

# Dockstore:

## Enhancing a community platform for sharing cloud-agnostic research tools

Louise Cabansay, UC Santa Cruz Genomics Institute

# What is Dockstore?

Dockstore is a free and open source platform for sharing scientific tools and workflows. It is a registry of Docker-based resources described using popular workflow languages CWL, WDL, and Nextflow.

- **Portability**
  - Run workflows in any environment that supports Docker
- **Interoperability**
  - Standardize computational analysis through GA4GH APIs
- **Reproducibility**
  - Create, Share, Use
  - Containers + Popular descriptor languages

Search Docker Tools and Workflows for the Sciences:

Enter Keyword...

*Dockstore, developed by the Cancer Genome Collaboratory, is an open platform used by the GA4GH for sharing Docker-based tools described with the Common Workflow Language (CWL), the Workflow Description Language (WDL), or Nextflow (NFL)*

VIDEO OVERVIEW

Sign up to Contribute >

Quick Start >

News and Events >

Discuss >

Browse Tools

Browse Workflows

A tool is a docker container with an associated descriptor describing how to run it.

Name	Author	Format	Project Links	Stars ↓
✓ pancancer/pcawg-dkzf-workflow	Brian O'Connor	CWL	GitHub Quay.io	4★
✓ pancancer/pcawg-sanger-cgp-workflow	Keiran Raine	CWL_WDL	GitHub Quay.io	3★
wtsicgp/dockstore-biobam2/bamtofastq	Keiran Raine	CWL	GitHub Quay.io	2★

Tweets by @DockstoreOrg

Dockstore Retweeted

Mike Lin @DNAm1n  
minivdl v0.3.0 is the first version with independent capability to run #CWLWDL workflows on the local host. Early testing, eager for community to try it out & report interop problems. @WDL\_dev @DockstoreOrg @GA4GH @czscience github.com/chanzuckerberg...

chanzuckerberg/ml...  
A static analysis tool...  
github.com

Jul 19, 2019

Now on version 1.6.0, first presented version 1.25 at BOSC2017

# What is Dockstore?

<https://dockstore.org/>

dockstore / dockstore

Code Issues 460 Pull requests 4 Projects 0 Wiki Security Insights

Our VM/Docker sharing infrastructure and management component <https://dockstore.org/>

dockstore docker cwl wdl containers bioinformatics nextflow workflow

2,158 commits 24 branches 178 releases 1 environment 22 contributors Apache-2.0

Branch: develop New pull request Create new file Upload files Find File Clone or download

Repository	Description	Latest commit
circleci	Feature/2374/token resource it (#2581)	15 days ago
github	Feature/dockstore classpath (#342)	3 years ago
THIRD-PARTY-NOTICES	Standardize model boilerplate (#1356)	last year
dockstore-client	[maven-release-plugin] prepare for next development iteration	2 days ago
dockstore-common	[maven-release-plugin] prepare for next development iteration	2 days ago
dockstore-event-consumer	[maven-release-plugin] prepare for next development iteration	2 days ago
dockstore-file-plugin-parent	[maven-release-plugin] prepare for next development iteration	2 days ago
dockstore-integration-testing	[maven-release-plugin] prepare for next development iteration	2 days ago
dockstore-language-plugin-parent	[maven-release-plugin] prepare for next development iteration	2 days ago
dockstore-webservice	Feature/2498/track tos acceptance (#2626)	5 hours ago
openapi-java-client	[maven-release-plugin] prepare for next development iteration	2 days ago
openapi-java-wes-client	[maven-release-plugin] prepare for next development iteration	2 days ago
reports	[maven-release-plugin] prepare for next development iteration	2 days ago

## Dockstore API <sup>1.6.0</sup>

Base URL: [dockstore.org/api/](https://dockstore.org/api/)  
<https://dockstore.org/swagger.json>

This describes the dockstore API, a webservice that manages pairs of Docker images and associated metadata such as CWL documents and Dockerfiles used to build those images

Terms of service  
Dockstore@ga4gh - Website  
Send email to Dockstore@ga4gh  
Apache License Version 2.0  
Dockstore documentation

Schemes: HTTPS Authorize

Filter by tag

### entries

Interact with entries in Dockstore regardless of whether they are containers or workflows

- PUT `/entries/{id}/aliases` Update the aliases linked to a entry in Dockstore.
- GET `/entries/{id}/collections` Get the collections and organizations that contain the published entry

### containers

List and register entries in the dockstore (pairs of images + metadata (CWL and Dockerfile))

- GET `/containers/{containerId}` Retrieve a tool.
- PUT `/containers/{containerId}` Update the tool with the given tool.
- DELETE `/containers/{containerId}` Delete a tool.
- GET `/containers/{containerId}/cwl` Get the primary CWL descriptor file on Github.
- GET `/containers/{containerId}/cwl/{relative-path}` Get the corresponding CWL descriptor file on Github.

Github: <https://github.com/dockstore/dockstore>

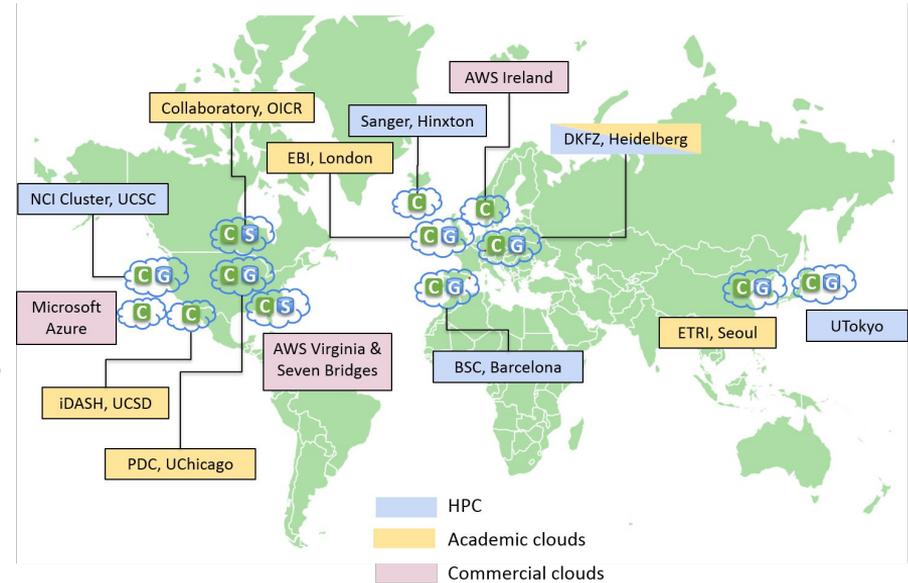
Swagger: <https://dockstore.org/api/static/swagger-ui/index.html#>

