Single cell RNA-seq 10x Genomics hands-on exercise

# By Gil Stelzer, June 2018

Loupe Cell Browser is a program created by 10x Genomics for visualizing Cell Ranger output. The 10x Genomics Cell Ranger is a pipeline that processes raw sequencing data (using the cellranger count program). This includes demultiplexing the libraries based on sample indices and converting the barcode and read data to FASTQ files. Thereupon, alignment is performed (using STAR and the relevant reference sequence), followed by filtering and unique molecular identifier (UMI) counting. Lastly, aggregation (using cellranger aggr) of several samples is accomplished by taking outputs from multiple runs, normalizing these runs to the same sequencing depth, recomputing the gene-barcode matrices and analyzing on the combined data.

## QC report (outs\web\_summary.html)

In the first part of the exercise, we will view a cellranger-generated QC (quality control) report

High as possible

>90%

>90%

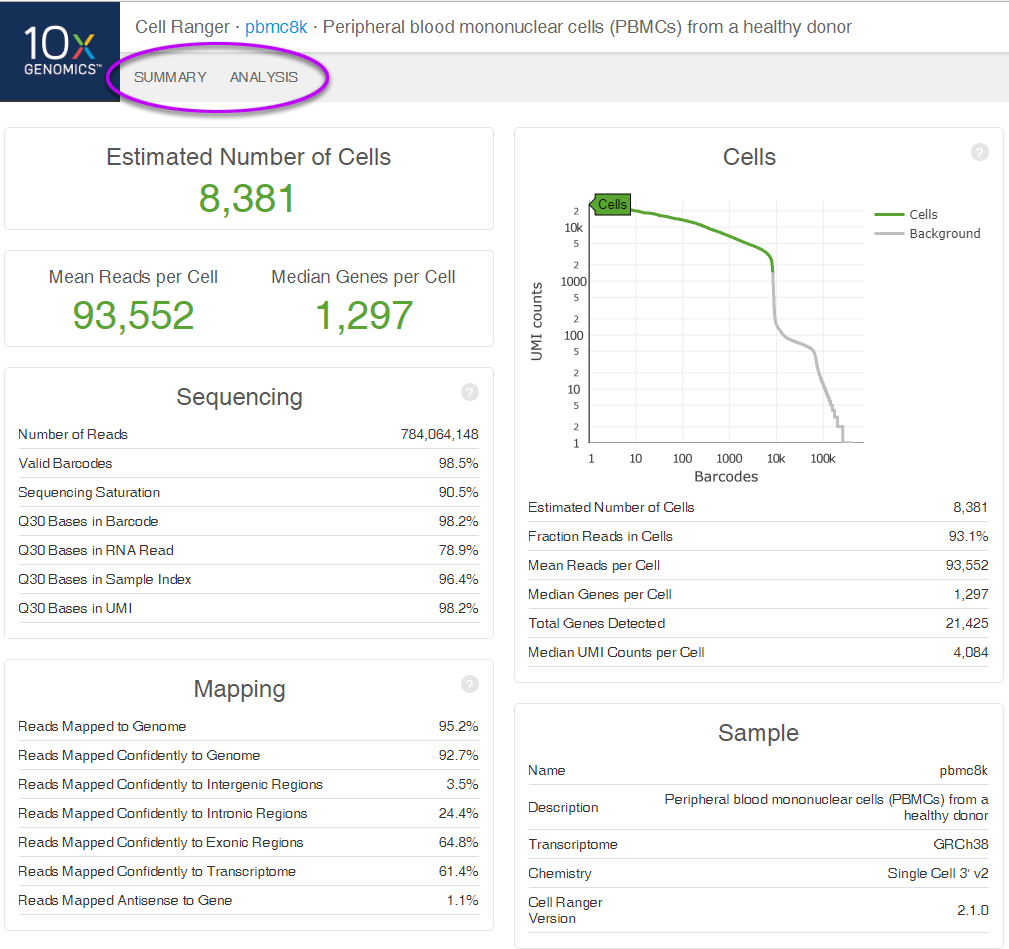
>90%

>70%

>80%

>75%

>30%



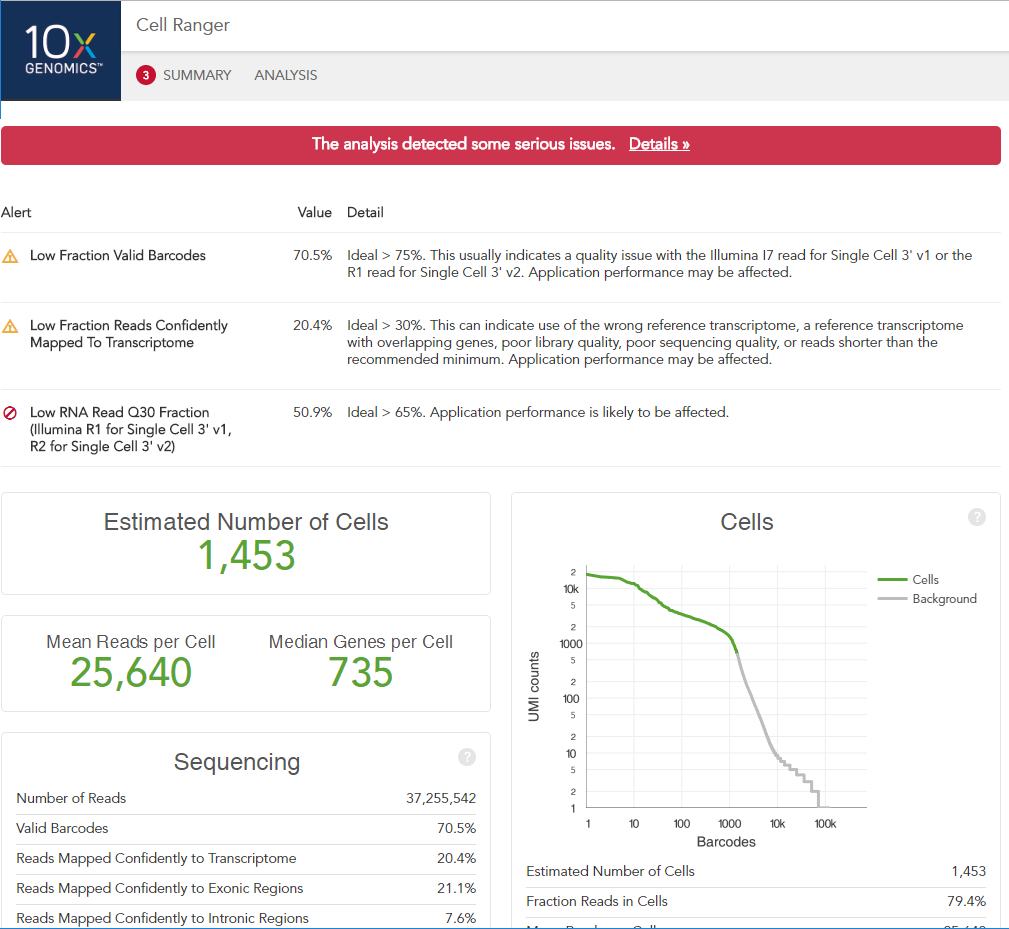
Valid barcodes - should be above %75

Reads Mapped Confidently to Transcriptome – if lower than 30% then the problem could be wrong reference transcriptome used for mapping, a reference transcriptome with overlapping genes, poor library quality, poor sequencing quality or reads shorter than the recommended minimum.

Sequencing saturation - describes how many reads are duplicates, based on UMI tags. If the sequencing is 90% saturated, it means that for every 10 reads that are sequenced, 9 are UMI duplicates. If the sequencing saturation is below 80% then it might be beneficial to sequence the sample more deeply to get additional UMI counts.

Q30 for R1 (Barcode and UMI) and I7 (Sample index) should be above 90%. Q30 for R2 (RNA read \ transcript) is usually less than 90% but above 70%.

