

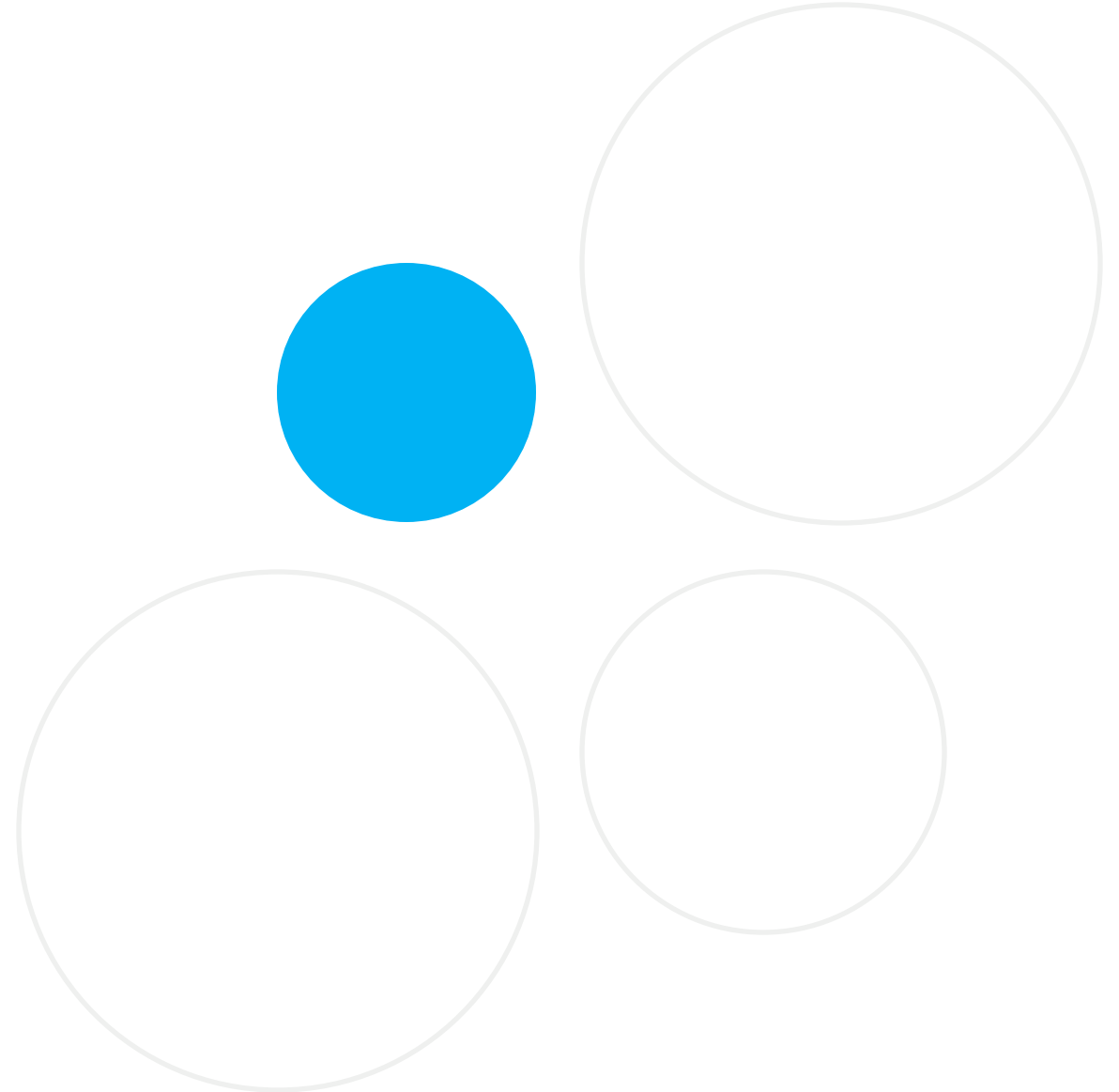
# **SC Jamboree**

## **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/



Name	
▶ Kallisto processed count matrix	
next1_B01_13A_R1.fastq.gz	
next1_B01_13A_R2.fastq.gz	
next1_B01_13B_R1.fastq.gz	
next1_B01_13B_R2.fastq.gz	
next1_B01_13C_R1.fastq.gz	
next1_B01_13C_R2.fastq.gz	
next1_B01_13D_R1.fastq.gz	

