Library of Integrated Network Based Cellular Signatures

(LINCS)

Aravind Subramanian

NIEHS Workshop

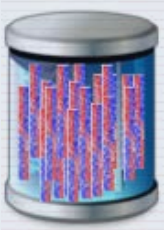
June 26, 2013

BROAD INSTITUTE

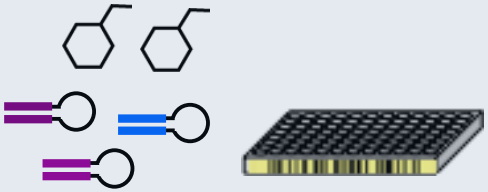
LINCS as a look-up table of cellular activity

COMMUNITY QUERIES PLATFORM-INDEPENDENT

database



GENOME SCALE GENETIC PHARMACOLOGIC



perturbations

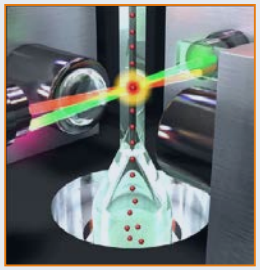
+

MODERATE COMPLEXITY 10'S

cell types

→

COMPLEX



read out

Connectivity Map

diseases

genes

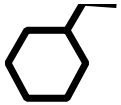
drugs

mRNA Expression Database

453 Affymetrix profiles
164 drugs
> 16,000 users
916 citations

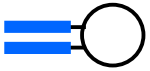
Lamb et al, Science (2006)

Expanding the Connectivity Map Concept



more small molecules

drugs, tools, natural products, libraries



genomic perturbations

shRNA, ORF, variants

cellular contexts

cell types, culture conditions

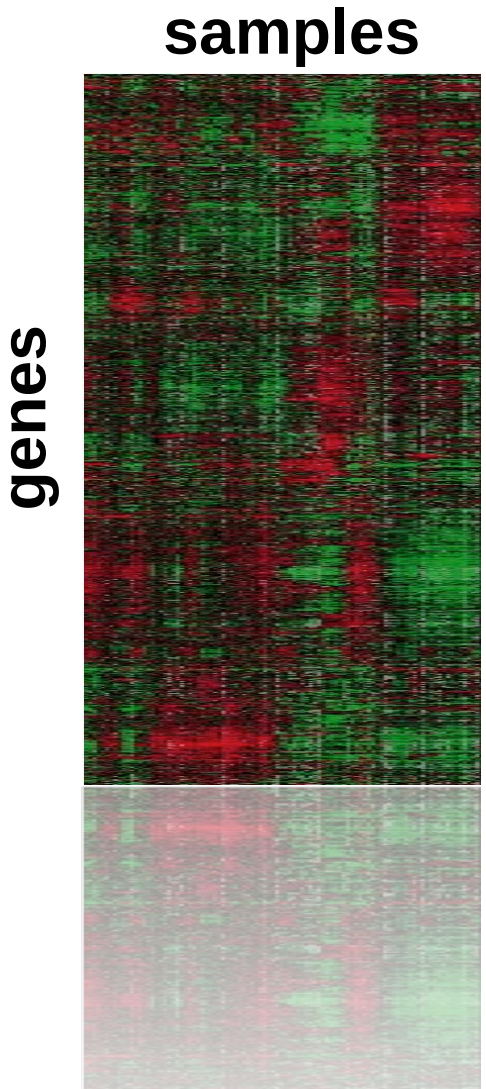
treatment parameters

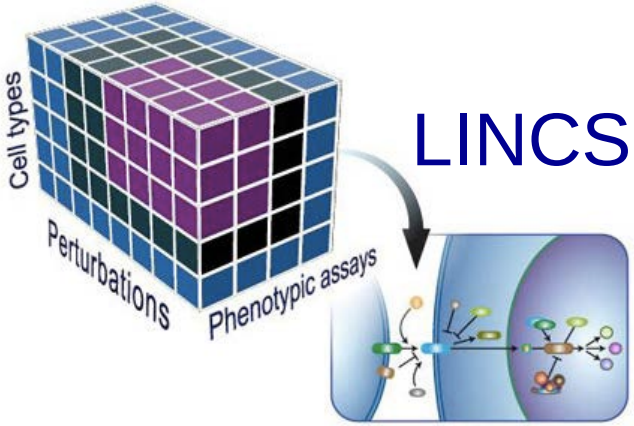
concentrations, durations, combinations

At ~ $500/array, approach does not scale

Toward reduced representation of transcriptome

gene expression is correlated





LINCS

Reduced Representation of Transcriptome

reduced representation transcriptome

'landmarks'

