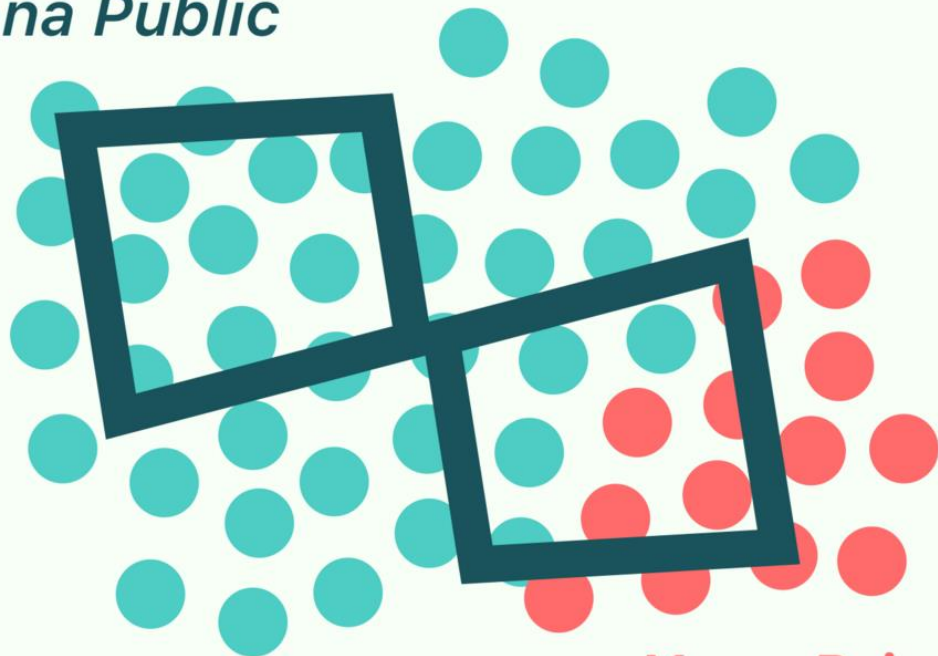


UCSC Xena

See the
bigger
picture

Xena Public



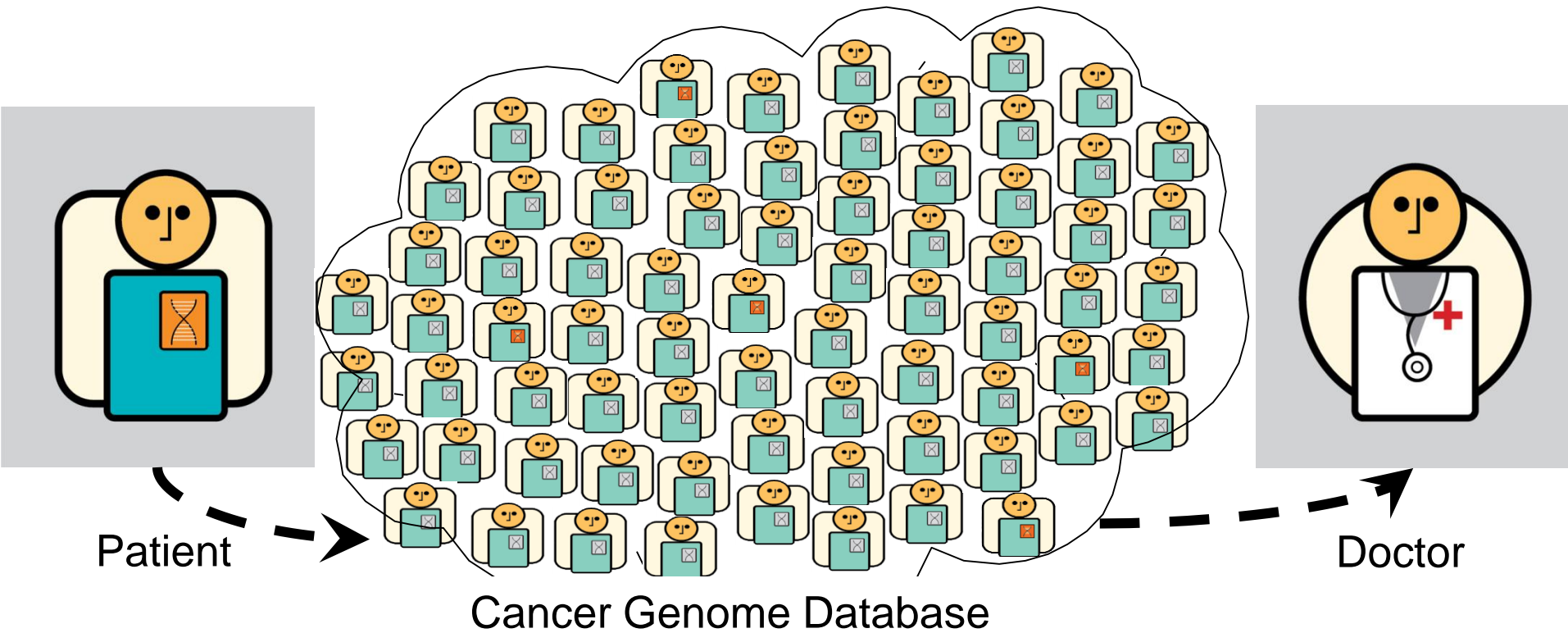
Xena Private

Jingchun Zhu and David Haussler

Genomics Institute, University of California Santa Cruz

ITCR annual meeting, June 13, 2016

Genomes are the key to the future of cancer treatment



Future requires a shared and standardized global network of data



Most are small studies with a few samples

Lab 1 5 cell lines before and after exposure to compound X

Lab 2 8 ER+ breast cancer patients who responded poorly to hormone therapy

...

Lab 5000 10 methylation profiles of patients with mutations in chromatin remodelers

Some important large datasets ...



20,000 adult tumors (U.S.)



17,000 donors (international)



2,800 whole genome tumors



>6,000 pediatric tumors (U.S.)



10,000 benign samples

Xena Architecture Overview

Federated Data Hubs

Distributed system to host cancer genomics and clinical/phenotype data
**SNPs, INDELs, large structural variants, CNV, gene-, transcript-, exon- protein-,
miRNA-expression, DNA methylation**

Xena Architecture Overview

Federated Data Hubs

Distributed system to host cancer genomics and clinical/phenotype data
**SNPs, INDELs, large structural variants, CNV, gene-, transcript-, exon- protein-,
miRNA-expression, DNA methylation**

Private Data Hubs

Easy to install on your laptop or
behind a firewall. Load your data
using our app or the command
line. **Mac, Windows and Linux.**
Can restrict access if desired

Xena Architecture Overview

Federated Data Hubs

Distributed system to host cancer genomics and clinical/phenotype data
**SNPs, INDELs, large structural variants, CNV, gene-, transcript-, exon- protein-,
miRNA-expression, DNA methylation**

Private Data Hubs

Easy to install on your laptop or
behind a firewall. Load your data
using our app or the command
line. **Mac, Windows and Linux.**
Can restrict access if desired

Public Data Hubs

Large datasets we already have



THE CANCER GENOME ATLAS

National Cancer Institute
National Human Genome Research Institute

20,000 adult tumors (U.S.)



17,000 donors (international)



PCA WG
PanCancer Analysis
OF WHOLE GENOMES

2,800 whole genome tumors



>6,000 pediatric tumors (U.S.)



