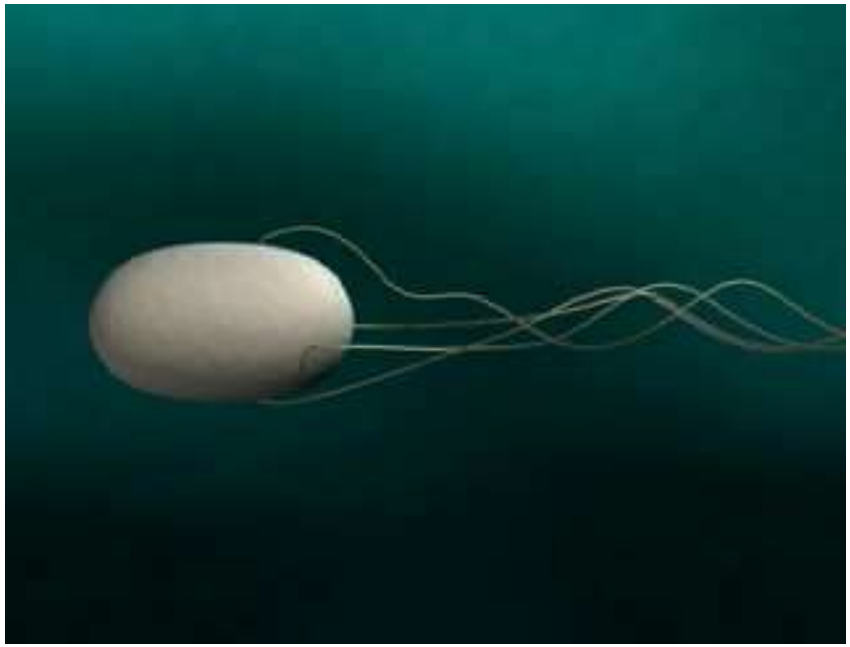
Targeted proteomics: informatics challenges before and after the measurements

Lars Malmström

IMSB, D-Biol



Perturbation

Fractionation

Modification

Genome sequencing

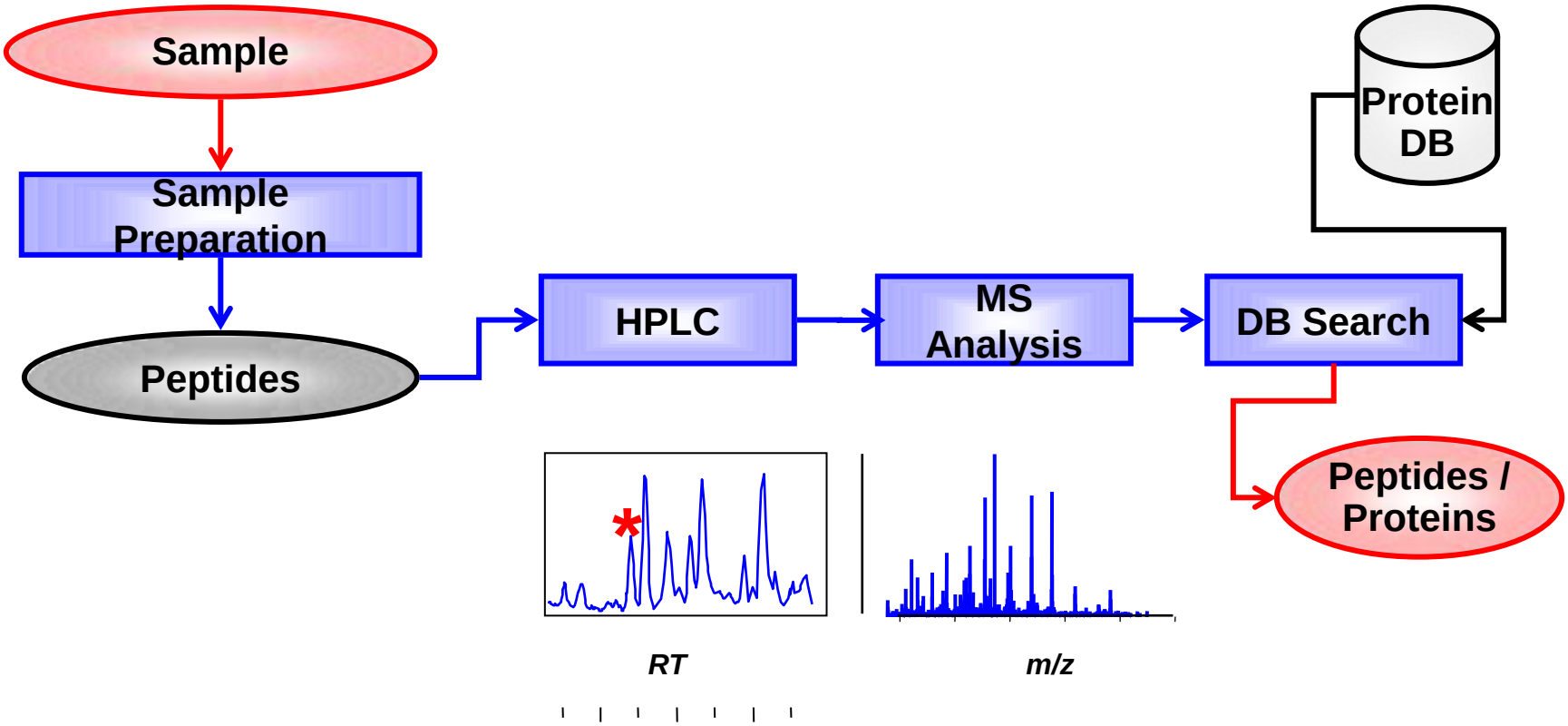
Mass spectrometry

Interpretation/Modeling: absolute component abundance, tertiary and quaternary protein structure, influence networks, interaction networks etc, etc.

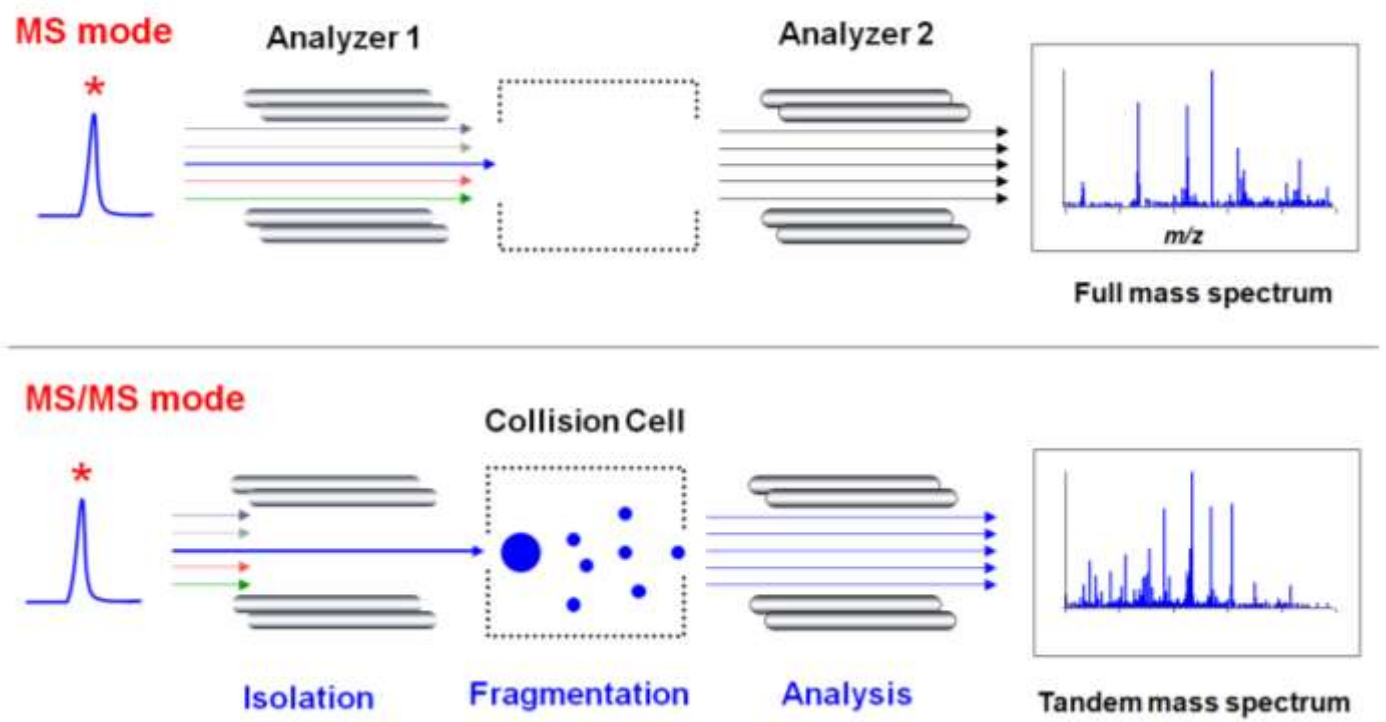
Data processing, data analysis, statistics and data integration

Data storage

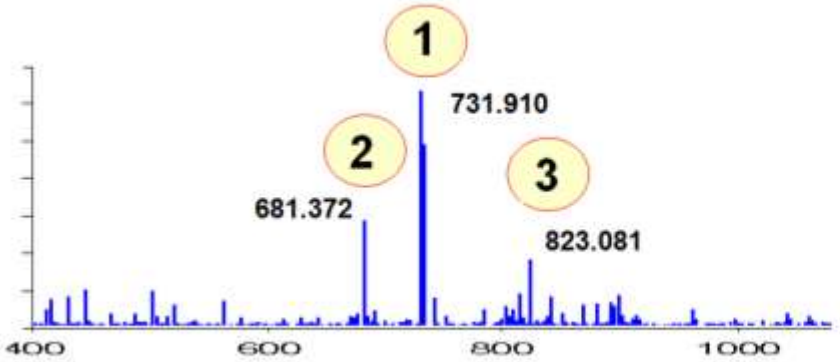
Peptide-based MS workflows can identify proteins

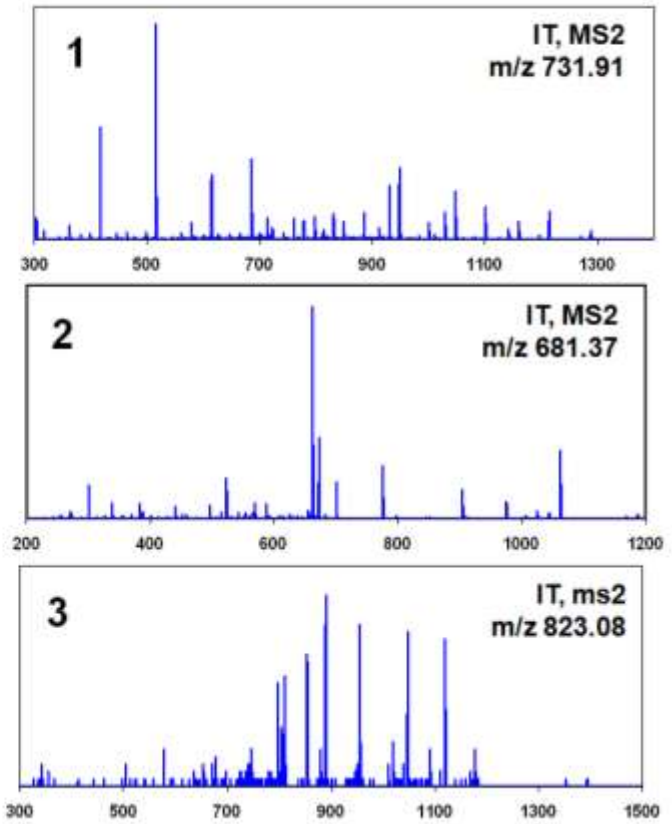


Isolated peptides are fragmented and measured in a second mass analyzer



High-intensity peptides are selected for fragmentation in Data Dependent Acquisition (DDA) MS





Protein list ← **Peptide list** ←

Redundancy in DDA MS data does not provide more information

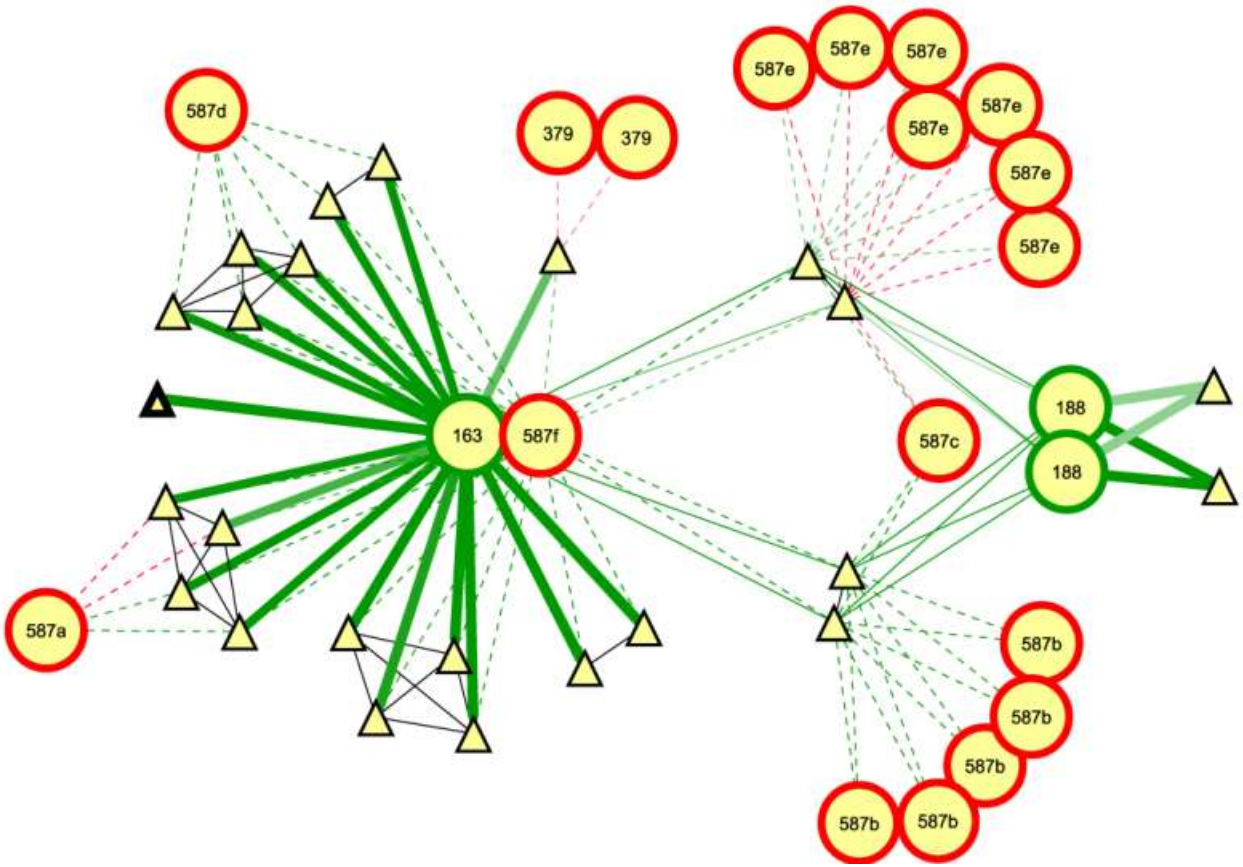
Peptide Accession	Pre AA	Peptide Sequence	Fol AA	Suitability Score	Best Prob	N Obs
PAp00001736	K	DLYANTVLSGGTTMYPGIADR	M	0.94	0.992	1355
PAp00004784	K	LCYVALDFEQEMATAASSSSLEK	S	0.86	0.992	1257
PAp00650061	R	TTGVVLDSGDGVTHTVPIYEGYALPHAILR	L	0.84	0.992	811
PAp00647657	K	SYELPDGQVITVGNER	F	0.81	0.992	773
PAp00008124	R	VAPEEHPVLLTEAPLNPK	A	0.83	0.992	763
PAp00004230	R	KDLYANTVLSGGTTMYPGIADR	M	0.52 [mc]	0.992	681
PAp00001808	K	DSYVGDEAQSK	R	0.68	0.992	356
PAp00003497	R	HQGVMVGMGQK	D	0.66	0.992	351
PAp00033507	K	EITALAPSTMK	I	0.65	0.992	306
PAp00653471	R	CPEAMFQPSFLGMESAGIHETSYNSIMK	C	0.62	1.000	287
PAp00000350	K	AGFAGDDAPR	A	0.68	0.992	280

