SC Jamboree Copying files from Synapse to Anvil workspace

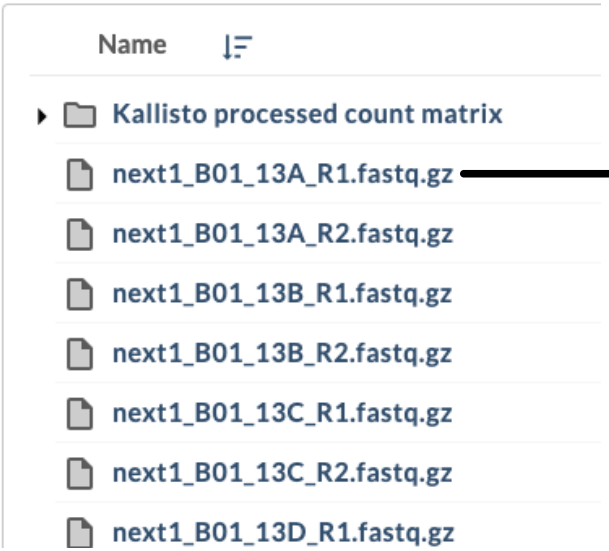
Kundaje Lab,

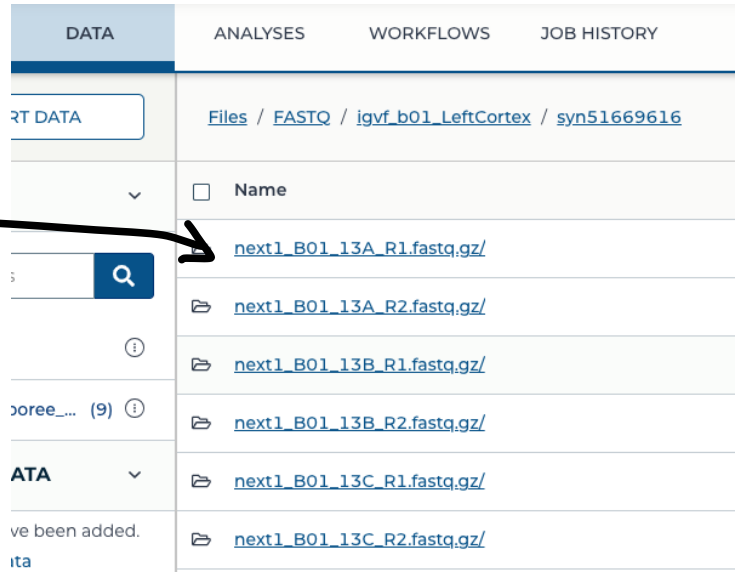
Eila Arich-Landkof

What?

https://www.synapse.org/#!Synapse:syn51669616

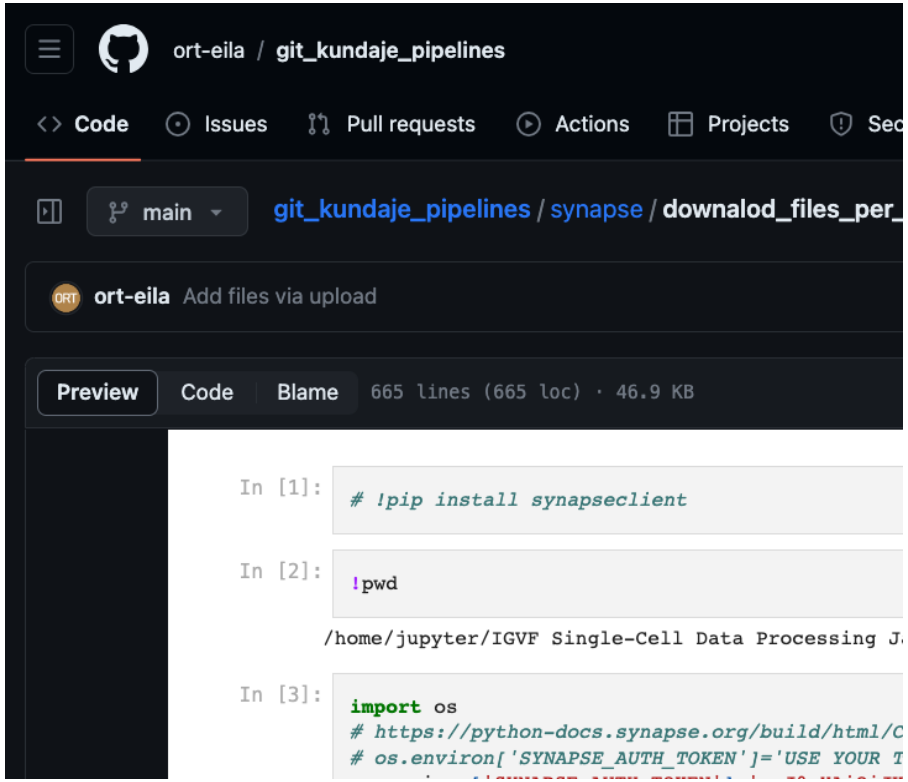
gs://fc-secure-0a879173-/FASTQ/igvf_b01_LeftCortex/syn51669616/