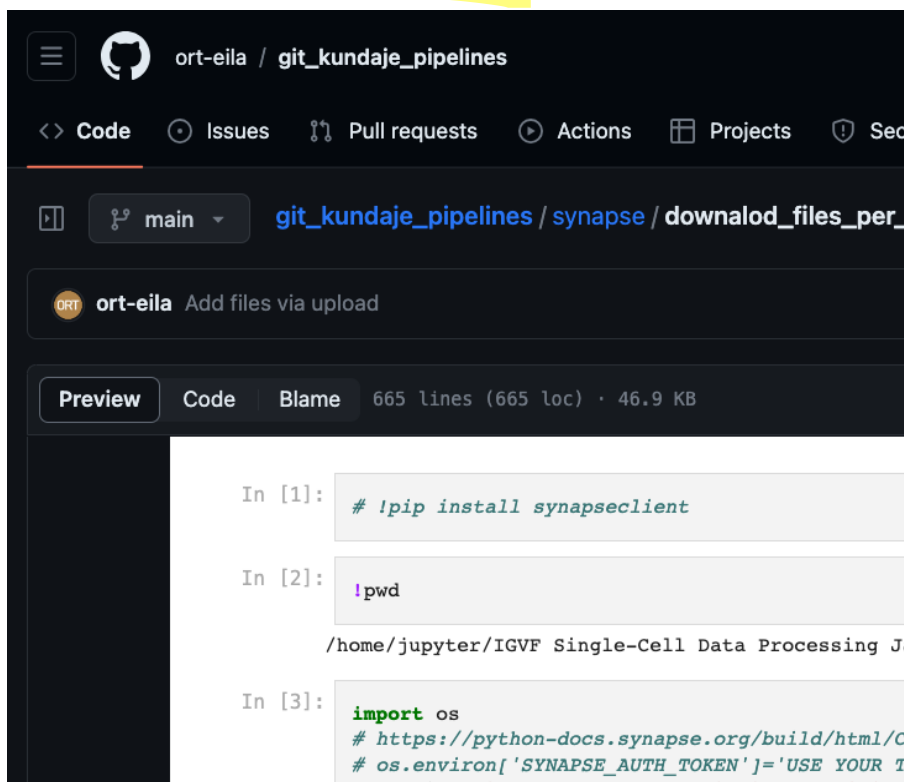


DATA	ANALYSES	WORKFLOWS	JOB HISTORY
RT DATA	Files / FASTQ / igvf_b01_LeftCortex / syn51669616		
	<input type="checkbox"/> Name		
	next1_B01_13A_R1.fastq.gz/		
	next1_B01_13A_R2.fastq.gz/		
	next1_B01_13B_R1.fastq.gz/		
	next1_B01_13B_R2.fastq.gz/		
	next1_B01_13C_R1.fastq.gz/		
	next1_B01_13C_R2.fastq.gz/		

# How? – programmatically, python

- Saves time & Bugs

[Github link](#)



The screenshot shows the GitHub interface for the repository 'ort-eila / git\_kundaje\_pipelines'. The 'Code' tab is selected, showing a file named 'download\_data.ipynb'. The file content is displayed in a code editor with syntax highlighting. The file has 665 lines (665 loc) and is 46.9 KB in size.

