

# Embryo polarity in moth flies and mosquitoes relies on distinct old genes with localized transcript isoforms

Yoseop Yoon<sup>1</sup>, Jeff Klomp<sup>1†</sup>, Ines Martin-Martin<sup>2</sup>, Frank Criscione<sup>2</sup>, Eric Calvo<sup>2</sup>, Jose Ribeiro<sup>2</sup>, Urs Schmidt-Ott<sup>1\*</sup>

<sup>1</sup>Department of Organismal Biology and Anatomy, University of Chicago, Chicago, United States; <sup>2</sup>Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, Rockville, United States

**Abstract** Unrelated genes establish head-to-tail polarity in embryos of different fly species, raising the question of how they evolve this function. We show that in moth flies (*Clogmia*, *Lutzomyia*), a maternal transcript isoform of *odd-paired* (*Zic*) is localized in the anterior egg and adopted the role of anterior determinant without essential protein change. Additionally, *Clogmia* lost maternal germ plasm, which contributes to embryo polarity in fruit flies (*Drosophila*). In culicine (*Culex*, *Aedes*) and anopheline mosquitoes (*Anopheles*), embryo polarity rests on a previously unnamed zinc finger gene (*cucoid*), or *pangolin* (*dTcf*), respectively. These genes also localize an alternative transcript isoform at the anterior egg pole. Basal-branching crane flies (*Nephrotoma*) also enrich maternal *pangolin* transcript at the anterior egg pole, suggesting that *pangolin* functioned as ancestral axis determinant in flies. In conclusion, flies evolved an unexpected diversity of anterior determinants, and alternative transcript isoforms with distinct expression can adopt fundamentally distinct developmental roles.

DOI: <https://doi.org/10.7554/eLife.46711.001>

\*For correspondence:  
uschmid@uchicago.edu

Present address: <sup>†</sup>University of North Carolina, Lineberger Comprehensive Cancer Center, Chapel Hill, United States

Competing interests: The authors declare that no competing interests exist.

Funding: See page 25

Received: 09 March 2019

Accepted: 06 September 2019

Published: 08 October 2019

Reviewing editor: Patricia J Wittkopp, University of Michigan, United States

© Copyright Yoon et al. This article is distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

## Introduction

The specification of the primary axis (head-to-tail) in embryos of flies (Diptera) offers important advantages for studying how new essential gene functions evolve in early development. This process rests on lineage-specific maternal mRNAs that are localized at the anterior egg pole ('anterior determinants'), which, surprisingly, have changed during the evolution of flies. While the anterior determinants of most flies remain unknown, they can be identified by comparing the transcriptomes of anterior and posterior egg halves (Klomp et al., 2015). Furthermore, their function can be analyzed in the syncytial early embryos of a broad range of species via microinjection, considering timing and subcellular localization. It is therefore possible to conduct phylogenetic comparisons at the functional level. Finally, when the function of anterior determinants is suppressed, embryos develop into an unambiguous, predictable phenotype: these embryos lack all anterior structures and develop as two outward facing tail ends ('double abdomen').

Anterior determinants can be encoded by new genes with a dedicated function in establishing embryonic polarity. One example is *bicoid* in the fruit fly *Drosophila melanogaster*. Maternal mRNA of *bicoid* is localized in the anterior pole of the egg and Bicoid protein is expressed in a gradient in the early embryo (Berleth et al., 1988). Bicoid-deficient embryos fail to develop anterior structures and instead form a second tail end, or a symmetrical double abdomen when the maternal activity gradient of another gene, *hunchback*, is disrupted simultaneously (Driever, 1993). The *bicoid* gene originated in the lineage of cyclorrhaphan flies more than 140 million years ago by duplication of *zen* (*knüllt* (*zen*; aka *Hox3*), which, in insects, plays an important role in extraembryonic tissue

**eLife digest** With very few exceptions, animals have ‘head’ and ‘tail’ ends that develop when they are an embryo. The genes involved in specifying these ends vary between species and even closely-related animals may use different genes for the same roles. For example, the products of two unrelated genes called *bicoid* in fruit flies and *panish* in common midges accumulate at one end of their respective eggs to distinguish head from tail ends. It remained unclear how other fly species, which have neither a *bicoid* nor a *panish* gene, distinguish the head from the tail end, or how genes can evolve the specific function of *bicoid* and *panish*.

Cells express genes by producing gene templates called messenger ribonucleic acids (or mRNAs for short). The central portions of mRNAs, known as protein-coding sequences, are then used to produce the protein. Proteins can play several distinct roles, which they acquire through evolution. This can happen in different ways, for example, genetic mutations in the part of a gene that codes for protein may alter the resulting protein, giving it a new activity. Alternatively, sequences at the beginning and the end of an mRNA molecule that do not code for protein, but regulate when and where proteins are made, can influence a protein’s role by changing its environment. Many genes produce mRNAs with alternative sequences at the beginning or the end, a process known as alternative transcription.

Here, Yoon et al. identified three unrelated genes that perform similar roles to *bicoid* and *panish* in the embryos of several different moth flies and mosquitoes. These genes appear to have acquired their activity because one of their alternative transcripts accumulated at the future head end, rather than through mutations in the protein-coding sequences. Studying multiple species also made it clear that *panish* inherited its function from a localized alternative transcript of an old gene that duplicated and diverged.

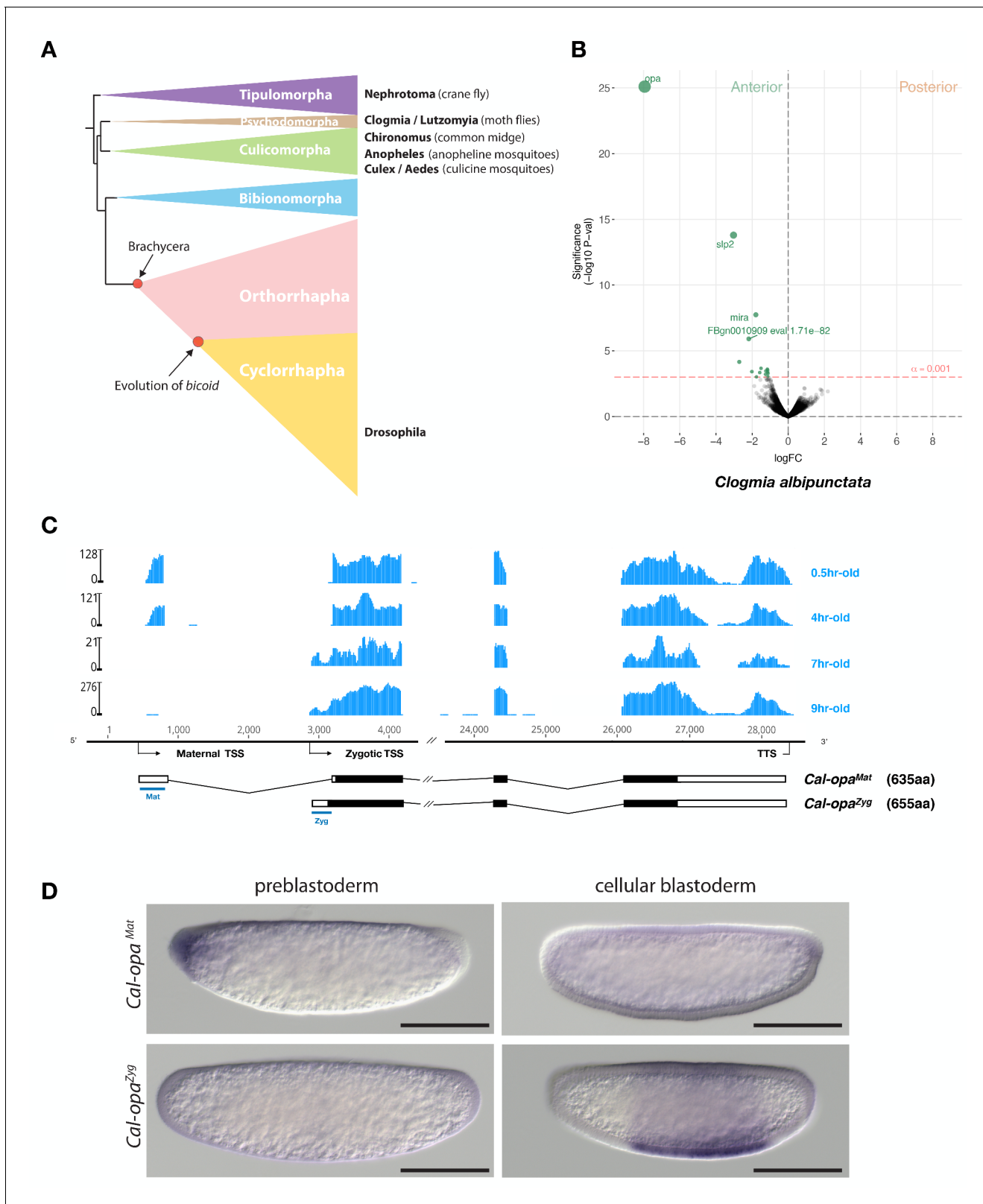
These findings suggest that alternative transcription may provide opportunities for genes to evolve new roles in fundamental processes in flies. Most animal genes use alternative start and stop sites for transcription, but the reasons for this remain largely obscure. This is especially the case in the human brain. The findings of Yoon et al., therefore, raise the question of whether alternative transcription has played an important role in the evolution of the human brain.

DOI: <https://doi.org/10.7554/eLife.46711.002>

development (**Schmidt-Ott et al., 2010**). The expression and function of cyclorrhaphan *bicoid* orthologs are conserved but *bicoid* has not been found outside this group, and has been lost in some lineages within the Cyclorrhapha.

Another example is *panish*, which encodes the anterior determinant of a midge, *Chironomus riparius*. This gene evolved by gene duplication of the *Tcf* homolog *pangolin* (*pan*) and capture of the maternal promoter of a nucleoside kinase gene, and has been called *panish* (for *pan*“ish”) (**Klomp et al., 2015**). Pangolin functions as the effector of  $\beta$ -catenin-dependent Wnt signaling pathway (‘canonical’ Wnt signaling) but Panish lacks the  $\beta$ -catenin domain of Pangolin, and sequence similarity between Pangolin and Panish is limited to the cysteine-clamp domain (30 amino acids). *panish* has not been found outside the family Chironomidae, suggesting that lower dipterans use different anterior determinants.

Here, we have used embryos of a wider range of dipteran species that lack *bicoid* and *panish* to address the question of how anterior determinants evolve. We started our analysis with moth flies (Psychodidae: *Clogmia albipunctata*, *Lutzomyia longipalpis*) and subsequently extended it to mosquitoes (Culicidae: *Culex quinquefasciatus*, *Aedes aegypti*, *Anopheles gambiae*, *Anopheles coluzzii*), and to crane flies (Tipulidae: *Nephrotoma suturalis*) (**Figure 1A**). Our results reveal three distinct old genes that evolved anterior determinants by localizing an alternative maternal transcript isoform at the anterior egg pole of the respective species. Therefore, alternative transcription might have played an important role in the evolution of this gene function and gene regulatory networks in fly embryos.



**Figure 1.** Expression of alternative *Cal-opa* transcripts in *Clogmia* embryos. (A) Phylogenetic relationship of fly species referred to in the text (Wiegmann et al., 2011). (B) Differential expression analysis of maternal transcripts between anterior and posterior halves of 1 hr-old *Clogmia* embryos. logFC: log fold-change. (C) Stage-specific RNA-seq read coverage of *Cal-opa* locus. Transcription start sites (TSS) and transcription termination sites (TTS) are indicated on the genomic scaffold (solid line with 1000 bp intervals marked) and were confirmed by RACE. Exon-intron sketches of *Cal-opa<sup>Mat</sup>* Figure 1 continued on next page

Figure 1 continued

and *Cal-opa<sup>Zyg</sup>* transcript variants are shown with the open reading frame in black and the position of in situ hybridization probes and dsRNA underlined in blue. (D) RNA in situ hybridization of *Cal-opa<sup>Mat</sup>* and *Cal-opa<sup>Zyg</sup>* transcripts in 1 hr-old preblastoderm and 7 hr-old cellular blastoderm embryos. Anterior is left and dorsal up. Scale bar: 100µm.

DOI: <https://doi.org/10.7554/eLife.46711.003>

The following figure supplement is available for figure 1:

**Figure supplement 1.** Protein alignment of predicted dipteran Odd-paired orthologs.

DOI: <https://doi.org/10.7554/eLife.46711.004>

## Results

### An alternative maternal transcript of the conserved segmentation gene *odd-paired* functions as anterior determinant in *Clogmia*

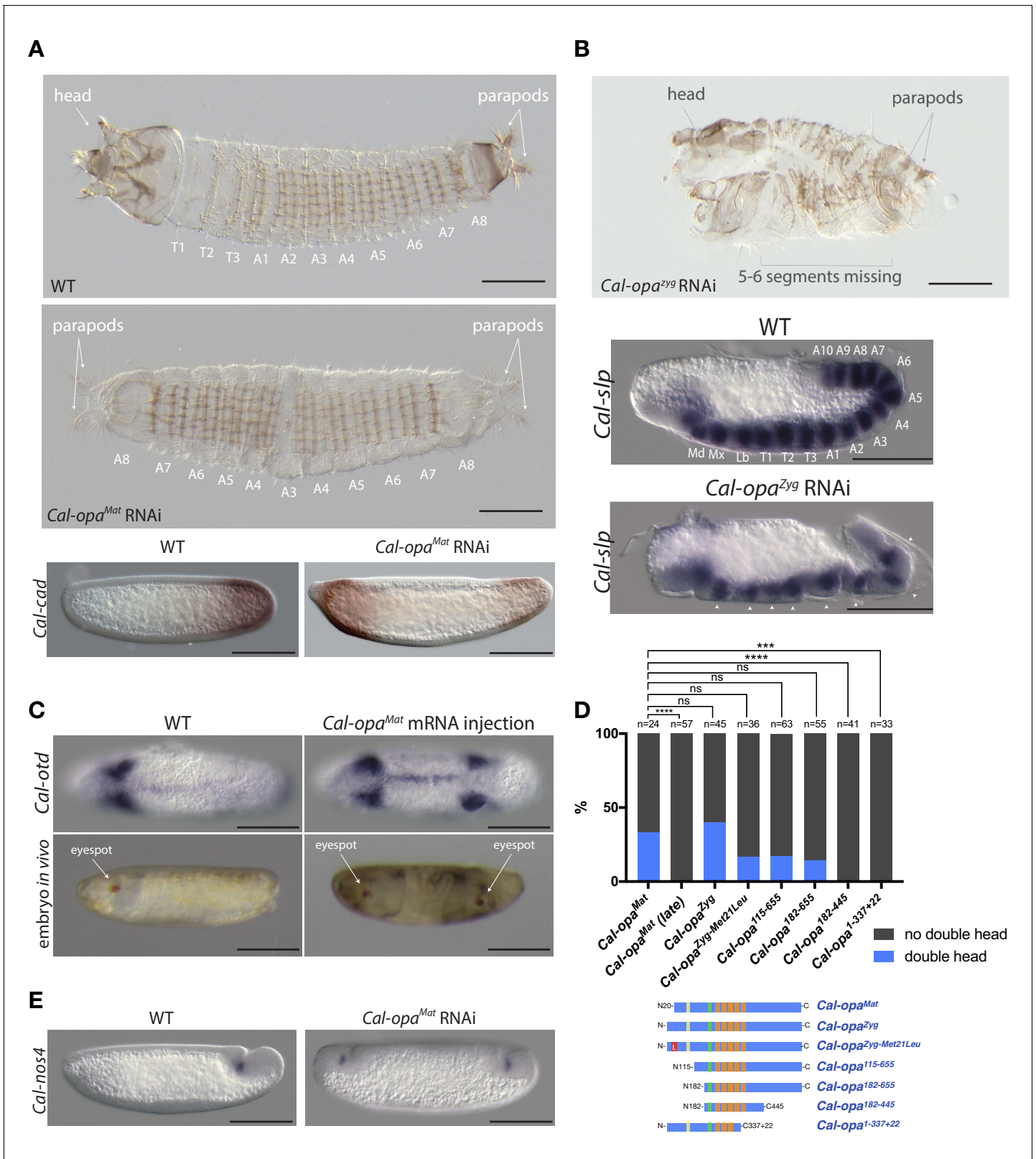
We annotated 5602 transcripts from the anterior and posterior transcriptomes of 1 hr-old bisected *Clogmia* embryos and ranked them according to the magnitude of their differential expression scores and P values (Figure 1B). In the anterior embryo, the most enriched transcript was homologous to *odd-paired*, the *Drosophila* homolog of mammalian *Zic* (zinc finger of the cerebellum) genes. ZIC proteins are known to function as transcription factors or co-factors (Houtmeyers et al., 2013). *odd-paired* was discovered in a screen for early *Drosophila* segmentation genes and subsequently classified as a 'pair-rule' gene, since *odd-paired* mutants fail to develop alternating segments (Jürgens et al., 1984). During the *Drosophila* segmentation process, *odd-paired* is expressed in a single broad domain and controls the 'frequency-doubling' of other pair-rule genes (Clark and Akam, 2016).

The *Clogmia* genome contains a single *odd-paired* locus (*Cal-opa*) (Vicoso and Bachtrog, 2015). Using RNA-seq data from preblastoderm and blastoderm embryos and Rapid Amplification of cDNA Ends (RACE), we identified maternal and zygotic *Cal-opa* transcripts with alternative first exons that we mapped onto a 54 kb genomic scaffold (Figure 1C). The maternal transcript (*Cal-opa<sup>Mat</sup>*) was detected in preblastoderm embryos (0.5 hr-old) and syncytial blastoderm embryos (4 hr-old). The zygotic transcript (*Cal-opa<sup>Zyg</sup>*) was found in cellularized blastoderm embryos (7 hr-old) and gastrulating embryos (9 hr-old). Protein alignments with homologs from other flies suggest that *Cal-opa<sup>Zyg</sup>* encodes the full-length Cal-Opa protein (655 amino acids), while *Cal-opa<sup>Mat</sup>* encodes a truncated protein variant (635 amino acids), lacking the N-terminal 20 amino acids of *Cal-opa<sup>Zyg</sup>* (Figure 1—figure supplement 1).

