Genomics 116 (2024) 110898

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Deciphering cellular heterogeneity in *Spodoptera frugiperda* midgut cell line through single cell RNA sequencing

Surjeet Kumar Arya [a], Douglas A. Harrison [b], Subba Reddy Palli [a,*]

[a] Department of Entomology, University of Kentucky, Lexington, KY 40546, USA
[b] College of Arts & Science Imaging Center & Department of Biology, University of Kentucky, Lexington, KY 40546, USA

ARTICLE INFO

Keywords:
Midgut
Single-cell sequencing
SfMG-0617
Spodoptera
Stem cells

ABSTRACT

Using the 10x Genomics Chromium single-cell RNA sequencing (scRNA-seq) platform, we discovered unexpected heterogeneity in an established cell line developed from the midgut of the Fall armyworm, *Spodoptera frugiperda,* a major global pest. We analyzed the sequences of 18,794 cells and identified ten unique cellular clusters, including stem cells, enteroblasts, enterocytes and enteroendocrine cells, characterized by the expression of specific marker genes. Additionally, these studies addressed an important knowledge gap by investigating the expression of genes coding for respiratory and midgut membrane insecticide targets classified by the Insecticide Resistance Action Committee. Dual-fluorescence tagging method, fluorescence microscopy and fluorescence-activated cell sorting confirmed the expression of midgut cell type-specific genes. Stem cells were isolated from the heterogeneous population of SfMG-0617 cells. Our results, validated by KEGG and Gene Ontology analyses and supported by Monocle 3.0, advance the fields of midgut cellular biology and establish standards for scRNA-seq studies in non-model organisms.

1. Introduction

The use of immortalized cell lines has exponentially increased over the past half-century, as these cells serve as indispensable models for many scientific inquiries. Using derivatives of diverse human organs and pathophysiological states, these cell lines facilitate the elucidation of organ-specific biological mechanisms and the pathogenesis of diseases. Moreover, they are instrumental in the biosynthesis of immunoglobulins and therapeutic proteins. For instance, an excess of 1270 lepidopteran cell lines have been characterized and harnessed for the mass production of bioinsecticides and recombinant proteins, as noted previously [1–3]. These cell lines are also employed in bioassays for insecticide discovery and development[4], as well as in investigations into virus-cell interactions and intracellular signal transduction pathways [4–7]. Single-cell RNA sequencing has revealed cellular heterogeneity within tissue-derived populations, revealing promising avenues for RNA interference screening and insecticide target identification. However, the cellular composition of insect-derived cell lines remains a subject of ongoing investigations.

Single-cell RNA sequencing (scRNA-seq) has recently emerged as a transformative technology capable of elucidating global transcriptomic profiles across thousands of isolated cells, thereby enabling the discovery of novel cellular subtypes, their respective physiological states, and their ontogenic origin [8]. The first cellular atlas of the adult *Drosophila melanogaster* midgut, reported by Hung et al. [9], included 22 discrete cell clusters representing intestinal stem cells (SC), enteroblasts (EB), enterocytes (EC), and enteroendocrine cells (EE), each with distinct gene expression signatures and marker genes.

The fall armyworm (FAW), *Spodoptera frugiperda*, is a polyphagous pest that has caused significant agricultural devastation in sub-Saharan Africa and South Asia since 2016 [10,11]. This species has remarkable host-plant diversity and feeds on 76 plant families, including key agricultural commodities [12]. Recently, the CT/BCIRL-SfMG1–0611-KZ (SfMG-0611) and CT/BCIRL-SfMG-0617-KZ (SfMG-0617) cell lines were established from the larval FAW midgut, augmenting the pool of resources available for global research endeavors [6].

To ascertain the cellular subtypes and corresponding marker genes within the SfMG-0617 midgut cell line, we employed the 10x Genomics-based scRNA-seq approach. This high-throughput analysis revealed ten distinct cell clusters within this established cell line. Subsequent gene expression profiling within these clusters identified specific marker genes. The promoters of these genes were used to drive the expression of

* Corresponding author.
E-mail address: rpalli@uky.edu (S.R. Palli).

https://doi.org/10.1016/j.ygeno.2024.110898
Received 2 May 2024; Received in revised form 24 June 2024; Accepted 16 July 2024
Available online 22 July 2024

the red fluorescent protein gene, facilitating the identification of cells belonging to each cluster. These results help in the separation of heterogeneous cell lines into homogeneous populations for use in insecticide discovery and production of recombinant proteins. This comprehensive analysis underscores the utility and inherent limitations of immortalized cell lines, further highlighting the invaluable role of scRNA-seq in advancing our understanding and potential improvement of these cellular resources.

2. Materials and methods

2.1. Cell culture maintenance and harvesting

The midgut cell line SfMG-0617 was used for scRNA-seq transcriptome analysis. These cells were maintained in TNM-FH (Sigma-Aldrich, USA) supplemented with 10% fetal bovine serum (FBS). The cells from the near-confluent flasks were harvested and plated. On day 3, when the cells reached ~75–90% confluency, they were detached from the surface, resuspended in TNM-FH medium, and centrifuged at 300 ×*g* for 5 min. The cell pellet was washed and resuspended in 0.5% BSA (Bovine serum albumin) containing PBS (Phosphate buffer saline). The cells were then checked for viability using trypan blue, passed through a 40 μm filter, and subjected to single-cell experiments.

2.2. Preparation of single-cell libraries using 10× genomics chromium technology

The cells were processed by the A&S Imaging Center at the University of Kentucky using microfluidic partitioning. Two independent samples were loaded into the 10x Genomics Chromium Controller, where cells were partitioned with barcoded GEMs into nanoliter emulsions. The cell loading was adjusted to target approximately 10,000 cells for each replicate. Using the 10x Genomics 3' Gene Expression Kit v3.1, uniquely barcoded partitioned cells were processed to produce and amplify cDNAs, from which libraries were prepared per the manufacturer's instructions (10x Genomics, Pleasanton, CA). The quality of the libraries was evaluated using an Agilent Bioanalyzer.

2.3. Sequencing and analysis

Libraries were sequenced using the Illumina NovaSeq 6000 platform. The gene expression libraries were sequenced to a target depth of >36-42,000 read pairs per cell. The resulting sequence read files (fastq) were analyzed using 10x Genomics Cell Ranger software (version 6.1.1 or later; 10x Genomics, Pleasanton, CA, USA). In brief, Cell Ranger was used for alignment, filtering, counting of barcodes and UMIs, cell calling, and production of an output containing cell-by-gene matrix counts. The raw data generated by Cell Ranger were subsequently imported into the R toolkit Seurat (v5.0.3) to compare the scRNA-seq datasets [13]. The baseline threshold was set to remove sequences from cells that fall below the registered threshold to ensure high-quality sequences. The sum of the counts (log-transformed) within each cell was used as a regressor for the linear regression model to control the cell number and sequencing depth. The FindIntegrationAnchors function was used to find anchors between the replicates. These anchors were used to integrate the two datasets using the IntegrateData function. Principal component (PC) analysis was used to reduce the dimensionality of the data and produce the top K principal components. These selected top PCs were clustered using the Louvain–Jaccard graph-based clustering approach. For visualization of the results, the t-SNE (t-distributed Stochastic Neighbor Embedding) and UMAP (Uniform Manifold Approximation and Projection) algorithms were used [14,15].

2.3.1. Data preprocessing and quality control

After obtaining single-cell RNA sequencing (scRNA-seq) datasets, the Seurat package (v5.0.3) was used to import matrix files produced by Cell Ranger. The data from two biological replicates were processed. Each replicate was encapsulated in a Seurat object and subjected to normalization. The normalized data from both replicates were then integrated into a single Seurat object for comprehensive analysis. We incorporated a sample identification column into the metadata and further partitioned it into 'Sampletype' and 'Barcode'. We also read a predefined list of mitochondrial genes for downstream filtering. Quality control measures were implemented to filter out poor-quality sequences from cells based on criteria such as the number of detected genes, the total number of UMIs, and the percentage of mitochondrial reads (**Table S1**). The resulting filtered Seurat objects retained cells that met stringent quality control criteria.

2.3.2. Feature selection and dimensionality reduction

Variable genes were identified using the variance-stabilizing transformation (VST) method, during which 2000 features were selected. The dataset was scaled, and principal component analysis (PCA) was performed. The significance of each principal component was determined using the JackStraw method, which provides information on the dimensions used for downstream analysis (**Supplementary file 1: Fig. S1**). Cells were clustered using the top 20 principal components with a resolution setting of 0.5. This process resulted in a multidimensional representation of cells in the UMAP (uniform manifold approximation and projection) space. The distribution of cells across clusters was visualized using various plotting functions. The DoubletFinder package (v2.0.4) was used to identify potential doublets in the data. Parameters such as pN, pK, and nExp were carefully selected based on the characteristics of the dataset [16]. After filtering, we retained only those cells classified as "Singlets" for downstream analyses (**Table S1**).

2.4. Testing reproducibility

In the context of single-cell datasets, which provide high-resolution insights into cellular heterogeneity and gene expression patterns, testing reproducibility is crucial to validate the accuracy and consistency of the data analysis pipelines used. One popular program used for testing reproducibility in single-cell datasets is Harmony (v1.2.0) [17]. We utilized the RunHarmony function to synchronize the datasets, ensuring that any systematic discrepancies across batches were minimized. The integrated data were subsequently updated with batch-corrected values. The batch-corrected datasets were visualized using UMAP and a feature plot. The analysis included comparing the raw and filtered data and downstream analysis using Seurat, and we observed high reproducibility in the identified midgut cell types in two replicates. Following batch effect correction, we performed SCT transformation on the integrated dataset [18]. SCT transformation is a standard preprocessing step in single-cell RNA-seq analysis, which accounts for technical noise and normalizes gene expression levels across cells. We have established a framework for reproducible analysis of cellular heterogeneity and gene expression patterns.

