Comparing methods for Single-Cell RNA Sequencing analysis in the characterization of mir-190 during cell fate selection

Ariana Adele Nagainis
Department of Molecular and Cellular Biology
University of Washington
nagainis@uw.edu

Shubhkarman Singh
Department of Computer Science
University of Washington
shubhs2@cs.washington.edu

Abstract

Stem cells are important in the generation of cell diversity. They must ensure a properly-timed cascade of transcription factors that determine cell fate. However, the mechanisms that balance the generation of cellular diversity is poorly understood. To elucidate these mechanisms, the Drosophila melanogaster larval central nervous system (CNS) has become a powerful model. The neuroblasts (NBs) in the larval central nervous system divide asymmetrically and give rise to cells with different developmental fates. Currently, there is little known about whether regulation of cell fate may involve microRNAs (miRNAs). These short non-coding RNA's silence gene expression of complementary mRNA transcripts. There is a possibility that temporal and spatial regulators at the post transcriptional level have potential to coordinate asymmetry of cell fate determinants. Previous research has used a single-cell RNA sequencing (scRNAseq) to compare the transcriptomics of individual cells between wildtype (wt) and mir-190 KO cells in drosophila. mir-190 deficient cells show a decrease in neuron production and an increase in differentiating cells, glial cells and kenyon cells compared to the control. This data suggests miR-190 plays a role in regulating the balance of glial cells and possibly affects self-renewal pathways. Currently, many scRNAseq publications use Seurat for their anaylsis. Although Seurat is the most popular package, there are a plethora of other scRNAseq packages that differ in their approach and can affect the accuracy and interpretability of the results. Thus our goal for this project is to compare the accuracy of results from different scRNA-seq data analysis methods, such as Seurat, scVI, constrastiveVI, and linearVI. We aim to determine if mir-190 deficient cells remain to show a decrease in neuron production compared to wildtype with newer and improved scRNAseq analysis.

1 Introduction

Single-cell RNA sequencing (scRNAseq) is a dynamic technique to examine transcriptomes in individual cells (Olsen et al., 2018). It is used to characterize and distinguish various cell types and examine cell-specific changes in the transcriptome. scRNAseq is useful in discovering new or rare cell types, identifying differences in cellular composition between mutant and non-mutant tissue, and understanding cellular differentiation during development (Hwang et. al., 2018).

In particular, scRNAseq is an incredibly useful tool for understanding the transcriptional difference between stem cells and their differentiated progeny during early development. The Drosophila central nervous system (CNS) is a well-described model for studying early development and cell fate selection (Harding and White, 2018). Within the Drosophila CNS are neural stem cells called neuroblasts which must undergo a properly-timed cascade of transcriptional changes to differentiate

34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.

into its proper cell fate. Currently, there are many scRNAseq datasets and atlases purely on the Drosophila CNS where all the various neural cell types are clustered and gene expression annotated.

Currently, many Drosophila larval brain scRNAseq publications use either Seurat or Monocle, both widely used in the single-cell genomics community, and both offer many similar features for preprocessing, dimensionality reduction, clustering, and visualization of scRNA-seq data. Although they are the most used packages for scRNAseq analysis, there are a plethora of other scRNAseq packages that differ in their approach and can affect the accuracy and interpretability of the results. Thus our goal for this project is to compare the accuracy of results from different scRNA-seq data analysis methods, such as Seurat, scVI, constrastiveVI, and linearVI.

1.1 Hypothesized mechanism for mir-190 regulation of cellular diversity

Asymmetric division is a key mechanism for stem cell differentiation. The Drosophila CNS is a well-described model to study the asymmetric division and cell fate selection (Harding and White, 2018). Within the Drosophila CNS are neural stem cells called neuroblasts, which can undergo two types of cell division, symmetrical and asymmetrical. Symmetrical cell division produces two equally partitioned neuroblasts, whereas asymmetrical cell division produces one neuroblast and one daughter cell (Egger et al., 2007). During asymmetric divisions, cell fate determinants are distributed during mitosis based on apical-basal cortical polarity. The apical end establishes a daughter cell that retains self-renewal neuroblast identity while the basal end differentiates the second daughter cell (Homem et al., 2013). Before the final differentiated state, the cell will enter an intermediate state called the ganglion mother cell (GMC). Cell fate determinants such as Prospero, Brat, and Numb must have proper apical-basal placement to get asymmetrically inherited into the daughter cells and give rise to a specific cell specialty. These cell fate determinants are highly conserved and tightly regulated thus dysregulation can result in stem cells dividing symmetrically thus leading to over proliferation of stem cells or differentiated cells.

