CANCER GENOMICS CLOUD

SEVEN BRIDGES

TCGA Meeting Workshop

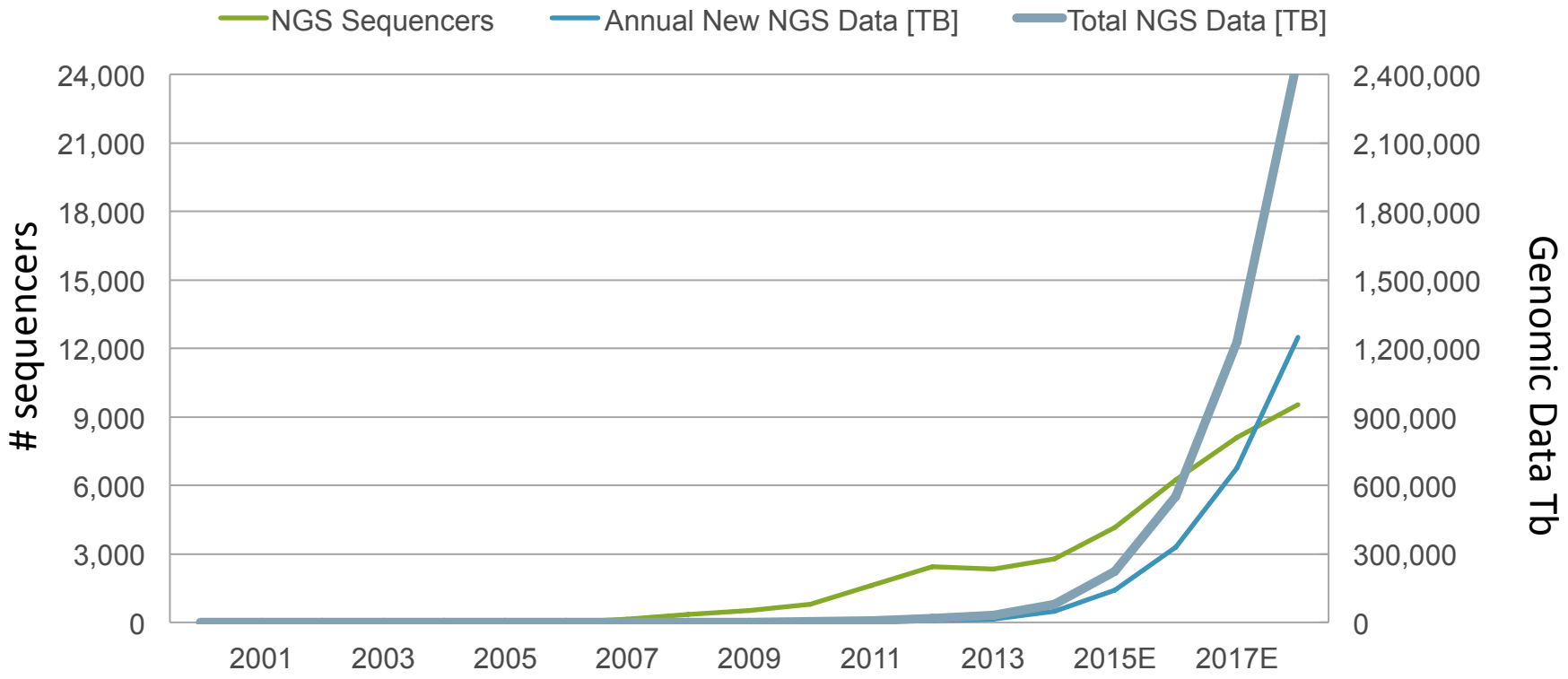
Deniz Kural, PhD

May 12, 2015

+ Brandi, Zeynep, Devin, Kate



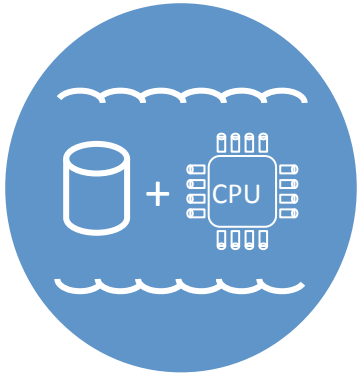

Amount of genomics data will exceed available resources



Between 2014-2018 production of new NGS data to exceed ***2 Exabytes***

NGS: Next Generation Sequencing
NGS sequencers include machines from Illumina, Life Technologies, and Pacific Biosciences. Human genome data based on estimates of whole human genomes sequenced
Sources: Financial reports of Illumina, Life Technologies, Pacific Biosciences; revenue guidances; JP Morgan; The Economist; Seven Bridges Analysis.

