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

Authors

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

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

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

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

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

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

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

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

* Authors for correspondence

Author affiliations

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

Author Contributions

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

Abstract

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

Introduction

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

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

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

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

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

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

Results

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

We generated a high quality *de novo* genome assembly of *Clogmia albipunctata* with 6 chromosome-size scaffolds (**Figure 1, Table 1, S1 Appendix**), consistent with the known chromosome number in this species (70, 71). The total length of the chromosome-size scaffolds is 304.3 Mb, close to genome size estimates based on flow cytometry data (316.6 Mb) (72). An additional 268 small scaffolds (< 1 MB) contained mostly repetitive sequences. We annotated this new reference genome using RNA-seq transcripts from embryonic, larval, pupal, and adult stages (**S1 Table**), protein sequences, and *ab initio* gene prediction software, and identified 15,047 protein coding genes. The annotated genome recovered 88%-94% of Complete Benchmarking Universal Single Copy Orthologs (BUSCO) (**Table 2**), in line with recently published dipteran genomes (73-75). While synteny between the chromosomes within the Cyclorrhapha clade can be highly conserved (time of divergence ~145 million years ago) (74), synteny between the chromosomes of *D. melanogaster* and *C. albipunctata* (divergence time ~250 million years ago) was essentially lost.

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

171 *knirps*), pair-rule genes (*even-skipped*, *fushi tarazu*, *odd-skipped*, *runt*, *sloppy-paired*, *paired,*
172 *hairy, odd-paired*), and other important regulators (*caudal*, *Dichaete*) (7, 9). Some of these genes
173 have more than one copy in the Clogmia genome, including 2 copies of *hunchback* (*Cal-hb1,*
174 *Cal-hb2*), *knirps/knirps-like* (*Cal-kni/knrl1, Cal-kni/knrl2*), *orthodenticle* (*Cal-otd1, Cal-otd2*),
175 *even-skipped* (*Cal-eve1, Cal-eve2*), *tailless* (*Cal-tll1, Cal-tll2*) and *caudal* (*Cal-cad 1, Cal-cad2*),
176 3 copies of *sloppy-paired1/2/fd19B* (*Cal-slp1, Cal-slp2, Cal-slp3*), and 4 copies of *empty*
177 *spiracles* (*Cal-ems1, Cal-ems2, Cal-ems3, Cal-ems4*) (**Figure 1, S2 Table**). Such gene
178 duplications are a potential source of redundancies in Clogmia's segmentation gene network (see
179 below).

180

181 **Distinct patterns of dynamic changes of ATAC-seq peaks during nuclear cycles 11 through**
182 **14 suggest coordinated regulation of associated genes**

183 To identify accessible chromatin before and during the main wave of zygotic genome activation,
184 we performed single embryo ATAC-seq at NC11, NC12, NC13, NC14 (before cellularization),
185 and Late NC14 approximately 20 minutes before gastrulation. The embryos were staged *in vivo*
186 using developmental time after egg activation at 25ºC and morphological criteria as previously
187 described (69) and validated under our laboratory conditions. At least three replicates were
188 performed for NC11 though NC14, and two replicates were performed for Late NC14. We
189 pooled the ATAC-seq data from all stages to create a master peak list of 32157 open chromatin
190 regions (peaks), using MACS2 (78). We also computed the number of peaks for each
191 developmental stage excluding stage-specific peaks that were not recovered as significant in the
192 master peak list. In this analysis, we found that the overall number of peaks increases with each
193 nuclear cycle, suggesting that the chromatin landscape is continuously changing (**Table 3**). Since
194 promoter accessibility does not necessarily indicate gene expression (79), we divided the
195 accessible chromatin regions into transcription start site (TSS) proximal regions (<500 bp of
196 TSS) and TSS distal regions (>500 bp) to facilitate the distinction of open promoters and other
197 open cis-regulatory elements. ATAC-seq peaks that only mapped to intergenic or intronic
198 sequence were treated as candidate enhancers or silencers (henceforth collectively referred to as
199 enhancers). The relative proportion of candidate enhancers increased by about 10% from NC11
200 (46.8%) to NC14 (58%) and then decreased from NC14 to Late NC14 (52.7%) (**S1 Figure**).
201 These observations suggest similar dynamic changes in chromatin accessibility during zygotic
202 genome activation in Clogmia and Drosophila (80, 81).

203 To identify peaks with dynamic patterns of chromatin accessibility, we compared the degree of
204 accessibility of each peak between successive timepoints (**Figure 2A-D**) using DESeq2 (82).
205 This analysis identified a total of 6930 dynamically regulated peaks (adjusted p-value ≤ 0.05 and
206 $|\log_2 \text{Fold Change}| \geq 1$; 21.6% of all 32157 peaks) and enabled us to quantitatively determine the
207 behavior of each peak between consecutive developmental stages and to group them according to
208 their accessibility changes over time (**S2 Figure**). The majority (5403 / 6930 dynamic peaks;
209 78%) fell into 8 groups (**Figure 2E)**. Many of the dynamic peaks closed (14.6%; Groups 5,13) or
210 opened (~20.4%; Groups 8,12) gradually over time. The remaining dynamic peak groups
211 exhibited more complex dynamics between NC11 and Late NC14, commonly with an inflection
212 at NC13 (e.g., Groups 3,1,15,14; see also **S2 Figure**). These observations suggest that
213 remodeling of chromatin accessibility between NC11 and Late NC14 at individual peaks often
214 follows patterns that are shared by many peaks and may reflect co-regulation of the associated
215 genes.

216

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

218 Previous studies in Drosophila have shown that temporal changes in chromatin accessibility are
219 regulated by the interplay of transcription factors that can interact with nucleosome-bound DNA
220 (pioneer factors) and transcription factors that have limited or no ability to initiate chromatin
221 accessibility but affect the accessibility of open chromatin (83, 84). The earliest changes in
222 chromatin accessibility are primarily driven by the pioneer factor Zld (46, 80, 85-87). As zygotic
223 genome activation accelerates, additional pioneer factors, such as GAF (88), CLAMP (89, 90),
224 and Opa (58, 68) are required for maintaining, modulating, or gaining chromatin accessibility. In
225 Clogmia, a motif enrichment analysis of peaks with the MEME software package (91, 92)
226 revealed an enrichment of a Zelda-binding motif (5'-CAGGTA), GA-rich motifs recognized by
227 GAF and CLAMP (88, 93-95), in addition to $(CA)_n$ motifs recognized by Combgap (Cg) (96),
228 and others (**S3 Figure**). Opa-binding motifs were not recovered in this analysis. These results are
229 comparable to findings in Drosophila (58, 62, 68) and suggest that key pioneer factors acting
230 during zygotic genome activation are likely conserved between Drosophila and Clogmia.

231 Homologs of *zelda* have been identified in many insects and some crustaceans (97), and several
232 transcriptomic studies suggest that Zld's function as pioneer transcription factor during zygotic
233 genome activation is conserved across insects (98-102). However, Zld's pioneer activity is stage-
234 specific in *Drosophila* (103), and these comparative studies did not directly test for Zld-
235 dependent effects on chromatin accessibility. We identified Clogmia *zld* (*Cal-zld*) (**S4 Figure**)
236 and confirmed high levels of *Cal-zld* expression at preblastoderm and blastoderm stages (**S5**
237 **Figure**). Loss of Zld in *Drosophila* results in a failure to activate a large cohort of early zygotic
238 genes, including those associated with the process of cellularization (46). To examine phenotypic
239 effects of *Cal-zld* depletion, batches of ~30 embryos were injected with *Cal-zld* dsRNA (double-
240 stranded RNA) within the first hour of development (NC2/3). A subset of embryos from each
241 batch (3 replicates) were monitored *in vivo* and allowed to develop until the end of NC14
242 (roughly 7-7.5 hours after injection). Many of the embryos in each batch failed to develop a
243 blastoderm while some developed normally, presumably because of insufficient *Cal-zld*
244 knockdown. A third class of embryos reached NC14 but failed to cellularize and developed
245 aberrantly (**S1 Movie**). This phenotype is comparable to the phenotype of *in vivo* imaged *zld*
246 mutant Drosophila embryos (**S2 Movie**) (86) and suggests that the role of Zld for the regulation
247 of early zygotic genes is conserved in *Clogmia.* We also quantified knockdown efficiencies of
248 *Cal-zld* transcript in individual embryos by RT-qPCR (**S3 Table**). In one quarter of the *Cal-zld*
249 RNAi embryos (5/20), *Cal-zld* mRNA levels were reduced by ~30-90%.

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

261 *zld* RNAi embryos to those of the alignment control group using DESeq2. Loss of *Cal-zld*
262 affected 2230 genes of which 865 (38.8%) were downregulated and 1365 (61.2%) were
263 upregulated. We then compared genes differentially expressed in *Cal-zld* RNAi embryos to
264 genes we found to be expressed maternally, zygotically, or both (**Figure 3A, S7A Figure** and
265 *Materials and Methods*). 691 genes were classified as zygotically expressed of which 209
266 (30.2%) were downregulated in *Cal-zld* RNAi embryos and 63 (9.1%) were upregulated
267 (corresponding to 24.2% of all 865 downregulated and 4.6% of all 1365 upregulated genes,
268 respectively). 6658 genes were classified as expressed maternally and zygotically of which 593
269 (8.9%) were downregulated and 1232 (18.5%) were upregulated. Finally, of the 227 genes
270 classified as maternally expressed, only 5 (2.2%) were found to be downregulated and 14 (6.2%)
271 upregulated in *Cal-zld* RNAi embryos. Since the strictly zygotic genes were ~3 times more often
272 downregulated than upregulated (209 versus 63), our results suggest that *Cal-zld* plays an
273 important role in activating zygotic genes (**Figure 3A),** consistent with the role of Zld during
274 zygotic genome activation in Drosophila embryos (46, 85). Changes in the levels of maternal
275 transcripts in response to *Cal-zld* RNAi could reflect indirect posttranscriptional roles of *Cal-zld*
276 in their regulation.

277 At the level of chromatin accessibility, *Cal-zld* depletion affected 4102 of all 32157 peaks
278 (12.8%). 2399 of them lost accessibility (58.4%) and 1703 gained accessibility (41.5%). The
279 majority (67%) of the peaks mapped to intergenic or intronic sequences (**Figure 3B)**. Depletion
280 of *Cal-zld* mRNA affected both dynamic peaks and non-dynamic peaks (**S7B Figure**). Dynamic
281 peaks were affected by Cal-zld knockdown in different ways (**Figure 3C**). Peaks in groups that
282 progressively gained accessibility in wild-type embryos (Groups 8, 12, 15 and 14) were more
283 likely to be *Cal-zld* dependent and to close in Cal-zld RNAi background (**Figure 3C, *left*** and
284 **S8A Figure**). This observation is consistent with an important role of Cal-Zld in opening new
285 chromatin regions during zygotic genome activation. Additionally, our data suggest that Cal-Zld
286 also plays an important role in closing chromatin regions during zygotic genome activation (**S8B**
287 **Figure**). For example, peaks in dynamic groups whose accessibility decreased in wild-type with
288 each nuclear cycle or that closed after NC13 (Groups 5 and 1) were more likely to require Cal-
289 Zld to close (**Figure 3C, *right*** and **S8B Figure**).

290 To better distinguish direct from indirect Cal-Zld targets, we performed a Zld motif search for all
291 *Cal-zld* sensitive peaks, using Zld motifs derived from our MEME analysis in Clogmia (5-
292 CAGGTA; cf. **S3 Figure)** and experimental data in *Drosophila melanogaster* (Fly Factor
293 Survey) (99). Overall, 2857 of the 4102 *Cal-zld* sensitive peaks contained at least one Zld motif.
294 Zld motifs were mostly found in the peaks that were reduced following *Cal-zld* knockdown
295 (1952 peaks) but also in peaks that increased following *Cal-zld* knockdown (905 peaks). In peaks
296 that lost accessibility following *Cal-zld* knockdown, Zld motifs were *centrally* enriched within
297 50 bp of the peak summit (**Figure 3D** and **S9 Figure)**. This was not the case for peaks that
298 gained accessibility following *Cal-zld* knockdown. The location of Cal-Zld binding within peaks
299 may therefore differ depending on the recruitment of repressors or activators. Taken together,
300 our findings suggest that Cal-Zld plays a central role in both opening and closing chromatin
301 accessibility in pre-gastrulation embryos.

302

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

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

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

325 At NC12, very few changes in chromatin accessibility were observed in response to *Cal-opa*
326 RNAi. Out of the 86 differentially accessible regions, 79 (91.9%) lost accessibility and 7 (8.1%)
327 gained accessibility (**Figure 4A**). At NC13, depletion of *Cal-opa* had a more pronounced effect
328 on chromatin accessibility, with 250/292 (85.6%) peaks losing accessibility and 42/292 (14.4%)
329 gaining accessibility **(Figure 4B**). Most of the differentially accessible regions mapped to
330 putative enhancer regions (**S11 Figure**). We also performed a motif enrichment analysis using
331 the MEME suite on all peaks as well as the peaks that either lost or gained accessibility in *Cal-*
332 *opa* RNAi embryos at either NC12 or NC13 (**Figure 4C**). In regions that lost accessibility in
333 *Cal-opa* RNAi embryos, the most enriched motifs were GAF/CLAMP and Opa binding motifs.
334 In addition, we recovered motifs with no associated transcription factor. Regions that gained
335 accessibility were enriched in DNA binding motifs of GAF/CLAMP, Abdominal-B (Abd-B),
336 Twist (Twi), and two motifs with no associated transcription factor.

