

Integrating the Connectivity Map and The Cancer Genome Atlas

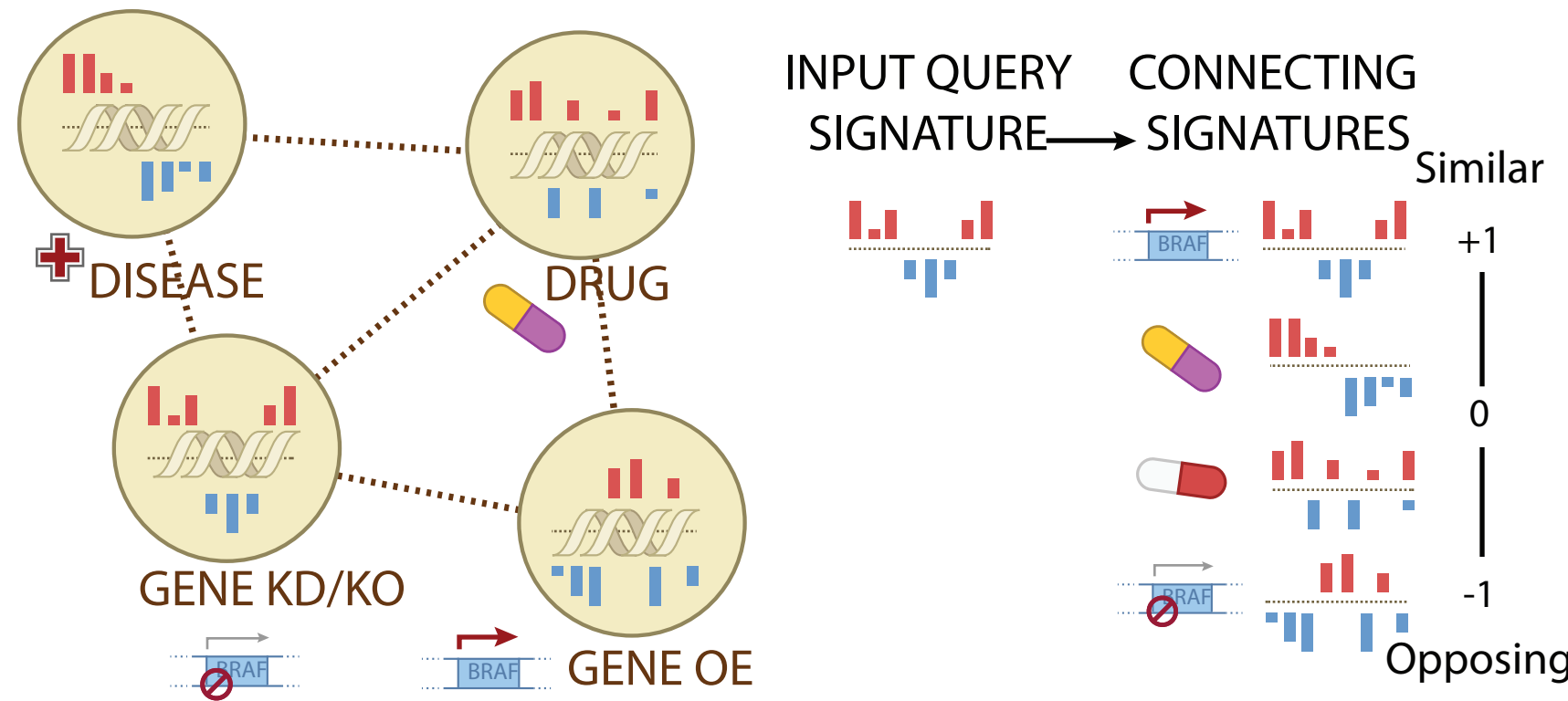
Ben Siranosian, Joseph Nasser, Rajiv Narayan, Aravind Subramanian, Todd Golub

The Broad Institute of MIT and Harvard, Cambridge, MA



The Connectivity Map (CMap)

mRNA expression database to connect genes to drugs to disease
Perturbation with gene knock-down, knock-out, overexpression or drug treatment in 9 core cell lines



The Cancer Genome Atlas (TCGA)