One such highly conserved cell fate determinant is Prospero, a transcription factor (TF) that acts as an intrinsic signal for the specification of cell fates (Ikeshima-Kataoka et al., 1997). Prospero is an important pan-neural TF that is needed for differentiation of GMC fate and neuron differentiation. Both Prospero mRNA and Prospero protein localize to differentiated cells. After cytokinesis, Prospero transcripts will translocate from the GMC cytoplasm to its nucleus. Low levels of Prospero lead to cells entering dormancy, however loss of Prospero will prevent that dormancy to occur (Lai and Doe, 2014). This role in dormancy hints at the importance of regulating Prospero during cell division. Currently, how Prospero mRNA is regulated in the cell during division and differentiation is unknown.

miRNAs have important roles in organizing asymmetric spatial localization of cell fate determinants in stem cell diversity (Stappert et al., 2015). miRNAs silence genes at the post-transcriptional level and facilitate characterization of various cell types. They inhibit expression of target mRNAs by binding to the 3' untranslated region to induce repression and prevent protein translation (O'Brien et al., 2018). One miRNA in particular is hypothesized to play a large role in determining cell fate of neurons, miRNA-190. Predictive algorithms have identified a conserved binding site across the Drosophila clade between Prospero mRNA and miR-190, suggesting a regulatory interaction. I hypothesize that before NB division, miR-190 could be binding to Prospero mRNA 3' UTR at the apical crescent to restrict Prospero expression to the basal crescent and to initiate the daughter cell into becoming a GMC. Prospero is usually expressed in the GMC in route to becoming a neuron or glial cell (Lai and Doe, 2014). Without a fully functional miR-190 to repress Prospero in the neuroblast, Prospero mRNA could be localized throughout the NB cell. Once the polarity of the cell has been established and Prospero could be localized on both the apical and basal crescent, Prospero may be symmetrically partitioned to both daughter cells. Based on what we know about Prospero, this symmetric partitioning of Prospero is capable of giving rise to two GMC after the division of the neuroblast. Furthermore, to support the hypothesis that miR-190 may be involved in neural stem cell division, RT-qPCR was performed to identify any targeted transcripts associated with miR-190. The preliminary data showed a derepression of Prospero compared to the other targets, ribosomal RNA was used to normalize RNA transcript levels. This indicates that Prospero could be interacting with miR-190.

Single-cell RNA sequencing (scRNAseq) is a dynamic technique to examine transcriptomes in individual cells (Olsen et al., 2018). It is used to characterize and distinguish various cell types and examine cell-specific changes in the transcriptome. scRNAseq is useful in discovering new or rare

cell types, identifying differences in cellular composition between mutant and non-mutant tissue, and understanding cellular differentiation during development (Hwang et. al., 2018). Previous work used scRNAseq with a mir-190 mutant to understand its role in cell fate selection and brain development. The results showed a significant increase in glial cells, neural progenitor cells, mushroom body cells, and undifferentiated cells in mir-190 mutants compared to wildtype. There was a decrease in neuron production suggesting a preference towards regulating self-renewal pathways and glial cell formation. Understanding the role of miRNAs in differentiation versus stemness can eventually aid in the development of stem cell therapies that can be applied to neurodegenerative diseases.

Although interesting results, Seurat has limited tools for a through analysis on mir190 affect on cellular diversity. Other scRNAseq packages can differ in their approach and can affect the accuracy and interpretability of the results. Thus our goal for this project is to compare the accuracy of results from different scRNA-seq data analysis methods, such as Seurat, scVI, constrastiveVI, and linearVI. We aim to determine if mir-190 deficient cells remain to show a decrease in neuron production compared to wildtype with newer and improved scRNAseq analysis.

1.2 Comparing Seurat, scVI and linearVI

Single-cell RNA sequencing (scRNA-seq) is a powerful technique for studying the transcriptomes of individual cells, allowing researchers to investigate the heterogeneity of gene expression within cell populations. There are several methods for analyzing scRNA-seq data, including Seurat, scVI, and linearVI.

Seurat

Seurat uses a novel method called "guided PCA" for clustering scRNA-seq data. This method uses a graph-based approach to identify clusters of cells, allowing for the integration of additional information, such as cell type annotations, into the clustering process.

Single Cell Variational Inference (scVI)

In contrast, scVI uses a probabilistic approach for clustering scRNA-seq data. This method models the gene expression data as a mixture of latent variables, using a variational autoencoder (VAE) to model the gene expression data to allow for the incorporation of uncertainty into the clustering process. This can help to identify cells with uncertain or mixed gene expression patterns, which may not be detected by traditional clustering methods.

