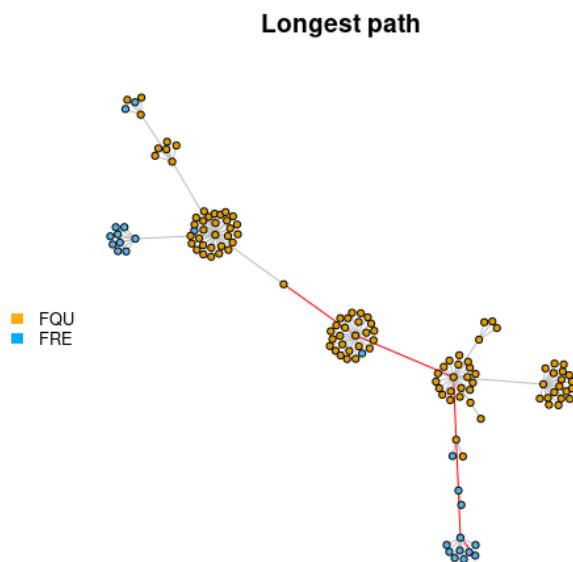
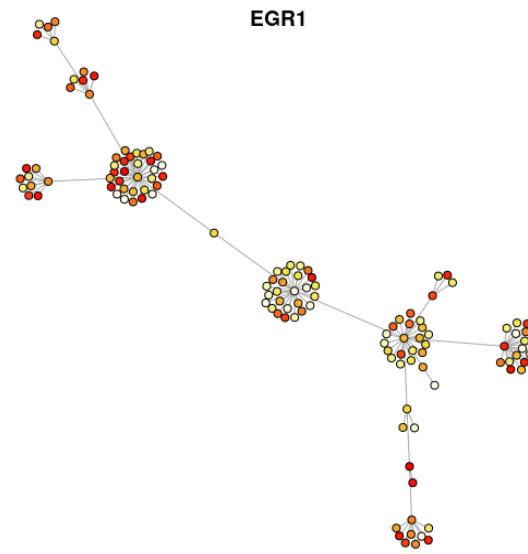


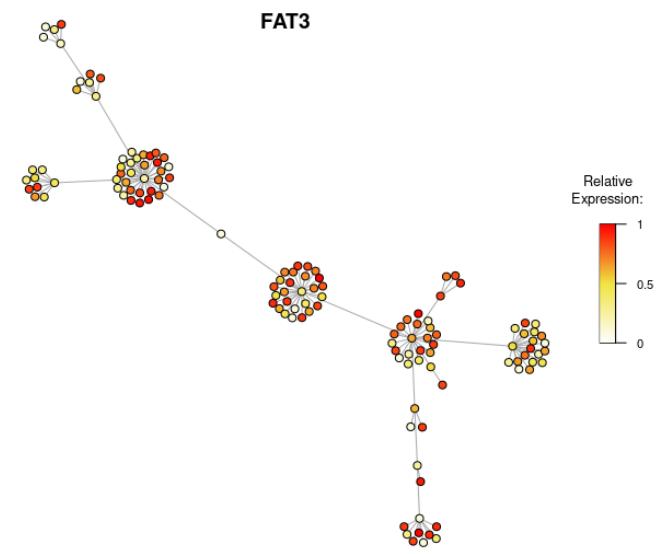
[A]



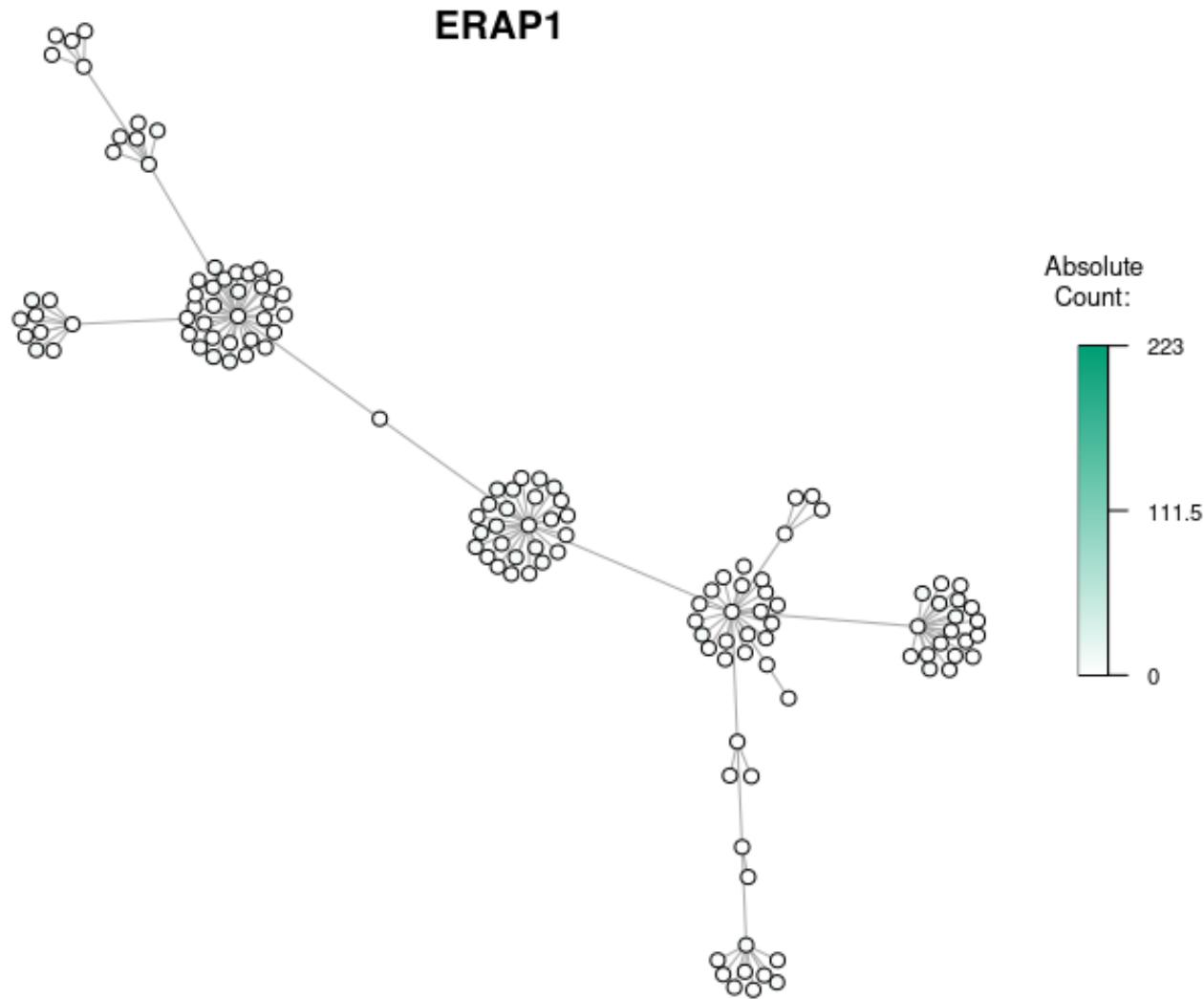
[B]



[C]

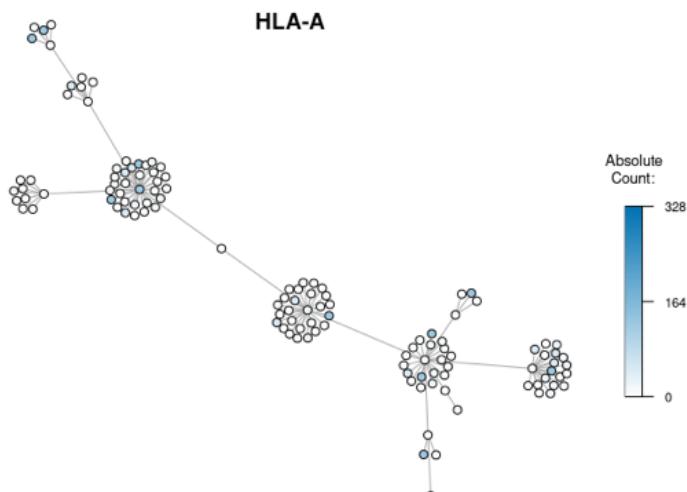


**Figure 1: Expression levels of FAT3 and EGR1 in fetal cells.** After generating a minimum spanning tree from the pairwise distances between the fetal cells and [A] looking at the longest path across the tree, two genes identified with a high correlation coefficient being [B] FAT atypical cadherin 3 (FAT3) and [C] early growth response 1 (EGR1). The key representing relative expression levels on the right correlates with the colour of the nodes. Each node is coloured according to its expression level relative to the other nodes in the plot. Nodes with a higher expression level are coloured red or orange, and those with medium expression levels are coloured yellow, and those with low expression levels are coloured white.

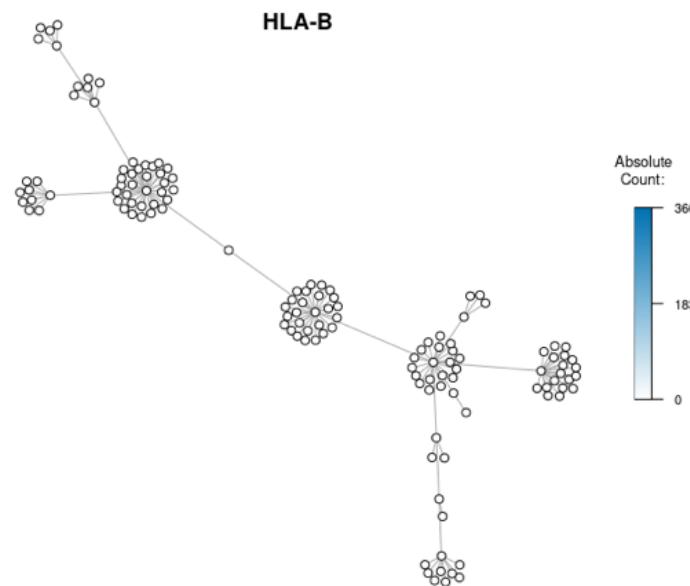


**Figure 2: Minimum spanning tree reflecting ERAP1's expression levels.** In this figure, the minimum spanning tree nodes are coloured according to their expression levels of the gene in fetal cells. The fetal cells were shown to lack the expression ERAP1, one of the 14 genes associated with MHC I. The absolute count represents the number of reads of the ERAP1 transcripts in the samples. The colour scale indicates and reflects the absolute count of ERAP1.

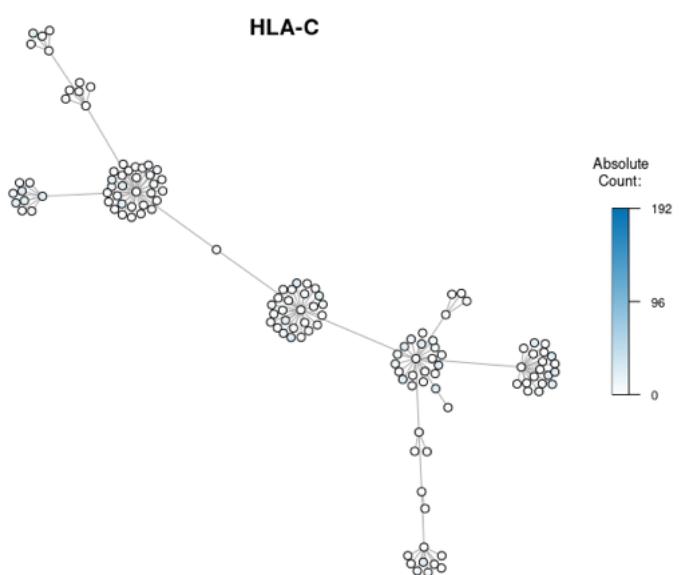
[A]



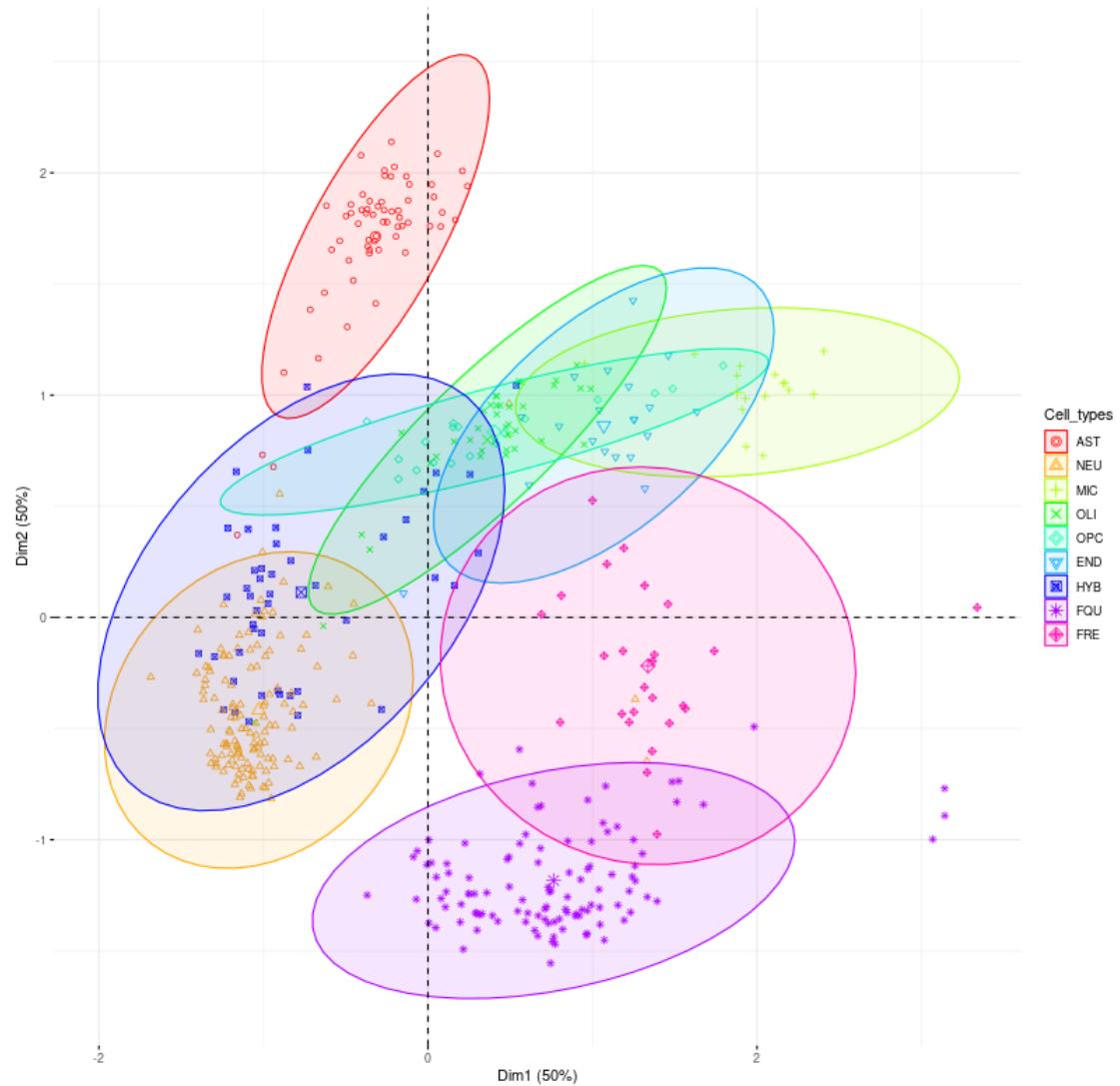
[B]



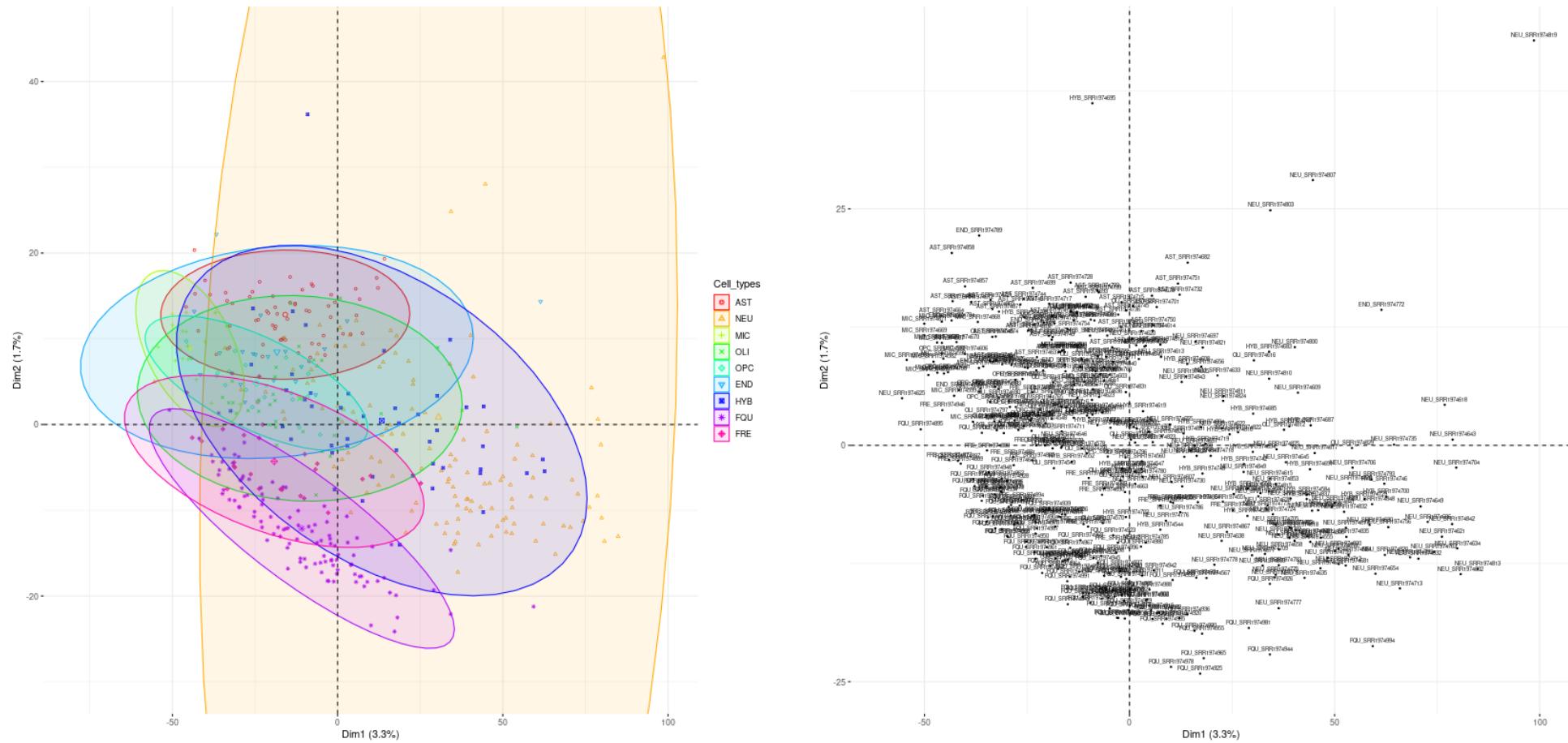
[C]



**Figure 3: Minimum spanning trees showing the fetal cells population's relative expression of HLA-A, -B, and -C.** The fetal cell populations showed a lack of expression of [A] HLA-A, [B] -B and [C] -C. The absolute count represents the number of reads of the genes in the samples. The colour scale indicates and reflects the absolute count, all of which showed none.

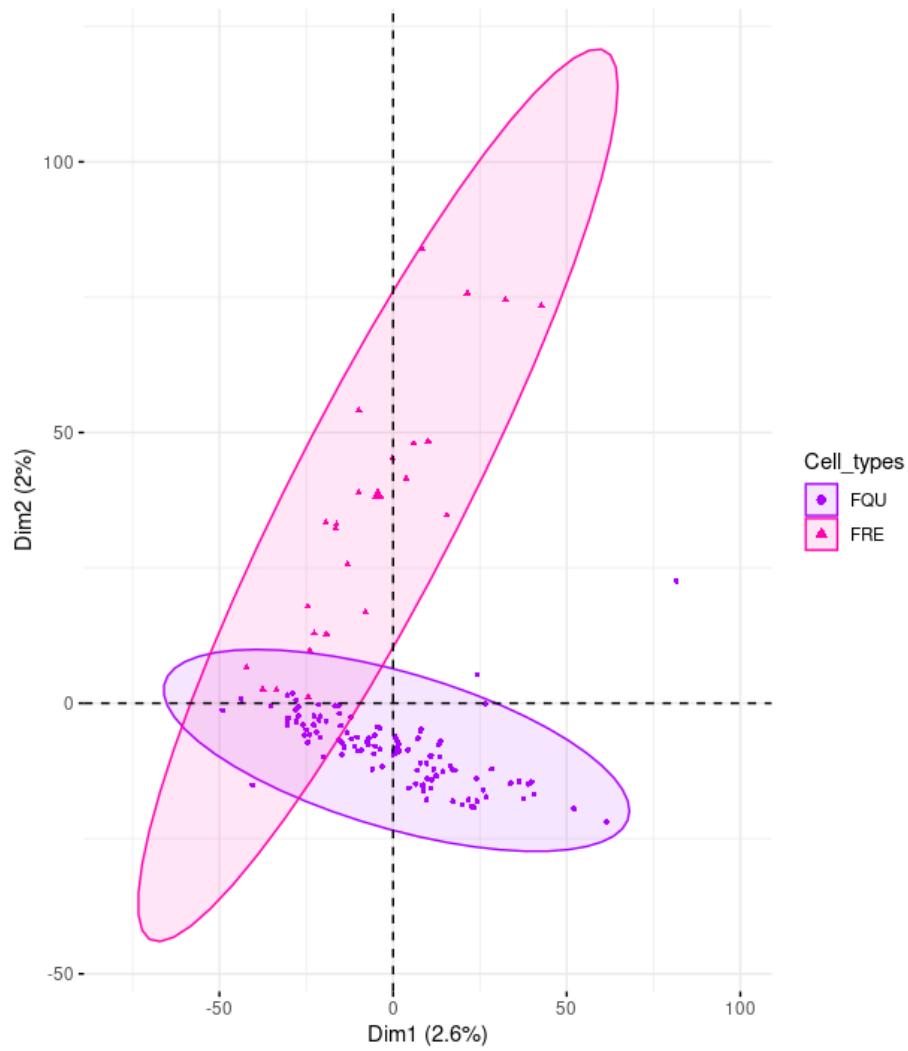


**Figure 4: Scatterplot showing the results of the PCA analysis.** Multidimensional scaling (MDS) and Principal Component Analysis (PCA) are performed on the distance matrix. AST, NEU, MIC, OLI, OPC, END, HYB, FQU, and FRE are abbreviations for different cell types and are colour co-ordinated accordingly. They stand for astrocytes, neurons, microglia, oligodendrocytes, oligodendrocyte progenitor cells, endothelial cells, hybrid cells, fetal quiescent cells, and adult quiescent cells, respectively. In the plot, each dot represents a cell, and the colour of the dot indicates its cell type. Each axis indicates how much variation of each of the dataset is represented by each of the dimensions. Together, the two dimensions explain 50% of the total variation in the dataset. The astrocyte group indicates a unique gene expression pattern compared to the others, as it is located slightly away from the rest of the group. In the PCA space, AST is not seen to overlap the centroid point for the group.

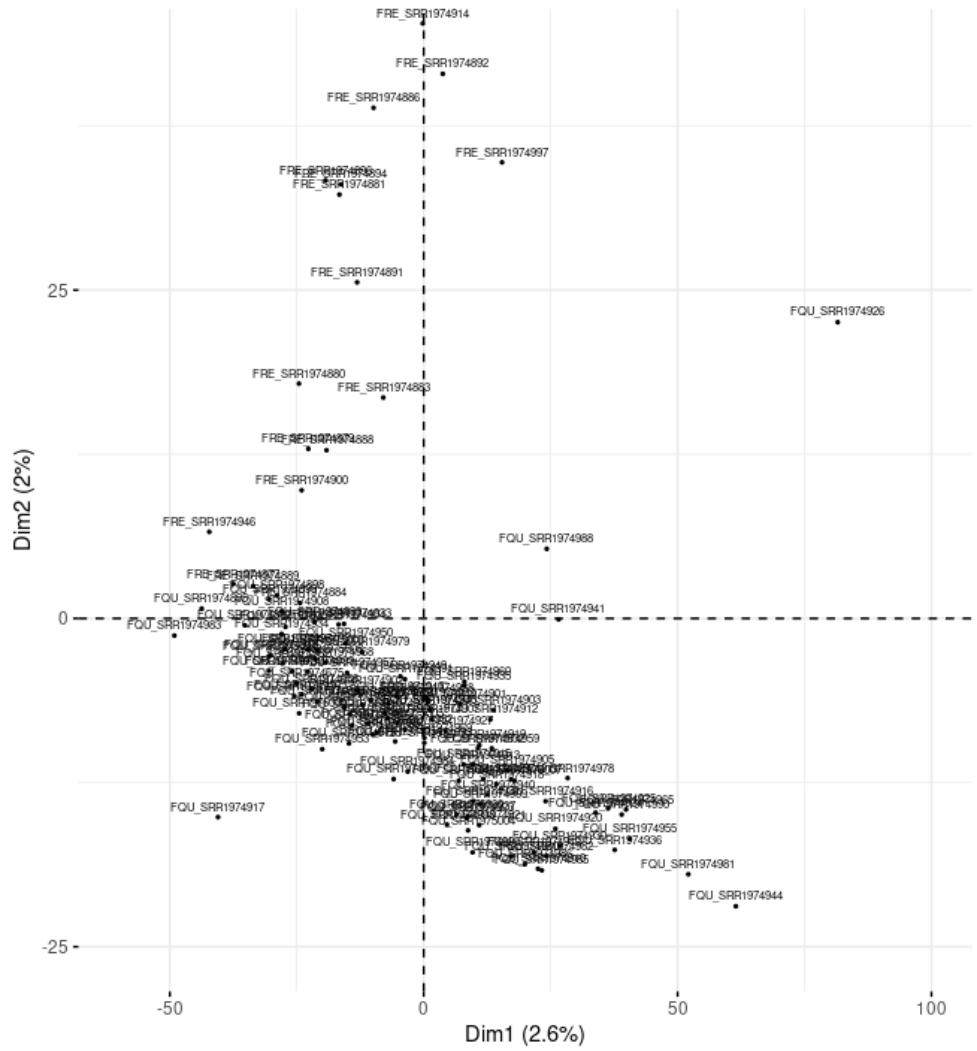


**Figure 5: PCA analysis plot of gene expression profiles from different cell types in the dataset.** The X- and Y-axes represent the two principal components obtained through PCA. Dimension 1 on the X-axis accounts for 3.3% of the total variance in the data, whereas dimension 2 on the Y-axis accounts for 1.7% of the total variance in the data. Each axis indicates how much variation of each of the datasets is represented by each of the dimensions. Each of the different cell types is seen to overlap the centroid. The data was filtered to filter out genes with low expression in fewer than 5 cells. Each point represents an individual cell, coloured according to its cell type (AST: astrocyte, NEU: neuron, MIC: microglia, OLI: oligodendrocyte, OPC: oligodendrocyte precursor cell, END: endothelial cell, HYB: hybrid cell, FOU: fibroblast/quiescent cell, and FRE: fibroblast/replicating cell) as shown on the cell type key. Both plots show the same results on gene expression data, they differ in the way the data is presented with the right plot labelling the plot with its cell type rather than colour coordinating it. The plots show the distribution of the cells in the first two principal components.