2.5. Analysis of gene expression signatures of midgut cell types

Following cluster identification, marker genes were determined for each cluster by assessing gene expression levels. These markers were utilized as features to predict cluster membership for individual cells. Subsequently, fold changes in gene expression across clusters were computed. To assign cell types, gene expression profiles within each cluster were compared to known midgut cell types in *D. melanogaster* and *Aedes ageypti* [9,19]. KOBAS (KEGG Orthology-Based Annotation System) 3.0 was used to obtain the GO (Gene Ontology) terms for the identified genes. The obtained GO terms were further analyzed using clusterProfiler, an R package, to reduce redundancy, prioritize the enriched/statistically significant terms, and display only their representatives to ease data interpretation [20,21]. Kyoto Encyclopedia of Genes and Genomes (KEGG) and GO enrichment analyses were

conducted to identify prevalent pathways under various cellular conditions. Enrichment scores were calculated based on gene proportions, and significant terms were retained using adjusted *p*-values. Key pathways were selected for visualization, and a heatmap was generated using the Seaborn library to illustrate variance in enrichment scores across conditions, providing a comprehensive view of KEGG and GO enrichment outcomes.

2.6. Identification of ISC, EC-like, and EE cell type marker gene promoters

To identify the promoter regions of the midgut cell type markers, we selected intergenic sequences upstream of the transcription start site of each marker gene. The promoter regions were amplified using gene-specific primers, genomic DNA, and PrimeSTAR GXL DNA Polymerase (TaKaRa, Japan). The promoter regions were subsequently cloned and inserted into the pBac: hr5ie1-EGFP-SV40:tdTomato-SV40 vector to produce the pBac:hr5ie1-EGFP-SV40:promoter-tdTomato-SV40 vector. The promoter region and PuroRed-SV40 cassette were cloned and inserted into the pBac: hr5ie1-EGFP-SV40 vector to generate the pBac: hr5ie1-EGFP-SV40:promoter-PuroRed-SV40 vector. All the vectors were constructed using Gibson assembly master mix (NEB, USA). The maps and primer sequences of the constructs are shown in **Supplementary File 1**.

2.7. Selection of SfMG-0617 cells expressing endogenous promoter constructs

SfMG-0617 cells were seeded in 6-well culture plates at a density of 3.0×10^6 cells per well and incubated at 27 °C overnight. The cells were transfected using 1 μg of promoter construct per well using Mirus-Transit Pro-Kit (mirus BIO) transfection reagent according to the manufacturer's instructions in 1000 μL of TNM-FH Insect medium. Subsequently, 1000 μL of TNM-FH Insect medium containing 10% FBS was added at four hours post-transfection. Fluorescence-activated cell sorting (FACS) was conducted at the Flow Cytometry and Immune Monitoring Core Facility, University of Kentucky, USA, employing a Sony sy3200 sorter. Sorted cells expressing Red fluorescence were resuspended in the conditioned medium supplemented with 20-hydroxyecdysone (20E) (20E, Catalog No: H5142, Sigma-Aldrich, St. Louis, MO) and collected in 5 mL falcon tubes. The sorted cells were cultured in 6-well plates to facilitate monitoring cell proliferation and differentiation. The selection of cells based on the puromycin resistance marker gene was also performed. A stock concentration of 10 mg/mL puromycin was prepared. Transfected cells were treated with puromycin at various concentrations (2 ng/μL, 5 ng/μL, 10 ng/μL, and 15 ng/μL), administered every third day. The surviving cells were monitored, and the selection continued until all surviving cells showed Red fluorescence.

3. Results

3.1. scRNA-seq of SfMG-0617 cell line

10× Genomics Chromium Technology was used to obtain sequences from 10,334 and 8460 SfMG-0617 cells in two replicates. >350 million sequence reads were obtained from each library, resulting in average read depths >36 and 42 thousand read pairs per cell, respectively, in two replicates. Nearly >15,000 protein-coding genes were represented in each replicate. Both samples exhibited similar sequencing and mapping metrics, as shown in Table 1. The mean reads per cell, valid barcodes, valid UMIs, and other metrics were nearly identical. These analyses confirm the high correlation and low variability between the two samples, supporting the robustness and reproducibility of our scRNA-seq data. The combination of sequencing data with sequencing depth, overall coverage, and the transcriptome mapping ratio demonstrated the robustness of our scRNA-seq experiments, providing a reliable basis for further analysis. The t-SNE projections of the cells were visualized based on UMI counts (**Supplementary File 1: Fig. S2**).

Table 1
Sequencing Parameters Comparison: Comparison of sequencing parameters between samples SfMG-0617 and SfMG-0617, detailing mean reads per cell, total reads, valid barcodes percentage, valid UMIs percentage, and mapping statistics across genomic regions.

Sequencing parameters	SfMG-0617-1	SfMG-0617 -2
Sequencing		
Mean Reads per cell	36,5590	42,326
Number of Reads	377,793,718	358,075,221
Valid Barcodes	97.7%	97.7%
Valid UMIs	100%	100%
Mapping		
Reads Mapped to Genome	90.7%	90.2%
Reads Mapped Confidently to Genome	66.3%	69.7%
Reads Mapped Confidently to Intergenic Regions	2.0%	2.0%
Reads Mapped Confidently to Intronic Regions	4.5%	4.8%
Reads Mapped Confidently to Exonic Regions	59.8%	62.9%
Reads Mapped Confidently to Transcriptome	62.4%	65.7%%
Reads Mapped Antisense to Gene	1.6%	1.7%
Cells		
Estimated Number of cells	10,334	8460
Median Genes per cell	2223	2513
Total genes detected	14,892	14,964

3.2. scRNA-seq analysis identified ten cell clusters in the SfMG-0617 midgut cell line

The integrated analysis identified ten cell clusters visualized with UMAP. The complete workflow profiles standardized for the SfMG-0617 single-cell RNA sequencing are shown in Fig. 1. These cell types, which included stem cells (SCs), enteroblast cells (EBs), enteroendocrine cells (EEs), enterocytes (ECs), and visceral muscles (VMs), were found to be identical between the two replicates. The marker genes that were highly expressed in each cluster were derived from the midgut cells of *D. melanogaster* [9,19,22–26]. Cluster #8, accounting for 7% of the cells in the SfMG-0617 cell line, was identified as SC based on the high expression levels of SC cell marker genes, the neurogenic locus Notch (gene-LOC118262320), the mushroom body large-type Kenyon cell-specific protein-encoding gene (gene-LOC118270242) and the headcase gene (gene-LOC118265494) (Fig. 2 **and Supplementary File 1: Fig. S3**). Cluster #1, accounting for 16% of the cells, contained EBs based on the high expression levels of the marker genes, the cdc42 homolog (gene-LOC118267256), and the CCHC-type zinc finger nucleic acid binding protein isoform X2 (gene-LOC118265866) (Fig. 2 **and Supplementary File 1: Fig. S3**). The other clusters (#0, 4, 7 and 9) sharing UMI coordinates with each other were identified as EC-like cells based on the expression of the chitinase-like protein EN03 isoform X4 (gene-LOC118276216), epidermal growth factor receptor kinase protein (gene-LOC126911560), the sugar transporter SWEET1 (gene-LOC118267186), and aldehyde dehydrogenase (gene-LOC118268490) genes. The cells in clusters #2 and #6 were identified as EE 1 and EE-like 1 based on the high expression of the homeobox protein prospero (gene-LOC118275652), tachykinins (gene-LOC118278153), dipeptidase 1 (gene-LOC118268421) and EE-like 1 because of the high expression of the protein turtle (gene-LOC118264893), protein eiger (gene-LOC118271688), and protein lethal (2) genes (gene-LOC118280771) (Fig. 2 **and Supplementary File 1: Fig. S3**). Clusters #3 and #5, accounting for 10% and 8%, respectively, of the cells were designated visceral muscle 1 and 2 based on the high expression of the marker genes kinesin-like protein (gene-LOC118278067), thymosin beta (gene-LOC118276119), and protein Skeletor (gene-LOC118270244) (Fig. 2 **and Supplementary File 1: Fig. S3)**.

