Title

A single-cell atlas of the *Culex tarsalis* midgut during West Nile virus infection

Authors

Emily A. Fitzmeyer[1], Taru S. Dutt[1], Silvain Pinaud[2], Barb Graham[1], Emily N. Gallichotte[1], Jessica Hill[3], Corey Campbell[1], Hunter Ogg[1], Virginia Howick, Mara Lawniczak, Erin Osborne Nishimura[3], Sarah Helene Merkling[4], Marcela Henao-Tamayo[1], Gregory D. Ebel[1]

Affiliations

[1]Department of Microbiology, Immunology and Pathology, College of Veterinary Medicine and Biomedical Sciences, Colorado State University, Fort Collins, Colorado, USA

[2] MIVEGEC, Université de Montpellier, IRD, CNRS, Montpellier, France

[3] Department of Biochemistry and Molecular Biology, College of Natural Sciences, Colorado State University, Fort Collins, Colorado, USA

[4] Institut Pasteur, Université Paris Cité, CNRS UMR2000, Insect-Virus Interactions Unit, 75015 Paris, France

Corresponding Author

Gregory D. Ebel, Sc.D.

Professor, Department of Microbiology, Immunology and Pathology

Director, Center for Vector-Borne Infectious Diseases

Colorado State University

Ft. Collins, CO 80526

gregory.ebel@colostate.edu

Abstract

The mosquito midgut functions as a key interface between pathogen and vector. However, studies of midgut physiology and associated virus infection dynamics are scarce, and in *Culex tarsalis* – an extremely efficient vector of West Nile virus (WNV) – nonexistent. We performed single-cell RNA sequencing on *Cx. tarsalis* midguts, defined multiple cell types, and determined whether specific cell types are more permissive to WNV infection. We identified 20 cell states comprised of 8 distinct cell types, consistent with existing

descriptions of *Drosophila* and *Aedes aegypti* midgut physiology. Most midgut cell populations were permissive to WNV infection. However, there were higher levels of WNV RNA (vRNA) in enteroendocrine cells and cells enriched for mitochondrial genes, suggesting enhanced replication in these populations. In contrast, proliferating intestinal stem cell (ISC) populations had the lowest levels of vRNA, a finding consistent with studies suggesting ISC proliferation in the midgut is involved in viral control. Notably, we did not detect significant WNV-infection induced upregulation of canonical mosquito antiviral immune genes (e.g., AGO2, R2D2, etc.) at the whole-midgut level. Rather, we observed a significant positive correlation between immune gene expression levels and vRNA in individual cells, suggesting that within midgut cells, high levels of vRNA may trigger antiviral responses. Our findings establish a *Cx. tarsalis* midgut cell atlas, and provide insight into midgut infection dynamics of WNV by characterizing cell-type specific enhancement/restriction of, and immune response to, infection at the single-cell level.

Introduction

Arthropod-borne viruses represent a severe and ever-growing public health threat (1). Mosquito-borne viruses alone are estimated to cause over 400 million infections globally each year (2). For transmission of a mosquito-borne virus to occur, a mosquito must first become infected with a virus via ingestion of an infectious bloodmeal after feeding on a viremic host (3,4). Said virus must establish infection in the mosquito midgut before it escapes the midgut and disseminates into the body cavity, and eventually enters the salivary glands and saliva – where transmission occurs (3,4). The mosquito midgut is a complex organ comprised of a variety of cell types with distinct functions including digestion, nutrient absorption, endocrine signaling, and innate immune activity (5,6). The midgut is also the site of infection and escape barriers that strongly influence virus population dynamics (4,5). Previous studies have demonstrated that successful infection of the midgut epithelium, and replication and immune evasion therein, is essential for establishing disseminated infection in an insect vector (3,7). In these ways, for hematophagous disease vectors like mosquitoes, the midgut serves as a critical interface between vector and pathogen.

Cx. tarsalis is the primary vector of West Nile virus (WNV) in much of North America (8–10). WNV is the most epidemiologically important arbovirus in N. America, causing ~2,700 deaths from 1999 to 2022 (8,9,11). Despite the importance of *Cx. tarsalis* as a vector of WNV and other important human viruses, studies

examining the cellular composition of its midgut, and WNV infection dynamics therein, are nonexistent. The recent publication of the full *Cx. tarsalis* genome, in conjunction with the growing body of work demonstrating the successful application of single-cell RNA sequencing methodologies in insect models has made it possible to address this significant knowledge gap (12–20). Therefore, we performed single-cell RNA sequencing (scRNA-seq) on dissociated midgut cells from both mock and WNV-infected *Cx. tarsalis* mosquitoes to gain a better understanding of how the midgut functions as the interface between vector and WNV.

We utilized a scRNA-seq approach previously demonstrated to be flavivirus RNA inclusive, which allowed us to detect WNV viral RNA (vRNA) in addition to host transcripts (21). Through this approach we identified distinct midgut populations corresponding with midgut cell types previously described in *Drosophila* and *Aedes aegypti* midguts – enterocyte (nutrient absorption cells), enteroendocrine (secretory cells), cardia (peritrophic matrix secreting cells), intestinal stem cell/enteroblast (undifferentiated progenitor cells), proliferating intestinal stem cell/enteroblast, visceral muscle cells, and hemocyte (immune cells) – and characterized the infection and replication dynamics of WNV within each population (5,6,14,16,17). We found that WNV infects most midgut cell types, with evidence suggesting enhanced replication in enteroendocrine cells and cells enriched for mitochondrial genes, and reduced replication in proliferating intestinal stem cells/enteroblasts. Additionally, we characterized the *Cx. tarsalis* immune response to WNV infection at both the whole-midgut and single-cell level. This study has bolstered our understanding of WNV midgut infection in a highly competent vector, and elucidated the midgut biology of *Cx. tarsalis*.

Results

Single-cell RNA sequencing of female *Cx. tarsalis* midguts identified 20 distinct cell populations*.* Using the 10X Genomics platform we performed scRNA-seq on dissociated mock and WNV-infected *Cx. tarsalis* midgut pools at 4 and 12 days post-infection (dpi). We recovered an average of 2,416 cells per pool with an average coverage of 255,000 reads per cell, which were mapped to the *Cx. tarsalis* genome (**Supplemental File 1**). Following quality control (QC) filtering, we retained data for 12,886 cells at 4dpi (7,386 WNV-infected, 5,500 mock), and 9,301 cells at 12dpi (4,609 WNV-infected, 4,692 mock) for downstream analyses (**Supplemental File 1**). Cells retained after QC contained an average of 597 (611 WNV-infected, 580 mock) and 407 (448 WNV-infected, 367 mock) unique genes per cell at days 4 and 12dpi respectively.

Guided clustering in Seurat (v4.3.0.1) generated 20 (4 dpi) and 17 (12 dpi) distinct clusters of cells (**Figure 1A-B**). The cell type of 15 cell clusters were identified using canonical gene markers and gene enrichment patterns previously identified in *Drosophila* and *Ae. aegypti* midguts (**Figure 1A-C**) (14,16–18,20). All cell-type identifications, with indicated exceptions, were based on conserved cluster markers between mock and WNV infection (**Supplemental File 2, 3**). We identified enterocytes (EC) by significant expression of POU2F1 (nubbin), PLA2G6 (phospholipase A2), and AGBL5 (zinc carboxypeptidase), and enteroendocrine cells (EE) by expression of PROX1 (homeo-prospero domain) (**Figure 2C, Supplemental Figure 3**). High expression of Mlc2 (light chain), Mhc (myosin), and ACTB (actin), allowed us to identify visceral muscle cells (VM); VM-1, VM-2 (**Figure 1C, Supplemental Figure 4**). We identified populations of hemocytes (HC) based on expression of NIMB2 and SPARC (HC-1) and high expression of two pebIII genes (HC-2) (**Figure 1C**) (16,18). EC-like cells (EC-like-1, EC-like-2, EC-like-3) were identified as such based on enrichment for several serine protease and alpha amylase genes (**Figure 1C**) (14,17). One cardia population (cardia-1) was identified by enrichment for sugar transport and chitin-binding genes, as well as several EC-like genes, and a second cardia population (cardia-2) was identified by expression of C-type lysozyme and sugar transporter genes (**Figure 1C**) (14,17). Intestinal stem cells/enteroblasts (ISC/EB) were identified by visualizing klumpfuss (klu) expression localized to these clusters via feature expression map (**Supplemental Figure 5**). Klu is a canonical marker for EBs not ISCs, however, EBs and ISCs are often indistinguishable by UMAP (14,17,20). One of the ISC/EB clusters was significantly enriched for PCNA and aurora kinases A and B – markers for cell proliferation and mitosis – and therefore named ISC/EB-prol to reflect this (**Figure 1C, Supplemental Figure 6**). A cluster that shared identical conserved markers with cardia-1 and was also significantly enriched for PCNA was identified as cardia-prol (**Figure 1C, Supplemental Figure 6**). A cluster of Malpighian tubule cells (MT) that was only present in one sample (mg5c) (**Supplemental Figure 1C**) was identified by significant enrichment for an inward rectifier potassium channel gene (irk-2) as well as several glutathione and vacuolar ATPase genes (**Figure 1C**). This indicates that Malpighian tubule tissue was inadvertently retained upon midgut collection for sample mg5c. Clusters without identifying markers are subsequently referred to by cluster number (e.g., cluster 4). Importantly, HCs and MT cells are not midgut cells, but considered associated with the midgut, while EC, EE, cardia, ISC/EB, and VM cell populations comprise the midgut (**Figure 2A, C**). We compared the proportion of each cluster between mock and WNV-infected replicates and found no significant differences

(**Supplemental Figure 1A-B**). The percent of the total population comprised by each cluster/cell-type can be found in **Supplemental Table 1**.

Characterization of *Cx. tarsalis* midgut secretory and immune cells*.* Enteroendocrine cells (EE) are the secretory cells of the midgut (**Figure 2A**) that, through the secretion of neuropeptides, regulate behavioral responses associated with feeding, satiety, stress, etc. (14,17,22). These cells, and the neuropeptides they secrete, have been previously characterized in *Ae. aegypti* and *Drosophila*, but never *Cx. tarsalis* (23,24). We identified *Cx. tarsalis* orthologs for previously described insect gut hormones found in EE cells - short neuropeptide (sNPF), bursicon (Burs), ion transport peptide (ITP), and tachykinin (Tk) receptor (14,17,20,23–25). However, tachykinin receptor was the only detectable neuropeptide/neuropeptide receptor identified in our EE populations (**Figure 2B**). The EE population was significantly enriched for canonical neuroendocrine genes (IA2, and SCG5) (26,27) and Syt1, and showed expression of Syt4, Syt6, Syx1A and nSyb (genes associated with vesicle docking and secretion) (**Figure 2B)** (14,28). Interestingly, the EE population showed strong enrichment for NEUROD6, a neuronal differentiation gene known to be involved in behavioral reinforcement in mammals (**Figure 2B)** (29).

Hemocytes (HC) are immune cells that circulate in the hemolymph (**Figure 2C**) and play a central role in the mosquito immune response – the exact nature of which varies by class of HC (15,16,18,30,31). Much like EEs, hemocytes have not been characterized in *Cx. tarsalis*. We distinguished the classes of our HC populations using identifying markers. The HC-1 population was identified as mature granulocytes due to expression of SCRASP1 (**Figure 2D**) and enrichment for c-type lectin, defensin, and cecropin genes (**Supplemental File 3**). The HC-2 population was identified as oenocytoids by expression of SCRB3 (**Figure 2D**) (15,16,18,30). SPARC was present in both HC classes, however NIMB2, a gene previously identified as a marker present in all hemocyte classes in *Anopheles gambiae* (18), was only detected in granulocytes (HC-1) (**Figure 2D**). The oenocytoid populations in *Cx. tarsalis* do not appear to express NIMB2 (**Figure 2D**).

COG profiles demonstrate homogeneity between midgut cell populations despite differences in conserved markers*.* We next examined the transcriptional profiles of each cluster to understand the function of unidentified clusters and compare the transcriptomes of distinct cell populations. We used cluster of orthologous genes (COG) categories, and a two-pronged approach to visualization – COG profiles of all genes

expressed in >75% of cells in a cluster (termed “base genes”) and COG profiles of all significant ($p<0.05$), conserved cluster markers with positive log_2 fold-changes (log_2FC) relative to the other clusters (**Supplemental Figure 1A, B**). Base gene and cluster marker gene profiles were derived from the total population at each timepoint. Despite the varying cell types, we noted homogeneity across base genes for each cluster, with the plurality of each profile for most clusters comprised of genes involved in translation and ribosomal biogenesis (J), and energy production and conversion (C) (**Supplemental Figure 1A**). However, ISC/EB-prol, cardia-2, and cardia-prol all possessed fewer ‘J’ COGs than other clusters. The COG profiles of EC-like populations show variability between their transcriptomes and ECs (**Supplemental Figure 1A-B**). As expected, VM populations contained the highest proportions of cytoskeletal genes (Z) in both base gene and cluster marker profiles compared to other cell types (**Supplemental Figure 1A-B**). Interestingly, base gene profiles differed dramatically between the HC-1 and HC-2 populations which reflects the distinct HC classes (granulocytes and oenocytoids) comprising each population (**Supplemental Figure 1A**).

