FIREBROWSE

Tutorial

Version 2016_04_17

gdac@broadinstitute.org

Introduction

FIREBROWSE

A simple and elegant way to explore cancer data.

Backed by a powerful computational infrastructure, application programming interface (API), graphical tools and online reports.

Sitting above one of the deepest and most integratively-characterized ***open*** cancer datasets in the world.

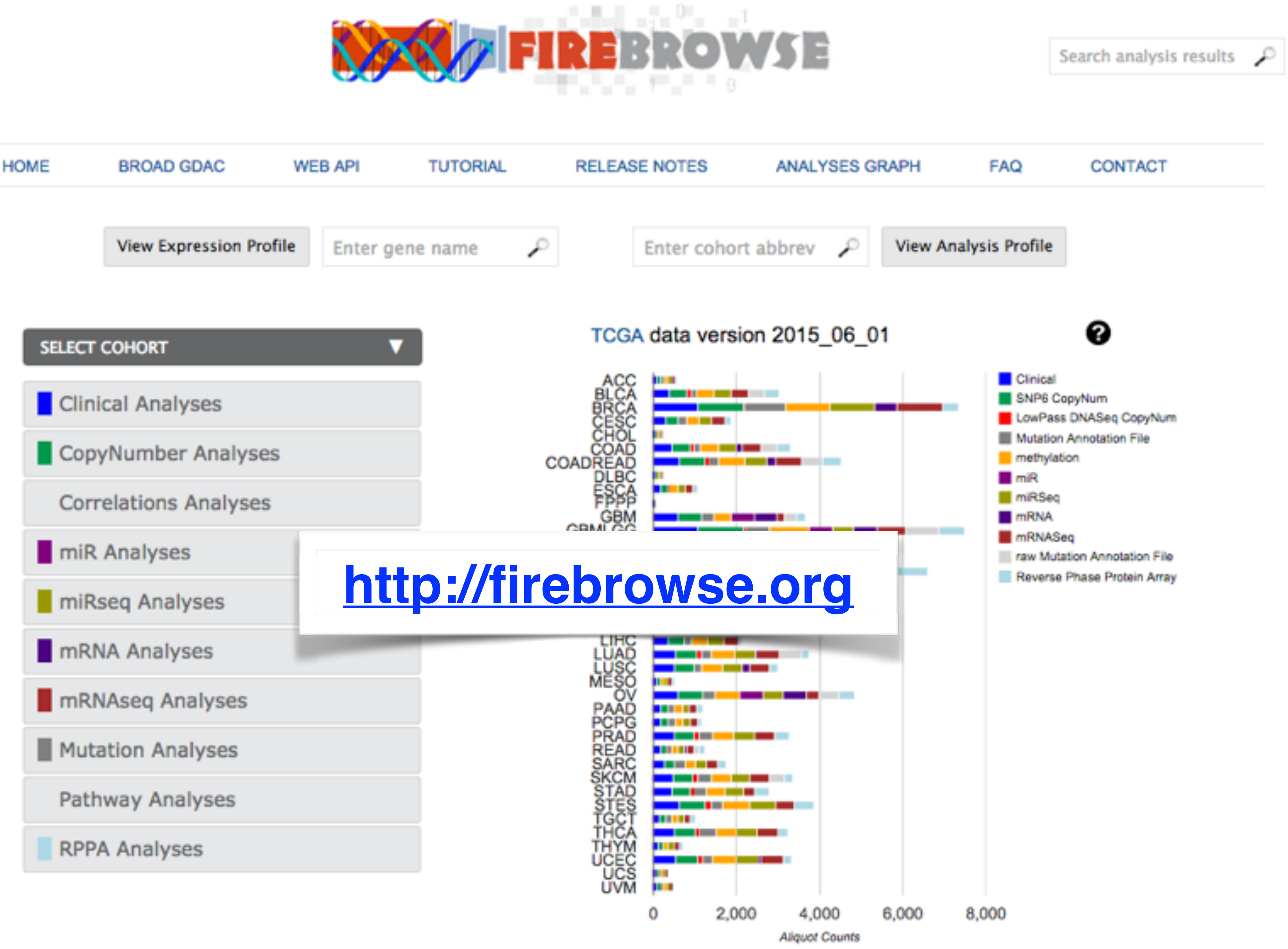
With over 80K sample aliquots from 11,000+ cancer patients, spanning 38 unique disease cohorts.

FIREBROWSE

Data and analyses utilized at numerous academic, research, and commercial sites around the world.

Example: cBio@MSKCC

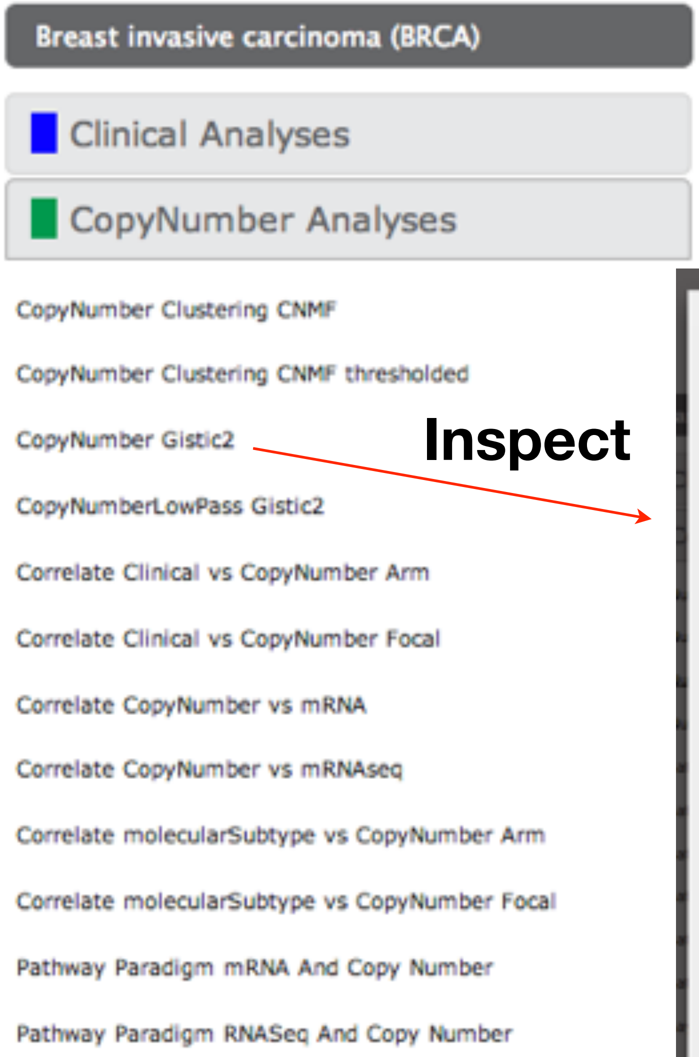
TCGA data & analyses in cBioPortal—expression, mutation, copy number, signficance analyses, and more—are loaded directly from Firehose.



http://firebrowse.org

~1500 Analyses (reports) per run
Find your favorite in 2 clicks

Choose Cohort



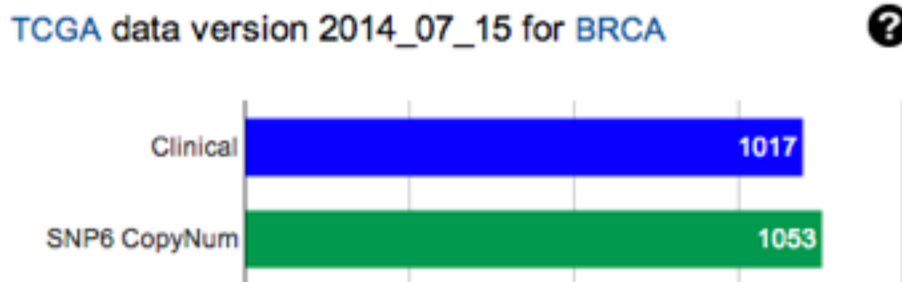
Breast invasive carcinoma (BRCA)

- 🟦 Clinical Analyses
- 🟩 CopyNumber Analyses

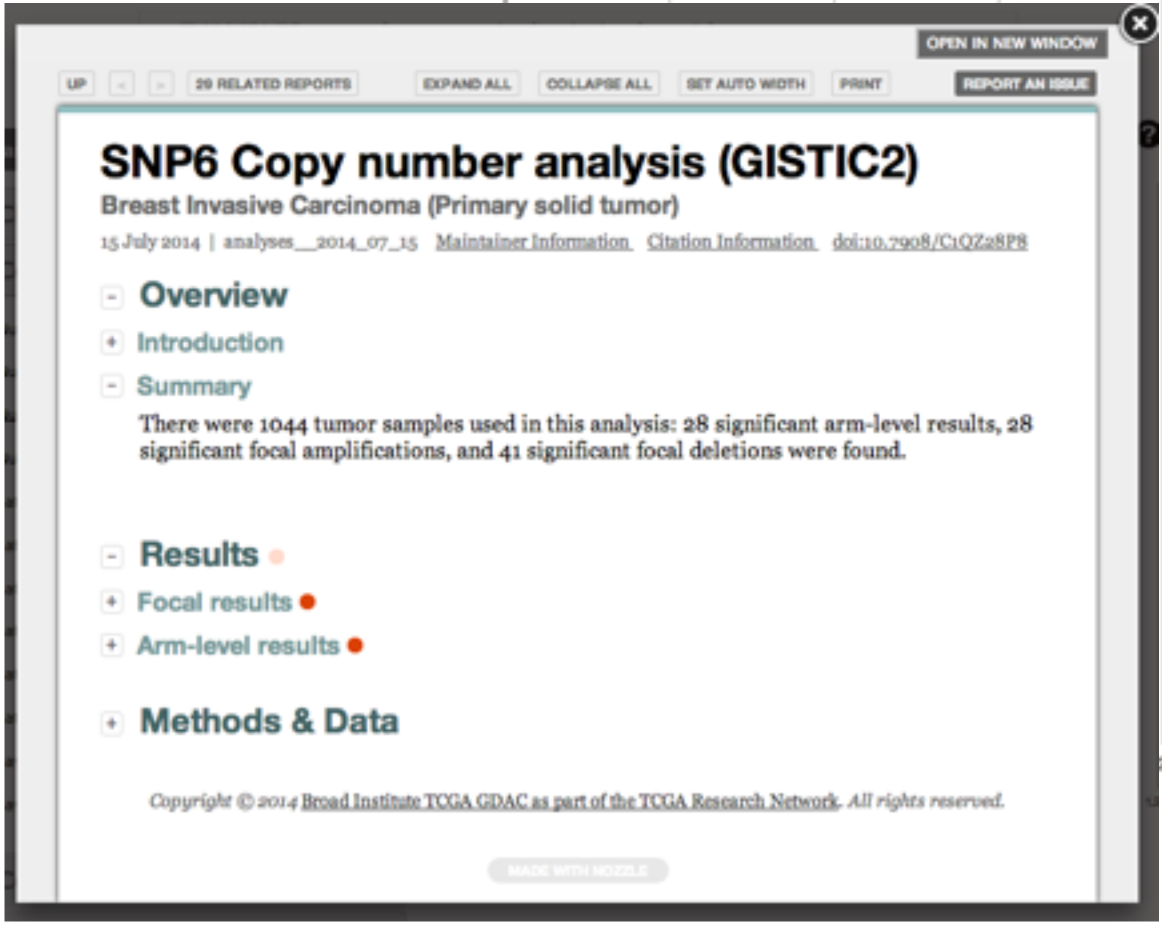
- CopyNumber Clustering CNMF
- CopyNumber Clustering CNMF thresholded
- CopyNumber Gistic2
- CopyNumberLowPass Gistic2
- Correlate Clinical vs CopyNumber Arm
- Correlate Clinical vs CopyNumber Focal
- Correlate CopyNumber vs mRNA
- Correlate CopyNumber vs mRNAseq
- Correlate molecularSubtype vs CopyNumber Arm
- Correlate molecularSubtype vs CopyNumber Focal
- Pathway Paradigm mRNA And Copy Number
- Pathway Paradigm RNASeq And Copy Number

Then DataType

Inspect



TCGA data version 2014_07_15 for BRCA



SNP6 Copy number analysis (GISTIC2)

Breast Invasive Carcinoma (Primary solid tumor)

15 July 2014 | analyses__2014_07_15 Maintainer Information Citation Information doi:10.7908/C1QZ28P8

- Overview

+ Introduction

- Summary

There were 1044 tumor samples used in this analysis: 28 significant arm-level results, 28 significant focal amplifications, and 41 significant focal deletions were found.

- Results

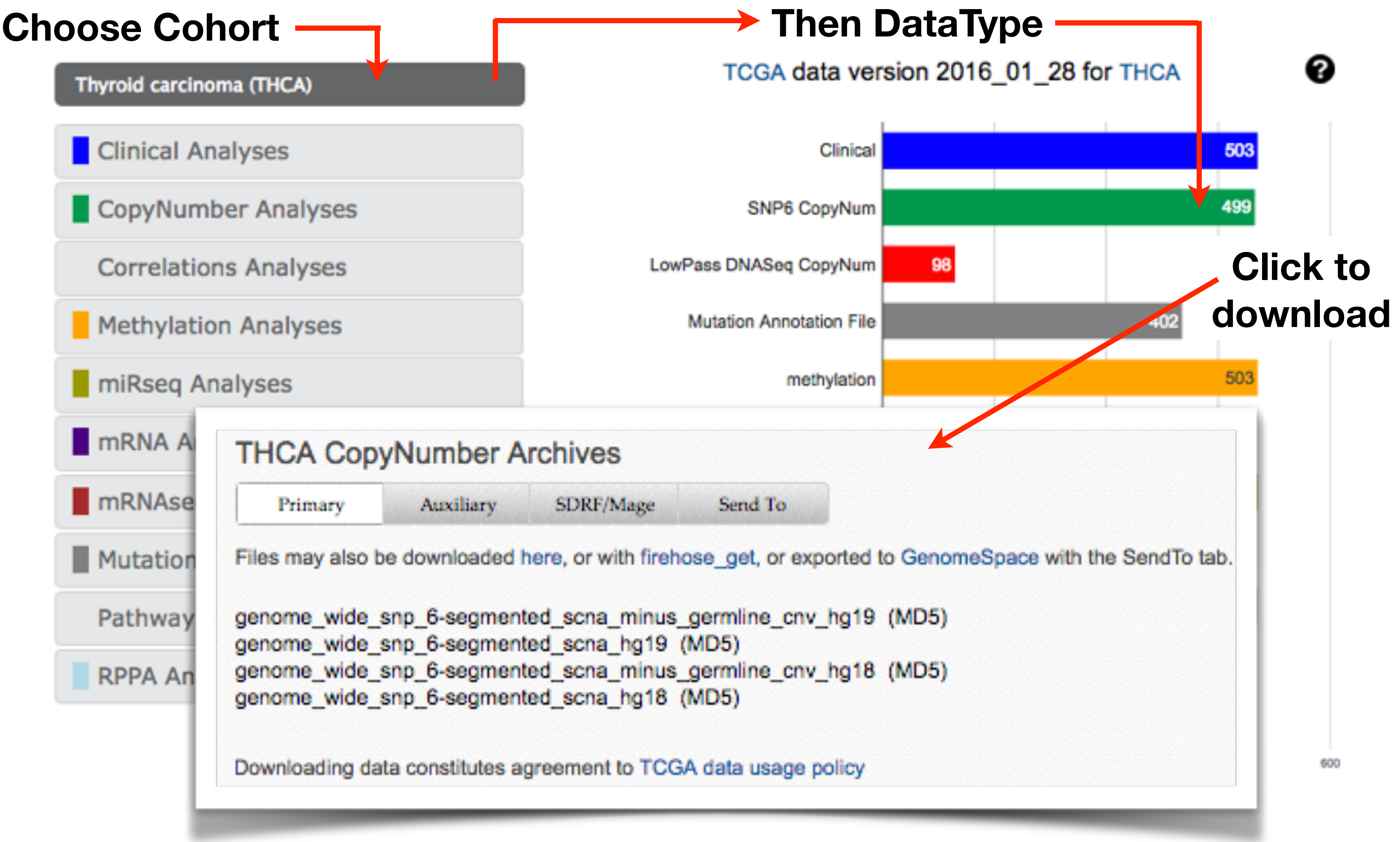
+ Focal results

+ Arm-level results

+ Methods & Data

Copyright © 2014 Broad Institute TCGA GDAC as part of the TCGA Research Network. All rights reserved.