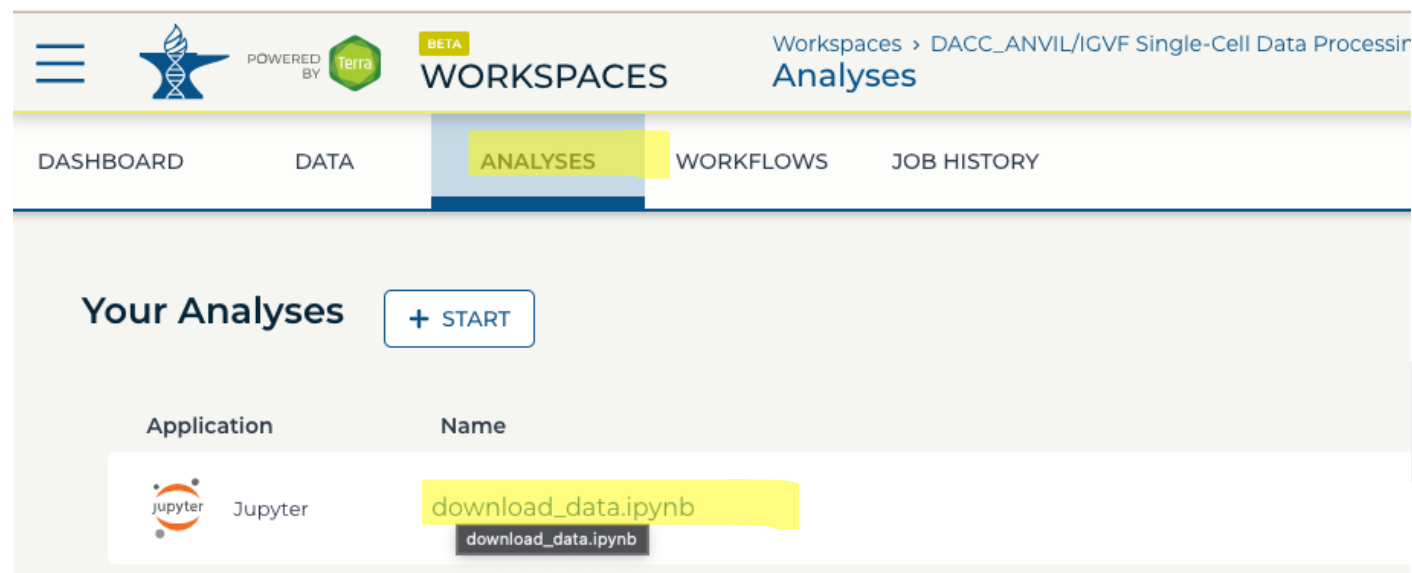
```
In [1]: # !pip install synapseclient

In [2]: !pwd

/home/jupyter/IGVF Single-Cell Data Processing J

In [3]: import os
# https://python-docs.synapse.org/build/html/C
# os.environ['SYNAPSE_AUTH_TOKEN']='USE YOUR T
```

## Anvil workspace



The screenshot shows the Anvil workspace interface. The top navigation bar includes a menu icon, the Anvil logo, 'POWERED BY Terra', a 'BETA' badge, and the text 'WORKSPACES'. The breadcrumb trail shows 'Workspaces > DACC\_ANVIL/IGVF Single-Cell Data Processing > Analyses'. The main navigation bar has tabs for 'DASHBOARD', 'DATA', 'ANALYSES' (which is highlighted), 'WORKFLOWS', and 'JOB HISTORY'. The main content area is titled 'Your Analyses' and features a '+ START' button. Below this, there is a table with two columns: 'Application' and 'Name'. The table contains one entry for 'Jupyter' with the name 'download\_data.ipynb' highlighted in yellow.

Application	Name
Jupyter	download_data.ipynb

# Script Steps

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 files in 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. Copy each file from the Synapse source to its workspace destination
  - If you run the script more than once, files that were downloaded, will not be downloaded again



# Destination path on Anvil workspace

- 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

**DASHBOARD**

**CLOUD INFORMATION**

Cloud Name	Google Cloud
Location	us-central1 (Iowa)
Google Project ID	terra-06940db2
Bucket Name	<b>fc-secure-0a879173-</b>
Estimated Storage Cost Updated on 6/27/2023	\$0.44
Bucket Size Updated on 6/27/2023	22.2 GiB
<a href="#">Open bucket in browser</a>	
<a href="#">Open project in Google Cloud Console</a>	