GTEx

10,000 benign samples

Xena Architecture Overview

Federated Data Hubs

Distributed system to host cancer genomics and clinical/phenotype data

Private Data Hubs

Your restricted-access data on your laptop or behind a firewall

Public Data Hubs

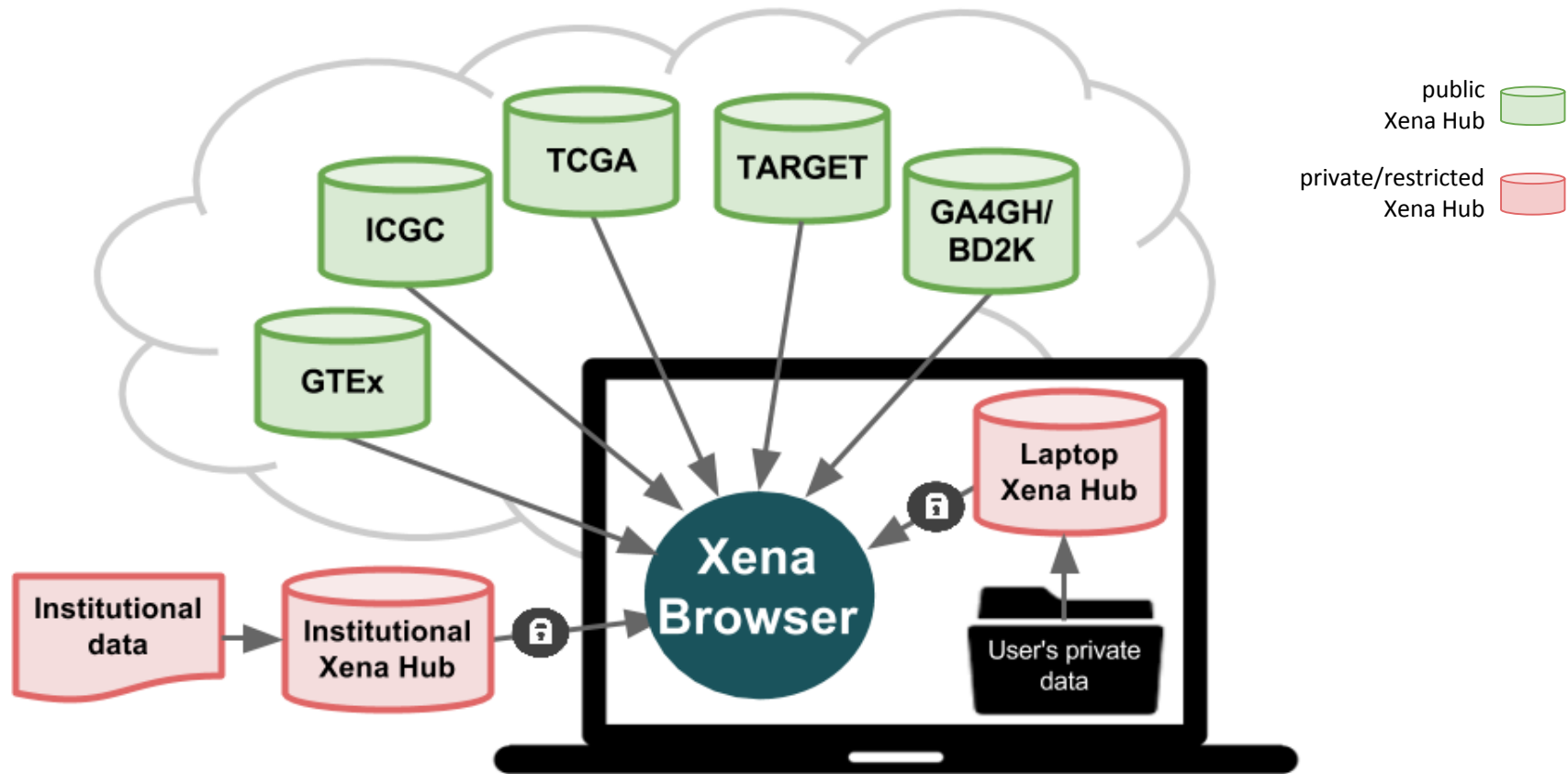
1400+ datasets from more than 35 cancer types, including Pan-Cancer.



The diagram shows two grey arrows pointing downwards from the 'Private Data Hubs' and 'Public Data Hubs' sections towards the 'Xena Browser' section, indicating that both types of data are accessed through the browser.

Xena Browser

Dynamically visualize and analyze cancer genomics data from this federated hub system together in your web browser. Secure, fast, and powerful. **Chrome, Firefox, and Safari.**



Data combined in the browser, ensuring data security

Collaborations – Use cases

- GA4GH / BD2K RNAseq recompute
 - Center for Big Data in Translational Genomics
- PCAWG Pan-cancer analysis of 2800 whole genomes
- Clinical Trial Use Cases
 - CKCC (California Kids Cancer Comparison)
 - ITOMIC (triple negative metastatic breast cancer trial)

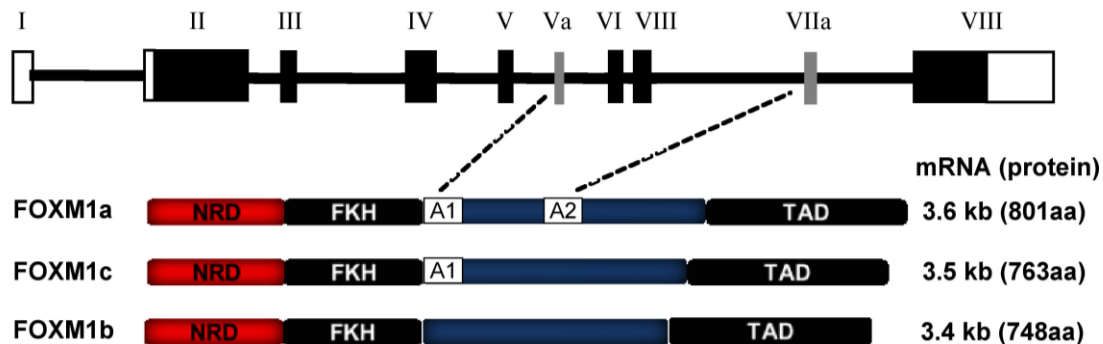
GA4GH BD2K RNASeq Recompute

- Uniformly analyzed RNAseq expression of 60,000 genes and 200,000 transcripts on 20,000 pan-cancer and pan-normal samples: TCGA, TARGET and GTEx
- Performed by TOIL (pipeline architecture for genomic workflows)
 - Fast, cheap (AWS spot market)
 - Add your data! Pipeline is open-source
- Xena hosts all recompute results on GA4GH-BD2K hub for visualization, analysis and download



FOXM1 and its isoforms

FOXM1 is a transcription factor involved in cell cycle progression



FOXM1a → transcriptional repressor

FOXM1b & *1c* → transcriptional activators

→ Use TOIL transcript-level data to **dynamically** view isoform expression across TCGA, TARGET and GTEx (~19,000 samples)

Pan-Cancer *FOXM1* Isoform Expression



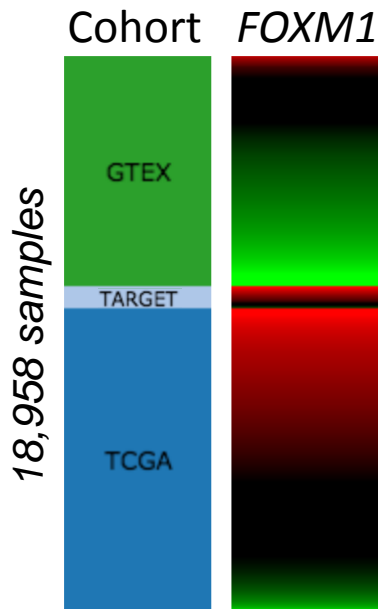
Visual Spreadsheet view

Each row is a single sample, like a spreadsheet

Data sorted left to right, any arbitrary user selected data

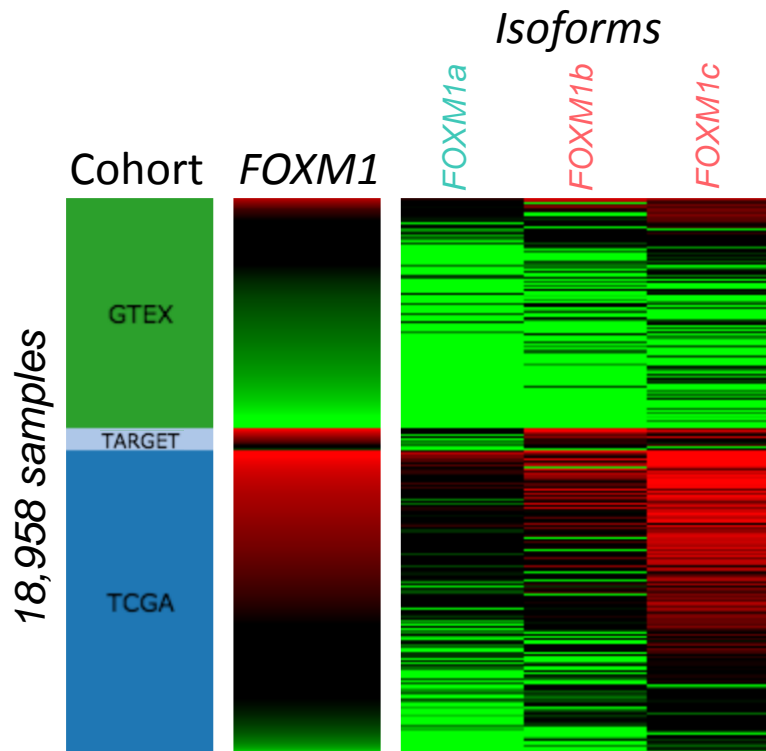
Data from multiple data hubs (public, private, in AWS, on laptop), combined in the browser

Pan-Cancer *FOXM1* Isoform Expression



Overall FOXM1 expression is higher in TCGA and TARGET tumors, and lower in normal samples.