WNV vRNA is detected at varying levels in the majority of midgut cell populations*.* In addition to characterizing the cellular heterogeneity of *Cx. tarsalis* midguts, we also sought to examine WNV infection dynamics at the single-cell level. The five-prime bias of the scRNA-seq chemistry captured and allowed us to detect the WNV 5’ UTR as a feature in our data. Importantly, WNV viral RNA (vRNA) was only detected in our WNV-infected samples, and was broadly detected across most cell populations at both time points (**Figure 3A**). Within WNV-infected replicates we compared the percent of cells with detectable vRNA (calculated as percent expressing) and the average vRNA level (calculated as average expression) in the total population for each timepoint (**Figure 3B**). We saw no significant difference in the total percent of WNV-infected cells between timepoints, but a significant increase in average total vRNA level by 12dpi (**Figure 3B**). Within clusters (replicates within time points combined), cells contained variable levels of vRNA, however some clusters (cluster 17, cardia-prol, etc.) were either not present or were comprised of ≤5 cells in the WNV-infected condition (**Figure 3C**). At both timepoints, cluster 4 contained both the highest average level of vRNA and was significantly enriched for vRNA relative to the other clusters (**Figure 3C-D**). Cluster 4 lacked canonical markers, however was significantly enriched for mitochondrial genes and mitochondrial tRNAs, suggesting these cells are in states of increased energy demand or stress (**Figure 3D**). There was minimal expression of

pro-apoptotic genes across all clusters – confirming that cell death is neither driving clustering nor causing the upregulation of mitochondrial genes and tRNAs in cluster 4 (**Supplemental Figure 7**).

WNV vRNA levels differ between epithelial cell populations*.* Next, we sought to compare the presence and level of vRNA in epithelial cell populations; EC and EC-like, EE, cardia, ISC/EB, and ISC/EB-prol. Average expression and percent expression values derived from clusters comprised of ≤5 cells in a given replicate were excluded from this comparison. At 4dpi the EC-like-2 population had the highest percentage of cells containing vRNA, significantly more than EC-like-1, EC, EE, ISC/EB-prol, and cardia populations (**Figure 5A**). Interestingly, the other EC-like population at 4dpi (EC-like-1) had significantly lower percentages of cells containing vRNA compared to other populations (**Figure 5A**). There were no significant differences in the percent of cells containing vRNA between any epithelial cell population at 12dpi (**Figure 5B**). At both time points, EC-like-2 and EE populations the highest average levels of vRNA (**Figure 4C**). At 12dpi EE populations had significantly higher levels of vRNA than all other epithelial cell populations (**Figure 4D**).

To further explore the epithelial cell populations and their involvement in WNV infection, we used slingshot (v2.10.0) to perform a trajectory inference and identify cell lineages. We identified two lineages: (1) ISC/EB → ISB/EB-prol → EE, and (2) ISC/EB → ISC/EB-prol → EC-like → EC (**Figure 4E**). As expected, we observed decreases in Klu and PCNA expression in both lineages before pseudotime 10, and saw an increase in expression of EC cell marker POU2F1 and EE cell marker PROX1 corresponding with the differentiation of lineages 1 and 2 respectively (**Supplemental Figure 12**). Plotting levels of WNV vRNA by lineage revealed that vRNA levels decrease in the ISC/EB-prol population and increase in fully differentiated EE and EC cells (**Figure 4F**).

Identification of genes associated with WNV infection at the whole-tissue and single-cell level*.* Bulk-RNA sequencing comparing WNV-infected to uninfected *Cx. tarsalis* midguts has not yet been described, so we performed a pseudo-bulk differential expression (DE) analysis to identify differentially expressed genes (DEGs) associated with mock and WNV-infected midguts at the whole-tissue level. We identified six significant DEGs at 4dpi; homocysteine S-methyltransferase, DMAS1 (aldo-keto reductase), and GBE1 (deltamethrin resistance-associated gene) were upregulated in response to WNV infection while BCAN (c-type lectin), uncharacterized gene11056, and a serine protease gene were downregulated (**Figure 5A**). At 12dpi we

identified 10 significant DEGs; an ML (MD-2 related lipid recognition) domain-containing gene, four CRYAB (heat shock protein) genes, and a chitin-binding domain-containing gene were upregulated in response to WNV infection, and three uncharacterized genes (gene13447, gene11056, gene9296) and fibrinogen/fibronectin were downregulated (**Figure 5B**). DEGs differed for each timepoint and, as such, we next examined DEGs in the WNV-infected condition between timepoints. We found many significant DEGs between timepoints; several leucine rich repeat containing genes were upregulated at 4dpi, and host immune gene LYSC4 was upregulated at 12dpi (**Figure 5C**).

To further examine genes associated with WNV infection, we performed a gene correlation on normalized counts for the top 500 variable genes (genes that have variation in expression across all cells) for each timepoint, determined significance via bootstrapping, then extracted and visualized characterized genes correlated (>0.65) with vRNA (**Figure 5D**, **Supplemental Table 2**). Transcription regulator ATRX, a cytochrome p450 gene and several uncharacterized genes were strongly correlated with vRNA at 4dpi (**Figure 5D, Supplemental Table 2**). At 12dpi, GSTE4, HAO1, METTLE20, PROX1, UROS, BCAN, DHDH, and CHKov1 were strongly correlated with vRNA along with serine protease, AMP dependent ligase, cytochrome p450, mitochondrial ribosomal S26, glutathione S-transferase, and aldo/keto reductase family genes (**Figure 5D, Supplemental Table 2**). Many of these genes have no documented roles in flavivirus infection. However, ATRX has been implicated in the cellular response to DNA damage, and chromatin remodeling – processes which many viruses exploit during infection (32–34). Further, cytochrome p450 enzymes and serine proteases have been purported to play a role in the mosquito response to viral infections (35–38).

Upon observing that PROX1, the canonical marker for EE cells, was significantly positively correlated with vRNA at 12dpi, we examined the correlation between vRNA and several previously described neuroendocrine genes (**Figure 2B**, **5E**). Two previously described housekeeping genes, RPL8 and RPL32 (39), were validated as having broad expression throughout the total population and used to both confirm that the high prevalence of vRNA in these populations was not confounding the results and provide a visual reference for a biologically insignificant correlation value (**Figure 5E**). Additionally, for each timepoint we determined the correlation between WNV vRNA and 1,000 random genes from the dataset (unlabeled solid line denotes the average of this calculation with 95% confidence intervals) (**Figure 5E**). At 4dpi PROX1, Syt6, and Syx1A, and at 12dpi PROX1, IA2, and Syx1A had strong positive correlations with vRNA (**Figure 5E**).

Characterization of the midgut immune response to WNV infection at the whole-tissue and single-cell level. While previous work demonstrated an increase in hemocyte proliferation upon bloodmeal ingestion and infection, there were no significant increases in the proportion of hemocyte populations associated with infection at either time point (**Figure 6A-B**) (15). Upon observing that no mosquito immune genes were identified as significantly upregulated in response to WNV infection by pseudo-bulk DEG and correlation analyses, we manually compared percent of cells expressing and expression level of key immune genes that have been implicated in viral control/infection response (19,30,40–43). We identified orthologs in the *Cx. tarsalis* genome for mosquito immune genes; DOME, NANOS1, MYD88, IMD, AGO2, R2D2, STAT5B, Cactus, PIAS1, SUMO2, LYSC4, MARCH8, PIWIL1, PIWIL2, DICER2, and NFKB1 and found no significant differences in the percent of cells expressing and average expression of these genes at either time point (**Figure 6C-D**). Next, we examined the relationships between expression of these immune genes and vRNA at the single-cell level in the WNV-infected population (**Figure 6E**). Interestingly, almost all genes were significantly positively correlated with vRNA at both timepoints (**Figure 6E**). To further confirm these findings, we visualized the relationship of the four most highly correlated immune genes (IMD, PIWIL1, PIAS1, and DOME) with vRNA in individual cells, and compared the expression level of each immune gene in both mock and WNV-infected conditions (timepoints combined) (**Figure 6F-G, Supplemental Figure 11**). These genes and vRNA were correlated, despite comparable expression levels of each gene across infection conditions, confirming that while vRNA load is correlated with specific immune genes at the individual cell level, it does not induce significant population level immune gene enrichment (**Figure 6F-G, Supplemental Figure 11**).

Discussion

In this study, we sought to generate a midgut cell atlas (i.e., map of cell type and function) for *Cx. tarsalis* and characterize WNV infection of the midgut at single-cell resolution by performing scRNA-seq on mock and WNV-infected midguts, collected at days 4 and 12 post infection. We identified and described nutrient absorptive (enterocyte), secretory (enteroendocrine), peritrophic matrix secreting (cardia), undifferentiated progenitor (intestinal stem cell/enteroblast), visceral muscle, and immune (hemocyte) cell populations (5,17,18). The distribution and proportion of each cell-type in the total population varied between timepoints, however we identified at least one cluster comprised of each cell-type at each timepoint. Several clusters were

precluded from identification due to either lack of canonical markers/enrichment patterns, or origination from a single replicate. Nonetheless, we have demonstrated that single-cell sequencing of *Cx. tarsalis* midguts is feasible and that distinct cell populations can be identified and characterized using previously described canonical cell-type markers and enrichment patterns (14,16–18,20).

We detected WNV RNA (vRNA) in the majority of midgut cells at both timepoints. While vRNA significantly increased in the total midgut by 12dpi, the percent of infected cells did not, demonstrating that the majority of midgut cells that will become infected are infected by 4dpi, while vRNA load increases as infection progresses. The high percentage of WNV-infected cells and permissibility of most cell populations to infection supports previous work demonstrating the extreme competence of *Cx. tarsalis* as a WNV vector (9,44–46).

Interestingly, while WNV infected almost all midgut cell populations, cluster 4 was significantly enriched for vRNA at both timepoints. This high WNV-expressing cluster was associated with very few (<15) defining cluster markers, precluding us from identifying its cell-type. The cluster markers associated with this cell state are comprised entirely of mitochondrial genes and mitochondrial tRNAs, suggesting a heightened state of cell stress and/or energy production. Importantly, this cluster was present and enriched for the same mitochondrial genes in the mock condition, suggesting that WNV was able to replicate to higher levels in cells enriched for mitochondrial genes, and not that viral replication induced significant stress and/or energy production responses in these clusters (47,48). Several previous studies have demonstrated that positive sense-single stranded RNA viruses like dengue virus (DENV), and SARS-CoV-2 modulate mitochondrial dynamics to facilitate replication and/or immune evasion (48–50), suggesting potential beneficial interactions between WNV and the mitochondria. Additionally, a study in *Lepidoptera* (moths and butterflies) purported that enrichment of mitochondrial genes is associated with insect stress resistance (47). Stress resistance responses modulate cell viability and it is known that the maintenance of cell viability is central to productive WNV infection (51,52). Moreover, we demonstrated that several heat shock genes (known to be protective against cell stress) were significantly upregulated in WNV-infected midguts, further suggesting interplay between WNV infection and the stress response (53–55). However, these heat shock genes were predominantly localized to cluster 8, an un-typed cluster that does not contain the high level of vRNA seen in cluster 4. Further work is needed to tease out the complexities of these cell states and their impact on WNV replication.

Our enteroendocrine (EE) cell populations only contained one of the previously described mosquito neuropeptides/neuropeptide receptors, tachykinin receptor, and only in a small subset of cells (14,17,20,23,25). This could be due to the known bias of scRNA-seq towards highly expressed genes, or due to *Cx. tarsalis* EE cell secretion of yet uncharacterized neuropeptides. High expression of NEUROD6 – a neurogenic differentiation factor frequently found in neurons involved in behavioral reinforcement – in EE populations at both timepoints supports the hypothesis that additional/uncharacterized neuropeptides may be present in *Cx. tarsalis* EE cells (29). Interestingly, EE populations contained more vRNA than other epithelial cell populations. Further, PROX1 (the canonical marker for EE cells) and select neuroendocrine and vesicle docking genes present in EE cells were strongly positively correlated with vRNA at both timepoints, supporting our hypothesis that EE cells serve as sites of enhanced WNV replication during midgut infection. This hypothesis is further supported by previous studies that suggest arboviruses preferentially replicate in highly polarized cell types, such as EE cells (56,57). Additionally, previous work with Sindbis virus (SINV) - a mosquito-borne alphavirus – identified EE cells as a site of infection initiation in *Ae. aegypti* (58).