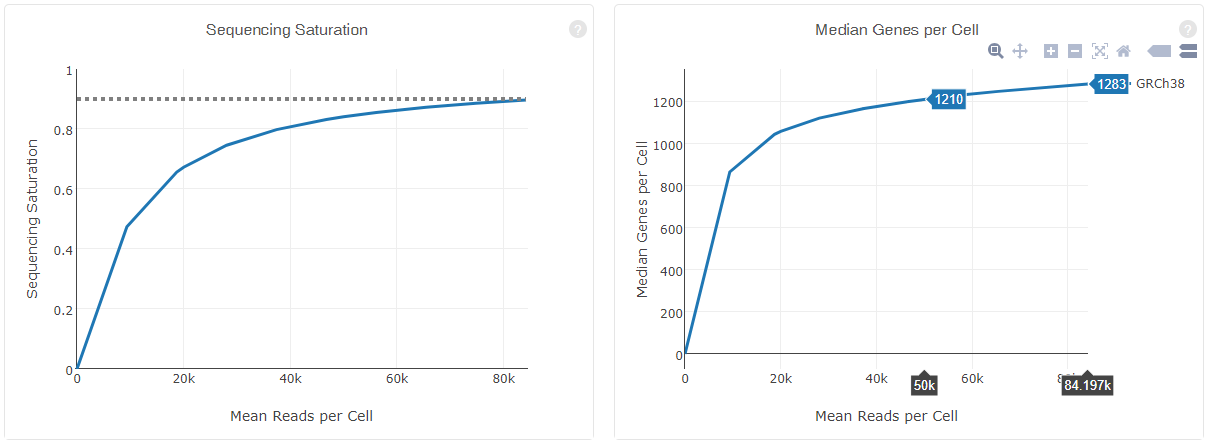
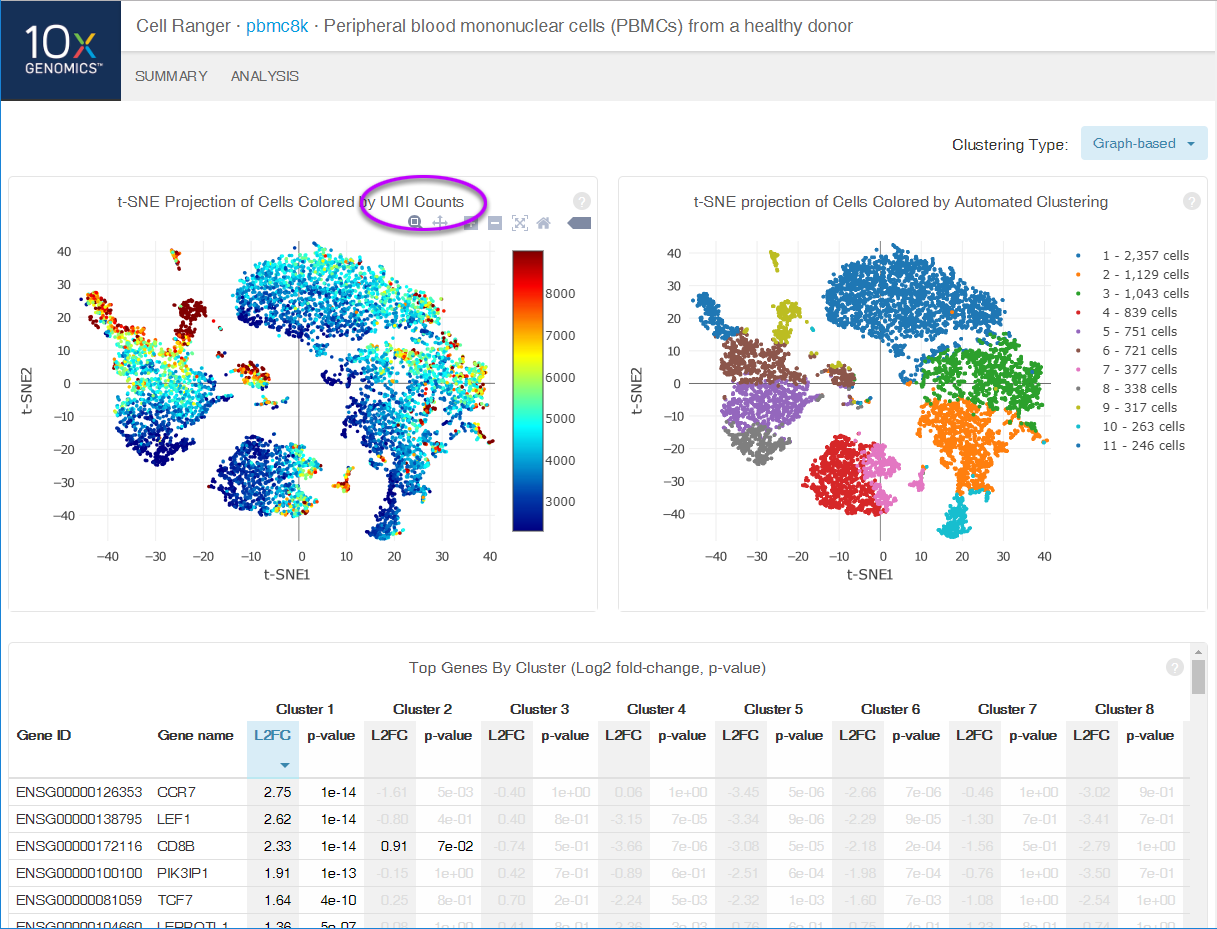
Fraction reads in cells - fraction of barcoded confidently mapped reads with cell-associated barcodes, derived from total cell associated UMI’s / total UMI’s. The higher this percentage, the lower (premature) cell lysis has occurred and the cleaner the data will be, i.e. less ambient RNA.



Problem during library creation such as ambient RNA and more.

## Analysis

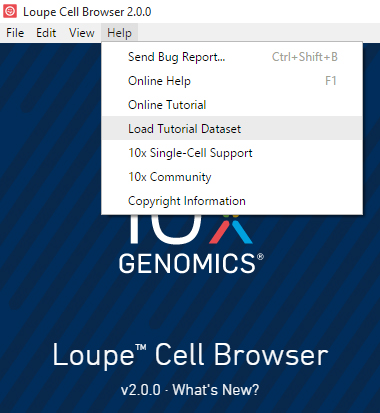
The analysis tab displays t-SNE plots, the left plot shows UMI distribution, whilst the right plot can be viewed also in the Loupe Cell Browser (which enables interactivity)



These graphs allow you plan the future experiment with the same type of samples. The plot on the right demonstrates that we could reduce sequencing depth and receive similar results.