+

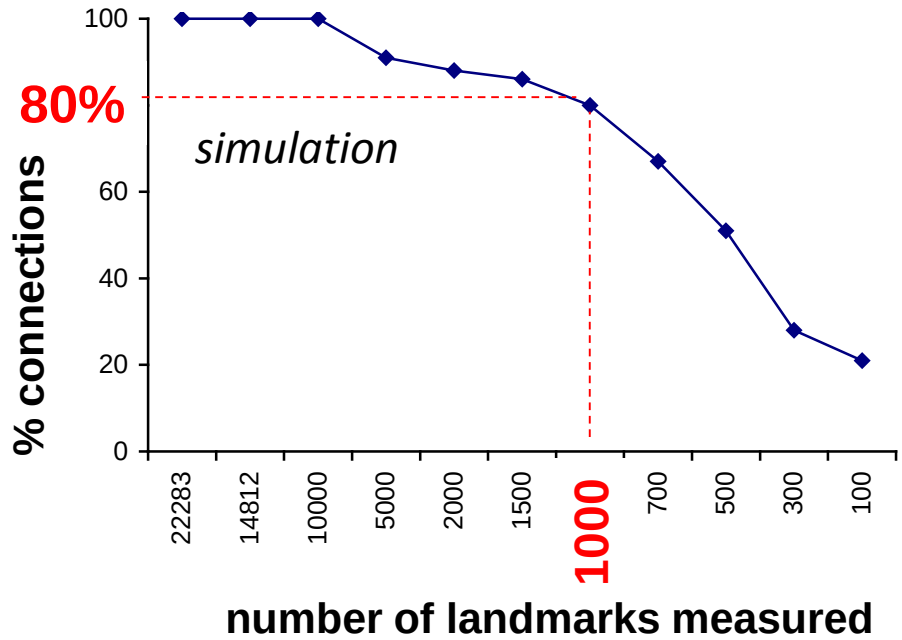
computational inference model

=

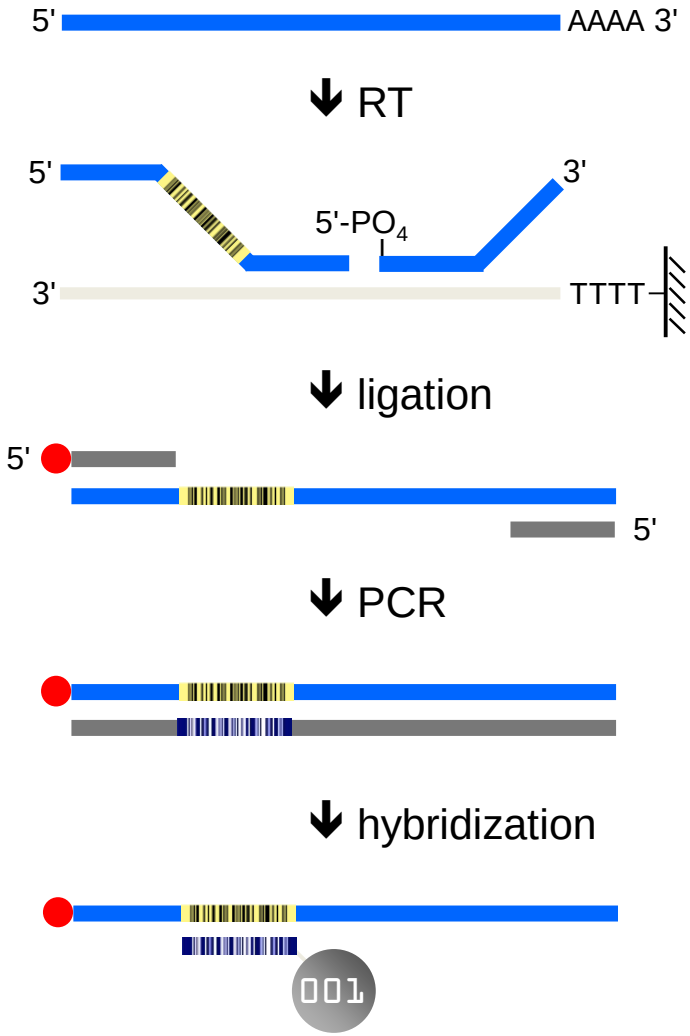
genome-wide expression profile

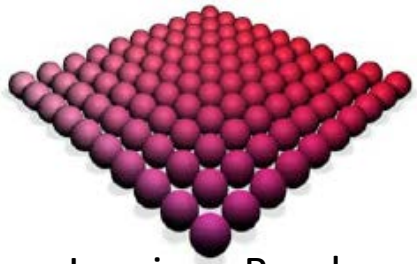
Gene Expression Omnibus

~ 130,000 profiles

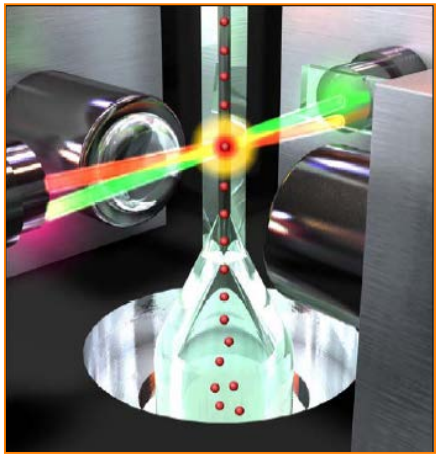


1000-plex Luminex bead profiling



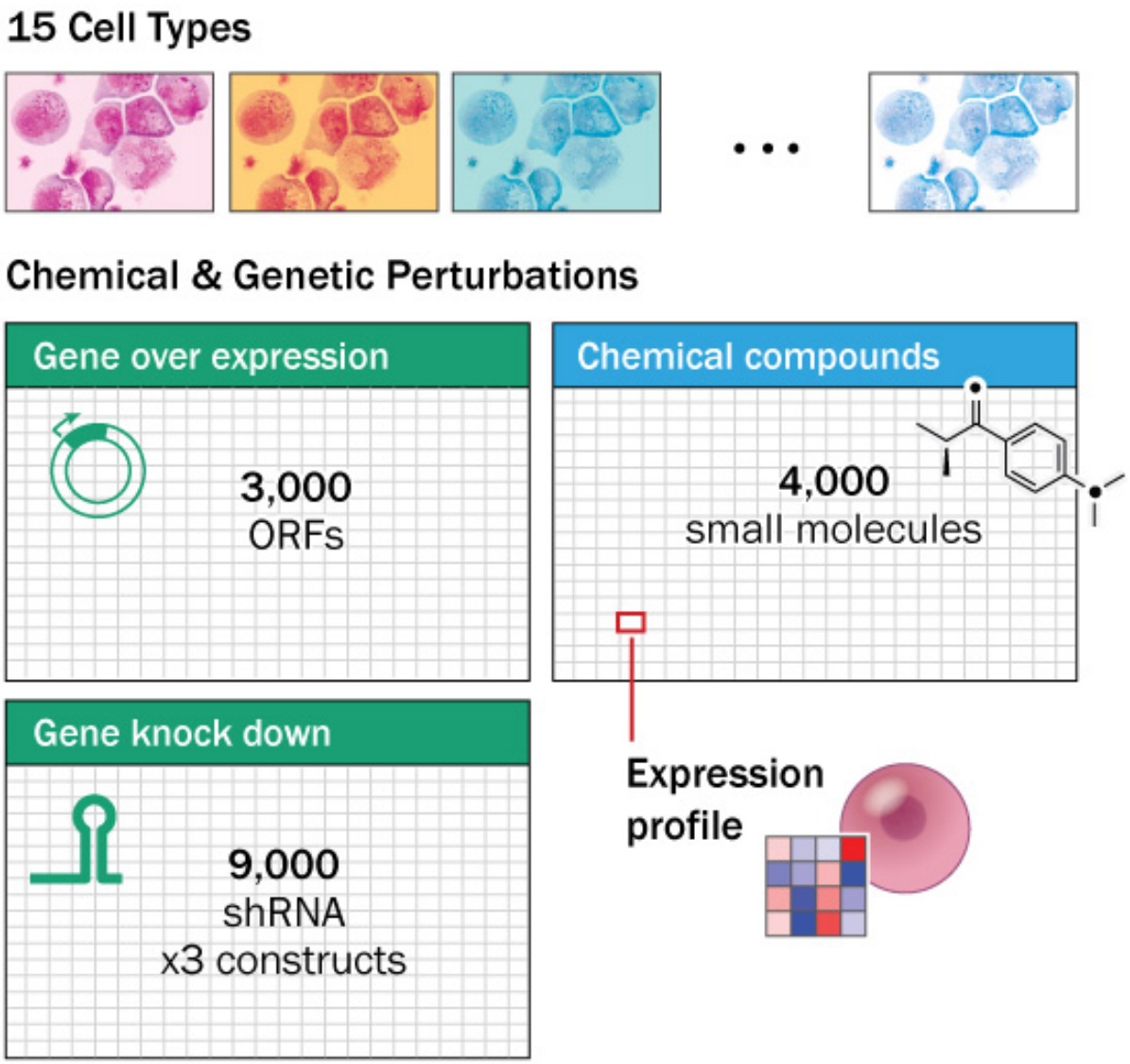


Luminex Beads
(500 colors,
2 genes/color)

