Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Single-cell transcriptomic characterization of AmE-711, a continuous honey bee (Apis mellifera) derived cell line

Melissa E. Flores

flores39@illinois.edu

University of Illinois Urbana-Champaign

Joel Rivera-Cardona

University of Illinois Urbana-Champaign

Michael Goblirsch

Agricultural Research Service - Southeast Area

Christopher B. Brooke

University of Illinois Urbana-Champaign

Gene E. Robinson

University of Illinois Urbana-Champaign

Research Article

Keywords: single-cell RNA-seq, in vitro, cell cycle, Deformed wing virus

Posted Date: September 8th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10706807/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background: AmE-711 was derived from honey bee embryonic tissues and established as a continuous cell line, the only one available to date. Morphologically, the cells display fibroblast-like qualities. AmE-711 is persistently infected with Deformed wing virus (DWV), a prevalent honey bee pathogen. We used single-cell RNA sequencing (scRNA-seq) to characterize its transcriptomic profile.

Results: We present a modification to the maintenance protocol for AmE-711 and a transcriptomic analysis. First, by increasing the temperature at which adherent cells are incubated during trypsinization to detach them from their substrate, we evenly split parent flasks to produce daughter flasks consistently seeded at high densities that reached confluence more quickly than previously reported. Using scRNA-seq analysis, we functionally annotated 13 clusters using *Drosophila melanogaster* protein-coding orthologs.

Cluster marker gene set enrichment analysis identified differentially expressed genes typically upregulated in nervous and muscle tissues. Markers were enriched for gene ontology terms including synaptic signaling, cell adhesion and localization, and sarcomere or cytoskeleton organization. Greater than 20% of cells expressed mitotic markers. All cells were infected with DWV, but clusters displayed heterogeneity in viral loads.

We used published honey bee scRNA-seq data to determine whether AmE-711 transcriptomes clustered with those of uninfected cells isolated from pupae undergoing metamorphosis or an adult worker brain. Despite clustering with mature Kenyon cells or neurons and muscles undergoing remodeling by plasmatocytes, the transcriptomes of most AmE-711 cells were distinct. Cell-cell communication analysis predicts that the third largest cluster may function as a hub for signaling events coordinating developmental, growth, or transcriptional programs that stabilize cellular identity.

Finally, we explored whether host gene expression levels were predictable in a DWV dose-sensitive manner. Our linear regression model identified that antiviral RNAi host defense and ribosomal gene expression is downregulated while cell adhesion and locomotion gene expression is upregulated when cellular DWV abundances increase.

Conclusions: As in previous studies, our results demonstrate the utility of scRNA-seq to characterize the heterogeneity of cell lines and indicate that AmE-711 is composed of DWV-infected cells of intermediate identities. We identified host cell gene programs that may help stabilize DWV infection, which could guide future approaches for treatment development.

Introduction

The western honey bee (*Apis mellifera*) is an excellent model for the study of genes and social behavior and a premier pollinating species. However, honey bees face several stressors, especially pesticides, pathogens, parasites, and poor nutrition due to loss of habitat and varied forage [1, 2]. Insect-derived cell

lines can serve as tools for the study of varied factors affecting honey bee health ranging from host-viral dynamics to sub-organismal processes such as the impact of pesticides on cellular stress responses. They also allow for the development of genetic tools including those that would be embryonically lethal [3]. Currently, most insect-derived cell lines use dipteran and lepidopteran tissues. The Biological Control of Insects Research Laboratory previously reported challenges when they were unable to establish a continuous honey bee-derived cell line despite initiating over 600 honey bee cell cultures [4].

To date, only one continuous cell line derived from honey bee embryonic tissues has been established. Goblirsch *et al*. [5] reported the establishment of AmE-711 (RRID: CVCL_2H47), a spontaneously immortalized cell line characterized morphologically to consist of fibroblast-like cells that were strongly adherent, not contact inhibited, and observed to occasionally contract. Carrillo-Tripp *et al*. [6] discovered that the cell line was persistently infected with an isolate of a highly prevalent pathogen of honey bees: Deformed wing virus (DWV), which they identified as being almost identical to isolate DWV-PA (GenBank Accession ID: AY292384).

DWV is transmitted in honey bee colonies both horizontally—as infected nurse bees feed the developing brood [7]—and vertically via eggs and sperm [8, 9]. Following a global spread of the honey bee ectoparasite and viral vector, the *Varroa destructor* mite, incidence of DWV, which encompasses three genetically similar positive-sense RNA viruses including the prevalent DWV Type A variant, Varroa Destructor Virus-1 (VDV-1) or DWV Type B variant, and Kakugo virus, has grown rapidly (reviewed in [10]). Whether the tissues used to establish the cell line were already infected or the cell line became infected due to cross contamination is unknown.

We were interested in determining whether AmE-711 cells were all terminally differentiated or whether any stem cell-like or reprogrammable lineage-specific progenitor subpopulations might be present. Moreover, AmE-711 cells have the potential to serve as a tool to investigate functional genomics despite persistent viral infection. In recent years, the cell line's potential has been demonstrated as studies have assessed the impact of insecticides and other pathogens of honey bees using AmE-711 viability and gene expression dynamics as readouts, showing its utility across a range of applications [11, 12].

Single-cell RNA-sequencing (scRNA-seq) allows for molecular characterization of even seemingly homogenous tissues. Its use has revealed unexpected heterogeneity in expression of cell lines. This includes heterogeneity of cell types in model cell lines used in breast cancer research [13] and heterogeneity of cellular and developmental states in more homogenous cell populations [14]. Luhur *et al*. [15] previously used a multi-omic approach to characterize novel intestinal-derived *Drosophila* cell lines, revealing heterogeneity attributable to the existence of both larval and adult precursors at various stages of differentiation during cell line establishment. In addition to canonical model organisms, scRNA-seq has revealed heterogeneity in an established cell line derived from the fall armyworm, which led to the identification of marker genes that can be used to select for homogeneous cell populations for downstream screening of insecticide targets [16].

Here, we first describe how a subtle change in the cell passaging protocol allowed us to successfully culture AmE-711 cells at a faster rate than described previously. This enabled us to use scRNA-seq to characterize 13 cell clusters within the cell line, identify cell line-specific marker genes, and use a comparative approach to glean cluster-specific functional information from *D. melanogaster* protein-coding orthologs. We also demonstrate how integration of published single-cell data from two samples of pooled early- and late-stage honey bee pupae as well as a whole brain sample from an adult worker honey bee [17, 18] can be used to compare *in vivo* gene expression in cells isolated from known tissues to the expression of heterogeneous cell programs in a previously uncharacterized *in vitro* culture environment. We further used our data to derive insights into the relationship between specific genes whose expression is sensitive to increasing viral loads in AmE-711 cells and compare these to genes encoding proteins previously demonstrated to co-immunoprecipitate with DWV [19].

Results

Modification of the trypsinization step during passaging produces daughter flasks with equal seeding densities.

AmE-711 cells display a fibroblast-like morphology with spindle-shaped, elongated cell bodies and cytoplasmic extensions (Figs. 1A–C). As cells proliferated, we observed an intricate network of short, thin projections connecting cells to one another along the peripheries of the flask where cell densities were decreased (Fig. 1A). They grow best when split heavily (at minimum 90% or greater than full confluence, Fig. 1B) and form long, branching extensions when seeded at too low a concentration. This often resulted in cells that failed to continue proliferating and eventually died. The cells are not contact inhibited and can form dense layers of cells strongly adherent to their substrate (Fig. 1C). However, AmE-711 cells proliferate slowly unlike other fibroblast cell lines, which can double within 24 to 72 h (e.g. see [20–23]).

The previously established protocol for passaging AmE-711 cells incubates them with trypsin at 32 ℃. This falls within the narrow temperature range at which honey bee colonies are optimally maintained by adult workers to promote healthy brood development [24–26] and is the temperature at which AmE-711 cells are routinely incubated. Because the cells are strongly adherent, trypsinization is necessary to detach them from their substrate as manual detachment alone is ineffective. However, three trials revealed that conducting this step at 37 ℃—the optimal temperature for trypsin, which can tolerate incubation temperatures up to 58 ℃ [27]—produced more robust daughter flasks with equally high seeding densities. On the other hand, trypsinization at 32 ℃ produced daughter flasks that split unevenly. Daughter flasks with low seeding densities grew slowly, due to the majority of cells remaining adherent to the parent flask.

We interpreted this to suggest that a sub-optimal seeding density stresses AmE-711 cells, causing them to become loosely adherent or die. Thus, the ability to consistently produce daughter flasks with high seeding densities can minimize this. Passaging the cells at 37 ℃ for no more than 10 min in 0.05%

trypsin-EDTA after flasks were fully or more confluent produced daughter flasks containing cells that grew faster compared to flasks passaged after cells reached 80% confluence or less. A heavy split allowed both parent and daughter flasks to reach 90% confluence or more in as quickly as six days.

Thus, a high seeding density in daughter flasks when splitting at a ratio of 1:2 or 1:3 appears to promote a faster doubling rate than the seven days reported previously [6]. No sign of cell stress (e.g., weak cell adherence, substantial cell death, or excess debris build-up) was observed after prolonged use of this passaging protocol. Using this temperature, cells continued to proliferate and form dense layers even after seven subsequent passages over the span of 2–3 months.

AmE-711 cells display heterogeneity in cluster-specific and mitotic markers as well as Deformed wing virus polyprotein RNA abundance

We prepared three dilutions of one sample for viability counts and scRNA-seq library preparation using a nearly confluent flask that had been split at a ratio of 1:2 nine days prior. 93.7% of sequenced reads were mapped to a concatenated viral (including DWV, VDV-1, sacbrood virus, Israeli acute paralysis virus, acute bee paralysis virus, and chronic bee paralysis virus) and *A. mellifera* reference genome (see Methods for alignment details). We observed an average ~ 102,000 reads/cell and median ~ 4,700 genes/cell. Prior to applying quality control metrics, 11,738 honey bee and viral genes were detected from 9,075 AmE-711 cells by CellRanger.

Following quality control to remove empty droplets, doublets, and damaged or dying cells from our sample matrix, *DWV polyprotein* (NCBI Gene ID: 1733515) RNA counts were detected in all cells (Fig. 1D) but counts for the other five viruses were not. Unsupervised clustering of 6,930 cells identified nine cell clusters (Fig. 1E). Because AmE-711 cells display fibroblast-like morphology, we expected that heterogeneity in the expression of genes encoding functionally significant extracellular matrix components would be among the most differentially expressed genes if no other terminally differentiated cell types were present in our sample. This would also depend on the expression profiles of the tissues from which these cells originated as has been demonstrated for other fibroblasts using scRNA-seq previously [28]. The “vst” method of Seurat’s “FindVariableFeatures” function uses Loess regression to fit a line for the relationship of the log of a gene’s variance and the log of its mean expression across cells. This makes it possible to transform raw RNA counts and stabilize variance in gene expression to identify genes that truly vary in their expression profile between cell types across a data set and minimize variance due to technical artefacts.

The most highly variable gene (HVG; Supplemental Fig. 1) was *no long nerve cord (nolo)*, which encodes a protein in fruit flies that is essential in ventral cord development by guiding extracellular matrix interactions with glia [29]. In fact, many among the top 5,000 most HVGs have known functions in muscle development, fibril and sarcomere organization, and cytoskeleton-associated functions based on gene ontology (GO) enrichment analysis (Table 1; Supplemental Fig. 1).

Marker analysis (Fig. 1I; Supplemental Data File 1) similarly identified differentially expressed marker orthologs that were enriched for GO terms centered on synaptic organization, neuronal development, and muscle cell differentiation in addition to terms centered on morphogenesis, development, and proliferation (Table 2; Supplemental Data File 1). Clusters with low orthologous coverage of markers were distinguished by species-specific and/or differentially expressed non-protein–coding genes; orthologous coverage ranged from 50 to 85%.