A previous study characterizing ISC dynamics in response to DENV in the *Ae. aegypti* midgut found that ISC proliferation increased refractoriness to infection, suggesting that cell renewal is an important part of the midgut immune response (59). Interestingly, we observed that proliferating ISC/EB populations consistently had the lowest levels of vRNA compared to other epithelial cell types. Further, cell lineage trajectory analysis showed that vRNA decreased to the lowest levels in the ISC/EB-prol population before rebounding in fully differentiated EE and EC cell types. Proliferating cell states expressed notably fewer base genes involved in translation, ribosomal structure, and biogenesis. These findings together suggest that the transcriptional state of proliferating ISC/EBs impedes WNV replication.

While the presence of vRNA alone does not signify active replication, average vRNA levels increased between timepoints in all epithelial cell populations, apart from the ISC/EB and ISC/EB-prol populations (**Supplemental Figure 13**), suggesting that the high vRNA levels in EE populations, and low vRNA levels in ISC/EB-prol populations, are the result of enhanced and restricted replication respectively. The lack of significant differences between the percent of cells in each population containing vRNA at 12dpi further supports that varying levels of vRNA in EE and ISC/EB-prol cells is due to permissivity to replication and not susceptibility to

infection. The factors that determine cell-type specific enhancement or suppression of WNV replication are not currently known, but could include more efficient evasion of antiviral pathways, more efficient mechanisms of midgut escape/dissemination, or abundance of pro/anti-viral genes.

We identified granulocyte and oenocytoid hemocyte populations – known to play important roles in the mosquito innate immune response – at both timepoints, allowing us to characterize a cellular immune component of mosquito midguts (5,15). There was no evidence of immune cell proliferation or immune gene upregulation in the total infected population compared to mock – suggesting little to no immune activation in the midgut upon WNV infection. This was surprising given the importance of the midgut as a site of innate immune activation (5,6,45,46). Although scRNA-seq of *Cx. tarsalis* after WNV infection has not been described, several previous studies in *Ae. aegypti* and *Cx. pipiens* have noted significant upregulation of IMD and Toll pathway genes in response to viral infection and highlighted that the innate immune response in mosquitoes is a strong determinant of vector competence (53,60,61). The absence of a notable immune response to WNV infection in *Cx. tarsalis* could be a determinant of the vector's extreme susceptibility and competence (9,44–46). However, in individual cells, most immune genes had some degree of significant positive correlation with vRNA suggesting that, while WNV infection does not cause significant enrichment of these genes in the total population, WNV infection and replication influences the expression of these genes at the single-cell level. This finding highlights that scRNA-seq is a powerful tool for characterizing infection dynamics that are not apparent when looking at the population average.

Limitations and Future Directions

Our inability to detect certain genes (i.e., neuropeptide genes and additional canonical markers we would expect to see) could be due to the low percentage (~30%) of reads mapped to the *Cx. tarsalis* genome (due to a predominance of reads that were too short to map), or the absence of those genes in the existing annotation file. Future scRNA-seq studies in *Culex* mosquitoes could potentially benefit from adjusting the fragmentation time recommended by 10X Genomics. Further, improvement of the existing *Cx. tarsalis* genome annotation would facilitate studies of gene expression in this species. A multitude of genes detected in our dataset remain uncharacterized due to a lack of appropriate orthologs, which could be explained by the evolutionary divergence between *Cx. tarsalis* and the species from which most gene orthologs were derived; *Cx.*

quinquefasciatus and *Ae. aegypti* which diverged 15-22 million years ago (MYA) and 148-216 MYA respectively (12). High levels of vRNA in specific cell types imply that replication is occurring/has occurred but it is important to note that the presence of vRNA is not analogous to active viral replication (e.g., the presence of vRNA could be the result of phagocytosis of an infected cell). Future studies could use qRT-PCR, focusing on our top genes of interest, to measure expression kinetics and levels following midgut infection in *Cx. tarsalis* and other relevant vectors. Finally, Early WNV infection in the *Cx. tarsalis* midgut will be further studied in our lab via immunofluorescence assays using cell-type specific RNA probes in whole midguts, putting the findings described here into a spatial context.

Conclusion

The work presented here demonstrates that WNV is capable of infecting most midgut cell types in *Cx. tarsalis*. Moreover, while most cells within the midgut are susceptible to WNV infection, we observed modest differences in virus replication efficiency that appeared to occur in a cell-type specific manner, with EE cells being the most permissive and proliferating ISC/EB cells being the most refractory. Our findings also strongly suggest interplay between WNV infection and the cell stress response, and we have provided evidence that WNV infection of the *Cx. tarsalis* midgut results in the upregulation of cell stress associated genes. We observed mild to no upregulation of key mosquito immune genes in the midgut as a whole, however, we show that immune gene expression is correlated with WNV vRNA level within individual cells. Additionally, we have generated a midgut cell atlas for *Cx. tarsalis* and, in doing so, improved the field's understanding of how WNV establishes infection in this highly competent vector.

Acknowledgements

The research reported in this publication was supported by National Institutes of Health grant R01-AI067380 (G.D.E), National Institute of Allergy and Infectious Diseases of the National Institutes of Health under Award Number T32-AI162691, and by Colorado State University's Office of the Vice President for Research's "Accelerating Innovations in Pandemic Disease" initiative, made possible through support from The Anschutz Foundation. Emily N Gallichotte was supported by funding to Verena (viralemergence.org) from the U.S. National Science Foundation, including NSF BII 2021909 and NSF BII 2213854. The content is solely the responsibility of the authors. Next Generation Sequencing was performed at the Genomics Shared

Resource (RRID: SCR_021984) at the University of Colorado. This resource is supported by Cancer Center Support Grant (P30CA046934).

Author Contributions

Emily A. Fitzmeyer – Conceptualization, validation, formal analysis, investigation, resources, data curation, writing – original draft, writing – review and editing, visualization, project administration, funding acquisition.

Taru S. Dutt – Validation, resources, investigation, writing – review and editing.

Silvain Pinaud – Methodology.

Barb Graham – Formal analysis.

Emily N. Gallichotte – Visualization, writing – review and editing.

Jessica Hill – Validation, resources.

Corey Campbell – Resources.

Hunter Ogg – Resources.

Virginia Howick – Methodology.

Mara Lawniczak – Methodology.

Erin Osborne Nishimura – Validation, resources.

Sarah Helene Merkling – Validation, resources.

Marcela Henao-Tamayo – Resources, supervision.

Gregory D. Ebel – Conceptualization, validation, resources, writing – review and editing, supervision, funding acquisition.

Declaration of interests

The authors declare no competing interests.

Inclusion and diversity

We support inclusive, diverse, and equitable conduct of research.

Legends

Figure 1. Single-cell sequencing of *Cx. tarsalis* midguts and cell-typing of midgut cell populations. Uniform Manifold Approximation and Projection (UMAP) reduction was used to visualize midgut cell

populations. (**A**) 20 distinct clusters were identified at 4dpi and (**B**) 17 distinct clusters were identified at 12dpi. (**C**) Expression of canonical marker genes that were conserved between infection condition (mock and WNV-infected) were used to determine cell type. Dot size denotes percent of cells expressing each gene, color denotes scaled gene expression. Panel C was derived from the total population.

Figure 2. Characterization of enteroendocrine cells and hemocytes. Gene counts were converted to binary (Blue = 1, Grey = 0) and visualized via UMAP to demonstrate proportion of EE cells (**A**) expressing neuropeptides, and secretory genes (**B**) and proportion of HC cells (**C**) expressing HC markers, as well as granulocyte and oenocytoid markers (**D**). Feature plots show binary expression for the total population.

Figure 3. WNV vRNA is detected in most midgut cell populations. (**A**) Levels of WNV vRNA across midgut cell populations at 4 and 12dpi. Purple to grey gradient denotes high to low levels of vRNA. (**B**) percent of cells in the total population containing WNV vRNA (percent expressing) and average WNV vRNA level in the total population (average expression) calculated for each WNV-infected replicate and compared between timepoints. Significance was determined by unpaired t-test, * = $p < 0.05$. (**C**) WNV vRNA level in individual cells of distinct clusters. (**D**) Enrichment for identical sets of mitochondrial genes and tRNAs specific to cluster 4. Dot size denotes percent of cells expressing each gene, color denotes scaled gene expression. “MT-“ prefix was added to tRNA gene names in this figure for clarity. Mitochondrial tRNAs were not included in mitochondrial gene estimation for QC filtering. Panels B, C, and D were derived from only WNV-infected samples.

Figure 4. Examining WNV vRNA levels in intestine epithelial cells. The percentage of cells containing WNV vRNA in each epithelial cell population at (**A**) 4dpi and (**B**) 12dpi. Average WNV vRNA level in epithelial cell populations at (**C**) 4dpi and (**D**) 12dpi. Panels **A-D** were derived from only WNV-infected replicates, and exclude replicate values derived from clusters of ≤5 cells. Significance determined by one-way ANOVA, * = $p < 0.05$, ** = $p < 0.005$, *** = $p < 0.0005$, **** = $p < 0.0001$. (**E**) Trajectory inference for ISC/EB and ISC/EB-prol populations identified two lineages. (**F**) vRNA levels in each lineage from pseudotime 0-30. Panels E-F were derived from the total population.

Figure 5. Identifying genes upregulated in response to WNV infection with differential expression and correlation analyses. DEGs associated with WNV-infected condition when compared to the mock condition at

4dpi (**A**) and 12dpi (**B**) were identified with DESeq2. Negative log_2FC indicates downregulation and positive log_2FC indicates upregulation associated with WNV infection. (**C**) Genes upregulated during WNV infection at 4dpi (negative log_2FC) and 12dpi (positive log_2FC). (**D**) Characterized genes that are significantly correlated with WNV vRNA at 4dpi (purple) and 12dpi (teal), with correlation values of >0.65. Only significant ($p < 0.05$) relationships shown. (**E**) Correlation of enteroendocrine specific genes and WNV vRNA at 4dpi (purple) and 12dpi (teal). Red triangular points denote nonsignificant correlation. Unlabeled solid line represents the mean correlation of WNV vRNA with 1000 randomly selected genes at the specified timepoint (unlabeled dotted lines represent the upper and lower 95% confidence intervals for this value). Labeled dotted lines denote vRNA correlation with RPL8 and RPL32 housekeeping reference genes. Panels C-E were derived from only WNV-infected replicates.

Figure 6. Lack of key immune gene upregulation in response to WNV infection.

(**A-B**) Percent of the total population of each replicate comprised by hemocytes compared between mock and WNV-infected conditions at each timepoint. Significance was determined by unpaired t-test. (**C-D**) Average expression of, and percent of cells expressing, mosquito innate immune genes compared between mock and WNV-infected replicates for each timepoint. Significance was determined by multiple unpaired t-tests. (**E**) Correlation of immune genes and WNV vRNA at 4dpi (purple) and 12dpi (teal). Red triangular points denote nonsignificant correlation. Unlabeled solid line represents the mean correlation of WNV vRNA with 1000 randomly selected genes at the specified timepoint (unlabeled dotted lines represent the upper and lower 95% confidence intervals for this value). Labeled dotted lines denote vRNA correlation with RPL8 and RPL32 housekeeping reference genes. (**F**) Feature scatter of IMD vs. WNV 5' UTR for both timepoints combined. (**G**) Expression level of IMD in mock and WNV-infected conditions for both timepoints combined. Panels E-F were derived from only WNV-infected replicates.

Methods

Virus*.* All infections were performed with a recombinant barcoded WNV (bcWNV) passage 2 stock (epidemic lineage I strain, 3356) grown on Vero cells. Titer for the stock was determined by standard Vero cell plaque assay (62).

