Conservation of symmetry breaking at the level of chromatin accessibility between fly species with unrelated anterior determinants

Authors

Ezra E. Amiri[1] <emsy@uchicago.edu> (ORCiD 0000-0002-6241-6656)

Ayse Tenger-Trolander[1] <atengertrolander@uchicago.edu> (ORCiD 0000-0001-8319-0273)

Muzi Li[1] <muzi1@uchicago.edu> (ORCiD 0000-0001-8194-5214)

Alexander Thomas Julian[2] < ajulian@hawk.iit.edu> (ORCiD 0000-0001-6268-5007)

Koray Kasan[1] <kkasan@uchicago.edu> (ORCiD 0000-0001-8098-0376)

Sheri A. Sanders[3] <Sheri.A.Sanders.57@nd.edu> (ORCiD 0000-0003-29692541)

Shelby Blythe[4,5*] <shelby.blythe@northwestern.edu> (ORCiD 0000-0003-4986-2579)

Urs Schmidt-Ott[1*] <uschmid@uchicago.edu> (ORCiD 0000-0002-1351-9472)

* Authors for correspondence

Author affiliations

1. The University of Chicago, Department of Organismal Biology and Anatomy, 1027 East 57th Street, Chicago, Illinois 60637, USA
2. Illinois Institute of Technology, Department of Biology, 3105 South Dearborn Street, Chicago, Illinois 60616, USA
3. Notre Dame University, 252 Galvin Life Science Center/Freimann Life Science Center, Notre Dame, Indiana 46556, USA
4. Northwestern University, Department of Molecular Biosciences, 2205 Tech Drive, Evanston, Illinois 60208, USA
5. Northwestern University and The University of Chicago, National Institute for Theory and Mathematics in Biology, 875 North Michigan Avenue, Suite 3500, Chicago, Illinois 60611, USA

Author Contributions

EA: Conceptualization, Methodology, Data curation, Investigation, Formal analysis, Visualization, Writing – original draft, and Writing – review and editing. ATT: Methodology, Formal analysis, Visualization, Writing – review and editing. ML: Data curation and Investigation. AJ: Formal analysis and Methodology. KK: Formal analysis and Visualization. SS: Project administration, Resources, Writing – review and editing. SB: Conceptualization, Resources, Funding acquisition, Project administration, Supervision, Writing – original draft, Writing – review and editing. USO: Conceptualization, Resources, Funding acquisition, Project administration, Supervision, Writing – original draft, Writing – review and editing.

Abstract

Establishing the anterior-posterior body axis is a fundamental process during embryogenesis, and the fruit fly, *Drosophila melanogaster*, provides one of the best-known case studies of this process. In Drosophila, localized mRNA of *bicoid* serves as anterior determinant (AD). Bicoid engages in a concentration-dependent competition with nucleosomes and initiates symmetry-breaking along the AP axis by promoting chromatin accessibility at the loci of transcription factor (TF) genes that are expressed in the anterior of the embryo. However, ADs of other fly species are unrelated and structurally distinct, and little is known about how they function. We addressed this question in the moth fly, *Clogmia albipunctata*, in which a maternally expressed transcript isoform of the pair-rule segmentation gene *odd-paired* is localized in the anterior egg and has been co-opted as AD. We provide a *de novo* assembly and annotation of the Clogmia genome and describe how knockdown of *zelda* and maternal *odd-paired* transcript affect chromatin accessibility and expression of TF-encoding loci. The results of these experiments suggest direct roles of Cal-Zld in opening and closing chromatin during nuclear cleavage cycles and show that Clogmia's maternal *odd-paired* activity promotes chromatin accessibility and anterior expression during the early phase of zygotic genome activation at Clogmia's *homeobrain* and *sloppy-paired* loci. We conclude that unrelated dipteran ADs initiate anterior-posterior axis-specification at the level of enhancer accessibility and that *homeobrain* and *sloppy-paired* homologs may serve a more widely conserved role in the initiation of anterior pattern formation given their early anterior expression and function in head development in other insects.

Introduction

Early embryos face the challenge of breaking symmetry and establishing spatially restricted domains of gene expression that define the body plan. Dipteran insects (flies, midges, mosquitoes) can achieve this through concentration gradients of transcription factors, referred to as morphogens, because their early embryos develop as a syncytium in which protein products from maternally localized mRNAs can diffuse (1, 2). Dipteran embryos undergo multiple rounds of synchronous nuclear division cycles (**NC**) before the blastoderm, a monolayer of nuclei between the cell membrane of the egg and the yolk, cellularizes and enters gastrulation to form the segmented germband. Because of its anatomical simplicity and experimental accessibility, the syncytial and cellular blastoderm of Drosophila has been widely used as a model for axis specification, pattern formation, and gene regulation and is among the best-known developmental transcription factor (**TF**) / gene networks in animals (3-10). Many aspects of embryogenesis and pattern formation are conserved between distantly related dipterans and provide a unifying framework for comparative studies (10-13). However, the maternally provided anterior determinants that establish anterior-specific zygotic gene expression and the embryo's overall head-to-tail polarity differ widely between species across the dipteran clade (14, 15). This finding raises the question of which other aspects of Drosophila's classic model of axial pattern formation (16) have diverged within the dipteran clade.

While the anterior determinant (**AD**) in Drosophila and many other higher flies (Cyclorrhapha) is encoded by a homeobox gene, *bicoid* (17-21), diverse zinc finger genes take on this role in lower dipterans, including *pangolin (Tcf* ortholog*)* in anopheline mosquitoes and crane flies, *cucoid* in culicine mosquitoes, *panish* in certain harlequin flies (Chironomini), and *odd-paired* (*Zic*

ortholog) in moth flies (Psychodidae) (14, 15). All these species deposit a maternally expressed transcript isoform of the respective AD gene in the anterior tip of the egg that, once translated, specifies anterior cell fates. Without the AD, embryos develop a bicaudal phenotype without head (double abdomen). The AD-encoded proteins (for simplicity we also refer to them as AD) differ widely in their structure and DNA-binding domains. Bicoid (Bcd) binds DNA via its homeodomain, Pangolin (Pan) through HMG and C-clamp domains, Panish through a C-clamp domain, and Opa and Cucoid through distinct C2H2 zinc finger domains (22-27). This molecular diversity prompted us to ask whether targets and mechanisms of action of dipteran ADs differ between species or whether they are conserved due to downstream network constraints.

Bcd binds to enhancers that promote the transcription of early segmentation genes, such as *hunchback* (28, 29), *Krüppel* (30), *knirps* (31), *giant* (32-34), *orthodenticle* (35), or *even-skipped* (36-39). This Bcd activity is complemented by the maternal TF gradients of Hunchback (Hb) (33, 40-42), which peaks in the anterior embryo and Caudal (Cad) (31, 43), which is repressed in the anterior embryo (44). Additionally, ubiquitous TFs, such as the pioneer factor Zelda (Zld) (45, 46), establish or promote chromatin accessibility at many sites that are targeted by Bcd (47-55). During early blastoderm stages, chromatin accessibility measured by ATAC-seq (assay for transposase-accessible chromatin with sequencing) is largely homogeneous across the embryo; however, a small cohort of early segmentation gene enhancers exhibits anteriorly restricted accessibility in response to high concentrations of Bcd (e.g., enhancers of "head gap genes", *knirps*, and *giant*, and *even-skipped* that drive expression in anterior domains) (56-60). Bcd engages in a concentration-dependent competition with nucleosomes to 'open' these otherwise inaccessible regulatory regions, effectively breaking symmetry along the AP axis at the level of chromatin accessibility (61), well before the bulk of zygotic genome activation and pattern formation occurs (62). While Bcd exerts a concentration-dependent function in gene regulation through its impact on chromatin states (56), it is not known whether this mechanism of action can be generalized for other dipteran ADs.

We have adressed this question by examining how the AD of *Clogmia albipunctata* affects chromatin accessibility and gene expression during the early phase of zygotic genome activation. Clogmia lacks *bicoid* and uses a maternal transcript isoform of *odd-paired* (*Cal-opa*) as anterior determinant (15). Knockdown of the maternal *Cal-opa* transcript by isoform-specific RNA interference (RNAi) results in the double abdomen phenotype, and posterior mis-expression of *Cal-opa* by mRNA injection causes a bicephalic (double head) phenotype. An alternative zygotic transcript isoform of Clogmia *odd-paired* appears during blastoderm cellularization at NC14 and exerts a conserved funtion in the late segmentation gene network but has no role in establishing head-to-tail polarity of the embryo (15, 63). However, prior to activation of zygotic *Cal-opa* expression, ectopic posterior *odd-paired* activity from either of the maternal or the zygotic isoform or of Drosphila *odd-paired* results in double heads in Clogmia (15). Early Cal-Opa activity is therefore both necessary and sufficient for establishing head-to-tail polarity of the Clogmia embryo. The symmetric double abdomen phenotype of Clogmia embryos with reduced maternal *Cal-opa* transcript is very similar to the phenotype of Drosophila embryos lacking the maternal transcripts of *bicoid* and *hunchback* (64). We note here that in Drosophila, maternal *hunchback* provides residual polarity to *bicoid*-deficient embryos (65). This is not the case in Clogmia, in which a gradient of Hb protein appears at NC12, which is later than in Drosophila but still before transcription intiation of several other "gap genes", including homologs of *Krüppel*, *giant*, *knirps*, and *tailless* (66, 67).

In Drosophila, zygotic Opa pioneers chromatin accessibility in late NC14 embryos to advance the progression of segmentation and dorso-vental patterning (58, 68). We therefore asked whether the maternal *Cal-opa* gradient of Clogmia breaks axial symmetry and initiates anterior pattern formation during zygotic genome activation by driving chromatin accessibility at the loci of select target genes. Like Drosophila, Clogmia forms a syncytial blastoderm at NC9 and cellularizes the blastoderm at NC14 shortly before entering gastrulation, although Clogmia's nuclear cycles last longer: NC10 ~2 times longer, NC11-13 ~3 times longer, and NC14 ~1.6 times longer (69). Here, we have used a *de novo* assembled and annotated genome sequence of *Clogmia albipunctata* to characterize chromatin accessibility in single wild-type embryos at NC11, NC12, NC13, and early and late NC14, using ATAC-seq. We then established a protocol for studying chromatin accessibility and gene expression changes in single blastoderm embryos that were subjected to RNA interference (RNAi), using Clogmia's ortholog of the pioneer factor gene *zelda* (*Cal-zld*) as target and confirmed that it has a pronounced impact on chromatin accessibility and gene expression. Finally, we examined how knockdown of maternal *odd-paired* transcript (*Cal-opa*) affected chromatin accessibility and gene expression during early stages of zygotic genome activation. Our results reveal groups of genes with similar dynamic changes in chromatin accessibility, point to direct roles of Cal-Zld in opening and closing chromatin, and support the hypothesis that the conserved feature of anterior determination between Clogmia and Drosophila is symmetry breaking at the level of chromatin accessibility, while key target genes may differ between these species.

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 (**Figure 1, Table 1, S1 Appendix**), consistent with the known chromosome number in this species (70, 71). 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) (72). 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 (73-75). While synteny between the chromosomes within the Cyclorrhapha clade can be highly conserved (divergence time ~145 million years ago) (74), synteny between the chromosomes of *D. melanogaster* and *C. albipunctata* (divergence time ~250 million years ago) was essentially lost.

Next, we interrogated the genome for the presence of predicted early segmentation genes. As expected, no *bicoid* ortholog was found but we recovered two putative Hox class 3 genes to which *bicoid* belongs (76), including the previously reported homolog of *zerknüllt* (*Cal-zen*) (77) on scaffold 7 and a more diverged potential *Cal-zen* paralog in position 3 of the Hox gene complex on scaffold 3 (evm.model.scaffold_3.2943.1.65be5b92). Our annotation also included homologs of *Drosophila melanogaster*'s early segmentation genes, including gap genes (*tailless*, *huckebein*, *orthodenticle/ocelliless*, *empty spiracles*, *buttonhead*, *hunchback*, *Krüppel*, *giant,*

knirps), pair-rule genes (*even-skipped*, *fushi tarazu*, *odd-skipped*, *runt*, *sloppy-paired*, *paired, hairy, odd-paired*), and other important regulators (*caudal*, *Dichaete*) (7, 9). Some of these genes have more than one copy in the Clogmia genome, including 2 copies of *hunchback* (*Cal-hb1, Cal-hb2*), *knirps/knirps-like* (*Cal-kni/knrl1, Cal-kni/knrl2*), *orthodenticle* (*Cal-otd1, Cal-otd2*), *even-skipped* (*Cal-eve1, Cal-eve2*), *tailless* (*Cal-tll1, Cal-tll2*) and *caudal* (*Cal-cad 1, Cal-cad2*), 3 copies of *sloppy-paired1/2/fd19B* (*Cal-slp1, Cal-slp2, Cal-slp3*), and 4 copies of *empty spiracles* (*Cal-ems1, Cal-ems2, Cal-ems3, Cal-ems4*) (**Figure 1, S2 Table**). Such gene duplications are a potential source of redundancies in Clogmia's segmentation gene network (see below).

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

To identify accessible chromatin before and during the main wave of zygotic genome activation, we performed single embryo ATAC-seq at NC11, NC12, NC13, NC14 (before cellularization), and Late NC14 approximately 20 minutes before gastrulation. The embryos were staged *in vivo* using developmental time after egg activation at 25ºC and morphological criteria as previously described (69) and validated under our laboratory conditions. At least three replicates were performed for NC11 though NC14, and two replicates were performed for Late NC14. We pooled the ATAC-seq data from all stages to create a master peak list of 32157 open chromatin regions (peaks), using MACS2 (78). We also computed the number of peaks for each developmental stage excluding stage-specific peaks that were not recovered as significant in the master peak list. In this analysis, we found that the overall number of peaks increases with each nuclear cycle, suggesting that the chromatin landscape is continuously changing (**Table 3**). Since promoter accessibility does not necessarily indicate gene expression (79), we divided the accessible chromatin regions into transcription start site (TSS) proximal regions (<500 bp of TSS) and TSS distal regions (>500 bp) to facilitate the distinction of open promoters and other open cis-regulatory elements. ATAC-seq peaks that only mapped to intergenic or intronic sequence were treated as candidate enhancers or silencers (henceforth collectively referred to as enhancers). The relative proportion of candidate enhancers increased by about 10% from NC11 (46.8%) to NC14 (58%) and then decreased from NC14 to Late NC14 (52.7%) (**S1 Figure**). These observations suggest similar dynamic changes in chromatin accessibility during zygotic genome activation in Clogmia and Drosophila (80, 81).

To identify peaks with dynamic patterns of chromatin accessibility, we compared the degree of accessibility of each peak between successive timepoints (**Figure 2A-D**) using DESeq2 (82). This analysis identified a total of 6545 dynamically regulated peaks (adjusted p-value ≤ 0.05 and $|\log_2 \text{foldchange}| \geq 1$; 20.4% of all 32157 peaks) and enabled us to quantitatively determine the behavior of each peak between consecutive developmental stages and to group them according to their accessibility changes over time (**S2 Figure**). The majority (5122 / 6545 dynamic peaks; 78%) fell into 8 groups (**Figure 2E)**. Many of the dynamic peaks closed (~18%; Groups 5,7) or opened (~15%; 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 3,4,15,17; see also **S2 Figure**). These observations suggest that remodeling of chromatin accessibility between NC11 and Late NC14 at individual peaks often follows patterns that are shared by many peaks and may reflect co-regulation of the associated genes.

215

216 ***Cal-zld* is required for promoting and closing chromatin at the blastoderm stage in Clogmia**

217 Previous studies in Drosophila have shown that temporal changes in chromatin accessibility are
218 regulated by the interplay of transcription factors that can interact with nucleosome-bound DNA
219 (pioneer factors) and transcription factors that have limited or no ability to initiate chromatin
220 accessibility but affect the accessibility of open chromatin (83, 84). The earliest changes in
221 chromatin accessibility are primarily driven by the pioneer factor Zld (46, 80, 85-87). As zygotic
222 genome activation accelerates, additional pioneer factors, such as GAF (88), Clamp (89, 90), and
223 Opa (58, 68) are required for maintaining, modulating, or gaining chromatin accessibility. In
224 Clogmia, a motif enrichment analysis of dynamic peaks with the MEME software package (91)
225 revealed an enrichment of a Zelda-binding motif (5'-CAGGTA), GA-rich motifs recognized by
226 GAF and Clamp, in addition to $(CA)_n$ motifs recognized by Combgap (Cg), and others (**S3**
227 **Figure**). Opa-binding motifs were not recovered in this analysis. These results are comparable to
228 findings in Drosophila (58, 62, 68) and suggest that key pioneer factors acting during zygotic
229 genome activation are likely conserved between Drosophila and Clogmia.

