A transcriptomic atlas of *Aedes aegypti* reveals detailed functional organization of major body parts and gut regional specializations in sugar-fed and blood-fed adult females

Bretta Hixson, Xiao-Li Bing[†], Xiaowei Yang[‡], Alessandro Bonfini, Peter Nagy, Nicolas Buchon*

Cornell Institute of Host-Microbe Interactions and Disease, Department of Entomology, Cornell University, Ithaca, United States

Abstract Mosquitoes transmit numerous pathogens, but large gaps remain in our understanding of their physiology. To facilitate explorations of mosquito biology, we have created Aegypti-Atlas (http://aegyptiatlas.buchonlab.com/), an online resource hosting RNAseq profiles of Ae. aegypti body parts (head, thorax, abdomen, gut, Malpighian tubules, ovaries), gut regions (crop, proventriculus, anterior and posterior midgut, hindgut), and a gut time course of blood meal digestion. Using Aegypti-Atlas, we provide insights into regionalization of gut function, blood feeding response, and immune defenses. We find that the anterior and posterior midgut possess digestive specializations which are preserved in the blood-fed state. Blood feeding initiates the sequential induction and repression/depletion of multiple cohorts of peptidases. With respect to defense, immune signaling components, but not recognition or effector molecules, show enrichment in ovaries. Basal expression of antimicrobial peptides is dominated by holotricin and gambicin, which are expressed in carcass and digestive tissues, respectively, in a mutually exclusive manner. In the midgut, gambicin and other effectors are almost exclusively expressed in the anterior regions, while the posterior midgut exhibits hallmarks of immune tolerance. Finally, in a cross-species comparison between Ae. aegypti and Anopheles gambiae midguts, we observe that regional digestive and immune specializations are conserved, indicating that our dataset may be broadly relevant to multiple mosquito species. We demonstrate that the expression of orthologous genes is highly correlated, with the exception of a 'species signature' comprising a few highly/disparately expressed genes. With this work, we show the potential of Aegypti-Atlas to unlock a more complete understanding of mosquito biology.

Editor's evaluation

Hixson et al. provide a large overview of gene expression level of the mosquito Aedes aegypti through the use of RNA-seq. They analyse gene expression changes in the digestive tract, as well as the 3 body regions and the ovaries in various conditions. These organ-specific transcriptomes fill a hole in our understanding of mosquito vector biology and will be an excellent starting point for many researchers to produce new projects.

Introduction

Hematophagous mosquito vectors contribute substantially to the world's disease burden through the transmission of *Plasmodium* parasites and arboviruses ("*World Health Organisation, 2020a* | Vector-borne diseases,"). With the rise of insecticide resistance (*Moyes et al., 2017*; *Ranson et al., 2011*; "WHO | Insecticide resistance," *World Health Organisation, 2020b*), and the expansion of mosquito species' ranges potentiated by climate change (*Liu et al., 2020*), new and more efficient means of vector control are needed, grounded in a solid understanding of mosquito physiology. The FlyAtlas project (*Chintapalli et al., 2007*; *Leader et al., 2018*) contributed to the field of insect physiology, not only by characterizing the functions associated with the major body parts and organs of *Drosophila melanogaster*, but by making it possible to quickly ascertain where in the body, and in what quantity, any gene of interest is transcribed. FlyAtlas also showed how the various contributions of different organs added up to constitute the whole fly transcriptome, providing an estimate of the relative transcriptional yield of the parts of the *Drosophila melanogaster* body. Similar anatomical datasets have been created for several species of *Anopheles* (*Baker et al., 2011*; *Martínez-Barnetche et al., 2012*; *Sreenivasamurthy et al., 2017*), but not for *Aedes* mosquitoes. While transcriptional profiles have been created for various *Ae. aegypti* body parts in diverse studies (*e.g* head (*Ptitsyn et al., 2011*), Malpighian tubules (*Li et al., 2017*), midgut (*Angleró-Rodríguez et al., 2017*; *Cui and Franz, 2020*; *Dong et al., 2017*; *Hyde et al., 2020*; *Liu et al., 2016*; *Raquin et al., 2017*; *Xiao et al., 2017*), ovaries (*Akbari et al., 2013*; *Nag et al., 2021*), carcass (*Akbari et al., 2013*; *Choi et al., 2012*), the lack of uniformity of strain and methodology between these studies makes direct comparison difficult.

The mosquito gut is of special interest to vector control, as it is central to the hematophagous lifestyle and serves as the first interface between a mosquito and the pathogens it transmits. Insect guts are highly regionalized organs, with specialized functions differentially distributed along their length (*Benguettat et al., 2018*; *Buchon and Osman, 2015*; *Buchon et al., 2013*; *Dutta et al., 2015*; *Lemaitre and Miguel-Aliaga, 2013*). The mosquito gut is divided into five anatomically distinct regions: foregut (comprising pharynx, dorsal diverticula, and crop), proventriculus, anterior midgut, posterior midgut, and hindgut. The crop and dorsal diverticula are muscular cuticle-lined sacs which store imbibed sugar for gradual release into the midgut (*Clements, 1992*). In *Anopheles gambiae*, the proventriculus and anterior midgut were found to exhibit an enrichment of transcripts encoding antimicrobial peptides (AMPs) and anti-*Plasmodium* factors (*Warr et al., 2007*) suggesting that these regions serve a special defensive function against orally acquired pathogens. Within the midgut, digestive functions are believed to be divided between the anterior and posterior regions, with the former specializing in the digestion of nectar but playing no direct role in the digestion of the blood meal, which is processed in the posterior midgut (*Hecker, 1977*). The hindgut receives and eliminates excreta from the midgut and urine from the Malpighian tubules (*Clements, 1992*). Genome-wide transcriptomic profiles have the potential to reveal much more detailed information about the regionalization of function in the mosquito gut. In recent years, many groups have used microarrays and RNAseq to create transcriptomic profiles of mosquito midguts, mostly in the context of infection (*Baker et al., 2011*; *Cui and Franz, 2020*; *Dong et al., 2017*; *Hyde et al., 2020*; *Raquin et al., 2017*; *Vedururu et al., 2019*; *Vlachou et al., 2005*; *Warr et al., 2007*). However, to our knowledge, only a single microarray-based study has ever characterized the transcriptomes of individual mosquito midgut regions (*Warr et al., 2007*) and no transcriptomic profiles have been generated for either foregut or hindgut tissues in any mosquito species.

The *Ae. aegypti* gut's response to the ingestion of a blood meal has been characterized as a biphasic process. In the first phase, mRNAs encoding a cohort of 'early' serine endopeptidases, transcribed in response to juvenile hormone (JH) secretion during previtellogenic maturation, are translated (*Bian et al., 2008*; *Felix et al., 1991*; *Jiang et al., 1997*; *Noriega et al., 1996*; *Noriega et al., 1997*; *Noriega et al., 2001*). In the second, ecdysone signaling mediates the transcription of a 'late' cohort of endo and exopeptidases, with the transcripts of key peptidases and overall proteolytic activity typically peaking around 24 hr post blood meal (pbm) (*Barillas-Mury et al., 1991*; *Brackney et al., 2010*; *Gooding, 1966*; *Graf and Briegel, 1982*; *Graf et al., 1998*; *He et al., 2021*; *Isoe et al., 2009b*; *Noriega et al., 2002*). The transcriptional timeline is, however, more complex than a simple early translation/late transcription paradigm might suggest. Several studies have, by RT-qPCR and microarray, documented de novo transcription of peptidases and other genes peaking much earlier than 24 hr pbm (*Brackney et al., 2010*; *Isoe et al., 2009b*; *Sanders et al., 2003*). However, we found

no RNAseq data to allow us to compare gut or midgut transcriptomes genome-wide in sugar-fed conditions and across the time-course of bloodmeal digestion.

Here, we aimed to complement existing datasets by, first, contextualizing the adult female gut's transcriptome with profiles of the whole body and other major body parts and organs (head, thorax, abdomen, Malpighian tubules, ovaries) with the dual goal of (a) creating an at-a-glance reference for the anatomical distribution of transcripts, genome-wide and (b) facilitating the identification of tissue-specific marker genes which may, in the future, be useful in the construction of tissue-specific expression systems. Second, we created profiles for the five anatomically distinct regions of the adult female gut (crop, proventriculus, anterior midgut, posterior midgut, hindgut) to facilitate explorations of the regionalization of gut function. Finally, we profiled transcriptional responses to blood feeding in the anterior and posterior midgut and at multiple timepoints (6, 24, and 48 hr) after blood meal ingestion. We opted to profile *Aedes aegypti*, as it is among the most important insect vectors (*Powell, 2018*), with an increasing worldwide range (*Iwamura et al., 2020*), and a recently updated genome assembly (*Matthews et al., 2018*). All the resulting transcriptional data can be accessed at the Aegypti-Atlas online database we constructed (http://aegyptiatlas.buchonlab.com/).

With RNAseq profiles for body parts, gut regions, and blood-fed guts in hand, we launched a broad investigation into multiple aspects of mosquito biology. Our inquiries were guided by the following questions: What functions are associated with each body part? What are the functional specializations of the gut regions? How does the gut transcriptome change in the context of blood feeding, and how are those changes distributed in the midgut? What can our atlas tell us about the organization of mosquitoes' antimicrobial defenses? Finally, in a cross-species comparison between the midguts of *Ae. aegypti* and *Anopheles gambiae* (*s.l.*), we asked whether the gut's structure-function relationship is conserved between two hematophagous mosquito vector species.

Results

The transcriptional landscape of an adult mosquito reflects the embryonic origins and functional signatures of its major tissues

In order to develop a transcriptomic atlas of some of the main body parts of *Ae. aegypti*, we generated RNAseq profiles from the whole bodies and dissected body parts (head, thorax, abdomen, gut, Malpighian tubules, and ovaries) of sucrose-fed adult female mosquitoes of a field-derived 'Thai' strain from a region where arbovirus transmission is endemic (*League et al., 2019*). For the dissected body parts (but not whole body), we omitted the anterior-most section of the thorax as well as the final segments of the abdomen to exclude transcripts from salivary glands and from sperm in the spermatheca (see *Figure 1—figure supplement 1A* for diagram). We first evaluated the relative variance between body part and whole-body profiles by Principal Component Analysis (PCA) (*Figure 1A*). All transcriptomes from the same body part clustered together closely, providing evidence for the replicability of our data. On PC1, body parts were grouped in a manner that accorded with their divergent embryonic origin (primarily mesoderm/ectoderm for head, thorax, and abdomen, endoderm for gut and Malpighian tubules) and whole-body transcriptomes were clustered in between. Ovaries, which contain the germline, strongly separated from the other, exclusively somatic, body parts on PC2. Altogether, PCA indicated that the transcriptomes of body parts are largely defined by their embryonic origin, and that germline tissues possess a highly distinctive transcriptional signature.

Each body part we sequenced expressed a set of unique or nearly unique transcripts some of which, we hypothesize, may be exclusive tissue markers. We conducted a census of putative tissue markers (here defined as genes expressed at five or more transcripts per million (TPM) in one body part and enriched at least 50-fold compared to all other body parts). *Figure 1B* presents a selection of markers from each body part (full census in *Supplementary file 1*). The quantity of markers ranged from 8 in the abdomen to 391 in the ovaries, which expressed more markers than all other body parts combined; nearly one in fifty of all genes in the genome qualified as an ovary marker. We further noted that ovaries simultaneously expressed the greatest number of highly enriched transcripts *relative to other dissected body parts* (*Figure 1—figure supplement 1B*) and the smallest number of highly enriched transcripts *relative to whole body* (*Figure 1—figure supplement 1C*). This apparent paradox led us to hypothesize that the ovaries make a disproportionately large contribution to the transcriptome of the whole female mosquito. A gene that is uniquely expressed in the ovaries will not
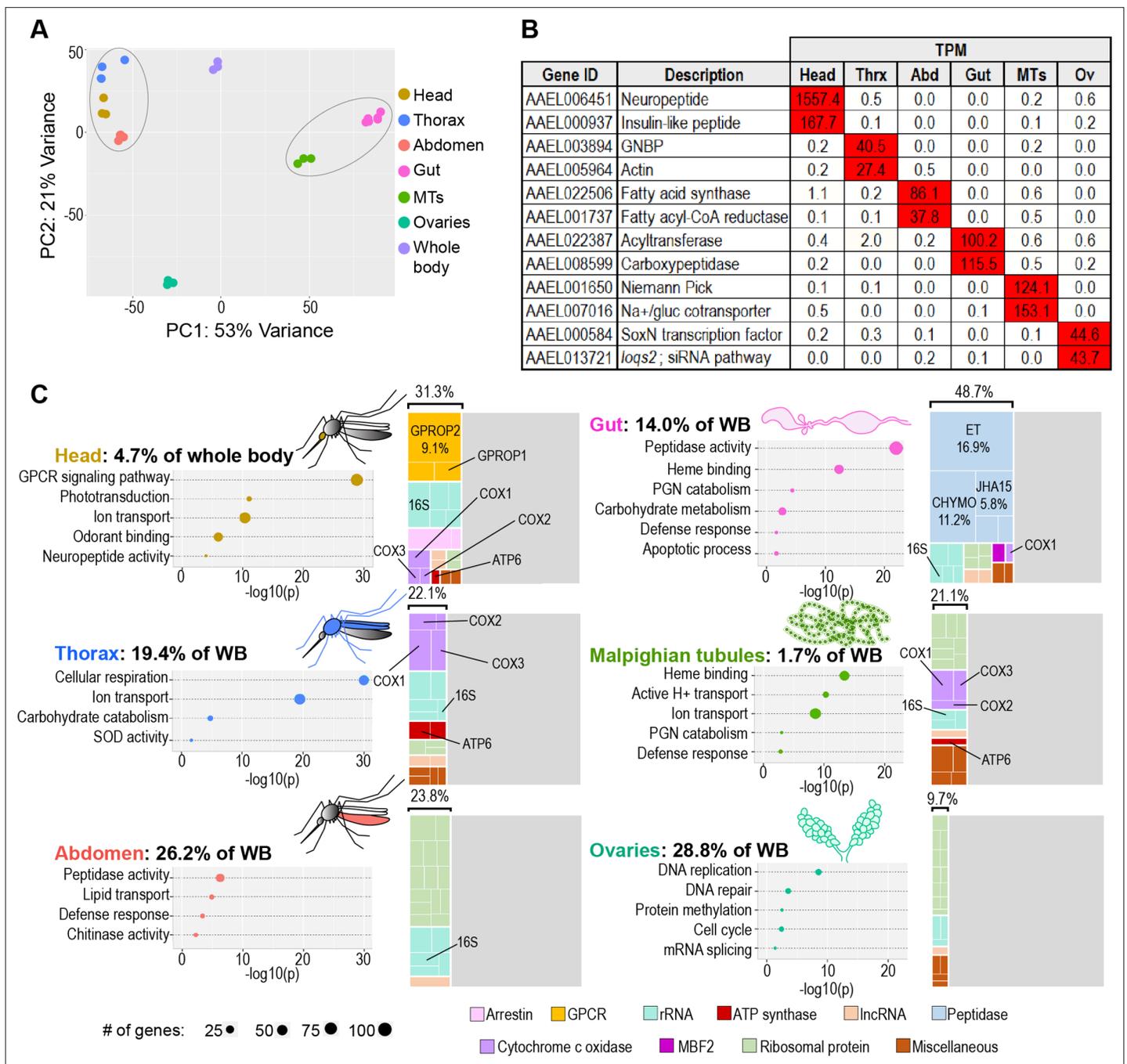
Figure 1. The transcriptomes of *Aedes aegypti* body parts reflect their embryonic origins and functional specializations. (**A**) Principal component analysis of the transcriptomes of whole body and body parts; 3–6 replicates per body part. Circled clusters contain body parts of dominantly endodermal (left) and mesodermal/ectodermal (right) derivation. (**B**) Expression of body-part-specific putative tissue markers (in TPM) (**C**) Calculated estimates of body parts' contributions to whole body transcriptome; Bubble plots: classic Fisher topGO gene ontology enrichment analysis of genes enriched 5 x (head, gut thorax, Malpighian tubules, abdomen) or 2 x (ovaries) relative to whole body (DESeq2, *padj* <0.05); Treemaps: 20 highest expressed genes per body part, grouped by category/function; percentages reflect the proportion of the transcriptome comprised by the displayed genes.

The online version of this article includes the following figure supplement(s) for figure 1:

Figure supplement 1. Ovaries possess many highly enriched genes relative to other body parts, but few relative to whole body.

Figure supplement 2. RT-qPCR confirms the anatomical specificity of putative marker genes.

Figure supplement 3. Calculated scaling factors accurately predict whole body gene expression.

be highly enriched relative to the whole mosquito body if the ovarian transcriptome comprises a large proportion of the whole body's transcriptome.

To validate our sequencing results, we performed RT-qPCR amplification of one putative marker gene from each body part. We selected a translation factor (AAEL013144) as a reference gene for our assay over more traditional candidates (*Dzaki et al., 2017*), on the basis of its robust expression and relatively small variation across samples *Figure 1—figure supplement 2A*. The results of our RT-qPCR confirmed the anatomical specificity of each selected marker (*Figure 1—figure supplement 2B* Fig).

