Human SRMAtlas: A resource of targeted assays to quantify the complete human proteome

Ulrike Kusebauch[1], **David S. Campbell**[1], **Eric W. Deutsch**[1], **Caroline S. Chu**[1], **Douglas A. Spicer**[1], **Mi-Youn Brusniak**[1], **Joseph Slagel**[1], **Zhi Sun**[1], **Jeffrey Stevens**[1], **Barbara Grimes**[1], **David Shteynberg**[1], **Michael R. Hoopmann**[1], **Peter Blattmann**[2], **Alexander V. Ratushny**[1,6], **Oliver Rinner**[2,3], **Paola Picotti**[2], **Christine Carapito**[2], **Chung-Ying Huang**[1], **Meghan Kapousouz**[1], **Henry Lam**[4], **Tommy Tran**[1], **Emek Demir**[5], **John D. Aitchison**[1,6], **Chris Sander**[5], **Leroy Hood**[1], **Ruedi Aebersold**[2,7,*], and **Robert L. Moritz**[1,*]

[1]Institute for Systems Biology, Seattle, WA 98109, USA [2]Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, 8093 Zurich, Switzerland [3]Biognosys AG, 8952 Schlieren, Switzerland [4]Department of Chemical and Biomolecular Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, China [5]Computational Biology Center, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA [6]Center for Infectious Disease Research, Seattle, WA 98109, USA [7]Faculty of Science, University of Zurich, 8006 Zurich, Switzerland

Summary

The ability to reliably and reproducibly measure any protein of the human proteome in any tissue or cell-type would be transformative for understanding systems-level properties as well as specific pathways in physiology and disease. Here we describe the generation and verification of a compendium of highly specific assays that enable quantification of 99.7% of the 20,277 annotated human proteins by the widely accessible, sensitive and robust targeted mass spectrometric method selected reaction monitoring, SRM. This human SRMAtlas provides definitive coordinates that conclusively identify the respective peptide in biological samples. We report data on 166,174 proteotypic peptides providing multiple, independent assays to quantify any human protein and numerous spliced variants, non-synonymous mutations and post-translational modifications. The data is freely accessible as a resource at www.srmatlas.org, and we demonstrate its utility by examining the network response to inhibition of cholesterol synthesis in liver cells and to docetaxel in prostate cancer lines.

[*]Correspondence: Robert L. Moritz, PhD, Institute for Systems Biology, 401 Terry Avenue N, Seattle, WA 98109, USA, rmoritz@systemsbiology.org, +1-206-732-1244 and Ruedi Aebersold, PhD, ETH Zurich, aebersold@imsb.biol.ethz.ch.

Author Contributions

RLM and RA conceived the project. UK, DSC, EWD, PB, OR, PP and RLM designed and interpreted experiments. UK, DSC, EWD, CSC, DAS, M-YB, JSl, ZS, JSt, BG, DSh, MRH, PB, AVR, CCa, CCH, MK, TT performed experiments and bioinformatics calculations. DSC, EWD, HL, UK, ED, CS carried out database design, integration and deployment. UK, DSC, RA and RLM wrote the manuscript. All authors contributed to the preparation of the manuscript.

Introduction

The ability to accurately and reproducibly detect and quantify any protein of the human proteome is a main objective in the life sciences. Achieving it would significantly contribute towards understanding the biochemical base of living cells (Edwards, 2011). In contrast to the human genome which has been determined in its entirety, the composition of the human proteome is still poorly defined. The prevalence of alternative splicing and post-translational modifications increase the complexity to an as yet unknown number of different proteoforms, and the annotation of protein-coding regions and experimental evidence for their validity are still being refined. Therefore, a well-defined protein sequence database with annotated functional information and the identification and quantification of at least one protein from every protein-coding gene offer a pragmatic and useful definition of a complete proteome (Mann et al., 2013).

The detection of proteins can be accomplished through mass spectrometric and affinity reagent based methods. The Human Protein Atlas, a systematic exploration of the human proteome using antibody-based reagents, is a unique effort attempting to characterize all human protein-coding genes (Uhlen et al., 2005). Since the initial release in 2005, the Protein Atlas evolved into a knowledge base that includes a diverse collection of 25,039 monoclonal and polyclonal antibodies, collectively targeting 17,005 proteins corresponding to ~84% of the predicted proteome (version 15).

For the mass spectrometric exploration of the proteome, a range of techniques have been developed and they can be broadly grouped into data dependent acquisition (DDA, also known as shotgun or discovery proteomics) and targeted mass spectrometry (MS) methods. Common to both methods is that the sample proteins are first converted into peptides by enzymatic digestion. They differ in the manner in which the mass spectrometer (MS) is used to analyze the resulting peptide mixtures. The majority of proteomic studies rely on the DDA strategy that selects peptide precursor ions for collision-induced dissociation (CID) from signals detected in a survey scan. The resulting fragment ion spectra are assigned to a peptide sequence by peptide spectrum matching (PSM), and proteins are inferred from confidently identified peptides. This workflow allows the identification of thousands of proteins in a sample and provides quantification via the presence of stable isotope labeled reference peptides or through label-free methods (Bantscheff et al., 2007). However, the biased precursor selection of the most abundant peptide ions in complex samples by DDA limits the reproducibility of data generated in repeat analyses. Also, to reach high proteome coverage, enormous numbers of peptides need to be sampled which, in turn, causes significant technical and computational challenges at the level of PSM and protein interference (Deutsch et al., 2015b). Therefore, DDA MS is well suited to discover the components of a sample but less so for the generation of reproducible quantitative data across many samples.

Selected Reaction Monitoring (SRM, also named multiple reaction monitoring, MRM) instead is a targeted, quantitative technique that is characterized by a lower limit of detection, a wider dynamic range and increased reproducibility. SRM is primarily performed on triple quadrupole (QQQ) MS instruments where the first quadrupole (Q1) filters the

precursor ions of a peptide, the second quadrupole (Q2) provides CID and the third quadrupole (Q3) isolates predetermined fragment ions. This process results in a quantifiable signal represented as a chromatographic trace. SRM is referred to as targeted approach as only predetermined ions are measured. The two-level mass selection and the non-scanning mode translate into increased specificity and sensitivity and, in presence of stable isotope labeled standards, in precise quantification. The pair of mass to charge (*m/z)* values that is isolated in Q1 and Q3 is referred to as a transition, and a set of transitions that determine a peptide signature is, in combination with the peptide's elution time, termed SRM assay. SRM has the unique capability of rapidly quantifying targeted proteins, their variants and modifications through the detection of suitable proteotypic surrogates as a multiplexed and cost-efficient alternative to antibody-based assays. In addition, peptide affinity reagents can be used to specifically capture analytes and enhance sensitivity (Anderson et al., 2004). SRM has been applied for decades in the pharmaceutical industry to quantify small molecules (Baty, 1977) and evolved recently into an established technique in the field of proteomics, due to advanced technology and reproducibility across instrument platforms and laboratories (Addona et al., 2009).

However, SRM requires defining *a priori* a set of target proteins, optimal peptides and assay parameters. This is not a trivial task as not every peptide is suitable for SRM, and assays need to be experimentally determined from selected peptides. The major challenge of SRM is the initial effort to develop high-quality SRM assays, which is, despite all progress, still a time-intensive process. Once an assay is developed, it can be applied perpetually in a variety of studies.

Recently, we developed SRM assays for *Saccharomyces cerevisiae* (Picotti et al., 2009; Picotti et al., 2013), *Streptococcus pyogenes* (Karlsson et al., 2012) and *Mycobacterium tuberculosis* (Schubert et al., 2013) proteomes and successfully applied these assays to a wide range of protein studies in the respective species (Ebhardt et al., 2015; Picotti and Aebersold, 2012). We and several other laboratories have also developed SRM assays for human proteins, typically for a small number of proteins in the context of a specific biological study. The targeted approach has progressively been applied towards the quantification of low abundant proteins in complex matrices, the verification of biomarker candidates, and has proven to be successful in clinical settings (Craciun et al., 2015; Gillette and Carr, 2013; Hüttenhain et al., 2012; Kennedy et al., 2014; Surinova et al., 2015). However, for the human species no proteome-wide assay resource has been available and experimental protein research has therefore remained substantially limited in scope.

As a consequence of these factors the majority of protein research is still focused on the same relatively small subset of proteins for which assays are readily available. Strikingly, the population of proteins most frequently reported in the scientific literature has not changed significantly since the publication of the human genome and thus the definition, in principle, of the proteome (Edwards, 2011). This indicates that the realization of the benefits of genomic knowledge for experimental protein research critically depends on the availability of assays supporting the quantification of any human protein.

In the present study we developed a complete proteome-centric database for the targeted identification and quantification of any human protein of interest via SRM. We present a unique compendium of SRM assays for essentially the entire human proteome consisting of verified high-resolution, high mass-accuracy MS fragment ions of each proteotypic peptide, the chromatographic behavior of each peptide as an SRM trace and the relative quantitative response, all of which constitute an SRM assay. We have compiled the data into a freely available web-accessible database providing multiple SRM assays for each protein, integrated with extensive bioinformatic knowledge bases to establish a resource of assays to unambiguously identify and quantify any protein of the human proteome. We expect that this resource will significantly advance protein based experimental biology because any human protein can now, in principle, be quantified in any sample. We also expect that the availability of reliable assays for the human proteome will significantly contribute to increase the reproducibility of research results on the human proteome.