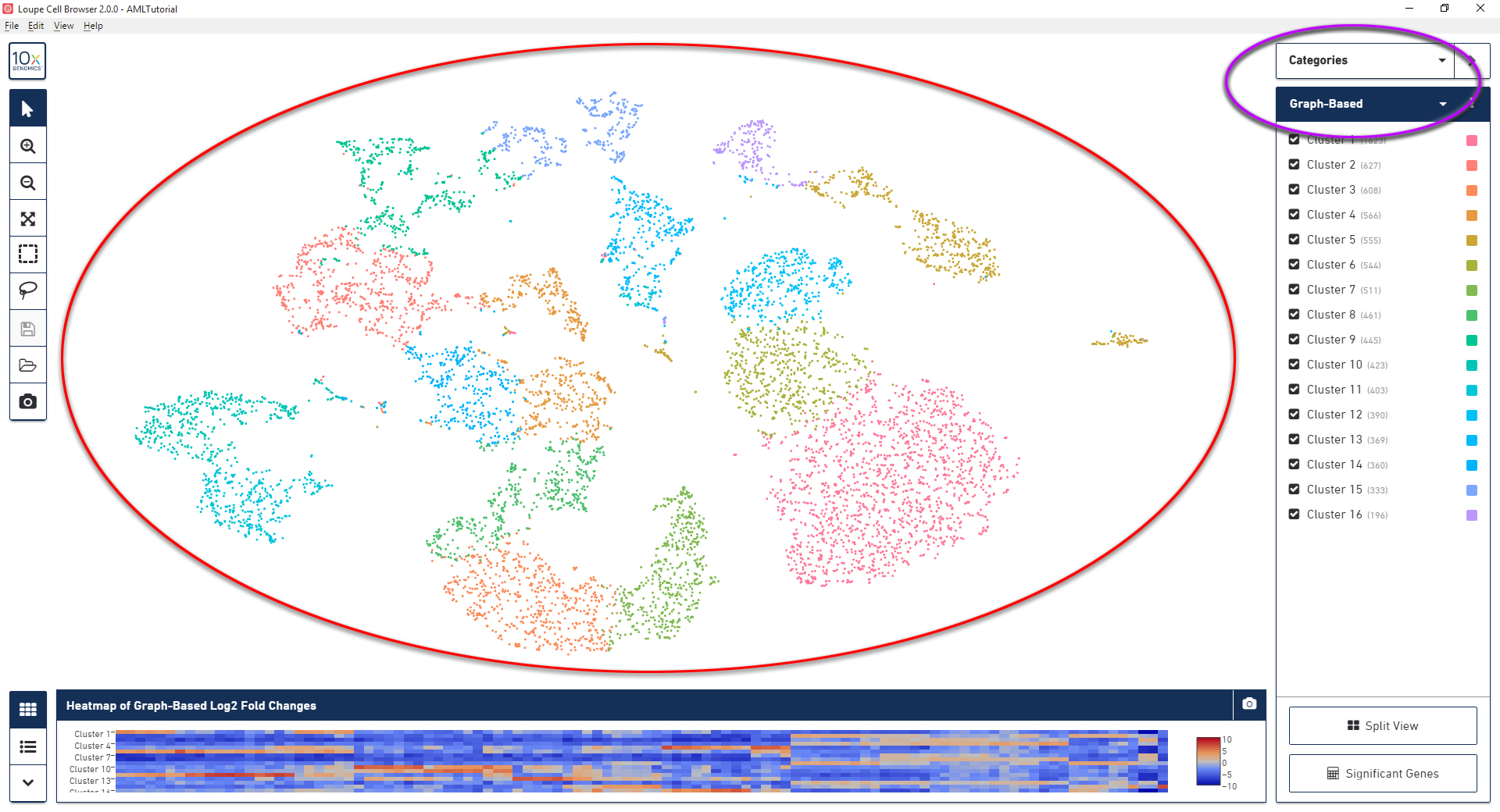
## Loupe Cell Browser

In the second part of the exercise we will demonstrate how to use the Loupe Cell Browser on an acute myeloid leukemia (AML) dataset. Open the Loupe Cell Browser found on your computer (local program). Click on 'Load Tutorial' from the Help menu in order to load the AMLTutorial file.



This dataset contains the results of a cellranger aggr run over three samples: two healthy control samples of frozen human bone marrow mononuclear cells, and a pre-transplant sample from a patient with acute myeloid leukemia (AML). This dataset was generated in collaboration with the Fred Hutchinson Cancer Research Center, and referenced in the Nature Communications publication, "Zheng et al, Massively parallel digital transcriptional profiling of single cells" (2017; doi:10.1038/ncomms14049).

Cell Ranger produces the two-dimensional scatter plot appearing in the center (marked in a red ellipse) by applying t-SNE dimensionality reduction on the most significant gene vectors using principal component analysis (using the first 10 PC’s). Each point represents a single barcode, the vast majority of which represent a single cell. By default, the Categories mode and Graph-Based clustering are selected (purple ellipse).



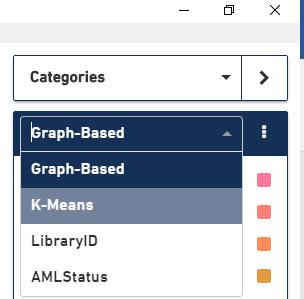
Briefly, the [Graph-Based clustering](https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/algorithms/overview#louvain) algorithm uses successive steps in order to cluster, starting by building a nearest-neighbor graph followed by Louvain Modularity Optimization and then cluster-merging by hierarchical clustering.

1. How many cells does cluster 5 include?

Right click cluster 13 and select “Edit color”. Change the color to RGB (140, 20, 170).

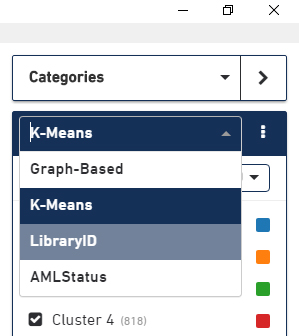
1. What percentage of cells does cluster 13 constitute (hover over the cell count)?

