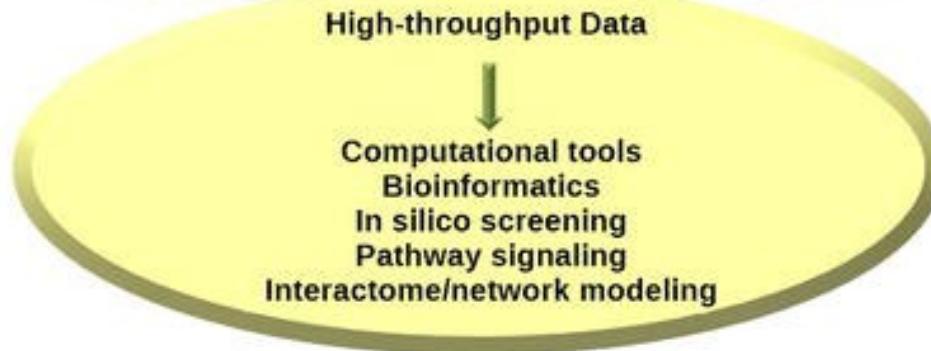
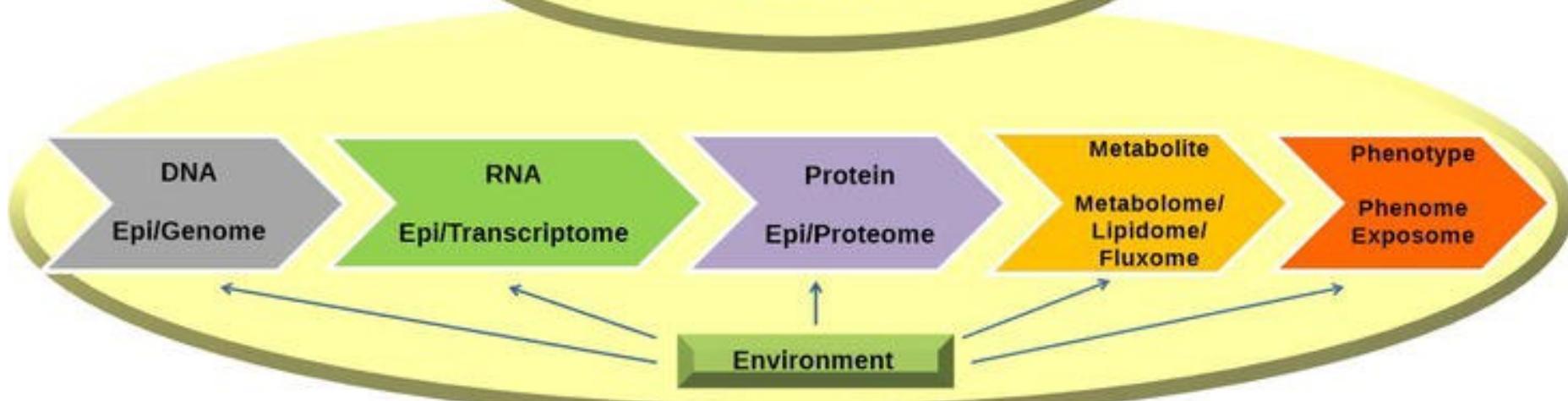
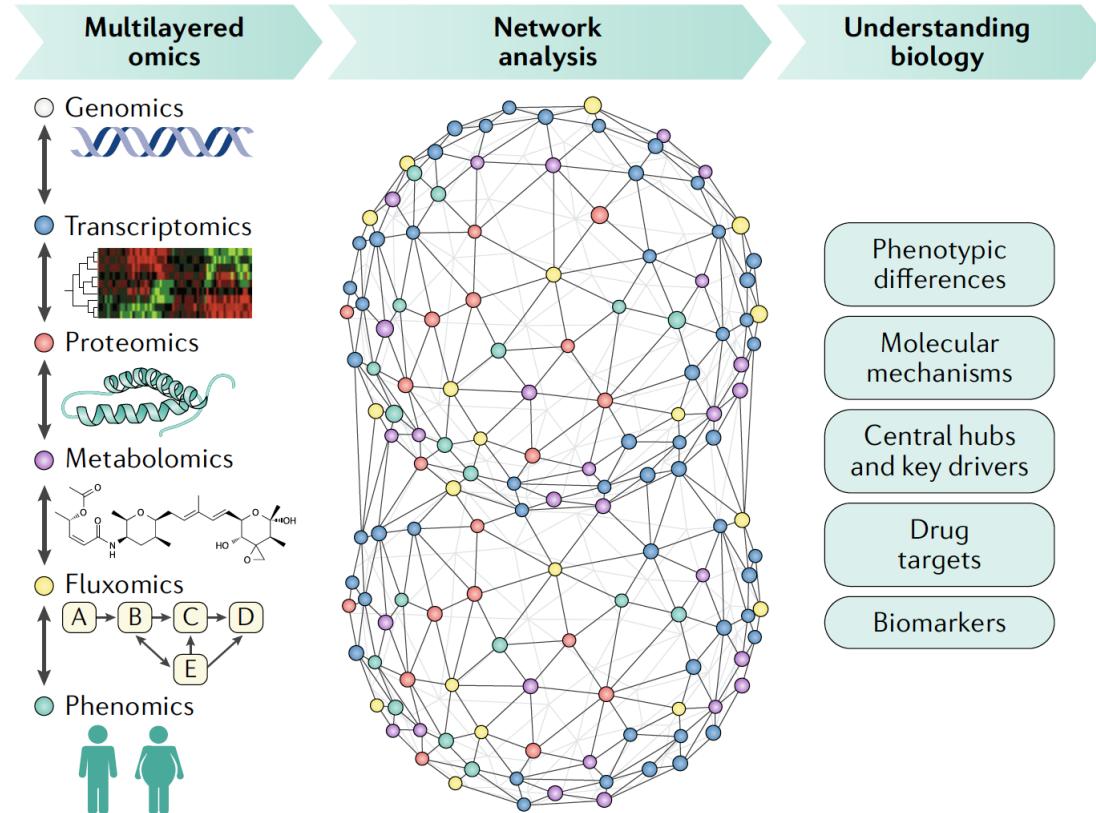


Proteomics

OMICS-based Systems



Systems Biology in hepatology

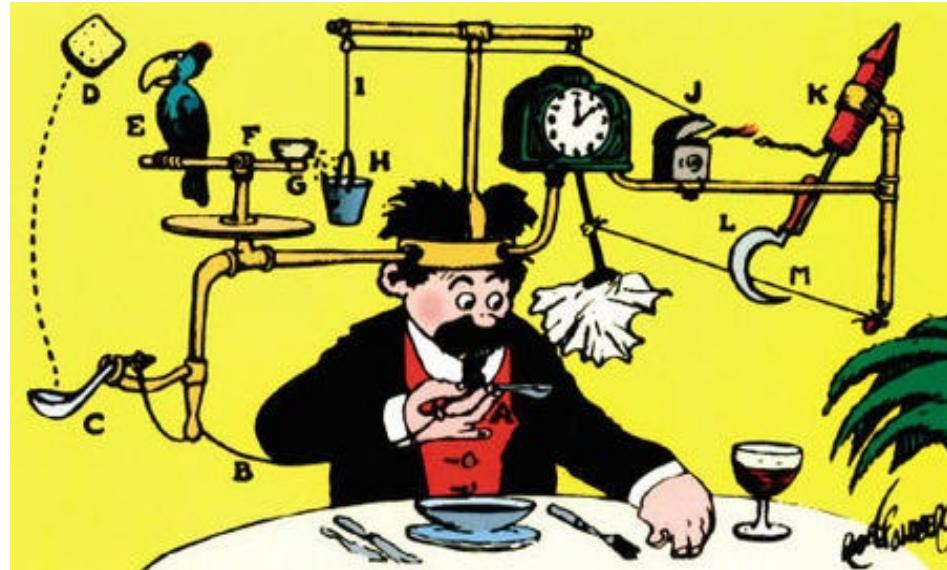


Systems Biology & Systems Medicine

- Evolution created enormously complicated organisms and this random and chaotic process only work in particular environment.
- Over the years complicated circuits are built on top of this complex structure.
- To study and understand this complexity is a challenging task.

Rube Goldberg's inventions

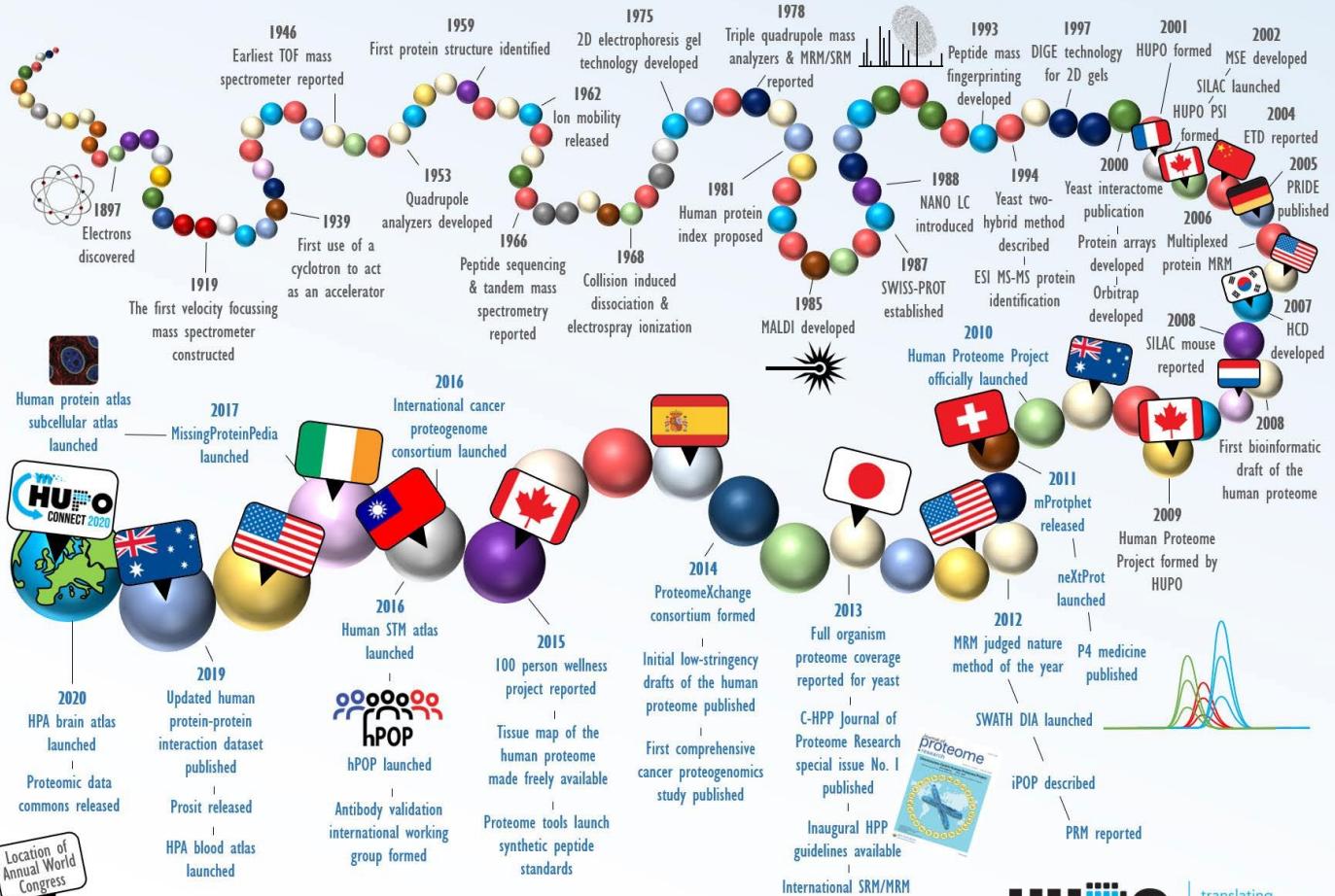
- ✓ Number of components
- ✓ Individual dynamics
- ✓ Their connection
- ✓ System dynamics



What is proteomics?

Definition – Large-scale study of proteins

In daily use – Studying a protein sample using mass spectrometry



<https://hupo.org/Proteomics-Timeline>



translating
the code of life

!

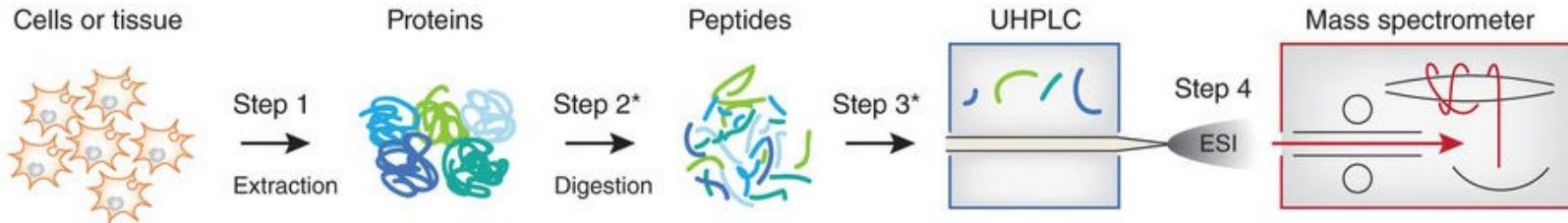
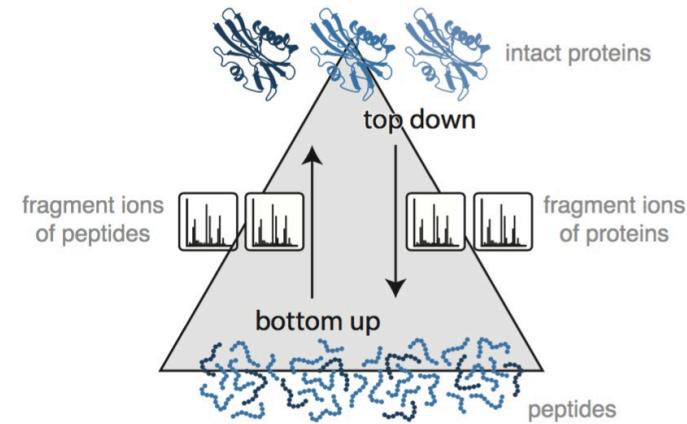
Mass spectrometry-based proteomics

Top down proteomics

- Analysis of intact proteins
- Complicated MS data analysis

Bottom up proteomics

- Peptides released by enzymatic degradation
- Measures peptide-ions and their fragmentation patterns
- By far the most used approach in proteomics





Different proteomics approaches

Data-dependent acquisition proteomics (DDA or Shotgun)

- No prior knowledge needed
- MS1 quantification

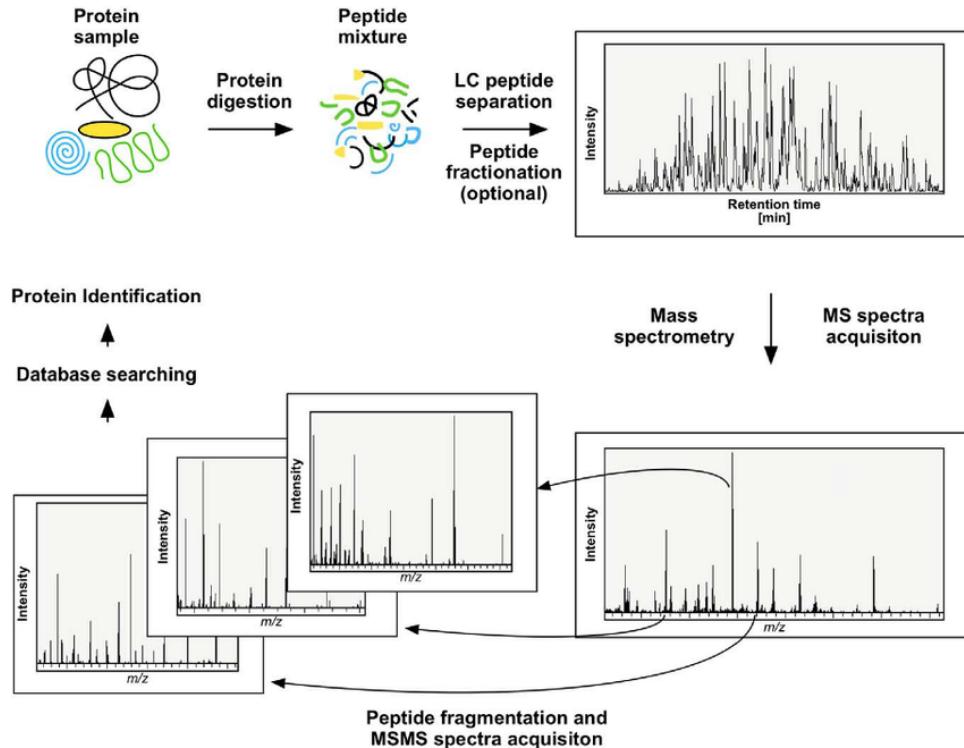
Data-independent acquisition proteomics (DIA or SWATH)

- Needs spectral libraries
- MS2 quantification

Targeted proteomics (SRM, MRM, PRM, HR-MRM etc.)

