NEU Proteomics Capstone Module: Case Studies in Quantitative Proteomics

Jacob D. Jaffe, Lev Litichevski, Jarrett Egerston,
Mike MacCoss, Meena Choi, Olga Vitek

w/ special guest: Ruedi Aebersold

Goals for this course

- Understand the concepts and value of a proteomic sentinel assay
- Understand the considerations in moving from discovery proteomics data to targeted proteomics
  - Practical: Selection of probes for a targeted proteomics sentinel assay
- Understand considerations for treatment and normalization of research grade targeted proteomics assays
  - Practical: Construct your own data processing pipeline
- Introduction to next-gen MS methods to expand targeted panels
- Understand the conceptual differences between using markers and profiles of activity
- Introduction to the concept of connectivity and exploration of the dataset in this space

Key Differences from Previous Modules

- This is the first time we have attempted this course
  - Expect some bumps in the road!

- Hackathons, not exercises
  - Put into practice all you have learned so far
  - Team-based work
  - There are no right or wrong answers
  - Key learning comes from sharing of experiences

Course Structure

	Wednesday (May 10)	Thursday (May 11)	Friday (May 12)
8:00-9:00		Finish Hackathon 1	
9:00-9:30		Compare, Contrast, and Discuss Results from Hackthon 1, Compare to P100 Set	Expansion of P100 via Next-Gen MS (DIA)
9:30-10:00			
10:00-10:30			Mini Hackathon: Hack 1 probe observability
10:30-11:00		refreshments	
11:00-11:30		The Targeted P100 Assay in Practice and Considerations for Reproducible Research	The Statistician's View: What's right and what's wrong with the P100 Methodologies
11:30-12:00			
12:00-12:30		Prep for Hackathon 2 - Data intro, code review	
12:30-13:00	Registration	Lunch	
13:00-13:30			
13:30-14:00	Keynote Address: Ruedi Aebersold	Hackathon 2 - Build your own data processing pipeline	Using Sentinel Profiles for Drug Characterization: Clustering and connectivity
14:00-14:30			
14:30-15:00			
15:00-15:30	refreshments		
15:30-16:00	Intro to Sentinel Assay Concept, Development of P100 Discovery Data Set	Keynote: Mike MacCoss - Proteomics Data Ecosystems	Closing Address: Ruedi Aebersold
16:00-16:30			
16:30-17:00	Prep for Hackathon 1 - Data intro, code review		
17:00-17:30	Begin Hackathon 1 - Probe Selection	Compare, Contrast, and Discuss Results from Hackthon 2	Wrap-up
17:30-18:00			

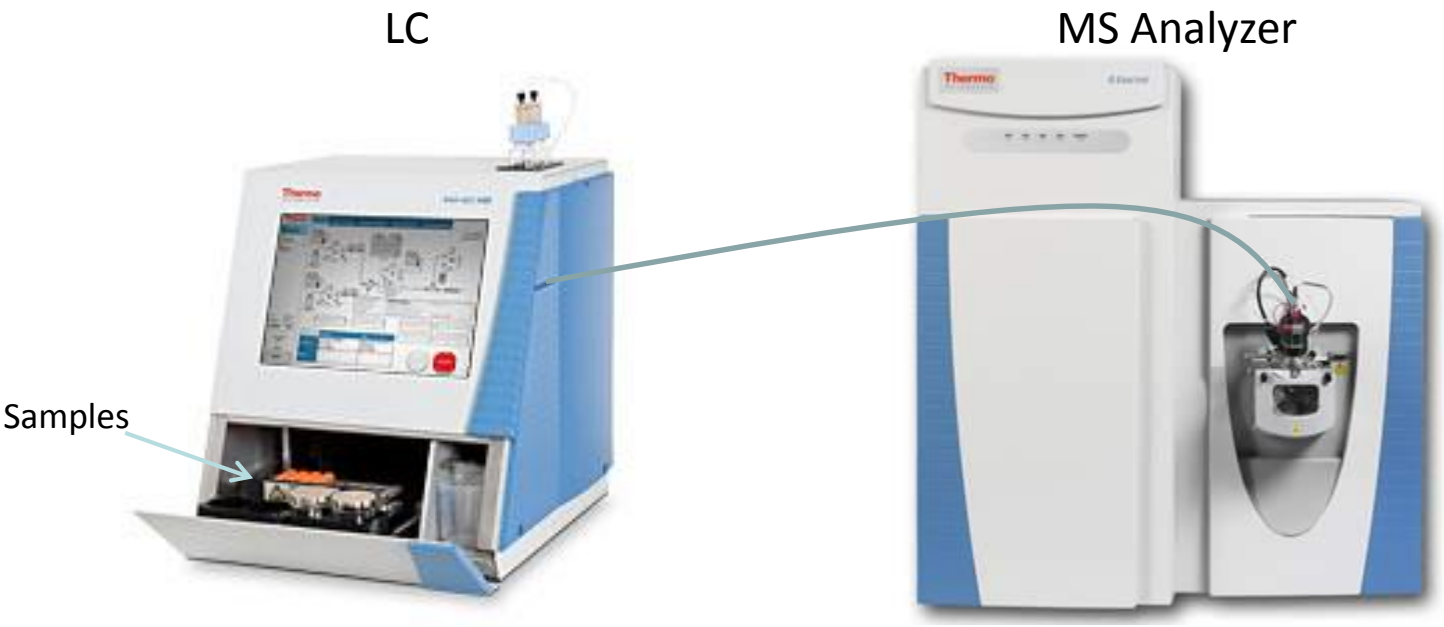
Everything you need:

bit.ly/PCCSEData

Lecture 1: Targeted Proteomics, the Sentinel Assay Concept, and Initial Development of the P100 Assay

Jacob D. Jaffe

Targeted vs. Discovery Proteomics



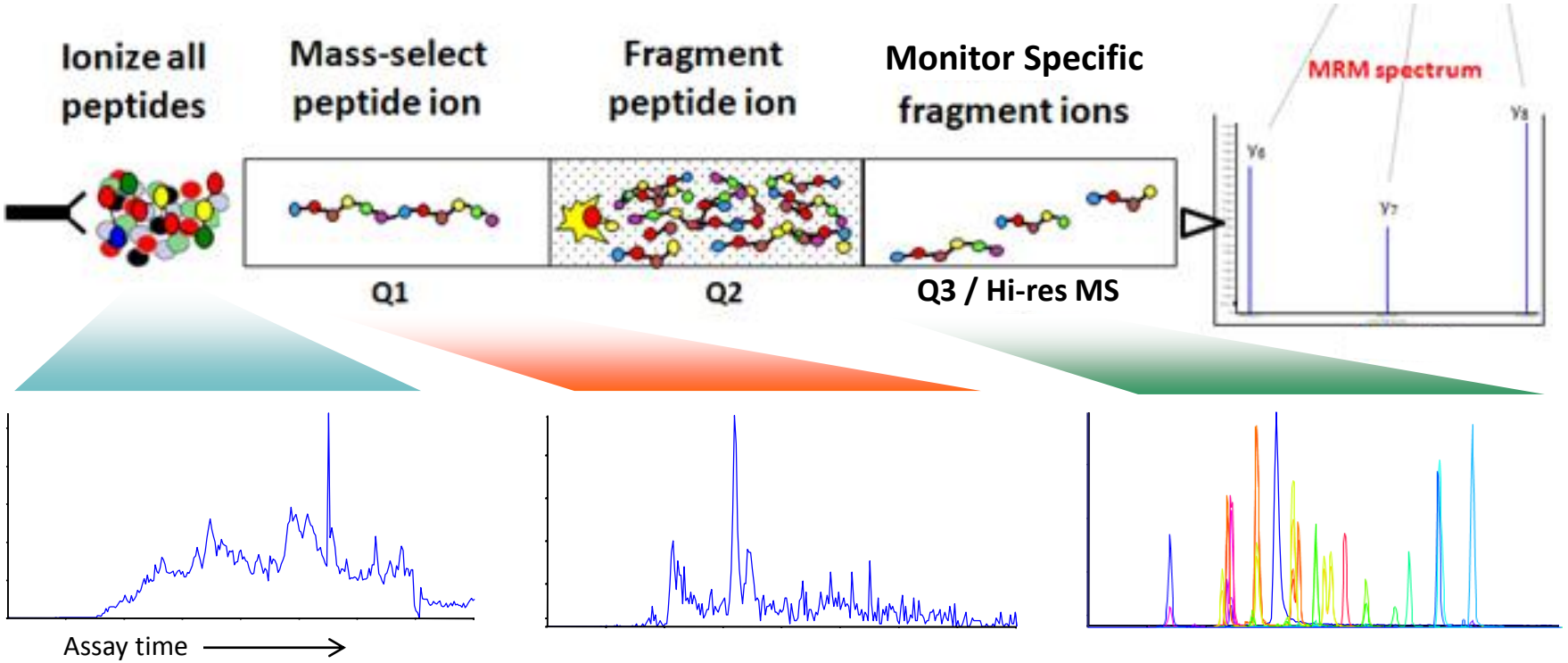
Shotgun Proteomics

- Complex sample is interrogated stochastically
- Get protein/peptide IDs
- Irreproducible repeats

Targeted Proteomics

- Focus efforts on relatively fewer analytes
- Verify ID, get quantification
- Consistency

Targeted MS: Experimental Implementation



- 100s of analytes in a single experiment
- **Synthetic peptide internal standards** – quantification and proof of ID
- ***When you want to guarantee you measure it each and every time!***

Why do targeted studies?