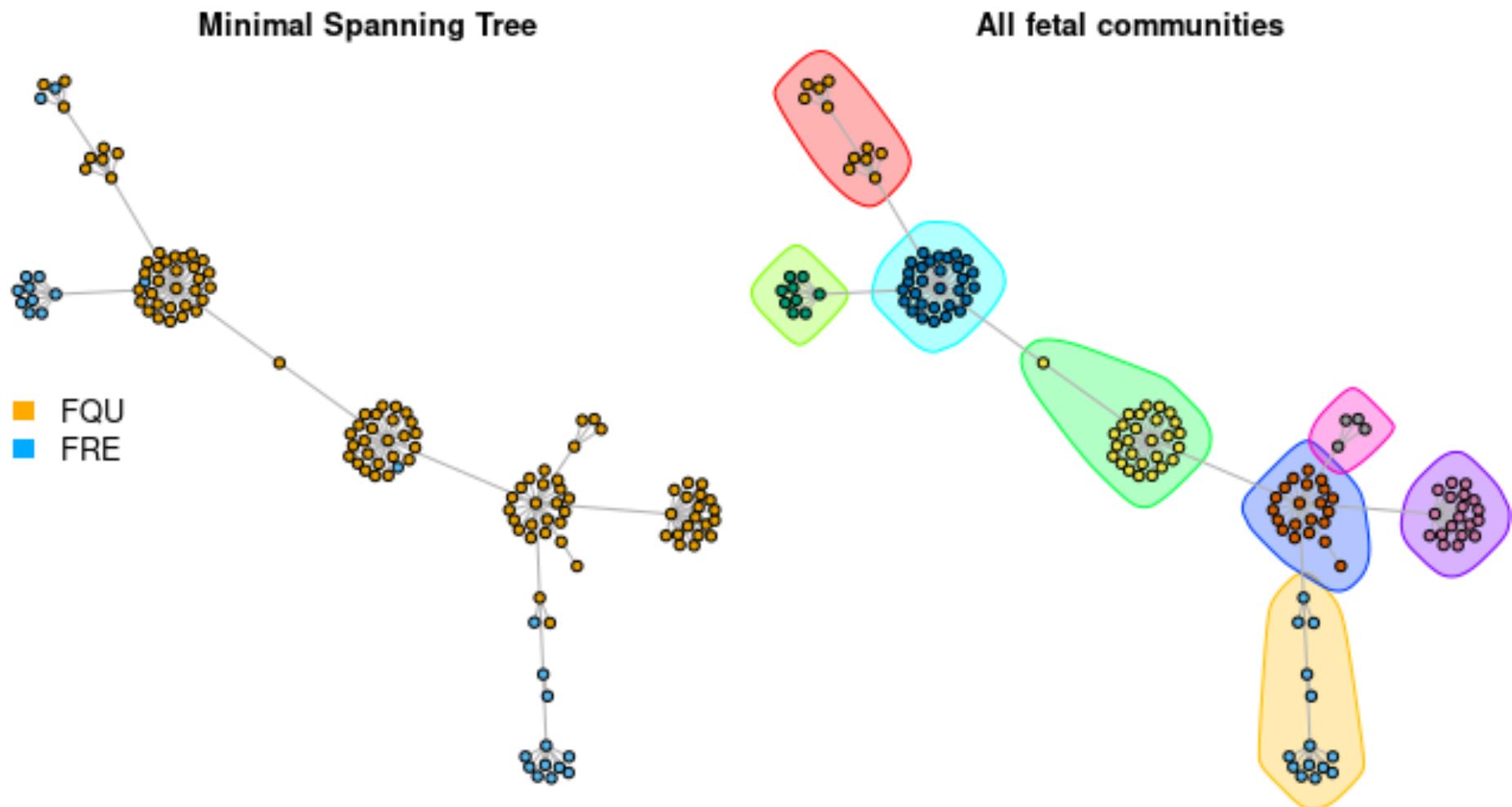
Individuals - PCA



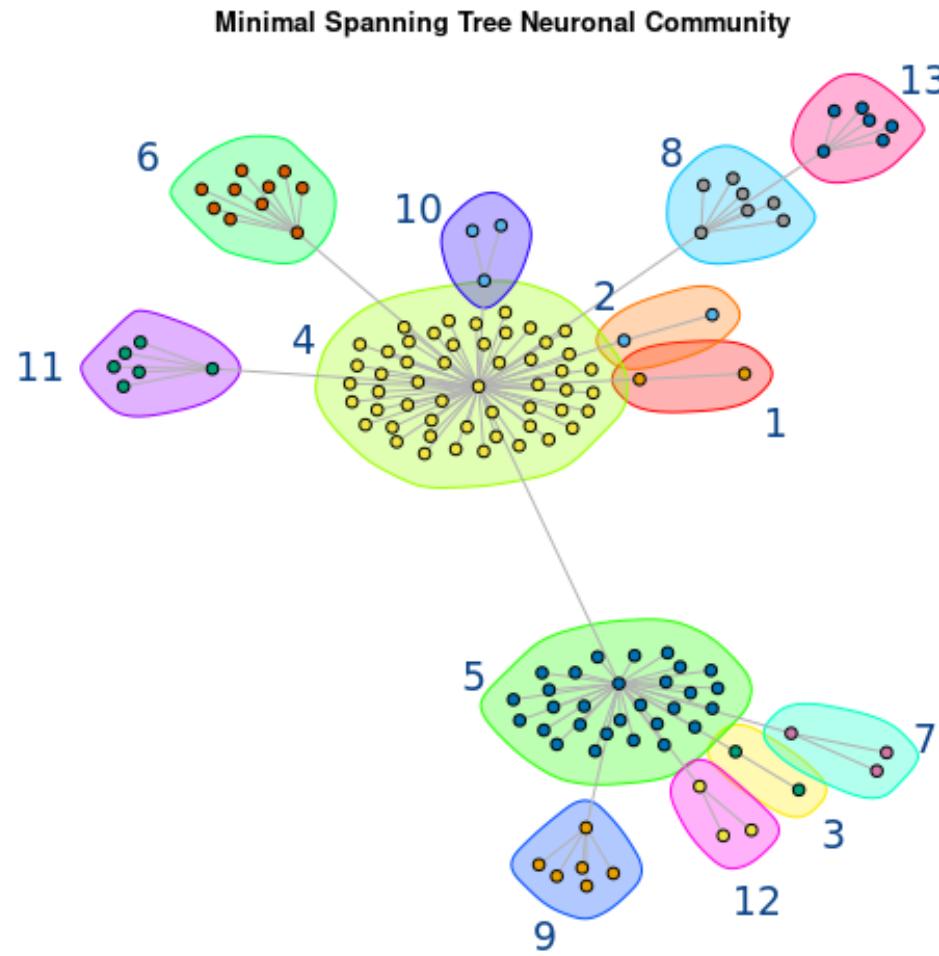
Individuals - PCA



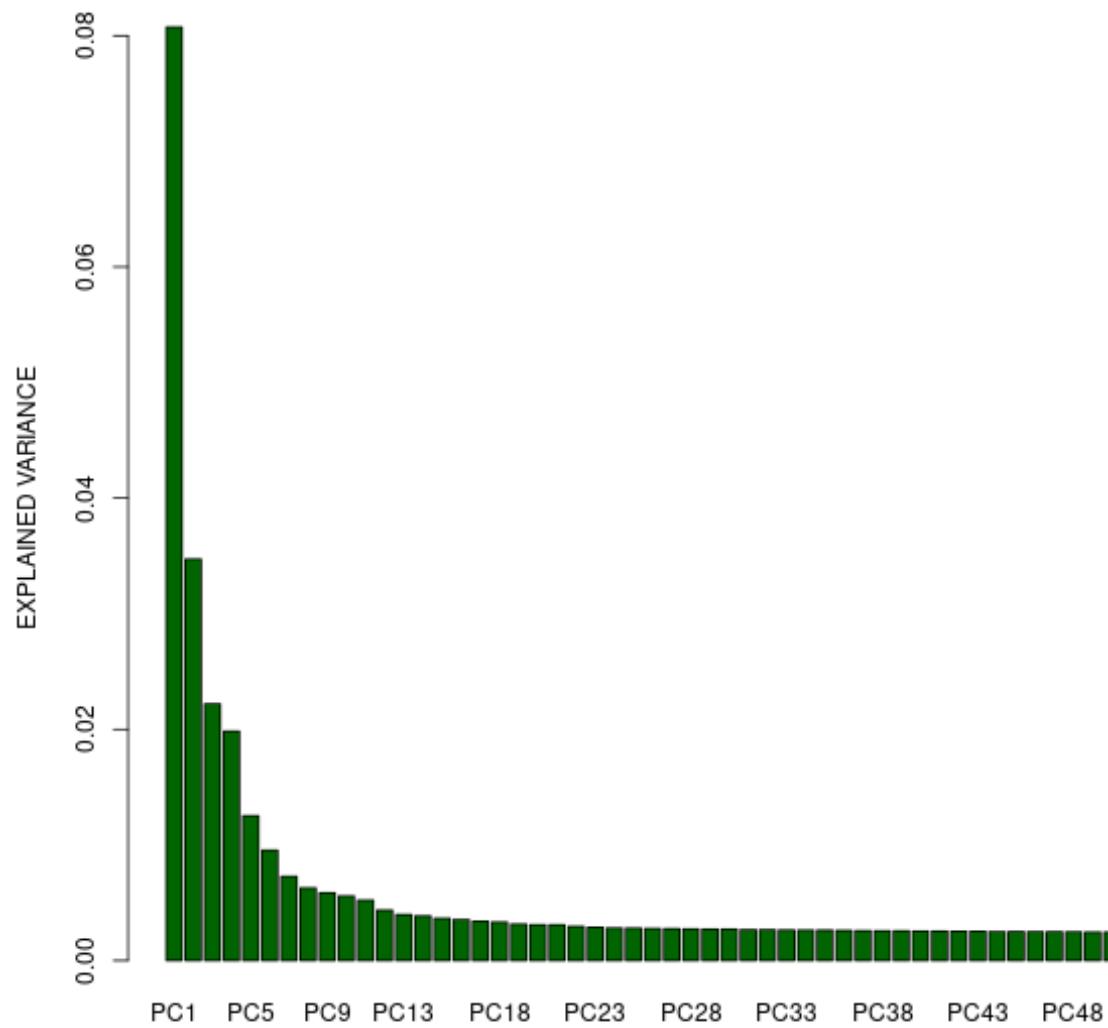
**Figure 6: PCA analysis plot of gene expression profiles of two types of fibroblasts, FQU and FRE.** In the figure, the plots show the distribution of the cells in the first two principal components. The X- and Y- axes represent the two principal components obtained through PCA. Dimension 1 on the X-axis accounts for 2.6% of the total variance in the data, whereas dimension 2 on the Y-axis accounts for 2.0% of the total variance in the data. Each axis indicates how much variation of each of the datasets is represented by each of the dimensions. . Each point represents an individual cell, coloured according to its cell type, either FQU (fibroblast/quiescent cell) in purple or FRE (fibroblast/replicating cell) in pink. . Both plots show the same results on gene expression data, they differ in the way the data is presented with the right plot labelling the plot with its cell type rather than colour coordinating it.



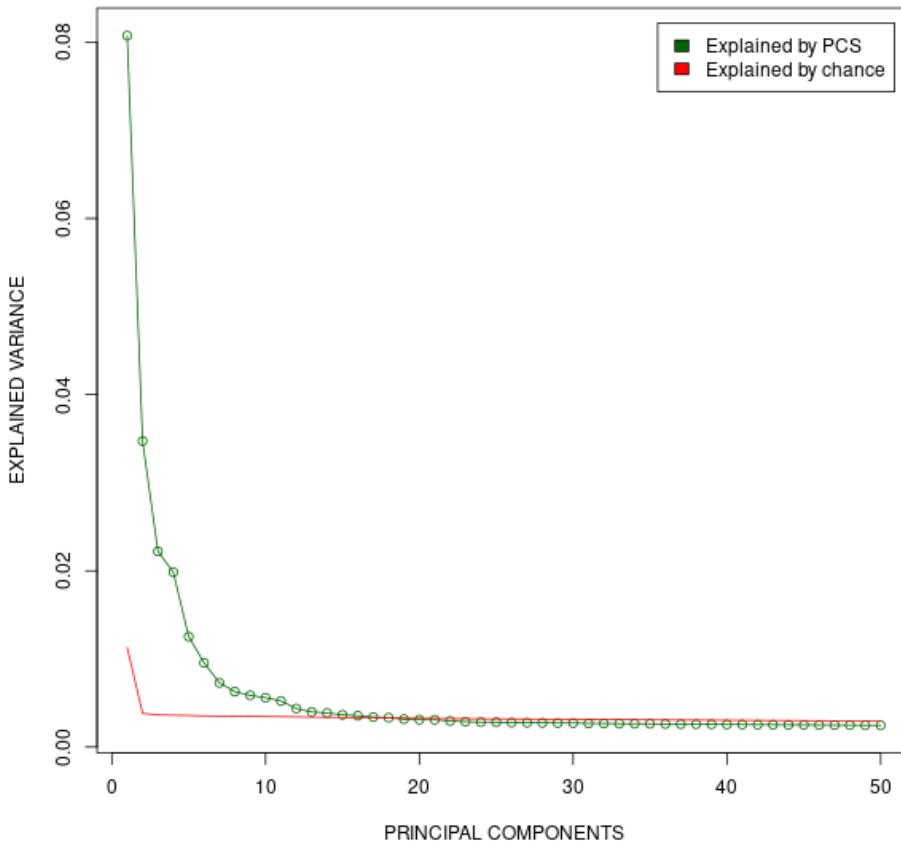
**Figure 7: Minimum spanning tree for all fetal cells.** The left of the figure shows the minimum spanning trees generated for all the fetal cells. All fetal communities were identified in the multiple spanning trees based on 'edge between-ness'. The 'edge between-ness' function measures the number of shortest paths that pass a given edge, allowing the identification of edges that are likely to connect different groups of nodes. The different colours correspond to the different fetal communities that were identified using this algorithm.



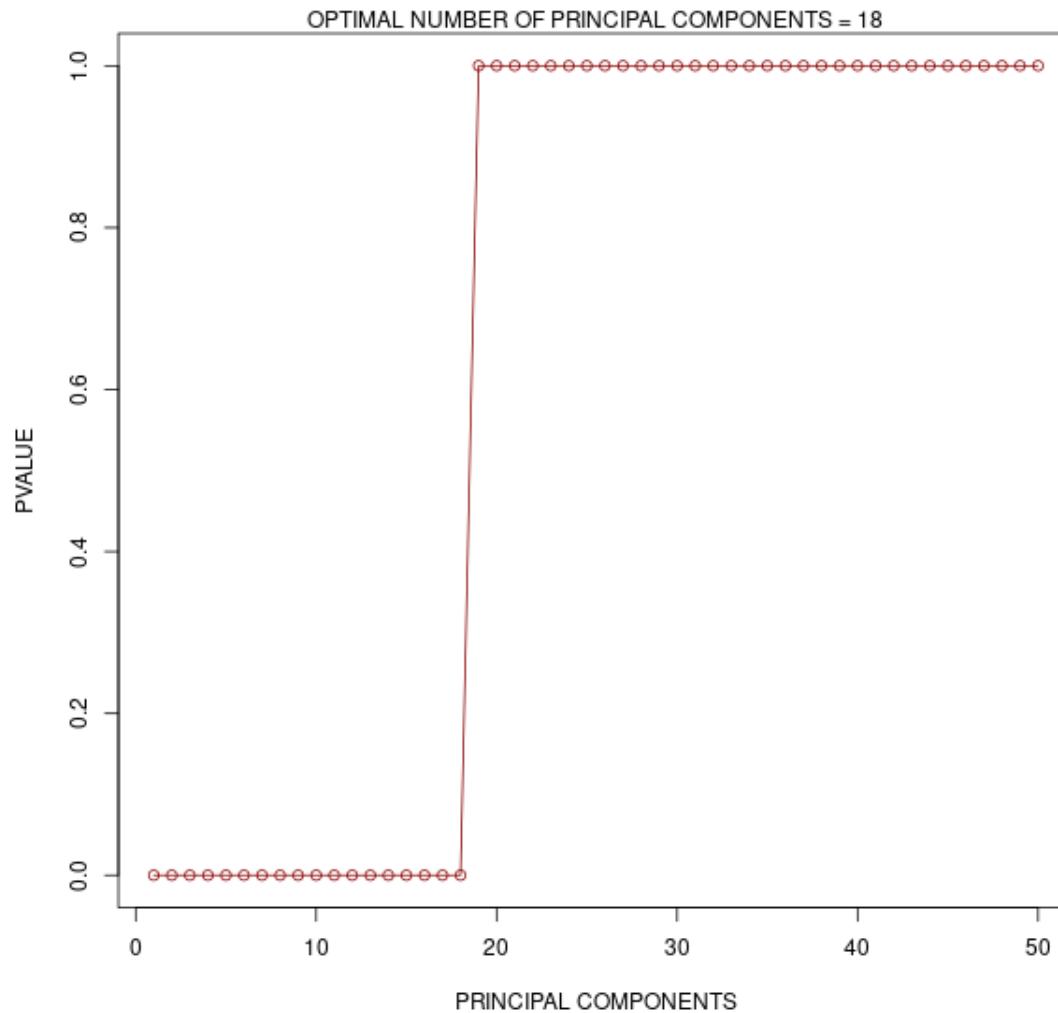
**Figure 8: Minimum spanning tree for all neuronal cells.** Colours indicates the different neuronal communities identified by the minimum-spanning tree.



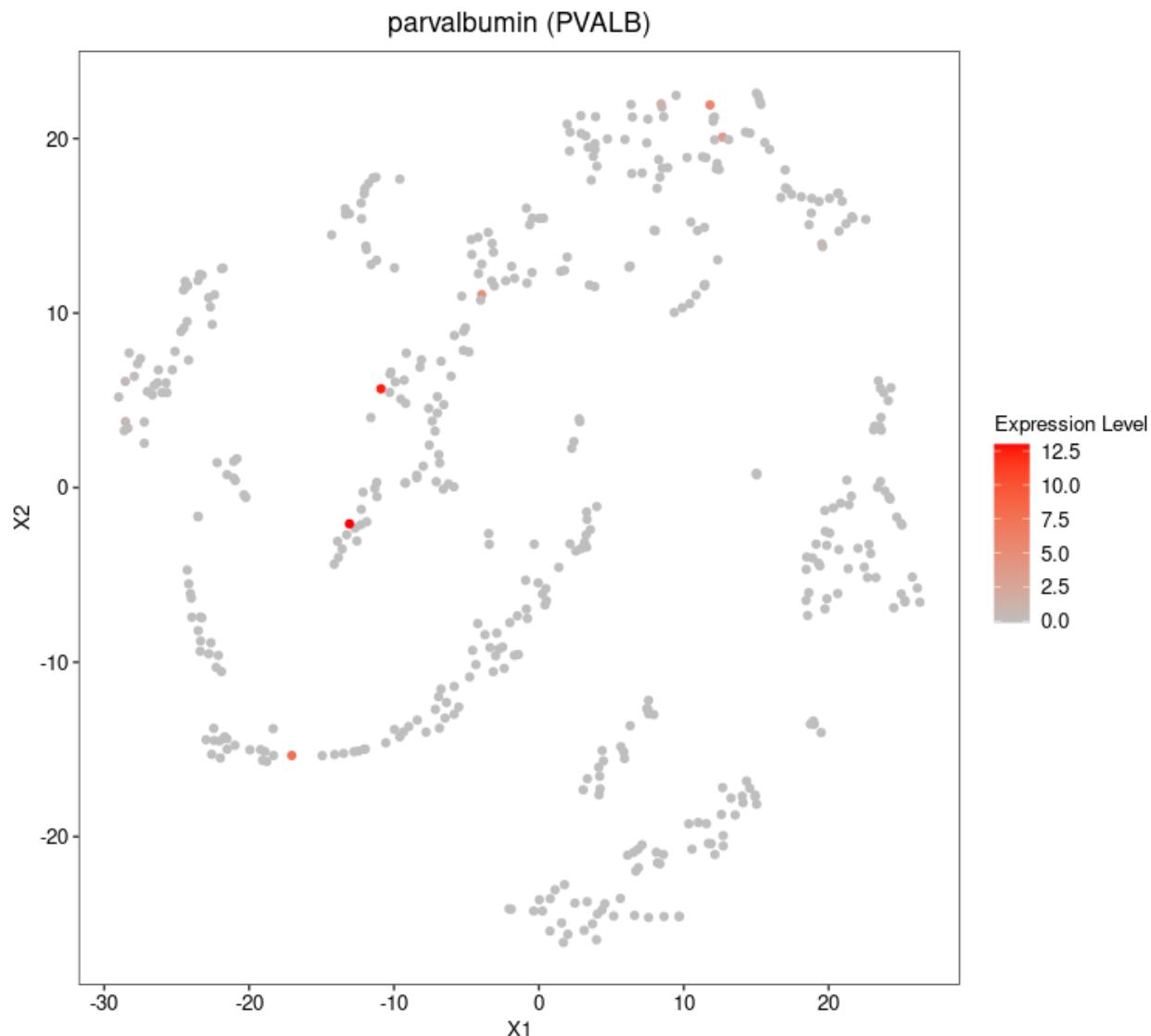
**Figure 9: Bar chart showing the explained variance of each principal component.** After calculating the principal components and the explained variance of each principal component. This bar chart shows the variance of the first 50 principal components. The calculation identifies the optimal number of principal components to use for initial dimensional reduction for the t-SNE analysis. It is important to identify the correct number of principal components as not enough may see the loss of important data and too many may lead to overfitting. This bar chart helps identify the correct number of principle components needed for the analysis.



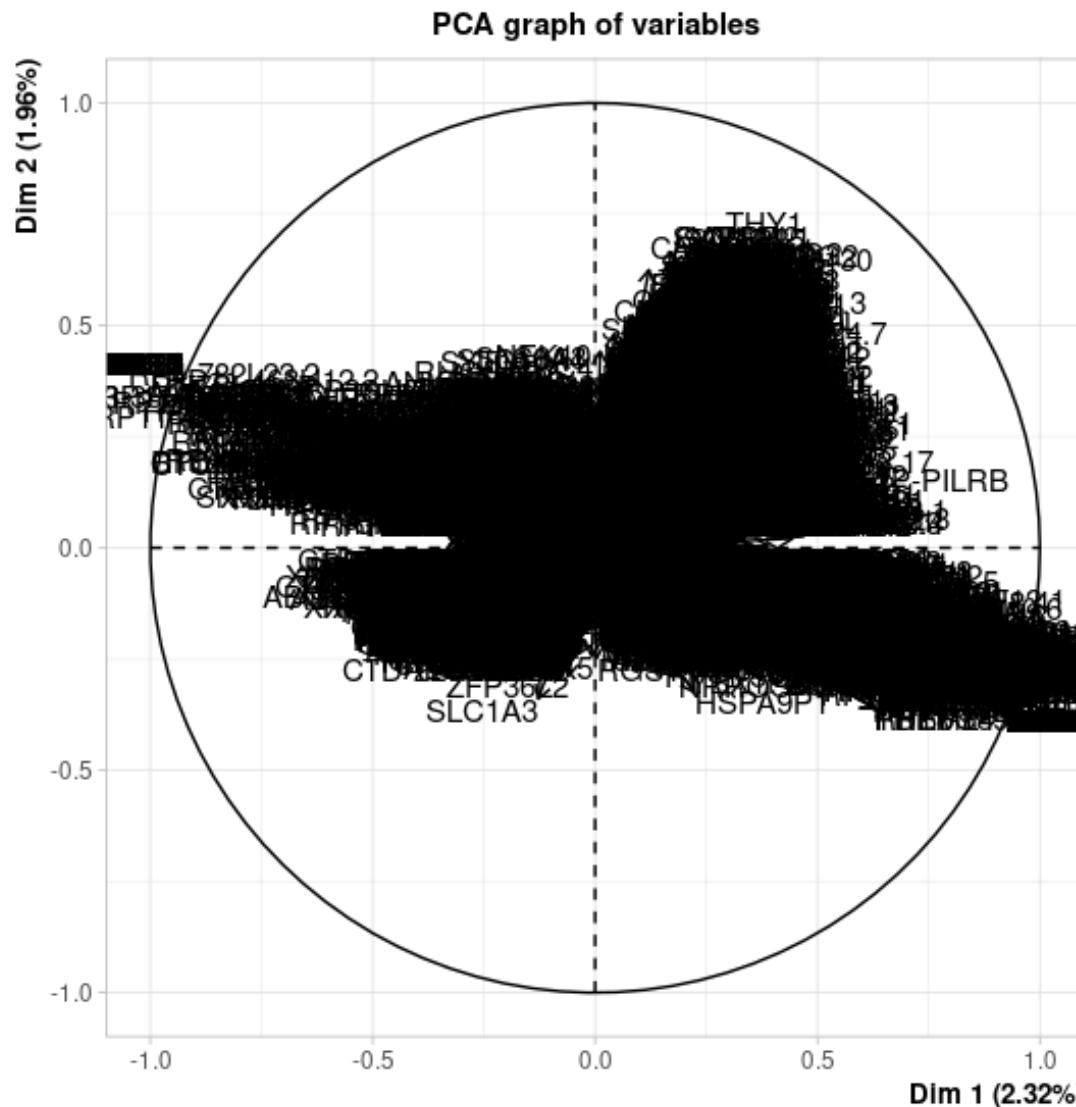
**Figure 10: Amount of variance explained by each principal component of a dataset using bootstrap analysis.** This plot has variance explained by the first 50 principal components. The green line connects the points that represent the variance explained by each principal component, while the red line represents the average explained variance by chance obtained from bootstrapping. 5 permutations of the data were done to calculate the distribution of the variance explained by chance.