230 Homologs of *zelda* have been identified in many insects and some crustaceans (92), and several
231 transcriptomic studies suggest that Zld's function as pioneer transcription factor during zygotic
232 genome activation is conserved across insects (93-97). However, Zld's pioneer activity is stage-
233 specific in *Drosophila* (98), and these comparative studies did not directly test for Zld-dependent
234 effects on chromatin accessibility. We also identified Clogmia *zld* (*Cal-zld*) (**S4 Figure**) and
235 confirmed high levels of *Cal-zld* expression at preblastoderm and blastoderm stages (**S5 Figure**).
236 Loss of Zld in *Drosophila* results in a failure to activate a large cohort of early zygotic genes,
237 including those associated with the process of cellularization (46). To examine phenotypic effects
238 of *Cal-zld* depletion, batches of ~30 embryos were injected with *Cal-zld* dsRNA (double-
239 stranded RNA) within the first hour of development (NC2/3). A subset of embryos from each
240 batch (3 replicates) were monitored *in vivo* and allowed to develop until the end of NC14
241 (roughly 7-7.5 hours after injection). Many of the embryos in each batch failed to develop a
242 blastoderm while some developed normally, presumably because of insufficient *Cal-zld*
243 knockdown. A third class of embryos reached NC14 but failed to cellularize and developed
244 aberrantly (**S1 Movie**). This phenotype is comparable to the phenotype of *in vivo* imaged *zld*
245 mutant Drosophila embryos (**S2 Movie**) (86), and suggests that the role of Zld for the regulation
246 of early zygotic genes is conserved in *Clogmia.* We also quantified knockdown efficiencies of
247 *Cal-zld* transcript in individual embryos by RT-qPCR (**S3 Table**). In one quarter of the *Cal-zld*
248 RNAi embryos (5/20), *Cal-zld* mRNA levels were reduced by ~30-90%.

249 To assess *Cal-zld*'s role in shaping chromatin accessibility and zygotic genome activation, we
250 adapted our ATAC-seq protocol to isolate RNA from single-embryo ATAC-seq library
251 preparations of Late NC14 *Cal-zld* RNAi embryos (8 replicates). In parallel, we processed stage-
252 matched untreated embryos that were allowed to develop under water in the dish used for egg
253 activation (wild-type controls; 5 replicates), and embryos that were prepared for injection under
254 oil but not injected (alignment controls; 5 replicates). The isolated RNA was used to generate
255 RNA-seq libraries to measure the efficiency of the *Cal-zld* knockdown as well as transcriptional
256 effects of the knockdown in individual embryos. To identify individual injected embryos with
257 strong *Cal-zld* knockdowns, we quantified *Cal-zld* expression levels using the RNA-seq data and
258 chose embryos with knockdown efficiencies greater than 75% (n= 4, **S6 Figure** and see
259 *Materials and Methods*). We then compared gene expression and ATAC-seq peaks of these *Cal-*

zld RNAi embryos to those of the alignment control group using DESeq2. Loss of *Cal-zld* affected 3575 genes of which 1693 (47.4%) were downregulated and 1882 (52.6%) were upregulated. We then compared genes differentially expressed in *Cal-zld* RNAi embryos to genes we found to be expressed maternally, zygotically, or both (**Figure 3A** and *Materials and Methods*). 691 genes were classified as zygotically expressed of which 286 (41.4%) were downregulated in *Cal-zld* RNAi embryos and 58 (8.4%) were upregulated (corresponding to 16.9% of all 1693 downregulated and 3.1% of all 1882 upregulated genes, respectively). 6658 genes were classified as expressed maternally and zygotically of which 1327 (19.9%) were downregulated and 1707 (25.6%) were upregulated. Finally, of the 227 genes classified as maternally expressed, only 2 (0.9%) were found to be downregulated and 28 (12.3%) upregulated in *Cal-zld* RNAi embryos. Since the strictly zygotic genes were ~5 times more often downregulated than upregulated (286 versus 58), our results suggest that *Cal-zld* plays an important role in activating zygotic genes (**Figure 3A, left),** consistent with the role of Zld during zygotic genome activation in Drosophila embryos (46, 85). Changes in the levels of maternal transcripts in response to *Cal-zld* RNAi could reflect indirect posttranscriptional roles of *Cal-zld* in their regulation.

At the level of chromatin accessibility, *Cal-zld* depletion affected 7181 of all 32,157 peaks (22.3%). 4272 of them lost accessibility (59.5%) and 2909 gained accessibility (40.5%). The majority (~70%) of the peaks mapped to intergenic or intronic sequences (**Figure 3B)**. Depletion of *Cal-zld* mRNA affected both dynamic peaks and non-dynamic peaks (**S7 Figure**). Dynamic peaks were affected by Cal-zld knockdown in different ways (**Figure 3C**). Peaks in groups that progressively gained accessibility in wild-type embryos (Groups 8, 12, 15 and 17) were more likely to be *Cal-zld* dependent and close in Cal-zld RNAi background (**Figure 3C, *left* and S8A Figure**). This observation is consistent with an important role of Cal-Zld in opening new chromatin regions during zygotic genome activation. Additionally, our data suggest that Cal-Zld also plays an important role in closing chromatin regions during zygotic genome activation (**S8B Figure**). For example, peaks in dynamic groups whose accessibility decreased in wild-type with each nuclear cycle (Groups 3, 5 and 7) were more likely to require Cal-Zld to close (**Figure 3C, *right* and S8B Figure**).

To better distinguish direct from indirect Cal-Zld targets, we performed a Zld motif search for all *Cal-zld* sensitive peaks, using Zld motifs derived from our MEME analysis in Clogmia (5-CAGGTA; cf. **S3 Figure)** and experimental data in *Drosophila melanogaster* (Fly Factor Survey) (99). Overall, 4763 of the 7181 *Cal-zld* sensitive peaks contained at least one Zld motif **(Figure 3D**). Zld motifs were mostly found in the peaks that were reduced following *Cal-zld* knockdown (3252 peaks) but also in peaks that increased following *Cal-zld* knockdown (1511). In peaks that gained accessibility following *Cal-zld* knockdown, Zld motifs were *centrally* enriched (within 50 bp of the peak summit; **S9 Figure)**. This was not the case for peaks that lost accessibility following *Cal-zld* knockdown. Central Cal-Zld binding may therefore facilitate the recruitment of repressors at these sites. Taken together, our findings suggest that Cal-Zld plays a central role in changing chromatin accessibility in pre-gastrulation embryos.

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

303 Having established that *Cal-zld* regulates chromatin accessibility in Clogmia embryos, we asked
304 whether maternal *Cal-opa* affects chromatin accessibility during early stages of axial pattern
305 formation. Since *Cal-opa* exerts its symmetry-breaking function in axis specification prior to the
306 onset zygotic *Cal-opa* expression during blastoderm cellularization at mid NC14 (15), we were
307 able to examine the function of *Cal-opa* without confounding effects of zygotic *Cal-opa* activity.

308 To assess the impact of maternal Cal-Opa activity on chromatin accessibility, we performed
309 ATAC-seq and RNA-seq on single *Cal-opa* RNAi embryos at NC12 (6 replicates) and NC13 (8
310 replicates). In parallel, we processed stage-matched untreated controls that were allowed to
311 develop under water in the dish used for egg activation (wild-type controls; 2 replicates for
312 NC12 and 3 replicates for NC13), as well as controls that were prepared for injection under oil
313 but not injected (alignment controls; 4 replicates for each developmental stage), and embryos
314 that were injected with dsRNA of the extraneous gene *DsRed* (injection controls; 3 replicates for
315 each stage). To identify RNAi embryos with maximally reduced *Cal-opa* mRNA levels, we used
316 the RNA-seq data to measure *Cal-opa* expression and selected embryos with knockdown
317 efficiencies greater than 75% (**S10 Figure** and see *Materials and Methods*). We identified 4
318 embryos at NC12 and 5 embryos at NC13 that were then compared to controls using DESeq2.
319 First, we compared the *Cal-opa* RNAi embryos to the alignment controls. This choice was made
320 to reduce noise in the data that could mask relevant targets. Below we report these results before
321 turning to the comparison of the *Cal-opa* RNAi embryos to the respective injection controls.

322 At NC12, very few changes in chromatin accessibility were observed in response to *Cal-opa*
323 RNAi. Out of the 91 differentially accessible regions, 85 (93.4%) lost accessibility and 6 (6.6%)
324 gained accessibility (**Figure 4A**). At NC13, depletion of *Cal-opa* had a more pronounced effect
325 on chromatin accessibility, with 206/230 (89.6%) peaks losing accessibility and 24/230 (10.4%)
326 gaining accessibility **(Figure 4B**). Most of the differentially accessible regions mapped to
327 putative enhancer regions (**S11 Figure**). We also performed a motif enrichment analysis using
328 the MEME suite on all peaks as well as the peaks that either lost or gained accessibility in *Cal-*
329 *opa* RNAi embryos at either NC12 or NC13 (**Figure 4C**). In regions that lost accessibility in
330 *Cal-opa* RNAi embryos, the most enriched motifs were GAF/Clamp and Opa binding motifs. In
331 addition, we recovered a motif with no associated transcription factor. Regions that gained
332 accessibility were enriched in DNA binding motifs of GAF/Clamp, Extradenticle (Exd),
333 Zerknüllt (Zen), and a fourth motif with no associated transcription factor.

334 Next, we determined the number of *Cal-opa* sensitive peaks that contained putative Cal-Opa
335 binding sites, including Opa motifs derived from our MEME analysis and a MEME analysis of
336 ATAC-seq data from Drosophila (58). At NC12, 54 of the 85 peaks losing accessibility and 1 of
337 the 6 peaks gaining accessibility in response to reduced *Cal-opa* activity contained at least one
338 Opa motif (**S2 Appendix**). At NC13, 94 of the 206 peaks losing accessibility and 3 of the 24
339 peaks gaining accessibility in response to reduced *Cal-opa* activity contained at least one Opa
340 motif (**S3 Appendix**). Overall, these results support the hypothesis that Opa promotes chromatin
341 accessibility at a subset of its binding sites during the early phase of zygotic genome activation.

342

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

345 The early zygotic segmentation gene network of dipterans is dominated by TF-encoding genes,
346 such as the gap genes and pair-rule genes (7, 9, 67, 100-104). Since Bcd breaks axial symmetry

347 by targeting these genes in a concentration-dependent manner, we focused our search of key Cal-
348 Opa targets on TF genes that are expressed during blastoderm formation. Using the RNA-seq
349 data from the same sets of embryos that were used to identify regions of differential chromatin
350 accessibility, we identified 104 differentially expressed genes at NC12, including 16 transcription
351 factor genes, all of which were downregulated upon *Cal-opa* knockdown (**Figure 5A and S3**
352 **Appendix**). At NC13, we identified 1165 differentially expressed genes, including 132 putative
353 transcription factor genes of which 77 were downregulated and 55 were upregulated in *Cal-opa*
354 RNAi embryos (**Figure 5B and S3 Appendix**). We then asked which of the differentially
355 expressed transcription factor genes are located near open chromatin regions (ATAC-seq peaks)
356 with an Opa motif. This analysis was restricted to peaks within a 40 kb region centered on each
357 gene body of interest and further narrowed the list of potential key target genes of Cal-Opa to 14
358 genes at NC12 and 86 genes at NC13 (**S2 Appendix and S3 Appendix)**. At NC12, 4 of the 14
359 candidate genes were associated with at least one Cal-Opa dependent ATAC-seq peak, including
360 *Cal-opa,* and homologs of *homeobrain* (*Cal-hbn*) and *sloppy-paired* (*Cal-slp1*, *Cal-slp2*) (**Figure**
361 **6, S4 Table, S12 Figure and S13 Figure**). At NC13, we thereby identified *Cal-opa* and *Cal-hbn*
362 in addition to Clogmia orthologs of *Dichaete* (*Cal-D*) and *Krüppel* (*Cal-Kr*) among the 86
363 candidates (**Figure 7, S4 Table and S13 Figure**). At this developmental stage, peaks at the loci
364 of *Cal-slp1* and *Cal-slp2* with reduced accessibility in *Cal-opa* RNAi embryos either did not
365 meet our significance threshold or were Opa motif negative. In the case of *Cal-opa*, a reduction
366 of expression was expected because the maternal *Cal-opa* transcript was targeted by RNAi and
367 the zygotic *Cal-opa* expression begins after NC13 (15). Nevertheless, autoregulation of *Cal-opa*
368 is a possibility given that we identified Cal-Opa sensitive peaks with Cal-Opa motifs at the *Cal-*
369 *opa* locus.

370 We tested the robustness of these results by repeating the analysis using the injection control
371 group as the reference. At NC12, we recovered 12 of the 14 initially identified candidate direct
372 targets (**S4 Table and S2 Appendix**), of which two, *Cal-hbn* and *Cal-slp1*, were associated with
373 Cal-Opa dependent ATAC-seq peaks, though only the peak at the *Cal-hbn* contained an Opa
374 motif (**Figure 6 and S12 Figure**). At NC13, we recovered 20 of the 86 candidate target genes
375 (**S4 Table and S3 Appendix**) of which 4 were associated with Cal-Opa dependent ATAC-seq,
376 including *Cal-hbn* and *Cal-slp1* (**Figure 7 and S12 Figure**).

377 *Cal-hbn* and *Cal-slp1* expression was strictly zygotic and detected during and after NC11 when
378 both genes were expressed in non-overlapping anterior domains (**Figure 8A, B**). Maternal *Cal-*
379 *slp2* transcript was previously found to be distributed in a shallow anterior-to-posterior gradient
380 and may contribute to the embryo's head-to-tail polarity in subtle ways (15). Zygotic *Cal-slp2*
381 expression during and after NC11 was detected in an anterior stripe (**Figure 8C**). A third *sloppy-*
382 *paired* homolog that we identified in the Clogmia genome (evm.TU.scaffold_5.987; **S14 Figure**
383 **and S15 Figure**) was expressed at low levels at NC12 and NC13 and robustly at NC14. The
384 expression of this gene was reduced in *Cal-opa* RNAi embryos but associated ATAC-seq peaks
385 did not contain Opa-binding motifs. Taken together, our findings suggest that *Cal-hbn*, and *Cal-*
386 *slp1* and *Cal-slp2* are direct early targets of Clogmia's AD and that Cal-Opa serves as their
387 transcriptional activator.

388

389 **Discussion**

390 **Roles of Zld and Opa in shaping chromatin accessibility and gene expression in Clogmia.**

391 Dipteran embryos are morphologically similar (13, 105), but their mechanisms of axis
392 specification (14, 15), primordial germ cell specification (15, 106), extraembryonic tissue
393 specification (107-109), morphogenesis (108, 110, 111), and sex determination (112-115) are not
394 conserved across the dipteran clade. This clade is therefore an attractive model taxon to study
395 how molecular mechanisms of embryogenesis evolve. The segmentation process of flies is
396 particularly suitable for studying the evolution of developmental TF gene networks because it
397 unfolds before the completion of cellularization of the multinucleated embryo when gene
398 expression can be readily visualized and perturbed in non-traditional model organisms. In this
399 study, we have established the resources for dissecting the developmental gene network of early
400 Clogmia embryos by providing an annotated chromosome-level genome sequence for our inbred
401 laboratory strain and by describing chromatin accessibility of precisely staged consecutive
402 blastoderm stages. We have used these resources to examine how *zelda* and maternal *odd-paired*
403 activities shape chromatin accessibility and gene expression in Clogmia.

