Case Studies in Quantitative Proteomics: When PRM is not enough – Hybrid PRM-DIA Assays

Jake Jaffe

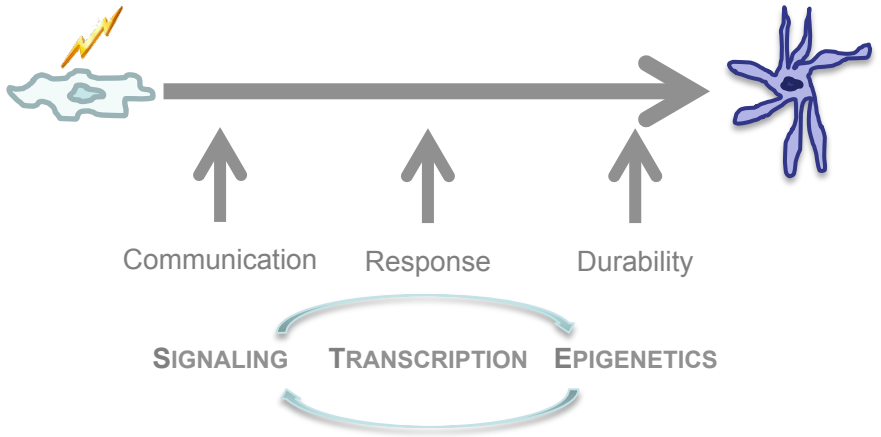
ASMS Short Course 2018

Outline

- Scientific Motivation
- Mechanics of implementation
- DIA data primary interpretation strategies
  - Data refinement
  - Spectral library importance
- Evaluating DIA results and gaining biological insights

LINCS: A library of perturbational signatures

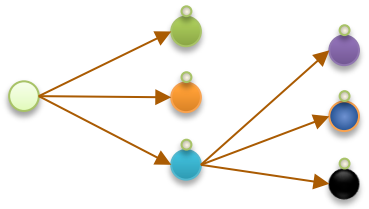
- Library of Integrated Network-based Cellular Signatures
  - Proteomic Characterization Center for Signaling and Epigenetics
  - "Poke" cells
    - Compounds (epi-active, pathways)
    - Genes
  - Measure molecular readouts
    - Epigenetics (GCP)
    - Signaling (P100)
  - Diverse model systems
    - Cancer cell lines, neuronal lineages, vascular primary cells



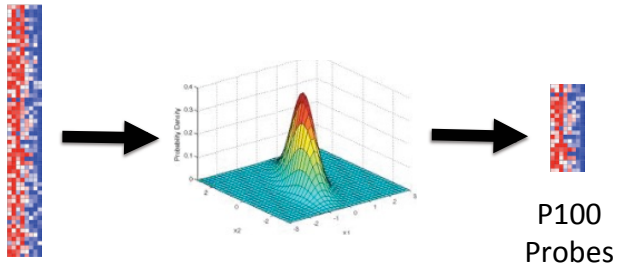
- Connectivity: The Master Reference of Cellular Activity
  - Drug to genes, drugs to other drugs, off targets effects, anticorrelations
  - Signaling to transcription to epigenetics

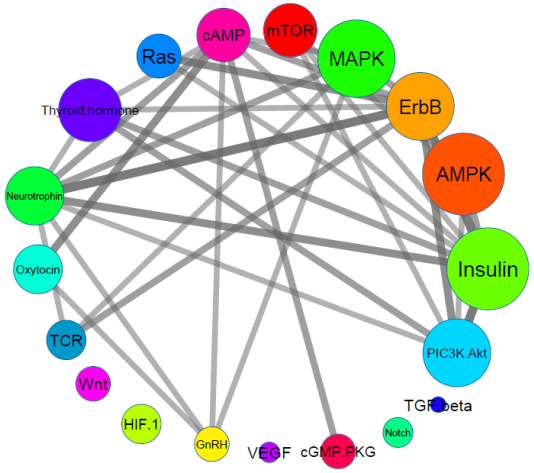
P100 Reduced-representation Phosphosignaling Assay

- The “P100” assay is our signature proxy for signaling
  - Biochemistry: automated phosphopeptide enrichment (x 96 samples)
  - 96 phosphopeptide targets
  - Labeled peptide internal standards for each analyte
  - Fully targeted and scheduled for heavy and light of all analytes (PRM)
  - Optimal transitions selected ***by hand*** from real data
  - Q-Exactive data acquisition



Kinase:Substrate :: 1:many





Pathways Represented
(by proxy)

The attraction and challenges of comprehensive MS

- We are creating a huge resource
  - > 10,000 samples in 6 years
  - All undergo phosphopeptide enrichment
  - Comprehensive MS places high value on these samples

- We could measure pathways directly rather than by proxy
  - And add new target analytes on demand
  - Generate arbitrary signature panels as necessary

- But it is still highly valuable to have consistent analytes across all samples
  - Measured and quantified in a reproducible manner, with internal standards

Solution: a *Hybrid P100-DIA* Phosphoproteomic Assay

- Continue to measure our core analytes

- Take advantage of existing assay infrastructure
  - Labeled internal standards
  - Bioinformatics and data reduction workflows

- **How can we insure that the answers we obtain with the hybrid assay are compatible with the original targeted assay?**
  - Are results equivalent?
  - Is quality maintained?

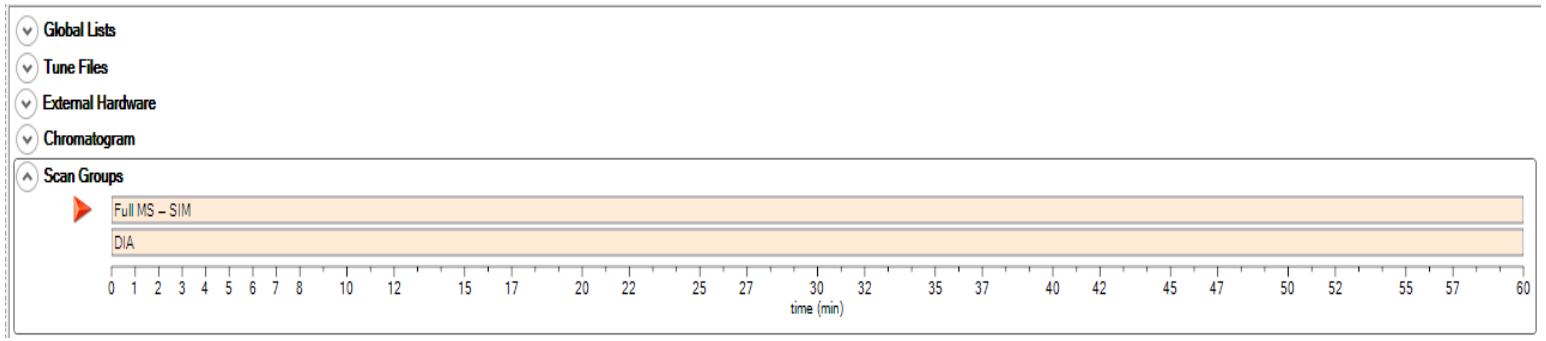
Major Challenges / Questions

- What is optimal acquisition method?
  - Duty cycle vs. sensitivity/selectivity
- How do we get same answer as with fully targeted HR-MRM?
  - Can’t necessarily just use same transitions
  - Some are forbidden!
  - But we can potentially use more!
- What gets lost?
  - Overall ability to detect probes
- How to mine data and quantify?
  - For the future!

This DIA method

- QExactive HF
- 22 m/z windows, 11 m/z offset between cycles
- Deconvolve windows to 11 m/z (effective)
- 30k resolution
  - Better impedance match to fill times
  - Increase signal-to-noise ≈√transient length
- m/z range 400-1000, 28 windows (x 2 for overlap)
- MS1 followed by 20 MS2
  - 0.26 minutes for 10 cycles
  - traverse range ~7 times
  - around 2.1 sec to traverse range
- ***Compare with***: Fully targeted 1.7 m/z isolation +0.3 m/z offset
  - ***Back-to-back sample injections***

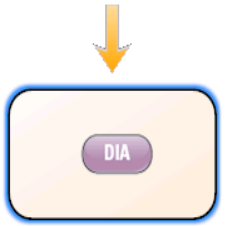
Overlap DIA Method: 400-1000 m/z mass range



Properties

Properties of the method

Global Settings	
User Role	Standard
Use lock ma	off
Chrom. peal	15 s
Time	
Method dura	60.00 min



Properties of Full MS —

General	
Runtime	0 to 60 min
Polarity	positive
Full MS — SIM	
Resolution	60,000
AGC target	3e6
Maximum IT	20 ms
Scan range	300 to 1200 m/z

Properties of DIA

General	
Runtime	0 to 60 min
Polarity	positive
Default char	4
DIA	
Resolution	30,000
AGC target	1e6
Maximum IT	50 ms
Loop count	27
MSX count	1
MSX isochro	on
Isolation wir	22.0 m/z
Fixed first m	—
NCE / stepp	27

56 Masses on Overlap Inclusion List

- 411-1005 mz (28 masses)
  - Δ22 mz between masses
- 400-994 mz (28 masses)
  - Δ22 mz between masses