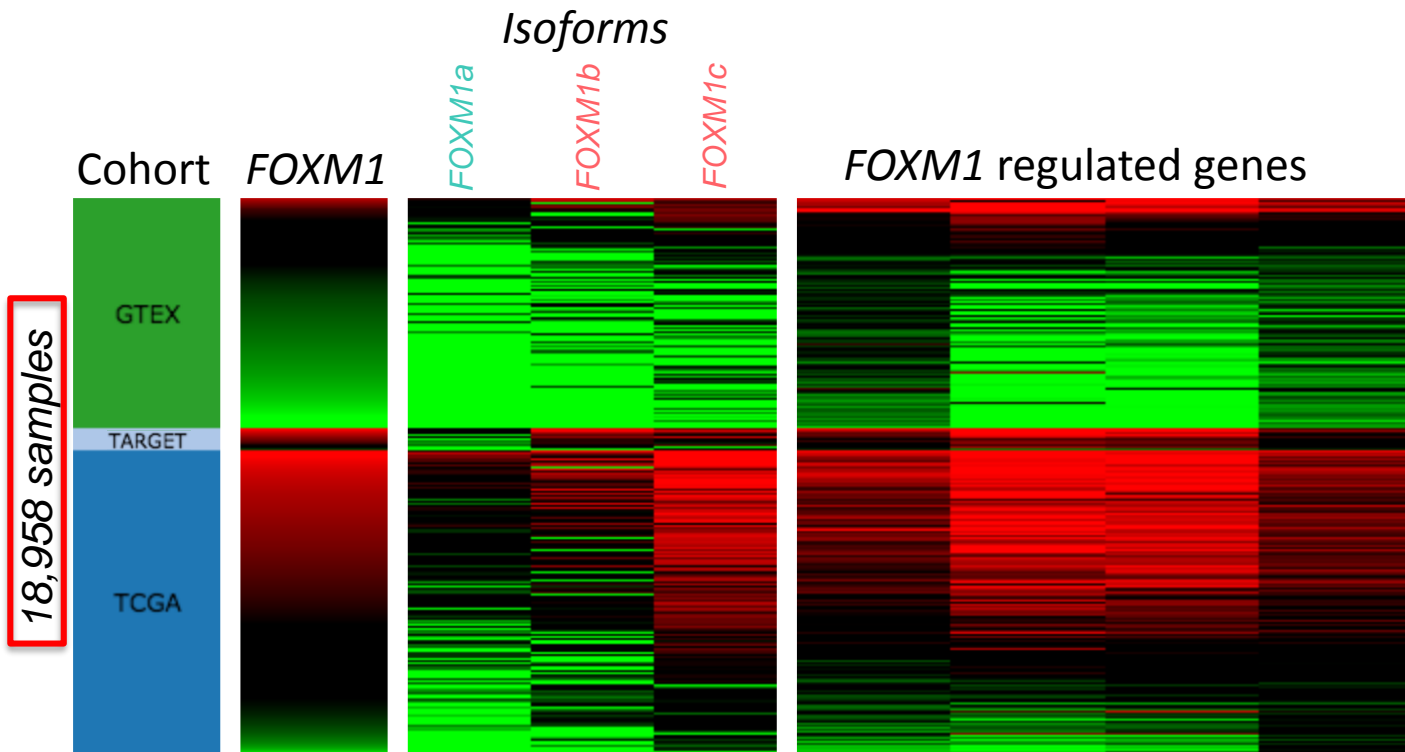
Pan-Cancer *FOXM1* Isoform Expression



FOXM1b, 1c isoforms (activators) are overexpressed in tumors compared to GTEx normal samples

FOXM1a isoform (repressor) does not show this pattern.

Pan-Cancer *FOXM1* Isoform Expression



We see a corresponding difference in *FOXM1* regulated genes.

20k samples, any user selected genes and transcripts, viewed in real-time on a web browser

