	Linking BaseSpace and Importing BaseSpace Reads to Terra Workspace	
	Document TG-TER-04, Version 2	
	Date:	Workflow Versions:
	4/1/2024	PHB v2

1.   PURPOSE/SCOPE

To standardize the process of uploading next generation sequencing (NGS) data from BaseSpace into Terra. See appendix 10.2 for details on initial setup to link BaseSpace and Terra accounts. Acceptable data types include Illumina raw sequencing file formats.

2.   REQUIRED RESOURCES

- Computer
- Internet connection: at least 10 and 5Mbps for download and upload speeds, respectively
- Internet browser
  - Google Chrome, Firefox, or Edge
- Google account
- Terra account, linked to Google account
- BaseSpace account, linked to Terra account; see Appendix 10.2 for details on linking accounts

3.   RELATED DOCUMENTS

Document Number	Document Name
None	None

4.   PROCEDURE

Initial Setup Required! Labs using BaseSpace Fetch for the first time must begin by importing the BaseSpace_Fetch_PHB wofklow and using command line to provide BaseSpace credentials to Terra to pull read files. **Start with appendices 10.1 and 10.2.**

4.1   DOWNLOAD BASESPACE SAMPLE SHEET AND IMPORT TO TERRA

1. *Login to the BaseSpace* account where samples were run at
   https://basespace.illumina.com/
2. In the runs tab, *select the run* to be analyzed (Fig 1)

Figure 1.

3. Within the files tab, scroll to the bottom and *download the sample sheet* (Fig 2 and 3)
4. *Open the sample sheet* in excel
5. Click the plus symbol at the bottom of the window to *open a new sheet* or open a new workbook (Fig 4)
6. *Copy the list of sample names* including the header "Sample_Name" beginning with cell A19; paste twice into the new sheet starting in cells A1 and B1 (Fig 5)
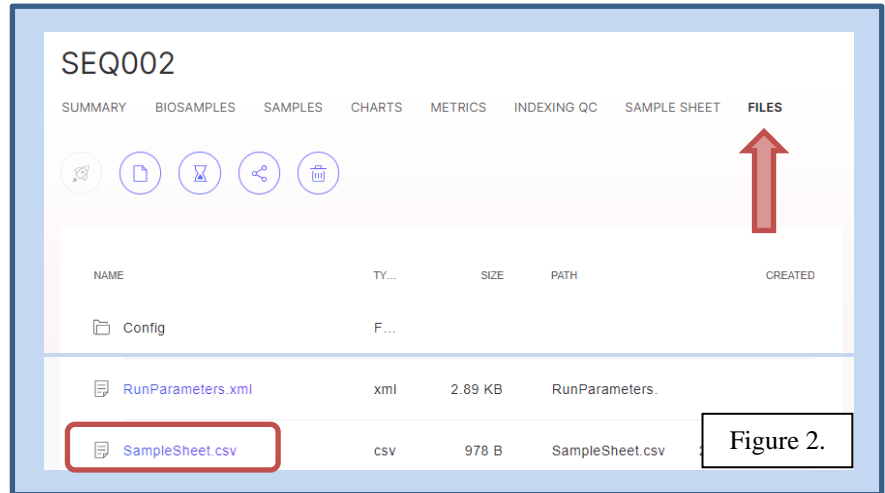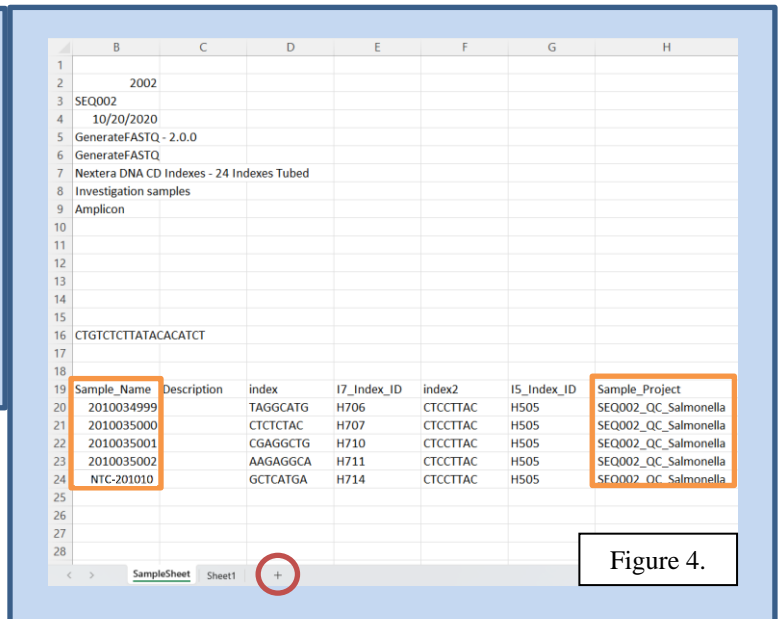

Figure 2.