Sequence

MSNVFIVGEL KDGELKKISR EITSAGRKIA DSIGGKVIAL LIGNGVEKHA PELAAVGADT ILTVNSGDYN
AETYSNQVAE VIKAQNPAVV LLPHTSQGKD YSPRVAVKVG AGIIADVVGF SVDGGKVVAK KPIYSGKAYA
NFKVTSPIQI FTVRPNSQEI TQKAGAGAVE ATSPSAGNAK VKIVSSDLSG GSKVQLTEAS IIVSGGRGIK
GPENWPILQE LADTLGAALG ASRAAVDAGW ISHSHQVGQT GKTVSPNCYI ACGISGAIQH LAGMGSSKYI
VAINKDGDAP IFKVATYGIV GDLFEVVPAV TAEFKKVLG

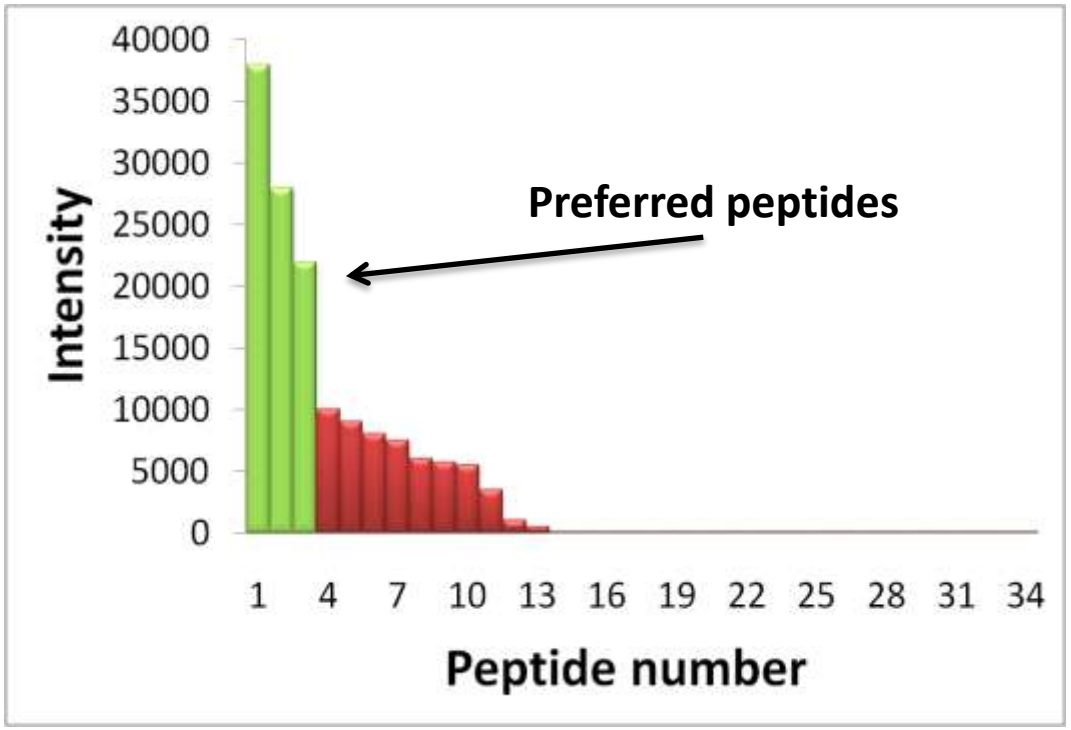
Protein Coverage = 85.8% (94.8% of likely observable sequence)

Some peptides are shared between multiple proteins



Some peptides are more easily detected by MS than others

- A *Proteotypic peptide* (PTP) is a peptide that can be detected with highest sensitivity and only match a single protein.
- A protein can produce many unique peptides when digested of which some are more easily detected in the MS.



Mallick et al. 2007 Nat Biotech

Targeted proteomics have some advantages over DDA

- Only measure proteins of interest
- Only measure the best PTPs from these proteins
- Only measure the most “relevant” fragments

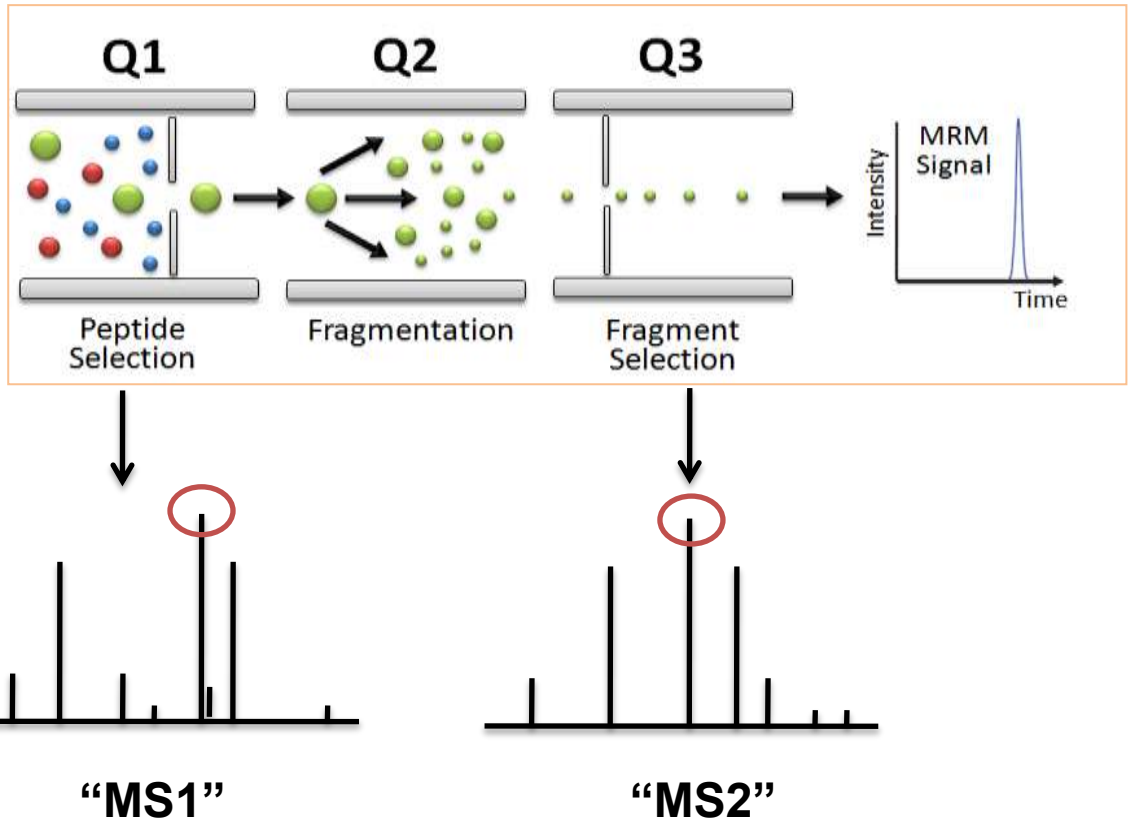
  $\Rightarrow$ Minimize redundancy and maximize the fraction relevant data

- Optimize the analysis for each peptide (collision energy, dwell time etc)

  $\Rightarrow$ Can achieve higher sensitivity

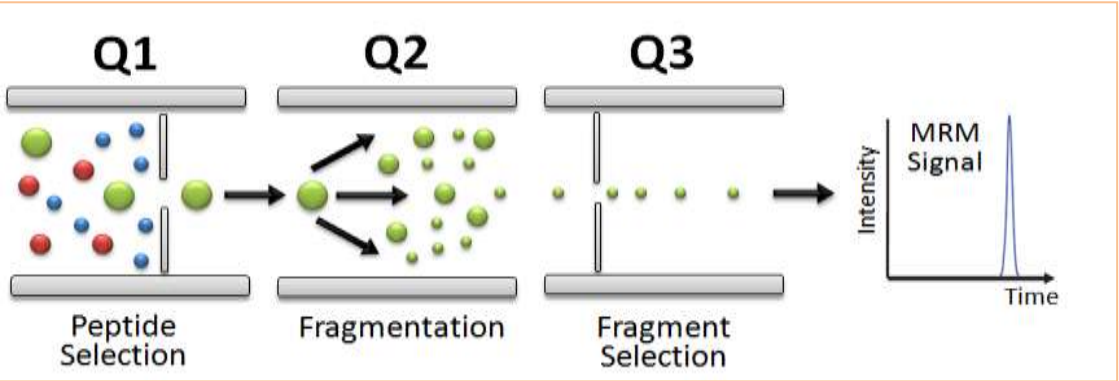
No discovery – have to know what you’re looking for.

Selected reaction monitoring (SRM) in a QQQ is one technology capable of targeted proteomics



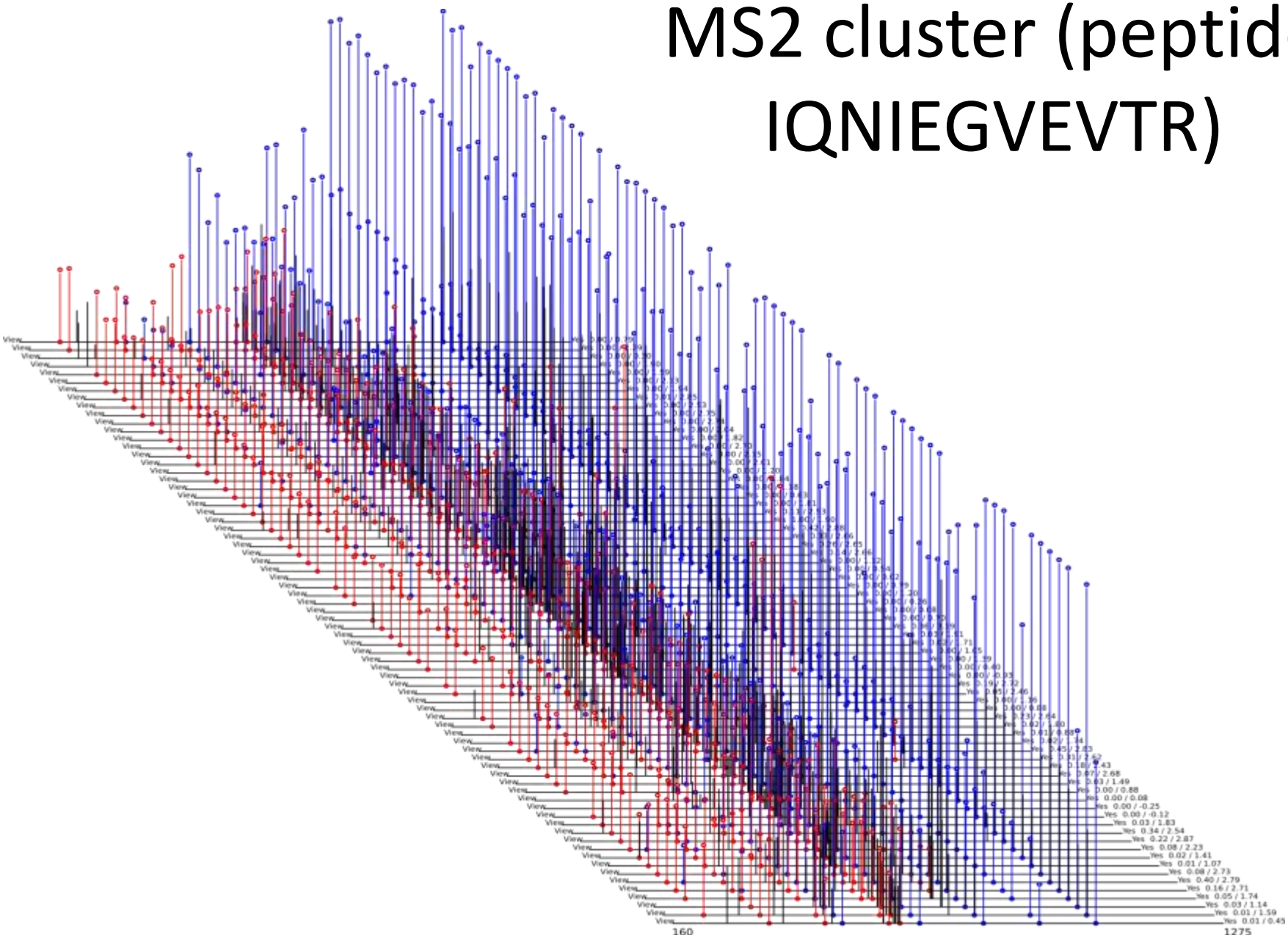
Thermo TSQ Quantum used for all measurements

