Going Wide with Targeted Proteomics: The Rise of Unhole-y Data

Jake Jaffe

The Broad Institute

LINCS Data Forum March 2013

Shotgun Proteomics Experiments are Stochastic

- Instrumentation inherently undersamples population
  - Speed
  - Sensitivity
  - Dynamic Range

Data Shapes in Proteomics

1. Long and skinny

Peptides/proteins

	1	2
###	###	###

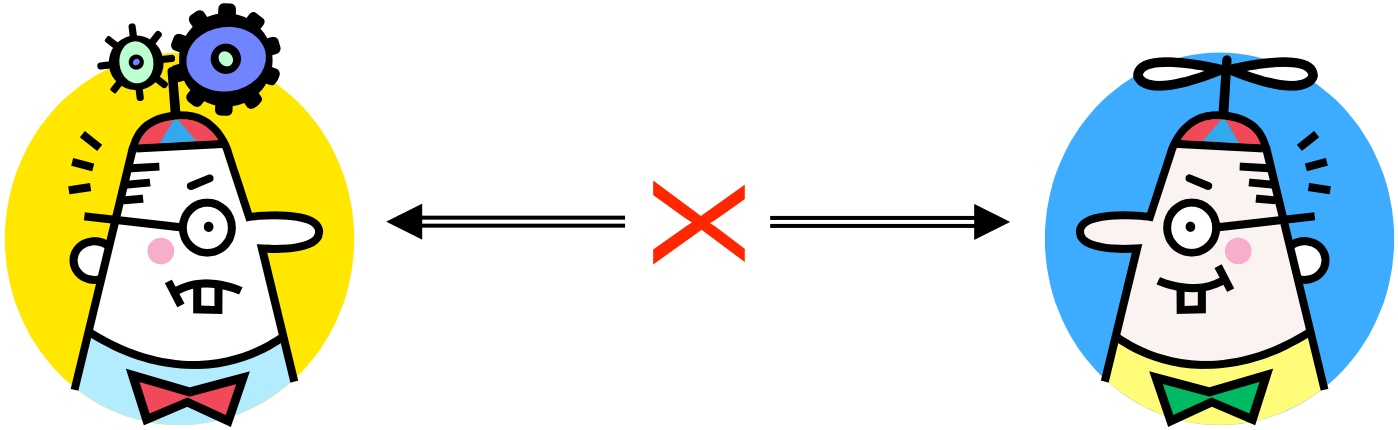
2. Hole-y

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
###	###	###	###	###	###	###	###	###	###	###	###	###	###	###

3. Short and wide

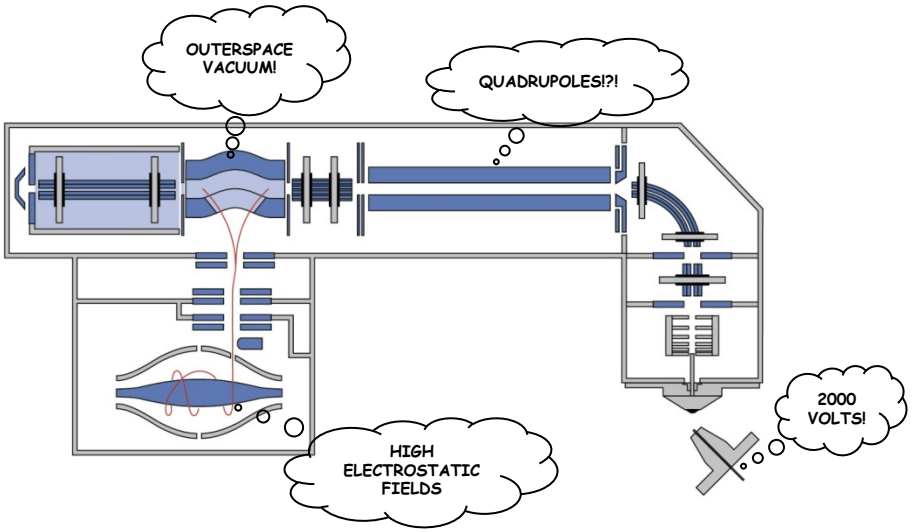
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
###	###	###	###	###	###	###	###	###	###	###	###	###	###	###