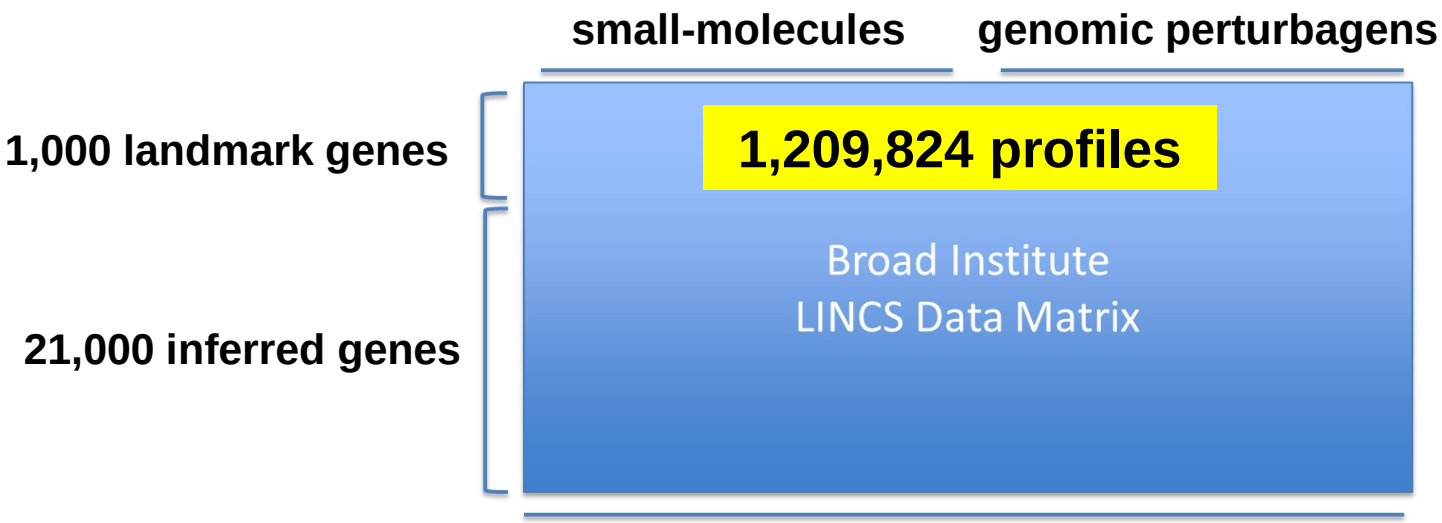


Reagent cost:
$5/sample

LINCS Goal


15 Cell Types
Chemical & Genetic Perturbations
Gene over expression
3,000
ORFs
Chemical compounds
4,000
small molecules
Gene knock down
9,000
shRNA
x3 constructs
Expression
profile

Current LINCS Dataset



15 cell types

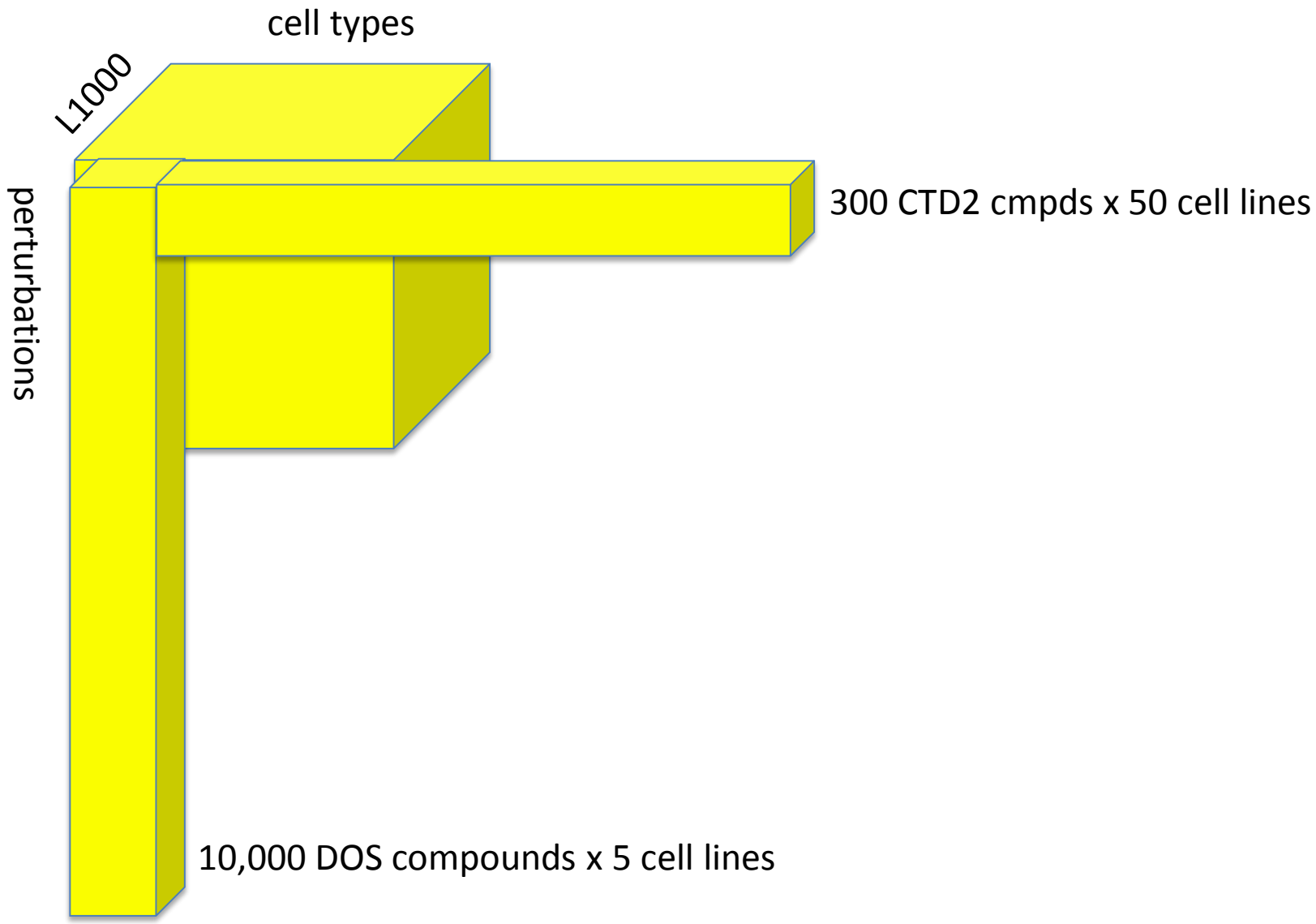
- Banked primary cell types
- Cancer cell lines
- Primary hTERT-immortalized
- Patient-derived iPS cells
- 5 community nominated