Select [K-Means](https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/algorithms/overview#louvain) clustering from the clustering selection menu.



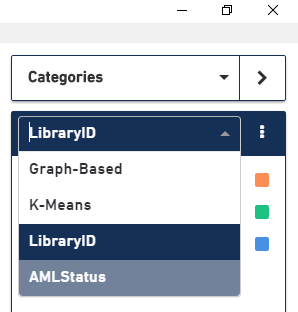
1. Which cluster (depicted as a certain color as determined by K-Means=10) is split \ scattered the most in two dimensions?
2. Change the default K-value to 6. Cluster number 10 (using K=10) is now merged into a larger cluster called 5 (using K=6). With which cluster(s) was it merged from K=10?

Select the LibraryID clustering



1. Are the normal samples good biological replicates? Explain your answer.

Select the AMLStatus clustering

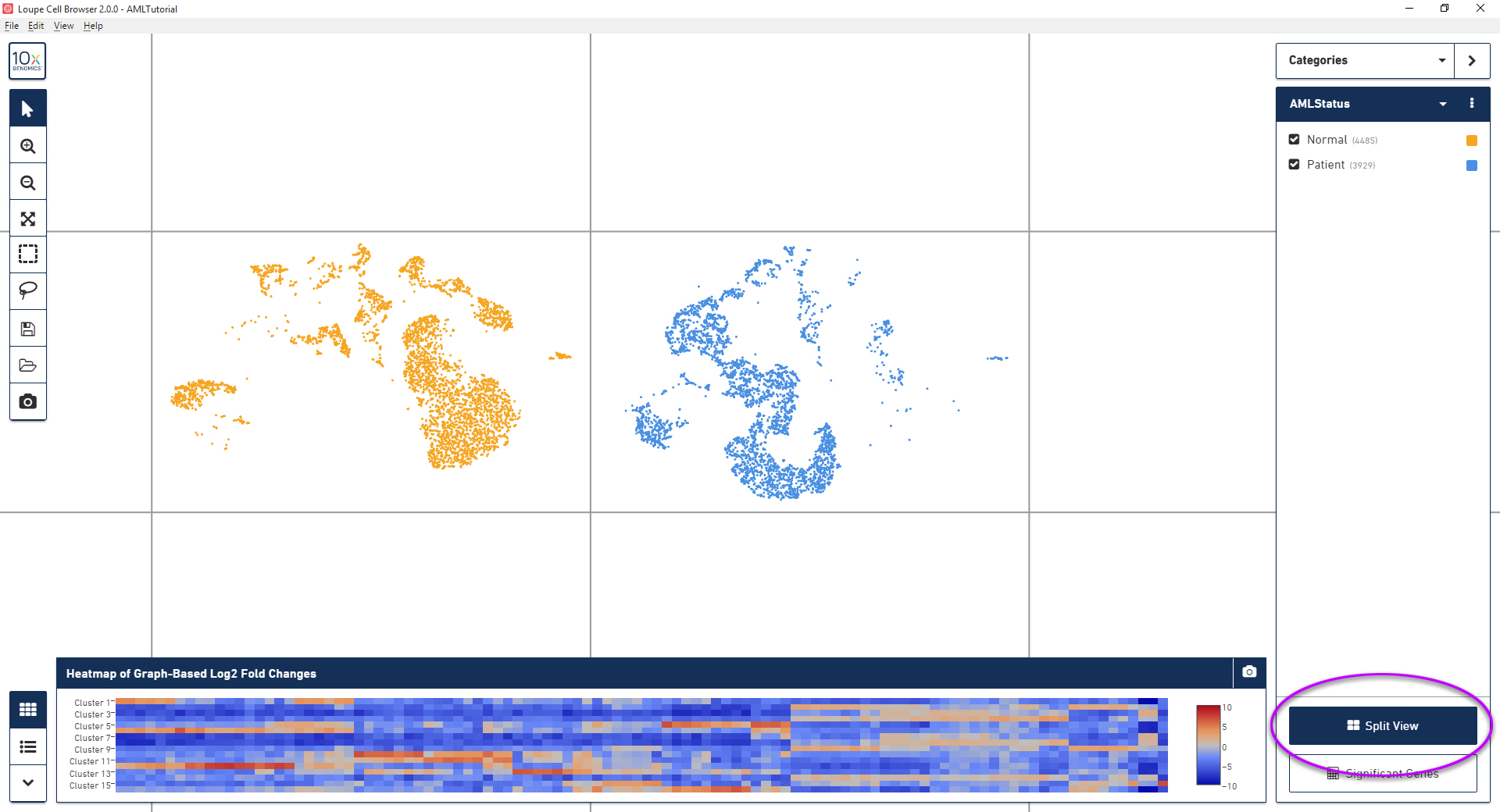


In this view, the samples are split into Normal vs. Patient.

We can see that there is an overlap in the gene expression between the patient and normal in some of the clusters.