The existence of transcripts specific to each body part allowed us to calculate their relative contribution to the whole body's transcriptome (see the 'mosquito equation' in Materials and methods). Expressed as percentages, the body part contributions we obtained were: head (4.7%), thorax (19.4%), abdomen (26.2%), gut (14.0%), Malpighian tubules (1.7%), and ovaries (28.8%). It should be noted that the percentage calculated for the thorax is likely to slightly underestimate its actual contribution, as a portion of the thorax was removed (*Figure 1—figure supplement 1A*) to eliminate salivary glands. This removal likely accounts, at least in part, for the fact that the sum of the calculated scaling factors is 94.9%, rather than 100%. We validated our estimates by using them to predict the whole-body expression of every gene in the genome, then plotting those predictions against observed values from the whole-body transcriptome (*Figure 1—figure supplement 3A*). Calculated and observed values were highly correlated (slope of 1.00, $R^2 = 0.95$) after removal of a single outlying gene (AAEL018689, a mitochondrial ribosomal RNA) which was underpredicted by our calculation. This analysis validates our hypothesis that the ovaries contribute a disproportionate number of transcripts to the whole-body transcriptome. Further, our data demonstrate that, due to the preponderance of ovarian contribution, multiple tissues have only a low representation in the body, severely limiting the ability to detect tissue-specific changes in whole-mosquito experiments.

To examine the biological processes and molecular functions enriched in each body part, we performed a Gene Ontology Enrichment Analysis (GOEA) of genes 5 x (or, for ovaries, 2 x) enriched in each body part (in comparison to whole body, DESeq2 *padj* <0.05). An extended list of GO categories may be found in *Supplementary file 2*. We complemented this approach by identifying the twenty highest expressed genes in each body part (treemaps in *Figure 1C*), reasoning that the highest expressed genes in a transcriptome may also give insight into function. In brief, the head showed enrichment for sensory machinery and neuronal signaling, the thoracic carcass (which houses flight muscle) showed signs of enhanced metabolic activity, and the abdominal carcass (which contains mostly fat body) displayed enrichment for regulatory proteolytic enzymes (CLIP-domain serine endopeptidases), defense, and lipid transport. The gut was enriched with defense-related genes including immune-activating peptidoglycan recognition proteins (PGRPs), AMPs, and immune-modulating amidase PGRPs. It also showed enrichment for digestive enzymes, especially peptidases. Notably, its three most prevalent transcripts were the non-CLIP serine endopeptidases early trypsin (*ET*), female-specific chymotrypsin (*CHYMO*), and juvenile hormone-regulated chymotrypsin (*JHA15*) which, together, accounted for more than one third of its transcriptome. (Note: Vectorbase gene IDs for all genes named in this text are listed in *Supplementary file 3*). The Malpighian tubules, which serve a function analogous to mammalian kidneys, expressed large quantities of ion transporters, especially vacuolar ATPases, which may be required to create proton gradients for use in secondary active transport (*Weng et al., 2003*). Because the ovaries did not express many highly enriched genes, we adjusted our analysis to examine genes expressed 2 x higher than in the whole body and discovered categories related to cell division and germline maintenance. We also observed that a small cohort of mitochondrial genes (*COX1*, *COX2*, *COX3*, *ATP6*, and a 16 s gene) were among the highest expressed transcripts in several body parts, especially those containing tissues that carry out energy-intensive functions (head, thorax, Malpighian tubules). Overall, we concluded that the most prominent transcripts and enriched GO categories in each body part were consistent with expectations for the tissues they house. We take these results as a validation of our data set, and as confirmation that the segments of the mosquito carcass (head, thorax, abdomen) perform well as proxies for important tissues (brain/eyes, flight muscle, fat body).

Gut region-specific transcriptomes reveal functional compartmentalization of the mosquito midgut

To evaluate the regionalization of gut function, we generated RNAseq profiles for the five main regions of the mosquito digestive tract: crop and dorsal diverticula (hereafter referred to as 'crop' for brevity), proventriculus, anterior midgut, posterior midgut, and hindgut. A PCA of these profiles (*Figure 2A*) demonstrated clear segregation between regions, and close clustering of all replicates from the same region, indicating clearly distinct transcriptomes. The anterior and posterior midgut, which are predominantly derived from endoderm, clustered together opposite the crop and hindgut, which are of primarily ectodermal origin. The proventriculus, which is partially derived from both germ layers, was intermediate between these two clusters on PC1. Whole-gut transcriptomes clustered closely with posterior midgut transcriptomes, suggesting that the posterior midgut contributes more to the transcriptome of the whole gut than do the other regions. Accordingly, very few genes in the posterior midgut showed enrichment compared to the whole gut (*Figure 2—figure supplement 1A*). GOEA on genes enriched in each region (*Figure 2B*), complemented by an examination of the twenty highest expressed genes in each (*Figure 2C*) revealed that the ectodermally derived regions of the gut (crop and hindgut) expressed genes involved in chitin metabolism, lipid metabolism, ion transport and numerous CLIP-domain serine endopeptidases. The proventriculus was notably enriched for wingless signaling and defensive genes (lysozymes, and the highly expressed AMP gambicin, *GAM1*). In the anterior midgut, carbohydrate metabolism and heme-binding proteins (predominantly cytochrome P450s) were the most prominently enriched categories. The highest expressed gene in this region was a member of the MBF2 family of transcription factors of unknown function in mosquitoes. As the posterior midgut transcriptome closely resembled that of the whole gut, only a handful of genes were enriched 5 x in this region relative to whole gut. We also noted that a large proportion of the genome was either unexpressed or little-expressed in the posterior midgut relative to the other gut regions (*Figure 2—figure supplement 1B*). To better identify genes enriched in the posterior midgut, we calculated enrichment against an aggregate of the other four gut regions. We found that the posterior midgut is highly enriched for peptidases (especially non-CLIP serine endopeptidases) and genes involved in translation. Other enriched categories include signal recognition particle (SRP)-dependent targeting, amino acid transport and amino acid metabolism. Together these categories describe a region that is primed to translate and secrete large quantities of peptidases, and to process and absorb the resulting free amino acids. Overall, our data demonstrate that the five gut regions are highly specialized functional units.

The central function of the gut is digestion, a highly sequential process. To better understand how digestion and nutrient absorption are distributed along the gut, we examined the regional cumulative expression of enzymes and transporters putatively involved in the digestion and absorption of lipids, carbohydrates, and proteins/amino acids. We postulated that the cumulative expression of a given set of digestive enzymes/transporters of common function reflects the *investment* of each region in that function (what fraction of their transcriptome they dedicate to that process). We first assembled lists of lipases, lipid-transporting proteins, amylases/maltases, glucosidases, sugar transporters, peptidases, and amino acid transporters with a probable digestive function (excluding genes with anticipated non-digestive function, *e.g.*, peptidases annotated as, or orthologous with, CLIP-domain serine endopeptidases, caspases, proteasome components, *etc.*). For complete lists and inclusion criteria, see *Supplementary file 4*. We summed the expression of the genes from each category (in TPM) and calculated their relative expression by region (*Figure 2D*). The highest investments in different categories of lipases were divided among several regions, with phospholipases receiving the greatest investment from crop and anterior midgut, sphingomyelinases from anterior midgut, and the remainder of lipases modestly more expressed in the posterior midgut and hindgut. The crop was the greatest investor in lipid carrier proteins and fatty acid transporters, while the anterior midgut showed the greatest investment in phospholipid and sterol transporters. The highest investments in sugar digestion/absorption and protein digestion were found in the anterior and posterior midgut, respectively. The hindgut was a disproportionate investor in amino acid transporters, suggesting that many of the products of protein digestion from the midgut are absorbed there. These data demonstrate that distinct gut regions invest different amounts of their respective transcriptomes in the digestion and absorption of different nutrients, hinting at a system of specialization/sequential processing of ingested materials.
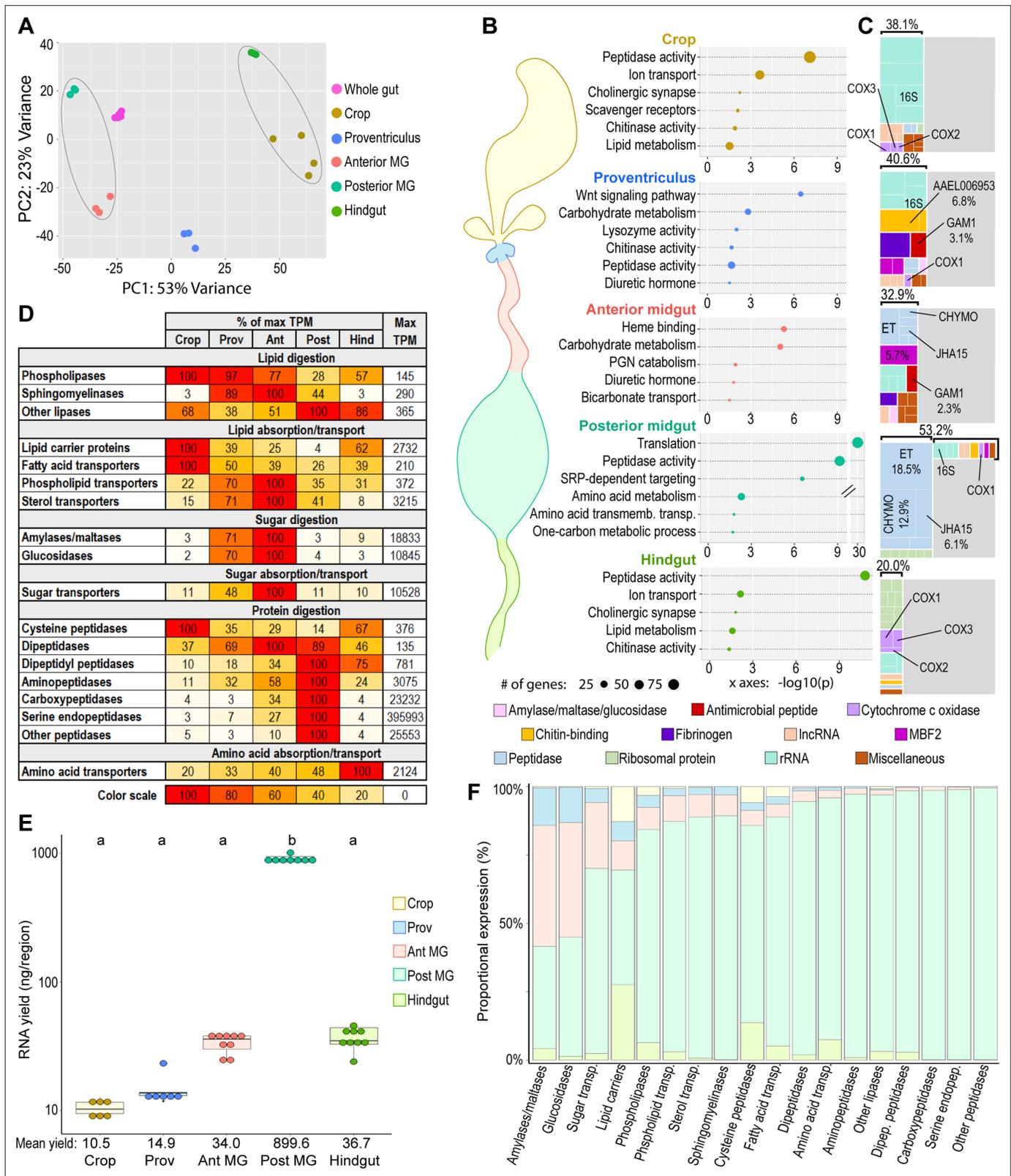
Figure 2. Mosquito gut regions' transcriptomic investments and outputs reveal digestive specializations. (**A**) Principal component analysis of the transcriptomes of whole gut and gut regions; 3–6 replicates. Circled clusters contain body parts of partial endodermal (left) and ectodermal (right) derivation. (**B**) Bubble plots: classic Fisher topGO gene ontology enrichment analysis of regionally enriched genes (for crop, proventriculus, and anterior midgut: 5 x enriched over whole-gut, DESeq2 padj < 0.05; for posterior midgut: enriched over all other gut regions, DESeq2 padj < 0.05) (**C**) Treemaps:

Figure 2 continued on next page

Figure 2 continued

20 highest expressed genes per region, grouped by category/function; percentages reflect the proportion of the transcriptome comprised by the displayed genes. (**D**) Regional investments in categories of digestive enzymes and transporters. Colored cells are scaled as percentages of the region with the highest transcriptional investment in the category (as measured by cumulative TPM). The region with the highest investment is scored as '100', with its cumulative TPM displayed in the rightmost column ('Max TPM'). (**E**) Estimates of regions' transcriptional output by empirical quantification of RNA yield; statistics: One-way ANOVA, Tukey HSD. (**F**) Estimated output of digestive enzymes/transporters by region (categorical investments weighted by regional RNA yield).

The online version of this article includes the following figure supplement(s) for figure 2:

Figure supplement 1. The transcriptome of the posterior midgut displays little deviation from the whole-gut transcriptome and low transcriptomic diversity.

Caution should be used in the interpretation of cumulative expression data, as *investment* (in TPM) is distinct from *output* (i.e. the number of transcripts of a given gene or category produced in a region). The five regions do not produce equal numbers of transcripts overall, so their transcriptional investments must be scaled by their total transcriptional *yield* to achieve an estimate of their relative contributions to specific functions. We measured the RNA content from each region (**Figure 2E**) and scaled the cumulative expression values from **Figure 2D** by yield to gain an estimate of the transcriptional output of each region with respect to each digestive category (**Figure 2F**). We found that the posterior midgut, by virtue of its outsized transcriptional yield, is the dominant contributor not only of peptidases, but of most categories of transcripts with digestive and absorptive functions. However, the proventriculus and anterior midgut together are the source of more than half of all amylases/maltases and glucosidases and contribute a sizeable minority of all sugar transporters. These observations cement our conclusion that the anterior midgut specializes in the digestion and absorption of carbohydrates, while the posterior midgut is responsible for most protein and lipid digestion.

Blood feeding initiates a series of transcriptomic shifts over the course of digestion

The gut of the hematophagous mosquito undergoes dramatic changes upon blood feeding. To better understand these changes at a transcriptome-wide level, we generated RNAseq profiles for sugar-fed guts as well as guts at 4–6 hr (hereafter referred to as 6 hr for brevity), 24 hr, and 48 hr after feeding on a live chicken. PCA (**Figure 3A**) demonstrates that the transcriptome changes throughout at least the first 24 hr pbm, before reverting to a near-basal condition by 48 hr pbm, suggesting that the transcriptional regulation of digestion is a very dynamic process.

To evaluate the functional changes to the blood-fed gut's transcriptome, we performed a GOEA (**Figure 3B**) comparing the gut at each time-point to its state at the time-point prior. Guts at the final timepoint (48 hr) were also compared to sugar-fed guts. The analysis was limited to genes that were up or downregulated at least two-fold (DESeq2 *padj* <0.05).

We found that, at 24 hr pbm, after ecdysone titers reach their peak, (**Hagedorn et al., 1975**), steroid receptor activity was significantly enriched. A total of five steroid receptor genes were upregulated, including *EcR*, *USP* (encoding the co-receptor to EcR), *Hnf4*, *ftz-f1*, and *HR3*. Multiple categories pertaining to protein synthesis and secretion were also regulated by blood-feeding: at 6 hr pbm, a small number of genes associated with mRNA splicing, protein folding, and SRP-dependent secretion were upregulated, while genes associated with miRNA inhibition of translation were downregulated. However, by 24 hr pbm, all these changes were reversed and, additionally, genes associated with peptide transport (39 genes) and translation (57 genes) were significantly downregulated. At 48 hr pbm, protein folding, peptide transport, and translation were all upregulated again. Notably, translation was one of very few categories that was significantly enriched in 48 hr blood-fed guts compared to exclusively sugar-fed guts.