Many 1000s of datasets per run
Find your favorite in 2 clicks



Choose Cohort

Then DataType

Click to download

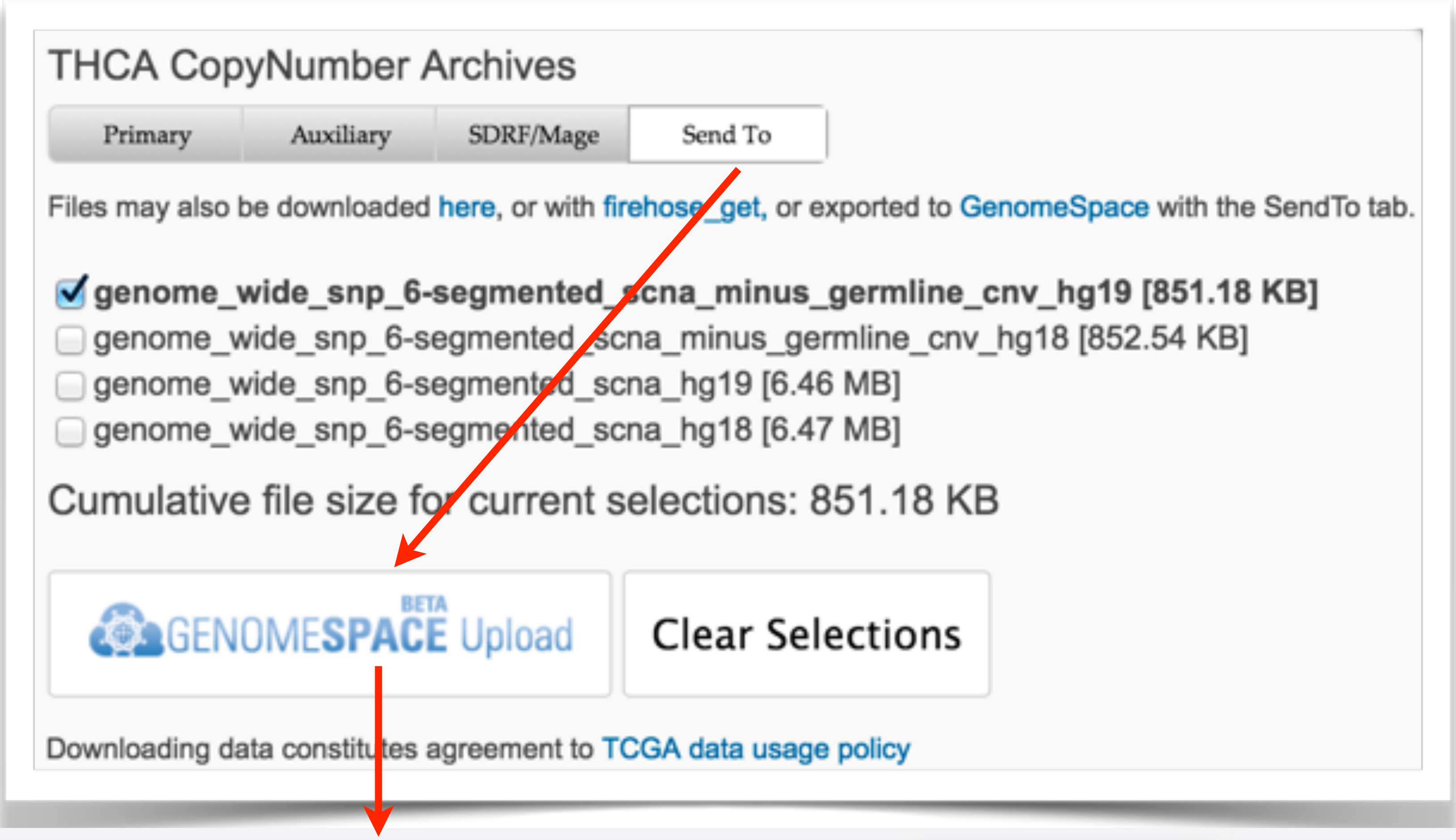
THCA CopyNumber Archives

Files may also be downloaded here, or with firehose_get, or exported to GenomeSpace with the SendTo tab.

genome_wide_snp_6-segmented_scna_minus_germline_cnv_hg19 (MD5)
genome_wide_snp_6-segmented_scna_hg19 (MD5)
genome_wide_snp_6-segmented_scna_minus_germline_cnv_hg18 (MD5)
genome_wide_snp_6-segmented_scna_hg18 (MD5)

Downloading data constitutes agreement to TCGA data usage policy

Or easily send to GenomeSpace for more analysis



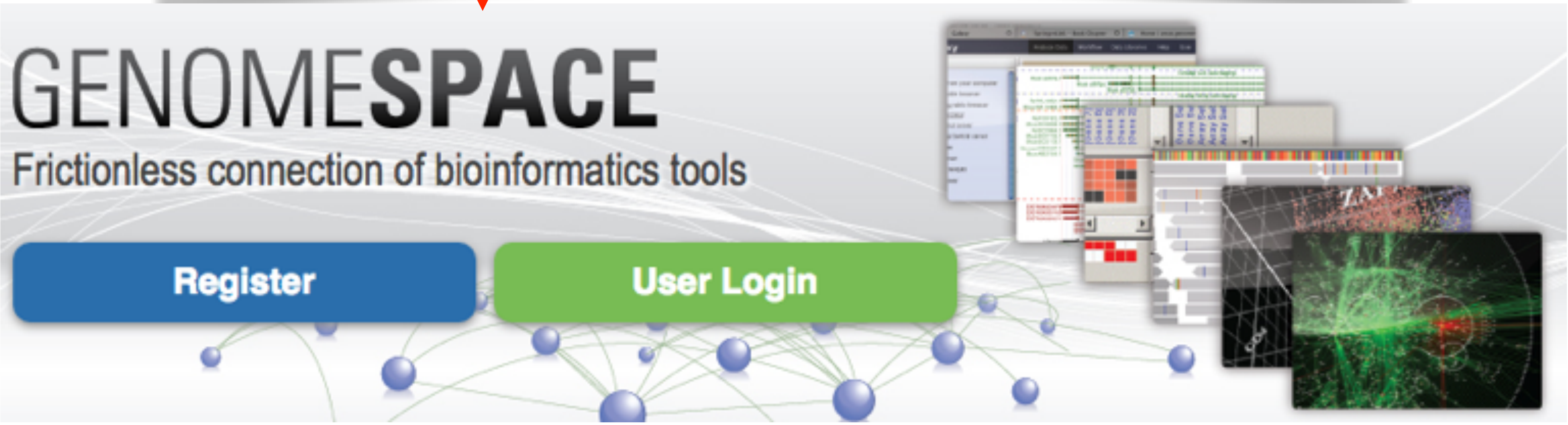
THCA CopyNumber Archives

Files may also be downloaded here, or with firehose_get, or exported to GenomeSpace with the SendTo tab.

☑ **genome_wide_snp_6-segmented_scna_minus_germline_cnv_hg19 [851.18 KB]**
□ genome_wide_snp_6-segmented_scna_minus_germline_cnv_hg18 [852.54 KB]
□ genome_wide_snp_6-segmented_scna_hg19 [6.46 MB]
□ genome_wide_snp_6-segmented_scna_hg18 [6.47 MB]

Cumulative file size for current selections: 851.18 KB

Downloading data constitutes agreement to TCGA data usage policy



GENOMESPACE

Frictionless connection of bioinformatics tools

Or download everything with 1 command

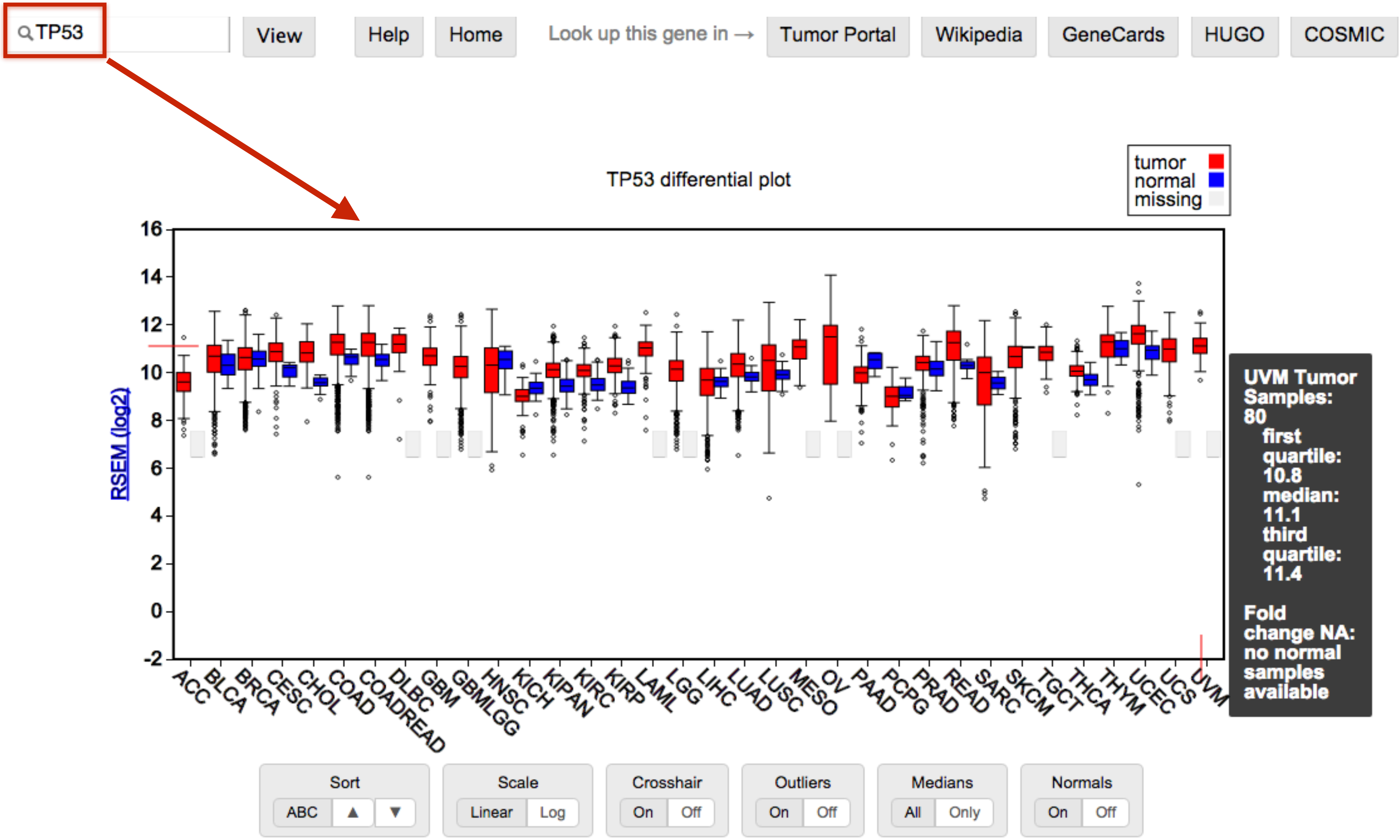
```
linux%  firehose_get  analyses  latest
```

Simple 20K bash script, just 1 moving part

Download Here

Graphical Tools

viewGene: expression level browser

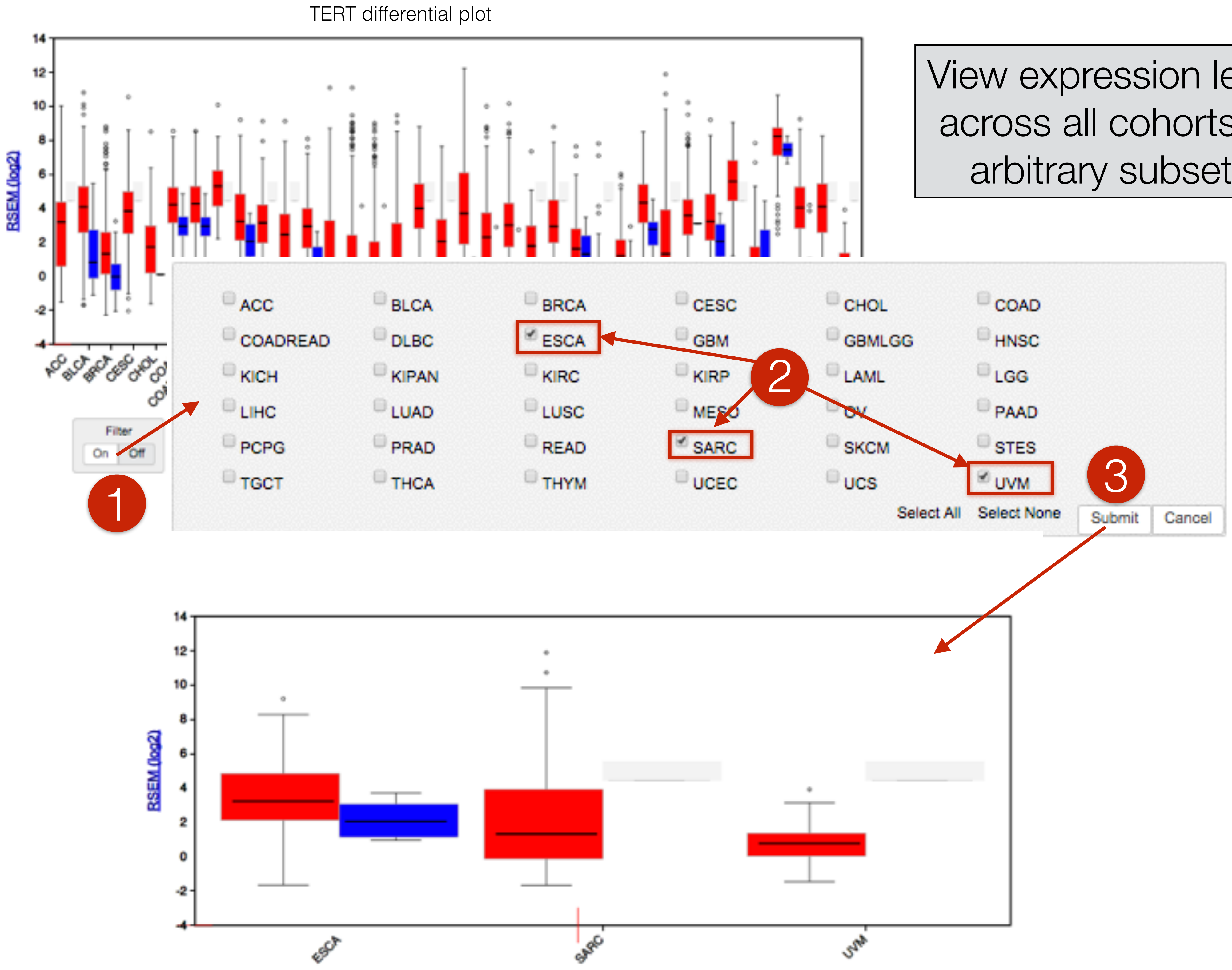


TP53 differential plot

UVM Tumor Samples: 80
first quartile: 10.8
median: 11.1
third quartile: 11.4

Fold change NA: no normal samples available

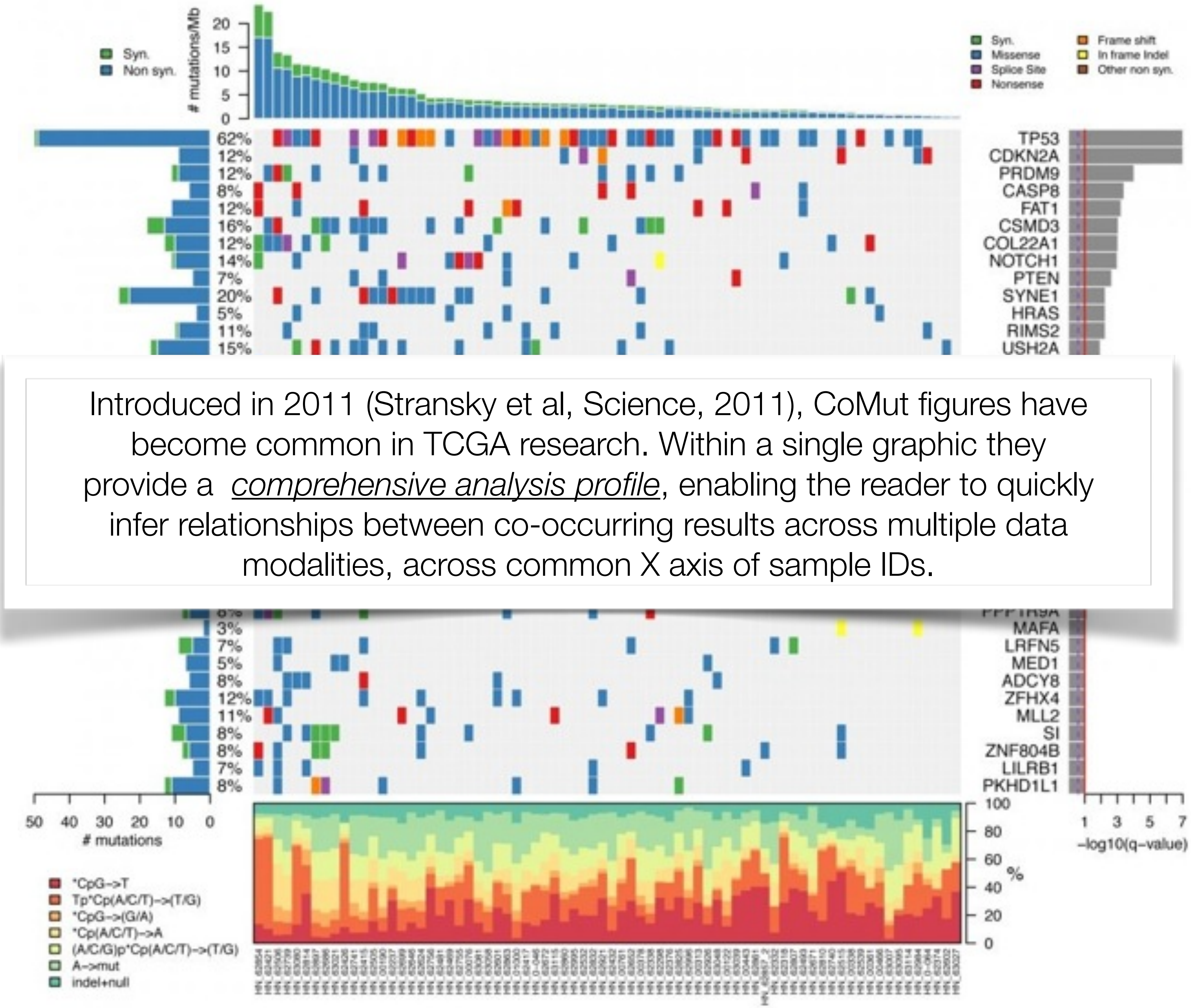
Quickly inspect mRNASeq expression levels for a selected gene



☐ ACC ☐ BLCA ☐ BRCA ☐ CESC ☐ CHOL ☐ COAD
☐ COADREAD ☐ DLBC ☑ ESCA ☐ GBM ☐ GBMLGG ☐ HNSC
☐ KICH ☐ KIPAN ☐ KIRC ☐ KIRP ☐ LAML ☐ LGG
☐ LIHC ☐ LUAD ☐ LUSC ☐ MESO ☐ OV ☐ PAAD
☐ PCPG ☐ PRAD ☐ READ ☑ SARC ☐ SKCM ☐ STES
☐ TGCT ☐ THCA ☐ THYM ☐ UCEC ☐ UCS ☑ UVM

View expression levels across all cohorts, or arbitrary subsets.