To confirm the alternative *Cal-opa<sup>Mat</sup>* and *Cal-opa<sup>Zyg</sup>* transcripts and their non-overlapping expression patterns, we performed whole mount RNA in situ hybridization experiments with transcript-specific probes. The *Cal-opa<sup>Mat</sup>* transcript was anteriorly localized in preblastoderm embryos but absent at the cellular blastoderm stage. Conversely, the *Cal-opa<sup>Zyg</sup>* transcript was absent in preblastoderm embryos but expressed broadly in the trunk region of 7 hr-old blastoderm embryos (Figure 1D), like *odd-paired* in *Drosophila*. These observations suggest that *Cal-opa* produces transcript isoforms with spatially and temporally distinct expression patterns.

To determine the function of *Cal-opa<sup>Mat</sup>* and *Cal-opa<sup>Zyg</sup>*, we established a protocol for microinjecting early *Clogmia* embryos and conducted transcript-specific RNA interference (RNAi) experiments. Injection of *Cal-opa<sup>Mat</sup>* double-stranded RNA (dsRNA) led to mirror-image duplications of the tail end (double abdomen; Figure 2A and Figure 2—figure supplement 1). In contrast, injection of dsRNA targeting *Cal-opa<sup>Zyg</sup>* resulted in half the number of segmental expression domains of *Cal-slp* (the ortholog of pair-rule gene *sloppy-paired*) and caused defects in segmentation, dorsal closure, and head development but did not alter embryo polarity (Figure 2B). Finally, injection of dsRNA targeting both *Cal-opa<sup>Mat</sup>* and *Cal-opa<sup>Zyg</sup>* resulted in double abdomens with missing segments (Figure 2—figure supplement 2). These observations indicate distinct roles of *Cal-opa<sup>Mat</sup>* and *Cal-opa<sup>Zyg</sup>* in specifying embryo polarity and in segmentation, respectively.

We noticed that maternal transcripts of *Cal-slp* and *Cal-mira*, a homolog of *miranda*, which encodes an adaptor protein for cell fate determinants in *Drosophila* (Ikeshima-Kataoka et al., 1997; Adams et al., 2000), were also slightly enriched in the anterior portion of embryo (Figure 1B). This observation was confirmed by RNA in situ hybridizations (Figure 2—figure supplement 3). Injection



**Figure 2.** Function of alternative *Cal-opa* transcripts in early *Clogmia* embryo. (A) 1 st instar larval cuticle of wild type (top) and following *Cal-opa<sup>Mat</sup>* RNAi (middle). RNA in situ hybridization of *Cal-cad* in a wild-type preblastoderm embryo (bottom left) and stage-matched *Cal-opa<sup>Mat</sup>* RNAi embryos (bottom right). Anterior is left and dorsal up. T: thoracic segment; A: abdominal segment. Segment numbers in *Cal-opa<sup>Mat</sup>* RNAi larval cuticle were assigned based on the assumption of polarity reversal. Scale bar: 100 $\mu$ m. (B) 1 st instar larval cuticle phenotype following *Cal-opa<sup>Zyg</sup>* RNAi (top) and Figure 2 continued on next page

Figure 2 continued

RNA in situ hybridization of *Cal-slp* in extending wild-type germband (middle) and stage-matched *Cal-opa<sup>Zyg</sup>* RNAi embryo (bottom). Anterior is left and dorsal up. Md: mandibular segment; Mx: maxillary segment; Lb: labial segment; T: thoracic segment; A: abdominal segment. Scale bar: 100µm. (C) RNA in situ hybridizations of *Cal-otd* in wild-type gastrula (top left) and stage-matched embryo following posterior *Cal-opa<sup>Mat</sup>* mRNA injection (top right) are shown in ventral view. A live wild-type embryo (bottom left) and a stage-matched embryo following posterior *Cal-opa<sup>Mat</sup>* mRNA injection (bottom right) in lateral view. Anterior is left. Scale bar: 100µm. (D) Posterior injection of *Cal-opa* mRNA and mutated variants. Complete, symmetrical duplication of the bilateral *Cal-otd* expression domain in gastrulating embryos was counted as double head (blue, see **Figure 2C**). All other phenotypes, including incomplete duplications and wild type, were conservatively counted as 'no double head' (black). Sketches of predicted Cal-Opa proteins are shown with ZIC/Opa conserved motif (ZOC) in yellow, the ZIC family protein N-terminal conserved domain (ZFNC) in green, and zinc finger domains in orange. The Met21Leu mutation in *Cal-Opa<sup>Zyg-Met21Leu</sup>* is marked in red. *Cal-opa<sup>Mat</sup>* (late): *Cal-opa<sup>Mat</sup>* was injected during the syncytial blastoderm stage (4 hr). ns:  $p > 0.05$ ; \*\*\*:  $p < 0.001$ ; \*\*\*\*:  $p < 0.0001$ , Fisher's exact test. (E) RNA in situ hybridization of *Cal-nos4* in a gastrulating embryo (left) and stage-matched *Cal-opa<sup>Mat</sup>* RNAi embryos (right).

DOI: <https://doi.org/10.7554/eLife.46711.005>

The following figure supplements are available for figure 2:

**Figure supplement 1.** Frequencies of strong *Cal-opa<sup>Mat</sup>* and *Cal-opa<sup>Zyg</sup>* RNAi phenotypes.

DOI: <https://doi.org/10.7554/eLife.46711.006>

**Figure supplement 2.** Cuticle phenotype of a 1st instar larva following *Cal-opa<sup>Mat</sup>* and *Cal-opa<sup>Zyg</sup>* double RNAi.

DOI: <https://doi.org/10.7554/eLife.46711.007>

**Figure supplement 3.** Maternal transcript localization and RNAi phenotypes of *Cal-slp* and *Cal-mira* in *Clogmia*.

DOI: <https://doi.org/10.7554/eLife.46711.008>

**Figure supplement 4.** Cuticle phenotype of a 1st instar *Clogmia* larva following *Cal-opa<sup>Mat</sup>* mRNA injection.

DOI: <https://doi.org/10.7554/eLife.46711.009>

**Figure supplement 5.** *Clogmia* homologs of *nanos*, *vasa*, *tudor*, and *germ cell-less*.

DOI: <https://doi.org/10.7554/eLife.46711.010>

of *Cal-slp* dsRNA resulted in head and dorsal closure defects while *Cal-mira* dsRNA caused labrum and antennal defects, but in both cases embryo polarity was retained (**Figure 2—figure supplement 3**).

To test whether *Cal-opa<sup>Mat</sup>* can induce head development ectopically, we injected *Cal-opa<sup>Mat</sup>* mRNA into the posterior pole of 1 hr-old embryos. These embryos expressed a head marker, *Cal-otd* (ortholog of *ocelliless/orthodenticle*), on both ends of the embryo and developed a symmetrical double head, including some duplicated thoracic elements (**Figure 2C** and **Figure 2—figure supplement 4** and **Video 1**). These observations suggest that anterior enrichment of maternal transcripts other than *Cal-opa<sup>Mat</sup>* mRNA is not essential for head development, and that *Cal-opa<sup>Mat</sup>* localization is sufficient for establishing embryo polarity.

### The anterior determinant function of *Cal-opa* is sensitive to expression timing but insensitive to 5' truncation of the open reading frame

Next, we asked whether the early timing of *odd-paired* expression is critical for its function as anterior determinant in moth flies. To test this hypothesis, we conducted posterior injections of *Cal-opa<sup>Mat</sup>* mRNA during the syncytial blastoderm stage (4 hr) and examined *Cal-otd* expression after gastrulation. These embryos developed with normal head-to-tail polarity (**Figure 2D**). This result places the requirement of *odd-paired* for axis specification prior to the syncytial blastoderm stage and suggests that early timing of *odd-paired* activity is essential for its function as anterior determinant.

*Cal-opa<sup>Mat</sup>* and *Cal-opa<sup>Zyg</sup>* mRNAs not only differ in the timing of expression, but also differ in their 5'UTRs and predicted N-terminal protein sequences, as mentioned above (**Figure 1C** and **Figure 1—figure supplement 1**). To test



**Video 1.** Live double head embryo in ventrolateral view. The larval cuticle of this embryo is shown in **Figure 3—figure supplement 1**.

DOI: <https://doi.org/10.7554/eLife.46711.011>

whether the open reading frame difference is required for the anterior determinant function, we injected *Cal-opa<sup>Zyg</sup>* mRNA at the posterior pole of preblastoderm embryos. Embryos from this experiment also developed as double heads (**Figure 2D**). Because the translation start site of *Cal-Opa<sup>Mat</sup>* is located downstream of the *Cal-Opa<sup>Zyg</sup>* translation start site, we also tested *Cal-opa<sup>Zyg</sup>* mRNA in which the putative start codon for *Cal-Opa<sup>Mat</sup>* was mutated to encode leucine (*Cal-opa<sup>Zyg-Met21Leu</sup>*). Posterior injection of *Cal-opa<sup>Zyg-Met21Leu</sup>* mRNA also resulted in double heads (**Figure 2D**). These findings indicate that the protein difference between *Cal-Opa<sup>Mat</sup>* and *Cal-Opa<sup>Zyg</sup>* is not essential for the anterior determinant function of *odd-paired* in *Clogmia*.

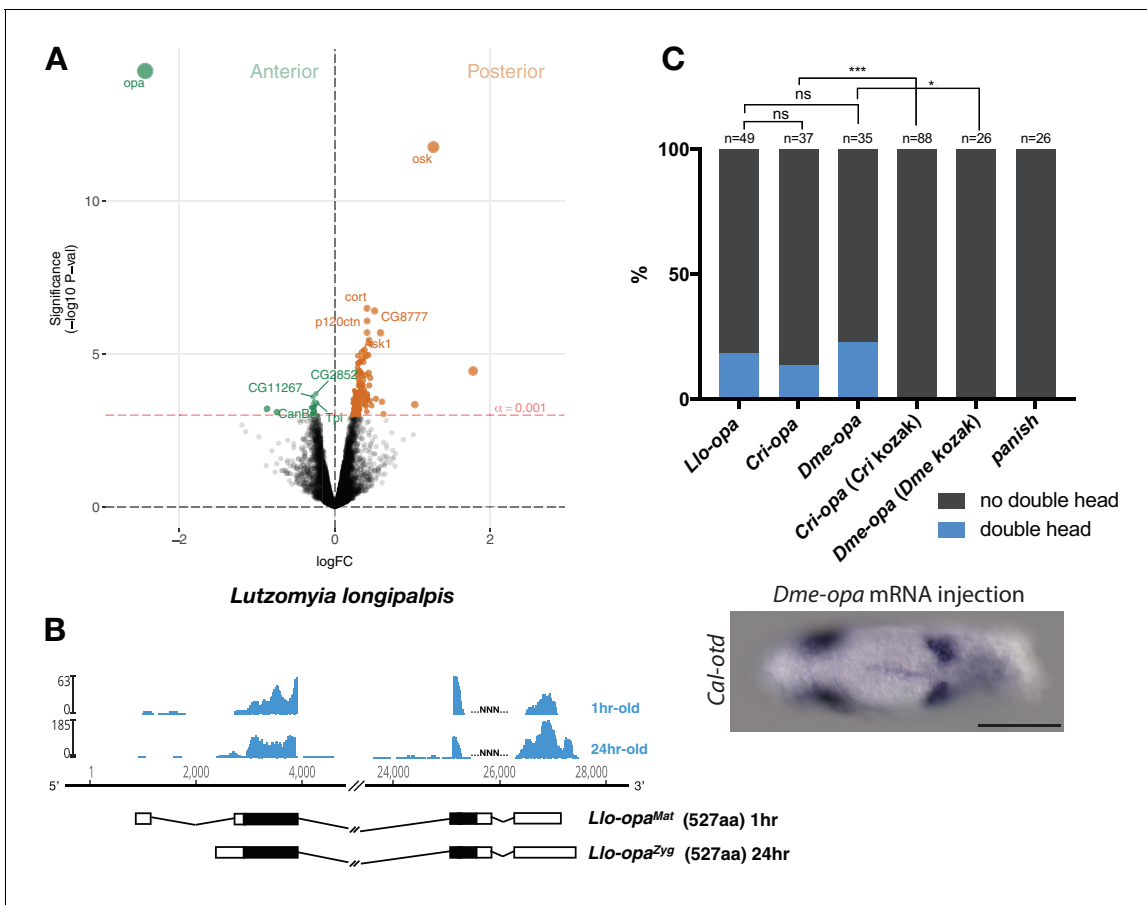
To test whether only a small portion of the *Cal-opa* open reading frame is required for its function as anterior determinant, we examined the ability of various truncated variants of *Cal-Opa* mRNAs (**Figure 1—figure supplement 1**) to induce head development at the posterior egg pole (**Figure 2D**). mRNAs of protein variants with large N-terminal truncation (*Cal-Opa<sup>115-655</sup>* and *Cal-Opa<sup>182-655</sup>*) retained the ability to induce double heads. However, mRNAs of protein variants with C-terminal truncation (*Cal-Opa<sup>182-445</sup>* and *Cal-Opa<sup>1-337+22</sup>*, a hypothetical splice variant) failed to induce double heads. These results indicate that the ability of *Cal-Opa* to specify embryo polarity requires the C-terminal portion of the protein but is largely insensitive to N-terminal truncation, corroborating our above conclusion that N-terminal differences between *Cal-Opa<sup>Mat</sup>* and *Cal-Opa<sup>Zyg</sup>* were not essential for evolving the anterior determinant function of *Cal-opa*.

### ***Cal-opa<sup>Mat</sup>* suppresses zygotic germ cell specification at the anterior pole and *Clogmia* lacks maternal germ plasm**

In *Drosophila* and other dipterans, maternal germ plasm in the posterior embryo not only specifies primordial germ cells but also contributes to and stabilizes embryo polarity via *nanos*, which suppresses the translation of anterior determinants in the posterior embryo (Tautz, 1988; Gavis and Lehmann, 1992; Struhl et al., 1992; Lemke and Schmidt-Ott, 2009). The activity of *nanos* in the posterior preblastoderm is dependent on *oskar* (Lehmann, 2016), which is conserved in many insects (Ewen-Campen et al., 2010). However, in *Clogmia*, expression profiling of anterior and posterior egg halves did not reveal any posteriorly localized maternal transcripts (**Figure 1B**,  $\alpha = 0.001$ , unadj.), and no *oskar* homolog was found in our *Clogmia* transcriptomes or the *Clogmia* genome (Vicoso and Bachtrog, 2015). To test whether *Clogmia* lacks maternal germ plasm, we examined the expression of candidate germ cell markers, including *Clogmia* homologs of *nanos* (*Cal-nos1*, *Cal-nos2*, *Cal-nos3*, and *Cal-nos4*), *vasa* (*Cal-vas*), *tudor* (*Cal-tud*), and *germ cell-less* (*Cal-gcl*). *Cal-nos1*, *Cal-nos3*, and *Cal-nos4* were not localized in the posterior of preblastoderm embryos but were expressed in a small set of cells at the posterior pole of cellular blastoderm and gastrulating embryos along with *Cal-vas*, *Cal-tud*, and *Cal-gcl* that were expressed more broadly (**Figure 2—figure supplement 5**). These observations suggest that *Clogmia* lacks maternal germ plasm and that *Clogmia* may induce the germ cell fate zygotically. To test this hypothesis, we examined *Cal-nos* expression in *Cal-opa<sup>Mat</sup>* RNAi embryos. *Cal-nos* positive cells were duplicated in double abdomens (**Figure 2E**), indicating that *Clogmia* uses an inductive mechanism for germ cell specification, which is repressed in the anterior embryo by *Cal-opa<sup>Mat</sup>*. Therefore, axis specification in the *Clogmia* embryo is independent from germ cell specification. To our knowledge, *Clogmia* is also the first example of inductive germ cell specification in flies.

### **Evolution of the anterior determinant function of moth fly *odd-paired***

Maternal *odd-paired* transcript is absent in freshly deposited eggs of chironomids (Klomp et al., 2015) and mosquitoes (Akbari et al., 2013), both of which belong to the Culicomorpha lineage (**Figure 1A**). To test whether localized maternal *odd-paired* transcript is broadly conserved in the Psychodomorpha lineage, we examined maternal transcript localization in the eggs of the sand fly *Lutzomyia longipalpis*, a moth fly species of public health concern due to its role in the transmission of visceral leishmaniasis. Of 5392 annotated transcripts, the most enriched maternal transcript in the anterior half of 1–2 hr old embryos was homologous to *odd-paired* and was therefore named *Llo-opa<sup>Mat</sup>* (**Figure 3A**). In the posterior *Lutzomyia* embryo, the most enriched transcript was homologous to *oskar*, indicating that *Lutzomyia* eggs contain maternal germ plasm at the posterior pole, unlike the *Clogmia* eggs. These findings suggest that a broad range of moth flies use *odd-paired* transcript as anterior determinant, and that maternal germ plasm was lost only in the *Clogmia*



**Figure 3.** Alternative *odd-paired* transcript isoforms in *Lutzomyia* and ectopic head induction by mRNAs derived from *Cal-opa* orthologs in *Clogmia*. (A) Differential expression analysis of maternal transcripts between anterior and posterior halves of 1–2 hr-old *Lutzomyia* embryos. logFC: log fold-change. (B) Stage-specific RNA-seq read coverage of *Llo-opa* genomic locus and sketches of *Llo-opa<sup>Mat</sup>* and *Llo-opa<sup>Zyg</sup>* transcripts (see also legend to **Figure 1A**). (C) Posterior injection of *odd-paired* mRNAs from *Lutzomyia*, *Chironomus*, and *Drosophila* in *Clogmia* embryo. Phenotypes were counted as in (**Figure 2D**) and double head from *Dme-opa* mRNA injection is shown as example. *Cri-opa* and *Dme-opa* mRNAs include Kozak sequence of *Cal-opa<sup>Mat</sup>* (TAAG upstream of the predicted translation start site). *Cri-kozak* and *Dme-kozak* refer to *odd-paired* sequences with donor-specific Kozak-sequences from *Chironomus* (AAAA) and *Drosophila* (GACC), respectively. ns.:  $p > 0.05$ ; \*:  $p < 0.05$ . \*\*\*:  $p < 0.001$ , Fisher's exact test.