Linear Variational Inference (LinearVI)

LinearVI is a linear method for analyzing scRNA-seq data. It uses a linear generative model to learn a low-dimensional representation of the data, which can then be used for downstream analysis. This method is computationally efficient and easy to implement, but may not capture the full complexity of the data compared to non-linear methods like scVI and constrastiveVI.

Overall, scVI, constrastiveVI, and linearVI are all methods for analyzing scRNA-seq data. However, they differ in their approach, with scVI using a probabilistic VAE, constrastiveVI using contrastive learning, and linearVI using a linear generative model. These differences can affect the accuracy and interpretability of the results.

Thus for our project, we are going to compare the accuracy of results from scRNA-seq data analysis methods, such as Seurat, scVI and linearVI using several metrics. One common metric is the adjusted rand index (ARI), which measures the similarity of the predicted clusters to the true labels, if known. A higher ARI indicates better clustering performance. Additionally, the quality of the learned low-dimensional representation of the data can be evaluated using visualization techniques, such as t-SNE or UMAP. These techniques can help to identify any patterns or structures in the data that are captured by the analysis method.

Overall, the accuracy of results from scRNA-seq data analysis methods can be measured using a variety of metrics, including the ARI and visualization techniques for the quality of the learned representation of the data.

2 Methods

2.1 Fly Strains

Drosophila melanogaster Oregon-R was used as the wild type strain. miR-190 mutant stock was obtained from Bloomington Drosophila Stock Center (BDSC): w[*]; TITImir-190[KO]/TM3, Pw[+mC]=GAL4-twi.G2.3, PUAS-2xEGFPAH2.3, Sb[1] Ser[1] All flies were kept at room temperature in vials containing conventional cornmeal agar medium.

2.2 10x Genomic Sequencing

A single-cell suspension concentration of 1000 cells/ul was submitted through UCSF Core Immunology Lab. scRNA-seq libraries were prepared using the Chromium Single Cell 3' Library and Gel Bead Kit v3 (10X Genomics). The pipeline consists of the following: sample preparation, single cell GEM generation using the 10x Single Cell Controller, sample cleanup, reverse transcription, cDNA preparation, sequencing library preparation, and sample sequencing. Sample preparation involved removal of dead cells and viability assessment using a flow cytometer. Sequencing was accomplished with HiSeq 4000 or Novaseq.

2.3 Seurat data processing

Seurat version 4.1.0 (Butler et al., 2018; Satija et al., 2015; Stuart et al., 2018) pipeline was adapted and executed on the normal and miR-190 mutant dataset. Quality control (QC) analysis on the control (WT) and miR-190 mutant datasets determined thresholds for filtered mitocondrial counts and unique features. The data was filtered for cells > 10 percent mitochondrial counts and cells that have unique feature counts between 1000 and 25000. I obtained a total of 11,580 cells; 3,951 wildtype (wt) cells with a median of 600 genes per cell and 7,629 miR-190 mutant cells with a median of 1200 genes per cell (Supplemental Figure 1A). Before integrating datasets, I preprocessed the datasets selecting 2000 features. Next, I merged datasets through the integration method of SCTransform (SCT). Integration among the datasets was performed by identifying common anchors between the two datasets and combining them into a single Seurat object. To define the dimensionality of the dataset, an ElbowPlot was used to determine [1:20] dimensions. This dimensionality was used to find integration anchors. Using 20 dimensions was also used when applying the "FindNeighbors" function, necessary for clustering. Resolution 0.08 was based on the number of identifiable clusters. UMAP was visualized using the reduction "pca". Cluster identities were determined based on previously identified gene markers for each cell type.

To determine if there is a difference in the number of cells in each cell type between WT and miR-190, I counted the number of cells per cluster. Since the number of cells in WT and miR-190 mutant is different (3,951 wildtype cells and 7,629 miR-190 mutant cells), I normalized the data by dividing WT 3951/ miR-190 7629 = 0.518. Each final total cell count of miR-190 clusters were multiplied by the difference (0.518) before graphing results.