# Motivations for Dockstore:

## The Pan Cancer Analysis of Whole Genomes (PCAWG)

### Cloud-Based, Distributed Collaboration

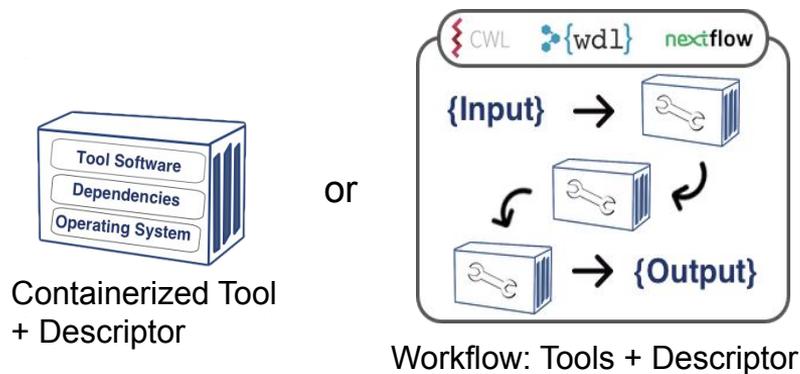
- International Cancer Genome Consortium (ICGC)
- ~5,800 Whole Genomes
  - ~2,800 Cancer Donors
- 8 sites storing and sharing data via GNOS
  - 300TB -> 900TB
- 14 Cloud (and HPC) environments
  - 3 Commercial, 7 OpenStack, 4 HPC
  - ~630 VMs, ~15K cores, ~60TB of RAM



# GA4GH Tool Registry Service (TRS) API

The Global Alliance for Genomics and Health (GA4GH) Tool Registry API standard for listing and describing available tools for exchange, indexing, and searching.

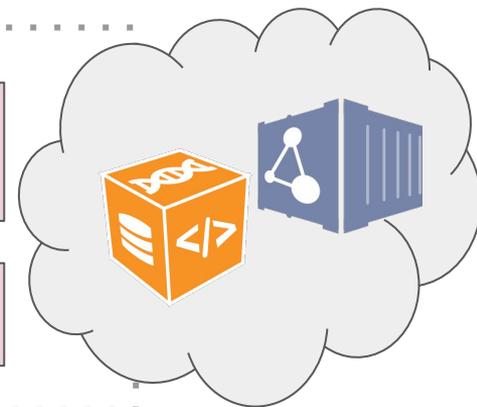
- Stand alone containerized tools
- Workflows with multiple tools wrapped in descriptor languages (Common Workflow Language, Workflow Descriptor Language, Nextflow)



## Sharing API

GET /api/ga4gh/v2/tools  
/{id}/versions/{version\_id}

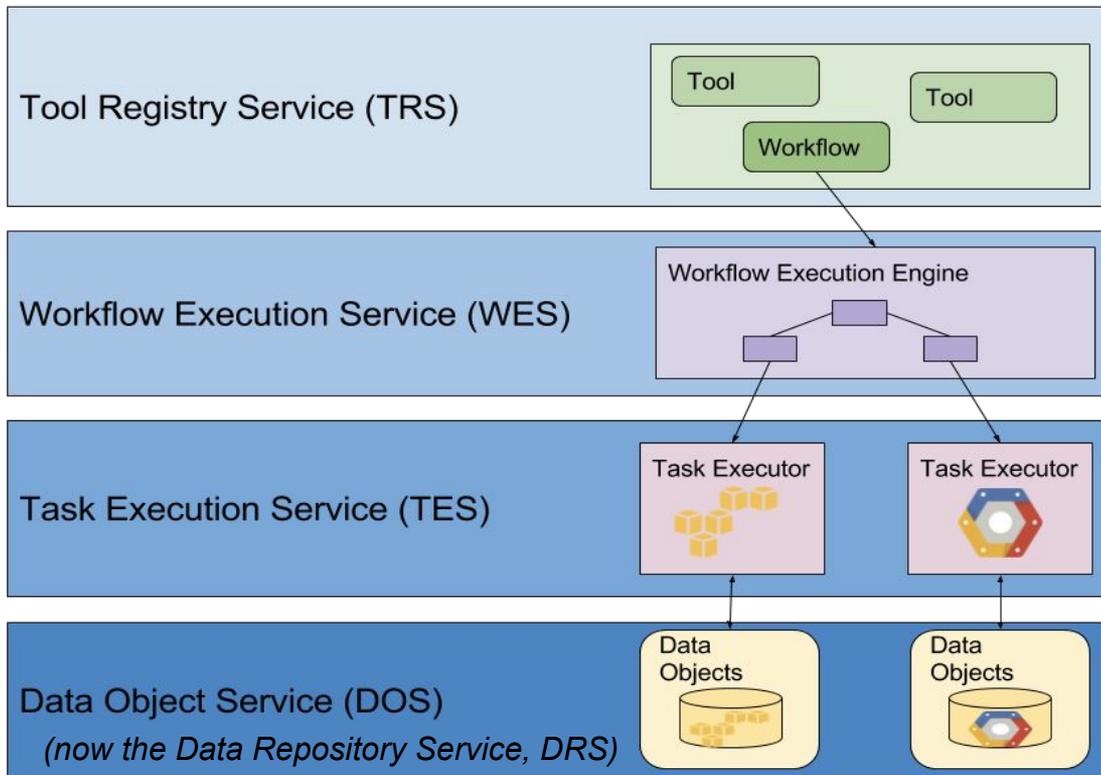
GET /api/ga4gh/v2/tools



- **GitHub page:** <https://github.com/ga4gh/tool-registry-service-schemas>
- **Latest draft release:** 1.0.0, working on 2.0.0 now



