Dockstore 2.0: Enhancing a community platform for sharing cloud-agnostic research tools

Denis Yuen[1], Brian O'Connor[2], Cricket Sloan[2], David Steinberg[2], Natalie Perez[2], Walt Shands[2], Gary Luu[1], Andrew Duncan[1], Charles Overbeck[2], Louise Cabansay[2], Lincoln Stein[1]

[1] Ontario Institute for Cancer Research, MaRS Centre, Toronto, Ontario   * Corresponding author (denis.yuen@oicr.on.ca)
[2] UC Santa Cruz Genomics Institute, University of California, Santa Cruz, CA, USA

Background

Dockstore is a platform that was created in response to the many challenges faced during the PCAWG study (Pan-Cancer Analysis of Whole Genomes). It is a repository of scientific workflows described using popular workflow languages with their dependencies distributed in Docker images. This powerful combination of technologies allows for improved sharing and portability of scientific workflows.

Docker is used to describe the environment that the workflow will run in along with all of the dependencies required.

Workflow languages are used to describe the steps involved in running a workflow and how they are dependent on each other, including all inputs and outputs. We currently support CWL, WDL and Nextflow. These descriptor documents can be stored on GitHub, Bitbucket, GitLab, or directly on dockstore.org.
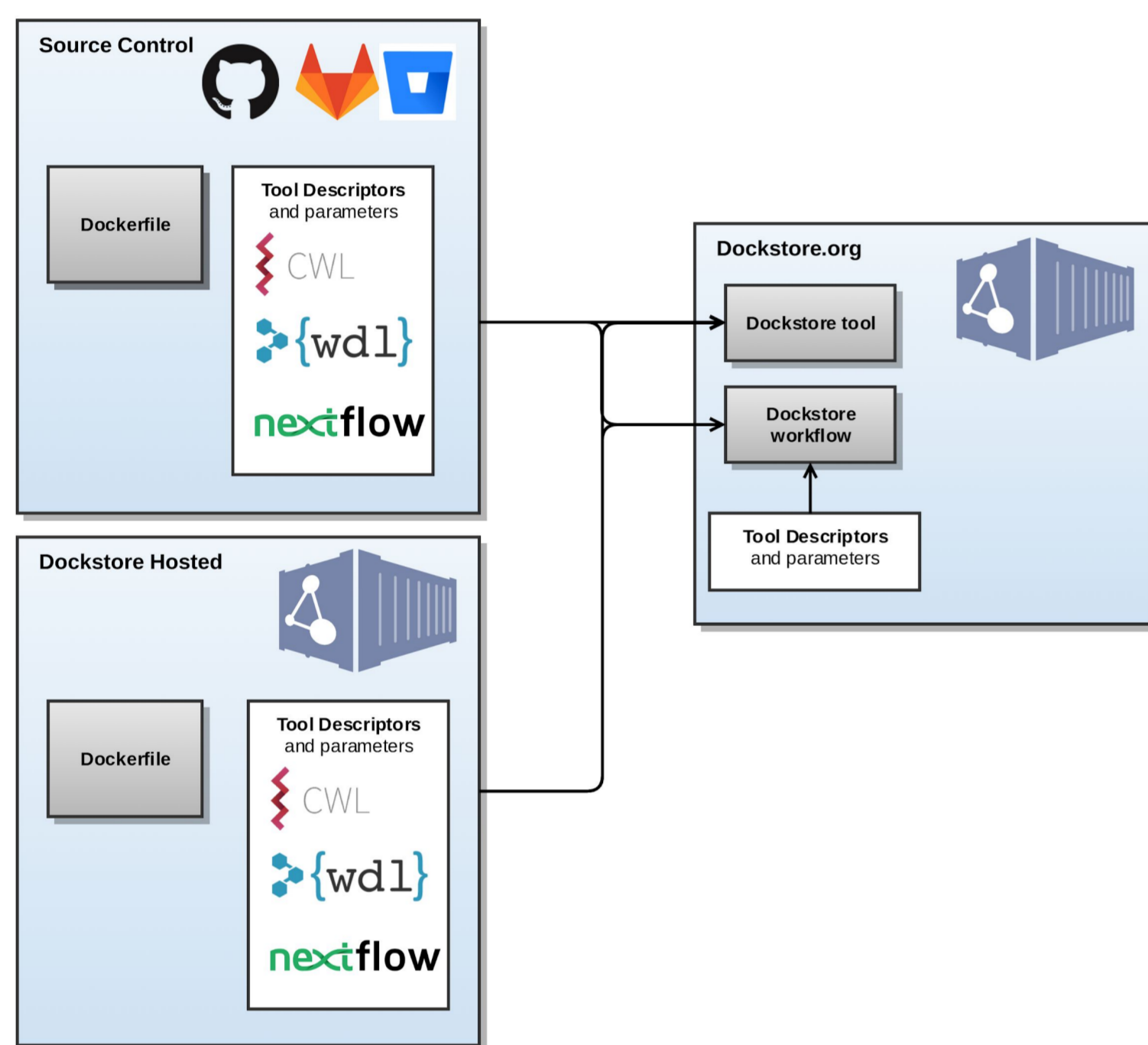
Benefits