CoMut: mutation co-occurence plots

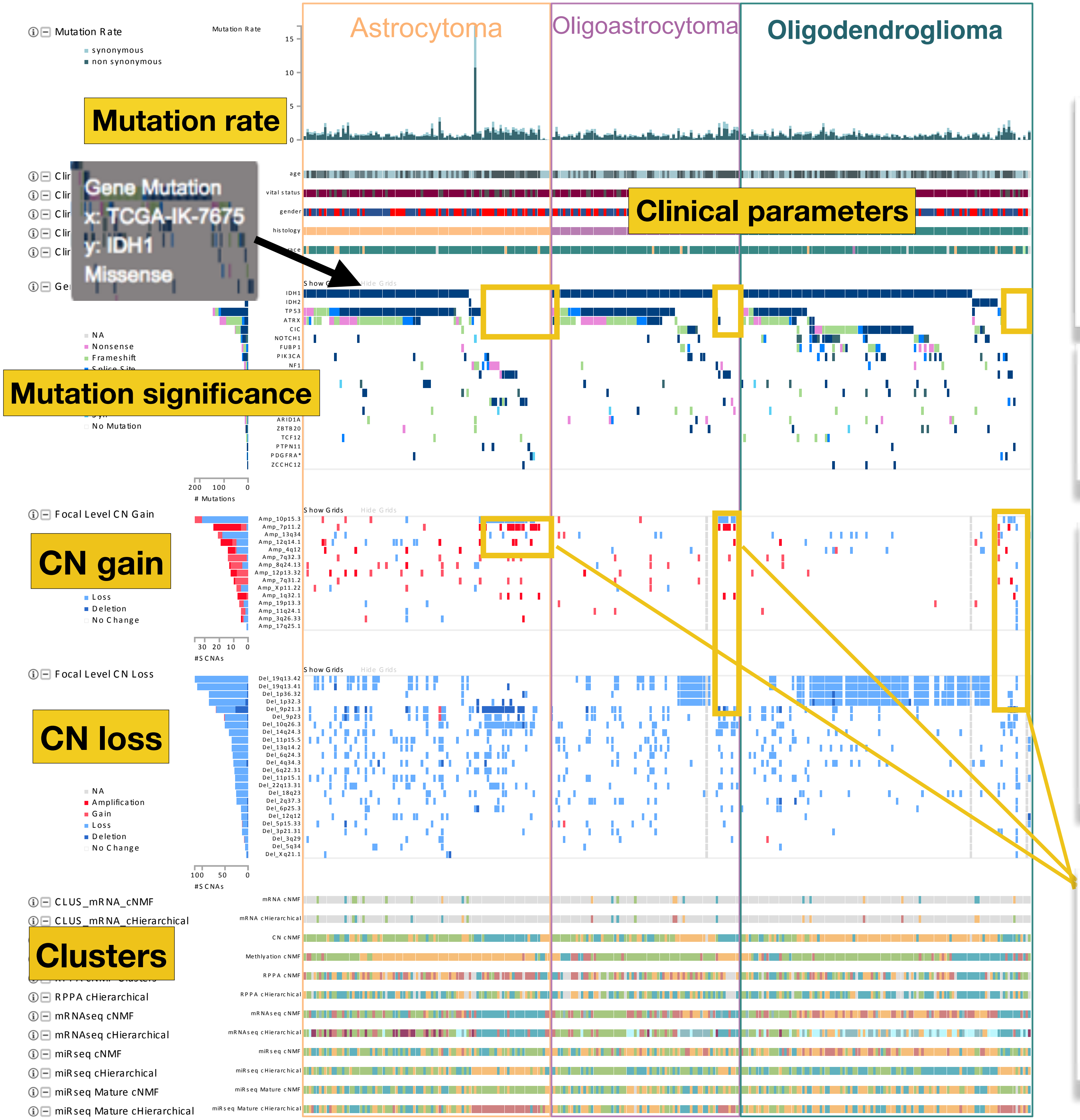


Introduced in 2011 (Stransky et al, Science, 2011), CoMut figures have become common in TCGA research. Within a single graphic they provide a *comprehensive analysis profile*, enabling the reader to quickly infer relationships between co-occurring results across multiple data modalities, across common X axis of sample IDs.

But in journals, figures are static and can be small and hard to read

And cannot be explored in real time

And reproducing them or investigating their implications can require substantial time for data retrieval, preparation and analysis



[iCoMut](#)

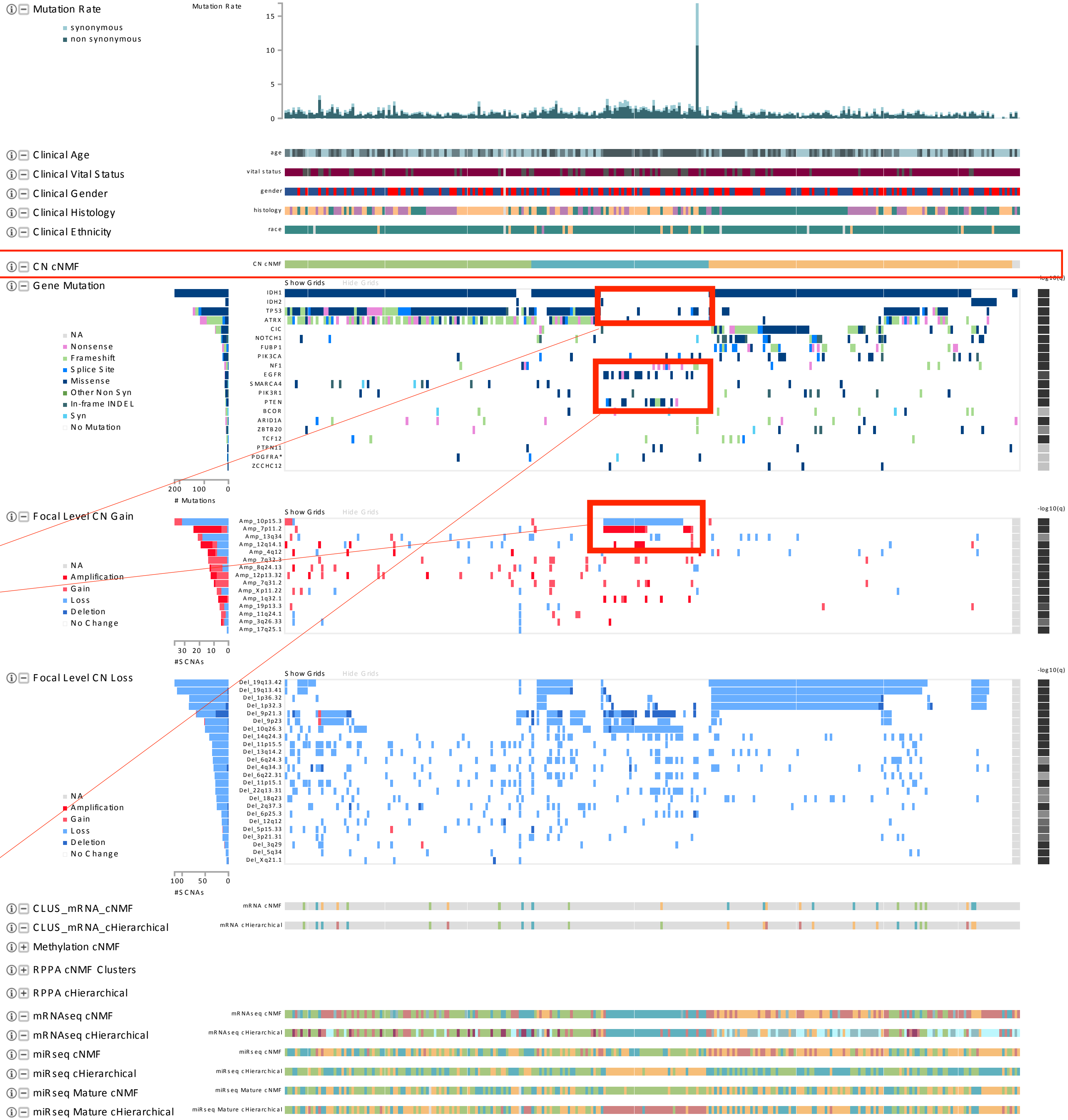
By making such figures interactive, allowing panels to be moved, sorted and searched, iCoMut dramatically enhances that process.

Example: hovering over pixels tells you about the underlying biology.

Here we show the TCGA LGG cohort: sorted first by clinical histology, then gene (descending order of mutation count). The clinical subtypes leap off the page at you.

As does the fact that the copy-number landscape differs when IDH1/2, TP53, and ATRX mutations drop off.

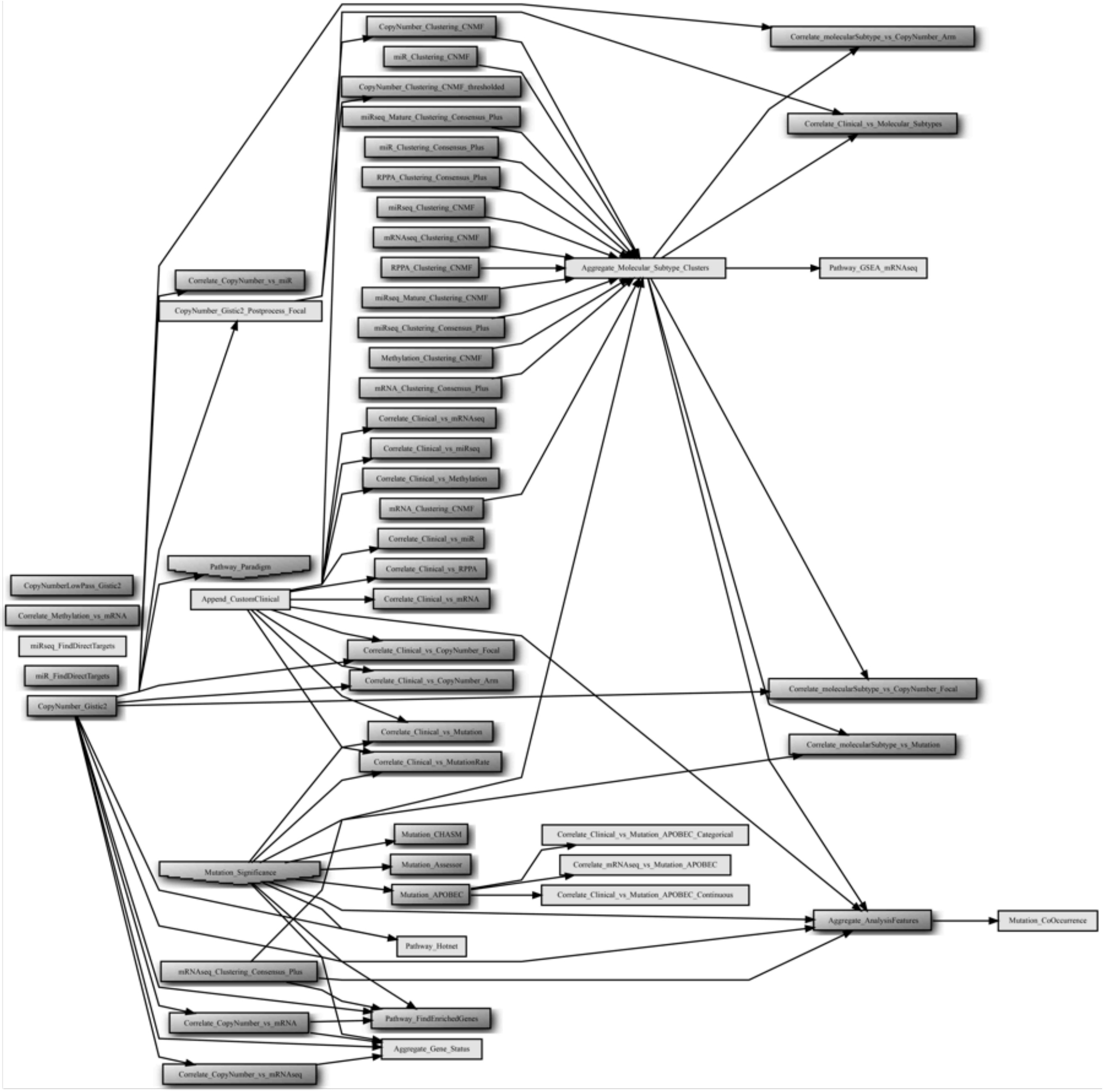
Now we've re-sorted by CNMF copy-number clustering, ***and dragged it from bottom of figure to top,*** just above mutation panel



Making it further apparent that the copy-number landscape differs as IDH1/2, TP53, and ATRX mutations diminish

Also shows apparent involvement with EGFR and PTEN.

iCoMut compressess an entire Firehose run into a single, interactive & reproducible figure