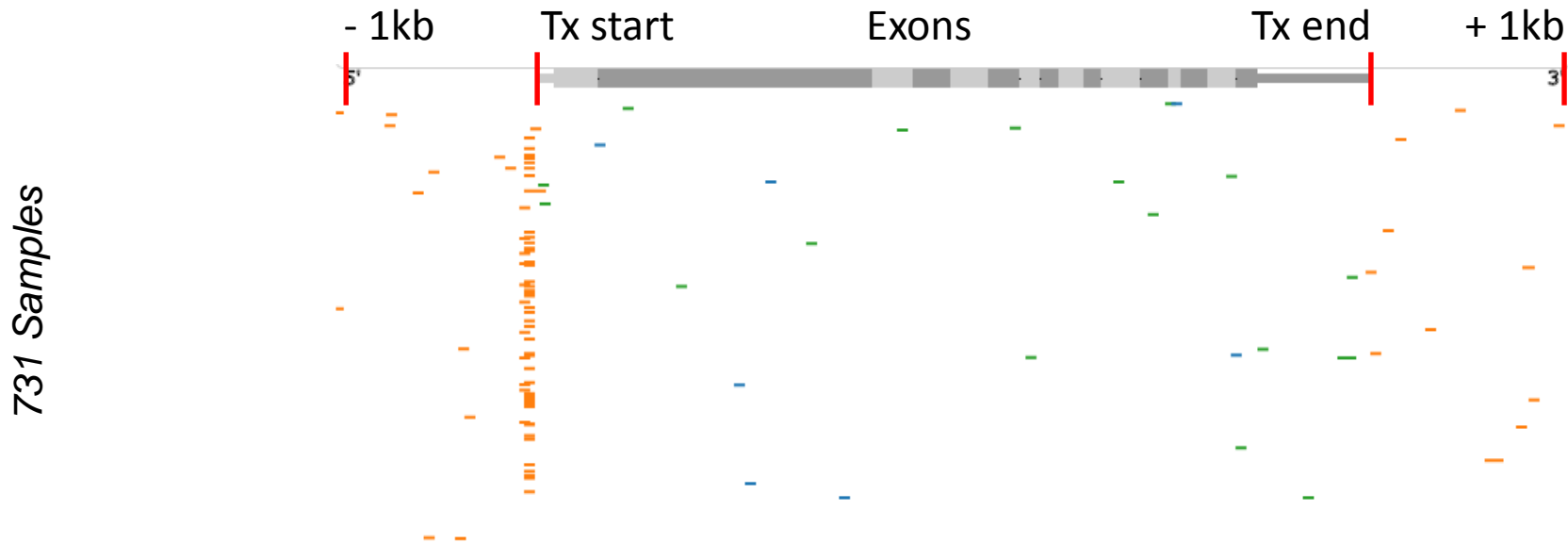


PCAWG
PanCancer Analysis
OF WHOLE GENOMES

Pan-Cancer Analysis of 2800 whole genomes

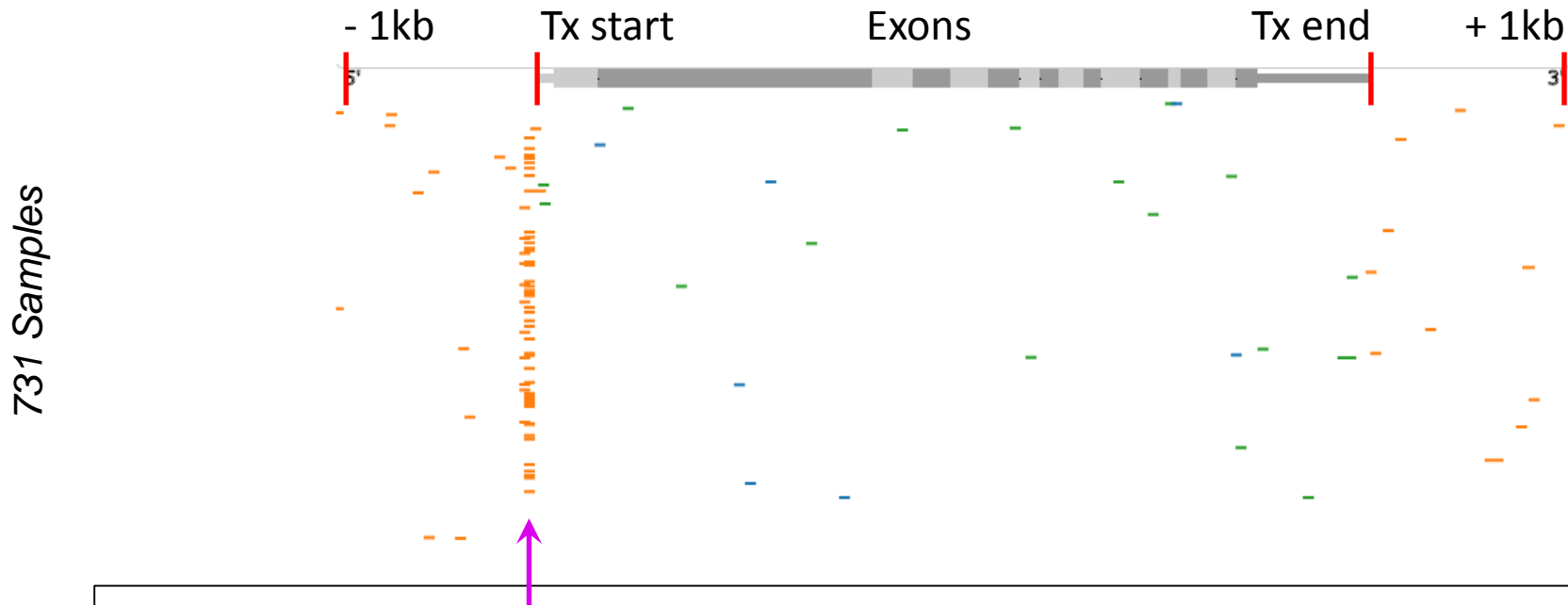
- Largest uniformly-analyzed whole genome dataset in the public domain
 - Variants within as well as outside exon regions
- We are part of the working group ‘Exploratory: portals, visualization and software infrastructure’

View *TERT* mutations along transcripts



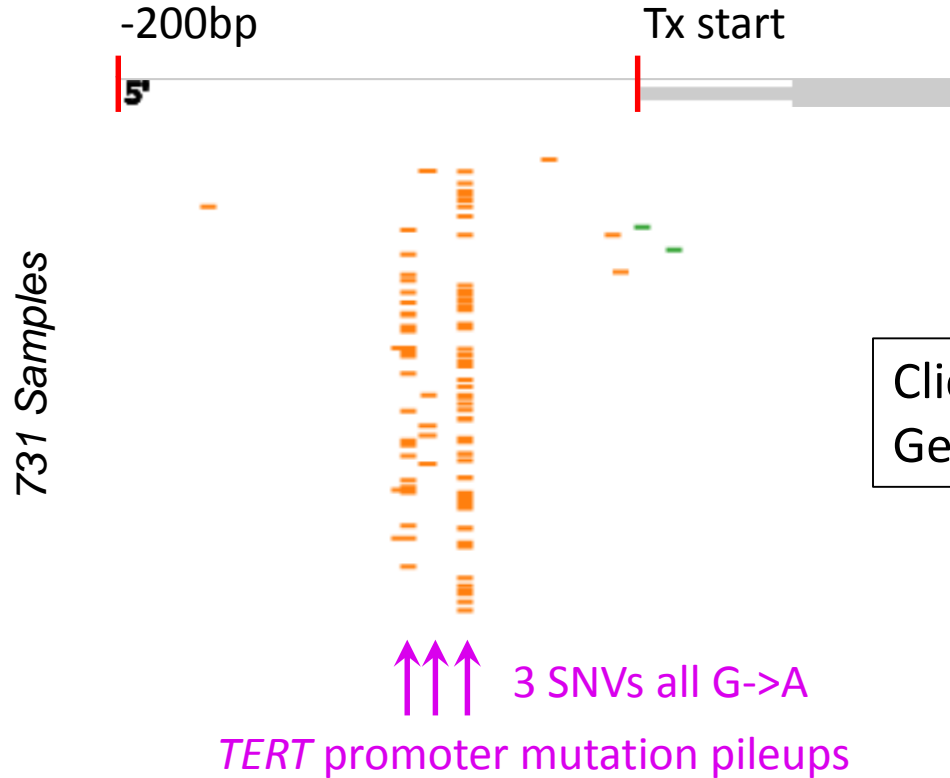
Sample-level mutation along transcripts and 1kb up- and down-stream

View *TERT* mutations along transcripts



TERT exon region has very few somatic mutations, but mutations pile up in the promoter region

Zooming in ...



Click through to UCSC
Genome Browser

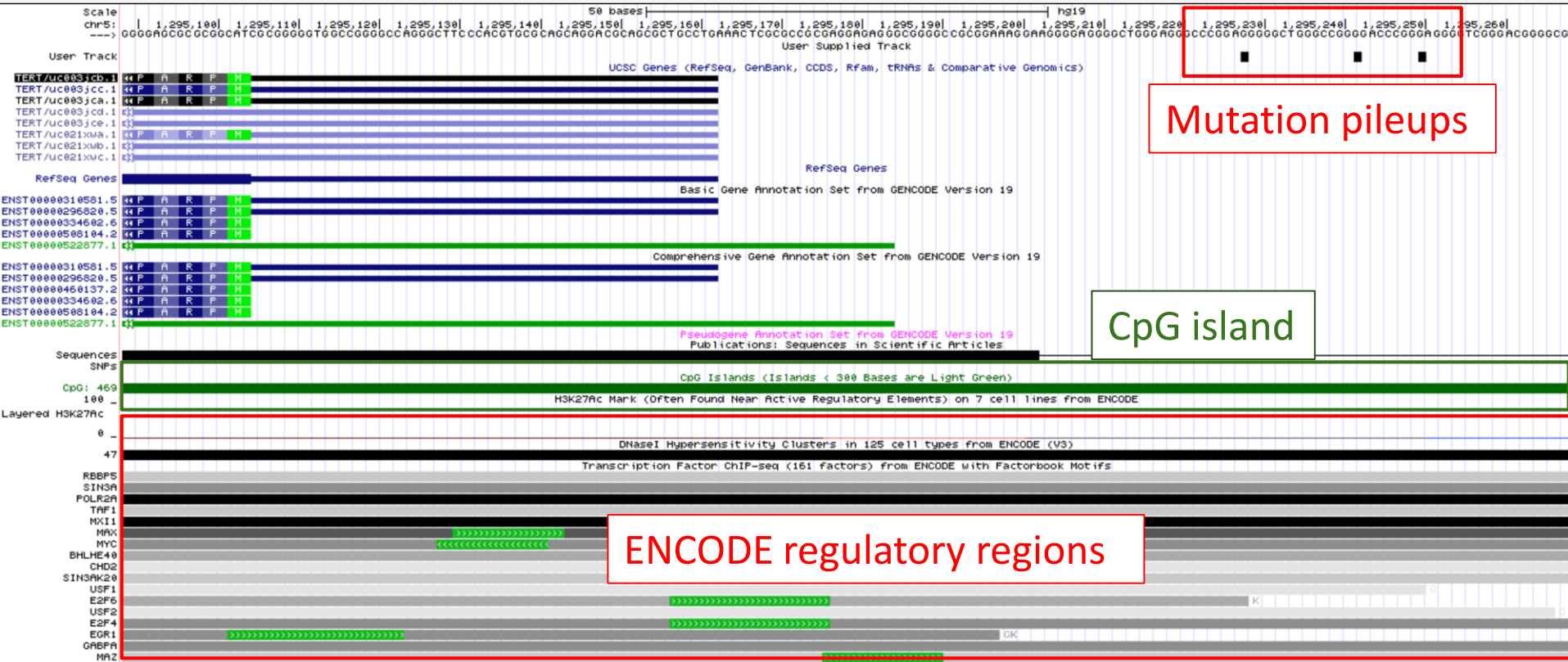
Click to Jump to Human Genome Browser

UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x

chr5:1,295,089-1,295,268 180 bp.