Results

To generate SRM assays for the entire human proteome we followed the process schematically illustrated in Figure 1. It consists of defining the target proteome, selection of proteotypic peptides, development of SRM assays via synthetic peptides and compiling the data into a web-accessible resource. Here, we describe each step of the process.

Step 1: Defining the Target Proteome

We used the 20,277 human protein sequences described in the manually annotated and reviewed UniProtKB/Swiss-Prot database (www.uniprot.org, release 2010-05) (Boutet, 2007) as reference to select peptides for each protein (see also Supplemental Information). We paid explicit attention to ensure that membrane bound proteins, large multi-domain proteins and protein activation events resulting in non-tryptic cleavage sites were equally represented. We extended this protein set to address known protein isoforms, peptides containing single nucleotide polymorphism (SNPs) and N-glycosylation sites. A database of 20,277 proteins and 14,677 isoforms formed the basis for the SRMAtlas (Figure 1, step one).

Step 2: Selection of Proteotypic Peptides

The selection of peptides that unambiguously identify each human protein is a key step in the development of SRM assays. We aimed to select at least the five best peptides for every human protein-coding gene using several criteria. Primarily we chose proteotypic peptides (PTPs) (Kuster et al., 2005) due to their high likelihood of being detected in subsequent measurements. We considered physiochemical properties including length, hydrophobicity and charge state, limitations in chemical synthesis, reactive amino acid residues susceptible to oxidation, pyroglutamate formation and deamidation and sequences that are likely modified by post-translational modifications or contain commonly occurring SNPs. These criteria are important and often overlooked in selecting PTPs, as SRM assay development and quantification generally depend on chemically synthesized peptides. To select the optimal set of PTPs that constitute the SRMAtlas we preferably relied on empirical data. For

those proteins for which no empirical data was available we computationally predicted the optimal set of peptides.

Selection of peptides from empirical data—PeptideAtlas (www.peptideatlas.org) (Desiere et al., 2004) is a major MS repository that accepts raw MS data acquired from biological samples generated by the scientific community and reanalyzes all MS data in a consistent process, including statistical validation of the results, using the Trans-Proteomic Pipeline (TPP) (Deutsch et al., 2015a). This database provides evidence of the most consistently detected peptides per protein and their confident detection at very low false-discovery rates (FDR, usually ~0.0002 at the PSM level corresponding to a 1% protein FDR). At the time we specified the peptide set underlying this study, the human PeptideAtlas (build 2010-05 internal) contained 106,184 distinct peptides identified in over 300 different experiments encompassing 59,142 MS runs of human cell lines, tissues and fluids.

In a first step we investigated how many of the 20,277 UniProtKB/Swiss-Prot proteins were represented by one to five distinct MS observed peptides in the human PeptideAtlas. 9,946 proteins were observed by at least one peptide while only 5,319 proteins (26% of the proteome) were identified by five or more distinct peptides. Taken together this demonstrated that observed peptides alone did not achieve full coverage of the predicted human proteome and that 51% of the human proteome had not yet been detected in MS approaches (Figure 2 blue bars, Table S1). Next, we screened the human PeptideAtlas for the best PTPs and annotated each PeptideAtlas observed peptide with an empirical suitability score (ESS). The ESS takes into account the peptide probability, the number of repeat identifications of the respective peptide and the selection criteria specified above (Figure 1, step two). The higher the ESS, the more suitable the peptide was deemed for assay development.

Selection of predicted peptides—For proteins that had no empirical evidence of being detected by MS or less than five PTPs in PeptideAtlas we used published and in-house algorithms to predict the best candidate peptides for assay development (Figure 1, step two). All algorithms provide a score based on the sequence and physicochemical properties of a peptide. Although each algorithm performed reasonably well, the set of peptides with the highest scores determined by each predictor overlapped less than expected. Therefore, we developed a predictive suitability score (PSS) that allowed us to computationally calculate MS suitable sequences for the entire human proteome (Sun et al., in preparation). Briefly, we retrained PeptideSieve (Mallick et al., 2007) and devised a composite scoring scheme considering the results of the individual algorithms to complement the observed peptides for assay development.

Subsequently, we applied the PeptideAtlas best SRM transition (PABST) algorithm (Deutsch et al., in preparation) to calculate an adjusted suitability score for both observed and predicted peptides by penalizing unfavorable sequence characteristics described above using a multiplicative weight scoring system. Finally, PABST ranked the adjusted scores of empirical observed and computational predicted peptides, and unique sequences with the highest scores were selected for SRM assay development (Figure 1, step two, Table S2).

This process was used to select at least five peptides per protein if allowed by sequence constraints. For higher molecular weight proteins (>50 kDa), we expanded the selection by dividing each protein sequence into 10 kDa segments and selected suitable peptides to provide assays for protein domains. For a small number of peptides we allowed less strict criteria with regard to length, hydrophobicity and charge state, to be able to select several peptides per protein and ensure as many proteins as possible are considered for SRM assay development. While we penalized sequences with unfavorable motifs and reactive amino acids, these were not entirely excluded, as otherwise assays could not have been developed for several proteins.

Step 3: Extension of Peptide Selection for Protein Isoforms, SNPs, and N-glycosylated proteins

To augment the assay development beyond a representative product of the 20,277 UniProtKB/Swiss-Prot proteins, we selected peptides identifying splice and sequence variants and N-glycosylation sites (Figure 1, step three). We attempted to select at least one peptide to specifically identify splice variants described in UniProtKB/Swiss-Prot Varsplic. Protein isoforms originating from differentially spliced versions of a particular mRNA are usually not characterized by several unique peptides, but with our selection approach we chose 11,309 peptides that allow the identification and quantification of unique splice forms. Further, we selected all suitable C-terminal peptides resulting in 6,820 additional peptides for the 20,277 proteins and 1,937 peptides for spliced variants. To account for sequence polymorphisms we extended the selection to include major SNPs resulting in non-synonymous mutations. We chose 3,662 peptides considering SNPs with a population frequency greater than 30% (=1831 SNPs) using NCBI dbSNP (build 131) and selected 3094 peptides (=1,547 SNPs) that fulfill the peptide selection criteria. To identify peptides representing N-glycosylated proteins, 5199 membrane proteins (Fagerberg et al., 2010), 1,748 secreted proteins (da Cunha et al., 2009) and 784 membrane proteins from 47 tissue types were used to select 10,938 peptides spanning N-glycosites located in the extracellular protein domain. Finally, for selecting peptides representing protein/peptide hormones, we targeted both the standard UniProt sequence as well as the mature form of these proteins considering their respective proteolytic cleavage sites to provide SRM assays for both the pre/pro-hormone and the activated form. We selected 142 peptides (124 distinct sequences) representing 129 proteins.

Overall, with this iterative and comprehensive selection process we determined 166,174 peptide sequences representing 20,255 proteins which constitute 99.9% of the predicted human proteome as defined in UniProtKB/Swiss-Prot. For 18,010 proteins (88.8% of the human proteome) we selected the best 5 PTPs, for 19,505 proteins (96.2%) we selected the best 3 PTPs and for 19,985 proteins (98.6% of the human proteome) the best 2 PTPs (Figure 2 red bars). Only 22 proteins remained inaccessible by tryptic peptides that pass the selection criteria and synthesis requirements, thus assays could not be developed for these proteins (Figure 2, Table S3).

Step 4: Development of SRM Assays and a Complete Human Peptide Library

To generate SRM assays, the peptides selected above were chemically synthesized and used to generate fragment ion spectra that were processed into consensus spectra and ultimately SRM assays.

Generation of fragment ion spectra—The 166,174 selected peptide sequences were individually chemically synthesized and used to generate high-resolution, high-mass accuracy reference fragment ion spectra. To process the large number of peptides we established an assay development pipeline including a robotics platform and multiple commonly used MS instruments duplicated at two geographical sites. Pools of 96 peptides each were analyzed on a quadrupole-time of flight MS (Agilent 6530 Q-TOF) in a data-directed approach using exclusive lists based on the expected charge state of a peptide as guidance for precursor selection. To increase the robustness of the fragment ion spectra we implemented a data acquisition strategy in which each precursor was fragmented exclusively at five different collision energies (CE) and at least five MS/MS spectra per CE were recorded (Figure 1, step four). The simultaneous acquisition of multiple CEs obviates the need for subsequent CE optimization, a time consuming aspect in the process of developing SRM assays. A set of peptides was used for strict retention time (RT) standardization across multiple MS instruments and to provide a catalogue with observed RTs and iRT values to enable multiplexed SRM analysis (Figure 1, Figure S1).