Figure 3.

Figure 4.

7. *Copy/paste the "Sample_Project"* column starting in cell C1 (Fig 5)
8. *Rename the column headers* (Fig 6):
   a. Column A: *entity:PulseNet_id*, where *PulseNet* is the name of the data tabe in Terra
   b. Column B: *basespace_sample_name*
   c. Column C: *basespace_collection_id*

Figure 5.
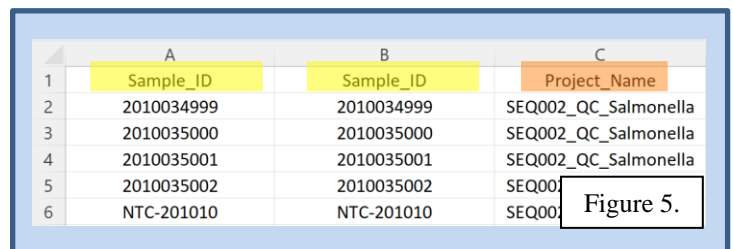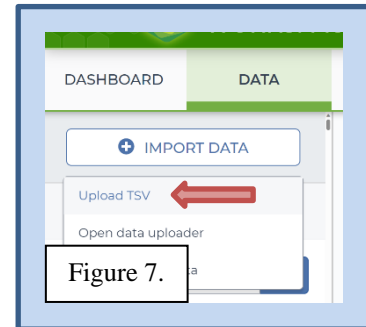
Figure 6.

Figure 7.

9. *Delete other tabs* in the workbook and *save the file* in .txt or .tsv format to a shared lab location

10. In the Terra workspace where data will be uploaded, click *import data* and *upload tsv* (Fig 7)

11. In the "Import Table Data" pop-up window, *drag and drop* or click to select the file from the shared lab location; click *start import job* (Fig 8)

   a. *If cell A1 nomenclature is not formatted beginning and ending with* *entity:* *and* *_id* , *respectively, an invalid format message will appear (Fig 9)*

12. *A new data table should now be listed in the tables sidebar or, if adding to an existing table, data should be populated in the relevant table*


Figure 8.
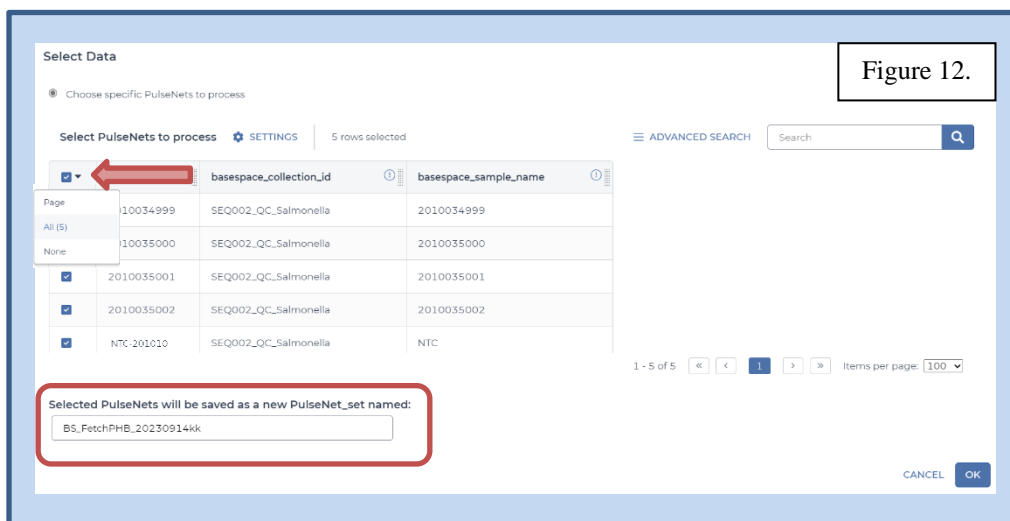

Figure 9.
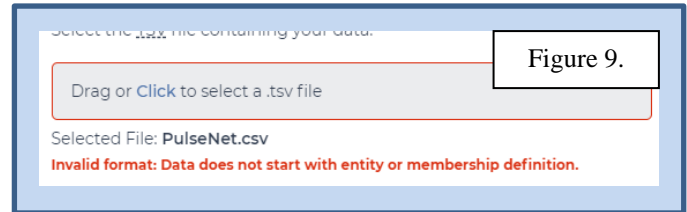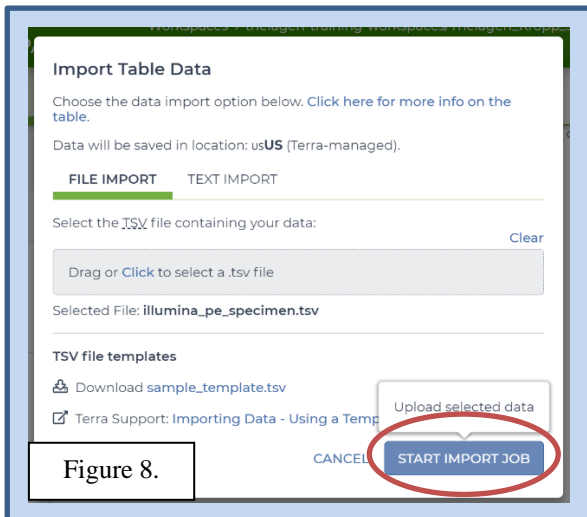
4.2   RUNNING THE BASESPACE WORKFLOW