337 Next, we determined the number of *Cal-opa* sensitive peaks that contained putative Cal-Opa
338 binding sites, including Opa motifs derived from our MEME analysis and a MEME analysis of
339 ATAC-seq data from Drosophila (58). At NC12, 63 of the 79 peaks losing accessibility and none
340 of the 7 peaks gaining accessibility in response to reduced *Cal-opa* activity contained at least one
341 Opa motif (**S2 Appendix**). At NC13, 112 of the 250 peaks losing accessibility and 6 of the 42
342 peaks gaining accessibility in response to reduced *Cal-opa* activity contained at least one Opa
343 motif (**S3 Appendix**). These results support the hypothesis that Opa directly promotes chromatin
344 accessibility during the early phase of zygotic genome activation.

345

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

348 The early zygotic segmentation gene network of dipterans is dominated by TF-encoding genes,
349 such as the gap genes and pair-rule genes (7, 9, 67, 104-108). Since Bcd breaks axial symmetry
350 by targeting these genes in a concentration-dependent manner, we focused our search of Cal-Opa
351 targets on TF genes that are expressed during blastoderm formation. Using the RNA-seq data
352 from the embryos that were used to identify regions of differential chromatin accessibility, we
353 identified 84 differentially expressed genes at NC12, including 11 transcription factor genes, all
354 of which were downregulated upon *Cal-opa* knockdown (**Figure 5A** and **S2 Appendix**). At
355 NC13, we identified 1047 differentially expressed genes, including 116 putative transcription
356 factor genes of which 67 were downregulated and 49 were upregulated in *Cal-opa* RNAi
357 embryos (**Figure 5B** and **S3 Appendix**). We then asked which of the differentially expressed
358 transcription factor genes are located near open chromatin regions (ATAC-seq peaks) with an
359 Opa motif. This analysis was restricted to peaks within a 40 kb region centered on each gene
360 body of interest and further narrowed the list of potential key target genes of Cal-Opa to 10
361 genes at NC12 and 72 genes at NC13 (**S2 Appendix and S3 Appendix)**. At NC12, 4 of the 10
362 candidate genes were associated with at least one Cal-Opa dependent ATAC-seq peak, including
363 *Cal-opa,* and homologs of *homeobrain* (*Cal-hbn*) and *sloppy-paired* (*Cal-slp1*, *Cal-slp2*) (**Figure**
364 **6, S4 Table,** and **S12 Figure**). At NC13, we thereby identified *Cal-opa, Cal-slp1*, *Cal-slp2* and
365 *Cal-hbn* in addition to Clogmia orthologs of *Dichaete* (*Cal-D*), *Krüppel* (*Cal-Kr*), and an
366 unidentified C2H2 zinc finger gene among the 72 candidates (**Figure 7, S4 Table, S12 Figure**
367 **and S13 Figure**).

368 We tested the robustness of these results by repeating the analysis using the injection control
369 group as reference. At NC12, we recovered *Cal-opa*, *Cal-hbn, Cal-slp1,* and *Cal-slp2* as
370 candidate targets, but only three of them (*Cal-opa*, *Cal-hbn, Cal-slp1*) were associated with Cal-
371 Opa dependent ATAC-seq peaks with Cal-Opa binding sites. In the case of *Cal-slp2*, the
372 reduction of these peaks was insignificant (**S4 Table** and **S2 Appendix**). At NC13, we recovered
373 *Cal-hbn, Cal-slp1 and Cal-slp2* (**S4 Table** and **S3 Appendix**). *Cal-opa*, *Cal-Kr*, *Cal-D* and the
374 unidentified C2H2 zinc finger gene were not recovered in this analysis at NC13 because they did
375 not meet the threshold for statistical significance (settings identical for NC12 and NC13). We
376 therefore focused on Clogmia's *homeobrain* and *sloppy-paired* homologs.

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

388

389 **Discussion**

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

391 Dipteran embryos are morphologically similar (13, 109), but their mechanisms of axis
392 specification (14, 15), primordial germ cell specification (15, 110), extraembryonic tissue
393 specification (111-113), morphogenesis (112, 114, 115), and sex determination (116-119) are not
394 conserved across the dipteran clade. This clade is therefore an attractive model taxon to study
395 how molecular mechanisms of embryogenesis evolve. The segmentation process of flies is
396 particularly suitable for studying the evolution of developmental TF gene networks because it
397 unfolds before the completion of cellularization of the multinucleated embryo when gene
398 expression can be readily visualized and perturbed in non-traditional model organisms. In this
399 study, we have established the resources for dissecting the developmental gene network of early
400 Clogmia embryos by providing an annotated chromosome-level genome sequence for our inbred
401 laboratory strain and by describing chromatin accessibility of precisely staged consecutive
402 blastoderm stages. We have used these resources to examine how *zelda* and maternal *odd-paired*
403 activities shape chromatin accessibility and gene expression in Clogmia.

404 We find that, as in Drosophila, zygotic genome activation in Clogmia is accompanied by large-
405 scale changes in chromatin accessibility. We demonstrate that Cal-Zld is necessary for activating
406 an early wave of zygotic genes and that reduced *Cal-zld* expression results in widespread losses
407 of chromatin accessibility, consistent with a conserved role for Zld to pioneer chromatin
408 accessibility within the context of zygotic genome activation. However, we also observe
409 differences compared with Drosophila in the early patterns of establishment and maintenance of
410 accessible states, notably losses in accessibility during the cleavage divisions. Further, we show
411 that a subset of these regions depends on Zld to lose accessibility during cleavage stages.
412 Clogmia's anterior determinant, Cal-Opa, likewise has the ability to influence chromatin
413 accessibility states, and loss of Cal-Opa also results in reduced chromatin accessibility, albeit at a
414 smaller cohort of regions compared with Cal-Zld. These Opa-dependent regions include zygotic
415 loci whose functions we further hypothesize are necessary for mediating anterior specification
416 downstream of Opa. Collectively, these results are consistent with a model where anterior
417 determination in Drosophila and Clogmia, while driven by unrelated transcription factors, share a
418 common feature of symmetry breaking at the level of cis-regulatory sequence accessibility
419 driven by the anterior determinant. Mechanistically, this could occur through either a pioneering
420 activity which requires binding of nucleosome occupied DNA or through competition with
421 nucleosomes to bind DNA. Bcd does not pioneer chromatin as observed with Zld and instead
422 outcompetes nucleosomes above a concentration threshold (61, 84). In contrast, a vertebrate
423 homolog of Opa, Zic3, was found to bind nucleosomal DNA *in vivo*, suggesting that Cal-Opa
424 may pioneer chromatin directly (120).

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

437 These regions could potentially control important regulatory events prior to large scale zygotic
438 genome activation.

439 In Drosophila, Zld almost exclusively drives gains in chromatin accessibility (56, 58, 87, 121)
440 through direct binding to DNA motifs independent of nucleosome association, consistent with its
441 role as a pioneer factor (85). In contrast, we noticed that in Clogmia a subset of the regions that
442 lose accessibility during the syncytial cleavage divisions do so in a Zld-dependent manner. Based
443 on the prevalence of Zld motifs within this class of regions (53%), these effects are often likely
444 direct. At present, we cannot distinguish between a mechanism where Cal-Zld functions directly
445 as a context-specific repressor or indirectly by permitting the binding of an independent
446 repressive factor. Further work, including confirming the direct targets of Cal-Zld, and
447 identification of cooperating repressors, will be needed to answer this question. It is worth
448 pointing out that the accessibility dynamics that we observe in Clogmia may reflect the
449 consequences of development with longer cell cycle timing. In Clogmia, NC11 through NC13
450 take ~3x longer than in Drosophila, where NC11 only lasts ~10 minutes, NC12 ~12 minutes, and
451 NC13 ~21 minutes (69), and the much shorter cell cycle times of Drosophila present significant
452 limitations to stable establishment and maintenance of chromatin states (62). Assuming that
453 kinetic rates for nuclear import, DNA replication, and DNA binding of transcription factors is
454 similar between Clogmia and Drosophila, the significantly longer syncytial interphase times in
455 Clogmia could alleviate the impact of cell cycle dependent constraints on transcriptional
456 regulation. Indeed, experimental lengthening of the Drosophila syncytial interphase can result in
457 premature transcriptional activation (122).

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

459 We find that homologs of *homeobrain* and *sloppy-paired* are targets of Cal-Opa to initiate
460 anterior pattern formation. These genes may play key roles in initiating anterior pattern
461 formation not just in Clogmia but also in other insects. In Drosophila, *homeobrain* is required for
462 labrum and brain development (123-125) and associated with enhancers that open in a Bcd-
463 dependent manner (56), and in the beetle *Tribolium castaneum*, *homeobrain* was found to play
464 an important zygotic role in establishing the embryo's anterior-posterior polarity (126). The
465 *sloppy-paired* gene, a member of the *forkhead* domain gene family (127), is best known for its
466 role as pair-rule segmentation gene (9, 128), but it is first expressed in a head domain of the
467 syncytial Drosophila embryo, like its closest paralogs, *sloppy-paired 2* and *fd19B* (129-132).
468 Partially redundant functions of sloppy-paired paralogs, which are common across insects (**S14**
469 **Figure and S15 Figure**), could have obscured a fundamental role of *sloppy-paired* activity in
470 head development.

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

482 expressed in the anterior region of the embryo (e.g., *orthodenticle*, *empty spiracles*, *buttonhead*,
483 *giant*, *knirps*, *hunchback*, *sloppy-paired 1*, *fd19B*, *even-skipped*, *paired*, *cap'n'collar*, *huckebein,*
484 *homeobrain*), and many of them correspond to enhancers that are known to drive expression of a
485 reporter gene in the anterior embryo (e.g., *giant* anterior (-6), *knirps* anterior, *orthodenticle*,
486 *buttonhead, even-skipped* stripe 1/5, *hunchback* P2, *hunchback* shadow enhancers, *eve-skipped*
487 stripe 2*,* an early *paired* enhancer, *giant* anterior (-10), *cap'n'collar*, *huckebein;* reviewed in
488 (56).

489 In Clogmia, the expression of both *hunchback* orthologs (*Cal-hb1*, *Cal-hb2*) was reduced in *Cal-*
490 *opa* RNAi embryos along with associated ATAC-seq peaks but these regions did not include
491 Opa motifs (**S16A Figure** and **S16B Figure**). Likewise, the expression of one of Clogmia's two
492 orthologs of *orthodenticle* (*Cal-otd1*) was reduced at NC13 in *Cal-opa* RNAi embryos but was
493 not associated with Opa-motif containing *Cal-opa* sensitive peaks (**S16C Figure**). The second
494 ortholog (*Cal-otd2*) was not expressed at NC12 and NC13. In the case of Clogmia *buttonhead*
495 (*Cal-btd*), we identified a *Cal-opa* sensitive peak with an Opa motif but the observed reduction
496 of *Cal-btd* expression was not significant (**S16D Figure**). *Cal-btd* expression could be under the
497 control of AD-dependent and AD independent enhancers, as previously reported for
498 *orthodenticle* in Drosophila (133). Clogmia orthologs of *giant* (*Cal-gt*), *knirps* (*Cal-kni/knrl1,*
499 *Cal-kni/knrl2*), and *empty spiracles* (*Cal-ems1-4*) were not expressed at NC12 or NC13 and no
500 significant changes in chromatin accessibility were detected at these loci at NC12 or NC13
501 following knockdown of *Cal-opa*. The ortholog of *huckebein* (*Cal-hkb*) was not expressed at
502 NC12 or NC13, either; however, its locus contained one Opa motif positive ATAC-seq peak that
503 lost accessibility following *Cal-opa* knockdown at NC13. The expression of Clogmia's
504 *cap'n'collar* (*Cal-cnc*) and *paired* (*Cal-prd*) was unaffected by *Cal-opa* RNAi, and the ATAC-
505 seq peaks at their loci were largely *Cal-opa* insensitive. Only one small intronic peak within the
506 Cal-*cnc* locus displayed reduced accessibility in *Cal-opa* RNAi embryos at NC13. The loci of
507 *even-skipped* orthologs (*Cal-eve1, Cal-eve2*) contain ATAC-seq peaks with Opa motifs that were
508 insensitive to *Cal-opa* RNAi. We note that the expression of *Cal-eve2* was elevated in *Cal-opa*
509 RNAi embryos at NC13. Taken together, these observations point to significant differences in
510 the key targets of Cal-Opa in Clogmia and Bcd in Drosophila and raise the question of how the
511 downstream segmentation gene networks of these species converge.