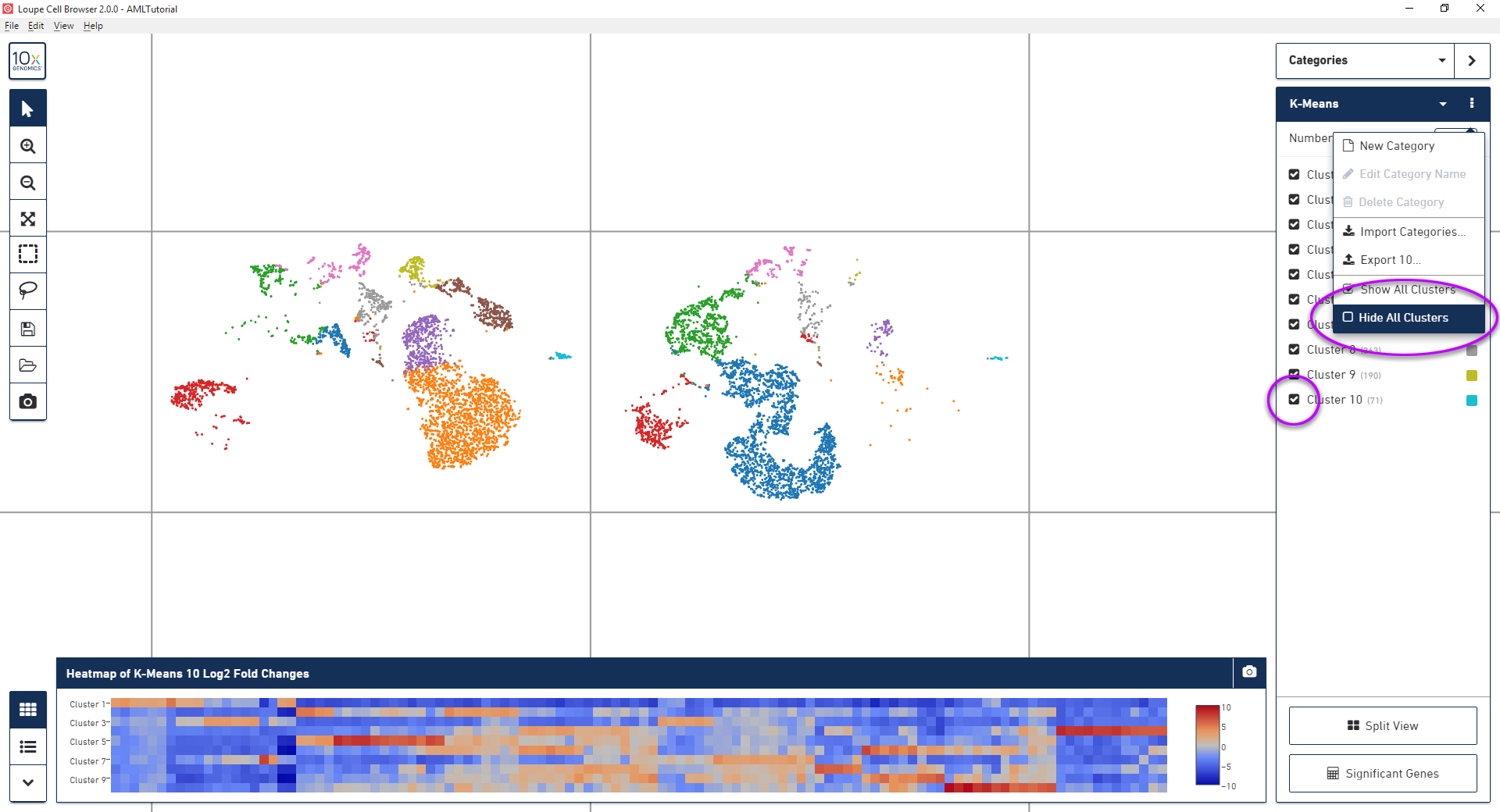
Use the split view button to view the samples in split panels



Please note that the normal sample appears in the left panel.

