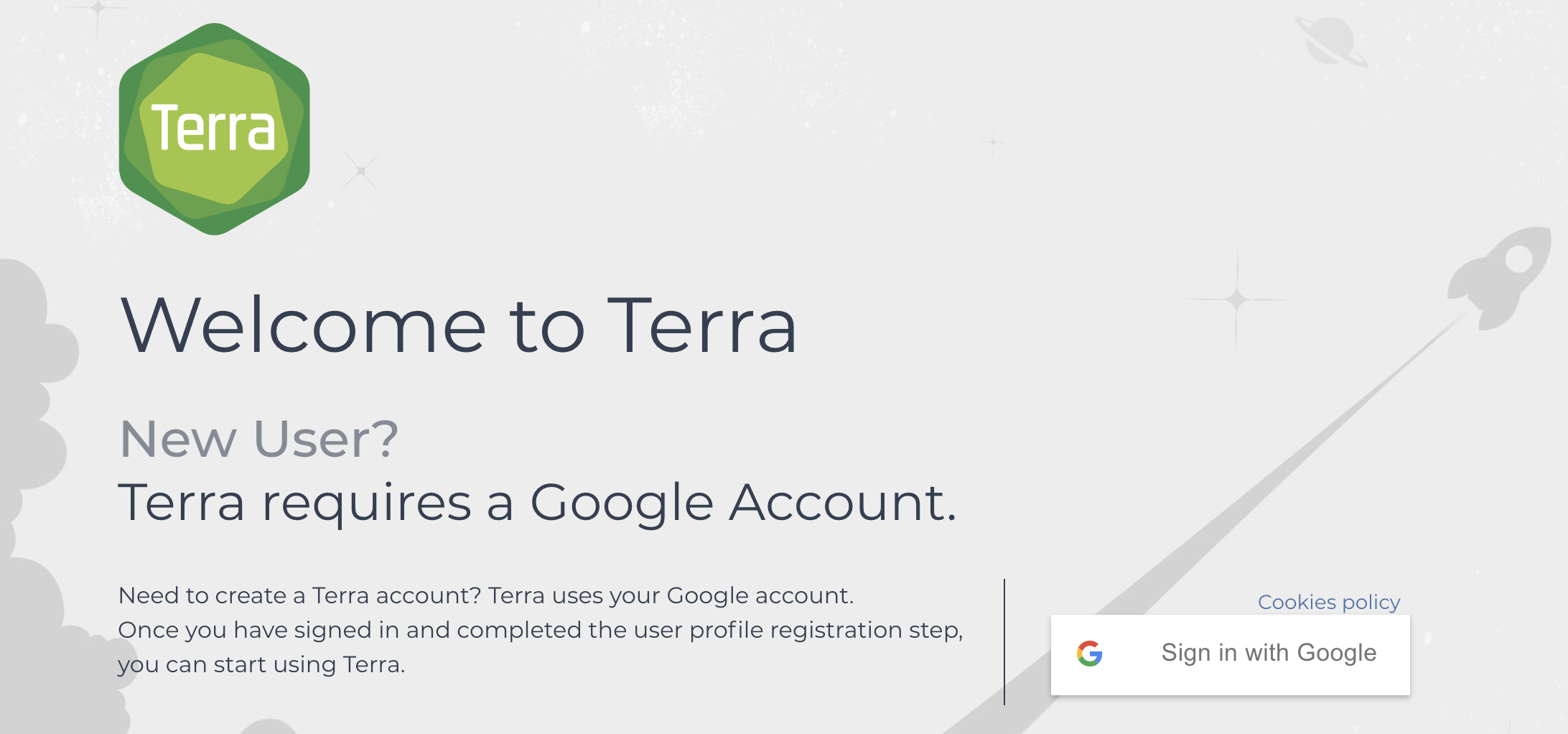
# Running dropseq\_scCloud\_workflow on Terra

The Terra version of the workflow has gives the user access to all workflow parameters as it is intended for users with advanced knowledge of the single-cell field.

