

# **Reviewing 10X RNA-Seq data with the Loupe Browser**

*Version 2024-02*

*(Loupe v7.0.1)*

## Licence

This manual is © 2019-24, Simon Andrews.

This manual is distributed under the creative commons Attribution-Non-Commercial-Share Alike 2.0 licence. This means that you are free:

- to copy, distribute, display, and perform the work
- to make derivative works

Under the following conditions:

- Attribution. You must give the original author credit.
- Non-Commercial. You may not use this work for commercial purposes.
- Share Alike. If you alter, transform, or build upon this work, you may distribute the resulting work only under a licence identical to this one.

Please note that:

- For any reuse or distribution, you must make clear to others the licence terms of this work.
- Any of these conditions can be waived if you get permission from the copyright holder.
- Nothing in this license impairs or restricts the author's moral rights.

Full details of this licence can be found at

<http://creativecommons.org/licenses/by-nc-sa/2.0/uk/legalcode>

In this exercise you are going to use the 10X Loupe Browser which is a desktop package to explore the output from CellRanger. One of the output files from the pipeline is a .cLoupe file which contains the expression data for all cells, a tSNE projection split into groups by two methods and some initial differential expression analysis.

Here you are going to load this data and see what you can learn about the same using the tools which loupe provides.

## Part 1: Loading data

- You can open the loupe browser 7 from your start menu. You can press the large “Open Loupe File” button and then select the COURSE .cLoupe file from the data folder to open the data in the browser.

The screenshot shows the Loupe Browser 7.0.1 interface. The main window displays a t-SNE plot of cells colored by cluster. The interface is divided into several sections:

- View Options:** Located at the top, it includes a dropdown for "Projection type" set to "t-SNE" and a "Split view" dropdown set to "No Split".
- Selection tools:** Located at the top right, it contains icons for pan, zoom, and other navigation tools.
- Left Sidebar:** Contains navigation options like "Home", "Clusters", "Features", "Reanalyze", "Advanced Selection", "V(D)J Clonotypes", and "Search for Features".
- Cluster Selection Panel:** A sidebar on the left lists "Pipeline-generated groups" under "Graph-Based", including "All" and "Cluster 1" through "Cluster 11", each with a checkbox and a cell count.
- Differential Expression Results:** A table at the bottom right shows "Up-regulated genes per selected cluster in group Graph-Based, compared to entire dataset." The table has columns for Feature, Cluster 1, Cluster 2, Cluster 3, Cluster 4, and CI.


Feature	Cluster 1	Cluster 2	Cluster 3	Cluster 4	CI
Name	L2FC	P-Value	L2FC	L2FC	L2
VCAN	3.03	8.30e-25	-5.49	-4.98	-6.47
CD14	3.01	9.89e-25	-5.34	-4.83	-4.96
S100A12	2.98	7.91e-24	-6.41	-6.01	-6.34

## Part 2: Basic Movement and Controls


- The main part of the view is a projection of your data. In the course data it defaults to tSNE but you can switch to UMAP using the “projection type” drop-down in the view options toolbar. In the view each dot is one cell and the colours represent what you have selected in your selection sidebar – usually clusters, or gene expression values.
- As long as you have the “pan” (hand shaped) tool selected in the selection tools toolbar you can move around in this view in the following ways
  - Click and drag within the main view to pan (drag all the cells) around
  - Use your mouse scrollwheel or trackpad to zoom in and out

- **Spend a couple of minutes getting used to the movement controls and looking at the structure of your data.**