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

Data Shapes in Proteomics

1. Long and skinny

Peptides/proteins

	1	2
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	###
≈≈≈	###	
≈≈≈	###	
≈≈≈		###
≈≈≈		###

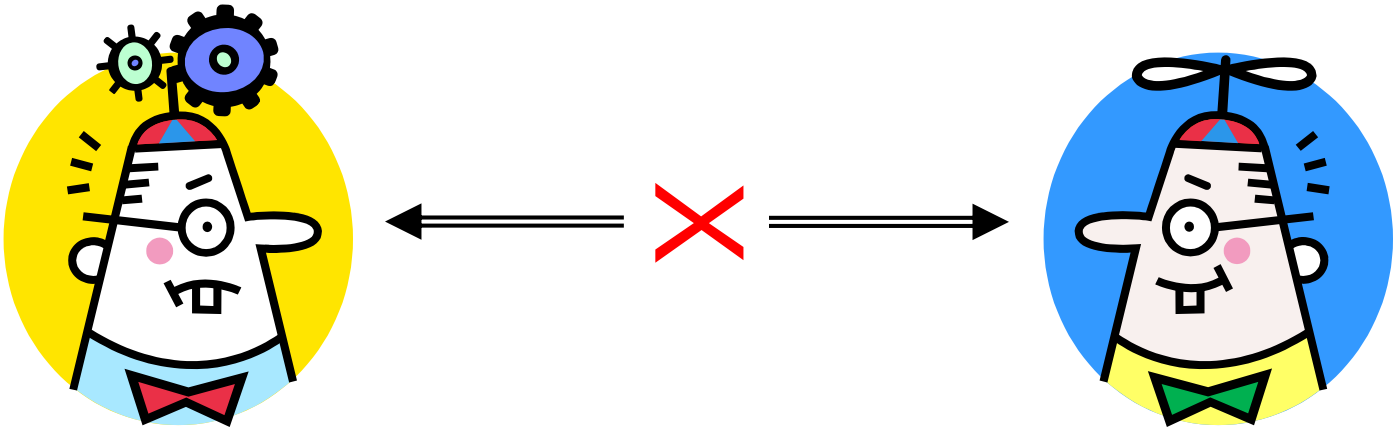
2. Short and wide

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

3. Hole-y

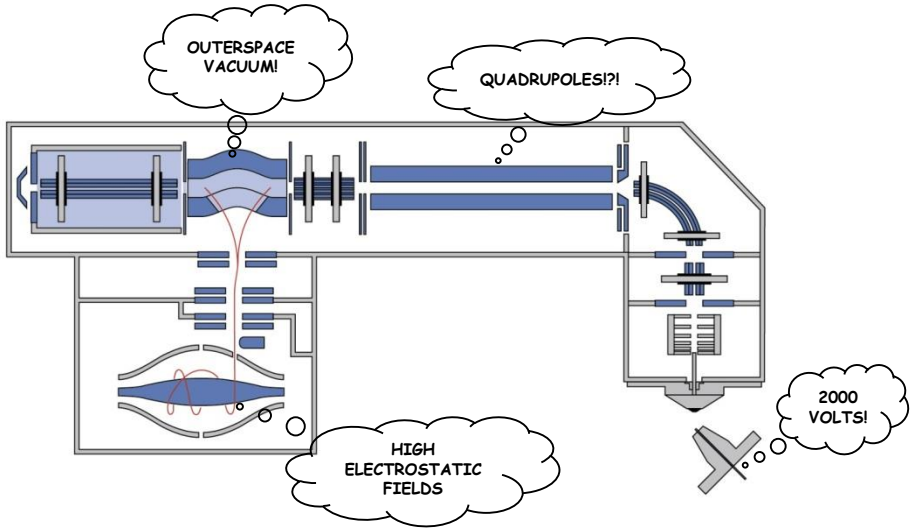
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
≈≈≈	###	###	###	###	###	###	###	###	###	###	###	###	###	###
≈≈≈	###	###	###	###	###	###	###	###	###		###	###	###	###
≈≈≈	###	###	###	###	###	###	###	###	###	###	###	###	###	###
≈≈≈	###	###	###	###	###	###	###	###	###	###	###	###	###	###
≈≈≈	###	###	###	###	###	###	###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

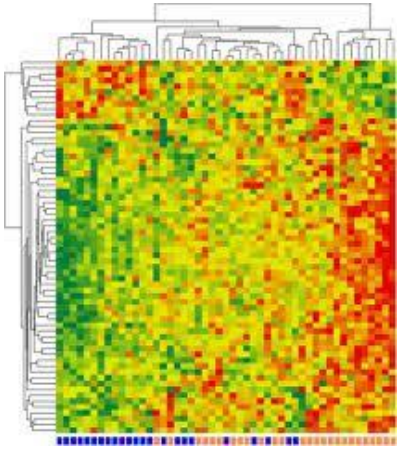
It’s hard to be hole-y



Instrument Jockeys

Data Geniuses





The Sentinel Assay Concept

A Sentinel Assay for Yeast Biology

A sentinel protein assay for simultaneously quantifying cellular processes

Martin Soste[1], Rita Hrabakova[1,2], Stefanie Wanka[3], Andre Melnik[1], Paul Boersema[1], Alessio Maiolica[4], Timon Wernas[1], Marco Tognetti[1], Christian von Mering[3] & Paola Picotti[1]

Nature Methods 11:1045-8 (2014)

ously, in high throughput. The approach we describe here is based on the concept of 'sentinel' proteins that report on the activation state of a specific biological process and are selected on the basis of literature evidence or computational prediction (**Fig. 1**). Our

- Two targeted proteomics assays:
  - Protein
  - Phosphopeptide
- Characterizes a range of biochemical responses to stimulation
- Targets nominated by literature, prior knowledge

A Sentinel Assay for Yeast Biology

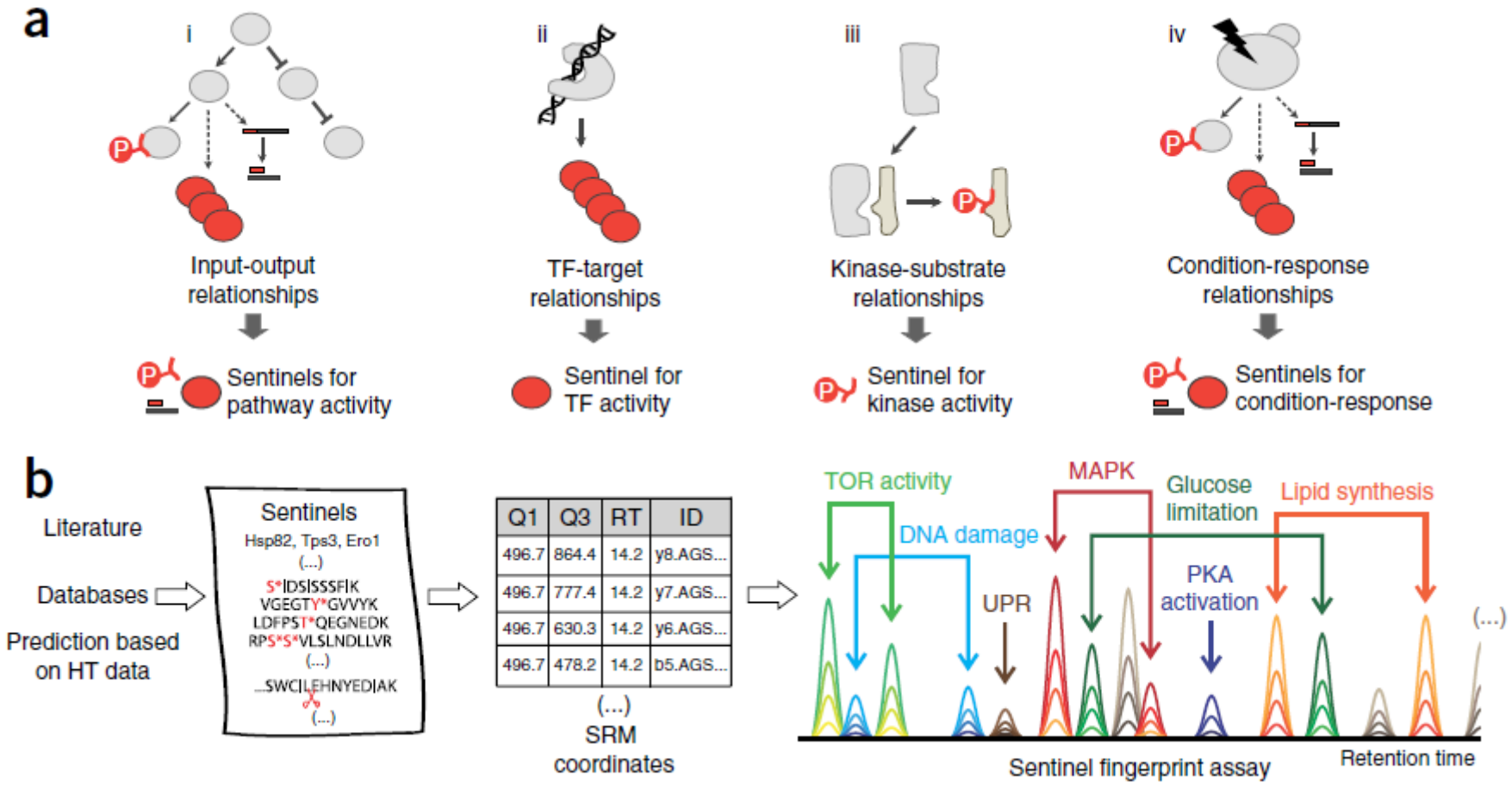
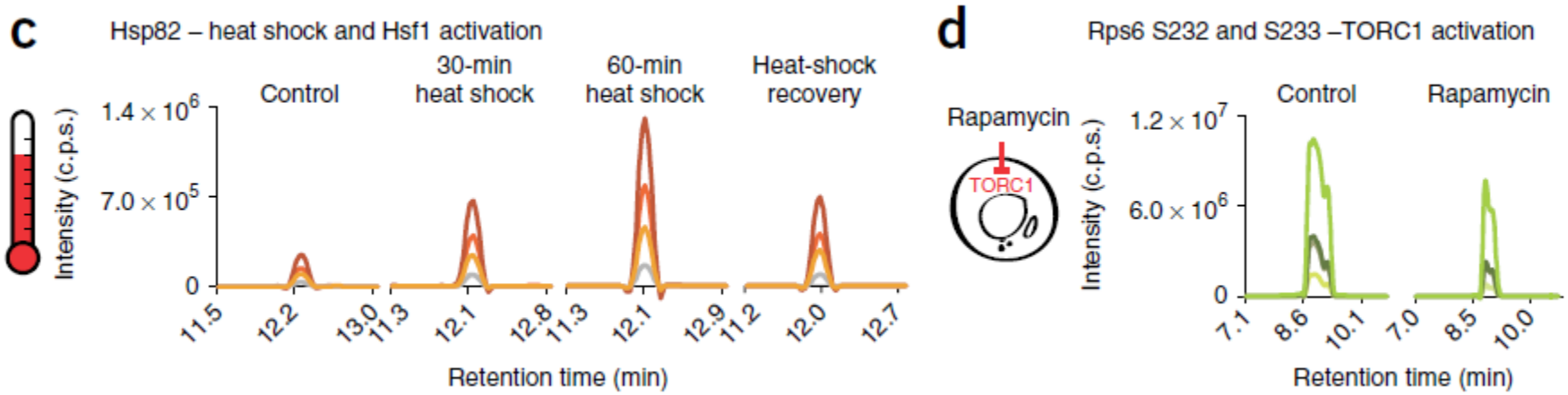