- Needs assay development
- Generally hypothesis driven
- MS2 quantification

Protein identification in shotgun proteomics



- Protein digestion into peptides
- Peptide separation using liquid chromatography
- Electrospray ionization
- Full scan of peptide ions (MS1)
- Selection of the most intense peptide ions to fragment and analyze (MS2)
- Fragment ion spectra are used for peptide identification
- Protein quantification based on peptide ion intensity (MS1 level)

Peptide fragmentation

Peptide ion fragmentation

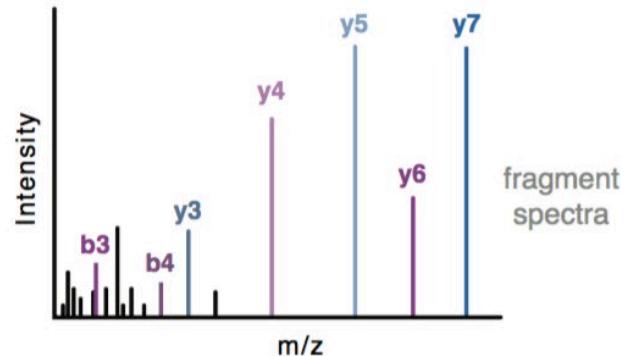
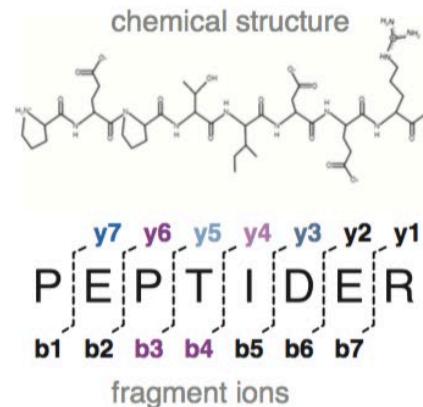
- Peptide ion isolation
- Collision with inert gas

Fragmentation pattern

- Based on sequence
- Depends on collision energy

Peptide identification

- Precursor mass
- Fragmentation pattern
- Match to data base



Data analysis in shotgun proteomics

Dedicated software and search algorithms (vendor specific and free software)

- Load raw data files
- Detect peptide peaks
- Match peptide fragmentation spectra against a data base
- Perform some kind of peptide/protein quantification
- Produce table of identified (and quantified) peptides or proteins

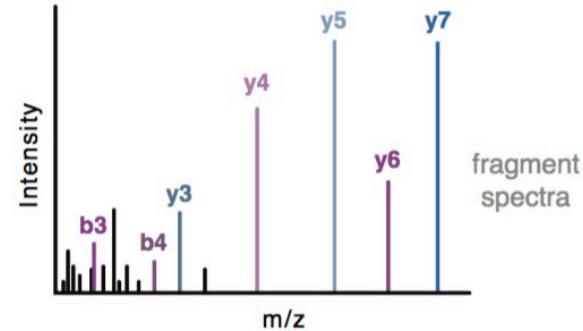
Matching peptide spectra

- Choose database to match against (species, subset of proteins)
- Choose your acceptable false discovery rate (FDR, typically 1%)

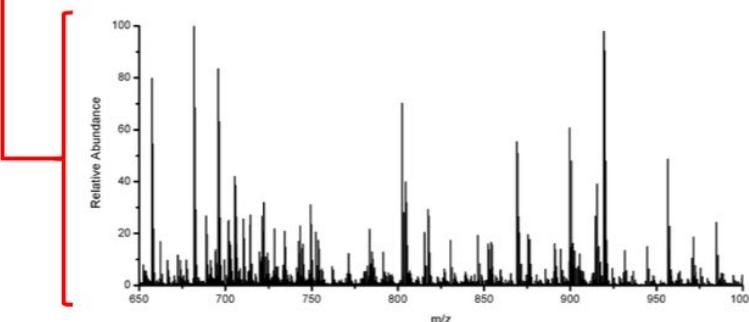
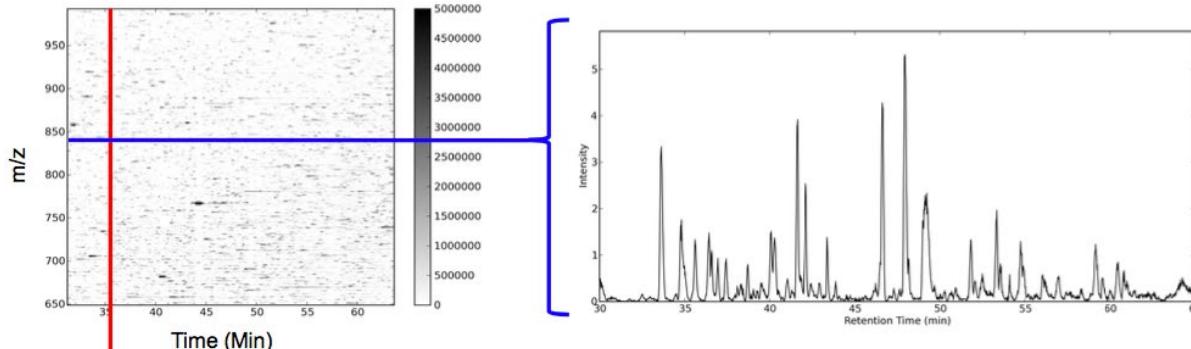
Protein quantification

- Choose quantification method (label free, isobaric tags, stable isotopes,

Unique peptide → Protein ID

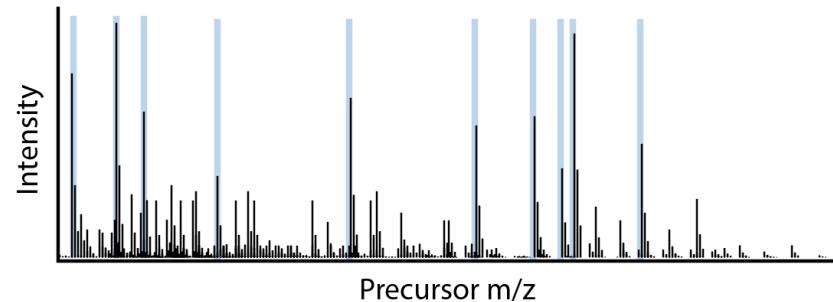
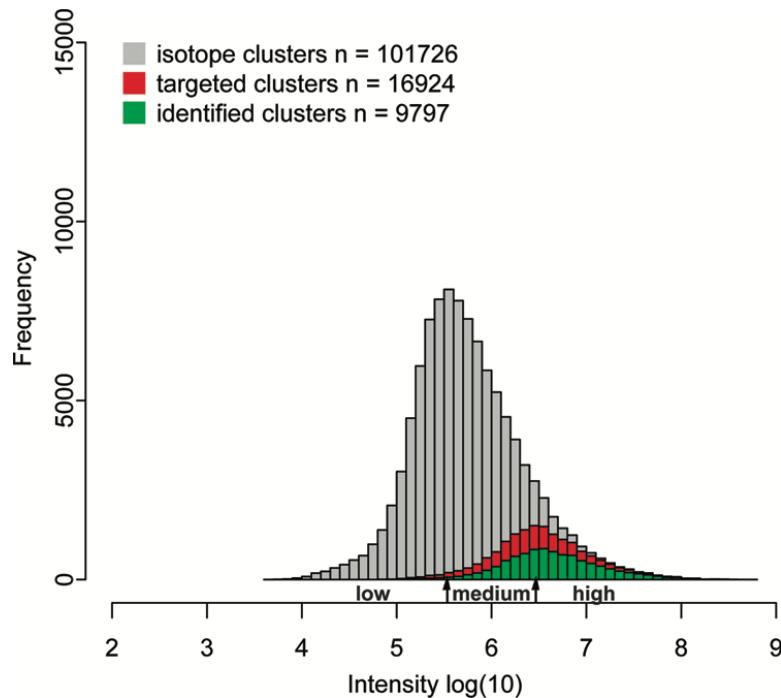


Proteomes are Dynamic and Complex



Data Dependent Acquisition will Always Under Sample

Isotope clusters identified from a shotgun experiment



How well does shotgun proteomics describe a sample?

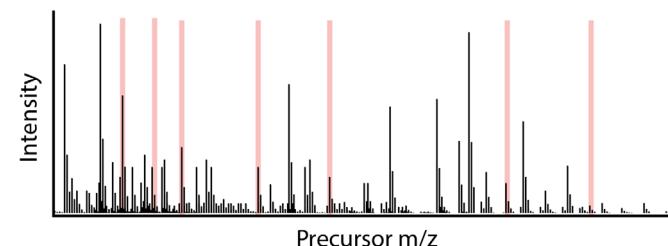
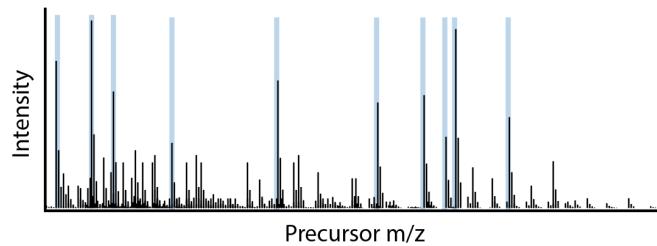
Imagine visiting Kyoto (Japan) and visiting the tallest buildings first



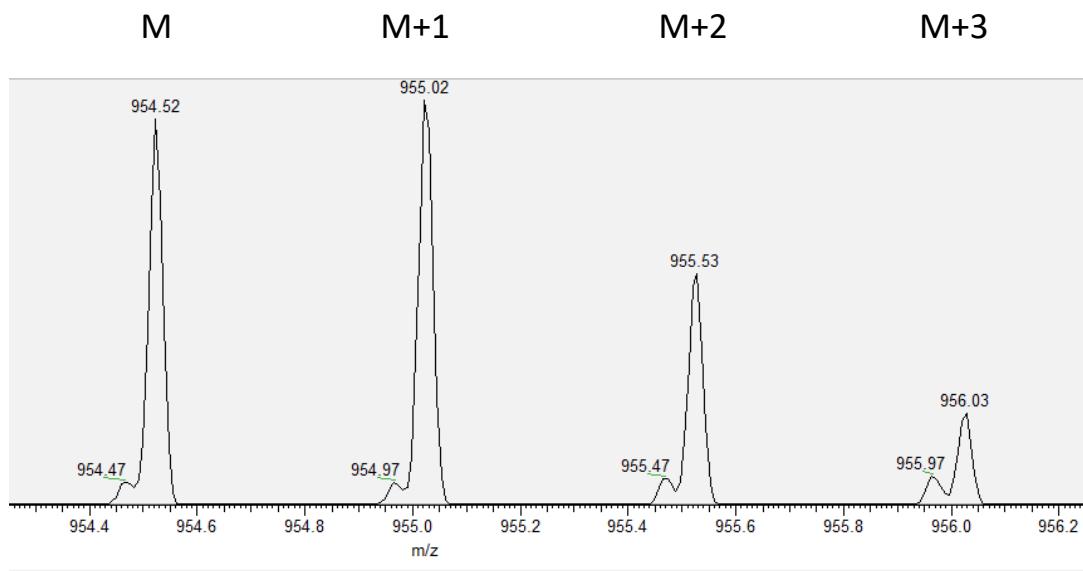
Copyright Rob Laddish

How well does shotgun proteomics describe a sample?

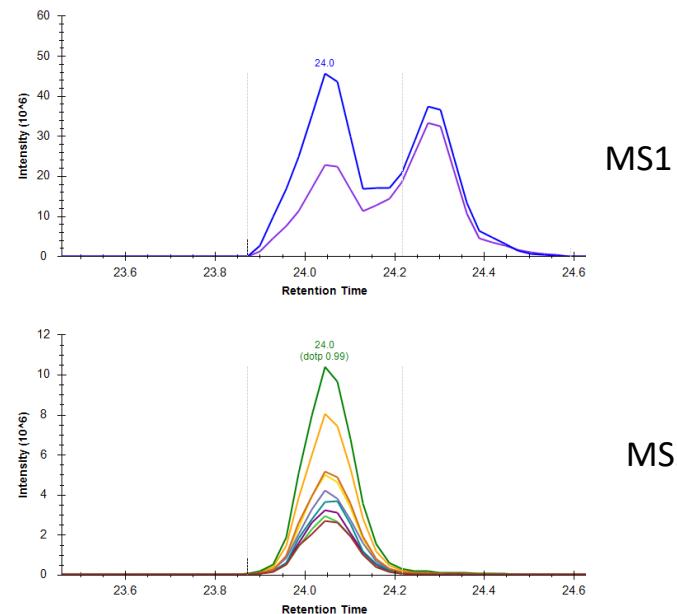
Imagine visiting Kyoto (Japan) and visiting the tallest buildings first



Interference



M: monoisotopic mass

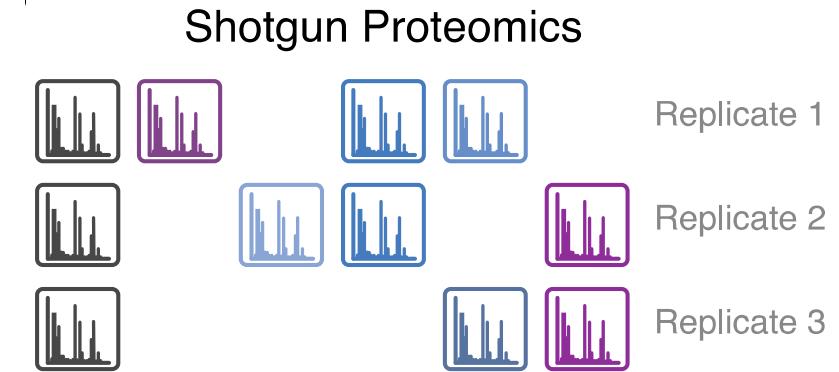


MS1

MS2

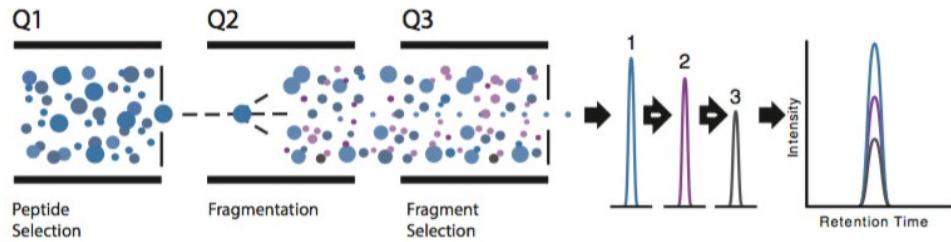
Limitations with shotgun proteomics

- “Semi-random” sampling for peptide fragmentation and identification
- Data sets with missing values
- Favors detection and quantification of high abundant proteins -> low sensitivity
- MS1 quantification generally less accurate than MS2 quantification

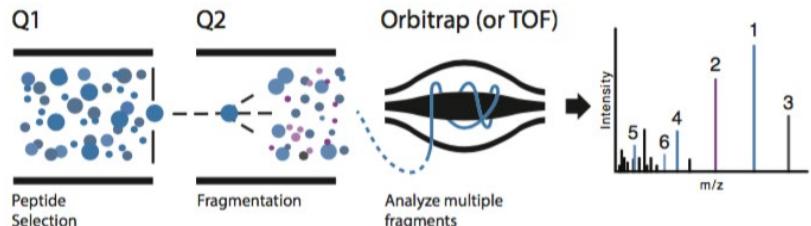


Targeted Proteomics

Selected Reaction Monitoring (SRM)/Multiple Reaction Monitoring (MRM)



Parallel Reaction Monitoring (PRM)



METHOD OF THE YEAR

NEWS FEATURE | SPECIAL FEATURE

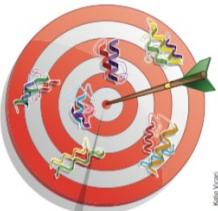
Targeted proteomics

Analysis of a preselected group of proteins delivers more precise, quantitative, sensitive data to more biologists. Vivien Marx reports.

