Toward Completion of the Human Proteome Parts List: Progress Uncovering Proteins That Are Missing or Have Unknown Function and Developing Analytical Methods

For the past 6 years, the *Journal of Proteome Research* has supported the Human Proteome Organization (HUPO) in dedicating a special issue of the journal to the Chromosome-centric Human Proteome Project (C-HPP); for the past 2 years, it has encompassed the whole HPP with contributions also from the Biology/Disease (B/D)-HPP and the HPP resource pillars. The 32 papers published this month further annotate the human protein parts list or present advances in analytical methods for the identification of proteoforms or new members of the human proteome. This Editorial presents highlights of selected papers that focus on the increasingly difficult task for the C-HPP international teams of identifying the remaining unannotated proteins in the human proteome that lack evidence of existence at the protein level (PE1) but have less direct evidence of their existence (PE2,3,4), the so-called missing proteins (MPs) in neXtProt. In total, 104 MPs have been found that meet the stringent HPP guidelines designed to mitigate against false discovery and incorrect annotation in neXtProt and other public databases (Table 1).

Table 1. List of Papers Presenting the Discovery of Credible Missing Proteins in This Issue

article in this issue	number of MPs and their cell or tissue source meeting HPP Guidelines v 2.10
Macron et al.[2,4]	13 cerebrospinal fluid
Clemente et al.[7]	1 mesenchymal stem cells
He et al.[17]	2 mirror proteases
Hwang et al.[5]	5 olfactory epithelium
Robin et al.[10]	1 HeLa cells
Ronci et al.[11]	3 mitochondria
Sun et al.[16]	14 multiple proteases
Weldemariam et al.[8]	26 embryonic stem cells
Zhang et al.[9]	30 membrane proteins
Sjöstedt et al.[6]	9 skin, retina, thymus, pituitary, hypothalamus, and lactating breast
Pullman et al.[32]	107 MassIVE database

In addition, 107 MPs identified in the MassIVE database will be processed through the Trans-Proteomic Pipeline of PeptideAtlas for submission to neXtProt. Many credible detections of MPs are only possible from analyses of rare or uncommonly analyzed cells and tissues, which often necessitate innovative sample preparation and data analysis workflows. Six papers present strategies and findings on the functionalization of some of the ~1260 uncharacterized "dark proteins" (uPE1) lurking in the human proteome yet lacking any experimentally determined or predicted function. Other papers report biomarker discovery, analyses of protein variants, glycan and other post-translational modifications (PTMs), and the mitochondrial database, proteome, and N-terminome (Figure 1). In summary, this Special Issue marks major advances in the journey toward completion of the human proteome parts list in

the context of proteome biology and disease; while tantalizingly near, this global journey must still overcome formidable challenges to complete the first accurate draft of the human proteome.

■ PROGRESS ON THE CREDIBLE DETECTION OF neXtProt PE2,3,4 MISSING PROTEINS

In the category of MPs, 14 papers present topics ranging from overall metrics for the Human Proteome and results from MS analyses of under-studied cells, organelles, or tissues to the application of novel or useful analytical methods. Omenn et al.[1] update the status of 17 470 credibly identified PE1 proteins in the neXtProt release 2018-01-17, including 462 previous MPs as of release 2017-01-23. This represents an authoritative 2018 draft of the human proteome. These PE1 proteins can be divided into two groups reflecting the method of identification: 16 092 proteins from MS analyses and 1378 proteins from other methods (protein−protein interactions, disease mutations, PTM and proteolytic processing, 3D structures, Edman degradation, antibody-based techniques, and other biochemical approaches). Meanwhile, PeptideAtlas expanded its count of canonical proteins by 625, including 269 from a single study of SUMOylation. The number of MPs has dropped from 5511 in 2012, to 2949 in 2016, to 2186 in the 2018-01 releases from neXtProt.

