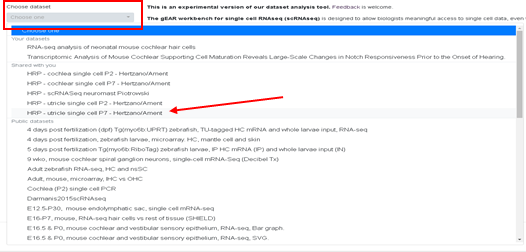
**gEAR Manual – scRNA-seq Workbench**

The gEAR scRNA-seq workbench enables users to analyze single cell RNA-sequencing (scRNA-seq) data without having to learn or use programming. Analyses can begin with a raw data file (e.g., a matrix of read counts as obtained from the cell ranger) or processed files. The steps range from quality control steps such as removing stressed or apoptotic cells to finding marker genes for cell clusters and comparing across clusters. Analyses can be saved and shared. The workbench is very user friendly. We do recommend reading the guide before using the workbench for the first time, to ascertain maximal benefit. To get to the scRNA-sequencing workbench, click on the picture below the scRNA-seq workbench description on the gEAR’s homepage.

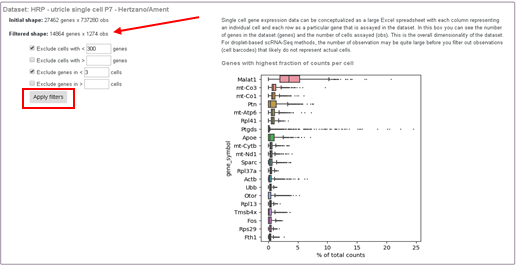


**Step 1 – Choosing a dataset:**

You can choose a dataset you wish to analyze by clicking on the drop down menu. Datasets are divided into three different categories; your datasets, datasets shared with you, and public datasets. To demonstrate the power of the workbench, we will be using the P7 utricle single cell dataset (HRP – utricle single cell P7 – Hertzano/Ament).



**Step 2 – Removing unwanted cells and genes:**

This first step is an important quality control step that can parse out low quality cells. These low quality cells are either empty captures or captures with multiple cells. This step can also exclude genes that are unlikely to be meaningful. To remove these cells and genes from the analysis, we can exclude cells with less than 300 genes and exclude genes that are in less than 3 cells (these numbers are arbitrarily set, meaning they can be changed depending on the dataset). You can also exclude cells with too many genes (removing captures with multiple cells) and genes that are in to many cells.

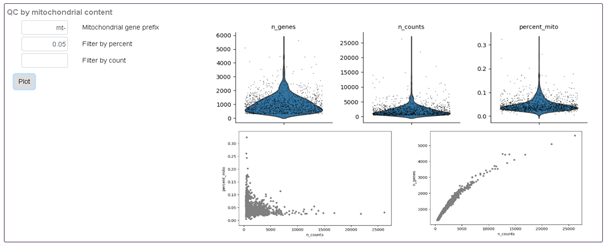
After clicking apply filters, we can see that genes and observations (cells) in the dataset have decreased. The gEAR also generates a plot showing the genes with the highest fraction of counts per cell.

**Step 3 – Filter by mitochondrial content:**

To get to this step, click on the slide bar next to that step on the side panel.



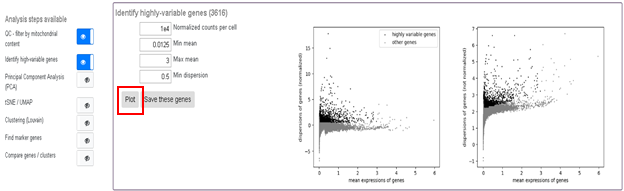
This step will allow you to remove cells that have high mitochondrial gene expression indicating stressed / apoptotic cells. Before choosing a percent to filter by, you first want to look at plots showing the number of genes in each cell, number of read counts per cell and percent mitochondrial content. To generate these plots, simply click on the plot button.



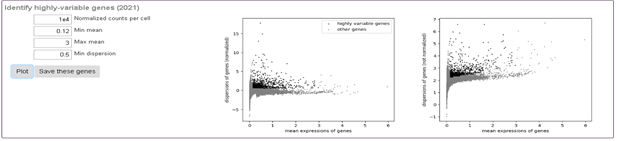
Depending on what is seen on the percent\_mito plot, we can adjust the percent mitochondrial we filter by. For this particular dataset, 0.05 (5%) works just fine.

