

Mass spectrometry-based proteomics (SWATH-MS / DIA) to identify and characterize biomarkers in complex samples

Dr. Peter Blattmann

blattmann@imsb.biol.ethz.ch

Aebersold group, ETH Zurich

Predicting the behavior of a system is a key scientific goal

- Provided that we know the **state of the system** at a particular time ($t=0$) and we have a theory **how the system works**, we should be able to **predict** the state of the system at any future time and its **reaction to perturbations** (e.g. longer pendulum, greater force, etc.)
- Works for systems with moderate number of parts and a basic model how they interact

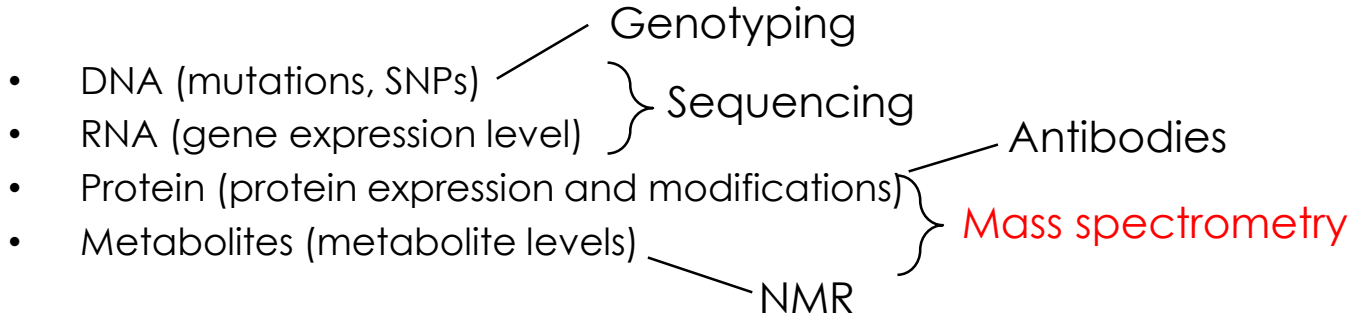
Two challenges in performing this in biology/medicine:

- Assess **the state of the system**
 - Understand **how the system works** or will respond
- Biomarkers are required to address both challenges



Proteins as biomarkers

Different classes of biomarkers exist in medicine and biology



Proteins are the main...

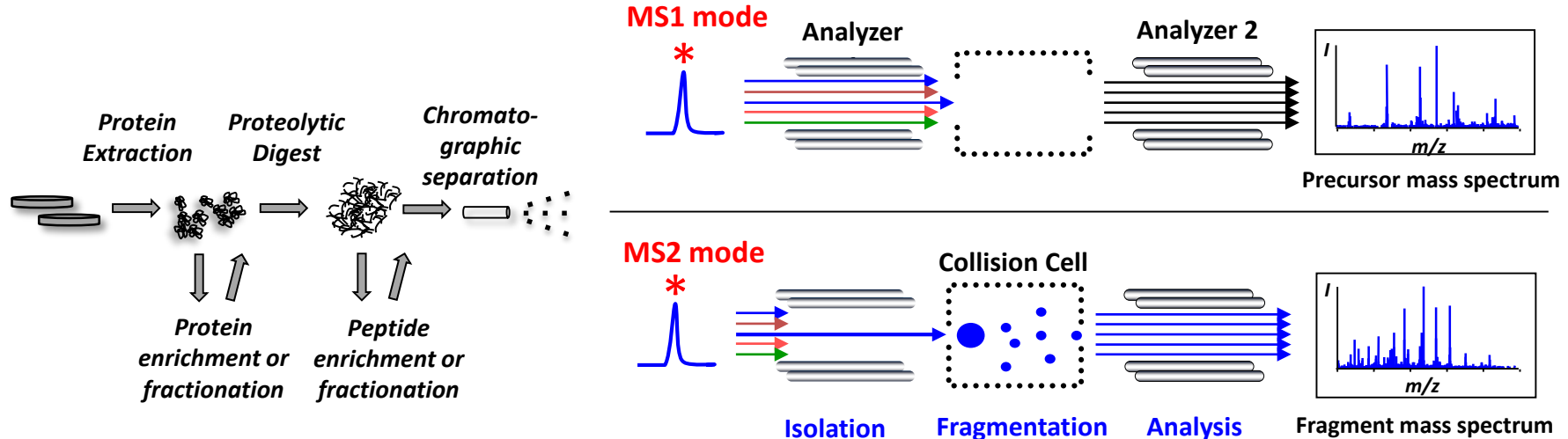
- building blocks of cells.
- mediators of signaling (receptors, kinases, post-translational modifications).
- entities that facilitate chemical reactions (enzymes).