SRM assays have to be developed and optimized

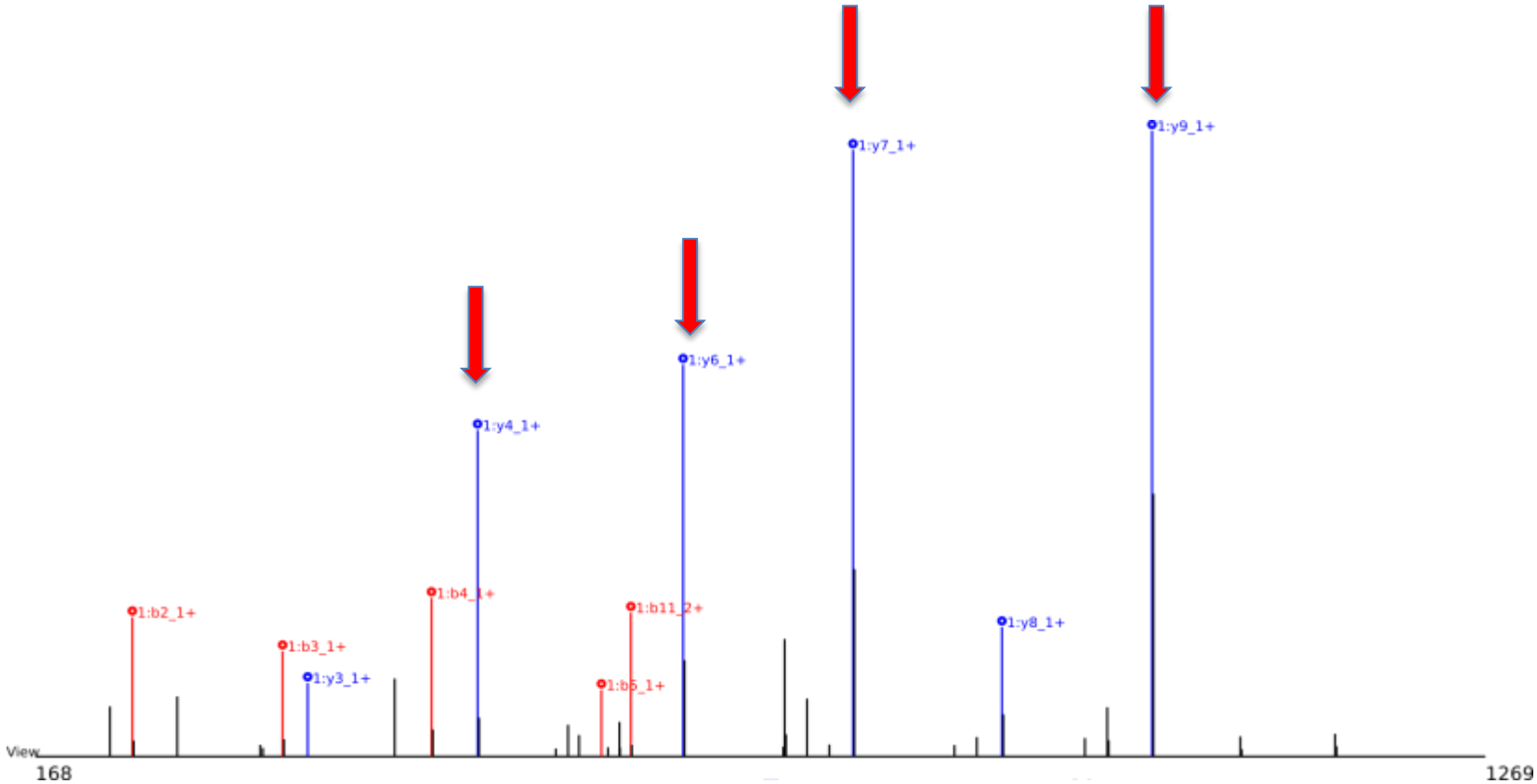


- *SRM assay* = set of optimal Q1/Q3 mass pairs for one peptide
- DDA data can assist in the selection of peptides and candidate Q1/Q3 mass pairs

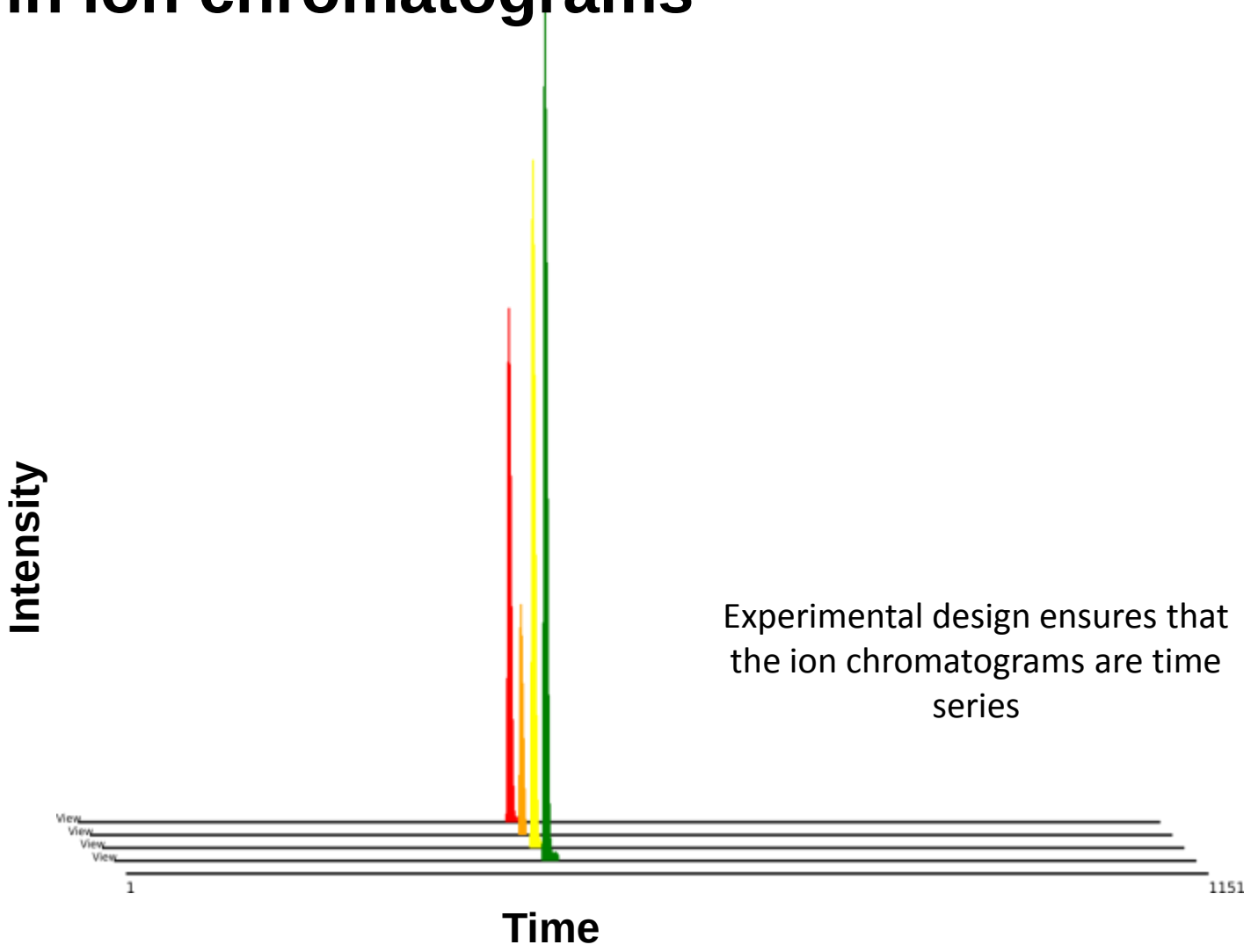
MS2 cluster (peptide IQNIEGVEVTR)



Consensus spectra of a cluster guides Q3 selection



The QQQ cycles through the Q1/Q3 pairs resulting in ion chromatograms



Experimental design ensures that the ion chromatograms are time series

Concepts of targeted proteomics

Targeted proteomics:

Analyze the proteins of interest – blind out irrelevant data

Optimize the analysis for each protein of interest once

Use the optimized settings for all subsequent experiments

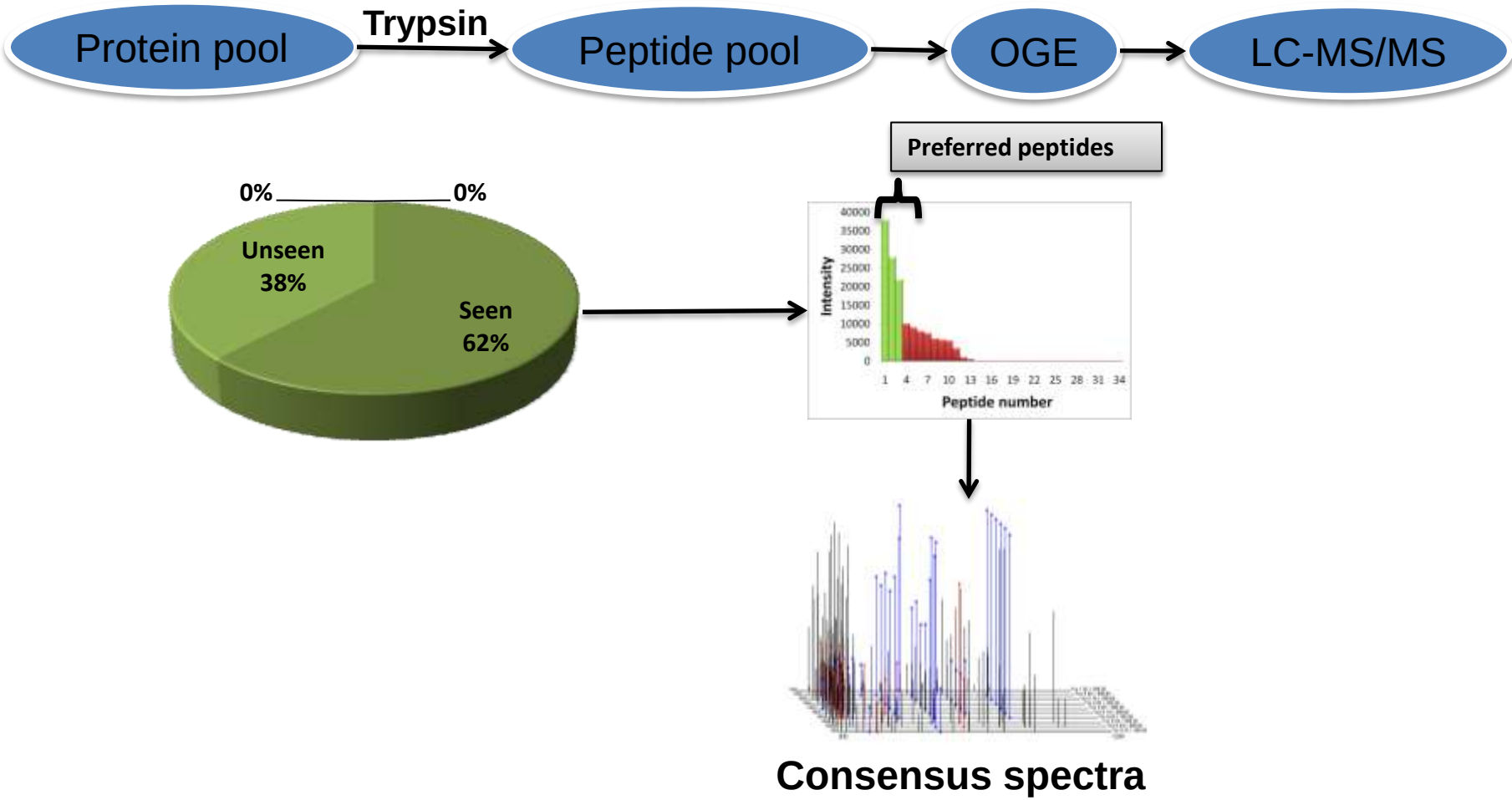
Benefits:

Highest achievable sensitivity and dynamic range

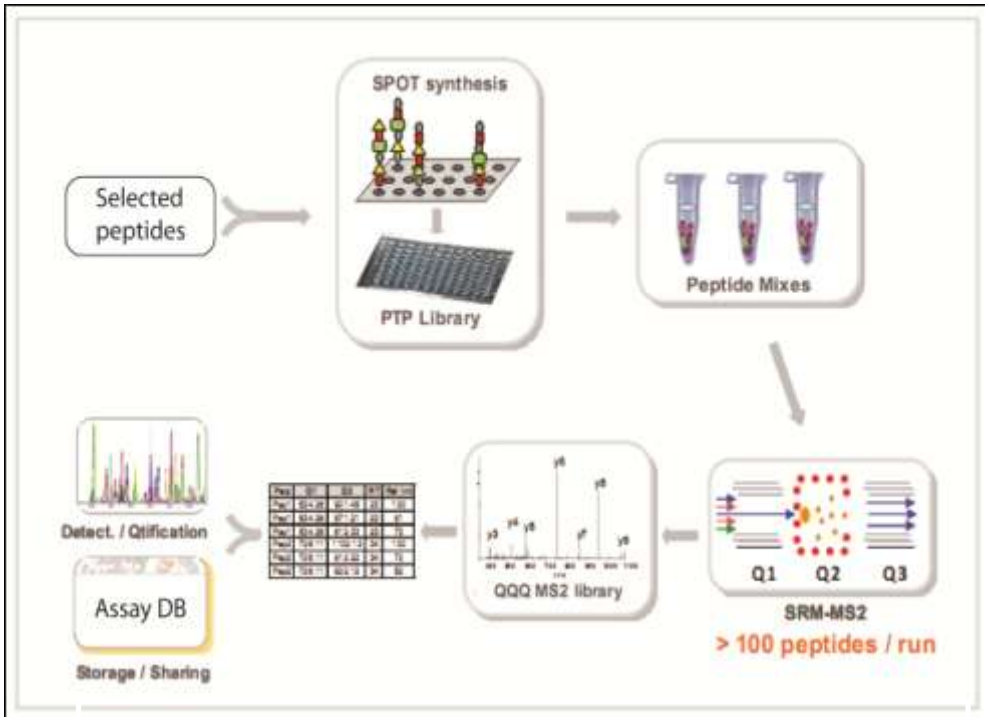
Very accurate and reproducible measurements

No random element as e.g. in shotgun proteomics – standardization, benchmarking possible