1. In the Terra workspace where the BaseSpace sample sheet was imported, *open the BaseSpace_Fetch_P* workflow within the workflows tab (Fig 10)

   a. For instructions on importing the BaseSpace_Fetch workflow, see appendix 10.1

Figure 10.
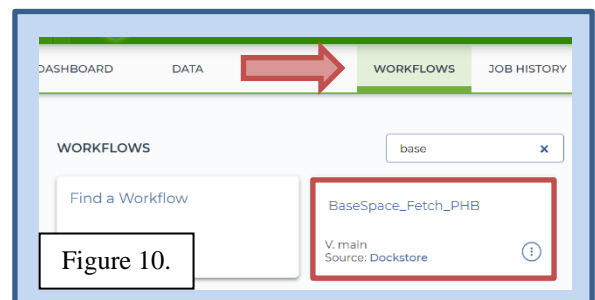
2. Choose the latest version of *v2* in the workflow version dropdown or the version used for internal validation (Fig 11, a)
3. Click the second bullet to *run workflow(s) with inputs defined by data table* (Fig 11, b)
4. Select the sample table to use under *select root entity type*, e.g. *PulseNet* (Fig 11, c)
5. Click *select data* (Fig 11, d)

Figure 11.
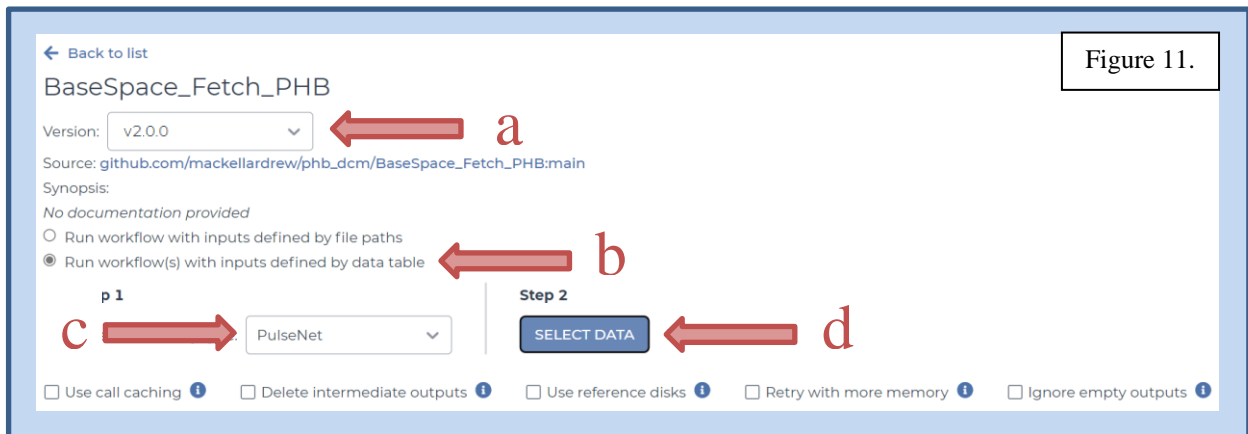
6. In the pop-up window, *choose the samples* for which reads should be imported, *modify the sample set name* if desired, and click *ok* (Fig 12)
   a. *Optional: to select all samples for sample sets greater than 100, click the down arrow and select all to include all samples in the table*
   b. *Optional: a name will be assigned to this set of samples chosen; for traceability it is useful to modify the job name to include the user's initials, e.g. BS_FetchPHB_YYYYMMDD[initials]*

Figure 12.