Although the number and identity of protein-coding genes in humans and many other organisms are known to a certain level of approximation, the numbers of proteins produced by each of these genes remains a mystery. Further complicating matters, given the many possible splice forms and post-translational modifications, the potential number of proteins is "staggering," says Arizona State University researcher Josh LaBaer, who is also president-elect of the US Human Proteome Organization. A protein is also dynamic. "It's phosphorylated this minute; it's not phosphorylated the next

"I personally can't wait until we stop hearing about someone describing how big of a list of proteins, peptides or phosphopeptides they detected," says one researcher critical of discovery proteomics who did not wish to be identified. Proteomics has been doing "my list is bigger than your list" for far too long. "It is way more important to measure the one right protein than 10,000 wrong ones."

Scientists wanting to follow well-founded hunches about dozens or hundreds of proteins seek a focused, reproducible, quantitative view of a small subset of the whole



Targeted Proteomics

Benefits

- Data for the same peptides are recorded in every run
- More consistent data
- Easier to compare between samples (e.g. case / control)

Drawbacks

- Peptides has to be well characterized
 - Ionization
 - Retention time
 - Fragmentation
- Tedious assay development
- Only provides data on a limited number of peptides

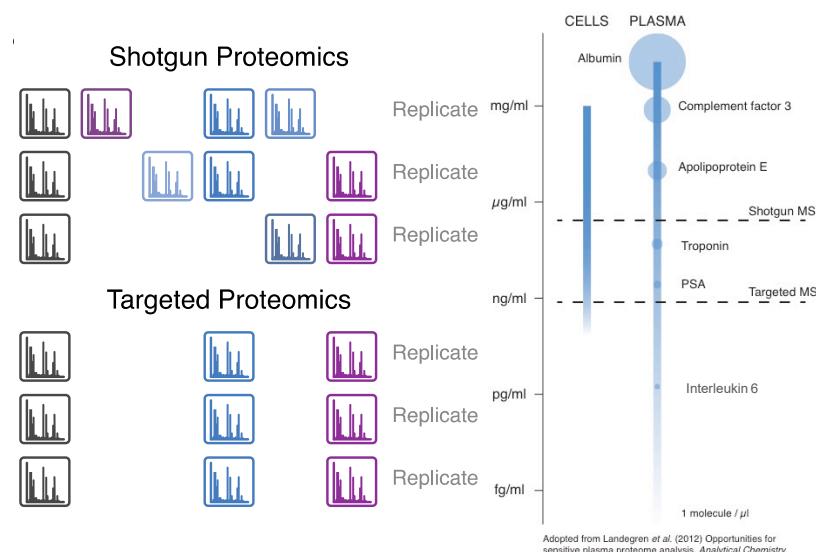
Targeted proteomics vs. shotgun proteomics

Shotgun proteomics

- Variable sampling of peptides for fragmentation – Replicate injections can detect different peptides
- Quantification on MS1-level (peptide ions)
- Generally used in discovery studies
- Limited sensitivity

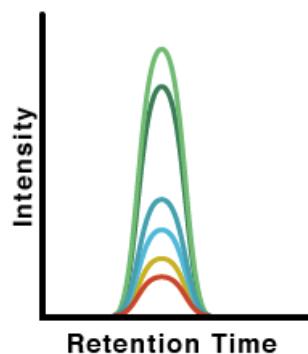
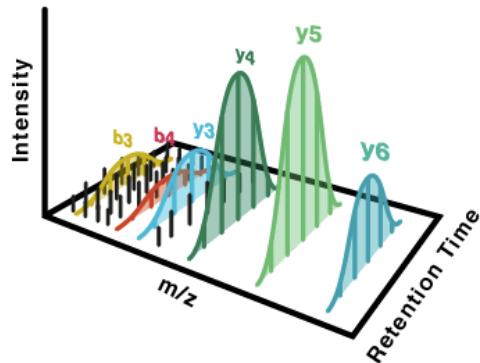
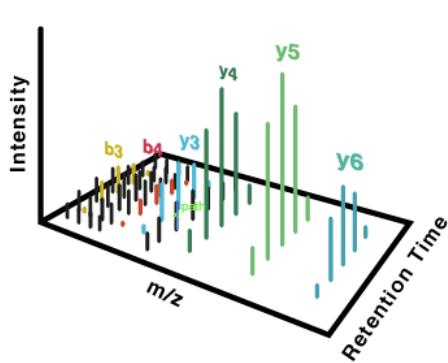
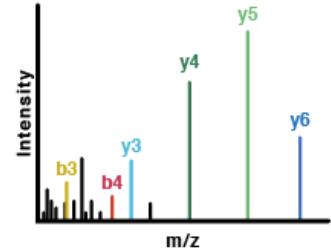
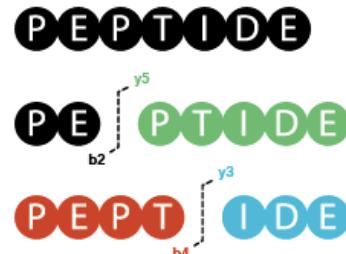
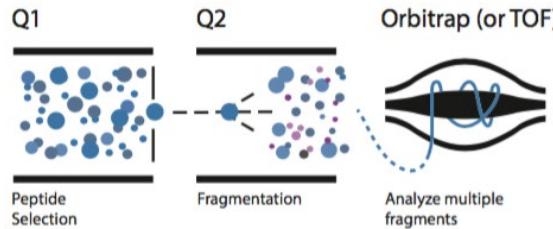
Targeted proteomics

- Predefined list of peptides to fragment – Good reproducibility
- Quantification on MS2-level (fragment ions)
- Generally used for verification of an hypothesis
- Higher sensitivity



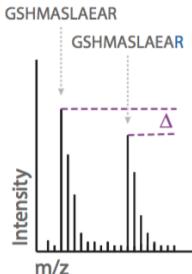
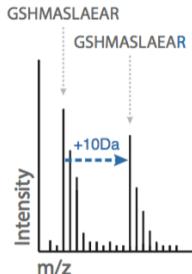
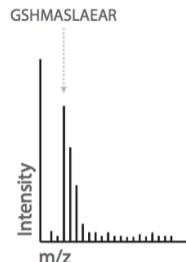
Chromatographic Peak Extraction

- Quantification of one peptide analyzed in PRM



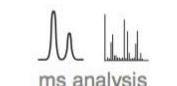
MS-based Peptide Quantification

Label Free

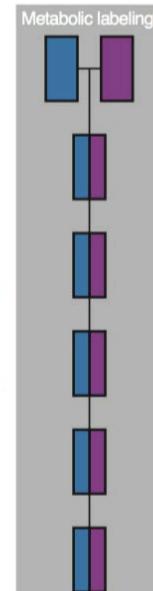


Stable isotopes

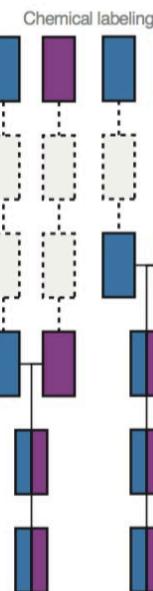
- Metabolic labeling
- Chemical labeling
- Spiked Standards



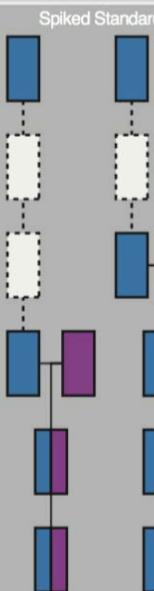
Metabolic labeling



Chemical labeling



Spiked Standards



Label free



Stable isotope-labeled standards

Heavy versions of peptides or proteins

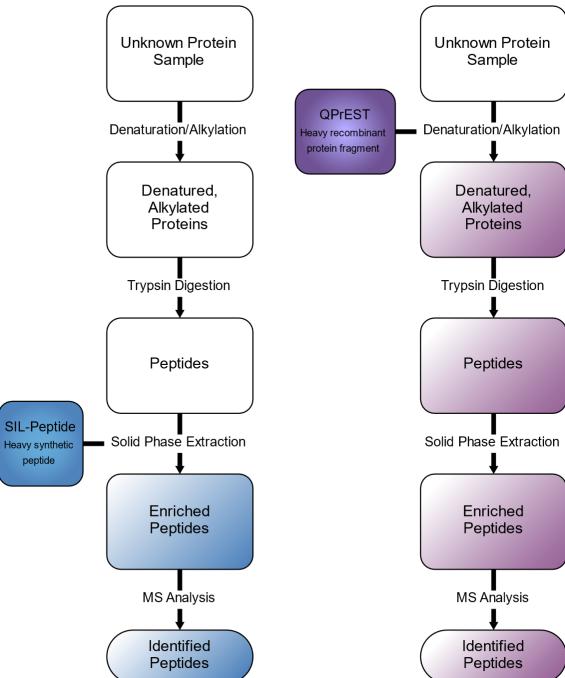
- Labeled with stable isotopes to introduce a small mass shift
- ^{13}C and ^{15}N most common
- Heavy Lys (+8 Da) or Arg (+10Da) most commonly used
- Have identical physiochemical properties as the light peptide, e.g. retention time, ionization, fragmentation

Peptides

- Chemical synthesis
- Many kinds of modifications possible
- Added late in the sample preparation

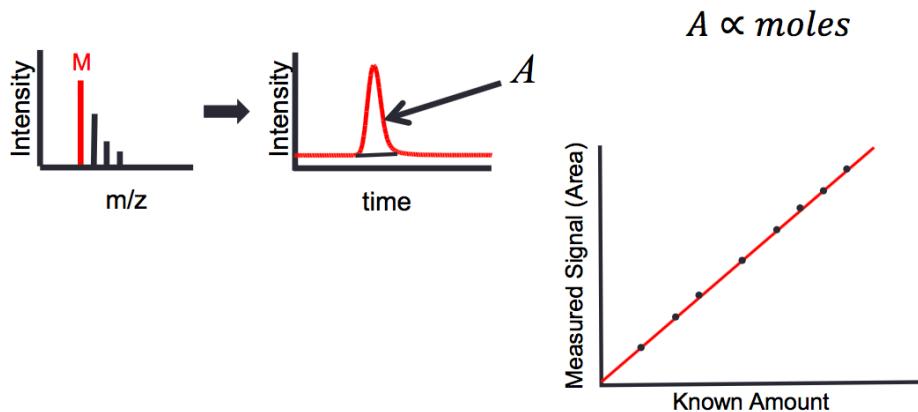
Proteins and protein fragments

- Metabolic labeling during protein expression
- Added early in the sample prep workflow
- Are enzymatically digested together with the sample

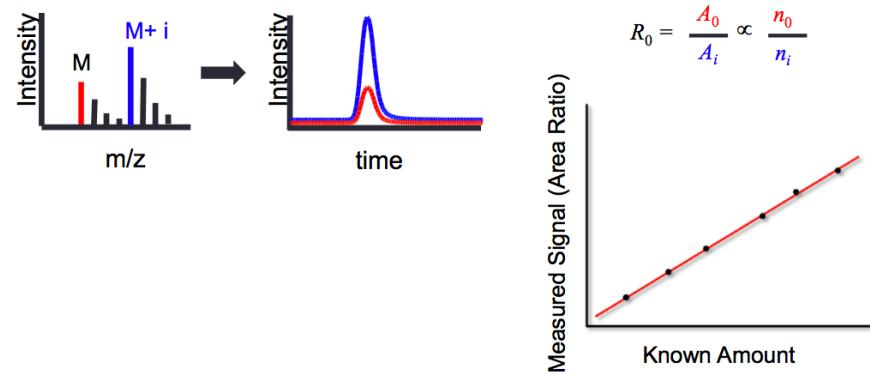


Quantification

Label free quantification
Analyte ($m/z = M$)



Quantification using isotope labeled standards
Sample ($m/z = M$).
Internal standard ($m/z = M + i$)



Accuracy and Precision

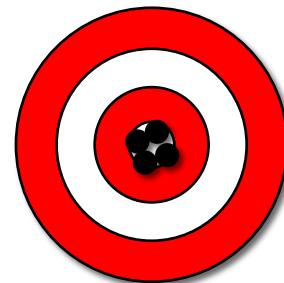
Accuracy

- How close we are to the real concentration
- Crucial when absolute quantities are needed

Precision

- How reproducible the quantification is
- Estimated by making replicate measurements
- Coefficient of variation (CV)