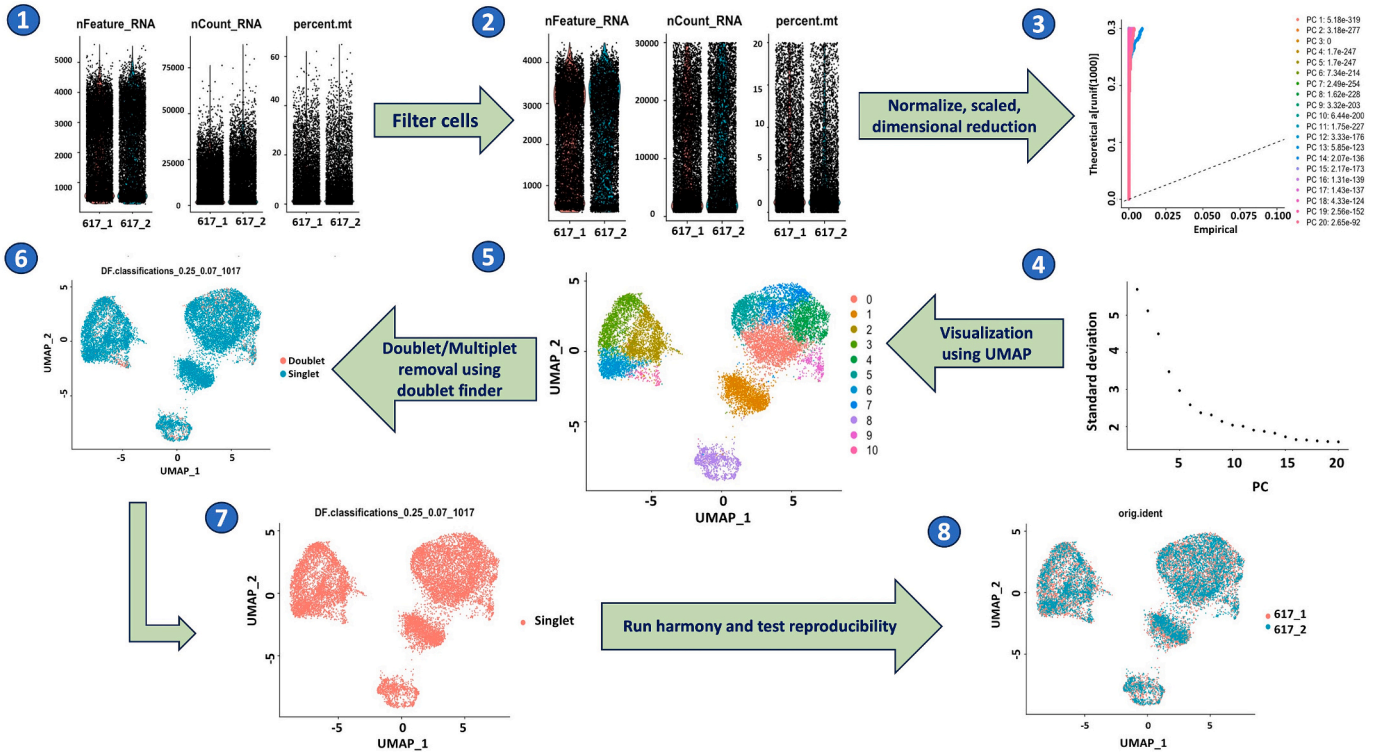

Fig. 1. Schematic representation of the standardized pipelines used for single-cell data analysis of the SfMG-0617 cells. 1) Initial single-cell profile of midgut cells, displaying nFeature_RNA, nCount_RNA, and percent.mt metrics before applying Seurat filtration. **2)** Cell profile post-filtration was achieved using the Seurat "subset" function. **3)** Subsequent normalization, scaling, and dimensionality reduction were performed on the filtered datasets, visualized via a jackstraw plot. **4)** Correlations between principal components and SDs were examined using an elbow plot. **5)** The identification and clustering of neighboring cells were visualized through a UMAP plot. **6)** Clusters were then filtered to remove doublets or multiplets using the DoubletFinder function. **7&8)** The remaining singlet cells were subjected to batch correction and reproducibility assessment using the RunHarmony function.

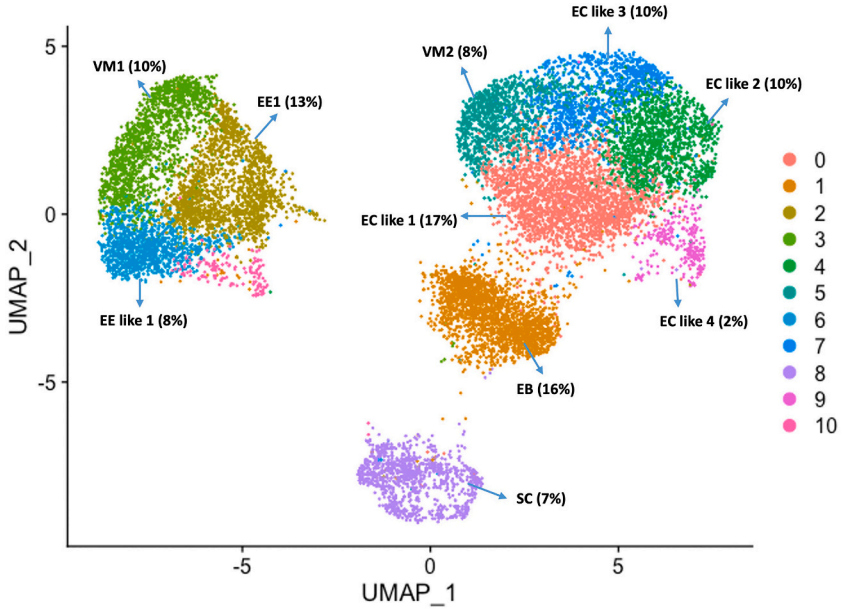

Fig. 2. Visualization of cell types in clusters using UMAP. We identified 10 distinct clusters, each expressing marker genes indicative of specific midgut cell types. These marker genes were identified based on the results from other insects, including *Drosophila melanogaster* and *Aedes aegypti*. EC, enterocytes; EE, enteroendocrine cells; EB, enteroblasts; SCs, intestinal stem cells; and VM, visceral muscles. The percentage of cells in each cluster is indicated in parentheses.

3.3. Novel marker identification using the FindAllMarker function

To delineate markers indicative of specific cellular clusters, we employed the 'FindAllMarkers' algorithm, which employs a statistical methodology grounded in a negative binomial distribution. This test mandated a minimum fold enrichment of 0.5 and a Bonferroni-adjusted p-value <0.05 to designate a gene as statistically significant. Due to the marginal variance in fold enrichment across divergent cellular clusters,

additional stratification was conducted based on the proportion of cells expressing a given gene within an individual cluster (pc.t1) versus across all analogous clusters (pc.t2). This augmented methodology facilitated the identification of marker genes with robust specificity for each cluster, thereby enhancing the precision of cellular categorization.

For stem cells, the most pertinent markers were identified as neuroligin-4 (gene-LOC118262926), a large-type Kenyon cell-specific protein in mushroom bodies (gene-LOC118270242), and a neurogenic locus Notch protein (gene-LOC118262320). Additional markers included a homeobox protein (gene-LOC118274815) and a headcase protein (gene-LOC118265494) (**See Supplementary Data 1 for a comprehensive list**). Among the enteroblast cells, the significant markers were neuronal acetylcholine receptor (gene-LOC118269784), endonuclease/exonuclease/phosphatase family protein (gene-LOC118273121), polyadenylate-binding protein (gene-LOC118279494), and eukaryotic translation initiation factor 5 A (gene-LOC118271679). For enterocytes (ECs), the following markers were used: chitinase-like protein (gene-LOC118276216), 3-oxoacyl-ACP synthase (gene-LOC118264942), spondin-2 (gene-LOC118276122), sugar transporter SWEET1-like (gene-LOC118267186), glutathione S-transferase 1 (gene-LOC118269546), and aldehyde dehydrogenase (gene-LOC118268490) (Fig. 3, **Supplementary Data 1**).

For enteroendocrine (EE) cells, the markers included proteoglycan 4 (gene-LOC118270259), dipeptidase 1 (gene-LOC118268421), tachykinins (gene-LOC118278153), and the protein turtle (gene-LOC118264893) (**Supplementary Data 1**). For Visceral Muscle Cells, the significant markers included thymosin beta (gene-LOC118276119), kinesin-like protein (gene-LOC118278067), protein Skeletor (gene-LOC118270244), and myotrophin (gene-LOC118271705). A comprehensive enumeration of these marker genes is available in **Supplementary Data 1**.

3.4. Gene expression signatures of identified cell types

The identified cell type-specific markers exhibiting an average natural logarithmic fold change (logFC) exceeding zero were utilized to interrogate the transcriptomic landscapes of individual clusters using KOBAS 3.0 and ClusterProfiler version 3.0 software [21,27]. The results of these analyses are shown in **Supplementary File 1, Fig. S4**. KEGG pathway enrichment analyses identified pathways, such as ribosomal biogenesis, ubiquitin-dependent proteolysis, Wnt signaling, and TGF-beta signaling, as markedly enriched in stem cell (SC) clusters (Fig. 4). Subsequent Gene Ontology (GO) analyses revealed a plethora of enriched biological processes, molecular functions, and cellular components, encompassing proteolytic activities, chromatin rearrangement, and protein–DNA complex assembly (Fig. 4). These GO terms are pivotal for the homeostasis, proliferation, self-renewal, and lineage specification of both intestinal and other stem cells, thereby corroborating that the molecular fingerprints of the FAW midgut SC clusters align with their anticipated functionalities.

In the EB cluster, the enriched KEGG pathways included ribosomal biogenesis, endoplasmic reticulum-associated protein processing, longevity-associated pathways, and mitophagy (Fig. 4). GO enrichment analyses revealed cellular phenomena such as cytoplasmic protein synthesis, cell cycle progression, and ribonucleoprotein complex subunit organization, implicating their roles in the further differentiation of enteroblast cells (Fig. 4). In the EE cell cluster, the enriched KEGG pathways included oxidative phosphorylation, ATP-dependent chromatin modulation, and ubiquitin-dependent proteolysis (Fig. 4). Comprehensive gene lists for these pathways are included in **Supplementary Data 2**. GO enrichment revealed cellular morphogenesis, neurite outgrowth, protein kinase enzymatic activities, and unfolded protein interactions, confirming that enteroendocrine cells primarily