**Step 4 – Identifying highly variable genes:**

This step can be accessed by clicking on the sliding bar icon next to the step on the side panel.



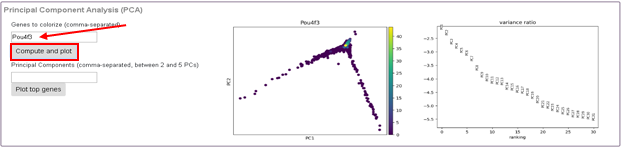
After clicking plot two graphs will display, the first one showing the dispersion of genes (normalized) vs mean expression of genes and the second one showing dispersion of genes (not normalized) vs mean expression of genes. Normalized counts per cell and Min dispersion can be changed via the corresponding boxes. The gEAR workbench will limit the maximum number of highly variable genes to 2,000. To reduce the number of genes, you can increase the minimal expression value of genes that may be considered as highly variable. After adjusting this value, we have reduced the genes to ~2000 and saved these genes by clicking on the box save these genes.



**Step 5 – Principle Component Analysis (PCA):**

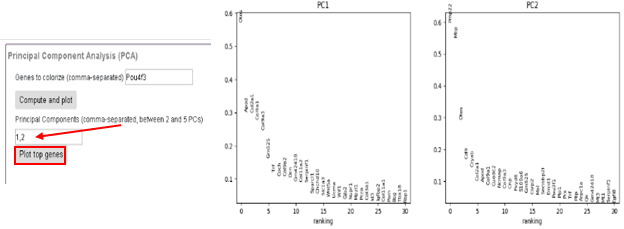
PCA simplifies the complexity of the dataset while retaining trends and patterns. For more information on PCA, please refer to this nature [paper](https://www.nature.com/articles/nmeth.4346). This step can be accessed by using the sliding bar in the side panel.

To visualize the PCA plots, you can simply click compute and plot. However, if you would like to visualize a gene(s), you can type it in the area provided (comma-separated). Here, we choose to visualize a hair cell marker gene, Pou4f3.



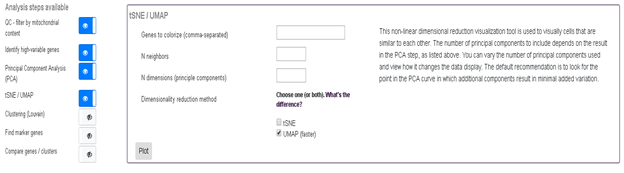
The first plot is showing PC1 vs PC2, with the gene of interest (Pou4f3) colored based on gene expression. The second plot is showing the amount of variance each principle component is contributing. For the next step (tSNE/uMAP), we will need to specify how many principle components we want to use. Use this graph to determine where the variance of the principle components are contributing minimal added variation. Here, we choose principle component 20 as the cut off.

To look at the top genes that are influencing some of the principal components, type 2 – 5 numbers (corresponding to the principal component number). Here we look at principal component 1 and 2.

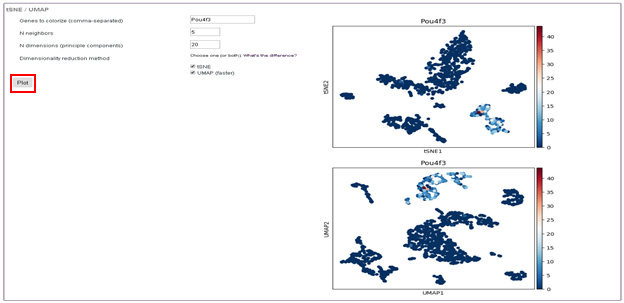


**Step 6 – tSNE/UMAP:**

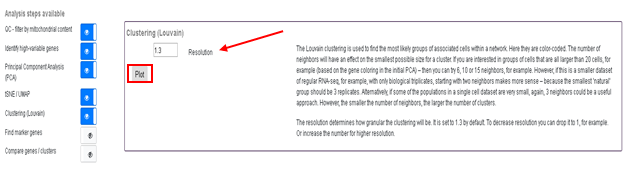
The next step is to generate a tSNE/UMAP of the dataset. These plots will allow us to visualize cells that are similar to each other (cluster together). For information on the differences between tSNE/UMAP, reference this [paper](https://www.biorxiv.org/content/early/2018/04/10/298430). This step can be accessed by using the sliding bar in the side panel.