### KNOWN BUGS IN SNAPSHOT 74:

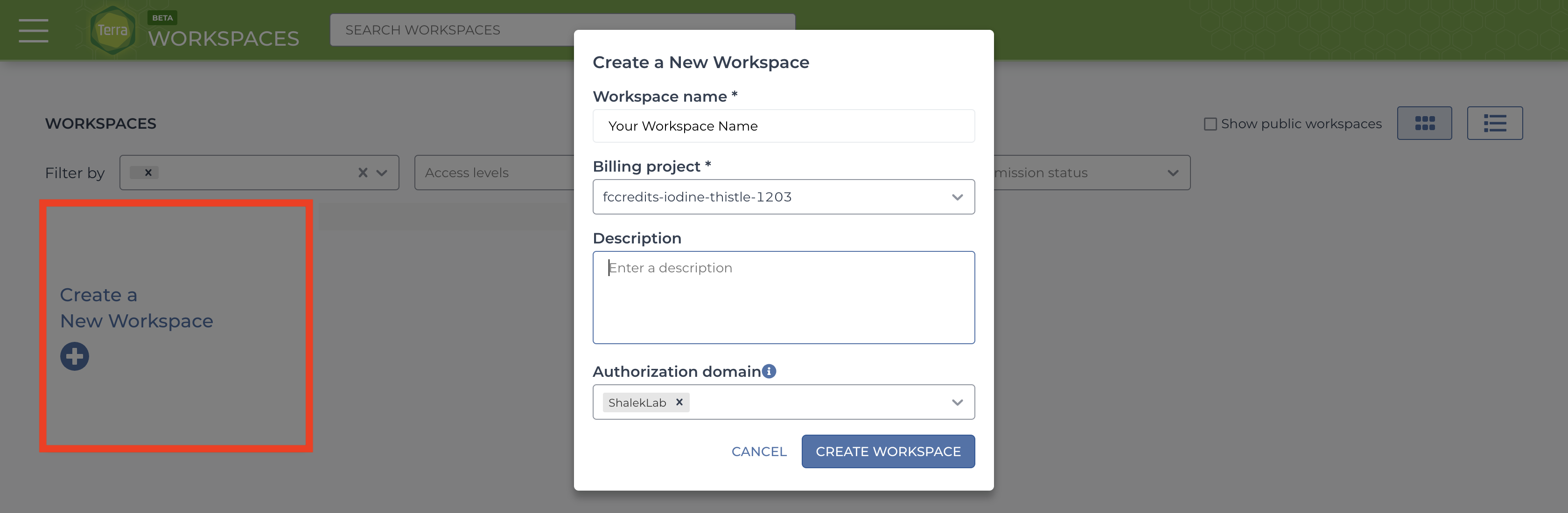
* Do not upload sequence files into the root of the bucket, instead use gsutil to move them into a folder on your bucket. Example: gsutil -m cp -r local/mouse\_fastqs gs://[bucket ID]/mouse\_fastqs. gsutil -m mv gs://[bucket ID]/\*.fastq.gz gs://[bucket ID]/mouse\_fastqs can be used to move all FASTQs already on the bucket to a directory mouse\_fastqs that will be automatically created.
* is\_bcl=true/run\_bcl2fastq=true is broken due to regevlab changing their docker repo for the tool.

### 1. Sign In

Visit the Terra by the Broad Institute and sign in using your preferred Google email account and complete the user profile registration steps if it is your first time using Terra. 

### 2. Create a New Workspace.

Skip this step if you have an existing workspace where you would like run dropseq\_scCloud.

After logging in/registration you will be brought to the “Your Workspaces” page, where you will see the "Create a New Workspace button. Click it and fill out the fields to your liking. 

