Babraham Bioinformatics

Reviewing 10X RNA-Seq data with the Loupe Browser

Version 2021-11

(Loupe v6.0.0)

Licence

This manual is © 2019-21, 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 from your start menu. You can press the large "Browser for a Loupe Cell Browser File" button and then select the COURSE.cloupe file from the data folder to open the data in the browser.

Part 2: Basic Movement and Controls

- The main part of the view is a projection of your data (usually defaulting to tSNE) where each dot is a cell. 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 to zoom in and out
- There are additional movement controls in the main toolbar. From the top down they are
  - Zoom In/Out
  - Autoscale
  - An option to save the current view

- The last tool in the toolbar is the marker (dot / cell) settings. These are usually automatically scaled, but you can use this tool to make them bigger or smaller. Personally, I find the default size to be a little small and often use this to enlarge them.

- Newer versions of CellRanger will also have performed a UMAP clustering. At the top, to the left of the Mode Selector is a projection selector where you can see the equivalent UMAP plot for the data. Have a look at both to see how different they look, but we will mostly be focussing on the tSNE plot for this exercise.

- **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
    - Categorical colouring where pre-defined groups of cells are coloured differently so you can tell them apart
    - Quantitative colouring where the projection is coloured by the quantitative expression level of one or more genes

- You can switch between these two colouring modes using the mode selector at the top right. For now we want to use "Categories"

- There will initially be two Categories defined. Both are clusters of cells defined from their expression using different methods.

    - Graph-Based uses a nearest neighbour approach to cluster cells
    - K-means uses a quantitative statistical clustering method to define groups

- You can select between the two groups using the menu at the top of the sidebar. You will see the cells being coloured according to the category you select. The k-means category has several sub-categories depending on the number of clusters which were defined and you can also change this from a drop down in the sidebar once you've selected K-means.

- 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.

- **Spend a couple of minutes looking at the clusters defined by the different methods.**

    - **Do both methods agree on the number and boundaries between clusters?**
    - **Are clusters always clearly physically separated by tSNE?**

Part 4: Working with Genes

- The other display mode is "Gene/Feature Expression". In this mode you can overlay quantitative expression onto the projection. When you select this mode the sidebar will change to show a list of genes of interest (which will initially be empty).

- 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

  - CD14 to find Monocytes
  - CD3E to find NKT cells
  - CD79A to find B cells
  - 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 can use the controls at the bottom of the sidebar to change the colouring and whether the expression range is linear or log scale.

- **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:

  - Select the gene you want to filter on from the sidebar (add it if it's not there already)

  - At the bottom of the sidebar where it says "Enter a Number" but in a minimum expression level on which you want to filter to select cells. You can look at the colour scale below to get an idea of what numbers make sense. You want to select as many physically connected cells as possible whilst picking up as few off target hits as possible.

  - After entering a number press return (or press the icon to the right of the number box). You will see the selected cells become highlighted in the main view. You'll also see an "Assign Selection" box appear. If you don't like the selection just press the X on the box. If you do want to keep it then select a category and cluster name for this set.

        - For the first cluster you'll need to make a "Cell Types" category, after that you can just select it.
        -
        - You'll need to make a new cluster name each time. Once you've entered the name press "Save" to save it.

    - After you've created a cluster the mode will switch back to "Categories" and show the category you just modified. To make a new selection you'll need to switch the mode back to Gene/Feature Expression.
- **Create the "Cell Type" category and use a suitable cutoff for all of the markers listed in part 4 to create clusters for each of those cell types.**

Part 6: Identifying Genes as Category Markers

- For the pre-defined categories (Graph Based and K-Means) the CellRanger pipeline will also have done some statistics to identify genes which are candidate markers for each cluster. You can access this information in the Data Panel.

- If the Data Panel isn't particularly large you can drag the top edge up to make it bigger.

- If you switch to the Category mode and select either K-means or Graph Based you will see that the Data View shows the set of clusters. If you click on a cluster name in the Data View you will see that you get a ranked lists of the most enriched genes in that cluster.

- If you click on a gene in the Data View you can either add it to your list of genes of interest in the Gene/Feature mode, or you immediately set it to be the active feature and colour the projection by that gene.

- You can click on the icons to the left of the Data Panel to switch between a table of hits and a heatmap to show the relative expression of genes in different clusters.

- **K-means Cluster 2 is split into two by the Graph Based clustering (into graph clusters 2 and 6).**

    - **Look at the genes which are predicted markers for Graph cluster 6 and see if you think it is reasonable to split up K-means cluster 2**

- **K-means Cluster 3 is split into two by the Graph Based clustering (into graph clusters 3 and 11).**

    - **Look at the genes which are predicted markers for Graph cluster 11 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 Categories mode and select the Category 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 select the type of analysis to perform
  - Globally distinguishing finds genes which distinguish these categories from all cells
  - Locally distinguishing finds genes which distinguish the selected categories from each other
- As before, results from the analysis appear in the Data 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 3 and 7 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 rectangular selection too which is mostly useless, and a lasso tool which lets you draw shapes around cells to select arbitrary shapes.
- **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 camera icon in the Data Panel to save an image of your heatmap

- You can use the export icon at the top right of the Data Panel to save your table of hits to a CSV file which you can open in a spreadsheet or an environment such as R.
- You can use the camera icon in the toolbar to save the main view to a file.
- From the dropdown at the top of the sidebar in Categories mode you can save the details of any clusters you have defined (either your own or the default ones).

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, either through the main interface, or by adding filters on numbers of reads, genes, or mitochondrial content.

There is a wizard to generate a new projection launched from the recluster button at the top of the sidebar when you're in Categories mode. 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.