Several of the categories regulated by blood feeding appear to reflect necessary adaptations to the changing luminal environment. Six hrs following blood meal ingestion, ferric iron-binding proteins (including three ferritin subunit precursors, AAEL004335, AAEL007383, and AAEL007385) were upregulated and two genes in the lysozyme activity category (*LYSC11*, and *LYSC4*) were also induced, possibly in response to the previously described increase of ROS and bacterial density pbm (**Gusmão et al., 2010**). At 24 hr pbm, the defense response category was also induced. Two defensins (*DEFA* and *DEFD*) were enriched, as well as two peptidoglycan-degrading amidases (*PGRP-SC1*, and
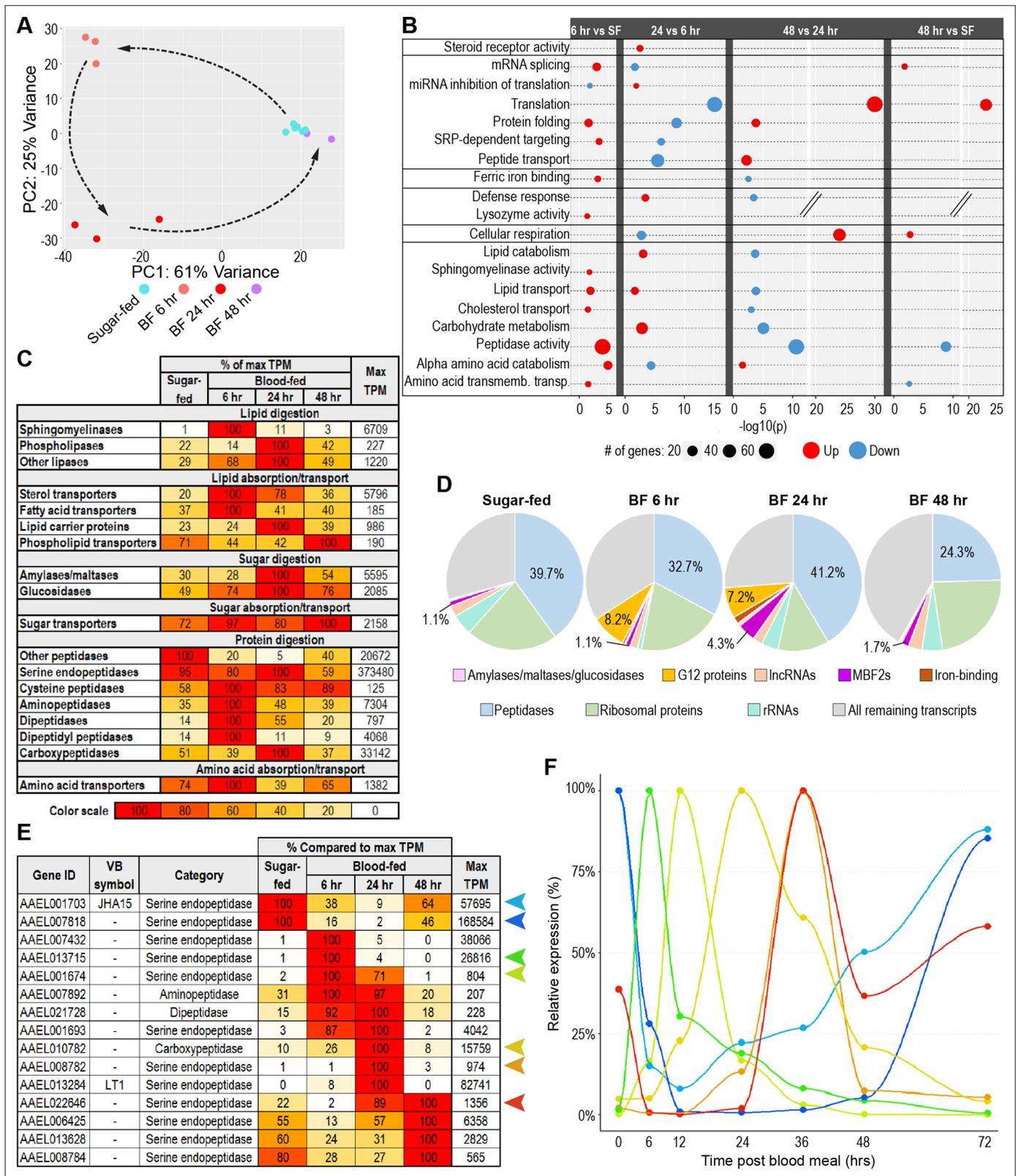
Figure 3. The gut's transcriptome shifts continuously throughout blood meal digestion. (**A**) Principal component analysis of the transcriptomes of sugar-fed and blood-fed guts; 3–6 replicates per condition. (**B**) Classic Fisher topGO gene ontology enrichment analysis of blood-fed guts at each timepoint pbm, in comparison to prior timepoint (DESeq2, *padj* <0.05, minimum of 2 x fold-change). (**C**) Investments in categories of digestive enzymes and transporters by timepoint/condition. Colored cells are scaled as percentages of the region with the highest transcriptional investment in the category

Figure 3 continued on next page

(as measured by cumulative TPM). The region with the highest investment is scored as '100', with its cumulative TPM displayed in the rightmost column. (**D**) Proportional investment (out of whole transcriptome) in selected gene categories/families by timepoint/condition. (**E**) Expression of selected peptidases, scaled as a percentage of maximum expression. The region with the highest expression (in TPM) is scored as '100', with its TPM displayed in the rightmost column. (**F**) Relative expression of selected peptidases at 0, 6, 12, 24, 36, 48, and 72 hrs post-blood meal measured by RT-qPCR; reference gene: AAEL009653; 3 replicates.

The online version of this article includes the following figure supplement(s) for figure 3:

Figure supplement 1. Increasing RNA yield drives higher gut peptidase output at 24 hr post blood meal.

Figure supplement 2. Gut peptidase transcripts are induced and depleted asynchronously during blood meal digestion.

Figure supplement 3. Temporal clustering of serine endopeptidases is independent of physical proximity and sequence similarity.

PGRPS4), orthologs of which, in *Drosophila*, modulate the activation of the immune deficiency (IMD) pathway (*Buchon et al., 2009*). At 48 hr pbm, both the ferric iron-binding and defense response categories were downregulated relative to 24 hr, which is unsurprising as most of the luminal content had been evacuated by that time (*Figure 3—figure supplement 1A*). We also noted a significant induction of genes involved in cellular respiration which could reflect the metabolic expenditure required for the peristaltic evacuation of the blood bolus. GO categories related to the digestion and absorption of macronutrients were nearly all upregulated at 6 hr pbm, with some upregulated again at 24 hr but downregulated to return to baseline by 48 hr. In summary, GOEA indicates that blood feeding regulates the gut's transcription of genes involved in diverse functions, including digestion, absorption, translation, and defense.

In adult mosquitoes, a peritrophic matrix is synthesized in the hours immediately following the ingestion of a blood meal. Previous studies have documented or proposed the involvement of several genes in peritrophic matrix synthesis in *Ae. aegypti*. These include the glucosamine-fructose-6-phosphate amino transferase *AeGfat-1*, which catalyzes the first step of chitin biosynthesis, the chitin synthase *AeCs*, which catalyzes the final step (*Kato et al., 2006*), and the peritrophins AAEL004798, AAEL006953, and *Aper50* (*Shao et al., 2005*; *Whiten et al., 2018*). All five of these genes were robustly expressed in our blood-fed guts, and two (*AeGfat-1* and *Aper50*) were significantly upregulated at 6 hr pbm. Notably, the peritrophin AAEL006953 was heavily concentrated in the proventriculus at baseline and, at a TPM of over 68,000, was that region's top expressed gene (*Figure 2B*). Other candidates, orthologous to *Drosophila* chitin synthases (AAEL002718) and peritrophins (AAEL005702, AAEL009585) showed little expression in the gut at any timepoint. Surprisingly, *DUOX*, the ortholog of a dual oxidase which catalyzes the formation of a dityrosine network in the mucin layer adjacent to the *An. gambia* peritrophic matrix (*Kumar et al., 2010*), was notably absent in the *Ae. aegypti* gut (*Figure 3—figure supplement 1B*).

To characterize the blood-fed gut's changes in digestion/absorption in more detail, we calculated the cumulative investment of the gut in each digestive category at each timepoint (*Figure 3C*). Overall, cumulative transcript analysis echoed the results of our GOEA. Both showed that the gut's transcriptional investment in sphingomyelinases and amino acid transporters peaked at 6 hr pbm, and that the expression of carbohydrate digestive enzymes was elevated at 24 hr pbm but declined again at 48 hr pbm. Our granular cumulative expression analysis revealed that aminopeptidases, dipeptidases, and dipeptidyl peptidases reached peak investment at 6 hr, and carboxypeptidases at 24 hr. Remarkably, the gut's investment in serine endopeptidases (non-CLIP), which constitute the majority of peptidase transcripts in the gut by nearly an order of magnitude, remained steady across sugar-fed, 6 hr, and 24 hr pbm guts, before plunging by approximately one third at 48 hr pbm.

We extended our cumulative investment approach to other functions in the gut by aggregating transcripts of other transcriptionally prominent families and comparing their relative proportions across all timepoints (*Figure 3D*). This revealed the dramatic temporary increase in the expression of a family of twelve genes containing the insect allergen domain (IPR010629) which are orthologous to the *An. gambiae* G12 gene and have been found to possess hemolytic, cytolytic, and antiviral properties (*Foo et al., 2021*). We also observed that the family of MBF2 transcription factors - one member of which was previously noted among the highest expressed genes of the anterior midgut (*Figure 2C*) - surged from just over 1% of the transcriptome to more than 4% at 24 hr pbm, partially subsiding to approximately 1.7% by 48 hr pbm. Our comparison of summed transcripts yielded an

apparent paradox: while GOEA demonstrated the upregulation of multiple peptidases at 6 hr pbm, the gut's overall investment in peptidases (cumulative TPM) had actually declined at this timepoint from 39.7% to 32.7% of the transcriptome, before rising to 41.2% at 24 hr pbm. This contradiction is explained by the depletion of transcripts from the small 'early' cohort of peptidases (*ET*, *CHYMO*, and *JHA15*) which compensated for the upregulation of other peptidases at 6 hr pbm. It should be noted that, while peptidase *investment* did not increase much in blood-fed guts, higher RNA yield *Figure 3—figure supplement 1C* caused peptidase transcript *output* to increase significantly at 24 hr pbm (*Figure 3—figure supplement 1D*).

To more closely examine the dynamics of peptidase transcription in the blood-fed gut, we performed a clustering analysis (*Figure 3—figure supplement 2A*) which revealed that the transcript levels of many peptidases in the gut change dramatically over the course of blood meal digestion, that these changes come in the form of both induction and repression/depletion, frequently followed by a return to baseline at the subsequent timepoint, and that they manifest in multiple successive waves. Our clustering analysis divided the dynamically expressed gut peptidases into three waves of induced transcripts (rapid, intermediate, and delayed), and three corresponding waves of repressed or depleted transcripts. Among induced peptidases, the rapid and delayed waves comprise genes that peaked in our data set at 6 hr and 24 hr pbm, (23 and 40 genes respectively). Our clustering analysis implied that another, 'intermediate' cohort of induced peptidases (16 genes) nests somewhere between these waves. These genes show sustained induction from 6 to 24 hr, and likely peak at some time between these two timepoints. The waves of rapid, intermediate, and delayed repressed/depleted peptidases mirror the expression patterns of their induced counterparts in inverse. *Figure 3E* shows the expression of representative peptidases from each of the six cohorts across the blood feeding kinetic. We validated our findings by performing RT-qPCR on selected peptidases (two repressed/depleted and five induced genes) at 6, 12, 24, 36, 48, and 72 hr pbm. We selected *RpS30* (AAEL009653), a housekeeping gene with robust and steady expression over the course of blood meal digestion, as a reference gene for our assay (*Figure 3—figure supplement 2B*). *Figure 3F* shows the relative expression of each peptidase (for unscaled data normalized only to the reference gene, see *Figure 3—figure supplement 2C*). At each timepoint we assayed, we captured one or more genes peaking sharply, further illustrating the dynamic nature of peptidase expression in the blood-fed gut. Most peptidases are organized in genomic clusters which, we hypothesized, could underlie co-regulation. We also examined the evolutionary relatedness of the peptidases from each temporal cluster by comparing their positions on a phylogenetic tree (composed using Geneious R11 software, *Figure 3—figure supplement 3A*). While sequence similarity corresponded closely with physical location in most cases, neither parameter appeared correlated with temporal clustering. Altogether, our data reveal that blood feeding initiates a well-orchestrated transcriptomic response in the gut, dominated by the dynamic up and down-regulation of peptidases throughout the entire process of digestion.

Regional digestive specializations are preserved in the blood-fed midgut

Considering the extensive functional regionalization of the *Ae. aegypti* gut and the profound reshaping of the whole-gut transcriptome upon blood feeding, we asked whether regional differences are preserved in the blood-fed gut. To evaluate the regionalized effects of blood feeding, we generated transcriptomes for dissected anterior midguts (without proventriculus) and posterior midguts at 24 hr pbm. PCA (*Figure 4A*) revealed that all profiles clustered both by region and by blood feeding status. As we previously noted (*Figure 2A*), there was little difference between sugar-fed whole gut and posterior midgut profiles. Upon blood feeding, both whole gut and posterior midgut shifted in tandem, maintaining proximity, indicating that blood feeding does not lessen the dominance of the posterior midgut's transcriptional yield over that of the anterior. Both midgut regions shifted in response to blood feeding; moreover, they moved in the same directions with respect to PC1 and PC2, suggesting that the overall transcriptional responses of the two regions share some commonality. We noted that the relative investments of these two regions in most of the prominent functional categories did not change dramatically upon blood feeding (*Figure 4B*). However, two of the most remarkable changes in the transcriptome of blood-fed mosquitoes, the upregulation of MBF2 transcription factors and G12 genes (*Figure 3D*), were mainly localized to the posterior midgut.

D

	% of max TPM				Max TPM
	Ant		Post		
	SF	BF	SF	BF	
Lipid digestion					
Sphingomyelinases	40	82	18	100	724
Phospholipases	39	44	14	100	285
Other lipases	12	21	24	100	1511
Lipid absorption/transport					
Sterol transporters	55	100	22	69	5890
Fatty acid transporters	86	100	57	77	96
Lipid carrier proteins	58	100	9	17	1199
Phospholipid transporters	80	100	27	24	468
Sugar digestion					
Amylases/maltases	61	100	2	10	31112
Glucosidases	66	100	3	8	16537
Sugar absorption/transport					
Sugar transporters	98	100	10	9	10701
Protein digestion					
Cysteine peptidases	65	100	32	25	170
Aminopeptidases	50	42	85	100	3603
Dipeptidyl peptidases	34	31	100	62	781
Dipeptidases	41	59	36	100	332
Carboxypeptidases	26	23	75	100	31031
Serine endopeptidases	27	16	100	83	395993
Other peptidases	10	2	100	8	25553
Amino acid absorption/transport					
Amino acid transporters	84	80	100	55	1018

Color scale: 100 | 80 | 60 | 40 | 20 | 0

Figure 4. Regional specializations in sugar and protein digestion are preserved in the blood-fed midgut. (**A**) Principal component analysis of the transcriptomes of sugar-fed and 24 hr blood-fed whole guts and anterior and posterior midguts; 3–6 replicates per condition. (**B**) Proportional expression (out of whole transcriptome) of selected gene categories/families by region/condition. (**C**) Classic Fisher topGO gene ontology enrichment analysis of blood-fed versus sugar-fed regions (DESeq2, *padj* <0.05, minimum of 2 x fold-change). Categories that are regulated in both regions are

Figure 4 continued on next page

Figure 4 continued

shown in red. (**D**) Investments in categories of digestive enzymes and transporters by region/condition. Colored cells are scaled as percentages of the region/condition with the highest transcriptional investment in the category (as measured by cumulative TPM). The treatment with the highest investment is scored as '100', with its cumulative TPM displayed in the rightmost column. (**E**) Estimates of midgut regions' transcriptional yield by empirical quantification of RNA in sugar-fed guts versus 24 hr after feeding with an artificial blood replacement diet; statistics: unpaired *t*-test. (**F**) Estimated output of digestive enzymes/transporters by midgut region in sugar-fed and blood-fed mosquitoes (categorical investments weighted by regional RNA yield).

A GOEA of categories upregulated and downregulated by blood feeding in the anterior and posterior midgut (*Figure 4C*) mainly recapitulated changes already noted in blood-fed whole guts (*Figure 3B*). Most categories that were upregulated in the anterior midgut were similarly upregulated in the posterior, albeit to differing extents. The fact that blood feeding promotes enrichment of many of the same digestive/absorptive categories in both anterior and posterior midgut could imply that the two regions are converging on a common role with respect to macronutrient exploitation. However, an analysis of cumulative expression (*Figure 4D*) demonstrated that the two regions' respective investments in digestive enzymes and transporters remained highly divergent. Consistent with the results of the GOEA, the posterior midgut substantially increased its investment in lipase transcripts, with a smaller increase in the anterior midgut. By contrast, the anterior midgut kept a substantial edge in its investment in lipid transporters and lipid carrier proteins. The posterior midgut increased its investment in amylases/maltases and glucosidases by approximately five and twofold, respectively, and the anterior midgut increased both by less than twofold. However, the difference in baseline expression was such that the blood-fed anterior midgut's investment was still ten times that of the blood-fed posterior midgut. Neither region substantially increased its investment in sugar transporters, peptidases, or amino acid transporters. When these regional investments were weighted by the increased RNA yield observed in both blood-fed midgut regions (*Figure 4E*), we found that the proportional output of the posterior midgut increased across most digestive categories, but that the anterior midgut still contributed a substantial minority of transcripts for sugar-digesting enzymes and sugar transporters (*Figure 4F*). Altogether, we concluded that both midgut regions participate in the transcriptional response to blood feeding, and that the blood-fed anterior and posterior midgut retain their regional specializations with respect to carbohydrate and protein digestion.

Immune gene patterning reveals areas of immune activity and immune tolerance in the mosquito body and gut

As mosquitoes are prolific vectors of important human pathogens, their defensive functions are of paramount interest. We assembled a table recording the expression of immune genes in each category by body part, gut region, and blood-fed timepoint (*Supplementary file 5*). Many genes associated with the siRNA and piRNA pathways display a strong tropism toward the ovaries (*Figure 5A*), possibly indicating that, in mosquitoes, siRNAs as well as piRNAs (*Siomi et al., 2011*) are required to protect the germline from selfish elements. Accordingly, the gene *loqs2*, which has been described as an essential component of *Ae. aegypti's* systemic siRNA response to Zika and dengue viruses (*Olmo et al., 2018*) was restricted to the ovaries to such an extent that it qualified as an ovary-specific marker (*Figure 1B*). The ovaries were also the site of the highest expression of genes coding for components of Toll, IMD, and JAK-STAT signaling. Remarkably, genes coding for extracellular regulators (such as most CLIP-domain serine endopeptidases) and pattern recognition proteins (PGRPs and Gram-negative bacteria-binding proteins, GNBPs) did not show this pattern of expression. Instead, they were generally highly expressed in the head, thorax and abdomen, the three compartments that contain fat body. Even more remarkably, AMPs (*Figure 5B*) showed negligible expression in the ovaries. Our results suggest that, in the female germline, portions of the IMD, Toll, and JAK-STAT pathways are uncoupled from upstream pattern recognition proteins and downstream AMP targets, possibly playing a regulatory role in some alternative process (e.g. development).

