

PCCSE



# Using Sentinel Profiles for Drug Characterization: Clustering and connectivity

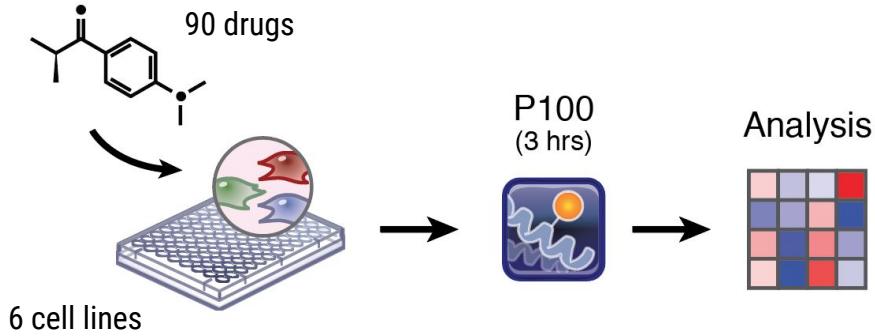
---

Lev Litichevskiy

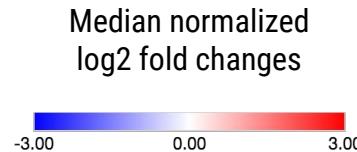
NEU PROTEOMICS CAPSTONE MODULE | MAY 12, 2017

# Profiles can be subtle

## Vemurafenib in 6 cell lines (3 replicates)



- Vemurafenib doesn't look remarkably different in a particular cell line
- But we know that it should...

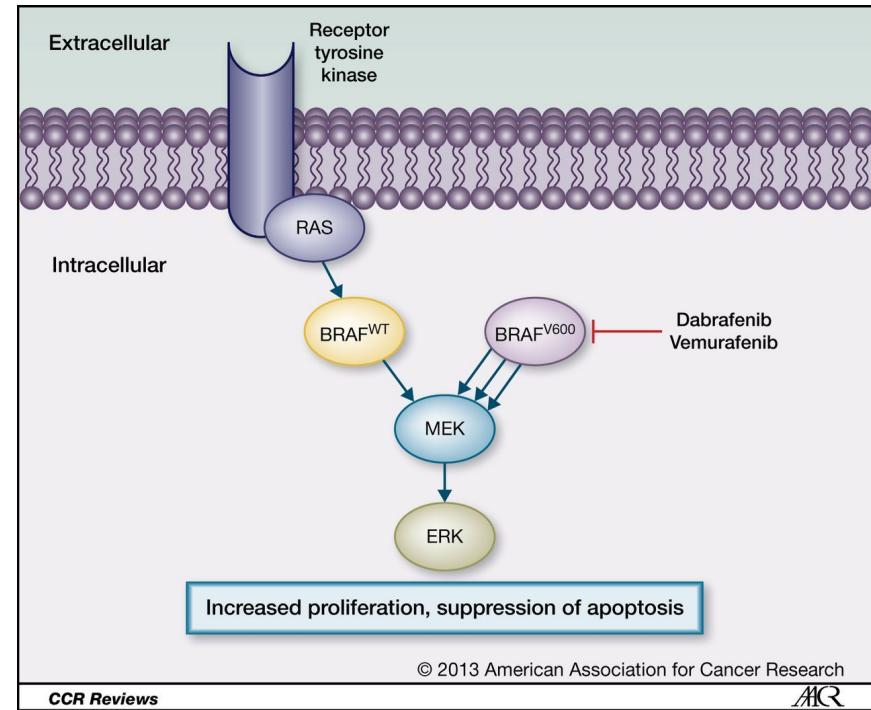


# Biology of vemurafenib

## A375 has a BRAF V600E mutation

V600E mutation in BRAF protein makes it constitutively active

Vemurafenib selectively binds to V600E form of BRAF protein

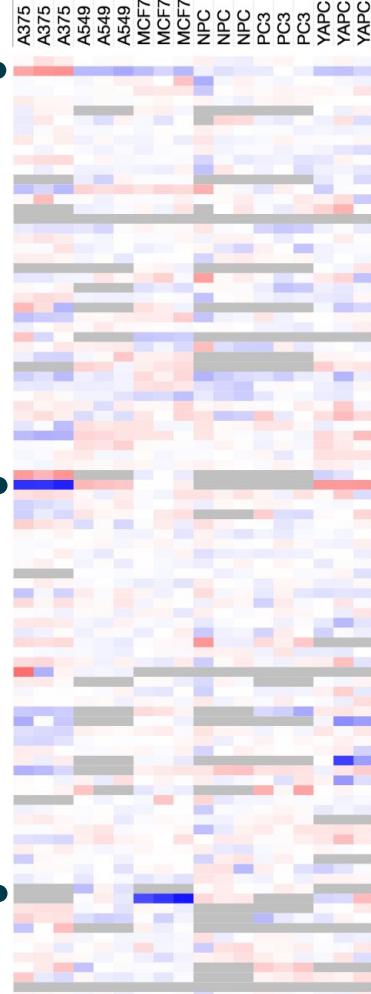


# What can we learn by looking at individual analytes?

RPS6KA3, S369



"May be phosphorylated at Thr-365 and **Ser-369** by MAPK1/ERK2 and MAPK3/ERK1."



AHNAK, S3426

N/A!

VPRBP, S1000



## In vivo Characterization

Methods used to characterize site *in vivo*:

mass spectrometry (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18)  
breast cancer (5, 12), breast ductal carcinoma (5), HER2 positive breast cancer (1), luminal A breast cancer (1), luminal B breast cancer (1), triple negative (1, 5), cervical cancer (15), cervical adenocarcinoma (15), leukemia (13), acute myelogenous leukemia (13), acute erythroid leukemias, including erythroleukemia (M6a) and very rare pure erythroid leukemia (M6b) (11), acute megakaryoblastic leukemia (M7) (11), acute monoblastic leukemia (M5a) or acute monocytic leukemia (M5b) (11), acute myeloblastic leukemia, with granulocytic maturation (M2) (11), acute myeloblastic leukemia, without maturation (M1) (11), lung cancer (12), non-small cell lung cancer (12), B cell lymphoma (11), non-Hodgkin's lymphoma (11), ovarian cancer (5), multiple myeloma (11), melanoma, skin cancer (3)

Relevant cell line - cell type - tissue:

293 (epithelial) [ATT (human), transfection] (14), 293T (epithelial) (2), A549 (pulmonary) (8), AML-193 (monocyte) (11), breast (1, 5), BT-20 (breast cell) (12), BT-549 (breast cell) (12), Calu 6 (pulmonary) (12), CMK (megakaryoblast) (11), CTS (myeloid) (11), DOHH2 ('B' lymphocyte) (11), H2009 (epithelial) (11), H2228 (epithelial) (11), H226 (epithelial) (12), H2887 (pulmonary) (12), H322 (pulmonary) (12), HS22M (uterine) (12), HCC1937 (breast cell) (12), HCC2279 (pulmonary) (12), HCC366 (pulmonary) (12), HCC78 (pulmonary) (12), HEL (erythroid) (11), HeLa (cervical) (4, 16, 17), HeLa S3 (cervical) (5), HOP62 (pulmonary) (12), Jurkat (T lymphocyte) (9, 18), K562 (erythroid) (10), Kasumi-1 (myeloid) (11), KG-1 (myeloid) (11, 13), liver (7), LOU-NH91 (squamous) (12), MCF-7 (breast cell) (12), MDA-MB231 (breast cell) (12), MDA-MB468 (breast cell) (12), MV4-11 (macrophage) (11), NCI-H157 (pulmonary) (12), NCI-H1648 (pulmonary) (12), NCI-H1666 (pulmonary) (12), NCI-H2172 (pulmonary) (12), NCI-H460 (pulmonary) (12), OPM-2 (plasma cell) (11), ovary (5), P31/FUJ (erythroid) (11), RL ('B' lymphocyte, precursor) (11), RPMI-8226 (plasma cell) (11), SH-SY5Y (neural crest) (6), SH-SY5Y (neural crest) (6), SU-DHL-6 (B lymphocyte) (11), U266 (plasma cell) (11), WM239A (epidermal) (3)

Controlled by

Treatments:

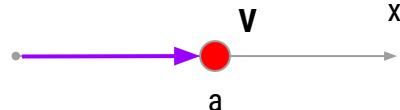
nocodazole (15)

# A profile is a vector in high dimensional space

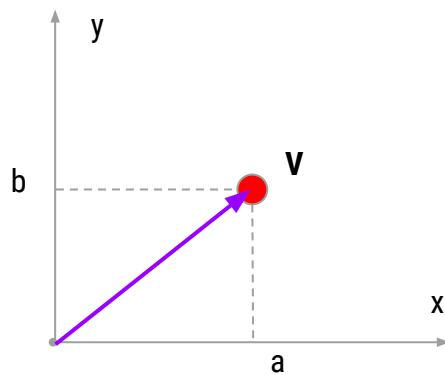
Our vectors have 96 dimensions

---

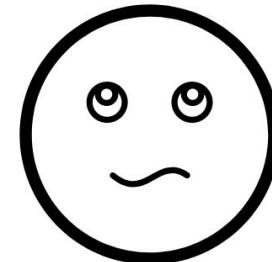
"We measured value  $a$  for analyte  $x$  when we treated with drug V."