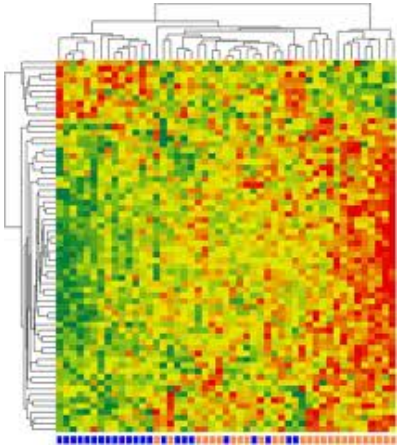
It’s hard to be hole-y



Instrument Jockeys

Data Geniuses

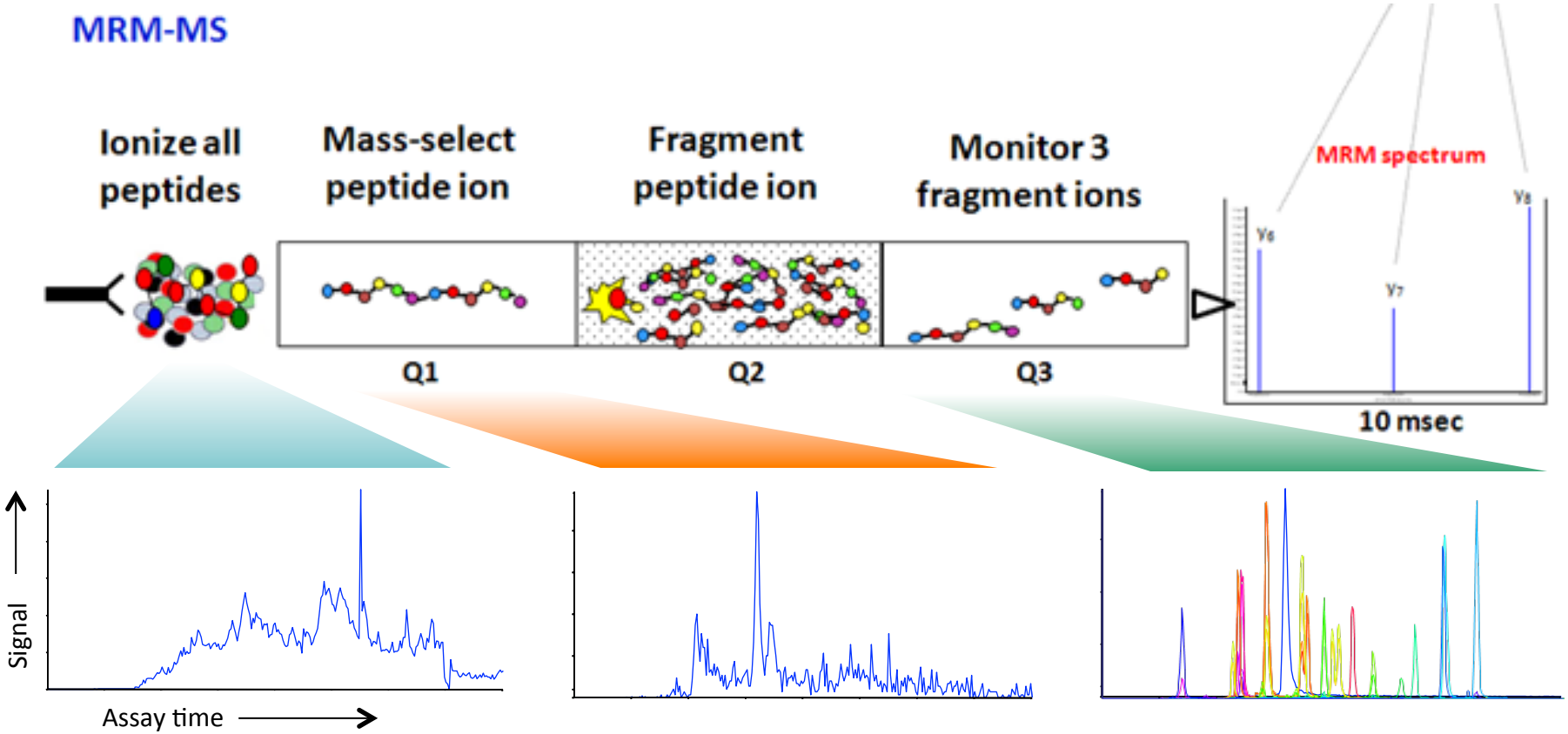




Solution: Targeted Proteomics

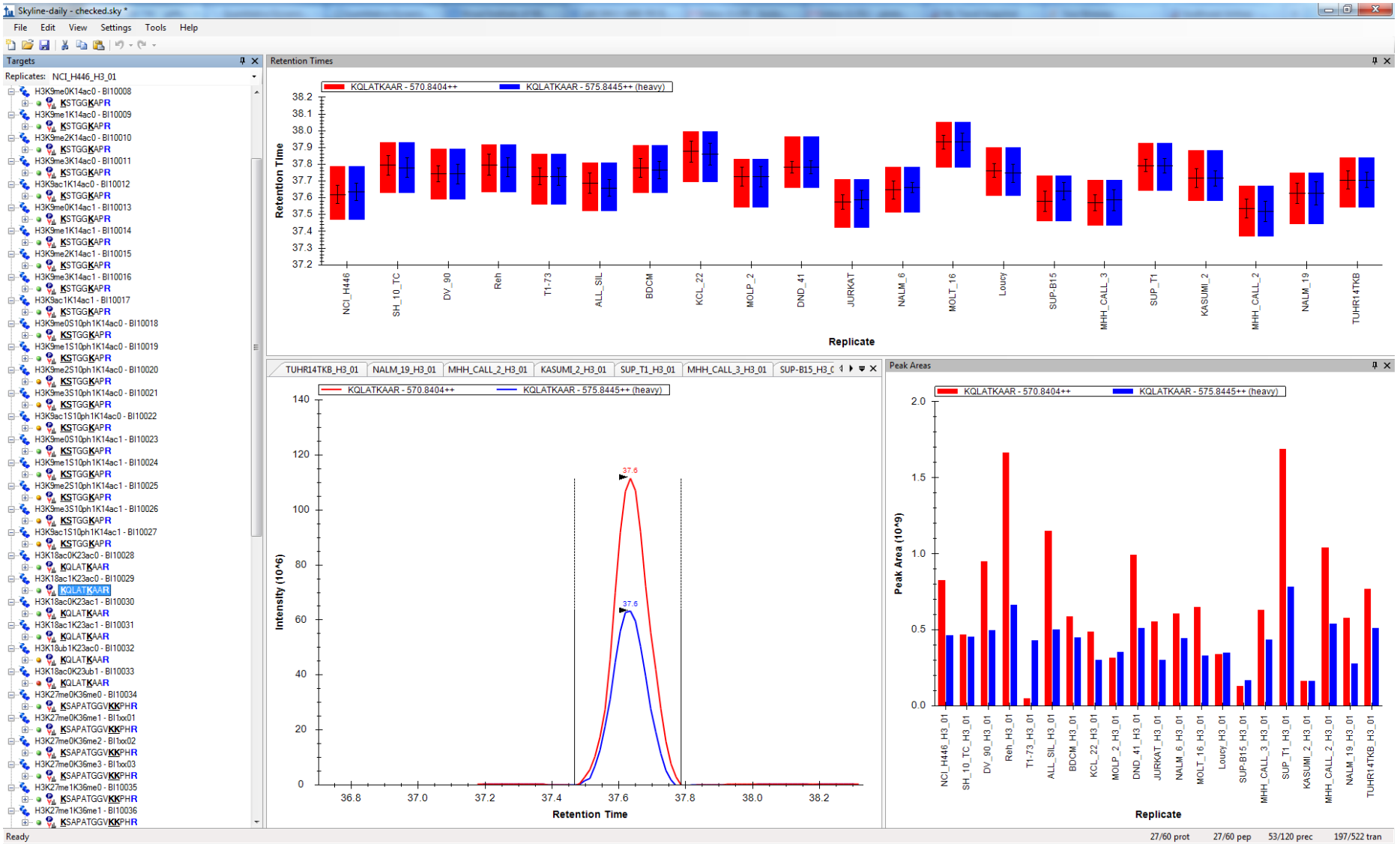
- Limit proteomic assay to finite number of meaningful analytes
  - Multiplex ~ 100-200
- Introduce internal standards for guaranteed observation of analytes
- Internal standards = better quantification
- When hole-yness is not an option

Targeted Proteomics Implementation

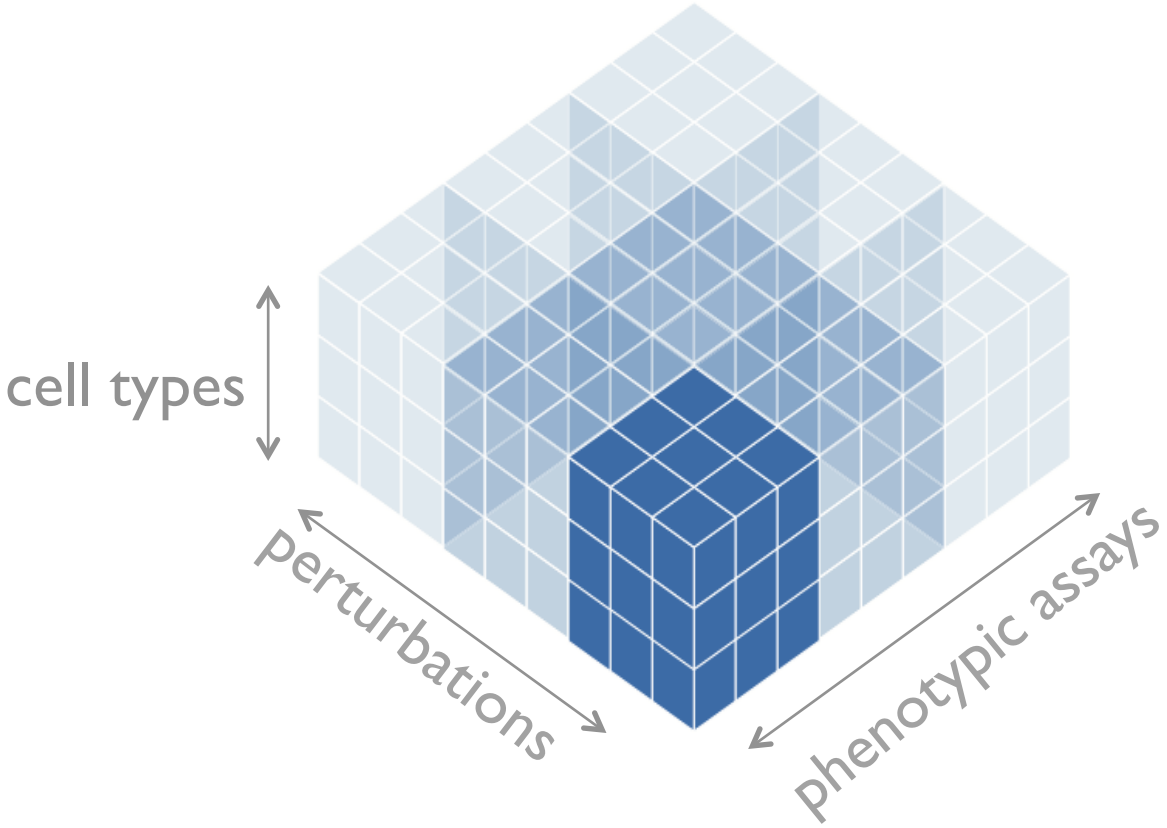


- Just need to know m/z, r.t., and fragment *m/z*s
- Generally cheaper and faster than shotgun proteomics