404 We find that, as in Drosophila, zygotic genome activation in Clogmia is accompanied by large-
405 scale changes in chromatin accessibility state. We demonstrate that Cal-Zld is necessary for
406 activating an early wave of zygotic genes and that reduced *Cal-zld* expression results in
407 widespread losses of chromatin accessibility at target genes, consistent with a conserved role for
408 Zld to pioneer chromatin accessibility within the context of zygotic genome activation. However,
409 we also observe differences compared with Drosophila in the early patterns of establishment and
410 maintenance of accessible states, notably losses in accessibility during the cleavage divisions.
411 Further, we show that a subset of these regions depends on Zld to lose accessibility during
412 cleavage stages. Clogmia's anterior determinant, Cal-Opa, likewise has the ability to influence
413 chromatin accessibility states, and loss of Cal-Opa also results in reduced chromatin
414 accessibility, albeit at a smaller cohort of regions compared with Cal-Zld. These Opa-dependent
415 regions however include zygotic loci whose functions we further hypothesize are necessary for
416 mediating anterior specification downstream of Opa. Collectively, these results are consistent
417 with a model where anterior determination in Drosophila and Clogmia, while driven by unrelated
418 transcription factors, share a common feature of symmetry breaking at the level of cis-regulatory
419 sequence accessibility driven by the anterior determinant. Mechanistically, this could occur
420 through either a pioneering activity which requires binding of nucleosome occupied DNA or
421 through competition with nucleosomes to bind DNA. Bcd does not pioneer chromatin as
422 observed with Zld and instead outcompete nucleosomes above a concentration threshold (61,
423 84). In contrast, a vertebrate homolog of Opa, Zic3, was found to bind nucleosomal DNA *in*
424 *vivo*, suggesting that Cal-Opa may pioneer chromatin directly (116).

425 However, unlike in Drosophila, where only gains in accessibility are observed during the
426 cleavage divisions (62), in Clogmia we observe a substantial cohort of regions that lose
427 accessibility over the final cleavage divisions. These regions likely represent sites that gain
428 accessibility prior to NC11 and thereafter lose accessibility. In Drosophila, losses in chromatin
429 accessibility are only observed following the lengthening of the cell cycle at NC14, which
430 coincides with the start of lineage restriction and cell fate specification (58). These programmatic
431 losses in accessibility are presumed to be driven by the activity of repressive factors, either
432 operating globally in all cells of the embryo, or conditionally as part of developmental programs
433 for cell fate specification. The period during which a region is accessible is also presumed to be
434 the period in which the region is functional and positively contributing to gene regulation. It is
435 surprising to observe in Clogmia losses in accessibility during the syncytial cleavage divisions
436 since this is thought to precede large scale zygotic genome activation and cell fate specification.

437 These regions could potentially control important regulatory events prior to large scale zygotic
438 genome activation. In Drosophila, Zld almost exclusively drives gains in chromatin accessibility
439 (56, 58, 87, 117) through direct binding to DNA motifs independent of nucleosome association,
440 consistent with its role as a pioneer factor (85). In contrast, we noticed that in Clogmia a subset
441 of the regions that lose accessibility during the syncytial cleavage divisions do so in a Zld-
442 dependent manner. Based on the prevalence of Zld motifs within this class of regions, these
443 effects are likely direct. At present, we cannot distinguish between a mechanism where Cal-Zld
444 functions directly as a context-specific repressor or indirectly by permitting the binding of an
445 independent repressive factor. Further work, including confirming the direct targets of Cal-Zld,
446 and identification of cooperating repressors, will be needed to answer this question. It is worth
447 pointing out that the accessibility dynamics that we observe in Clogmia may reflect the
448 consequences of development with longer cell cycle timing. In Clogmia, NC11 through NC13
449 take ~3x longer than in Drosophila, where NC11 only lasts ~10 minutes, NC12 ~12 minutes, and
450 NC13 ~21 minutes (69), and the much shorter cell cycle times of Drosophila present significant
451 limitations to stable establishment and maintenance of chromatin states (62). Assuming that
452 kinetic rates for nuclear import, DNA replication, and DNA binding of transcription factors is
453 similar between Clogmia and Drosophila, the significantly longer syncytial interphase times in
454 Clogmia could alleviate the impact of cell cycle dependent constraints on transcriptional
455 regulation. Indeed, experimental lengthening of the Drosophila syncytial interphase can result in
456 premature transcriptional activation (118).

457 **Functional comparison of the ADs in Clogmia and Drosophila**

458 We find that homologs of *homeobrain* and *sloppy-paired* are targets of Cal-Opa to initiate
459 anterior pattern formation. These genes may play key roles in initiating anterior pattern
460 formation not just in Clogmia but also in other insects. In Drosophila, *homeobrain* is required for
461 labrum and brain development (119-121) and associated with enhancers that open in a Bcd-
462 dependent manner (56), and in the beetle *Tribolium castaneum*, *homeobrain* was found to play
463 an important zygotic role in establishing the embryo's anterior-posterior polarity (122). The
464 *sloppy-paired* gene, a member of the *forkhead* domain gene family (123), is best known for its
465 role as pair-rule segmentation gene (9, 124), but it is first expressed in a head domain of the
466 syncytial Drosophila embryo, like its closest paralogs, *sloppy-paired 2* and *fd19B* (125-128).
467 Partially redundant functions of sloppy-paired paralogs, which are common across insects (**S14**
468 **Figure and S15 Figure**), could have obscured a fundamental role of *sloppy-paired* activity in
469 head development.

470 An important limitation of our study is that we cannot rule out that maternal Cal-Opa targets
471 additional genes via constitutively accessible cis-regulatory DNA. Testing this possibility will
472 require new genetic tools or reagents to assess where maternal Cal-Opa binds to genomic DNA
473 *in vivo*. In the meantime, we provide here a preliminary discussion of how targets of Cal-Opa
474 and Bcd compare. Bcd contributes to the activation of many early segmentation genes but its role
475 in the posterior embryo is much more subtle than in the anterior embryo, where all expression
476 domains directly or indirectly critically depend on Bcd. Target enhancers of Bcd have been
477 classified based on how their accessibility depends on Bcd and Zld (56). The accessibility of
478 enhancers targeted by Bcd that drive expression in the posterior region of the embryo is
479 independent of Bcd (e.g., *Krüppel* CD1, *knirps* posterior, *giant* posterior). However, multiple
480 enhancers that open in a Bcd-dependent manner were found to be associated with genes that are
481 expressed in the anterior region of the embryo (e.g., *orthodenticle*, *empty spiracles*, *buttonhead*,

482 *giant*, *knirps*, *hunchback*, *sloppy-paired 1*, *fd19B*, *even-skipped*, *paired*, *cap'n'collar*, *huckebein,*
483 *homeobrain*), and many of them correspond to enhancers that are known to drive expression of a
484 reporter gene in the anterior embryo (e.g., *giant* anterior (-6), *knirps* anterior, *orthodenticle*,
485 *buttonhead, even-skipped* stripe 1/5, *hunchback* P2, *hunchback* shadow enhancers, *eve-skipped*
486 stripe 2*,* an early *paired* enhancer, *giant* anterior (-10), *cap'n'collar*, *huckebein;* reviewed in
487 (56).

488 In Clogmia, the expression of both *hunchback* orthologs (*Cal-hb1*, *Cal-hb2*) was reduced in *Cal-*
489 *opa* RNAi embryos along with associated ATAC-seq peaks but these regions did not include
490 Opa motifs (**S16A Figure and S16B Figure**). One of Clogmia's two orthologs of *orthodenticle*
491 (*Cal-opa1*) was not associated with Opa-motif containing *Cal-opa* sensitive peaks and its
492 expression level was not significantly reduced in *Cal-opa* RNAi embryos (**S16C Figure**). The
493 second ortholog (*Cal-opa2*) was not expressed at NC12 and NC13. In the case of Clogmia
494 *buttonhead* (*Cal-btd*), we identified a *Cal-opa* sensitive peak with an Opa motif but the observed
495 reduction of *Cal-btd* expression was not significant (**S16D Figure**). Both *Cal-otd1* and *Cal-btd*
496 expression could be under the control of AD-dependent and AD independent enhancers, as
497 previously reported for *orthodenticle* in Drosophila (129). Clogmia orthologs of *giant* (*Cal-gt*),
498 *knirps* (*Cal-kni/knrl1, Cal-kni/knrl2*), and *empty spiracles* (*Cal-ems1-4*) were not expressed at
499 NC12 or NC13 and no significant changes in chromatin accessibility were detected at these loci
500 at NC12 or NC13 following knockdown of *Cal-opa*. The ortholog of *huckebein* (*Cal-hkb*) was
501 not expressed at NC12 or NC13, either; however, its locus contained one Opa motif positive
502 ATAC-seq peak that lost accessibility following *Cal-opa* knockdown at NC13. Clogmia's
503 *cap'an'collar* (*Cal-cnc*) and *paired* (*Cal-prd*) loci were expressed but their expression was
504 unaffected by *Cal-opa* RNAi, and the ATAC-seq peaks at their loci were *Cal-opa* insensitive.
505 The loci of *even-skipped* orthologs (*Cal-eve1, Cal-eve2*) contain ATAC-seq peaks with Opa
506 motifs, however, all the ATAC-seq peaks associated with these loci were insensitive to *Cal-opa*
507 RNAi, and the expression of *Cal-eve2* was slightly increased at NC13 in *Cal-opa* RNAi
508 embryos. Taken together, these observations point to significant differences in the key targets of
509 Cal-Opa in Clogmia and Bcd in Drosophila and raise the question of how the downstream
510 segmentation gene networks of these species converge.

511

512 **Materials and Methods**

513 **Clogmia culture and embryo collection**

514 The Clogmia culture was maintained in 100mm x15mm Petri Dishes (Fisher Scientific
515 FB0875712) on moist, compressed cotton (Jorvet SKU: J0197) sprinkled with parsley leaf
516 powder (Starwest Botanicals SKU: 209895-5). For embryo collections, ~2-day old females were
517 transferred to round 250 mL glass bottles with a ~1 inch bottom layer of moistened and
518 compressed cotton and mated at room temperature for 2-3 days. Female ovaries were dissected
519 with a pair of fine forceps (Dumont 5) and eggs released into water for egg activation and
520 embryogenesis. For staged egg collections, embryos were released into flat bottom glass 100mm
521 x15mm Petri Dishes (Corning CM: 101425255) containing 15ml of water. Care was taken to
522 disperse the eggs on the bottom of the dish as crowding can delay development. Petri Dishes
523 were covered with a lid and moved to an incubator set at 25° C until the desired stage. Egg
524 harvesting post activation was on average 47 minutes for NC2/3, 261 minutes for NC11, 292

525 minutes for NC12, 332 minutes for NC13, 401 minutes for NC14, and 465 minutes for Late
526 NC14.

527

528 **Embryo fixation**

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

539

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

541 *Clogmia albipunctata* embryo injection was done as previously described (Yoon et al. 2019). We
542 note here that RNAi efficiency in Clogmia embryos varies (15). This variability could be due to
543 technical variables (e.g., injection needle, amount injected) or biological variables (e.g.,
544 differences in the RNAi response between individual embryos, or the general health or genetic
545 background of the embryos or their mothers) or a combination of such effects. Therefore, we
546 took care to control the age of the females and selected all experimental embryos from females
547 with healthy looking ovaries and eggs. To assess maternal *Cal-opa* RNAi efficiency, we injected
548 embryos in the first hour of development with dsRNA of the isoform-specific first exon of the
549 maternal *Cal-opa* isoform. In three independently conducted experiments we reproduced the
550 double abdomen phenotype, confirming that our knockdown reagent was satisfactory. We also
551 quantified knockdown efficiencies of *Cal-opa* transcript in individual embryos by RT-qPCR and
552 performed FC calculations of *Cal-opa* mRNA levels with the average level in alignment controls
553 as reference (**S5 Table**). This analysis suggests that *Cal-opa* RNAi can nearly abolish the *Cal-*
554 *opa* transcript.

555

556 **dsRNA synthesis**

557 A PCR template was made using primers specific for the target gene with T7 overhangs and
558 either cDNA, gDNA or a plasmid depending on the target sequence. The PCR product was run
559 on a gel and purified using the Zymo Gel DNA Recovery Kit (D4007/D4008) following the
560 manufactures instructions. The PCR template was used as the input for the transcription reaction
561 producing dsRNA. dsRNA synthesis and purification were done using the Invitrogen
562 MEGAscript RNAi Kit (AM1626) following the manufacturer's instructions. Purified dsRNA
563 was eluted in injection buffer (0.1 mM NaH_2PO_4 Ph 6.8; 5 mM KCl).

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

566 *Cal-opa,* maternal transcript (201 bp):
567 5'-CAGAGATGCATAATACGACTCACTATAGGGAGAAAACAATTGTGAAGTGCGACA
568 5'-
569 CAGAGATGCATAATACGACTCACTATAGGGAGACAAATTTCCAAACGATGACAGA

570 *Cal-zld* (a combination of two dsRNAs were used 579 bp and 668 bp):
571 5'-TAATACGACTCACTATAGGGAGAAGTCCCGCAATTGATACAGC
572 5'-TAATACGACTCACTATAGGGAGAAGGATGTTGGTGGACACTCC
573 5'-TAATACGACTCACTATAGGGAGACGGAATGGTGTGTGAAACAG
574 5'-TAATACGACTCACTATAGGGAGATTCGAGGGGTTATTGTCCTG

575 dsRed (273 bp):
576 5'-CAGAGATGCATAATACGACTCACTATAGATGCAGAAGAAGACTATGG
577 5'- CAGAGATGCATAATACGACTCACTATAGCTACAGGAACAGGTGGTG

578

579 **RT-qPCR**

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

582 Forward and reverse primer sequences for targets (lengths of amplicon in brackets):
583 *Cal-opa*, maternal transcript (112 bp):
584 5'- TTGCGCTTCATTCTGTCATCG
585 5'- CACTGAGGGCTGATTCTCGG
586
587 *Cal-zld* (138 bp):
588 5'-AAAGATGAACCACCGTGCGA
589 5'-CATTGATGGCAGTGGACCCT
590
591 *Cal-rpl35* (119 bp):
592 5'- CTGTCCAAAATCCGAGTCGT
593 5'- AGGTCGAGGGGCTTGTATTT
594

595 **Fluorescent HCR *in situ* hybridizations**

596 Gene specific probes and amplifier-fluorophore sets were designed and synthesized by Molecular
597 Instruments Inc (Los Angeles, CA). The amplifier-fluorophore set used was B1-647 for *Cal-slp1*,
598 B2-546 for *Cal-hbn*, and B3-488 for *Cal-slp2*. Buffers (probe hybridization, probe wash and
599 amplification buffers) throughout the protocol were purchased from Molecular Instruments Inc.
600 The HCR *in situ* hybridizations were performed as described (130, 131). Briefly, fixed embryos
601 were rehydrated and washed 3 times in PBS-Tween20 (1x PBS, 0.1% Tween-20). Embryos were
602 permeabilized for 30 minutes at room temperature (100 ml Permeabilization buffer: 100 µL
603 100% Triton X-100, 50 µL Igepal CA-630, 50 mg Sodium Deoxycholate (Sigma D6750-25G),
604 50 mg Saponin (Sigma 47036-50G-F), 200 mg BSA Fraction V (Sigma A5611-5G). Following
605 embryos were pre-hybridized in probe hybridization buffer for 30 minutes in a 37° C water bath.
606 Once pre-hybridization was complete, fresh hybridization solution containing gene specific
607 probes were added and the embryos were left to incubate overnight in a 37° C water bath. On the

608 second day the hybridization solution was removed, and the embryos were washed 4 times with
609 pre-warmed probe wash buffer followed by two washes in 5x SSCT (5x SSC, 0.1% Tween-20).
610 An amplifier hairpin solution was prepared following the manufacturer's instructions. Embryos
611 were incubated with pre-warmed amplification buffer for 30 minutes at room temperature before
612 adding amplifier hairpin solution. Embryos were left to incubate overnight at room temperature
613 in the dark. The following day the embryos were washed 5 times in 5x SSCT at room
614 temperature. The embryos were then counter stained with DAPI and mounted in Aqua-
615 Polymount for imaging. Confocal imaging was performed on the Zeiss LSM 900.

616

617 **Image Acquisition and Processing**

618 Confocal imaging was performed on the Zeiss LSM 900 using a Plan-Apochromat 20x/0.8 M27
619 objective. Pixel-density of 1024 x 1024. Excitation/Emission wavelengths (in nanometers) were:
620 493/517 for Alexa Flour 488, 653/668 for Alexa Flour 647, 401/422 for DAPI, and 548/561 for
621 Alexa Flour 546. For each image a Z-stack of images was captured from the embryo's surface to
622 approximately midway through the embryo. Then a max projection was created from this Z-stack
623 using FIJI/Image J. Brightness was increased in Photoshop. No further adjustments to the images
624 were made.