"We measured value  $a$  for analyte  $x$  and value  $b$  for analyte  $y$  when we treated with drug V."



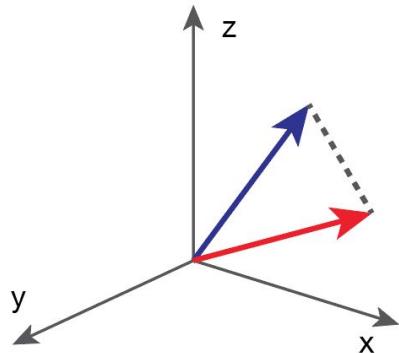
"We measured these 96 values for these 96 analytes for drug V."



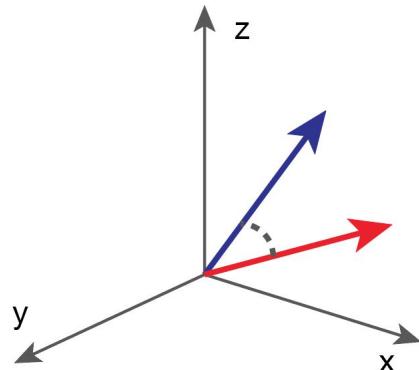
# How can we compare vectors?

## Examples of similarity metrics

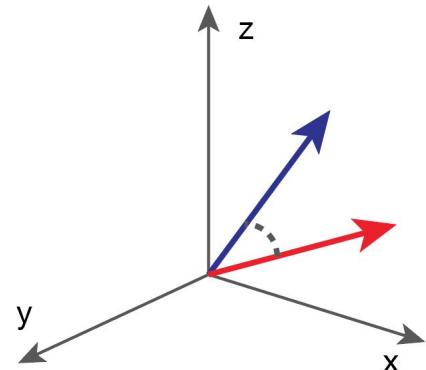
Euclidean distance



Pearson correlation



Spearman correlation

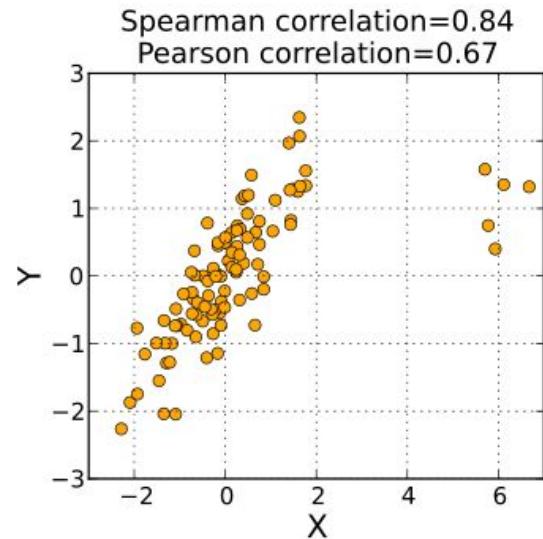
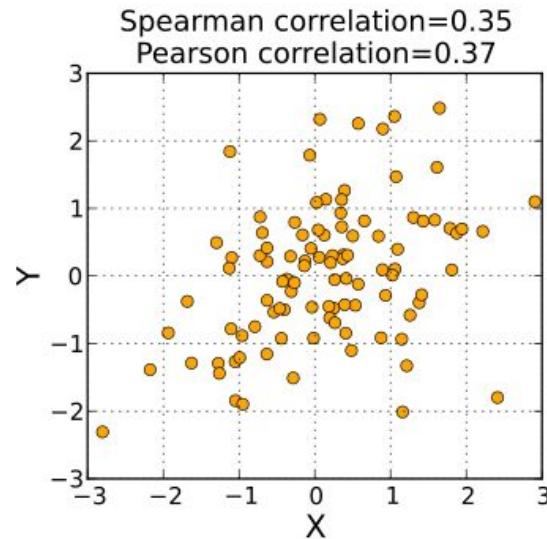
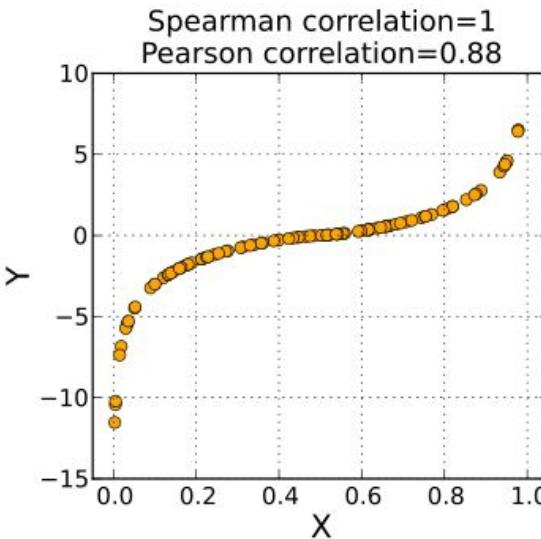