Target Proteomics Quantification

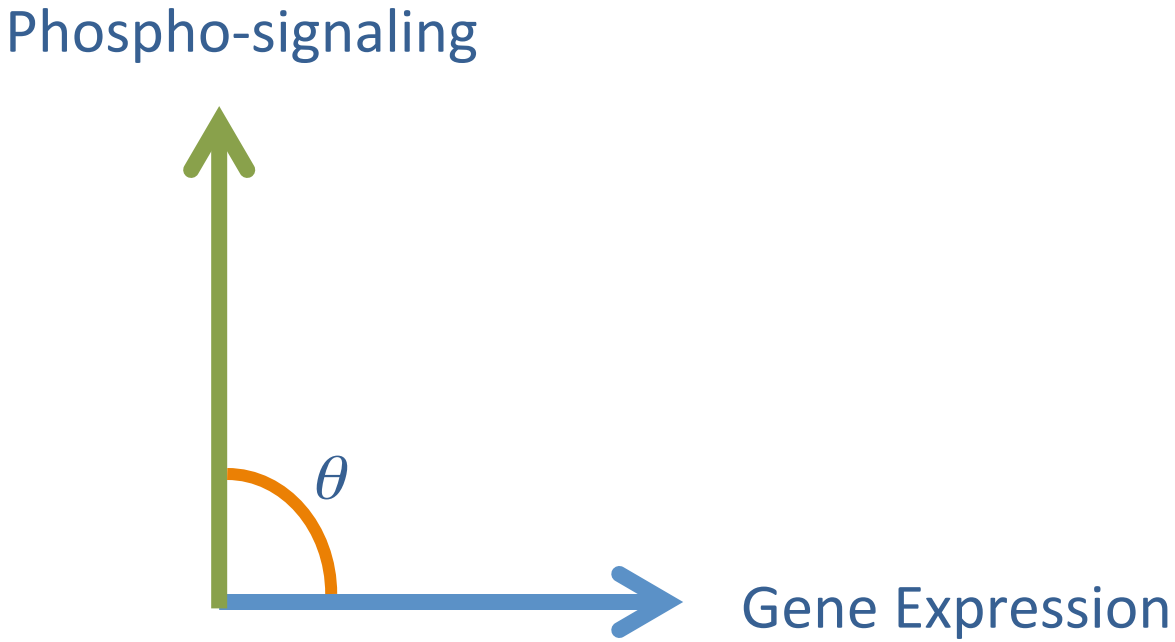


Targeted Proteomics and LINCS



- Desirable to have “complete” data in every assay

Targeted Phosphoproteomics Application: P100

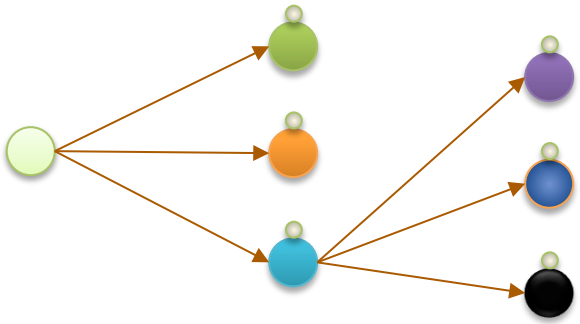


- θ may be large
- Phosphosignaling is inaccessible through expression profiling
- Phosphosignaling can occur on different time scales than gene expression

Reduced Representation Signaling Set: P100

- We are building a phosphosignaling analog of the L1000 assay – the P100

- Phosphosignaling is likely coordinated
  - Kinase-substrate relationships tend to be 1-to-many
  - *A priori* expectation of coordinate regulation of sites

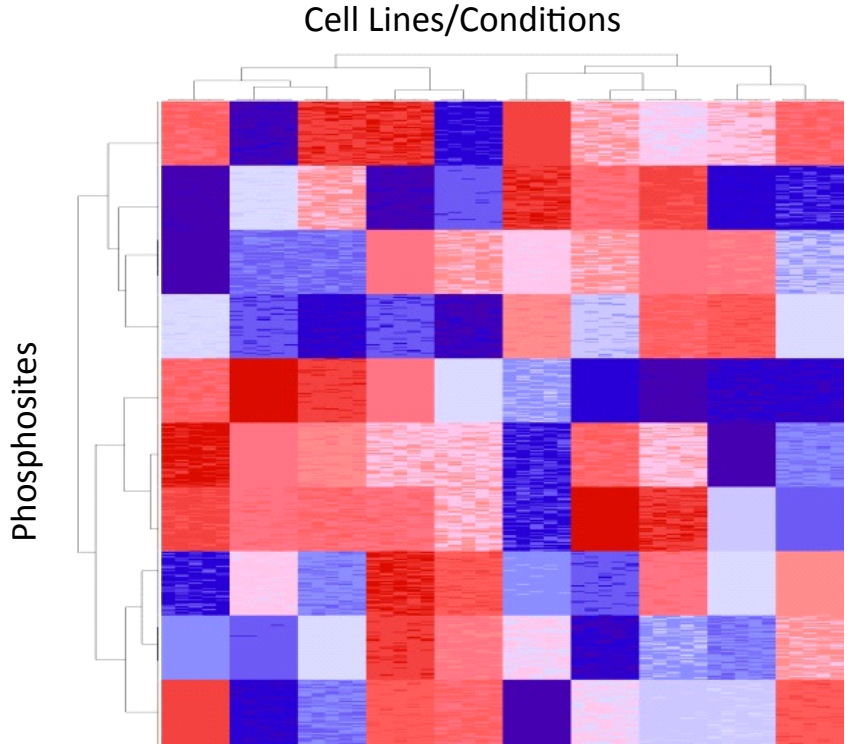


- Don't need to monitor every phosphosite in the assay
  - Amenable to L1000-like reductionist approaches

Generation of source data from LINCS Lines and Compounds

2 x MCF7 / PC3 / HL60 x

Compound
captopril
ambroxol
paclitaxel
carmustine
fulvestrant
clindamycin
chlortetracycline
chlorzoxazone
doxorubicin
daunorubicin
GW-8510
staurosporine
dexverapamil
mesalazine
anisomycin
digitoxigenin
digoxigenin
digoxin
lanatoside C
irinotecan
scriptaid
trichostatin A
MS-275
H-7
geldanamycin
valproic acid