Peptide selection



Rapid generation of MRM assays with synthetic peptides



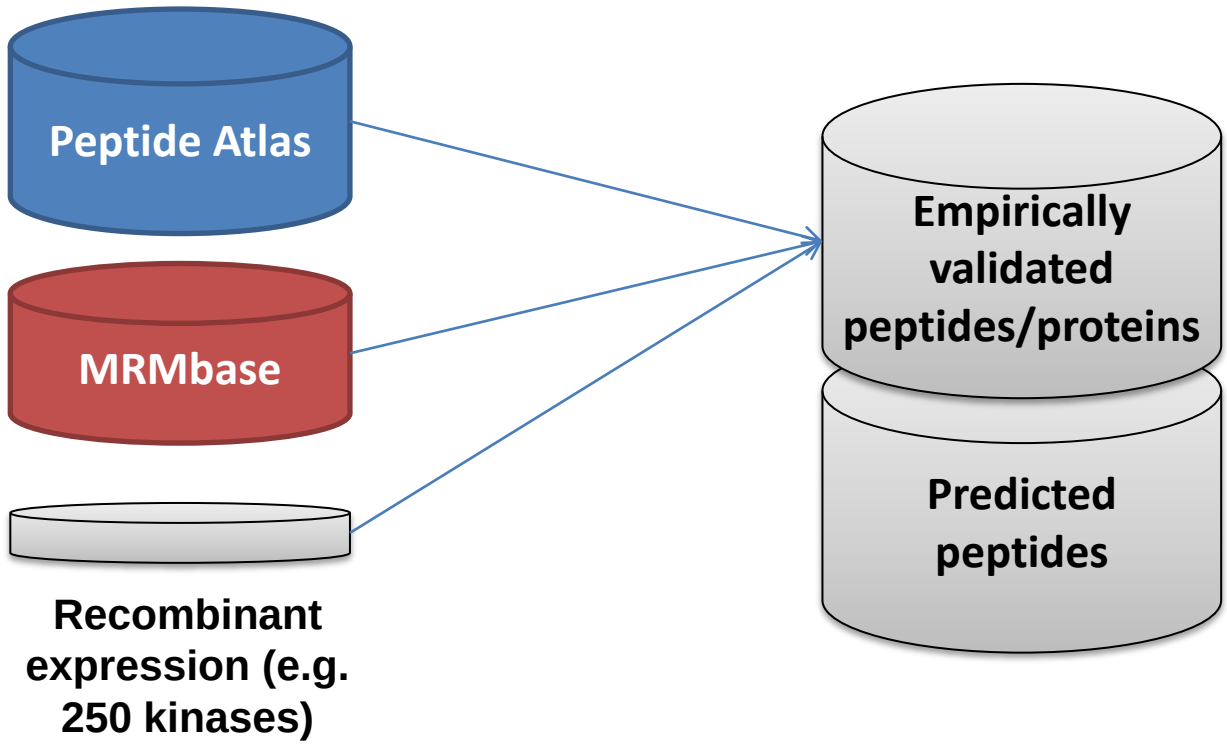
Picotti, Rinner…,Aebersold; Nat. Methods 2009

High-throughput generation of selected reaction-monitoring assays for proteins and proteomes

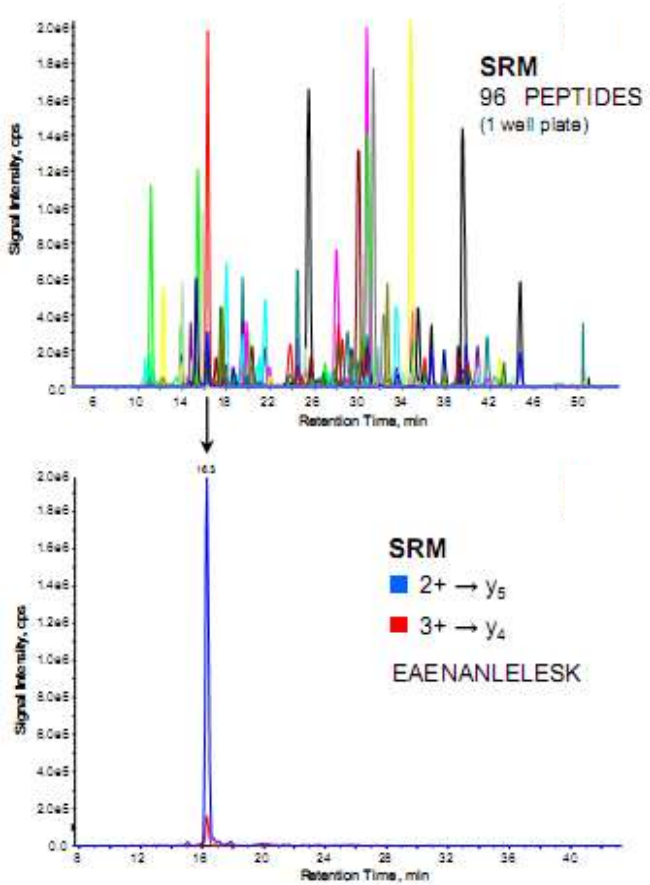
Paola Picotti[1], Oliver Rinner[1,2], Robert Stallmach[1], Franziska Dautel[3], Terry Farrah[4], Bruno Domon[1], Holger Wenschuh[5] & Ruedi Aebersold[1,4,6,7]

Selected reaction monitoring (SRM) uses sensitive and specific mass spectrometric assays to measure target analytes across multiple samples, but it has not been broadly applied in proteomics owing to the tedious assay development process for each protein. We describe a method based on crude synthetic peptide libraries for the high-throughput development of SRM assays. We illustrate the power of the approach by generating and applying validated SRM assays for all *Saccharomyces cerevisiae* kinases and phosphatases.

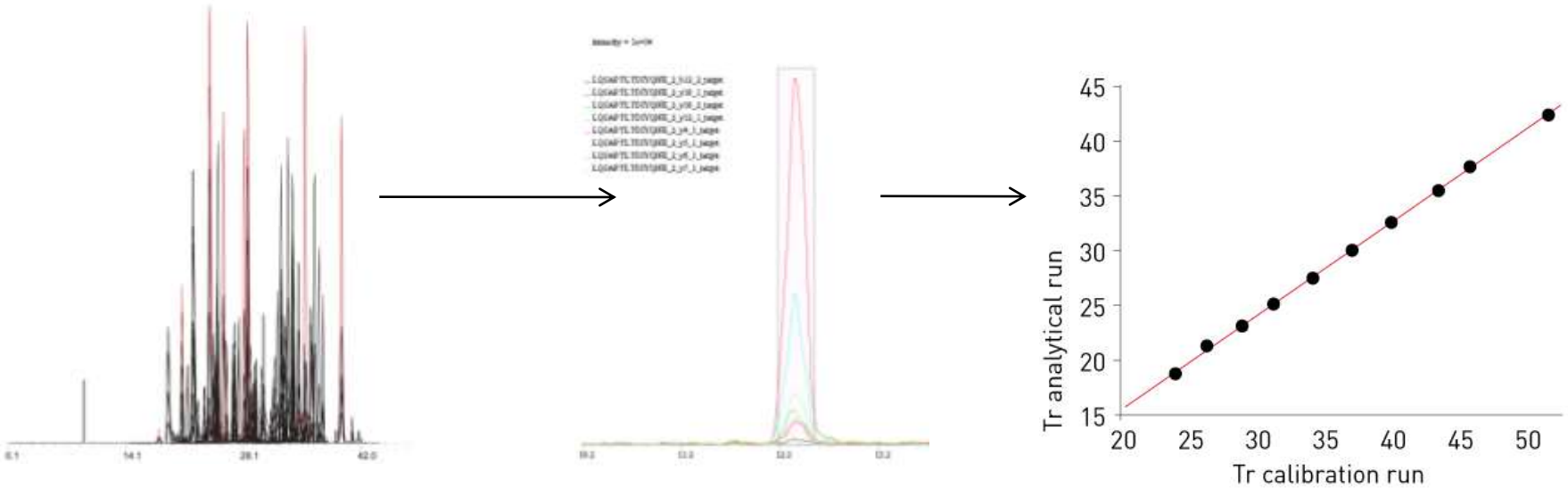
Finding the highflyers



Schedule SRM (Retention time calibration)
- multiplexed SRM



Retention time normalization – direct scheduled MRM across methods/machines



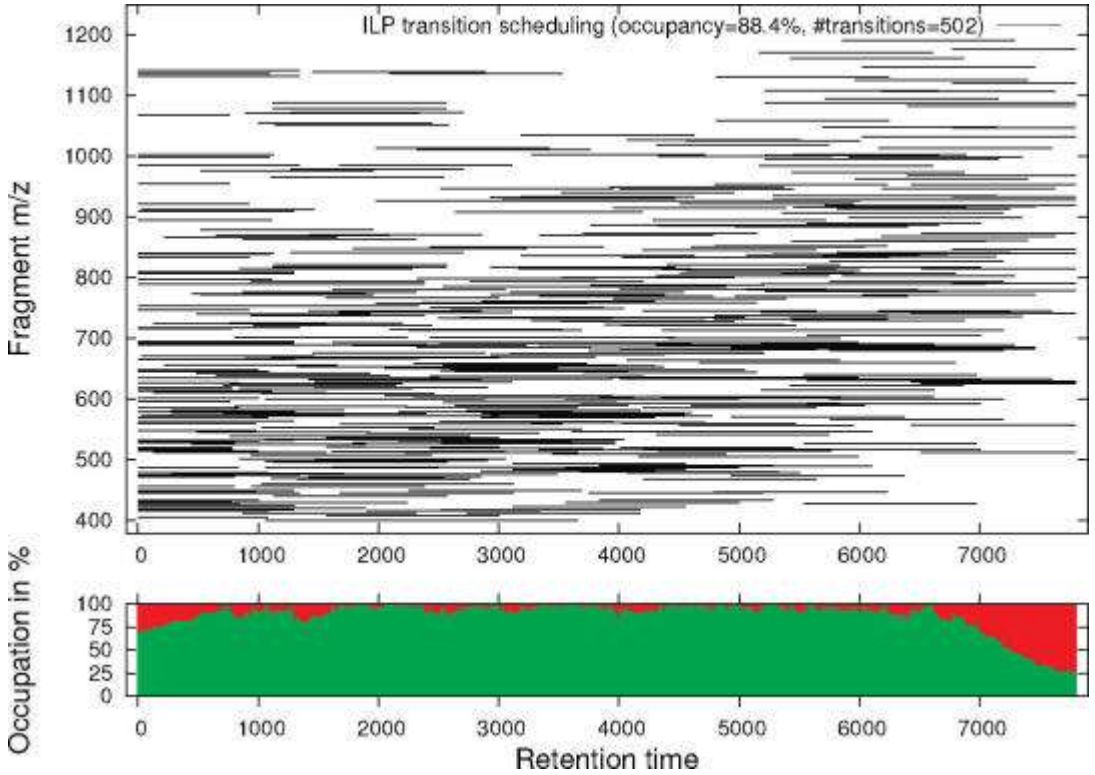
1.Synthetic peptides are run with RT normalization peptides (RT-kit)

2.Assay coordinates are recalculated in absolute RT space (iRT)

