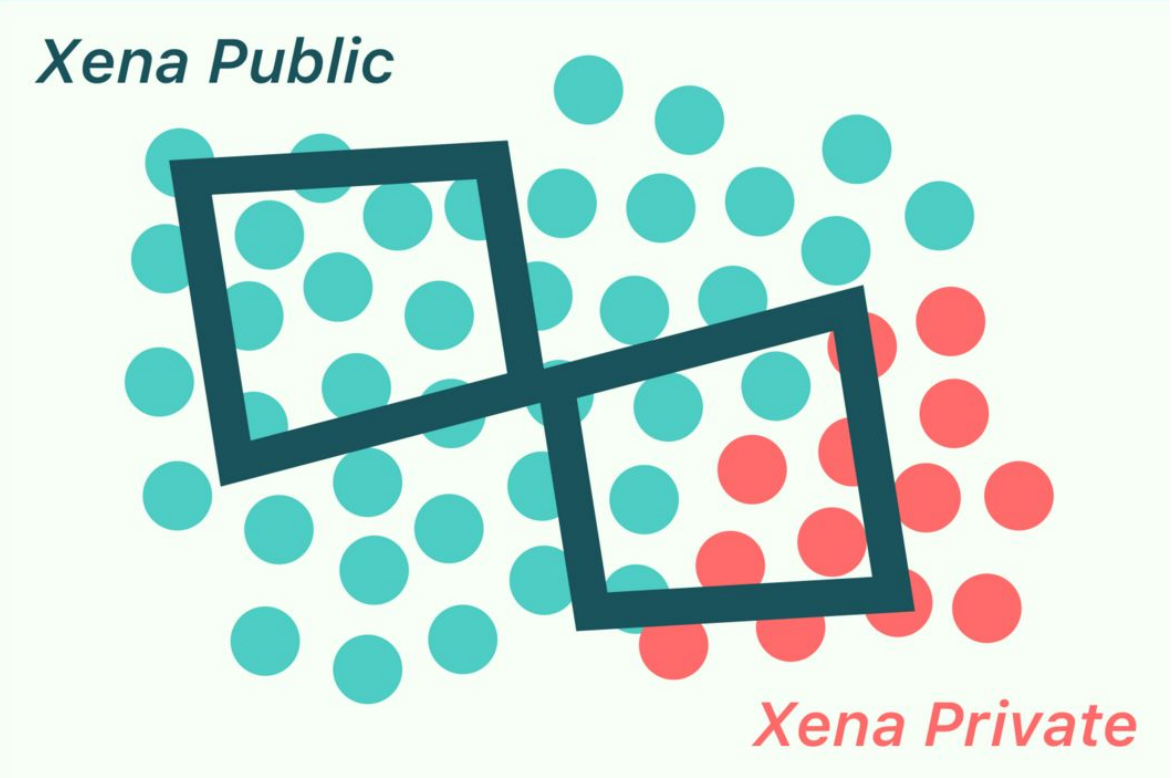
UCSC Xena

See the bigger picture



Jingchun Zhu, David Haussler

University of California Santa Cruz Genomics Institute

ITCR Annual Meeting, June 1, 2017

Outline

Xena overview

New visualizations

New browser features

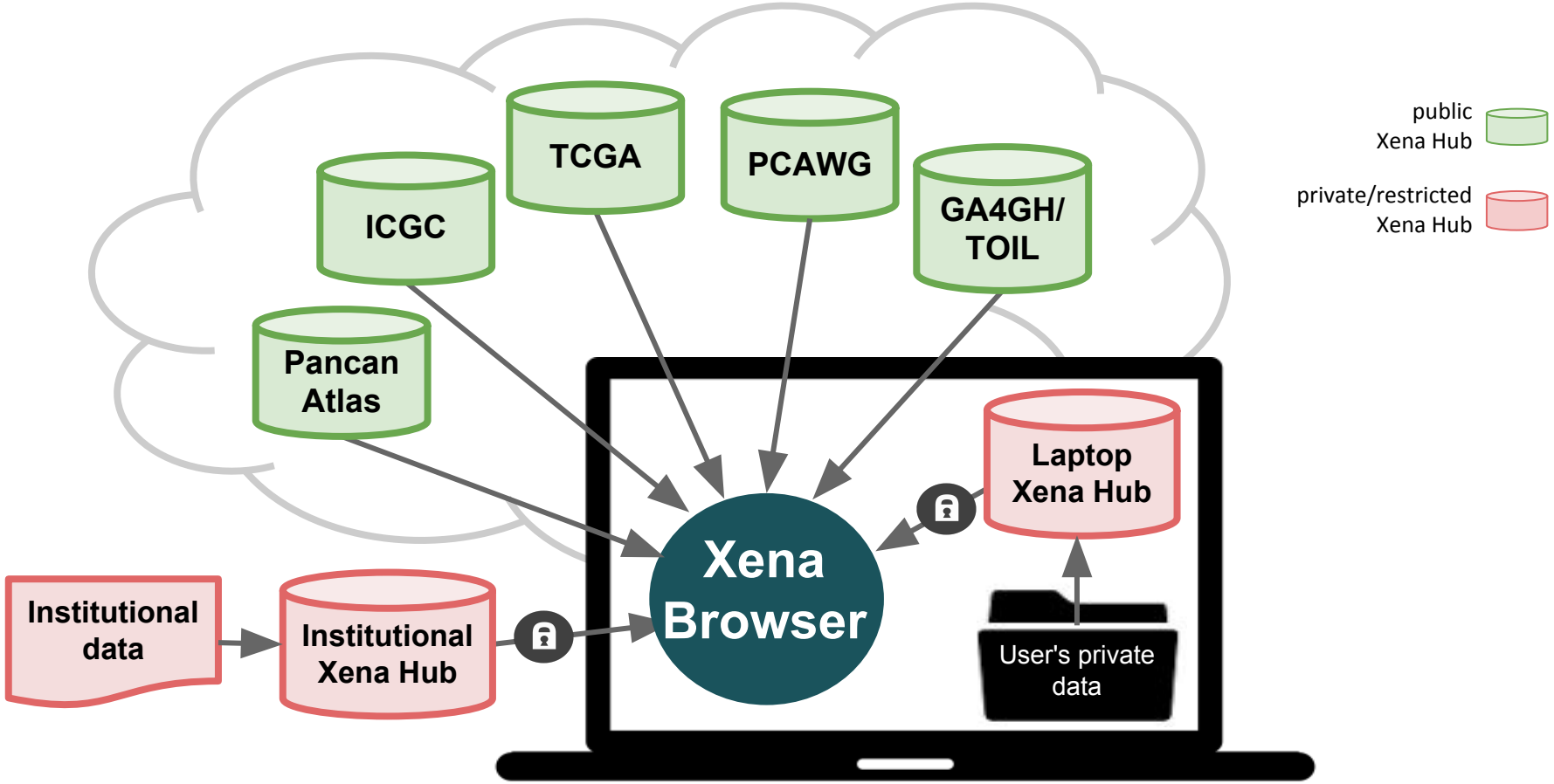
Integration with other tools

Future work

Xena overview

We need a shared and standardized global network of genomic data

Both of large repositories and small studies



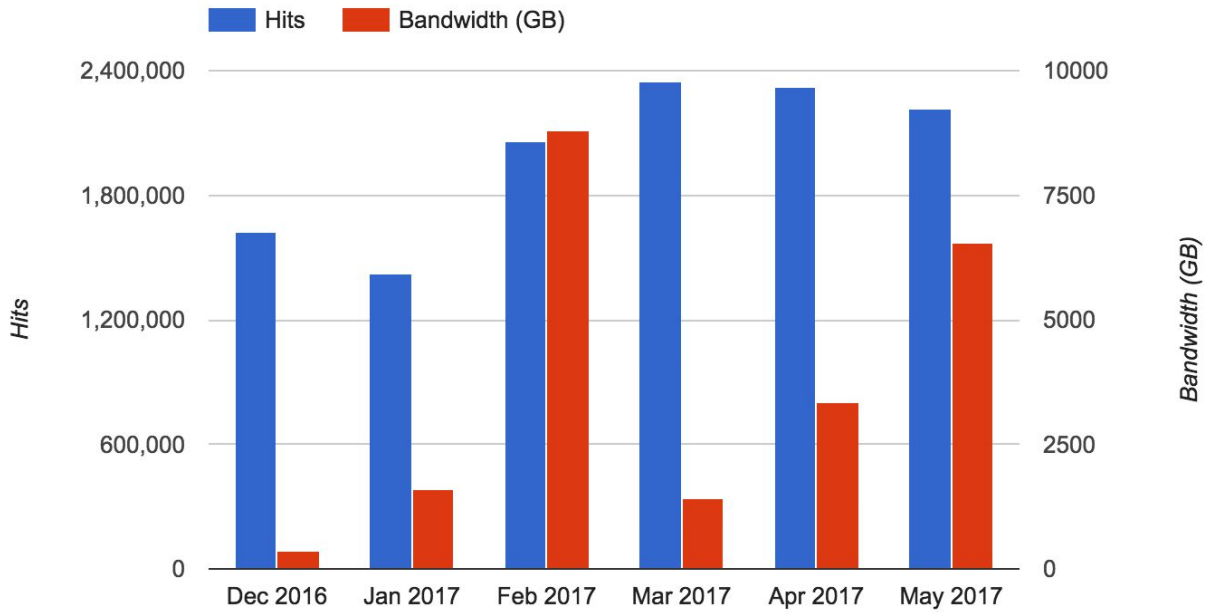
- Federated data hubs
- Data combined in the browser, ensuring data security

Data types Xena visualizes

- SNPs and small INDELs
- Structural variants
- Segmented copy number, gene-level copy number