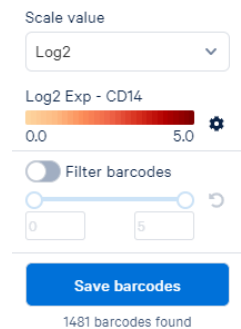
### Part 3: Working with Categories

- One of the main functions of Loupe is to be able to colour your projection by other useful information. There are two main types of colouring you can apply
  1. Categorical colouring where pre-defined groups of cells are coloured differently so you can tell them apart. This could be automatically generated clusters, or manually annotated groups of known cell types.
  2. Quantitative colouring where the projection is coloured by the quantitative expression level of one or more genes
- We can look at the pre-calculated clusters by using the  icon in the mode toolbar.
- There will initially be two Categories defined. Both are clusters of cells defined from their expression using different methods.
  1. Graph-Based uses a nearest neighbour approach to cluster cells
  2. K-means uses a quantitative statistical clustering method to define groups
- You can select between the two groups by clicking on their names in the Selections sidebar. You will see the cells being coloured according to the category you select. In k-means analysis you have to say how many clusters you want to define, so you will also see a drop down box to see the clustering when selecting between 2 and 10 kmeans clusters.
- Within each category you can turn groups on/off by ticking/unticking the boxes next to them. When you mouse over cells in the main view it will tell you which category they belong to.
- In the view options toolbar at the top you can also split the view by one of the categories to more clearly see the different sub-groups. You don't have to split by the same category you're colouring by – so you can colour by graph-based clustering and split by kmeans for example.
- **Spend a couple of minutes looking at the clusters defined by the different clustering methods.**
  1. **Do both methods agree on the number of clusters and the boundaries between clusters?**
  2. **Are clusters always clearly physically separated by tSNE?**
  3. **Are there ever cells which are classified in the same cluster, but which are separated in the tSNE?**

## Part 4: Working with Genes

- The other display mode is “Features” . In this mode you can overlay quantitative gene expression onto the main tSNE/UMAP projection. When you select this mode the sidebar will change to show a list of genes of interest (which will initially be empty).
- You can create named subgroups of genes to use later, but we’re just going to be looking at individual genes so it’s fine to use the default “Untitled List” group for our initial exploration.
- We will later use this mode to highlight genes identified from statistical analyses, but for the moment we don’t have those. You will often have known marker genes though and will want to see how they behave in your sample.
- In this data we are going to consider the following markers
  1. CD14 to find Monocytes
  2. CD3E to find NKT cells
  3. CD79A to find B cells
  4. FCGR3A to find NK cells

- To select a new gene start typing its name into the search box at the top of the sidebar. Once you see the gene you want then click on it and it will be added to the sidebar.
- Once a gene is in the sidebar you can click on it to have the projection coloured by the expression for that gene. You will also see a popup menu which lets you adjust the scaling and colouring of the main view.



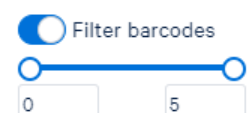
- **Find all of the marker genes listed above and see if you can use them to spot where the different groups of cells are within your projection.**

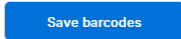
## Part 5: Creating new Categories with Filters

- The cloupe file comes with the Graph and Kmeans categories, but you can also define your own. Here we are going to make a new set of categories called “Cell Type” where we will use the markers from the last part to define the clusters in the category.
- We can make a new cluster category by applying a filter to the expression data. The process is as follows:

1. Select the gene you want to filter on from the sidebar (add it if it’s not there already)

2. In the popup which appears turn on the “Filter Barcodes” option



3. Drag the lower (leftmost) handle up to start selecting only cells with higher levels of expression of that gene. You should see that one group of cells in the main view becomes highlighted. Once you have a threshold which gives you a clean view of one group of cells you can press the  button.
4. When saving barcodes you need to provide two pieces of information
  1. The group of categories you want to use (the lower of the two boxes). For the first category there won't be a group, so create a group called "Cell Types". For all subsequent groups you can select the group you initially created.

Add to a new or existing group

 X ▼

2. The name for this category – this will be the name of the Cell Type for the marker you used to filter the cells.

Name the selected cluster

 X ▼

Add to a new or existing group

 X ▼

5. After you've created a cluster the mode will switch back to "Clusters" and show the category you just modified. To make a new selection you'll need to switch the mode back to Features.