Outline

- Introduction to mass spectrometry-based proteomics (DIA/SWATH-MS)
- Three studies from our lab (published and unpublished)
 - Blood N-Glycoproteins in Five Solid Carcinomas (Saijc et al. 2018)
 - Using ratio of protein abundance as biomarkers (Buljan et al. in preparation)
 - Systems approach to understand what determines variability to drug response (Blattmann et al. 2017)
- Conclusion and Outlook

Mass spectrometry-based quantification of proteins

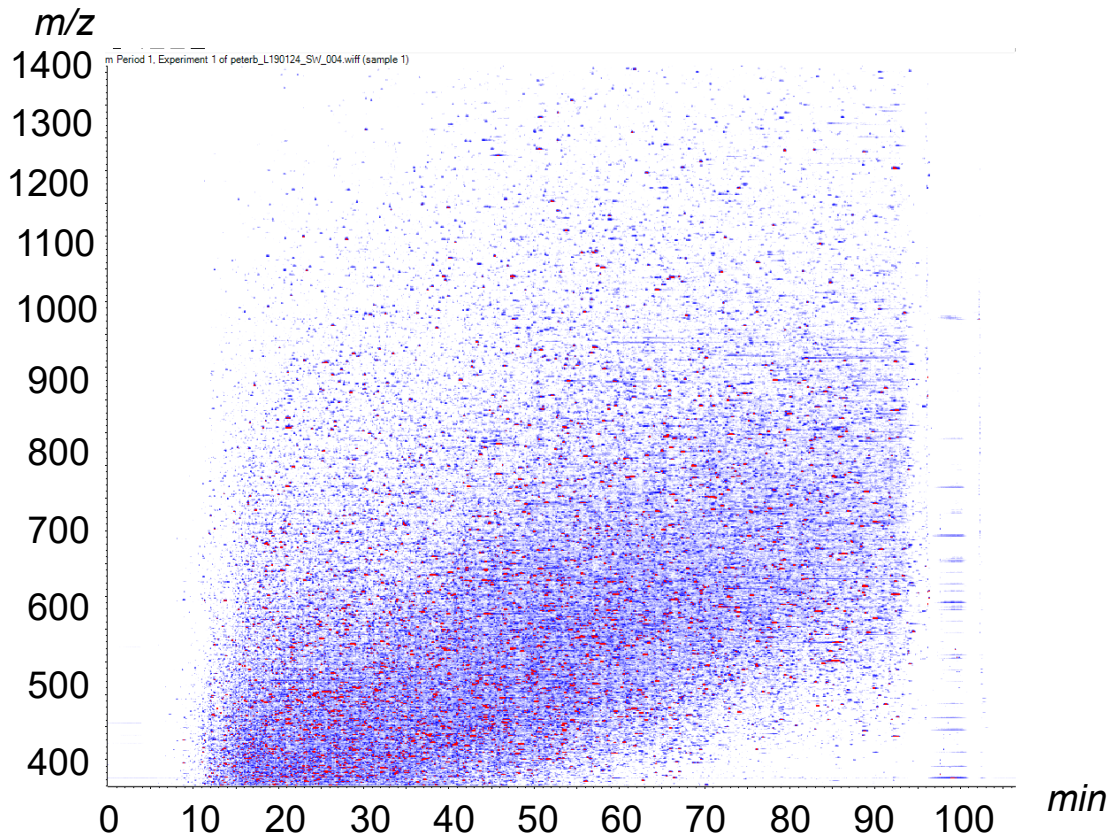
- Proteomics uses tandem mass spectrometers, that can be operated in different modes.
- From the acquired mass spectra, identification and abundance of peptides and proteins can be obtained with unprecedented specificity and accuracy.