## Under DATA / FILES

**WORKSPACES**

**DATA**

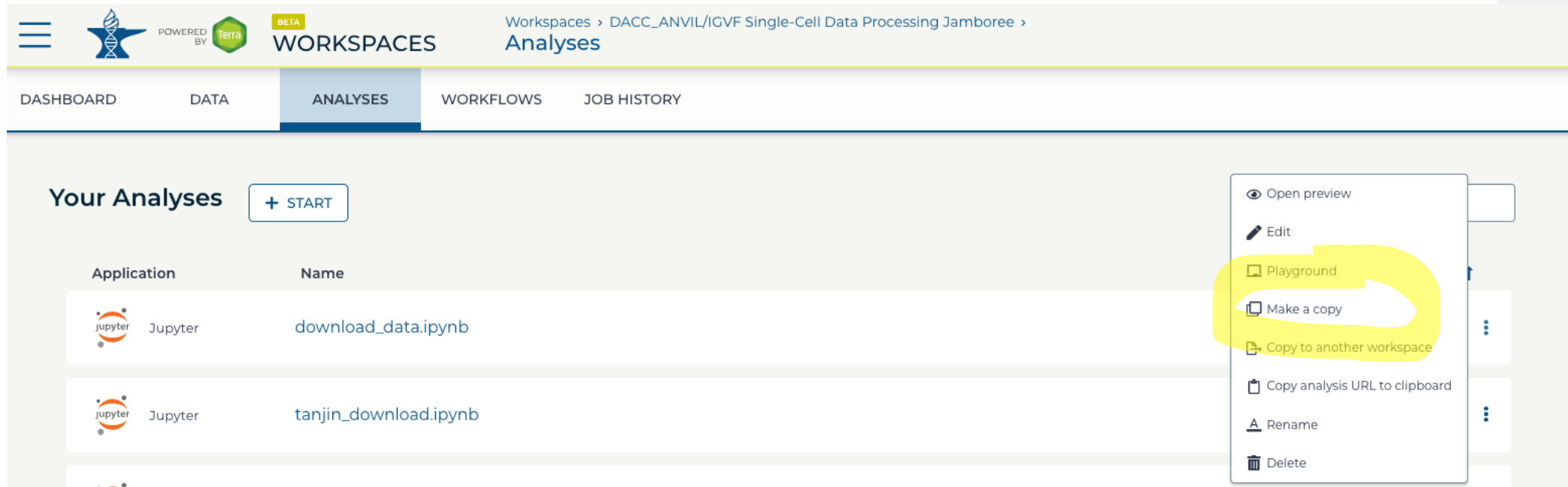
**FILES**

Name
ENCODE_FASTQ/
<b>FASTQ/</b>
ISO9601_example/
annotations/
biocontrol/
corebooks/
submissions/
whitetail/

**Files:** this is not a folder, only a link to the folders under the bucket.

# How to run the cop script?

- Step 1: **Make a copy** of the script from the Anvil workflow OR from git to your local machine

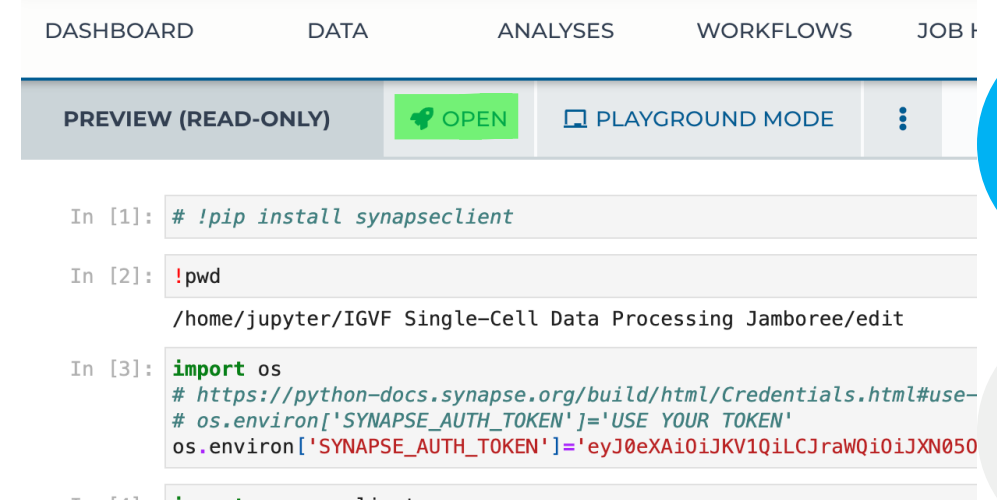
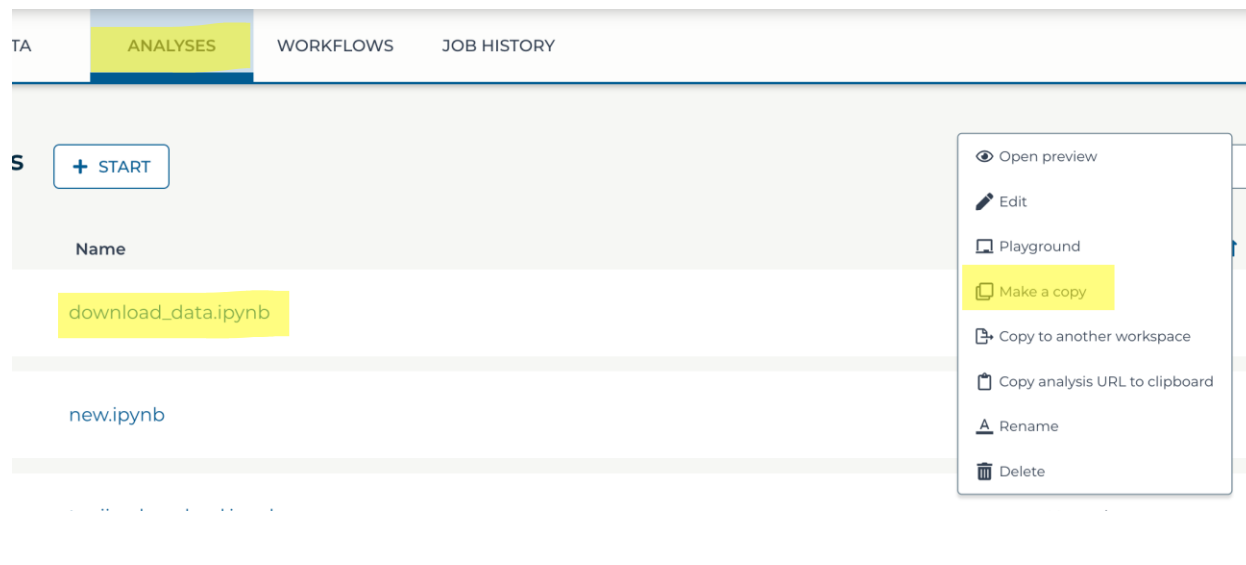