3.Retention times are normalized to analytical runs for all assays together

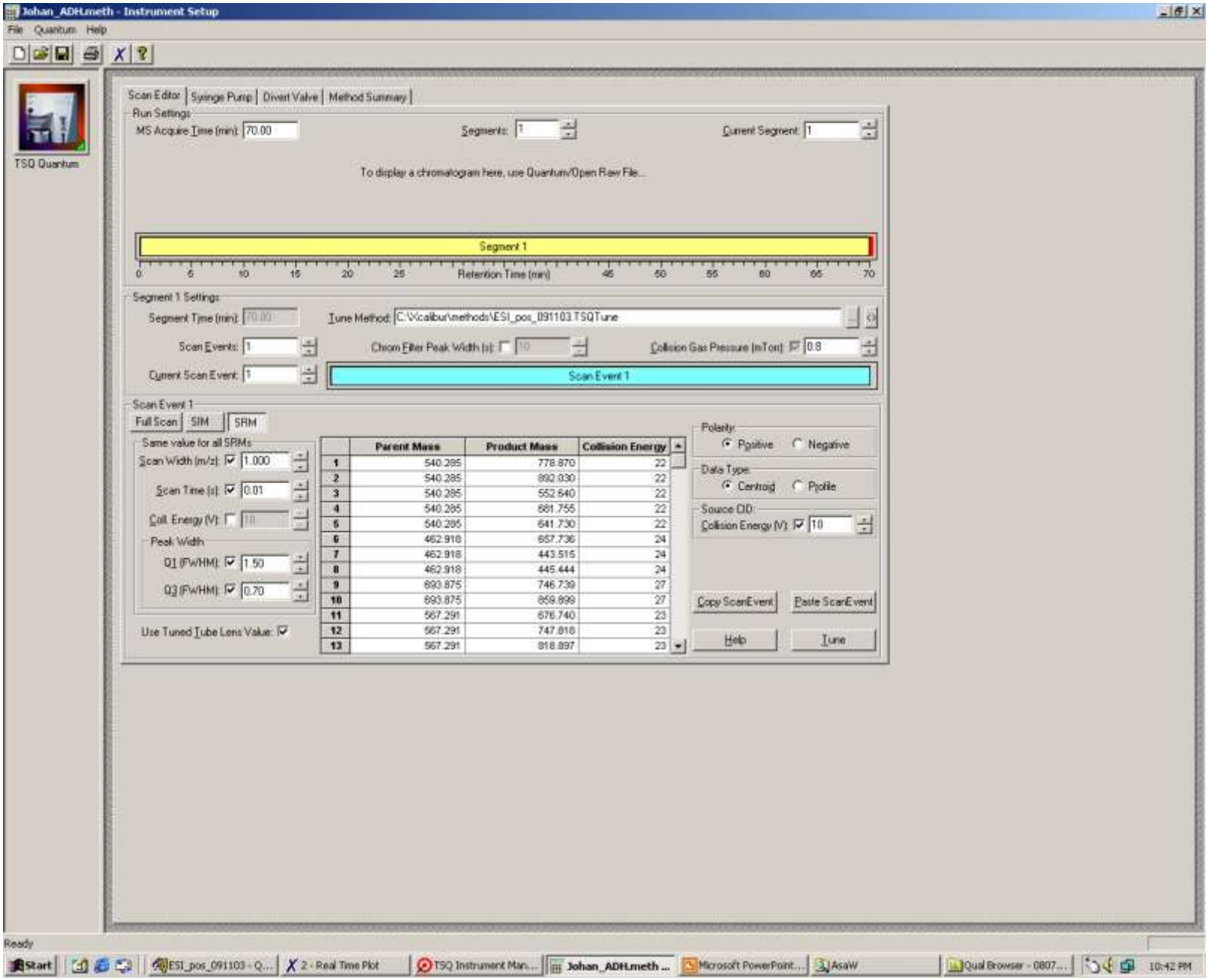
-> direct scheduled MRM

Scheduled SRM

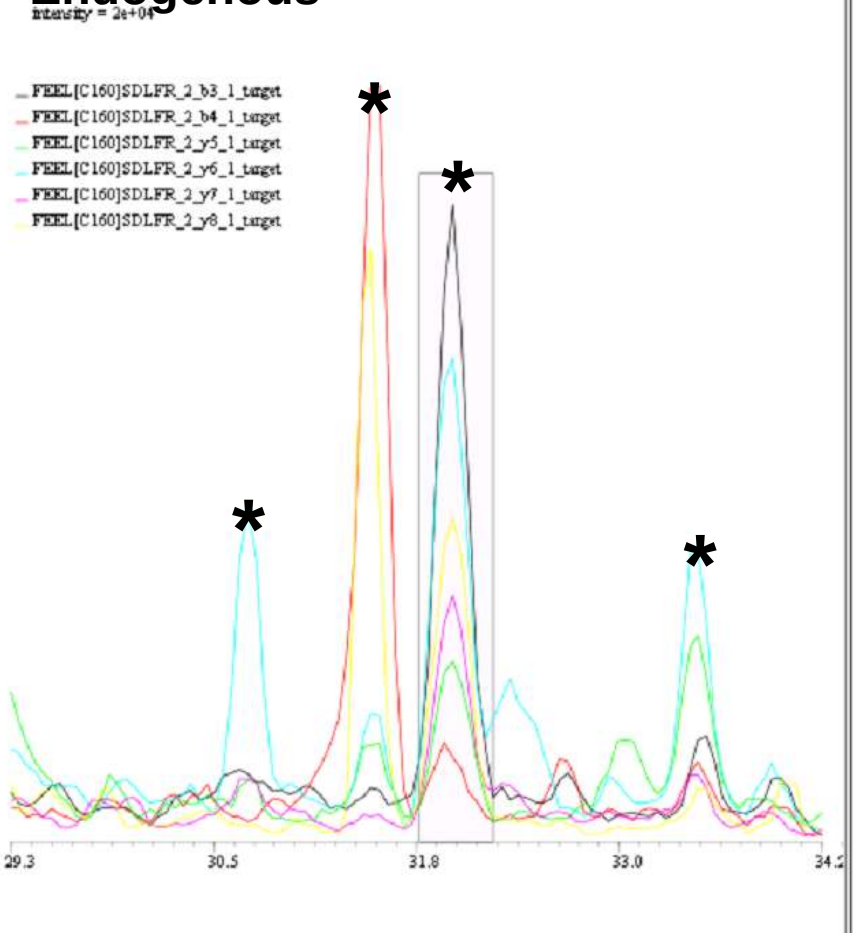


Transition list

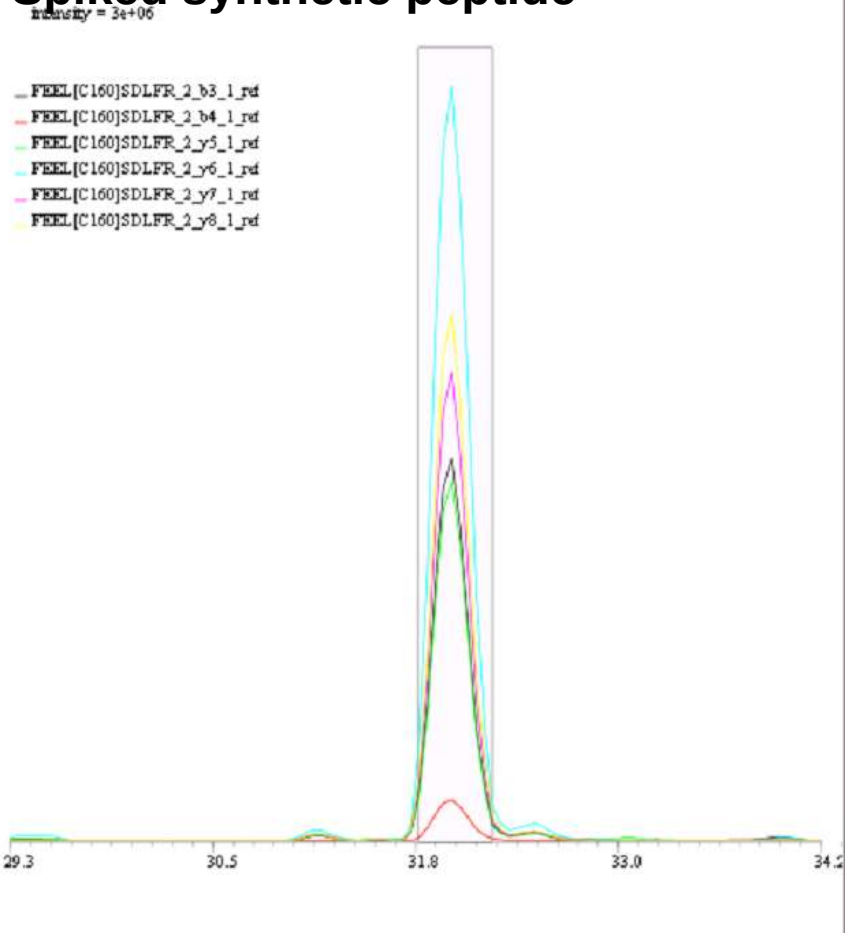
1	Q1	Q3	CE	desc
2	540.285	778.8722	21.68377602	STLPEIYEK__2_y6_1
3	540.285	892.0316	21.68377602	STLPEIYEK__2_y7_1
4	540.2849	552.64	21.68377602	STLPEIYEK__2_y4_1
5	540.285	681.7555	21.68377602	STLPEIYEK__2_y5_1
6	540.2851	641.7343	21.68377602	STLPEIYEK__2_b6_1
7	462.9185	657.7367	23.68251065	ANGTTVLVGMPAGAK__3_b7_1
8	462.9185	443.5153	23.68251065	ANGTTVLVGMPAGAK__3_y5_1
9	462.9185	445.4447	23.68251065	ANGTTVLVGMPAGAK__3_b5_1
10	693.8754	746.7398	26.90581953	ANGTTVLVGMPAGAK__2_y8_1
11	693.8747	859.8992	26.90581953	ANGTTVLVGMPAGAK__2_y9_1
12	567.2931	676.7395	22.60207811	VSEAAIEASTR__2_y6_1
13	567.2931	747.8183	22.60207811	VSEAAIEASTR__2_y7_1
14	567.2941	818.8971	22.60207811	VSEAAIEASTR__2_y8_1
15	567.2945	948.0126	22.60207811	VSEAAIEASTR__2_y9_1
16	567.3036	1035.091	22.60207811	VSEAAIEASTR__2_y10_1
17	418.7294	723.7957	17.55089446	IGDYAGIK__2_y7_1
18	418.7294	666.7438	17.55089446	IGDYAGIK__2_y6_1
19	418.7294	551.6552	17.55089446	IGDYAGIK__2_y5_1
20	418.7293	520.5547	17.55089446	IGDYAGIK__2_b5_1
21	418.7294	690.766	17.55089446	IGDYAGIK__2_b7_1
22	406.2141	549.5936	17.12535692	YVVDTSK__2_y5_1
23	406.2134	450.461	17.12535692	YVVDTSK__2_y4_1
24	406.2136	648.7262	17.12535692	YVVDTSK__2_y6_1
25	406.2135	578.6349	17.12535692	YVVDTSK__2_b5_1
26	406.2136	477.5298	17.12535692	YVVDTSK__2_b4_1
27	472.7568	732.7602	19.38780362	VLGIDGGEGK__2_y8_1
28	472.7568	562.5489	19.38780362	VLGIDGGEGK__2_y6_1
29	472.7568	845.9196	19.38780362	VLGIDGGEGK__2_y9_1


Johan_ADH.meth - Instrument Setup
File Quantum Help
TSQ Quantum
Scan Editor | Syringe Pump | Divert Valve | Method Summary
Run Settings
MS Acquire Time (min): 70.00
Segments: 1
Current Segment: 1
To display a chromatogram here, use Quantum/Open Raw File...
Segment 1
0 5 10 15 20 25 Retention Time (min) 45 50 55 60 65 70
Segment 1 Settings
Segment Time (min): 70.00
Tune Method: C:\Xcalibur\methods\ESI_pos_091103.TSQTune
Scan Events: 1
Chrom Filter Peak Width (s): 10
Collision Gas Pressure (mTorr): 0.8
Current Scan Event: 1
Scan Event 1
Scan Event 1
Full Scan | SIM | SRM
Same value for all SRMs
Scan Width (m/z): 1.000
Scan Time (s): 0.01
Coll. Energy (V): 10
Peak Width
Q1 (FWHM): 1.50
Q3 (FWHM): 0.70
Use Tuned Tube Lens Value:
Parent Mass | Product Mass | Collision Energy
1 540.285 778.870 22
2 540.285 892.030 22
3 540.285 552.640 22
4 540.285 681.755 22
5 540.285 641.730 22
6 462.918 657.736 24
7 462.918 443.515 24
8 462.918 445.444 24
9 693.875 746.739 27
10 693.875 859.899 27
11 567.291 676.740 23
12 567.291 747.818 23
13 567.291 818.897 23
Polarity
Positive Negative
Data Type
Centroid Profile
Source CID
Collision Energy (V): 10
Copy ScanEvent
Paste ScanEvent
Help
Tune
Ready
Start
ESI_pos_091103 - Q...
2 - Real Time Plot
TSQ Instrument Man...
Johan_ADH.meth ...
Microsoft PowerPoint...
AsaW
Qual Browser - 0807...
10:42 PM