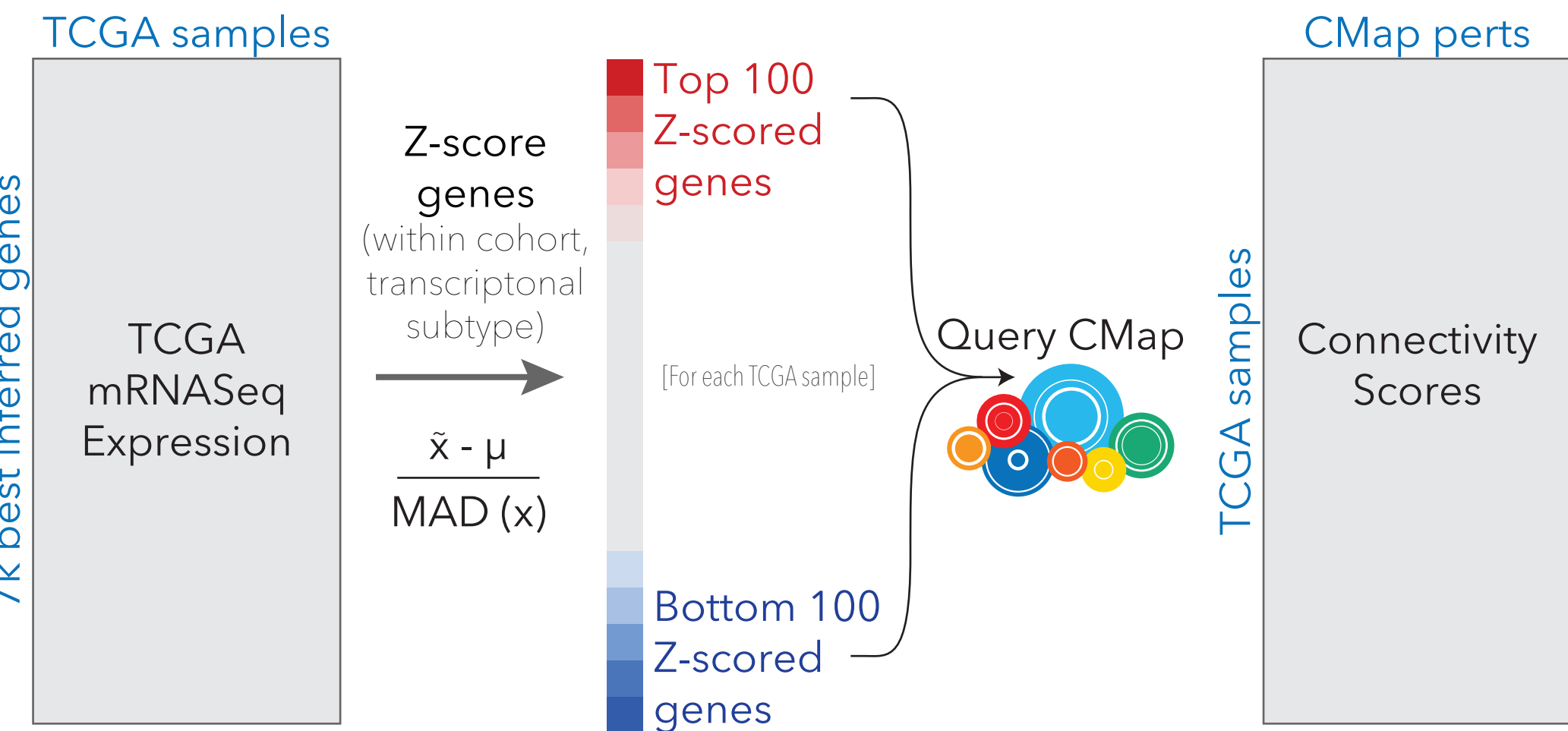
Database of key genomic changes across major types of cancer
- mRNASeq, mutations, copy number alterations
Publicly accessible to researchers around the world
Patient-derived data
High sample numbers - 1093 sequenced breast cancer tumors

How can we integrate the two databases?

