**FireCloud :: Cell Ranger mkfastq/count :: Worksheet**

**September 2018**

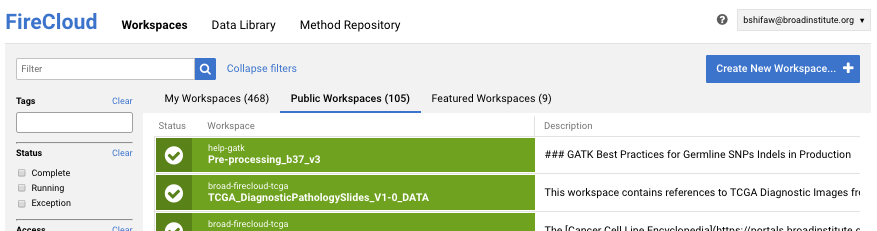
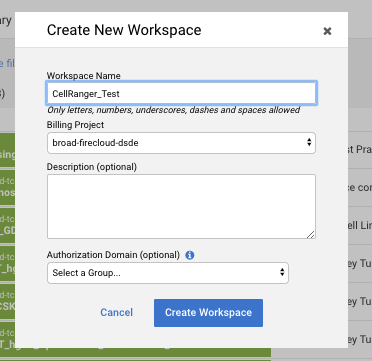
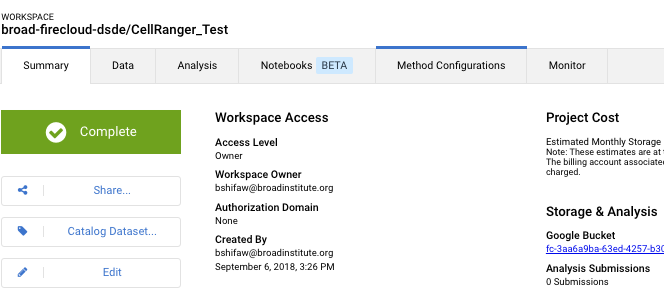
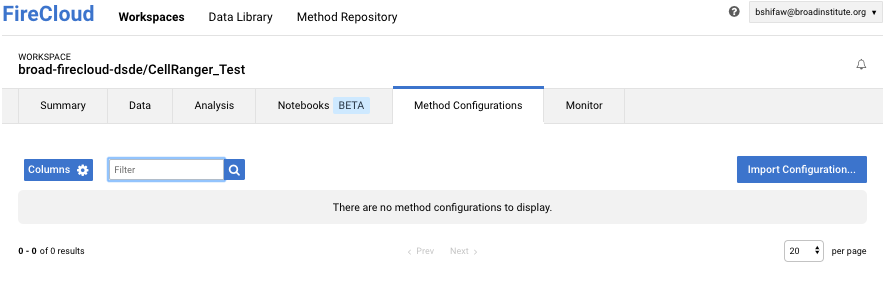
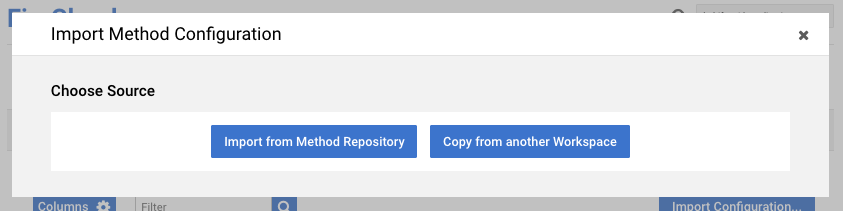