512

513 **Materials and Methods**

514 **Clogmia culture and embryo collection**

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

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

528

529 **Embryo fixation**

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

540

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

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

556

557 **dsRNA synthesis**

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

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

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

571 *Cal-zld* (a combination of two dsRNAs were used 579 bp and 668 bp):
572 5'-TAATACGACTCACTATAGGGAGAAGTCCCGCAATTGATACAGC
573 5'-TAATACGACTCACTATAGGGAGAAGGATGTTGGTGGACACTCC
574 5'-TAATACGACTCACTATAGGGAGACGGAATGGTGTGTGAAACAG
575 5'-TAATACGACTCACTATAGGGAGATTCGAGGGGTTATTGTCCTG

576 dsRed (273 bp):
577 5'-CAGAGATGCATAATACGACTCACTATAGATGCAGAAGAAGACTATGG
578 5'- CAGAGATGCATAATACGACTCACTATAGCTACAGGAACAGGTGGTG

579

580 **RT-qPCR**

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

583 Forward and reverse primer sequences for targets (lengths of amplicon in brackets):
584 *Cal-opa*, maternal transcript (112 bp):
585 5'- TTGCGCTTCATTCTGTCATCG
586 5'- CACTGAGGGCTGATTCTCGG
587
588 *Cal-zld* (138 bp):
589 5'-AAAGATGAACCACCGTGCGA
590 5'-CATTGATGGCAGTGGACCCT
591
592 *Cal-rpl35* (119 bp):
593 5'- CTGTCCAAAATCCGAGTCGT
594 5'- AGGTCGAGGGGCTTGTATTT
595

596 **Fluorescent HCR *in situ* hybridizations**

597 Gene specific probes and amplifier-fluorophore sets were designed and synthesized by
598 Molecular Instruments Inc (Los Angeles, CA). The amplifier-fluorophore set used was B1-647
599 for *Cal-slp1*, B2-546 for *Cal-hbn*, and B3-488 for *Cal-slp2*. Buffers (probe hybridization, probe
600 wash and amplification buffers) throughout the protocol were purchased from Molecular
601 Instruments Inc. The HCR *in situ* hybridizations were performed as described (134, 135).
602 Briefly, fixed embryos were rehydrated and washed 3 times in PBS-Tween20 (1x PBS, 0.1%
603 Tween-20). Embryos were permeabilized for 30 minutes at room temperature (100 ml
604 Permeabilization buffer: 100 µL 100% Triton X-100, 50 µL Igepal CA-630, 50 mg Sodium
605 Deoxycholate (Sigma D6750-25G), 50 mg Saponin (Sigma 47036-50G-F), 200 mg BSA
606 Fraction V (Sigma A5611-5G). Following embryos were pre-hybridized in probe hybridization
607 buffer for 30 minutes in a 37° C water bath. Once pre-hybridization was complete, fresh
608 hybridization solution containing gene specific probes were added and the embryos were left to

609 incubate overnight in a 37º C water bath. On the second day the hybridization solution was
610 removed, and the embryos were washed 4 times with pre-warmed probe wash buffer followed by
611 two washes in 5x SSCT (5x SSC, 0.1% Tween-20). An amplifier hairpin solution was prepared
612 following the manufacturer's instructions. Embryos were incubated with pre-warmed
613 amplification buffer for 30 minutes at room temperature before adding amplifier hairpin solution.
614 Embryos were left to incubate overnight at room temperature in the dark. The following day the
615 embryos were washed 5 times in 5x SSCT at room temperature. The embryos were then counter
616 stained with DAPI and mounted in Aqua-Polymount for imaging. Confocal imaging was
617 performed on the Zeiss LSM 900.

618

619 **Image Acquisition and Processing**

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

627

628 **ATAC-seq library preparation**

629 Single embryo ATAC-seq library preparations were done essential as described previously in
630 *Drosophila melanogaster* (58, 62). Embryos developed in an incubator at 25ºC until the desired
631 nuclear cycle. Nuclear cycle was confirmed by time after egg activation and the morphology was
632 checked under water using a transmission light microscope (Zeiss Stemi 2000-C). Once at the
633 desired nuclear cycle individual embryos were placed in the lid of a 1.5 ml low-retention
634 Eppendorf tube and homogenized in 10 µl of cold lysis buffer (10 mM Tris-HCl pH 7.4; 10 mM
635 NaCl; 3 mM MgCl2; 1% Igepal CA-630) (136), using a fire-sealed microcapillary tube
636 (Drummond Microcap 25). Once homogenized, 40 µl of lysis buffer was added to the lid, and the
637 body of the tube was carefully closed over the lid. Samples were place on ice until all embryos
638 were processed. The nuclei were pelleted by centrifugation at 800 RCF at 4ºC for 10 minutes.
639 Supernatant removal was observed under a dissection scope to ensure that the nuclei pellet was
640 not dislodged. The nuclear pellet was resuspended in 7.5 µl of Tagmentation buffer (Illumina
641 Tagment DNA TDE1 Enzyme and Buffer Kit) and placed on ice until all samples were
642 resuspended. Then 2.5 µl of Tn5 Transposase was added and mixed by pipetting up and down.
643 Tagmentation and amplification of ATAC-seq libraries were performed as described previously
644 (Buenrostro et al., 2015). The tagmentation reaction was performed at 37ºC and 800 RPM for 30
645 minutes using an Eppendorf Thermomixer. Immediately after this step, the reactions were
646 purified using the Qiagen Minelute kit following the manufacturer's instructions. Purified
647 tagmented DNA was eluted in 10 µl of buffer EB. Library amplification was performed as
648 described previously (Buenrostro et al., 2015). Typically, 12-14 total PCR were performed.
649 Amplified libraries were purified using 1.8x Ampure SpRI beads following the manufacturer's
650 instructions. Purified ATAC-seq libraries were eluted in 15 ul. The ATAC-seq library profile
651 was evaluated using Agilent high sensitivity bio-analyzer. Concentrations were estimated using a
652 Qubit fluorometer. Libraries from the same experiment were pooled and sequenced together.

653 Sequencing was done at the University of Chicago Genomics Core (Chicago, IL, USA) using the
654 Illumina NextSeq (PE 75 bp) and at Admera Health (South Plainfield, NJ, USA) using the
655 Illumina HiSeq (PE 150 bp) and the Illumina NovaSeq X Plus (PE 150 bp).

656 ATAC-seq libraries were amplified using a set of modified Buenrostro primers that introduced
657 Unique Dual Indexes (UDI) (58). The generalized UDI ATAC primer sequences are:

658 ATAC UDI Index Read 2 (i5):
659 5'- AATGATACGGCGACCACCGAGATCTACACnnnnnnnnTCGTCGGCAGCGTCAGATG
660 T*G -3'

661 ATAC UDI Index Read 1 (i7):
662 5'- CAAGCAGAAGACGGCATACGAGATnnnnnnnnGTCTCGTGGGCTCGGAGATG*T -3'

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

665

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

667 Embryos developed in an incubator at 25ºC until the desired nuclear cycle. Nuclear cycle was
668 confirmed by time after egg activation and the morphology was checked under water using a
669 transmission light microscope (Zeiss Stemi 2000-C). RNA isolation was performed using the
670 Zymo Quick-RNA Tissue/Insect Microprep Kit following the manufacturer's instructions except
671 for the tissue homogenization step. Instead, homogenization was done as follows. Once at the
672 desired nuclear cycle individual embryos were placed in the lid of a 1.5 ml low-retention
673 Eppendorf tube and homogenized in 10 µl of cold lysis buffer (10 mM Tris-HCl pH 7.4; 10 mM
674 NaCl; 3 mM MgCl2; 1% Igepal CA-630) (136), using a fire-sealed microcapillary tube
675 (Drummond Microcap 25). Once homogenized the lysate was added to 450 µl of Zymo RNA
676 Lysis Buffer and thoroughly mixed. RNA isolation resumed following the manufacturer's
677 instructions including DNAse treatment. Purified RNA was eluted in DNase/RNase-Free Water.

678

679 **RNA-seq library preparation**

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

688

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

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

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

698

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

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

706

707 **ATAC-seq data processing**

708 Demultiplexed reads were trimmed of adapters using TrimGalore! (137) (version 0.6.10,
709 cutadapt version 4.4) and mapped to the *Clogmia albipunctata* assembly (calbi-uni4000) using
710 Bowtie2 (138) (version 2.4.1) with option -X 2000. Suspected optical and PCR duplicates were
711 marked by Picard MarkDuplicates (version 2.21.4) using default parameters
712 (https://broadinstitute.github.io/picard/). Mapped, trimmed, duplicate marked reads were
713 imported into R using the GenomicAlignments (139) (version 1.38.2) and Rsamtools (140)
714 (version 2.18.0). Reads that mapped, were properly paired, non-secondary and had a map quality
715 scores ≥ 10 were imported. Only reads mapping to scaffolds 1-5 and scaffold 7 were used for
716 downstream analysis. The start and end coordinates of the reads were adjusted to account for the
717 Tn5 interaction site. Watson strand start sites had four base pairs subtracted, and Crick-strand
718 start sites had five base pairs subtracted (136). Reads with mapped length ≤ 120 bp were
719 considered to have originated from 'open' chromatin.

720 *MACS2*

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

726 *DESeq2*

727 A count matrix for differential enrichment analyses using DESeq2 (82) (version 1.42.1) was
728 generated by counting the number of open ATAC reads (length ≤ 120 bp) that overlapped with a
729 MAC2 identified peak. Before running DESeq2, peaks with low count values were removed.
730 DESeq2 design parameters were specific to the experimental data set. For the WT stage-wise
731 comparisons (e.g. NC12 vs NC11, NC13 vs NC12, NC14 vs NC13 and LNC14 vs NC14) the
732 design parameters passed to DESeq2 were stage and sequencing platform (~instrument.atac.data
733 + stage). For the comparison between RNAi embryos and alignment control embryos (e.g., *Cal-*
734 *zld* RNAi vs alignment control and *Cal-opa* RNAi vs alignment control) the design components
735 were embryo batch and genotype (~Experiment.Date + genotype). All control groups were given

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

Dynamic peaks were defined as peaks that underwent a statistically significant change in accessibility in at least one of the WT stage wise comparisons. Different dynamic peak classes were found using the R package DEGreport (141) (R package version 1.38.5) with the parameter of the minimum group size set to 10.

MEME

MEME (91) (version 5.4.1) analysis was performed using the following options -order 1 -meme-mod zoops -maxw 14 -meme-nmotifs 15 -meme-searchsize 100000 -centrimo-score 5.0 -centrimo-ethresh 10.0. Motifs were reported from all enrichment or discovery programs within the MEME suite (i.e., MEME, STREME (92), CentriMo), unless specified otherwise. MEME was given a set of sequences containing the 100 bp sequence centered on the summit of the peak. A set of background sequences were also given to MEME. To determine motifs enriched in dynamic peaks a set of non-dynamic peak sequences were provided as background sequences. To determine motifs enriched in *Cal-zld* sensitive peaks a set of sequences from peaks unaffected by *Cal-zld* were provided as background sequences. To determine motifs enriched in *Cal-opa* sensitive peaks a set of sequences from peaks unaffected by *Cal-opa* were provided as background sequences. Identified motifs were compared to known *Drosophila melanogaster* motifs from the following databases: OnTheFly_2014, Fly Factor Survey, FLYREG, iDMMPMM, and DMMPMM databases (142-146).

RNA-seq data processing

Demultiplexed reads were trimmed of adapters using TrimGalore! (137) (version 0.6.10, cutadapt version 4.4; https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) and mapped to the *Clogmia albipunctata* assembly (calbi-uni4000) using Bowtie2 (138) (version 2.4.1) with default parameters. Suspected optical and PCR duplicates were marked by Picard MarkDuplicates (version 2.21.4) using default parameters. Mapped, trimmed, duplicate marked reads were imported into R using the GenomicAlignments (139) (version 1.38.2) and Rsamtools (140) (version 2.18.0). Reads that mapped, were properly paired, non-secondary and had a map quality scores ≥ 10 were imported.

DESeq2

A gene count matrix for differential enrichment analyses was generated by counting the number of RNA-seq reads that overlapped with a transcript of a gene. Before running DESeq2, genes with low count values were removed. DESeq2 design parameters were specific to the experimental data set. For the comparison between RNAi embryos and alignment control embryos (e.g., *Cal-zld* RNAi vs alignment control and *Cal-opa* RNAi vs alignment control) the design components were embryo batch and genotype (~Experiment.Date + genotype). All control groups were given their own genotype to account for differences in embryo manipulation, and the genotype for the RNAi embryos was determined by the strength of knockdown (see *Fold change calculations*). The criteria for a statistically significant change

were a $\log_2$ Fold Change ≥ 1 and an adjusted p-value ≤ 0.05 or a $\log_2$ Fold Change ≤ -1 and an adjusted p-value ≤ 0.05.

Maternal, zygotic, maternal and zygotic classification

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

Fold change calculations

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

Genome Assembly

We used Cantata Bio's genome assembly services to generate a *de novo* reference genome for *Clogmia albipunctata* from ~70 freshly eclosed males of a 12-generation inbred line of an established laboratory culture (147). High-fidelity (HiFi) PacBio reads were used to generate an initial draft assembly subsequently refined with Omni-C reads (**S1 Appendix**). Details on sequencing and software used to assemble the genome have been described elsewhere (74). We estimated genome heterozygosity and sequencing error rates using jellyfish to count 21-mers and GenomeScope2.0 to analyze the k-mer frequencies.

Genome Annotation

We used *RepeatModeler* (148) (version v2.0.4) to identify repeat elements in the *Clogmia albipunctata* genome and *RepeatMasker*'s (149) (version v4.1.5) *fambd.py* script to extract *Arthropoda* records from the *Dfam* database. These sequences were combined to generate a custom repeat library, which was then used to soft-mask the genome. To predict gene structures,

we used RNA-seq and protein evidence. mRNA data from 26 individual embryos of various stages, pooled samples of first and fourth-instar larvae, one female and one male pupa generated in this study were mapped to the genome (**S1 Table**). Additionally, 11 RNA samples (9 embryonic and two adult) from other studies were downloaded from NCBI (**S1 Table**). For protein evidence, we downloaded all available Dipteran sequences from NCBI's RefSeq (January 2024) and UniProt's complete protein sequences (Release 2023_04). A detailed description of the annotation pipeline is available elsewhere (74). Briefly, we assembled and mapped the RNA transcripts and protein sequences to the reference and used these data as evidence of gene structures and to predict gene models. *Evidence Modeler* (150) (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* (151, 152) (version 2.1.12), assigning gene names based on orthology. We assessed the final annotation quality using BUSCO (Benchmarking Universal Single-Copy Orthologs) (153, 154) on the predicted protein sequences. Finally, we analyzed synteny between the *C. albipunctata* and *D. melanogaster* genomes using *MCScanX* (155) (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 (156) (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*. This work used Jetstream2 at Indiana University through allocation BIO220075 from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program, which is supported by National Science Foundation grants #2138259, #2138286, #2138307, #2137603, and #2138296. Bioinformatic work was supported in part by the Notre Dame University Genomics and Bioinformatics Core Facility.

Competing interests

The authors declare no competing interests.