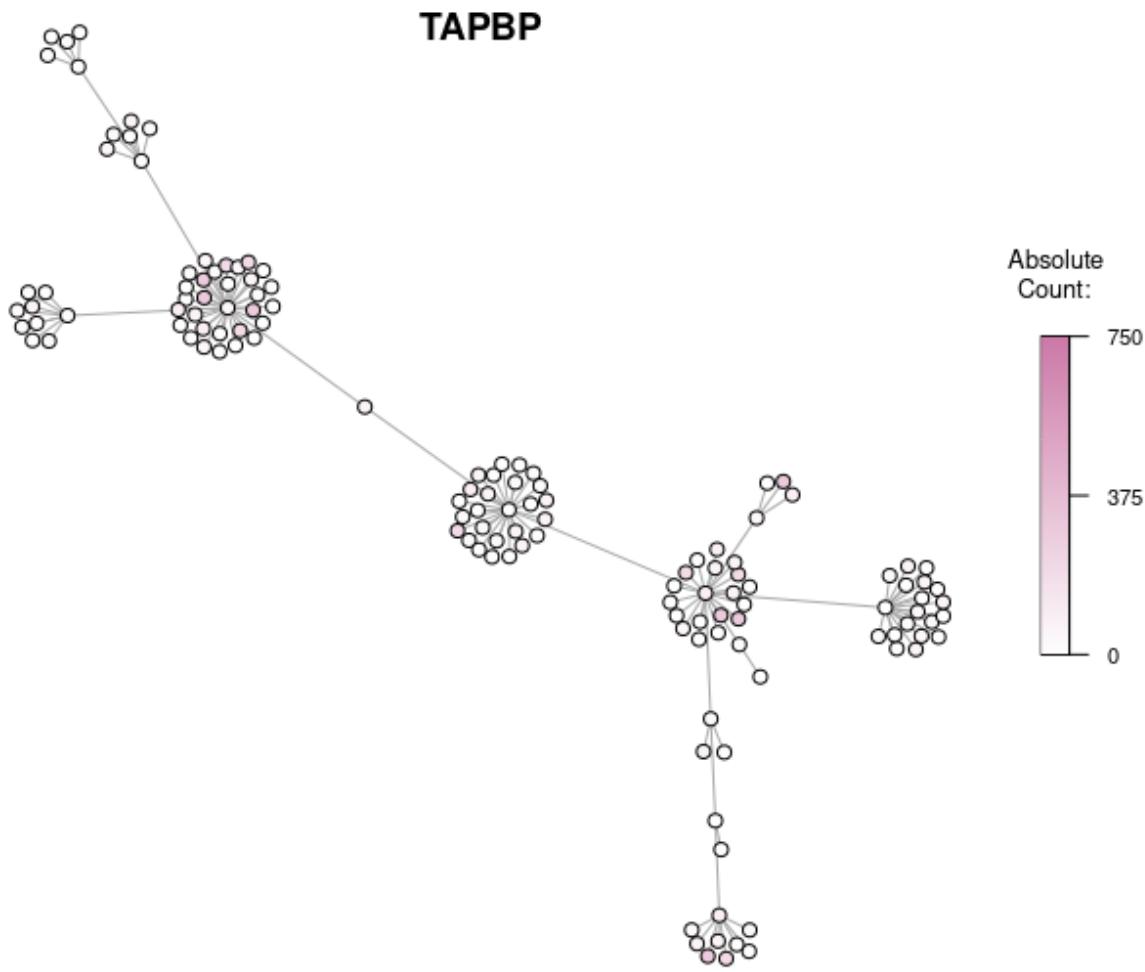
**Figure 11: Bootstrapping to determine the optimal number of principal components to use for dimensional reduction.** The p-values of the principal components (Y-axis) from the bootstrapped data are plotted against the principal components (X-axis). The optimal number of principal components is based on a p-value with a set threshold of 0.05. The optimal number of components is identified to be 18.



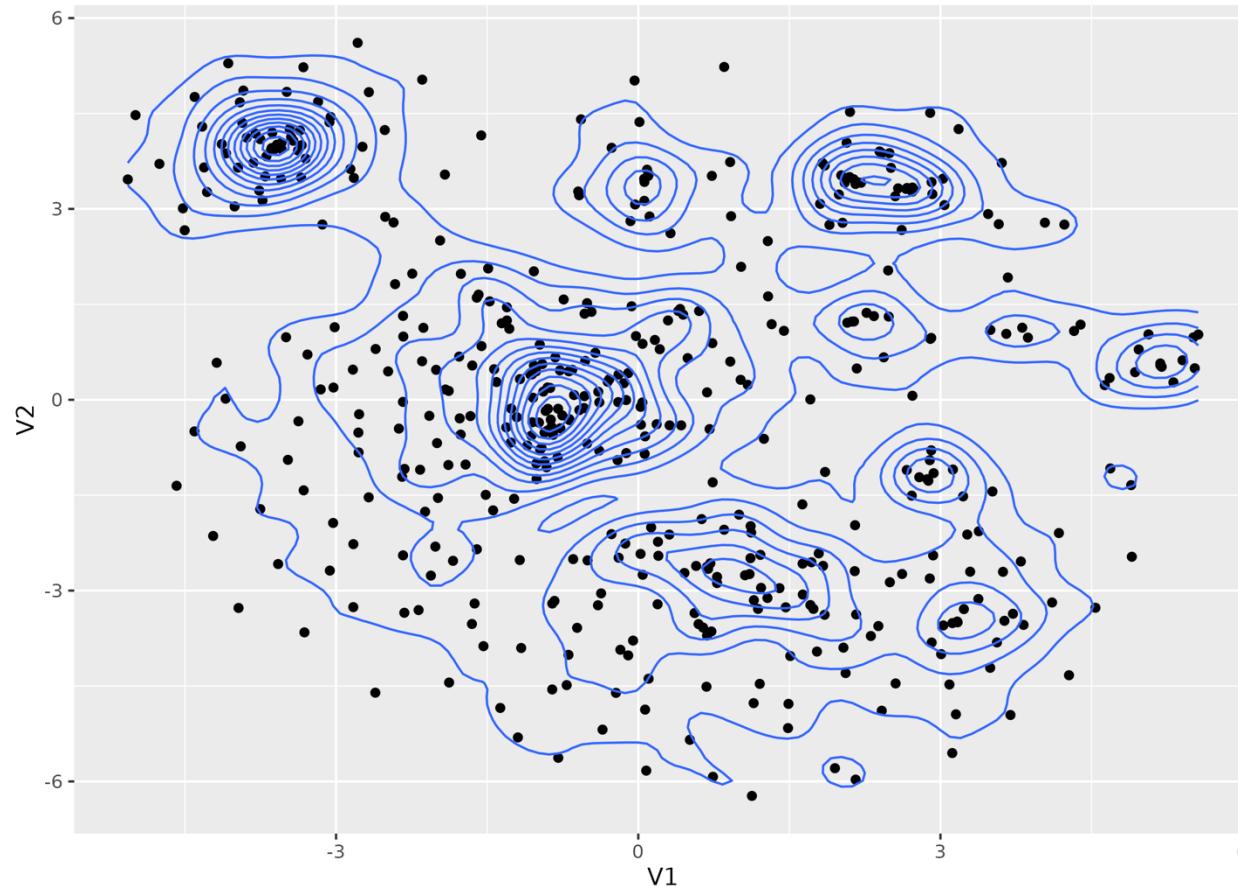
**Figure 12:** t-SNE plot showing the expression levels of PVALB. The colour scale ranges from 0 to 12.5, with the highest expression level represented by the colour red and the lowest the colour grey. X1 and X2 on the X- and Y- axes of this t-SNE plot refer to the two dimensions of the reduced-dimensional space.



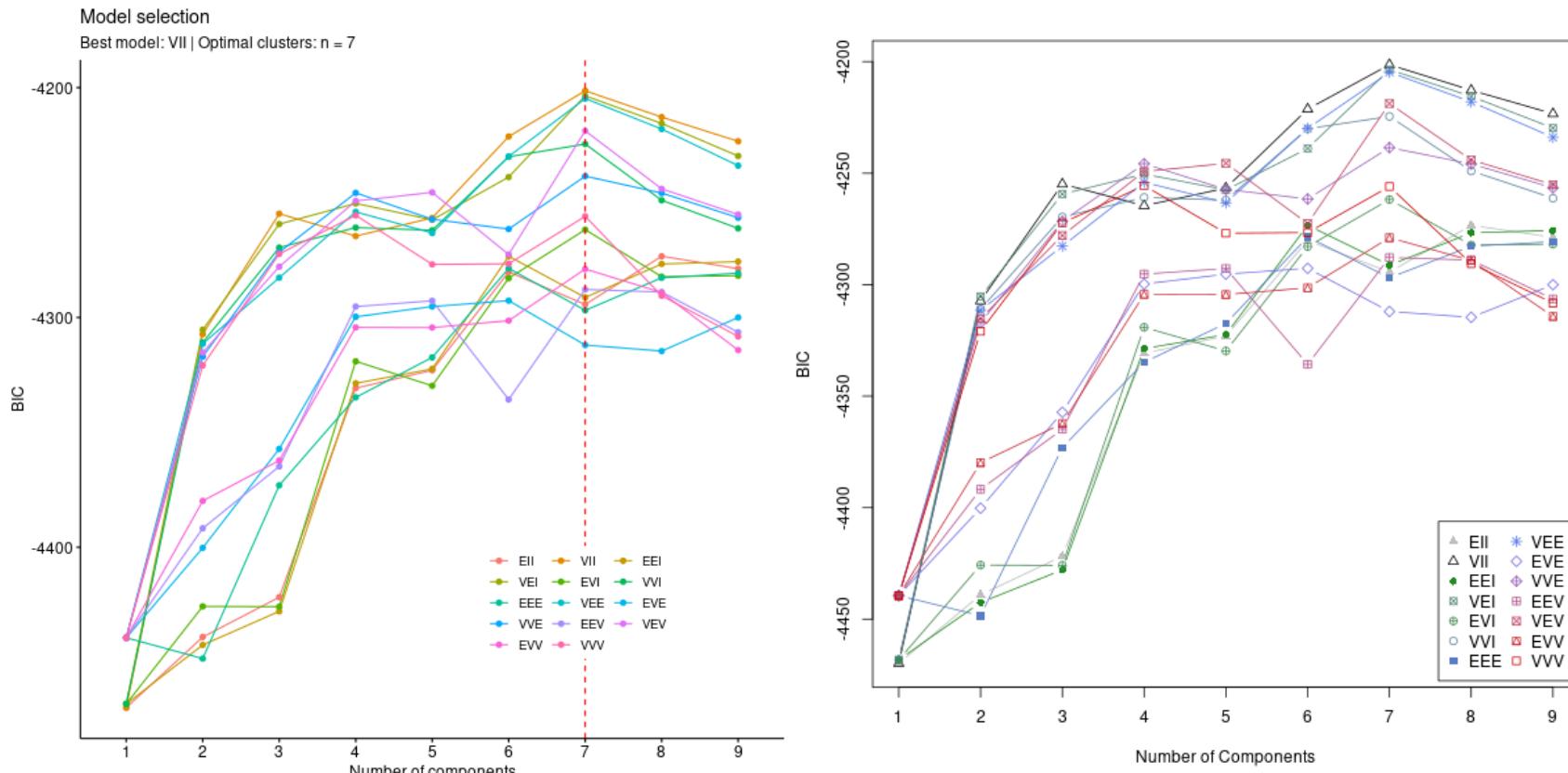
**Figure 13: PCA plot of variables.** The plot shows the relationships between all variables. Dim1 on the X-axis represents principal component 1 and Dim2 on the Y-axis represents principal component 2 with their respective percentages. The graph of variables is very clouded with each cell layered on-top of each other.



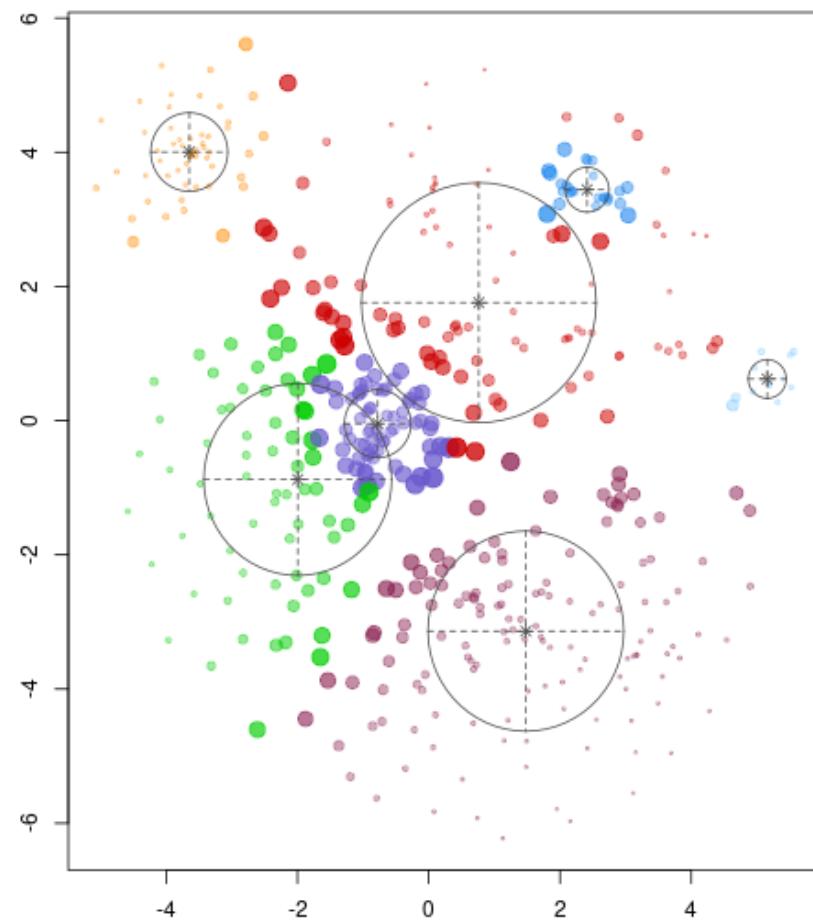
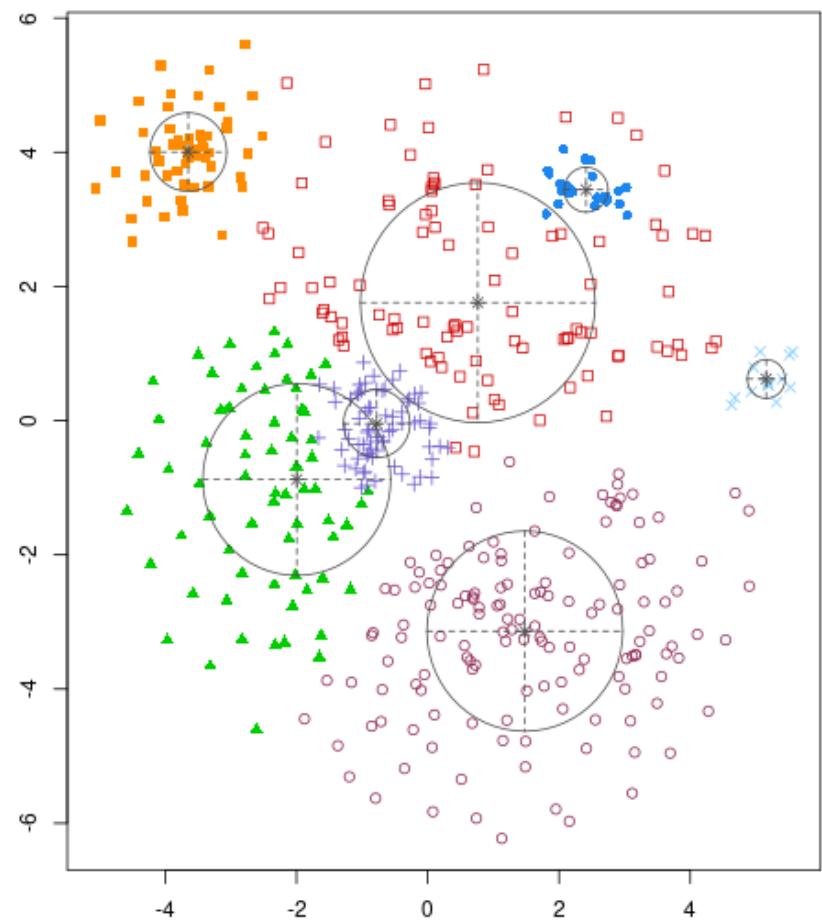
**Figure 14: Minimum spanning tree of TAPBP expression level in MHC-associated genes.** In this figure, is the expression level of TAPBP. Using the bar on the right, the colour gradient is shown to range from white to dark pink reflecting the expression levels, with dark pink being a high expression, the maximum count for TAPBP, and white being low expression levels. As seen in the figure, a few MHC-associated genes have higher expression levels of TAPBP.



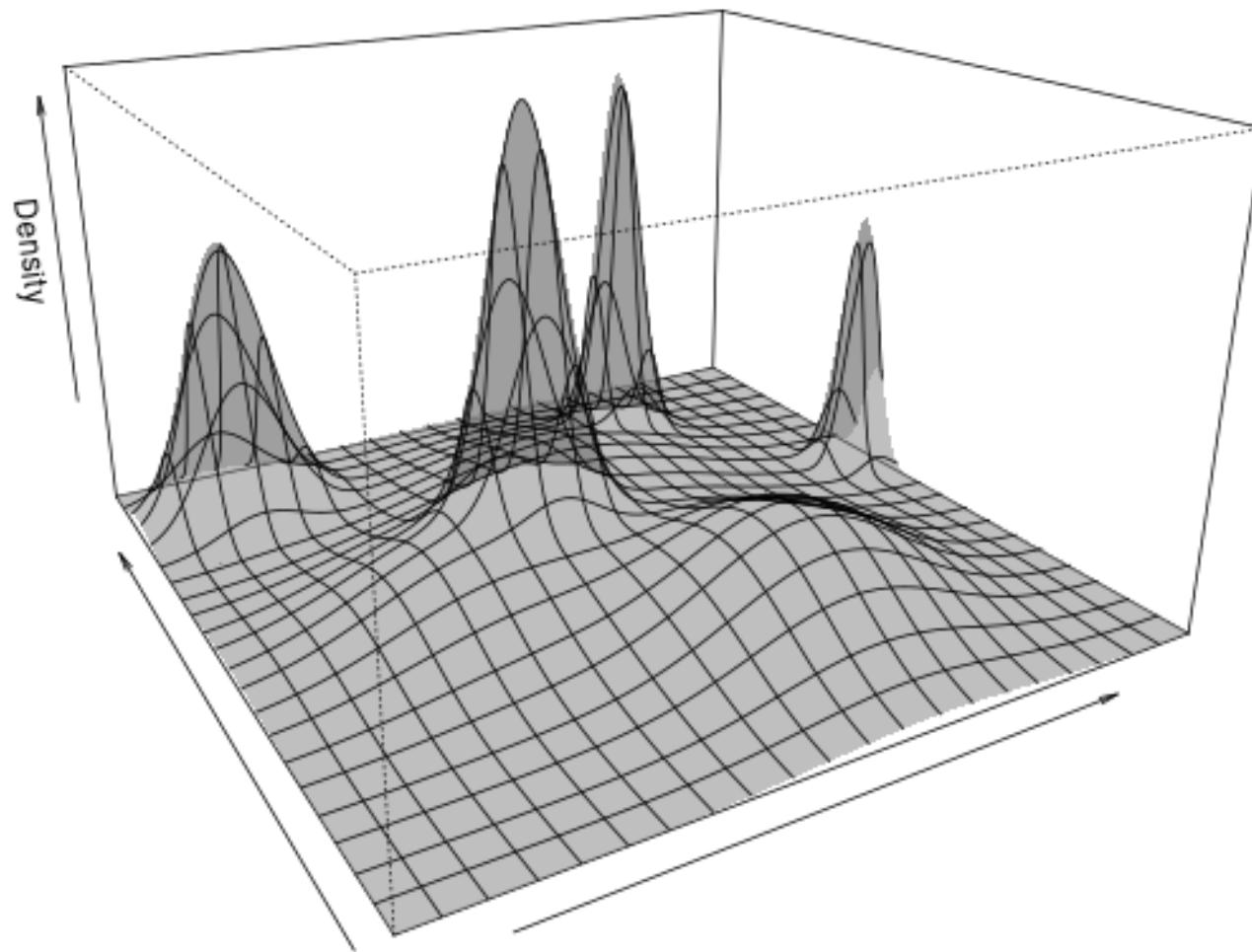
**Figure 15: Geometric density on the t-SNE results.** A two-dimensional distribution t-SNE plot is shown in this figure, whereby the geometric density is overlaid on the plot of gene expression data using an unbiased error model. Each plot represents a gene and its expression profile, providing insight into the distribution of data points in the two-dimensional space. This plot highlights areas with higher density and potential clustering patterns. V1 and V2 on the X- and Y-axes represent two-dimensional coordinates of each data point in the plot.



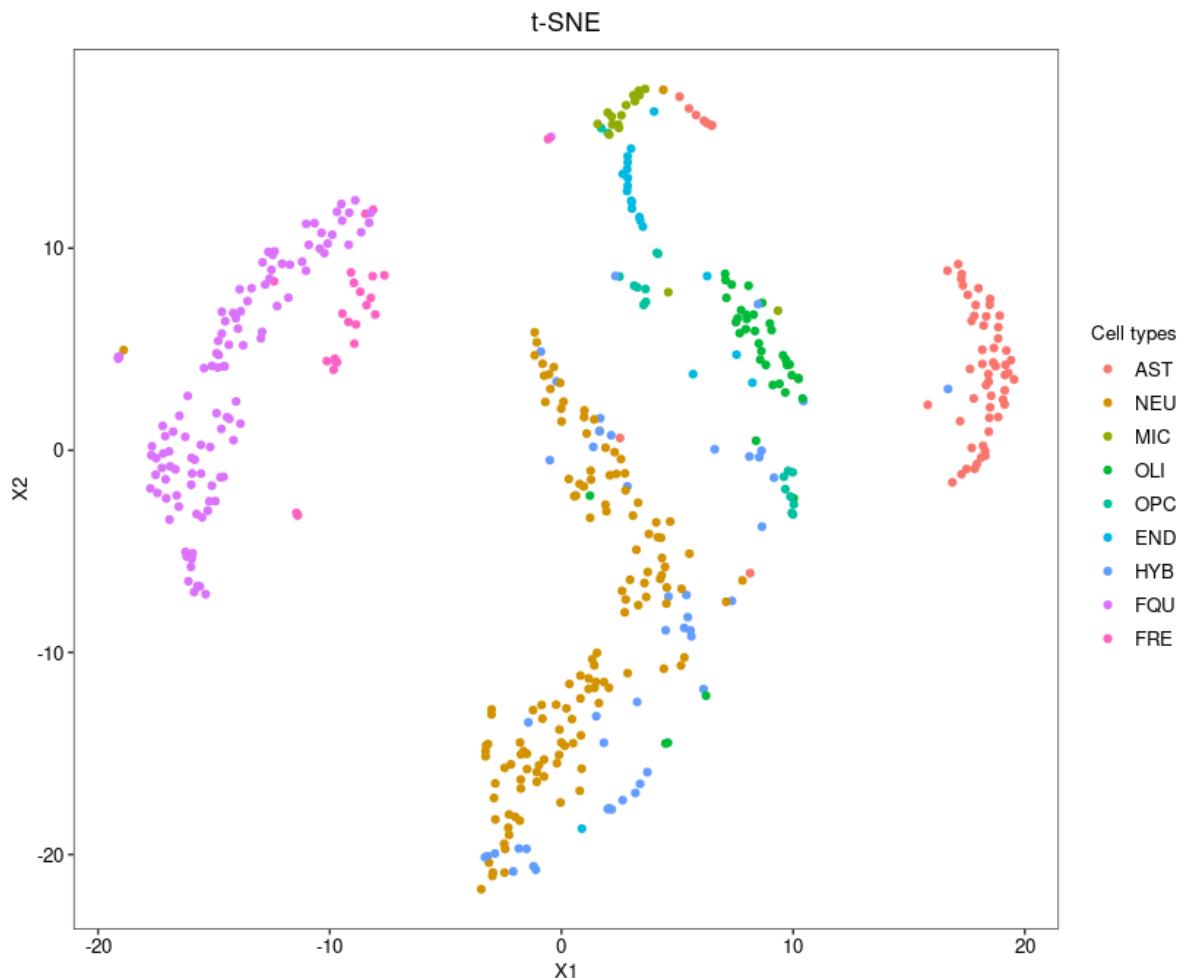
**Figure 16: BIC classification identifies an optimal number of 7 clusters with the best model being VII.** The plots visually show the optimal number of clusters for the data based on the BIC score. The x-axis represents the number of components in the model, and the y-axis represents the BIC score. As the BIC value decrease, the number of components increase. Each coloured line corresponds to a different model type. The letter combinations E (equal variance), V(variable variance), and I (independence) specify a model type. EII corresponds to a spherical model with equal volume and equal shape, VII corresponds to a spherical model with variable volume and equal shape, EEI corresponds to a diagonal model with equal volume and equal shape, VEI corresponds to a diagonal model with variable volume and equal shape, EVI corresponds to a diagonal model with equal volume and variable shape, VVI corresponds to a diagonal model with variable volume and variable shape, EEE corresponds to an ellipsoidal model with equal volume and equal shape, EEV corresponds to an ellipsoidal model with equal volume and equal shape, VEV corresponds to an ellipsoidal model with variable volume and equal shape and VVV corresponds to an ellipsoidal model with variable volume and variable shape (Darmanis et al., 2015). The models are compared based on their BIC values, with lower BIC values indicating a better model. The best model has been identified to be VII. The vertical red dotted line indicates the optimal number of components based on the BIC criterion.