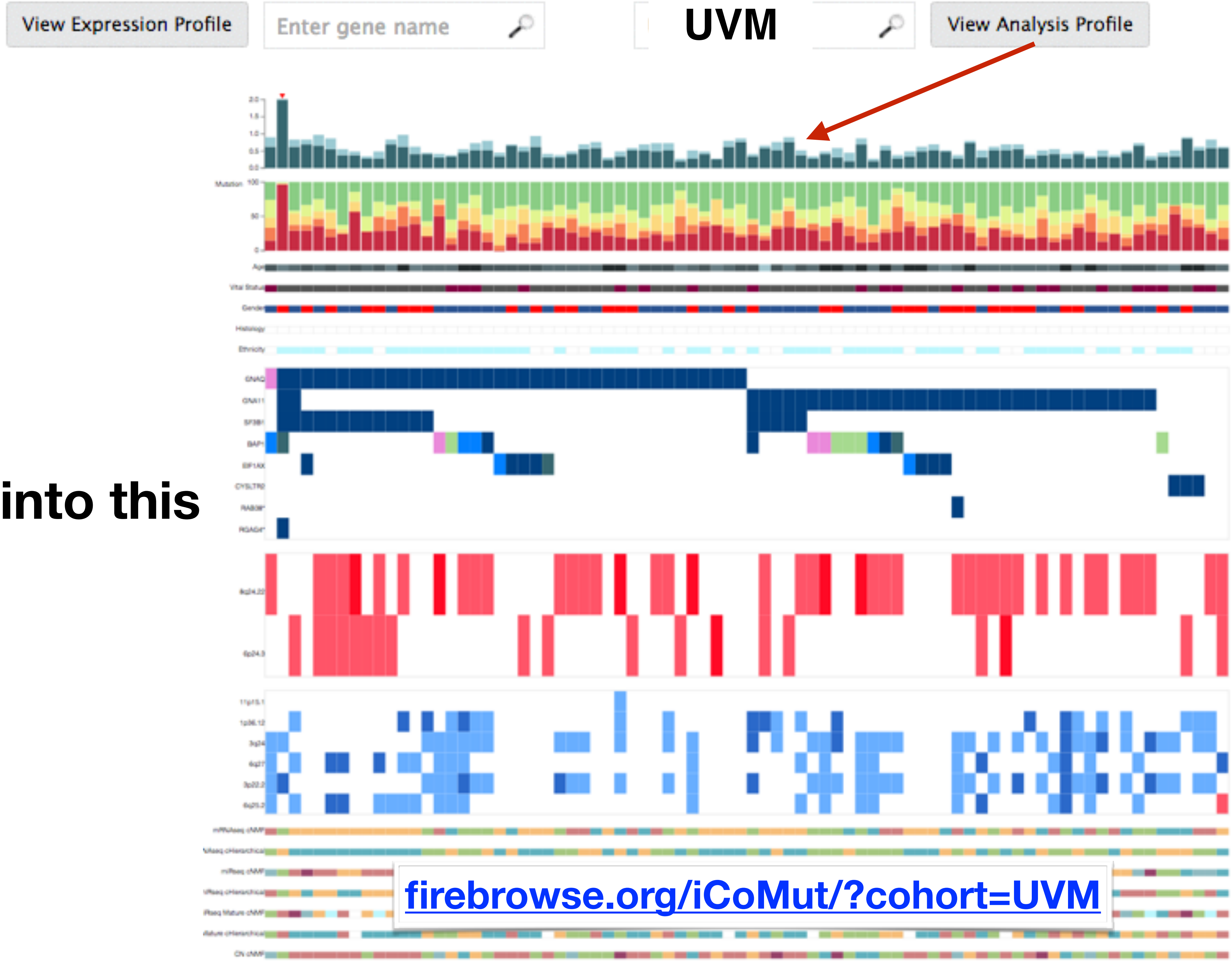

CopyNumber_Clustering_CNMF
miR_Clustering_CNMF
CopyNumber_Clustering_CNMF_thresholded
miRseq_Mature_Clustering_Consensus_Plus
miR_Clustering_Consensus_Plus
RPPA_Clustering_Consensus_Plus
miRseq_Clustering_CNMF
mRNAseq_Clustering_CNMF
RPPA_Clustering_CNMF
miRseq_Mature_Clustering_CNMF
miRseq_Clustering_Consensus_Plus
Methylation_Clustering_CNMF
mRNA_Clustering_Consensus_Plus
Correlate_Clinical_vs_mRNAseq
Correlate_Clinical_vs_miRseq
Correlate_Clinical_vs_Methylation
mRNA_Clustering_CNMF
Correlate_Clinical_vs_miR
Correlate_Clinical_vs_RPPA
Correlate_Clinical_vs_mRNA
Correlate_CopyNumber_vs_miR
CopyNumber_Gistic2_Postprocess_Focal
Pathway_Paradigm
Append_CustomClinical
CopyNumberLowPass_Gistic2
Correlate_Methylation_vs_mRNA
miRseq_FindDirectTargets
miR_FindDirectTargets
CopyNumber_Gistic2
Correlate_molecularSubtype_vs_CopyNumber_Arm
Correlate_Clinical_vs_Molecular_Subtypes
Aggregate_Molecular_Subtype_Clusters
Pathway_GSEA_mRNAseq
Correlate_Clinical_vs_CopyNumber_Focal
Correlate_Clinical_vs_CopyNumber_Arm
Correlate_molecularSubtype_vs_CopyNumber_Focal
Correlate_Clinical_vs_Mutation
Correlate_Clinical_vs_MutationRate
Correlate_molecularSubtype_vs_Mutation
Mutation_CHASM
Mutation_Assessor
Mutation_APOBEC
Correlate_Clinical_vs_Mutation_APOBEC_Categorical
Correlate_mRNAseq_vs_Mutation_APOBEC
Correlate_Clinical_vs_Mutation_APOBEC_Continuous
Mutation_Significance
Aggregate_AnalysisFeatures
Mutation_CoOccurrence
Pathway_Hotnet
mRNAseq_Clustering_Consensus_Plus
Pathway_FindEnrichedGenes
Correlate_CopyNumber_vs_mRNA
Aggregate_Gene_Status
Correlate_CopyNumber_vs_mRNAseq

Turning this …

Firehose analysis workflow

Run on 38 TCGA cohorts
> 100 tasks per run
~10 datatypes

Distills 50 TB of input data into 10GB of results (5000x)

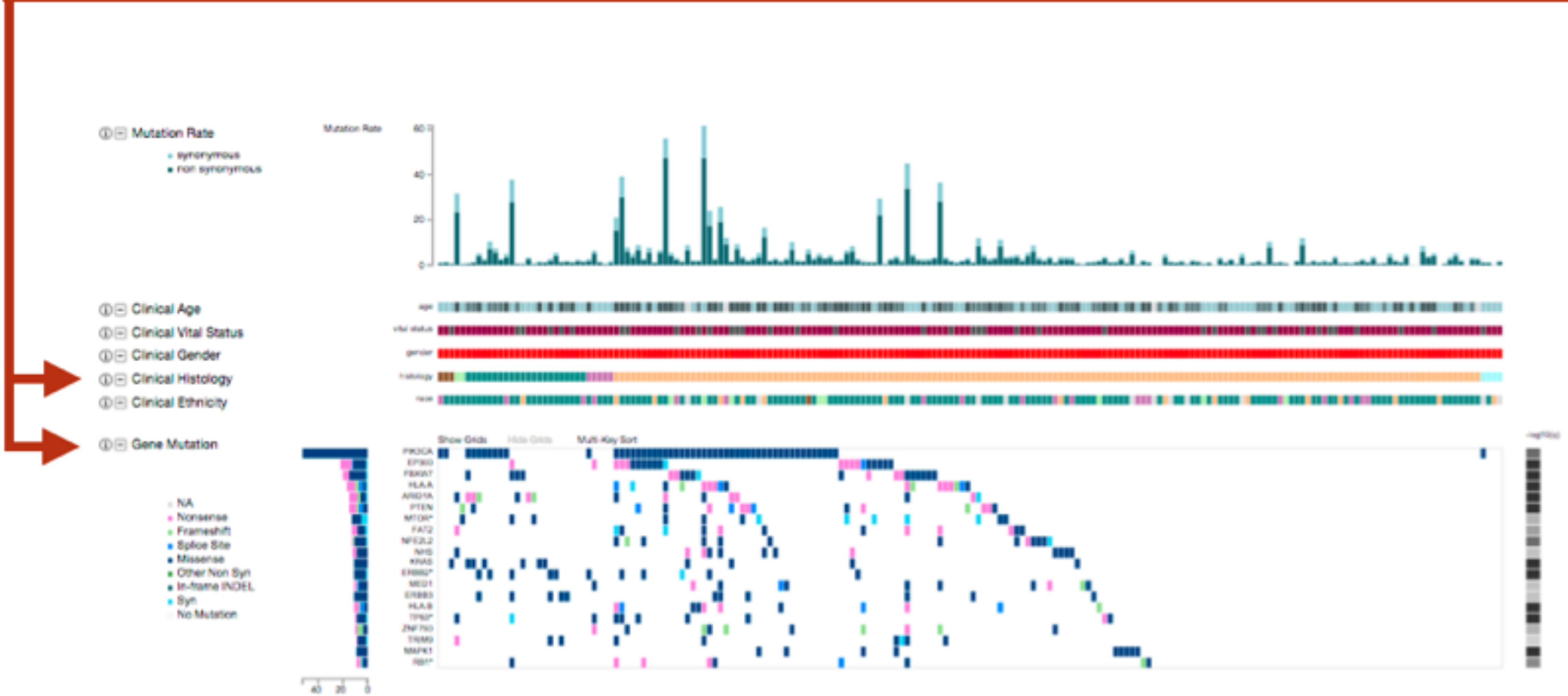


UVM

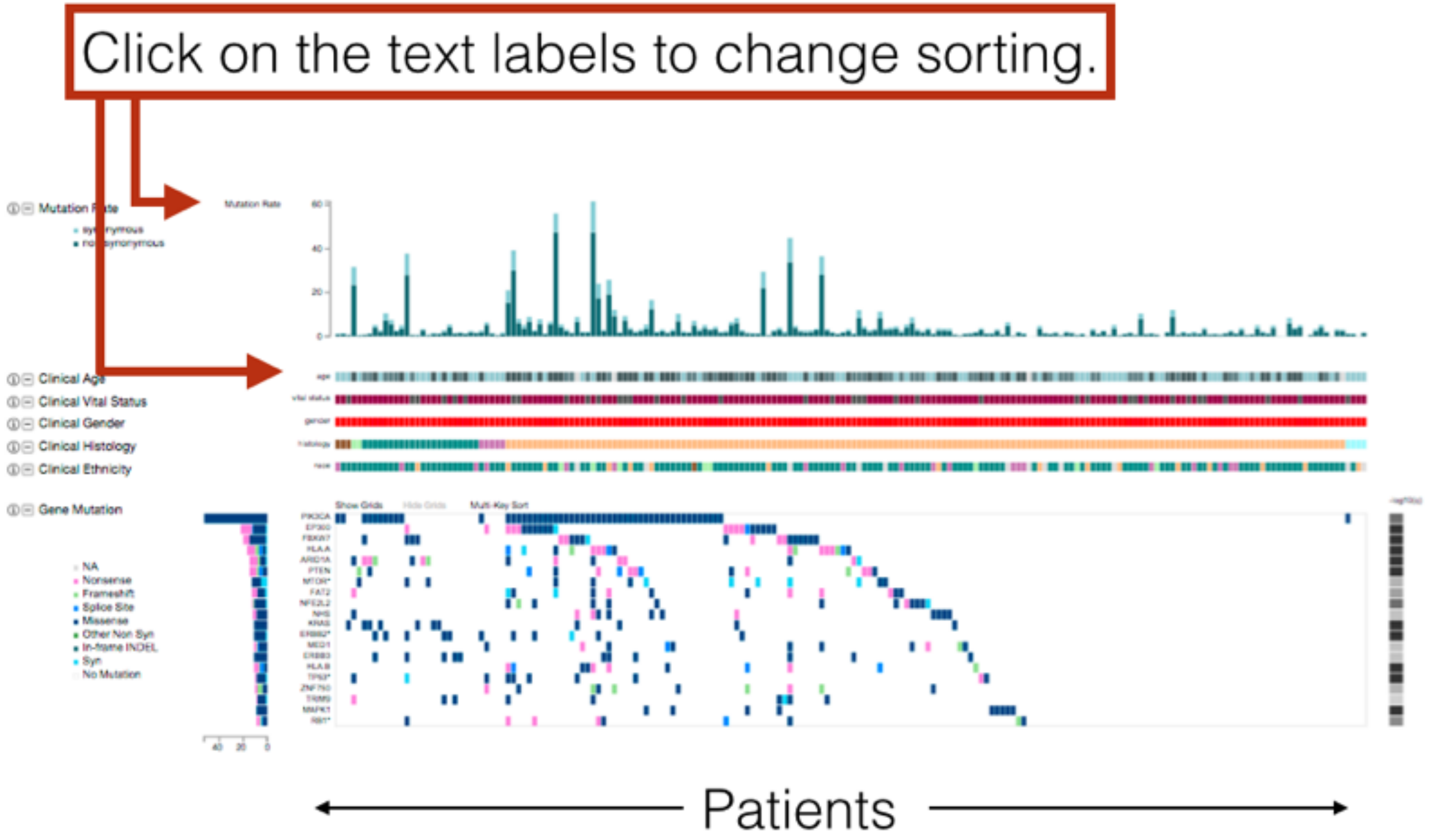
… into this

firebrowse.org/iCoMut/?cohort=UVM

By default, patients are sorted by histology and gene mutation



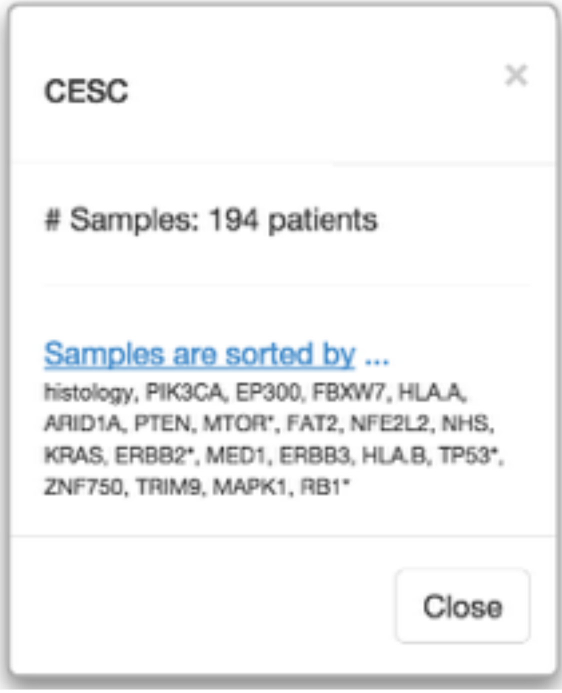
← Patients →



Click on the text labels to change sorting.

Patients

Sort status of samples is reported in the info box



CESC

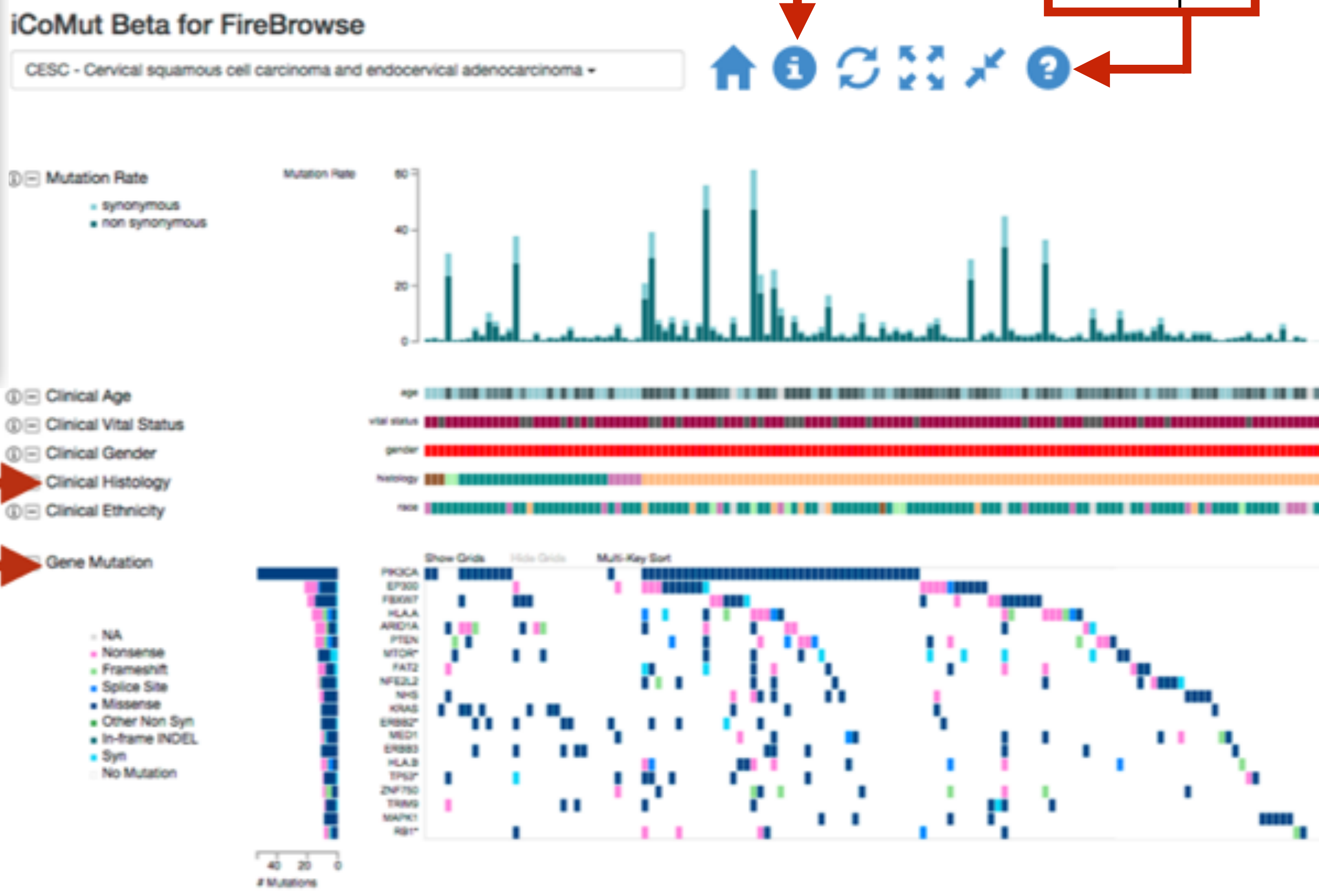
Samples: 194 patients

Samples are sorted by ...