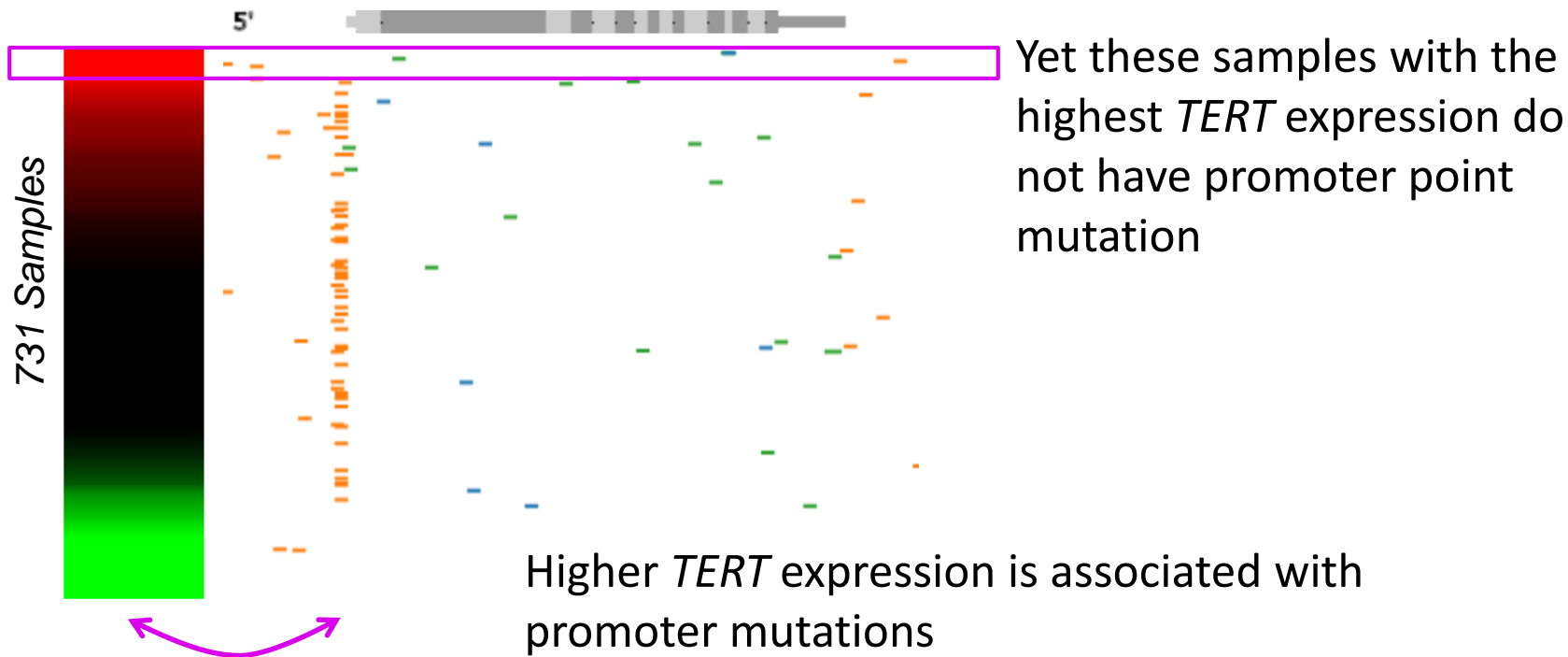
go



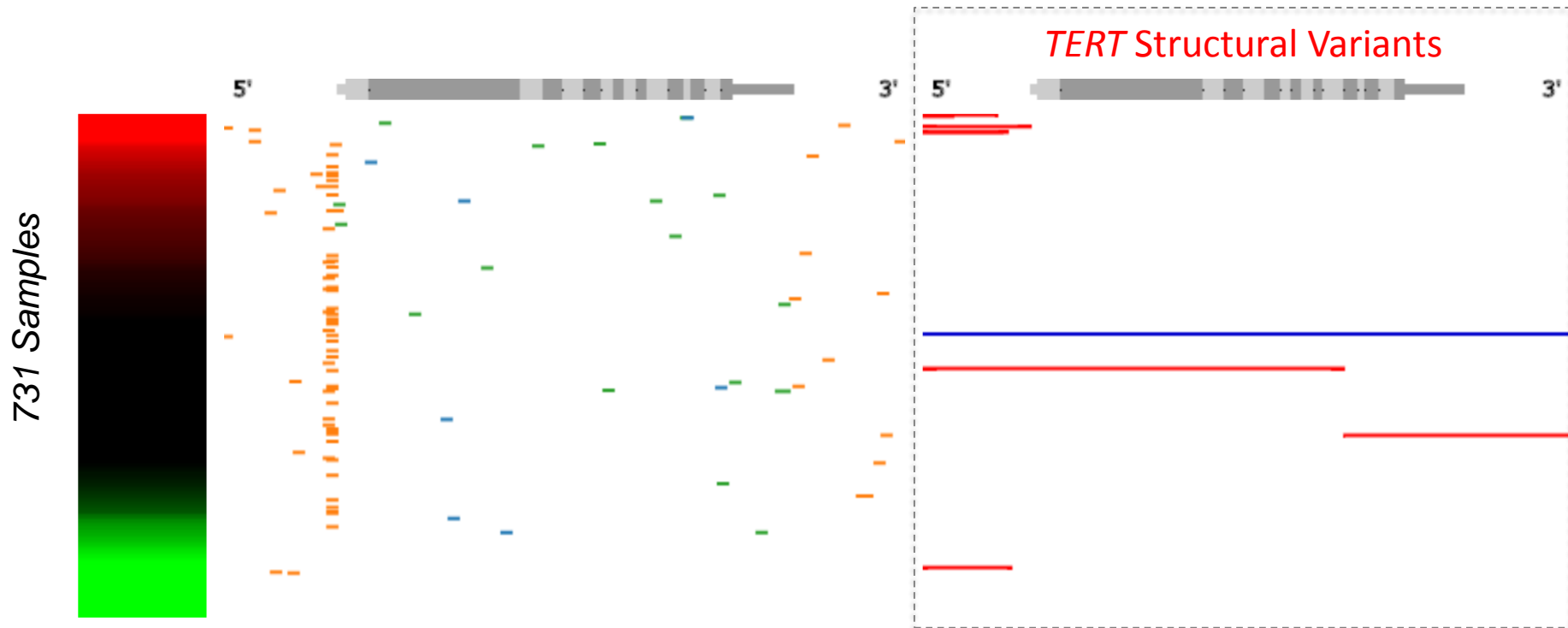
Explore patterns between mutation and expression



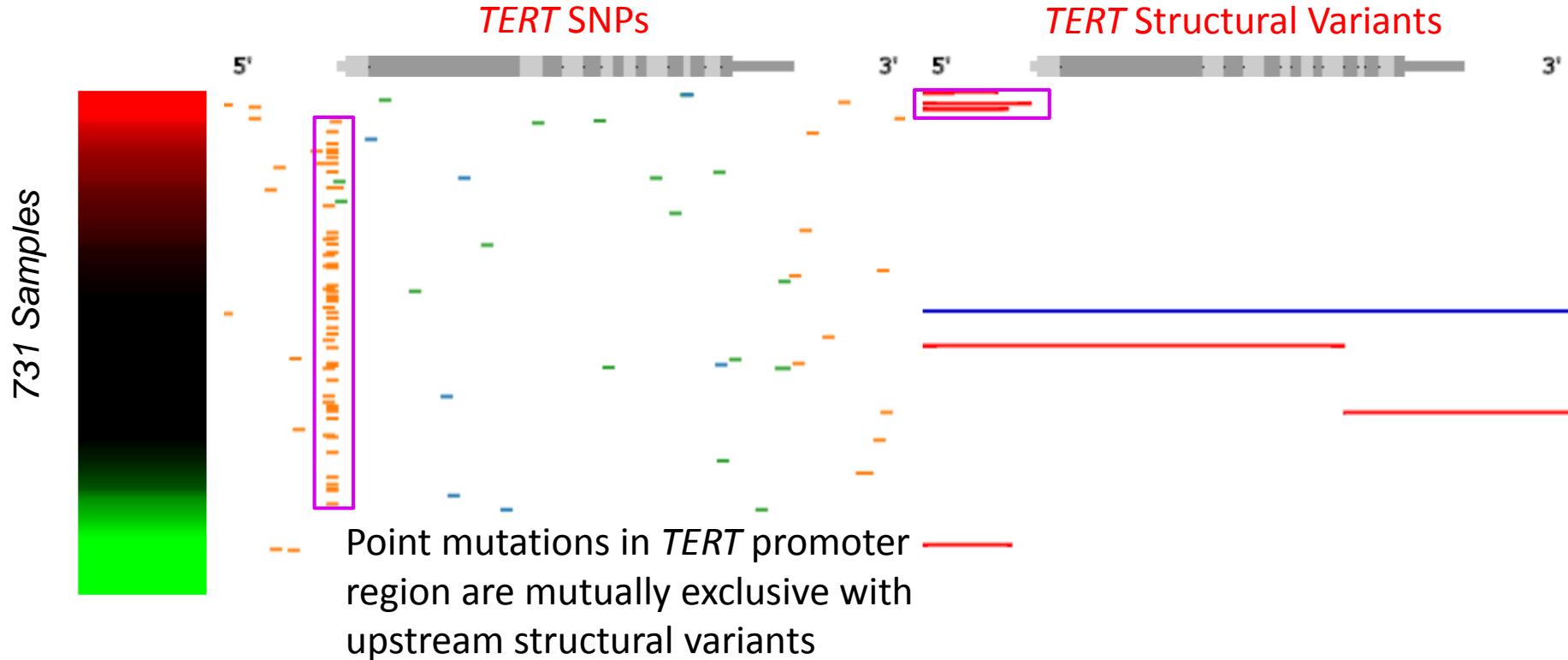
Explore patterns between mutation and expression



