Breaking anterior-posterior symmetry in the moth fly *Clogmia albipunctata*

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 (AP) body axis is a fundamental process during embryogenesis, and the fruit fly, *Drosophila melanogaster*, provides one of the best-known case studies. But for unknown reasons, different species of flies (Diptera) establish the AP axis through unrelated, structurally distinct anterior determinants (ADs). The AD of Drosophila, Bicoid (Bcd), initiates symmetry-breaking during nuclear cleavage cycles (NCs) when ubiquitous pioneer factors, such as Zelda (Zld), drive zygotic genome activation (ZGA) at the level chromatin accessibility by nucleosome depletion. While Bcd engages in a concentration-dependent competition with nucleosomes at the loci of a small set of transcription factor (TF) genes that are expressed in the anterior embryo, it remains unknown whether unrelated ADs of other fly species function in the same way and target homologous genes. We have examined the symmetry-breaking mechanism of a moth fly, *Clogmia albipunctata*, in which a maternally expressed transcript isoform of the pair-rule segmentation gene *odd-paired* serves as AD. We provide a *de novo* assembly and annotation of the Clogmia genome and describe how Clogmia's orthologs of *zelda* (*Cal-zld*) and *odd-paired* (*Cal-opa*) affect chromatin accessibility and gene expression. Our results suggest direct roles of *Cal-zld* in opening and closing chromatin during nuclear cleavage cycles (NCs) and show that during the early phase of ZGA maternal *Cal-opa* activity promotes chromatin accessibility and anterior expression at Clogmia's *homeobrain* and *sloppy-paired* loci. These genes are not known as key targets of Bcd but may serve a more widely conserved role in the initiation of anterior pattern formation given their early anterior expression and function in head development in insects. We conclude that the ADs of Drosophila and Clogmia differ in their target genes but share the mechanism of concentration-dependent nucleosome depletion.

Introduction

The insect order of flies (Diptera) offers excellent opportunities for comparative study of developmental mechanisms, as many dipteran species can be reared cost-effectively and because the fruit fly, *Drosophila melanogaster*, provides a well-established model for functional comparisons at the molecular and cellular level. Dipteran embryos are morphologically similar across the clade (1, 2). However, the mechanisms of axis specification (3, 4), primordial germ cell specification (3), extraembryonic tissue specification (5, 6), morphogenesis (5, 7-10), and sex determination (11-14) differ between dipteran species. Explaining such unexpected plasticity in developmental gene networks requires functional comparisons, which can reveal underlying general principles and mechanisms of evolutionary change. Anterior-posterior (**AP**) axis specification, a long-standing model for pattern formation and gene regulation in Drosophila (15-23), is particularly suitable for comparative functional studies because the syncytial nature of early dipteran embryos facilitates visualization and perturbation of gene activity in non-traditional model organisms. Many dipterans establish the embryo's head-to-tail polarity through transcription factors (**TF**s) that diffuse from maternally localized mRNA and form AP concentration gradients that regulate gene expression in a dose-dependent manner (24, 25). These anterior determinants (**AD**s) prevent the formation of bicaudal embryos that lack head and thorax (double abdomen). Irrespective of this fundamental role, unrelated ADs have been found in different fly species. While the AD of Drosophila and many other cyclorrhaphan fly species is encoded by a homeobox gene, *bicoid* (26-30), diverse zinc finger genes take on this role in lower dipterans, including *pangolin* (dipteran ortholog *Tcf* gene family) in anopheline mosquitoes and

crane flies, *cucoid* in culicine mosquitoes, *panish* in certain harlequin flies (Chironomini), and *odd-paired* (dipteran ortholog of *Zic* gene family) in moth flies (Psychodidae) (3, 4). The phylogenetic occurrence of these ADs indicates that they have been frequently replaced during the dipteran radiation and that only a small fraction of dipteran ADs has been identified to date. Moreover, the AD proteins identified so far differ widely in their structure and DNA-binding domains. Bicoid (**Bcd**) binds DNA via its homeodomain, Pangolin (**Pan**) through HMG and C-clamp domains, Panish through a C-clamp domain, and Odd-paired (**Opa**) and Cucoid through distinct C2H2 zinc finger domains (31-36). This molecular diversity prompted us to ask whether the mechanisms of dipteran ADs and their target genes differ between species or whether they are conserved due to contextual or downstream network constraints.

We have addressed these questions by analyzing the mechanism of action of the AD of the moth fly *Clogmia albipunctata*, a C2H2 zinc finger protein encoded by the *zic* gene family member *odd-paired* (3) and comparing it to the AD of *Drosophila melanogaster*, a homeodomain protein encoded by *bicoid* (26), which is missing in Clogmia and other midges. The embryos of Drosophila and Clogmia undergo 13 nuclear division cycles (**NC**) before they cellularize and initiate gastrulation (37, 38). By the end of NC9 (Clogmia) or the beginning of NC10 (Drosophila), most nuclei of the embryo have arranged in a monolayer between the egg membrane and the yolk. This 'syncytial blastoderm' undergoes four additional nuclear cleavage cycles before it cellularizes during NC14, thereby ending the syncytial phase of embryogenesis. During this period, a segmental body plan has been established by a regulatory network of transcription factor encoding genes, known as gap and pair-rule genes (19, 21, 39, 40). In Drosophila, only a few hundred genes are expressed before the midblastula transition (NC14) (41-43), when mitotic cycles become desynchronized and the cell cycle lengthens partly due to the addition of gap phases, as the embryo transition from maternal to zygotic control of development (44). Genes expressed during NC8-NC13 include regulators of the cell cycle, sex determination, dosage compensation, and early pattern formation (45-47). Widespread zygotic genome activation (**ZGA**), however, is delayed until NC14 due to constraints on transcription imposed by rapid cell cycling and the requirement of global changes in chromatin accessibility (47-49). After each DNA replication cycle, the formation of nucleosomes reduces chromatin accessibility. This barrier is overcome by a special class of sequence-specific TFs capable of binding nucleosome occupied DNA and destabilizing nucleosomes (50, 51). The major driver of chromatin accessibility during Drosophila's nuclear cleavage cycles is Zelda (**Zld**), a ubiquitously expressed zinc finger protein (46, 52-55). Zld's activity is required throughout ZGA and enables the binding of other TFs including those involved in embryonic patterning (53, 56-60). Two additional pioneer TFs, GAGA Factor (GAF) and CLAMP, act independently and cooperatively with Zld to establish open chromatin and drive zygotic gene expression once major ZGA is underway (61-64).

Bcd's role in AP patterning becomes indispensable after blastoderm formation at the beginning of NC10 (65). During early blastoderm stages, chromatin accessibility measured by ATAC-seq (assay for transposase-accessible chromatin with sequencing) is largely homogeneous across the embryo. Indeed, Zld establishes chromatin accessibility ubiquitoulsy in the embryo at many sites that are targeted by Bcd (57, 59, 66-72). However, a small cohort of early segmentation gene enhancers that drive Bcd-dependent anterior expression exhibits anteriorly restricted accessibility (33, 73-79). The underlying mechanism has been inferred by measuring accessibility and activity of the Bcd-dependent proximal P2 enhancer element of *hunchback*, an early target of Bcd, under experimentally modulated Bcd input and nucleosome stability at the enhancer (80). Bcd engages

128 in a concentration-dependent competition with nucleosomes to 'open' otherwise inaccessible
129 regulatory regions. Bcd activity thereby effectively breaking symmetry along the AP axis at the
130 level of chromatin accessibility, well before the bulk of zygotic genome activation and pattern
131 formation occurs (81).

132 In Clogmia, knockdown of the maternal *Cal-opa* transcript by isoform-specific RNA interference
133 (RNAi) results in a symmetrical double abdomen phenotype, and posterior mis-expression of
134 *Cal-opa* by mRNA injection prior to blastoderm formation causes a bicephalic (double head)
135 phenotype. Early Cal-Opa activity is therefore both necessary and sufficient for establishing
136 head-to-tail polarity of the Clogmia embryo. An alternative zygotic transcript isoform of
137 Clogmia *odd-paired* appears during blastoderm cellularization at NC14 and exerts a conserved
138 function in the late segmentation gene network but has no role in establishing head-to-tail
139 polarity of the embryo (3, 82). No overlapping functions have been observed between the
140 maternal and the zygotic *Cal-opa* isoforms, which encode nearly identical proteins, suggesting
141 that Clogmia's AD functions exclusively before NC14, in contrast to Bcd, which remains
142 essential for AP patterning until the end of NC14 (65). However, in Drosophila, zygotic Opa
143 pioneers chromatin accessibility in late NC14 embryos to advance the progression of
144 segmentation and dorsoventral patterning (78, 83). We therefore asked whether the maternal *Cal-*
145 *opa* gradient of Clogmia breaks axial symmetry by driving chromatin accessibility at gene loci
146 that are transcribed in the anterior blastoderm before the beginning of NC14 and whether these
147 genes are homologous to Bcd target genes.

148 To set the stage for addressing these questions we first assembled and annotated the Clogmia
149 genome and identified homologs of Drosophila's segmentation genes. We then characterized
150 chromatin accessibility in single wild-type embryos at NC11, NC12, NC13, and both early and
151 late NC14 to predict putative enhancer and promoter regions and to capture their stage-
152 dependent accessibility. We then assessed how chromatin accessibility patterns and transcription
153 are affected by *Cal-zld* activity. Finally, we examined how reduced maternal *Cal-opa* expression
154 affects chromatin accessibility and TF gene expression during NC12 and NC13. Our results
155 reveal groups of gene loci with similar dynamic changes in chromatin accessibility, point to
156 direct roles of Cal-Zld in opening and closing chromatin, and provides evidence for a role of
157 maternal Cal-Opa in opening chromatin during early blastoderm stages at a small set of loci with
158 anterior expression, including homologs of *sloppy-paired* and *homeobrain.* We conclude that the
159 conserved feature of anterior determination in Clogmia and Drosophila is symmetry breaking at
160 the level of chromatin accessibility, even though specific target genes may differ between the
161 two species.

162

163 **Results**

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

166 We generated a high quality *de novo* genome assembly of *Clogmia albipunctata* with 6
167 chromosome-size scaffolds (**Figure 1, Table 1, S1 Appendix**), consistent with the known
168 chromosome number in this species (84, 85). The total length of the chromosome-size scaffolds
169 is 304.3 Mb, close to genome size estimates based on flow cytometry data (316.6 Mb) (86). An
170 additional 268 small scaffolds (< 1 MB) contained mostly repetitive sequences. We annotated
171 this new reference genome using RNA-seq transcripts from embryonic, larval, pupal, and adult

172 stages (**S1 Table**), protein sequences, and *ab initio* gene prediction software, and identified
173 15,047 protein coding genes. The annotated genome recovered 88%-94% of Complete
174 Benchmarking Universal Single Copy Orthologs (BUSCO) (**Table 2**), in line with recently
175 published dipteran genomes (87-89). While synteny between the chromosomes within the
176 Cyclorrhapha clade can be highly conserved (time of divergence ~145 million years ago) (88),
177 synteny between the chromosomes of *D. melanogaster* and *C. albipunctata* (divergence time
178 ~250 million years ago) was essentially lost (**Figure 1**).

179 Next, we interrogated the genome for the presence of predicted early segmentation genes. As
180 expected, no *bicoid* ortholog was found but we recovered two putative Hox class 3 genes to
181 which *bicoid* belongs (90), including the previously reported homolog of *zerknüllt* (*Cal-zen*) (91)
182 on scaffold 7 and a more diverged potential *Cal-zen* paralog in position 3 of the Hox gene
183 complex on scaffold 3 (evm.TU.scaffold_3.2943.1.65be5b92). Our annotation also included
184 homologs of *Drosophila melanogaster*'s early segmentation genes, including gap genes (*tailless*,
185 *huckebein*, *orthodenticle/ocelliless*, *empty spiracles*, *buttonhead*, *hunchback*, *Krüppel*, *giant,*
186 *knirps*), pair-rule genes (*even-skipped*, *fushi tarazu*, *odd-skipped*, *runt*, *sloppy-paired*, *paired,*
187 *hairy, odd-paired*), and other important regulators of the segmentation network (*caudal*,
188 *Dichaete*) (19, 21). Some of these genes have more than one copy in the Clogmia genome,
189 including 2 copies of *hunchback* (*Cal-hb1, Cal-hb2*), *knirps/knirps-like* (*Cal-kni/knrl1, Cal-*
190 *kni/knrl2*), *orthodenticle* (*Cal-otd1, Cal-otd2*), *even-skipped* (*Cal-eve1, Cal-eve2*), *tailless* (*Cal-*
191 *tll1, Cal-tll2*) and *caudal* (*Cal-cad 1, Cal-cad2*), 3 copies of *sloppy-paired1/2/fd19B* (*Cal-slp1,*
192 *Cal-slp2, Cal-slp3*), and 4 copies of *empty spiracles* (*Cal-ems1, Cal-ems2, Cal-ems3, Cal-ems4*)
193 (**Figure 1, S2 Table**). Such gene duplications are a potential source of redundancies in
194 Clogmia's segmentation gene network (see below).

195

196 **Dynamic patterns of chromatin accessibility during consecutive blastoderm stages suggest**
197 **coordinated regulation of associated genes**

198 To identify accessible chromatin before and during the main wave of zygotic genome activation,
199 we performed single embryo ATAC-seq at NC11, NC12, NC13, NC14 (before cellularization),
200 and Late NC14 approximately 20 minutes before gastrulation. The embryos were staged *in vivo*
201 using developmental time after egg activation at 25ºC and morphological criteria as previously
202 described (38) and validated under our laboratory conditions. At least three replicates were
203 performed for NC11 though NC14, and two replicates were performed for Late NC14. We
204 pooled the ATAC-seq data from all stages to create a master peak list of 32157 open chromatin
205 regions (peaks), using MACS2 (92). We also computed the number of peaks for each
206 developmental stage excluding stage-specific peaks that were not recovered as significant in the
207 master peak list. In this analysis, we found that the overall number of peaks increases with each
208 nuclear cycle, suggesting that the chromatin landscape is continuously changing (**Table 3**). Since
209 promoter accessibility does not necessarily indicate gene expression (93), we divided the
210 accessible chromatin regions into transcription start site (TSS) proximal regions (within 500 bp
211 of TSS) and TSS distal regions (>500 bp) to facilitate the distinction of open promoters and other
212 open cis-regulatory elements. ATAC-seq peaks that only mapped to intergenic or intronic
213 sequence were treated as candidate enhancers or silencers (henceforth collectively referred to as
214 enhancers). The relative proportion of candidate enhancers increased by about 10% from NC11
215 (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 (94, 95).

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 (96). This analysis identified a total of 6930 dynamically regulated peaks (adjusted p-value ≤ 0.05 and $|\log_2 \text{Fold Change}| \geq 1$; 21.6% of all 32157 peaks) and enabled us to quantitatively determine the behavior of each peak (i.e., maintaining, gaining or losing accessibility) between consecutive developmental stages. We then clustered dynamic peaks into groups based on shared changes in accessibility from stage to stage and delineated 21 dynamic groups (**S2 Figure**). Nearly 80% (5403 / 6930) of the dynamic peaks fell into one of 8 groups, which reveal the major accessibility trends we observe in the data (**Figure 2E)**. Many of the dynamic peaks closed (14.6%; Groups 5,13) or opened (~20.4%; Groups 8,12) gradually over time. The remaining dynamic peak groups exhibited more complex dynamics between NC11 and Late NC14, commonly with an inflection at NC13 (e.g., Groups 3,1,15,14; see also **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.

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

Previous studies in Drosophila have shown that temporal changes in chromatin accessibility are regulated by the interplay of transcription factors that can interact with nucleosome-bound DNA (pioneer factors) and transcription factors that have limited or no ability to initiate chromatin accessibility (49, 56). The earliest changes in chromatin accessibility are primarily driven by the pioneer factor Zld (46, 54, 58, 94, 97). As zygotic genome activation accelerates, additional pioneer factors, such as GAF (63), CLAMP (61, 62), and Opa (78, 83) are required for maintaining, modulating, or gaining chromatin accessibility. In Clogmia, a motif enrichment analysis of peaks with the MEME software package (98, 99) revealed an enrichment of a Zelda-binding motif (5'-CAGGTA), GA-rich motifs recognized by GAF and CLAMP (63, 100-102), in addition to $(CA)_n$ motifs recognized by Combgap (Cg) (103), and others (**S3 Figure**). Opa-binding motifs were not recovered in this analysis. These results are comparable to findings in Drosophila (78, 81, 83) and suggest that motif enrichment within chromatin accessible at ZGA is very similar between Drosophila and Clogmia.

Homologs of *zelda* have been identified in many insects and some crustaceans (104), and several transcriptomic studies suggest that Zld's function as pioneer transcription factor during zygotic genome activation is conserved across insects (105-109). However, these comparative studies did not directly test for Zld-dependent effects on chromatin accessibility. We identified Clogmia *zld* (*Cal-zld*) (**S4 Figure**) and confirmed high levels of *Cal-zld* expression at preblastoderm and blastoderm stages (**S5 Figure**). Loss of Zld in *Drosophila* results in a failure to activate a large cohort of early zygotic genes, including those associated with the process of cellularization (46). To examine phenotypic effects of *Cal-zld* depletion, batches of ~30 embryos were injected with *Cal-zld* dsRNA (double-stranded RNA) within the first hour of development (NC2/3). Injected embryos were monitored *in vivo* and allowed to develop until the end of NC14 (roughly 7-7.5 hours after injection). Across three independent replicates, a subset of injected embryos reached

260 NC14 but failed to cellularize and developed abnormally (**S1 movie**). This phenotype is
261 comparable to the phenotype of *zld* mutant Drosophila embryos (**S2 movie**) (58) and suggests
262 that the role of Zld for the regulation of early zygotic genes is conserved in Clogmia. More than
263 half of the injected embryos (33/59) did not reach the blastoderm stage while early mortality of
264 an uninjected control group was ~27% (6/22), but it is unclear whether this difference reflects a
265 requirement for Zelda earlier in development or an artifact of the injection procedure.