histology, PIK3CA, EP300, FBXW7, HLA.A, ARID1A, PTEN, MTOR*, FAT2, NFE2L2, NHS, KRAS, ERBB2*, MED1, ERBB3, HLA.B, TP53*, ZNF750, TRIM9, MAPK1, RB1*

Close

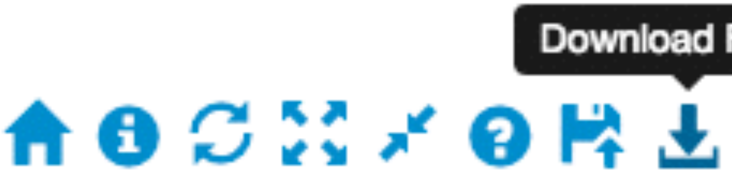
General Help



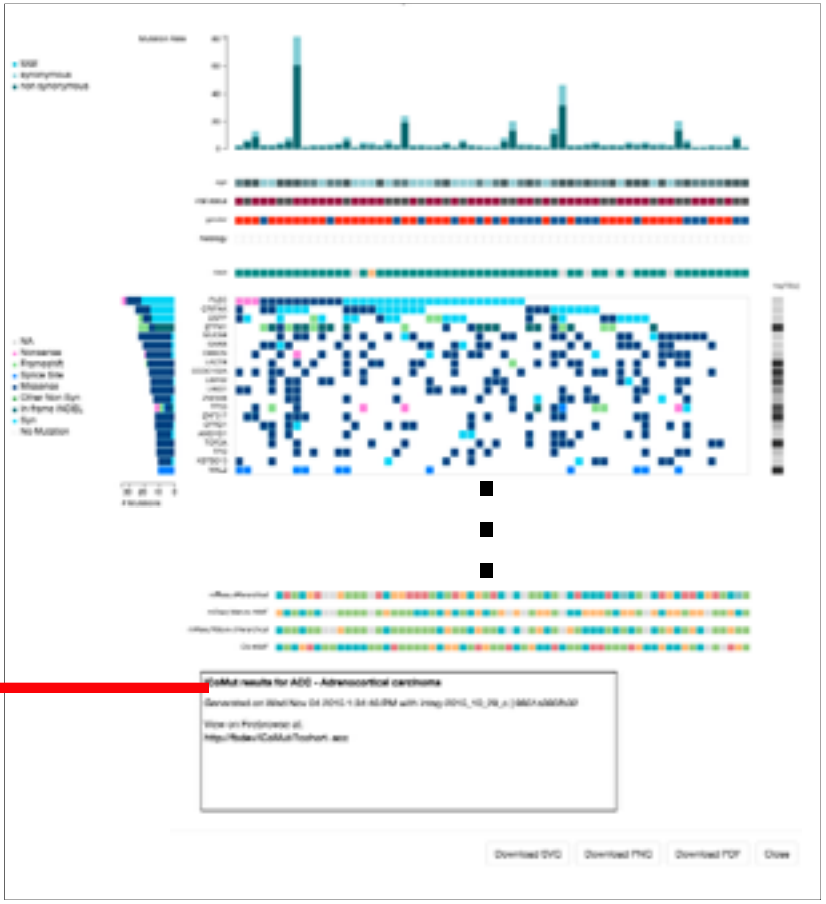
iCoMut Beta for FireBrowse

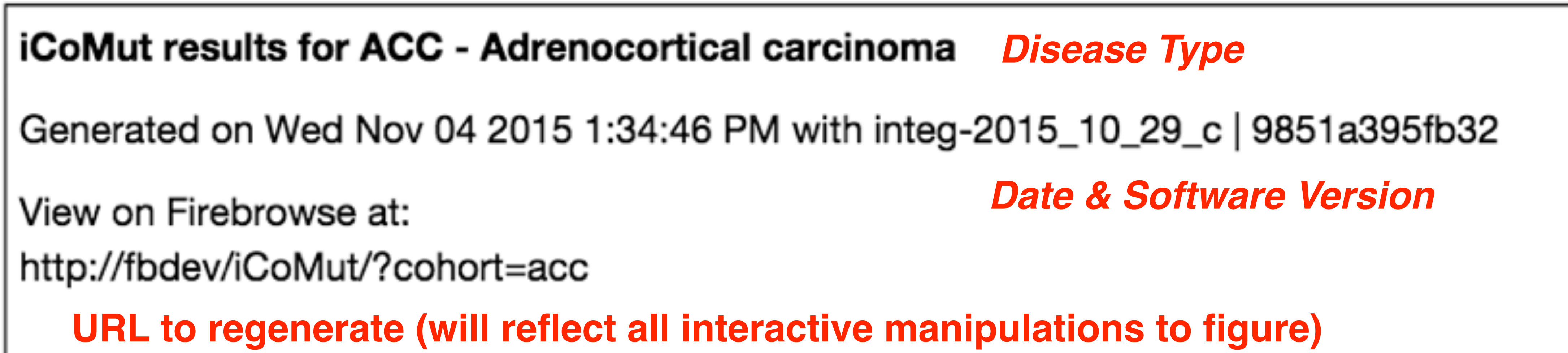
CESC - Cervical squamous cell carcinoma and endocervical adenocarcinoma ▾

Push-Button Publication Figure Reproducibility



SVG





iCoMut results for ACC - Adrenocortical carcinoma

Disease Type

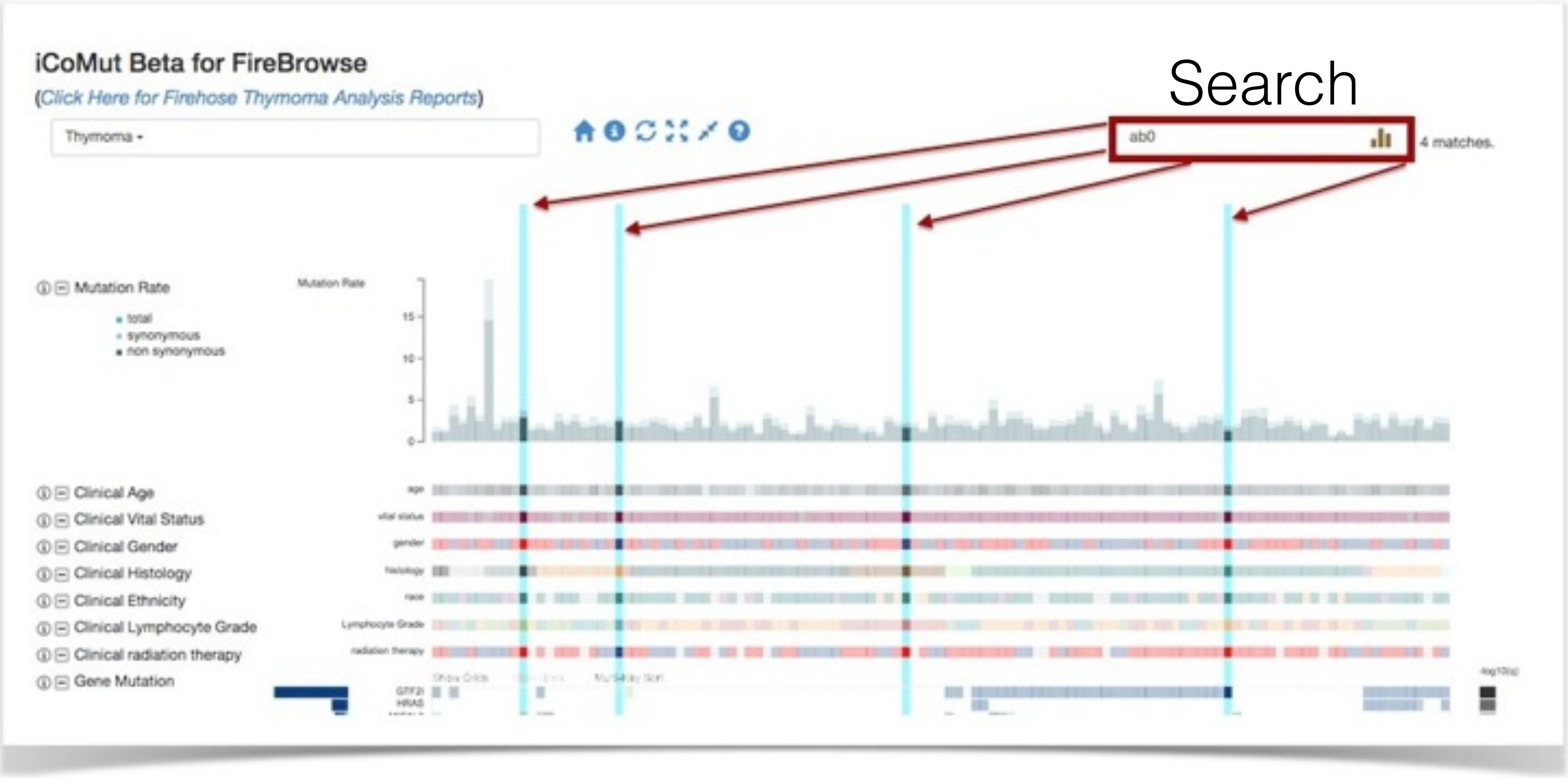
Generated on Wed Nov 04 2015 1:34:46 PM with integ-2015_10_29_c | 9851a395fb32

Date & Software Version

View on Firebrowse at:
http://fbdev/iCoMut/?cohort=acc

URL to regenerate (will reflect all interactive manipulations to figure)

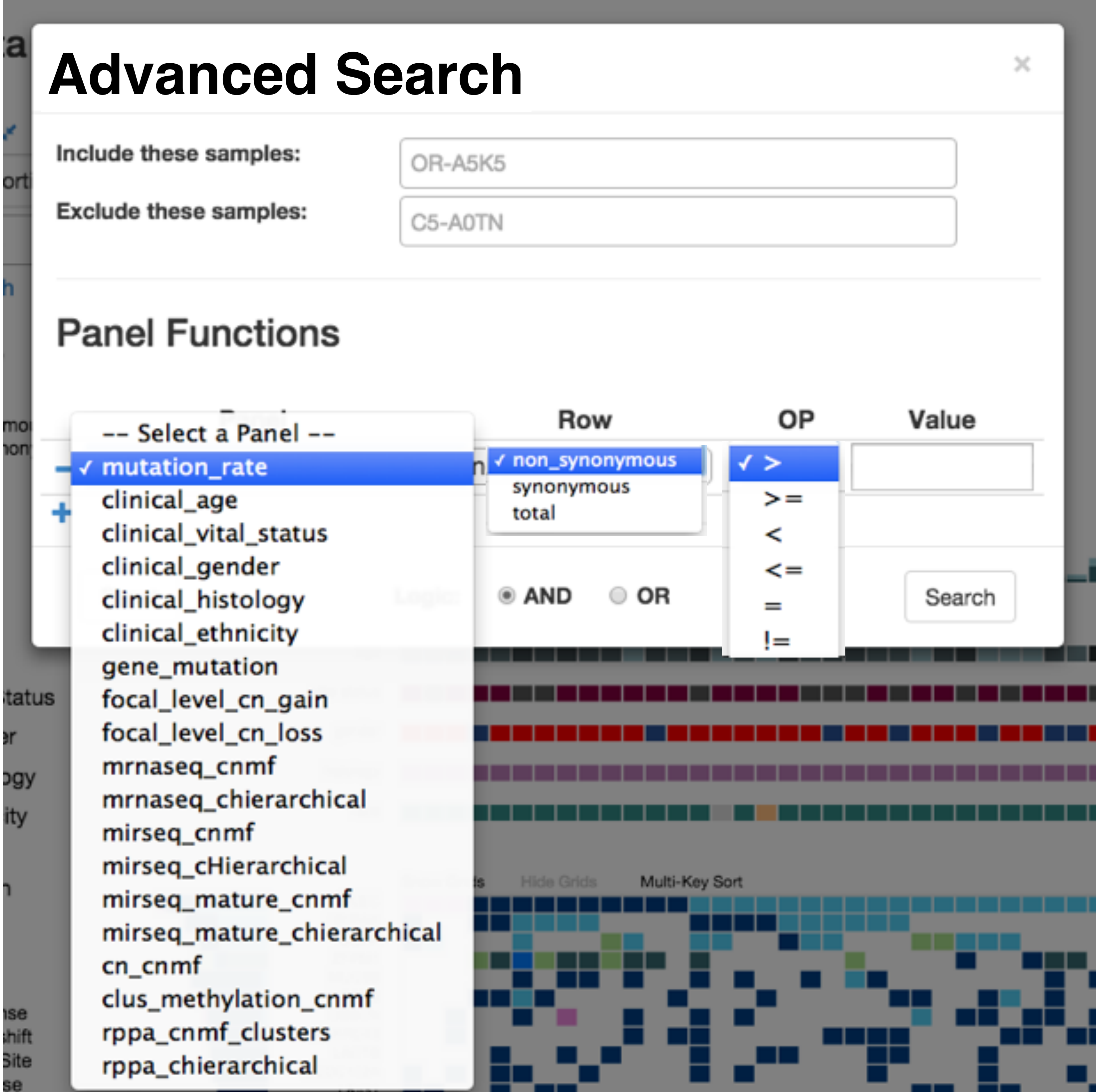
Many more graphical controls …



Example: locate patient/sample of interest

Collaboratively explore questions in realtime on telecons: in what expression cluster does patient X fall?