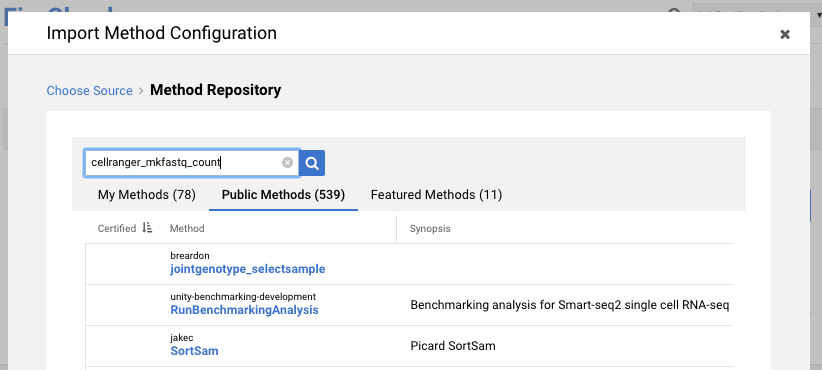
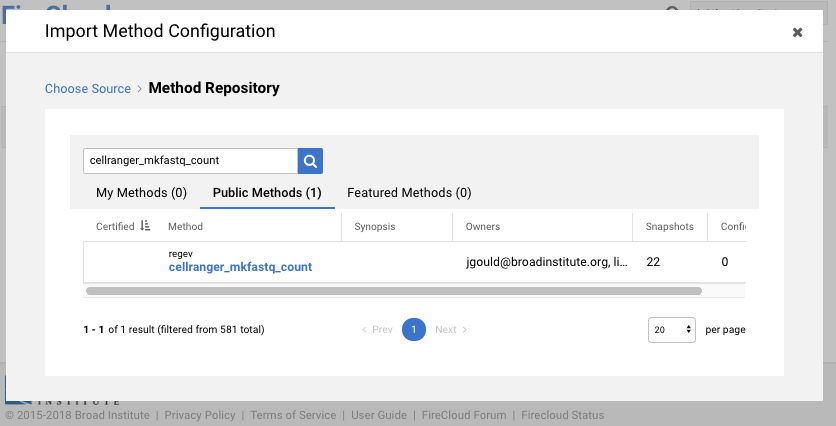
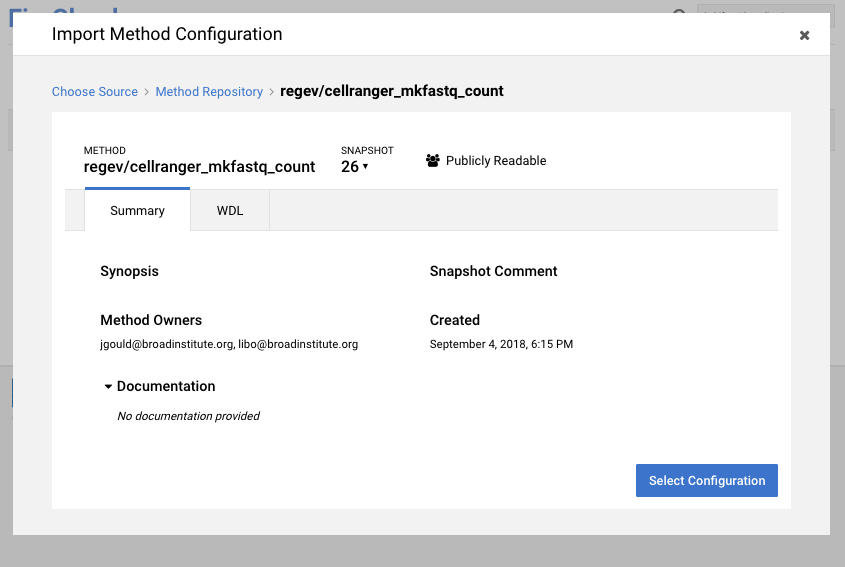
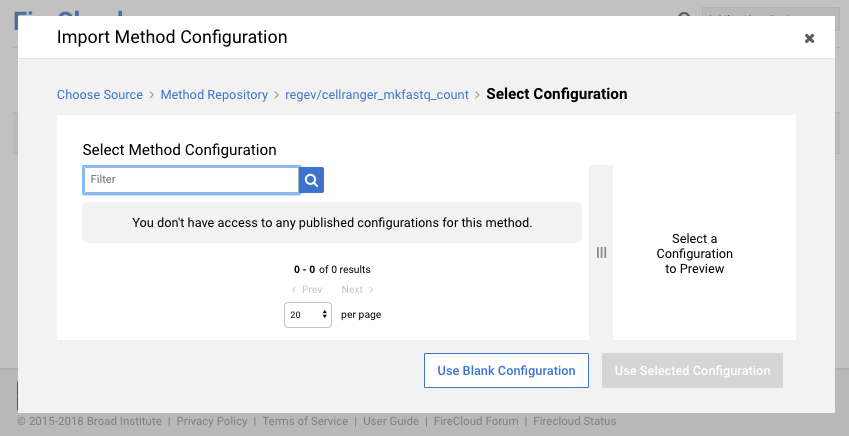
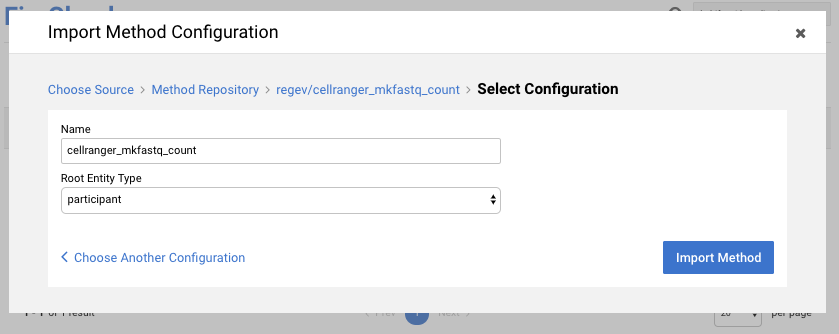
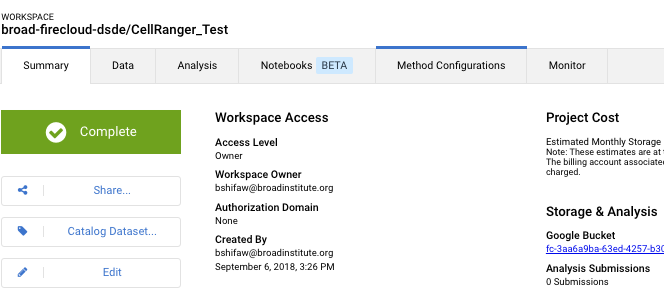