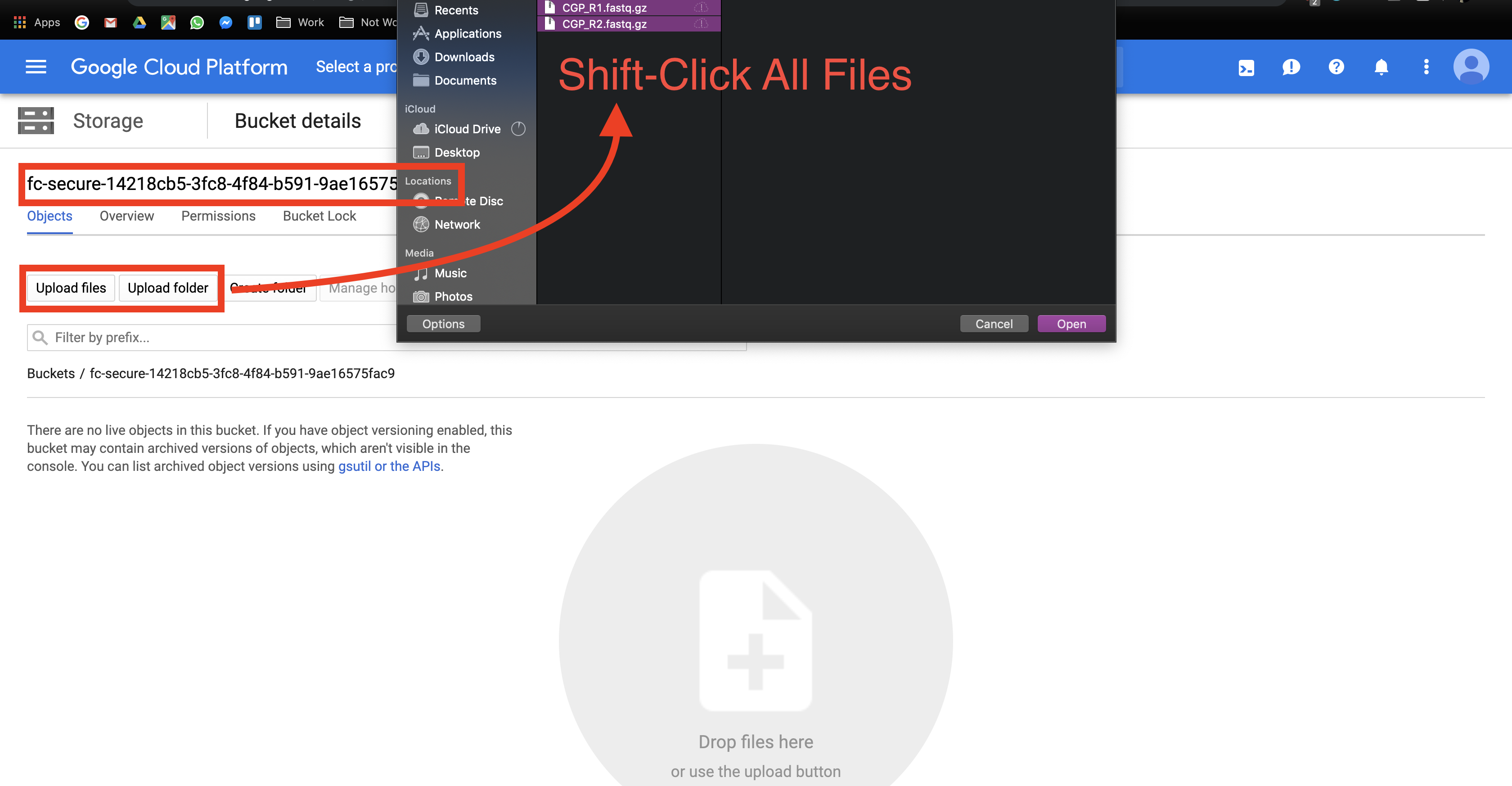
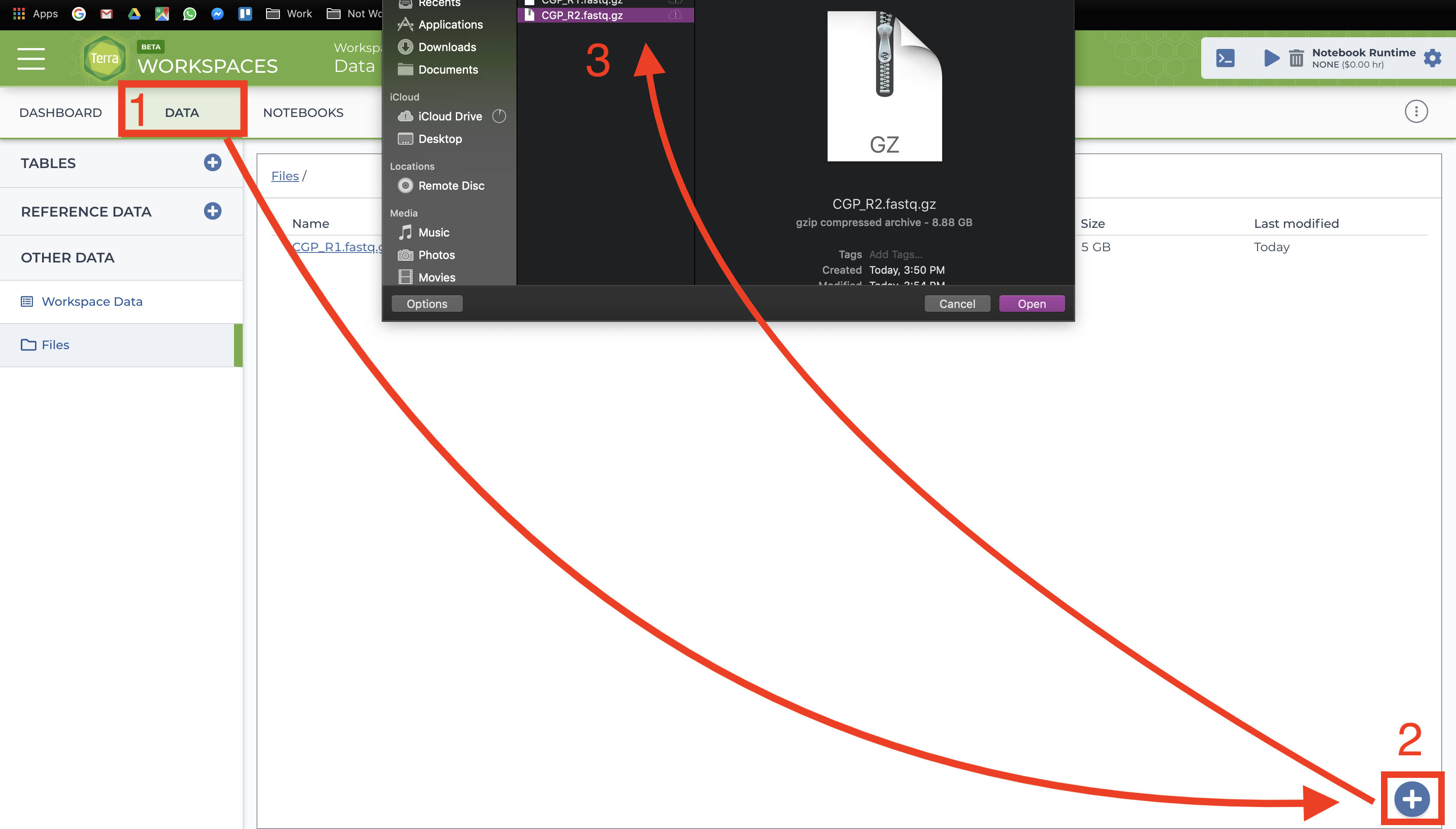
If you do not have a billing group, look [into creating a billing account](https://software.broadinstitute.org/firecloud/documentation/article?id=9762).

