Library of Integrated Network-based Cellular Signatures

Proteomic Characterization Center for Signaling and Epigenetics

Scientific Advisory Board Meeting
&
Mid-Course Review

December 9th, 2016

Jacob D. Jaffe – Broad Institute, PI

Li-Huei Tsai – MIT, co-investigator

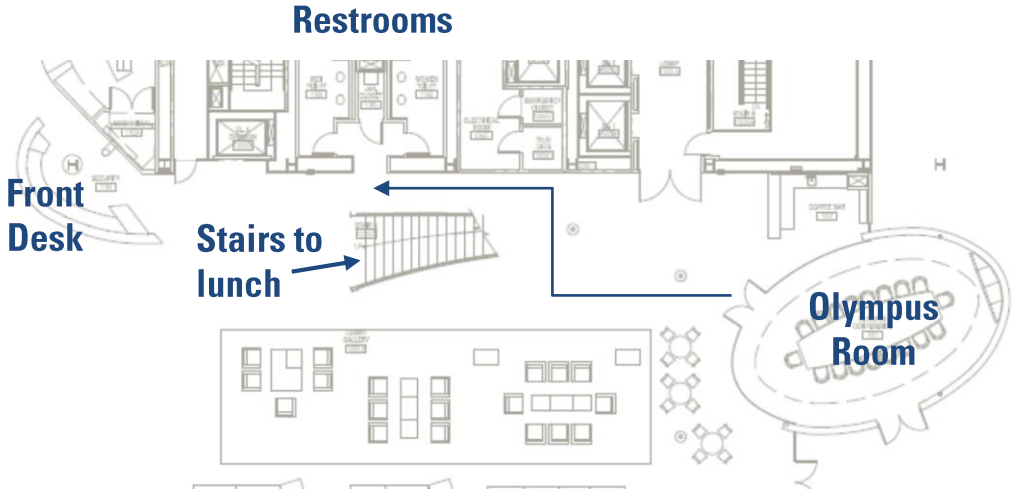
Michael MacCoss – U. of Washington, co-investigator

Welcome!

Agenda & Logistics

8:30 – 9:00 AM	Introductions and Overviews
	Welcome – Jacob Jaffe, Broad Institute
	Intro to Mid-Course Review Goals – Ajay Pillai, NIH
	Center Overview – Jacob Jaffe, Broad Institute; Li-Huei Tsai, MIT; Michael MacCoss, Univ. of Washington
9:00 – 10:20 AM	Session I: Data Generation & Analysis
	Data Generation Overview (Malvina Papanastasiou, Broad Institute)
	Neural Lineages and Models (Jennie Young, MIT)
	Analysis (Lev Litichevskiy & Ryan Peckner, Broad Institute)
	DIA (Michael MacCoss, Univ. of Washington; Sebastian Vaca, Broad Institute)
	Break – Coffee served
10:30 – 10:50 AM	Session II: Data Release & Access
	Traceability of Experimental Workflows and Metadata (Adam Officer, Broad Institute)
	Public Portal Access (Dave Lahr, Broad Institute)
10:50 – 11:40 AM	Session III: Community Inreach and Outreach
	LINCS Consortium Common Project (Desiree Davison & Shawn Egri, Broad Institute)
	MEMA Arrays (Fatema Abdurrob, MIT)
	Cardiovascular Models (Srila Gopal & Iris Jaffe, Tufts Medical Center)
	Break – Coffee served
11:40 – 2:00 PM	Session IV: Roundtable Discussions
	Framing Questions for Future Directions – Jacob Jaffe, Broad Institute
12:00 PM	Lunch served (Board Room, 1001M)
12:15 PM	SAB Executive Session and Drafting of Report
1:30 PM	SAB Feedback and Concluding Remarks
2:00 PM	Optional Lab Tours and Further Engagement with LINCS PCCSE Members

- Wifi: **Broad**
  - (no password)



- Teleconference address:
  - https://broadinstitute.zoom.us/j/6177147638

- Logistical assistance:
  - Amy Galaviz
    - agalaviz@broadinstitute.org
    - +1-617-714-7631

Our team will highlight our progress…


Fatema Abdurrob


Mukta Bagul


Joel Blanchard


Amanda Creech


Desiree Davison


Katherine DeRuff


Jarrett Egertson


Shawn Egri


Todd Golub


Jake Jaffe

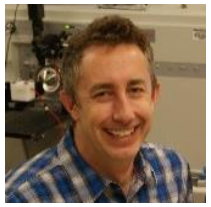

Tak Ko


Dave Lahr


Dan Lam


Lev Litichevskiy


Xiaodong Lu


Mike MacCoss


Brendan MacLean


Ted Natoli


Adam Officer


Ryan Peckner


Malvina Papanastasiou


Brian Searle


Vagisha Sharma

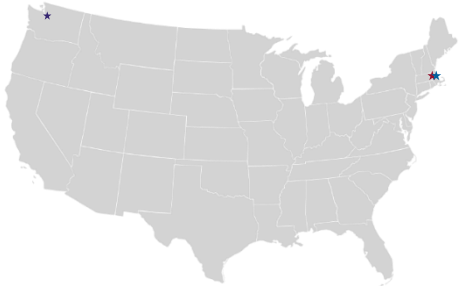

Aravind Subramanian


Sonia Ting


Li-Huei Tsai


Sebastian Vaca


Jennie Young



BROAD INSTITUTE

Massachusetts Institute of Technology

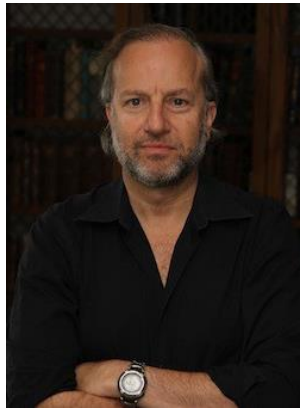
UNIVERSITY of WASHINGTON

Some key outreach collaborators…

Srila Gopal
Tufts Medical Center

Iris Jaffe
Tufts Medical Center

Stephen Haggarty
Mass. Gen. Hospital



Isaac Kohane
Harvard Medical School

Roy Perlis
Mass. Gen. Hospital

Joe Gray
OHSU / LINCS

Laura Heiser
OHSU / LINCS

THE MID-COURSE REVIEW

PURPOSE AND PROCESS

Mid-course review criteria

1. Feedback on Data Generation & Analysis
2. Feedback on Data Release and Accessibility
3. Feedback on Community Outreach
4. Feedback on major adaptations going forward
5. Other comments/advice

LINCS PCCSE

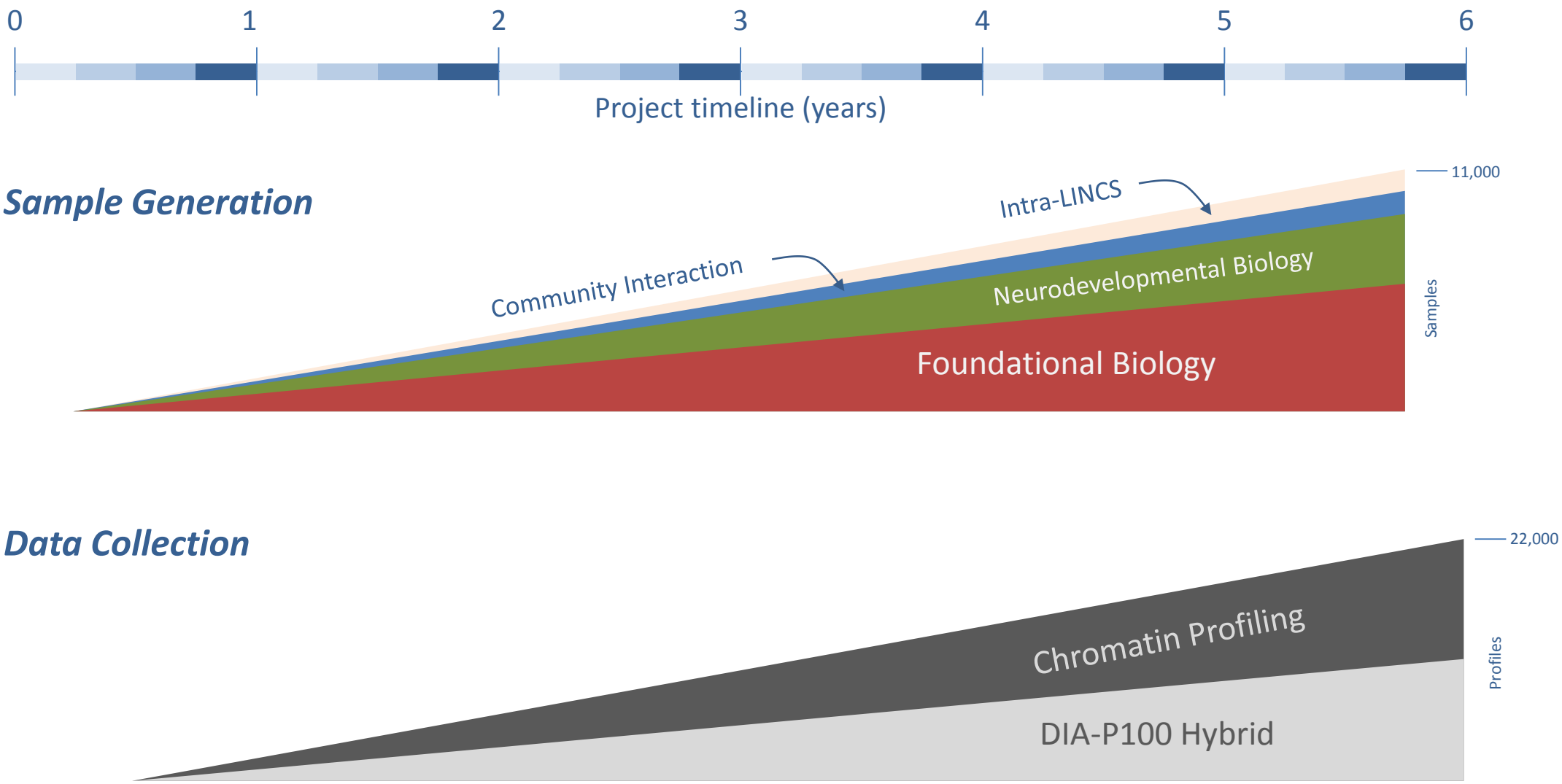
01 - CENTER OVERVIEW

Mission and methods of the LINCS PCCSE:

The LINCS Proteomic Characterization Center for Signaling and Epigenetics (PCCSE) is dedicated to understanding the therapeutic potential of drugs and unlocking the fundamental mechanisms of action of drugs, genes, and diseases through the lens of their cellular effects on phosphosignaling and epigenetic processes.

- Chemical and genetic perturbations
- Multiple biological models
- Distinct proteomic profiling assays
- Next-generation mass spectrometry
- Extensive collaboration with partners

LINCS is an audacious program; PCCSE has audacious goals



Assembly of the proper team is critical to impact

Massachusetts Institute of Technology

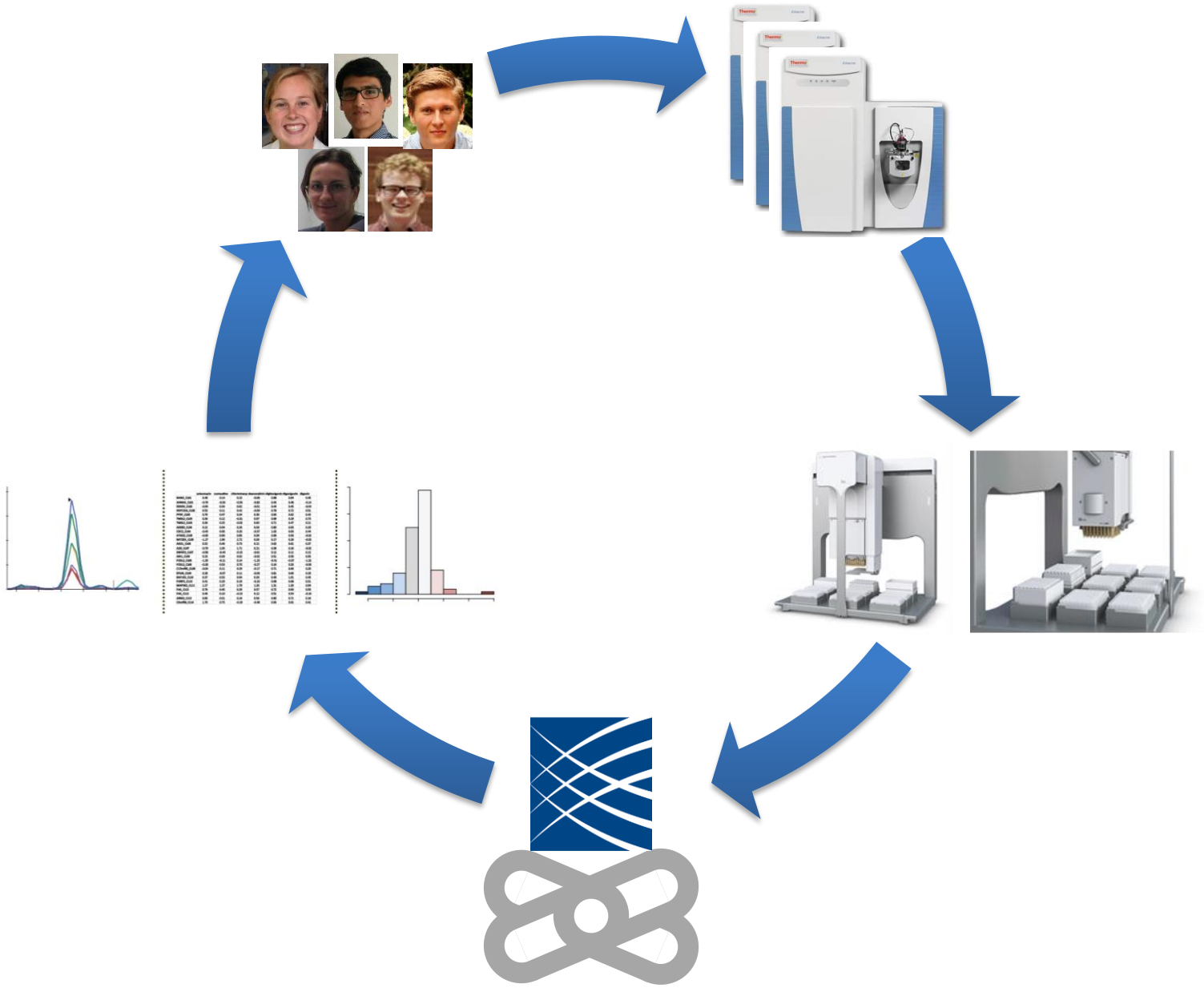
- Neurobiology expertise
- Stem-cell and differentiation experience
- Molecular and cellular biology techniques for neuronal lineages

UNIVERSITY *of* WASHINGTON

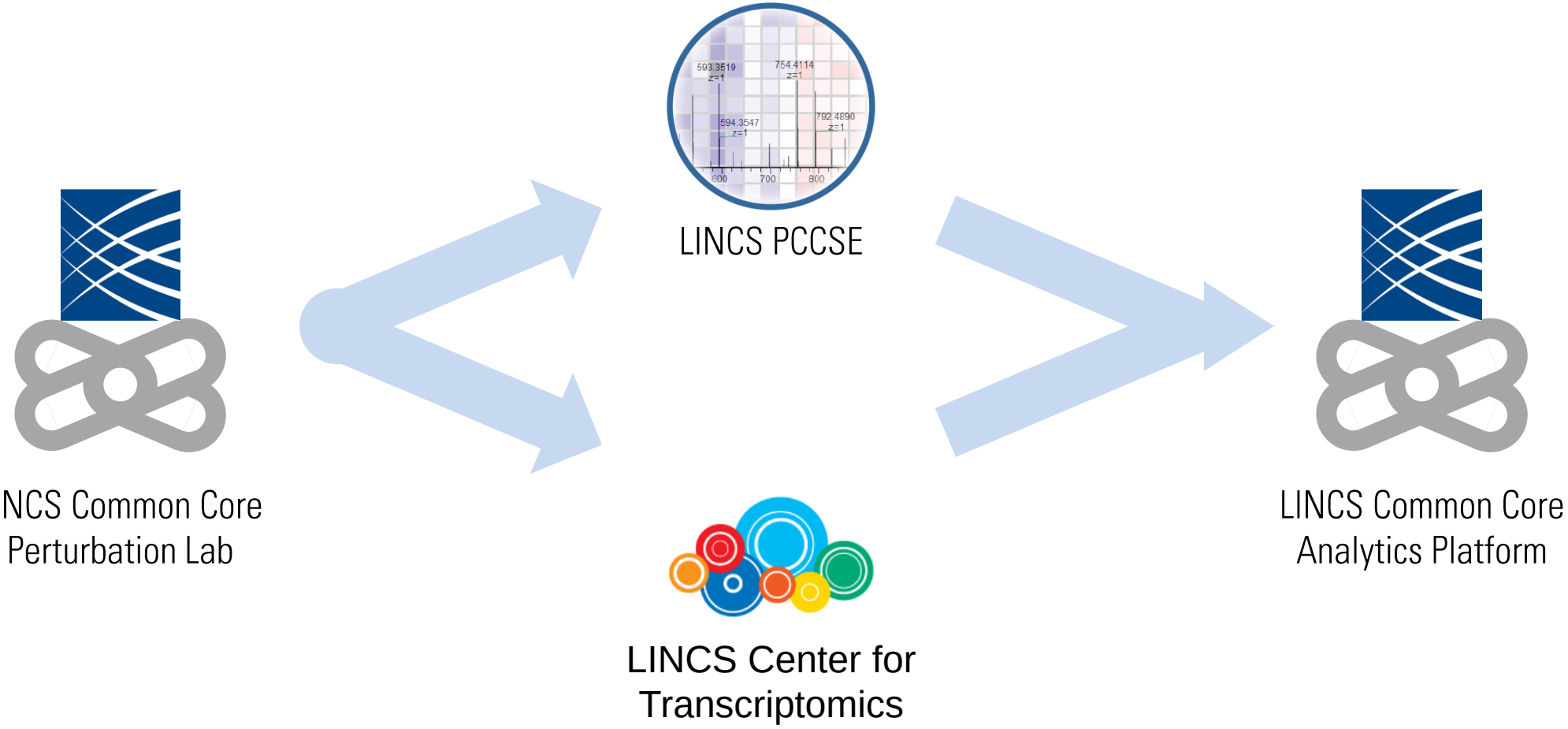
- Next-gen MS methods & analysis dev
- Key software components
- Data dissemination infrastructure
- Deep connections to computational community

- High-throughput proteomics
- Reagents and robust profiling infrastructure
- Experienced analytics teams
- Project coordination and execution

Broad Institute Proteomics enables high throughput profiling

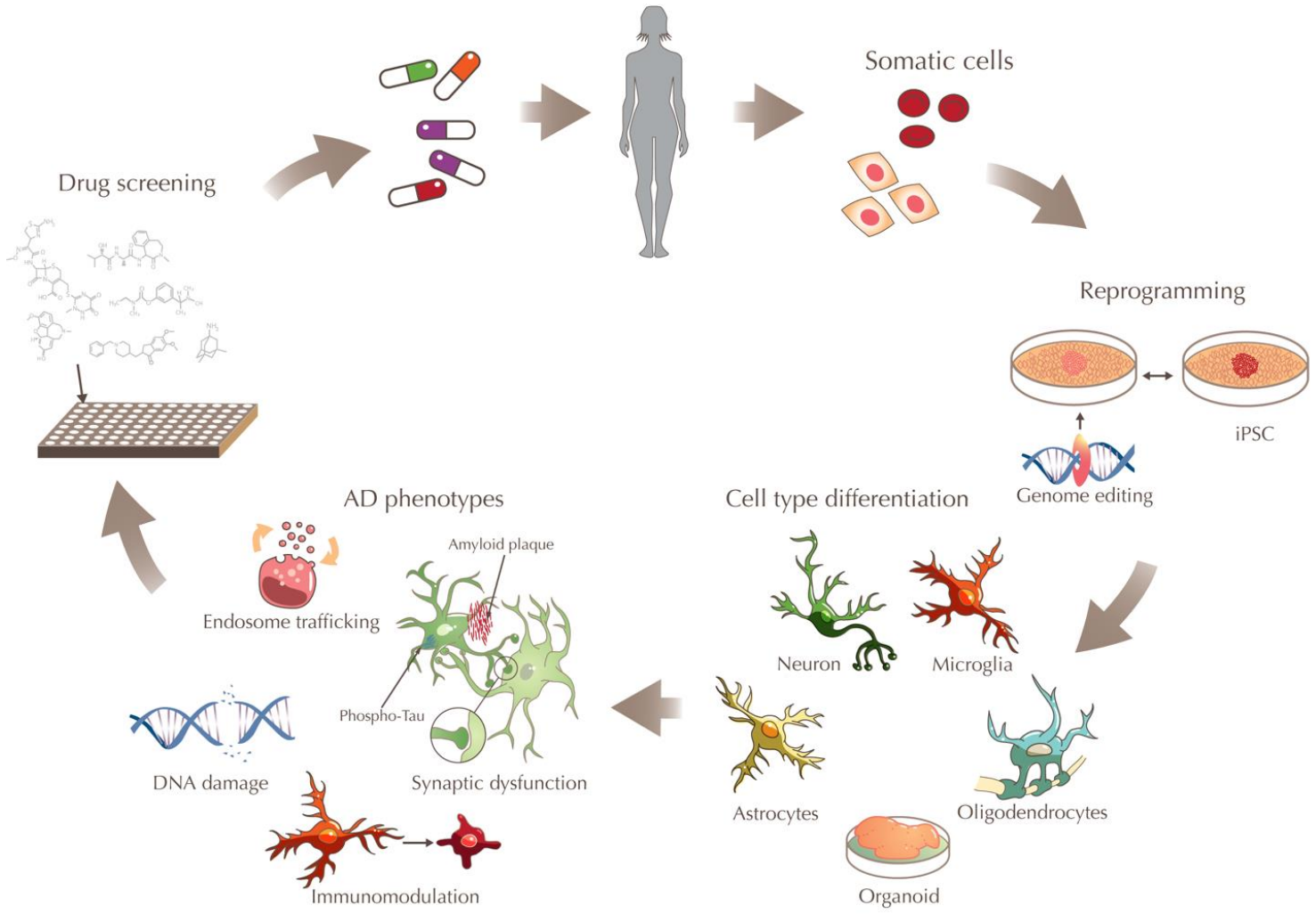


Broad Institute LINCS Common Core creates synergy and cooperation



TSAI LAB - MIT

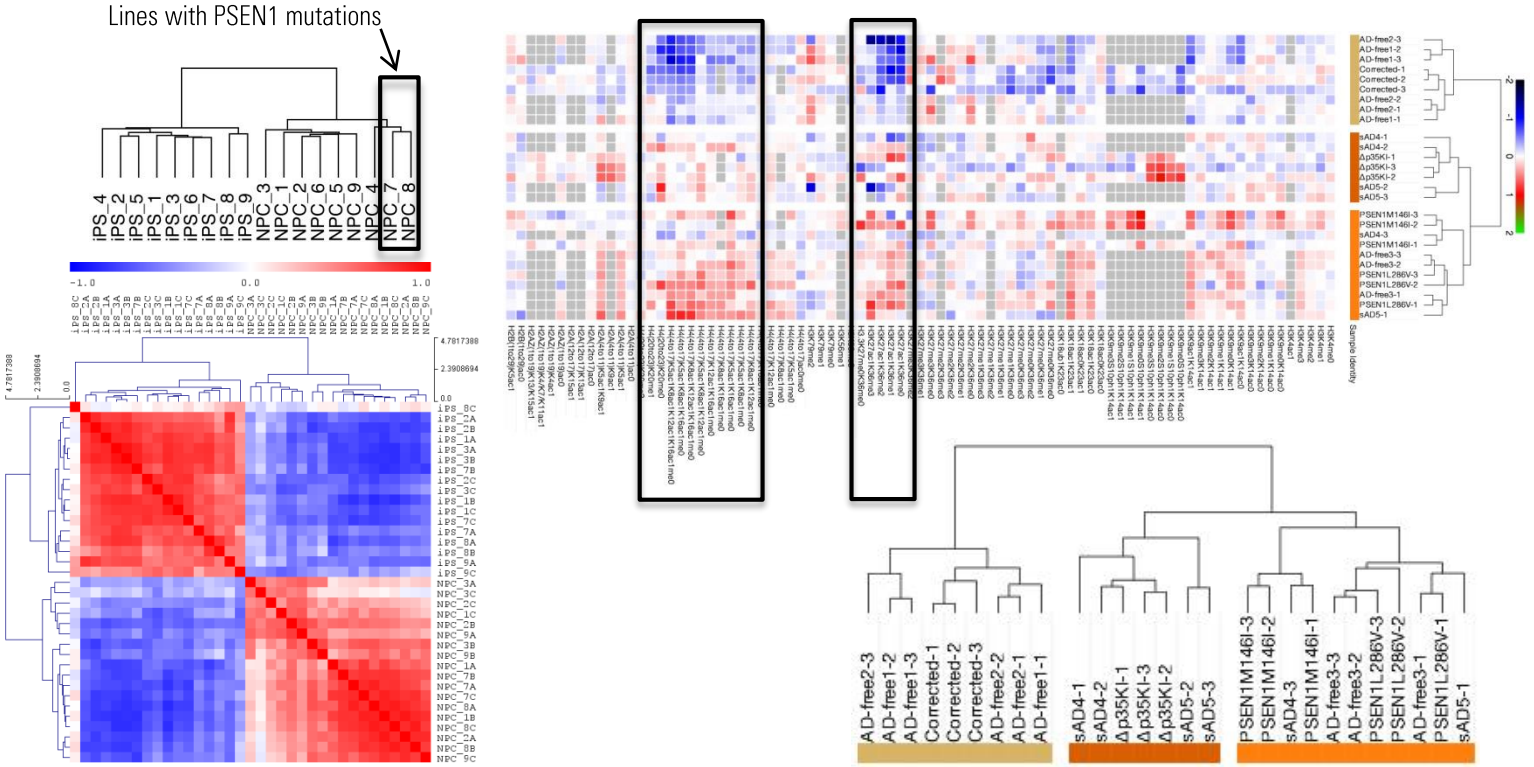
Using human reprogrammed stem cells (iPSCs) to model disease



- Complex neurological disorders are not a single disease at the molecular level
- Combinatorial analysis of phosphosites and histone marks give idea to phosphosignaling pathways and epigenetic modifications, and how these processes intersect
- Mechanistic differences via multiple signature assays

Mungenast *et al.*, 2015, Mol Cell Neurosci

LINCS enables systematic study of genetic basis of disease



Drug	Target/Mechanism of Action	Drug	Target/Mechanism of Action
Ex527	**SIRT1 and SIRT2 inhibitor**	Everolimus	mTOR inhibitor
ruxolitinib	JAK inhibitor	BMS906024	pan-Notch inhibitor
sirolimus	mTOR inhibitor	losmapimod	p38 inhibitor
salermide	**SIRT1 and 2 inhibitor**	Nilotinib	tyrosine kinase inhibitors
calpain inhibitor II	calpain inhibitor	lenalidomide	immune suppressor, anti-angiogenesis
Tretinonin	Retinoic acid	AR A014418	GSK-3 inhibitor
TBB	CK2/Dyrk1a inhibitor	Selumetinib	MEK1/2 inhibitor
Tofacitinib	JAK inhibitor	**VX-970**	**ATR inhibitor**
Olaparib	**PARP inhibitor**	afuresertib	AKT inhibitor
KN-93	CAMKII/IV inhibitor	Ipl145	PI3K delta, gamma inhibitor
KU55933	**ATM inhibitor**	BMS-345541	nf-KB inhibitor
1271738-62-5	MLL inhibitor	PS-1145	nf-KB inhibitor

- Distinct chromatin states in iPS cells and iPS-derived neural progenitors from Alzheimer's disease patients
- Analysis of GCP signatures from drug perturbations that cluster with PSEN1 isogenic lines
  - genome integrity
  - immune activation

LINCS enables systematic study of genetic basis of autism

- Epigenetic processes are disrupted in neurological disorders such as autism spectrum disorders (ASDs)

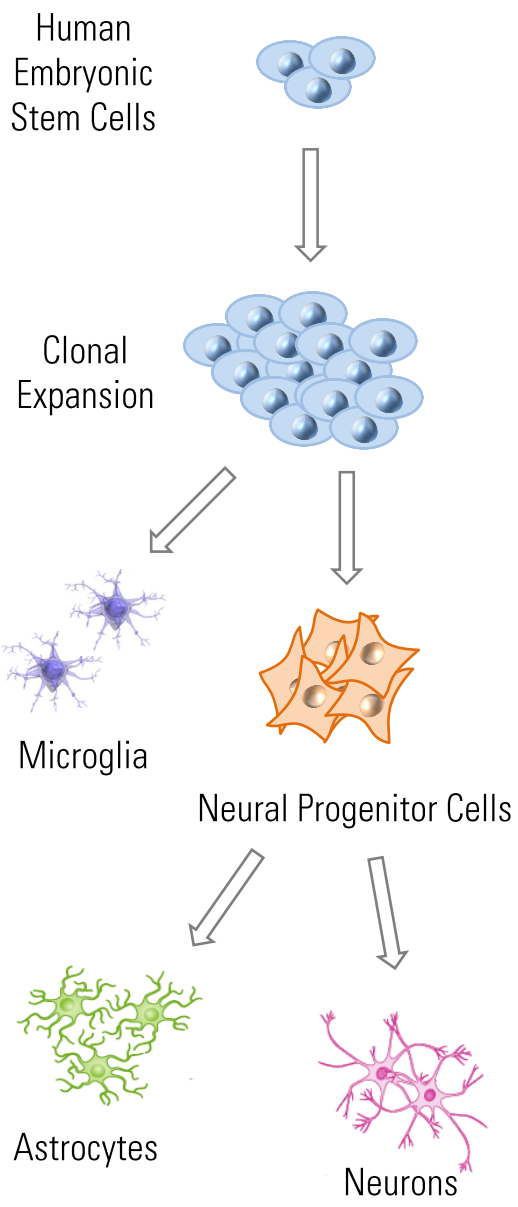
- Causal mutations have been found in chromatin-associated enzymes in ASD patients

- Aim: Study neural cells during development and across cellular subtypes
  - establish epigenetic and phosphosignaling signatures from compound perturbations
  - study effects of knocking down chromatin-modifying enzymes
  - re-create perturbations in ASD risk genes

Autism Spectrum Disorder Genes			
ADNP	DYRK1a	MECP2	SEMA5A
ARID1B	EHMT1	NLGN3	SHANK3
CACNA1C	EIF4G1	NLGN4X	SYNGAP1
CNTN4	EPC2	NRXN1	TLK2
CNTNAP2	FMR1	PTCHD1	TSC1
CREBBP	GRIN2B	PTEN	TSC2
CTCF	GRIP1	SCN1A	UBE3a
CTNNB1	GRM5	SCN2a	UBE3b

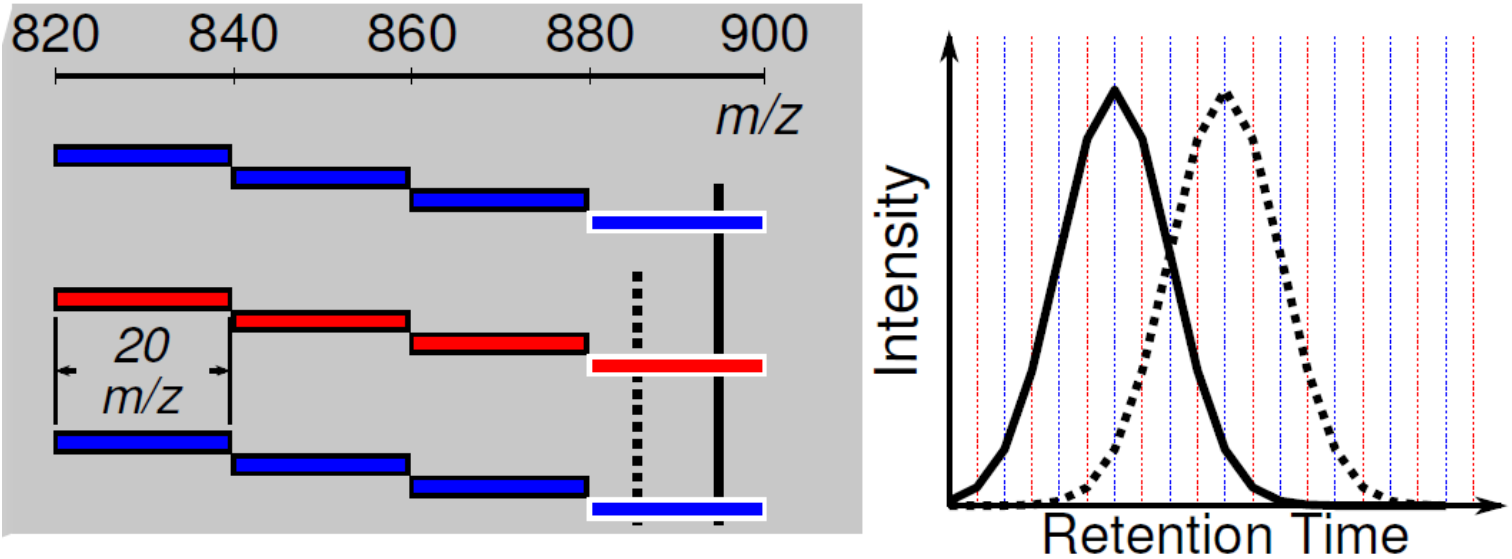
Current efforts to establishing neurobiology models

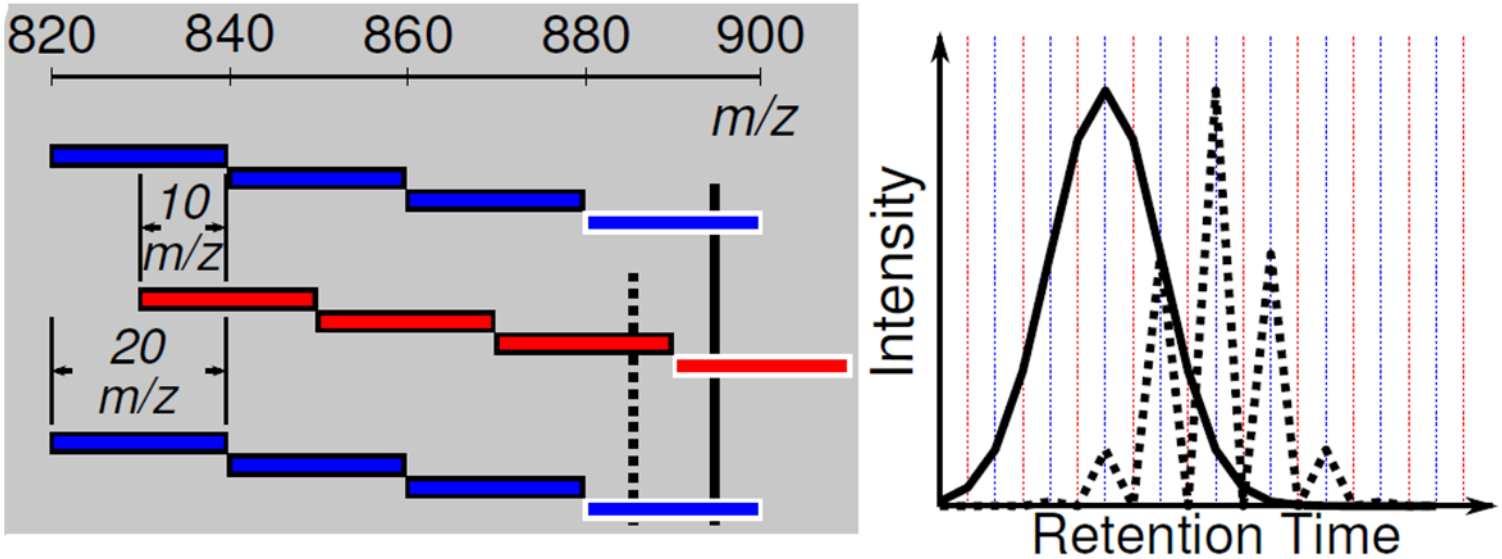
- Creating neural cell types for high-content proteomics assays
  - Generated neural progenitor cells (NPCs) from human embryonic stem (ES) cells
  - Generating neurons with passive differentiation via growth factor withdrawal
  - Future: Will generate microglia and astrocytes
- Establishing a CRISPR/Cas9 neural system
  - H9 Cas9 ES cell lines
  - HUES8 iCas9 ES cell lines
- Outreach with MEP LINCS at OHSU



MACCOSS LAB – UNIV. WASHINGTON

Novel MS data acquisition technologies

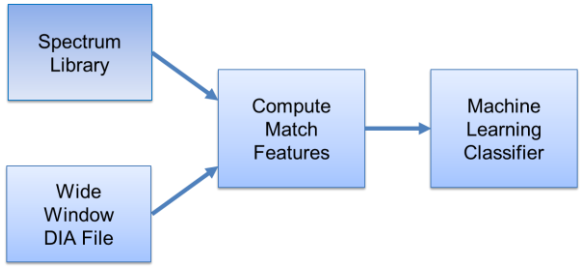




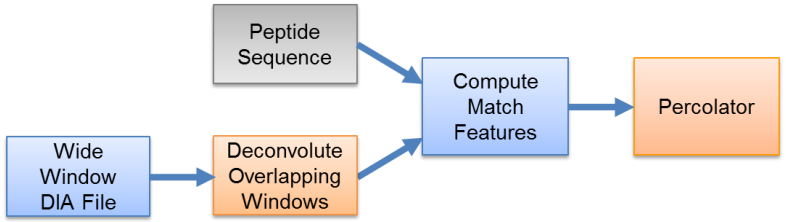
Jarrett Egertson
Austin Keller

Novel analytical methods for Next-Gen MS

- Typical Workflow

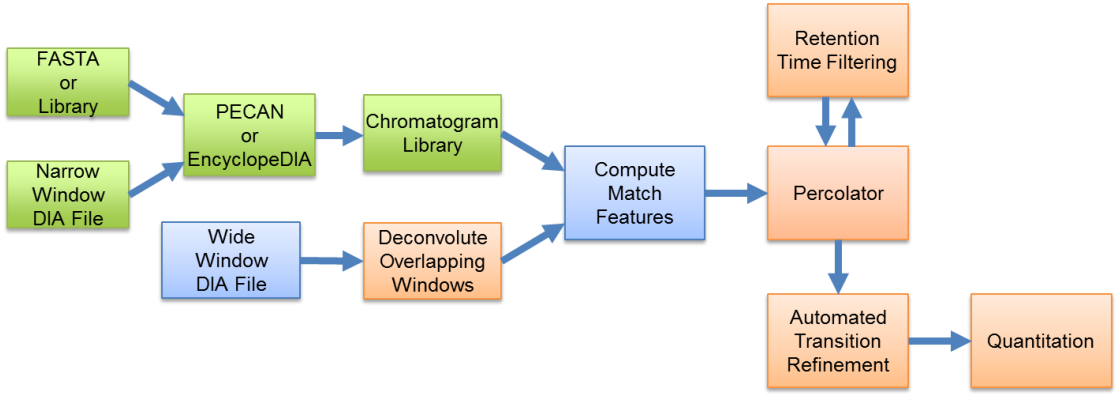


- Tool 1: Pecan
  - Uses only peptide sequences
  - No spectral or retention time information

