



Workflows in AnVIL

What are workflows?

For computational biology tasks, workflows help organize a set of connected instructions to process large datasets at once. For example, a simple workflow could involve quality control of raw sequence data followed by alignment using existing bioinformatics tools.

Consists of

Data and Resources

Input

Example

Exome Sequence Data

WORKFLOWS

General, Automated

Processing

Mapping and Variant Discovery

Specific, Manual

Analysis

Effective Prediction and Variant Load Analysis

Materials and Methods

Sharing

Preprint in *bioRxiv*

Workflow components



Terra User Interface



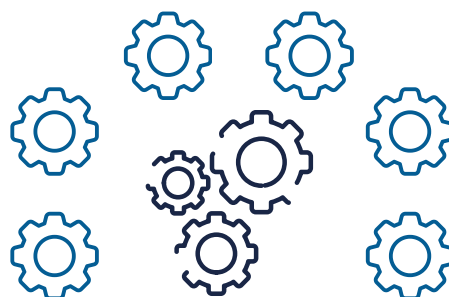
Google Compute Engine runs jobs within a container



WDL script + Data input(s)



Researcher launches workflow



Cromwell execution engine creates individual jobs to run workflow

Workflows In AnVIL

Write your own WDL



Find a *docker image* that has the tools you need



Write your WDL workflow. Specify the container's image in the WDL:

```
runtime {  
  docker: 'ubuntu:latest'  
}
```



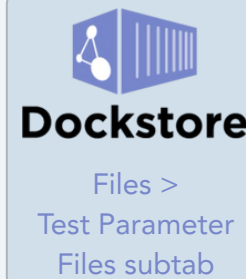
Specify workflow Inputs

Images can be used interactively or in a WDL!

```
runtime{  
  docker: "my_project/do_stuff: 1.2.0"  
}
```

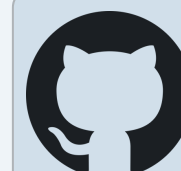
Import workflows

Import a publicly available workflow



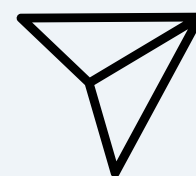
On *Dockstore*, you can find a test parameter file (JSON) which you can then upload in Terra after Importing your workflow

Use a pre-configured workflow in the Terra Repository and import it to a workspace within AnVIL



When exporting from outside sources, note that exporting workflows from these sources is a two-step process!

1. Copy the workflow's script to a new method in the Terra Repository or a Dockstore repository
2. Export the workflow to an AnVIL-powered-by-Terra workspace.



Submit workflow on Terra



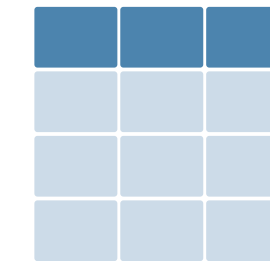
Monitor workflow and outputs



Visualize and analyze outputs using *RStudio* or *Jupyter* notebook

Configure Inputs

Workflow inputs can be specified via



Data Tables within the workspace



A JSON file:

Upload your own or use a pre-configured one on Terra

Containers for reproducibility

Find existing images containing bioinformatics tools from public registries (e.g., DockerHub, Quay)

OR

Build your own and push to DockerHub or another registry

Pro tip: Use Github Actions to build and push Docker images! Find more on this [Github article](#).

Public registries with Images:



Google Cloud Platform (GCP)



Additional Docker information

- A Docker **file** defines the container's dependencies, environment variables, file system, and applications.
- A Docker **image** builds and runs a container, which contains everything defined in the Docker file.

For more information, refer to [this AnVIL resource](#).