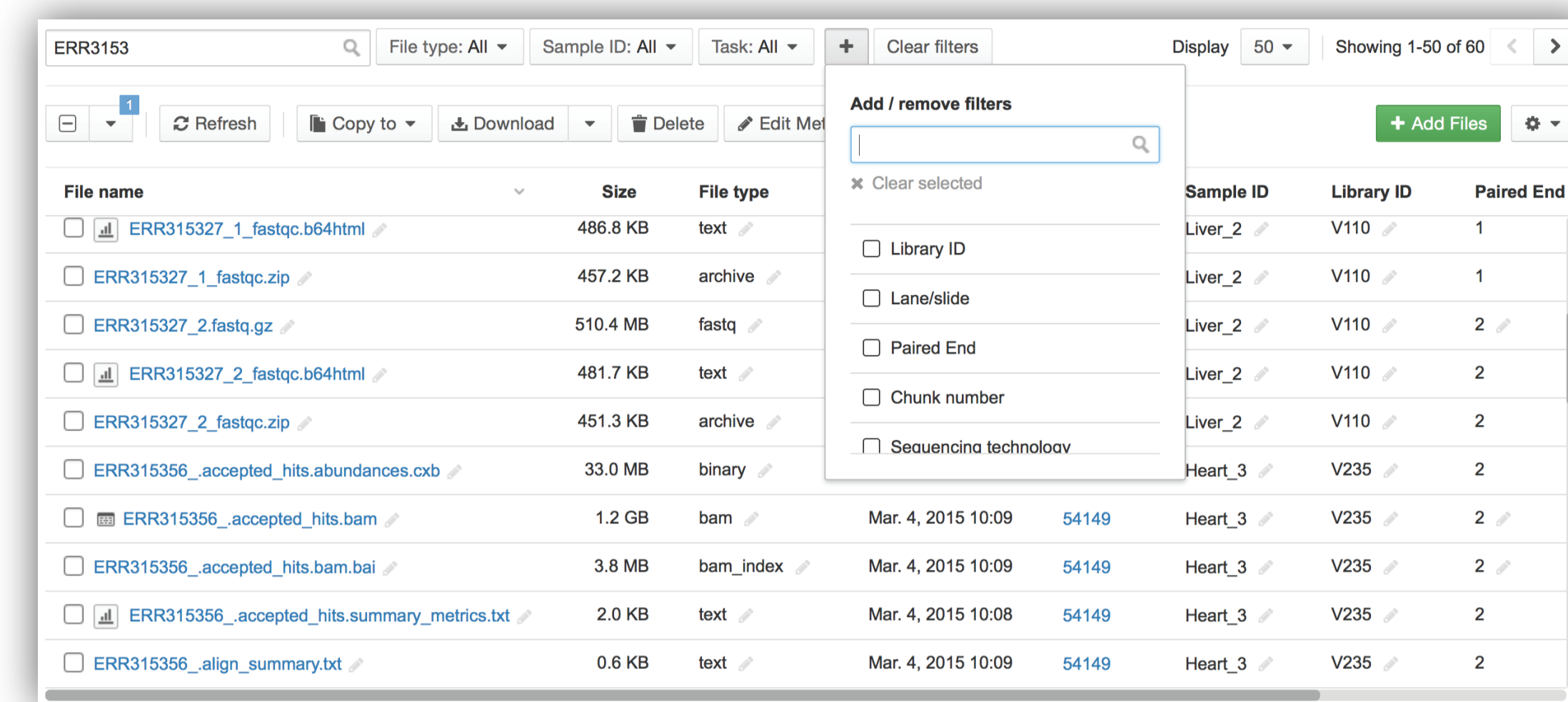


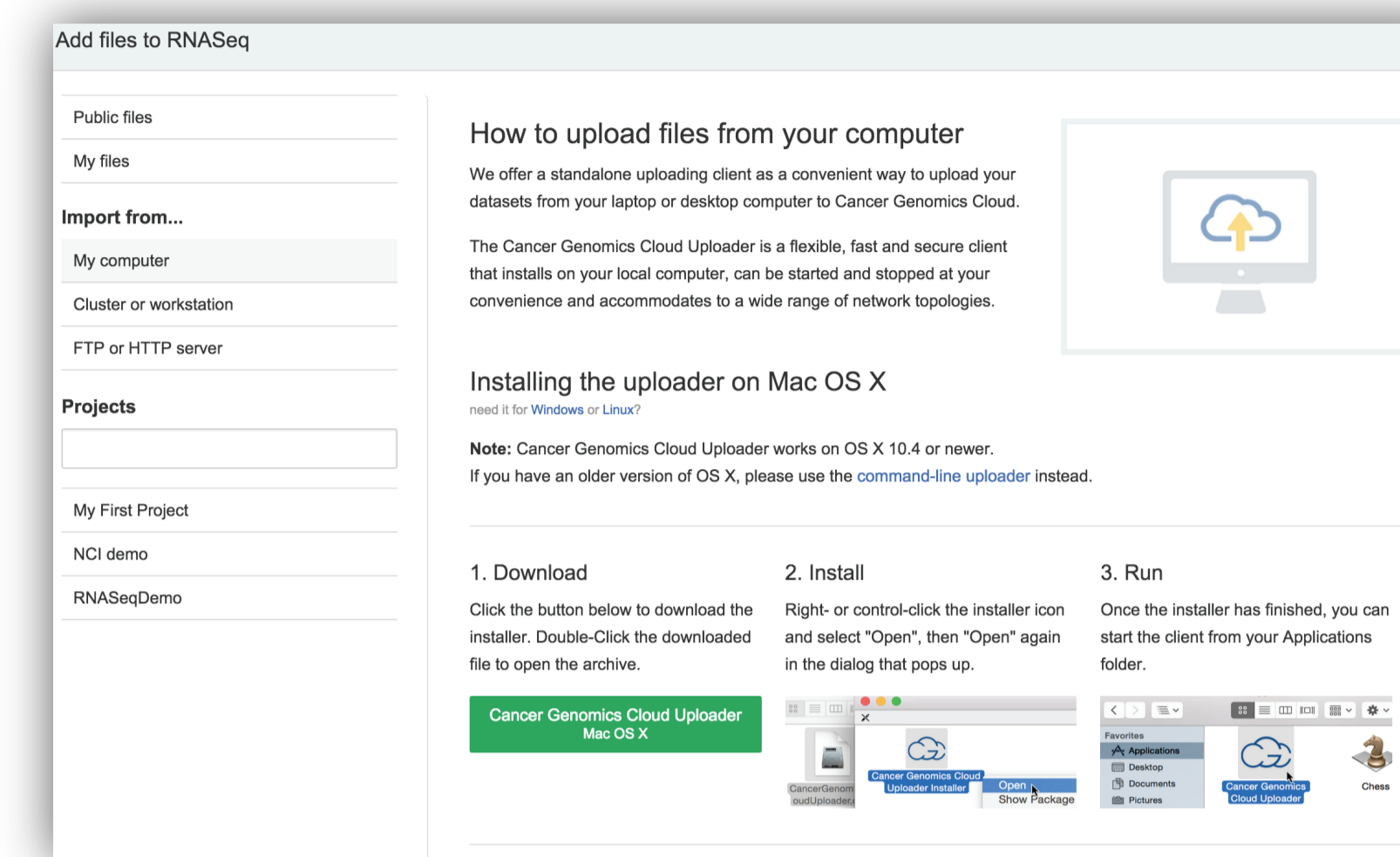
## Background

The advent of next generation sequencing has transformed our ability to generate genomic data. Today, cancer researchers have access to petabytes of multi-dimensional information from thousands of patients. However, analysis of this information only becomes more challenging as the amount of data continues to increase. This difficulty is exemplified when we consider data generated by the efforts of The Cancer Genomics Atlas (TCGA) network. Simply downloading the complete TCGA repository would require several weeks with a highly optimized network connection. Once downloaded, integrated analysis of this data remains out of reach for any researcher without access to the largest institutional compute clusters. The Cancer Genomics Cloud (CGC) Pilots project seeks to directly address these challenges by co-localizing data with the computational resources to analyze it, without the need to wait in a queue. The CGC will enable researchers to securely leverage the power of cloud computing to gain biologically relevant and actionable insights from massive public datasets including TCGA. Reproducible analysis of public (including controlled and open-access data) as well as private data can be performed using both application program and graphical user interfaces. The CGC features Rabix, a powerful implementation of the Common Workflow Language which enables developers to easily create and share portable and reproducible tools and pipelines. The CGC will be released to the community for evaluation and feedback by the end of 2015. **Pre-register at [cancergenomicscloud.org](http://cancergenomicscloud.org)**