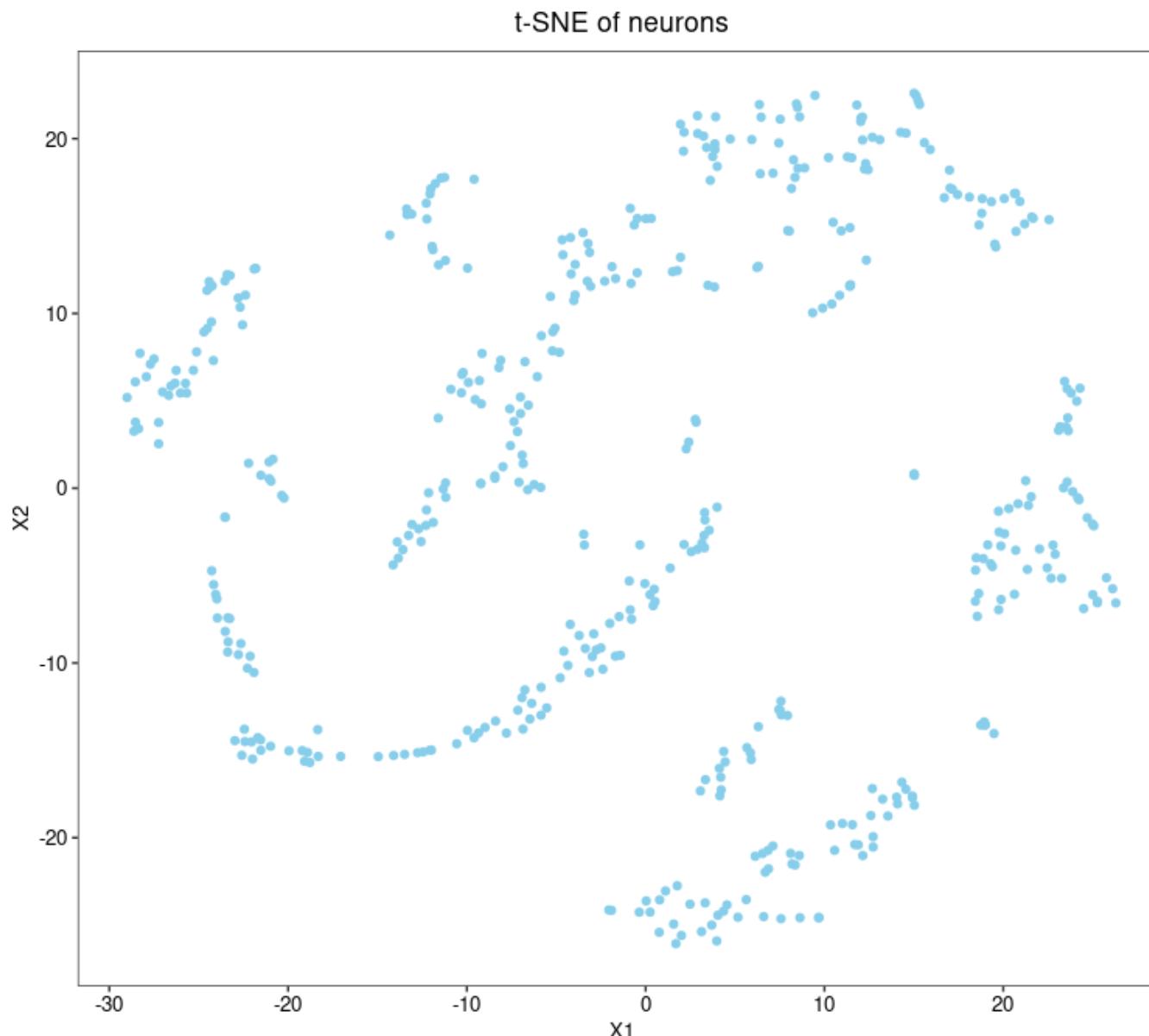
**Figure 17: The Mclust algorithm was run with the best model, Mclust VII as determined by BIC.** Each of the points on the plot represents a cell and its colour represents the cluster it belongs to. The X- and Y- axes represents the first and second principal components of the t-SNE results. The image on the left shows the classification plot obtained after clustering the t-SNE results using the Mclust algorithm showing which cluster has aligned well with the axes of the plot. The plot on the right of the figure shows the uncertainty plot to highlighting the parts of the data that is poorly aligned.



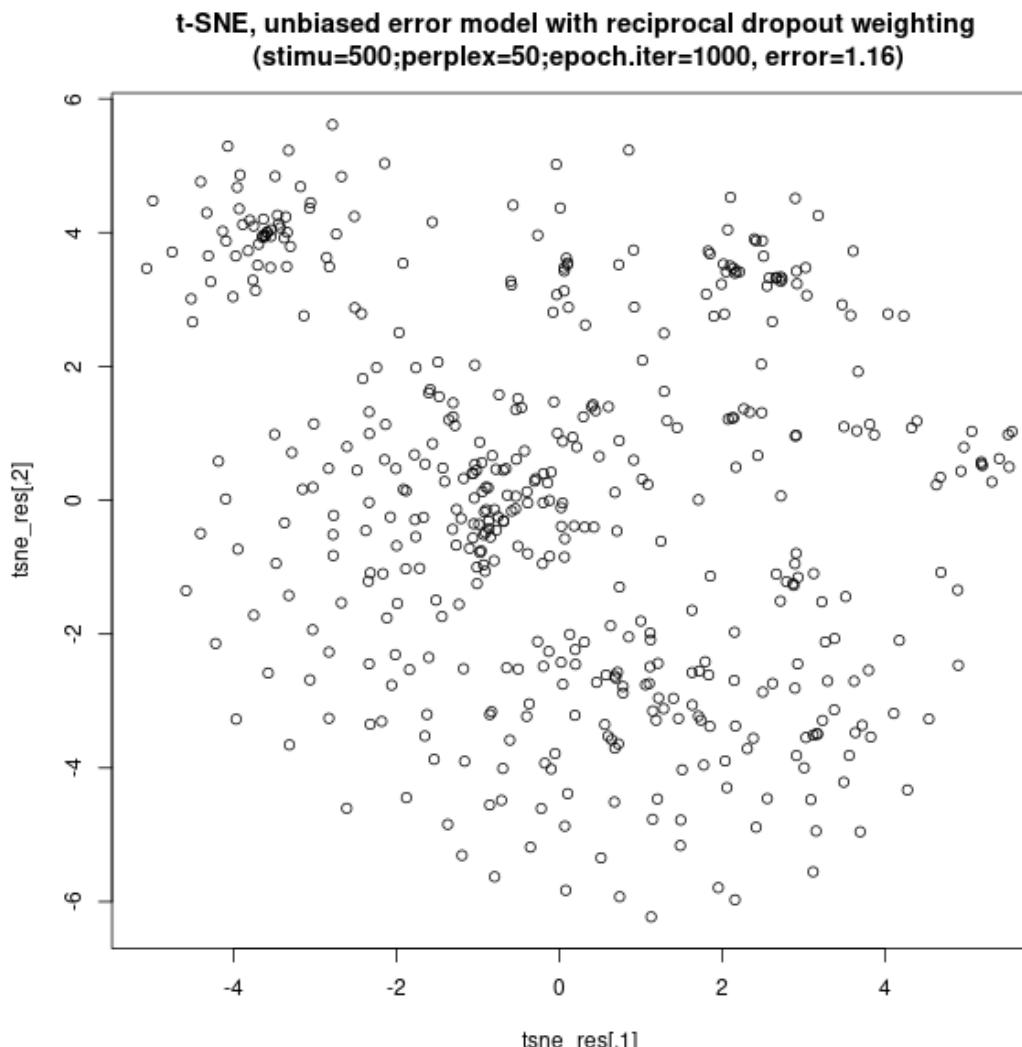
**Figure 18: Density perspective plot of the t-SNE results for the unbiased model.** The figure is visualisation of the density distribution of the clusters of data points in a three-dimensional space. The z-axis represents the density of the data points in that region, whilst the X- and Y- axes are the two dimensions of the t-SNE space. Regions of high and low density are highlighted in this plot.



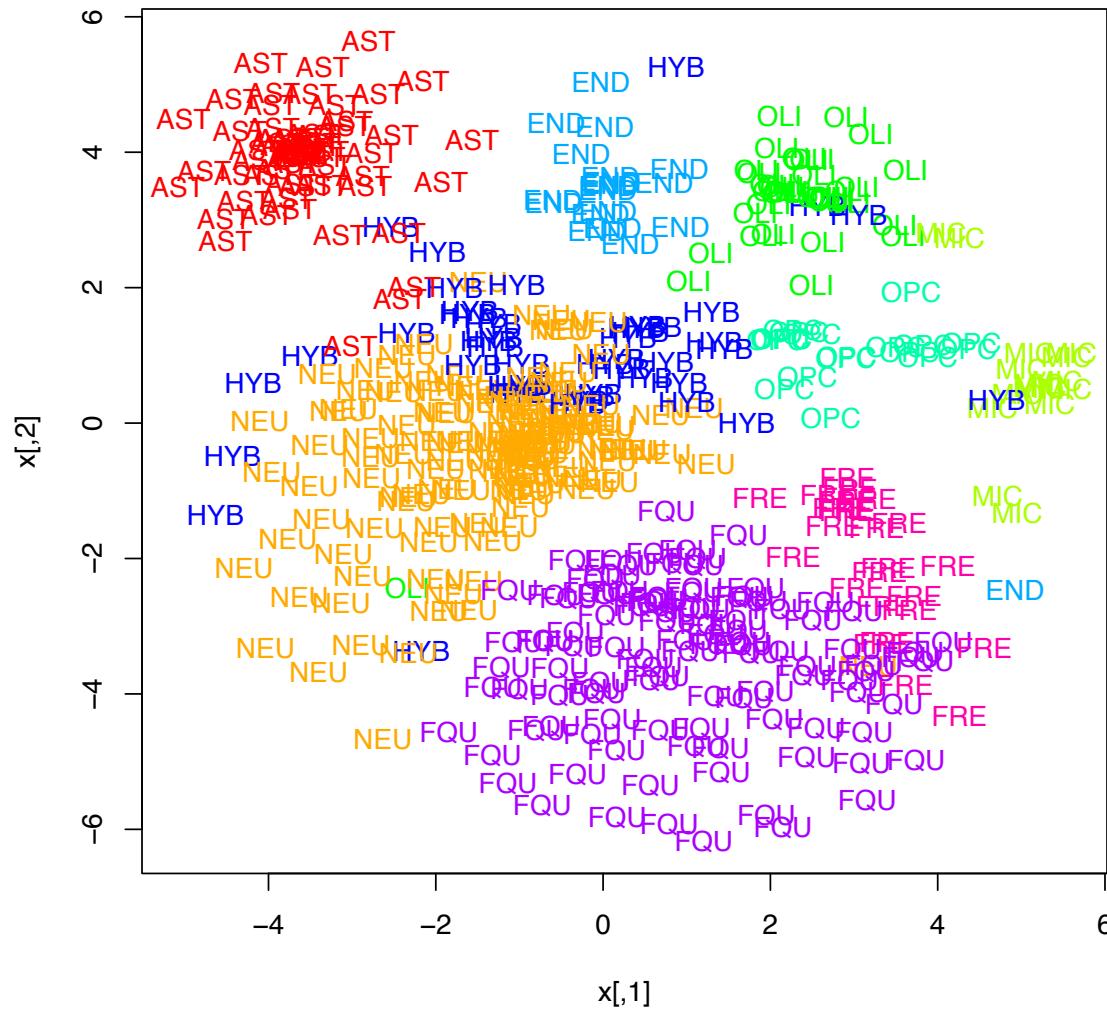
**Figure 19: t-SNE plot of single-cell RNA-seq data.** t-SNE was performed to reduce dimensions. Each dot represents an individual cell which is colour coded according to its cell type which is shown on the key on the right-hand side of the figure.



**Figure 20: Gene expression across the neurones, visualized by t-SNE.** This figure displays a two-dimensional representation of high-dimensional single-cell gene expression data, with each point representing a single cell.



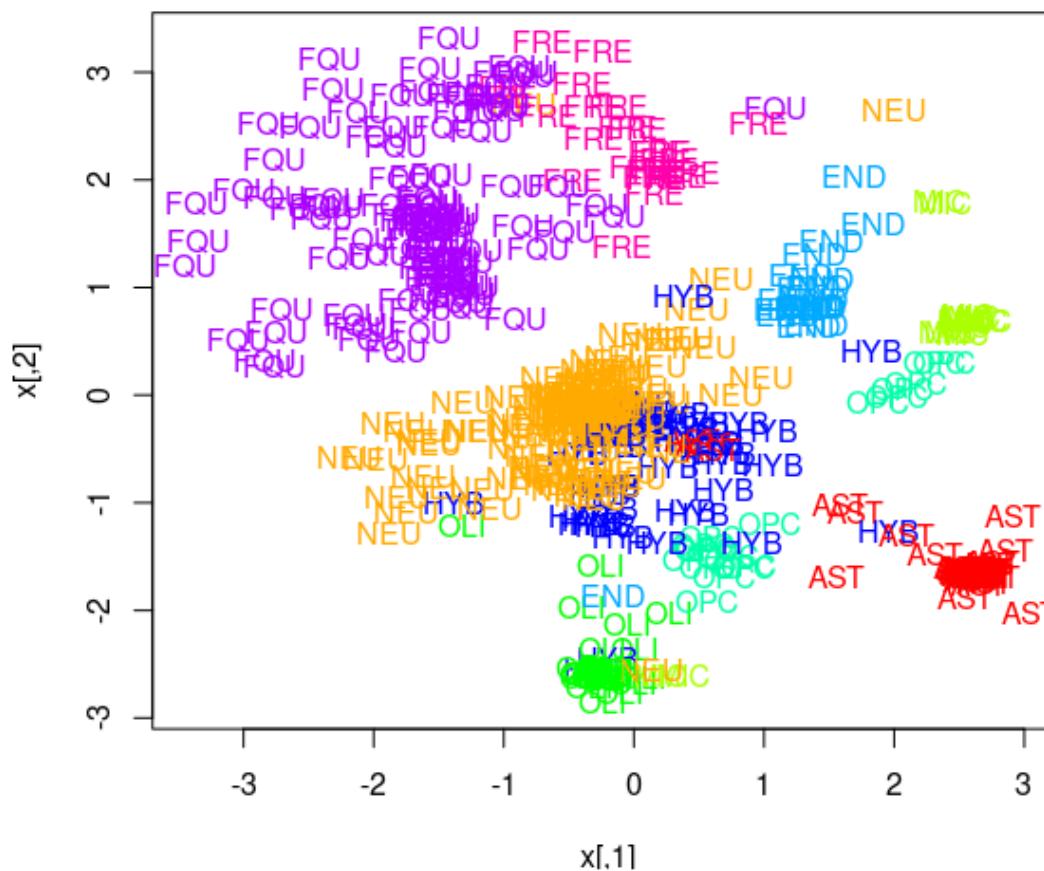
**Figure 21: Two-dimensional t-SNE plot.** This figure shows the distribution of data points in a two-dimensional scatterplot. This dataset was created using an unbiased error model with reciprocal dropout weighting. The x and y-axes represent the two dimensions obtained from the t-SNE algorithm. The data was generated with a perplexity of 80 giving a relatively greater emphasis on global structure, with the t-SNE algorithm iterating over the dataset 1000 times. Each point represents a single observation with the density of the points increasing where there are more observations.



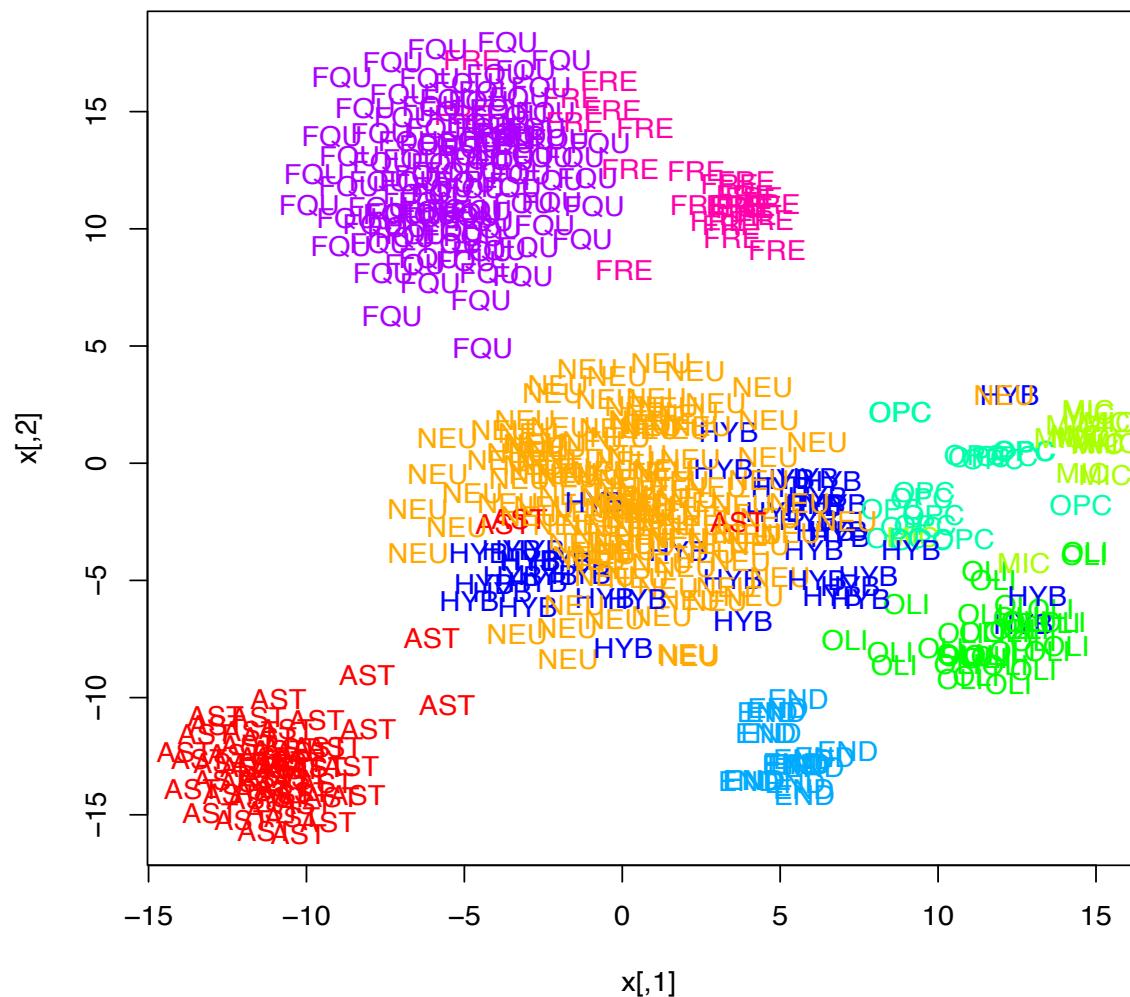
**Figure 22: t-SNE plot showing unbiased clustering of data points based on reciprocal dropout weighting.** Each point in the figure represents the data point. Clustering was performed using an error model with reciprocal dropout weighting and a perplexity value of 80. The plots are the relative distances of data points, with more similar data points clustered together. The map is coloured according to the distinct groupings identified by the clustering study, with darker regions signifying higher densities.

## Standard and late exaggeration

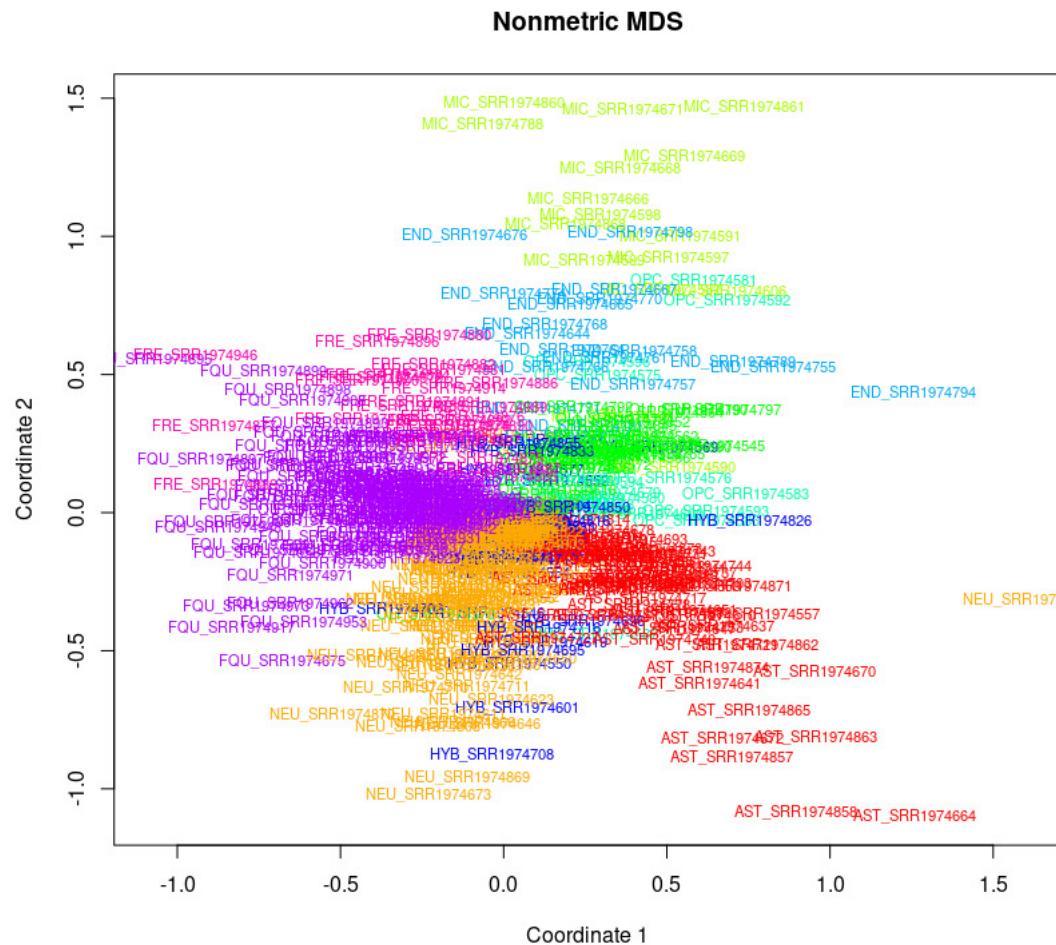
perplexity = 50, exaggeration\_factor = 12, stop\_lying\_iter = 250,  
late\_exaggeration\_factor = 1.5, start\_late\_lying\_iter = 750



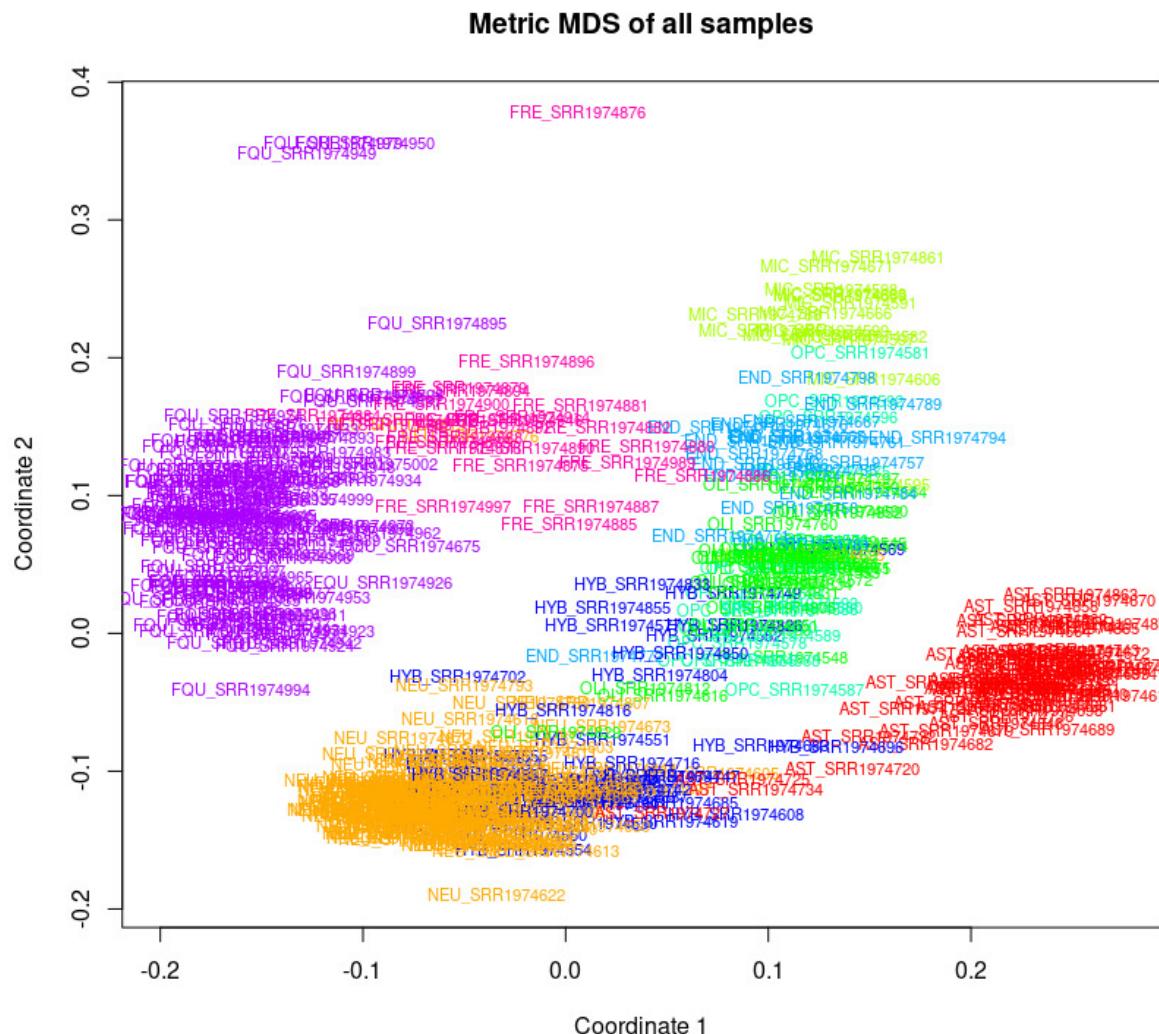
**Figure 23: Results of a force clustering experiment using t-SNE visualisation in the original pipeline.** The t-SNE visualization was performed using the smallvis algorithm with a perplexity value of 50, eta value of 100, and exaggeration factor of 12. The figure is the results under the condition of standard and late exaggeration. Each data point represents a cell type. The cells are colour coded and labelled based on their clusters.