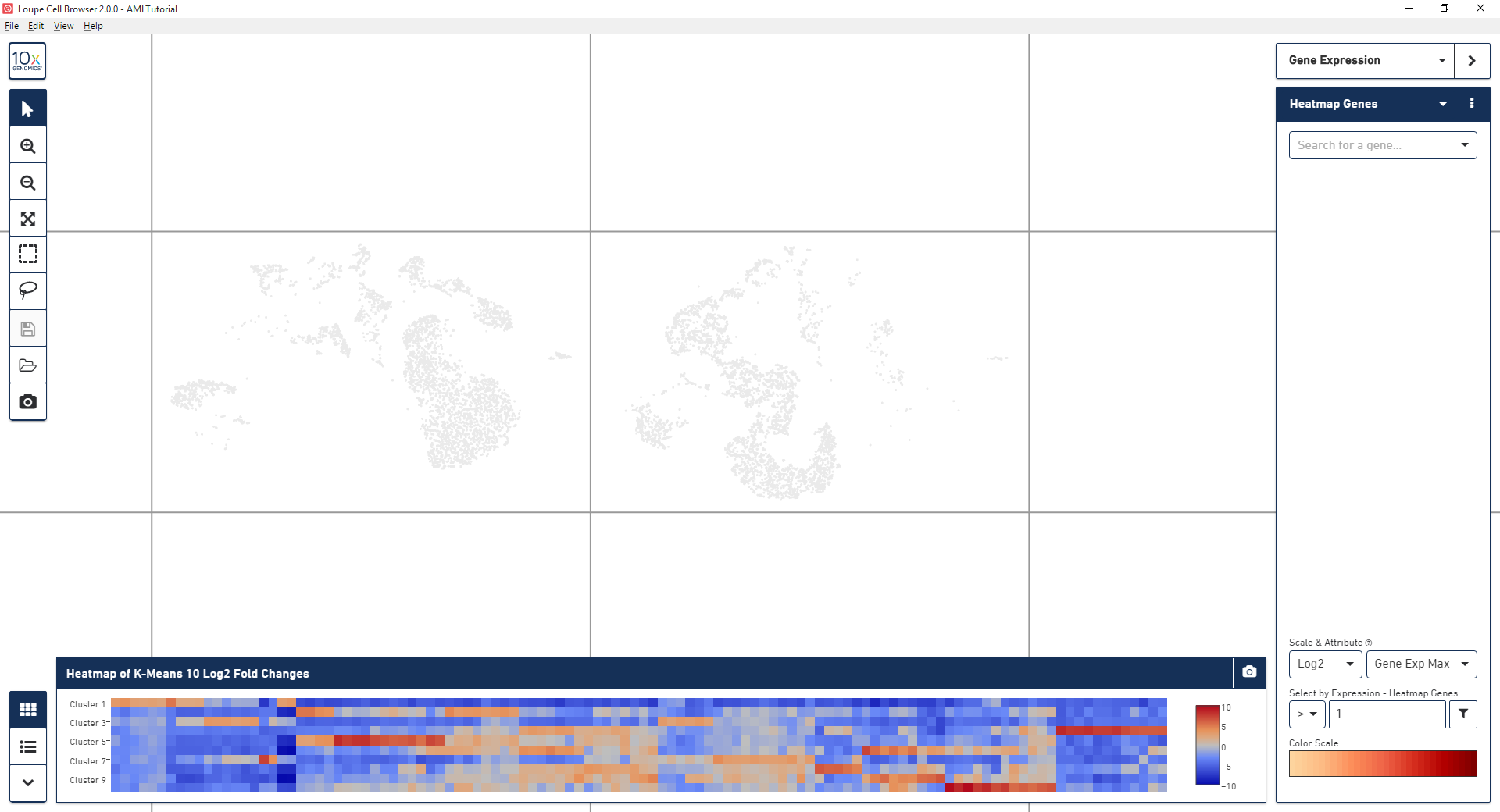
Select the K-Means clustering. Change the number of clusters to 10.

1. Do all the K-Means clusters appear in both normal and patient panels in the same proportion? In order to view this easily you can hide all clusters and then show each one by clicking their checkbox.



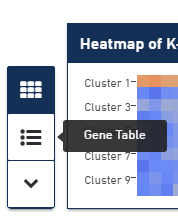
The bottom heatmap panel displays differential gene log2 fold change (average expression within the cluster cells compared to all the rest of the cells). Cells are clustered based on the selected clustering algorithm (K-Means in this case). Clusters appear in the rows and genes appear in the columns.

Looking at the heatmap, select the highest expressed gene in cluster 7 (using the scale bar on the heatmap panel). Once clicked (the colored square), the gene expression mode will be displayed –



1. Which gene did you select as the highest expressed gene for cluster 7?
2. Is there a difference in this gene’s expression in both treatments?

Return to the categories mode and select the gene table view



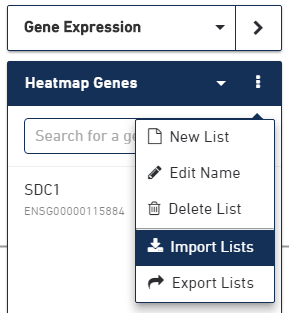
Select cluster 10 and sort the genes by descending P-Value



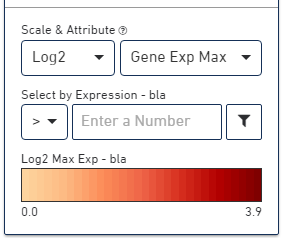
Click the top scoring gene (most significant with lowest P-Value, symbol on left) and view in which cells the gene is expressed

Download the “b\_cells.csv” file from “/course\_2018/single\_cell\_RNA\_seq\_exercise”

Import this gene list into Loupe by navigating to the downloaded location

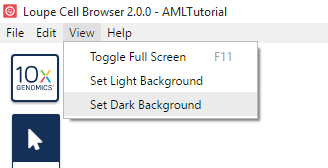


By default, “Gene Exp Max” is selected



When no gene is selected from the list, then cells, which express **any** of the genes in the list, are colored according the maximally expressed gene for a given cell

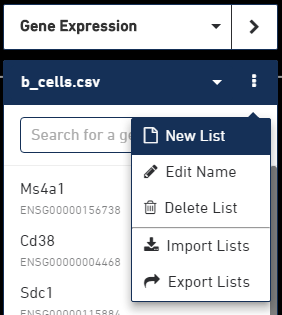
Select Ms4a1 to view its expression.   
In order to better view all cells expressing this gene, first unsplit the view (by selecting the Split view in the categories mode) and then select the “set dark background” option.