A major strategy to detect the expression of MPs is the analysis of under-studied or rare tissue specimens (Figure 1). Examples here are cerebrospinal fluid (CSF), olfactory epithelium, human mesenchymal stem cells (hMSCs) from adipose tissue, and human embryonic stem cells (hESCs). Macron et al.[2] of the Swiss Chromosome 2 team analyzed a commercial pool of CSF samples and identified 2281 proteins according to the stringent HPP guidelines;[3] two PE2 proteins, augurin (Q9H1Z8) and shadow of prion protein (Q5BIV9), were identified from five and two uniquely mapping peptides of at least nine amino acids (aa) in length, respectively, and then confirmed using spectra of their cognate synthetic peptides. For 14 additional MPs, there was only one such peptide; however, a second uniquely mapping "stranded peptide" was found in PeptideAtlas for 5 of these, making a total of 7 MPs with good spectra. Four of these proteins are exclusively or mostly expressed in brain. The same authors[4] performed a "deep dive" on a different pool of normal CSF specimens, detecting 3379 proteins overall and 12 MPs, including 3 transmembrane protocadherins; 8 were based on 2−6 uniquely mapping peptides, whereas 4 more had a single peptide that was combined with a stranded peptide in PeptideAtlas from previous CSF data sets. However, 6 were
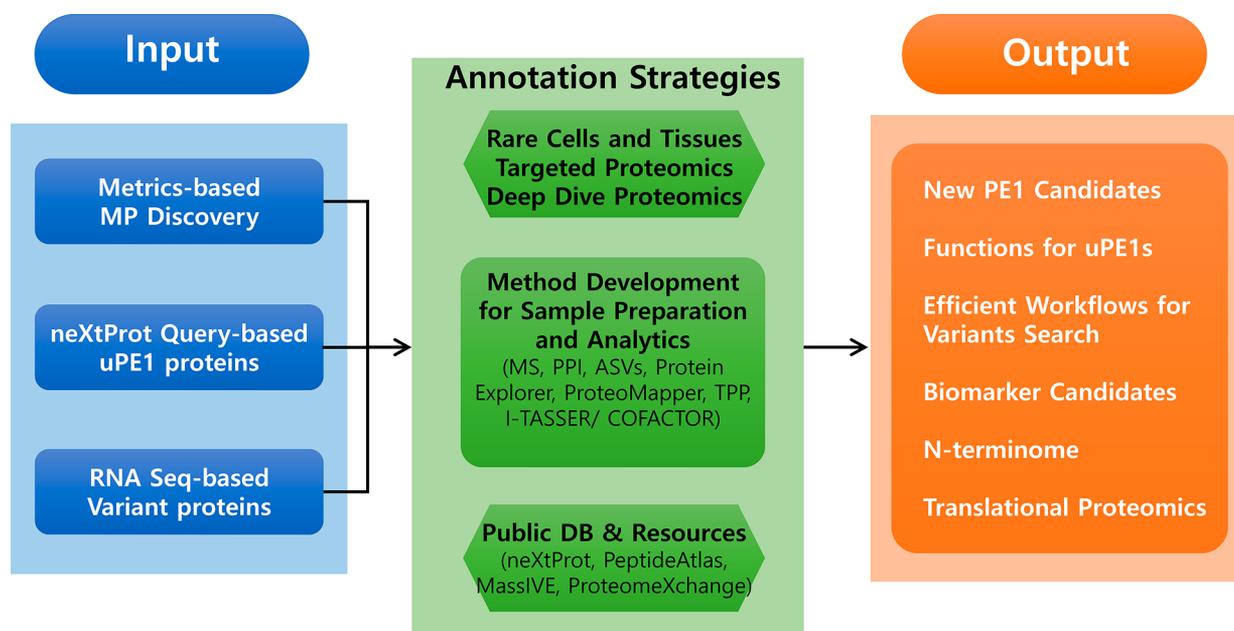
Figure 1. Overall flow of research works presented in this Special Issue. The "Input" part is composed of metrics for counting missing proteins, the updated release of neXtProt proteome information, and various cell-type RNA-Seq databases. Annotation strategies cover several components that can be developed or upgraded, leading to the production of various deliverables as outputs. Abbreviations used here: ASVs, alternative spliced variants; PE, protein evidence or existence; PPI, protein–protein interaction; TPP, Trans-Proteomic Pipeline; uPE1, uncharacterized PE1.

already found in the first Macron et al. paper,[2] so the combined total of newly found MPs from CSF is 13.