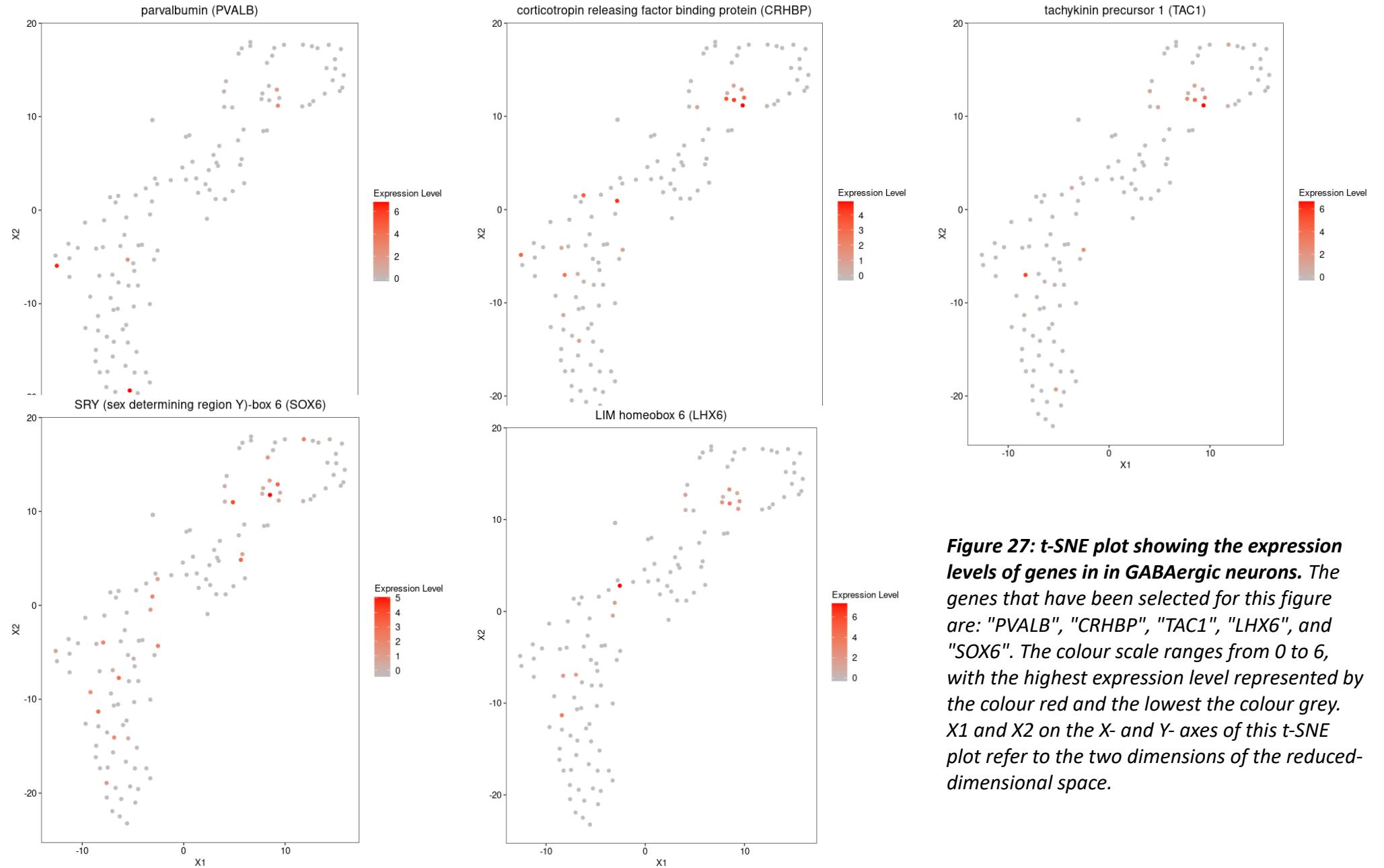
**Figure 24: t-SNE plot showing biased clustering from biased error model.** The figure was created by applying t-SNE to biased gene expression profiles that only contained a selection of genes with high variance. Each data point is represented as a point on the two-dimensional t-SNE space in the figure. Points with similarities are grouped. The colour of the plot is determined by the cluster assignments produced using a model-based clustering approach.



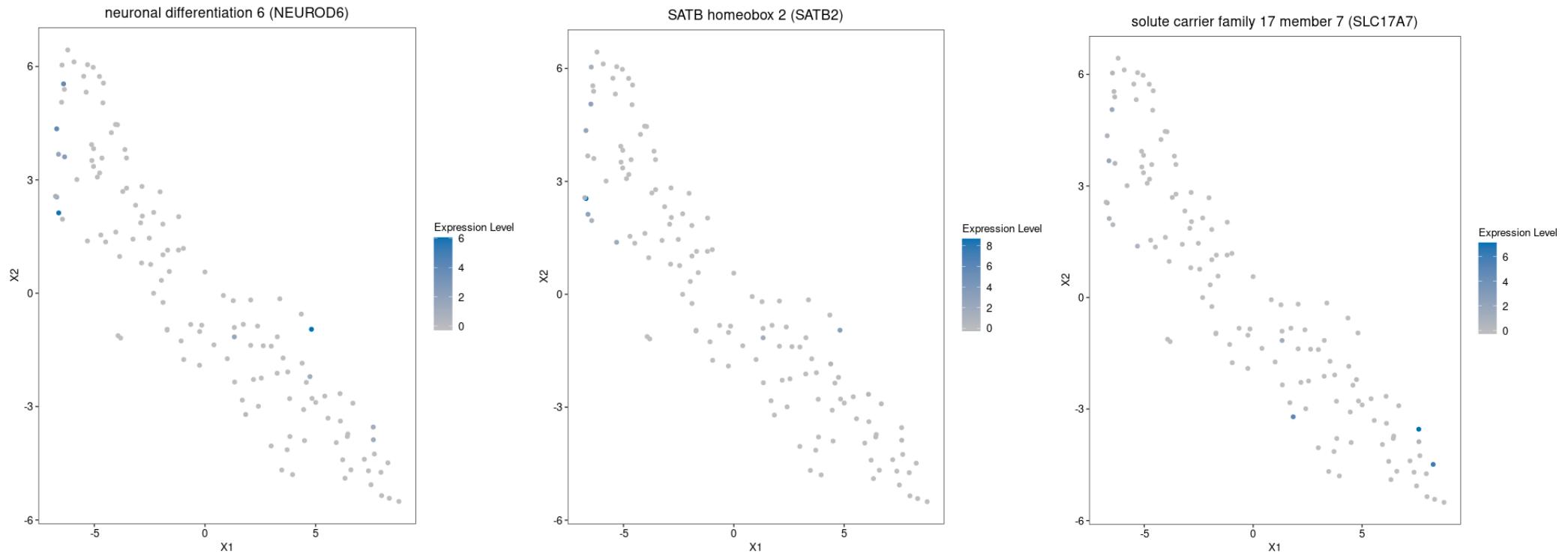
**Figure 25: Nonmetric MDS plot.** The figure shows the distribution using nonmetric MDS. The positions of the samples were obtained through iterative MDS. Each of the colours of the points belongs to different cell types and they are labelled accordingly. The clustering of samples is done in a two-dimensional space based on their pairwise distances. The x and y axes correspond to the first and second coordinates of the nonmetric MDS solution.



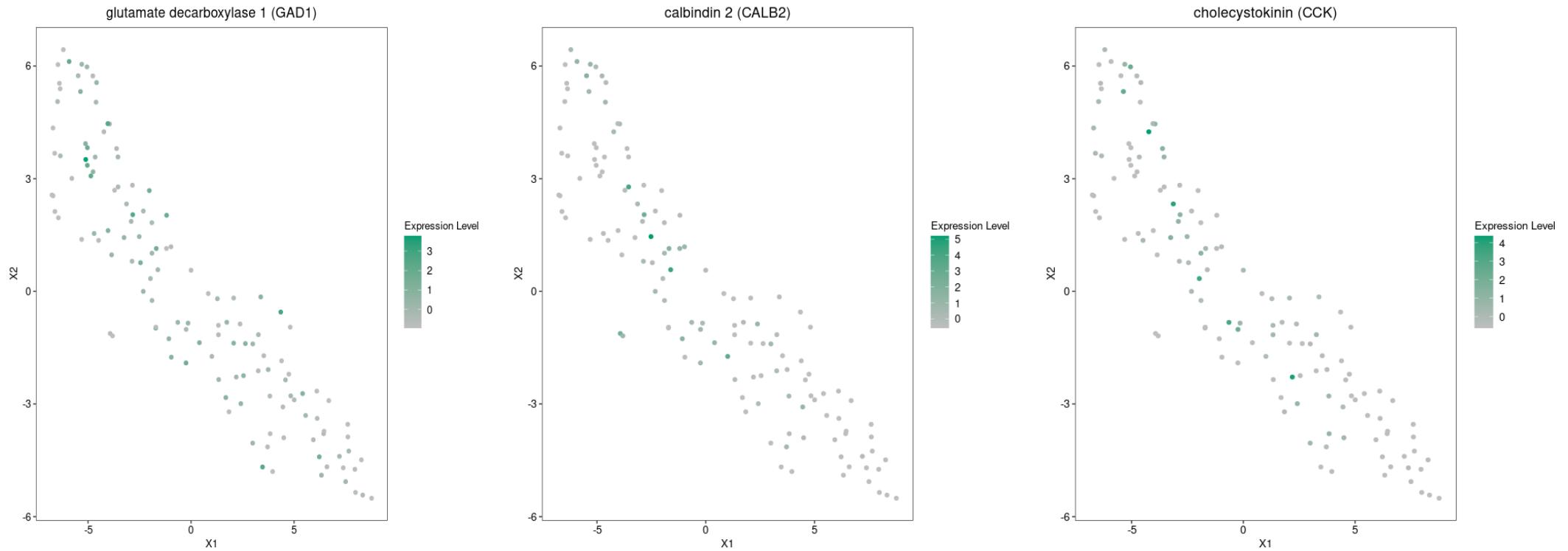
**Figure 26: Metric MDS of all samples showing their relation based on gene expression.** This figure shows the result of metric multidimensional scaling (MDS) performed on a gene expression dataset from different cell types, including astrocytes (AST), neurons (NEU), microglia (MIC), oligodendrocytes (OLI), oligodendrocyte progenitor cells (OPC), endothelial cells (END), hybrid cells (HYB), and two types of fibroblasts (FQU and FRE). Based on each of their gene expression profiles, the MDS analysis produced a 2D coordinate system that represents the relationship between samples in this plot with the X-axis representing Coordinate 1 and the Y-axis representing Coordinate 2. The Euclidean distances between samples were converted into a matrix, which was then utilised for principal component analysis (PCA). The colours of the labels correspond to the different cell types.



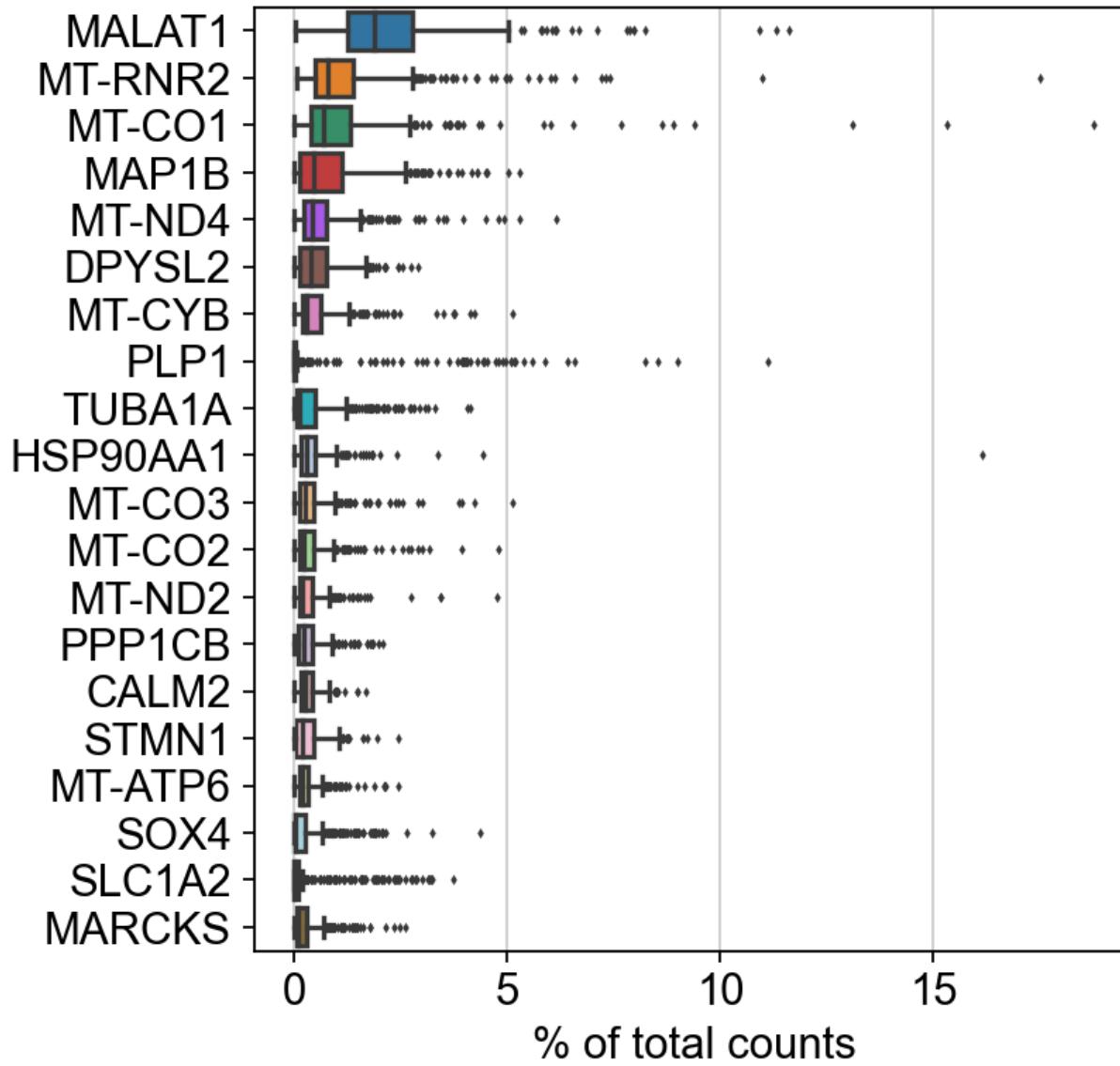
**Figure 27: t-SNE plot showing the expression levels of genes in GABAergic neurons.** The genes that have been selected for this figure are: "PVALB", "CRHBP", "TAC1", "LHX6", and "SOX6". The colour scale ranges from 0 to 6, with the highest expression level represented by the colour red and the lowest the colour grey. X1 and X2 on the X- and Y-axes of this t-SNE plot refer to the two dimensions of the reduced-dimensional space.



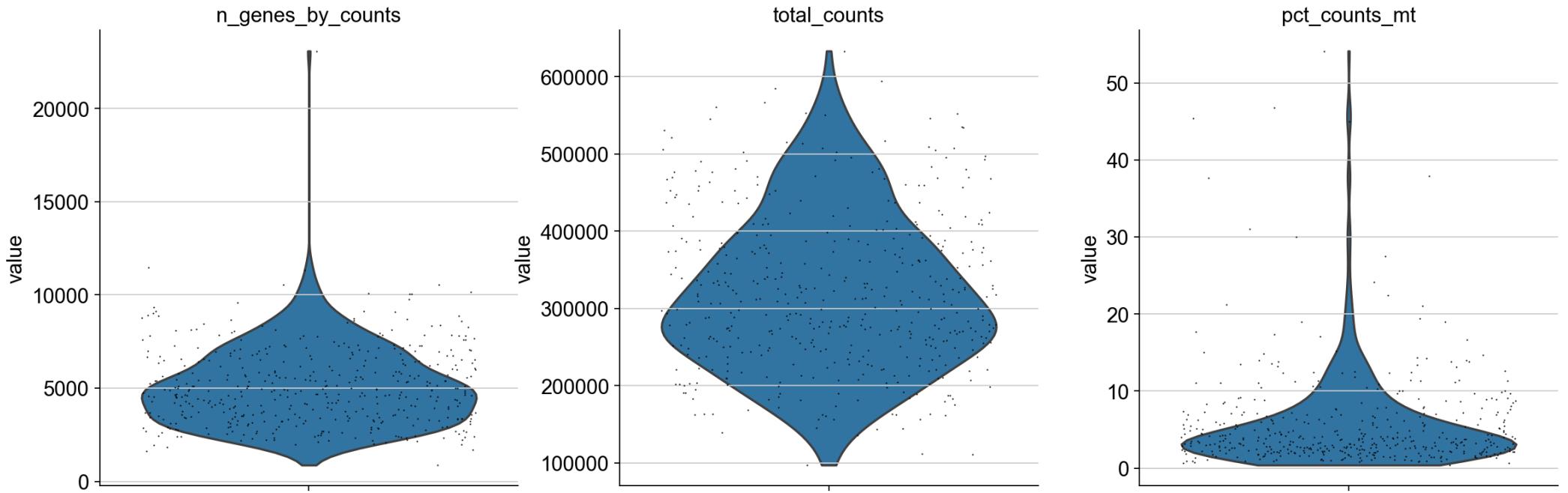
**Figure 28: t-SNE plot showing the expression levels of genes in distinct populations of excitatory communities.** SLC17A7, NEUROD6, and SATB2 are the genes that have been selected for this figure, all of which are markers for excitatory neurons. Cells are coloured according to the expression levels of the genes NEUROD6, SATB2, and SLC17A7. The colour scale ranges from 0 to 6, with the highest expression level represented by the colour blue and the lowest the colour grey. X1 and X2 on the X- and Y- axes of this t-SNE plot refer to the two dimensions of the reduced-dimensional space.



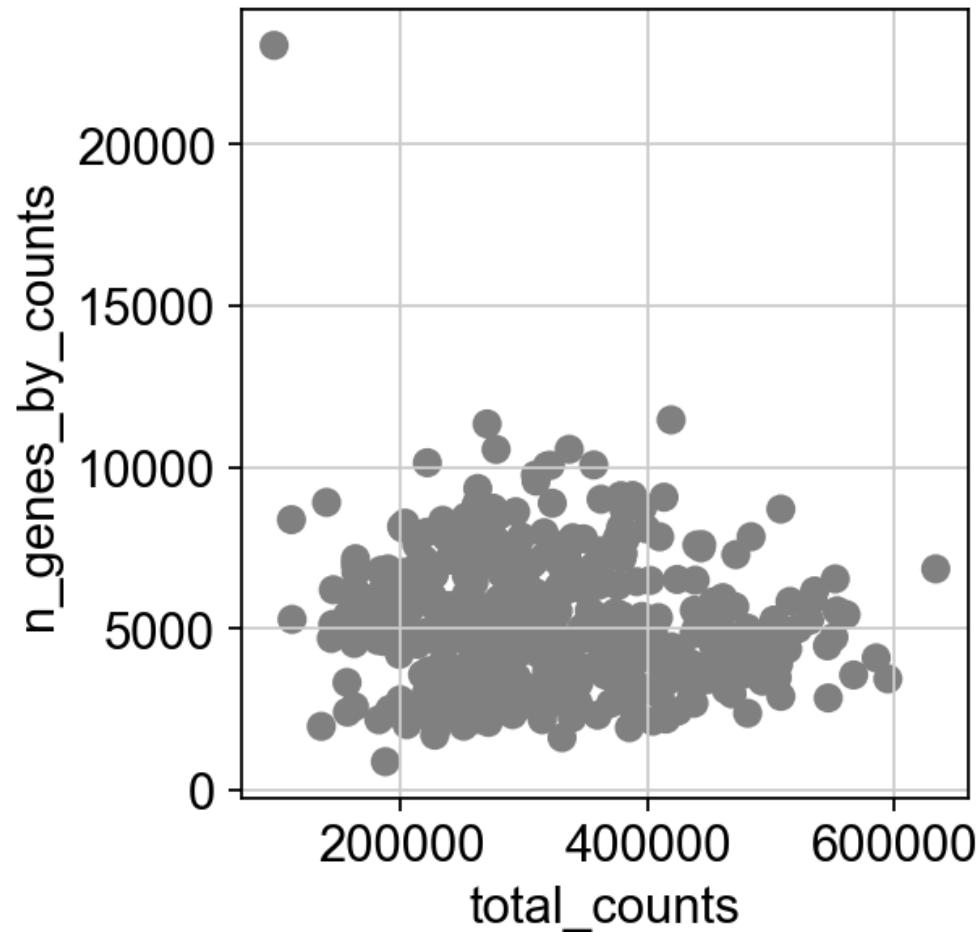
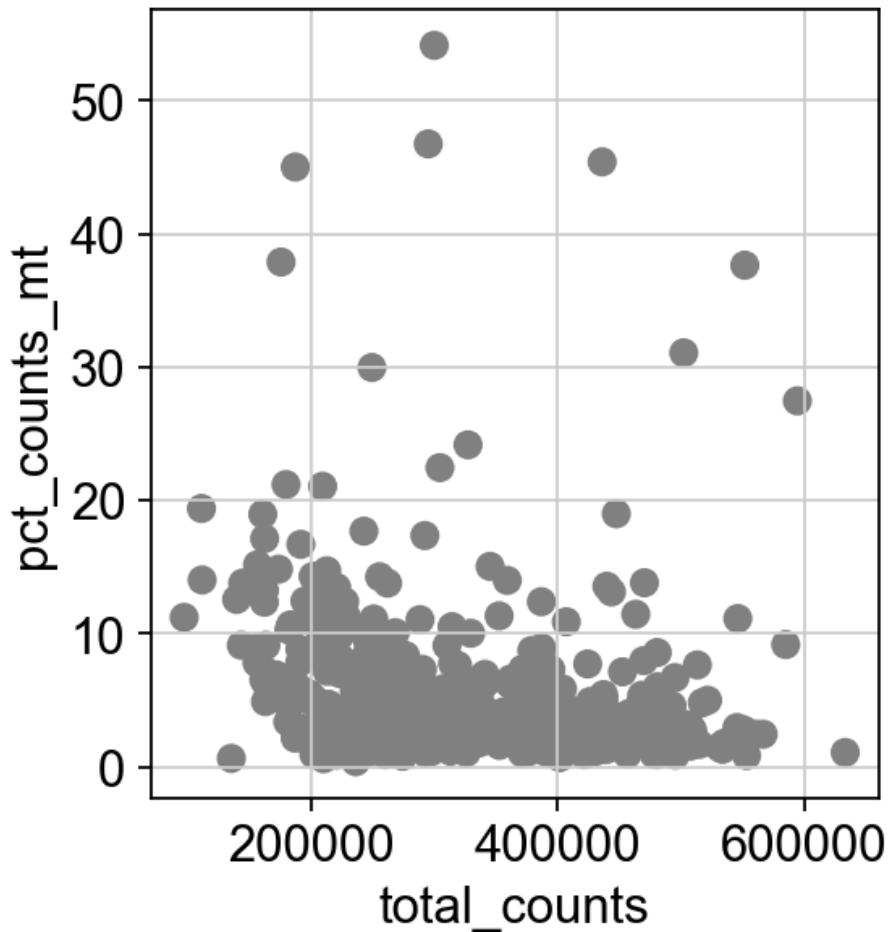
**Figure 29: t-SNE plot showing the expression levels of genes in distinct populations of cells.** On the plot, cells are coloured according to the expression levels of three genes: GAD1, CCK, and CALB2, all of which are markers for inhibitory neurons. The colour scale explains the expression levels of each ranging from 0 to 5, with the highest expression level represented by the colour green and the lowest the colour grey. X1 and X2 on the X- and Y- axes of this t-SNE plot refer to the two dimensions of the reduced-dimensional space.



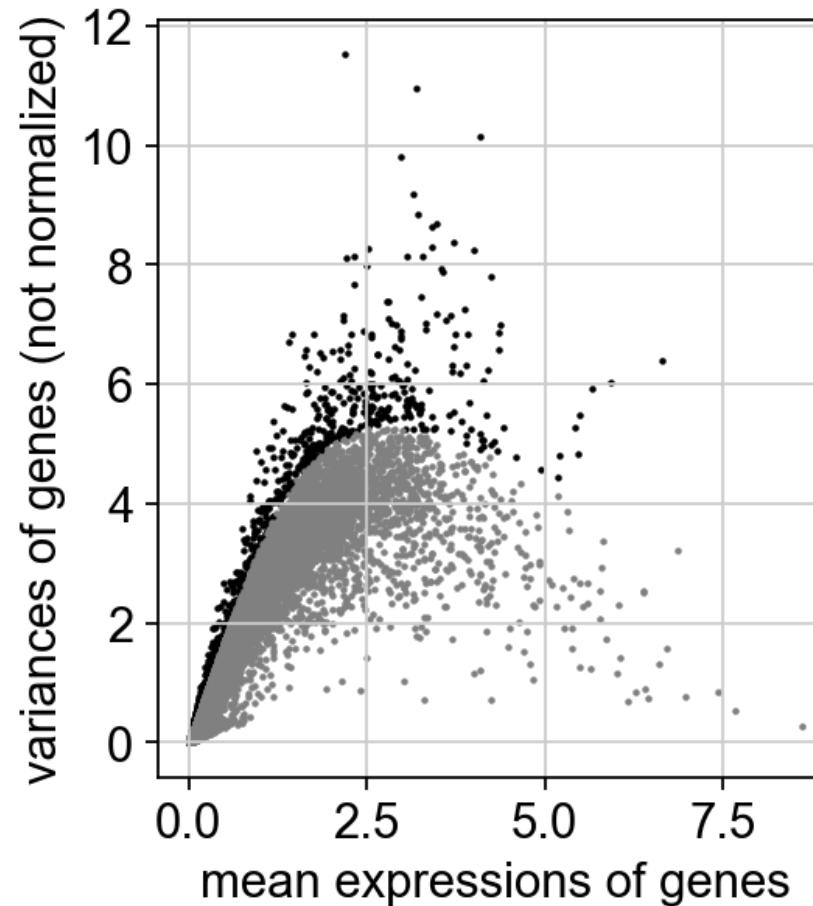
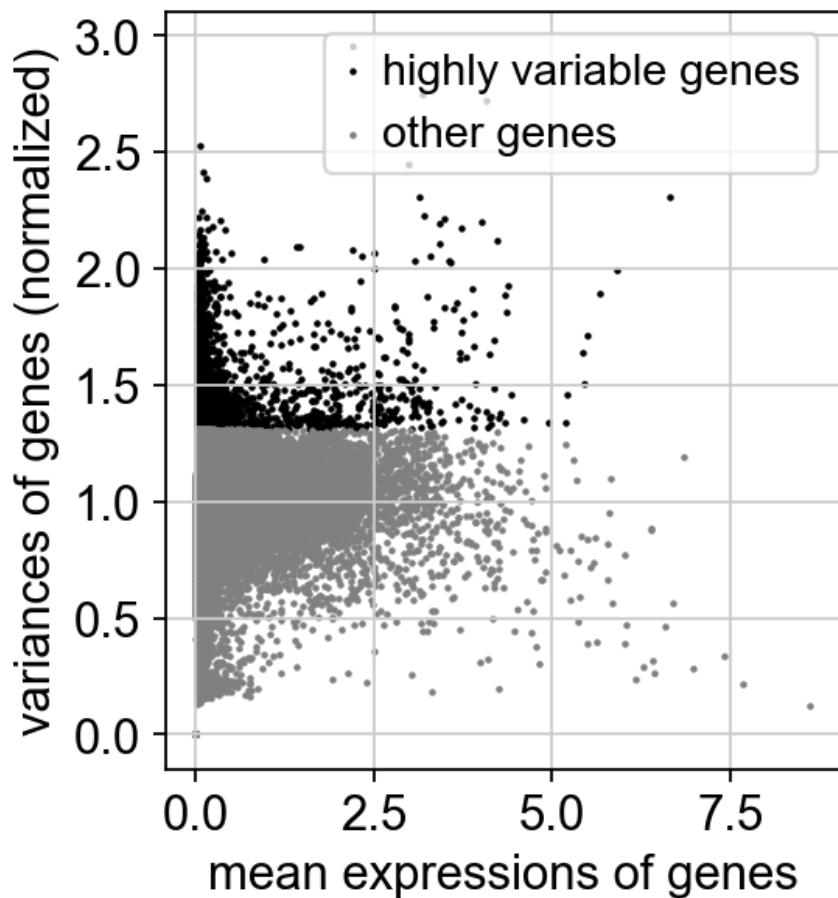
**Figure 30: Visualisation of the top 20 highest expressed genes with a set of boxplots.** In this figure, the Y-axis shows the expression level of each gene in percent of total counts, while the X-axis ranks the genes from highest to lowest expression. Each boxplot reflects the expression value distribution for each gene across all cells in the dataset. The box reflects the expression values' interquartile range (IQR), with the horizontal line inside the box denoting the median expression value. The whiskers that extend from the box reflect the range of expression values that are within 1.5 times the IQR. Outliers are points that fall outside of this range and are represented as isolated dots.