An examination of the distribution of AMPs and lysozymes in the body (*Figure 5B*) revealed that two AMPs, holotricin (*GRRP*) and gambicin, together accounted for most AMP transcripts in the mosquito body. These two AMPs were expressed in different body parts in a near mutually exclusive manner. The head, thorax, and abdomen of the mosquito predominantly expressed holotricin transcripts, with
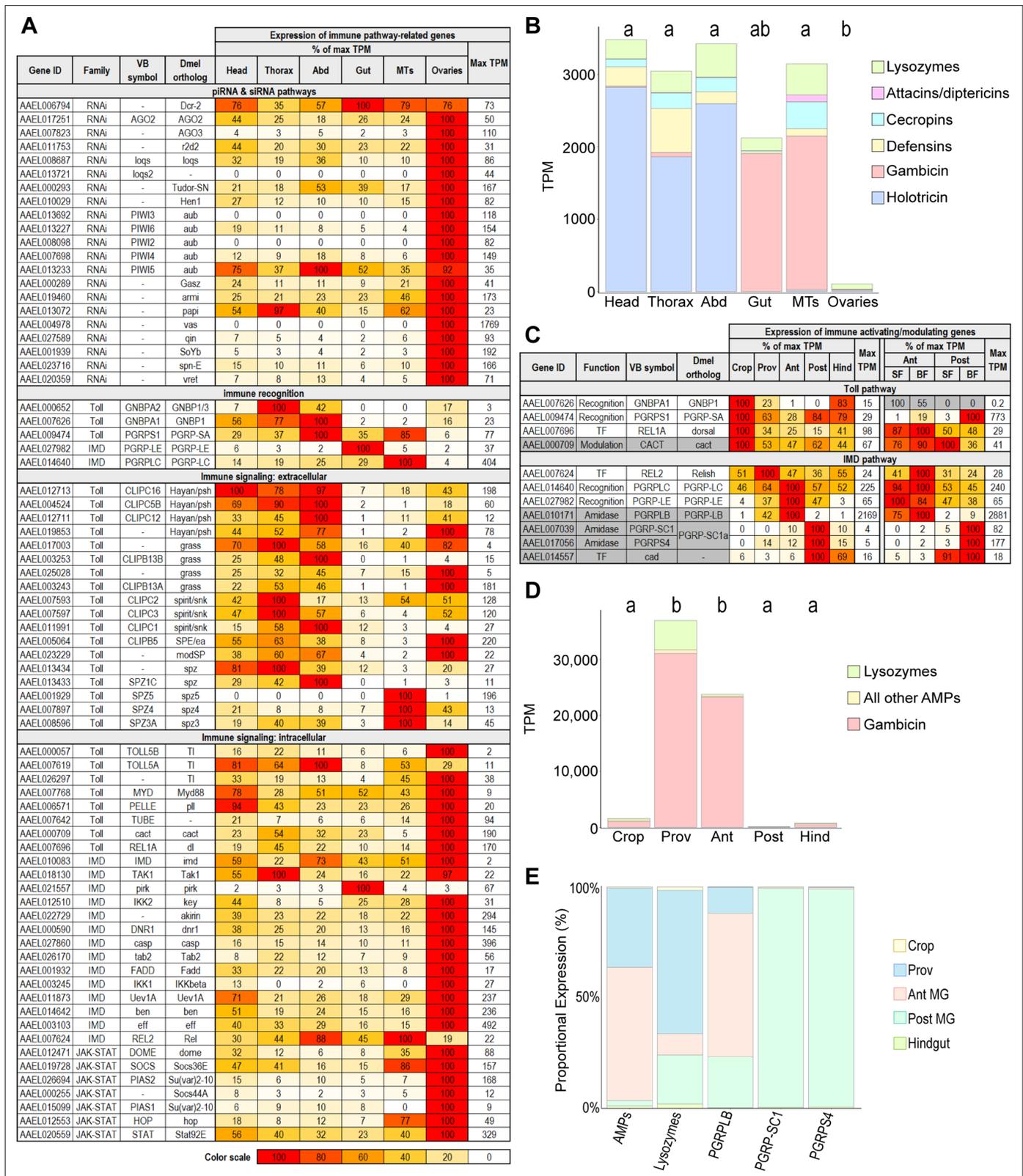
A

Expression of immune pathway-related genes

Gene ID	Family	VB symbol	Dmel ortholog	Head	Thorax	Abd	Gut	MTs	Ovaries	Max TPM
				% of max TPM						
piRNA & siRNA pathways										
AAEL006794	RNAi	-	Dcr-2	76	35	57	100	79	76	73
AAEL017251	RNAi	AGO2	AGO2	44	25	18	26	24	100	50
AAEL007823	RNAi	-	AGO3	4	3	5	2	3	100	110
AAEL011753	RNAi	-	r2d2	44	20	30	23	22	100	31
AAEL008687	RNAi	loqs	loqs	32	19	36	10	10	100	86
AAEL013721	RNAi	loqs2	-	0	0	0	0	0	100	44
AAEL000293	RNAi	-	Tudor-SN	21	18	53	39	17	100	167
AAEL010029	RNAi	-	Hen1	27	12	10	10	15	100	82
AAEL013692	RNAi	PIWI3	aub	0	0	0	0	0	100	118
AAEL013227	RNAi	PIWI6	aub	19	11	8	5	4	100	154
AAEL008098	RNAi	PIWI2	aub	0	0	0	0	0	100	82
AAEL007698	RNAi	PIWI4	aub	12	9	18	8	6	100	149
AAEL013233	RNAi	PIWI5	aub	75	37	100	52	35	92	35
AAEL000289	RNAi	-	Gasz	24	11	11	9	21	100	41
AAEL019460	RNAi	-	armi	25	21	23	23	46	100	173
AAEL013072	RNAi	-	papi	54	97	40	15	62	100	23
AAEL004978	RNAi	-	vas	0	0	0	0	0	100	1769
AAEL027589	RNAi	-	qin	7	5	4	2	6	100	93
AAEL001939	RNAi	-	SoYb	5	3	4	2	3	100	192
AAEL023716	RNAi	-	spn-E	15	10	11	6	10	100	166
AAEL020359	RNAi	-	vret	7	8	13	4	5	100	71
immune recognition										
AAEL000652	Toll	GNBPA2	GNBP1/3	7	100	42	0	0	17	3
AAEL007626	Toll	GNBPA1	GNBP1	56	77	100	2	2	16	23
AAEL009474	Toll	PGRPS1	PGRP-SA	29	37	100	35	85	6	77
AAEL027982	IMD	PGRP-LE	PGRP-LE	6	3	2	100	5	2	37
AAEL014640	IMD	PGRPLC	PGRP-LC	14	19	25	29	100	4	404
Immune signaling: extracellular										
AAEL012713	Toll	CLIPC16	Hayan/psh	100	78	97	7	18	43	198
AAEL004524	Toll	CLIPC5B	Hayan/psh	69	90	100	2	1	18	60
AAEL012711	Toll	CLIPC12	Hayan/psh	33	45	100	1	11	41	12
AAEL019853	Toll	-	Hayan/psh	44	52	77	1	2	100	78
AAEL017003	Toll	-	grass	70	100	58	16	40	82	4
AAEL003253	Toll	CLIPB13B	grass	25	48	100	0	1	4	15
AAEL025028	Toll	-	grass	25	32	45	7	15	100	5
AAEL003243	Toll	CLIPB13A	grass	22	53	46	1	1	100	181
AAEL007593	Toll	CLIPC2	spirit/snk	42	100	17	13	54	51	128
AAEL007597	Toll	CLIPC3	spirit/snk	47	100	57	6	4	52	120
AAEL011991	Toll	CLIPC1	spirit/snk	15	58	100	12	3	4	27
AAEL005064	Toll	CLIPB5	SPE/ea	55	63	38	8	3	100	220
AAEL023229	Toll	-	modSP	38	60	67	4	2	100	22
AAEL013434	Toll	-	spz	81	100	39	12	3	20	27
AAEL013433	Toll	SPZ1C	spz	29	42	100	0	1	3	11
AAEL001929	Toll	SPZ5	spz5	0	0	0	0	100	1	196
AAEL007897	Toll	SPZ4	spz4	21	8	8	7	100	43	13
AAEL008596	Toll	SPZ3A	spz3	19	40	39	3	100	14	45
Immune signaling: intracellular										
AAEL000057	Toll	TOLL5B	Tl	16	22	11	6	6	100	2
AAEL007619	Toll	TOLL5A	Tl	81	64	100	8	53	29	11
AAEL026297	Toll	-	Tl	33	19	13	4	45	100	38
AAEL007768	Toll	MYD	Myd88	78	28	51	52	43	100	9
AAEL006571	Toll	PELLE	pll	94	43	23	23	26	100	20
AAEL007642	Toll	TUBE	-	21	7	6	6	14	100	94
AAEL000709	Toll	cact	cact	23	54	32	23	5	100	190
AAEL007696	Toll	REL1A	dl	19	45	22	10	14	100	170
AAEL010083	IMD	IMD	imd	59	22	73	43	51	100	2
AAEL018130	IMD	TAK1	Tak1	55	100	24	16	22	97	22
AAEL021557	IMD	pirk	pirk	2	3	3	100	4	3	67
AAEL012510	IMD	IKK2	key	44	8	5	25	28	100	31
AAEL022729	IMD	-	akirin	39	23	22	18	22	100	294
AAEL000590	IMD	DNR1	dnr1	38	25	20	13	16	100	145
AAEL027860	IMD	casp	casp	16	15	14	10	11	100	396
AAEL026170	IMD	tab2	Tab2	8	22	12	7	9	100	56
AAEL001932	IMD	FADD	Fadd	33	22	20	13	8	100	17
AAEL003245	IMD	IKK1	IKKbeta	13	0	2	6	0	100	27
AAEL011873	IMD	Uev1A	Uev1A	71	21	26	18	29	100	237
AAEL014642	IMD	ben	ben	51	19	24	15	16	100	236
AAEL003103	IMD	eff	eff	40	33	29	16	15	100	492
AAEL007624	IMD	REL2	Rel	30	44	88	45	100	19	22
AAEL012471	JAK-STAT	DOME	dome	32	12	6	8	35	100	88
AAEL019728	JAK-STAT	SOCS	Socs36E	47	41	16	15	86	100	157
AAEL026694	JAK-STAT	PIAS2	Su(var)2-10	15	6	10	5	7	100	168
AAEL000255	JAK-STAT	-	Socs44A	8	3	2	3	5	100	12
AAEL015099	JAK-STAT	PIAS1	Su(var)2-10	6	9	10	8	0	100	9
AAEL012553	JAK-STAT	HOP	hop	18	8	12	7	77	100	49
AAEL020559	JAK-STAT	STAT	Stat92E	56	40	32	23	40	100	329

Color scale: 100 | 80 | 60 | 40 | 20 | 0

B

C

Expression of immune activating/modulating genes

Gene ID	Function	VB symbol	Dmel ortholog	Crop	Prov	Ant	Post	Hind	Max TPM	Ant SF	Ant BF	Post SF	Post BF	Max TPM
				% of max TPM						% of max TPM				
Toll pathway														
AAEL007626	Recognition	GNBPA1	GNBP1	100	23	1	0	83	15	100	55	0	0	0.2
AAEL009474	Recognition	PGRPS1	PGRP-SA	100	63	28	84	79	29	1	19	3	100	773
AAEL007696	TF	REL1A	dorsal	100	34	25	15	41	98	87	100	50	48	29
AAEL000709	Modulation	CACT	cact	100	53	47	62	44	67	76	90	100	36	41
IMD pathway														
AAEL007624	TF	REL2	Relish	51	100	47	36	55	24	41	100	31	24	28
AAEL014640	Recognition	PGRPLC	PGRP-LC	46	64	100	57	52	225	94	100	53	45	240
AAEL027982	Recognition	PGRP-LE	PGRP-LE	4	37	100	47	3	65	100	84	47	38	65
AAEL010171	Amidase	PGRPLB	PGRP-LB	1	42	100	2	1	2169	75	100	2	9	2881
AAEL007039	Amidase	PGRP-SC1	PGRP-SC1a	0	0	10	100	10	82	4	0	2	5	82
AAEL017056	Amidase	PGRPS4	PGRP-SC1a	0	14	12	100	15	5	0	2	3	100	177
AAEL014557	TF	cad	-	6	3	6	100	69	16	5	3	91	100	18

D

E

Figure 5. Immune gene patterning reveals zones of immune activity and tolerance in the mosquito body and gut. (**A**) Body parts' investments in immune pathway-related genes, scaled as a percentage of maximum expression. The body part with the highest expression (in TPM) is scored as '100' with its TPM displayed in the rightmost column. Note: immune-related genes may serve an alternate, developmental role in ovaries. (**B**) Cumulative antimicrobial peptide expression by body part (TPM); statistics: One-way ANOVA, Tukey HSD. (**C**) Expression of immune activating (white) /modulating

Figure 5 continued on next page

Figure 5 continued

(gray) genes, scaled as a percentage of maximum expression. The region with the highest expression (in TPM) is scored as '100' with its TPM displayed in the rightmost column of each half of the table. For genes with very low expression (Max TPM <1) color scale is replaced with gray. (**D**) Cumulative antimicrobial peptide expression by gut region (TPM); statistics: One-way ANOVA, Tukey HSD. (**E**) Output of immune-related genes and gene categories by gut region (TPM weighted by regional RNA yield).

very little gambicin, while the gut and Malpighian tubules were dominated by gambicin expression and expressed very few holotricin transcripts.

Immune activity in the gut is tightly regulated as this organ interfaces with both commensal and pathogenic microbes. In *Drosophila*, the IMD pathway is the main pathway controlling AMPs in the midgut, and its activity is strongly constrained by expression of amidase PGRPs (*Buchon et al., 2009*), while the Toll pathway is expressed mostly in the ectodermal portions of the gut. We found a similar pattern in the gut of *Ae. aegypti* (*Figure 5C*) with orthologs of Toll pathway recognition proteins most strongly expressed in the crop while IMD-activating PGRPs (*PGRPLC* and *PGRP-LE*) and IMD pathway components were enriched in the midgut. Investment in these IMD-activating PGRPs was highest in the anterior midgut, while immune-modulating PGRPs showed more divided expression. The amidase *PGRPLB* was most prominently expressed in the anterior midgut, but the short amidase PGRPs (*PGRP-SC1* and *PGRPS4*, orthologs of *Drosophila PGRP-SC1a* and *PGRP-SC1b*) were most strongly expressed in the posterior midgut. Likewise, an ortholog of caudal (*cad*), a transcription factor that limits AMP expression in the *Drosophila* midgut (*Clayton et al., 2013*; *Ryu et al., 2008*), was profoundly enriched in the posterior midgut. Altogether, we found that the expression patterns of these key genes suggested enhanced immune vigilance in the anterior portion of the midgut, with hallmarks of immune tolerance prominent in the posterior midgut. We also observed that blood feeding enhanced the expression of the short amidase PGRPs in the posterior midgut, sharpening the regional divide.

A cumulative analysis of the expression of immune effectors (AMPs and lysozymes) in the gut showed that the proventriculus and anterior midgut, respectively, invest 3.6% and 2.3% of their transcriptomes in the expression of these effectors – especially gambicin (*Figure 5D*). The posterior midgut, by contrast, expressed just 110 TPM of AMPs and lysozymes combined, supporting our hypothesis that the posterior midgut is characterized by immune tolerance toward microbes. When AMP and lysozyme expression is weighted by gut regional RNA yield (*Figure 5E*), it is apparent that the anterior midgut and proventriculus contribute the majority of transcripts in these categories despite their relatively small transcriptional yields. Overall, the expression patterns of immune activators, modulators, and effectors suggests that the anterior portions of the midgut may exert strong selection of microbes upon entry, while the posterior midgut is a region of greater immune tolerance.

Gut digestive and defensive regionalization is well conserved between the midguts of *Aedes aegypti* and *Anopheles gambiae*

Ae. aegypti and *An. gambiae* are both important hematophagous vectors, and their digestive tracts bear a clear anatomical similarity. To assess how similar these organs are at the transcriptomic level, we created RNAseq profiles for whole guts (comprising all regions from crop to hindgut) and midgut regions (proventriculus, anterior midgut and posterior midgut) from *An. gambiae* (*s.l.*) G3 mosquitoes and compared them to their *Ae. aegypti* counterparts. In a clustering analysis (*Figure 6A*), we found that the profiles of midgut regions and whole guts from each species were grouped similarly, with anterior midguts intermediate between proventriculi and posterior midguts. Also, in both species, whole guts clustered between anterior and posterior midgut regions but displayed more similarity to the latter, suggesting that for *An. gambiae*, as for *Ae. aegypti*, the transcriptional yield of the posterior midgut greatly exceeds that of the anterior.