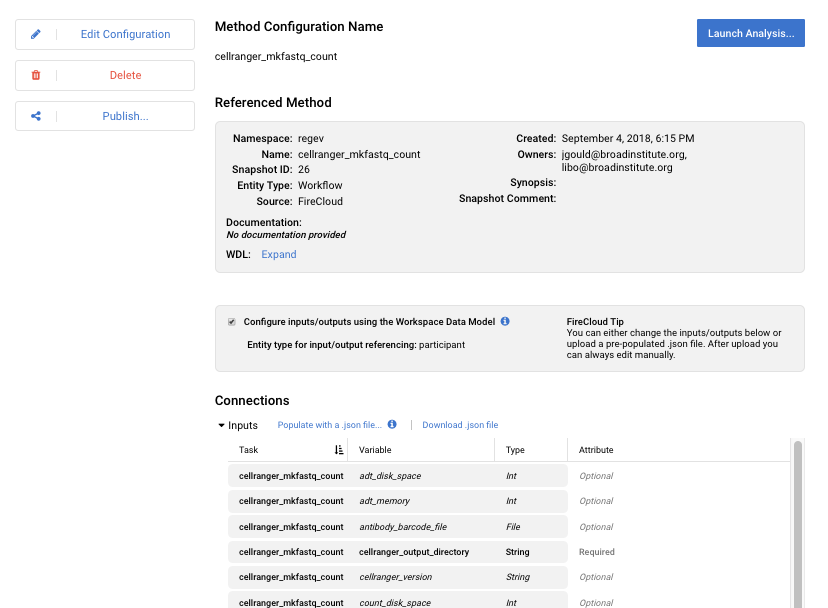
Cell Ranger is a suite of analysis pipelines that process single-cell RNA-seq output from the 10X Genomics Chromium platform to align reads, generate gene-cell matrices and perform clustering and gene expression analysis.