Mosquito infection*.* Mosquito studies were conducted using laboratory colony-derived *Cx. tarsalis* mosquitoes (>50 passages) WNV infections in mosquitoes were performed under A-BSL3 conditions. Larvae were raised on a diet of powdered fish food. Mosquitoes were maintained at 26°C with a 16:8 light:dark cycle and maintained at 70–80% relative humidity, with water and sucrose provided ad libitum. *Cx. tarsalis* mosquitoes were transferred to A-BSL3 conditions 48 hours prior to blood feeding, and dry starved 20-24 hours before blood feeding. Seven days after pupation (6-7 days after emergence) mosquitoes were exposed to an infectious bloodmeal containing a 1:1 dilution of defibrinated calf's blood and bcWNV stock diluted in infection media (Dulbecco's Modified Eagle's Medium, 5% penicillin-streptomycin, 2% amphotericin B, and 1% fetal bovine serum (FBS)) for a final concentration of 3-6e^{7} PFU/mL, or a mock bloodmeal containing a 1:1 dilution of defibrinated calf's blood and infection media. All bloodmeals were provided in a hog's gut glass membrane feeder, warmed by circulating 37°C water. Following 50-60 minutes of feeding, mosquitoes were cold-anesthetized, and engorged females were separated into cartons and maintained on sucrose.

Collection of mosquito tissues*.* At indicated time points, mosquitoes were cold-anesthetized and transferred to a dish containing Sf900III insect cell culture media (Gibco) with 5% FBS. Midguts were dissected, and transferred to tubes containing 500uL Sf900III media + 5% FBS and kept on ice for the duration of dissections. Ten pooled midguts per sample/tube were collected for dissociation and sequencing.

Midgut dissociation and single-cell suspension preparation*.* A dissociation buffer containing *Bacillus licheniformis* protease (10mg/mL) and DNAse I (25U/mL) was prepared in Sf900III media (Gibco). Pooled midguts were resuspended in dissociation media, transferred to a 96-well culture dish, and triturated with a p1000 pipet at 15-20 minute intervals for 105 minutes. At each interval 100-125µL containing dissociated single-cells was collected (with replacement) from the top of the dissociation reaction and transferred to 25mL of Sf900III + 5% FBS on ice. Dissociation reactions were kept covered at 4°C between trituration. Upon complete tissue dissociation the entire remaining volume of each reaction was transferred to Sf900III + 5% FBS on ice. Collection tubes with dissociated cells were centrifuged at 700xg for 10 minutes at 4°C, resuspended in 500uL of Sf900III + 5% FBS, and passed through a 40um small volume filter (PluriSelect). Immediately prior to loading on the Chromium Controller, cell suspensions were spun down at 700xg for 10 minutes and washed twice in 1mL PBS + 0.04% bovine serum albumin (BSA), and resuspended in 50µL of

PBS + 0.04% BSA. Cell concentration was determined using the Countess II Automated Cell Counter (Thermo Fisher Scientific) and the appropriate cell suspension volume (target recovery of 10,000 cells) was loaded on the Chromium controller.

Gel Bead-In Emulsions (GEM) generation and cDNA synthesis*.* GEM generation and cDNA synthesis were performed using the Next GEM Single Cell 5' GEM kit v2 (PN-1000266) and Next GEM Chip K Single Cell Kit (PN-1000286) (10X Genomics). Reactions for GEM generation were prepared according to the Chromium Next GEM Single Cell 5' Reagent Kit v2 (dual index) user guide with one alteration; a primer specific to the WNV envelope region of the genome was added at a concentration of 10nM (7.5µl – displacing 7.5µl of the total H_2O added to each reaction). GEMs were generated using the 10X Chromium controller X series and cDNA was synthesized according to the Chromium Next GEM Single Cell 5' Reagent Kit v2 (dual index) user guide.

Library preparation and sequencing*.* Libraries were prepared using the Next GEM single cell 5' v2 library construction kit and Dual Index Kit TT Set A (10X Genomics, PN-1000190 and PN-1000215 respectively). Library construction was carried out according to the Chromium Next GEM Single Cell 5' Reagent Kit v2 (dual index) user guide. Library concentration was determined by KAPA Library Quantification Kit (Roche). Libraries were then diluted to 15nM, pooled by volume, and sequenced at the CU Anschutz Genomics and Microarray core on the NovaSeq 6000 (150x10x10x150) (Illumina) at a target coverage of $4.0e^8$ read pairs per sample, equating to 40,000 read pairs per cell. Average sample coverage and cell recovery per sample was $5.3e^8$ read pairs and 2417.9 cells respectively (**Supplemental File 1**).

Reference generation and sample processing with Cell Ranger*.* The gene feature files associated with the *Cx. tarsalis* and WNV genomes were converted to gene transfer format (gtf) using AGAT (v1.0.0) prior to being filtered with the CellRanger (v7.0.1) mkgtf function (63). An 'MT-' prefix was manually added to all non-tRNA features located in the mitochondrial chromosome of the *Cx. tarsalis* genome. The contents of the filtered WNV genome feature file, and fasta file, were appended to the *Cx. tarsalis* feature and fasta files respectively, and run through CellRanger::mkref. All sequencing data were processed and mapped to the aforementioned *Cx. tarsalis* reference genome using CellRanger::count with the following parameters: --include-introns=true \ --expect-cells=10000.

Quality control and Seurat workflow*.* Cell Ranger output files were individually read into RStudio (RStudio - v2023.09.0+463, R – v4.3.2) as SingleCellExperiment objects using the singleCellTK package (v2.12.0). Doublet identification and ambient RNA estimation were performed with singleCellTK::runCellQC using the algorithms “scDblFinder” and “DecontX” respectively. Samples were filtered for doublets and ambient RNA contamination by keeping cells with the following metrics: decontX_contamination < 0.6, scDblFinder_doublet_score < 0.9. Samples were then converted to Seurat objects, log normalized, and merged into one Seurat object, with columns pertaining to sample of origin, infection condition, and timepoint added to the object metadata prior to processing as described in the Seurat (v4.3.0.1) guided clustering tutorial (64). Briefly, mitochondrial gene percentage for each cell was calculated, and cells with the following metrics were retained: nFeature_RNA > 100, nFeature_RNA < 2500, percent_mt < 25 (14). Cell retention metrics were informed by a previous study of the Drosophila midgut (14). Percent of cells retained after QC for each sample can be found in **Supplemental File 1**. Features were log normalized, variable features were identified, the top 2000 variable features were scaled, a principle component (PC) analysis dimensionality reduction was run, and the number of PCs needed to adequately capture variation in the data was determined via elbow plot. Nearest neighbors were computed, and appropriate clustering granularity was determined with Clustree (v0.5.0). Uniform Manifold Approximation and Projection (UMAP) dimensional reduction was performed, and clusters were visualized with UMAP reduction. Cluster markers were identified with Seurat::FindConservedMarkers using default parameters and infection condition as the grouping variable. For clusters that had very few conserved markers we split the dataset by infection condition and used Seurat::FindMarkers on clusters in either the mock or WNV-infected condition. WNV vRNA as a cluster marker was identified by splitting the dataset by infection condition and using Seurat::FindAllMarkers on the WNV-infected samples. In all cases where calculations were performed on individual replicates or individual conditions, the merged Seurat object was split by sample or condition using Seurat::SplitObject so that calculations performed on subsets of the data or individual replicates were derived from a dataset that had been normalized as one. Percent expression and average expression were calculated with scCustomize::Percent_Expressing (v1.1.3) and Seurat::AverageExpression respectively. Feature expression levels were visualized with Seurat::FeaturePlot and Seurat::VlnPlot.

Trajectory inference with Slingshot. Lineage structure and pseudotime inference was performed using the Slingshot (v2.10.0) and tradeSeq (v1.16.0) functions getLineages, getCurves, and fitGAM successively on the dataset containing the total population. ISC/EB and ISC/EB-prol populations were specified as the start state and EC and EE populations were specified as the end state for lineage determination.

Pseudo-bulk differential expression analysis with DESeq2*.* Pseudo-bulk dataset was generated and differential expression analysis performed as described by Khushbu Patel's (aka bioinformagician) pseudo-bulk analysis for single-cell RNA-Seq data workflow tutorial (65). Briefly, raw counts were aggregated at the sample level using Seurat::AggregateExpression, and the aggregated counts matrix extracted and used to create a DESeq2 object (dds). The dds object was filtered to retain genes with counts >=10 prior to running DESeq2 and extracting results for the appropriate contrast. DESeq2 results were visualized via EnhancedVolcano (v1.20.0). We identified infection associated DEGs between timepoints by performing a pseudo-bulk DE analysis between 4dpi and 12dpi WNV-infected samples and mock samples separately. We then filtered out DEGs associated only with bloodmeal consumption (genes that came up as significantly differentially expressed between timepoints in our mock condition) leaving only DEGs associated with infection.

Gene correlation with vRNA level*.* Raw gene counts for each timepoint were extracted and normalized using scLink::sclink_norm (v1.0.1) (66). Correlation matrices were then generated for the top 500 variable genes (identified during the Seurat guided clustering workflow) using scLink::sclink_corr (66). For specific neuroendocrine and immune gene correlations a vector of specific gene names was supplied in lieu of the top 500 variable genes. P-values associated with the correlation values were determined via bootstrapping.

Ortholog identification with EggNOG Mapper*.* Coding sequence (CDS) genome coordinates were extracted from the *Cx. tarsalis* gene transfer format file and corresponding genome sequences were extracted from the available fasta file using Bedtools (v2.26.0). CDS were then assigned orthologs via eggNOG-mapper v2 (web version) using default parameters (**Suplemental File 4**) (67,68). gtf file gene IDs were merged with the eggNOG output file using custom R scripts. Gene ID, ortholog seed, preferred gene name, COG category notation, PFAM information, and gene description were retained in a gene name and ID file within which information pertaining to identical gene IDs was aggregated using a custom R script. This aggregated gene

name and ID file (**Supplemental File 5**) was used to assign gene names to the marker files used for cell-typing, DESeq2 results files, and scLink correlation matrices.

Statistical analyses*.* Statistical analyses were performed in GraphPad Prism version 10.0.3. Differences in vRNA level in known clusters were measured by one-way ANOVA with Tukey's multiple comparisons test. Differences in average expression and percent expression of mosquito immune genes between mock and WNV-infected conditions were measured by multiple unpaired t-tests. Differences in hemocyte population proportion were measured by unpaired t-test. All scripts used for data processing, analysis, and visualization are available on GitHub: https://github.com/fitz-meyer/scRNA_seq_fitz

Supplemental Material

Supplemental Figure 1. COG profiles of midgut cell populations. Cluster of orthologous gene (COG) profiles for (**A**) genes expressed in ≥75% of cells in each cluster and (**B**) cluster marker genes were visualized as percentage of total for each cluster/population. Colors represent COG notation A-Z, NA and DIV as shown in the notation key embedded in the figure. Where applicable, marker gene COG profiles were derived from cluster markers that are conserved between infection conditions.

Supplemental Table 1. Cluster proportion. Percent of the total midgut cell population each distinct cluster comprises.

Supplemental Figure 2. Cluster proportion and grouping by condition and replicate. Proportion of the total population comprised by each cluster compared between mock and WNV-infected conditions at 4dpi (**A**) and 12dpi (**B**). Only significant comparisons shown. Significance determined by multiple unpaired t-tests. Bar = mean, error bars = SD. (**C**) Cluster grouping and composition by sample – mock and infected samples both plotted. Different colors denote different samples. (**D**) Cluster grouping and composition by infection condition. Salmon = mock, blue = WNV-infected.

Supplemental Figure 3. Confirming enteroendocrine cell-type by visualizing PROX1 expression. PROX1 expression visualized in the total population via uniform manifold approximation and projection (UMAP) feature map and violin plot. Color in feature map denotes expression level.

Supplemental Figure 4. Confirming visceral muscle cell-type by visualizing cytoskeletal gene expression. Actin (ACTB), myosin (Mhc), and light chain (Mlc2) expression visualized in the total population via uniform manifold approximation and projection (UMAP) feature map and violin plot. Color in feature map denotes expression level.

Supplemental Figure 5. Identifying intestinal stem cell/enteroblast (ISC/EB) cell-type by visualizing canonical marker gene expression. Klumpfuss (klu) expression visualized in the total population via UMAP feature map. Color in feature map denotes expression level.

Supplemental Figure 6. Proliferation and mitotic markers in proliferating intestinal stem cell/enteroblast (ISC/EB-prol) cell-type. Identifying proliferation in ISC/EB-prol populations by visualizing PCNA, and aurora kinases A and B in the total population via uniform manifold approximation and projection (UMAP) feature map. Color in feature maps denotes expression level.

Supplemental Figure 7. Confirming that cell death does not drive clustering. Expression of apoptotic and anti-apoptotic genes visualized in the total population via UMAP feature plot. Color in feature map denotes expression level.

Supplemental Figure 8. S phase markers. Expression of S phase gene markers visualized in the total population via UMAP feature plot. Color in feature map denotes expression level.

Supplemental Figure 9. G2/M phase markers. Expression of G2/M phase gene markers visualized in the total population via UMAP feature plot. Color in feature map denotes expression level.