- Gene-, Transcript-, Exon-, Protein-, and miRNA-expression
- DNA methylation (genes and probes)
- Phenotype, clinical data
- Signature scores, classifications, derived parameters

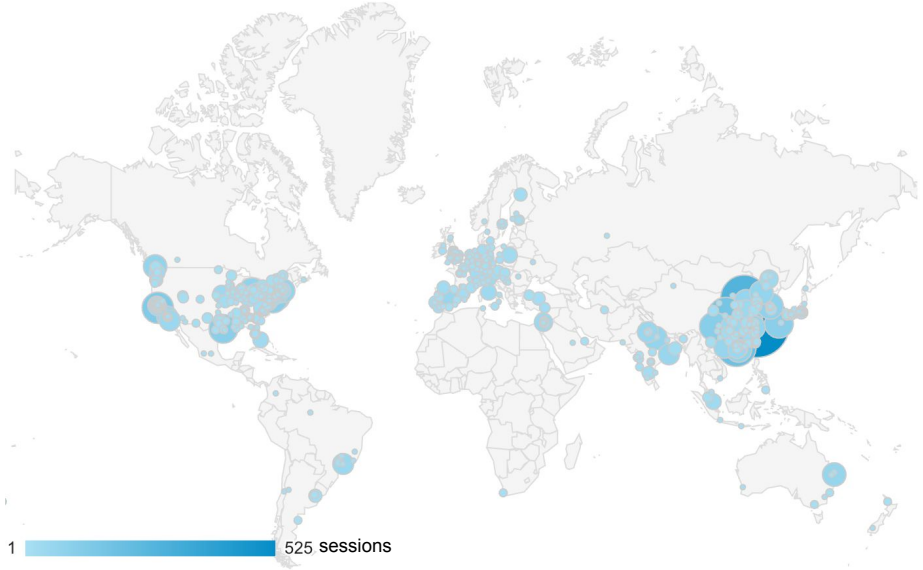
Xena Hub Usage

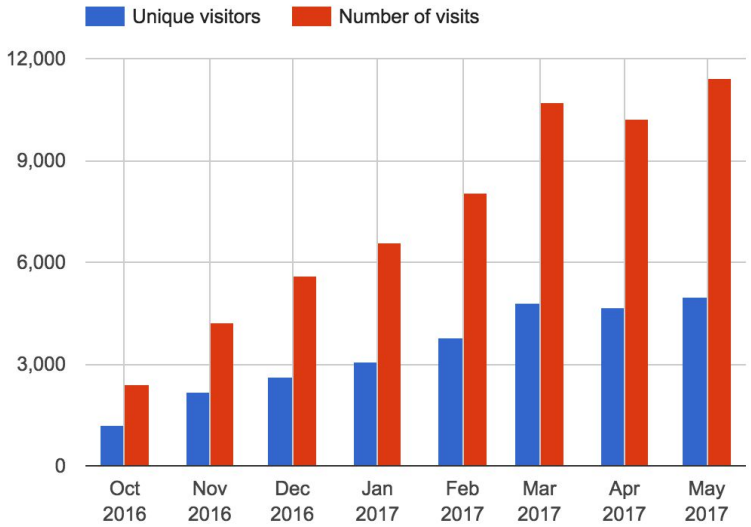
- Millions of hits per month
- TB data visualized/ downloaded
- 430 laptop hubs 'phoned home' in April 2017



Xena Browser usage

- 6,915 sessions in the past month
- Average 06:56 min per session





New Whole Genome Data Visualizations

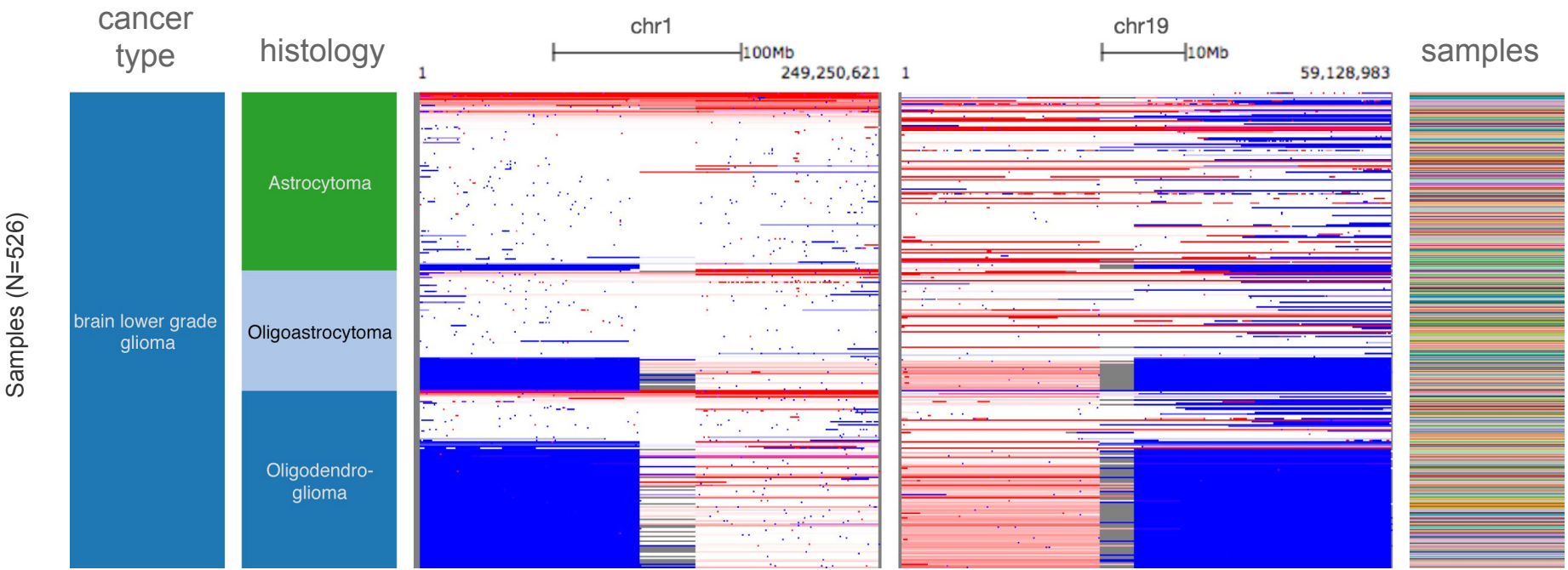
Driven by the PCAWG Project

- Coding & non-coding regions
- Gene- & coordinate-centric views
- Copy number variations
- Simple mutations
- Structural variants