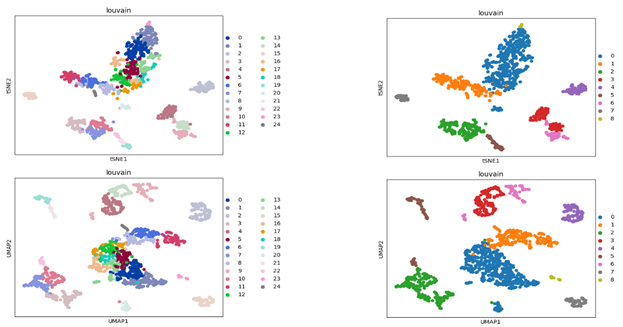
There are some variables that can change the tSNE/UMAP plot. This first one is colorizing the plot using gene expression. This section can be left blank, but it is very useful to specify genes that can distinguish between different cell types (marker genes – here we use Pou4f3 to visualize hair cells). The next variable is the number of neighbors, meaning how many neighbors (other cells) a cell has to balance attention between global and local aspects of the data. The third variable is the number of principle components used (dimensions); we discussed how to determine this in the previous step. You can also choose to generate a tSNE, UMAP or both by clicking on the boxes next to each type of plot. Each variable can be changed and the tSNE can be reran as many times as needed.



**Step 6 – Louvain Clustering:**

The next step of the process is using Louvain clustering to group cells together based on gene expression. The number of principle components and number of neighbors used in the previous step can affect the way the Louvain clustering looks. To access this step, slide the bar on the side panel.

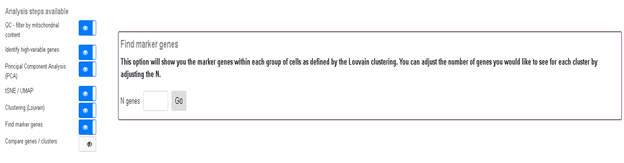
The resolution of the clustering determines how granular the clustering will be. For instance, using the 1.3 default will give more clusters than needed for our initial analysis, but bumping it down to 0.2 gives us a much more reasonable number of clusters.



Resolution = 0.2

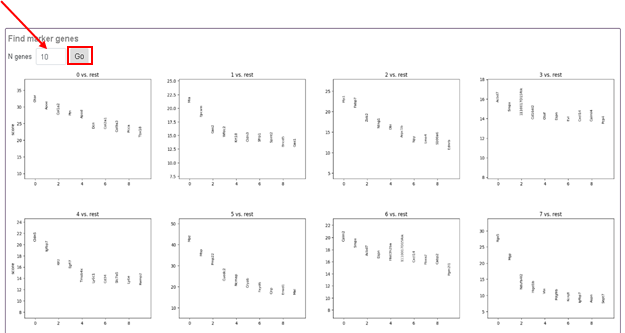
Resolution = 1.3

**Step 7 – Find Marker Genes:**

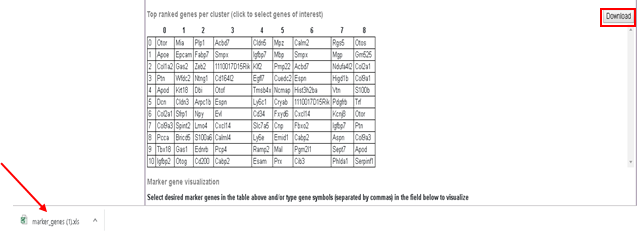
Once you find the resolution that works best for your dataset, the next step is to find marker genes for each cluster. To access this step, slide the bar on the side panel.

This tool shows genes that are highly expressed in each cluster. However, this does not mean the genes are only expressed in that particular cluster but simply that the genes are expressed more in that cluster compared to the other clusters. To obtain marker genes, enter the number of genes you want to see for each cluster, here we choose to display the top 10 genes.

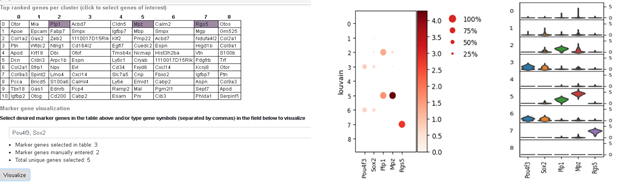
The gEAR will generate a plot for each cluster vs the rest of the clusters. These plots show the most influential genes for each cluster.



