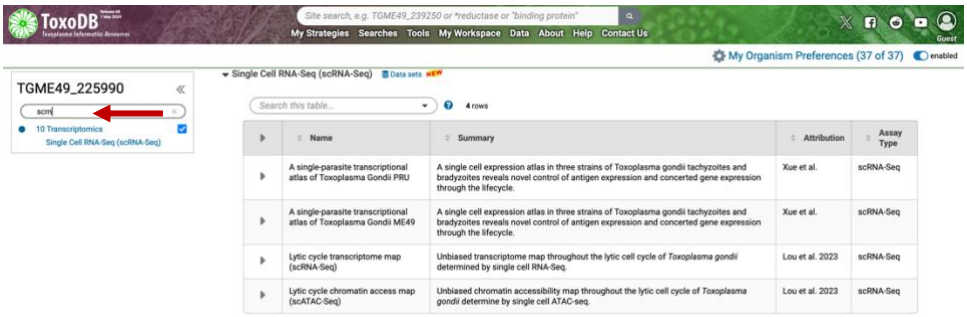
Single Cell RNA-Sequencing (scRNA-seq)

Note: *this exercise uses Toxo.org as an example database, but the same functionality is available on all VEuPathDB resources where this type of data is present.*

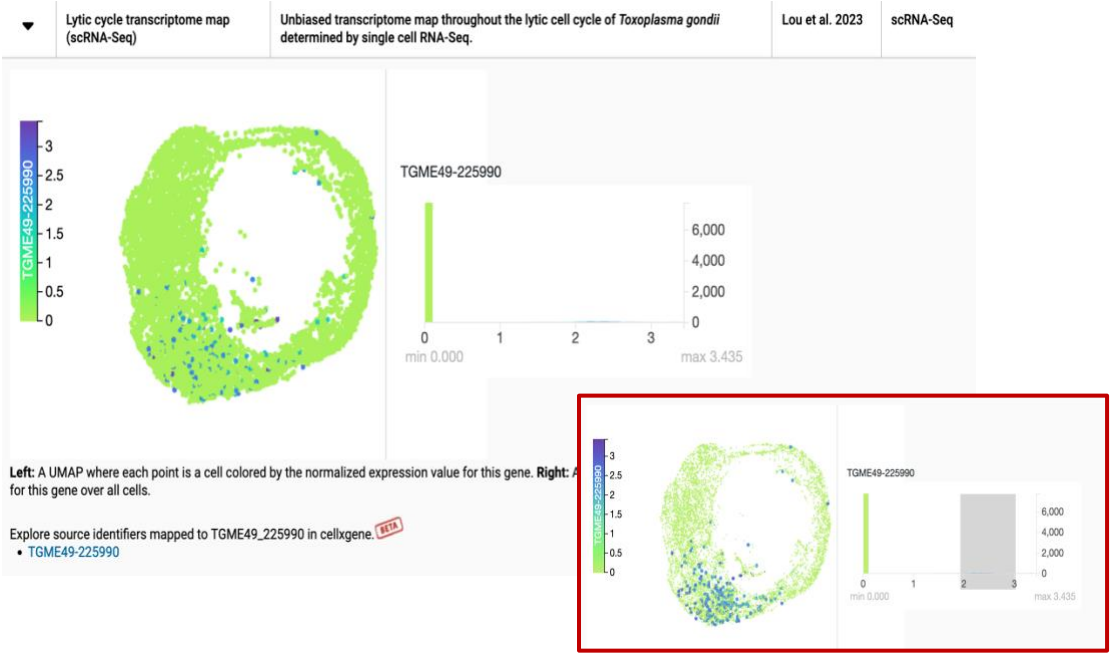
Learning objectives:

- Explore scRNA-seq data on specific gene pages.
- Explore scRNA-seq data using the cellxgene application.