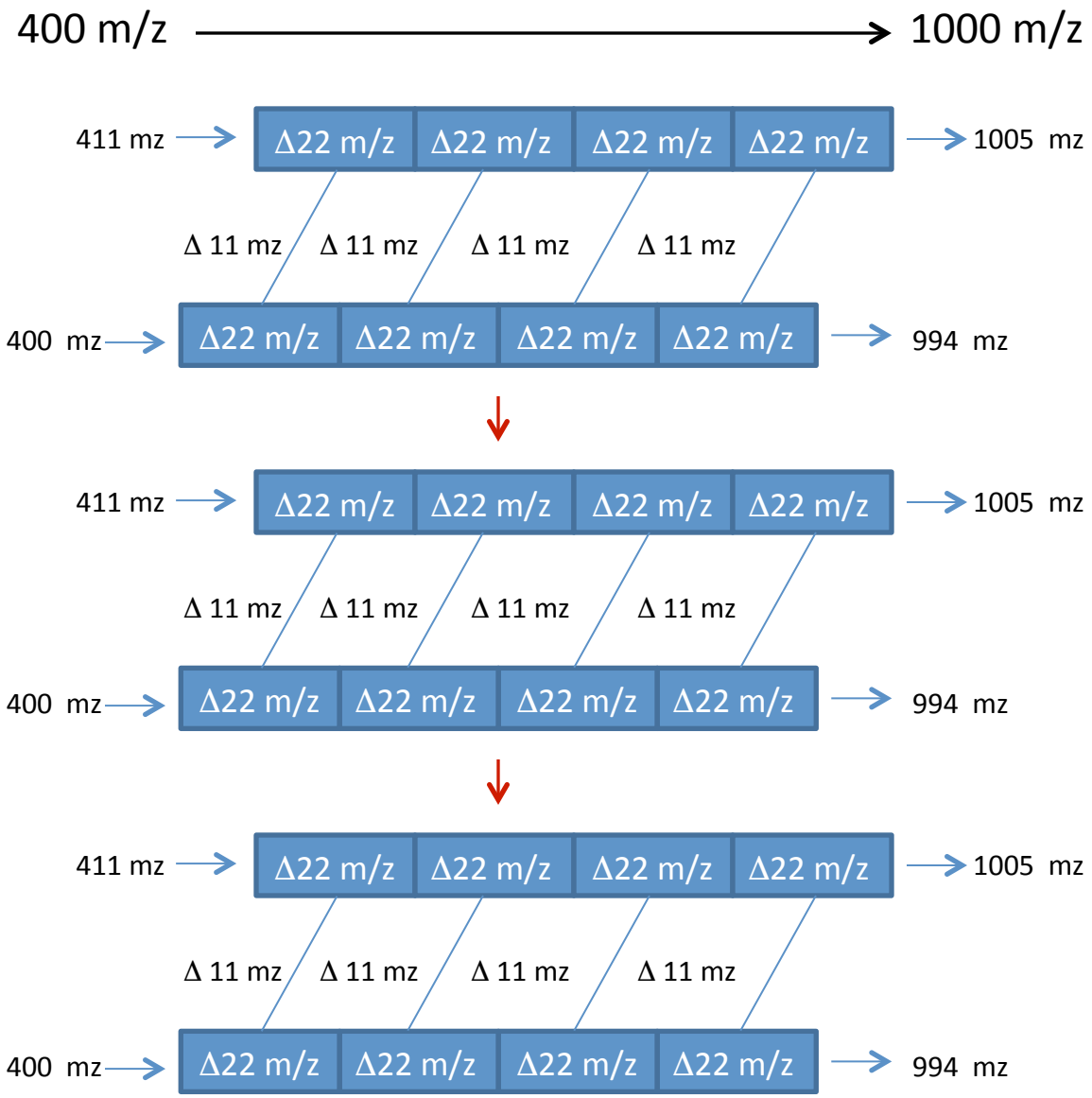
Jarrett Egertson, Jenn Abelin

Overlap DIA Method: 400-1000 m/z mass range

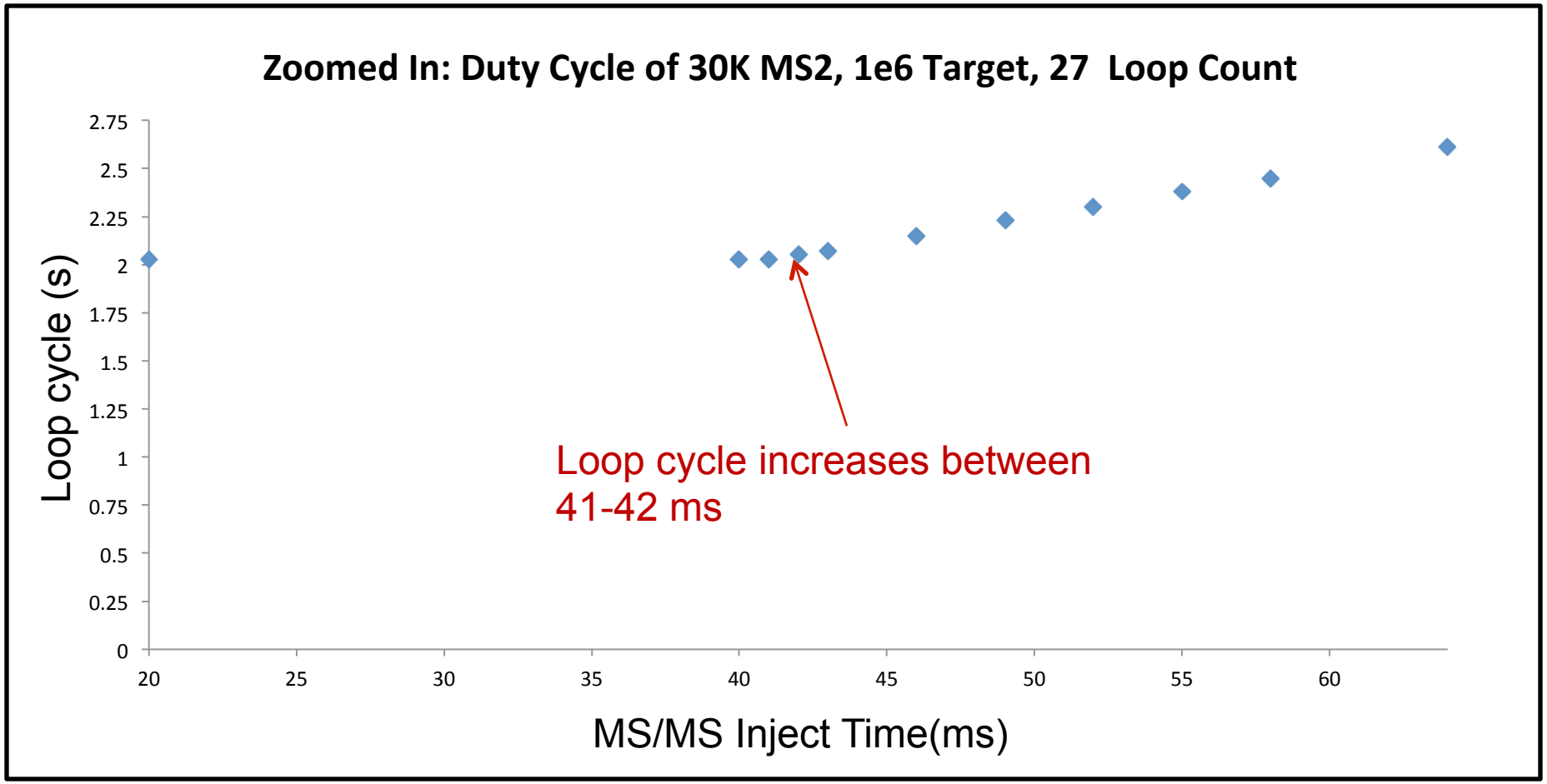
56 Masses

Mass [m/z	Formula [	Formula t	Species	CS [z]	Polarity	Start [min	End [min]	NCE	Comment
411.4369					Positive				
433.4469					Positive				
455.4569					Positive				
477.4669					Positive				
499.4769					Positive				
521.4869					Positive				
543.4969					Positive				
565.5069					Positive				
587.5169					Positive				
609.5269					Positive				
631.5369					Positive				
653.547					Positive				
675.557					Positive				
697.567					Positive				
719.577					Positive				
741.587					Positive				
763.597					Positive				
785.607					Positive				
807.617					Positive				
829.627					Positive				
851.637					Positive				
873.647					Positive				
895.657					Positive				
917.667					Positive				
939.677					Positive				
961.687					Positive				
983.697					Positive				
1,005.71					Positive				
400.4319					Positive				
422.4419					Positive				
444.4519					Positive				
466.4619					Positive				
488.4719					Positive				
510.4819					Positive				
532.4919					Positive				
554.5019					Positive				
576.5119					Positive				
598.5219					Positive				
620.5319					Positive				
642.542					Positive				
664.552					Positive				
686.562					Positive				
708.572					Positive				
730.582					Positive				
752.592					Positive				
774.602					Positive				
796.612					Positive				
818.622					Positive				
840.632					Positive				
862.642					Positive				
884.652					Positive				
906.662					Positive				
928.672					Positive				
950.682					Positive				
972.692					Positive				
994.702					Positive				

DIA Overlap Window Scheme



Parallel operation parameters affect choice of duty cycle

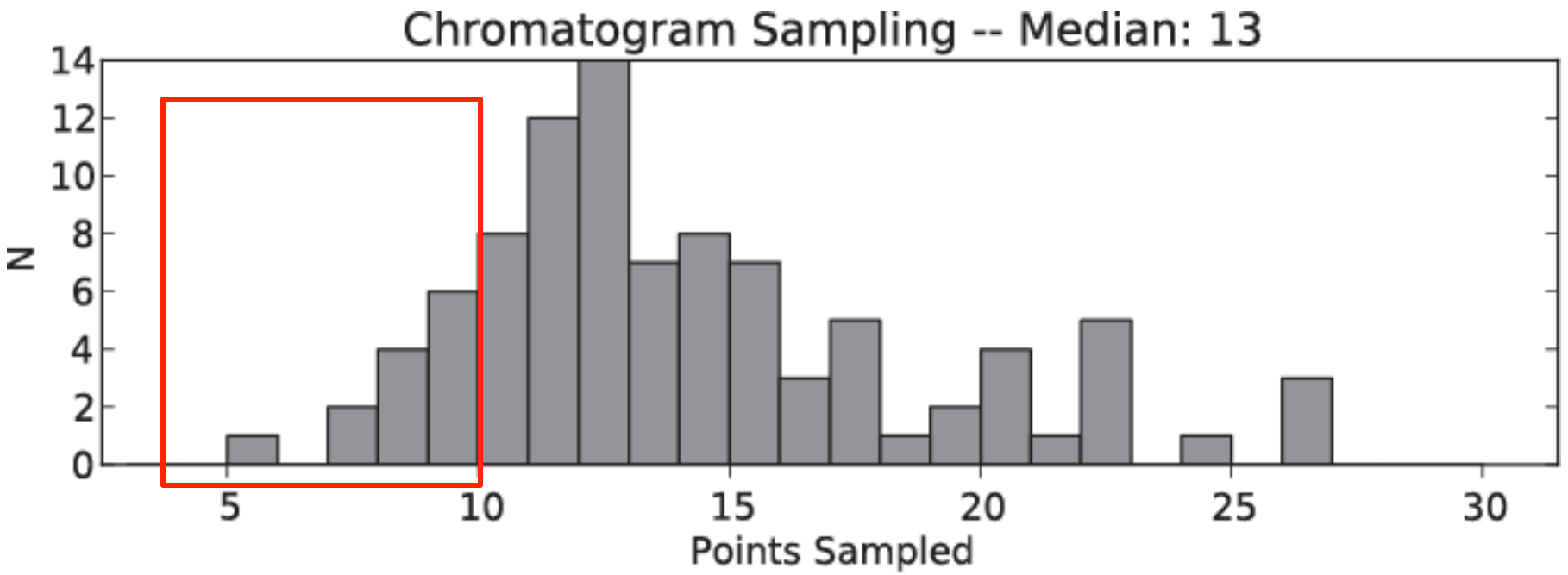


41 ms max IT allows you to maintain parallel operations, but additional inject time will improve MS2 quality with only a small increase in duty cycle time