# GA4GH Cloud Work Stream APIs



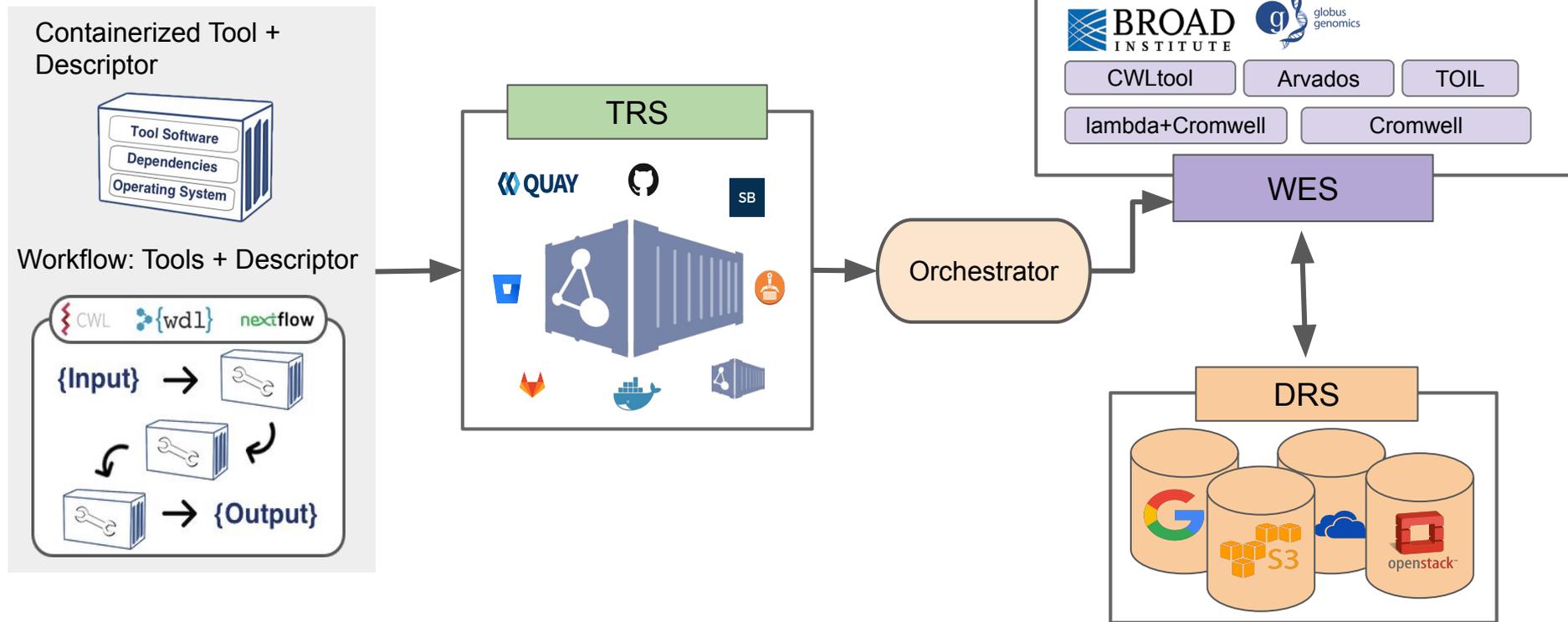
**Sharing Tools  
and Workflows**

**Executing  
Workflows**

**Executing Tasks**

**Accessing Data**

# Dockstore & GA4GH Cloud Workstream



# Dockstore, going beyond TRS

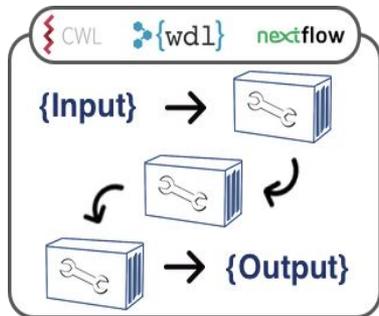
- GA4GH - TRS API
  - Standardized subset of Dockstore and Biocontainers
  - We provide a [validator](#) to help those implementing TRS
- Dockstore API
  - Extends TRS with functionality specific to Dockstore
    - Catalog that brings together multiple external repositories
    - Authenticated access
    - Advanced language support
    - Launch analysis
    - Community and collaboration oriented features
      - Organizations, starring, platform verification

# Containers, Descriptors, and Parameter Files

**Tool:** Container + Descriptor



**Workflow:** Tool(s) + Descriptor



**Descriptor:**

The workflow language used to describe how to run your tool or workflow/pipeline.

- Which tools/containers
- What steps and when
- Where I/O parameters
- How to allocate compute requirements
- Can also include metadata information



**Parameter File** (wdl, cwl):

- Specifies the actual input/output files and where to find them (local, ftp, http, or cloud)
- JSON, YAML

```
{
  "bam_input": {
    "class": "File",
    "format": "http://edamontology.org/format_2572",
    "path": "/tmp/NA12878.chrom20.ILLUMINA.bwa.CEU.low_coverage.20121211.bam"
  },
  "bamstats_report": {
    "class": "File",
    "path": "/tmp/bamstats_report.zip"
  }
}
```

# Dockstore as a Registry

- Searchable and Centralized Catalogue
- Many ways to get your tool or workflow into Dockstore!
  - External Hosting: Github, Bitbucket, Quay.io, Docker Hub
  - Direct Hosting (currently just workflows)



## GitHub Account Login Method

GitHub credentials are used for login purposes as well as for pulling source code from GitHub  
One of GitHub or Google is required.

[Link Account](#)

## Google Account Login Method

Google credentials are used for login purposes and integration with FireCloud.  
One of GitHub or Google is required.

[Link Account](#)

## Quay Account

Quay.io credentials are used for pulling information about Docker images and automated builds.

[Link Account](#)

## Bitbucket Account

Bitbucket credentials are used for pulling source code from Bitbucket.

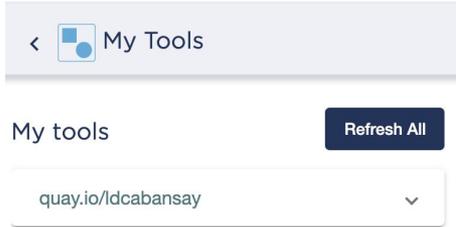
[Link Account](#)

## GitLab Account

GitLab credentials are used for pulling source code from GitLab.

[Link Account](#)

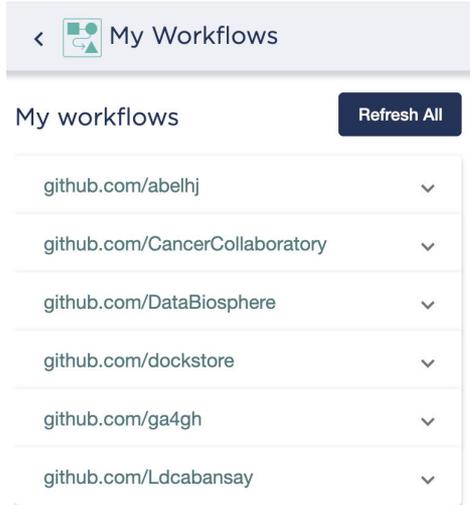
# Ways to Register to Dockstore



< My Tools

My tools Refresh All

quay.io/lcabansay



< My Workflows

My workflows Refresh All

- github.com/abelhj
- github.com/CancerCollaboratory
- github.com/DataBiosphere
- github.com/dockstore
- github.com/ga4gh
- github.com/Ldcabansay

Through linked account  
via supported public  
APIs

## Register Workflow

- 1 Workflow storage type
- 2 Create a workflow

You can write your CWL/WDL and store it on Dockstore.org, or you can register a CWL/WDL/NextFlow from an existing third party repository, such as GitHub, Bitbucket, etc.