Supplemental Figure 10. Percent of cells in each cluster containing WNV vRNA. Calculated from total population, not divided by replicates.

Supplemental Figure 11. Visually confirming vRNA level is correlated with select immune gene expression without significantly increasing expression in the total population. Correlation between vRNA and 3 of the most highly correlated immune genes (as determined by scLink) confirmed by feature scatter. Equivalent expression levels of immune genes between mock and WNV-infected conditions confirmed by violin plot. Scatter plots derived from only WNV-infected replicates.

Supplemental Table 2. Complete list of genes identified as significantly positively correlated with WNV vRNA. Gene name, ID, correlation value, p-value, and description for all genes significantly ($p < 0.05$) positively correlated (> 0.65) with vRNA. Significance was determined by bootstrapping.

References

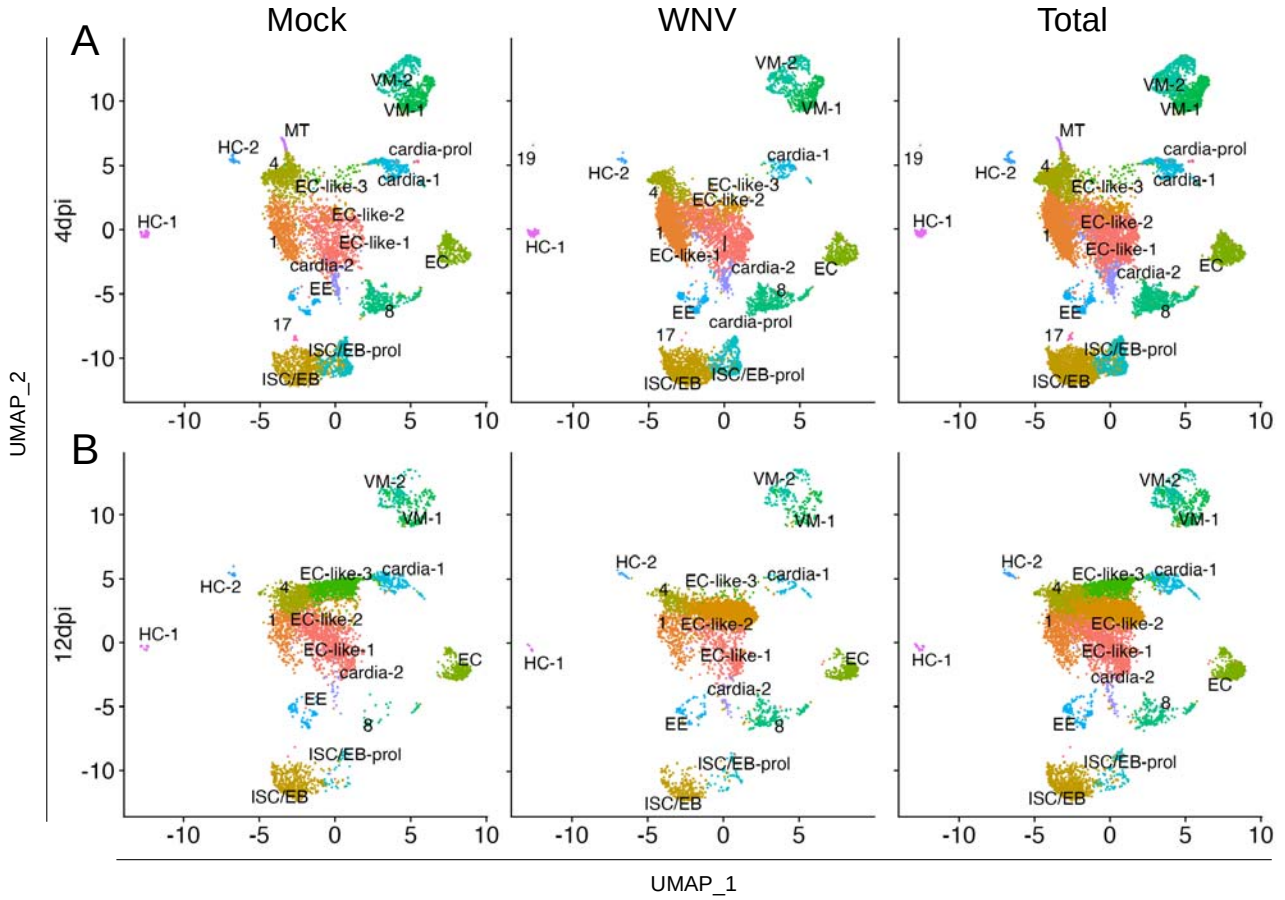
1. Girard M, Nelson CB, Picot V, Gubler DJ. Arboviruses: A global public health threat. Vaccine. 2020 May;38(24):3989–94.
2. Pierson TC, Diamond MS. The continued threat of emerging flaviviruses. Nat Microbiol. 2020 May 4;5(6):796–812.
3. Carpenter A, Bryant WB, Santos SR, Clem RJ. Infection of Aedes aegypti Mosquitoes with Midgut-Attenuated Sindbis Virus Reduces, but Does Not Eliminate, Disseminated Infection. J Virol. 2021 Jun 10;95(13).
4. Weaver SC, Forrester NL, Liu J, Vasilakis N. Population bottlenecks and founder effects: implications for mosquito-borne arboviral emergence. Vol. 19, Nature Reviews Microbiology. Nature Research; 2021. p. 184–95.
5. Hixson B, Taracena ML, Buchon N. Midgut Epithelial Dynamics Are Central to Mosquitoes' Physiology and Fitness, and to the Transmission of Vector-Borne Disease. Front Cell Infect Microbiol. 2021 Mar 25;11.
6. Caccia S, Casartelli M, Tettamanti G. The amazing complexity of insect midgut cells: types, peculiarities, and functions. Cell Tissue Res. 2019 Sep 29;377(3):505–25.
7. Lan H, Chen H, Liu Y, Jiang C, Mao Q, Jia D, et al. Small Interfering RNA Pathway Modulates Initial Viral Infection in Midgut Epithelium of Insect after Ingestion of Virus. J Virol. 2016 Jan 15;90(2):917–29.
8. https://www.cdc.gov/westnile/statsmaps/data-and-maps.html [Internet]. 2023. Center for Disease Control and Prevention, West Nile virus.
9. Dunphy BM, Kovach KB, Gehrke EJ, Field EN, Rowley WA, Bartholomay LC, et al. Long-term surveillance defines spatial and temporal patterns implicating Culex tarsalis as the primary vector of West Nile virus. Sci Rep. 2019 Apr 29;9(1):6637.
10. Turell MJ, O'Guinn ML, Dohm DJ, Webb JP, Sardelis MR. Vector Competence of *Culex tarsalis* from Orange County, California, for West Nile Virus. Vector-Borne and Zoonotic Diseases. 2002 Sep;2(3):193–6.
11. Artsob H, Gubler DJ, Enria DA, Morales MA, Pupo M, Bunning ML, et al. West Nile Virus in the New World: Trends in the Spread and Proliferation of West Nile Virus in the Western Hemisphere. Zoonoses Public Health. 2009 Aug 9;56(6–7):357–69.
12. Main BJ, Marcantonio M, Johnston JS, Rasgon JL, Brown CT, Barker CM. Whole-genome assembly of *Culex tarsalis*. G3 Genes|Genomes|Genetics. 2021 Apr 12;11(2).
13. Merkling SH, Lambrechts L. Taking Insect Immunity to the Single-Cell Level. Trends Immunol. 2020 Mar;41(3):190–9.
14. Hung RJ, Hu Y, Kirchner R, Liu Y, Xu C, Comjean A, et al. A cell atlas of the adult *Drosophila* midgut. Proceedings of the National Academy of Sciences. 2020 Jan 21;117(3):1514–23.
15. Severo MS, Landry JJM, Lindquist RL, Goosmann C, Brinkmann V, Collier P, et al. Unbiased classification of mosquito blood cells by single-cell genomics and high-content imaging. Proceedings of the National Academy of Sciences. 2018 Aug 7;115(32).
16. Raddi G, Barletta ABF, Efremova M, Ramirez JL, Cantera R, Teichmann SA, et al. Mosquito cellular immunity at single-cell resolution. Science (1979). 2020 Aug 28;369(6507):1128–32.
17. Cui Y, Franz AWE. Heterogeneity of midgut cells and their differential responses to blood meal ingestion by the mosquito, Aedes aegypti. Insect Biochem Mol Biol. 2020 Dec;127:103496.

18. Kwon H, Mohammed M, Franzén O, Ankarklev J, Smith RC. Single-cell analysis of mosquito hemocytes identifies signatures of immune cell subtypes and cell differentiation. Elife. 2021 Jul 28;10.
19. Lee WS, Webster JA, Madzokere ET, Stephenson EB, Herrero LJ. Mosquito antiviral defense mechanisms: a delicate balance between innate immunity and persistent viral infection. Parasit Vectors. 2019 Dec 11;12(1):165.
20. Hung RJ, Li JSS, Liu Y, Perrimon N. Defining cell types and lineage in the Drosophila midgut using single cell transcriptomics. Curr Opin Insect Sci. 2021 Oct;47:12–7.
21. Sanborn MA, Li T, Victor K, Siegfried H, Fung C, Rothman AL, et al. Analysis of cell-associated DENV RNA by oligo(dT) primed 5' capture scRNAseq. Sci Rep. 2020 Jun 3;10(1):9047.
22. Schoofs L, De Loof A, Van Hiel MB. Neuropeptides as Regulators of Behavior in Insects. Annu Rev Entomol. 2017 Jan 31;62(1):35–52.
23. Predel R, Neupert S, Garczynski SF, Crim JW, Brown MR, Russell WK, et al. Neuropeptidomics of the Mosquito *Aedes aegypti*. J Proteome Res. 2010 Apr 5;9(4):2006–15.
24. Chen J, Kim S min, Kwon JY. A Systematic Analysis of Drosophila Regulatory Peptide Expression in Enteroendocrine Cells. Mol Cells. 2016 Apr;39(4):358–66.
25. Christ P, Hill SR, Schachtner J, Hauser F, Ignell R. Functional characterization of mosquito short neuropeptide F receptors. Peptides (NY). 2018 May;103:31–9.
26. TORII S. Expression and Function of IA-2 Family Proteins, Unique Neuroendocrine-specific Protein-tyrosine Phosphatases. Endocr J. 2009;56(5):639–48.
27. MBIKAY M, SEIDAH NG, CHRÉTIEN M. Neuroendocrine secretory protein 7B2: structure, expression and functions. Biochemical Journal. 2001 Jul 15;357(2):329.
28. Shao X, Li C, Fernandez I, Zhang X, Südhof TC, Rizo J. Synaptotagmin–Syntaxin Interaction: The C2 Domain as a Ca2+-Dependent Electrostatic Switch. Neuron. 1997 Jan;18(1):133–42.
29. Bimpisidis Z, König N, Stagkourakis S, Zell V, Vlcek B, Dumas S, et al. The NeuroD6 Subtype of VTA Neurons Contributes to Psychostimulant Sensitization and Behavioral Reinforcement. eNeuro. 2019 May;6(3):ENEURO.0066-19.2019.
30. Kumar A, Srivastava P, Sirisena P, Dubey S, Kumar R, Shrinet J, et al. Mosquito Innate Immunity. Insects. 2018 Aug 8;9(3):95.
31. Cardoso-Jaime V, Tikhe CV, Dong S, Dimopoulos G. The Role of Mosquito Hemocytes in Viral Infections. Viruses. 2022 Sep 20;14(10):2088.
32. Aguilera P, López-Contreras AJ. ATRX, a guardian of chromatin. Trends in Genetics. 2023 Jun;39(6):505–19.
33. Lilley CE, Chaurushiya MS, Weitzman MD. Chromatin at the intersection of viral infection and DNA damage. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms. 2010 Mar;1799(3–4):319–27.
34. Tsai K, Thikmyanova N, Wojcechowskyj JA, Delecluse HJ, Lieberman PM. EBV Tegument Protein BNRF1 Disrupts DAXX-ATRX to Activate Viral Early Gene Transcription. PLoS Pathog. 2011 Nov 10;7(11):e1002376.
35. Gorman MJ, Paskewitz SM. Serine proteases as mediators of mosquito immune responses. Insect Biochem Mol Biol. 2001 Mar;31(3):257–62.
36. Cao X, Gulati M, Jiang H. Serine protease-related proteins in the malaria mosquito, Anopheles gambiae. Insect Biochem Mol Biol. 2017 Sep;88:48–62.
37. Brackney DE, Foy BD, Olson KE. The effects of midgut serine proteases on dengue virus type 2 infectivity of Aedes aegypti. Am J Trop Med Hyg. 2008 Aug;79(2):267–74.
38. Skorokhod O, Vostokova E, Gilardi G. The role of <scp>P450</scp> enzymes in malaria and other vector-borne infectious diseases. BioFactors. 2024 Jan 9;50(1):16–32.
39. Dzaki N, Ramli KN, Azlan A, Ishak IH, Azzam G. Evaluation of reference genes at different developmental stages for quantitative real-time PCR in Aedes aegypti. Sci Rep. 2017 Mar 16;7(1):43618.