How the mass spectrometer sees the proteome

per sample

- 1 MS1 map
- 64 MS2 maps
- about 4 GB data
- $10^6 - 10^8$ features

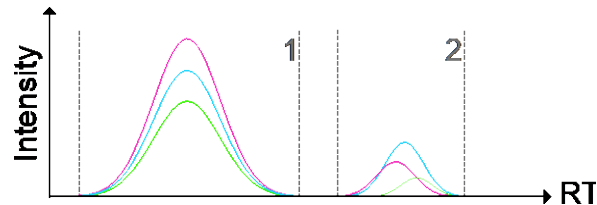
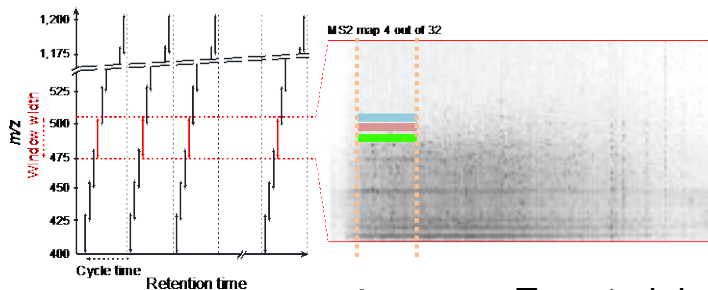


OpenSWATH: Targeted extraction of fragment ion signals

PEPEIDEK

Precursor	Fragment	RT	Intensity
478.73224	430.20586	21.9	1000
478.73224	405.67948	21.9	670
478.73224	365.68456	21.9	450

Rosenberger *et al.*, Sci. Data, 2014



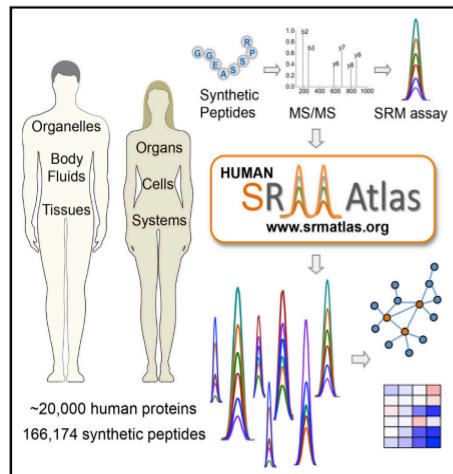
Targeted data extraction

Gillet, *et al.*, Mol. Cell. Proteomics, 2012
Navarro *et al.*, Nat. Biotechnol, 2016

Röst & Rosenberger, *et al.*, Nat. Biotechnol., 2014
Rosenberger *et al.*, Nat Biotechnol, 2017

Human SRMatlas: A Resource of Targeted Assays to Quantify the Complete Human Proteome

Graphical Abstract



Highlights

- Human SRMatlas: 166,174 proteotypic peptides representing the human proteome
- Resource of verified high-resolution spectra and multiplexed SRM assays
- Supports proteome-scale quantification as well as hypothesis-driven research
- Web database with free unlimited access

Authors

Ulrike Kusebauch, David S. Campbell, Eric W. Deutsch, ..., Leroy Hood, Ruedi Aebersold, Robert L. Moritz

Correspondence

aebersold@imbs.biol.ethz.ch (R.A.), rmoritz@systemsbiology.org (R.L.M.)