Funding

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

Data and Resource Availability

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

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

Genome Browser: The genomic data produced in this study can be further interrogated using a genome browser ecosystem centered on JBrowse2 (157) and delivered as a cloud image designed for individual use, currently available on NSF's Jetstream2 cloud platform, with a portable Docker image in development. Jetstream2's user-friendly interface (image: *Clogmia albipunctata* Genome Resources) (158, 159) integrates tools for genomic analysis, including BLAST search (SequenceServer2.0) (160), CRISPR guide RNA design (modified crisprDesigner) (161), R shiny-driven differential gene expression (DGE) analysis (freecount) (162), and synteny mapping for comparative genomics (ShinySyn) (163). Each tool is connected back to the browser, leveraging JBrowse2's ability to dynamically visualize and contextualize the genomic data. BLAST results link directly to genomic regions, enabling users to explore expression tracks, variants, and synteny relationships. CRISPR tools provide sgRNA design with metrics for selection and visualization of target regions alongside expression data, off-target effects, and variants. DGE analysis visualizations are linked back to the genome browser for further exploration. All gene model annotations in the browser connect to common external databases, such as NCBI, Uniprot, FlyBase, and Google Scholar, for further exploring and interpreting the data. Future enhancements will include expanded sgRNA profiling, support for Docker to enable deployment on commercial cloud platforms, and further optimization of workflows, ensuring this ecosystem remains a powerful and accessible solution for the use of our genomic data presented here and elsewhere (74) for research and education. To access the *Clogmia albipunctata* genome browser and related tools, create an ACCESS ID and use this ID to create an ACCESS account to log in to Jetstream2 to apply for an 'allocation' of credits that can be used on Jetstream2. Detailed instructions on the use of Jetstream2 can be found at https://jetstream-cloud.org/get-started/index.html. This on-demand model reduces resource costs and security concerns compared to hosting a centralized site. Administrative tools, (e.g. automated reboot scripts, system service management, and load balancing for multi-user groups or workshops) enhance usability while minimizing operational overhead.

904 **Tables**

905

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

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

907

908

909 **Table 2: *Clogmia albipunctata* annotation statistics and BUSCO scores.** The quality metrics
910 include the percentage of complete universal orthologs identified in the annotation across five
911 BUSCO datasets. The number of genes in each lineage-specific BUSCO database is shown in
912 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%

913

914

915

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

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

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

922 **Figures**
923

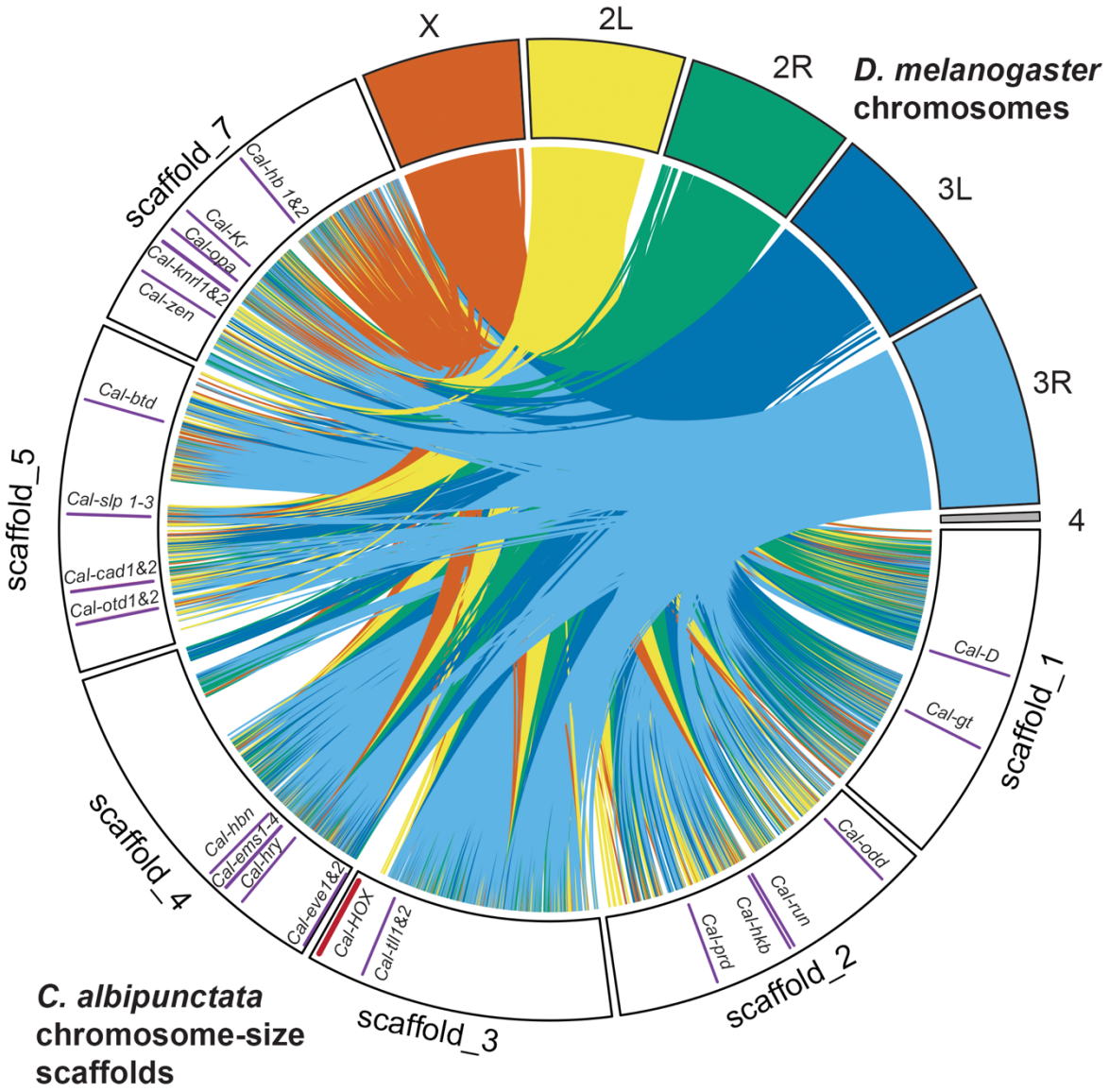

Figure 1: Synteny analysis of *Drosophila melanogaster* chromosomes and *Clogmia albipunctata* scaffolds. Synteny analysis between *Clogmia albipunctata* chromosomes-sized scaffolds *and Drosophila melanogaster* chromosomes. Groups of collinear genes are represented by contact lines connecting the positions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

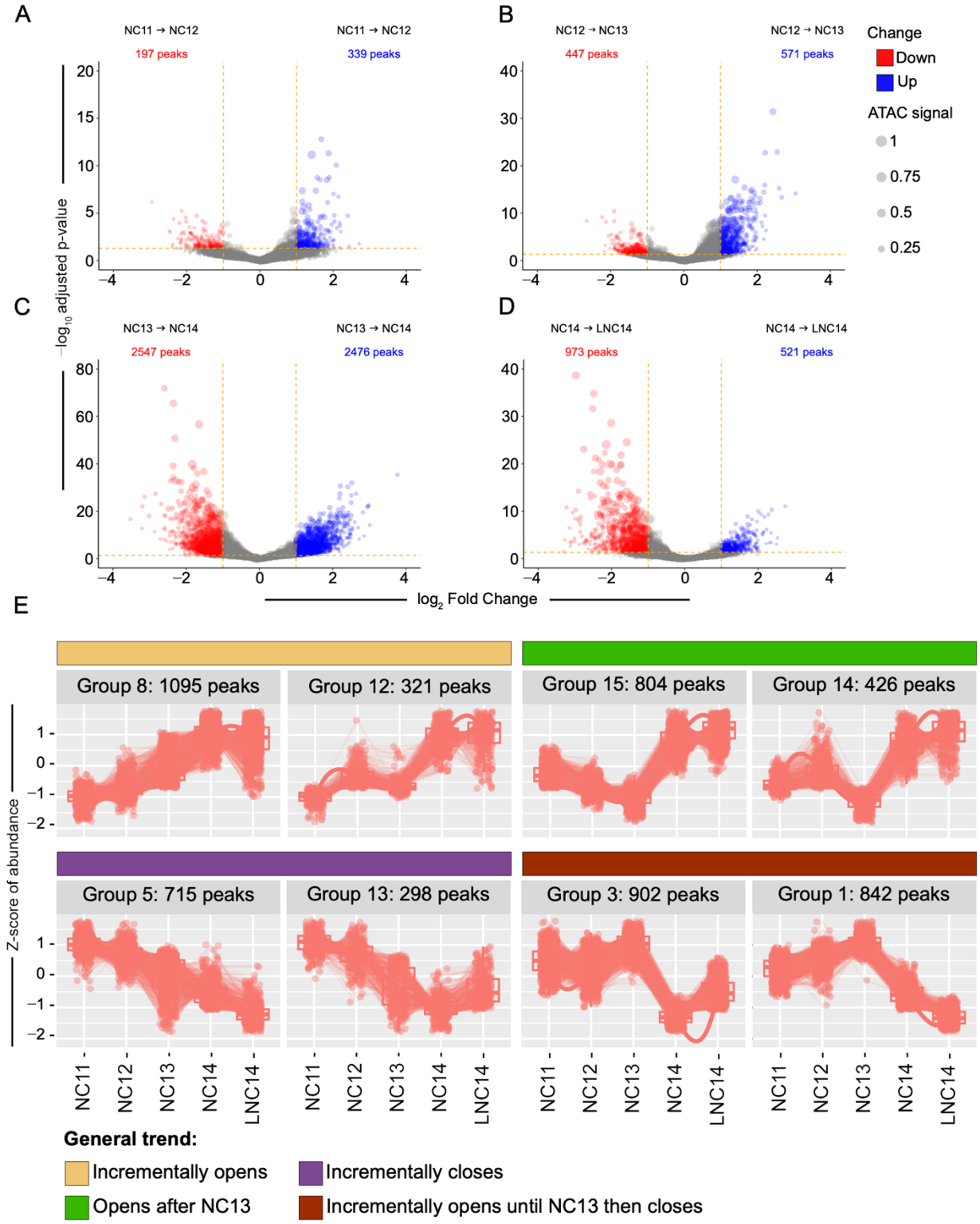

Figure 2: **Changes in chromatin accessibility during NC11 to Late NC14.** A-D) Volcano plots of differentially accessible ATAC-seq peaks between consecutive developmental stages. Significant change of adjusted p-value ≤ 0.05 and $|\log_2$ Fold Change$| \geq 1$ highlighted; red ($\log_2$ Fold Change ≤ -1) and blue ($\log_2$ Fold Change ≥ 1). E) Dynamic peak groups produced using DEGreport (141) (R package version 1.38.5). Y-axis, Z-score of abundance. X-axis, stage.

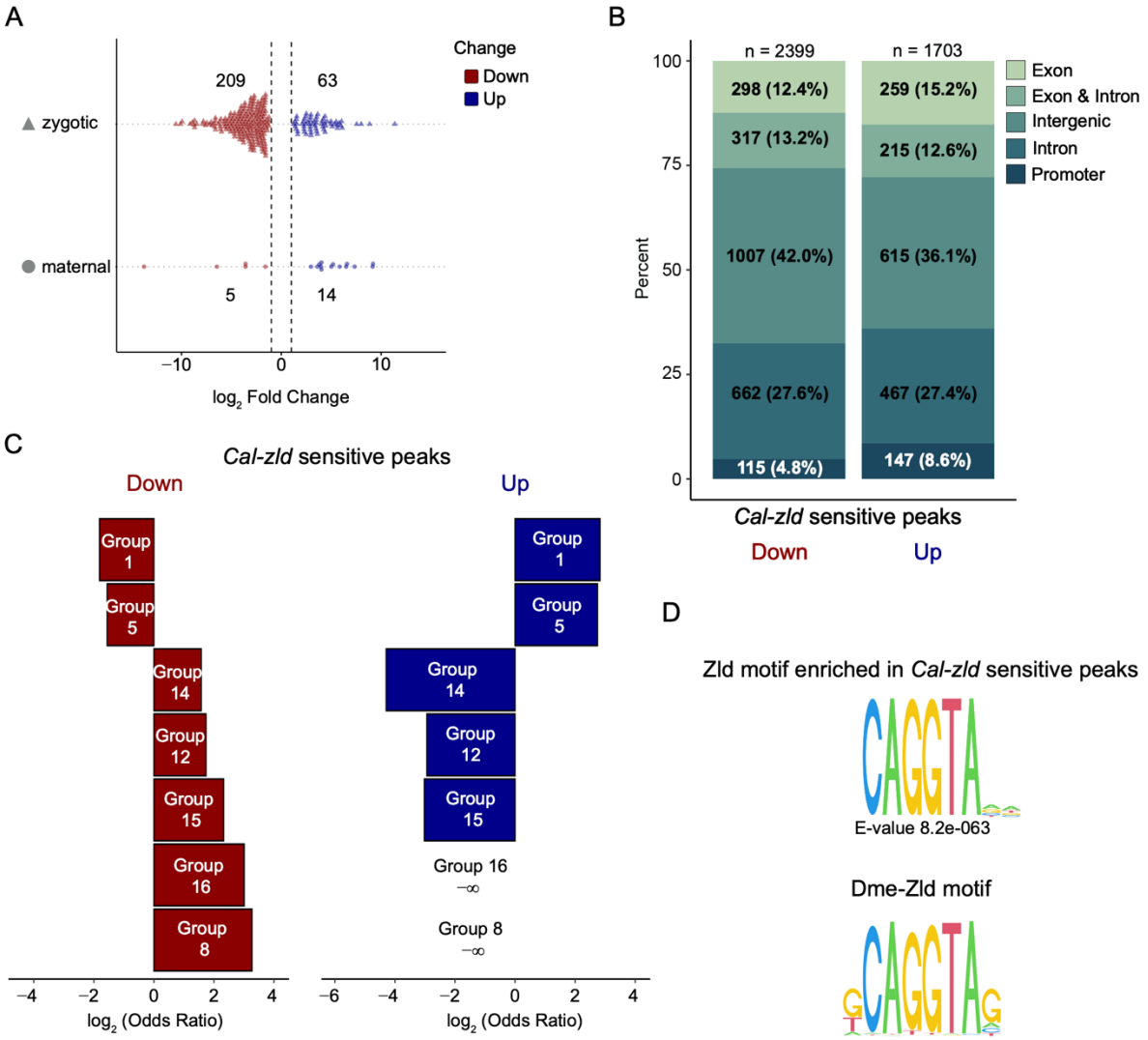

Figure 3: Cal-Zld regulates chromatin accessibility and zygotic transcription positively and negatively. A) Observed expression change of maternal or zygotic genes that are differentially expressed in *Cal-zld* RNAi embryos. Genes are grouped by maternal or zygotic classification on the y-axis. Log_2 Fold Change on the x-axis. Only genes with an adjusted p-value ≤ 0.05 and $|\log_2$ Fold Change$| \geq$ are shown. 1B) 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.