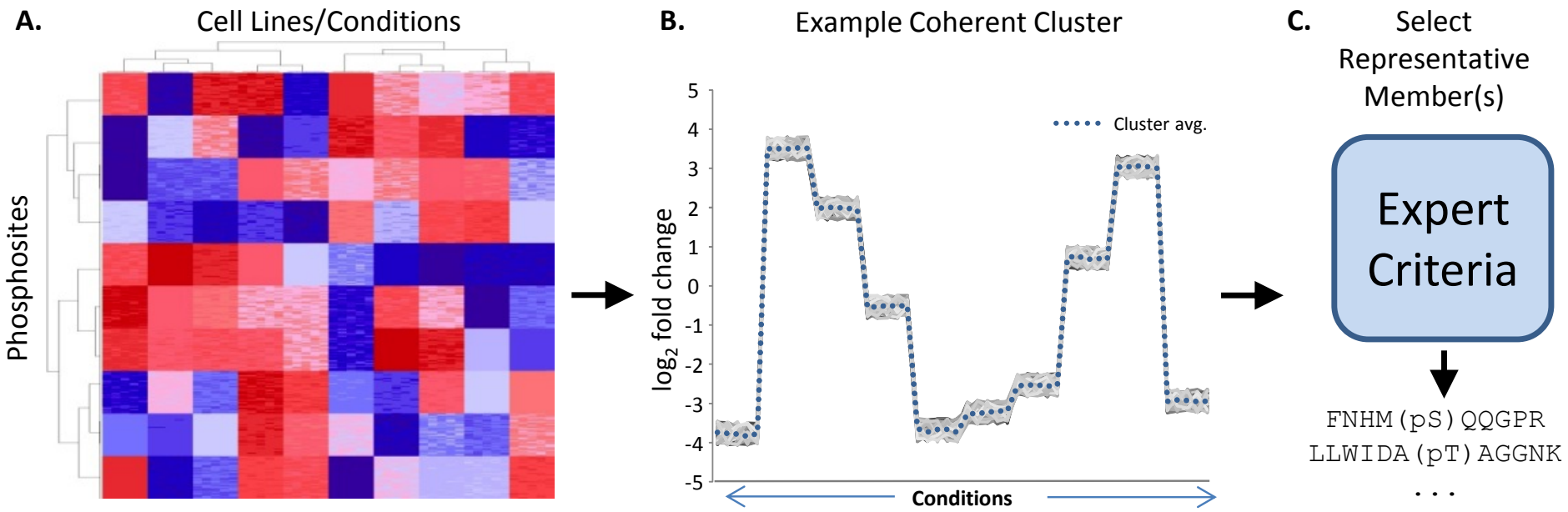


Chosen based on:

- **Modulates kinase/phosphatase expression**
- **Concentration from CMAP**
- **Have matching gene expression data**
- **Several same class examples**
- **PosCon = staurosporine**

Jinal Patel, Xiaodong Lu

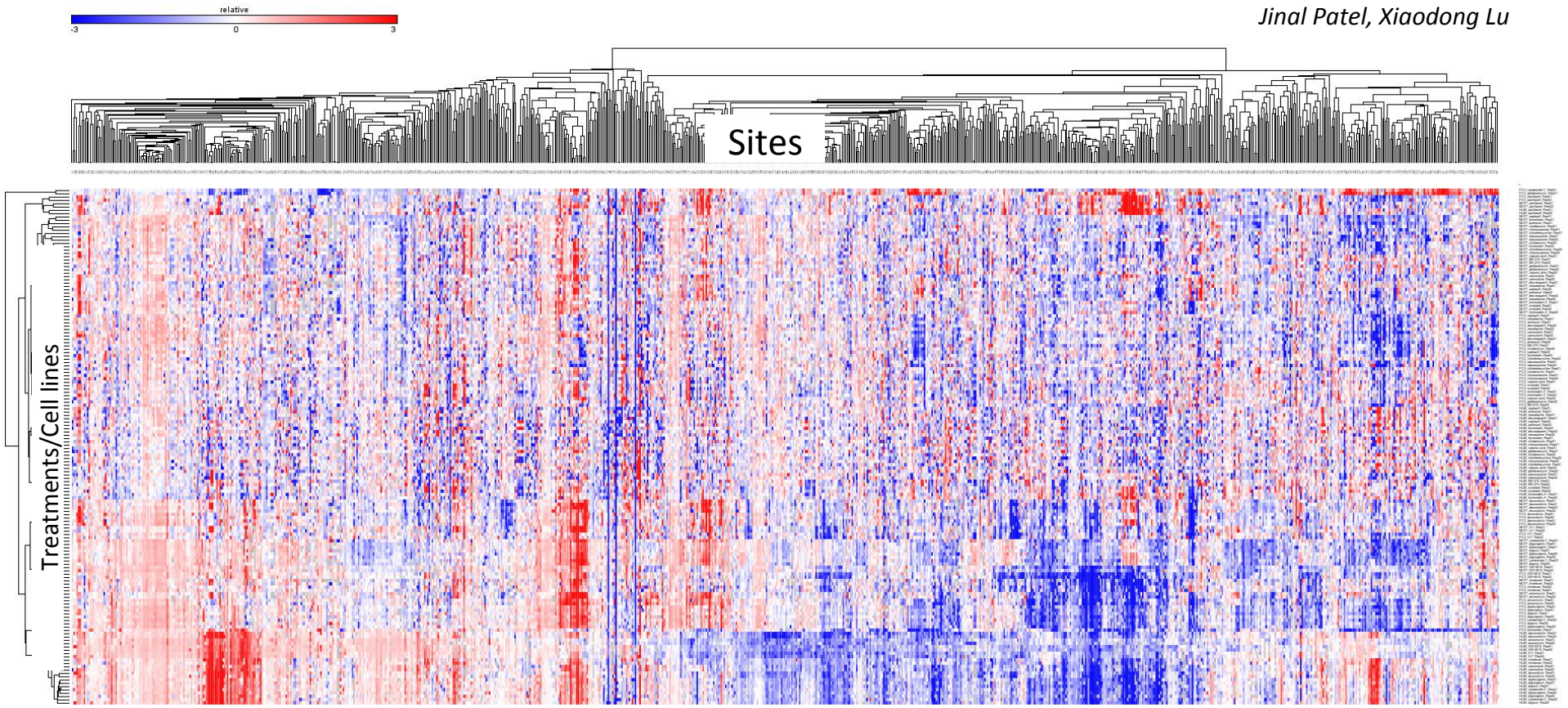
Learn a practical reduced probe set from the data



- Natural synergy between L1000 and P100 projects
- Exploit existing robust methods