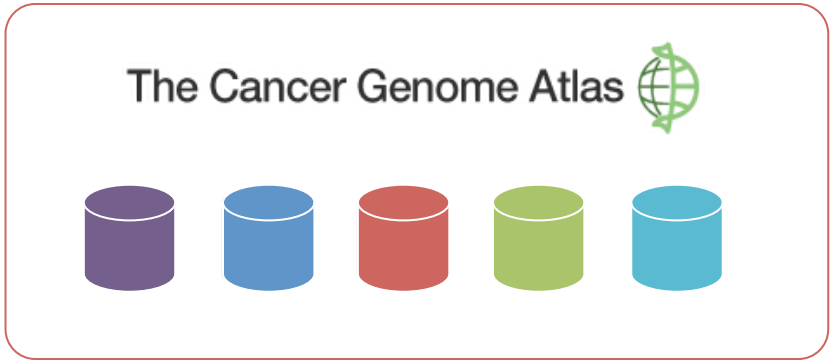
Large-scale cancer genomics will shift how computation is done

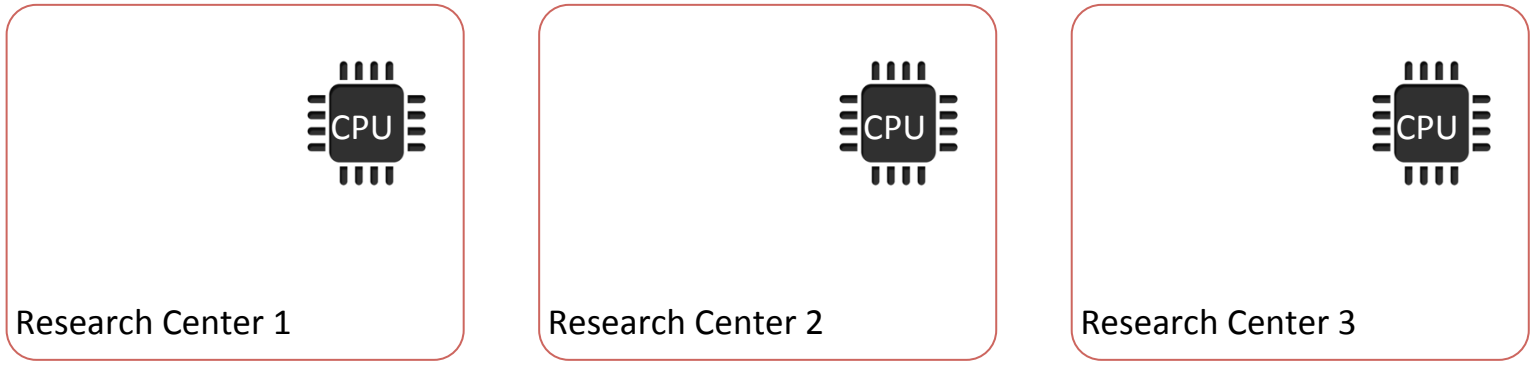


CLOUD COMPUTATION

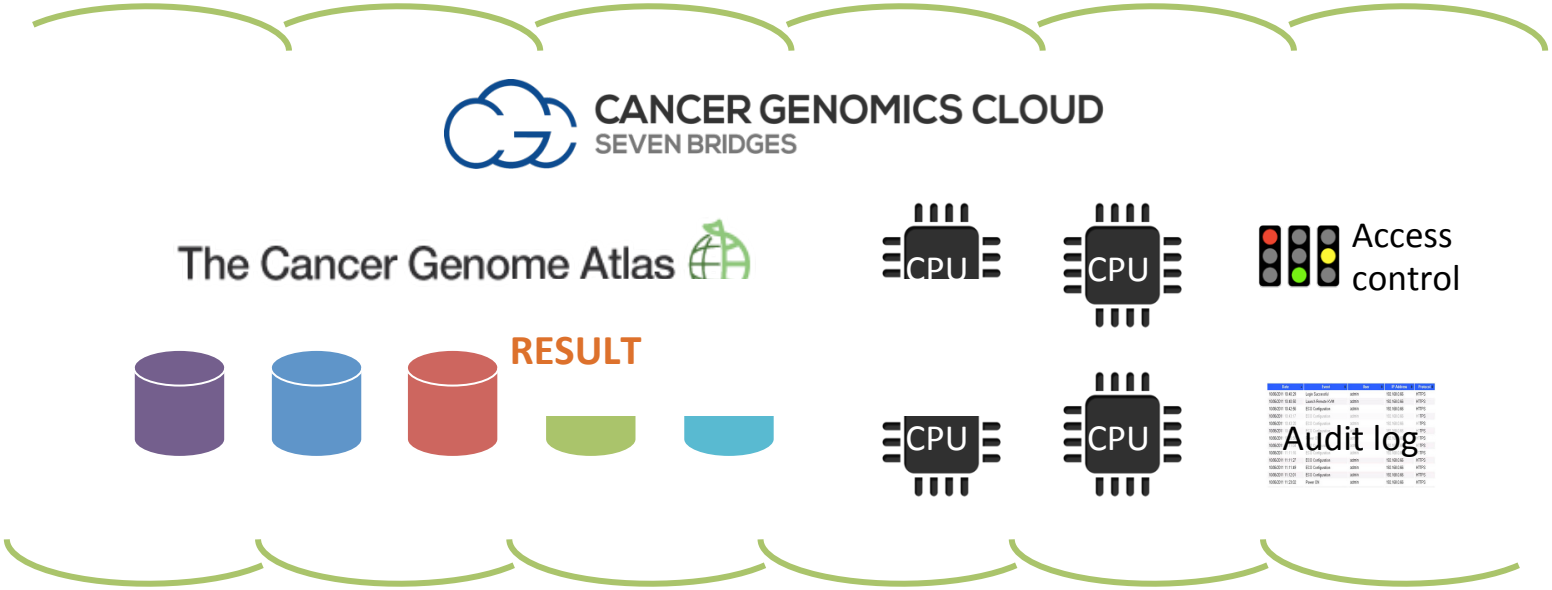
PORTABLE WORKFLOWS
replace
data transfers

Standard Model: Data comes to computation





Cloud: Algorithms come to the data



Workflow **Research Center 1**

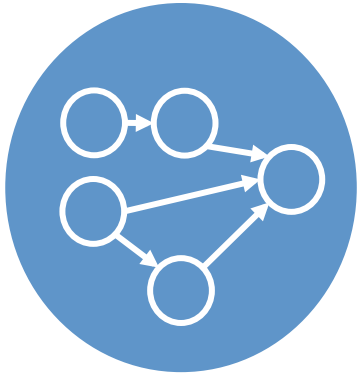
Workflow **Research Center 2**

Workflow **Research Center 3**

Other awardees of the NCI Cancer Cloud pilots: **Institute for Systems Biology, The Broad Institute**