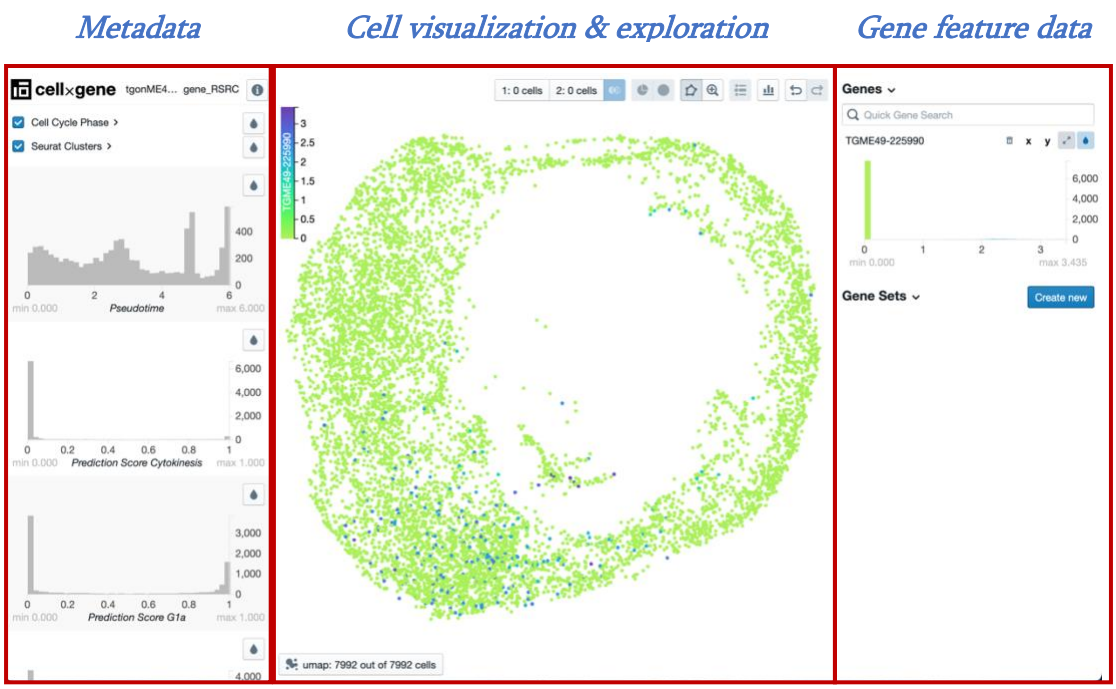
1. Go to the gene page for the gene **TGME49_225990** and scroll down to the single cell section. You can quickly do this by filtering the categories on the left side of the gene page.



2. Expand the experiment called “Lytic cycle transcriptome map (scRNA-Seq)”. What does the UMAP plot show? Where are the cells with the highest expression of this gene? You can click and drag in the histogram panel on the right to highlight cells in the left panel. Choose the area between 2 and 3 on the histogram to highlight high expressing cells on the graph.



3. Explore scRNA-Seq data in the cellxgene application. Cellxgene (cell-by-gene) is an open-source data visualization and exploration tool designed to help interrogate high dimensional data. We use cellxgene in VEuPathDB as a supplement to allow investigators to explore scRNA-Seq data.
    a. Click on the link “Explore source identifiers mapped to TGME49_225990 in cellxgene.”
    b. Your initial view will be a UMAP plot of all cells from this experiment. This may be black and white, or may be colored to show expression of a specific gene depending on how you got there. In this case you should see green cells and blue cells. What do the blue cells represent?
    c. The left-hand panel includes ***metadata*** while the right-hand panel includes ***gene feature data*** where data for any gene measured in the dataset can be explored. The central area is the ***cell visualization and exploration*** panel.