Exactly cosine similarity if  
vectors are centered

Values first converted to ranks

# Pearson or Spearman?

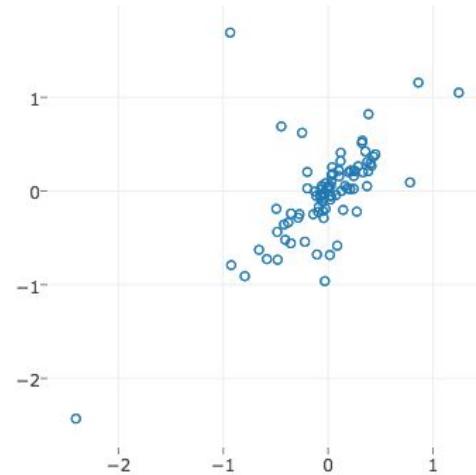
## Spearman tolerates outlier values



# Pearson v. Spearman on real data

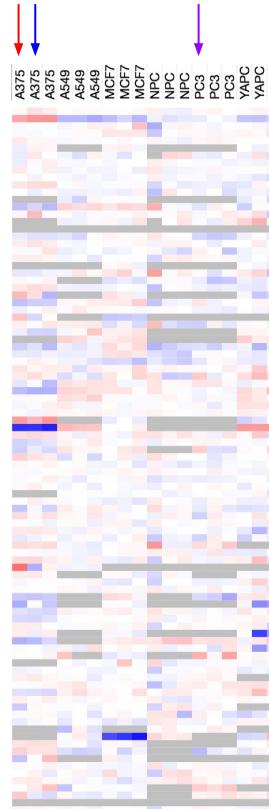
Typically we get similar results

Vemurafenib, A375,  
Replicate 1

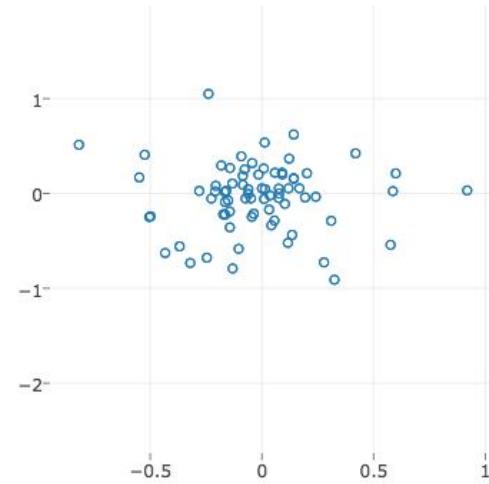


Vemurafenib, A375,  
Replicate 2

Pearson correlation: 0.65  
Spearman correlation: 0.66



Vemurafenib, A375,  
Replicate 1

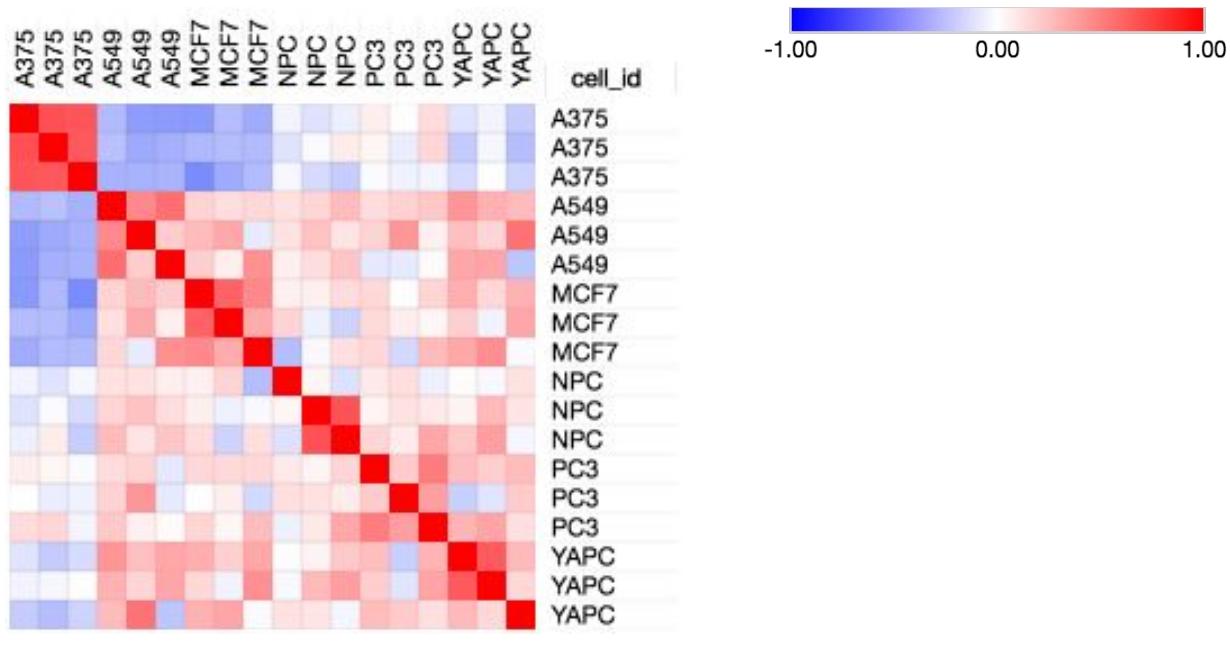


Vemurafenib, PC3,  
Replicate 1

Pearson correlation: -0.03  
Spearman correlation: 0.07

# Similarity matrix for vemurafenib

Appears that A375 samples stand out

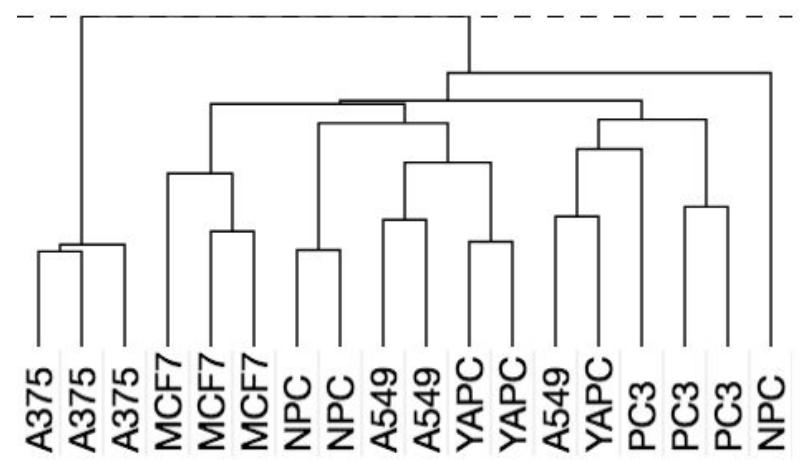


# Hierarchical clustering

Uses similarities to visualize structure in data

Clustering allows us to see more clearly that A375 is in its own cluster.

- Similarity metric: Spearman
- Linkage method: Average



# Two shortcomings of clustering

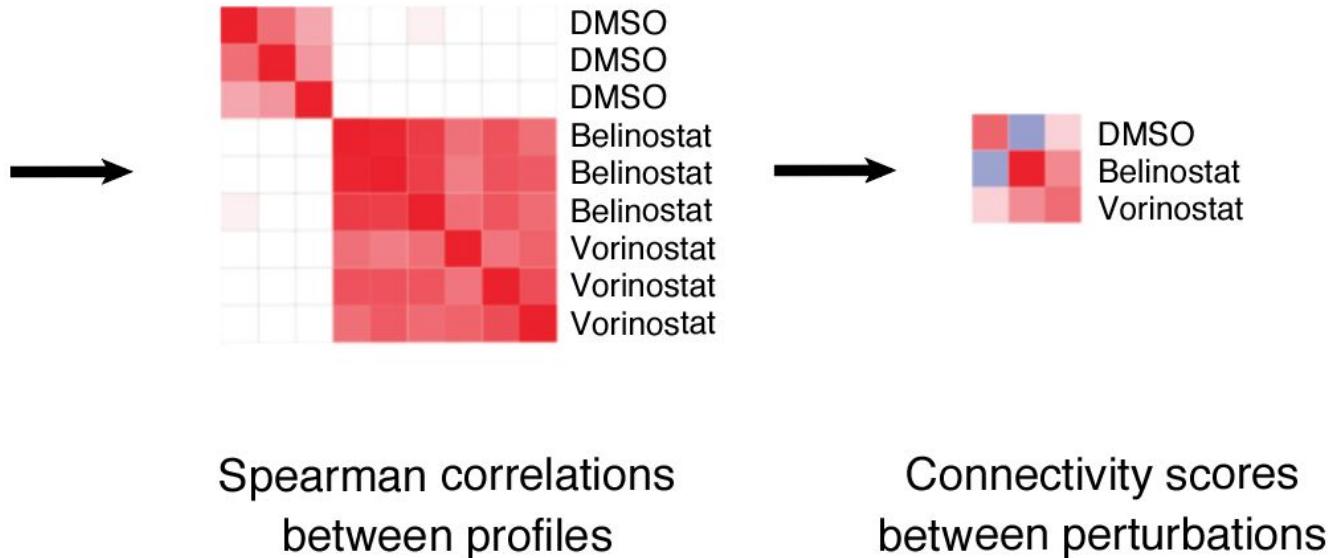
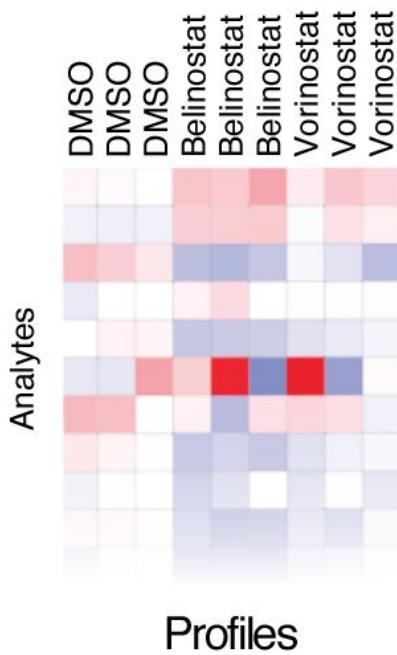
We address them with *connectivity*

---

1. We still need to deal with replicates
2. How do we know if a particular correlation is significant?

# Connectivity framework

---



# How we compute connectivity

First, we compute similarity

	A1	A2	A3	B1	B2	B3	C1	C2	C3
A1	1	.91	.83	.34	.39	.62	.43	.51	.58
A2	.91	1	.74	.24	.31	.40	.57	.61	.52
A3	.83	.74	1	.50	.46	.27	.39	.38	.71
B1	.34	.24	.50	1	.88	.72	-.40	-.31	-.64
B2	.39	.31	.46	.88	1	.79	-.39	-.71	-.21
B3	.62	.40	.27	.72	.79	1	.11	-.04	-.10
C1	.43	.57	.39	-.40	-.39	.11	1	.43	.55
C2	.51	.61	.38	-.31	-.71	-.04	.43	1	.70
C3	.58	.52	.71	-.64	-.21	-.10	.55	.70	1

# How we compute connectivity

Consider a given pair of drugs:

- Query = drug A
- Target = drug B

Targets

Queries

	A1	A2	A3	B1	B2	B3	C1	C2	C3
A1	1	.91	.83	.34	.39	.62	.43	.51	.58
A2	.91	1	.74	.24	.31	.40	.57	.61	.52
A3	.83	.74	1	.50	.46	.27	.39	.38	.71
B1	.34	.24	.50	1	.88	.72	-.40	-.31	-.64
B2	.39	.31	.46	.88	1	.79	-.39	-.71	-.21
B3	.62	.40	.27	.72	.79	1	.11	-.04	-.10
C1	.43	.57	.39	-.40	-.39	.11	1	.43	.55
C2	.51	.61	.38	-.31	-.71	-.04	.43	1	.70
C3	.58	.52	.71	-.64	-.21	-.10	.55	.70	1

# How we compute connectivity

Extract test and background distributions

Targets

Queries

	A1	A2	A3	B1	B2	B3	C1	C2	C3
A1	1	.91	.83	.34	.39	.62	.43	.51	.58
A2	.91	1	.74	.24	.31	.40	.57	.61	.52
A3	.83	.74	1	.50	.46	.27	.39	.38	.71
B1	.34	.24	.50	1	.88	.72	-.40	-.31	-.64
B2	.39	.31	.46	.88	1	.79	-.39	-.71	-.21
B3	.62	.40	.27	.72	.79	1	.11	-.04	-.10
C1	.43	.57	.39	-.40	-.39	.11	1	.43	.55
C2	.51	.61	.38	-.31	-.71	-.04	.43	1	.70
C3	.58	.52	.71	-.64	-.21	-.10	.55	.70	1