Query examples:

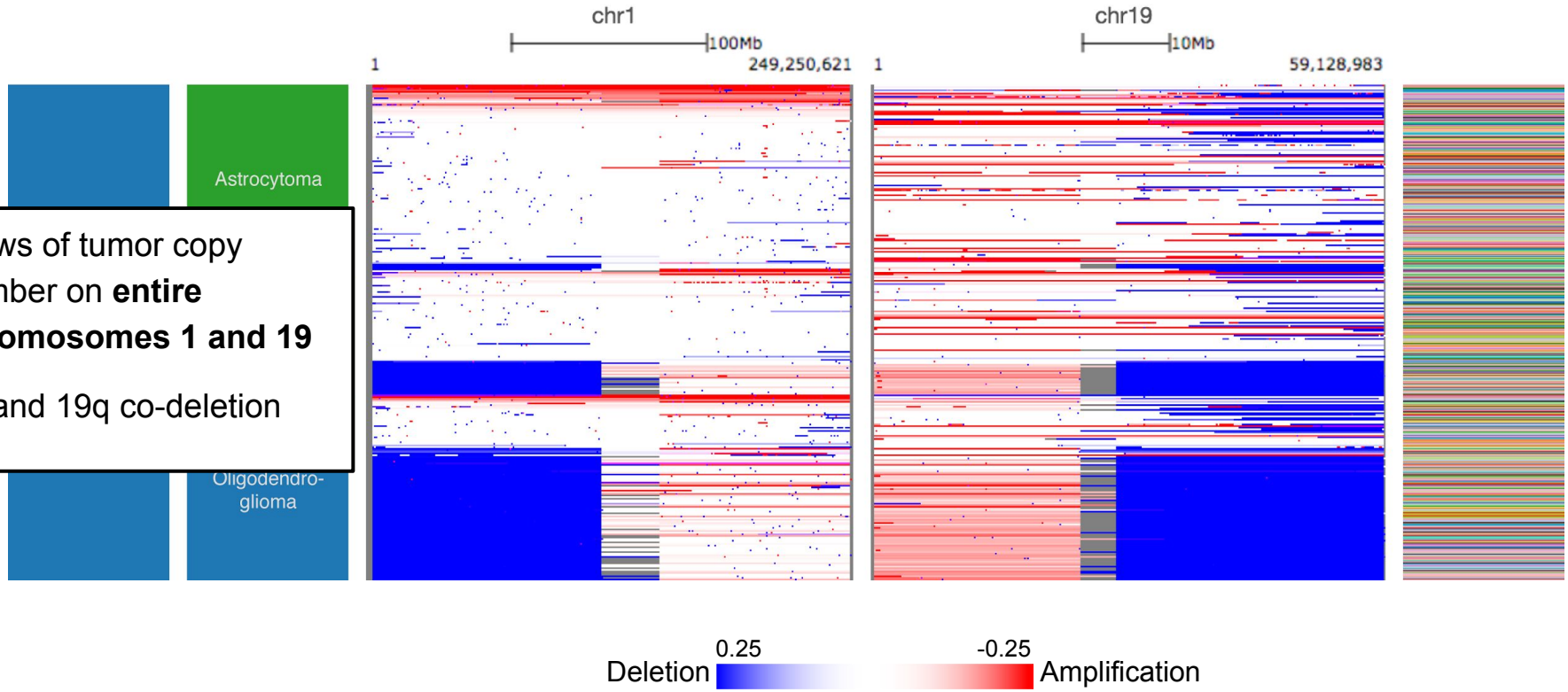
'chr1' 'chr19q' 'chr21:42870119-42870526' 'TP53'

1p/19q Co-deletion in lower grade glioma



https://xenabrowser.net/heatmap/?bookmark=b34336d5c5cff6583f2f617aea022d10

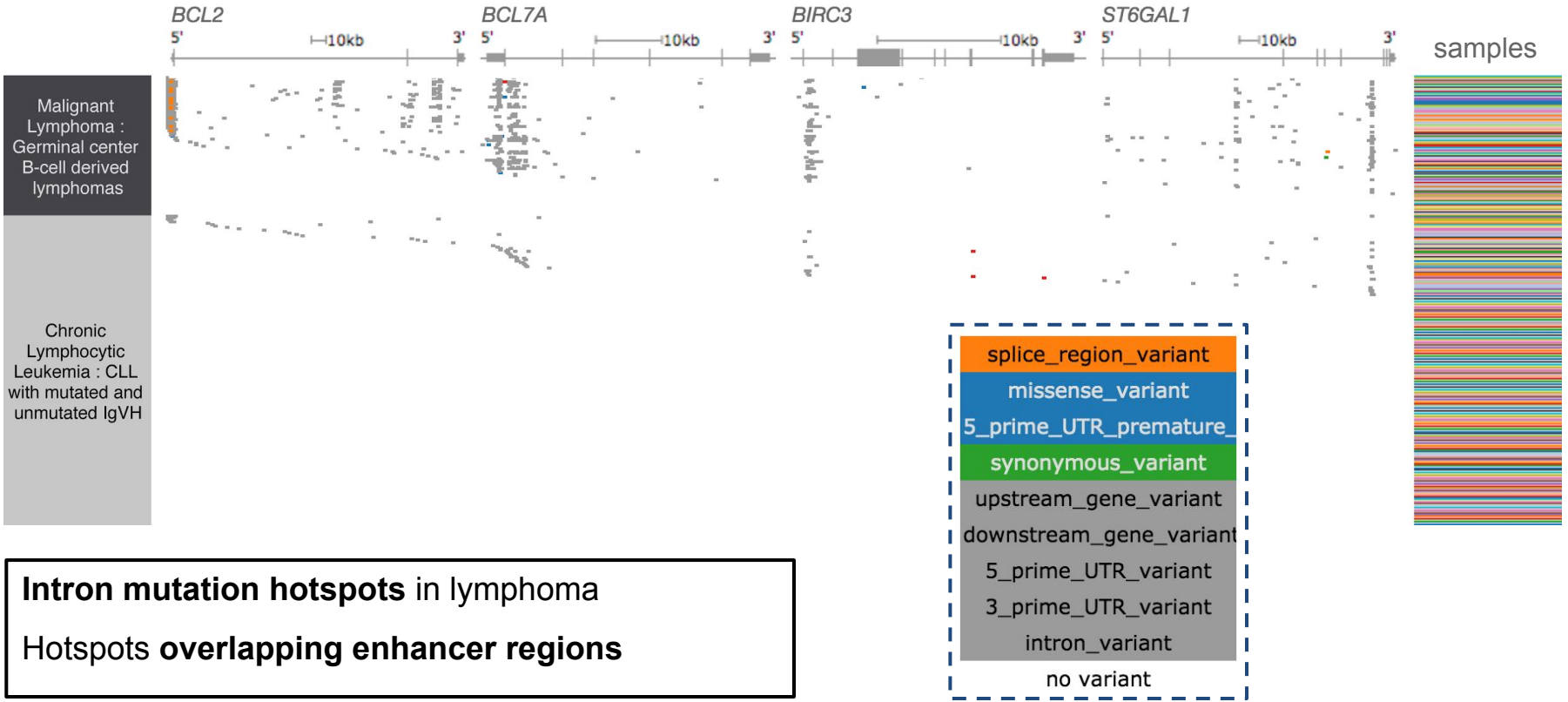
1p/19q Co-deletion in lower grade glioma