An additional four subclusters were annotated for 6,122 AmE-711 cells following a comparison to expression profiles from previously published honey bee single cell data sets in an integrated "tissue comparison" Seurat object. Integration allowed us to transfer cell type and state annotations based on clustering of AmE-711 cells with *in vivo* derived cells isolated from whole pupae and an adult whole brain (more details below; Fig. 1F; Supplemental Data File 1; Supplemental Fig. 2A). We created a reduced AmE-711 Seurat object that excluded AmE-711 cells for which transfer of annotations resulted in sparse clustering in the original AmE-711 Uniform Manifold Approximation and Projection (UMAP) [30] space or from cells in clusters dominated by pupal or brain sample cells (i.e., where AmE-711 cells made up less than 10% of a cluster; Fig. 2A; Supplemental Fig. 2A; Supplemental Data File 1).

7 out of 24 clusters (IntC04–09 and IntC24) were exclusively composed of or dominated by AmE-711 cells. GO enrichment analysis of cluster markers for these seven clusters revealed that genes active in morphogenesis, aerobic respiration, cytoskeleton processes including cell-cell junction formation, synaptic organization, and microtubule motor transporter activity distinguished these AmE-711 dominated clusters (Table 3; Supplemental Data File 1). Comparison of cluster marker expression for markers identified as differentially expressed in the integrated object showed subtle differences when their expression was examined in the AmE-711 Seurat object (Supplemental Fig. 2B; Supplemental Data File 1), supporting that these additional four communities represented transcriptionally distinct sub-communities.

AmE-711 post-comparison cluster C09 was distinguished by high *DWV polyprotein* RNA levels in cells with low total RNA counts and sparse expression of other host genes (Fig. 1D; Supplemental Figs. 3A–B). Just under 25% of cells were predicted to be in the G2/M cell cycle phase and just under 30% were predicted to be in the S phase based on their expression of 21 orthologs for canonical *D. melanogaster* G2/M cell cycle genes and 19 canonical S cell cycle genes. G2/M- and S-scored cells were concentrated in the post-comparison clusters C01, C02, C03, C08, and C09 (Figs. 1G–H; Supplemental Data File 1). This proliferative state distinguished the cells in C08 and C09, where no cells were scored as being in the G1 phase. Additionally, G2/M and S cell cycle markers comprised a high proportion of counts in the cells of the most highly DWV-infected cells in C09 where DWV may be co-opting genes involved in the G2/M or S phases for its own replication (Fig. 1H; Supplemental Figs. 3C–D). Cluster C02 was distinguished by cells with substantially larger proportions of ribosomal genes and mitochondrial genes despite lower transcript and gene abundances (Supplemental Fig. 3E–F).

Integration of four scRNA-seq samples confirms that AmE-711 cells do not express canonical cell fate commitment markers as strongly as cells isolated from organismal tissues

Given our findings from our analysis of AmE-711 cluster markers, we were uncertain as to whether the expression of canonical marker genes used to identify neurons and myocytes, such as *fne* and *Mef2*, was comparable to expression levels in cells isolated from organismal tissues. Lacking a healthy uninfected control sample due to a lack of treatment options for DWV removal, we decided to compare the expression in AmE-711 to expression of cells derived from animal tissues.

There are currently few available single-cell datasets from *A. mellifera* tissues. Previous work identified DWV-induced changes in brain tissues that seem to drive expression profiles overlapping with healthy glial cells [17]. Because our sample might contain proliferating and/or differentiating neuron-like cells, we chose to include the whole brain sample from this study to compare the transcriptomes of mature post-mitotic neurons to AmE-711 cells. We also included the two pupal samples sequenced by Patir *et al*. [18] who characterized honey bee metamorphosis-associated changes in pupae transitioning from larval to adult body plans (morphs) using pre-pupae and pale-eyed pupae.

We expected that cells in the pupal samples would capture transcriptional programs of cells undergoing massive remodeling during holometabolous development, or metamorphosis. Cells from the whole brain would capture the mature, postmitotic states of adult neurons and glia. We additionally hypothesized that the embryonically derived AmE-711 cells may reflect transcriptomic profiles akin to larval brain or muscle cells that express transcriptional programs driving differentiation of adult pre-cursor cells. Thus, we reasoned that fully differentiated cells in our AmE-711 sample should cluster with maturing or mature cell types from the pale-eyed pupae and/or the 7-days-post-eclosion adult whole brain. Additionally, pupal samples had been isolated and pooled from whole individuals, allowing us to compare cell line expression profiles to a greater assortment of cell types.

We followed the same workflow and used the reference genome we had created to align the raw sequences for each of the three samples from the published datasets that we had used for our AmE-711 sample. We aimed to minimize differences attributable to reference genome versions for *A. mellifera* and the six honey bee viruses described above, or software versions used by the authors of the original publications (see Methods). After alignment but prior to integration, we merged the four Seurat objects we created for each sample and clustered the merged cells following the same workflow used for the AmE-711 object in isolation to inspect the full dataset. Batch effects were evident between experiments that produced each sample dataset (Supplemental Fig. 4). Thus, we performed reciprocal PCA (RPCA) integration as this method is most effective when cells are in different biological states and less likely to align appropriately after integration due to overcorrection when compared to other anchor identification methods.

AmE-711 cells made up a substantial proportion of the integrated Seurat object (Figs. 2A–B). Just under 25% of the cells in the integrated object were scored as being in the G2/M phase, ~ 32% in the S phase, and ~ 43% in G1 (Figs. 2C–D). Unsupervised clustering of 12,328 cells in our integrated Seurat object

identified 24 clusters (Fig. 2E; Supplemental Fig. 5). Reads aligned to *DWV polyprotein* were detected only in AmE-711 cells (Fig. 2F; Fig. 3; Supplemental Data File 1) whereas no reads aligned to VDV-1, sacbrood virus, Israeli acute bee paralysis virus, acute bee paralysis virus, nor chronic bee paralysis virus.

To check for over-integration, we first used a local inverse Simpson's Index (iLISI) test [31] using the sample IDs to identify batch effects. We followed this with a cell-type LISI (cLISI) test [31]. The results supported the preservation of expected biological variation while reducing technical artefacts (Supplemental Data File 2).

We annotated clusters based on differential expression and GO enrichment analyses of cluster marker orthologs (Fig. 3A; Supplemental Data File 1) in addition to expression of canonical cell type orthologs (Fig. 3B; Supplemental Data File 1). Examination of the *z*-scaled expression of markers for mature neurons and muscles showed a clear demarcation between AmE-711 cells expressing these markers and tissue-derived cells (Fig. 3B; Supplemental Data File 1). For example, the expression of commonly used neuronal marker genes showed that cells strongly expressing these markers lacked *DWV polyprotein* expression whereas the same cells in those clusters that faintly expressed these markers did express *DWV polyprotein* RNA, identifying them as AmE-711 cells (Fig. 3B; Supplemental Data File 1). The same was true for muscle markers (Figs. 3A–B; Supplemental Data File 1). Interestingly, tissue comparison cluster IntC24, characterized by high *DWV polyprotein* RNA levels, displayed uniquely downregulated expression of canonical markers (Fig. 3B).

The same quality control metrics checked for the AmE-711 data revealed substantial variation for the cell line-derived cells compared to the tissue derived cells (Supplemental Fig. 6). For example, clusters dominated by AmE-711 cells such as IntC04–IntC07 expressed a greater number of genes per cell. Many pupal and AmE-711 dominated clusters expressed greater proportions of ribosomal and mitochondrial genes compared to adult brain clusters, likely due to cells undergoing massive amounts of energetically demanding remodeling.

Several neuronal and neuromuscular intermediate-identified cells clustered together, suggesting that AmE-711 transcriptomic profiles were most similar to those of mature or maturing mushroom body neurons called Kenyon cells and actively remodeling pupal neurons and muscles (Supplemental Fig. 7). Marker genes identifying Kenyon cell subtypes have been extensively studied in the mushroom bodies of adult worker bees, including identification through *in situ* hybridization, RT-qPCR, and single-nuclei RNA-sequencing [32–40]. The AmE-711 dominated clusters IntC04–09 and IntC24 often overlapped in their expression of markers for neurons, muscles, and plasmatocytes with clusters dominated by pupal isolated cells. Plasmatocytes are typically highly active during metamorphosis as they engulf remodeling tissues including larval neurons in addition to their role in insect innate immunity [41].

After reducing the gene matrix of our reduced AmE-711 Seurat object to 6,755 *D. melanogaster* orthologs expressed in the 6,122 cells with 13 transferred cluster identities, FlyPhoneDB2 was used to predict signaling activity between the 13 post-comparison designated cluster identities using the moderate-high

confidence curated database of ligand-receptor pairs mined from the literature and predicted by AlphaFold-Multimer [42].

We examined mean interaction scores among the top 25% highest scores to explore the predicted cell-cell signaling architecture between clusters. We identified one cluster, post-comparison cluster C04, predicted to act as a primary hub coordinating signaling events as it is predicted to act both as a cluster secreting ligands and as a cluster expressing receptors that bind ligands secreted by other clusters (Fig. 4A; Supplemental Data File 3).

Average expression of FlyPhone pathway genes was organized into five categories based on known functions in fruit flies. This highlighted two clusters that strongly expressed genes with known functions in the regulation of cellular development and growth as well as adhesion and cytoskeleton dynamics typical of mesenchymal cells. Meanwhile, genes with known functions in neurogenic axon guidance and myogenic priming decreased in expression in clusters with increasing viral RNA (Fig. 4B; Supplemental Data File 3). We observed ubiquitous expression of genes encoding important extracellular matrix components with consequences on neural and muscular development in fruit fly knockout embryos [43–46] (Supplemental Fig. 8). This supports the FlyPhone predictions for these gene modules focused on the highest quantile of mean interaction scores.

Increasing DWV may suppress cellular differentiation programs and promote metabolic and cytoskeletal remodeling

Finally, we explored whether there were host genes that varied in their expression as a function of DWV viral load and, if so, whether there might be functional significance in the types of genes that are dose sensitive. To do this, we first fit a two-part hurdle model using the MAST package's (v 1.36.0) zero-inflated regression model (*zlm*) for each of 10,142 genes evaluated in order to derive continuous coefficients and calculate variance-covariance matrices [47]. The first part of the model derived the probability that a gene was expressed at all followed by modeling the log-normalized counts in the presence or absence of the predictor. Each test produced a regression coefficient to capture the expected change in expression per 1 unit increase of *DWV polyprotein* RNA abundance (conditional on whether the gene's expression was > 0).

We identified 6,022 genes whose expression decreases with higher *DWV* RNA abundance and 4,117 genes whose expression increases (Supplemental Data File 4). 1,642 of these genes met our replication criteria (see Methods). Changes in host cell gene expression were small in magnitude (β_c slopes). Most genes exhibited modest changes across the observed range of cellular viral loads when we plotted the $\log_2$-FC between model-predicted gene expression levels in cells with the top and bottom 10% of cellular *DWV polyprotein* RNA abundances (Fig. 5).

GO term enrichment was performed for the 514 upregulated and 623 downregulated dose-sensitive fruit fly orthologs that met our requirements for replication. Dose-sensitive genes that decreased with increasing viral loads were enriched for GO biological process terms centered on translation, differentiation, development, morphogenesis, and RISC complex assembly (Supplemental Data File 4). Genes whose expression increased as viral loads increased were especially enriched for cytoskeleton-related and electron transport chain terms (Supplemental Data File 4).

Becchimanzi *et al.* [19] previously identified 144 proteins that co-immunoprecipitate with DWV. 52 of the genes encoding these proteins were among our replicated dose-sensitive genes, including *Arginine kinase (ArgK)*. The authors observed upregulation of this gene in infected larvae and adults and reported that DWV fitness decreased with *ArgK* knockdown. Model predicted values indicated a $\log_2$-FC > 0.25 for this gene (Fig. 5), suggesting that viral-induced host gene expression changes may not need to be large in magnitude to have substantial effects on DWV fitness.

When comparing bulk-RNA seq data of DWV-infected workers to a single cell dataset of uninfected brains, Traniello *et al*. [17] identified eight transcription factors (TFs) with target genes enriched in the bulk-level data, implicating these TFs as significant regulators of DWV infection on worker brain gene expression. Some of these same TFs were among our dose-sensitive genes significantly downregulated with increasing viral loads (Supplemental Data File 4)—including *protein hairy (hry)*, *abrupt*, and *homeobox protein Nkx-6.1*. Traniello *et al*. [17] had additionally identified seven immune function-associated genes that were shared among four datasets of DWV-infected workers that they analyzed. Two of these, *Idgf4* and *Desat1*, were also among our significantly upregulated genes sensitive to increasing viral loads. Furthermore, in AmE-711 cells, *Desat1* and *hry* were among the top 25 most differentially upregulated and downregulated genes respectively (Fig. 5; Supplemental Data File 4).