266 To assess *Cal-zld*'s role in shaping chromatin accessibility and zygotic transcription, we adapted
267 our ATAC-seq protocol to isolate RNA from single-embryo ATAC-seq library preparations of
268 Late NC14 *Cal-zld* RNAi embryos (see *Materials and Methods*). To identify RNAi embryos
269 with maximally reduced *Cal-zld* mRNA levels, we used the RNA-seq data to measure *Cal-zld*
270 expression and selected embryos with knockdown efficiencies greater than 75% (n= 4/8, **S6**
271 **Figure** and see *Materials and Methods*). We then compared gene expression and ATAC-seq
272 peaks of these *Cal-zld* RNAi embryos to those of uninjected control embryos using DESeq2.
273 Loss of *Cal-zld* affected 2230 genes of which 865 (38.8%) were downregulated and 1365
274 (61.2%) were upregulated. We then compared genes differentially expressed in *Cal-zld* RNAi
275 embryos to genes we found to be expressed maternally, zygotically, or both (**Figure 3A, S7A**
276 **Figure** and *Materials and Methods*). 691 genes were classified as zygotically expressed of which
277 209 (30.2%) were downregulated in *Cal-zld* RNAi embryos and 63 (9.1%) were upregulated
278 (corresponding to 24.2% of all 865 downregulated and 4.6% of all 1365 upregulated genes,
279 respectively). 6658 genes were classified as expressed maternally and zygotically of which 593
280 (8.9%) were downregulated and 1232 (18.5%) were upregulated. Finally, of the 227 genes
281 classified as maternally expressed, only 5 (2.2%) were found to be downregulated and 14 (6.2%)
282 upregulated in *Cal-zld* RNAi embryos. Since the strictly zygotic genes were ~3 times more often
283 downregulated than upregulated (209 versus 63), our results suggest that *Cal-zld* plays an
284 important role in activating zygotic genes (**Figure 3A),** consistent with the role of Zld during
285 zygotic genome activation in Drosophila embryos (46, 54). Changes in the levels of maternal
286 transcripts in response to *Cal-zld* RNAi could reflect indirect posttranscriptional roles of *Cal-zld*
287 in their regulation.

288 At the level of chromatin accessibility, *Cal-zld* depletion affected 4102 of all 32157 peaks
289 (12.8%). 2399 of them lost accessibility (58.4%) and 1703 gained accessibility (41.5%). The
290 majority (67%) of the peaks mapped to intergenic or intronic sequences (**Figure 3B)**. Depletion
291 of *Cal-zld* mRNA affected both dynamic peaks and non-dynamic peaks (**S7B Figure**). Dynamic
292 peaks were affected by *Cal-zld* knockdown in different ways (**Figure 3C**). Peaks in groups that
293 progressively gained accessibility in wild-type embryos (Groups 8, 12, 15 and 14) were more
294 likely to be *Cal-zld* dependent and to close in *Cal-zld* RNAi background (**Figure 3C, *left*** and
295 **S8A Figure**). This observation is consistent with an important role of Cal-Zld in opening new
296 chromatin regions during zygotic genome activation. Additionally, our data suggest that Cal-Zld
297 also plays an important role in closing chromatin regions during zygotic genome activation (**S8B**
298 **Figure**). For example, peaks in dynamic groups whose accessibility decreased in wild-type with
299 each nuclear cycle or that closed after NC13 (Groups 5 and 1) were more likely to require Cal-
300 Zld to close (**Figure 3C, *right*** and **S8B Figure**).

301 To assess if the changes in accessibility we observed were a direct consequence of *Cal-zld*
302 knockdown, we performed two types of motif analyses. We first performed a motif enrichment
303 analysis with MEME on *Cal-zld* sensitive peaks to identify motifs enriched centrally within 50
304 bp of the peak summit (**S9 Figure**). The Zld motif was enriched in peaks that lost accessibility

following *Cal-zld* knockdown, consistent with a direct role of Cal-Zld in nucleosome depletion. A Zld motif was not enriched in peaks that gained accessibility following *Cal-zld* knockdown (**Figure 3D** and **S9 Figure)**. The enrichment of Zld motifs in only one class of *Cal-zld* sensitive peaks may reflect the prevalence of Cal-Zld binding in those regions. In support of this hypothesis, we note that both motif affinity and occupancy positively correlate with Zld's pioneering activity in Drosophila (56, 110). Next, we extended the search for Cal-Zld motifs to the full width of each peak, using Zld motifs derived from our MEME analysis in Clogmia (5-CAGGTA; cf. **S3 Figure)** and experimental data in *Drosophila melanogaster* (Fly Factor Survey) (111). Overall, 2857 of the 4102 *Cal-zld* sensitive peaks contained at least one Zld motif within the full peak width. Consistent with the motif enrichment analysis, a Zld motif was prevalent in peaks whose accessibility was lost following *Cal-zld* depletion (81%, 1952/2399 peaks). However, we also observed Zld motifs in peaks whose accessibility increased following *Cal-zld* knockdown (53%, 905/1703 peaks). Therefore, these regions may close due to Cal-Zld activity directly. Taken together, our findings suggest that Cal-Zld plays a central role in both opening and closing chromatin accessibility in pre-gastrulation embryos.

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

Having established that *Cal-zld* regulates chromatin accessibility in Clogmia embryos, we asked whether maternal *Cal-opa* affects chromatin accessibility during early stages of axial pattern formation. Since *Cal-opa* exerts its symmetry-breaking function in axis specification prior to the onset of zygotic *Cal-opa* expression during blastoderm cellularization at mid NC14 (3), we were able to examine the function of *Cal-opa* without confounding effects of zygotic *Cal-opa* activity. To assess the impact of maternal Cal-Opa activity on chromatin accessibility, we performed ATAC-seq and RNA-seq on single *Cal-opa* RNAi embryos at NC12 (6 replicates) and NC13 (8 replicates). In parallel, we processed stage-matched embryos that were prepared for injection under oil but not injected (alignment controls; 4 replicates for each developmental stage), and embryos that were injected with dsRNA of the extraneous gene *DsRed* (injection controls; 3 replicates for each stage). To identify RNAi embryos with maximally reduced *Cal-opa* mRNA levels, we used the RNA-seq data to measure *Cal-opa* expression and selected embryos with knockdown efficiencies greater than 75% (**S10 Figure** and see *Materials and Methods*). We identified 4 embryos at NC12 and 5 embryos at NC13 with strong *Cal-opa* knockdowns and assessed changes in chromatin accessibility in these embryos by performing DESeq2 analysis with the alignment control group as reference.

At NC12, very few changes in chromatin accessibility were observed in response to *Cal-opa* RNAi. Out of the 86 differentially accessible regions, 79 (91.9%) lost accessibility and 7 (8.1%) gained accessibility (**Figure 4A**). At NC13, depletion of *Cal-opa* had a more pronounced effect on chromatin accessibility, with 250/292 (85.6%) peaks losing accessibility and 42/292 (14.4%) gaining accessibility **(Figure 4B**). Most of the differentially accessible regions mapped to putative enhancer regions (**S11 Figure**). We also performed a motif enrichment analysis using the MEME suite on all peaks as well as the peaks that either lost or gained accessibility in *Cal-opa* RNAi embryos at either NC12 or NC13 (**Figure 4C**). In regions that lost accessibility in *Cal-opa* RNAi embryos, the most enriched motifs were GAF/CLAMP and Opa binding motifs. In addition, we recovered motifs that have not been shown to be bound by a TF in Drosophila.

350 Regions that gained accessibility were enriched in DNA binding motifs of GAF/CLAMP,
351 Abdominal-B (Abd-B), Twist (Twi), and two motifs with no associated transcription factor.

352 Next, we determined the number of *Cal-opa* sensitive peaks that contained putative Cal-Opa
353 binding sites, including Opa motifs derived from our MEME analysis and a MEME analysis of
354 ATAC-seq data from Drosophila (78). At NC12, 63 of the 79 peaks losing accessibility and none
355 of the 7 peaks gaining accessibility in response to reduced *Cal-opa* activity contained at least one
356 Opa motif (**S2 Appendix**). At NC13, 112 of the 250 peaks losing accessibility and 6 of the 42
357 peaks gaining accessibility in response to reduced *Cal-opa* activity contained at least one Opa
358 motif (**S3 Appendix**). These results support the hypothesis that Opa directly promotes chromatin
359 accessibility during the early phase of zygotic genome activation.

360

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

363 The early zygotic segmentation gene network of dipterans is dominated by TF-encoding genes,
364 such as the gap genes and pair-rule genes (19, 21, 29, 39, 112-115). Since Bcd breaks axial
365 symmetry by targeting these genes in a concentration-dependent manner, we focused our search
366 of Cal-Opa targets on TF genes that are expressed during blastoderm formation. Using the RNA-
367 seq data from the embryos that were used to identify regions of differential chromatin
368 accessibility, we identified 84 differentially expressed genes at NC12, including 11 transcription
369 factor genes, all of which were downregulated upon *Cal-opa* knockdown (**Figure 5A** and **S2**
370 **Appendix**). At NC13, we identified 1047 differentially expressed genes, including 116 putative
371 transcription factor genes of which 67 were downregulated and 49 were upregulated in *Cal-opa*
372 RNAi embryos (**Figure 5B** and **S3 Appendix**). We then asked which of the differentially
373 expressed transcription factor genes are located near open chromatin regions (ATAC-seq peaks)
374 with an Opa motif. This analysis was restricted to peaks within a 40 kb region centered on each
375 gene body of interest and further narrowed the list of potential key target genes of Cal-Opa to 10
376 genes at NC12 and 72 genes at NC13 (**S2 Appendix and S3 Appendix)**. At NC12, 4 of the 10
377 candidate genes were associated with at least one Cal-Opa dependent ATAC-seq peak, including
378 *Cal-opa,* and homologs of *homeobrain* (*Cal-hbn*) and *sloppy-paired* (*Cal-slp1*, *Cal-slp2*) (**Figure**
379 **6, S3 Table,** and **S12 Figure**). At NC13, we thereby identified *Cal-opa, Cal-slp1*, *Cal-slp2* and
380 *Cal-hbn* in addition to Clogmia orthologs of *Dichaete* (*Cal-D*), *Krüppel* (*Cal-Kr*), and an
381 unidentified C2H2 zinc finger gene among the 72 candidates (**Figure 7, S3 Table, S12 Figure**
382 **and S13 Figure**).

383 We tested the robustness of these results by repeating the analysis using the injection control
384 group as reference. At NC12, we recovered *Cal-opa*, *Cal-hbn, Cal-slp1,* and *Cal-slp2* as
385 candidate targets, but only three of them (*Cal-opa*, *Cal-hbn, Cal-slp1*) were associated with Cal-
386 Opa dependent ATAC-seq peaks with Cal-Opa binding sites. In the case of *Cal-slp2*, the
387 reduction of these peaks was insignificant (**S3 Table** and **S2 Appendix**). At NC13, we recovered
388 *Cal-hbn, Cal-slp1 and Cal-slp2* (**S3 Table** and **S3 Appendix**). *Cal-opa*, *Cal-Kr*, *Cal-D* and the
389 unidentified C2H2 zinc finger gene were not recovered in this analysis at NC13 because they did
390 not meet the threshold for statistical significance (settings identical for NC12 and NC13). We
391 therefore focused on Clogmia's *homeobrain* and *sloppy-paired* homologs.

392 *Cal-hbn* and *Cal-slp1* expression was strictly zygotic and detected during and after NC11 when
393 both genes were expressed in non-overlapping anterior domains (**Figure 8A, B**). Maternal *Cal-*

394 *slp2* transcript was previously found to be distributed in a shallow anterior-to-posterior gradient
395 and may contribute to the embryo's head-to-tail polarity in subtle ways (3). Zygotic *Cal-slp2*
396 expression during and after NC11 was detected in an anterior stripe (**Figure 8C**). A third *sloppy-*
397 *paired* homolog that we identified in the Clogmia genome (evm.TU.scaffold_5.987; **S14 Figure**
398 **and S15 Figure**) was expressed at low levels at NC12 and NC13 and robustly at NC14. The
399 expression of this gene was reduced in *Cal-opa* RNAi embryos but associated ATAC-seq peaks
400 did not contain Opa-binding motifs. Taken together, our findings suggest that *Cal-hbn*, and *Cal-*
401 *slp1* and *Cal-slp2* are direct early targets of Clogmia's AD and that Cal-Opa serves as their
402 transcriptional activator.

403

404 **Discussion**

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

406 In this study, we have established the resources for dissecting developmental gene networks of
407 early Clogmia embryos by providing an annotated chromosome-level genome sequence for our
408 inbred laboratory strain and by describing chromatin accessibility of precisely staged consecutive
409 blastoderm stages. We have used these resources to examine how the actives of *zelda* and
410 maternal *odd-paired* shape chromatin accessibility and gene expression in Clogmia. We found
411 that, as in Drosophila, the period spanning zygotic genome activation is accompanied by large-
412 scale changes in chromatin accessibility. However, there are notable differences between
413 Clogmia and Drosophila in the early patterns of establishment and maintenance of accessible
414 states before NC14. In Clogmia, the chromatin landscape is highly dynamic in that gains as well
415 as losses occur during the nuclear cleavage cycles of the blastoderm stage. In Drosophila, only
416 gains in accessibility are observed until widespread ZGA and lengthening of the cell cycle at
417 NC14 (76, 78, 79, 81, 116). This difference could be a consequence of cell cycle timing. In
418 Clogmia, NC11 through NC13 last ~3x longer than in Drosophila (38), in which NC11 only lasts
419 ~10 minutes, NC12 ~12 minutes, and NC13 ~21 minutes (37). The much shorter cell cycle times
420 of Drosophila present significant limitations to establish and maintain chromatin states necessary
421 for transcriptional activity (81). Assuming that kinetic rates for nuclear import, DNA replication,
422 and DNA binding of transcription factors is similar between Clogmia and Drosophila, the
423 significantly longer syncytial interphase times in Clogmia could alleviate constraints on the
424 initiation of zygotic transcription. However, the timing of AP pattering between the two species
425 appears comparable (112, 117), suggesting that the timing of ZGA might also be conserved.
426 Alternatively, the longer syncytial interphase times in Clogmia could alleviate constraints on
427 maintaining or silencing active chromatin states. In Drosophila, Polycomb group (PcG) proteins
428 begin to establish heritable, transcriptionally silenced chromatin states during NC14 (118), but in
429 Clogmia, the longer nuclear cleavage cycles may provide time for the maturation of repressive
430 histone marks at accessible chromatin regions before NC14. This scenario would be consistent
431 with the observed requirement of Cal-Zld for closing chromatin regions, as discussed in the next
432 section.

433

434 **A unique role of Zld in regulating chromatin accessibility transitions in the Clogmia**
435 **embryo**

436 We demonstrate that reduced *Cal-zld* expression results in widespread losses of chromatin
437 accessibility, consistent with a conserved role for Zld to pioneer chromatin accessibility within
438 the context of zygotic genome activation. This study also describes a function of Cal-Zld in
439 closing open chromatin, consistent with the unique chromatin dynamics of Clogmia.
440 Specifically, many regions that lose accessibility during the syncytial cleavage divisions do so in
441 a Zld-dependent manner. Based on the frequent occurrence of Zld motifs within this class of
442 regions (53%), many of these effects are likely direct. We therefore propose that Cal-Zld
443 functions not only to open and maintain chromatin accessibility but also to regulate the closure
444 of many regions during nuclear cleavage cycles. This contrasts with Drosophila, in which Zld
445 almost exclusively drives gains in chromatin accessibility during nuclear cleavage cycles (73, 78,
446 97, 119, 120). However, two recent studies found that during ZGA, Zld activity directs how
447 repressive chromatin marks are deposited. Loss of *zld* results in context-specific losses and gains
448 of H3K27me3 within specific PcG domains (118). Zld also enables histone acetylation at
449 promoters and enhancers through its interaction with Chip, Drosophila's homolog of LIM-
450 domain-binding protein 1 (Ldb1), which in turn recruits the histone acetyltransferase CBP (121).
451 The longer nuclear cleavage cycles in Clogmia might enable similar roles of Cal-Zld before
452 NC14. In this case, Clogmia and Drosophila may differ in the timing of when specific histone
453 marks are deposited and affected by Zld depletion. Differences in the activities of Zld and Cal-
454 Zld could also reflect differences in their protein structures or isoforms. For example, Cal-Zld
455 contains an additional zinc-finger that has been eroded in Drosophila (**S4 Figure**) (104).
456 Additionally, stage-specific isoforms of Zld have been described in Drosophila, including a
457 variant lacking the C-terminal three zinc-fingers (122-125). This isoform functions as a dominant
458 negative form in cell culture although flies in which the expression of this isoform is suppressed
459 are viable and fertile (124). Therefore, the molecular underpinnings of the distinct functional
460 readouts of Zld and Cal-Zld could be complex and likely involve both direct and indirect
461 mechanisms.