957 Only groups with a $\log_2$(Odds Ratio) > 1.5 in at least one comparison are shown. D) Position
958 weight matrix (PWM) logo of the Zld motif enriched in *Cal-zld* sensitive peaks as identified by
959 MEME (top) and the canonical Zld motif in *Drosophila melanogaster*, shown for reference
960 (bottom).
961

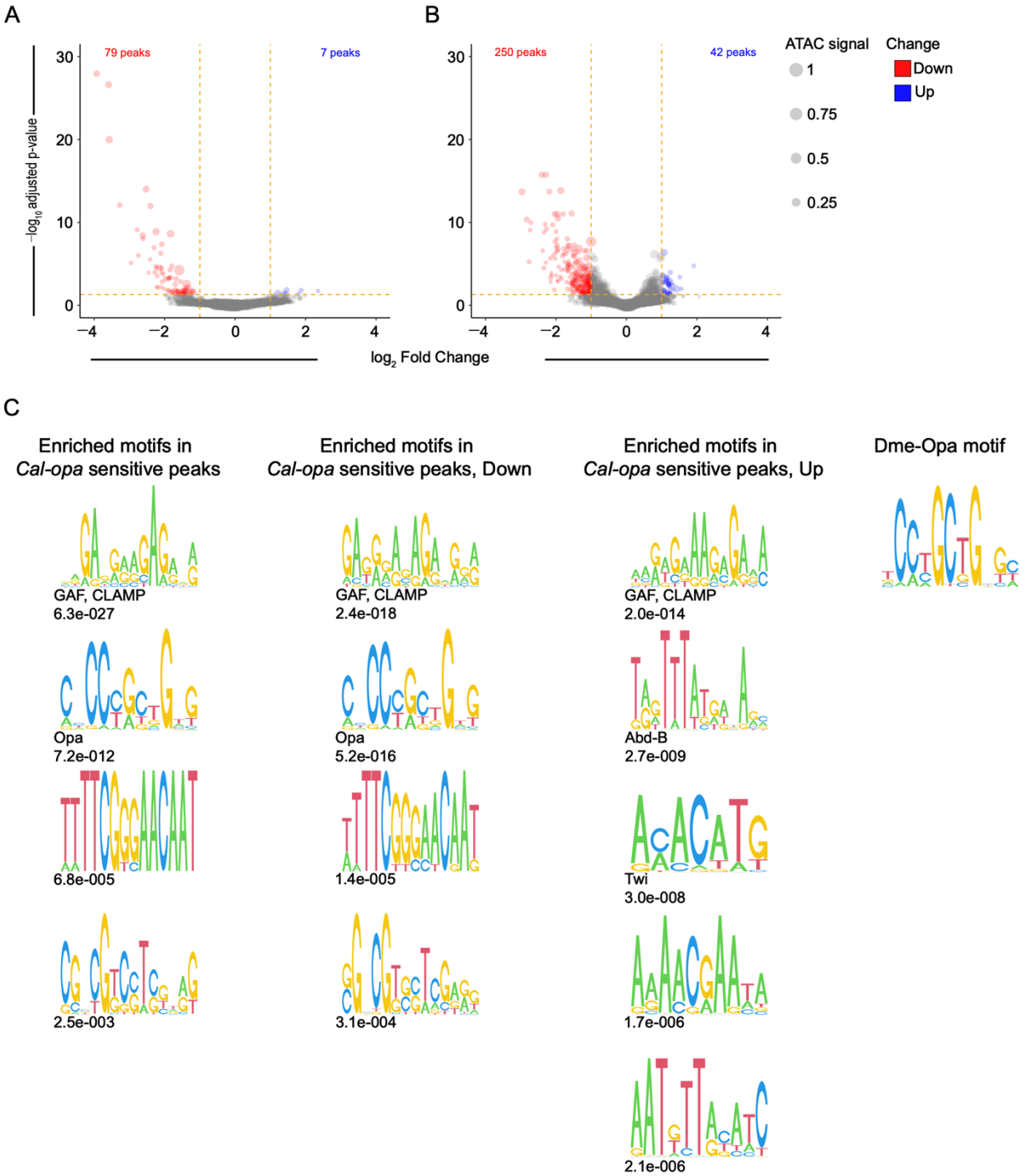

Figure 4: Cal-Opa is required for establishing a selective set of chromatin accessibility peaks. Volcano plots of differentially accessible peaks between the *Cal-opa* RNAi group and the alignment control group at A) NC12 and B) NC13. Significant change of adjusted p-value ≤ 0.05 and |$\log_2$ Fold Change| ≥ 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

970 stages. DNA binding proteins of *Drosophila melanogaster* that bind to these or similar motifs as
971 identified by MEME are indicated. We include CLAMP in this list as it has been shown to bind
972 GA-rich motifs (see text). E-value denotes the significance of recovered motifs. The canonical
973 Opa motif in *Drosophila melanogaster* is shown for reference.
974

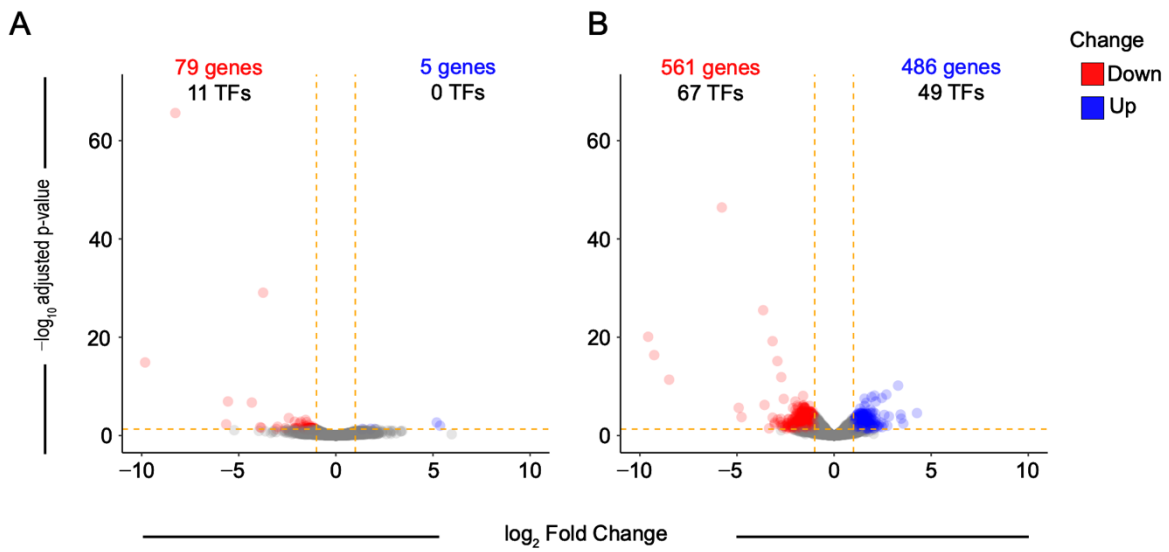

Figure 5: *Cal-opa* knockdown leads to the miss-regulation of a small set of transcription factor genes. Volcano plots of differentially expressed genes between the *Cal-opa* RNAi group and the alignment control group at A) NC12 and B) NC13. Significant change of adjusted p-value ≤ 0.05 and |$\log_2$ 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.

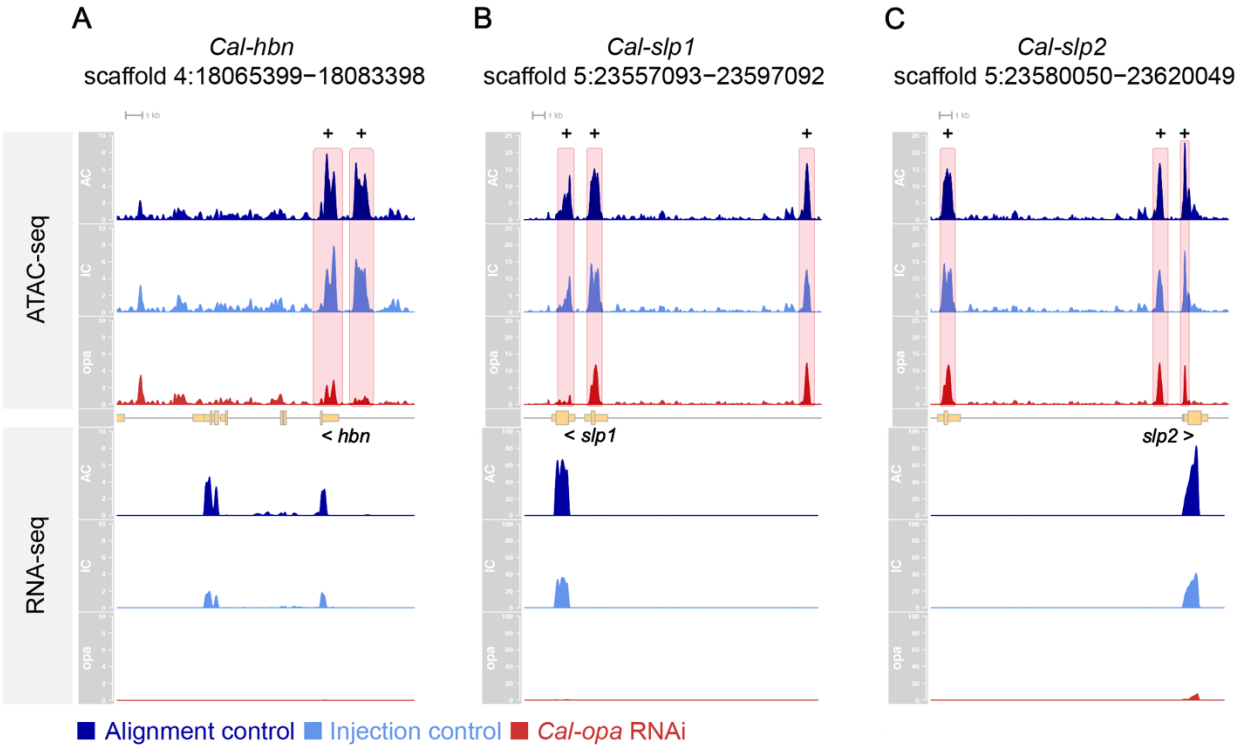

Figure 6: Key Cal-Opa targets at NC12. Gene loci of A) *Cal-hbn*, B) *Cal-slp1*, and C) *Cal-slp2* at NC12. Top tracks: CPM of ATAC-seq data (Alignment control in blue, Injection control in red, *Cal-opa* RNAi in yellow). Middle tracks show gene models in pale yellow. Bottom tracks: CPM of RNA-seq data.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

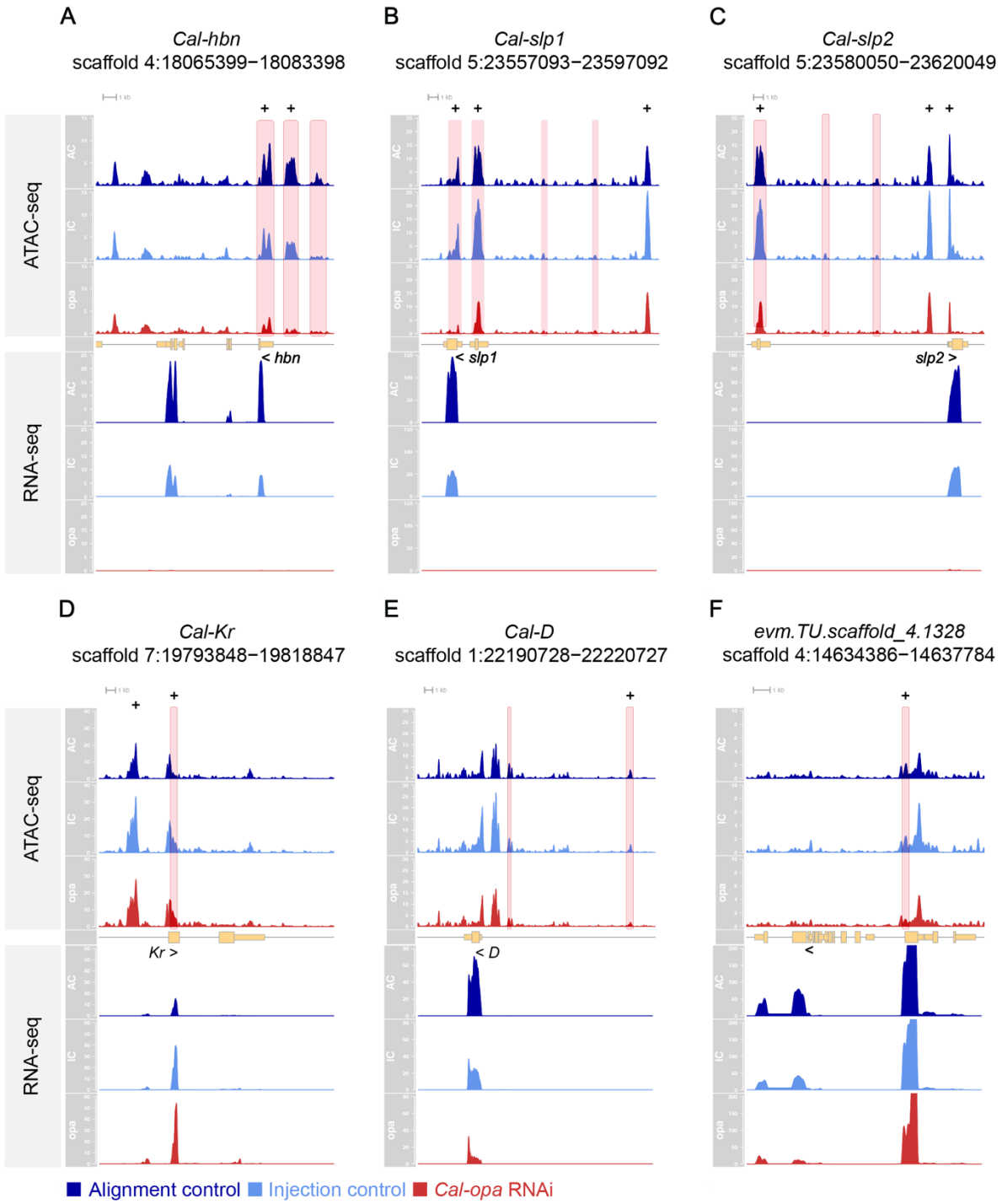

Figure 7: Key Cal-Opa targets at NC13. Gene loci of A) *Cal-hbn*, B) *Cal-slp1*, C) *Cal-slp2*, D) *Cal-Kr*, E) *Cal-D* and F) *evm.TU.scaffold_4.1328* at NC13. Top tracks: CPM of ATAC-seq data. Middle track gene model in pale yellow. Bottom tracks: CPM of RNA-seq data. Peaks with significant changes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