We repeated our analysis of regional transcriptomic investment in nutrient digestion and absorption by generating lists of genes encoding enzymes and transporters, in the same categories as in *Figure 2C*, summing their regional expression in TPM, and calculating their relative abundance. Since we lacked transcriptomes for whole body, crop, and hindgut for *An. gambiae*, it was necessary to adjust the inclusion criteria we employed (*Supplementary file 4*) for our previous investment analysis. Complete lists of genes in each category with adjusted inclusion criteria can be found in *Supplementary file 6* for *Ae. aegypti* and *Supplementary file 7* for *An. gambiae*. Our comparison across the
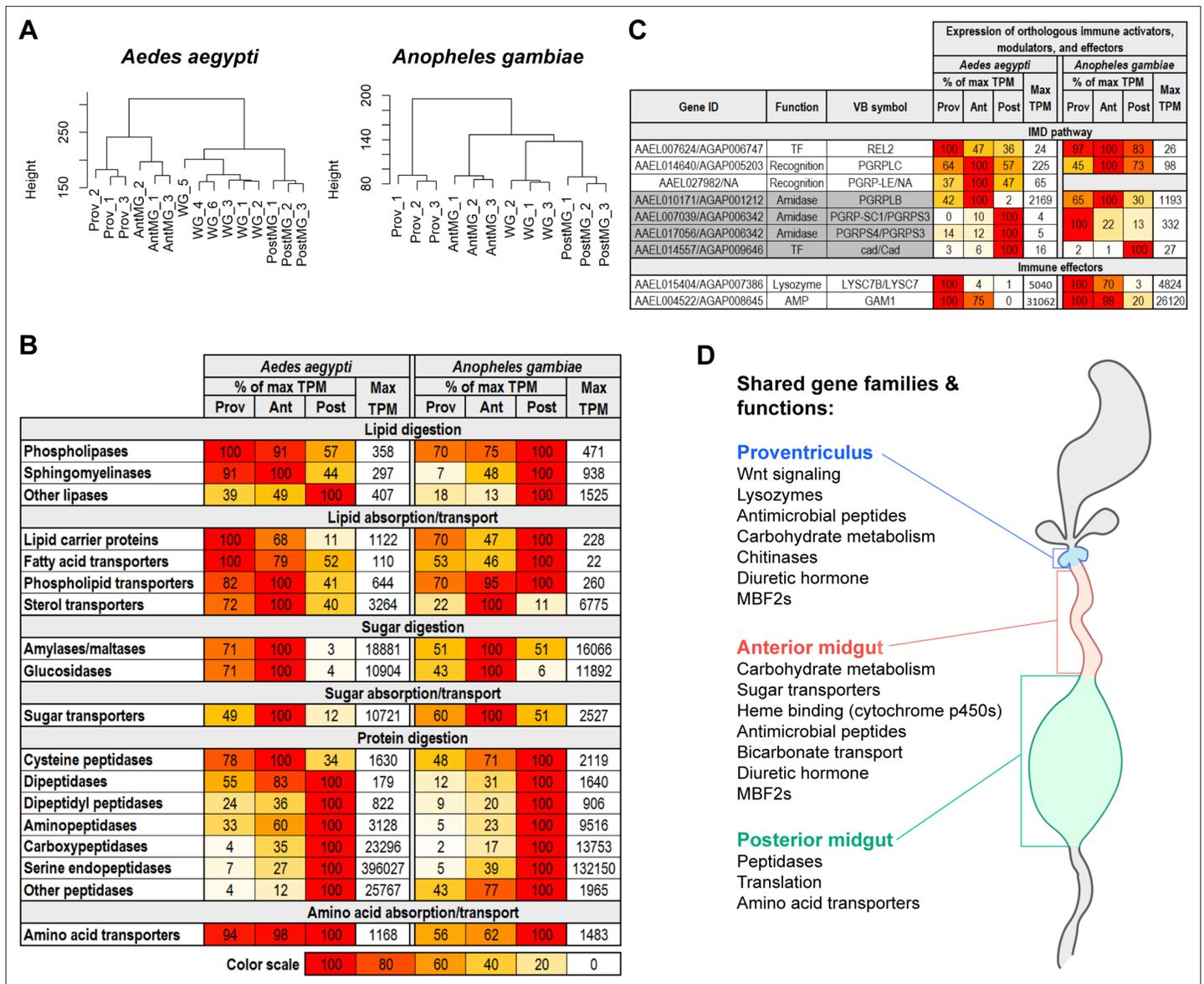
Figure 6. Digestive and defensive regional specializations are well conserved between the *Aedes aegypti* and *Anopheles gambiae* midguts. (**A**) Clustering analyses of the transcriptomes of *Aedes aegypti* and *Anopheles gambiae* whole gut and midgut regions; 3–6 replicates (**B**) Regional investments in categories of digestive enzymes and transporters. Colored cells are scaled as percentages of the region with the highest transcriptional investment in the category (as measured by cumulative TPM). The region with the highest investment is scored as '100', with its cumulative TPM displayed in the rightmost column of each half of the table. (**C**) Expression of immune activating (white)/modulating (gray) genes and effectors, scaled as a percentage of maximum expression. The region with the highest expression (in TPM) is scored as '100' with its TPM displayed in the rightmost column of each half of the table. (**D**) Conserved regional enrichment of gene families and functions in *Ae. aegypti* and *An. gambiae* midgut regions, as evaluated by classic Fisher topGO gene ontology enrichment analysis, investment analysis, and a comparison of the twenty highest expressed transcripts.

The online version of this article includes the following figure supplement(s) for figure 6:

Figure supplement 1. Midgut regions in *Aedes aegypti* and *Anopheles gambiae* display similar regional specializations, but disparate amplitudes of investment in key functions.

two species (*Figure 6B*) revealed overall conservation of gut structure-function. This was particularly evident for the anterior and posterior midguts which, respectively, remained the strongest investors in sugar and protein digestion/absorption. However, in *An. gambiae*, the posterior midgut appears to make a greater proportional investment in the digestion/absorption of both lipids and sugars. We also noted that, while the posterior midgut in both species was the dominant investor in peptidases,

the amplitude of the *Ae. aegypti* investment far outstripped that of *An. gambiae* (*Figure 6—figure supplement 1A*), particularly with respect to serine endopeptidases.

To compare the distribution of defensive functions in the two midguts, we examined the expression patterns of important orthologous immune activators, modulators, and effectors between the two mosquitoes (*Figure 6C*) and found clear similarities. In both species, the IMD-activating *PGRPLC* receptor as well as the immune-modulating *PGRPLB* amidase were most strongly expressed in the anterior midguts, the transcription factor *cad* was most strongly expressed in the posterior midguts, and the orthologs of *GAM1* and *LYSC7B* (respectively the highest expressed AMP and lysozyme in the *Ae. aegypti* midgut) were most strongly expressed in the proventriculi. This distribution of key genes leads us to hypothesize that the immune regionalization of the *An. gambiae* midgut is likely similar to that of *Ae. aegypti*. It should be noted, however, that the short amidase, *PGRPS3*, a putative negative regulator of the IMD pathway, is robustly expressed in the *An. gambiae* proventriculus, in striking contrast to its *Ae. aegypti* counterparts. We further note that, while the amplitude of *GAM1* and *LYSC7B*/*LYSC7* investment is comparable in the two species, the *An. gambaie* proventriculus and anterior midgut transcriptomes contained large quantities of transcripts for cecropins and defensins, rendering their overall AMP investment more than five times higher than the same regions in *Ae. aegypti* (*Figure 6—figure supplement 1B*). In both species, the posterior midgut's investment in AMPs was negligible. Altogether, the expression of digestive (*Figure 6B*) and defensive genes (*Figure 6C*) as well as a GOEA (*Figure 6—figure supplement 1C*) of the *An. gambiae* midgut regions confirm that the midgut structure-function relationship is well conserved between the two mosquito species. *Figure 6D* provides a summary of enriched functions and prominently expressed gene families that are shared by the midgut regions of *Ae. aegypti* and *An. gambiae*.

A small number of disparately expressed genes differentiate *Aedes aegypti* and *Anopheles gambiae* midgut transcriptomes

The conservation of gut functional regionalization may be due either to conserved expression patterns among orthologs across evolutionary distance, or to convergent patterning of non-homologous genes of shared function. To evaluate how well the relative expression of individual orthologous genes has been maintained between *Ae. aegypti* and *An. gambiae* midgut regions, we performed a series of correlation analyses using only the 7,430 genes classified by Orthogroups Analysis (*Supplementary file 8*) as one-to-one orthologs. Scaled expression values for the orthologs were poorly correlated both in whole guts (*Figure 7A*), and in dissected regions (*Figure 7—figure supplement 1A*), with R^2 values ranging from 0.08 (proventriculus) to 0.56 (posterior midgut). However, the distribution of data points in the correlation graphs suggested that a few highly/disparately expressed genes substantially depressed correlation. To test the effect of these genes on overall correlation, we sequentially censored genes that were strongly and significantly disparately expressed between both species (DESeq2, *padj* <0.05), rescaled the expression (TPM) of the remaining one-to-one orthologs, and plotted the resulting changes in slope and correlation coefficient (*Figure 7B*, *Figure 7—figure supplement 1B*). For the proventriculus, anterior midgut, posterior midgut, and whole gut the censorship of, respectively 4, 10, 20, and 35 1-to-1 orthologous pairs was sufficient to reveal strong correlation between the remaining orthologs, with slopes ranging from 0.78 to 1.1 and R^2 values ranging from 0.77 to 0.89 (*Figure 7C*, *Figure 7—figure supplement 1C*). From this, we conclude that a 'species signature' masking the transcriptional similarity of *Ae. aegypti* and *An. gambiae* midguts is the product of a small number of highly/disparately expressed genes.

Figure 7D displays the genes censored from our correlation analysis in the three midgut regions. The most disparately expressed gene in both the *An. gambiae* proventriculus and anterior midgut encodes an uncharacterized protein containing a domain (IPR007931) from the Tsetse EP protein which, in tsetse flies has been shown to play a role in resisting the establishment of trypanosome infections (*Haines et al., 2010*). MBF2 transcription factors were also censored in each region: twice they were expressed higher in *Ae. aegypti* (proventriculus and anterior midgut) and once in *An. gambiae* (posterior midgut). Among the disparately expressed orthologs, the most abundant category was ribosomal proteins. We observed that in both the anterior and posterior midgut, *Ae. aegypti* had significantly increased the expression of some ribosomal proteins relative to their one-to-one orthologs in *An. gambiae*, and *vice versa*. Phylogenetic analysis of ribosomal proteins (*Figure 7—figure supplement 2A*) confirmed that the putatively one-to-one orthologous ribosomal genes identified
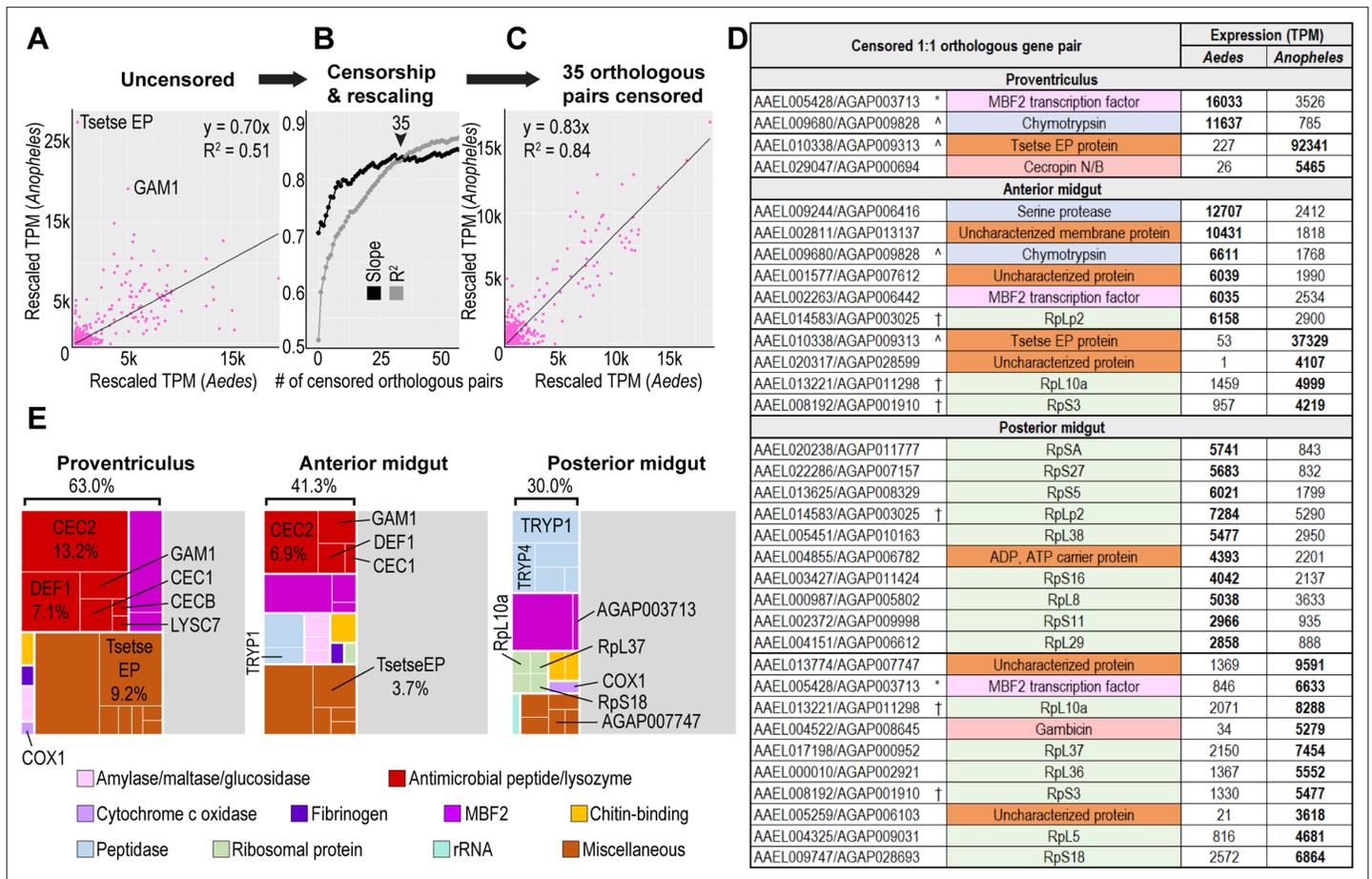
Censored 1:1 orthologous gene pair		Expression (TPM)	
		Aedes	Anopheles
Proventriculus			
AAEL005428/AGAP003713 *	MBF2 transcription factor	**16033**	3526
AAEL009680/AGAP009828 ^	Chymotrypsin	**11637**	785
AAEL010338/AGAP009313 ^	Tsetse EP protein	227	**92341**
AAEL029047/AGAP000694	Cecropin N/B	26	**5465**
Anterior midgut			
AAEL009244/AGAP006416	Serine protease	**12707**	2412
AAEL002811/AGAP013137	Uncharacterized membrane protein	**10431**	1818
AAEL009680/AGAP009828 ^	Chymotrypsin	**6611**	1768
AAEL001577/AGAP007612	Uncharacterized protein	**6039**	1990
AAEL002263/AGAP006442	MBF2 transcription factor	**6035**	2534
AAEL014583/AGAP003025 †	RpLp2	**6158**	2900
AAEL010338/AGAP009313 ^	Tsetse EP protein	53	**37329**
AAEL020317/AGAP028599	Uncharacterized protein	1	**4107**
AAEL013221/AGAP011298 †	RpL10a	1459	**4999**
AAEL008192/AGAP001910 †	RpS3	957	**4219**
Posterior midgut			
AAEL020238/AGAP011777	RpSA	**5741**	843
AAEL022286/AGAP007157	RpS27	**5683**	832
AAEL013625/AGAP008329	RpS5	**6021**	1799
AAEL014583/AGAP003025 †	RpLp2	**7284**	5290
AAEL005451/AGAP010163	RpL38	**5477**	2950
AAEL004855/AGAP006782	ADP, ATP carrier protein	**4393**	2201
AAEL003427/AGAP011424	RpS16	**4042**	2137
AAEL000987/AGAP005802	RpL8	**5038**	3633
AAEL002372/AGAP009998	RpS11	**2966**	935
AAEL004151/AGAP006612	RpL29	**2858**	888
AAEL013774/AGAP007747	Uncharacterized protein	1369	**9591**
AAEL005428/AGAP003713 *	MBF2 transcription factor	846	**6633**
AAEL013221/AGAP011298 †	RpL10a	2071	**8288**
AAEL004522/AGAP008645	Gambicin	34	**5279**
AAEL017198/AGAP000952	RpL37	2150	**7454**
AAEL000010/AGAP002921	RpL36	1367	**5552**
AAEL008192/AGAP001910 †	RpS3	1330	**5477**
AAEL005259/AGAP006103	Uncharacterized protein	21	**3618**
AAEL004325/AGAP009031	RpL5	816	**4681**
AAEL009747/AGAP028693	RpS18	2572	**6864**

Figure 7. A small number of disparately expressed genes constitute a 'species signature' differentiating midgut transcriptomes in *Aedes aegypti* and *Anopheles gambiae.* (**A–C**) Correlation of one-to-one orthologous genes before and after censorship of the 35 most disparately expressed orthologous pairs. (**D**) One-to-one orthologs censored in a regional correlation analysis; (^) signifies orthologous pairs censored in both proventriculus and anterior midgut, (†) signifies orthologous pairs censored in both anterior midgut and posterior midgut (*) signifies orthologous pairs censored in both proventriculus and posterior midgut. (**E**) 20 highest expressed transcripts per region, grouped by category/function; percentages reflect the proportion of the transcriptome comprised by the displayed genes.

The online version of this article includes the following figure supplement(s) for figure 7:

Figure supplement 1. A small number of highly/disparately transcribed genes obscure the transcriptional correlation between one-to-one orthologous genes in *Aedes aegypti* and *Anopheles gambiae* midgut regions.

Figure supplement 2. Phylogenetic analysis confirms disparately expressed one-to-one orthologous ribosomal proteins are sister sequences.

Figure supplement 3. A model for regionalized microbe selection and immune tolerance in the mosquito gut.

by Orthogroups Analysis were sister sequences and that the disparately expressed pairs of orthologous genes were widely distributed throughout the phylogenetic tree. We therefore conclude that evolutionary changes between the two species have reshuffled the expression of ribosomal proteins, significantly reducing the roles of some while increasing those of others.