625

626 **ATAC-seq library preparation**

627 Single embryo ATAC-seq library preparations were done essential as described previously in
628 *Drosophila melanogaster* (58, 62). Embryos developed in an incubator at 25ºC until the desired
629 nuclear cycle. Nuclear cycle was confirmed by time after egg activation and the morphology was
630 checked under water using a transmission light microscope (Zeiss Stemi 2000-C). Once at the
631 desired nuclear cycle individual embryos were placed in the lid of a 1.5 ml low-retention
632 Eppendorf tube and homogenized in 10 µl of cold lysis buffer (10 mM Tris-HCl pH 7.4; 10 mM
633 NaCl; 3 mM MgCl2; 1% Igepal CA-630) (132), using a fire-sealed microcapillary tube
634 (Drummond Microcap 25). Once homogenized, 40 µl of lysis buffer was added to the lid, and the
635 body of the tube was carefully closed over the lid. Samples were place on ice until all embryos
636 were processed. The nuclei were pelleted by centrifugation at 800 RCF at 4ºC for 10 minutes.
637 Supernatant removal was observed under a dissection scope to ensure that the nuclei pellet was
638 not dislodged. The nuclear pellet was resuspended in 7.5 µl of Tagmentation buffer (Illumina
639 Tagment DNA TDE1 Enzyme and Buffer Kit) and placed on ice until all samples were
640 resuspended. Then 2.5 µl of Tn5 Transposase was added and mixed by pipetting up and down.
641 Tagmentation and amplification of ATAC-seq libraries were performed as described previously
642 (Buenrostro et al., 2015). The tagmentation reaction was performed at 37ºC and 800 RPM for 30
643 minutes using an Eppendorf Thermomixer. Immediately after this step, the reactions were
644 purified using the Qiagen Minelute kit following the manufacturer's instructions. Purified
645 tagmented DNA was eluted in 10 µl of buffer EB. Library amplification was performed as
646 described previously (Buenrostro et al., 2015). Typically, 12-14 total PCR were performed.
647 Amplified libraries were purified using 1.8x Ampure SpRI beads following the manufacturer's
648 instructions. Purified ATAC-seq libraries were eluted in 15 ul. The ATAC-seq library profile was
649 evaluated using Agilent high sensitivity bio-analyzer. Concentrations were estimated using a
650 Qubit fluorometer. Libraries from the same experiment were pooled and sequenced together.
651 Sequencing was done at the University of Chicago Genomics Core (Chicago, IL, USA) using the

652 Illumina NextSeq (PE 75 bp) and at Admera Health (South Plainfield, NJ, USA) using the
653 Illumina HiSeq (PE 150 bp) and the Illumina NovaSeq X Plus (PE 150 bp).

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

657 ATAC UDI Index Read 2 (i5):
658 5'- AATGATACGGCGACCACCGAGATCTACACnnnnnnnnTCGTCGGCAGCGTCAGATGT*
659 G -3'

660 ATAC UDI Index Read 1 (i7):
661 5'- CAAGCAGAAGACGGCATACGAGATnnnnnnnnGTCTCGTGGGCTCGGAGATG*T -3'

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

664

665 **RNA isolation for RT-qPCR and RNA-seq experiments**

666 Embryos developed in an incubator at 25ºC until the desired nuclear cycle. Nuclear cycle was
667 confirmed by time after egg activation and the morphology was checked under water using a
668 transmission light microscope (Zeiss Stemi 2000-C). RNA isolation was performed using the
669 Zymo Quick-RNA Tissue/Insect Microprep Kit following the manufacturer's instructions except
670 for the tissue homogenization step. Instead, homogenization was done as follows. Once at the
671 desired nuclear cycle individual embryos were placed in the lid of a 1.5 ml low-retention
672 Eppendorf tube and homogenized in 10 µl of cold lysis buffer (10 mM Tris-HCl pH 7.4; 10 mM
673 NaCl; 3 mM MgCl2; 1% Igepal CA-630) (132), using a fire-sealed microcapillary tube
674 (Drummond Microcap 25). Once homogenized the lysate was added to 450 µl of Zymo RNA
675 Lysis Buffer and thoroughly mixed. RNA isolation resumed following the manufacturer's
676 instructions including DNAse treatment. Purified RNA was eluted in DNase/RNase-Free Water.

677

678 **RNA-seq library preparation**

679 RNA quality and quantity was assessed using the Agilent bio-analyzer. Non-Strand-specific
680 RNA-seq libraries were prepared using the Smarter v4 Ultra-Low Input protocol from Takara for
681 cDNA synthesis and Nextera XT from Illumina for Library prep (protocols provided by Takara
682 and Illumina). Library quality and quantity was assessed using the Agilent bio-analyzer. Libraries
683 from the same experiment were pooled and sequenced together. Sequencing was done at the
684 University of Chicago Genomics Core (Chicago, IL, USA) using the Illumina NovaSeq X (PE
685 150 bp) and at Admera Health (South Plainfield, NJ, USA) using the Illumina NovaSeq X Plus
686 (PE 150bp).

687

688 **ATAC-seq library preparation and RNA isolation**

689 The first steps are the same as the standard ATAC-seq protocol above. Individual embryos are
690 homogenized in cold lysis buffer, and a centrifugation step is performed to pellet the nuclei.
691 Following the supernatant (~47 ul) is removed and mixed thoroughly with 450 µl Zymo RNA
692 lysis buffer in a fresh low-bind tube. The nuclear pellet is resuspended in 7.5 µl Tag mentation

693 buffer and the ATAC-seq protocol continues as above with no deviations. RNA isolation
694 proceeded immediately, or the RNA lysate was stored at temperately at -20ºC. RNA isolation
695 was performed using the Zymo Quick-RNA Tissue/Insect Microprep Kit following the
696 manufacturer's instructions. Isolated RNA was for RT-qPCR and RNA-seq experiments.

697

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

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

705

706 **ATAC-seq data processing**

707 Demultiplexed reads were trimmed of adapters using TrimGalore! (133) (version 0.6.10,
708 cutadapt version 4.4) and mapped to the *Clogmia albipunctata* assembly (calbi-uni4000) using
709 Bowtie2 (134) (version 2.4.1) with option -X 2000. Suspected optical and PCR duplicates were
710 marked by Picard MarkDuplicates (version 2.21.4) using default paraments
711 (https://broadinstitute.github.io/picard/). Mapped, trimmed, duplicate marked reads were
712 imported into R using the GenomicAlignments (135) (version 1.38.2) and Rsamtools (136)
713 (version 2.18.0). Reads that mapped, were properly paired, non-secondary and had a map quality
714 scores ≥ 10 were imported. Only reads mapping to scaffolds 1-5 and scaffold 7 were used for
715 downstream analysis. The start and end coordinates of the reads were adjusted to account for Tn5
716 interaction site. Watson strand start sites had four base pairs subtracted, and Crick-strand start
717 sites had five base pairs subtracted (132). Reads with mapped length ≤ 120 bp were considered
718 to have originated from 'open' chromatin.

719 *MACS2*

720 Accessible chromatin peaks from NC11 through NC14 were determined using MACS2 (78)
721 (version 2.2.7.1) with options -f BEDPE -q 1e-25 on a merged dataset comprising all ATAC
722 reads corresponding to open chromatin (length ≤ 120 bp). Peaks were assigned to a genomic
723 feature on the basis of overlapping with one or more feature of a gene (e.g. exon, intron,
724 promoter). Peaks that overlapped with no gene feature were classified as intergenic.

725 *DESeq2*

726 A count matrix for differential enrichment analyses using DESeq2 (82) (version 1.42.1) was
727 generated by counting the number of open ATAC reads that overlapped with a MAC2 identified
728 peak. For the WT stage-wise comparisons (e.g. NC12 vs NC11, NC13 vs NC12, NC14 vs NC13
729 and LNC14 vs NC14) the design parameter passed to DESeq2 was ~stage. For the comparison
730 between RNAi embryos and alignment control embryos (e.g. *Cal-zld* RNAi vs alignment control
731 and *Cal-opa* RNAi vs alignment control) the design component was ~genotype. The criteria for a
732 statistically significance change was a $\log_2$ foldchange ≥ 1 and an adjusted p-value ≤ 0.05 or a
733 $\log_2$foldchange ≤ -1 and an adjusted p-value ≤ 0.05.

Dynamic peaks were defined as a peak that underwent a statistically significant change in accessibility in at least one of the WT stage wise comparisons. Different dynamic peak classes were found using the R package DEGreport (137) (R package version 1.38.5) with the parameter of the minimum groups size set to 20.

MEME

MEME (91) (version 5.4.1) analysis was performed using the following options -order 1 -meme-mod zoops -maxw 14 -meme-nmotifs 15 -meme-searchsize 100000 -centrimo-score 5.0 -centrimo-ethresh 10.0. MEME was given a set of sequencings containing the 100 bp sequence centered on the summit of the peak. A set of background sequences were also given to MEME. To determine motif enriched in dynamic peaks a set of non-dynamic peak sequences were provided as background sequences. To determine motifs enriched in *Cal-zld* sensitive peaks a set of sequences from peaks unaffected by *Cal-zld* were provided as background sequences. To determine motifs enriched in *Cal-opa* sensitive peaks a set of sequences from peaks unaffected by *Cal-opa* were provided as background sequences. Identified motifs were compared to known *Drosophila melanogaster* motifs from the following databases: OnTheFly_2014, Fly Factor Survey, FLYREG, iDMMPMM, and DMMPMM databases (99, 138-141).

RNA-seq data processing

Demultiplexed reads were trimmed of adapters using TrimGalore! (133) (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 (version 2.4.1) with default parameters. Suspected optical and PCR duplicates were marked by Picard MarkDuplicates (version 2.21.4) using default paraments. Mapped, trimmed, duplicate marked reads were imported into R using the GenomicAlignments (135) (version 1.38.2) and Rsamtools (136) (version 2.18.0). Reads that mapped, were properly paired, non-secondary 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 a transcript of a gene. For the comparison between RNAi embryos and alignment control embryos (e.g. *Cal-zld* RNAi vs alignment control and *Cal-opa* RNAi vs alignment control) the design component was ~genotype. For the comparison between *Cal-opa* RNAi vs injection control the design component was ~genotype. The criteria for a statistically significance change was a $\log_2$ foldchange ≥ 1 and an adjusted p-value ≤ 0.05 or a $\log_2$foldchange ≤ -1 and an adjusted p-value ≤ 0.05.

Maternal, zygotic, maternal and zygotic classification

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

Fold change calculations

Fold change was calculated as the sample's counts per million divided by the control's counts per million for the gene of interest. First a counts per million (CPM) matrix was made by counting the number of RNA-seq reads that overlapped with a gene's transcript and then normalizing to counts per million. Fold change calculations were done with embryos grouped by experimental date (i.e. embryos from the same egg packet). For each embryo the CPM was divided by the average CPM of wild-type and alignment control embryos. If a date did not have any wild-type or alignment control embryos fold change was calculated using the average CPM of all wild-type and alignment control embryos in the data set. For *Cal-zld* RNAi experiments a fold change (FC) calculation of *Cal-zld* mRNA levels was conducted with the average of control embryos (untreated wild-type and alignment controls) from the same experimental date as reference. *Cal-zld* RNAi embryos with a FC < 0.25 (knockdown efficiency greater than 75%) were used for further analysis. Similarly, for *Cal-opa* RNAi experiments FC calculations were done separately for each stage. The fold change (FC) calculation of *Cal-opa* mRNA levels was conducted with the average of control embryos (untreated wild-type and alignment controls) from the same experimental date as reference. *Cal-opa* RNAi embryos with a FC < 0.25 (knockdown efficiency greater than 75%) were used for further analysis.

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 (142). 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 (74). 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* (143) (version v2.0.4) to identify repeat elements in the *Clogmia albipunctata* genome and *RepeatMasker*'s (144) (version v4.1.5) *fambd.py* script to extract *Arthropoda* records from the *Dfam* database. These sequences were combined to generate a custom repeat library, which was then used to soft-mask the genome. To predict gene structures, we used RNA-seq and protein evidence. mRNA data from 26 individual embryos of various stages, pooled samples of first and fourth-instar larvae, one female and one male pupa generated in this study were mapped to the genome (**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 (74). 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* (145) (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* (146, 147) (version 2.1.12), assigning gene

names based on orthology. We assessed the final annotation quality using BUSCO (Benchmarking Universal Single-Copy Orthologs) (148, 149) on the predicted protein sequences. Finally, we analyzed synteny between the *C. albipunctata* and *D. melanogaster* genomes using *MCScanX* (150) (primary release) with parameters: match_score = 50 (default), match size = 20 (minimum genes per syntenic block), gap_penalty = -1 (default), and max_gaps = 100. Results were visualized using the *circlize* R package (151) (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.

Transcription factor identification

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

Acknowledgments

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

Competing interests

The authors declare no competing interests.

Funding

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

Data and Resource Availability

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

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

Genome Browser: The genomic data produced in this study can be further interrogated using a genome browser ecosystem centered on JBrowse2 (152) and delivered as a cloud image designed for individual use. Currently available on NSF's Jetstream2 cloud platform, with a portable Docker image in development. Jetstream2's user-friendly interface (image: *Clogmia albipunctata* Genome Resources) (153, 154) integrates tools for genomic analysis, including BLAST search (SequenceServer2.0) (155), CRISPR guide RNA design (modified crisprDesigner) (156), R shiny-driven differential gene expression (DGE) analysis (freecount) (157), and synteny mapping for comparative genomics (ShinySyn) (158). Each tool is connected back to the browser, leveraging JBrowse2's ability to dynamically visualize and contextualize the genomic data. BLAST results link directly to genomic regions, enabling users to explore expression tracks, variants, and synteny relationships. CRISPR tools provide sgRNA design with metrics for selection and visualization of target regions alongside expression data, off-target effects, and variants. DGE analysis visualizations are linked back to the genome browser for further exploration. All gene model annotations in the browser connect to common external databases, such as NCBI, Uniprot, FlyBase, and Google Scholar, for further exploring and interpreting the data. Future enhancements will include expanded sgRNA profiling, support for Docker to enable deployment on commercial cloud platforms, and further optimization of workflows, ensuring this ecosystem remains a powerful and accessible solution for the use of our genomic data presented here and elsewhere (74) for research and education. To access the *Clogmia albipunctata* genome browser and related tools, create an ACCESS ID and use this ID create an ACCESS account to log in to Jetstream2 to apply for an 'allocation' of credits that can be used on Jetstream2. Detailed instructions on the use of Jetstream2 can be found at https://jetstream-cloud.org/get-started/index.html. This on-demand model reduces resource costs and security concerns compared to hosting a centralized site. Administrative tools, (e.g. automated reboot scripts, system service management, and load balancing for multi-user groups or workshops) enhance usability while minimizing operational overhead.

893 **Tables**

894

895 **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%
Eukaroyotic BUSCO's recovered (%)	98.04%

896

897

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

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

902

903

904

Table 3: **Comparison of chromatin accessibility peaks, nuclear cycle length in minutes, and haploid genome size of *Clogmia albipunctata* and *Drosophila melanogaster***. Peaks for *Drosophila melanogaster* are based on (56, 62) and nuclear cycle lengths are based on (69).