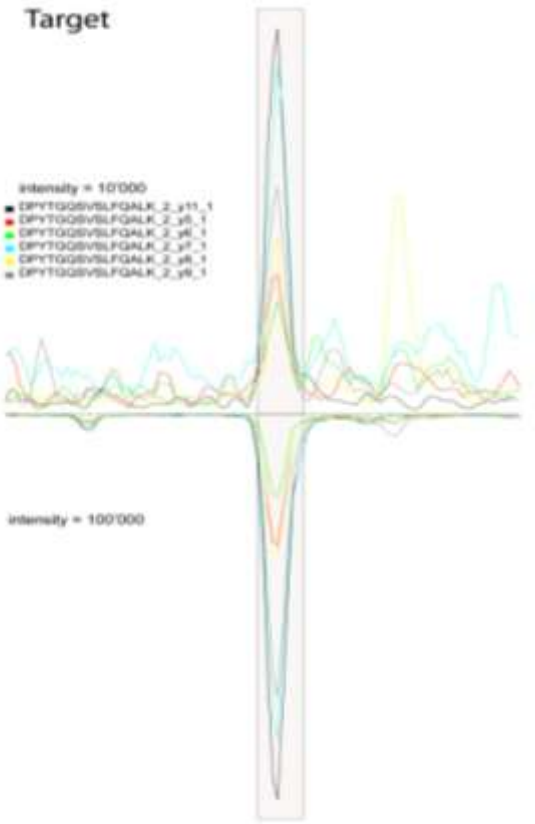
Endogenous

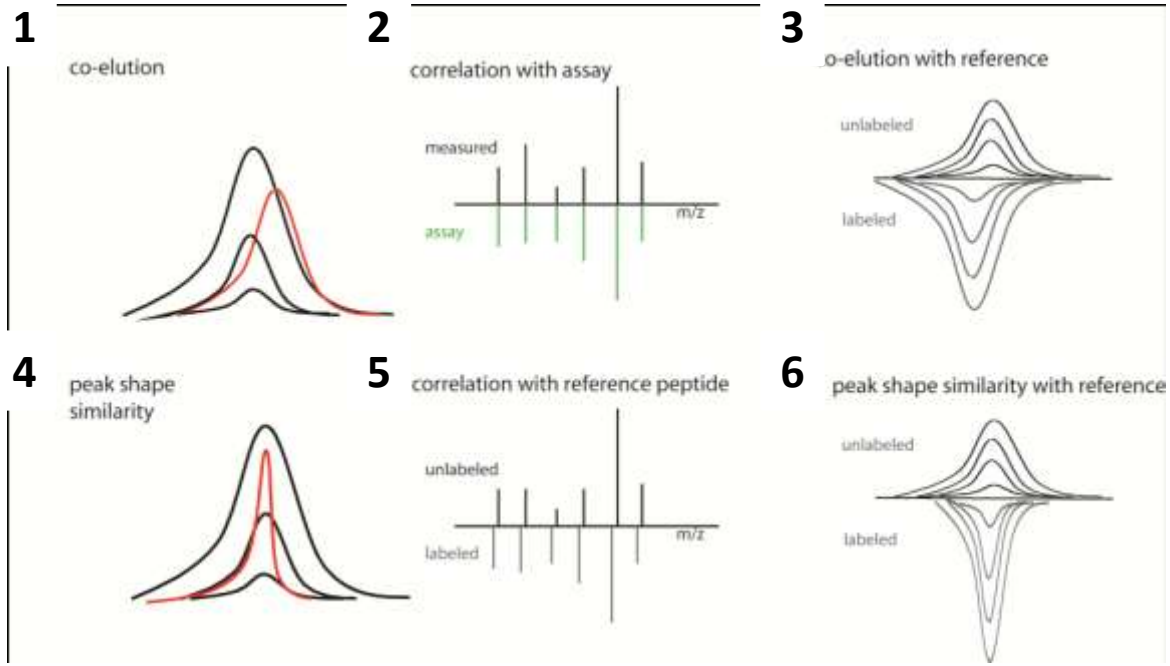


Spiked synthetic peptide

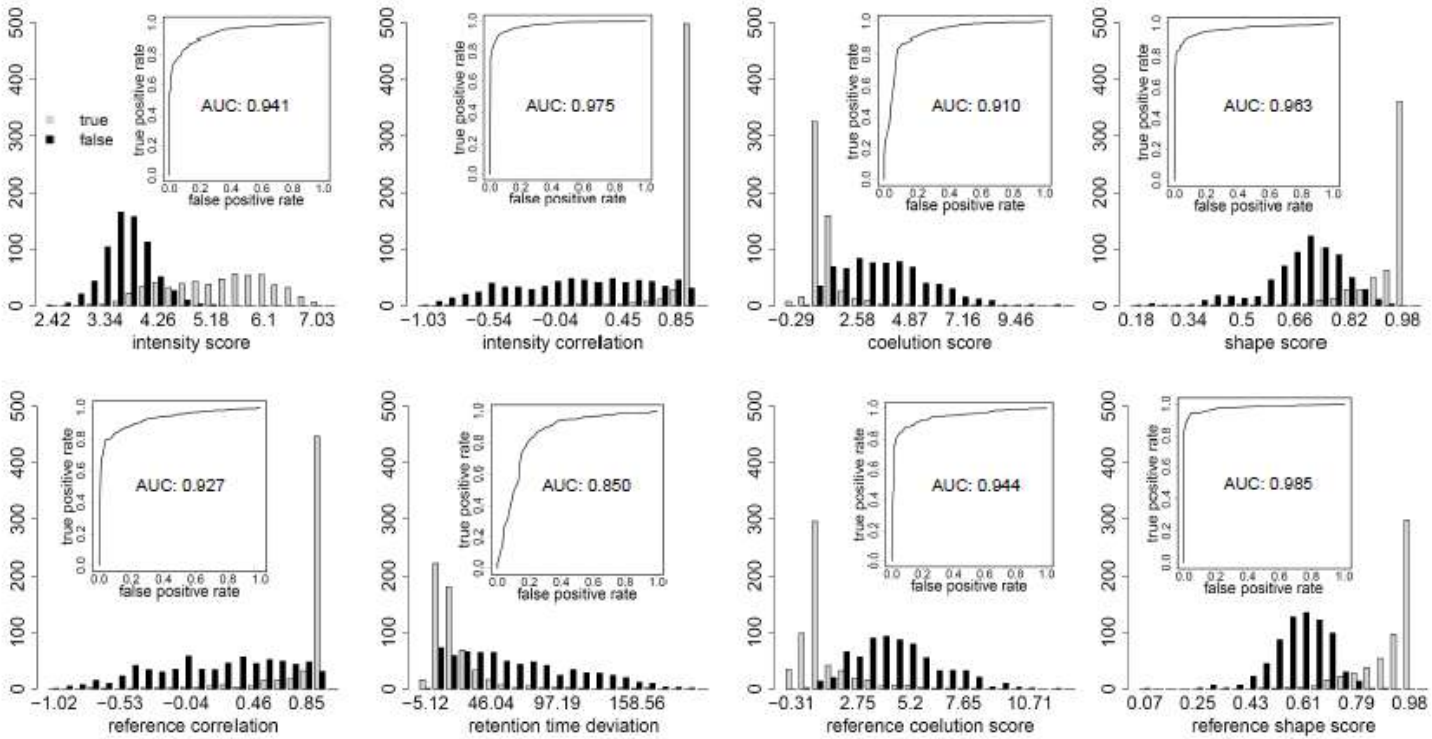


A comprehensive scoring model for sensitive and accurate scoring of MRM data

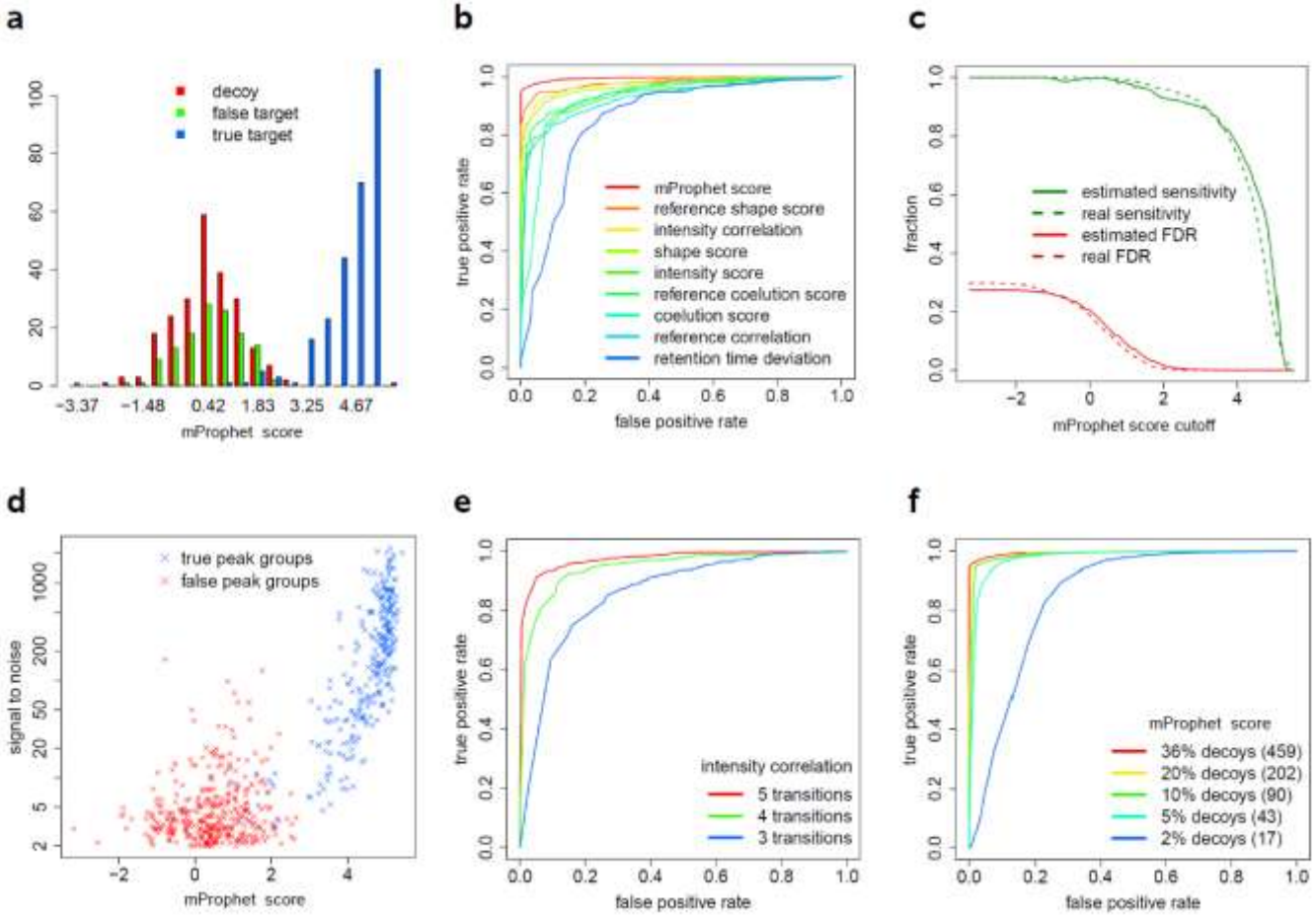




Separation power of the uncombined scores

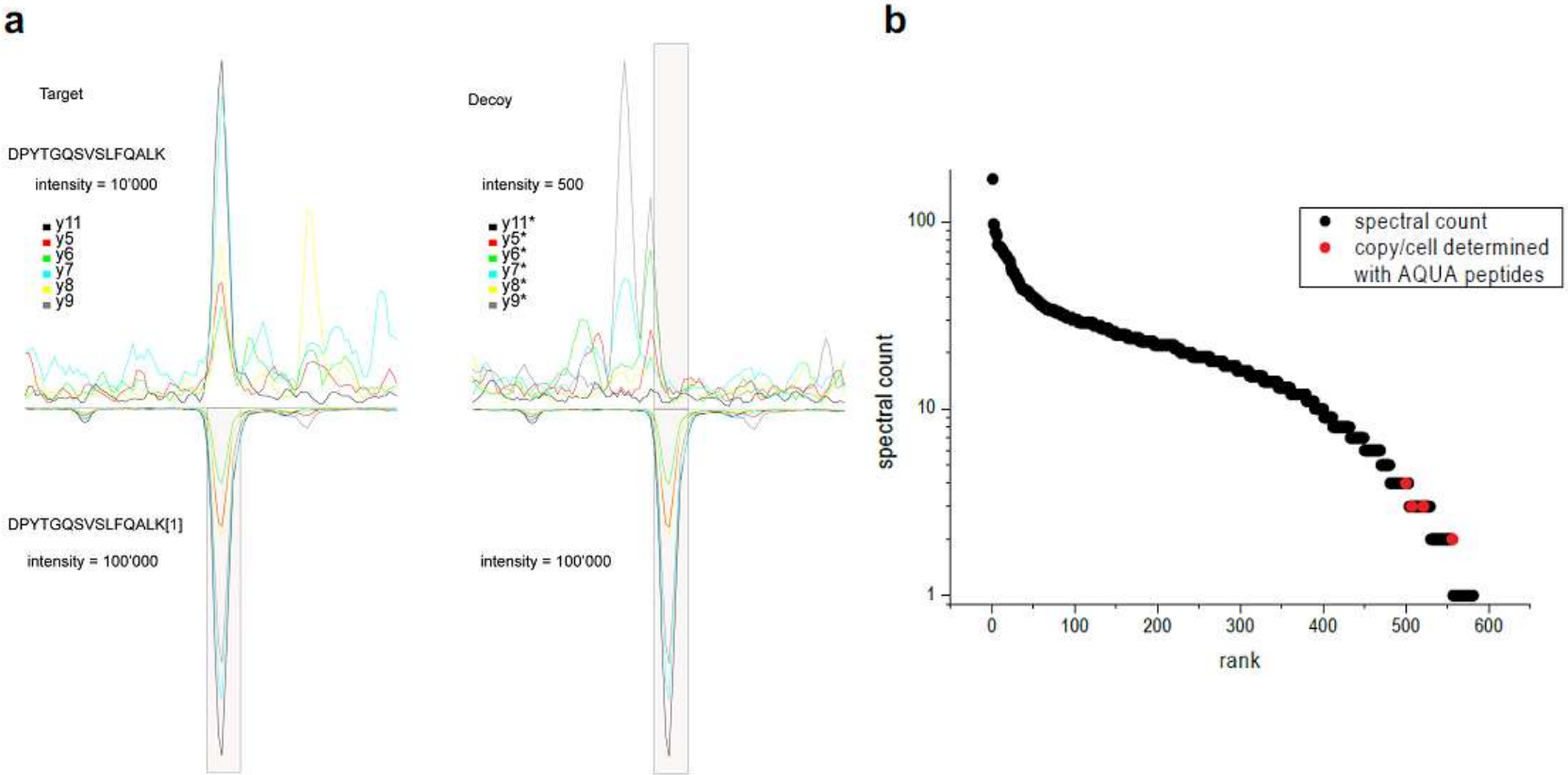


Training mQuest/mProphet with a standard data set

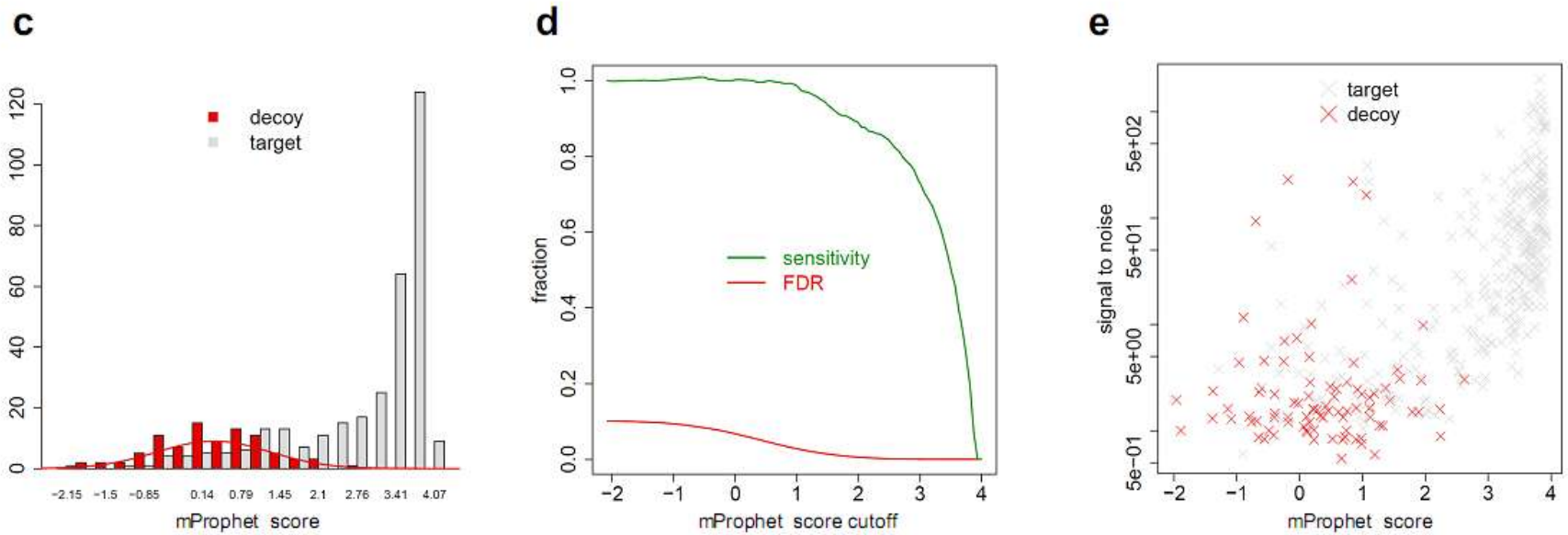


Reiter, Rinner, {…}, Aebersold (2011)

Verification of endogenous peptides in unfractionated lysates

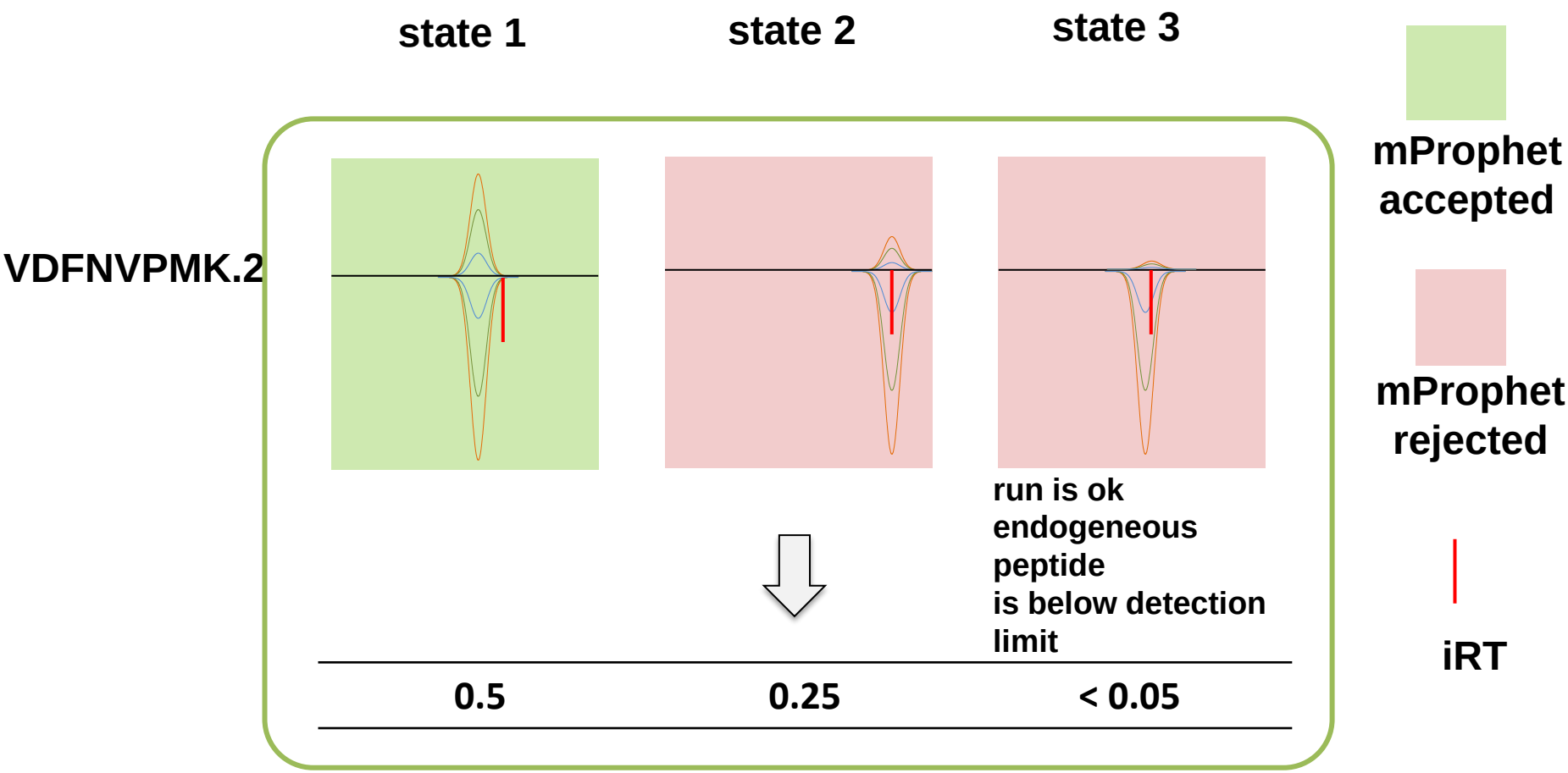


Excellent separation of decoy signals and true identifications



Decoy transitions are transitions for non-existing peptides – serving as null hypothesis