DOI: <https://doi.org/10.7554/eLife.46711.012>

The following figure supplement is available for figure 3:

**Figure supplement 1.** Expression level of select *Lutzomyia* gap gene (red) and pair-rule gene (blue) homologs.

DOI: <https://doi.org/10.7554/eLife.46711.013>

lineage. Close examination of *Lutzomyia* transcriptomes from 1 hr-old and 24 hr-old embryos also revealed zygotic *odd-paired* transcript (*Llo-opa<sup>Zyg</sup>*) (**Figure 3B**). *Llo-opa<sup>Mat</sup>* and *Llo-opa<sup>Zyg</sup>* share the same open reading frame but differ at their untranslated 5' and 3' ends. Since the N-terminal ends of *Llo-Op<sup>Mat</sup>*/*Llo-Op<sup>Zyg</sup>* and *Cal-Op<sup>Zyg</sup>* proteins are homologous (**Figure 1—figure supplement 1**), we infer that the N-terminal truncation of *Cal-Op<sup>Mat</sup>* occurred after the transcript had evolved maternal expression and anterior localization. The detection of *Llo-opa<sup>Zyg</sup>* transcript in 24 hr-old embryos coincided with that of gap and pair-rule segmentation gene homologs (**Figure 3—figure supplement 1**), indicating that *Llo-opa<sup>Zyg</sup>* functions during segmentation.

The *odd-paired* gene of ancestral moth flies could have evolved the ability to establish the embryo polarity via specific amino acid substitutions. In this case, *odd-paired* homologs from species with a different anterior determinant, such as *Drosophila* or *Chironomus*, should not induce ectopic head development in *Clogmia* embryos. Alternatively, *odd-paired* could have evolved its role as axis determinant in moth flies independent of any amino acid substitution via co-option. In this case, *odd-paired* homologs from *Drosophila* or *Chironomus* could have the ability to induce head



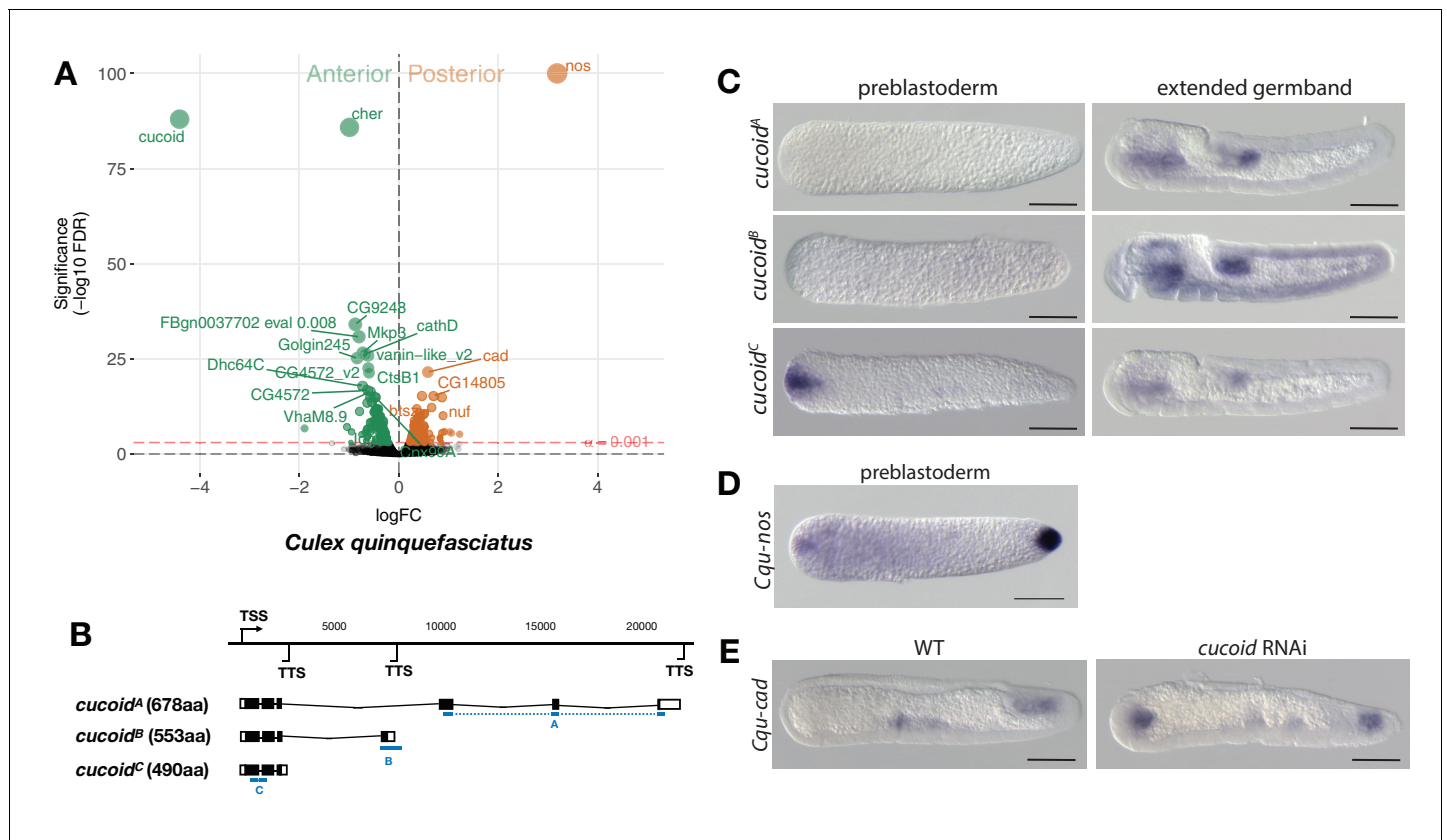
development in *Clogmia* embryos when appropriately expressed. To test this possibility, we injected *odd-paired* mRNA from *Lutzomyia*, *Chironomus*, or *Drosophila* into the posterior pole of early *Clogmia* embryos. All of these *odd-paired* homologs induced double heads in *Clogmia*, provided that the endogenous kozak sequence of *Cal-opa<sup>Mat</sup>* was used for optimal translation efficiency (**Figure 3C**). Since neither *Drosophila* nor *Chironomus* uses *odd-paired* for specifying embryo polarity, these results suggest that amino acid substitutions were not essential for the evolution of the anterior determinant function of *odd-paired* in moth flies. We therefore propose that this gene function evolved via co-option when alternative maternal transcript of moth fly *odd-paired* became enriched at the anterior egg pole.

## A previously uncharacterized C2H2 zinc finger gene, named *cucoid*, functions as anterior determinant in culicine mosquitoes

Given that freshly deposited mosquito eggs lack maternal *odd-paired* transcript orthologs (**Akbari et al., 2013**), we extended our search for anterior determinants to mosquitoes. Initially, we focused on the Southern House Mosquito *Culex quinquefasciatus*, a vector of West Nile virus (the leading cause of mosquito-borne disease in the continental United States) and of *Wuchereria bancrofti* (the major cause of lymphatic filariasis). This species was chosen because their eggs are large and have clearly distinguishable anterior and posterior egg poles. We annotated 8239 *Culex* transcripts from the pooled anterior and posterior transcriptomes of 1 hr-old preblastoderm embryos and ranked them according to the magnitude of their differential expression scores and P values (**Figure 4A**). In the posterior embryo, the most enriched transcript was related to *nanos*, consistent with the presence of maternal germ plasm in this species (**Figure 4D**) (**Juhn et al., 2008**). The most enriched transcript in the anterior embryo was closely related to an uncharacterized gene of *Drosophila* (CG9215). However, reciprocal BLAST searches suggest that CG9215 belongs to a poorly defined larger gene family in *Drosophila* that might be represented by a single gene in mosquitoes. We named this gene *cucoid* to reflect its *bicoid*-like function in a culicine mosquito. *cucoid* encodes a protein with five C2H2 zinc finger domains (**Figure 4—figure supplement 1**). RACE experiments with cDNA from 0 to 7 hr-old embryos revealed three alternative *cucoid* transcripts with distinct 3' ends (*cucoid<sup>A</sup>*, *cucoid<sup>B</sup>*, and *cucoid<sup>C</sup>*) (**Figure 4B**), but only *cucoid<sup>B</sup>* and *cucoid<sup>C</sup>* were recovered from cDNA of 0–2 hr-old preblastoderm embryos, suggesting that one or both these transcripts might be maternally localized at the anterior pole. To test this hypothesis, we performed RNA in situ hybridizations with specific probes for *cucoid<sup>A</sup>* (probe A) or *cucoid<sup>B</sup>* (probe B) and, due to the very short sequence unique to *cucoid<sup>C</sup>* (121 nucleotides), a probe against all three isoforms (probe C). *cucoid<sup>A</sup>* and *cucoid<sup>B</sup>* expression was detected in the fore and hind gut of extended germbands but not in 1 hr-old preblastoderm embryos. In contrast, the probe against all three isoforms detected maternally localized *cucoid* transcript at the anterior pole in addition to the zygotic expression pattern (**Figure 4C**). Taken together, these results suggest that only the *cucoid<sup>C</sup>* isoform is maternally localized at the anterior pole and could function as anterior determinant. To test this hypothesis, we injected *cucoid* dsRNA from the shared 5' region and examined the expression of a posterior marker (*Cqu-cad*) in gastrulating embryos. Many of these embryos expressed *Cqu-cad* in the anterior and underwent ectopic gastrulation at the anterior pole, suggesting that normal head-to-tail polarity was lost (**Figure 4E**). Taken together, our results suggest that *cucoid* acquired the anterior determinant function via the localization of a maternal transcript isoform with an alternative 3' end.

We obtained similar results in another culicine mosquito, the Yellow Fever Mosquito *Aedes aegypti*, which transmits Dengue, Chikungunya, and Zika viruses. In this species, expression profiling of 5802 transcripts from the anterior and posterior transcriptomes of 1 hr-old preblastoderm embryos also identified *cucoid* (*Aae-cucoid*) as the gene with the most significantly enriched transcript in the anterior embryo (**Figure 5A**). RACE experiments with cDNA from 1 hr to 6 hr-old embryos revealed three similar *Aae-cucoid* transcripts with alternative 3' ends (*Aae-cucoid<sup>A</sup>*, *Aae-cucoid<sup>B</sup>*, and *Aae-cucoid<sup>C</sup>*) (**Figure 5B**), and RNA in situ hybridization experiments with a probe against all three isoforms confirmed the anterior localization of *Aae-cucoid* transcript in preblastoderm embryos (**Figure 5C**). *Aae-cucoid* expression in *Aedes* germbands could not be examined for technical reasons.

In the posterior embryo, no highly enriched transcripts were observed. This was unexpected given that whole mount in situ hybridizations revealed posterior localized transcript of *Aedes nanos* in 1 hr-old embryos (**Figure 5C**) and that maternal transcript of *Aedes oskar* is also localized at the



**Figure 4.** Expression and function of alternative *cucooid* isoforms in *Culex*. (A) Differential expression analysis of maternal transcripts between anterior and posterior halves of 1 hr-old *Culex* embryos. logFC: log fold-change. (B) Sketches of *cucooid* transcript isoforms based on RNA-seq and RACE experiments (see also legend to **Figure 1A**). Blue lines the position of in situ hybridization probes. (C) RNA in situ hybridization of *cucooid*<sup>A</sup>, *cucooid*<sup>B</sup>, and *cucooid*<sup>C</sup> transcripts in 1 hr-old preblastoderm and germband extending embryos. Anterior is left and dorsal up. Scale bar: 100µm. (D) RNA in situ hybridization of *Cqu-nos* in 1 hr-old preblastoderm embryo. Anterior is left. Scale bar: 100µm (E) RNA in situ hybridization of *Cqu-cad* in a wild-type gastrulating embryo (left) and in a stage-matched *cucooid* RNAi embryos (right; 16/48 with *cucooid* dsRNA versus 0/25 with lacZ control dsRNA; p<0.0005). Anterior is left and dorsal up. Scale bar: 100µm.

DOI: <https://doi.org/10.7554/eLife.46711.014>

The following figure supplement is available for figure 4:

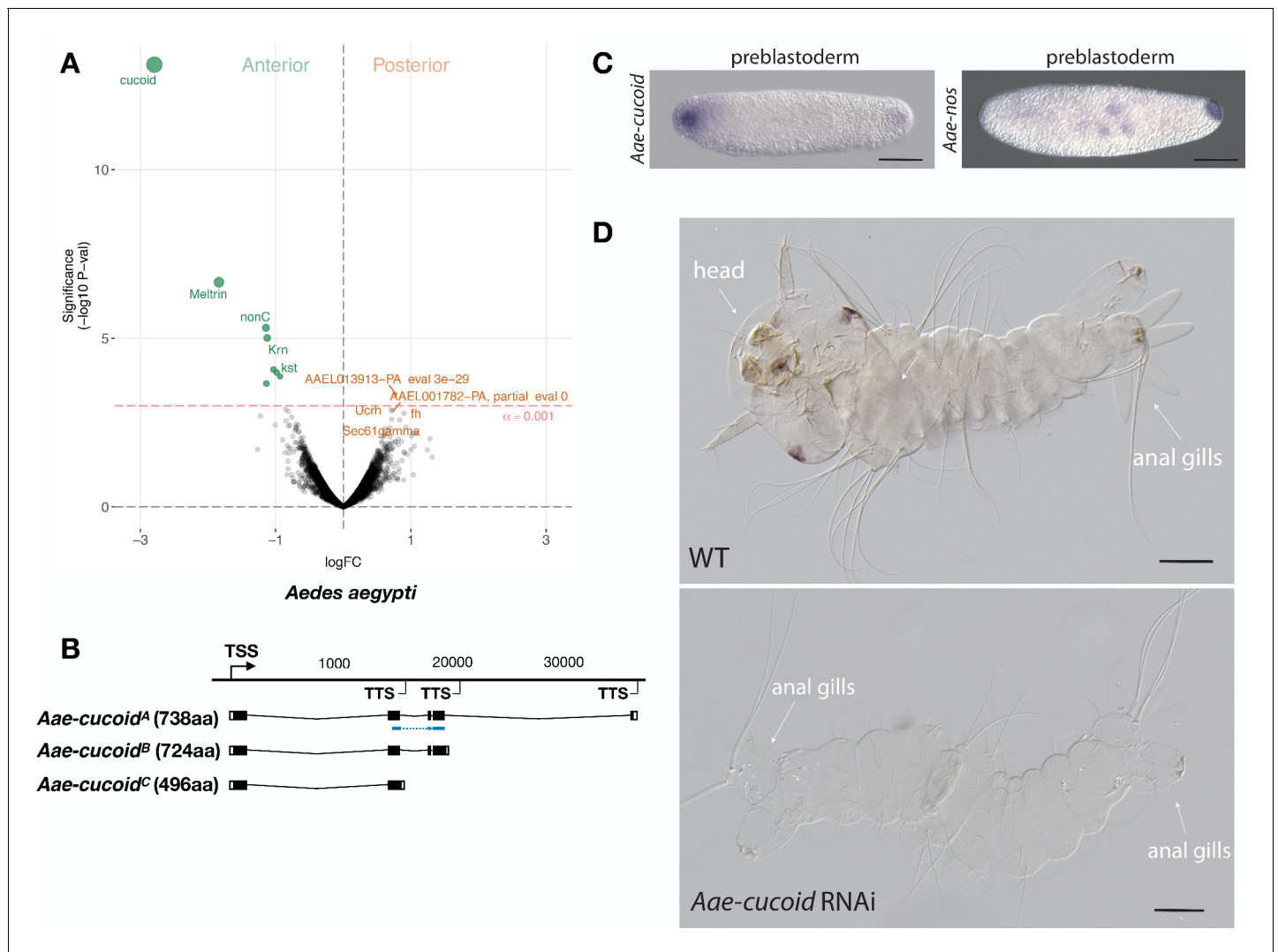
**Figure supplement 1.** Protein alignment of dipteran *cucooid* orthologs and a *Drosophila* paralog CG4424.

DOI: <https://doi.org/10.7554/eLife.46711.015>

posterior pole (*Juhn and James, 2006*). Low statistical power of our differential expression analysis in *Aedes* might explain this discrepancy, since we could have confounded the anterior and posterior pole in some of the bisected *Aedes* eggs (see Materials and methods). Alternatively, only a small portion of these transcripts might be localized at the posterior pole. Injection of *Aae-cucooid* dsRNA against the shared region of all transcripts resulted in double abdomens (**Figure 5D**), suggesting that *cucooid* evolved its function as anterior determinant prior to the divergence of the *Culex* and *Aedes* lineages.