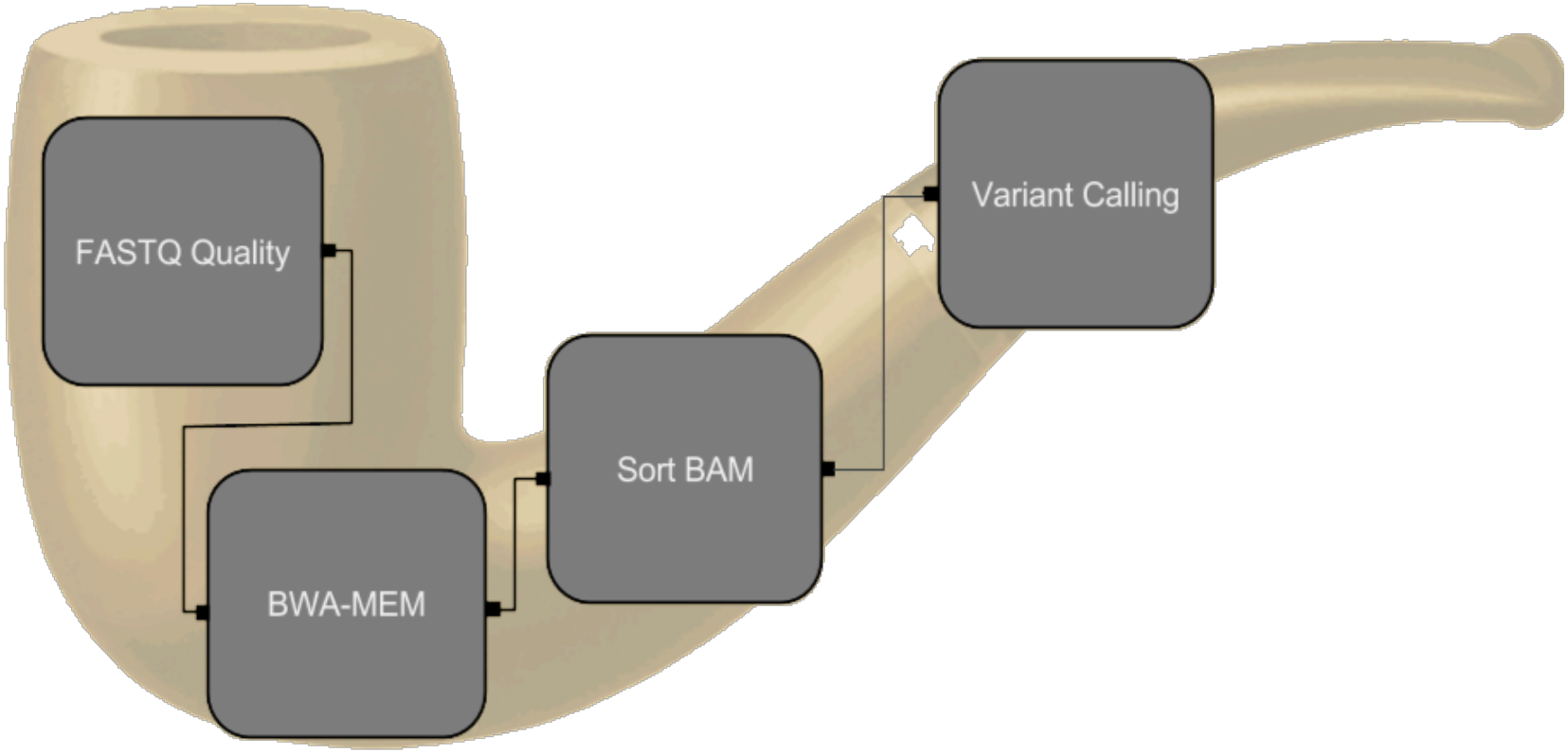
Millions of genomes will shift how computation is done

CLOUD
COMPUTE



PORTABLE WORKFLOWS
replace
data transfers

Ceci n’est pas une pipeline!?



Replicating computations is a hard problem

	Scripts	Machine Images	App containers + Pipeline-Language
Easy distribution			
Simple to run			
Compatible across environments			
100% reproducible			
Modular			
Automated optimization			
Distributed computing			

PORTABLE WORKFLOWS
will replace data transfers

enabled by **open standards** that…

……are independent of the computation environment

…ensure 100% reproducibility and auditability

…are platform independent

Interaction Methods

Programmatic Access:

- API
  - Includes Data API
- SDK
  - Upload & Execute own code

Graphical Analysis Tools:

- Data-Mining Tools
- Genome Browser
- QC tools

Upload additional datasets, analyze in tandem with TCGA

Public files

My files

Import from...

My computer

Cluster or workstation

FTP or HTTP server

Projects

My First Project

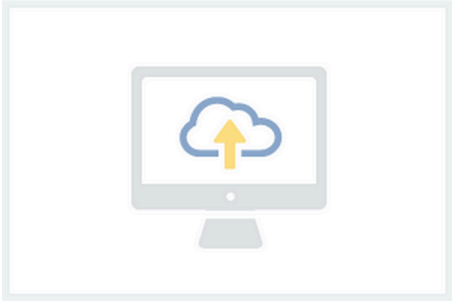
NCI demo

RNASeqDemo

How to upload files from your computer

We offer a standalone uploading client as a convenient way to upload your datasets from your laptop or desktop computer to Cancer Genomics Cloud.