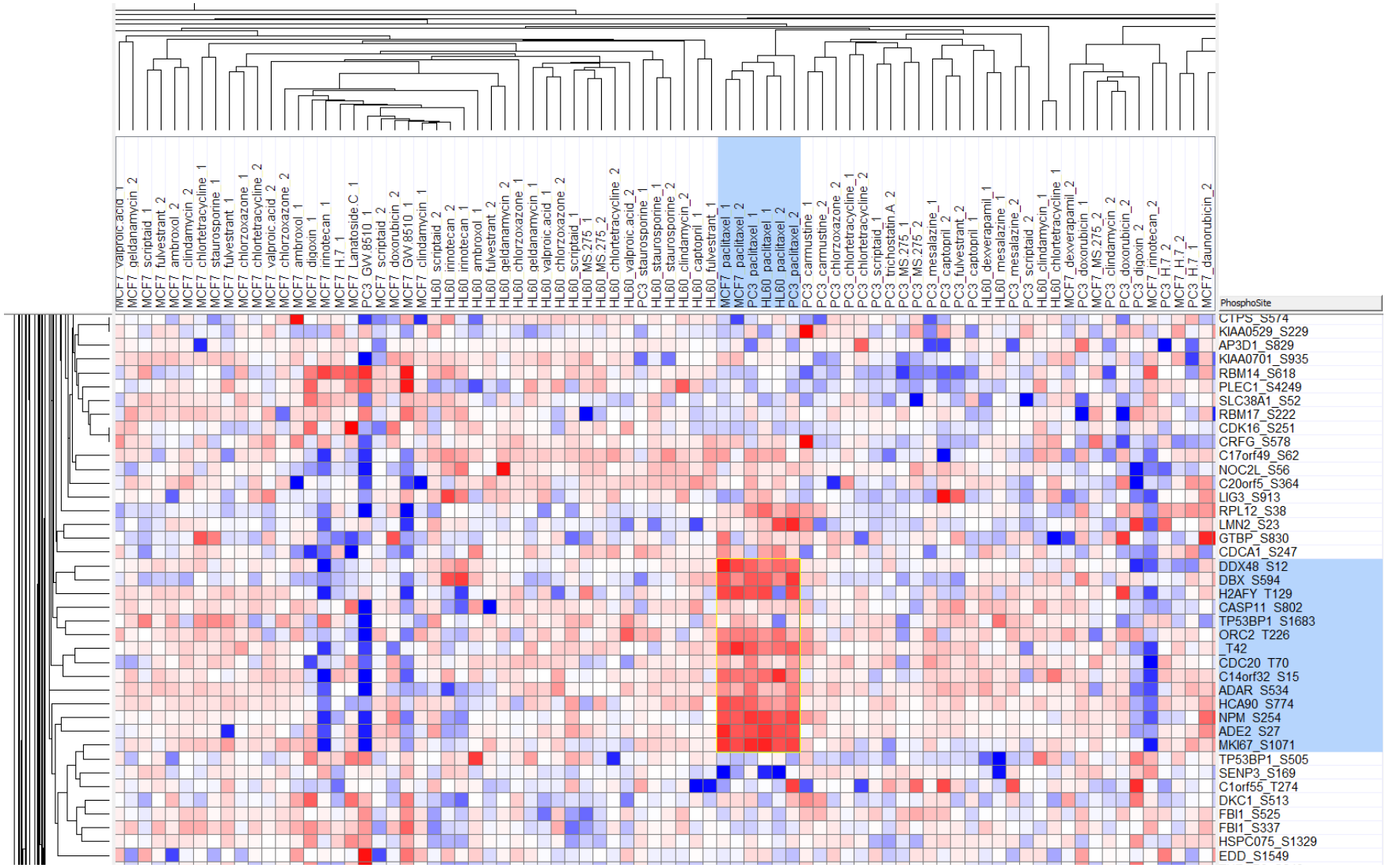
P100 Source Dataset

Jinal Patel, Xiaodong Lu



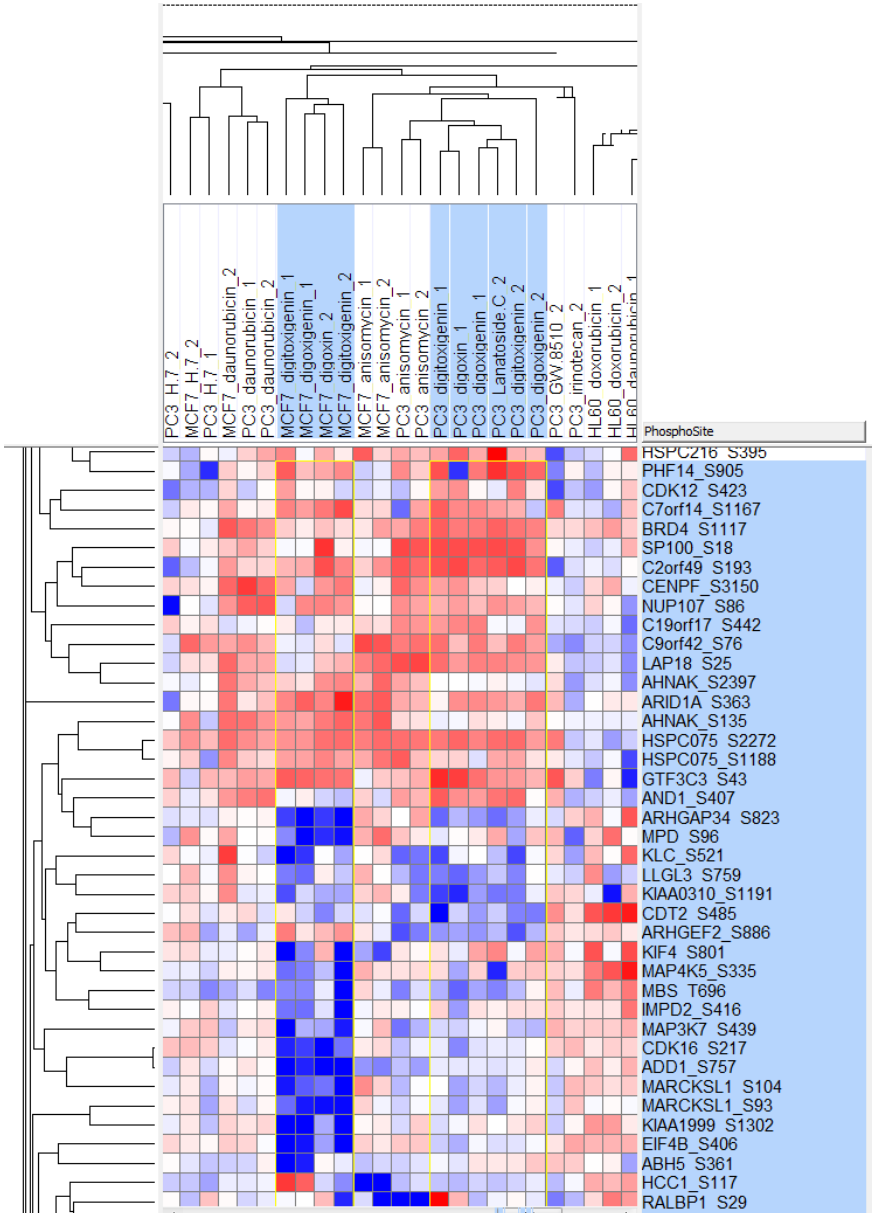
- Largest systematic set of perturbations with phosphoproteomics readout in existence
  - Over 10,000 phosphosites observed
- Over 1,200 sites present in >75% of all experiments

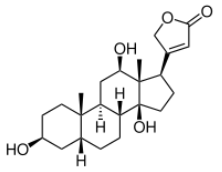
Signals span lineages



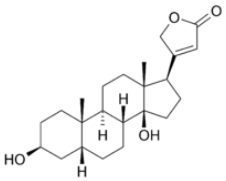
- All 6 Taxol treatments cluster – regardless of lineage