Views of tumor copy number on **entire chromosomes 1 and 19**

1p and 19q co-deletion

Intron Mutations in Lymphoma genes



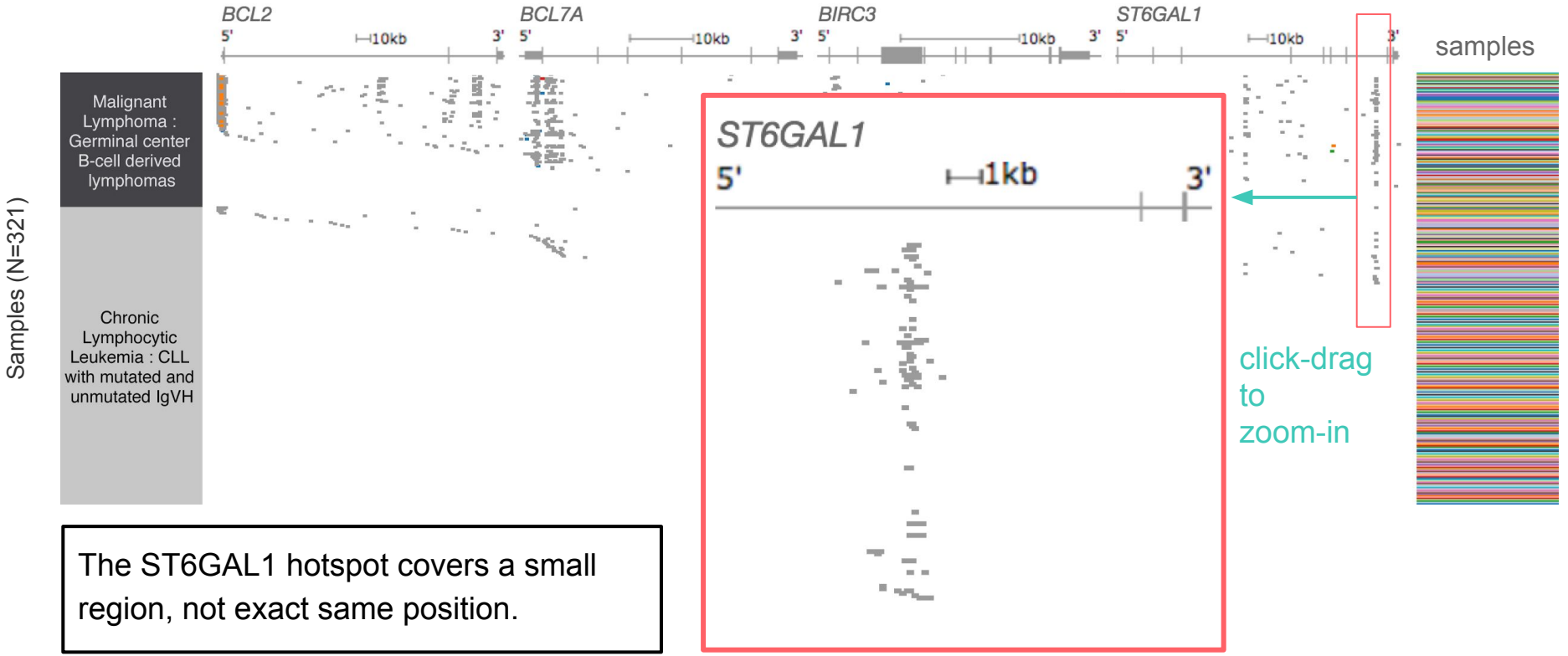
Intron mutation hotspots in lymphoma

Hotspots **overlapping enhancer regions**

Genes from Mathelier A et. al. Genome Biol. 2015

https://xenabrowser.net/heatmap/?bookmark=e9166aa24e55dd6bff3cf907a63fab32

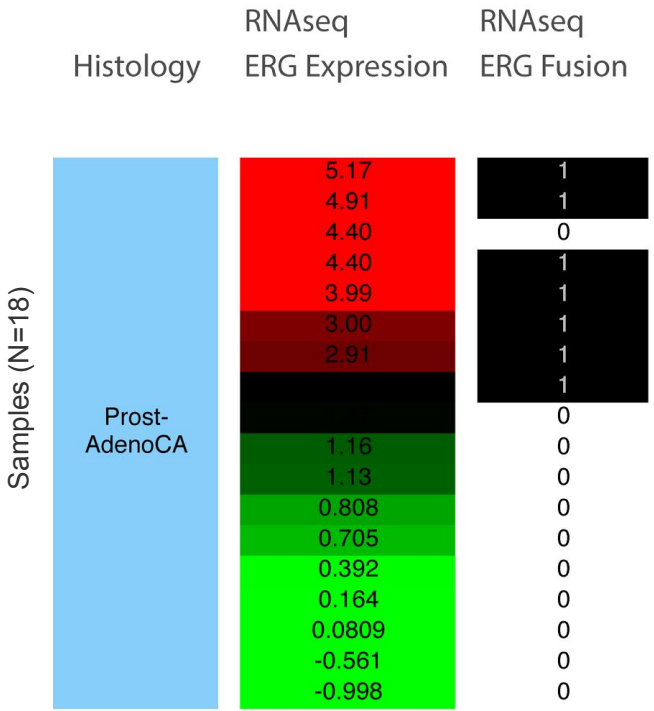
Intron Mutations in Lymphoma genes



The ST6GAL1 hotspot covers a small region, not exact same position.

https://xenabrowser.net/heatmap/?bookmark=b8a268baa9feb53cd15e046f94d940f0

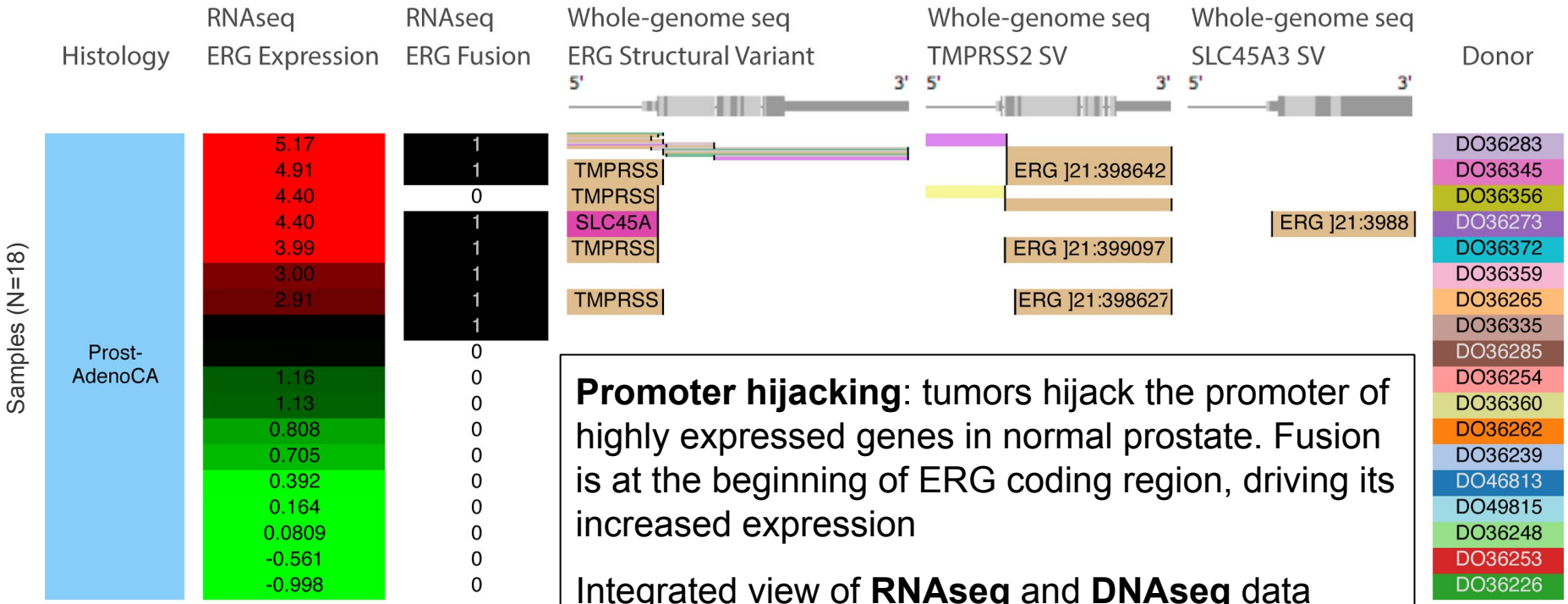
Structural Variant: ERG Fusion in Prostate Cancer