Discussion

Using single-cell RNA sequencing, we provide the first comprehensive transcriptomic characterization of the continuous honey bee cell line, AmE-711. Our results confirmed that all AmE-711 cells in our sample were persistently infected with DWV. The majority of cells in our sample exhibited a mesenchymal cell-like transcriptional state with expression of marker genes typically involved in basement membrane extracellular matrix component secretion [48, 49] which is vital for morphogenesis of nervous and muscular tissues (reviewed in [44]). This likely underlies their fibroblast-like morphology and suggests that mesodermal tissues fated to become muscle- or nervous-system associated cell types may have been prevalent in the tissues used during establishment of AmE-711.

A minority of the cells profiled here strongly expressed markers typically observed in maturing muscles and sensory neurons. When we compared the expression profiles of AmE-711 cells to those isolated from organismal tissues, they tended to cluster with pupal isolated muscles and neurons undergoing remodeling by plasmatocytes during metamorphosis. Plasmatocyte-like profiles may indicate host cell innate immune activation by the presence of *DWV* RNA or reflect transcriptional programs that are highly

active during pupal neuronal and muscular remodeling. Our molecular characterization of this cell line additionally showed that just under 25% of the isolated cells were actively proliferating, in line with the mitotic index of other immortalized insect cell lines.

An important observation consistent across our analyses is that AmE-711 cells express mixed-lineage gene programs typically associated with developing or mature fruit fly muscle and neuron identities. Integration of our cell line sample with cells isolated from adult and pupal tissues showed that pan-muscular and pan-neuronal marker genes are expressed at levels that distinguish them from expression *in vivo* by their downregulated relative expression. Instead, expression levels of these marker genes are more consistent with incomplete differentiation or lineage priming as the cells appear to lack expression levels typical for those committed to canonical cell fates. The co-expression of markers associated with neuronal, muscular, and immune-like (plasmatocyte) functions further support a model in which AmE-711 cells occupy a transitional transcriptional state rather than discrete, fully specified identities.

Mammalian fibroblasts are structural cells that are important for organ architecture and consequently are highly heterogeneous in their transcriptomes depending on their source [28, 50, 51]. Mesenchymal-like cell states are associated with cytoskeletal remodeling and cellular plasticity, enabling cells to maintain proliferative capacity while resisting terminal differentiation, in line with our marker gene expression and GO enrichment analyses. Our findings therefore place AmE-711 within a broader framework of heterogeneous cell lines that retain developmental flexibility while exhibiting stable growth *in vitro*.

Persistent infection seems to be a central factor shaping the transcriptomic landscape of AmE-711 cells. DWV viral reads have been detected throughout the body of infected larvae, pupae, and workers [8, 52–54]. Wu *et al*. [55] used an *in vitro* system derived from honey bee pupal heads and abdomens to demonstrate that DWV preferentially replicates in undifferentiated cells of pale-eyed pupae but becomes inefficient at replicating in pupal tissues derived later in development. The dampening of genes driving morphogenesis and antiviral responses with increasing viral loads predicted by our linear regression model as well as by FlyPhone may confer viral replication advantage that does not trigger apoptosis while maintaining cells in a mesenchymal-like state.

Downregulated genes with the greatest fold-changes between cells expressing the top and bottom 10% of *DWV polyprotein* RNA abundances included *Argonaute-2 (AGO2)*, *carbonic anhydrase-related protein 10 isoform X1 (CARPA)*, and two uncharacterized lncRNAs. Induction of stress-associated genes including immune-activated ones like *vir-1*, *hymenoptaecin*, *AGO2, Dicer-2*, and *Pgrp-s2*, which have all been previously shown to be upregulated in response to DWV infection [56–60] were among those downregulated as *DWV polyprotein* abundances increased.

Upregulated genes with the greatest fold-changes included two uncharacterized honey bee genes and *Papilin (Ppn)*, which encodes a secreted sulfated glycoprotein that is an important component of basement membranes. *Ppn* is known to affect muscle patterning and hematopoietic precursor cell maintenance but does not drive muscle differentiation [61–63]. Changes in large magnitudes in the

expression of these genes may be a response to viral replication, initiating apoptosis. Additionally, several of the targets downregulated as DWV proliferates are repressors of cell proliferation and guide embryonic patterning, including *hry*, *salm*, and *abrupt*.

Our analysis of predicted cell-cell communication further supports the presence of coordinated developmental and growth-related signaling within this heterogeneous population. Clusters enriched for intermediate neuronal-muscular phenotypes and those with higher infection levels emerged as primary hubs of inferred ligand–receptor interactions. The dominant pathways identified were associated with developmental regulation, cytoskeletal dynamics, and extracellular matrix interactions rather than canonical cell fate commitment. Notably, signaling programs linked to adhesion and axon guidance (reviewed in [64]) were among the most prominent, reinforcing the importance of cytoskeletal remodeling and spatial organization in these cells. Although these inferences are based on transcriptomic data and require experimental validation, they provide a coherent framework connecting cellular heterogeneity, intermediate identity, and functional signaling states.

Genes expressed in the embryonic material used to establish this cell line (eggs between 48- and 72-hours-post-oviposition) may have been favorable for viral proliferation. The proteome of honey bee embryos of this age are enriched for proteins involved in morphogenesis and larval organ development [65]. Viruses like SARS-CoV-2, HIV-1, and the oncogenic human herpesvirus 8 (HHV-8) have been previously shown to co-opt cytoskeletal regulatory mechanisms to favor their own replication [66]. Most often, this occurs through post-translational modifications during the initial stages of infection. Therefore, one interpretation may be that DWV infection contributes to the maintenance of an intermediate, proliferative cell state by dampening differentiation programs while sustaining metabolic and cytoskeletal activity that favors viral proliferation.

In addition to providing biological insight, our study has practical implications for the use of AmE-711 as a research tool. Our modification to the maintenance protocol for these cells should allow future researchers to consistently culture AmE-711 cells at high seeding densities. Higher infection levels may shift cells towards aerobic respiration while suppressing host functions like protein transport and ribosome biogenesis. This suggests that AmE-711 cells respond by limiting metabolic stress demands imposed by viral replication. Optimizing nutritive status that supports cell growth while limiting the energetic demands of viral replication may further improve doubling times.

Several limitations of this study should be acknowledged. First, the absence of an uninfected AmE-711 population restricts our ability to directly disentangle the effects of viral infection from intrinsic properties of this spontaneously immortalized cell line. Second, our analysis is based on a single biological sample, which limits the assessment of variability across passages. Third, functional validation of inferred signaling interactions and gene regulatory mechanisms will be necessary to confirm our hypotheses. Future work aimed at removing DWV infection, modulating viral load, or inducing differentiation along specific lineages will be critical to validate our predictions.

Conclusions

We successfully maintained and subcultured the continuous honey bee cell line, AmE-711 and report the first transcriptomic characterization of the cell line since its establishment in 2011. We confirm that the cell line is persistently infected with DWV and reveal that AmE-711 cells are a heterogeneous population exhibiting transcriptional signatures of multiple lineages while remaining in an intermediate, mesenchymal-like state. We show that persistent DWV infection is associated with coordinated but modest shifts in host gene expression that favor cytoskeletal remodeling, metabolic activity, and continued proliferation while suppressing differentiation and antiviral pathways. Our results support that scRNA-seq is a powerful tool for molecular characterization of heterogeneity at the single-cell level, even for cells that are expected to be largely homogenous.

Methods

Cell line maintenance and passage

AmE-711 cells were maintained at 32 ℃ in Schneider's Insect Medium (Sigma Aldrich) supplemented with 10% FBS (Sigma Aldrich). AmE-711 cultures were maintained in Corning™ 75 cm^2 U-shaped cell culture flasks (Fisher Scientific) tightly sealed with phenolic caps. Media warmed to 37 ℃ was replaced every 4–6 days. Cells were passaged when they had reached 95% or greater than full confluence using warm 0.05% trypsin-EDTA (Gibco) to release cells from the culture flask following a 10-minute trypsin incubation period at 37 ℃. We cryopreserved cells for long-term storage in a liquid nitrogen tank using a freezing medium consisting of Schneider's Insect Medium mixed with 10% FBS and 10% DMSO.

scRNA-seq AmE-711 library preparation

A medium consisting of a mixture of Schneider's Insect Medium (+ 10% FBS), 25 µg/mL DNase I, and 5 mM EDTA was prepared and vacuum-filtered for use in single cell isolation. After washing the cells with dPBS and detaching them from the culture flask by trypsinization, cells were counted using a hemacytometer to isolate three technical replicates of 0.1, 0.5, and 1.0 x 10^6 cells, respectively. Cells were stored on ice and immediately transported to the Roy J. Carver Biotechnology Center at the University of Illinois Urbana-Champaign. The three cell suspensions were subjected to a total cell count and %viability assay with AOPI staining on a Nexcelom Cellometer Spectrum Image Cytometry System (Nexcelom Biosciences). Viability was > 97% for each sample.

AmE-711 libraries were constructed using a NextGEM Single-Cell 3 v3.1 kit (10X Genomics). The library pool was amplified via qPCR and loaded on two SP lanes of a NovaSeq 6000 S4 flow cell (Illumina). Over 928 million cDNA single reads (28 x 150 nt) were produced.

Identification of 1:1 D. melanogaster *orthologs by conditional reciprocal best hits (CRBH) codon and protein (BLASTp) Alignments*

We used a comparative approach to annotate our clusters with functional annotations known in fruit flies, identifying 1:1 protein-coding orthologs using two methods. First, we used the coding sequence (CDS) analysis pipeline CRBHits for R [67], which implements the conditional reciprocal best hit (CRBH) algorithm “crb-blast” to identify orthologous codons between *A. mellifera* (GCF003254395.2_Amel_HAv3.1_CDS_from_genomic.fna from the NCBI’s taxon 7460 genome FTP site) and *D. melanogaster* (GCF000001215.4_Release6_plusISO1MT _CDS_from_genomic.fna from the NCBI’s taxon 7227 genome FTP site). We kept only the Entrez Gene IDs for “rbh”-scored codon alignments as high-confidence 1:1 orthologous pairs and excluded secondary “sec”-scored hits.

We then used the NCBI BLAST+ tool [68, 69] to conduct protein BLAST queries by searching protein databases. We first built our databases using protein sequences for both *A. mellifera* (GCF003254395.2_Amel_HAv3.1_protein.faa) and *D. melanogaster* (GCF000001215.4_Release6_plusISO1MT_protein.faa). We submitted a BLASTp query for alignments from *A. mellifera* to our *D. melanogaster* protein database and subsequently from *D. melanogaster* to our *A. mellifera* protein database that met the following parameters: minimum e-value of 1e-5; best hit score edge of 0.1; best hit overhang of 0.1, and max target sequences of 1.

We proceeded with entries that had a ≥ 30% identity score and kept those with maximum bit scores for each query ID in both queries. This identity threshold was sufficient to identify the same ortholog identified by CRBHits using codon alignments, highlighting how orthologs identified by coding sequence similarity and protein sequences can validate orthologous pairs that may otherwise have been filtered for low alignment at the other level. We then kept only pairs where the query ID for the *A. mellifera* query matched the subject ID for the *D. melanogaster* match and vice versa, assigning each pair a unique ID. Finally, we identified the Entrez Gene IDs encoding each protein by matching IDs for each species from the GFF files available at the FTP sites referenced above. We merged this list of Entrez Gene IDs with the one from our CRBHits results, keeping unique pairs identified by only one method or by both but excluding any with results from both approaches that were not in agreement. This approach produced a list of 6,808 high-confidence orthologous pairs of protein-coding genes.