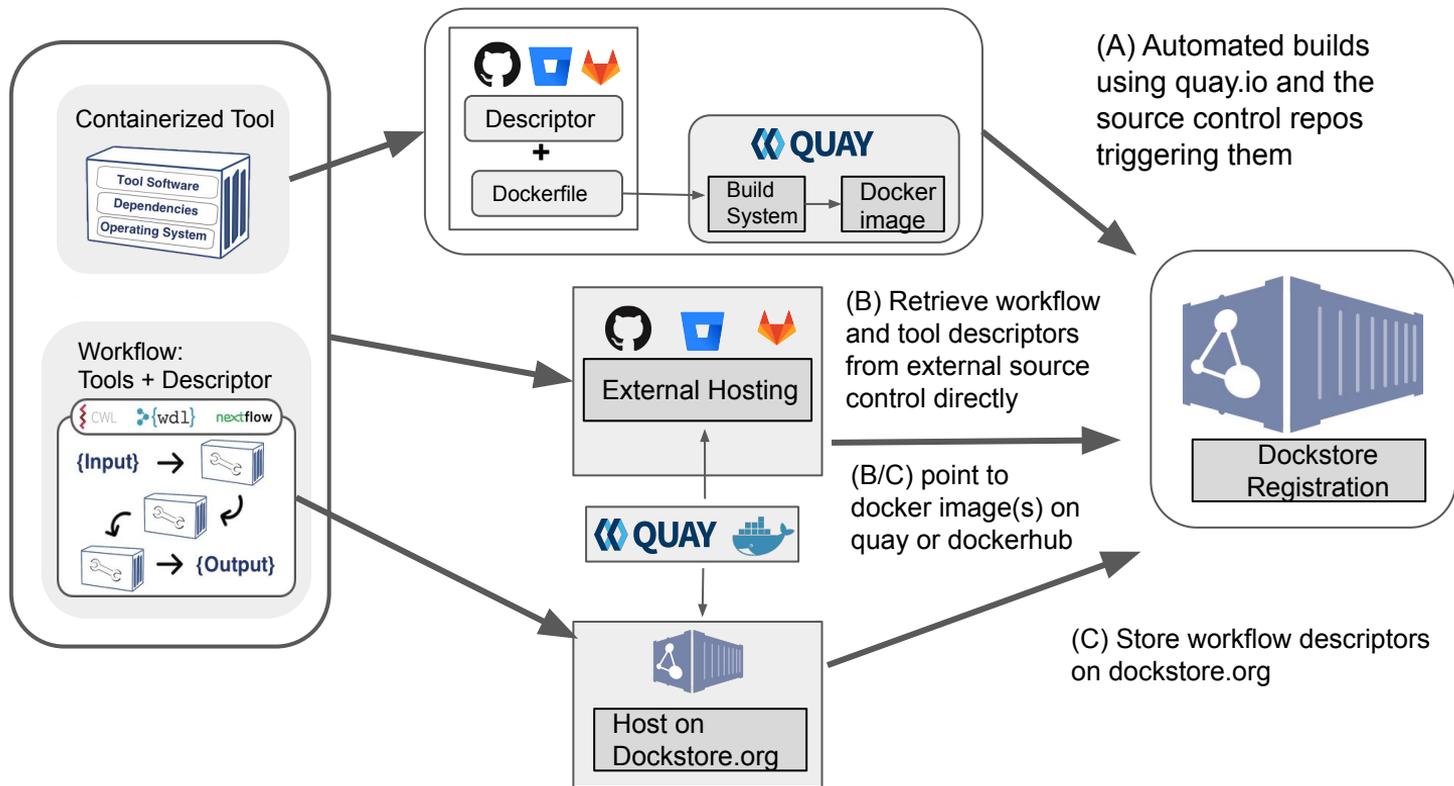
- Use CWL, WDL or Nextflow from GitHub, Bitbucket, etc.
- Create and save CWL, WDL, or Nextflow on Dockstore.org

Close

Next

OR through manual registration  
pointing to external source repo or  
directly on Dockstore.org

# Ways to Register to Dockstore

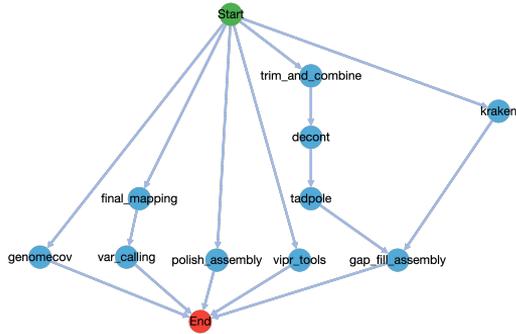


# Containers and Descriptor Languages

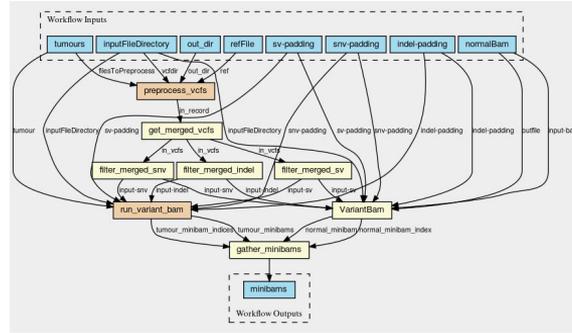
- Containers
  - Docker
- Descriptor Language Support
  - CWL, WDL, and now Nextflow!
    - Syntax Highlighting, Validation
    - Visualization
      - Including additional visualization options from [view.commonwl.org](http://view.commonwl.org) and EPAM WDL Viewer
    - Metadata parsing
      - Authorship, contact info, description
      - Markdown syntax (CWL, WDL)
      - I/O Filetypes (CWL)