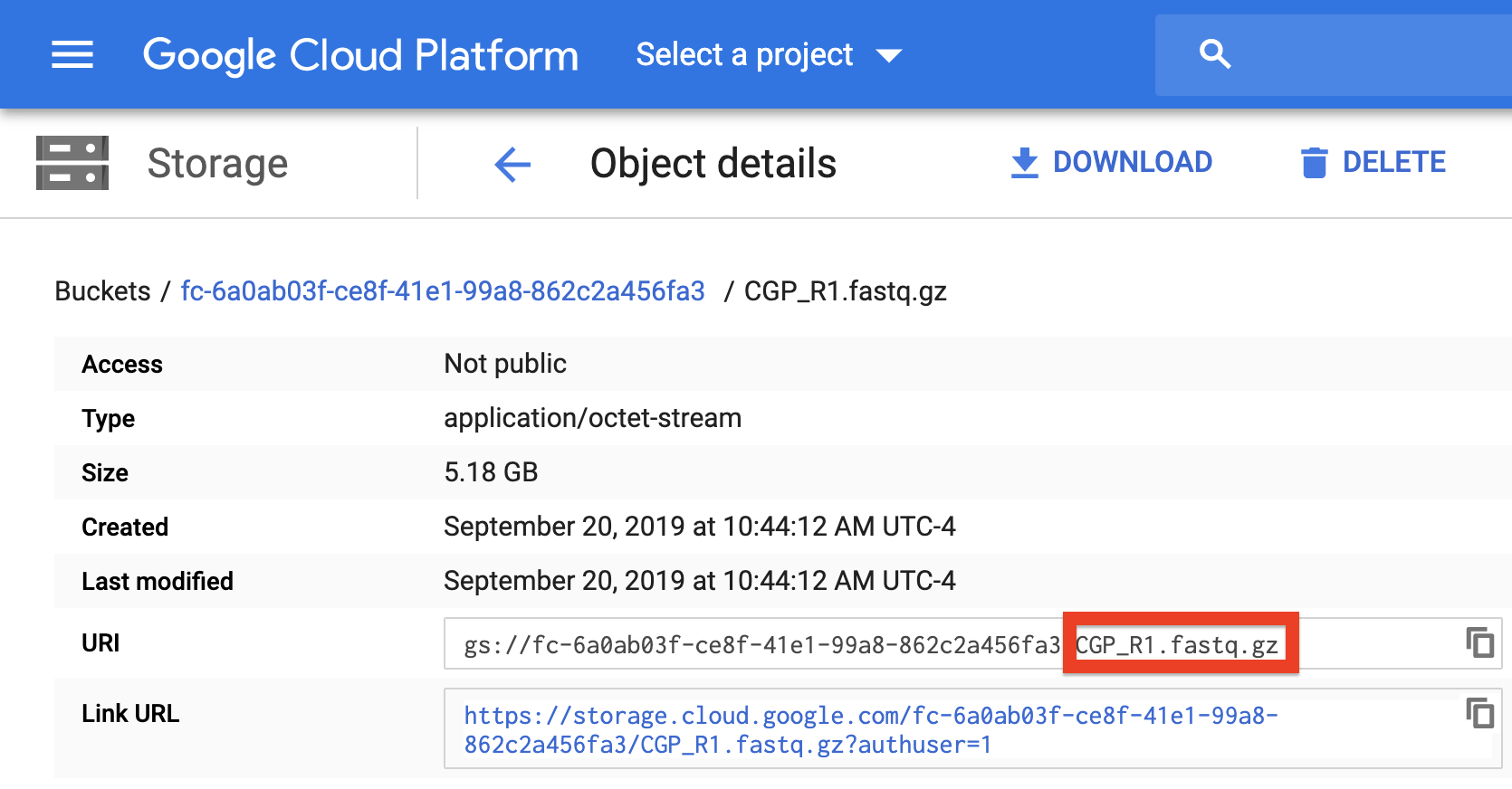
The authorization domain] is the group of users who will be able to access your workspace. You can create an authorization domain by following the instructions [here](https://support.terra.bio/hc/en-us/articles/360026775691).

### 3. Add Your Sequence Data and Input CSV File

There are three methods of uploading files to your workspace Google Bucket. Before proceeding, find your workspace bucket by visiting your workspace’s “Dashboard” tab. In the bottom-right corner of the dashboard, you will see your Google Bucket ID which you can copy by clicking the adjacent clipboard button. You can visit the bucket interface by clicking the “Open in Browser” hyperlink.

First upload your sequence data files using one of the methods below.