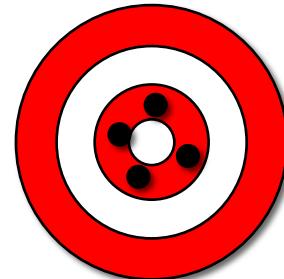
Accurate and precise



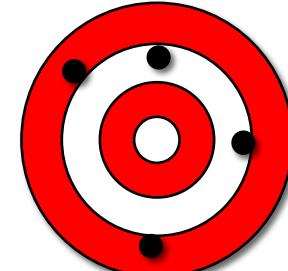
Precise, but not accurate



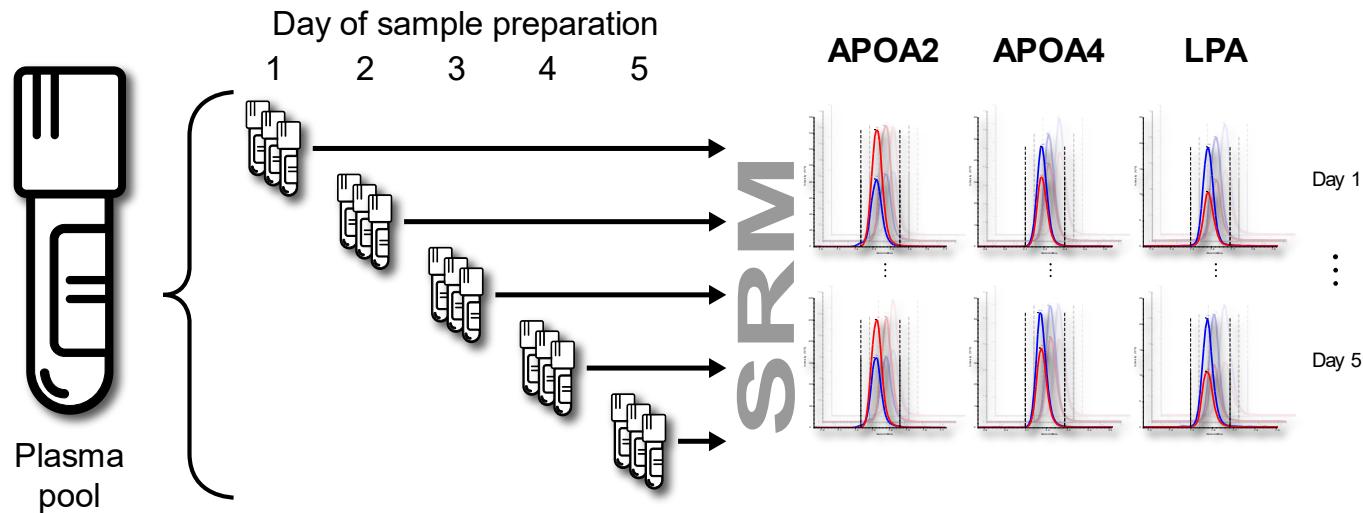
Accurate, but not precise



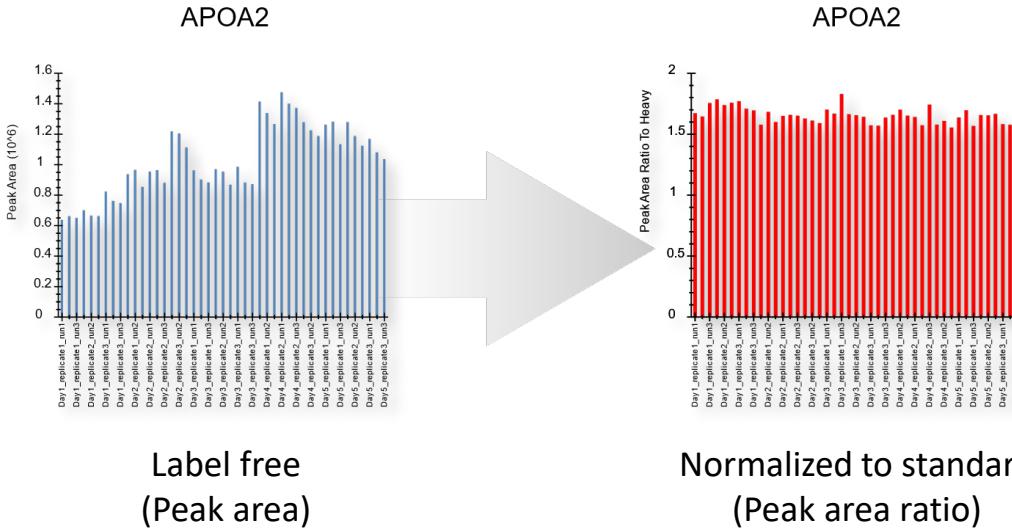
Neither accurate nor precise



Assay validation – Apolipoprotein SRM



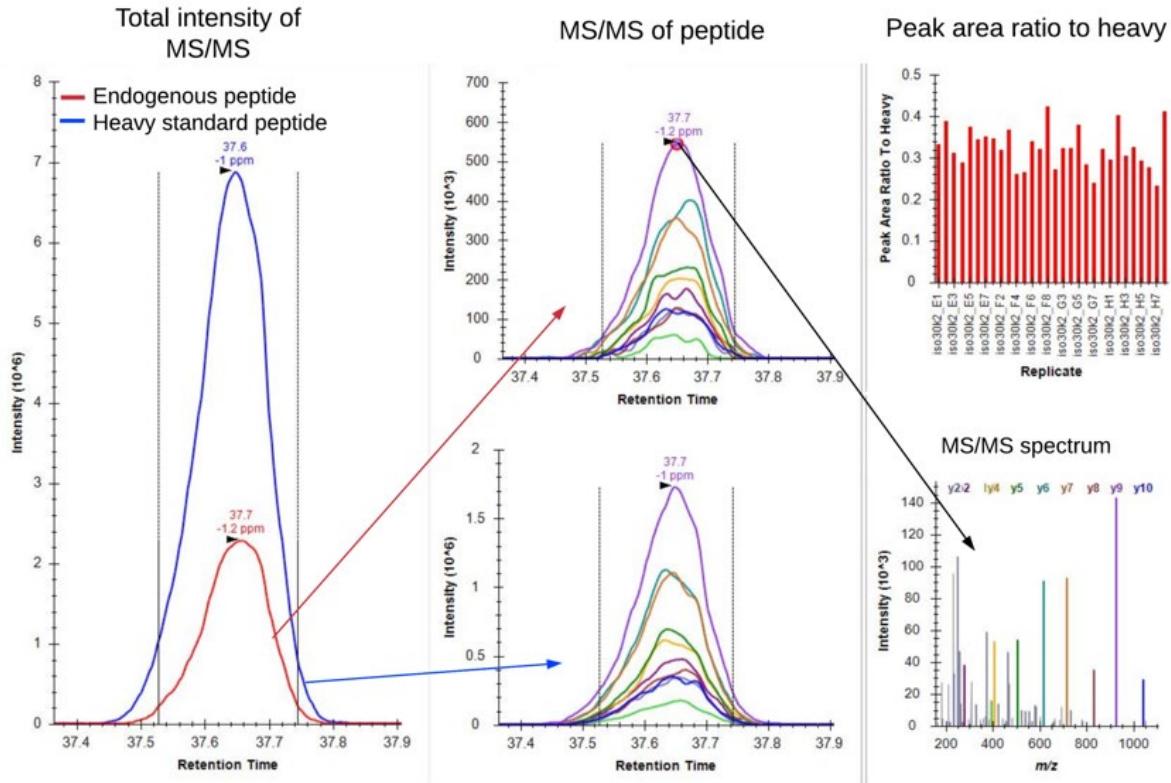
Assay validation – Apolipoprotein SRM



Robustness across five days for 13 proteins

- CV: 4.5-8.8 %
- Median CV: 5.9 %
- Label free CV: approx. 30 %

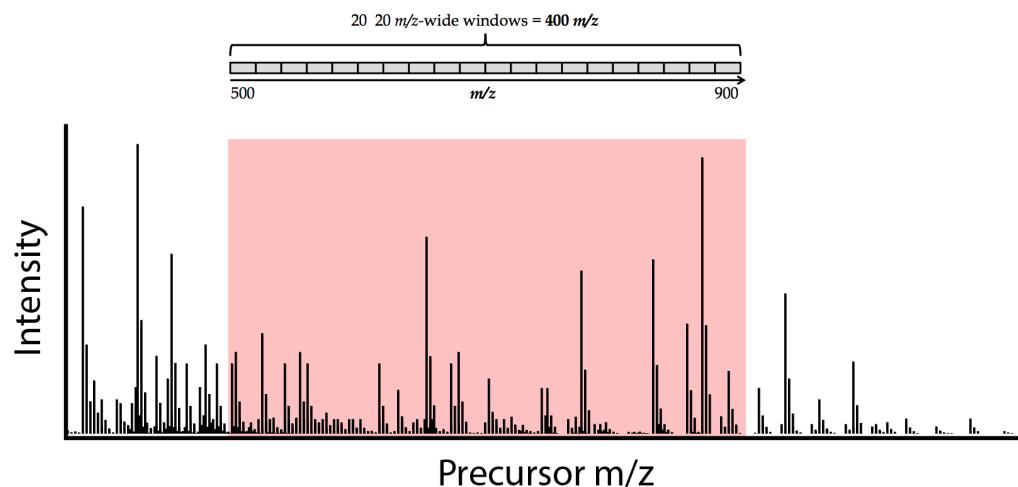
Targeted proteomics data analysis in Skyline



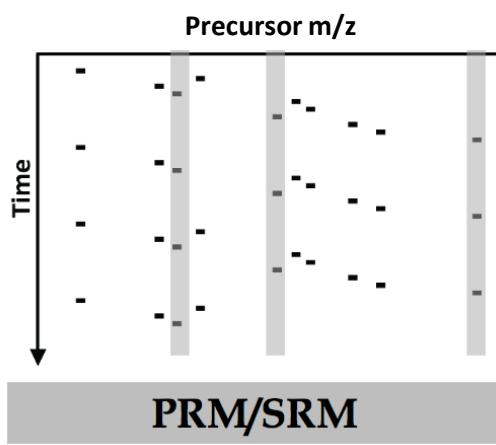
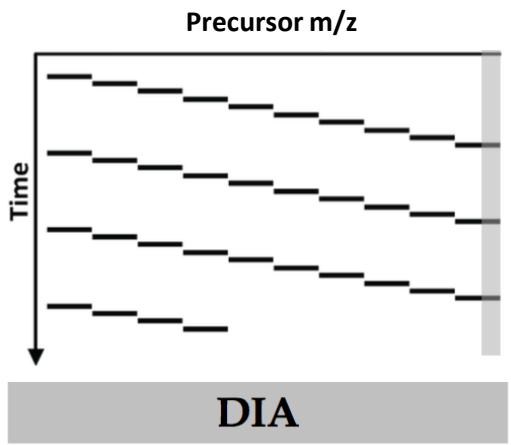
- Add list of the analyzed peptides
- Read raw data files
- Integrate peaks (automatic)
- Remove fragments with interference
- Export peak areas or peak area ratio to standard if isotope labeled standards were used

Data Independent Acquisition (DIA or SWATH)

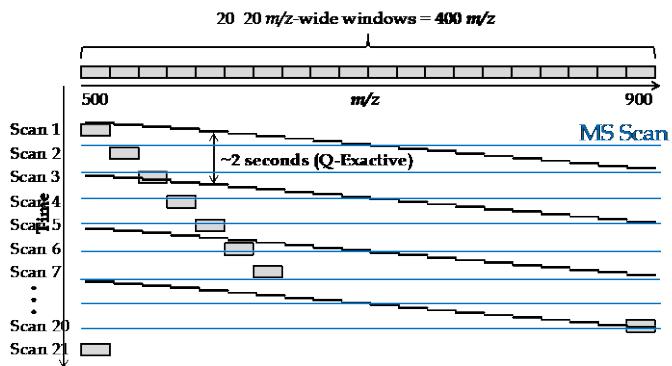
- DIA or Sequential Windowed Acquisition of all Theoretical Fragment Ion Mass Spectra (SWATH)
- Measure all fragment ions for all precursors within a limited mass range
- Can analyze thousands of peptides in MS2 in one single run
- Need known fragmentation patterns and RT for data analysis



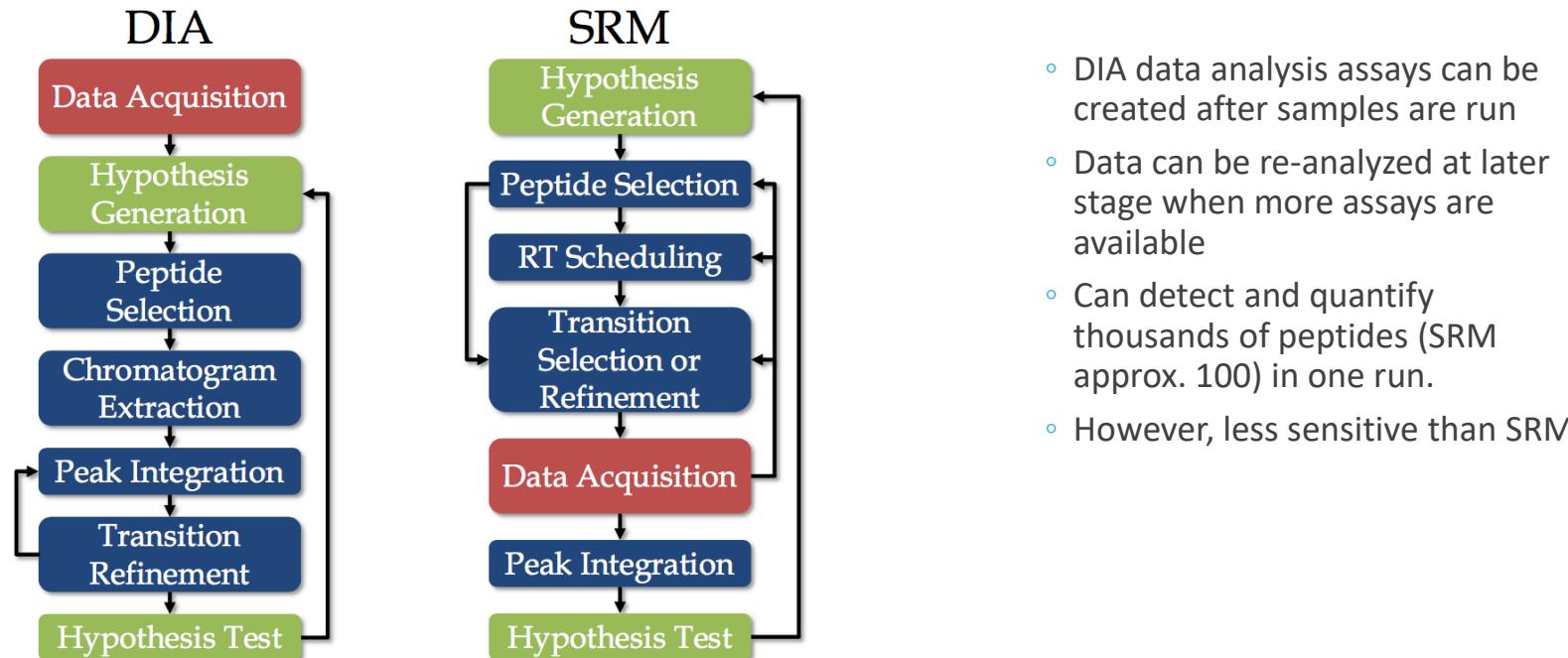
“SRM” on “All”



Comprehensive MS/MS Acquisition



Advantages of DIA over SRM / PRM



DDA vs DIA vs SRM

DDA vs. DIA vs. SRM

DDA



MS Quantification
on semi-random subset

DIA



MS/MS Quantification
on almost everything

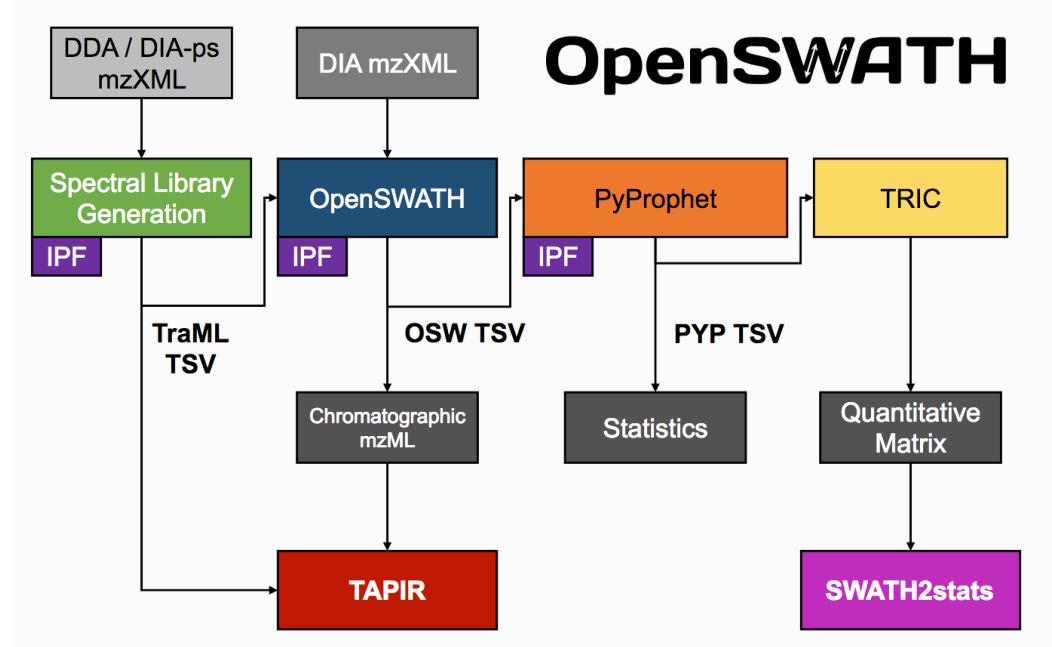
SRM/PRM



High Selectivity MS/MS
Quantification
on targets

DIA data analysis

- Need for spectral libraries to identify peptide peaks
- Fragmentation spectra with relative fragment intensities
- Indexed Retention Time (iRT)
- Statistical analysis to exclude miss identified peptides and proteins



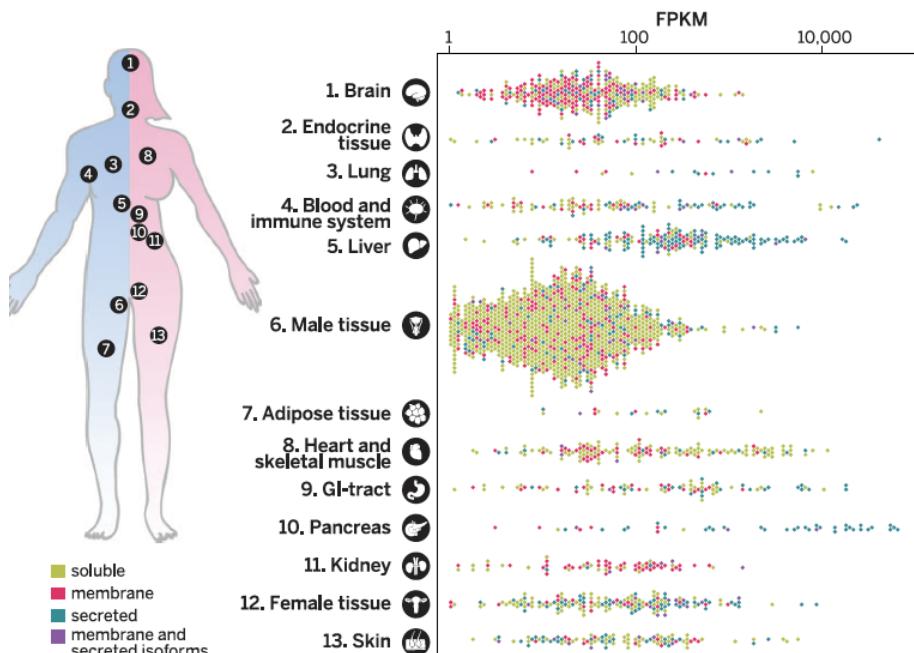
Tissue based map of the human proteome

PROTEOMICS

Tissue-based map of the human proteome

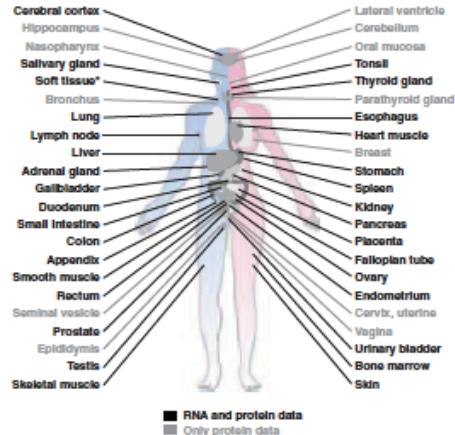
Mathias Uhlén,^{1,2,3*} Linn Fagerberg,¹ Björn M. Hallström,^{1,2} Cecilia Lindskog,⁴
Per Oksvold,¹ Adil Mardinoglu,⁵ Åsa Sivertsson,¹ Caroline Kampf,⁴ Evelina Sjöstedt,^{1,4}
Anna Asplund,⁴ IngMarie Olsson,⁴ Karolina Edlund,⁶ Emma Lundberg,¹ Sanjay Navani,⁷
Cristina Al-Khalili Szgyarto,² Jacob Odeberg,¹ Dijana Djureinovic,⁴
Jenny Ottosson Takanen,² Sophia Hober,² Tove Alm,¹ Per-Henrik Edqvist,⁴
Holger Berling,² Hanna Tegel,² Jan Mulder,⁸ Johan Rockberg,² Peter Nilsson,¹
Jochen M. Schwenk,¹ Marica Hamsten,² Kalle von Feilitzen,¹ Mattias Forsberg,¹
Lukas Persson,¹ Fredric Johansson,¹ Martin Zwahlen,¹ Gunnar von Heijne,⁹
Jens Nielsen,^{3,5} Fredrik Pontén⁴

Resolving the molecular details of proteome variation in the different tissues and organs of the human body will greatly increase our knowledge of human biology and disease. Here, we present a map of the human tissue proteome based on an integrated omics approach that involves quantitative transcriptomics at the tissue and organ level, combined with tissue microarray-based immunohistochemistry, to achieve spatial localization of proteins down to the single-cell level. Our tissue-based analysis detected more than 90% of the putative protein-coding genes. We used this approach to explore the human secretome, the membrane proteome, the druggable proteome, the cancer proteome, and the metabolic functions in 32 different tissues and organs. All the data are integrated in an interactive Web-based database that allows exploration of individual proteins, as well as navigation of global expression patterns, in all major tissues and organs in the human body.