7. Set the first five input settings as follows (Fig 13):

    a. *workspace.access_token*

    b. *workspace.api_server*

    c. *this.basespace_collection_id*

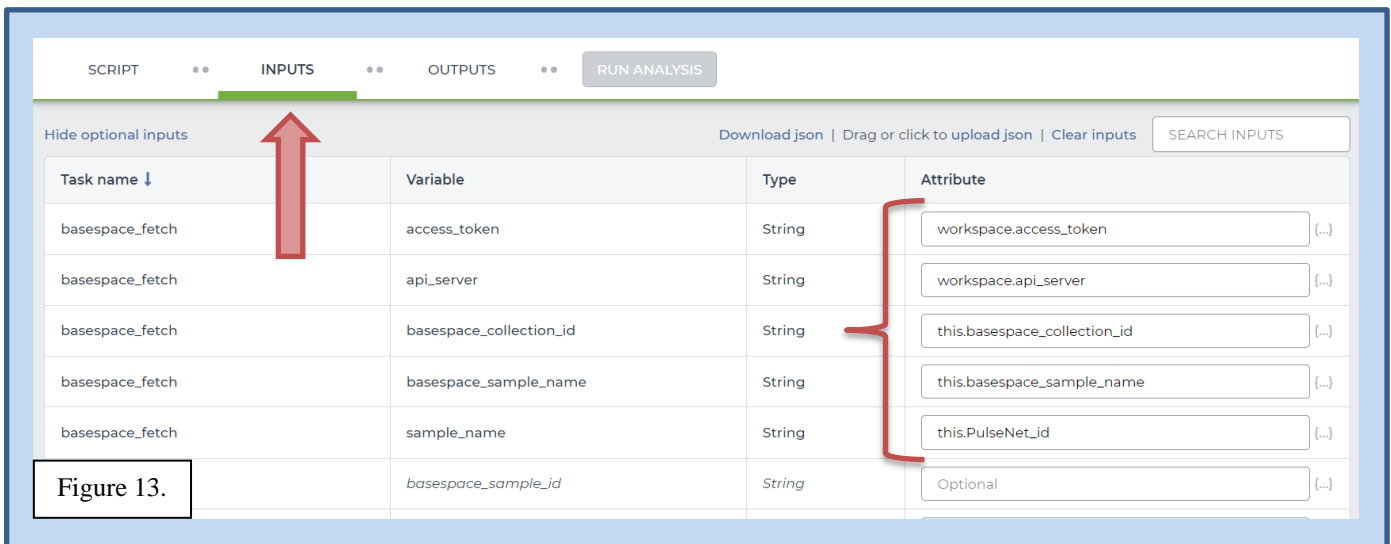   d. *this.basespace_sample_name*

    e. *this.PulseNet_id*

Figure 13.

8. In the outputs tab, click *use defaults* and *save* the workflow (Fig 14)

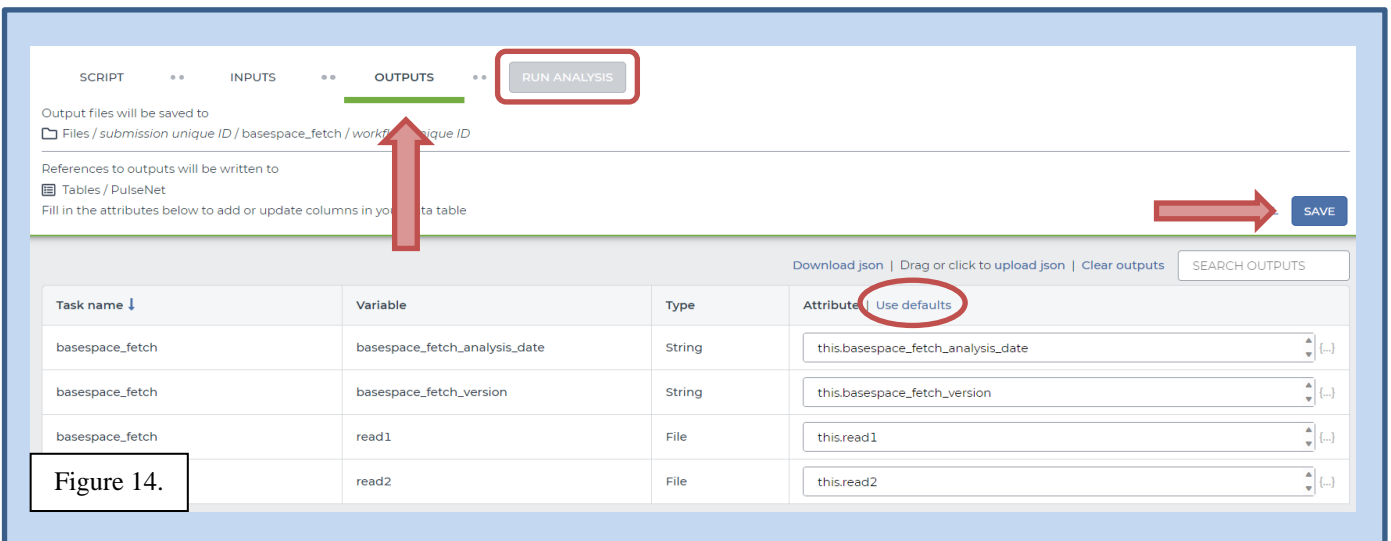9. Then click *run analysis*, enter any comments if desired, and click *launch*

Figure 14.