1001 *Kr* ATAC-seq: 0-40 CPM. *Cal-Kr* RNA-seq: 0-60 CPM. *Cal-D* ATAC-seq: 0-30 CPM. *Cal-D*
1002 RNA-seq: 0-80 CPM. *evm.TU.scaffold_4.1328* ATAC-seq: 0-10 CPM. *evm.TU.scaffold_4.1328*
1003 RNA-seq: 0-200 CPM. Blue: Alignment control. Light blue: Injection control. Red: *Cal-opa*
1004 RNAi.
1005

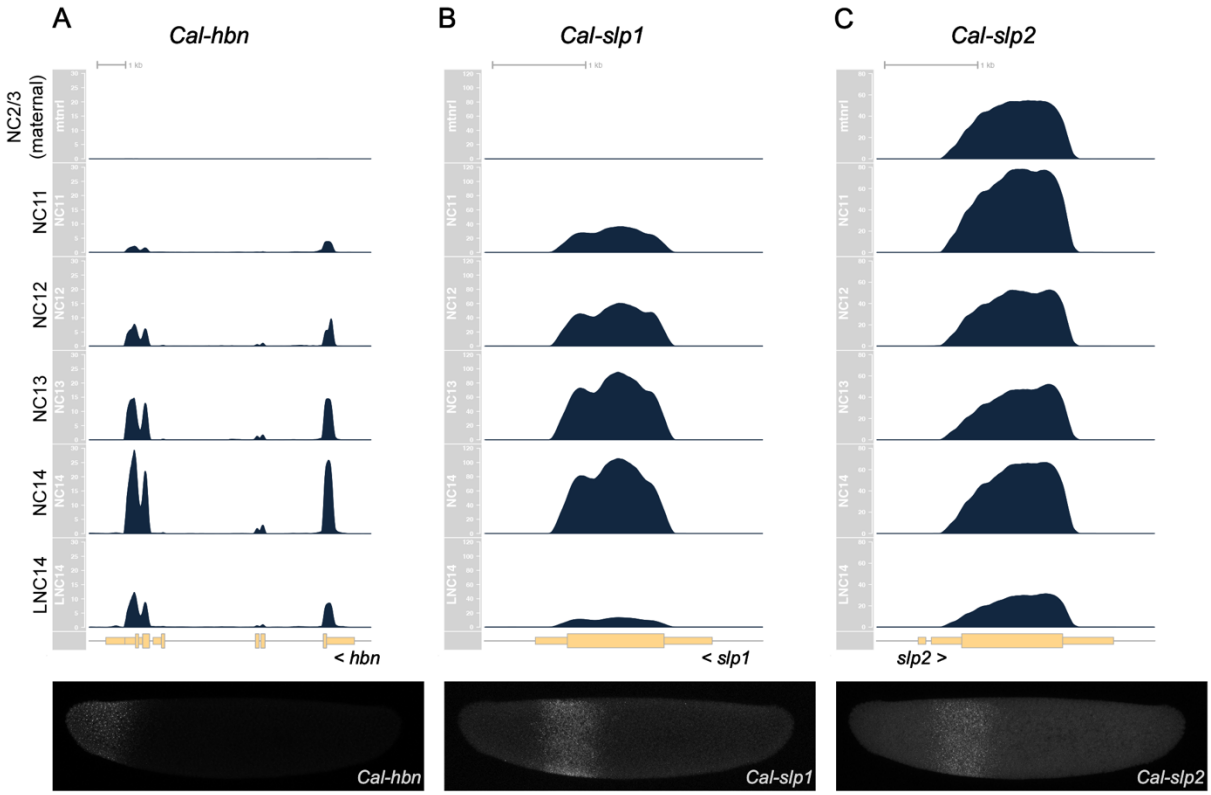

Figure 8: Time-course RNA-seq data for *Cal-hbn*, *Cal-slp1*, and *Cal-slp2*. Expression levels (CPM) of A) *Cal-hbn*, B) *Cal-slp1*, C) *Cal-slp2* at 6 consecutive stages, including NC2/3 (maternal, ~45-minute old embryos), NC11, NC12, NC13, NC14 and Late NC14 are shown above fluorescent *in situ* hybridization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

1015 **References**
1016

1017 1. Driever W, Nüsslein-Volhard C. The bicoid protein determines position in the Drosophila
1018 embryo in a concentration-dependent manner. Cell. 1988;54(1):95–104.

1019 2. Driever W, Nüsslein-Volhard C. A gradient of bicoid protein in Drosophila embryos.
1020 Cell. 1988;54(1):83–93.

1021 3. St Johnston D, Nüsslein-Volhard C. The origin of pattern and polarity in the Drosophila
1022 embryo. Cell. 1992;68(2):201–19.

1023 4. Surkova S, Golubkova E, Mamon L, Samsonova M. Dynamic maternal gradients and
1024 morphogenetic networks in Drosophila early embryo. Bio Systems. 2018;173:207–13.

1025 5. Ma J, He F, Xie G, Deng GX. Maternal AP determination in the Drosophila oocyte and
1026 embryo. Wiley Interdiscip Rev Dev Biol. 2016;5:562–81.

1027 6. Wieschaus E. Positional Information and Cell Fate Determination in the Early Drosophila
1028 Embryo. Current topics in developmental biology. 2016;117:567–79.

1029 7. Jaeger J. The gap gene network. Cellular and molecular life sciences : CMLS.
1030 2011;68(2):243–74.

1031 8. Tkačik G, Gregor T. The many bits of positional information. Development. 2021;148(2).

1032 9. Clark E, Peel AD, Akam M. Arthropod segmentation. Development. 2019;146(18).

1033 10. Crombach A, Jaeger J. Life's attractors continued: progress in understanding
1034 developmental systems through reverse engineering and in silico evolution. In: Crombach A,
1035 editor. Evolutionary Systems Biology: Springer Nature Switzerland AG; 2021.

1036 11. Schmidt-Ott U, Lynch JA. Emerging developmental genetic model systems in
1037 holometabolous insects. Curr Opin Genet Dev. 2016;39:116–28.

1038 12. Lemke S, Kale G, Urbansky S. Comparing gastrulation in flies: Links between cell
1039 biology and the evolution of embryonic morphogenesis. Mechanisms of Development. 2020;164.

1040 13. Anderson DT. Comparative Embryology of Diptera. Annu Rev Entomol. 1966;11:23–+.

1041 14. Klomp J, Athy D, Kwan CW, Bloch NI, Sandmann T, Lemke S, et al. Embryo
1042 development. A cysteine-clamp gene drives embryo polarity in the midge Chironomus. Science.
1043 2015;348(6238):1040–2.

1044 15. Yoon Y, Klomp J, Martin-Martin I, Criscione F, Calvo E, Ribeiro J, et al. Embryo
1045 polarity in moth flies and mosquitoes relies on distinct old genes with localized transcript
1046 isoforms. eLife. 2019;8.

1047 16. Schmidt-Ott U, Yoon Y. Evolution and loss of beta-catenin and TCF-dependent axis
1048 specification in insects. Curr Opin Insect Sci. 2022;50:100877.

1049 17. Berleth T, Burri M, Thoma G, Bopp D, Richstein S, Frigerio G, et al. The role of
1050 localization of bicoid RNA in organizing the anterior pattern of the Drosophila embryo. EMBO
1051 J. 1988;7(6):1749–56.

1052 18. Stauber M, Taubert H, Schmidt-Ott U. Function of bicoid and hunchback homologs in
1053 the basal cyclorrhaphan fly Megaselia (Phoridae). Proc Natl Acad Sci U S A.
1054 2000;97(20):10844–9.

1055 19. Shaw PJ, Salameh A, McGregor AP, Bala S, Dover GA. Divergent structure and function
1056 of the bicoid gene in Muscoidea fly species. Evolution & development. 2001;3(4):251–62.

1057 20. Lemke S, Busch SE, Antonopoulos DA, Meyer F, Domanus MH, Schmidt-Ott U.
1058 Maternal activation of gap genes in the hover fly Episyrphus. Development. 2010;137(10):1709–
1059 19.

1060 21. Wotton KR, Jimenez-Guri E, Jaeger J. Maternal co-ordinate gene regulation and axis
1061 polarity in the scuttle fly Megaselia abdita. PLoS Genet. 2015;11(3):e1005042.

1062 22. Hursh DA, Stultz BG. Odd-Paired: The Drosophila Zic Gene. Advances in experimental
1063 medicine and biology. 2018;1046:41–58.

1064 23. Ravindranath AJ, Cadigan KM. Structure-function analysis of the C-clamp of
1065 TCF/Pangolin in Wnt/ss-catenin signaling. PloS one. 2014;9(1):e86180.

1066 24. Liu Q, Onal P, Datta RR, Rogers JM, Schmidt-Ott U, Bulyk ML, et al. Ancient
1067 mechanisms for the evolution of the bicoid homeodomain's function in fly development. eLife.
1068 2018;7.

1069 25. Nitta KR, Jolma A, Yin Y, Morgunova E, Kivioja T, Akhtar J, et al. Conservation of
1070 transcription factor binding specificities across 600 million years of bilateria evolution. eLife.
1071 2015;4.

1072 26. Duan B, Fu D, Zhang C, Ding P, Dong X, Xia B. Selective Nonmethylated CpG DNA
1073 Recognition Mechanism of Cysteine Clamp Domains. J Am Chem Soc. 2021;143(20):7688–97.

1074 27. Li M, Kasan K, Saha Z, Yoon Y, Schmidt-Ott U. Twenty-seven ZAD-ZNF genes of
1075 Drosophila melanogaster are orthologous to the embryo polarity determining mosquito gene
1076 cucoid. PloS one. 2023;18(1):e0274716.

1077 28. Driever W, Nüsslein-Volhard C. The bicoid protein is a positive regulator of hunchback
1078 transcription in the early Drosophila embryo. Nature. 1989;337(6203):138–43.

1079 29. Struhl G, Struhl K, Macdonald PM. The gradient morphogen bicoid is a concentration-
1080 dependent transcriptional activator. Cell. 1989;57(7):1259–73.

1081 30. Hoch M, Seifert E, Jäckle H. Gene expression mediated by cis-acting sequences of the
1082 Krüppel gene in response to the Drosophila morphogens bicoid and hunchback. Embo j.
1083 1991;10(8):2267–78.

1084 31. Rivera-Pomar R, Lu X, Perrimon N, Taubert H, Jäckle H. Activation of posterior gap
1085 gene expression in the *Drosophila* blastoderm. Nature. 1995;376:253–6.

1086 32. Berman BP, Nibu Y, Pfeiffer BD, Tomancak P, Celniker SE, Levine M, et al. Exploiting
1087 transcription factor binding site clustering to identify cis-regulatory modules involved in pattern
1088 formation in the Drosophila genome. Proc Natl Acad Sci U S A. 2002;99(2):757–62.

1089 33. Schroeder MD, Pearce M, Fak J, Fan H-Q, Unnerstall U, Emberly E, et al.
1090 Transcriptional control in the segmentation gene network of *Drosophila*. PLOS. 2004;2:e271.

1091 34. Ochoa-Espinosa A, Yucel G, Kaplan L, Pare A, Pura N, Oberstein A, et al. The role of
1092 binding site cluster strength in Bicoid-dependent patterning in Drosophila. Proc Natl Acad Sci U
1093 S A. 2005;102(14):4960–5.

1094 35. Gao Q, Finkelstein R. Targeting gene expression to the head: the Drosophila
1095 orthodenticle gene is a direct target of the Bicoid morphogen. Development. 1998;125(21):4185–
1096 93.

1097 36. Stanojevic D, Small S, Levine M. Regulation of a segmentation stripe by overlapping
1098 activators and repressors in the Drosophila embryo. Science. 1991;254(5036):1385–7.

1099 37. Arnosti DN, Barolo S, Levine M, Small S. The eve stripe 2 enhancer employs multiple
1100 modes of transcriptional synergy. Development. 1996;122(1):205–14.

1101 38. Small S, Blair A, Levine M. Regulation of even-skipped stripe 2 in the Drosophila
1102 embryo. EMBO J. 1992;11(11):4047–57.

1103 39. Small S, Kraut R, Hoey T, Warrior R, Levine M. Transcriptional regulation of a pair-rule
1104 stripe in Drosophila. Genes Dev. 1991;5(5):827–39.