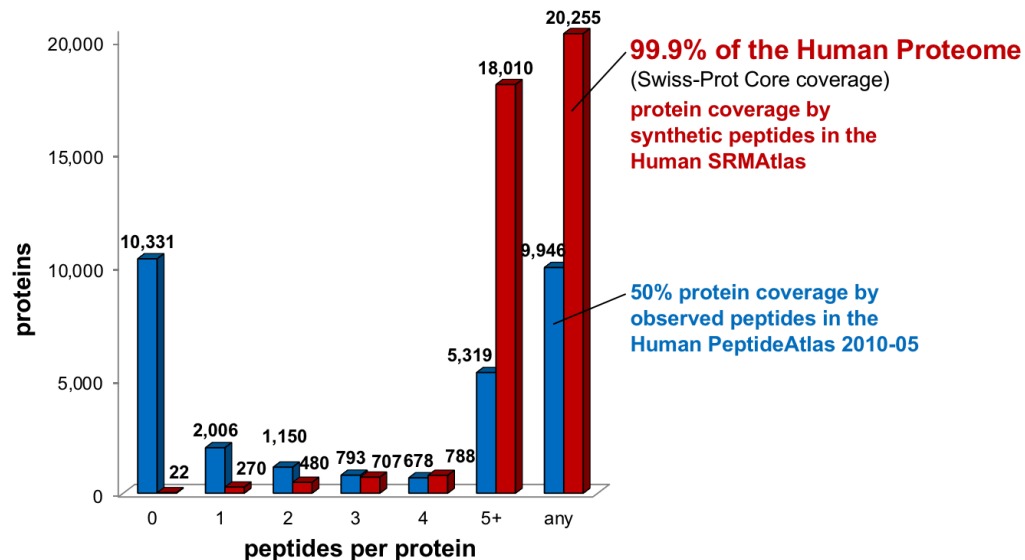
In Brief

This resource enables the accurate detection and quantification of any known or predicted human protein from complex biological samples.

Accession Numbers

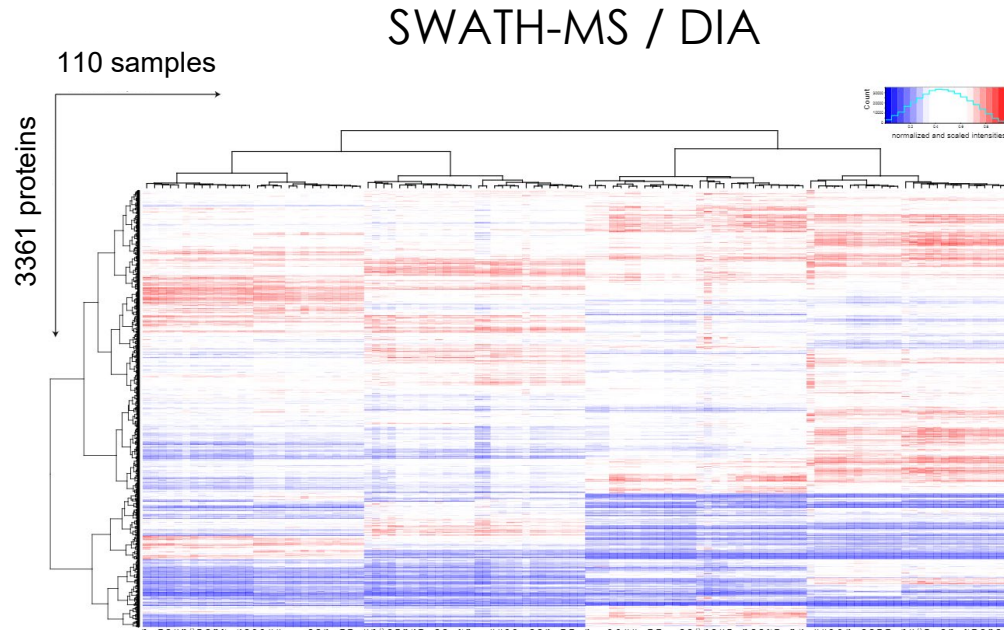
GSE83654

For every human protein one or several definitive mass spectrometric measurement assays are publicly available supporting peptide centric analyses



Consistent measurements across 100s of samples

- Measurement of > 12'000 peptides from > 3000 of proteotypic proteins across 100s of samples (newer machines achieve even higher rates).