Without database lookup or scripting, etc



Advanced Search

Include these samples: OR-A5K5

Exclude these samples: C5-A0TN

Panel Functions

Panel: ✓ mutation_rate

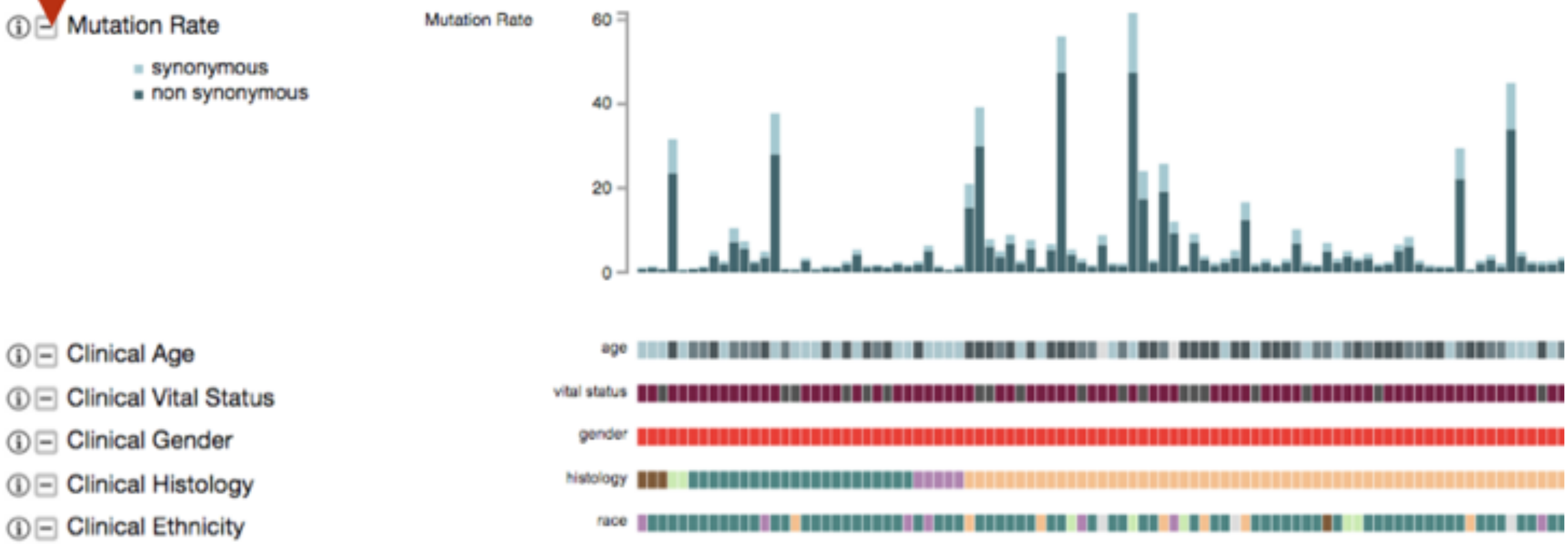
Row: ✓ non_synonymous

OP: ✓ >

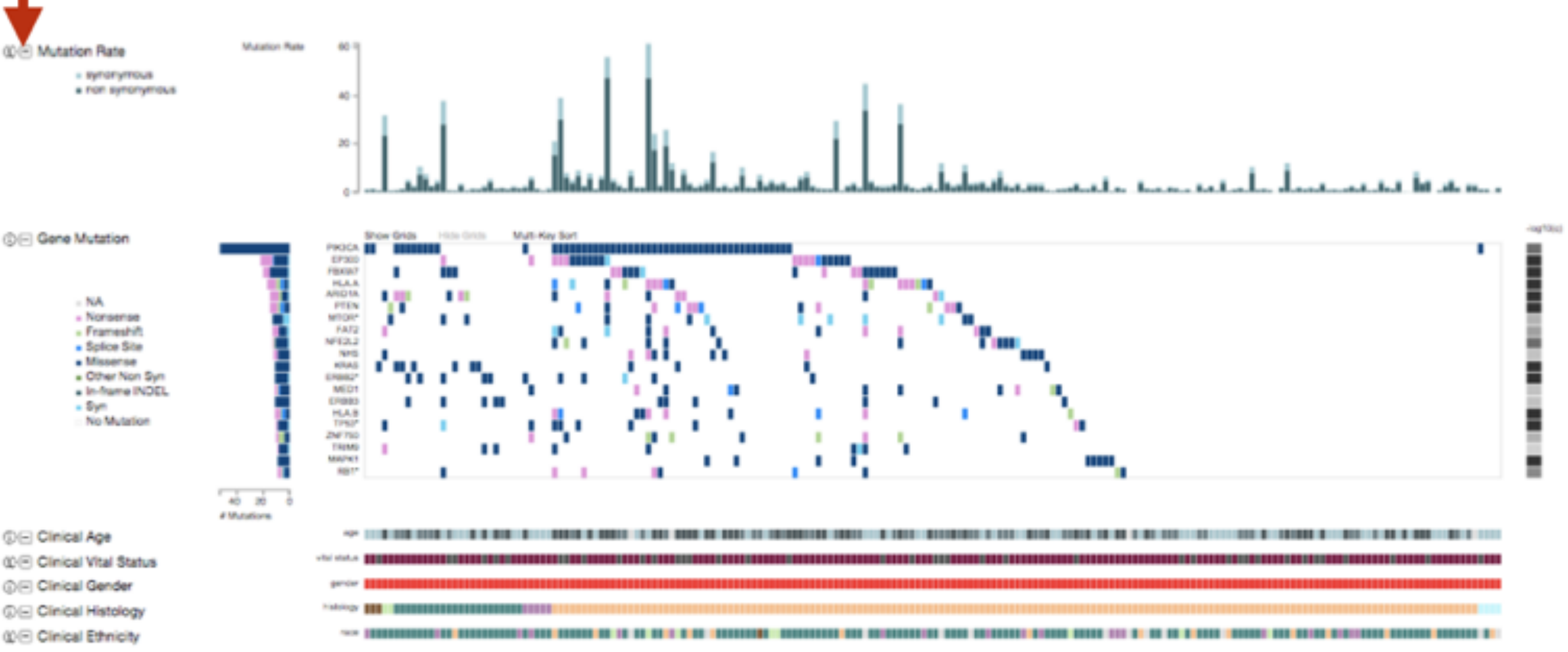
Value:

Logic: ● AND ○ OR

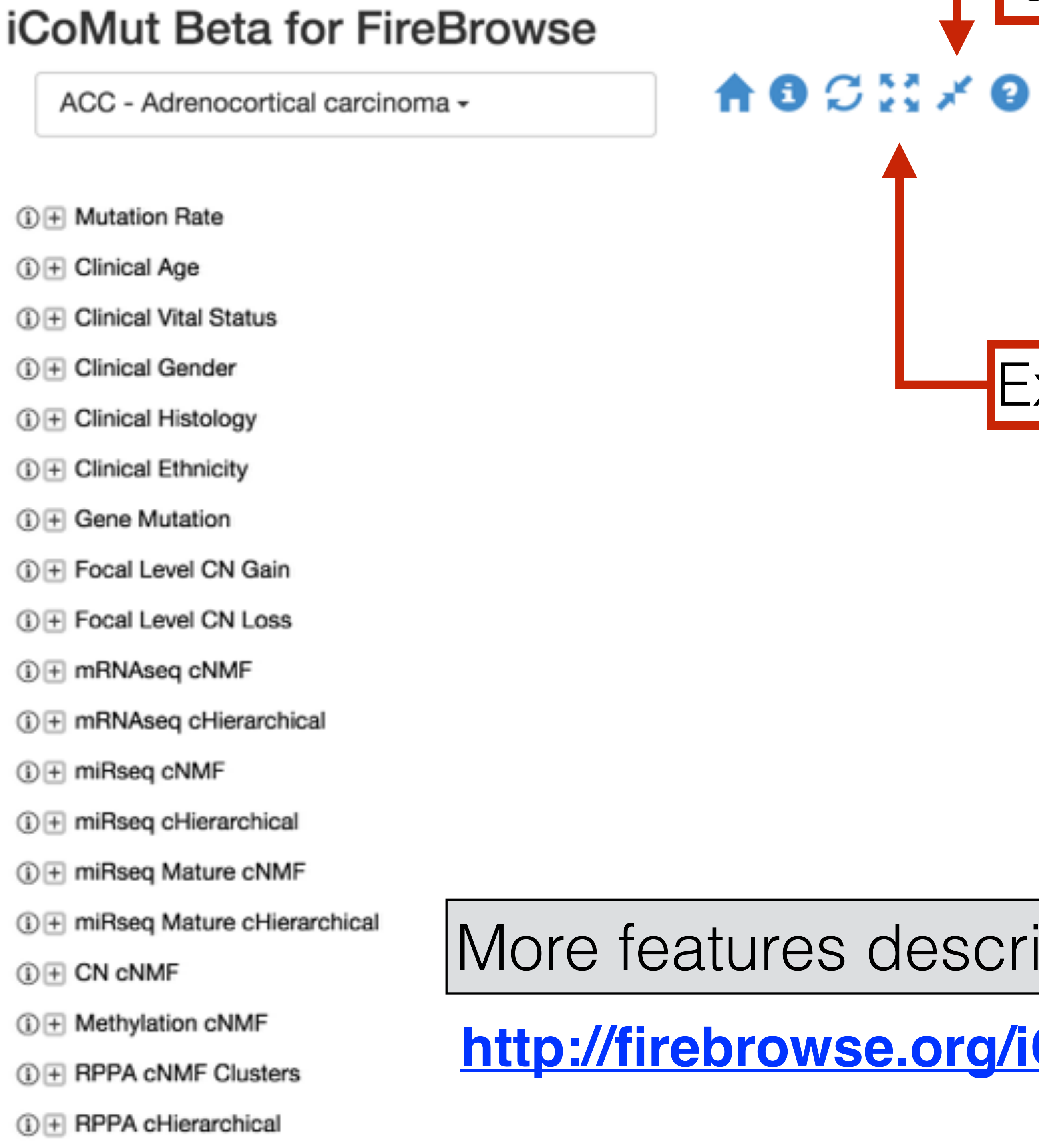
Click on ⊟ to collapse a panel



Drag and drop the ⊟ or ⊞ icon to rearrange the panels



Rearranged panels



iCoMut Beta for FireBrowse

ACC - Adrenocortical carcinoma ▾

- Mutation Rate
- Clinical Age
- Clinical Vital Status
- Clinical Gender
- Clinical Histology
- Clinical Ethnicity
- Gene Mutation
- Focal Level CN Gain
- Focal Level CN Loss
- mRNAseq cNMF
- mRNAseq cHierarchical
- miRseq cNMF
- miRseq cHierarchical
- miRseq Mature cNMF
- miRseq Mature cHierarchical
- CN cNMF
- Methylation cNMF
- RPPA cNMF Clusters
- RPPA cHierarchical

Collapse all panels

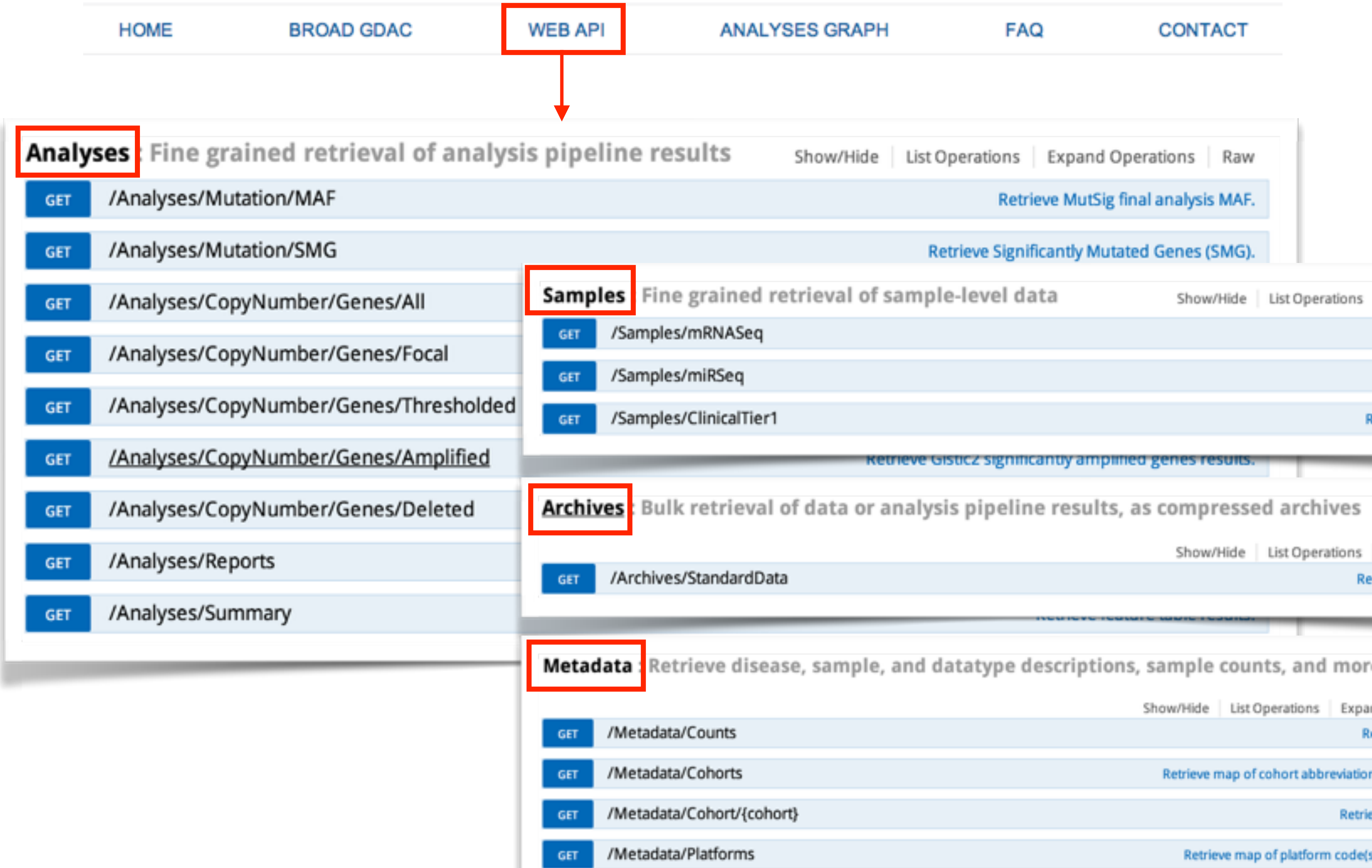
Expand them again

More features described in online help

http://firebrowse.org/iCoMut/#icomutHelp

Programmatic Tools

API-Powered : 25+ RESTful apis in 4 categories



HOME | BROAD GDAC | WEB API | ANALYSES GRAPH | FAQ | CONTACT

Analyses : Fine grained retrieval of analysis pipeline results — Show/Hide | List Operations | Expand Operations | Raw

- GET /Analyses/Mutation/MAF — Retrieve MutSig final analysis MAF.
- GET /Analyses/Mutation/SMG — Retrieve Significantly Mutated Genes (SMG).
- GET /Analyses/CopyNumber/Genes/All
- GET /Analyses/CopyNumber/Genes/Focal
- GET /Analyses/CopyNumber/Genes/Thresholded
- GET /Analyses/CopyNumber/Genes/Amplified — Retrieve Gistic2 significantly amplified genes results.
- GET /Analyses/CopyNumber/Genes/Deleted
- GET /Analyses/Reports
- GET /Analyses/Summary — Retrieve feature table results.

Samples : Fine grained retrieval of sample-level data — Show/Hide | List Operations

- GET /Samples/mRNASeq
- GET /Samples/miRSeq
- GET /Samples/ClinicalTier1 — R

Archives : Bulk retrieval of data or analysis pipeline results, as compressed archives — Show/Hide | List Operations

- GET /Archives/StandardData — Re

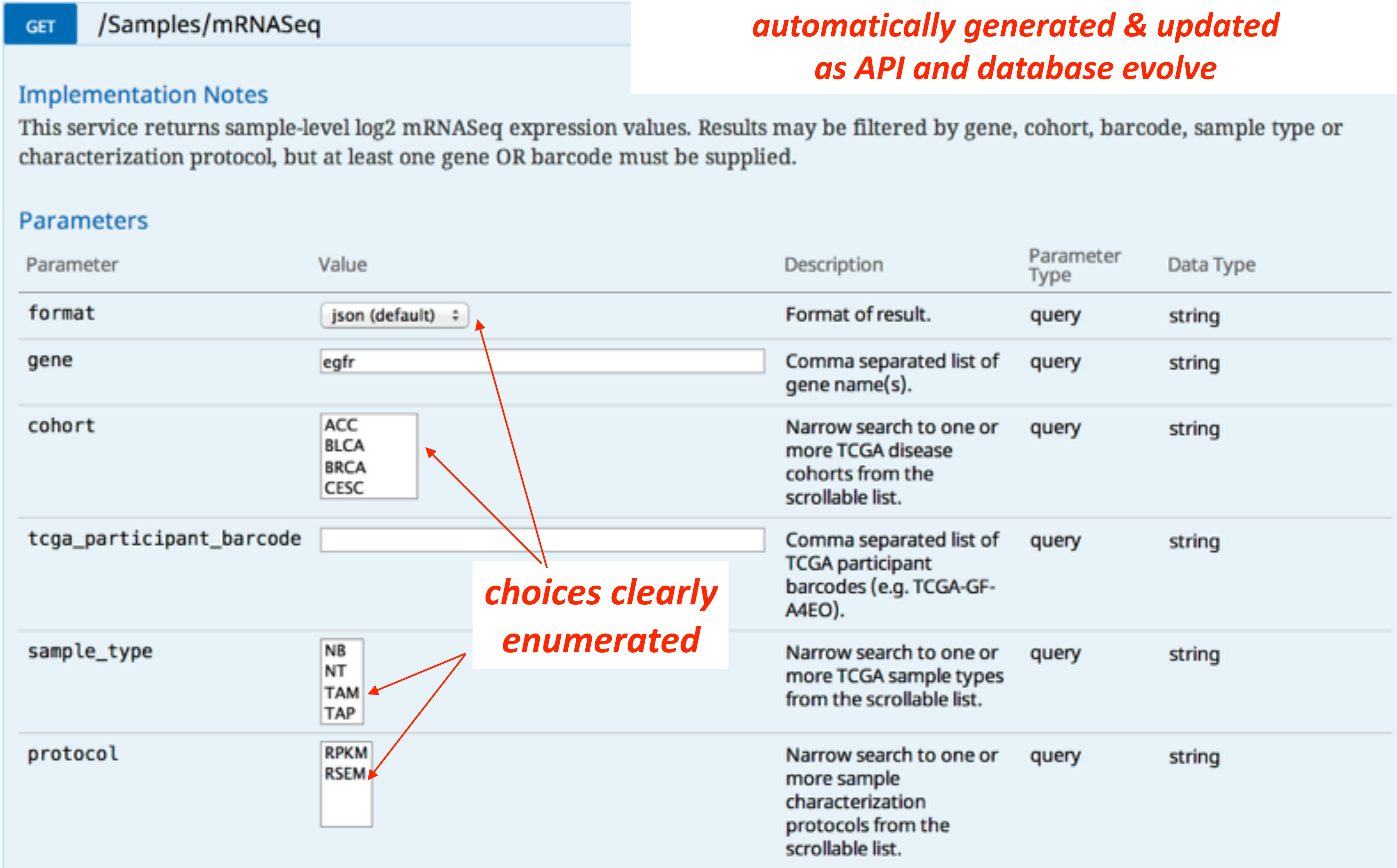
Metadata : Retrieve disease, sample, and datatype descriptions, sample counts, and mor — Show/Hide | List Operations | Expa

- GET /Metadata/Counts — R
- GET /Metadata/Cohorts — Retrieve map of cohort abbreviatior
- GET /Metadata/Cohort/{cohort} — Retrie
- GET /Metadata/Platforms — Retrieve map of platform code(s

Interactive Docs

learn APIs and explore data
by playing in real time
instead of cut/paste from static HTML or PDF

automatically generated & updated
as API and database evolve



GET /Samples/mRNASeq

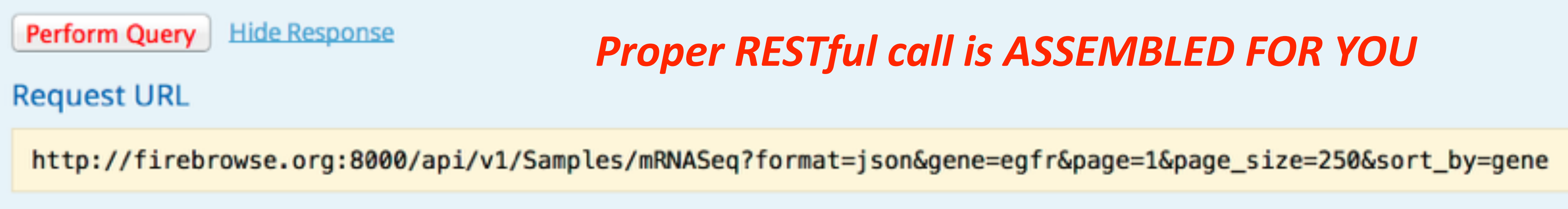
Implementation Notes

This service returns sample-level log2 mRNASeq expression values. Results may be filtered by gene, cohort, barcode, sample type or characterization protocol, but at least one gene OR barcode must be supplied.