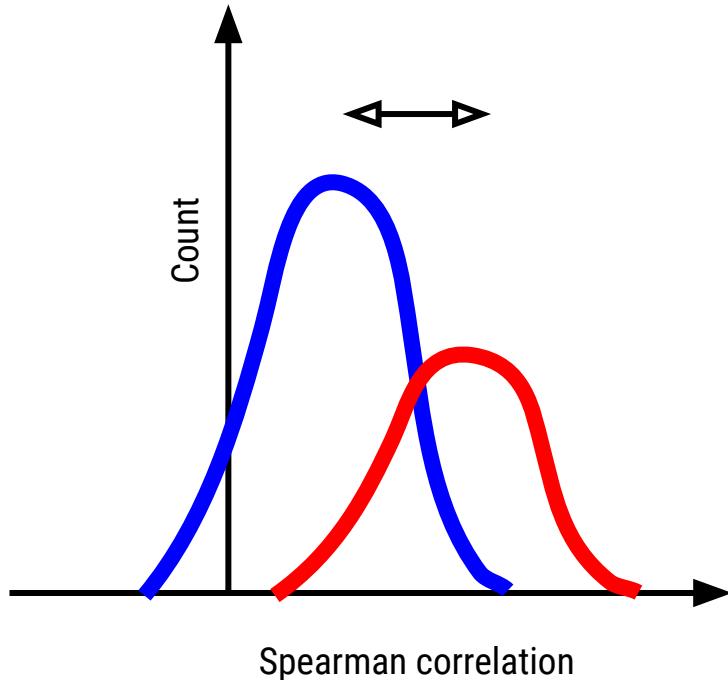
# How we compute connectivity

Compare test and background distributions

Test

.34	.24	.50	.39	.31	.46	.62	.40	.27
.88	.72	-.40	-.31	-.64	.79	-.39	-.71	-.21

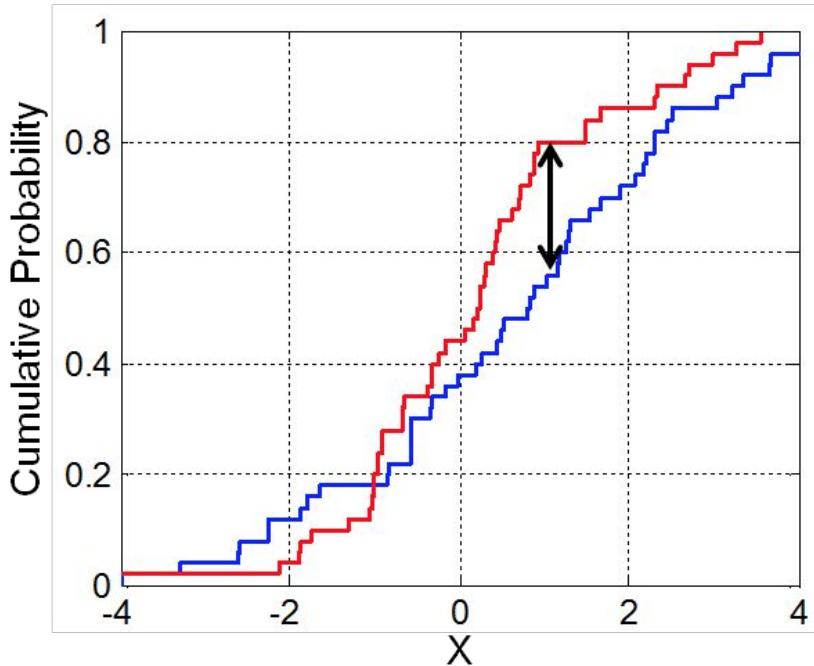
Background



# How we compute connectivity

Compare 2 distributions using Kolmogorov-Smirnov two-sample (KS) test

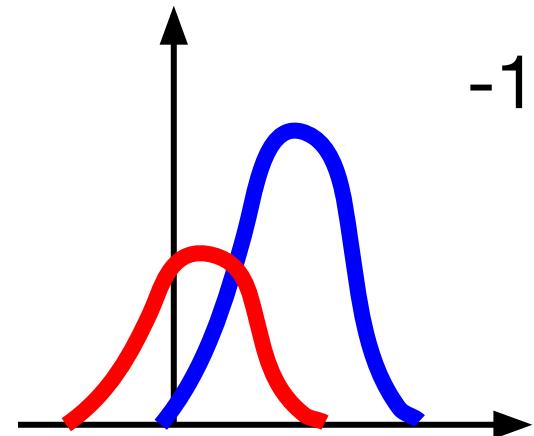
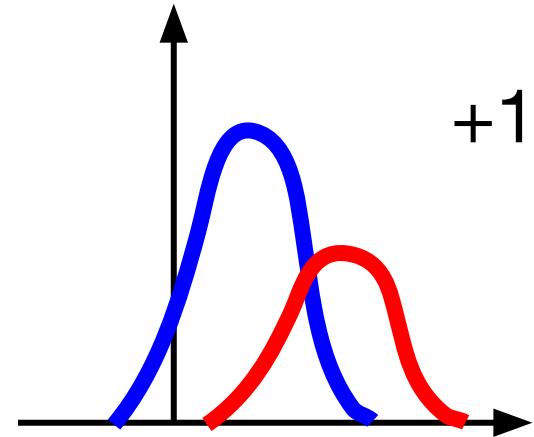
- Test statistic ranges from 0 to 1



# How we compute connectivity

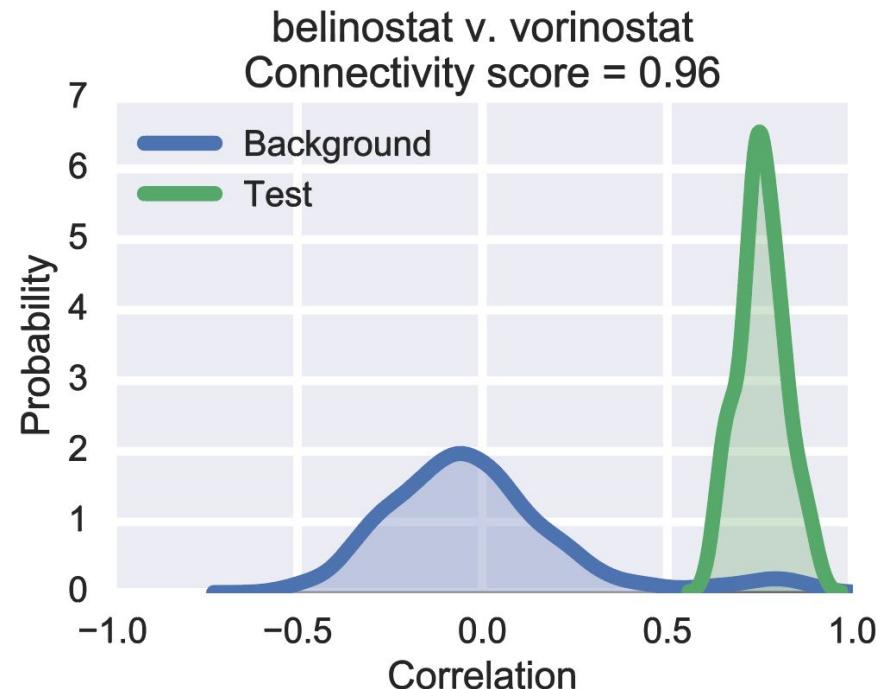
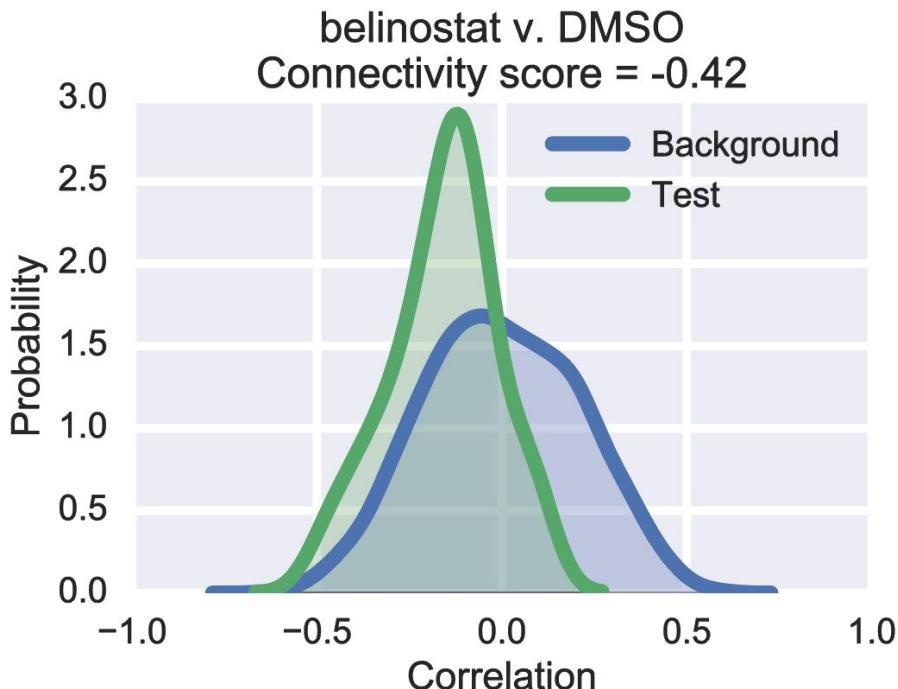
Add sign to result of KS test