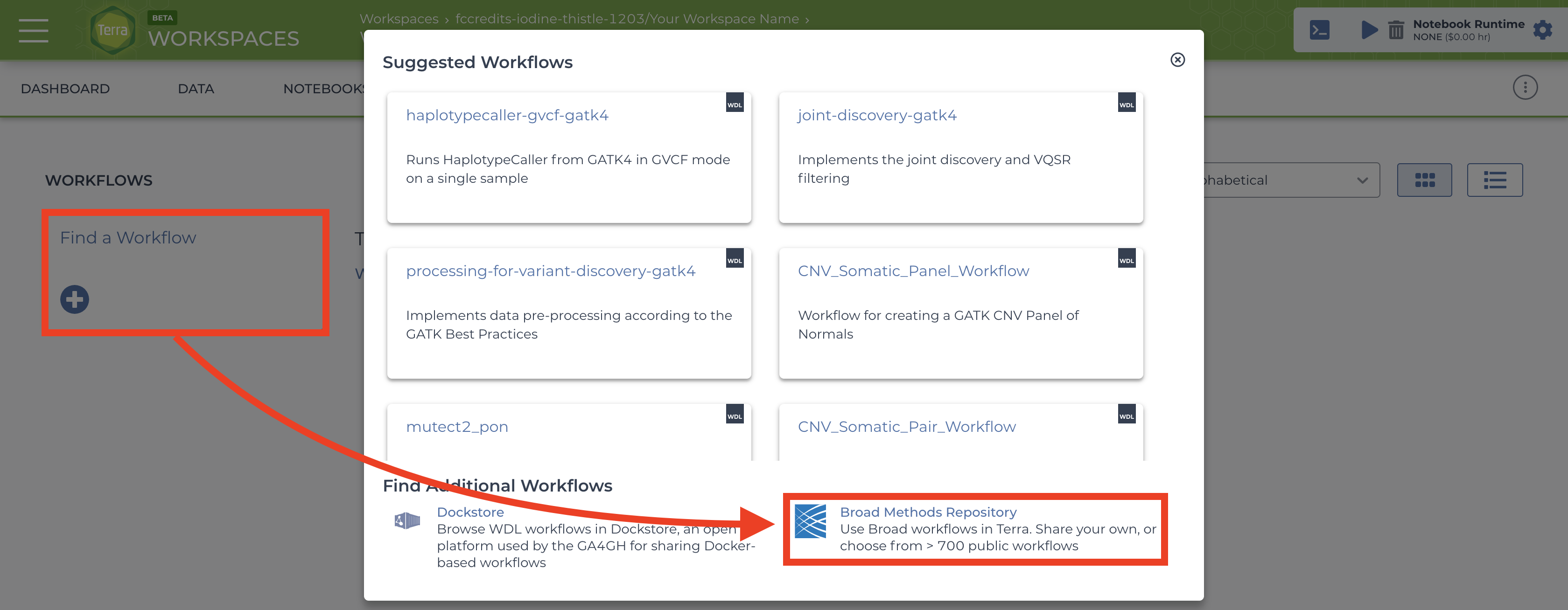
1. gsutil **HIGHLY RECOMMENDED**: Through your computer’s console, install the gsutil tool by following the [installation guide](https://cloud.google.com/storage/docs/gsutil_install). An example command that would transfer files from your computer to the workspace bucket would be:  
   gsutil -m cp local/path/to/file.fastq.gz gs://[Bucket ID]/destination/directory/  
   For an entire folder of sequence data, copy it recursively through using the command as such:  
   gsutil -m cp -r local/path/to/folder gs://[Bucket ID]/destination/directory.
2. You can also manually upload data to the Google Bucket, but note that this process take much more time than gsutil. At the bottom right of your workspace’s “Dashboard” tab, click the “Open in Browser” hyperlink to visit your bucket. Click either the “Upload File” or “Upload Folder” button to navigate to and upload your files or a folder containing your files respectively. 
3. Alternatively to gsutil and Google Bucket file uploading, users can manually upload data one file at a time through the Terra interface. This method is painfully slow so you should strongly consider using gsutil. Go to the “Data” tab and click the plus button towards the bottom-right of the page. Navigate to and select your file and hit open. Repeat the process for however many files. 

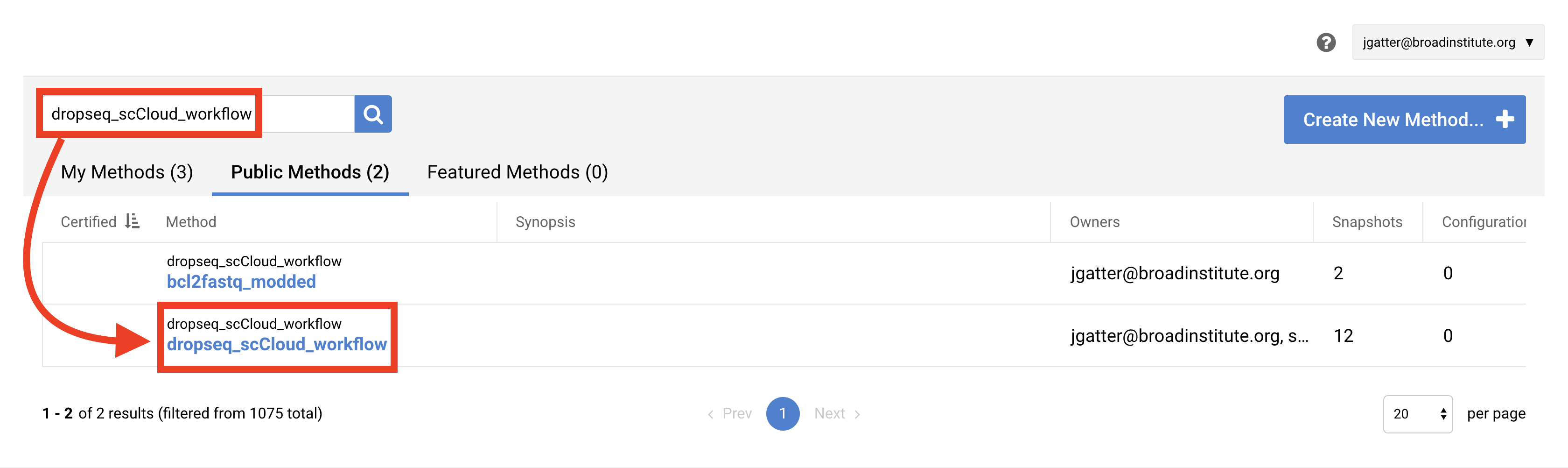
Before uploading your input CSV file, it is recommended that you make certain that it adheres to the criteria specified in the [documentation](/dropseq_scCloud/#formatting-your-input_csv_file). To verify that the paths you listed in the file are correct, navigate to your bucket using the instructions listed [above](/terra/#3-add-your-sequence-data-and-input-csv-file) and locate your sequence data files. Click on each file to view its URI (gsURL), which should resemble the format gs://<bucket ID>/path/to/file.fastq.gz in the case of gzip-compressed FASTQ files. The locations you should enter in the path columns of your input CSV file should be all of the characters following the bucket ID and trailing slash, in this case path/to/file.fastq.gz. 