Signals group related compounds and replicates

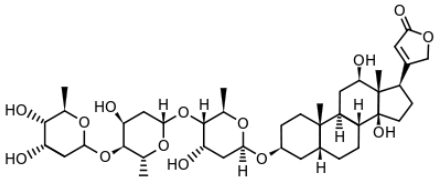




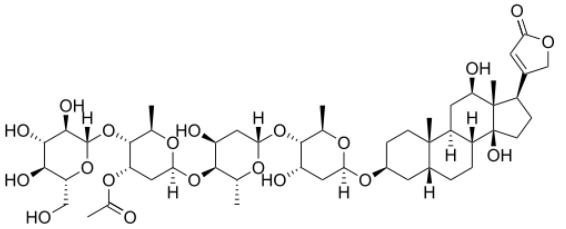
Digoxigenin



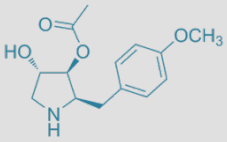
Digitoxigenin



Digoxin

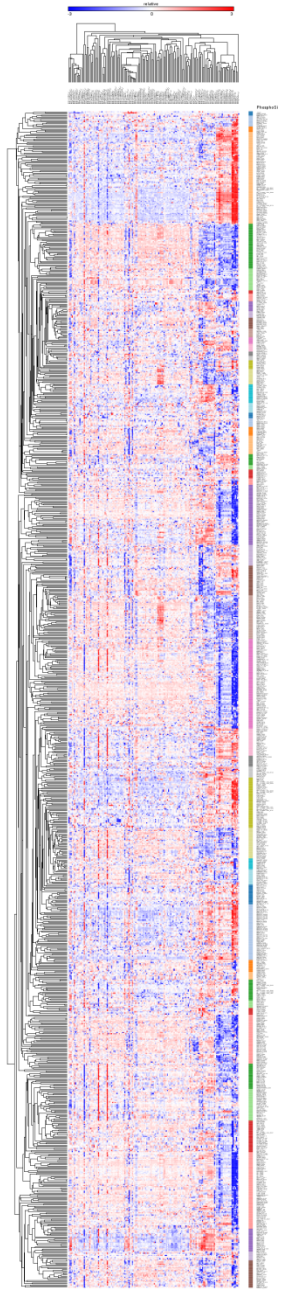


Lanatoside C

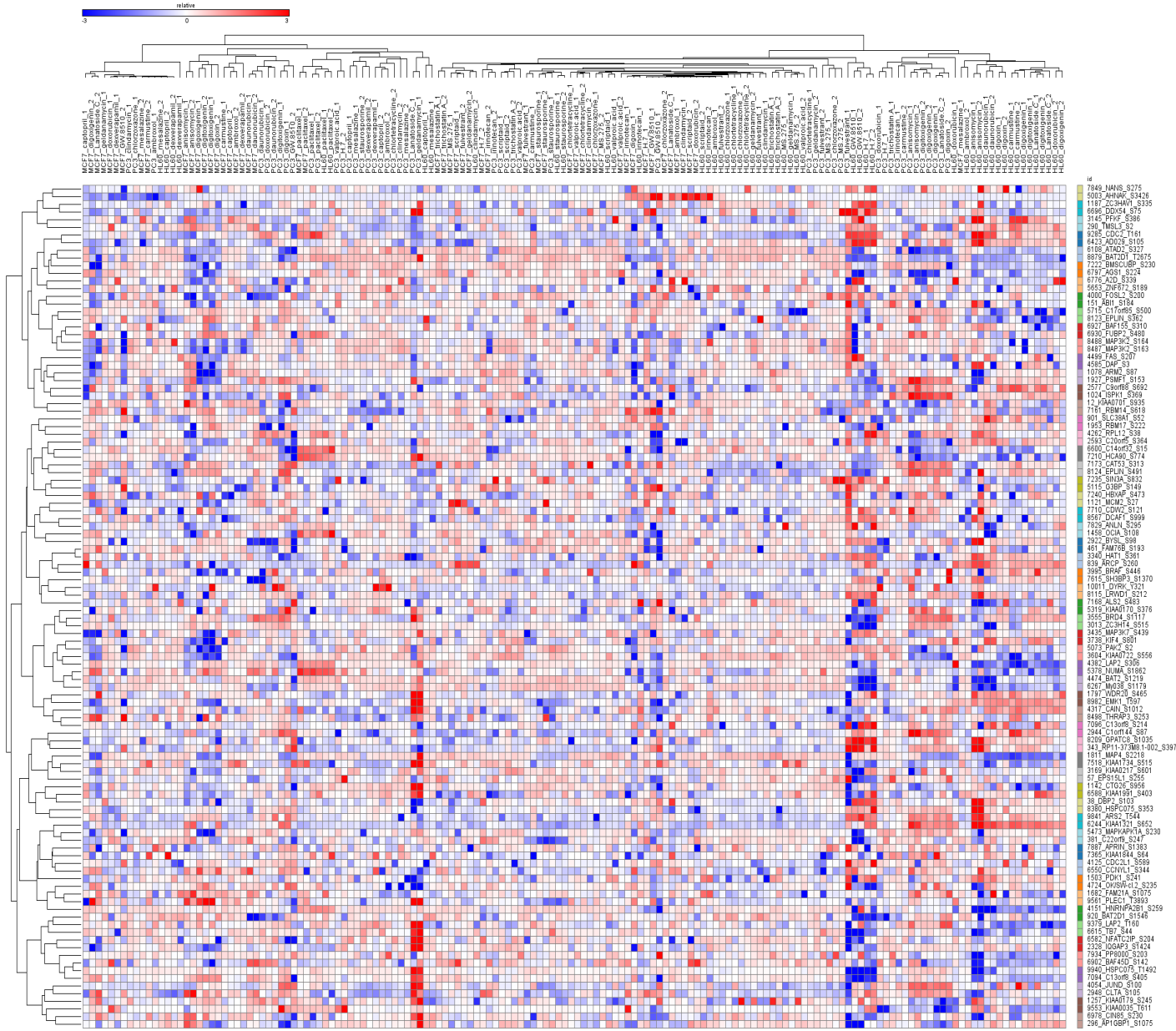


Anisomycin

Dimensionality reduction: >1000 to 55 with 2x probe redundancy



Roger Hu



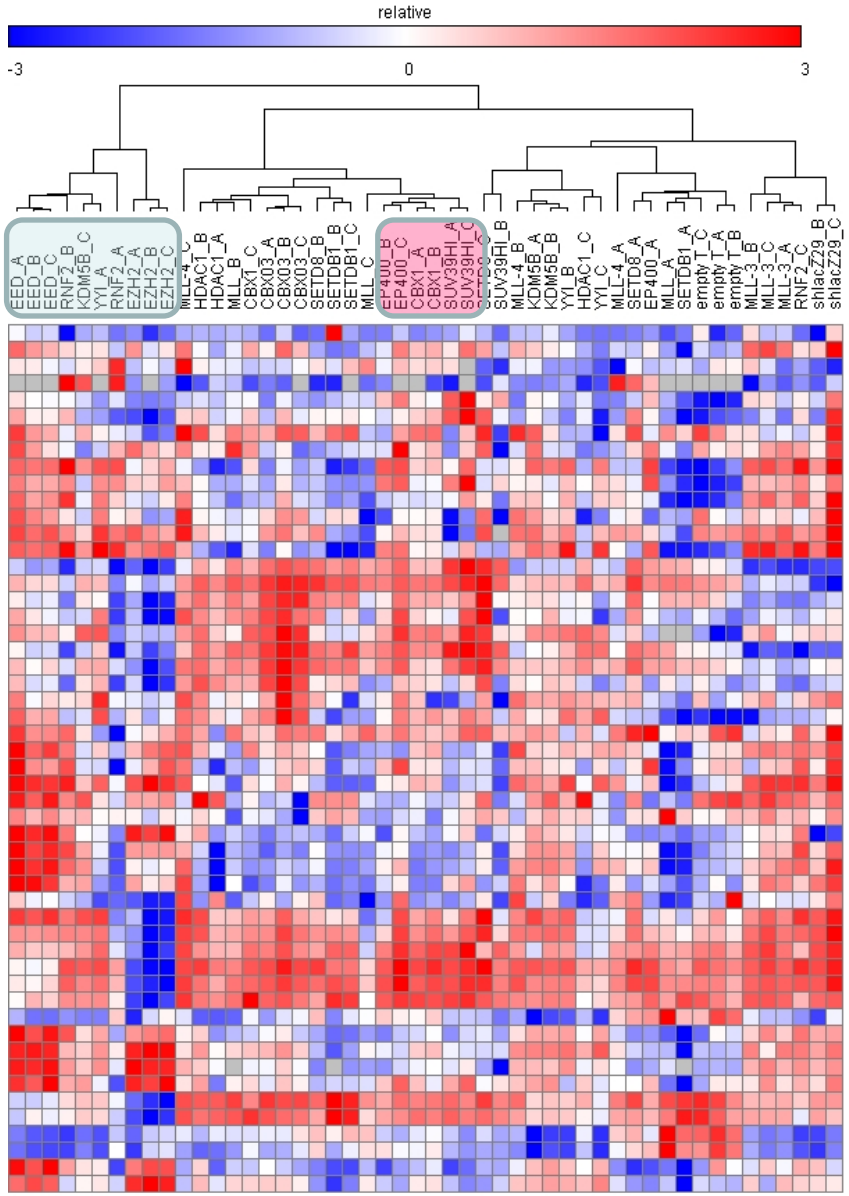
P100 Progress…

- Reduced Representation Targeted Phosphoproteomics Assay
- Large data set collected
  - Available at LINCS website
- Internal standards in hand
- Final assay configuration in progress
  - Expected to be deployable in multiple laboratories