40. Cai D, Liu L, Tian B, Fu X, Yang Q, Chen J, et al. Dual-Role Ubiquitination Regulation Shuttling the Entire Life Cycle of the Flaviviridae. Front Microbiol. 2022 May 6;13.
41. Pujhari S, Hughes GL, Pakpour N, Suzuki Y, Rasgon JL. Wolbachia-induced inhibition of O'nyong nyong virus in Anopheles mosquitoes is mediated by Toll signaling and modulated by cholesterol. bioRxiv. 2023 Jun 1;
42. Stokes S, Almire F, Tatham MH, McFarlane S, Mertens P, Pondeville E, et al. The SUMOylation pathway suppresses arbovirus replication in Aedes aegypti cells. PLoS Pathog. 2020 Dec 22;16(12):e1009134.
43. Souza-Neto JA, Sim S, Dimopoulos G. An evolutionary conserved function of the JAK-STAT pathway in anti-dengue defense. Proceedings of the National Academy of Sciences. 2009 Oct 20;106(42):17841–6.
44. Fikrig E, Main AJ, Cheng G, Anderson JF, Ferrandino FJ. Horizontal and Vertical Transmission of West Nile Virus Genotype NY99 by Culex salinarius and Genotypes NY99 and WN02 by Culex tarsalis. Am J Trop Med Hyg. 2012 Jan 1;86(1):134–9.
45. Grubaugh ND, Weger-Lucarelli J, Murrieta RA, Fauver JR, Garcia-Luna SM, Prasad AN, et al. Genetic Drift during Systemic Arbovirus Infection of Mosquito Vectors Leads to Decreased Relative Fitness during Host Switching. Cell Host Microbe. 2016 Apr 13;19(4):481–92.
46. Fitzmeyer EA, Gallichotte EN, Weger-Lucarelli J, Kapuscinski ML, Abdo Z, Pyron K, et al. Loss of West Nile virus genetic diversity during mosquito infection due to species-dependent population bottlenecks. iScience. 2023 Oct;26(10):107711.
47. Pandey A, Suman S, Chandna S. Predictive role of mitochondrial genome in the stress resistance of insects and nematodes. Bioinformation. 2010 Jun 24;5(1):21–7.
48. Gatti P, Ilamathi HS, Todkar K, Germain M. Mitochondria Targeted Viral Replication and Survival Strategies—Prospective on SARS-CoV-2. Front Pharmacol. 2020 Aug 28;11.
49. Barbier V, Lang D, Valois S, Rothman AL, Medin CL. Dengue virus induces mitochondrial elongation through impairment of Drp1-triggered mitochondrial fission. Virology. 2017 Jan;500:149–60.
50. Chatel-Chaix L, Cortese M, Romero-Brey I, Bender S, Neufeldt CJ, Fischl W, et al. Dengue Virus Perturbs Mitochondrial Morphodynamics to Dampen Innate Immune Responses. Cell Host Microbe. 2016 Sep;20(3):342–56.
51. Muralidharan S, Mandrekar P. Cellular stress response and innate immune signaling: integrating pathways in host defense and inflammation. J Leukoc Biol. 2013 Aug 29;94(6):1167–84.
52. Ambrose RL, Mackenzie JM. ATF6 Signaling Is Required for Efficient West Nile Virus Replication by Promoting Cell Survival and Inhibition of Innate Immune Responses. J Virol. 2013 Feb 15;87(4):2206–14.
53. Evans CG, Chang L, Gestwicki JE. Heat Shock Protein 70 (Hsp70) as an Emerging Drug Target. J Med Chem. 2010 Jun 24;53(12):4585–602.
54. Li DC, Yang F, Lu B, Chen DF, Yang WJ. Thermotolerance and molecular chaperone function of the small heat shock protein HSP20 from hyperthermophilic archaeon, Sulfolobus solfataricus P2. Cell Stress Chaperones. 2012 Jan;17(1):103–8.
55. Qian J, Vafiadaki E, Florea SM, Singh VP, Song W, Lam CK, et al. Small Heat Shock Protein 20 Interacts With Protein Phosphatase-1 and Enhances Sarcoplasmic Reticulum Calcium Cycling. Circ Res. 2011 Jun 10;108(12):1429–38.
56. Tamhankar M, Patterson JL. Directional entry and release of Zika virus from polarized epithelial cells. Virol J. 2019 Dec 8;16(1):99.
57. Ramirez L, Betanzos A, Raya-Sandino A, González-Mariscal L, del Angel RM. Dengue virus enters and exits epithelial cells through both apical and basolateral surfaces and perturbs the apical junctional complex. Virus Res. 2018 Oct;258:39–49.
58. Ahearn YP, Saredy JJ, Bowers DF. The Alphavirus Sindbis Infects Enteroendocrine Cells in the Midgut of Aedes aegypti. Viruses. 2020 Aug 4;12(8):848.

59. Taracena ML, Bottino-Rojas V, Talyuli OAC, Walter-Nuno AB, Oliveira JHM, Angleró-Rodriguez YI, et al. Regulation of midgut cell proliferation impacts Aedes aegypti susceptibility to dengue virus. PLoS Negl Trop Dis. 2018 May 21;12(5):e0006498.
60. Yang C, Wang F, Huang D, Ma H, Zhao L, Zhang G, et al. Vector competence and immune response of Aedes aegypti for Ebinur Lake virus, a newly classified mosquito-borne orthobunyavirus. PLoS Negl Trop Dis. 2022 Jul 18;16(7):e0010642.
61. Chen TY, Bozic J, Mathias D, Smartt CT. Immune-related transcripts, microbiota and vector competence differ in dengue-2 virus-infected geographically distinct Aedes aegypti populations. Parasit Vectors. 2023 May 19;16(1):166.
62. Sexton NR, Bellis ED, Murrieta RA, Spangler MC, Cline PJ, Weger-Lucarelli J, et al. Genome Number and Size Polymorphism in Zika Virus Infectious Units [Internet]. 2021. Available from: http://jvi.asm.org/
63. Dainat J. AGAT: Another Gff Analysis Toolkit to handle annotations in any GTF/GFF format.
64. Satija R. Seurat - Guided Clustering Tutorial. 2023.
65. Patel K. https://www.youtube.com/watch?v=04gB2owLKus. 2022. DESeq2 workflow tutorial | Differential Gene Expression Analysis | Bioinformatics 101.
66. Li WV, Li Y. scLink: Inferring Sparse Gene Co-expression Networks from Single-cell Expression Data. Genomics Proteomics Bioinformatics. 2021 Jun;19(3):475–92.
67. Cantalapiedra CP, Hernandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. 2021.
68. Huerta-Cepas J, Szklarczyk D, Heller D, Herńandez-Plaza A, Forslund SK, Cook H, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. 2019.

Figure 1.


A
B
Mock
WNV
Total
4dpi
12dpi
UMAP_2
UMAP_1
VM-1
VM-2
MT
HC-1
HC-2
cardia-prol
cardia-1
cardia-2
EC-like-1
EC-like-2
EC-like-3
EC
EE
ISC/EB-prol
ISC/EB