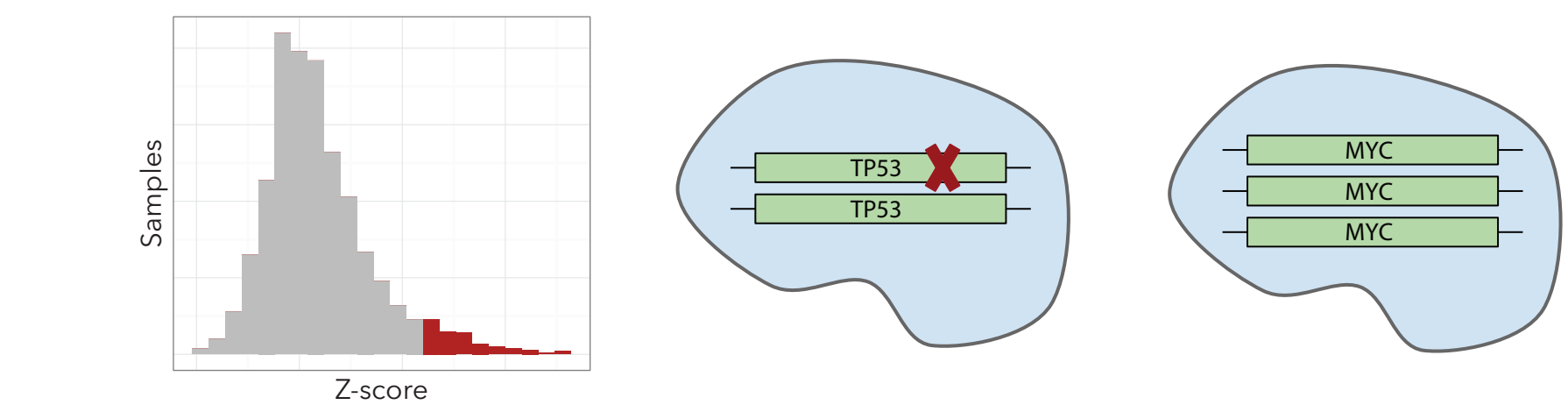
- Method development
 - Data processing, statistics, visualization of results
- Are cell-derived CMap signatures applicable to patient-derived data?
 - Different system, experimental methods
- Can a comparison tell us something new about cancer biology?
 - Sets of patients with alterations
 - Individuals with exceptional characteristics

Methods

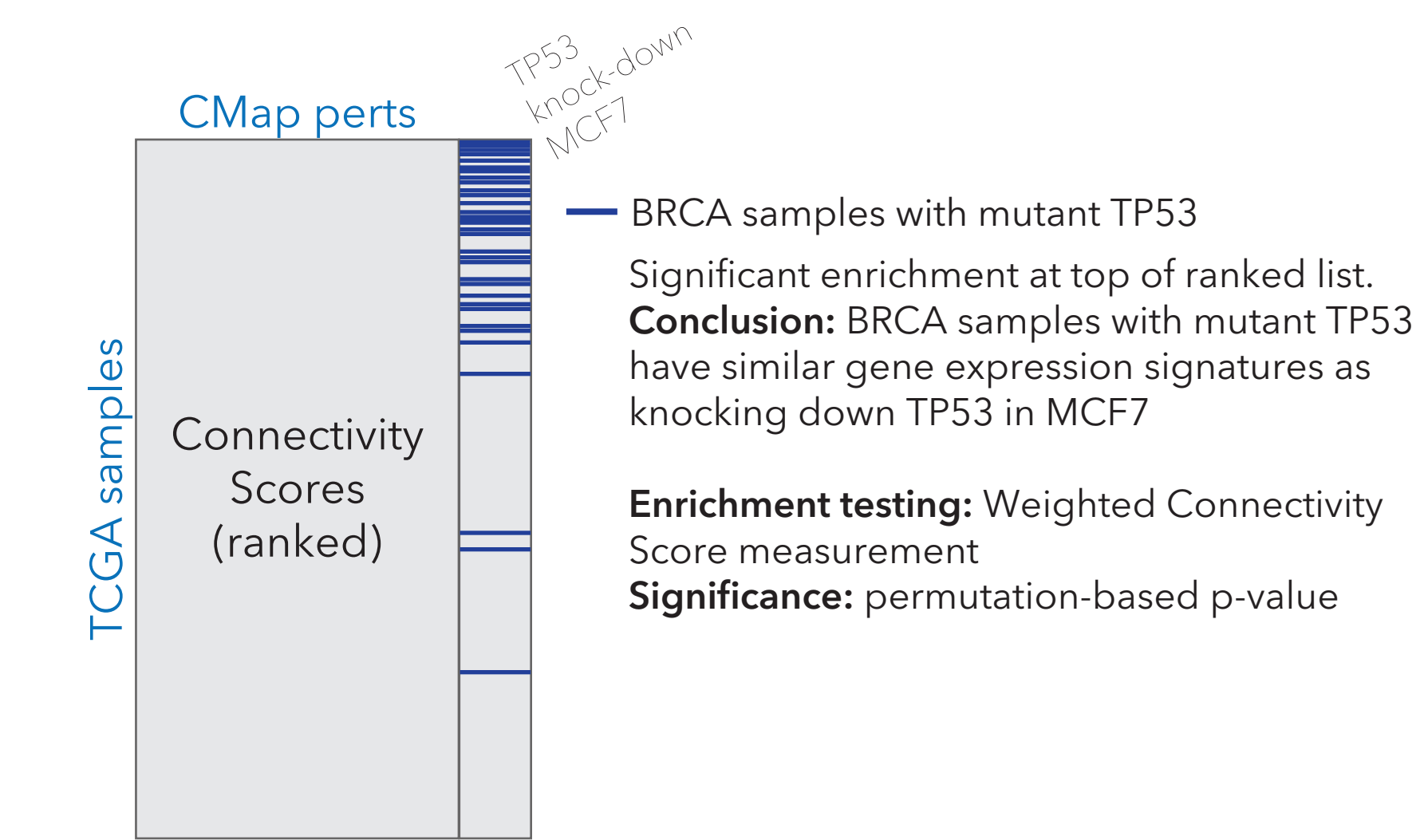
- 1) Define a differential expression signature from each TCGA sample
- 2) CMap query gives a connectivity score to each perturbation experiment in each cell line