Many of the genes censored in our correlation analysis were among the highest expressed genes in the three *Ae. aegypti* and *An. gambiae* midgut regions. A follow-up comparison of the 20 highest expressed genes in the midgut regions of the two species (*Figures 2C and 7E*) revealed striking differences in both amplitude and function. In the proventriculus, *An. gambiae* devoted a higher proportion of its transcriptome (63% vs 41% in *Ae. aegypti*) to the expression of the 20 highest expressed genes. The three highest expressed, including an uncharacterized gene (AGAP013543), the Tsetse EP gene (AGAP009313), and a cecropin (*CEC2*) together composed more than one third of the transcriptome. Both species' proventriculi shared expression of two MBF2 transcription factors (AAEL005428 and AAEL013885 in *Ae. aegypti*, AGAP011630 and AGAP000570 in *An. gambiae*), as well as the AMP

gambicin (*GAM1*). However, the *An. gambiae* proventriculus also expressed three cecropins (*CEC2*, *CEC1*, *CECB*), one defensin (*DEF1*) and a lysozyme (*LYSC7*) among its top 20 genes, together totaling more than 25% of its transcriptome. Cecropins (*CEC1*, *CEC2*) and a defensin (*DEF1*) were likewise among the 20 highest expressed genes in the *An. gambiae* but not the *Ae. aegypti* anterior midgut. In the posterior midgut, the highest expressed peptidases in *An. gambiae* were much lower expressed than their counterparts in *Ae. aegypti*, and the top 20 genes in the *An. gambiae* posterior midgut included two MBF2 transcription factors (AGAP000570, AGAP003713) that were absent among the highest expressed genes in the corresponding region in *Ae. aegypti*. Altogether, we find that while gut regional specializations have been broadly conserved between *Ae. aegypti* and *An. gambiae*, and while most one-to-one orthologous genes maintain closely correlated expression patterns, this correlation is lost among a few of the highest expressed genes in the midgut, creating a 'species signature' that differentiates the two midguts' transcriptomes.

Discussion

Aegypti-Atlas is a new resource which contextualizes and dissects the transcriptome of the sugar and blood-fed *Aedes aegypti* gut. Aegypti-Atlas will allow exploration of gene expression across the main parts/regions of the female mosquito body/gut in an easily accessible online database. In addition, we have performed a first analysis of this new dataset, examining how biological functions are distributed in the parts of the mosquito body and the regions of the gut, as well as how they change in the gut over the course of blood meal digestion. Our study demonstrates the importance of estimating investment, yield, and output in transcriptomic studies. In addition, this research has yielded new insights into such diverse areas as the regionalization of digestive function, the temporal patterning of peptidase transcription, the organization of immune defenses, and the similarities and differences between the guts of two hematophagous vectors.

Assessments for highest expressed genes and categorical investment/output are valuable complements to GO enrichment analysis

In bioinformatic analysis of large-scale transcriptomic data sets, GOEA is often the preferred method for determining the functional role of a tissue or body part, or for detecting functional differences across conditions. We have employed GOEA to derive valuable information about the enrichment of functions in mosquito body parts, gut regions, and blood-fed conditions. However, enrichment analysis alone cannot capture all the complexities of comparative transcriptomics, as it is primarily sensitive not to quantities of transcripts from a given category, but to quantities of genes belonging to that category. Furthermore, GOEA is substantially less robust in non-model species, owing to incomplete annotations which impact many key genes. We propose that complementary approaches, such as identifying highest expressed genes and examining transcriptomic profiles through the lenses of *investment* and *output*, can help to strengthen conclusions and capture dynamics which might otherwise be overlooked. By examining the twenty highest expressed genes in each body part and gut region, we were able to identify genes that account for large proportions of the transcriptome, but which were missed by enrichment analysis (e.g. *GAM1* in the proventriculus, MBF2 factors in the anterior midgut). Comparing the top-expressed genes across transcriptomes allowed us to see that some body parts are dominated by a small number of highly expressed genes (e.g. the gut) while others spread their transcriptome more equally over the genome (e.g. the ovaries). Comparing the relative expression of one-to-one orthologs between species allowed us to identify genes that are disproportionately expressed in one species (e.g. the Tsetse EP protein, a putative antimicrobial effector in the *An. gambiae* proventriculus) and to detect a reshuffling of expression among ribosomal proteins across the two species.

Another approach we adopted – summing transcripts belonging to a functional category and comparing the sums across body parts, regions, or timepoints – allowed us to assess relative investment in the given category. We confined the use of this method to categories of genes encoding proteins that are all presumed to have similar molecular function, narrowly defined (e.g. peptidases cleaving proteins, or antimicrobial peptides killing pathogens). The utility of evaluating regional or temporal specialization by summing transcripts (investment) as opposed to counting genes (enrichment), is evident when we consider that using GOEA, the highest expressed peptidase in the gut

carries no more weight than the fifth highest – despite a difference of over 165,000 TPM, or 16.5% of the total gut transcriptome. Using this investment approach, we were also able to break large categories (e.g. peptidases) into smaller categories (e.g. carboxypeptidases, dipeptidases, etc.) for a fine-grained picture of where and when each was most expressed. There is, however, an important caveat to the interpretation of cumulative analysis as it fails to account for differential transcriptional yields between profiled body parts, regions, or conditions. We overcame this caveat by weighting transcriptional investment by RNA yield to calculate estimated regional/temporal output of digestive transcripts, from which we were able to infer how certain tasks (*e.g.*, sugar digestion) are divided between the regions of the gut, as well as how peptidase transcript output changes over the course of blood meal digestion.

Spatial organization of gut function is conserved across mosquito species; few highly/disparately expressed genes drive transcriptomic differences

The *Ae. aegypti* gut is linearly divided into five anatomically distinct regions. With the caveat that the crop only admits sugar meals, ingested materials encounter each of these regions sequentially. The regions of the gut therefore possess an ordinal quality, corresponding to stages (either transient or prolonged) of the ingestion/digestion process. Through GOEA and quantitative evaluation of transcripts belonging to digestive categories, we discerned a clear pattern of strong investment in sugar digestion/absorption in the anterior midgut, as hypothesized by *Hecker, 1977*. Peptidases, by contrast, were predominantly expressed in the posterior midgut. These specializations were maintained under blood-fed conditions, and largely conserved in *An. gambiae* (*s.l.*) midguts. Other notable instances of conserved regional function are apparent in the proventriculus, where both species share enrichment of wingless signaling components (*Buchon et al., 2013*) and AMPS (*Hao et al., 2003*) with other dipteran species.

While the midgut regions of *Ae. aegypti* and *An. gambiae* shared many important functions, and their one-to-one orthologs maintained strikingly close transcriptional correlation, we noted that differences in the expression of a small number of highly expressed genes created large disparities in their overall categorical apportionment of transcripts. Most notably, the *An. gambiae* proventriculus and anterior midgut expressed far greater quantities of AMPs – possibly in response to some microbial presence – as well as a handful of highly expressed but poorly characterized genes, including an apparent ortholog of a tsetse midgut protein with a putative defensive function (*Haines et al., 2010*). Meanwhile, the *An. gambiae* posterior midgut expressed far fewer peptidase transcripts than its *Ae. aegypti* counterpart, and far more of an MBF2 family of transcription cofactors which, in the *Ae. aegypti* gut, was only prominent in the anterior regions of the sugar-fed midgut and, intriguingly, the blood-fed posterior midgut. In *Drosophila* and *Bombyx mori*, MBF2 factors have been shown to complex with the transcription factor FTZ-F1 and cofactor MBF1 to activate transcription of target genes (*Li et al., 1997*; *Liu et al., 2000*). We cannot say whether this role is conserved in *Ae. aegypti* and *An. gambiae*. However, the expression of MBF2 factors in both species outstripped the expression of their orthologs of *FTZ-F1* (AAEL019863, AAEL026810, AGAP005661) and *MBF1* (AAEL008768, AGAP004990) by several orders of magnitude in multiple conditions and/or gut regions (see Aegypti-Atlas website), suggesting MBF2 may not be confined to cofactoring FTZ-F1 in these species. Altogether, this analysis yielded multiple 'species signature' genes worthy of closer scrutiny and functional study.

The anterior midgut participates in the transcriptional response to blood feeding

The anterior midgut has been held to play little or no role in the process of blood meal digestion, as the blood bolus is sealed into the posterior midgut by the formation of the peritrophic matrix shortly after ingestion (*Billingsley, 1990*). However, late in the blood meal response (24 hr) we found that the anterior midgut had increased its investment in some categories of digestive enzymes and transporters (e.g. sphingomyelinases, sterol transporters, amylases/maltases, glucosidases, *Figure 4D*) as well as its overall transcriptional yield (*Figure 4E*). While the proportional output of the anterior midgut drops relative to the posterior midgut (*Figure 4F*), it still contributes approximately one third of the transcripts for sugar-digesting enzymes in the whole midgut (exclusive of proventriculus).

This finding is congruent with Billingsley and Hecker's observation that alpha-glucosidase activity was elevated in homogenates of *Anopheles stephensi* anterior midguts at 24 hr pbm (*Billingsley and Hecker, 1991*). As the majority of amylases/maltases and glucosidases in the *Ae. aegypti* genome possess signal peptides but lack transmembrane domains, it is possible that enzymes secreted in the anterior midgut are capable of diffusing into the extra-peritrophic area in the posterior midgut where they may participate in blood meal digestion.

Gut peptidases are dynamically up and downregulated in sequential transcriptional waves upon blood feeding

Peptidase expression in the *Ae. aegypti* gut unfolds in a series of phases, coordinated by rising and falling titers of JH and ecdysone. In the post-emergence/pre-vitellogenic phase, JH drives the transcription of a cohort of "early" peptidases, priming the gut for its first blood meal (*Bian et al., 2008*; *Jiang et al., 1997*; *Noriega et al., 2001*; *Noriega et al., 1997*). Within hours of blood feeding, translation of this pool of transcripts is initiated and peptidase activity begins to increase (*Felix et al., 1991*). Next, rising ecdysone titers drive the transcription of a "late" cohort of peptidases, which complete the digestive process (*He et al., 2021*). Finally, early peptidases are transcribed anew (*Bian et al., 2008*; *Noriega et al., 1996*) amid a postprandial surge of JH (*Hernández-Martínez et al., 2015*; *Lucas et al., 2015*; *Shapiro et al., 1986*; *Zhao et al., 2016*), restoring the gut to a state of readiness for its next blood meal.

Peptidase activity peaks somewhere between 18 and 36 hr pbm (*Gooding, 1966*; *Graf and Briegel, 1982*; *Noriega et al., 2002*), as do transcripts for some of the highest expressed 'late' phase peptidases (*Brackney et al., 2010*; *Isoe et al., 2009a*; *Isoe et al., 2009b*). However, a few publications have documented that some peptidases peak at earlier timepoints (*Brackney et al., 2010*; *Isoe et al., 2009b*; *Sanders et al., 2003*), and that at least one of these (*SPI*) is directly responsive to/ dependent on ecdysone signaling (*He et al., 2021*). Here, our genome-wide multi-timepoint series has demonstrated that these early-peaking genes are not isolated and exceptional but are, rather, members of large transcriptional cohorts. We show that peptidases in the blood-fed *Ae. aegypti* gut are expressed not only in two phases: 'early' (translational) and 'late' (transcriptional), but that the 'late' transcriptional response manifests in multiple successive waves or shifts. We have divided these into three ('rapid', 'intermediate', and 'delayed'), but because we only obtained RNAseq profiles for guts at baseline and three blood-fed timepoints, we can only speculate how many of these shifts are actually triggered over the course of blood meal digestion, and how synchronously any of the peptidases that our clustering analysis grouped together are actually regulated. Our timepoints were too broadly spaced to say at what time each gene reached its true peak, or whether peptidases that were apparently quiescent throughout blood meal digestion in fact participated in a shift that was too transient to be detected by our experimental design. Future work will likely uncover even more complexity than we have described.

The sharp temporal patterning of peptidase transcription in the blood-fed mosquito gut raises intriguing questions. How is this transcriptional choreography achieved? Is the transcription of all 'late' phase peptidases ecdysone-dependent, and are rising ecdysone levels sufficient to drive them? To what extent is the transcriptional induction of succeeding waves dependent on the activity of prior ones? Why are some 'delayed' peptidases so slow to show any transcriptional response when ecdysone titers are ascendent (e.g. AAEL008782, *Figure 3E*)? And what mechanism effects the precipitous drop in the 'rapid' cohort of induced peptidases (ascendent at 6 hr pbm) at a timepoint (24 hr pbm) when ecdysone titers are still high (e.g. AAEL013715) (*Hagedorn et al., 1975*)? What is the adaptive significance of the rapid up and down-regulation of so many genes of shared function? We observed that the gut's investment in aminopeptidases, dipeptidases, and carboxypeptidases has distinct temporal peaks, suggesting that polypeptides may be attacked more from one terminus or the other at different times over the course of digestion. However, most of the peptidases that participate in shift-changes are endopeptidases, and it is not clear how the sequential substitution of one set of endopeptidases for another – as opposed to simultaneous expression – helps to move the process of digestion forward. We speculate that rapidly changing conditions in the blood-fed gut may alter the stability and kinetics of specific peptidases, and that different peptidases predominate at different times because they are adapted to function in specific, transient conditions obtaining at corresponding stages of the digestive process.

The anterior and posterior midgut regions are, respectively, characterized by signs of immune activity and immune tolerance

In the midguts of both *Ae. aegypti* and *An. gambiae*, the majority of AMP transcripts are contributed by the proventriculus and anterior midgut (*Figure 5D*, *Figure 6—figure supplement 1B*) with minimal expression in the posterior midgut. Also, in both species, the transcription factor *caudal* which, in *Drosophila* (*Ryu et al., 2008*) and *An. gambiae* (*Clayton et al., 2013*) has been shown to repress AMP expression, is dominantly expressed in the posterior midgut. The expression of immune activating recognition proteins (*PGRPLC* and, in *Ae. aegypti*, *PGRP-LE*) is less profoundly patterned, but still displays some tropism toward the anterior midgut in both mosquitoes. Altogether, these patterns suggest that ingested microbes are subjected to heightened immune surveillance and antimicrobial activity in the first regions of the midgut but are thereafter well tolerated by the posterior midgut, where they may serve some mutualistic functions. We speculate that selection by gambicin in the proventriculus and anterior midgut of *Ae. aegypti* and *An. gambiae* may play an important role in determining the composition of the microbiota that seed the posterior midgut (*Figure 7—figure supplement 3A*).

Holotricin and gambicin are highly expressed under baseline conditions in mutually exclusive regions of the body

Our examination of the expression of immune effectors yielded the observation that at baseline, AMP expression in *Ae. aegypti* is dominated by two genes: holotricin and gambicin. We observed that holotricin predominated in the carcass, while gambicin was highly expressed in the gut and Malpighian tubules. While it is not uncommon for antimicrobial effectors to display strong tropisms to specific tissues, we are not aware of comparable instances where a single AMP exhibits such overwhelming transcriptional dominance. In future work, it might be interesting to compare the activity of holotricin and gambicin against different types of microbes (Gram-positive, Gram-negative, fungi, etc.), and to examine the effects that silencing these genes has on midgut communities, and on mosquitoes' survival in the contexts of oral and systemic infection.

Conclusion

In this manuscript we introduce Aegypti-Atlas, a repository of RNAseq data, and demonstrate how these data can yield insights into tissue function, the organization of digestive specializations, regulatory networks, and immune effectors, as well as the changes wrought by diet and evolution in mosquitoes. This resource may also be useful for the creation of functional genomic tools (e.g. tissue-specific expression systems) which will afford researchers a greater degree of control in experiments and allow for more confidence in the interpretation of results. It is our hope that Aegypti-Atlas will be valuable to other investigators in many areas of mosquito biology.

Materials and methods

Mosquito provenance and rearing

For all experiments, we used 5- to 12-day-old female mosquitoes of the Thai strain (*Ae. aegypti*) or the G3 strain (*An. gambiae, s.l.*) kindly provided by Laura Harrington. The Thai colony has been maintained at Cornell University since 2011 and is outcrossed annually with field-collected eggs from Thailand to relieve inbreeding and maintain field relevance. Mosquitoes were reared at a density of 200 larvae per 1 liter tray. *Ae. aegypti* received 720 mg of fish food (Hikari #04428). *An. gambiae* received 50 mg of ground fish food per day from day 1 to 4 of development, and 150 mg per day thereafter until pupation. Adults were maintained in humidified chambers (RH 75% ± 5%) at 29 °C on a diet of 10% sucrose ad libitum. Females were reared and maintained in approximately equal proportion with males for a minimum of 5 days prior to blood-feeding and/or dissection and were presumed to be mated.

Blood feeding and mock blood feeding

For all RNA-seq experiments involving blood-fed guts, mosquitoes were starved for twenty-four hours, then blood-fed on live chickens without anesthetic in accordance with Cornell University

IACUC approved protocol #01–56. For our extended blood feeding time series by RT-qPCR, mosquitoes were fed through a membrane on rooster blood treated with the anticoagulant sodium citrate (Lampire Biological Laboratories, Pipersville, PA, catalogue #7208806). For experiments where RNA yield was estimated in guts and midgut regions under blood-fed conditions, we employed an artificial formulation (SkitoSnack) (*Gonzales et al., 2018*) fed through a membrane. This formulation was used in order to avoid any skewing of yield which might result from RNA content in a natural blood meal. For RNA yield experiments, boluses were removed during dissection.