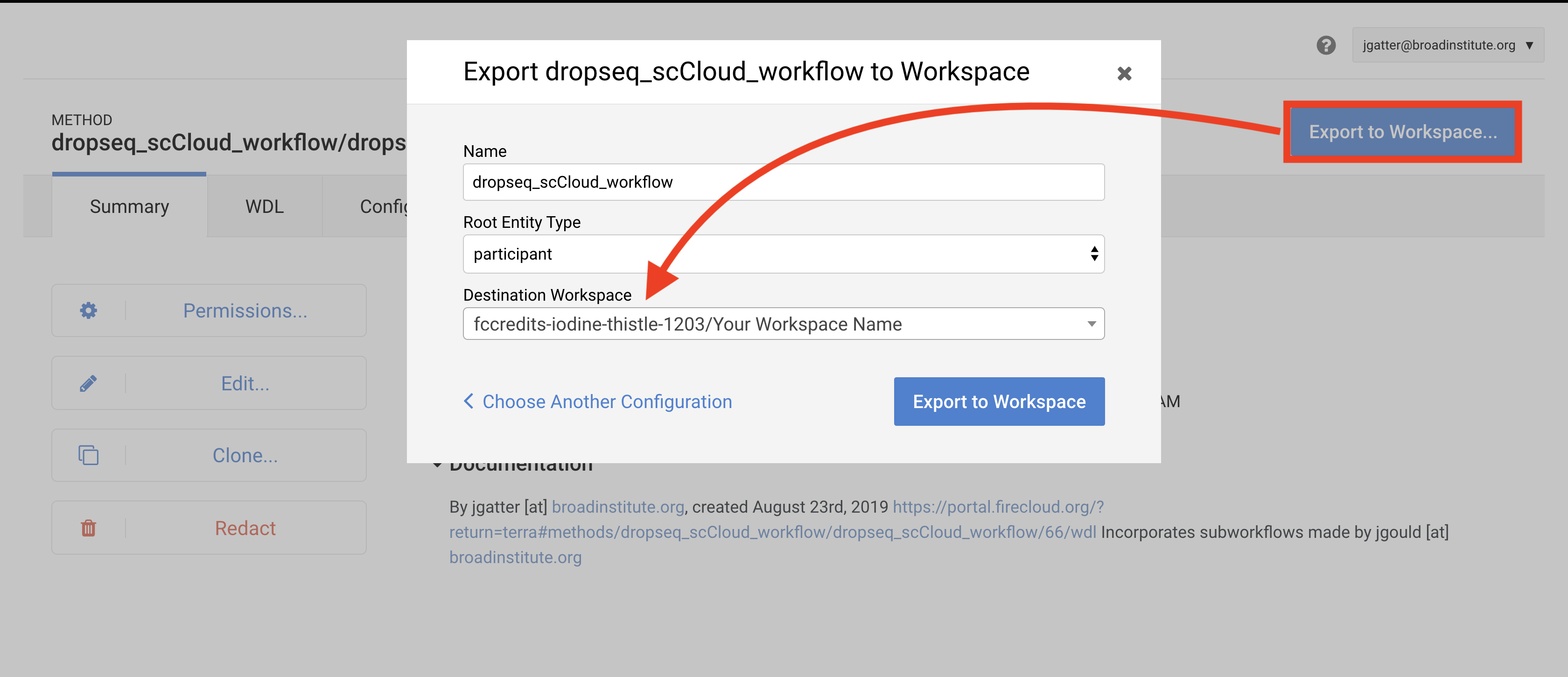
When finished, upload your input CSV file to the bucket.

### 4. Import the dropseq\_scCloud Workflow

This section may be subject to change in the recent future as Terra is still undergoing development and soon Firecloud Methods will be integrated to Terra rather than remain in the legacy application.