Fig. 1 from Soste et al., *Nature Methods* 11:1045-8 (2014)

Examples of Specific Sentinel Behaviors

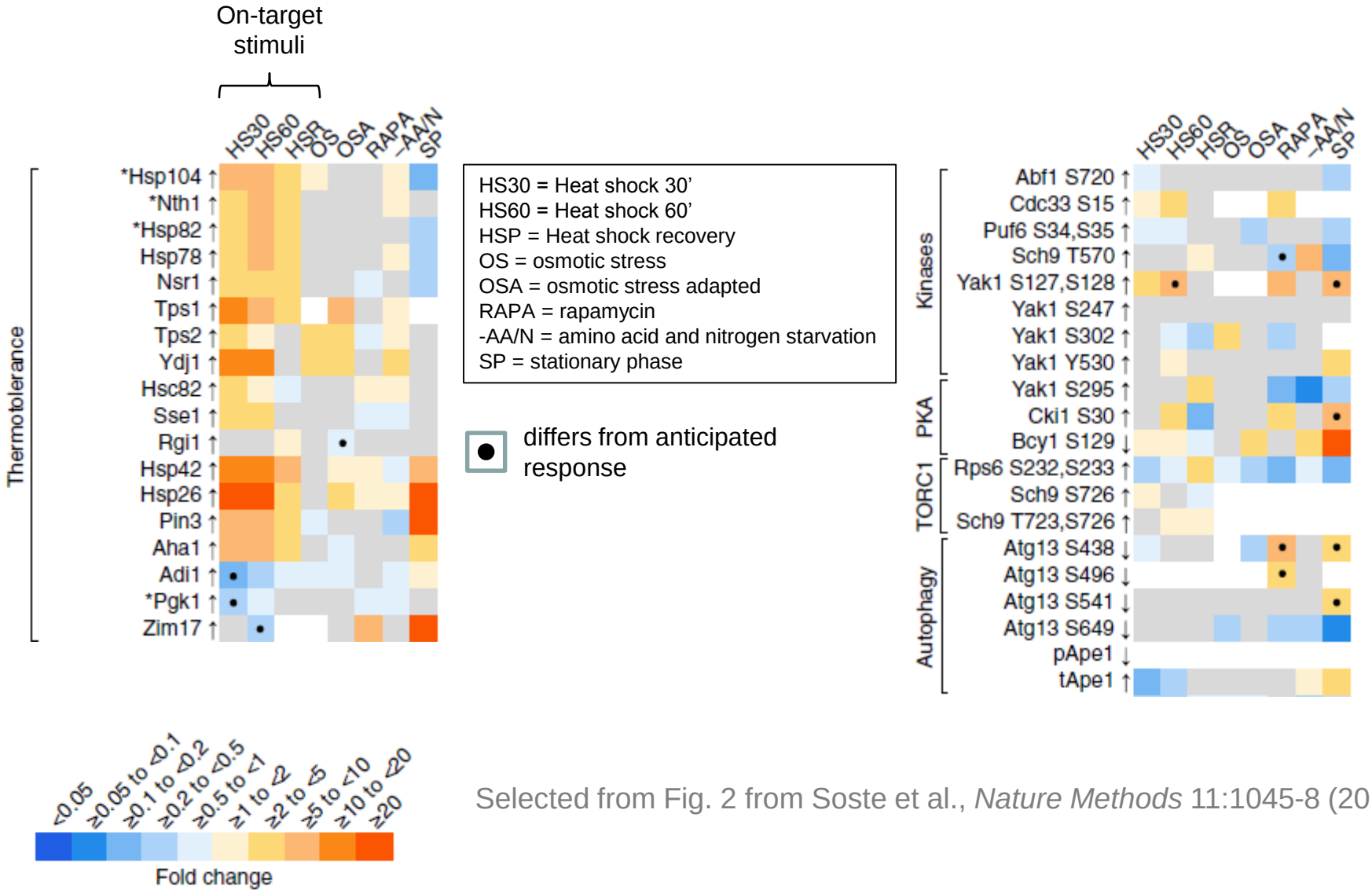


Selected from Fig. 3 from Soste et al., *Nature Methods* 11:1045-8 (2014)

Examples of Sentinel Reporting Profiles

Protein Level Reporters

Phospho Level Reporters



Selected from Fig. 2 from Soste et al., *Nature Methods* 11:1045-8 (2014)

Considerations for Translating to a Human Sentinel Assay

- Human proteome/phosphoproteome is bigger
- Pathway architecture in human cells is more complex
- Probably less direct knowledge as to what the sentinels would be
- Proteomic accessibility of desired sentinel peptides
- Diversity of biology and cell types coupled with proteome dynamic range might make life difficult

“P100” – A different kind of sentinel assay

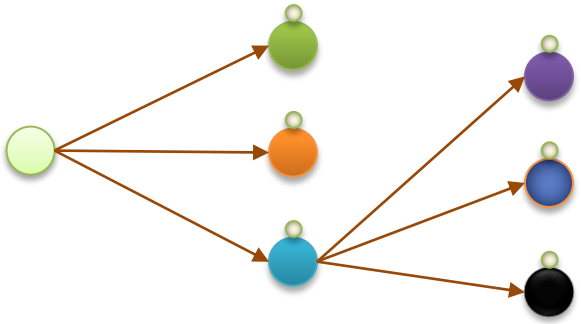
Reduced Representation Phosphosignaling

Why phosphoproteomics?

- Classic “pathways” are commonly represented as signaling cascades of phosphoproteomic events
- Phosphoproteomic events are the “first responders” of cells and occur on a range of time scales from minutes to days
- Phosphoproteins are not easily quantified using nucleic acid based techniques
- Kinase inhibitors are a major class of emerging therapeutics

Reduced Representation Signaling Set: 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 all the time?
  - Amenable to reductionist approaches?