Generation of SRM assays—To convert the fragment ion spectra into SRM assays we subsequently generated consensus spectra. 3,250,015 spectra from the 6530 Q-TOF (base CE only) were confidently assigned to 149,265 peptide sequences out of the 166,174 synthesized peptides (89.8%). The five CE events for each peptide and their high-quality PSMs provided 14,970,896 spectra for use in monitoring differential fragmentation at multiple low and high CE values. We then generated consensus spectral libraries from each CE event to provide plots for every peptide and charge state that visualize optimal CE values for each individual fragment ion. The base CE, i.e. the CE value calculated from the default CE vs. precursor ion mass function, provided the highest abundance signal for the majority of fragment ions. However, for some fragments the selection of a lower or higher CE than the calculated base CE resulted in increased fragment ion signal intensities.

Next, we extracted from the 6530 Q-TOF base CE spectral library for each peptide and charge state SRM assay coordinates to acquire the peptides' chromatographic traces on a triple quadrupole MS (Agilent 6460 QQQ). Fragment ions with the highest signal intensities and above the precursor *m/z* were preferably selected to obtain assays with optimal sensitivity and selectivity. SRM chromatographic traces were successfully acquired for 126,712 peptides corresponding to a success rate of 84.9% based on the 6530 Q-TOF verified peptides which served as input to generate these SRM assays.

In addition, we determined the SRM signatures for all peptides on a quadrupole-linear ion trap MS (Sciex QTrap 5500) instrument by acquiring SRM traces and full MS/MS spectra upon the detection of a transition (Figure 1, step four). We generated a QTrap 5500 spectral

library from 1,789,651 high-quality PSMs that were assigned to 149,961 peptide sequences (90.2% of the human proteome).

SRM assay success—Whereas excellent peptide recoveries were achieved with each quadrupole type instrument, the combined recovery exceeded the results from each instrument type. In total, the recovery yielded 158,015 peptides with verified fragment ion spectra and SRM assay coordinates corresponding to 95.1% of all selected peptides (Figure 3A). Peptides of 7–20 amino acids length constitute 91.4% of the Human SRMAtlas and were identified with a 96% success rate, while peptides with 21–30 amino acids resulted in an 83% success rate in qualifying fragmentation spectra. Peptides with an expected precursor charge state (z) of 2 (61.3%), 3 (28.5%) and 4 (6.3%) were preferably selected and performed generally better compared to a small number of peptides which fragmented with $z = 1$ (C-terminal peptides) or z 5 (long peptides with several basic residues). Further, we found that peptides with an SSRCalc value of 7–46 performed best and that cysteine containing peptides showed a decreased success rate compared to all other peptide sequences (Figure 3B–G, Figure S2).

Next, we reassessed the protein coverage achieved by the 158,015 successfully developed SRM assays, taking into account that some peptides failed to result in the correct synthesis product or to fragment with sufficient quality. The generated assays covered 99.7% of the predicted human proteome with at least one SRM assay per protein and provide a minimum of three assays for 19,337 (95.4%) of all UniProtKB/Swiss-Prot annotated human proteins (Figure 4). We were able to develop a minimum of four assays for 91.5% and at least five assays for 85.3% of the proteome, respectively. Taken together, on average, each protein of the human proteome is represented by eight SRM assays per protein and some by more than 25 peptides. The assessment of the SRM assay chromatographic performance utilizing the 6460 QQQs was as successful. For 98.9% of the predicted human proteome we were able to acquire high-quality SRM traces with at least one peptide per protein while 90.3% of the proteome is represented by three SRM assays.

During the course of the project, updated versions of the UniProtKB/Swiss-Prot database were released. To account for new protein entries, we developed 443 additional SRM assays for 162 entries and included these assays in our database to provide SRM assays for updated human reference proteomes 2014 (20,193 proteins) and 2015 (20,203 proteins) (Figure S3, Table S4). Overall, we have successfully developed 158,015 mass spectrometric assays based on high-quality MS/MS spectra and subsequent QQQ deployment with the use of 166,174 chemically synthesized peptides to reliably identify and quantify essentially any human protein. The database of SRM assays can be adapted to changes in genome annotation with modest effort.

Assessing the Peptide Selection Success in the Context of Recent Public Data

Recent technical advances in high-resolution MS and efforts to discover complete proteomes of mammalian cells and tissues have led to a substantial increase in discovery proteomics data. The state of the human proteome as viewed through PeptideAtlas in 2015 (Deutsch et al., 2015b), which incorporates data from large scale proteomic measurements, reports ~133

million high-quality PSMs, identifying more than 1 million distinct peptides which collectively represent 14,070 (70%) confidently identified human proteins, 5% ambiguous and 9% redundant detections, leaving 16% (3,166 proteins) undetected. Given a large number of peptides discovered since the peptide selection for the SRMAtlas was performed, we retrospectively investigated the success of selecting suitable peptides that were observed in the recent PeptideAtlas and were not available in the initial selection database. We found that 84% of the newly observed peptides, that fulfill the selection criteria described above, were selected for the comprehensive human peptide SRM assay development. Further, we ranked all observed peptides based on spectral count and determined that we selected 85% of the most abundant peptides by using our predictive algorithm, indicating the robustness of the computational peptide selection algorithms used.

Step 5: Data Access through the Human SRMAtlas Resource

With the intent to facilitate life science research, we developed SRMAtlas (www.srmatlas.org), a freely available resource providing unlimited access to this unique compendium of targeted assays (Figure 1, step five). A web-interface allows researchers to query assays for their targets of interest. The query returns verified assays including peptide sequence, precursor and fragment ions with their charge states, fragment ion rank order, collision energy for different MS instruments, retention time, hydrophobicity and peptide uniqueness within the annotated human proteome. All MS/MS spectra and SRM chromatograms are displayed together with collision energy plots for optimal CE selection. We provide various assay download options such as instrument specific transition lists for immediate import to the MS method and subsequent acquisition. Default query settings are provided for ease of use and all queries can be customized. For workflows including the quantification with labeled standards, we implemented queries for transitions of the light endogenous peptide and the heavy isotope labeled analogue. The result page in the human SRMAtlas not only reports verified assay coordinates but also integrates with external knowledge bases including neXtProt, PeptideAtlas, the Human Protein Atlas, Pathway Commons and SRMCollider offering comprehensive information on a protein of interest.

SRMAtlas Application

To demonstrate the utility of the SRMAtlas resource, we carried out two studies. In study I, we chose cellular cholesterol regulation as an example for a clinically relevant pathway that can be perturbed using drugs. The transcription factor *SREBP2* induces expression of genes in the cholesterol biosynthesis pathway if the endogenous levels of cholesterol are depleted by inhibition with statins (Goldstein and Brown, 2015). A drug-induced gene module enriched for SREBP target genes was identified that also contained unknown targets of this pathway (Iskar et al., 2013). Hence, the objective of our test case was twofold: i) to perform, using the SRMAtlas as a resource, a systematic proteomic quantification of enzymes in the cholesterol synthesis pathway upon drug treatment and ii) to assess if the regulation of putative *SREBP* target genes identified in Iskar et al. translates to the protein level upon classical perturbation of the *SREBP* pathway with statin.

To test for differential drug-induced regulation of protein levels, we selected the two liver cell lines Huh7 and HepG2, treated them with lipoprotein-deficient serum (LPDS) and

atorvastatin, and subsequently quantified the relative abundance of target proteins using the assays obtained from the SRMAtlas. We targeted 64 proteins with the SRMAtlas assays and unequivocally quantified 1–3 proteotypic peptides for 33 proteins (74 peptides total) in unfractionated total cell lysate of Huh7 and HepG2 cells. After perturbation with LPDS and atorvastatin, 32 out of the 33 proteins showed regulation (Figure 5A, Table S5). This included detecting peptides for 18 out of the 22 enzymes in the cholesterol synthesis pathway (Figure 5B).

All enzymes of the cholesterol synthesis pathway, except of *LBR*, increased their abundance upon stimulation of *SREBP2*. *HMGCR*, *FDFT1*, and *DHCR7* showed the strongest response with an up to 16-fold increase in abundance. The measured SRM chromatograms of peptide TQNLPNCQLISR and LFSASEFEDPLVGEDTER from protein *FDTF1* show the difference in signal abundance between treated and untreated cells as representative examples (Figure 5C). The absence in regulation of *LBR* confirmed a previous report showing no increased *LBR* expression in HepG2 cells upon *SREBP* activation (Bennati et al., 2006) and *LBR* was also not present in the co-regulated module (Iskar et al., 2013). In addition to the proteins in the cholesterol synthesis pathway, additional 14 proteins that were part of the targeted gene module and present in other cellular pathways also changed substantially their expression. Most of these proteins are not established *SREBP* targets and this represents therefore an important confirmation of their regulation downstream of *SREBP*. Interestingly, the response in Huh7 and HepG2 cells differed for some proteins and thus the proteins could be divided into clusters based on their response to the drug treatment (Figure 5A). Cluster I consisted of proteins that were increased similarly in both cell lines and contained most of the enzymes in the cholesterol synthesis pathway. The proteins in cluster II were mostly regulated in HepG2 cells and contained proteins present in the mevalonate pathway, the first part of the cholesterol biosynthesis. Hence, using the SRMAtlas it was possible to efficiently profile the changes in protein abundance along a whole pathway, to confirm the co-regulation of novel putative *SREBP* target genes and to examine the relationship of protein regulation in different pathways.