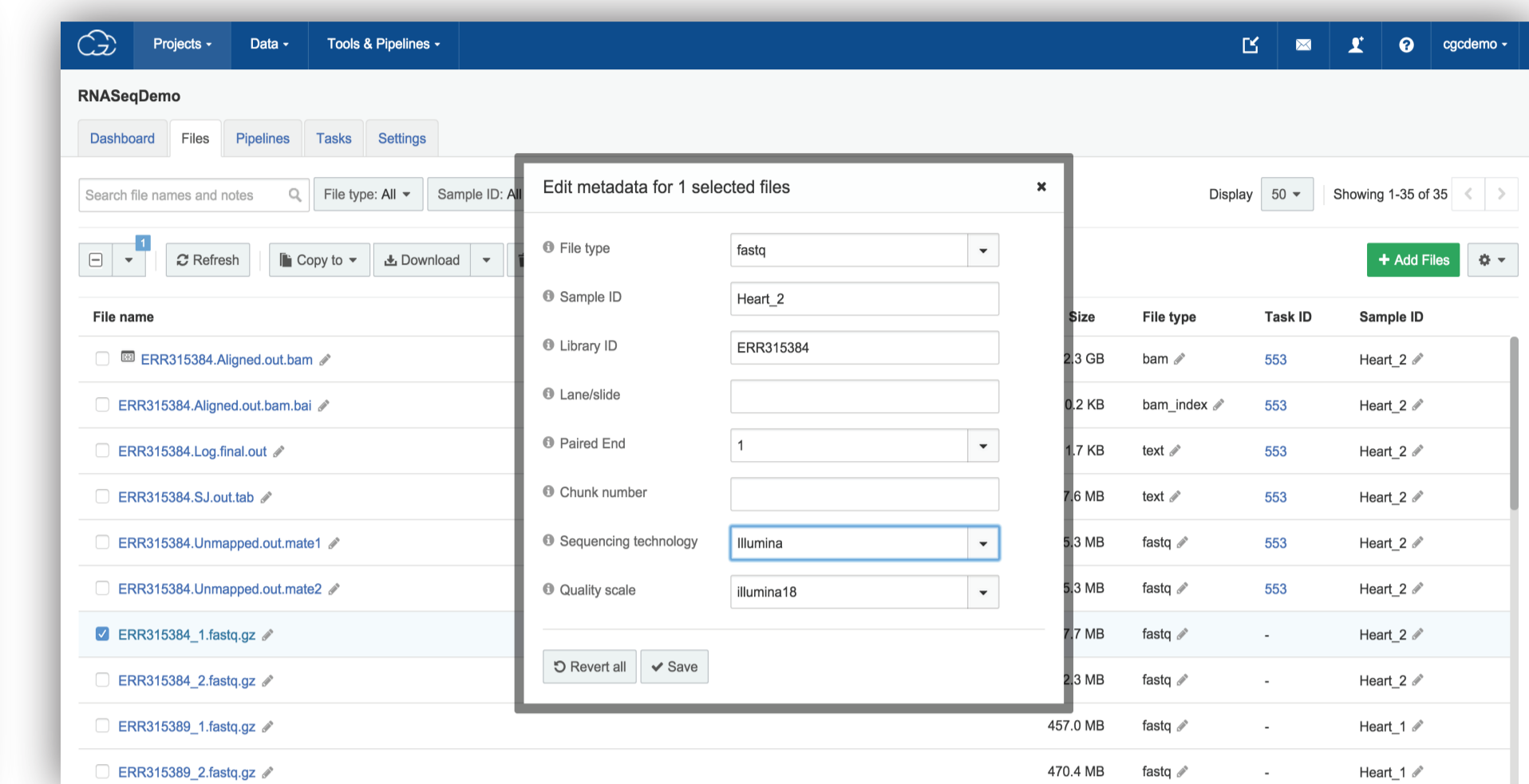
## The Seven Bridges Cancer Genomics Cloud Pilot helps you:



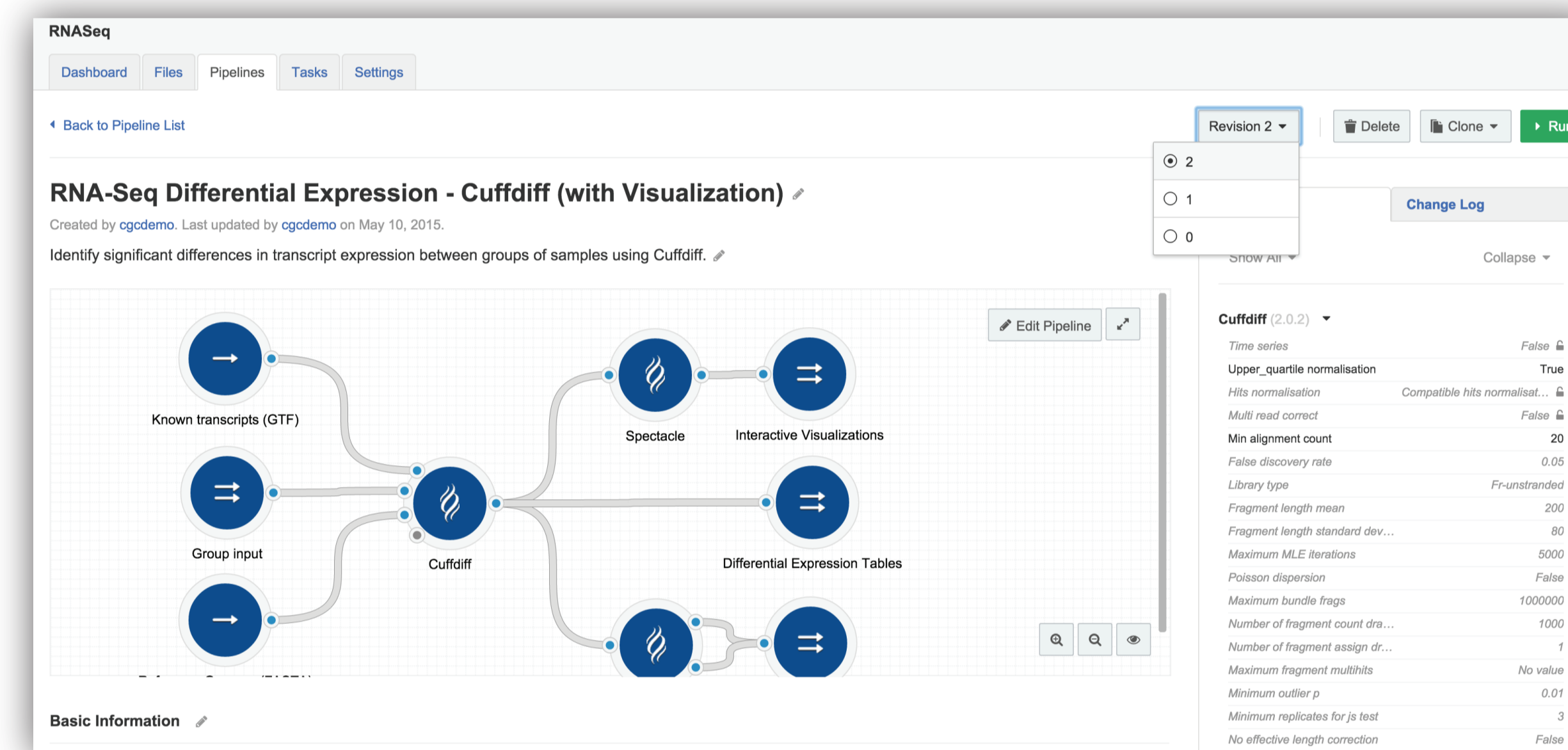
Find and discover Controlled & Public Access TCGA data