Final goals of P100 assay development

- Drug responsiveness
- Universal (-ish)
- Ser/Thr focused
- Compact
- Operates at scale
- Longitudinally comparable

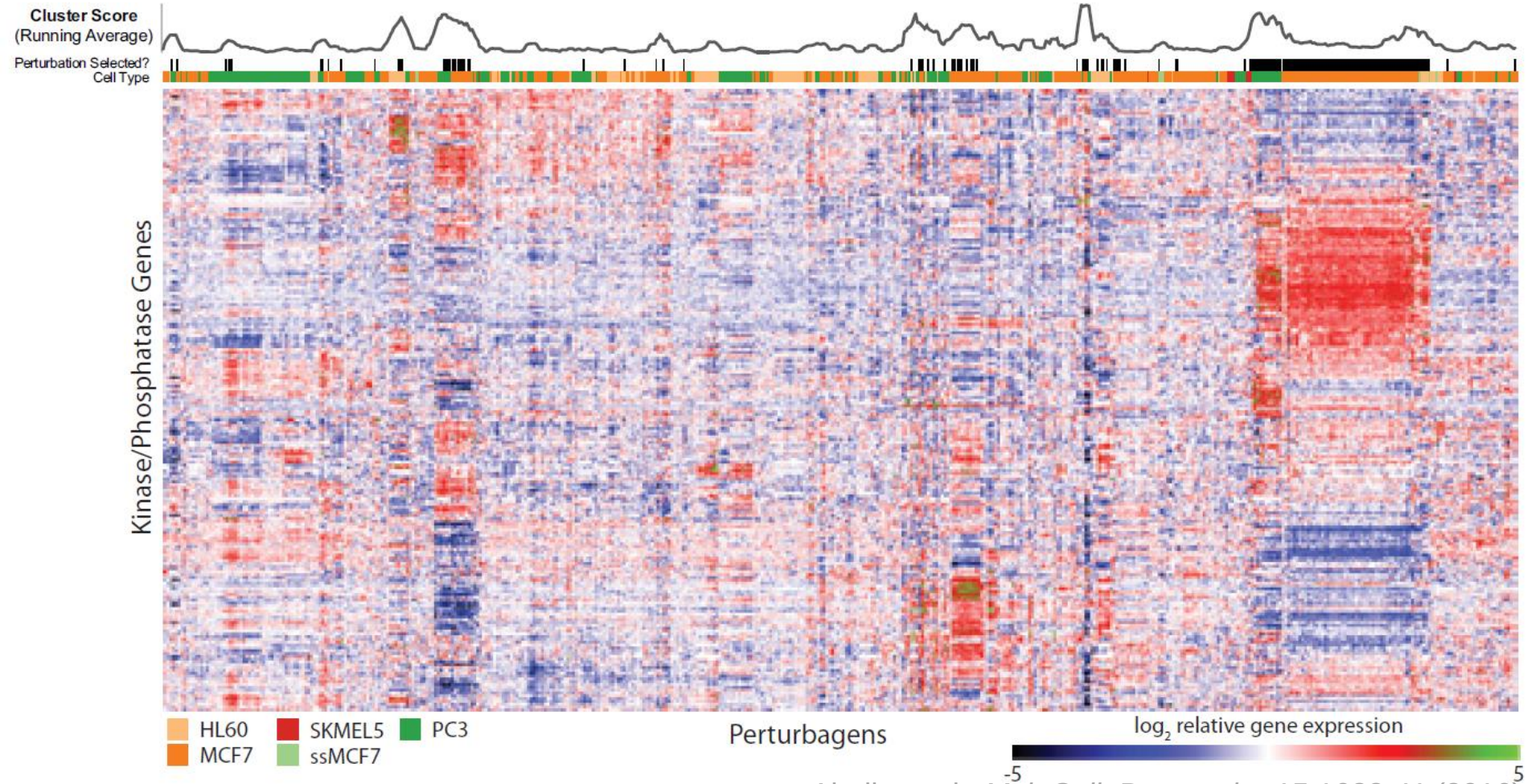
Major Problem #1: No easily usable prior data

- Many piecemeal, *ad hoc* drug studies
  - 1 or 2 drugs
  - Many focus on pTyr signaling
  - Wide range of biological models
  - Inconsistent data reporting formats
  - General lack of systematic studies

- We decided to make our own
  - Discovery proteomics
  - Quantitation by SILAC

Selection of Compounds for Systematic Data Generation

- Start with gene expression profiles induced by drug treatment
  - Multiple cell types
  - Extract kinase/phosphatase genes



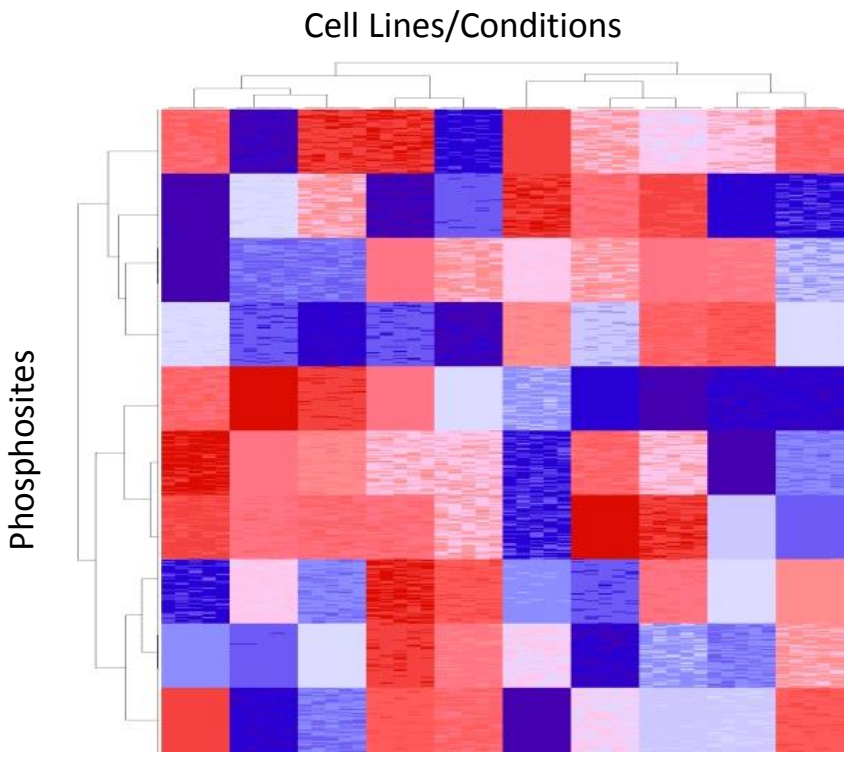
Abelin et al., *Mol. Cell. Proteomics* 15:1622-41 (2016)

Generation of foundational data

2 x MCF7 (breast cancer) / PC3 (prostate cancer) / HL60 (acute promyelocytic leukemia) 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

Select representative phosphopeptide probes from coherent clusters

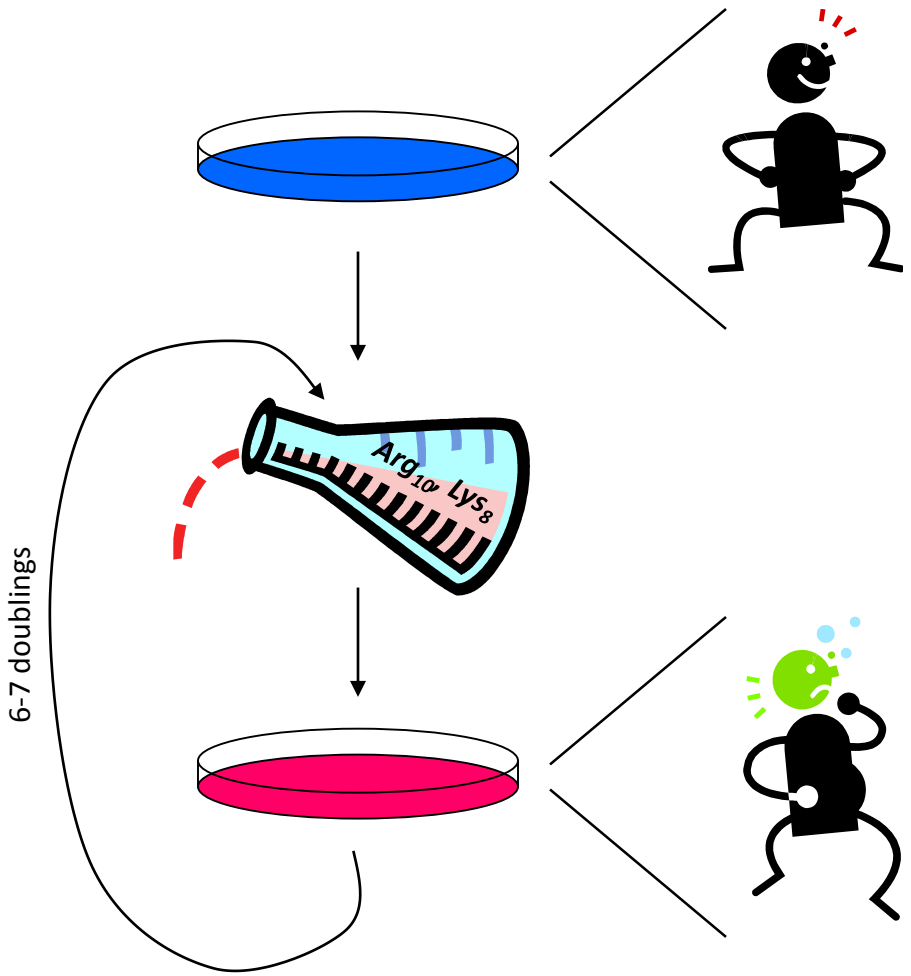


Chosen based on:

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

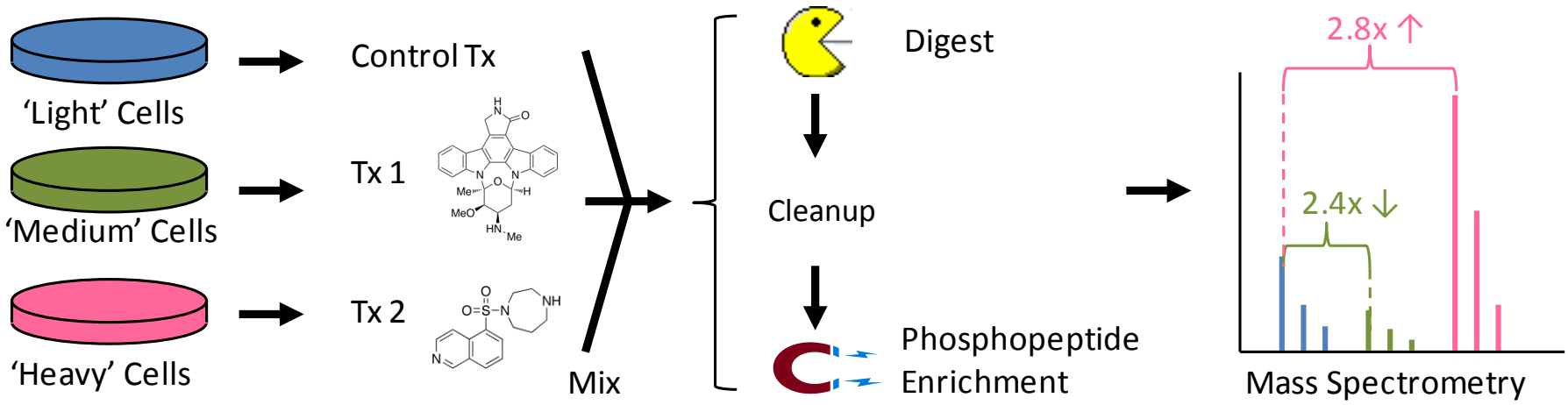
Jinal Patel, Xiaodong Lu

SILAC and Stable Isotopes: Let cells do the work



- Feed cells growth medium with “heavy” amino acids
- Proteins incorporate these amino acids
- All proteins are slightly “heavier” but otherwise biochemically equivalent
- Choice of Lys and Arg typical because of trypsin

SILAC-based discovery profiling

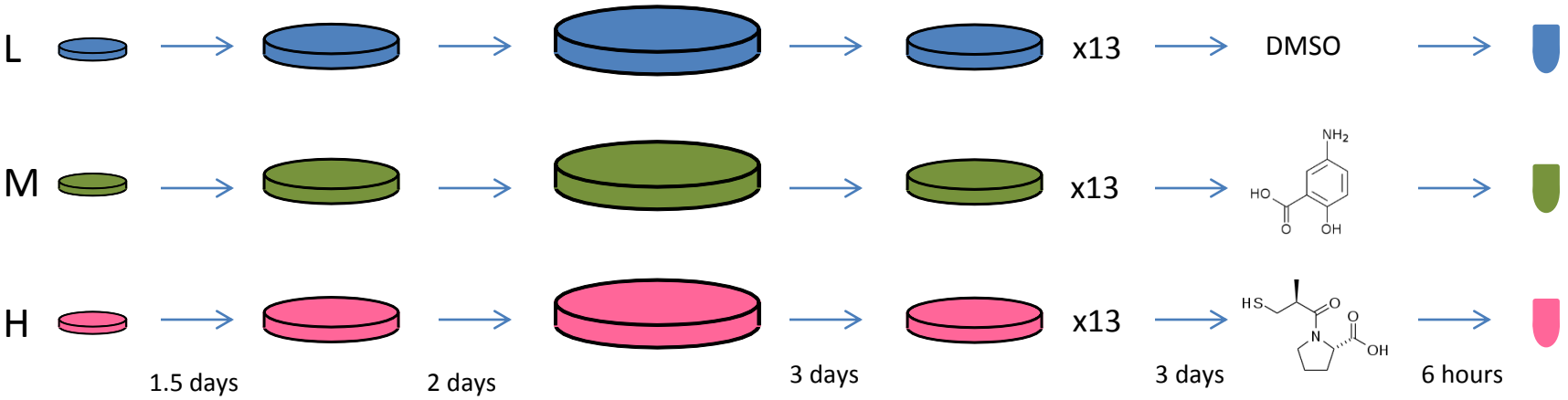


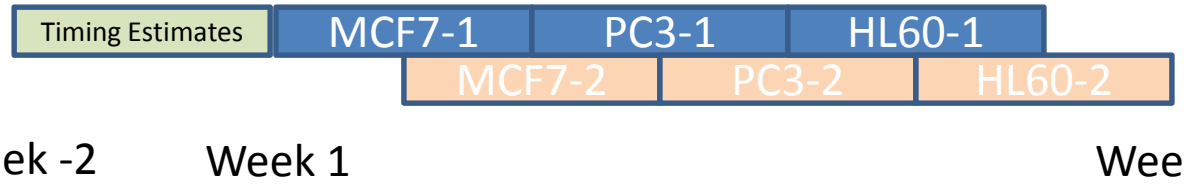
- Cells pre-labeled by SILAC
- Treatments: 6 hours, concentrations from gene expression study
- IMAC-based phosphopeptide enrichment
  - Primary yield: Ser/Thr
- Mass spec provides both ID (MS2) and quantification (MS1)

Considerations for successful assay configuration

- Are there a set of phosphosites that we can (almost) always observe?
- Do these phosphosites behave in interesting ways in response to perturbations?
- Do we gain anything over gene expression measurements?
- Can we pick a (small) set of highly informative phosphopeptides?
- Could profiling these peptides reproduce the observations of a deeper experiment?
- Is the assay generally useful?

Experimental Implementation



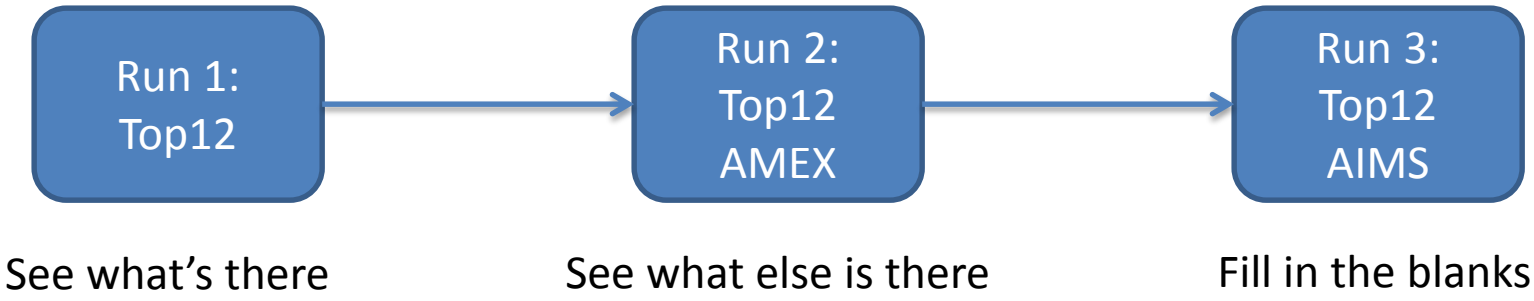


More than 7.5 x 10^8 cells grown for experiments!
True bio-replicates

Xiaodong Lu, Jinal Patel

Computational Fractionation Acquisition Strategy

- All experiments performed on a Q-Exactive mass spectrometer
  - 2 hour, “Top 12” shotgun proteomics experiments
  - Each sample analyzed 3 times:

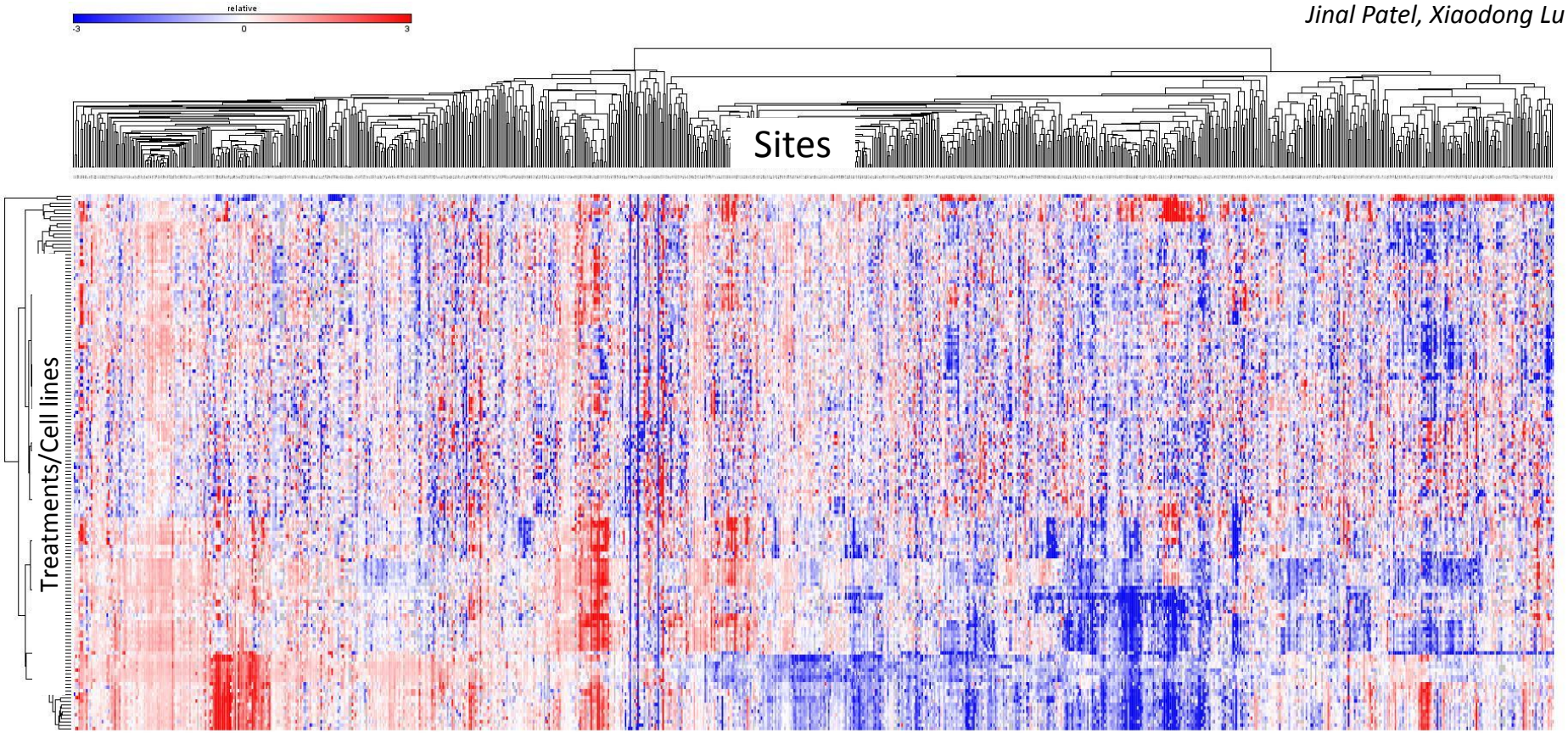


- Primary interpretation and quantification in MaxQuant

Rudomin et al., *J. Proteome Res* 8:3145-60 (2009)
Jaffe et al., *Mol. Cell. Proteomics* 7:1952-62 (2008)

P100 Source Dataset

Jinal Patel, Xiaodong Lu



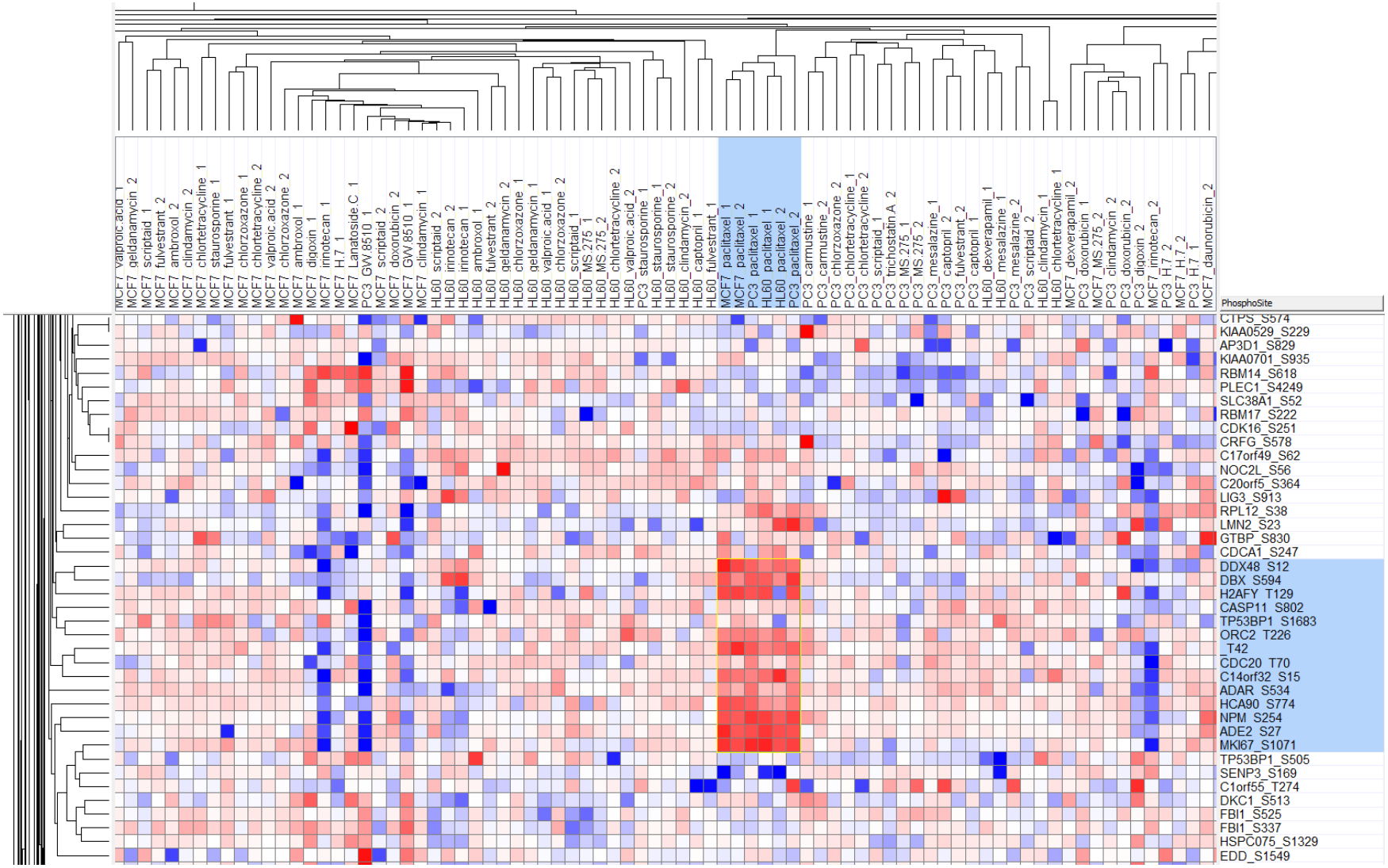
- Large systematic set of perturbations with phosphoproteomics readout
  - Over 10,000 phosphosites observed
- Over 1,200 sites present in >75% of all experiments

Will it work?

- ✓ **Are there a set of phosphosites that we can (almost) always observe?**

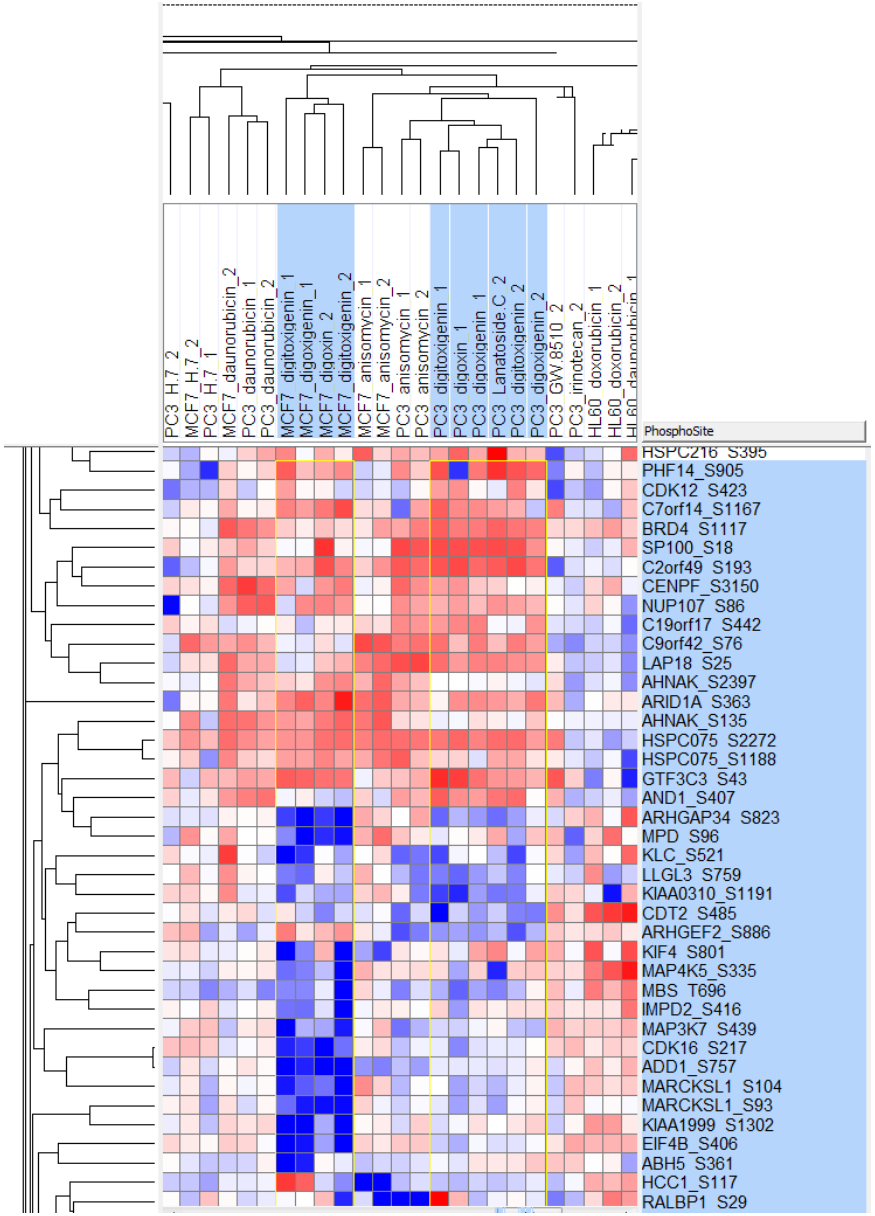
- Do these phosphosites behave in interesting ways in response to perturbations?
- Do we gain anything over gene expression measurements?
- Can we pick a (small) set of highly informative phosphopeptides?
- Could profiling these peptides reproduce the observations of a deeper experiment?
- Is the assay generally useful?