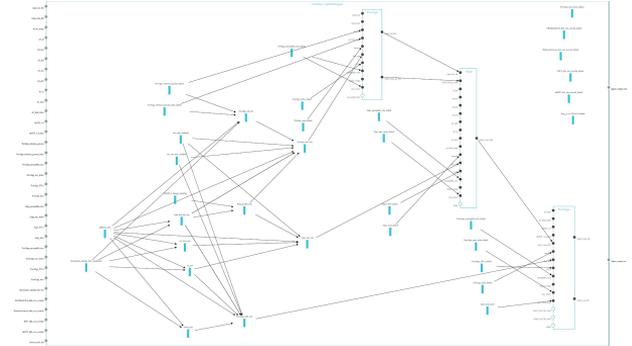
# Visualizations



Built-in for CWL, WDL,  
and Nextflow (pictured)



[view.commonwl.org](http://view.commonwl.org) for  
CWL



EPAM Pipeline Builder for WDL

Search Organizations Docs Login/Register

Available Workflows

github.com/DataBiosphere/topmed-workflows/UM\_aligner\_wdl:1.32.0

Last Modified: 19 days ago

datacommons datstage topmed u-of-michigan

Info Launch Versions Files Tools

Descriptor Files Test Parameter Files

u\_of\_michigan\_aligner.wdl

```

256 echo "Running pre-alignment"
257
258 samtools view -T ${ref_fasta} -uh -F 0x900 ${input_cram} \
259 | bam-ext-mem-sort-manager squeeze --in -.ubam --keepDups --rmTags AS:i,
260 | samtools sort -l 1 -@ 1 -n -T ${pre_output_base}.samtools_sort_tmp - \
261 | samtools fixmate - - \
262 | bam-ext-mem-sort-manager bam2fastq --in -.bam --outBase ${pre_output_b
263
264 }
265
266 output {
267   File output_list_file = "${pre_output_base}.list"
268   # Capture all the files mentioned in the pre_output_base.list file
269   # So they will be present for the Align task
270   Array[File] output_fastq_gz_files = glob("${pre_output_base}.*")
271 }
272
273 runtime {
274   maxRetries: max_retries
275   preemptible: preemptible_tries
276   #memory: "6.5 GB"
277   memory: sub(memory, "\\..*", "") + " GB"
278   cpu: sub(CPU, "\\..*", "")
279   disks: "local-disk " + sub(disk_size, "\\..*", "") + " HDD"
280   zones: "us-central1-a us-central1-b us-east1-d us-central1-c us-central1-f"
281   docker: docker_image
282 }
283
284 task Align {
285   File input_list_file
286   Array[File] input_fastq_gz_files
287
288   File ref_alt
289   File ref_bwt
290   File ref_pac
291   File ref_peg

```

Recent Versions

1.32.0 Mar 11, 2019

✓ 1.29.0 Sep 27, 2018

feature/freeze8-wdl Jun 24, 2019

master Mar 11, 2019

develop Mar 11, 2019

feature/variant-discovery-task Nov 11, 2018

See all versions

Source Repositories

GitHub

Collections

TOPMed Alignment Pipeline

Launch with

DNAnexus

FireCloud

DNAnexus

Terra

Sharing

f t in

# Launching Analysis

- GA4GH TRS-WES Orchestrator
  - FireCloud
  - Terra
- Partner Platforms
  - DNAnexus
  - DNAnexus
  - Terra
  - Local Execution via Dockstore CLI

Launch with

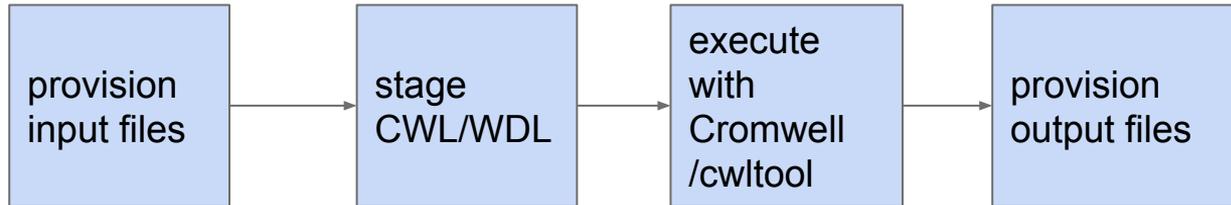


# Dockstore CLI

A handy command line resource for developers to play with tools/workflows locally before running them on a cloud

- DRS file provisioning
- WES beta support
- Locally
  - Cromwell
  - Cwltool

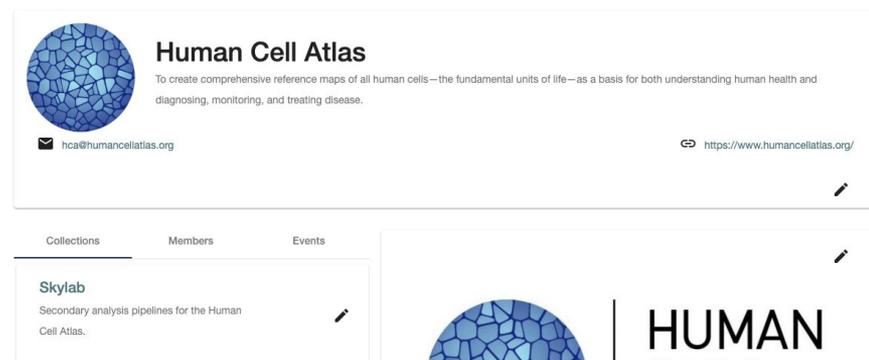
## Execution with the Dockstore Command Line Interface (CLI)



*Simple Dockstore Command Line*

# Organizations and Collections

- Organizations
  - A place for groups, labs, consortiums, etc to showcase their projects, collaborate, and group sets of tools/workflows into ‘collections’
  - Markdown descriptions
  - Membership roles
- Collections
  - Playlist of workflows or tools highlighted by an Organization
  - Markdown descriptions

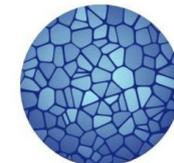


**Human Cell Atlas**  
To create comprehensive reference maps of all human cells—the fundamental units of life—as a basis for both understanding human health and diagnosing, monitoring, and treating disease.  
✉ [hca@humancellatlas.org](mailto:hca@humancellatlas.org) <https://www.humancellatlas.org/>

Collections Members Events

**Skylab**  
Secondary analysis pipelines for the Human Cell Atlas.

+ Create collection



## HUMAN CELL ATLAS

### ABOUT HUMAN CELL ATLAS

In London on 13 and 14 October, 2016, a collaborative community of world-leading scientists met and discussed how to build a Human Cell Atlas—a collection of maps that will describe and define the cellular basis of health and disease.

Cells are the most fundamental unit of life, yet we know surprisingly little about them. They vary enormously within the body, and express different sets of genes. Without maps of different cell types and where they are located in the body, we cannot describe all their functions and understand the biological networks that direct their activities.