- This is called the *connectivity score*
- Ranges from -1 to 1



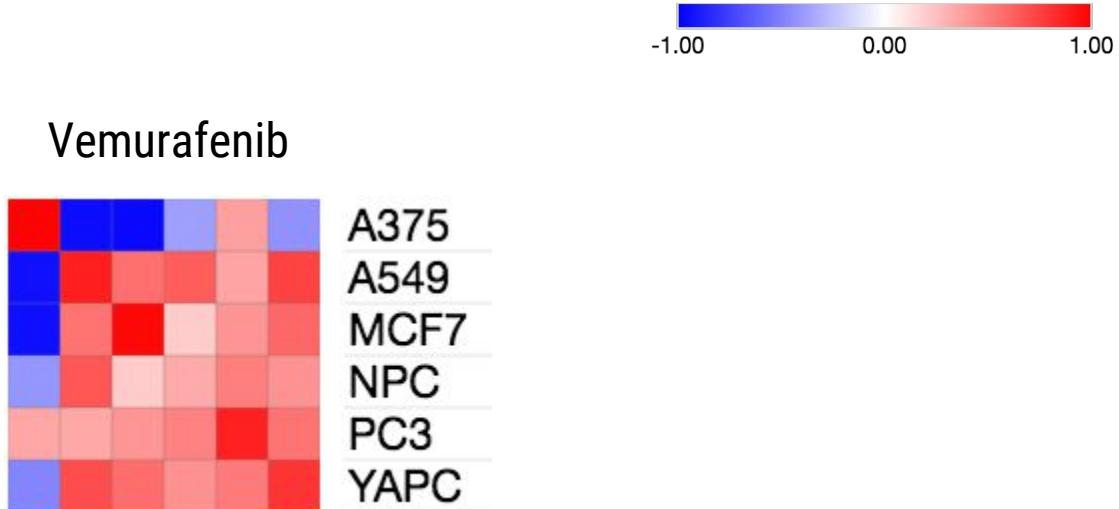
# Test and background distributions on real data

More separation is a higher connectivity score



# Connectivity matrix for vemurafenib

Very clear that A375 stands out



# Summary

## Why we find connectivity useful

---

1. Collapses replicates
2. Compares similarities (Spearman correlations) to a background, which helps evaluate significance

### BONUS: Connectivity allows comparison across assays

- e.g. Global chromatin profiling (GCP) measures 59 features of chromatin modifications
- e.g. L1000 measures gene expression for ~1000 genes

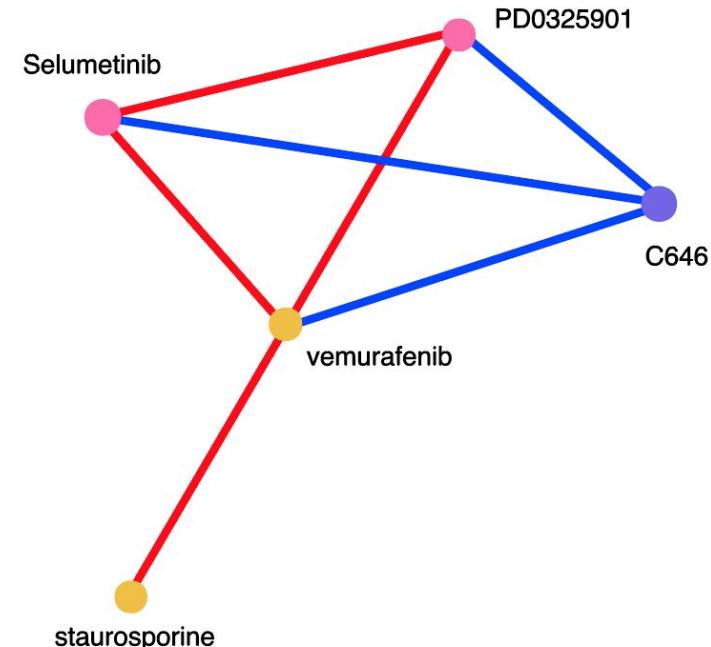
# Network visualizations

Can see second-order connections

---

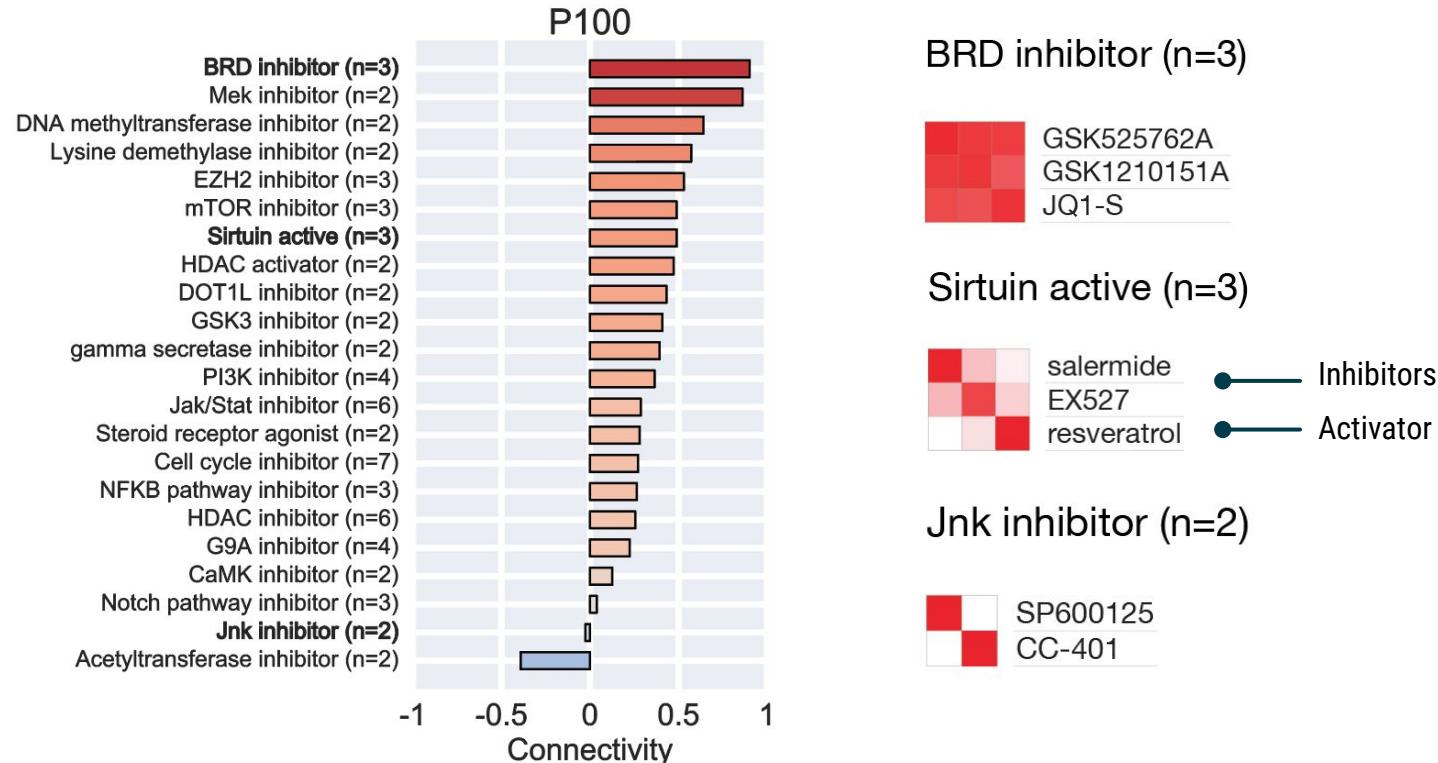
## Undirected graphs

1. Symmetrize connections
2. Pick a threshold; throw out connections below threshold
3. Keep connections related to a query of interest
4. Keep also connections also between first-order connections