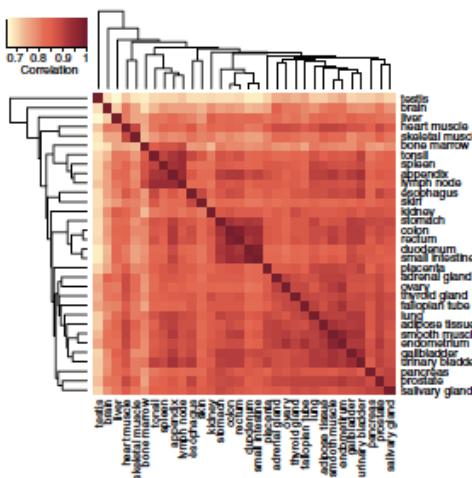


Classification of the human proteome

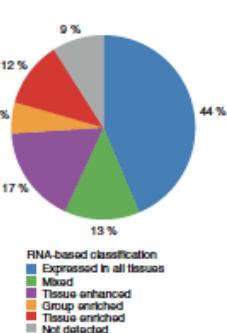
A



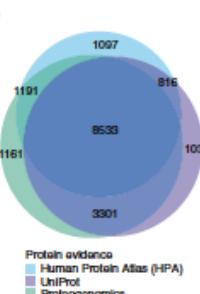
B



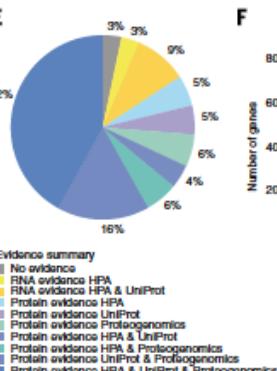
C



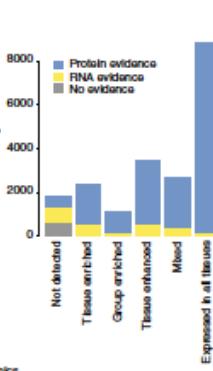
D



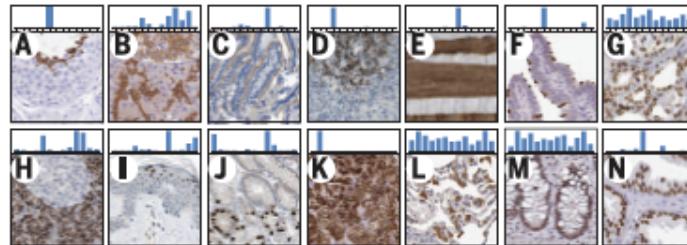
E



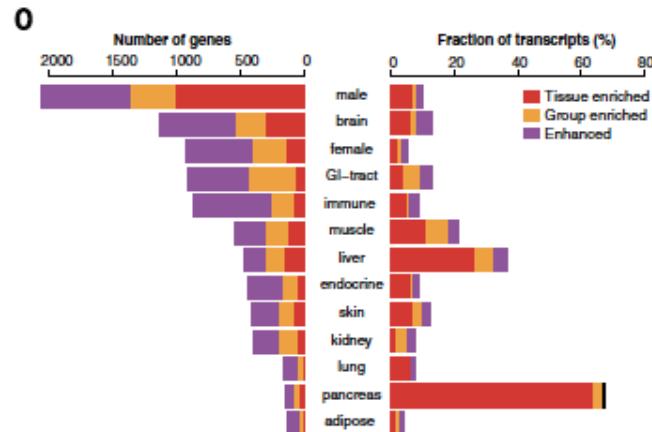
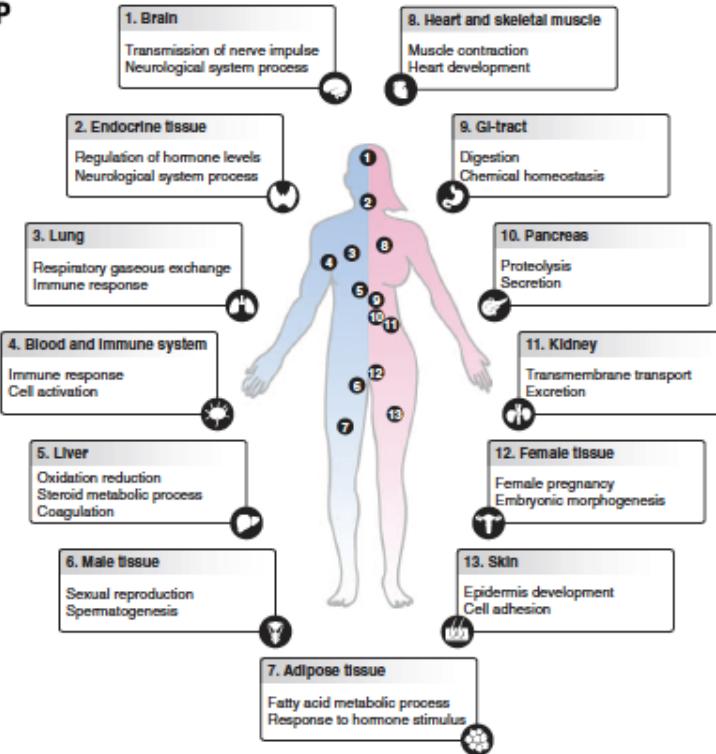
F



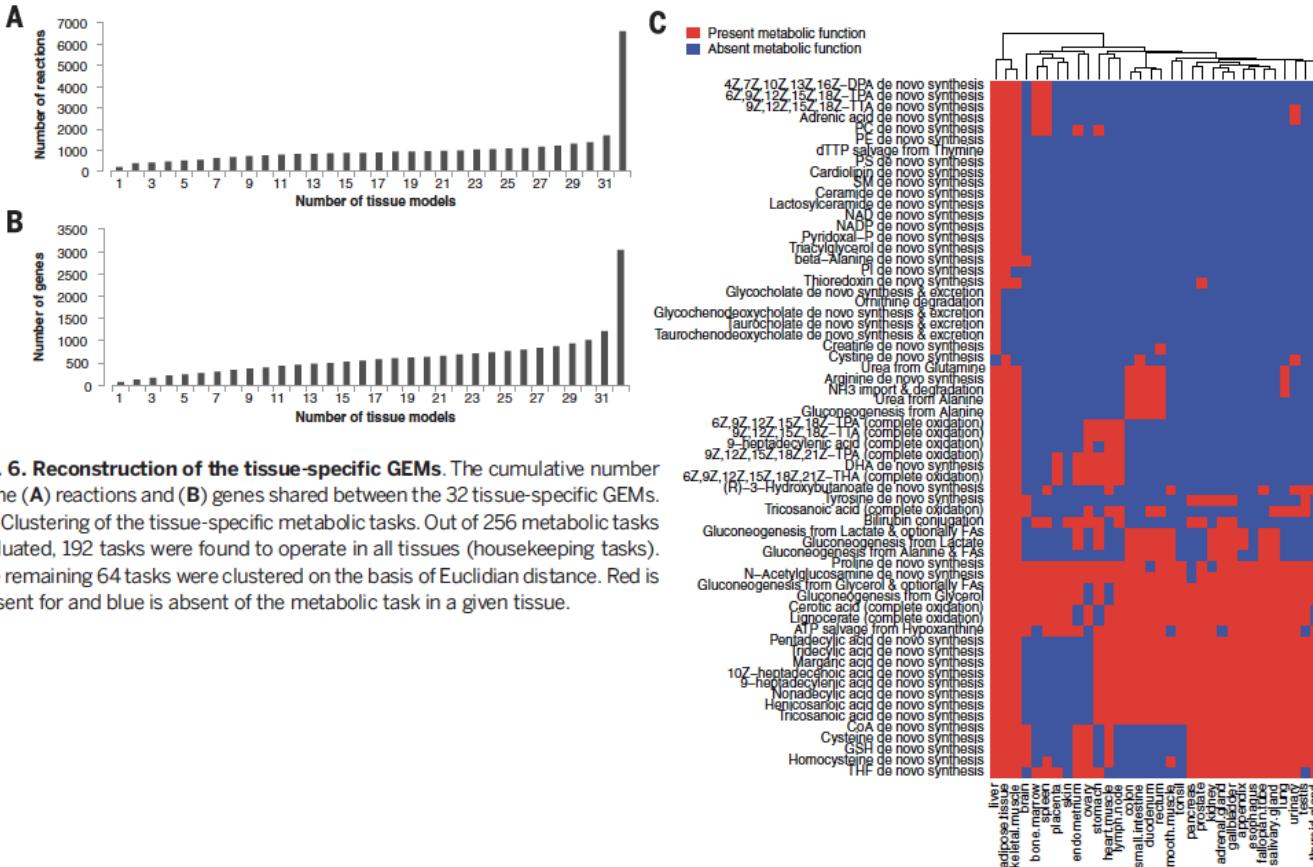
Analysis of Tissue elevated proteins



P



Generation of metabolic models



A deep proteome and transcriptome abundance atlas of 29 healthy human tissues

Dongxue Wang^{1,†} , Basak Eraslan^{2,3,†}, Thomas Wieland⁴, Björn Hallström⁵, Thomas Hopf⁴, Daniel Paul Zolg¹, Jana Zecha¹, Anna Asplund⁶, Li-hua Li¹, Chen Meng¹, Martin Frejno¹ , Tobias Schmidt¹, Karsten Schnatbaum⁷, Mathias Wilhelm¹, Frederik Ponten⁶ , Mathias Uhlen⁵, Julien Gagneur^{2,*} , Hannes Hahne^{4,**}  & Bernhard Kuster^{1,8,***} 

Abstract

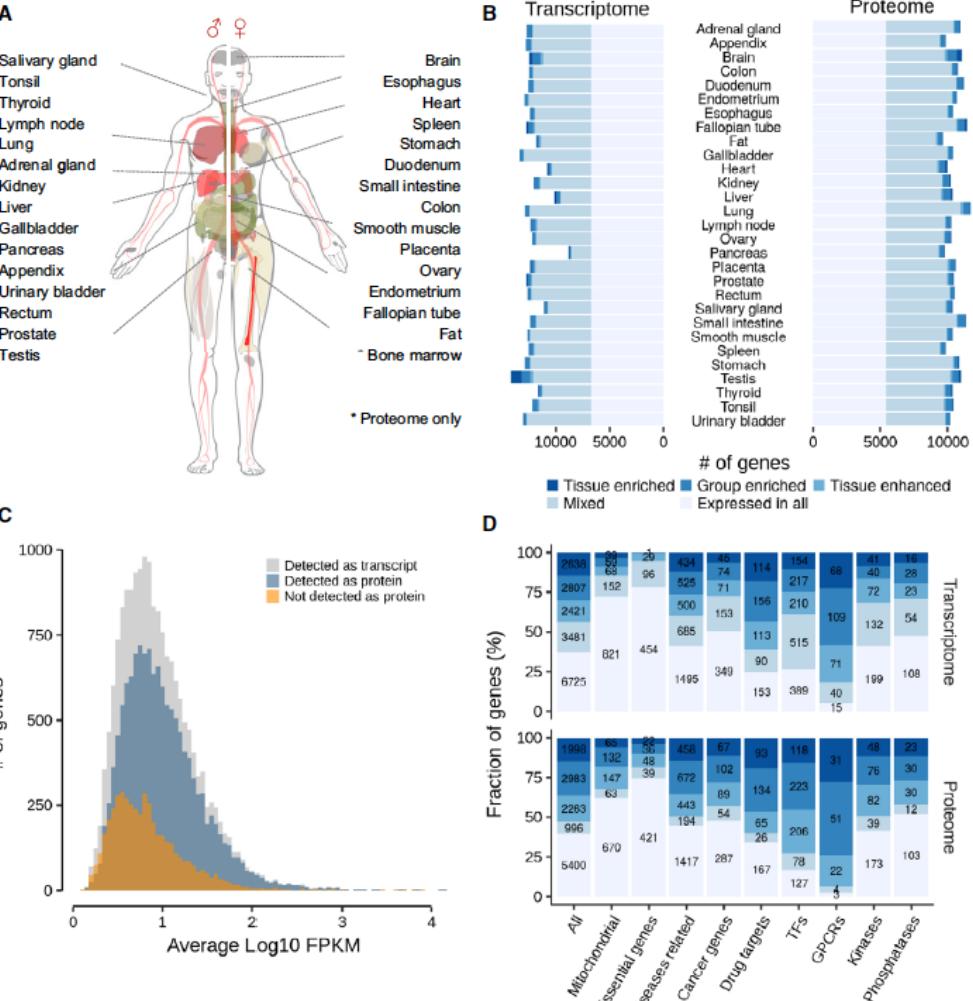
Genome-, transcriptome- and proteome-wide measurements provide insights into how biological systems are regulated. However, fundamental aspects relating to which human proteins exist, where they are expressed and in which quantities are not fully understood. Therefore, we generated a quantitative proteome and transcriptome abundance atlas of 29 paired healthy human tissues from the Human Protein Atlas project representing human genes by 18,072 transcripts and 13,640 proteins including 37 without prior protein-level evidence. The analysis revealed that hundreds of proteins, particularly in testis, could not be detected even for highly expressed mRNAs, that few proteins show tissue-specific expression, that strong differences between mRNA and protein quantities within and across tissues exist and that protein expression is often more stable across tissues than that of transcripts. Only 238 of 9,848 amino acid variants found by exome sequencing could be confidently detected at the protein level showing that proteogenomics remains challenging, needs better computational methods and requires rigorous validation. Many uses of this resource can be envisaged including the study of gene/protein expression regulation and biomarker specificity evaluation.