5.   QUALITY RECORDS

- Raw read files
- Metadata results table
- Workspace elements and files

6.   TROUBLESHOOTING

- If the first cell in the metadata table does not start with *entity:* and end with _*id*, an error message will prevent file import; adjust the metadata text in cell A1 and re-upload
- Metadata column headers and workflow input text indicated in gray in this SOP are customizable; black text is required
- Terra data table column headers become available as workflow inputs when running workflows; search for them in workflow input dropdowns using the prefix this. To fileter
- Filter for workspace data and files in workflow input dropdowns using the prefix workspace.
- If any workflow, input, or output settings are entered incorrectly, the analysis will not run as expected; verify all settings are correct and re-launch analysis
- If analysis fails, navigate to the job history in the workspace and click on the job submission for details; for help resolving run failures, email support@theiagen.com

7.   LIMITATIONS
None

8.   REFERENCES

1. Theiagen's BaseSpace Fetch Setup (initial setup)
2. Theiagen's BaseSpace_Fetch (workflow resource page)

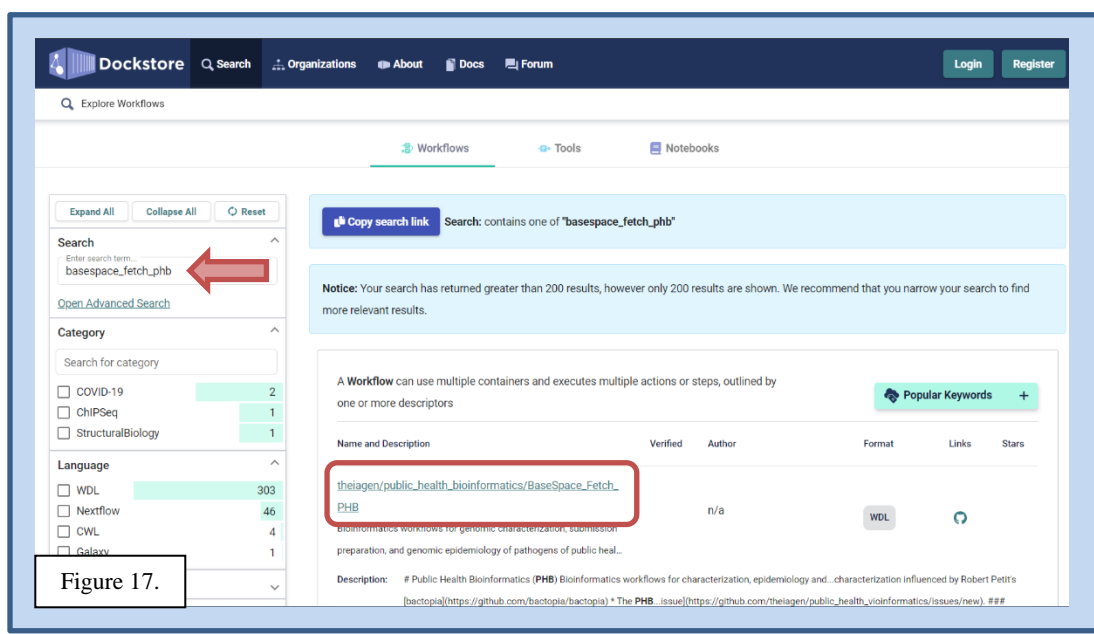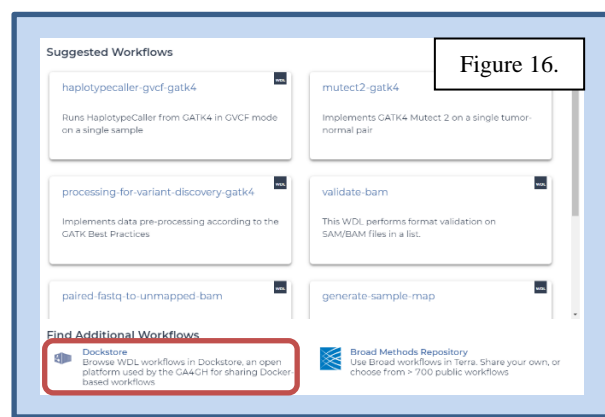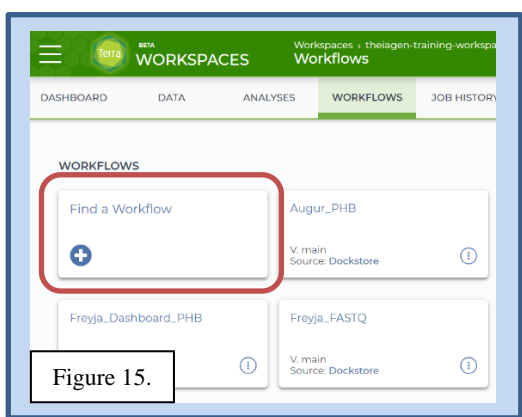## 9.   REVISION HISTORY