- Searchable catalogue of tools and workflows accessible via a proposed GA4GH standard API
- Launch workflows locally and in a variety of cloud platforms
- Create organization landing pages for your team, lab, grant, or institution
- Social features such as starring, labels, and discussion threads

Feature	CWL	WDL	Nextflow
Tool registration	Yes	Yes	No
Workflow registration	Yes	Yes	Yes
Hosted workflows	Yes	Yes	Coming soon!
DAG Display	Yes[1]	Yes[2]	Limited support
Tool Tab Display	Yes[1]	Yes[2]	Yes
Launch-with Platforms	Not yet!	FireCloud[3] DNAstack[3] DNAnexus[3]	Not yet!
Metadata Display	Yes	Yes[2]	Yes

https://docs.dockstore.org/docs/user-tutorials/language-support/
[1] CWL version 1.0
[2] WDL version draft-2
[3] workflows only

Community

Dockstore is thankful to its many contributors, users, and partners. Here we present a few to give a sense of what is going on in this space.
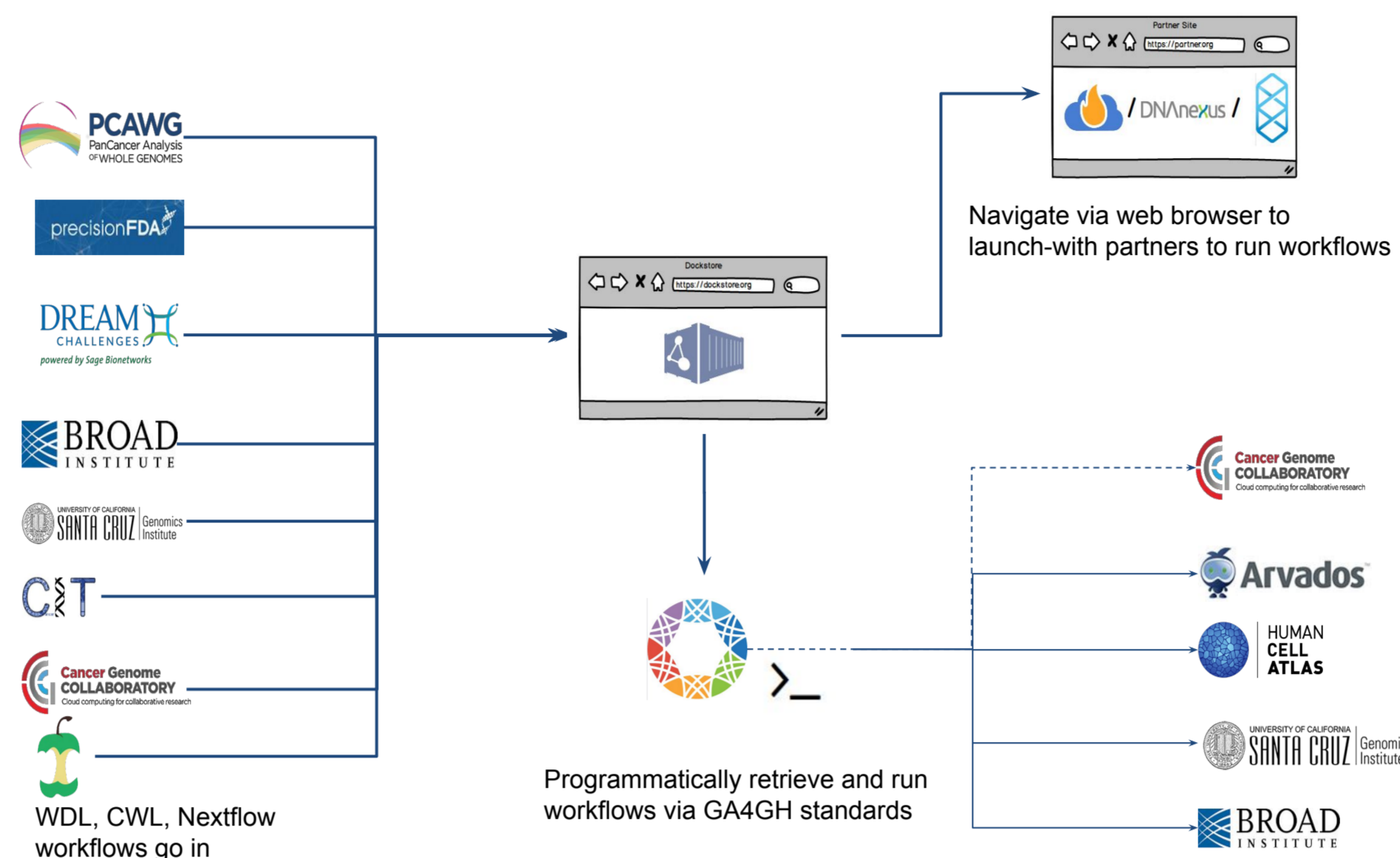
The GA4GH-DREAM tool and workflow series of infrastructure challenges sought to promote the description of tools and workflows using CWL and WDL and enumerate the platforms that could run them

**2017** challenge: participation-driven

2018 challenge: forged onward with an API-based (TRS, WES) pilot

The CWL reference implementation can run tools directly from tool registry services in your terminal-based development environment. See Use with GA4GH Tool Registry API

The nf-core series of high quality Nextflow workflows has been registered on Dockstore allowing users to see them alongside WDL and CWL workflows