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

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

911 **Figures**
912

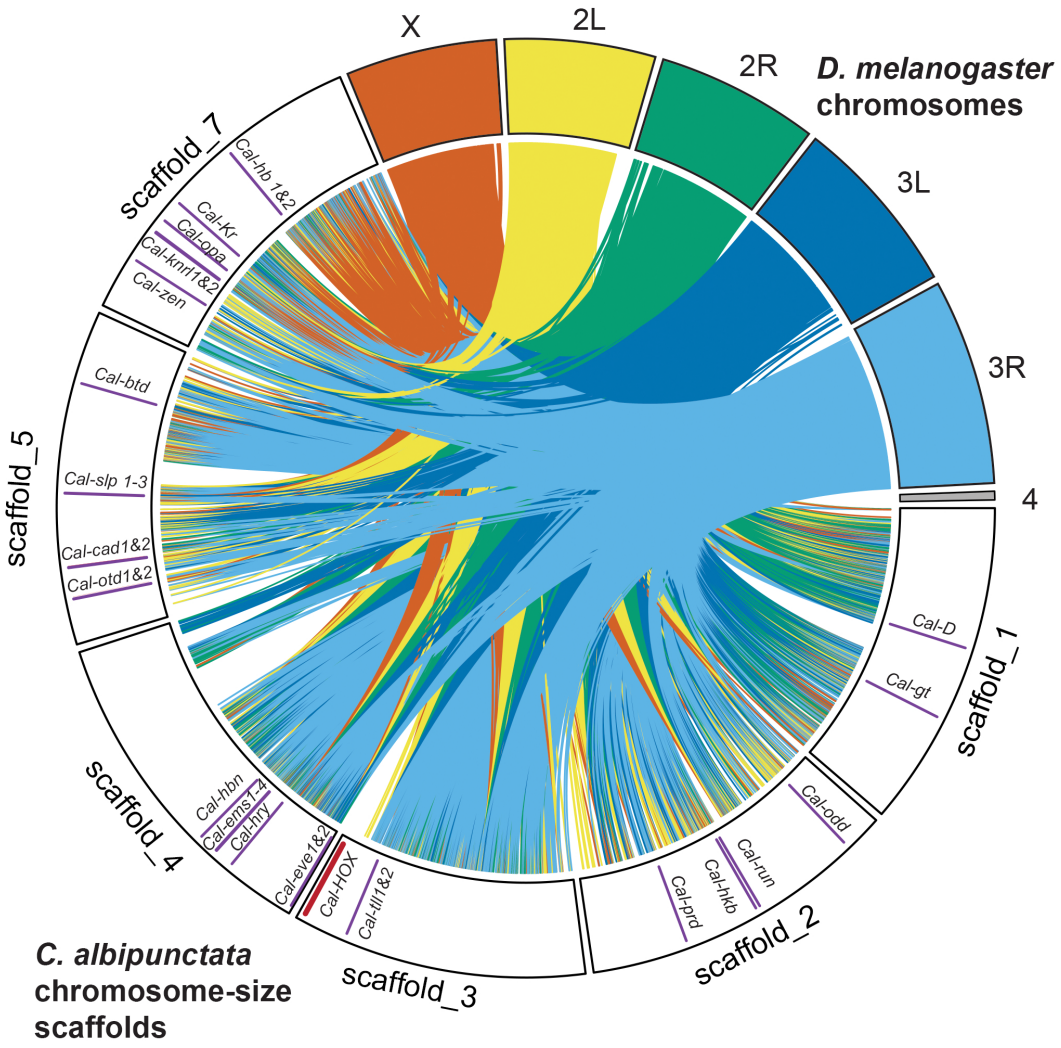

Figure 1: Synteny analysis of *Drosophila melanogaster* chromosomes and *Clogmia albipunctata* scaffolds. Synteny analysis between *Clogmia albipunctata* chromosomes-sized scaffolds *and Drosophila melanogaster* chromosomes. Groups of collinear genes are represented by contact lines connecting the positions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

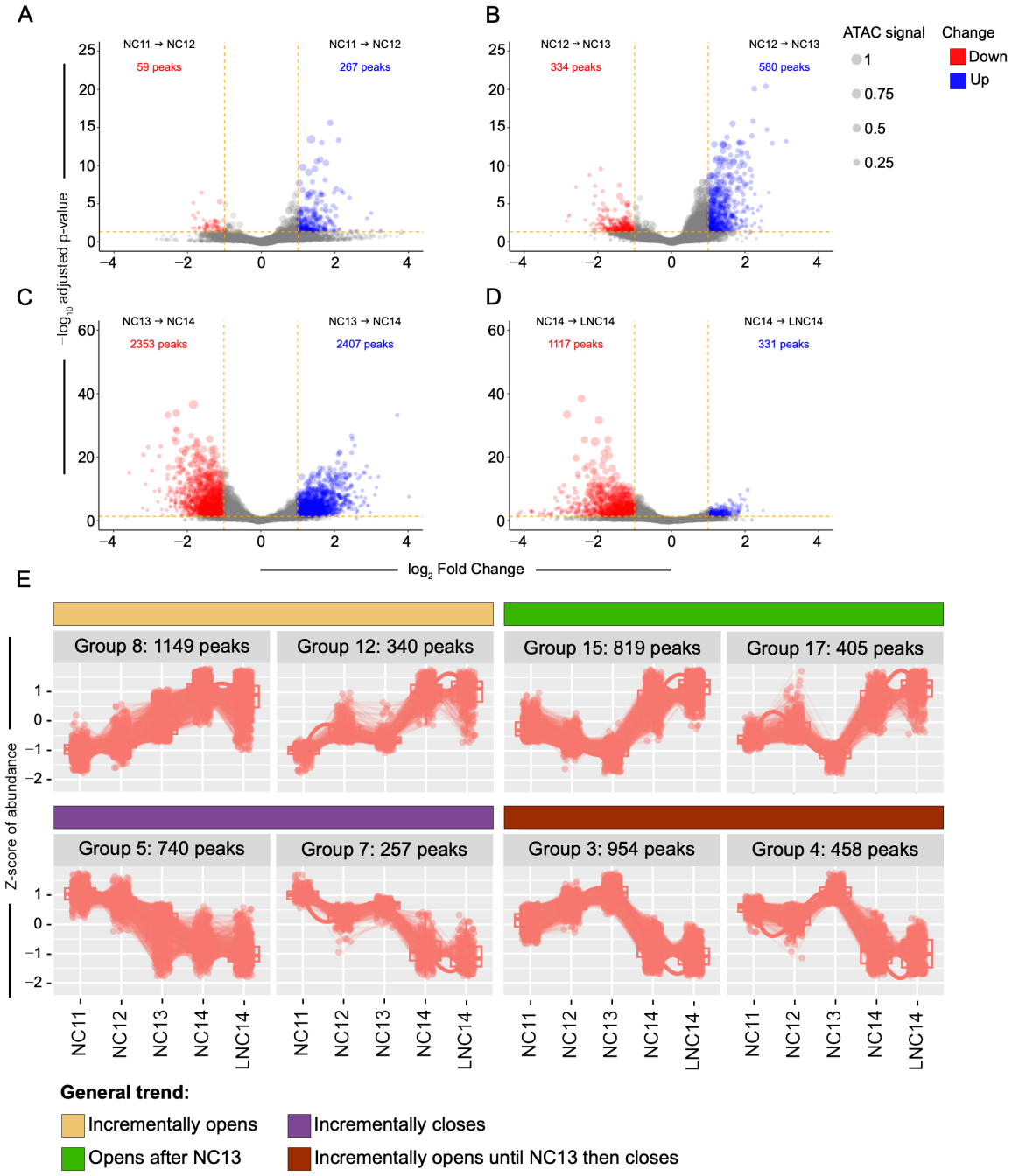

Figure 2: **Changes in chromatin accessibility during NC11 to Late NC14.** A-D) Volcano plots of differentially accessible ATAC-seq peaks between consecutive developmental stages. Significant change of p-adjusted ≤ 0.05 and |log2FC| ≥ 1 highlighted; red (log2FC ≤ -1) and blue (log2FC ≥ 1). E) Dynamic peak groups produced using DEGreport (137) (R package version 1.38.5). Y-axis, Z-score of abundance. X-axis, stage.

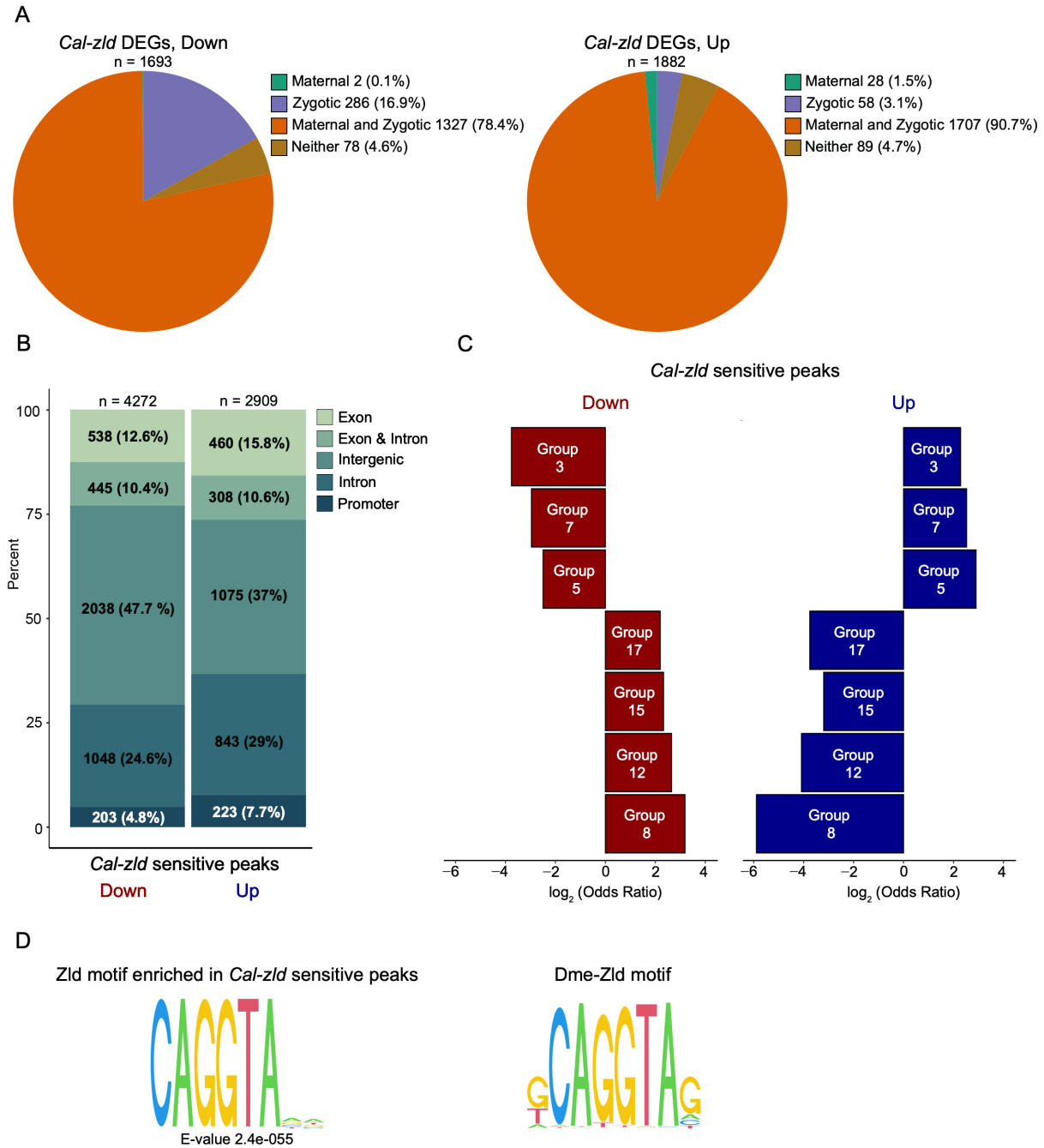

Figure 3: Cal-Zld regulates chromatin accessibility and zygotic transcription positively and negatively. A) Percentage of down-regulated (left) and up-regulated (right) genes in *Cal-zld* RNAi embryos classified as maternal, zygotic, maternal and zygotic or with little to no expression (neither) in wild type. Classifications were determined by expression analysis of genes at NC2/3 (~45-minute old embryos) or at syncytial blastoderm stages (NC11 through Late NC14, see *Materials and Methods*). B) Genomic distribution of *Cal-zld* sensitive peaks that were up-regulated (blue) and down-regulated (red). C) Dynamic groups whose accessibility decreases over time are more likely to be *Cal-zld* sensitive peaks that gain accessibility (Group 3 p-val, 5.7e-84; Group 5 p-val, 3.6e-120; Group 7 p-val, 2.6e-32) unlike dynamic groups whose accessibility increases over time (Group 8, p-val 7.0e-45; Group 12, p-val 1.2e-11; Group 15 p-

val 8.3e-23; Group 17; 3.5e-13). Dynamic groups whose accessibility increases over time are more likely to be *Cal-zld* sensitive peaks that lose accessibility (Group 8, p-val 1.8e-257; Group 12, p-val 2.9e-55; Group 15 p-val 7.7e-94; Group 17;1.1e-42) unlike dynamic groups whose accessibility decrease over time (Group 3 p-val, 3.7e-44; Group 5 p-val, 6.3e-24; Group 7 p-val, 2.1e-10). Odds ratios and p-values were calculated by a two-sided Fisher's exact test on contingency tables constructed by the presence/absence of dynamic groups in *Cal-zld* sensitive peaks up or down. D) Position weight matrix (PWM) logo of the Zld motif enriched in *Cal-zld* sensitive peaks as identified by MEME (left) and the canonical Zld motif in *Drosophila melanogaster*, shown for reference (right).

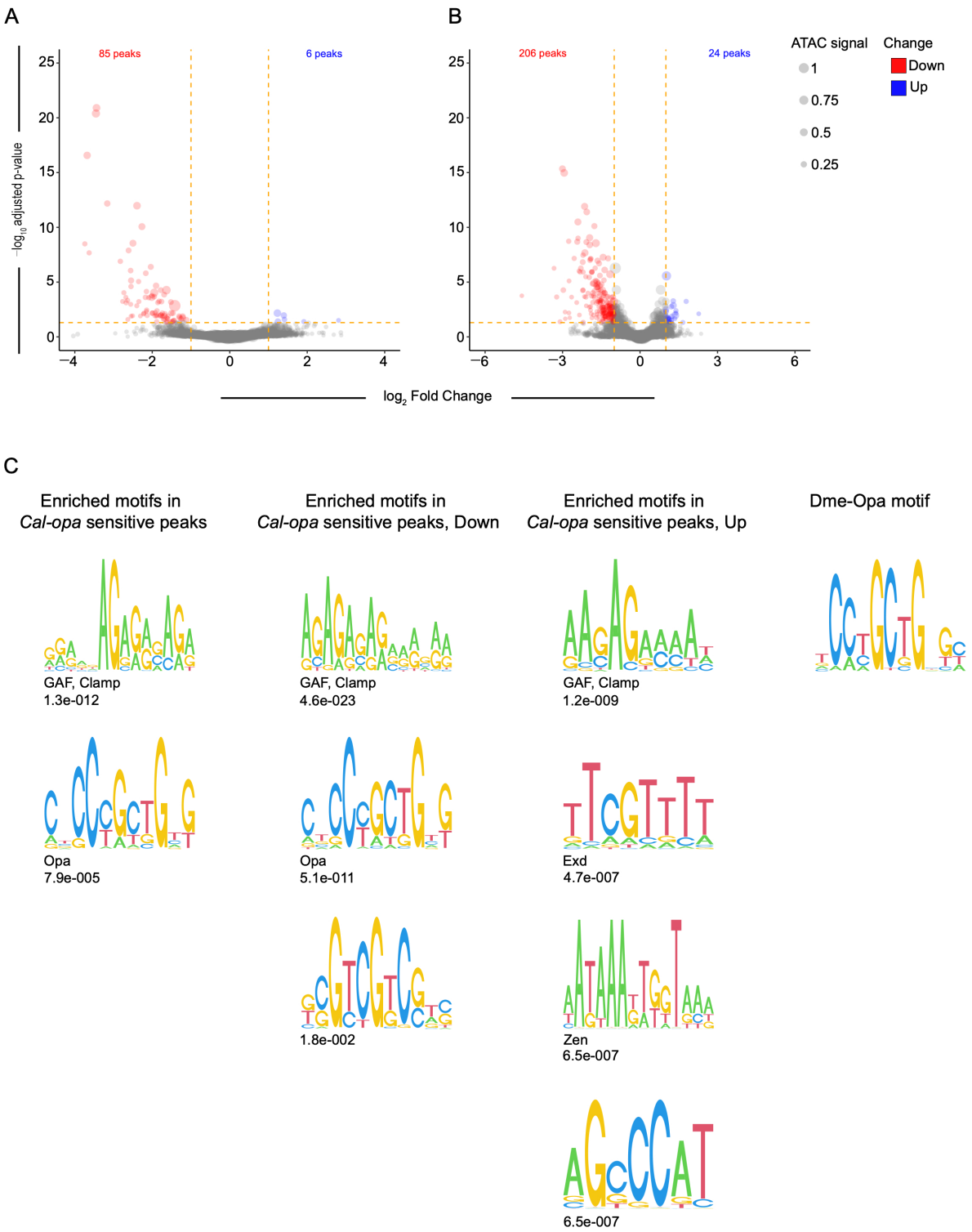

Figure 4: Cal-Opa is required for establishing a selective set of chromatin accessibility peaks. Volcano plots of differentially accessible peaks between the *Cal*-opa RNAi group and the alignment control group at A) NC12 and B) NC13. Significant change of p-adjusted ≤ 0.05 and |log2FC| ≥ 1 highlighted; red (log2FC ≤ -1) and blue (log2FC ≥ 1). C) PWM logos of motifs enriched in all *Cal-opa* sensitive peaks (n = 255), in *Cal-opa* sensitive peaks that lose

accessibility (n = 225), and in *Cal-opa* sensitive peaks that gain accessibility (n = 30), as identified by MEME. Data combined for both NC12 and NC13 stages. When available factors that bind to the motif or a similar motif are listed. E-value, denoting the significance of a motif is provided. The canonical Opa motif in *Drosophila melanogaster* is shown for reference.