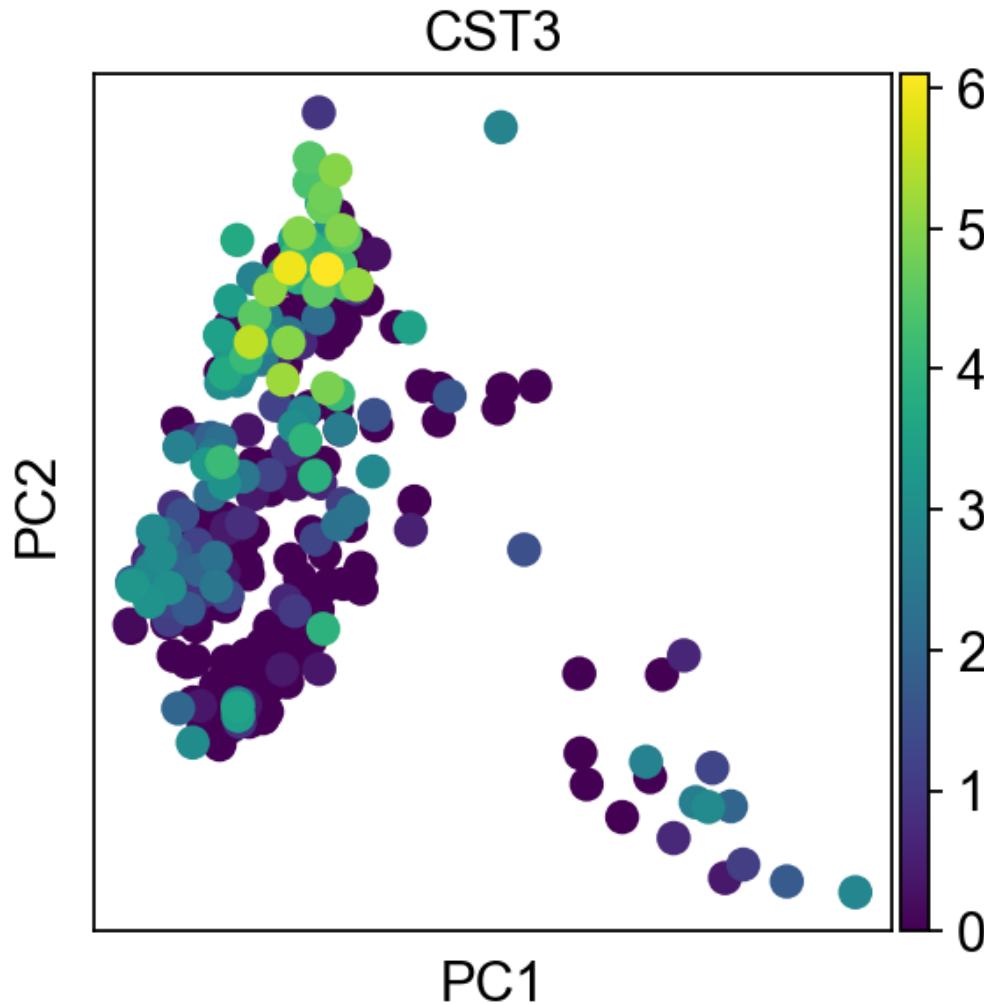
**Figure 31: Violin plot shows the distribution of single-cells gene expression.** The multi-panel figure shows each of the three violin plots side-by-side. Violin plots are used to visualize a summary of the distribution of numerical data. The violin plot on the left of the figure shows the number of genes expressed in each cell, the middle figure shows the total counts of RNA molecules detected in each cell and the right figure shows the percentage of counts of RNA molecules that are mitochondrial. The Y-axis of each shows the density or frequency of cells within each category. The width of each violin plot corresponds to the density of cells with that metric value. The violin's height represents the frequency of cells with the given expression value. The points across the plot indicate the individual cells' metric values.



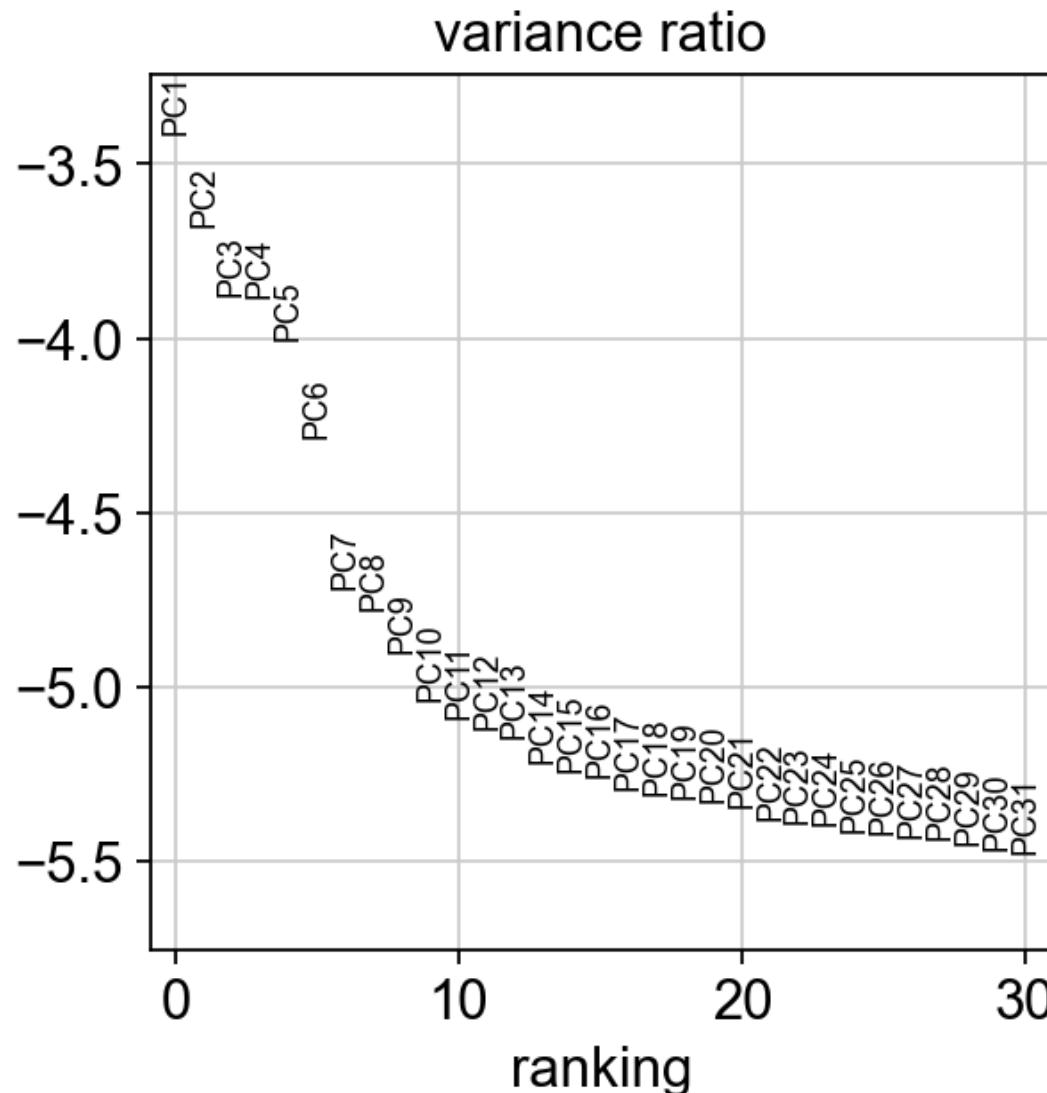
**Figure 32: Scatterplots showing the relationship between two variables in the single-cell RNA sequencing data.** The scatterplot on the left shows the relationship between the total number of counts in each cell and the percentage of counts that come from mitochondrial genes. The X-axis represents the total number of counts. The Y-axis represents the percentage of counts that come from mitochondrial genes. The scatterplot on the right shows the relationship between the total number of counts in each cell and the number of genes expressed in each cell. This figure is useful for inspecting data quality and highlighting cells that should be filtered out due to low gene expression levels. The X-axis represents the total number of counts. The Y-axis represents the number of genes expressed. Each point on the scatterplot represents individual cells.



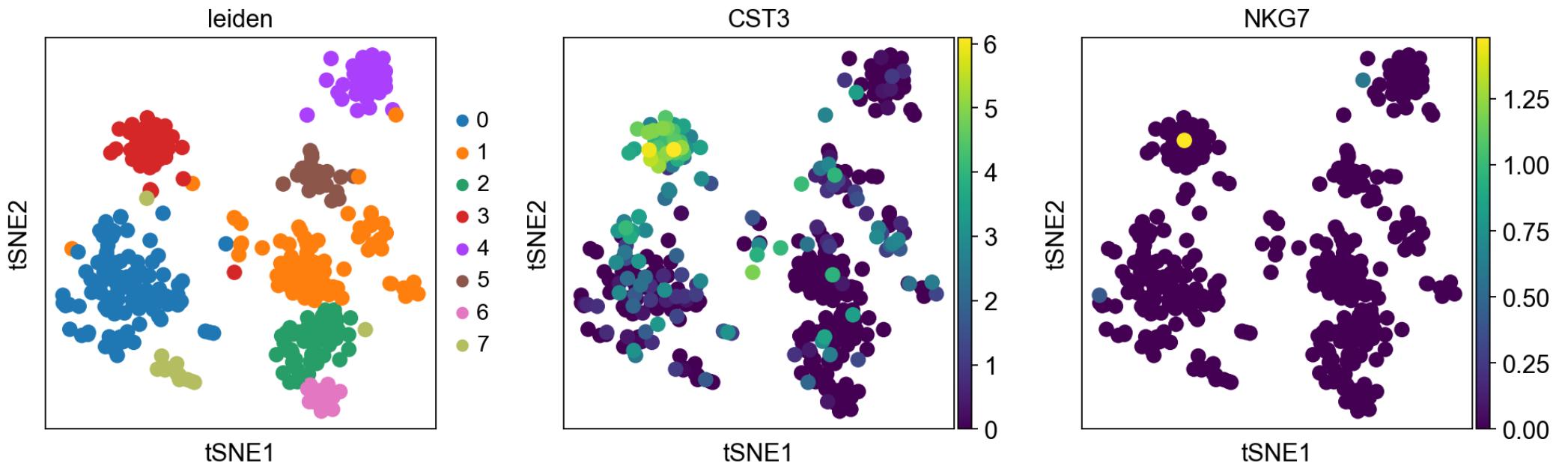
**Figure 33: The identification of highly variable genes and their distribution.** After selecting the top 4000 genes with less than 10% mitochondrial count. The figure shows a plot with a normalised and not normalised gene expression dispersion (variance) plotted against mean expression. The highly variable genes are highlighted as black dots.



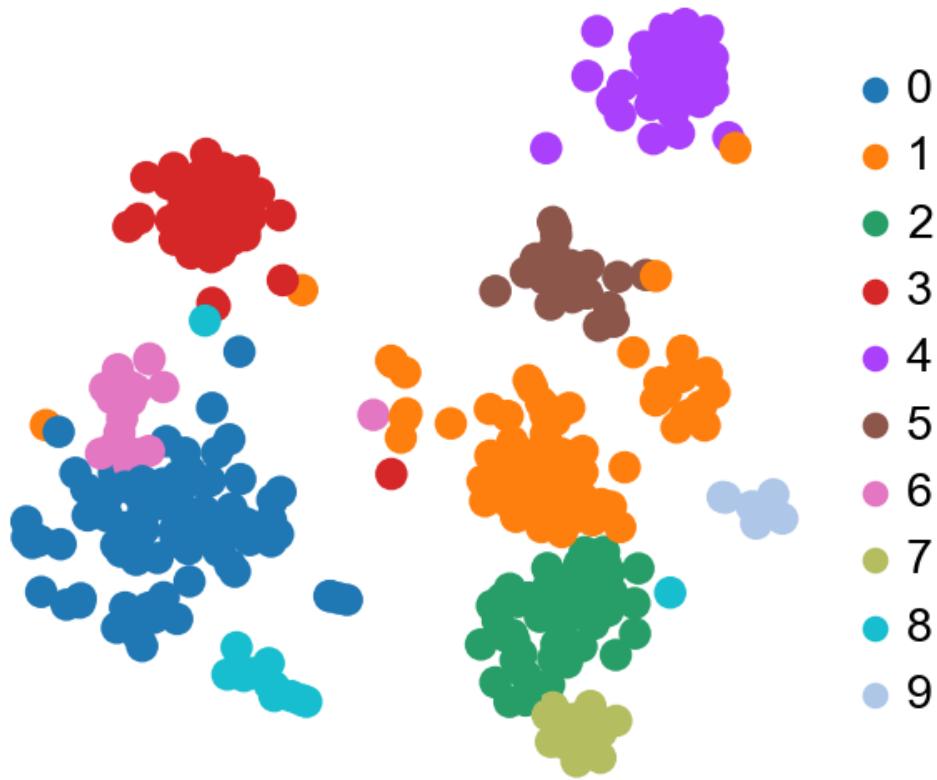
**Figure 34: Results of PCA on highly variable genes shown by a scatterplot.** In the figure, each point in the scatter plot represents a cell with the colour of the data points representing the expression level of the gene Cystatin C (CST3) in that cell. The scale explains the expression levels of each ranging from 0 to 6, with the highest expression level represented by the colour yellow and the lowest the colour purple.



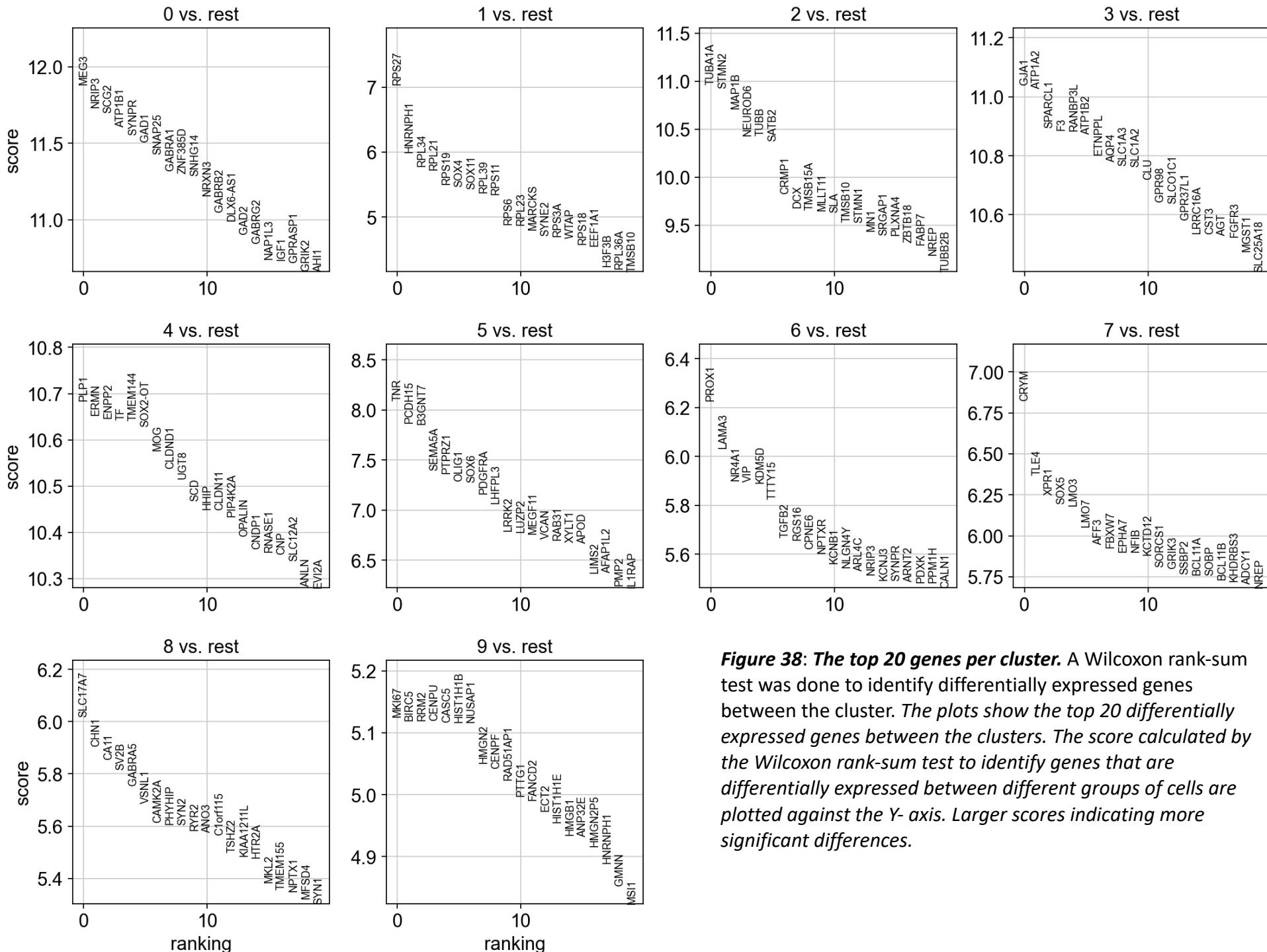
**Figure 35: The total variance in the data explained by each principal component in PCA.** In this figure, the x-axis represents the number of principal components, whilst the y-axis represents the proportion of variance explained by each PC on a logarithmic scale. PC1, PC2, PC3, PC4, etc. refer to the principal components 1, 2, 3, 4, etc., respectively. Principal components are usually ordered from left to right in terms of their contribution to the total variance in the data. PC1 is the leftmost component and has the highest variance, where PC31 explains the least variance.



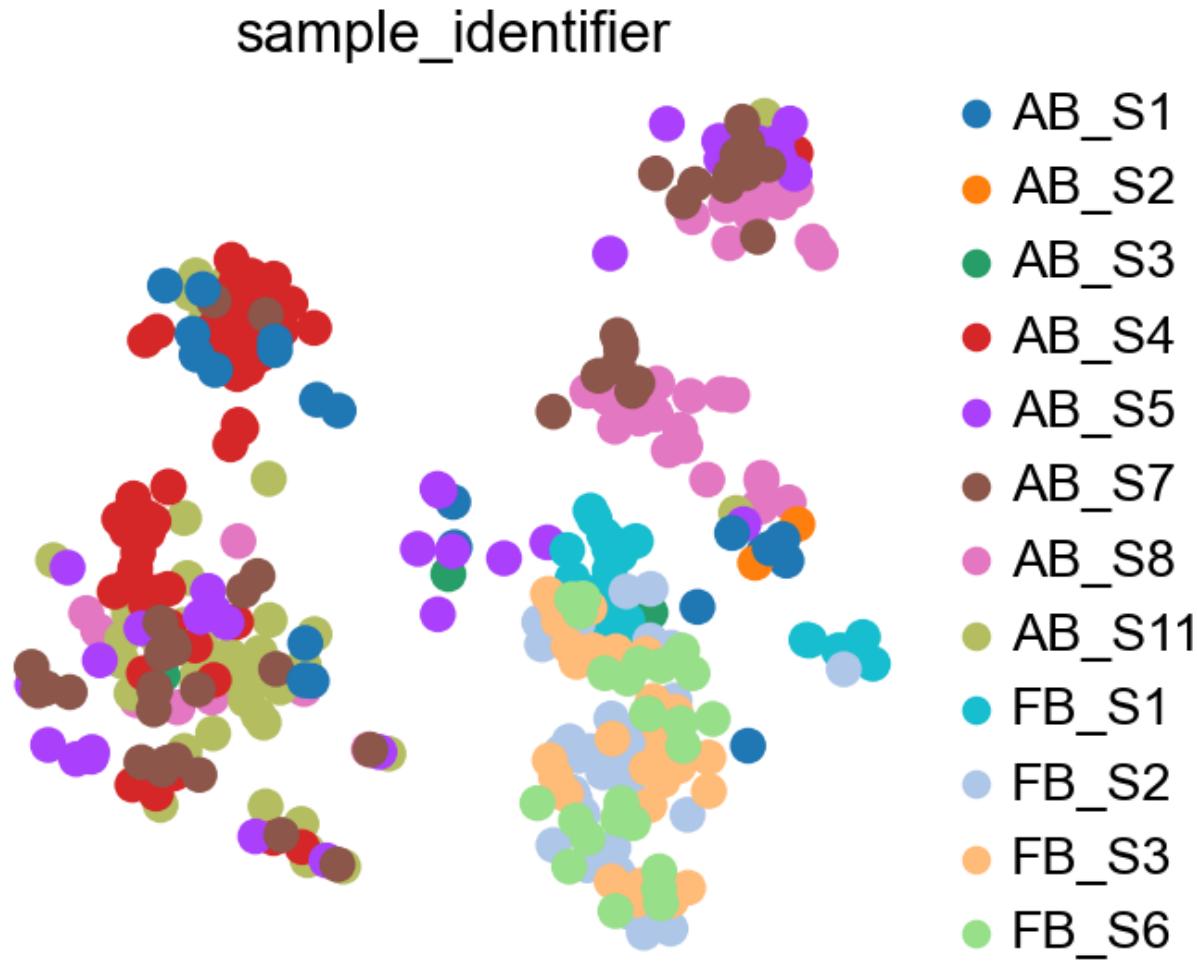
**Figure 36: t-SNE dimensionality reduction plot.** In this figure, the t-SNE dimensionality reduction plot shows the distribution of cells in the reduced space based on the expression levels of the top 40 principal components. Each dot represents a cell. The dots are coloured according to the expression level of the genes CST3 (in the middle) and NKG7 (on the right). The scale explains the expression levels of each ranging from 0 to 6 and 0 to 1.25, with the highest expression level represented by the colour yellow and the lowest by the colour purple. The plot on the right shows a t-SNE plot coloured by the Leiden clustering results. The Leiden algorithm was used to cluster cells based on the similarity in their gene expression profiles. Each point in the figure on the right represents a cell, and their colours indicate its cluster assignment, and 8 clusters were identified.



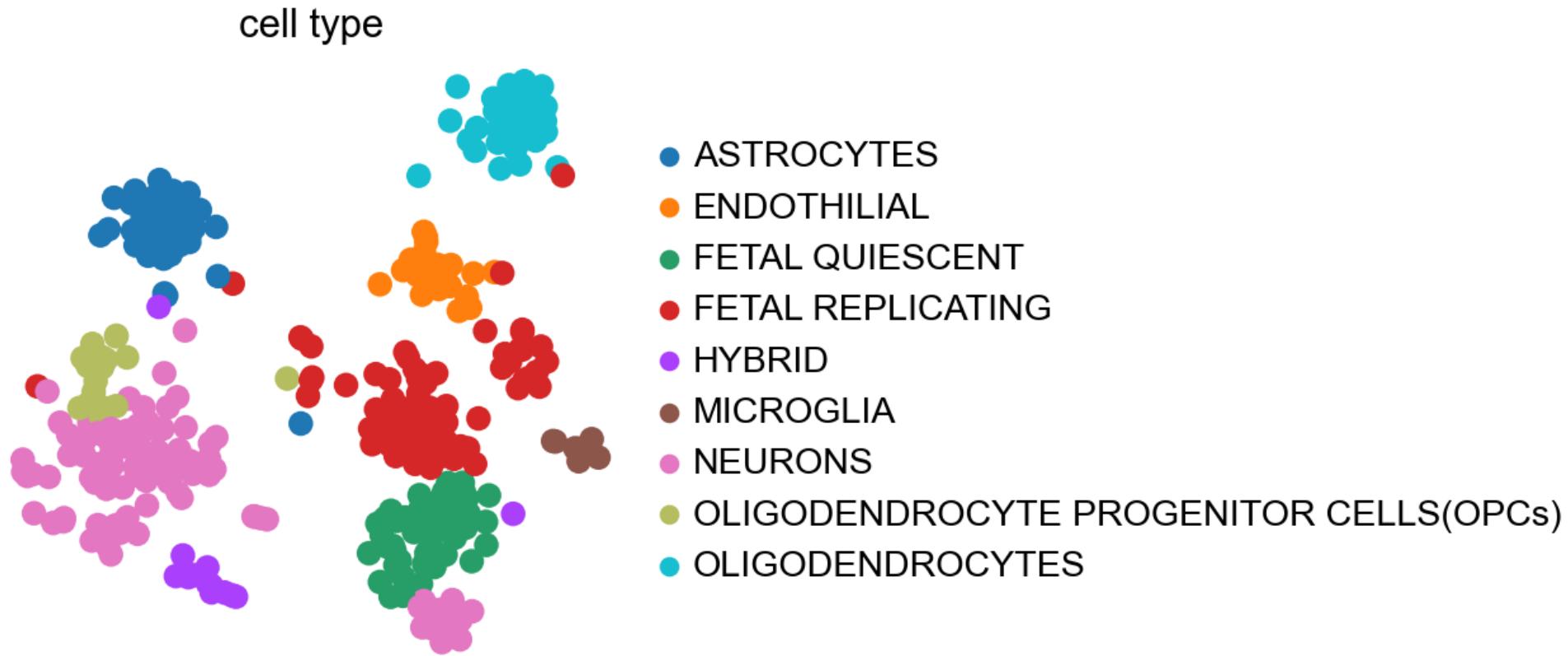
**Figure 37: Clustering results obtained from the Leiden algorithm.** This figure is the result of performing clustering using the Leiden algorithm with a resolution value of 1.3. The clustering results are visualised in a two-dimensional t-SNE plot. Data points are represented as a dot, coloured according to the cluster assignment. 10 clusters were identified in total.