2.4 Annotation of Cell Clusters

Cells were characterized using previously known markers and created a dot plot showing the gene expression patterns of each of the known marker genes with our dataset (Figure x). The following markers were confirmed by Avalos et al., 2019: Differentiated neurons could be defined by Glutamic acid decarboxylase 1 (Gad1), an enzyme for the main inhibitory neurotransmitter gamma-Aminobutyric acid (GABA), knot (kn), a transcription factor required for wing disc patterning in larvae, and abnormal chemosensory jump 6 (acj6). Undifferentiated neurons could be defined by headcase (hdc), a known gene for controlling proper timing of neural differentiation and Thor which contributes to translational regulation and cell growth regulation. Peptidergic neurons were determined by target of wit (twit), a regulator of neurotransmitter release during maturation of larval neuromuscular junctions, and dimmed (dimm), a transcription factor that promotes cellular properties for cell fate selection. Neural progenitor cells (NPC) were distinguished by the following: asense (ase), Notch (N), deadpan (D), miranda (mira), and grainyhead (grh). Glial cells were defined by hoepel1 (hoe1), believed to play a crucial role in transmembrane transport, wrapper, important for axon ensheathment and glial apoptosis, and CG6126. Kenyon cells (KC) were characterized by the

expression of the following: short neuropeptide F precursor (sNPF), Neprilysin 1 (Nep1), portabella (prt), Prospero (pros), eyeless (ey).

2.5 Single-Cell Variational Inference (scVI) data processing

scVI version 0.19.0 (Lopez, R., Regier, J., Cole, M.B. et al., 2018) pipeline was adapted and executed on the normal and miR-190 mutant dataset. scVI model runs more efficiently on Python, so the Seurat object from R was converted to an AnnData file to be used in Python. Converting the Seurat object to an AnnData file is a two-step process. First, the Seurat object is saved as an h5Seurat file then convert to an AnnData file for use in Scanpy. Scanpy is a scalable toolkit for analyzing single-cell gene expression data built jointly with AnnData (Wolf, F., Angerer, P. & Theis, F., 2018). AnnData is a Python package for handling annotated data matrices in memory and on disk, positioned between pandas and xarray (Isaac Virshup, Sergei Rybakov, Fabian J. Theis, Philipp Angerer, F. Alexander Wolf, 2021). Next, the data was preprocessed to remove genes that are very lowly expressed, specifically setting the minimum number of counts required for a gene to 3. Then the data was normalized and ran through feature selection to reduce the number of features (genes) down to 1200 highly variable genes. Lastly, AnnData was configured and ran setup_anndata, which alerts scvi-tools to the locations of various matrices inside the anndata. It's important to run this function with the correct arguments so scvi-tools is notified that your dataset has batches, annotations, etc. I create the scVI model with the following parameters:

n_hidden: 128, n_latent: 10, n_layers: 1, dropout_rate: 0.1, dispersion: gene, gene_likelihood: zinb, latent_distribution: normal

The model was trained across 400 epochs. For running the UMAP visualization without batch correction, I computed a neighborhood graph of observations with 20 principal components and 20 neighboring data points. I embed the neighborhood graph using UMAP with 0.3 as the effective minimum distance between embedded points. Using this embedding, I created a UMAP for cell type, control/mutant, and mitochondrial gene percentage annotations. The same process for determining the difference in the number of cells in each cell type between WT and miR-190 as the Seurat data processing was used.

Finally, scVI latent space was used to generate the same UMAP plots to observe scVI successfully accounts for batch effects in the data, and followed with the same process for cell counts as above.

We can cluster cells using the Leiden algorithm [Traag, 2018], an improved version of the Louvain algorithm [Blondel, 2008]. It has been proposed for single-cell analysis by [Levine, 2015], using 0.09 as the resolution to maintain seven clusters for the seven cell types.

Using the marker genes determined in the Annotation of Cell Cluster section, I created a dot plot and heat map showing the gene expression patterns of each of the known marker genes with out data set.

2.6 Linearly-decoded VAE scVI data processing

The scVI model learns low-dimensional latent representations of cells which get mapped to parameters of probability distributions which can generate counts consistent to what is observed from data. In the standard version of scVI these parameters for each gene and cell arise from applying neural networks to the latent variables. Neural networks are flexible and can represent non-linearities in the data. This comes at a price, there is no direct link between a latent variable dimension and any potential set of genes which would covary across it. The LDVAE model replaces the neural networks with linear functions. Now a higher value along a latent dimension will directly correspond to higher expression of the genes with high weights assigned to that dimension. This leads to a generative model comparable to probabilistic PCA or factor analysis, but generates counts rather than real numbers.

I use the same scVI version, the same process for converting the Seurat object into an AnnData file, and the same preprocessing steps as for scVI data processing described above. I initialized the LinearSCVI model with 20 latent dimensions with the following parameters:

n_hidden: 128, n_latent: 20, n_layers: 1, dropout_rate: 0.1, dispersion: gene, gene_likelihood: nb, latent_distribution: normal

I trained the model for 250 epochs. I created a UMAP for the cell clustering based on the Leiden algorithm to visualize the LinearSCVI latent space, similar to what was described earlier.

The same process for determining the difference in the number of cells in each cell type between WT and miR-190 as above was used.