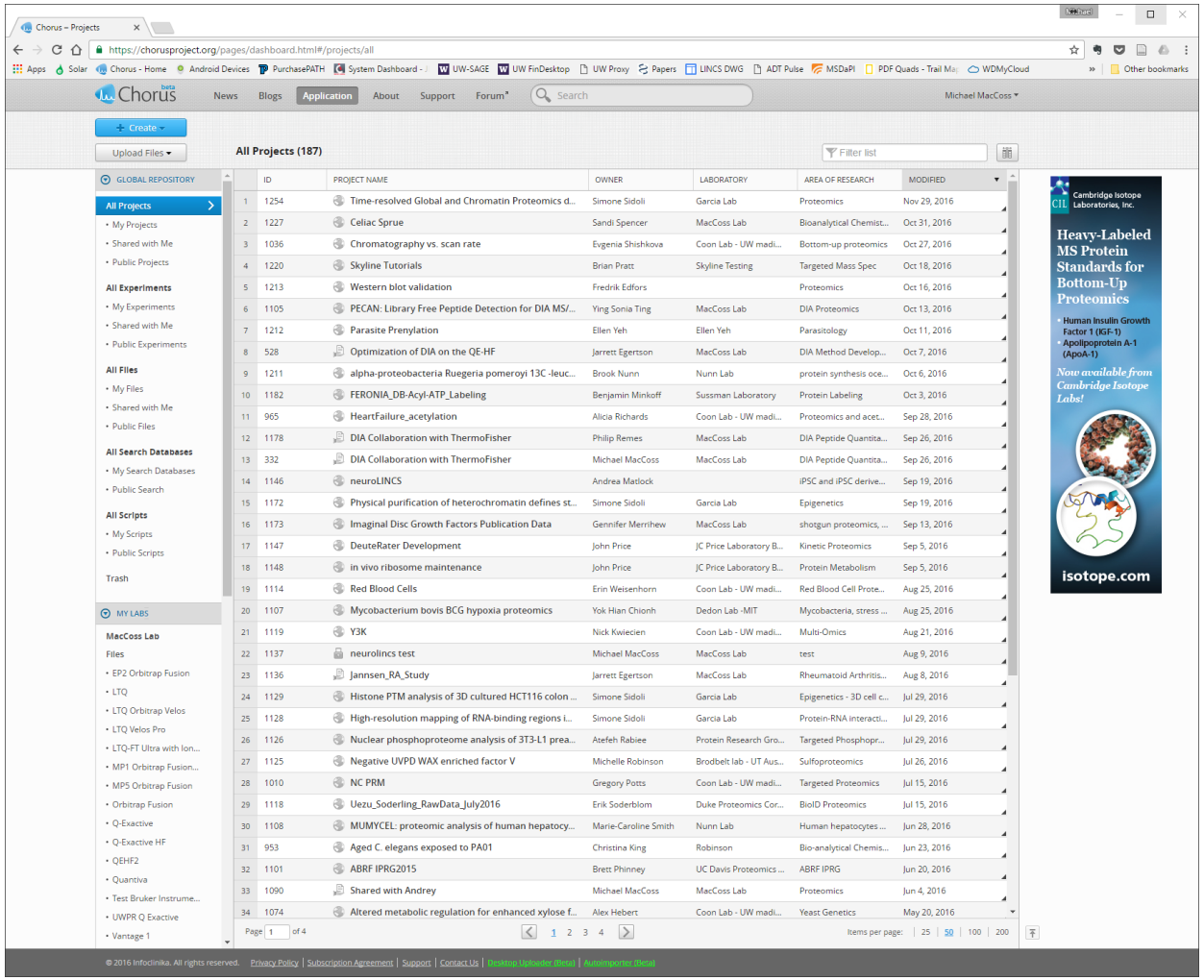


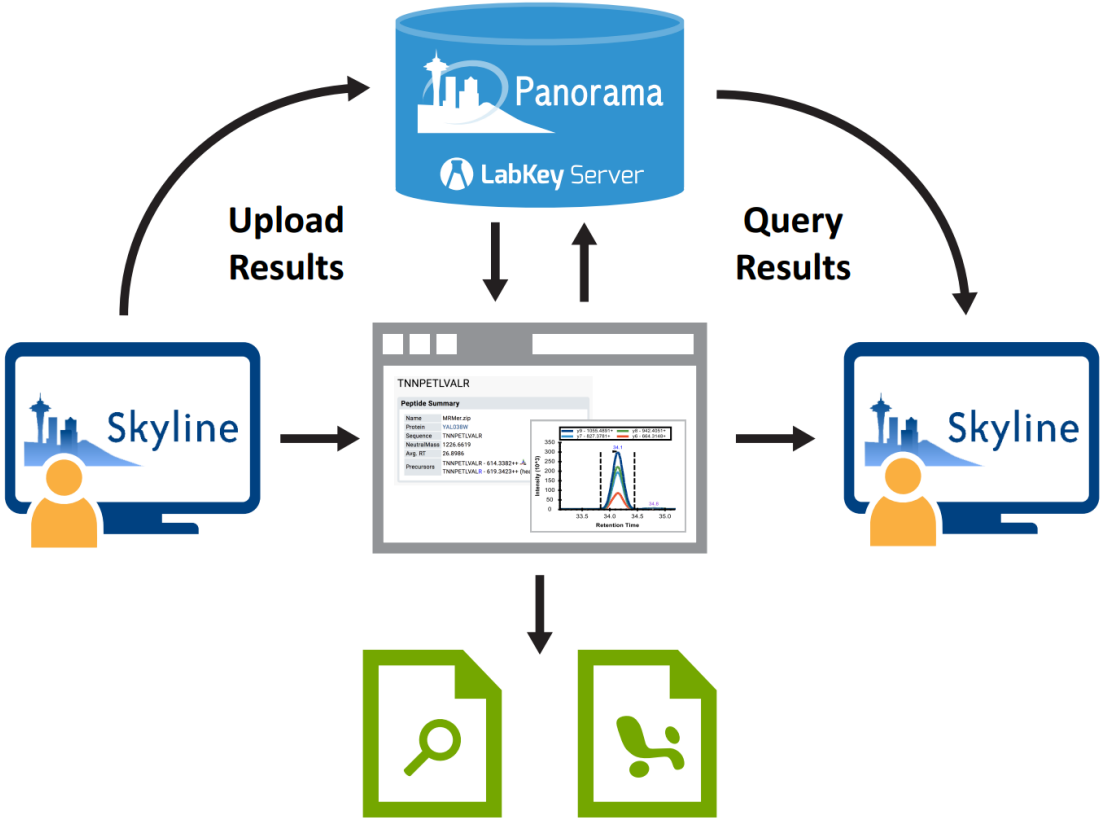
Jarrett Egertson
Brendan MacLean
Brian Searle
Sonia Ting

- Tool 2: EncyclopeDIA
  - Uses peptide sequences
  - Spectrum or Chromatogram libraries
  - Positional isomers



Robust tools for data management and sharing





Andrey Bondarenko

Vagisha Sharma
Brendan MacLean

LINCS PCCSE

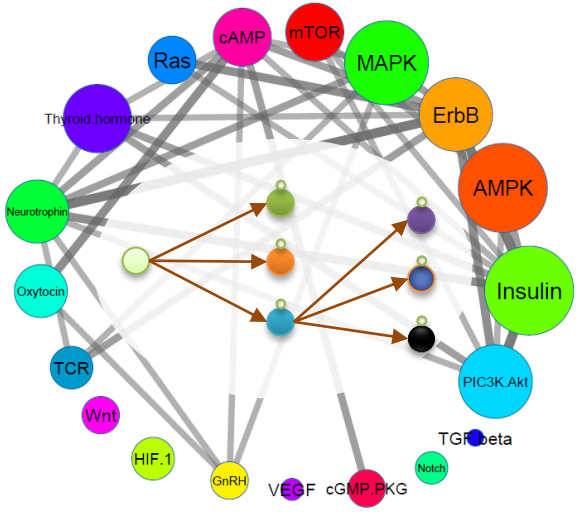
I: DATA GENERATION & ANALYSIS

LINCS PCCSE

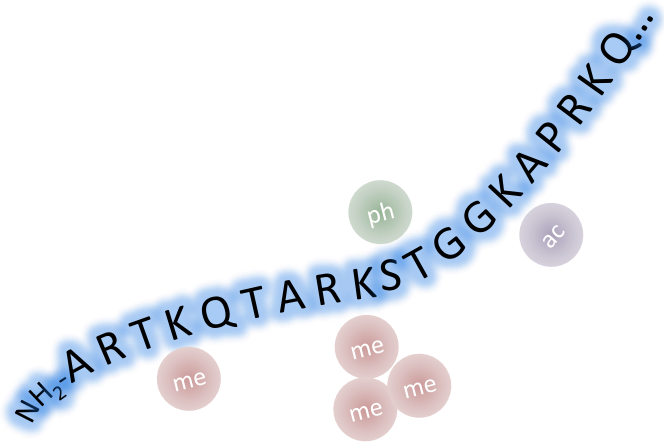
02 - DATA GENERATION OVERVIEW

Targeted proteomics assays

P100; 96 phosphopeptide probes commonly observed and modulated in diverse cell types

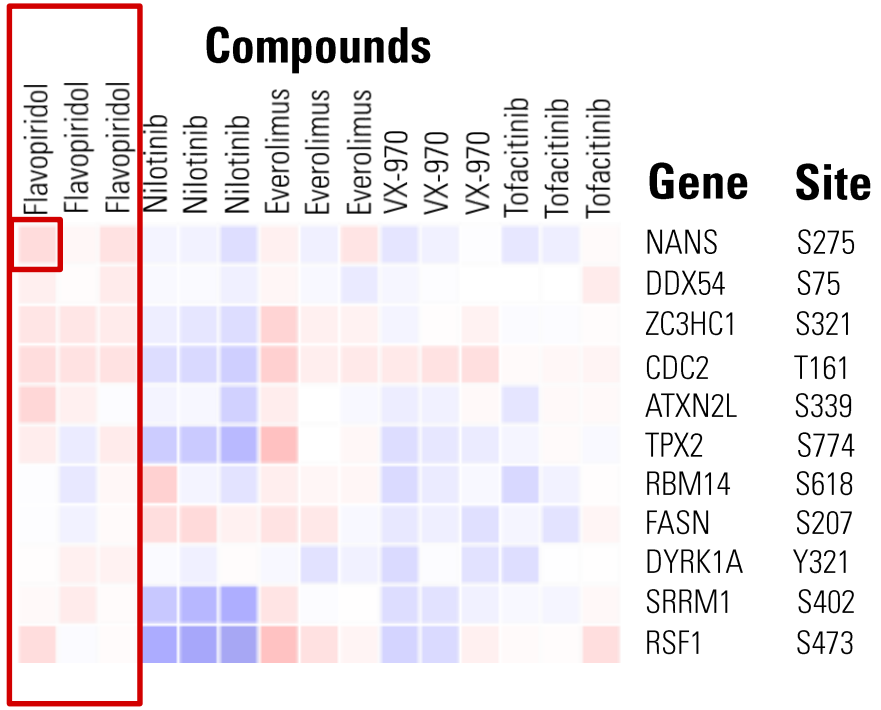


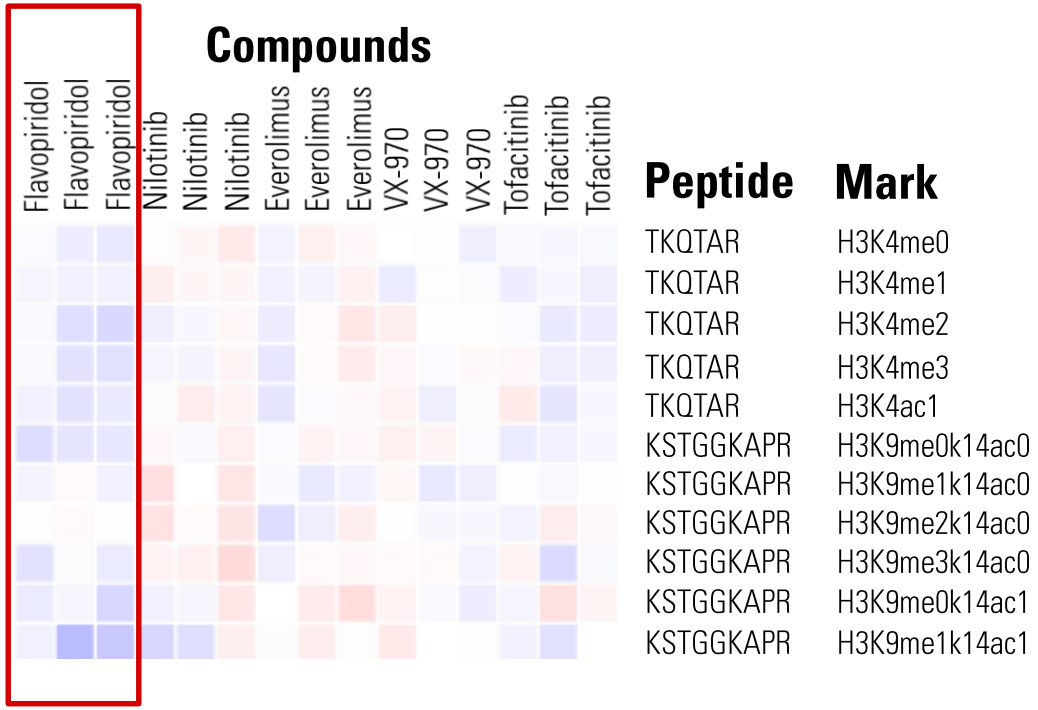
GCP; ~ 60 probes corresponding to combinatorial post-translational modifications on histone tails



- Monitor the impact of *drug* and *genetic* perturbations in a range of disease and developmental models
- Understanding *processes underlying disease* that will lead to *novel therapeutic intervention*

Perturbation *profiles* are composed of individual quantitative readouts





Key innovations that have enabled our progress

- Establish production-level workflows

  cellular treatments, sample collection, lysis buffers, sample preparation, phosphopeptide enrichment

- Incorporate Quality Control procedures
- Develop automated workflows (BRAVO Liquid Handling Platform)
- Develop Mass Spectrometry analysis methods (Q-exactive series)
- Develop data analysis templates (Skyline)

Harmonized orchestration of data generation

Weekly Meeting

Sample Preparation

- Cell Lines, Primary Cells, Neural Lineages
  - Drug Perturbations
  - Gene Disruption/Editing
  - Neuronal Fates
- Sample Collection (96-well plates, triplicates)

- P100[1]
  - Cell Lysis
  - Protein Quantitation
  - Protein Level Normalization
  - Protein Digestion*
  - Protein Desalting
  - Tip Preparation*
  - Sample Resuspension*
  - Phosphopeptide Enrichment*
  - Peptide Desalting
  - Sample Resuspension
- GCP[2]
  - Histone Extraction
  - Protein Quantitation
  - Protein Level Normalization
  - Primary Propionylation*
  - Protein Desalting
  - Protein Digestion*
  - Secondary Propionylation*
  - Peptide Desalting
  - Sample Resuspension

*automated

Sample Analysis

- Mass Spectrometry
- Skyline

Batch & Integrative Analysis

- Data Integration
- Proteomics Signatures of Perturbations
- Connections between Perturbations & Signatures
- Integration of Biology

…etc…

[1]Abelin *et al*, 2016, MCP, [2]Creech *et al*, 2015, Methods

Standard Operating Procedures (SOPs) in Panorama, LINCS

Panorama PanoramaWeb

LINCS | LINCS PCCSE Overview

Overview Information

Start Page >

SOPs

P100

- 1-1 P100 Lysis Buffer Prep.pdf
- 1-2 P100 Automated Protein Assay.pdf
- 1-3 P100 Protein Level Normalization.pdf
- 1-4 P100 Automated Protein Digestion.pdf
- 2 P100 Positive Pressure 96-Well Desalt.pdf
- 3-1 P100 Tip Prep.pdf
- 3-2 P100 Sample Resuspension.pdf
- 3-3 P100 IMAC Enrichment-Agilent-NTA.pdf
- 3-4 P100 RPS Desalt.pdf
- 4 P100 Final Sample Resuspension.pdf
- A1 P100 Synthetic Heavy Mix Stock Creation.pdf
- A2 P100 Data Acquisition Guidelines.pdf

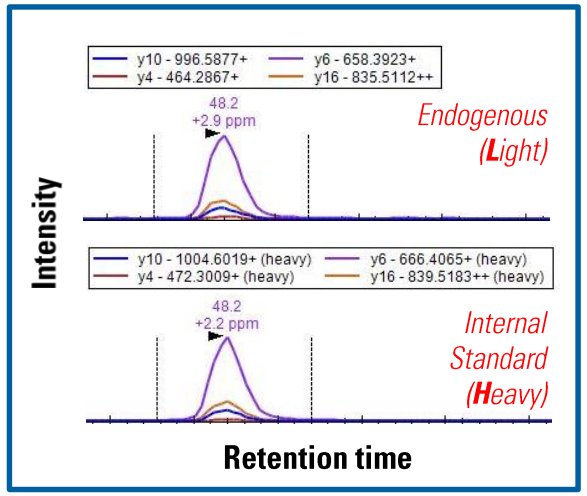
GCP

- 1 GCP Histone Extraction.pdf
- 2 GCP Protein Quantification and Purity Determination.pdf
- 3 GCP Primary Propionylation.pdf
- 4 GCP Protein Oasis Desalt.pdf
- 5 GCP Trypsin Digest.pdf
- 6 GCP Secondary Propionylation.pdf
- 7 GCP Final Desalt.pdf
- 8 GCP Mass Spec Analysis.pdf
- A1 GCP Mastermix and Equimix Formulation Manual.pdf
- A2 GCP Assay Lysis Buffer and Additive Prep.pdf

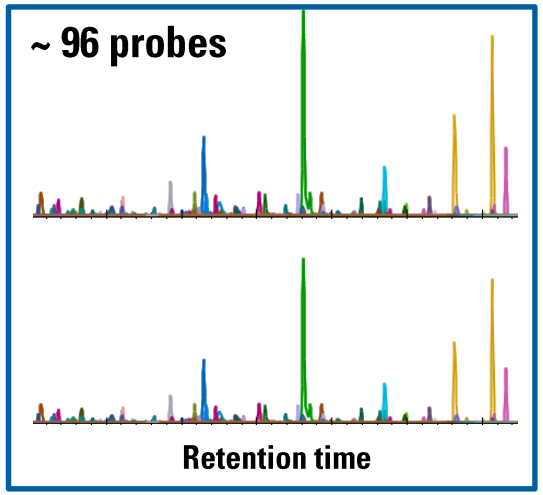
DISCUSSION

https://panoramaweb.org/labkey/wiki/LINCS/Overview%20Information/page.view?name=sops

Quantitative readouts of phosphosignaling and chromatin modifications

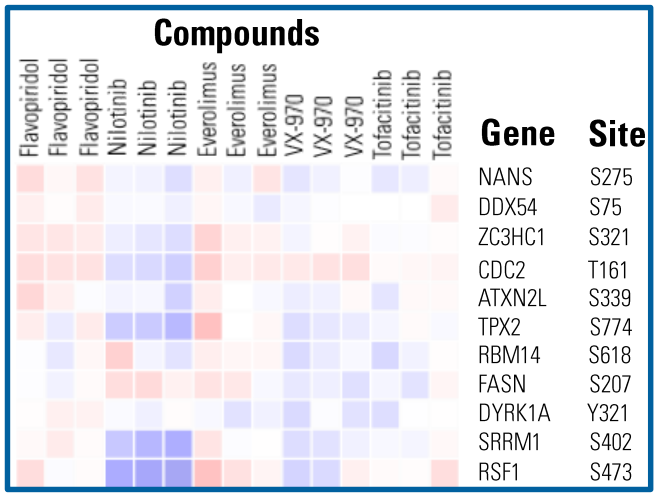


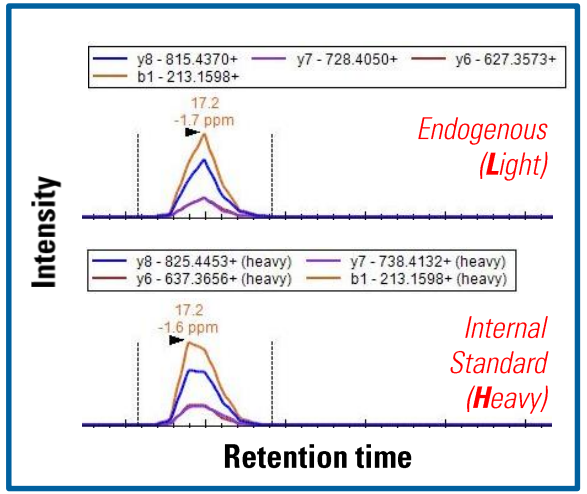
$\frac{L}{H}$



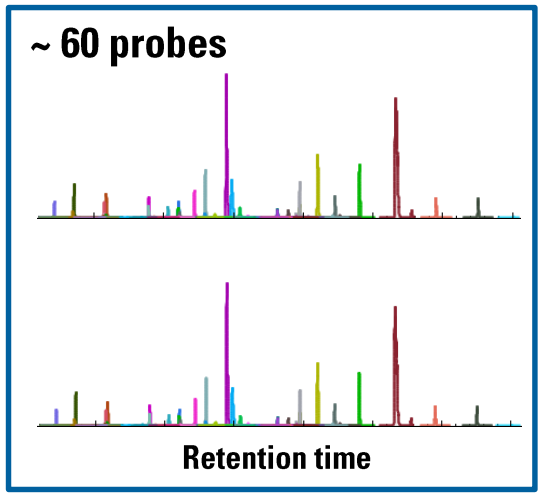
$\left(\frac{L_1}{H_1}, \frac{L_2}{H_2}, \ldots, \frac{L_{96}}{H_{96}}\right)$

Automated
Data
Processing



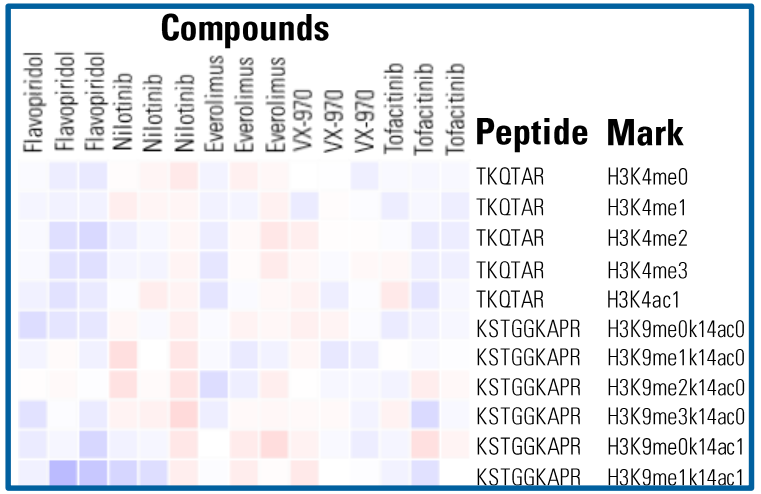


$\frac{L}{H}$



$\left(\frac{L_1}{H_1}, \frac{L_2}{H_2}, \ldots, \frac{L_{96}}{H_{96}}\right)$

Automated
Data
Processing



Unprecedented amount of *drug perturbation* data generated

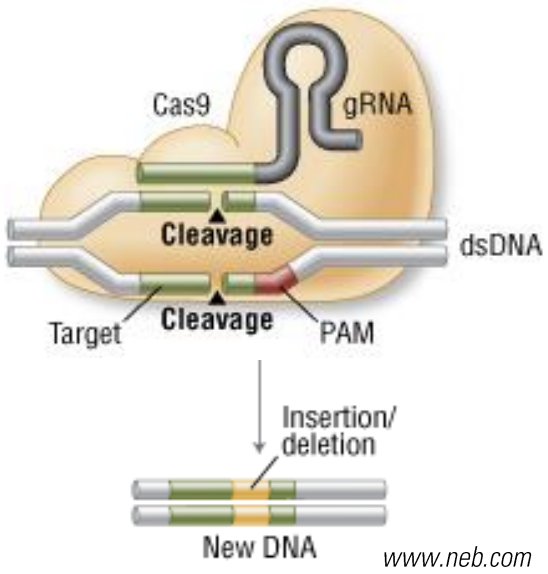
Cells	Provenance	Institute	Epigenetically Active	Neuroactive	Kinase/Pathway Inhibitors	
A375	Skin Cancer	Broad	●	●	●	3400+
YAPC	Pancreatic Cancer	Broad	●	●	●	
A549	Lung Cancer	Broad	●	●	●	
MCF7	Breast Cancer	Broad	●	●	●	
PC3	Prostate Cancer	Broad	●	●	●	
H9	NPCs (ESC-derived)	Broad	●	●	●	
HUES8	NPCs & Neurons	MIT	●	●	●	
iPS	NPCs (patient derived)	Mass General Hospital	◐	◐	◐	600+
HUVEC, HAoSMC	Cardiovascular	Tufts Medical Center	◐	◐	◐	

● *Completed*
● *In Progress*

*https://panoramaweb.org/labkey/project/LINCS

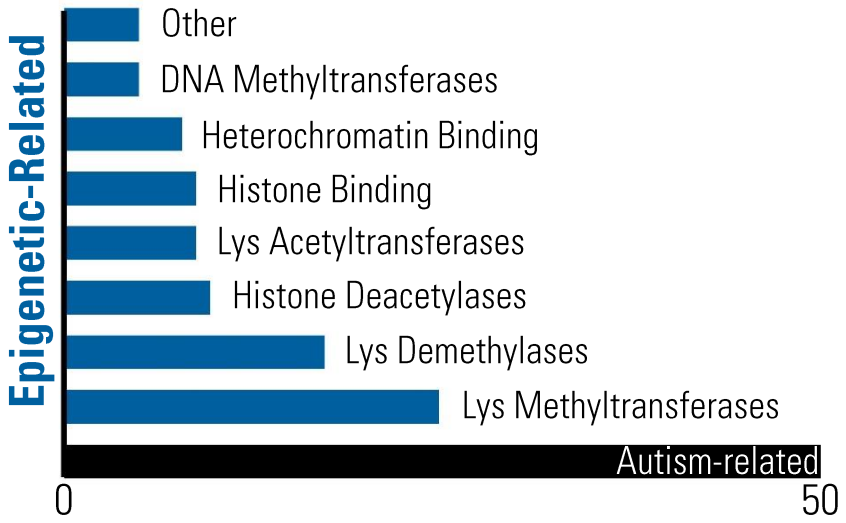
Genome editing leads next phase of data generation

- CRISPR/Cas9 technology for disrupting genes:
  - modifying the epigenetic landscape
  - implicated in neurodevelopmental disorders
- Proteomic readouts for 1150+ samples to date



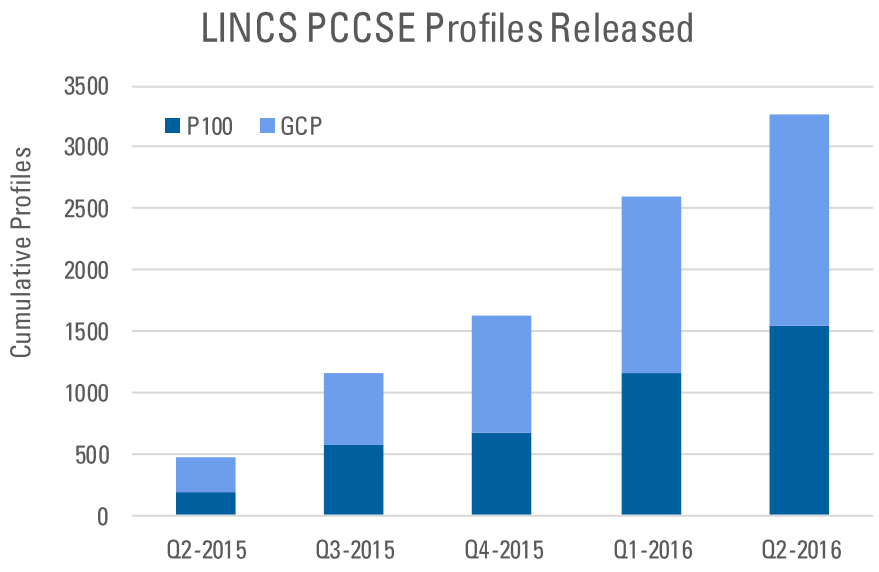
Cells	Provenance	Institute	Epigenetic-Related	Autism-related
A375	Skin Cancer	Broad	●	
YAPC	Pancreatic Cancer	Broad	●	
A549	Lung Cancer	Broad	●	
MCF7	Breast Cancer	Broad	●	
PC3	Prostate Cancer	Broad	●	
HUES8	NPCs & Neurons	MIT	●	●

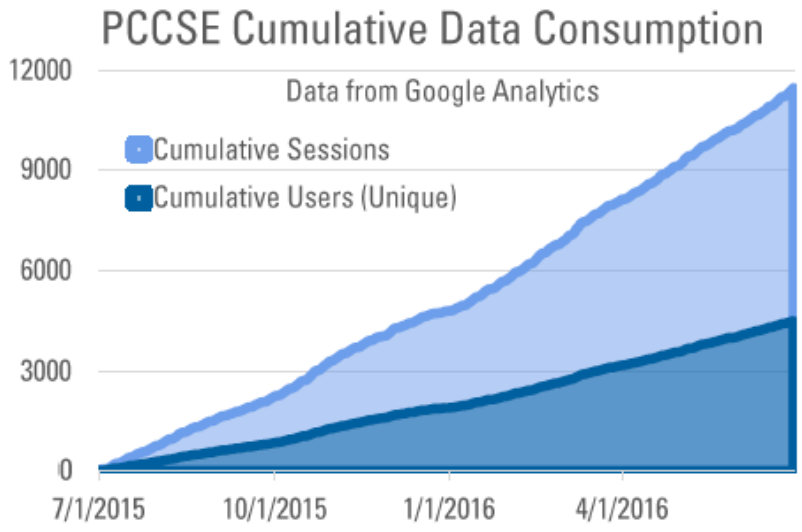
● *Completed*
● *In Progress*



Highlighted achievements of data generation

- Developed reproducible, robust and automated workflows for sample preparation and analysis
- Completely documented SOPs
- Publication of GCP and P100 assays
- Vast amount of data generated & released
- Public Panorama data repository established*
- Increasing number of collaborations





*https://panoramaweb.org/labkey/project/LINCS

The even more promising future

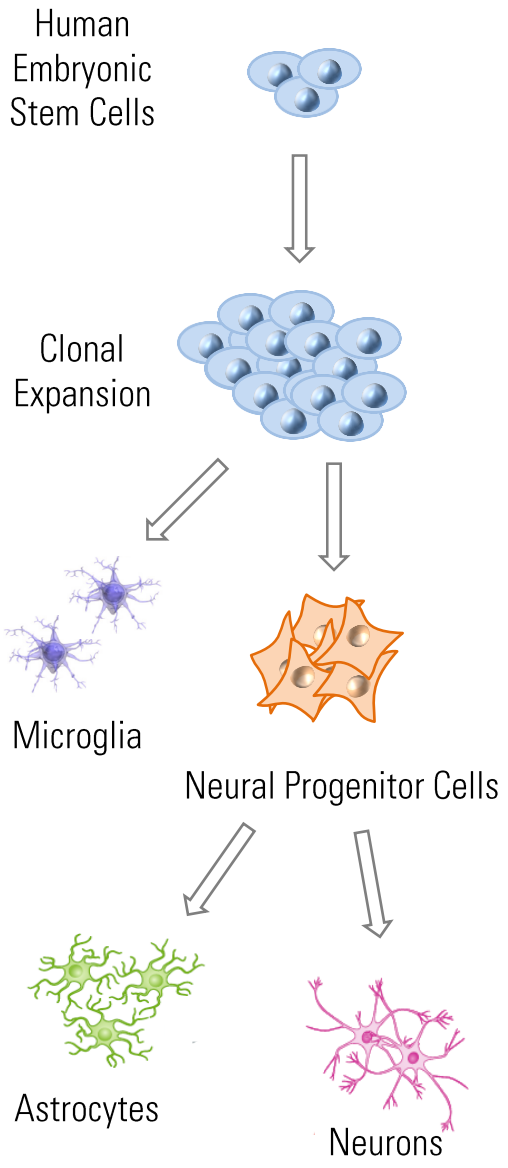
- Gene editing using the CRISPR/Cas9 system
- Increased neurobiology focus
  - Epigenetic & phosphosignaling signatures at different stages of neural development and cell types
  - CRISPR/Cas9 for genes implicated in neurodevelopmental disorders
- Interactive tools for data analysis
  - Improve our understanding of how perturbations are related to one another
  - Establish connectivity queries for any given perturbation
- Next-generation MS profiling for phosphosignaling
  - Comprehensive mass spectrometry analysis (DIA) for greater phosphosite coverage

LINCS PCCSE

03 - NEURAL LINEAGES AND MODELS

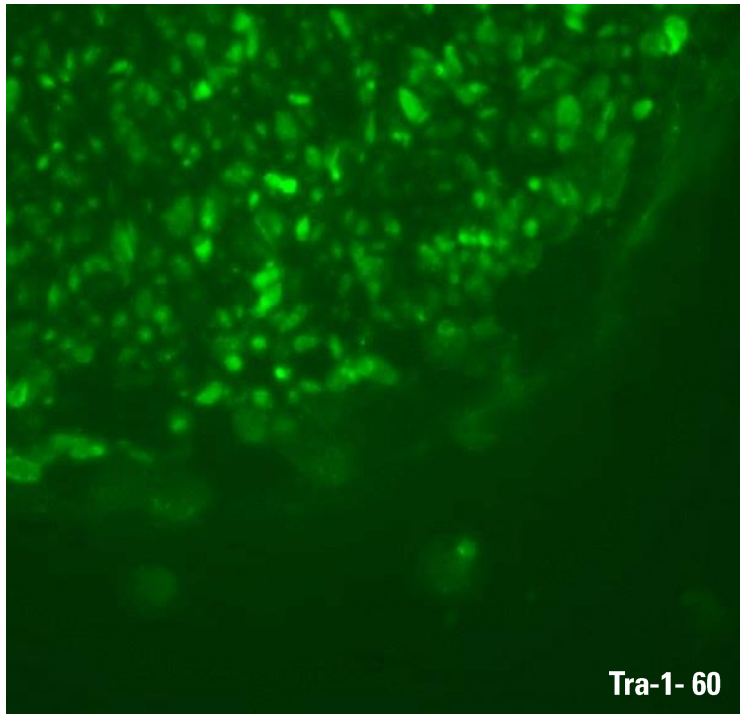
Current efforts in establishing neurobiology models

- Creating neural cell types for high-content proteomics assays
  - Generated neural progenitor cells (NPCs) from human embryonic stem (hES) cells
  - Generating neurons with passive differentiation via growth factor withdrawal
  - Future: Will generate microglia and astrocytes
- Establishing a CRISPR/Cas9 neural system
  - H9 Cas9 ES cell lines
  - HUES8 iCas9 ES cell lines

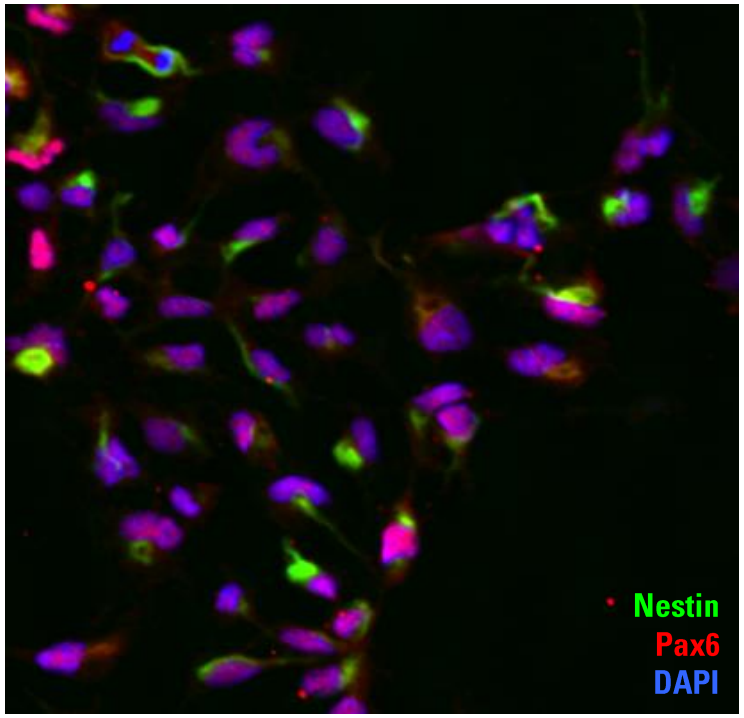


Current efforts in establishing neurobiology models

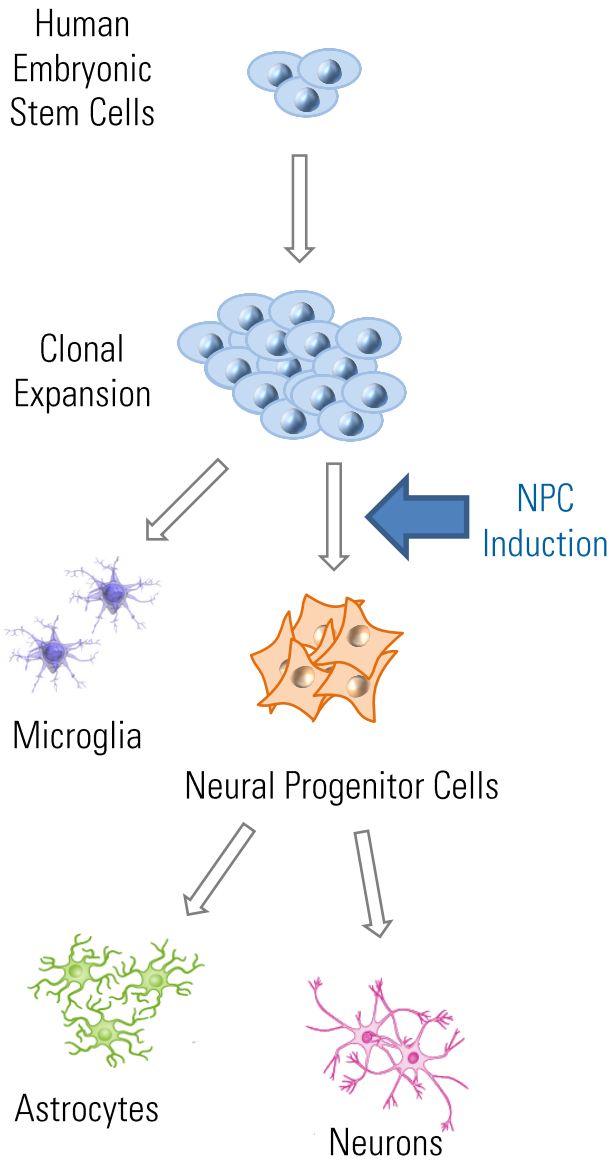
- Creating neural cell types for high-content proteomics assays
  - Generated neural progenitor cells (NPCs) from human embryonic stem (hES) cells
    - P100 and GCP signatures from compound treatments



H9 ES Cells

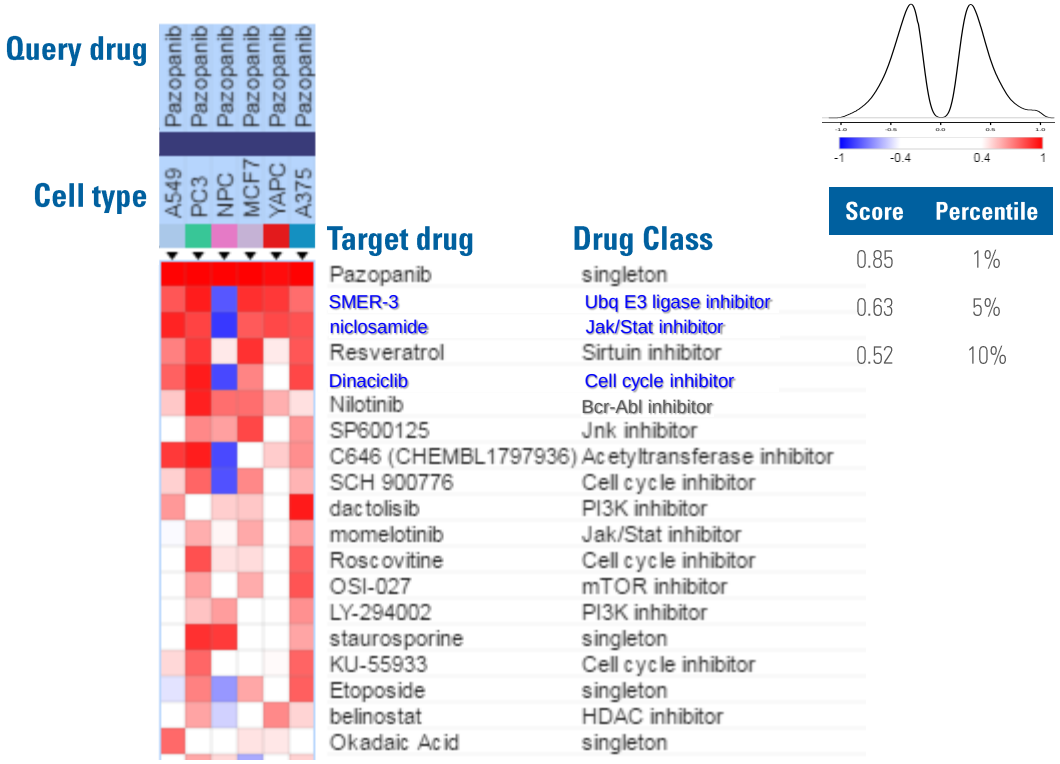


H9 ES-derived neural progenitors

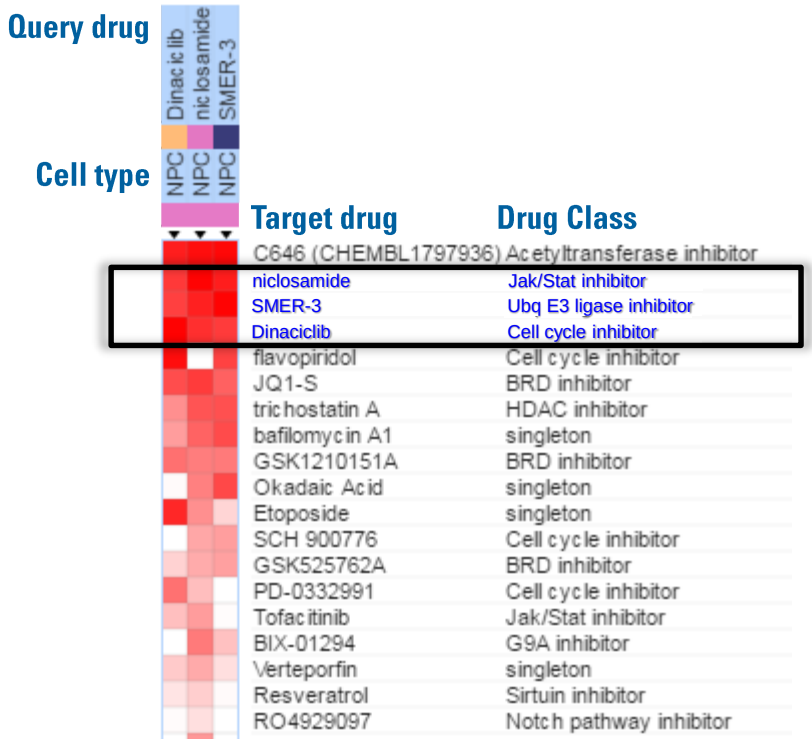


P100 analysis reveals an NPC-specific divergence of signaling pathways

P100 Connectivities in Cancer Cells and NPCs



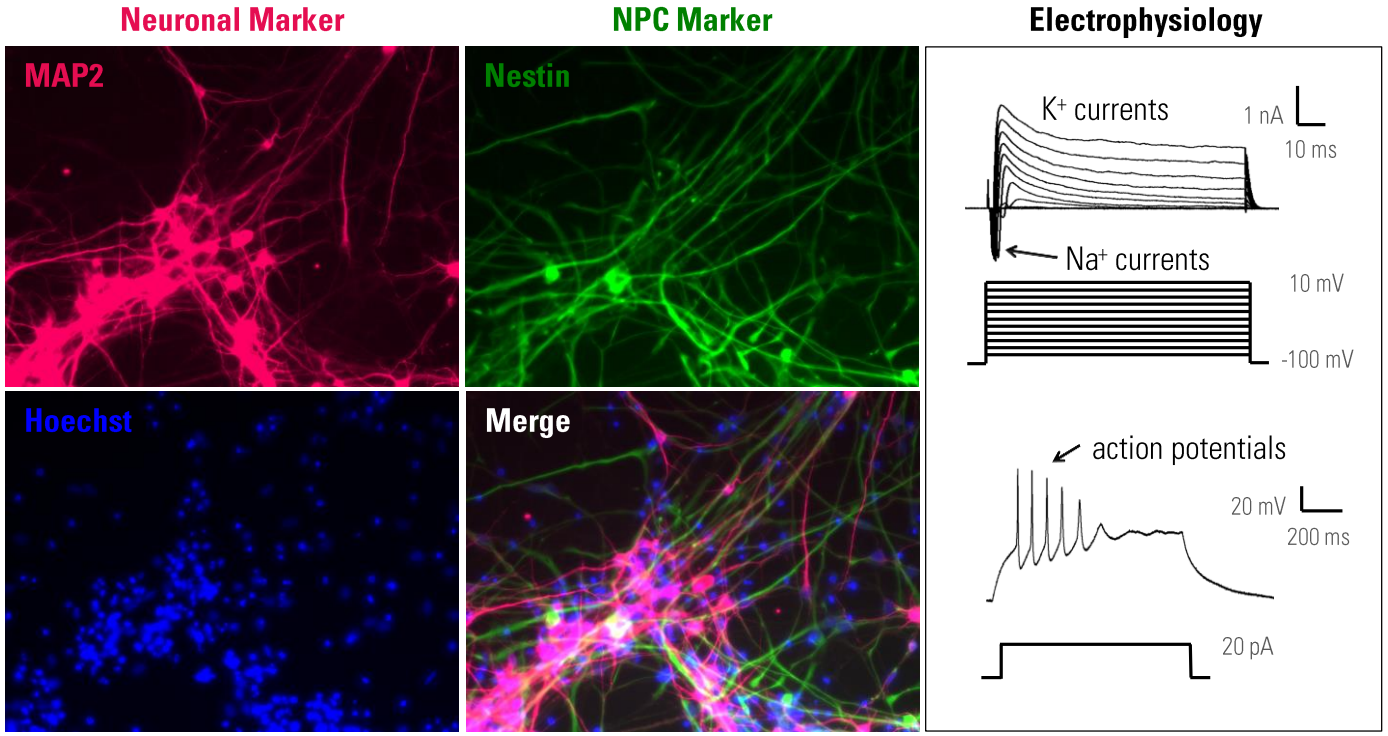
Connectivity of Pazopanib Cancer Connections in NPCs