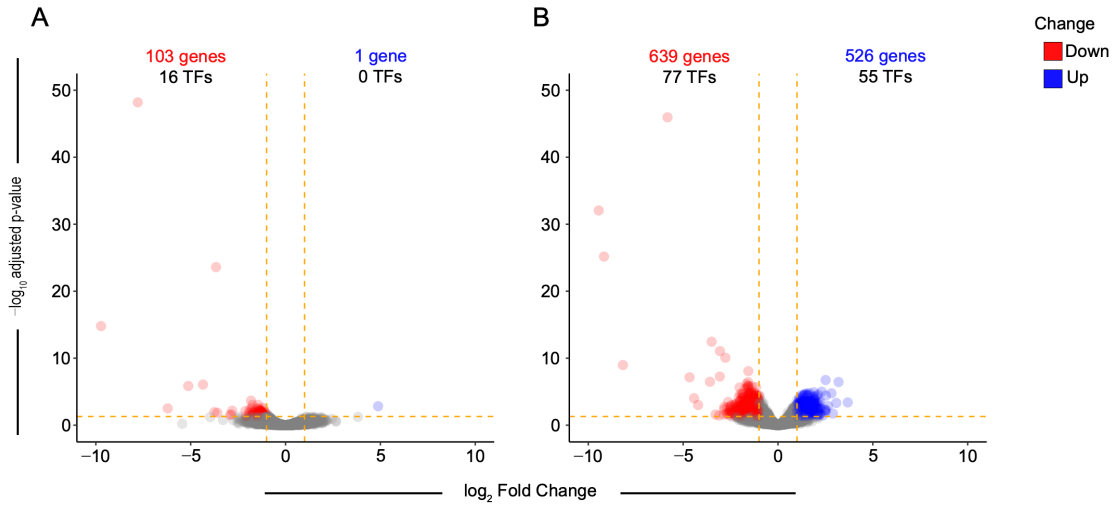

Figure 5: *Cal-opa* knockdown leads to the miss-regulation of a small set of transcription factor genes. Volcano plots of differentially expressed genes between the *Cal*-opa RNAi group and the alignment control group at A) NC12 and B) NC13. Significant change of p-adj ≤ 0.05 and |log2FC| ≥ 1 highlighted; red (log2FC ≤ -1) and blue (log2FC ≥ 1). Number of transcription factors (TFs) is indicated in the text.

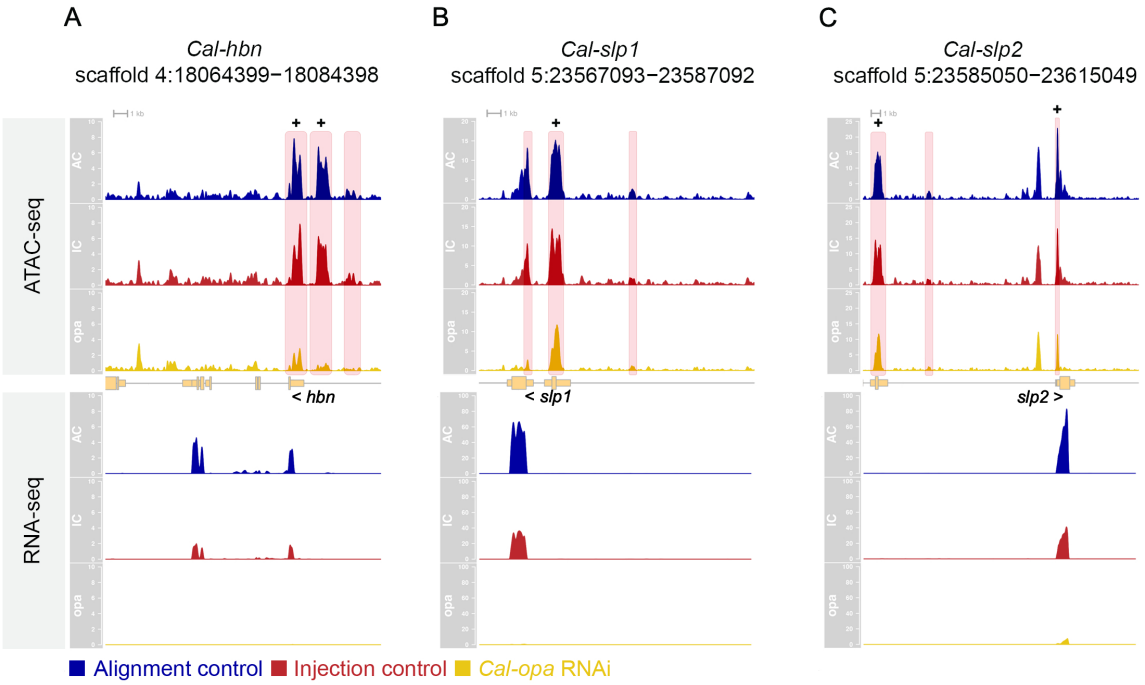

Figure 6: Key Cal-Opa targets at NC12. Gene loci of A) *Cal*-hbn, B) *Cal*-slp1, and C) *Cal-slp2* at NC12. Top tracks: CPM of ATAC-seq data (Alignment control in blue, Injection control in red, *Cal-opa* RNAi in yellow). Middle tracks show gene models in pale yellow. Bottom tracks: CPM of RNA-seq data. Peaks with significant changes in accessibility are highlighted in pink. Peaks containing Opa motifs are marked with a plus sign. The y-axis is the same for all tracks of the same data type at each locus. *Cal-hbn* ATAC-seq: 0-10 CPM. *Cal-hbn* RNA-seq: 0-10 CPM. *Cal-slp1* ATAC-seq: 0-20 CPM. *Cal-slp1* RNA-seq: 0-100 CPM. *Cal-slp2* ATAC-seq: 0-25 CPM. *Cal*-slp2 RNA-seq: 0-100 CPM. Blue: Alignment control. Red: Injection control. Yellow: *Cal-opa* RNAi.

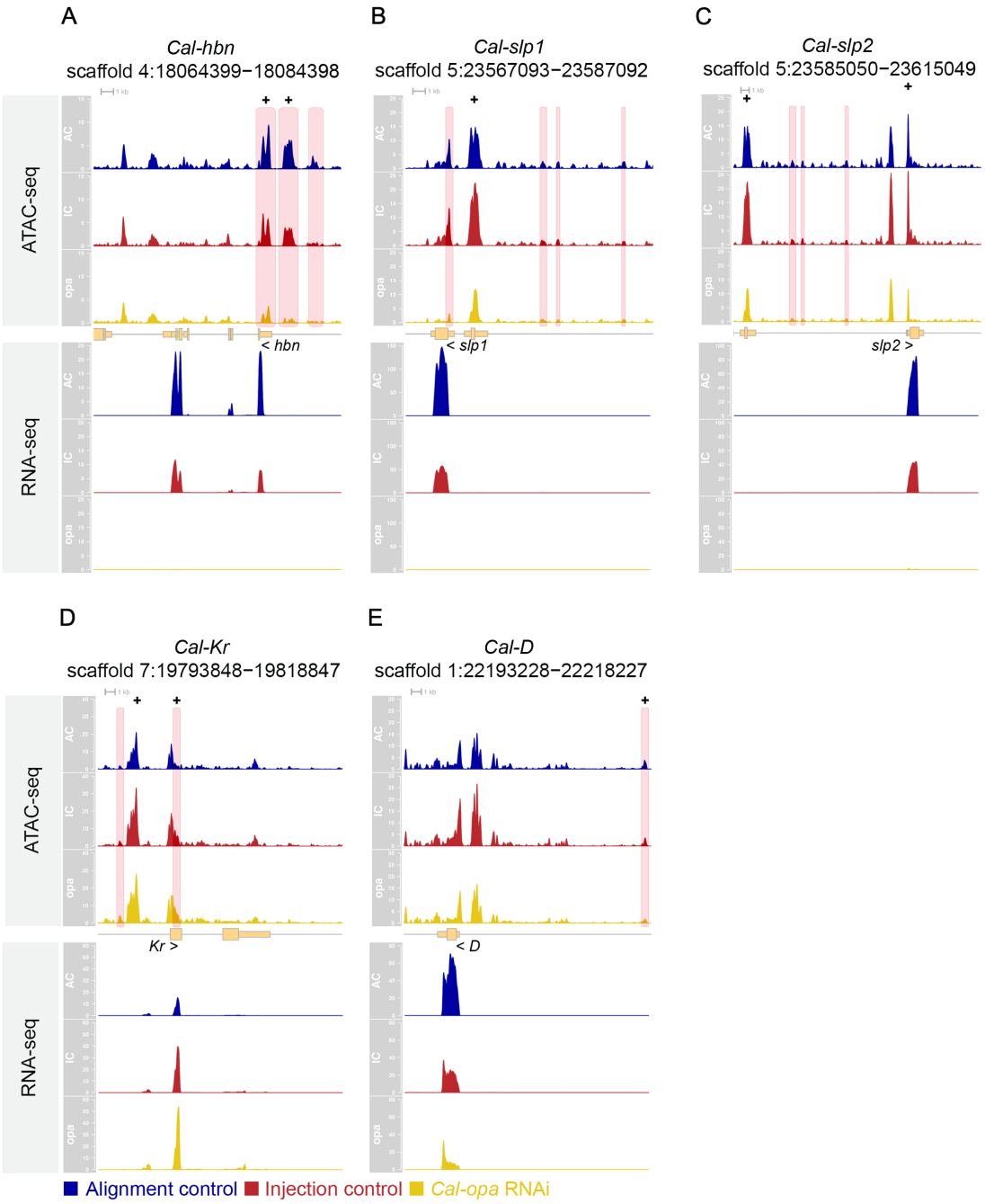

Figure 7: Key Cal-Opa targets at NC13. Gene loci of A) *Cal-hbn*, B) *Cal-slp1*, C) *Cal-slp2*, D) *Cal-Kr* and E) *Cal-D* at NC13. Top tracks: CPM of ATAC-seq data. Middle track gene model in pale yellow. Bottom tracks: CPM of RNA-seq data. Peaks with significant changes in accessibility highlighted in pink. Peaks containing Opa motifs are marked with a plus sign. The y-axis is the same for all tracks of the same data type at each locus. *Cal*-hbn ATAC-seq: 0-15 CPM. *Cal*-hbn RNA-seq: 0-25 CPM. *Cal-slp1* ATAC-seq: 0-25 CPM. *Cal*-slp1 RNA-seq: 0-150 CPM. *Cal-slp2*: ATAC-seq: 0-25 CPM. *Cal-slp2* RNA-seq: 0-100 CPM. *Cal-Kr* ATAC-seq: 0-40 CPM. *Cal-Kr* RNA-seq: 0-60 CPM. *Cal-D* ATAC-seq: 0-30 CPM. *Cal-D* RNA-seq: 0-80 CPM. Blue: Alignment control. Red: Injection control. Yellow: *Cal-opa* RNAi.

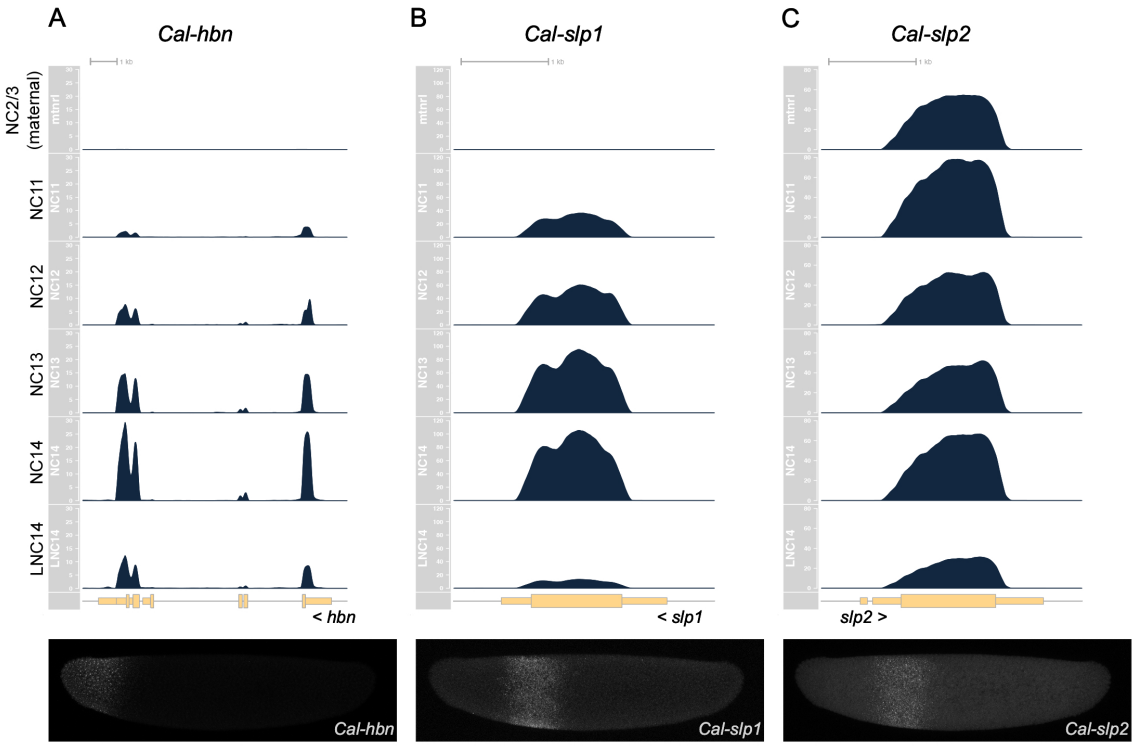

Figure 8: Time-course RNA-seq data for *Cal-hbn*, *Cal-slp1*, and *Cal-slp2*. Expression levels (CPM) of A) *Cal-hbn*, B) *Cal-slp1*, C) *Cal-slp2* at 6 consecutive stages, including NC2/3 (maternal, ~45-minute old embryos), NC11, NC12, NC13, NC14 and Late NC14 are shown above fluorescent *in situ* hybridization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

1000 **References**

1001 1. Driever W, Nüsslein-Volhard C. The bicoid protein determines position in the Drosophila
1002 embryo in a concentration-dependent manner. Cell. 1988;54(1):95-104.

1003 2. Driever W, Nüsslein-Volhard C. A gradient of bicoid protein in Drosophila embryos. Cell.
1004 1988;54(1):83-93.

1005 3. St Johnston D, Nüsslein-Volhard C. The origin of pattern and polarity in the Drosophila
1006 embryo. Cell. 1992;68(2):201-19.

1007 4. Surkova S, Golubkova E, Mamon L, Samsonova M. Dynamic maternal gradients and
1008 morphogenetic networks in Drosophila early embryo. Bio Systems. 2018;173:207-13.

1009 5. Ma J, He F, Xie G, Deng GX. Maternal AP determination in the Drosophila oocyte and
1010 embryo. Wiley Interdiscip Rev Dev Biol. 2016;5:562-81.

1011 6. Wieschaus E. Positional Information and Cell Fate Determination in the Early Drosophila
1012 Embryo. Curr Top Dev Biol. 2016;117:567-79.

1013 7. Jaeger J. The gap gene network. Cellular and molecular life sciences : CMLS.
1014 2011;68(2):243-74.