Jenn Abelin

Points across peak are generally sufficient in DIA

Number of Points Across Peak

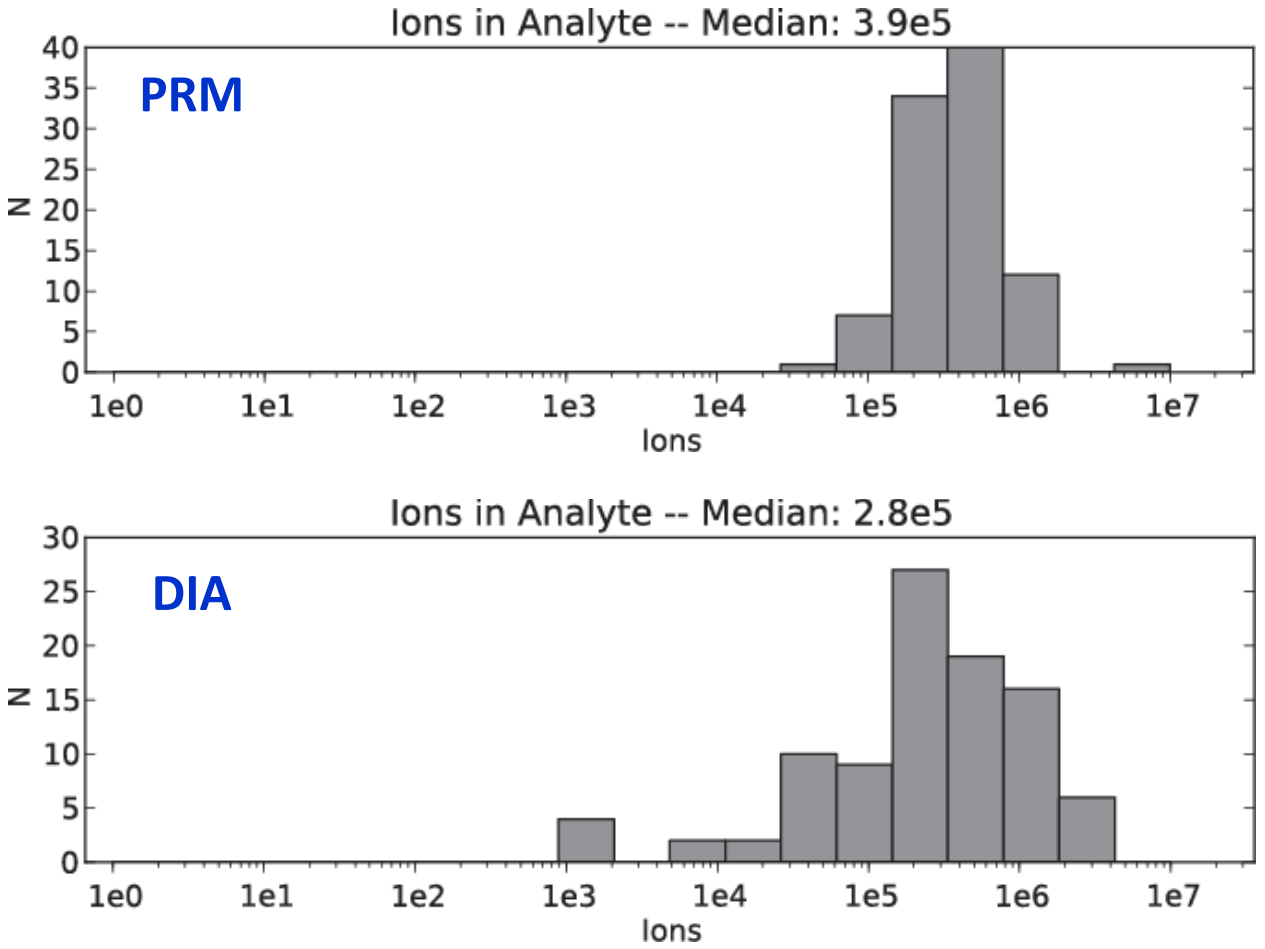


Aggregate statistics across 95 peptide analytes in a typical LCMS run

Jarrett Egertson

Fewer ions in MS/MS scans derived from desired analytes

Estimates of actual number of (unscaled) charges detected from target analytes:



Aggregate statistics across 95 peptide analytes in a typical LCMS run

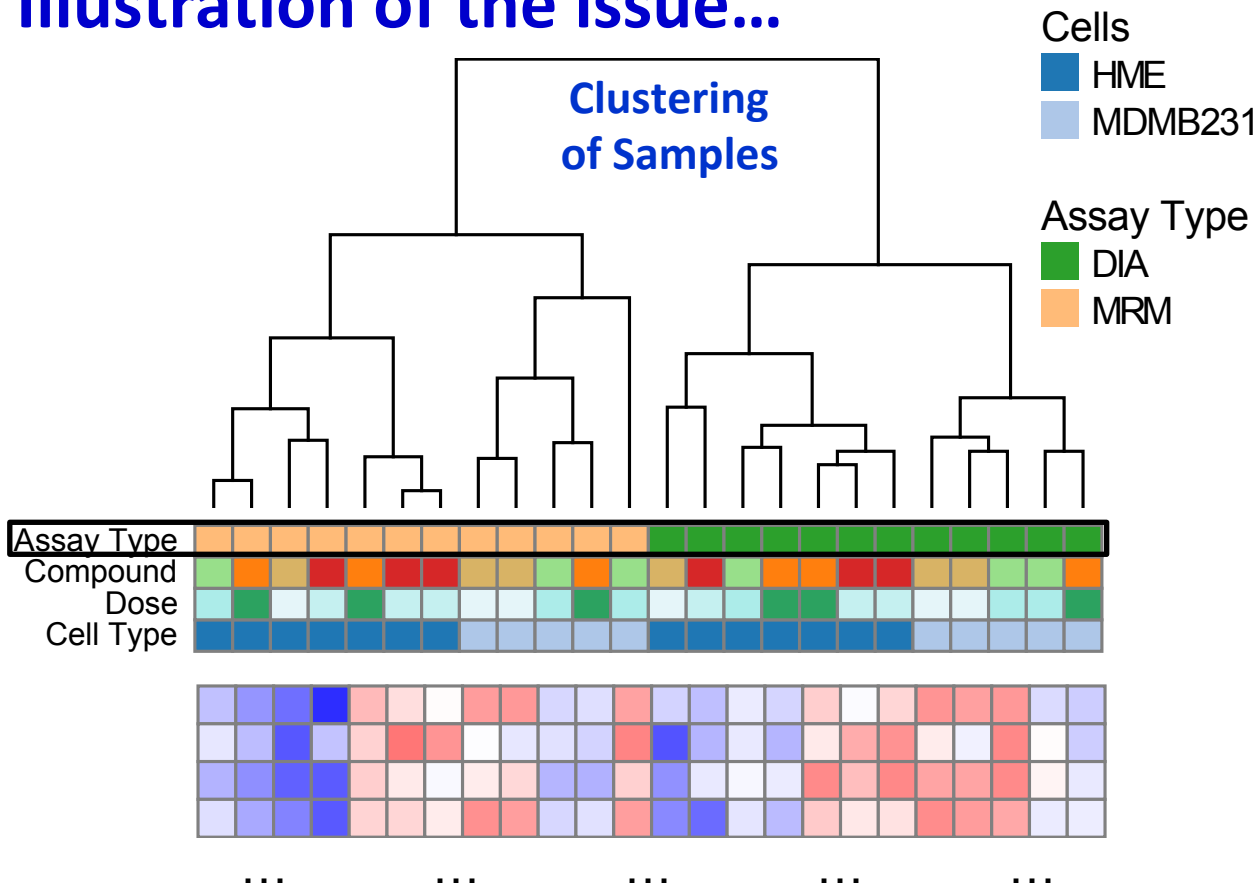
Jarrett Egertson

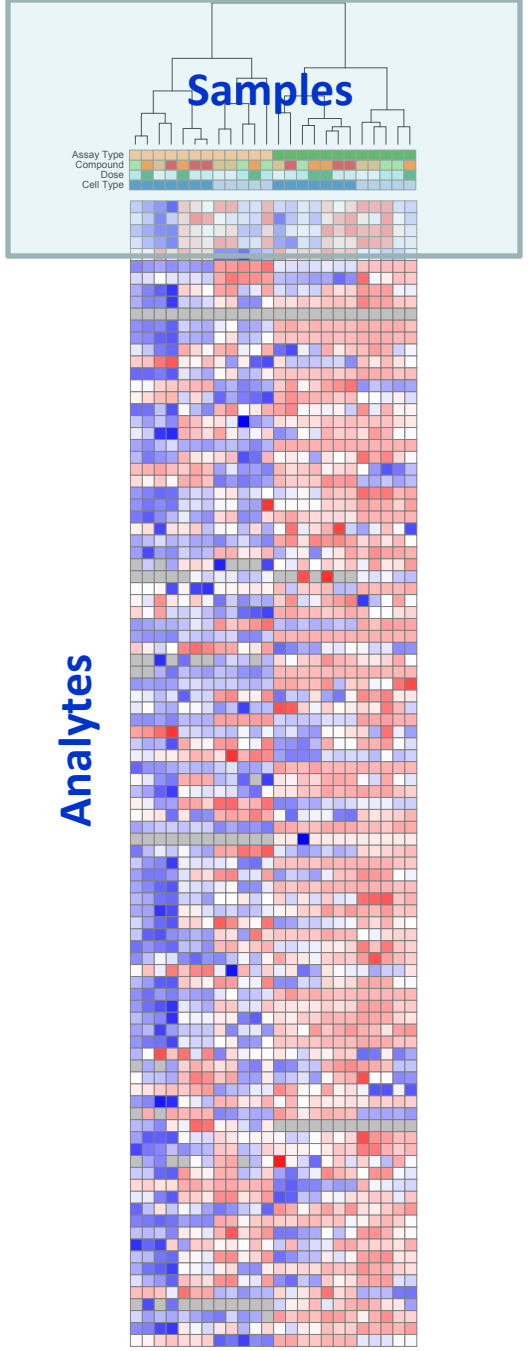
$64,000 Question:

- Do the assays give the same answer?