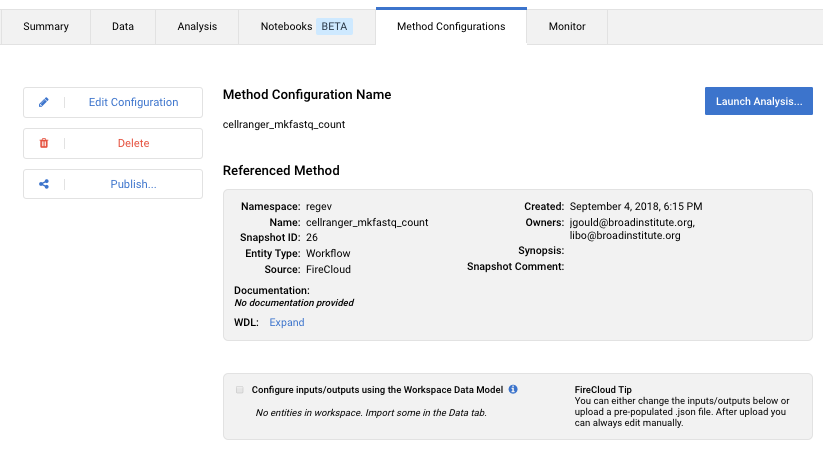
Broad scientists (Bo Li, Joshua Gould, et al.) have created a method in FireCloud that runs two of the Cell Ranger pipelines in the cloud:

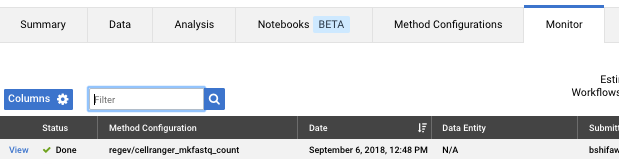
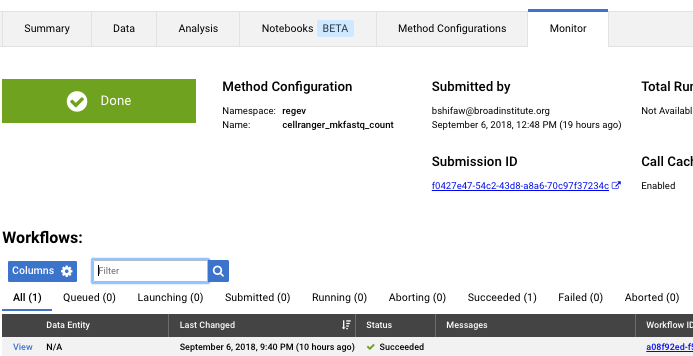
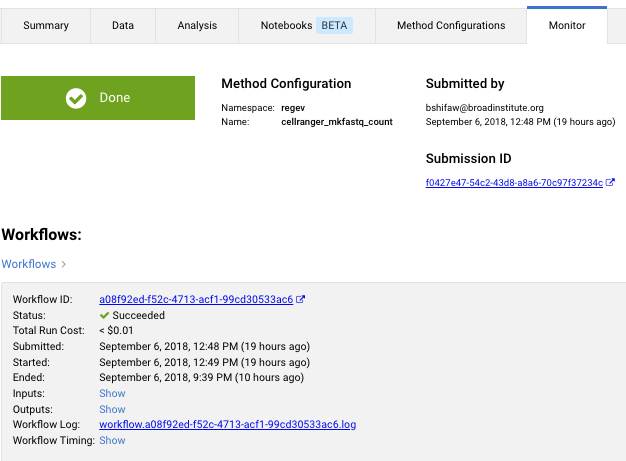
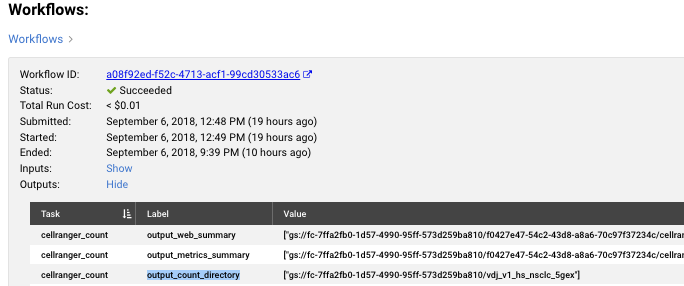
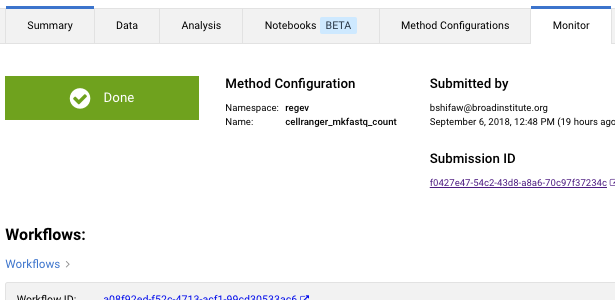
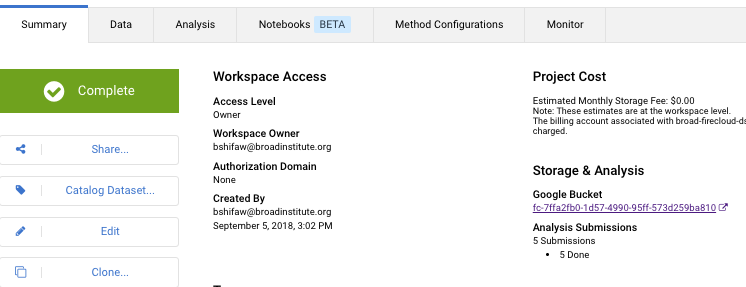
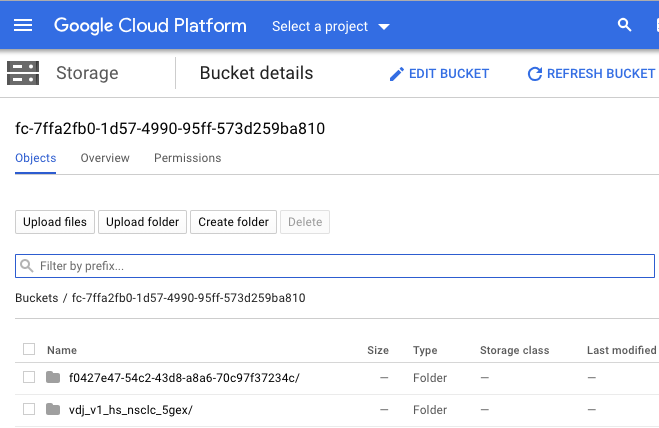
* The **cellranger mkfastq** pipeline, which demultiplexes raw base call files generated by Illumina sequencers into FASTQ files,
* And the **cellranger count** pipeline, that takes FASTQ files from cellranger mkfastq and performs alignment, filtering, barcode counting, and UMI counting.