1105 40. Simpson-Brose M, Treisman J, Desplan C. Synergy between the hunchback and bicoid
1106 morphogens is required for anterior patterning in Drosophila. Cell. 1994;78(5):855–65.

1107 41. Wimmer EA, Carleton A, Harjes P, Turner T, Desplan C. *bicoid*-independent formation
1108 of thoracic segments in *Drosophila*. Science. 2000;287:2476–9.

1109 42. Manu, Surkova S, Spirov AV, Gursky VV, Janssens H, Kim AR, et al. Canalization of
1110 gene expression and domain shifts in the Drosophila blastoderm by dynamical attractors. PLoS
1111 Comput Biol. 2009;5(3):e1000303.

1112 43. Rivera-Pomar R, Jackle H. From gradients to stripes in Drosophila embryogenesis: filling
1113 in the gaps. Trends Genet. 1996;12(11):478–83.

1114 44. Macdonald PM. Translational repression by Bicoid: competition for the cap. Cell.
1115 2005;121(3):321–2.

1116 45. Staudt N, Fellert S, Chung HR, Jackle H, Vorbruggen G. Mutations of the Drosophila
1117 zinc finger-encoding gene vielfaltig impair mitotic cell divisions and cause improper
1118 chromosome segregation. Mol Biol Cell. 2006;17(5):2356–65.

1119 46. Liang HL, Nien CY, Liu HY, Metzstein MM, Kirov N, Rushlow C. The zinc-finger
1120 protein Zelda is a key activator of the early zygotic genome in Drosophila. Nature.
1121 2008;456(7220):400–3.

1122 47. Foo SM, Sun Y, Lim B, Ziukaite R, O'Brien K, Nien CY, et al. Zelda potentiates
1123 morphogen activity by increasing chromatin accessibility. Curr Biol. 2014;24(12):1341–6.

1124 48. Xu Z, Chen H, Ling J, Yu D, Struffi P, Small S. Impacts of the ubiquitous factor Zelda on
1125 Bicoid-dependent DNA binding and transcription in Drosophila. Genes Dev. 2014;28(6):608–21.

1126 49. Mir M, Stadler MR, Ortiz SA, Hannon CE, Harrison MM, Darzacq X, et al. Dynamic
1127 multifactor hubs interact transiently with sites of active transcription in Drosophila embryos.
1128 eLife. 2018;7.

1129 50. Bishop TR, Onal P, Xu Z, Zheng M, Gunasinghe H, Nien CY, et al. Multi-level
1130 regulation of even-skipped stripes by the ubiquitous factor Zelda. Development. 2023;150(23).

1131 51. Crocker J, Stern DL. Functional regulatory evolution outside of the minimal even-
1132 skipped stripe 2 enhancer. Development. 2017;144(17):3095–101.

1133 52. Ling J, Umezawa KY, Scott T, Small S. Bicoid-Dependent Activation of the Target Gene
1134 hunchback Requires a Two-Motif Sequence Code in a Specific Basal Promoter. Mol Cell.
1135 2019;75(6):1178–87 e4.

1136 53. Eck E, Liu J, Kazemzadeh-Atoufi M, Ghoreishi S, Blythe SA, Garcia HG. Quantitative
1137 dissection of transcription in development yields evidence for transcription-factor-driven
1138 chromatin accessibility. eLife. 2020;9.

1139 54. Fernandes G, Tran H, Andrieu M, Diaw Y, Perez Romero C, Fradin C, et al. Synthetic
1140 reconstruction of the hunchback promoter specifies the role of Bicoid, Zelda and Hunchback in
1141 the dynamics of its transcription. eLife. 2022;11.

1142 55. Harden TT, Vincent BJ, DePace AH. Transcriptional activators in the early Drosophila
1143 embryo perform different kinetic roles. Cell Syst. 2023;14(4):258–72 e4.

1144 56. Hannon CE, Blythe SA, Wieschaus EF. Concentration dependent chromatin states
1145 induced by the bicoid morphogen gradient. eLife. 2017;6.

1146 57. Haines JE, Eisen MB. Patterns of chromatin accessibility along the anterior-posterior axis
1147 in the early Drosophila embryo. PLoS Genet. 2018;14(5):e1007367.

1148 58. Soluri IV, Zumerling LM, Payan Parra OA, Clark EG, Blythe SA. Zygotic pioneer factor
1149 activity of Odd-paired/Zic is necessary for late function of the Drosophila segmentation network.
1150 eLife. 2020;9.

1151 59. Cusanovich DA, Reddington JP, Garfield DA, Daza RM, Aghamirzaie D, Marco-
1152 Ferreres R, et al. The cis-regulatory dynamics of embryonic development at single-cell
1153 resolution. Nature. 2018;555(7697):538–42.

1154 60. Calderon D, Blecher-Gonen R, Huang X, Secchia S, Kentro J, Daza RM, et al. The
1155 continuum of Drosophila embryonic development at single-cell resolution. Science.
1156 2022;377(6606):eabn5800.

1157 61. Degen EA, Croslyn C, Mangan NM, Blythe SA. Bicoid-nucleosome competition sets a
1158 concentration threhold for transcription constrained by genome replication. bioRxiv. 2024.

1159 62. Blythe SA, Wieschaus EF. Establishment and maintenance of heritable chromatin
1160 structure during early Drosophila embryogenesis. eLife. 2016;5.

1161 63. Clark E, Akam M. Odd-paired controls frequency doubling in Drosophila segmentation
1162 by altering the pair-rule gene regulatory network. eLife. 2016;5.

1163 64. Hülskamp M, Pfeifle C, Tautz D. A morphogenetic gradient of Hunchback protein
1164 organizes the expression of the gap genes *Krüppel* and *knirps* in the early *Drosophila* embryo.
1165 Nature. 1990;346:577–80.

1166 65. Tautz D. Regulation of the *Drosophila* segmentation gene *hunchback* by two maternal
1167 morphogenetic centers. Nature (London). 1988;332:281–4.

1168 66. Janssens H, Siggens K, Cicin-Sain D, Jimenez-Guri E, Musy M, Akam M, et al. A
1169 quantitative atlas of Even-skipped and Hunchback expression in Clogmia albipunctata (Diptera:
1170 Psychodidae) blastoderm embryos. EvoDevo. 2014;5(1):1.

1171 67. Garcia-Solache M, Jaeger J, Akam M. A systematic analysis of the gap gene system in
1172 the moth midge *Clogmia albipunctata*. Dev Biol. 2010;344:306–18.

1173 68. Koromila T, Gao F, Iwasaki Y, He P, Pachter L, Gergen JP, et al. Odd-paired is a
1174 pioneer-like factor that coordinates with Zelda to control gene expression in embryos. eLife.
1175 2020;9.

1176 69. Jimenez-Guri E, Wotton KR, Gavilan B, Jaeger J. A staging scheme for the development
1177 of the moth midge Clogmia albipunctata. PloS one. 2014;9(1):e84422.

1178 70. Amabis JM, Simoes LCG. Chromosome studies in a species of *Telmatoscopus* sp.
1179 Caryologia. 1972;25(2).

1180 71. Amabis JM. Cytological evidence of male sex heterogamety in sex determination of
1181 *Telmatoscopus albipunctatus* (Diptera: Psychodidae). Chromosoma. 1977;62:133–8.

1182 72. Schmidt-Ott U, Rafiqi AM, Sander K, Johnston JS. Extremely small genomes in two
1183 unrelated dipteran insects with shared early developmental traits. Dev Genes Evol.
1184 2009;219(4):207–10.

1185 73. Generalovic TN, McCarthy SA, Warren IA, Wood JMD, Torrance J, Sims Y, et al. A
1186 high-quality, chromosome-level genome assembly of the Black Soldier Fly (Hermetia illucens
1187 L.). G3 (Bethesda). 2021;11(5).

1188 74. Tenger-Trolander A, Amiri E, Gantz V, Kwan CW, Sanders SA, Schmidt-Ott U.
1189 Genomic resources for the scuttle fly Megaselia abdita: A Model Organism for Comparative
1190 Developmental Studies in Flies bioRxiv. 2025.

1191 75. Ji J, Gao Y, Xu C, Zhang K, Li D, Li B, et al. Chromosome-level genome assembly of
1192 marmalade hoverfly Episyrphus balteatus (Diptera: Syrphidae). Sci Data. 2024;11(1):844.

1193 76. Stauber M, Jäckle H, Schmidt-Ott U. The anterior determinant *bicoid* of *Drosophila* is a
1194 derived Hox class 3 gene. Proc Natl Acad Sci USA. 1999;96:3786–9.

1195 77. Stauber M, Prell A, Schmidt-Ott U. A single *Hox3* gene with composite *bicoid* and
1196 *zerknüllt* expression characteristics in non-Cyclorrhaphan flies. Proc Natl Acad Sci U S A.
1197 2002;99:274–9.

1198 78. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based
1199 analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9):R137.

1200 79. Reddington JP, Garfield DA, Sigalova OM, Karabacak Calviello A, Marco-Ferreres R,
1201 Girardot C, et al. Lineage-Resolved Enhancer and Promoter Usage during a Time Course of
1202 Embryogenesis. Dev Cell. 2020;55(5):648–64 e9.

1203 80. Schulz KN, Harrison MM. Mechanisms regulating zygotic genome activation. Nat Rev
1204 Genet. 2019;20(4):221–34.

1205 81. Hamm DC, Harrison MM. Regulatory principles governing the maternal-to-zygotic
1206 transition: insights from Drosophila melanogaster. Open Biol. 2018;8(12):180183.

1207 82. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for
1208 RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.

1209 83. Larson ED, Marsh AJ, Harrison MM. Pioneering the developmental frontier. Mol Cell.
1210 2021;81(8):1640–50.

1211 84. Brennan KJ, Weilert M, Krueger S, Pampari A, Liu HY, Yang AWH, et al. Chromatin
1212 accessibility in the Drosophila embryo is determined by transcription factor pioneering and
1213 enhancer activation. Dev Cell. 2023;58(19):1898–916 e9.

1214 85. Harrison MM, Li XY, Kaplan T, Botchan MR, Eisen MB. Zelda binding in the early
1215 Drosophila melanogaster embryo marks regions subsequently activated at the maternal-to-
1216 zygotic transition. PLoS Genet. 2011;7(10):e1002266.

1217 86. McDaniel SL, Gibson TJ, Schulz KN, Fernandez Garcia M, Nevil M, Jain SU, et al.
1218 Continued Activity of the Pioneer Factor Zelda Is Required to Drive Zygotic Genome
1219 Activation. Mol Cell. 2019;74(1):185–95 e4.

1220 87. Sun Y, Nien CY, Chen K, Liu HY, Johnston J, Zeitlinger J, et al. Zelda overcomes the
1221 high intrinsic nucleosome barrier at enhancers during Drosophila zygotic genome activation.
1222 Genome research. 2015;25(11):1703–14.

1223 88. Gaskill MM, Gibson TJ, Larson ED, Harrison MM. GAF is essential for zygotic genome
1224 activation and chromatin accessibility in the early Drosophila embryo. eLife. 2021;10.

1225 89. Duan J, Rieder L, Colonnetta MM, Huang A, McKenney M, Watters S, et al. CLAMP
1226 and Zelda function together to promote Drosophila zygotic genome activation. eLife. 2021;10.

1227 90. Colonnetta MM, Abrahante JE, Schedl P, Gohl DM, Deshpande G. CLAMP regulates
1228 zygotic genome activation in Drosophila embryos. Genetics. 2021;219(2).

1229 91. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME Suite. Nucleic Acids Res.
1230 2015;43(W1):W39–49.

1231 92. Bailey TL. STREME: accurate and versatile sequence motif discovery. Bioinformatics.
1232 2021;37(18):2834–40.

1233 93. Soruco MM, Chery J, Bishop EP, Siggers T, Tolstorukov MY, Leydon AR, et al. The
1234 CLAMP protein links the MSL complex to the X chromosome during Drosophila dosage
1235 compensation. Genes & development. 2013;27(14):1551–6.

1236 94. Kuzu G, Kaye EG, Chery J, Siggers T, Yang L, Dobson JR, et al. Expansion of GA
1237 Dinucleotide Repeats Increases the Density of CLAMP Binding Sites on the X-Chromosome to
1238 Promote Drosophila Dosage Compensation. PLoS genetics. 2016;12(7):e1006120.

1239 95. Kaye EG, Booker M, Kurland JV, Conicella AE, Fawzi NL, Bulyk ML, et al. Differential
1240 Occupancy of Two GA-Binding Proteins Promotes Targeting of the Drosophila Dosage
1241 Compensation Complex to the Male X Chromosome. Cell Rep. 2018;22(12):3227–39.

1242 96. Ray P, De S, Mitra A, Bezstarosti K, Demmers JA, Pfeifer K, et al. Combgap contributes
1243 to recruitment of Polycomb group proteins in Drosophila. Proceedings of the National Academy
1244 of Sciences of the United States of America. 2016;113(14):3826–31.

1245 97. Ribeiro L, Tobias-Santos V, Santos D, Antunes F, Feltran G, de Souza Menezes J, et al.
1246 Evolution and multiple roles of the Pancrustacea specific transcription factor zelda in insects.
1247 PLoS genetics. 2017;13(7):e1006868.

1248 98. Biedler JK, Hu W, Tae H, Tu Z. Identification of early zygotic genes in the yellow fever
1249 mosquito Aedes aegypti and discovery of a motif involved in early zygotic genome activation.
1250 PloS one. 2012;7(3):e33933.

1251 99. Arsala D, Lynch JA. Ploidy has little effect on timing early embryonic events in the
1252 haplo-diploid wasp Nasonia. Genesis. 2017;55(5).