Dissections and RNA extraction

Mosquitoes were sacrificed by submersion in 70% ethanol and dissected in sterile PBS. For body part samples, the last 2–3 abdominal segments were removed to eliminate the sperm-containing spermatheca. Whole guts, Malpighian tubules, and ovaries were then dissected out. The residual carcass was divided into head, thorax, and abdomen. A small anterior section of the thoracic carcass containing the salivary glands was excised to exclude salivary gland transcripts from thoracic profiles. For RNAseq experiments, a minimum of 10–20 individuals per replicate was used for large body parts (e.g. whole body, abdominal carcass, whole guts) and ≥100 individuals per replicate for small gut regions (e.g. crop, proventriculus). For additional information about replicates, see *Supplementary file 9*. For RT-qPCR time-series experiments, a minimum of 10 individuals was dissected per replicate. For all experiments, a minimum of three replicates was prepared per part and/or condition.

All samples for RNAseq were homogenized in 600 µl TRIzol and stored at –80 °C. For RNAseq experiments, RNA was extracted via a modified phenol-chloroform method (*Troha et al., 2018*). In brief, samples in TRIzol were thawed and combined with 120 µl of chloroform, followed by vortexing and centrifugation for 15 min at 12,000 g, 4 °C. Between 200 and 250 µl of aqueous supernatant was removed to a fresh tube, where it was mixed first with 700 µl Buffer RLT (Qiagen RNeasy kit, cat #74004) then 500 µl 100% ethanol. This mixture was passed through a RNeasy spin column (700 µl at a time) with centrifugations of 20 s at 10,000 g, followed by a final spin for 1 min at 10,000 g to dry the column. The membrane was then washed twice with 500 µl of Buffer RPE (20 second centrifugation, 10,000 g), followed by a 1 min centrifugation at 16,000 g to dry the column. Samples were eluted in 30 µl of RNase-free water, applied directly to the membrane.

For RNA yield experiments, dissected gut regions were pooled (10 for posterior guts, 20–50 for other regions) in 1 mL Trizol, and the number in each sample was recorded. Samples were then extracted by a standard phenol-chloroform procedure. Exactly 400 µl of aqueous supernatant was recovered from each tube, and RNA was precipitated overnight with isopropanol at –20 °C. Pellets were resuspended in 20–50 µl of nuclease-free water, and concentrations were measured by Qubit. Yield was then calculated as follows: (concentration * suspension volume / # of dissected regions).

Library preparation and sequencing

Libraries were prepared using the Quantseq 3' mRNA-seq prep kit from Lexogen according to the manufacturer's instructions. Sample quality was evaluated before and after library preparation using a fragment analyzer (Advanced Analytical). Libraries were pooled and sequenced on the Illumina Nextseq 500 platform using standard protocols for 75 bp single-end read sequencing at the Cornell Life Sciences Sequencing Core. Sequences have been deposited on NCBI (BioProject ID: PRJNA789580).

Analysis pipelines

Quality control of raw reads was performed with fastqc (https://github.com/s-andrews/FastQC) (*Andrews et al., 2020*) and reads were trimmed by BBMap (https://jgi.doe.gov/data-and-tools/bbtools/) and then mapped to the *Ae. aegypti* transcriptome or the *An. gambiae* transcriptome *Aedes aegypti* LVP_AGWG *Aaeg*L5.2 and *Anopheles gambiae* PEST *Agam*P4.12, VectorBase, https://www.vectorbase.org/ (*Giraldo-Calderón et al., 2022*) using Salmon version 0.9.1 with default parameters. Salmon's transcript-level quantifications were then aggregated to the gene level for gene-level differential expression analysis with the "tximport" package on the R version 3.5. DEseq2 (*Love et al., 2014*) was used for differential expression analysis and PCAs were performed using custom R scripts (available upon request). GOEA was performed using the topGO package (classic Fisher method). Scripts for Salmon, topGO, Orthogroups analysis, and DESeq2 analysis, as well as DESeq2 outputs

can be accessed at GitHub (https://github.com/bingdun/RNAseq_analysis_of_Aedes_aegypti; *Bing, 2022*).

All phylogenetic trees were constructed using Geneious software version R11, with the following settings. Alignment type: Global alignment, Cost Matrix: Blosum62, Genetic Distance Model: Jukes-Cantor, Tree Build Method: Neighbor-Joining. Clustering was performed using the Pheatmap package in R with gene expression scaled by row. To calculate z-scores, we first censored all genes expressed at less than 2 TPM in the relevant body parts or gut regions, then employing the following expression for each gene in each body part or gut region: $z = (x - \mu)/s$, where x is the expression in the body part or gut region, μ represents the mean expression of all body parts or gut regions, and s is the standard deviation. The peptide sequences of *Aaeg*L5.2 and *Agam*P4.12 were used to find orthogroups and orthologs in the two species using OrthoFinder v2.3 (*Emms and Kelly, 2019*). The longest transcript variant of each gene was extracted to run the OrthoFinder with the following options: '-M msa -S blast -A muscle -T iqtree'.

Calculating the 'mosquito equation'

If X is defined as the expression of a specific gene in a given body part (in TPM), and the part's initial stands for a scaling factor equal to the proportion of transcripts in the whole body derived from that part, the relationship between the expression of that gene in each body part versus in the whole body can be expressed as.

$$(X_{Head} * H) + (X_{Thorax} * T) + (X_{Abdomen} * A) + (X_{Gut} * G) + (X_{MT} * MT) + (X_{Ovaries} * O) = X_{WholeBody}$$

where $H + T + A + G + MT + O = 100\%$ Whole Body expression.

For marker genes, the expression in other body parts may be considered to be negligible, (i.e. for a head marker, X_{Thorax}, $X_{Abdomen}$, X_{Gut}, X_{MT}, and $X_{Ovaries}$ are all ≈ 0). Therefore.

$$X_{Head} * H \approx X_{WholeBody}$$

We solved for the estimated scaling factor for each of our qualifying markers, then averaged the values we obtained for each body part to create an estimate of the percent of transcripts in the whole body that are contributed by each part.

RT-qPCR

For all RT-qPCR amplifications, RNA was pretreated with DNase (Quantabio 95150–01 K) and cDNA was prepared using the qScript cDNA synthesis kit (Quantabio 95047–100) and amplified using Perfecta SYBR Green Fastmix (Quantabio 95072–012) in a Bio-Rad CFX-Connect Instrument. The primer sequences used in this study are available in *Supplementary file 10*.

Additional information

Author contributions

Bretta Hixson, Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Validation, Visualization, Writing - original draft, Writing – review and editing; Xiao-Li Bing, Xiaowei Yang, Data curation, Formal analysis, Writing – review and editing; Alessandro Bonfini, Peter Nagy, Formal analysis, Investigation; Nicolas Buchon, Conceptualization, Data curation, Formal analysis, Funding acquisition, Project administration, Resources, Supervision, Writing - original draft, Writing – review and editing

Author ORCIDs

Bretta Hixson http://orcid.org/0000-0001-7983-5104
Alessandro Bonfini http://orcid.org/0000-0001-6642-8665
Peter Nagy http://orcid.org/0000-0002-5053-0646
Nicolas Buchon http://orcid.org/0000-0003-3636-8387

Decision letter and Author response

Decision letter https://doi.org/10.7554/eLife.76132.sa1
Author response https://doi.org/10.7554/eLife.76132.sa2

Additional files

Supplementary files

• Supplementary file 1. TPMs for putative tissue markers in major body parts of *Aedes aegypti*.

• Supplementary file 2. Gene ontology categories enriched in Aedes aegypti body parts, gut regions, and blood-fed guts; and in Anopheles gambiae (s.l.) midgut regions.

• Supplementary file 3. Index of VectorBase IDs for genes named in the text.

• Supplementary file 4. TPMs for digestive enzymes and transporters expressed in the Aedes aegypti gut.

• Supplementary file 5. Relative expression (investment) and TPMs for immune-related genes in Aedes aegypti body parts, gut regions, and blood-fed guts; and in Anopheles gambiae (s.l.) midgut regions.

• Supplementary file 6. TPMs for digestive enzymes and transporters expressed in the Aedes aegypti midgut.

• Supplementary file 7. TPMs for digestive enzymes and transporters expressed in the Anopheles gambiae (s.l.) midgut.

• Supplementary file 8. Output of orthogroups analysis relating genes from Aedes aegypti and *Anopheles gambiae*.

• Supplementary file 9. Details of replicates for RNAseq experiments in *Aedes aegypti* and Anopheles gambiae (s.l.).

• Supplementary file 10. Sequences of primers for RTqPCR assays.

Data availability

Data have been submitted to NCBI. The BioProject ID is PRJNA789580.

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Hixson B, Bing XL, Yang X, Bonfini A, Nagy P, Buchon N	2022	A transcriptomic atlas of Aedes aegypti reveals detailed functional organization of major body parts and gut regional specializations in sugar-fed and blood-fed adult females	https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA789580	NCBI BioProject, PRJNA789580

References

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

Andrews S, Mahé F, Yi H, Brokamp J, mareq. 2020. FastQC. 36532ce. GitHub. https://github.com/s-andrews/FastQC

Angleró-Rodríguez YI, MacLeod HJ, Kang S, Carlson JS, Jupatanakul N, Dimopoulos G. 2017. Aedes aegypti Molecular Responses to Zika Virus: Modulation of Infection by the Toll and Jak/Stat Immune Pathways and Virus Host Factors. *Frontiers in Microbiology* **8**:02050. DOI: https://doi.org/10.3389/fmicb.2017.02050, PMID: 29109710

Baker DA, Nolan T, Fischer B, Pinder A, Crisanti A, Russell S. 2011. A comprehensive gene expression atlas of sex- and tissue-specificity in the malaria vector, Anopheles gambiae. *BMC Genomics* **12**:296. DOI: https://doi.org/10.1186/1471-2164-12-296, PMID: 21649883

Barillas-Mury C, Graf R, Hagedorn HH, Wells MA. 1991. cDNA and deduced amino acid sequence of a blood meal-induced trypsin from the mosquito, Aedes aegypti. *Insect Biochemistry* **21**:825–831. DOI: https://doi.org/10.1016/0020-1790(91)90089-W

Benguettat O, Jneid R, Soltys J, Loudhaief R, Brun-Barale A, Osman D, Gallet A. 2018. The DH31/CGRP enteroendocrine peptide triggers intestinal contractions favoring the elimination of opportunistic bacteria. *PLOS Pathogens* **14**:e1007279. DOI: https://doi.org/10.1371/journal.ppat.1007279, PMID: 30180210

Bian G, Raikhel AS, Zhu J. 2008. Characterization of a juvenile hormone-regulated chymotrypsin-like serine protease gene in Aedes aegypti mosquito. *Insect Biochemistry and Molecular Biology* **38**:190–200. DOI: https://doi.org/10.1016/j.ibmb.2007.10.008, PMID: 18207080

Billingsley PF. 1990. The Midgut Ultrastructure of Hematophagous Insects. *Annual Review of Entomology* **35**:219–248. DOI: https://doi.org/10.1146/annurev.en.35.010190.001251

Billingsley PF, Hecker H. 1991. Blood digestion in the mosquito, Anopheles stephensi Liston (Diptera: Culicidae): activity and distribution of trypsin, aminopeptidase, and alpha-glucosidase in the midgut. *Journal of Medical Entomology* **28**:865–871. DOI: https://doi.org/10.1093/jmedent/28.6.865, PMID: 1770523

Bing X. 2022. RNAseq_analysis_of_Aedes_aegypti. 0755da9. GitHub. https://github.com/bingdun/RNAseq_analysis_of_Aedes_aegypti

Brackney DE, Isoe J, W C B, Zamora J, Foy BD, Miesfeld RL, Olson KE. 2010. Expression profiling and comparative analyses of seven midgut serine proteases from the yellow fever mosquito, Aedes aegypti. *Journal of Insect Physiology* **56**:736–744. DOI: https://doi.org/10.1016/j.jinsphys.2010.01.003, PMID: 20100490

Buchon N., Broderick NA, Poidevin M, Pradervand S, Lemaitre B. 2009. *Drosophila* intestinal response to bacterial infection: activation of host defense and stem cell proliferation. *Cell Host & Microbe* **5**:200–211. DOI: https://doi.org/10.1016/j.chom.2009.01.003, PMID: 19218090

Buchon N, Osman D, David FPA, Fang HY, Boquete JP, Deplancke B, Lemaitre B. 2013. Morphological and molecular characterization of adult midgut compartmentalization in *Drosophila*. *Cell Reports* **3**:1725–1738. DOI: https://doi.org/10.1016/j.celrep.2013.04.001, PMID: 23643535

Buchon N., Osman D. 2015. All for one and one for all: Regionalization of the *Drosophila* intestine. *Insect Biochemistry and Molecular Biology* **67**:2–8. DOI: https://doi.org/10.1016/j.ibmb.2015.05.015, PMID: 26044368

Chintapalli VR, Wang J, Dow JAT. 2007. Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. *Nature Genetics* **39**:715–720. DOI: https://doi.org/10.1038/ng2049, PMID: 17534367

Choi YJ, Fuchs JF, Mayhew GF, Yu HE, Christensen BM. 2012. Tissue-enriched expression profiles in Aedes aegypti identify hemocyte-specific transcriptome responses to infection. *Insect Biochemistry and Molecular Biology* **42**:729–738. DOI: https://doi.org/10.1016/j.ibmb.2012.06.005, PMID: 22796331

Clayton AM, Cirimotich CM, Dong Y, Dimopoulos G. 2013. Caudal is a negative regulator of the Anopheles IMD Pathway that controls resistance to *Plasmodium falciparum* infection. *Developmental & Comparative Immunology* **39**:323–332. DOI: https://doi.org/10.1016/j.dci.2012.10.009, PMID: 23178401

Clements AN. 1992. The Biology of Mosquitoes Volume 1: Development, Nutrition and Reproduction. London: Chapman & Hall.

Cui Y, Franz AWE. 2020. Heterogeneity of midgut cells and their differential responses to blood meal ingestion by the mosquito, Aedes aegypti. *Insect Biochemistry and Molecular Biology* **127**:103496. DOI: https://doi.org/10.1016/j.ibmb.2020.103496, PMID: 33188922

Dong S, Behura SK, Franz AWE. 2017. The midgut transcriptome of Aedes aegypti fed with saline or protein meals containing chikungunya virus reveals genes potentially involved in viral midgut escape. *BMC Genomics* **18**:e6. DOI: https://doi.org/10.1186/s12864-017-3775-6, PMID: 28506207

Dutta D, Buchon N, Xiang J, Edgar BA. 2015. Regional Cell Specific RNA Expression Profiling of FACS Isolated *Drosophila* Intestinal Cell Populations. *Current Protocols in Stem Cell Biology* **34**:2F. DOI: https://doi.org/10.1002/9780470151808.sc02f02s34, PMID: 26237570

Dzaki N, Ramli KN, Azlan A, Ishak IH, Azzam G. 2017. Evaluation of reference genes at different developmental stages for quantitative real-time PCR in Aedes aegypti. *Scientific Reports* **7**:43618. DOI: https://doi.org/10.1038/srep43618, PMID: 28300076

Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology* **20**:238. DOI: https://doi.org/10.1186/s13059-019-1832-y, PMID: 31727128

Felix CR, Betschart B, Billingsley PF, Freyvogel TA. 1991. Post-feeding induction of trypsin in the midgut of Aedes aegypti L. (Diptera: Culicidae) is separable into two cellular phases. *Insect Biochemistry* **21**:197–203. DOI: https://doi.org/10.1016/0020-1790(91)90050-O

Foo ACY, Thompson PM, Chen SH, Jadi R, Lupo B, DeRose EF, Arora S, Placentra VC, Premkumar L, Perera L, Pedersen LC, Martin N, Mueller GA. 2021. The mosquito protein AEG12 displays both cytolytic and antiviral properties via a common lipid transfer mechanism. *PNAS* **118**:e2019251118. DOI: https://doi.org/10.1073/pnas.2019251118, PMID: 33688047

Giraldo-Calderón GI, Harb OS, Kelly SA, Rund SS, Roos DS, McDowell MA. 2022. VectorBase.org updates: bioinformatic resources for invertebrate vectors of human pathogens and related organisms. *Current Opinion in Insect Science* **50**:100860. DOI: https://doi.org/10.1016/j.cois.2021.11.008, PMID: 34864248

Gonzales KK, Rodriguez SD, Chung H-N, Kowalski M, Vulcan J, Moore EL, Li Y, Willette SM, Kandel Y, Van Voorhies WA, Holguin FO, Hanley KA, Hansen IA. 2018. The Effect of SkitoSnack, an Artificial Blood Meal Replacement, on Aedes aegypti Life History Traits and Gut Microbiota. *Scientific Reports* **8**:1–14. DOI: https://doi.org/10.1038/s41598-018-29415-5, PMID: 30038361

Gooding RH. 1966. Physiological Aspects Of Digestion Of The Blood Meal By Aedes Aegypti (Linnaeus) And Culex Fatigans Wiedemann1, 2. *Journal of Medical Entomology* **3**:53–60. DOI: https://doi.org/10.1093/jmedent/3.1.53, PMID: 5941568

Graf R, Briegel H. 1982. Comparison between aminopeptidase and trypsin activity in blood-fed females of Aedes aegypti. *Revue Suisse de Zoologie* **89**:845–850. DOI: https://doi.org/10.5962/bhl.part.82479

Graf R, Lea AO, Briegel H. 1998. A temporal profile of the endocrine control of trypsin synthesis in the yellow fever mosquito, Aedes aegypti. *Journal of Insect Physiology* **44**:451–454. DOI: https://doi.org/10.1016/S0022-1910(98)00042-0, PMID: 12770164

Gusmão DS, Santos AV, Marini DC, Bacci M, Berbert-Molina MA, Lemos FJA. 2010. Culture-dependent and culture-independent characterization of microorganisms associated with Aedes aegypti (Diptera: Culicidae) (L.) and dynamics of bacterial colonization in the midgut. *Acta Tropica* **115**:275–281. DOI: https://doi.org/10.1016/j.actatropica.2010.04.011, PMID: 20434424