Recurrent fusion-driven **ERG overexpression** drives a subset of prostate cancer

Displaying **RNAseq** analysis results

Structural Variant: ERG Fusion in Prostate Cancer



Promoter hijacking: tumors hijack the promoter of highly expressed genes in normal prostate. Fusion is at the beginning of ERG coding region, driving its increased expression

Integrated view of **RNAseq** and **DNAseq** data

New Browser Features

Search, Filter and Group samples

Powerful, text-based search interface

- Simple search:
  - 'Make two groups of samples: those that have a ATRX missense mutation and those that do not' => 'missense'
- Complicated search:
  - 'Show me all samples that are stage III or IV *with* MYC amplification or KRAS mutation' => *‘(stage III OR stage IV) AND (B:>0.5 OR KRAS)’*

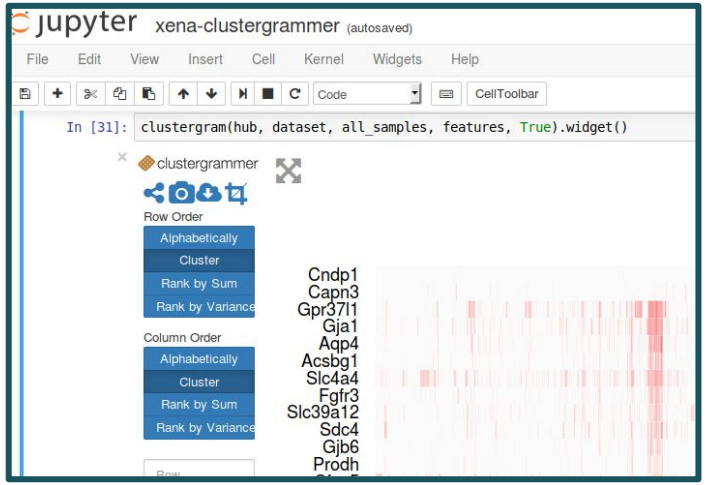
https://xenabrowser.net/heatmap/?bookmark=8363fde2c753009079a862fd41479244

Other New Browser Features

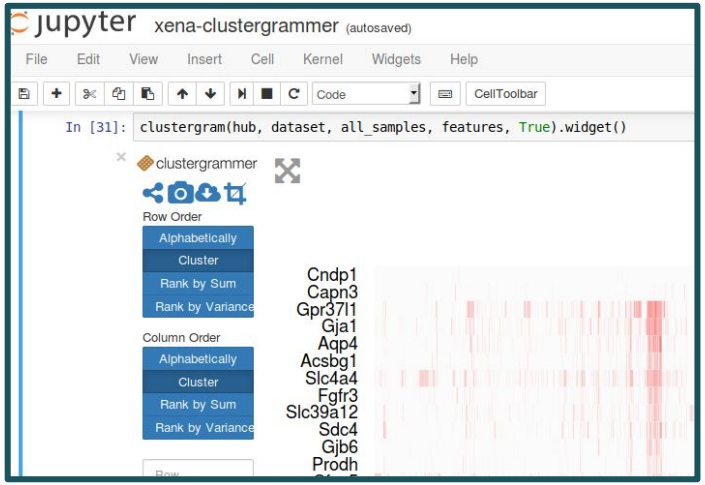
- Adjustable KM plot time axis: e.g. 1, 3, 5 years
- Coloring in linear or log scale
- Statistics for box plot, scatter plot and KM plot
- Enhanced PDF
- Download the entire spreadsheet at once
- Bookmarks