Parameters

Parameter	Value	Description	Parameter Type	Data Type
format	json (default)	Format of result.	query	string
gene	egfr	Comma separated list of gene name(s).	query	string
cohort	ACC BLCA BRCA CESC	Narrow search to one or more TCGA disease cohorts from the scrollable list.	query	string
tcga_participant_barcode		Comma separated list of TCGA participant barcodes (e.g. TCGA-GF-A4EO).	query	string
sample_type	NB NT TAM TAP	Narrow search to one or more TCGA sample types from the scrollable list.	query	string
protocol	RPKM RSEM	Narrow search to one or more sample characterization protocols from the scrollable list.	query	string

choices clearly enumerated

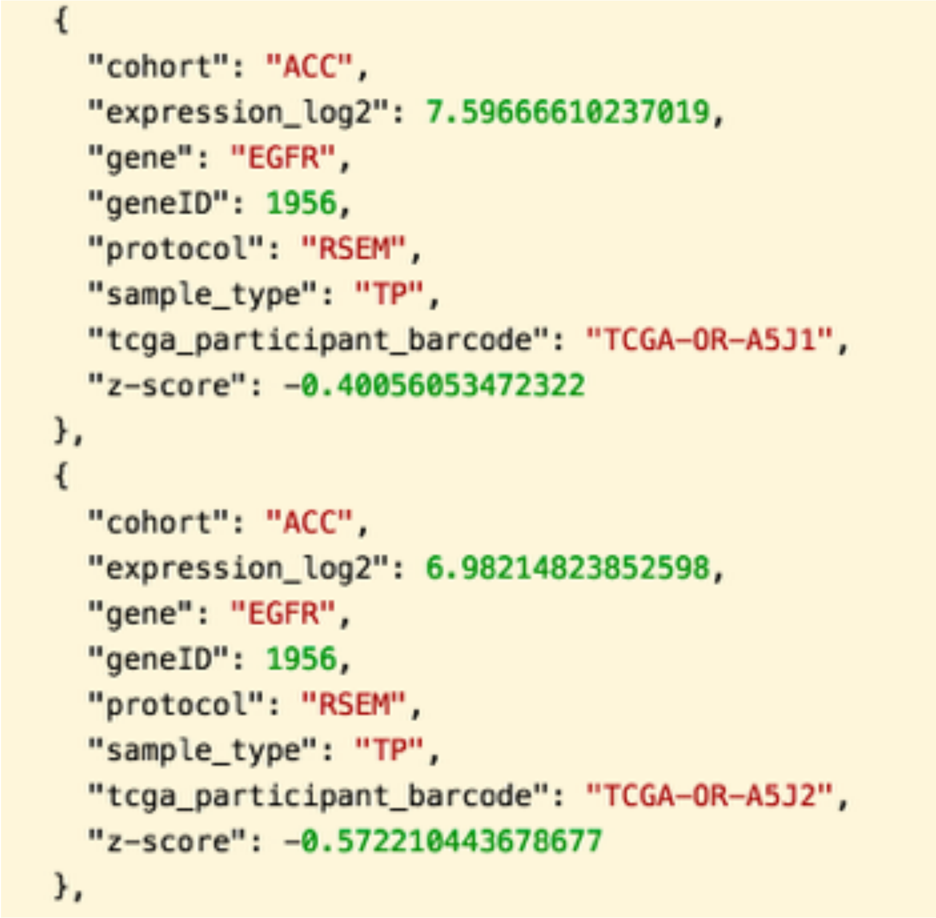


Request URL

```
http://firebrowse.org:8000/api/v1/Samples/mRNASeq?format=json&gene=egfr&page=1&page_size=250&sort_by=gene
```

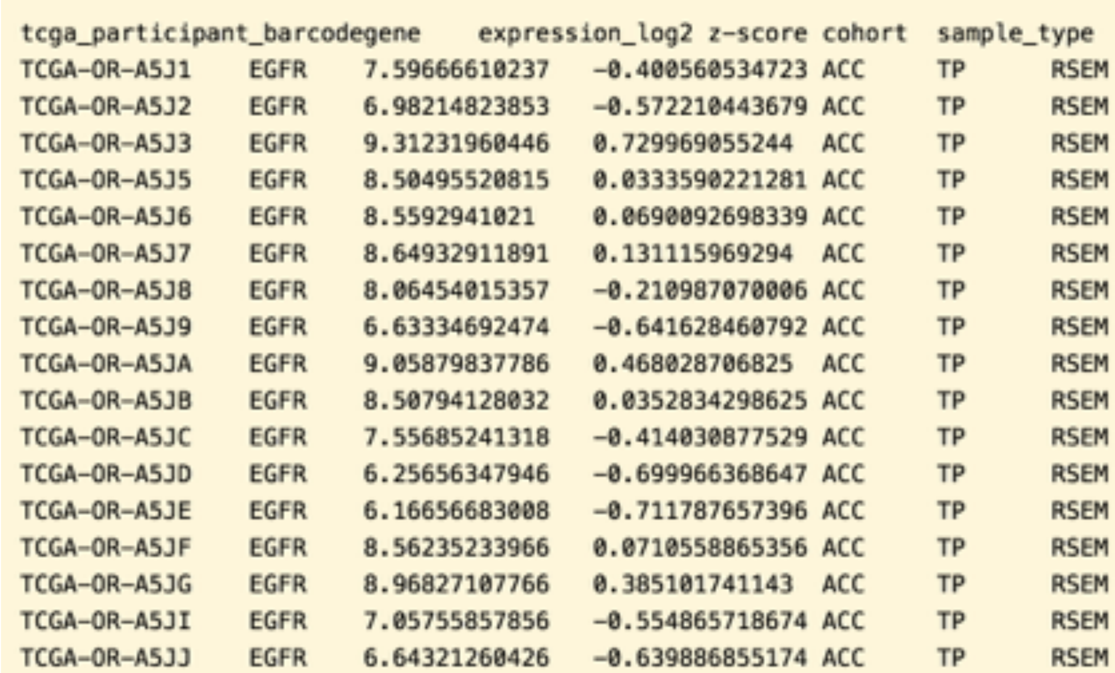
Proper RESTful call is ASSEMBLED FOR YOU

Results returned in multiple formats



```
  {
    "cohort": "ACC",
    "expression_log2": 7.59666610237019,
    "gene": "EGFR",
    "geneID": 1956,
    "protocol": "RSEM",
    "sample_type": "TP",
    "tcga_participant_barcode": "TCGA-OR-A5J1",
    "z-score": -0.40056053472322
  },
  {
    "cohort": "ACC",
    "expression_log2": 6.98214823852598,
    "gene": "EGFR",
    "geneID": 1956,
    "protocol": "RSEM",
    "sample_type": "TP",
    "tcga_participant_barcode": "TCGA-OR-A5J2",
    "z-score": -0.572210443678677
  },
```

JSON for computers/programmers



```
tcga_participant_barcodegene    expression_log2 z-score cohort  sample_type
TCGA-OR-A5J1    EGFR    7.59666610237   -0.400560534723 ACC     TP      RSEM
TCGA-OR-A5J2    EGFR    6.98214823853   -0.572210443679 ACC     TP      RSEM
TCGA-OR-A5J3    EGFR    9.31231960446   0.729969055244  ACC     TP      RSEM
TCGA-OR-A5J5    EGFR    8.50495520815   0.0333590221281 ACC     TP      RSEM
TCGA-OR-A5J6    EGFR    8.5592941021    0.0690092698339 ACC     TP      RSEM
TCGA-OR-A5J7    EGFR    8.64932911891   0.131115969294  ACC     TP      RSEM
TCGA-OR-A5J8    EGFR    8.06454015357   -0.210987070006 ACC     TP      RSEM
TCGA-OR-A5J9    EGFR    6.63334692474   -0.641628460792 ACC     TP      RSEM
TCGA-OR-A5JA    EGFR    9.05879837786   0.468028706825  ACC     TP      RSEM
TCGA-OR-A5JB    EGFR    8.50794128032   0.0352834298625 ACC     TP      RSEM
TCGA-OR-A5JC    EGFR    7.55685241318   -0.414030877529 ACC     TP      RSEM
TCGA-OR-A5JD    EGFR    6.25656347946   -0.699966368647 ACC     TP      RSEM
TCGA-OR-A5JE    EGFR    6.16656683008   -0.711787657396 ACC     TP      RSEM
TCGA-OR-A5JF    EGFR    8.56235233966   0.0710558865356 ACC     TP      RSEM
TCGA-OR-A5JG    EGFR    8.96827107766   0.385101741143  ACC     TP      RSEM
TCGA-OR-A5JI    EGFR    7.05755857856   -0.554865718674 ACC     TP      RSEM
TCGA-OR-A5JJ    EGFR    6.64321260426   -0.639886855174 ACC     TP      RSEM
```

TSV, CSV for scientists, algorithms

Even Easier in Python, R, and UNIX

fbget

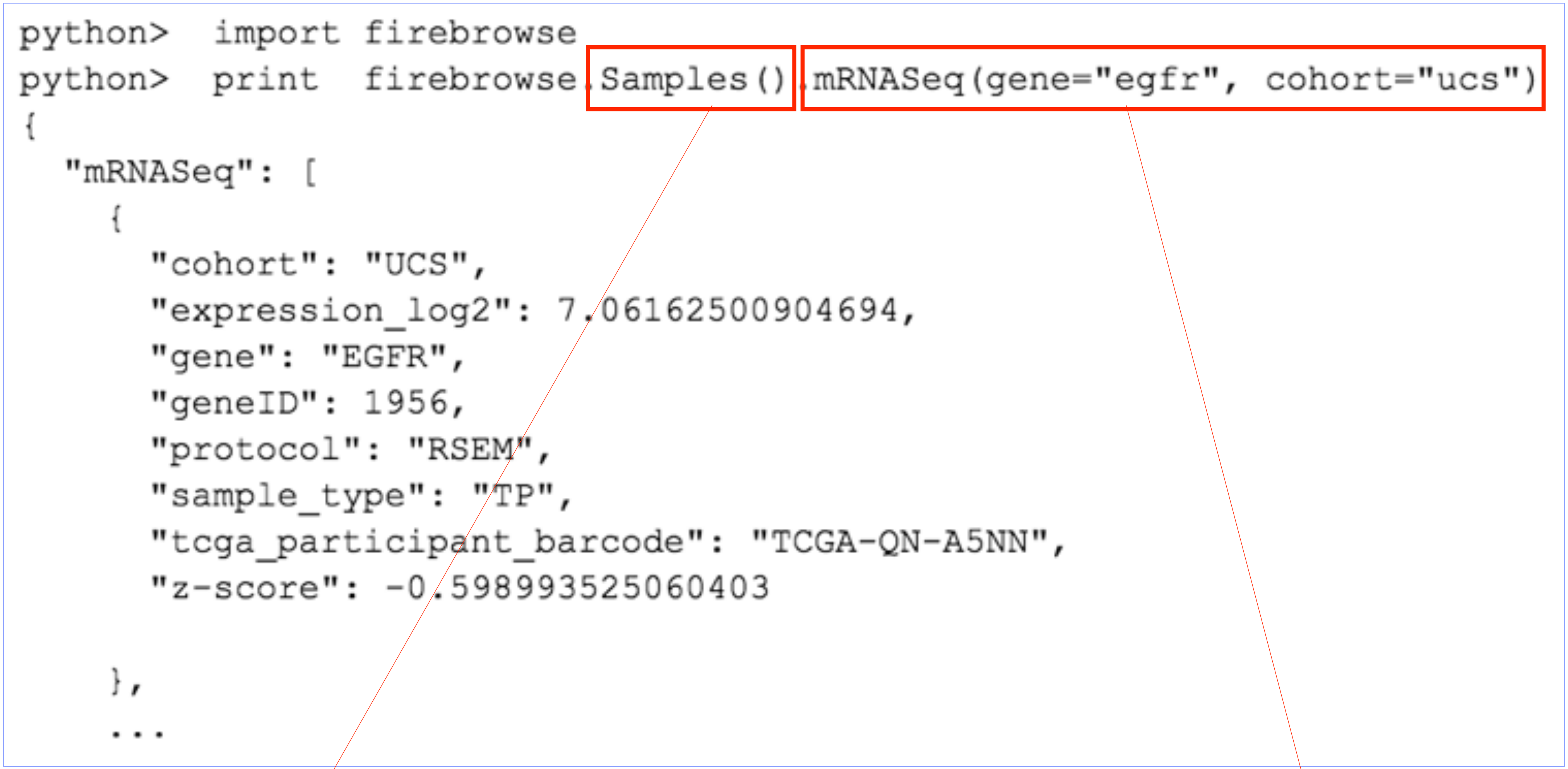
- Low-level Python bindings: 1-1 with RESTful api
- Higher-level interface, for easy/common bioinformatics
- UNIX command line interface, too
- Automatically generated, easily synched with RESTful API
- Flexible, copiously documented and tested
- BSD-style open source license

Download

FireBrowseR : bindings for R

https://github.com/mariodeng/FirebrowseR

fbget : low level interface

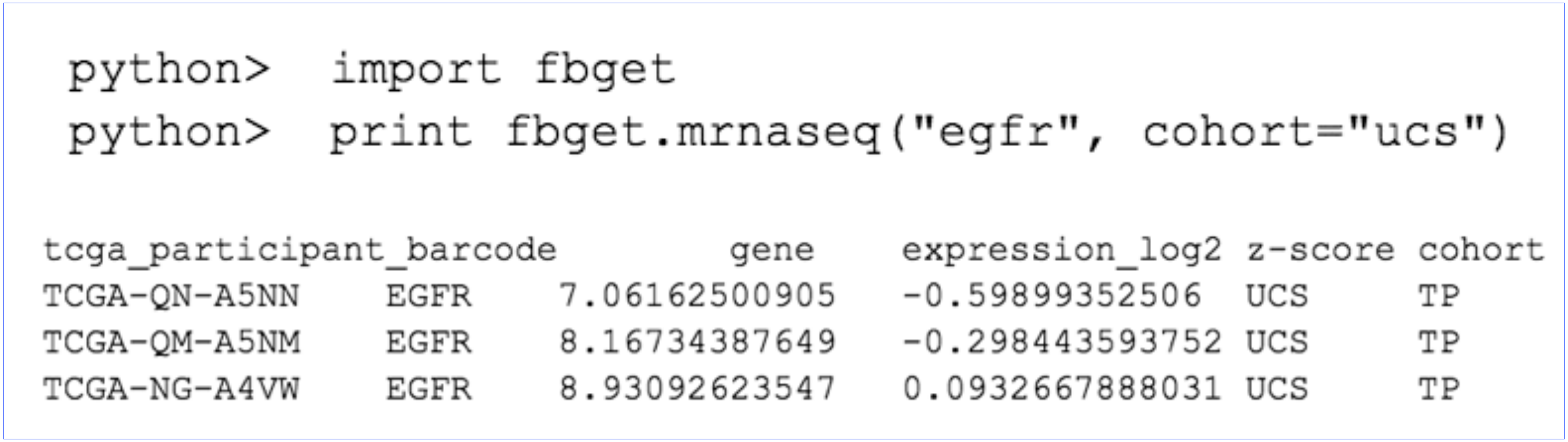


```
python>  import firebrowse
python>  print  firebrowse.Samples().mRNASeq(gene="egfr", cohort="ucs")
{
  "mRNASeq": [
    {
      "cohort": "UCS",
      "expression_log2": 7.06162500904694,
      "gene": "EGFR",
      "geneID": 1956,
      "protocol": "RSEM",
      "sample_type": "TP",
      "tcga_participant_barcode": "TCGA-QN-A5NN",
      "z-score": -0.598993525060403

    },
    ...
```