[For further reading, 10X Genomics provides documentation on Cell Ranger (<https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/what-is-cell-ranger>), and Li, Gould et al. have documented the FireCloud method (<https://kco-cloud.readthedocs.io/en/latest/cellranger.html>)]

In this worksheet, we will teach you how to create a workspace for the cellranger\_mkfastq\_count method in FireCloud and launch an analysis.

1. Navigate to FireCloud (<https://software.broadinstitute.org/firecloud/>) in your preferred web browser. Click the `Use FireCloud` link.
2. After logging in to FireCloud, make sure you are viewing the Workspaces page and click on `Create New Workspace` at the upper right.   
   
3. A window will appear asking you to provide a name for your new workspace and to define the Billing Project. The name of the workspace is up to you. For the Billing Project, you may use the free credits provided to you when originally signed up for FireCloud or one provided by the workshop instructors. The other fields are optional. Once you are ready to continue, click `Create Workspace`.  
   
4. You will now be viewing your newly created workspace. Click on the `Method Configurations` tab within the workspace.  
   
5. The Method Configuration tab lists the methods in your workspace, but there won’t be any listed yet. We’ll need to load the Cell Ranger method from the Method Repository. To do this, click on the `Import Configuration...` button on the top right.  
   
6. A window will appear, and select `Import from Method Repository` to continue.  
   
7. The next window will list the available methods. Head to the `Public Methods` tab and use the filter box on the top left to filter for `cellranger\_mkfastq\_count` method.  
   
8. Click on `cellranger\_mkfast\_count`, owned by jgould@broadinstitute.org.  
   
9. In the next window, click the `Select Configuration` button.  
   
10. Looks like there isn’t any configurations we can borrow. No worries, we’ll set it ourselves later. Click `Use Blank Configuration` to continue.   
    
11. The next window will allow you select the root entity type for the workflow, however in this particular exercise and in contrast to other FireCloud exercises, our workflow will not be using the Data Model in the workspace, so no need to modify anything. Click on `Import Method` to continue.   
    
12. You will now be viewing the `Method Configuration` for the cellranger\_mkfastq\_count workflow. Before we continue with editing inputs and variables, let’s head back to the `Summary` tab. We need to copy the URL of the Google Cloud bucket for our Workspace. You’ll see it under the `Storage & Analysis` heading on the right side. Highlight and copy that string of letters and numbers to your clipboard. Be sure not to copy any trailing characters or whitespace.  
    
13. Let’s navigate back to the Method Configurations tab.To edit the inputs and variables you see further down the page, click on `Edit Configuration` on the left side. Next, it’s critical to uncheck the `Configure inputs/outputs using the Workspace Data Model` checkbox so that the method does not expect a root entity from the Data Model.   
    
14. Fill in the the following three variable configurations as shown here, including the quotation marks around the gs:// Google Cloud bucket URLs (if you forget to include quotation marks, the analysis will fail to launch):
    1. Cellranger\_output\_directory = “gs://<paste the workspace bucket URL here>”
    2. Input\_csv\_file = "gs://gatk-test-data/rna-seq-cellranger/samplesheet\_cellranger.csv"
    3. run\_mkfastq = false
15. Click the green `Save` button on the left. You are now ready to run the Cell Ranger method. Click on the `Launch Analysis...` button on the top right. A window will appear, and go ahead and click the `Launch` button to start the workflow.



1. The workflow will take some time to complete, likely on the order of several hours. To check the status of your job click on the `Monitor` tab. Your launched workflow will display “Done” under the `Status` column once it completes.   
   
2. Click on `View` in the status column, and then click on `View` again in the Data Entity column.   
   
3. Click on `Show` in the shaded box next to the row labeled `Outputs:` to show you a list and location of outputs generated by the workflow.   
   
4. One of the outputs is `output\_count\_directory`, which holds the aligned bam along with other metrics. In this case it would be our workspace bucket because we set it to be so in our inputs variable `cellranger\_output\_directory`.  
   
5. To view the what's inside the output directory `gs://<your workspace bucket>/vdj\_v1\_hs\_nsclc\_5gex` click on the `Summary` tab of your workspace at the top left.  
    
6. Click on the workspace bucket link in the `Summary` tab on the right side of the screen. In the image below it would be fc-7ffa2fb0-1d57-4990-95ff-573d259ba810.  
   
7. A new tab or window will open in your browser asking you to confirm your Google account. Confirm your identity to proceed to the the workspace bucket.
8. The workspace bucket will list the working directory of the submitted workflow (name is unique to each job) and the output directory of Cell Ranger which is the name of our sample. Click on Cell Ranger’s output directory `vdj\_v1\_hs\_nsclc\_5gex`  
   
9. You should now see the contents of Cell Ranger output directory.  
   