C

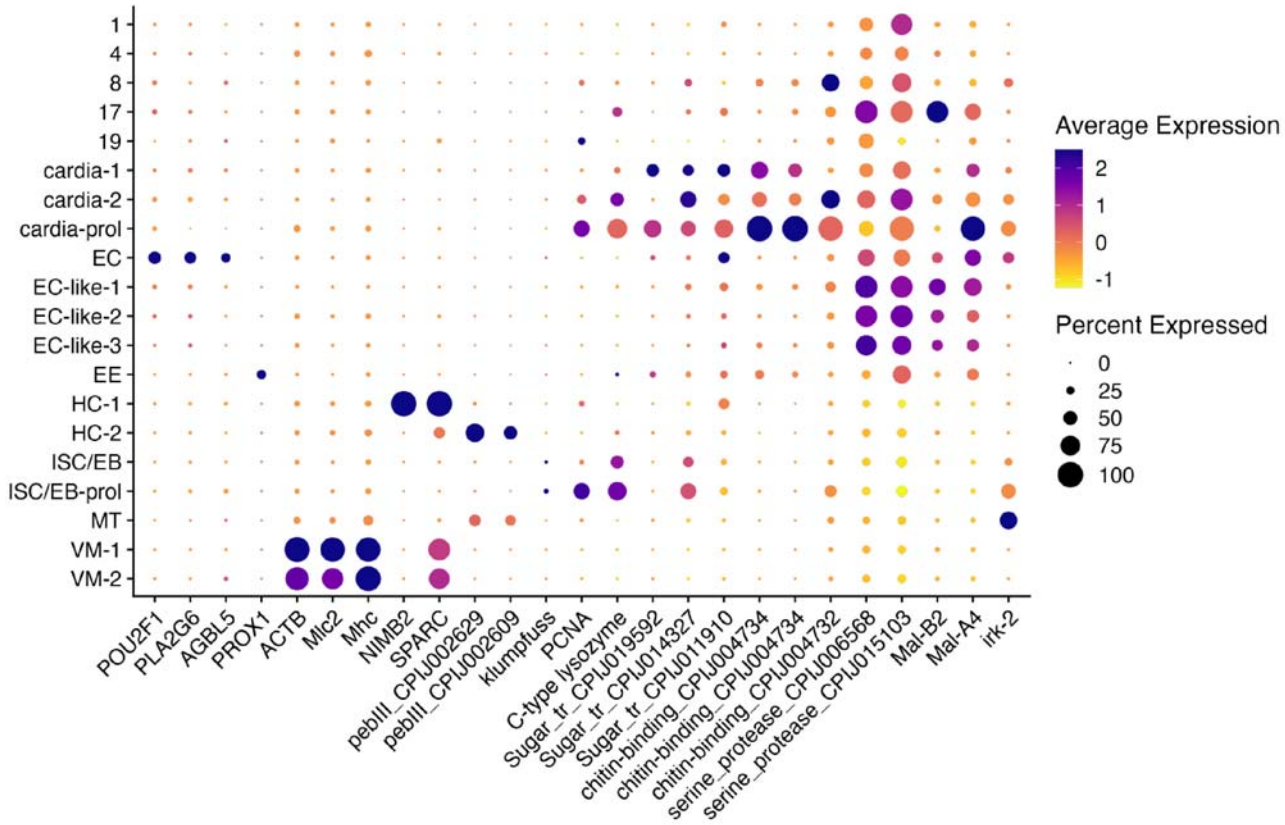
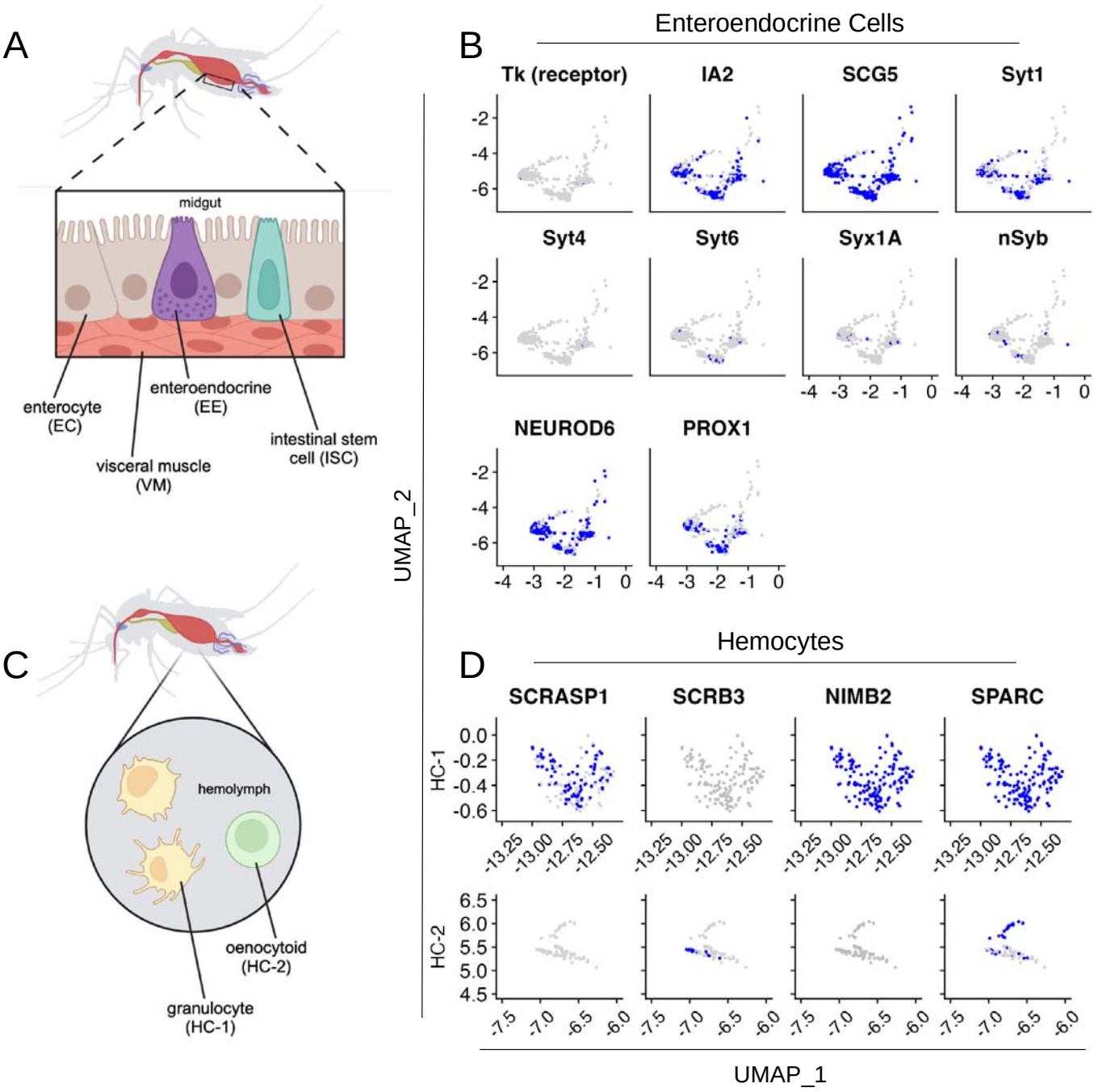
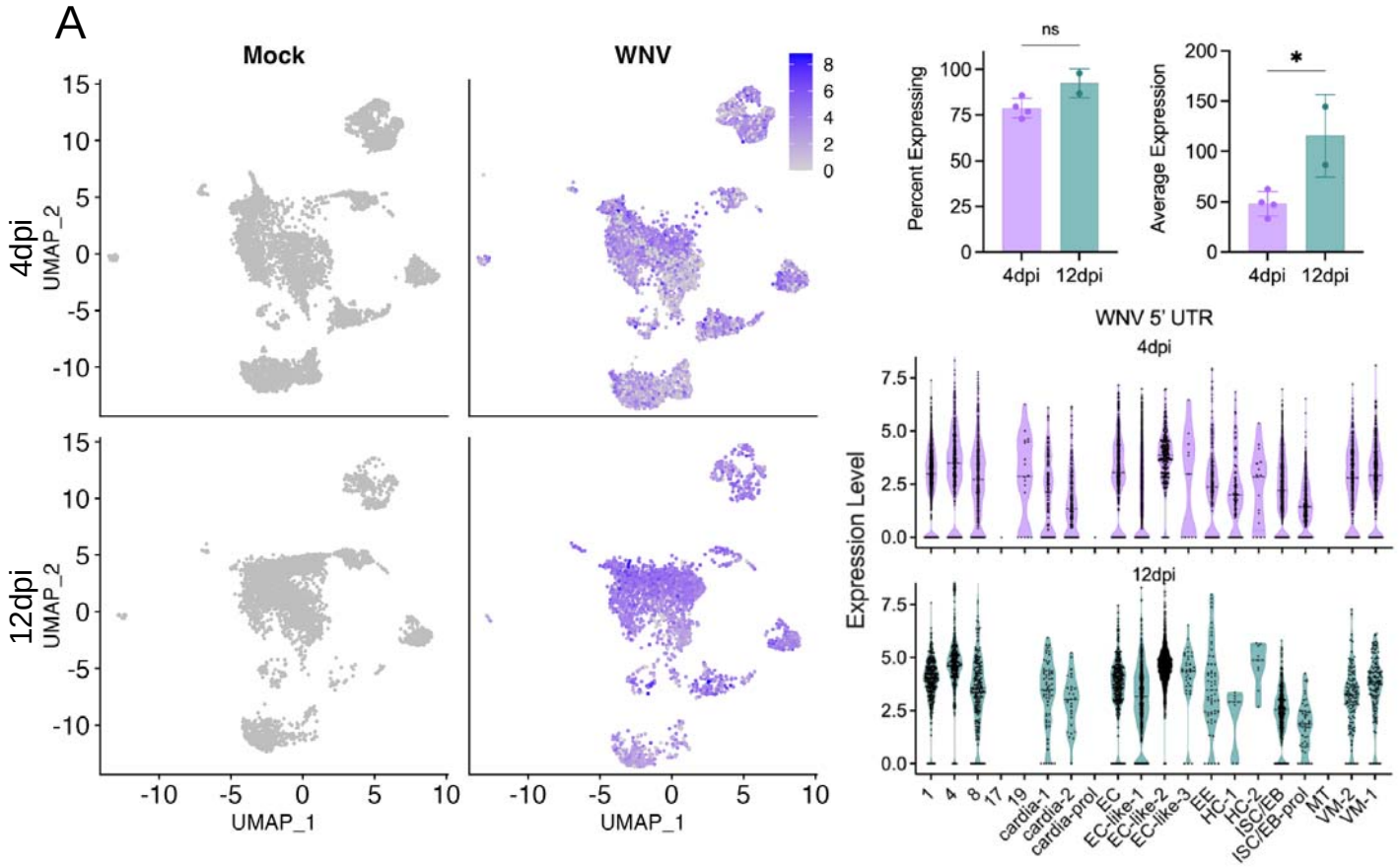

1
4
8
17
19
cardia-1
cardia-2
cardia-prol
EC
EC-like-1
EC-like-2
EC-like-3
EE
HC-1
HC-2
ISC/EB
ISC/EB-prol
MT
VM-1
VM-2
POU2F1
PLA2G6
AGBL5
PROX1
ACTB
Mlc2
Mhc
NIMB2
SPARC
pebIII_CPIJ002629
pebIII_CPIJ002609
klumpfuss
PCNA
C-type lysozyme
Sugar_tr_CPIJ019592
Sugar_tr_CPIJ014327
Sugar_tr_CPIJ011910
chitin-binding_CPIJ004734
chitin-binding_CPIJ004734
chitin-binding_CPIJ004732
serine_protease_CPIJ006568
serine_protease_CPIJ015103
Mal-B2
Mal-A4
irk-2
Average Expression
2
1
0
-1
Percent Expressed
0
25
50
75
100

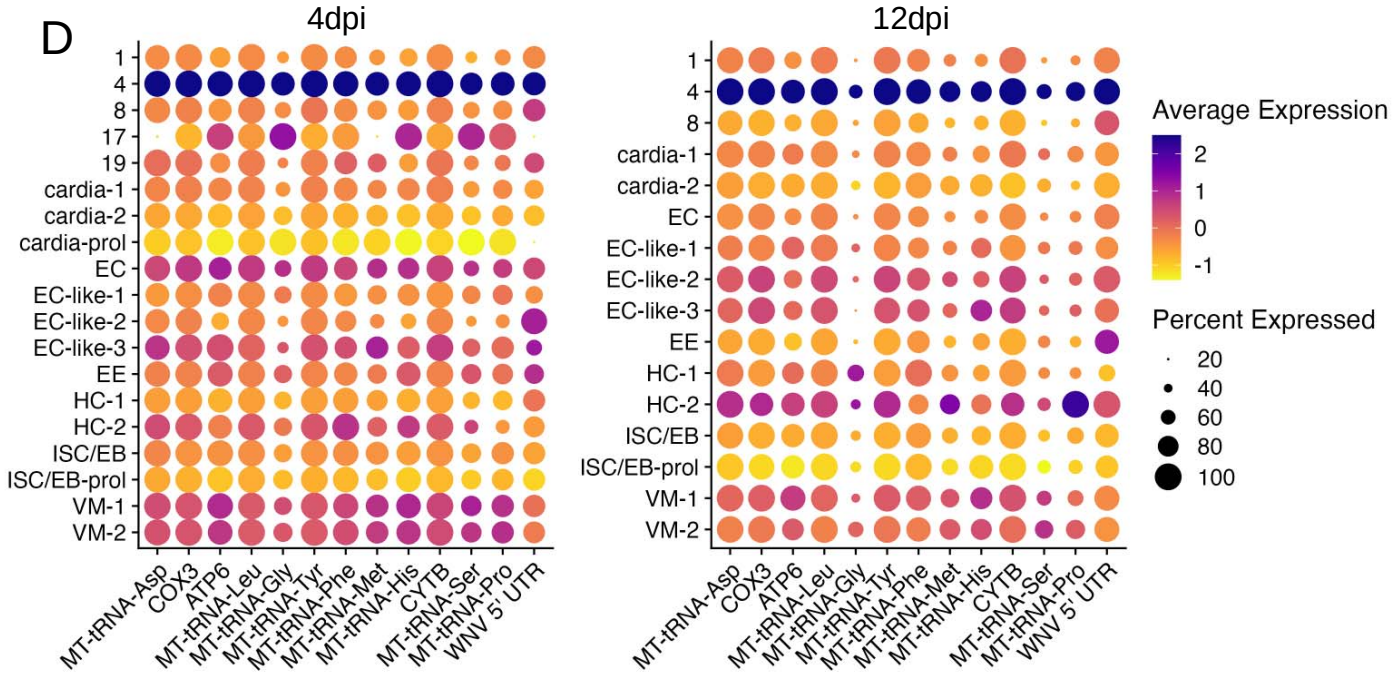
Figure 2.


A
midgut
enterocyte (EC)
enteroendocrine (EE)
visceral muscle (VM)
intestinal stem cell (ISC)
B
Enteroendocrine Cells
Tk (receptor)
IA2
SCG5
Syt1
Syt4
Syt6
Syx1A
nSyb
NEUROD6
PROX1
C
hemolymph
oenocytoid (HC-2)
granulocyte (HC-1)
D
Hemocytes
SCRASP1
SCRB3
NIMB2
SPARC
HC-1
HC-2
UMAP_2
UMAP_1

37

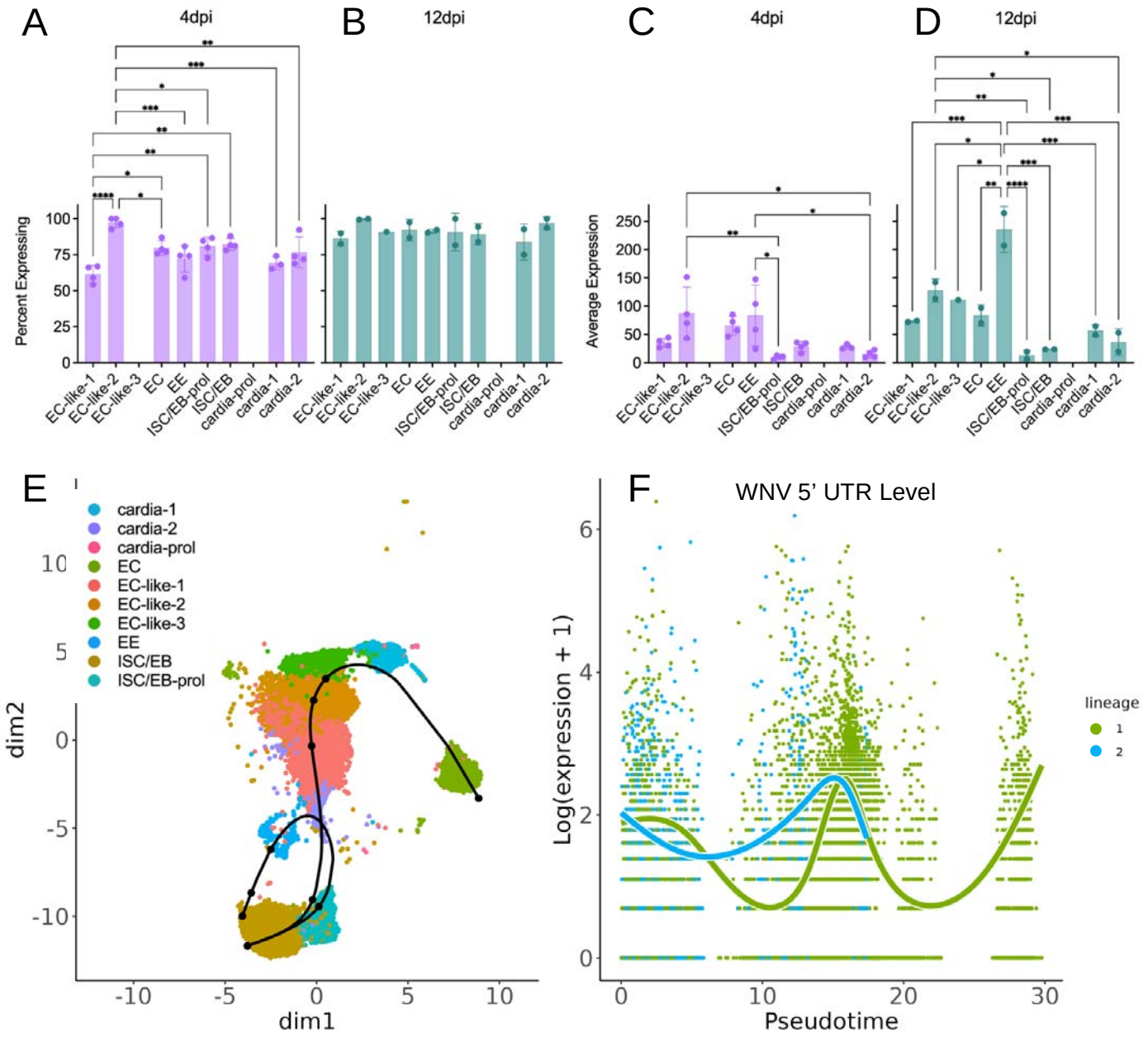
Figure 3.