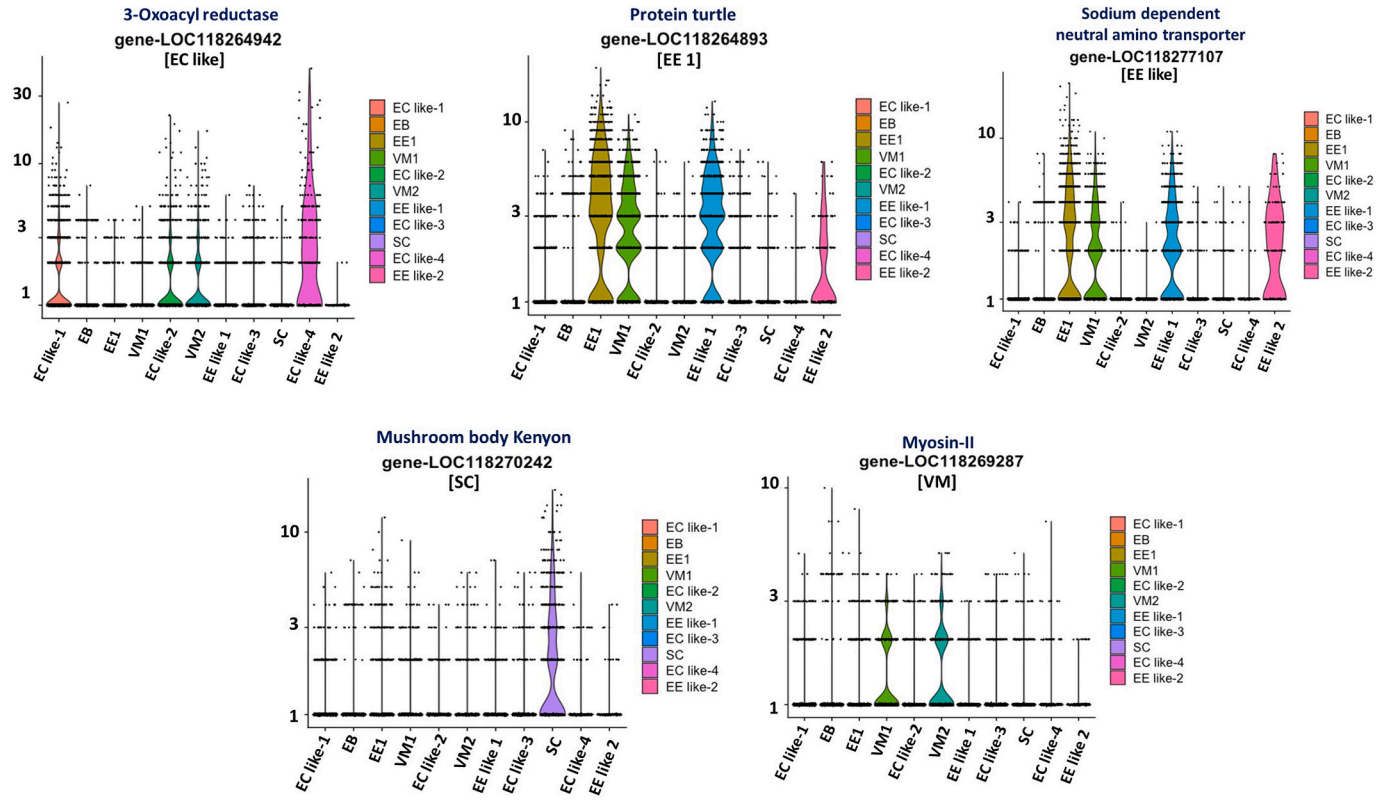

Fig. 3. Identification of cell-type specific markers in midgut cells using violin and feature plots. Both violin and feature plots were generated to visualize the expression patterns of novel marker genes across different midgut cell types. These marker genes were identified using the 'FindAllMarkers' algorithm based on a negative binomial distribution test requiring a minimum fold enrichment of 0.5 and a Bonferroni-adjusted *p*-value of <0.05. Further stratification considered the proportion of cells expressing a gene in individual clusters (pc.t1) versus across analogous clusters (pc.t2) to ensure robust specificity.

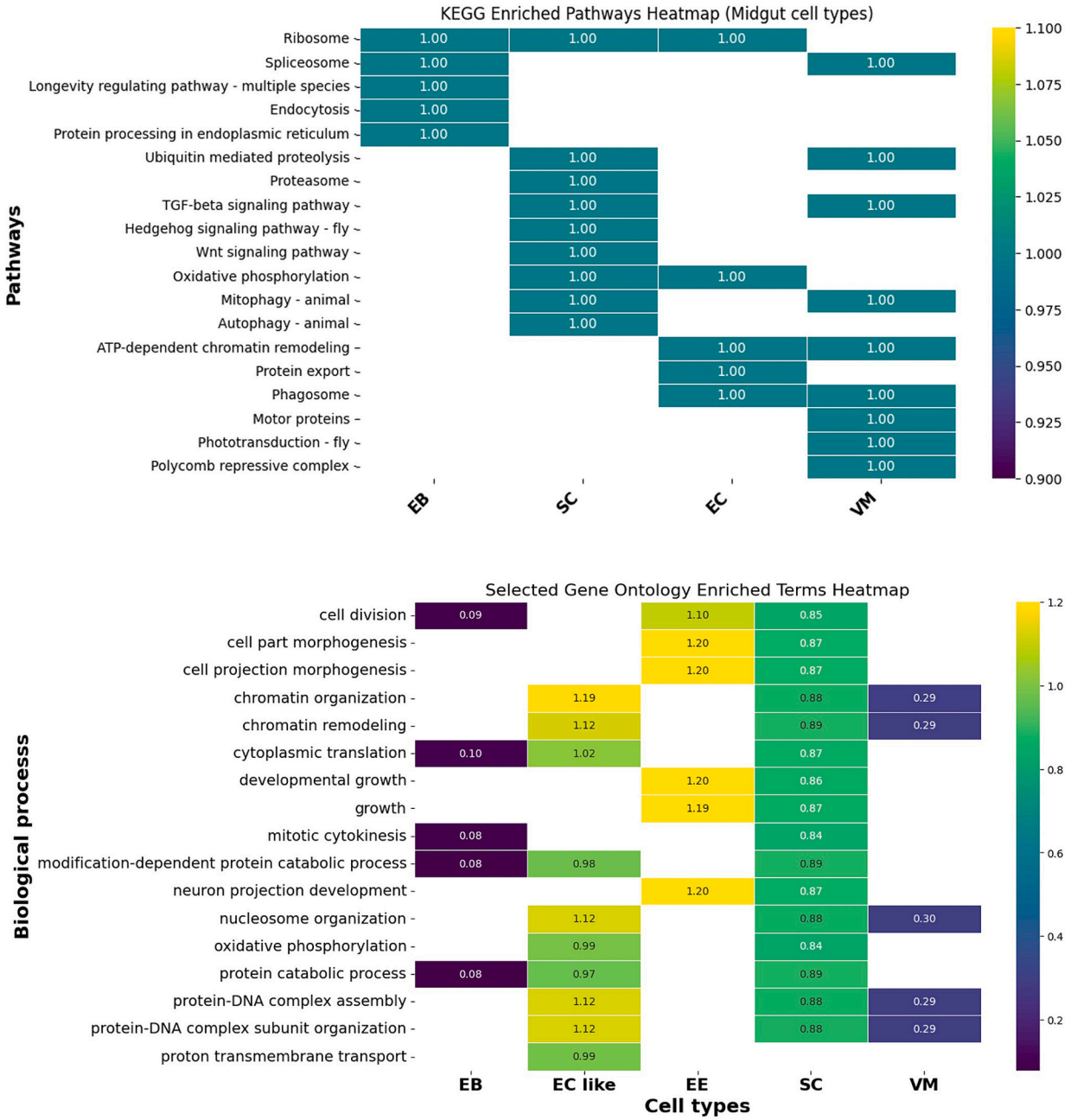

Fig. 4. Heatmap showing KEGG and GO enrichment in different midgut cells. The results are divided into **(1)** KEGG pathway enrichment specific to midgut cell types and **(2)** GO enrichment focused on biological processes. Enrichment scores are visually represented through color gradations. The data for each cellular condition were subjected to individual processing to maintain data uniformity and subsequently amalgamated and sieved based on an adjusted *p* value for discerning pivotal enriched terms. The derived 'Enrichment Score', which incorporates gene proportions from the sample and the background, elucidates variances in chosen pathways spanning midgut cell types. This heatmap, constructed using the Seaborn library, epitomizes the differences in enrichment scores for pivotal pathways among the detected clusters.

operate as secretory units (Fig. 4).

Utilizing uniform manifold approximation and projection (UMAP) analyses, additional clusters, specifically clusters 0, 2, 7, and 9 (termed EC-like 1), were found to be enriched in critical biological pathways, including oxidative phosphorylation, glutathione biosynthesis, and pyruvate metabolism (Fig. 4). The associated GO terms pertained to cytoplasmic protein synthesis, post-translational protein modification via small protein conjugation, and modification-mediated protein degradation (Fig. 4). These pathways are implicated in modulating absorption activities in midgut tissues.

Another distinct cluster, designated Visceral Muscles, demonstrated enrichment in pathways involving motor proteins, Wnt signaling, mitophagy, and ATP-dependent chromatin modification (Fig. 4). The GO terms enriched within this cluster were related to aerobic cellular respiration, proton transmembrane flux, and the respiratory electron transport chain, among others (Fig. 4). These pathways are known to modulate cellular activities in the midgut. A comprehensive inventory of genes implicated in cell type-specific KEGG enrichment is shown in **Supplementary Data 2**.

3.5. *Constructing lineage trajectories*

To investigate the relationships among distinct cellular clusters and elucidate the mechanisms of cellular differentiation, we employed

Monocle 3.0, a computational framework designed to analyze single-cell RNA sequencing data. This tool facilitates both differential gene expression and pseudotemporal ordering analyses [28–30]. Monocle implements generalized additive models to assess the association between gene expression levels and pseudotime, thereby allowing inference of developmental trajectories. For this lineage reconstruction, we targeted cell populations identified as SCs, EBs, EEs, ECs, and VMs. Our analysis revealed three distinct cellular differentiation pathways: **1)** SC → EB, **2)** EB → EE, and **3)** EB → EC-like. These trajectories corroborate existing biological paradigms, which predict that EBs differentiate from SCs and that both EEs and ECs originate from EBs (Fig. 5).

These studies employed graph-based tests to identify genes differentially expressed during pseudotemporal progression. This approach validated marker gene expression specific to each cellular subtype, confirming the veracity of the constructed developmental trajectories. These insights are particularly pertinent given that the SfMG-0617 cell line, the subject of our investigation, was developed from the midgut that contains muscle cells. Consequently, this comprehensive analysis substantiates the lineage relationships among these cellular types and validates the robustness of Monocle in accurately mapping cellular differentiation pathways (Fig. 5).

3.6. Expression of insecticide target genes in midgut cells

The expression of insecticide target genes within specific midgut cell types is not known. We examined the expression patterns of insecticide target genes, focusing on respiratory and midgut targets by analyzing scRNA-seq data from the SfMG-0617 cells utilizing sequences of known respiratory (e.g., ATP synthase subunit beta, cytochrome *b*-c1 complex, NADH-ubiquinone oxidoreductase, acetyl-CoA carboxylase) and targets of *Bacillus thuringiensis* toxin (e.g., aminopeptidase, alkaline phosphatase, cadherins, ABC transporters).