- 3) Define a sample set based on characteristic of interest
[outlier gene expression | mutation | copy number alteration]

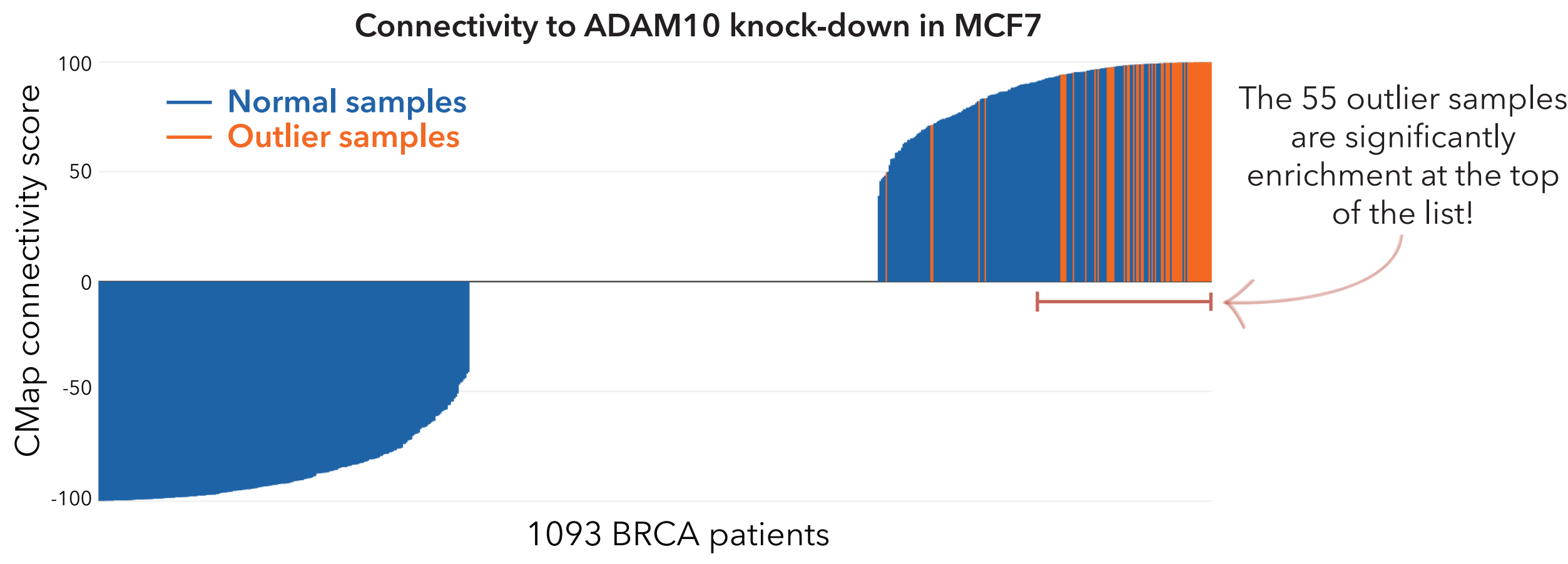
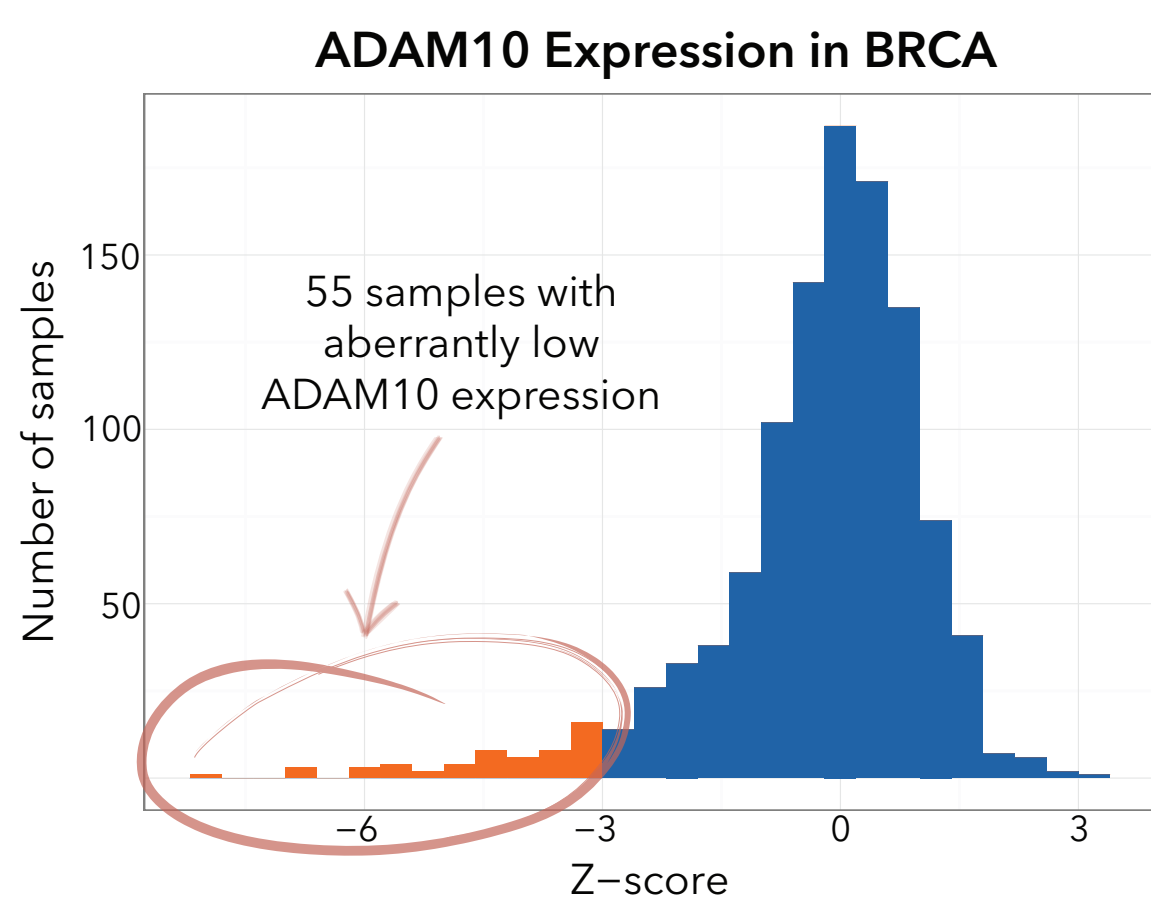