For years, we have yearned for samples of olfactory epithelium to seek the expression of some of the 404 predicted proteins from olfactory receptor genes. Hwang et al.[5] from Korea identified 3731 proteins, among which were five MPs (P0C7M7, P46721, P59826, Q658L1, Z8N434) previously identified only by mRNA (PE2) and validated with synthetic peptides in a trans-sphenoidal surgical sample of ethmoid mucosa of the nasal cavity in a patient with meningioma of the anterior skull base. The MPs met the HPP Guidelines v2.1.0,[3] requiring two uniquely mapping and non-nested peptides of at least 9 aa, confirmed with the neXtProt uniqueness checker. Sadly, none were olfactory receptor proteins. However, P59826 (BPIFB3, BPI fold-containing family B member 3) is annotated in neXtProt as recognizing and binding specific classes of odorants, acting as a carrier molecule transporting odorants across the mucus layer to access receptor sites. BPIFB3 transcripts were expressed in the olfactory epithelium at 10 times the levels in 17 control tissues. The authors also identified 50 splice isoforms and 2000 single aa sequence variants.

Sjöstedt et al.,[6] from the Human Protein Atlas and the HPP Antibody Profiling resource pillar, introduce tissue profiling with a combination of transcription and protein expression/localization with immunohistochemistry to identify MPs in skin and five understudied tissue types: retina, thymus, pituitary, hypothalamus, and lactating breast. Transcription data were available from HPA, the GTEx consortium, and the FANTOM5 consortium to identify genes selectively expressed in specialized tissues. With this workflow, which went well beyond standard tissue microarray assays, the authors verified targets that showed enhanced or enriched expression in certain tissues and their histological structures with antibody (Ab)-based proteomics techniques. Among the 300 proteins sought, 9 MPs were detected and localized: POU4F1, FRMD1,

ARHGEF33, GABRG1, KRTAP2-1, BHLHE22, SPRR4, AVPR1B, and DCLK3. This approach could be enhanced by adding MS/MS analyses of the same specimens.

Clemente et al.,[7] the Chromosome 16 team from Spain, present proteomic profiling of about 6000 proteins (false discovery rate (FDR) 1%) from ten million hMSCs derived from adipose tissue or umbilical cord using RP-QTOF or PAGE-Q Exactive analyses. They identified peptides from 13 MPs after filtering with the neXtProt uniqueness checker. Only one had more than a single unique peptide; they showed compelling evidence of the expression of hyaluronan synthase 1 (Q92839-1), with four peptides of 9+ aa and coelution of heavy-labeled synthetic peptides for three of them. The one peptide that was not confirmed, due to different retention time, was attributed to the reverse sequence of −LV− from a different protein. It is surprising that this membrane-localized enzyme with seven transmembrane regions was an MP, as hyaluronan is a well-studied and critical extracellular matrix component, especially in cartilage. The additional 12 proteins reserved for further analysis may be good prospects. For 3 of these 12 proteins (Q6TDP4-1, Q3ZCT8-1, Q96HZ4-2), they found a stranded uniquely mapping second peptide in PeptideAtlas, and they are now pursuing their confirmation with synthetic peptides and selected reaction monitoring (SRM) approaches.

Weldemariam et al.[8] of the Chromosome 4 team from Taiwan performed subcellular analyses of membrane proteins from hESCs using their reported two-step centrifugation with sucrose to obtain nuclear, cytoplasmic, and membrane fractions, followed by gel-assisted digestion with trypsin, peptide fractionation, and LC−MS/MS analysis. According to the 2018-01-17 release from neXtProt, membrane proteins represent 48% of the 2186 PE2,3,4 MPs. This paper reports 11 970 unique proteins with 1% protein FDR, including 6138 membrane proteins, with 26 PE2,3,4 MPs meeting HPP Guidelines and 21 verified by MRM−MS or matching spectra

from synthetic peptides in PeptideAtlas. The identification of these MPs and their further analysis revealed potential roles in pluripotency-related pathways and lineage- and tissue-specific differentiation. GPR19 was highlighted as the most positively significant gene association for the maintenance or differentiation of hESCs. Single unique peptides for 87 MPs were detected, so the stranded peptide approach could be applied. They also prepared to address the uPE1/neXt-CP50 challenge, noting the identification of 570 uPE1 proteins, of which 28 were coded on Chromosome 4.