Blattmann et al. 2017

SRM / MRM

- higher throughput
- less peptides

Conclusion: MS-based proteomics (DIA/SWATH-MS)

- DIA /SWATH-MS is a powerful proteomic technique to quantify peptides/proteins across 100s of samples with high precision and accuracy.
- DIA/SWATH-MS acquires a proteomic fingerprint of a peptide sample that can always be re-interrogated with new hypotheses (targeted data extraction).
- New mass spectrometers will greatly increase sensitivity, specificity, and throughput even further.

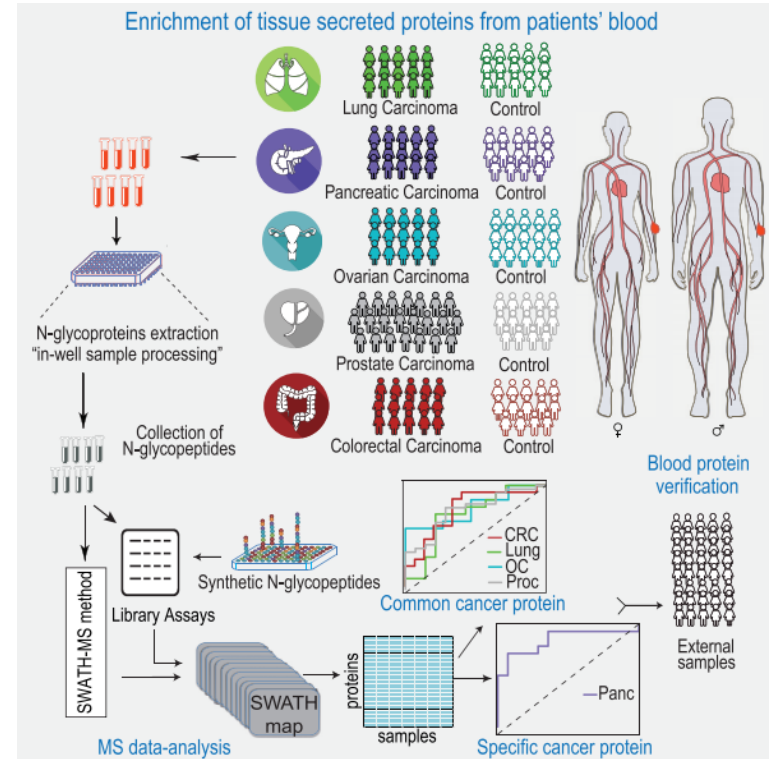
Three studies from our lab

Two challenges:

- Assess **the state of the system**
 - Understand **how the system works** or will respond
-
1. Blood N-glycoproteins across five carcinomas (Saijc et al. 2018)
 2. Ratio of protein abundance as biomarkers (Buljan et al. in preparation)
 3. Systems approach to understand what determines variability to drug response (Blattmann et al. 2017)

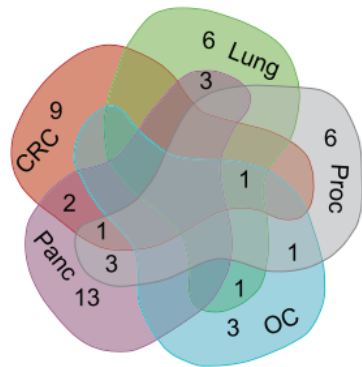
Biomarker discovery of plasma proteins

- Enrichment for N-glycoproteins from blood serum
- Quantified 1,444 distinct N-glycopeptides from 272 proteins
- Serum samples from 284 patients and controls with different localized carcinomas
- Mean CV of 19% across experimental replicates