5,178 compounds

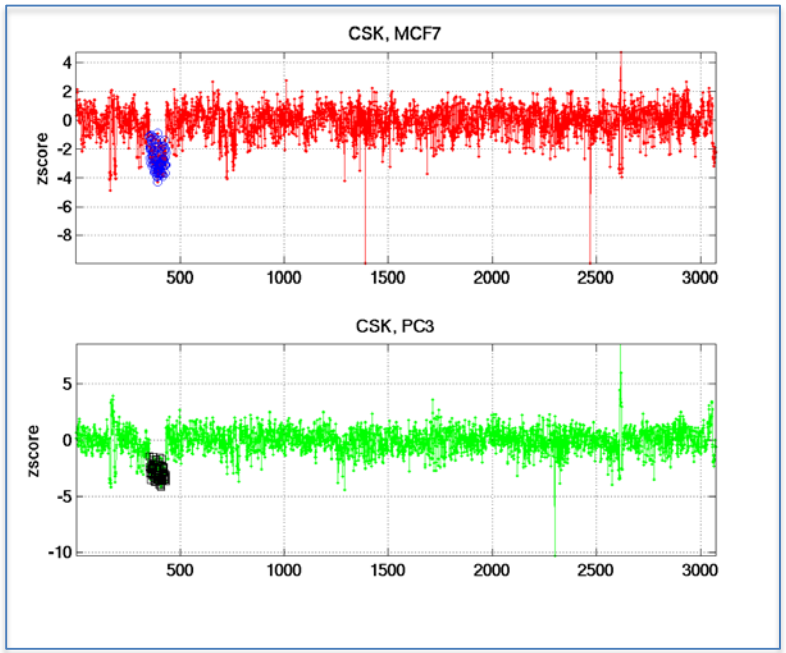
- 1,300 off-patent FDA-approved drugs
- 700 bioactive tool compounds
- 2,000+ screening hits (MLPCN + others)

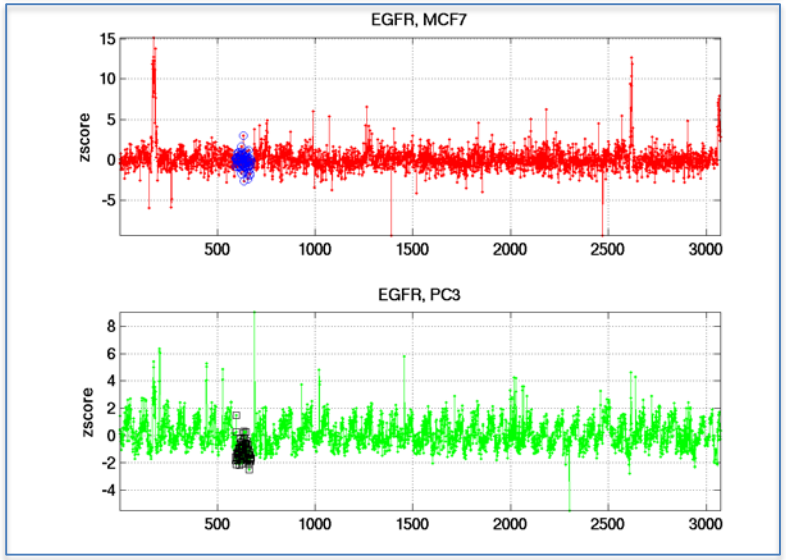
3,712 genes (shRNA + cDNA)

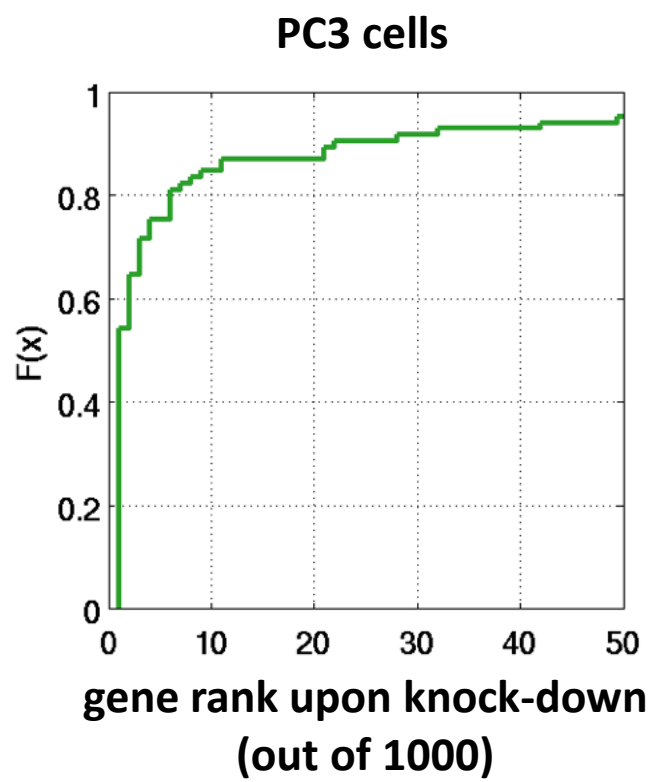
- targets/pathways of FDA-approved drugs (n=900)
- candidate disease genes (n=600)


cell types
L1000
perturbations
300 CTD2 cmpds x 50 cell lines
10,000 DOS compounds x 5 cell lines

Do shRNAs knock down their target?



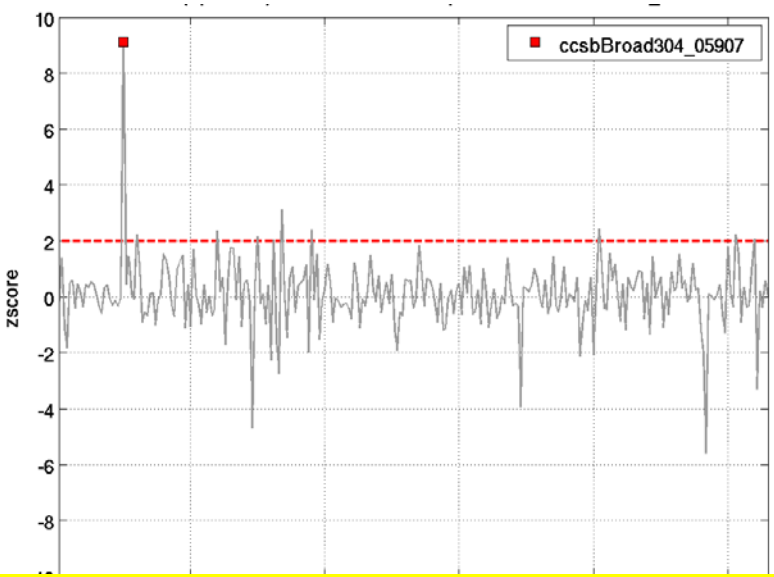


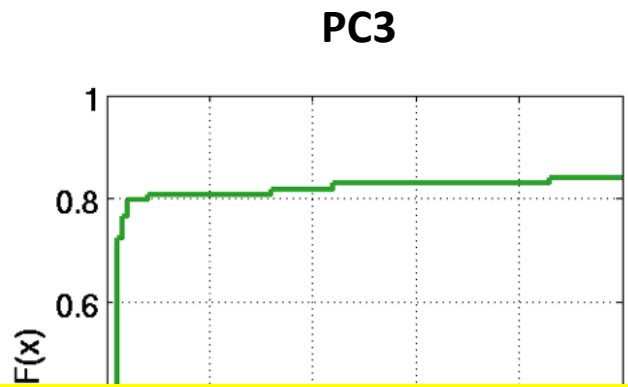


85% in top 1 percentile in single cell line

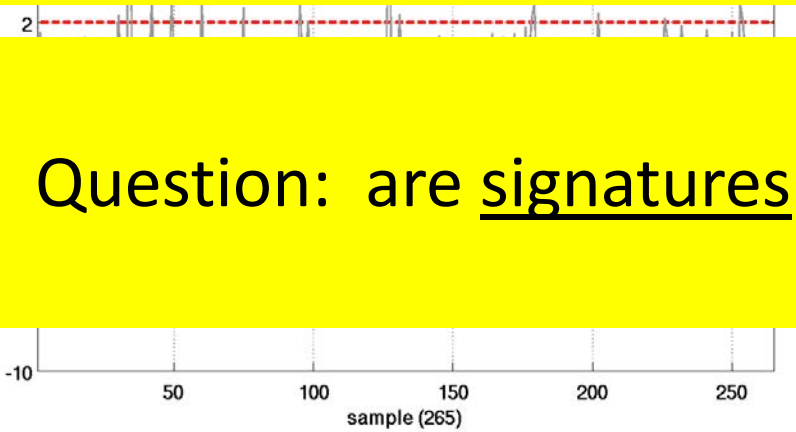
98% in top 1 percentile across 9 lines

Do ORF constructs overexpress their target?