462

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

464 Bcd contributes to the activation of many early segmentation genes; however, its role in the
465 posterior embryo is much more subtle than in the anterior embryo, where all expression domains
466 directly or indirectly depend on Bcd. Bcd target enhancers have been classified based on how
467 their accessibility depends on Bcd and Zld (73). The accessibility of Bcd target enhancers that
468 drive expression in the posterior of the embryo is established independent of Bcd (e.g., *Krüppel*
469 CD1, *knirps* posterior, *giant* posterior). However, enhancers that open in a Bcd-dependent
470 manner are associated with genes expressed in the anterior of the embryo (e.g., *orthodenticle*,
471 *empty spiracles*, *buttonhead*, *giant*, *knirps*, *hunchback*, *sloppy-paired 1*, *fd19B*, *even-skipped*,
472 *paired*, *cap'n'collar*, *huckebein, homeobrain*). Moreover, many correspond to enhancers that
473 drive reporter gene expression in the anterior embryo including *giant* anterior (-6), *giant* anterior
474 (-10), *knirps* anterior, *orthodenticle*, *buttonhead, even-skipped* stripe 1/5, *hunchback* P2,
475 *hunchback* shadow enhancers, *eve-skipped* stripe 2, an early *paired* enhancer, *cap'n'collar*, and
476 *huckebein* (reviewed in (73)). The majority of Bcd-dependent anterior expression domains first
477 appear during NC13, with the noteworthy exception the P2 transcript of *hunchback*. The
478 promoter of this transcript is in an open formation as early NC10 (81) and drives Bcd-dependent
479 expression throughout the anterior half starting at NC11(47, 69, 126).

480 Clogmia's anterior determinant is not required for *hunchback* expression before NC13 (**S16A**
481 **Figure** and **S16B Figure**), and maternal *hunchback* activity does not contribute to embryo
482 polarity given that knockdown of maternal *Cal-opa* transcripts in early embryos results in
483 symmetrical double abdomens (3). However, Cal-Opa initiates symmetry-breaking at the level of
484 chromatin accessibility and gene expression before NC12, notably at the loci of TF gene
485 homologs of *sloppy-pai*red and *homeobrain*. These genes are expressed in complementary,
486 evolutionarily conserved domains across the anterior half of the embryo. We note that these
487 genes are among the Bcd targets discussed above; both loci contain Bcd-bound enhancers whose
488 accessibility is sensitive to the concentration of Bcd. Mechanistically, Cal-Opa could prime
489 accessibility at target loci through a canonical pioneering mechanism which requires binding of
490 nucleosome occupied DNA and promoting nucleosome instability (49, 50, 58, 62, 127, 128). A
491 vertebrate homolog of Opa, Zic3, was found to bind nucleosomal DNA *in vivo*, suggesting that
492 Cal-Opa may pioneer chromatin directly (129). Alternatively, Cal-Opa binding may compete
493 with nucleosomes to bind unoccupied DNA during each nuclear cleavage cycle (130), similar to
494 Bcd (56, 73, 80).

495 Maternal Cal-Opa may target additional genes. Rigorously testing this possibility will require
496 additional genetic tools or reagents to determine where maternal Cal-Opa binds *in vivo* and
497 whether its target chromatin regions that become accessible independent of maternal Cal-Opa
498 activity. We note that maternal Cal-Opa activity is expected to cease before the onset of zygotic
499 *Cal-opa* expression during NC14. Bcd is still active at NC14, and because of this difference, the
500 AD of Clogmia may target fewer genes than Bcd. We examined this possibility by individually
501 analyzing our data on Clogmia orthologs of Bcd target genes other than *sloppy-paired* and
502 *homeobrain*. The expression of Clogmia's two *hunchback* orthologs (*Cal-hb1*, *Cal-hb2*) was
503 reduced in *Cal-opa* RNAi embryos at NC13. We also observe a reduction in a few of the
504 associated ATAC-seq peaks at these loci, but these regions did not include Opa motifs (**S16C**
505 **Figure** and **S16D Figure**). Likewise, the expression of one of Clogmia's two orthologs of
506 *orthodenticle* (*Cal-otd1*) was reduced at NC13 in *Cal-opa* RNAi embryos, but this reduction was
507 not associated with *Cal-opa* sensitive peaks containing Opa motifs (**S17A Figure**). The second
508 ortholog (*Cal-otd2*) was not expressed at NC12 and NC13. In the case of Clogmia *buttonhead*
509 (*Cal-btd*), we identified a *Cal-opa* sensitive peak with an Opa motif, but the observed reduction
510 of *Cal-btd* expression was not significant (**S17B Figure**). *Cal-btd* expression could be under the
511 control of AD-dependent and AD-independent enhancers, as previously reported for
512 *orthodenticle* in Drosophila (131). Clogmia orthologs of *giant* (*Cal-gt*), *knirps* (*Cal-kni/knrl1,*
513 *Cal-kni/knrl2*), and *empty spiracles* (*Cal-ems1-4*) were not expressed at NC12 or NC13, and
514 none of these loci had significant changes in chromatin accessibility following knockdown of
515 *Cal-opa*. The ortholog of *huckebein* (*Cal-hkb*) was not expressed at NC12 or NC13, either;
516 however, its locus contained one Opa motif positive peak that lost accessibility following *Cal-*
517 *opa* knockdown at NC13. The expression of Clogmia's *cap'n'collar* (*Cal-cnc*) and *paired* (*Cal-*
518 *prd*) was unaffected by *Cal-opa* RNAi, and the ATAC-seq peaks at their loci were largely *Cal-*
519 *opa* insensitive. Only one small intronic peak within the Cal-*cnc* locus displayed reduced
520 accessibility in *Cal-opa* RNAi embryos at NC13. The loci of *even-skipped* orthologs (*Cal-eve1,*
521 *Cal-eve2*) contain ATAC-seq peaks with Opa motifs that were insensitive to *Cal-opa* RNAi, and
522 the expression of *Cal-eve2* was elevated in *Cal-opa* RNAi embryos at NC13. Taken together,
523 these observations point to significant differences in the target genes of Cal-Opa and Bcd, and
524 suggest that *sloppy-paired* and *homeobrain* are key zygotic initiators of AP patterning - a role

525 that is likely complemented by the distribution of maternal *Cal-slp2* transcript in an AP gradient
526 (3).

527

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

529 Neither *homeobrain* nor *sloppy-paired* has received much attention as target genes of Bcd.
530 However, their identification in Clogmia as the only TF genes with strictly AD-dependent
531 expression raises the question of whether they serve a more broadly conserved role in the
532 initiation of anterior pattern formation. In Drosophila, *homeobrain* mutant embryos lack the
533 labrum and anterior portions of the brain (132-135), indicating that *homeobrain* is essential for
534 specifying the most anterior portion of the embryo. A more extreme *homeobrain* loss-of-function
535 phenotype has been observed in the beetle *Tribolium castaneum*, where this gene plays an
536 important zygotic role in establishing the embryo's anterior-posterior polarity (136). In
537 Tribolium, knockdown of *homeobrain* results in loss of the entire head and – in extreme cases –
538 duplications of the tail end. The duplicated abdomen was defective, and thoracic structures
539 persisted in these embryos. However, complete double abdomens could be obtained by
540 simultaneously suppressing the expression *Tc-zen1*, thereby preventing the anterior repression of
541 *caudal*, an important gene for posterior patterning (137, 138)

542 The *sloppy-paired* gene, a member of the *forkhead* domain gene family (139), is best known for
543 its role as a pair-rule segmentation gene (21, 40), but it is initially expressed in a head domain of
544 the syncytial Drosophila embryo, like its lineage-specific paralogs *sloppy-paired 2* and *fd19B*
545 (138, 140-142). The partially redundant functions of *sloppy-paired* paralogs, which are common
546 across insects (**Figure S2.14** and **Figure S2.15**), may have obscured a fundamental role of
547 *sloppy-paired* activity in initiating anterior pattern formation in Drosophila. This also applies to
548 the two *sloppy-paired* homologs of Tribolium, in which only the gnathocephalon (mandibular,
549 maxillary, and labial segment) and even-numbered trunk segments are lost when *Tca-slp1* is
550 knocked down (143, 144). We therefore propose that along with *homeobrain*, *sloppy-paired* may
551 play a more widely conserved role in establishing the AP axis in insect embryos.

552

553 **Materials and Methods**

554 **Clogmia culture and embryo collection**

555 The Clogmia culture was maintained in 100mm x15mm Petri Dishes (Fisher Scientific
556 FB0875712) on moist, compressed cotton (Jorvet SKU: J0197) sprinkled with parsley leaf
557 powder (Starwest Botanicals SKU: 209895-5). For embryo collections, ~2-day old females were
558 transferred to round 250 mL glass bottles with a ~1 inch bottom layer of moistened and
559 compressed cotton and mated at room temperature for 2-3 days. Female ovaries were dissected
560 with a pair of fine forceps (Dumont 5) and eggs released into water for egg activation and
561 embryogenesis. For staged egg collections, embryos were released into flat bottom glass 100mm
562 x15mm Petri Dishes (Corning CM: 101425255) containing 15ml of water. Care was taken to
563 disperse the eggs on the bottom of the dish as crowding can delay development. Petri Dishes
564 were covered with a lid and moved to an incubator set at 25° C until the desired stage. Egg
565 harvesting post activation was on average 47 minutes for NC2/3, 261 minutes for NC11, 292
566 minutes for NC12, 332 minutes for NC13, 401 minutes for NC14, and 465 minutes for Late
567 NC14.

568

569 **Embryo fixation**

570 *Clogmia* Embryos were dechorionated using a 10% dilution of commercial bleach (8.25%
571 sodium hypochloride) for 90 seconds. Dechorionated embryos were fixed in a 50 mL falcon
572 tube, using 20 mL of boiling salt/detergent-solution (100 µL 10% triton-X; 500 µL 28% NaCl;
573 up to 20 mL of water). The mixture was gently swirled for 10 seconds. To stop the heat fixation,
574 30 mL of deionized water was added. Embryos were transferred to 1.5 mL reaction vials and
575 devitellinized in a 1:1 mixture of n-heptane and methanol by vigorous shaking. This step was
576 followed by 3 washes with cold methanol (stored at -80C). Embryos were inspected under a
577 stereomicroscope and manual devitellinizations were performed as needed using sharp tungsten
578 needles in a 3% agar plate covered with methanol. Devitellinized embryos were stored in 100%
579 methanol at −20°C.

580

581 **Microinjections of embryos and RNAi efficiency**

582 *Clogmia albipunctata* embryo injection was done as previously described (3). We note here that
583 RNAi efficiency in Clogmia embryos varies (3). This variability could be due to technical
584 variables (e.g., injection needle, amount injected) or biological variables (e.g., differences in the
585 RNAi response between individual embryos, or the general health or genetic background of the
586 embryos or their mothers) or a combination of such effects. Therefore, we took care to control
587 the age of the females and selected all experimental embryos from females with healthy looking
588 ovaries and eggs.

589 To examine phenotypic effects of *Cal-zld* depletion, batches of ~30 embryos were injected with
590 *Cal-zld* dsRNA within the first hour of development (NC2/3). A subset of embryos from each
591 batch (3 replicates) were monitored *in vivo* and allowed to develop until the end of NC14
592 (roughly 7-7.5 hours after injection). We also quantified knockdown efficiencies of the *Cal-zld*
593 transcript in individual embryos by RT-qPCR and performed FC calculations of *Cal-zld* mRNA
594 levels with the average level in uninjected control embryos as reference (**S4 Table**). In one
595 quarter of the *Cal-zld* RNAi embryos (5/20), *Cal-zld* mRNA levels were reduced by ~30-90%.

596 To assess maternal *Cal-opa* RNAi efficiency, we injected embryos in the first hour of
597 development (NC2/3) with dsRNA of the isoform-specific first exon of the maternal *Cal-opa*
598 isoform. In three independently conducted experiments we reproduced the double abdomen
599 phenotype, confirming that our knockdown reagent was satisfactory. We also quantified
600 knockdown efficiencies of the *Cal-opa* transcript in individual embryos by RT-qPCR and
601 performed FC calculations of *Cal-opa* mRNA levels with the average level in uninjected control
602 embryos as reference (**S5 Table**). This analysis suggests that *Cal-opa* RNAi can nearly abolish
603 the *Cal-opa* transcript.

604

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

606 To assess *Cal-zld*'s role in shaping chromatin accessibility and zygotic transcription, we adapted
607 our ATAC-seq protocol to isolate RNA from single-embryo ATAC-seq library preparations of
608 Late NC14 *Cal-zld* RNAi embryos (8 replicates). In parallel, we processed stage-matched
609 untreated embryos that were allowed to develop under water in the dish used for egg activation
610 (wild-type controls; 5 replicates), and embryos that were prepared for injection under oil but not

injected (alignment controls; 5 replicates). The isolated RNA was used to generate RNA-seq libraries to measure the efficiency of the *Cal-zld* knockdown as well as transcriptional effects of the knockdown in individual embryos. To identify individual injected embryos with strong *Cal-zld* knockdowns, we quantified *Cal-zld* expression levels using the RNA-seq data and chose embryos with knockdown efficiencies greater than 75% (n= 4, **S6 Figure** and see *Fold change calculations*).

To assess the impact of maternal Cal-Opa activity on chromatin accessibility and zygotic transcription, we performed ATAC-seq and RNA-seq on single *Cal-opa* RNAi embryos at NC12 (6 replicates) and NC13 (8 replicates). In parallel, we processed stage-matched untreated controls that were allowed to develop under water in the dish used for egg activation (wild-type controls; 2 replicates for NC12 and 3 replicates for NC13), as well as controls that were prepared for injection under oil but not injected (alignment controls; 4 replicates for each developmental stage), and embryos that were injected with dsRNA of the extraneous gene *DsRed* (injection controls; 3 replicates for each stage). To identify RNAi embryos with maximally reduced *Cal-opa* mRNA levels, we used the RNA-seq data to measure *Cal-opa* expression and selected embryos with knockdown efficiencies greater than 75% (**S10 Figure** and see *Fold change calculations*). We identified 4 embryos at NC12 and 5 embryos at NC13 with strong *Cal-opa* knockdowns and assessed changes in chromatin accessibility in these embryos by performing DESeq2 analysis with the alignment control group as reference.

dsRNA synthesis

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

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

Cal-opa, maternal transcript (201 bp):
5'-CAGAGATGCATAATACGACTCACTATAGGGAGAAAACAATTGTGAAGTGCGACA
5'-
CAGAGATGCATAATACGACTCACTATAGGGAGACAAATTTCCAAACGATGACAGA

Cal-zld (a combination of two dsRNAs were used 579 bp and 668 bp):
5'-TAATACGACTCACTATAGGGAGAAGTCCCGCAATTGATACAGC
5'-TAATACGACTCACTATAGGGAGAAGGATGTTGGTGGACACTCC
5'-TAATACGACTCACTATAGGGAGACGGAATGGTGTGTGAAACAG
5'-TAATACGACTCACTATAGGGAGATTCGAGGGGTTATTGTCCTG

dsRed (273 bp):
5'-CAGAGATGCATAATACGACTCACTATAGATGCAGAAGAAGACTATGG
5'- CAGAGATGCATAATACGACTCACTATAGCTACAGGAACAGGTGGTG

653

654 **RT-qPCR**

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

657 Forward and reverse primer sequences for targets (lengths of amplicon in brackets):
658 *Cal-opa*, maternal transcript (112 bp):
659 5'- TTGCGCTTCATTCTGTCATCG
660 5'- CACTGAGGGCTGATTCTCGG
661
662 *Cal-zld* (138 bp):
663 5'-AAAGATGAACCACCGTGCGA
664 5'-CATTGATGGCAGTGGACCCT
665
666 *Cal-rpl35* (119 bp):
667 5'- CTGTCCAAAATCCGAGTCGT
668 5'- AGGTCGAGGGGCTTGTATTT
669

670 **Fluorescent HCR *in situ* hybridizations**

671 Gene specific probes and amplifier-fluorophore sets were designed and synthesized by
672 Molecular Instruments Inc (Los Angeles, CA). The amplifier-fluorophore set used was B1-647
673 for *Cal-slp1*, B2-546 for *Cal-hbn*, and B3-488 for *Cal-slp2*. Buffers (probe hybridization, probe
674 wash and amplification buffers) throughout the protocol were purchased from Molecular
675 Instruments Inc. The HCR *in situ* hybridizations were performed as described (145, 146).
676 Briefly, fixed embryos were rehydrated and washed 3 times in PBS-Tween20 (1x PBS, 0.1%
677 Tween-20). Embryos were permeabilized for 30 minutes at room temperature (100 ml
678 Permeabilization buffer: 100 µL 100% Triton X-100, 50 µL Igepal CA-630, 50 mg Sodium
679 Deoxycholate (Sigma D6750-25G), 50 mg Saponin (Sigma 47036-50G-F), 200 mg BSA
680 Fraction V (Sigma A5611-5G). Following embryos were pre-hybridized in probe hybridization
681 buffer for 30 minutes in a 37° C water bath. Once pre-hybridization was complete, fresh
682 hybridization solution containing gene specific probes were added and the embryos were left to
683 incubate overnight in a 37° C water bath. On the second day the hybridization solution was
684 removed, and the embryos were washed 4 times with pre-warmed probe wash buffer followed by
685 two washes in 5x SSCT (5x SSC, 0.1% Tween-20). An amplifier hairpin solution was prepared
686 following the manufacturer's instructions. Embryos were incubated with pre-warmed
687 amplification buffer for 30 minutes at room temperature before adding amplifier hairpin solution.
688 Embryos were left to incubate overnight at room temperature in the dark. The following day the
689 embryos were washed 5 times in 5x SSCT at room temperature. The embryos were then counter
690 stained with DAPI and mounted in Aqua-Polymount for imaging. Confocal imaging was
691 performed on the Zeiss LSM 900.

692

693 **Image Acquisition and Processing**

694 Confocal imaging was performed on the Zeiss LSM 900 using a Plan-Apochromat 20x/0.8 M27
695 objective. Pixel-density of 1024 x 1024. Excitation/Emission wavelengths (in nanometers) were:

493/517 for Alexa Flour 488, 653/668 for Alexa Flour 647, 401/422 for DAPI, and 548/561 for Alexa Flour 546. For each image a Z-stack of images was captured from the embryo's surface to approximately midway through the embryo. Then a max projection was created from this Z-stack using FIJI/Image J. Brightness was increased in Photoshop. No further adjustments to the images were made.