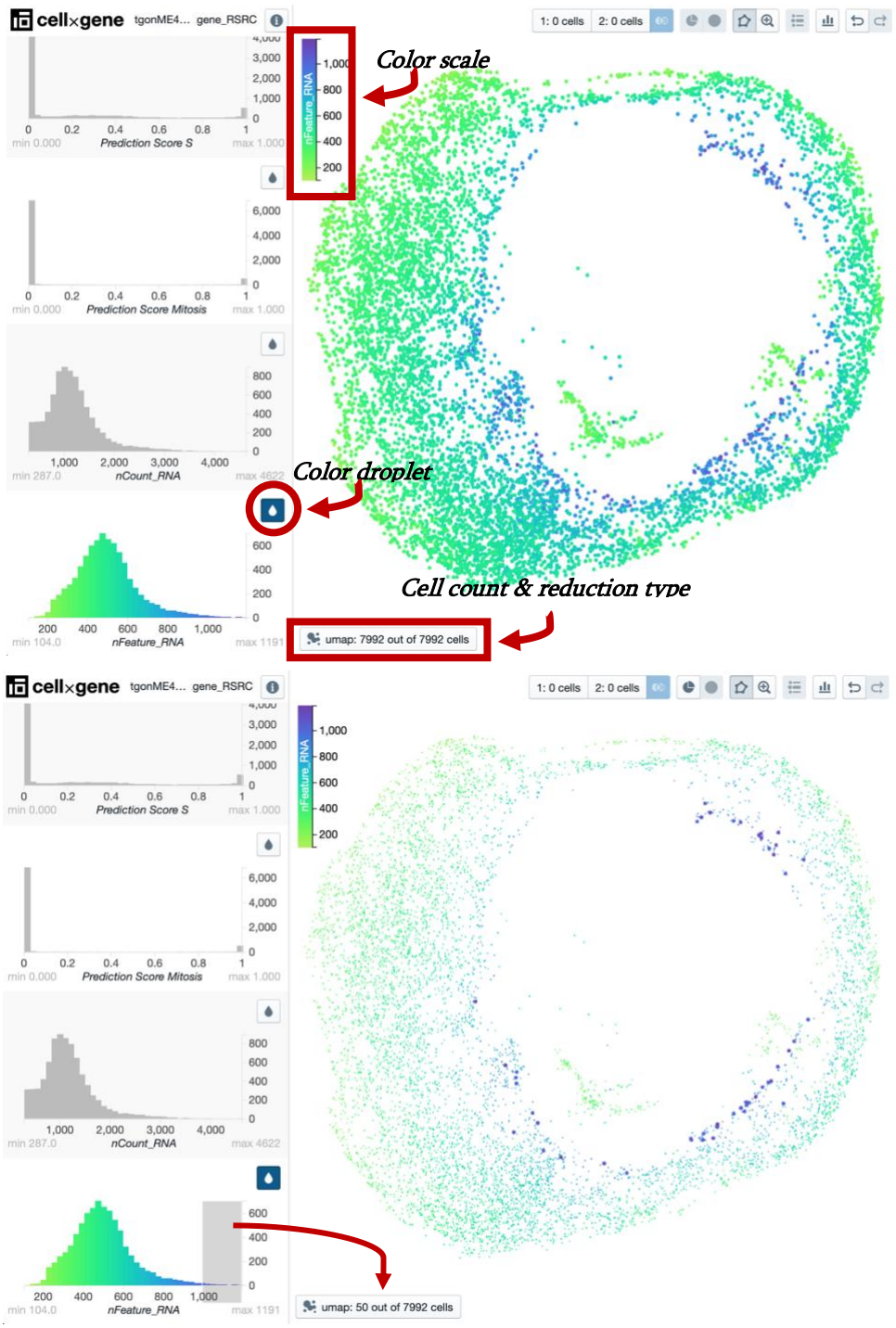


    d. Note that the metadata section includes numerical metadata represented as interactive histograms and categorical metadata such as the cluster assignments or replicates. The exact data shown here will vary by experiment.
    e. The droplet icon can be used to color the cells in the central panel with metadata from the left panel or gene expression data from the right panel. Try this:

- Expand the “Cluster” metadata category to see the cluster names. Note that these have been annotated by the author of the dataset
- Use the droplet icon to color the cells by cell cycle phase. Which cell cycle phase represent the cells where this gene is expressed?
- Hover over the cluster names to bring them into focus in the UMAP.
- Label the UMAP with the cluster names by clicking on the labels button in the central panel menu.

f. The droplet icon can also be used to color cells based on continuous metadata. Generally, the continuous metadata available is provided for QC purposes. Try this: click on the droplet icon for the nFeature_RNA (number of genes detected in each cell). How many cells are displayed?

g. In single cell data, it is common to capture a variable number of genes from each cell. How many cells were captured in which 1000 or more genes were observed? To find this, click and drag the histogram area in the left panel to highlight the area representing 1000 and above. Note: don't worry about being exact here, you are just trying to get an idea of what the data looks like.