The screenshot displays the Anvil Workspaces interface. At the top, there's a navigation bar with a menu icon, the Terra logo, and the text 'POWERED BY Terra'. Below this, the 'WORKSPACES' section is active, showing the path 'Workspaces > DACC\_ANVIL/IGVF Single-Cell Data Processing Jamboree > Analyses'. The main navigation bar includes 'DASHBOARD', 'DATA', 'ANALYSES' (highlighted), 'WORKFLOWS', and 'JOB HISTORY'. The 'Your Analyses' section features a '+ START' button and a table of analyses. The table has two columns: 'Application' and 'Name'. Two analyses are listed: 'Jupyter' with 'download\_data.ipynb' and 'Jupyter' with 'tanjin\_download.ipynb'. A context menu is open over the 'download\_data.ipynb' file, listing actions: 'Open preview', 'Edit', 'Playground', 'Make a copy' (highlighted with a yellow circle and an arrow), 'Copy to another workspace', 'Copy analysis URL to clipboard', 'Rename', and 'Delete'.

Application	Name
Jupyter	download_data.ipynb
Jupyter	tanjin_download.ipynb

# How to run – Step 2 – edit mode

1. **Rename your script:** XXXX\_download\_file.ipynb
2. Open your script in **edit mode.**



# How to run – Step 2 – Update variables

- Personalize the script with your values.
- Search for "TODO" in the script.

1. Synapse ID to copy from

SynID syn51669616   
Access 

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

2. Create local folder and replace the path below

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

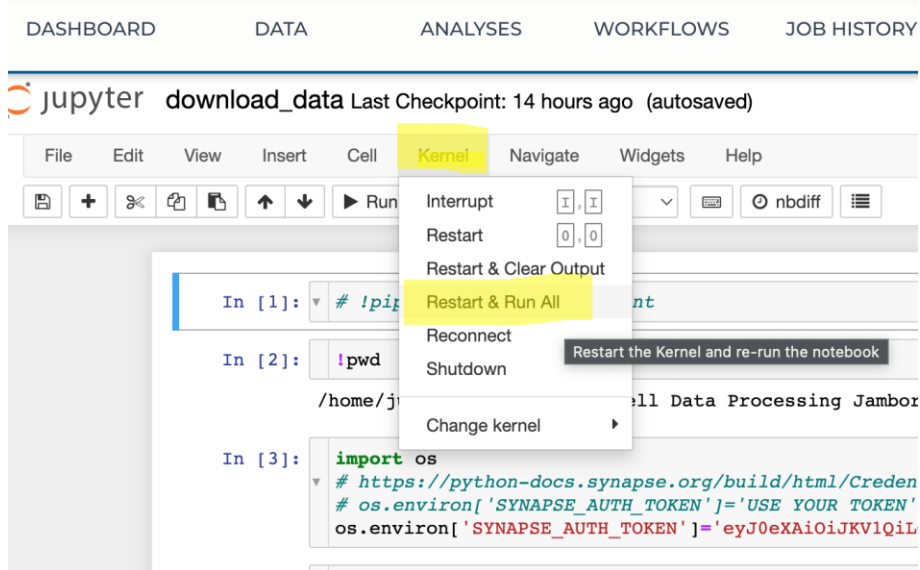
3. Update the path on Anvil with your project name

 igvf\_b01\_LeftCortex/

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



# Execute & Compare files



- Compare the copied files with your data on Synapse files
  - Source:  
<https://www.synapse.org/#!/Synapse:syn51669616>
  - Destination: under Data => Files in your workspace
- Jupyter is not stable and might terminate
  - Just execute it again
  - The code will not repeat the upload process for files that .

**Let me know if you have any  
issue**