In study II, we measured the effect of docetaxel treatment on three differentially responsive prostate cancer cell lines, LNCaP, DU145 and PC3. To select target proteins, we first determined a transcriptional time course response by microarray analysis. Prostate cancer is a leading cause of mortality in males. While many men present with localized and curable disease, a large number of deaths are driven by the development of metastatic prostate cancer and low curative options. Docetaxel is the first line drug treatment for metastatic castrate-resistant prostate cancer but ~50% of patients develop resistance (Antonarakis and Armstrong, 2011). Docetaxel acts mainly through the significant uptake in cells and the inhibition of microtubule function leading to mitotic arrest in the cell cycle and cell death. While the main action of taxanes like docetaxel is in overall cell cycle arrest, there are many unknown activities of the drug that contribute to its antitumor effects.

Performing microarray analysis of mRNA transcripts, we identified a dysregulated network of genes associated with docetaxel perturbation of the cell cycle and used SRMAtlas assays to target the corresponding protein products at four time points post-treatment (0–72h) and in untreated controls to investigate comparative proteome and mRNA transcript abundance

changes over time in each of these three cell lines. We targeted 36 proteins spanning nuclear proteins, cytosolic and membrane proteins with SRMAtlas assays and unambiguously quantified 33 proteins with 1–4 proteotypic peptides each (87 peptides total, *HIST2H2AC* through shared peptides) in unfractionated total cell lysates of these prostate cancer cell lines along the time course. Analysis of the combined mRNA and protein abundance data was overall in agreement with each other, showing larger abundance changes at 48 and 72h compared to 8 and 24h, but also highlighted the differences of the three cell lines and some discordance for mRNA and protein abundance involved in the cell-cycle response (Figure 6, Table S5). Cluster analysis also highlighted differences in mRNA and protein abundance especially for later time points (Figure S4). These include concordantly decreasing abundance in proteins involved in cell cycle, DNA repair and nucleotide base synthesis (*RRM2*, *RFC3*, *TMYS* and *MCM's* and *UBE2C*). Discordance between mRNA and protein abundance was observed for scaffolding and structural stabilization proteins (*KIF23* and *NUSAP1*). These results detail regulation within the cell to stabilize cells undergoing stress and the concordant reduction in proteins involved in normal cell cycle action. The difference in timescales shows the effect in transcript abundance occurring first, followed by reduction in protein abundance as would be expected in normal transcription/translation timescales. Of note, the discordance of the Kinesin KIF23 abundance at the protein level for DU145 and PC3 is in agreement with previous studies detailing kinesin overexpression and increased resistance to docetaxel in breast cancer (De et al., 2009) and glioma cells (Takahashi et al., 2012). Kinesins are key components in spindle movements during the cell cycle and can presumably meliorate the action of docetaxel. Inhibition of the kinesin complex or key members may interfere with the resistance mechanism to docetaxel and highlight a possible avenue for therapeutic intervention. Additionally, with the deployment of SRM assays for KIF23, this assay could be developed further to provide a new prognostic and diagnostic marker of therapeutic resistance. This analysis demonstrates the ease of rapidly deploying targeted quantitative assays to study discrete sets of proteins without relying on the vagaries of antibodies which is both laborious and costly.

Discussion

We generated high-quality fragmentation spectra and verified SRM assays for all accessible proteins described in the UniProtKB/Swiss-Prot database, a selection of spliced variants, non-synonymous SNPs and N-glycosylated proteins. This should enable reliable and reproducible quantification of all annotated proteins in the human proteome.

From a systems perspective, biological processes constitute networks of interacting molecules and changes in network state are informative about the biochemical state of the cell. To determine the state of a network and to compare network states between samples, the parts of a network need to be consistently detected in sample cohorts and quantified, at least at the level of relative quantification. An incomplete parts list may not explain observed processes. The Human SRMAtlas assays provide the tools to reliably navigate predicted or experimentally derived protein networks, to rapidly probe promising interaction partners and to perform relative or, in the presence of isotope labeled reference peptides, absolute quantification and to thus provide new insight into complex biological mechanisms.

The developed assays can be universally applied in different sample types to target any protein of interest and its abundance changes. The assays are beneficial in hypothesis-driven experiments focusing on a relatively small number of proteins, e.g. those that carry out a specific function or to probe a mechanism but also in large association studies to monitor protein panels across individuals. We demonstrated the utility of the resource in two studies and a multitude of different applications can be envisioned. Specific enzyme classes and signaling pathways can be studied. Proteins operating cooperatively in networks can be measured across many samples to define, for instance, their dynamic spatiotemporal changes in order to help identify the current health status or disease state. The assays can be used to investigate the response to system perturbations, enriched sub-proteomes, to confirm protein interaction partners or to analyze protein complexes to determine their stoichiometry. The SRM assays are particularly useful in a clinical setting where large volumes of genomic data suggest aberrations or dysregulation in biochemical pathways, thus providing functional hypotheses that are testable by the SRMAtlas resource. Current biomarker verification strategies rely primarily on the development of antibodies for western blots and ELISA tests, a time-intensive and expensive process which limits the verification of many candidate markers. Our assay resource can close this gap by allowing immediate assay implementation and subsequent high-throughput and cost-efficient verification of numerous markers in large cohorts of patient samples. We developed molecular assays with high specificity for the entire proteome as an emerging alternative and complement to conventional antibody screenings. While it can be challenging to verify the specificity of an antibody in discriminating antigenic variants, SRM provides high specificity by molecular determination through the proteotypicity of the selected peptides and several independent assays per protein.

N-glycosylated proteins are secreted or located on the cell surface and constitute a clinically interesting group of proteins that is investigated for biomarkers and drug targets. We developed assays to target N-glycosylated proteins by specifically selecting peptides that span the N-glycosylated sequence motif for use in studies involving N-glycosylation affinity approaches. SNPs have primarily been investigated in genome-wide association studies to classify different traits and gained interest as markers to diagnose diseases and predict drug response. To investigate SNPs resulting in non-synonymous sequence variants, we developed assays for the most frequent mutations which typically result in protein malfunction and may expose an altered phenotype.

Although the detection of target peptides by SRM is highly sensitive and specific, it can be challenging due to the complexity and large dynamic range in the human proteome, potentially resulting in interference of peptides with similar mass and chromatographic properties. Such interference can lead to a failure in detecting the target peptide. The SRM assays can be deployed in any sample but the occurrence of potentially interfering transitions needs to be assessed in each experiment to minimize false positives and imprecise quantification. Depending on the scope of a study, different levels of assay validation are recommended (Carr et al., 2014). While SRMAtlas assays are robust and powerful to assess the protein abundance in biological samples for research purposes, assays intended to support clinical decisions would require further clinical validation for robustness in large patient cohorts. Currently the technique cannot attain the entire range of proteins in higher

eukaryotes, therefore optimized protein extraction, enrichment and fractionation can further improve sensitivity.

While we present a truly comprehensive and highly characterized resource of assays for all human proteins, their variants and some post-translational modifications, the human proteome is complex and the coverage by empirical observed peptides is evolving and by no means complete. We provide very high coverage (>99%) at the protein level but the protein sequence coverage by peptides, in their native and post-translational modified form, will increase as more data become available. Our established pipeline will allow us to develop SRM assays for newly identified peptides and different post-translational modifications in future efforts with ease. The human SRMAtlas is based on the UniProtKB/Swiss-Prot database 2010 and 2015 but as the proteome continues to be refined, new assays can be added.

The human SRMAtlas not only facilitates the reliable identification of proteins, but also provides coordinates for the reproducible quantification of analytes. Absolute quantification of peptides can be accomplished by adding known amounts of stable isotope labeled standards. Label free quantification can also be performed by SRM but is only accurate for relative quantification given there is no standardization anchor point. Alternatively, absolute label-free quantification based on few anchor points can be pursued (Ludwig et al., 2012). Currently, the resource does not provide experimentally determined limits of detection (LOD) and limits of quantification (LOQ) for every SRM assay as these parameters strongly depend on the individual setting in which the assay is deployed (e.g., cell lysate versus plasma, type of sample preparation, chromatographic performance of LC system). In order to obtain the most accurate LOD/LOQ, these values need to be determined in the individual sample type in the context of each study. We introduced PASSEL (Farrah et al., 2012) as a repository for SRM data to share deployed assays and their performance in different matrices and similar databases followed (Sharma et al., 2014; Whiteaker et al., 2014). While SRM assays developed in different laboratories may be available as part of a publication or through these repositories, the information is scattered, time-consuming to gather, and currently limited to a small number of proteins (<1000). In contrast, the human SRMAtlas provides high-quality SRM assays including their spectral libraries for the entire predicted human proteome developed in a consistent manner. These assays are generic and are not based on any sample type or biological context but have proven to work in a number of settings (cell lines, plasma, urine protein digests, etc.).