3 Results

3.1 Differences in cell counts among Seurat, scVI, and linear VI

After clustering the data with each scRNAseq package, we counted the number of cells in each cell type between the control and mir-190. The dataset for mir-190 was adjusted for the excess amount of cell sequence compared to the control (see methods section for details the data normalization). Using Seurat, the cell counts of mir-190 deficient cells show a decrease in neuron production and an increase in differentiating cells, glial cells and kenyon cells compared to the control (Figure 2A). Although the cell counts were slightly different when processed with scVI, we were able to see the same general trend with a decrease in neuron production and an increase in differentiating cells, glial cells and kenyon cells (Figure 2B). However, when preforming cell counts with linearVI, our cell counts show a decrease in Kenyon cells, opposite of the cell counts and thus clustering of seurat and scVI (Figure 2C).

Among all three analysis, we are still able to confer that miR-190 may play a role in regulating an appropriate balance of pan-neuronal cells and glial cells. Without miR-190, neuronal differentiation could be difficult and remain as NPC or differentiate into glial cells. However, we need to further examine the discrepancy between the packages with Kenyon cells.

3.2 Evaluation of Method Clustering

We need to evaluate the clustering of the two clustering methods compared with Seurat to see how similar each of the clusterings are with one another. I decided to use the adjusted rand index from the Scikit-learn Python library. Rand Index is a function that computes a similarity measure between two clustering. For this computation rand index considers all pairs of samples and counting pairs that are assigned in the similar or different clusters in the predicted and true clustering. Afterwards, the raw Rand Index score is 'adjusted for chance' into the Adjusted Rand Index score by using the following formula:

$$Adjusted\_RI = \frac{RI - Expected\_RI}{max(RI) - Expected\_RI} \tag{1}$$

It has two parameters namely labels_true, which is ground truth class labels, which we will use the Seurat clustering for, and labels_pred, which are clusters label to evaluate, which are the two scVI clusterings.

The ARI lies between 0 and 1, with higher scores indicating better separation. Random labelings have an ARI close to 0.0. 1.0 stands for perfect match.

A

Clustering comparison between methods (ARI)

	Seurat	scVI	LD-SCVI
Seurat	1	0.47059	0.91837
scVI	0.47059	1	0.46785
LD-SCVI	0.91837	0.46785	1

Using the adjusted rand index, we notice that LinearSCVI clustering was able to match the Seurat clustering well, but scVI couldn't match as well. This could be due to the linear decoder being able to explicitly link latent cells to genes, whereas there isn't a direct link between a latent variable dimension and any potential set of genes in the standard scVI.

4 Conclusion

Overall, comparing Seurat, scVI, and linearVI, there are slight differences between data analysis and conclusions. Seurat and scVI had slightly different results when counting the number of cells in each cell type, however, showed similar trends, thus we were able to make the same conclusions with both analysis. However, linearVI had different cell count for Kenyon cells. Either better characterization of the cell type needs to occur or we need to reassess how the data is being clustered. Interestingly enough, when running a Rand Index, we see that the clustering similarity is higher between seurat and linearVI than seurat and scVI. This is surprisingly considering Seurat and scVI had more similar cell counts. Further analysis into how the packages are processing the data as well as user cross examination would need to take place to conclude which package is giving us more accurate results. Overall though, using all three analysis, we are still able to confer that miR-190 may play a role in regulating an appropriate balance of pan-neuronal cells and glial cells. Without miR-190, neuronal differentiation could be difficult and remain as NPC or differentiate into glial cells.

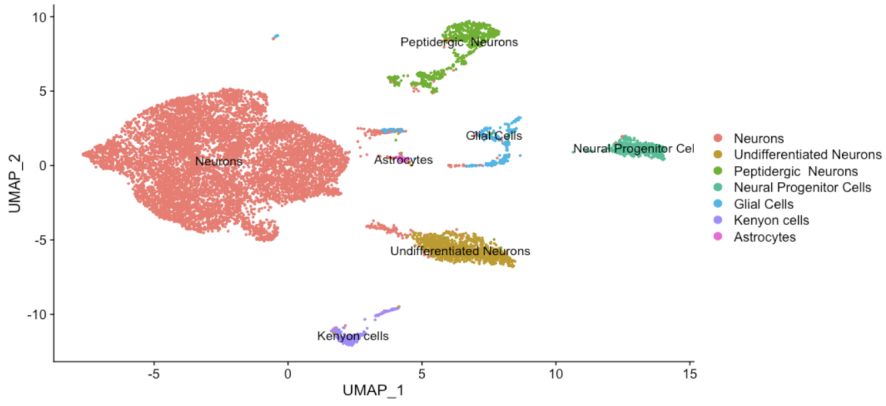
5 Discussion

Future directions would be to double check how the data is being processed and clustered among all three packages. Contacting the package developers is also an option is understand how each package may be concluding different cell counts for kenyon cells. There are other scRNAseq packages such as Monocle and Scanpy as well that we could quickly run the data through and see if we get the same cell counts although they are very similar to Seurat.