### ***pangolin/Tcf* functions as anterior determinant in anopheline mosquitoes**

The *Anopheles gambiae* species complex constitutes an outgroup to the *Culex-Aedes* clade. It includes eight or more sub-Saharan species that are difficult to distinguish due to widespread genealogical heterogeneity across the genome, incomplete lineage sorting and introgression (*Thawornwattana et al., 2018*). We interchangeably used *A. gambiae* and *A. coluzzii*, two sibling species within this species complex that are responsible for the majority of malaria transmission in Africa, to identify the anterior determinant of this mosquito lineage. Whole mount RNA in situ

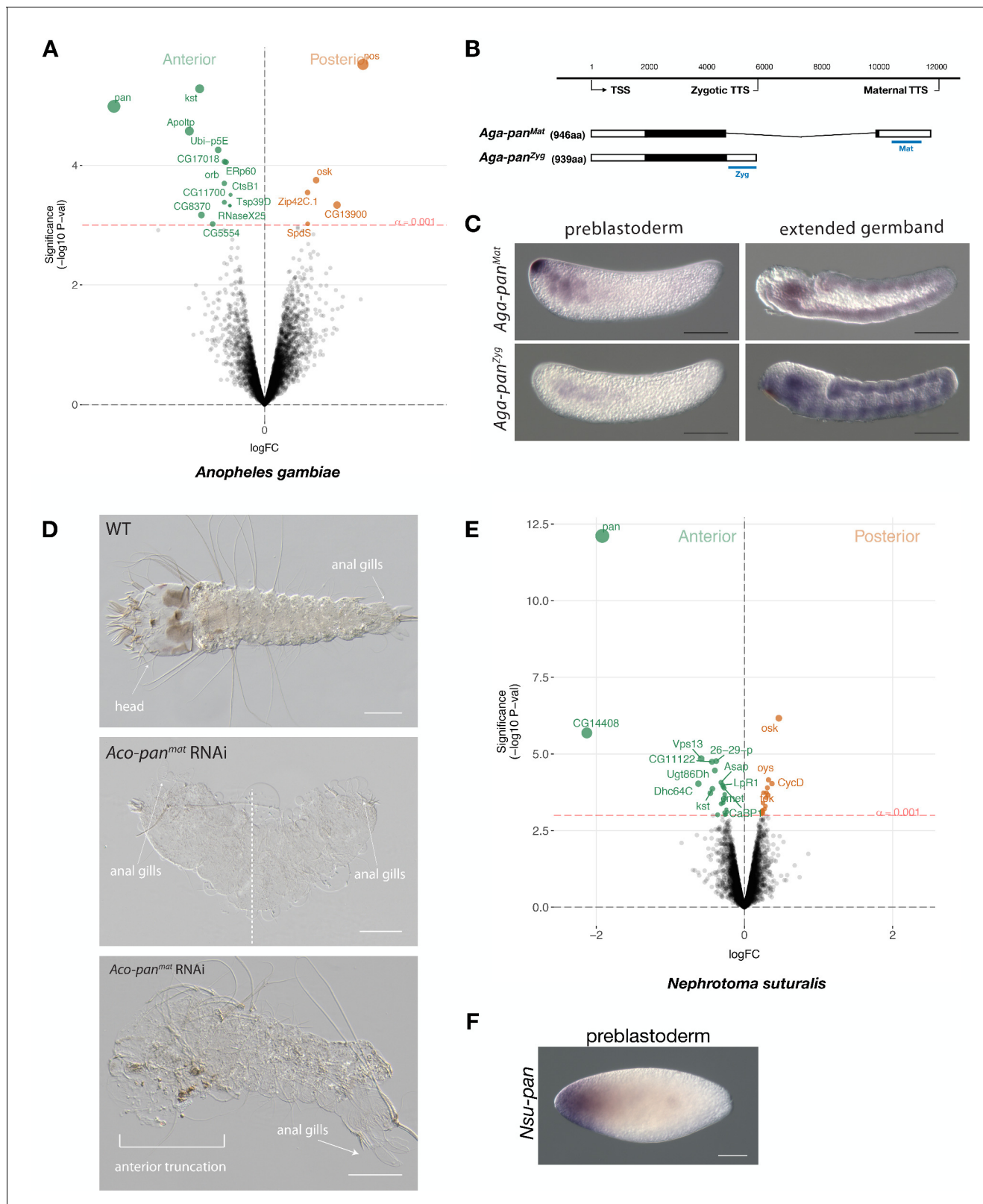


**Figure 5.** Expression and function *cucooid* in *Aedes*. (A) Differential expression analysis of maternal transcripts between anterior and posterior halves of 1 hr-old *Aedes* embryos.  $\log_{2} FC$ :  $\log_{2}$  fold-change. (B) Sketches of *Aae-cucooid* transcript isoforms based on RNA-seq and RACE experiments with the position of the in situ hybridization probe and dsRNA underlined in blue (see also legend to **Figure 1A**). (C) RNA in situ hybridization *Aae-cucooid* (left) and *Aae-nos* (right) transcript in 1 hr-old preblastoderm embryo. Anterior is left and dorsal up. Scale bar: 100  $\mu$ m. (D) 1st instar *Aedes* larval cuticle of wild type (top) and following *Aae-cucooid* RNAi (bottom; 9/26 versus 0/22 with control dsRNA;  $p < 0.005$ ). Scale bar: 100  $\mu$ m.

DOI: <https://doi.org/10.7554/eLife.46711.016>

hybridizations with a probe against the *Anopheles gambiae* ortholog of *cucooid* did not detect any anterior localized transcript in 1 hr-old embryos, suggesting that *Anopheles* uses a different anterior determinant than *Culex* and *Aedes*.

To further test this possibility, we sequenced the anterior and posterior transcriptomes 1 hr-old preblastoderm embryos of *A. gambiae* and ranked 9353 transcripts according to the magnitude of their differential expression scores and P values. In the posterior embryo, the most enriched transcript was homologous to *nanos*. In the anterior embryo, the most enriched transcript was homologous to *pangolin* (also known as *Tcf*) (**Figure 6A**). To test for potential alternative maternal and zygotic isoforms of *pangolin* in *Anopheles*, we mapped the assembled transcripts and 5' and 3' RACE products from 1 to 6 hr-old embryos onto an available *A. gambiae* genome assembly (AgamP4). We identified two alternative transcript variants with non-overlapping 3'UTRs but nearly identical open reading frames that we named *Aga-pan<sup>Mat</sup>* and *Aga-pan<sup>Zyg</sup>*, respectively (**Figure 6B** and **Figure 6—figure supplement 1**). *Aga-pan<sup>Mat</sup>* was tightly localized at the anterior pole of 1–2



**Figure 6.** Expression and function of alternative *pangolin* isoforms in *Anopheles*. (A) Differential expression analysis of maternal transcripts between anterior and posterior halves of 1 hr-old *Anopheles* embryos. logFC: log fold-change. (B) Sketches of *Aga-pan<sup>Mat</sup>* and *Aga-pan<sup>Zyg</sup>* transcripts based on RNA-seq and RACE experiments (see also legend to **Figure 1A**). Blue lines the position of in situ hybridization probes and dsRNA. (C) RNA in situ hybridization of *Aga-pan<sup>Mat</sup>* and *Aga-pan<sup>Zyg</sup>* transcripts in 1 hr-old preblastoderm and germband extending embryos of *A.gambiae*. Anterior is left and Figure 6 continued on next page

Figure 6 continued

dorsal up. Scale bar: 100µm. (D) 1 st instar *A.coluzzii* larval cuticle of wild type (top). Cuticle phenotypes of double abdomen (middle; 3/37) and intermediate, anterior truncation phenotype (bottom; 8/37) following *Aco-pan<sup>Mat</sup>* RNAi. Scale bar: 100µm. (E) Differential expression analysis of maternal transcripts between anterior and posterior halves of 1 hr-old *Nephrotoma* (Tipulidae) embryos. logFC: log fold-change. (F) RNA in situ hybridization of *Nsu-pan<sup>Mat</sup>*.

DOI: <https://doi.org/10.7554/eLife.46711.017>

The following figure supplements are available for figure 6:

**Figure supplement 1.** Protein alignment of dipteran Pangolin orthologs and Panish.

DOI: <https://doi.org/10.7554/eLife.46711.018>

**Figure supplement 2.** RNA in situ hybridization of *Aco-pan<sup>Mat</sup>* and *Aco-pan<sup>Zyg</sup>* transcripts in 1 hr-old *A.coluzzii* preblastoderm embryos.

DOI: <https://doi.org/10.7554/eLife.46711.019>

**Figure supplement 3.** Stage-specific RNA-seq read coverage of *Ast-pan* genomic locus and sketches of *Ast-pan<sup>Mat</sup>* and *Ast-pan<sup>Zyg</sup>* transcripts.

DOI: <https://doi.org/10.7554/eLife.46711.020>

**Figure supplement 4.** Cuticle phenotype of a *A.coluzzii* 1 st instar larva following *Aco-pan<sup>Zyg</sup>* RNAi (11/23).

DOI: <https://doi.org/10.7554/eLife.46711.021>

**Figure supplement 5.** *panish* RNAi rescue experiments with *Cri-pan* mRNA.

DOI: <https://doi.org/10.7554/eLife.46711.022>

hr-old preblastoderm embryos and only weakly expressed in elongated germbands, whereas *Aga-pan<sup>Zyg</sup>* was expressed segmentally in elongated germbands but not in embryos younger than 2 hr-old preblastoderm embryos (**Figure 6C**). Both *pangolin* isoforms were conserved in *Anopheles coluzzii* with sequence identity above 99% and the maternal isoform (*Aco-pan<sup>Mat</sup>*) was localized at the anterior pole (**Figure 6—figure supplement 2**). The stage-specific expression of both *pangolin* isoforms was also conserved outside the *Anopheles gambiae* species complex in *Anopheles stephensi* (**Figure 6—figure supplement 3**). Alignments of dipteran Pangolin proteins suggest that, in *Anopheles*, the maternal variant includes an additional seven amino acids at C-terminal end due to alternative polyadenylation and splicing (**Figure 6—figure supplement 1**).

To investigate the function of localized maternal *pangolin* expression in *Anopheles*, we specifically targeted this isoform in *Anopheles coluzzii*. Injection of *Aco-pan<sup>Mat</sup>*-specific dsRNA into several hundred 1 hr-old *A. coluzzii* embryos resulted in only 37 cuticles with variable phenotypes, including anterior truncations and, in extreme cases, double abdomens (**Figure 6D**). We noticed perturbed segmentation boundaries in the double abdomens, suggesting that *Aga-pan<sup>Mat</sup>* may also function in segmentation, as suggested by its weak zygotic expression pattern. Injection of *Aco-pan<sup>Zyg</sup>*-specific dsRNA into 1 hr-old *A. coluzzii* embryos resulted in severe segmentation defects that were difficult to characterize, but no double abdomens or anterior-specific truncation defects were found (**Figure 6—figure supplement 4**). Taken together with the isoform-specific transcript localization data presented above, these RNAi results support the hypothesis that *pangolin* acquired the anterior determinant function via the localization of a maternal transcript isoform with an alternative 3' end.

### Localization of maternal *pangolin* transcript in crane flies suggests that *pangolin* functioned as anterior determinant in ancestral flies

Anterior-localized maternal *pangolin* (*Tc-pan*) transcript has also been observed in the eggs of a beetle (*Tribolium castaneum*) (**Bucher et al., 2005**), but the function of this transcript remains unknown. Previous *Tc-pan* RNAi experiments targeted both maternal and zygotic transcripts and only revealed a function in posterior development, due to the role of zygotic *Tc-pan* in canonical Wnt signaling in the posterior growth zone (**Bolognesi et al., 2008; Fu et al., 2012; Prühs et al., 2017; Ansari et al., 2018**). To test whether ancestral dipterans localized maternal *pangolin* transcript at the anterior pole of the egg, we established a culture of the crane fly *Nephrotoma suturalis* (Tipulidae), which belongs to the Tipulomorpha, one of the oldest branches of dipterans (**Grimaldi and Engel, 2005; Wiegmann et al., 2011**). We sequenced the anterior and posterior transcriptomes of freshly deposited *Nephrotoma* egg halves and ranked 5371 transcripts according to the magnitude of their differential expression scores and P values (**Figure 6E**). The most enriched transcript in the posterior embryo was related to *oskar*, suggesting that crane fly eggs contain maternal germ plasm at the posterior pole. The most enriched transcript in the anterior embryo was homologous to *pangolin* and therefore named *Nsu-pan*. The anterior localization of this transcript was confirmed by RNA in

situ hybridization (**Figure 6F**). RACE experiments with cDNA from 1 hr-old embryos identified multiple isoforms with slightly variable 5' ends but the same open reading frame. An alignment of the predicted Nsu-Pan protein from this open reading frame with other dipteran Pangolin homologs revealed conserved N-terminal and C-terminal ends in Nsu-Pan. Taken together with our *Anopheles* data, our results in *Nephrotoma* suggest that ancestral dipteran insects localized maternal *pangolin* transcript in the anterior egg pole, where this transcript may have functioned as anterior determinant.

### Pangolin cannot substitute for Panish in Chironomus

In the midge *Chironomus*, the ortholog of *pangolin* (*Cri-pan*) is not expressed maternally but its diverged paralog *panish* functions maternally as anterior determinant (*Klomp et al., 2015*). Given that *panish* evolved from *pangolin* via gene duplication, *panish* probably inherited its role from *pangolin*. Therefore, it is possible that *Cri-pan* and *panish* are still functionally equivalent when expressed at the anterior pole of preblastoderm *Chironomus* embryos. Alternatively, *Panish* may have co-evolved with the targets required for anterior patterning and *Pangolin* can no longer interact with those targets. In this case, *Cri-pan* should no longer be able to fulfill the function of *panish*. To distinguish between these possibilities, we examined the ability of *panish* and *Cri-pan* mRNAs to rescue the RNAi phenotype of *panish*. We have previously shown that dsRNA of the *panish* 3'UTR can induce the double abdomen phenotype with a penetrance of nearly 100%, and that this phenotype can be rescued in roughly half of the embryos by injecting *panish* mRNA with heterologous UTRs at the anterior pole, shortly after the injection of dsRNA (*Klomp et al., 2015*). We used this assay to compare the functions of *panish* mRNA (positive control), frame shifted *panish* mRNA (negative control), *Cri-pan* mRNA, and a modified *Cri-pan* mRNA designed to better resemble *panish* mRNA (*Cri-pan<sup>trunc.</sup>* mRNA). *Cri-pan<sup>trunc.</sup>* mRNA encodes a N-terminal truncated Cri-Pan variant, lacking the  $\beta$ -Catenin binding and HMG box domains, with two mutations in the cysteine-clamp domain to mimic conserved changes of the *Panish* cysteine-clamp (**Figure 6—figure supplement 5A**). Only *panish* mRNA rescued *panish* RNAi embryos (**Figure 6—figure supplement 5B**), suggesting that *panish* co-evolved with its targets and functionally diverged after its origin via gene duplication from *pangolin*.

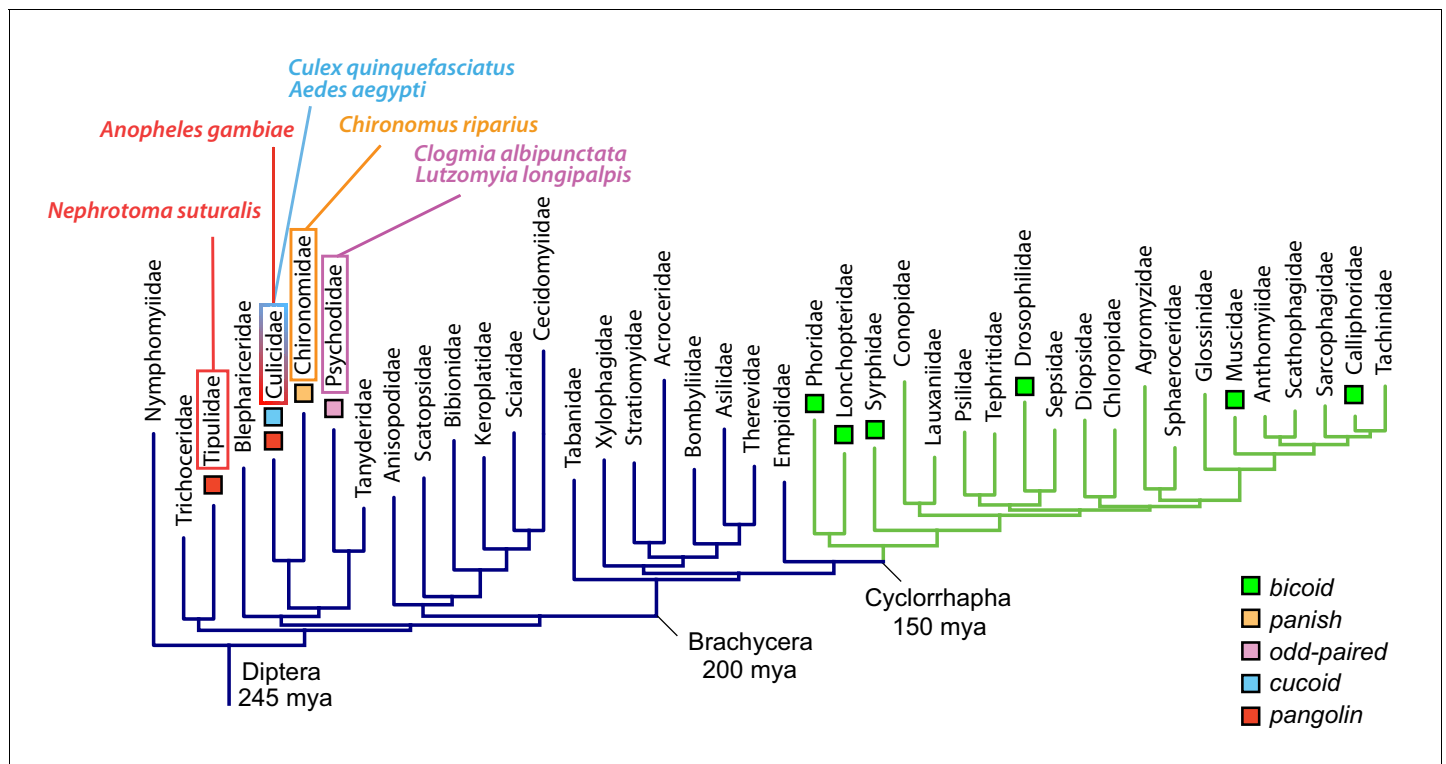
## Discussion

### Role of alternative transcription in the evolution of embryonic axis determinants from old genes

In this study, we have identified three unrelated old genes that encode the anterior determinant in moth flies, culicine mosquitoes, and anopheline mosquitoes, respectively (**Figure 7**). All three genes not only localize their maternal transcript at the anterior egg pole; they also are subject to alternative transcription, which allows a single gene to generate multiple transcript isoforms with distinct 5' and 3' ends through the use of alternative promoters (alternative transcription initiation) and polyadenylation signals (alternative transcription termination).