The SRM assays presented here also provide a unique resource for efforts seeking to provide protein-level evidence of any human protein either previously observed or never observed to date to advance our knowledge in human biology and complex diseases. The high-resolution, high-mass accuracy MS/MS spectra generated from synthetic peptides constitute a 'gold-standard' fragmentation database which can provide additional confidence by spectral comparison with MS/MS spectra derived from discovery proteomic studies. Recently, the concept of SWATH-MS was introduced (Gillet et al., 2012), a data-independent MS technique which is less sensitive than SRM but capable of generating a digital map of a large fraction of a proteome. The analysis of SWATH data requires

information from high-quality fragment ion spectral libraries and assays such as the ones developed here to mine these complex fragment ion maps.

In conclusion, the human SRMAtlas provides verified MS assays based on SRM technology developed in a uniform and consistent process for essentially every protein of the human proteome. These assays can be rapidly deployed in systems biology and biomedical studies to identify and quantify any human protein with high sensitivity and high selectivity, and to navigate complete proteome maps to understand their biological functions.

Experimental Procedures

Peptide Selection

For every human protein in UniProtKB/Swiss-Prot release 2010-05 a set of PTPs was selected by mining PeptideAtlas (build 2010-05 internal) or by bioinformatic prediction. The selection of peptides for new protein entries in release 2014-11 and 2015-08 was based on PeptideAtlas build 2014-08. Each peptide obtained a suitability score using the PABST algorithm. Selection criteria include: fully tryptic, 7–30 amino acids, SSRCalc 10–46, expected z of 2 to 4 and unique within the human proteome. Peptide selection for spliced isoforms was based on UniProtKB/Swiss-Prot Varsplic (2010-06), for SNPs on the subset of NCBI dbSNP (build 131) entries annotated in the UniProt feature tables.

Peptide Synthesis

Peptides were synthesized via solid phase (Thermo-Fisher) or SPOT synthesis (JPT Peptide Technologies). 96 peptides were pooled and subjected to MS analysis.

MS Analysis

Peptides were analyzed on a G6530A nano HPLC Chip Cube Q-TOF LC-MS system (Agilent Technologies). Spectra were acquired in a data-directed approach using an exclusive precursor selection Auto MS/MS mode. Each *m/z* was fragmented over a wide range of CEs. Further, peptides were analyzed on a QTrap 5500 LC-MS with a Nano Spray Source III and Tempo nano MDLC (Sciex) in SRM mode, triggering the acquisition of a full MS/MS spectrum upon the detection of a transition.

Data Analysis

MS/MS spectra were searched with X!Tandem and Mascot against an artificial protein database consisting of the 166,174 selected peptides. The search results were validated with the TPP. Consensus spectral libraries were created with SpectraST. For each peptide precursor up to ten fragment ions from the 6530 Q-TOF consensus spectrum were extracted for transition verification.

SRM

Peptides were analyzed with the selected transitions on a G6460A nano HPLC Chip Cube QQQ LC-MS (Agilent Technologies). Data were acquired in dynamic MRM mode using the base CE.

Resource

The human SRMAtlas is available at www.srmatlas.org, SRMAtlas build: 'Complete Human SRMAtlas'.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Addona TA, Abbatiello SE, Schilling B, Skates SJ, Mani DR, Bunk DM, Spiegelman CH, Zimmerman LJ, Ham AJL, Keshishian H, et al. Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. Nat Biotech. 2009; 27:633–641.

Anderson NL, Anderson NG, Haines LR, Hardie DB, Olafson RW, Pearson TW. Mass Spectrometric Quantitation of Peptides and Proteins Using Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA). Journal of Proteome Research. 2004; 3:235–244. [PubMed: 15113099]

Antonarakis ES, Armstrong AJ. Evolving standards in the treatment of docetaxel-refractory castration-resistant prostate cancer. Prostate Cancer Prostatic Diseases. 2011; 14:192–205. [PubMed: 21577234]

Bantscheff M, Schirle M, Sweetman G, Rick J, Kuster B. Quantitative mass spectrometry in proteomics: a critical review. Anal Bioanal Chem. 2007; 389:1017–1031. [PubMed: 17668192]

Baty JDJ. Single and multiple ion recording techniques for the analysis of diphenylhydantoin and its major metabolite in plasma. Biomedical mass spectrometry. 1977; 4:36. [PubMed: 836941]

Bennati AM, Castelli M, Della Fazia MA, Beccari T, Caruso D, Servillo G, Roberti R. Sterol dependent regulation of human TM7SF2 gene expression: Role of the encoded 3β-hydroxysterol 14-reductase in human cholesterol biosynthesis. Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids. 2006; 1761:677–685. [PubMed: 16784888]

Boutet EL, Damien, Tognolli Michael, Schneider Michael, Bairoch Amos. UniProtKB/Swiss-Prot. The Manually Annotated Section of the UniProt KnowledgeBase Methods in Molecular Biology. 2007; 406:89–112. [PubMed: 18287689]

Carr SA, Abbatiello SE, Ackermann BL, Borchers C, Domon B, Deutsch EW, Grant RP, Hoofnagle AN, Hüttenhain R, Koomen JM, et al. Targeted Peptide Measurements in Biology and Medicine: Best Practices for Mass Spectrometry-based Assay Development Using a Fit-for-Purpose Approach. Molecular & Cellular Proteomics. 2014; 13:907–917. [PubMed: 24443746]

Craciun FL, Bijol V, Ajay AK, Rao P, Kumar RK, Hutchinson J, Hofmann O, Joshi N, Luyendyk JP, Kusebauch U, et al. RNA Sequencing Identifies Novel Translational Biomarkers of Kidney Fibrosis. J Am Soc Nephrol. 2015; 27:1702–1713. [PubMed: 26449608]

da Cunha JPC, Galante PAF, de Souza JE, de Souza RF, Carvalho PM, Ohara DT, Moura RP, Oba-Shinja SM, Marie SKN, Silva WA, et al. Bioinformatics construction of the human cell surfaceome. Proceedings of the National Academy of Sciences. 2009; 106:16752–16757.

De S, Cipriano R, Jackson MW, Stark GR. Overexpression of Kinesins Mediates Docetaxel Resistance in Breast Cancer Cells. Cancer Research. 2009; 69:8035–8042. [PubMed: 19789344]

Desiere F, Deutsch E, Nesvizhskii A, Mallick P, King N, Eng J, Aderem A, Boyle R, Brunner E, Donohoe S, et al. Integration with the human genome of peptide sequences obtained by high-throughput mass spectrometry. Genome Biology. 2004; 6:R9. [PubMed: 15642101]

Deutsch EW, Mendoza L, Shteynberg D, Slagel J, Sun Z, Moritz RL. Trans-Proteomic Pipeline, a standardized data processing pipeline for large-scale reproducible proteomics informatics. PROTEOMICS – Clinical Applications. 2015a; 9:745–754. [PubMed: 25631240]

Deutsch EW, Sun Z, Campbell D, Kusebauch U, Chu CS, Mendoza L, Shteynberg D, Omenn GS, Moritz RL. State of the Human Proteome in 2014/2015 As Viewed through PeptideAtlas: Enhancing Accuracy and Coverage through the AtlasProphet. Journal of Proteome Research. 2015b; 14:3461–3473. [PubMed: 26139527]

Ebhardt HA, Root A, Sander C, Aebersold R. Applications of targeted proteomics in systems biology and translational medicine. PROTEOMICS. 2015; 15:3193–3208. [PubMed: 26097198]

Edwards AM. Too many roads not taken. Nature. 2011; 470:163–165. [PubMed: 21307913]

Fagerberg L, Jonasson K, von Heijne G, Uhlén M, Berglund L. Prediction of the human membrane proteome. PROTEOMICS. 2010; 10:1141–1149. [PubMed: 20175080]

Farrah T, Deutsch EW, Kreisberg R, Sun Z, Campbell DS, Mendoza L, Kusebauch U, Brusniak MY, Hüttenhain R, Schiess R, et al. PASSEL: The PeptideAtlas SRMexperiment library. Proteomics. 2012; 12:1170–1175. [PubMed: 22318887]

Gillet LC, Navarro P, Tate S, Röst H, Selevsek N, Reiter L, Bonner R, Aebersold R. Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. Molecular & Cellular Proteomics. 2012; 11 O111.016717.

Gillette MA, Carr SA. Quantitative analysis of peptides and proteins in biomedicine by targeted mass spectrometry. Nat Meth. 2013; 10:28–34.

Goldstein JL, Brown MS. A Century of Cholesterol and Coronaries: From Plaques to Genes to Statins. Cell. 2015; 161:161–172. [PubMed: 25815993]

Hüttenhain R, Soste M, Selevsek N, Röst H, Sethi A, Carapito C, Farrah T, Deutsch EW, Kusebauch U, Moritz RL, et al. Reproducible Quantification of Cancer-Associated Proteins in Body Fluids Using Targeted Proteomics. Science Translational Medicine. 2012; 4:142ra194.