5.1 Figures

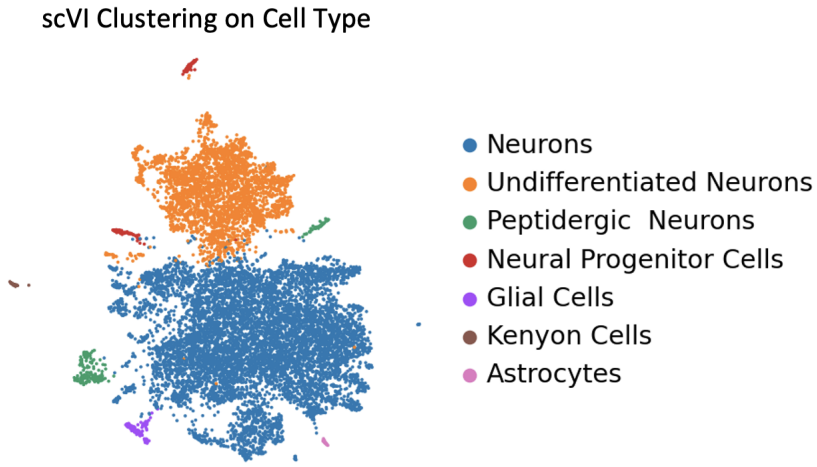
A

Seurat



B

scVI



C

Linear-decoding scVI

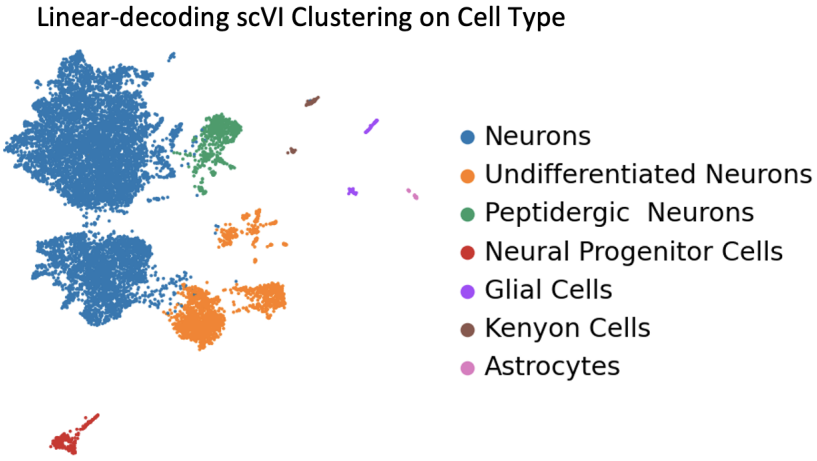
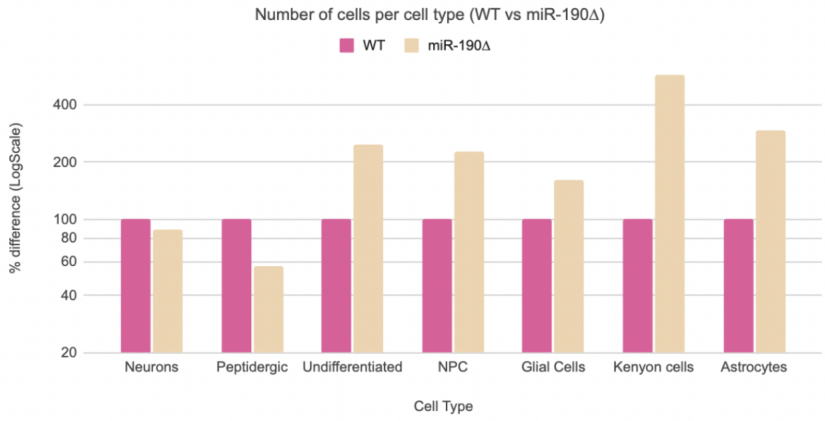