Example: A targeted chromatin modification assay

- Targeted high resolution assay against chromatin modification states
  - Over 60 possible combinations of histone marks are monitored
- We can query chromatin signatures from the MRM assay and infer function
  - Compare laboratory perturbations
  - Compare differing genotypes
  - Compare compound treatments
  - Compare across these experiment types
- The “CMap“ of chromatin-space

Molecular Chromatin Signatures of knockdowns



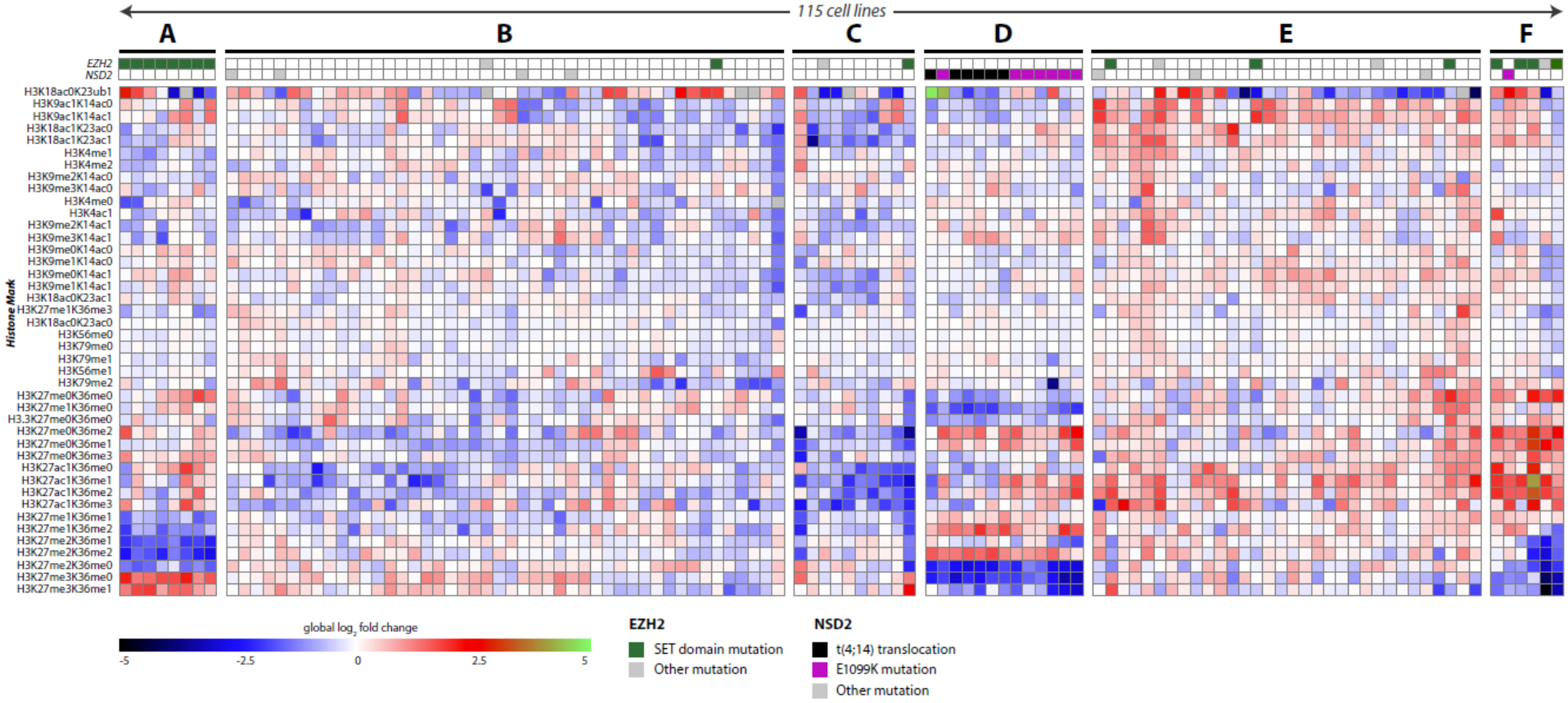
H3K4me0 - BI10003
H3K4me1 - BI10004
H3K4me2 - BI10005
H3K4me3 - BI10006
H3K9me0K14ac0 - BI10008
H3K9me1K14ac0 - BI10009
H3K9me2K14ac0 - BI10010
H3K9me3K14ac0 - BI10011
H3K9ac1K14ac0 - BI10012
H3K9me0K14ac1 - BI10013
H3K9me1K14ac1 - BI10014
H3K9me2K14ac1 - BI10015
H3K9me3K14ac1 - BI10016
H3K9ac1K14ac1 - BI10017
H3K9me0S10ph1K14ac0 - BI10018
H3K9me1S10ph1K14ac0 - BI10019
H3K9me2S10ph1K14ac0 - BI10020
H3K9me3S10ph1K14ac0 - BI10021
H3K9ac1S10ph1K14ac0 - BI10022
H3K9me0S10ph1K14ac1 - BI10023
H3K9me1S10ph1K14ac1 - BI10024
H3K9me2S10ph1K14ac1 - BI10025
H3K9me3S10ph1K14ac1 - BI10026
H3K9ac1S10ph1K14ac1 - BI10027
H3K18ac0K23ac0 - BI10028
H3K18ac1K23ac0 - BI10029
H3K18ac0K23ac1 - BI10030
H3K18ac1K23ac1 - BI10031
H3K18ub1K23ac0 - BI10032
H3K18ac0K23ub1 - BI10033
H3K27me0K36me0 - BI10034
H3K27me1K36me0 - BI10035
H3K27me1K36me1 - BI10036
H3K27me1K36me2 - BI10037
H3K27me1K36me3 - BI10038
H3K27me2K36me0 - BI10039
H3K27me2K36me1 - BI10040
H3K27me2K36me2 - BI10041
H3K27me3K36me0 - BI10043
H3K27me3K36me1 - BI10044
H3K27me3K36me2 - BI10045
H3K27me3K36me3 - BI10046
H3K27ac1K36me0 - BI10047
H3K27ac1K36me1 - BI10048
H3K27ac1K36me2 - BI10049
H3.3K27me0K36me0 - BI10051
H3K56me0 - BI10054
H3K56ac1 - BI10058
H3K56ub1 - BI10059
H3K79me0 - BI10060
H3K79me1 - BI10061
H3K79me2 - BI10062