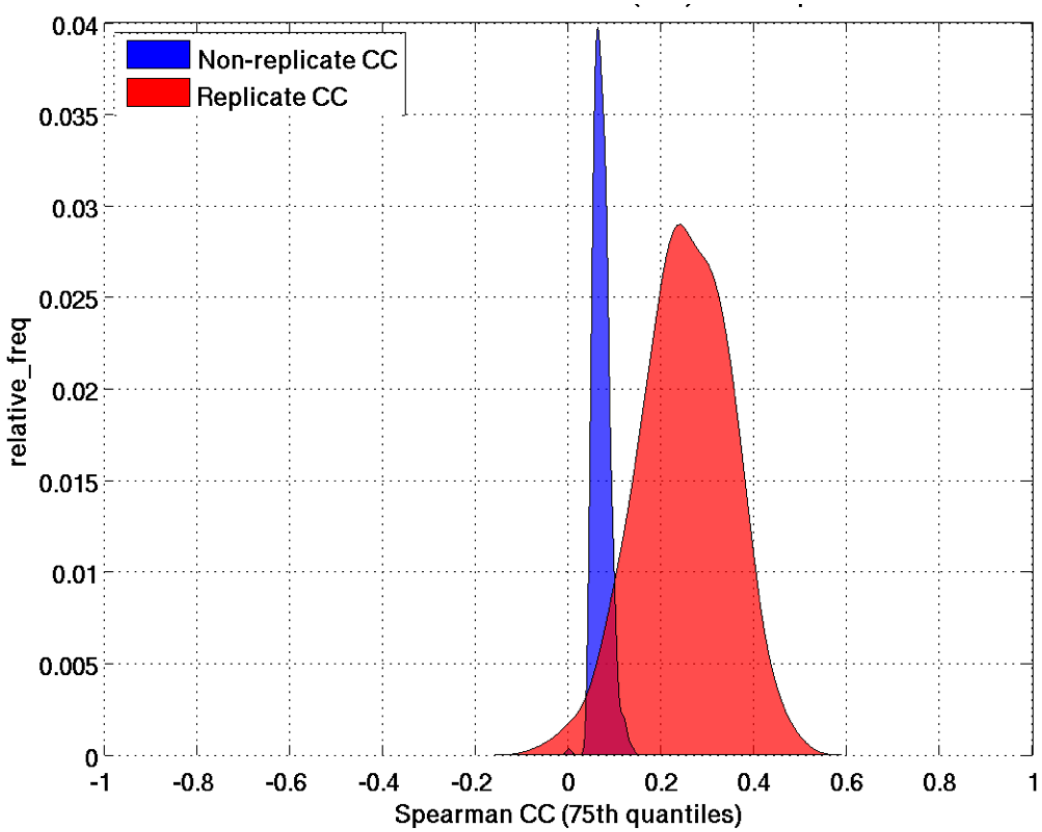


Conclusion: most genetic perturbagens modulate target

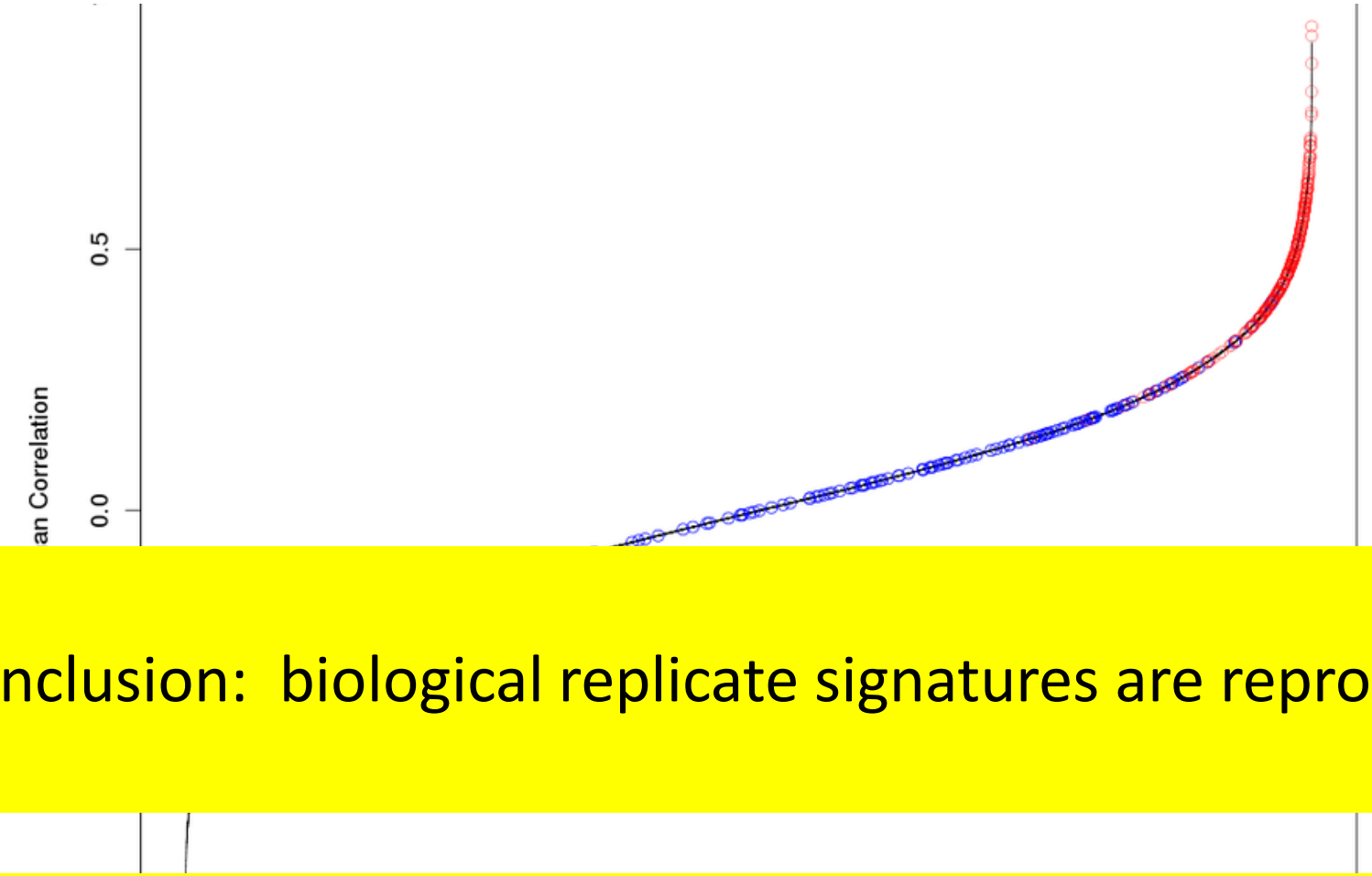


Question: are <u>signatures</u> reproducible?