In moth flies, the localized maternal *odd-paired* transcript that functions as anterior determinant has an alternative first exon compared to the canonical isoform (**Figure 1B–D**, **Figure 2**, and **Figure 3A–B**). In the case of mosquitoes, maternal transcript isoforms of *cucoïd* (in culicine mosquitoes) or *pangolin* (in anopheline mosquitoes) with alternative last exons are localized at the anterior pole of the egg and function as anterior determinant (**Figure 4**, **Figure 5**, and **Figure 6**). Since anterior determinants are localized in the anterior egg, and signals for the subcellular localization of transcripts are typically found in UTRs (*Holt and Bullock, 2009*), it is possible that alternative transcription facilitates the evolution of anterior determinants by providing the UTR sequence for isoform-specific localization signals that do not interfere with other gene functions. For example, it has been shown that alternative last exons of transcript isoforms confer isoform-specific localization in neurons (*Taliaferro et al., 2016*; *Ciulli Mattioli et al., 2019*). Additional experiments will be needed to test whether the unique UTR sequences of anterior determinants are essential for their localization at the anterior egg pole.

In addition to changes in UTR sequences, alternative transcription also can result in the truncation or elongation of the open reading frame. For example, the anterior determinant of *Clogmia* (Cal-



**Figure 7.** Anterior axis determinants in Diptera. The phylogenetic tree of dipteran families is based on published data and shows Cyclorrhapha in green (Wiegmann et al., 2011). Mya, million years ago.

DOI: <https://doi.org/10.7554/eLife.46711.023>

The following figure supplements are available for figure 7:

**Figure supplement 1.** Occurrence of *panish* in Chironomidae genomes.

DOI: <https://doi.org/10.7554/eLife.46711.024>

**Figure supplement 2.** Maternal and zygotic transcript variant of *Clogmia zerknullt* (*Cal-zen*).

DOI: <https://doi.org/10.7554/eLife.46711.025>

$Opa^{Mat}$ ) lacks the N-terminal 20 amino acids (Figure 1—figure supplement 1), and the anterior determinant of *Anopheles* ( $Aga-pan^{Mat}$ ) encodes protein that includes additional seven amino acids at the C-terminal end (Figure 6—figure supplement 1). However, these changes to the protein may not have been important for adopting a function as anterior determinant. The truncation in the maternal Odd-paired protein that we observed in *Clogmia* is not conserved in *Lutzomyia*, in which  $Llo-opa^{Mat}$  and  $Llo-opa^{Zyg}$  encode the same protein, and full-length Odd-paired homologs from these and other species can function as anterior determinant in *Clogmia* (Figure 3B–C). Also, the elongation of  $Aga-Pan^{Mat}$  protein is not conserved in *Nephrotoma*, in which the localized  $Nsu-pan^{Mat}$  transcript encodes a Pangolin protein with a conserved C-terminal end (Figure 6—figure supplement 1). Therefore, modifications in the open reading frame of these genes may reflect secondary changes.

### Evolution of new genes that encode embryonic axis determinants

Unlike the anterior determinants identified in this study, the previously described anterior determinants of *Drosophila* and *Chironomus* are encoded by newly evolved genes, *bicoid* and *panish*. These genes seem to be dispensable outside the context of axis specification (Driever, 1993; Klomp et al., 2015), suggesting that they evolved specifically for this function. They could have acquired their function de novo via protein evolution or via inheritance from the progenitor gene. Our findings suggest that the role of *pangolin* in axis specification was already present in ancestral dipterans (Figure 7). We therefore propose that *panish*, which evolved from *pangolin* via gene duplication in the Chironominae lineage (Figure 7—figure supplement 1), inherited its function from

*pangolin*. Future examinations of *pangolin* isoforms and their expression in the eggs of chironomids that lack *panish* orthologs (species representing basal chironomid lineages) could reveal intermediate steps in this process, such as a localized truncated *pangolin* isoform.

Similarly, *bicoid* could have acquired its function de novo via protein evolution or via inheritance from its progenitor gene, *zerknüllt*. Several previous studies have hypothesized that Bicoid replaced Orthodenticle, a conserved homeodomain protein with similar DNA-binding affinity that functions in animal head development (Wimmer et al., 2000; Schröder, 2003; Lynch et al., 2006; Datta et al., 2018). Ancestrally reconstructed homeodomains confirmed that a single amino acid change in the homeodomain of Bicoid (Q50K), which is shared by Bicoid and Orthodenticle, caused a dramatic shift of Bicoid's DNA-binding affinity in vitro and target recognition in vivo (Liu et al., 2018).

We cannot rule out that *orthodenticle* functioned as anterior determinant in ancestral brachyceran flies. However, in analogy with our findings, it is also possible that a maternal *zerknüllt* isoform became localized at the anterior pole of the egg and acquired the role of anterior determinant via co-option, prior to the origin of *bicoid* via gene duplication (Figure 7—figure supplement 2). Maternal *zerknüllt* expression is common in lower Diptera but was lost in Cyclorrhapha (Stauber et al., 2002). If *bicoid* inherited its function from *zerknüllt*, the Q50K mutation in the homeodomain of Bicoid must have been a secondary, potentially maladaptive change. In this case, it may have been fixed in the cyclorrhaphan stem lineage via a compensatory or balancing mechanism and would have driven co-evolution of its targets.

It may be objected that, in *Drosophila*, reverting the K50 residue of the Bicoid homeodomain to Q50 is lethal and results in a *bicoid* null phenotype (Liu et al., 2018). However, how the ancestral gene network responded to the Q50K mutation of Bicoid cannot be inferred from observations in *Drosophila*. Moreover, published biochemical data suggest that the Q50K mutation increases interaction with the consensus Bicoid binding DNA motif much stronger than it reduces interaction with the consensus *Zerknüllt* binding DNA motif. It is therefore conceivable that the Q50K mutation had a less dramatic effect in ancestral flies, in which the target genes of the anterior determinant were activated via *Zerknüllt* binding sites, than in *Drosophila*, in which the target genes of the anterior determinant are activated via Bicoid binding sites. Examination of the mechanisms that determine embryo polarity in non-cyclorrhaphan Brachycera flies might help to test this hypothesis. If *panish* and *bicoid* inherited their functions, their evolutionary origin and divergence could have served the purpose of reducing the pleiotropy of their progenitor genes, *pangolin* and *zerknüllt*, rather than allowing them to take on an entirely new function in development.

## Role of alternative transcription in the evolution

Recent genome-wide analyses have shown that alternative transcription is a widespread phenomenon. For example, there are on average four alternative transcription start sites per gene in humans (Forrest et al., 2014) and at least 50–70% of mammalian genes are subject to alternative polyadenylation (Shepard et al., 2011; Adams et al., 2000; Derti et al., 2012). Alternative transcript isoforms can be tightly regulated in a cell or tissue specific manner and can affect transcription and translation efficiency as well as splicing (An et al., 2008; Davuluri et al., 2008; Lau et al., 2010; Pinto et al., 2011; Smibert et al., 2012; Anvar et al., 2018; Tushev et al., 2018; Taliaferro et al., 2016; Ciolli Mattioli et al., 2019; Shabalina et al., 2010). Functional studies in model organisms have shown that alternative transcription can generate dominant negative and alternatively localized protein isoforms (Davuluri et al., 2008; Bharti et al., 2008; Vacik et al., 2011; Berkovits and Mayr, 2015; Schrankel et al., 2016), while misregulation of alternative transcript isoforms has been associated with human diseases including cancer (Mayr and Bartel, 2009; Wiesner et al., 2015; Pal et al., 2012; Shapiro et al., 2011). However, the contribution of alternative promoters (alternative transcription initiation) and polyadenylation signals (alternative transcription termination) to the evolution of new gene functions and regulatory networks remains poorly understood (Carroll et al., 2005; Peter and Davidson, 2011; Wittkopp and Kalay, 2011; de Klerk and 't Hoen, 2015).

The results of several previous studies suggest that alternative transcription may underlie the evolutionary diversification of gene functions. For example, a large fraction of alternative promoter sequences is conserved between human and mice, but those with cell or tissue restricted expression have frequently changed during mammalian evolution (Baek et al., 2007; Forrest et al., 2014), suggesting that alternative promoters may have played a significant role in cell type evolution. Another likely substrate for evolutionary diversification are the protein terminal ends generated by alternative



transcription in conjunction with alternative splicing. These terminal ends commonly contain intrinsically disordered regions which are enriched in sites that mediate protein-protein interactions (Buljan et al., 2013; Shabalina et al., 2014). A recent case study found that a light fur color variant of beach mice evolved repeatedly via selection for an alternative *agouti* isoform with increased translocation efficiency (Linnen et al., 2013; Mallarino et al., 2017).

Our in vivo study revealed three old genes that evolved the anterior determinant function by localizing an alternative transcript isoform at the anterior pole of the egg. Therefore, we propose that differential expression of alternative transcript isoforms can result in the evolution of new gene functions, independent of, and prior to gene duplication and sub-functionalization. Given that alternative transcription is a widespread phenomenon, it could play an important role in the evolution of gene regulatory networks.

## Materials and methods

### Key resources table

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Strain, strain background ( <i>Nephrotoma suturalis</i> )	Nsu	<a href="https://doi.org/10.1083/jcb.25.1.95">https://doi.org/10.1083/jcb.25.1.95</a>	NA	
Strain, strain background ( <i>Chironomus riparius</i> )	Cri	<a href="https://doi.org/10.1126/science.aaa7105">https://doi.org/10.1126/science.aaa7105</a> ; <a href="https://doi.org/10.1111/mec.1411">https://doi.org/10.1111/mec.1411</a>	CRIP_Laufer	
Strain, strain background ( <i>Clogmia albipunctata</i> )	Cal	<a href="https://doi.org/10.1007/s004270050238">https://doi.org/10.1007/s004270050238</a>	NA	
Strain, strain background ( <i>Culex quinquefasciatus</i> )	Cqu	NIAID	NA	
Strain, strain background ( <i>Anopheles gambiae</i> )	A.gambiae	NIAID	G-3 strain	
Strain, strain background ( <i>Anopheles coluzzii</i> )	A.coluzzii	Insect transformation facility, University of Maryland	NA	
Strain, strain background ( <i>Aedes aegypti</i> )	Aae	NIAID	Liverpool 'Black eye'	
Gene ( <i>Clogmia albipunctata</i> )	<i>Cal-opa<sup>Mat</sup></i>	This paper	MN122104	See Materials and methods - Cloning procedures and mRNA/dsRNA synthesis; RNA in situ hybridization; Rapid Amplification of cDNA Ends (RACE)
Gene ( <i>Clogmia albipunctata</i> )	<i>Cal-opa<sup>Zyg</sup></i>	This paper	MN122105	See Materials and methods - Cloning procedures and mRNA/dsRNA synthesis; RNA in situ hybridization; Rapid Amplification of cDNA Ends (RACE)

Continued on next page

Continued

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Gene ( <i>Clogmia albipunctata</i> )	<i>Cal-slp</i>	This paper	MN122106	See Materials and methods - Cloning procedures and mRNA/dsRNA synthesis; RNA in situ hybridization
Gene ( <i>Clogmia albipunctata</i> )	<i>Cal-mira</i>	This paper	MN122107	See Materials and methods - Cloning procedures and mRNA/dsRNA synthesis; RNA in situ hybridization
Gene ( <i>Clogmia albipunctata</i> )	<i>Cal-otd</i>	This paper	MN122108	See Materials and methods - RNA in situ hybridization
Gene ( <i>Clogmia albipunctata</i> )	<i>Cal-nos1</i>	This paper	MN122109	See Materials and methods - RNA in situ hybridization
Gene ( <i>Clogmia albipunctata</i> )	<i>Cal-nos2</i>	This paper	MN122110	See Materials and methods - RNA in situ hybridization
Gene ( <i>Clogmia albipunctata</i> )	<i>Cal-nos3</i>	This paper	MN122111	See Materials and methods - RNA in situ hybridization
Gene ( <i>Clogmia albipunctata</i> )	<i>Cal-nos4</i>	This paper	MN122112	See Materials and methods - RNA in situ hybridization
Gene ( <i>Clogmia albipunctata</i> )	<i>Cal-vas</i>	This paper	MN122113	See Materials and methods - RNA in situ hybridization
Gene ( <i>Clogmia albipunctata</i> )	<i>Cal-tud</i>	This paper	MN122114	See Materials and methods - RNA in situ hybridization
Gene ( <i>Clogmia albipunctata</i> )	<i>Cal-gcl</i>	This paper	MN122115	See Materials and methods - RNA in situ hybridization
Gene ( <i>Lutzomyia longipalpis</i> )	<i>Llo-opa<sup>Mat</sup></i>	This paper	MN122116	See Materials and methods - Rapid Amplification of cDNA Ends (RACE)
Gene ( <i>Lutzomyia longipalpis</i> )	<i>Llo-opa<sup>Zyg</sup></i>	This paper	MN122117	See Materials and methods - Rapid Amplification of cDNA Ends (RACE)
Gene ( <i>Lutzomyia longipalpis</i> )	<i>Llo-osk</i>	This paper	MN122118	See Materials and methods - RNA in situ hybridization
Gene ( <i>Chironomus riparius</i> )	<i>Cri-opa</i>	This paper	MN122119	See Materials and methods - RNA in situ hybridization
Gene ( <i>Culex quinquefasciatus</i> )	<i>cucoi<sup>dA</sup></i>	This paper	MN122120	See Materials and methods - RNA in situ hybridization
Gene ( <i>Culex quinquefasciatus</i> )	<i>cucoi<sup>dB</sup></i>	This paper	MN122121	See Materials and methods - RNA in situ hybridization

Continued on next page

Continued

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Gene ( <i>Culex quinquefasciatus</i> )	<i>cucoid<sup>C</sup></i>	This paper	MN122122	See Materials and methods - Cloning procedures and mRNA/dsRNA synthesis; RNA in situ hybridization
Gene ( <i>Culex quinquefasciatus</i> )	<i>Cqu-nos</i>	This paper	MN122123	See Materials and methods - RNA in situ hybridization
Gene ( <i>Aedes aegypti</i> )	<i>Aae-cucoid<sup>A</sup></i>	This paper	MN122124	See Materials and methods - Rapid Amplification of cDNA Ends (RACE)
Gene ( <i>Aedes aegypti</i> )	<i>Aae-cucoid<sup>B</sup></i>	This paper	MN122125	See Materials and methods - Rapid Amplification of cDNA Ends (RACE)
Gene ( <i>Aedes aegypti</i> )	<i>Aae-cucoid<sup>C</sup></i>	<a href="http://vectorbase.org">vectorbase.org</a>	AAEL013321	See Materials and methods - Rapid Amplification of cDNA Ends (RACE)
Gene ( <i>Aedes aegypti</i> )	<i>Aae-nos</i>	<a href="http://vectorbase.org">vectorbase.org</a>	AAEL012107	See Materials and methods - RNA in situ hybridization
Gene ( <i>Anopheles gambiae</i> )	<i>Aga-pan<sup>Mat</sup></i>	This paper	MN122126	See Materials and methods - Cloning procedures and mRNA/dsRNA synthesis; RNA in situ hybridization; Rapid Amplification of cDNA Ends (RACE)
Gene ( <i>Anopheles gambiae</i> )	<i>Aga-pan<sup>Zyg</sup></i>	This paper	MN122127	See Materials and methods - Cloning procedures and mRNA/dsRNA synthesis; RNA in situ hybridization; Rapid Amplification of cDNA Ends (RACE)
Gene ( <i>Anopheles gambiae</i> )	<i>Aga-cad</i>	This paper	MN122128	See Materials and methods - RNA in situ hybridization
Gene ( <i>Nephrotoma suturalis</i> )	<i>Nsu-pan</i>	This paper	MN122129	See Materials and methods - RNA in situ hybridization
Sequence-based reagent	dsRNAs	This paper		see Materials and methods - Cloning procedures and mRNA/dsRNA
Sequence-based reagent	RNA probes	This paper		See Materials and methods - RNA in situ hybridization
Commercial assay or kit	SMARTer RACE 5'/3' Kit	Takara	634858	
Commercial assay or kit	mMESSAGE mMACHINE SP6	Thermo Fisher	AM1340	

Continued on next page

Continued

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Commercial assay or kit	QuikChange Lightning Site-Directed Mutagenesis Kit	Agilent	210515	
Software, algorithm	Geneious	<a href="https://www.geneious.com">https://www.geneious.com</a>	RRID: SCR_010519	version 11.1.5
Software, algorithm	GraphPad Prism	<a href="https://www.graphpad.com">https://www.graphpad.com</a>	RRID: SCR_015807	version 1.4

## Cloning procedures and mRNA/dsRNA synthesis

Coding sequences from *Clogmia*, *Lutzomyia*, *Chironomus*, *Anopheles*, *Nephrotoma*, *Culex*, and *Aedes* were amplified from embryonic cDNA with primers constructed from RNA-seq data. Coding sequence of *odd-paired* was amplified from cDNA (FI01113) that was obtained from the BDGP Gold collection of the Drosophila Genomics Resource Center. Amplified cDNA was cloned into the expression vector pSP35T (Amaya et al., 1991), using In-Fusion HD Cloning Kit (Clontech), and PstI- or EcoRI-linearized vector was used for mRNA synthesis using mMACHINE SP6. *panish*, *Cri-pan*, and *Cri-pan<sup>trunc</sup>* mRNAs were synthesized from PCR template (containing T7 polymerase binding site) using mMACHINE T7ultra. Mutations in the open reading frame (*Cal-opa<sup>Zyg-Met21Leu</sup>*, *panish FS*, *Cri-pan<sup>trunc</sup>*) and in the kozak sequences of *Dme-opa* and *Cri-opa* were generated using QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent). Double-stranded RNA (dsRNA) was generated from PCR-amplified templates using embryonic cDNA and primers containing T7 polymerase binding sites as described (Klomp et al., 2015).

