Supplemental Materials

I. Meeting Agenda and list of questions

II. Speakers slide presentations

III. Discussion group leaders and participants

Targeted Peptide Measurements in Biology and Medicine:
Guidance for Assay Development Using a "Fit-for-Purpose" Approach

June 18 - 19, 2013
NIH Campus
Bldg. 60
Bethesda, MD 20892

TUESDAY: JUNE 18, 2013

08:00 am – 09:15 am	**Introduction**
08:00 am – 08:30 am	**Registration**
08:30 am – 08:45 am	**Welcome, Overview and Meeting Charge** *Steven A. Carr, Ph.D., The Broad Institute of MIT & Harvard, Boston, MA*
08:45 am – 09:15 am	**Overview of Fit-for-Purpose Assay Development** *Andrew Hoofnagle, M.D., Ph.D., University of Washington*
09:15 am – 09:45 am	**TIER 1: Clinical Lab/FDA**
09:15 am – 09:45 am	**How Good is Good Enough?** *Russell P. Grant, Ph. D., Laboratory Corporation of America*
09:45 am – 12:45 pm	**TIER 2: Biomarker Verification**
09:45 am – 10:15 am	**Multiplexed Biomarker Verification in Plasma** *Susan Abbatiello, Ph.D., Broad Institute of MIT and Harvard*
10:15 am – 10:45 am	**Break**
10:45 am – 11:15 am	**Targeted MRM/MS Experiments** *Christoph Borchers, Ph.D., University of Victoria*
11:15 am – 11:45 am	**LC-MRM MS in Cancer Biology and Translational Research** *John M. Koomen, Ph.D., Moffitt Cancer Center*
11:45 am – 12:15 pm	**Best Practices for PRISM-SRM Assay Development** *Tao Liu, Ph.D., Pacific Northwest National Laboratory*
12:15 pm – 12:45 pm	**Targeted Peptide Measurement in Biology and Medicine** *Brad Ackermann, Ph.D., Laboratory for Experimental Medicine, Eli Lilly Company*
12:45 pm – 01:30 pm	**Lunch** (on your own)
01:30 pm – 02:00 pm	**Protein Target and Biomarker Quantitation in Translational Research of Biologics** *Hendrik Neubert, Ph.D. Pfizer Inc.*

01:30 pm – 03:30 pm	**TIER 3: Biology Focus**
02:00 pm – 02:30 pm	**Multiplexed Quantitative Analyses conducted in PRM Mode** *Bruno Domon, Ph.D., Luxembourg Clinical Proteomics Center*
02:30 pm – 03:00 pm	**Applying SRM for the quantification of proteins in basic biological/clinical research** *Ruth Huttenhain, Ph.D., UCSF*
03:00 pm – 03:30 pm	**Labeled Reference Peptide Method** *Daniel C. Liebler, Ph.D., Vanderbilt University*
03:30 pm – 05:00 pm	**Data Analysis**
03:30 pm – 04:00 pm	**Statistical Design and Analysis of Targeted Quantitative Proteomic Experiments** *Olga Vitek, Ph.D., Purdue University*
04:00 pm – 04:30 pm	**Automated & Reproducible Data Analysis Tools for Targeted Proteomics** *DR Mani, Ph.D., Broad Institute of MIT and Harvard*
04:30 pm – 05:00 pm	**Data Analysis** *Lukas Reiter, Ph.D., BiognoSYS AG*
05:00 pm – 06:00 pm	**Software for Data Acquisition and Databasing**
05:00 pm – 05:30 pm	**Software for Data Acquisition and Databasing** *Eric Deutsch, Ph.D., Institute for Systems Biology*
05:30 pm – 06:00 pm	**Targeted Proteomics Environment: Tools for targeted assay development and data analysis** *Brendan MacLean, University of Washington*
06:00 pm	Adjourn

WEDNESDAY, JUNE 19, 2013

08:30 am - 12:30 pm	Program Details for Day 2 Session
08:30 am – 10:30 am	Group leaders will convene their break-out groups to review topic areas and begin to formulate best practices and guidance for each defined Tier of targeted measurement/assay development. Groups will also address what information authors must provide in manuscripts.
10:30 am – 11:00 am	Break
11:00 am – 12:30 pm	Reconvene for read-back/presentation/discussion of each group's proposals and recommendations
12:30 pm	Adjourn

Questions Speakers are to Address

- What are the goals of your targeted MS experiments or software? Who are the "customers" or likely users of the methods and results?
    - e.g., precisely quantify the concentration(s) of peptide(s) in a tryptic digest of a complex protein matrix (e.g. human serum, human tissue, cell culture lysates), but not the accurate amount of the protein in the original sample. Use in candidate biomarker Verification or to study response to perturbation in a biological system.
    - e.g., develop software to facilitate and automate collection of targeted MS experimental data by research community
- What range of analyte plex-level per-injection do you typically use in your experiments? What is the impact of the plex-level used on the robustness and figures of merit (CV, LOD, LOQ) of the developed MRM/SRM assays?
- Explain how you establish confidence that what is being measured is the analyte of interest (e.g., match to spectra of an internal standards, match to reference spectra from discovery experiments, RT, etc.). How do these methods differ from "Discovery Proteomics" using data-dependent or data-independent experiments?
- Explain your method of quantification, how many transitions you monitor and which ones are chosen to quantify. If you are using internal standards describe in detail how they are used. If you are not using internal standards, explain how you are quantifying. Discuss the capabilities and limitations of your approach.
- If you generate standard curves (calibration or response curve), explain how you use them to assess the quantitative accuracy of the assay (e.g., are the slope and y-intercept from the curve regression used in calculating the analyte concentration in a sample? Is an external calibration curve used?).
- Can you provide a useful estimate or accurately determine the amount of protein in the matrix based on the measured levels of peptides? Explain how/why. Indicate experimental parameters such as number of peptides per protein and the criteria/computational tools applied. If you have multiple peptides from the same protein and each gives a different answer for the extrapolated protein level, how do you deal with this?
- Describe methods you use to establish presence of interferences and how you deal with them if detected
- How do you account for suppression of ionization in your quantification method?
- How do you "qualify" your measurements/assays, i.e., what criteria do you use or think appropriate to say that your measurements/assays has been successfully developed?
- What software and analytical tools do you use in your studies and why?
- What information do authors need to provide in their manuscripts/supplement to enable reviewers and readers to understand what was done and to be able to judge the confidence of the measurements made?

Targeted Peptide Measurements in Biology and Medicine: Best Practices for Assay Development Using a "Fit-for-Purpose" Approach

Steven A. Carr, Broad Institute of MIT and Harvard

June 18 and 19, 2013 NIH, Bethesda

Why have this meeting? Why now?

- In the field of Proteomics, umbrella terms like ”MRM/SRM” and “targeted MS” can convey the erroneous message that the results are unquestionably correct with respect to what is being detected and how much is present
- New methods (e.g., HR-PRM) are blurring lines between targeted quantification and discovery proteomics: what do we need to watch out for?
- Many targeted-MS papers are being published without documenting what was done and what results were obtained that justify the claims made
  - difficult for reviewers and readers to assess quality and reliability

Why have this meeting? Why now?

Parallel to situation in discovery proteomics

- MCP developed and published the first set of guidelines for publication of peptide and protein identification data: *Mol Cell Proteomics* (2004) 3: 531
  - try to insure that high quality, significant data are entering the proteomics literature
  - guidelines should not be burdensome nor should they dictate what tools to use: avoid stifling innovation
  - Initiate process requiring submission of data as a condition for acceptance of manuscript
- 2009/2010: Guidelines revised and updated
  - Statistical methods of data analysis
  - quantitative methods used in discovery proteomics

Top Level Meeting Goals

I. Describe what are people doing and why ("fit for purpose")

Strawman: "Tier" approach to measurements

II. Define what information (e.g., experimental methods and results) is needed for reviewers and readers to understand and assess the capabilities and limitations of the study

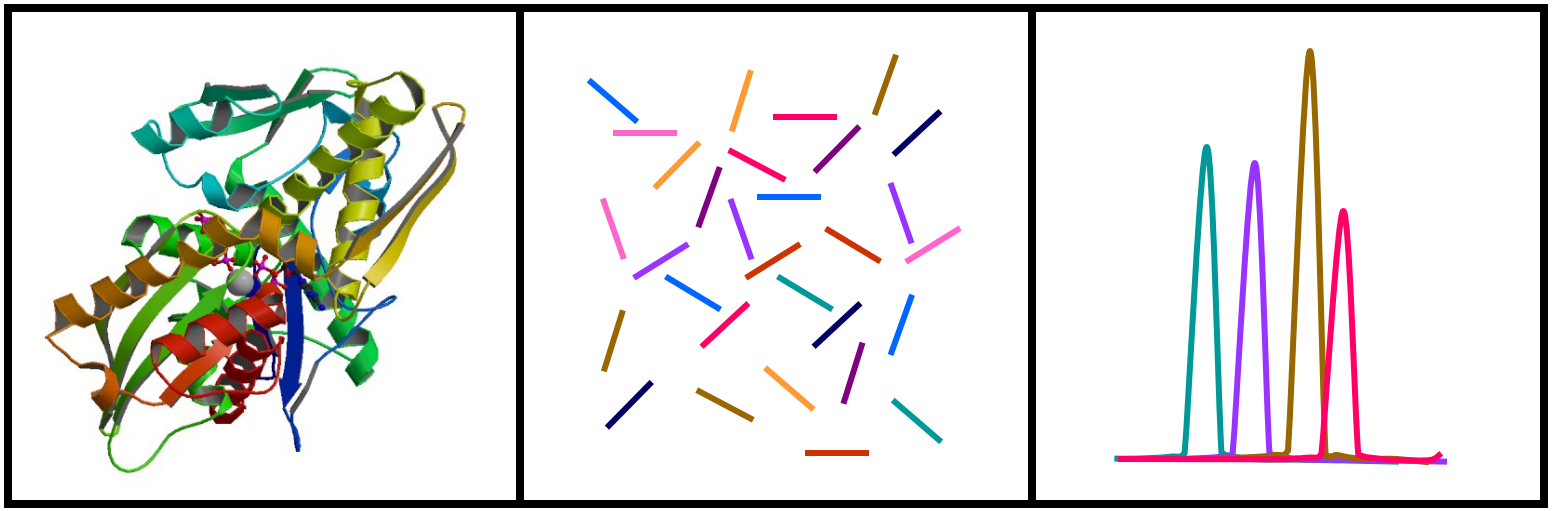
Fitness-for-purpose Concept

- ". . .the property of data produced by a measurement process that enables the user of the data to make technically correct decisions for a stated purpose."
  - Thompson and Ramsey (1995) Analyst; Bethem and Boyd (1998) J Amer Soc Mass Spectrom; Bethem et al.(2003) JASMS
- ". . .the magnitude of the uncertainty associated with a measurement in relation to the needs of the application area."
  - Kaiser in Baldwin et al. (1997) JASMS
- "…assay validation should be tailored to meet the intended purpose of the biomarker study, with a level of rigor commensurate with the intended use of the data."
  - Lee et al. (2006) Pharm. Research
- To establish fitness for purpose, ask "Is the uncertainty of the method within the data recipient's tolerance of uncertainty?"

Slide adapted with permission from Bob Bethem

Targeted Peptide Measurements in Biology and Medicine

What Are the Options



Andy Hoofnagle, MD PhD
Department of Laboratory Medicine
University of Washington

Practical Definitions

Validation:

A sincere attempt to demonstrate the robustness of a system

Quality control:

Ensuring that a validated system is working as it should

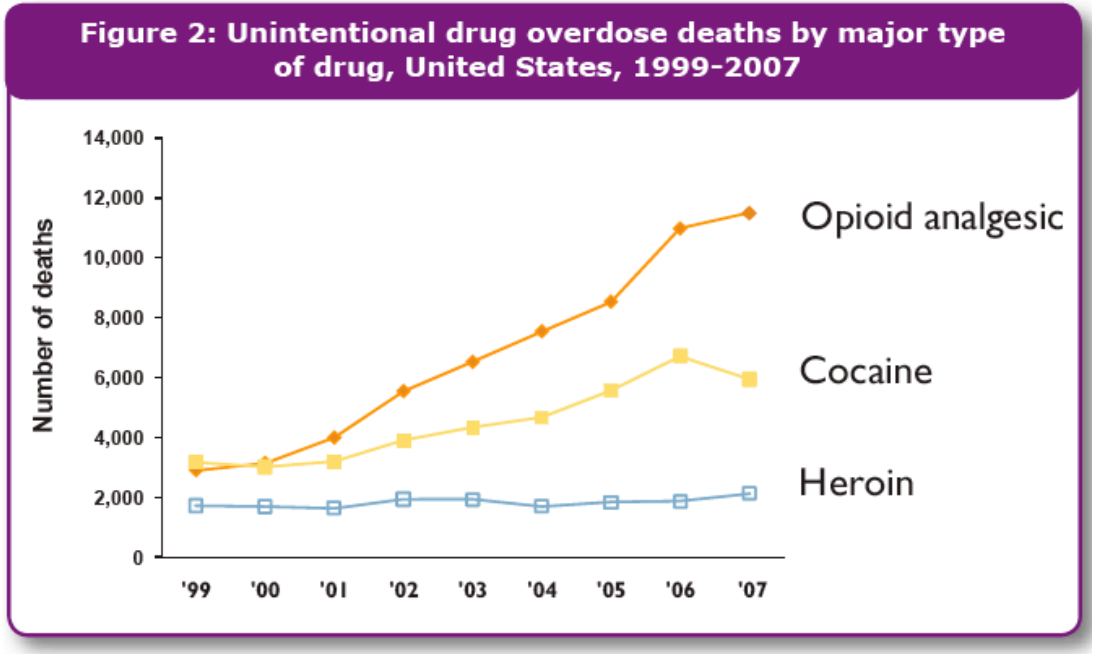
(fit-for-purpose)

Opioid Overdose is a Huge Problem

#1 killer in 2011

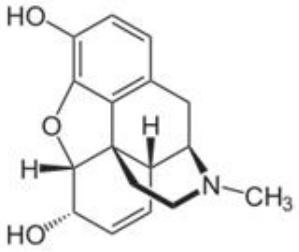
Surpassed car accidents

5x deaths *vs.* 1990s

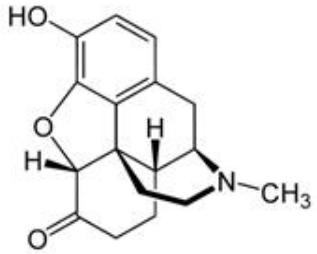


http://www.cdc.gov/HomeandRecreationalSafety/pdf/poison-issue-brief.pdf

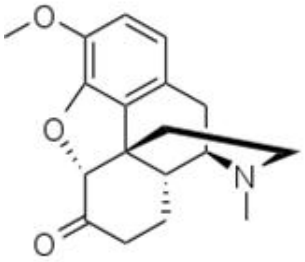
Molecular Basis of Cross-Reactivity in Immunoassays



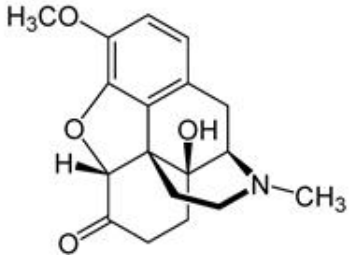
Morphine



Hydromorphone

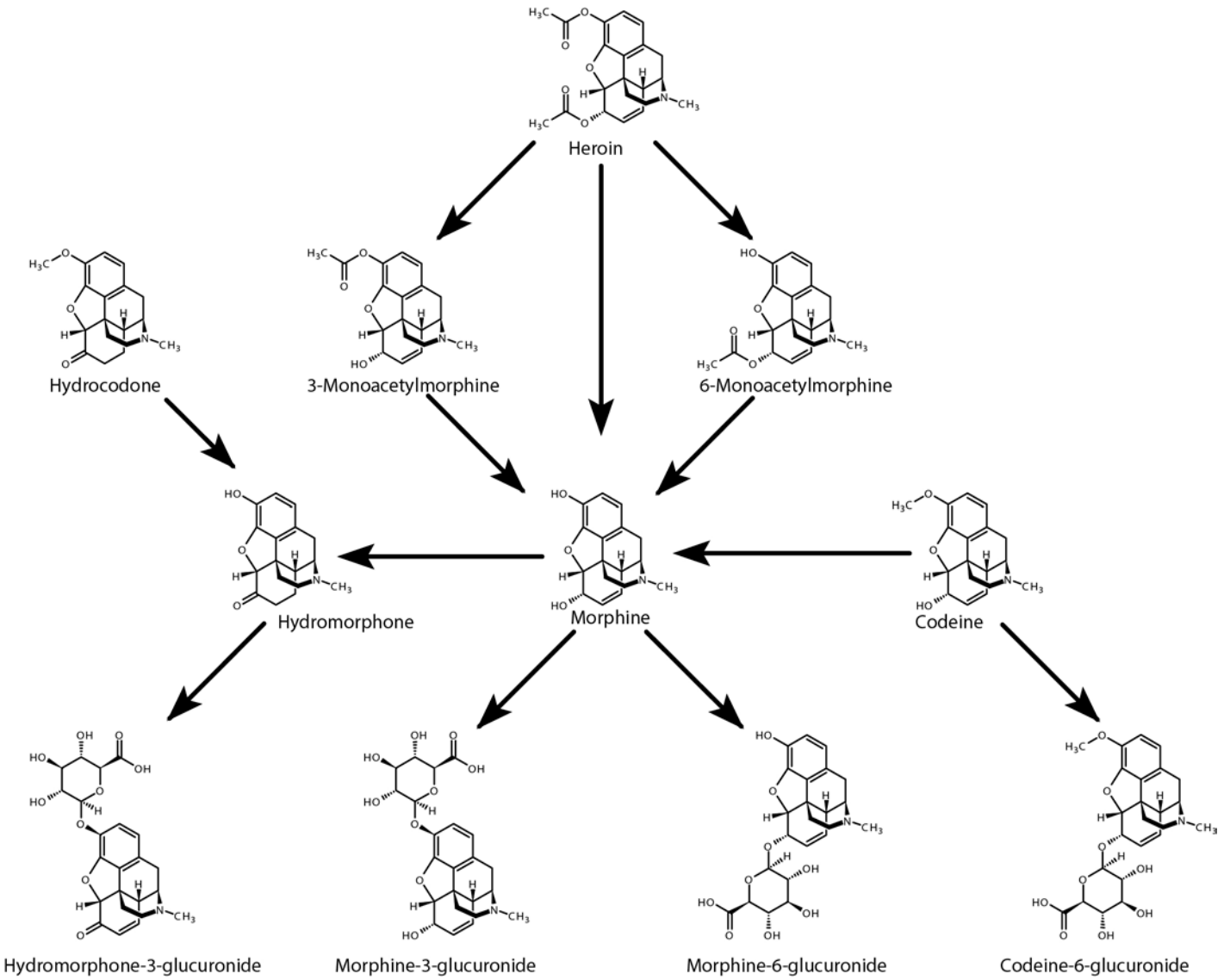


Hydrocodone

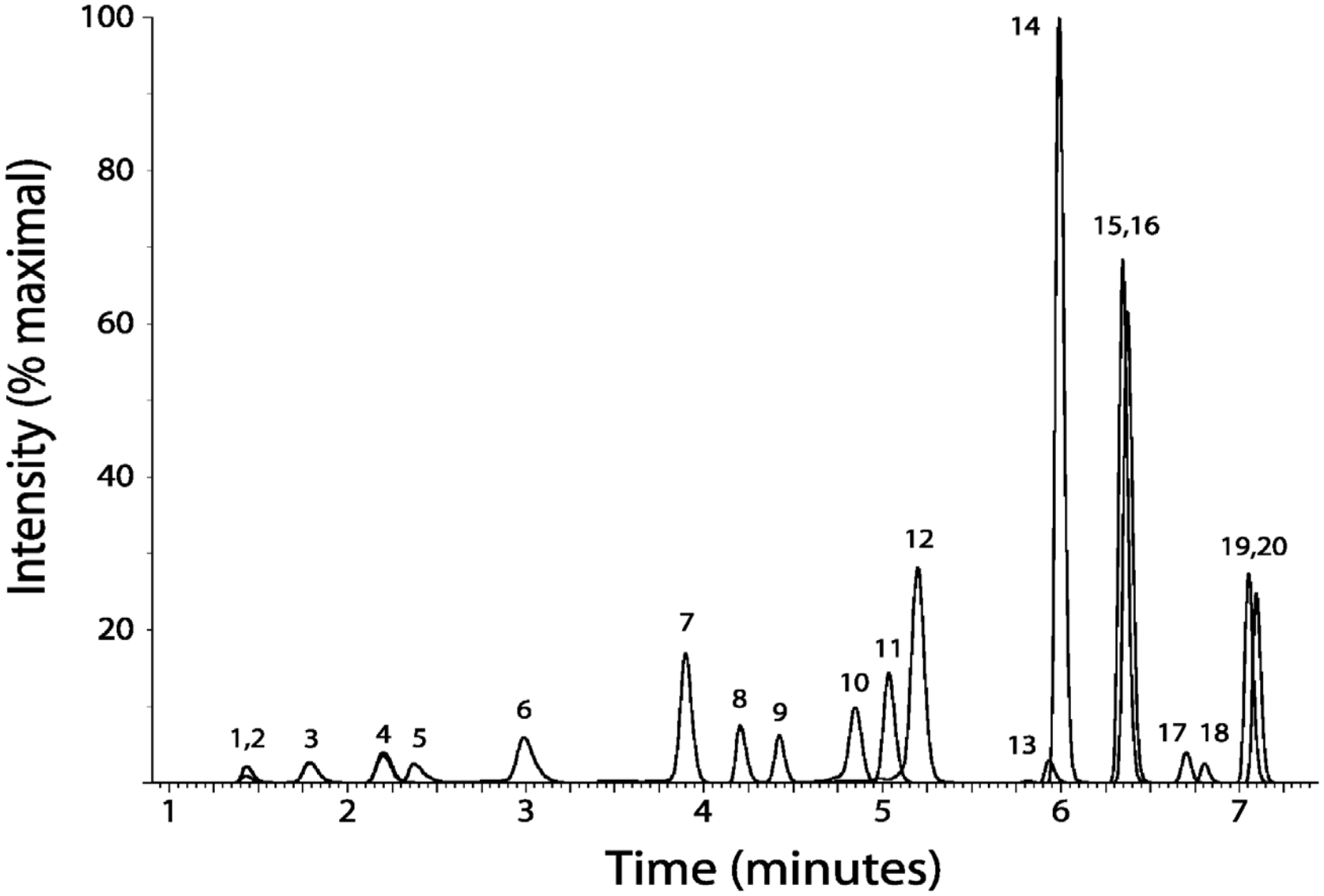


Oxycodone

Opioid Metabolism



Chromatogram of 20 Opioids and Metabolites



Quality Control

Specificity and Sensitivity

Batch specific

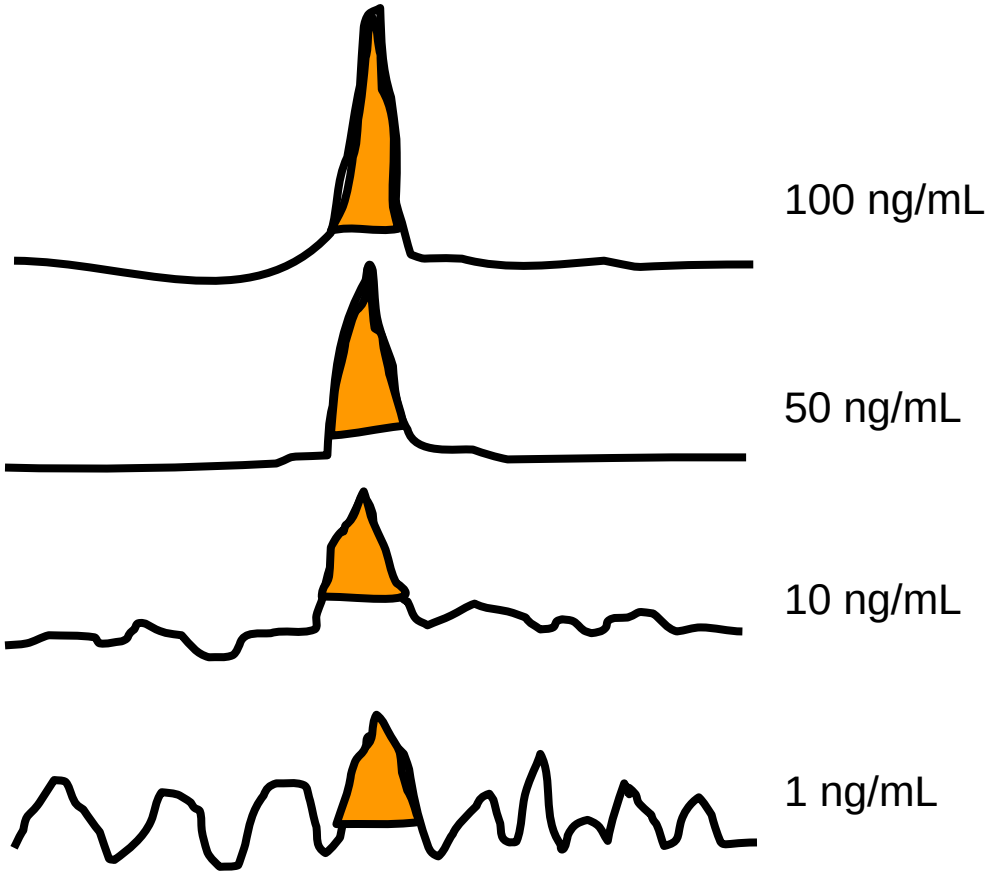
- Lowest calibrator (peak area or signal-to-noise)
- System suitability (signal intensity)
- Carryover
- Internal standard purity
- Calibration curve linearity
- Quality control materials

Specimen specific

- Relative retention time
- Confirmatory ion ratios
- Minimum internal standard (peak area or signal-to-noise)

Quality Control

Instrument Sensitivity



Integrating noise?

Solution:

Ensure that lowest standard has an acceptable peak area or signal-to-noise.

Quality Control

Instrument Sensitivity

Date	Unextract A D2	Unextract A D3	Std A D2	IS A D2	Std A D3	IS A D3
1/2/2008			300	12495	435	15744
1/4/2008	2015	2467	258	12828	374	16164
1/6/2008			292	13820	424	17413
1/8/2008			305	12584	442	15856
1/9/2008			284	13391	412	16873
1/11/2008	2410	2943	307	12664	445	15957
1/13/2008			296	13804	430	17393
1/15/2008			289	12706	419	16010
1/16/2008			245	13454	355	16952
1/18/2008	1913	2312	234	10090	339	12713
1/20/2008			198	9034	287	11383
1/22/2008	1208	1315	130	7023	189	8849
1/23/2008			98	5012	142	6315

Quality Control

Specificity

Sample	Analyte	Conc	RT	Analyte Peak Area	IS	Response	Calc conc	Ion Ratio	Ratio Flag?
Std D	D3	200	2.81	12870	12370	1.04	198.9	1.5	NO
Std C	D3	100	2.82	5990	12335	0.49	100.1	1.5	NO
Std B	D3	30	2.88	2288	13237	0.17	32.9	1.5	NO
Std A	D3	1	2.88	252	13704	0.02	0.9	1.4	NO
Ctrl HI	D3		2.90	780	13109	0.06	11.6	1.4	NO
Ctrl LO	D3		2.86	3501	13932	0.25	48.7	1.5	NO
W54634	D3		2.83	1650	13081	0.13	24.5	1.5	NO
W58132	D3		2.86	1249	12703	0.10	19.1	1.5	NO
W66023	D3		2.89	1641	13211	0.12	24.1	1.5	NO
W62743	D3		2.81	2423	12636	0.19	37.2	1.5	NO
W66817	D3		2.86	2208	13543	0.16	31.6	1.5	NO
H66438	D3		2.83	2359	12163	0.19	37.6	1.4	NO
H62633	D3		2.81	2057	13559	0.15	29.4	1.5	NO
H78948	D3		2.82	1593	12762	0.12	24.2	1.4	NO
H78388	D3		2.82	2836	12824	0.22	42.9	1.4	NO
H74054	D3		2.87	1257	12469	0.10	19.6	1.5	NO
H87727	D3		2.67	1362	12944	0.11	20.4	1.9	YES
H89927	D3		2.85	1851	12528	0.15	28.7	1.4	NO
H88721	D3		2.89	2731	13284	0.21	39.9	1.4	NO

Quality Control

Internal Standard

Sample	Analyte	Conc	RT	Analyte Peak Area	IS	Response	Calc conc	Ion Ratio	Ratio Flag?
Std D	D3	200	2.84	13902	12597	1.10	201.1	1.4	NO
Std C	D3	100	2.82	5210	10349	0.50	99.8	1.3	NO
Std B	D3	30	2.88	2128	11640	0.18	33.3	1.4	NO
Std A	D3	1	2.87	249	10274	0.02	1.1	1.4	NO
Ctrl HI	D3		2.88	808	13878	0.06	10.7	1.4	NO
Ctrl LO	D3		2.88	3190	12186	0.26	48.4	1.4	NO
M64598	D3		2.89	2809	8234	0.34	63.0	1.3	NO
M67899	D3		2.86	1408	13192	0.11	19.7	1.5	NO
M70133	D3		2.84	1504	13469	0.11	20.6	1.5	NO
M70134	D3		2.82	1443	10960	0.13	24.3	1.2	NO
M70135	D3		2.90	1703	12058	0.14	26.1	1.3	NO
M71222	D3		2.80	1774	13004	0.14	25.2	1.4	NO
M72090	D3		2.84	1932	11747	0.16	30.4	1.4	NO
T80031	D3		2.83	2034	2312	0.88	162.6	1.4	NO
T80100	D3		2.83	2178	13075	0.17	30.8	1.3	NO
T81070	D3		2.89	4508	12678	0.36	65.7	1.2	NO
T90909	D3		2.83	3488	11378	0.31	56.7	1.4	NO
T91002	D3		2.80	2912	12790	0.23	42.1	1.4	NO
T91114	D3		2.84	2600	12610	0.21	38.1	1.5	NO

Quality Control

One Step Further

Matrix effects (Ion suppression)

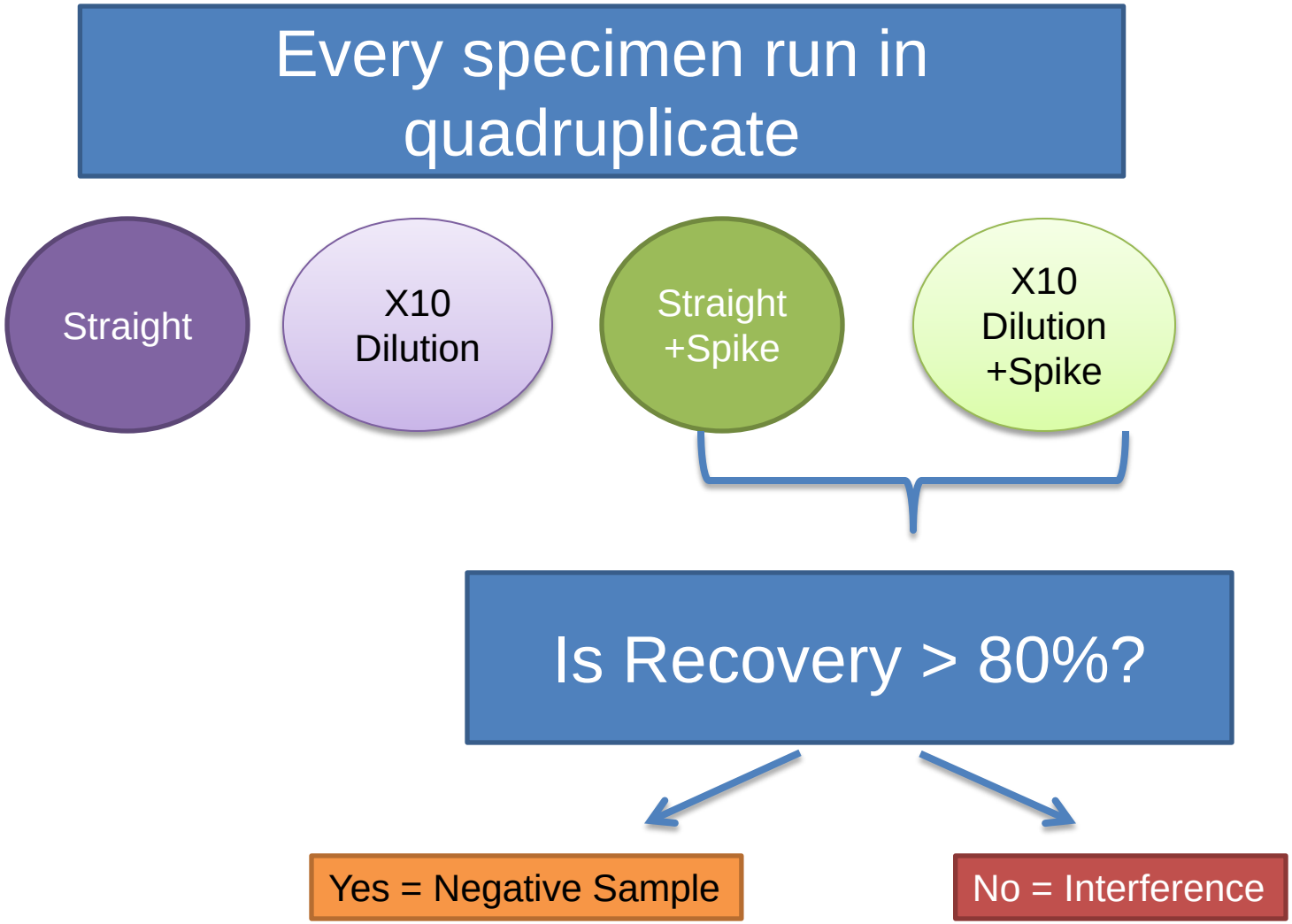
What do we do if useful internal standards are not available?

Standard addition experiments
Determine that negative results are truly negative

Dickerson, J Anal Toxicol, 2012
Dickerson, Clin Chim Acta, 2013

Quality Control

One Step Further



Quality Control

Specificity and Interference

In the absence of stable isotope labeled analog

Lack of interference

- Happens during validation
- Spike recovery or mixing studies in as many different biological replicates of matrix as possible/reasonable
- Constant transition ion ratios across the experiment
- Reliable recovery of analyte across the experiment (requires calibration materials)

Specificity

- Rely on (relative) retention time and transition ion ratios in each sample
- If there is a question, acquire MS/MS to compare with pure material

Different Uses of Targeted Assays

Distribution of reagents to clinical laboratories (think FDA)

Patients (laboratory developed tests, think CLIA)

Verification of biomarkers

Discovery of biomarkers

Mechanistic studies in basic science

Regulatory Statutes and Guidance Documents

Food and Drug Administration

Clinical Laboratory Improvement Amendments of 1988

Clinical Laboratory Standards Institute

International Organization for Standards (ISO)

New York State

Clinical society recommendations

Journal guidelines

Some Examples of Guidance Documents

FDA

Guidance for Industry, Bioanalytical Method Validation

2001

CLSI (C50)

Mass Spectrometry in the Clinical Laboratory

2007

ISO (17511)

Metrological traceability of values assigned to calibrators and control materials

2003

Examples of Targeted Methods

Target during analysis

Selected reaction monitoring (SRM, MRM)

Parallel reaction monitoring

Isotope labeled peptides (partial *vs.* complete coverage of targets)

Winged labeled peptides

Isotope labeled proteins

Purified *vs.* crude internal standard peptides

Target post hoc

Data-independent acquisition

Examples of Targeted Methods

Calibration

None

Internal standard

External calibration materials

Purified peptides

Recombinant proteins

Native human matrix

External calibration can improve between-day precision

Defining Fit-for-purpose

Tier 1

Clinical laboratory, FDA submission

Guidelines and regulation

Tier 2

e.g. Biomarker verification

Labeled internal standards for all analytes

Tier 3

e.g. Biology focus

Limited or no use of internal standards

What is the Measurand?

Peptide concentration

the concentration of a peptide in a tryptic digest of a complex protein matrix (e.g. human serum, human tissue, cell culture lysates), not the amount of a protein in the original sample

Protein concentration

the concentration of protein in a complex protein matrix

When Does It Become An Assay?

Definition of MRM transitions for a peptide

Detection of a peptide spiked into a matrix

Quantification of a peptide spiked into a matrix

Quantification of protein spiked into a matrix

Quantification of an endogenous protein

Tiers of Targeted Assays

Validation Experiments Performed

Things we might choose from:

Recovery, repeatability, reproducibility, linearity, interferences and matrix effects, stability, LLOQ, LOD, carryover

Examples:

(1) Bilirubin is spiked into a sample at multiple concentrations. Measured analyte concentration is compared with unspiked sample.

(2) Two concentrations of peptide are spiked into digested human plasma and injected in triplicate on each of 10 days.

Tiers of Targeted Assays
Validation Criteria Met

Examples of what we might consider acceptable:

(1) The amount of peptide detected before and after the addition of bilirubin to digests of five different human serum samples are within 15% (for each sample and bilirubin concentration tested)

(2) The average within-day variability for each spike is <10% and the between-day variability for each spike is <15%

Tier 1 Assay

The Highest Bar

Oversight:	Food & Drug Administration
Guidance documents:	CLSI (nothing specific for LC-MS yet)
Expected throughput:	10,000s
Number of labs deploying:	100s
Clinical validation required:	Yes
Assay needed:	Calibrated (measuring protein concentration), internal standards, batch quality controls

Tier 1 Assay

The Highest Bar

Simplified example of what might be acceptable to FDA:

Precision: <8% within-day variability, <12% between-day variability

Bias: <5% on each of five days

Calibration curve slope: <5% difference over five days

Interference and Matrix effects: Blank samples (with no spiked internal standard peptide) and double blanks (with no spiked peptide or spiked internal standard peptide) contribute less than 5% of LLOQ signal, recovery of analyte spiked into 60 samples is 85-115% for all samples, three transitions monitored and the two transition ratios are within 25% of mean for all 60 samples and are monitored for all samples in production as QC

Tier 1 Assay

The Highest Bar

Simplified example of what might be acceptable to FDA (cont'd):

LLOQ validation: A sample run consecutively for 25 days at a level 50% above the LLOQ has a precision <15%

Carryover: Blank samples run after a matrix-matched highest calibrator have less than 5% of the signal at the LLOQ for the endogenous peptide and internal standard channels

Stability and sample type: different collection and storage conditions are evaluated for the effect on the measurement of the endogenous anlayte concentration, no effect is >15%

Clinical validation: safe and effective (PMA, 100s-1000s of samples), equivalence (510k, 100s of samples)

Tier 1 Assay

For Clinical Use

Oversight:	CLIA
Guidance documents:	CLSI, accrediting organizations
Expected throughput:	1,000s-10,000s
Number of labs deploying:	one (or more if published)
Clinical validation required:	Yes
Assay needed:	Calibrated (measuring protein concentration), internal standards, batch quality controls

Tier 1 Assay

For Clinical Use

Consider two different categories:

Category A:

The same acceptance criteria as for a Level 1 Assay

Category B:

Precision: <15% within-day variability (N=20), <20% between-day variability (N=25) at a concentration 50% above the lower end of the analytical measurement range

Interference and Matrix effects: recovery of analyte spiked into 10 samples is 80-110% for all samples, three transitions are monitored and the two transition ratios are within 25% of mean for all 10 samples and are monitored for all samples in production as QC

Tier 1 Assay

For Clinical Use

Category B (cont'd):

Linearity: The relationship between expected and observed concentration is linear across the analytical measurement range (for each point, observed concentration is within 10% of expected)

LLOQ: Linear dilution of samples with endogenous analyte are each run 5 times and the concentration at which the CV rises to 20% is identified by interpolation

Carryover: A blank run after the highest calibrator contributes to less than 5% of the LLOQ

Stability and sample type: Different collection and storage conditions are evaluated for the effect on the measurement of the endogenous analyte concentration, no effect is >15%

Tier 2 Assay

Biomarker Verification

Oversight:	Journals, investors
Expected throughput:	100s
Number of labs deploying:	2-4
Clinical validation required:	No
Assay needed:	Internal standards

Different Tier 2 Assays

Consider the following experiment:

Seven logarithmic concentrations of peptide are spiked into a matrix (e.g. digested ovarian stromal tissue) and run in duplicate. Constant amount of internal standard peptide is added to each of the samples.