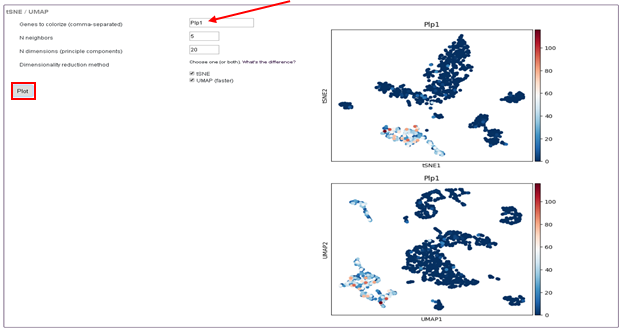
Also, the top ranked genes per cluster will be listed in table form below. You can download this table by clicking on the download button.



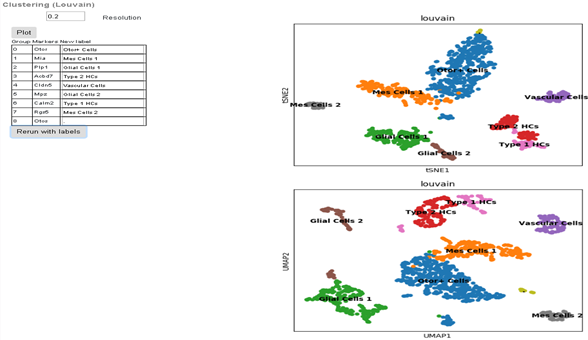
To visualize some of the marker genes, you can click on the genes you want to visualize in the table or type in a particular gene you are interested in. Here, we chose a couple genes from the table and typed in some already known marker genes for hair cells (Pou4f3) and supporting cells (Sox2).



After clicking visualize, you are presented with a dot plot (left) and violin plots (right). The dot plot displays gene expression based on color and how many cells express that gene in that cluster is shown by the size of the dot. Violin plots are generated for each gene and each cluster. You can also visualize clusters by going back to the tSNE step and typing in the particular gene you are interested in. For example, Plp1 is a great marker for one of the clusters.



After using these marker genes to identify each cluster, they can now be labeled with unique names with the table now present at the Louvain clustering step. After entering a unique name for each cluster, click ‘Rerun with Labels’. Your labels are now presented on the tSNE and UMAP.

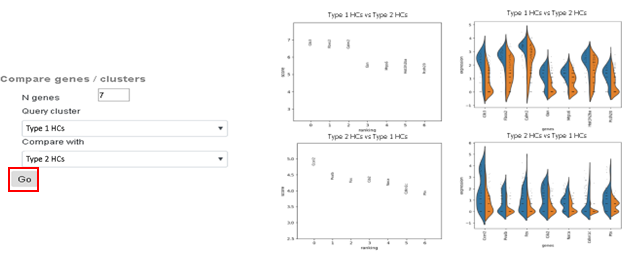


**Step 8 – Compare Genes / Clusters:**

The final feature of the scRNA-seq workbench allows you to compare genes between clusters. You can access this feature by using the slide bar on the side panel.

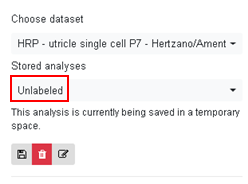


This tool allows you to compare the genes that are highly differentially expressed between two clusters. You can adjust the number of genes you want plotted via the number of genes box. Here we compare the differences between Type 1 HC and Type 2 HC genes.

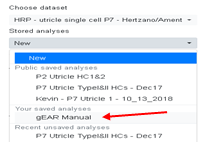


**Step 9 – Saving Analyses:**

You can save or delete your analysis at any time during the process. This is done by clicking on the save analysis or delete analysis buttons found on the top left hand side of the workbench. You can rename it after saving by clicking on the rename analysis button.



Once an analysis is saved, it can be accessed it anytime by looking in the drop down box under stored analyses. This analysis was saved as ‘gEAR Manual’.



After an analysis is saved, that analysis can be made public. This will allow anyone that has access to the dataset to see your analysis allowing for easy collaboration between researchers. Once an analysis is public, it will be displayed under the public datasets portion under stored analyses. Also if you change anything to the public dataset, it will spawn a local copy within your profile.