ATAC-seq library preparation

Single embryo ATAC-seq library preparations were done essential as described previously in *Drosophila melanogaster* (78, 81). Embryos developed in an incubator at 25ºC until the desired nuclear cycle. Nuclear cycle was confirmed by time after egg activation and the morphology was checked under water using a transmission light microscope (Zeiss Stemi 2000-C). Once at the desired nuclear cycle individual embryos were placed in the lid of a 1.5 ml low-retention Eppendorf tube and homogenized in 10 µl of cold lysis buffer (10 mM Tris-HCl pH 7.4; 10 mM NaCl; 3 mM MgCl2; 1% Igepal CA-630) (147), using a fire-sealed microcapillary tube (Drummond Microcap 25). Once homogenized, 40 µl of lysis buffer was added to the lid, and the body of the tube was carefully closed over the lid. Samples were place on ice until all embryos were processed. The nuclei were pelleted by centrifugation at 800 RCF at 4ºC for 10 minutes. Supernatant removal was observed under a dissection scope to ensure that the nuclei pellet was not dislodged. The nuclear pellet was resuspended in 7.5 µl of Tagmentation buffer (Illumina Tagment DNA TDE1 Enzyme and Buffer Kit) and placed on ice until all samples were resuspended. Then 2.5 µl of Tn5 Transposase was added and mixed by pipetting up and down. Tagmentation and amplification of ATAC-seq libraries were performed as described previously (Buenrostro et al., 2015). The tagmentation reaction was performed at 37ºC and 800 RPM for 30 minutes using an Eppendorf Thermomixer. Immediately after this step, the reactions were purified using the Qiagen Minelute kit following the manufacturer's instructions. Purified tagmented DNA was eluted in 10 µl of buffer EB. Library amplification was performed as described previously (Buenrostro et al., 2015). Typically, 12-14 total PCR were performed. Amplified libraries were purified using 1.8x Ampure SpRI beads following the manufacturer's instructions. Purified ATAC-seq libraries were eluted in 15 ul. The ATAC-seq library profile was evaluated using Agilent high sensitivity bio-analyzer. Concentrations were estimated using a Qubit fluorometer. Libraries from the same experiment were pooled and sequenced together. Sequencing was done at the University of Chicago Genomics Core (Chicago, IL, USA) using the Illumina NextSeq (PE 75 bp) and at Admera Health (South Plainfield, NJ, USA) using the Illumina HiSeq (PE 150 bp) and the Illumina NovaSeq X Plus (PE 150 bp).

ATAC-seq libraries were amplified using a set of modified Buenrostro primers that introduced Unique Dual Indexes (UDI) (78). The generalized UDI ATAC primer sequences are:

ATAC UDI Index Read 2 (i5):
5'- AATGATACGGCGACCACCGAGATCTACACnnnnnnnnTCGTCGGCAGCGTCAGATG T*G -3'

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

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

739

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

741 Embryos developed in an incubator at 25ºC until the desired nuclear cycle. Nuclear cycle was
742 confirmed by time after egg activation and the morphology was checked under water using a
743 transmission light microscope (Zeiss Stemi 2000-C). RNA isolation was performed using the
744 Zymo Quick-RNA Tissue/Insect Microprep Kit following the manufacturer's instructions except
745 for the tissue homogenization step. Instead, homogenization was done as follows. Once at the
746 desired nuclear cycle individual embryos were placed in the lid of a 1.5 ml low-retention
747 Eppendorf tube and homogenized in 10 µl of cold lysis buffer (10 mM Tris-HCl pH 7.4; 10 mM
748 NaCl; 3 mM MgCl2; 1% Igepal CA-630) (147), using a fire-sealed microcapillary tube
749 (Drummond Microcap 25). Once homogenized the lysate was added to 450 µl of Zymo RNA
750 Lysis Buffer and thoroughly mixed. RNA isolation resumed following the manufacturer's
751 instructions including DNAse treatment. Purified RNA was eluted in DNase/RNase-Free Water.

752

753 **RNA-seq library preparation**

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

762

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

764 The first steps are the same as the standard ATAC-seq protocol above. Individual embryos are
765 homogenized in cold lysis buffer, and a centrifugation step is performed to pellet the nuclei.
766 Following the supernatant (~47 ul) is removed and mixed thoroughly with 450 µl Zymo RNA
767 lysis buffer in a fresh low-bind tube. The nuclear pellet is resuspended in 7.5 µl Tag mentation
768 buffer and the ATAC-seq protocol continues as above with no deviations. RNA isolation
769 proceeded immediately, or the RNA lysate was stored at temperately at -20ºC. RNA isolation
770 was performed using the Zymo Quick-RNA Tissue/Insect Microprep Kit following the
771 manufacturer's instructions. Isolated RNA was for RT-qPCR and RNA-seq experiments.

772

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

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

780

781 **ATAC-seq data processing**

782 Demultiplexed reads were trimmed of adapters using TrimGalore! (148) (version 0.6.10,
783 cutadapt version 4.4) and mapped to the *Clogmia albipunctata* assembly (calbi-uni4000) using
784 Bowtie2 (149) (version 2.4.1) with option -X 2000. Suspected optical and PCR duplicates were
785 marked by Picard MarkDuplicates (version 2.21.4) using default parameters
786 (https://broadinstitute.github.io/picard/). Mapped, trimmed, duplicate marked reads were
787 imported into R using the GenomicAlignments (150) (version 1.38.2) and Rsamtools (151)
788 (version 2.18.0). Reads that mapped, were properly paired, non-secondary and had a map quality
789 scores ≥ 10 were imported. Only reads mapping to scaffolds 1-5 and scaffold 7 were used for
790 downstream analysis. The start and end coordinates of the reads were adjusted to account for the
791 Tn5 interaction site. Watson strand start sites had four base pairs subtracted, and Crick-strand
792 start sites had five base pairs subtracted (147). Reads with mapped length ≤ 120 bp were
793 considered to have originated from 'open' chromatin.

794 *MACS2*

795 Accessible chromatin peaks from NC11 through NC14 were determined using MACS2 (92)
796 (version 2.2.7.1) with options -f BEDPE -q 1e-25 on a merged dataset comprising all ATAC
797 reads corresponding to open chromatin (length ≤ 120 bp). Peaks were assigned to a genomic
798 feature on the basis of overlapping with one or more features of a gene (e.g. exon, intron,
799 promoter). Peaks that overlapped with no gene feature were classified as intergenic.

800 *DESeq2*

801 A count matrix for differential enrichment analyses using DESeq2 (96) (version 1.42.1) was
802 generated by counting the number of open ATAC reads (length ≤ 120 bp) that overlapped with a
803 MAC2 identified peak. Before running DESeq2, peaks with low count values were removed.
804 DESeq2 design parameters were specific to the experimental data set. For the WT stage-wise
805 comparisons (e.g. NC12 vs NC11, NC13 vs NC12, NC14 vs NC13 and LNC14 vs NC14) the
806 design parameters passed to DESeq2 were stage and sequencing platform (~instrument.atac.data
807 + stage). For the comparison between RNAi embryos and alignment control embryos (e.g., *Cal-*
808 *zld* RNAi vs alignment control and *Cal-opa* RNAi vs alignment control) the design components
809 were embryo batch and genotype (~Experiment.Date + genotype). All control groups were given
810 their own genotype to account for differences in embryo manipulation, and the genotype for the
811 RNAi embryos was determined by the strength of knockdown (see *Fold change calculations*).
812 The criteria for a statistical significance change were a $\log_2$ Fold Change ≥ 1 and an adjusted p-
813 value ≤ 0.05 or a $\log_2$ Fold Change ≤ -1 and an adjusted p-value ≤ 0.05.

814 Dynamic peaks were defined as peaks that underwent a statistically significant change in
815 accessibility in at least one of the WT stage wise comparisons. Different dynamic peak classes
816 were found using the R package DEGreport (152) (R package version 1.38.5) with the parameter
817 of the minimum group size set to 10.

818 *MEME*

819 MEME (98) (version 5.4.1) analysis was performed using the following options -order 1 -meme-
820 mod zoops -maxw 14 -meme-nmotifs 15 -meme-searchsize 100000 -centrimo-score 5.0 -
821 centrimo-ethresh 10.0. Motifs were reported from all enrichment or discovery programs within
822 the MEME suite (i.e., MEME, STREME (99), CentriMo), unless specified otherwise. MEME

823 was given a set of sequences containing the 100 bp sequence centered on the summit of the peak.
824 A set of background sequences were also given to MEME. To determine motifs enriched in
825 dynamic peaks a set of non-dynamic peak sequences were provided as background sequences.
826 To determine motifs enriched in *Cal-zld* sensitive peaks a set of sequences from peaks unaffected
827 by *Cal-zld* were provided as background sequences. To determine motifs enriched in *Cal-opa*
828 sensitive peaks a set of sequences from peaks unaffected by *Cal-opa* were provided as
829 background sequences. Identified motifs were compared to known *Drosophila melanogaster*
830 motifs from the following databases: OnTheFly_2014, Fly Factor Survey, FLYREG,
831 iDMMPMM, and DMMPMM databases (111, 153-156).

832

833 **RNA-seq data processing**

834 Demultiplexed reads were trimmed of adapters using TrimGalore! (148) (version 0.6.10,
835 cutadapt version 4.4; https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) and
836 mapped to the *Clogmia albipunctata* assembly (calbi-uni4000) using Bowtie2 (149) (version
837 2.4.1) with default parameters. Suspected optical and PCR duplicates were marked by Picard
838 MarkDuplicates (version 2.21.4) using default parameters. Mapped, trimmed, duplicate marked
839 reads were imported into R using the GenomicAlignments (150) (version 1.38.2) and Rsamtools
840 (151) (version 2.18.0). Reads that mapped, were properly paired, non-secondary and had a map
841 quality scores ≥ 10 were imported.

842 *DESeq2*

843 A gene count matrix for differential enrichment analyses was generated by counting the number
844 of RNA-seq reads that overlapped with a transcript of a gene. Before running DESeq2, genes
845 with low count values were removed. DESeq2 design parameters were specific to the
846 experimental data set. For the comparison between RNAi embryos and alignment control
847 embryos (e.g., *Cal-zld* RNAi vs alignment control and *Cal-opa* RNAi vs alignment control) the
848 design components were embryo batch and genotype (~Experiment.Date + genotype). All
849 control groups were given their own genotype to account for differences in embryo
850 manipulation, and the genotype for the RNAi embryos was determined by the strength of
851 knockdown (see *Fold change calculations*). The criteria for a statistically significant change
852 were a $\log_2$ Fold Change ≥ 1 and an adjusted p-value ≤ 0.05 or a $\log_2$ Fold Change ≤ -1 and an
853 adjusted p-value ≤ 0.05.

854 *Maternal, zygotic, maternal and zygotic classification*

855 Gene expression was compared in NC2/3 (maternal, ~45-minute old) and syncytial stage
856 embryos (NC11- Late NC14). We assume here that NC2/3 embryos have little to no active
857 transcription and instead are enriched in maternally deposited mRNAs. The distribution of gene
858 expression was plotted for both stages, and a gene was determined to be expressed if its
859 expression level was above 2^5 counts. Then genes were classified based on their expression
860 profile in NC2/3 and syncytial stage embryos. We found 227 maternal genes, 691 zygotic genes,
861 6658 maternal and zygotic genes, and 7473 genes not expressed during either stage.

862 *Fold change calculations*

863 Fold change was calculated as the sample's counts per million divided by the control's counts
864 per million for the gene of interest. First a counts per million (CPM) matrix was made by
865 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 (157). 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 (88). 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* (158) (version v2.0.4) to identify repeat elements in the *Clogmia albipunctata* genome and *RepeatMasker*'s (159) (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 (88). 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* (160) (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* (161, 162) (version 2.1.12), assigning gene names based on orthology. We assessed the final annotation quality using BUSCO (Benchmarking Universal Single-Copy Orthologs) (163, 164) on the predicted protein sequences. Finally, we analyzed synteny between the *C. albipunctata* and *D. melanogaster*

genomes using *MCScanX* (165) (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 (166) (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, Natasha Megherea for managing the *Clogmia albipunctata* culture, and Suryadi Mcqueen for the illustrations of *Clogmia albipunctata* and *Drosophila melanogaster*. Edwin "Chip" Ferguson critically reviewed versions of this manuscript and provided valuable feedback. 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

949 for the annotation of the genome are listed in S1 Table, including those samples that were not
950 generated in this study but used as evidence for the annotation of the genome.
951 Genome Browser: The genomic data produced in this study can be further interrogated using a
952 genome browser ecosystem centered on JBrowse2 (167) and delivered as a cloud image designed
953 for individual use, currently available on NSF's Jetstream2 cloud platform, with a portable
954 Docker image with documentation and BSGenome package created for R available at Clogmia -
955 github: https://github.com/kallistaconsulting/genomic_resources_clogmia. Jetstream2's user-
956 friendly interface (image: *Clogmia albipunctata* Genome Resources) (168, 169) integrates tools
957 for genomic analysis, including BLAST search (SequenceServer2.0) (170), CRISPR guide RNA
958 design (modified crisprDesigner) (171), R shiny-driven differential gene expression (DGE)
959 analysis (freecount) (172), and synteny mapping for comparative genomics (ShinySyn) (173).
960 Each tool is connected back to the browser, leveraging JBrowse2's ability to dynamically
961 visualize and contextualize the genomic data. BLAST results link directly to genomic regions,
962 enabling users to explore expression tracks, variants, and synteny relationships. CRISPR tools
963 provide sgRNA design with metrics for selection and visualization of target regions alongside
964 expression data, off-target effects, and variants. DGE analysis visualizations are linked back to
965 the genome browser for further exploration. All gene model annotations in the browser connect
966 to common external databases, such as NCBI, Uniprot, FlyBase, and Google Scholar, for further
967 exploring and interpreting the data. Future enhancements will include expanded sgRNA
968 profiling, support for Docker to enable deployment on commercial cloud platforms, and further
969 optimization of workflows, ensuring this ecosystem remains a powerful and accessible solution
970 for the use of our genomic data presented here and elsewhere (88) for research and education. To
971 access the *Clogmia albipunctata* genome browser and related tools, create an ACCESS ID and
972 use this ID to create an ACCESS account to log in to Jetstream2 to apply for an 'allocation' of
973 credits that can be used on Jetstream2. Detailed instructions on the use of Jetstream2 can be
974 found at https://jetstream-cloud.org/get-started/index.html. This on-demand model reduces
975 resource costs and security concerns compared to hosting a centralized site. Administrative tools,
976 (e.g. automated reboot scripts, system service management, and load balancing for multi-user
977 groups or workshops) enhance usability while minimizing operational overhead.
978

Tables

Table 1: *Clogmia albipunctata* reference assembly statistics and quality metrics. Completeness and continuity of the assembly are assessed by Benchmarking Universal Single-Copy Orthologs (BUSCO) and scaffold N50 and L50, respectively. Eukaryotic BUSCOs recovered: the percentage of complete universal orthologs from the eukaryotic BUSCO dataset identified in the assembly. Scaffold N50: The length (in bp) of the shortest scaffold such that all scaffolds of equal or greater length account for at least 50% of the total assembly length (174). L50: The minimum number of scaffolds whose collective length accounts for 50% of the total assembly length (175).

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

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

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

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 (73, 81) and nuclear cycle lengths are based on (38).