Bioinformatic analysis of AmE-711 cells

The FastQ files generated were demultiplexed with CellRanger (v7.1.0, 10X Genomics)[70]. The data were aligned to the *A. mellifera* HAv3.1 reference genome (NCBI Annotation Release 104, GCF_003254395.2) which was concatenated with seven viral genomes for: acute bee paralysis virus (GCF_000856345.1), chronic bee paralysis virus RNAs 1 and 2 (GCF_000875145.1), Deformed wing virus (GCF_000852585.1), Israeli acute paralysis virus (GCF_000870485.1), sacbrood virus (GCF_000847625.1), and varroa destructor virus-1 (GCF_000856945.1). We then converted the GFF and assembly report files from NCBI into GTF files using the rtracklayer package in R (v4.5.3) for compatibility with CellRanger and to identify mitochondrial gene (mtDNA) accessions for downstream quality control. Only protein-coding and lncRNA regions were considered for alignment.

Downstream analyses were conducted with Seurat (v5.4.0) [71] using the filtered feature matrices produced by CellRanger. To create a Seurat Object, we required genes to be detected in at least 10 cells and for each cell to contain at least 200 genes to remove lowly expressed genes and empty droplets. To assure cell viability, percent mtDNA was calculated to filter out dying/damaged cells. Transcript counts (nCount_RNA) were used to identify and filter doublets.

Cells with greater than 3 median absolute deviations (MAD) of mtDNA (> 7.17%) and above 3 MAD unique molecular identifiers (UMI) were removed from subsequent analysis using the “isOutlier” function from the package scater (v1.34) [72]. Counts were log-normalized and scaled with default parameters. We then identified the top 5,000 HVGs using the “vst” selection method to identify stably expressed genes with high degrees of variation between cells. We then performed linear dimension reduction using principal component analysis (PCA) of the top 50 components. We performed GO enrichment analysis to assign HVGs with functional annotations after filtering the gene set to the 53% of genes with 1:1 *D. melanogaster* orthologs. The clusterProfiler package (v4.18.4) [73] identified GO terms with Benjamini-Hochberg adjusted *p*-values < 0.05.

An elbow plot was used to determine the dimensionality of the dataset and this parameter was set to 30 based on the observed plateau. A *k*-nearest-neighbor graph was constructed based on Euclidean distance in PCA space followed by application of the Louvain algorithm. Unsupervised hierarchical clustering was performed for resolution parameters ranging from 0.1 to 1.0 increasing by intervals of 0.1 at a time. A resolution of 0.3 was chosen based on the cluster plot from the “Clustree” package (Supplemental Fig. 5A) [74]. This package visualizes how the membership of individual cells to a given cluster changes at higher resolutions to identify unstable clusters. Non-linear dimensional reduction for visualization was then performed with UMAP [30] to calculate cell embedding coordinates using 10 for the n.neighbors parameter to prioritize preservation of local structure for our single AmE-711 sample.

Cell cycle scoring was conducted on the normalized data based on expression of orthologous canonical G2/M and S phase marker genes available on the Tinyatlas Github Repository [75]. Seurat’s “CellCycleScoring” function assigns scores to each cell based on its expression of canonical markers during the S and G2/M phases as these marker sets are expected to be anti-correlated in their expression levels. A G1 phase score is then assigned to cells that could not be confidently assigned to either G2/M or S phases.

First, to identify marker genes for each cluster, differential expression analysis was performed for each of the 9 clusters (pct.1) against all other clusters (pct.2). We used the default Wilcoxon test with a fast implementation by the Presto package for the Seurat “FindAllClusterMarkers” function. We focused our test to identify marker genes with a minimum log-fold change threshold of 0.5 between clusters and that showed a minimum difference of 0.25 in the fraction of detection (min.pct.diff) between the two groups. After filtering out genes with an adjusted *p*-value ≥ 0.05 we used 7,803 marker orthologs for downstream functional analyses. This represented 74% coverage with *Drosophila* orthologs for our full list of 10,569 marker genes.

Cluster annotation was performed using a combination of approaches (results for all tests performed are available in Supplemental Data File 1). First, we used our orthologous cluster markers to perform GO enrichment analysis to provide functional annotation for the 9 clusters. We next mined canonical marker genes from the honey bee and fruit fly literature, including the expression of curator markers and top computed markers for various tissues in the Fly Cell Atlas [76]. We further mined fruit fly developmental characterization studies, including single-cell sequencing data, to identify orthologous markers for processes including neuronal, muscular, and hemocytic maturation and immune-signaling.

Finally, the cluster marker ortholog gene sets were used to perform marker enrichment analysis using the Multi-Gene List Enrichment tools of the *Drosophila* single-cell Database (DRscDB) [77]. This tool identified whether orthologous markers were enriched among the top 100 marker genes in all available *Drosophila* single-cell gene sets for matches with Benjamini-Hochberg adjusted *p*-values < 0.01 (Supplemental Data File 1).

Bioinformatic analysis of published whole pupal and adult brain samples for in vivo *tissue comparison to* in vitro *AmE-711 cell line gene expression profiles*

To compare cell-line expression to *in vivo* tissue expression, FastQ files from two previously published single-cell honey bee studies [17, 18] were downloaded from the European Nucleotide Archive (ENA Project PRJEB45881) and NCBI Gene Expression Omnibus (GEO ID GSE130785) repositories and aligned to the same concatenated reference genome described above. These three samples included: 1) a sample of four pooled pre-pupae collected 2 days after their brood cells had been capped inside the colony (hereafter referred to as D11 pre-pupae); 2) a sample of four pooled pale-eyed pupae collected 6 days after brood cells had been capped inside the colony (D15 pale-eyed pupae); and 3) a whole brain sample from an adult worker honey bee collected 7 days after eclosion from a source colony with a queen instrumentally inseminated with semen from a single drone.

Patir *et al*. [18] reported that approximately 7,000 cells from each set of pooled samples were used for cDNA library preparation using the Chromium platform v2.0 (10X Genomics), generating 180 million reads. Using the concatenated honey bee and viral reference genome we created for alignment of AmE-711 reads, we mapped 93.6% of raw sequenced D11 pre-pupal reads with an average of ~ 24,651 reads/cell and a median of ~ 1,850 genes/cell. From an estimated 2,808 cells, 11,204 genes were detected. We mapped 93.0% of raw sequenced D15 pale-eyed pupal reads with an average of ~ 20,300 reads/cell and a median of ~ 2,300 genes/cell. From an estimated 2,689 cells, 11,394 genes were detected.

Traniello *et al*. [17] reported that approximately 1,300 cells from an adult whole brain were used to construct cDNA libraries. In their experiment, they used a Chromium Single Cell 3′ Library and Gel Bead Kit v2 loaded on a HiSeq 4000 lane (Illumina), generating ~ 186 million reads. We mapped 93.8% of raw sequenced reads to our concatenated honey bee and viral reference genome with an average of ~ 139,000 reads/cell and a median of ~ 1,500 genes/cell. 10,538 genes were detected from an estimated 1,337 cells. [78].

The filtered feature matrices were pre-processed as described above for our AmE-711 sample. Cells with greater than 3 MAD of mtDNA (> 2.19% for D11 cells; > 6.96% for D15 cells; and > 7.17% for adult brain cells) and greater than 3 MAD RNA counts were removed from subsequent analysis. Cell filtering of low-quality cells allowed us to integrate 6,930 AmE-711 cells with reads for 11,398 genes; 1,079 D11 pre-pupal cells with reads for 8,726 genes; 2,093 D15 pale-eyed pupal cells with reads for 9,818 genes; and 2,226 adult whole brain cells with reads for 10,254 genes.

Data were normalized and the variance due to molecular count data stabilized using the Seurat "SCTransform" v2 regularization function using the glmGamPoi method (v1.18.0) [79, 80] for our merged Seurat object. We used all cells rather than a subset to fit the negative binomial regression model and did not subset the Pearson residuals from the model to include only highly variable features when using the "ScaleData" function in Seurat. A benchmarking study [81] of atlas-level data integration methods cautioned that HVG selection during pre-processing can improve batch removal but at the cost of conserving biologically meaningful information. For example, it can skew integrated data such that some biological functions such as changes in cell cycle state are only best captured when a full gene set is used [82]. Data stabilization was followed by linear dimension reduction of the first 50 principal components.

We next used the RPCA integration method, which projects each dataset onto the PCA space of the others to constrain the anchors selected by the same mutual neighborhood requirement. The faster and more conservative RPCA-based integration approach makes it possible to integrate data from samples that are biologically very different, such as between cells isolated from an immortalized cell line and cells isolated from organismal tissues. Similarly, we anticipated that mitotic and differentiating pupal brain cells would be less likely to cluster with post-mitotic brain cells isolated from the 7-day-old adult worker bee. As such, when integrating our datasets, we increased the k.anchor parameter to 10. We again selected a dimensionality of 30 based on the results of an Elbow Plot and repeated our process for identifying an optimal resolution for clustering (Supplemental Fig. 5B) and non-linear dimensional reduction.

We calculated cell cycle scores after re-joining our layers and log-normalizing the integrated data. We identified differentially expressed markers for each of the 24 clusters following the workflow described previously. We again achieved 74% coverage of the full list of 23,977 markers. These 17,681 orthologous cluster markers were thus used to perform GO enrichment analysis and then used to mine the aforementioned databases (Supplemental Data File 1).

FlyPhoneDB2: Signaling pathway analysis

To explore transcriptomic patterns associated with cell-cell communication in a given sample source, we used the FlyPhoneDB2 tool [42, 83]. The FlyPhone R package (v 0.1) can be used to predict communication events between cell types by using the expression patterns between known *Drosophila* ligand and receptor pairs. Building on its original model [84], FlyPhoneDB2 infers cell-cell communication

through the co-expression of 1,804 ligand-receptor pairs. These draw from a manually curated database (v2.5) compiled via a thorough literature review and those predicted by AlphaFold-Multimer, including from orthologs of ligands and receptors in mammalian species from CellTalkDB [85]. We limited our analysis to the ligand-receptor pairs in the “high–moderate” confidence knowledgebase performing 2,000 permutation tests to generate *p*-value scores for genes with a percent expression greater than 25% across all cells.

MAST Regression: DWV dose-sensitive differential gene expression in AmE-711 cells

First, we extracted the log-normalized counts of *DWV polyprotein* (NCBI Gene ID: 1733515, denoted as the log(1 + *x*) of the DWV counts below) stored in the RNA data layer of the AmE-711 cells to use as a continuous covariate. Doing so considers viral expression as a predictor of the expression of a given host cell gene. We adjusted our model to include normalized % mitochondrial reads and the number of unique genes detected per cell as covariates to control for technical artefacts. We additionally included cluster membership to account for the influence of cell type or cell state on gene expression profiles.

We performed the same test 20 times on genes detected in at least 75 cells using a random sub-sample of cells with a cap of 300 cells from any one cluster during each replication to control for the overrepresentation of large clusters. This allowed us to perform our test for 10,142 genes. With only one biological replicate (and thus because we consequently could not perform differential expression analysis with pseudobulking), we wanted to perform this test in as rigorous and unbiased a way possible to avoid pseudo-inflated results.

A likelihood ratio test for the DWV coefficient produced a joint hurdle *p*-value based on whether the model was improved by the presence of the viral term. These *p*-values were subsequently adjusted using a Benjamini-Hochberg correction. We computed a replication score (n_sig) that qualified the number of seeds (rep_fraction) in which a given gene met a hurdle *p*-value < 0.05 and the proportion out of 20 that this was the case. We also examined the regression coefficients to see whether their direction was consistent across the majority of replications by calculating whether model-predicted slopes were most often negative, positive, or tied (majority_sign), evaluating the consistency of the direction of the slope coefficient. We reasoned that if DWV loads were influencing the expression of a given gene, expression should be consistently changing in response to DWV and not another factor and thus negative slopes would detect genes downregulated by increasing viral loads and positive slopes would detect those upregulated with increasing viral loads. Thus, we excluded genes designated as “tied” for the majority_sign. Finally, we calculated a Fisher meta-analysis FDR for each gene across the 20 replications of the test.