- No.

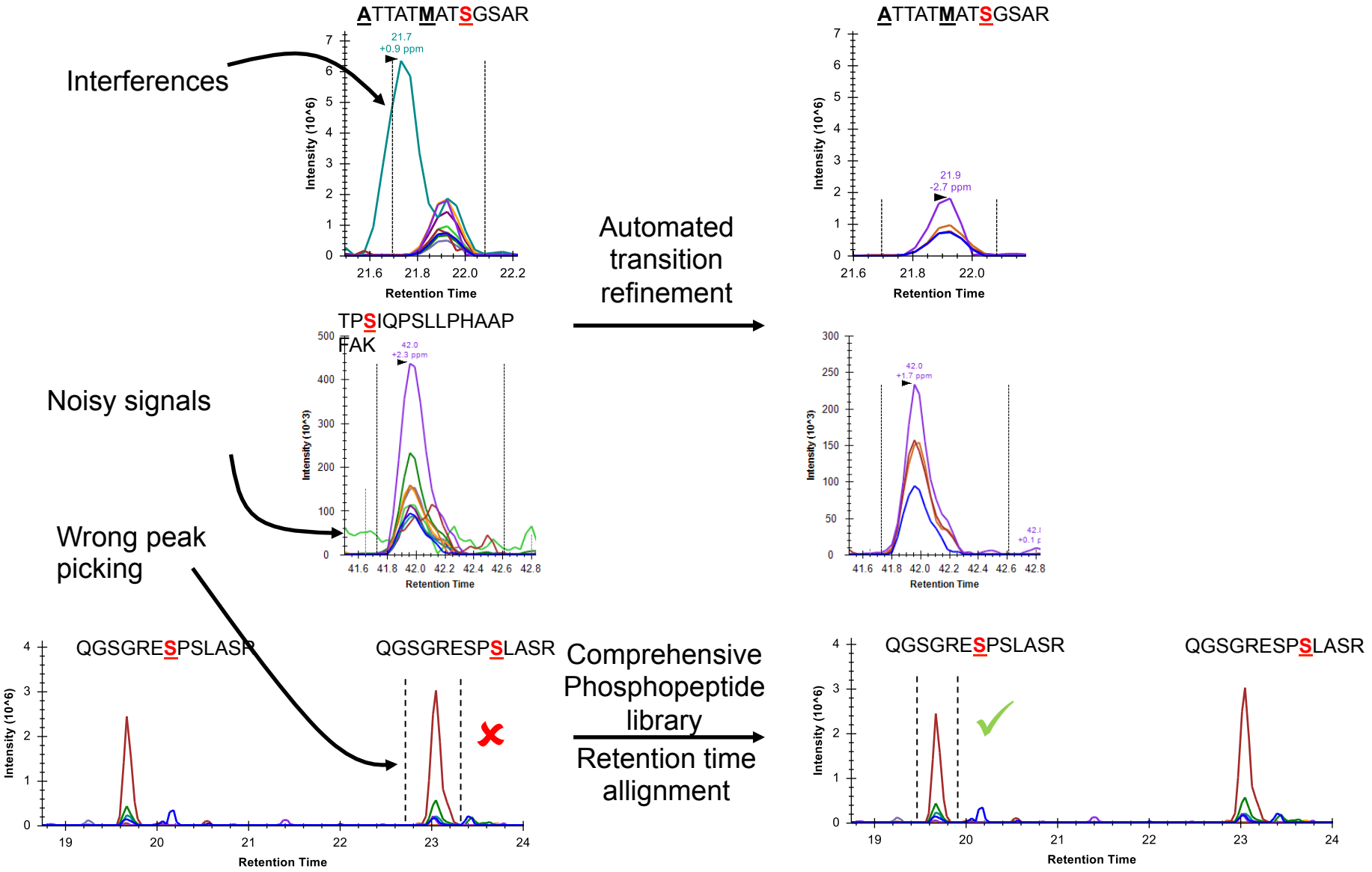
Illustration of the issue…





- 12 samples acquired with both PRM and DIA methodologies
- 2 different cell types
- Cells also treated with compounds

DIA data can be extremely noisy



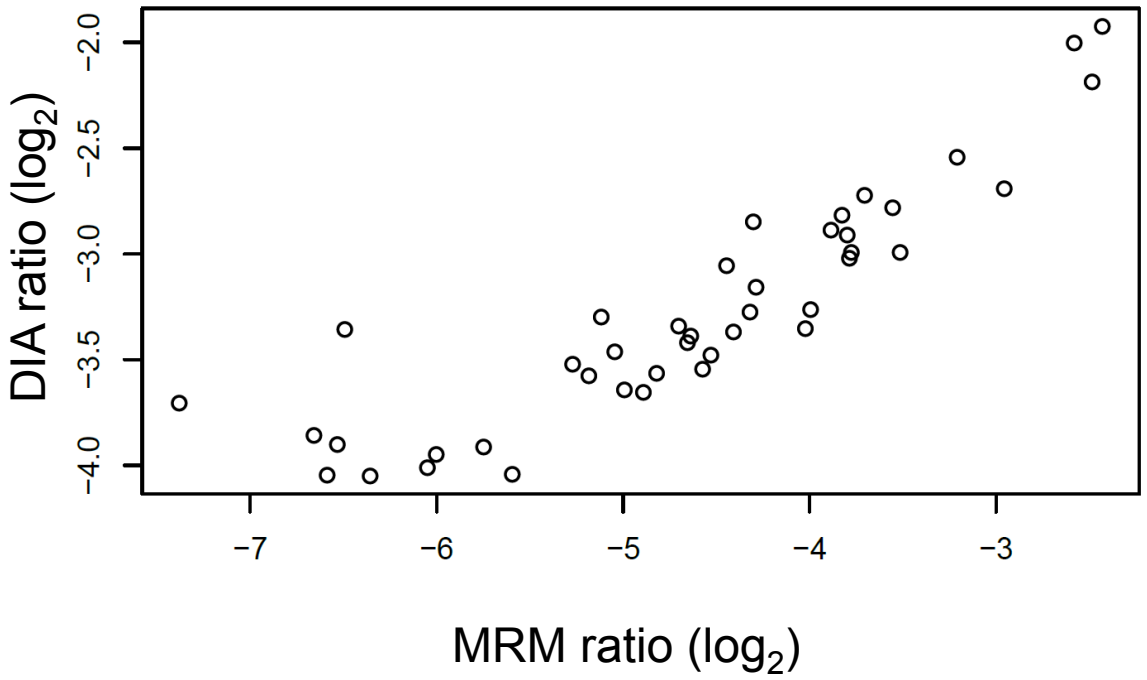
Sebastian Vaca

Some of the issues

- Many transitions now prone to interference in DIA setting
  - Low b, y
  - We sometimes use these for site localization
- Many heavy standards now co-isolating with light endogenous analytes
  - Almost all b ions invalid unless there is a missed cleavage
- Lower overall signal for any given analyte due to co-isolation of many peptides

DIA ≠ PRM

(Example: ALGS[+80]PTKQLLPC[+57]EMAC[+57]NEK)

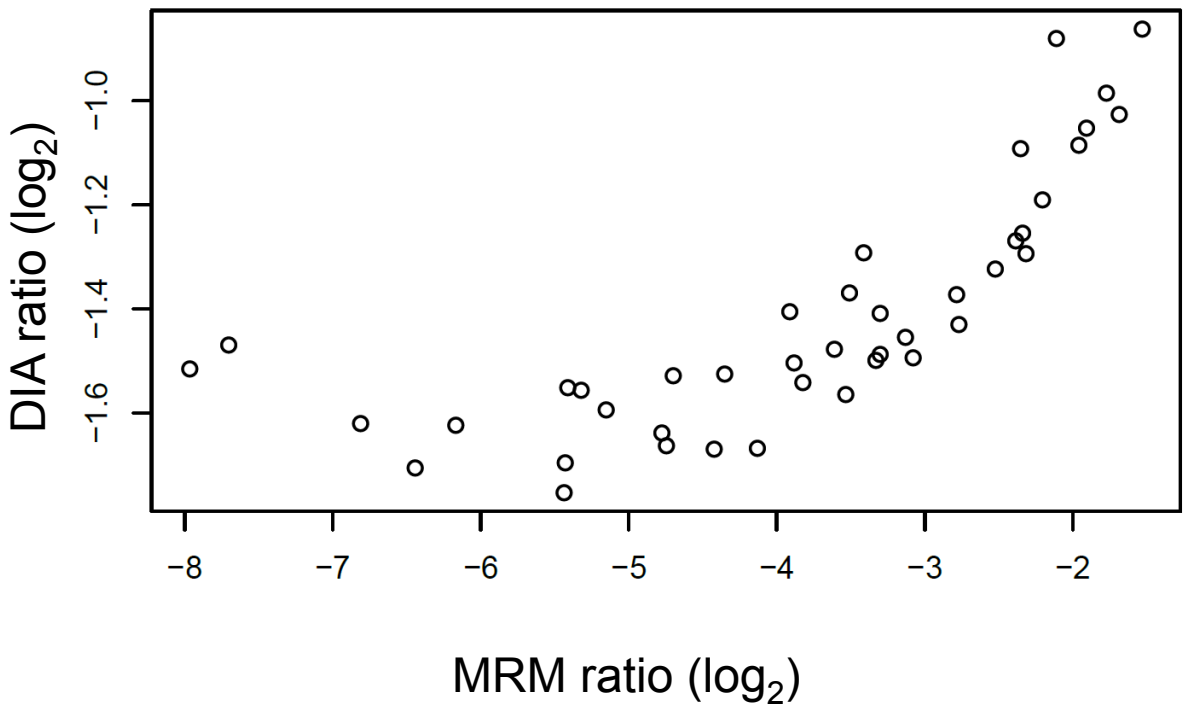


$y = 0.39x - 1.43$

$r^2 = 0.78$

DIA ≠ PRM

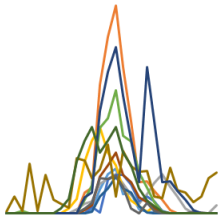
(Example: SPS[+80]PAHLPDDPKVAEK)