The results revealed distinct expression patterns of target genes in midgut cell subtypes. Aminopeptidase was predominantly expressed in EC-like cells, SCs, and VMs. ATP-binding cassette proteins exhibited expression across all midgut cell types, except in enteroblasts, and the same trend was observed for Cadherin proteins (Fig. 6 and **Supplementary File 1: Fig. S5**). Moreover, the phospholipid-transporting ATPase ABCA1 was predominantly expressed in EE-like and EC-like cells. The respiratory target ATP synthase and the cytochrome *b*-c1 subunit were broadly expressed across all cell types except for differentiating EB cells (**Fig. S5**). Similar expression trends were detected for NADH-ubiquinone oxidoreductase and acetyl-CoA carboxylase.

3.7. Isolation and enrichment of distinct cell types in the SfMG-0617 midgut cell line

The presence of different types of cells in the SfMG-0617 midgut cell line was confirmed using the marker genes identified. Promoters of marker genes were cloned into RFP (Red fluorescent protein) reporter vectors and transfected into SfMG-0617 cells. Notch and Forkhead promoters were used to mark stem cells and enteroblasts. Prospero and esterase promoters were employed to label enteroendocrine and enterocyte cells. EGFP expression was used to monitor for transfection efficiency (**Supplementary File 1: Fig. S6**). Cells transfected with constructs containing the Notch and Forkhead promoters exhibited more intense red fluorescence compared to cells transfected with constructs containing the Prospero and esterase promoters (Fig. 7). Notch promoter construct was used to isolate SCs and EBs cells using fluorescence-activated cell sorting (FACS) (**Supplementary File1: Figs. S7 and S8**) and puromycin antibiotic selection (**Fig. S7**). The FACS-sorted cells exhibited red fluorescence. After each round of puromycin selection, surviving cells exhibiting red fluorescence and morphological characteristics consistent with stem cells were identified until 100% of cells were RFP positive. These cells were then transferred to the conditioned medium supplemented with 20-hydroxyecdysone to stimulate differentiation and proliferation (**Supplementary File 1: Fig. S9**). Ongoing research aims to further characterize these cells in

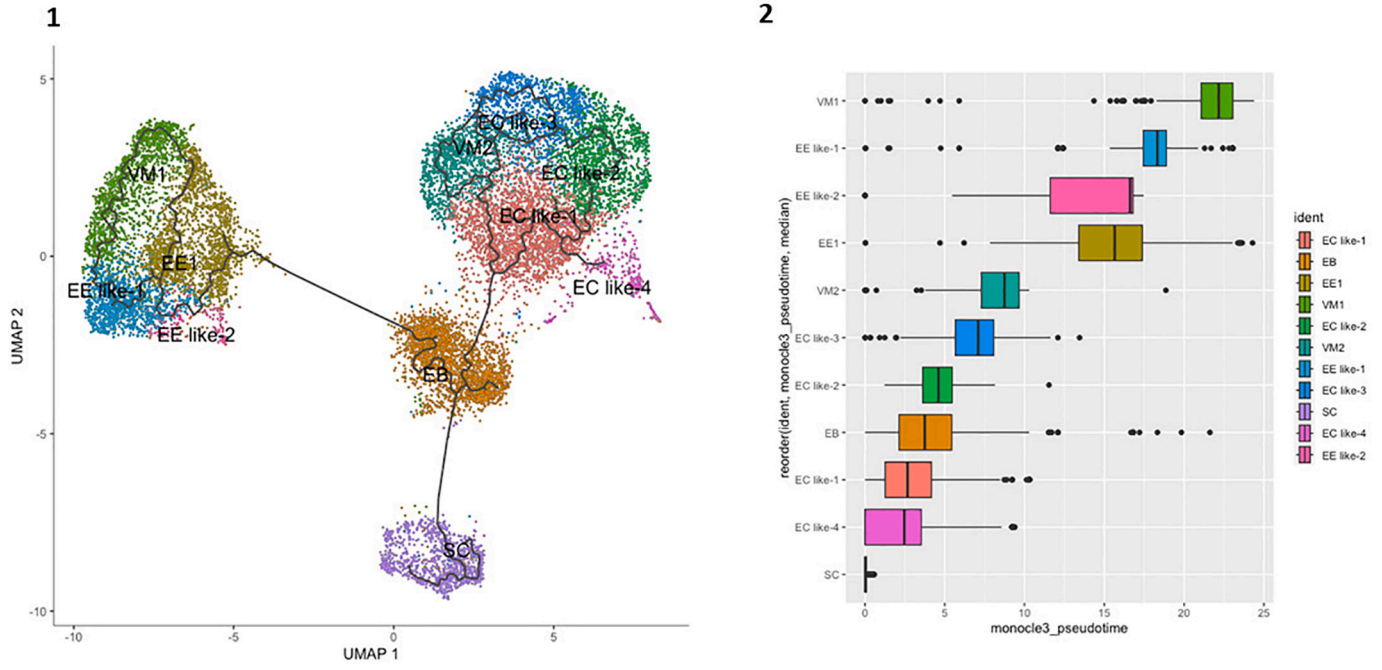

Fig. 5. Lineage reconstruction of cellular differentiation pathways using Monocle 3.0. 1. Overview of the cellular differentiation trajectories of intestinal cell populations, including stem cells (SC), enteroblasts (EB), enterocytes (EC), enteroendocrine (EE) and visceral muscle (VM) cells. Monocle 3.0 was utilized for single-cell RNA-sequencing data analysis, and generalized additive models were used to associate the gene expression levels with pseudotime. Three distinct differentiation pathways were identified: 1) SC → EB, 2) SC → EB → EE, and 3) EB → EC-like. These trajectories are consistent with existing biological paradigms, indicating that EB differentiates from SC and serves as a precursor for both EEs and ECs. **2)** The study also employed graph-based tests to identify genes differentially expressed during pseudotemporal progression, enabling the validation of marker gene expression specific to each cellular subtype.

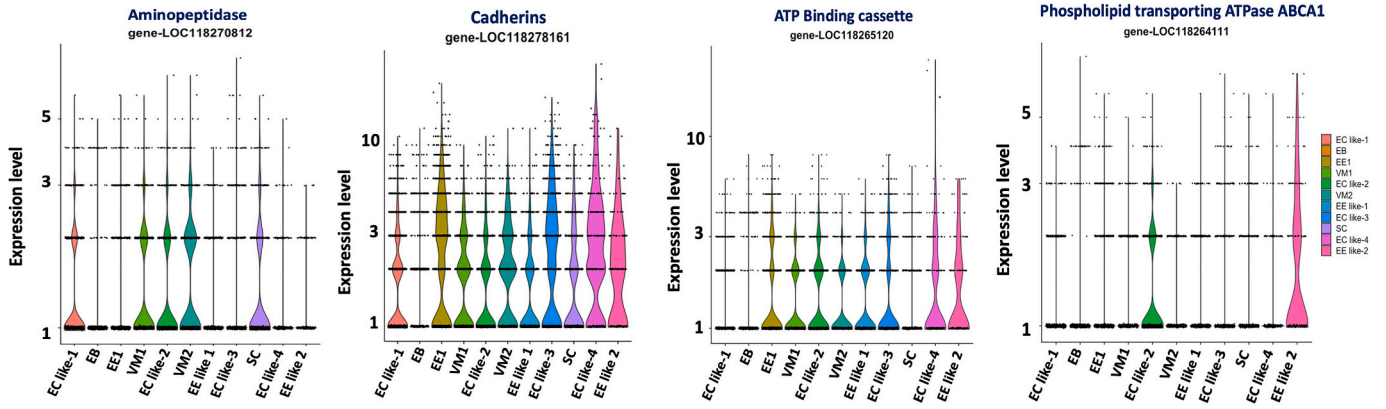

Fig. 6. Localization expression of the gene coding for *Bacillus thuringiensis* toxin target proteins in midgut cells. This violin and feature plot illustrates the expression patterns of three Bt targets: aminopeptidase, Cadherins, ATP binding Cassette, and phospholipid transporting ATPase ABCA1, in SfMG-0617 cell clusters.