- Pazopanib is a top ranked connection in all cell lines but downstream effectors are strong negative connections in NPCs

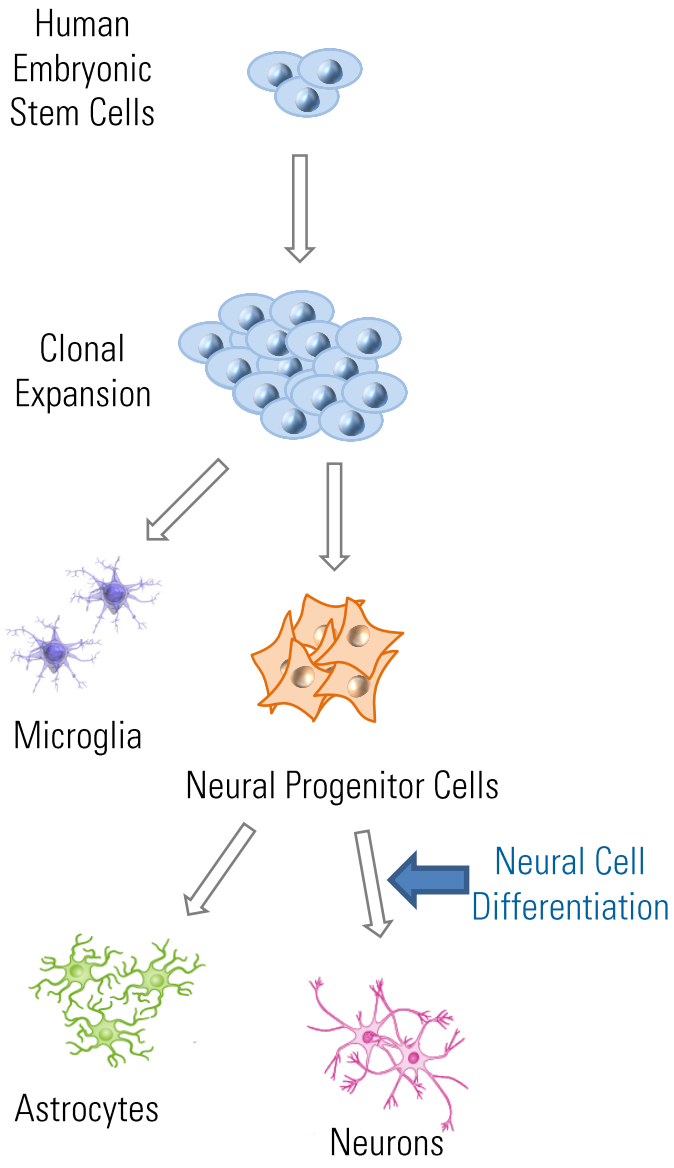
Growth factor withdrawal enables large-scale production of mature neurons

- Passive differentiation of NPCs into neurons through growth factor withdrawal (GFW)
  - mature neurons in ~8 weeks
  - *heterogeneity of culture*

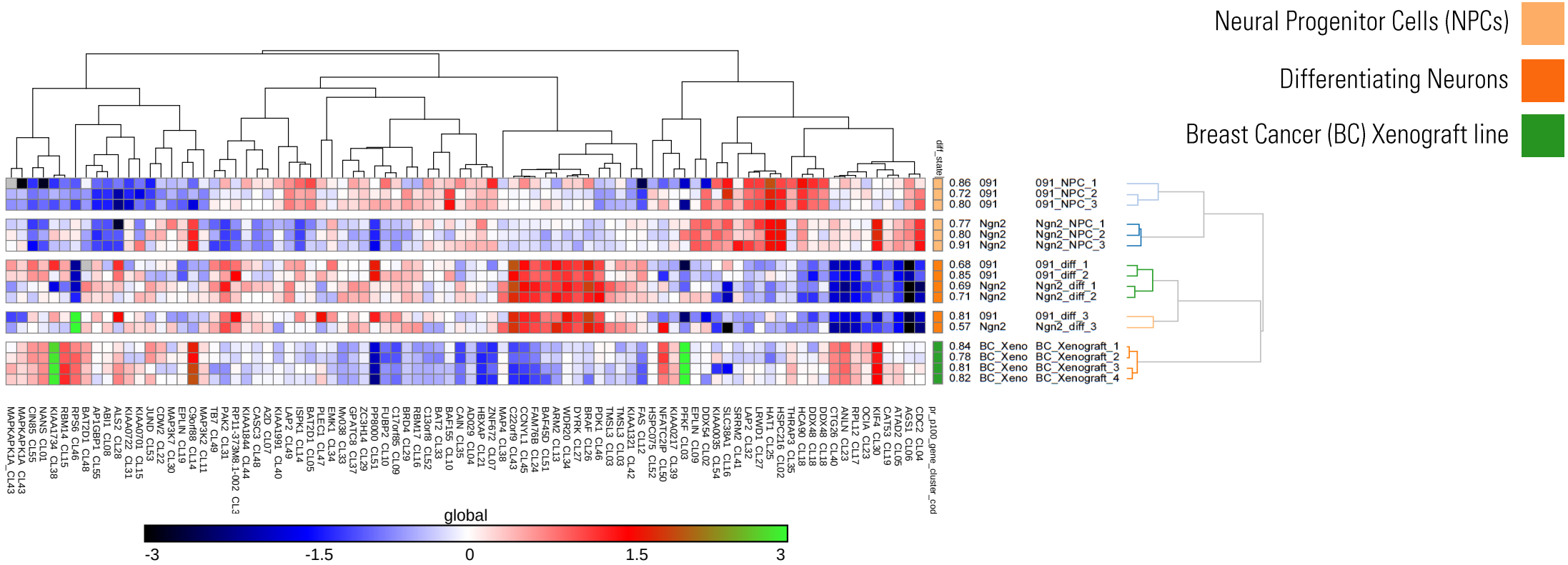


6.5 weeks of GFW

8 weeks of GFW



P100 analysis reveals a convergence of signaling states from NPCs to neurons



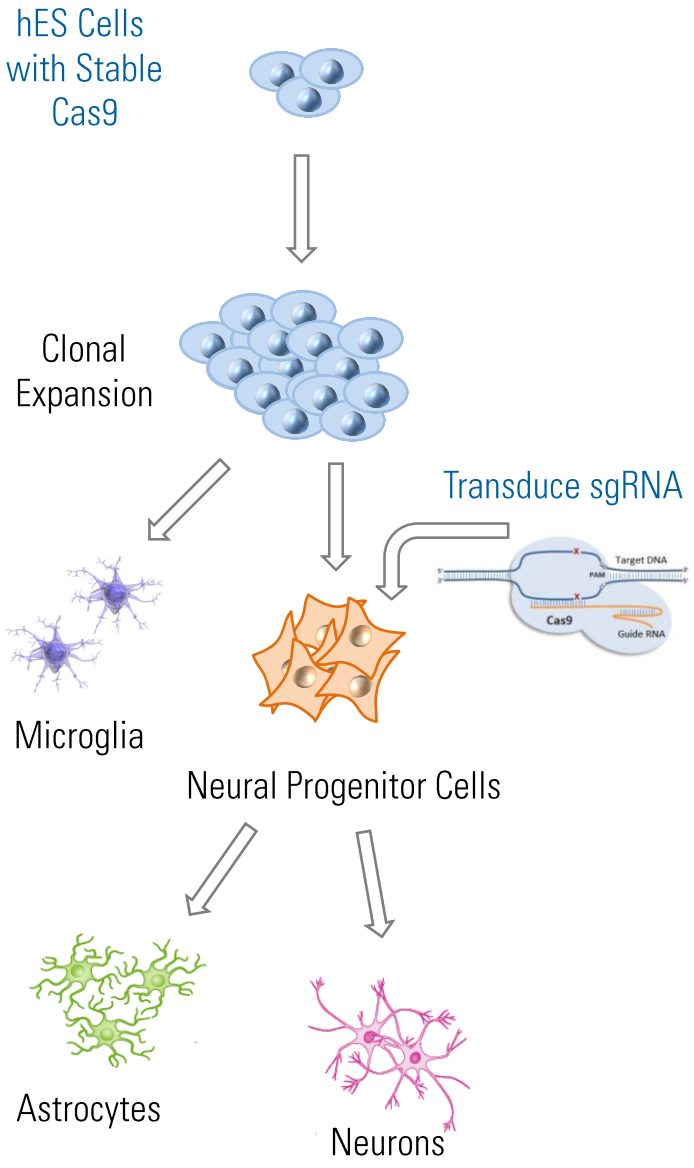
- Committed neurons cluster together in their phosphoproteomic signatures
- P100 signatures are distinct between NPCs, differentiating neurons, and breast cancer cells

MGH 1811
N-GRID
Steve Haggarty, Isaac Kohane, Roy Perlis

Current efforts in establishing neurobiology models

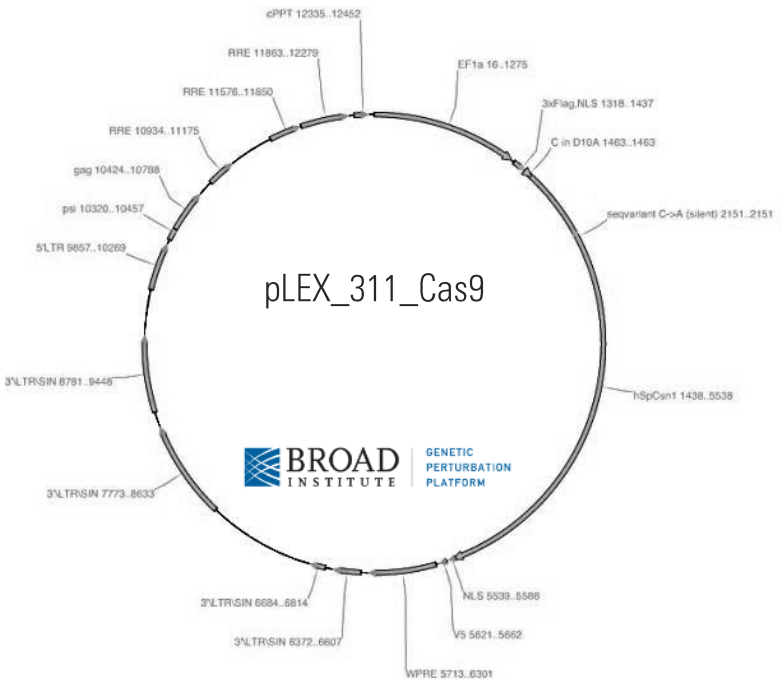
- Establishing a CRISPR/Cas9 neural system
  - H9 Cas9 ES cell lines
  - HUES8 iCas9 ES cell lines

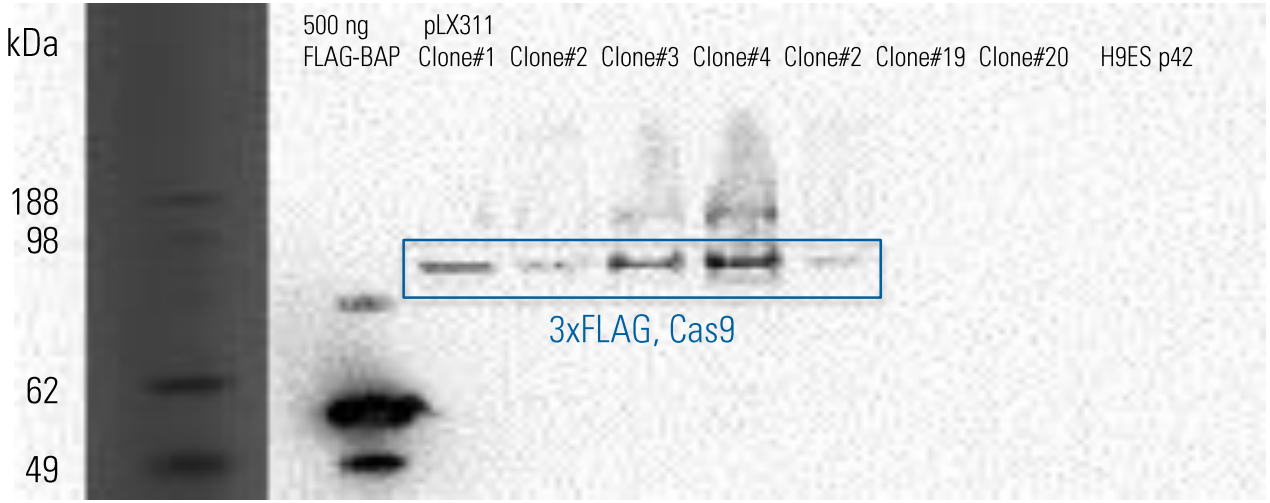
Autism Spectrum Disorder Genes			
ADNP	DYRK1a	MECP2	SEMA5A
ARID1B	EHMT1	NLGN3	SHANK3
CACNA1C	EIF4G1	NLGN4X	SYNGAP1
CNTN4	EPC2	NRXN1	TLK2
CNTNAP2	FMR1	PTCHD1	TSC1
CREBBP	GRIN2B	PTEN	TSC2
CTCF	GRIP1	SCN1A	UBE3a
CTNNB1	GRM5	SCN2a	UBE3b



Production of H9 ESCs with stable Cas9

- H9 embryonic stem (ES) cells transduced with FLAG-tagged Cas9 lentivirus





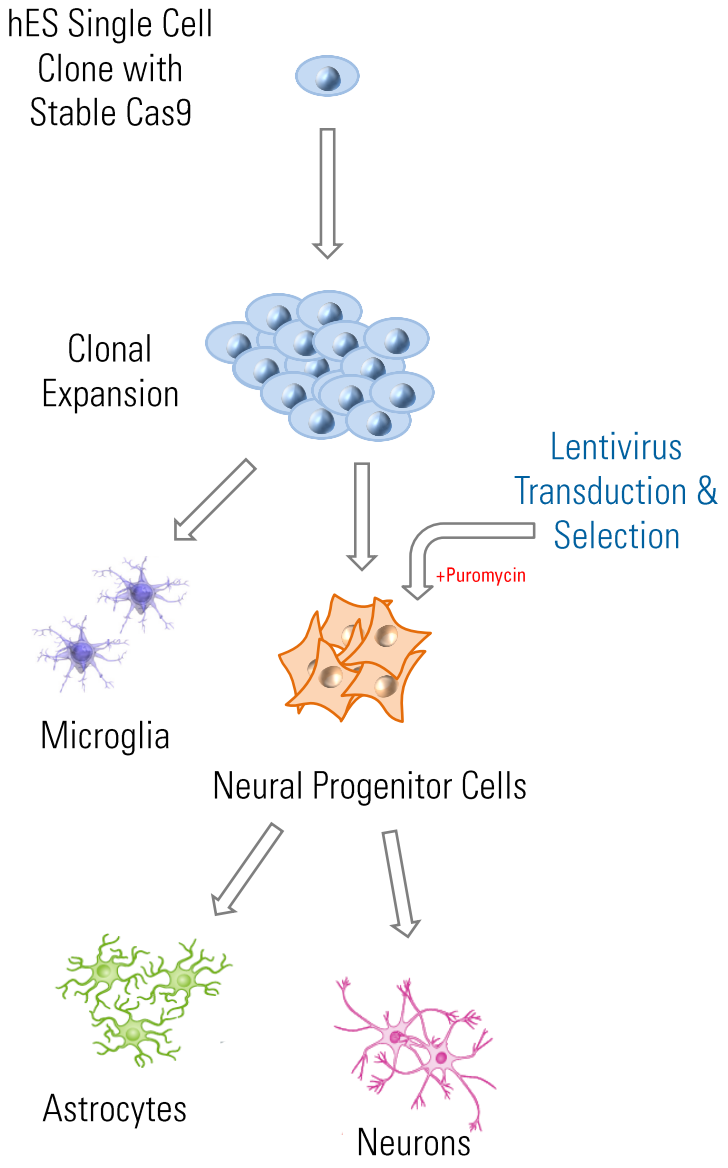
- Dual-Smad inhibition used to differentiate parental H9 ES cells into neural progenitor cells (NPCs)

H9 Cas9 ES-derived NPCs fail validation

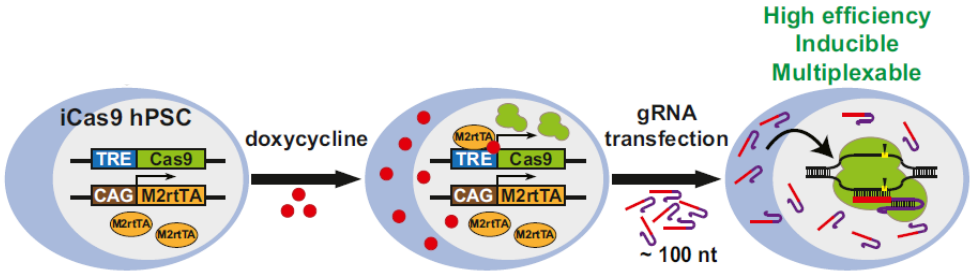
- Establishing a CRISPR/Cas9 neural system
  - H9 Cas9 ES-derived NPCs
    - resistant to lentiviral transduction
    - very sensitive to puromycin (0.2 µg/ml)
    - low CRISPR/Cas9-editing

Criteria	Test
Inducible Cas9 Expression	●
sgRNA Transduction	●
Puromycin Selectability	●
CRISPR/Cas9-Editing	●

● *Passed validation*
● *Failed validation*

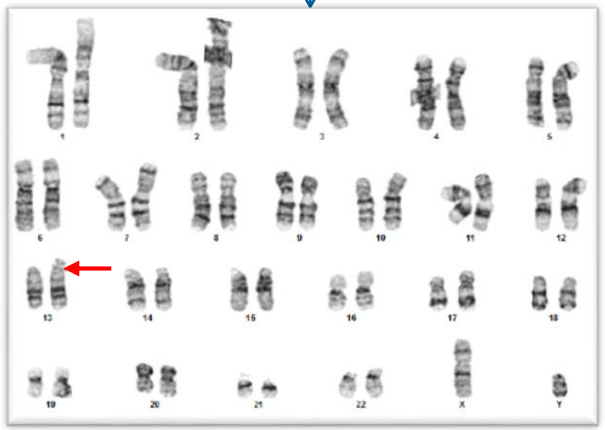


HUES8 with inducible Cas9 are a promising model system

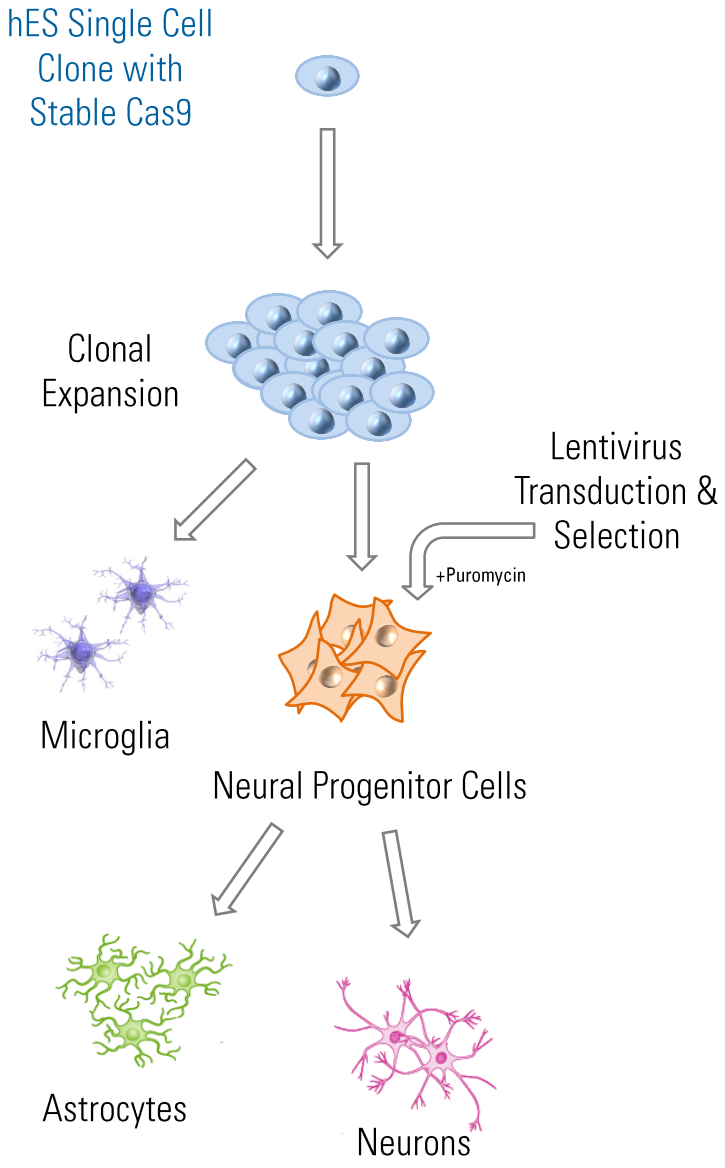


# Metaphase Analyzed	Chromosomes	Considered
17	46, XY	Normal Male
1	92, XXYY	Abnormal/Non-Clonal

Clone #1 *normal karyotype*



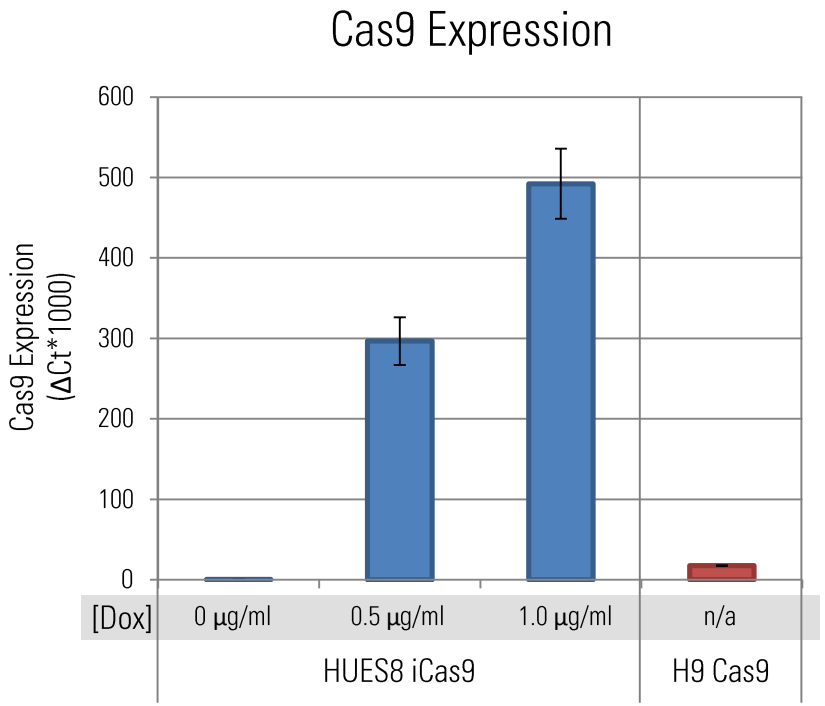
Clone #2 *abnormal karyotype*

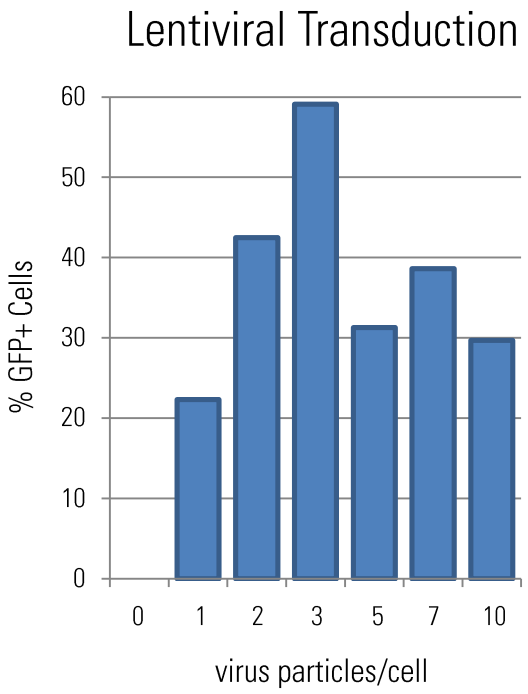


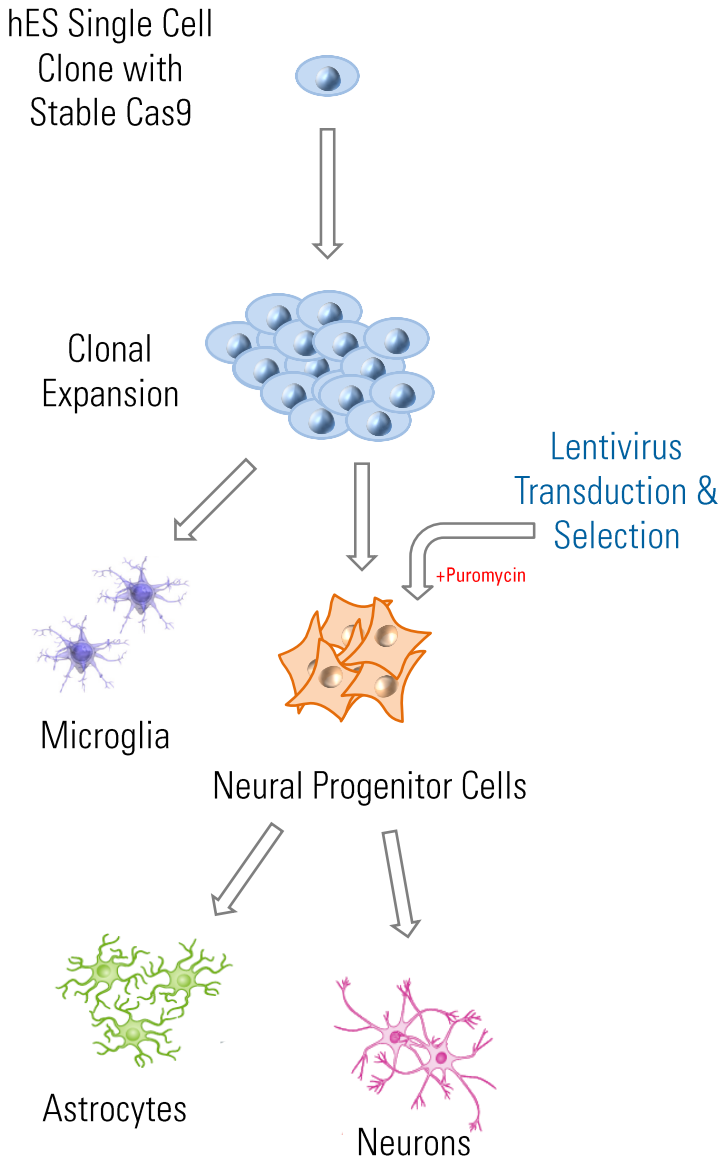
Memorial Sloan-Kettering
Cancer Center
Danwei Huangfu

Optimization of HUES8-iCas9 NPCs for CRISPR/Cas9-editing

- Establishing a CRISPR/Cas9 neural system
  - HUES8-iCas9 NPCs
    - good Cas9 expression in response to doxycycline (Dox)
    - tolerant of lentiviral transduction







Optimization of HUES8-iCas9 NPCs for CRISPR/Cas9-editing

- Induction of Cas9 expression in HUES8-iCas9 NPCs
  - Dox concentration: 1 µg/mL
- Transduction efficiency with lentivirus for sgRNA delivery
  - at given seeding density and lentiviral titer: 3 virus particles/cell
- Sensitivity to puromycin to select for sgRNA transduction
  - puromycin kill curve: 2 µg/mL
- Assessment of CRISPR/Cas9-mediated editing
  - Broad GPP lentiviral Cas9 activity assay

Timeline		Test
Day 0	Seeding NPCs	●
Day 1	sgRNA Transduction	●
	2 days	●
Day 3	Puromycin Selection	●
	…	●
Day ?	Doxycycline → Cas9	●
	…	

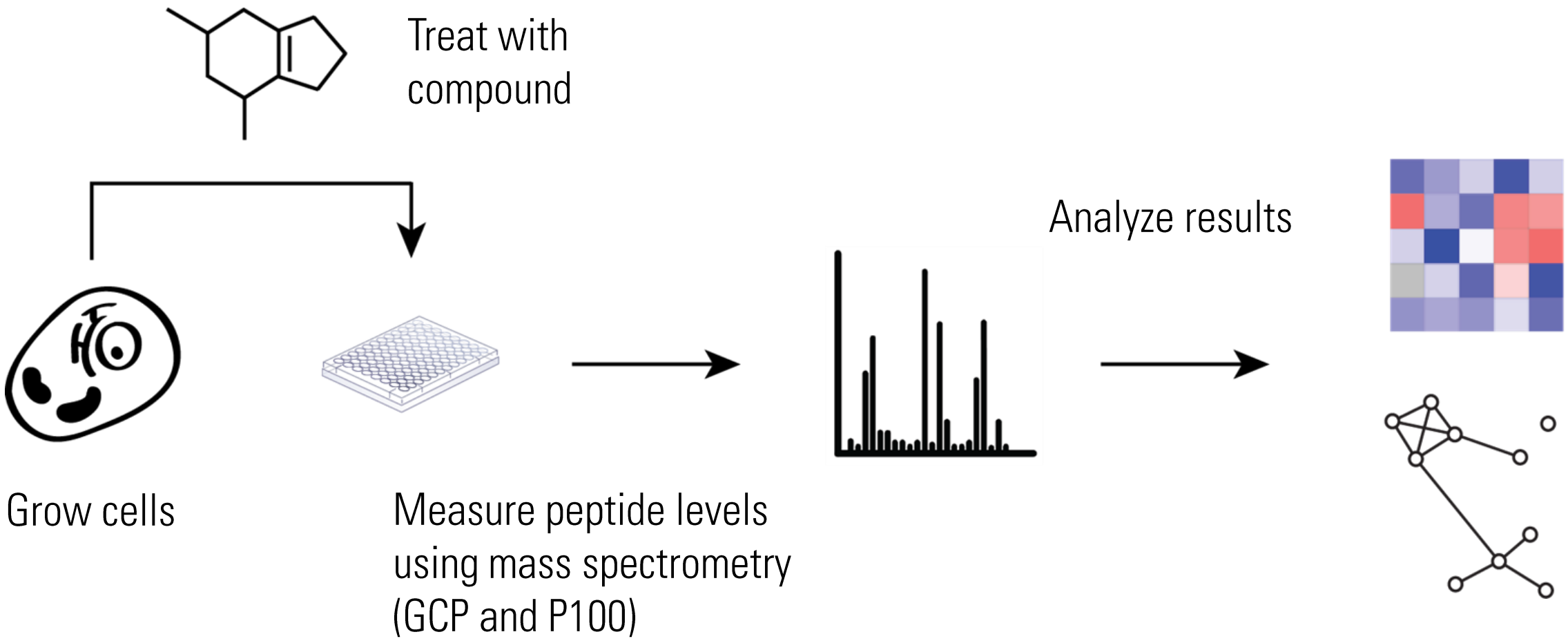
● *Completed*
● *In Progress*

LINCS PCCSE

04 - ANALYSIS

Small-molecule dataset contains >3400 samples

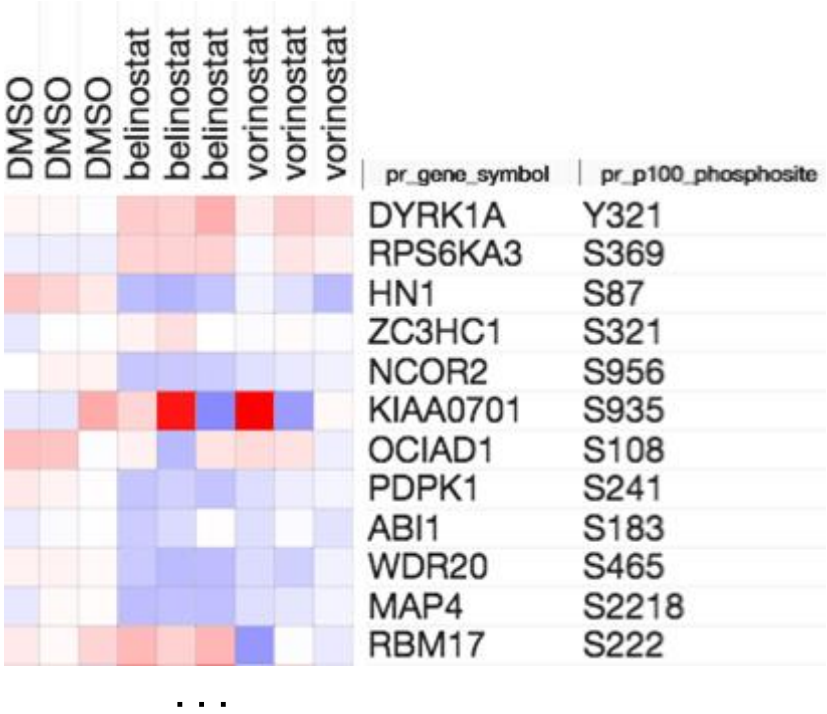
- 2 assays x 6 cell lines x 90 compounds (including controls) x 3 replicates



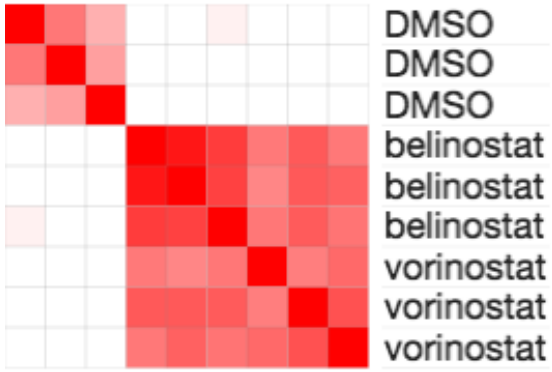
3 ways to look at these data: profiles, similarities, and connectivities

- Later today: data accessibility by Dave Lahr, metadata by Adam Officer

Profiles



Similarity is between profiles



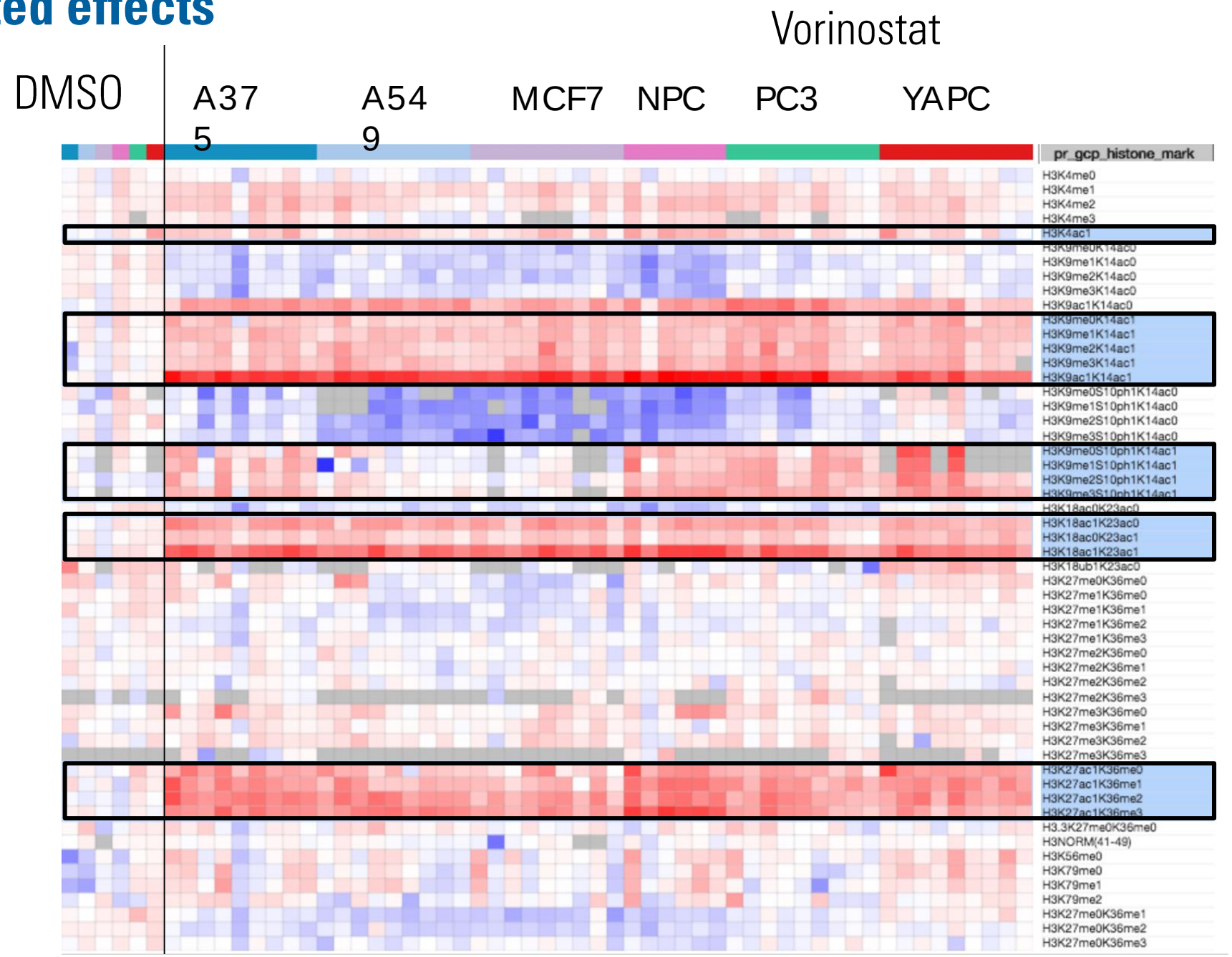
Connectivity is between perturbagens

Proteomics Signature Pipeline (PSP)

GCP profiles show expected effects

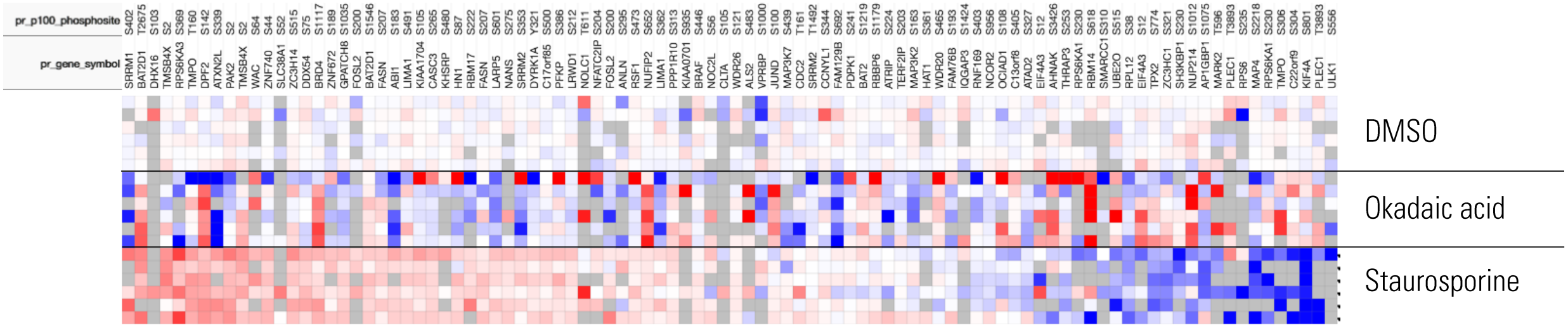
- Histone deacetylase (HDAC) inhibitors cause increased acetylation