1253 100. Pires CV, Freitas FC, Cristino AS, Dearden PK, Simoes ZL. Transcriptome Analysis of
1254 Honeybee (Apis Mellifera) Haploid and Diploid Embryos Reveals Early Zygotic Transcription
1255 during Cleavage. PloS one. 2016;11(1):e0146447.

1256 101. Gao S, Xue S, Gao T, Lu R, Zhang X, Zhang Y, et al. Transcriptome analysis reveals the
1257 role of Zelda in the regulation of embryonic and wing development of Tribolium castaneum.
1258 Bull Entomol Res. 2023;113(5):587–97.

1259 102. Ventos-Alfonso A, Ylla G, Belles X. Zelda and the maternal-to-zygotic transition in
1260 cockroaches. FEBS J. 2019;286(16):3206–21.

1261 103. Gibson TJ, Larson ED, Harrison MM. Protein-intrinsic properties and context-dependent
1262 effects regulate pioneer factor binding and function. Nat Struct Mol Biol. 2024;31(3):548–58.

1263 104. Nüsslein-Volhard C, Wieschaus E. Mutations affecting segment number and polarity in
1264 *Drosophila*. Nature. 1980;287:795–801.

1265 105. Goltsev Y, Hsiong W, Lanzaro G, Levine M. Different combinations of gap repressors
1266 for common stripes in *Anopheles* and *Drosophila* embryos. Dev Biol. 2004;275:435–46.

1267 106. Lemke S, Busch SE, Antonopoulos DA, Meyer F, Domanus MH, Schmidt-Ott U.
1268 Maternal activation of gap genes in the hover fly *Episyrphus*. Development. 2010;137:1709–19.

1269 107. Wotton KR, Jimenez-Guri E, Crombach A, Janssens H, Alcaine-Colet A, Lemke S, et al.
1270 Quantitative system drift compensates for altered maternal inputs to the gap gene network of the
1271 scuttle fly Megaselia abdita. eLife. 2015;4.

1272 108. Bullock SL, Stauber M, Prell A, Hughes JR, Ish-Horowicz D, Schmidt-Ott U.
1273 Differential cytoplasmic mRNA localization adjusts pair-rule transcription factor activity to
1274 cytoarchitecture in dipteran evolution. Development. 2004;131:4251–61.

1275 109. Campos-Ortega JA, Hartenstein V. The embryonic development of *Drosophila*
1276 *melanogaster*. 2 ed. Berlin, Heidelberg, New York: Springer-Verlag; 1997.

1277 110. Yoon Y. Evolution of embryonic axis determinants via alternative transcription:
1278 University of Chicago; 2019.

1279 111. Rafiqi AM, Park CH, Kwan CW, Lemke S, Schmidt-Ott U. BMP-dependent serosa and
1280 amnion specification in the scuttle fly Megaselia abdita. Development. 2012;139(18):3373–82.

1281 112. Rafiqi AM, Lemke S, Ferguson S, Stauber M, Schmidt-Ott U. Evolutionary origin of the
1282 amnioserosa in cyclorrhaphan flies correlates with spatial and temporal expression changes of
1283 zen. Proc Natl Acad Sci U S A. 2008;105(1):234–9.

1284 113. Kwan CW, Gavin-Smyth J, Ferguson EL, Schmidt-Ott U. Functional evolution of a
1285 morphogenetic gradient. eLife. 2016;5.

1286 114. Caroti F, Gonzalez Avalos E, Noeske V, Gonzalez Avalos P, Kromm D, Wosch M, et al.
1287 Decoupling from yolk sac is required for extraembryonic tissue spreading in the scuttle fly
1288 Megaselia abdita. eLife. 2018;7.

1289 115. Urbansky S, Gonzalez Avalos P, Wosch M, Lemke S. Folded gastrulation and T48 drive
1290 the evolution of coordinated mesoderm internalization in flies. eLife. 2016;5.

1291 116. Vicoso B, Bachtrog D. Numerous transitions of sex chromosomes in Diptera. PLoS Biol.
1292 2015;13(4):e1002078.

1293 117. Hall AB, Basu S, Jiang X, Qi Y, Timoshevskiy VA, Biedler JK, et al. SEX
1294 DETERMINATION. A male-determining factor in the mosquito Aedes aegypti. Science.
1295 2015;348(6240):1268–70.

1296 118. Krzywinska E, Dennison NJ, Lycett GJ, Krzywinski J. A maleness gene in the malaria
1297 mosquito Anopheles gambiae. Science. 2016;353(6294):67–9.

1298 119. Saccone G. A history of the genetic and molecular identification of genes and their
1299 functions controlling insect sex determination. Insect Biochem Mol Biol. 2022;151:103873.

1300 120. Grand RS, Pregnolato M, Baumgartner L, Hoerner L, Burger L, Schubeler D. Genome
1301 access is transcription factor-specific and defined by nucleosome position. Mol Cell.
1302 2024;84(18):3455–68 e6.

1303 121. Schulz KN, Bondra ER, Moshe A, Villalta JE, Lieb JD, Kaplan T, et al. Zelda is
1304 differentially required for chromatin accessibility, transcription factor binding, and gene
1305 expression in the early Drosophila embryo. Genome research. 2015;25(11):1715–26.

1306 122. Seller CA, O'Farrell PH. Rif1 prolongs the embryonic S phase at the Drosophila mid-
1307 blastula transition. PLoS Biol. 2018;16(5):e2005687.

1308 123. Walldorf U, Kiewe A, Wickert M, Ronshaugen M, McGinnis W. Homeobrain, a novel
1309 paired-like homeobox gene is expressed in the Drosophila brain. Mech Dev. 2000;96(1):141–4.

1310 124. Kolb D, Kaspar P, Kloppel C, Walldorf U. The Drosophila homeodomain transcription
1311 factor Homeobrain is involved in the formation of the embryonic protocerebrum and the
1312 supraesophageal brain commissure. Cells Dev. 2021;165:203657.

1313 125. Hildebrandt K, Kolb D, Kloppel C, Kaspar P, Wittling F, Hartwig O, et al. Regulatory
1314 modules mediating the complex neural expression patterns of the homeobrain gene during
1315 Drosophila brain development. Hereditas. 2022;159(1):2.

1316 126. Ansari S, Troelenberg N, Dao VA, Richter T, Bucher G, Klingler M. Double abdomen in
1317 a short-germ insect: Zygotic control of axis formation revealed in the beetle Tribolium
1318 castaneum. Proc Natl Acad Sci U S A. 2018;115(8):1819–24.

1319 127. Schomburg C, Janssen R, Prpic NM. Phylogenetic analysis of forkhead transcription
1320 factors in the Panarthropoda. Dev Genes Evol. 2022;232(1):39–48.

1321 128. Clark E. Dynamic patterning by the Drosophila pair-rule network reconciles long-germ
1322 and short-germ segmentation. PLoS Biol. 2017;15(9):e2002439.

1323 129. Cadigan KM, Grossniklaus U, Gehring WJ. Functional redundancy: the respective roles
1324 of the two sloppy paired genes in Drosophila segmentation. Proc Natl Acad Sci U S A.
1325 1994;91(14):6324–8.

1326 130. Grossniklaus U, Pearson RK, Gehring WJ. The Drosophila sloppy paired locus encodes
1327 two proteins involved in segmentation that show homology to mammalian transcription factors.
1328 Genes Dev. 1992;6(6):1030–51.

1329 131. Fujioka M, Jaynes JB. Regulation of a duplicated locus: Drosophila sloppy paired is
1330 replete with functionally overlapping enhancers. Dev Biol. 2012;362(2):309–19.

1331 132. Lee HH, Frasch M. Survey of forkhead domain encoding genes in the Drosophila
1332 genome: Classification and embryonic expression patterns. Dev Dyn. 2004;229(2):357–66.

1333 133. Gao Q, Finkelstein R. Targeting gene expression to the head: the *Drosophila*
1334 *orthodenticle* gene is a direct target of the Bicoid morphogen. Development. 1998:4185–93.

1335 134. Choi HMT, Schwarzkopf M, Fornace ME, Acharya A, Artavanis G, Stegmaier J, et al.
1336 Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive,
1337 versatile, robust. Development. 2018;145(12).

1338 135. Bruce HS, Jerz G, Kelly SR, McCarthy J, Pomerantz A, Senevirathne G, et al.
1339 Hybridization chain reaction (HCR) in situ protocol V.1.; 2021.

1340 136. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native
1341 chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins
1342 and nucleosome position. Nat Methods. 2013;10(12):1213–8.

1343 137. Krueger F. Trim Galore! : Babraham Bioinformatics; 2012.

1344 138. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods.
1345 2012;9(4):357–9.

1346 139. Lawrence M, Huber W, Pages H, Aboyoun P, Carlson M, Gentleman R, et al. Software
1347 for computing and annotating genomic ranges. PLoS Comput Biol. 2013;9(8):e1003118.

1348 140. Morgan M, Pages H, Obenchain V, Hayden N. Rsamtools: Binary alignment (BAM),
1349 FASTA, variant call (BCF), and tabix file import.: Bioconductor; 2013.

1350 141. Pantano L. DEGreport: report of DEG analysis 2017 [Available from:
1351 http://lpantano.github.io/DEGreport/.

1352 142. Bergman CM, Carlson JW, Celniker SE. Drosophila DNase I footprint database: a
1353 systematic genome annotation of transcription factor binding sites in the fruitfly, Drosophila
1354 melanogaster. Bioinformatics. 2005;21(8):1747–9.

1355 143. Kulakovskii IV, Makeev V. [Integration of data obtained by different experimental
1356 methods to determine the motifs in DNA sequences recognized by transcription-regulating
1357 factors]. Biofizika. 2009;54(6):965–74.

1358 144. Kulakovskiy IV, Favorov AV, Makeev VJ. Motif discovery and motif finding from
1359 genome-mapped DNase footprint data. Bioinformatics. 2009;25(18):2318–25.

1360 145. Zhu LJ, Christensen RG, Kazemian M, Hull CJ, Enuameh MS, Basciotta MD, et al.
1361 FlyFactorSurvey: a database of Drosophila transcription factor binding specificities determined
1362 using the bacterial one-hybrid system. Nucleic Acids Res. 2011;39(Database issue):D111–7.

1363 146. Shazman S, Lee H, Socol Y, Mann RS, Honig B. OnTheFly: a database of Drosophila
1364 melanogaster transcription factors and their binding sites. Nucleic Acids Res. 2014;42(Database
1365 issue):D167–71.

1366 147. Rohr KB, Tautz D, Sander K. Segmentation gene expression in the mothmidge Clogmia
1367 albipunctata (Diptera, psychodidae) and other primitive dipterans. Dev Genes Evol.
1368 1999;209(3):145–54.

1369 148. Smit A, Hubley R. RepeatModeler 2008 [Available from: https://www.repeatmasker.org/.

1370 149. Smit A, Hubley R, Green P. RepeatMasker 2013 [Available from:
1371 https://www.repeatmasker.org/.

1372 150. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, et al. Automated eukaryotic
1373 gene structure annotation using EVidenceModeler and the Program to Assemble Spliced
1374 Alignments. Genome Biol. 2008;9(1):R7.

1375 151. Huerta-Cepas J, Szklarczyk D, Heller D, Hernandez-Plaza A, Forslund SK, Cook H, et al.
1376 eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource
1377 based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 2019;47(D1):D309–D14.

1378 152. Cantalapiedra CP, Hernandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-
1379 mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the
1380 Metagenomic Scale. Mol Biol Evol. 2021;38(12):5825–9.

1381 153. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO:
1382 assessing genome assembly and annotation completeness with single-copy orthologs.
1383 Bioinformatics. 2015;31(19):3210–2.

1384 154. Manni M, Berkeley MR, Seppey M, Zdobnov EM. BUSCO: Assessing Genomic Data
1385 Quality and Beyond. Curr Protoc. 2021;1(12):e323.

1386 155. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, et al. MCScanX: a toolkit for
1387 detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res.
1388 2012;40(7):e49.

1389 156. Gu Z, Gu L, Eils R, Schlesner M, Brors B. circlize Implements and enhances circular
1390 visualization in R. Bioinformatics. 2014;30(19):2811–2.

1391 157. Diesh C, Stevens GJ, Xie P, De Jesus Martinez T, Hershberg EA, Leung A, et al.
1392 JBrowse 2: a modular genome browser with views of synteny and structural variation. Genome
1393 Biol. 2023;24(1):74.

1394 158. Hancock DY, Fischer J, Lowe JM, Fischer J, Lowe JM, Snapp-Childs W, et al.
1395 Jetstream2: Accelerating cloud computing via Jetstream. . Practice and Experience in Advanced
1396 Research Computing (PEARC '21). New York, NY: Association for Computing Machinery;
1397 2021.

1398 159. Boerner TJ, Deems S, Furlani TR, Knuth SL, Towns J. ACCESS: Advancing Innovation:
1399 NSF's Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support. Practice
1400 and Experience in Advanced Research Computing (PEARC '23). New York, NY: Association
1401 for Computing Machinery; 2023.

1402 160. Priyam A, Woodcroft BJ, Rai V, Moghul I, Munagala A, Ter F, et al. Sequenceserver: A
1403 Modern Graphical User Interface for Custom BLAST Databases. Mol Biol Evol.
1404 2019;36(12):2922–4.

1405 161. Beeber D, Chain FJ. crispRdesignR: A Versatile Guide RNA Design Package in R for
1406 CRISPR/Cas9 Applications. J Genomics. 2020;8:62–70.

1407 162. Brooks EM, Sanders SA, Pfrender ME. FreeCount: A coding-free framework for guided
1408 count data visualization and analysis. Practice and Experience in Advanced Research
1409 Computing 2024: Human Powered Computing (PEARC '24). New York, NY: Association for
1410 Computing Machinery; 2024.

1411 163. Xiao Z, Lam HM. ShinySyn: a Shiny/R application for the interactive visualization and
1412 integration of macro- and micro-synteny data. Bioinformatics. 2022;38(18):4406–8.

1413