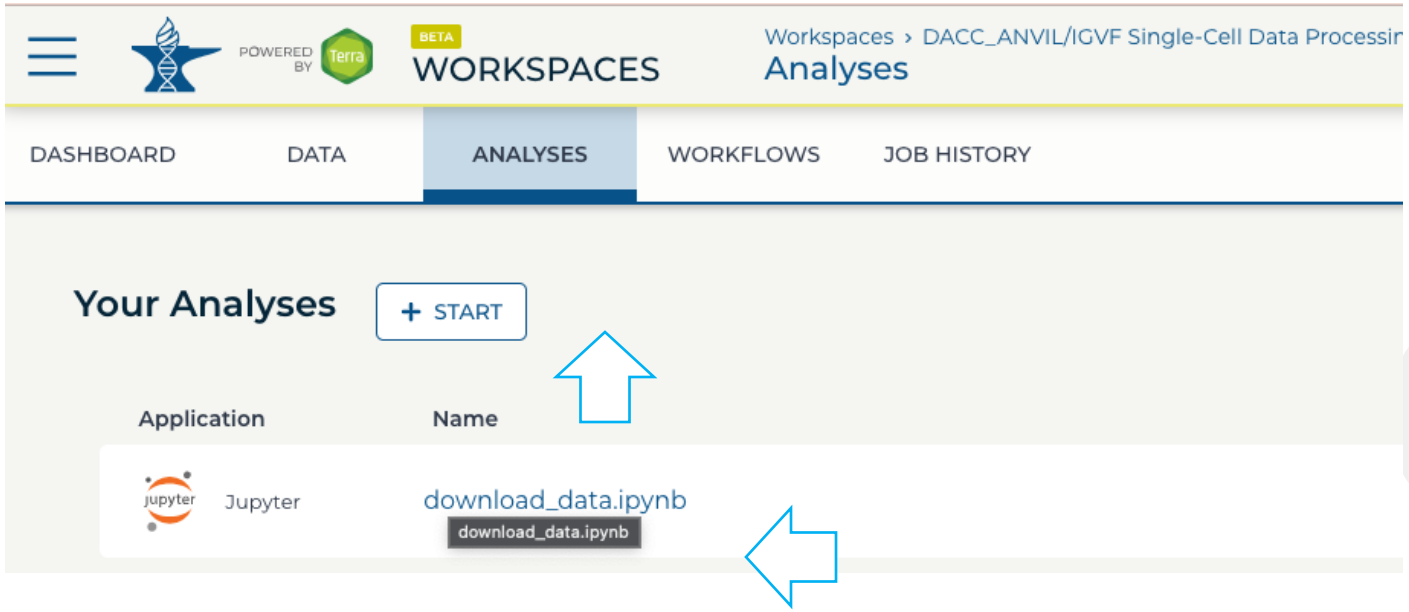


How? – programmatically

[Github link]