We classified a gene as “replicated” when the Fisher FDR < 0.05 and the rep_fraction was significant in $k \geq 7$ seeds. We performed GO enrichment analysis for the intersection of genes identified by our linear regression model with those encoding proteins previously co-immunoprecipitated with DWV [19],

separating upregulated from those downregulated to identify functions associated with DWV-sensitive genes or protein-encoding genes in both *in vitro* (in AmE-711 cells) and *in vivo* (in infected animals from Becchimanzi *et al*.) conditions. To determine whether β_c coefficients from the continuous part of our two-part model translated to small or large fold-changes when comparing between the top and bottom 10% of infected cells, we calculated the $\log_2$ transformed product of a gene's average regression slope (mean β_c) and the expression range of 4.47 between the DWV quantiles.

Abbreviations

AmE-711
Apis mellifera embryo-derived continuous cell line established July 2011
DWV
Deformed wing virus
VDV-1
varroa destructor virus
scRNA-seq
single-cell RNA-Sequencing
HVG
highly variable genes
GO
gene ontology

Declarations

Ethics, consent to participate, and consent to publish declarations: not applicable

Disclaimer

Reference to any specific commercial products, processes, or services by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply endorsement, recommendation, or favoring by the U.S. Government. The views and opinions of the authors expressed herein do not necessarily state or reflect those of the U.S. Government and shall not be used for advertising or product endorsement purposes.

Funding

This work was funded by the National Science Foundation (NSF) DBI Biology Integration Institutes Program (Award #2022049), U.S. Department of Agriculture-Agricultural Research Service (USDA-ARS)-Congressional grant "Bee Genomics in the Midwest" (Award #AJ978), an NSF Graduate Research Fellowship, and a University of Illinois Urbana-Champaign Aspire Predoctoral Fellowship.

Author Contribution

M.G. provided AmE-711 aliquots, per material transfer agreement, and cell culture technical expertise, made possible by funding from C.B.B. and G.E.R. M.E.F. performed cell culture maintenance and passaging with technical expertise guidance from J.R.-C. M.E.F. and J.R.-C. performed the molecular isolation of single cell aliquots for cell counts and subsequent library preparation conducted at the DNA Services Laboratory within the Roy J. Carver Biotechnology Center at the University of Illinois Urbana-Champaign. M.E.F. performed the bioinformatic analysis of the cell line single-cell RNA-Sequencing data produced in addition to analysis of previously generated single-cell RNA-Sequencing datasets. M.E.F. and G.E.R. interpreted the results and wrote the manuscript. All authors have read and approved the final manuscript.

Acknowledgement

We are grateful to Dr. Alvaro Hernandez, Chris Wright, and Dr. Fangxiu Xu of the DNA Services Laboratory for their expert assistance in experimental planning in addition to preparation and sequencing of single-cell RNA-Sequencing libraries. We would also like to thank Dr. Jenny Drnevich, Jessica Holmes, and Dr. Yifei Kang of the High Performance Computing in Biology group, also housed in the Roy J. Carver Biotechnology Center, for their expertise in data analysis offered in their Advanced Single-cell RNA-Sequencing Analysis Workshop. We thank Dr. Adam Dolezal for valuable insights into cell line culturing best practices during initial conceptualization of this project. We also thank Drs. Anna Sokac and Paul Bonthuis, members of M.E.F.'s Diagnostic Committee, for their thoughtful feedback and evaluation of the initial draft of this manuscript presented to the Neuroscience Program at the University of Illinois Urbana-Champaign for M.E.F.'s First Year Project. Finally, we also thank members of the Robinson Lab for their valuable feedback improving this manuscript.

Data Availability

The datasets analyzed during the current study are available in the European Nucleotide Archive (ENA Project: PRJEB45881) and National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO ID: GSE130785) repositories. Raw sequences for the AmE-711 sample have been deposited at the NCBI Sequence Read Archive in the BioProject database (BioProject ID: PRJNA1492309).

Competing interests

The authors declare no competing interests.

References

1. Branchiccela B, Castelli L, Corona M, Díaz-Cetti S, Invernizzi C, Martínez de la Escalera G et al. Impact of nutritional stress on the honeybee colony health. Sci Rep. 2019;9:10156. https://dx.doi.org10.1038/s41598-019-46453-9.
2. O'Neal ST, Anderson TD, Wu-Smart JY. Interactions between pesticides and pathogen susceptibility in honey bees. Curr Opin Insect Sci. 2018;26:57–62. https://dx.doi.org10.1016/j.cois.2018.01.006.
3. Adolfi A, Lycett GJ. Opening the toolkit for genetic analysis and control of *Anopheles* mosquito vectors. Curr Opin Insect Sci. 2018;30:8–18. https://dx.doi.org10.1016/j.cois.2018.07.014.
4. Guo Y, Goodman CL, Stanley DW, Bonning BC. Cell Lines for Honey Bee Virus Research. Viruses. 2020;12. https://dx.doi.org10.3390/v12020236.
5. Goblirsch MJ, Spivak MS, Kurtti TJ. A cell line resource derived from honey bee (*Apis mellifera*) embryonic tissues. PLoS One. 2013;8:e69831. https://dx.doi.org10.1371/journal.pone.0069831.
6. Carrillo-Tripp J, Dolezal AG, Goblirsch MJ, Miller WA, Toth AL, Bonning BC. *In vivo* and *in vitro* infection dynamics of honey bee viruses. Sci Rep. 2016;6:22265. https://dx.doi.org10.1038/srep22265.
7. Tentcheva D, Gauthier L, Zappulla N, Dainat B, Cousserans F, Colin ME et al. Prevalence and seasonal variations of six bee viruses in *Apis mellifera L.* and Varroa destructor mite populations in France. Appl Environ Microbiol. 2004;70:7185–91. https://dx.doi.org10.1128/aem.70.12.7185-7191.2004.
8. Yue C, Schröder M, Gisder S, Genersch E. Vertical-transmission routes for Deformed wing virus of honeybees (*Apis mellifera*). J Gen Virol. 2007;88:2329–36. https://dx.doi.org10.1099/vir.0.83101-0.
9. Yue C, Genersch E. RT-PCR analysis of Deformed wing virus in honeybees (*Apis mellifera*) and mites (*Varroa destructor*). J Gen Virol. 2005;86:3419–24. https://dx.doi.org10.1099/vir.0.81401-0.
10. Woodford L, Evans DJ. Deformed wing virus: using reverse genetics to tackle unanswered questions about the most important viral pathogen of honey bees. FEMS Microbiol Rev. 2021;45. https://dx.doi.org10.1093/femsre/fuaa070.
11. Goblirsch M, Adamczyk JJ. Using the honey bee (*Apis mellifera L.*) cell line AmE-711 to evaluate insecticide toxicity. Environ Toxicol Chem. 2023;42:88–99. https://doi.org/10.1002/etc.5500.
12. Kumar D, Goblirsch M, Adamczyk J, Karim S. Acute bee paralysis virus regulation of microRNA expression dynamics in the insect host (*Apis mellifera*) cell line, AmE-711. Front Cell Infect Microbiol. 2025;15:1709271. https://dx.doi.org10.3389/fcimb.2025.1709271.
13. Dave A, Charytonowicz D, Francoeur NJ, Beaumont M, Beaumont K, Schmidt H et al. The Breast Cancer Single-Cell Atlas: Defining cellular heterogeneity within model cell lines and primary tumors to inform disease subtype, stemness, and treatment options. Cell Oncol (Dordr). 2023;46:603–28. https://dx.doi.org10.1007/s13402-022-00765-7.
14. Osorio D, Yu X, Yu P, Serpedin E, Cai JJ. Single-cell RNA sequencing of a European and an African lymphoblastoid cell line. Sci Data. 2019;6:112. https://dx.doi.org10.1038/s41597-019-0116-4.
15. Luhur A, Mariyappa D, Bohall P, Multini L, Elkins M, Zelhof AC. Establishment of Drosophila intestinal cell lines as tools for multiomic screening and deciphering intestinal biology. Sci Rep. 2025;15:32291. https://dx.doi.org10.1038/s41598-025-17336-z.

16. Arya SK, Harrison DA, Palli SR. Deciphering cellular heterogeneity in *Spodoptera frugiperda* midgut cell line through single cell RNA sequencing. Genomics. 2024;116:110898. https://doi.org/10.1016/j.ygeno.2024.110898.
17. Traniello IM, Bukhari SA, Kevill J, Ahmed AC, Hamilton AR, Naeger NL, et al. Meta-analysis of honey bee neurogenomic response links Deformed wing virus type A to precocious behavioral maturation. Sci Rep. 2020. https://doi.org/10.1038/s41598-020-59808-4.
18. Patir A, Raper A, Fleming R, Henderson BEP, Murphy L, Henderson NC et al. Cellular heterogeneity of the developing worker honey bee (*Apis mellifera*) pupa: a single cell transcriptomics analysis. G3 Genes|Genomes|Genetics. 2023;13:jkad178. https://doi.org/10.1093/g3journal/jkad178
19. Becchimanzi A, De Leva G, Mattossovich R, Camerini S, Casella M, Jesu G, et al. Deformed wing virus coopts the host arginine kinase to enhance its fitness in honey bees (*Apis mellifera*). BMC Biol. 2025;23:12. https://doi.org/10.1186/s12915-025-02117-x.
20. Karatza C, Stein WD, Shall S. Kinetics of *in vitro* ageing of mouse embryo fibroblasts. J Cell Sci. 1984;65:163–75. https://dx.doi.org10.1242/jcs.65.1.163.
21. Tamm I, Kikuchi T, Wang E, Pfeffer LM. Growth rate of control and beta-interferon-treated human fibroblast populations over the course of their in vitro life span. Cancer Res. 1984;44(6):2291–6. PMID: 6722770.
22. Wang J, Su W, Nie W, Wang J, Xiao W, Wang D. Establishment and characterization of fibroblast cell lines from the skin of the Yangtze finless porpoise. In Vitro Cell Dev Biol Anim. 2011;47:618–30. https://dx.doi.org10.1007/s11626-011-9448-x.
23. Amand MMS, Hanover JA, Shiloach J. A comparison of strategies for immortalizing mouse embryonic fibroblasts. J Biol Methods. 2016;3:e41. https://dx.doi.org10.14440/jbm.2016.110.
24. Jones JC, Helliwell P, Beekman M, Maleszka R, Oldroyd BP. The effects of rearing temperature on developmental stability and learning and memory in the honey bee, *Apis mellifera*. J Comp Physiol A Neuroethol Sens Neural Behav Physiol. 2005;191:1121–9. https://dx.doi.org10.1007/s00359-005-0035-z.
25. Medina RG, Paxton RJ, Hernández-Sotomayor SMT, Pech-Jiménez C, Medina-Medina LA, Quezada-Euán JJG. Heat stress during development affects immunocompetence in workers, queens and drones of Africanized honey bees (*Apis mellifera L.*) (Hymenoptera: Apidae). J Therm Biol. 2020;89:102541. https://dx.doi.org10.1016/j.jtherbio.2020.102541.
26. Tautz J, Maier S, Groh C, Rossler W, Brockmann A. Behavioral performance in adult honey bees is influenced by the temperature experienced during their pupal development. Proc Natl Acad Sci U S A. 2003;100:7343–7. https://dx.doi.org10.1073/pnas.1232346100.
27. Finehout EJ, Cantor JR, Lee KH. Kinetic characterization of sequencing grade modified trypsin. Proteomics. 2005;5:2319–21. https://dx.doi.org10.1002/pmic.200401268.
28. Muhl L, Genové G, Leptidis S, Liu J, He L, Mocci G et al. Single-cell analysis uncovers fibroblast heterogeneity and criteria for fibroblast and mural cell identification and discrimination. Nat Commun. 2020;11:3953. https://dx.doi.org10.1038/s41467-020-17740-1.