Specific and common candidate biomarkers for cancers

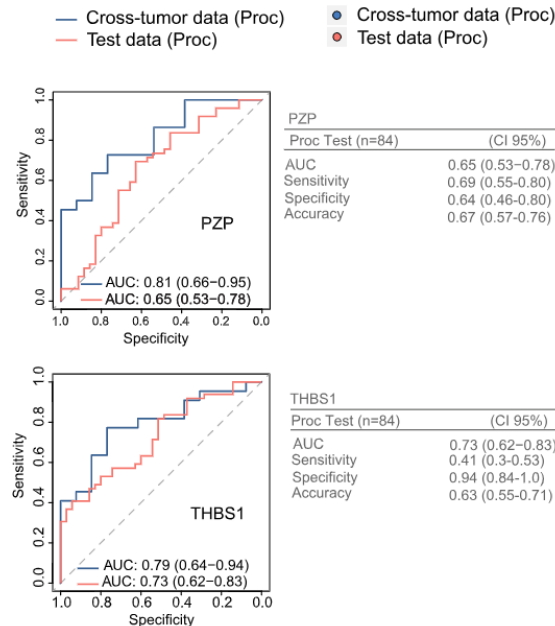
- Data resource that assesses biomarkers across several cancers within same study.
- THBS1 was regulated in 4 out of 5 cancers
- Validation study confirmed effect for PZP and THBS1 in Prostate cancer



Venn diagram summary of shared proteins

Carcinomas	Shared elements
CRC / Lung / OC / Proc	(1) THBS1
CRC / Panc / Proc	(1) ANGPTL7
CRC / Panc	(2) A2M, APOB
Lung / OC	(1) SELP
Lung / Panc	(3) HRG, F13B, F12
OC / Proc	(1) GP5
Panc / Proc	(3) ITIH3, FN1, IGFB3