Category A: >95% *pure* internal standards, between-day imprecision <20% *(N=10)*, two transition ratios are monitored and should be within 25% of the mean over time

Category B: same as category A, *unpurified* internal standards

Category C: purified or unpurified internal standard peptides, between-day imprecision <20% *(N=5)*

Tier 3 Assay

Biology Focus

Oversight:	Journals, grant reviewers
Expected throughput:	10s
Number of labs deploying:	1
Clinical validation required:	No
Assay needed:	To be discussed

Tier 2: Multiplexed Biomarker Verification in Plasma

Susan Abbatiello, Ph.D.
The Broad Institute of MIT and Harvard
June 18, 2013

BROAD INSTITUTE

CPTAC VWG Study 9 – Targeting 34 Proteins in Depleted Plasma, 125 Peptide Targets

unlabeled protein

^{15}N labeled protein

34 proteins

8-point Curve low amol – 100 fmol/ug plasma

27 proteins

Fixed Spike Level

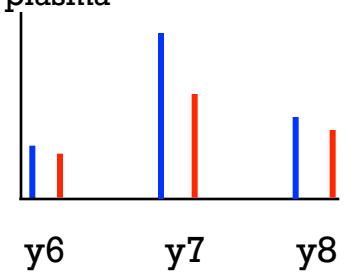
depleted plasma

Trypsin

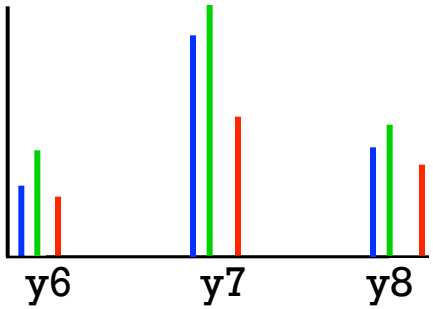
^{13}C/^{15}N labeled peptides

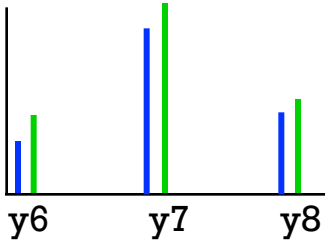
115 peptides

Fixed Spike Level



Figures of Merit (LOD, LOQ)

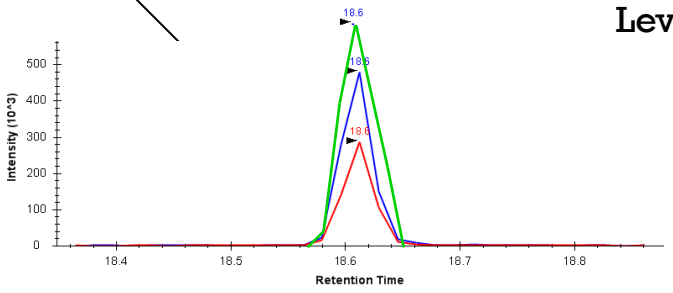




Recovery from Assay

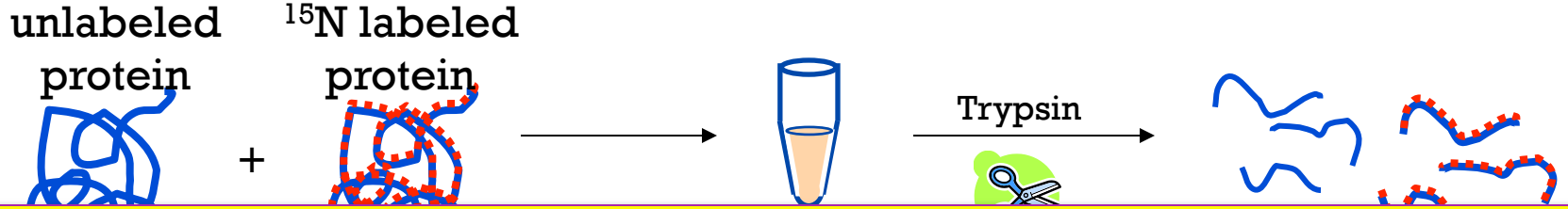
9.1 = 750 transitions

9.2 = 1095 transitions



LC-MRM-MS

CPTAC VWG Study 9 – Targeting 34 Proteins in Depleted Plasma, 125 Peptide Targets



Goals:

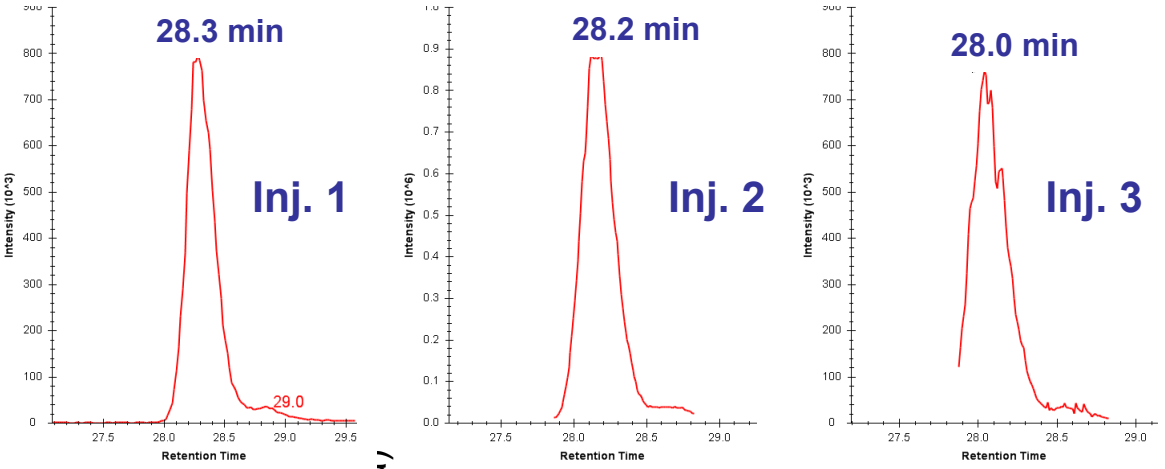
- **Design an assay with low technical variation that can be used to assess biological variation in a large number of samples by precise, relative quantitation**
- Prove feasibility of > 100-plex (34 proteins) assays in plasma
- Improve LOD and LOQ by depleting abundant proteins
- Evaluate true quantitative accuracy and digestion recovery using heavy labeled proteins
- Conduct blinded verification study to assess accuracy, precision and reproducibility across multiple sites and instrument platforms
- Evaluate system suitability test in context of this large-scale inter-lab study

34 proteins, 10 participating sites, 15 instruments, 4 Vendors

Multiplexed MRM Assays Require Good Quality Data

Chromatography

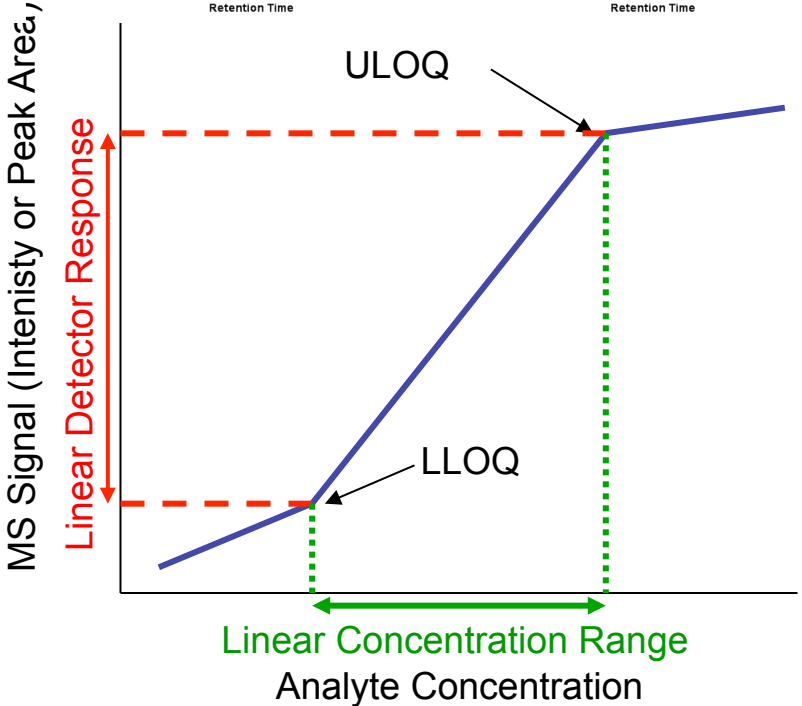
- Scheduling puts rigorous demands on RT reproducibility
- Peak width and RT drift are often limiting factors
- Environmental Factors and LC plumbing play key roles



MS Signal

- What is the minimum MS signal required to satisfy LLOQ criteria?
- How does peak shape influence observed MS signal?
- How important is ESI stability?

Prior to analyzing complex samples, are LC-MRM-MS systems running in optimal condition?



Analytes

- Typical multiplex levels range from 10s to 160 peptide targets
- Study 9 pushed the upper end of analyte multiplexing
  - 125 peptide targets
  - 1-2 types of isotopically labeled internal standards
  - 750 or 1095 transitions monitored in a single injection (1 hour gradient, 80 min LC cycle time)

Points to Consider:

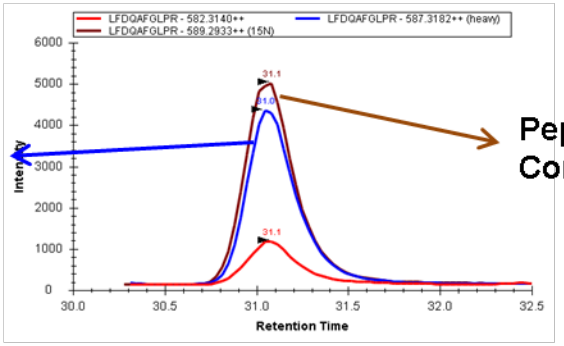
- Minimum dwell time
- Minimum cycle time
- Retention time scheduling
- Number of coeluting transitions

- 10 msec (*13-15 msec total)
- 1-2 sec (FWHM dependent)
- 1.5-2 min
- 60-120 transitions
  - ~7 - 20 ananlytes

^{15}N Labeled Proteins Improve Quantitative Accuracy*

^{13}C/^{15}N Peptide Internal Standards

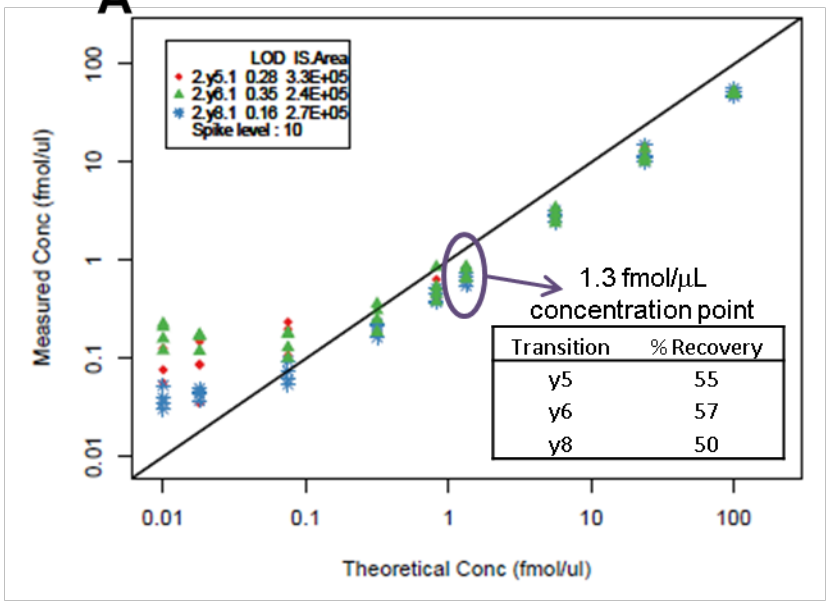
$$\text{Peptide Conc}_{(13C/15N)} = \frac{\text{Light Peak Area}}{^{13}\text{C}/^{15}\text{N Peak Area}} \times \frac{10\ \text{fmol}}{\mu\text{L}}$$



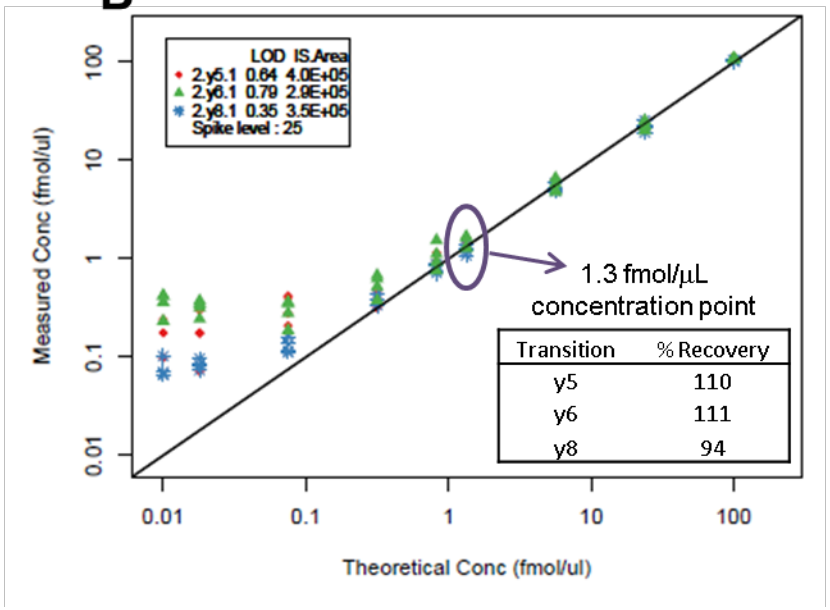
^{15}N Protein Internal Standards

$$\text{Peptide Conc}_{(15N)} = \frac{\text{Light Peak Area}}{^{15}\text{N Peak Area}} \times \frac{25\ \text{fmol}}{\mu\text{L}}$$

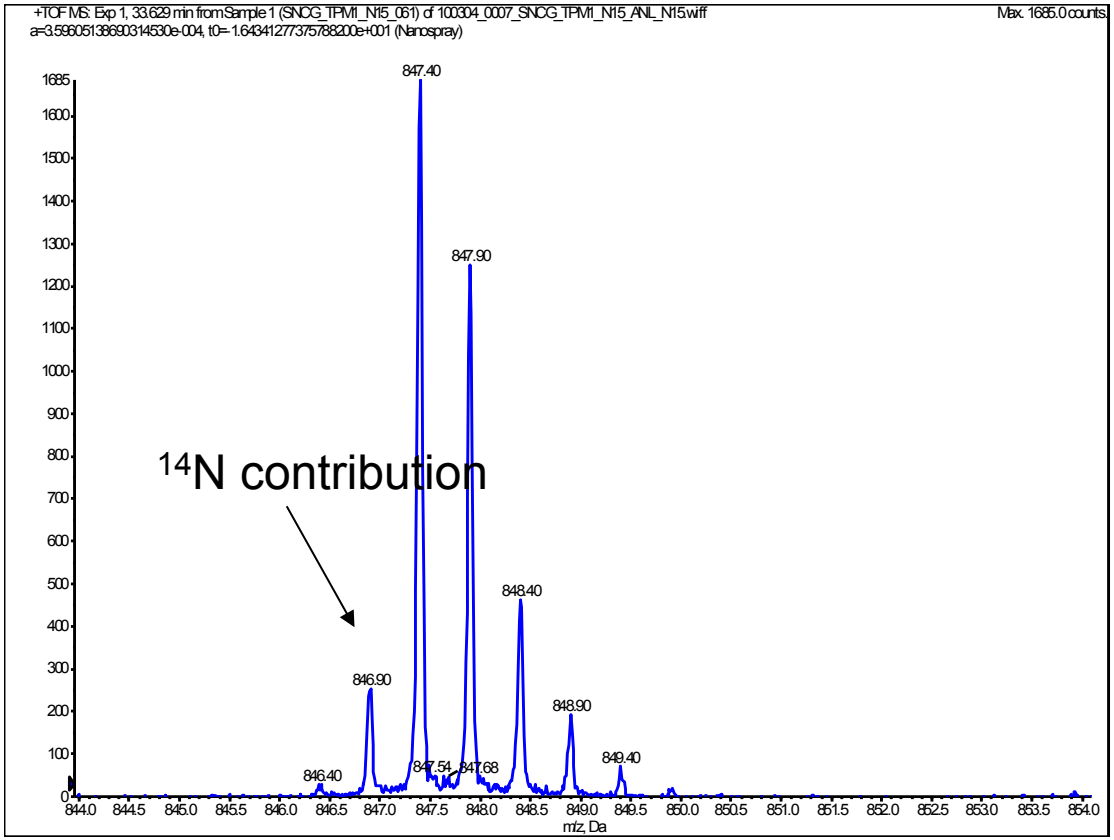
A



B



Limitations to ^{15}N Proteins and Isotopically Labeled Standards in General



- **Determined ^{15}N incorporation is 99.10 % for synuclein gamma peptide TVEEAENIAVTSGVVR**
- **What is the peptide-level ^{15}N incorporation?**
- **Do protein standards digest the same way as native proteins?**
- **How pure are the proteins?**

- How confidently do we quantify our standards?

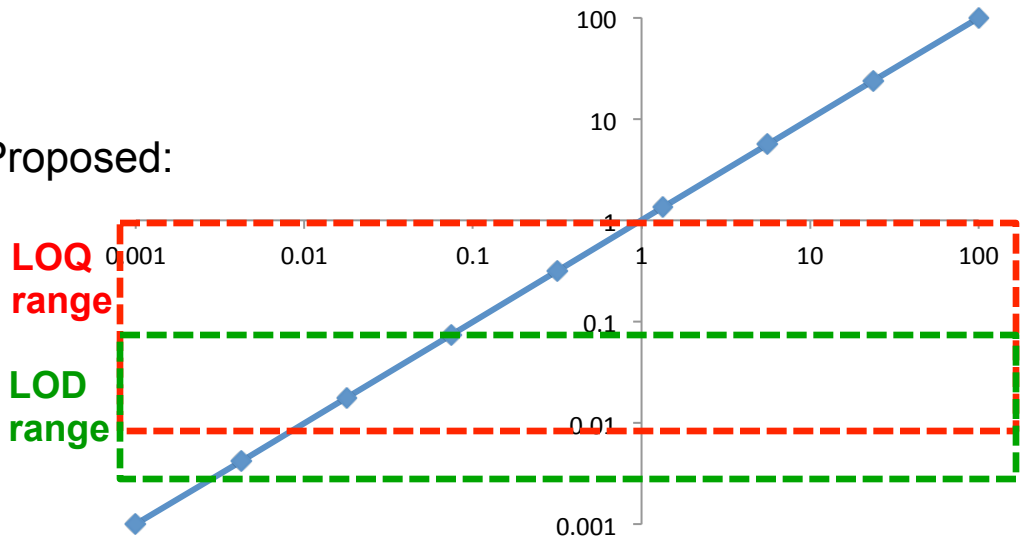
LOD/LOQ Calculations: How Many Points in the Curve are Needed?

What is the ideal concentration range?

$$LOD = \bar{s}_{blank} + t_{0.95} \times (s_{blank} + s_{low})/\sqrt{n}$$

(fmol/µL)
250
113
51
23
10
4.6
2.0
0.9
0.42
0.19
0.09
0.04
0.017
0.008
0.004
0.002

Proposed:



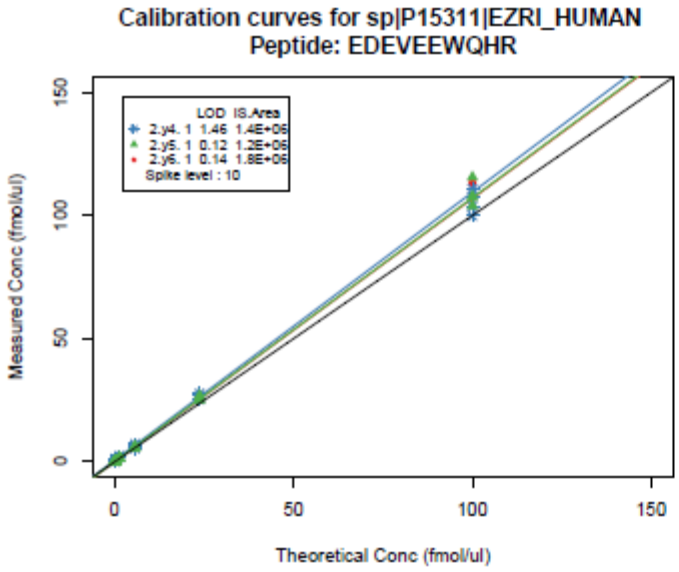
- Generate preliminary curves (16 pts)
- Pick a range and number of points to cover most peptides
- 4 process replicates

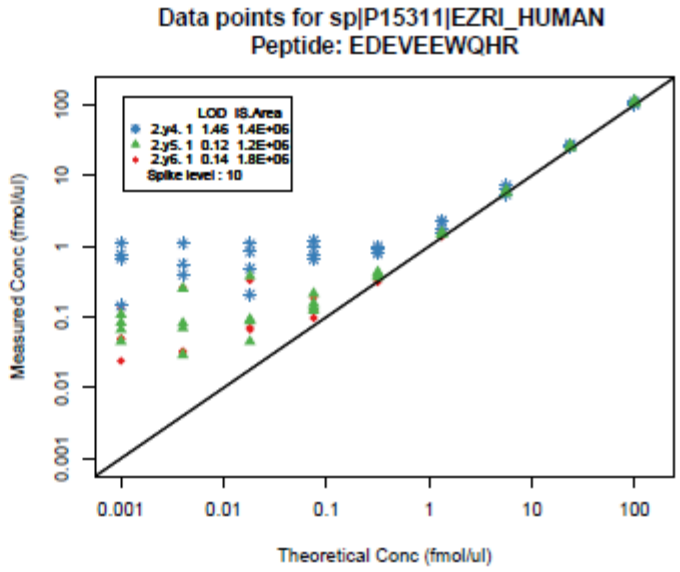
Linnet & Kondratovich, (2004) Clin Chem

Keshishian et al, (2009) MCP

Use of Response Curves

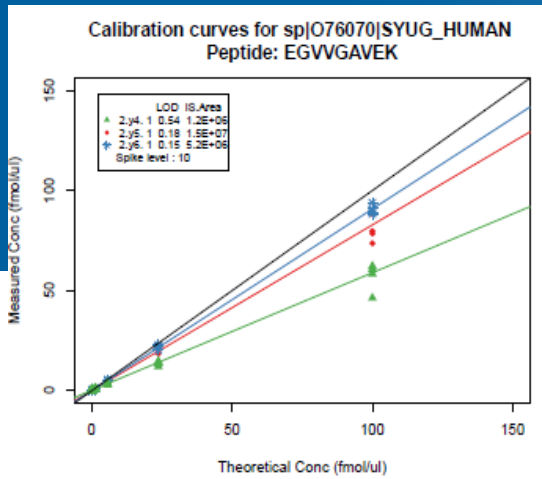
- Check for linearity
- Check for interferences
- Slope = 1; y-intercept = 0 for linear plots
- We do not currently use slope to correct quantitation values
- If peptide standards are used, assumptions are made about protein digestion efficiency

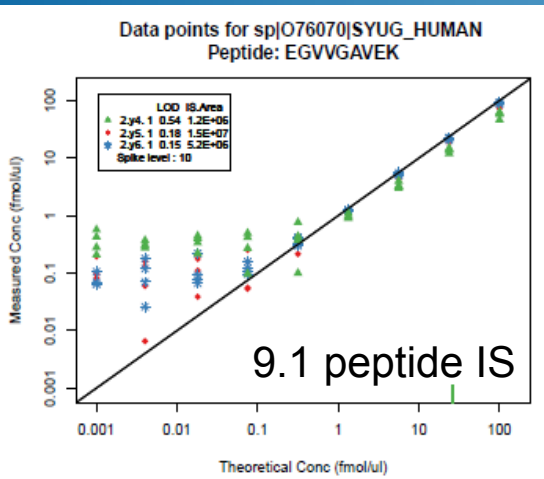


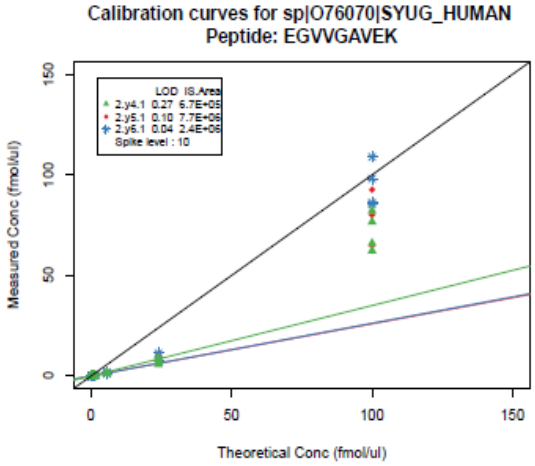


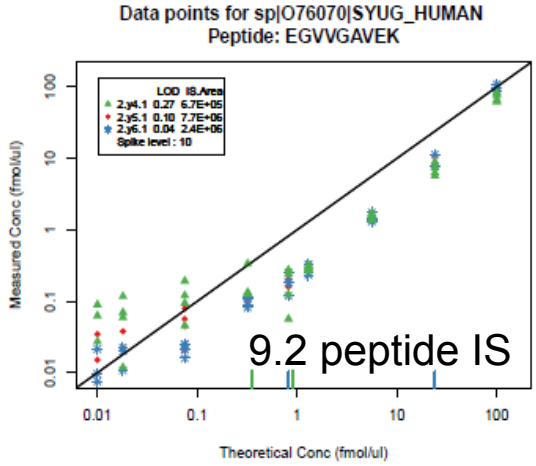
Quantitation at the Transition Level

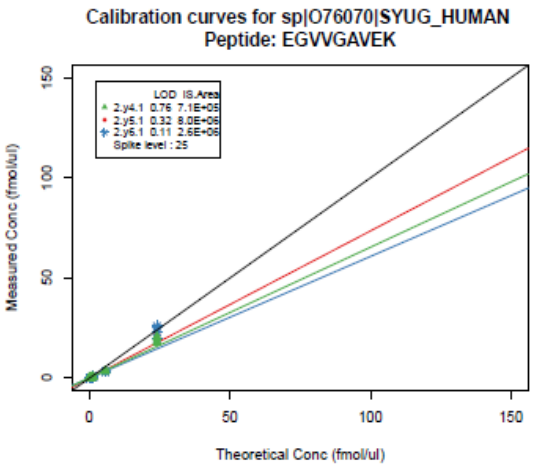
- Single “best” transition used
  - Stable
  - Intense
  - Interference-free
- Remaining transitions used for ID and interference check
- Is using only 1 transition acceptable or should the sum of transitions be used?

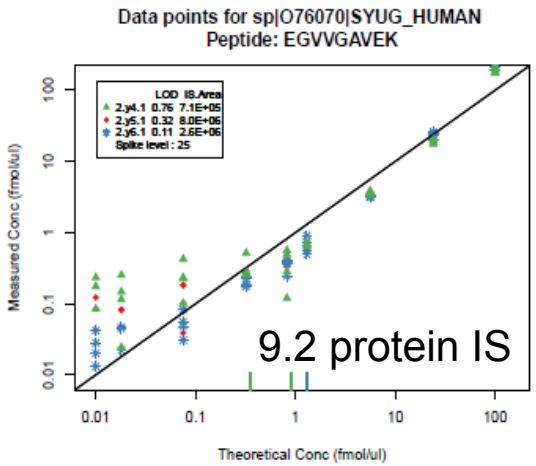








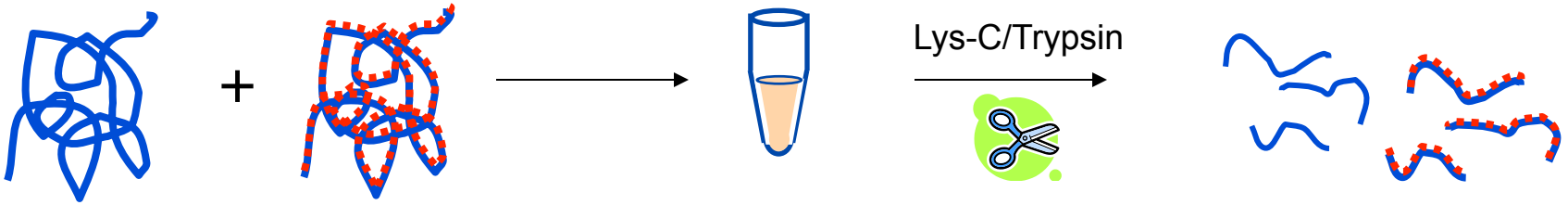




Qualifying Assay Criteria

- LC-MRM-MS passes system suitability specifications
- Internal standards meet minimum signal threshold
- Peptides co-elute with internal standards
- Analytes and internal standards are interference-free
- Instrument performance and quantitative response is similar in inter-lab studies
- Digestion controls must be consistent in all samples

Software



Skyline

DDA LC-MS/MS, Database search

Skyline

Spectral Library

Selection of 123 target peptides

Skyline

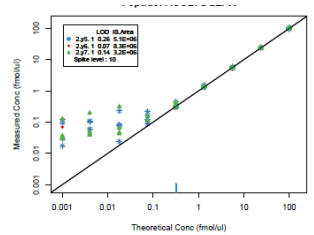
Top 5 Product Ions

MRM-MS Data Acquisition

Skyline

- Selection of best 3 ions
- CE Calculation

QuaSAR



- LODs
- CVs
- Curves
- Interferences

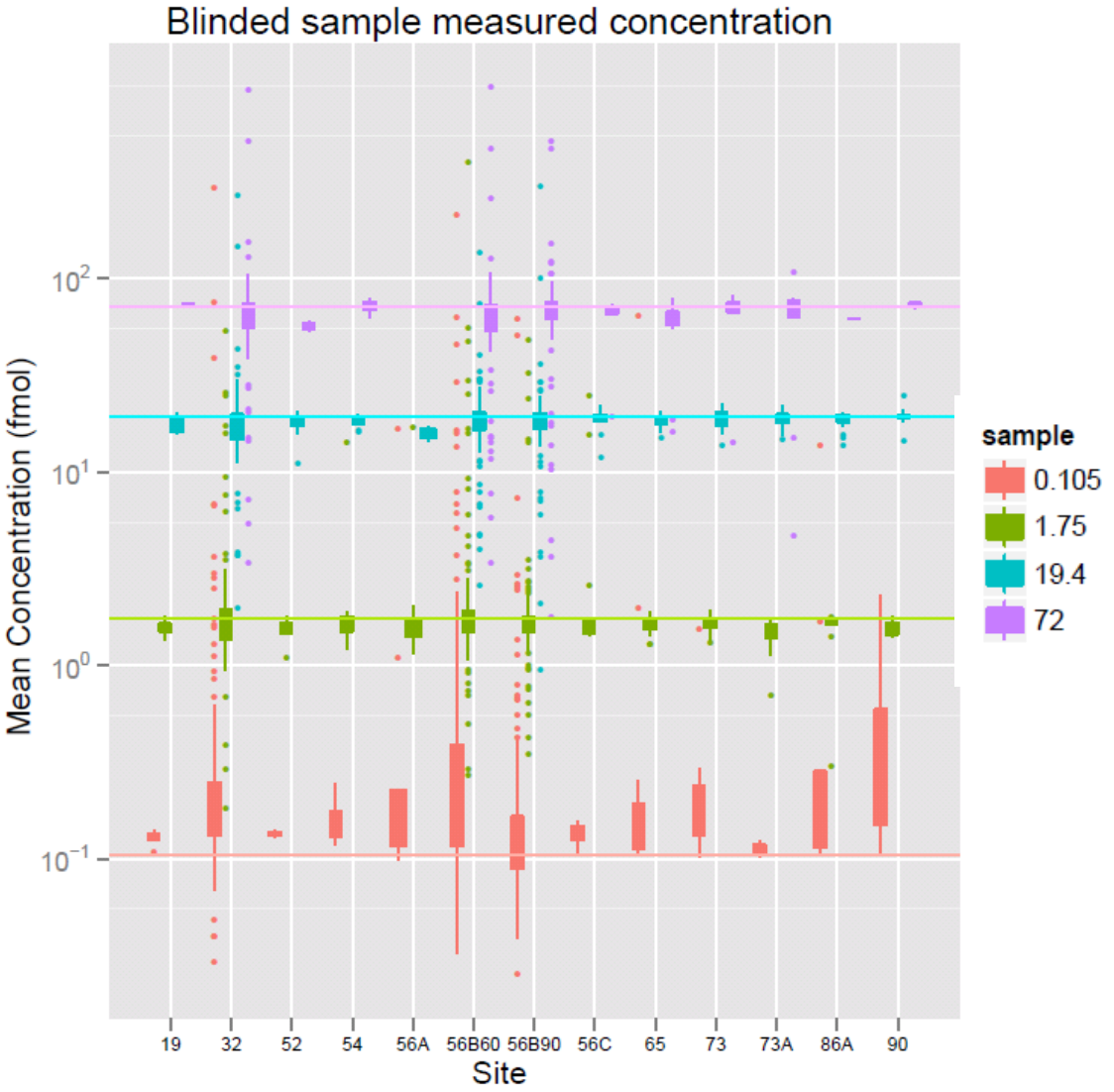
Advantages and Disadvantages of This Approach

- Peptide ID is straight forward
- Interferences can be objectively determined
- Quantitation is based on a single transition
- Protein standards bring us closer to “accuracy”
- Multiplexing allows us to target many analytes in 1 injection
- External calibration is not used

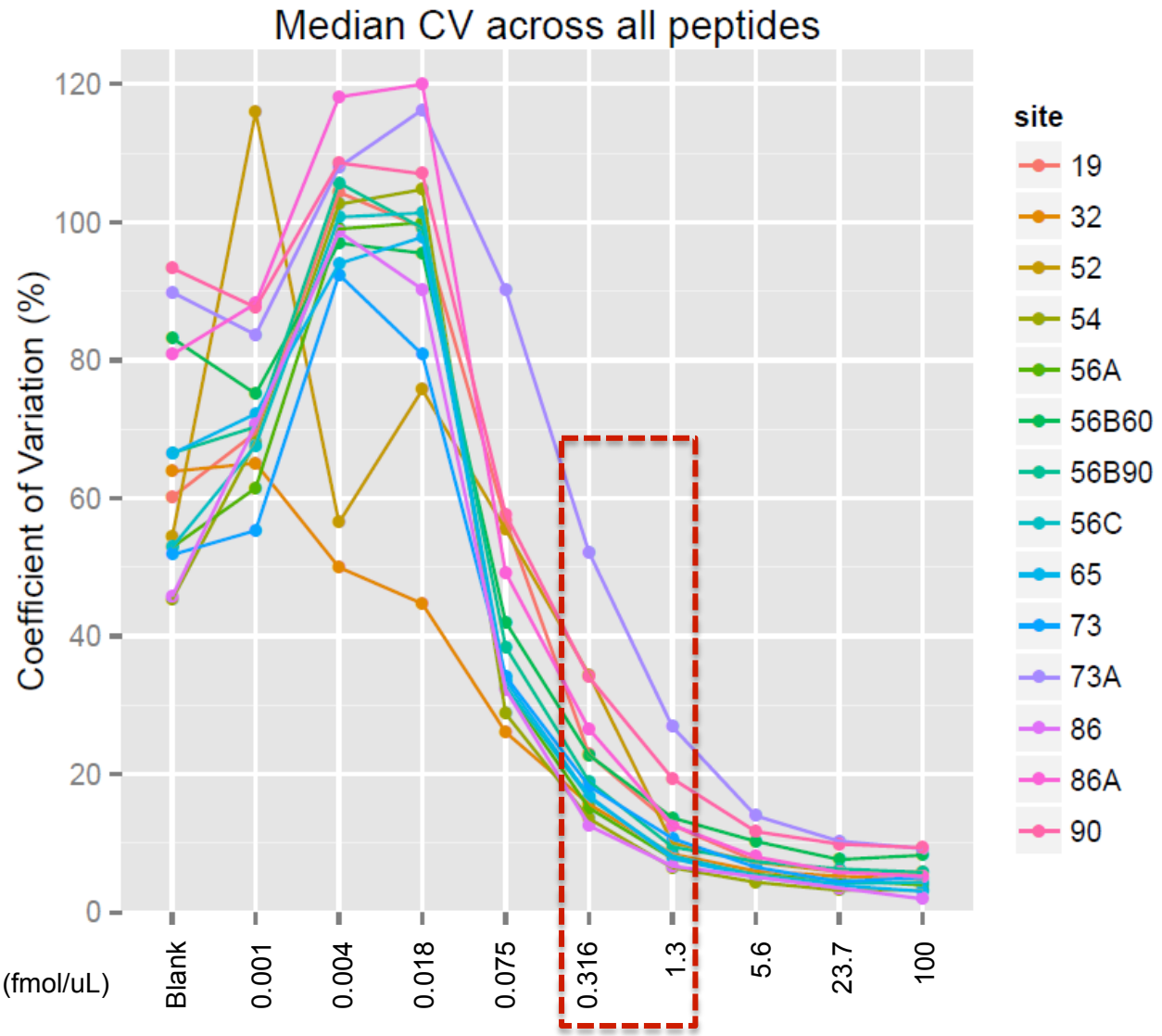
- Cost of internal standards
- Accurately quantifying internal standards
- Quantitation is based on a single transition
- Protein standards are hard to come by
- Data quality must be high to quantify all analytes in 1 injection
- Assumptions regarding digestions are made
- Curves can take 3+ days to complete

Study 9.1 – Blinded Sample Results

- Results across 13 sites showed good reproducibility
- Some sites showed differences in precision and accuracy
- Use of the system suitability and curves will help identify sources of variation



“Fit-for-purpose” Tier 2 Biomarker Verification



Minimum Information Needed in Publications

- Complete disclosure of LC-MS methods used
  - Dwell time, cycle time, RT window, AGC target, etc
- Statistics or "calculations" section in the Methods
  - LODs, LOQs, precision, accuracy, response curves, etc
  - Cite methods of calculation
- How is quantitation being done?
  - # transitions used for quantitation
  - Peptides per protein: combined or reported separately
  - What assumptions are made?
  - How is the curve being used?
- Address interferences
  - How are data evaluated for interferences?

CPTAC VWG Participants & Acknowledgements

Broad Institute: Susan Abbatiello, Terri Addona, Michael Burgess, Steven A. Carr, Hasmik Keshishian, Eric Kuhn, D.R. Mani, James Markell
Buck Institute for Age Research: Michael P. Cusack, Bradford W. Gibson Jason M. Held, **Birgit Schilling**
Fred Hutchinson Cancer Research Center: Amanda G. Paulovich, Jeffrey R. Whiteaker, Shucha Zhang
Indiana University: Mu Wang, Jong-Won Kim, Jimsan You
Massachusetts General Hospital: Steven J. Skates
Memorial Sloan-Kettering Cancer Center: Paul Tempst, Mousumi Ghosh
National Cancer Institute: Emily Boja Tara Hiltke, Christopher Kinsinger, Mehdi Mesri, Henry Rodriguez, Robert Rivers
NISS: Xingdong Feng, Nell Sedransk, Jessie Xia
NIST: Paul Rudnick
New York University: John Lyssand, Thomas A. Neubert, Åsa Wahlander, Sofia Waldemarson, Pawel Sadowski
Plasma Proteome Institute: N. Leigh Anderson
Purdue University: Charles Buck, Fred Regnier, Dorota Inerowicz, Vicki Hedrick
University of California, San Francisco: Simon Allen, Susan J. Fisher, **Steven C. Hall**,
University of North Carolina: David Ransohof
University of Victoria: Christoph H. Borchers, Angela Jackson, Derek Smith
University of Washington: Michael MacCoss, Brendan MacLean, Daniela Tomazela
Vanderbilt University: Daniel Liebler, Kent Shaddox, Corbin Whitwell, Lisa Zimmerman

Funding: National Cancer Institute

Targeted Proteomic Measurements in Biology and Medicine

Christoph Borchers

June 18, 2013

Targeted MRM/MS Experiments

Our Goals

- To rapidly and precisely quantitate a multiplexed panel of proteins in human biofluids using an absolute quantitative proteomic strategy.
- To verify and validate the candidate disease biomarkers toward clinical use.

Projected Users

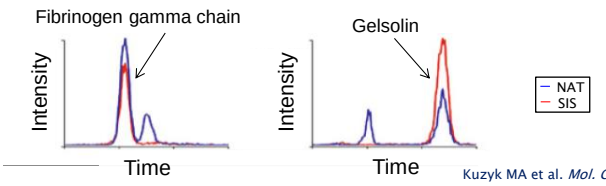
- Verification and validation ⇒ quantitative proteomic researchers in academic and industrial laboratories.
- Clinical implementation ⇒ technicians in clinical and pharmaceutical laboratories.

Q/A

- Will MRM-based technology eventually replace ELISAs?
  MRM-based techniques are ideal for expediting the verification and validation phases of the protein biomarker pipeline. These will complement, but not replace ELISAs in the clinic.

Confirmation of Analyte Signal

- We use stable isotope-labeled standard (SIS) peptides.
  - $^{13}C/^{15}N$ labeled standards are chemically identical to unlabeled counterpart and are distinguishable by mass only.



Kuzyk MA et al. *Mol. Cell Proteomics* 8 '09 1860.

- 3 transitions/peptide empirically targeted for RT verification and interference screening in the control.
- 1 transition/peptide in the final MRM method.

Q/A

- How else can the target peptide identities be confirmed?
  Automated transition building from the intensities of untargeted discovery experiments (*e.g.*, iSRM software Kiyonami et al. MCP '11).

Ionization Suppression

- Correct for ion suppression and matrix effects through the use of $^{13}C/^{15}N$ labeled analogs of the unlabeled analytes (peptides in our case).
- SIS peptides behave identically to their NAT counterpart in terms of chromatographic retention, electrospray ionization, and gas-phase fragmentation.
  - Produce same pattern of product ions.
  - Distinguishable by precursor and/or product ion *m/z* only.

Question

- If labeled internal standards are so effective at alleviating ion suppression, why then is protein quantitation with them not yet universal?
  Cost of the standards is a deterrent, while the reproducibility and transferability of the technique is a misconceived limitation.

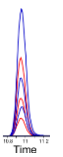
Interference Assessment

Approach with Normal Biofluid

- Monitor 3 MRM transitions/peptide form under matrix and matrix-free conditions.

Criteria for Interference-free Transitions

- identical SIS and NAT retention times
- symmetrical and similar SIS and NAT peak shapes
- absence of co-eluting ion pairs
- 2 of 3 transitions must obey the above and have av. relative ratio between SIS in buffer, SIS in biofluid, and NAT in biofluid <20% CV



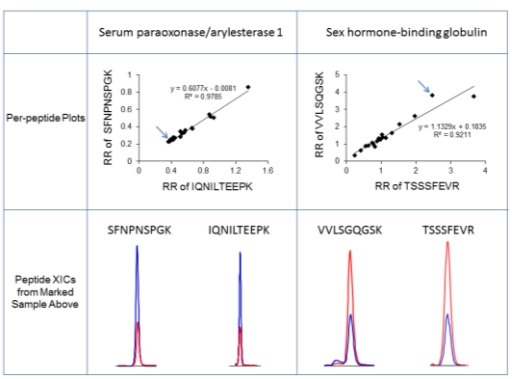
Peptide (Protein)	Product Ion	Relative Ratios to y_7^+			Av. Relative Ratio (%CV)
		SIS in buffer	SIS in plasma	NAT in plasma	
LSITGTYDLK (AAT)	y_6^+	0.56	0.55	0.54	0.55 (2)
LSITGTYDLK (AAT)	y_7^+	1.00	1.00	1.00	1.00 (0)
LSITGTYDLK (AAT)	y_8^{2+}	0.35	0.35	0.40	0.37 (7)

Quantifier

Interference Assessment

Approach with Unknown Biofluid Samples

- 1 transition/peptide
- ≥2 peptides/protein



Q/A

- What are the limitations of this approach?
  1. Lengthy and manual data analysis.
  2. Insufficient no. of peptides/protein may cause reliable peptides to be disqualified.

Analyte Plex-level per Injection

- We typically target >100 peptides in a single run.
  - 348 peptides (149 plasma proteins) were recently targeted in a quantitative MRM analysis.

    Percy AJ et al. *Biochim. Biophys. Acta* '13 in-press.
- Robustness is independent of plex-level.

  ➡ av. CVs <10% for signal and <0.05% for RT
- Equivalent protein concentration and LOQs are obtained.

Q/A

- Is the ability to reach higher plex-levels a current technological limitation?

  It appears so. The 348 peptides targeted in a single run is pushing the current limits, whereby cycle times are <1 s and dwell times are sufficient to obtain ~10 points across the chromatographic profile of the reconstructed peptide.

Assay Qualification Criteria

Our Strategy

- rigorous 2-stage qualification of normal biofluid:

  interference screening

  std curve generation
- qualification of unknown samples:

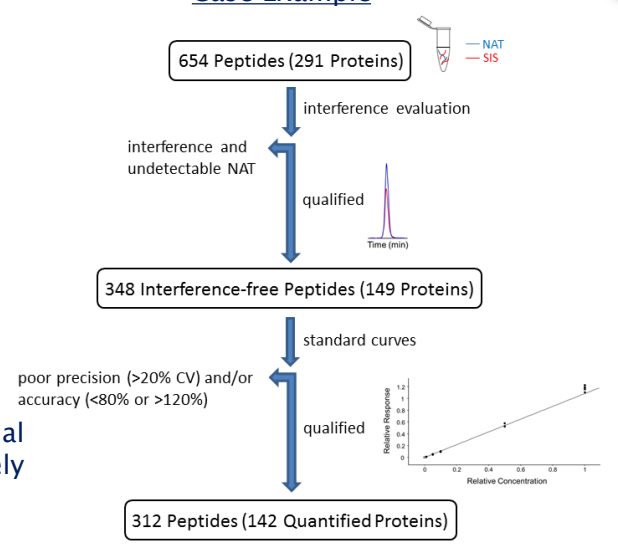
  interference screening

Q/A

- How can the internal standards be effectively quantified?

  AAA and CZE

Case Example



Protein Quantitation Method

- We target each peptide's highest responding, interference-free MRM transition and multiple peptides per protein.
- Peptide standard curves are generated with control biofluid.

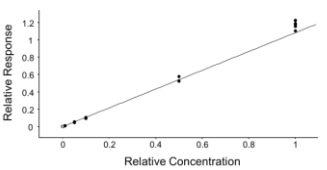
- A min. of 3 precise (<20% av. CV/level) and accurate (av. 80-120%/level) concentration levels must qualify to quantify.
- merits: multiplexing ability and sample throughput
- limitations: manual and laborious data analysis

Q/A

- Is a single MRM transition sufficient for accurate quantitation after preliminary interference screening?
  Yes, since our qualification criteria for generating curves is strict. This enables precise and accurate protein quantitation.

Use of Standard Curves

- Protein concentrations (NAT_{conc}) in healthy and diseased samples are determined from regression equations of control.



$$NAT_{conc} = \frac{SIS_{conc} \times m}{RR - b}$$

- RR is measured from the sample, while SIS concentration, slope, and y-intercept are all known values.

Question

- What number of replicates for the concentration levels of the control and the unknown samples is appropriate?
  We use 5 replicates/level for the control and 1 replicate for the unknown, with blank injections in between.