Figure 2

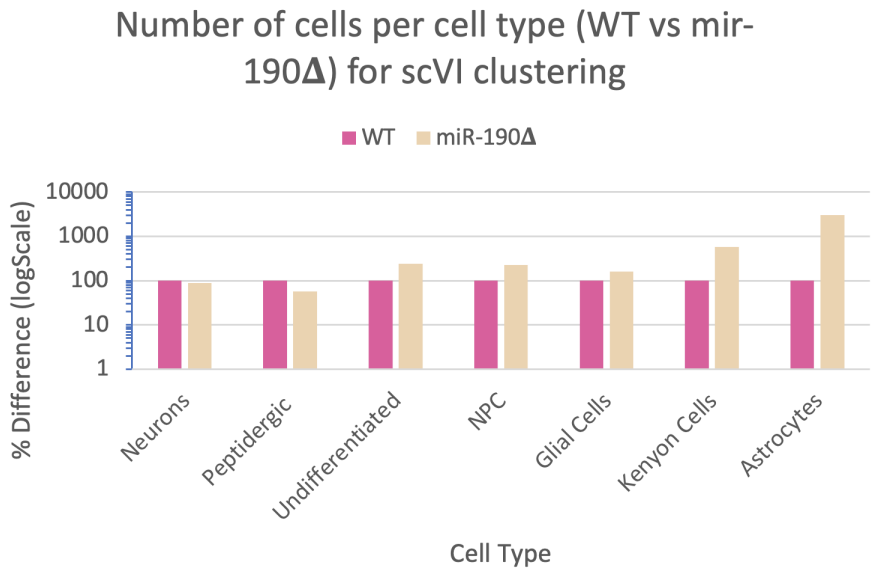
A

Seurat



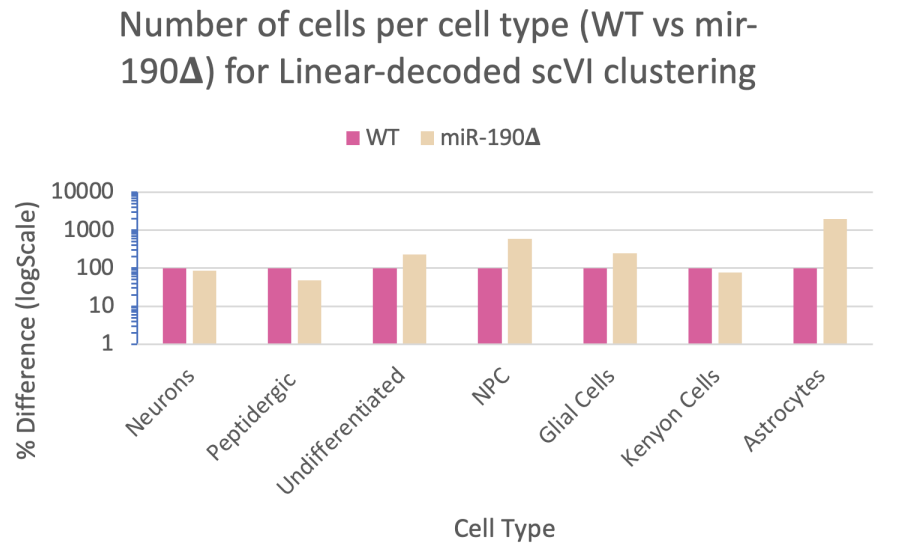
B

scVI



C

Linear-decoding scVI



References

Ahn, H. J., Jeon, S. H., Kim, S. H. (2014). Expression of a set of glial cell-specific markers in the Drosophila embryonic central nervous system. BMB reports, 47(6), 354–359. https://doi.org/10.5483/bmbrep.2014.47.6.177

Allen, A. M., Neville, M. C., Birtles, S., Croset, V., Treiber, C. D., Waddell, S., Goodwin, S. F. (2020). A single-cell transcriptomic atlas of the adult Drosophila ventral nerve cord. eLife, 9, e54074. https://doi.org/10.7554/eLife.54074

Brunet Avalos, C., Maier, G. L., Bruggmann, R., Sprecher, S. G. (2019). Single cell transcriptome atlas of the Drosophila larval brain. eLife, 8, e50354. https://doi.org/10.7554/eLife.50354

Davie, K., Janssens, J., Koldere, D., De Waegeneer, M., Pech, U., Kreft, Ł., Aibar, S., Makhzami, S., Christiaens, V., Bravo González-Blas, C., Poovathingal, S., Hulselmans, G., Spanier, K. I., Moerman, T., Vanspauwen, B., Geurs, S., Voet, T., Lammertyn, J., Thienpont, B., Liu, S., … Aerts, S. (2018). A Single-Cell Transcriptome Atlas of the Aging Drosophila Brain. Cell, 174(4), 982–998.e20. https://doi.org/10.1016/j.cell.2018.05.057