A complete Human Cell Atlas would give us a unique ID card for each cell type, a three-dimensional map of how cell types work together to form tissues, knowledge of how all body systems are connected, and insights into how changes in the map underlie health and disease. It would allow us to identify which genes associated with disease are active in our bodies and where, and analyze the regulatory mechanisms that govern the production of different cell types.

# Usability Improvements

- Improved Published Tool/Workflow Interface
  - Labels - improve searchability
  - Metadata and markdown display
    - Authorship, contact info, descriptions
    - Readme like descriptions (CWL, WDL)
  - Versioning: display all tags/versions, set defaults, check validity, verify compatible platforms
  - Link to checker workflow

github.com/DataBiosphere/topmed-workflows/UM\_variant\_caller\_wdl:1.32.0

Last Modified: 21 days ago

datacommons datastage topmed u-of-michigan

Git Reference	Date Modified	Valid	Verified Platforms	Actions
feature/freeze8-wdl	Jun 24, 2019, 9:12:14 AM	✓	✗	View
<b>Default</b> 1.32.0	Mar 11, 2019, 4:10:28 PM	✓	✗	View
pmaster	Mar 11, 2019, 4:10:28 PM	✓	✗	View
pdevelop	Mar 11, 2019, 3:28:36 PM	✓	✗	View
feature/variant-discovery-task	Nov 11, 2018, 8:47:04 PM	✓	✗	View
feature/scatter-merge	Nov 11, 2018, 2:09:25 PM	✗	✗	View
1.31.0	Sep 30, 2018, 5:35:11 PM	✓	✗	View
1.30.0	Sep 28, 2018, 5:06:04 PM	✓	✗	View
1.29.0	Sep 27, 2018, 9:46:10 PM	✓	✓ Cromwell	View
1.28.0	Sep 18, 2018, 4:39:53 PM	✓	✗	View
feature/scatter-discovery	Sep 13, 2018, 3:30:44 PM	✓	✗	View

delly dkfz embl nci pcawg sevenbridges

Info Launch Versions Files Tools

### Workflow Information

**GitHub:** sbg/sbg\_dockstore\_tools:v1.0.1  
**TRS:** #workflow/github.com/sbg/sbg\_dockstore\_tools/icgc\_pcawg\_dkfz\_embl\_sbg\_modified  
**Workflow Path:** /pcawg/vc/icgc\_pcawg\_dkfz\_embl\_workflow.cwl  
**Test File Path:**  
**Checker Workflow:** n/a  
**Descriptor Type:** CWL

### Workflow Version Information

v1.0.1

**Author:** Seven Bridges  
**E-mail:** support@sbgenomics.com  
[Export as ZIP](#)

**Description:**

The DKFZ/EMBL workflow from the ICGP PanCancer Analysis of Whole Genomes (PCAWG) project.



This workflow is a combination of the **ICGP PCAWG Seqware Delly Workflow** (EMBL workflow) for structural variation and the **ICGP PCAWG DKFZ Workflow** workflow for SNVs, indels, and copy number variants.

For more information see the PCAWG project [page](#) and the GitHub [page](#) for the code.

*A list of all inputs and parameters with corresponding descriptions can be found at the bottom of the page.*

### Common Use Cases

This tool is used to call variants from Tumor/Normal pair of BAM files, previously generated by using the **ICGP-PCAWG-Seqware-BWA-Workflow** tool available in **Public Apps**, in the same manner as the VCF files in the ICGP PCAWG dataset. To do this, set:

- Tumor/Normal pair in BAM format, provided via **Input BAM Files** port.

# Usability Improvements

- **Search Interface**
  - Search name, author, organization, descriptor, etc
  - Filter results using facets
    - Descriptor Language
    - Labels

Advanced Search

Find tools and workflows with **files** that have ...

**All these words:**

**The exact word or phrase:**

**Any of these words:**

**None of these words:**

Expand All

Search

Enter search term  
Topmed

Entry Type

workflow (7)

Language

CWL (4)  
 WDL (3)

Author

Wait Shands (2)  
 Ruchi Munshi (1)  
 Seven Bridges (1)  
 Vladimir Obucina (1)  
 Yilin Xu (1)

Workflow: Source Control

github.com (7)

Workflow: Organization

DataBiosphere (7)

Search: contains one of "Topmed"

Tag Cloud

A workflow is a series of tools strung together, with an associated descriptor describing how to run it.

Name	Author	Format	Project Links	Stars
✓ DataBiosphere/topmed-workflows/UM_variant_caller_wd	Wait Shands	WDL	GitHub	6★
✓ DataBiosphere/topmed-workflows/UM_aligner_wdl	Wait Shands	WDL	GitHub	2★
DataBiosphere/topmed-workflows/CCDG_aligner_functional	Ruchi Munshi	WDL	GitHub	2★
✓ DataBiosphere/topmed-workflows/UM_aligner_cwl	Seven Bridges	CWL	GitHub	1★
DataBiosphere/topmed-workflows/gatk-vcf-comparator	n/a	CWL	GitHub	
DataBiosphere/topmed-workflows/CCDG_aligner_functional	Yilin Xu	CWL	GitHub	
DataBiosphere/topmed-workflows/UM_variant_caller_cwl	Vladimir Obucina	CWL	GitHub	

Items per page: 10 1 - 7 of 7

# Usability Improvements

- **Lots of new documentation and tutorials!**
- Example Topics:
  - Getting Started With...(Docker, Nextflow, CWL, WDL)
  - Launching Tools and Workflows
  - Writing checker workflows
  - Developing File Provisioning Plugins
  - Creating Organizations
  - And many more!

<https://docs.dockstore.org/>

## Developer Tutorial

Go through the process of creating a tool and registering it on Dockstore.

## End User Tutorials

Learn how to use Dockstore from the perspective of a user who runs tools and workflows.

## Advanced Tutorials

A collection of articles and tutorials regarding advanced Dockstore topics

[Documentation](#) / [Advanced Developer](#) / [Organizations and Collections](#)

## Organizations and Collections

### Organizations

Organizations are landing pages for collaborations, institutions, consortiums, companies, etc. that allow users to showcase tools and workflows. This is achieved through the creation of collections, which are groupings of related tools and workflows. The users of an organization do not need to own the tools or workflows in any way; the tools and workflows just have to be published. Collections can be thought of as a playlist on a music streaming service where tools and workflows are analogous to individual songs. They can be shared publicly, and the user does not need to own them.

### Creating an organization

To create an organization request, go to the [organizations](#) page and select [Create Organization Request](#). Any user can request to create an organization by filling out the following form. For now, the request must be approved by a Dockstore curator in order to be public. Until it is approved, you are still able to edit it, add collections, add members, etc.