4 classes, one per API category:
Samples, Analyses,
Archives, Metadata

N methods per class, matching
RESTful API; each defaults
to returning 1 page, in JSON

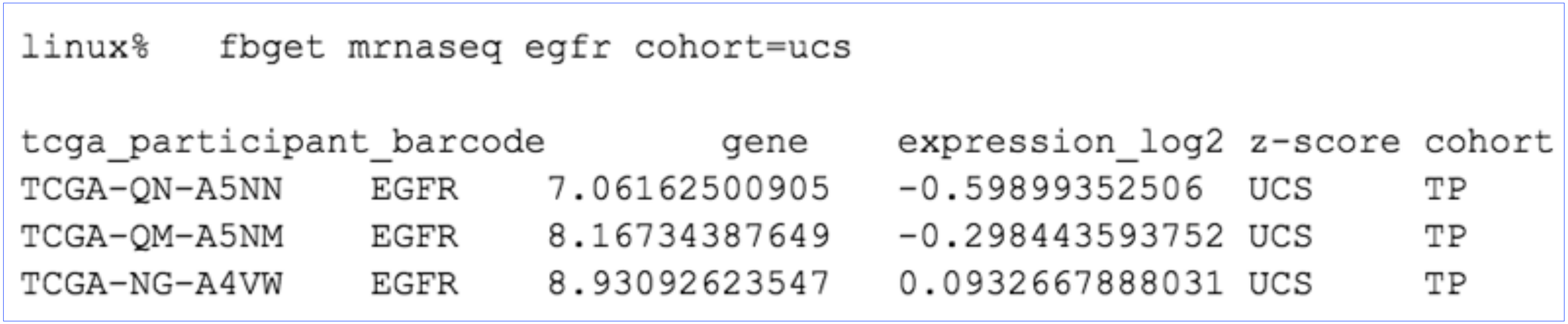
fbget : high level interface



```
 python>  import fbget
 python>  print fbget.mrnaseq("egfr", cohort="ucs")

tcga_participant_barcode        gene    expression_log2 z-score cohort
TCGA-QN-A5NN    EGFR    7.06162500905   -0.59899352506  UCS     TP
TCGA-QM-A5NM    EGFR    8.16734387649   -0.298443593752 UCS     TP
TCGA-NG-A4VW    EGFR    8.93092623547   0.0932667888031 UCS     TP
```

- Simpler, e.g. objects do not need to be instantiated
- Intuitive defaults for common bioinformatic use cases
- Transparently iterates:
  - ✓ To retrieve all pages of results in 1 call
  - ✓ In TSV format

fbget : UNIX CLI interface



```
linux%   fbget mrnaseq egfr cohort=ucs

tcga_participant_barcode        gene    expression_log2 z-score cohort
TCGA-QN-A5NN    EGFR    7.06162500905   -0.59899352506  UCS     TP
TCGA-QM-A5NM    EGFR    8.16734387649   -0.298443593752 UCS     TP
TCGA-NG-A4VW    EGFR    8.93092623547   0.0932667888031 UCS     TP
```

Because sometimes even writing just a couple of lines of Python takes too long

Example: quickly list patients

All of TCGA

```
linux%  fbget patients

tcga_participant_barcode    date     cohort
TCGA-PK-A5H9    2015-04-02 00:00:00 ACC
TCGA-PA-A5YG    2015-04-02 00:00:00 ACC
TCGA-OR-A5JD    2015-04-02 00:00:00 ACC
TCGA-P6-A5OF    2015-04-02 00:00:00 ACC
TCGA-P6-A5OG    2015-04-02 00:00:00 ACC
```

Or just GBM

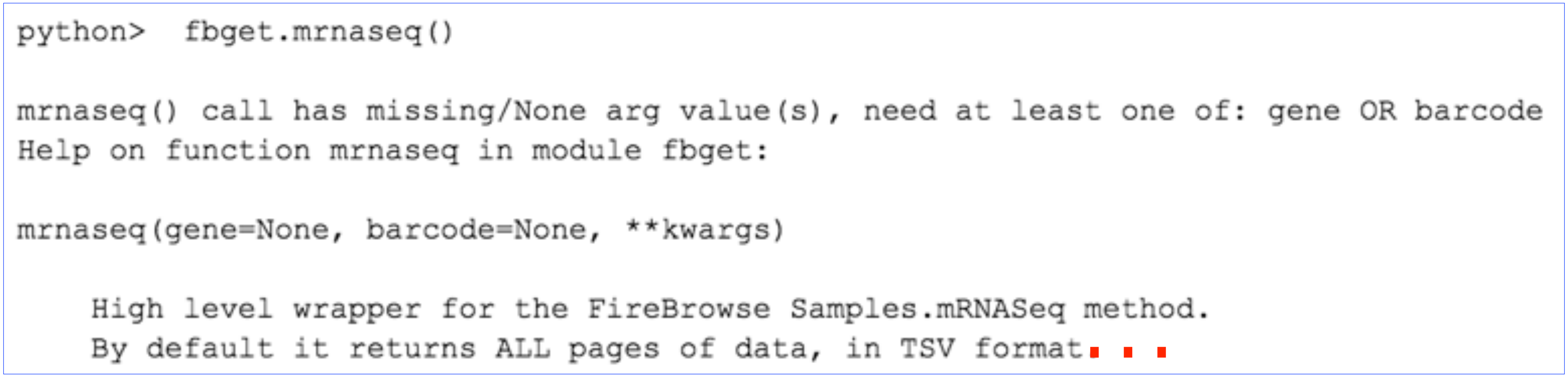
```
linux%  fbget patients  cohort=gbm

tcga_participant_barcode    date     cohort
TCGA-19-4065    2015-04-02 00:00:00 GBM
TCGA-81-5911    2015-04-02 00:00:00 GBM
TCGA-81-5910    2015-04-02 00:00:00 GBM
TCGA-12-1089    2015-04-02 00:00:00 GBM
```

This can be enhanced to yield platform data matrix, like AWG freeze list

fbget Documentation

- Website
- fbget —examples
- Python help

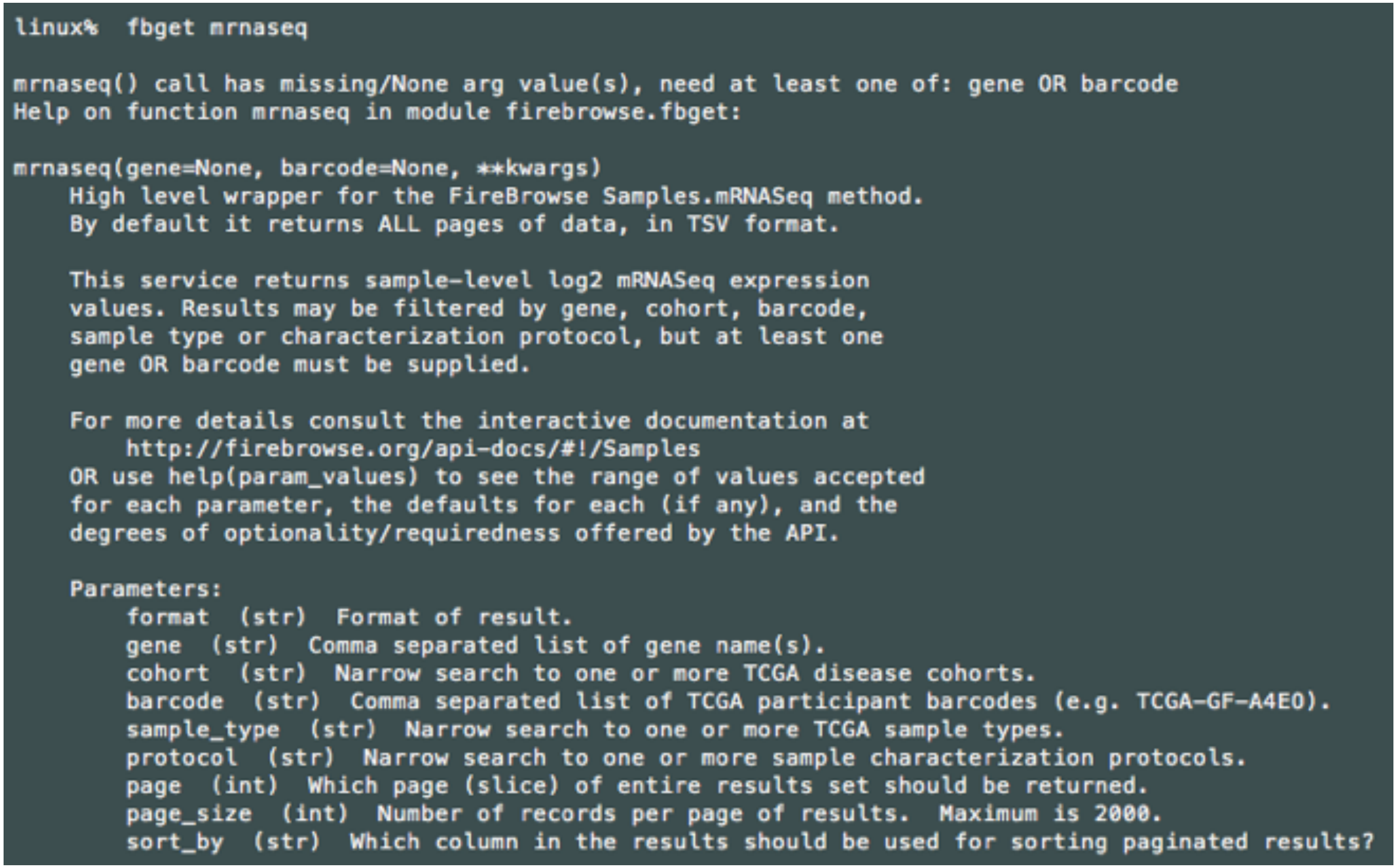
Docs for almost all class methods and functions can also be obtained by invoking the function with zero arguments.



```
python>  fbget.mrnaseq()

mrnaseq() call has missing/None arg value(s), need at least one of: gene OR barcode
Help on function mrnaseq in module fbget:

mrnaseq(gene=None, barcode=None, **kwargs)

    High level wrapper for the FireBrowse Samples.mRNASeq method.
    By default it returns ALL pages of data, in TSV format. . .
```

Better than an inscrutable stack trace, don't you think?

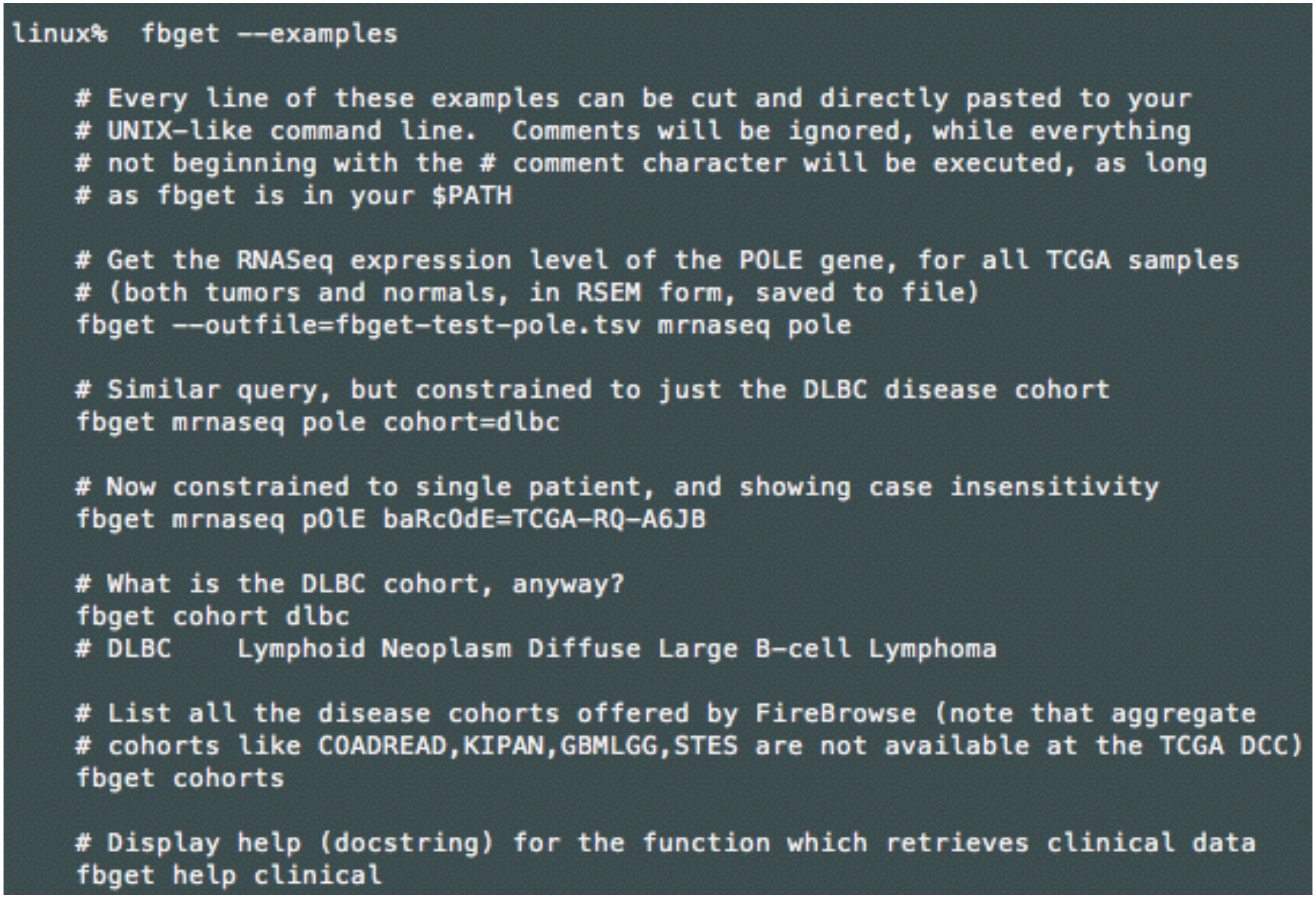
Same is true on UNIX command line



```
linux%  fbget mrnaseq

mrnaseq() call has missing/None arg value(s), need at least one of: gene OR barcode
Help on function mrnaseq in module firebrowse.fbget:

mrnaseq(gene=None, barcode=None, **kwargs)
    High level wrapper for the FireBrowse Samples.mRNASeq method.
    By default it returns ALL pages of data, in TSV format.

    This service returns sample-level log2 mRNASeq expression
    values. Results may be filtered by gene, cohort, barcode,
    sample type or characterization protocol, but at least one
    gene OR barcode must be supplied.

    For more details consult the interactive documentation at
        http://firebrowse.org/api-docs/#!/Samples
    OR use help(param_values) to see the range of values accepted
    for each parameter, the defaults for each (if any), and the
    degrees of optionality/requiredness offered by the API.

    Parameters:
        format  (str)  Format of result.
        gene  (str)  Comma separated list of gene name(s).
        cohort  (str)  Narrow search to one or more TCGA disease cohorts.
        barcode  (str)  Comma separated list of TCGA participant barcodes (e.g. TCGA-GF-A4EO).
        sample_type  (str)  Narrow search to one or more TCGA sample types.
        protocol  (str)  Narrow search to one or more sample characterization protocols.
        page  (int)  Which page (slice) of entire results set should be returned.
        page_size  (int)  Number of records per page of results.  Maximum is 2000.
        sort_by  (str)  Which column in the results should be used for sorting paginated results?
```

Docs obtained by invoking functions with zero arguments

Examples Embedded Directly in Tool



```
linux%  fbget --examples

    # Every line of these examples can be cut and directly pasted to your
    # UNIX-like command line.  Comments will be ignored, while everything
    # not beginning with the # comment character will be executed, as long
    # as fbget is in your $PATH

    # Get the RNASeq expression level of the POLE gene, for all TCGA samples
    # (both tumors and normals, in RSEM form, saved to file)
    fbget --outfile=fbget-test-pole.tsv mrnaseq pole

    # Similar query, but constrained to just the DLBC disease cohort
    fbget mrnaseq pole cohort=dlbc

    # Now constrained to single patient, and showing case insensitivity
    fbget mrnaseq pOlE baRcOdE=TCGA-RQ-A6JB

    # What is the DLBC cohort, anyway?
    fbget cohort dlbc
    # DLBC    Lymphoid Neoplasm Diffuse Large B-cell Lymphoma

    # List all the disease cohorts offered by FireBrowse (note that aggregate
    # cohorts like COADREAD,KIPAN,GBMLGG,STES are not available at the TCGA DCC)
    fbget cohorts

    # Display help (docstring) for the function which retrieves clinical data
    fbget help clinical
```

Fin