Forward and reverse primer sequences for generating templates for mRNA synthesis were:

Cal-*opa*<sup>Mat</sup>  
 5'-TAAGATGAGTCCGAATCACTTACTGGCC  
 5'-TTAATAGGCCGTCGCTGCACC  
 Cal-*opa*<sup>Zy9</sup>  
 5'-CAACATGATGATGAACGCTTTTATGGAA  
 5'-TTAATAGGCCGTCGCTGCACC  
 Cal-*opa*<sup>115-655</sup>  
 5'-ATGCTCTTCTCAAATCACTCTTCAGC  
 5'-TTAATAGGCCGTCGCTGCACC  
 Cal-*opa*<sup>182-655</sup>  
 5'-ATGAACCCGGGAACCTTGGG  
 5'-TTAATAGGCCGTCGCTGCACC  
 Cal-*opa*<sup>182-445</sup>  
 5'-ATGAACCCGGGAACCTTGGG  
 5'-TTACGCGGGATTGACTGACTATG  
 Cal-*opa*<sup>1-337+22</sup>  
 5'-CAACATGATGATGAACGCTTTTATGGAA  
 5'-TCACAAAATTTCACTGAATCCGTCAAAATATCACTAGA  
 Llo-*opa*  
 5'-AAAGATGATGATGAATGCATTTATGGACACAG  
 5'-TCAGTACGCCGTGGCGGCG  
 Dme-*opa*<sup>Dme kozak</sup>  
 5'-GACCATGATGATGAACGCCTTCA  
 5'-GTCAATACGCCGTCGCTGCGCCGGG  
 Cri-*opa*<sup>Cri kozak</sup>  
 5'-AAAAATGATGATGAATGGTTTTATGGACACA  
 5'-TCAATAAGCTGTCGTTGGACCGTGAT  
 Cri-*pan*<sup>trunc</sup>  
 5'-AAAAATGTATCCAGATTGGAGCTCGC  
 5'-TTACGTCACACTAATAGCATTTCATCATCCC

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

*Cal-opa*<sup>Mat</sup> (222 bp):

5'-CAGAGATGCATAATACGACTCACTATAGGGAGAAAACAATTGTGAAGTGCAGACA  
5'-CAGAGATGCATAATACGACTCACTATAGGGAGACAAATTTCCAAACGATGACAGA

*Cal-opa*<sup>Zyg</sup> (315 bp):

5'- CAGAGATGCATAATACGACTCACTATAGGGAGAACTACCGCCGCGAACACACCG  
5'-CAGAGATGCATAATACGACTCACTATAGGGAGAGTCCAGTTCGATTCCATAAAAAGC

*Cal-slp* (927 bp):

5'- CAGAGATGCATAATACGACTCACTATAGGGAGATCGATCAGCTCCCTTTTGCC  
5'-CAGAGATGCATAATACGACTCACTATAGGGAGATGAGATCGTTCCTGGAC

*Cal-mira* (993 bp):

5'- CAGAGATGCATAATACGACTCACTATAGGGAGAACAGCAAAAAGGAAGCGAAA  
5'-CAGAGATGCATAATACGACTCACTATAGGGAGAGGGATTCAATTTGCCTTTGA

*Aga-pan*<sup>Mat</sup> (976 bp):

5'- CAGAGATGCATAATACGACTCACTATAGGGAGACACACAGGGCACAATAATCG  
5'- CAGAGATGCATAATACGACTCACTATAGGGAGAGACTGCATGTCCGTCTA

*Aga-pan*<sup>Zyg</sup> (843 bp):

5'- CAGAGATGCATAATACGACTCACTATAGGGAGAACATCACACACCCACACAC  
5'- CAGAGATGCATAATACGACTCACTATAGGGAGATTGGTCCGTTCGTATTGTA

*cucooid* (926 bp):

5' - CAGAGATGCATAATACGACTCACTATAGGGAGACGAGGATGTTGCTGGAGAAT  
5'- CAGAGATGCATAATACGACTCACTATAGGGAGAACTCCCGAAATCGGAAAAC

*Aae-cucooid* (857 bp):

5'-CAGAGATGCATAATACGACTCACTATAGGGAGACGACAAGCCCTACAAATGCT  
5'-CAGAGATGCATAATACGACTCACTATAGGGAGATGATCTGGATGTTGCCGTAG

## Microinjection of embryos

*Chironomus* embryo injection was done as previously described (Klomp et al., 2015). *Clogmia* eggs were dissected from ovaries and activated under water. Eggs of *Aedes*, *Culex*, and *Anopheles* were collected in a dark chamber on a moist filter paper for about 30 min (Fisher Scientific, Cat. No 09–795C). These eggs were transferred to another filter paper cut into 4 cm x 2 cm pieces and aligned perpendicularly to the edge of a cover glass (Fisher Scientific, Cat. No 12-648-5C) with the prospective injection side pointing towards the glass edge. We noticed that injecting eggs near the anterior or posterior pole was critical for survival of the procedure. During the alignment procedure, water was applied to the filter paper as needed to prevent eggs from desiccation. After aligning the eggs, the cover glass was removed and excess water on the filter paper was absorbed using filter paper. A second cover glass with a layer of double-sided tape (Scotch 3M) was slightly pressed against the aligned eggs to transfer the eggs to the double-sided tape. The embryos were then immediately covered with halocarbon oil to prevent desiccation. For cuticle preparations, the embryos were injected under halocarbon oil 27 (Sigma, MKBZ7202V). The oil was washed off under a gentle stream of water immediately after injection. In the case of *Clogmia*, *Aedes*, and *Culex* the cover glass was transferred to a moist chamber (petri dish with wet kimwipe paper) and kept at 28 °C, and water was added every day to prevent desiccation. In the case of *Anopheles*, the eggs were allowed to develop under water. Removal of the halocarbon oil was critical to ensure embryo survival until late developmental stages and hatching. For eggs to be fixed within a day following injection, we used a 1:1 mixture of halocarbon oil 27 and halocarbon oil 700 (Sigma, MKCB5817) and left the injected eggs immersed in the oil until fixation. Embryos were injected with quartz needles using a Narishige IM-300 microinjector. Quartz capillaries (Sutter Instruments Q100-70-10) were pulled with a Sutter instrument P-2000 laser-based micropipette puller. Our settings for the needle puller were: Heat 645, Fil 4, Vel 40, Del 125, Pul 130. Needles were back-filled and the tip was broken open at the time of injection by slightly touching the first egg.

## Embryo fixation

*Clogmia* Embryos were dechorionated using a 10% dilution of commercial bleach (8.25% sodium hypochloride) for 3 min. For *Nephrotoma* embryos, a 25% dilution was used for 3 min until the

chorion became slightly transparent. Embryos of *Aedes*, *Culex*, and *Anopheles* were dechorionated as described (*Juhn and James, 2012*). Dechorionated embryos were fixed in a 50 mL falcon tube, using 20 mL of boiling salt/detergent-solution (100  $\mu$ L 10% triton-X, 500  $\mu$ L 28% NaCl, up to 20 mL of water). After 10 s, water was applied to the tube to cool down the embryos. If needed, the embryos were devitellinized in a 1:1 mixture of n-heptane and methanol by gentle shaking. Embryos with vitelline membrane attached were further devitellinized using sharp tungsten needles in an agar plate covered with methanol. Devitellized embryos were stored in 100% methanol at  $-20^{\circ}\text{C}$ .

## RNA in situ hybridization

RNA in situ hybridizations were conducted as described (*Klomp et al., 2015*), using digoxigenin (DIG)-labeled probes and Fab fragments from anti-DIG antibodies conjugated with alkaline phosphatase (AP) (Roche, IN, USA). Probes were prepared from PCR templates, using sequence-specific forward primers and reverse primers with T7 promoter sequence (see above for *Cal-opa<sup>Mat</sup>*, *Cal-opa<sup>Zyg</sup>*, *Cal-cad*, *Cal-slp*, *Cal-mira*, *Aga-pan<sup>Mat</sup>*, *Aga-pan<sup>Zyg</sup>*, and *Aae-cuicoid*; gene specific sequence underlined).

*Cal-nos1* (450 bp):

5'-AGCACTTTTCCCCCAAGAGT

5'-CAGAGATGCATAATACGACTCACTATAGGGGAGAGGCATTCATATTTCTCAGCA

*Cal-nos2* (475 bp):

5'-AATTATTCTGTTCCAAAGTTGAGATT

5'-CAGAGATGCATAATACGACTCACTATAGGGGAGACCCCAGACTGGTGACAAAT

*Cal-nos3* (548 bp):

5'-TGAGTTAAATAGAGTGAAAACAGCAA

5'-CAGAGATGCATAATACGACTCACTATAGGGGAGATACCGTCTCGTGCTTAATCG

*Cal-nos4* (440 bp):

5'-GGCAAAATTTCCAAGTGAA

5'-CAGAGATGCATAATACGACTCACTATAGGGGAGACGTGTCCTCAAGCGTGTAGAT

*Cal-vas* (938 bp):

5'-CTGAGGCGAACTTGTGTGAA

5'-CAGAGATGCATAATACGACTCACTATAGGGGAGAATTGGCAATGTCCAGTCCTC

*Cal-tud* (921 bp):

5'-ATTCTGCAAGTCGTGAGGT

5'-CAGAGATGCATAATACGACTCACTATAGGGGAGACCTGTACCAGCCATTGTCCT

*Cal-gcl* (450 bp):

5'-GCAGAACCCCTTGGACATTA

5'-CAGAGATGCATAATACGACTCACTATAGGGGAGAGTAACGCCCAATTCGTCT

*cuicoid<sup>A</sup>* (939 bp):

5'-ACGATGAGGAGGAGGTTCT

5'-CAGAGATGCATAATACGACTCACTATAGGGGAGACGCACTTCACCGTGTGTAAC

*cuicoid<sup>B</sup>* (717 bp):

5'-GGGGCGACATCTATATCTCACT

5'-CAGAGATGCATAATACGACTCACTATAGGGGAGAACAGTGAGAAAAATCCCAACTTTAGT

*cuicoid<sup>C</sup>* (926 bp):

5'-CGAGGATGTTGCTGGAGAAT

5'-CAGAGATGCATAATACGACTCACTATAGGGGAGAACTCCCGAAATCGGAAAAT

*Cqu-cad* (956 bp):

5'-CACGTGTTCCATCAGTCCAG

5'-CAGAGATGCATAATACGACTCACTATAGGGGAGAATGAGGCTTAACGAGGATGG

*Cqu-nos* (927 bp):

5'-AAGTGCCGTGAATTTGTCC

5'-CAGAGATGCATAATACGACTCACTATAGGGGAGAGCGAAACCAATTCGACAGTT

*Nsu-pan* (966 bp):

5'-TCGCGGCAAGATCATAGTCC

5'-CAGAGATGCATAATACGACTCACTATAGGGGAGACCTGCAGGGTTTACACCACT

*Aae-nos* (914 bp):

5'-CAAACGTGAAGCGGAAGATT

5'-CAGAGATGCATAATACGACTCACTATAGGGGAGAATTACGTCCGGAAGTGTTCG

## Rapid Amplification of cDNA Ends (RACE)

Total RNA was phenol/chloroform extracted from *Clogmia* (1 hr-old and 9 hr-old embryos), *Anopheles* (1–6 hr-old embryos), *Culex* (0–7 hr old embryos), and *Nephrotoma* (1–29 hr-old embryos) fixed in TRIzol Reagent (Invitrogen) and precipitated with isopropanol. 5'/3' RACE was performed using SMARTer RACE 5'/3' Kit (Clontech) with the custom-made primers (including at the 5' end 15 nucleotides of pRACE vector sequence). Gene specific sequences are underlined.

*Cal-opa* 5'RACE primer: 5'-GATTACGCCAAGCTTCTGGGTGACGCCGTGGGCAAGGACGTCA

*Cal-opa* 3'RACE primer: 5'-GATTACGCCAAGCTTCGCGTCGATCGTCACGCCCCAAATTCCG

*Aga-pan* 5'RACE primer: 5'-GATTACGCCAAGCTTCGAATCTCCGGCCCGGAATTGAGACTT

*Aga-pan* 3'RACE primer: 5'-GATTACGCCAAGCTTAGCTTCACGCGACCAGCAAAACCAACGG

*cucoid* 5'RACE primer: 5'- GATTACGCCAAGCTTCGTGACGGCTTCGATGGTTGGTTTTCC

*cucoid* 3'RACE primer: 5'- GATTACGCCAAGCTTCGCACGTGTTGAACAGTCACATGTTGAC

*Aae-cucoid* 5'RACE primer: 5'- GATTACGCCAAGCTTGATCCGGTGGATCGGACTTGCCGAGAT  
T

*Aae-cucoid* 3'RACE primer: 5'- GATTACGCCAAGCTTAACCTCCCTCGGGGTTGAACGTGAAGC  
T

*Aae-cucoid* 5'RACE primer: 5'- GATTACGCCAAGCTTGATCCGGTGGATCGGACTTGCCGAGAT  
T

*Aae-cucoid* 3'RACE primer: 5'- GATTACGCCAAGCTTAACCTCCCTCGGGGTTGAACGTGAAGC  
T

*Nsu-pan* 5'RACE primer: 5'-GATTACGCCAAGCTTCTGGTCTGTGCGACGTTCTTCCAAATCG

*Nsu-pan* 3'RACE primer: 5'-GATTACGCCAAGCTTTCCCCTGGTGCAAATCCACGAGATGTG

## Cuticle preparations

Cuticles were prepped four to five days after injection. Eggshells was removed with tungsten needles and the embryos were transferred to a glass block dish with a drop of 1:4 glycerol/acetic acid. Following incubation in 1:4 glycerol/acetic acid overnight at room temperature, the cuticles were transferred onto a glass slide, oriented, mounted in 1:1 Hoyer's medium/lactic acid (*Stern and Sucena, 2000*), covered with a cover glass, and dried overnight at 65 °C.

## RNA-seq sample preparation and sequencing

Bisection of anterior and posterior embryo halves, RNA extraction, and sequencing were conducted as described (*Klomp et al., 2015*). In the case of *Clogmia*, anterior or posterior embryo halves from three 1 hr-old embryos were pooled and RNA-seq data were obtained from two replicates. In case of *Lutzomyia*, embryo halves from ten 1–2 hr-old embryos were pooled four replicates were generated. In case of *Anopheles* (G-3 strain), embryo halves from five 1 hr-old embryos were pooled and three replicates were generated. In the case of *Culex*, embryo halves from seven 1 hr-old embryos were pooled and three replicates were generated. In case of *Aedes* (Liverpool 'black eye' strain), embryo halves from five 1 hr-old embryos were pooled and four replicates were generated. In case of *Nephrotoma*, embryo halves from nine 1 hr-old embryos were pooled and three replicates were generated. Stage-specific *Clogmia* transcriptomes were generated from the offspring of a single mother and total RNA from five embryos was used for each stage. In the case of *Lutzomyia*, about 100 staged embryos were pooled for RNA extraction, and two independent RNA extractions from each time point were combined and submitted for sequencing.

Prior to library construction, RNA integrity, purity, and concentration were assessed using an Agilent 2100 Bioanalyzer with an RNA 6000 Nano Chip (Agilent Technologies, USA). Purification of messenger RNA (mRNA) was performed using the oligo-dT beads provided in the Illumina TruSeq mRNA RNA-SEQ kit (Illumina, USA). Complementary DNA (cDNA) libraries for Illumina sequencing were constructed using the Illumina TruSeq mRNA RNA-SEQ kit (Illumina, USA), using the manufacturer-specified protocol. Briefly, the mRNA was chemically fragmented and primed with random oligos for first strand cDNA synthesis. Second strand cDNA synthesis was then carried out with dUTPs to preserve strand orientation information. The double-stranded cDNA was then purified, end repaired, and 'a-tailed' for adaptor ligation. Following ligation, the samples were selected a final library size (adapters included) of 400–550 bp using sequential AMPure XP bead isolation (Beckman Coulter, USA). The libraries were sequenced in an Illumina HiSeq 4000 DNA sequencer, utilizing a pair end sequencing flow cell with a HiSeq Reagent Kit v4 (Illumina, USA).

## RNA-seq data preprocessing

The TrimGalore (*Krueger, 2012*) wrapper for Cutadapt (*Martin, 2011*) and FastQC (*Andrews, 2010*) was used to remove adapters and low quality sequences from raw fastq files. Overlapping reads were combined with Flash (*Magoč and Salzberg, 2011*) prior to assembly.