Determining Protein Concentrations

- Strict qualification criteria enables interference-free peptides to be precisely quantitated.
- Peptide acts as a surrogate for intact protein.
- Multiple peptides per protein are targeted with the highest determined concentration deemed representative of the protein concentration.
- Quantitative accuracy, based on ELISAs, is a challenge.

Question

- Is applying correction factors to compensate for sample loss or poor digestion efficiency acceptable at the quantitation stage?
  Yes, if the recovery is known.

Analytical Tools and Software

Analytical Tools

- SPE → desalt and concentrate
- 1D and 2D RPLC → improve peak capacity
- MRM/MS on triple quadrupole → sensitive with wide DR

Software

- vendor-specific → freeware cannot generate plots with NAT and SIS reversed
- Excel or IgorPro → figure and table generation

Question

- Should vendor independent software for generating curves with quantitation based on regression equations be created?
  Yes, we are planning to develop this.

Detail Required in Manuscripts

UVic Genome BC PROTEOMICS CENTRE

Experimental Conditions

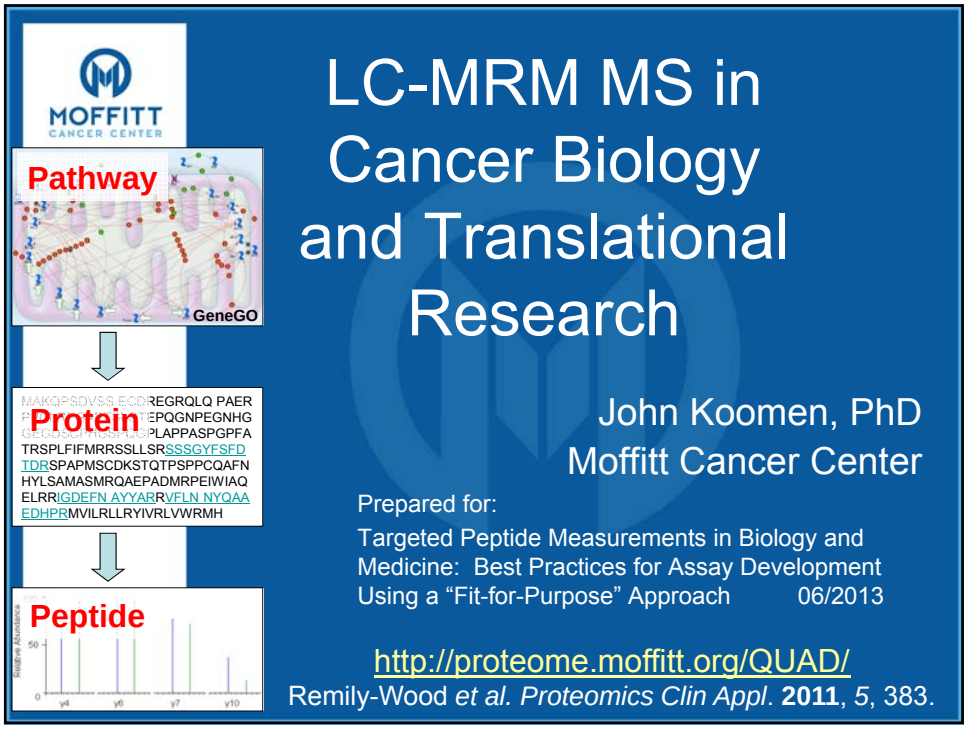
- Peptide/protein list and Uniprot accession numbers.
- Chromatographic details.
- Acquisition parameters and MRM transition lists.
- Linearity, sensitivity, and precision of quantitation.
- Protein concentrations.

Verification of Results

- Qualification and quantitation criteria.

Question

- Should raw data be uploaded into public databases?
  Yes.

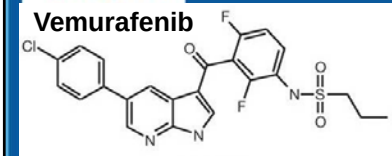

MOFFITT
CANCER CENTER
Pathway
GeneGO
Protein
TRSPLFIFMRRSSLLSRSSSGYFSFD
TDRSPAPMSCDKSTQTPSPPCQAFN
HYLSAMASMRQAEPADMRPEIWIAQ
ELRRIGDEFN AYYARRVFLN NYQAA
EDHPRMVILRLLRYIVRLVWRMH
Peptide
LC-MRM MS in Cancer Biology and Translational Research
John Koomen, PhD
Moffitt Cancer Center
Prepared for:
Targeted Peptide Measurements in Biology and Medicine: Best Practices for Assay Development Using a "Fit-for-Purpose" Approach 06/2013
http://proteome.moffitt.org/QUAD/
Remily-Wood et al. Proteomics Clin Appl. 2011, 5, 383.

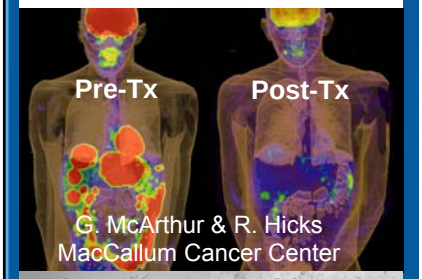

MOFFITT
CANCER CENTER
Customers
• Cancer Biologists
– Evaluate Model Systems/Perturbations
• Compare Cell Lines
• Drug Effects
• Pathway Activation
• Confirmatory to Westerns, More Quant. Info
• No Antibody, No Problem-Assay Development for New Analyte
• Clinicians
– Measure Correlates in Clinical Trials
– Develop Diagnostics/Prognostics
– Signature of Drug Response/Therapeutic Escape
• Sponsored Research Agreement
– Develop Commercial Assays for Implementation in Models & Tissues
• Compare Against Other Methods: Precision, Accuracy, Linearity, Sensitivity

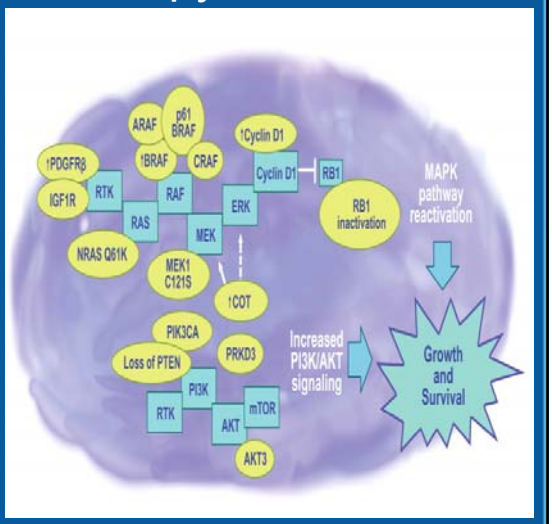
Resistance to B-RAF V600E Targeted Therapy in Melanoma

MOFFITT CANCER CENTER

Vemurafenib



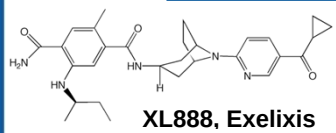




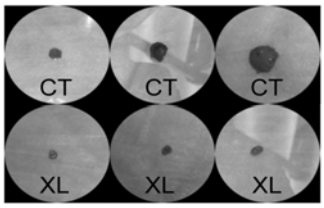
Sosman *et al. N Engl J Med*. **2012**, *366*, 707.
Paraiso & Smalley. *Cancer Discov*. **2012,** *2*, 390.

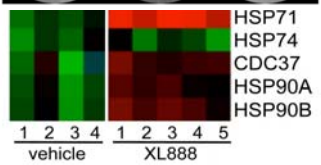
LC-MRM Quantification of Biomarkers of HSP90 Inhibition

MOFFITT CANCER CENTER



XL888, Exelixis



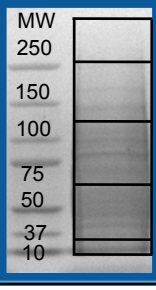


Paraiso *et al. Clin Cancer Res*. **2012**, *18*, 2502.

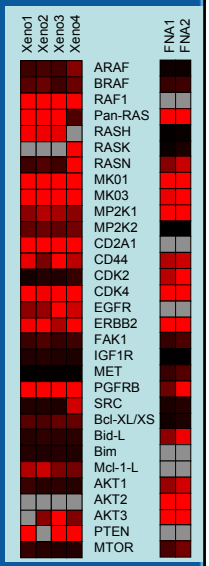
ClinicalTrials.gov: NCT01657591

HSP90 Clients

- **MAPK**: ARAF, BRAF, CRAF, COT, p90RSK
- **RTK**: IGF1R, PDGFR
- **PI3K/Akt**: Akt, Raptor, S6, PDK1



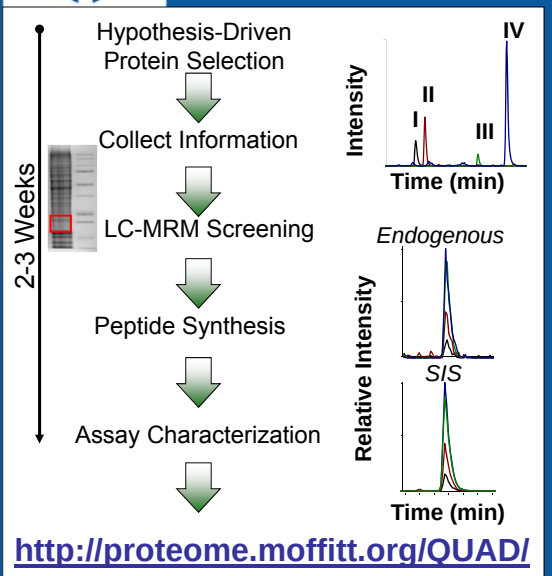
Peptides	Transitions
33	178
145	697
250	1254
194	935
92	514



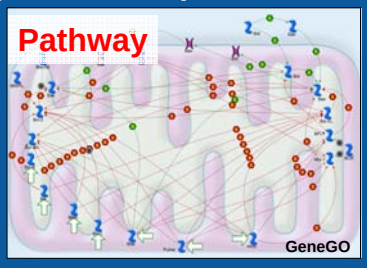
Distribution of Experiments

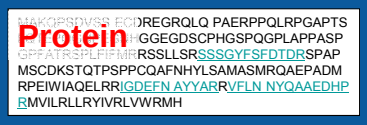
- Cancer Biologists:
  - Measurements
    - Expression 80%
    - Modification (PO_4, ABPP) 15%
    - Mutation 5%
  - Assay Type
    - Detection (not in LC-MS/MS) 5%
    - Relative Quantification 70%
    - Absolute Quantification 25%
      - Normalized to Housekeeping Protein (GAPDH) to Compare to Western
- Clinicians and Commercial Development
  - Measurements
    - Expression 95%
    - Phosphorylation 5%
  - Assay Type
    - Absolute Quantification 100%
      - Evaluation against Quality Control Standards (Normalize and/or Drop Data)

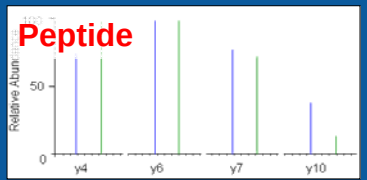
LC-MRM Assay Development



http://proteome.moffitt.org/QUAD/







Remily-Wood *et al. Proteomics Clin Appl*. **2011**, *5*, 383.

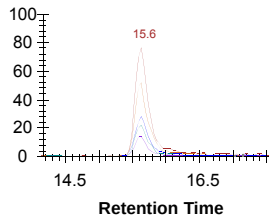
Software

- Assay Setup
  - TSQ and Xcalibur (Historical)
  - Skyline[1]
  - LC-MS/MS Data Pipeline: Sequest, TPP[2], Skyline
  - Uniqueness: iPIP (in-house), BLAST, Databases
- Data Acquisition
  - TSQ and Xcalibur (Thermo)
- Data Analysis
  - QuanBrowser (Historical, Mainly Cal. Curves)
  - MRMer (Historical)[3]
  - Skyline
- Statistical Analysis
  - MatLab, R (Y. Ann Chen)

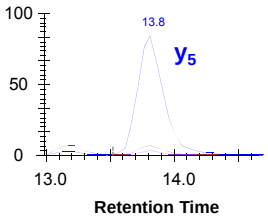
1. MacLean B, *et al*. ***Bioinformatics*** 2010, *26*, 966.
2. Deutsch EW, *et al*. **Proteomics** 2010, *10*, 1150.
3. Martin DB, *et al*. ***Mol Cell Proteomics*** 2008, *7*, 2270.

Standards and Transitions

BRAF-IGDFGLATVKSR



CTNB1-AIPELTK



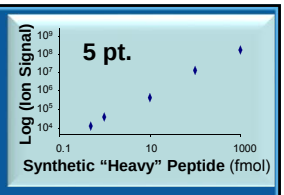
- SIS Peptides for Best Protein Expression Assay (1°) and for Unmodified and Modified Peptide Pairs
- Single Conservative Amino Acid Replacement for 2° Protein Expression Assays*
- Exploring Use of SIS Proteins and Mutant Proteins as Standards
- 3-5 Transitions Per Peptide
  - Detection > Peak Shape > Ratios Correct
- Transitions Selected for Quantification
  - **All**: Detected and Each Above 10% Base Peak
  - **Single Most Intense**: Other Fragments Detected at Threshold or Poor Peak Shape
  - **Just Those Detected**: No Better Choice and No More Sample

Remily-Wood and Koomen.** ***J Mass Spectrom. 2012, ***47*, 188.**

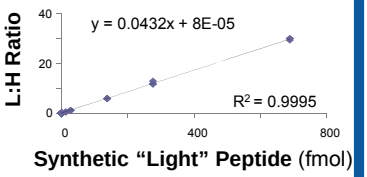
Standard Curves:

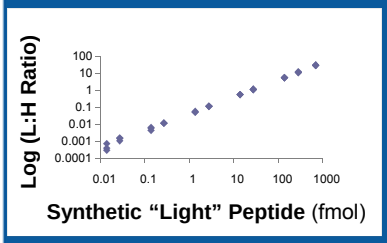
ERα-LASTNDKGSMAMESAK

12 Point Calibration Curves with Synthetic Peptides

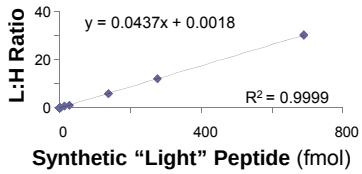


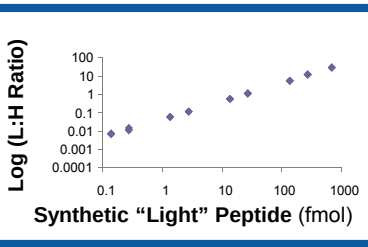
In Buffer



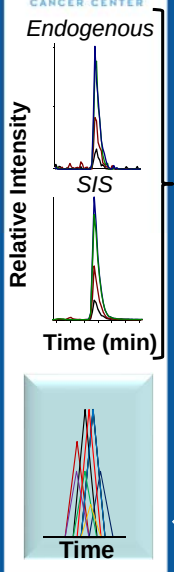


In Matrix





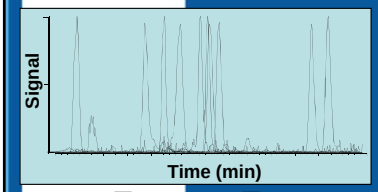
Peptide ID and Verification



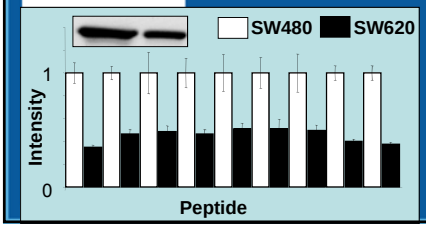
- Start from Defined Biological Question (➔ peptide)
- LC-MRM Screening Method
  - Enrich Protein
  - Screen "Usual Suspects"
- Depending on "Poor MS/MS with Few Fragments"
  - Comparison to Internal Standards (Tier 2)
  - Transition Ratios
    - Tier 2: Comparison Against Standard and/or Quantification (+/- 5%)
    - Tier 3: Ranking or "Eyeball Test" in Skyline
    - Differences between LC-MS/MS and LC-MRM
    - Loss of Transitions at Low Abundance
    - Higher CV for Lower Transitions
  - Retention Time
    - Tier 2: Comparison Against Standard
    - Tier 2/3: Correlation of LC-MS/MS and LC-MRM Elution Times
    - Tier 3: LC-MRM Data "Eyeball Test" - Plot Data and Peaks from All Samples Must Overlap

Clinical Translation

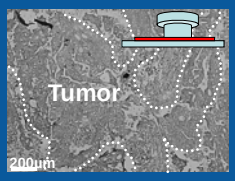
Assay Development

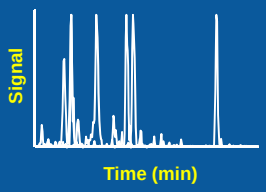


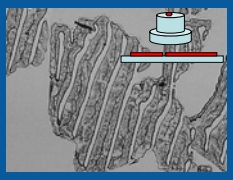
Verification in Cell Lines

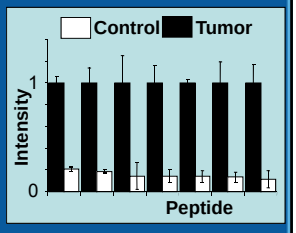


Clinical Implementation









Chen *et al.* *J Proteome Res.* **2010**, *9*, 4215.

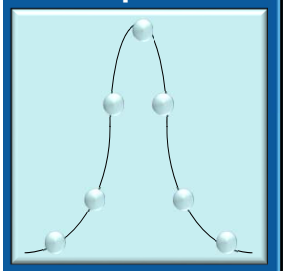
Development Milestones

- Matching Endogenous and SIS Peptide RT and Transition Ratios
- Compare Data Against Other Method in Known Biological System
  - Westerns
  - ELISA
  - Nephelometry (Other Antibody-Based Techniques)
- Translation to Clinical Specimens

Analyte Multiplexing

- Unscheduled (TSQ Quantum Ultra and Vantage, Thermo)
  - Limit of Peak Sampling: ~30 peptides
  - Gradient (30 min)
- Scheduled (Vantage, Thermo)
  - Limits of Peak Sampling
    - 400 Peptides for Relative Quantification (Tier 3)
    - 32 Proteins: 250 Peptides with 1,254 Transitions
    - Dependent on Peptide Hydrophobicity
  - Gradients up to 90 min (Tier 3 Assays)
- Combining Assays
  - Prefractionated Protein
    - SDS-PAGE
    - Affinity Enrichment
  - Loss from Less Material, Not Sampling and Not Interferences
  - CV, LOD, LOQ due more to Sample Type/Matrix Background, not Multiplexing

20 ms/Transition
>7 Samples/Peak



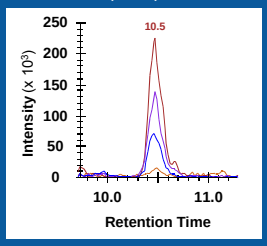
20 seconds

Detection of Interference

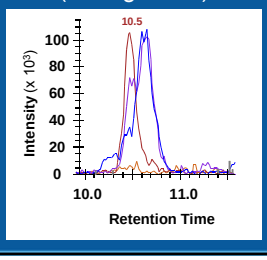
- Peak Shape
- Transition Ratio Measurements

- Interventions:
  - Omit Transitions
  - Add New Transitions
  - Lengthen Gradient
  - Narrow Quad Resolution?
  - Drop Data and Redefine Strategy for Protein Enrichment

CADH1-VTEPLDR
(SIS)



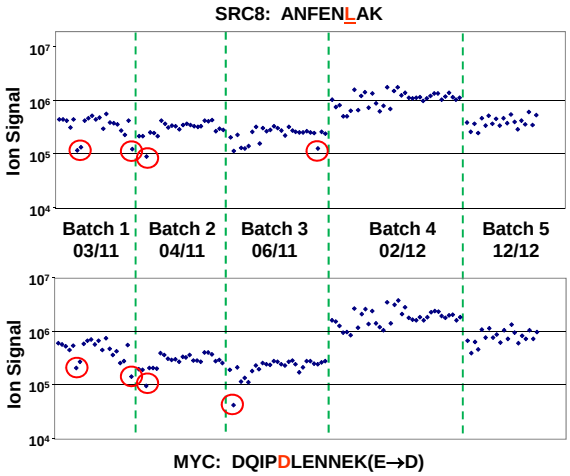
CADH1-VTEPLDR
(Endogenous)



Ion Suppression

- Compare to SIS/Review SIS Peptide Signal

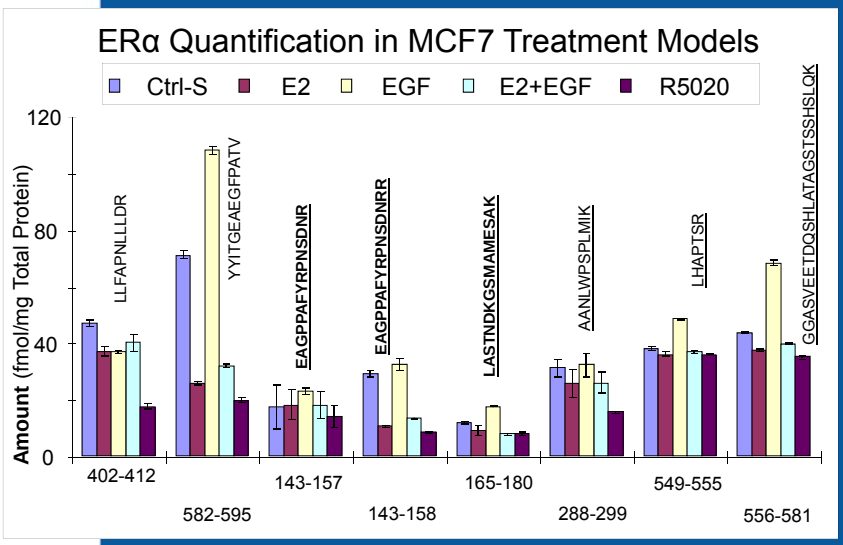
- Enrichment
  - Fractionation
  - IP
  - IMAC
  - IEF
  - ABPP



Which Peptide(s) are Best?

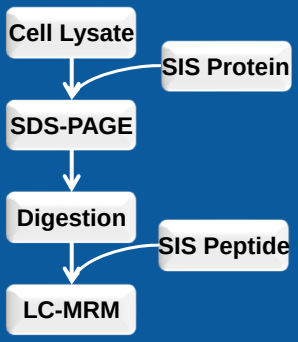
- Peptides per Protein
  - n = 1-14 including Modified Peptides
- If they Disagree?
  - Homework:
    - Modified or Mutated?
    - Location in protein? Accessible for Digestion?
    - Size? Hydrophobicity? Difficulty in Recovering Peptide?
    - Repeated Sequence?
  - Significance by T-test or ANOVA
  - Monitor Ratio to Others
- Defer to Best Peptide (with SIS)

LC-MRM of ER Peptides



Accuracy of Quantification

- Use of SIS Proteins
- Comparison against ELISA for Multiple Proteins
  - Slopes, Intercepts, R^2 Values



Quantitative Comparison	SIS Peptide	SIS Protein
SRC Amt (fmol/mg)	19.17 ± 0.32	28.36 ± 1.14
CV	1.6%	6.0%
Difference	---	+ 48%
p value	---	0.008

LC-MRM vs. ELISA

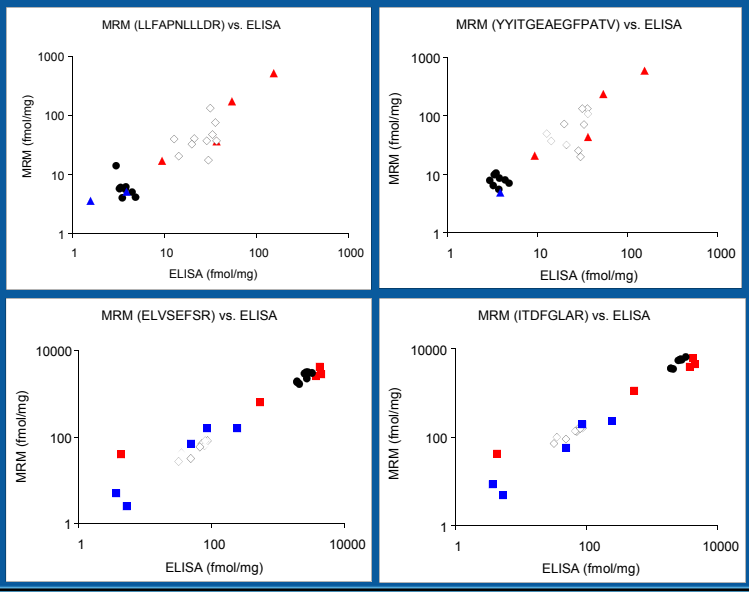
MOFFITT CANCER CENTER

ER

- Patient (ER+)
- Patient (ER-)
- MCF7
- BT474

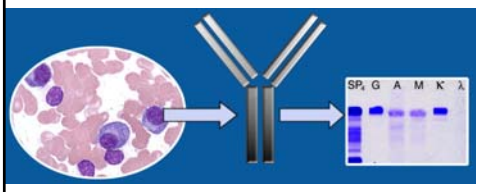
Her2

- Patient (Her2+)
- Patient (Her2-)
- MCF7
- BT474

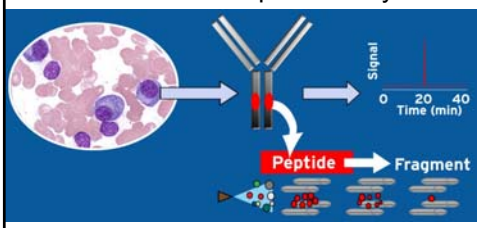


Immunoglobulin Quantification

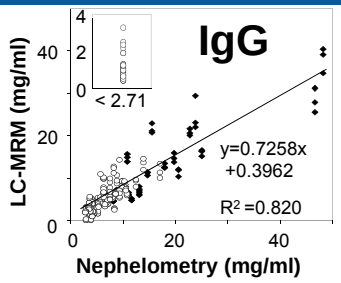
Antibody-Based Protein Assay

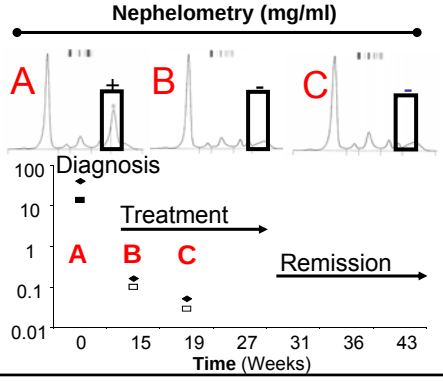


MS-Based Peptide Assay



Koomen J, *et al. Mol Cell Proteomics* 2008, 7, 1780.





MIAMRME

- Pathway, Protein, Peptide, Transitions
  - Method for Peak Selection/*in silico* evaluation of uniqueness?
  - Export Transition List from Acquisition Software
- Internal Standard Sequence, Mass, *etc*.
- Level of Multiplexing
- Purification/Enrichment Method
- Digestion Method
- QA/QC Standards and Metrics
- Instrument Settings
- Peaks Selected for Quantification and Rationale
- Data Normalization Method
- Evaluation of Sample-to-Sample and Batch-to-Batch Variability
- Comparison to Other Techniques (when available)

Acknowledgments

Koomen Lab/Proteomics
Yi Chen, PhD
Elizabeth Remily-Wood
Yun Xiang, PhD
Bin Fang, PhD
Wei Guan
Victoria Izumi
Robert Sprung, PhD

Colon Cancer
David Shibata, MD
Erin Siegel, PhD
Timothy Yeatman, MD
Domenico Coppola, MD
Emina Huang, MD (UF)

Lung Cancer
Eric Haura, MD
Jiannong Li, PhD
Guolin Zhang, PhD
Lanxi Song, MS

Multiple Myeloma
William S. Dalton, MD PhD
Ken Shain, MD PhD
Lori Hazlehurst, PhD
Rachid Baz, MD
Kaaron Benson, MD

Biostatistics and Cancer Informatics
Y. Ann Chen, PhD
Kate Fisher
Steven Eschrich, PhD
Richard Liu
Eric Welsh, PhD

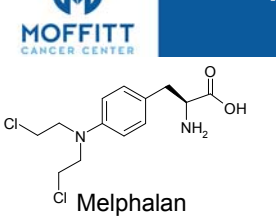
Breast/Proteome Sciences
Anthony Magliocco, MD
Heather Han, MD
David Britton, PhD
Ian Pike, PhD

Promega
Mike Rosenblatt, PhD
Robin Hurst, PhD

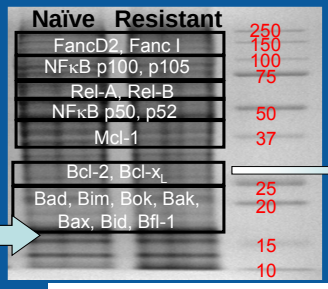
Melanoma
Keiran Smalley, PhD
Vernon Sondak, MD
Jane Messina, MD
Kim Paraiso
Eirik Haarberg

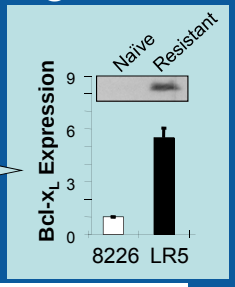
Funding: R01-CA077859, R21-CA152345, U54-CA143970, R01-CA161107, R01-CA123174, Lung SPORE P50-CA119997, R21-CA141285, CCSG P30 CA076292, UF-Moffitt Partnership, US Army DoD National Functional Genomics Center, Florida DoH Bankhead Coley Cancer Research Program, Proteome Sciences, Promega

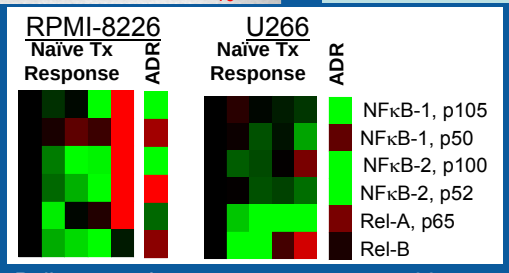
Melphalan Resistance Signature



Melphalan
↓
DNA Damage
↓
NFκB Signaling
↓
Fanconi Anemia → Repair DNA
Bcl-2 Family → Evade Apoptosis
→ Survival

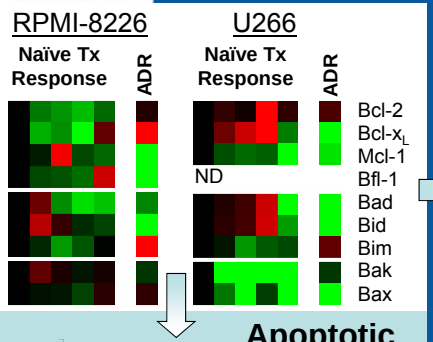


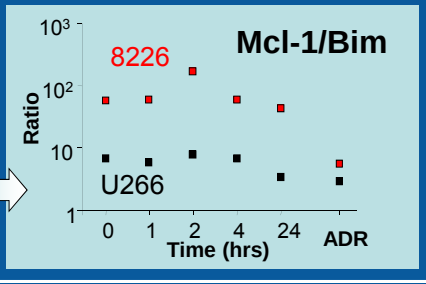


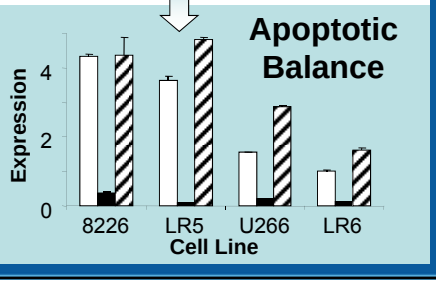


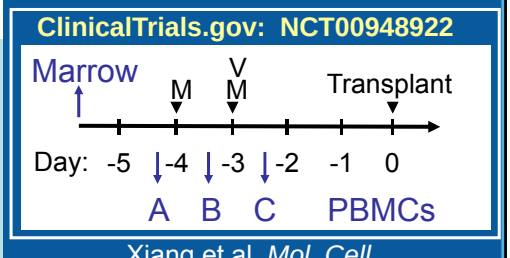
Bellamy *et al*. *Cancer Res*. **1991**, *51*, 995.
Hazlehurst *et al*. *Cancer Res*. **2003**, *63*, 7900.
Yarde *et al*. *Cancer Res*. **2009**, *69*, 9367.

Melphalan Resistance Signature









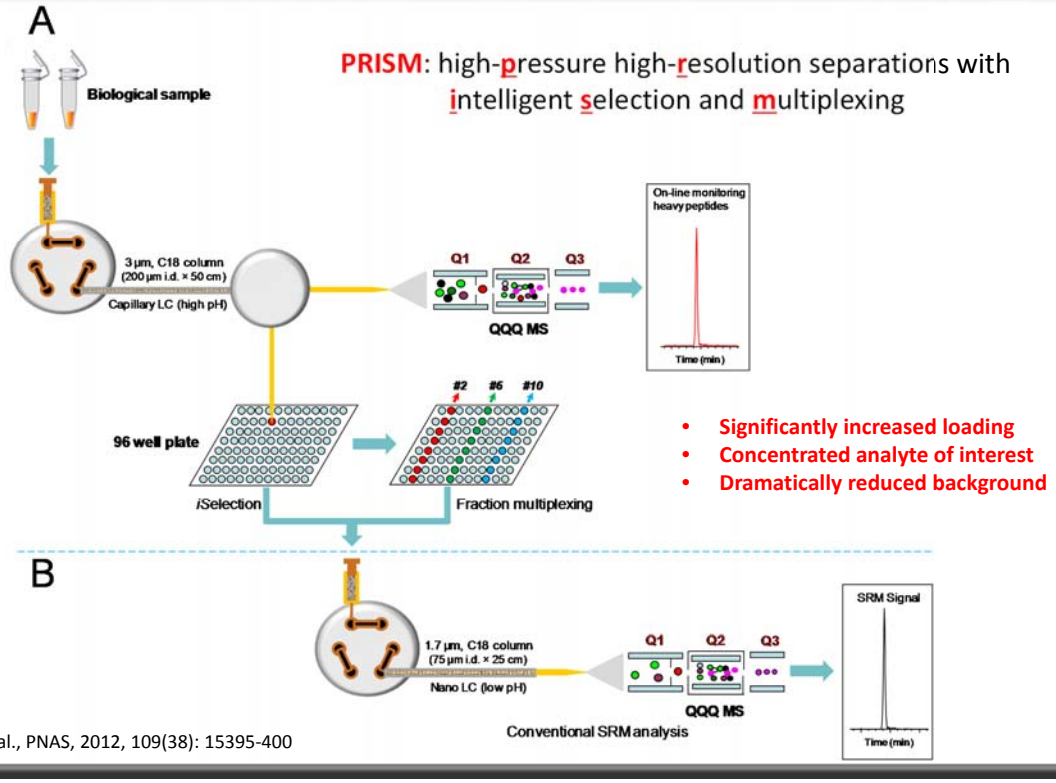
Xiang et al. *Mol. Cell. Proteomics* **2011**, *10*, M110.0055.

CLINICAL PROTEOMIC TUMOR ANALYSIS CONSORTIUM

Best Practices for PRISM-SRM Assay Development

Tao Liu, Ph.D.
Pacific Northwest National laboratory
6/18/2013

PRISM-SRM workflow



Shi et al., PNAS, 2012, 109(38): 15395-400

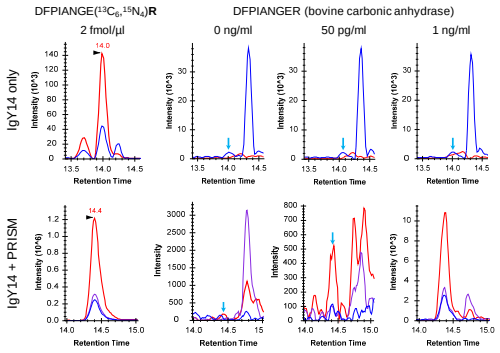
Typical figures of merit for PRISM-SRM assays

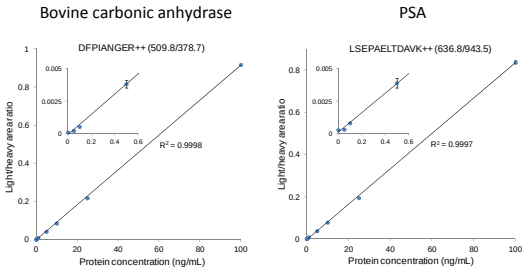
- **LOD** (S/N>3):
  - 50-100 pg/mL with IgY14
  - 0.3-1 ng/mL without IgY14
- **LOQ** (S/N>10):
  - 50-300 pg/mL with IgY14
  - 0.5-5 ng/mL without IgY14
- **CV**: <10%

Table 1. Summary of the LOD and LOQ of four target proteins in female plasma with IgY14 only and IgY14-PRISM

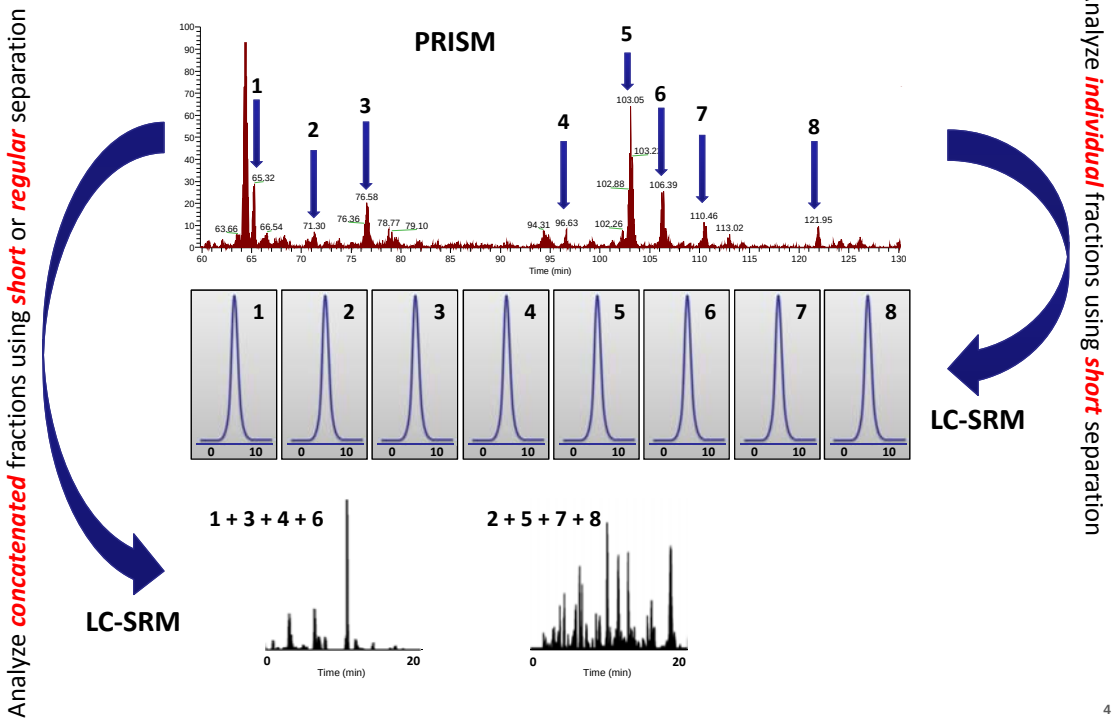
Proteins	Surrogate peptides	Fractionation strategies	LOD, ng/mL	LOQ, ng/mL
Bovine carbonic anhydrase	DGPLTGTYR	IgY14 only	5	10
		IgY14 plus PRISM	0.05	0.5
	DFPIANGER	IgY14 only	5	10
		IgY14 plus PRISM	<0.05	0.05
Bovine β-lactoglobulin	VLVLDTDYKK	IgY14 only	0.5	5
		IgY14 plus PRISM	<0.05	<0.05
	VYVEELKPTPEGDLEILLQK	IgY14 only	5	10
		IgY14 plus PRISM	<0.05	<0.05
E. coli β-galactosidase	VDEDQPFPAVPK	IgY14 only	10	25
		IgY14 plus PRISM	0.5	5
	LWSAEIPNLYR	IgY14 only	10	100
		IgY14 plus PRISM	0.05	0.1
PSA	IVGGWEcamCEK*	IgY14 only	25	100
		IgY14 plus PRISM	0.05	0.1
	LSEPAELTDAVK	IgY14 only	5	10
		IgY14 plus PRISM	<0.05	<0.05

*Cysteine was synthesized as carbamidomethyl cysteine.