Other highlighted contributors to Dockstore include workflows from the Cancer IT (Sanger Institute), the Broad Institute, the TOPMed (Trans-Omics for Precision Medicine) program and DataSTAGE (Storage, Toolspace, Access and analytics for biG data Empowerment).
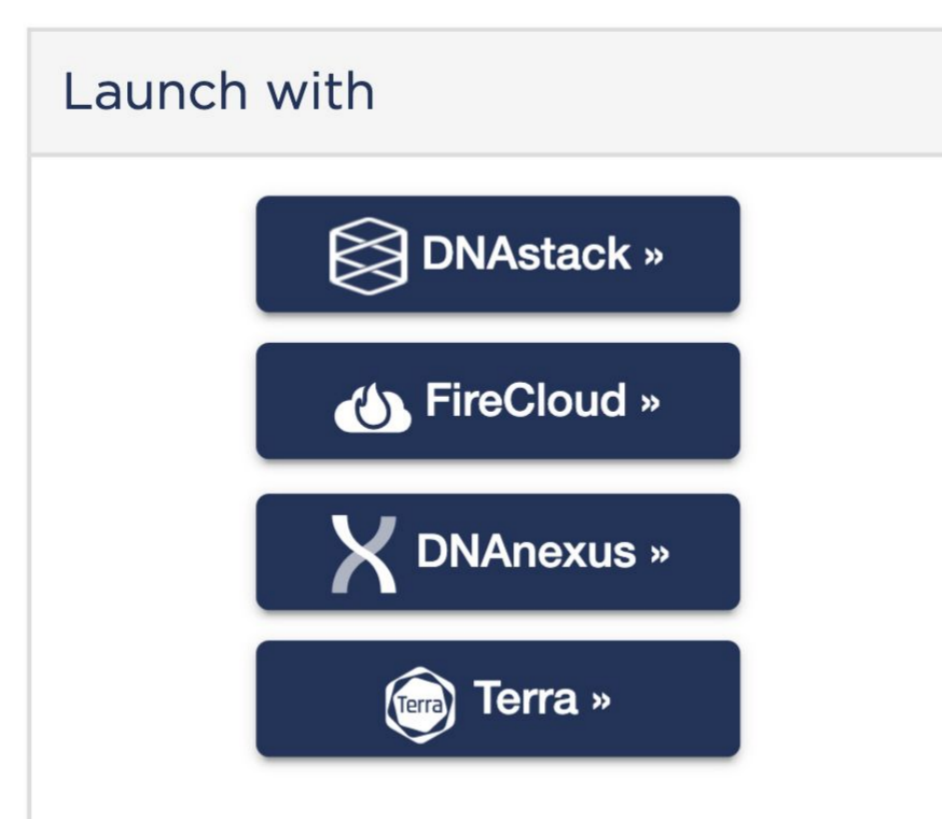