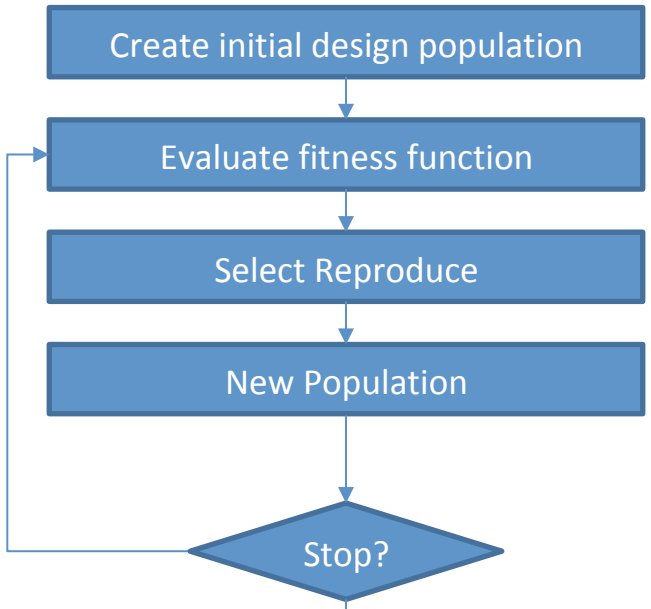


$y = 0.11x - 1.01$

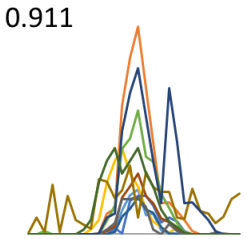
$r^2 = 0.55$

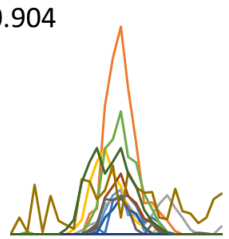
Genetic algorithm transition refinement: Evolving towards accurate measurements

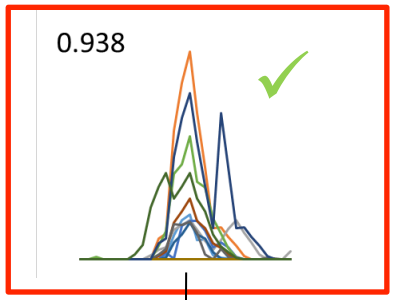




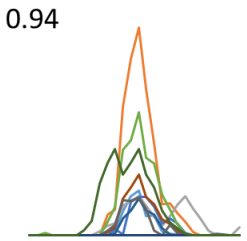
Generation 1

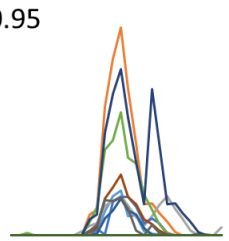


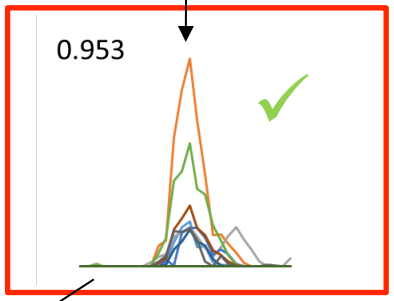




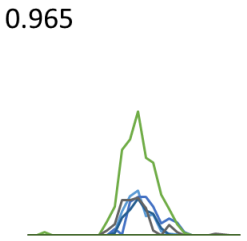
Generation 2

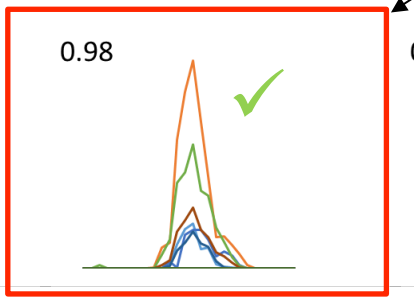


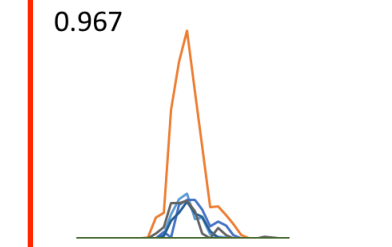




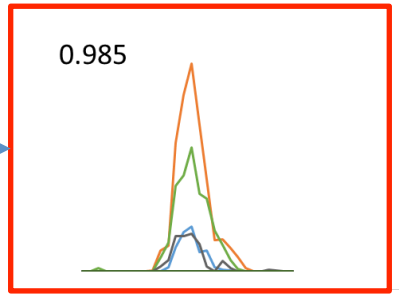
Generation 3





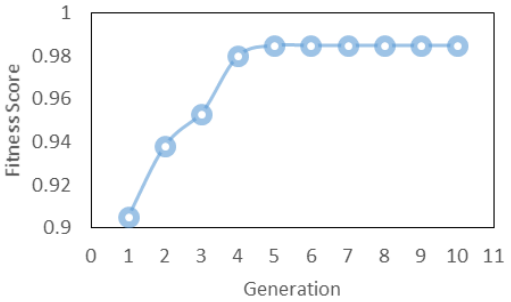


Generation N



Fitness function:

- Intensity
- Similarity between transitions

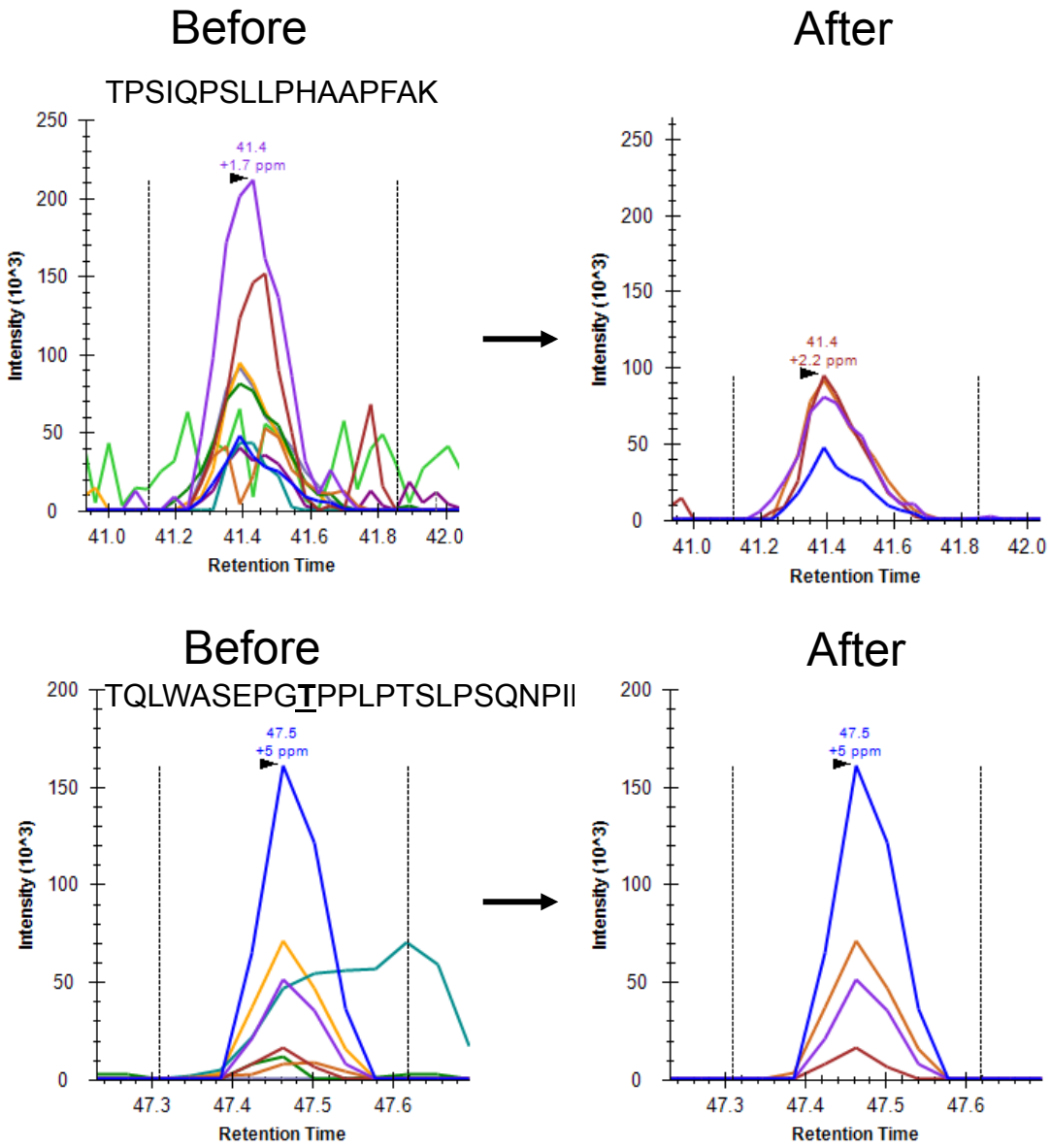


Best set of transitions

- High intensity
- Clean signals

Sebastian Vaca

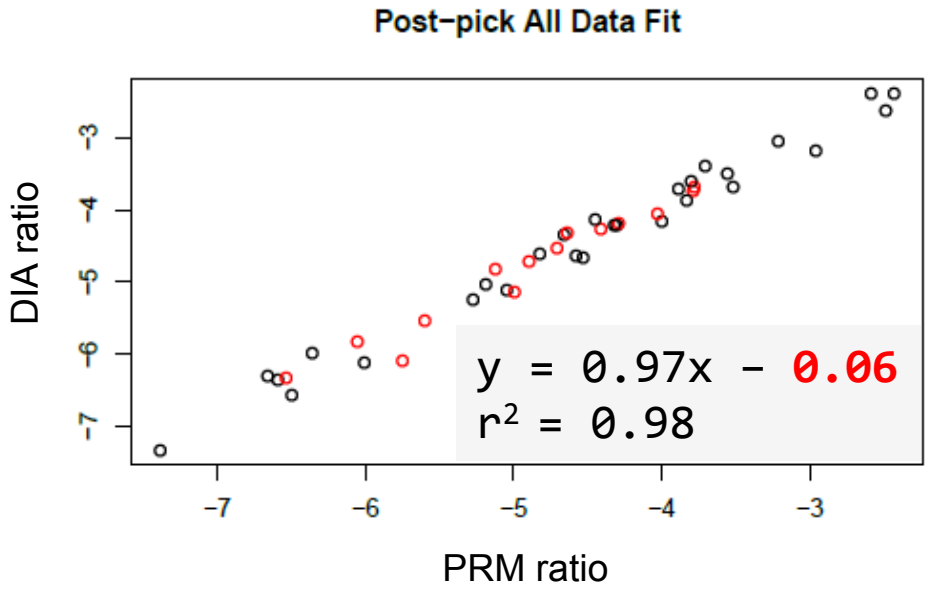
Genetic algorithm transition refinement: Evolving towards accurate measurements

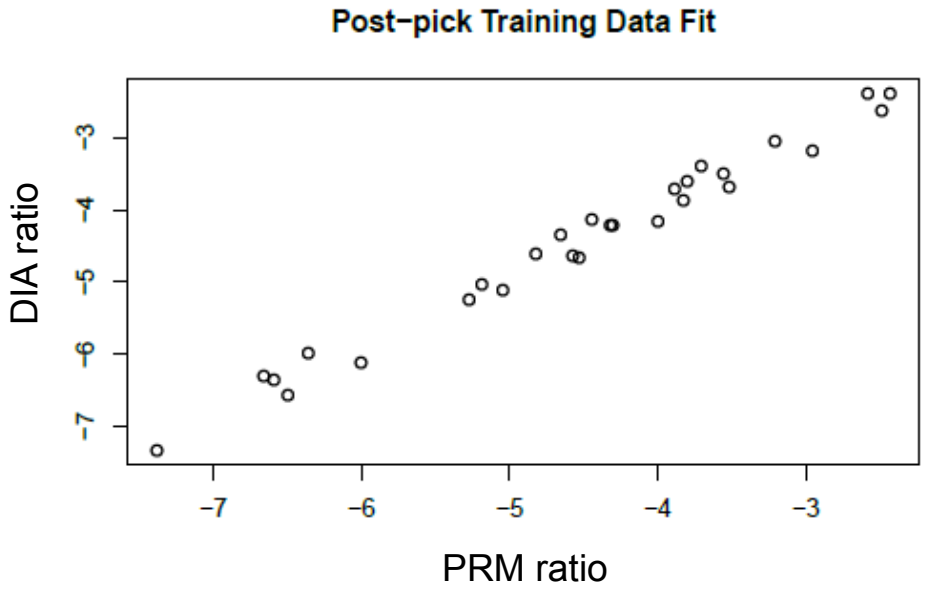


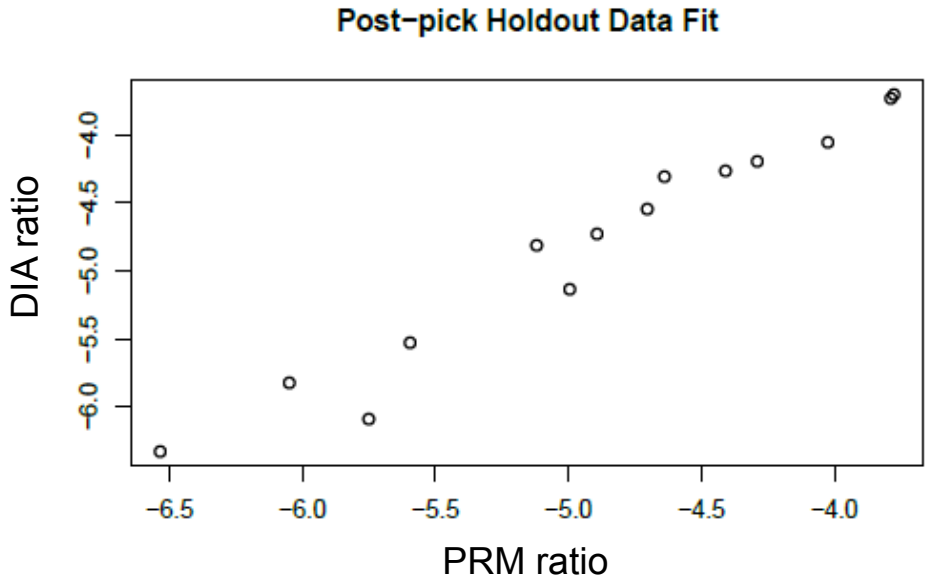
See poster: Sebastian Vaca "Avant-garde" MP 357

Sebastian Vaca

Now DIA = PRM (Example: ALGS[+80]PTKQLLPC[+57]EMAC[+57]NEK)





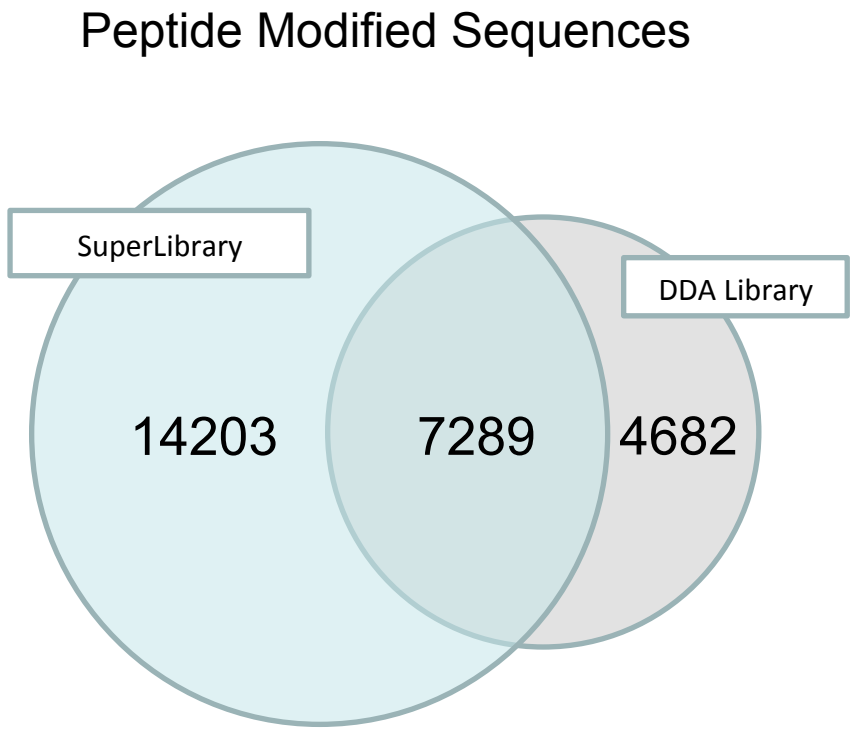


Model contains
33 transitions:

y11 z=1	y15 z=2	b8 [-98] z=1	b9 [-98] z=2	b11 [-98] z=3
y10 z=1	y14 z=2	b10 [-98] z=1	b11 z=2	b12 z=3
y8 z=1	y10 z=2	b11 z=1	b12 [-98] z=2	b13 z=3
y5 z=1	y9 z=2	b12 z=1	b14 z=2	b14 z=3
y4 z=1	y8 z=2	b7 z=2	b16 z=2	b14 [-98] z=3
y3 z=1	y7 z=3	b8 z=2	b17 [-98] z=2	
y17 z=2	b7 [-98] z=1	b8 [-98] z=2	b9 z=3	

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-labelled standard peptides
- Approach:
  - Inject sample 6 times, each with 100 m/z precursor range
    - 400-500 m/z, 500-600 m/z, etc.
  - Sample across range with 4 m/z windows, 50% overlap alternating
- Identification of:
  - 21492 modified peptides sequences (~x4 additional IDs compared to DDA)
  - 23945 precursor ions (~x3 additional IDs compared to DDA)
  - 17429 Phosphopeptides (+61% compared to DDA)
  - 5543 Confidently localized phosphosites (+31% compared to DDA)
  - **Good MS/MS even when MS1 precursor is poor!**
  - **Great identification of positional isomers!**

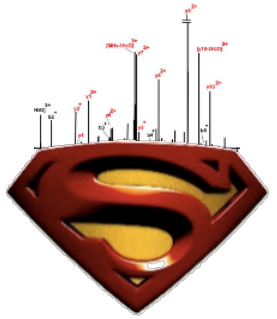


Sebastian Vaca

High-quality data enabled by adapted quantitative proteomics tools

Spectral library generation

MS Method	Search Engine	Validation	Formatting
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)	**Skyline** • Generate spectral library