Anvil workspace



- Saves time & Bugs

Script Steps

https://www.synapse.org/#!Synapse:syn51669616

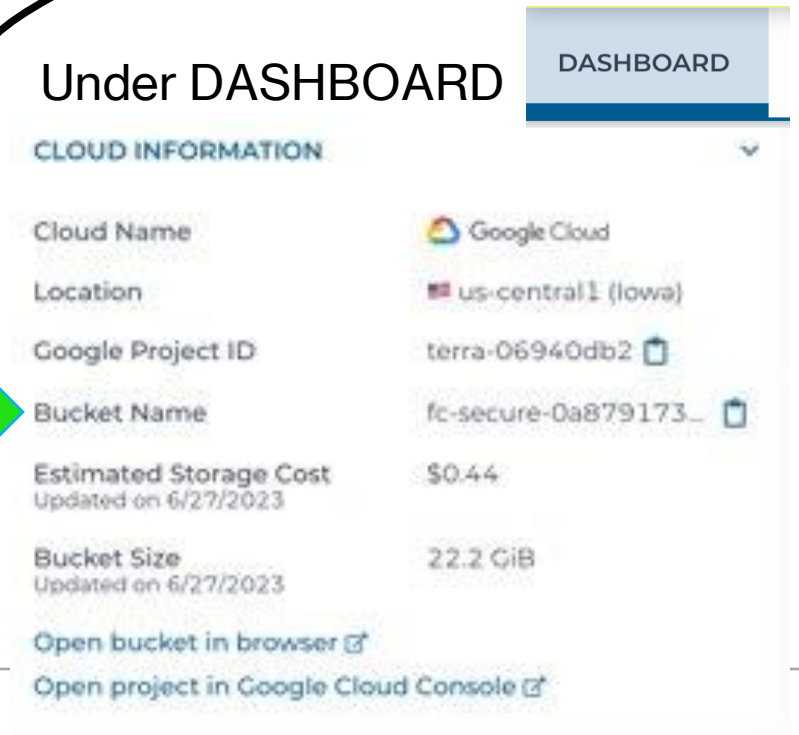
gs://bucket name/FASTQ/project name/synapse ID/

1. Get the list of files from synapse based on its dataset id
   - https://www.synapse.org/#!Synapse:syn51669616
2. Create the destination path for the Anvil workspace
   - gs://bucket name/FASTQ/project name/synapse ID/
   - gs://fc-secure-0a879173-62d3-4c3a-8fc3-e35ee4248901/FASTQ/igvf_b01_LeftCortex/syn51669616/
3. Loop on the list of files that are available on synapse dataset and copy to destination
   1. If you run it more than once, files that were downloaded, will not be downloaded again

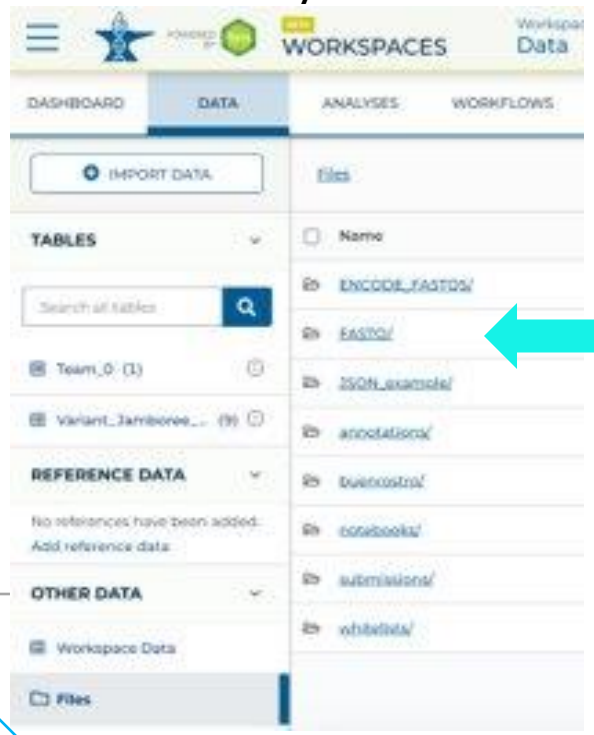
Destination path on Anvil

- A path on Google Cloud Platform (GCP)
  - gs://bucket name/FASTQ/project name/synapse ID/
  - gs://fc-secure-0a879173-62d3-4c3a-8fc3-e35ee4248901/FASTQ/igvf_b01_LeftCortex/syn51669616/