The Cancer Genomics Cloud Uploader is a flexible, fast and secure client that installs on your local computer, can be started and stopped at your convenience and accommodates to a wide range of network topologies.



Installing the uploader on Mac OS X

need it for Windows or Linux?

Note: Cancer Genomics Cloud Uploader works on OS X 10.4 or newer.
If you have an older version of OS X, please use the command-line uploader instead.

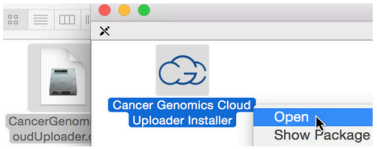
1. Download

Click the button below to download the installer. Double-Click the downloaded file to open the archive.

Cancer Genomics Cloud Uploader
Mac OS X

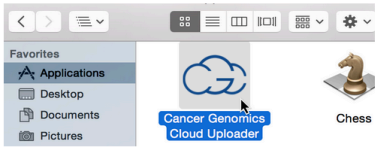
2. Install

Right- or control-click the installer icon and select "Open", then "Open" again in the dialog that pops up.

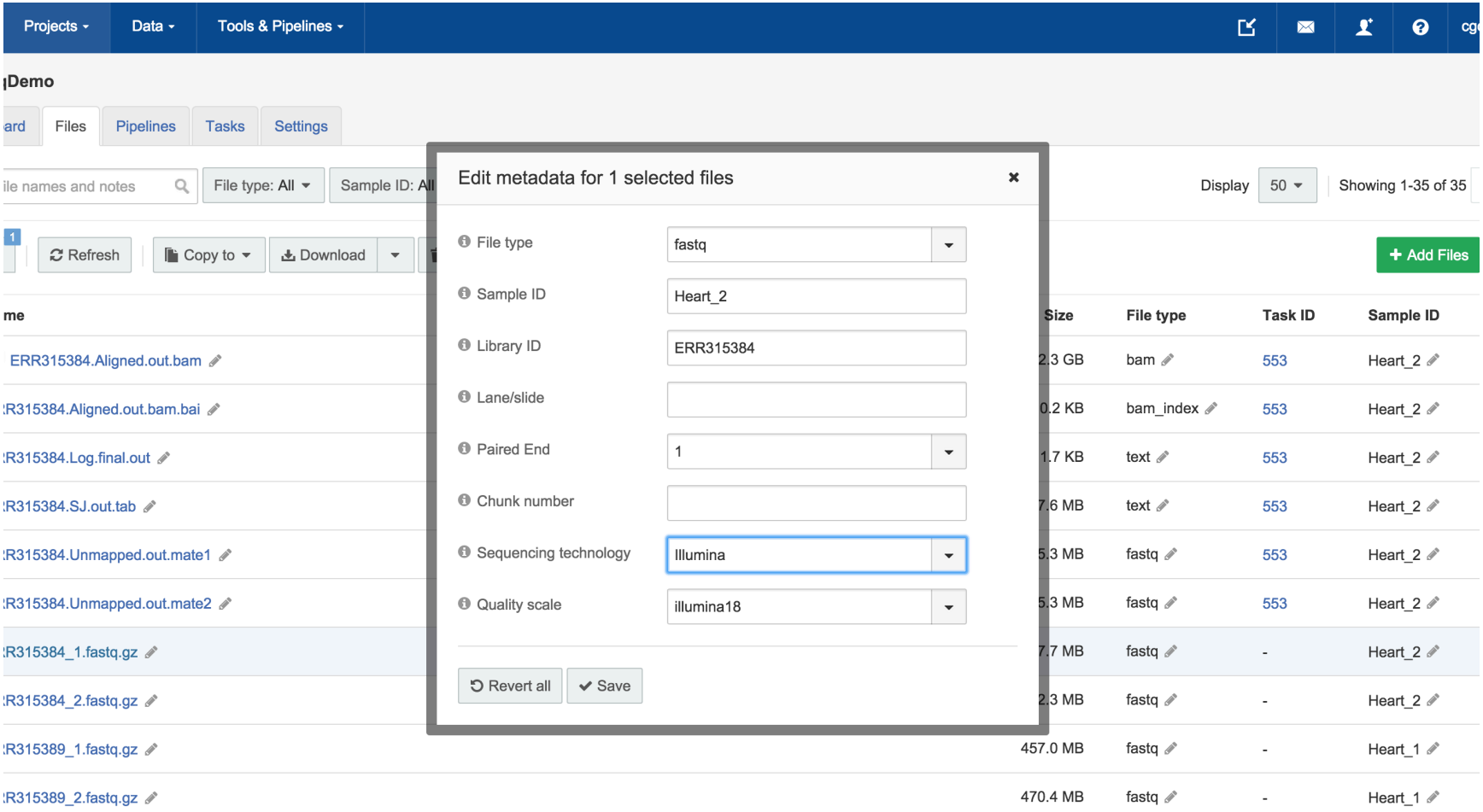


3. Run

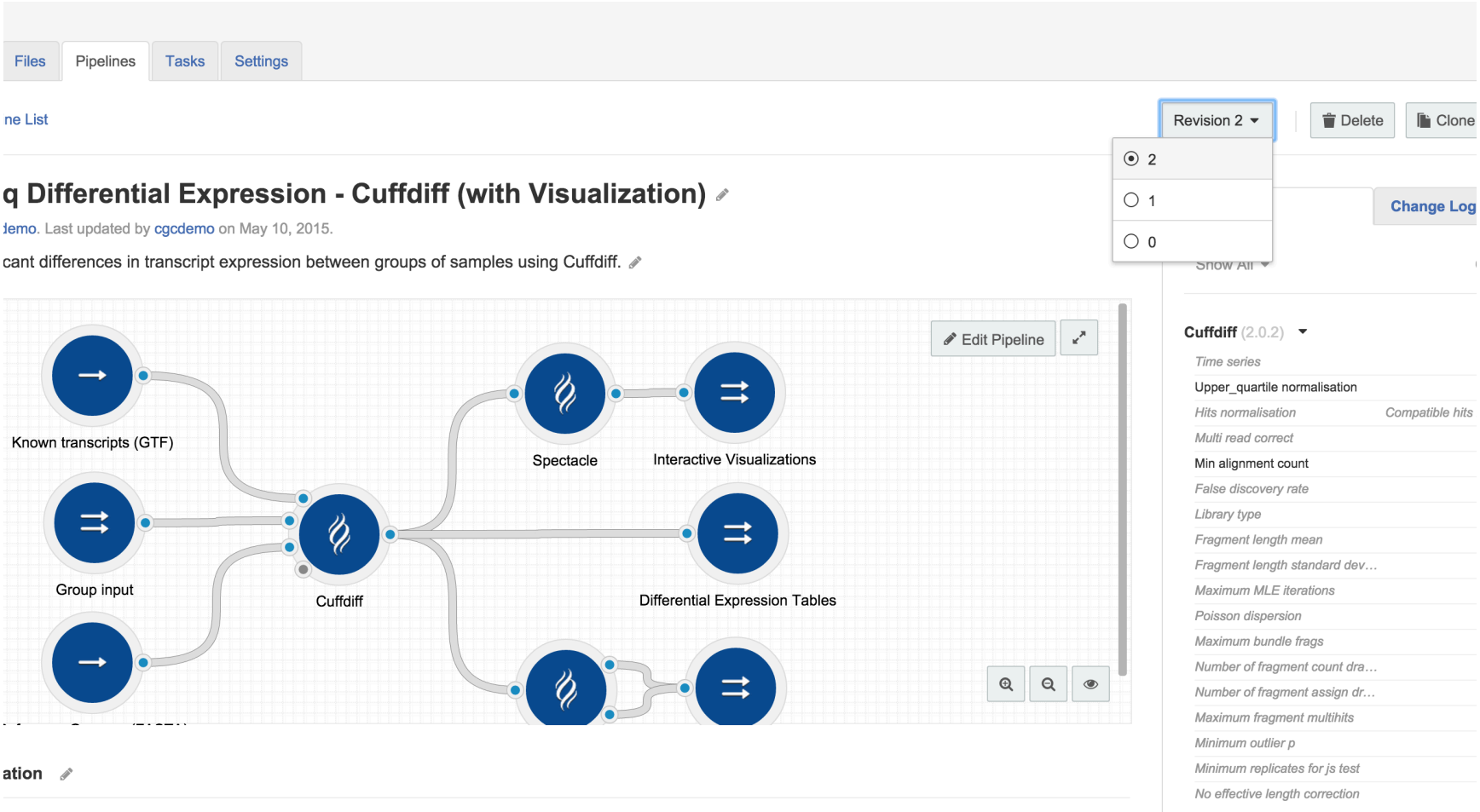
Once the installer has finished, you can start the client from your Applications folder.



