GenAP Online Tutorials

Single-Cell Galaxy tools & workflows

Last updated: October 1st, 2020
Summary

- **Galaxy** is a powerful tool with a graphical user interface for running complex analysis workflows.

- Galaxy containers in GenAP come with a large suite of tools available, no installation of tools is required.

- We have installed many popular single-cell tools such as Salmon-Alevin, Seurat, Scanpy, UCSC-cell browser (credit to the [EBI-gene-expression group](http://www.ebi.ac.uk) for developing galaxy wrappers for many of these tools). We are constantly working to integrate more tools for Galaxy. Check back frequently to find the most recent tools available.
Example workflow: CellRanger output processing with Seurat

Let’s walk through an example workflow.

For this tutorial, we will use the same dataset used in the Seurat vignette. We will start from the quantified CellRanger output for the PBMC 3k v3 Dataset from 10x Genomics and use Seurat to process the data.

We have uploaded all the data to a public datahub for easy download. You can find all of the original files for this dataset () from 10x here.

You can find all workflows currently available in GenAP at the end of this presentation.
3k PBMCs from a Healthy Donor
Single Cell Gene Expression Dataset by Cell Ranger 1.1.0
Peripheral blood mononuclear cells (PBMCs) from a healthy donor (same donor as pbmc3k).

- 2,720 cells detected
- Sequenced on a high-throughput Sequencer 500 with ~40k reads per cell
- 70bp read1 (barcoded) and 100bp read2 (UMI)
- Analysis run with v1.2.0

This dataset is licensed under the Creative Commons Attribution License.

To batch download the files, copy and paste the commands below:

```
# Input Files
wget https://support.10xgenomics.com/single-cell-gene-expression/datasets/1.1.0/pbmc3k
```

```
# Output Files
wget https://support.10xgenomics.com/single-cell-gene-expression/datasets/1.1.0/pbmc3k
```

**Input Files**
- 7390605744115779301757644968e54G

**Output Files**
- 7390605744115779301757644968e54G

<table>
<thead>
<tr>
<th>File Type</th>
<th>Size</th>
<th>millions</th>
</tr>
</thead>
<tbody>
<tr>
<td>FASTQs</td>
<td>17.30 GB</td>
<td>7390605744115779301757644968e54G</td>
</tr>
<tr>
<td>Genome aligned GFF</td>
<td>15.96 GB</td>
<td>15.96 GB</td>
</tr>
<tr>
<td>Genome aligned GFF index</td>
<td>7.18 MB</td>
<td>7.18 MB</td>
</tr>
<tr>
<td>Per molecule read information</td>
<td>704.25 MB</td>
<td>704.25 MB</td>
</tr>
<tr>
<td>Gene / cell matrix (filtered)</td>
<td>7.27 MB</td>
<td>7.27 MB</td>
</tr>
<tr>
<td>Gene / cell matrix (gene)</td>
<td>11.72 MB</td>
<td>11.72 MB</td>
</tr>
<tr>
<td>Clustering analysis</td>
<td>12.52 MB</td>
<td>12.52 MB</td>
</tr>
</tbody>
</table>
Public datahub with all necessary files

We have also uploaded all the required data to a public datahub for easy usage in GenAP.


The next step, is to upload the data into your Galaxy instance.
A standard scRNA-seq analysis of 10x CellRanger data using Seurat inside GenAP

In this workflow, we will reproduce the analysis performed in the original tutorial for Seurat (v2.3.4).

Upload files to Galaxy → 10x CellRanger output → Seurat workflow (CRout_seurat_2.3.4) → Processed Seurat object → Visualize & annotate

Filter cells → Normalise data → ScaleData → RunPCA → FindClusters → RunTsne → FindMarkers
Uploading single-cell data to Galaxy: Option 1: URL link

Click the Download from URL button at the top left corner inside of your Galaxy instance.
Uploading single-cell data to Galaxy: Option 1: URL link

Click the Download from URL button at the top left corner inside of your Galaxy instance.

Click Paste/Fetch data three times. Enter the respective URL links for the 10x Cellranger input files from the public datahub into the fields and name the files with a unique and meaningful name.

(You can get the links for the files by right-clicking and then click “copy link address”.)

Click Start to upload the files into Galaxy.
Uploading single-cell data to Galaxy: Option 2: Upload from the FileBrowser

Another possibility to upload your data, is using the FileBrowser app in GenAP.

Open the FileBrowser app and upload your files into the FileBrowser.

Make sure the files are found in the FileBrowser under /ftp/share.
Uploading single-cell data to Galaxy: Option 2: Upload from the FileBrowser

Click on **Upload file** option under **Get Data** on the left hand side of Galaxy.
Uploading single-cell data to Galaxy: Option 2: Upload from the FileBrowser

Select the three files found in your /ftp/share folder and press Start.

(If your files are not visible here, they are likely not under /ftp/share!)
Files inside of Galaxy

If the file upload worked successfully, you should see 3 green files in the right sidebar inside Galaxy.

If the files are red, there was a problem with the upload. Check the error message carefully and see whether you can fix the upload issue yourself.

If you have any remaining problems with uploading files, send us a ticket at via:
https://genap.ca/p/home/contact-us
Uploading single-cell workflows inside Galaxy

Now that we have our single-cell data uploaded into galaxy, let’s load the workflow.

To load a workflow into Galaxy, click on workflow at the top of the navbar in Galaxy.
Uploading single-cell workflows inside Galaxy

Now that we have our single-cell data uploaded into galaxy, let’s load the workflow.

To load a workflow into Galaxy, click on workflow at the top of the navbar in Galaxy.