	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			

1005 **Figures**
1006

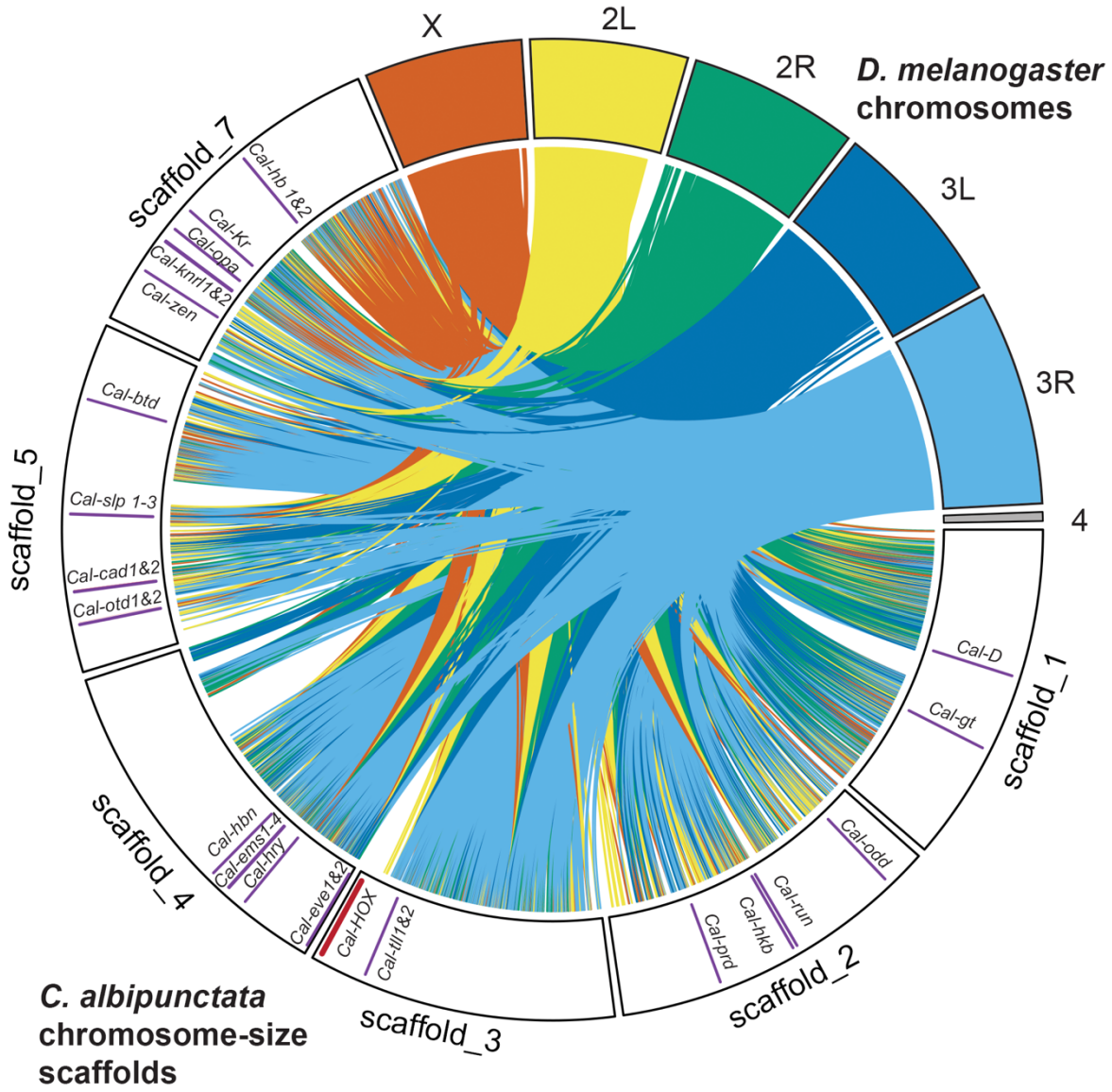

Figure 1: Synteny analysis of *Drosophila melanogaster* chromosomes and *Clogmia albipunctata* scaffolds. Synteny analysis between *Clogmia albipunctata* chromosomes-sized scaffolds *and Drosophila melanogaster* chromosomes. Groups of collinear genes are represented by contact lines connecting the positions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

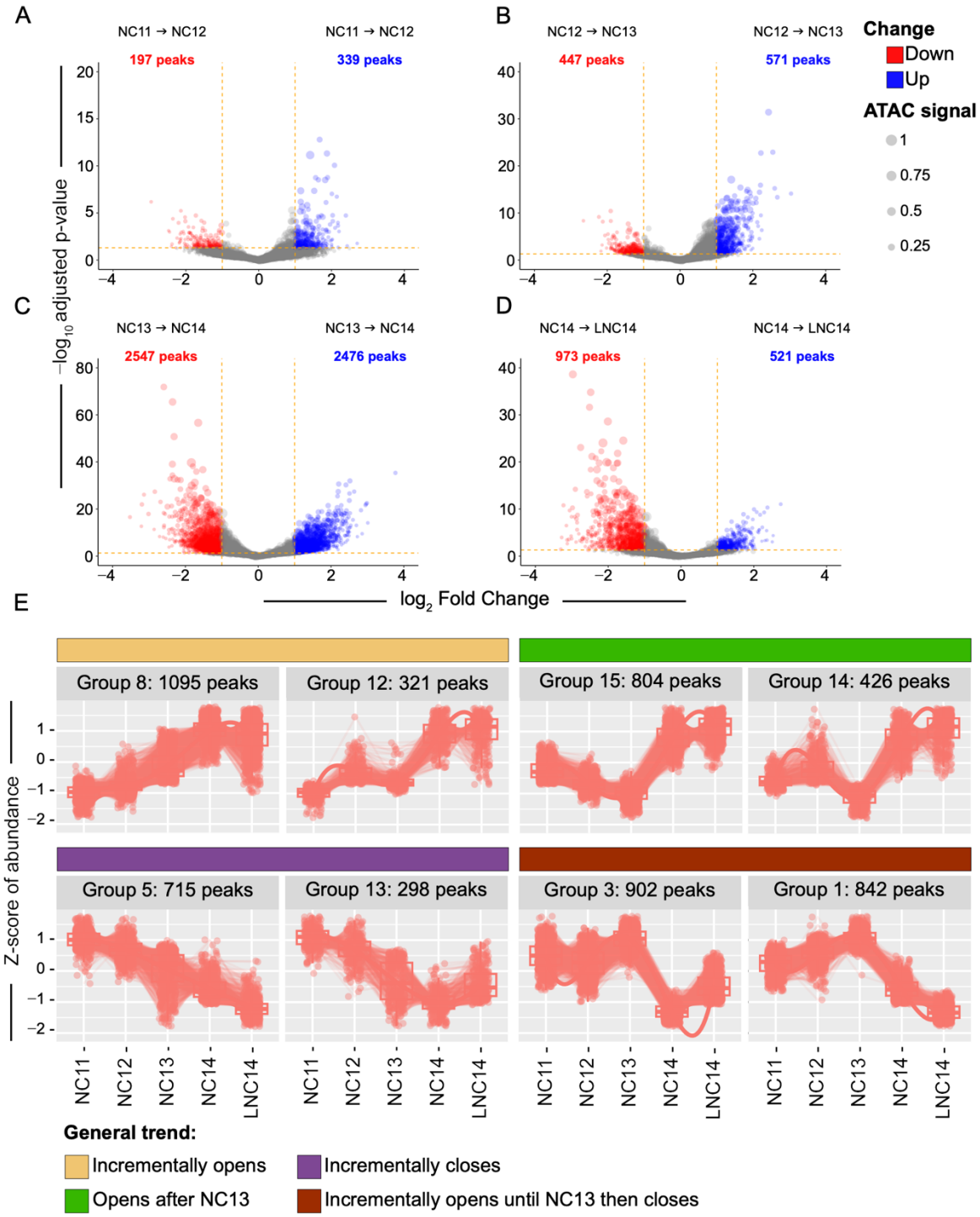

Figure 2: **Changes in chromatin accessibility during NC11 to Late NC14.** A-D) Volcano plots of differentially accessible ATAC-seq peaks between consecutive developmental stages. Significant change of adjusted p-value ≤ 0.05 and |$\log_2$ Fold Change| ≥ 1highlighted; red ($\log_2$ Fold Change ≤ -1) and blue ($\log_2$ Fold Change ≥ 1). E) Dynamic peak groups produced using DEGreport (Pantano, 2017) (R package version 1.38.5). Y-axis, Z-score of abundance. X-axis, stage.

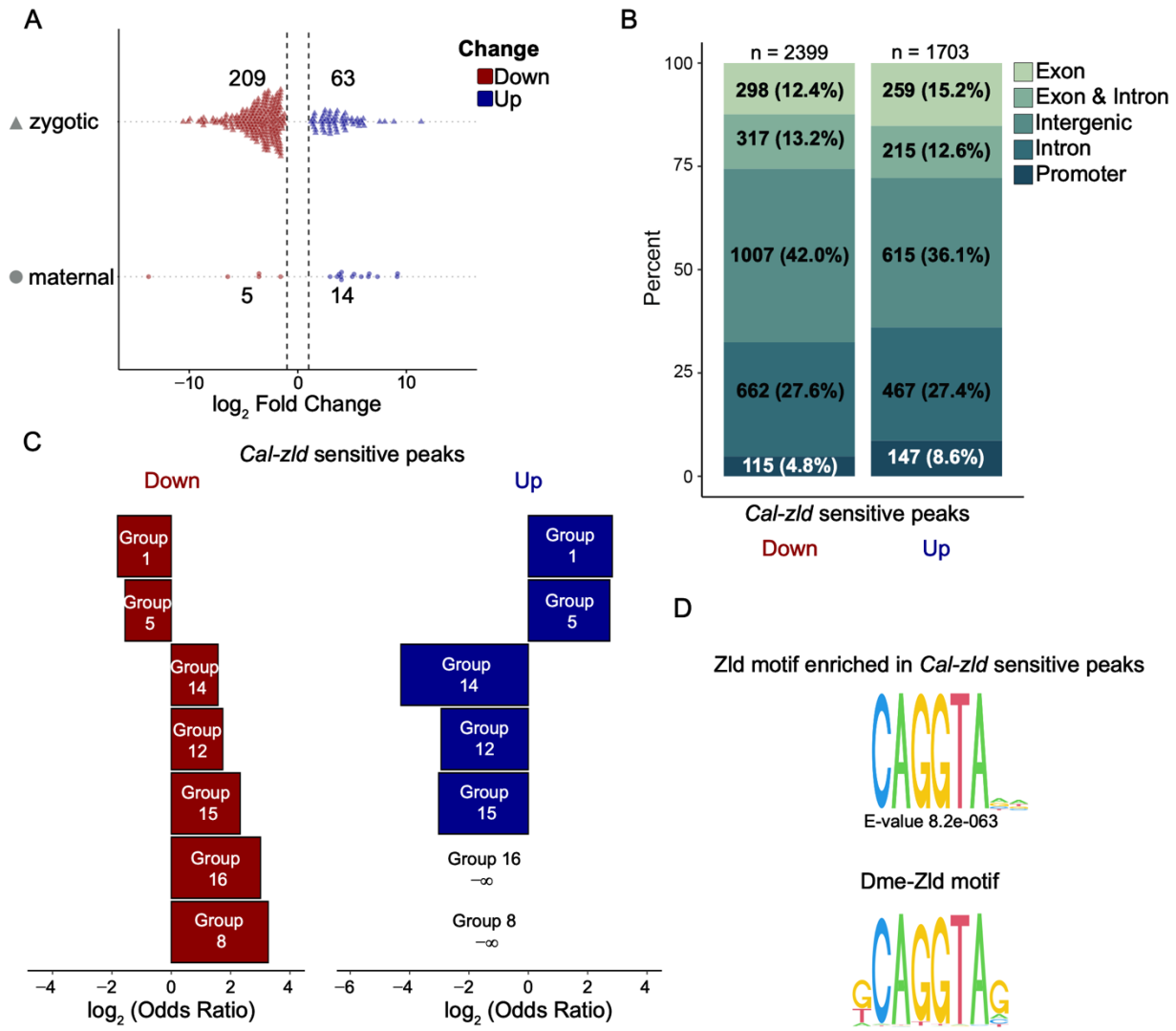

Figure 3: Cal-Zld regulates chromatin accessibility and zygotic transcription positively and negatively. A) Observed expression change of maternal or zygotic genes that are differentially expressed in *Cal-zld* RNAi embryos. On the y-axis genes are grouped by maternal or zygotic classification (see *Materials and Methods*). Log_2 Fold Change on the x-axis. Only genes with an adjusted p-value ≤ 0.05 and $|\log_2 \text{Fold Change}| \geq 1$ are shown. B) Genomic distribution of *Cal-zld* sensitive peaks that were up-regulated (blue) and down-regulated (red). C) Dynamic groups whose accessibility increases over time are more likely to be *Cal-zld* sensitive peaks that lose accessibility (Group 8 p-value, 3.4e-176; Group 16 p-value, 8.7e-05; Group 15 p-value, 1.3e-55; Group 12 p-value, 4.6e-12; Group 14 p-value, 7.2e-12) unlike dynamic groups whose accessibility decrease over time or after NC13 (Group 5 p-value, 3.5e-08; Group 1 p-value, 2.7e-12). Dynamic groups whose accessibility decreases over time are more likely to be *Cal-zld* sensitive peaks that gain accessibility (Group 5 p-value, 1.7e-70; Group 1 p-value, 1.4e-95), unlike dynamic groups whose accessibility increases over time (Group 8 p-value, 3.4e-25; Group 16 p-value, 6.3e-01; Group 15 p-value, 1.8e-11; Group 12 p-value, 9e-05; Group 14 p-value, 2.6e-07). Odds ratios and p-values were calculated by a two-sided Fisher's exact test on contingency tables constructed by the presence/absence of dynamic groups in *Cal-zld* sensitive peaks up or down. Only groups with a $\log_2(\text{Odds Ratio}) > 1.5$ in at least one comparison are

1042 shown. D) Position weight matrix (PWM) logo of the Zld motif enriched in *Cal-zld* sensitive
1043 peaks as identified by MEME (top) and the canonical Zld motif in *Drosophila melanogaster*,
1044 shown for reference (bottom).
1045

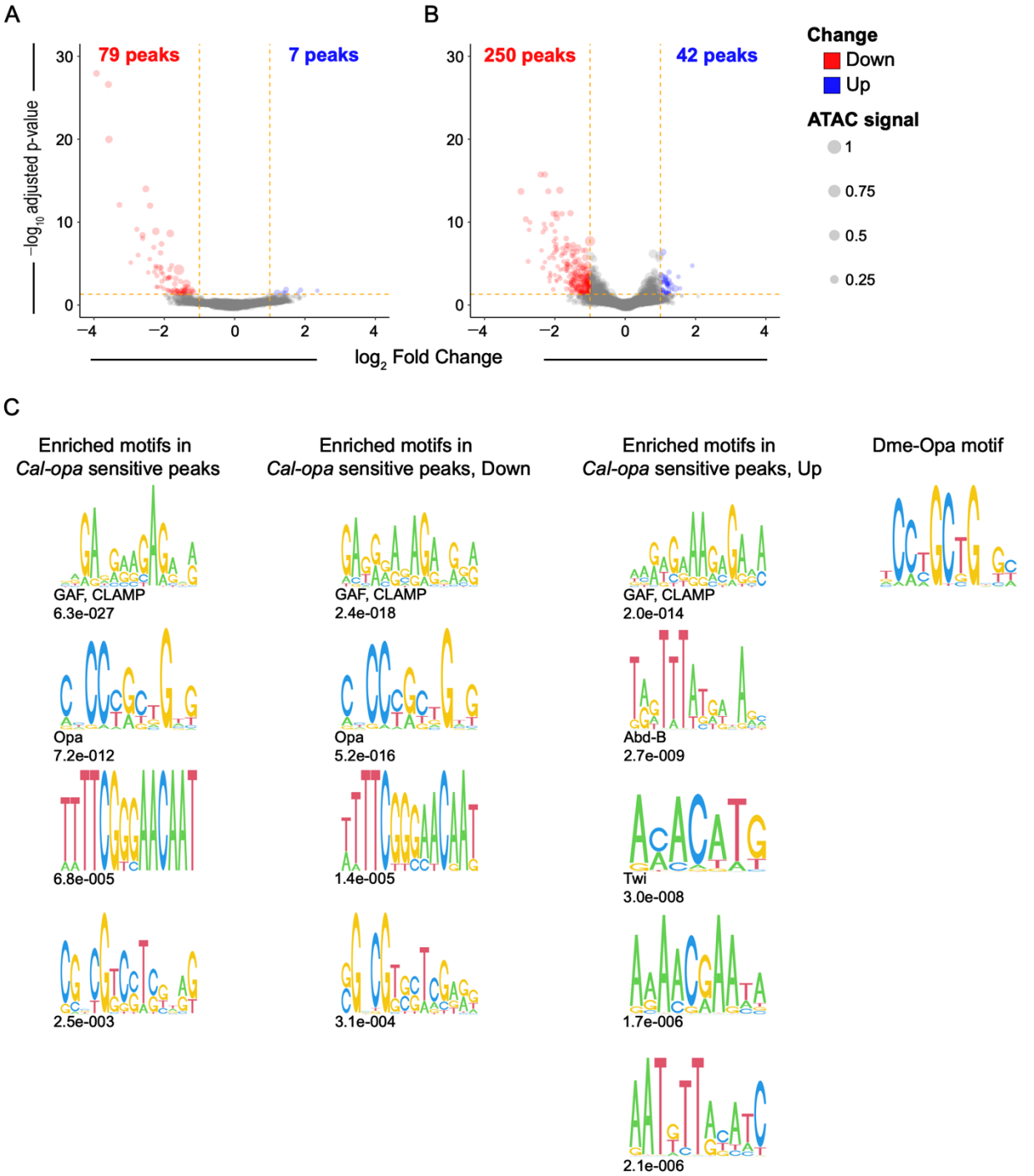

Figure 4: Cal-Opa is required for establishing a selective set of chromatin accessibility peaks. Volcano plots of differentially accessible peaks between the *Cal-opa* RNAi group and the alignment control group at A) NC12 and B) NC13. Significant change of adjusted p-value ≤ 0.05 and |$\log_2$ Fold Change| ≥ 1 highlighted; red ($\log_2$ Fold Change ≤ -1) and blue ($\log_2$ Fold Change ≥ 1). C) PWM logos of motifs enriched in all *Cal-opa* sensitive peaks (n = 326), in *Cal-opa* sensitive peaks that lose accessibility (n = 277), and in *Cal-opa* sensitive peaks that gain accessibility (n = 49), as identified by the MEME suite. Data combined for both NC12 and NC13 stages. DNA binding proteins of *Drosophila melanogaster* that bind to these or similar motifs as

identified by MEME are indicated. We include CLAMP in this list as it has been shown to bind GA-rich motifs (see text). E-value denotes the significance of recovered motifs. The canonical Opa motif in *Drosophila melanogaster* is shown for reference.

