1–4. <https://doi.org/10.1145/3626203.3670605>
185. Xiao Z, Lam H-M. ShinySyn: a Shiny/R application for the interactive visualization and integration of macro- and micro-synteny data. *Bioinformatics*. 2022;38(18):4406–8. <https://doi.org/10.1093/bioinformatics/btac503> PMID: [35866686](https://pubmed.ncbi.nlm.nih.gov/35866686/)
186. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009;10:421. <https://doi.org/10.1186/1471-2105-10-421> PMID: [20003500](https://pubmed.ncbi.nlm.nih.gov/20003500/)
187. Li R, Grimm SA, Wade PA. Low-input ATAC&mRNA-seq protocol for simultaneous profiling of chromatin accessibility and gene expression. *STAR Protoc*. 2021;2(3):100764. <https://doi.org/10.1016/j.xpro.2021.100764> PMID: [34485936](https://pubmed.ncbi.nlm.nih.gov/34485936/)
188. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform*. 2019;20(4):1160–6. <https://doi.org/10.1093/bib/bbx108> PMID: [28968734](https://pubmed.ncbi.nlm.nih.gov/28968734/)
189. Austin M, Clifford C, Rutt G, Price BW, Natural History Museum Genome Acquisition Lab, Wellcome Sanger Institute Tree of Life programme, et al. The genome sequence of a caddisfly, *Limnephilus lunatus* (Curtis, 1834). *Wellcome Open Res*. 2023;8:25. <https://doi.org/10.12688/wellcome-openres.18752.1> PMID: [37408608](https://pubmed.ncbi.nlm.nih.gov/37408608/)
190. Sivell D, Mitchell R, Crowley LM, Natural History Museum Genome Acquisition L, University of O, Wytham Woods Genome Acquisition L. The genome sequence of the German scorpionfly, *Panorpa germanica* Linnaeus, 1758. *Wellcome Open Res*. 2024;9:285.
191. Crowley LM, Wawman DC, University of O, Wytham Woods Genome Acquisition L, Darwin Tree of Life Barcoding c, Wellcome Sanger Institute Tree of Life Management S, et al. The genome sequence of the spotted crane fly, *Nephrotoma appendiculata* (Pierre, 1919). *Wellcome Open Res*. 2024;9:38.
192. Rambaut A. Figtree ver 1.4.4. Edinburgh, UK: Institute of Evolutionary Biology, University of Edinburgh; 2018.