#### Create Organization Request

Fill out the form to send an organization request to a Dockstore curator to approve. Once approved, the organization will be publicly visible.

Name \*

OICR

The name of the organization  
Display Name \*

Ontario Institute for Cancer Research

The display name of the organization  
Topic \*

OICR is a collaborative, not-for-profit research institute accelerating the development of n

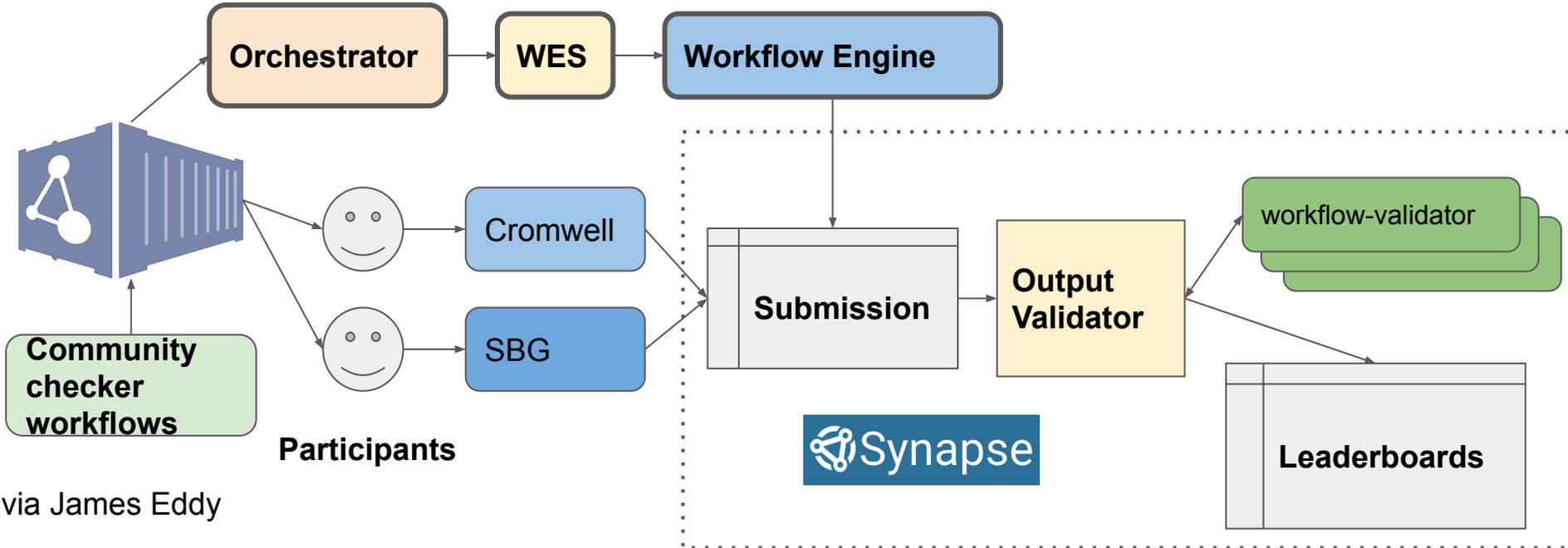
A short description of the organization  
Organization website

<https://oicr.on.ca/>

Link to organization website  
Location

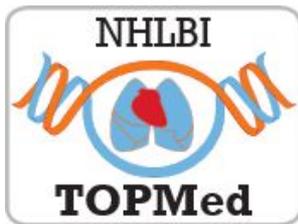
# Highlighted Community Content

GA4GH-DREAM Workflow Testbed (2018)

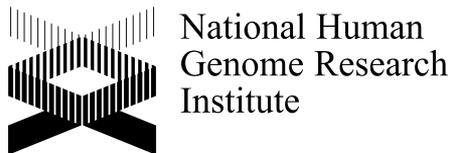


via James Eddy

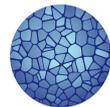
# Collaborations - *The Commons Alliance and Other Projects*



**NHLBI Data  
STAGE**



**NHGRI  
AnVIL**



**HUMAN  
CELL  
ATLAS**



*And Many Others*

*Dockstore will serve as an official workflow provider for the Commons Alliance, a common infrastructure collaboration to facilitate combining **Data + Cloud Compute***

# Next Release - 1.7.0 (Late August/September)

- Prototype for services/apps
  - Dockerized genome browsers, jupyter notebooks, and reference data providers
  - Playground for Github Apps and automatic update syncing
- Immutable workflow versions
  - DOI issuing via linked Zenodo credentials
- Language support updates
  - WDL draft-3 to 1.0
  - cwltool update (possibly CWL 1.1 support)
- Display of logs from when we test and verify workflows
- Improved security of your data and logging, monitoring of the site
- Summary of changes at <https://docs.dockstore.org/changelog/>

# Future Work

- Signing of Entries on Dockstore
  - Verify ownership and integrity of Docker images
- Alternative Containerization Support
  - Singularity and/or uDocker support
- Additional Workflow Languages
- Services/“Apps”
- Testbed Expansion
  - Automated testing for workflows across cloud platforms
- **Open to collaborations with partners for any of these kind of projects!**

# The Dockstore Team



Lincoln Stein

Denis Yuen

Andrew Duncan

Gary Luu



UNIVERSITY OF CALIFORNIA  
**SANTA CRUZ** | Genomics  
Institute

Brian O'Connor

Benedict Paten

Cricket Sloan

Charles Overbeck

Walt Shands

David Steinberg

Louise Cabansay

Natalie Alvarez

Ryan Bautista

Melaina Galaspie

Shreya Radesh

Emily Soth

# Acknowledgements



This work was funded by the Government of Canada through Genome Canada and the Ontario Genomics Institute (OGI-168).