Data quality:
correlation between biological replicate signatures



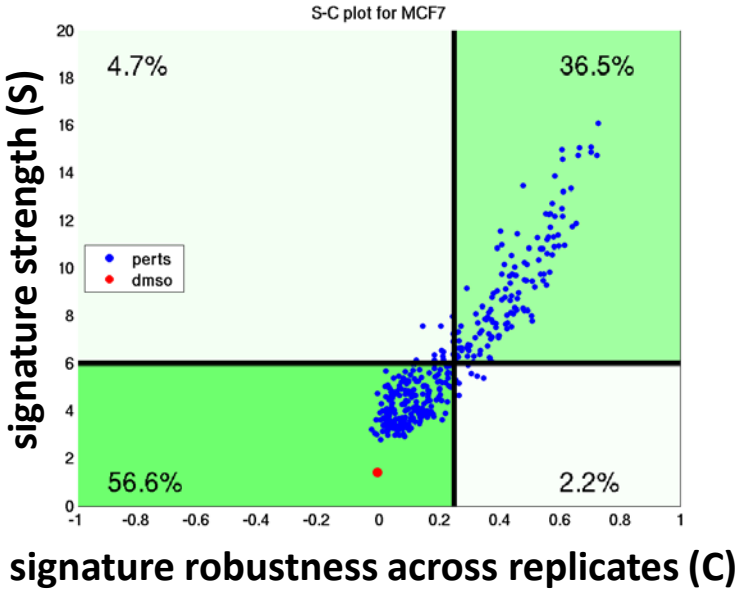
Biological replicate signatures are highly correlated

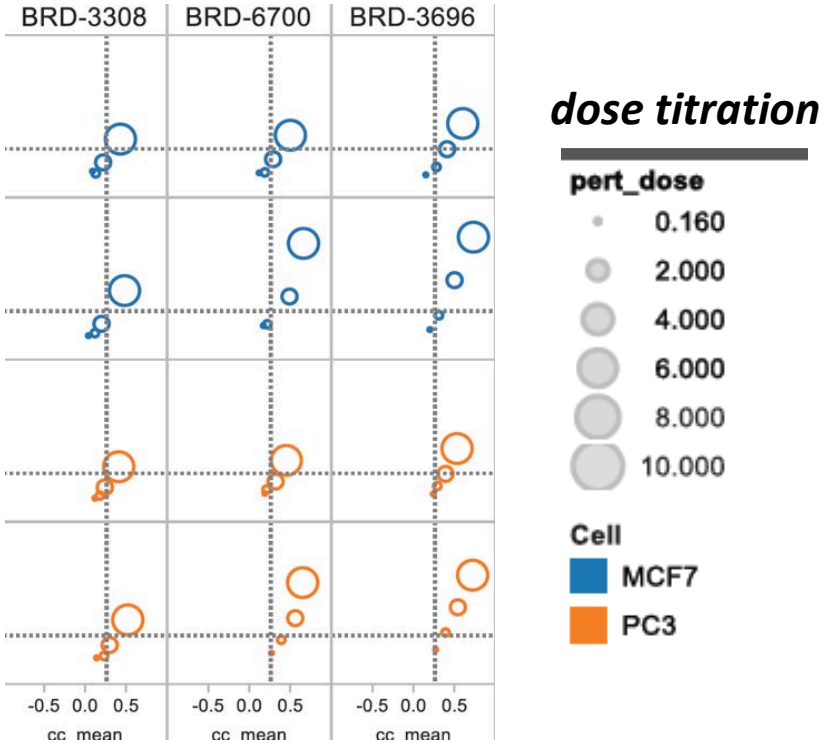


Conclusion: biological replicate signatures are reproducible

Question: are signatures reproducibly on target?

S-C Plots: Signature Strength and Correlation





1% Club

- Ask whether a given query signature returns expected connections within top 1% of all signatures

Conclusion: different shRNAs targeting the same gene often have different cellular effects

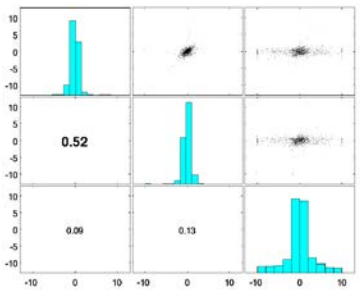
Question: what explains this divergence?

variable knock-down of target? X

off-target effects based on seed sequence? ✓

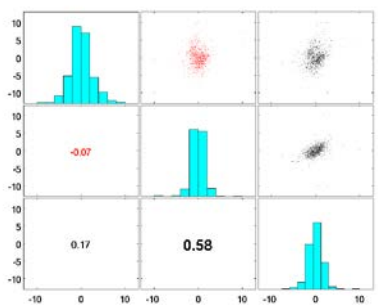
Need consensus (core) signatures

Rep1 Rep2 Rep3



Hairpin A

Rep1 Rep2 Rep3



Hairpin B

Moderated Z-score

Emphasize consistent signatures

Minimize non-shared off-target effects

$$MODZ = \frac{w_a.z_a + w_b.z_b + w_c.z_c}{\sqrt{w_a^2 + w_b^2 + w_c^2}}$$

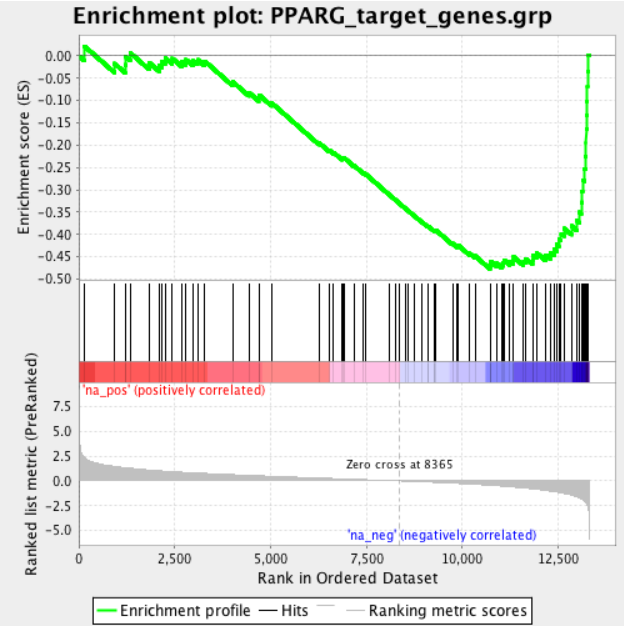
where,

$$w_a = \frac{R_{AB} + R_{AC}}{2} \quad w_b = \frac{R_{AB} + R_{BC}}{2} \quad \ldots$$

In general,

$$MODZ = \frac{\sum_i w_i.z_i}{\sqrt{\sum_i w_i^2}}$$

Example: PPARγ consensus shRNA signature connected to PPARγ pathway (literature)