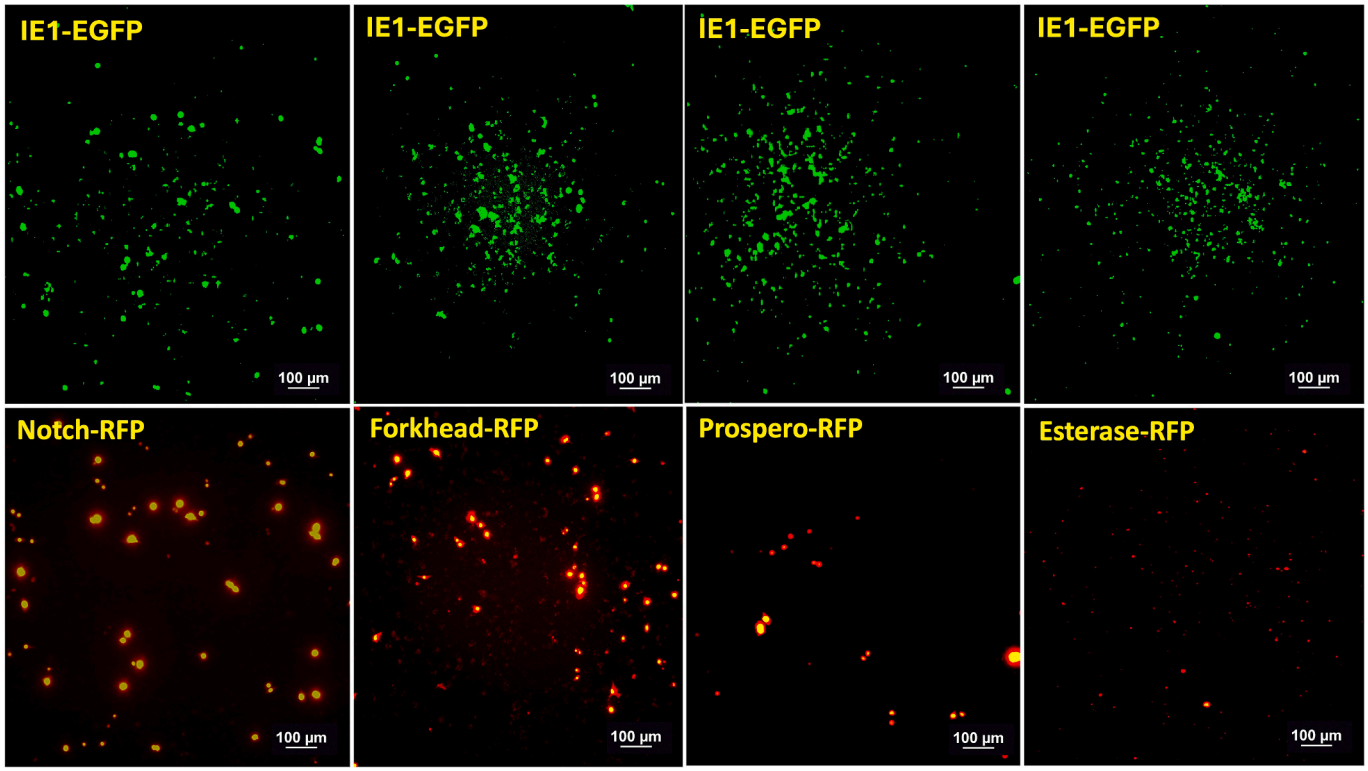

Fig. 7. Photographs of midgut cells transfected with cell-type specific marker gene promoter constructs taken under a fluorescence microscope for red and green fluorescent protein signals. The Notch gene (specifically labels stem cells, A), Forkhead gene (marks enterocytes or columnar cells, B), Prospero gene (marks stem cells, C) and the Esterase gene (marks enteroendocrine cells, D) promoter constructs and IE1-GFP construct were transfected into SfMG-0617 cells. The cells were photographed under a fluorescence microscope. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

terms of their proliferative and differentiative capacities, as well as to identify additional factors that may modulate these processes.

4. Discussion

Cellular heterogeneity research is crucial for understanding the unique functions and adaptations of non-model insects. These insects, unlike well-studied model organisms, have specialized physiological processes due to their distinct evolution. Studies on cellular heterogeneity could help understand their development, physiology, and responses to environmental challenges. Some results have been reported from non-model organisms. For example, Vigneron et al. [31] used single-cell RNA sequencing to analyze *Trypanosoma brucei*, a parasite transmitted by tsetse flies. This study provided a detailed transcriptomic profile of individual cells, revealing the process of metacyclogenesis, which is crucial for understanding how the parasite becomes infectious. This insight is essential for developing strategies to block disease transmission. Cui and Franz (2020) [19] applied single-nucleus RNA sequencing to the mosquito *Aedes aegypti* and identified 20 distinct midgut cell types and significant changes in cell composition and gene expression following blood meal ingestion. This cellular diversity is key to efficient digestion and nutrient absorption. The scRNA-seq, was employed to study hemocytes in *Anopheles gambiae* and *Aedes aegypti* [32] and *Marsupenaeus japonicus* [33]. These studies highlighted the

complexity and specialization of immune cells. The scRNA-seq was used to create a comprehensive cell atlas of the adult female midgut [34]. These studies identified cell types and showed how different regions of the midgut interact with arboviruses, crucial for understanding the mosquito's vector competence and immune response. These studies underscore the value of cellular heterogeneity research. Advanced techniques like scRNA-seq allow researchers to identify novel cell types and understand their roles in development, immune response, and host-pathogen interactions. This knowledge is essential for developing targeted strategies to control vector-borne diseases, improve pest management, and enhance the ecological and agricultural benefits.

Insect cell lines have become indispensable tools in molecular biology, virology, and bioprocessing research. The ability of these peptides to express recombinant proteins makes them valuable assets for the biopharmaceutical industry and beyond. However, understanding and managing cellular heterogeneity within these cell lines is crucial for maximizing their utility. Cellular heterogeneity in insect cell lines is important in biology research and biotechnology applications. Recent advances have introduced novel methods for examining cellular heterogeneity. Fluorescent reporter proteins have revolutionized our ability to monitor gene expression driven by specific promoters in real-time [35,36]. This approach allows non-invasive evaluation of the impact of specific conditions on individual cells, providing opportunities for engineering and precise isolation of particular cell populations. As the field of cellular heterogeneity evolves, these innovative approaches promise to revolutionize our understanding of cell behavior and expand their applications beyond bioprocessing. In 2007, Stockholm et al. [37]. discussed phenotypic heterogeneity within clonal cell populations in vitro, emphasizing the role of local stochastic interactions between phenotypically identical cells in initiating phenotypic switches. The use of live-cell imaging techniques allowed the identification of distinct phenotypes within the population of baculovirus-infected Sf9 cells. It allowed the classification of infected cells based on the timing and level of gene expression, revealing the dynamic responses of insect cell cultures to infection [38].

Recent technological advances have facilitated studies on cellular heterogeneity. Methods such as microfluidic-based cell isolation, next-generation sequencing, and innovative labeling strategies for proteomics enable the comprehensive characterization of DNA, RNA, and proteins at single-cell resolution [39,40]. These single-cell omics profiling strategies have far-reaching implications, including delineating cellular diversity, tracing cell lineages, identifying new cell types, and deciphering regulatory mechanisms between omics layers. Single-cell analysis revealed heterogeneity among different cell types within tissues and clonal cell populations. Our studies used 10× Genomics Chromium technology to investigate the heterogeneity of the SfMG-0617 cell line. Surprisingly, ten different midgut cell types were detected in a cell line that had been in culture for many years and had undergone >50 passages. Notably, 23% of the sequenced SfMG-0617 cells were identified as stem cells or enteroblasts, challenging the widely held belief that foundational cells inevitably differentiate after multiple passages in culture. The assumption that cells undergo “development” in culture to transition between identities has been a fundamental premise in cell culture research for decades.

Recent studies in cell biology and related areas have highlighted the significance of heterogeneity within cell populations. For instance, in studies on pancreatic β cells, large-scale CRISPR screens combined with single-cell RNA sequencing (scRNA-seq) revealed genetic heterogeneity in the MIN6 cell line. This pancreatic β-cell line increases with passage number. This genetic heterogeneity led to distinct functional clusters, including endocrine, basal, epithelial, and neuroendocrine cells, with a unique lncRNA-enriched cluster showing differentially expressed insulin transcription. These findings challenge the assumption of homogeneity within cell lines and suggest that individual cells within a population can exhibit diverse characteristics and behaviors [41]. Similarly, in the case of human pluripotent stem cells (hPSCs), traditionally considered homogeneous, recent advancements in single-cell technologies have revealed a high degree of variability between individual cells within hPSC populations. This variability can arise from genetic and epigenetic abnormalities associated with long-term in vitro culture and somatic cell reprogramming. Some of these variations can confer growth advantages to specific cells, altering cellular phenotypes and posing concerns in hPSC applications. On the other hand, intrinsic heterogeneity within hPSCs, such as an asynchronous cell cycle and spatial asymmetry in cell adhesion, has been found to produce multiple lineages during differentiation, redefining the concept of pluripotency [42].

Moreover, in cancer research, it has become increasingly evident that multilevel heterogeneity is a fundamental feature often overlooked or underestimated. Cancer cell populations with high genomic instability exhibit karyotype heterogeneity, which makes cloning cells challenging and leads to growth heterogeneity, where outliers dominantly contribute to population growth [43]. This complexity challenges the accuracy of conventional methods and underscores the need for single-cell analysis when cells are not genetically identical. Our studies on SfMG-0617 cell line heterogeneity align with recent research across various domains, highlighting the critical importance of acknowledging and addressing heterogeneity within cell populations. These findings emphasize that the assumption of uniform cell development in culture may not hold true and underscore the need for more nuanced approaches to studying and manipulating cells, with potential implications for insecticide discovery and development and broader insights into cell biology, disease modeling, and cancer research.

We also report on the lineage trajectories shedding light on the mechanisms of cellular differentiation. We employed Monocle 3.0 computational framework designed to analyze scRNA-seq data to achieve this goal. This tool allowed us to identify cell types and infer developmental trajectories, a novel aspect that has yet to be extensively explored in previous studies. Our lineage reconstruction analysis revealed three distinct cellular differentiation pathways: 1) SC → EB, 2) EB → EE, and 3) EB → EC-like. These trajectories align with existing biological paradigms that predict the differentiation of EB from SC and the subsequent development of both EE and EC from EB [5,9,19,22,44–48].

The interpretation of these lineage trajectories assumes that cells within our culture system are developing and transitioning in a manner that mirrors natural development within the whole organism. Recent studies across various fields have supported this assumption. For instance, research in mouse embryonic stem (ES) cells [49] revealed stochastic and reversible transitions along a linear chain of states. This suggests that living cells transition between specific molecularly and developmentally distinct states. Similarly, a microfluidic platform enabling single-cell RNA–seq after lineage tracking [50] demonstrated that cells within lineages exhibit greater intralineage transcriptional similarity, highlighting the relevance of lineage-based analysis in understanding heterogeneous cell populations. Furthermore, the LINNAEUS (lineage tracing by nuclease-activated editing of ubiquitous sequences) strategy [51] allowed the simultaneous tracking of lineage and transcriptome profiles in thousands of single cells, providing a systematic approach for tracing the developmental origin of novel cell types. Drawing parallels with these studies, we emphasize the relevance of considering cell development in culture and lineage-based analysis in elucidating complex cellular differentiation pathways. These data contribute to the broader fields of cell and developmental biology. These findings underscore the idea that cell cultures may faithfully recapitulate natural developmental processes, paving the way for more comprehensive investigations into cell identity and lineage trajectories across various biological systems.