Funded by:



# Extra Slides

Handy stuff in case of questions

# Language Support

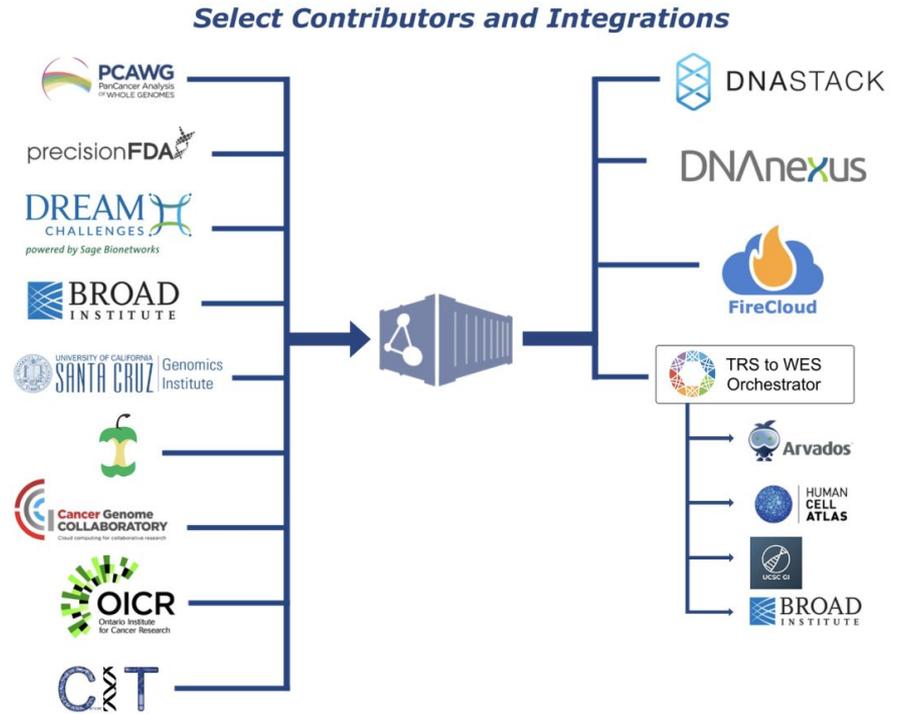
Feature	CWL	WDL	Nextflow
<b>Dockstore site</b>			
Tool registration	Yes	Yes	No
Workflow registration	Yes	Yes	Yes
Hosted Workflows	Yes	Yes	Yes
DAG Display	Yes (cwl version=>1.0) <sup>[0]</sup>	Yes (wdl version<=draft-2) <sup>[1]</sup>	Limited support
Tool Tab Display	Yes (cwl version=>1.0)	Yes (wdl version<=draft-2)	Yes
Launch-with Platforms	Not yet!	FireCloud (workflows only) <sup>[2]</sup> DNASTack (workflows only) <sup>[3]</sup> DNAnexus (workflows only) Terra (workflows only)	Not yet!
Metadata Display	Yes	Yes (wdl version<=draft-2)	Yes
<b>Dockstore CLI</b>			
Local workflow engines	cwltool, Cromwell	Cromwell <sup>[4]</sup>	Nextflow
File Provisioning In	Local File System HTTP FTP S3 via plugins Data Object Service	Local File System HTTP FTP S3 via plugins Data Object Service	Local File System HTTP FTP S3
Plugins Support	s3 s3cmd icgc-get Data Object Service	s3 s3cmd icgc-get Data Object Service	No
File Provisioning Out	Local File System HTTP FTP S3 via plugins	Local File System	Local File System S3
Notifications	Yes	Yes	No

More info:

<https://docs.dockstore.org/docs/user-tutorials/language-support/>

# Dockstore Ecosystem

Dockstore is thankful to its many contributors, users, and partners. This community has pulled together a library of over 450 tools and workflows. In the diagram to the right we've highlighted a few select contributors to give a sense of what has been occurring in this space.





BETA

## Importing from Dockstore

github.com/DataBiosphere/topmed-workflows/UM\_variant\_caller\_wdl  
V. 1.32.0



Please note: Dockstore cannot guarantee that the WDL and Docker image referenced by this Workflow will not change. We advise you to review the WDL before future runs.

```
1 import "https://raw.githubusercontent.com/DataBiosphere/topmed-workflows
2
3
4 ## This is the U of Michigan variant caller workflow WDL for the workflo
5 ## https://github.com/statgen/topmed_freeze3_calling
6 ##
7 ## It uses a Docker image built with software tools that can reproduce
8 ## variant calls compatible to TopMed Freeze 3a
9 ##
10 ## NOTE: This workflow assumes that input CRAM files have been built wit
11 ## human reference genome. In particular for the TopMed CRAM files the
12 ## reference genome files to use are located here:
13 ## ftp://share.sph.umich.edu/gotcloud/ref/hs38DH-db142-v1.tgz
14 ##
15 ##
16
17 workflow TopMedVariantCaller {
18   Boolean? calculate_DNA_contamination
19   Boolean calculate_contamination = select_first([calculate_DNA_contamin
20
21   Boolean? dynamically_calculate_file_size
22   Boolean dynamically_calculate_disk_requirement = select_first([dynamic
```

## Destination Workspace

[Or create a new workspace](#)

# Running Dockstore Tools - Launch With FireCloud

## Importing github.com/DataBiosphere/topmed-workflows/Functional\_Equivalence:1.30.0 from Dockstore

⚠ Please note: Dockstore cannot guarantee that the WDL and Docker image referenced by this Workflow will not change. We advise you to review the WDL before future runs.

Name

Functional\_Equivalence

Destination Workspace

Select a workspace ▾

Export

```
1 ## Copyright Broad Institute, 2017
2 ##
3 ## This WDL pipeline implements CCGD Pipeline standards to process
4 ## high-throughput sequencing data for downstream processing.
5 ##
6 ## Requirements/expectations :
7 ## - Human whole-genome pair-end sequencing data in unmapped BAM (uBAM) format
8 ## - One or more read groups, one per uBAM file, all belonging to a single sample (SM)
9 ## - Input uBAM files must additionally comply with the following requirements:
10 ## - file names all have the same suffix (we use ".unmapped.bam")
11 ## - files must pass validation by ValidateSamFile
12 ## - reads are provided in query-sorted order
13 ## - all reads must have an RG tag
14 ## - Reference genome must be Hg38 with ALT contigs
15 ##
16 ## Runtime parameters are optimized for Broad's Google Cloud Platform implementation.
17 ## For program versions, see docker containers.
18 ##
19 ## LICENSING :
20 ## This script is released under the WDL source code license (BSD-3) (see LICENSE in
21 ## https://github.com/broadinstitute/wdl). Note however that the programs it calls may
22 ## be subject to different licenses. Users are responsible for checking that they are
23 ## authorized to run all programs before running this script. Please see the docker
24 ## page at https://hub.docker.com/r/broadinstitute/genomes-in-the-cloud/ for detailed
25 ## licensing information pertaining to the included programs.
26 ##
27 # WORKFLOW DEFINITION
28 workflow PairedEndSingleSampleWorkflow {
29
30   File wgs_evaluation_interval_list
31   File wgs_coverage_interval_list
32
33   String sample_name
34   String base_file_name
35   Array[File] flowcell_unmapped_bams
36
```