Click this button to upload a new workflow
Uploading single-cell workflows inside Galaxy

In the **Import Workflow** window, paste the URL for the workflow of your choice. In this tutorial, we will use the workflow for CRout_seurat_2.3.4:

https://datahub-330-pd6.p.genap.ca/galaxy_workflows/genap_galaxy.CRout_seurat_2.3.4.ga
Uploading single-cell workflows inside Galaxy

Click on Import workflow and you will be redirected to the previous window, where you will now see the workflow in your list.
Setting up single-cell workflows inside Galaxy

Now that we have our desired workflow inside Galaxy, let’s run the workflow.

To run the workflow we just uploaded, click on the arrow next to the workflow name and then click on Run. This will bring you to a new page, listing all of the steps of the workflow.
Setting up single-cell workflows inside Galaxy

This is the workflow menu. Here you configure the run parameters for this specific run of the workflow.

Please note that any changes to parameters you are making in this menu, will only apply to this run of the workflow.

If you want to make general changes to the workflow that apply to all future runs of your workflow, please perform these changes in the workflow editor.
Setting up single-cell workflows inside Galaxy

Setting this to “Yes” will send the results to a new History, rather than putting them in your current history.
Setting up single-cell workflows inside Galaxy

These are all of the different steps of the workflow. Click on any particular step to change the settings for this run.
Setting up single-cell workflows inside Galaxy

Drag and drop the files you have previously uploaded here. Make sure to enter the correct file for each option.
Now we will modify some of the parameters of the pipeline, to follow the original Seurat tutorial as close as possible. We don’t have a function to calculate % mitochondrial genes in Galaxy, so we will omit this filter for now.
First, we will modify the Genen high threshold for the number of genes identified per cell.

This is set to 2500 in the original Seurat tutorial.
Setting up single-cell workflows inside Galaxy

Next, we are going to modify the selection of highly variable genes.
Once we have adjusted all of the input parameters, we can start the workflow by clicking “Run workflow”.

Running single-cell workflows inside Galaxy
Running single-cell workflows inside Galaxy

Once you have submitted the workflow, you should see a list of jobs on the right hand side of galaxy.
Running single-cell workflows inside Galaxy

The jobs will start running one after the other. Intermediate objects that are not needed will be removed from the history. Jobs finishing successfully are marked green. If there was a problem with a job, it will be marked red, show an error message and all following jobs will be paused.
Running single-cell workflows inside Galaxy

When all of the jobs have finished green, the workflow run is complete.
Checking results of the workflow

After the workflow has finished, you should have the following items produced in your history. Please note that the numbering of the output files might be different. This is not important, the number at the beginning is just an internal object number for Galaxy!

- Log from exporting Seurat RDS object back to FileBrowser
- Data frame with marker genes for all clusters identified by Seurat
- Final Seurat object containing cells embedded in tSNE space
Checking results of the workflow

After the workflow has finished, you should have the following items produced in your history. Please note that the numbering of the output files might be different. This is not important, the number at the beginning is just an internal object number for Galaxy!

- 2D tSNE plot
- Data frame containing 2D tSNE coordinates for all cells
Checking results of the workflow

You can directly check the results of the clustering analysis by clicking on the view Data button of the Plot dimension job:

Click on the “View Data” button of the Plot dimension job to see this plot.
Checking results of the workflow

We can also process the Seurat results using **Seurat Export2CellBrowser** tool, followed by the **UCSC Cell Browser** tool. This will create an interactive UCSC Cellbrowser viewer inside of Galaxy:

First, input the Seurat RDS object and the marker list from FindMarkers into **Seurat Export2CellBrowser**
Checking results of the workflow

Then, use the output of the **Seurat Export2CellBrowser** tool as input for **UCSC Cell Browser**. Make sure to change the file format of the expression data to **Tar file with CellBrowser files to execute cBuild**.
Checking results of the workflow inside Galaxy

Once the job has finished, look at the clustering in UCSC CellBrowser:

Click on the “View Data” button of the UCSC Cell Browser job.
Checking results of the workflow inside Galaxy

This will create an interactive CellBrowser view inside of Galaxy. Click on View Data to see the CellBrowser:
Checking results of the workflow inside Galaxy

With the CellBrowser, we can quickly check whether the clustering corresponds with the original Seurat tutorial. Let’s check the B-cell marker MS4A1:

UCSC CellBrowser results in Galaxy

https://satijalab.org/seurat/v2.4/pbmc3k_tutorial.html
Analyzing and annotating results in SCAP

For a more in-depth analysis and annotation of the dataset, you can use the Rshiny app SCAP with your processed data.

To learn more about SCAP and the loom file format, please consult the SCAP tutorial.
Modifying a workflow

If you want to permanently change settings in your workflow or add or remove a tool, go to the Workflow section and click on the workflow you want to edit. This will open the workflow manager:
Modifying a workflow

If you want to permanently change settings in your workflow or add or remove a tool, go to the Workflow section and click on the workflow you want to edit. This will open the workflow manager:

For more details on workflow management, check the official Galaxy workflow tutorial:

https://galaxyproject.org/learn/advanced-workflow/
Single-cell workflows currently available in GenAP

All GenAP workflows can be found on the Single-cell Academy github page. Additionally, all workflows can be loaded into Galaxy from this public datahub URL:


Starting your analysis from FASTQ files

- genap_galaxy.alevin_scanpy_1.4.3.ga
- genap_galaxy.alevin_seurat_2.3.4.ga

Starting your analysis from CellRanger output or DGE matrix

- genap_galaxy.10xCR_scanpy_1.4.3.ga
- genap_galaxy.10xCR_seurat_2.3.4.ga
- genap_galaxy.DGE_scanpy_1.4.3.ga
- genap_galaxy.DGE_seurat_2.3.4.ga