29. Meyer S, Schmidt I, Klämbt C. Glia ECM interactions are required to shape the *Drosophila* nervous system. Mech Dev. 2014;133:105–16. https://dx.doi.org10.1016/j.mod.2014.05.003.
30. McInnes L, Healy J, Melville JUMAP. Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv [stat.ML]. 2018. http://arxiv.org/abs/1802.03426
31. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods. 2019;16:1289–96. https://dx.doi.org10.1038/s41592-019-0619-0.
32. Kaneko K, Hori S, Morimoto MM, Nakaoka T, Paul RK, Fujiyuki T et al. *In situ* hybridization analysis of the expression of *futsch*, *tau*, and *MESK2* homologues in the brain of the European honeybee (*Apis mellifera L.*). PLoS One. 2010;5:e9213. https://dx.doi.org10.1371/journal.pone.0009213.
33. Takeuchi H. Gene expression in the honeybee mushroom body and its gene orthologues. Evolution of Nervous Systems. Elsevier; 2007. pp. 457–69. https://dx.doi.org10.1016/B0-12-370878-8/00188-9.
34. Paul RK, Takeuchi H, Kubo T. Expression of two ecdysteroid-regulated genes, *Broad-Complex* and *E75*, in the brain and ovary of the honeybee (*Apis mellifera L.*). Zoolog Sci. 2006;23:1085–92. https://dx.doi.org10.2108/zsj.23.1085.
35. Duer A, Paffhausen BH, Menzel R. High order neural correlates of social behavior in the honeybee brain. J Neurosci Methods. 2015;254:1–9. https://dx.doi.org10.1016/j.jneumeth.2015.07.004.
36. Kamata S, Kubo T, Kohno H. Analysis of molecular and cellular bases of honey bee mushroom body development. Sci Rep. 2025;15:21462. https://doi.org/10.1038/s41598-025-06268-3.
37. Kaneko K, Suenami S, Kubo T. Gene expression profiles and neural activities of Kenyon cell subtypes in the honeybee brain: identification of novel 'middle-type' Kenyon cells. Zoological Lett. 2016;2:14. https://doi.org/10.1186/s40851-016-0051-6.
38. Suenami S, Oya S, Kohno H, Kubo T. Kenyon Cell subtypes/populations in the honeybee mushroom bodies: Possible function based on their gene expression profiles, differentiation, possible evolution, and application of genome editing. Front Psychol. 2018;9. https://doi.org/10.3389/fpsyg.2018.01717.
39. Zhang W, Wang L, Zhao Y, Wang Y, Chen C, Hu Y et al. Single-cell transcriptomic analysis of honeybee brains identifies vitellogenin as caste differentiation-related factor. iScience. 2022;25:104643. https://dx.doi.org10.1016/j.isci.2022.104643.
40. Mu X, Zhang Z, Liu Q, Ma J, Qin Y, Lang H et al. Single-nucleus and spatial transcriptomics identify brain landscape of gene regulatory networks associated with behavioral maturation in honeybees. Nat Commun. 2025;16:3343. https://dx.doi.org10.1038/s41467-025-58614-8.
41. Hultmark D, Andó I. Hematopoietic plasticity mapped in *Drosophila* and other insects. Elife. 2022;11:e78906. https://dx.doi.org10.7554/eLife.78906.
42. Qadiri M, Liu Y, Kim A-R, Han M, Zhou E, Veal A et al. FlyPhoneDB2: A computational framework for analyzing cell-cell communication in *Drosophila* scRNA-seq data integrating AlphaFold-multimer

predictions. Comput Struct Biotechnol J. 2025;27:2814–22. https://dx.doi.org10.1016/j.csbj.2025.06.032.

43. Broadie K, Baumgartner S, Prokop A. Extracellular matrix and its receptors in *Drosophila* neural development. Dev Neurobiol. 2011;71:1102–30. https://dx.doi.org10.1002/dneu.20935.
44. Isabella AJ, Horne-Badovinac S. Building from the ground up: Basement membranes in Drosophila development. Curr Top Membr. 2015;76:305–36. https://dx.doi.org10.1016/bs.ctm.2015.07.001.
45. Kelemen-Valkony I, Kiss M, Csiha J, Kiss A, Bircher U, Szidonya J et al. *Drosophila* basement membrane collagen col4a1 mutations cause severe myopathy. Matrix Biol. 2012;31:29–37. https://dx.doi.org10.1016/j.matbio.2011.09.004.
46. Ramos-Lewis W, Page-McCaw A. Basement membrane mechanics shape development: Lessons from the fly. Matrix Biol. 2019;75–76:72–81. https://dx.doi.org10.1016/j.matbio.2018.04.004.
47. Finak G, McDavid A, Yajima M, Deng J, Gersuk V, Shalek AK et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. Genome Biol. 2015;16:278. https://dx.doi.org10.1186/s13059-015-0844-5.
48. Cho B, Yoon S-H, Lee D, Koranteng F, Tattikota SG, Cha N et al. Single-cell transcriptome maps of myeloid blood cell lineages in *Drosophila*. Nat Commun. 2020;11:4483. https://dx.doi.org10.1038/s41467-020-18135-y.
49. Tattikota SG, Cho B, Liu Y, Hu Y, Barrera V, Steinbaugh MJ, et al. A single-cell survey of *Drosophila* blood. Elife. 2020;9:e54818. https://doi.org/10.7554/eLife.54818.
50. Lendahl U, Muhl L, Betsholtz C. Identification, discrimination and heterogeneity of fibroblasts. Nat Commun. 2022;13:3409. https://dx.doi.org10.1038/s41467-022-30633-9.
51. Buechler MB, Pradhan RN, Krishnamurty AT, Cox C, Calviello AK, Wang AW et al. Cross-tissue organization of the fibroblast lineage. Nature. 2021;593:575–9. https://dx.doi.org10.1038/s41586-021-03549-5.
52. Lanzi G, de Miranda JR, Boniotti MB, Cameron CE, Lavazza A, Capucci L et al. Molecular and biological characterization of Deformed wing virus of honeybees (*Apis mellifera L.*). J Virol. 2006;80:4998–5009. https://dx.doi.org10.1128/JVI.80.10.4998-5009.2006.
53. Gusachenko ON, Woodford L, Balbirnie-Cumming K, Campbell EM, Christie CR, Bowman AS et al. Green bees: Reverse genetic analysis of Deformed wing virus transmission, replication, and tropism. Viruses. 2020;12:532. https://dx.doi.org10.3390/v12050532.
54. Ryabov EV, Christmon K, Heerman MC, Posada-Florez F, Harrison RL, Chen Y et al. Development of a honey bee RNA virus vector based on the genome of a Deformed wing virus. Viruses. 2020;12:374. https://dx.doi.org10.3390/v12040374.
55. Wu Y, Yuan X, Li J, Kadowaki T. DWV infection *in vitro* using honey bee pupal tissue. Front Microbiol. 2021;12:631889. https://dx.doi.org10.3389/fmicb.2021.631889.
56. Norton AM, Buchmann G, Ashe A, Watson OT, Beekman M, Remnant EJ. Deformed wing virus genotypes A and B do not elicit immunologically different responses in naïve honey bee hosts. Insect Mol Biol. 2025;34:33–51. https://dx.doi.org10.1111/imb.12948.

57. Pizzorno MC, Field K, Kobokovich AL, Martin PL, Gupta RA, Mammone R et al. Transcriptomic responses of the honey bee brain to infection with Deformed wing virus. Viruses. 2021;13. https://dx.doi.org10.3390/v13020287.
58. Silva D, Traverso L, Dominguez G, Becerra F, Fuentes Y, Ons S et al. Transcriptomic analysis of genes related to neuronal functions in *Apis mellifera*: expression dynamics during the progression of DWV-A infection. Research Square. Research Square; 2026. https://doi.org/10.21203/rs.3.rs-9087825/v1
59. Loughran SE, Dingle L, Bowman AS, Manfredini F. Neurotranscriptomic profiling of Deformed wing virus-infected honey bee foragers (*Apis mellifera*) with different cognitive abilities. Biol Open. 2025;14(11):bio062204. https://doi.org/10.1242/bio.062204.
60. Chang ZT, Huang YF, Chen YW, Yen MR, Hsu PY, Chen TH et al. Transcriptome-level assessment of the impact of Deformed wing virus on honey bee larvae. Sci Rep. 2021;11:15028. https://dx.doi.org10.1038/s41598-021-94641-3.
61. Lee JI, Park S, Park H, Lee Y, Park J, Lee D, et al. The matrix glycoprotein *Papilin* maintains the haematopoietic progenitor pool in *Drosophila* lymph glands. Development. 2025;152:dev204367. https://dx.doi.org/10.1242/dev.204367.
62. Kramerova IA, Kawaguchi N, Fessler LI, Nelson RE, Chen Y, Kramerov AA et al. Papilin in development; a pericellular protein with a homology to the ADAMTS metalloproteinases. Development. 2000;127:5475–85. https://dx.doi.org10.1242/dev.127.24.5475.
63. Fessler JH, Kramerova I, Kramerov A, Chen Y, Fessler LI. Papilin, a novel component of basement membranes, in relation to ADAMTS metalloproteases and ECM development. Int J Biochem Cell Biol. 2004;36:1079–84. https://dx.doi.org10.1016/j.biocel.2003.12.010.
64. Tong M, Jun T, Nie Y, Hao J, Fan D. The role of the Slit/Robo signaling pathway. J Cancer. 2019;10:2694–705. https://dx.doi.org10.7150/jca.31877.
65. Fang Y, Feng M, Han B, Lu X, Ramadan H, Li J. In-depth proteomics characterization of embryogenesis of the honey bee worker (*Apis mellifera ligustica*). Mol Cell Proteomics. 2014;13:2306–20. https://dx.doi.org10.1074/mcp.M114.037846.
66. De Conto F, Mancabelli L, Prandini E, Ventura M. Highly dynamic cytoskeletal networks support productive viral infection and host innate immune response activation. Curr Clin Microbiol Rep. 2025;12:5. https://dx.doi.org10.1007/s40588-025-00241-4.
67. Ullrich K, CRBHits. From conditional reciprocal best hits to Codon alignments and Ka/ks in R. J Open Source Softw. 2020;5:2424. https://doi.org/10.21105/joss.02424.
68. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421. https://dx.doi.org10.1186/1471-2105-10-421.
69. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10. https://dx.doi.org10.1016/S0022-2836(05)80360-2.
70. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R et al. Massively parallel digital transcriptional profiling of single cells. Nat Commun. 2017;8:14049. https://dx.doi.org10.1038/ncomms14049.

71. Hao Y, Stuart T, Kowalski MH, Choudhary S, Hoffman P, Hartman A et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nat Biotechnol. 2024;42:293–304. https://dx.doi.org10.1038/s41587-023-01767-y.
72. McCarthy DJ, Campbell KR, Lun AT, Wills QF. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data. R Bioinf. 2017;33:1179–86. https://academic.oup.com/bioinformatics/article-abstract/33/8/1179/2907823.
73. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16:284–7. https://dx.doi.org10.1089/omi.2011.0118.
74. Zappia L, Oshlack A. Clustering trees: a visualization for evaluating clusterings at multiple resolutions. Gigascience. 2018;7. https://dx.doi.org10.1093/gigascience/giy083.
75. Kirchner R, Barrera V, Tinyatlas. GitHub Repository. https://github.com/hbc/tinyatlas/tree/master/cell_cycle
76. Li H, Janssens J, De Waegeneer M, Kolluru SS, Davie K, Gardeux V et al. Fly Cell Atlas: A single-nucleus transcriptomic atlas of the adult fruit fly. Science. 2022;375:eabk2432. https://dx.doi.org10.1126/science.abk2432.
77. Hu Y, Tattikota SG, Liu Y, Comjean A, Gao Y, Forman C et al. DRscDB: A single-cell RNA-seq resource for data mining and data comparison across species. Comput Struct Biotechnol J. 2021;19:2018–26. https://dx.doi.org10.1016/j.csbj.2021.04.021.
78. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32:3047–8. https://dx.doi.org10.1093/bioinformatics/btw354.
79. Ahlmann-Eltze C, Huber W. glmGamPoi: fitting Gamma-Poisson generalized linear models on single cell count data. Bioinformatics. 2021;36:5701–2. https://dx.doi.org/10.1093/bioinformatics/btaa1009.
80. Choudhary S, Satija R. Comparison and evaluation of statistical error models for scRNA-seq. Genome Biol. 2022;23:27. https://dx.doi.org10.1186/s13059-021-02584-9.
81. Luecken MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. Mol Syst Biol. 2019;15:e8746. https://doi.org/10.15252/msb.20188746.
82. Hrovatin K, Sikkema L, Shitov VA, Heimberg G, Shulman M, Oliver AJ et al. Considerations for building and using integrated single-cell atlases. Nat Methods. 2025;22:41–57. https://dx.doi.org10.1038/s41592-024-02532-y.
83. Attrill H, Antonazzo G, Goodman JL, Thurmond J, Strelets VB, Brown NH, et al. A new experimental evidence-weighted signaling pathway resource in FlyBase. Development. 2024;151:dev202255. https://dx.doi.org/10.1242/dev.202255.
84. Liu Y, Li JSS, Rodiger J, Comjean A, Attrill H, et al. FlyPhoneDB: An integrated web-based resource for cell-cell communication prediction in *Drosophila*. Genetics. 2022;220(3):iyab235. https://doi.org/10.1093/genetics/iyab235.