- 4) Test for enrichment at top/bottom of ranked connectivity scores
 - Sets with strong enrichment: alteration is similar to CMap signature



Validation of cell-derived CMap signatures

Expect sets of outlier gene expression to connect strongly to signatures of knock-down or overexpression of the same gene



Low ADAM10 expression looks similar to ADAM10 knock-down

Enrichment score	0.849	- Positive enrichment = similarity
P-value	3.6×10^{-4}	- Validation of CMap signature
Percentile rank	99.6%	- Applicable to patient-derived data
Mean connectivity score	93.42	

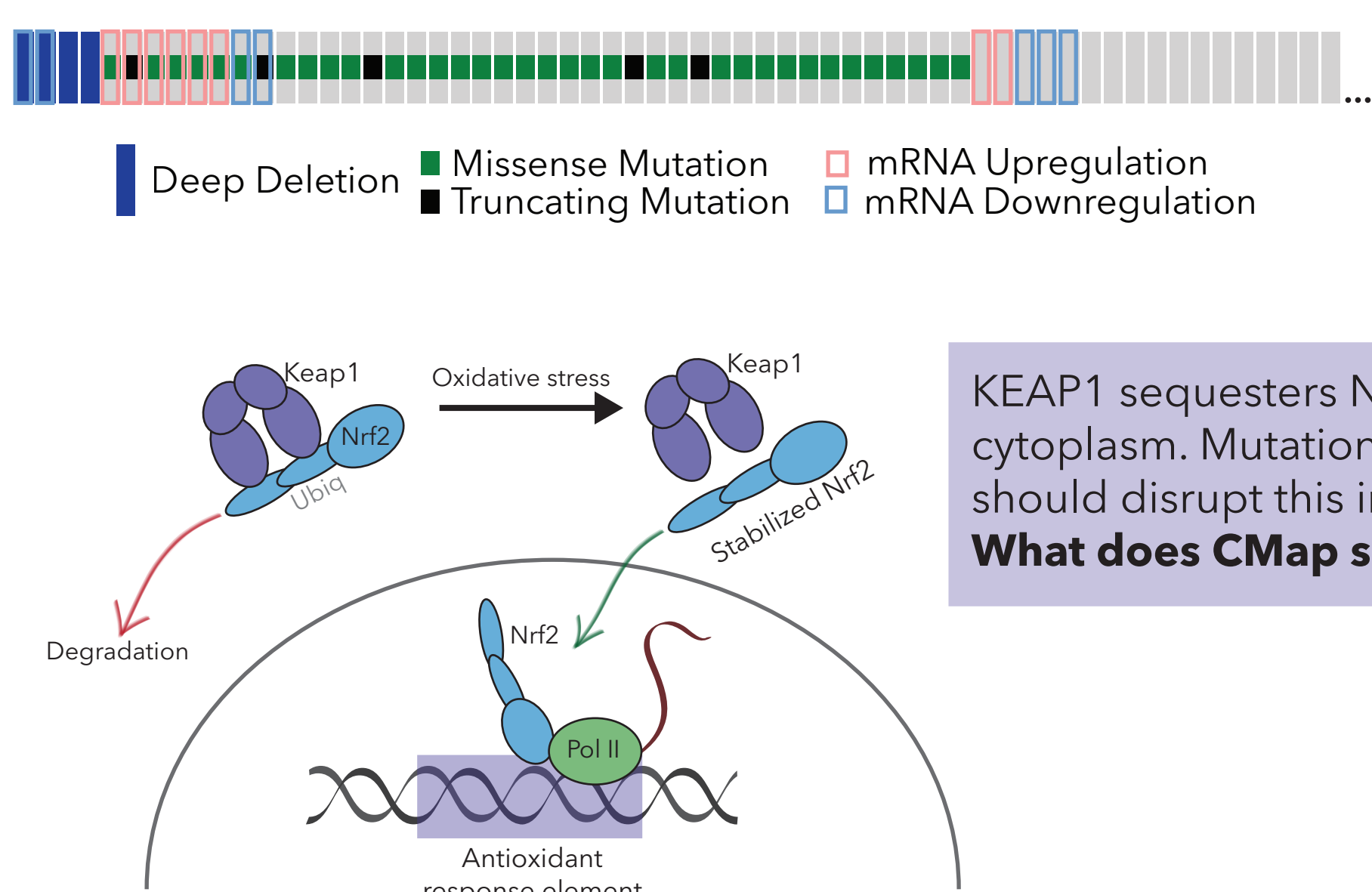
Connection recovery

Outlier sample sets with a matched knock-down or overexpression signature in at least one cell line number significant at $p=0.01$ / total number

BRCA	699/1212	58%
PRAD	613/1131	54%
LUAD	326/847	38%
COAD	78/562	14%
SKCM	131/1058	12%

CMap gives a clue into cancer biology

KEAP1: Altered in 20% of lung adenocarcinoma tumors



Comparing LUAD samples with KEAP1 mutations to CMap

Perturbagen	Enrichment	P-value
KEAP1 Knock-down	0.669	0.0014
NRF2 Knock-down	-0.616	0.0131
NRF2 Overexpression	0.736	0.0605
STK11 Knock-down	0.734	0.0004

Validation of signature: KEAP1 mutants are similar to KEAP1 knock-down
Confirmation of biological interaction: KEAP1 mutants are opposite to NRF2 knock-down, similar to NRF2 overexpression

The #1 connection in all of CMap aggregated signatures: STK11 knock-down. STK11 regulates AMPK signaling - effects on metabolism, energy homeostasis

Is there a link between KEAP1 and STK11 signaling?
Collaborate and experiment to find out.