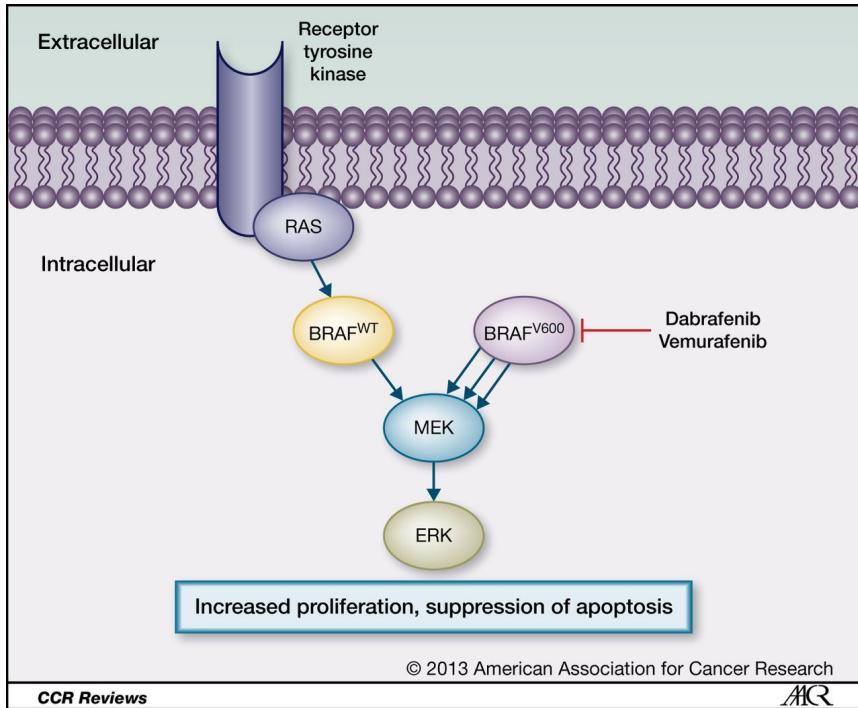


## **(5) Biological applications of connectivity**

# 1. Compounds with the same annotated mechanism of action (MoA) generally have high connectivity to each other

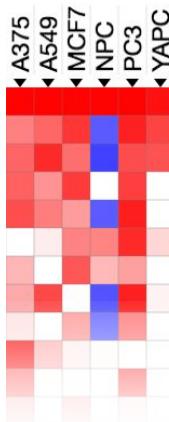


## 2. Vemurafenib connects to Mek inhibitors in only A375 cells



### 3. Pazopanib has different connectivities in NPCs

Pazopanib



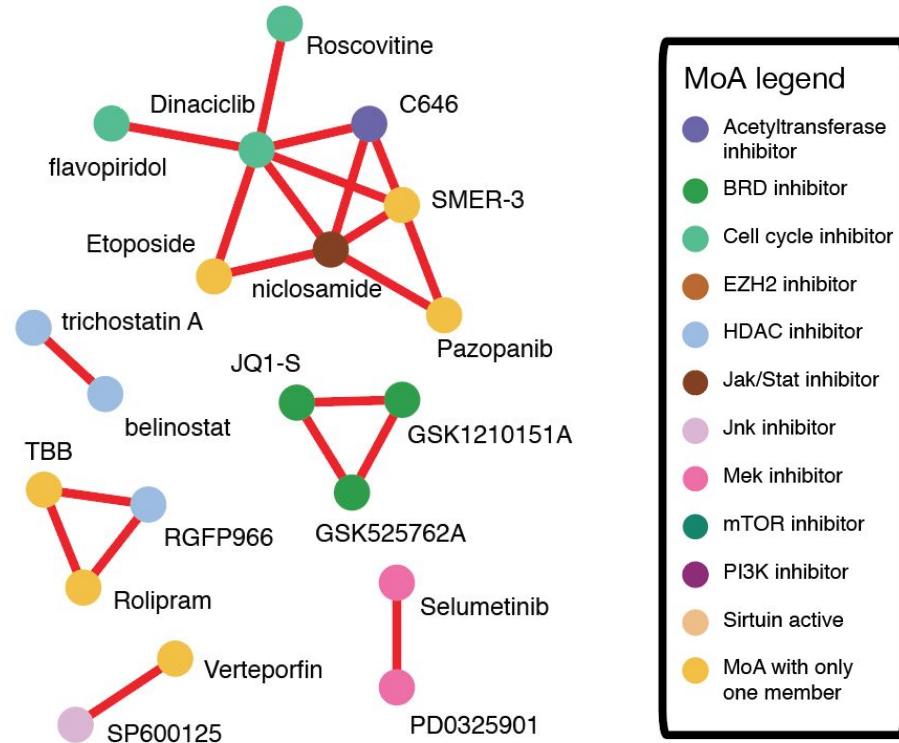
#### Mechanism of Action

- PDGFR, VEGFR (and c-KIT, FGFR) inhibitor
- \* E3 ligase inhibitor
- \* Jak/Stat inhibitor
- Sirtuin inhibitor
- \* Cell cycle inhibitor
- Multikinase inhibitor (Bcr-Abl)
- Jnk inhibitor
- \* Acetyltransferase inhibitor
- \* Cell cycle inhibitor
- PI3K inhibitor
- Cell cycle inhibitor
- Jak/Stat inhibitor

Drug	Mechanism of Action
Pazopanib	PDGFR, VEGFR (and c-KIT, FGFR) inhibitor
TG101348	Jak/Stat inhibitor
staurosporine	Kinase inhibitor; general
Nilotinib	Multikinase inhibitor (Bcr-Abl)
UNC-0646	G9A inhibitor
methylstat	Lysine demethylase inhibitor
zebularine	DNA methyltransferase inhibitor
tacrolimus	Calcineurin inhibitor
SP600125	Jnk inhibitor
LY-294002	PI3K inhibitor
SCH 900776	Cell cycle inhibitor
bafilomycin A1	Inhibitor of the vacuolar-type H <sup>+</sup> -ATPase
JQ1-S	BRD inhibitor
Tofacitinib	Jak/Stat inhibitor
BIX-01294	G9A inhibitor
Dinaciclib	Cell cycle inhibitor
SMER-3	E3 ligase inhibitor
C646	Acetyltransferase inhibitor
trichostatin A	HDAC inhibitor
niclosamide	Jak/Stat inhibitor