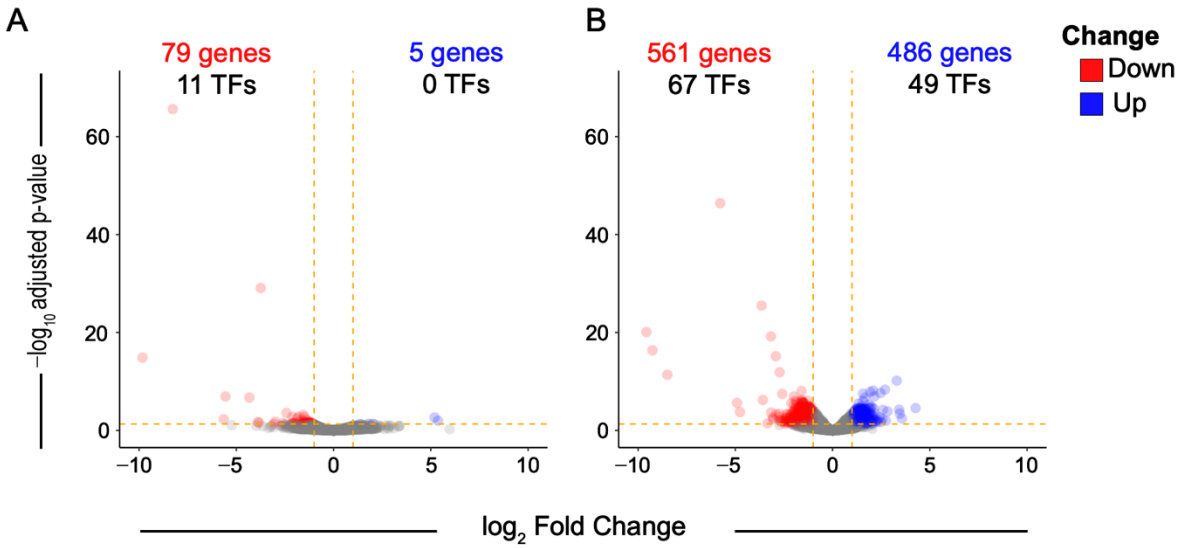

Figure 5: *Cal-opa* knockdown leads to the mis-regulation of a small set of transcription factor genes. Volcano plots of differentially expressed genes between the *Cal-opa* RNAi group and the alignment control group at A) NC12 and B) NC13. Significant change of adjusted p-value ≤ 0.05 and |$\log_2$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

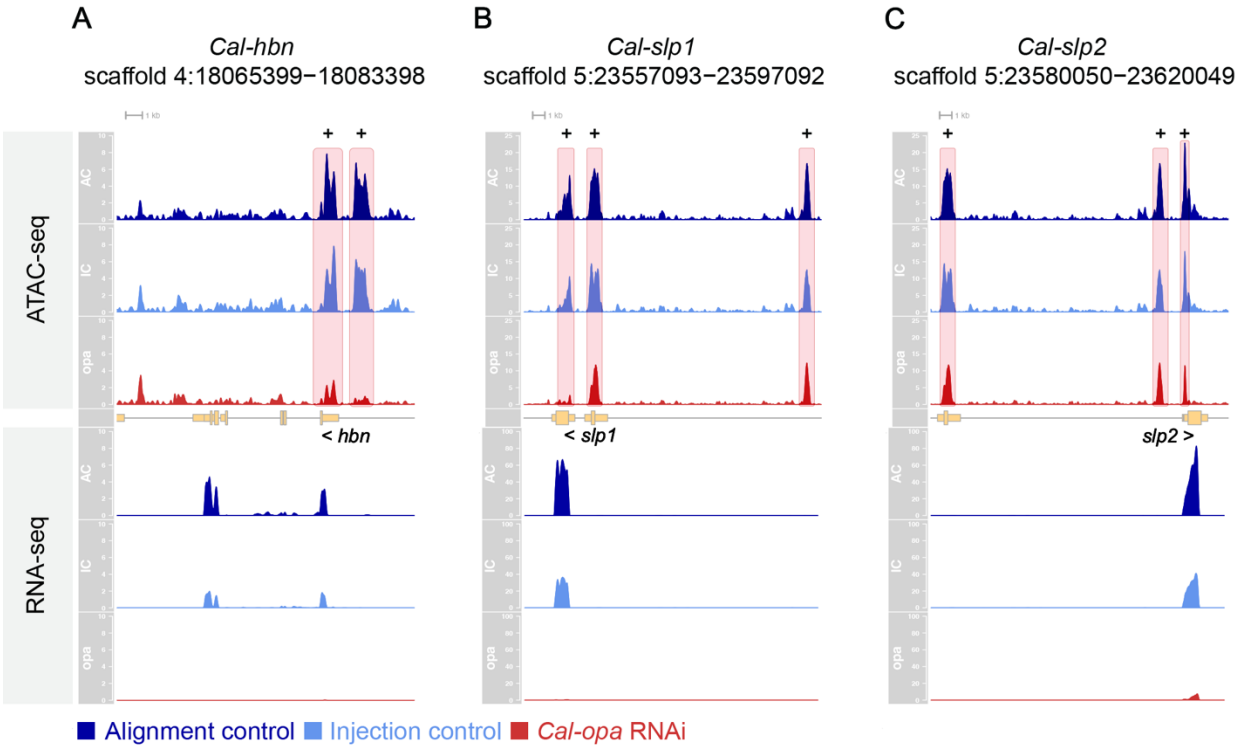

Figure 6: Key Cal-Opa targets at NC12. Gene loci of A) *Cal-hbn*, B) *Cal-slp1*, and C) *Cal-slp2* at NC12. Top tracks: CPM of ATAC-seq data (Alignment control in blue, Injection control in red, *Cal-opa* RNAi in yellow). Middle tracks show gene models in pale yellow. Bottom tracks: CPM of RNA-seq data. 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: 0-25 CPM. *Cal-slp1* RNA-seq: 0-100 CPM. *Cal-slp2* ATAC-seq: 0-25 CPM. *Cal*-slp2 RNA-seq: 0-100 CPM. Blue: Alignment control. Light blue: Injection control. Red: *Cal-opa* RNAi.

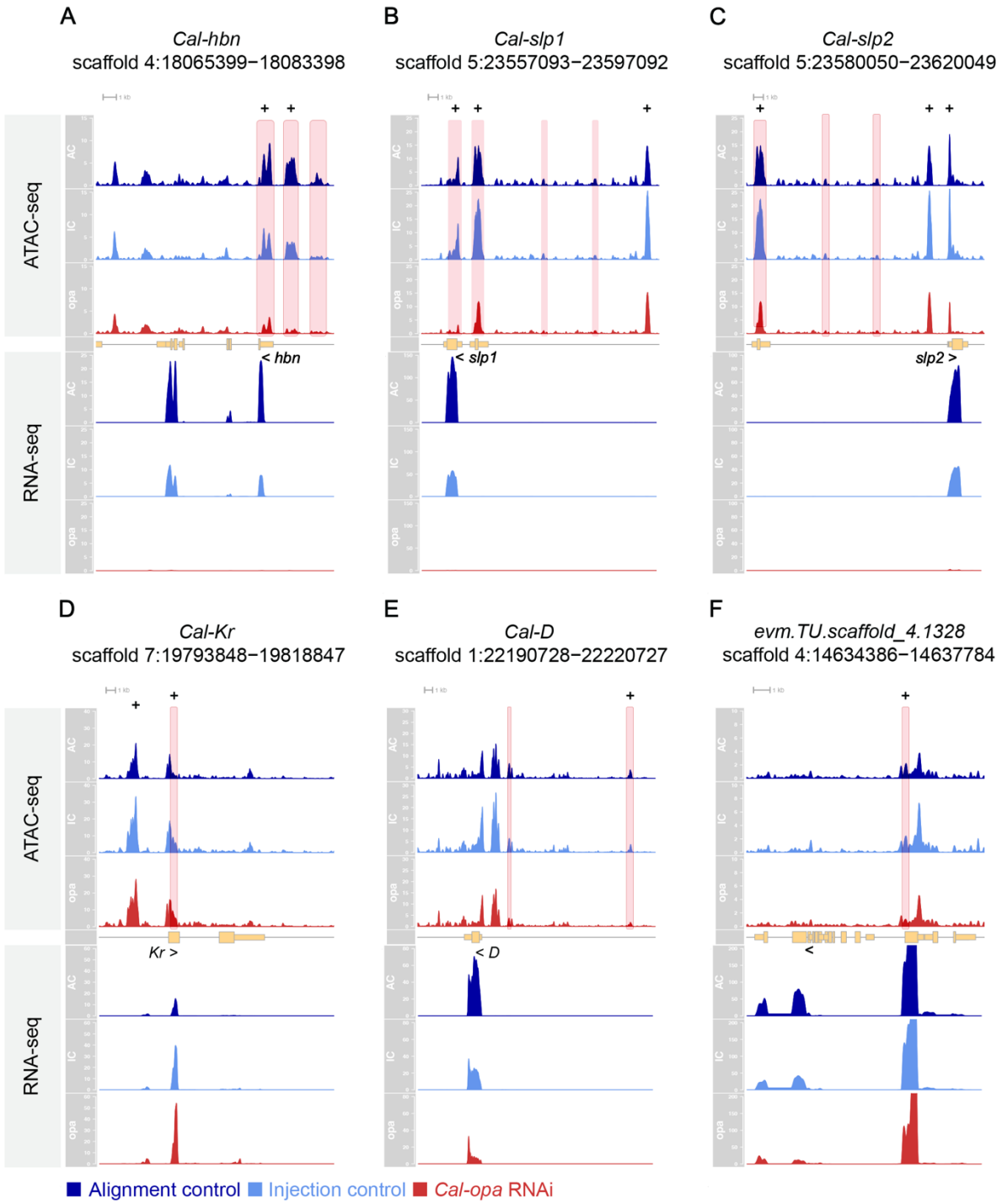

Figure 7: Key Cal-Opa targets at NC13. Gene loci of A) *Cal-hbn*, B) *Cal-slp1*, C) *Cal-slp2*, D) *Cal-Kr*, E) *Cal-D* and F) *evm.TU.scaffold_4.1328* at NC13. Top tracks: CPM of ATAC-seq data. Middle track gene model in pale yellow. Bottom tracks: CPM of RNA-seq data. Peaks with significant changes in accessibility highlighted in pink. Peaks containing Opa motifs are marked with a plus sign. The y-axis is the same for all tracks of the same data type at each locus. *Cal-hbn* ATAC-seq: 0-15 CPM. *Cal-hbn* RNA-seq: 0-25 CPM. *Cal-slp1* ATAC-seq: 0-25 CPM. *Cal-slp1* RNA-seq: 0-150 CPM. *Cal-slp2*: ATAC-seq: 0-25 CPM. *Cal-slp2* RNA-seq: 0-100 CPM. *Cal-*

1085 *Kr* ATAC-seq: 0-40 CPM. *Cal-Kr* RNA-seq: 0-60 CPM. *Cal-D* ATAC-seq: 0-30 CPM. *Cal-D*
1086 RNA-seq: 0-80 CPM. *evm.TU.scaffold_4.1328* ATAC-seq: 0-10 CPM. *evm.TU.scaffold_4.1328*
1087 RNA-seq: 0-200 CPM. Blue: Alignment control. Light blue: Injection control. Red: *Cal-opa*
1088 RNAi.
1089

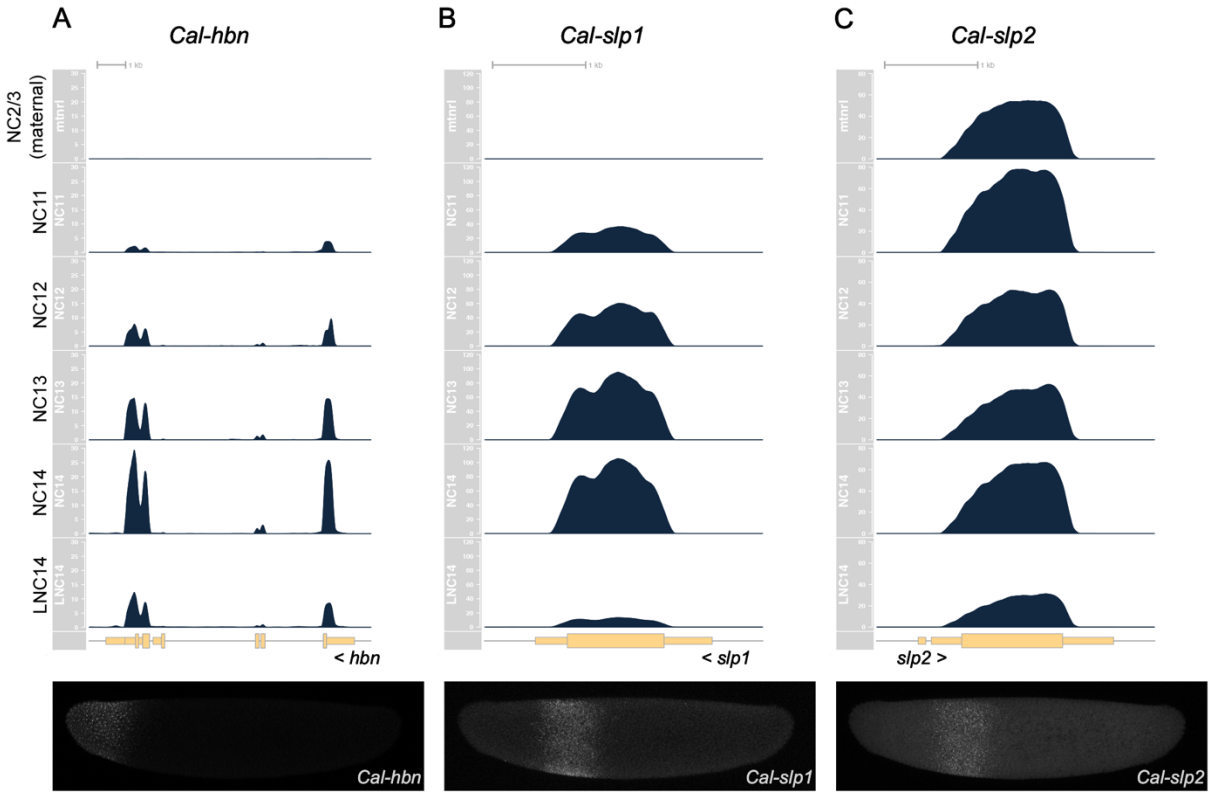

Figure 8: Time-course RNA-seq data for *Cal-hbn*, *Cal-slp1*, and *Cal-slp2*. Expression levels (CPM) of A) *Cal-hbn*, B) *Cal-slp1*, C) *Cal-slp2* at 6 consecutive stages, including NC2/3 (maternal, ~45-minute old embryos), NC11, NC12, NC13, NC14 and Late NC14 are shown above fluorescent *in situ* hybridization 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.

1099 **References**
1100

1101 1. Anderson DT. The comparative embryology of the Diptera. Annu Rev Entomol.
1102 1966;11:23-64.

1103 2. Campos-Ortega JA, Hartenstein V. The embryonic development of *Drosophila*
1104 *melanogaster*. 2nd ed. Berlin, Heidelberg, New York: Springer-Verlag; 1997. 405 p.

1105 3. Yoon Y, Klomp J, Martin-Martin I, Criscione F, Calvo E, Ribeiro J, et al. Embryo
1106 polarity in moth flies and mosquitoes relies on distinct old genes with localized transcript
1107 isoforms. eLife. 2019;8.

1108 4. Klomp J, Athy D, Kwan CW, Bloch NI, Sandmann T, Lemke S, et al. Embryo
1109 development. A cysteine-clamp gene drives embryo polarity in the midge Chironomus. Science.
1110 2015;348(6238):1040-2.

1111 5. Rafiqi AM, Lemke S, Ferguson S, Stauber M, Schmidt-Ott U. Evolutionary origin of the
1112 amnioserosa in cyclorrhaphan flies correlates with spatial and temporal expression changes of
1113 *zen*. Proceedings of the National Academy of Sciences of the United States of America.
1114 2008;105(1):234-9.

1115 6. Kwan CW, Gavin-Smyth J, Ferguson EL, Schmidt-Ott U. Functional evolution of a
1116 morphogenetic gradient. eLife. 2016;5.

1117 7. Caroti F, Gonzalez Avalos E, Noeske V, Gonzalez Avalos P, Kromm D, Wosch M, et al.
1118 Decoupling from yolk sac is required for extraembryonic tissue spreading in the scuttle fly
1119 Megaselia abdita. eLife. 2018;7.

1120 8. Urbansky S, Gonzalez Avalos P, Wosch M, Lemke S. Folded gastrulation and T48 drive
1121 the evolution of coordinated mesoderm internalization in flies. eLife. 2016;5.

1122 9. Fraire-Zamora JJ, Jaeger J, Solon J. Two consecutive microtubule-based epithelial
1123 seaming events mediate dorsal closure in the scuttle fly Megaselia abdita. eLife. 2018;7.

1124 10. Dey B, Kaul V, Kale G, Scorcelletti M, Takeda M, Wang YC, et al. Divergent
1125 evolutionary strategies preempt tissue collision in fly gastrulation. bioRxiv2023.

1126 11. Hall AB, Basu S, Jiang X, Qi Y, Timoshevskiy VA, Biedler JK, et al. A male-
1127 determining factor in the mosquito Aedes aegypti. Science. 2015;348(6240):1268-70.

1128 12. Krzywinska E, Dennison NJ, Lycett GJ, Krzywinski J. A maleness gene in the malaria
1129 mosquito Anopheles gambiae. Science. 2016;353(6294):67-9.

1130 13. Saccone G. A history of the genetic and molecular identification of genes and their
1131 functions controlling insect sex determination. Insect biochemistry and molecular biology.
1132 2022;151:103873.

1133 14. Vicoso B, Bachtrog D. Numerous transitions of sex chromosomes in Diptera. PLoS
1134 biology. 2015;13(4):e1002078.

1135 15. St Johnston D, Nüsslein-Volhard C. The origin of pattern and polarity in the *Drosophila*
1136 embryo. Cell. 1992;68:201-20.

1137 16. Surkova S, Golubkova E, Mamon L, Samsonova M. Dynamic maternal gradients and
1138 morphogenetic networks in Drosophila early embryo. Bio Systems. 2018;173:207-13.

1139 17. Ma J, He F, Xie G, Deng GX. Maternal AP determination in the Drosophila oocyte and
1140 embryo. Wiley interdisciplinary reviews Developmental biology. 2016;5:562-81.