Acetylation marks highlighted



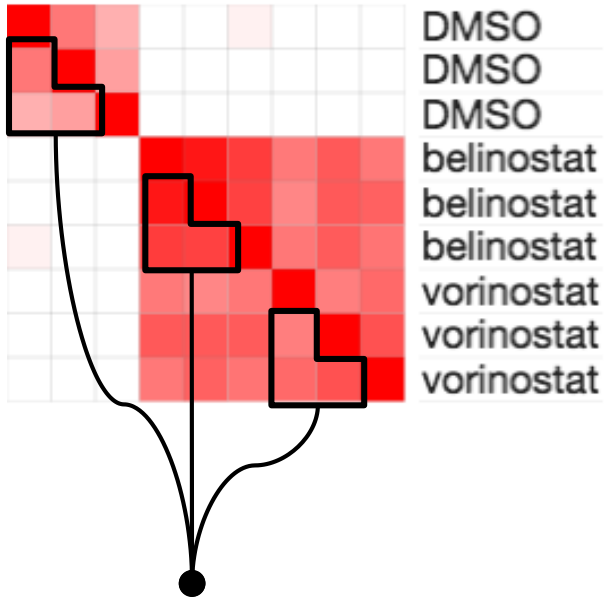
P100 profiles show expected effects

- Staurosporine and okadaic acid have widespread effects

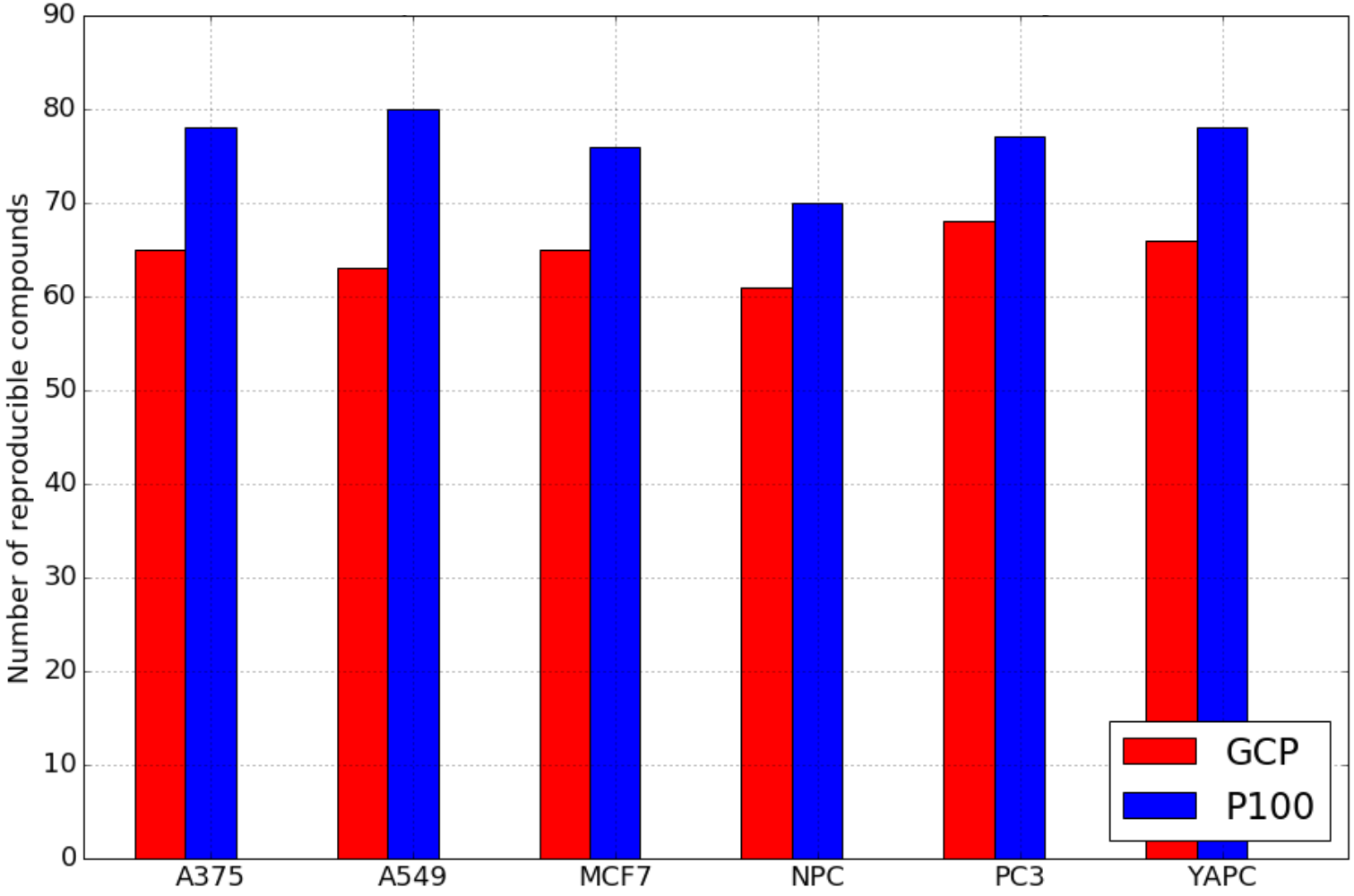


N.B. Matrix transposed

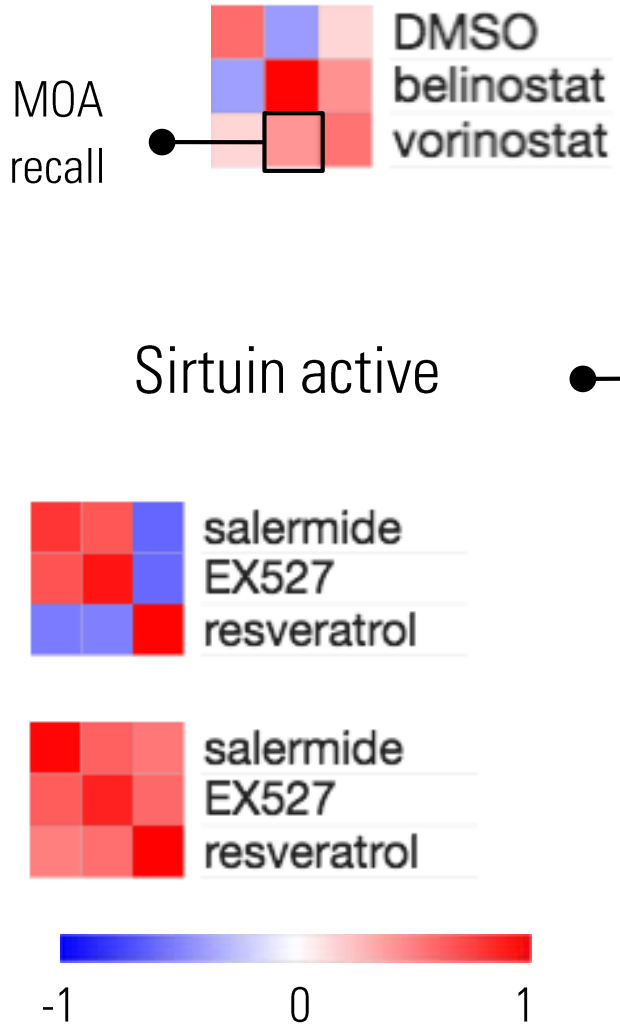
Similarity between biological replicates is high

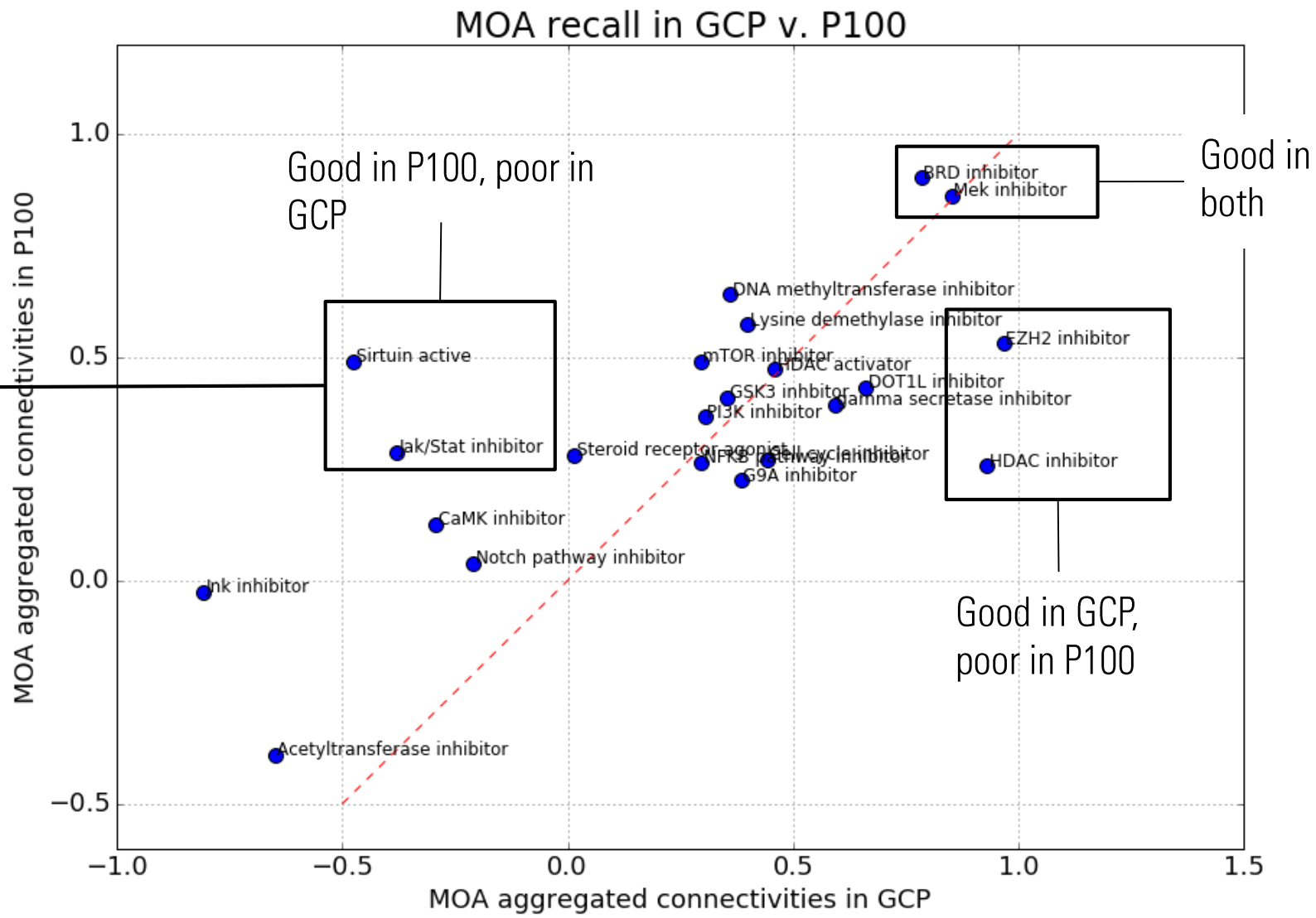


Replicate recall



Looking at connectivity b/w compounds with the same mechanism of action (MOA) helps compare & contrast assays

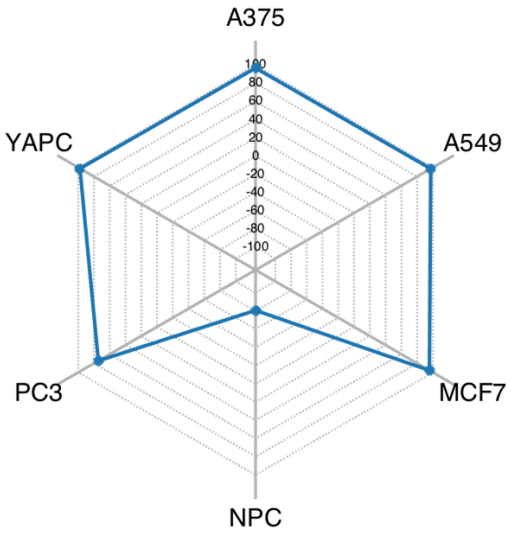




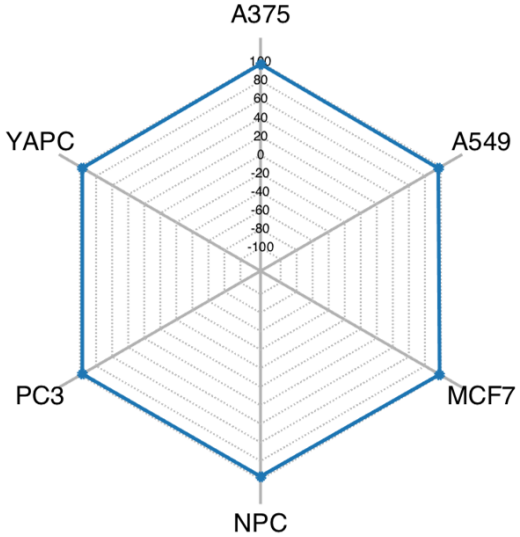
MOA recall can be cell line-specific

- For example, Mek inhibitors (n=2) in GCP

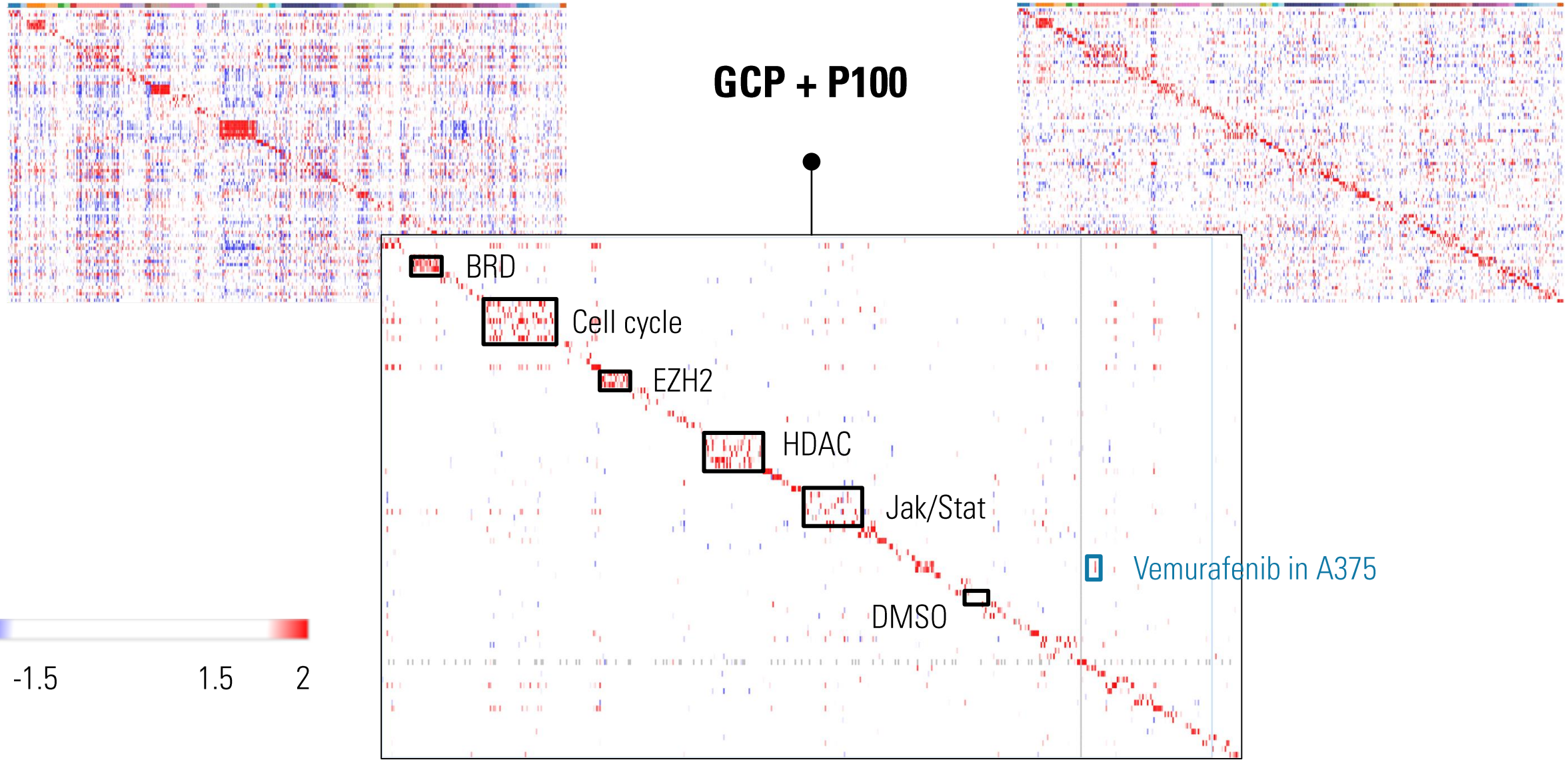
GCP



P100

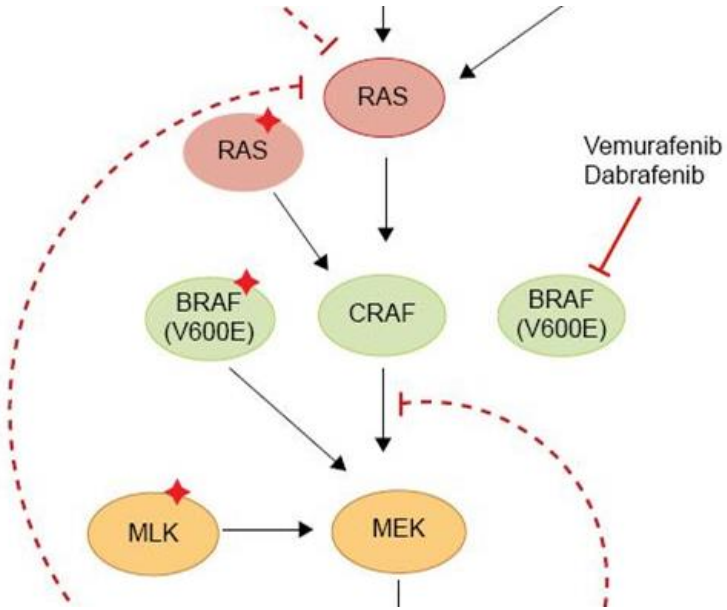


Integrating the 2 assays increases our confidence in connections

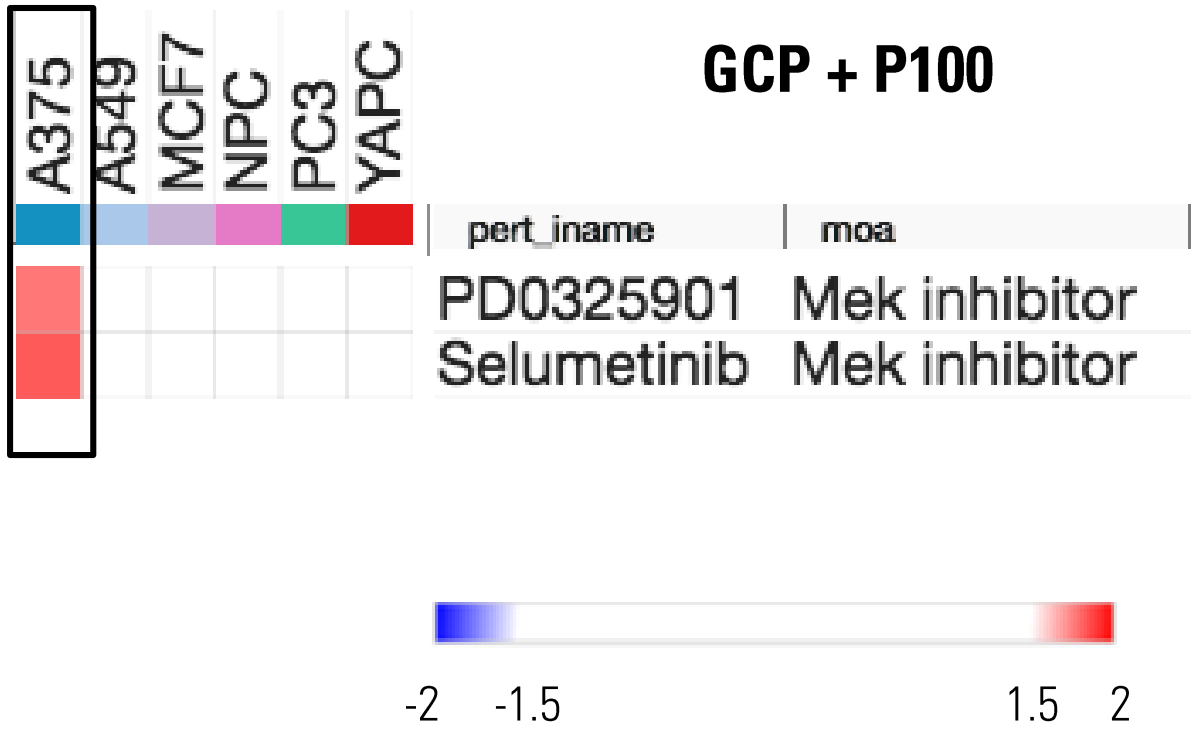


Vemurafenib connects to Mek inhibitors in A375 only

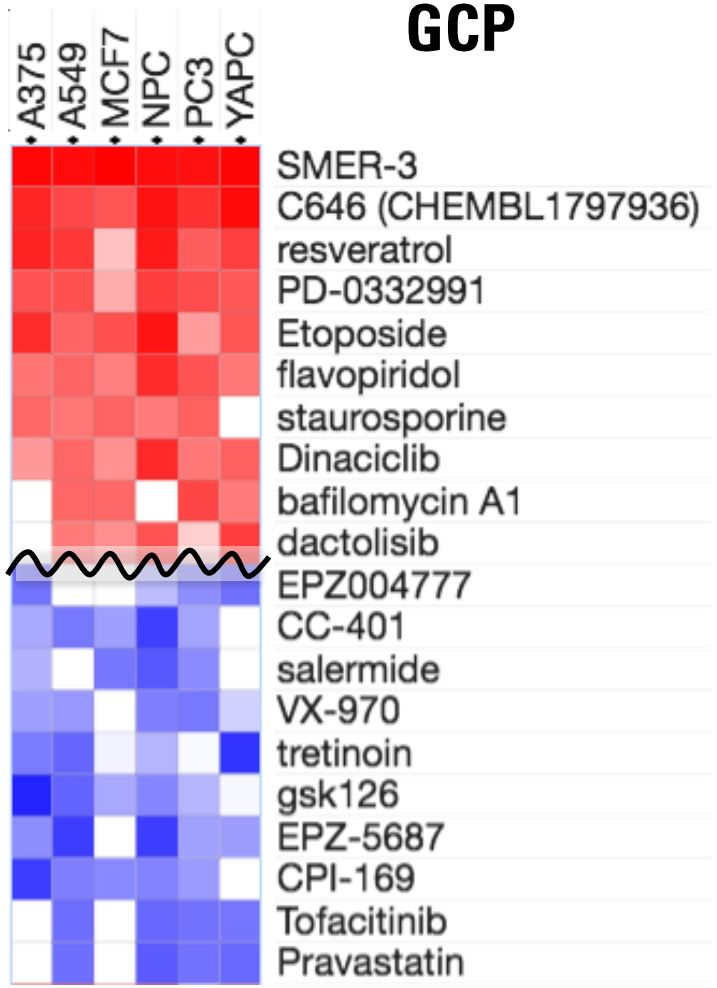
- Only A375 cells have the V600E BRAF mutation that makes them vulnerable to treatment with vemurafenib



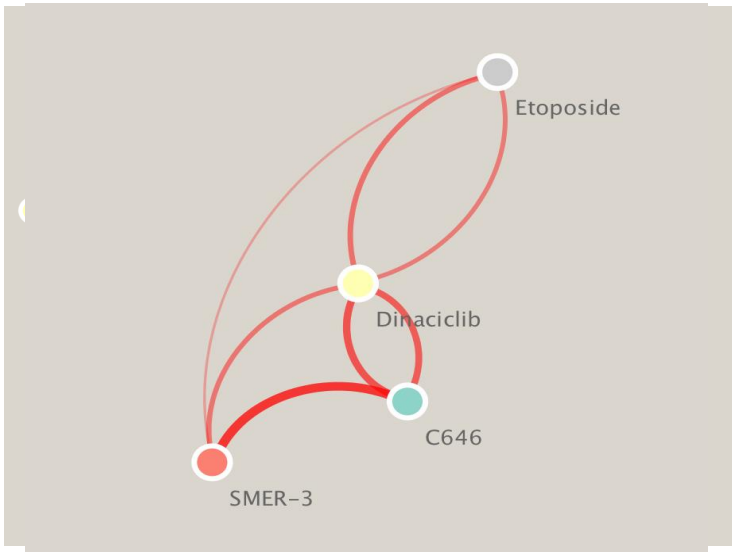
Tran, *Drug Design, Dev., and Therapy*, 2015



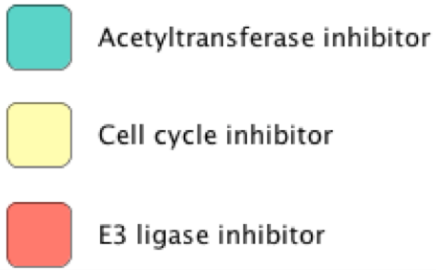
C646, SMER-3, and cell cycle inhibitors consistently connect

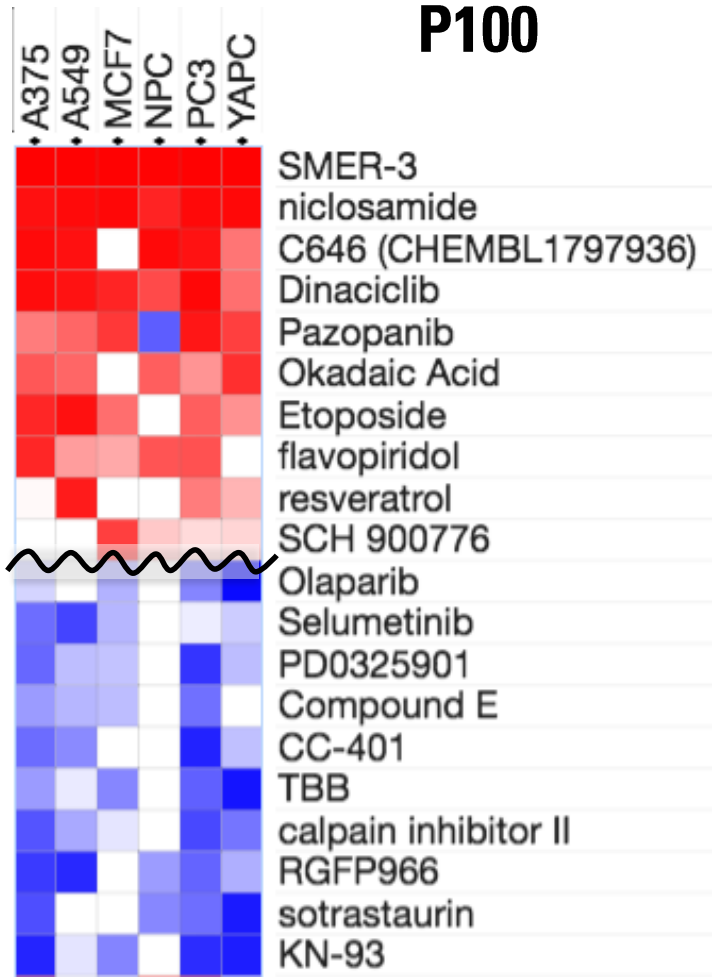


Query = SMER-3



GCP+P100 MEDIAN



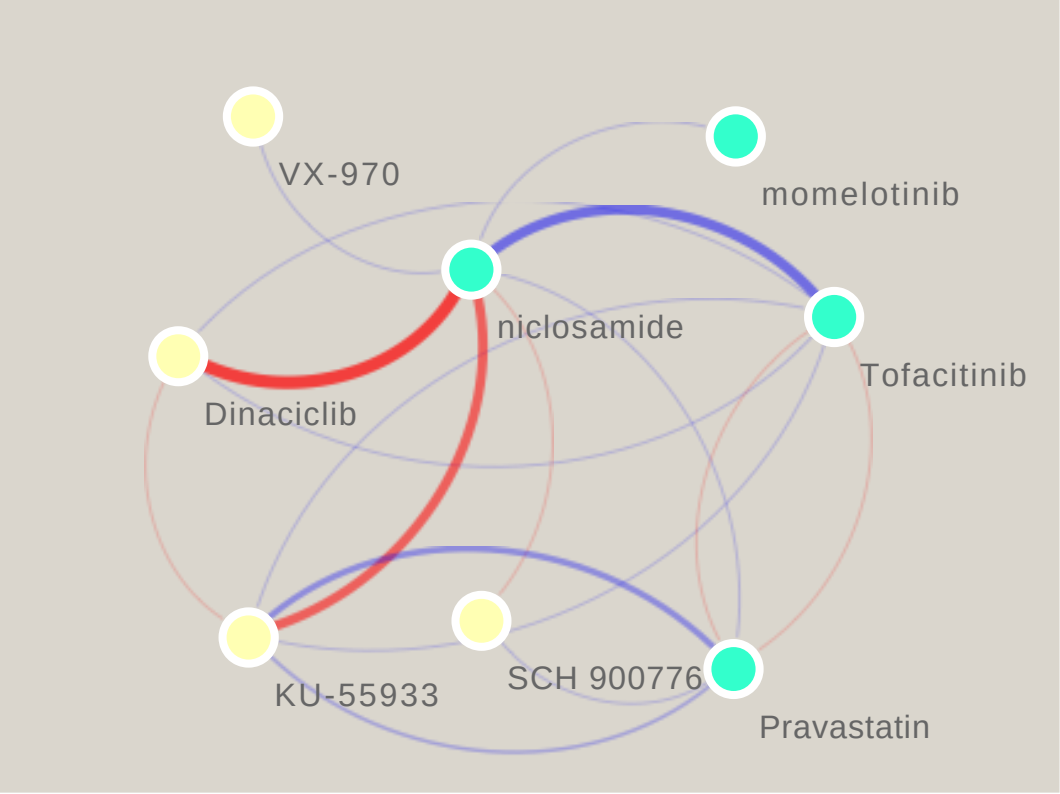


Query = SMER-3

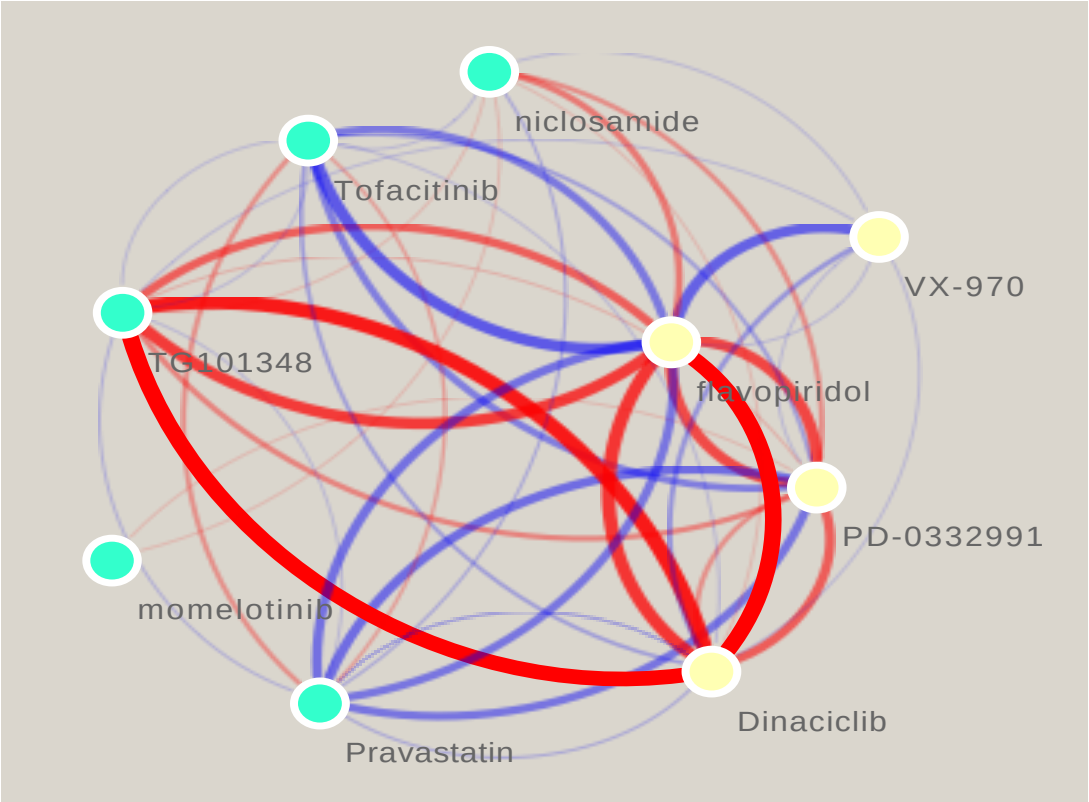
Visualizing networks of connectivities increases the informativity of queries

- Querying niclosamide within a network of connectivities shows that it connects to cell cycle inhibitors instead of its fellow Jak/Stat inhibitors.

P100 PC3

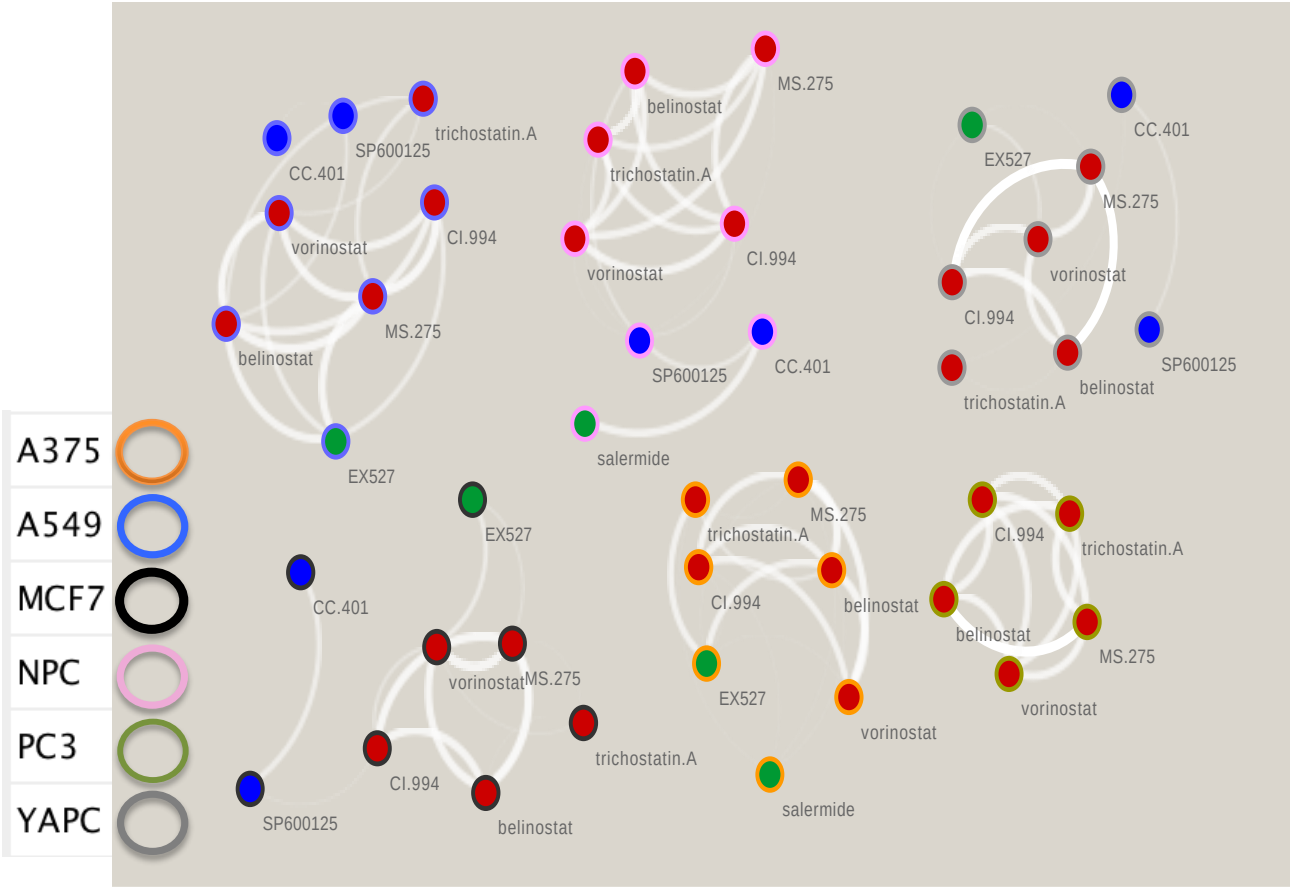


GCP PC3



The network perspective allows for querying of extra-LINCS information

- Connectivites can be combined with GO annotations, MOAs and known gene targets to allow query of biological terms not directly within the scope of LINCS.
- Query *histone deacetylation* in GCP (only positive connections shown)



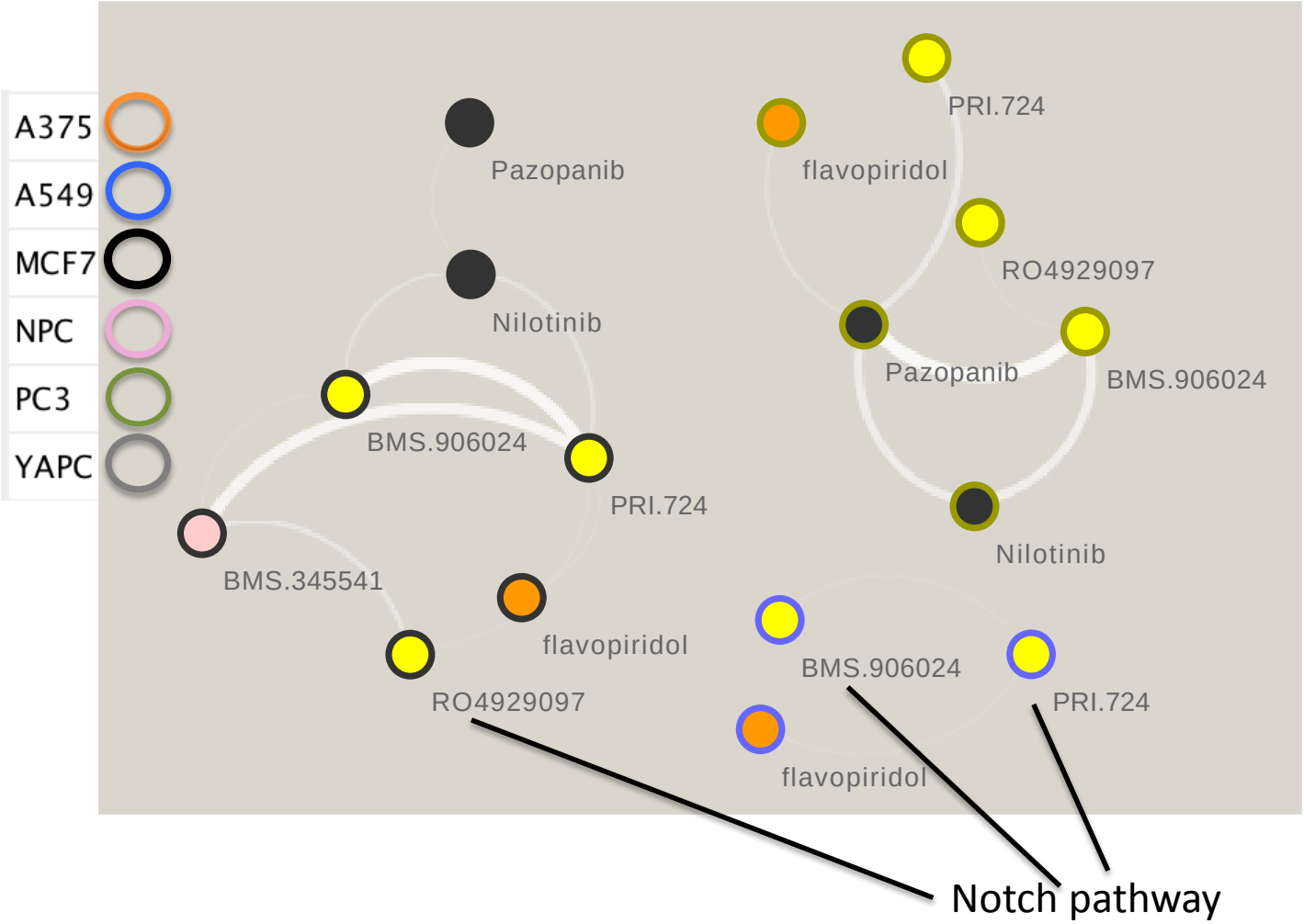
Viewing connectivities within a network allows results to be ranked a la Google:

A375	A549	MCF7	PC3	NPC	YAPC
belinostat	belinostat	MS.275	MS.275	belinostat	MS.275
MS.275	MS.275	vorinosta	belinostat	CI.994	vorinosta
CI.994	vorinosta	belinostat	trich A	vorinosta	CI.994
vorinosta	CI.994	CI.994	CI.994	MS.275	belinostat
trich A	EX.527	trich A	vorinosta	trich A	EX.527

These ranks are derived directly from connectivity scores.