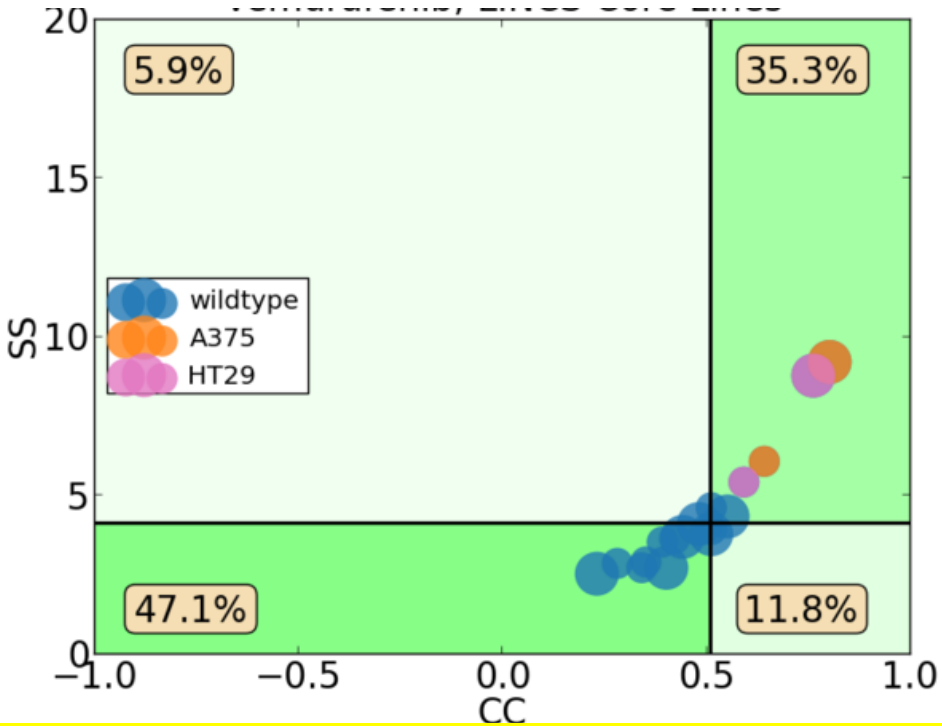


576 transcription factor target gene sets from MSigDB

PPARγ signature: 107 genes
4/107 measured ('landmarks')
Remainder inferred

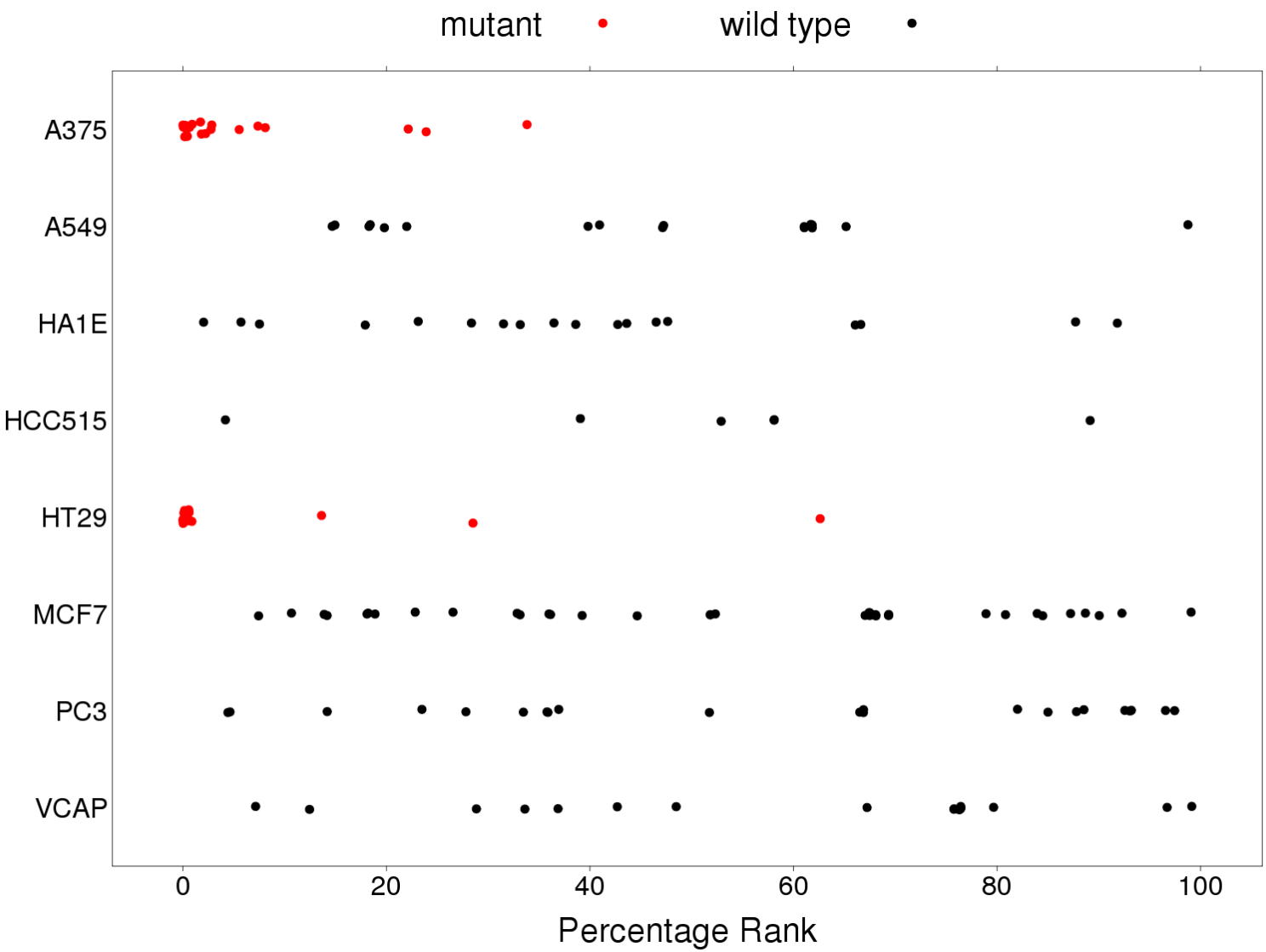
PPARγ gene set most connected to PPARγ consensus signature (rank: 1/576)