Integration with other tools

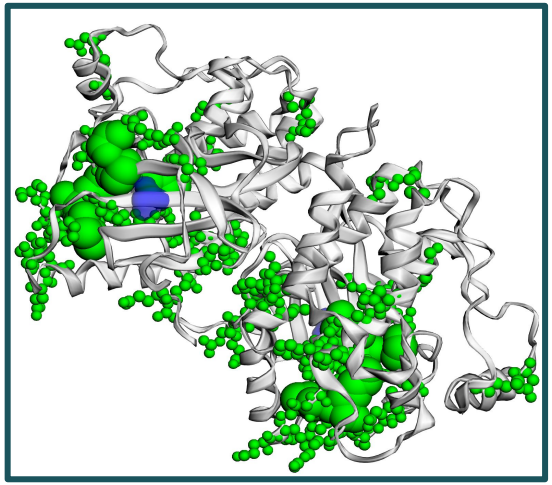
Jupyter Notebook/Python API



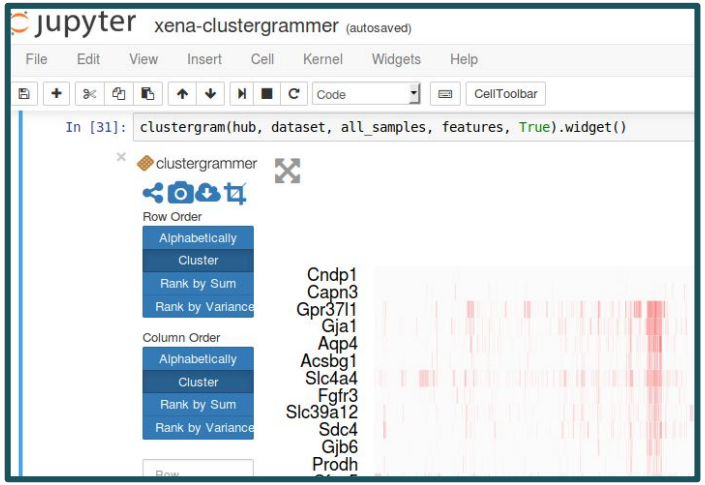
Jupyter Notebook/Python API



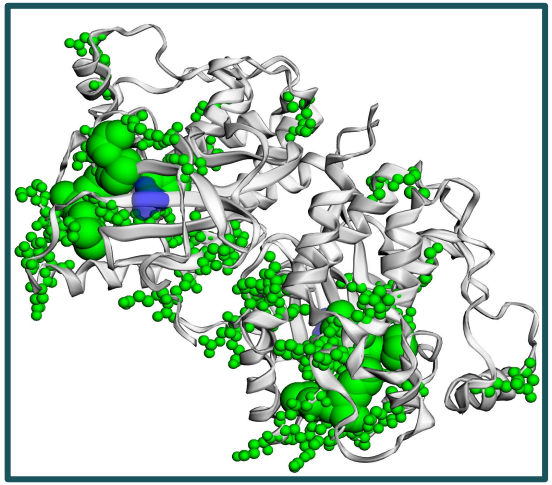
MuPIT



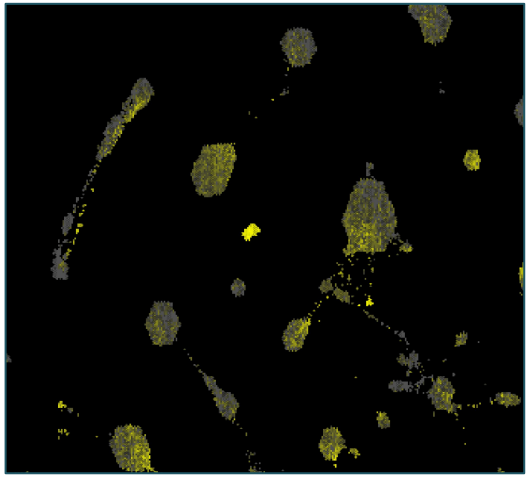
Jupyter Notebook/Python API



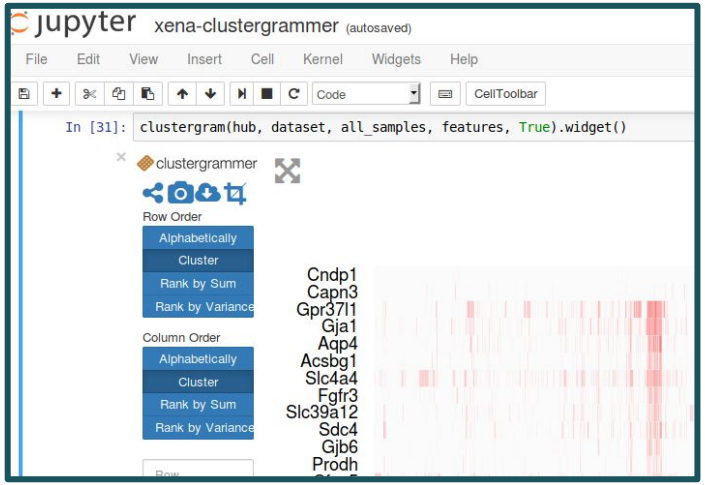
MuPIT



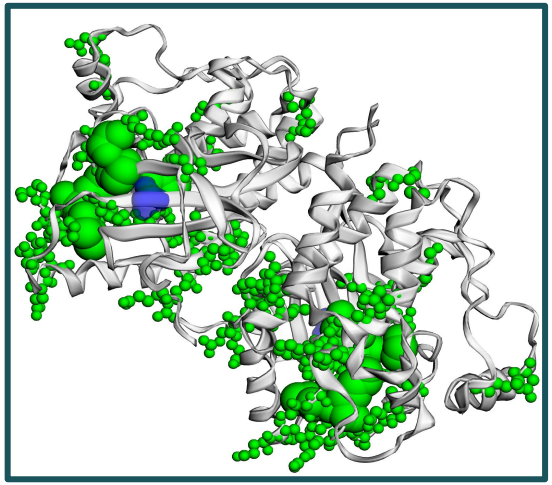
TumorMap



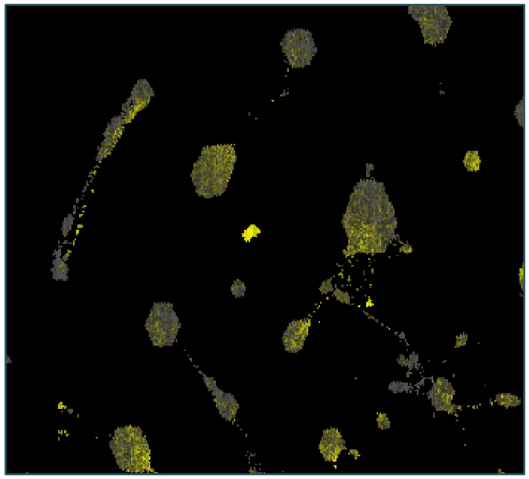
Jupyter Notebook/Python API