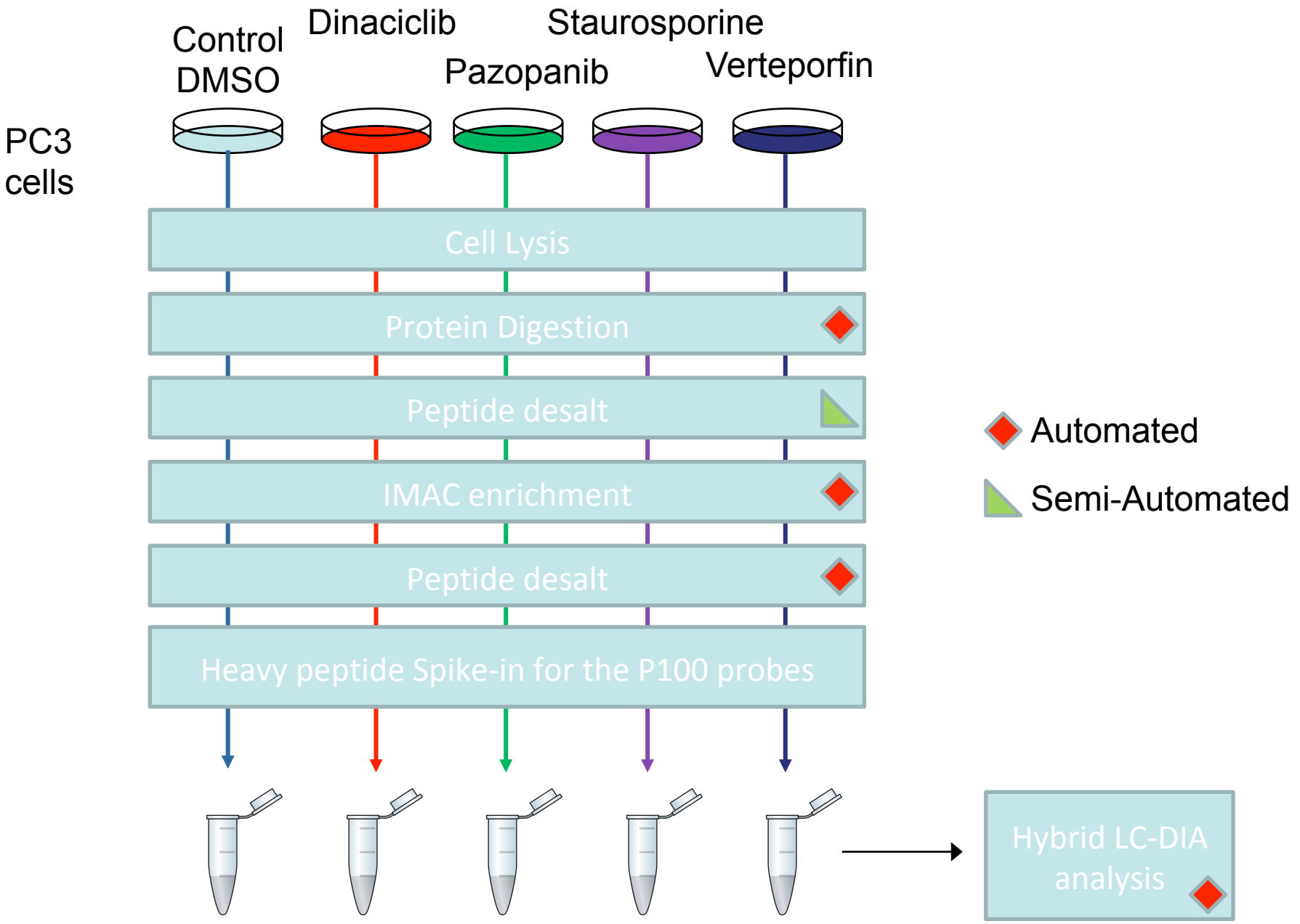
Super Spectral Library

DIA peptide query and data refinement

MS Method	Peptide Query	Signal Extraction	Refinement
P100 DIA runs • Overlapped DIA • 28 x 22 m/z	**EncyclopeDIA** • Peptide identification • Retention time alignment • FDR<1%	**Skyline** • Chromatogram extraction	**GA algorithm** • Transition refinement • Filtering data

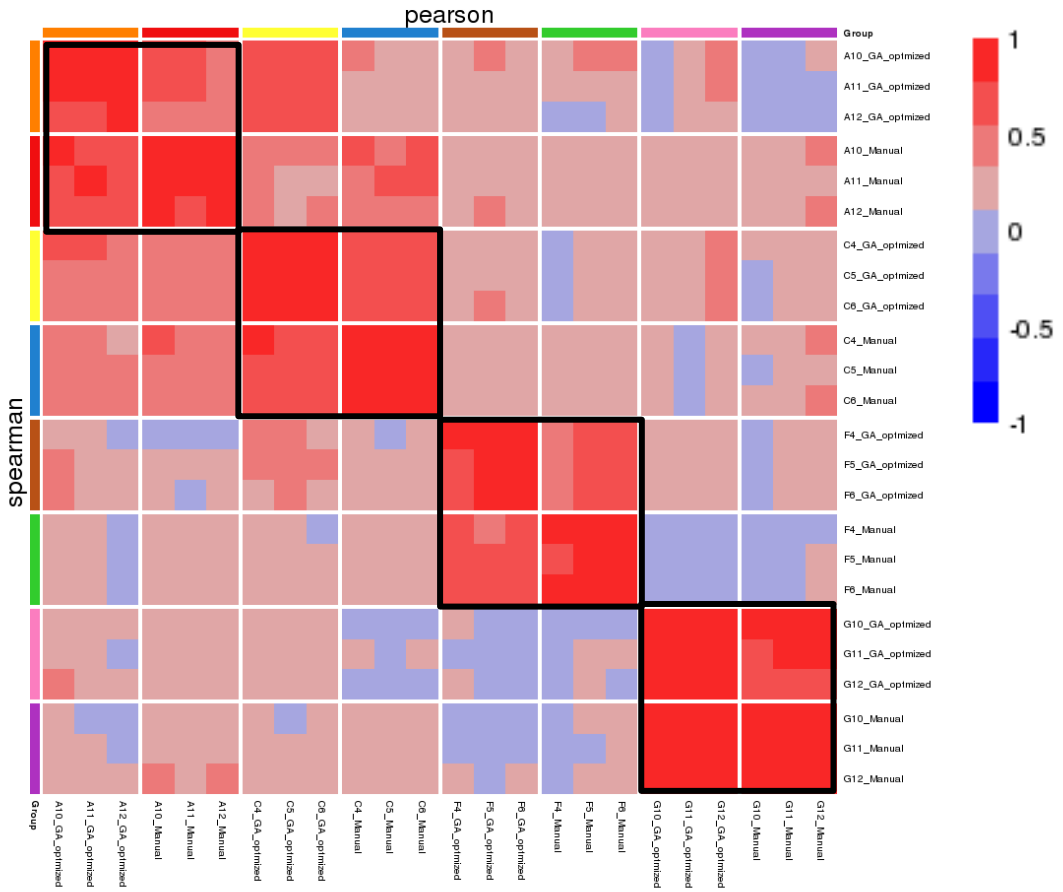
Sebastian Vaca

Validating our approach with manually curated P100 data

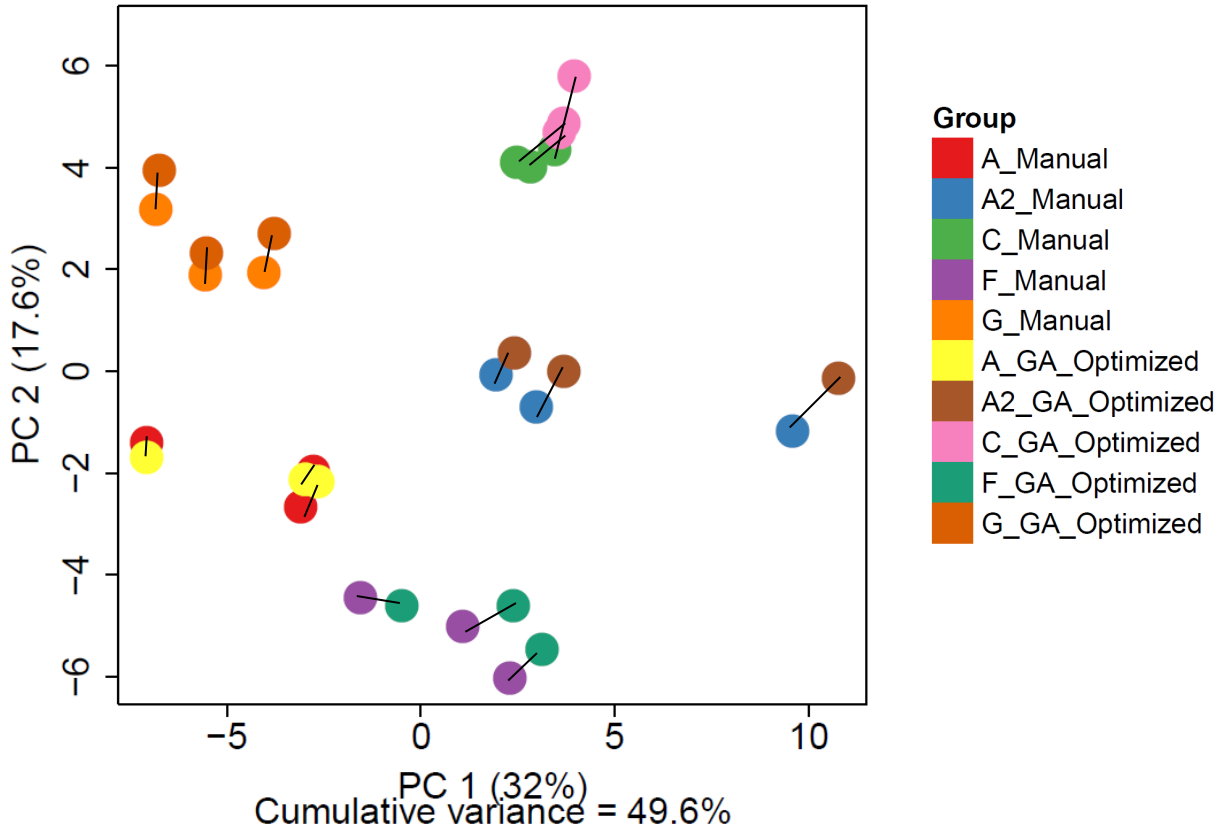


Automated data refinement matches manually validated results

- Expert Manual validation dataset vs. Automated transition refinement dataset



Good correlation between corresponding samples with manual and automated approach