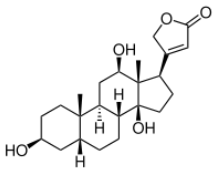
Signals span lineages



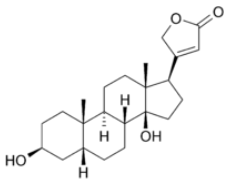
- All 6 Taxol treatments cluster – regardless of lineage

Signals group related compounds and replicates

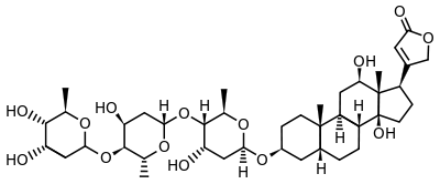




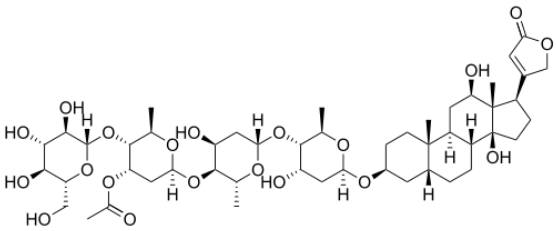
Digoxigenin



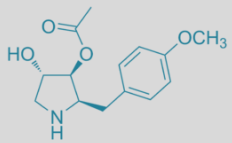
Digitoxigenin



Digoxin



Lanatoside C

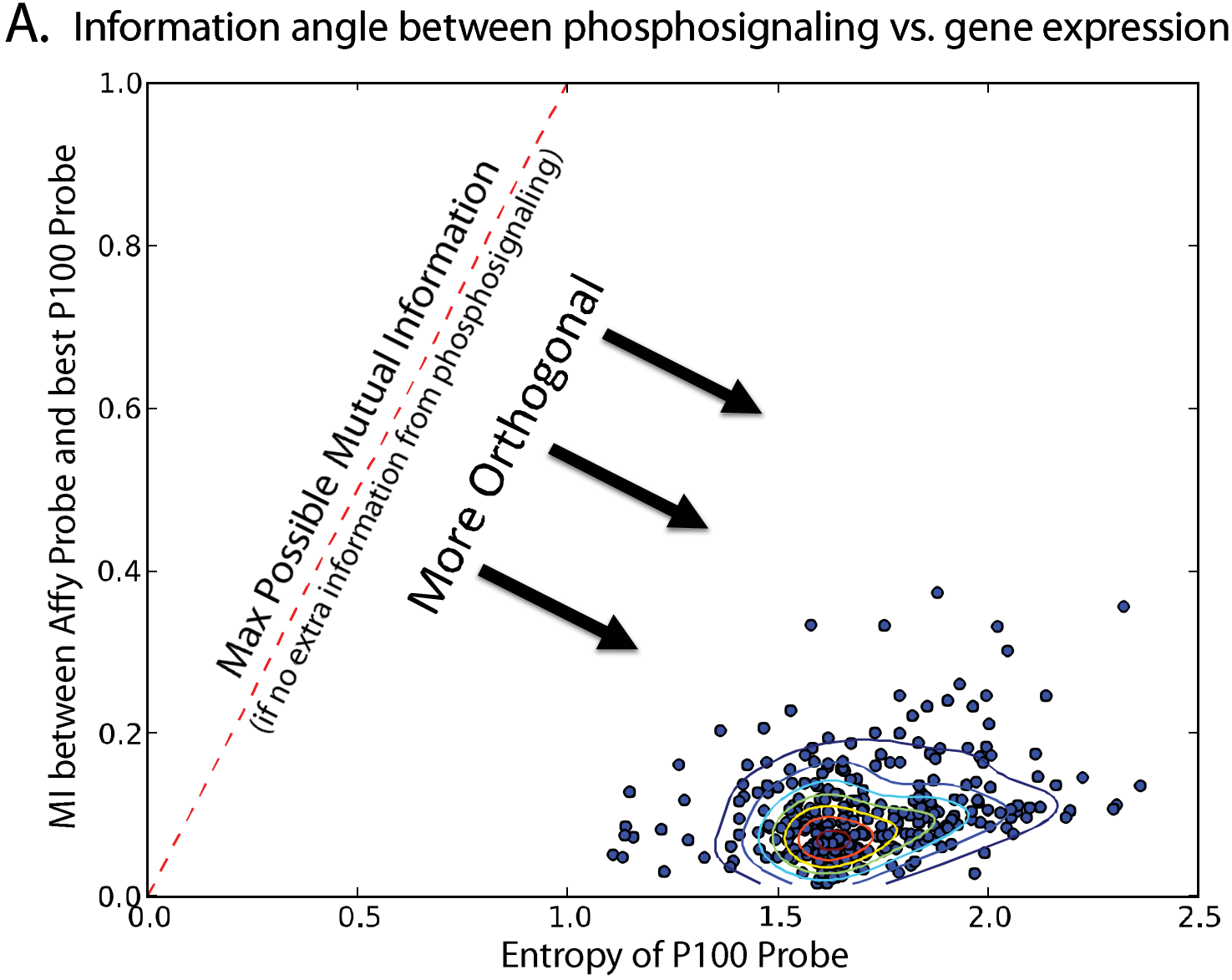


Anisomycin

Will it work?

- ✓ **Are there a set of phosphosites that we can (almost) always observe?**
- ✓ **Do these phosphosites behave in interesting ways in response to perturbations?**
- Do we gain anything over gene expression measurements?
- Can we pick a (small) set of highly informative phosphopeptides?
- Could profiling these peptides reproduce the observations of a deeper experiment?
- Is the assay generally useful?

Phospho-state adds value over gene expression

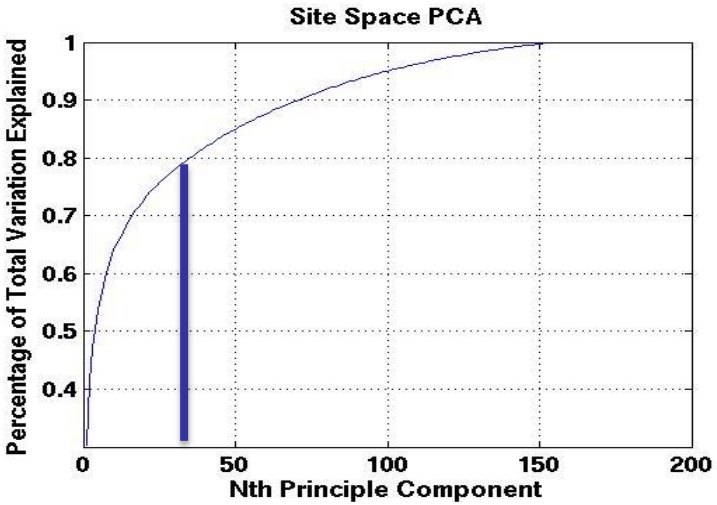


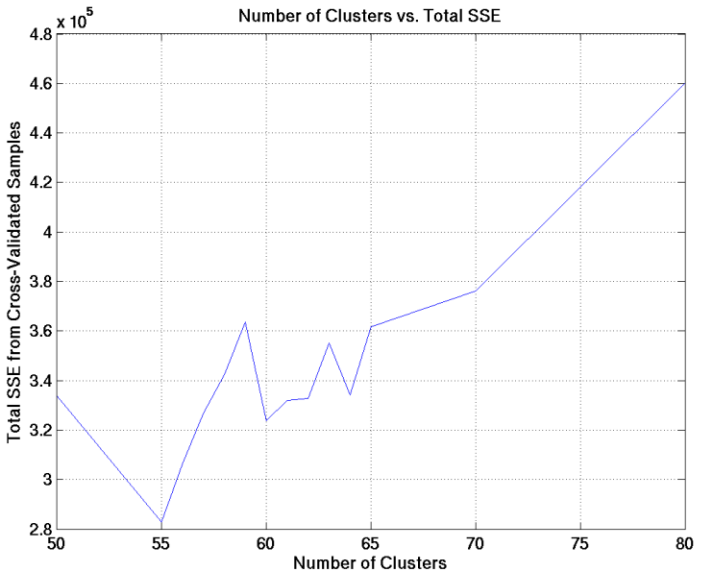
Will it work?

- ✓ Are there a set of phosphosites that we can (almost) always observe?
- ✓ Do these phosphosites behave in interesting ways in response to perturbations?
- ✓ Do we gain anything over gene expression measurements?
- Can we pick a (small) set of highly informative phosphopeptides?
- Could profiling these peptides reproduce the observations of a deeper experiment?
- Is the assay generally useful?