Hagedorn HH, O'Connor JD, Fuchs MS, Sage B, Schlaeger DA, Bohm MK. 1975. The ovary as a source of alpha-ecdysone in an adult mosquito. *PNAS* **72**:3255–3259. DOI: https://doi.org/10.1073/pnas.72.8.3255, PMID: 1059110

Haines LR, Lehane SM, Pearson TW, Lehane MJ. 2010. Tsetse EP protein protects the fly midgut from trypanosome establishment. *PLOS Pathogens* **6**:e1000793. DOI: https://doi.org/10.1371/journal.ppat.1000793, PMID: 20221444

Hao Z, Kasumba I, Aksoy S. 2003. Proventriculus (cardia) plays a crucial role in immunity in tsetse fly (Diptera: Glossinidiae). *Insect Biochemistry and Molecular Biology* **33**:1155–1164. DOI: https://doi.org/10.1016/j.ibmb.2003.07.001, PMID: 14563366

He YZ, Ding Y, Wang X, Zou Z, Raikhel AS. 2021. E93 confers steroid hormone responsiveness of digestive enzymes to promote blood meal digestion in the midgut of the mosquito Aedes aegypti. *Insect Biochemistry and Molecular Biology* **134**:103580. DOI: https://doi.org/10.1016/j.ibmb.2021.103580, PMID: 33901693

Hecker H. 1977. Structure and function of midgut epithelial cells in culicidae mosquitoes (insecta, diptera). *Cell and Tissue Research* **184**:321–341. DOI: https://doi.org/10.1007/BF00219894, PMID: 21747

Hernández-Martínez S, Rivera-Perez C, Nouzova M, Noriega FG. 2015. Coordinated changes in JH biosynthesis and JH hemolymph titers in Aedes aegypti mosquitoes. *Journal of Insect Physiology* **72**:22–27. DOI: https://doi.org/10.1016/j.jinsphys.2014.11.003, PMID: 25445664

Hyde J, Correa MA, Hughes GL, Steven B, Brackney DE. 2020. Limited influence of the microbiome on the transcriptional profile of female Aedes aegypti mosquitoes. *Scientific Reports* **10**:10880. DOI: https://doi.org/10.1038/s41598-020-67811-y, PMID: 32616765

Isoe J, Rascón AA, Kunz S, Miesfeld RL. 2009a. Molecular genetic analysis of midgut serine proteases in Aedes aegypti mosquitoes. *Insect Biochemistry and Molecular Biology* **39**:903–912. DOI: https://doi.org/10.1016/j.ibmb.2009.10.008, PMID: 19883761

Isoe J, Zamora J, Miesfeld RL. 2009b. Molecular analysis of the Aedes aegypti carboxypeptidase gene family. *Insect Biochemistry and Molecular Biology* **39**:68–73. DOI: https://doi.org/10.1016/j.ibmb.2008.09.006, PMID: 18977440

Iwamura T, Guzman-Holst A, Murray KA. 2020. Accelerating invasion potential of disease vector Aedes aegypti under climate change. *Nature Communications* **11**:2130. DOI: https://doi.org/10.1038/s41467-020-16010-4, PMID: 32358588

Jiang Q, Hall M, Noriega FG, Wells M. 1997. cDNA cloning and pattern of expression of an adult, female-specific chymotrypsin from Aedes aegypti midgut. *Insect Biochemistry and Molecular Biology* **27**:283–289. DOI: https://doi.org/10.1016/s0965-1748(97)00001-5, PMID: 9134710

Kato N, Mueller CR, Fuchs JF, Wessely V, Lan Q, Christensen BM. 2006. Regulatory mechanisms of chitin biosynthesis and roles of chitin in peritrophic matrix formation in the midgut of adult Aedes aegypti. *Insect Biochemistry and Molecular Biology* **36**:1–9. DOI: https://doi.org/10.1016/j.ibmb.2005.09.003, PMID: 16360944

Kumar S, Molina-Cruz A, Gupta L, Rodrigues J, Barillas-Mury C. 2010. A peroxidase/dual oxidase system modulates midgut epithelial immunity in Anopheles gambiae. *Science (New York, N.Y.)* **327**:1644–1648. DOI: https://doi.org/10.1126/science.1184008, PMID: 20223948

Leader DP, Krause SA, Pandit A, Davies SA, Dow JAT. 2018. FlyAtlas 2: A new version of the *Drosophila melanogaster* expression atlas with RNA-Seq, miRNA-Seq and sex-specific data. *Nucleic Acids Research* **46**:D809–D815. DOI: https://doi.org/10.1093/nar/gkx976, PMID: 29069479

League GP, Baxter LL, Wolfner MF, Harrington LC. 2019. Male accessory gland molecules inhibit harmonic convergence in the mosquito Aedes aegypti. *Current Biology* **29**:R196–R197. DOI: https://doi.org/10.1016/j.cub.2019.02.005, PMID: 30889386

Lemaitre B, Miguel-Aliaga I. 2013. The digestive tract of *Drosophila melanogaster*. *Annual Review of Genetics* **47**:377–404. DOI: https://doi.org/10.1146/annurev-genet-111212-133343, PMID: 24016187

Li FQ, Takemaru KI, Goto M, Ueda H, Handa H, Hirose S. 1997. Transcriptional activation through interaction of MBF2 with TFIIA. *Genes to Cells* **2**:143–153. DOI: https://doi.org/10.1046/j.1365-2443.1997.1090306.x, PMID: 9167971

Li Y, Piermarini PM, Esquivel CJ, Price DP, Drumm HE, Schilkey FD, Hansen IA. 2017. RNA-Seq Comparison of Larval and Adult Malpighian Tubules of the Yellow Fever Mosquito Aedes aegypti Reveals Life Stage-Specific Changes in Renal Function. *Frontiers in Physiology* **8**:283. DOI: https://doi.org/10.3389/fphys.2017.00283, PMID: 28536536

Liu QX, Ueda H, Hirose S. 2000. MBF2 is a tissue- and stage-specific coactivator that Is regulated at the step of nuclear transport in the silkworm Bombyx mori. *Developmental Biology* **225**:437–446. DOI: https://doi.org/10.1006/dbio.2000.9836, PMID: 10985861

Liu J, Liu Y, Nie K, Du S, Qiu J, Pang X, Wang P, Cheng G. 2016. Flavivirus NS1 protein in infected host sera enhances viral acquisition by mosquitoes. *Nature Microbiology* **1**:16087. DOI: https://doi.org/10.1038/nmicrobiol.2016.87, PMID: 27562253

Liu Y, Lillepold K, Semenza JC, Tozan Y, Quam MBM, Rocklöv J. 2020. Reviewing estimates of the basic reproduction number for dengue, Zika and chikungunya across global climate zones. *Environmental Research* **182**:109114. DOI: https://doi.org/10.1016/j.envres.2020.109114, PMID: 31927301

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**:1–21. DOI: https://doi.org/10.1186/s13059-014-0550-8, PMID: 25516281

Lucas KJ, Roy S, Ha J, Gervaise AL, Kokoza VA, Raikhel AS. 2015. MicroRNA-8 targets the Wingless signaling pathway in the female mosquito fat body to regulate reproductive processes. *PNAS* **112**:1440–1445. DOI: https://doi.org/10.1073/pnas.1424408112, PMID: 25605933

Martínez-Barnetche J, Gómez-Barreto RE, Ovilla-Muñoz M, Téllez-Sosa J, García López DE, Dinglasan RR, Ubaida Mohien C, MacCallum RM, Redmond SN, Gibbons JG, Rokas A, Machado CA, Cazares-Raga FE, González-Cerón L, Hernández-Martínez S, Rodríguez López MH. 2012. Transcriptome of the adult female malaria mosquito vector Anopheles albimanus. *BMC Genomics* **13**:1–17. DOI: https://doi.org/10.1186/1471-2164-13-207, PMID: 22646700

Matthews BJ, Dudchenko O, Kingan SB, Koren S, Antoshechkin I, Crawford JE, Glassford WJ, Herre M, Redmond SN, Rose NH, Weedall GD, Wu Y, Batra SS, Brito-Sierra CA, Buckingham SD, Campbell CL, Chan S, Cox E, Evans BR, Fansiri T, et al. 2018. Improved reference genome of Aedes aegypti informs arbovirus vector control. *Nature* **563**:501–507. DOI: https://doi.org/10.1038/s41586-018-0692-z, PMID: 30429615

Moyes CL, Vontas J, Martins AJ, Ng LC, Koou SY, Dusfour I, Raghavendra K, Pinto J, Corbel V, David J-P, Weetman D, Sinnis P. 2017. Contemporary status of insecticide resistance in the major Aedes vectors of arboviruses infecting humans. *PLOS Neglected Tropical Diseases* **11**:e0005625. DOI: https://doi.org/10.1371/journal.pntd.0005625, PMID: 28727779

Nag DK, Dieme C, Lapierre P, Lasek-Nesselquist E, Kramer LD. 2021. RNA-Seq analysis of blood meal induced gene-expression changes in Aedes aegypti ovaries. *BMC Genomics* **22**:e551. DOI: https://doi.org/10.1186/s12864-021-07551-z, PMID: 34044772

Noriega FG, Pennington JE, Barillas-Mury C, Wang XY, Wells MA. 1996. Aedes aegypti midgut early trypsin is post-transcriptionally regulated by blood feeding. *Insect Molecular Biology* **5**:25–29. DOI: https://doi.org/10.1111/j.1365-2583.1996.tb00037.x, PMID: 8630532

Noriega FG, Shah DK, Wells MA. 1997. Juvenile hormone controls early trypsin gene transcription in the midgut of Aedes aegypti. Insect Molecular Biology **6**:63–66. DOI: https://doi.org/10.1046/j.1365-2583.1997.00154.x, PMID: 9013256

Noriega FG, Edgar KA, Goodman WG, Shah DK, Wells MA. 2001. Neuroendocrine factors affecting the steady-state levels of early trypsin mRNA in Aedes aegypti. *Journal of Insect Physiology* **47**:515–522. DOI: https://doi.org/10.1016/s0022-1910(00)00145-1, PMID: 11166316

Noriega FG, Edgar KA, Bechet R, Wells MA. 2002. Midgut exopeptidase activities in Aedes aegypti are induced by blood feeding. *In Journal of Insect Physiology* **48**:e2. DOI: https://doi.org/10.1016/S0022-1910(01)00165-2

Olmo RP, Ferreira AGA, Izidoro-Toledo TC, Aguiar ERGR, de Faria IJS, de Souza KPR, Osório KP, Kuhn L, Hammann P, de Andrade EG, Todjro YM, Rocha MN, Leite THJF, Amadou SCG, Armache JN, Paro S, de Oliveira CD, Carvalho FD, Moreira LA, Marois E, et al. 2018. Control of dengue virus in the midgut of Aedes aegypti by ectopic expression of the dsRNA-binding protein Loqs2. *Nature Microbiology* **3**:1385–1393. DOI: https://doi.org/10.1038/s41564-018-0268-6, PMID: 30374169

Powell JR. 2018. Mosquito-Borne Human Viral Diseases: Why Aedes aegypti? *The American Journal of Tropical Medicine and Hygiene* **98**:1563–1565. DOI: https://doi.org/10.4269/ajtmh.17-0866

Ptitsyn AA, Reyes-Solis G, Saavedra-Rodriguez K, Betz J, Suchman EL, Carlson JO, Black WC. 2011. Rhythms and synchronization patterns in gene expression in the Aedes aegypti mosquito. *BMC Genomics* **12**:153. DOI: https://doi.org/10.1186/1471-2164-12-153, PMID: 21414217

Ranson H, N'guessan R, Lines J, Moiroux N, Nkuni Z, Corbel V. 2011. Pyrethroid resistance in African anopheline mosquitoes: what are the implications for malaria control? *Trends in Parasitology* **27**:91–98. DOI: https://doi.org/10.1016/j.pt.2010.08.004, PMID: 20843745

Raquin V, Merkling SH, Gausson V, Moltini-Conclois I, Frangeul L, Varet H, Dillies MA, Saleh MC, Lambrechts L. 2017. Individual co-variation between viral RNA load and gene expression reveals novel host factors during early dengue virus infection of the Aedes aegypti midgut. *PLOoS Neglected Tropical Diseases* **11**:6152. DOI: https://doi.org/10.1371/journal.pntd.0006152, PMID: 29261661

Ryu JH, Kim SH, Lee HY, Bai JY, Nam YD, Bae JW, Lee DG, Shin SC, Ha EM, Lee WJ. 2008. Innate immune homeostasis by the homeobox gene caudal and commensal-gut mutualism in *Drosophila. Science (New York, N.Y.)* **319**:777–782. DOI: https://doi.org/10.1126/science.1149357, PMID: 18218863

Sanders HR, Evans AM, Ross LS, Gill SS. 2003. Blood meal induces global changes in midgut gene expression in the disease vector, Aedes aegypti. *Insect Biochemistry and Molecular Biology* **33**:1105–1122. DOI: https://doi.org/10.1016/s0965-1748(03)00124-3, PMID: 14563362

Shao L, Devenport M, Fujioka H, Ghosh A, Jacobs-Lorena M. 2005. Identification and characterization of a novel peritrophic matrix protein, Ae-Aper50, and the microvillar membrane protein, AEG12, from the mosquito, Aedes aegypti. *Insect Biochemistry and Molecular Biology* **35**:947–959. DOI: https://doi.org/10.1016/j.ibmb.2005.03.012, PMID: 15978997

Shapiro AB, Wheelock GD, Hagedorn HH, Baker FC, Tsai LW, Schooley DA. 1986. Juvenile hormone and juvenile hormone esterase in adult females of the mosquito Aedes aegypti. *Journal of Insect Physiology* **32**:867–877. DOI: https://doi.org/10.1016/0022-1910(86)90102-2

Siomi MC, Sato K, Pezic D, Aravin AA. 2011. PIWI-interacting small RNAs: the vanguard of genome defence. *Nature Reviews. Molecular Cell Biology* **12**:246–258. DOI: https://doi.org/10.1038/nrm3089, PMID: 21427766

Sreenivasamurthy SK, Madugundu AK, Patil AH, Dey G, Mohanty AK, Kumar M, Patel K, Wang C, Kumar A, Pandey A, Prasad TSK. 2017. Mosquito-Borne Diseases and Omics: Tissue-Restricted Expression and Alternative Splicing Revealed by Transcriptome Profiling of Anopheles stephensi. *Omics* **21**:488–497. DOI: https://doi.org/10.1089/omi.2017.0073, PMID: 28708456

Troha K, Im JH, Revah J, Lazzaro BP, Buchon N. 2018. Comparative transcriptomics reveals CrebA as a novel regulator of infection tolerance in *D. melanogaster. PLOS Pathogens* **14**:e1006847. DOI: https://doi.org/10.1371/journal.ppat.1006847, PMID: 29394281

Vedururu RK, Neave MJ, Tachedjian M, Klein MJ, Gorry PR, Duchemin JB, Paradkar PN. 2019. RNASeq Analysis of Aedes albopictus Mosquito Midguts after Chikungunya Virus Infection. *Viruses* **11**:513. DOI: https://doi.org/10.3390/v11060513, PMID: 31167461

Vlachou D, Schlegelmilch T, Christophides GK, Kafatos FC. 2005. Functional Genomic Analysis of Midgut Epithelial Responses in Anopheles during Plasmodium Invasion. *Current Biology* **15**:1185–1195. DOI: https://doi.org/10.1016/j.cub.2005.06.044, PMID: 16005290

Warr E, Aguilar R, Dong Y, Mahairaki V, Dimopoulos G. 2007. Spatial and sex-specific dissection of the Anopheles gambiae midgut transcriptome. *BMC Genomics* **8**:37. DOI: https://doi.org/10.1186/1471-2164-8-37, PMID: 17261194

Weng XH, Huss M, Wieczorek H, Beyenbach KW. 2003. The V-type H+-ATPase in Malpighian tubules of *Aedes aegypti* : localization and activity . *Journal of Experimental Biology* **206**:2211–2219. DOI: https://doi.org/10.1242/jeb.00385, PMID: 12771170

Whiten SR, Ray WK, Helm RF, Adelman ZN, Weiss BL. 2018. Characterization of the adult Aedes aegypti early midgut peritrophic matrix proteome using LC-MS. *PLOS ONE* **13**:e0194734. DOI: https://doi.org/10.1371/journal.pone.0194734, PMID: 29570734

World Health Organisation. 2020a. Vector-borne diseases. https://www.who.int/en/news-room/fact-sheets/detail/vector-borne-diseases [Accessed July 22, 2020].

World Health Organisation. 2020b. Insecticide resistance. http://www.who.int/malaria/areas/vector_control/insecticide_resistance/en/ [Accessed November 20, 2020].

Xiao X, Yang L, Pang X, Zhang R, Zhu Y, Wang P, Gao G, Cheng G. 2017. A Mesh-Duox pathway regulates homeostasis in the insect gut. *Nature Microbiology* **2**:17020. DOI: https://doi.org/10.1038/nmicrobiol.2017.20, PMID: 28248301

Zhao B, Hou Y, Wang J, Kokoza VA, Saha TT, Wang X-L, Lin L, Zou Z, Raikhel AS. 2016. Determination of juvenile hormone titers by means of LC-MS/MS/MS and a juvenile hormone-responsive Gal4/UAS system in Aedes aegypti mosquitoes. *Insect Biochemistry and Molecular Biology* **77**:69–77. DOI: https://doi.org/10.1016/j.ibmb.2016.08.003, PMID: 27530057