1015 8. Tkačik G, Gregor T. The many bits of positional information. Development. 2021;148(2).

1016 9. Clark E, Peel AD, Akam M. Arthropod segmentation. Development. 2019;146(18).

1017 10. Crombach A, Jaeger J. Life's attractors continued: progress in understanding
1018 developmental systems through reverse engineering and in silico evolution. In: Crombach A,
1019 editor. Evolutionary Systems Biology: Springer Nature Switzerland AG; 2021.

1020 11. Schmidt-Ott U, Lynch JA. Emerging developmental genetic model systems in
1021 holometabolous insects. Curr Opin Genet Dev. 2016;39:116-28.

1022 12. Lemke S, Kale G, Urbansky S. Comparing gastrulation in flies: Links between cell
1023 biology and the evolution of embryonic morphogenesis. Mechanisms of Development. 2020;164.

1024 13. Anderson DT. Comparative Embryology of Diptera. Annu Rev Entomol. 1966;11:23-64.

1025 14. Klomp J, Athy D, Kwan CW, Bloch NI, Sandmann T, Lemke S, et al. Embryo
1026 development. A cysteine-clamp gene drives embryo polarity in the midge Chironomus. Science.
1027 2015;348(6238):1040-2.

1028 15. Yoon Y, Klomp J, Martin-Martin I, Criscione F, Calvo E, Ribeiro J, et al. Embryo polarity
1029 in moth flies and mosquitoes relies on distinct old genes with localized transcript isoforms. eLife.
1030 2019;8.

1031 16. Schmidt-Ott U, Yoon Y. Evolution and loss of beta-catenin and TCF-dependent axis
1032 specification in insects. Curr Opin Insect Sci. 2022;50:100877.

1033 17. Berleth T, Burri M, Thoma G, Bopp D, Richstein S, Frigerio G, et al. The role of
1034 localization of bicoid RNA in organizing the anterior pattern of the Drosophila embryo. EMBO
1035 J. 1988;7(6):1749-56.

1036 18. Stauber M, Taubert H, Schmidt-Ott U. Function of bicoid and hunchback homologs in the
1037 basal cyclorrhaphan fly Megaselia (Phoridae). Proc Natl Acad Sci U S A. 2000;97(20):10844-9.

1038 19. Shaw PJ, Salameh A, McGregor AP, Bala S, Dover GA. Divergent structure and function
1039 of the bicoid gene in Muscoidea fly species. Evol Dev. 2001;3(4):251-62.

1040 20. Lemke S, Busch SE, Antonopoulos DA, Meyer F, Domanus MH, Schmidt-Ott U.
1041 Maternal activation of gap genes in the hover fly Episyrphus. Development. 2010;137(10):1709-
1042 19.

1043 21. Wotton KR, Jimenez-Guri E, Jaeger J. Maternal co-ordinate gene regulation and axis
1044 polarity in the scuttle fly Megaselia abdita. PLoS genetics. 2015;11(3):e1005042.

1045 22. Hursh DA, Stultz BG. Odd-Paired: The Drosophila Zic Gene. Advances in experimental
1046 medicine and biology. 2018;1046:41-58.

1047 23. Ravindranath AJ, Cadigan KM. Structure-function analysis of the C-clamp of
1048 TCF/Pangolin in Wnt/ss-catenin signaling. PloS one. 2014;9(1):e86180.

1049 24. Liu Q, Onal P, Datta RR, Rogers JM, Schmidt-Ott U, Bulyk ML, et al. Ancient
1050 mechanisms for the evolution of the bicoid homeodomain's function in fly development. eLife.
1051 2018;7.

1052 25. Nitta KR, Jolma A, Yin Y, Morgunova E, Kivioja T, Akhtar J, et al. Conservation of
1053 transcription factor binding specificities across 600 million years of bilateria evolution. eLife.
1054 2015;4.

1055 26. Duan B, Fu D, Zhang C, Ding P, Dong X, Xia B. Selective Nonmethylated CpG DNA
1056 Recognition Mechanism of Cysteine Clamp Domains. J Am Chem Soc. 2021;143(20):7688-97.

1057 27. Li M, Kasan K, Saha Z, Yoon Y, Schmidt-Ott U. Twenty-seven ZAD-ZNF genes of
1058 Drosophila melanogaster are orthologous to the embryo polarity determining mosquito gene
1059 cucoid. PLoS One. 2023;18(1):e0274716.

1060 28. Driever W, Nüsslein-Volhard C. The bicoid protein is a positive regulator of hunchback
1061 transcription in the early Drosophila embryo. Nature. 1989;337(6203):138-43.

1062 29. Struhl G, Struhl K, Macdonald PM. The gradient morphogen bicoid is a concentration-
1063 dependent transcriptional activator. Cell. 1989;57(7):1259-73.

1064 30. Hoch M, Seifert E, Jäckle H. Gene expression mediated by cis-acting sequences of the
1065 Krüppel gene in response to the Drosophila morphogens bicoid and hunchback. Embo j.
1066 1991;10(8):2267-78.

1067 31. Rivera-Pomar R, Lu X, Perrimon N, Taubert H, Jäckle H. Activation of posterior gap
1068 gene expression in the *Drosophila* blastoderm. Nature. 1995;376:253-6.

1069 32. Berman BP, Nibu Y, Pfeiffer BD, Tomancak P, Celniker SE, Levine M, et al. Exploiting
1070 transcription factor binding site clustering to identify cis-regulatory modules involved in pattern
1071 formation in the Drosophila genome. Proc Natl Acad Sci U S A. 2002;99(2):757-62.

1072 33. Schroeder MD, Pearce M, Fak J, Fan H-Q, Unnerstall U, Emberly E, et al.
1073 Transcriptional control in the segmentation gene network of *Drosophila*. PLOS. 2004;2:e271.

1074 34. Ochoa-Espinosa A, Yucel G, Kaplan L, Pare A, Pura N, Oberstein A, et al. The role of
1075 binding site cluster strength in Bicoid-dependent patterning in Drosophila. Proc Natl Acad Sci U
1076 S A. 2005;102(14):4960-5.

1077 35. Gao Q, Finkelstein R. Targeting gene expression to the head: the Drosophila
1078 orthodenticle gene is a direct target of the Bicoid morphogen. Development. 1998;125(21):4185-
1079 93.

1080 36. Stanojevic D, Small S, Levine M. Regulation of a segmentation stripe by overlapping
1081 activators and repressors in the Drosophila embryo. Science. 1991;254(5036):1385-7.

1082 37. Arnosti DN, Barolo S, Levine M, Small S. The eve stripe 2 enhancer employs multiple
1083 modes of transcriptional synergy. Development. 1996;122(1):205-14.

1084 38. Small S, Blair A, Levine M. Regulation of even-skipped stripe 2 in the Drosophila
1085 embryo. EMBO J. 1992;11(11):4047-57.

1086 39. Small S, Kraut R, Hoey T, Warrior R, Levine M. Transcriptional regulation of a pair-rule
1087 stripe in Drosophila. Genes Dev. 1991;5(5):827-39.

1088 40. Simpson-Brose M, Treisman J, Desplan C. Synergy between the hunchback and bicoid
1089 morphogens is required for anterior patterning in Drosophila. Cell. 1994;78(5):855-65.

1090 41. Wimmer EA, Carleton A, Harjes P, Turner T, Desplan C. *bicoid*-independent formation of
1091 thoracic segments in *Drosophila*. Science. 2000;287:2476-9.

1092 42. Manu, Surkova S, Spirov AV, Gursky VV, Janssens H, Kim AR, et al. Canalization of
1093 gene expression and domain shifts in the Drosophila blastoderm by dynamical attractors. PLoS
1094 Comput Biol. 2009;5(3):e1000303.

1095 43. Rivera-Pomar R, Jackle H. From gradients to stripes in Drosophila embryogenesis: filling
1096 in the gaps. Trends Genet. 1996;12(11):478-83.

1097 44. Macdonald PM. Translational repression by Bicoid: competition for the cap. Cell.
1098 2005;121(3):321-2.

1099 45. Staudt N, Fellert S, Chung HR, Jackle H, Vorbruggen G. Mutations of the Drosophila
1100 zinc finger-encoding gene vielfaltig impair mitotic cell divisions and cause improper
1101 chromosome segregation. Mol Biol Cell. 2006;17(5):2356-65.

1102 46. Liang HL, Nien CY, Liu HY, Metzstein MM, Kirov N, Rushlow C. The zinc-finger
1103 protein Zelda is a key activator of the early zygotic genome in Drosophila. Nature.
1104 2008;456(7220):400-3.

1105 47. Foo SM, Sun Y, Lim B, Ziukaite R, O'Brien K, Nien CY, et al. Zelda potentiates
1106 morphogen activity by increasing chromatin accessibility. Curr Biol. 2014;24(12):1341-6.

1107 48. Xu Z, Chen H, Ling J, Yu D, Struffi P, Small S. Impacts of the ubiquitous factor Zelda on
1108 Bicoid-dependent DNA binding and transcription in Drosophila. Genes Dev. 2014;28(6):608-21.

1109 49. Mir M, Stadler MR, Ortiz SA, Hannon CE, Harrison MM, Darzacq X, et al. Dynamic
1110 multifactor hubs interact transiently with sites of active transcription in Drosophila embryos.
1111 eLife. 2018;7.

1112 50. Bishop TR, Onal P, Xu Z, Zheng M, Gunasinghe H, Nien CY, et al. Multi-level regulation
1113 of even-skipped stripes by the ubiquitous factor Zelda. Development. 2023;150(23).

1114 51. Crocker J, Stern DL. Functional regulatory evolution outside of the minimal even-
1115 skipped stripe 2 enhancer. Development. 2017;144(17):3095-101.

1116 52. Ling J, Umezawa KY, Scott T, Small S. Bicoid-Dependent Activation of the Target Gene
1117 hunchback Requires a Two-Motif Sequence Code in a Specific Basal Promoter. Mol Cell.
1118 2019;75(6):1178-87 e4.

1119 53. Eck E, Liu J, Kazemzadeh-Atoufi M, Ghoreishi S, Blythe SA, Garcia HG. Quantitative
1120 dissection of transcription in development yields evidence for transcription-factor-driven
1121 chromatin accessibility. Elife. 2020;9.

1122 54. Fernandes G, Tran H, Andrieu M, Diaw Y, Perez Romero C, Fradin C, et al. Synthetic
1123 reconstruction of the hunchback promoter specifies the role of Bicoid, Zelda and Hunchback in
1124 the dynamics of its transcription. Elife. 2022;11.

1125 55. Harden TT, Vincent BJ, DePace AH. Transcriptional activators in the early Drosophila
1126 embryo perform different kinetic roles. Cell systems. 2023;14(4):258-72 e4.

1127 56. Hannon CE, Blythe SA, Wieschaus EF. Concentration dependent chromatin states
1128 induced by the bicoid morphogen gradient. eLife. 2017;6.

1129 57. Haines JE, Eisen MB. Patterns of chromatin accessibility along the anterior-posterior axis
1130 in the early Drosophila embryo. PLoS genetics. 2018;14(5):e1007367.

1131 58. Soluri IV, Zumerling LM, Payan Parra OA, Clark EG, Blythe SA. Zygotic pioneer factor
1132 activity of Odd-paired/Zic is necessary for late function of the Drosophila segmentation network.
1133 eLife. 2020;9.

1134 59. Cusanovich DA, Reddington JP, Garfield DA, Daza RM, Aghamirzaie D, Marco-Ferreres
1135 R, et al. The cis-regulatory dynamics of embryonic development at single-cell resolution. Nature.
1136 2018;555(7697):538-42.

1137 60. Calderon D, Blecher-Gonen R, Huang X, Secchia S, Kentro J, Daza RM, et al. The
1138 continuum of Drosophila embryonic development at single-cell resolution. Science.
1139 2022;377(6606):eabn5800.

1140 61. Degen EA, Croslyn C, Mangan NM, Blythe SA. Bicoid-nucleosome competition sets a
1141 concentration threhold for transcription constrained by genome replication. bioRxiv. 2024.

1142 62. Blythe SA, Wieschaus EF. Establishment and maintenance of heritable chromatin
1143 structure during early Drosophila embryogenesis. eLife. 2016;5.

1144 63. Clark E, Akam M. Odd-paired controls frequency doubling in Drosophila segmentation
1145 by altering the pair-rule gene regulatory network. Elife. 2016;5.

1146 64. Hülskamp M, Pfeifle C, Tautz D. A morphogenetic gradient of Hunchback protein
1147 organizes the expression of the gap genes *Krüppel* and *knirps* in the early *Drosophila* embryo.
1148 Nature. 1990;346:577-80.

1149 65. Tautz D. Regulation of the *Drosophila* segmentation gene *hunchback* by two maternal
1150 morphogenetic centers. Nature (London). 1988;332:281-4.

1151 66. Janssens H, Siggens K, Cicin-Sain D, Jimenez-Guri E, Musy M, Akam M, et al. A
1152 quantitative atlas of Even-skipped and Hunchback expression in Clogmia albipunctata (Diptera:
1153 Psychodidae) blastoderm embryos. EvoDevo. 2014;5(1):1.

1154 67. Garcia-Solache M, Jaeger J, Akam M. A systematic analysis of the gap gene system in the
1155 moth midge *Clogmia albipunctata*. Developmental biology. 2010;344:306-18.

1156 68. Koromila T, Gao F, Iwasaki Y, He P, Pachter L, Gergen JP, et al. Odd-paired is a pioneer-
1157 like factor that coordinates with Zelda to control gene expression in embryos. eLife. 2020;9.

1158 69. Jimenez-Guri E, Wotton KR, Gavilan B, Jaeger J. A staging scheme for the development
1159 of the moth midge Clogmia albipunctata. PloS one. 2014;9(1):e84422.

1160 70. Amabis JM, Simoes LCG. Chromosome studies in a species of *Telmatoscopus* sp.
1161 Caryologia. 1972;25(2).

1162 71. Amabis JM. Cytological evidence of male sex heterogamety in sex determination of
1163 *Telmatoscopus albipunctatus* (Diptera: Psychodidae). Chromosoma. 1977;62:133-8.

1164 72. Schmidt-Ott U, Rafiqi AM, Sander K, Johnston JS. Extremely small genomes in two
1165 unrelated dipteran insects with shared early developmental traits. Development genes and
1166 evolution. 2009;219(4):207-10.

1167 73. Generalovic TN, McCarthy SA, Warren IA, Wood JMD, Torrance J, Sims Y, et al. A high-
1168 quality, chromosome-level genome assembly of the Black Soldier Fly (Hermetia illucens L.). G3
1169 (Bethesda). 2021;11(5).

1170 74. Tenger-Trolander A, Amiri E, Gantz V, Kwan CW, Sanders SA, Schmidt-Ott U. Genomic
1171 resources for the scuttle fly Megaselia abdita: A Model Organism for Comparative
1172 Developmental Studies in Flies bioRxiv. 2025.

1173 75. Ji J, Gao Y, Xu C, Zhang K, Li D, Li B, et al. Chromosome-level genome assembly of
1174 marmalade hoverfly Episyrphus balteatus (Diptera: Syrphidae). Scientific data. 2024;11(1):844.

1175 76. Stauber M, Jäckle H, Schmidt-Ott U. The anterior determinant *bicoid* of *Drosophila* is a
1176 derived Hox class 3 gene. Proc Natl Acad Sci USA. 1999;96:3786-9.

1177 77. Stauber M, Prell A, Schmidt-Ott U. A single *Hox3* gene with composite *bicoid* and
1178 *zerknüllt* expression characteristics in non-Cyclorrhaphan flies. Proceedings of the National
1179 Academy of Sciences of the United States of America. 2002;99:274-9.

1180 78. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based
1181 analysis of ChIP-Seq (MACS). Genome biology. 2008;9(9):R137.

1182 79. Reddington JP, Garfield DA, Sigalova OM, Karabacak Calviello A, Marco-Ferreres R,
1183 Girardot C, et al. Lineage-Resolved Enhancer and Promoter Usage during a Time Course of
1184 Embryogenesis. Dev Cell. 2020;55(5):648-64 e9.

1185 80. Schulz KN, Harrison MM. Mechanisms regulating zygotic genome activation. Nat Rev
1186 Genet. 2019;20(4):221-34.