The network perspective allows for querying of extra-LINCS information

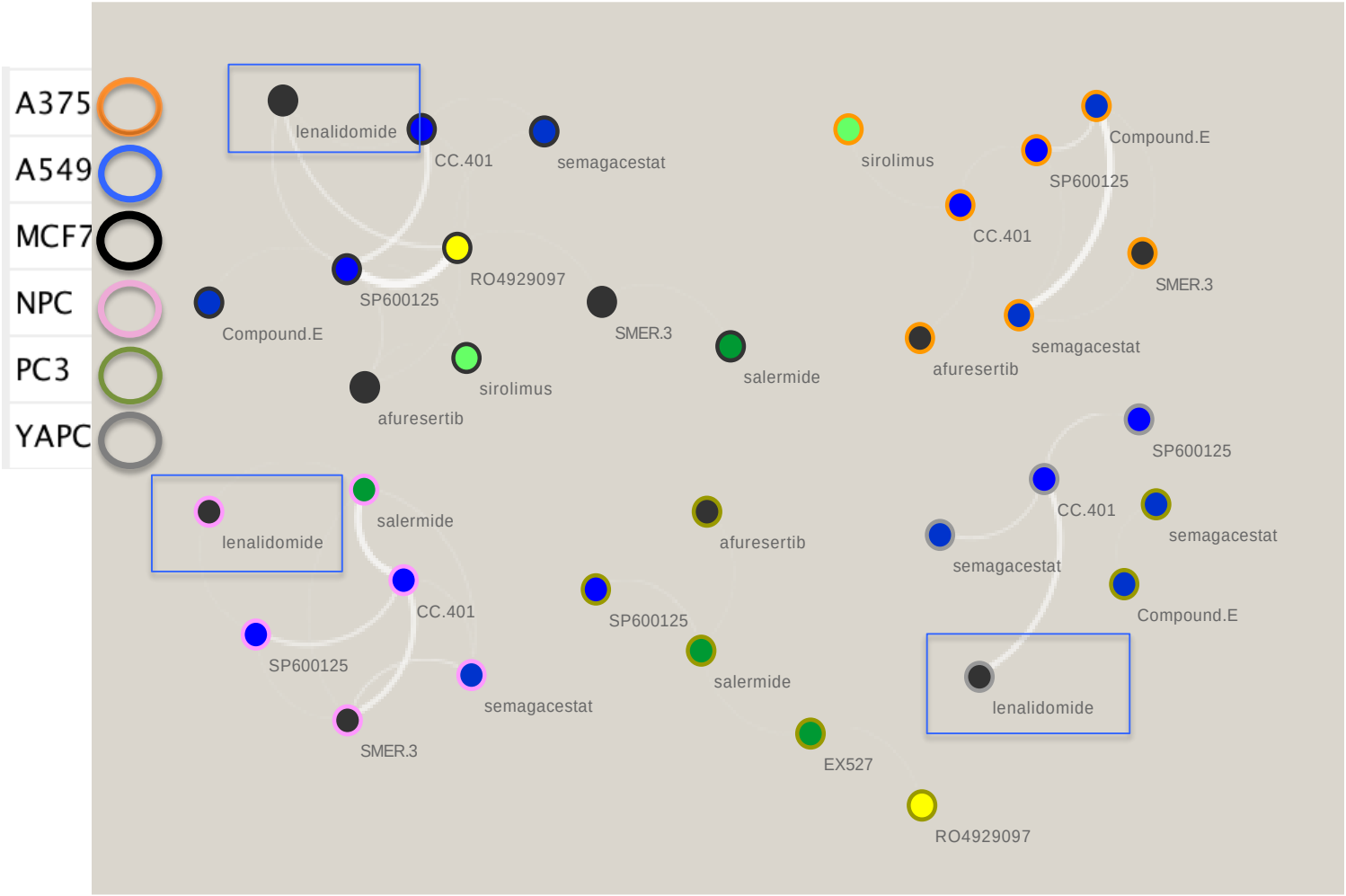
- Query *epithelial cell proliferation* in P100



Of eight profiled compounds whose gene targets relate to epithelial cell proliferation, Notch pathway inhibitors are the only MOA class to have multiple members appear with significant connections.

The network perspective allows for querying of extra-LINCS information

- Query *protein ubiquitination* in GCP



Lenalidomide, which has recently been shown to modulate ubiquitin pathways, appears consistently .

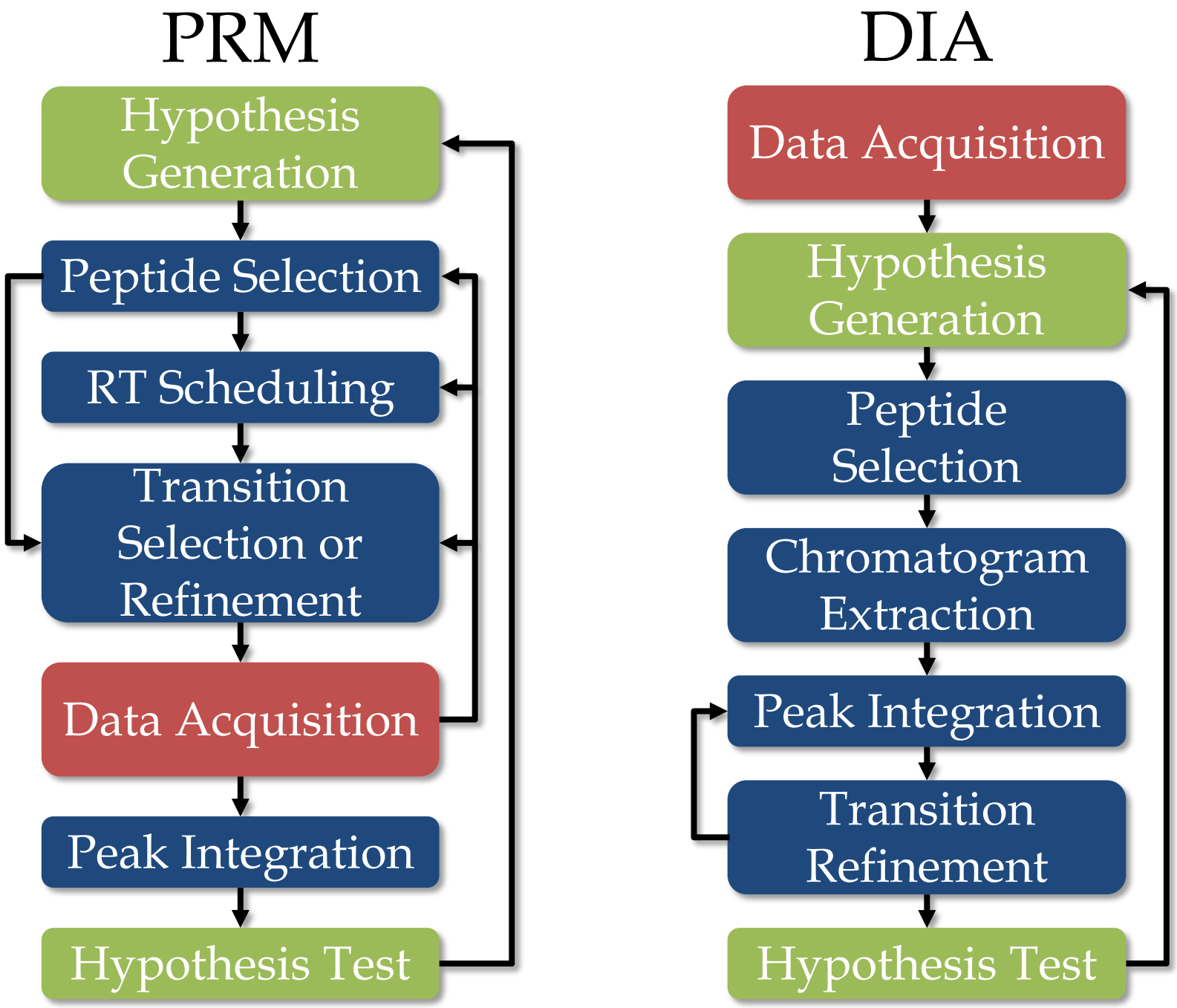
Conclusions

- Analysis of LINCS proteomics data reveals both known and potentially novel biology, as well as robustness of the assays
- Connections between perturbations take on new dimensions of meaning when viewed through the lenses of multiple assays
- Visualization and interpretation of networks of connectivities opens the door to Google-type searching of LINCS proteomics data by non-LINCS queries

LINCS PCCSE

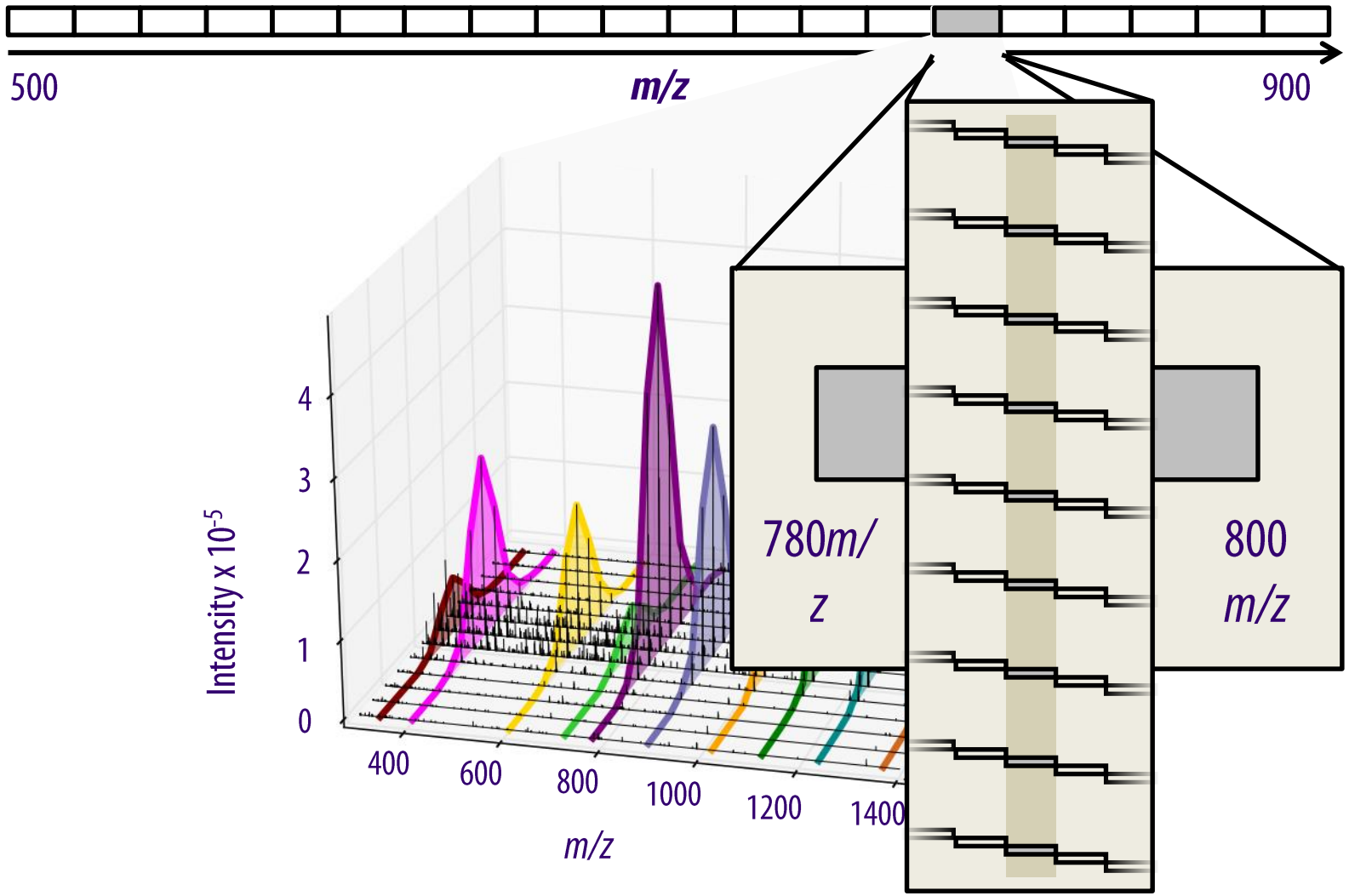
05 – DATA INDEPENDENT ACQUISITION (DIA)

Why DIA?

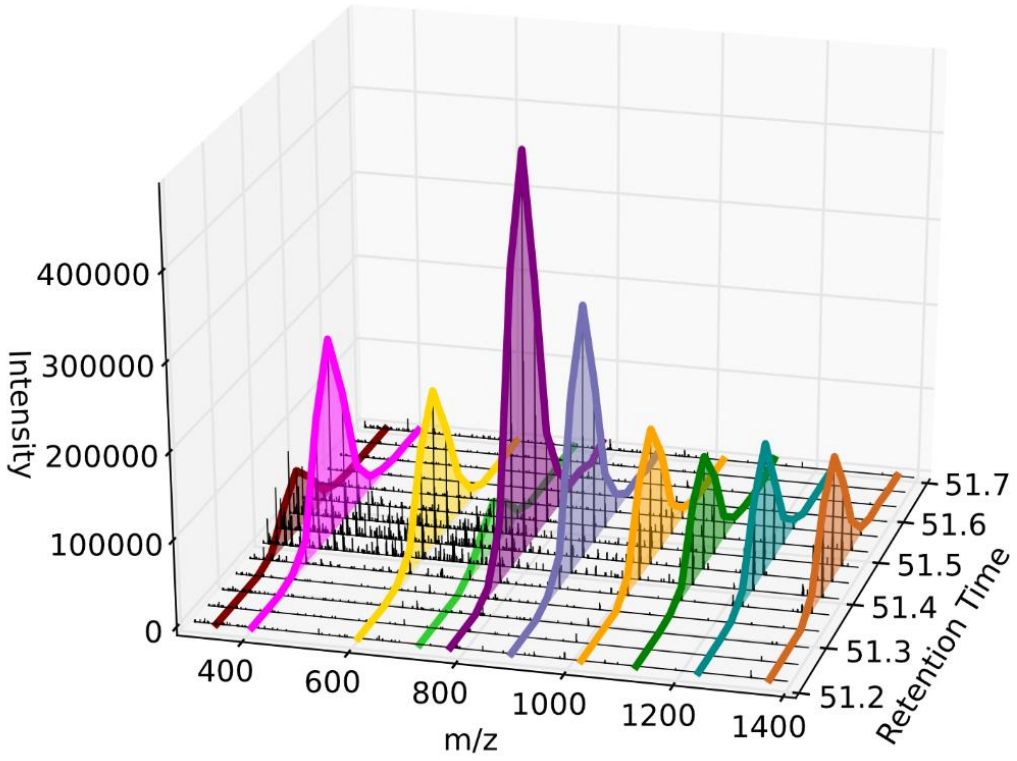

PRM
Hypothesis Generation
Peptide Selection
RT Scheduling
Transition Selection or Refinement
Data Acquisition
Peak Integration
Hypothesis Test
DIA
Data Acquisition
Hypothesis Generation
Peptide Selection
Chromatogram Extraction
Peak Integration
Transition Refinement
Hypothesis Test

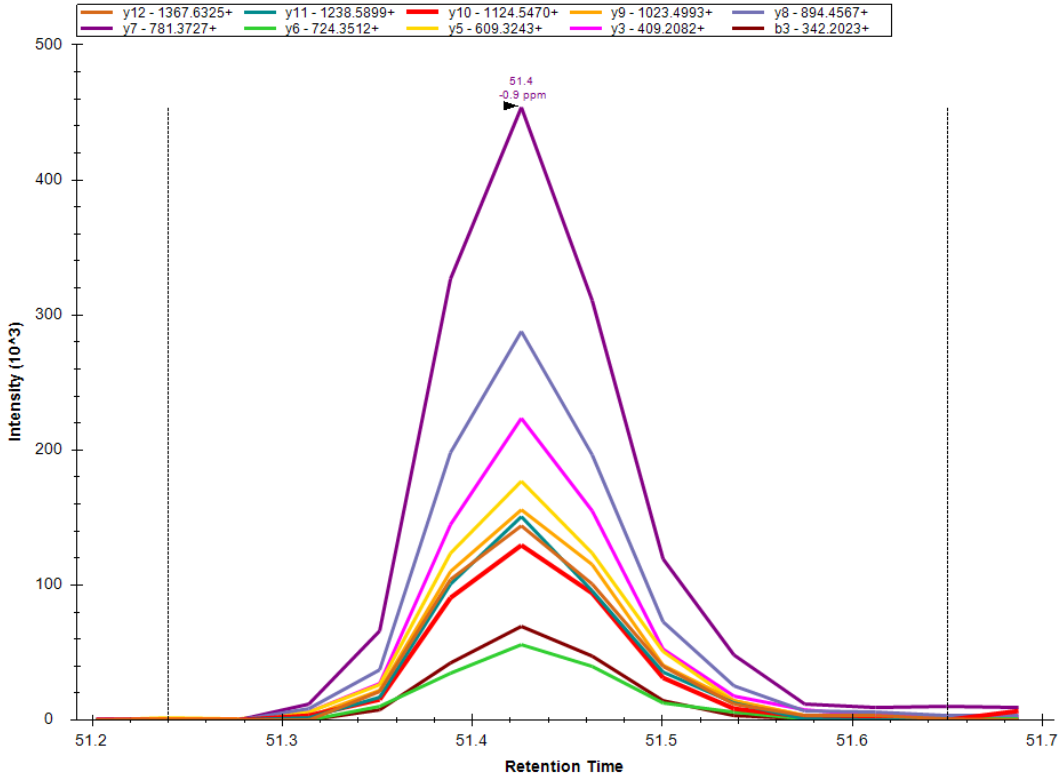
Data Independent Acquisition – targeted chromatogram extraction

VLENTFEIGSDSIFDK++ (790.4 *m/z*)



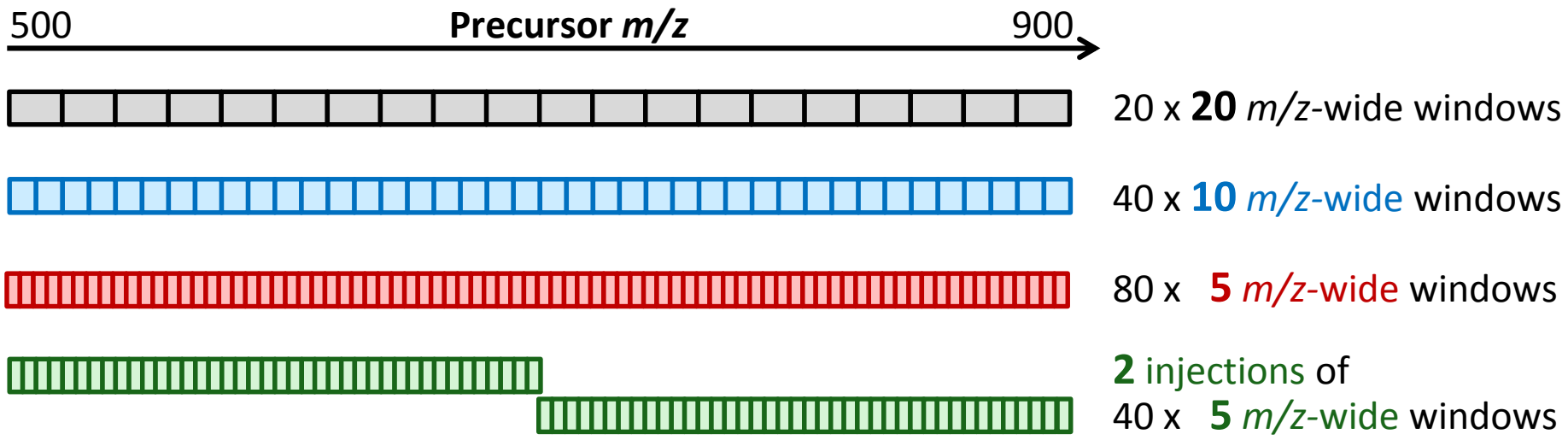
Extraction of fragment ions from DIA data



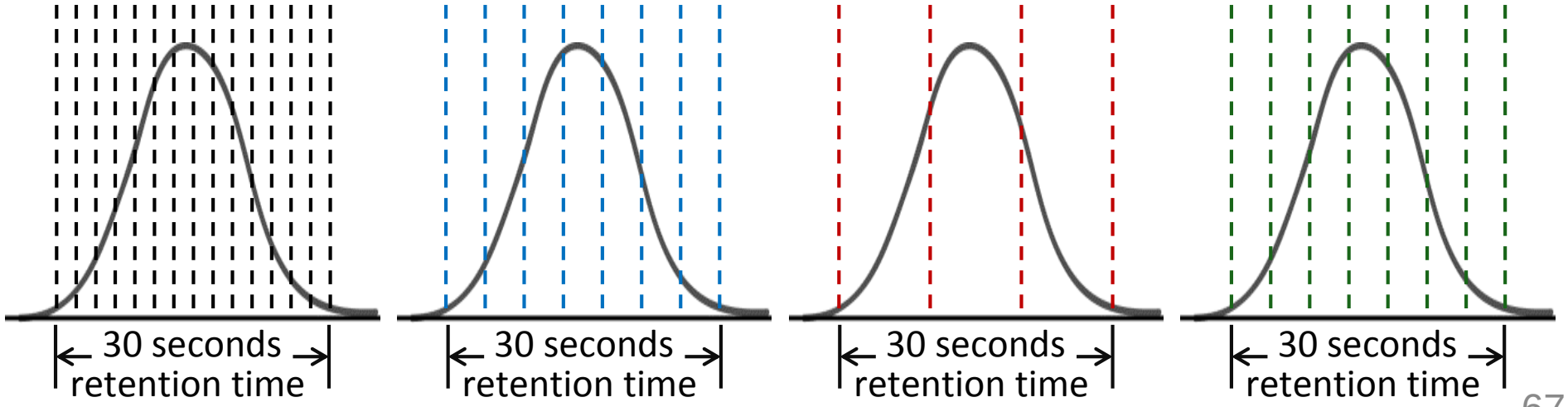


Venable et al Nat Methods 2004
Dong et al Science 2007

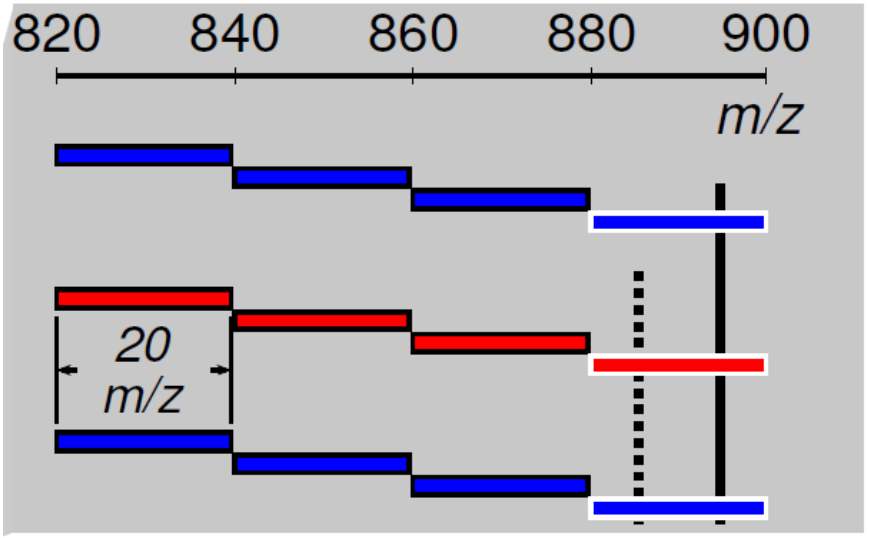
DIA is all about balance (and sacrifice)

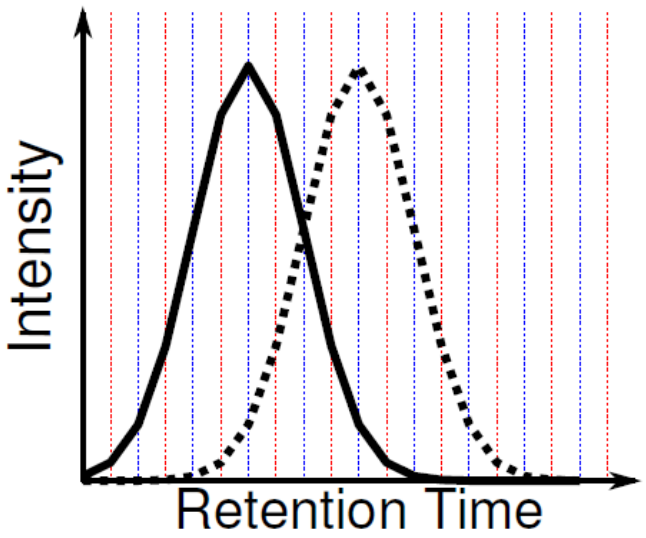


Instrument MS/MS scan speed: 10 Hz



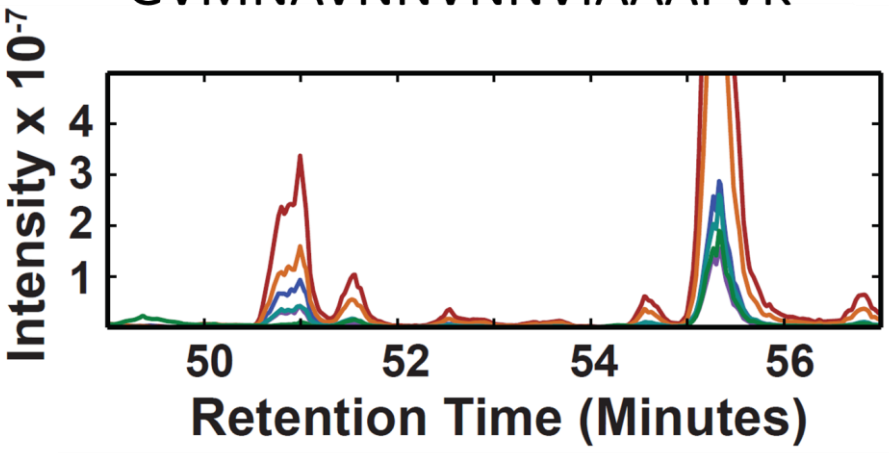
Improving precursor selectivity with overlapping DIA windows

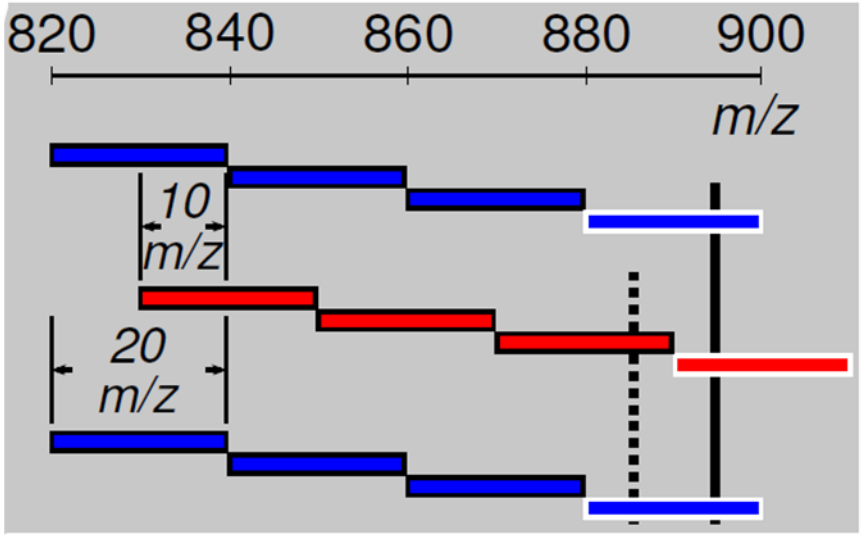


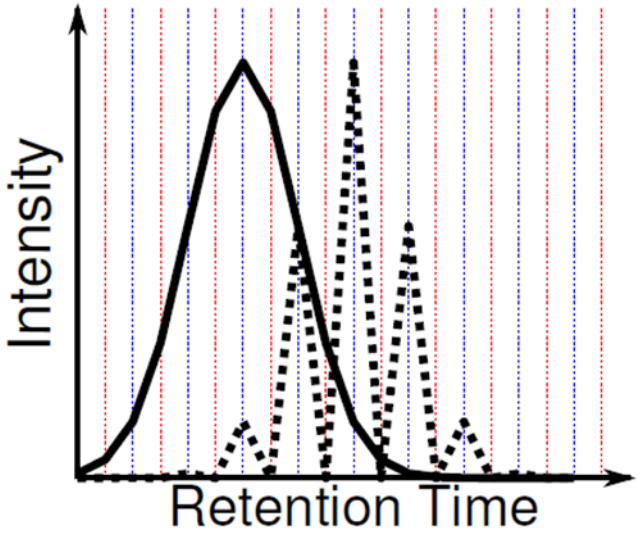


20 *m/z* DIA

GVMNAVNNVNNVIAAAFVK

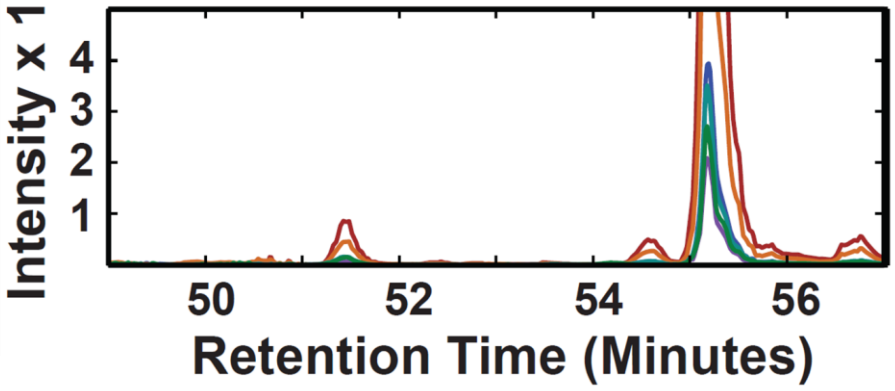






20 *m/z* DIA + Overlap + Demultiplexing

GVMNAVNNVNNVIAAAFVK



LINCS PCCSE

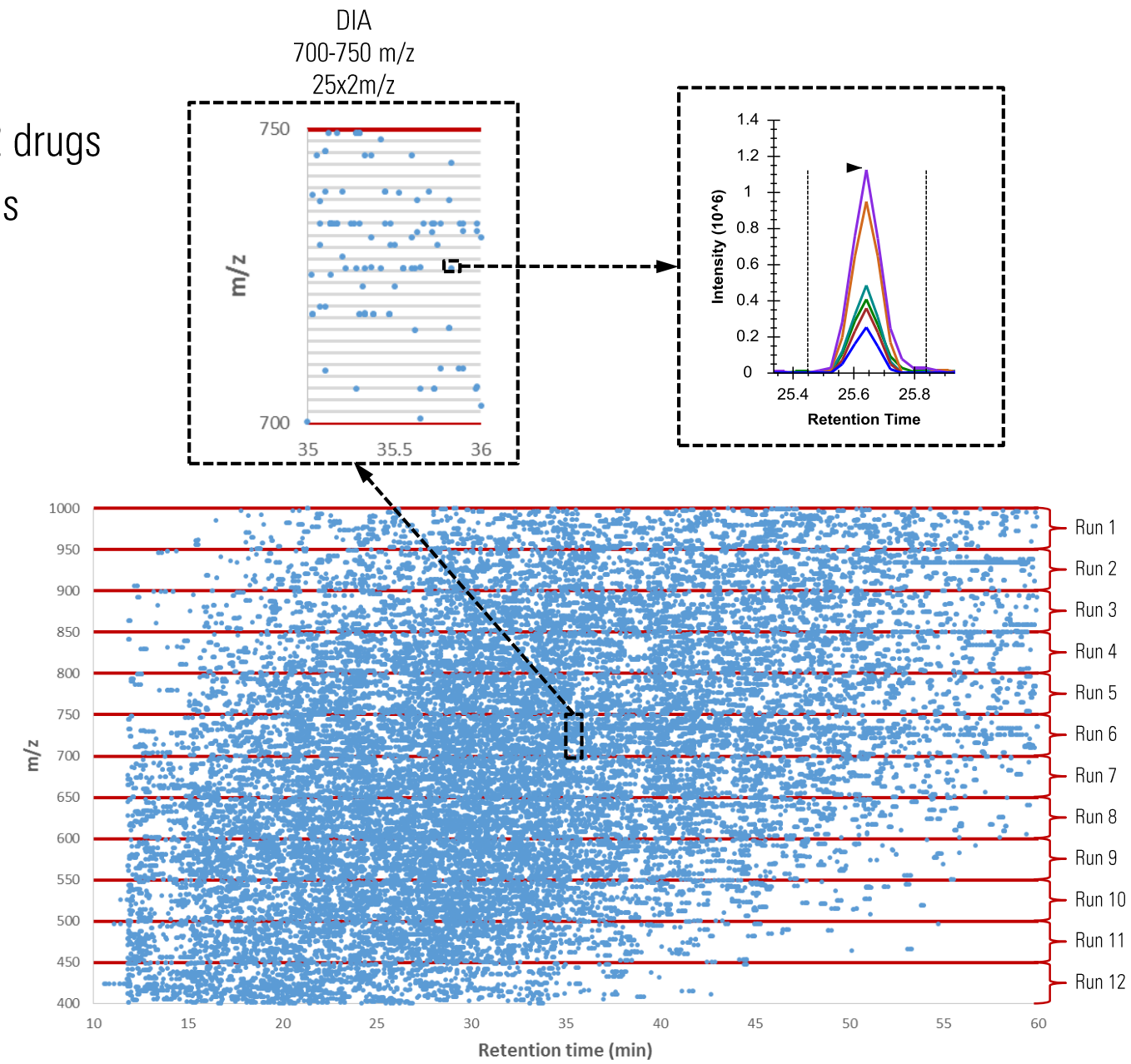
BUILDING THE PHOSPHO-SUPERLIBRARY

Comprehensive spectral/chromatographic libraries are needed to improve DIA analysis

- Spectrum-centric DIA data analysis:
  - Needs spectral/chromatographic library
  - High-quality libraries are necessary
- DDA-based spectral libraries are affected by the semi-stochastic nature of the parent ion selection
  - Spectra used for peptide identification might not be informative enough for peptide quantification
  - DDA produces incomplete datasets
- Comprehensive analysis of proteomes by DIA
  - DIA promises to acquire a signal for all peptides above the instrument's LOD
- Challenges:
  - Design the right DIA method for the spectral/chromatographic library
  - Data-treatment workflow for DIA data

Narrow-window DIA enables a comprehensive proteome analysis

- Sample:
  - Pool of 32 samples (PC-3 cells) treated with 32 drugs
  - Includes P100 heavy-labelled standard peptides
- MS Method:
  - Thermo Fisher Orbitrap Fusion Lumos
  - DIA method:
    - 12 DIA LC-MS runs each covering only 50 m/z
    - Narrow-window DIA: 25 x 2m/z
    - Cycle time <3 sec



High-quality data enabled by adapted quantitative proteomics tools

Narrow-window DIA

- 12 LC-MS runs
- 25 x 2m/z windows

→

SpectrumMIII

- DDA-like database search
- precursor ions: ± 1 m/z
- product ions : ± 10 ppm

→

Percolator

- Semi-supervised machine learning
- User-defined features
- FDR 1% at PSM level (q-values<0.01)

→

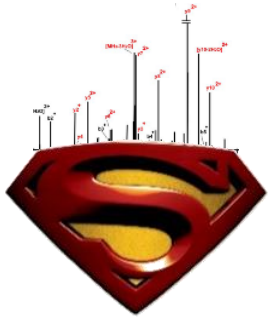
SpectrumMIII

- Export PepXML

Skyline

- Generate spectral library

→



Super Spectral Library

P100 DIA runs

→

EncyclopeDIA

- Peak group identification
- Retention time alignment
- Remove interfered transitions

→

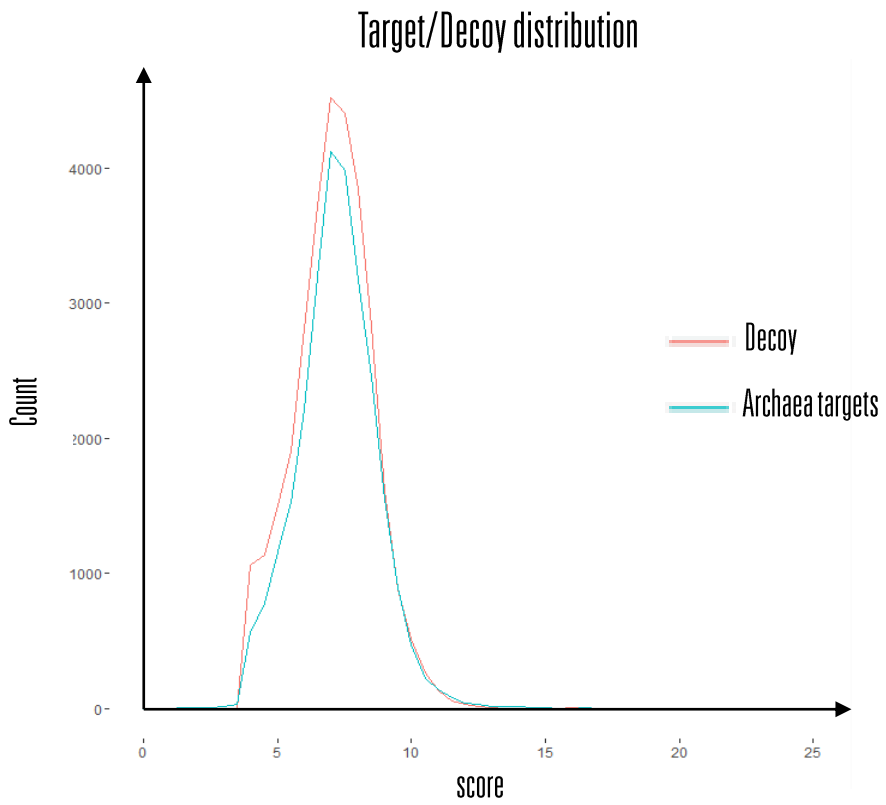
Skyline

- Filtering data
- Peptide quantification

A “Decoy-Decoy” approach proves the low false discovery rate of the strategy

- Searching for what is not there
  - Sample: Human cell line (Jurkat) whole digest
  - Database search: Archaea database (~19000 proteins)
  - Validation using Percolator
- No hits should be found in these conditions

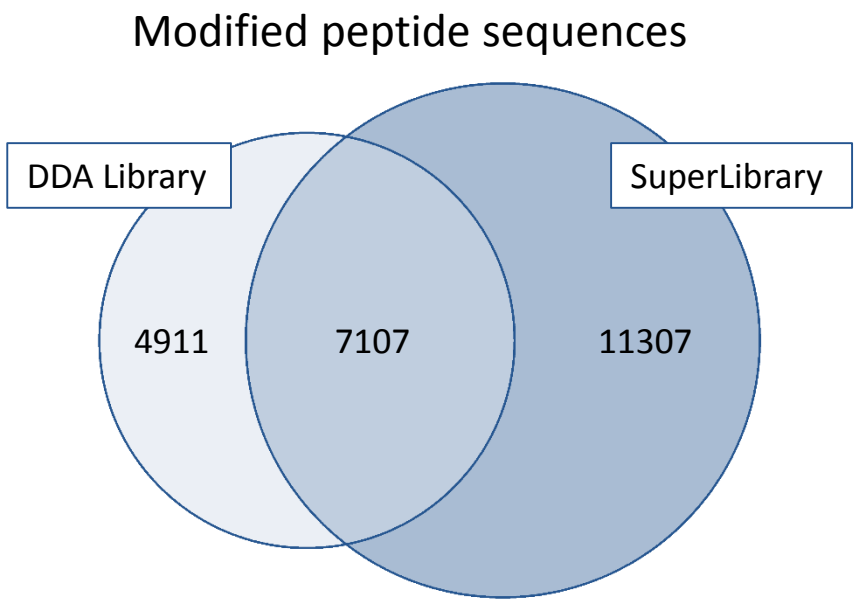
- The SpectrumMill + Percolator workflow produces a very low false discovery rate
- Extremely conservative



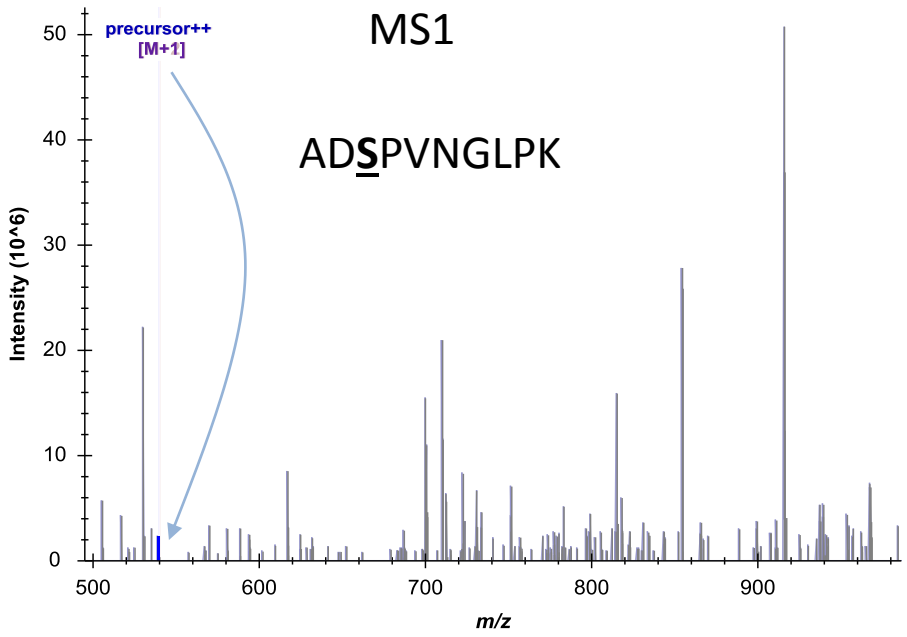
- Only 4 peptides were identified
- All shared peptides between the archaea and the human proteome