- **Create the "Cell Type" group and use a suitable cutoff for all of the markers listed in part 4 to create categories for each of those cell types.**

## Part 6: Identifying Genes as Category Markers

- For the pre-defined categories (Graph Based and K-Means) the Cell Ranger pipeline will also have done some statistics to identify genes which are candidate markers for each cluster. You can access this information in the Differential Expression Results panel at the bottom.
- If the panel isn't particularly large you can drag the top edge up to make it bigger.
- If you switch to the Clusters mode and select either K-means or Graph Based you will see that the Differential Expression Results shows the set of clusters. If you click on a cluster name in the this panel you will see that you get a ranked lists of the most enriched genes in that cluster, showing their log fold change and p-value.
- You can click on a gene in the Differential Results Table and add it to your list of genes of interest so that you can then colour the projection by that gene.

- You can click on the buttons at the top of the Differential Expression Panel to switch between a table of hits and a heatmap to show the relative expression of genes in different clusters.

Feature TableHeat Map

- You can select a gene in the features mode and then select the “Expression Distribution” tab at the top of the Differential Expression Panel to see a violin plot of the levels of expression of that gene across the currently selected clusters.
- K-means(10) Cluster 4 is split into two by the Graph Based clustering (into graph clusters 4 and 10).**
  - 1. Look at the genes which are predicted markers for Graph clusters 4 and 10 and see if you think it is reasonable to split up K-means cluster 4**
- K-means Cluster 3 is split into two by the Graph Based clustering (into graph clusters 7 and 11).**
  - 1. Look at the genes which are predicted markers for Graph cluster 7 and see if you think it is reasonable to split up K-means cluster 3**

## Part 7: Calculating specific differences between categories

- The general pre-calculated analysis only finds markers for each of the pre-generated clusters. Loupe also has the ability to perform custom statistics to find either general or specifically enriched genes between any clusters.
- To calculate new statistically significant genes go to the Clusters mode and select the Group from which you want to select clusters. Use the tick boxes to select only the clusters you want to use for the analysis. At the bottom of the sidebar you can then press

Run Differential Expression

In this single dataset there are two types of comparison you can perform


- Globally distinguishing finds genes which distinguish these categories from the rest of the entire dataset
  - Locally distinguishing finds genes which distinguish the selected categories from each other
- In studies with multiple samples you can also run a comparison for the selected clusters between different sample groups.
  - As before, results from the analysis appear in the Differential Expression Results panel and you can switch between the tabular view and a heatmap.

- **Calculate the set of genes which distinguish Monocytes from NK cells and review the results.**
- **In Part 6 you should have seen that the distinction between graph clusters 4 and 10 was pretty sketchy. Try using the locally distinguishing selection to find genes which separate these two groups. Do you now think they should be separated or combined?**

## Part 8: Manual selections

- Another option within the program is that instead of using automatically defined clusters, or clusters based on gene expression that you can manually select groups of cells from the projection and then build a cluster from these and find the genes which define them.
- To do this you need to use the selection tools from the toolbar. There is a lasso tool which lets you draw shapes around cells to select arbitrary shapes, and a paint tool which lets you colour in the cells you want to select.
- **Go back to your Cell-Types groups. Just above, and to the left of the Monocytes is a small outgroup of unassigned cells. Use the lasso tool to select these cells and add them to your cell types group under a new name.**
- **Use the statistical tools to identify genes which are specifically expressed in the group of cells you just selected.**

## Part 9: Saving and Exporting

- Once you're finished with your analysis you can save your work in a few ways
  - You can use File > Save to save to a new cloupe file so you can open what you've done directly in Loupe in future.
  - You can use the Export button at the top right to save the current view to a graphics file (either PNG or SVG), or to save the data in the main view to a CSV data file.
  - From the Cluster mode you can save the details of any groups you have defined (either your own or the default ones) by clicking on the  icon.

## Part 10 (Optional): Re-clustering

Newer versions of loupe offer the chance to recalculate a tSNE or UMAP plot after selecting a subset of cells. You can also apply filters based on numbers of reads, genes, or mitochondrial content.



There is a wizard to generate a new projection launched from the **Reanalyze** button in the mode sidebar. Press this and follow through the wizard to create a new projection. You'll probably find it better to put the graphs into 'log' mode when setting the filter thresholds, and for the mitochondrial genes you can select that this is a human data set.