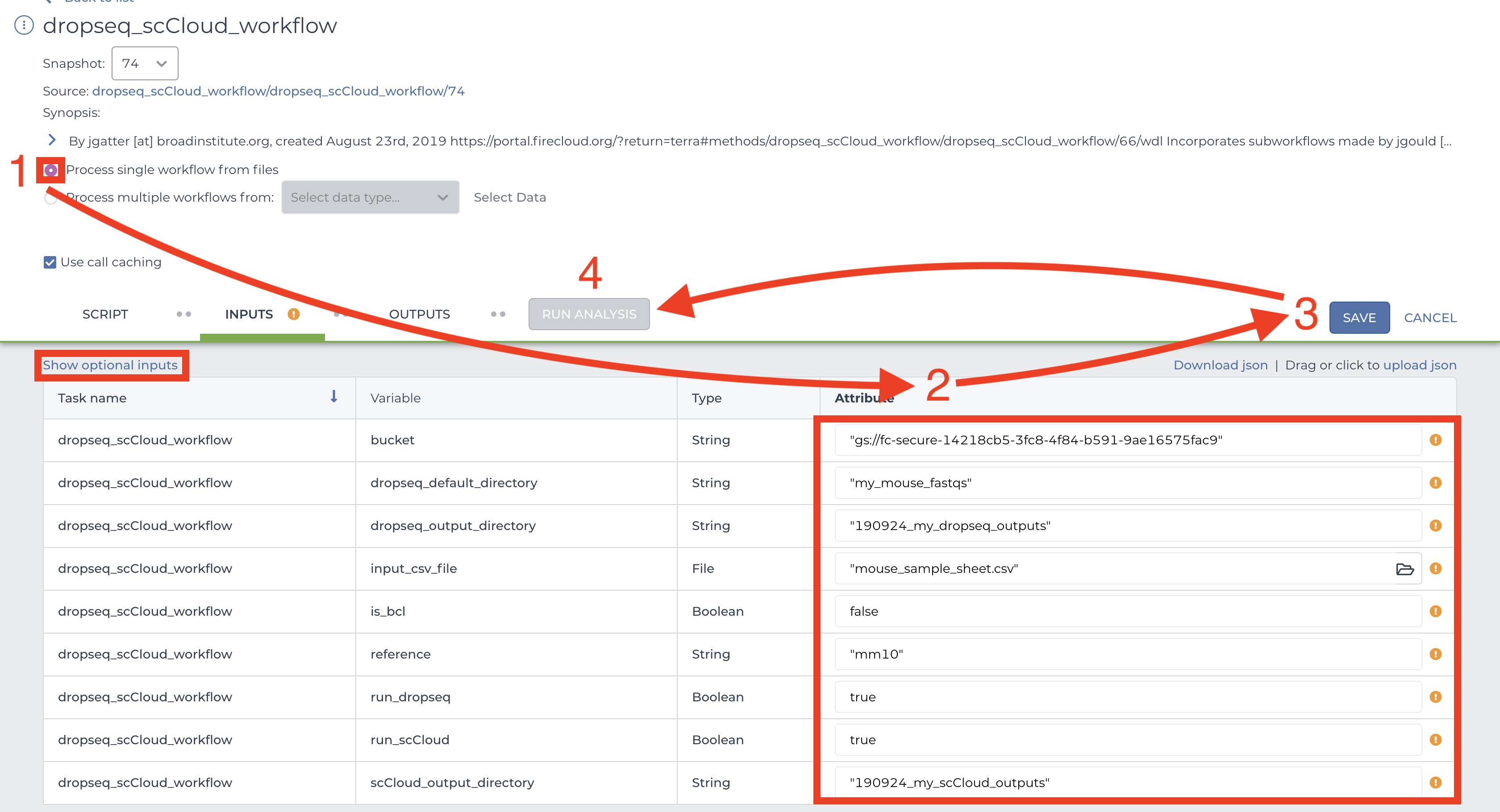
To import workflows for running jobs in your workspace, click the “Workflow” tab and click the “Find a Workflow” button. On the window that pops up, click the “Broad Method Repository” hyperlink to be brought to FireCloud Methods. .

In the search bar, enter “dropseq\_scCloud\_workflow”, hit search, and click on the dropseq\_scCloud\_workflow hyperlink when it appears. .

Towards the top-right of the workflow page, click “Export to Workspace…” button and then click the “Use Blank Configuration” button. Select your workspace as the destination workspace via the bottom-most dropdown menu. Click the “Export to Workspace” button and when it offers you to visit the edit page, click “Yes”. 