85. Shao X, Liao J, Li C, Lu X, Cheng J, Fan X. CellTalkDB: a manually curated database of ligand-receptor interactions in humans and mice. Brief Bioinform. 2021;22:bbaa269. https://dx.doi.org/10.1093/bib/bbaa269.

Tables

Tables 1 to 3 are available in the Supplementary Files section.

Figures

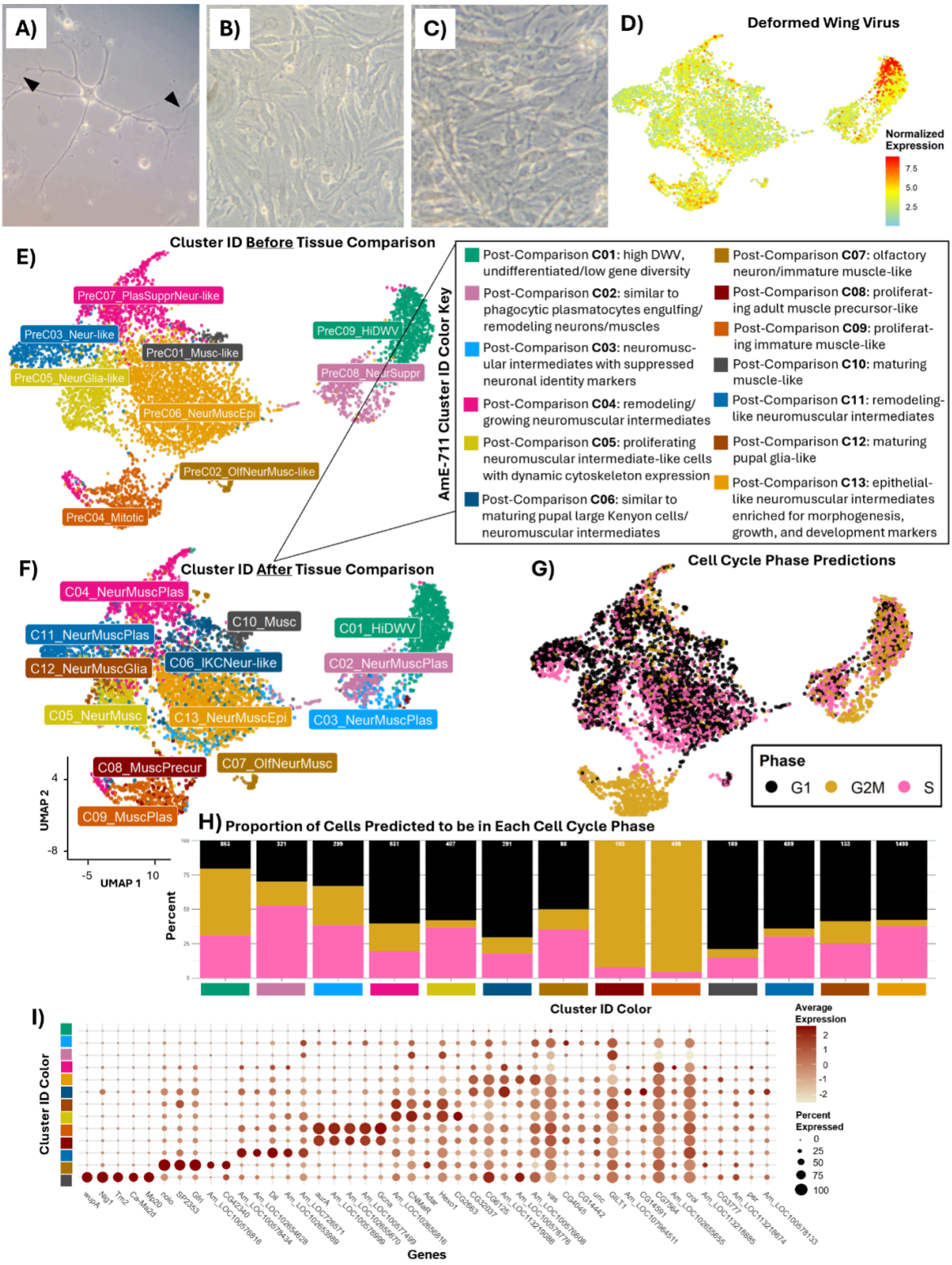


Figure 1

AmE-711 cell line morphological and transcriptomic characteristics.

We isolated 6,930 cells of the honey bee embryo-derived cell line, AmE-711, using single-cell RNA sequencing to characterize their heterogeneous transcriptomic state and persistent infection with deformed wing virus (DWV). A) The cells display cytoplasmic extensions (black triangles) and strong

adherence both to their substrate and to one another but lack contact inhibition. D) All cells are infected with DWV, a prevalent pathogen of honey bees, but RNA abundances varied in our sample. We annotated E) 9 clusters with identities that were expanded to F) include 4 additional subcluster identities after a comparison of cell line single-cell gene expression profiles to previously published adult and pupal tissue-derived single-cell gene expression profiles. G) Just over 20% of cells were mitotic based on expression of orthologous canonical cell cycle markers from *D. melanogaster*, including H) 2 out of 13 clusters that distinctly lacked cells scored to be in the G1 phase. Cell-line cluster markers showed differences in their expression when comparing cluster identity annotations and I) after comparing cell line gene expression to tissue gene expression. Gene symbols for orthologous *D. melanogaster* genes lack “Am_” preceding them.

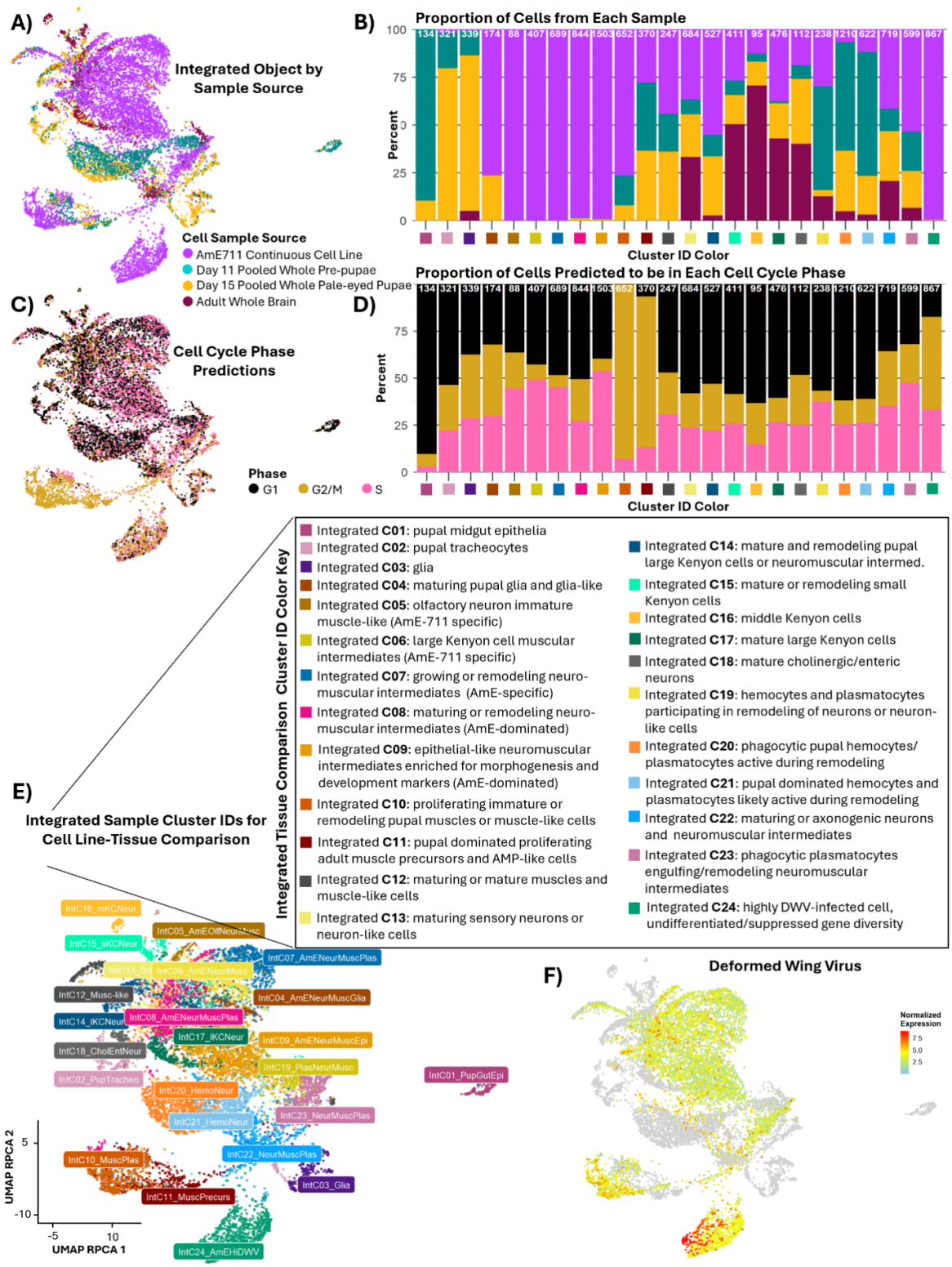


Figure 2

Integrated Seurat object of AmE-711 and published honey bee datasets.

Integration of three previously published honey bee single-cell data sets with our AmE-711 data set captured the transcriptomic profiles of 12,328 cells. Successful integration showed adequate mixing of some of the 24 cell clusters that contained pupal and adult brain cells and some mixing between with

these and cell line isolated cells, but the majority of AmE-711 cells only clustered with one another. A–B) AmE-711 cells made up about half of all cells in the integrated Seurat object. C–D) About 25% of cells in the integrated object were scored as mitotic (G2/M or S Phase), largely excluding cells isolated from adult brains, in line with previous work that has noted an absence of neurogenesis in the adult honey bee brain. E) Our analysis of cluster marker gene sets and well known honey bee cell type specific marker genes allowed us to annotate 24 clusters. F) Only AmE-711 cells expressed *DWV polyprotein*.