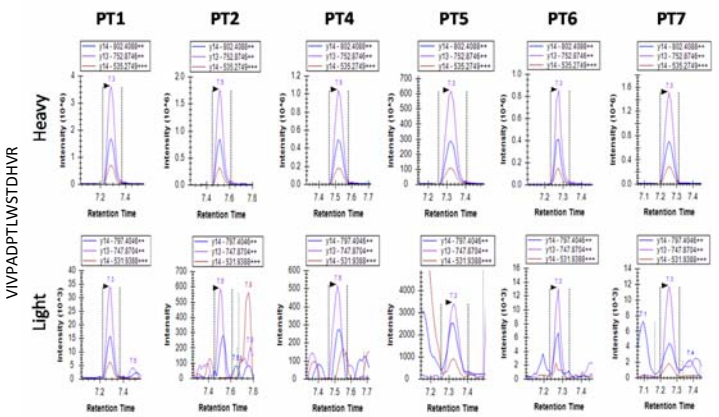
Improved throughput in PRISM-SRM



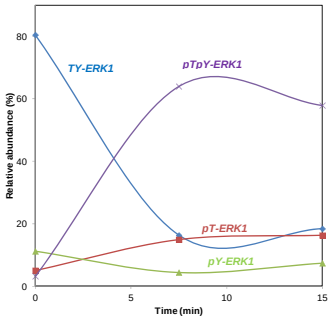
Q/A: Experimental goals and applications

- **What are the goals of the targeted MS experiments? Who are the likely users of the methods and results?**
  - **Experimental goals**: To sensitively and precisely quantify the concentration(s) of peptide(s) in a tryptic digest of a complex protein matrix, but not the accurate amount of the protein in the original sample
  - **Applications**: Verification of low abundance biomarker candidates (e.g., fusion protein products); highly sensitive quantification of proteins and PTMs (e.g., phosphorylation) in systems biology

Quantitation of TMPRSS2-ERG fusion proteins in prostate tumor tissues



Phosphorylation stoichiometry of ERK1

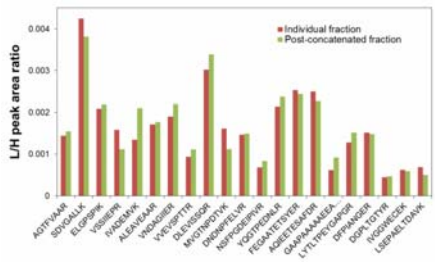


TY-ERK1: VADPDHDHTGFLTEYVATR
*p*T-ERK1: VADPDHDHTGFL*T*EYVATR
*p*Y-ERK1: VADPDHDHTGFLTE*Y*VATR
*pTp*Y-ERK1: VADPDHDHTGFL*T*E*Y*VATR

Q/A: Effects of concatenation/pooling

- **What is the typical range of analyte plex-level per-injection? What is the impact of the plex-level used on performance of the PRISM-SRM assays?**

Depleted plasma, regular gradient

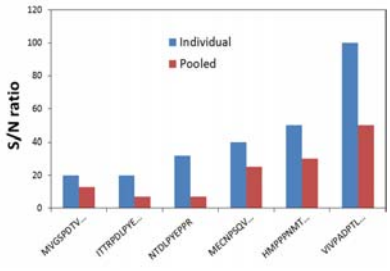


Non-depleted plasma; regular gradient

Surrogate peptide (best transition)	Pooling fraction			Individual fraction			Relative error (%)
	S/N	L/H	CV (%)	S/N	L/H	CV (%)	
DGPLTGTYR (490.2/597.3)	2.7	0.00380	25.4	2.8	0.00323	11.1	17.6
DFPIANGER (509.8/378.7)	13.1	0.00130	3.6	17.1	0.00138	1.5	5.7
VLVLDTDYKK (597.3/981.5)	33.4	0.0253	7.6	82.5	0.0279	15.9	9.0
IVGGWECEK (539.2/865.3)	18.3	0.00224	28.6	25.9	0.00210	1.2	6.7
LSEPAELTDAVK (636.8/943.5)	20.9	0.00167	18.8	30.8	0.00204	17.0	18.5

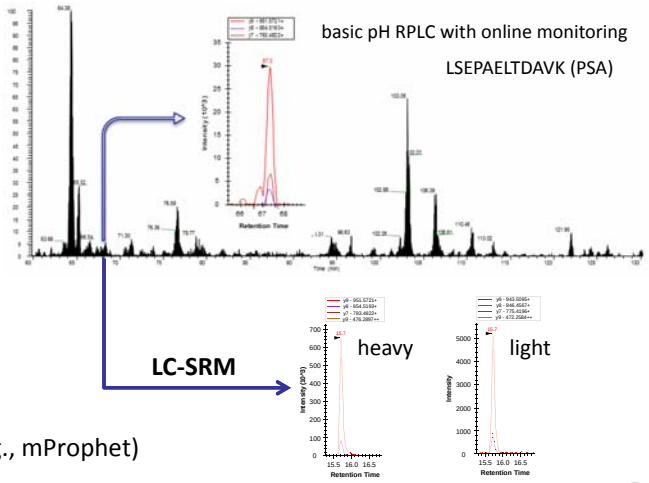
- Typical plex-level in PRISM-SRM: 5-30 (can be much larger)
- Concatenation to >10 fractions has minimal impact
- Accuracy is largely unaffected
- Impact on precision is evident for LLOQ level analytes
- Impact on sensitivity is matrix- and separation-dependent:
  - potential non-characteristic loss in pooling and concentration steps
  - less impact in depleted plasma (vs. non-depleted plasma)
  - pooling followed by short separation diminishes signal by 1/3 to 2/3

Tumor tissue; short gradient



Q/A: Confidence of measurements

- **How is the confidence of measurements established?**
  - Through matching to transition profiles (>3 transitions) and LC elution of heavy isotope-labeled internal standards
  - PRISM separation (basic pH RPLC) with online monitoring adding another level of confidence (i.e., orthogonal with 2nd dimension acidic pH RPLC separation)
  - Increased confidence by monitoring more transitions (e.g., iSRM, QTRAP)
  - Confidence further established through titration (response curve)



- **How do these methods differ from those in "Discovery Proteomics"?**
  - Confidence typically governed by use of internal standard
  - Usually does not have full spectrum
  - No well-established scoring algorithm (e.g., FDR); new tools are emerging (e.g., mProphet)

Q/A: Quantification method

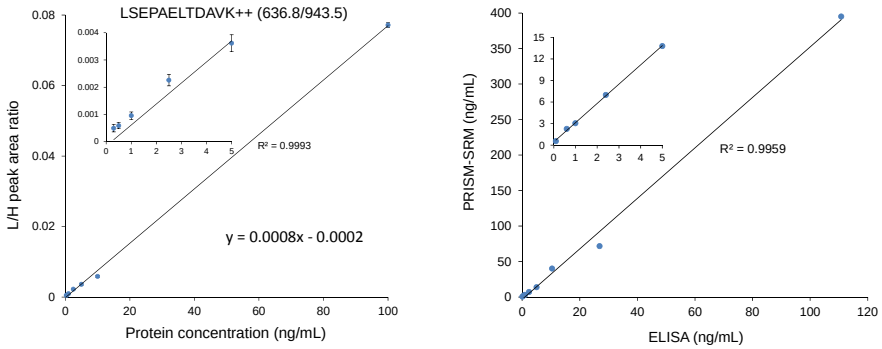
- **Explain the method of quantification.**
  - Peptides are quantified through Light/Heavy ratios (peak area) and response curves
  - All transitions are checked for potential interference (e.g., AuDIT)
  - The most sensitive transition is used for quantification
  - "Crude" peptides can be used for comparative studies

- **Discuss the capabilities and limitations of the approach.**
  - **Strength**: use of internal standards mitigates many issues that could affect quantitation accuracy (e.g., sample prep and instrument analysis reproducibility)
  - **Limitations**: 1) synthetic heavy peptide standards are required for PRISM-SRM; 2) cost is high if purified and accurately quantified internal standards are used; 3) 3-5 weeks lead time for peptide synthesis; 4) not "true" absolute quantitation

Q/A: Use of standard curves

- **If you generate standard curves (calibration or response curve), explain how you use them to assess the quantitative accuracy of the assay.**
  - The slope and y-intercept from the curve regression are used in calculating the analyte concentration in a sample (***process*** replicates for at least one of the data points)
  - External calibration can be used, but is not required
  - Correlation to other measurements (e.g., ELISA) can be made

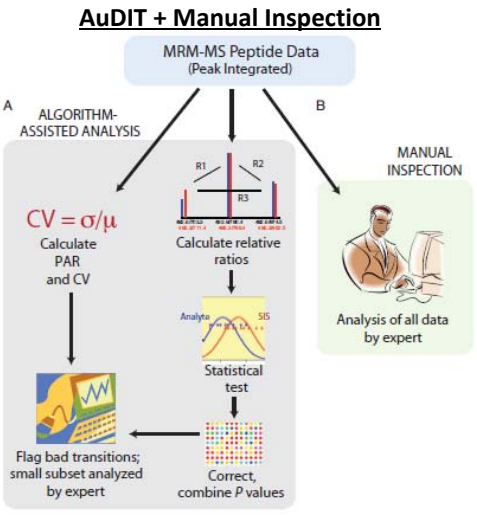
PRISM-SRM analysis of PSA in non-depleted female plasma



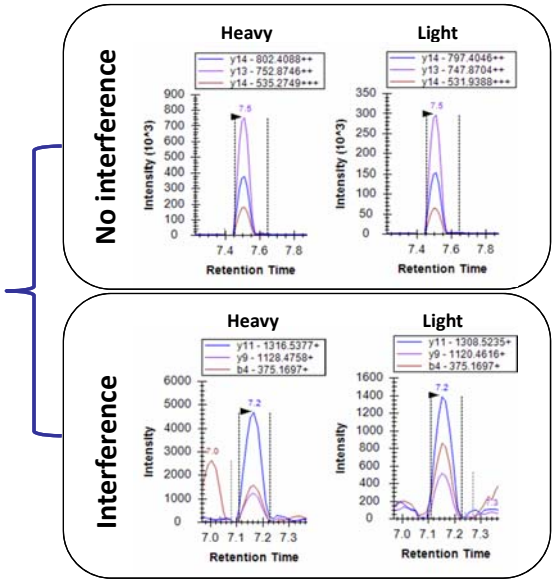
*60% of immunoreactivity of spiked PSA could be lost due to complexation to A2M in plasma; SRM is detecting true "total PSA"

Q/A: Detecting interference

- **What methods are use to establish presence of interferences? How do you deal with them if detected?**
  - Check heavy transitions in both buffer and matrix; compare light and heavy in matrix using AuDIT
  - If interference detected, either change transitions or refine LC gradient (e.g., long gradient)



Abbatiello et al., Clin Chem, 2010, 56:2, 291-305



Q/A: Data analysis tools; ionization suppression

- **What software and analytical tools do you use in your studies and why?**
  - In-house programs are used for peptide selection from our own data repository
  - Skyline and in-house algorithms are used for transition selection (e.g., analysis of Orbitrap LC-MS/MS data) and CE optimization (e.g., direct infusion, CE ramping in scheduled SRM)
  - Skyline is used for LC-SRM scheduling and method generation
  - Skyline is used for data visualization and quantification
  - Skyline is used for data sharing (Panorama is being tested)

- **How do you account for suppression of ionization in your quantification method?**
  - Because heavy isotope-labeled internal standards are used throughout PRISM-SRM analysis, we do not need to account for suppression of ionization

Q/A: "Bottom-up" protein quantification

- **Can you provide a useful estimate or accurately determine the amount of protein in the matrix based on the measured levels of peptides? Explain how/why. Indicate experimental parameters such as number of peptides per protein and the criteria/computational tools applied.**
  - The amount of proteins in the matrix can be estimated:
    - use protein and heavy isotope-labeled peptide when building response curve (2-3 peptides; spike-in samples go through the entire sample prep process)
    - if protein standard is not available, measure the same actual clinical samples using both PRISM-SRM and other measurements (e.g., ELISA) and establish calibration curve

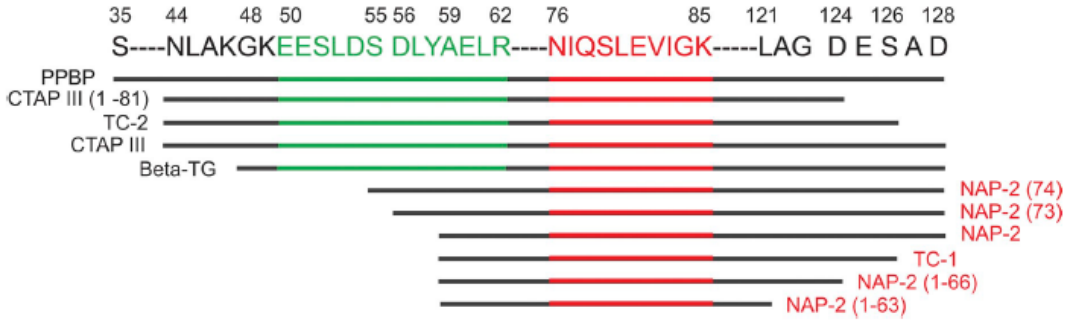
- **If you have multiple peptides from the same protein and each gives a different answer for the extrapolated protein level, how do you deal with this?**
  - Make sure there is no potential PTM site (e.g., phosphorylation, N-glycosylation)
  - Check protein isoform information
  - Check if there is potential motif that inhibits trypsin digestion
  - Cheer up, some great discovery might have been made!!!

An example: protein isoforms

Pro-platelet basic protein (PPBP)

		100 HC, 50 T1D			10 HC, 10 T1D				50 T2D			
Protein	Peptide	DASP Verification Cohort			DASP Blind Validation Cohort				T2D Specificity Cohort			
		FC[a]	P-value[a]	AUC	Specificity	Sensitivity	FC[a]	P-value[a]	FC[b]	P-value[b]	FC[c]	P-value[c]
PPBP	EESLDSDLYAELR	1.6	8.41E-08	0.77	0.3	1	1.6	4.93E-02	−1.7	1.71E-04	11.0	4.44E-12
PPBP	NIQSLEVIGK	8.5	6.62E-18	0.93	1	1	30.4	1.08E-05	−1.2	7.44E-02	32.6	4.44E-12

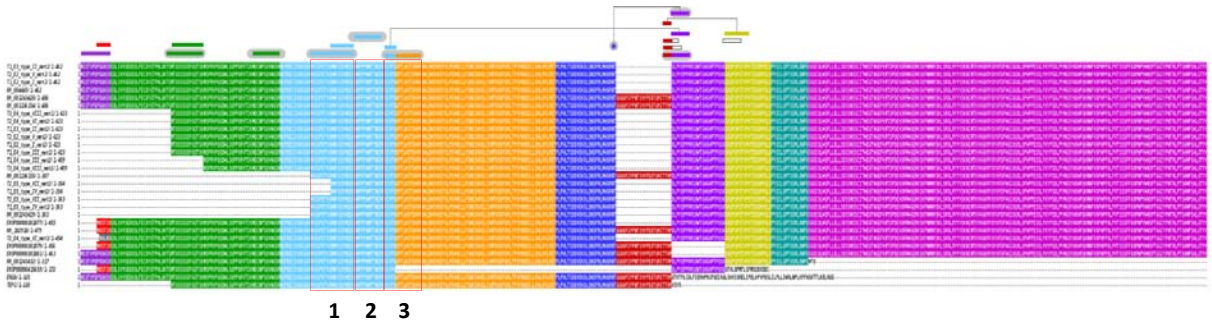
[a]T1D in respect to HC.
[b]T1D of blind set in respect to T2D.
[c]T2D in respect to HC of blind set.



Zhang et al., J Exp Med. 2013, 210(1): 191-203

TMPRSS2-ERG fusion protein products

All possible TMPRSS2-ERG fusion protein products in prostate cancer

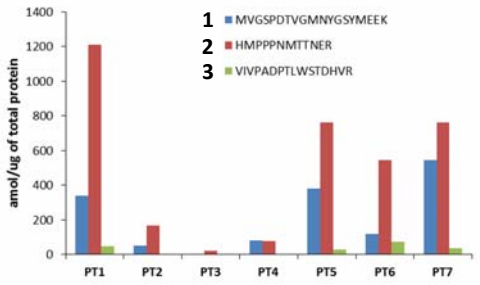


1 2 3

Multiple protein products are detected in TMPRSS2-ERG fusion positive tumors

Sequence	amol/μg of total protein						
	PT1	PT2	PT3	PT4	PT5	PT6	PT7
TEMTASSSSDYGQTSK							287
MVGSPDTVGMNYGSYMEEK	339	50		80	382	117	546
HMPPPNMTTNER	1210	169	20	78	760	543	762
VIVPADPTLWSTDHVR	48	NQ		NQ	30	74	35
ITTRPDLPYEPPR*	31	10					
NTDLPYEPPR*	26	10		NQ			

*mutually exclusive peptides



He et al., manuscript in preparation

Q/A: "Qualify" the assays

- **How do you "qualify" your measurements/assays, i.e., what criteria do you use or think appropriate to say that your measurements/assays has been successfully developed?**
  - Detection of endogenous signal in intended biological matrix
  - Response curve is linear spanning >=3 orders (reversed curve OK)
  - CV<20%
  - Endogenous level: >2~5 X LLOQ, to be clinically useful
  - If to be used for relative comparison, recovery evaluation can be done, but not required

Q/A: Min. requirement for publication

- **What information do authors need to provide in their manuscripts/supplement to enable reviewers and readers to understand what was done and to be able to judge the confidence of the measurements made?**
  - PRISM separation conditions: column dimensions, packing material, gradient, loading, fraction collection, concatenation
  - LC-SRM conditions in details: column dimensions, packing material, gradient, loading, MS conditions
  - Extracted ion chromatograms (XICs) for both light and heavy signals at claimed detection level (for at least three transitions; if not available at this level, show simultaneous detection of >3 transitions at a higher concentration level)
  - S/N ratios for each transition measured
  - CV for each transition measured
  - Response curve for each transition used for quantification

Acknowledgement

- Tujin Shi
- Weijun Qian
- Jintang He
- Chaochao Wu
- Qibin Zhang
- Thomas Fillmore
- Yuqian Gao
- Athena Schepmoes
- Mahmud Hossain
- Joe Brown
- Karin Rodland
- David Camp
- Dick Smith

- NCI CPTAC
- NCI EDRN
- NIGMS P41 BTRC (NCRR)
- NIH Director's New Innovator Award

Targeted Peptide Measurement in Biology and Medicine

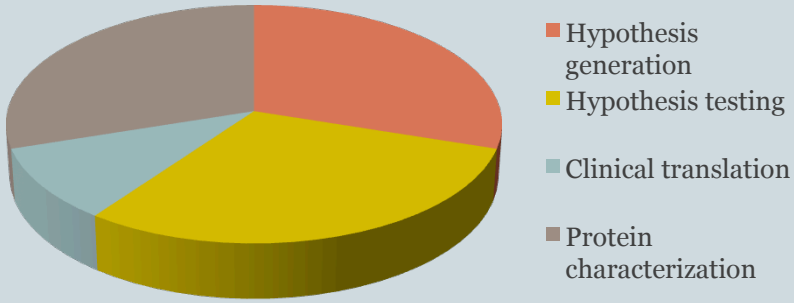
Brad Ackermann, Ph.D.

Biological Mass Spectrometry
Laboratory for Experimental Medicine
Eli Lilly and Company
19-June-2013

Background

- Group charged with performing biological protein MS for Lilly Research Labs
- Applications
  - Biomarker discovery
  - Target validation/characterization
  - Biomarker verification/clinical translation
  - Reagent characterization and other forms of protein analysis
- Instrumentation: Orbitraps, triple quads, MALDI-ToF
- Customers: Program Teams (discovery, pre-clinical, clinical)
- Therapeutic areas: Onco, AI, MSK, NS, Endo
- Goal: use targeted MS for rapid, selective protein analysis
  - Improve biochemical understanding of drug targets and disease
  - Verify and translate protein biomarkers (PD, tailoring, disease)

Protein-MS Work Distribution



Protein Biomarker Workflow

Discovery → Qualification → Verification → Validation

Tier I **Tier II** **Tier III**

Q1: Level of Multiplexed Analysis

What range of analyte plex-level per-injection do you typically use in your experiments? What is the impact of the plex-level used on the robustness and figures of merit (CV, LOD, LOQ) of the developed MRM/SRM assays?

Level of Multiplexed Analysis

- We typically do not use LC/MS/MS for *de novo* hypothesis generation; rather, we use it to understand biology surrounding a target that is of interest to our portfolio.
- Most assays have fewer than 10 analytes; targeted assays MRM are developed using the fewest analytes possible to interrogate target biology.
- The influence of plex multiple on assay robustness/performance has not been investigated.
- Formal acceptance criteria are not applied in early stages.
- LLOQ is based on precision and accuracy (if protein std available). LOD is considered to be 3x std dev of blank.

Q2: Confidence in Protein Assignment

Explain how you establish confidence that what is being measured is the analyte of interest (e.g., match to spectra of an internal standards, match to reference spectra from discovery experiments, RT, etc.). How do these methods differ from "Discovery Proteomics" using data-dependent or data-independent experiments?

Confidence in Protein Assignment

Targeted MRM

- MRM methods are not prepared without peptide standards.
- We synthesize standards for all surrogate peptides of interest and use full scan MS/MS, RT, and exact *m/z* (if available) to confirm assignments.
- Blast searches are performed to verify the uniqueness of all surrogate peptides used.

Non-Targeted Label-Free Analysis

- *De novo* hypothesis generation work done using label-free methods on an Orbitrap; MS1 used for quantification with data-dependent MS/MS for peptide identification.
- Workflow for discovery proteomics using label-free methodology has been published.[1,2] Measures taken to confirm protein assignments are discussed.

1. Higgs RE, et al., *J. Proteom.* Res. **4**, 1442-50 (2005)
2. Higgs RE et al., *Int. J. Proteom.* 674282, 1-10 (2013)

Q3: Method of Quantification

Explain your method of quantification, how many transitions you monitor and which ones are chosen to quantify. If you are using internal standards describe in detail how they are used. If you are not using internal standards, explain how you are quantifying. Discuss the capabilities and limitations of your approach.

Method of Quantification

- Labeled protein IS rarely used (availability/purity).
- SIL-IS prepared for all surrogate peptides, typically with 2-3 flanking residues.
- One to three transitions monitored depending on matrix complexity and method for clean-up.
- Extra transitions often used for confirmation only.
- *Relative* assay[1]– peptide std curves
- *Definitive* assay[1] – protein std curves
- Protein standards are well characterized.
- Peptide stds are qualified by AAA for peptide content (on fit-for-purpose basis).
- For *definitive* assays, spike-recovery and std addition are used to qualify a surrogate matrix.

1. Lee JW, et al. *Pharm. Res.* **23**, 312-328 (2006).

Q4: Standard Curves

If you generate standard curves (calibration or response curve), explain how you use them to assess the quantitative accuracy of the assay (e.g., are the slope and y-intercept from the curve regression used in calculating the analyte concentration in a sample? Is an external calibration curve used?).

Standard Curves

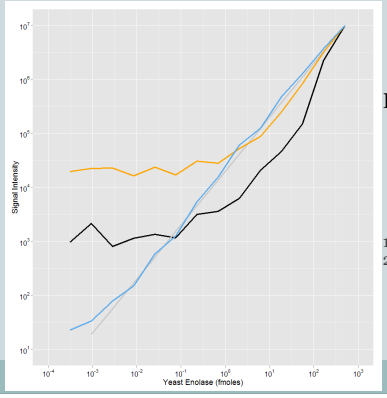
- External calibration using either peptide or protein standards are analyzed using peptide SIL internal standards.
- Standard curves are prepared in surrogate matrix $\geq$ 5 pts.
- Various fitting methods applied (1/x. $1/x^2$, 4PL).
- Authentic samples used whenever possible to define the working range of the assay.
- Duplicate std curves which bracket samples used to verify assay performance. Back-calculated values for individual stds gives crude estimate of accuracy.
- As assays progress, it's important to understand the validity of the surrogate matrix used (i.e. demonstrate parallelism).
- Parallelism between surrogate and authentic matrix can be assessed for both *relative* and *definitive* assays.

Q5: Protein Estimation

Can you provide a useful estimate or accurately determine the amount of protein in the matrix based on the measured levels of peptides? Explain how/why. Indicate experimental parameters such as number of peptides per protein and the criteria/computational tools applied. If you have multiple peptides from the same protein and each gives a different answer for the extrapolated protein level, how do you deal with this?

Protein Estimation

- Linear statistical models used to roll up transitions to peptides and peptides to proteins.[1,2]
  - Peptide area = Peptide Constant * [Protein] → $\log(AUC_{ij}) = Peptide_j + Sample_i + e_{ij}$
- Dilution can be used to identify peptides or transitions.



Dilution of yeast enolase in UPS1 background
Black = average of all peptides
Orange = model estimate
Blue = dilution optimal model

1. Higgs RE et al., *Int. J. Proteom.* 674282, 1-10 (2013)
2. Chang C-Y, et al., *Mol. Cell. Proteomics* **11**, 1-12 (2012)

Q6: Interferences

Describe methods you use to establish presence of interferences and how you deal with them if detected.

Interferences

- Appropriate blank samples are included with all sample analysis runs.
- SRM ratios in biological matrix are compared to synthetic peptide standard.
- Linear dilution of signal in surrogate matrix or buffer also used.
- High reliance on anti-protein IP
  - Improved enrichment with reduced interference
  - Conventional sample prep (SPE, IEX) only used on as needed basis
- Interferences addressed by:
  - Changes in sample prep
  - Choose alternate transition or peptide
  - Change LC column or gradient
  - Use high resolution SIM instead of MRM (Q-Exactive)
- Important: watch out for endogenous ligand interference when using high affinity antibodies.

Q7: Ionization Suppression

How do you account for suppression of ionization in your quantification method?

Ionization Suppression

- Not a major issue owing to high reliance on SIL internal standards.
- Anti-peptide IP employed where ever possible over conventional methods (faster method development and reduced ion suppression).
- We do not specifically quantify ion suppression; it is therefore rolled up into our estimation of spike-recovery.
- Single biggest concern comes from detergents needed for protein handling and IP. We usually modify the sample preparation method to address this issue (change detergent, increase washing steps, etc.).

Q8: Assay Qualification

How do you "qualify" your measurements/assays, i.e., what criteria do you use or think appropriate to say that your measurements/assays has been successfully developed?

Assay Qualification

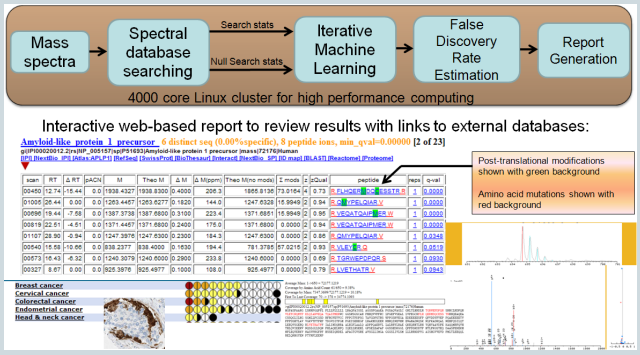
- Depends on the use of the assay (Fit-for-Purpose).
- Both *pre-study* and *in-study* validation considered.
- *Pre-study:* Total Error of 30% is considered the default for acceptance
  - Not applied to early work
  - Accuracy measured by spike-recovery using a protein std (*definitive*)
  - Accuracy can be estimated from peptide std or QCs (*relative*)
- *In-study*: bracketed std curves used to assess run acceptance.
- For later stage work, qualification of the surrogate matrix is performed:
  1. Std addition – does the surrogate and authentic matrix give similar slopes?
  2. Dilutional linearity – does the signal for an authentic biological sample decrease linearly when diluted in surrogate matrix?
  3. Spiked recovery – is recovery within $\pm$ 20% of theoretical?
- Important to address source of determinant error (e.g. stability or PAE).

Q9: Software Tools

What software and analytical tools do you use in your studies and why?

Software Tools

- In-house proteomics pipeline
  - Utilizes OMSSA, X! Tandem, Protein Pilot, and Peaks de-novo for peptide ID
  - High resolution label free MS1 quantification
  - MALDI-quant (Gutierrez *et al.* 2005)
- JMP, SAS, R for statistical analyses, Multiquant and Pinpoint for MRMs



Q10: Author recommendations

What information do authors need to provide in their manuscripts/ supplement to enable reviewers and readers to understand what was done and to be able to judge the confidence of the measurements made?

Author recommendations

- Sufficient detail needed to allow replication of work.
- More extensive supplemental sections recommended.
- Authors should include discussion on how/why they selected the level of validation used (explain fit-for-purpose rationale).
- Surrogate matrix qualification should be documented.
- Consistent nomenclature should be adopted for key terms regarding validation or quality assessment.
- Validation data should be included in publication of measurements.

Author recommendations

- *Nature* **496**, 398 (2013) – *Editorial announcement*

  To ease the interpretation and improve the reliability of published results we will more systematically ensure that key methodological details are reported, and we will give more space to methods sections. We will examine statistics more closely and encourage authors to be transparent, for example by including their raw data.

  Central to this initiative is a checklist intended to prompt authors to disclose technical and statistical information in their submissions, and to encourage referees to consider aspects important *for research reproducibility (go.nature.com/oloeip).*

Biomarker Verification
Recommendations for Tier 2

- Tier 2 assays do not need to be in full control; however, the major contributions to assay variation should be identified during development and addressed as an assay progresses.
- Total error of 30% is recommended for assay acceptance.
- Methods can be either *relative* or *definitive* based on the standards used.
- Assay precision should be measured for all methods and can be used for LLOQ assignment.
- Accuracy (bias) is measured in *definitive* methods by spike-recovery and requires a characterized protein standard.
- Spiked recovery using interleaving validation dilutions can be used to estimate bias without a protein std (*relative* methods).
- Parallelism should be demonstrated between the surrogate matrix used to prepare standards (and QCs) and the authentic biological matrix.
- Total error profile should be considered when selecting method for curve fitting.

Biomarker Verification

Recommendations for Tier 2 - continued

- SIL peptide internal standards should be made for each analyte.
- IS with flanking residues recommended, but not required.
- Std curves must have a minimum of 5 pts.
- In-study validation:
  - Duplicate std curves (bracketed) recommended
  - QC samples optional
- Selectivity blanks should be included in each run.
- More than one transition per analyte depending on sample prep and matrix.
- Fit-for-Purpose experiments to address pre-analytical effects need to be considered to avoid large sources of determinant error.

Protein Target and Biomarker Quantitation in Translational Research of Biologics

Targeted Peptide Measurements in Biology and Medicine:
Best Practices for Assay Development Using a "Fit-for-Purpose" Approach

Dr Hendrik Neubert
Associate Research Fellow
Head of Biomarkers & Mass Spectrometry
PDM NBE, Pfizer Inc; Andover MA
Bethesda – 18/19 June 2013

WORLDWIDE RESEARCH & DEVELOPMENT
Pharmacokinetics, Dynamics & Metabolism – NBE

Goals of Targeted MS Assays

WHAT:

- Quantify proteins targeted by therapeutics (preclinically 20% / clinically 80%)
  - Basal levels in normal and disease; all fluids and tissues
    - Parameter determination for mechanistic PK/PD models supporting target evaluation and FIH dose projections (translational pharmacology)
    - Precision medicine/patient stratification
  - Longitudinal assessment of target engagement (free, complex, total), in preclinical, FIH and POC studies
    - How hard did the drug hit the target to test the mechanism?
- Longitudinal assessment of mechanistic PD protein biomarkers (proximal to target/site of action)
  - Did we test the proposed mechanism of action?

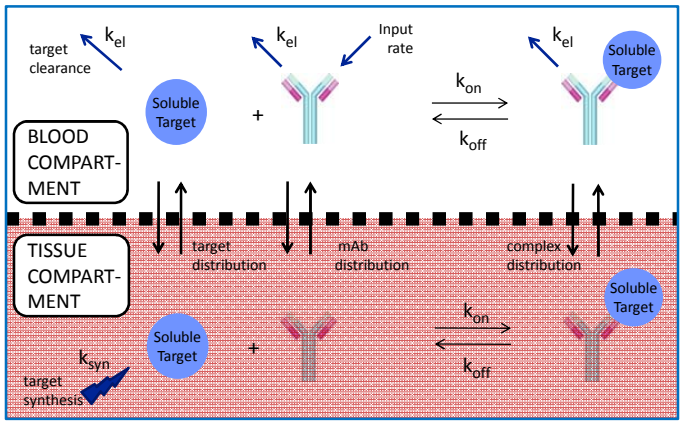
CUSTOMERS:

- PK/PD and Clinical Pharmacology modelers
- Precision Medicine/Clinical Scientists

WORLDWIDE RESEARCH & DEVELOPMENT
Pharmacokinetics, Dynamics & Metabolism – NBE

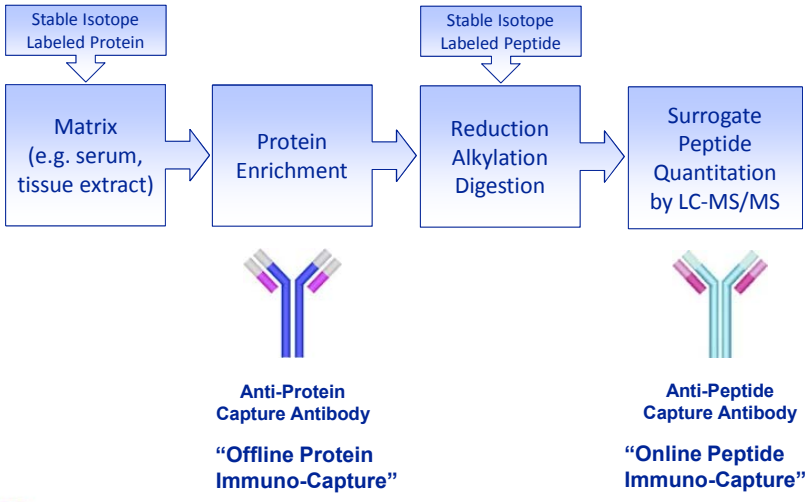
2

Mechanistic 'Site-Of-Action' PK/PD Model



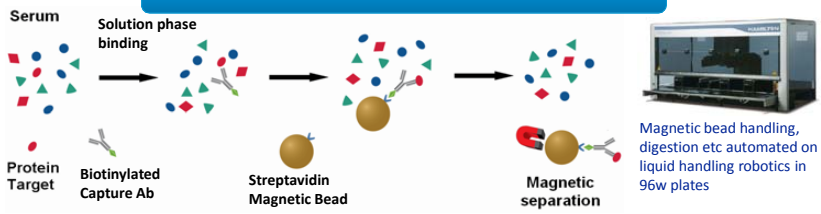
Immunoaffinity (IA)-LC-MS/MS Assays

configurations for protein quantification



Mechanics of Immunoaffinity Workflows

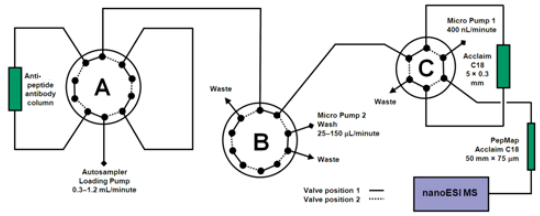
PROTEIN IMMUNOAFFINTY-ENRICHMENT



Magnetic bead handling, digestion etc automated on liquid handling robotics in 96w plates

(1) Fernández Ocaña, Anal. Chem.; 2012;84(14):5959 **(2)** Neubert, Anal. Chem.; 2013;85(3):1719

PEPTIDE IMMUNOAFFINITY-ENRICHMENT



NanoSpray Multiple Reaction Monitoring

(1) Neubert, Clin. Chem .2010;56(9):1413 **(2)** Palandra, Anal. Chem.; 2013;85(11):5522

5

Confidence in what's being measured

Assay development

- Protein and/or peptide immunoaffinity enrichment increases selectivity
- SIL peptide (coelution with native peptide; matching MRM transition ratios)
- MRM triggered MS/MS (if enough signal); match spectra
- Increase MS signal of native peptide by spiking…
  - *Recombinant protein prior to digestion*
  - *Digest of recombinant protein (or cell lysate, tissue digest containing protein of interest)*
  - *Light synthetic peptides*
- When possible: induce increase in protein in biological sample (cell line, primary cells, whole blood, etc) using external stimulus

Assay validation/implementation

- Assess assay selectivity as part of qualification/validation by analyzing protein in normal/disease populations:
  - *Evaluate potential for variable matrix-related interferences in independent sources of matrix*
  - *Recovery of spiked recombinant/endogenous pool into individual samples*
    - *E.g. Unspiked and spiked matrix (n>6) are analyzed in replicates and results are considered acceptable if 80% of samples are within +/- 25% % of the target value.*

Pfizer WORLDWIDE RESEARCH & DEVELOPMENT
Pharmacokinetics, Dynamics & Metabolism – NBE

6

Method of Quantification

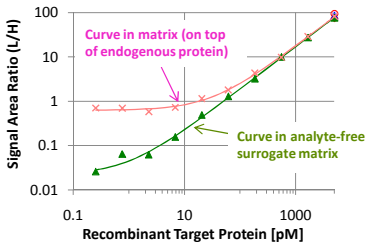
- Standard curve with recombinant protein in analyte-free surrogate matrix
- SIL peptides (typically use extended sequence containing trypsin cleavage sites) added to all samples, standards and QCs
- Monitor 2-3 transitions per analyte
  - usually use single transition (most intense) for calculating signal-area-ratio (native/SIL)
  - occasionally summing transitions if S/N advantage achieved
- Back-calculate samples against standard curve

Advantage: recombinant protein curve is proven approach in many assay formats including immunoassays; akin to small molecule LCMS; SIL peptide provides additional normalization for part of the workflow

Disadvantage: recombinant ≠ endogenous protein

Standard Curves

- Standard curve with recombinant protein in analyte-free surrogate matrix (buffer, depleted matrix from same species, same matrix from different species)
  - Demonstrate parallelism (dilution linearity of incurred sample)!!
  - Duplicate standards and blanks: one set of standards run before samples, one after



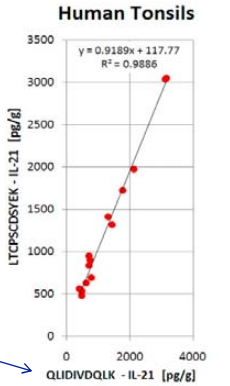
- Linear or non-linear curve fit
  - Non-linear fit common for many IA approaches
  - Depending of affinity of capture Ab
- Assess goodness of curve fit by back-calculating standards against curve
  - For assay validation/implementation -- if needed remove outliers using predefined criteria and document analytical reason; e.g. +/- 25%; *x* number of stds included in final curve; 75% of total numbers of stds included)
- Don't extrapolate above highest or below lowest standard (unless during exploratory method development)

Multiplexing vs Single Measurement

- Typically 1 protein analyzed (sometimes 2, rarely 3 or more)
- Typically 2-3 peptides per protein (sometimes 1, rarely more than 3)
  - Decide upfront what to do with the data from several peptides
    1. Use highest responding peptide
    2. Calculate average if difference no more than e.g. 20%
    3. Report different concentrations based on individual peptides for confirmed isoforms

Confirmatory peptide (good correlation between two peptides confirms validity to quantify with one peptide)

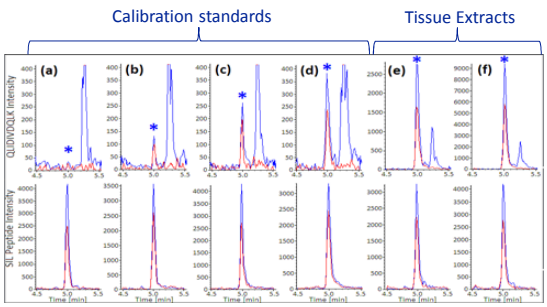
Lead peptide used for quantification



Palandra, Anal. Chem.; 2013;85(11):5522

Interferences

- Establish during method development using pooled sample(s) and assess spike recovery
- Assess assay selectivity using protein spike recovery as part of assay qualification/ validation from multiple matrix sources (different individuals; matching disease state, etc)
- Consistency of response of SIL peptide (e.g. across a 96w plate)
- Consistency of response of SIL peptide between buffer and matrix samples (ionization suppression; matrix effect?)



Palandra, Anal. Chem.; 2013;85(11):5522

Assay Qualification/Validation

- Fit-for-purpose assay qualification/validation
  - Stage-gate dependent increase of analytical rigor (early disc., lead development, preclinical tox, FIH, POC ….)
- How is the data used? … Exploratory biomarker versus defined clinical end-point?
  - Non-regulated versus regulated (GCP and GLP)
- Assay qualification/validation plan predefines acceptance criteria for each experiment
- Example of validation experiments performed for target biomarker assay for FIH study (GLP/GCP)
  - *Primary standard stock solution stability*
  - *Back-calculated calibration standards & calibration parameters*
  - *Intra-batch and inter-batch precision and accuracy (different target levels, different days, many replicates)*
  - *Matrix inter-lot accuracy precision*
  - *Effect of therapeutic on recovery*
  - *Dilution linearity/parallelism*
  - *Confirmation of LLOQ*
  - *Autosampler re-injection reproducibility*
  - *Benchtop (4C) matrix stability*
  - *Freeze/thaw matrix stability at -20C and -70C*
  - *Long term stability (1m, 3m, 6m, 1y …)*
  - *….*

Software

- Skyline (early MRM method development)
- Mostly vendor specific quantification software such as
  - Analyst Quantitation Wizard and MultiQuant (Sciex)
  - LCQuan and PinPoint (Thermo)
- Sometimes Excel and relevant data-analysis Add-on's for better multi-parametric curve fitting of immunoaffinity LC-MS/MS data

Publications

- Information on all topics discussed in this presentation should be provided (either for manuscript or supplement):
  - *Goal of measurement*
  - *Why the chosen assay format is most relevant (contrast to other assay formats) and suitable for the intended purpose*
  - *Rationale for selection of all standards (recombinant proteins, SIL peptide etc), capture Ab, surrogate peptides, MRM transitions.*
  - *Confidence in the measurement; highlight what makes the method specific*
  - *If single peptide is measured: why this is good enough?*
  - *If multiple peptides are measured: what to do with the data and why?*
  - *Illustrate analytical rigor and show fit-for-purpose assay qualification dependant on use of data*

Multiplexed Quantitative Analyses Conducted in PRM Mode

Targeted Peptide Measurements in Biology and Medicine:
Best Practices for Assay Development
Using a "Fit-for-Purpose" Approach

Bethesda, MD, June 18-19, 2013.

Bruno Domon, PhD
Head Luxembourg Clinical Proteomics Center
Invited Professor University of Luxembourg

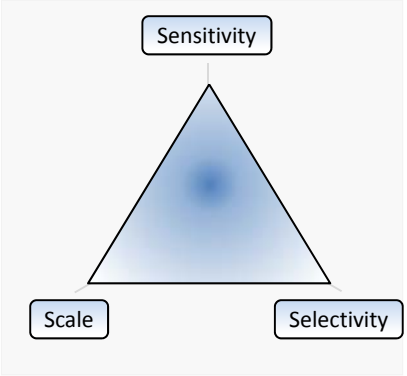
Challenges in Quantitative Proteomics

- **Complexity of proteomic samples**
  - Reduce sample complexity (*i.e.* biochemical background)
  - High selectivity of measurements
- **Wide dynamic range of protein concentrations**
  - High sensitivity (LOD / LOQ)
  - Cover a wide dynamic range
- **Experiment at scale**
  - Biological variability; *e.g.* assess "*sensitivity" and "selectivity" of markers*
  - Multiplexing capability, concomitant analysis of large panels of peptides
  - Throughput and robustness of platform to measure large cohorts

Domon - 2013 ...

Requirements for a Targeted Experiment

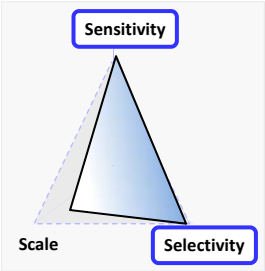
- **Selectivity**
  - Background interferences
  - Monitor multiple transitions
  - *Confirm identity of the analyte !*
    Full MS/MS spectrum (or surrogate)
- **Sensitivity**
  - Low limit of detection (LOD)
  - Measure low abundance components
- **Multiplexing capability**
  - Number of peptides in one LC-MS run
  - Schedule measurements *(elution time)*



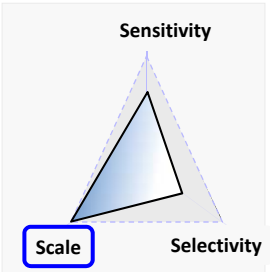
3

Targeted Proteomic Experiments

- **Precise Quantification**
  - Determine precisely the amount of analyte (e.g. *biomarker verification study*)
  - **Internal standards** (calibrated amount)
  - Limited number of analytes



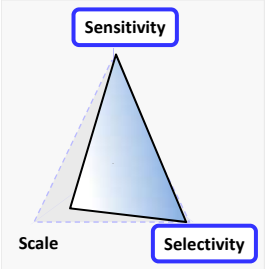
- **Screening Experiment**
  - **Detection** of peptides in biological sample
  - Differential analysis *(biology study / discovery)*
  - **Large-scale** (*hundred of candidates*)



Gallien et al.; J. Mass Spectrom., 2011

Domon - 2013 ... 4

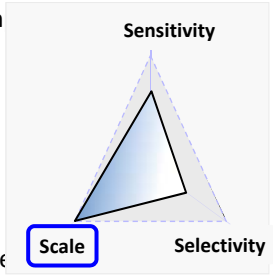
Targeted Experiment: Precise Quantification



- **Precise Quantification**
  - Determine precisely the amount of analyte (e.g. *biomarker verification study*)
  - **Internal standards** (calibrated amount)
  - Limited number of analytes
- **Requirements:**
  - Calibrated **internal standards**
  - Dilution curves to determine LOD and LOQ
  - Replication of analyses (analytical precision, CVs)
  - Confirmation of the analyte **identity** based on evidences *{Reference MS/MS spectrum; accurate mass}*
  - Assess presence of **interferences**
  - Rigorous QC protocols, and data review

Domon - 2013 ... 5

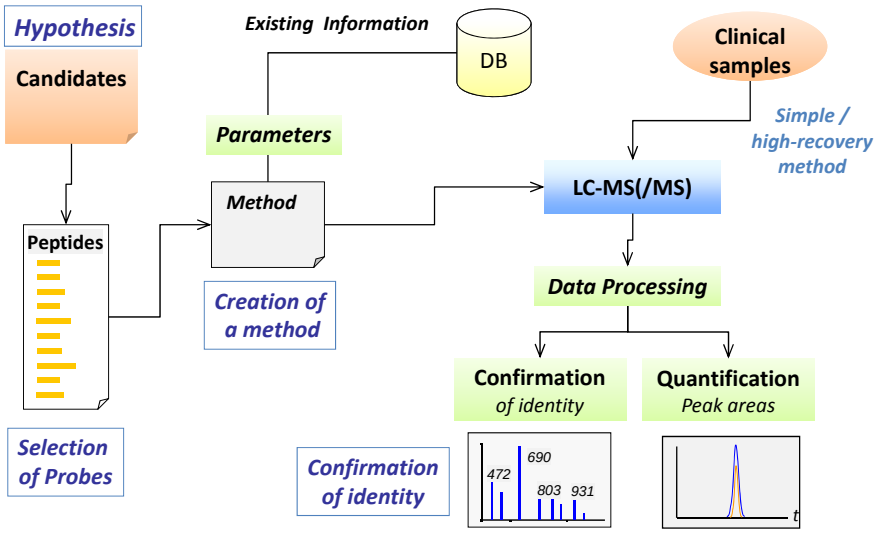
Targeted Experiment: Screening Mode



- **Screening Experiment**
  - **Detection:** presence of peptides in biological sam
  - Differential analysis *(biology study / discovery)*
  - **Large-scale** (*hundred of candidates*)
- **Requirements:**
  - Reference peptide (/spectra) desired
  - Focus on detection / concentration estimation
  - Relative quantification : changes in expression leve
  - Selectivity desired, however false positive are acceptable *{Confirmation in a secondary screen}*
  - Typically highly multiplexed experiments *{Precise scheduling of measurement required}*
  - Automated data analysis *{subsequent validation of results}*

Domon - 2013 ...