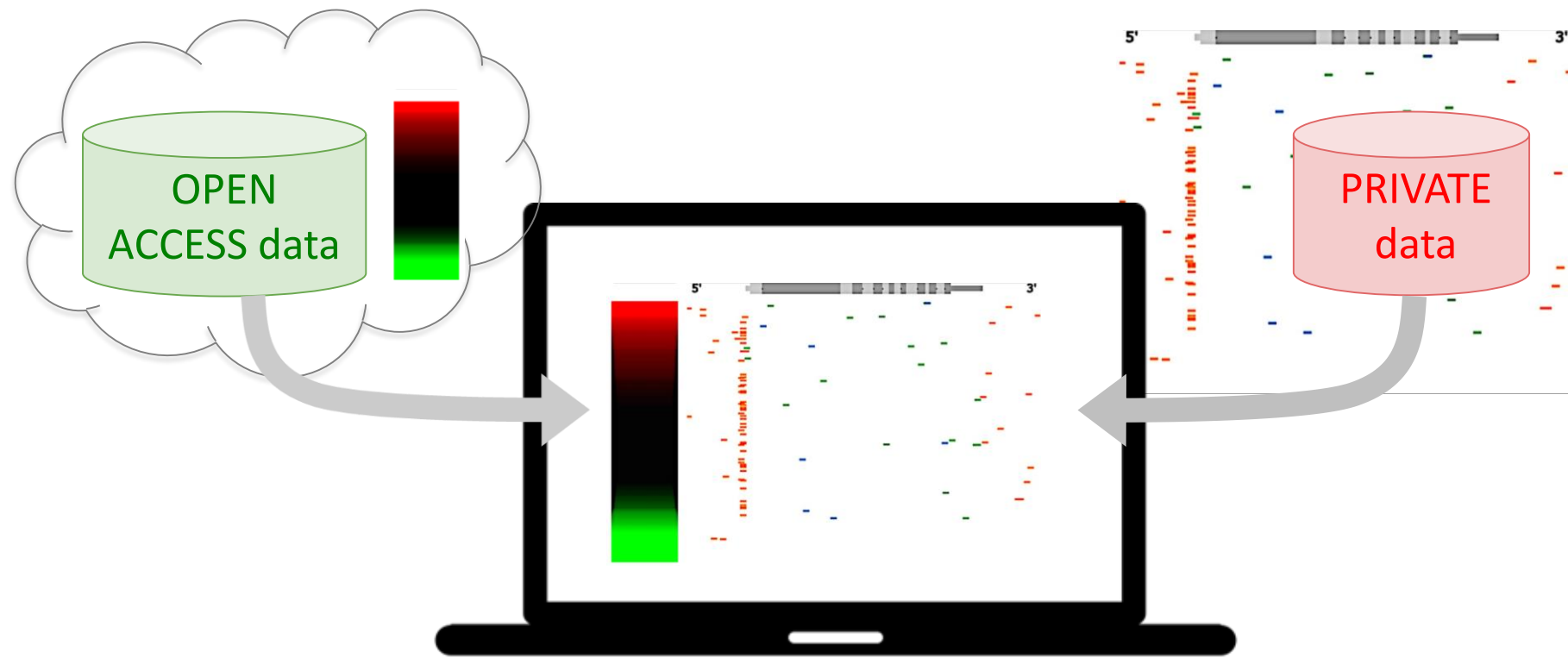
TERT promoter point mutations, SVs, and expression



TERT promoter point mutations, SV, and expression



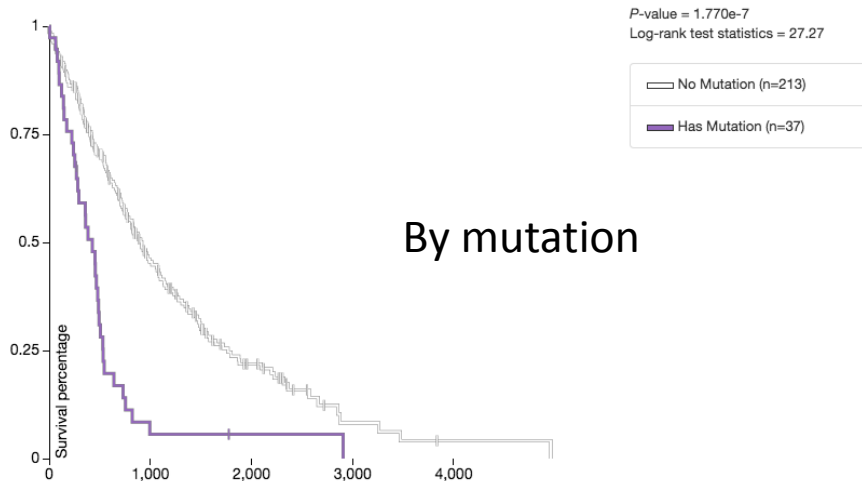
Xena Browser links data across public and private data hubs