Run workflows on a variety of cloud platforms using a web browser

Integrated with community visualizers like view.commonwl.org for CWL and EPAM for WDL.

Pictured is a Nextflow visualization with the Docker image for a step highlighted
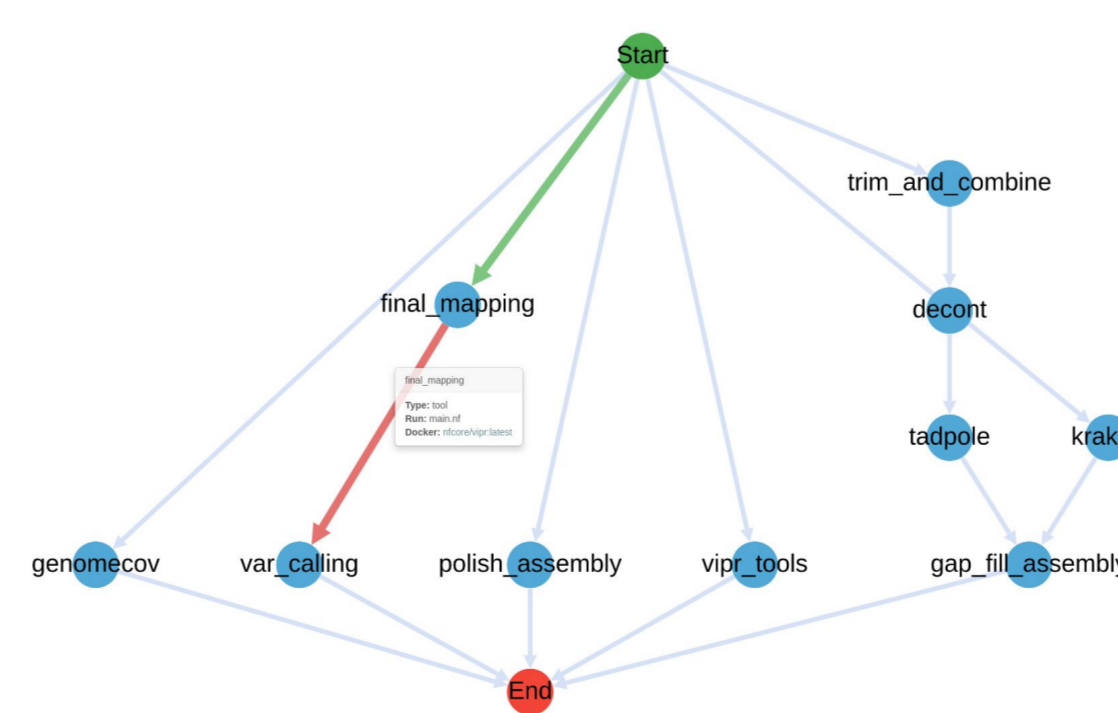
Highlighted New Features