Example: activity of BRAF inhibitor vemurafenib in different cellular contexts

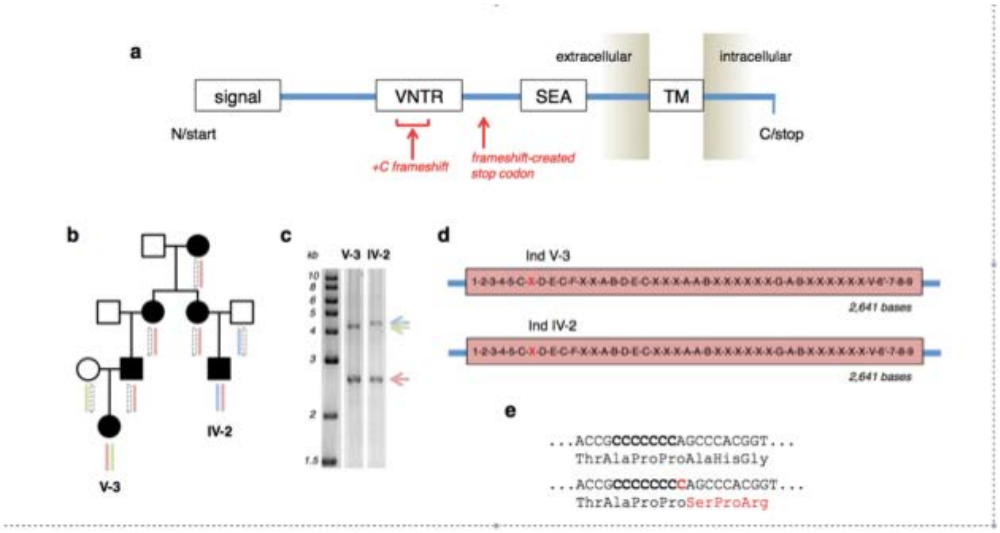


Question: can vemurafenib activity be connected to its target via consensus shRNA signature?

Rank vemurafenib connectivity to 3000 consensus shRNA signatures

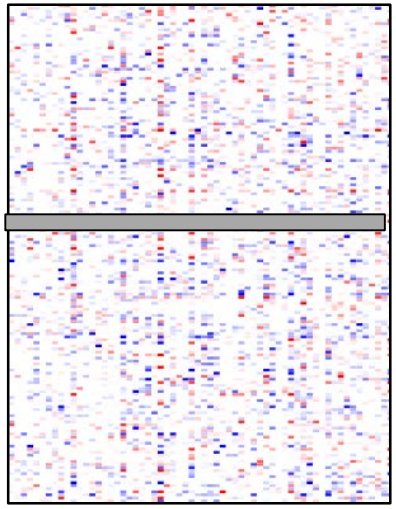


Example: Discovering repressors of MUC1 in MCKD1



Kirby A., et al. Mutations causing medullary cystic kidney disease type 1 (MCKD1) lie in a large VNTR in MUC1 missed by massively parallel sequencing. March 2013, *Nature Genetics*

LINCS DB



MUC1 →

Top MUC1 repressors

Compound	Dose	Cell Line	CMap Z-score	Specificity	% inhibition
Anti-infective	10 uM	MCF7	-5.41	0.34	-43%
Kinase inhibitor	10 uM	MCF7	-4.86	0.28	-46%
GPCR ligand	10 uM	MCF7	-4.76	0.31	-73%
Steroid Hormone ligand	10 uM	MCF7	-4.4	0.12	-60%
GPCR ligand	10 uM	MCF7	-4.29	0.17	-39%
Steroid Hormone ligand	10 uM	MCF7	-3.59	0.19	-71%

LINCS Focus

- Generate initial data set ✓
- Make raw data (and QC) available ✓
- Develop first gen. analytical tools ✓
- Begin to understand data *(in progress)*
- Develop biologist-friendly tools *(beta release)*
- Engage in use-case collaborations *(ongoing)*

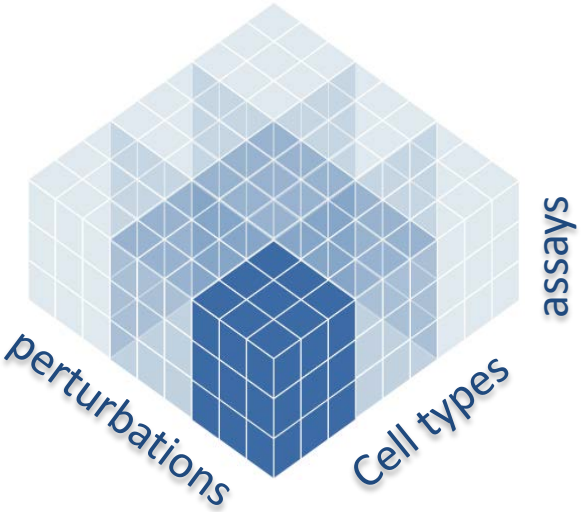
Data Release and Access Plans

	desc	format	availability
level 1	**Raw** data	Plate folders with	**now: 3,430 folders**
level 2	**Normalized** dataset	Matrix: GCTX	**now: 1,209,824 profiles**
level 3	**Signatures** (differentially expressed genes)	1. mongo DB 2. Matrix: GCTX	**now: 369,564 signatures (beta release)**
level 4	**Queries**	Biologist-friendly web apps	**Q1 2013 (beta release)**

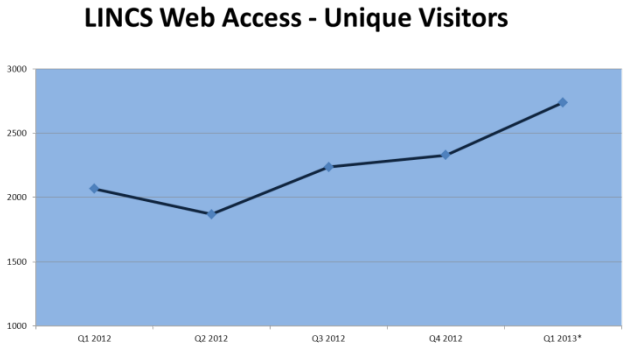
http://www.lincscloud.org

The LINCS Challenge & Opportunity

- Opportunity:
  - New understanding of health and disease via **identifying common patterns** (signatures) in molecular and cellular responses to perturbations at "genome scale".



- Challenges:
  - Integrating and visualizing **diverse data types and signatures**
  - Holistic understanding of cell responses
  - Adoption and use of the LINCS approach by multiple research communities



lincsproject.org

LINCS Perturbations: Drugs and other small molecules; genetic (shRNA, ORFs); disease (primary & iPS cells).

The LINCS Network

Data Production/Analysis Centers

Broad Institute
Harvard Medical School

Technology Development Centers

Arizona State
Broad Institute
Columbia
Yale

Computational Tools Centers

Columbia
U. Cincinnati
Miami School of Medicine
Wake Forest

External Collaborations

- Snyder Lab, Sanford-Burnham Medical Research Institute
- FDA
- GTEx
- Mootha Lab, Massachusetts General Hospital
- ENCODE/Epigenomics
- Rao Lab, NIH CRM:
- Scadden Lab, Massachusetts General Hospital
- McCray Lab, University of Iowa
- Loring Lab, Scripps Research Institute
- Edenberg Lab, Indiana University
- Spria Lab, Boston University
- Pandolfi Lab, BIDMC
- Chen Lab, NHLBI
- Ebert Lab, Brigham and Women's Hospital
- Kotton Lab, Boston University

LINCS Team at Broad Institute

Steven Corsello
John Davis
John Doench
Melanie Donahue
Corey Flynn
Peyton Greenside
Joshua Gould
Larson Hogstrom
Roger Hu
Arthur Liberzon
Xiaodong Lu

Rajiv Narayan
Ted Natoli
David Peck
Federica Piccioni
David Root
Victor Rusu
Ian Smith
David Wadden
Bang Wong
Xiaohua Wu

Todd Golub