See also: B Eraslan *et al* (February 2019)

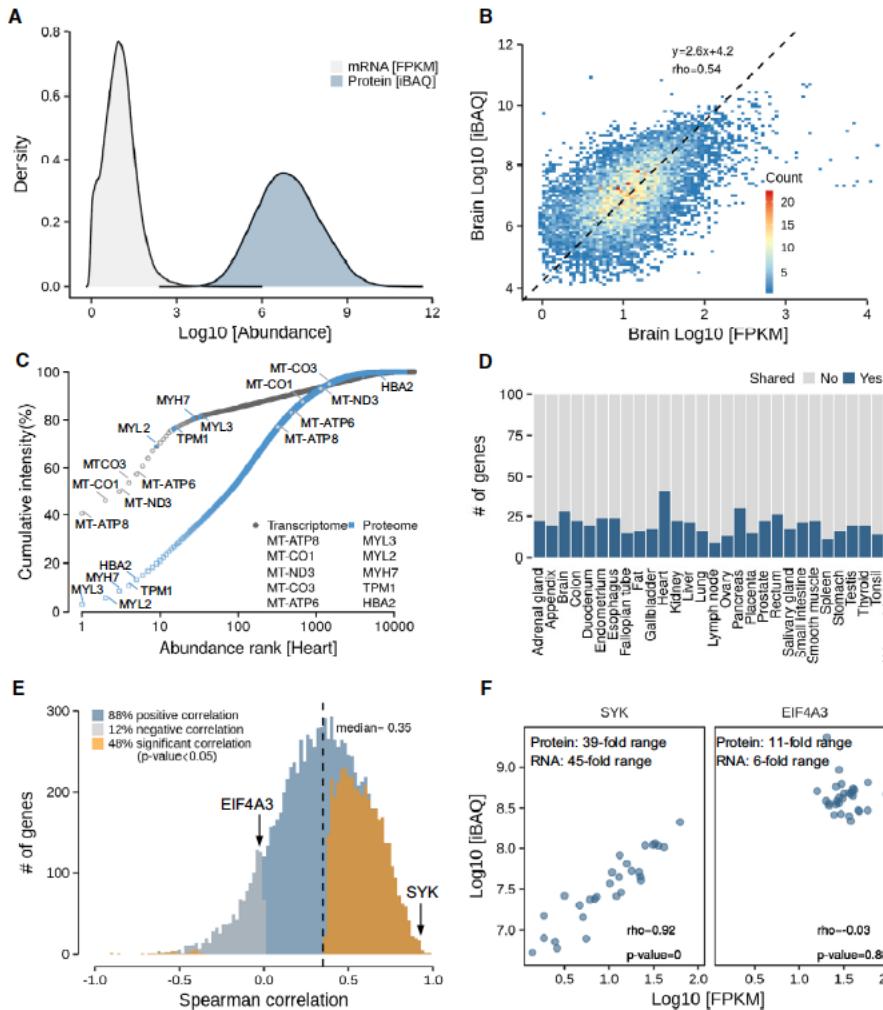
Introduction

Delineating the factors that govern protein expression and activity in cells is among the most fundamental research topics in biology. Although the number of potential protein-coding genes in the human genome is stabilizing at about 20,000, high-quality evidence for their physical existence has not yet been found for all and intense efforts are ongoing to identify these currently ~13% “missing proteins” (Ommen *et al*, 2017). While it is also generally accepted that the quantities of proteins vary greatly within and across different cell types, tissues and body fluids (Kim *et al*, 2014; Wilhelm *et al*, 2014), this has not been analysed systematically for many human tissues. Furthermore, it is not very clear yet how the many anabolic and catabolic processes are coordinated to give rise to the often vast differences in the levels of proteins. Messenger RNA levels are important determinants for protein abundance (Vogel *et al*, 2010; Schwahnhäusser *et al*, 2011), and extensive mRNA expression maps of human cell types and tissues have been generated as proxies for estimating protein abundance (GTEx Consortium, 2013; Uhlen *et al*, 2015; Thul *et al*, 2017). However,

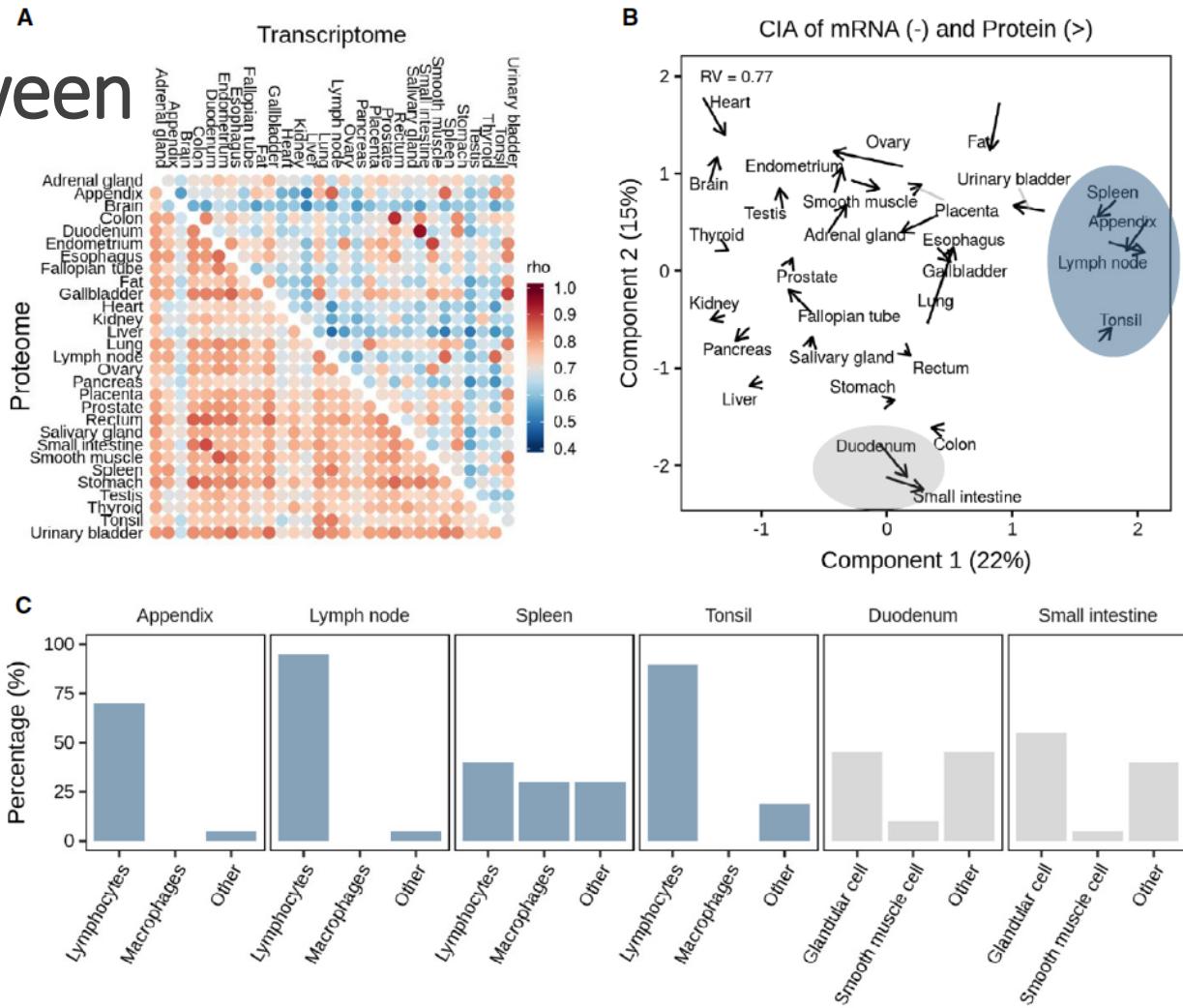
Proteomics and transcriptomics analysis of human tissues



Proteomics and transcriptomics expressions in human tissues



Correlations between Proteomics and transcriptomics expressions



Resource

A Quantitative Proteome Map of the Human Body

Lihua Jiang,^{1,3} Meng Wang,^{1,3} Shin Lin,² Ruiqi Jian,¹ Xiao Li,¹ Joanne Chan,¹ Guanlan Dong,¹ Huaying Fang,¹ Aaron E. Robinson,¹ GTEx Consortium, and Michael P. Snyder^{1,4,*}

¹Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA

²Division of Cardiology, University of Washington, Seattle, WA 98195, USA

³These authors contributed equally

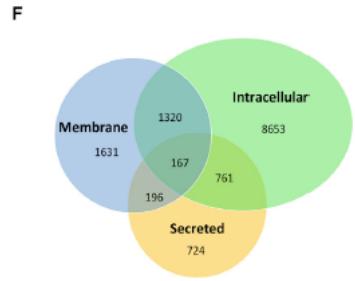
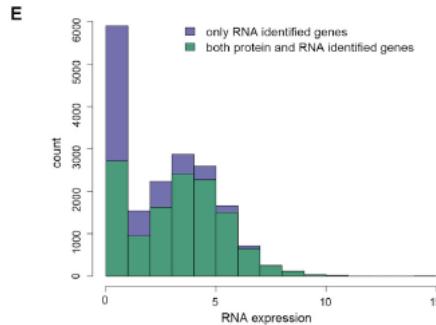
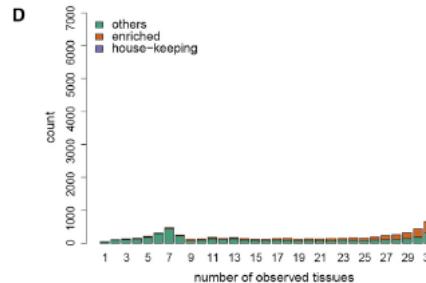
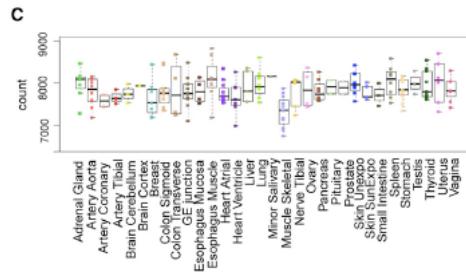
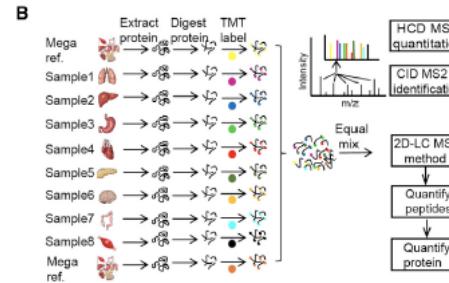
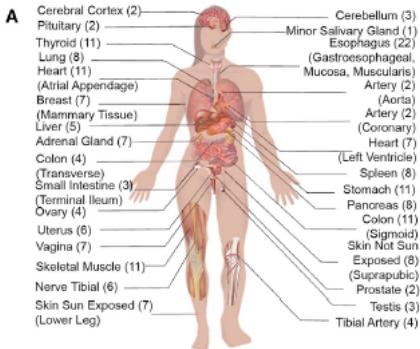
⁴Lead Contact

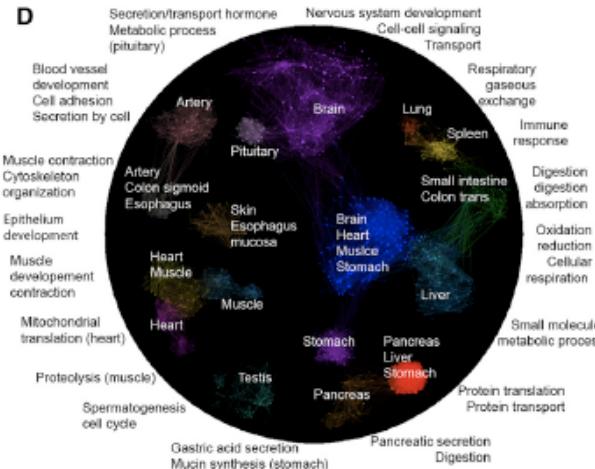
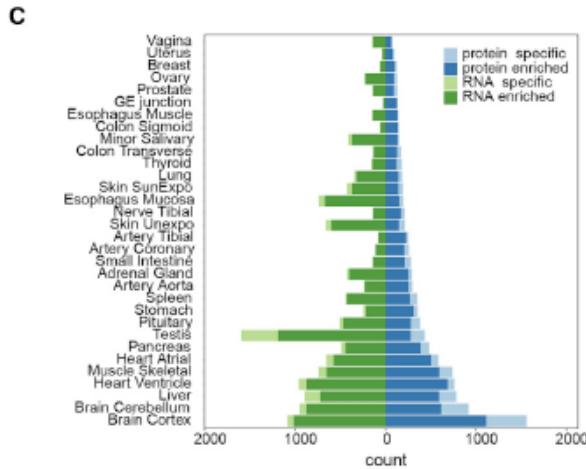
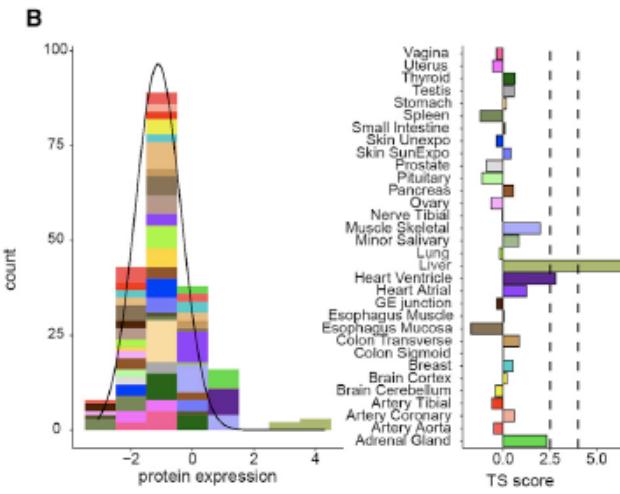
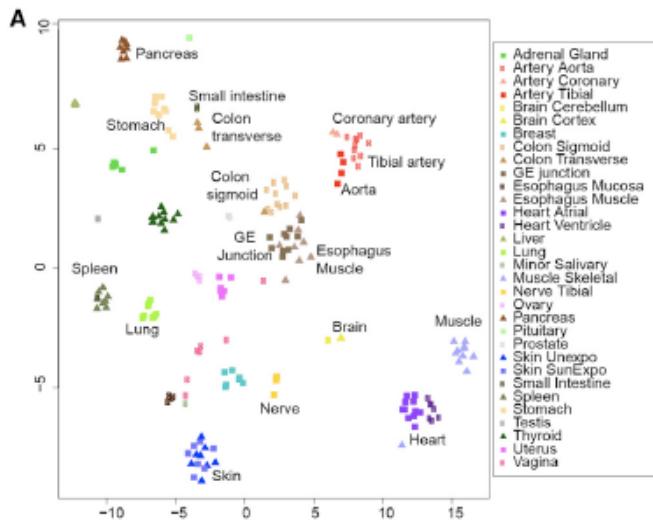
*Correspondence: mpsnyder@stanford.edu

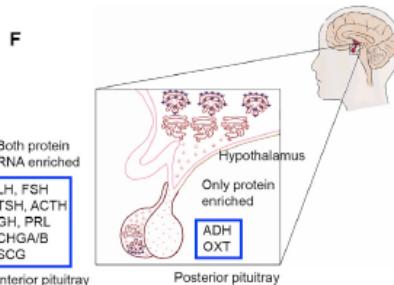
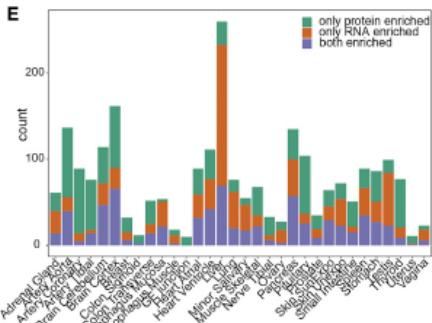
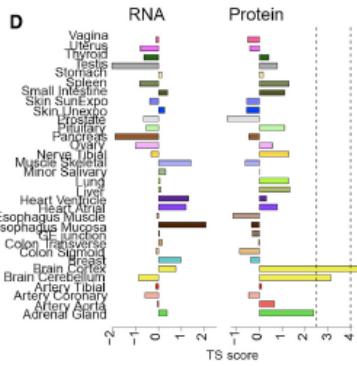
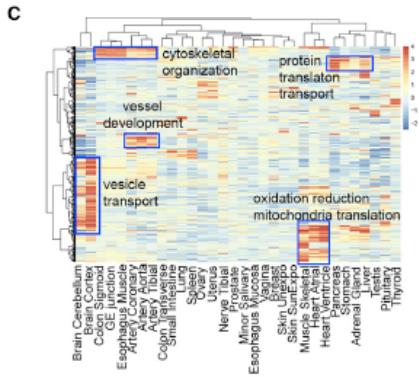
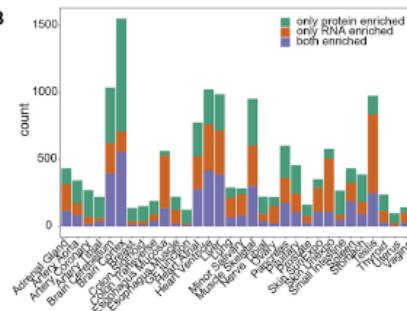
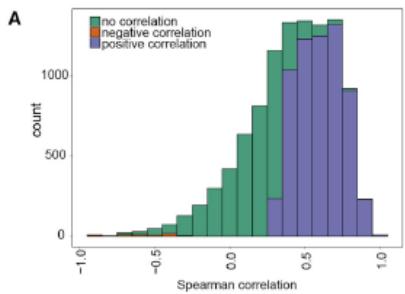
<https://doi.org/10.1016/j.cell.2020.08.036>

SUMMARY

Determining protein levels in each tissue and how they compare with RNA levels is important for understanding human biology and disease as well as regulatory processes that control protein levels. We quantified the relative protein levels from over 12,000 genes across 32 normal human tissues. Tissue-specific or tissue-enriched proteins were identified and compared to transcriptome data. Many ubiquitous transcripts are found to encode tissue-specific proteins. Discordance of RNA and protein enrichment revealed potential sites of synthesis and action of secreted proteins. The tissue-specific distribution of proteins also provides an in-depth view of complex biological events that require the interplay of multiple tissues. Most importantly, our study demonstrated that protein tissue-enrichment information can explain phenotypes of genetic diseases, which cannot be obtained by transcript information alone. Overall, our results demonstrate how understanding protein levels can provide insights into regulation, secretome, metabolism, and human diseases.