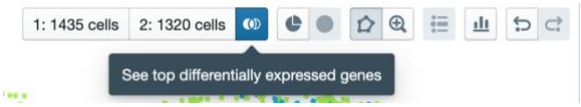


h. Do you think there is a bias in the number of genes assayed per cell in the different cell cycle stages? Can you get the number of cells in each of the stage that met the >/= 1000 genes representation (See step f)? To do this, use the checkboxes next to the cell cycle stage labels to deselect all but one stage.

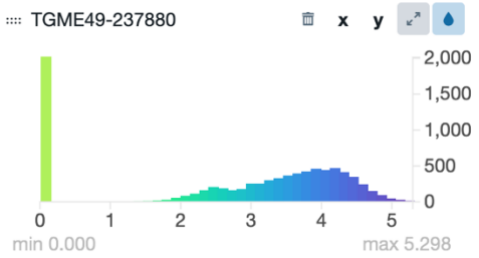
i. Do you see a difference in the number of cells? Don't forget to take account of the overall number of cells in each population. You can see this in the left panel.

j. Now let us identify genes that differentiate between the S and C stages? Follow these steps to do this:

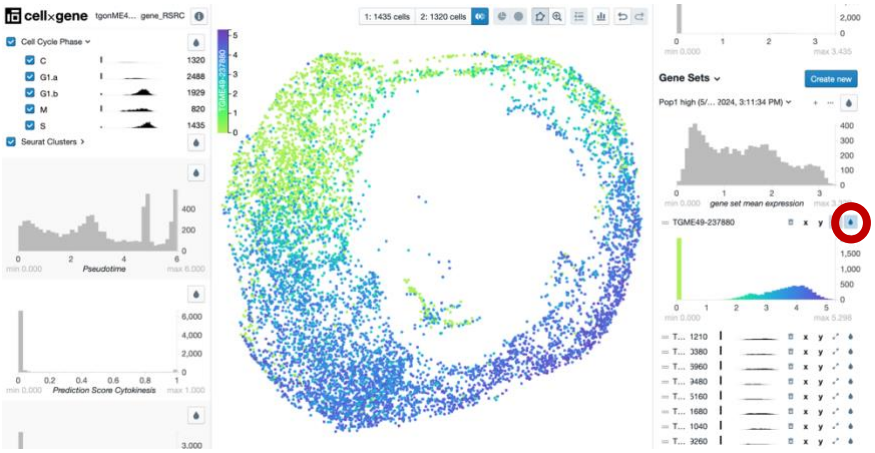
- Select the S cell cycle phase by using the check boxes in the left pane.
- Click on population 1 in the menu bar to save the selection for differential expression.
- Repeat the same process to select the C cell cycle population and save it as population 2.
- When done with your selections and saving populations, click on the differential expression icon.



- Click on population 1 in the right-hand gene feature panel to reveal the top S phase genes. Click on the expand icon to view a gene more clearly.
- The histogram in the right panel shows the expression of this gene over all the cells. You can color the UMAP by clicking on the droplet icon next to each gene. The expression of this gene in each cluster can be viewed as histograms in the left panel.



- Copy one of the gene IDs and explore it in ToxoDB. Can you come up with a rational reason why your selected gene might be important in this cell cycle phase? Note that copying gene IDs from cellxgene is frustrating. If you click on the expand icon for the individual gene, it becomes easier to copy the gene ID. Also, you might have to modify the ID because single cell analysis software often requires changing ID formats. In this case, the dash '-' needs to be converted to an underscore '_'.



- Repeat this for the C cell cycle phase.