### 5. Configure and Launch dropseq\_scCloud\_workflow

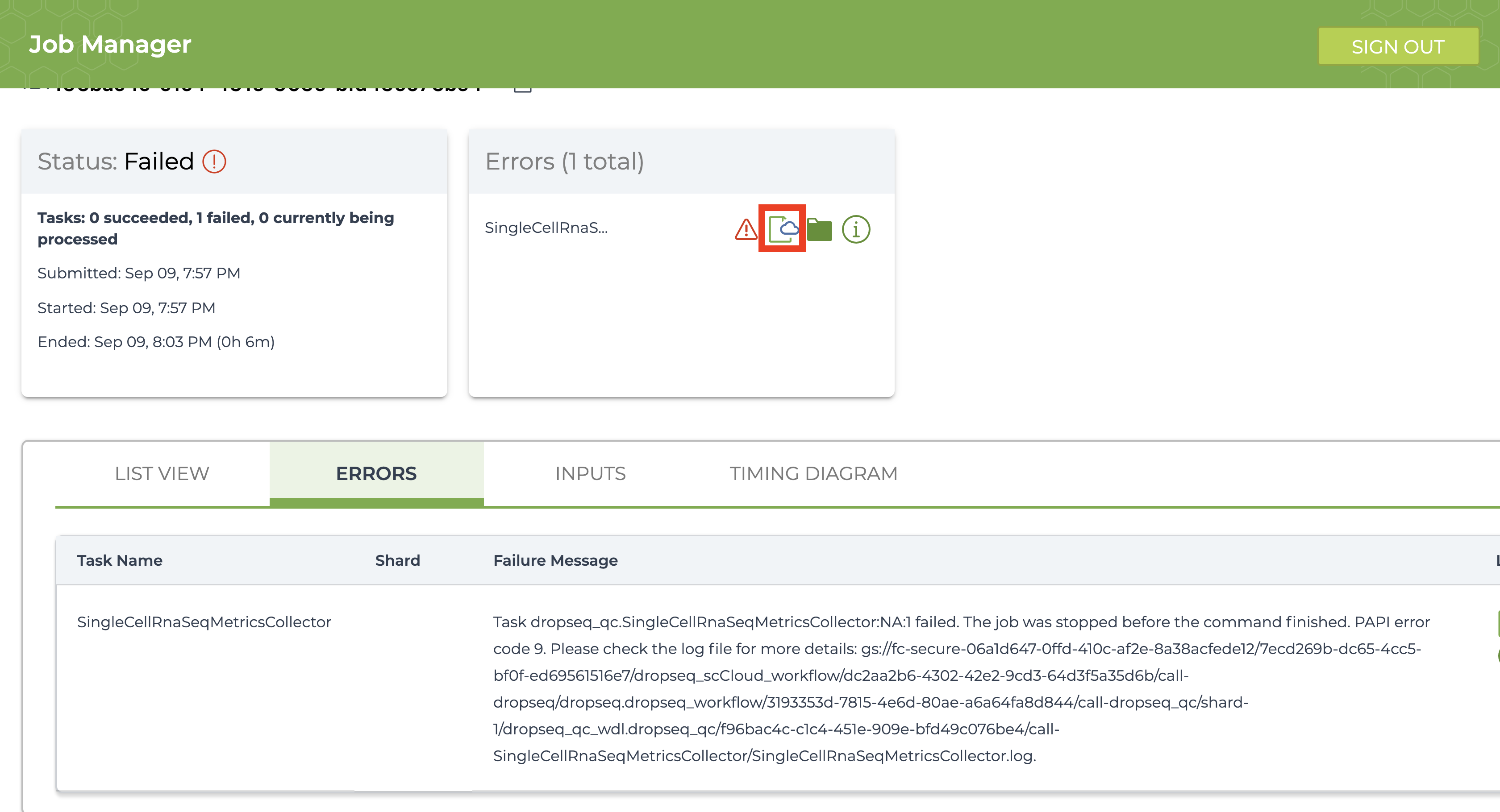
You should be on the edit page of the tool now, but to access this page of the tool in the future, you can go to the Workflow tab of your Terra Workspace and click on the dropseq\_scCloud\_workflow button.

This page is where you set variables for dropseq\_scCloud\_workflow, Drop-seq pipeline workflow, and scCloud workflow. First click the bubble for “Process single workflow from files.” Scroll down to fill out required inputs. Note that Strings must be surrounded by quotation marks and typing true and false (don’t surround with quotation marks) for Booleans are case-sensitive. 

To set values for optional parameters, click the “Show optional inputs” text above the “Task name” column and scroll down. These are variables that have default values if the boxes are left empty, so there is no need to fill each out unless it suits your needs. To better understand the “dropseq” and “scCloud” inputs, read the [Drop-seq pipeline](https://sccloud.readthedocs.io/en/latest/drop_seq.html#inputs) and [scCloud](https://sccloud.readthedocs.io/en/latest/scCloud.html#aggregate-matrix) documentations. It should be mentioned that each workflow’s required inputs have been overridden by dropseq\_scCloud workflow (e.g. input\_csv\_file, output\_directory, etc.), so refer to the dropseq\_scCloud documentation for those related variables.

When satisfied with your inputs, save them using the button in the upper-right and and then hit the “RUN ANALYSIS” button to the left. You will be brought to the job history page where your job has been logged. Here you will eventually know if the job ran succesfully. Provided the job does not fail within 10 minutes, the job will take about a variable amount of time to complete depending on the tasks you are running and the amount of data you gave it. Drop-seq pipeline usually will take 20 to 45 hours to run while scCloud should take an hour or less.

### 6. Advice for Troubleshooting

If the job fails it is recommended you navigate to and read the log file of the task that failed.  
To do this, click the failed job in “Job History” tab and click “View” on the next page. Then on the workflow status page under the “List View” tab, click the hyperlink of the failed task and repeat if there are failed subtasks. On the failed subtask that has no children, click on the log file button and skim to gain a better understanding. 

Evaluate the error based on the message and decide whether you need to alter variables, move files in your bucket, or change and reupload your input CSV file.

### 7. Uploading scCloud Output Files to Alexandria for Visualization

The explicit workflow outputs of dropseq\_scCloud\_workflow are the alexandria metadata file (alexandria\_metadata.txt), the dense expression matrix (ends with .scp.expr.txt), and the two coordinate files (end with .scp.X\_diffmap\_pca.coords.txt and .scp.X\_fitsne.coords.txt). Download these to your computer by visiting your Google Bucket through the workspace dashboard and then reupload via the first four tabs of the Upload/Edit Study page of the Alexandria workspace (See [Running dropseq\_scCloud on the Alexandria](/alexandria/) for a better understanding). Alternatively, use gsutil to transfer the files from your Terra bucket to your Alexandria bucket

Synchronize your Single Cell Portal study to account for the added files. Visualize the study by clicking the “Explore” tab and then the “View Options” hyperlink to gain more options for analysis. 