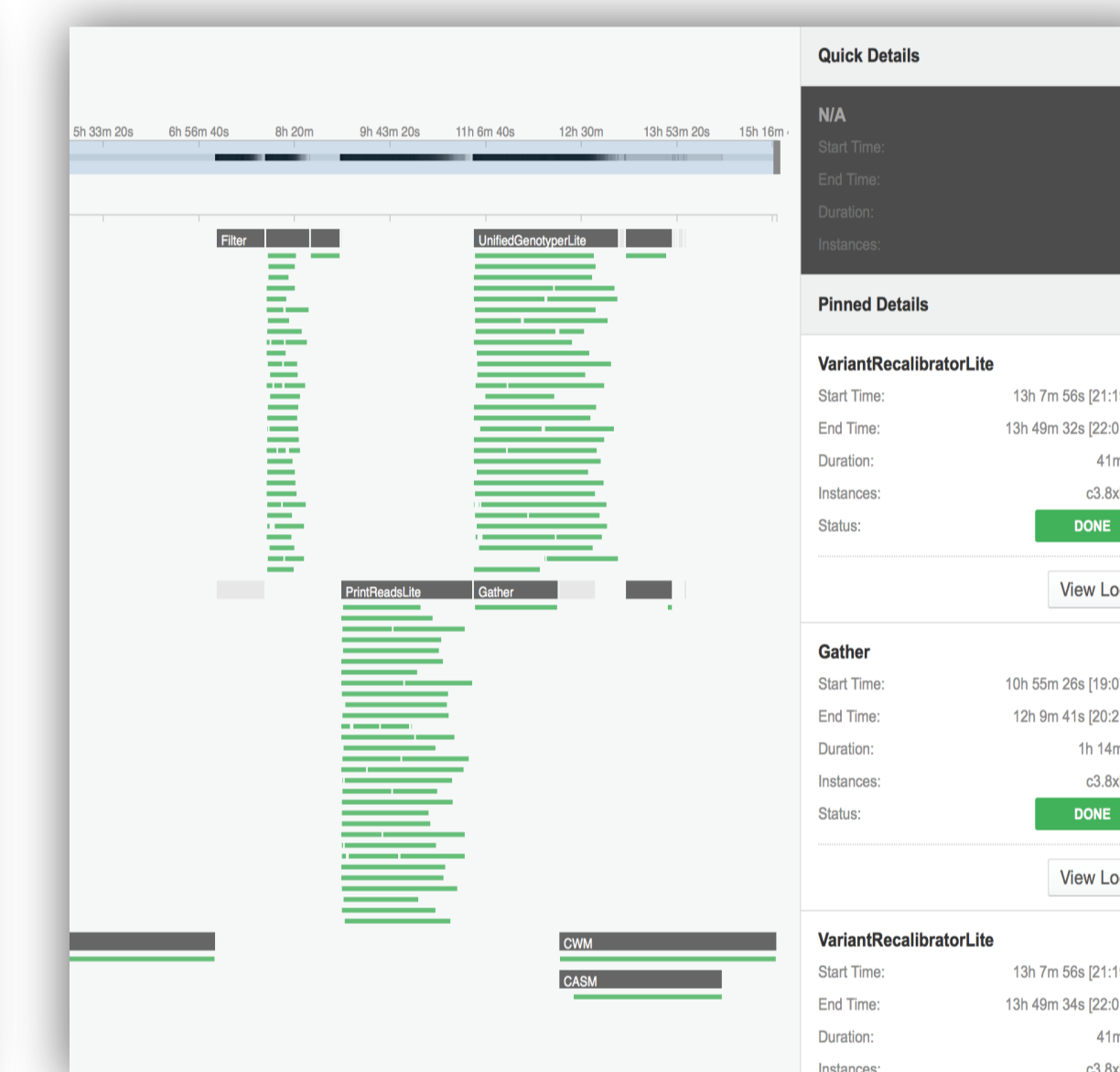
...add your own data



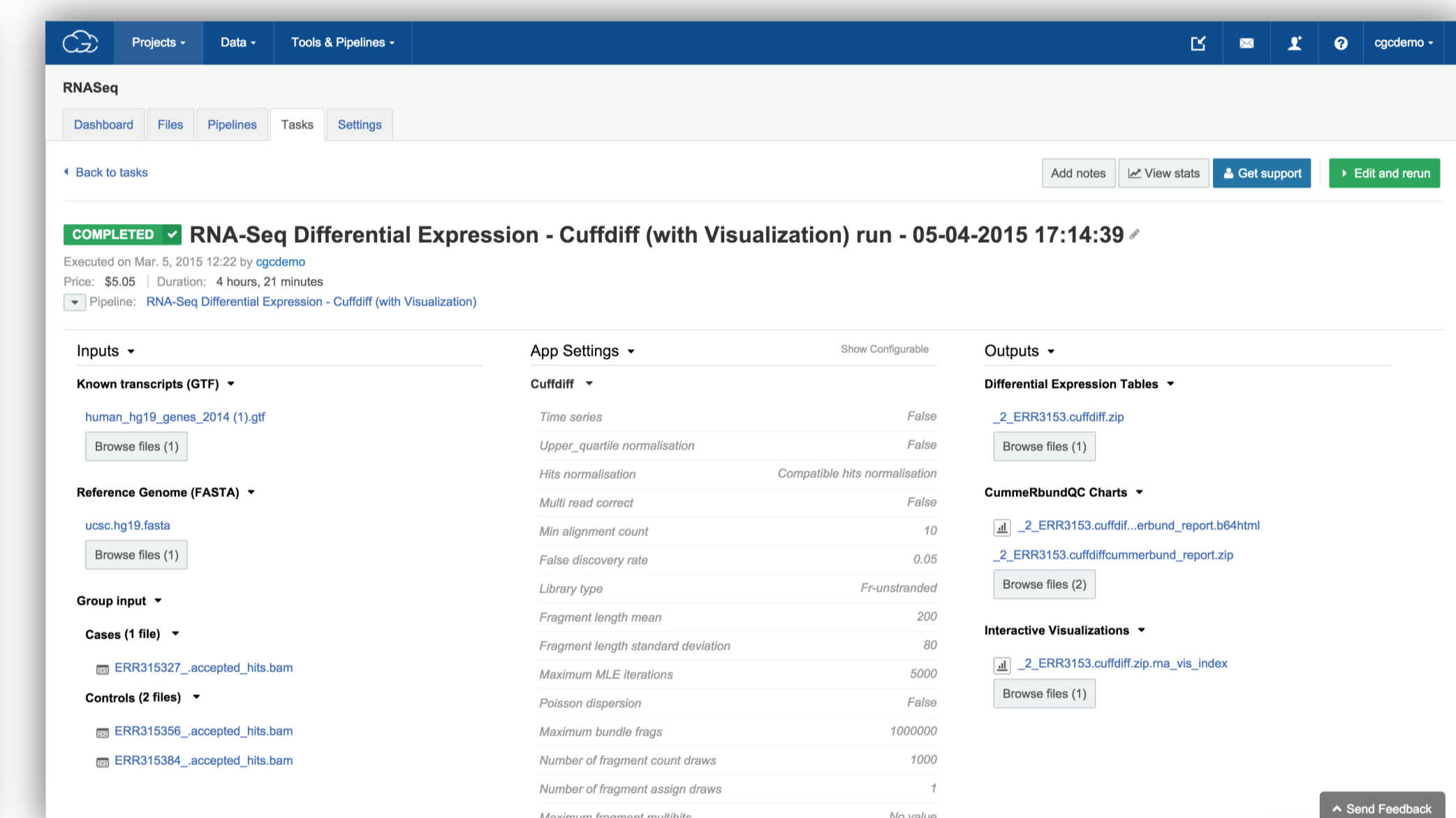
...and easily manage metadata properties.