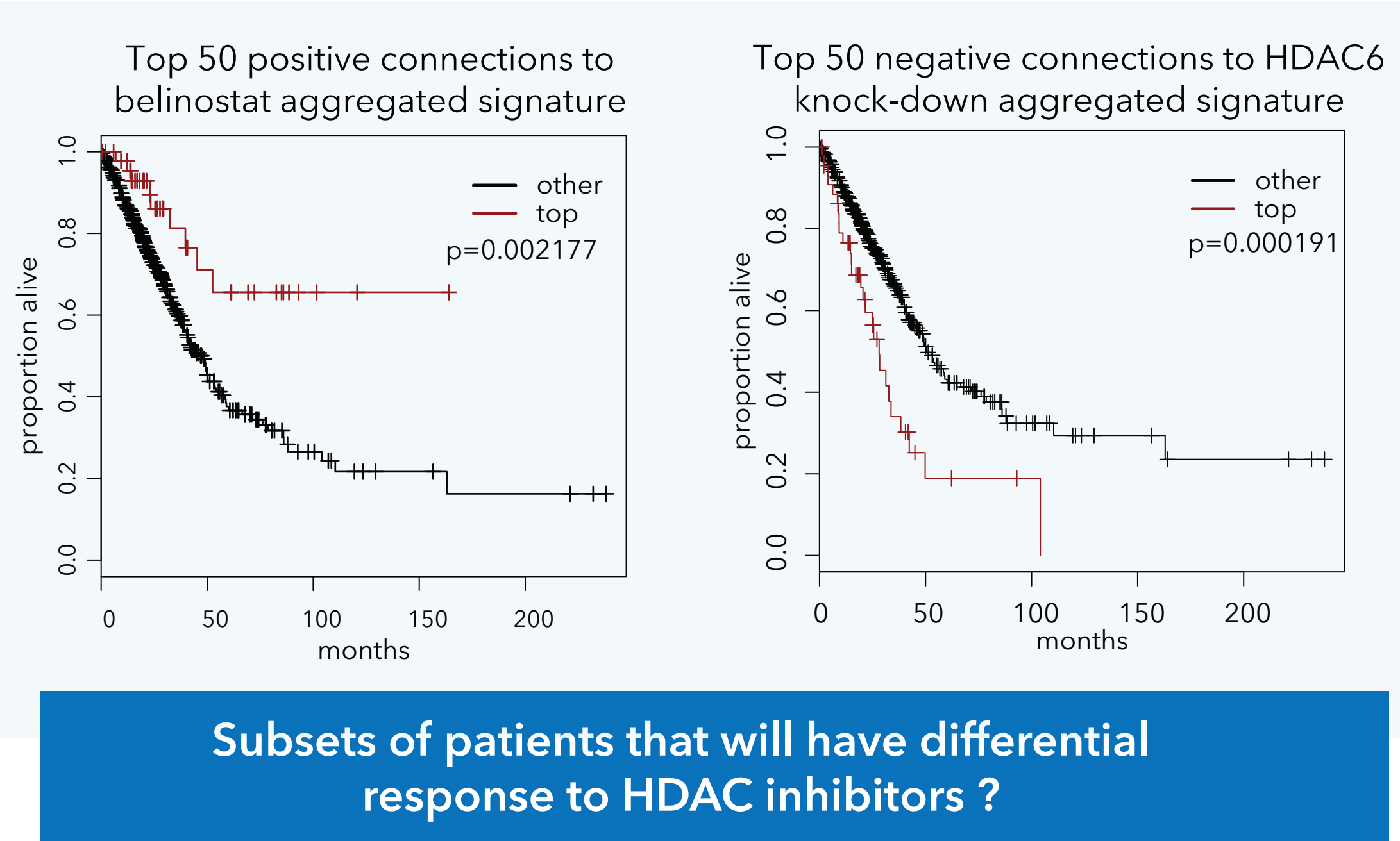
A signature of HDAC activation is predictive of survival in LUAD

Connectivity scores to aggregated CMap signatures tested for correlation with survival (cox regression)