Column cluster complete linkage by Pearson correlation

EHZ2 knockdown signature
K27me3 down
K27ac1 up
K79me1,me2 up

SUV39H1 knockdown signature
K9me2,me3 down
S10ph up

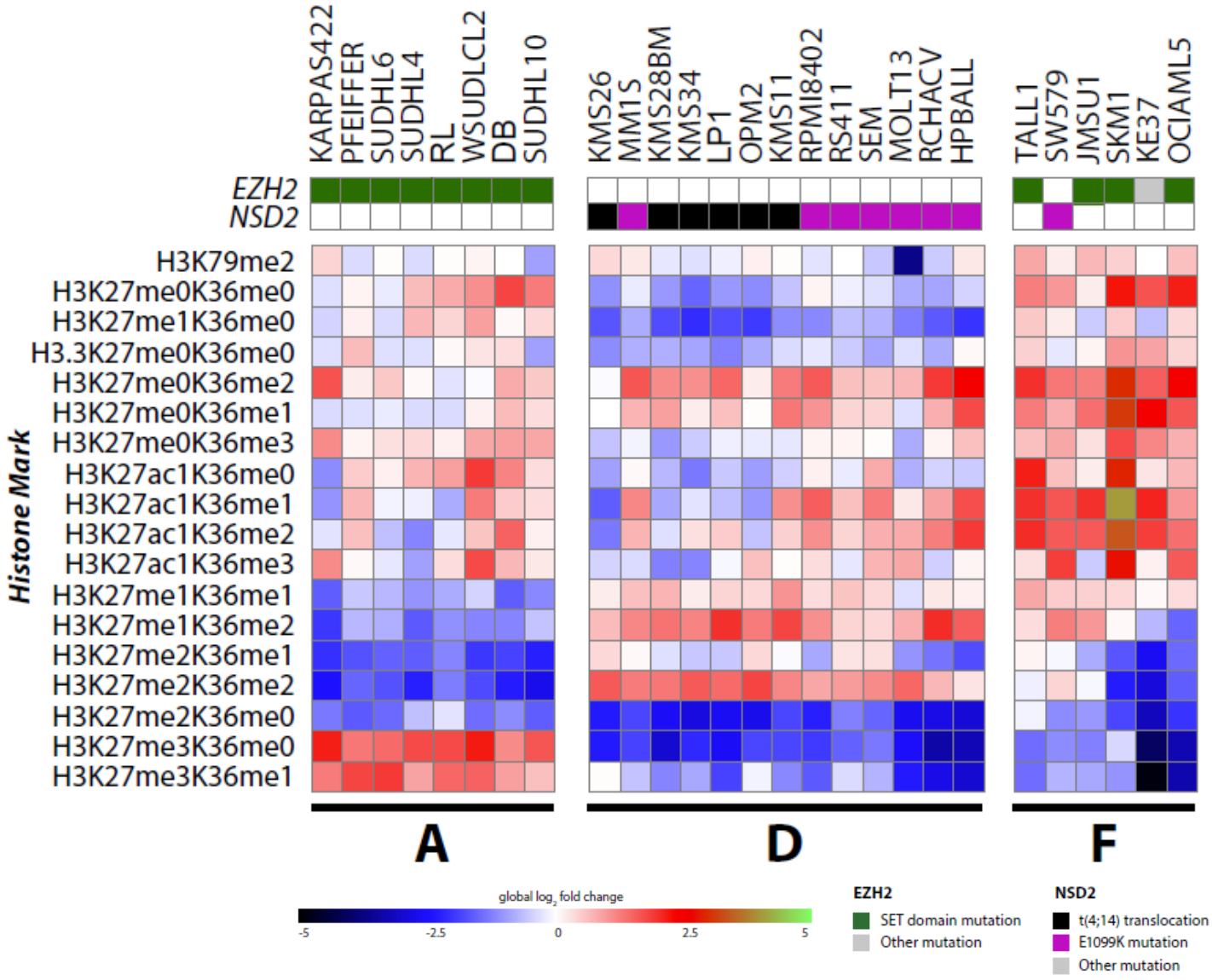
Molecular Chromatin Signatures of Genotypes



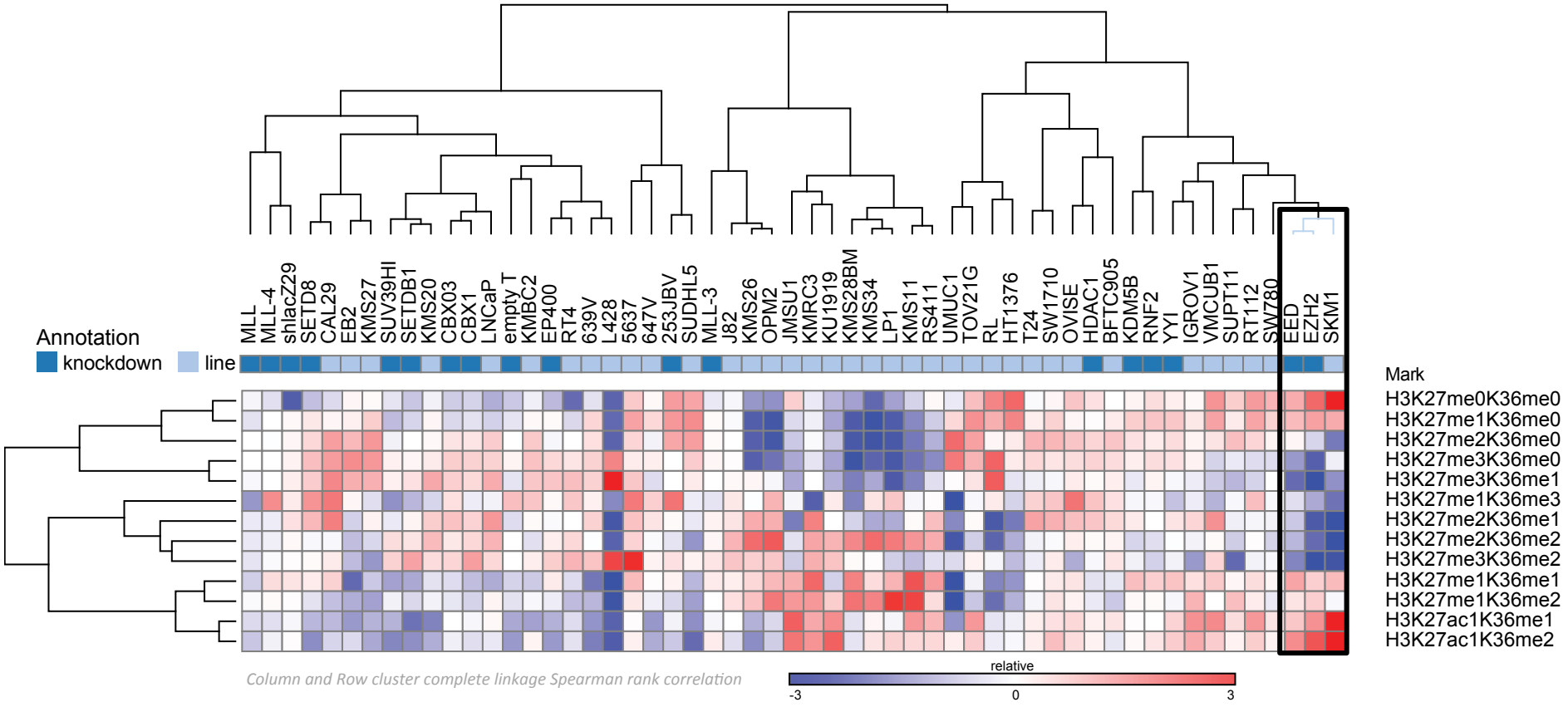
- Steady-state chromatin profiles of cell line in the Cancer Cell Line Encyclopedia

Jordan Taylor, CCLE team

Molecular Chromatin Signatures of Genotypes



Queries across data types in chromatin space



- Can co-cluster perturbation signature and mutation signature
- EZH2 mutation in SKM1 behaves like loss of function

Conclusions and Thoughts

- Proteomics must adapt to play nicely with functional genomics
- Major challenge is in generating (unhole-y) data at scale and cost
- Targeted proteomics offers a potential solution
  - Goal: universal cross-sample queries via standardization
- We are building a targeted phosphosignaling assay
- Targeted chromatin modification assay has already yielded promising results
- End-of-days for hole-y proteomics data!

Acknowledgements

- LINCS Program and Program Officers
  - U01 CA164186-01/Jaffe
- MacCoss Lab, Univ. of Washington
  - ***Mike MacCoss*** – co-PI
  - Brendan MacLean
- Broad Institute Proteomics Platform
  - Philipp Mertins
  - ***Jinal Patel***
  - Steve Carr
- Broad Institute LINCS Centers
  - Todd Golub
  - Aravind Subramanian
  - ***Xiaodong Lu***
  - ***Roger Hu***
- CCLE Project Team
  - ***Jordan Taylor*** (Broad Proteomics)
  - Frank Stegmeier (Novartis)
  - HoMan Chan (Novartis)
  - Ellen Gelfand
  - Greg Kryukov
  - Levi Garraway
- Target Accelerator Team
  - Xiaoyun Wu
  - Yashaswi Shrestha
  - Jesse Boehm