Revision	Version	Release Date
Document Creation	1	9/2023
Aligning with PHB v2.0.0 release; slight format and figure changes	2	4/2024

10.   APPENDICES

10.1   IMPORTING THE BASESPACE_FETCH_PHB WORKFLOW FROM DOCKSTORE

1. In the *Terra workspace* of interest, open the *workflows* tab and click *find a workflow* (Fig 15)
2. In the pop-up window, click *dockstore* (Fig 16)
3. To find the BaseSpace_Fetch_PHB workflow, type *basespace_fetch_phb* in the search bar (Fig 17)
    i. Looking at the workflow suffix, click on the *BaseSpace_Fetch_PHB* workflow (Fig 17)
        1. **Note**: *BaseSpace_Fetch is an older workflow iteration and not recommended*

Figure 15.

Figure 16.

Figure 17.

ii. Click *Terra* to launch the workflow in Terra (Fig 18)
iii. Choose the *destination workspace* in the dropdown and click import or create a new workspace (Fig 19)
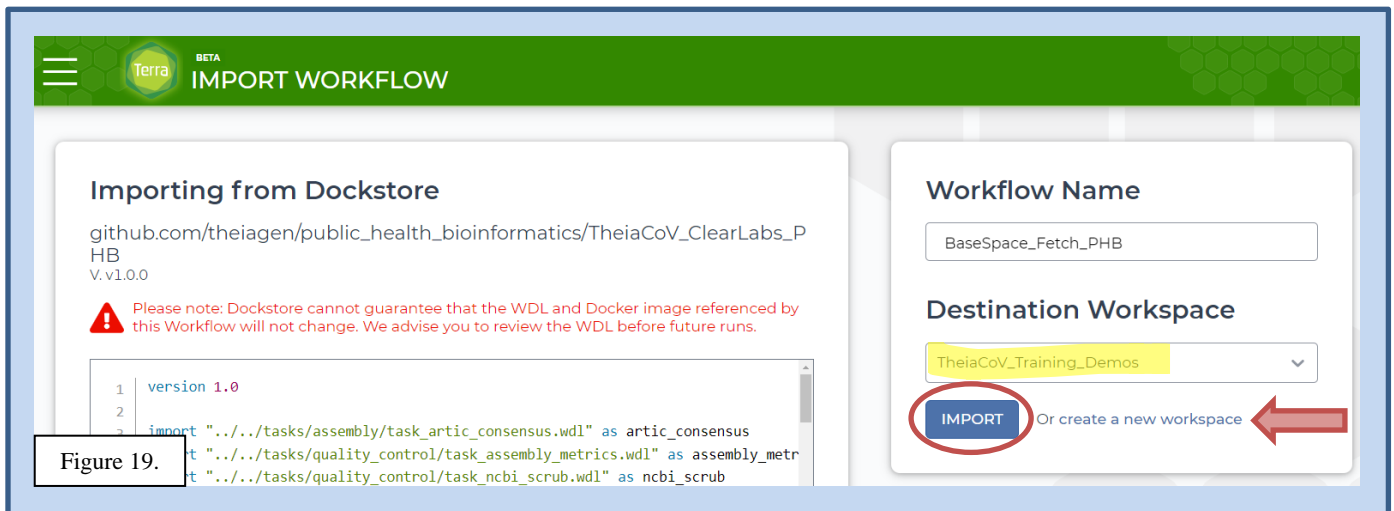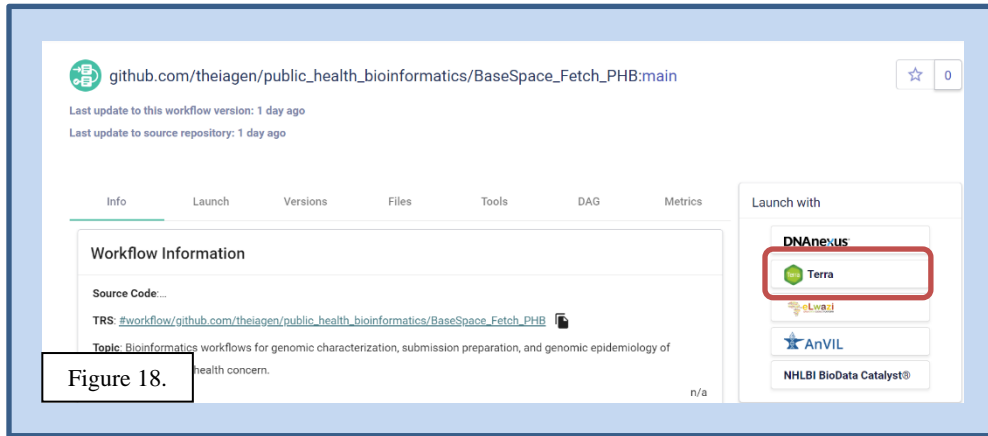
Figure 18.

Figure 19.