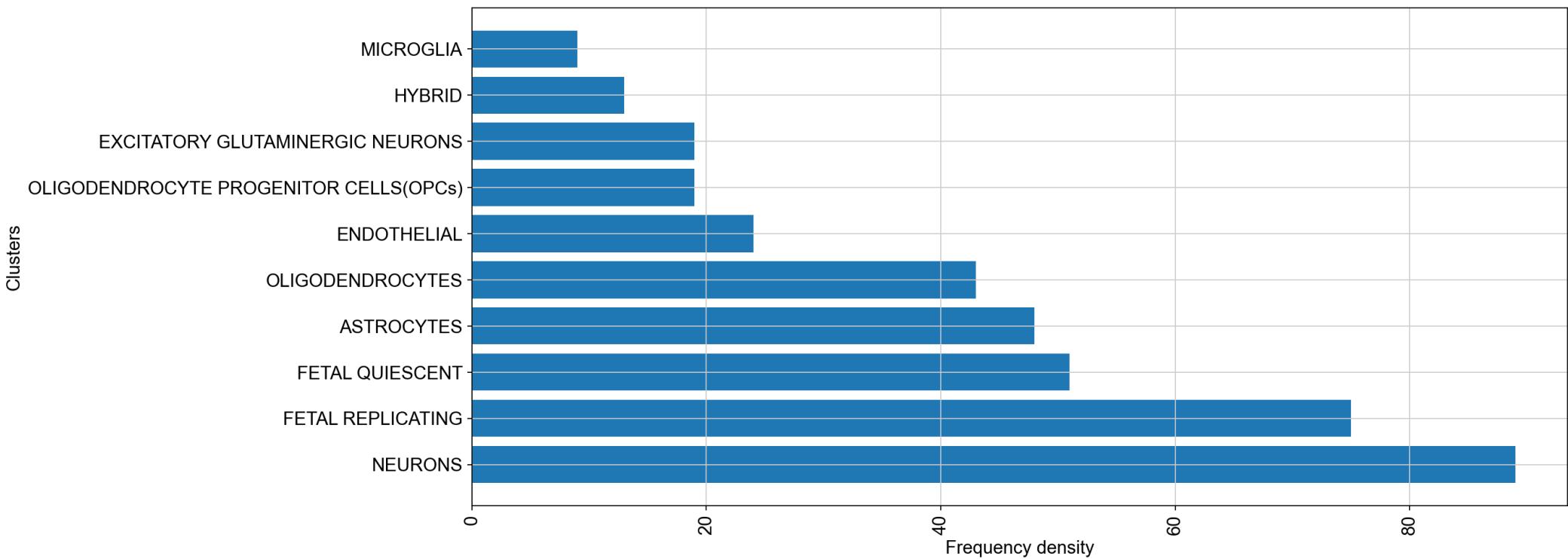
**Figure 38: The top 20 genes per cluster.** A Wilcoxon rank-sum test was done to identify differentially expressed genes between the cluster. The plots show the top 20 differentially expressed genes between the clusters. The score calculated by the Wilcoxon rank-sum test to identify genes that are differentially expressed between different groups of cells are plotted against the Y- axis. Larger scores indicating more significant differences.



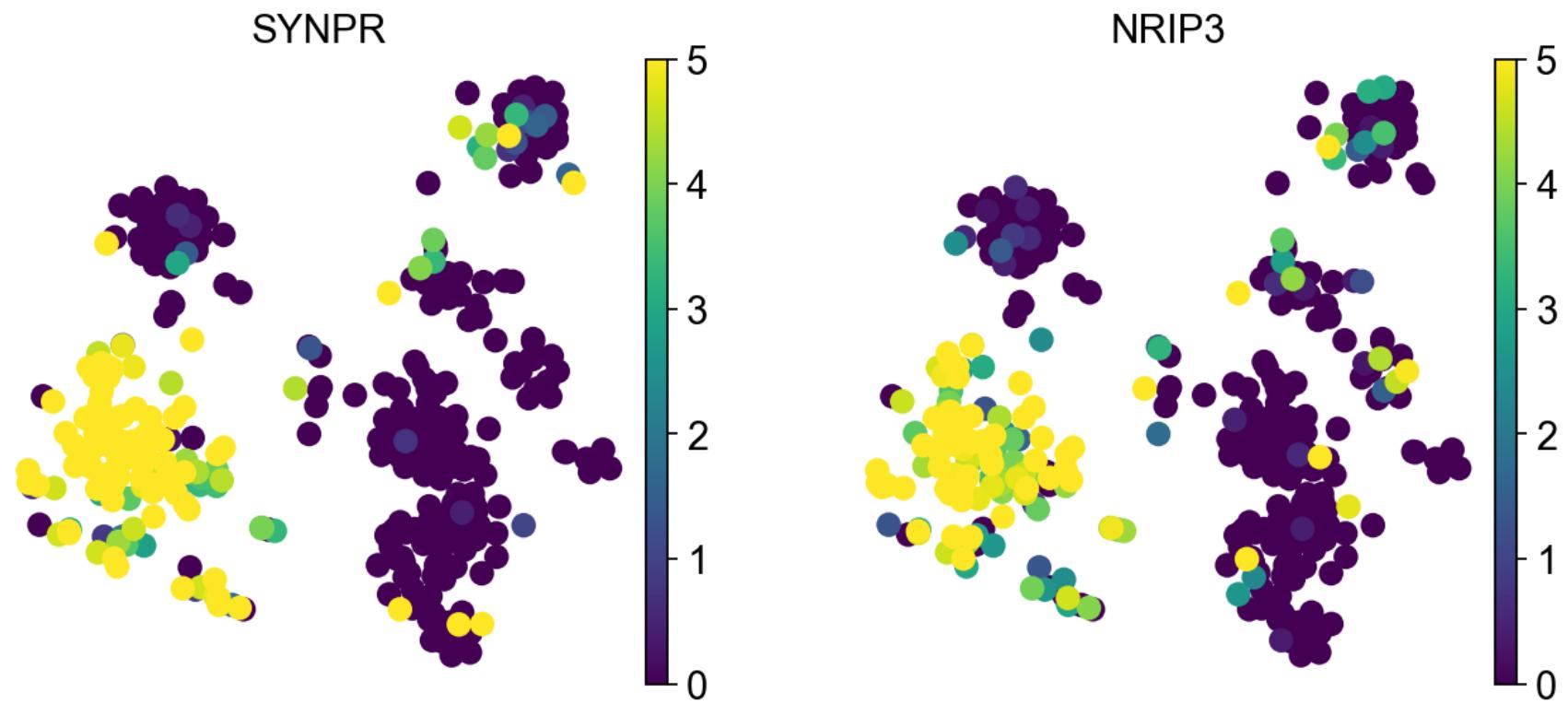
**Figure 39: t-SNE plot distinguishing between adult and fetal samples.** This plot is useful to identify samples via Leiden clustering, distinguishing between each adult and fetal brain surgical sample. In this figure, each dot represents a cell, and the colour corresponds to the sample in the count matrix it was taken from. The plot was generated using the Scanpy Python library. Adult brain surgical samples are represented as AB\_S1, AB\_S2, AB\_S3, AB\_S4, AB\_S5, AB\_S7, AB\_S8 and AB\_S11. Fetal brain surgical samples are represented by FB\_S1, FB\_S2, FB\_S3 and FB\_S6.



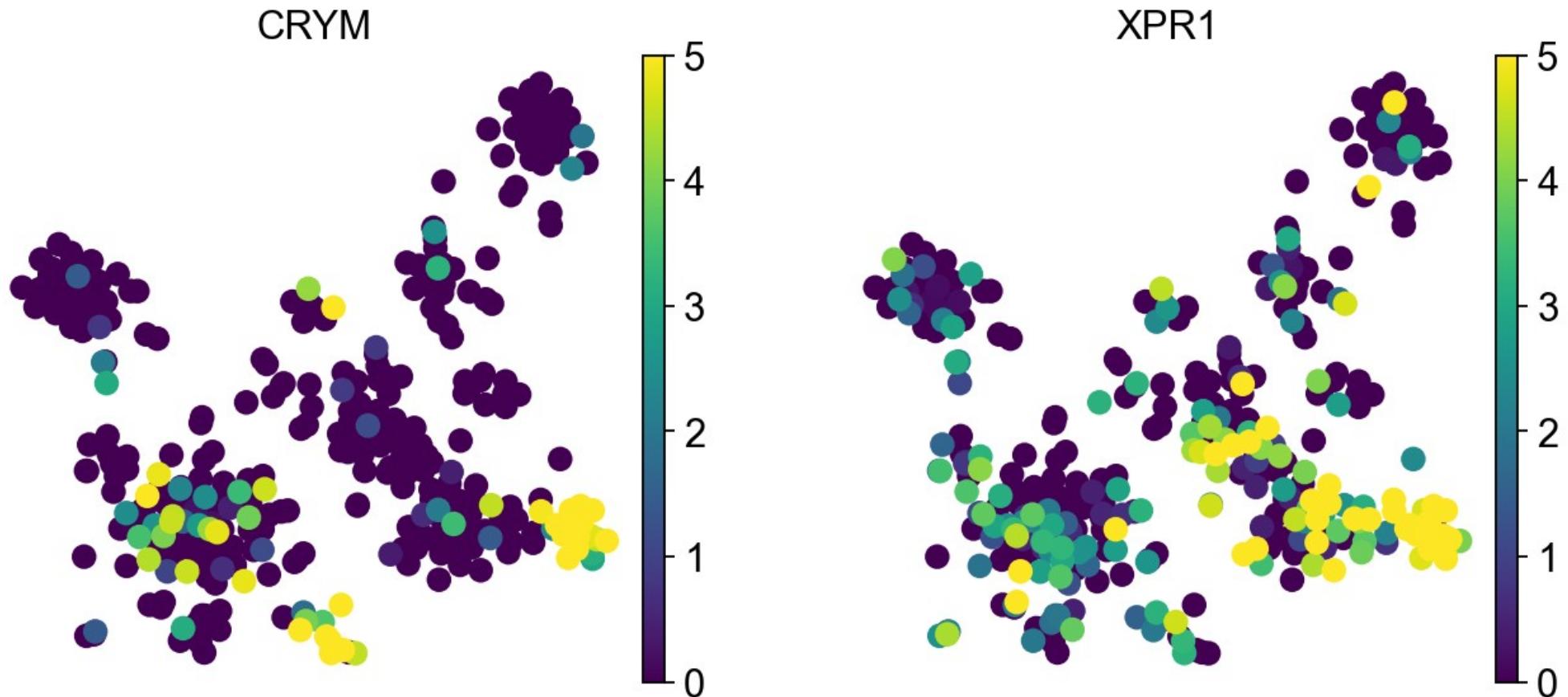
**Figure 40: t-SNE plot of sc-RNA seq data.** In this figure, the plot shows the t-SNE results of sc-RNA seq analysis. Each dot represents a single cell coloured according to its assigned cell type, which is indicated by the key to the right of the plot.



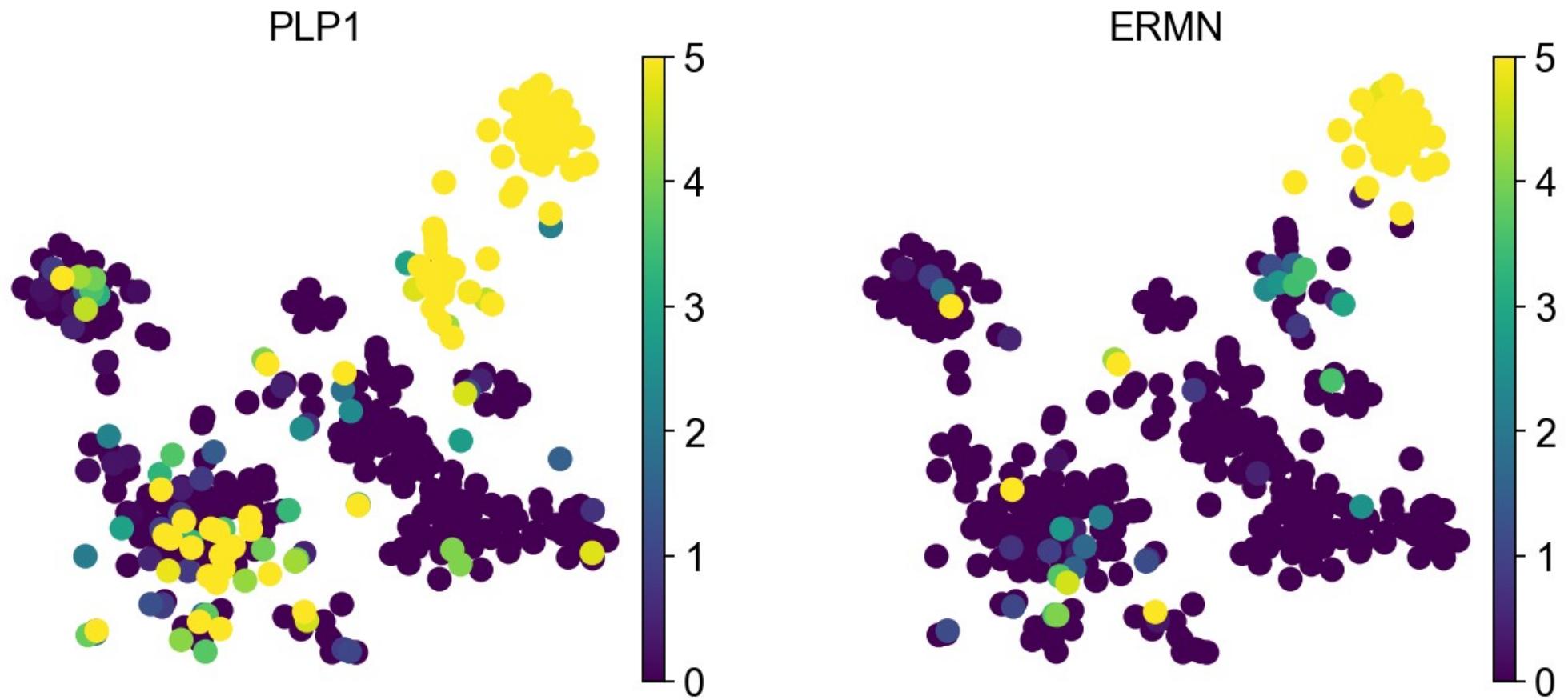
**Figure 41: t-SNE histogram showing the distribution of cell types.** The Y-axis represents a cell type, and the length of the bar reflects the frequency density of cells belonging to that cell type/cluster as shown on the X- axis.



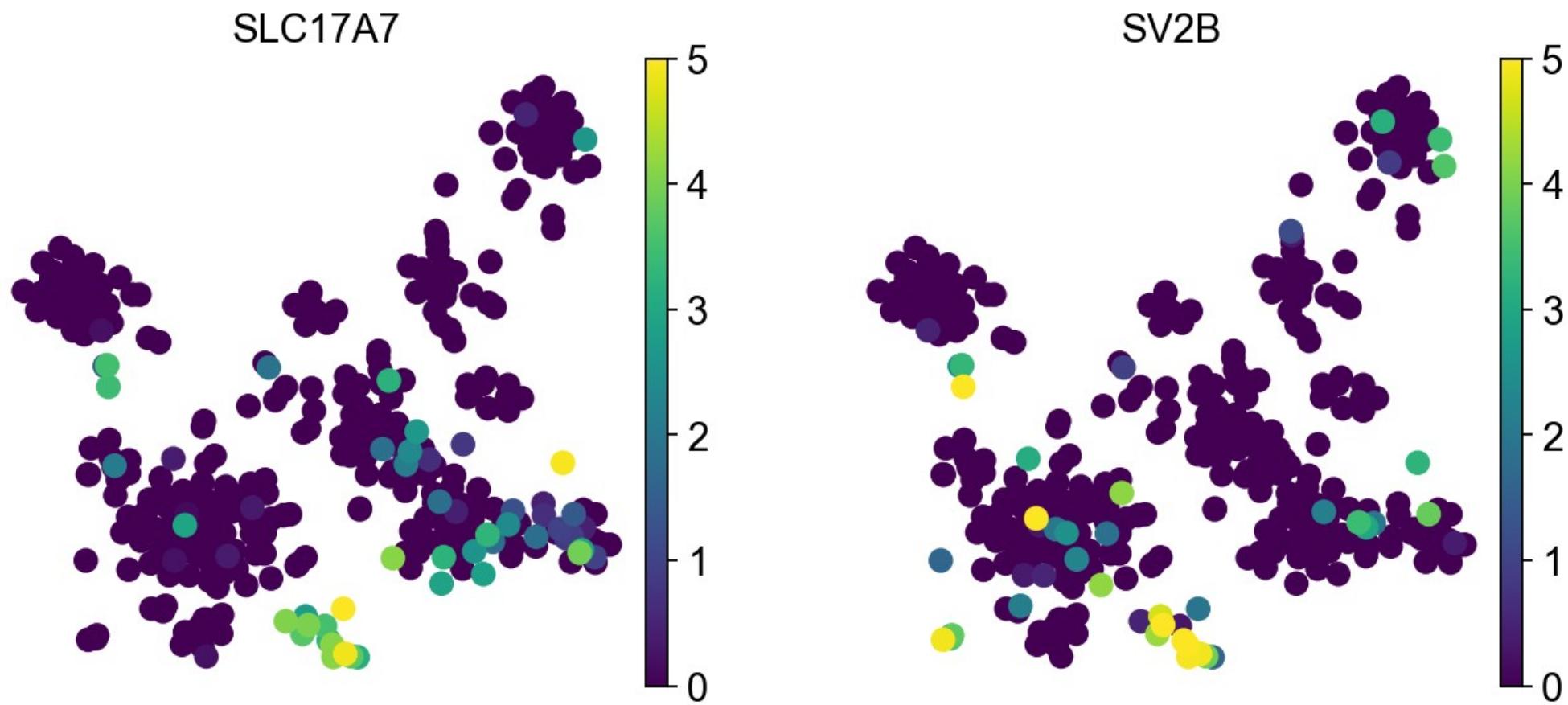
**Figure 42: t-SNE expression of genes SYNPR and NRIP3 in Neuron cell cluster.** This figure represents the distribution of the marker genes based on their expression levels. The dots are coloured based on their expression levels of the genes SYNPR (on the left) and NRIP3 (on the right). These expressions of genes are from the neuron cell cluster. The scale describes the range of expression for both, from 0 to 5, with yellow being the highest expression and purple the lowest.



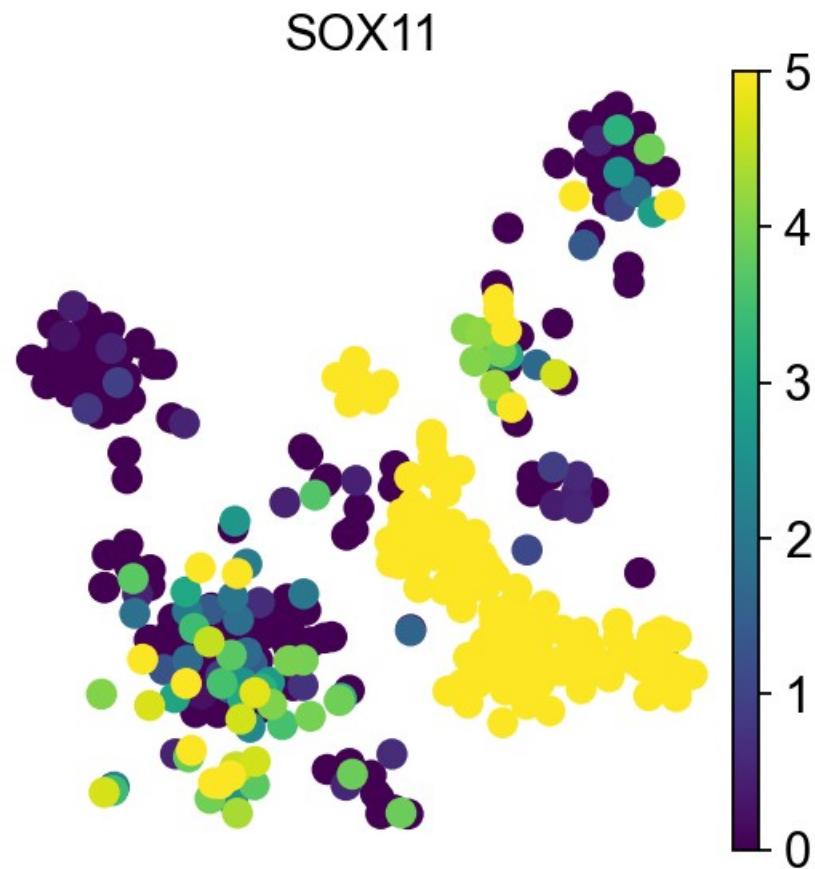
**Figure 43: t-SNE expression of genes CRYM and XPR1 in Neuron cell cluster.** This figure represents the distribution of the marker genes based on their expression levels. The dots are coloured based on their expression levels of the genes CRYM (on the left) and XPR1 (on the right). These expressions of genes are from the cell cluster. The scale describes the range of expression for both, from 0 to 5, with yellow being the highest expression and purple the lowest.



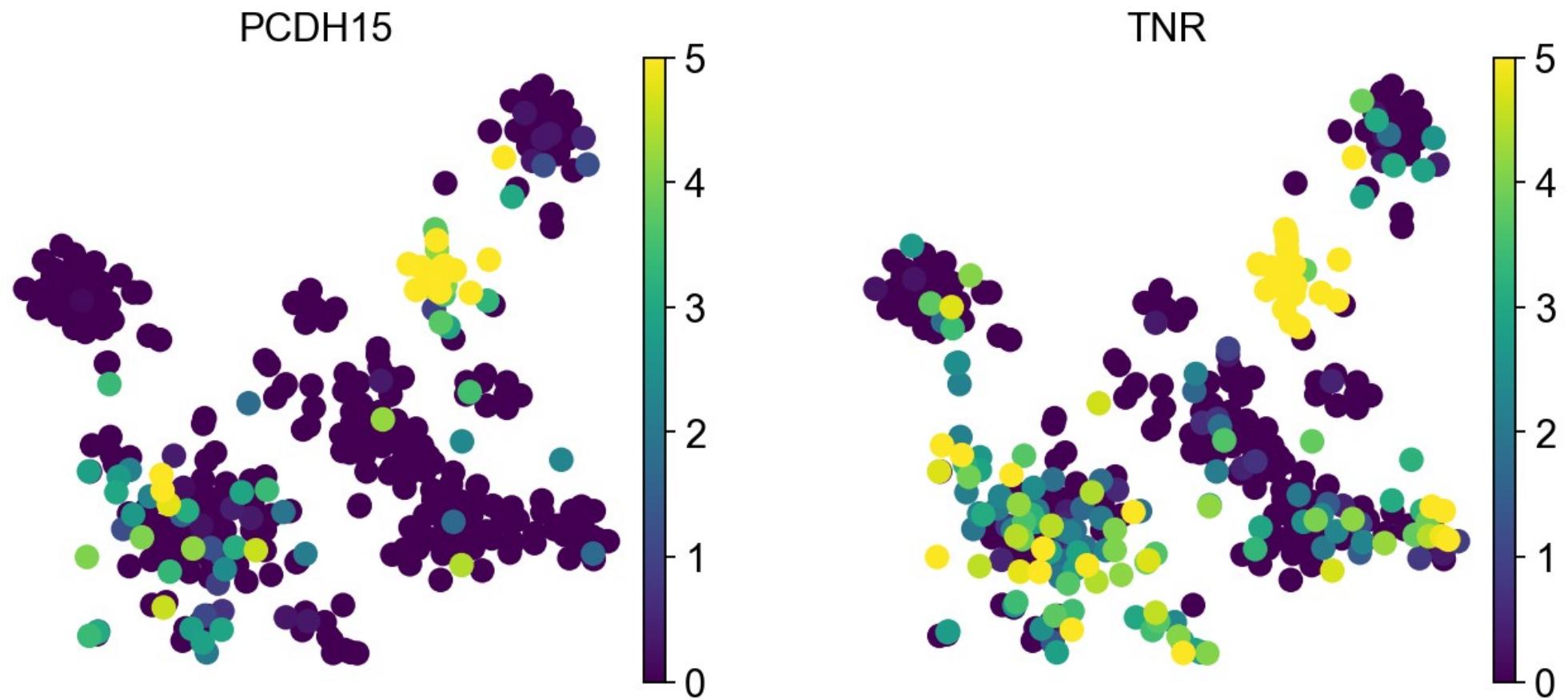
**Figure 44: t-SNE expression of genes PLP1 and ERMN in Oligodendrocytes cell cluster.** This figure represents the distribution of the marker genes based on their expression levels. The dots are coloured based on their expression levels of the genes PLP1 (on the left) and ERMN (on the right). These expressions of genes are from the cell cluster. The scale describes the range of expression for both, from 0 to 5, with yellow being the highest expression and purple the lowest



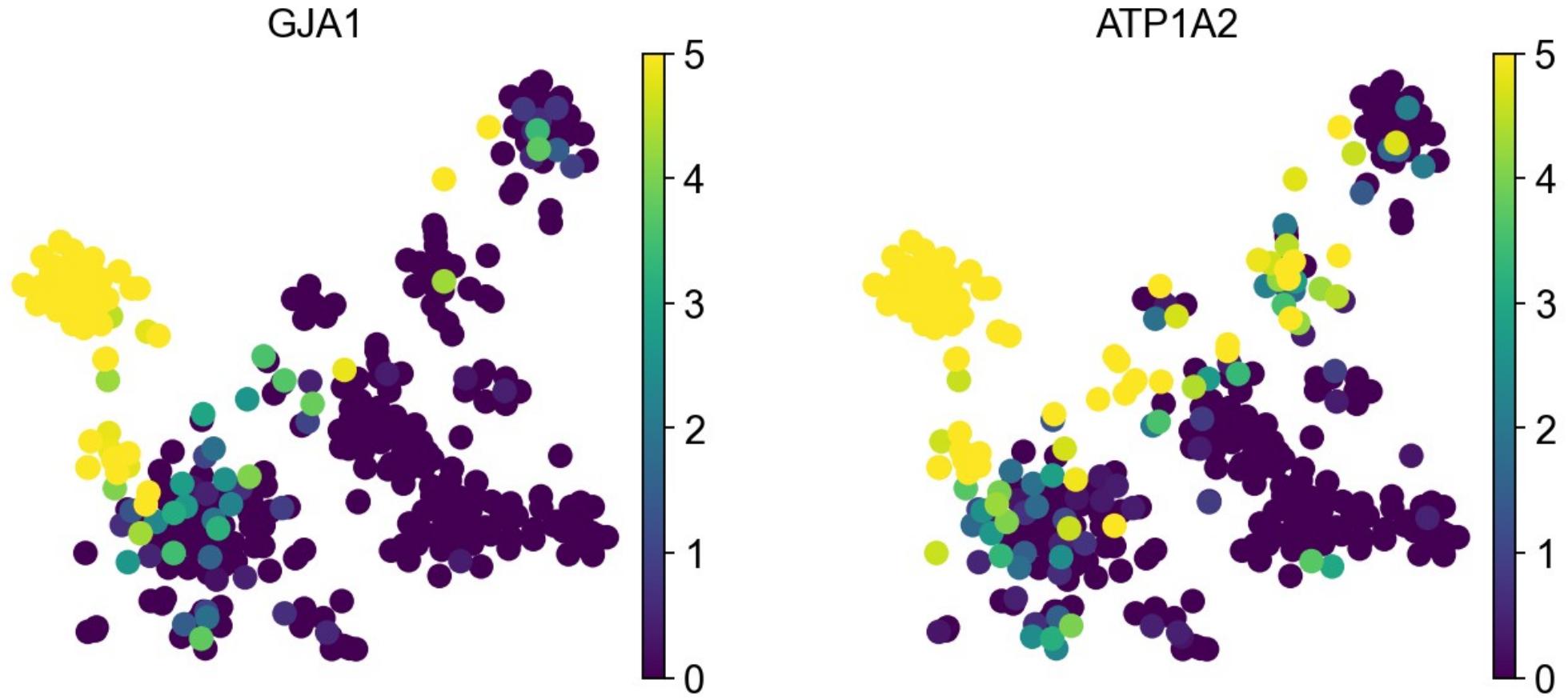
**Figure 45:** t-SNE expression of genes *SLC17A7* and *SV2B* in hybrid cell cluster. This figure represents the distribution of the marker genes based on their expression levels. The dots are coloured based on their expression levels of the genes *SLC17A7* (on the left) and *SV2B* (on the right). These expressions of genes are from the cell cluster. The scale describes the range of expression for both, from 0 to 5, with yellow being the highest expression and purple the lowest.