Iskar M, Zeller G, Blattmann P, Campillos M, Kuhn M, Kaminska KH, Runz H, Gavin AC, Pepperkok R, van Noort V, et al. Characterization of drug-induced transcriptional modules: towards drug repositioning and functional understanding. Molecular Systems Biology. 2013; 9:622.

Karlsson C, Malmström L, Aebersold R, Malmström J. Proteome-wide selected reaction monitoring assays for the human pathogen Streptococcus pyogenes. Nat Commun. 2012; 3:1301. [PubMed: 23250431]

Kennedy JJ, Abbatiello SE, Kim K, Yan P, Whiteaker JR, Lin C, Kim JS, Zhang Y, Wang X, Ivey RG, et al. Demonstrating the feasibility of large-scale development of standardized assays to quantify human proteins. Nat Meth. 2014; 11:149–155.

Kuster B, Schirle M, Mallick P, Aebersold R. Scoring proteomes with proteotypic peptide probes. Nature Reviews Molecular Cell Biology. 2005; 6:577–583. [PubMed: 15957003]

Ludwig C, Claassen M, Schmidt A, Aebersold R. Estimation of Absolute Protein Quantities of Unlabeled Samples by Selected Reaction Monitoring Mass Spectrometry. Molecular & Cellular Proteomics. 2012; 11

Mallick P, Schirle M, Chen SS, Flory MR, Lee H, Martin D, Ranish J, Raught B, Schmitt R, Werner T, et al. Computational prediction of proteotypic peptides for quantitative proteomics. Nature Biotechnology. 2007; 25:125–131.

Mann M, Kulak NA, Nagaraj N, Cox J. The Coming Age of Complete, Accurate, and Ubiquitous Proteomes. Molecular Cell. 2013; 49:583–590. [PubMed: 23438854]

Picotti P, Aebersold R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. Nat Meth. 2012; 9:555–566.

Picotti P, Bodenmiller B, Mueller LN, Domon B, Aebersold R. Full Dynamic Range Proteome Analysis of S. cerevisiae by Targeted Proteomics. Cell. 2009; 138:795–806. [PubMed: 19664813]

Picotti P, Clement-Ziza M, Lam H, Campbell DS, Schmidt A, Deutsch EW, Rost H, Sun Z, Rinner O, Reiter L, et al. A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis. Nature. 2013; 494:266–270. [PubMed: 23334424]

Schubert OT, Mouritsen J, Ludwig C, Röst HL, Rosenberger G, Arthur PK, Claassen M, Campbell DS, Sun Z, Farrah T, et al. The Mtb Proteome Library: A Resource of Assays to Quantify the Complete Proteome of Mycobacterium tuberculosis. Cell Host & Microbe. 2013; 13:602–612. [PubMed: 23684311]

Sharma V, Eckels J, Taylor GK, Shulman NJ, Stergachis AB, Joyner SA, Yan P, Whiteaker JR, Halusa GN, Schilling B, et al. Panorama: A Targeted Proteomics Knowledge Base. Journal of Proteome Research. 2014; 13:4205–4210. [PubMed: 25102069]

Surinova S, Radová L, Choi M, Srovnal J, Brenner H, Vitek O, Hajdúch M, Aebersold R. Non-invasive prognostic protein biomarker signatures associated with colorectal cancer. EMBO Molecular Medicine. 2015; 7:1153–1165. [PubMed: 26253080]

Takahashi S, Fusaki N, Ohta S, Iwahori Y, Iizuka Y, Inagawa K, Kawakami Y, Yoshida K, Toda M. Downregulation of KIF23 suppresses glioma proliferation. Journal of Neuro-Oncology. 2012; 106:519–529. [PubMed: 21904957]

Uhlen M, Bjoerling E, Agaton C, Szigyarto CAK, Amini B, Andersen E, Andersson AC, Angelidou P, Asplund A, Asplund C, et al. A Human Protein Atlas for Normal and Cancer Tissues Based on Antibody Proteomics. Molecular & Cellular Proteomics. 2005; 4:1920–1932. [PubMed: 16127175]

Whiteaker JR, Halusa GN, Hoofnagle AN, Sharma V, MacLean B, Yan P, Wrobel JA, Kennedy J, Mani DR, Zimmerman LJ, et al. CPTAC Assay Portal: a repository of targeted proteomic assays. Nat Meth. 2014; 11:703–704.
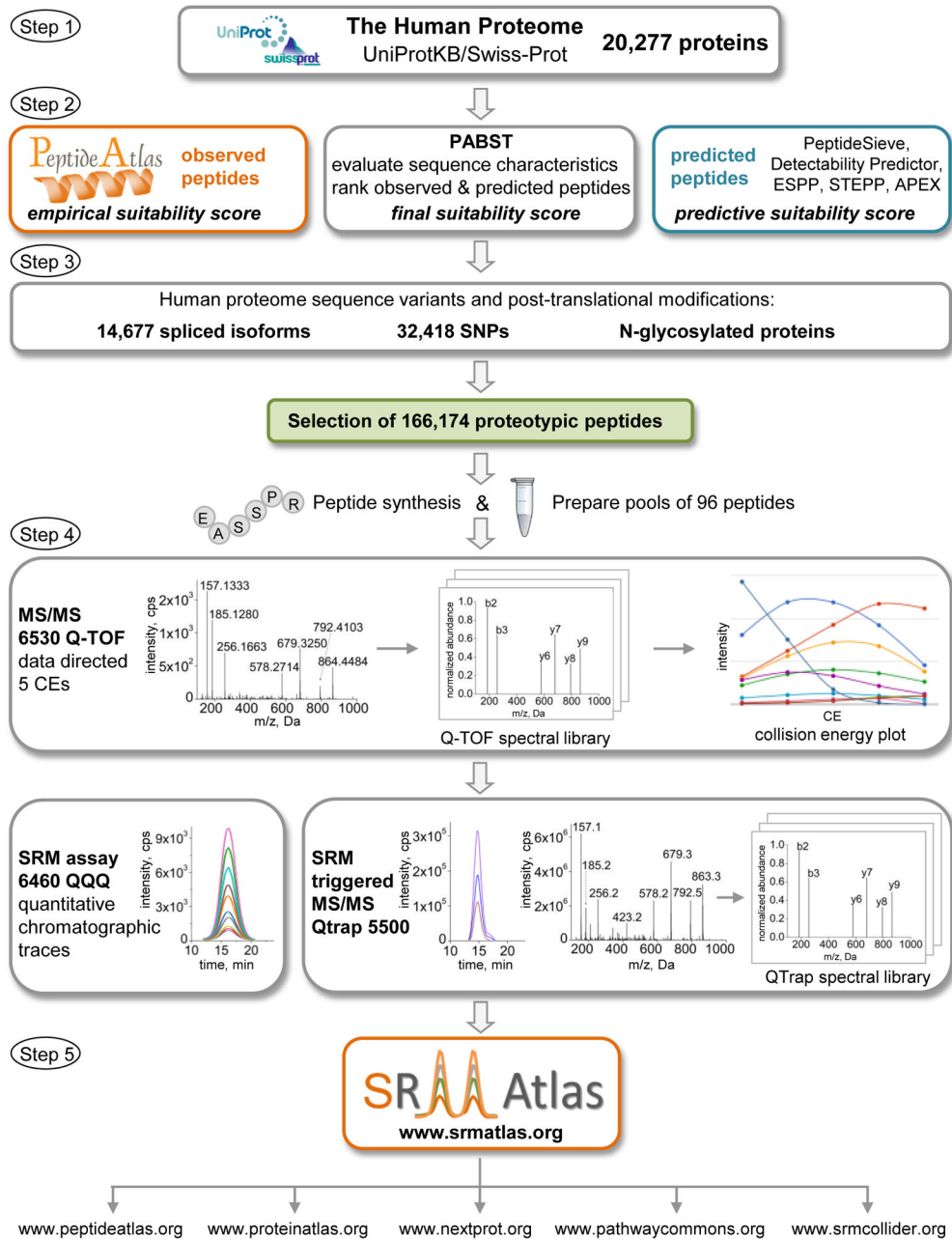
Figure 1. Human SRMAtlas development