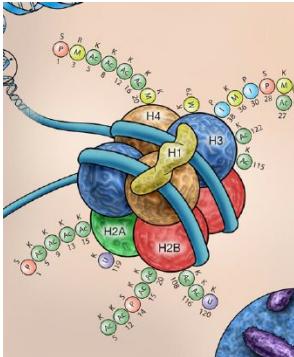
# 4. Strongest connections in P100

- Aggregated across cell lines
- 0.5% cutoff



# 5. Connecting perturbations to baseline states

GCP performed on 90 compounds x 6 cell lines



## Example GCP features

H3K4me0

H3K4me1

H3K4me2

H3K4ac1

H3K18ub1K23ac0

H3K27me1K36me3

H3K27me3K36me2

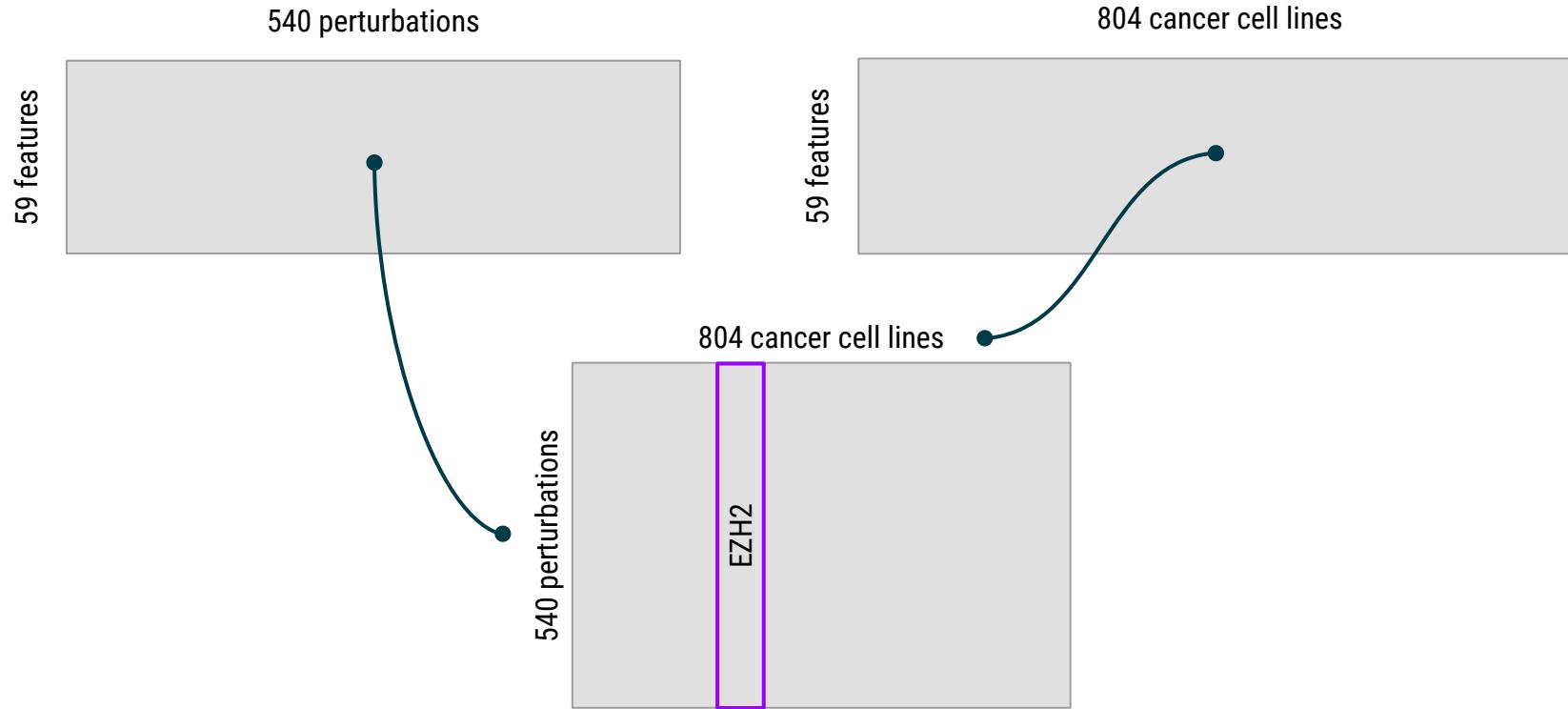
H3K27ac1K36me0

59 features

540 perturbations

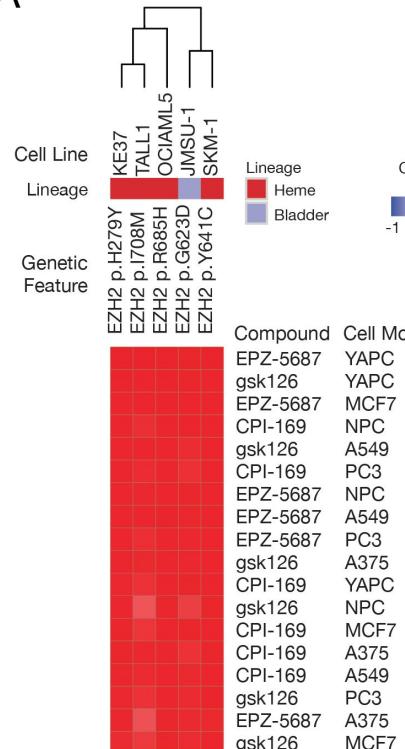
# 5. Connecting perturbations to baseline states

GCP also performed on 804 cell lines

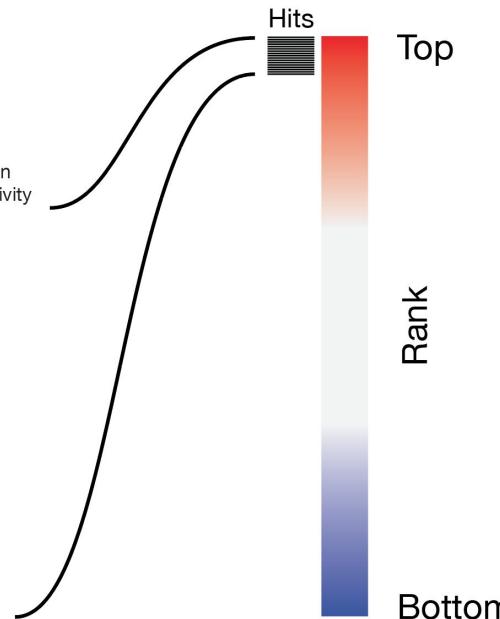


# EZH2 mutants connect to EZH2 inhibition

A

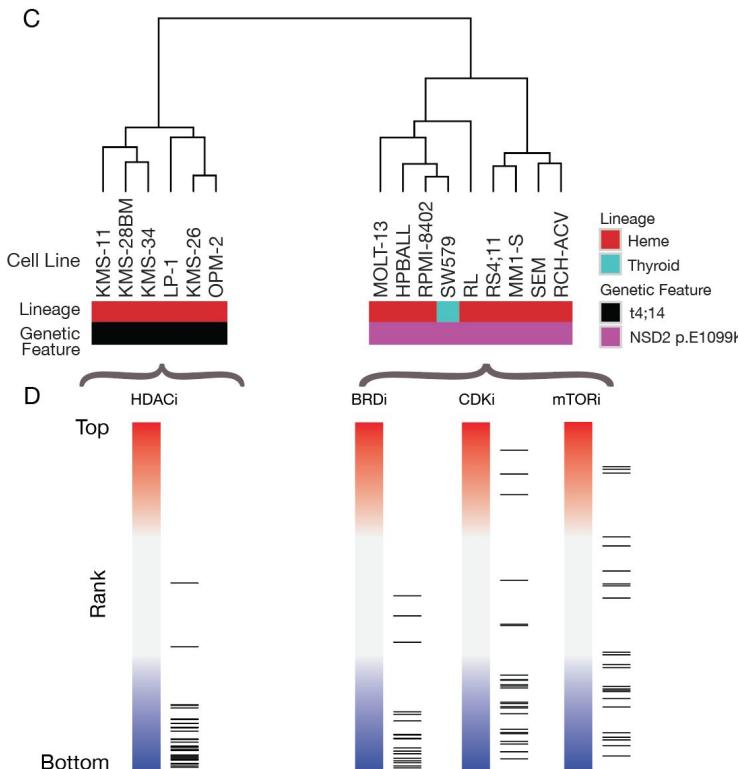


B



# Two clusters of NSD2 mutants

## Negative connections to perturbations suggest therapeutics



Expert Rev Hematol. 2017 Mar;10(3):229-237. doi: 10.1080/17474086.2017.1280388. Epub 2017 Feb 1.

### Deacetylase inhibitors: an advance in myeloma therapy?

Laubach JP<sup>1</sup>, San-Miguel JF<sup>2</sup>, Hungria V<sup>3</sup>, Hou J<sup>4</sup>, Moreau P<sup>5</sup>, Lonial S<sup>6</sup>, Lee JH<sup>7</sup>, Einsele H<sup>8</sup>, Alsina M<sup>9</sup>, Richardson PG<sup>1</sup>.

ricolinostat are also discussed. Expert commentary: DACi are a unique and effective new class of agents for the treatment of MM, with panobinostat being the first to have clinically meaningful benefit for patients with relapsed or refractory MM. Optimization of dose and

J Hematol Oncol. 2016 Feb 18;9:12. doi: 10.1186/s13045-016-0241-x.

### Synergism between the mTOR inhibitor rapamycin and FAK down-regulation in the treatment of acute lymphoblastic leukemia.

Shi PJ<sup>1</sup>, Xu LH<sup>2</sup>, Lin KY<sup>3</sup>, Weng WJ<sup>4</sup>, F...

### Potent efficacy of combined PI3K/mTOR and JAK or ABL inhibition in murine xenograft models of Ph-like acute lymphoblastic leukemia.

Tasian SK<sup>1,2</sup>, Teachey DT<sup>1,2</sup>, Li Y<sup>1</sup>, Shen F<sup>1</sup>, Harvey RC<sup>3</sup>, Chen IM<sup>3</sup>, Ryan T<sup>1</sup>, Vincent TL<sup>1</sup>, Willman CL<sup>3</sup>, Dard AE<sup>2,4</sup>, Hwang SP<sup>1,2</sup>, Lab ML<sup>5,6</sup>, Carroll M<sup>2,4</sup>, Cordon SA<sup>1,2</sup>

Oncotarget. 2015 Oct 13;6(31):32089-103. doi: 10.18633/oncotarget.5156.

### Co-targeting of Bcl-2 and mTOR pathway triggers synergistic apoptosis in BH3 mimetics resistant acute lymphoblastic leukemia.

Iacovelli S<sup>1</sup>, Ricciardi MR<sup>2</sup>, Allegretti M<sup>1</sup>, Mirabili S<sup>1</sup>, Licchetta R<sup>2</sup>, Bergamo P<sup>1</sup>, Rinaldo C<sup>3</sup>, Zeuner A<sup>4</sup>, Foà R<sup>1</sup>, Milella M<sup>5</sup>, McCubrey JA<sup>6</sup>, Martelli AM<sup>7</sup>, Tafuri A<sup>2</sup>.