1141 18. Wieschaus E. Positional Information and Cell Fate Determination in the Early Drosophila
1142 Embryo. Curr Top Dev Biol. 2016;117:567-79.

1143 19. Jaeger J. The gap gene network. Cellular and molecular life sciences : CMLS.
1144 2011;68(2):243-74.

1145 20. Tkačik G, Gregor T. The many bits of positional information. Development. 2021;148(2).

1146 21. Clark E, Peel AD, Akam M. Arthropod segmentation. Development. 2019;146(18).

1147 22. Crombach A, Jaeger J. Life's attractors continued: progress in understanding
1148 developmental systems through reverse engineering and in silico evolution. In: Crombach A,
1149 editor. Evolutionary Systems Biology: Springer Nature Switzerland AG; 2021.

1150 23. Irizarry J, Stathopoulos A. Dynamic patterning by morphogens illuminated by cis-
1151 regulatory studies. Development. 2021;148(2).

1152 24. Driever W, Nüsslein-Volhard C. A gradient of bicoid protein in Drosophila embryos.
1153 Cell. 1988;54:83-93.

1154 25. Driever W, Nüsslein-Volhard C. The Bicoid protein determines position in the
1155 *Drosophila* embryo in a concentration-dependent manner. Cell. 1988;54:95-104.

1156 26. Berleth T, Burri M, Thoma G, Bopp D, Richstein S, Frigerio G, et al. The role of
1157 localization of *bicoid* RNA in organizing the anterior pattern of the *Drosophila* embryo. EMBO
1158 J. 1988;7:1749-56.

1159 27. Stauber M, Taubert H, Schmidt-Ott U. Function of *bicoid* and *hunchback* homologs in
1160 the basal cyclorrhaphan fly *Megaselia* (Phoridae). Proc Natl Acad Sci USA. 2000;97:10844-9.

1161 28. Shaw PJ, Salameh A, McGregor AP, Bala S, Dover GA. Divergent structure and function
1162 of the *bicoid* gene in Muscoidea fly species. Evol Dev. 2001;in press.

1163 29. Lemke S, Busch SE, Antonopoulos DA, Meyer F, Domanus MH, Schmidt-Ott U.
1164 Maternal activation of gap genes in the hover fly *Episyrphus*. Development. 2010;137:1709-19.

1165 30. Wotton KR, Jimenez-Guri E, Jaeger J. Maternal co-ordinate gene regulation and axis
1166 polarity in the scuttle fly Megaselia abdita. PLoS genetics. 2015;11(3):e1005042.

1167 31. Hursh DA, Stultz BG. Odd-Paired: The Drosophila Zic Gene. Advances in experimental
1168 medicine and biology. 2018;1046:41-58.

1169 32. Ravindranath AJ, Cadigan KM. Structure-function analysis of the C-clamp of
1170 TCF/Pangolin in Wnt/ss-catenin signaling. PloS one. 2014;9(1):e86180.

1171 33. Liu Q, Onal P, Datta RR, Rogers JM, Schmidt-Ott U, Bulyk ML, et al. Ancient
1172 mechanisms for the evolution of the bicoid homeodomain's function in fly development. eLife.
1173 2018;7.

1174 34. Nitta KR, Jolma A, Yin Y, Morgunova E, Kivioja T, Akhtar J, et al. Conservation of
1175 transcription factor binding specificities across 600 million years of bilateria evolution. eLife.
1176 2015;4.

1177 35. Duan B, Fu D, Zhang C, Ding P, Dong X, Xia B. Selective Nonmethylated CpG DNA
1178 Recognition Mechanism of Cysteine Clamp Domains. J Am Chem Soc. 2021;143(20):7688-97.

1179 36. Li M, Kasan K, Saha Z, Yoon Y, Schmidt-Ott U. Twenty-seven ZAD-ZNF genes of
1180 Drosophila melanogaster are orthologous to the embryo polarity determining mosquito gene
1181 cucoid. PloS one. 2023;18(1):e0274716.

1182 37. Foe VE, Odell GM, Edgar BA. Mitosis and morphogenesis in the Drosophila embryo:
1183 point and counterpoint. In: Bate M, Martinez-Arias A, editors. The development of *Drosophila*
1184 *melanogaster*. 1. Cold Spring Harbor: Cold Spring Harbor Laboratory Press; 1993. p. 149-300.

1185 38. Jimenez-Guri E, Wotton KR, Gavilan B, Jaeger J. A staging scheme for the development
1186 of the moth midge Clogmia albipunctata. PloS one. 2014;9(1):e84422.

1187 39. Nüsslein-Volhard C, Wieschaus E. Mutations affecting segment number and polarity in
1188 *Drosophila*. Nature. 1980;287:795-801.

1189 40. Clark E. Dynamic patterning by the Drosophila pair-rule network reconciles long-germ
1190 and short-germ segmentation. PLoS biology. 2017;15(9):e2002439.

1191 41. De Renzis S, Elemento O, Tavazoie S, Wieschaus EF. Unmasking activation of the
1192 zygotic genome using chromosomal deletions in the Drosophila embryo. PLoS biology.
1193 2007;5(5):e117.

1194 42. Lott SE, Villalta JE, Schroth GP, Luo S, Tonkin LA, Eisen MB. Noncanonical
1195 compensation of zygotic X transcription in early *Drosophila melanogaster* development revealed
1196 through single-embryo RNA-seq. PLoS biology. 2011;9(2):e1000590.

1197 43. Kwasnieski JC, Orr-Weaver TL, Bartel DP. Early genome activation in Drosophila is
1198 extensive with an initial tendency for aborted transcripts and retained introns. Genome research.
1199 2019;29(7):1188-97.

1200 44. Tadros W, Lipshitz HD. The maternal-to-zygotic transition: a play in two acts.
1201 Development. 2009;136(18):3033-42.

1202 45. ten Bosch JR, Benavides JA, Cline TW. The TAGteam DNA motif controls the timing of
1203 Drosophila pre-blastoderm transcription. Development. 2006;133(10):1967-77.

1204 46. Liang HL, Nien CY, Liu HY, Metzstein MM, Kirov N, Rushlow C. The zinc-finger
1205 protein Zelda is a key activator of the early zygotic genome in Drosophila. Nature.
1206 2008;456(7220):400-3.

1207 47. Chen K, Johnston J, Shao W, Meier S, Staber C, Zeitlinger J. A global change in RNA
1208 polymerase II pausing during the Drosophila midblastula transition. eLife. 2013;2:e00861.

1209 48. Blythe SA, Wieschaus EF. Zygotic genome activation triggers the DNA replication
1210 checkpoint at the midblastula transition. Cell. 2015;160(6):1169-81.

1211 49. Larson ED, Marsh AJ, Harrison MM. Pioneering the developmental frontier. Molecular
1212 cell. 2021;81(8):1640-50.

1213 50. Iwafuchi-Doi M, Zaret KS. Pioneer transcription factors in cell reprogramming. Genes &
1214 development. 2014;28(24):2679-92.

1215 51. Zaret KS. Pioneer Transcription Factors Initiating Gene Network Changes. Annual
1216 review of genetics. 2020;54:367-85.

1217 52. Staudt N, Fellert S, Chung HR, Jackle H, Vorbruggen G. Mutations of the Drosophila
1218 zinc finger-encoding gene vielfaltig impair mitotic cell divisions and cause improper
1219 chromosome segregation. Molecular biology of the cell. 2006;17(5):2356-65.

1220 53. Nien CY, Liang HL, Butcher S, Sun Y, Fu S, Gocha T, et al. Temporal coordination of
1221 gene networks by Zelda in the early Drosophila embryo. PLoS genetics. 2011;7(10):e1002339.

1222 54. Harrison MM, Li XY, Kaplan T, Botchan MR, Eisen MB. Zelda binding in the early
1223 Drosophila melanogaster embryo marks regions subsequently activated at the maternal-to-
1224 zygotic transition. PLoS genetics. 2011;7(10):e1002266.

1225 55. Satija R, Bradley RK. The TAGteam motif facilitates binding of 21 sequence-specific
1226 transcription factors in the Drosophila embryo. Genome research. 2012;22(4):656-65.

1227 56. Brennan KJ, Weilert M, Krueger S, Pampari A, Liu HY, Yang AWH, et al. Chromatin
1228 accessibility in the Drosophila embryo is determined by transcription factor pioneering and
1229 enhancer activation. Developmental cell. 2023;58(19):1898-916 e9.

1230 57. Foo SM, Sun Y, Lim B, Ziukaite R, O'Brien K, Nien CY, et al. Zelda potentiates
1231 morphogen activity by increasing chromatin accessibility. Current biology : CB.
1232 2014;24(12):1341-6.

1233 58. McDaniel SL, Gibson TJ, Schulz KN, Fernandez Garcia M, Nevil M, Jain SU, et al.
1234 Continued Activity of the Pioneer Factor Zelda Is Required to Drive Zygotic Genome
1235 Activation. Molecular cell. 2019;74(1):185-95 e4.

1236 59. Mir M, Stadler MR, Ortiz SA, Hannon CE, Harrison MM, Darzacq X, et al. Dynamic
1237 multifactor hubs interact transiently with sites of active transcription in Drosophila embryos.
1238 eLife. 2018;7.

1239 60. Mir M, Reimer A, Haines JE, Li XY, Stadler M, Garcia H, et al. Dense Bicoid hubs
1240 accentuate binding along the morphogen gradient. Genes & development. 2017;31(17):1784-94.

1241 61. Duan J, Rieder L, Colonnetta MM, Huang A, McKenney M, Watters S, et al. CLAMP
1242 and Zelda function together to promote Drosophila zygotic genome activation. eLife. 2021;10.

1243 62. Colonnetta MM, Abrahante JE, Schedl P, Gohl DM, Deshpande G. CLAMP regulates
1244 zygotic genome activation in Drosophila embryos. Genetics. 2021;219(2).

1245 63. Gaskill MM, Gibson TJ, Larson ED, Harrison MM. GAF is essential for zygotic genome
1246 activation and chromatin accessibility in the early Drosophila embryo. eLife. 2021;10.

1247 64. Moshe A, Kaplan T. Genome-wide search for Zelda-like chromatin signatures identifies
1248 GAF as a pioneer factor in early fly development. Epigenetics Chromatin. 2017;10(1):33.

1249 65. Huang A, Amourda C, Zhang S, Tolwinski NS, Saunders TE. Decoding temporal
1250 interpretation of the morphogen Bicoid in the early Drosophila embryo. eLife. 2017;6.

1251 66. Xu Z, Chen H, Ling J, Yu D, Struffi P, Small S. Impacts of the ubiquitous factor Zelda on
1252 Bicoid-dependent DNA binding and transcription in Drosophila. Genes & development.
1253 2014;28(6):608-21.

1254 67. Bishop TR, Onal P, Xu Z, Zheng M, Gunasinghe H, Nien CY, et al. Multi-level
1255 regulation of even-skipped stripes by the ubiquitous factor Zelda. Development. 2023;150(23).

1256 68. Crocker J, Stern DL. Functional regulatory evolution outside of the minimal even-
1257 skipped stripe 2 enhancer. Development. 2017;144(17):3095-101.

1258 69. Ling J, Umezawa KY, Scott T, Small S. Bicoid-Dependent Activation of the Target Gene
1259 hunchback Requires a Two-Motif Sequence Code in a Specific Basal Promoter. Molecular cell.
1260 2019;75(6):1178-87 e4.

1261 70. Eck E, Liu J, Kazemzadeh-Atoufi M, Ghoreishi S, Blythe SA, Garcia HG. Quantitative
1262 dissection of transcription in development yields evidence for transcription-factor-driven
1263 chromatin accessibility. eLife. 2020;9.

1264 71. Fernandes G, Tran H, Andrieu M, Diaw Y, Perez Romero C, Fradin C, et al. Synthetic
1265 reconstruction of the hunchback promoter specifies the role of Bicoid, Zelda and Hunchback in
1266 the dynamics of its transcription. eLife. 2022;11.

1267 72. Harden TT, Vincent BJ, DePace AH. Transcriptional activators in the early Drosophila
1268 embryo perform different kinetic roles. Cell systems. 2023;14(4):258-72 e4.

1269 73. Hannon CE, Blythe SA, Wieschaus EF. Concentration dependent chromatin states
1270 induced by the bicoid morphogen gradient. eLife. 2017;6.

1271 74. Datta RR, Ling J, Kurland J, Ren X, Xu Z, Yucel G, et al. A feed-forward relay integrates
1272 the regulatory activities of Bicoid and Orthodenticle via sequential binding to suboptimal sites.
1273 Genes & development. 2018;32(9-10):723-36.

1274 75. Haines JE, Eisen MB. Patterns of chromatin accessibility along the anterior-posterior axis
1275 in the early Drosophila embryo. PLoS genetics. 2018;14(5):e1007367.

1276 76. Cusanovich DA, Reddington JP, Garfield DA, Daza RM, Aghamirzaie D, Marco-
1277 Ferreres R, et al. The cis-regulatory dynamics of embryonic development at single-cell
1278 resolution. Nature. 2018;555(7697):538-42.

1279 77. Bozek M, Cortini R, Storti AE, Unnerstall U, Gaul U, Gompel N. ATAC-seq reveals
1280 regional differences in enhancer accessibility during the establishment of spatial coordinates in
1281 the Drosophila blastoderm. Genome research. 2019;29(5):771-83.

1282 78. Soluri IV, Zumerling LM, Payan Parra OA, Clark EG, Blythe SA. Zygotic pioneer factor
1283 activity of Odd-paired/Zic is necessary for late function of the Drosophila segmentation network.
1284 eLife. 2020;9.

1285 79. Calderon D, Blecher-Gonen R, Huang X, Secchia S, Kentro J, Daza RM, et al. The
1286 continuum of Drosophila embryonic development at single-cell resolution. Science.
1287 2022;377(6606):eabn5800.

1288 80. Degen EA, Croslyn C, Mangan NM, Blythe SA. Bicoid-nucleosome competition sets a
1289 concentration threhold for transcription constrained by genome replication. bioRxiv. 2024.

1290 81. Blythe SA, Wieschaus EF. Establishment and maintenance of heritable chromatin
1291 structure during early Drosophila embryogenesis. eLife. 2016;5.

1292 82. Clark E, Akam M. Odd-paired controls frequency doubling in Drosophila segmentation
1293 by altering the pair-rule gene regulatory network. eLife. 2016;5.

1294 83. Koromila T, Gao F, Iwasaki Y, He P, Pachter L, Gergen JP, et al. Odd-paired is a
1295 pioneer-like factor that coordinates with Zelda to control gene expression in embryos. eLife.
1296 2020;9.

1297 84. Amabis JM, Simoes LCG. Chromosome studies in a species of *Telmatoscopus* sp.
1298 Caryologia. 1972;25(2).

1299 85. Amabis JM. Cytological evidence of male sex heterogamety in sex determination of
1300 *Telmatoscopus albipunctatus* (Diptera: Psychodidae). Chromosoma. 1977;62:133-8.

1301 86. Schmidt-Ott U, Rafiqi AM, Sander K, Johnston JS. Extremely small genomes in two
1302 unrelated dipteran insects with shared early developmental traits. Development genes and
1303 evolution. 2009;219(4):207-10.

1304 87. Generalovic TN, McCarthy SA, Warren IA, Wood JMD, Torrance J, Sims Y, et al. A
1305 high-quality, chromosome-level genome assembly of the Black Soldier Fly (Hermetia illucens
1306 L.). G3 (Bethesda). 2021;11(5).

1307 88. Tenger-Trolander A, Amiri E, Gantz V, Kwan CW, Sanders SA, Schmidt-Ott U.
1308 Genomic resources for the scuttle fly Megaselia abdita: A Model Organism for Comparative
1309 Developmental Studies in Flies bioRxiv. 2025.

1310 89. Ji J, Gao Y, Xu C, Zhang K, Li D, Li B, et al. Chromosome-level genome assembly of
1311 marmalade hoverfly Episyrphus balteatus (Diptera: Syrphidae). Scientific data. 2024;11(1):844.

1312 90. Stauber M, Jäckle H, Schmidt-Ott U. The anterior determinant *bicoid* of *Drosophila* is a
1313 derived Hox class 3 gene. Proc Natl Acad Sci USA. 1999;96:3786-9.

1314 91. Stauber M, Prell A, Schmidt-Ott U. A single *Hox3* gene with composite *bicoid* and
1315 *zerknüllt* expression characteristics in non-Cyclorrhaphan flies. Proceedings of the National
1316 Academy of Sciences of the United States of America. 2002;99:274-9.

1317 92. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based
1318 analysis of ChIP-Seq (MACS). Genome biology. 2008;9(9):R137.

1319 93. Reddington JP, Garfield DA, Sigalova OM, Karabacak Calviello A, Marco-Ferreres R,
1320 Girardot C, et al. Lineage-Resolved Enhancer and Promoter Usage during a Time Course of
1321 Embryogenesis. Developmental cell. 2020;55(5):648-64 e9.

1322 94. Schulz KN, Harrison MM. Mechanisms regulating zygotic genome activation. Nature
1323 reviews Genetics. 2019;20(4):221-34.

1324 95. Hamm DC, Harrison MM. Regulatory principles governing the maternal-to-zygotic
1325 transition: insights from Drosophila melanogaster. Open Biol. 2018;8(12):180183.

1326 96. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for
1327 RNA-seq data with DESeq2. Genome biology. 2014;15(12):550.

1328 97. Sun Y, Nien CY, Chen K, Liu HY, Johnston J, Zeitlinger J, et al. Zelda overcomes the
1329 high intrinsic nucleosome barrier at enhancers during Drosophila zygotic genome activation.
1330 Genome research. 2015;25(11):1703-14.

1331 98. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME Suite. Nucleic acids research.
1332 2015;43(W1):W39-49.