Determining the Right Number of Groups to Monitor

- Each group should contain phosphopeptides with similar patterns
- PCA and HC methods suggest ~50 clusters are optimal
- Further refinement by hierarchical clustering leads to 55 clusters
- We can choose 2 probes per cluster to build redundancy into assay

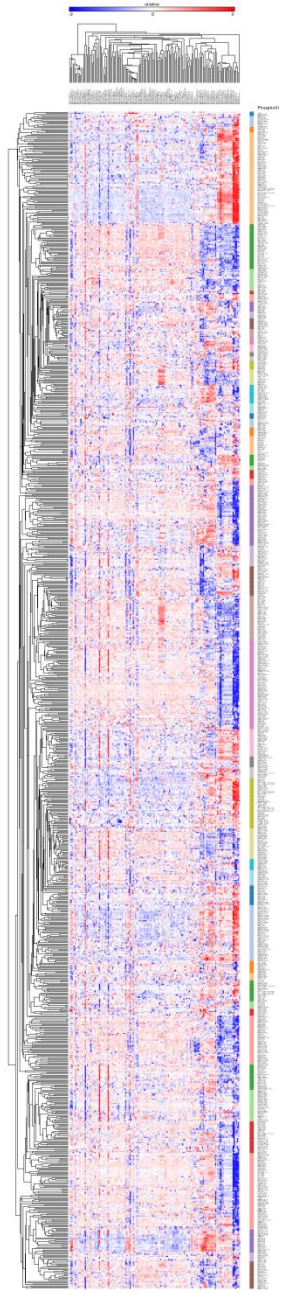




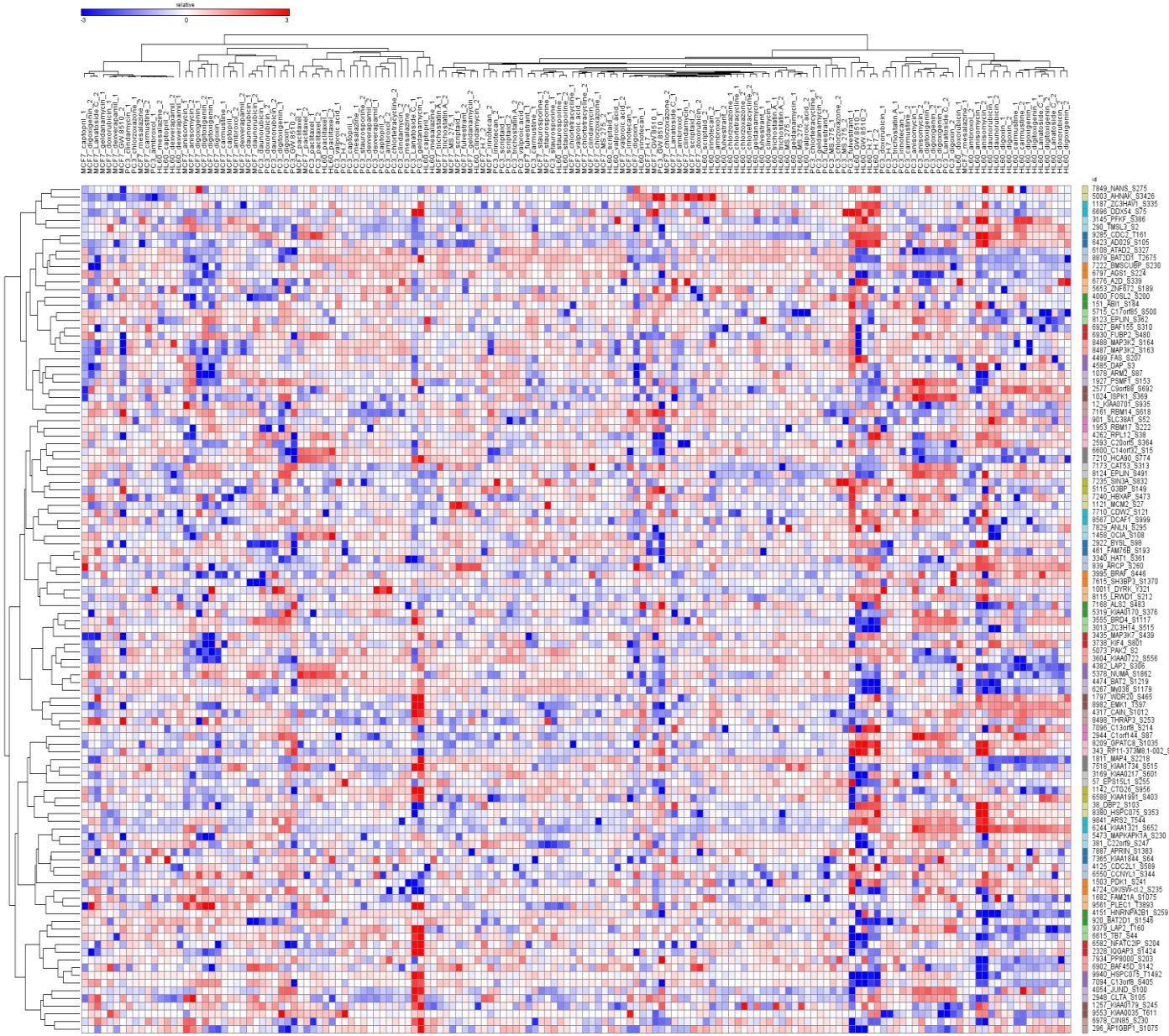
Principles for picking group representatives

1. # sites (1 is best)
2. Missed cuts (0 is best)
3. C and M count (0 is best)
4. Observability (156 is best)
5. Localization (1 is best)
6. Length (shorter is better)

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

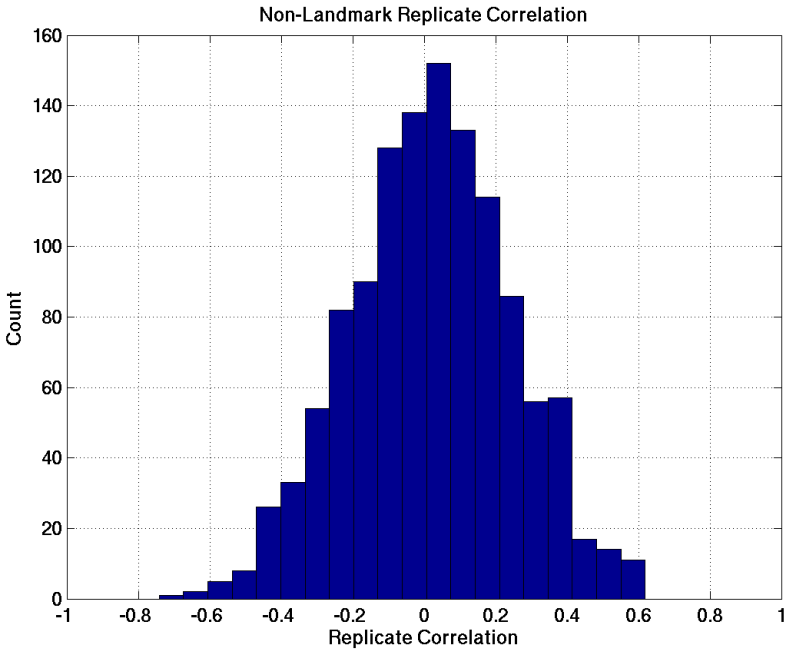


Roger Hu

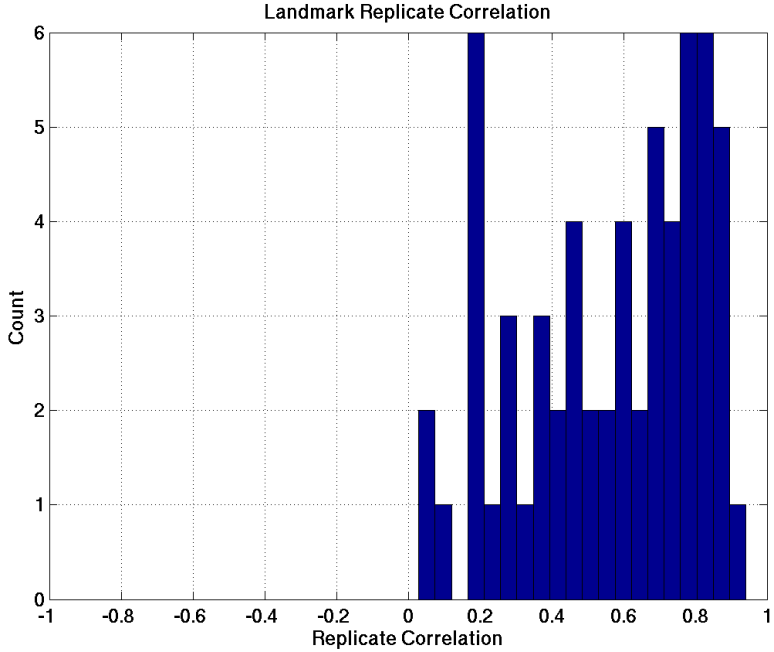


Probeset recalls replicates significantly better than random

Random Probes



P100 probes



- Looks like it is going to work.

Will it work?

- ✓ Are there a set of phosphosites that we can (almost) always observe?
- ✓ Do these phosphosites behave in interesting ways in response to perturbations?
- ✓ Do we gain anything over gene expression measurements?
- ✓ Can we pick a (small) set of highly informative phosphopeptides?
- Could profiling these peptides reproduce the observations of a deeper experiment?
- Is the assay generally useful?

To be continued…

- But first, a hackathon