Under DASHBOARD

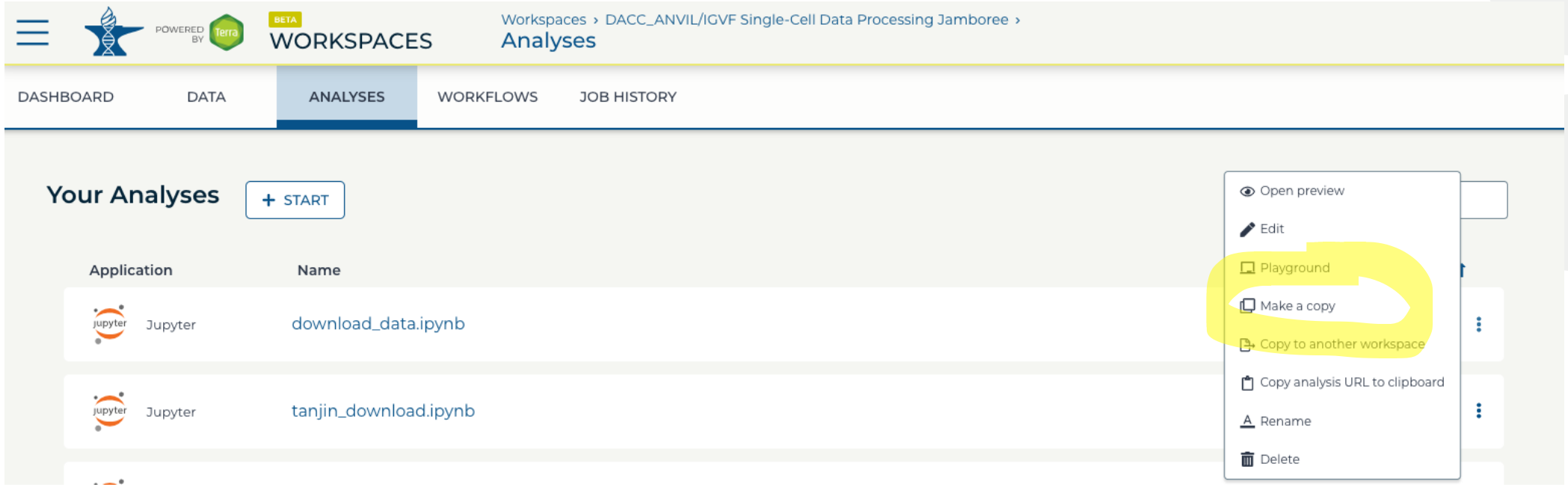


Under DATA / FILES



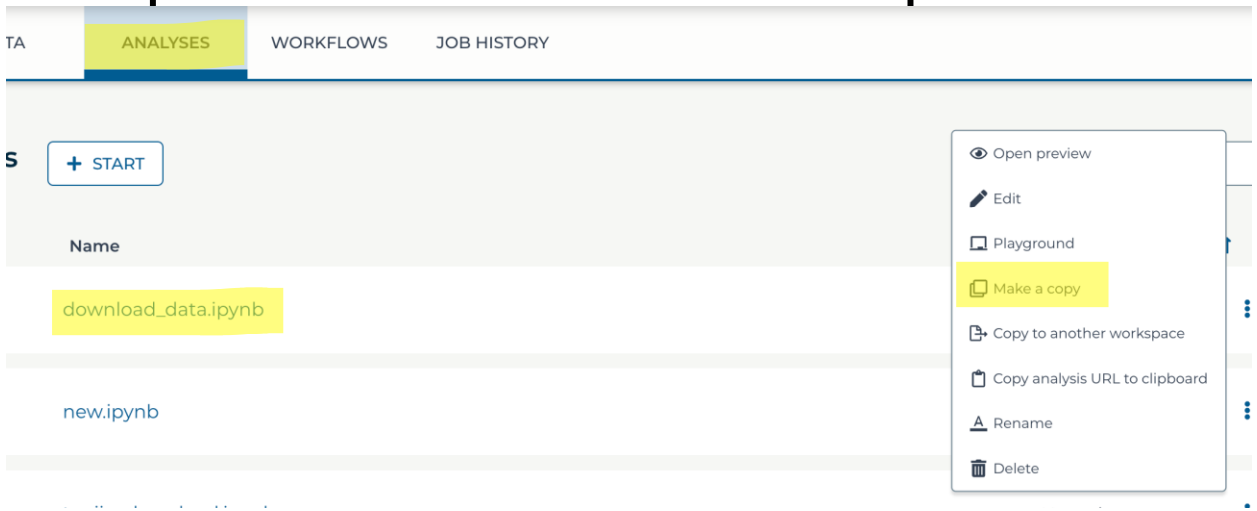
How to run - Step 1

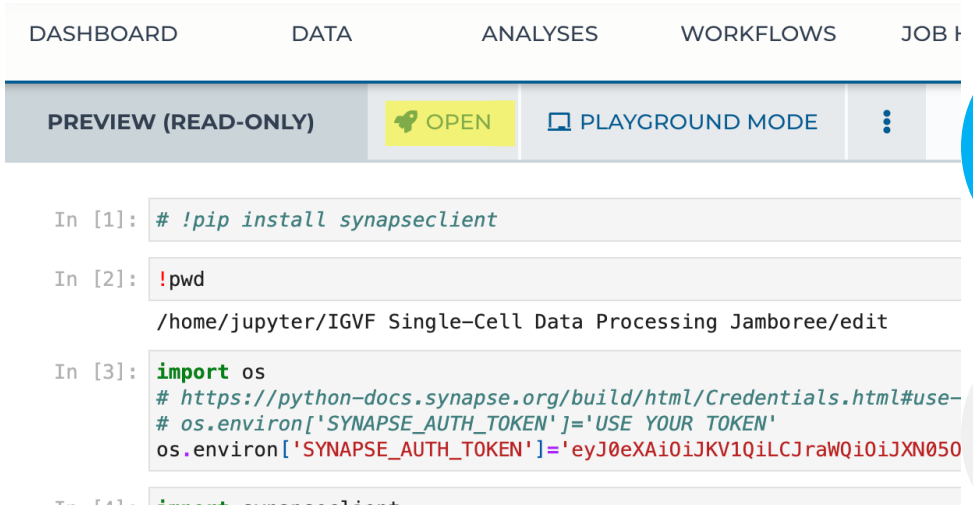
- Make a copy of the script from the Anvil workflow OR from git to your local machine



How to run – Step 2 - Make your own copy for the script

- Make a copy for the script and rename it: XXXX_download_file.ipynb
- Open the file and click on open

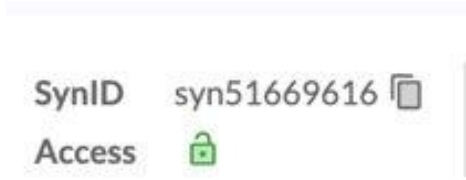




How to run – Step 2 – Update variables

- Search for TODO on the script

Synapse ID to copy from




```
# TODO: change the dataset id
dataset_ids = list(set(["syn51669616","syn51730625"]))
print(len(dataset_ids))
```

Create local folder and look for the following code to change

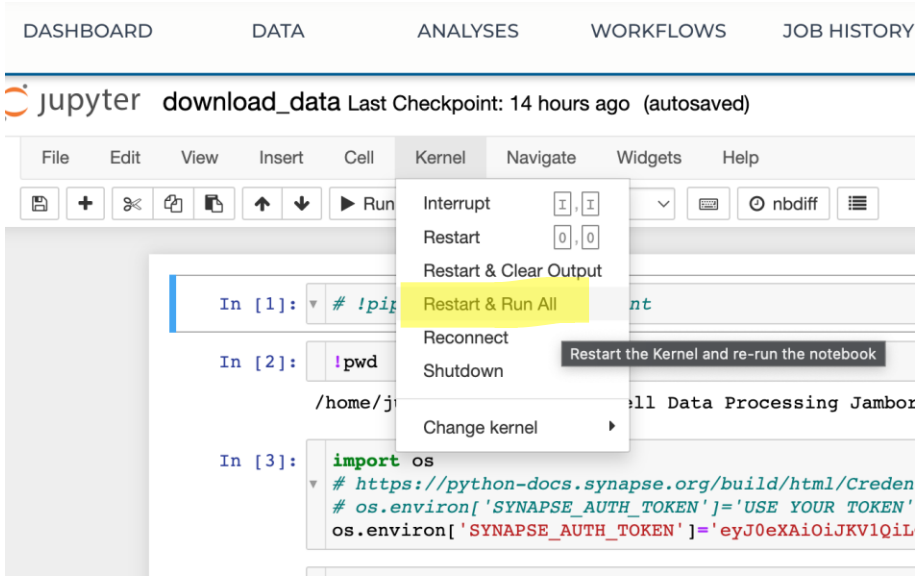
```
# TODO: change this to a local folder that is on your machine
local_file_path="/home/jupyter/eila/"
```

Project name to use under FASTQ

```
# TODO: change here the name of the dataset that you are downloading. For example: "igvf_b01_LeftCortex"
gs_file_path = os.path.join(gs_root_file,"igvf_b01_LeftCortex",dataset_id,dst_file_name)
```

Execute & Compare

- Compare the files
  - https://www.synapse.org/#!Synapse:syn51669616 (synapse – source)
  - https://anvil.terra.bio/#workspaces/DACC_ANVIL/IGVF%20Single-Cell%20Data%20Processing%20Jamboree/data (under workspace data - destination)
- Sometime the Jupyter is not stable and will not complete the upload
  - Just rerun it!
  - The code will not repeat upload that was already completed



Let me know if you have any issue