10.2 OPEN COMMAND LINE TERMINAL TO PROVIDE BASESPACE CREDENTIALS TO TERRA

*This setup is up only required once per BaseSpace account. Additionally, the access token and api server keys may be copied into multiple workspaces without having to re-run command line

Figure 20.

Figure 21.
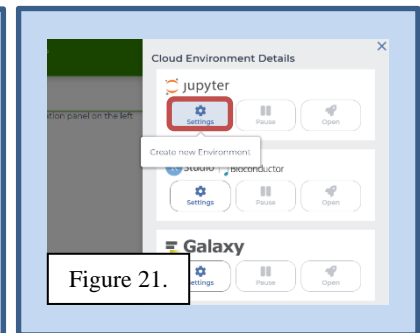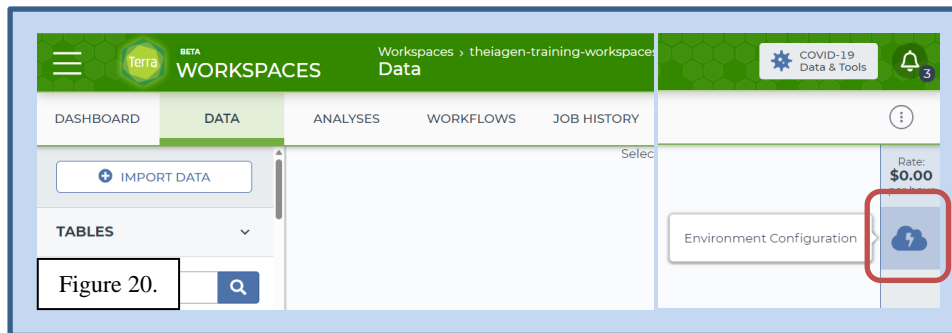
1. In the workspace to be used to import BaseSpace reads, open a cloud environment in Terra by clicking on the *enivironment configuration* (Fig 20)
2. Click on the *Jupyter settings* button (Fig 21)
3. *Set environment configurations* as follows (Fig 22):
   a. Application configuration: *Default: (GATK 4.3.3.0, Python 3.10.11, R 4.3.1)*
      i. *The version numbers may vary slightly as updates are incorporated into the default application configuration*
   b. Cloud compute profile: CPUs *2*, Memory *7.5*
   c. Persistent disk: Disk Type *Standard*, Disk Size (GB) *250*
   d. Click *create*
4. *Click on the open terminal button* located in the right sidebar (Fig 23)
5. *A message will appear that says "Creating cloud environment. You can navigate away and return in 3-5 minutes."*

Figure 23.

Figure 22.

6. Test to see if the terminal is working by typing *ls* and *hit enter*

   a. *Verify the output* looks like the following:

   ```
   jupyter@2a9974687b45:~$ ls
   gitignore_global  jupyter.log  lost+found  packages
   ```

   b. *If everything looks okay, proceed with the command line; if not, try closing and re-opening the terminal and re-testing the* ls *command*

7. Enter the following to create a bin: *mkdir ~/bin* and *hit enter*

   a. *Ctrl + c and Ctrl + v do not work in this terminal, but right clicking to copy/paste should work*

8. Type the command *ls* and *hit enter* to verify a new bin has been created; *you should see "bin" in addition to the following:*

   ```
   jupyter@2a9974687b45:~$ ls
   bin  gitignore_global  jupyter.log  lost+found  packages
   ```

9. Enter the command: *wget "https://launch.basespace.illumina.com/CLI/latest/amd64-linux/bs" -O $HOME/bin/bs*

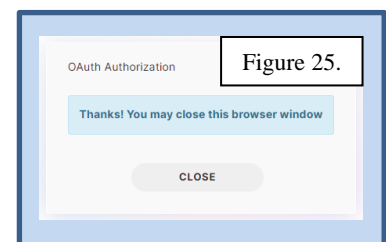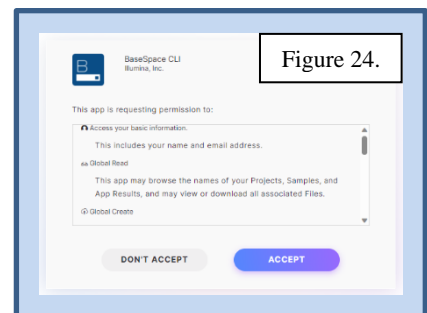   a. *You should get the following returned:*

   ```
   --2023-09-14 22:42:27--  https://launch.basespace.illumina.com/CLI/latest/amd64-linux/bs
   Resolving launch.basespace.illumina.com (launch.basespace.illumina.com)... 18.160.200.113, 18.160.200.107,
    18.160.200.47, ...
   Connecting to launch.basespace.illumina.com (launch.basespace.illumina.com)|18.160.200.113|:443... connect
   ed.
   HTTP request sent, awaiting response... 200 OK
   Length: 12566528 (12M) [binary/octet-stream]
   Saving to: '/home/jupyter/bin/bs'

   /home/jupyter/bin/bs       100%[====================================>]  11.98M  --.-KB/s   in 0.1s

   2023-09-14 22:42:28 (112 MB/s) - '/home/jupyter/bin/bs' saved [12566528/12566528]
   ```