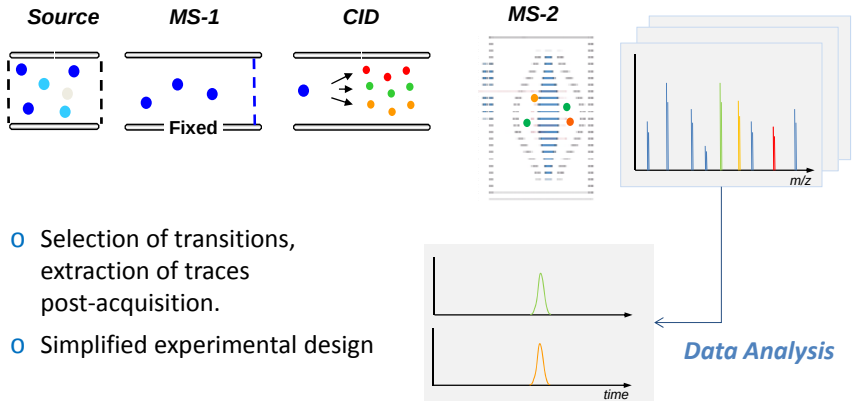
Targeted Proteomic Experiment



Domon - 2013 Kiyonami et al.; MCP (2011) ... 7

Parallel Reaction Monitoring (PRM)

- Performed on a quadrupole / orbitrap instrument (high-resolution)

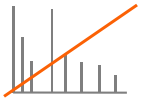


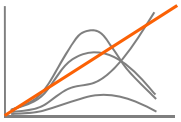
- Selection of transitions, extraction of traces post-acquisition.
- Simplified experimental design

Design of a PRM Experiment

Quadrupole – Orbitrap

PRM mode





Peptide Selection → *Transition Selection* → *Optimization*

Reference Conditions

> Protein_1

> Protein_2

> Protein_i

Peptides

Pep	Q1	Frag	CE	Time	Win	Fill time	Area
Pep_A	724	786	30	25.3	0.7	50	
Pep_A		887	30				
Pep_A		966	30				
Pep_A		1087	30				
Pep_B	863	754	35	17.5	0.7	50	
Pep_B		917	35				
Pep_B		1018	35				
Pep_B		1132	35				
.....							
Pep_N	563	797	25	24.3	0.7	50	
Pep_N		899	25				
Pep_N		999	25				
Pep_N		1101	25				

Design of Targeted Proteomic Experiments

SRM: *Peptide Selection* → *Properties* → *Transitions* → *Optimization* → *Acquisition* → *Data Processing*

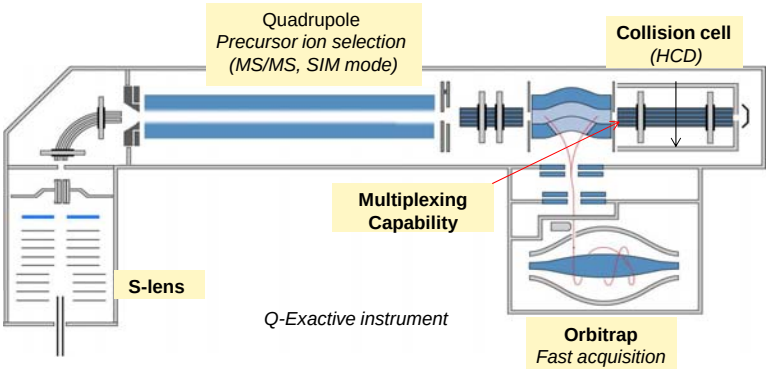
Pep	Q1	Q3	Time	CE	Win	Dwell	Int.
Pep_A	724	786	25.3	30	0.7	20	
Pep_A		887		30		20	
Pep_A		966		30		20	
Pep_A		1087		30		20	
Pep_B	863	754	17.5	35	0.7	50	
Pep_B		917		35		50	
Pep_B		1018		35		50	
Pep_B		1132		35		50	
.....							

PRM: *Peptide Selection* → *Properties* → *Optimization* → *Acquisition* → *Transitions* → *Data Processing*

Pep	Q1	Time	CE	Win	Fill	Frag	Int.
Pep_A	724	25.3	30	0.7	50	786	
Pep_A			30			887	
Pep_A			30			966	
Pep_A			30			1087	
Pep_B	863	17.5	35	0.7	50	754	
Pep_B			35			917	
Pep_B			35			1018	
Pep_B			35			1132	
.....							

10

Quadrupole / Orbitrap Mass Spectrometer



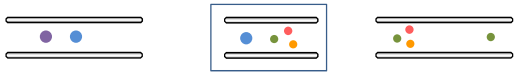
Main characteristics:

- **High resolution** (> 100,000)
- **Accurate mass** (ppm range)
- **Trapping** devise

Targeted *modes of operation:*

- *Method requires predefined analytes*
- ***SIM*** *mode: relies on precursor ions*
- ***MS/MS*** *(HCD) mode: relies on fragment ions*

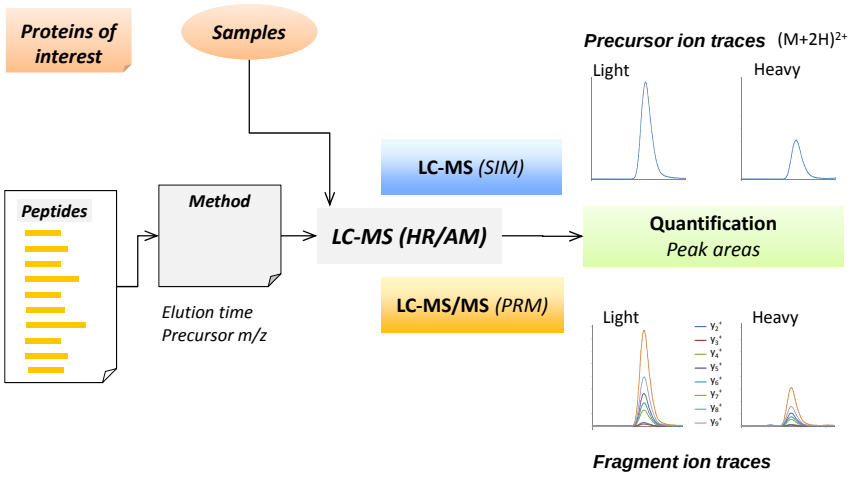
Targeted Proteomics Strategies



	Isolation	CID	Analyzer
SRM	1 u	Quadrupole	Quad (LR)
	(default)	*In-beam: flux*	*Static*
Primary *data*			*Traces [set m/z]*
Acquisition			***Sequential***
PRM	1 u	Quadrupole	HR/AM
	(default)	***Trapping*** *device*	
Primary data			*Full spectra*
Acquisition		***Multiplexing***	***Parallel***

12

HR/AM Targeted Proteomics



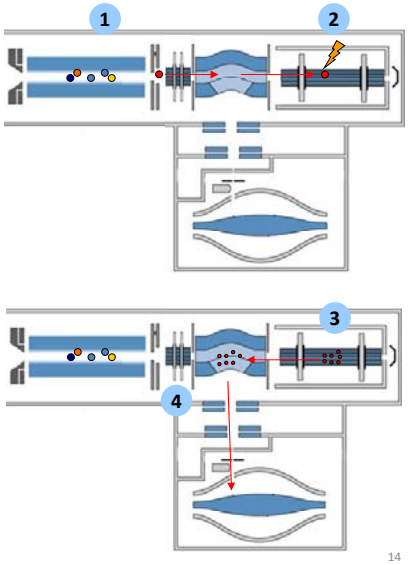
Gallien et al. Proteomics (2012), 11: 1709-1723

Quantification in PRM Mode: Principle

Targeted HCD Mode

- ***Quantification based on fragment ions***
- ***Acquisition method:***

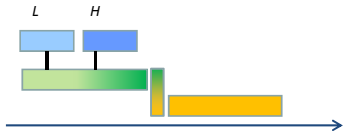
1. Selection precursor ion by quadrupole
2. Fragmentation in the collision cell (HCD)
3. Accumulation of fragment ions (HCD cell)
4. Transfer of fragment ions to the orbitrap (via C-trap).

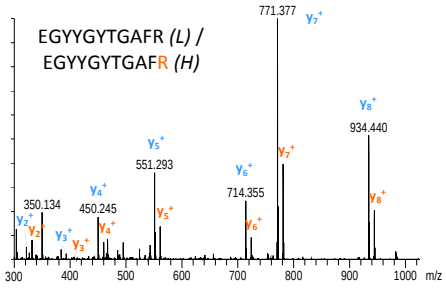


Gallien et al.; MCP (2012)

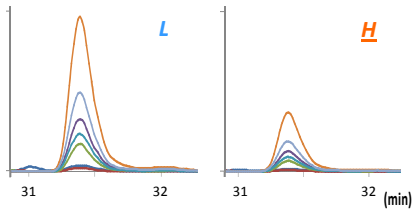
PRM Mode: Multiplexed Analysis

- Sequential isolation of L/H precursors
- Fragmentation and storage in HCD cell
- One orbitrap detection scan





- Quantification based fragment ions
- Selection of ions post-acquisition

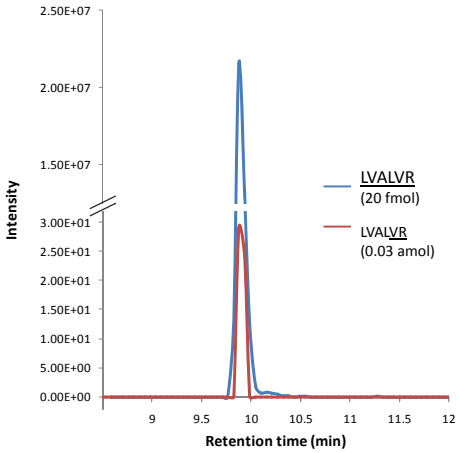


> *SRM-like quantification but with high resolution fragment ions*

Performance: Ultimate Sensitivity

- ***Sample***
  - Two isotopologous peptides:
    LVALVR and LVAL**VR**
  - Amounts : 20 fmol : 0.01 amol
    20 fmol : 0.03 amol
    20 fmol : 0.10 amol
    20 fmol : 0.30 amol
    20 fmol : 0.90 amol
- ***Acquisition method***
  - Resolution: 70,000
  - Maximum injection time: 3 s
  - Multiplexed SIM method

> ***Sensitivity : LOD of 30 zeptomol***
> *Intra-scan dynamic range > 4.5 logs*



Gallien et al.; MCP (2012)

Large-Scale Experiment – Critical Parameters

- **Chromatography:**
  - **Cycle time** (sampling across the elution profile: *8 – 10 data points*)
- **Mass spectrometer resolving power** *(orbitrap)*
  - **Resolution** ~ length of the transient
    e.g. 35,000 resolution ~ 128 ms transient
  - Number of peptides = Cycle time / Transient time
- **Ion Trapping**
  - Enrichment for low abundance precursor ions
  - Max fill time < Transient length

17

PRM Large-Scale Screening Experiment

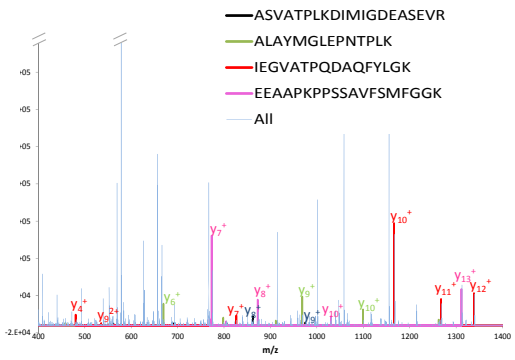
Parameters	***Multiplexed PRM analysis***						
Multiplexing: Nb precursors	**1**	1	2	**4**	4	8	**8**
Max injection time/pep [ms]	**< 60**	< 120	< 30	**< 30**	< 30	< 30	**< 15** ←
Resolution	**17 k**	35k	17 k	**35k**	35k	70k	**35k**
Transient time *[ms]*	**64**	128	64	**128**	128	256	**128**
Cycle time *[s]*	**2.0**	2.0	2.0	**2.0**	3.0	3.0	**3.0**
Isolation window/peptide *[Th]*	**2**	2	2	**2**	2	2	**1**
'Total' isolation window *[Th]*	**2**	2	4	**8**	8	16	**8**
Number peptides / cycle	**32**	16	64	**64**	92	92	**184** ←

Domon - 2013 ... 18

Multiplexed PRM Analysis (4-plex)

Method:

- Sequential isolation 4 precursor ions
- Fragmentation /storage in HCD cell
- One detection scan (orbitrap); Parallel detection of fragment ions.



Targeted Data Analysis:

> ***Verify identity***

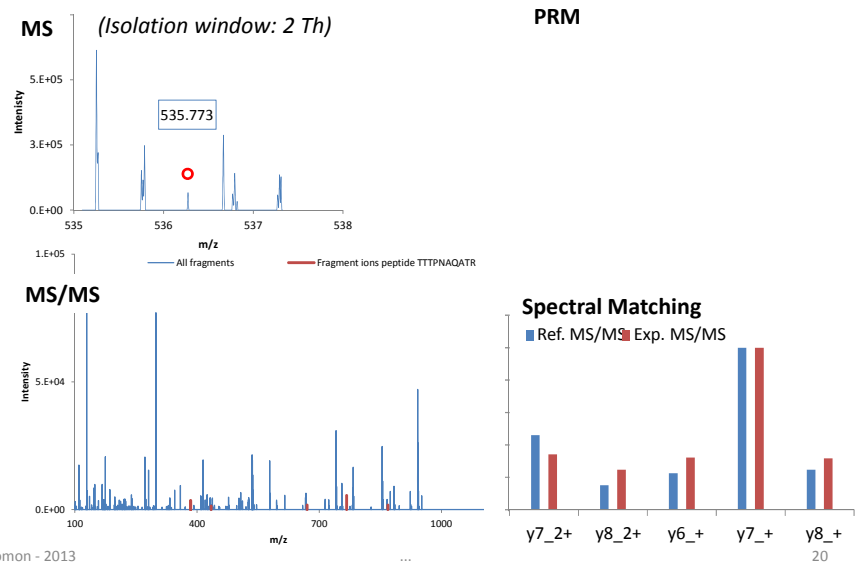
> **Quantify**

Results:

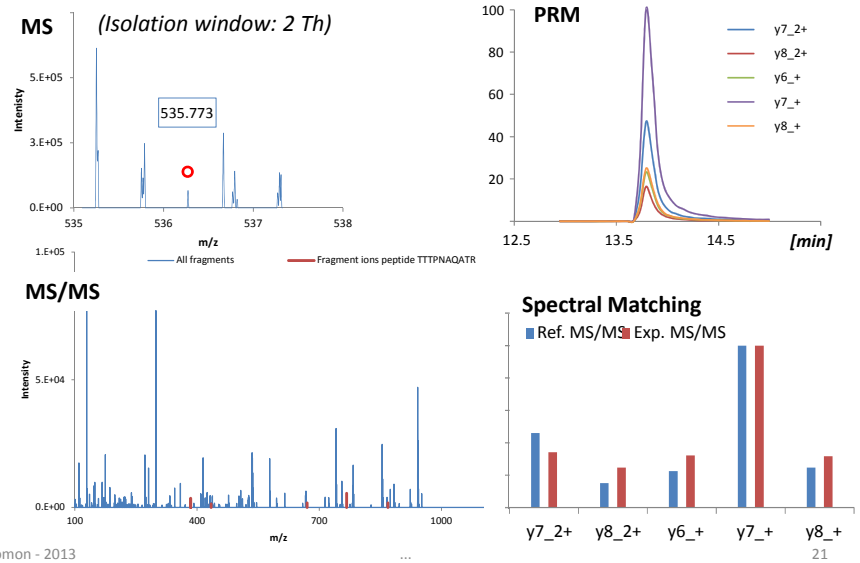
> **605** peptides (~80%) verified by **MS/MS spectra** (*p-value ≤ 0.1*)

> Consistent **quantification** results (*CV < 10 % for 95 % of peptides*)

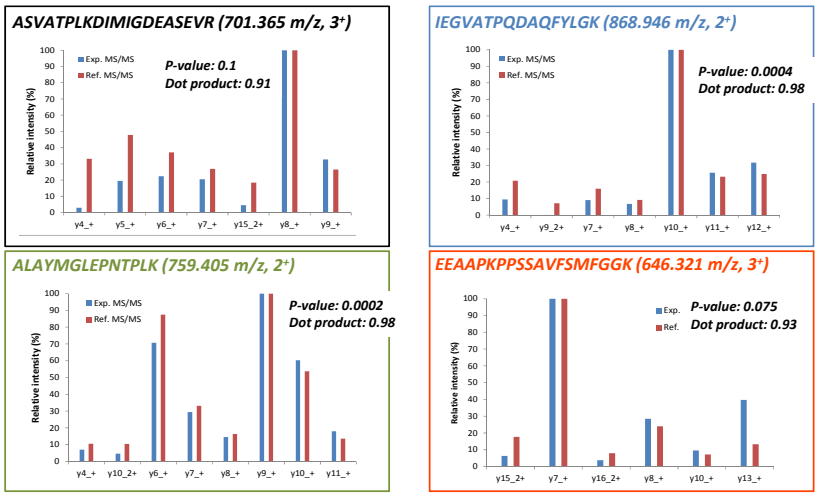
Example: Peptide *TTTPNAQATR*



Example: Peptide *TTTPNAQATR*



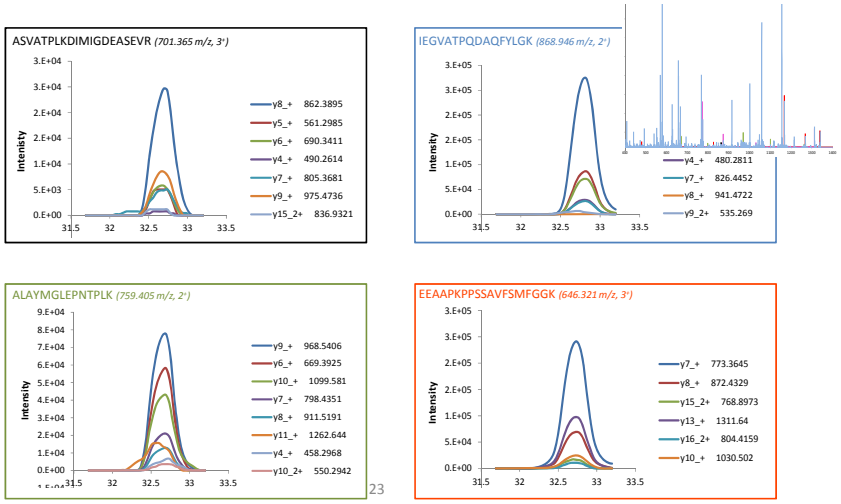
Multiplexed PRM Analysis *(4-plex)*



> Confirmation of the identity of the targeted peptides by spectral matching

22

Multiplexed PRM Analysis *(4-plex)*



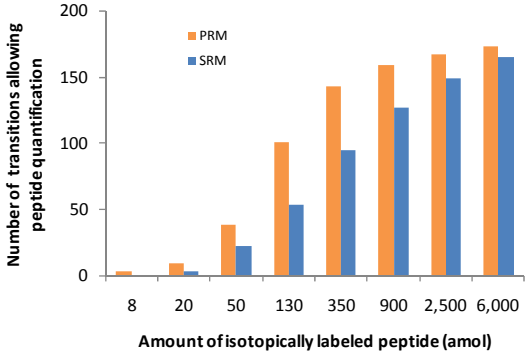
23

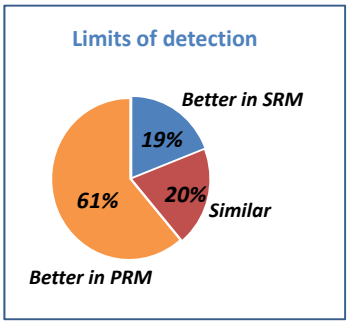
> *Extraction of the ion chromatograms of the 6-8 most intense ions for each peptide*

Gallien et al.; MCP (2012)

Comparison of SRM and PRM Performances

- *35 IS spiked in urine sample*
- *Triplicated analysis*
- *175 transitions were evaluated*

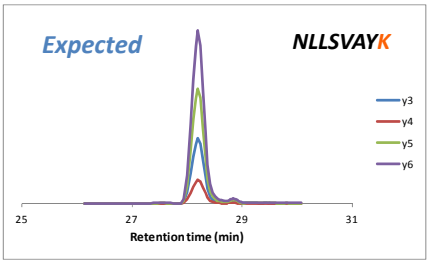




> High resolution of PRM analysis often provides more selectivity
> Extension of linearity range *(not systematic)*

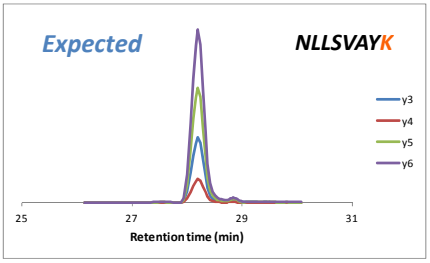
24

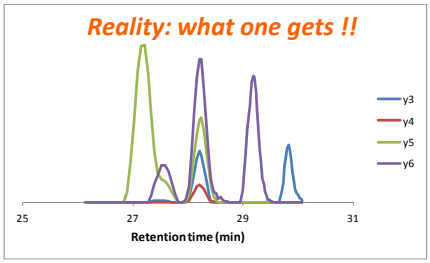
Selectivity Issue …..



Reality:

Selectivity Issue …..

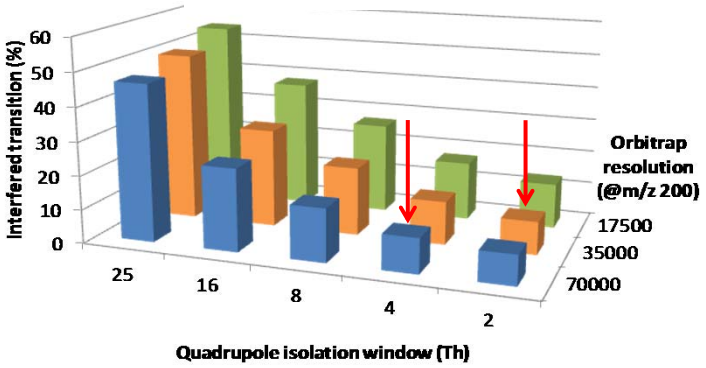




 Kim et al.; Proteomics – Clinical Applications (2012)

Selectivity of PRM Analyses

o *2172 fragment ions evaluated (122 peptide pairs: IS +endo)*



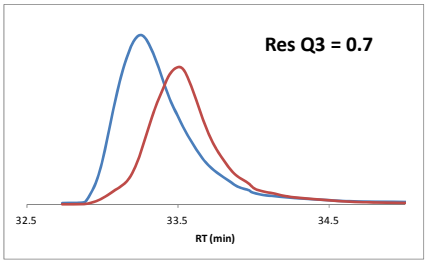
> *The selectivity of measurements is dramatically affected by the isolation window*
> *Limited benefit of increasing the nominal orbitrap resolution (17.5 k to 70k)*

Domon - 2013

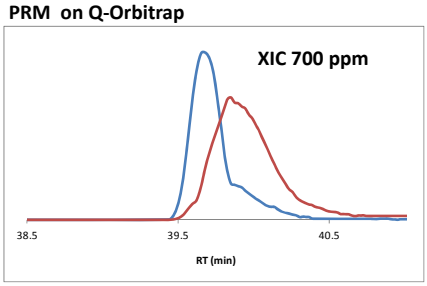
Gallien et al.; J. Proteomics (2012)

Selectivity in HR/AM Mode /1

SRM on Triple quadrupole

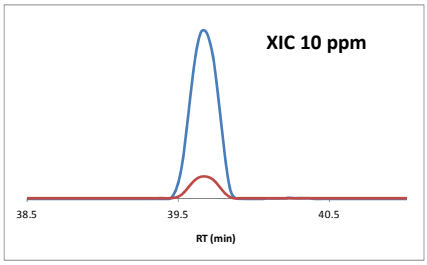


PRM on Q-Orbitrap



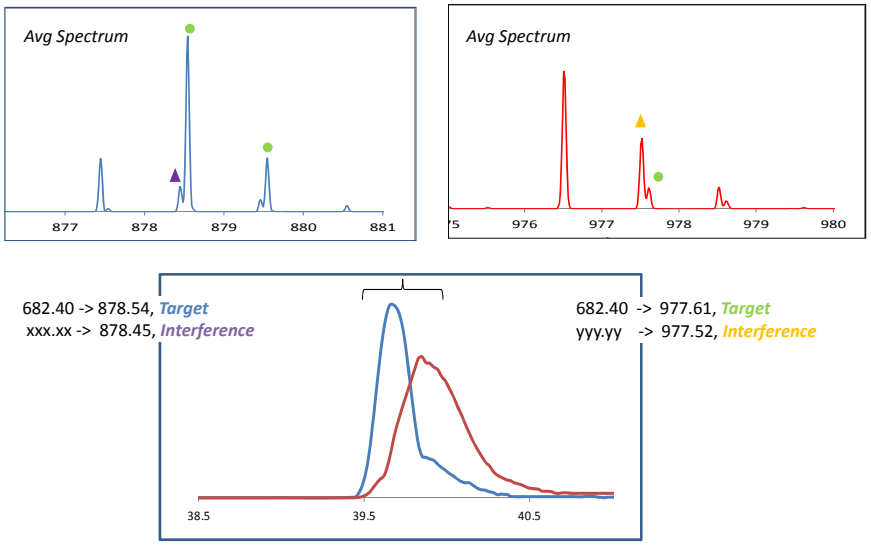
SDLAVPSELALLKY**K**
spiked in urine samples

Transitions : 682.40->878.54
682.40->977.61



Gallien et al.; J. Proteomics (2012)

Interferences



Requirements for Quantitative Analyses

- **Determination of LOD and LOQ**
  - In the matrix of interest *(multiple samples)*
- **Assessments of level of interferences**
  - Specify complexity of sample (ratio sample/analyte)
  - Window width of precursor mass selection
  - Resolving power of the second mass analyzer
- **Evidences to confirm the identity of the analyte**
  - Number of transitions measured
  - Accurate mass of the precursor and/or fragment ions
  - Fragmentation pattern (ion ratios or MS/MS fingerprint)
- **Proposed Quantification Scheme:**
  1. Confirm identity
  2. Assess signal-to-noise ratio to ensure signal > LOQ
  3. ***Quantify*** only, and only if **1** and **2** are fulfilled.

Domon - 2013 ... 30

Conclusions

- **Quantification in complex biological samples is challenging**
  - It requires high selectivity and sensitivity
- **HR/AM targeted methods as alternatives to *conventional* SRM**
  - Selected ion monitoring (SIM) mode, for precursor ions
  - Parallel reaction monitoring (PRM), for fragment ions
- **Figures of merit**
  - High sensitivity and selectivity (increased analytical precision)
  - Increased confidence in assignment (accurate mass)
  - Faster /simpler method development than SRM
  - Multiplexing capabilities enabling large-scale experiments
  - The data analysis is straightforward (quantification algorithms)
- **Hotspot**
  - Sample preparation is critical to control the background (LOD)

Acknowledgements

- Sebastien Gallien, *LCP*
- Adele Bourmaud, LCP
- Yeoun Jin Kim , *LCP*
- Nina Khristenko, *LCP*
- Jan van Oostrum, *LCP*
- Stephane Trevisiol, *LCP*
- Elodie Duriez , *LCP*
- Panchali Goswami, LCP
- Sang-Yoon Kim, LCP
- Kevin Demeure, LCP
- Guy Berchem, CHL - PPM
- Markus Kellmann, Thermo
- Thomas Moehring, Thermo
- Andreas Huhmer, Thermo
- Catharina Crone, Thermo

- **Funding:**
  - Fonds National de la Recherche, (FNR) G.-D. Luxembourg
  - Ministry of Research and Higher Education (MESR), G.-D. Luxembourg
  - ThermoFisher Scientific (Bremen and San Jose)

Targeted Peptide Measurements in Biology and Medicine: Best Practices for Assay Development

Applying SRM for the quantification of proteins in basic biological/clinical research

Ruth Hüttenhain, PhD

Aebersold lab

NIH workshop, June 17-18, Bethesda, MD

GOALS OF THE TARGETED MS EXPERIMENTS

Quantification of a defined set of proteins (~ 20 - 100) in complex samples across different biological/clinical conditions

Requirements for protein quantification in basic research:

- Sensitivity
- Consistency
- Multiplexed quantification of 20 – 100 proteins
- Relative quantification
- Accuracy & reproducibility

But…

⇒ Requirements for basic applications are less stringent than for clinical assays (current state-of-the art is western blot)

SRM Measurements for Basic Research Applications

Step 1: Generation of SRM assays for target proteins

Step 2: Testing the detectability of proteins in the target sample

Step 3: Preparing the final quantification method

Step 4: Data analysis of large-scale SRM datasets

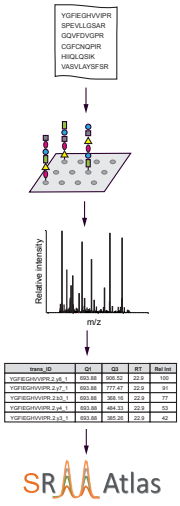
Example applications:

Response of *M. tuberculosis* to hypoxia in a time-course experiment

Quantification of cancer-associated proteins in a case-control study

Schubert *et al.* The Mtb Proteome Library: A Resource of Assays to Quantify the Complete Proteome of Mycobacterium tuberculosis. *Cell Host Microbe* (2013)
Hüttenhain *et al.* Reproducible Quantification of Cancer-Associated Proteins in Body Fluids Using Targeted Proteomics. *Sci Transl Med* (2012)

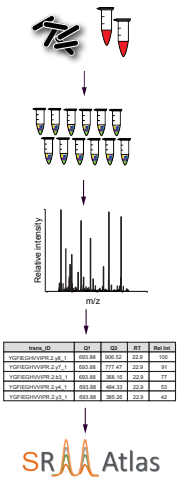
Step 1: SRM Assay Generation I



1. Selection of proteotypic peptides (up to 5 peptides):
   - 6 – 21 amino acids (avoiding Cys, Try and Met if possible)
   - Uniqueness
   - Peptides previously observed (PeptideAtlas) or predicted (hydrophobicity/length)
2. Peptide synthesis
3. SRM triggered MS2 of synthetic peptide libraries
4. Database search for peptide sequence identification (FDR < 1%)
   - Confident assignment of peptide sequence to fragment ion spectra
5. Extraction of SRM assay coordinates
   - Up to 8 transitions (based on intensity), RT, rel. intensity
6. SRM assay storage and dissemination (www.srmatlas.org)

Picotti *et al.* High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. *Nat Methods* (2010)

STEP 1: SRM ASSAY GENERATION II



1. Protein/peptide extraction from target samples
2. Sample fractionation (Off-gel electrophoresis or SCX)
3. MS analysis (Triple TOF or Orbitrap HCD)
4. Database search for peptide sequence identification (FDR < 1%)
   - Confident assignment of peptide sequence to fragment ion spectra
5. Extraction of SRM assay coordinates
   - Up to 8 transitions (based on intensity), RT, rel. intensity
6. SRM assay storage and dissemination (www.srmatlas.org)

Schubert *et al.* The Mtb Proteome Library: A Resource of Assays to Quantify the Complete Proteome of Mycobacterium tuberculosis. *Cell Host Microbe* (2013)
Picotti *et al.* A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis. *Nature* (2013)

STEP 1: SRM ASSAY LIBRARIES IN SRMATLAS

M. tuberculosis

- 15,679 peptides
- 3,894 proteins (97% of the annotated Mtb proteins)

S. cerevisiae

- 28,216 peptides
- ~6,400 proteins (97% of the yeast proteome)

Cancer-associated proteins

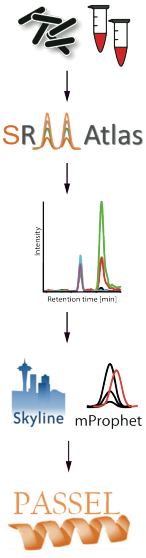
- 3,996 peptides
- 1,157 proteins

Human *N*-glycoproteins

- 4,421 *N*-glycosites
- 2,007 *N*-glycoproteins

⇒ **Libraries comprising SRM assay coordinates extracted from crude synthetic peptide measurements** (without validation in complex sample background)

Step 2: Detectability Test of SRM Assays in Target Sample

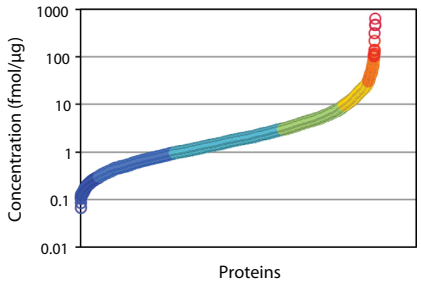


1. Protein/peptide extraction from target samples
2. Extraction of SRM assay coordinates from SRMAtlas
   - Up to 3 peptides/protein
   - 5 transitions/peptide
3. SRM measurements of target proteins in target sample without internal standard
   - Inclusion of decoy transitions for data evaluation
4. Data evaluation
   - Visualization in Skyline
   - mProphet analysis to match SRM data to initial assay coordinates (FDR < 1 – 2% for confident identification)
5. Data storage in PASSEL supporting future experimental designs

Step 2: Detectable Concentration Range

M. tuberculosis

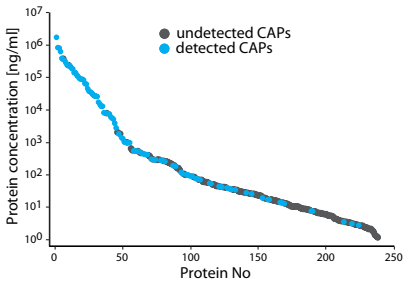
Detection of 2,884 proteins across 4 orders of magnitude in unfractionated Mtb digest



Absolute label-free abundance estimates

Cancer-associated proteins

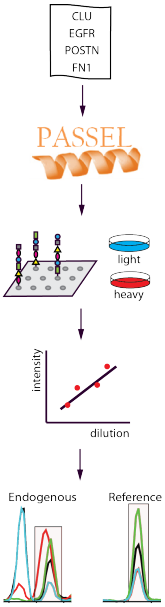
Detection of 162 proteins across 5 orders of magnitude in depleted plasma



Estimated abundances from PeptideAtlas

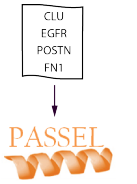
Schubert *et al.* The Mtb Proteome Library: A Resource of Assays to Quantify the Complete Proteome of Mycobacterium tuberculosis. *Cell Host Microbe* (2013)
Hüttenhain *et al.* Reproducible Quantification of Cancer-Associated Proteins in Body Fluids Using Targeted Proteomics. *Sci Transl Med* (2012)

Step 3: Preparing Final Quantification Method

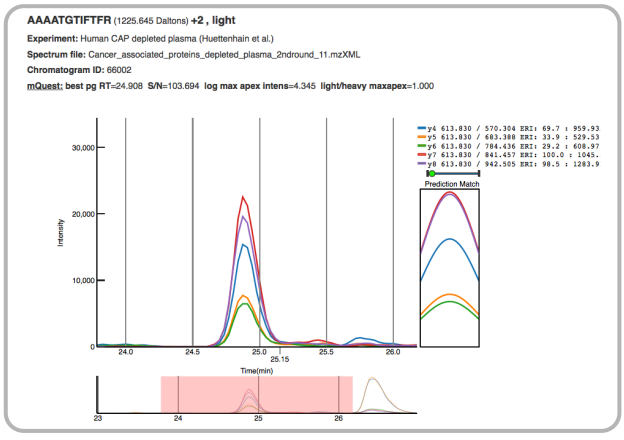


1. Defining target proteins
2. Selection of up to 3 peptides/protein and 5 transitions/peptide (PASSEL, SRMAtlas)
3. Synthesis of crude heavy isotope labeled reference peptides or metabolically labeled sample as internal standard for relative quantification
4. Adjustment of internal standard to the endogenous peptide amount (dilution series measured in the background of the sample), but no CV, LOQ and LOD
5. Selection of best 3 transitions for quantification method (no interferences)

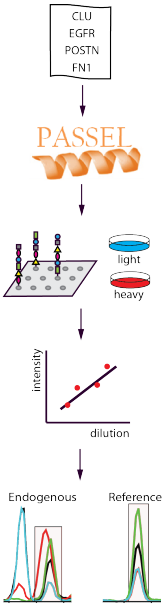
Step 3: Preparing Final Quantification Method



1. Defining target proteins
2. Selection of up to 3 peptides/protein and 5 transitions/peptide (PASSEL, SRMAtlas)

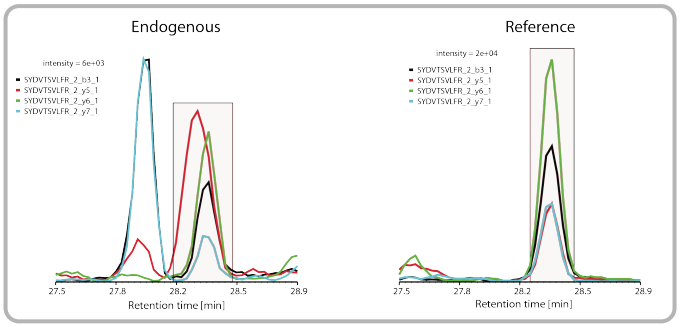


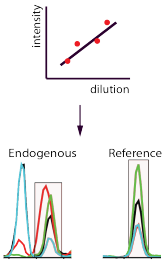
Step 3: Preparing Final Quantification Method



1. Defining target proteins
2. Selection of up to 3 peptides/protein and 5 transitions/peptide (PASSEL, SRMAtlas)
3. Synthesis of crude heavy isotope labeled reference peptides or metabolically labeled sample as internal standard for relative quantification
4. Adjustment of internal standard to the endogenous peptide amount (dilution series measured in the background of the sample), but no CV, LOQ and LOD
5. Selection of best 3 transitions for quantification method (no interferences)

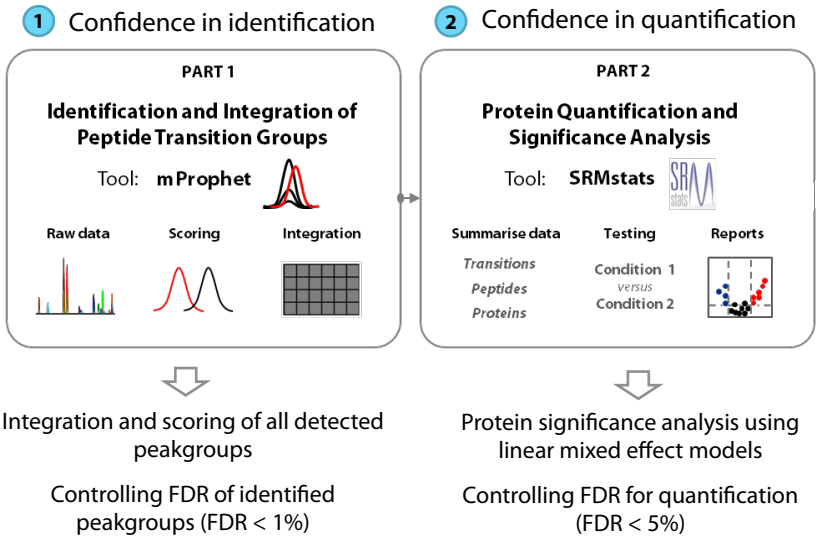
Step 3: Preparing Final Quantification Method





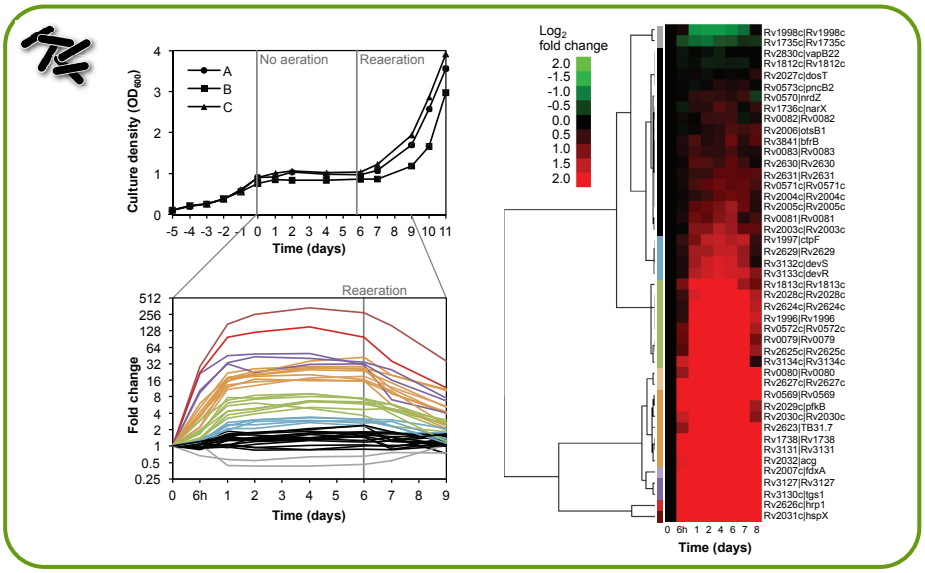
4. Adjustment of internal standard to the endogenous peptide amount (dilution series measured in the background of the sample), but no CV, LOQ and LOD
5. Selection of best 3 transitions for quantification method (no interferences)

Step 4: SRM Data Analysis



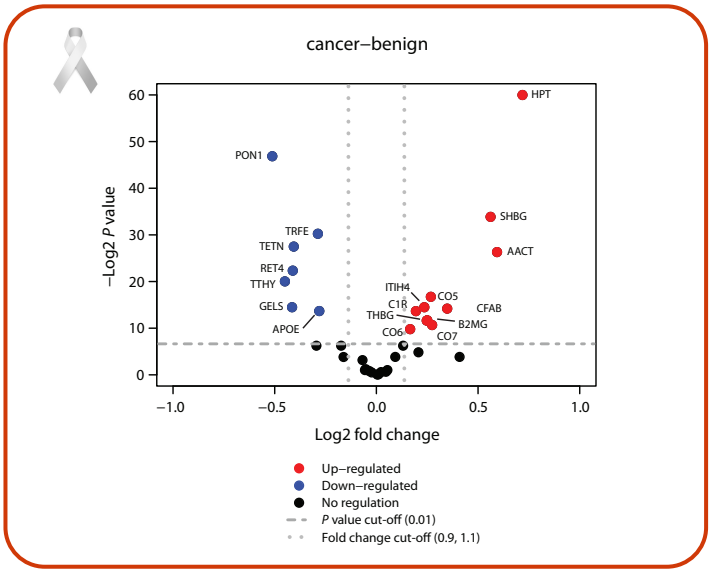
Surinova *et al.* Automated SRM data analysis workflow for large scale targeted proteomic studies. *Nat Protocols, in press*