Another view: Kaplan-Meier analysis

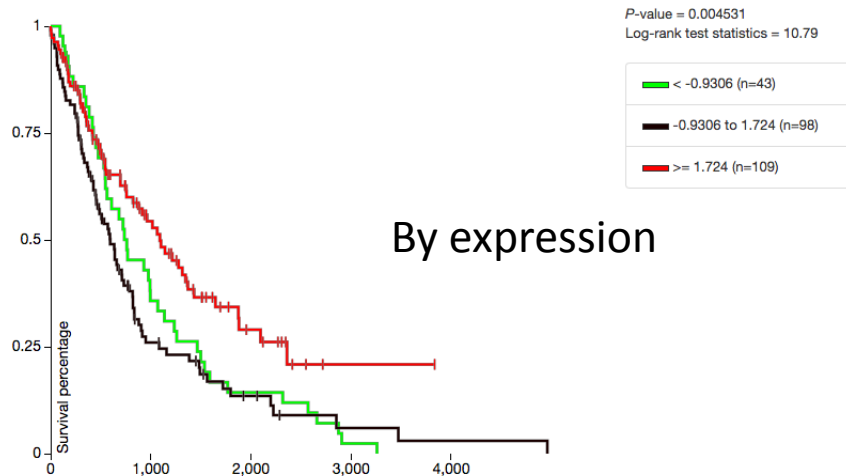
Kaplan Meier

TERT_SNP

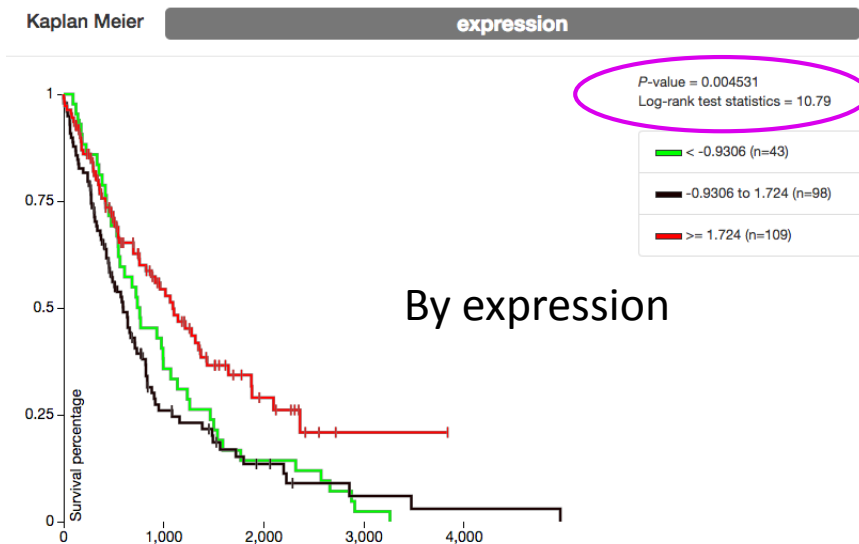
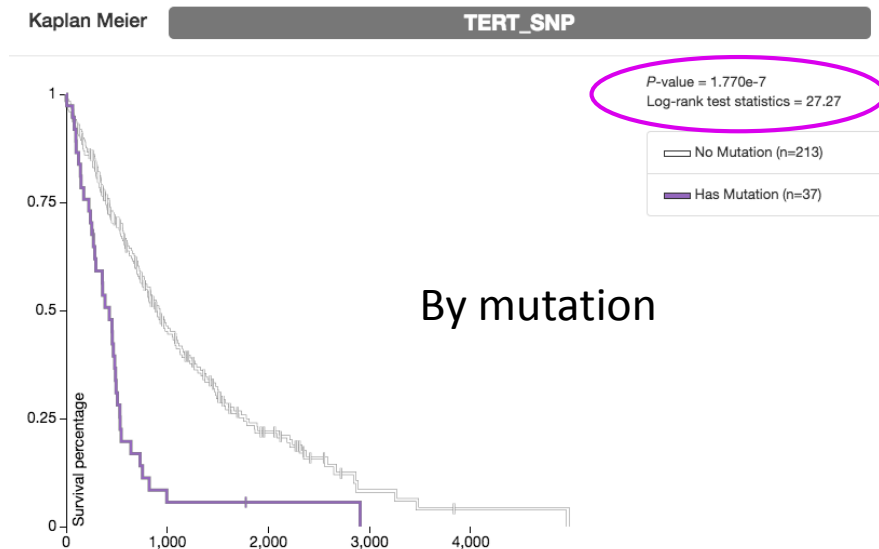


Kaplan Meier

expression

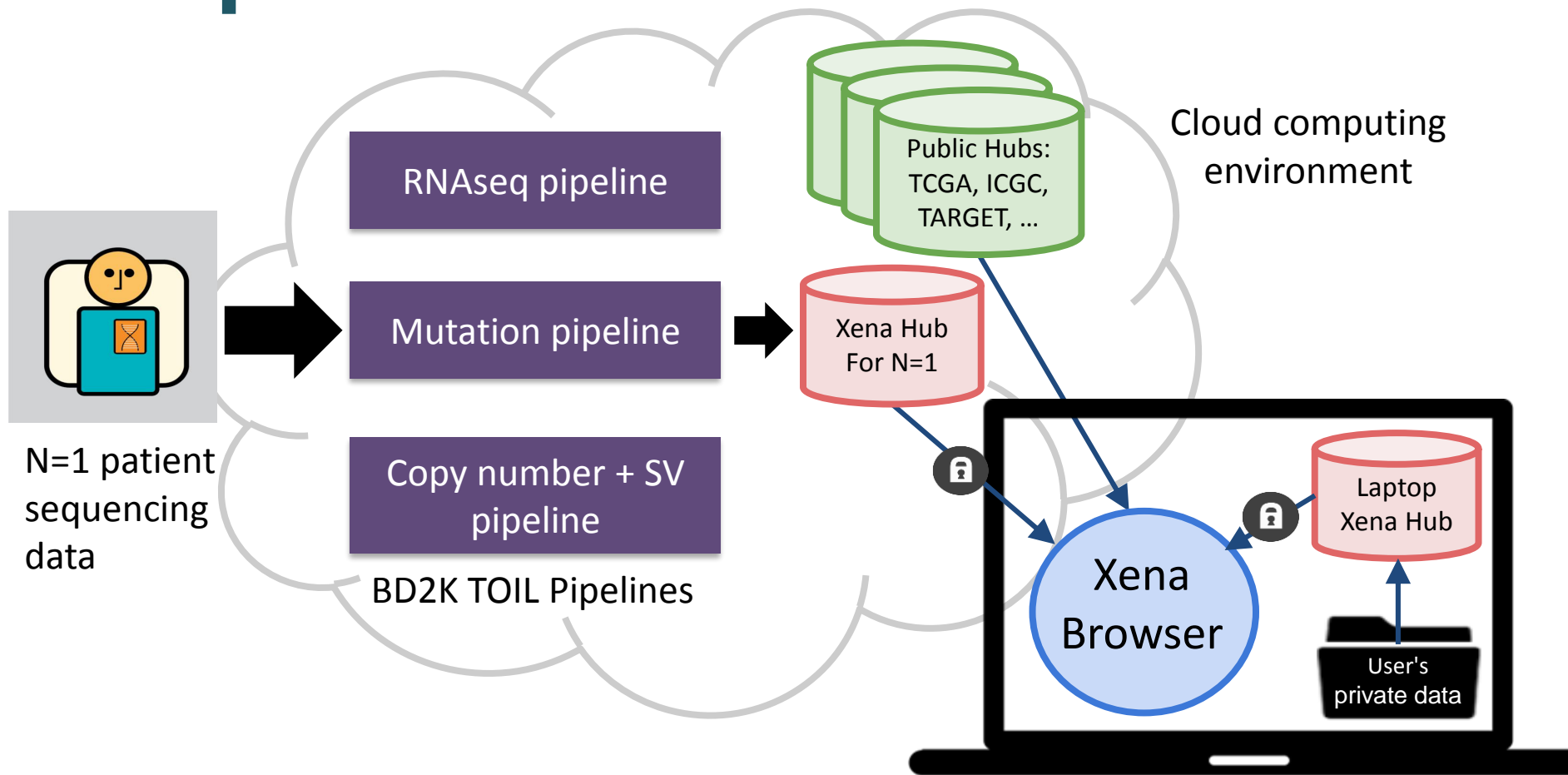


Another view: Kaplan-Meier analysis



Test statistics computed in the browser – data is not sent to a server, maintaining data privacy

Interpretation of N=1 case



California Kids Cancer Comparison (CKCC)



The California Kids Cancer Comparison (CKCC) project provides the power of large-scale bioinformatics to improve patient outcome. It analyzes each patient's tumor in the context of thousands of pediatric and adult tumors that have undergone similar characterization.