Hatch, H., Belalcazar, H. M., Marshall, O. J., Secombe, J. (2021). A KDM5-Prospero transcriptional axis functions during early neurodevelopment to regulate mushroom body formation. eLife, 10, e63886. https://doi.org/10.7554/eLife.63886

Harding, K., White, K. (2018). Drosophila as a Model for Developmental Biology: Stem Cell-Fate Decisions in the Developing Nervous System. Journal of developmental biology, 6(4), 25. https://doi.org/10.3390/jdb6040025

Huang, S., Hua, X., Kuang, M. et al. miR-190 promotes malignant transformation and progression of human urothelial cells through CDKN1B/p27 inhibition. Cancer Cell Int21, 241 (2021). https://doi.org/10.1186/s12935-021-01937-5

Hwang, B., Lee, J.H. Bang, D. Single-cell RNA sequencing technologies and bioinformatics pipelines. Exp Mol Med 50, 1–14 (2018). https://doi.org/10.1038/s12276-018-0071-8

Kato, K., Forero, M. G., Fenton, J. C., Hidalgo, A. (2011). The glial regenerative response to central nervous system injury is enabled by pros-notch and pros-NFB feedback. PLoS biology, 9(8), e1001133. https://doi.org/10.1371/journal.pbio.1001133

Kim NC, Marqués G. The Ly6 neurotoxin-like molecule target of wit regulates spontaneous neurotransmitter release at the developing neuromuscular junction in Drosophila. Dev Neurobiol. 2012 Dec;72(12):1541-58. doi: 10.1002/dneu.22021. Epub 2012 Jul 27. PMID: 22467519.

Kunz T, Kraft KF, Technau GM, Urbach R. Origin of Drosophila mushroom body neuroblasts and generation of divergent embryonic lineages. Development. 2012 Jul;139(14):2510-22. doi: 10.1242/dev.077883. Epub 2012 Jun 6. PMID: 22675205.

Lai, S. L., Doe, C. Q. (2014). Transient nuclear Prospero induces neural progenitor quiescence. eLife, 3, e03363. https://doi.org/10.7554/eLife.03363

Lee, M. H., Salvaterra, P. M. (2002). Abnormal chemosensory jump 6 is a positive transcriptional regulator of the cholinergic gene locus in Drosophila olfactory neurons. The Journal of neuroscience : the official journal of the Society for Neuroscience, 22(13), 5291–5299. https://doi.org/10.1523/JNEUROSCI.22-13-05291.2002

Li, H., Horns, F., Wu, B., Xie, Q., Li, J., Li, T., Luginbuhl, D. J., Quake, S. R., Luo, L. (2017). Classifying Drosophila Olfactory Projection Neuron Subtypes by Single-Cell RNA Sequencing. Cell, 171(5), 1206–1220.e22. https://doi.org/10.1016/j.cell.2017.10.019

Merrill, C.B., Pabon, M.A., Montgomery, A.B. et al. Optimized assay for transposase-accessible chromatin by sequencing (ATAC-seq) library preparation from adult Drosophila melanogaster neurons. Sci Rep 12, 6043 (2022). https://doi.org/10.1038/s41598-022-09869-4

Mohler J, Seecoomar M, Agarwal S, Bier E, Hsai J. Activation of knot (kn) specifies the 3-4 intervein region in the Drosophila wing. Development. 2000 Jan;127(1):55-63. doi: 10.1242/dev.127.1.55. PMID: 10654600.

O'Brien J, Hayder H, Zayed Y and Peng C (2018) Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. Front. Endocrinol. 9:402. doi: 10.3389/fendo.2018.00402

Olsen TK, Baryawno N. Introduction to Single-Cell RNA Sequencing. Curr Protoc Mol Biol. 2018 Apr;122(1):e57. doi: 10.1002/cpmb.57. PMID: 29851283.

Park, D., Shafer, O. T., Shepherd, S. P., Suh, H., Trigg, J. S., Taghert, P. H. (2008). The Drosophila basic helix-loop-helix protein DIMMED directly activates PHM, a gene encoding a neuropeptide-amidating enzyme. Molecular and cellular biology, 28(1), 410–421. https://doi.org/10.1128/MCB.01104-07

Squair, J.W., Gautier, M., Kathe, C. et al. Confronting false discoveries in single-cell differential expression. Nat Commun 12, 5692 (2021). https://doi.org/10.1038/s41467-021-25960-2

Stappert, L., Roese-Koerner, B., Brüstle, O. (2015). The role of microRNAs in human neural stem cells, neuronal differentiation and subtype specification. Cell and tissue research, 359(1), 47–64. https://doi.org/10.1007/s00441-014-1981-y