Dynamic human liver proteome atlas reveals functional insights into disease pathways

Lili Niu^{1,2} , Philipp E Geyer^{2,†}, Rajat Gupta^{1,‡}, Alberto Santos^{1,3,4}, Florian Meier^{2,§}, Sophia Doll^{2,†}, Nicolai J Wewer Albrechtsen^{1,5} , Sabine Klein^{6,7}, Cristina Ortiz⁶, Frank E Uschner^{6,7}, Robert Schierwagen^{6,7} , Jonel Trebicka^{6,7,8,*}  & Matthias Mann^{1,2,**} 

Abstract

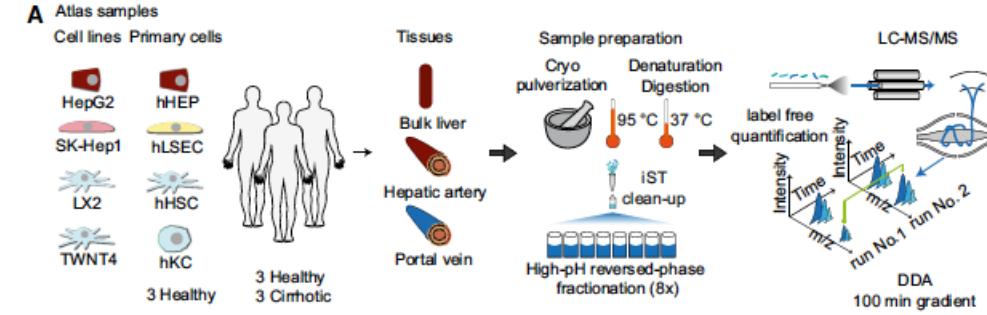
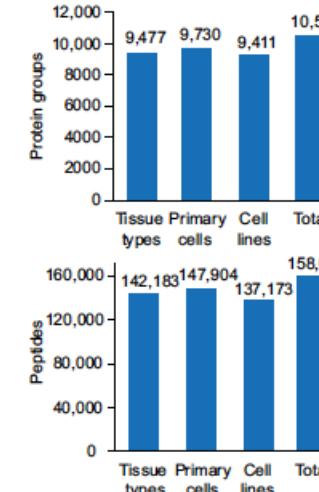
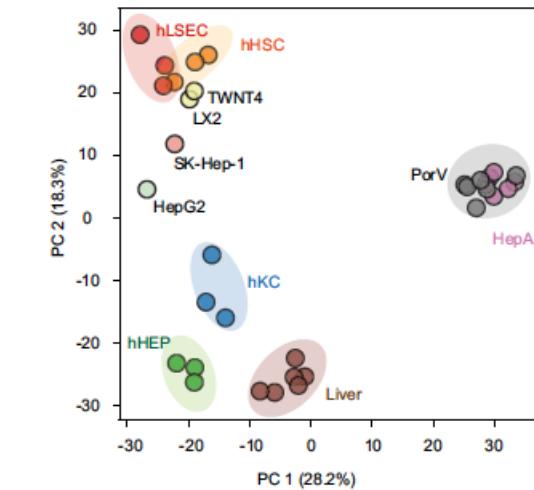
Deeper understanding of liver pathophysiology would benefit from a comprehensive quantitative proteome resource at cell type resolution to predict outcome and design therapy. Here, we quantify more than 150,000 sequence-unique peptides aggregated into 10,000 proteins across total liver, the major liver cell types, time course of primary cell cultures, and liver disease states. Bioinformatic analysis reveals that half of hepatocyte protein mass is comprised of enzymes and 23% of mitochondrial proteins, twice the proportion of other liver cell types. Using primary cell cultures, we capture dynamic proteome remodeling from tissue states to cell line states, providing useful information for biological or pharmaceutical research. Our extensive data serve as spectral library to characterize a human cohort of non-alcoholic steatohepatitis and cirrhosis. Dramatic proteome changes in liver tissue include signatures of hepatic stellate cell activation resembling liver cirrhosis and providing functional insights. We built a web-based dashboard application for the interactive exploration of our resource (www.liverproteome.org).

Introduction

The liver is essential for the human body's homeostasis and maintains a well-orchestrated network of parenchymal and non-parenchymal cell types, interconnecting the vascular and biliary system. While hepatocytes perform key metabolic functions, detoxification, and protein synthesis, the non-parenchymal cells provide a microenvironment for substance exchange and promote inflammatory and immunological responses (Kmiec, 2001; Shetty *et al.*, 2018). The liver is constantly exposed to gut-derived dietary antigens, microbial products and toxic substances such as alcohol, drugs, and excess lipids, all of which can induce liver damage. Chronic liver injury results in persistent hepatic inflammation, which can further progress to fibrosis and eventually cirrhosis—the common end-stage of chronic liver disease (CLD). CLD—including alcohol-related and non-alcoholic fatty liver disease (ALD and NAFLD)—is a major global health problem affecting approximately 1.5 billion people and causing more than two million deaths annually due to complications of cirrhosis and hepatocellular carcinoma (Loomba & Sanyal, 2013; Asrani *et al.*, 2019; Moon *et al.*, 2020). Liver disease is also important

Figure 1. In-depth characterization of the human liver proteome.

- A Overview of biological material used for generating the liver proteome atlas. (hHSC: hepatic stellate cell, hHEP: hepatocyte, hKC: Kupffer cell, hLSEC: liver sinusoidal endothelial cell, TWNT4 and LX2: immortalized human hepatic stellate cell line, SK-Hep-1: human hepatic adenocarcinoma cell line, HepG2: human liver cancer cell line). Number of biological replicates is $n = 6$ for bulk liver, hepatic artery and portal vein; $n = 3$ for hHEP, hLSEC, hHSC, hKC and $n = 1$ for HepG2, SK-Hep1, LX2 and TWNT4. No additional replications of the experiment was done in laboratory.
- B Total quantified proteome depth in tissues ($n = 18$), primary cells ($n = 12$), immortalized cell lines ($n = 4$) and all samples ($n = 34$). In all cases, n means biological replicates unless otherwise indicated. The upper and lower panel shows the number of quantified protein groups and peptides, respectively.

**B** Quantified proteins in the atlas samples**D** Principal component analysis

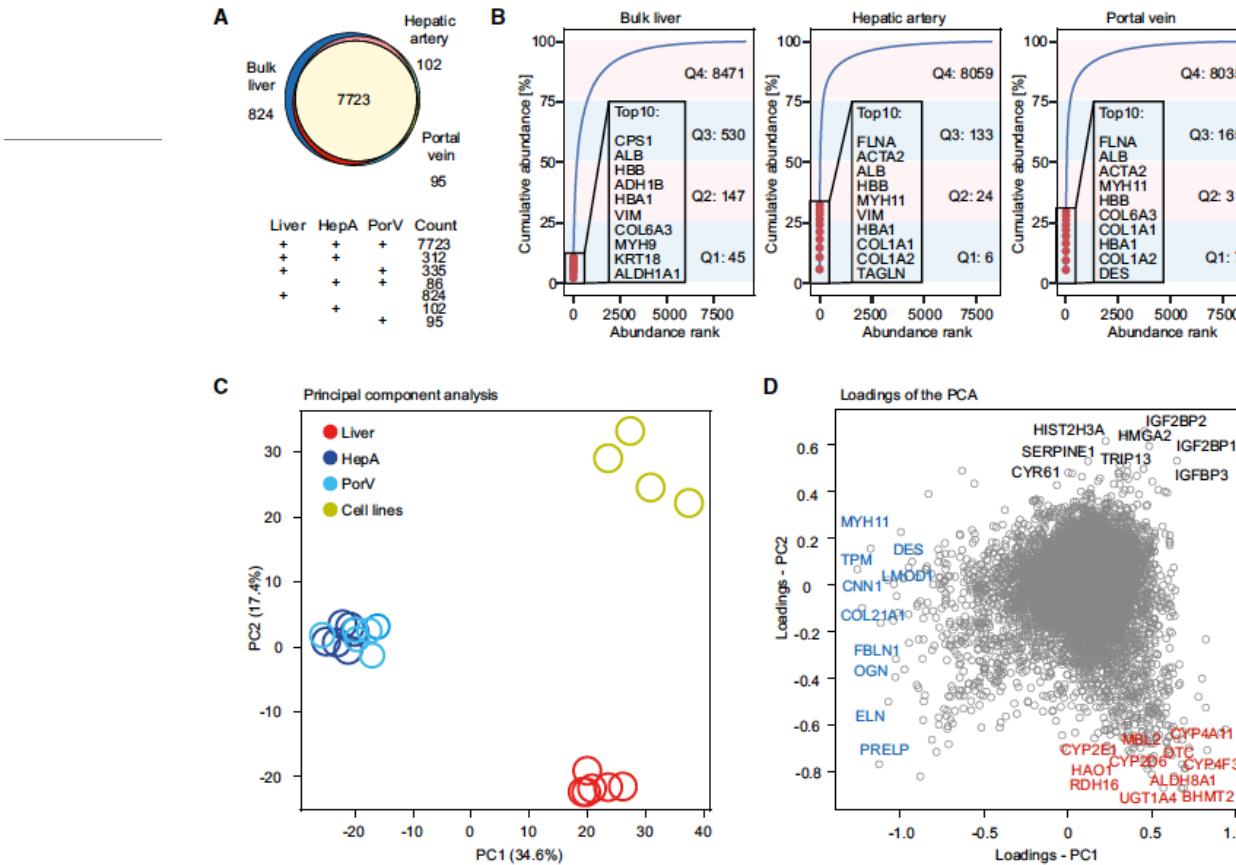


Figure 2. Comparative analysis of the liver tissue proteomes.

- A Commonly and exclusively quantified proteins in bulk liver tissue, hepatic artery, and portal vein.
- B Cumulative protein abundance of liver biopsy, hepatic artery, and portal vein as a function of protein rank, with the total of the top 10 abundant proteins and number of proteins that comprise four quartiles indicated.
- C PCA of liver, hepatic artery, portal vein, and cell lines.
- D Loadings of the PCA in Panel (C) with proteins that contribute most to the variance for the three clusters annotated.

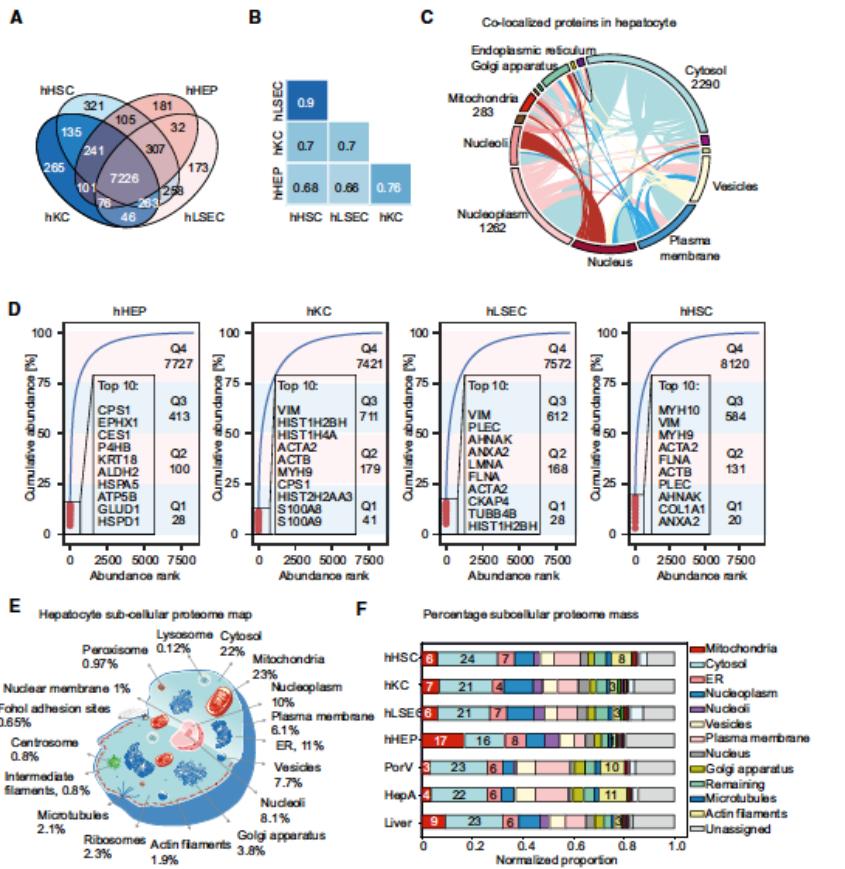


Figure 3. Comparative analysis of liver cell type proteomes.

- Commonly and exclusively quantified proteins in liver cell types (HHSC: hepatic stellate cell, hHEP: hepatocyte, hKC: Kupffer cell, hLSEC: liver sinusoidal endothelial cell).
- Pair-wise correlation of the proteomes of the four primary cell types, with Pearson correlation coefficients noted.
- Circos plot representing proteins predicted to be co-localized in subcellular compartments.
- Cumulative protein abundance of liver cell types as a function of protein rank, indicating the total of the top 10 abundant proteins and number of proteins that comprise the four quartiles.
- Schematic representation of the sub-cellular mass composition of an average hepatocyte.
- The bar plot shows the contribution of each organelle to total cellular protein mass, also accounting for unassigned proteins. Percentage of cytosol, mitochondria, nucleoli, plasma membrane, and actin filaments is indicated.

Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation

Andreas-David Brunner¹ , Marvin Thielert¹, Catherine Vasilopoulou¹, Constantin Ammar¹ , Fabian Coscia² , Andreas Mund² , Ole B Hoerning³, Nicolai Bache³, Amalia Apalategui⁴, Markus Lubeck⁴, Sabrina Richter^{5,6} , David S Fischer^{5,6} , Oliver Raether⁴, Melvin A Park⁷, Florian Meier^{1,8} , Fabian J Theis^{5,6}  & Matthias Mann^{1,2,*} 

Abstract

Single-cell technologies are revolutionizing biology but are today mainly limited to imaging and deep sequencing. However, proteins are the main drivers of cellular function and in-depth characterization of individual cells by mass spectrometry (MS)-based proteomics would thus be highly valuable and complementary. Here, we develop a robust workflow combining miniaturized sample preparation, very low flow-rate chromatography, and a novel trapped ion mobility mass spectrometer, resulting in a more than 10-fold improved sensitivity. We precisely and robustly quantify proteomes and their changes in single, FACS-isolated cells. Arresting cells at defined stages of the cell cycle by drug treatment retrieves expected key regulators. Furthermore, it highlights potential novel ones and allows cell phase prediction. Comparing the variability in more than 430 single-cell proteomes to transcriptome data revealed a stable-core proteome despite perturbation, while the transcriptome appears stochastic. Our technology can readily be applied to ultra-high sensitivity analyses of tissue material, posttranslational modifications, and small molecule studies from small cell counts to gain unprecedented insights into cellular heterogeneity in health and disease.