Quantification accross conditions

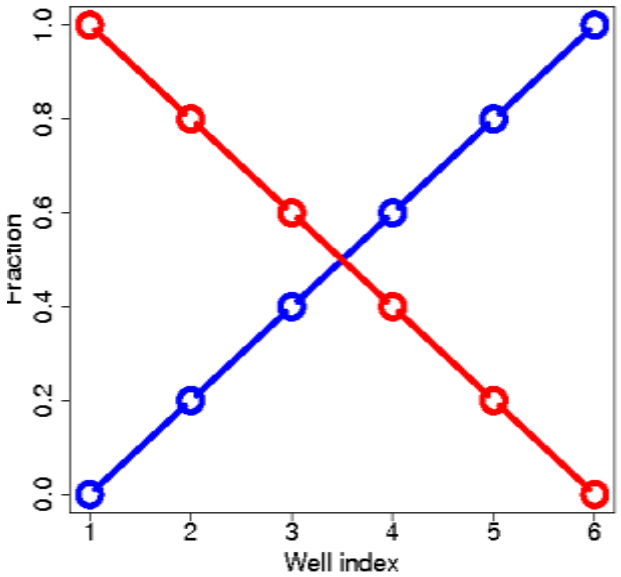


ATAQS and Skyline

- Designed to facilitate the experimentalist from design to data analysis
- Skyline is a windows-application
- ATAQS is a web-based application

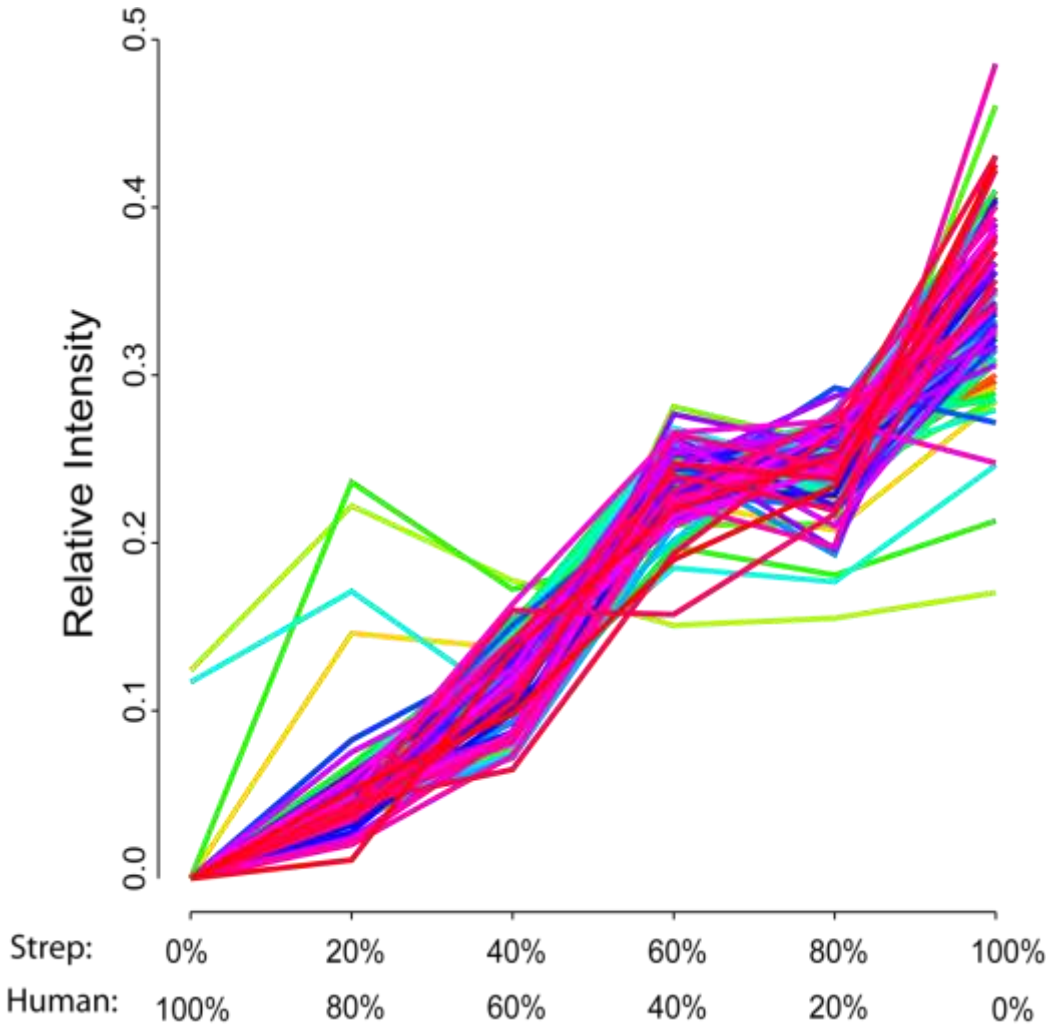
SRM label-free relative quantification demonstrated by a dilution experiment

Experimental design



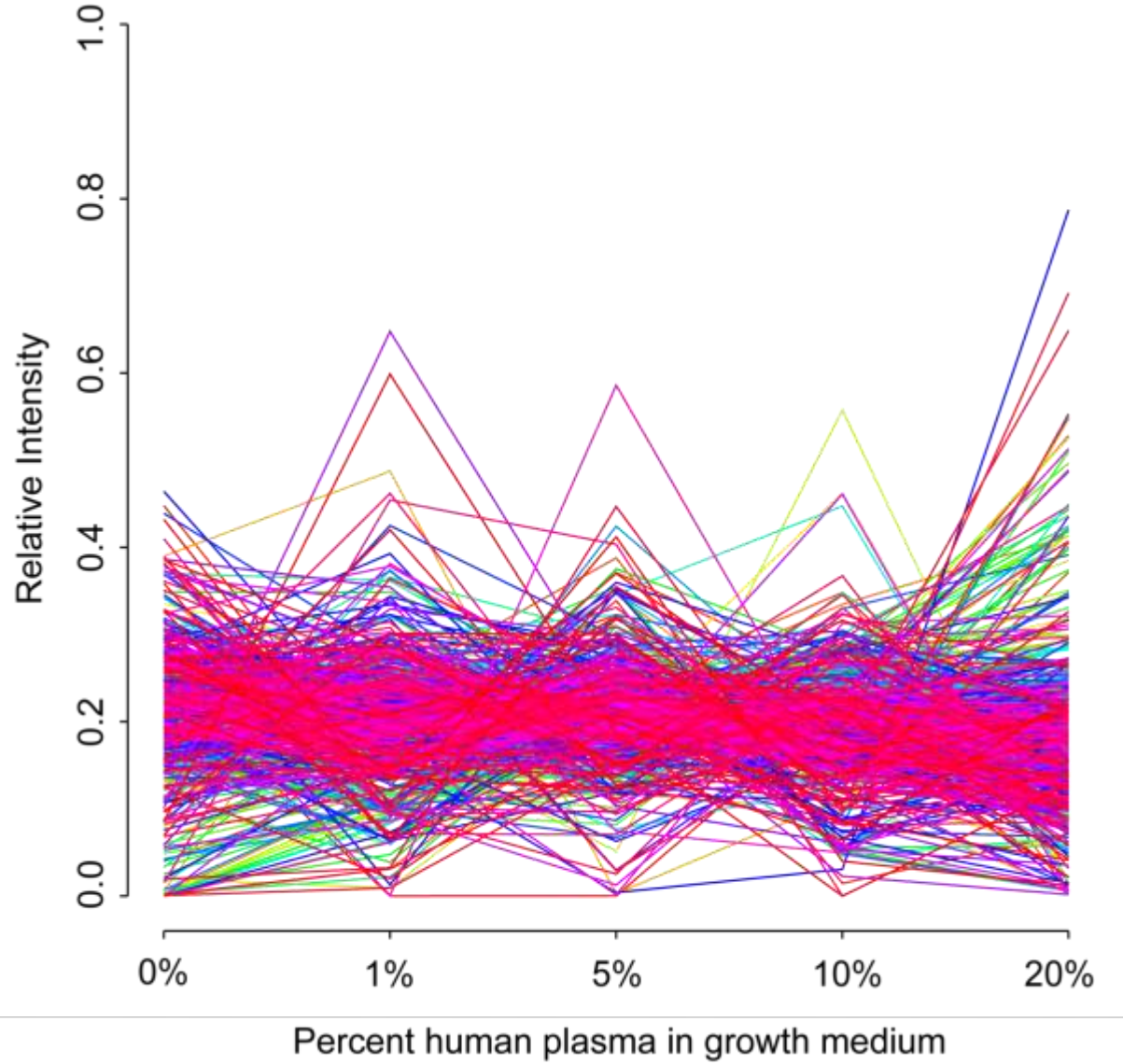
-Strep
-Human

Strep SRM assay measurements

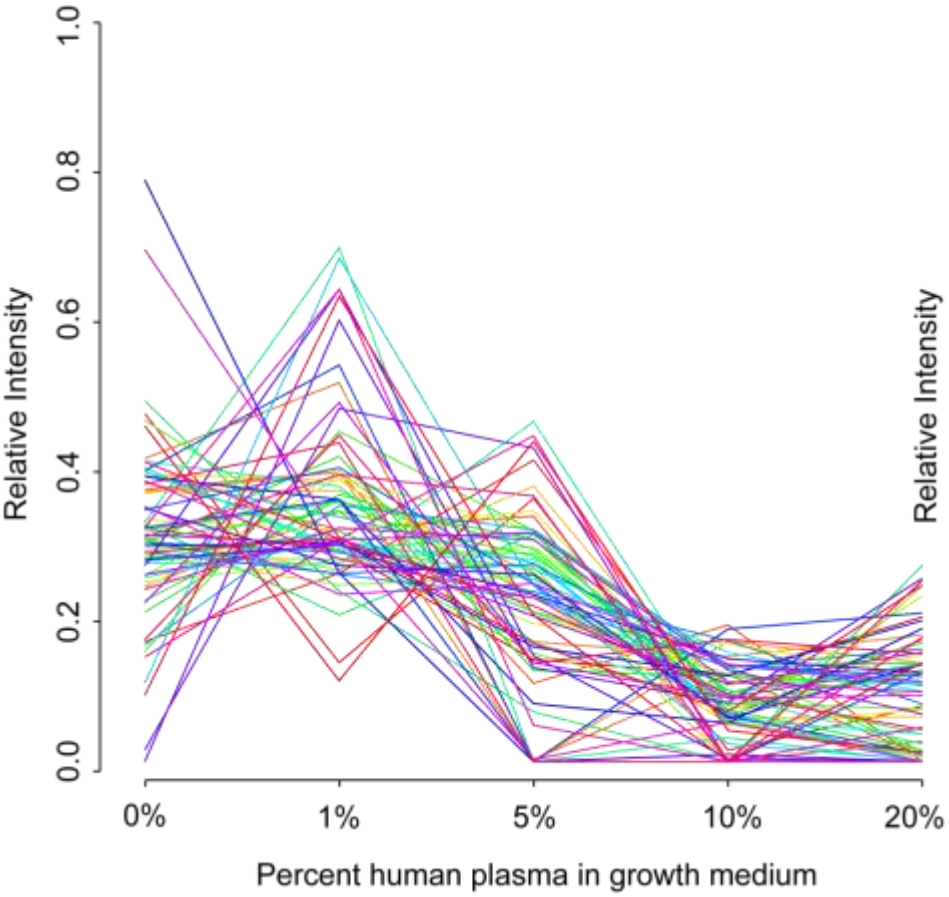


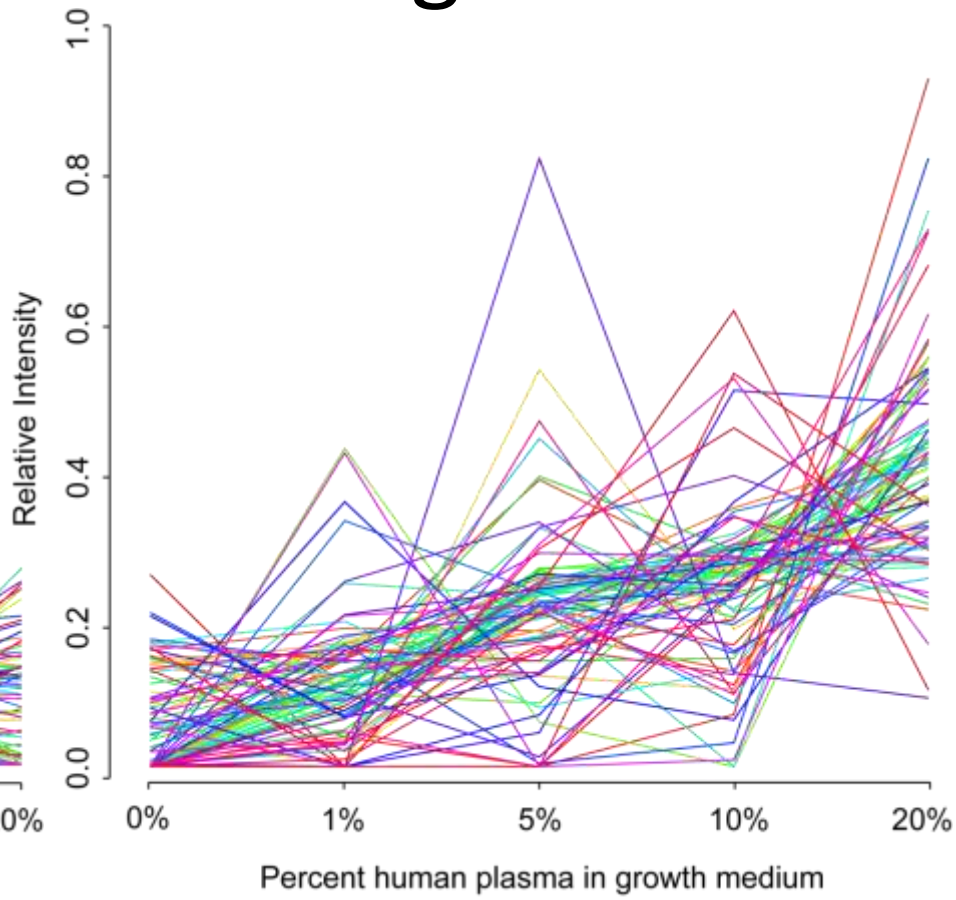
Complete proteome attempted in a dose-response experiment

705 proteins measured in all plasma levels, < 1% FDR.