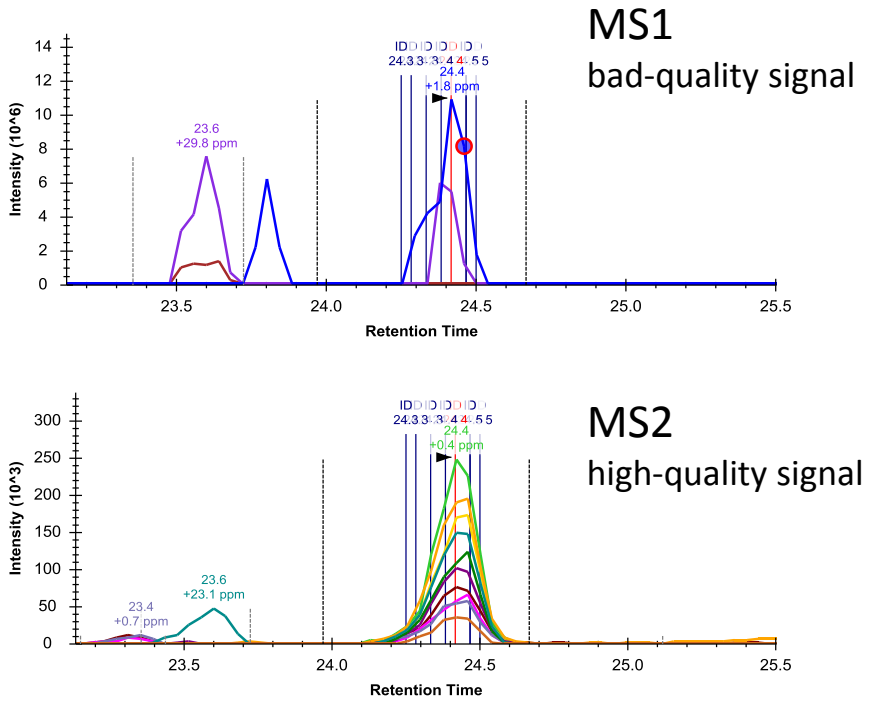
A “Super Spectral Library” more than doubles potential signaling analytes

- Sample:
  - Pool of 32 samples (PC-3 cells) treated with 32 drugs
  - Includes P100 heavy-labeled standard peptides
- Identification of:
  - 18408 modified peptides sequences (3x additional IDs compared to DDA)
  - 20570 precursor ions (3x additional IDs compared to DDA)
  - 14363 Phosphopeptides (+49% compared to DDA)
  - 4968 Localized phosphosites (+29% compared to DDA)



The high selectivity and untargeted acquisition of narrow-window DIA increases sensitivity

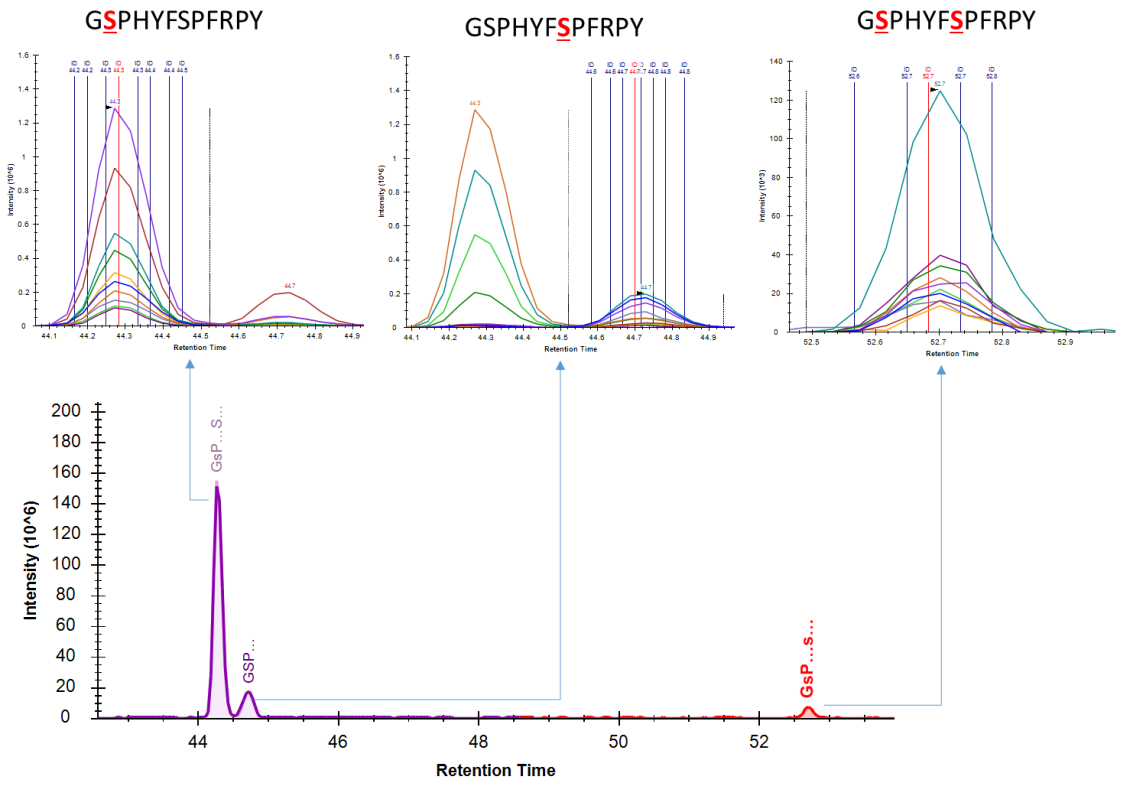




- This peptide would be difficult (not possible) to identify with DDA
- The narrow-window DIA enables its identification

Narrow-window DIA improves quantification of phosphorylated isomers

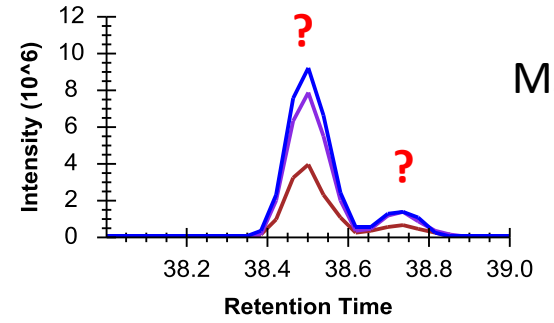
In the super library



Confident localization of phosphorylation sites

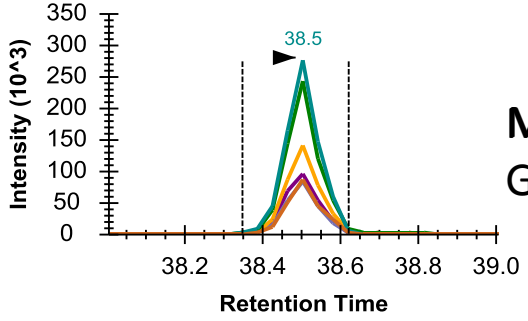
- Retention times
- Fragmentation patterns
- Relative intensities
- Well-responding specific-transitions

In the queried data

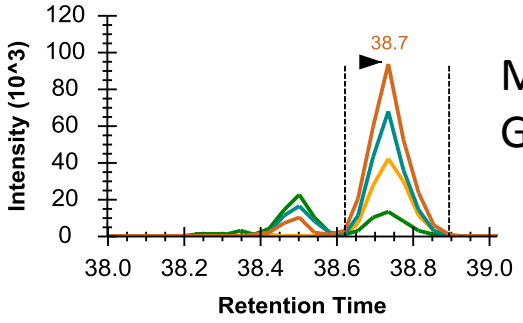


MS1

GSPHYFSPFRPY
or
GSPHYFSPFRPY
?



MS2
GSPHYFSPFRPY



MS2
GSPHYFSPFRPY

Chromatograms of site-specific fragment ions enable the characterization of phosphorylation positions

Our strategy significantly increases signaling pathway coverage

Highly conservative filtering criteria

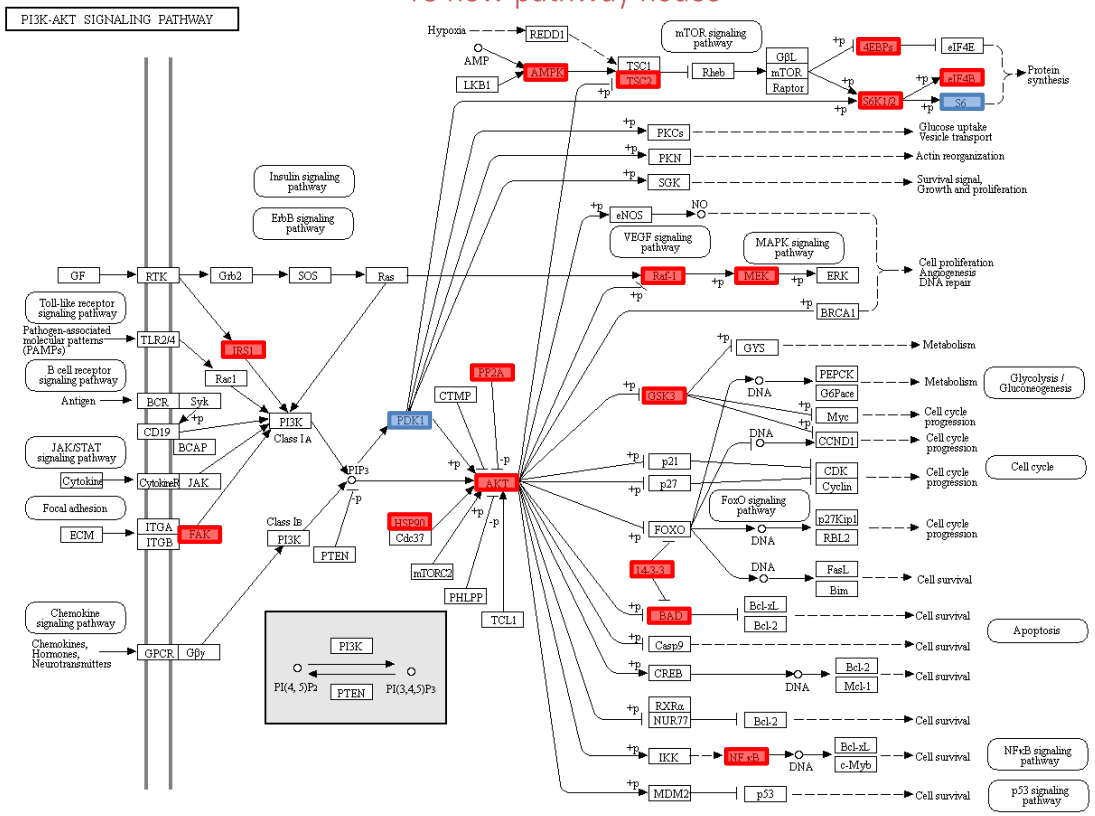
P100 samples acquired in DIA → Queried: Super library (20570 precursors)
Detected: 11341 candidate precursors
Quantified: 2841 peptides / 3539 precursors

Top of the basket filtered peptides:

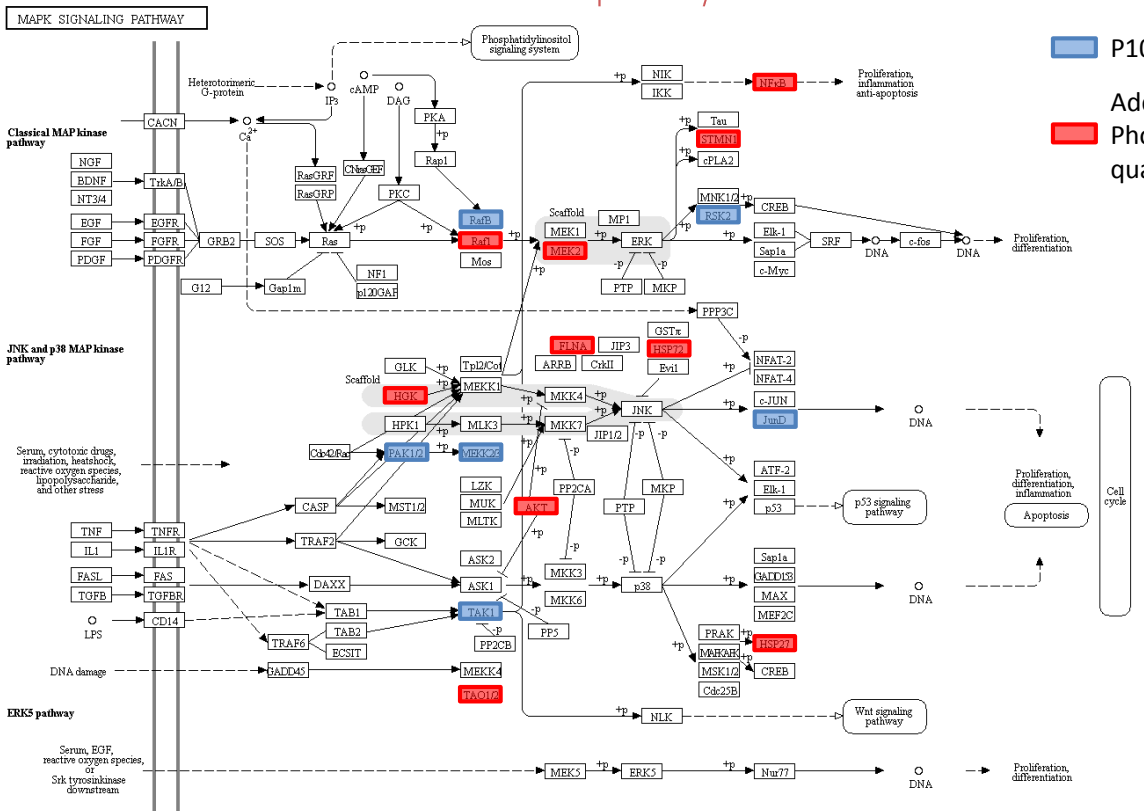
- 5 transitions
- Mean dot-product among all samples >0.75
- At least one replicate with mean dotp>0.90
- At least one replicate with mean intensity higher than the median of all intensities

Increased coverage of signaling pathways

2 pathway nodes
16 new pathway nodes



6 pathway nodes
10 new pathway nodes



P100 probes

Additional Phosphopeptides quantified by DIA

Conclusion

- Narrow-window DIA enables a comprehensive proteome analysis
- Developed a data-analysis workflow that insures high-quality and confident identifications
- We can demonstrate that the approach generates low false discovery rate
- Increased the number of identified phosphopeptides and phosphorylation sites
- The “Super Spectral Library” significantly increases the number of potential signaling analytes that can be quantified
- Our strategy significantly increases signaling pathway coverage

In the future

- There’s a need to obtain an automatized, accurate and robust DIA data analysis pipeline
- Querying the raw data with peptide-centric analysis (PECAN)
  - Chromatographic libraries would provide more specificity

- P100 DIA datasets are already acquired and ready to be analyzed

LINCS PCCSE

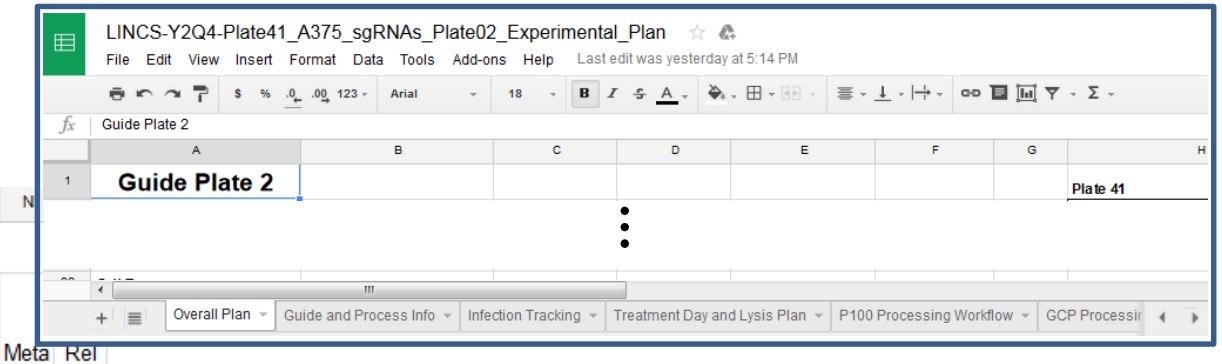
II: DATA RELEASE AND ACCESS

LINCS PCCSE

06 – TRACEABILITY OF EXPERIMENTAL WORKFLOWS AND METADATA

Central organization enables quick access to experimental plans

- LINCS Experimental Tracking
  - Stored on Google Drive for easy access

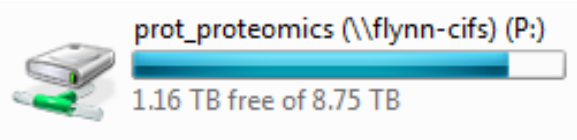


	A	B	C	D	F	G	H	I	J	K	L	M	N
1					P100								
2	Plate	Cell Type	Description	Collection Date	Harv	Pre	Dig	Enr	Res	Anlz	MQC	QC	Meta	Rel
30														
31	Plate 32	YAPC	CMPD Tranche 3	20151022	3	3	3	3	3	3	3	3	3	3
32	Plate 33	A549	CMPD Tranche 3	20151118	3	3	3	3	3	3	3	3	3	3
33	Plate 34	PC3	CMPD Tranche 3	20151118	3									
34	Plate 35	MCF10a	LINCS Common Project	20160218	3									
35	Plate 36	iPS	Haggarty Pilot	20160104	3									
36	Plate 37	A375-311	Cas9/sgRNA pilot	20160202	3									
37	Plate 38	HUVEC	Y1 Outreach	20160411	3									
38	Plate 39a	HAoSMC	Y1 Outreach	20160524	n/a	n								
39	Plate 39b	HAoSMC	Y1 Outreach	20160627	3									
40	Plate 40	A375-311	Cas9 Plate 1	20160426	3									
41	Plate 41	A375-311	Cas9 Plate 2	20160531	3									
42	Plate 42	A375-311	Cas9 Plate 3	20160613	3									
43	Plate 43	MCF10a	Time-course pilot	20160719-22	3									
44	Plate 44	YAPC-311	Cas9 Plate 1	20160816	3									
45	Plate 45,46	mixed iPS	Haggarty Pilot	?	3									

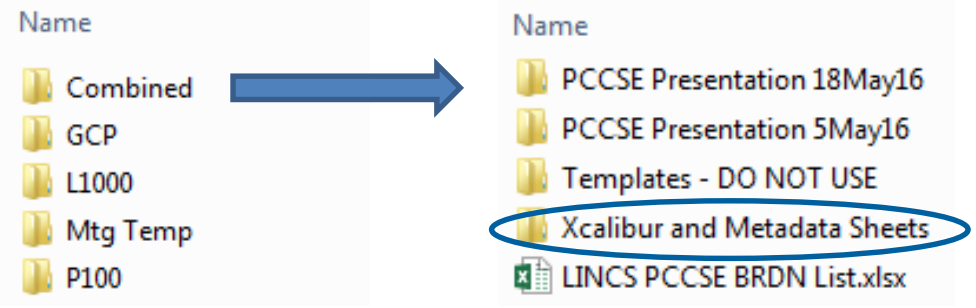
Process	Date Completed	Initials	Notes/Details
Harvest	20160613	AO, DD, AC, XL, SE, DL	See "Overall Plan"
Histone Extraction	20160830	SE, AC	
Protein Assay	20160830	SE	Samples below 10ug: none
Primary Propionylation	20160915	SE	BRAVO Used: Marvin
Oasis Desalt	20160915	SE	
Trypsin Digest	20160920	AC	BRAVO Used: Marvin
Secondary Propionylation	20160922	AC	BRAVO Used: Marvin
SepPak Desalt	20160923	AC	
Resuspension	20160929	SE	
Data Acquisition	20161007	AC	LC-MS system used: LC B, Tesla
Data QC	20161016	AC	
Metadata Attachment			

Consistent server organization allows for facile data access

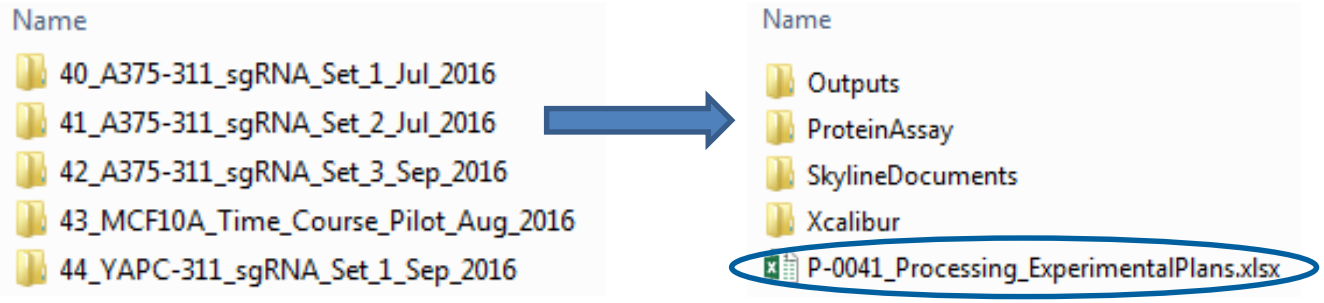
- Actual experimental data stored on our server
  - Data far too large to store on Google Drive
  - Raw MS files alone for one experiment are > 500 GB



- Folders organized by assay type
  - Metadata kept in shared location
  - Allows for mutual access and maximizes inter-assay agreement



- Further segmented into experiment ID and data types
  - Template folders to keep data management consistent
  - Unified naming system
  - Additional experimental tracking information stored here

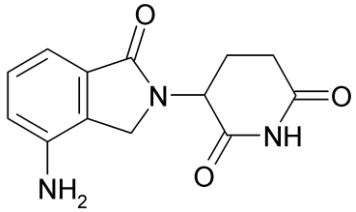


DCIC involved in curation of metadata for interoperability and future integration

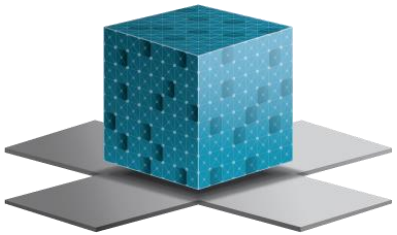
Genetic Perturbagens

- PAM sequence
- Target gene information
- Internal ID

Small Molecule Perturbagens

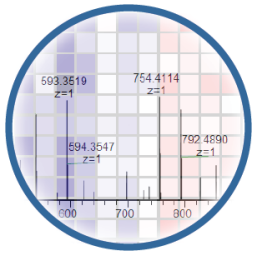


- Compound Name
- PubChem CID
- Batch information
- Internal ID



Information registered with DCIC who assign unique LSM/LNA identifiers

- Used in data integration across multiple assay types
- All metadata verified with the DCIC prior to sample annotation



Collect metadata for compounds and viruses, verify and register with DCIC and input into our databases

- Aim to minimize human intervention in this process
- Assign unique sample IDs

- More than 20 metadata fields are assigned to each sample prior to release
- Retain several more for internal records

Metadata organization allows for quick attachment and verification

- Curated master compound/virus metadata database
  - Sourced from the DCIC
- Metadata and sequence generation
  - Minimized number of input fields
  - Used by MS for acquisition
- MS data acquisition
  - Maximize redundancy and traceability
  - Metadata incorporated into MS files
- MS data QC via Skyline
  - Reattach metadata and integrate preliminary results
- Export GCT for data analysis

Internal ID#	Compound Name	SM_Center_Sample_ID	PubChem ID	Isomeric SMILES	LSM ID	Treatment Conc
12-1-1	staurosporine	BRD-K17953061	44259	C[C@@]12[C@@H]	LSM-1103	1
12-1-2	staurosporine	BRD-K17953061	44259	C[C@@]12[C@@H]	LSM-1103	1

Automated generation to reduce errors

Internal ID#	Compound Name	LSM ID	BRD# of Compound	Pubchem CID	Isomeric SMILES string
89-1-5	DMSO	LSM-36361	BRD-K08970894	679	CS(=O)C
89-1-5	DMSO	LSM-36361	BRD-K08970894	679	CS(=O)C
89-1-5	DMSO	LSM-36361	BRD-K08970894	679	CS(=O)C

Acquisition Instrument	Curie
Acquisition LC	N
Acquisition Date	9/15/2016
Study Identifier	P100_Plate39b_HAoSMC_CardioOutreach
Path	C:\Curie_Public\2016_Sept\LINCS C:\Xcalibur\methods\LINCS\P100_DIA_11amu_Overlap_22a
Method	muwindow_60min_50msIT_400-1000_27loopcount

Unique Sample IDs generated, integrated into raw files

```
CM20160908_P100_Plate39b_HAoSMC_CardioOutreach_P-0039b_A01_acq_01
Type: Unknown ID: PHC-44828-001A01 Row: 0
Sample Name:
Comments:                BRD-K08970894_DMSO_0_uM_3_h_rep_01
```

Attachment of metadata and reverification

Document Grid: p100_samples_IMPORTCHECK

Views | 1 of 80 | Export... | Find:

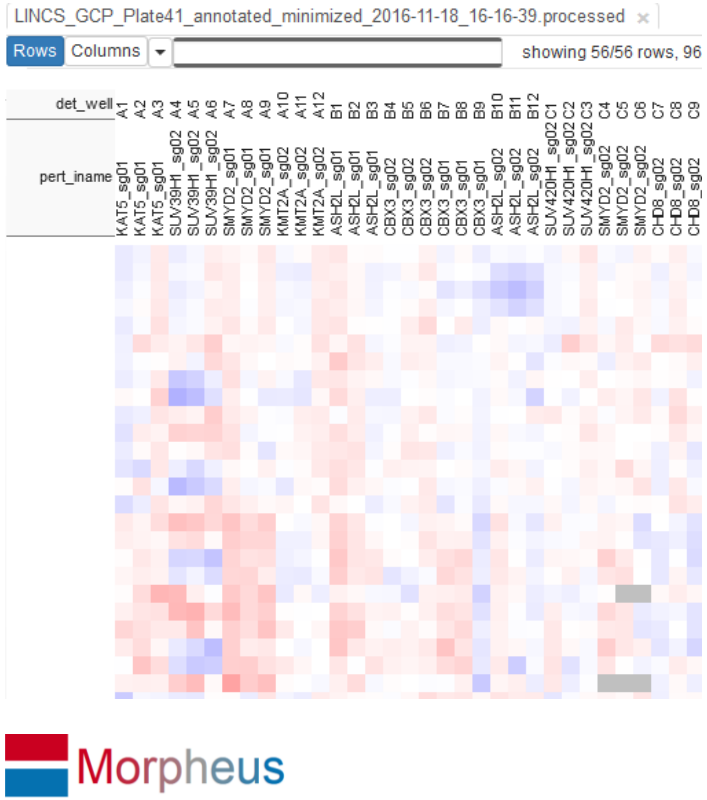
Replicate Name	id	cell_id	det_plate	det_well	det_well_enrichmer	det_normalization_g	det_filename	pert_id	lsm_id	pubchem_cid
A01_acq_01	PHC-44828-001A...	HAoSMC	P-0039b	A1	0.73	1	CN20160915_P1...	BRD-K08970894	LSM-36361	679
A02_acq_01	PHC-44828-002A...	HAoSMC	P-0039b	A2	0.85	1	CN20160915_P1...	BRD-K08970894	LSM-36361	679
A03_acq_01	PHC-44828-003A...	HAoSMC	P-0039b	A3	0.8	1	CN20160915_P1...	BRD-K08970894	LSM-36361	679

Data export

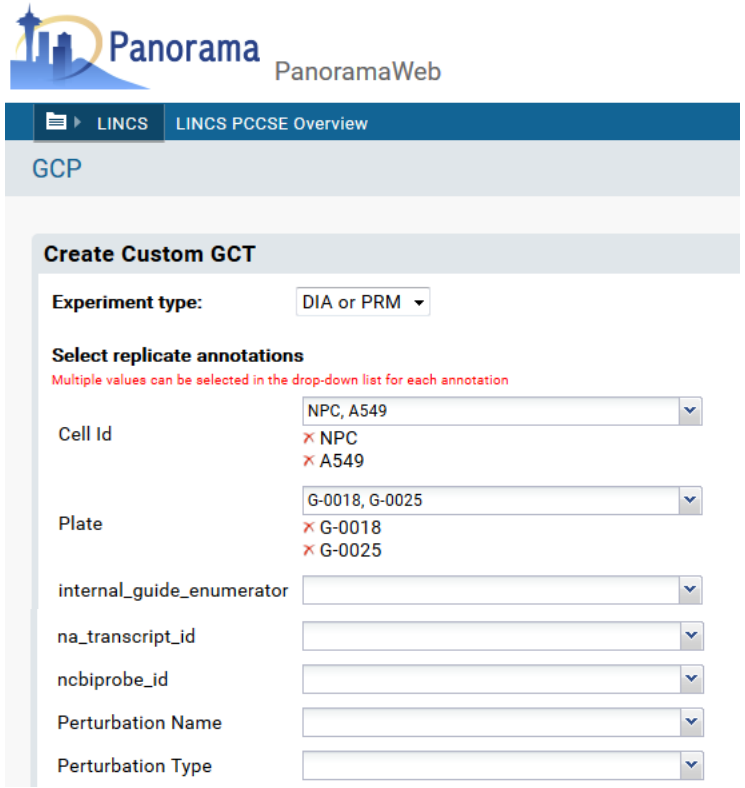
LINCS_P100_Plate39_DIA_Quant.gct
GCT File
182 KB

Easy integration with existing tools and frameworks

CMap/L1000 tools



Skyline/Proteomics tools



- Direct integration into existing pipelines through GCT file format
- Panorama allows for easy subsetting of PCCSE data
- **Careful tracking and integration of metadata is critical for later analysis!**
  - From trivial (like process tracking) to complicated (leveraging outside information for “Google-like” search)

LINCS PCCSE

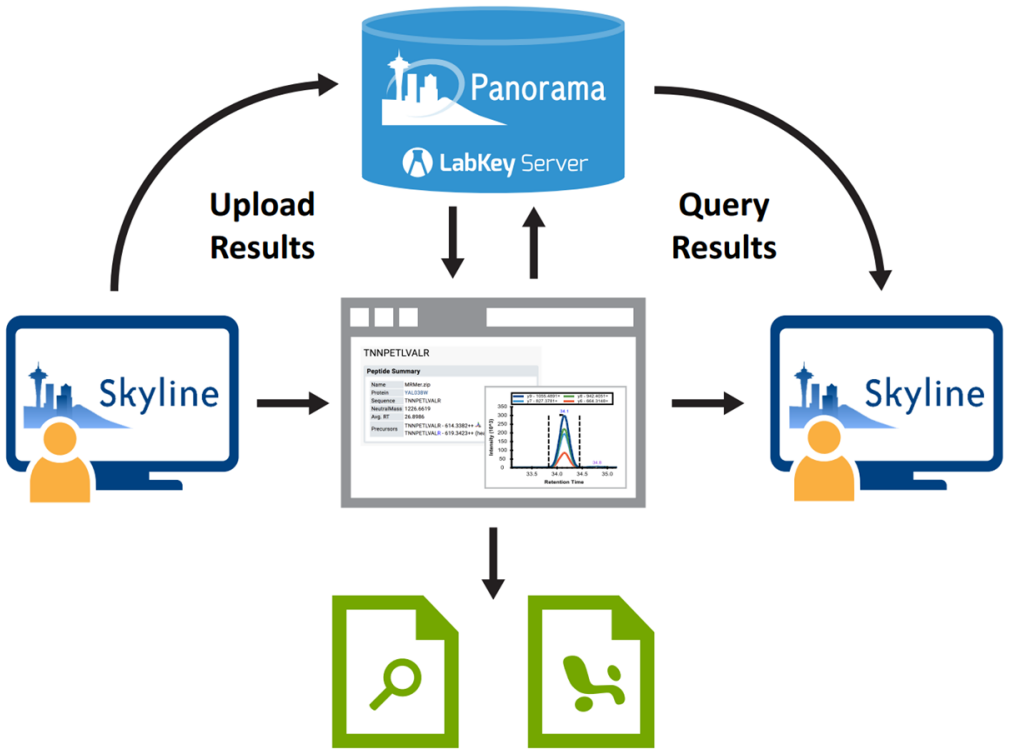
07 – PUBLIC PORTAL ACCESS

Data release & access*iblity* – providing public portal & tools

- Access vs. Accessibility
  - Access: it is is possible for someone to retrieve the data
  - Accessibility: making it as easy as possible for someone to discover and use the data
- Ongoing goal: provide tools to visualize & analyze PCCSE data
  - Quick wins: reuse & extend existing CMap / L1000 tools
- How: align data levels, metadata and file formats

Panorama is the main LINCS PCCSE data repository

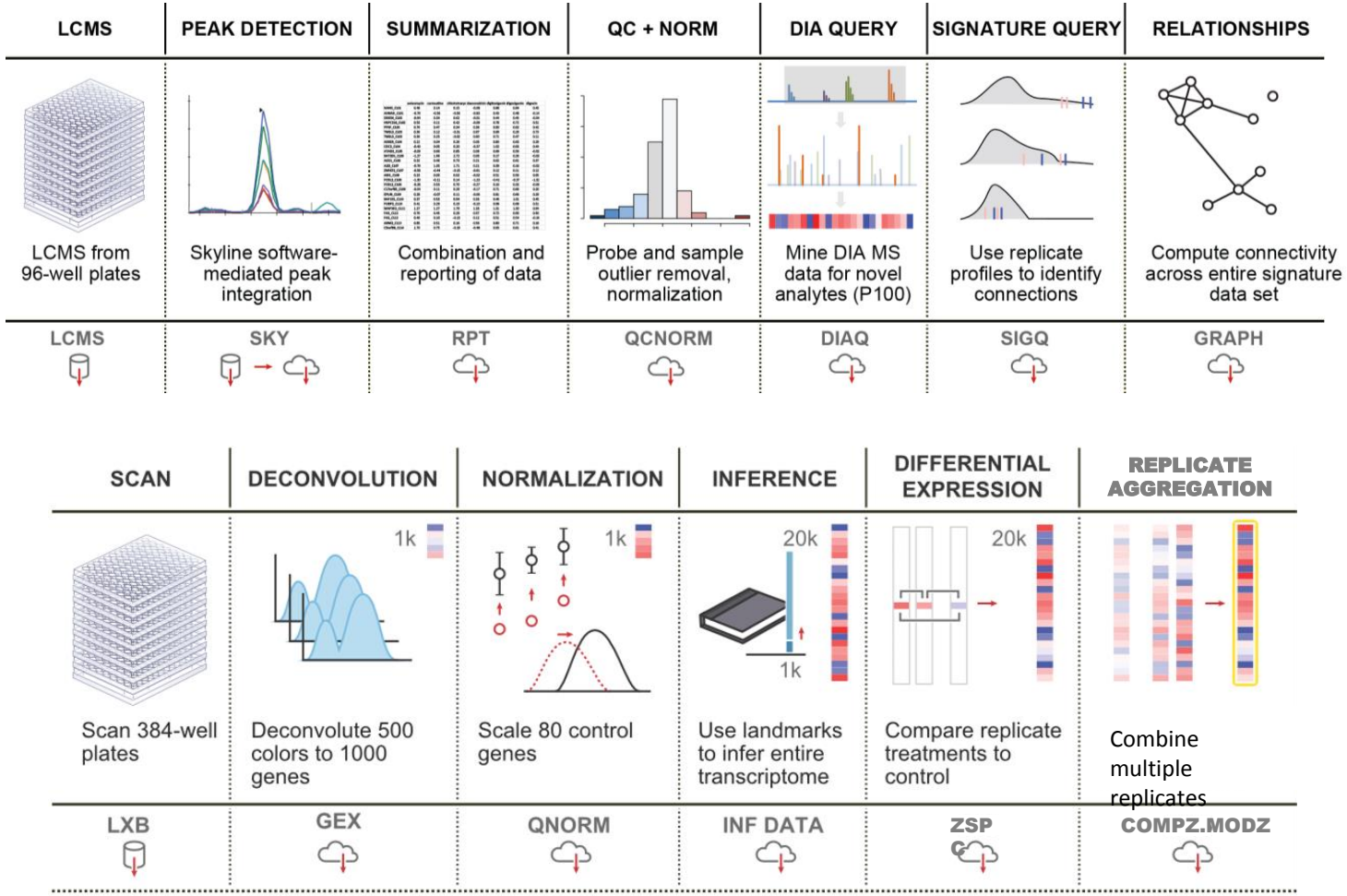
- Publish fully annotated Skyline documents
- Build chromatogram libraries
- Aggregate lab QC data (future)
- Free hosted version (http://panoramaweb.org)
  - 206 separate projects so far (LINCS, CPTAC, & ABRF sPRG)
  - >7000 data sets uploaded
  - User controlled security
- Locally installable server application
- Free and open source (Apache 2.0)



Sharma, et. al, *J. Proteome Res. 2014*

Leveraging aligned data levels, metadata and file formats to quickly achieve accessible data

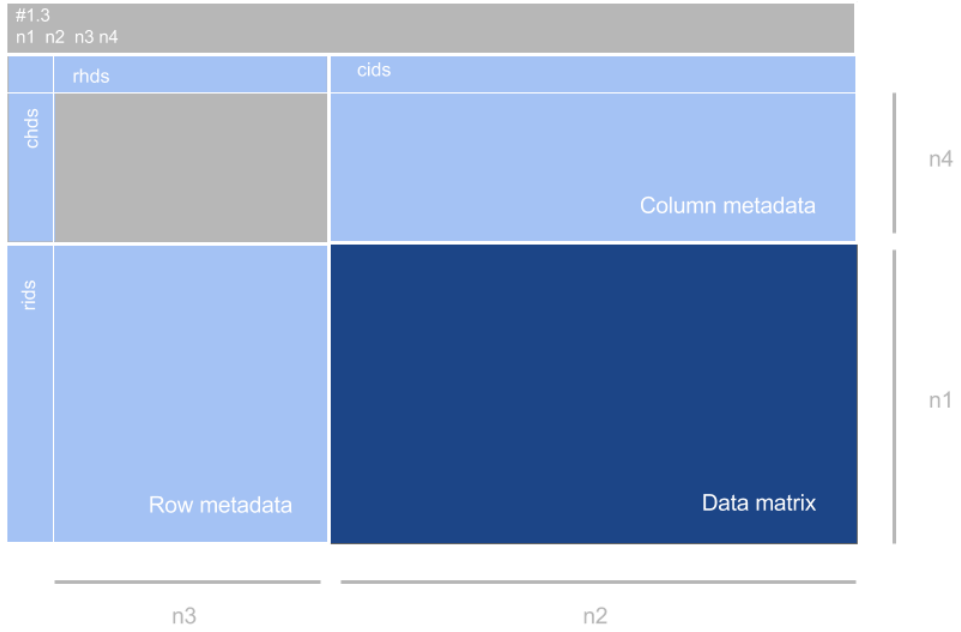
Aligned data levels for Proteomics and L1000:



Consistent metadata fields and values:

pert_id	BRD-K12345678
pert_type	trt_cp
pert_iname	vemurafenib

Reused GCT file format:



Panorama website allows searching/browsing, visualization and analysis

- Demo!

The Chorus Project

Chorus

Usage Statistics

1,596 Users

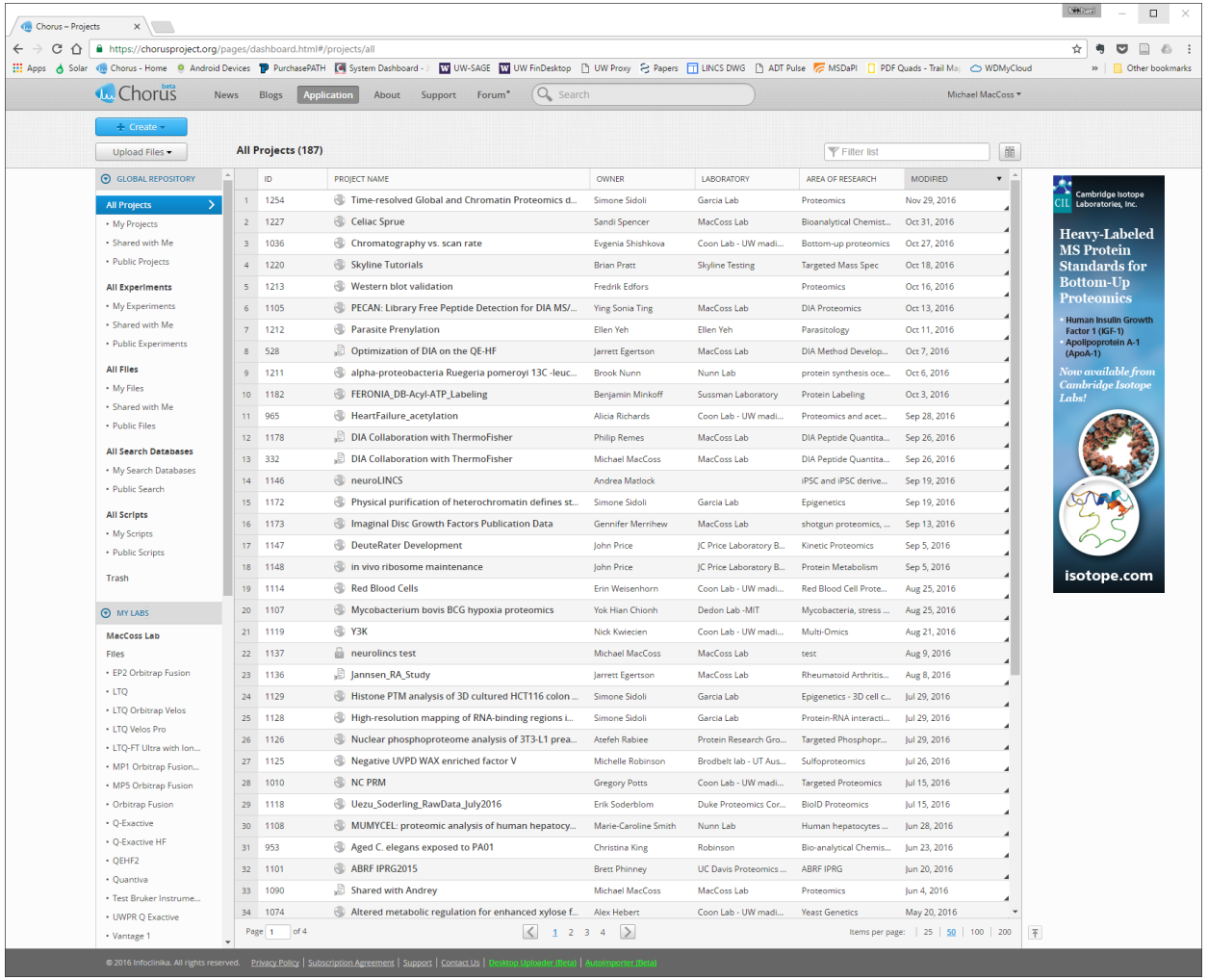
43.88 Data volume (TB)

125,401 Data files

161 Public projects

406 Public Experiments

http://chorusproject.org



New paradigm for data storage and processing of MS “big” data

- Proteomics data needs to be “un-siloed”
  - A single place for all data
- Volume of high value data is increasing
- Crucial for peer review and critical analysis
- Economics are driving demand for data reuse
- Cultural shift: Public access to data collected with public funds
- We need to bring algorithms to the data
- We want to be able to do computations across all data and not a tiny subset

LINCS PCCSE

III: COMMUNITY INREACH & OUTREACH

PCCSE community inreach and outreach efforts

- Drug signatures in patient iPS-derived neuronal models (Steve Haggarty, MGH)
- **Drug signatures in breast epithelial models (LINCS) MCF10A common project**
- piLINCS, and leadership in LINCS proteomic working group (DCIC/Mike MacCoss, UW)
- **Testing of microenvironment microarrays for neuronal differentiation (OHSU LINCS Center)**
- Profiling of common conditions in cortical and motor neurons (NeuroLINCS Center)
- **Signatures of vascular toxicity (Iris Jaffe/Srila Gopal, Tufts Medical Center)**
- Dedicated short course on LINCS PCCSE methods and data (Olga Vitek, NEU, May 2017)
- National and international invited podium presentations to discuss LINCS proteomics efforts
  - ETH Zurich, Targeted Proteomics International Symposium, USHUPO, & more.