We also employed graph-based tests to identify genes differentially expressed along these pseudotemporal progression pathways. This approach validated marker gene expression specific to each cellular subtype, reinforcing the robustness of the constructed developmental trajectories. In conjunction with our cell type identification, this

advancement in lineage reconstruction provides a comprehensive understanding of the dynamic processes governing midgut cell differentiation [44,45]. We report on examining the localization of insecticide targets within specific midgut cells. Previous research has focused primarily on identifying these target sites, but their expression within distinct cell subtypes have largely been unexplored [52]. We analyzed scRNA-seq data from the SfMG-0617 cell line to investigate the expression patterns of target sites of insecticides. Distinct expression patterns for these targets across midgut cell subtypes were detected. We conducted statistical analyses to evaluate the significance of expression patterns of genes coding for insecticide among the different cell types. We employed the 'FindAllMarkers' algorithm, which is based on a negative binomial distribution. A minimum fold enrichment of 0.5 and a Bonferroni-adjusted *p*-value of <0.05 were used as criteria for statistical significance. This analysis validated the observed distinct expression patterns and provided evidence for the differential expression of insecticide target genes in different cell types. For instance, aminopeptidase is predominantly expressed in enterocyte-like cells, stem cells, and visceral muscles, suggesting potential roles in digestion and muscle function. Cadherin and ATP-binding cassette proteins exhibit ubiquitous expression patterns across most midgut cell types, implicating their involvement in diverse cellular processes. Regarding respiratory targets, genes such as ATP synthase beta and the cytochrome *b*-c1 subunit exhibited broad expression across various cell types, highlighting their fundamental roles in cellular energy production. This in-depth analysis advances our understanding of the distribution of insecticide targets within the midgut and has significant implications for developing targeted insecticides and resistance management strategies.

Our studies employed innovative methods for isolating distinct cells from the heterogeneous midgut cell population. We designed plasmid constructs to express fluorescent proteins under the control of specific promoters in stem cells, enteroblasts, enteroendocrine cells, and enterocytes. This approach allowed us to evaluate the activity of promoter regions for these genes and led to the selection of the notch promoter for controlling the expression of a puromycin resistance cassette. Subsequent puromycin selection and fluorescence-activated cell sorting (FACS) enabled the isolation of stem cells. This approach facilitated the isolation and characterization of specific midgut cell types and opened avenues for further research into the factors that modulate their differentiation.

To further investigate the stability and differentiation trajectories of cell-like subtypes, we are currently exploring the effect of 20-Hydroxyecdysone (20E) on isolated stem cells. The 20E has previously been reported by many researchers and has demonstrated its effectiveness in inducing significant morphological changes and promoting differentiation in insect cells. Early foundational studies provided the basis for our experimental design. In 1981, research on the *Manduca sexta* cell line (MRRL-CH1) showed that 20E induced elongation and clumping responses, highlighting the importance of the cytoskeleton in mediating these effects [53]. This work laid the groundwork for understanding how 20E influences cellular morphology. Further studies in *Spodoptera frugiperda* and *Plodia interpunctella* cell lines demonstrated that 20E facilitates differentiation into specialized cell types [54]. This study helped illustrate the broader applicability of 20E across different insect cell lines. The 20E induced cell clumping, elongation, and differentiation, leading to increased cellular complexity and functional specialization [55]. 20E could trigger morphogenetic and secretory processes, as shown in the Indian meal-moth cell line (IAL-PID2) [56]. The cells formed pseudo-epithelial aggregates and synthesized glycoproteins essential for cellular expansion and membrane integrity, further supporting 20E role in cellular differentiation. Studies on the *Spodoptera frugiperda* (Sf21) cell line combined 20E with insulin, producing a neuronal phenotype and enhancing the differentiation process. This synergistic effect provided a robust model for high-throughput insecticide screening and reinforced the importance of 20E in differentiation [55]. The latest research confirmed the role of 20E in inducing significant morphological changes, such as cell aggregation, elongation, and the formation of axon-like processes, indicating neuronal differentiation and functional specialization [55]. By building on this foundational research, we aimed to further elucidate the role of 20E in maintaining and inducing specific cell states in cultured insect cells.

5. Challenges and study limitations

While our study represents a significant advancement in understanding cellular heterogeneity in non-model insects, several challenges remain. One limitation is the reliance on scRNA-seq data, which, although comprehensive, may not capture all aspects of cellular heterogeneity due to technical limitations such as dropout events and sequencing depth. Future studies could integrate other omics data, such as proteomics and metabolomics, to provide a more holistic view of cellular functions. Understanding cellular heterogeneity in non-model insects has broad implications for pest management and vector control. Insights gained from these studies can inform the development of targeted insecticides, strategies for managing insecticide resistance, and novel biotechnological applications, such as recombinant protein production.

6. Relevance and future applications

Our findings provide insights into basic biology and have implications for targeted interventions in pest management, especially in dealing with the economic and ecological impact of the Fall armyworm. Understanding the cellular landscape of the midgut can be crucial for developing new strategies to control this global pest. This study highlights the power of advanced single-cell RNA sequencing methods and bioinformatics approaches for revealing cellular heterogeneity and functionality. These findings validate the existing biological paradigms and pave the way for in-depth investigations into the functional genomics of this important cell line.

7. Conclusions

In conclusion, research into cellular heterogeneity in non-model insects is essential for unraveling the complexities of their biology. The insights gained from such studies not only advance our fundamental understanding of insect physiology but also have practical applications in disease control, agriculture, and biotechnology. By exploring the cellular diversity and function in these organisms, we can develop innovative solutions to global challenges related to health and environmental sustainability.

CRediT authorship contribution statement

Surjeet Kumar Arya: Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Conceptualization. **Douglas A. Harrison:** Writing – review & editing, Resources, Methodology. **Subba Reddy Palli:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors did not generative AI and AI-assisted technologies in the writing process and take full responsibility for the content of the published article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This material is based on the work supported by the National Science Foundation I/UCRC, the Center for Arthropod Management Technologies (under grant no. IIP-1821936 and by industry partners), and the US Department of Agriculture (under HATCH Project 2353057000). We thank the College of Arts & Sciences Imaging Center and the Flow Cytometry and Immune Monitoring Core for technical assistance with scRNA-seq and flow cytometry analysis, respectively. The A&S Imaging Center is supported by the UK College of Arts & Sciences and the Department of Biology. The UK Flow Cytometry & Immune Monitoring core facility is partly supported by the Office of the Vice President for Research, the Markey Cancer Center, and an NCI Center Core Support Grant (P30 CA177558) to the University of Kentucky Markey Cancer Center.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2024.110898.