The scheme outlines the workflow steps to generate SRM assays for every human protein. Peptides were selected for 20,277 proteins in the UniProtKB/Swiss-Prot database as well as for spliced isoforms, SNPs and modifications. Selection of PTPs was an iterative process by mining MS observed peptides in PeptideAtlas and the use of prediction tools. The PABST algorithm evaluated sequence constraints and ranked observed and predicted peptides, the highest scoring peptides for each protein were selected for SRM assay development. Peptides were individually synthesized and pooled in sets of 96. Peptides were analyzed on an Agilent 6530 Q-TOF with five CEs to acquire high-resolution MS/MS spectra to create

spectral libraries and CE plots. SRM coordinates were extracted from the spectral library to acquire chromatographic traces on an Agilent 6460 QQQ. SRM assays were also developed on a Sciex QTrap 5500, upon the detection of a transition a full MS/MS spectrum was acquired to create a QTrap spectral library. SRM assay parameters including precursor and fragment ion type, charge state and rank order, elution time as well as chromatograms, MS/MS spectra and CE plots are provided in the human SRMAtlas resource. The human SRMAtlas is integrated with external knowledge bases providing comprehensive information on a protein of interest. See also Figure S1 and Table S2.
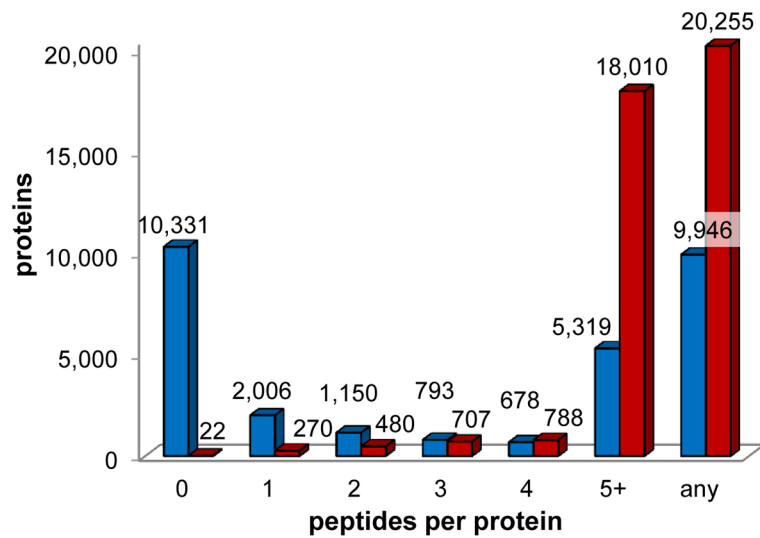
Figure 2. Human proteome coverage

The graph details the number of peptides per protein by empirically observed peptides in the human PeptideAtlas (build 2010-05, blue) and by PTPs selected for the human SRMAtlas (red). '5+' specifies five or more peptides. 'any' shows the number of proteins for which at least one peptide is available. 9946 proteins (49.1% of the predicted human proteome) were described by MS observed peptides in PeptideAtlas 2010. SRMAtlas provides with 99.9% proteome coverage for 20,255 proteins by synthetic peptides. See also Table S1 and Table S3.
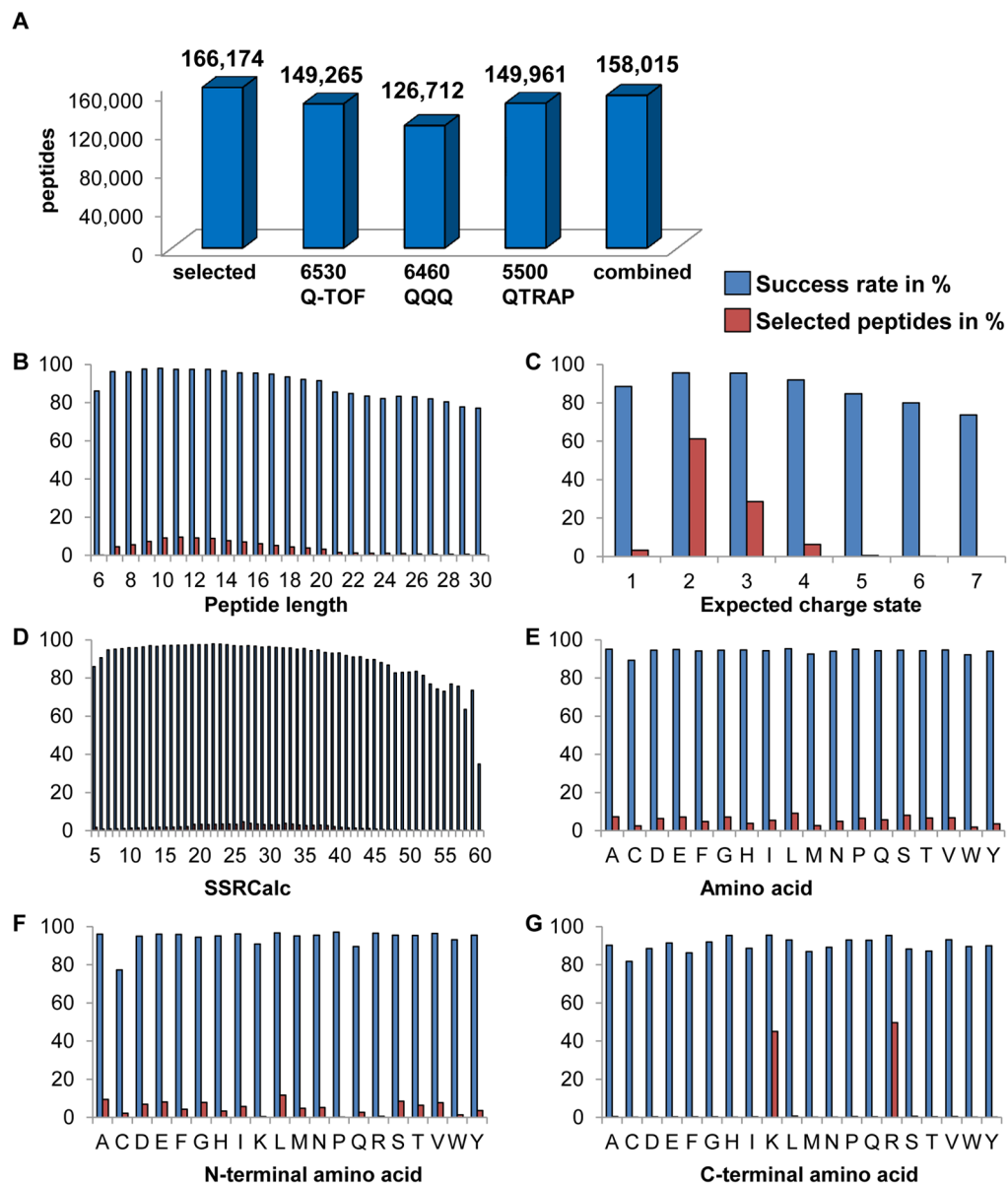
Figure 3. SRM assay success