10. Enter the command: *chmod u+x $HOME/bin/bs*

    a. *Nothing will be returned for this command*

11. Enter the command: *source ~/.profile*

    a. *The following will be returned:*

    ```
    bash: /home/jupyter/.profile: No such file or directory
    ```

12. Enter the command: *~/bin/bs auth*

    a. *The following will be output:*

    ```
    Please go to this URL to authenticate:  https://basespace.illumina.com/oauth/device?code=Uj3D3
    ```

    b. *Click the link and login* to the BaseSpace account to link to Terra

       i. *If the BaseSpace account is already logged in, the login process may be skipped*

    c. *Accept the permission request* from Terra in BaseSpace (Fig 24)

13. *An authorization message will be received from BaseSpace (Fig 25)*

Figure 24.

Figure 25.

14. *Return to the terminal* window; *a welcome message should be returned for the authorized user similar to the following:* `Welcome, Kelsey Kropp`

15. *Verify the basespace access* has been added by entering the command *ls -a*

    a. *The following should be returned – notably, the .basespace bin*

    ```
    .           bin       .entrypoint.out   .ipython     .local      .welder.log
    ..          .cache    .gitconfig        .jupyter     lost+found  .wget-hsts
    .basespace  .config   gitignore_global  jupyter.log  packages
    ```

16. Enter the following command: *cat ~/.basespace/default.cfg*

    a. *The following should be returned:*
    ```
    apiServer   = https://api.basespace.illumina.com
    accessToken = b2aac9f782404625ace2c6ca83e43187
    ```

17. *Copy the apiServer key* https://api.basespace.illumina.com; *copy by highlighting, right click, copy*
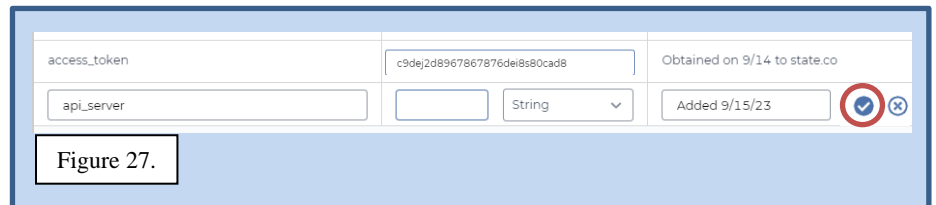
18. *Open a new Terra window* for the relevant workspace

19. Scroll to the bottom of the left sidebar and *click on workspace data* (Fig 26)
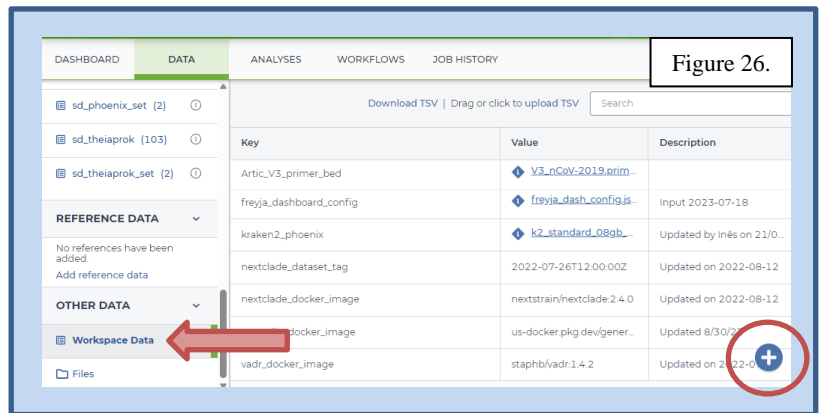
20. Add a workspace data key by *clicking the blue plus symbol* (Fig 26); specify the following (Fig 27):

    a. Key: *api_server*

    b. Value: *paste the apiServer key* obtained above in step 17

    c. Description: <u>optional</u>

Figure 26.

Figure 27.

21. *Click the blue checkmark* to save (Fig 27)

22. In the terminal window *copy the accessToken* value; don't include spaces

23. Return to the Terra workspace, re-open the workspace data tab, and *add a new workspace data key* by clicking the blue plus symbol (Fig 26); specify the following (Fig 27), then *click the blue checkmark to save* (Fig 27):

    a. Key: *access_token*

    b. Value: *paste the accessToken* value copied from the terminal window obtained in step 21

    c. Description: <u>optional</u>

24. Continue with section 4.1 to download and import the BaseSpace sample sheet and section 4.2 to run the BaseSpace_Fetch_PHB workflow to import sample reads