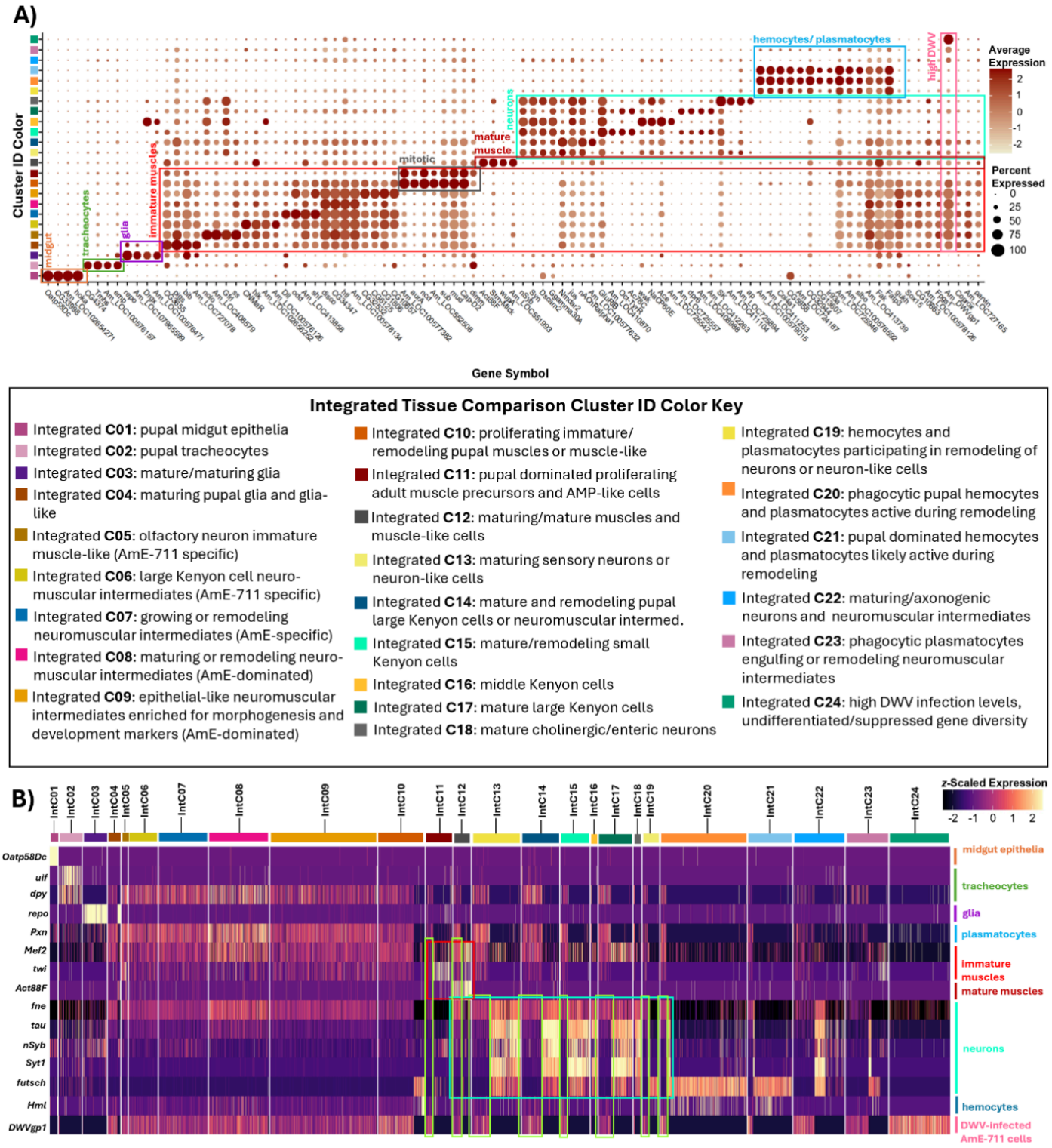


Figure 3

Cell line and tissue-specific marker genes.

The integrated dataset allowed us to identify AmE-711 and tissue specific marker genes in addition to examining expression of canonical cell type markers. Our analysis confirmed that most cell-line isolated cells have unique transcriptomes, even for cells that cluster together with pupal-derived muscle cells and adult-derived neurons. A) Cluster markers for six cell types, mitotic cells, and highly DWV-infected AmE-711 cells showed average marker expression by AmE-711 clusters enriched among those identified as maturing muscles and some maturing or remodeling neurons. This suggests these cells have transcriptomes in intermediate neuromuscular states. B) However, *z*-scaled expression of log-normalized canonical markers for these cell types and *DWV polyprotein*RNA showed that for muscle- (highlighted in red) and neuron-annotated clusters (highlighted in aqua), infected AmE-711 cells express these genes in distinct patterns (highlighted in green). Marker expression was downregulated when compared to expression levels for known cell types isolated from organismal tissues. In both plots, gene symbols for orthologous *D. melanogaster*genes lack “Am_” preceding them.

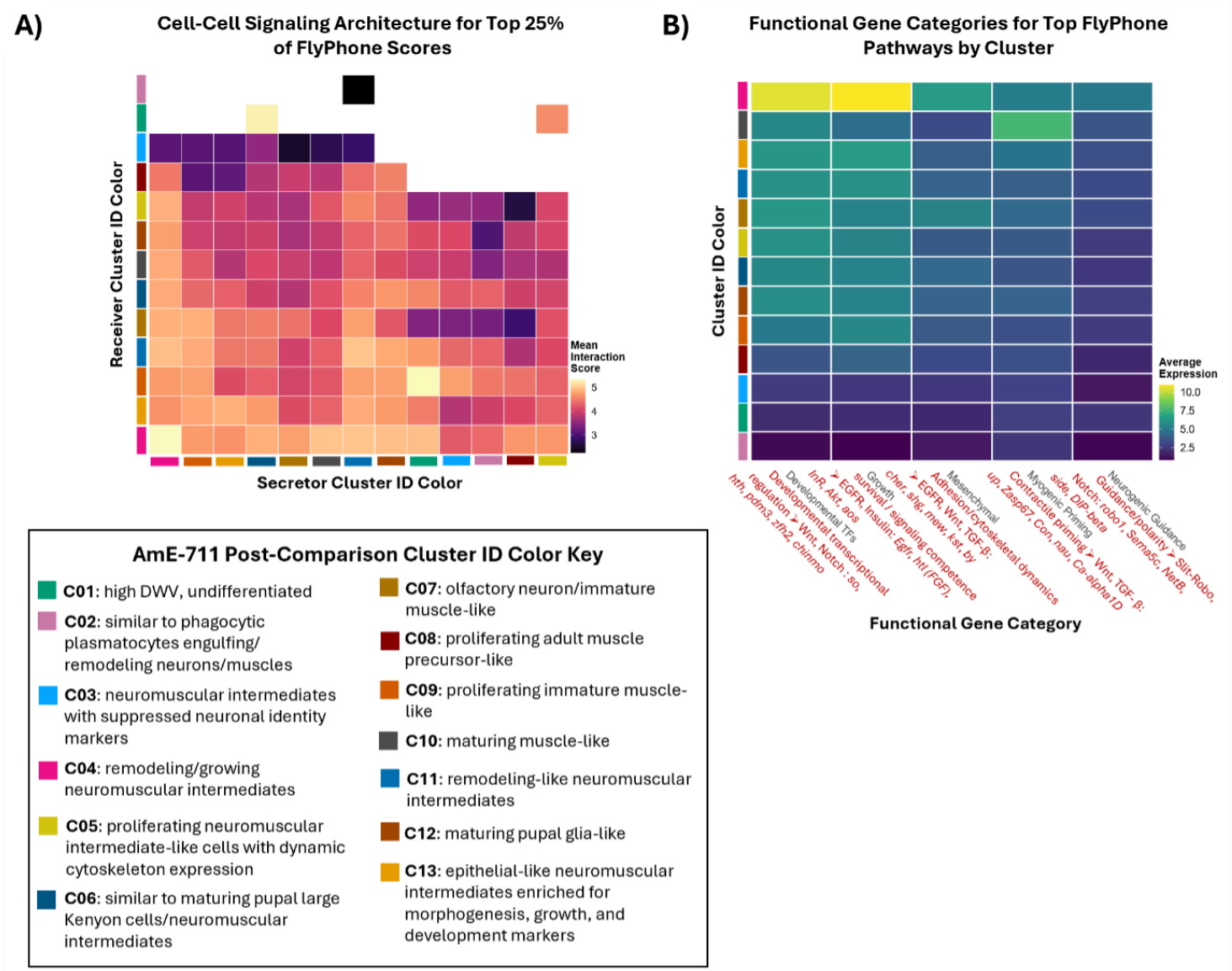


Figure 4

FlyPhone predicted cell-cell communication activity and pathways active in 13 AmE-711 clusters for the top quantile of FlyPhone ligand-receptor interaction scores.

The top 25% of FlyPhone scores suggest that signaling events between clusters are centered around five functional *D. melanogaster*gene categories primarily coordinated by cells in the third largest cluster. A) Based on the mean interaction scores in the third quantile, the third largest cluster (C04) is predicted to function as a signaling hub secreting ligands and binding those secreted by 11 out of 13 other clusters. B) Ligand-receptor pairs and reporter gene average expression was grouped across five functional gene categories to highlight nine clusters with strongest expression of transcription factors that regulate development, genes coordinating growth, and genes that promote mesenchymal-like cell identities coordinating adhesion and cytoskeletal dynamics. On the other hand, expression of genes involved in myogenic and neurogenic priming was lower in clusters with the highest *DWV polyprotein* RNA levels (C01–C03).

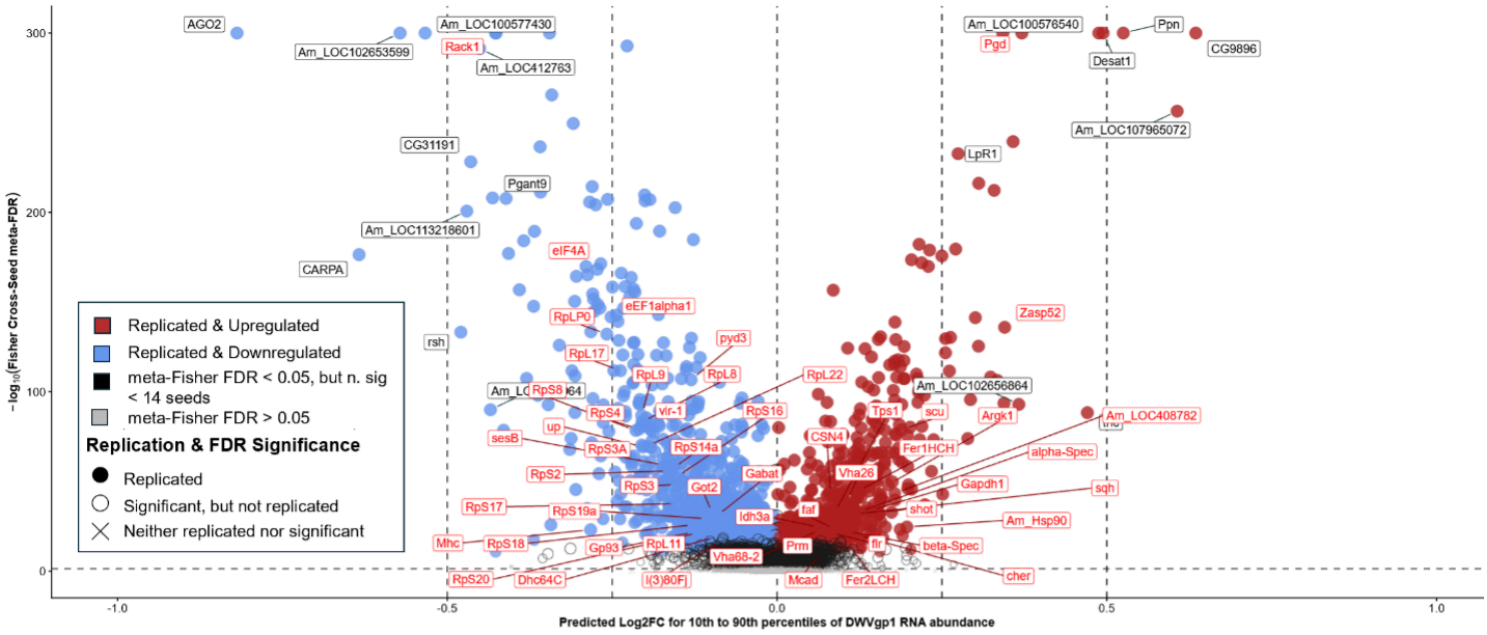


Figure 5

Volcano plot of model-predicted expression differences for Deformed wing virus (DWV) dose-sensitive genes comparing cells expressing the top and bottom 10% of DWV polyprotein RNA abundance levels.

A zero inflated linear regression model predicted that for 1,642 host cell genes, increasing viral RNA abundances could predict changes in expression even after controlling for technical artefacts or gene expression changes due to cluster membership. The genes highlighted in red encode 53/144 significantly enriched proteins co-immunoprecipitated with DWV in a previous study. Effect sizes are shown as model predicted expression differences for a 10th–90th percentile DWV viral load range. Gene symbols for orthologous protein-coding D. melanogaster genes lack “Am_” preceding them.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FloresetalAmE711BMCgenomicsSupplInfov6.docx
- FloresetalAmE711BMCgenomicsSupplData3FlyPhoneFINAL.xlsx
- FloresetalAmE711BMCgenomicsSupplData2LISIFINAL.xlsx
- FloresetalAmE711BMCgenomicsSupplData4MASTFINAL.xlsx
- FloresetalAmE711BMCgenomicsSupplData1HVGsMarkersFINAL.xlsx
- Tables.docx