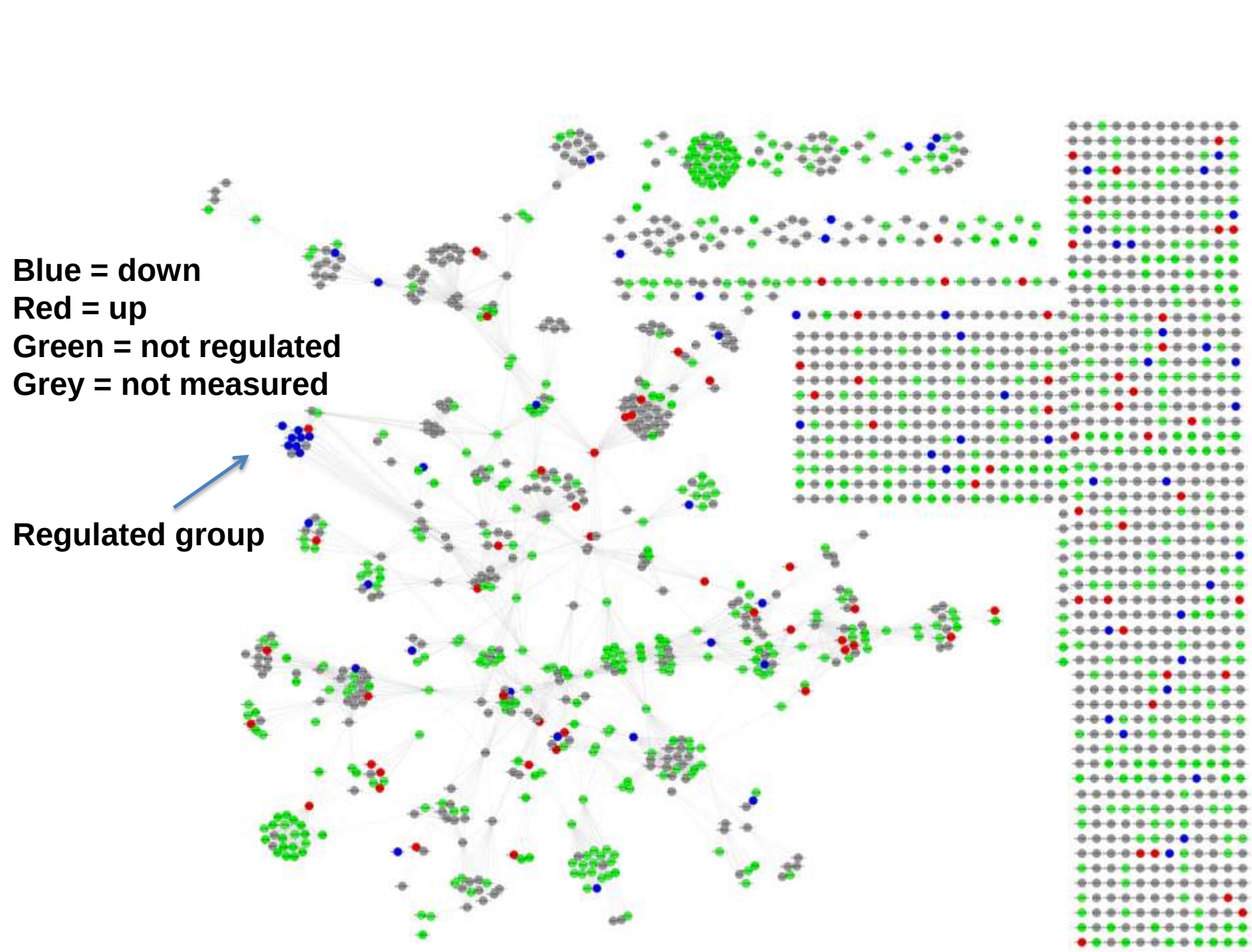
Identify regulated proteins using PCA and k-mean clustering





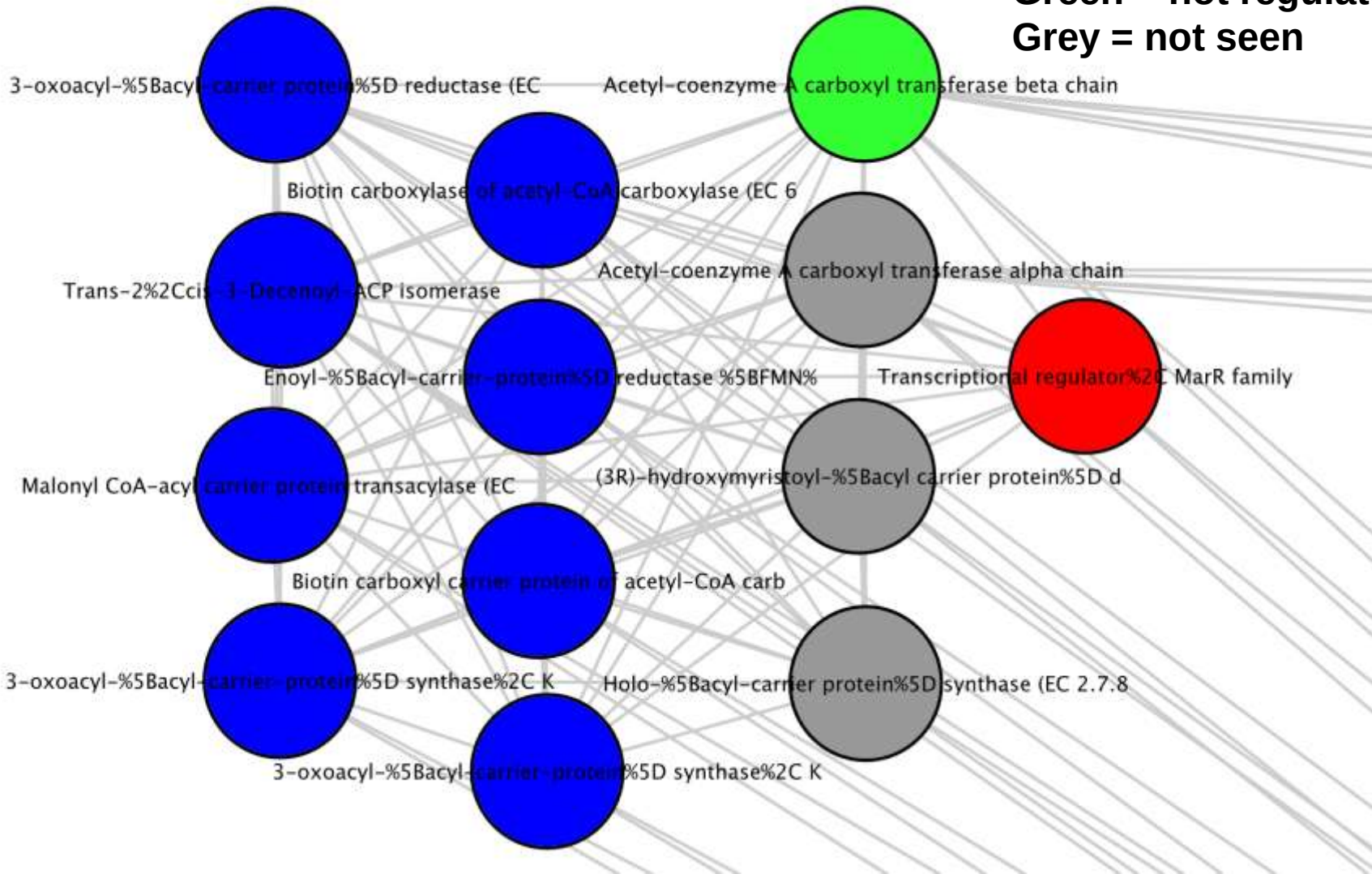
69 proteins down-regulated

84 proteins up-regulated

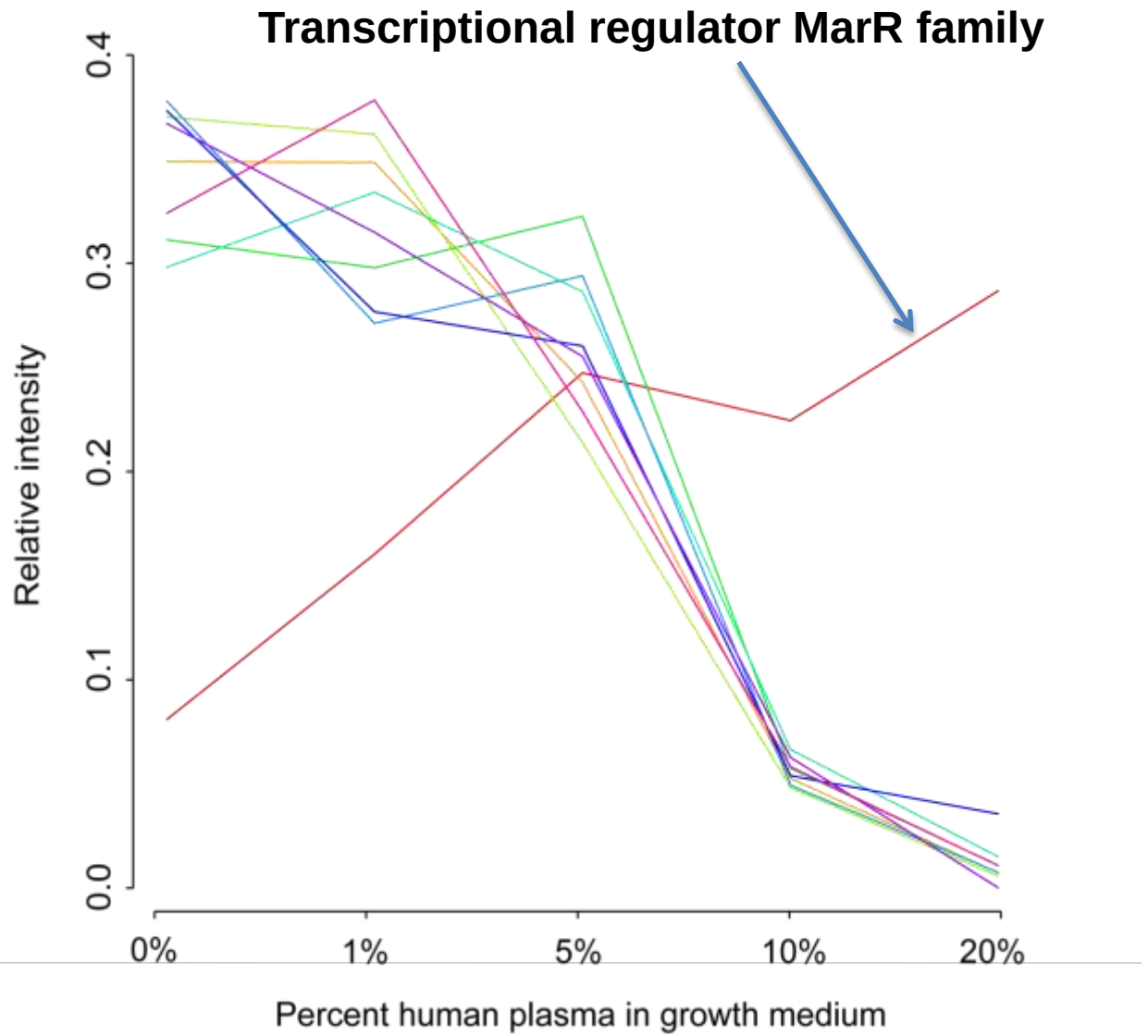

Blue = down
Red = up
Green = not regulated
Grey = not measured
Regulated group

Fatty Acid Biosynthesis FASII

Blue = down
Red = up
Green = not regulate
Grey = not seen



Regulated proteins in FASII group reveal a potential



Vol 458|5 March 2009|doi:10.1038/nature07772 nature

LETTERS

Type II fatty acid synthesis is not a suitable antibiotic target for Gram-positive pathogens

Sophie Brinster[1,2], Gilles Lamberet[3], Bart Staels[4], Patrick Trieu-Cuot[5], Alexandra Gruss[3] & Claire Poyart[1,2,5,6]

Antimicrobial drugs targeting the reportedly essential type II fatty acid synthesis (FASII) pathway[1–5] have been recently acclaimed for their efficacy against infections caused by multiresistant Gram-positive bacteria[6–8]. Our findings show that the strategy for antibiotic development based on FASII pathway targets is fundamentally flawed by the fact that exogenous fatty acids fully bypass inhibition of this pathway in both *in vitro* and *in vivo* conditions. We demonstrate that major Gram-positive pathogens—such as streptococci, pneumococci, enterococci and staphylococci—overcome drug-induced FASII pathway inhibition when supplied with exogenous fatty acids, and human serum proves to be a highly

several major low-GC Gram-positive pathogens. It would not concern bacteria with requirements for fatty acids missing in serum, or drugs that target subsequent activities, for example, those involved in phospholipid or lipid A synthesis (see refs 8 and 19).

Clinical isolates of Gram-positive pathogens *S. aureus*, *E. faecalis*, *E. faecium*, *S. pneumoniae*, *Streptococcus pyogenes* and *S. agalactiae*, including singly or multiply drug resistant strains, were tested for sensitivity to FASII inhibitors. We chose two commercially available anti-FASII drugs: cerulenin, which like platensimycin and platencin inhibits FabB/F, and triclosan, which targets FabI and other FASII components[8,20]. Bacteria were challenged with these drugs in the presence or

Conclusions

- Possible to generate SRM assays for almost an entire proteome by using synthetic peptides
- SRM can be used for label-free relative quantification

- Verified that the FASII pathway is down-regulated when the bacteria is presented fatty acids in the medium