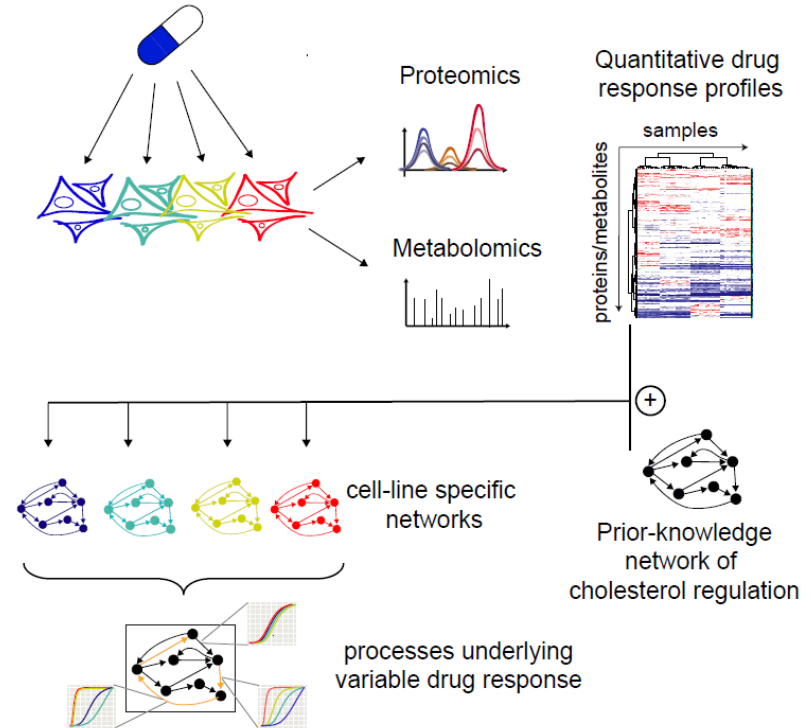
*platelet-related protein



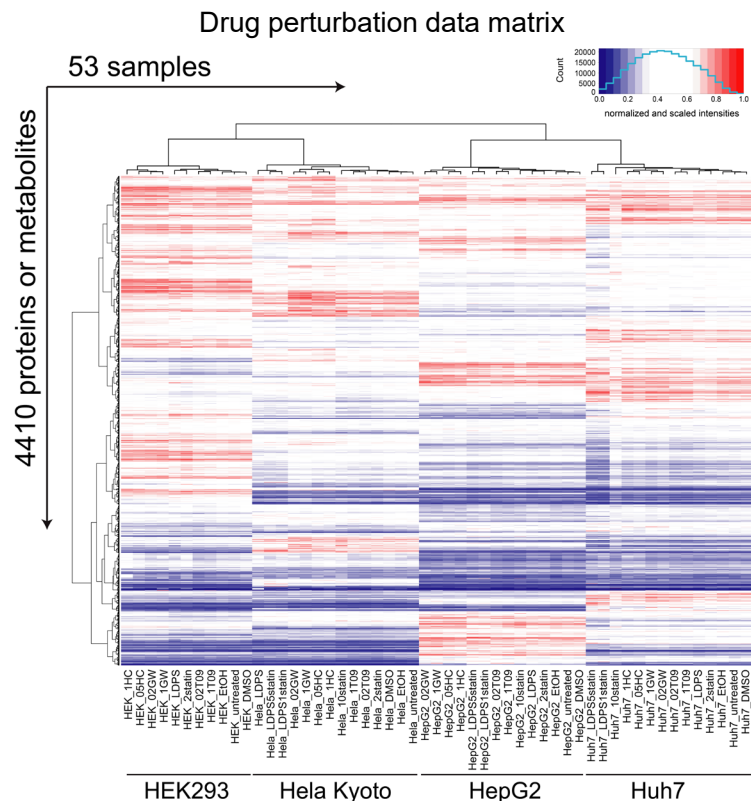
Understand what determines variability to cellular drug response

Cholesterol regulation as a clinical relevant complex biological process

1. Perturb 4 different cell lines with the same drugs or siRNAs
2. Measure quantitative proteomic and metabolomic response
3. Generate cell-line specific mathematical models based on prior knowledge and quantitative data to identify the processes that vary.



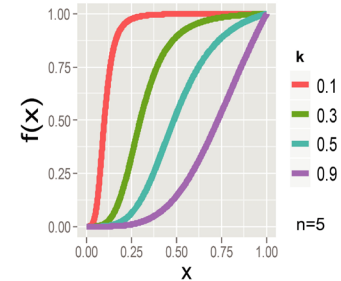
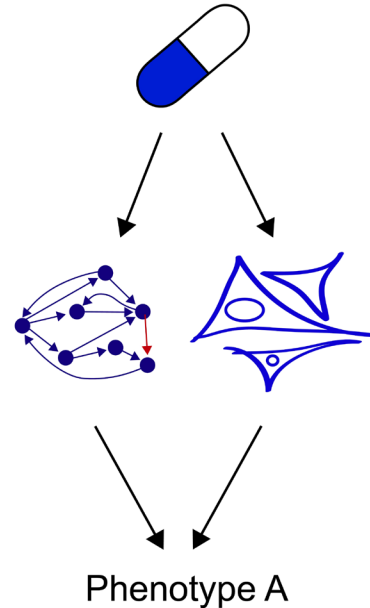
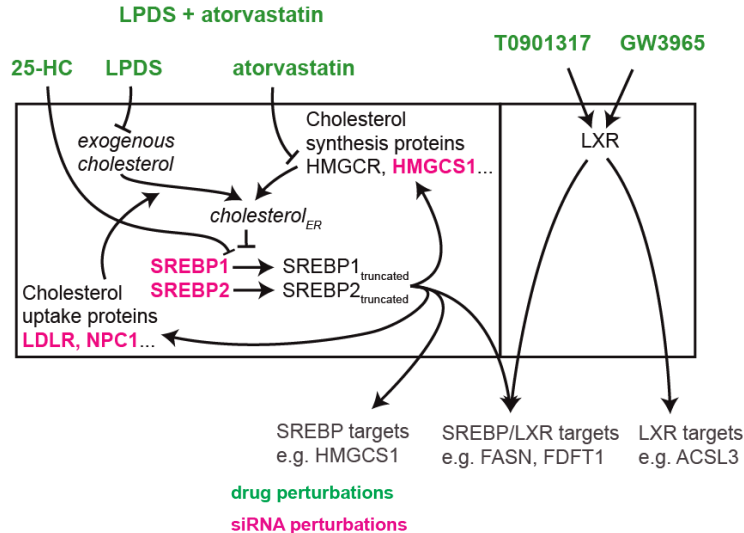
Summary of results