(A) Number of developed SRM assays per instrument type in comparison to the number of synthesized peptides. The 6530 Q-TOF extracted coordinates served as input for the 6460 QQQ derived SRM assays with a success of 84.9%. 6530 Q-TOF and QTrap 5500 combined result in 158,015 targeted assays constituting 95.1% of the selected peptides. (B–G) Selected peptides (red) and their assay success rate (blue) in percent are displayed by (B) peptide length, (C) expected charge state, (D) hydrophobicity as SSRCalc value, (E) amino acid, (F) N-terminal amino acid and (G) C-terminal amino acid. See also Figure S2.
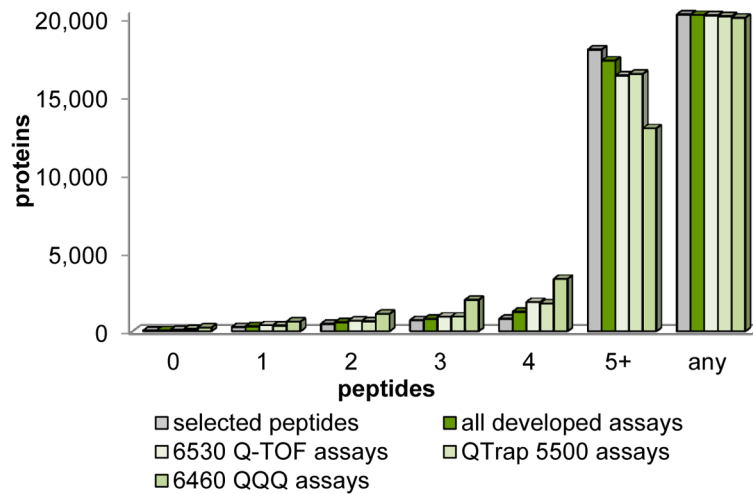
Figure 4. SRM assay coverage in the human SRMAtlas
Assay coverage by peptides per protein and instrument is displayed in green shades, selected peptides are shown in grey. 158,015 successfully developed assays represent 99.7% (20,225 proteins) of the human proteome (dark green). 95.4% of the human proteome is presented by at least three assays. 22 proteins are inaccessible. See also Figure S3 and Table S4.
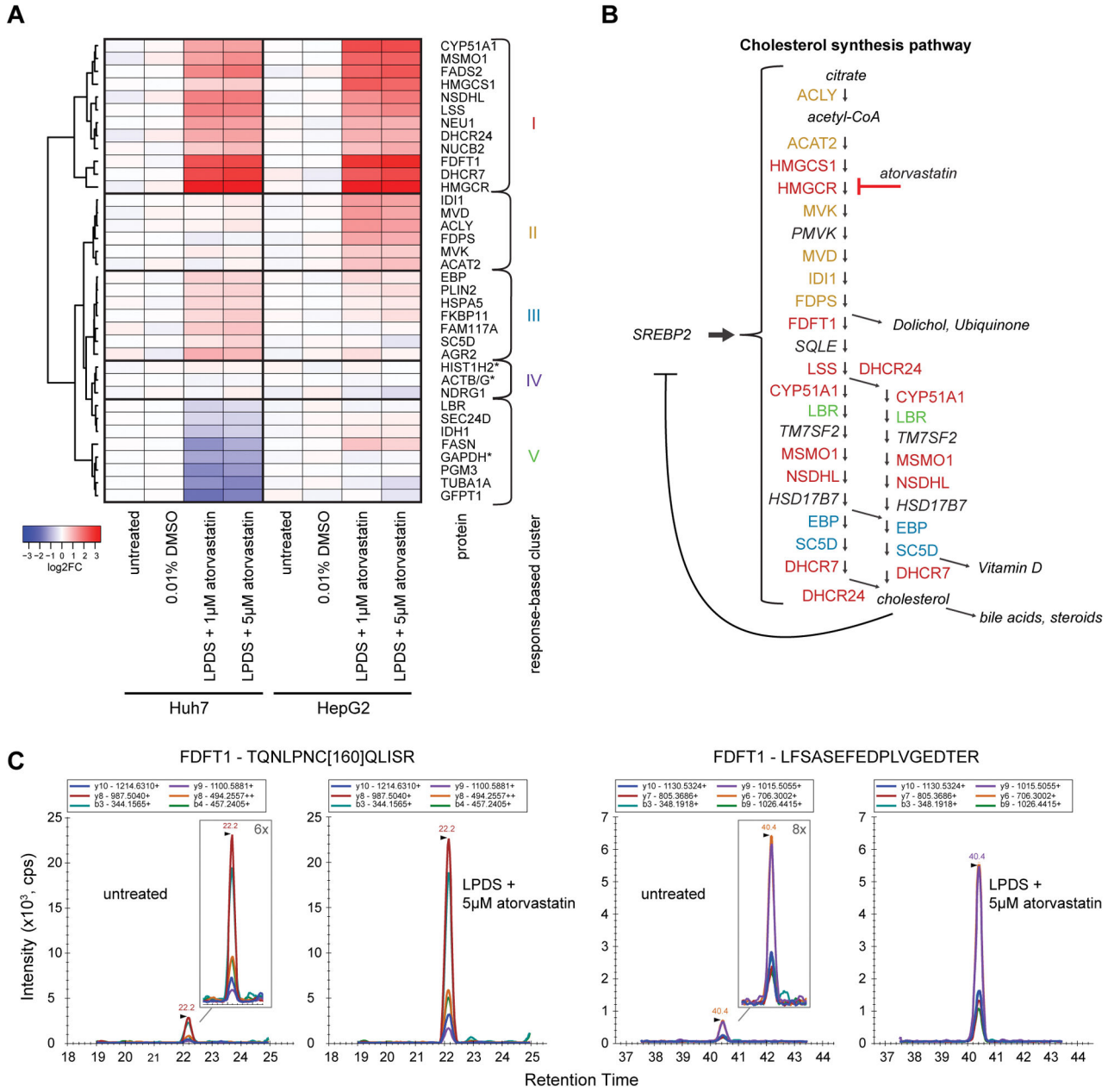
Figure 5. Drug-induced inhibition of cholesterol synthesis

Systematic proteomic quantification of proteins from a reported gene module and enzymes in the cholesterol synthesis pathway upon drug treatment. (A) The heatmap shows the change in protein abundance following the treatment with lipoprotein deficient serum (LPDS) and atorvastatin compared to control conditions (untreated and 0.01% DMSO) of the same cell line. The signal represents the mean result from three independent biological experiments and three SRM analyses per sample. Proteins were hierarchically clustered according to the elicited response with the Ward2 algorithm and euclidean distance. Based on the clustering tree and the protein regulation we defined five different clusters of proteins showing similar response (I–V). Proteins marked with asterisks were included as

"housekeeping" proteins for normalization. (B) Shown are the measured proteins from the pathway synthesizing cholesterol from acetyl-CoA. The enzymes are sorted by their position in the pathway and the proteins are color-coded according to the cluster in (A) they belong to (I–IV). The proteins or metabolites in italics have not been measured and are included for completeness. The inhibition of *HMGCR* by atorvastatin and the negative feedback of SREBP2 are depicted. (C) SRM chromatograms of peptide TQNLPNC[160]QLISR and LFSASEFEDPLVGEDTER from protein *FDTF1* showing the difference in signal abundance between untreated cells and cells treated with LDPS + 5 μM atorvastatin as representative examples. The lower signal in untreated cells is magnified in the inset. See also Table S5.
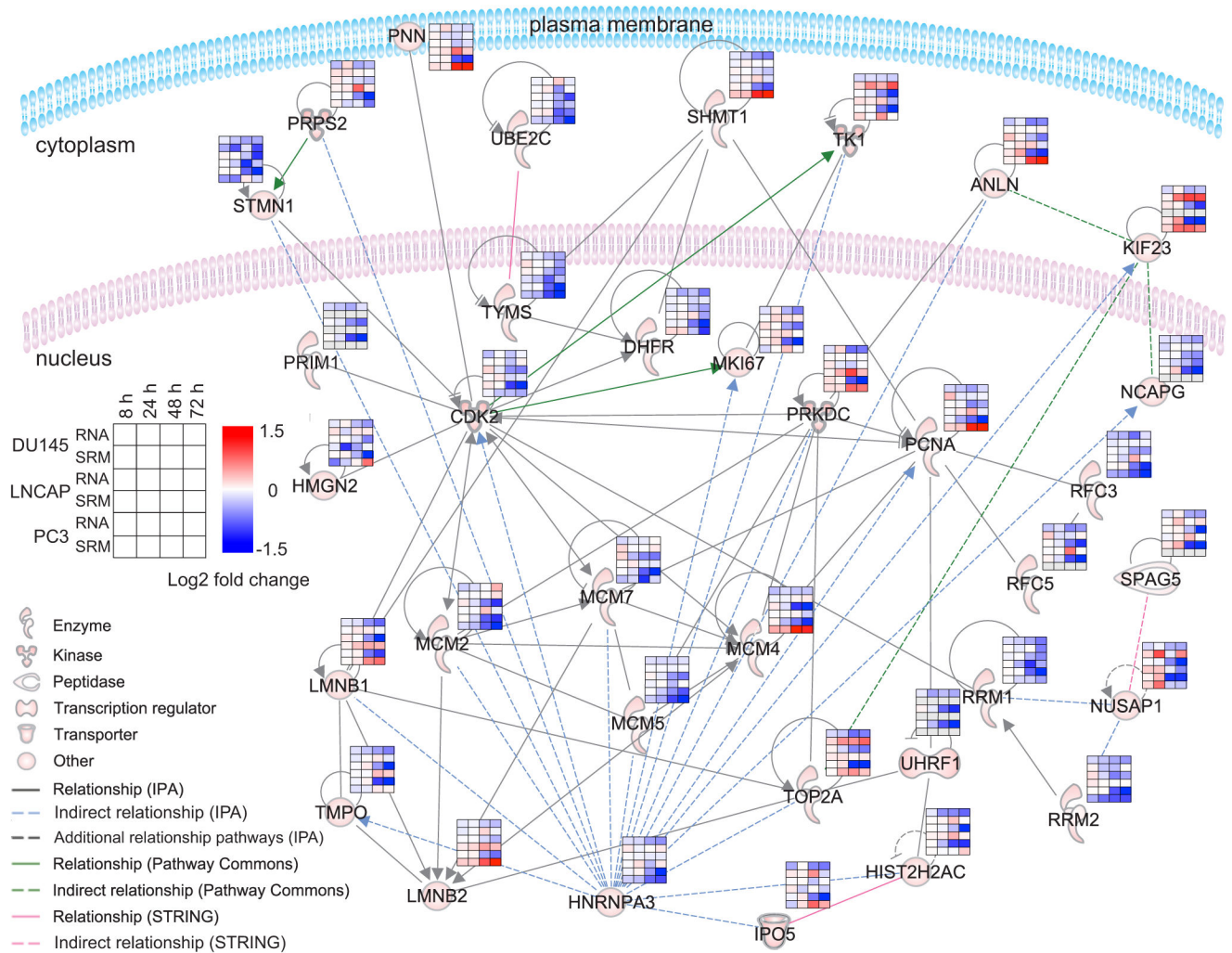
Figure 6. Network of proteins associated with docetaxel perturbation of the cell cycle
SRM-based quantification of a protein network in three prostate cancer cell lines, DU145, LNCaP and PC-3, at four time points post-treatment with docetaxel and in untreated controls in comparison to mRNA abundance changes. The microarray-derived functional network was visualized with Ingenuity Pathway Analysis (IPA). The structure of the network is based on the IPA Core Analysis, STRING and Pathway Commons derived direct interactions and indirect relationships. Each heatmap visualizes the log2 abundance change of treated versus control cells for each time in each cell line at the transcript (mRNA) and protein (SRM) level. The signal represents the mean result from two technical replicates at the transcript level and three SRM analyses per sample. See also Figure S4 and Table S5.