A
Mock
WNV
4dpi
12dpi
UMAP_2
UMAP_1
ns
*
Percent Expressing
Average Expression
WNV 5' UTR
Expression Level
1
4
8
17
19
cardia-1
cardia-2
cardia-prol
EC
EC-like-1
EC-like-2
EC-like-3
EE
HC-1
HC-2
ISC/EB
ISC/EB-prol
MT
VM-2
VM-1


D
4dpi
12dpi
MT-tRNA-Asp
COX3
ATP6
MT-tRNA-Leu
MT-tRNA-Gly
MT-tRNA-Tyr
MT-tRNA-Phe
MT-tRNA-Met
MT-tRNA-His
CYTB
MT-tRNA-Ser
MT-tRNA-Pro
WNV 5' UTR
Average Expression
Percent Expressed

38

Figure 4.


A
4dpi
B
12dpi
C
4dpi
D
12dpi
Percent Expressing
Average Expression
EC-like-1
EC-like-2
EC-like-3
EC
EE
ISC/EB-prol
ISC/EB
cardia-prol
cardia-1
cardia-2
E
cardia-1
cardia-2
cardia-prol
EC
EC-like-1
EC-like-2
EC-like-3
EE
ISC/EB
ISC/EB-prol
dim1
dim2
F
WNV 5' UTR Level
Log(expression + 1)
Pseudotime
lineage
1
2

39

Figure 5.

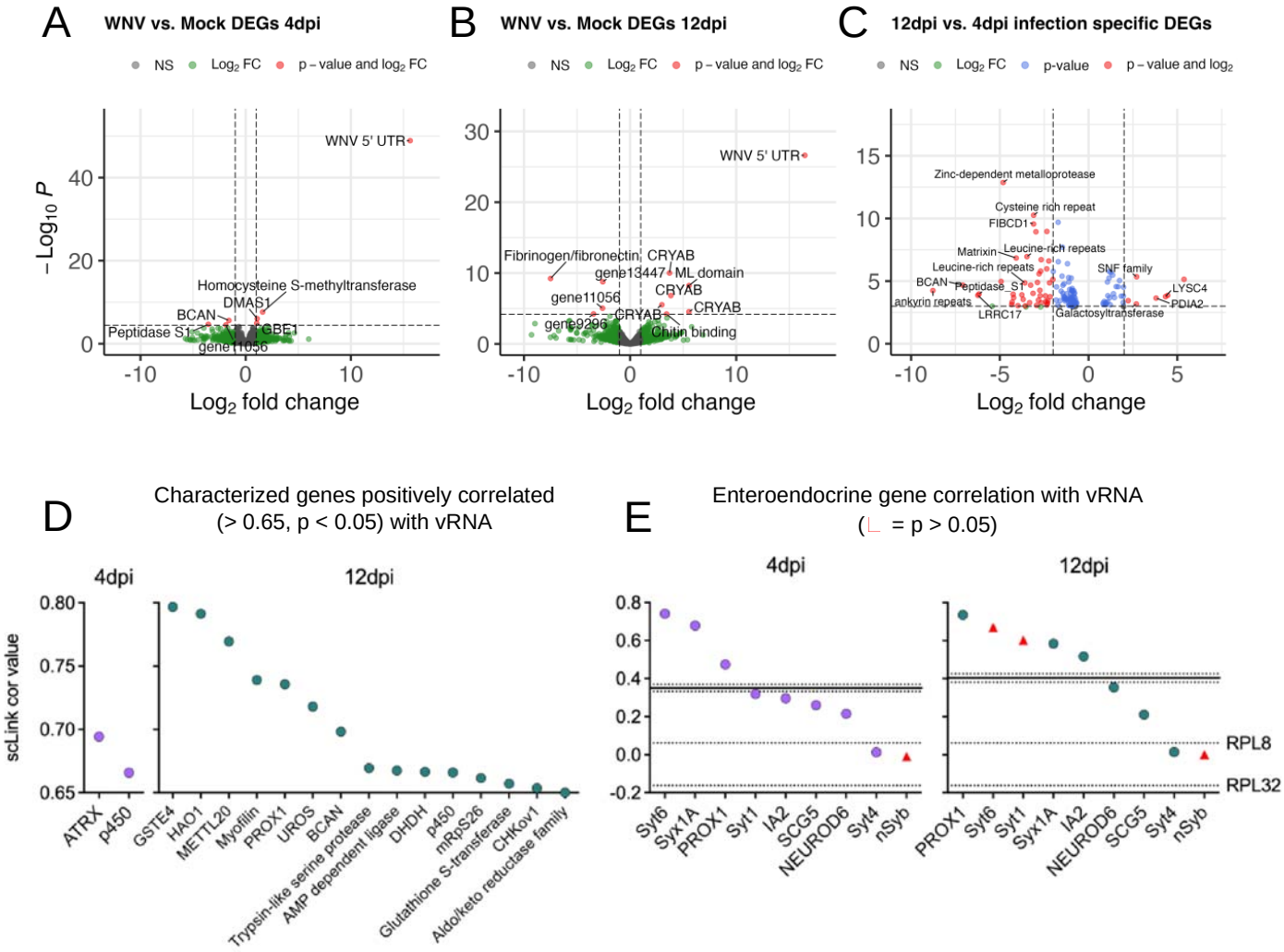
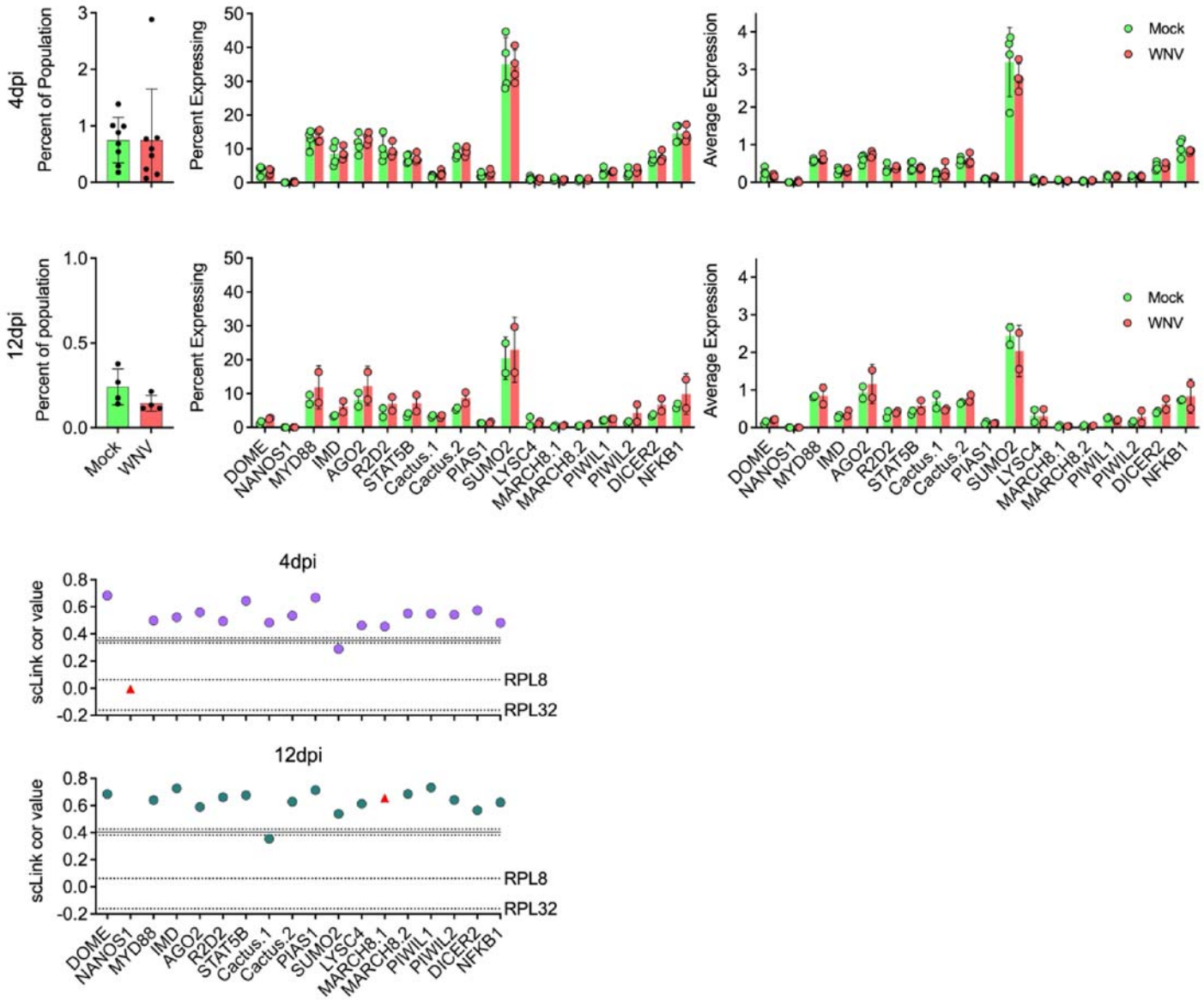

A
WNV vs. Mock DEGs 4dpi
NS
Log2 FC
p – value and log2 FC
WNV 5' UTR
Homocysteine S-methyltransferase
DMAS1
BCAN
Peptidase S1
GBE1
gene11056
– Log10 P
Log2 fold change
B
WNV vs. Mock DEGs 12dpi
WNV 5' UTR
Fibrinogen/fibronectin
CRYAB
gene13447
ML domain
gene11056
CRYAB
CRYAB
gene9296
CRYAB
Chitin binding
Log2 fold change
C
12dpi vs. 4dpi infection specific DEGs
NS
Log2 FC
p-value
p – value and log2
Zinc-dependent metalloprotease
Cysteine rich repeat
FIBCD1
Matrixin
Leucine-rich repeats
Leucine-rich repeats
SNF family
BCAN
Peptidase_S1
LYSC4
ankyrin repeats
PDIA2
LRRC17
Galactosyltransferase
Log2 fold change
D
Characterized genes positively correlated
(> 0.65, p < 0.05) with vRNA
4dpi
12dpi
scLink cor value
ATRX
p450
GSTE4
HAO1
METTL20
Myofilin
PROX1
UROS
BCAN
Trypsin-like serine protease
AMP dependent ligase
DHDH
p450
mRpS26
Glutathione S-transferase
CHKov1
Aldo/keto reductase family
E
Enteroendocrine gene correlation with vRNA
(= p > 0.05)
4dpi
12dpi
RPL8
RPL32
Syt6
Syx1A
PROX1
Syt1
IA2
SCG5
NEUROD6
Syt4
nSyb

Figure 6.


4dpi
Percent of Population
Percent Expressing
Average Expression
Mock
WNV
12dpi
Percent of population
DOME
NANOS1
MYD88
IMD
AGO2
R2D2
STAT5B
Cactus.1
Cactus.2
PIAS1
SUMO2
LYSC4
MARCH8.1
MARCH8.2
PIWIL1
PIWIL2
DICER2
NFKB1
scLink cor value
RPL8
RPL32