Step 4: Time Course in MTB under Hypoxic Stress



Schubert *et al.* The Mtb Proteome Library: A Resource of Assays to Quantify the Complete Proteome of Mycobacterium tuberculosis. *Cell Host Microbe* (2013)

STEP 4: CASE-CONTROL STUDY FOR OVARIAN CANCER



Hüttenhain *et al.* Reproducible Quantification of Cancer-Associated Proteins in Body Fluids Using Targeted Proteomics. *Sci Transl Med* (2012)

QUESTIONS/ANSWERS I

1. Goal of the targeted MS experiment and plex-level
   - Consistent, quantitative measurements across large number of samples
   - Relative quantification of 20-100 proteins in a tryptic digest of complex biological/clinical samples (total cell lysates, tissue lysates, plasma) for basic research questions
2. ID confidence
   - SRM assay coordinates derived from confidently identified spectra
   - mQuest/mProphet analysis to match to assay coordinates and internal standard (if available)
   - FDR cutoff < 1% for confident identification
3. Quantification
   - Up to 3 peptides/protein (if possible) and 3 transitions/peptide
   - Crude heavy labeled peptides used for relative quantification

Questions/Answers II

4. Interference detection
    - 5 transitions/peptide initially tested in the sample of interest to select the transitions without interferences
    - Detection of the interferences using mProphet scoring and/or manual inspection in Skyline
5. Protein quantification based on measured peptide levels
    - Transition/peptide level measurements are combined into a protein measurements using statistical models implemented in SRMstats
6. Software and analytical tools
    - SRMAtlas: Dissemination and extraction of SRM assays
    - PASSEL: Dissemination and extraction of experimental SRM data
    - Skyline: Visualization of SRM data
    - mQuest/mProphet: ID confidence
    - SRMstats: statistical analysis/confidence in quantification

Information Required For Publication

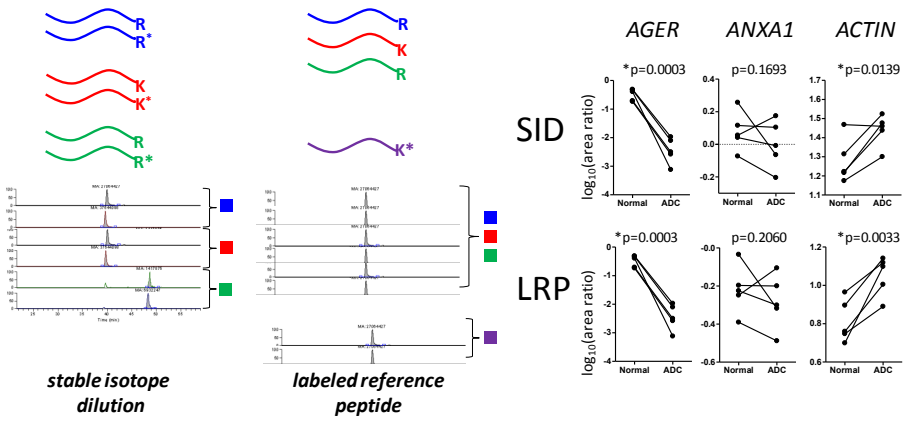
1. Assay generation
    - Description of the assay generation process
    - Selection criteria and number of peptides and transitions
    - Confidence measure for assay generation
    - Transition list including CE and RT/iRT
2. Quantification
    - Data evaluation
    - Confidence measure for identification/quantification
3. Significance Analysis
    - Statistical test
    - p-value/fold-change cutoff
4. Public availability/dissemination of data
    - SRMAtlas for SRM assays
    - PASSEL for SRM dataset
    - Skyline file

Tier 3 Assays: Labeled Reference Peptide Method

Daniel C. Liebler
Lisa J. Zimmerman

Vanderbilt University School of Medicine

Labeled reference peptide method



Zhang et al. (2011) Mol. Cellular Proteomics 10: M110.006593

Q. What are the goals of our targeted MS experiments or software?

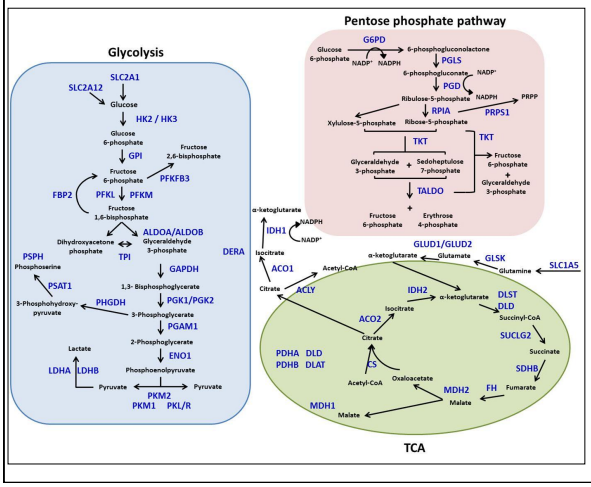
- Comparative quantitation between specimens and experiments
- Verify measurements from shotgun datasets
- Preliminary screening of biomarker candidates
- Pathway and system profiling

Q. Who are the "customers" or likely users of the methods and results?

- Basic and translational research laboratories

Metabolism Assay Panel

- 58 proteins in glycolysis, TCA and pentose phosphate pathways
- 216 peptides (4 transitions monitored per peptide)= 864 total transitions

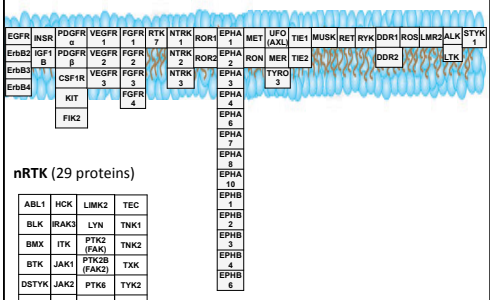


Protein Name	Gene Symbol
ATP-citrate synthase	ACLY
aconitate hydratase, cytopasmic	ACO1
Fructose-bisphosphate aldolase A	ALDOA
Fructose-bisphosphate aldolase B	ALDOB
Citrate synthase, mitochondrial	CS
Putative deoxyribose-phosphate aldolase	DERA
Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydroge	DLAT
Dihydrolipoyl dehydrogenase, mitochondrial	DLD
Dihydrolipoyl dehydrogenase	DLD
Alpha-ketoglutarate dehydrogenase complex dihydrolipoyl succinyltransferase	DLST
Alpha-enolase	ENO1
Fructose-1,6-bisphosphatase 1	FBP1
Fructose-1,6-bisphosphatase isozyme 2	FBP2
Isoform Long of Glucose-6-phosphate 1-dehydrogenase	G6PD
Glyceraldehyde-3-phosphate dehydrogenase	GAPDH
glutaminase kidney form	GLSK
Glutamate dehydrogenase 1, mitochondrial	GLUD1
Glutamate dehydrogenase 2, mitochondrial	GLUD2
Hexokinase-2	HK2
Hexokinase-3	HK3
Isocitrate dehydrogenase [NADP] cytoplasmic	IDH1
Isocitrate dehydrogenase [NADP]	IDH2
Isoform 1 of Isocitrate dehydrogenase [NAD] subunit alpha, mitochondrial	IDH3A
Isoform 1 of Protein tyrosine kinase 2 beta	IDH3A
Isoform 1 of Isocitrate dehydrogenase [NAD] subunit alpha,	IDH3A
Isocitrate dehydrogenase [NAD] subunit gamma,	IDH3G
L-lactate dehydrogenase A	LDHA
L-lactate dehydrogenase B	LDHB
Malate dehydrogenase	MDH1
Malate dehydrogenase, mitochondrial	MDH2
Isoform 1 of Pyruvate dehydrogenase E1 component subunit beta	PDHB
6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3	PFKFB3
6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 4	PFKFB4
Phosphofructokinase	PFKM
Phosphoglycerate mutase 1	PGAM1
6-phosphogluconate dehydrogenase, decarboxylating	PGD
Phosphoglycerate kinase 1	PGK1
Phosphoglycerate kinase 2	PGK2
6-phosphogluconolactonase	PGLS
phosphoglycerate dehydrogenase	PHGDH
Isoform R-type of Pyruvate kinase isozymes R/L	PKLR
Pyruvate kinase	PKM2
Ribose-phosphate pyrophosphokinase 1	PRPS1
phosphoserine aminotransferase 1	PSAT1
phosphoserine phosphatase	PSPH
Ribose-5-phosphate isomerase	RPIA
Succinate dehydrogenase [ubiquinone] iron-sulfur subunit,	SDHB
neutral amino acid transporter B	SLC1A5
Solute carrier family 2, facilitated glucose transporter member 1	SLC2A1
Solute carrier family 2, facilitated glucose transporter member 1	SLC2A1
Solute carrier family 2, facilitated glucose transporter member 12	SLC2A12
Succinyl-CoA ligase [GDP-forming] subunit beta,	SUCLG2
Transaldolase	TALDO1
Transketolase	TKT
triosephosphate isomerase 1 isoform 1	TPI1
triosephosphate isomerase 1 isoform 2	TPI2
fumarate hydratase	FH
glucose-6-phosphate isomerase	GPI
Isoform 1 of Pyruvate dehydrogenase E1 component subunit beta	PDHB
Pyruvate kinase	PKM1
isocitrate dehydrogenase 3 (NAD+) alpha	IDH3A
isocitrate dehydrogenase 3 (NAD+) beta	IDH3B
isocitrate dehydrogenase 3 (NAD+) gamma	IDH3G
phosphofructokinase, liver	PFKL

Receptor tyrosine kinase/non-receptor tyrosine kinase panel

- PRM assays on Q-Exactive
- 83 proteins in RTK/nRTK pathways
- 310 peptides (4 transitions monitored per peptide) =1240 total transitions

RTK (20 subfamilies; 54 proteins)



nRTK (29 proteins)

ABL1	HCK	LIMK2	TEC
BLK	IRAK3	LYN	TNK1
BMX	ITK	PTK2 (FAK)	TNK2
BTK	JAK1	PTK2B (FAK2)	TXK
DSTYK	JAK2	PTK6	TYK2
FER	JAK3	SRC	YES
FES	LCK	SYK	ZAP70
FYN			

Protein Description	Gene Name	Protein Description	Gene Name
Epidermal growth factor receptor	EGFR	Ephrin type-A receptor 3	EPHA3
Receptor tyrine-protein kinase erbB-2	ERBB2	Ephrin type-A receptor 4	EPHA4
Receptor tyrine-protein kinase erbB-3	ERBB3	Ephrin type-A receptor 6	EPHA6
Receptor tyrine-protein kinase erbB-4	ERBB4	Ephrin type-A receptor 7	EPHA7
Fibroblast growth factor receptor 2	FGFR2	Ephrin type-B receptor 1	EPHB1
Fibroblast growth factor receptor 4	FGFR4	Ephrin type-B receptor 2	EPHB2
insulin receptor	INSR	Ephrin type-B receptor 3	EPHB3
Insulin-like growth factor 1 receptor	IGF1R	Ephrin type-B receptor 4	EPHB4
Tyrine-protein kinase receptor UFO	AXL	Ephrin type-B receptor 6	EPHB6
Tyrine-protein kinase Mer	MERTK	Vascular endothelial growth factor receptor 1	FLT1
Hepatocyte growth factor receptor	MET	Vascular endothelial growth factor receptor 2	KDR
FL cytokine receptor	FLT3	Vascular endothelial growth factor receptor 3	FLT4
Alpha-type platelet-derived growth factor receptor	PDGFRA	Tyrine-protein kinase receptor Tie-1	TIE1
Beta-type platelet-derived growth factor receptor	PDGFRB	Angiopoietin-1 receptor	TEK
Macrophage-stimulating protein receptor	MST1R	Tyrine-protein kinase transmembrane receptor ROI	ROR1
Mast/stem cell growth factor receptor	KIT	Proto-oncogene tyrine-protein kinase receptor Ret	RET
Proto-oncogene tyrine-protein kinase Src	SRC	High affinity nerve growth factor receptor	NTRK1
Tyrine-protein kinase Lyn	LYN	BDNF/NT-3 growth factors receptor	NTRK2
Ephrin type-A receptor 1	EPHA1	NT-3 growth factor receptor	NTRK3
Ephrin type-A receptor 2	EPHA2	Tyrine-protein kinase receptor TYRO3	TYRO3
Muscle, skeletal receptor tyrine-protein kinase	MUSK	Tyrine-protein kinase JAK1	JAK1
Fibroblast growth factor receptor 1	FGFR1	Tyrine-protein kinase JAK2	JAK2
Fibroblast growth factor receptor 3	FGFR3	Tyrine-protein kinase JAK3	JAK3
Macrophage colony-stimulating factor 1 receptor	CSF1R	Tyrine-protein kinase SYK	SYK
Ephrin type-A receptor 8	EPHA8	Tyrine-protein kinase ZAP-70	ZAP70
Ephrin type-A receptor 10	EPHA10	Tyrine-protein kinase Fer	FER
Tyrine-protein kinase transmembrane receptor RC	ROR2	Tyrine-protein kinase BTK	BTK
Epithelial discoidin domain-containing receptor 1	DDR1	Interleukin-1 receptor-associated kinase 3	IRAK3
Discoidin domain-containing receptor 2	DDR2	Serine/threonine-protein kinase LMTK2	LMTK2
Leukocyte tyrine kinase receptor	LTK	Tyrine-protein kinase Tec	TEC
ALK tyrine kinase receptor	ALK	Tyrine-protein kinase Blk	BLK
Inactive tyrine-protein kinase 7	PTK7	Tyrine-protein kinase TXK	TXK
Tyrine-protein kinase Fyn	FYN	Dual serine/threonine and tyrine protein kinase	DSTYK
Tyrine-protein kinase Yes	YES1	LIM domain kinase 2	LIMK2
Tyrine-protein kinase Lck	LCK	Tyrine-protein kinase RYK	RYK
Tyrine-protein kinase HCK	HCK	Proto-oncogene tyrine-protein kinase R	R1
Tyrine-protein kinase ITK/TSK	ITK	Tyrine-protein kinase STYK1	STYK1
Focal adhesion kinase 1	PTK2	Tyrine-protein kinase Fes/Fps	FES
Protein-tyrine kinase 2-beta	PTK2B	Epidermal growth factor receptor	EGFR
Protein-tyrine kinase 6	PTK6	Hepatocyte growth factor receptor	MET
Non-receptor tyrine-protein kinase TYK2	TYK2	Mast/stem cell growth factor receptor	KIT
Non-receptor tyrine-protein kinase TNK1	TNK1	Proto-oncogene tyrine-protein kinase Src	SRC
Activated CDC42 kinase 1	TNK2	Ephrin type-B receptor 2	EPHB2
Tyrine-protein kinase ABL1	ABL1	Ephrin type-B receptor 3	EPHB3
Cytoplasmic tyrine-protein kinase BMX	BMX	Protein-tyrine kinase 2-beta	PTK2B

Q. What range of analyte plex-level per-injection do you typically use in your experiments?

Plex-Level	Proteins	Method	Example
Low	< 5	Unscheduled	SID methods (i.e. KRAS, BRAF mutants)
Medium	10-15	Unscheduled	LRP approach
High	> 15	Scheduled	Thematic Panels (metabolism, RTK/nRTK, etc) with LRP approach

Q.What is the impact of the plex-level used on the robustness and figures of merit (CV, LOD, LOQ) of the developed MRM/SRM assays?

- CV independent of plex level under above scenarios
- LOD/LOQ not determined

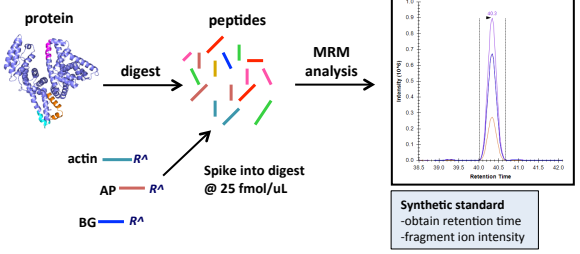
Q.Explain how you establish confidence that what is being measured is the analyte of interest.

- Synthetic peptides are obtained for all analytes
  - New England Peptide FLASHPURE™ Peptide Array; 96-well plate format; 2-3 mg lyophilized material per well; no AAA; moderate purity (50-85%); MALDI-MS QC analysis for each peptide; $3,300 per plate
- Preliminary analyses in our laboratory
  - master mix prepared; analyzed on MRM/SRM/PRM instrument to optimize transitions and obtain retention time
  - Standards enable exclusion of poor performers; development of scheduled methods

Q. How do these methods differ from "Discovery Proteomics" using data-dependent or data-independent experiments?

- analyses are targeted to specific analytes (*m/z* targets)

Q. Explain your method of quantification, how many transitions you monitor and which ones are chosen to quantify.



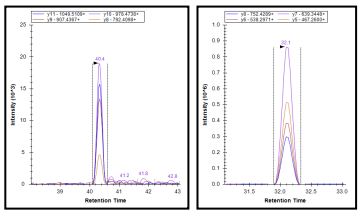
Synthetic standard
-obtain retention time
-fragment ion intensity

Assay development
- choose 4-5 transitions/peptide
- choose 3-4 peptides/ protein
- retention time for method

LRP peptide standards
Human β-actin GYSFTTTAER*
E. coli alkaline phosphatase AAQGDITAPGGAR*
E. coli β-galactosidase APLDNDIGVSEATR*

$$\text{Normalized Peak Area} = \frac{\Sigma \text{ peak area of all measured transitions per peptide}}{\Sigma \text{ peak area of all measured transitions of LRP}}$$

Protein "value" is sum of normalized peptide peak areas



Endogenous peptide
assay target

LRP peptide
reference/normalization standard

Q. If you generate standard curves (calibration or response curve), explain how you use them to assess the quantitative accuracy of the assay (e.g., are the slope and y-intercept from the curve regression used in the calculating the analyte concentration in the sample? Is an external calibration curve used?

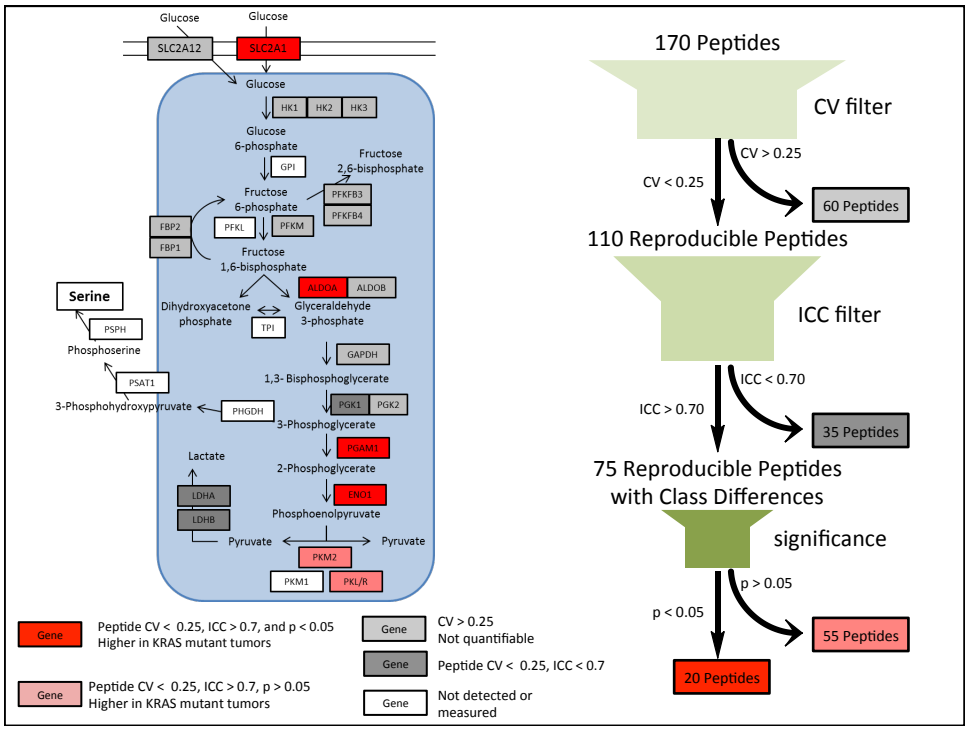
- ***Not applicable***

Q. Can you provide a useful estimate or accurately determine the amount of protein in the matrix based on the measured levels of peptides? Explain how/why. Indicate experimental parameters such as number of peptides per protein and the criteria/computational tools applied. If you have multiple peptides from the same protein and each gives a different answer for the extrapolated protein level, how do you deal with this?

- ***Not applicable***

Q. Describe methods you use to establish presence of interferences and how you deal with them if detected.

- Synthetic peptide standards provide transition intensities under run conditions.
- Compare detected fragment ion patterns with those from authentic standard.
- Skyline and visual inspection of the fragment ion patterns
- Interferences in a specific transition lead to selction of an alternative fragment ion.
- AuDIT – developed at Broad Institute) Automated Detection of Inaccurate and imprecise Transitions (AuDIT) algorithm. Assists in detecting interferences by comparing the relative product ion intensities of the analyte peptide to those of the SIS.
- Typically spike three LRP standards; if there is an interference problem with one, the others can be used.



Q. How do you account for suppression of ionization in your quantification method?

- Ionization suppression affecting whole runs or large segments may be evident from LRP peptide signal
- Transient suppression effects not detected or corrected in LRP method

Q. What software and analytical tools do you use in your studies and why?

- Skyline
- QuaSAR
- AuDIT
- Prototype data analysis package in "R" for CV and ICC determination

Q. What information do authors need to provide in their manuscripts/supplement to enable reviewers and readers to understand what was done and to be able to judge the confidence of the measurements made?

- List target peptides and transitions
- Peptide standards vs. predicted transitions
- CV for all peptide measurements
- CV for LRP peptides
- ICC for peptides across sample groups

Acknowledgements

- Lisa Zimmerman
- Jaime Hutton
- Ming Li
- Rob Slebos
- De Lin
- Hye-Jung Kim
- Haixia Zhang
- Mike MacCoss
- Brendan MacLean
- U24CA59988 Clinical Proteomic Tumor Analysis Consortium

Statistical design and analysis of targeted quantitative proteomic experiments

Olga Vitek

Statistics and computer science
Purdue University

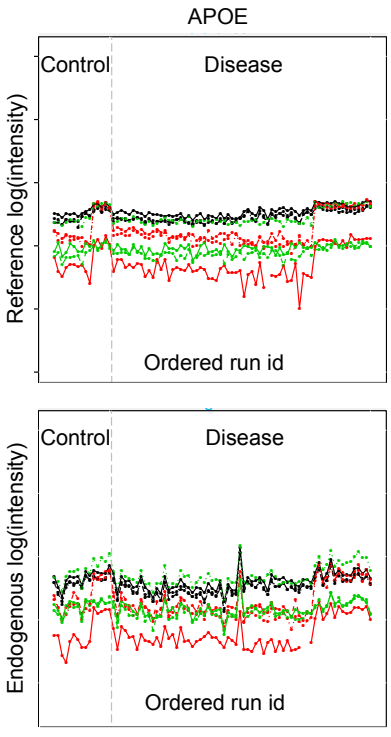
Special thank you:
Veavi Chang and Meena Choi, Purdue University

Customers and goals

- Customers
  - All users of targeted MS
    - biology & biomarker discovery
  - Not necessarily statisticians
- Biotechnological goal
  - Find differentially abundant proteins
- Statistical goals
  - Design: minimize bias & maximize efficiency
  - Analysis: effectively summarize all data
- Overall goal
  - Maximize reproducibility

Label-based SRM study of ovarian cancer

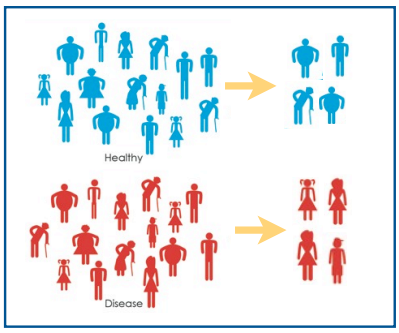
R. Hüttenhain et al.



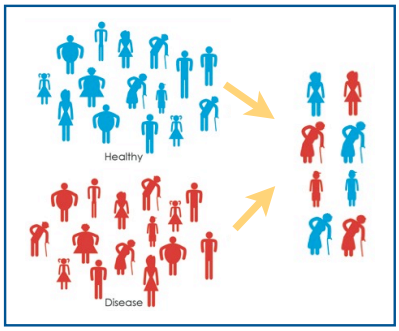
Impact of multiplexing

(Study design)

- Impact of multiplexing
  - − Multiple comparisons
  - − Interferences and larger variation
  - \+ Learn variance from all proteins
- More important: assay design
  - Consistent protocol
  - Randomization & blocking
- Even more important: study design
  - *Subject selection*
    - Defined populations
    - Matching for confounding factors
    - Defined scope of conclusions
  - *Protein selection*
    - Biological variation
    - Expected changes



Complete randomization
= inflated variance



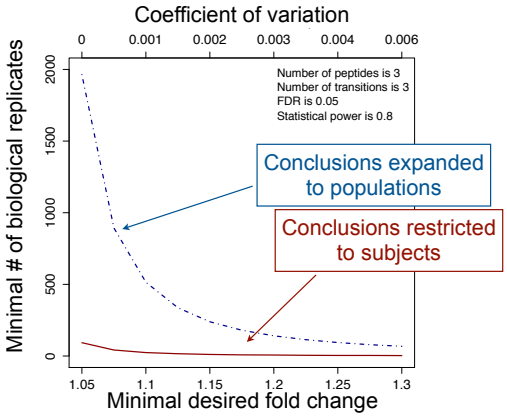
Matching
= restricted randomization

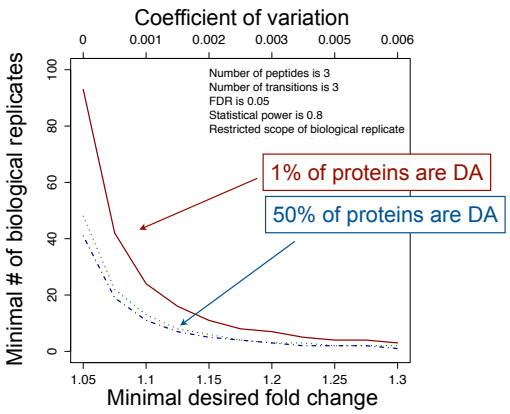
Käll and Vitek, *PLoS Computational Biology*, 2011

Impact of multiplexing

(Study design)

- Impact of multiplexing
  - − Multiple comparisons
  - − Interferences and larger variation
  - \+ Learn variance from all proteins
- More important: assay design
  - Consistent protocol
  - Randomization & blocking
- Even more important: study design
  - *Subject selection*
    - Defined populations
    - Matching for confounding factors
    - Defined scope of conclusions
  - *Protein selection*
    - Biological variation
    - Expected changes

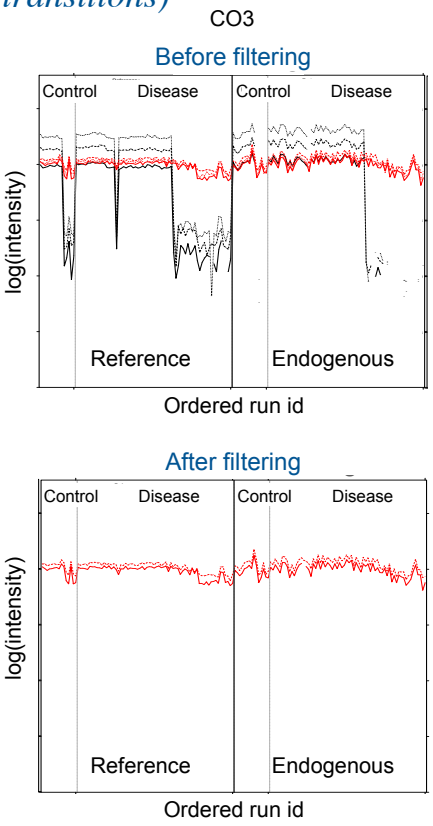




Confidence in identification and quantification

(1: Identify and quantify the right transitions)

- Current practice:
  - *Consistency of*
    - Retention times of endogenous fragments
    - Relative intensities of fragments
    - Co-elution with reference peptides
- Proposed, in addition:
  - *Interferences propagate into between-run variation of intensities*
    - Weight each transition by the variation
    - Extra weight if needed (e.g. mProphet)
    - Set small weights to 0



	Sensitivity N = 15, 66	Specificity N = 7, 8
All	**0.71**	**0.53**
Proposed	**0.88**	**0.75**

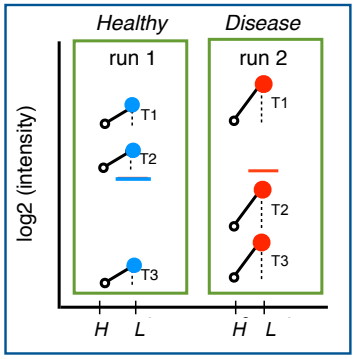
Sensitivity: *literature*
Specificity: *controls vs controls*

Chang et al., *ASMS* 2013

Confidence in identification and quantification

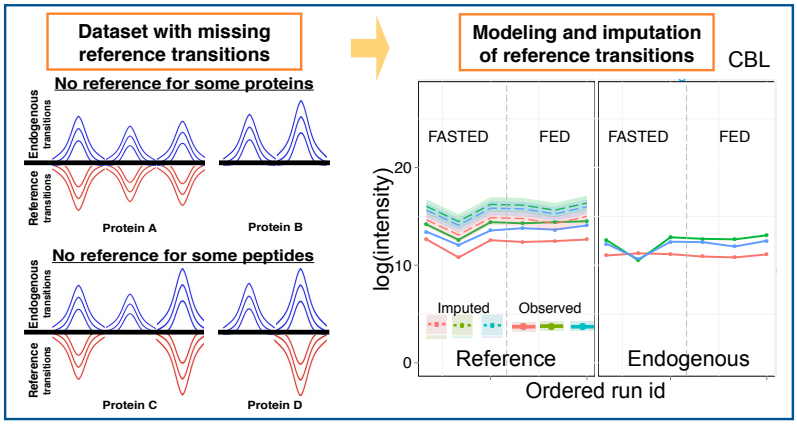
(2: Enhance the accuracy with internal standards)

- Current practice:
  - Internal standards enhance our ability to detect DA proteins
- Proposed:
  - The full collection of reference transitions is unnecessary
  - Can predict intensities of missing reference transitions from the observed
    - Particularly helpful in large-scale confirmatory studies



Chang et al., *MCP* 2012
Chang et al., *ASMS* 2012

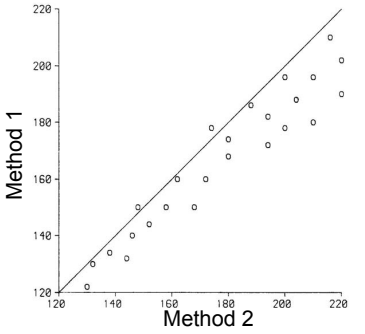
Label-based SRM study of mouse metabolism E. Sabidó et al.

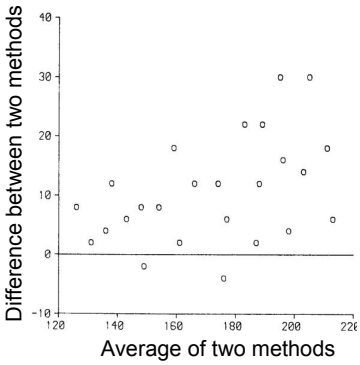


Accuracy assessment with standard curves

(Agreement with known concentrations, or with other methods)

- Current practice
  - Correlation & regression are flawed
  - Measure association, not agreement
    - arbitrary high correlations and R^2 by increasing the spread on X axis
    - more noise leads to better apparent agreement with the 45° line
- Better metrics:
  - Bias and variance wrt known concentrations
  - Paired differences wrt alternative methods
- LOD and LOQ are not always needed
  - Assays with an acceptable dynamic range can focus on changes in abundance
    - Standard curves are not informative of biological variation



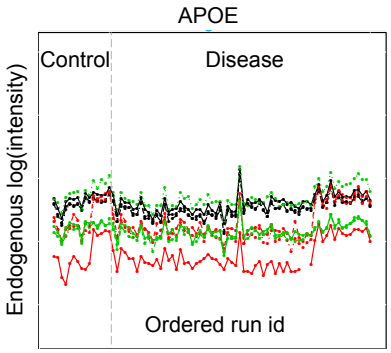


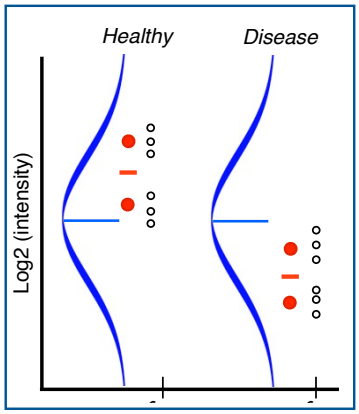
Altman and Bland, *The Statistician*, 1983
Altman and Bland, *The Lancet*, 1986

Protein-level quantification

(Detection of changes in protein abundance, summarized over peptides)

- Linear mixed effects model
  - Jointly model all measurements of a protein
  - Input: log(intensity of transition) (i.e. not ratio)
- User assumption: nature of interferences
  - *Additive model: interferences are noise*
    - Less sensitive, more specific
  - *Model with interactions: systematic interferences*
    - More sensitive, less specific
- User assumption: scope of conclusions
  - *Expanded to the underlying population*
    - Less sensitive, more specific
    - Needs random sampling and large sample size to characterize the variation in the population
  - *Restricted to the subjects in the study*
    - More sensitive, less specific



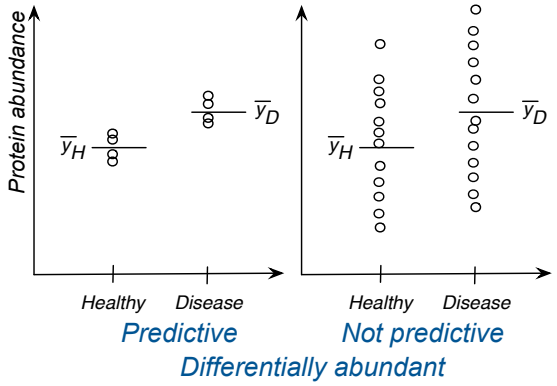


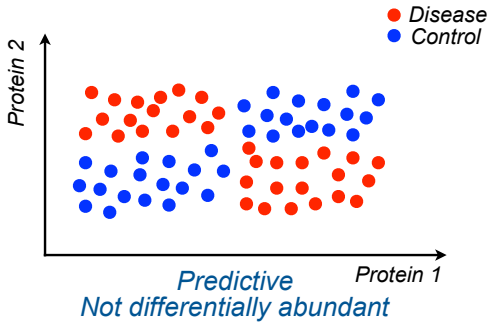
Chang et al., *MCP* 2012

What is a good assay?

(Does it answer the research question?)

- Accurately quantifies analytes?
  - LOQ and LOD, dynamic range
- Detects differential abundance?
  - *Dynamic range & technical variation*
    - Versus expected biological change & technical variation
    - On average over subjects
- Can be used as a biomarker?
  - *Is predictive?*
    - For an individual subject
  - *Is accurate?*
    - Sensitivity & specificity of classification
    - Evaluated on an independent validation set





Software and analytical tools

(Statistical software for relative protein quantification)

- Input
  - Statement of the problem
    - Groups to compare, scope of conclusions
  - Experimental design
  - Identified and quantified transitions
- Data analysis
  - Label-free, label-based; LC-MS, SRM, DIA
  - Group comparison, time course
- Experimental design
  - # of biological samples, peptides, transitions
  - *New:* # of reference proteins & peptides
  - *New:* a subset of informative features
- Availability
  - R-based, open-source

www.stat.purdue.edu/~ovitek

Specific to SRM

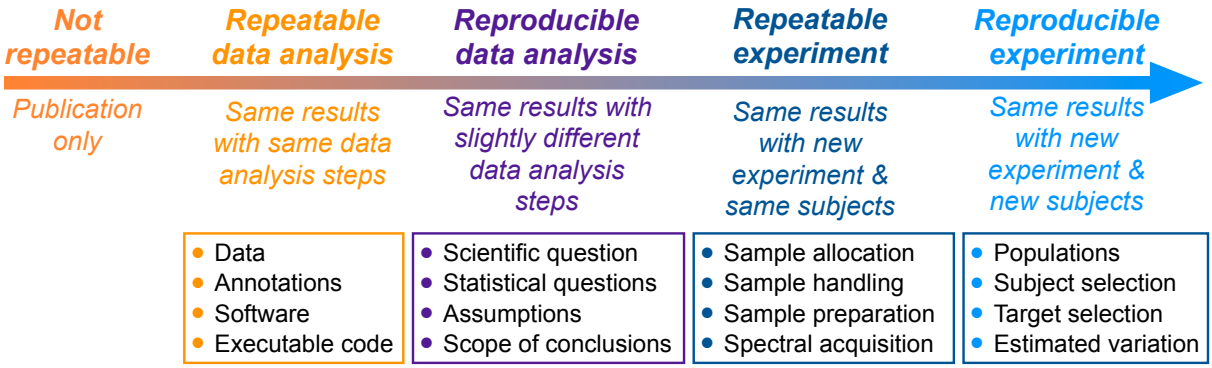
All functionalities
Is being integrated with Skyline
Will supersede SRMstats

External tool

Information in manuscripts

(Ensure reproducible research)

Reporting determines the scope of reproducibility



Example guidelines:

REporting recommendations for tumor MARKer prognostic studies (REMARK). *Nature Clinical Practice Oncology*. 2005

Transparent Reporting of Clinical Trials (CONSORT). www.consort-statement.org/consort-statement

Nature Journals: Reporting checklist for life sciences articles. www.nature.com/authors/policies/checklist.pdf

Large literature available

Automated & Reproducible Data Analysis Tools for Targeted Proteomics:

The QuaSAR Pipeline

[Quantitative and Statistical Analysis of Reaction Monitoring Results]

D. R. Mani
manidr@broadinstitute.org

Representing
Susan Abbatiello, Rushdy Ahmad, Deepak Mani and Steven Carr

Fit-for-purpose Assay Workshop
Jun 18, 2013

Goals: QuaSAR as an automated pipeline for MRM-MS data analysis

- Create an easy to use, automated platform for analysis of MRM-MS data
  - Generally applicable when multiple precursor-product pairs (transitions) are measured (e.g., parallel reaction monitoring)
  - Enable laboratory technicians or scientists to quickly perform standard analysis without statistical / computing expertise
- Encapsulate the best algorithms for commonly performed analyses
  - Calibration curves
  - Assay characterization and metrics
- Enable reproducible data analysis

Some Features under Construction

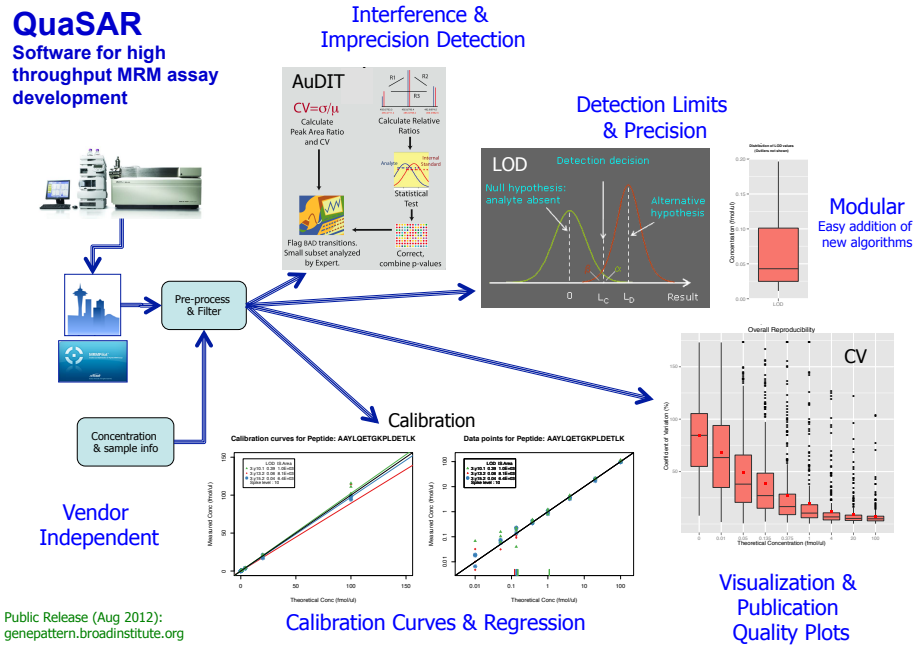
QuaSAR is optimized for targeted analysis with isotopically labeled internal standards

- Use of isotopically labeled peptide standards (with multiple transitions monitored per precursor) allows:
  - Confident identification of analyte being measured
  - Identification of interferences
  - More precise quantitation
    - Accounts for
      - » Ionization efficiency
      - » Run-to-run variation

- Some QuaSAR features are available only when internal standards are present

QuaSAR

Software for high throughput MRM assay development



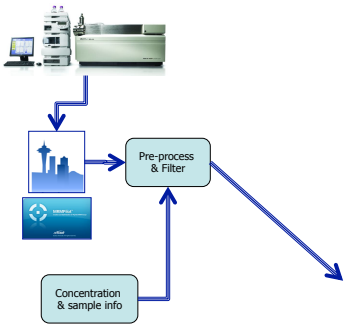
Public Release (Aug 2012): genepattern.broadinstitute.org

Mani, Abbatiello, Ahmad, Mani and Carr, 2013. Manuscript in Preparation

QuaSAR Input Data:
Exports from peak integration software

- Raw MRM-MS data is processed using peak integration software
  - E.g.: Skyline, Multi-Quant, Pinpoint
- Peak integration software determines:
  - Peak location
  - Identity / match to standard
  - Peak area, height, FWHM, and other peak characteristics
- QuaSAR input:
  - Exported table from peak integration software
  - Concentration, sample grouping and replicate info

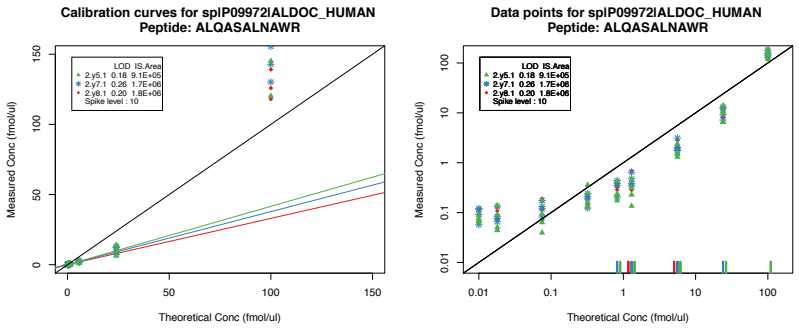
Calibration / Response Curves



- For each peptide QuaSAR outputs:
  - Linear calibration curves for every monitored transition for the peptide;
  - Log plots to visualize all raw data points, and identify trends.
- Calibration curves are fitted using robust, weighted regression.
  - Tables provide:
    - Slope and intercept along with confidence intervals
- Calibration curves can be based on:
  - Concentration (= PAR * IS Concentration)
  - Peak area ratio (= Analyte Peak Area / IS Peak Area)
  - Analyte peak intensity (area)

Calibration / Response Curves

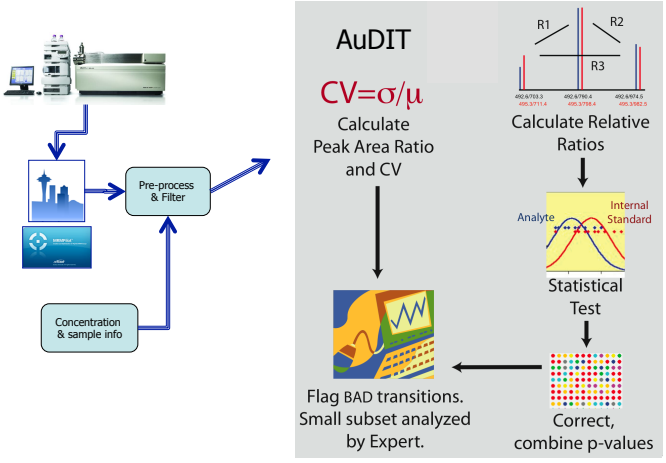
QuaSAR Output



peptide	transition.id	slope	slope stderr	y-intercept	y-intercept stderr
AAYLQETGKPLDETLK	2.y7.1	0.01472712	0.003205305	-0.000119491	8.35E-05
AAYLQETGKPLDETLK	2.y9.1	0.01078705	0.003212164	-0.000314087	8.36E-05
AEVNGLAAQGK	2.y5.1	0.31772655	0.049735831	-0.001208592	0.001294958
AEVNGLAAQGK	2.y8.1	0.25086684	0.027051172	-0.003264394	0.000704324
AEVNGLAAQGK	2.y7.1	0.28192988	0.049231984	-0.002154075	0.00128184
AGLCQTFVYGGCR	2.y8.1	-0.6687997	0.44916624	29.75333306	0.011694819
AGLCQTFVYGGCR	2.y7.1	-0.1765	0.548283165	32.56094202	0.014275499

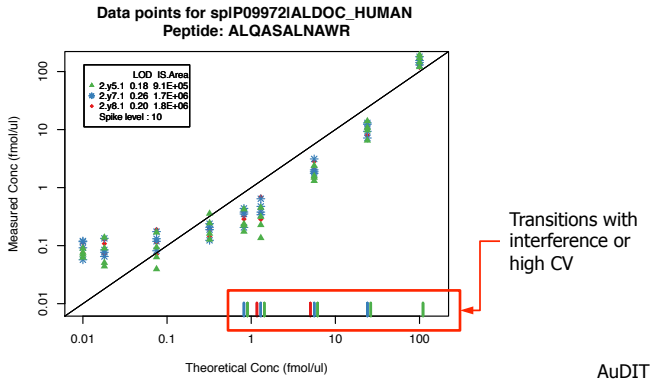
Regression Table

Interference & Imprecision Detection using AuDIT



Interference & Imprecision Detection using AuDIT

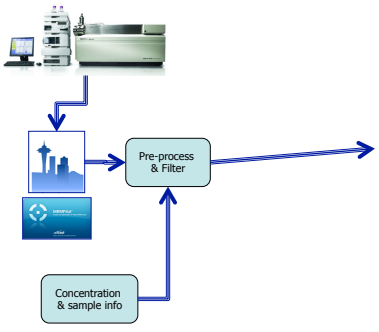
QuaSAR Output

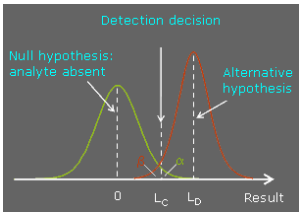


AuDIT Table

peptide	sample	transition.id	pvalue.final	status	cv	cv.status	final.call
AEVNGLAAQGK	I	2.y5.1	0.999436761	good	0.26378578	bad	bad
AEVNGLAAQGK	I	2.y7.1	0.999973076	good	0.17551101	good	good
AEVNGLAAQGK	I	2.y8.1	0.999998359	good	0.34918799	bad	bad
AEVNGLAAQGK	J	2.y5.1	0.529145605	good	0.11329347	good	good
AEVNGLAAQGK	J	2.y7.1	0.558946276	good	0.09584199	good	good

Limits of Detection (LOD) & Quantification (LOQ)

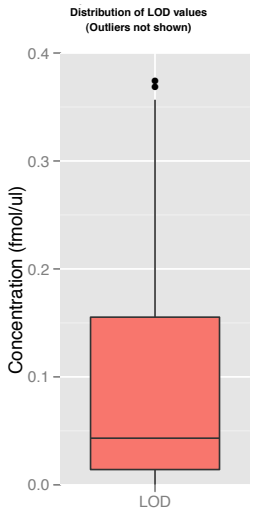




- Default method for calculating LOD and LOQ uses the blank (b) and a low concentration sample (s):
  - LOD = $\text{mean}_b + t_{(1-\alpha)}\, s.d_b + t_{(1-\beta)}\, s.d_S$
    LOQ = 3 × LOD
  - Other methods—ranging from algorithms using only the blank sample, to those that use the entire calibration curve, are also available for calculating LOD and LOQ in specific experimental settings.