## Transcriptome assembly and annotation

Trinity 2.4.0 (*Grabherr et al., 2011*) on the Indiana University Karst high-performance computing cluster was used for assembling contiguous sequences (contigs) from the paired end (PE) sequence data of *Clogmia*, *Lutzomyia*, *Anopheles*, and *Nephrotoma*. ABySS 2.0 (*Jackman et al., 2017*) was used for assembling contigs from *Culex* and *Aedes* data. Only contigs of 200 nucleotides or greater were retained. BLAST+ tools (*Camacho et al., 2009*) were used to annotate contigs by conducting best-reciprocal-blast first against the *Drosophila melanogaster* transcriptome (BDGP6) peptide sequences (blastx/tblastn) and then the coding sequence (tblastx) with a maximum threshold evalue of 1e-10. Biomart and AnnotationDbi packages were used for gene ids and names. The longest open reading frames (ORFs) of unannotated transcripts were compared to the RefSeq invertebrate protein database (downloaded 4-1-2017) using blastp (max evalue 1e-10) followed by a similar comparison to remove transcripts with ORFs matching RefSeq plant, protozoan, archaea, bacteria, fungi, plasmid, or viral sequences (downloaded 6-1-2017). Remaining transcripts were designated by the top BLAST hit in *D. melanogaster*.

## Alignment and differential expression analysis

Cleaned paired-end read data was aligned and analyzed using R base (*Ihaka and Gentleman, 1996*) and Bioconductor (*Gentleman et al., 2004*) software packages. Sequence alignment was conducted with the seed-and-vote aligner, Subread, as implemented in the Rsubread package (*Liao et al., 2013*) with up to five multi-mapping locations, six mismatches, and 20 subreads/seeds per read. Sequence file manipulation, including sorting and indexing of '.bam' files, was done using Rsamtools (*Morgan et al., 2013*).

To avoid potential biases in transcript localization unrelated to anterior-posterior axis formation, transcripts annotated with mitochondrial, ribosomal, or ambiguous status (e.g., predicted, hypothetical, or uncharacterized) were filtered out prior to the differential expression comparisons. Transcripts with 20 or fewer counts in any of the A-P pairs were also excluded from the analysis prior to library normalization. Lower scoring, potentially related transcripts matching a given gene from the *D. melanogaster* transcriptome were retained for initial differential expression comparisons but removed for clarity of presentation in subsequent analyses and volcano plots. Trimmed mean of M-values (TMM) (*Robinson and Oshlack, 2010*) was used for normalization and EdgeR (*Robinson et al., 2010*) was used to perform quasi-likelihood F-tests between A-P samples, corrected for multiple testing using FDR (Benjamini-Hochberg). Following filtering based on annotation and detection of >20 counts per paired samples, we used the following number of transcripts for differential expression comparisons: 5602 for *Clogmia*; 5392 for *Lutzomyia*; 8239 for *Culex*; 5802 for *Aedes*; 9353 for *Anopheles*; 5371 for *Nephrotoma*.

## Mapping RNA-seq reads to genomic loci

RNA-seq reads from stage-specific transcriptomes were mapped to genomic scaffolds containing a gene of interest using TopHat RNA-seq aligner (*Trapnell et al., 2009*). Publicly available *Anopheles stephensi* transcriptomes used in this paper were: SRR515316, SRR515341, SRR514863, and SRR515304.

## Data availability

This project was deposited at the National Center for Biotechnology Information under Bioproject ID PRJNA454000 and the reads were deposited in the Short Reads Archives under accessions SRR7132661, SRR7132662, SRR7132659, SRR7132660, SRR7132665, SRR7132666, SRR7132663 and SRR7132664 for *Clogmia*, SRR7134470, SRR7134469, SRR7134472, SRR7134471, SRR7134468, and SRR7134467 for *Lutzomyia*, SRR8729860, SRR8729859, SRR8729858, SRR8729857, SRR8729856 and SRR8729855 for *Anopheles*, SRR8729854, SRR8729853, SRR8729852 and SRR8729851 for *Aedes*, SRR8729868, SRR8729867, SRR8729870, SRR8729869, SRR8729864 and SRR8729863 for *Culex* and



SRR8729866, SRR8729865, SRR8729872, SRR8729871, SRR8729861 and SRR8729862 for *Nephrotoma*. Transcript sequences are listed on the Key Resources Table.

## Acknowledgements

We thank Chun Wai Kwan, Claudia Vacca, and Nicole Horio for technical assistance, Arthur Forer (York University, Canada) for the *Nephrotoma* culture, Robert Harrell and Channa Aluvihare (Insect Transformation Facility, University of Maryland, MD) for shipping blood-fed mosquitoes, Vanessa Macias and Anthony James (University of California at Irvine, CA) for *Aedes* and *Culex* reagents, Molly Duman Scheel (Indiana University, IN) for logistic support, and Edwin L Ferguson (University of Chicago, IL) and M Feder (University of Chicago, IL) for laboratory equipment. EL Ferguson provided detailed comments on a manuscript draft. This work was supported by funds from the National Science Foundation (IOS-1355057), the National Institute of General Medical Science (5R01GM127366-02), the University of Chicago, and the National Center for Advancing Translational Sciences of the National Institutes of Health (UL1 TR000430) to U S-O, and the Intramural Program of the National Institute of Allergy and Infectious Diseases to J R YY was the recipient of an award of University of Chicago Henry Hinds Funds for Graduate Student Research in Evolutionary Biology. Transcriptomic data are available at the National Center for Biotechnology Information Sequence Read Archive (Bio-project ID PRJNA454000). This work utilized the computational resources of the NIH HPC Biowulf cluster.

## Additional information

### Funding

Funder	Grant reference number	Author
National Science Foundation	IOS-1355057	Urs Schmidt-Ott
National Institute of General Medical Sciences	R01 GM127366-01A1	Urs Schmidt-Ott
National Center for Advancing Translational Sciences	UL1 TR000430	Urs Schmidt-Ott
University of Chicago	Institutional fund	Urs Schmidt-Ott
National Institute of Allergy and Infectious Diseases	Intramural Program	Jose Ribeiro
University of Chicago	Henry Hinds Funds for Graduate Student Research in Evolutionary Biology	Yoseop Yoon

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

### Author contributions

Yoseop Yoon, Conceptualization, Formal analysis, Funding acquisition, Investigation, Visualization, Writing—original draft, Writing—review and editing; Jeff Klomp, Conceptualization, Formal analysis, Investigation, Visualization, Writing—original draft, Writing—review and editing; Ines Martin-Martin, Formal analysis, Investigation, Visualization; Frank Criscione, Investigation; Eric Calvo, Supervision; Jose Ribeiro, Resources, Formal analysis, Supervision, Funding acquisition; Urs Schmidt-Ott, Conceptualization, Supervision, Funding acquisition, Writing—original draft, Project administration, Writing—review and editing

### Author ORCIDs

Yoseop Yoon  <https://orcid.org/0000-0002-6062-0979>

Urs Schmidt-Ott  <https://orcid.org/0000-0002-1351-9472>

### Decision letter and Author response

Decision letter <https://doi.org/10.7554/eLife.46711.038>

Author response <https://doi.org/10.7554/eLife.46711.039>

## Additional files

### Supplementary files

- Supplementary file 1. Annotation and quantitation of *Clogmia* transcriptome. Gene names are those indicated as best-reciprocal-blast hits (see Methods). Contig names correspond to transcriptome assembly. Read counts are given based on non-unique mapping of pre-processed RNA-seq data to annotated transcriptome.

DOI: <https://doi.org/10.7554/eLife.46711.026>

- Supplementary file 2. Annotation and quantitation of *Lutzomyia* transcriptome. Gene names are those indicated as best-reciprocal-blast hits (see Methods). Contig names correspond to transcriptome assembly. Read counts are given based on non-unique mapping of pre-processed RNA-seq data to annotated transcriptome.

DOI: <https://doi.org/10.7554/eLife.46711.027>

- Supplementary file 3. Annotation and quantitation of *Culex* transcriptome. Gene names are those indicated as best-reciprocal-blast hits (see Methods). Contig names correspond to transcriptome assembly. Read counts are given based on non-unique mapping of pre-processed RNA-seq data to annotated transcriptome.

DOI: <https://doi.org/10.7554/eLife.46711.028>

- Supplementary file 4. Annotation and quantitation of *Aedes* transcriptome. Gene names are those indicated as best-reciprocal-blast hits (see Methods). Contig names correspond to transcriptome assembly. Read counts are given based on non-unique mapping of pre-processed RNA-seq data to annotated transcriptome.

DOI: <https://doi.org/10.7554/eLife.46711.029>

- Supplementary file 5. Annotation and quantitation of *Anopheles* transcriptome. Gene names are those indicated as best-reciprocal-blast hits (see Methods). Contig names correspond to transcriptome assembly. Read counts are given based on non-unique mapping of pre-processed RNA-seq data to annotated transcriptome.

DOI: <https://doi.org/10.7554/eLife.46711.030>

- Supplementary file 6. Annotation and quantitation of *Nephrotoma* transcriptome. Gene names are those indicated as best-reciprocal-blast hits (see Methods). Contig names correspond to transcriptome assembly. Read counts are given based on non-unique mapping of pre-processed RNA-seq data to annotated transcriptome.

DOI: <https://doi.org/10.7554/eLife.46711.031>

- Transparent reporting form

DOI: <https://doi.org/10.7554/eLife.46711.032>

### Data availability

Sequencing data have been deposited at the National Center for Biotechnology Information Sequence Read Archive (Bioproject ID PRJNA454000).

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Yoon Y, Klomp J, Martin-Martin I, Criscione F, Calvo E, Ribeiro J, Schmidt-Ott U	2018	Evolution of an Embryonic Axis Determinant via Alternative Transcription	<a href="https://www.ncbi.nlm.nih.gov/bioproject/PRJNA454000/">https://www.ncbi.nlm.nih.gov/bioproject/PRJNA454000/</a>	NCBI Bioproject, PRJNA454000

The following previously published dataset was used:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Xiaofang Jiang	2012	Anopheles stephensi strain:Indian	<a href="https://www.ncbi.nlm.nih.gov/bioproject/PRJNA454000/">https://www.ncbi.nlm.nih.gov/bioproject/PRJNA454000/</a>	NCBI Bioproject,

## References

- Adams MD**, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, George RA, Lewis SE, Richards S, Ashburner M, Henderson SN, Sutton GG, Wortman JR, Yandell MD, Zhang Q, Chen LX, et al. 2000. The genome sequence of *Drosophila melanogaster*. *Science* **287**:2185–2195. DOI: <https://doi.org/10.1126/science.287.5461.2185>, PMID: 10731132
- Akbari OS**, Antoshechkin I, Amrhein H, Williams B, Diloreto R, Sandler J, Hay BA. 2013. The developmental transcriptome of the mosquito *Aedes aegypti*, an invasive species and major arbovirus vector. *G3: Genes/Genomes/Genetics* **3**:1493–1509. DOI: <https://doi.org/10.1534/g3.113.006742>, PMID: 23833213
- Amaya E**, Musci TJ, Kirschner MW. 1991. Expression of a dominant negative mutant of the FGF receptor disrupts mesoderm formation in xenopus embryos. *Cell* **66**:257–270. DOI: [https://doi.org/10.1016/0092-8674\(91\)90616-7](https://doi.org/10.1016/0092-8674(91)90616-7), PMID: 1649700
- An JJ**, Gharami K, Liao GY, Woo NH, Lau AG, Vanevski F, Torre ER, Jones KR, Feng Y, Lu B, Xu B. 2008. Distinct role of long 3' UTR BDNF mRNA in spine morphology and synaptic plasticity in hippocampal neurons. *Cell* **134**:175–187. DOI: <https://doi.org/10.1016/j.cell.2008.05.045>, PMID: 18614020
- Andrews S**. 2010. FastQC. *Babraham Bioinformatics*.
- Ansari S**, Troelenberg N, Dao VA, Richter T, Bucher G, Klingler M. 2018. Double abdomen in a short-germ insect: zygotic control of Axis formation revealed in the beetle *Tribolium castaneum*. *PNAS* **115**:1819–1824. DOI: <https://doi.org/10.1073/pnas.1716512115>, PMID: 29432152
- Anvar SY**, Allard G, Tseng E, Sheynkman GM, de Klerk E, Vermaat M, Yin RH, Johansson HE, Ariyurek Y, den Dunnen JT, Turner SW, 't Hoen PAC. 2018. Full-length mRNA sequencing uncovers a widespread coupling between transcription initiation and mRNA processing. *Genome Biology* **19**:46. DOI: <https://doi.org/10.1186/s13059-018-1418-0>, PMID: 29598823
- Baek D**, Davis C, Ewing B, Gordon D, Green P. 2007. Characterization and predictive discovery of evolutionarily conserved mammalian alternative promoters. *Genome Research* **17**:145–155. DOI: <https://doi.org/10.1101/gr.5872707>, PMID: 17210929
- Berkovits BD**, Mayr C. 2015. Alternative 3' UTRs act as scaffolds to regulate membrane protein localization. *Nature* **522**:363–367. DOI: <https://doi.org/10.1038/nature14321>, PMID: 25896326
- Berleth T**, Burri M, Thoma G, Bopp D, Richstein S, Frigerio G, Noll M, Nüsslein-Volhard C. 1988. The role of localization of *bicoid* RNA in organizing the anterior pattern of the *Drosophila* embryo. *The EMBO Journal* **7**:1749–1756. DOI: <https://doi.org/10.1002/j.1460-2075.1988.tb03004.x>, PMID: 2901954
- Bharti K**, Liu W, Csermely T, Bertuzzi S, Arnheiter H. 2008. Alternative promoter use in eye development: the complex role and regulation of the transcription factor MITF. *Development* **135**:1169–1178. DOI: <https://doi.org/10.1242/dev.014142>, PMID: 18272592
- Bolognesi R**, Farzana L, Fischer TD, Brown SJ. 2008. Multiple wnt genes are required for segmentation in the short-germ embryo of *tribolium castaneum*. *Current Biology* **18**:1624–1629. DOI: <https://doi.org/10.1016/j.cub.2008.09.057>, PMID: 18926702
- Bucher G**, Farzana L, Brown SJ, Klingler M. 2005. Anterior localization of maternal mRNAs in a short germ insect lacking bicoid. *Evolution and Development* **7**:142–149. DOI: <https://doi.org/10.1111/j.1525-142X.2005.05016.x>, PMID: 15733312
- Buljan M**, Chalancon G, Dunker AK, Bateman A, Balaji S, Fuxreiter M, Babu MM. 2013. Alternative splicing of intrinsically disordered regions and rewiring of protein interactions. *Current Opinion in Structural Biology* **23**:443–450. DOI: <https://doi.org/10.1016/j.sbi.2013.03.006>, PMID: 23706950
- Bullock SL**, Ish-Horowicz D. 2001. Conserved signals and machinery for RNA transport in *Drosophila* oogenesis and embryogenesis. *Nature* **414**:611–616. DOI: <https://doi.org/10.1038/414611a>
- Camacho C**, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* **10**:421. DOI: <https://doi.org/10.1186/1471-2105-10-421>, PMID: 20003500
- Carroll SB**, Grenier JK, Weatherbee SD. 2005. From DNA to diversity: . In: *Molecular Genetics and the Evolution of Animal Design*. Blackwell: Oxford Press.
- Ciulli Mattioli C**, Rom A, Franke V, Imami K, Arrey G, Terne M, Woehler A, Akalin A, Ulitsky I, Chekulaeva M. 2019. Alternative 3' UTRs direct localization of functionally diverse protein isoforms in neuronal compartments. *Nucleic Acids Research* **47**:2560–2573. DOI: <https://doi.org/10.1093/nar/gky1270>, PMID: 30590745
- Clark E**, Akam M. 2016. Odd-paired controls frequency doubling in *Drosophila* segmentation by altering the pair-rule gene regulatory network. *eLife* **5**:e18215. DOI: <https://doi.org/10.7554/eLife.18215>, PMID: 27525481
- Cranston PS**, Hardy NB, Morse GE. 2012. A dated molecular phylogeny for the Chironomidae (Diptera). *Systematic Entomology* **37**:172–188. DOI: <https://doi.org/10.1111/j.1365-3113.2011.00603.x>
- Datta RR**, Ling J, Kurland J, Ren X, Xu Z, Yucel G, Moore J, Shokri L, Baker I, Bishop T, Struffi P, Levina R, Bulyk ML, Johnston RJ, Small S. 2018. A feed-forward relay integrates the regulatory activities of bicoid and orthodenticle via sequential binding to suboptimal sites. *Genes & Development* **32**:723–736. DOI: <https://doi.org/10.1101/gad.311985.118>, PMID: 29764918