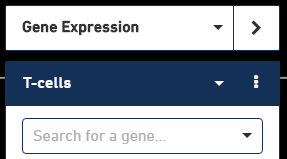
Split the view according to LibraryID and return to the gene expression mode.

1. Is there a difference in the expression of this gene between the samples?

Create a new gene list and name it T-cells



Search for the TCRE gene, which is a marker for T-cells

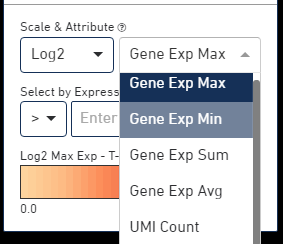


Use GeneCards or any other system to find the official symbol for this gene and search for it in Loupe

1. What is the official symbol for Tcre?
2. Is there a difference in this gene’s expression in both samples?

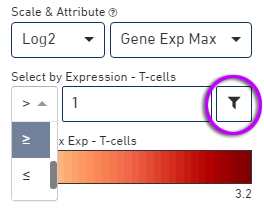
Search for Cd3g which is also a marker for T-cells

In order to see the co-expression of the two genes in the “T-cells” list in the same cells select “Gene Exp Min” (non-intuitive labeling)



1. Did the amount of highlighted cells increase or decrease?

Change the scalar selection back to “Gene Exp Max” and filter for cells with a log2 fold change greater or equal to 1



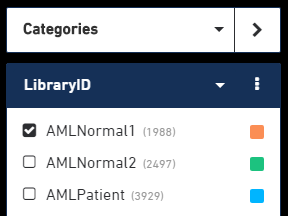
Create a new category called Mononuclear cells(or any other name) and a cluster called T-cells

Notice that the mode changed to categories and that only the T-cells that are marked in blue appear on the right panel

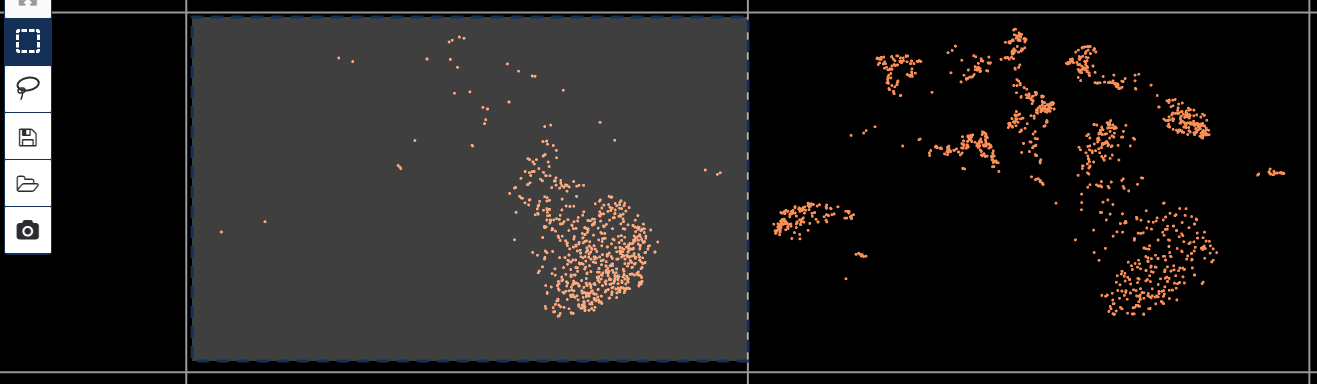
Click split view (twice) until the T-cells are isolated in their own panel

In order to find the proportion of cells in each sample (normal1, normal2 and patient) we will count how many T-cells appear in each using the LibraryID category. Click the “Split View” button to split according to T-cells vs all other cells (left panel now shows T-cells and the right one shows all the rest). When you select LibraryID you will see that the T-cells are a mixture of them.

Uncheck the check-boxes for AMLNormal2 and AMLPatient so now the left panel will only display T-cells belonging to AMLNormal1.



Use the Rectangular selection tool to select all cells in the left panel containing T-cells from the normal1 sample.



Add the new cluster to the Monouclear cells category and name it “AMLNormal1 T-cells” (or any other name)

Click split view twice until you have 3 panels

Select the LibraryID category again and this time check only AMLNormal2. Use the Rectangular selection tool again to mark “AMLNormal2 T-cells”.

Click split view twice until you have 4 panels

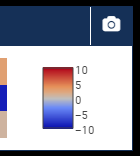
The top left panel now contain only AMLPatient so rename it to “AMLPatient T-cells”

1. How many T-cells are included in each of the original samples?
2. Export a plot of the panels using the “Export Plot to Image” tool (when naming the file include the file type – e.g. jpg)



Click the “Significant Genes” button and then “Globally distinguishing”

1. Export a plot of the heatmap using the “Export Heatmap to Image” tool (when naming the file include the file type – e.g. jpg)



Since the heatmap expression pattern for AMLNormal2 resembles AMLPatient more than when comparing AMLNormal1 to AMLPatient, we will compare only the latter two. Uncheck AMLNormal2 T-cells, click the “Significant Genes” button and then “Locally distinguishing” which will compare only the checked samples.

1. Select the “Gene Table” view and select the “Top 100 Genes” then click the “Export Table to CSV”