10/12 best survival-correlated compound perturbagens are HDAC inhibitors

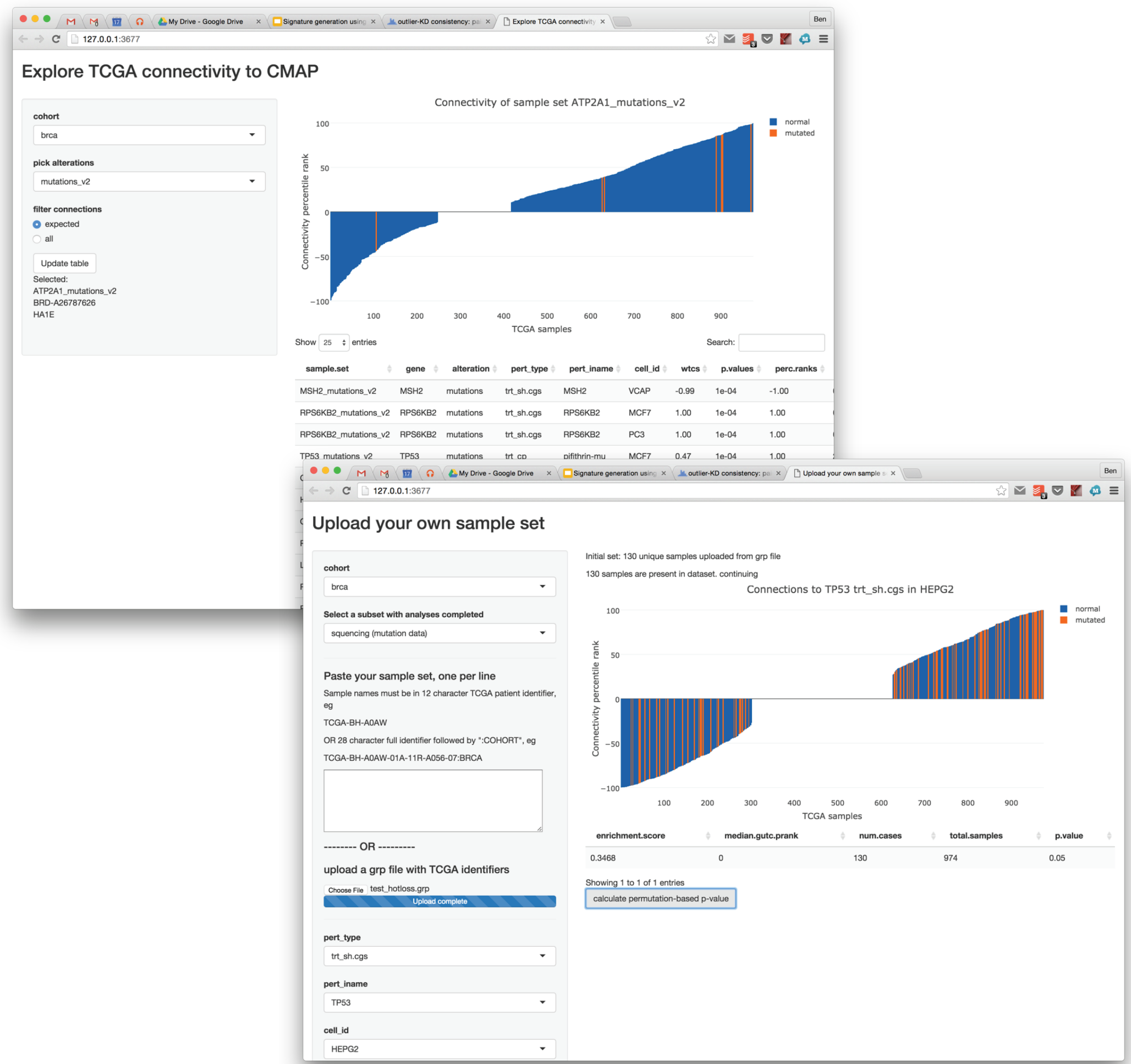
- Positive connection to HDAC inhibitors correlated with survival
- Negative connections to HDAC6 knock-down correlated with poor prognosis

compound	coef	p	fdr	annotation
belinostat	-0.0059	9.41E-06	0.012	HDAC inhibitor, cell cycle inhibitor
trichostatin-a	-0.0052	1.42E-05	0.013	HDAC inhibitor, CDK expression enhancer
HC-toxin	-0.0055	1.62E-05	0.013	HDAC inhibitor
panobinostat	-0.0050	2.88E-05	0.016	HDAC inhibitor, apoptosis stimulant
givinostat	-0.0063	6.24E-05	0.029	HDAC inhibitor, interleukin receptor antagonist
trichostatin-a	-0.0049	8.78E-05	0.030	HDAC inhibitor, CDK expression enhancer
vorinostat	-0.0046	1.24E-04	0.033	HDAC inhibitor, cell cycle inhibitor
THM-I-94	-0.0048	1.32E-04	0.035	HDAC inhibitor, apoptosis stimulant
acepromazine	-0.0108	1.81E-04	0.042	dopamine receptor antagonist
SB-202190	-0.0059	2.22E-04	0.049	p38 MAPK inhibitor, interleukin inhibitor
scriptaid	-0.0049	2.36E-04	0.051	HDAC inhibitor
ISOX	-0.0040	2.98E-04	0.057	HDAC inhibitor



Subsets of patients that will have differential response to HDAC inhibitors ?

Web apps accelerate discovery



Three use cases

- 1) Explore pre-computed results
Users can explore the results of our investigation with a set of default sample sets.
- 2) Query with custom sample sets
Rapid hypothesis testing with a new sample set of interest. Useful for biologists to explore new ideas.
- 3) Use external expression data
Investigate novel data with our methods. Bring in your own expression data and sample sets.

Future directions

- Pilot investigation handled 5 cohorts
 - Expand to the rest of TCGA
- Other large-scale expression databases
 - GEO data, TOX21
- Release and promote web apps
 - Open source soon!
- Dive further into biological findings
 - KEAP1/STK11 findings
- Compile list of best CMap perturbagens

Acknowledgments

CMap lab and data team
Golub lab members
Broad ACB resources
Kayleigh and Cancer program admins
cbioportal.org
R / shiny / plotly
NIH LINCS program
TCGA