- Davuluri RV**, Suzuki Y, Sugano S, Plass C, Huang TH. 2008. The functional consequences of alternative promoter use in mammalian genomes. *Trends in Genetics* **24**:167–177. DOI: <https://doi.org/10.1016/j.tig.2008.01.008>, PMID: 18329129
- de Klerk E**, 't Hoen PA. 2015. Alternative mRNA transcription, processing, and translation: insights from RNA sequencing. *Trends in Genetics* **31**:128–139. DOI: <https://doi.org/10.1016/j.tig.2015.01.001>, PMID: 25648499
- Derti A**, Garrett-Engle P, Macisaac KD, Stevens RC, Sriram S, Chen R, Rohl CA, Johnson JM, Babak T. 2012. A quantitative atlas of polyadenylation in five mammals. *Genome Research* **22**:1173–1183. DOI: <https://doi.org/10.1101/gr.132563.111>, PMID: 22454233
- Driever W**. 1993. 'Maternal control of anterior development in the *Drosophila* embryo.'. In: *The Development of Drosophila Melanogaster*. Laboratory Press: Cold Spring Harbor.
- Ewen-Campen B**, Schwager EE, Extavour CG. 2010. The molecular machinery of germ line specification. *Molecular Reproduction and Development* **77**:3–18. DOI: <https://doi.org/10.1002/mrd.21091>, PMID: 19790240
- Forrest AR**, Kawaji H, Rehli M, Baillie JK, de Hoon MJ, Haberle V, Lassmann T, Kulakovskiy IV, Lizio M, Andersson R, Mungall CJ, Meehan TF, Schmeier S, Bertin N, Jørgensen M, Dimont E, Arner E, Schmid C, Schaefer U, et al. 2014. A promoter-level mammalian expression atlas. *Nature* **507**:462–470. DOI: <https://doi.org/10.1038/nature13182>, PMID: 24670764
- Fu J**, Posnien N, Bolognesi R, Fischer TD, Rayl P, Oberhofer G, Kitzmann P, Brown SJ, Bucher G. 2012. Asymmetrically expressed axin required for anterior development in tribolium. *PNAS* **109**:7782–7786. DOI: <https://doi.org/10.1073/pnas.1116641109>, PMID: 22552230
- Gavis ER**, Lehmann R. 1992. Localization of *Nanos* RNA controls embryonic polarity. *Cell* **71**:301–313. DOI: [https://doi.org/10.1016/0092-8674\(92\)90358-J](https://doi.org/10.1016/0092-8674(92)90358-J), PMID: 1423595
- Gentleman RC**, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, et al. 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology* **5**:R80. DOI: <https://doi.org/10.1186/gb-2004-5-10-r80>, PMID: 15461798
- Grabherr MG**, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**:644–652. DOI: <https://doi.org/10.1038/nbt.1883>, PMID: 21572440
- Grimaldi D**, Engel MS. 2005. *Evolution of Insects*. Cambridge: Cambridge University Press.
- Gusev O**, Suetsugu Y, Cornette R, Kawashima T, Logacheva MD, Kondrashov AS, Penin AA, Hatanaka R, Kikuta S, Shimura S, Kanamori H, Katayose Y, Matsumoto T, Shagimardanova E, Alexeev D, Govorun V, Wisecaver J, Mikhayev A, Koyanagi R, Fujie M, et al. 2014. Comparative genome sequencing reveals genomic signature of extreme desiccation tolerance in the anhydrobiotic midge. *Nature Communications* **5**:4784. DOI: <https://doi.org/10.1038/ncomms5784>, PMID: 25216354
- Holt CE**, Bullock SL. 2009. Subcellular mRNA localization in animal cells and why it matters. *Science* **326**:1212–1216. DOI: <https://doi.org/10.1126/science.1176488>, PMID: 19965463
- Houtmeyers R**, Souopgui J, Tejpar S, Arkell R. 2013. The ZIC gene family encodes multi-functional proteins essential for patterning and morphogenesis. *Cellular and Molecular Life Sciences* **70**:3791–3811. DOI: <https://doi.org/10.1007/s00018-013-1285-5>, PMID: 23443491
- Ihaka R**, Gentleman R. 1996. R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics* **5**:299–314. DOI: <https://doi.org/10.2307/1390807>
- Ikeshima-Kataoka H**, Skeath JB, Nabeshima Y, Doe CQ, Matsuzaki F. 1997. Miranda directs Prospero to a daughter cell during *Drosophila* asymmetric divisions. *Nature* **390**:625–629. DOI: <https://doi.org/10.1038/37641>, PMID: 9403694
- Jackman SD**, Vandervalk BP, Mohamadi H, Chu J, Yeo S, Hammond SA, Jahesh G, Khan H, Coombe L, Warren RL, Birol I. 2017. ABySS 2.0: resource-efficient assembly of large genomes using a bloom filter. *Genome Research* **27**:768–777. DOI: <https://doi.org/10.1101/gr.214346.116>, PMID: 28232478
- Juhn J**, Marinotti O, Calvo E, James AA. 2008. Gene structure and expression of *Nanos* (*nos*) and *Oskar* (*osk*) orthologues of the vector mosquito, *Culex quinquefasciatus*. *Insect Molecular Biology* **17**:545–552. DOI: <https://doi.org/10.1111/j.1365-2583.2008.00823.x>, PMID: 18828840
- Juhn J**, James AA. 2006. *Oskar* gene expression in the vector mosquitoes, *Anopheles gambiae* and *Aedes aegypti*. *Insect Molecular Biology* **15**:363–372. DOI: <https://doi.org/10.1111/j.1365-2583.2006.00655.x>, PMID: 16756555
- Juhn J**, James AA. 2012. Hybridization in situ of salivary glands, ovaries, and embryos of vector mosquitoes. *Journal of Visualized Experiments*:3709. DOI: <https://doi.org/10.3791/3709>, PMID: 22781778
- Jürgens G**, Wieschaus E, Nüsslein-Volhard C, Kluding H. 1984. Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster*: II. Zygotic loci on the third chromosome. *Wilhelm Roux's Archives of Developmental Biology* **193**:283–295. DOI: <https://doi.org/10.1007/BF00848157>, PMID: 28305338
- Kaiser TS**, Poehn B, Szkiba D, Preussner M, Sedlazeck FJ, Zrim A, Neumann T, Nguyen LT, Betancourt AJ, Hummel T, Vogel H, Dorner S, Heyd F, von Haeseler A, Tessmar-Raible K. 2016. The genomic basis of circadian and circalunar timing adaptations in a midge. *Nature* **540**:69–73. DOI: <https://doi.org/10.1038/nature20151>, PMID: 27871090
- Kelley JL**, Peyton JT, Fiston-Lavier AS, Teets NM, Yee MC, Johnston JS, Bustamante CD, Lee RE, Denlinger DL. 2014. Compact genome of the antarctic midge is likely an adaptation to an extreme environment. *Nature Communications* **5**:4611. DOI: <https://doi.org/10.1038/ncomms5611>, PMID: 25118180

- Kim S**, Oh M, Jung W, Park J, Choi H-G, Shin SC. 2017. Genome sequencing of the winged midge, *Parochlus steinenii*, from the antarctic peninsula. *GigaScience* **6**. DOI: <https://doi.org/10.1093/gigascience/giw009>
- Klomp J**, Athy D, Kwan CW, Bloch NI, Sandmann T, Lemke S, Schmidt-Ott U. 2015. Embryo development. A cysteine-clamp gene drives embryo polarity in the midge *chironomus*. *Science* **348**:1040–1042. DOI: <https://doi.org/10.1126/science.aaa7105>, PMID: 25953821
- Krueger F**. 2012. Trim galore!. Babraham Bioinformatics.
- Larkin MA**, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. 2007. Clustal W and clustal X version 2.0. *Bioinformatics* **23**: 2947–2948. DOI: <https://doi.org/10.1093/bioinformatics/btm404>, PMID: 17846036
- Lau AG**, Irier HA, Gu J, Tian D, Ku L, Liu G, Xia M, Fritsch B, Zheng JQ, Dingledine R, Xu B, Lu B, Feng Y. 2010. Distinct 3'UTRs differentially regulate activity-dependent translation of brain-derived neurotrophic factor (BDNF). *PNAS* **107**:15945–15950. DOI: <https://doi.org/10.1073/pnas.1002929107>, PMID: 20733072
- Lehmann R**. 2016. Germ plasm biogenesis—an Oskar-Centric perspective. *Current Topics in Developmental Biology* **116**:679–707. DOI: <https://doi.org/10.1016/bs.ctdb.2015.11.024>, PMID: 26970648
- Lemke S**, Schmidt-Ott U. 2009. Evidence for a composite anterior determinant in the hover fly *Episyrphus balteatus* (Syrphidae), a cyclorrhaphan fly with an anterodorsal serosa anlage. *Development* **136**:117–127. DOI: <https://doi.org/10.1242/dev.030270>, PMID: 19060334
- Liao Y**, Smyth GK, Shi W. 2013. The subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Research* **41**:e108. DOI: <https://doi.org/10.1093/nar/gkt214>, PMID: 23558742
- Linnen CR**, Poh YP, Peterson BK, Barrett RD, Larson JG, Jensen JD, Hoekstra HE. 2013. Adaptive evolution of multiple traits through multiple mutations at a single gene. *Science* **339**:1312–1316. DOI: <https://doi.org/10.1126/science.1233213>, PMID: 23493712
- Liu Q**, Onal P, Datta RR, Rogers JM, Schmidt-Ott U, Bulyk ML, Small S, Thornton JW. 2018. Ancient mechanisms for the evolution of the bicoid homeodomain's function in fly development. *eLife* **7**:e34594. DOI: <https://doi.org/10.7554/eLife.34594>, PMID: 30298815
- Lynch JA**, Brent AE, Leaf DS, Pultz MA, Desplan C. 2006. Localized maternal orthodenticle patterns anterior and posterior in the long germ wasp *Nasonia*. *Nature* **439**:728–732. DOI: <https://doi.org/10.1038/nature04445>, PMID: 16467838
- Magoč T**, Salzberg SL. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**:2957–2963. DOI: <https://doi.org/10.1093/bioinformatics/btr507>, PMID: 21903629
- Mallarino R**, Linden TA, Linnen CR, Hoekstra HE. 2017. The role of isoforms in the evolution of cryptic coloration in *Peromyscus* mice. *Molecular Ecology* **26**:245–258. DOI: <https://doi.org/10.1111/mec.13663>, PMID: 27105018
- Martin M**. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**:10–12. DOI: <https://doi.org/10.14806/ej.17.1.200>
- Mayr C**, Bartel DP. 2009. Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* **138**:673–684. DOI: <https://doi.org/10.1016/j.cell.2009.06.016>, PMID: 19703394
- Morgan M**, Pages H, Obenchain V, Hayden N. 2013. Rsamtools: Binary alignment (BAM), FASTA, variant call (BCF), and tabix file import. In: *Bioconductor*.
- Pal S**, Gupta R, Davuluri RV. 2012. Alternative transcription and alternative splicing in Cancer. *Pharmacology & Therapeutics* **136**:283–294. DOI: <https://doi.org/10.1016/j.pharmthera.2012.08.005>, PMID: 22909788
- Peter IS**, Davidson EH. 2011. Evolution of gene regulatory networks controlling body plan development. *Cell* **144**:970–985. DOI: <https://doi.org/10.1016/j.cell.2011.02.017>, PMID: 21414487
- Pinto PA**, Henriques T, Freitas MO, Martins T, Domingues RG, Wyrzykowska PS, Coelho PA, Carmo AM, Sunkel CE, Proudfoot NJ, Moreira A. 2011. RNA polymerase II kinetics in polo polyadenylation signal selection. *The EMBO Journal* **30**:2431–2444. DOI: <https://doi.org/10.1038/emboj.2011.156>, PMID: 21602789
- Prühs R**, Beermann A, Schröder R. 2017. The roles of the Wnt-Antagonists axin and Lrp4 during embryogenesis of the red flour beetle *tribolium castaneum*. *Journal of Developmental Biology* **5**:10. DOI: <https://doi.org/10.3390/jdb5040010>
- Robinson MD**, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**:139–140. DOI: <https://doi.org/10.1093/bioinformatics/btp616>, PMID: 19910308
- Robinson MD**, Oshlack A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology* **11**:R25. DOI: <https://doi.org/10.1186/gb-2010-11-3-r25>, PMID: 20196867
- Schmidt-Ott U**, Rafiqi AM, Lemke S. 2010. *Hox3/zen* and the evolution of extraembryonic epithelia in insects. *Advances in Experimental Medicine and Biology* **689**:133–144. DOI: [https://doi.org/10.1007/978-1-4419-6673-5\\_10](https://doi.org/10.1007/978-1-4419-6673-5_10), PMID: 20795328
- Schrankel CS**, Solek CM, Buckley KM, Anderson MK, Rast JP. 2016. A conserved alternative form of the purple sea urchin HEB/E2-2/E2A transcription factor mediates a switch in E-protein regulatory state in differentiating immune cells. *Developmental Biology* **416**:149–161. DOI: <https://doi.org/10.1016/j.ydbio.2016.05.034>, PMID: 27265865
- Schröder R**. 2003. The genes *orthodenticle* and *hunchback* substitute for *bicoid* in the beetle *tribolium*. *Nature* **422**:621–625. DOI: <https://doi.org/10.1038/nature01536>, PMID: 12687002
- Shabalina SA**, Spiridonov AN, Spiridonov NA, Koonin EV. 2010. Connections between alternative transcription and alternative splicing in mammals. *Genome Biology and Evolution* **2**:791–799. DOI: <https://doi.org/10.1093/gbe/evq058>, PMID: 20889654

- Shabalina SA**, Ogurtsov AY, Spiridonov NA, Koonin EV. 2014. Evolution at Protein ends: major contribution of alternative transcription initiation and termination to the transcriptome and proteome diversity in mammals. *Nucleic Acids Research* **42**:7132–7144. DOI: <https://doi.org/10.1093/nar/gku342>, PMID: 24792168
- Shapiro IM**, Cheng AW, Flytzanis NC, Balsamo M, Condeelis JS, Oktay MH, Burge CB, Gertler FB. 2011. An EMT-driven alternative splicing program occurs in human breast Cancer and modulates cellular phenotype. *PLoS Genetics* **7**:e1002218. DOI: <https://doi.org/10.1371/journal.pgen.1002218>, PMID: 21876675
- Shepard PJ**, Choi EA, Lu J, Flanagan LA, Hertel KJ, Shi Y. 2011. Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. *RNA* **17**:761–772. DOI: <https://doi.org/10.1261/rna.2581711>, PMID: 21343387
- Smibert P**, Miura P, Westholm JO, Shenker S, May G, Duff MO, Zhang D, Eads BD, Carlson J, Brown JB, Eisman RC, Andrews J, Kaufman T, Cherbas P, Celniker SE, Graveley BR, Lai EC. 2012. Global patterns of tissue-specific alternative polyadenylation in *Drosophila*. *Cell Reports* **1**:277–289. DOI: <https://doi.org/10.1016/j.celrep.2012.01.001>, PMID: 22685694
- Stauber M**, Jäckle H, Schmidt-Ott U. 1999. The anterior determinant bicoid of *Drosophila* is a derived hox class 3 gene. *PNAS* **96**:3786–3789. DOI: <https://doi.org/10.1073/pnas.96.7.3786>, PMID: 10097115
- Stauber M**, Prell A, Schmidt-Ott U. 2002. A single Hox3 gene with composite bicoid and zerknullt expression characteristics in non-Cyclorrhaphan flies. *PNAS* **99**:274–279. DOI: <https://doi.org/10.1073/pnas.012292899>
- Stern DL**, Sucena E. 2000. Preparation of larval and adult cuticles for light microscopy. In: Sullivan W, Ashburner M, Hawley RS (Eds). *Drosophila Protocols*. Cold Spring Harbor: CSHL Press.
- Struhl G**, Johnston P, Lawrence PA. 1992. Control of *Drosophila* body pattern by the hunchback morphogen gradient. *Cell* **69**:237–249. DOI: [https://doi.org/10.1016/0092-8674\(92\)90405-2](https://doi.org/10.1016/0092-8674(92)90405-2), PMID: 1568245
- Taliaferro JM**, Vidaki M, Oliveira R, Olson S, Zhan L, Saxena T, Wang ET, Graveley BR, Gertler FB, Swanson MS, Burge CB. 2016. Distal alternative last exons localize mRNAs to neural projections. *Molecular Cell* **61**:821–833. DOI: <https://doi.org/10.1016/j.molcel.2016.01.020>, PMID: 26907613
- Tautz D**. 1988. Regulation of the *Drosophila* segmentation gene hunchback by two maternal morphogenetic centres. *Nature* **332**:281–284. DOI: <https://doi.org/10.1038/332281a0>, PMID: 2450283
- Thawornwattana Y**, Dalquen D, Yang Z. 2018. Coalescent analysis of phylogenomic data confidently resolves the species relationships in the anopheles gambiae species complex. *Molecular Biology and Evolution* **35**:2512–2527. DOI: <https://doi.org/10.1093/molbev/msy158>, PMID: 30102363
- Trapnell C**, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**:1105–1111. DOI: <https://doi.org/10.1093/bioinformatics/btp120>, PMID: 19289445
- Tushev G**, Glock C, Heumüller M, Biever A, Jovanovic M, Schuman EM. 2018. Alternative 3' UTRs modify the localization, regulatory potential, stability, and plasticity of mRNAs in neuronal compartments. *Neuron* **98**:495–511. DOI: <https://doi.org/10.1016/j.neuron.2018.03.030>, PMID: 29656876
- Vacik T**, Stubbs JL, Lemke G. 2011. A novel mechanism for the transcriptional regulation of wnt signaling in development. *Genes & Development* **25**:1783–1795. DOI: <https://doi.org/10.1101/gad.17227011>, PMID: 21856776
- Vicoso B**, Bachtrog D. 2015. Numerous transitions of sex chromosomes in diptera. *PLOS Biology* **13**:e1002078. DOI: <https://doi.org/10.1371/journal.pbio.1002078>, PMID: 25879221
- Wiegmann BM**, Trautwein MD, Winkler IS, Barr NB, Kim JW, Lambkin C, Bertone MA, Cassel BK, Bayless KM, Heimberg AM, Wheeler BM, Peterson KJ, Pape T, Sinclair BJ, Skevington JH, Blagoderov V, Caravas J, Kutty SN, Schmidt-Ott U, Kampmeier GE, et al. 2011. Episodic radiations in the fly tree of life. *PNAS* **108**:5690–5695. DOI: <https://doi.org/10.1073/pnas.1012675108>, PMID: 21402926
- Wiesner T**, Lee W, Obenauf AC, Ran L, Murali R, Zhang QF, Wong EW, Hu W, Scott SN, Shah RH, Landa I, Button J, Lailler N, Sboner A, Gao D, Murphy DA, Cao Z, Shukla S, Hollmann TJ, Wang L, et al. 2015. Alternative transcription initiation leads to expression of a novel ALK isoform in Cancer. *Nature* **526**:453–457. DOI: <https://doi.org/10.1038/nature15258>, PMID: 26444240
- Wimmer EA**, Carleton A, Harjes P, Turner T, Desplan C. 2000. *Bicoid-independent formation of thoracic segments in Drosophila*. *Science* **287**:2476–2479. DOI: <https://doi.org/10.1126/science.287.5462.2476>, PMID: 10741965
- Wittkopp PJ**, Kalay G. 2011. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nature Reviews Genetics* **13**:59–69. DOI: <https://doi.org/10.1038/nrg3095>, PMID: 22143240