**<https://software.broadinstitute.org/morpheus/>**

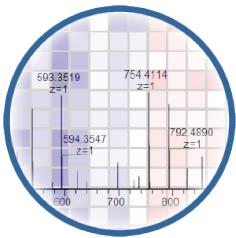
**<https://clue.io/morpheus>**

# Questions?

Lev Litichevskiy  
[lev@broadinstitute.org](mailto:lev@broadinstitute.org)

# Acknowledgements

---



PCCSE



Amanda  
Creech



Desiree Davison



Shawn Egri



Jake Jaffe



Katherine  
DeRuff



Xiaodong Lu



Adam Officer



Malvina  
Papanastasiou



Ryan Peckner



Sebastian  
Vaca

# Appendix

# Differential expression

**Plate control rather than vehicle control**

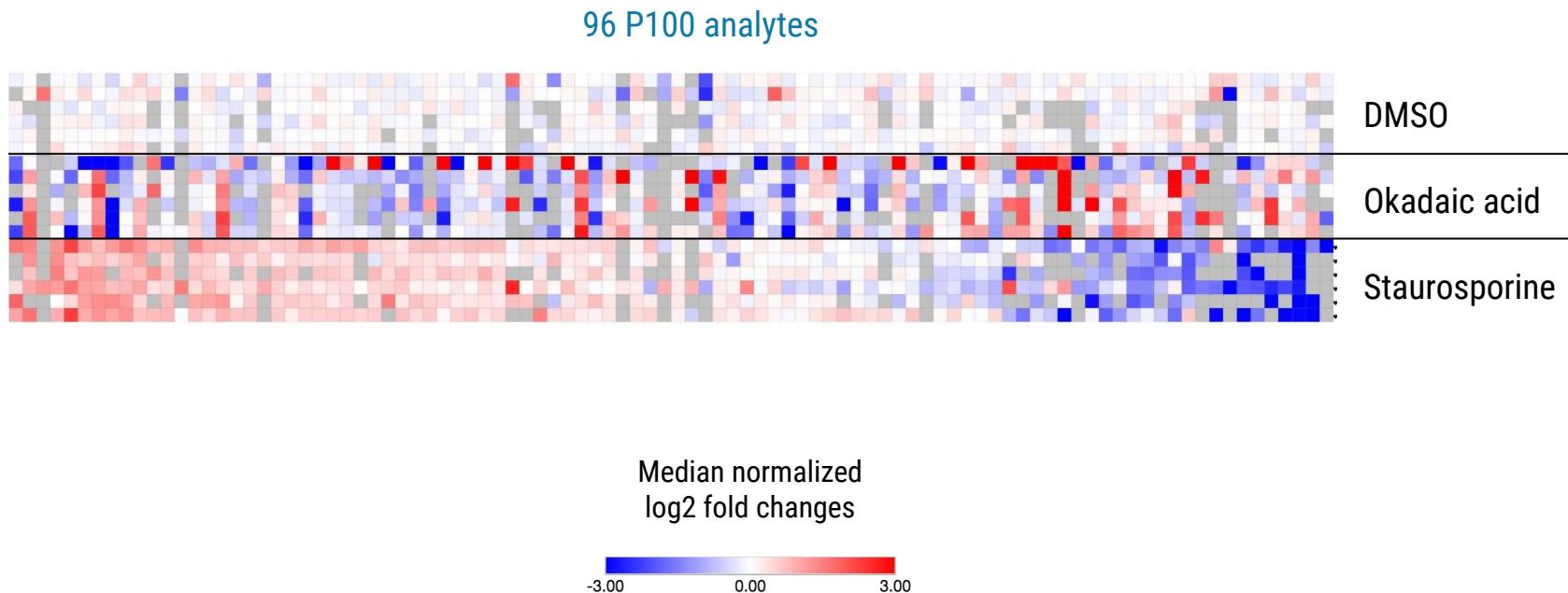
---

Typically, perturbation is compared to DMSO (i.e. vehicle)

We compare our treatments to the median profile on the plate (i.e. row median normalization)

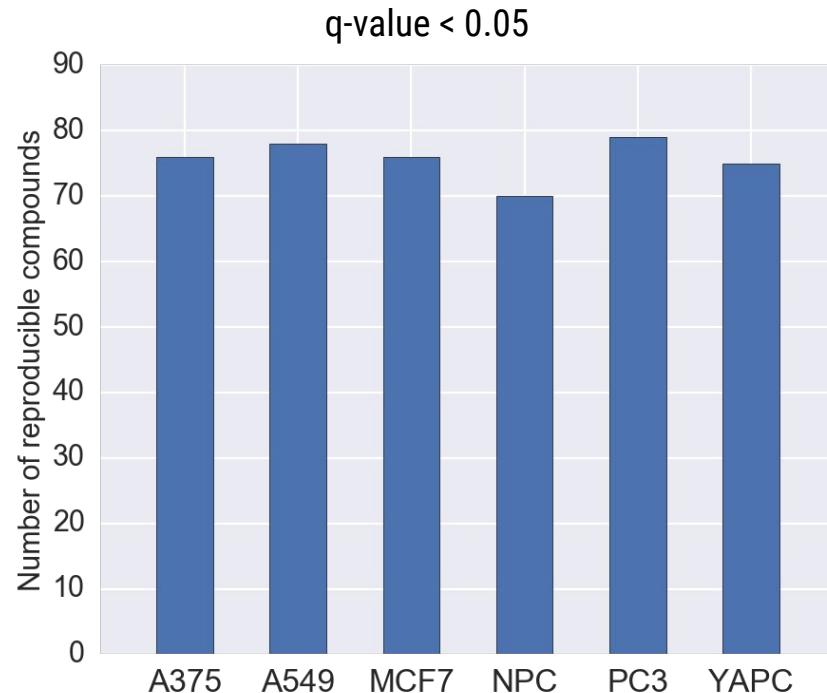
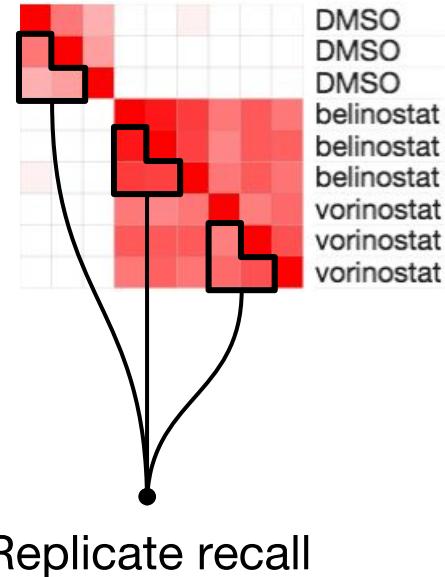
# Profiles show expected effects

Staurosporine and okadaic acid have widespread effects



# Similarity between biological replicates is high

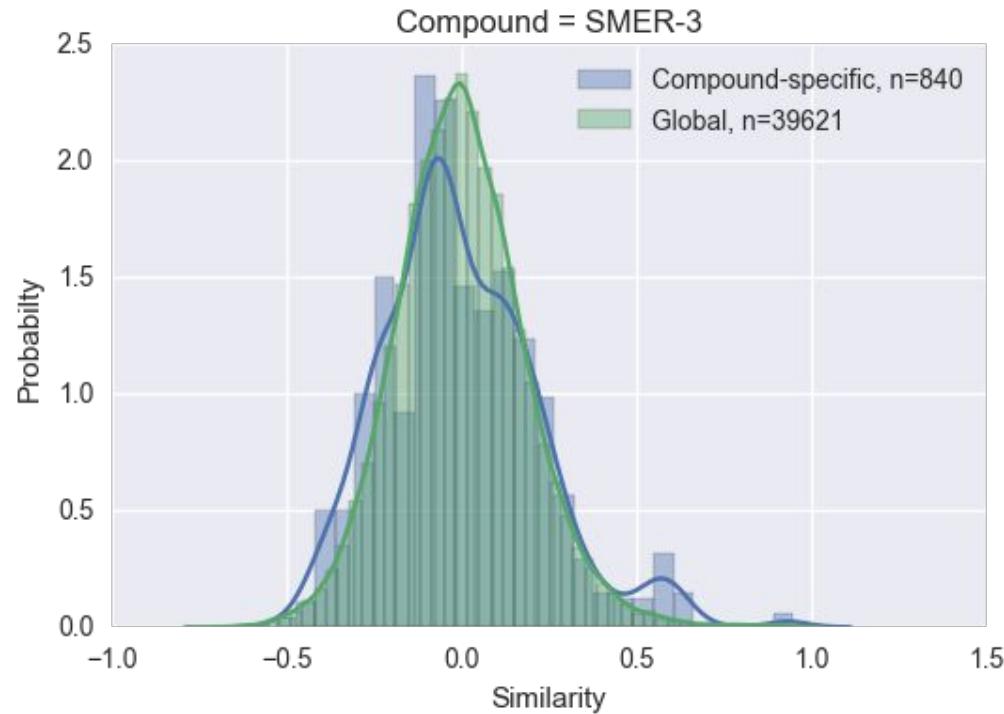
---



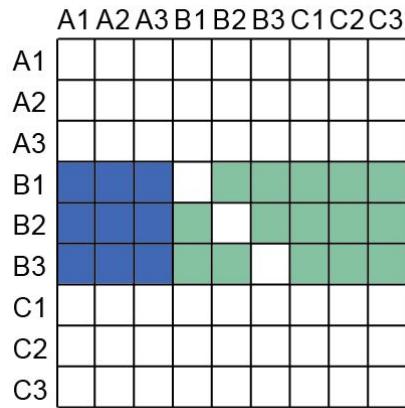
# Why a compound-specific background?

Theoretically allows us to find more nuanced connections

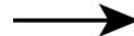
In reality, the compound-specific background is similar to the global background



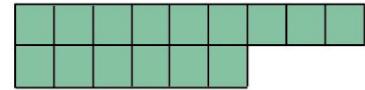
# Example: connectivity between compounds A and B



Similarities



Background distribution



Background distribution is  
for compound B



KS-test