Build reproducible computational pipelines by combining public tools



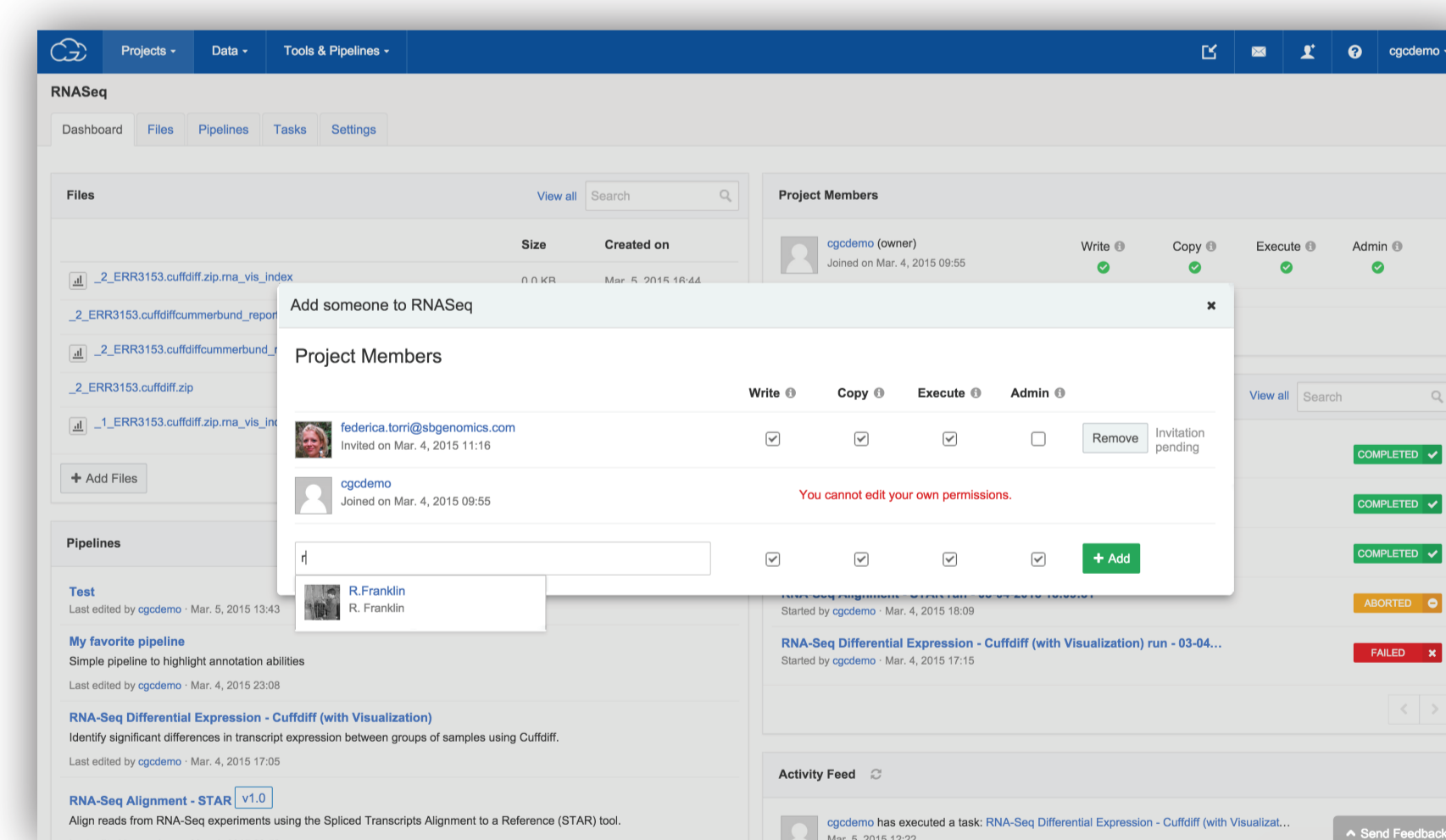
...monitor and optimize executions



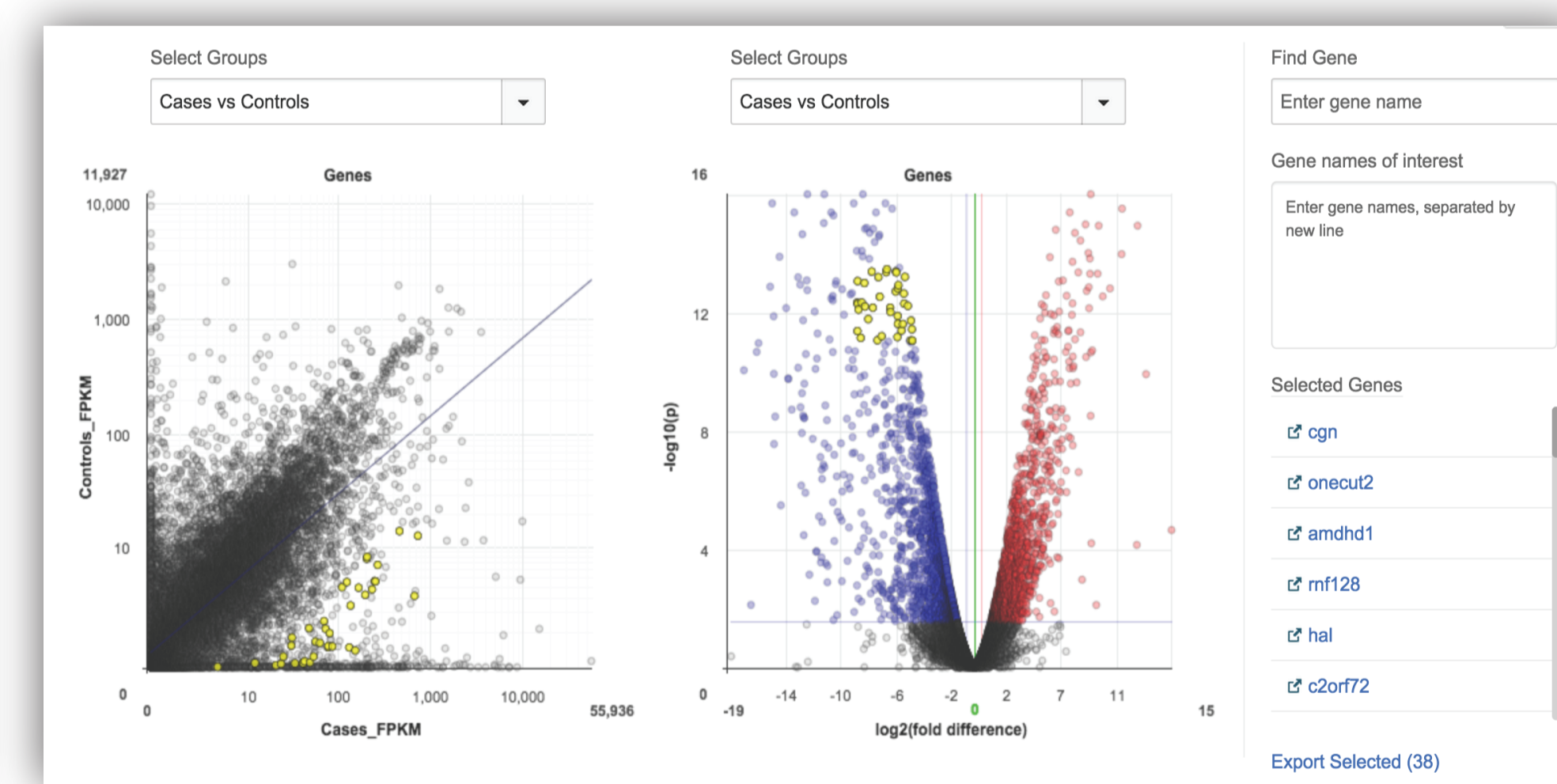
...and capture the complete provenance of every output.

## How much data is in a petabyte?

1 byte	A single character	!
2 kilobytes	One type-written page	📄
100 megabytes	Two-volume encyclopedia	📚
100 gigabytes	A floor of academic journals	📖
10 terabytes	The print collections of the Library of Congress	🏛️
2 petabytes	Print material in all US academic research libraries	🇺🇸
2.5+ petabytes	Data generated by TCGA	☁️