1187 81. Hamm DC, Harrison MM. Regulatory principles governing the maternal-to-zygotic
1188 transition: insights from Drosophila melanogaster. Open Biol. 2018;8(12):180183.

1189 82. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for
1190 RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.

1191 83. Larson ED, Marsh AJ, Harrison MM. Pioneering the developmental frontier. Mol Cell.
1192 2021;81(8):1640-50.

1193 84. Brennan KJ, Weilert M, Krueger S, Pampari A, Liu HY, Yang AWH, et al. Chromatin
1194 accessibility in the Drosophila embryo is determined by transcription factor pioneering and
1195 enhancer activation. Dev Cell. 2023;58(19):1898-916 e9.

1196 85. Harrison MM, Li XY, Kaplan T, Botchan MR, Eisen MB. Zelda binding in the early
1197 Drosophila melanogaster embryo marks regions subsequently activated at the maternal-to-
1198 zygotic transition. PLoS Genet. 2011;7(10):e1002266.

1199 86. McDaniel SL, Gibson TJ, Schulz KN, Fernandez Garcia M, Nevil M, Jain SU, et al.
1200 Continued Activity of the Pioneer Factor Zelda Is Required to Drive Zygotic Genome Activation.
1201 Mol Cell. 2019;74(1):185-95 e4.

1202 87. Sun Y, Nien CY, Chen K, Liu HY, Johnston J, Zeitlinger J, et al. Zelda overcomes the
1203 high intrinsic nucleosome barrier at enhancers during Drosophila zygotic genome activation.
1204 Genome research. 2015;25(11):1703-14.

1205 88. Gaskill MM, Gibson TJ, Larson ED, Harrison MM. GAF is essential for zygotic genome
1206 activation and chromatin accessibility in the early Drosophila embryo. Elife. 2021;10.

1207 89. Duan J, Rieder L, Colonnetta MM, Huang A, McKenney M, Watters S, et al. CLAMP
1208 and Zelda function together to promote Drosophila zygotic genome activation. Elife. 2021;10.

1209 90. Colonnetta MM, Abrahante JE, Schedl P, Gohl DM, Deshpande G. CLAMP regulates
1210 zygotic genome activation in Drosophila embryos. Genetics. 2021;219(2).

1211 91. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME Suite. Nucleic acids research.
1212 2015;43(W1):W39-49.

1213 92. Ribeiro L, Tobias-Santos V, Santos D, Antunes F, Feltran G, de Souza Menezes J, et al.
1214 Evolution and multiple roles of the Pancrustacea specific transcription factor zelda in insects.
1215 PLoS genetics. 2017;13(7):e1006868.

1216 93. Biedler JK, Hu W, Tae H, Tu Z. Identification of early zygotic genes in the yellow fever
1217 mosquito Aedes aegypti and discovery of a motif involved in early zygotic genome activation.
1218 PLoS One. 2012;7(3):e33933.

1219 94. Arsala D, Lynch JA. Ploidy has little effect on timing early embryonic events in the
1220 haplo-diploid wasp Nasonia. Genesis. 2017;55(5).

1221 95. Pires CV, Freitas FC, Cristino AS, Dearden PK, Simoes ZL. Transcriptome Analysis of
1222 Honeybee (Apis Mellifera) Haploid and Diploid Embryos Reveals Early Zygotic Transcription
1223 during Cleavage. PLoS One. 2016;11(1):e0146447.

1224 96. Gao S, Xue S, Gao T, Lu R, Zhang X, Zhang Y, et al. Transcriptome analysis reveals the
1225 role of Zelda in the regulation of embryonic and wing development of Tribolium castaneum. Bull
1226 Entomol Res. 2023;113(5):587-97.

1227 97. Ventos-Alfonso A, Ylla G, Belles X. Zelda and the maternal-to-zygotic transition in
1228 cockroaches. Febs J. 2019;286(16):3206-21.

1229 98. Gibson TJ, Larson ED, Harrison MM. Protein-intrinsic properties and context-dependent
1230 effects regulate pioneer factor binding and function. Nature structural & molecular biology.
1231 2024;31(3):548-58.

1232 99. Zhu LJ, Christensen RG, Kazemian M, Hull CJ, Enuameh MS, Basciotta MD, et al.
1233 FlyFactorSurvey: a database of Drosophila transcription factor binding specificities determined
1234 using the bacterial one-hybrid system. Nucleic acids research. 2011;39(Database issue):D111-7.

1235 100. Nüsslein-Volhard C, Wieschaus E. Mutations affecting segment number and polarity in
1236 *Drosophila*. Nature. 1980;287:795-801.

1237 101. Goltsev Y, Hsiong W, Lanzaro G, Levine M. Different combinations of gap repressors for
1238 common stripes in *Anopheles* and *Drosophila* embryos. Dev Biol. 2004;275:435-46.

1239 102. Lemke S, Busch SE, Antonopoulos DA, Meyer F, Domanus MH, Schmidt-Ott U.
1240 Maternal activation of gap genes in the hover fly *Episyrphus*. Development. 2010;137:1709-19.

1241 103. Wotton KR, Jimenez-Guri E, Crombach A, Janssens H, Alcaine-Colet A, Lemke S, et al.
1242 Quantitative system drift compensates for altered maternal inputs to the gap gene network of the
1243 scuttle fly Megaselia abdita. eLife. 2015;4.

1244 104. Bullock SL, Stauber M, Prell A, Hughes JR, Ish-Horowicz D, Schmidt-Ott U. Differential
1245 cytoplasmic mRNA localization adjusts pair-rule transcription factor activity to cytoarchitecture
1246 in dipteran evolution. Development. 2004;131:4251-61.

1247 105. Campos-Ortega JA, Hartenstein V. The embryonic development of *Drosophila*
1248 *melanogaster*. 2 ed. Berlin, Heidelberg, New York: Springer-Verlag; 1997.

1249 106. Yoon Y. Evolution of embryonic axis determinants via alternative transcription:
1250 University of Chicago; 2019.

1251 107. Rafiqi AM, Park CH, Kwan CW, Lemke S, Schmidt-Ott U. BMP-dependent serosa and
1252 amnion specification in the scuttle fly Megaselia abdita. Development. 2012;139(18):3373-82.

1253 108. Rafiqi AM, Lemke S, Ferguson S, Stauber M, Schmidt-Ott U. Evolutionary origin of the
1254 amnioserosa in cyclorrhaphan flies correlates with spatial and temporal expression changes of
1255 zen. Proc Natl Acad Sci U S A. 2008;105(1):234-9.

1256 109. Kwan CW, Gavin-Smyth J, Ferguson EL, Schmidt-Ott U. Functional evolution of a
1257 morphogenetic gradient. eLife. 2016;5.

1258 110. Caroti F, Gonzalez Avalos E, Noeske V, Gonzalez Avalos P, Kromm D, Wosch M, et al.
1259 Decoupling from yolk sac is required for extraembryonic tissue spreading in the scuttle fly
1260 Megaselia abdita. Elife. 2018;7.

1261 111. Urbansky S, Gonzalez Avalos P, Wosch M, Lemke S. Folded gastrulation and T48 drive
1262 the evolution of coordinated mesoderm internalization in flies. eLife. 2016;5.

1263 112. Vicoso B, Bachtrog D. Numerous transitions of sex chromosomes in Diptera. PLoS
1264 biology. 2015;13(4):e1002078.

1265 113. Hall AB, Basu S, Jiang X, Qi Y, Timoshevskiy VA, Biedler JK, et al. SEX
1266 DETERMINATION. A male-determining factor in the mosquito Aedes aegypti. Science.
1267 2015;348(6240):1268-70.

1268 114. Krzywinska E, Dennison NJ, Lycett GJ, Krzywinski J. A maleness gene in the malaria
1269 mosquito Anopheles gambiae. Science. 2016;353(6294):67-9.

1270 115. Saccone G. A history of the genetic and molecular identification of genes and their
1271 functions controlling insect sex determination. Insect Biochem Mol Biol. 2022;151:103873.

1272 116. Grand RS, Pregnolato M, Baumgartner L, Hoerner L, Burger L, Schubeler D. Genome
1273 access is transcription factor-specific and defined by nucleosome position. Molecular cell.
1274 2024;84(18):3455-68 e6.

1275 117. Schulz KN, Bondra ER, Moshe A, Villalta JE, Lieb JD, Kaplan T, et al. Zelda is
1276 differentially required for chromatin accessibility, transcription factor binding, and gene
1277 expression in the early Drosophila embryo. Genome research. 2015;25(11):1715-26.

1278 118. Seller CA, O'Farrell PH. Rif1 prolongs the embryonic S phase at the Drosophila mid-
1279 blastula transition. PLoS biology. 2018;16(5):e2005687.

1280 119. Walldorf U, Kiewe A, Wickert M, Ronshaugen M, McGinnis W. Homeobrain, a novel
1281 paired-like homeobox gene is expressed in the Drosophila brain. Mech Dev. 2000;96(1):141-4.

1282 120. Kolb D, Kaspar P, Kloppel C, Walldorf U. The Drosophila homeodomain transcription
1283 factor Homeobrain is involved in the formation of the embryonic protocerebrum and the
1284 supraesophageal brain commissure. Cells Dev. 2021;165:203657.

1285 121. Hildebrandt K, Kolb D, Kloppel C, Kaspar P, Wittling F, Hartwig O, et al. Regulatory
1286 modules mediating the complex neural expression patterns of the homeobrain gene during
1287 Drosophila brain development. Hereditas. 2022;159(1):2.

1288 122. Ansari S, Troelenberg N, Dao VA, Richter T, Bucher G, Klingler M. Double abdomen in
1289 a short-germ insect: Zygotic control of axis formation revealed in the beetle Tribolium
1290 castaneum. Proceedings of the National Academy of Sciences of the United States of America.
1291 2018;115(8):1819-24.

1292 123. Schomburg C, Janssen R, Prpic NM. Phylogenetic analysis of forkhead transcription
1293 factors in the Panarthropoda. Dev Genes Evol. 2022;232(1):39-48.

1294 124. Clark E. Dynamic patterning by the Drosophila pair-rule network reconciles long-germ
1295 and short-germ segmentation. PLoS Biol. 2017;15(9):e2002439.

1296 125. Cadigan KM, Grossniklaus U, Gehring WJ. Functional redundancy: the respective roles
1297 of the two sloppy paired genes in Drosophila segmentation. Proc Natl Acad Sci U S A.
1298 1994;91(14):6324-8.

1299 126. Grossniklaus U, Pearson RK, Gehring WJ. The Drosophila sloppy paired locus encodes
1300 two proteins involved in segmentation that show homology to mammalian transcription factors.
1301 Genes Dev. 1992;6(6):1030-51.

1302 127. Fujioka M, Jaynes JB. Regulation of a duplicated locus: Drosophila sloppy paired is
1303 replete with functionally overlapping enhancers. Dev Biol. 2012;362(2):309-19.

1304 128. Lee HH, Frasch M. Survey of forkhead domain encoding genes in the Drosophila
1305 genome: Classification and embryonic expression patterns. Dev Dyn. 2004;229(2):357-66.

1306 129. Gao Q, Finkelstein R. Targeting gene expression to the head: the *Drosophila*
1307 *orthodenticle* gene is a direct target of the Bicoid morphogen. Development. 1998:4185-93.

1308 130. Choi HMT, Schwarzkopf M, Fornace ME, Acharya A, Artavanis G, Stegmaier J, et al.
1309 Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive,
1310 versatile, robust. Development. 2018;145(12).

1311 131. Bruce HS, Jerz G, Kelly SR, McCarthy J, Pomerantz A, Senevirathne G, et al.
1312 Hybridization chain reaction (HCR) in situ protocol V.1.; 2021.

1313 132. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native
1314 chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins
1315 and nucleosome position. Nature methods. 2013;10(12):1213-8.

1316 133. Krueger F. Trim Galore! : Babraham Bioinformatics; 2012.

1317 134. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature methods.
1318 2012;9(4):357-9.

1319 135. Lawrence M, Huber W, Pages H, Aboyoun P, Carlson M, Gentleman R, et al. Software
1320 for computing and annotating genomic ranges. PLoS computational biology.
1321 2013;9(8):e1003118.

1322 136. Morgan M, Pages H, Obenchain V, Hayden N. Rsamtools: Binary alignment (BAM),
1323 FASTA, variant call (BCF), and tabix file import.: Bioconductor; 2013.

1324 137. Pantano L. DEGreport: report of DEG analysis 2017 [

1325 138. Bergman CM, Carlson JW, Celniker SE. Drosophila DNase I footprint database: a
1326 systematic genome annotation of transcription factor binding sites in the fruitfly, Drosophila
1327 melanogaster. Bioinformatics. 2005;21(8):1747-9.

1328 139. Kulakovskii IV, Makeev V. [Integration of data obtained by different experimental
1329 methods to determine the motifs in DNA sequences recognized by transcription-regulating
1330 factors]. Biofizika. 2009;54(6):965-74.

1331 140. Kulakovskiy IV, Favorov AV, Makeev VJ. Motif discovery and motif finding from
1332 genome-mapped DNase footprint data. Bioinformatics. 2009;25(18):2318-25.

1333 141. Shazman S, Lee H, Socol Y, Mann RS, Honig B. OnTheFly: a database of Drosophila
1334 melanogaster transcription factors and their binding sites. Nucleic acids research.
1335 2014;42(Database issue):D167-71.

1336 142. Rohr KB, Tautz D, Sander K. Segmentation gene expression in the mothmidge Clogmia
1337 albipunctata (Diptera, psychodidae) and other primitive dipterans. Dev Genes Evol.
1338 1999;209(3):145-54.

1339 143. Smit A, Hubley R. RepeatModeler 2008 [Available from: https://www.repeatmasker.org/.

1340 144. Smit A, Hubley R, Green P. RepeatMasker 2013 [Available from:
1341 https://www.repeatmasker.org/.

1342 145. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, et al. Automated eukaryotic
1343 gene structure annotation using EVidenceModeler and the Program to Assemble Spliced
1344 Alignments. Genome biology. 2008;9(1):R7.

1345 146. Huerta-Cepas J, Szklarczyk D, Heller D, Hernandez-Plaza A, Forslund SK, Cook H, et al.
1346 eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource
1347 based on 5090 organisms and 2502 viruses. Nucleic acids research. 2019;47(D1):D309-D14.

1348 147. Cantalapiedra CP, Hernandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-
1349 mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the
1350 Metagenomic Scale. Molecular biology and evolution. 2021;38(12):5825-9.

148. Simao 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.

149. Manni M, Berkeley MR, Seppey M, Zdobnov EM. BUSCO: Assessing Genomic Data Quality and Beyond. Curr Protoc. 2021;1(12):e323.

150. 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 research. 2012;40(7):e49.

151. Gu Z, Gu L, Eils R, Schlesner M, Brors B. circlize Implements and enhances circular visualization in R. Bioinformatics. 2014;30(19):2811-2.

152. 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 biology. 2023;24(1):74.

153. Hancock DY, Fischer J, Lowe JM, Fischer J, Lowe JM, Snapp-Childs W, et al. Jetstream2: Accelerating cloud computing via Jetstream. . Practice and Experience in Advanced Research Computing (PEARC ’21). New York, NY: Association for Computing Machinery; 2021.

154. Boerner TJ, Deems S, Furlani TR, Knuth SL, Towns J. ACCESS: Advancing Innovation: NSF’s Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support. Practice and Experience in Advanced Research Computing (PEARC ’23). New York, NY: Association for Computing Machinery; 2023.

155. Priyam A, Woodcroft BJ, Rai V, Moghul I, Munagala A, Ter F, et al. Sequenceserver: A Modern Graphical User Interface for Custom BLAST Databases. Molecular biology and evolution. 2019;36(12):2922-4.

156. Beeber D, Chain FJ. crispRdesignR: A Versatile Guide RNA Design Package in R for CRISPR/Cas9 Applications. J Genomics. 2020;8:62-70.

157. Brooks EM, Sanders SA, Pfrender ME. FreeCount: A coding-free framework for guided count data visualization and analysis. Practice and Experience in Advanced Research Computing 2024: Human Powered Computing (PEARC '24). New York, NY: Association for Computing Machinery; 2024.

158. Xiao Z, Lam HM. ShinySyn: a Shiny/R application for the interactive visualization and integration of macro- and micro-synteny data. Bioinformatics. 2022;38(18):4406-8.