Introduction

In single-cell analysis, biological variability can directly be attributed to individual cells instead of being averaged over an ensemble or complex tissue (Regev *et al.*, 2017). While microscopy has always been single-cell based, specialized deep sequencing technologies have achieved this for systems biological approaches (Smith *et al.*, 2010; Ramsköld *et al.*, 2012; Jaitin *et al.*, 2014; Schnitzbauer *et al.*, 2017; Schaum *et al.*, 2018; Lundberg & Borner, 2019). At the level of proteins, the functional actors of cells, single cells are currently studied by antibody-based technologies, which are by necessity directed against previously chosen targets (Uhlén *et al.*, 2015; Stoeckius *et al.*, 2017; Jackson *et al.*, 2020). In contrast, mass spectrometry (MS)-based proteomics is unbiased in the sense that it measures all proteins within its range of detection (Larance & Lamond, 2015; Aebersold & Mann, 2016). Thus, it would be highly desirable to apply this technology to single cells if the required sensitivity and robustness could be achieved. Previous approaches that employed chemical multiplexing of peptides have labeled a small number of single cells but combined them with a dominant booster channel for MS analysis (Budnik *et al.*, 2018; Tsai *et al.*, 2020; Schoof *et al.*, 2021), which can hamper signal deconvolution

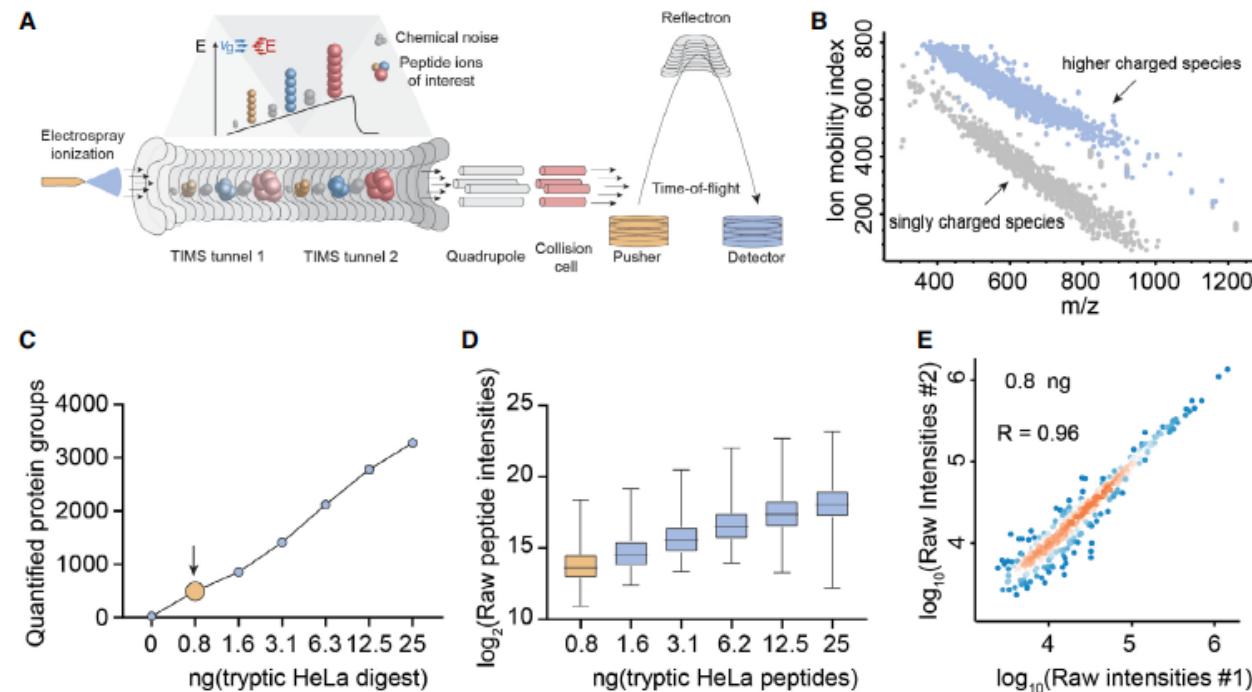


Figure 1. TIMS enables virtually noise-free spectra and ultra-high sensitivity proteomics.

- A, B** The TIMS-qTOF principle separating singly charged background peaks from multiply charged peptide precursor ions, making precursor ions visible at extremely low signal levels (0.8 ng HeLa digest).
- C** Quantified proteins from a HeLa digest dilution series from 25 ng peptide material down to 0.8 ng (arrow), roughly corresponding to the protein amount contained in three HeLa cells on our initial LC-MS setup (See Material and Methods).
- D** Linear quantitative response curve of the HeLa digest experiment in C (Box and Whiskers; The middle represents the median, the top and the bottom of the box represent the upper and lower quartile values of the data, and the whiskers represent the maximum and minimum value of the data).
- E** Quantitative reproducibility of two successive HeLa digest experiments at the lowest dilution (technical LC-MS/MS replicates).

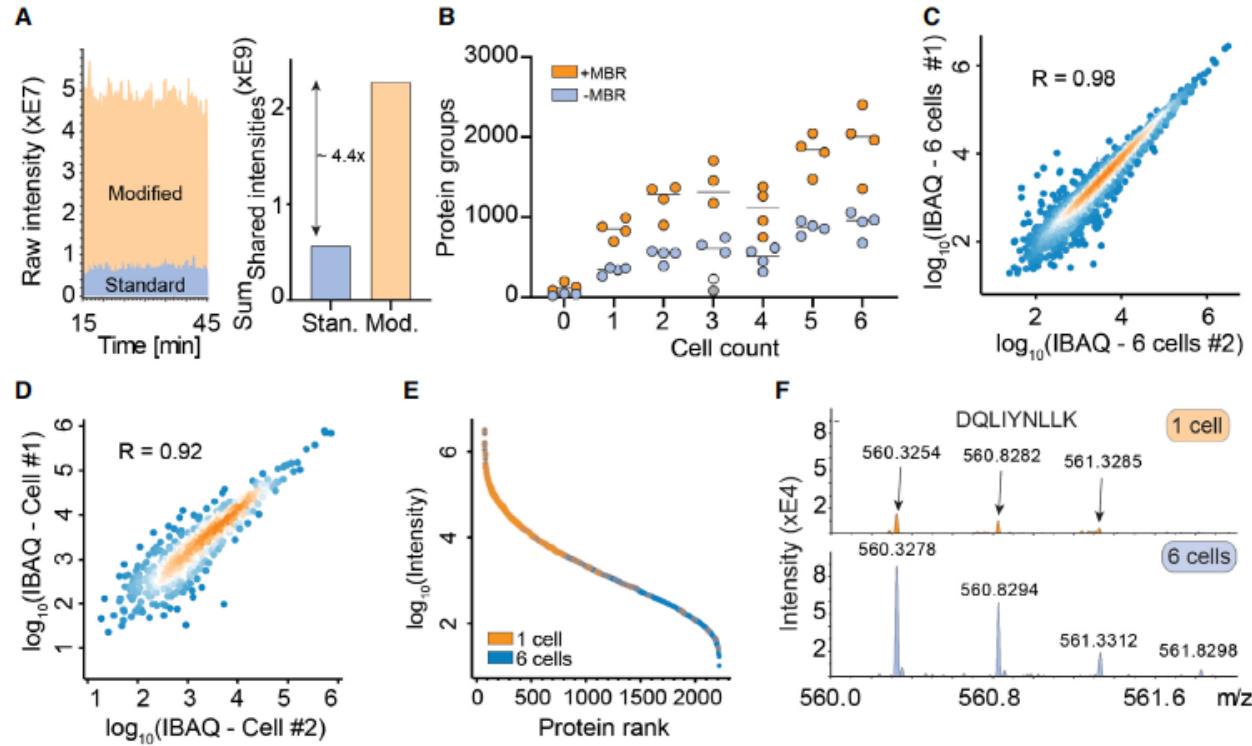


Figure 2. A novel mass spectrometer allows the analysis of true single-cell proteomes.

- A Raw signal increase from standard versus modified TIMS-qTOF instrument (left) and at the evidence level (quantified peptide features in MaxQuant) (right).
- B Proteins quantified from one to six single HeLa cells, either with “matching between runs” (MBR) in MaxQuant (orange) or without matching between runs (blue). The outlier in the three-cell measurement in grey (no MBR) or white (with MBR) is likely due to failure of FACS sorting as it identified a similar number of proteins as blank runs (Horizontal lines within each respective cell count indicate median values).
- C Quantitative reproducibility in a rank order plot of a six-cell replicate experiment.
- D Same as C for two independent single cells.
- E Rank order of protein signals in the six-cell experiment (blue) with proteins quantified in a single cell colored in orange.
- F Raw MS1-level spectrum of one precursor isotope pattern of the indicated sequence and shared between the single-cell (top) and six-cell experiments (bottom).

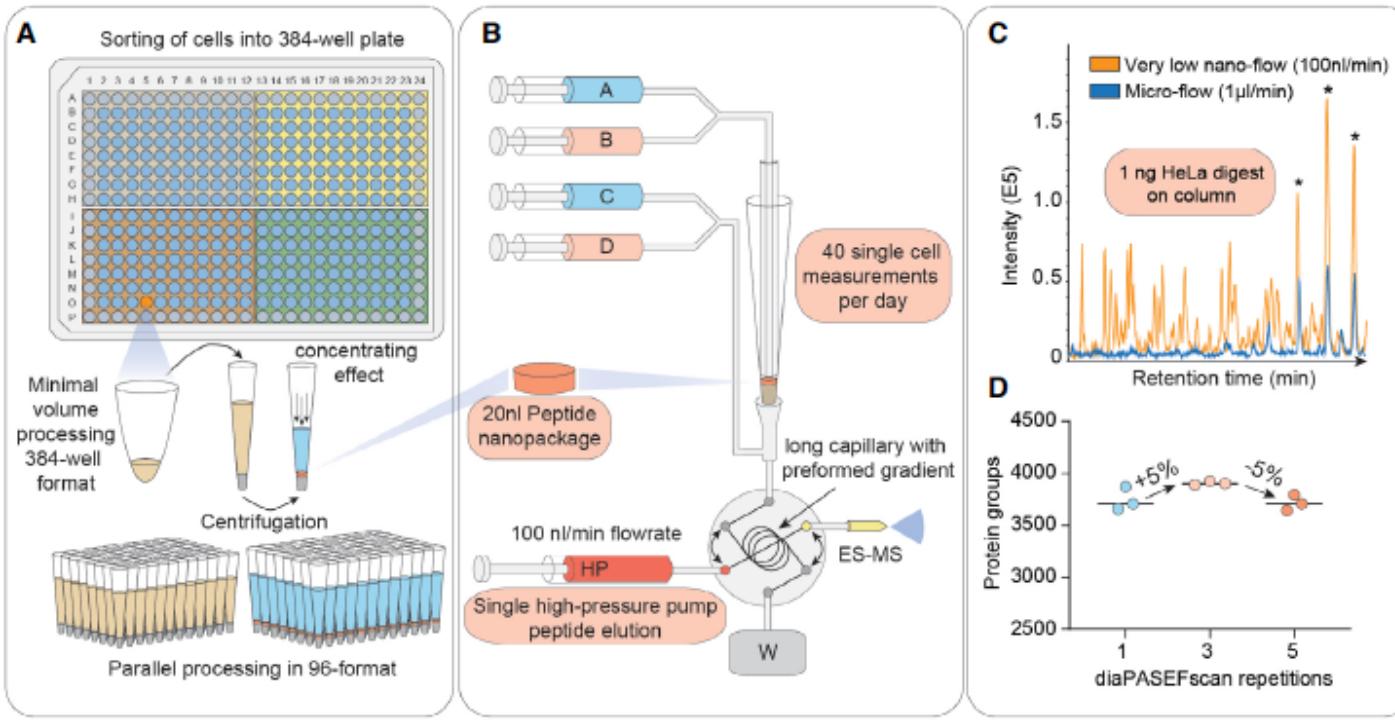


Figure 3. Miniaturized sample preparation coupled to very low-flow chromatography and diaPASEF.

- A Single cells are sorted in a 384-well format into 1 μ l lysis buffer by FACS with outer wells serving as qualitative and quantitative controls. Single cells are lysed and proteins are solubilized at 72°C in 20% acetonitrile, and digested at 37°C. Peptides are concentrated into 20 nl nanopackages in StageTips in a 96-well format.
- B These tips are automatically picked and peptide nanopackages are eluted in a sub-100-nl volume. After valve switching, the peptide nanopackage is pushed on the analytical column and separated, fully controlled by the single high-pressure pump at 100 nl/min.
- C Base-peak chromatogram of the standardized nanoflow (100 nl/min, orange) and microflow (1 μ l/min, blue) gradients with 1 ng of HeLa digest on the StageTip. Asterisks indicate polyethylene glycole contaminants in both runs.
- D Nanoflow (100 nl/min) and short-gradient diaPASEF method combined. Summation of one to five diaPASEF scan repetitions was used to find the optimum for high-sensitivity measurements at 1 ng of HeLa digest.

Figure 5. Single cells have a stable-core proteome but not transcriptome.

- A Gene or protein expression completeness per cell for T-SCP (Cells \times Proteins: 424 \times 2,480), SMARTseq2 (Cells \times Genes: 720 \times 24,990), or Drop-seq (Cells \times Genes: 5,022 \times 41,161) shown as violin plot; middle points represent the data set median.
- B Principal component analysis of single-cell gene and protein expression measurements (1,672 shared genes).
- C Heat map of cell-cell correlations across individual cells measured by proteomics and by both transcriptome technologies (1,672 shared genes).
- D Coefficient of variation of single-cell protein expression levels in LC-MS based proteomics as a function of mean expression levels with the “core proteome” colored in orange.
- E Boxplot of coefficient of variation of protein and transcript expression levels in LC-MS based proteomics, SMARTseq2, and Drop-seq technologies with a separate “core proteome” colored in orange (Box and Whiskers; The middle represents the median, the top and the bottom of the box represent the upper and lower quartile values of the data, and the whiskers represent the 1.5 \times IQR).
- F Rank order abundance plot for the core proteome with color-coded protein classes (Red: SUMO2 and TDP52L2 proteins; Turquoise: Chaperonin and folding machinery-associated proteins. Orange Translation initiation and elongation; Yellow: Structural proteins; Blue: DEAD box helicase family members).

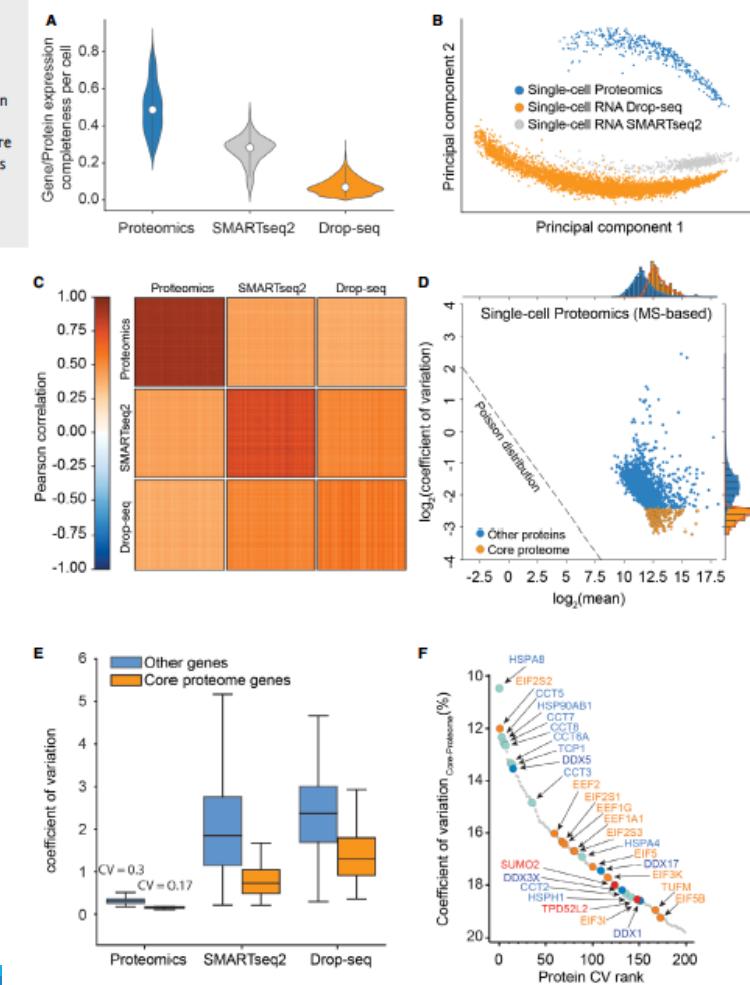


Figure 5.