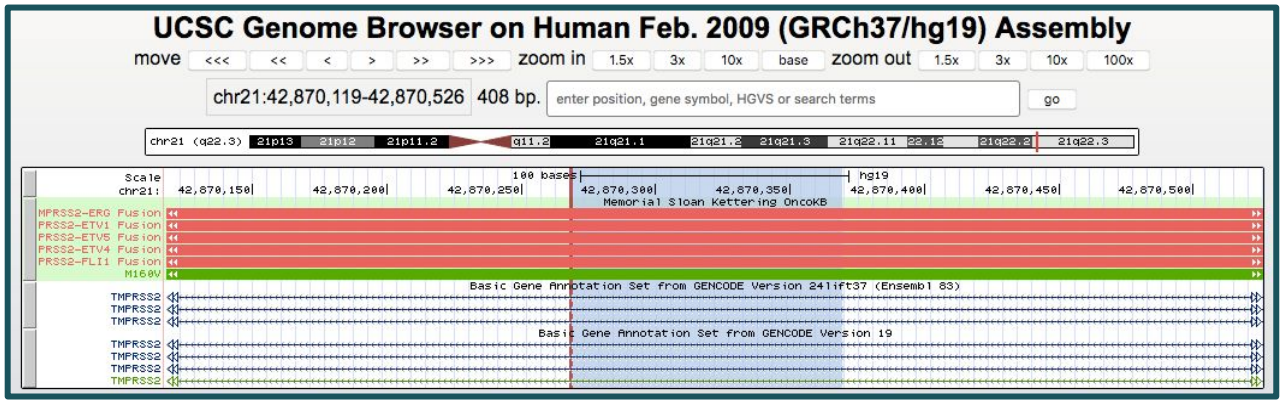
MuPIT



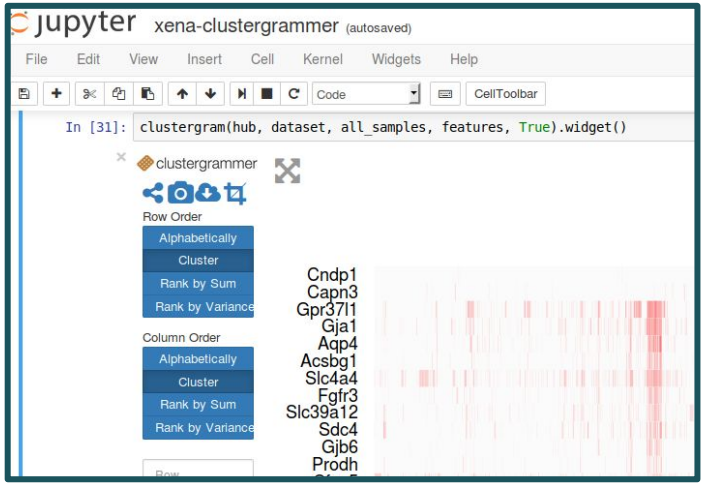
TumorMap



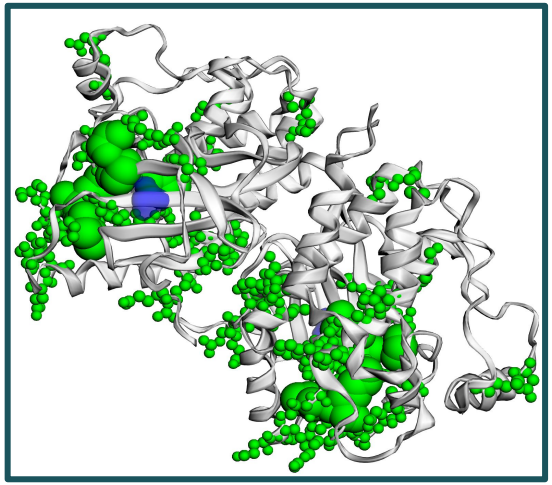
UCSC Genome Browser



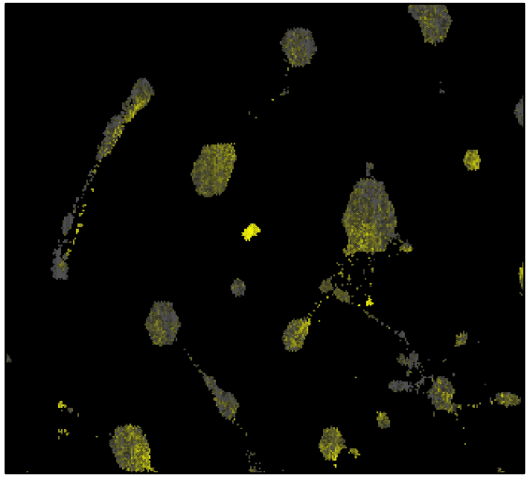
Jupyter Notebook/Python API



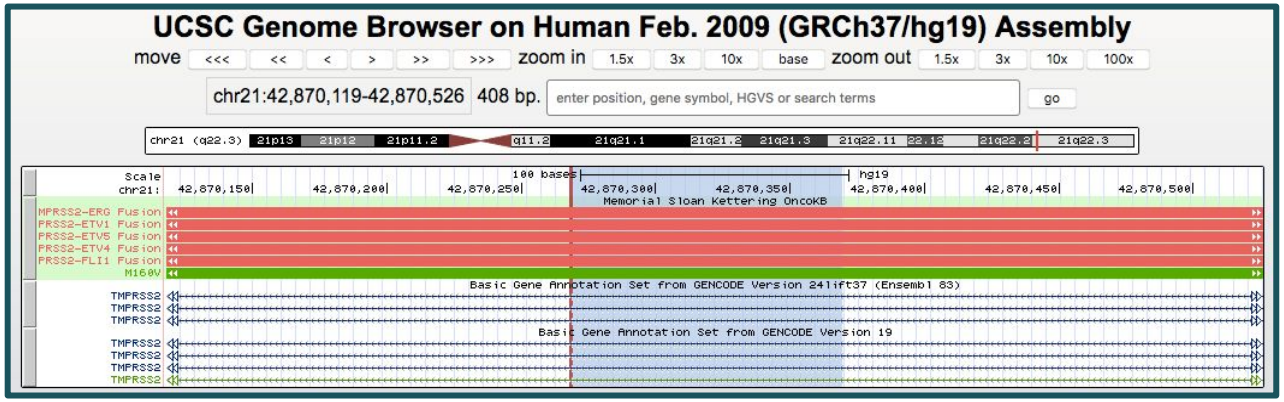
MuPIT



TumorMap



UCSC Genome Browser

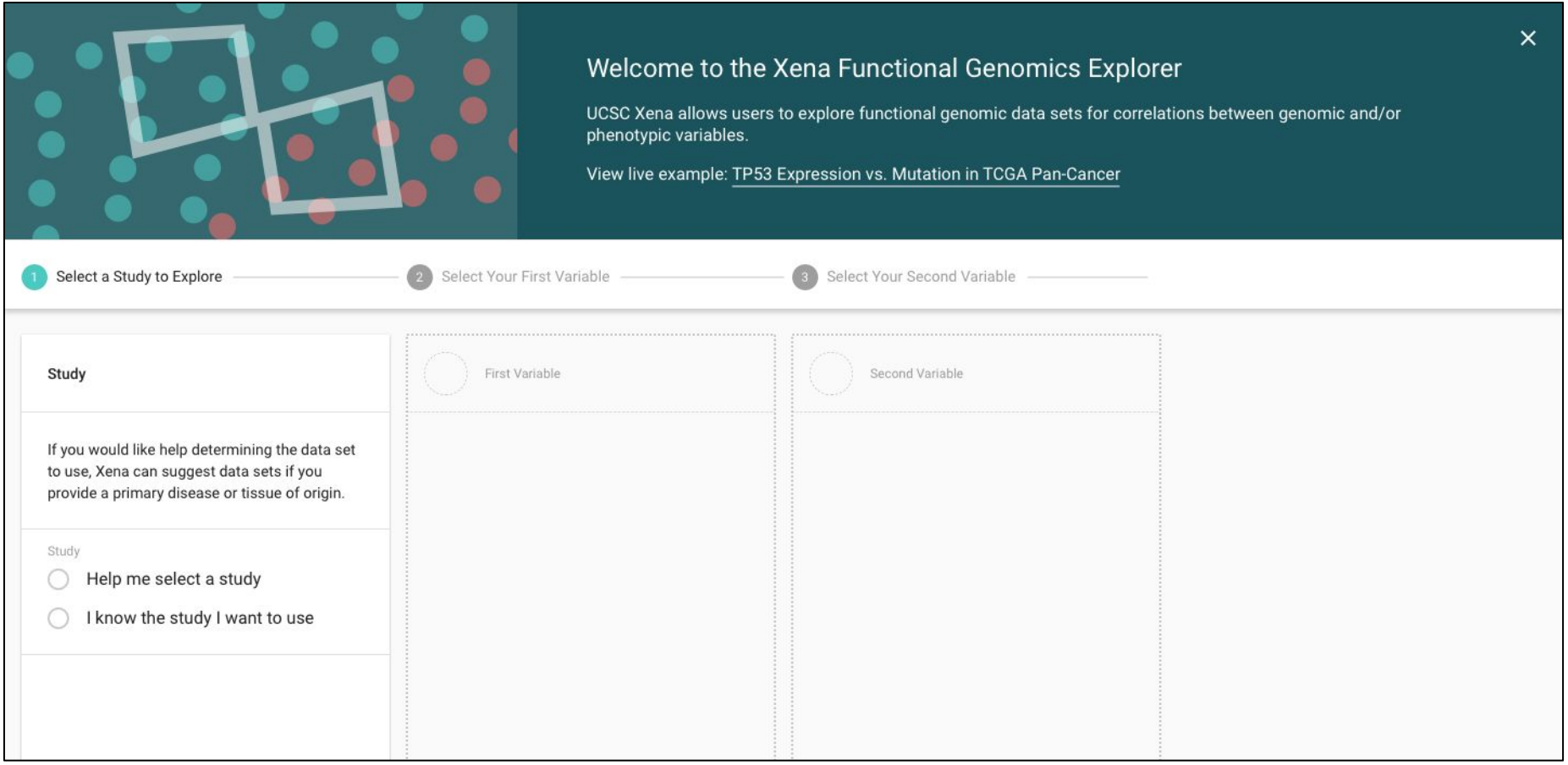


Genomic Data Commons

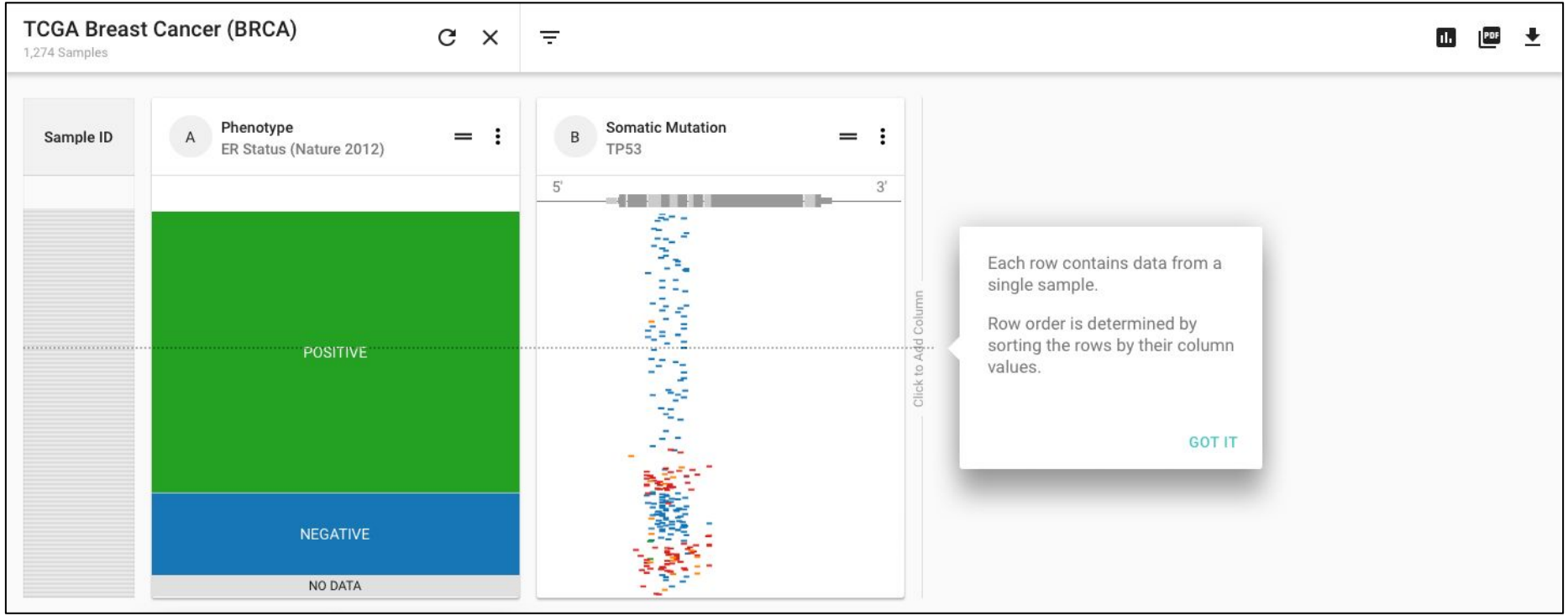
data coming soon!