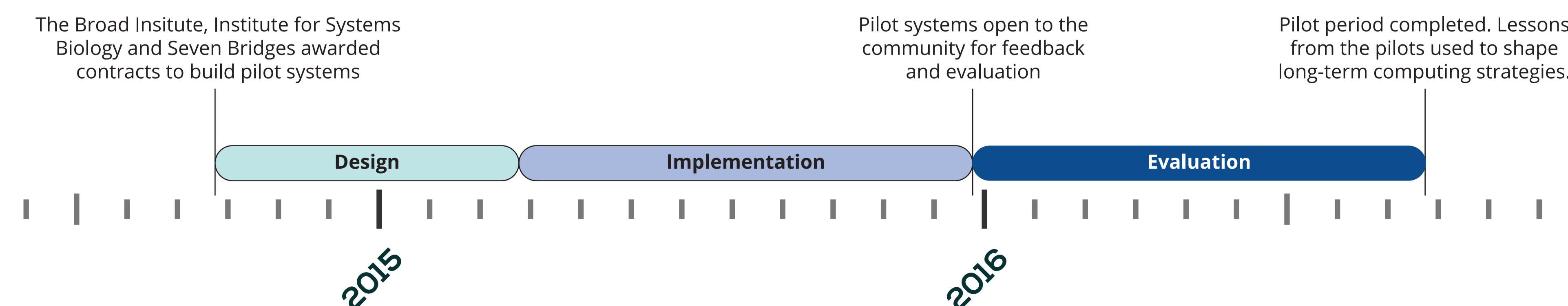
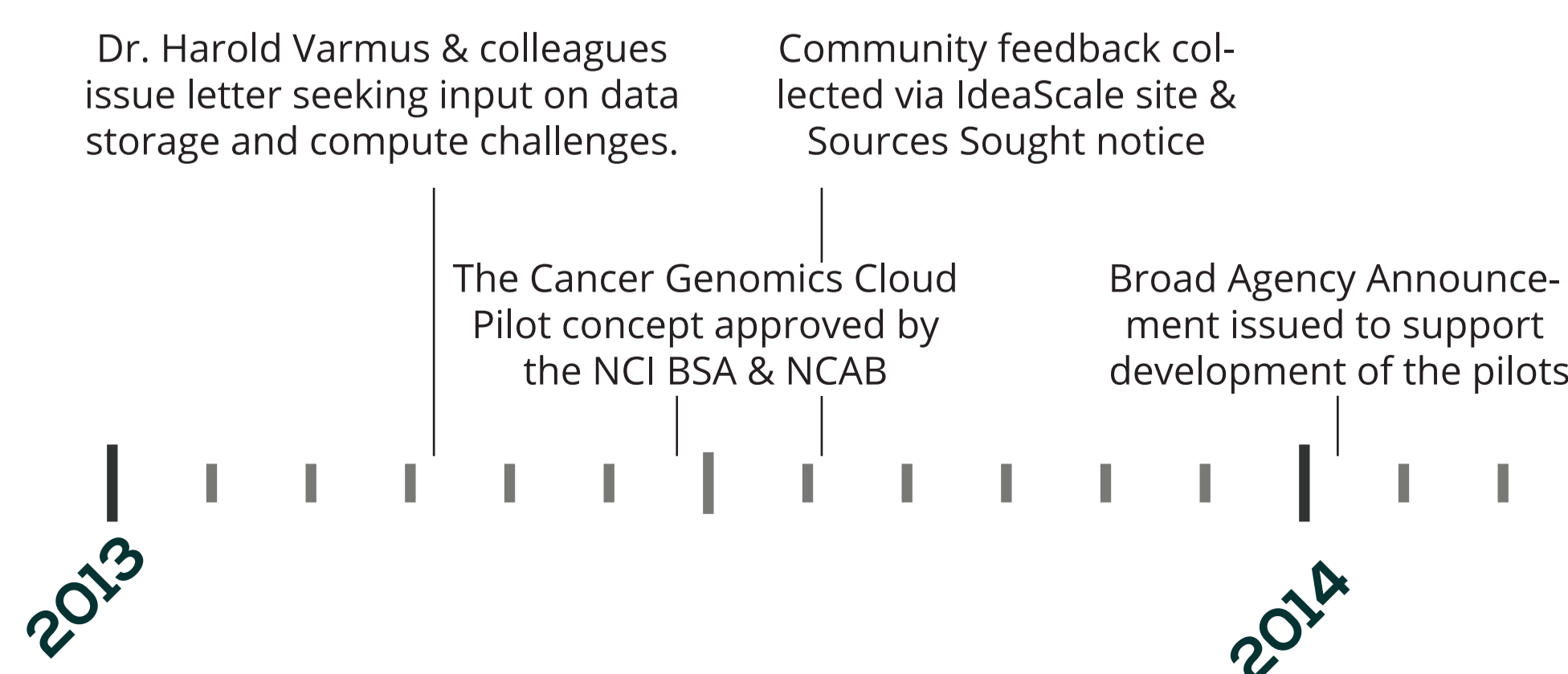


Easily and securely collaborate



...and deeply explore your data.

## Timeline



## For Developers

- Deploy and share portable tools and pipelines using Rabix (the docker-based implementation of CWL)
- Automate workflows, manage permissions, and search for data via API
- Iterate tool parameters and computational instances to optimize executions

Be one of the first.

[cancergenomicscloud.org](http://cancergenomicscloud.org)

@genomicscloud | [cg@sbgenomics.com](mailto:cg@sbgenomics.com)