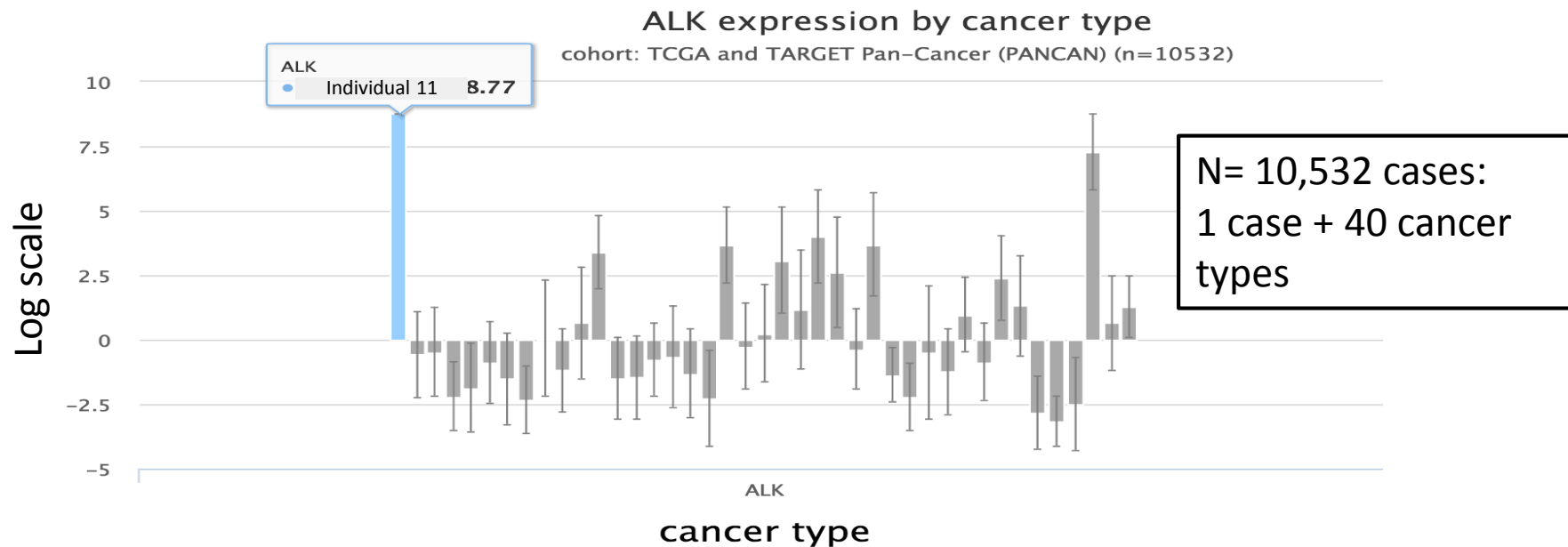
Olena Morozova & CKCC Team

Clinical Trial – Individual 011

- 26-year-old non-Hispanic Caucasian male
- July 11, 2015: presented with shortness of breath, elevated heart rate and blood pressure, as well as blurry vision
- August 2, 2015: diagnosed with a pons/DIPG (Diffuse Intrinsic Pontine Glioma) based on brain MRI
- August 10, 2015: having met eligibility criteria, consent was obtained and patient was enrolled
- August 11, 2015: underwent surgery to remove the tumor
- Per institutional protocol, surgery was followed by standard radiation therapy
- Current status: living
- **Tumor is not a “typical” DIPG because of the individual’s age**

What are unique features of Individual 11 as compared to other tumors in the reference compendium?

GENE EXPRESSION OUTLIER ANALYSIS REVEALS HIGH ALK EXPRESSION IN INDIVIDUAL 11 AS COMPARED TO CKCC REFERENCE COMPENDIUM



ALK expression is hugely higher than any other cancer except neuroblastoma, a disease where *ALK* is a frequently mutated gene

Cohort CKCC_VC Samples in Please select... 10532 samples

Matching samples: 1

[Help with search](#)

X: B: _primary_disease

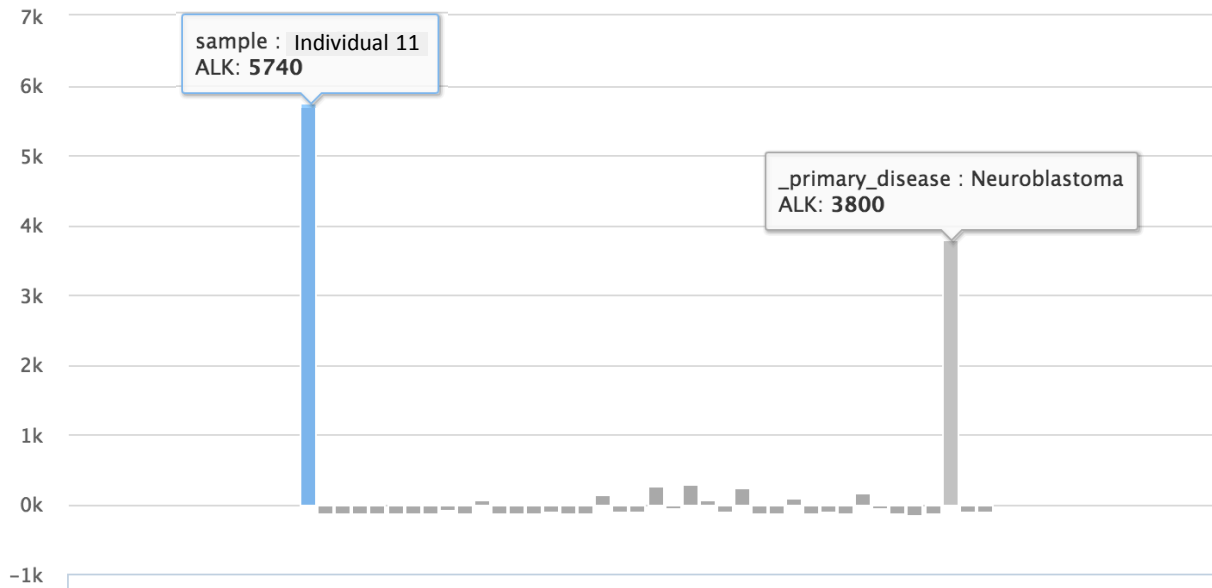
Y: A: ALK

ALK by _primary_disease

cohort: CKCC_VC (n=10532)

RSEM Normalized Count

before-log_{ALK}



_primary_disease

(Click
to hide)

- Individual 11
 - acute myeloid leukemia
 - adrenocortical cancer
 - cholangiocarcinoma
 - bladder urothelial carcinoma
 - breast invasive carcinoma
 - cervical & endocervical cancer
 - colon adenocarcinoma
 - uterine corpus endometrioid carcinoma
 - esophageal carcinoma
 - glioblastoma multiforme
 - head & neck squamous cell carcinoma
- ▲ 1/4 ▼

Supported by Xena Composite Cohort

- Combine user-selected samples from any number of studies, from public or private hubs on the Xena Browser
- No need to download large datasets to combine! Data stays where it is while still maintaining data privacy
- Clinical Driver Projects: N=1 trial case combined with all TCGA and TARGET cases

Our tools are connectable and reusable

Modules that are embeddable in other web applications

- **KM statistics** - compute right in user's web browser, no need to send data to R server
- **Interval tree** - fast genomic coordinate overlapping query
- **Visual spreadsheet** – embedding inside **open source ICGC portal** underway
- Contribute to community through **BioJS**

Turn-key functional genomics data hub

- Installable across all platforms and computing environments
- **Javascript and Python API** to connect to other applications

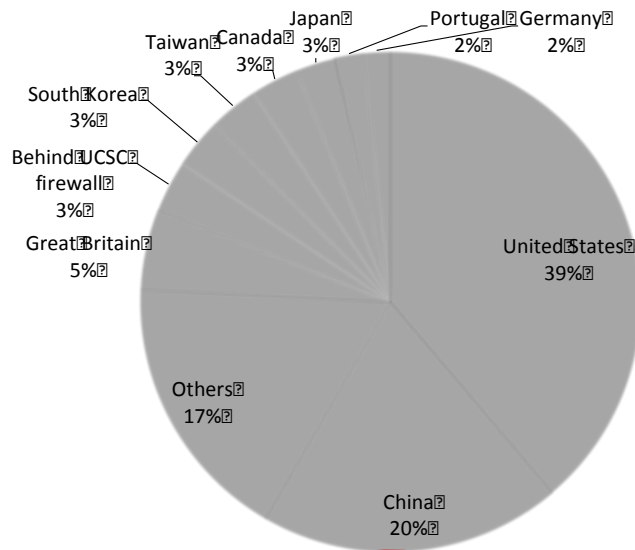
Latest Usage (May 2016)

- Xena Browser

- 5700 users
- Average duration: 10 min
- Page hits: 1 million
- Data download: 230GB

- Xena Hub

- downloads: 73
- Mac: 58 Linux: 10 Windows: 5



User outreach and support

- Working with individual biology research projects
- Email support, google-group, twitter, mailing list, newsletter, help doc
- Video tutorial (708 views)
- Workshops, conferences (~6/year)

Our Team

Brian Craft

Mary Goldman

Jing Zhu

David Haussler



UNIVERSITY OF CALIFORNIA
SANTA CRUZ | Genomics
Institute

Collaborators

TOIL Team & UCSC BD2K center

CKCC Team

TumorMap Team

ITOMIC trial

MuPIT/CRAVAT

BioJS

PCAWG working groups

UCSC Genomics Core



PCAWG
PanCancer Analysis
of WHOLE GENOMES



Big Data to
Knowledge (BD2K)



Global Alliance
for Genomics & Health



BIOJS



NATIONAL CANCER INSTITUTE
Informatics Technology for
Cancer Research



Thanks

<http://xena.ucsc.edu>

<https://www.youtube.com/watch?v=go38U6iLjsw>