1333 99. Bailey TL. STREME: accurate and versatile sequence motif discovery. Bioinformatics.
1334 2021;37(18):2834-40.

1335 100. Soruco MM, Chery J, Bishop EP, Siggers T, Tolstorukov MY, Leydon AR, et al. The
1336 CLAMP protein links the MSL complex to the X chromosome during Drosophila dosage
1337 compensation. Genes & development. 2013;27(14):1551-6.

1338 101. Kuzu G, Kaye EG, Chery J, Siggers T, Yang L, Dobson JR, et al. Expansion of GA
1339 Dinucleotide Repeats Increases the Density of CLAMP Binding Sites on the X-Chromosome to
1340 Promote Drosophila Dosage Compensation. PLoS genetics. 2016;12(7):e1006120.

1341 102. Kaye EG, Booker M, Kurland JV, Conicella AE, Fawzi NL, Bulyk ML, et al. Differential
1342 Occupancy of Two GA-Binding Proteins Promotes Targeting of the Drosophila Dosage
1343 Compensation Complex to the Male X Chromosome. Cell reports. 2018;22(12):3227-39.

1344 103. Ray P, De S, Mitra A, Bezstarosti K, Demmers JA, Pfeifer K, et al. Combgap contributes
1345 to recruitment of Polycomb group proteins in Drosophila. Proceedings of the National Academy
1346 of Sciences of the United States of America. 2016;113(14):3826-31.

1347 104. Ribeiro L, Tobias-Santos V, Santos D, Antunes F, Feltran G, de Souza Menezes J, et al.
1348 Evolution and multiple roles of the Pancrustacea specific transcription factor zelda in insects.
1349 PLoS genetics. 2017;13(7):e1006868.

1350 105. Biedler JK, Hu W, Tae H, Tu Z. Identification of early zygotic genes in the yellow fever
1351 mosquito Aedes aegypti and discovery of a motif involved in early zygotic genome activation.
1352 PloS one. 2012;7(3):e33933.

1353 106. Arsala D, Lynch JA. Ploidy has little effect on timing early embryonic events in the
1354 haplo-diploid wasp Nasonia. Genesis. 2017;55(5).

1355 107. Pires CV, Freitas FC, Cristino AS, Dearden PK, Simoes ZL. Transcriptome Analysis of
1356 Honeybee (Apis Mellifera) Haploid and Diploid Embryos Reveals Early Zygotic Transcription
1357 during Cleavage. PloS one. 2016;11(1):e0146447.

1358 108. Gao S, Xue S, Gao T, Lu R, Zhang X, Zhang Y, et al. Transcriptome analysis reveals the
1359 role of Zelda in the regulation of embryonic and wing development of Tribolium castaneum.
1360 Bull Entomol Res. 2023;113(5):587-97.

1361 109. Ventos-Alfonso A, Ylla G, Belles X. Zelda and the maternal-to-zygotic transition in
1362 cockroaches. The FEBS journal. 2019;286(16):3206-21.

1363 110. Gibson TJ, Larson ED, Harrison MM. Protein-intrinsic properties and context-dependent
1364 effects regulate pioneer factor binding and function. Nature structural & molecular biology.
1365 2024;31(3):548-58.

1366 111. Zhu LJ, Christensen RG, Kazemian M, Hull CJ, Enuameh MS, Basciotta MD, et al.
1367 FlyFactorSurvey: a database of Drosophila transcription factor binding specificities determined
1368 using the bacterial one-hybrid system. Nucleic acids research. 2011;39(Database issue):D111-7.

1369 112. Garcia-Solache M, Jaeger J, Akam M. A systematic analysis of the gap gene system in
1370 the moth midge *Clogmia albipunctata*. Developmental biology. 2010;344:306-18.

1371 113. Goltsev Y, Hsiong W, Lanzaro G, Levine M. Different combinations of gap repressors
1372 for common stripes in *Anopheles* and *Drosophila* embryos. Dev Biol. 2004;275:435-46.

1373 114. Wotton KR, Jimenez-Guri E, Crombach A, Janssens H, Alcaine-Colet A, Lemke S, et al.
1374 Quantitative system drift compensates for altered maternal inputs to the gap gene network of the
1375 scuttle fly Megaselia abdita. eLife. 2015;4.

1376 115. Bullock SL, Stauber M, Prell A, Hughes JR, Ish-Horowicz D, Schmidt-Ott U.
1377 Differential cytoplasmic mRNA localization adjusts pair-rule transcription factor activity to
1378 cytoarchitecture in dipteran evolution. Development. 2004;131:4251-61.

1379 116. Cardamone F, Piva A, Loser E, Eichenberger B, Romero-Mulero MC, Zenk F, et al.
1380 Chromatin landscape at cis-regulatory elements orchestrates cell fate decisions in early
1381 embryogenesis. Nature communications. 2025;16(1):3007.

1382 117. Janssens H, Siggens K, Cicin-Sain D, Jimenez-Guri E, Musy M, Akam M, et al. A
1383 quantitative atlas of Even-skipped and Hunchback expression in Clogmia albipunctata (Diptera:
1384 Psychodidae) blastoderm embryos. EvoDevo. 2014;5(1):1.

1385 118. Gonzaga-Saavedra N, Degen EA, Soluri IV, Croslyn C, Blythe SA. Nucleation-
1386 dependent propagation of Polycomb modifications emerges during the Drosophila maternal to
1387 zygotic transition. bioRxiv. 2025.

1388 119. Schulz KN, Bondra ER, Moshe A, Villalta JE, Lieb JD, Kaplan T, et al. Zelda is
1389 differentially required for chromatin accessibility, transcription factor binding, and gene
1390 expression in the early Drosophila embryo. Genome research. 2015;25(11):1715-26.

1391 120. Li XY, Harrison MM, Villalta JE, Kaplan T, Eisen MB. Establishment of regions of
1392 genomic activity during the Drosophila maternal to zygotic transition. eLife. 2014;3.

1393 121. Galouzis CC, Kherdjemil Y, Forneris M, Viales RR, Marco-Ferreres R, Furlong EEM.
1394 Chip (Ldb1) is a putative cofactor of Zelda forming a functional bridge to CBP during zygotic
1395 genome activation. Molecular cell. 2025;85(12):2425-41 e9.

1396 122. Giannios P, Tsitilou SG. The embryonic transcription factor Zelda of Drosophila
1397 melanogaster is also expressed in larvae and may regulate developmentally important genes.
1398 Biochemical and biophysical research communications. 2013;438(2):329-33.

1399 123. Hamm RL, Meisel RP, Scott JG. The evolving puzzle of autosomal versus Y-linked male
1400 determination in Musca domestica. G3 (Bethesda). 2015;5(3):371-84.

1401 124. Hamm DC, Larson ED, Nevil M, Marshall KE, Bondra ER, Harrison MM. A conserved
1402 maternal-specific repressive domain in Zelda revealed by Cas9-mediated mutagenesis in
1403 Drosophila melanogaster. PLoS genetics. 2017;13(12):e1007120.

1404 125. Pearson JC, Watson JD, Crews ST. Drosophila melanogaster Zelda and Single-minded
1405 collaborate to regulate an evolutionarily dynamic CNS midline cell enhancer. Developmental
1406 biology. 2012;366(2):420-32.

1407 126. Schröder C, Tautz D, Seifert E, Jäckle H. Differential regulation of the two transcripts
1408 from the *Drosophila* gap segmentation gene *hunchback*. EMBO J. 1988;7:2881-7.

1409 127. Judd J, Duarte FM, Lis JT. Pioneer-like factor GAF cooperates with PBAP (SWI/SNF)
1410 and NURF (ISWI) to regulate transcription. Genes & development. 2021;35(1-2):147-56.

1411 128. Tsukiyama T, Becker PB, Wu C. ATP-dependent nucleosome disruption at a heat-shock
1412 promoter mediated by binding of GAGA transcription factor. Nature. 1994;367(6463):525-32.

1413 129. Grand RS, Pregnolato M, Baumgartner L, Hoerner L, Burger L, Schubeler D. Genome
1414 access is transcription factor-specific and defined by nucleosome position. Molecular cell.
1415 2024;84(18):3455-68 e6.

1416 130. Polach KJ, Widom J. Mechanism of protein access to specific DNA sequences in
1417 chromatin: a dynamic equilibrium model for gene regulation. Journal of molecular biology.
1418 1995;254(2):130-49.

1419 131. Gao Q, Finkelstein R. Targeting gene expression to the head: the *Drosophila*
1420 *orthodenticle* gene is a direct target of the Bicoid morphogen. Development. 1998:4185-93.

1421 132. Hildebrandt K, Kolb D, Kloppel C, Kaspar P, Wittling F, Hartwig O, et al. Regulatory
1422 modules mediating the complex neural expression patterns of the homeobrain gene during
1423 Drosophila brain development. Hereditas. 2022;159(1):2.

1424 133. Kolb D, Kaspar P, Kloppel C, Walldorf U. The Drosophila homeodomain transcription
1425 factor Homeobrain is involved in the formation of the embryonic protocerebrum and the
1426 supraesophageal brain commissure. Cells Dev. 2021;165:203657.

1427 134. Walldorf U, Kiewe A, Wickert M, Ronshaugen M, McGinnis W. Homeobrain, a novel
1428 paired-like homeobox gene is expressed in the Drosophila brain. Mechanisms of development.
1429 2000;96(1):141-4.

1430 135. Ransick A, Davidson EH. A complete second gut induced by transplanted micromeres in
1431 the sea urchin embryo. Science. 1993;259(5098):1134-8.

1432 136. Ansari S, Troelenberg N, Dao VA, Richter T, Bucher G, Klingler M. Double abdomen in
1433 a short-germ insect: Zygotic control of axis formation revealed in the beetle Tribolium
1434 castaneum. Proceedings of the National Academy of Sciences of the United States of America.
1435 2018;115(8):1819-24.

1436 137. Schoppmeier M, Fischer S, Schmitt-Engel C, Lohr U, Klingler M. An ancient anterior
1437 patterning system promotes *caudal* repression and head formation in Ecdysozoa. Current biology
1438 : CB. 2009;19(21):1811-5.

1439 138. Grossniklaus U, Pearson RK, Gehring WJ. The Drosophila sloppy paired locus encodes
1440 two proteins involved in segmentation that show homology to mammalian transcription factors.
1441 Genes & development. 1992;6(6):1030-51.

1442 139. Schomburg C, Janssen R, Prpic NM. Phylogenetic analysis of forkhead transcription
1443 factors in the Panarthropoda. Development genes and evolution. 2022;232(1):39-48.

1444 140. Cadigan KM, Grossniklaus U, Gehring WJ. Functional redundancy: the respective roles
1445 of the two sloppy paired genes in Drosophila segmentation. Proceedings of the National
1446 Academy of Sciences of the United States of America. 1994;91(14):6324-8.

1447 141. Fujioka M, Jaynes JB. Regulation of a duplicated locus: Drosophila sloppy paired is
1448 replete with functionally overlapping enhancers. Developmental biology. 2012;362(2):309-19.

1449 142. Lee HH, Frasch M. Survey of forkhead domain encoding genes in the Drosophila
1450 genome: Classification and embryonic expression patterns. Developmental dynamics : an official
1451 publication of the American Association of Anatomists. 2004;229(2):357-66.

1452 143. Choe CP, Miller SC, Brown SJ. A pair-rule gene circuit defines segments sequentially in
1453 the short-germ insect Tribolium castaneum. Proceedings of the National Academy of Sciences of
1454 the United States of America. 2006;103(17):6560-4.

1455 144. Choe CP, Brown SJ. Evolutionary flexibility of pair-rule patterning revealed by
1456 functional analysis of secondary pair-rule genes, paired and sloppy-paired in the short-germ
1457 insect, Tribolium castaneum. Developmental biology. 2007;302(1):281-94.

1458 145. Choi HMT, Schwarzkopf M, Fornace ME, Acharya A, Artavanis G, Stegmaier J, et al.
1459 Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive,
1460 versatile, robust. Development. 2018;145(12).

1461 146. Bruce HS, Jerz G, Kelly SR, McCarthy J, Pomerantz A, Senevirathne G, et al.
1462 Hybridization chain reaction (HCR) in situ protocol V.1.; 2021.

1463 147. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native
1464 chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins
1465 and nucleosome position. Nature methods. 2013;10(12):1213-8.

1466 148. Krueger F. Trim Galore! : Babraham Bioinformatics; 2012.

1467 149. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature methods.
1468 2012;9(4):357-9.

1469 150. Lawrence M, Huber W, Pages H, Aboyoun P, Carlson M, Gentleman R, et al. Software
1470 for computing and annotating genomic ranges. PLoS computational biology.
1471 2013;9(8):e1003118.

1472 151. Morgan M, Pages H, Obenchain V, Hayden N. Rsamtools: Binary alignment (BAM),
1473 FASTA, variant call (BCF), and tabix file import.: Bioconductor; 2013.

1474 152. Pantano L. DEGreport: report of DEG analysis 2017 [

1475 153. Bergman CM, Carlson JW, Celniker SE. Drosophila DNase I footprint database: a
1476 systematic genome annotation of transcription factor binding sites in the fruitfly, Drosophila
1477 melanogaster. Bioinformatics. 2005;21(8):1747-9.

1478 154. Kulakovskii IV, Makeev V. [Integration of data obtained by different experimental
1479 methods to determine the motifs in DNA sequences recognized by transcription-regulating
1480 factors]. Biofizika. 2009;54(6):965-74.

1481 155. Kulakovskiy IV, Favorov AV, Makeev VJ. Motif discovery and motif finding from
1482 genome-mapped DNase footprint data. Bioinformatics. 2009;25(18):2318-25.

1483 156. Shazman S, Lee H, Socol Y, Mann RS, Honig B. OnTheFly: a database of Drosophila
1484 melanogaster transcription factors and their binding sites. Nucleic acids research.
1485 2014;42(Database issue):D167-71.

1486 157. Rohr KB, Tautz D, Sander K. Segmentation gene expression in the mothmidge Clogmia
1487 albipunctata (Diptera, psychodidae) and other primitive dipterans. Dev Genes Evol.
1488 1999;209(3):145-54.

1489 158. Smit A, Hubley R. RepeatModeler 2008 [Available from: https://www.repeatmasker.org/.

1490 159. Smit A, Hubley R, Green P. RepeatMasker 2013 [Available from:
1491 https://www.repeatmasker.org/.

1492 160. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, et al. Automated eukaryotic
1493 gene structure annotation using EVidenceModeler and the Program to Assemble Spliced
1494 Alignments. Genome biology. 2008;9(1):R7.

1495 161. Huerta-Cepas J, Szklarczyk D, Heller D, Hernandez-Plaza A, Forslund SK, Cook H, et al.
1496 eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource
1497 based on 5090 organisms and 2502 viruses. Nucleic acids research. 2019;47(D1):D309-D14.

1498 162. Cantalapiedra CP, Hernandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-
1499 mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the
1500 Metagenomic Scale. Molecular biology and evolution. 2021;38(12):5825-9.

1501 163. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO:
1502 assessing genome assembly and annotation completeness with single-copy orthologs.
1503 Bioinformatics. 2015;31(19):3210-2.

1504 164. Manni M, Berkeley MR, Seppey M, Zdobnov EM. BUSCO: Assessing Genomic Data
1505 Quality and Beyond. Curr Protoc. 2021;1(12):e323.

1506 165. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, et al. MCScanX: a toolkit for
1507 detection and evolutionary analysis of gene synteny and collinearity. Nucleic acids research.
1508 2012;40(7):e49.

1509 166. Gu Z, Gu L, Eils R, Schlesner M, Brors B. circlize Implements and enhances circular
1510 visualization in R. Bioinformatics. 2014;30(19):2811-2.

1511 167. Diesh C, Stevens GJ, Xie P, De Jesus Martinez T, Hershberg EA, Leung A, et al.
1512 JBrowse 2: a modular genome browser with views of synteny and structural variation. Genome
1513 biology. 2023;24(1):74.

1514 168. Hancock DY, Fischer J, Lowe JM, Fischer J, Lowe JM, Snapp-Childs W, et al.
1515 Jetstream2: Accelerating cloud computing via Jetstream. . Practice and Experience in Advanced
1516 Research Computing (PEARC '21). New York, NY: Association for Computing Machinery;
1517 2021.

1518 169. Boerner TJ, Deems S, Furlani TR, Knuth SL, Towns J. ACCESS: Advancing Innovation:
1519 NSF's Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support. Practice
1520 and Experience in Advanced Research Computing (PEARC '23). New York, NY: Association
1521 for Computing Machinery; 2023.

1522 170. Priyam A, Woodcroft BJ, Rai V, Moghul I, Munagala A, Ter F, et al. Sequenceserver: A
1523 Modern Graphical User Interface for Custom BLAST Databases. Molecular biology and
1524 evolution. 2019;36(12):2922-4.

1525 171. Beeber D, Chain FJ. crispRdesignR: A Versatile Guide RNA Design Package in R for
1526 CRISPR/Cas9 Applications. J Genomics. 2020;8:62-70.

1527 172. Brooks EM, Sanders SA, Pfrender ME. FreeCount: A coding-free framework for guided
1528 count data visualization and analysis. Practice and Experience in Advanced Research
1529 Computing 2024: Human Powered Computing (PEARC '24). New York, NY: Association for
1530 Computing Machinery; 2024.

1531 173. Xiao Z, Lam HM. ShinySyn: a Shiny/R application for the interactive visualization and
1532 integration of macro- and micro-synteny data. Bioinformatics. 2022;38(18):4406-8.

1533 174. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial
1534 sequencing and analysis of the human genome. Nature. 2001;409(6822):860-921.

1535 175. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for
1536 genome assemblies. Bioinformatics. 2013;29(8):1072-5.

1537