LINCS PCCSE

08 – LINCS CONSORTIUM MCF10A COMMON PROJECT

Why MCF10A as a model system?

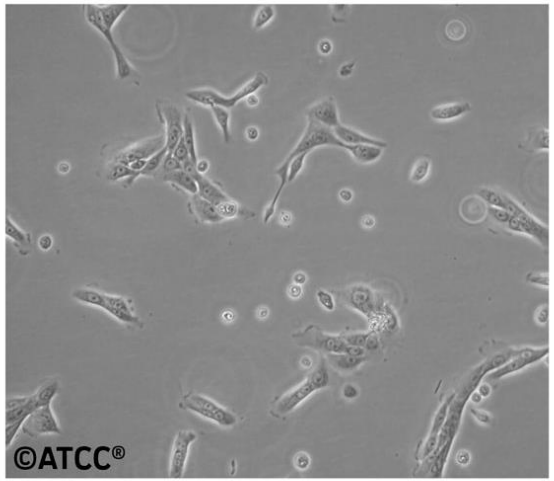
MCF10A is a non-tumorigenic mammary epithelial cell line

Has few genetic abnormalities

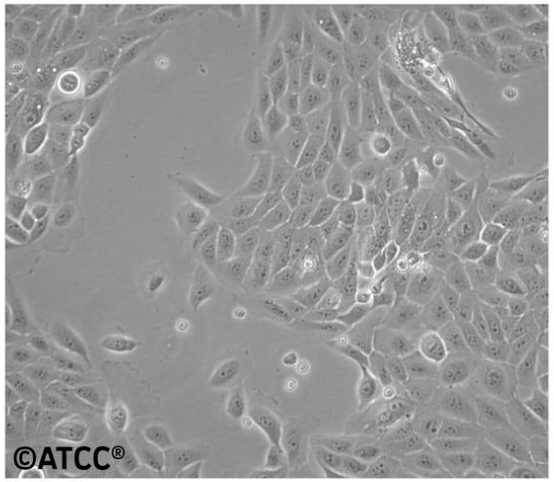
Commonly used to model normal human breast cells

Is amenable to different culture modes

Has a history of use in wide range of research



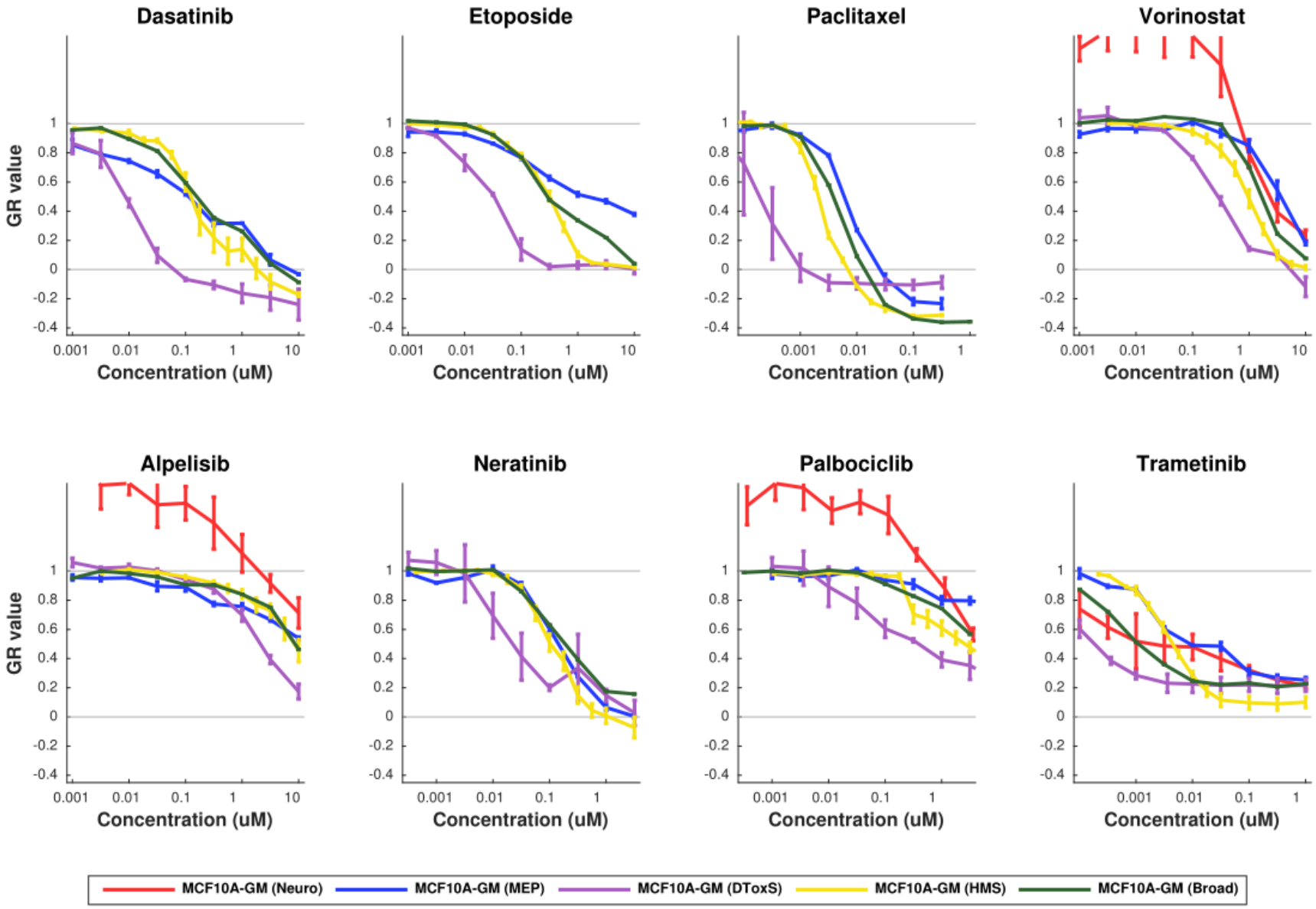
Low Density

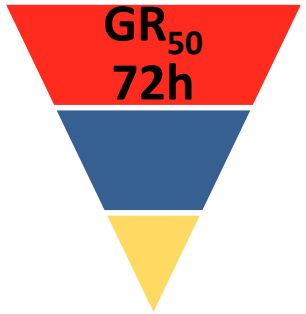


High Density

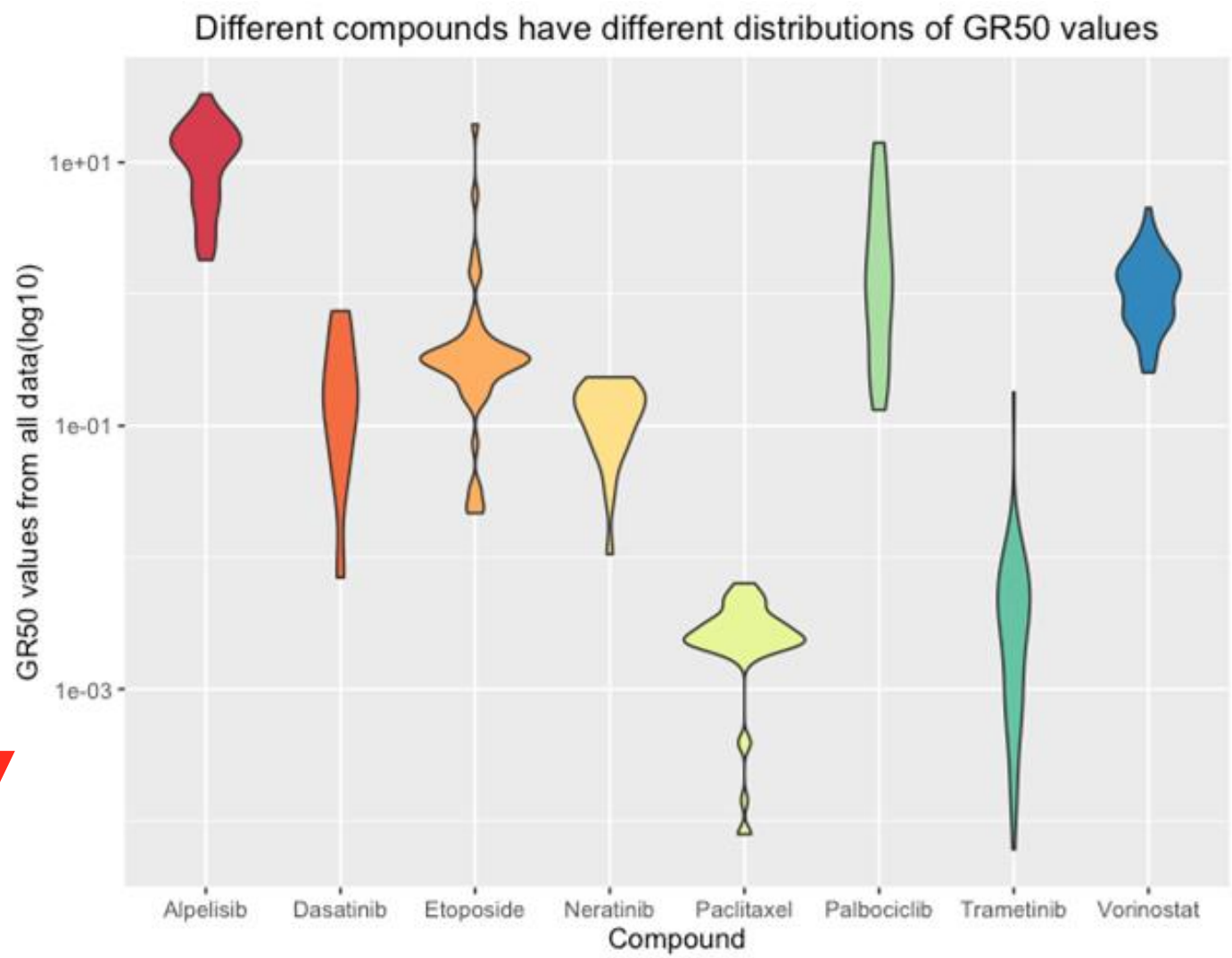
Drug Name	Mechanism of Action
Alpelisib (BYL719)	Ras/PI3K-P110a inhibitor
Dasatinib	Tyrosine kinase inhibitor
Etoposide	Top2B inhibitor
Neratinib	Tyrosine kinase inhibitor
Palbociclib	CDK4/CDK6 inhibitor
Paclitaxel	Microtubule breakdown inhibitor
Trametinib	MEK1/2 inhibitor
Vorinostat	HDAC inhibitor

Inter-center analysis reveals differences in basic kill curves

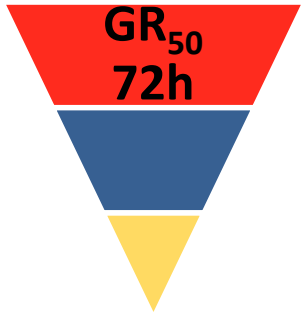




Analysis reveals a range of measured GR_{50}



Compound	Concentration
DMSO	0.1pct
Paclitaxel	0.003 µM
Alpelisib	10 µM
Neratinib	0.1 µM
Dasatinib	0.316 µM, 0.1 µM, 0.032 µM
Trametinib	0.01 µM, 0.003 µM, 0.001 µM
Palbociclib	3.16 µM
Etoposide	0.316 µM

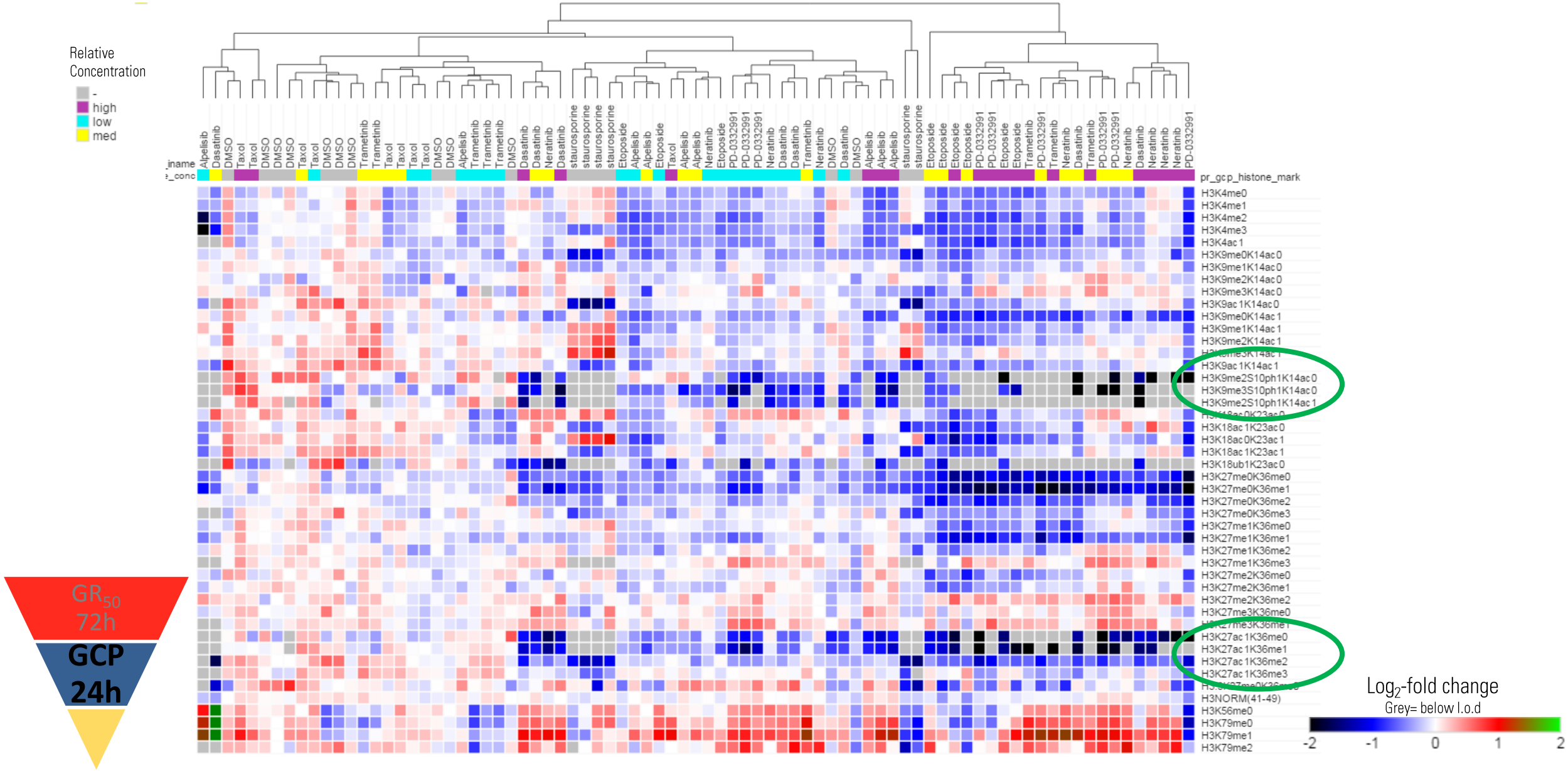


Drug concentrations for GCP and P100 assays defined by average of GR_{50} curves

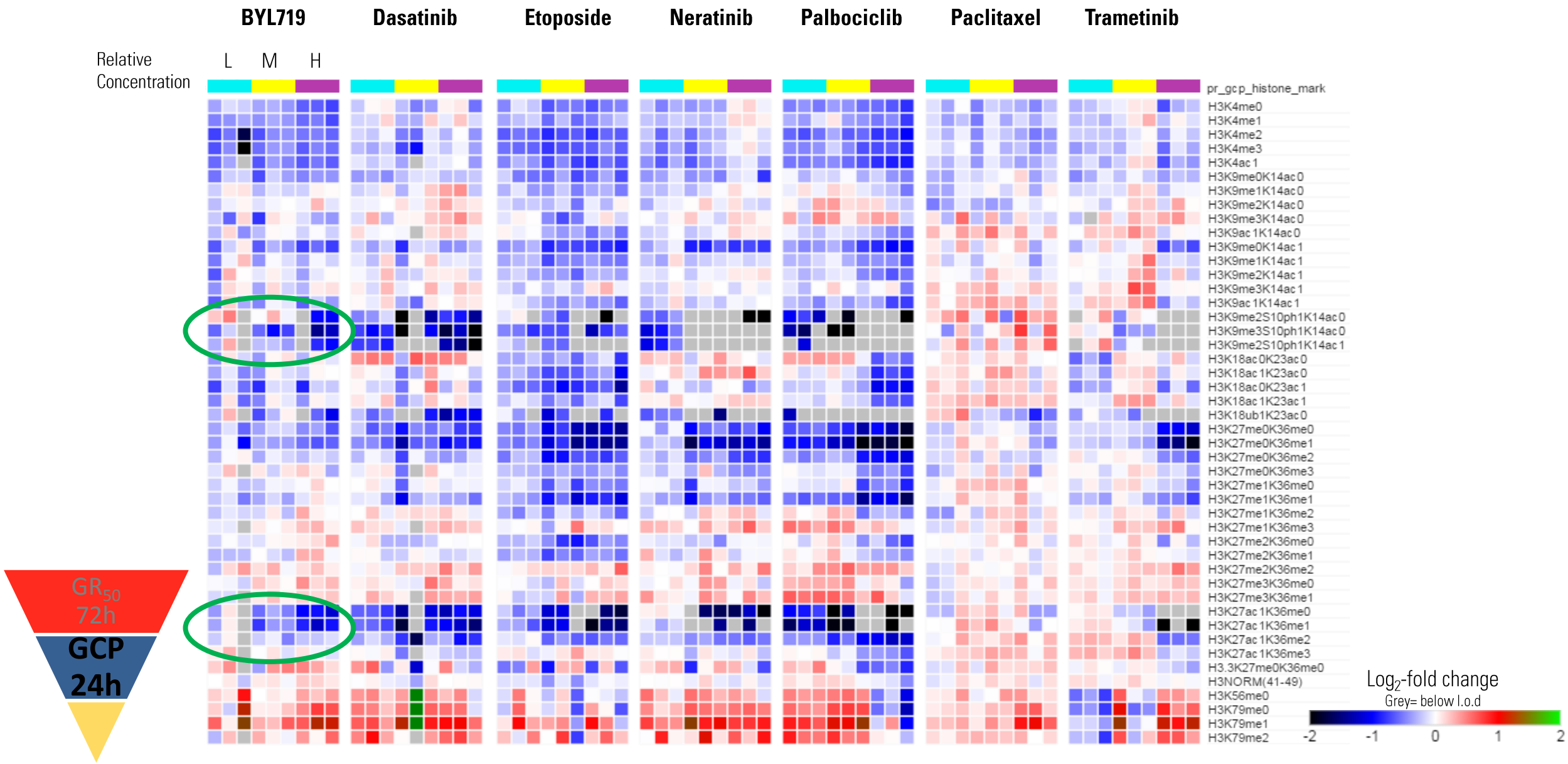
Drug	Low Conc (µM)	Medium Conc (µM)	High Conc (µM)
Neratinib	0.032	0.1	0.316
Dasatinib	0.1	0.316	1
Palbociclib	0.316	0.999	3.16
Paclitaxel	0.001	0.003	0.01
Trametinib	0.001	0.003	0.01
Alpelisib	1	3.162	10
Etoposide	0.1	0.316	1
Vorinostat	1	3.162	10

- Medium concentration was defined as equal to average GR_{50} value

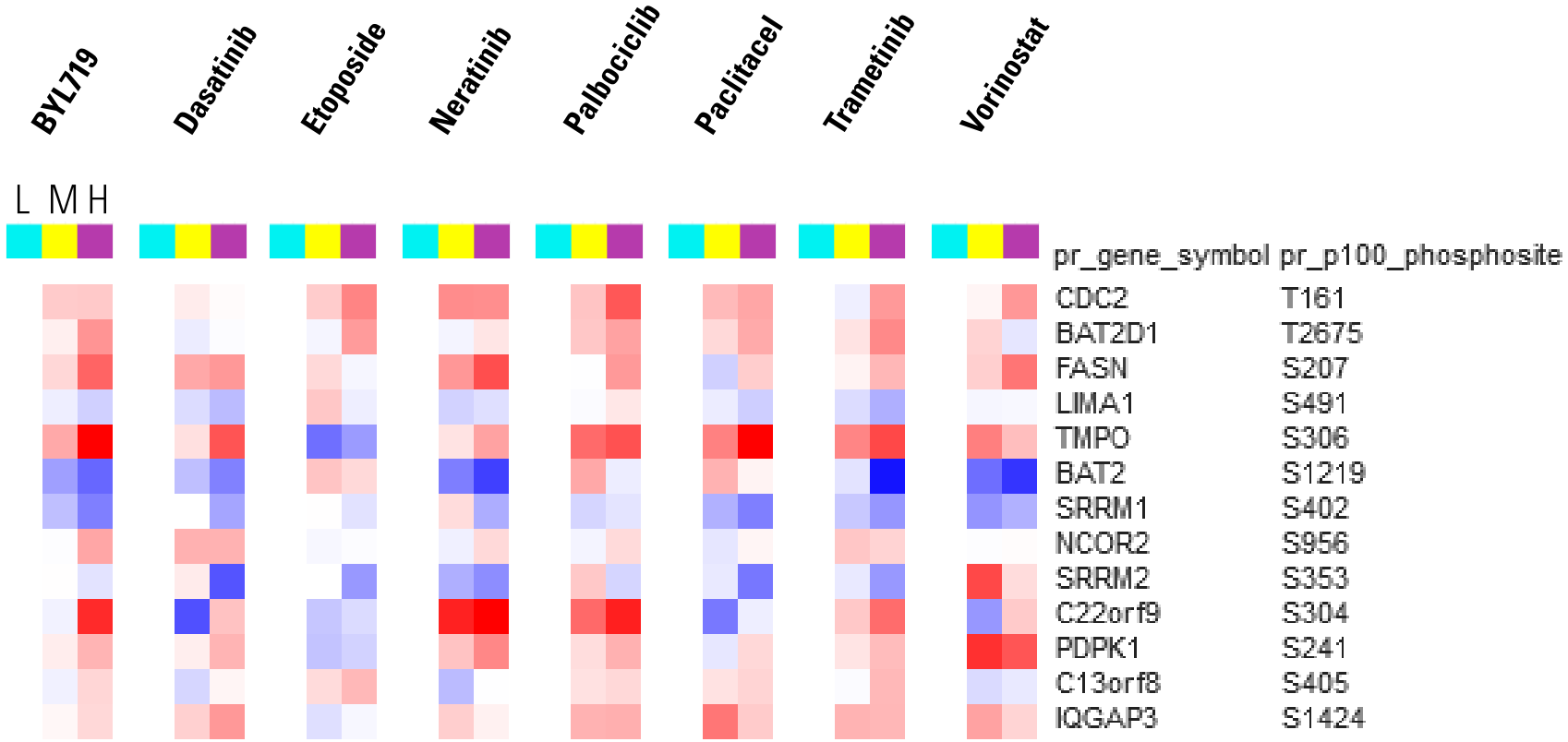
Clustering reveals a subset of histone PTMs as markers of cellular stress

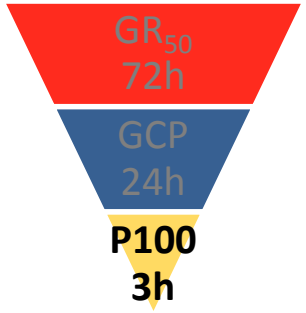


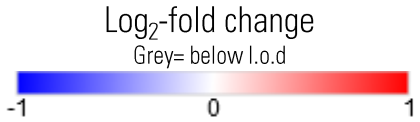
Intensity of signature of cellular stress is proportional to drug concentration



P100 probes show signature of cellular stress proportional to drug concentration



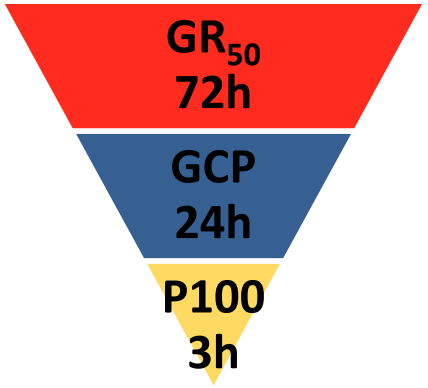




MCF10A common project conclusions

- More stringent protocols need to be developed in order to achieve multi-center data concordance

- The GR_{50} assay (72 h) can be used to assess cellular stress, however the GCP (24 h) and P100 (3 h) provide a rapid readout of cellular stress far before cellular death actually occurs

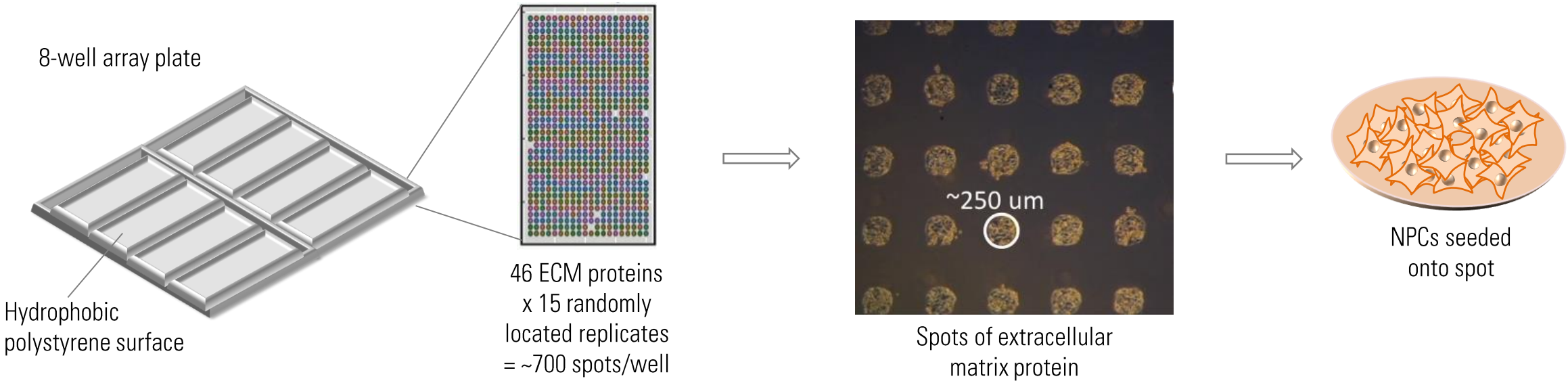


- Follow-up MCF10A study will investigate the evolution of the 'Signature of Stress' in P100 assay as a function of time (t=6, 24, 48, 72 h), as opposed to drug concentration
  - Will the same phosphopeptide set prove to be representative of cellular stress?
  - Will different modes of stress be stratified?

LINCS PCCSE

09 – MICROENVIRONMENT MICROARRAYS (MEMA) IN NEURONAL DEVELOPMENT

MEMA plates with arrays of extracellular matrix proteins



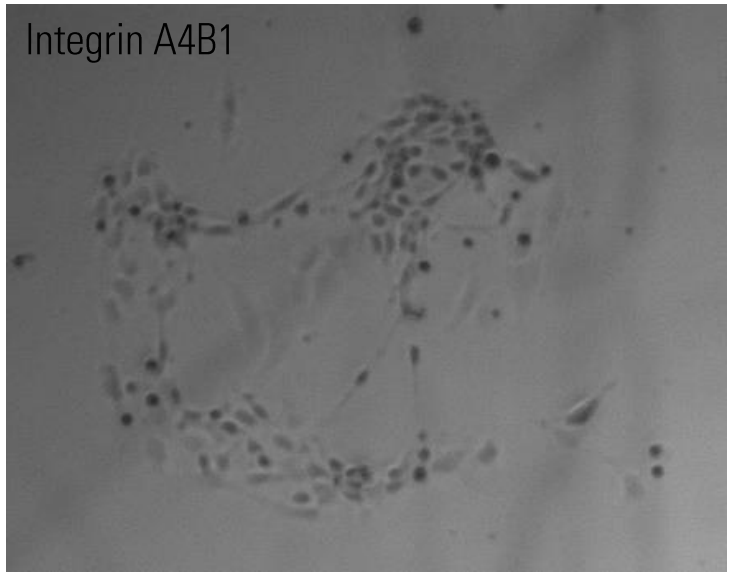
MEMA Substrates				
ALCAM	Integrin αMβ2	E-cadherin Fc Chimera	vitronectin	PECAM1
Cadherin-20 (CAD-20)	Integrin αVβ6	ECM1	Biglycan	Tenascin C
Cadherin-6/KCAD Fc Chimera	Laminin	Fibronectin	Decorin	VCAM1
Cadherin-8 (CAD-8) Fc Chimera	Laminin-5	GAP43	Periostin	Collagen Type V
CD44	Lumican	HyA-500K	SPARC/osteonectin	Collagen XXIII α1/COL23A1
CEACAM6	M-Cad	HyA-50K	Thrombospondin-1/2	Desmoglein 2
Collagen I	Nidogen-1	ICAM-1	Brevican	Osteopontin
Collagen Type II	Osteoadherin/OSAD	Integrin α10β1	Collagen Type III	P-Cadherin Fc Chimera
			Collagen Type IV	

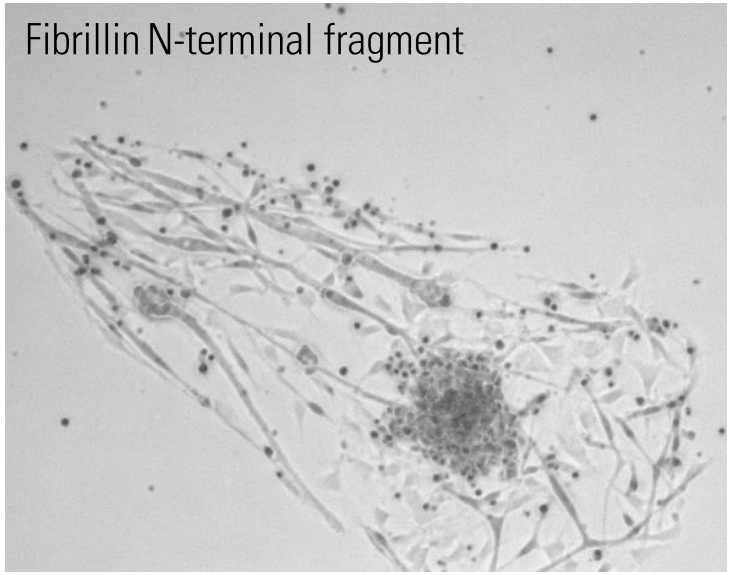
MEMA applications

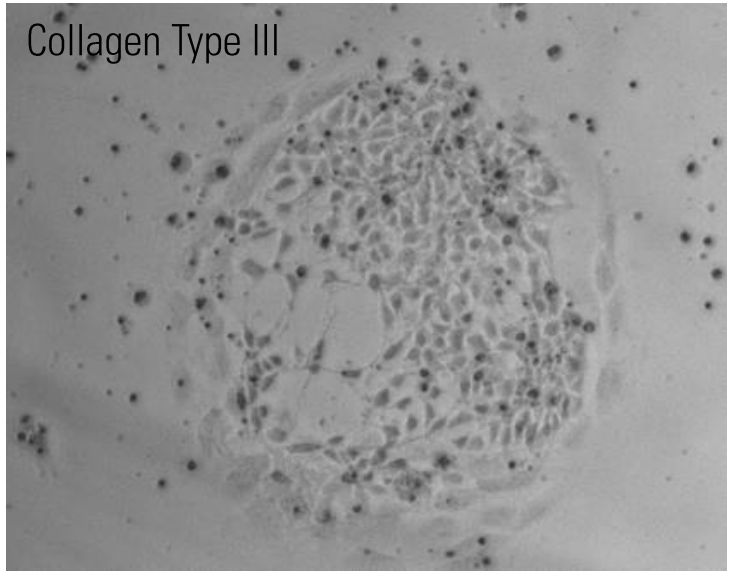
- Can we find substrates that will promote neuronal differentiation in high-content hES-derived NPC cultures?
- How do cells respond to chemical perturbations in the presence of ECM?
- Are there substrates that can rescue neurodevelopmental phenotypes that occur upon the sgRNA-mediated knockout of ASD implicated genes?
- Cell painting as an unbiased approach for characterizing cell morphology alterations induced by disease states or perturbations

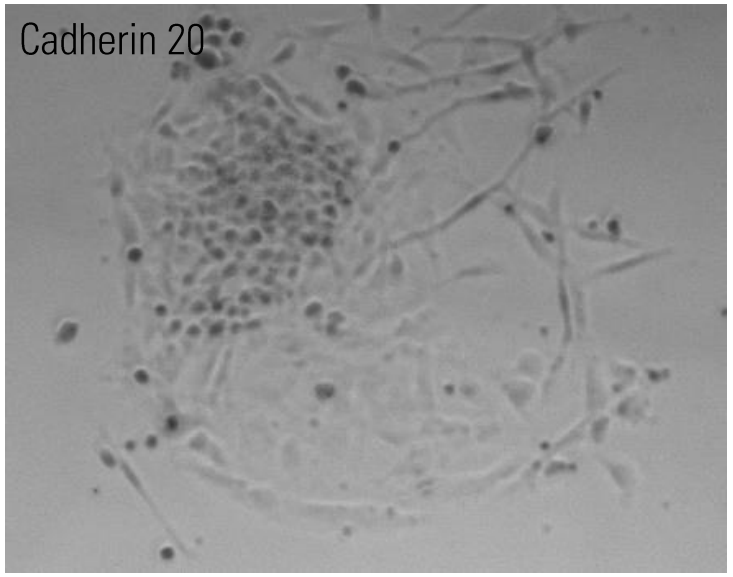
Can MEMA substrates mediate differentiation?