In-progress ...
“Building a Spreadsheet” Wizard

Welcome screen / select study

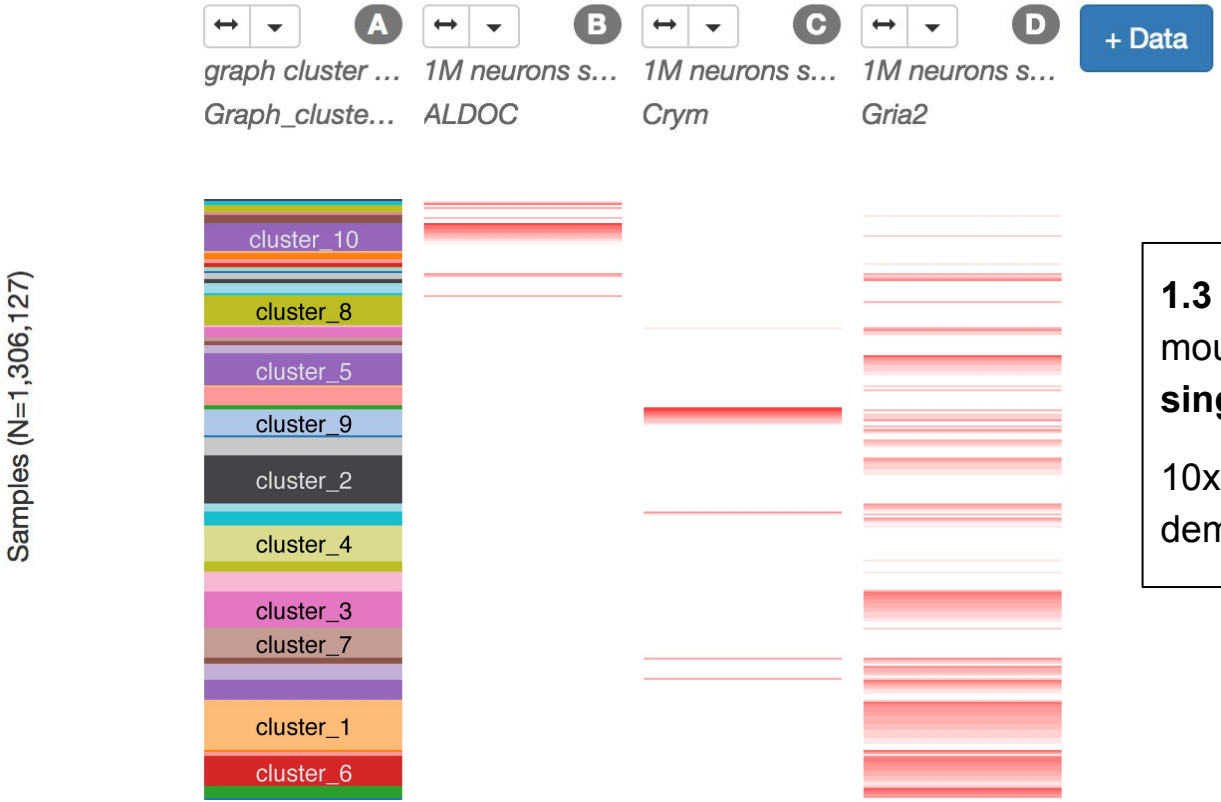


Displaying two variables / Done



New adventure …
Single Cell RNAseq Visualization

1.3 million mouse brain cells



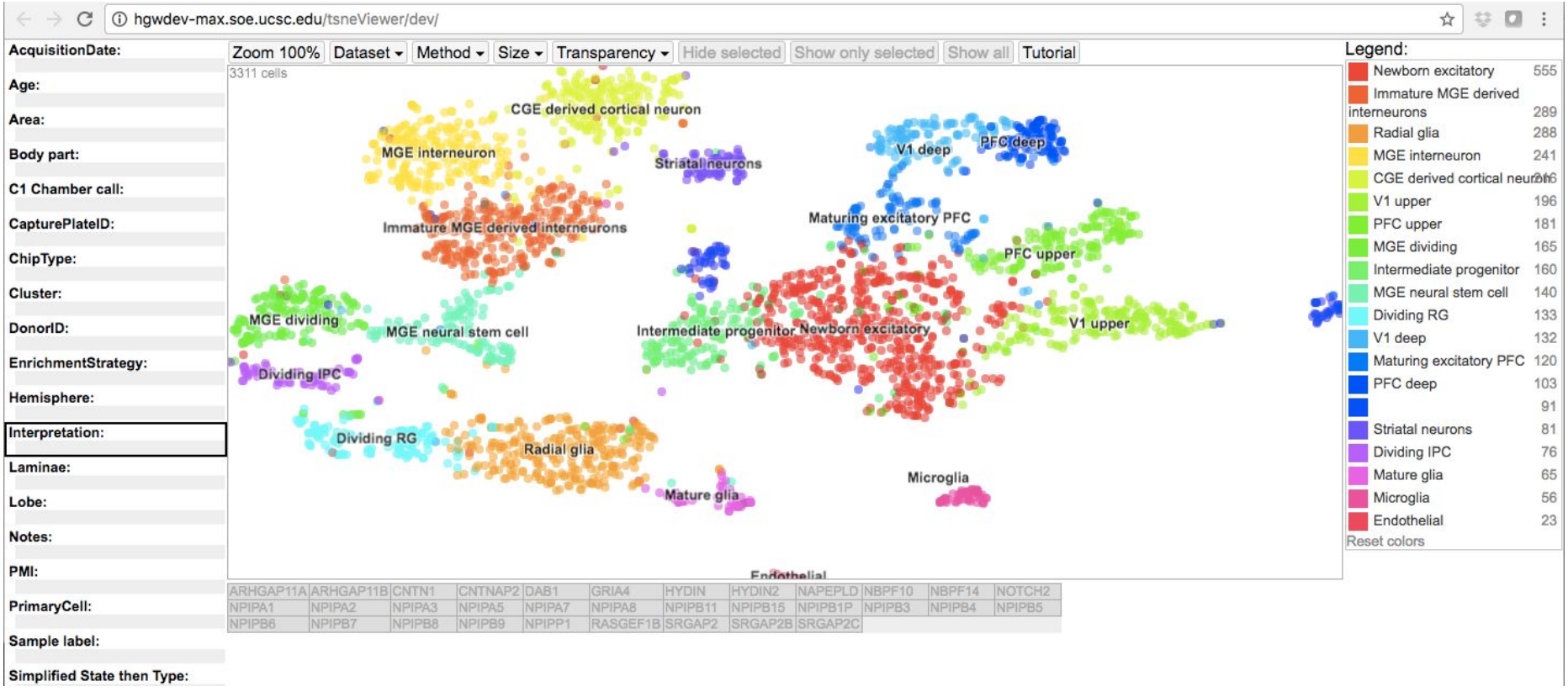
1.3 million embryonic mouse brain cells **single-cell RNAseq**

10x genomics demonstration dataset

Stem Cell Genomics

Kent, Haussler, Salama, Kriegstein, Pollen, Nowakowski

CIRM
CALIFORNIA'S STEM CELL AGENCY



Funding

National Cancer Institute
Amazon Web Services

Collaborators

TOIL Team
UCSC Genomics Core
UCSC Genome Browser
PCAWG consortium
ITOMIC trial
Treehouse
Angela Brooks's Lab
Eric Collison
Tumor Map - Stuart Lab
MuPIT/CRAVAT
BioJS

UNIVERSITY OF CALIFORNIA SANTA CRUZ Genomics Institute

PCAWG PanCancer Analysis OF WHOLE GENOMES

Global Alliance for Genomics & Health

BIOJS

NATIONAL CANCER INSTITUTE
Informatics Technology for Cancer Research

http://xena.ucsc.edu

https://github.com/ucscXena

Jing Zhu jzhu@soe.ucsc.edu

Mary Goldman mary@soe.ucsc.edu

Brian Craft craft@soe.ucsc.edu