Dockstore 1.7.0 (projected)

- **Services Prototype**: experimental support for long-lived services such as genome browsers and reference data servers for running workflows
- **Easy Registration Prototype:** registration using GitHub Apps, automatically keep repos up to date via GitHub hooks
- **Immutable and DOI issuing**: freeze workflow releases and issue DOIs for them using Zenodo
- **Improved support for CWL, WDL**
  - testing with a cwltool supporting CWL 1.1
  - testing with WDL 1.0 parsing libraries
- **Display testing logs for verified workflows**

Dockstore 1.6.0 (2019-04)

- **Dockstore Organizations**: Create landing pages to describe and group workflows based on institution, grant, theme, etc.
- **Beta WES Client**: launch a Dockstore workflow on any GA4GH WES-compatible platform
- **Better CWL, WDL, and Nextflow Language Support:**
  - Nextflow parsing improvements
  - cwltool and Cromwell compatibility update

Dockstore 1.5.0 (2018-09)

- **Hosted Tools/Workflows**: also store tool/workflow descriptors and test parameter files directly on Dockstore.org
- **Nextflow Support**: register, search, and visualize Nextflow workflows
- **DRS File Plugin**: provision files via the GA4GH-DRS standard for our CLI

Future Work

Looking for collaborators / testers

- **Alternative Containerization Support**: Singularity and/or uDocker support
- **Additional Workflow Languages**: Support for more languages and platforms
- **Signing of entries on Dockstore**: Verify ownership and integrity of Docker images
- **CI Environment**: Automated testing for workflows across cloud platforms

References

- O'Connor BD, Yuen D, Chung V *et al.* The Dockstore: enabling modular, community-focused sharing of Docker-based genomics tools and workflows [version 1; referees: 2 approved]. *F1000Research* 2017, **6**:52 (doi: 10.12688/f1000research.10137.1)
- Denis Yuen, Andrew Duncan, Victor Liu, Brian O'Connor, Gary Luu, Charles Overbeck, … Abraham. (2019, April 5). ga4gh/dockstore: 1.6.0 (Version 1.6.0). Zenodo. http://doi.org/10.5281/zenodo.2630727
- Gary Luu, Andrew Duncan, Denis Yuen, Kitty Cao, JWKaiqi, Charles Overbeck, … angular-cli. (2019, March 20). dockstore/dockstore-ui2: 2.3.0-rc.0 (Version 2.3.0-rc.0). Zenodo. http://doi.org/10.5281/zenodo.2600372
- Amstutz, Peter; Crusoe, Michael R.; Tijanić, Nebojša; Chapman, Brad; Chilton, John; Heuer, Michael; Kartashov, Andrey; Leehr, Dan; Ménager, Hervé; Nedeljkovich, Maya; Scales, Matt; Soiland-Reyes, Stian; Stojanovic, Luka (2016): Common Workflow Language, v1.0. figshare. https://doi.org/10.6084/m9.figshare.3115156.v2
- Voss K, Gentry J and Van der Auwera G. Full-stack genomics pipelining with GATK4 + WDL + Cromwell [version 1; not peer reviewed]. *F1000Research* 2017, **6**(ISCB Comm J):1379 (poster) (https://doi.org/10.7490/f1000research.1114631.1)
- Di Tommaso, P., Chatzou, M., Floden, E. W., Barja, P. P., Palumbo, E., & Notredame, C. (2017). *Nextflow enables reproducible computational workflows*. Nature Biotechnology, 35(4), 316 doi:10.1038/nbt.3820

dockstore