References

[1] A. Bairoch, The cellosaurus, a cell-line knowledge resource, J. Oiomol. Techniq.: JBT 29 (2018) 25.
[2] G.F. Caputo, S.S. Sohi, S.V. Hooey, L.J. Gringorten, Insect cell lines and baculoviruses as effective biocontrol agents of forest pests, in: BMC Proceedings, Springer, 2011, p. P71.
[3] S.K. Arya, C.L. Goodman, D. Stanley, S.R. Palli, A database of crop pest cell lines, In Vitro Cellular & Developmental Biology-Animal 58 (2022) 719–757.
[4] L. Swevers, J. Vanden Broeck, G. Smagghe, The possible impact of persistent virus infection on the function of the RNAi machinery in insects: a hypothesis, Front. Physiol. 4 (2013) 66388.
[5] R.S. Hakim, S. Caccia, M. Loeb, G. Smagghe, Primary culture of insect midgut cells, In Vitro Cellular & Developmental Biology-Animal 45 (2009) 106–110.
[6] K. Zhou, C.L. Goodman, J. Ringbauer, Q. Song, B. Beerntsen, D. Stanley, Establishment of two midgut cell lines from the fall armyworm, Spodoptera frugiperda (Lepidoptera: Noctuidae), In Vitro Cellular & Developmental Biology-Animal 56 (2020) 10–14.
[7] A. Khorramnejad, H.D. Perdomo, U. Palatini, M. Bonizzoni, L. Gasmi, Cross talk between viruses and insect cells cytoskeleton, Viruses 13 (2021) 1658.
[8] C. Trapnell, Defining cell types and states with single-cell genomics, Genome Res. 25 (2015) 1491–1498.
[9] R.-J. Hung, Y. Hu, R. Kirchner, Y. Liu, C. Xu, A. Comjean, S.G. Tattikota, F. Li, W. Song, S. Ho Sui, A cell atlas of the adult Drosophila midgut, Proc. Natl. Acad. Sci. 117 (2020) 1514–1523.
[10] R.N. Nagoshi, N.N. Htain, D. Boughton, L. Zhang, Y. Xiao, B.Y. Nagoshi, D. Mota-Sanchez, Southeastern Asia fall armyworms are closely related to populations in Africa and India, consistent with common origin and recent migration, Sci. Rep. 10 (2020) 1421.
[11] E. Njuguna, P. Nethononda, K. Maredia, R. Mbabazi, P. Kachapulula, A. Rowe, D. Ndolo, Experiences and perspectives on *Spodoptera frugiperda* (Lepidoptera: Noctuidae) management in sub-Saharan Africa, J. Integr. Pest Manag. 12 (2021) 7.
[12] K. Overton, J.L. Maino, R. Day, P.A. Umina, B. Bett, D. Carnovale, S. Ekesi, R. Meagher, O.L. Reynolds, Global crop impacts, yield losses and action thresholds for fall armyworm (*Spodoptera frugiperda*): a review, Crop Prot. 145 (2021) 105641.
[13] A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, Integrating single-cell transcriptomic data across different conditions, technologies, and species, Nat. Biotechnol. 36 (2018) 411–420.
[14] L. Maaten, Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (2008) 2579.
[15] E. Becht, L. McInnes, J. Healy, C.-A. Dutertre, I.W. Kwok, L.G. Ng, F. Ginhoux, E. W. Newell, Dimensionality reduction for visualizing single-cell data using UMAP, Nat. Biotechnol. 37 (2019) 38–44.
[16] C.S. McGinnis, L.M. Murrow, Z.J. Gartner, DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors, Cell systems 8 (2019) 329–337, e324.
[17] I. Korsunsky, N. Millard, J. Fan, K. Slowikowski, F. Zhang, K. Wei, Y. Baglaenko, M. Brenner, P.-R. Loh, S. Raychaudhuri, Fast, sensitive and accurate integration of single-cell data with harmony, Nat. Methods 16 (2019) 1289–1296.
[18] C. Hafemeister, R. Satija, Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression, Genome Biol. 20 (2019) 296.
[19] Y. Cui, A.W. Franz, Heterogeneity of midgut cells and their differential responses to blood meal ingestion by the mosquito *Aedes aegypti*, Insect Biochem. Mol. Biol. 127 (2020) 103496.
[20] T. Wu, E. Hu, S. Xu, M. Chen, P. Guo, Z. Dai, T. Feng, L. Zhou, W. Tang, L., Zhan, clusterProfiler 4.0: a universal enrichment tool for interpreting omics data, Innov 2 (2021).
[21] G. Yu, L.-G. Wang, Y. Han, Q.-Y., He, clusterProfiler: an R package for comparing biological themes among gene clusters, Omics 16 (2012) 284–287.
[22] C.N. Perdigoto, F. Schweisguth, A.J. Bardin, Distinct levels of notch activity for commitment and terminal differentiation of stem cells in the adult fly intestine, Development 138 (2011) 4585–4595.
[23] D. Dutta, A.J. Dobson, P.L. Houtz, C. Gläßer, J. Revah, J. Korzelius, P.H. Patel, B. A. Edgar, N. Buchon, Regional cell-specific transcriptome mapping reveals regulatory complexity in the adult Drosophila midgut, Cell Rep. 12 (2015) 346–358.
[24] M. Nászai, L.R. Carroll, J.B. Cordero, Intestinal stem cell proliferation and epithelial homeostasis in the adult Drosophila midgut, Insect Biochem. Mol. Biol. 67 (2015) 9–14.
[25] X. Guo, C. Yin, F. Yang, Y. Zhang, H. Huang, J. Wang, B. Deng, T. Cai, Y. Rao, R. Xi, The cellular diversity and transcription factor code of *Drosophila* enteroendocrine cells, Cell Rep. 29 (2019) 4172–4185 (e4175).
[26] G. Lin, N. Xu, R. Xi, Paracrine wingless signalling controls self-renewal of Drosophila intestinal stem cells, Nature 455 (2008) 1119–1123.
[27] D. Bu, H. Luo, P. Huo, Z. Wang, S. Zhang, Z. He, Y. Wu, L. Zhao, J. Liu, J. Guo, KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis, Nucleic Acids Res. 49 (2021) W317–W325.
[28] X. Qiu, A. Hill, J. Packer, D. Lin, Y.-A. Ma, C. Trapnell, Single-cell mRNA quantification and differential analysis with census, Nat. Methods 14 (2017) 309–315.
[29] X. Qiu, Q. Mao, Y. Tang, L. Wang, R. Chawla, H.A. Pliner, C. Trapnell, Reversed graph embedding resolves complex single-cell trajectories, Nat. Methods 14 (2017) 979–982.
[30] C. Trapnell, D. Cacchiarelli, J. Grimsby, P. Pokharel, S. Li, M. Morse, N.J. Lennon, K.J. Livak, T.S. Mikkelsen, J.L. Rinn, Pseudo-temporal ordering of individual cells reveals dynamics and regulators of cell fate decisions, Nat. Biotechnol. 32 (2014) 381.
[31] A. Vigneron, M.B. O'Neill, B.L. Weiss, A.F. Savage, O.C. Campbell, S. Kamhawi, J. G. Valenzuela, S. Aksoy, Single-cell RNA sequencing of Trypanosoma brucei from tsetse salivary glands unveils metacyclogenesis and identifies potential transmission blocking antigens, Proc. Natl. Acad. Sci. 117 (2020) 2613–2621.
[32] H. Kwon, M. Mohammed, O. Franzén, J. Ankarklev, R.C. Smith, Single-cell analysis of mosquito hemocytes identifies signatures of immune cell subtypes and cell differentiation, Elife 10 (2021) e66192.
[33] K. Koiwai, T. Koyama, S. Tsuda, A. Toyoda, K. Kikuchi, H. Suzuki, R. Kawano, Single-cell RNA-seq analysis reveals penaeid shrimp hemocyte subpopulations and cell differentiation process, Elife 10 (2021) e66954.
[34] S. Wang, Y. Huang, F. Wang, Q. Han, N. Ren, X. Wang, Y. Cui, Z. Yuan, H. Xia, A cell atlas of the adult female *Aedes aegypti* midgut revealed by single-cell RNA sequencing, Scientific Data 11 (2024) 587.
[35] D.L. Coutu, T. Schroeder, Probing cellular processes by long-term live imaging–historic problems and current solutions, J. Cell Sci. 126 (2013) 3805–3815.
[36] S. George, A.M. Jauhar, J. Mackenzie, S. Kieβlich, M.G. Aucoin, Temporal characterization of protein production levels from baculovirus vectors coding for GFP and RFP genes under non-conventional promoter control, Biotechnol. Bioeng. 112 (2015) 1822–1831.
[37] D. Stockholm, R. Benchaouir, J. Picot, P. Rameau, T.M.A. Neildez, G. Landini, C. Laplace-Builhé, A. Paldi, The origin of phenotypic heterogeneity in a clonal cell population in vitro, PLoS One 2 (2007) e394.
[38] M. Silvano, R. Correia, N. Virgolini, C. Clarke, P.M. Alves, I.A. Isidro, A. Roldão, Gene expression analysis of adapted insect cells during influenza VLP production using RNA-sequencing, Viruses 14 (2022) 2238.
[39] J. Lee, D.Y. Hyeon, D. Hwang, Single-cell multiomics: technologies and data analysis methods, Exp. Mol. Med. 52 (2020) 1428–1442.
[40] P. Hu, W. Zhang, H. Xin, G. Deng, Single cell isolation and analysis, Front. Cell Develop. Biol. 4 (2016) 116.
[41] R. Zhao, J. Lu, Q. Li, F. Xiong, Y. Zhang, J. Zhu, G. Peng, J. Yang, Single-cell heterogeneity analysis and CRISPR screens in MIN6 cell line reveal transcriptional regulators of insulin, Cell Cycle 20 (2021) 2053–2065.
[42] S. Yang, Y. Cho, J. Jang, Single cell heterogeneity in human pluripotent stem cells, BMB Rep. 54 (2021) 505.
[43] B.Y. Abdallah, S.D. Horne, J.B. Stevens, G. Liu, A.Y. Ying, B. Vanderhyden, S. A. Krawetz, R. Gorelick, H.H. Heng, Single cell heterogeneity: why unstable genomes are incompatible with average profiles, Cell Cycle 12 (2013) 3640–3649.
[44] M.J. Loeb, R.S. Hakim, Insect midgut epithelium in vitro: an insect stem cell system, J. Insect Physiol. 42 (1996) 1103–1111.
[45] B. Ohlstein, A. Spradling, The adult *Drosophila* posterior midgut is maintained by pluripotent stem cells, Nature 439 (2006) 470–474.
[46] B. Biteau, H. Jasper, Slit/Robo signaling regulates cell fate decisions in the intestinal stem cell lineage of Drosophila, Cell Rep. 7 (2014) 1867–1875.
[47] Z. Guo, B. Ohlstein, Bidirectional notch signaling regulates *Drosophila* intestinal stem cell multipotency, Science 350 (2015) aab0988.
[48] S. Caccia, M. Casartelli, G. Tettamanti, The amazing complexity of insect midgut cells: types, peculiarities, and functions, Cell Tissue Res. 377 (2019) 505–525.

[49] S. Hormoz, Z.S. Singer, J.M. Linton, Y.E. Antebi, B.I. Shraiman, M.B. Elowitz, Inferring cell-state transition dynamics from lineage trees and endpoint single-cell measurements, Cell Syst. 3 (2016) 419–433 (e418).
[50] R.J. Kimmerling, G. Lee Szeto, J.W. Li, A.S. Genshaft, S.W. Kazer, K.R. Payer, J. de Riba Borrajo, P.C. Blainey, D.J. Irvine, A.K. Shalek, A microfluidic platform enabling single-cell RNA-seq of multigenerational lineages, Nat. Commun. 7 (2016) 10220.
[51] B. Spanjaard, B. Hu, N. Mitic, P. Olivares-Chauvet, S. Janjuha, N. Ninov, J. P. Junker, Simultaneous lineage tracing and cell-type identification using CRISPR–Cas9-induced genetic scars, Nat. Biotechnol. 36 (2018) 469–473.
[52] R. Nauen, R. Slater, T.C. Sparks, A. Elbert, A. Mccaffery, IRAC: insecticide resistance and mode-of-action classification of insecticides, Modern crop protection compounds 3 (2019) 995–1012.
[53] D.E. Lynn, H. Oberlander, The effect of cytoskeletal disrupting agents on the morphological response of a cloned Manduca sexta cell line to 20-hydroxy-ecdysone, Wilhelm Roux's Archiv.Develop. Biol. 190 (1981) 150–155.
[54] D. Lynn, H. Oberlander, The establishment of cell lines from imaginal wing discs of Spodoptera frugiperda and Plodia interpunctella, J. Insect Physiol. 29 (1983) 591–596.
[55] N. Kislev, I. Segal, M. Edelman, Ecdysteroids induce morphological changes in continuous cell lines of Lepidoptera, Wilhelm Roux's Archiv.Develop. Biol. 193 (1984) 252–256.
[56] P. Cassier, P. Serrant, R. Garcia, N. Coudouel, M. André, D. Guillaumin, P. Porcheron, H. Oberlander, Morphological and cytochemical studies of the effects of ecdysteroids in a lepidopteran cell line (IAL-PID2), Cell Tissue Res. 265 (1991) 361–369.