- Unique baseline profile of each cell line is modulated by drug perturbation
- Widespread drug response
 - 694 (21%) of proteins and 435 (42%) of metabolites were regulated ($|\log_2FC| > 0.5$ & $FDR < 0.01$)
- Combining proteomics and metabolomics allowed quantifying the response from orthogonal levels
 - Metabolomics: drug metabolites and inhibition of pathways
 - Proteomics: regulated pathways
- Qualitative conservation but quantitative variability in response phenotypes

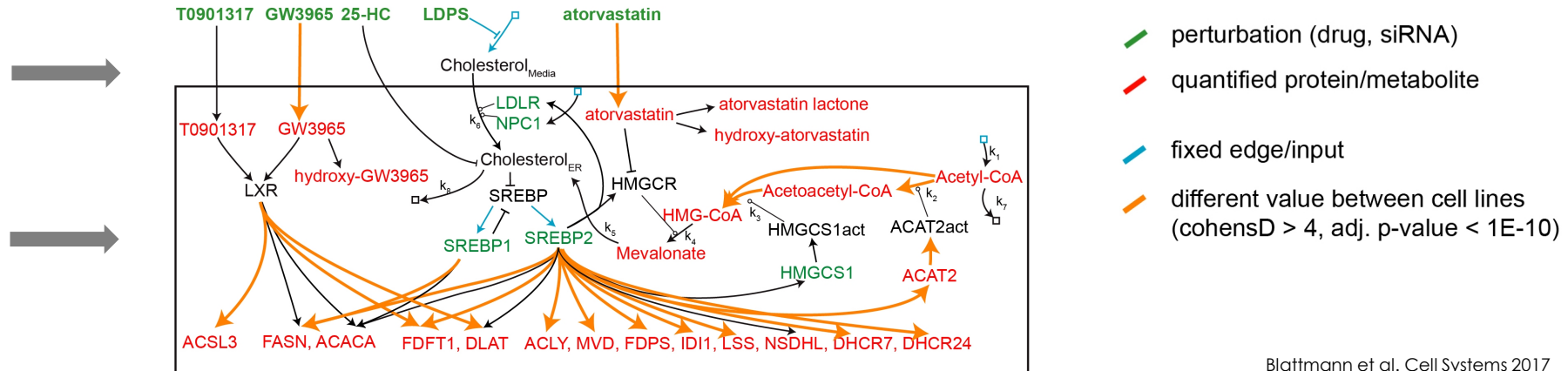
Mathematical model of cholesterol regulation

- Combining an prior-knowledge network with estimating the model edges from the experimental data (CellNOpt, Terfve et al. 2012)



The variability cannot be reduced to one major factor

- Comparison of the various cell-line specific models identifies the origin for the variability in drug response
- Differences in drug uptake and in the effect of transcription factors underlie the heterogeneity in drug response
- The variability in response phenotypes cannot be explained by a few different factors but only with the network as a whole.



Conclusion

- DIA/SWATH-MS can be used to both measure protein biomarkers and to gain a better understanding how the system responds to perturbations.
- DIA/SWATH-MS can be combined with fractionation techniques (Affinity-purification, Size-exclusion chromatography, etc.).

Key technology to support both goals :

- Assess **the state of the system**
- Understand **how the system works** or will respond

Acknowledgements

Ruedi Aebersold

Aebersold lab

Matthias Gstaiger

Tatjana Sajic

Maria Buljan

Fabian Frommelt

Mass-spec operating team

Ludovic Gillet

Alex Leitner

Tatjana Sajic

Collaborations

Julio Saez-Rodriguez lab

Bernard Mallisen lab

Marc Rubin lab

Zlatko Trajanoski lab



Aebersold lab, Tschamut 2019



SWISS NATIONAL SCIENCE FOUNDATION



SystemsX.ch

The Swiss Initiative in Systems Biology