Limits of Detection (LOD) & Quantification (LOQ)

QuaSAR Output



peptide	transition.id	LOD	LOQ
ALQASALNAWR	2.y8.1	0.20200176	0.60600528
ALQASALNAWR	2.y5.1	0.18219591	0.54658772
ALQASALNAWR	2.y7.1	0.26102768	0.78308304
ALYEAGER	2.y7.1	0.00664957	0.01994872
ALYEAGER	2.y5.1	0.10399518	0.31198554
ALYEAGER	2.y4.1	0.03927739	0.11783216
AMVALIDVFHQYSGR	3.y9.1	0.04975135	0.14925406
AMVALIDVFHQYSGR	3.y7.1	0.02702137	0.08106411
AMVALIDVFHQYSGR	3.y6.1	0.0360675	0.1082025

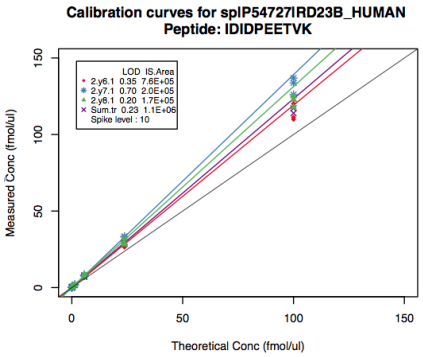
All Transitions

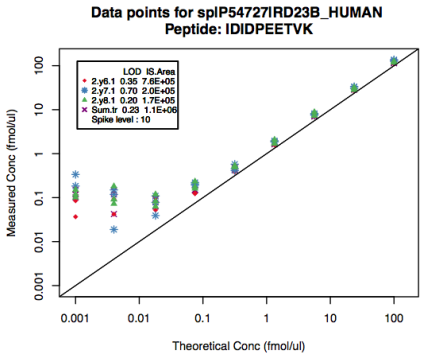
Quantifying Transitions

peptide	transition.id	LOD	LOQ
ALQASALNAWR	2.y5.1	0.18219591	0.54658772
ALYEAGER	2.y7.1	0.00664957	0.01994872
AMVALIDVFHQYSGR	3.y7.1	0.02702137	0.08106411

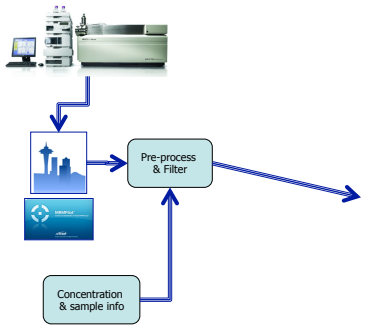
Choice of quantifying transition

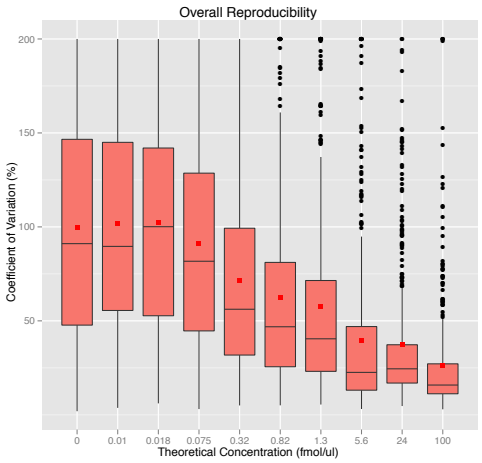
- Lowest LOD
  - Uses only a single transition to quantify peptide; other transitions are ignored
- Sum of transitions with no interferences
  - More robust
  - If new samples have different interferences, transitions constituting the sum can change, resulting in inconsistencies.





Assay Precision: Coefficient of Variation





CV Table for Quantifying Transitions

sample	peptide	transition.id	concentration	cv	mean.concentration
D	AEVNGLAAQGK	2.y7.1	0.075	155	0.01567
E	AEVNGLAAQGK	2.y7.1	0.32	101.5	0.1243
F	AEVNGLAAQGK	2.y7.1	0.82	54.42	0.2574
G	AEVNGLAAQGK	2.y7.1	1.3	88.33	0.1077
H	AEVNGLAAQGK	2.y7.1	5.6	52.39	1.216
I	AEVNGLAAQGK	2.y7.1	24	17.55	7.154
J	AEVNGLAAQGK	2.y7.1	100	9.584	48.01

New / Unknown Samples

- Configured assays can be run on new samples and the resulting data processed using QuaSAR
  - New samples are indicated using a missing value for the sample concentration in the concentration map
- QuaSAR calculates mean concentration and CV for each monitored transition
- Response curve is currently ignored during new sample concentration determination
  - Need to address response curve application when new samples are run separately from calibration curve samples.

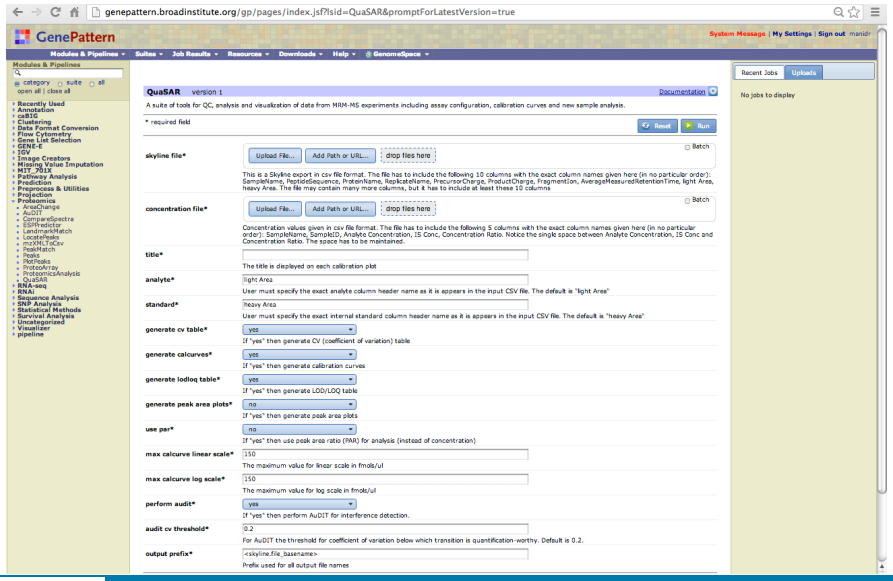
sample	peptide	transition.id	concentration	cv	mean.concentration
4	AAYLQETGKPLDETLK	2.y10.1	NA	140.6	0.007957
25	AAYLQETGKPLDETLK	2.y10.1	NA	108.3	0.3747
75	AAYLQETGKPLDETLK	2.y10.1	NA	65.8	1.057

CV Table for New Samples (Quantifying Transitions)

QuaSAR supports alternate workflows

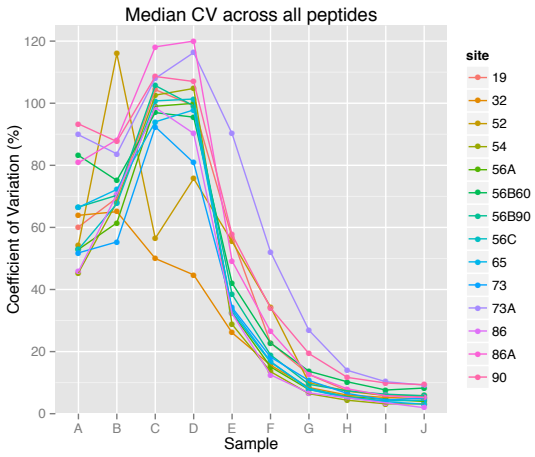
- "Reverse Calibration Curves"
  - Internal standard concentration is varied, keeping the analyte level constant
  - Used in situations where background matrix has endogenous analyte present
- Alternate internal standards
  - E.g.: Pre-digestion protein standards + post-digestion peptide standards to account for digestion efficiency / recovery
- No internal standards
  - Calculations based on peak area
  - QuaSAR does not confirm confident identification of peptide
    - Identification step delegated to peak integration software
- Sample processing variants with MRM-MS quantification
  - E.g.: iMRM-MS

QuaSAR in GenePattern



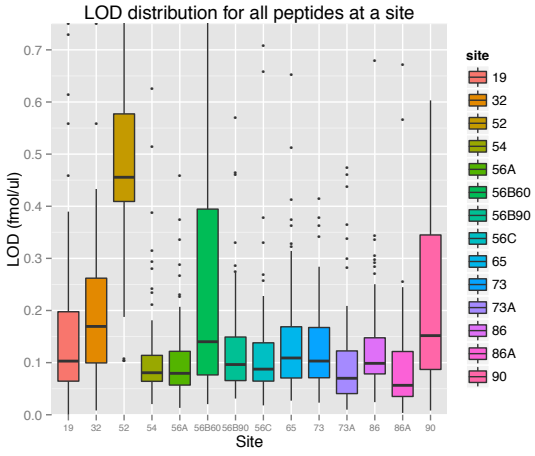
QuaSAR enables high-throughput analysis of multiple MRM-MS experiments

- Plots show median CV for a set of calibration curve samples with increasing concentration.
- Identical samples with pre-specified SOP were analyzed on 14 different instruments spanning 9 laboratories
- Data were processed using QuaSAR and plots created from QuaSAR output
  - QuaSAR analysis + plots took less than a day!



QuaSAR enables high-throughput analysis of multiple MRM-MS experiments

- Plots show LOD distribution for peptides at each site.
- Identical samples with pre-specified SOP were analyzed on 14 different instruments spanning 9 laboratories
- Data were processed using QuaSAR and plots created from QuaSAR output
  - QuaSAR analysis + plots took less than a day!



Summary of QuaSAR features

- QuaSAR implements a comprehensive, easy to use pipeline for MRM-MS data analysis including:
  - ***Statistics***: For every monitored peptide
    - Coefficient of variation (CV)
    - Regression slope and intercept (with confidence intervals)
    - Interference detection, and
    - Limits of detection (LOD) and quantification (LOQ)
  - ***Visualization***: Succinct visual summaries of various results including reproducibility, interferences and detection limits are generated.
  - ***Interpretation***: Statistics and visualization are integrated to enable effective data interpretation, understanding and insight.

QuaSAR can promote standardization of data analysis in manuscripts

- Use of QuaSAR results in reproducible data analysis:
  - QuaSAR version + parameter settings completely specifies analysis
  - Easy replication of data analysis at other sites/laboratories
- Custom analysis would require authors to specify:
  - Determination and use of response/calibration curves
  - Interference detection and action taken
  - Assessment of assay precision
  - Specification of assay sensitivity (limits of detection/ quantification)

BIOGNOSYS
PROTEOMICS – NEXT GENERATION

National Institute of Health
Assay Workshop – Data Analysis

Lukas Reiter

1) Goals and Customers

- What are the goals of your targeted MS experiments or software?
- Who are the “customers” or likely users of the methods and results?

1) Goals and Customers

Goals

As a Company
- Robust and easy to use products

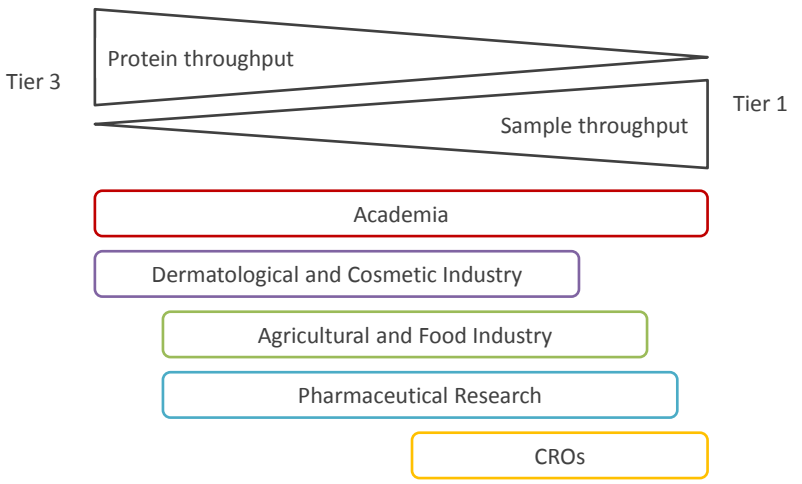
For Technology
- Suitability to answer question
- Sensitivity
- Quantitative accuracy/precision
- Sample throughput
- Analyte throughput

Of Experiments
- Discovery (DIA)
- Mechanistic insight
- Biomarker validation
- Absolute quantification
- Monitor specific modification

For Software
- Automated
- Objective and results comparable
- Robust and flexible (chromatography, technology, plex-level,...)

1) Goals and Customers

Customers



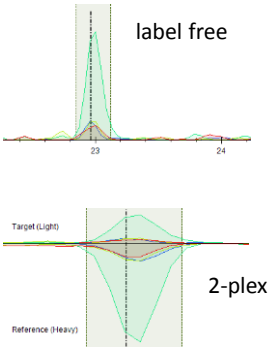
2) Plex Level

- What range of analyte plex-level per-injection do you typically use in your experiments?
- What is the impact of the plex-level used on the robustness and figures of merit (CV, LOD, LOQ) of the developed MRM/SRM assays?

2) Plex Level

Range

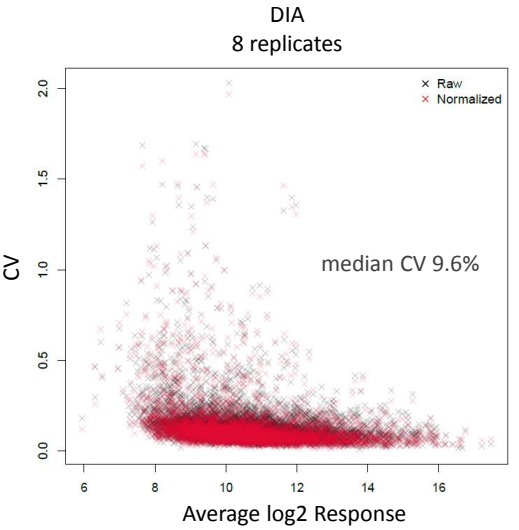
- DIA
  - Label free
  - Label (2-plex)
- MRM/PRM
  - Plex level 1 or 2



2) Plex Level

Impact

Influence on LOD, LOQ and CV not systematically checked



BIOGNOSYS
PROTEOMICS – NEXT GENERATION

7

3) Analyte Identification

- Explain how you establish confidence that what is being measured is the analyte of interest (e.g., match to spectra of an internal standards, match to reference spectra from discovery experiments, RT, etc.).
- How do these methods differ from “Discovery Proteomics” using data-dependent or data-independent experiments?

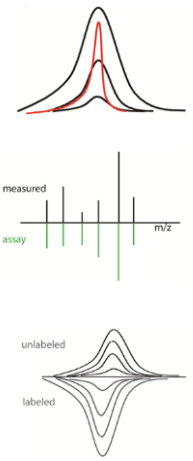
BIOGNOSYS
PROTEOMICS – NEXT GENERATION

8

3) Analyte Identification

Method – Scoring

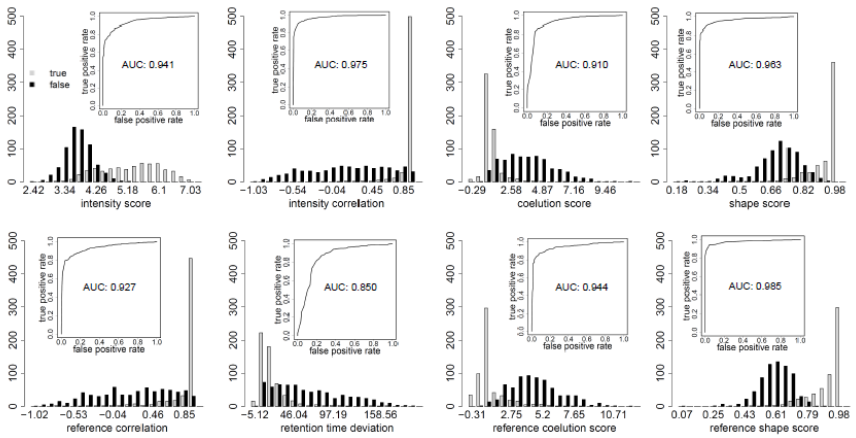
- Internal scores
  - Shape correlation
- External scores
  - Relative fragment ion current
  - iRT (retention time)
- Plex scores
  - Relative fragment ion current
  - iRT (retention time)
  - Shape correlation



BIOGNOSYS
PROTEOMICS – NEXT GENERATION

3) Analyte Identification

Method – Scoring



True signal
False signal

Reiter et al. NM 2011

BIOGNOSYS
PROTEOMICS – NEXT GENERATION

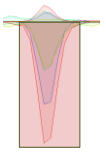
3) Analyte Identification

Method – Scoring

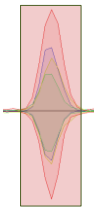
Workflows

- Label free
- 2-plex
  - "Spike in" (typically SIS)
    - Peak detection on internal standard
    - Peak scoring on endogenous
  - "Label"
    - Peak detection and scoring on both channels

spike in

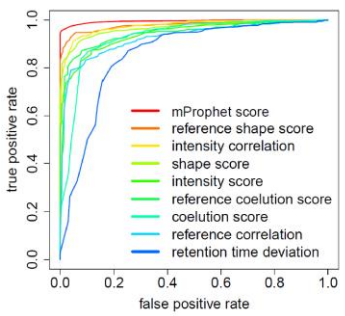


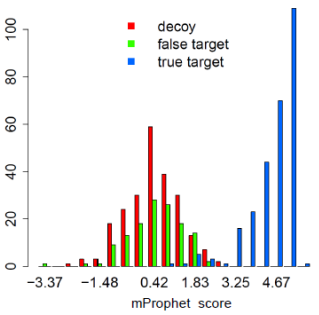
label



3) Analyte Identification

Method – Scoring





Scores are optimally combined using semi-supervised machine learning

Reiter et al. NM 2011

3) Analyte Identification

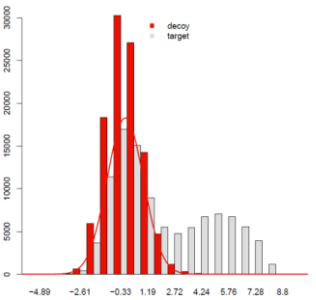
Method – Qvalues or false discovery rates

How to estimate Qvalues (FDR) in MRM?

- Statistically characterize the signals that result from an assay being measured in a matrix where the analyte is not detectable

Decoy transitions

- Assays for molecules that do not exist
- Negative controls



Assay signal
Decoy signal

3) Analyte Identification

Method – Comparison to discovery proteomics

- Can be applied to DIA, MRM and PRM
- Comparison to shotgun proteomics
  - Identical goal to estimate a Qvalue/FDR
  - Distinct method
    - FDR = PSM_{decoy}/PSM_{target} does not work

3) Analyte Identification

Method – Use

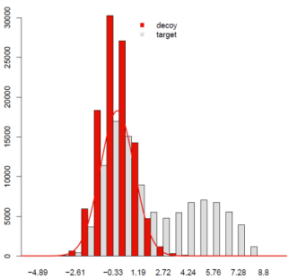
Qvalues

- Always calculated
- Especially useful for
  - Label free
  - DIA and large MRM/PRM data sets
- Combined with manual data review in small MRM/PRM data sets
  - i.e. dilution series

3) Analyte Identification

Method – Cons

- Decoys have to be available
- Decoys have to be representative
- Parameterization of decoy distribution
- Cutting away the left tail of the target distribution
- Does not solve issue to get complete quantification matrices



3) Analyte Identification

Method – Pros

- Large data sets (DIA)
- Why helpful in case internal standards are available?
  - Detection of internal standard signal
  - If LOD is missing
  - Can be hard to determine a “real” LOD
    - Matrix just missing the endogenous analyte required, pure light and heavy analyte, one of the two with known absolute quantity needed
    - Variability of sample matrix (e.g. human blood)
  - Transferability to other laboratories, machines,…

4) Analyte Quantification

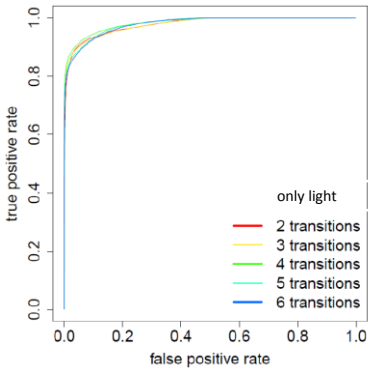
- Explain your method of quantification, how many transitions you monitor and which ones are chosen to quantify.
- If you are using internal standards describe in detail how they are used.
- If you are not using internal standards, explain how you are quantifying.
- Discuss the capabilities and limitations of your approach.

4) Analyte Quantification

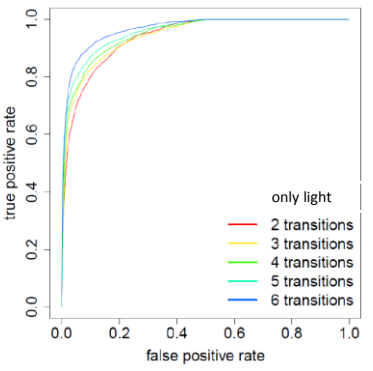
MRM

Identifications in MRM (n=591)

2-plex ROC curves



label free ROC curves



Reiter et al. NM 2011

4) Analyte Quantification

MRM/PRM

- Transitions with highest response
- Transitions with interferences removed
- 1-plex
  - ~5 transitions
  - Relative quantification
- 2-plex
  - ~3 transitions (total 6)
  - Relative and absolute quantification
- Sum of peak areas (or peak heights) for an analyte

4) Analyte Quantification

MRM/PRM – Internal standards

- Same number of transitions monitored for endogenous as for internal standard
- S/N of internal standard ~500
- Sum of transition areas for endogenous and internal standard
- Ratio of sums endogenous to internal standard

4) Analyte Quantification

DIA – Label free

- ~6-10 most responsive fragment ions
- Sums of peak areas (or peak heights)
- Normalization
- Relative quantification

4) Analyte Quantification

Capabilities and limitations

"Quantify the analyte in the MS sample vial"

- Absolute analyte quantification with internal standard
  - Losses of internal standard (absorption,…)
- Relative quantification with internal standard
  - Variation of ion spray, MS sensitivity etc. can be balanced out
- Label free relative quantification
  - Experiment design and normalization
- Carry over should be considered for experiments where this can be an issue

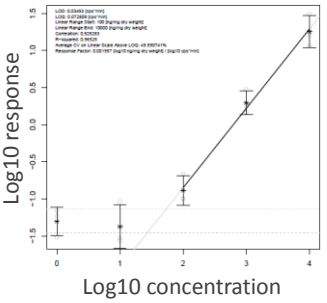
5) Standard Curves

- If you generate standard curves (calibration or response curve), explain how you use them to assess the quantitative accuracy of the assay.
- Are the slope and y-intercept from the curve regression used in calculating the analyte concentration in a sample?
- Is an external calibration curve used?

5) Standard Curves

Handling and Use

- Exponential dilution (n=3)
- Calibration curves
  - LOD/LOQ (absolute or relative)
  - CV in linear range
  - Specification sheet
  - Quality control
- Accuracy difficult to determine (->ELISA)
- Precision estimated based on CVs in linear range
- Calibration curves not used to determine analyte concentration



6) Protein Quantitation Using Peptides

- Can you provide a useful estimate or accurately determine the amount of protein in the matrix based on the measured levels of peptides?
- Explain how/why.
- Indicate experimental parameters such as number of peptides per protein and the criteria/computational tools applied.
- If you have multiple peptides from the same protein and each gives a different answer for the extrapolated protein level, how do you deal with this?

6) Protein Quantitation Using Peptides

Challenges

Challenges for absolute quantitation with internal standard

- Determination of amount of starting material
  - e.g cell volume, number of cells, total protein
- Recovery of protein from sample preparation protocol
- Protein digest
- Protein modifications
- Proteotypicity of peptides (specific to protein or gene)
- Losses of internal standard

6) Protein Quantitation Using Peptides

Measures

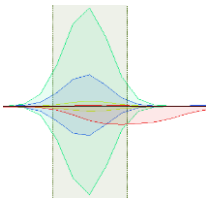
- Repeat digest and sample preparation and monitor variance
- Multiple peptides per protein
  - DIA: 5 to as many as possible peptides per protein
  - MRM: start with 3-5 peptides per protein, scheduling
- Consider proteotypicity
- Consistency
  - DIA: average
  - MRM: Select most stable peptides
- Recombinant protein
  - Modifications (glycosylation), fold,…?

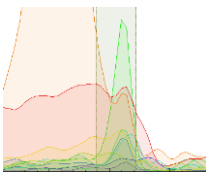
7) Interferences

Describe methods you use to establish presence of interferences and how you deal with them if detected.

7) Interferences

- Manual detection and removal
- Automatic refinement with interference score*
  - Based on shape correlations





*Similar as in Abbatiello et al. 2010 Clinical Chemistry

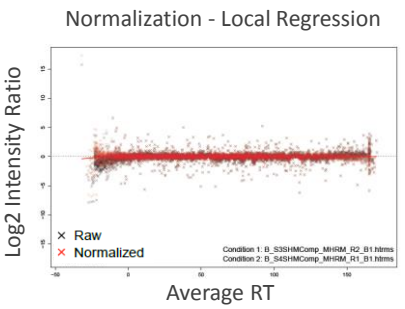
8) Ion suppression

How do you account for suppression of ionization in your quantification method?

8) Ion suppression

Label free

- Stable spray conditions
- Blocking and randomization of acquisition sequence
- Normalization*
  - Account for time drifts in ionization affecting all/most eluting peptides
  - Central tendency
  - Local regression
- Use several peptides per protein



*Callister et al. 2006 JPR

9) Assay Quality Control

How do you “qualify” your measurements/assays, i.e., what criteria do you use or think appropriate to say that your measurements/assays has been successfully developed?

9) Assay Quality Control

- Protein relative response factor
- Proteotypicity for protein or gene
- Accurate iRT (indexed retention time of apex)
- Most responsive fragment ions and relative intensities (no interferences)
- Low sample preparation variance
- Label free
  - Detectability (Qvalue)
- Absolute quantification with internal standards
  - LOD/LOQ
  - Linear range

10) Software

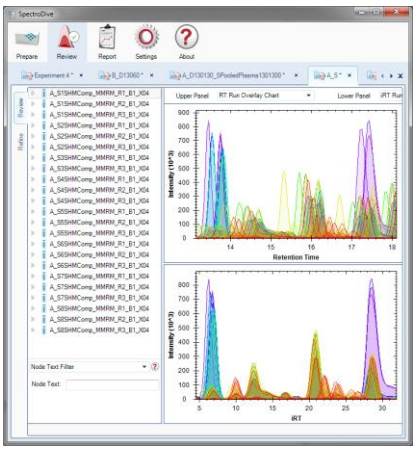
What software and analytical tools do you use in your studies and why?

10) Software

MRM/PRM

SpectroDive

- Biognosys internal software
- Scheduling with iRT
- Automated parameter calibration
- Scoring and Qvalues
- Profiling across runs
- Automated assay refinement and panel versioning



10) Software

DIA

Spectronaut

- Biognosys software
- Automated parameter calibration
- Fast and low memory requirements
- Scoring and Qvalues
- Profiling across runs
- Targeted data extraction with iRT

11) Information for Publication

What information do authors need to provide in their manuscripts/supplement to enable reviewers and readers to understand what was done and to be able to judge the confidence of the measurements made?

11) Information for Publication

- Algorithm/software version and parameters
- Acquisition sequence (randomization/blocking)
- Identification criteria
  - Algorithm description
  - Criteria for manual peak identification/integration
- Normalization
- Proteotypicity of peptides
- Transitions per peptide
- Peptides per protein
- Discussion of known challenges regarding absolute quantification

Seattle Proteome Center

INSTITUTE FOR Systems Biology
Revolutionizing science. Enhancing life.

Targeted Peptide Measurements in Biology and Medicine: Best Practices for Assay Development

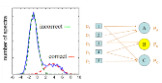
Software for Data Acquisition and Databasing

Eric Deutsch

PeptideAtlas Trans Proteomic Pipeline

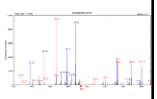
NIGMS NHGRI European Commission

Overview of PeptideAtlas Project Components



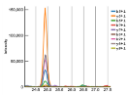
Trans Proteomic Pipeline — Complete software suite of tools for processing MS/MS data

PeptideAtlas — Compendium of results from uniform processing of many MS/MS datasets



SRMAtlas — Compendium of ranked peptides and transitions for all proteins in select proteomes

PASSEL — Repository for submitted SRM experimental datasets and results



Proteomics Standards Initiative

ProteomeXchange

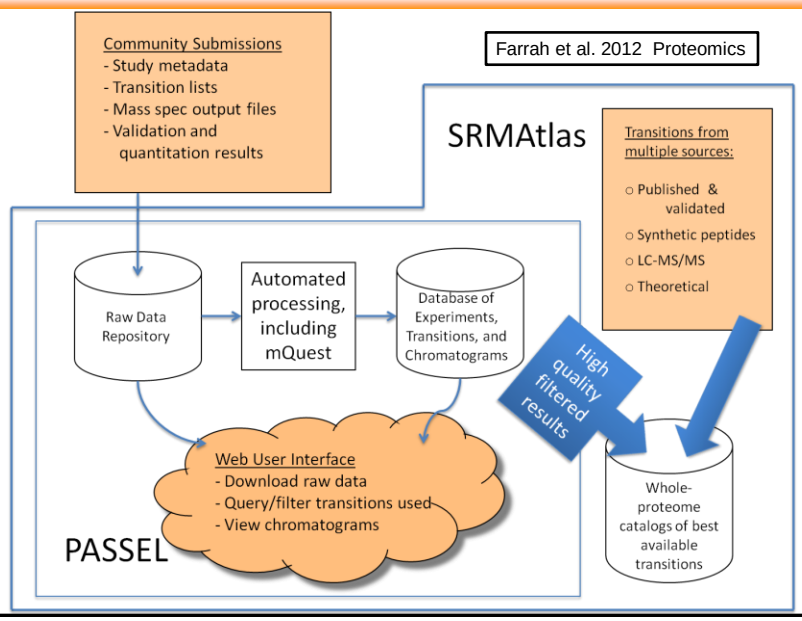
PRIDE Tranche

Human Proteome Project

2

PeptideAtlas SRM Experiment Library (PASSEL)

Data repository specifically designed for SRM experimental results

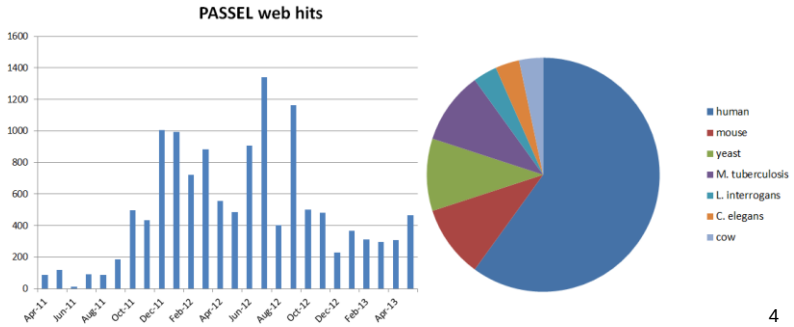


3

Q/A: Goals, Customers

What are the goals of your targeted MS experiments or software? Who are the "customers" or likely users of the methods and results?

- Develop software and resources to facilitate design and accessibility of SRM experiments
  - SRMAtlas catalogs best-available information on potential assays
  - PASSEL stores resulting experimental data
  - Experimental results feed back into SRMAtlas



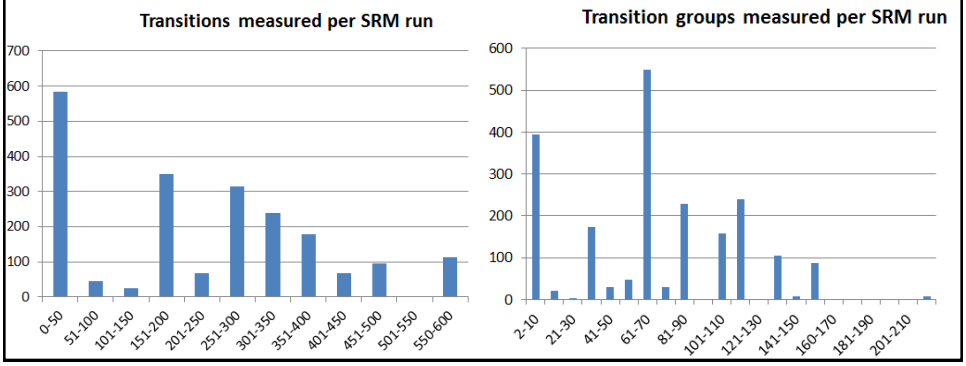
PeptideAtlas

4

Q/A: Analyte Plex Level

What range of analyte plex-level per injection do you typically use in your experiments?

- PASSEL current has 20 public experiments. 3 about to be released.
- Another 5 that are under review and thus still private
- 2163 total SRM runs
- 592709 total (redundant) Q1/Q3 pairs measured



Q/A: ID confidence

Explain how you establish confidence that what is being measured is the analyte of interest (e.g., match to spectra of an internal standards, match to reference spectra from discovery experiments, RT, etc.). How do these methods differ from "Discovery Proteomics" using data-dependent or data-independent experiments?

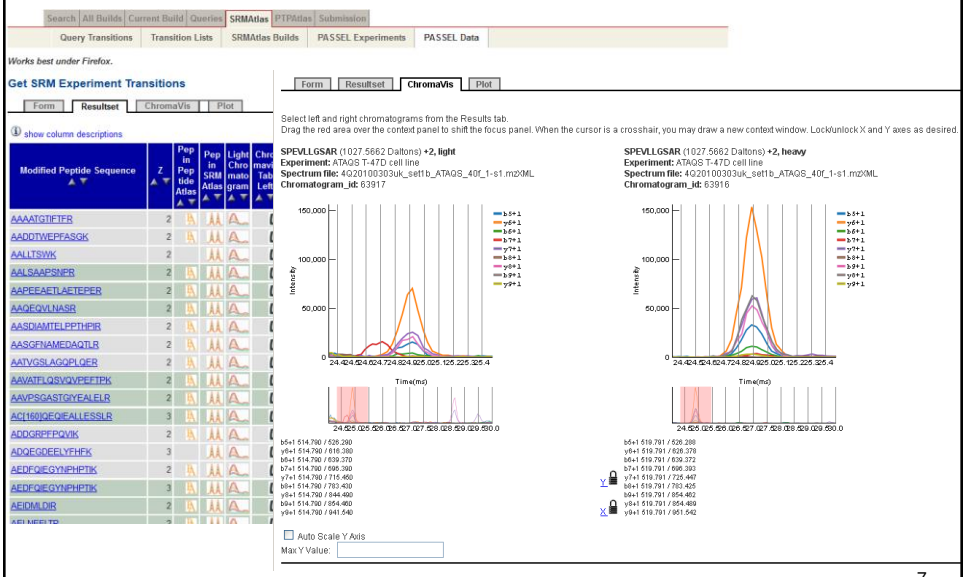
- PASSEL processing pipeline is still an evolving process
- Currently uses mQuest and mProphet for statistical analysis
- mProphet scores are loaded into PASSEL for viewing when possible
- Requires decoy transitions

- This pipeline is similar to the PeptideAtlas – shotgun pipeline, which uses decoy and model-based confidence metrics

PeptideAtlas

6

PASSEL Transition Browser

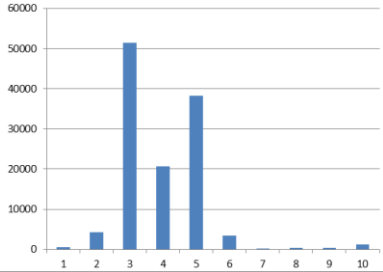


7

Q/A: Quantification

Explain your method of quantification, how many transitions you monitor and which ones are chosen to quantify. If you are using internal standards describe in detail how they are used. If you are not using internal standards, explain how you are quantifying. Discuss the capabilities and limitations of your approach.

- Goal of PASSEL is not to reproduce/automatically derive quantification
- Goal of PASSEL is to disseminate experimental data along with uniformly derived identification confidence for reuse in assay development and experimental design



PeptideAtlas

8

Q/A: Calibration Curves

If you generate standard curves (calibration or response curve), explain how you use them to assess the quantitative accuracy of the assay (e.g., are the slope and y-intercept from the curve regression used in calculating the analyte concentration in a sample? Is an external calibration curve used?)

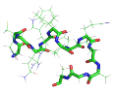
- A new component of SRMAtlas: SRMQuantAtlas
- Asseses peptides and transitions suitability via calibration curves

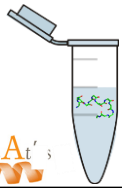
- Initially human peptides in simple mixture, plasma, urine matrix

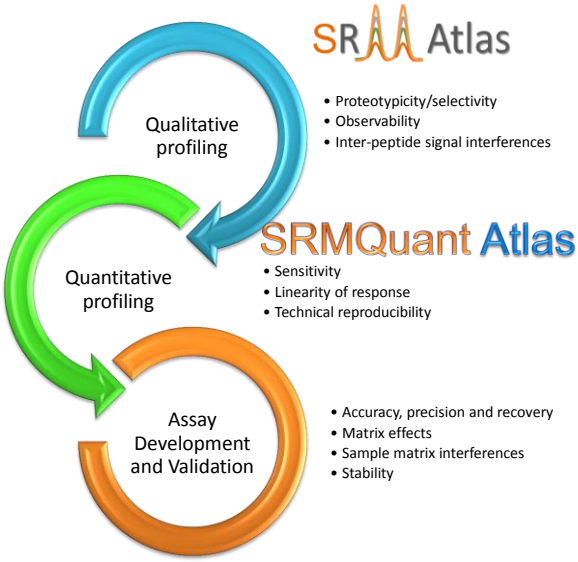
PeptideAtlas

9

Quantitative SRM Assay Development







PeptideAtlas

10

SRMQuantAtlas Workflow

Peptide pool
(95 plus standard)

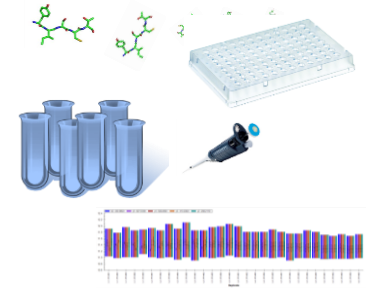
Dilution series
(L1-L6; 1024-fold)

Replicate data
(n=5)

Quantitative analysis
(<20% peak area CV)

Calibration curve
min. L4-L6

LOD/LOQ



$$LOD = \frac{3s}{m}; LOQ = \frac{10s}{m}$$

PeptideAtlas

SRMQuantAtlas Interface

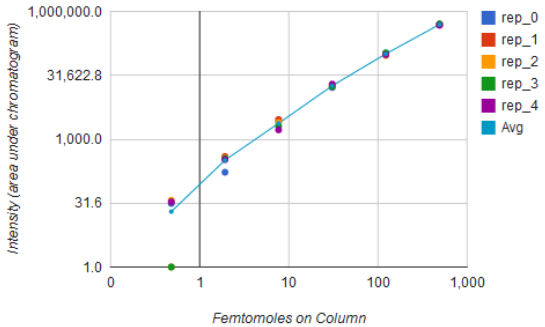
Search | All Builds | Current Build | Queries | SRMAtlas | PTPAtlas | Submission

show row descriptions

Quantitation information

Peptide Sequence	FMTDPHAMR
Precursor m/z	372.505
Titration range	0.18-179.3
R-squared value	0.9995
LOD	0.06
LOQ	0.20
Linear range	0.20-179.3
Fold change	24.0
Is quantitative	Yes

Content-type: text/html



PeptideAtlas

Q/A: Protein Abundance

Can you provide a useful estimate or accurately determine the amount of protein in the matrix based on the measured levels of peptides? Explain how/why. Indicate experimental parameters such as number of peptides per protein and the criteria/computational tools applied. If you have multiple peptides from the same protein and each gives a different answer for the extrapolated protein level, how do you deal with this?