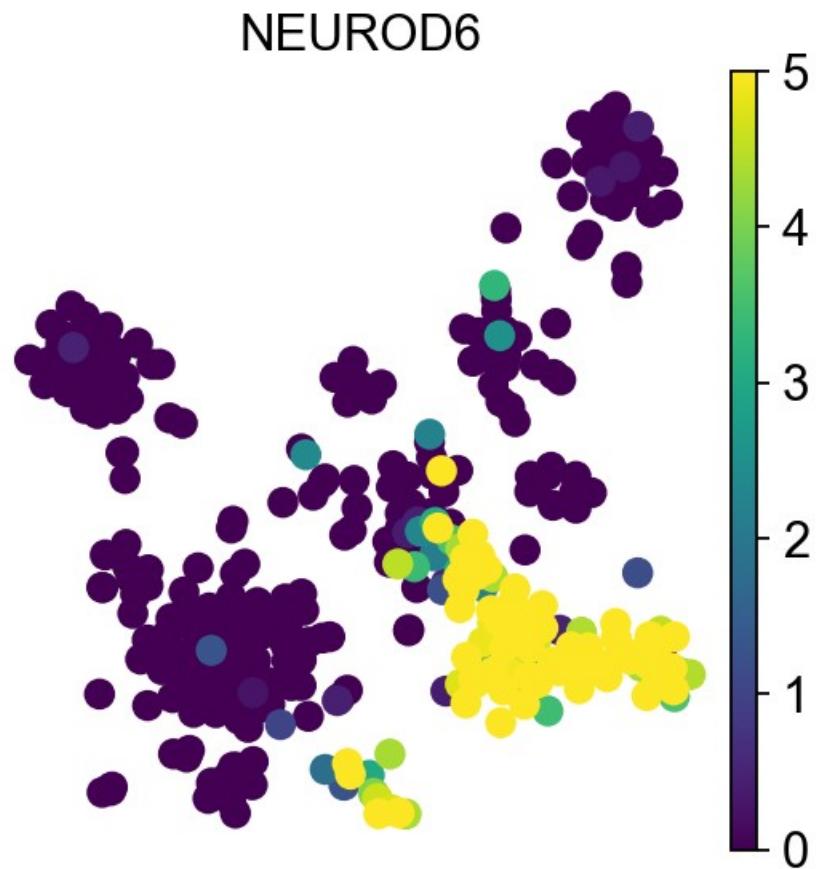
**Figure 46: t-SNE expression of genes SOX11 in Fetal replicating cluster.** This figure represents the distribution of the marker genes based on their expression levels. The dots are coloured based on their expression levels of the gene SOX11. These expressions of genes are from the fetal replicating cluster. The scale describes the range of expression for both, 0 to 5, with yellow being the highest expression and purple the lowest.



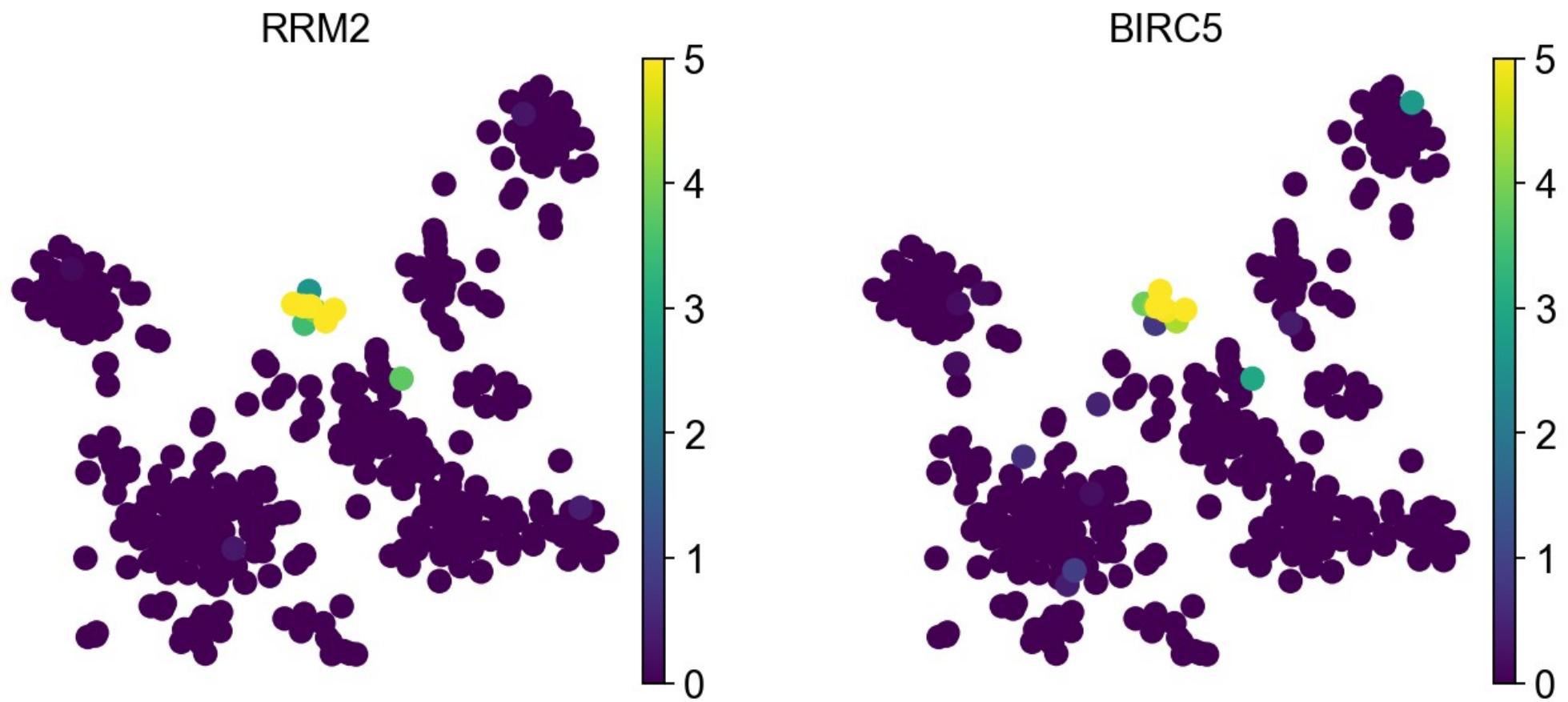
**Figure 47: t-SNE expression of genes PCDH15 and TNR in the endothelial cell cluster.** This figure represents the distribution of the marker genes based on their expression levels. The dots are coloured based on their expression levels of the genes PCDH15 (left) and TNR (right). These expressions of genes are from the endothelial cell cluster. The scale describes the range of expression for both, 0 to 5, with yellow being the highest expression and purple the lowest.



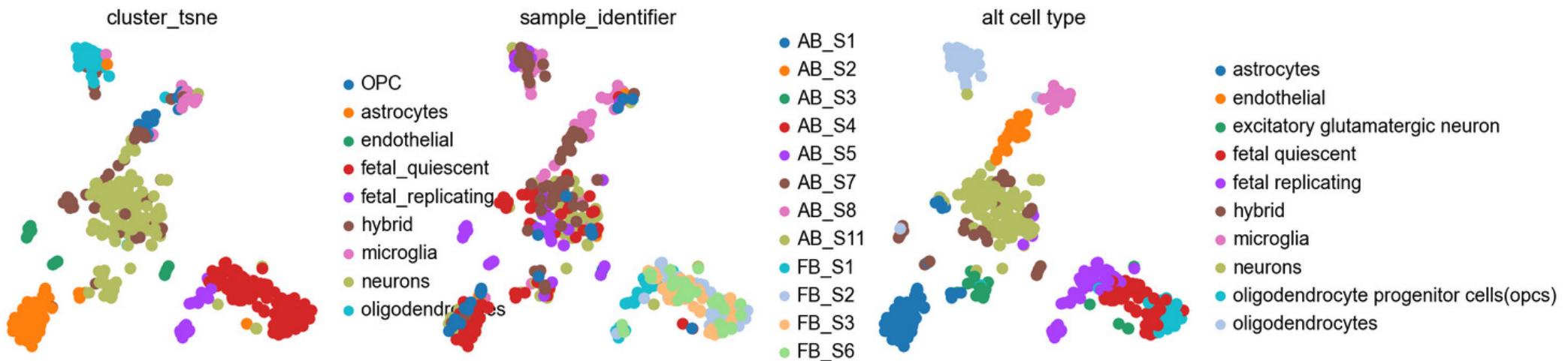
**Figure 48: t-SNE expression of genes GJA1 and ATP1A2 in the Astrocytes cell cluster.** This figure represents the distribution of the marker genes based on their expression levels. The dots are coloured based on their expression levels of the genes GJA1 (left) and ATP1A2 (right). These expressions of genes are from the astrocyte cell cluster. The scale describes the range of expression for both, 0 to 5, with yellow being the highest expression and purple the lowest.



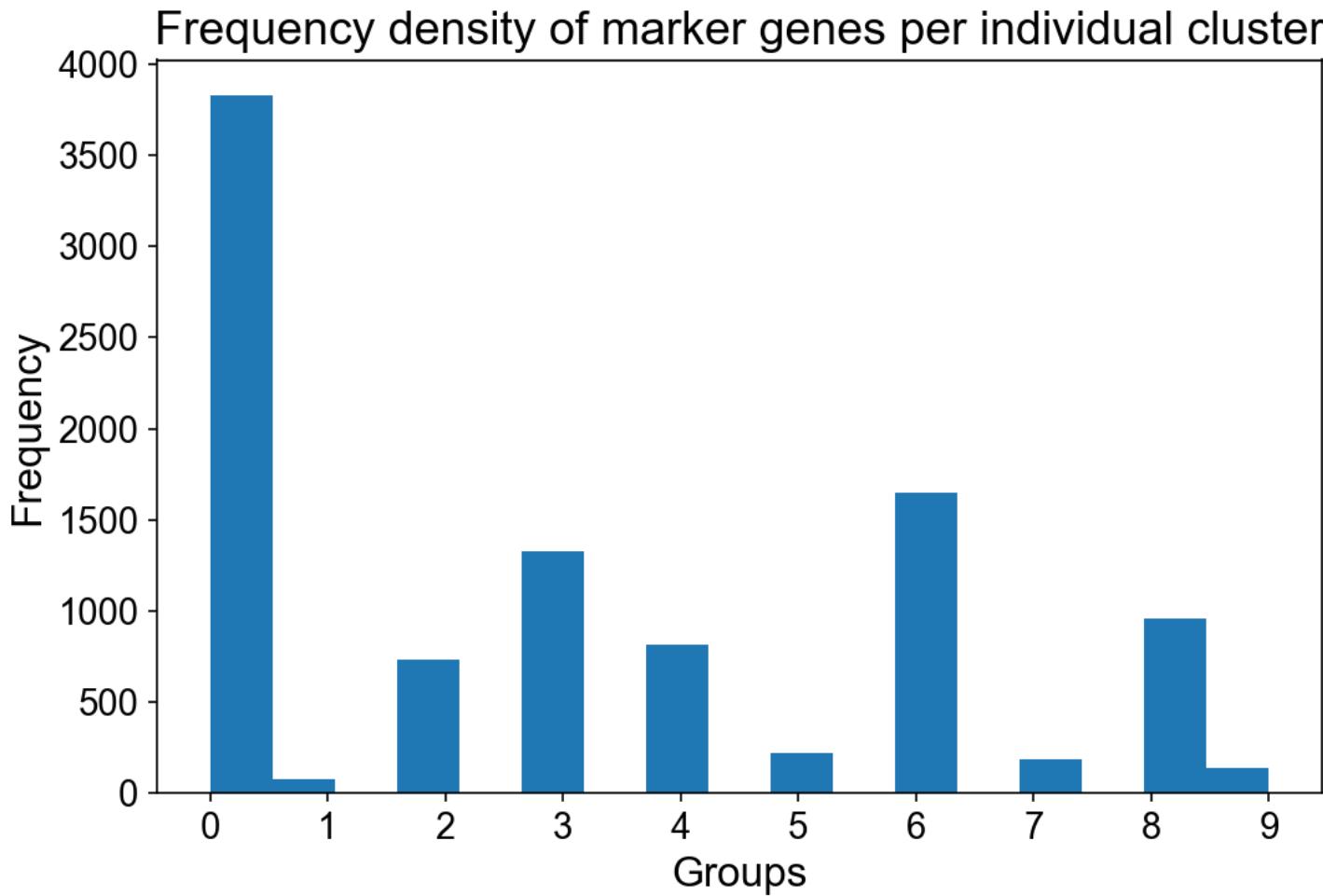
**Figure 49: t-SNE expression of genes NEUROD6 in Fetal quiescent cluster.** This figure represents the distribution of the marker genes based on their expression levels. The dots are coloured based on their expression levels of the gene NEUROD6. These expressions of genes are from the fetal quiescent cluster. The scale describes the range of expression for both, 0 to 5, with yellow being the highest expression and purple the lowest.



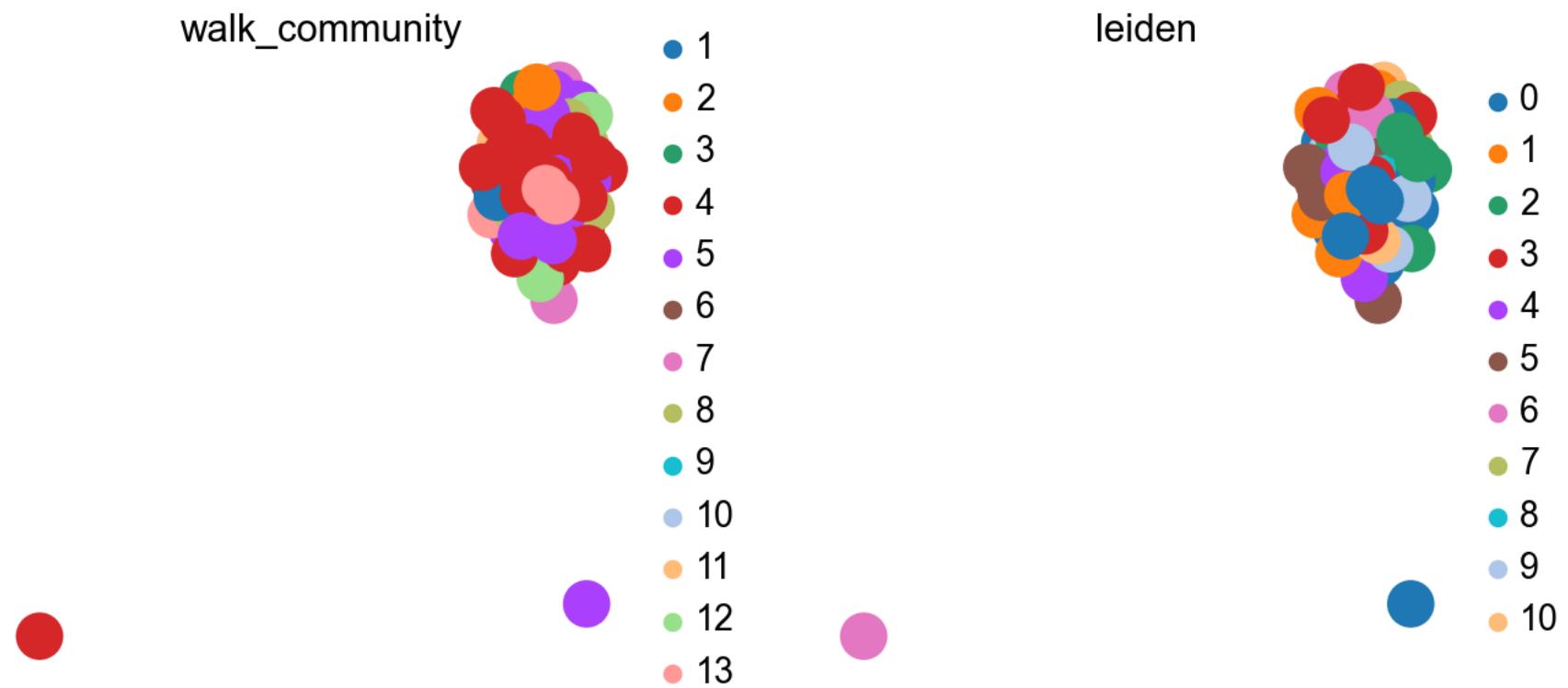
**Figure 50: t-SNE expression of genes RRM2 and BIRC5 in Microglia cell cluster.** This figure represents the distribution of the marker genes based on their expression levels. The dots are coloured based on their expression levels of the genes RRM2 (on the left) and BIRC5 (on the right). These expressions of genes are from the Microglia cell cluster. The scale describes the range of expression for both, 0 to 5, with yellow being the highest expression and purple the lowest.



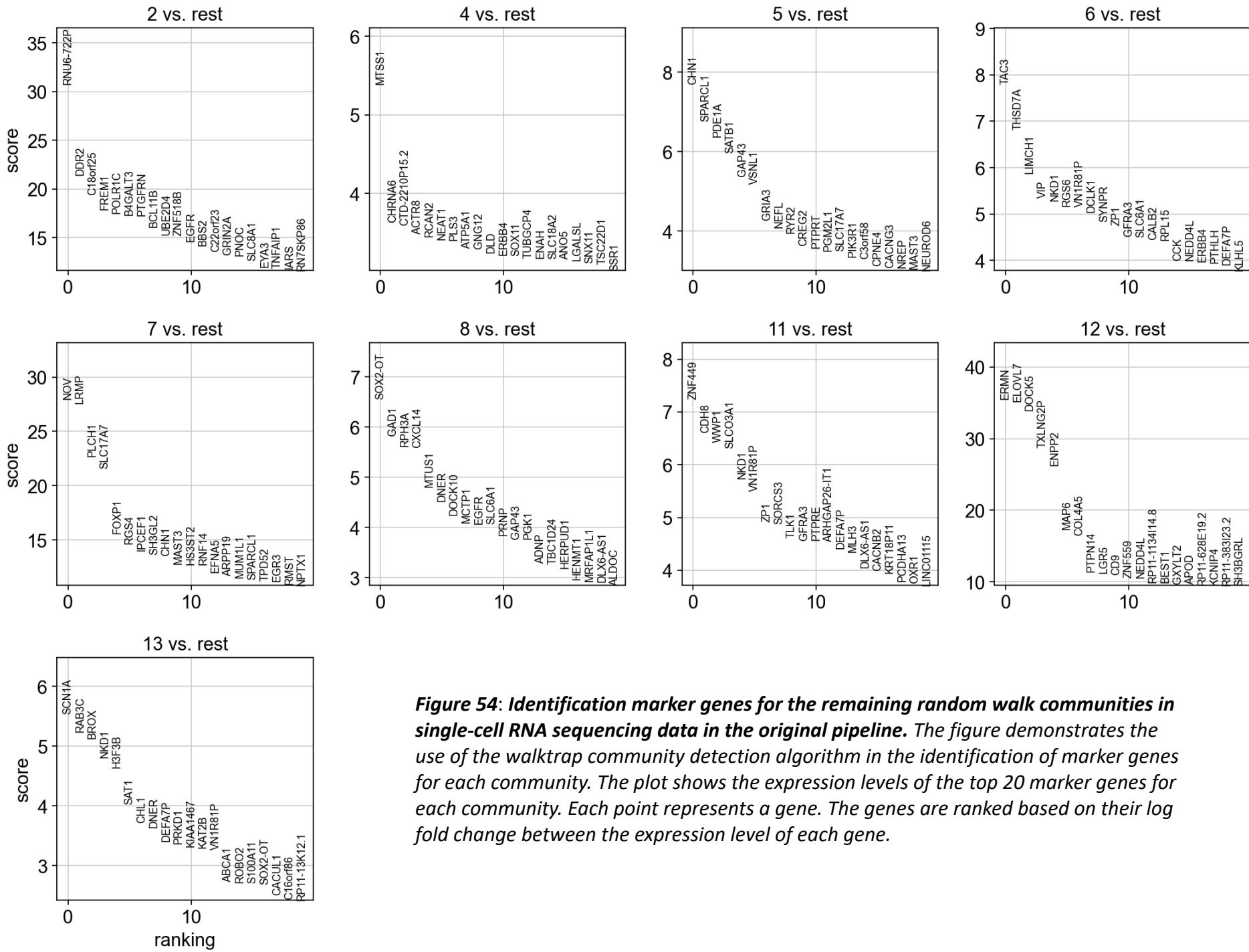
**Figure 51: Comparison of t-SNE across both pipelines.** When comparing the t-SNE results across the original and alternative pipelines, there is a high concordance between the mapping techniques.



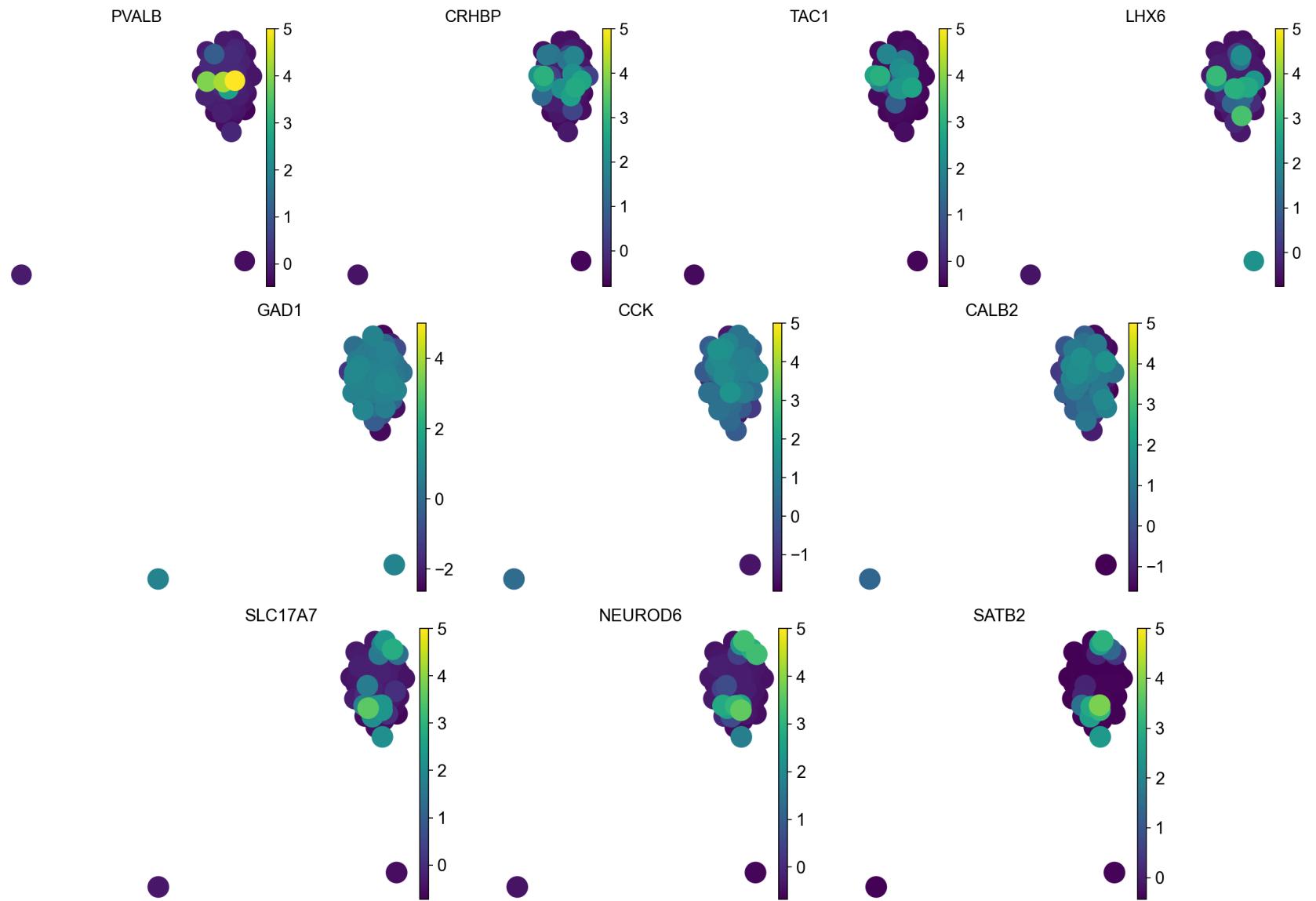
**Figure 52: Histogram of marker genes per individual cluster.** The Y-axis represents the frequency of density marker genes per cluster which is reflected by the height of the bar of each group. The X-axis is the groups of the 10 individual clusters that were identified.



**Figure 53: t-SNE plot shows the distribution cells and comparison of clustering methods using the Leiden and walktrap community detection algorithm in the original pipeline.** The graph was calculated based on 15 principal components with 30 nearest neighbours. Then, t-SNE was performed on the data. The Leiden algorithm was used to obtain clusters with a resolution of 1.2. The community structure of the dataset was also determined using the walktrap community detection algorithm. The plots show the distribution of cells in the two clustering methods. The cells are coloured based on their cluster assignment. The results of the two algorithms were generally consistent, but some differences can be observed in certain regions of the t-SNE plot. The walktrap algorithm detected more communities than the Leiden algorithm.



**Figure 54: Identification marker genes for the remaining random walk communities in single-cell RNA sequencing data in the original pipeline.** The figure demonstrates the use of the walktrap community detection algorithm in the identification of marker genes for each community. The plot shows the expression levels of the top 20 marker genes for each community. Each point represents a gene. The genes are ranked based on their log fold change between the expression level of each gene.



**Figure 55: t-SNE plot showing the expression levels of genes in distinct populations of excitatory, Inhibitory/Interneurons and GABAergic neurons communities.** The colour scale ranges from 0 to 5, with the highest expression level represented by the colour yellow and the lowest the colour purple. SLC17A7, NEUROD6, and SATB2 are marker genes for the excitatory community. PVALB, CRHBP, TAC1, LHX6, and SOX6 are all markers for the GABAergic neuron community and GAD1, CCK, and CALB2 are markers for inhibitory neurons.