HUES8-iCas9 NPCs
TWO WEEKS of GFW



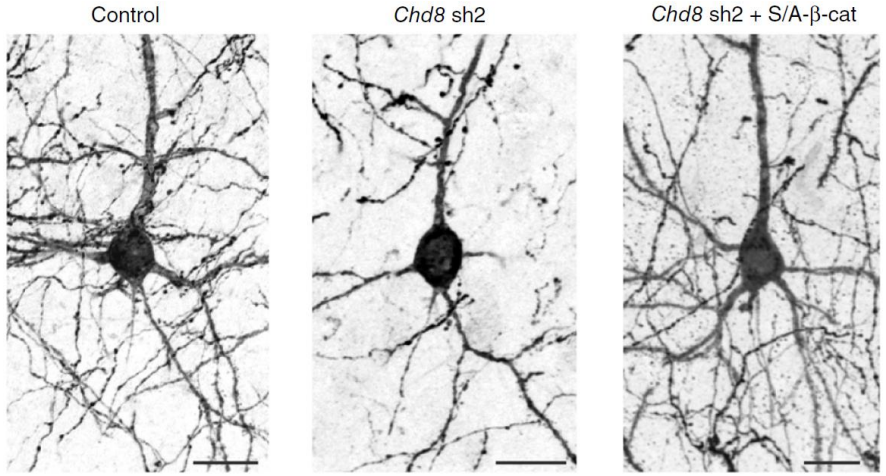


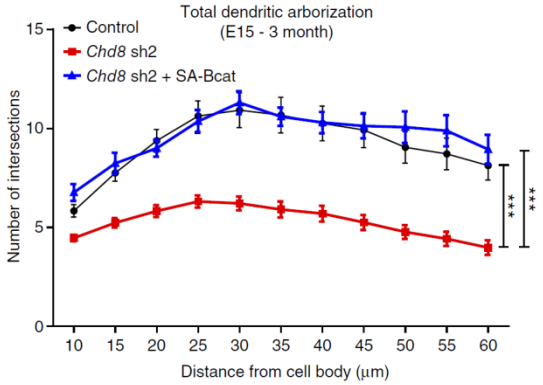




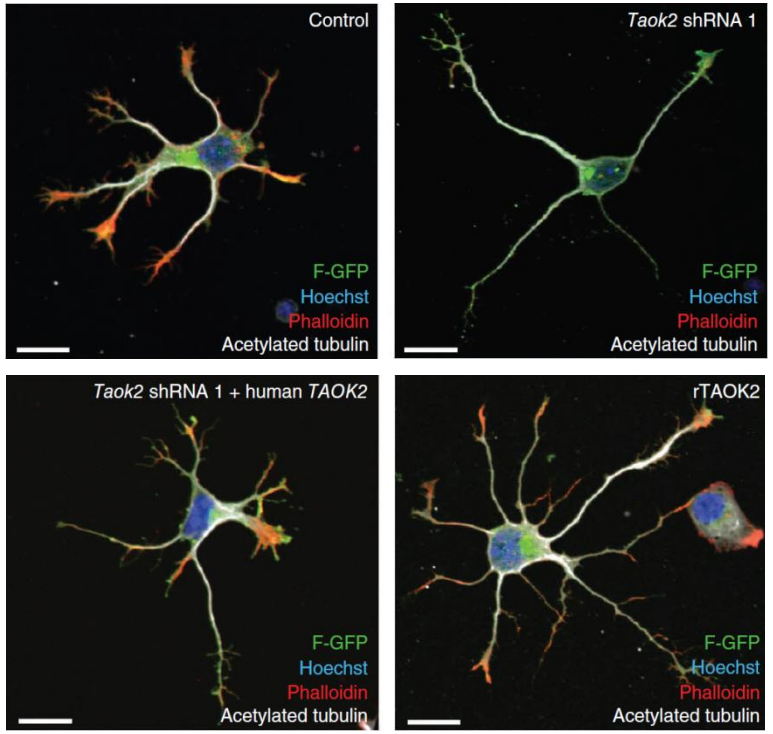
ASD risk genes affect morphology of developing neurons

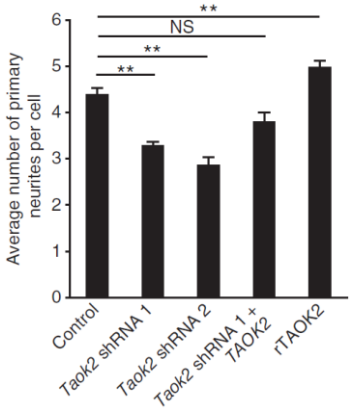
- CHD8: chromatin remodeling protein
- Knockdown in the developing brain leads to *abnormal neuronal morphology* and behavior in mice via disruption of Wnt–β-catenin signaling

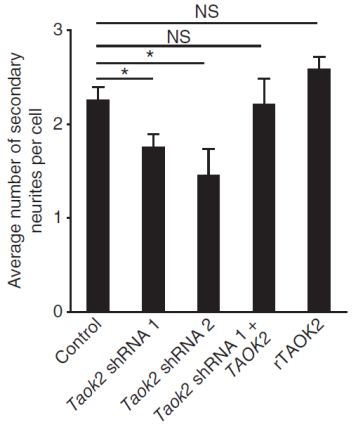




- TAOK2: Ser/Thr kinase
- Down-regulation or overexpression affects differentiation of cortical pyramidal neurons
- Basal dendrite complexity decreases with knockdown and increases with overexpression via JNK pathway

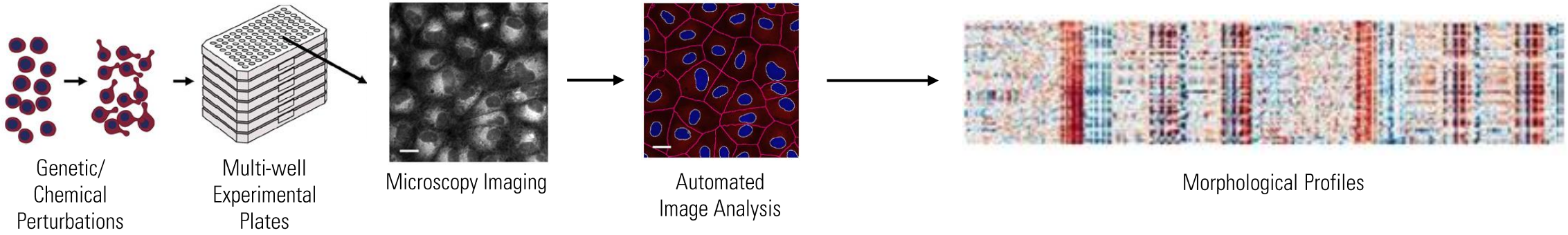


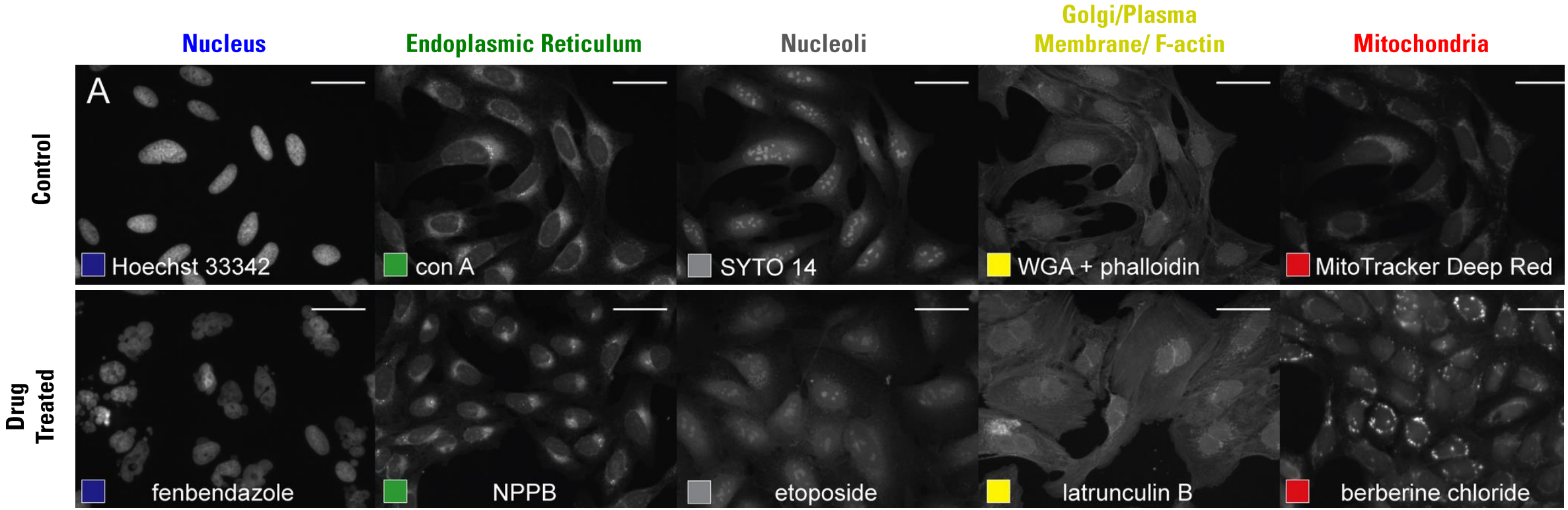




Durak *et al*, 2016, Nat Neurosci; de Anda *et al*, 2012, Nat Neurosci

Cell painting characterizes subcellular structure



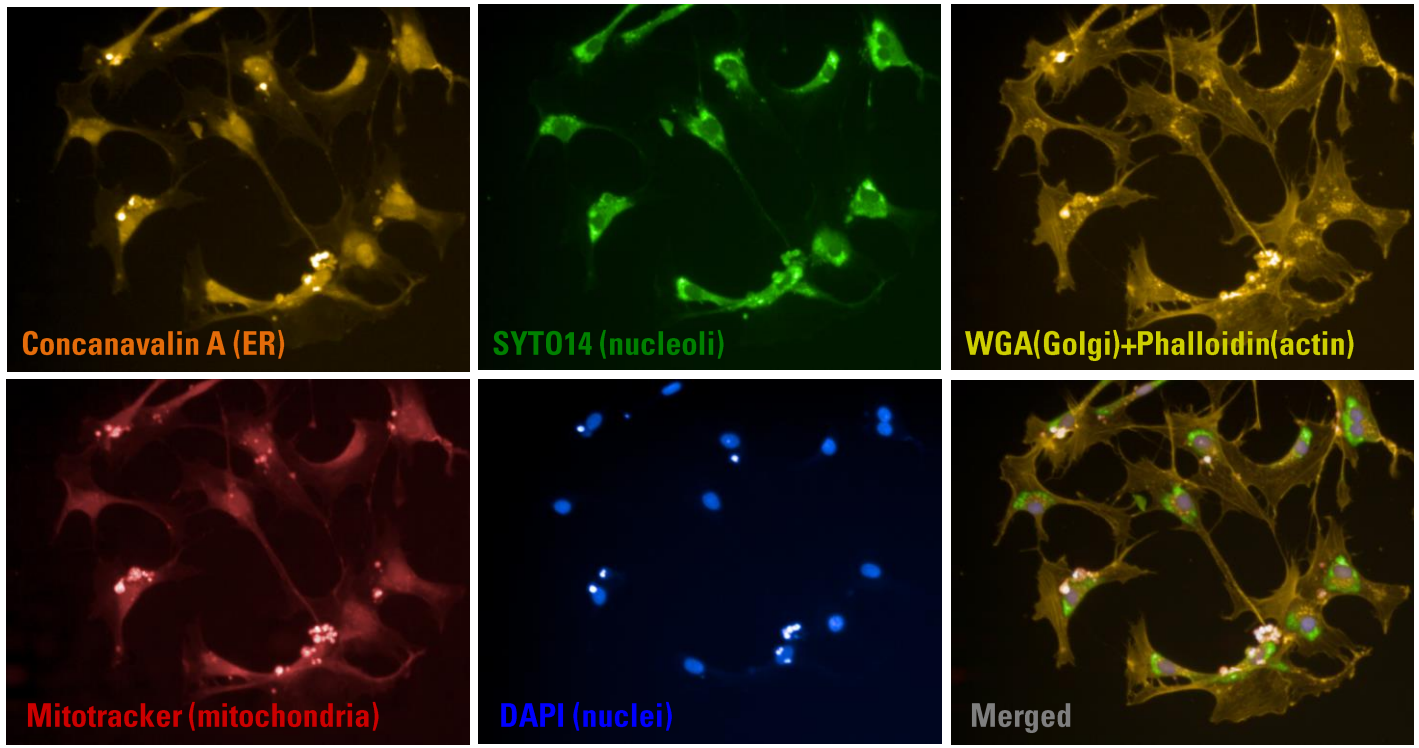


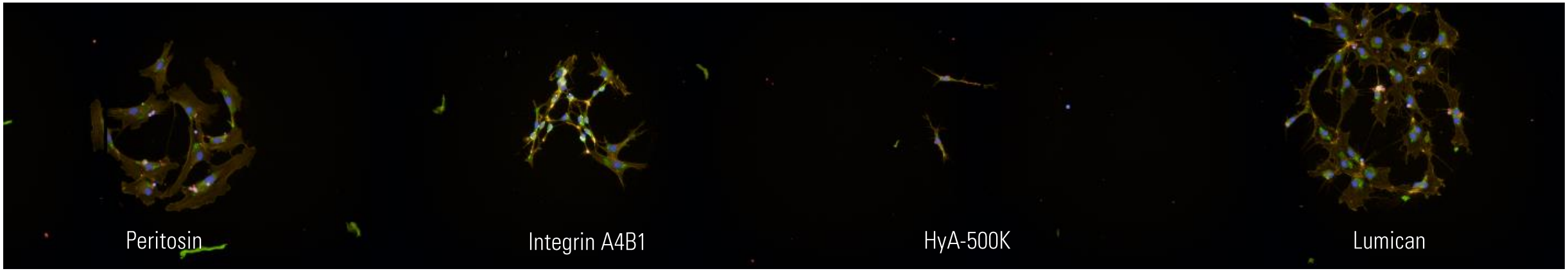
Gustafsdottir *et al*, 2013, PLOS One

Cell painting highlights MEMA substrate-mediated morphologies in hNPCs

Substrate: **PECAM1**

HUES8-iCas9 NPCs
6 weeks of GFW



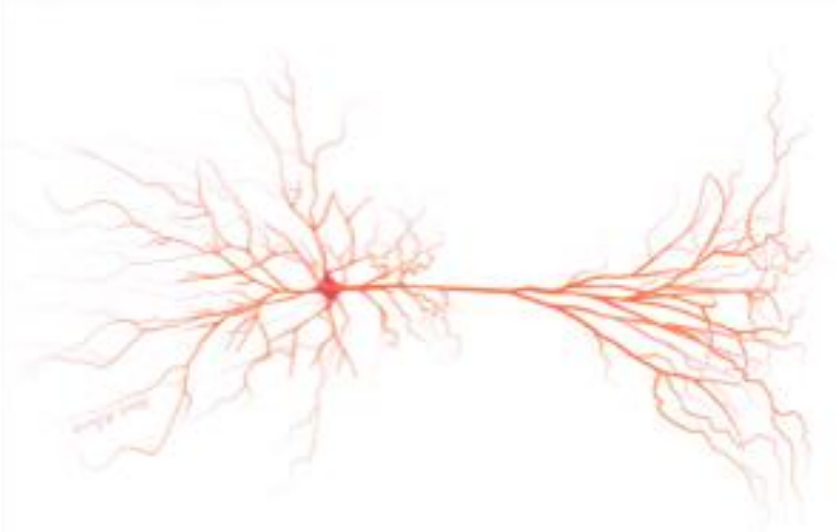


BROAD INSTITUTE
Anne Carpenter, Maria Alimova

Future work

- Introduce ECM substrates which encourage healthy neuron morphology to GFW hNPC differentiation protocol
  - Validate via immunohistochemistry for neuron, NPC, and glia markers

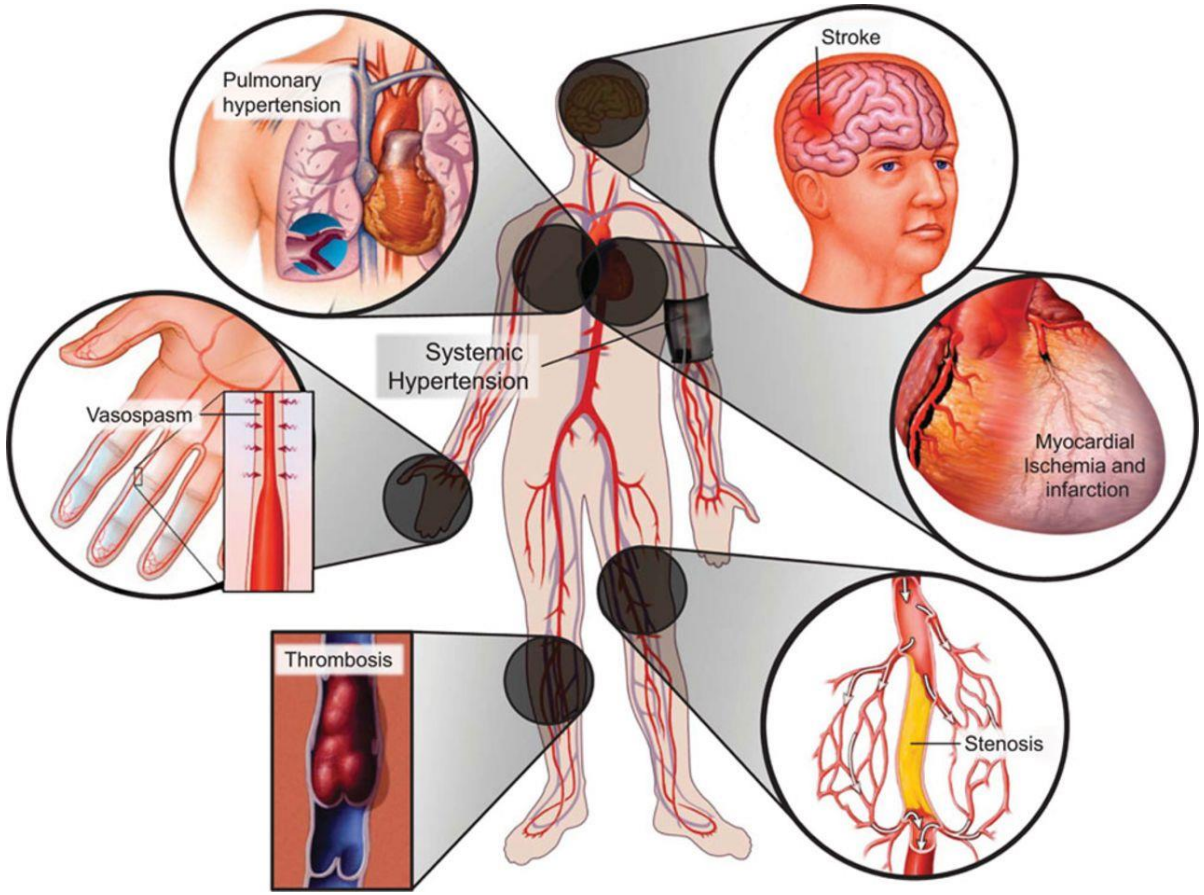
- Develop cell painting (Anne Carpenter, Broad) algorithms to generate morphological profiles of genetic and epigenetic cell states



LINCS PCCSE

10 – CARDIOVASCULAR MODELS

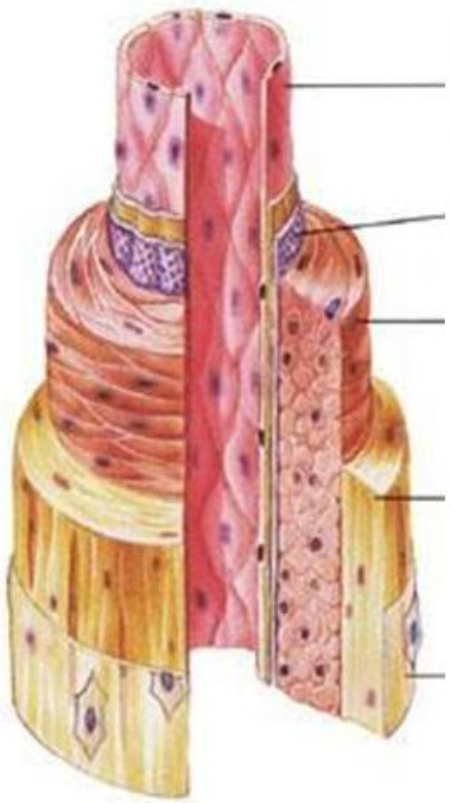
Impact of novel anti-cancer agents on the vasculature



Tyrosine kinase inhibitors	HTN	Angina	MI	Stroke	PAD	Pulm HTN	DVT/PE
Sorafenib	X	X	X	X			X
Sunitinib	X	X	X	X			X
Pazopanib	X	X	X	X			X
Axitinib	X	X	X	X			X
Regorafenib	X	X	X				
Cabozantinib	X		X	X			X
Vandetanib	X			X			
Lenvatinib	X		X	X			X
Nilotinib		X	X	X	X		X
Ponatinib	X	X	X	X	X		X
Dasatinib						X	

Joerg Herrmann et al. Circulation. 2016;133:1272-1289

Structure of the blood vessel



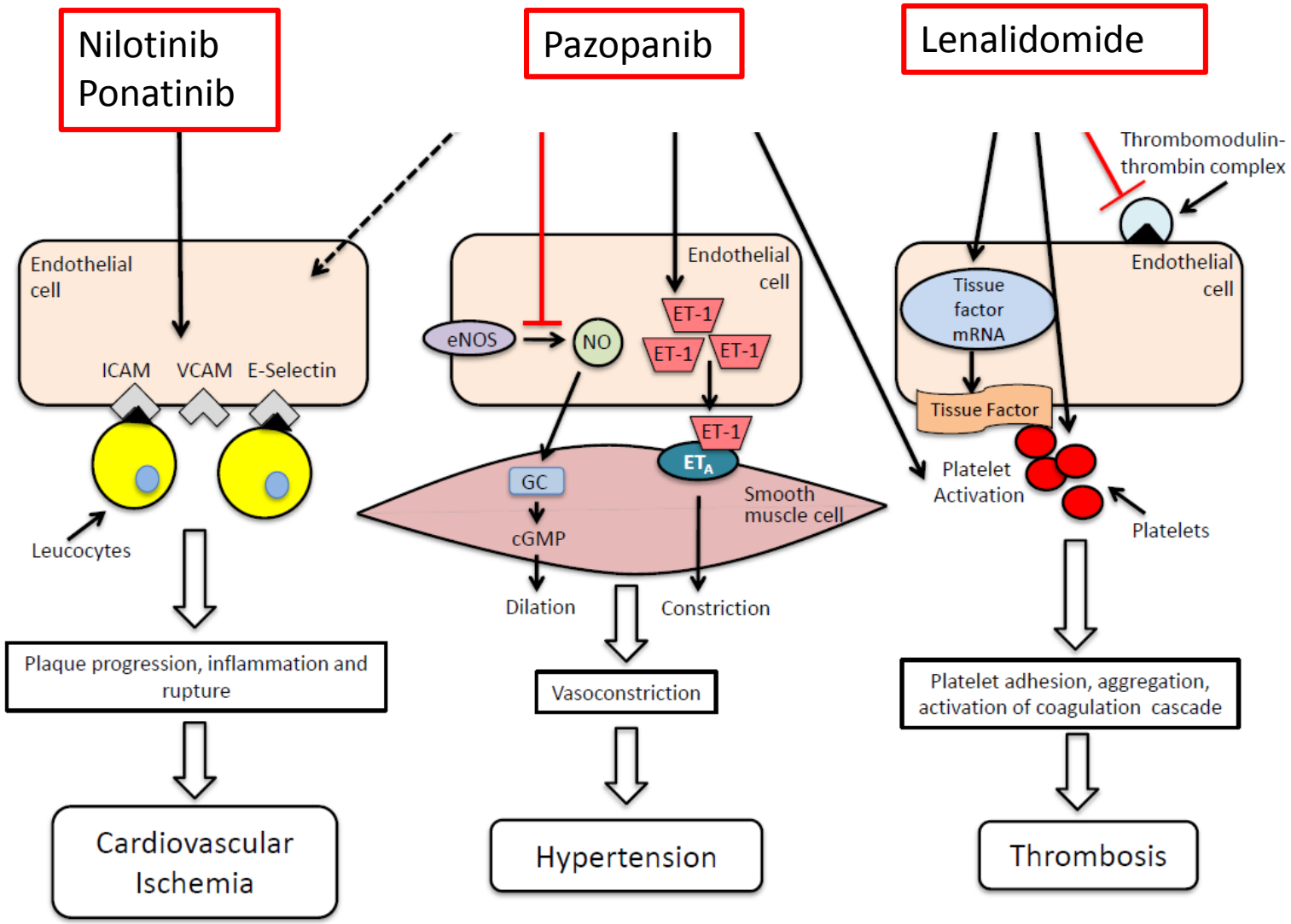
Endothelium – smooth inner lining, repels blood cells, secretes vasodilators and vasoconstrictors

Elastic tissue

Smooth muscle- contraction and relaxation of blood vessel

Connective tissue

The endothelial cell: proposed central mediator of vascular toxicity of chemotherapeutics



Gopal S, Miller KB, Jaffe IZ, Clin Sci (Lond). 2016

Hypothesis

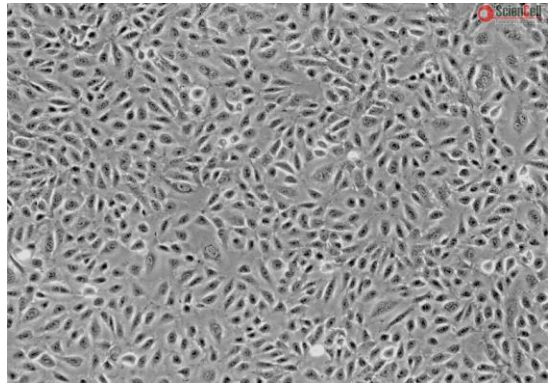
- Anti cancer agents cause vascular toxicity by altering signaling and epigenetics in vascular cells.
- Pathways altered in vascular cells are different from those in cancer cells and understanding these mechanisms can help design therapeutic targets to attenuate toxicity while maintaining anti cancer benefits.

Aims

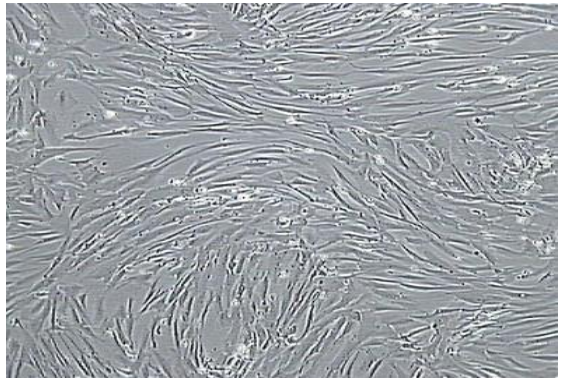
- Determine the chromatin profile and phosphoproteomic profile of primary human vascular cells (human aortic SMCs and ECs) treated with small molecule inhibitors.
- Compare the epigenetic and phosphoproteomic profiles in vascular cells to those already being generated in cancer cell lines (A549 (lung), A375 (skin), PC3 (prostate) and MCF7 (breast) and identify areas of similarities and differences.

Study design

HUVEC



HuAoSMC



Treated with subset of drugs from the LINCS panel and other cardiovascular drugs

P100 and global chromatin profiling (GCP)

Inhibitor panel used in vascular cell lines

Epigenetic modulators	
Drug	Mechanism
Vorinostat	HDAC inhibitor
Decitabine	DNMT inhibitor
JQ1-S	Bromodomain inhibitor
GSK126	EZH-2 inhibitor
UNC-0646	G9a inhibitor
GSK-J4	Histone demethylase inhibitor
Resveratrol	SirT1 activator
Geldanamycin	HSP 90- inhibitor

Kinase inhibitors	
Drug	Mechanism
Pazopanib	VEGFR, PDGFR inhibitor
Nilotinib	BCR-Abl inhibitor
LY-294002	PI3-K inhibitor
Okadaic Acid	PP1 and PP2a inhibitor
Tofacitinib	Jak3 inhibitor
SP600125	Jnk inhibitor
Losmapimod	p38 MAPK
AR A014418	GSK3 inhibitor
Staurosporine:	Pan kinase inhibitor

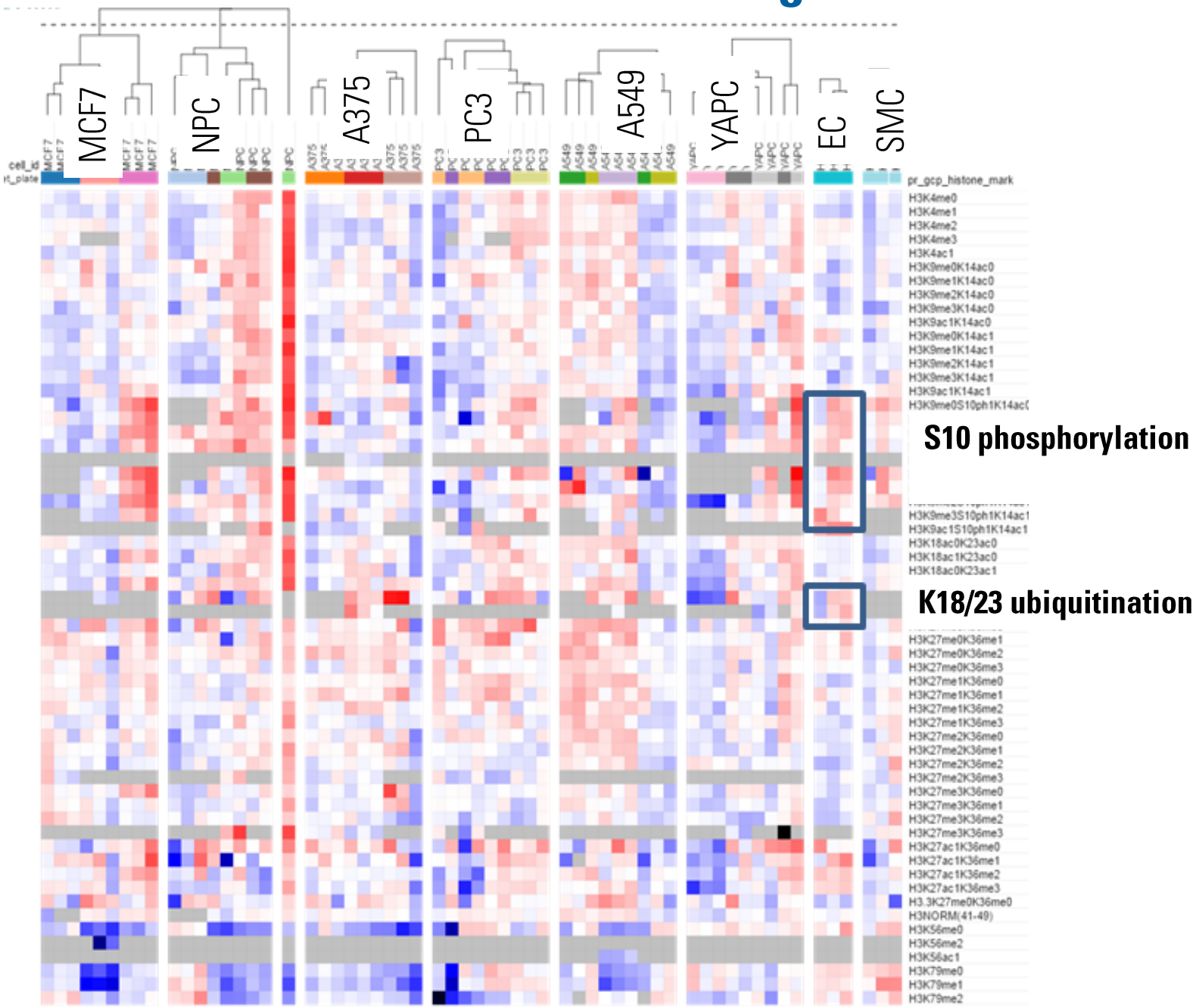
Others	
Drug	Mechanism
Tretnoin	RAR agonist
Tacrolimus	Calcineurin inhibitor
Sirolimus	mTOR inhibitor
Lenalidomide	Immunomodulator
Curcumin	NFKB inhibitor
KN-62	CaMKIIalpha inhibitor
Pravastatin	HMGCoA reductase inhibitor
Rolipram	PDE4 inhibitor

Cardiovascular drugs
Aldosterone
Angiotensin II
Eplerenone
Losartan
Spironolactone
Dexamethasone

Red: used in oncology, no vascular toxicity reported
Green: used in oncology, known vascular toxicity
Black: other compounds

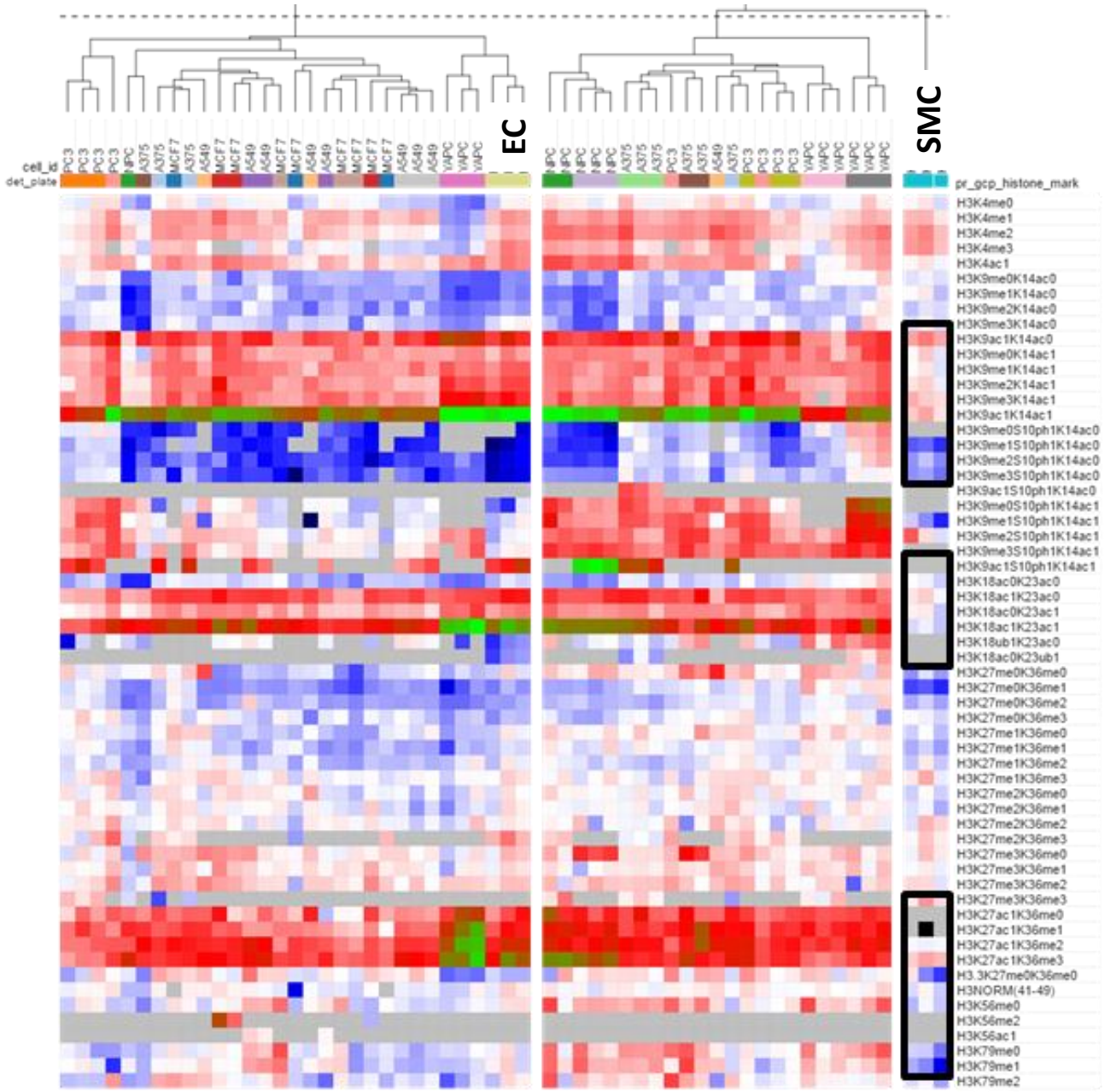
Chromatin profiles of vascular endothelial cells show distinct signatures

- At baseline, HUVECs have a distinct chromatin profile compared to other cancer cell lines and neuronal cells.
- HUVECs are most similar to SMCs.



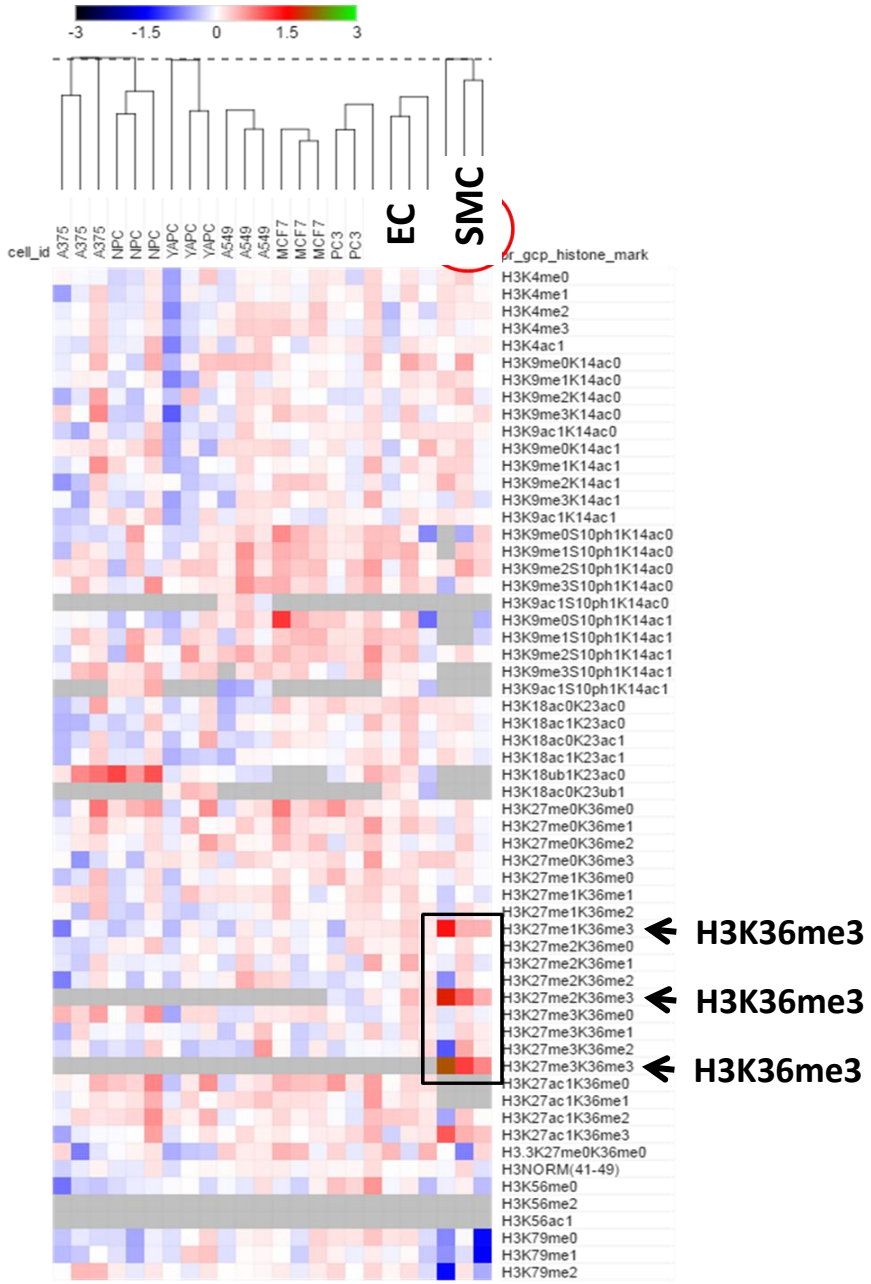
Aortic smooth muscle cells generate a unique chromatin signature to vorinostat

- Vorinostat, an HDAC inhibitor has the same signature in all cell types except SMCs.



Lenalidomide specifically increases H3K36 trimethylation in smooth muscle cells

- Lenalidomide is associated with increased venous thrombosis and used in the treatment of myeloma.



Pazopanib develops a multi-modal signature in vascular endothelial cells

Figure 1

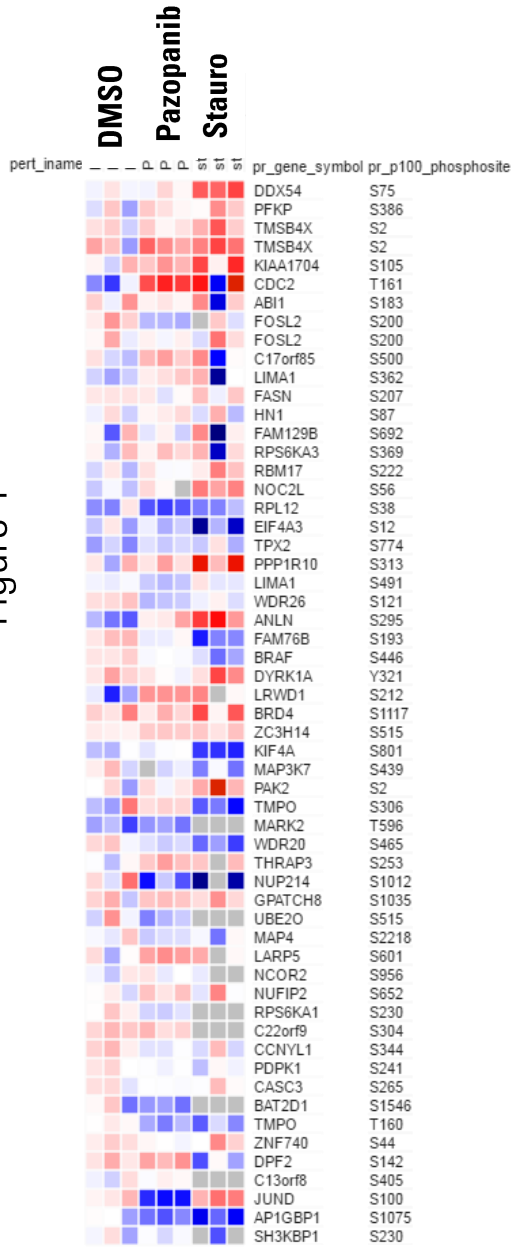


Figure 2

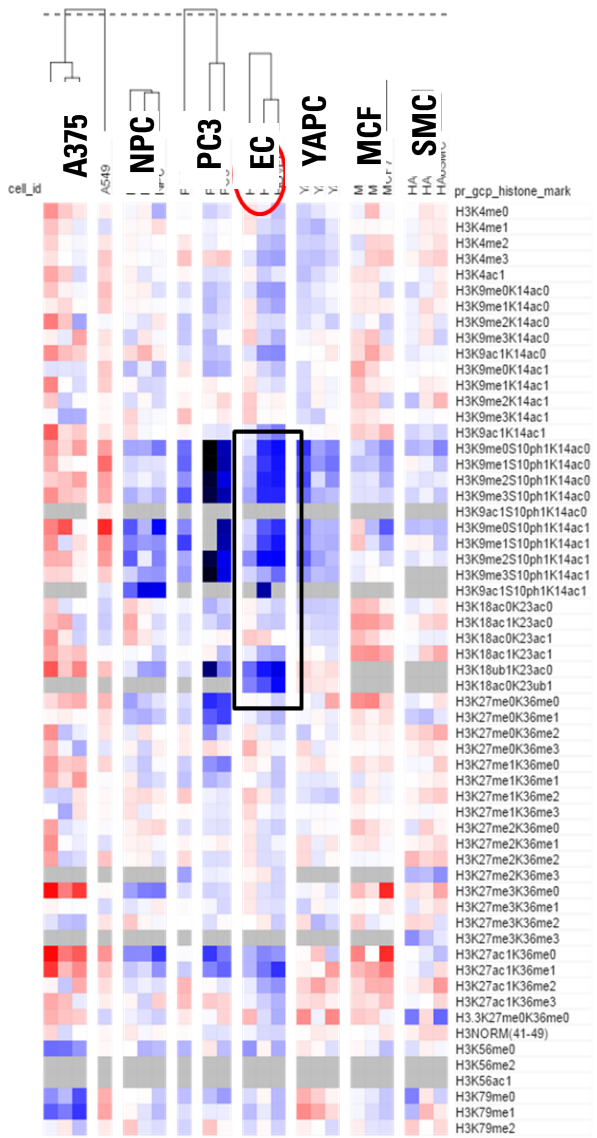


Figure 1: The P100 signaling signature of Pazopanib in HUVECs bears hallmarks of staurosporine (*within HUVECs*)

Figure 2: Chromatin signatures exhibit decreased S10 phosphorylation and K18/K23 ubiquitination (*compared across many cell lines*)

Taken together, these data suggest significant cytotoxicity of Pazopanib in HUVECs

Conclusions & future directions

- Anti cancer agents have direct effects on the chromatin profile and signaling pathways in vascular cells.
- Some drugs(vorinostat, lenalidomide, pazopanib) cause distinct effects in vascular cells when compared to cancer cells.
- Hypothesis generated from these assays will be tested in other models to identify critical pathways in vascular cells and identify possible mechanisms of vascular toxicity.
- Identification of critical pathways involved in vascular cells may help in the development of drugs to attenuate vascular effects without mitigating anti cancer effect.

Thank you!

LINCS PCCSE

IV: ROUNDTABLES

LINCS PCCSE publication roadmap

Title/Idea	Key Activities to Complete	Principal Authors	Timeframe
"Data Library" paper of all drug treatments in all cell types / Joint signatures GCP+P100 => crosstalk b/t Epi and Signaling; Signatures that unexpectedly cross P100 to GCP and vice versa; divergence among biological models, especially neural models	Finish making data! Find vignettes that highlight the multi-assay dimensionality of the data	Lev + All	Q1 2017
Universal Network-based Query for Integrating Omics Signatures	Algorithms are there, need some story or showcase	Ryan	6 months
Characterization of Cas9 neurons	Finish making them; demonstrate Cas9 activity; deeper characterization	Fatema, Jennie, Tak, Joel	6 months
Demonstration of DIA data mineability?	Duct taping algorithms into something that people can actually use	Jarrett/Sonia/Brian/Mike/Jake /Sebastian/Ryan	Start putting together in summer
PECAN - DIA Peptide Centric Search	None	Sonia	Submitted
DIA Overlap Paper	Comparative analysis	Jarrett	
AutoQC Panorama		Vagisha/Brendan/Adam/Josh Eckels	
DIA Superlibrary Generation	Comparative Statistics, expand examples	Jarrett/Sebastian	
CHORUS as a LINCS Data Repository and Analysis Platform (for DIA, for PRM?) Tutorial? Application Note? => Interactivity with Skyline (including direct interaction)	Lots of software development needed	Mike/Jarrett/Brendan	
Super-next-gen DIA Query algorithm	Proof-of-principle experiments, hardening of code	Ryan	
MultiOMICS signature data ecosystem		Together with Transcriptomics	
Look for focused neuro-results in LINCS data for specific follow up in LHT lab			

SAB guidance questions

- What are the key impact opportunity areas for our center going forward? How should we allocate resources?
  - Short term?
  - Longer term?
- How do we catalyze use of the data we have generated by others?
- Are there key areas of focus that have been overlooked? Could these lead to greater engagement with key communities?
- Have we addressed all major mid-course review criteria? Are there major course corrections needed?