- Not currently within scope of PASSEL

PeptideAtlas

13

Q/A: Interferences

Describe methods you use to establish presence of interferences and how you deal with them if detected.

- Interference taken in account in mProphet modeling for ID confidence calculation, but otherwise no PASSEL metric for interference severity

PeptideAtlas

14

Q/A: Ion Suppression

How do you account for suppression of ionization in your quantification method?

- Note in scope of PASSEL

Q/A: Assay Qualification

How do you "qualify" your measurements/assays, i.e., what criteria do you use or think appropriate to say that your measurements/assays has been successfully developed?

- My opinion:

- All measurements should be qualified with a probability that the measured peak group is indeed the intended analyte
- Threshold can be authors' choice

- Quantitative measurements should have associated uncertainties
  - CVs are nice
  - But how about a real uncertainty: ± n.nn
    - Relative to other measurements
    - Absolute if appropriate
    - Maybe still to hard. Need better software.

Q/A: Software

What software and analytical tools do you use in your studies and why?

- mProphet toolset
- Skyline
- Vendor software

- SRMAtlas & related in-house tools
- SRMQuantAtlas & related in-house tools

PeptideAtlas

17

Q/A: Metadata Dissemination

What information do authors need to provide in their manuscripts/ supplement to enable reviewers and readers to understand what was done and to be able to judge the confidence of the measurements made?

- Currently PASSEL has rather modest requirement to encourage use:
  - Mass spec output files (vendor format or mzML)
  - Transition lists (TraML or whatever)
  - Basic metadata in a submission form (basic for now)
  - Analysis results in whatever format (not currently used, but available for download)

- MIAPE – Quant is a PSI Minimum Information specification for all quantitative study data, including SRM
- TraML is a PSI data format for encoding SRM transitions
- mzQuantML is a PSI data format for all quantitative study data, including SRM
- mzTab is a PSI format under development for simplified quant data

PeptideAtlas

18

MIAPE – Quant

- MIAPE = Minimum Information About a Proteomics Experiment
- Set of minimum information elements for the quantitative aspect of a proteomics experiment
- Complements MIAPE – MS + MIAPE – MSI

- Submitted through the PSI document process
- Explicitly sought input from:
  - Data metrics workshop participants
  - HUPO New Technologies Initiative members
  - Journal editors

- Published: Martínez-Bartolomé, S., Deutsch, E. W., et al. 2013, Journal of Proteomics, S1874-3919, 102-4

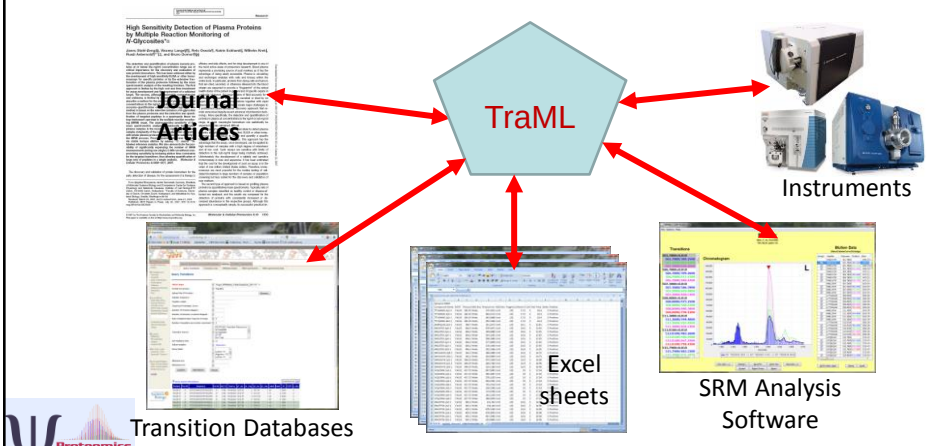
19

1. General features:
   - 1.1. Experiment identifier or name
   - 1.2. Date stamp
   - 1.3. Responsible person or role
   - 1.4. Associated MIAPE documents
   - 1.5. Quantitative approach
2. Experimental design and sample description:
   - 2.1. Experimental design
     - 2.1.1. Groups
     - 2.1.2. Biological and technical replicates
   - 2.2. Sample description
     - 2.2.1. Labelling protocol
     - 2.2.2. for each sample:
       - 2.2.2.1. Sample name
       - 2.2.2.2. Sample labelling with reporting ion mass, reagent or isotope labelled amino acid
       - 2.2.2.3. Replicate and/or group
     - 2.2.3. Isotopic corrections
   - 2.3. Internal references
3. Input data:
   - 3.1. Input data type
   - 3.2. Input data format
   - 3.3. Availability of the input data
4. Protocol:
   - 4.1. Quantification software
     - 4.1.1. Software name and version
     - 4.1.2. Parameters used for the quantitative process
   - 4.2. Confidence filter of features prior to quantification
   - 4.3. Description of the method of the quantitative values calculation for each feature
   - 4.4. Combination of quantification values
   - 4.5. Calibration curves of standards
5. Resulting data:
   - 5.1. Quantification values at peptide level:
     - 5.1.1. Quantification values
     - 5.1.2. Statistical values associated with the quantification value
   - 5.2. Quantification values at protein level:
     - 5.2.1. Quantification values
     - 5.2.2. Statistical values associated with the quantification value

20

Unifying exchange of transitions with TraML

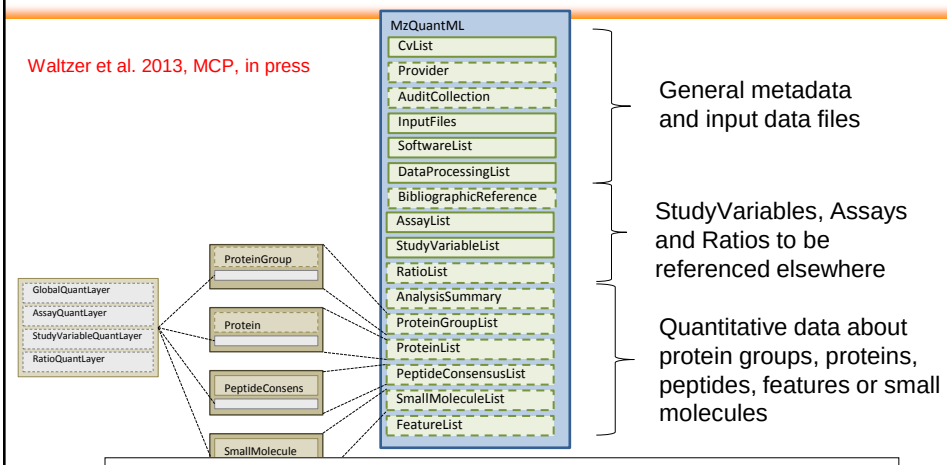
- PSI's TraML (Transitions Markup Language) (similar to mzML et al.)
- Format for encoding SRM transitions
- Version 1.0.0 now finalized and published in MCP



21

mzQuantML - for encoding the quantitative aspects of MS1, MS2, SRM, etc. experiments

Waltzer et al. 2013, MCP, in press



Important features:
- Captures evidence trail from quantitative values about:
  protein group → protein → peptide → features on 2D LC-MS maps
- Currently this evidence trail is lost in almost all workflows
- Without this, very difficult to visualise each step of the analysis workflow

22

mzTab - Aims

- To provide a lightweight format for exchanging proteomics data
  - Simple summary report of the experimental results
  - Peptides identified in a given experimental setting
  - Reported quantification values
  - Metadata
  - Not ‘direct’ support for technical/biological replicates
- Easy to use by the proteomics community, systems biologists as well as providers of knowledge bases
- It can be used by non-experts in bioinformatics

23

mzTab – current draft

- What the format does NOT aim to do:
  - Replace mzIdentML or mzQuantML
  - Contain the complete data of a MS proteomics experiment
  - Provide detailed evidence for the data
  - Allow a researcher to recreate the process which led to the results
  - Be requirements conforming (MIAPE, journal guidelines, etc.)
  - In short: be complete in any way
- But does provide a standardized format for users who are going to transform mzIdentML, mzQuantML into TSV anyway
- Currently in the PSI document process

24

ProteomeXchange Consortium

- **Goal:** Sharing proteomics data between main existing proteomics repositories.

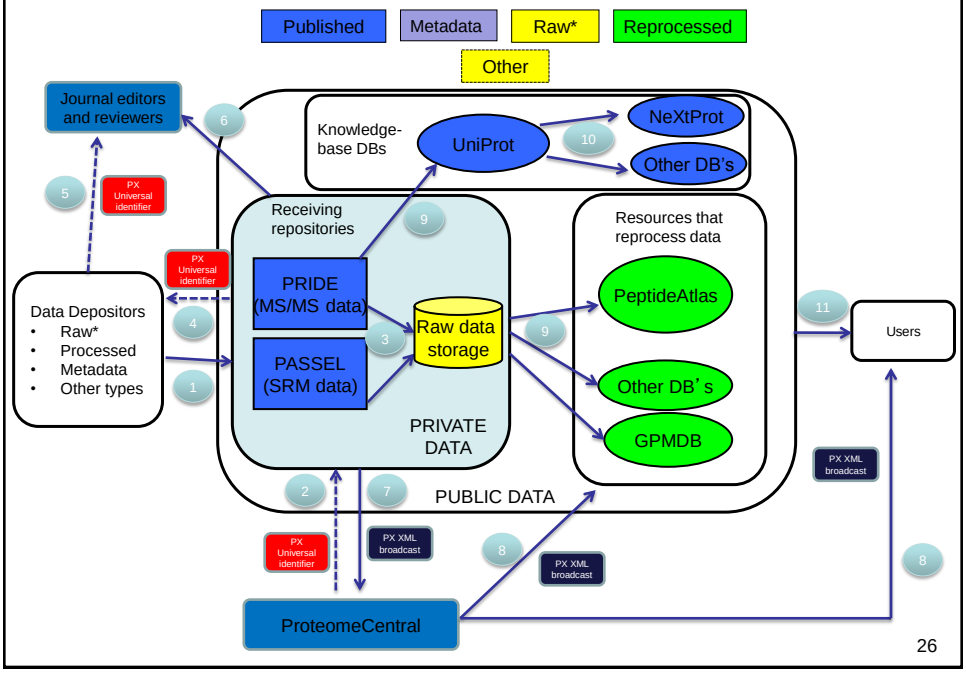
- Includes PeptideAtlas and PRIDE, with data sharing infrastructure to be provided at the EBI and ISB.

- ProteomeXchange is primarily **user-oriented**: the idea is to provide a **single point of submission**, but **multiple points of data visualization and analysis.**

- Since January 2011, supported by the European Commission (FP7 framework). SIB (UniProt) also involved in the grant.

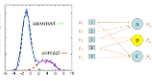
http://www.proteomexchange.org

25



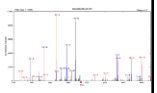
26

Overview of PeptideAtlas Project Components



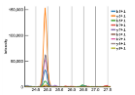
Trans Proteomic Pipeline — Complete software suite of tools for processing MS/MS data

PeptideAtlas — Compendium of results from uniform processing of many MS/MS datasets



SRMAtlas — Compendium of ranked peptides and transitions for all proteins in select proteomes

PASSEL — Repository for submitted SRM experimental datasets and results



SRMQuantAtlas — Compendium of peptides and transitions with calibration curves

Proteomics Standards Initiative

ProteomeXchange

PRIDE

Tranche

Human Proteome Project

Thank you!

Questions?

PeptideAtlas

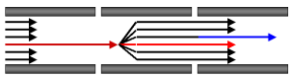
Skyline

Targeted Proteomics Environment

Tools for targeted assay development and data analysis

Brendan MacLean
MacCoss Lab

UNIVERSITY of WASHINGTON



Chromatography-based Quantification

- SRM
- MS1 chromatogram extraction
- Targeted MS/MS
- Data independent acquisition (DIA/SWATH)

Skyline

- 1000+ registered users, 4500+ instances each week
- Supporting 5 instrument vendors, 4 funding development
- Supporting 15 peptide search engines, 4 library formats
- Free and open source (Apache 2.0)

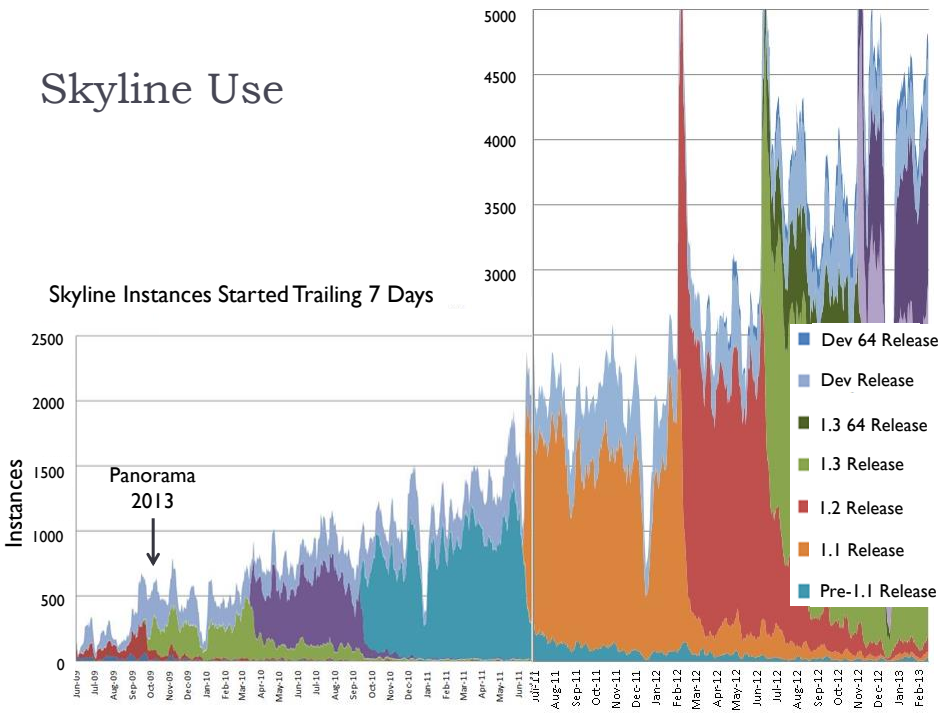
Aggregating and Publishing

- Publish fully annotated Skyline documents
- Build chromatogram libraries
- Aggregate lab QC data (future)

Panorama

- Free hosted version (http://panoramaweb.org)
  - 23 separate projects so far (CPTAC and ABRF sPRG)
  - User controlled security
- Locally installable server application
- Free and open source (Apache 2.0)

Skyline Use



Analyte Plex Level

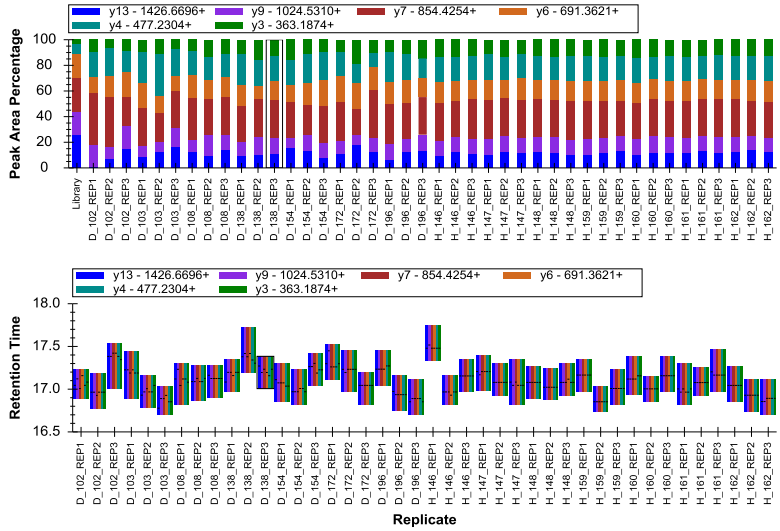
- Pushing vendor limits
  - >320 transitions on Thermo
  - >1000 transitions on AB SCIEX
  - >256 light-heavy precursor pairs on Waters
- Vendors pushing the limits with Skyline
  - Agilent and AB SCIEX triggered MRM
  - Thermo iSRM
  - Agilent, AB SCIEX and Thermo scheduled targeted MS/MS
  - DIA/SWATH

Prior Knowledge and Consistency

- Powerful enough to be used cross-lab / cross experiment
- More powerful run-to-run
- Relative ion abundance
  - Spectral and chromatogram libraries
- Retention time
  - iRT
- Does ensuring comparable measurements require ID?

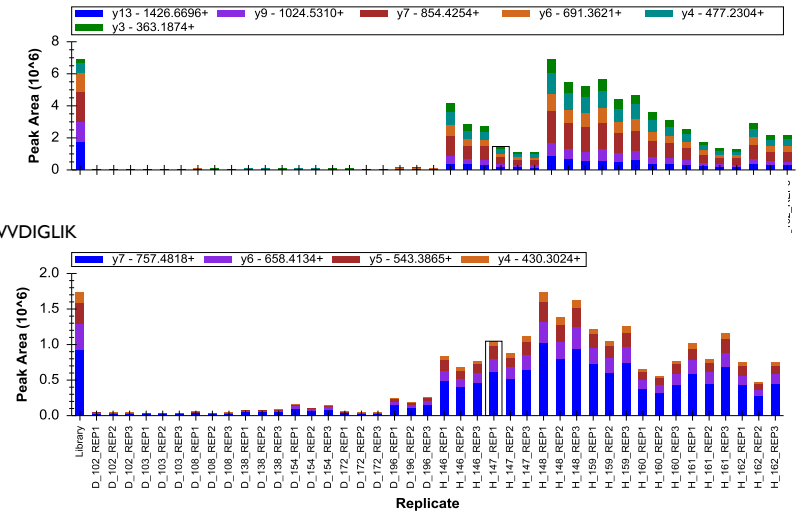
Haptoglobin

LQTEGDGIYTLNSEK



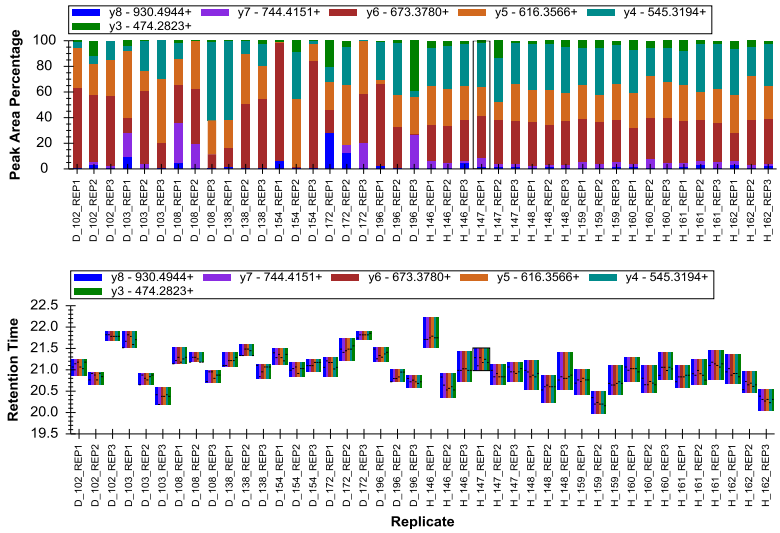
Haptoglobin

LQTEGDGIYTLNSEK



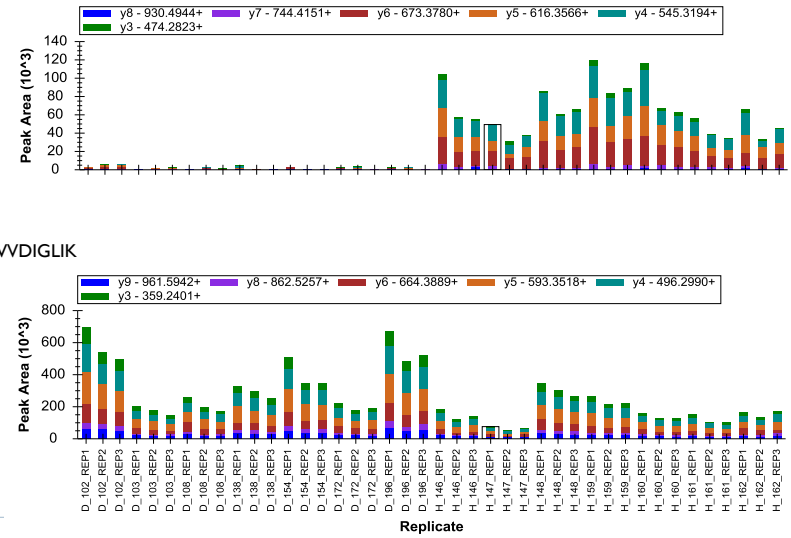
Mitochondrial 39S ribosomal protein L9

CSSLLWAGAAWLR



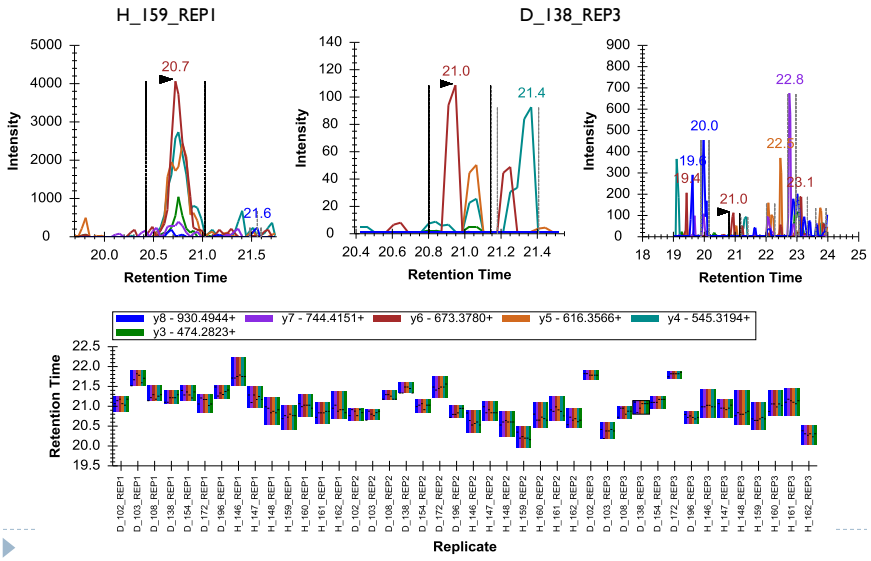
Mitochondrial 39S ribosomal protein L9

CSSLLWAGAAWLR



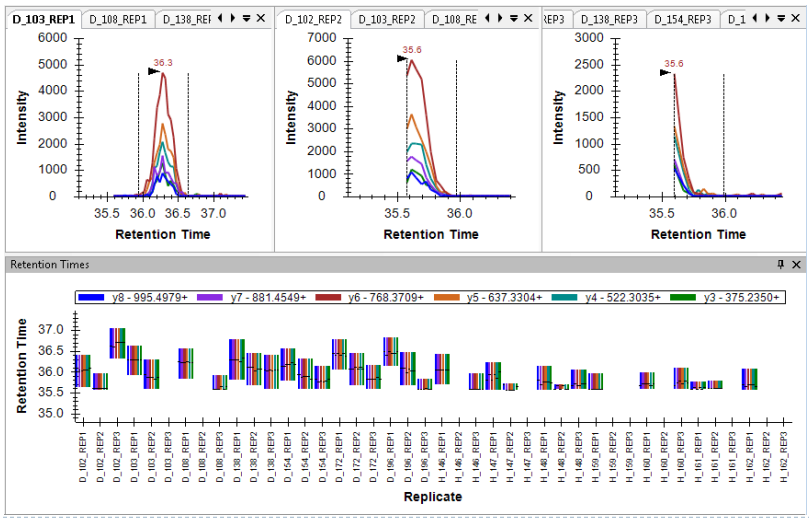
SVVDIGLIK

CSSLLWAGAAWLR

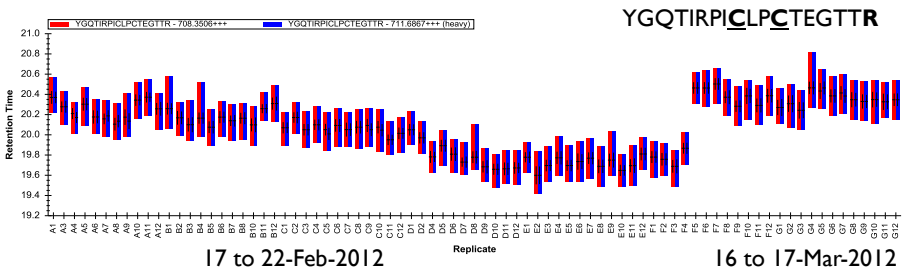


Truncated and Missing Peaks

TGTNLMDFLSR



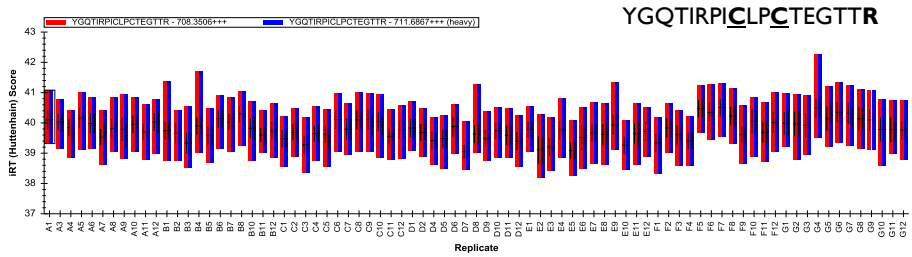
Deciphering the Unexpected



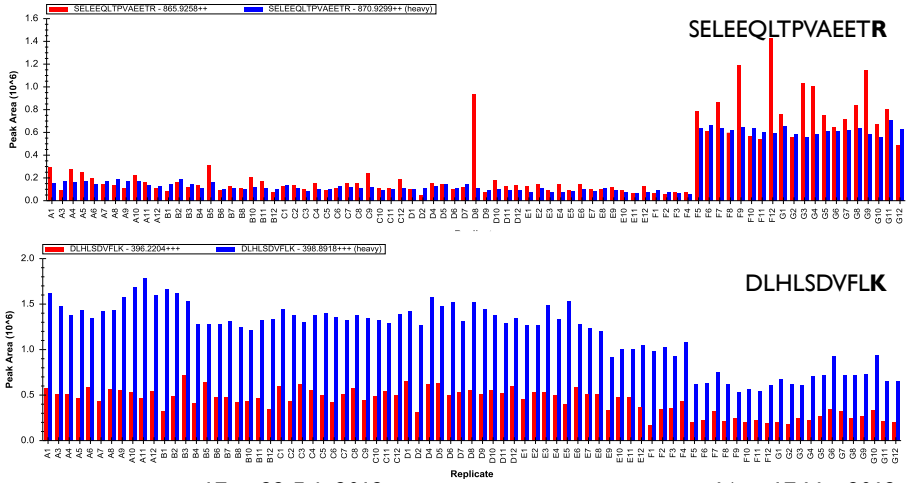
Results Grid

	Replicate Name	Acquired Time	Peptide Peak Found Ratio	Peptide Retention Time	Ratio To Standard	BioReplicate	Run	Condition
	F2	2/21/2012 9:02:23 PM	0.83	19.46	0.7895	59	59	Disease
	F3	2/21/2012 10:15:23 PM	0.83	19.39	1.1252	60	60	Disease
	F4	2/22/2012 12:41:17 AM	1	19.66	1.2937	61	61	Disease
	F5	3/16/2012 6:47:27 AM	1	20.13	1.2389	62	62	Healthy
	F6	3/16/2012 8:00:25 AM	1	20.18	0.9268	63	63	Healthy
	F7	3/16/2012 9:13:23 AM	1	20.17	1.3614	64	64	Healthy

Aligned by iRT



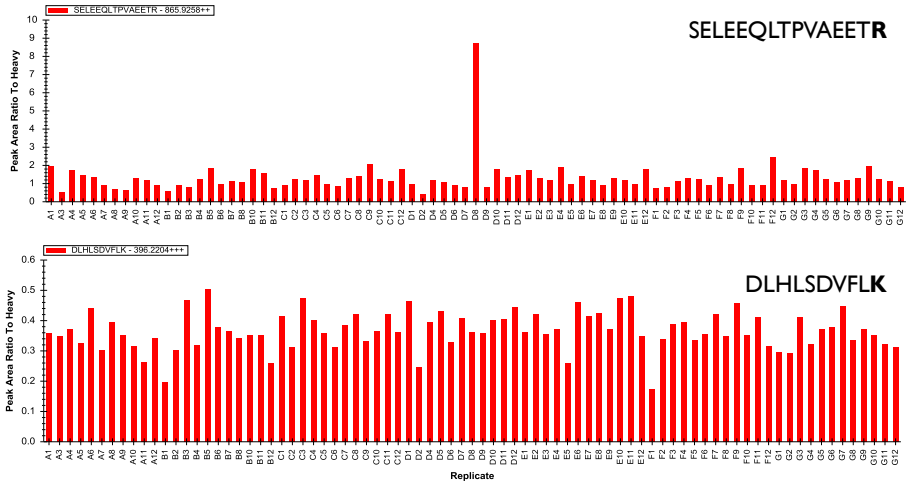
Global Normalization Problem



17 to 22-Feb-2012

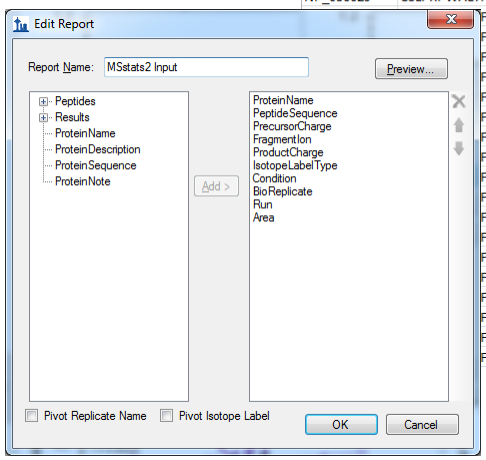
16 to 17-Mar-2012

Normalization by Reference Standard

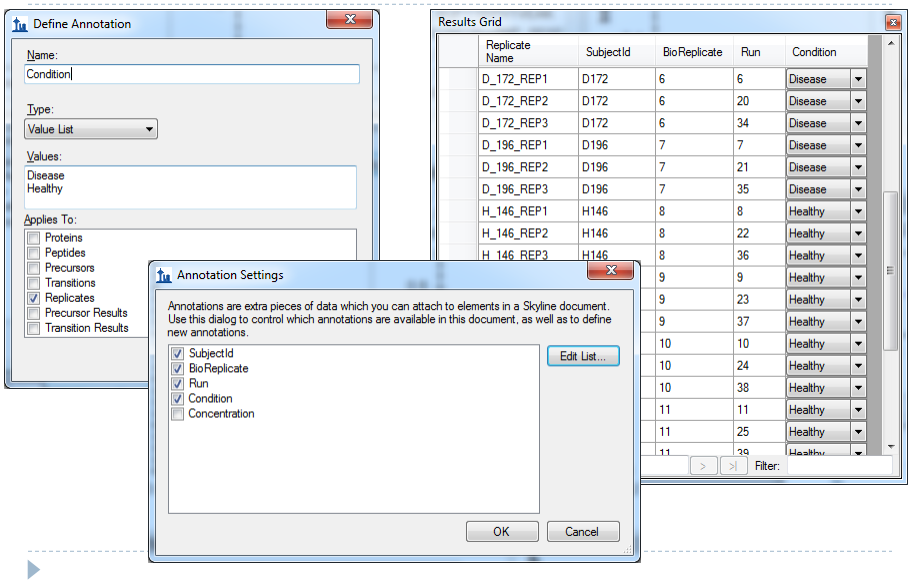


Custom Reports

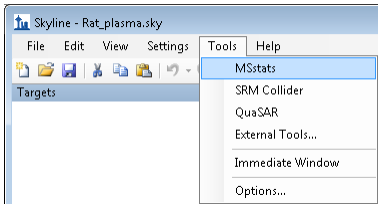
ProteinName	PeptideSequence	Precurs	Frag	Produ	IsotopeLa	Condition	Bio	Run	Area
NP_036629	CSLPRPWALTFSYGR	2	y10	1	light	Disease	1	1	14516
NP_036629	CSLPRPWALTFSYGR	2	y10	1	light	Disease	1	15	9607
NP_036629	CSLPRPWALTFSYGR	2	y10	1	light	Disease	1	29	7480
NP_036629	CSLPRPWALTFSYGR	2	y10	1	light	Disease	2	2	5692
NP_036629	CSLPRPWALTFSYGR	2	y10	1	light	Disease	2	16	5953
NP_036629	CSLPRPWALTFSYGR	2	y10	1	light	Disease	2	30	649
NP_036629	CSLPRPWALTFSYGR	2	y10	1	light	Disease	3	3	10476
NP_036629	CSLPRPWALTFSYGR	2	y10	1	light	Disease	3	17	3952
NP_036629	CSLPRPWALTFSYGR	2	y10	1	light	Disease	3	31	3165
NP_036629	CSLPRPWALTFSYGR	2	y10	1	light	Disease	4	4	9830
NP_036629	CSLPRPWALTFSYGR	2	y10	1	light	Disease	4	18	10671
NP_036629	CSLPRPWALTFSYGR	2	y10	1	light	Disease	4	32	6369
NP_036629	CSLPRPWALTFSYGR	2	y10	1	light	Disease	5	5	15037
NP_036629	CSLPRPWALTFSYGR	2	y10	1	light	Disease	5	19	9128
NP_036629	CSLPRPWALTFSYGR	2	y10	1	light	Disease	5	33	6918
NP_036629	CSLPRPWALTFSYGR	2	y10	1	light	Disease	6	6	11991
NP_036629	CSLPRPWALTFSYGR	2	y10	1	light	Disease	6	20	8630
NP_036629	CSLPRPWALTFSYGR	2	y10	1	light	Disease	6	34	6896
NP_036629	CSLPRPWALTFSYGR	2	y10	1	light	Disease	7	7	13061
NP_036629	CSLPRPWALTFSYGR	2	y10	1	light	Disease	7	21	12258
NP_036629	CSLPRPWALTFSYGR	2	y10	1	light	Disease	7	35	9037
NP_036629	CSLPRPWALTFSYGR	2	y10	1	light	Healthy	8	8	7891
NP_036629	CSLPRPWALTFSYGR	2	y10	1	light	Healthy	8	22	3362
NP_036629	CSLPRPWALTFSYGR	2	y10	1	light	Healthy	8	36	4448

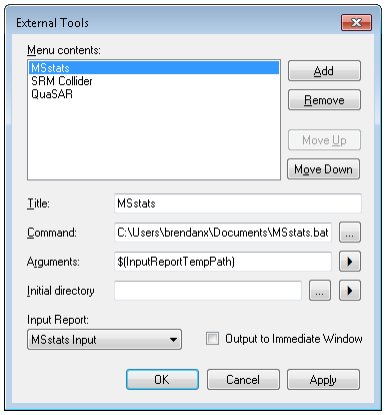


Replicate Annotations



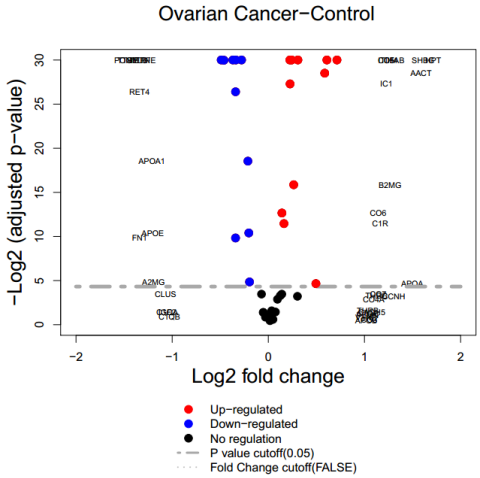
External Tool MSstats





Downstream Analysis with Statistical Tools

- Analysis of reports with R – MSstats



Full Cycle – Reusing Targeted Results

1. Create Folder
2. Users / Permissions
3. **Configure Targeted MS Folder**

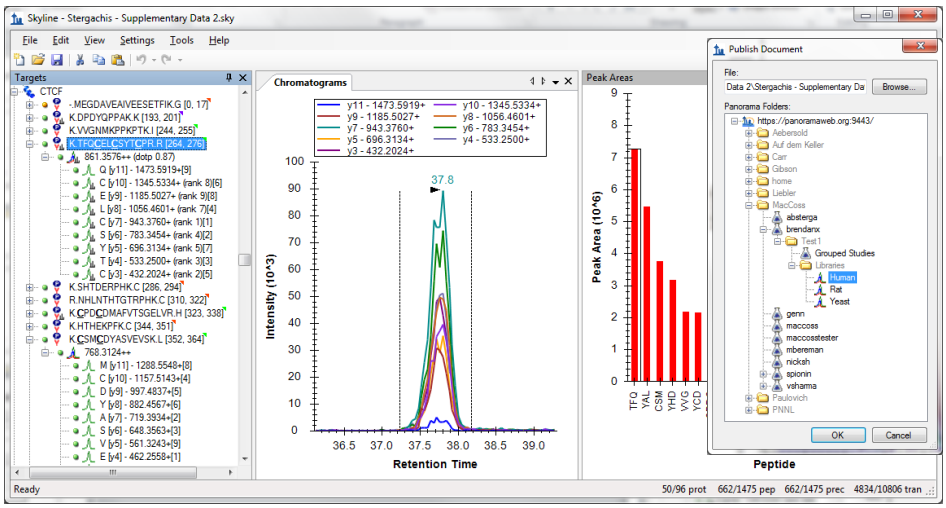
Panorama

Configure Targeted MS Folder

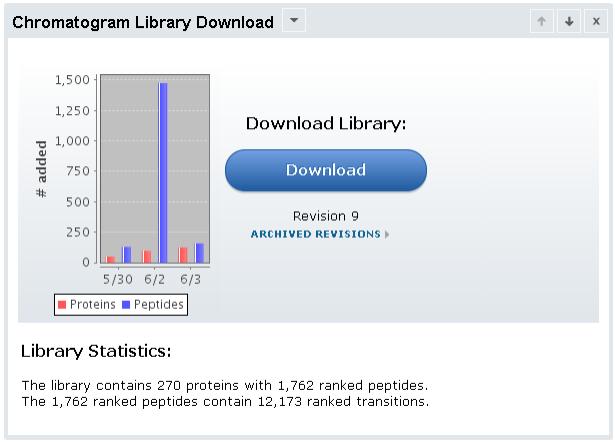
- **Experimental data** - a collection of published Skyline documents for various experimental designs
- **Chromatogram library** - curated precursor and product ion expression data for use in designing and validating future experiments
  - Rank peptides within proteins by peak area

FINISH

Publish to Panorama Library

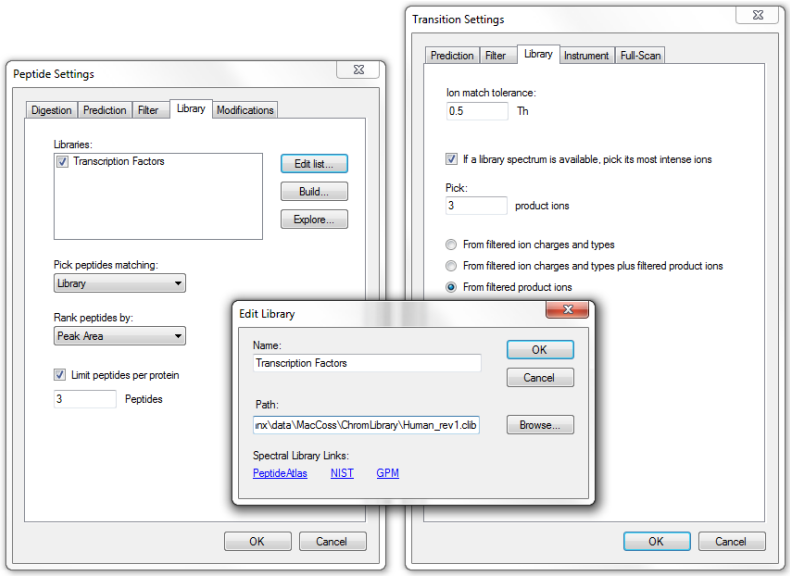


Download from Panorama

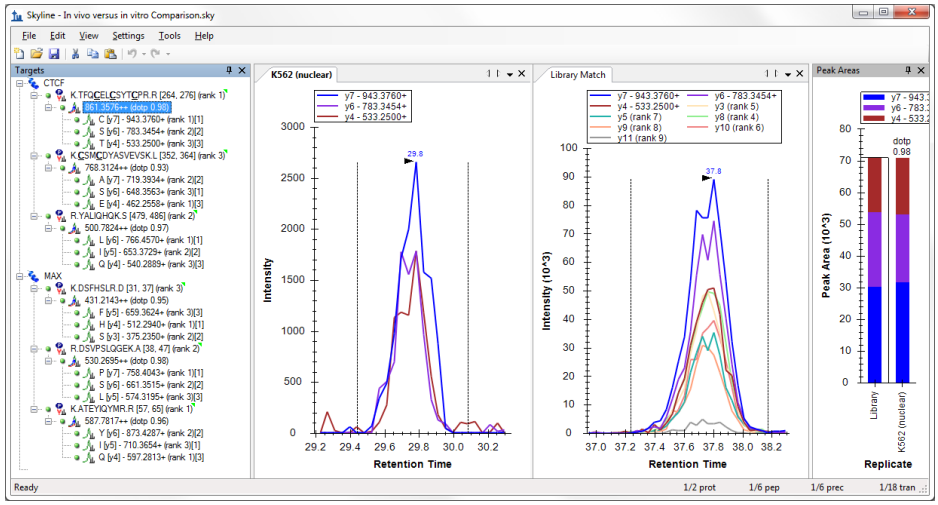


Panorama

Chromatogram Library in Skyline



Chromatogram Library in Skyline



Information Provided by Authors

- Information rich
  - Fully annotated Skyline documents
  - Custom report templates
  - Published to Panorama
- Do it yourself
  - Raw data
  - Transition lists
  - FASTA

III. List of discussion group leaders and participants for each Tier

Tier I – Clinical Discussion Leaders: Andy Hoofnagle, University of Washington and Russell Grant, Laboratory Corporation of America

Participants: Leigh Anderson. Hendrik Neubert, Elizabeth Mansfield, Josip Blonder, Nadir Rifai, James Ritchie, Dan Chan, Emily Boja, Julianne Botelho, Michael Boyne, Gordon Whitely, Sang Won Lee

Tier 2: Candidate Verification Discussion Leaders: Brad Ackerman, Eli Lilly and Susan Abbatiello, Broad

Participants: Christoph Borchers, Brad Ackerman, John Koomen, Tao Liu, Mandy Paulovich, Eric Kuhn, Hasmik Keshishian, R. Reid Townsend, DR Mani, Brendan MacLean, Jennifer Van Eyk, Arun Wiita

Tier 3: Biology Focus Discussion Leaders: Bruno Domon, Luxembourg CRP-Santé and Ruth Hüttenhain, UCSF

Participants: Dan Liebler, Olga Vitek, Lukas Reiter, Robert Moritz, Eric Deutsch, Sue Weintraub, Ralph Bradshaw, Steve Carr, Juan Oses, Jerry Lee