Metadata & finding the right datasets



Versioning & Reproducibility of pipelines

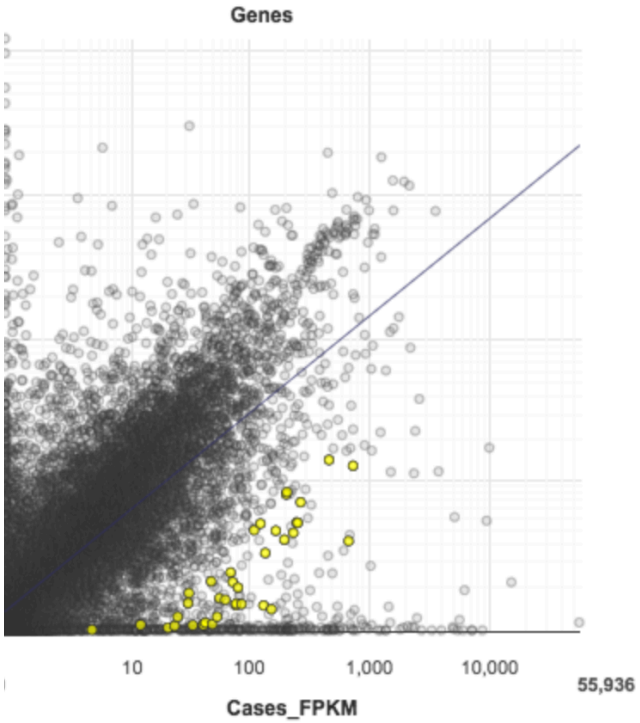


Monitor and optimize computational resource usage.