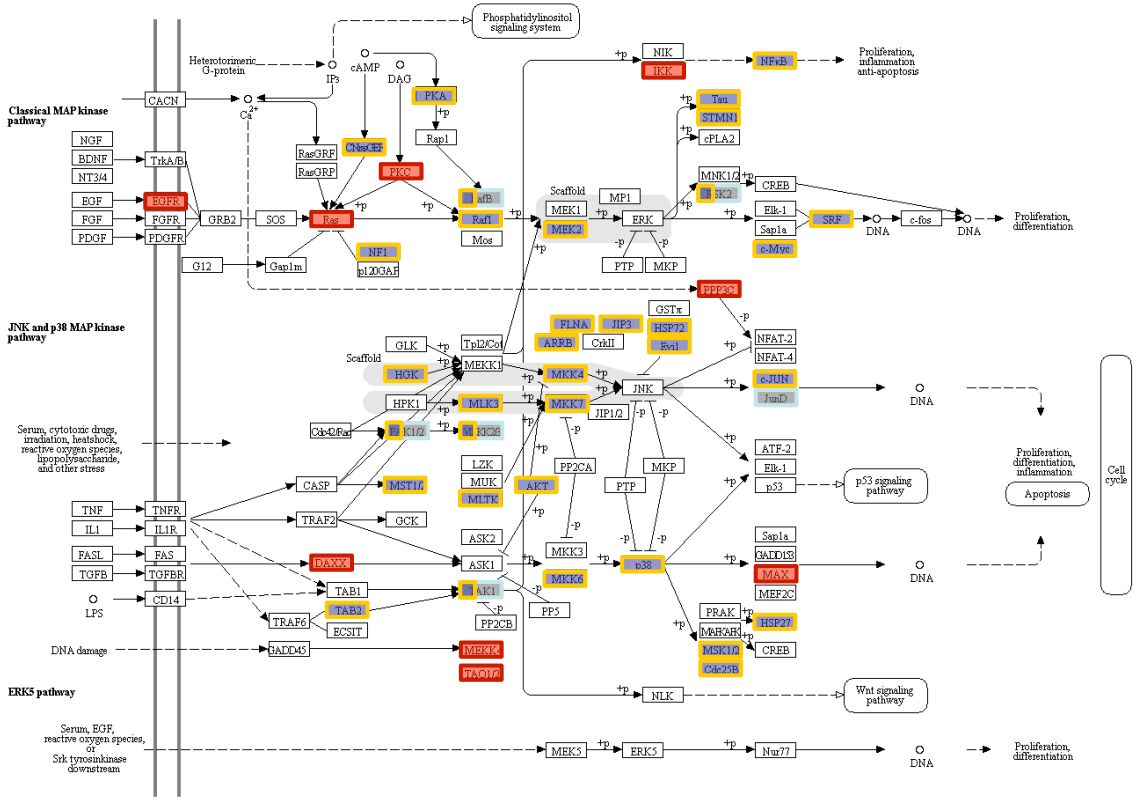


Corresponding samples with manual and automated approach cluster together

Sebastian Vaca

Our strategy significantly increases signaling pathway coverage

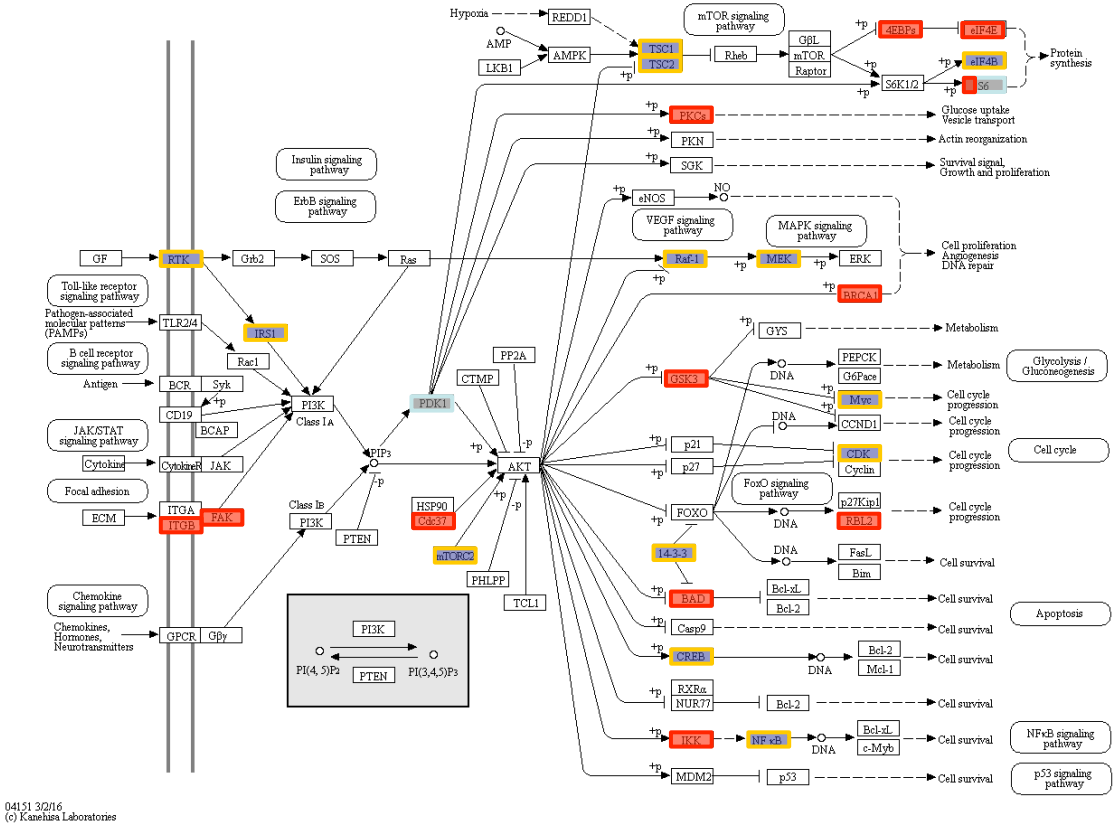
MAPK SIGNALING PATHWAY



6 pathway nodes

33 new pathway nodes

PI3K SIGNALING PATHWAY



2 pathway nodes

21 new pathway nodes

P100 probes

Additional targets from DIA analysis (DDA Library)

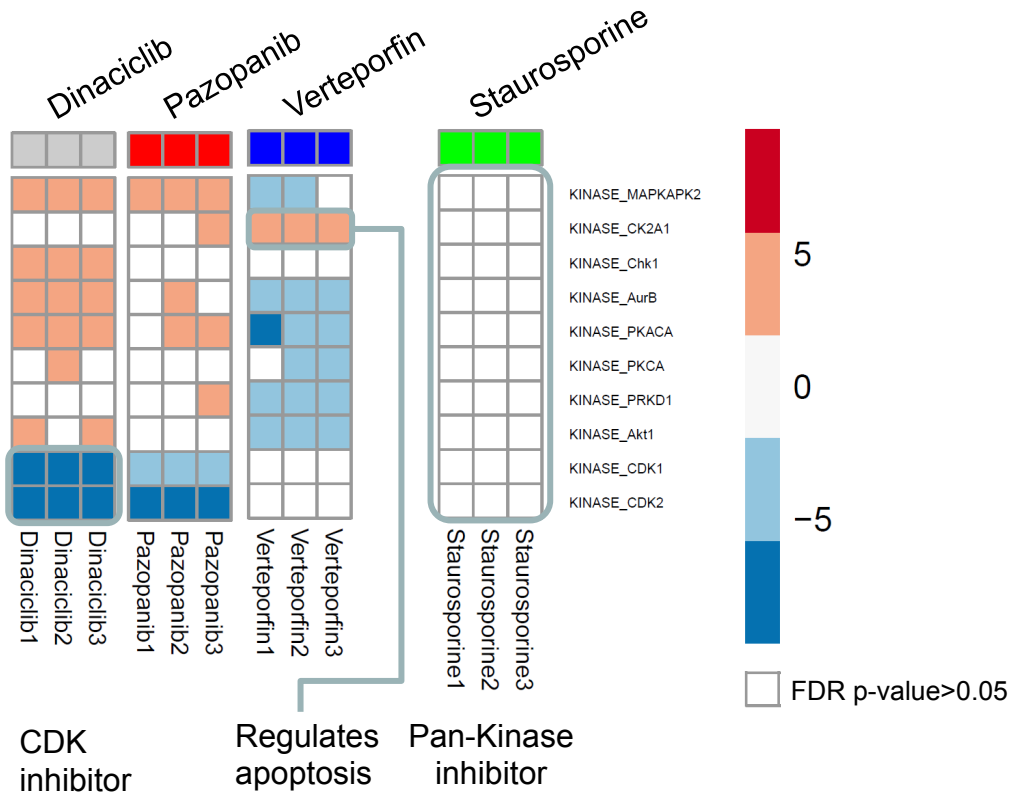
Additional targets from DIA analysis (Superlibrary)

See poster: Sebastian Vaca “Avant-garde” MP 357

Moving from gene-centric to phosphosite-centric analysis

- Gene Set Enrichment Analysis (GSEA) is a computational method for doing functional enrichment analysis at the ***gene*** level
  - Gene sets are groups of genes that share a common biological function, chromosomal location or regulation (MSigDB)
- GSEA can be adapted to consider data at the ***phosphosite*** level (ptmGSEA) and queried against phosphosite sets (PTMSigDb)

ptmGSEA vs. PTMSigDB



See poster: Karsten Krug “PTMsigDB” MP 696

Conclusions

- Porting PRM assays to Comprehensive MS assays (DIA, SWATH, etc.) is attractive
  - High value on samples
  - Re-usable, minable data
  - Still get most of the benefits of PRM on super high value analytes with internal standards
- Current instrumentation and methodologies are poised to potentiate this option
  - Duty cycle generally acceptable
  - Faster filling and/or more capacity would be better
- Transition optimization and high quality spectral library generation is really important for DIA workflows!
  - Check results vs. orthogonal method if possible!
- Real biological insights can be obtained through DIA phosphoproteomics workflows

Acknowledgments

Broad Institute – Jaffe Lab and Proteomics Platform

Sebastian Vaca
Karsten Krug
Kat DeRuff
Shawn Egri
Adam Officer
Malvina Papanastasiou
Jenn Abelin
DR Mani
Jinal Patel
Caitlin Feeney
Sue Abbatiello
Lindsay Pino
Ryan Peckner
Steve Carr

University of Washington – MacCoss Lab

Jarrett Egertson
Brendan MacLean
Brian Searle
Vagisha Sharma
Sonia Ting
Mike MacCoss

Library of Integrated Network-based Cellular Signatures (LINCS)
U01 CA164186
U54 HG008097