Visual interface for comparative analyses

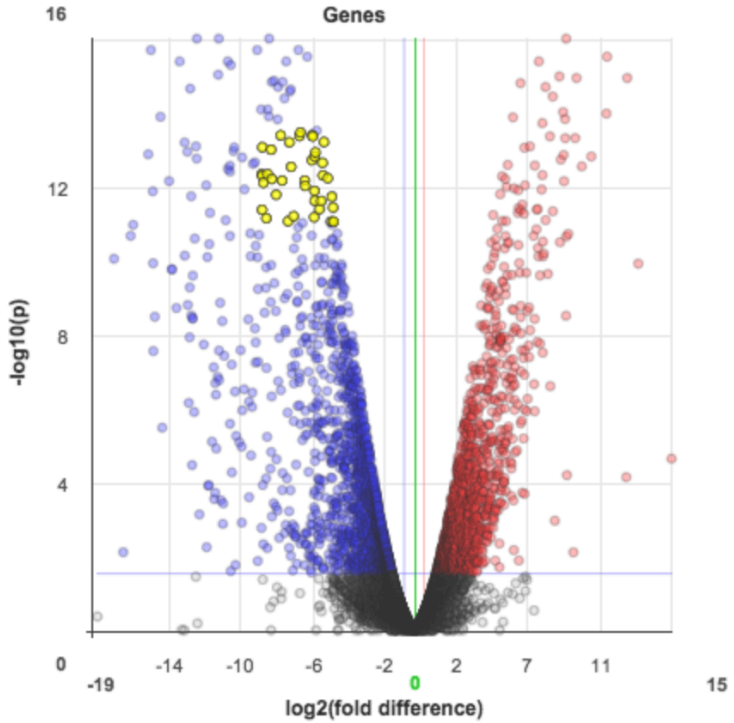
Select Groups

Cases vs Controls



Select Groups

Cases vs Controls



Find Gene

Enter gene name

Gene names of interest

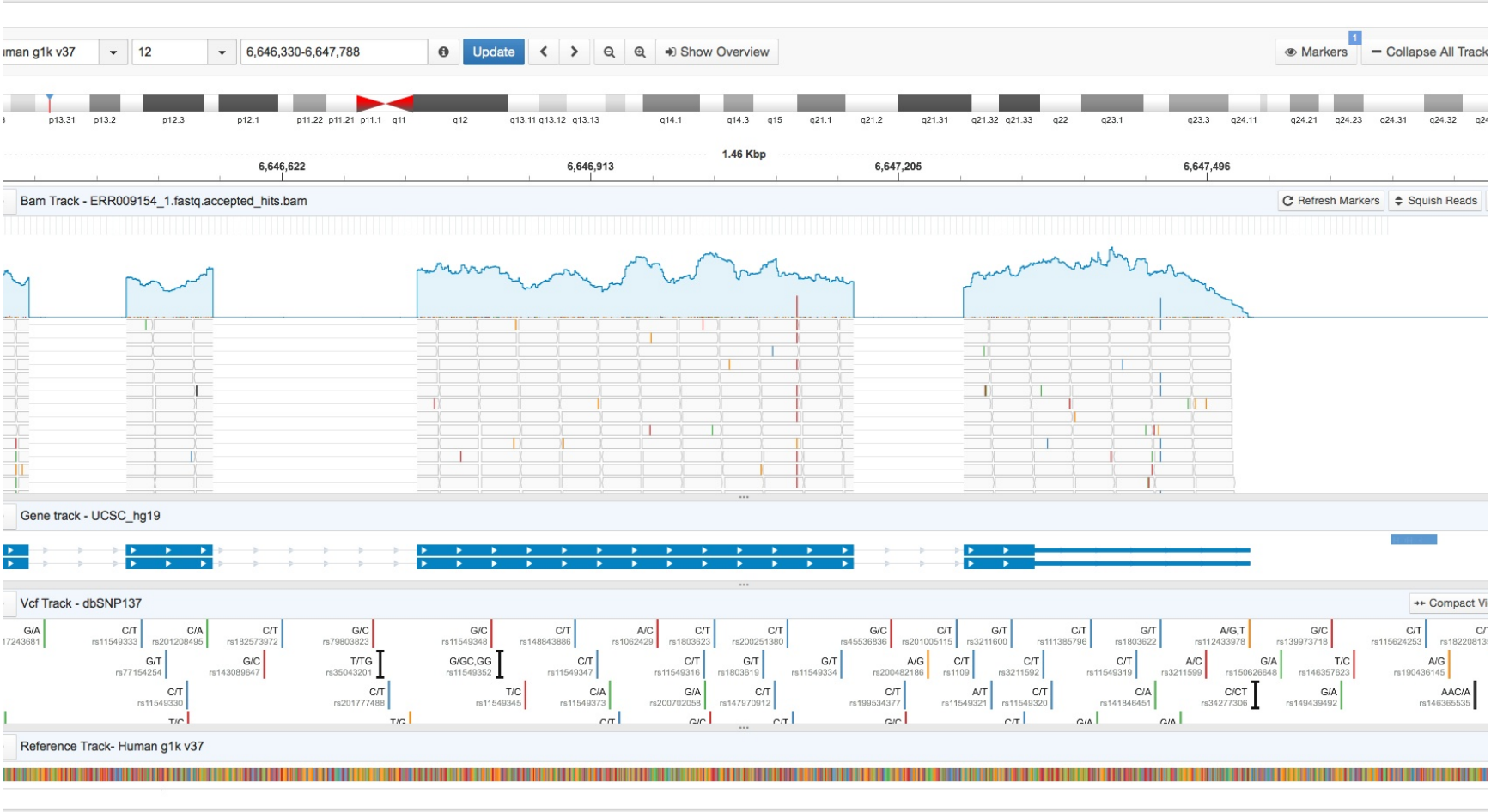
Enter gene names, separat
new line

Selected Genes

- cgn
- onecut2
- amdhd1
- rnf128
- hal
- c2orf72

Export Selected (38)

Integrated Genome Browser with Annotations



Please help us develop the tools useful to you

- www.CancerGenomicsCloud.org
- ~$1M in funding to support computation and private data storage in evaluation period.
- Please share feedback!

- .. Or if you forget www.tcga.ninja