Zhang et al.[9] used a hydrophobic peptide separation workflow to identify membrane-associated MPs and found 30 MPs (≥2 non-nested unique peptides with ≥9 amino acids) from the membrane preparations of K562 and HeLa cell lines and human placenta. Nineteen of the 30 were categorized as membrane proteins or extracellular matrix proteins. Robin et al.[10] reanalyzed 41 publicly available HeLa cell data sets, using a custom-made quantitative spectral counting approach to find just 1 MP candidate (FRAT2) plus 189 new phosphorylation sites, 392 N-terminal acetylation sites, and 21 SAAVs. Ronci et al.[11] reported a sequential fractionation strategy that identified and verified three low-abundance nuclear-coded MPs (METTL12, FAM163A, RGS13) in mitochondria from five cell line samples after filtering with the neXtProt uniqueness checker; no MP candidates were found with HeLa cells. The lack of MPs remaining to be identified in HeLa cells is not surprising because these cells are one of the most commonly used and well-studied over the years.

A very productive line of investigation over the past few years has been the exploitation of knowledge that transcript expression is highest for nearly 1000 protein-coding genes in the male reproductive tract. Hundreds of PE2,3,4 MPs have been detected in sperm and testis, led by the Chromosome 2/14 consortium in Switzerland and France[12,13] and the Chromosome 1 team in China.[14,15] Sun et al.[16] identified 11 558 proteins in testis using SDS-PAGE fractionation, digestion with a combination of the proteases trypsin, LysC, and GluC, and high-pH reverse-phase fractionation; this total is 3700 more proteins than the previous year.[15] They identified 114 MP candidates and verified 14 with 2 unique peptides of at least 9 aa by spectrum quality checking, isobaric PTM and SAAV filtering, and synthesized peptide matching. Of the 14 MPs, 3 were testis-specific proteins and 10 were closely related to liver cancer, potentially useful for biomarker development in hepatocellular carcinoma. He et al.[17] used the "mirror protease" approach, whereby the new protease LysargiNase (Huesgen et al.),[18] which cleaves N-terminal to lysine and arginine and so mirrors trypsin (cutting C-terminal to lysine and arginine), was used with Tricine-SDS-PAGE enrichment for low-molecular-weight (LMW) proteins and identified 4063 proteins and 2565 LMW (<40 kDa) proteins, of which 1130 had paired peptides generated from the two proteases. Among seven MP candidates, they verified two with a series of continuous and complementary b/y-product ions in the paired spectra: beta-defensin 123 (Q8N688) and C17orf64 (Q86WR6). C17orf64 was independently prioritized by Siddiqui et al.[19] for detection on Chr 17 (below). The other five have only one proteotypic peptide and require further verification. Attention to LMW proteins may be wise because there is a much higher proportion of MPs to PE1 proteins when the MW is <40 kDa.

An example for all of the HPP chromosome teams is the analysis by Siddiqui et al.[19] of the basis for progress toward the

October 2016 neXt-MP50 challenge for Chromosome 17. The number of MPs from Chr 17 was reduced from 148 in neXtProt 2016-01 to 105 in neXtProt 2018-01. The evidence of PE1 was by mass spectrometry for 25, protein–protein interaction for 12, a combination of the two approaches for 3, and other methods for the remaining 3, across the global proteomics community. Notable among these new PE1 proteins were five keratin-associated proteins, a single olfactory receptor, and five additional membrane-embedded proteins. The Chr 17 team then evaluated the remaining 105 MPs for the best prospects, of which 88 are PE2 with transcript data. All but 16 have 2 or more potential proteotypic sequences of at least 9 aa. They identified the top 10 and next 15 for MS and the top 10 for protein–protein interaction assays. One MS candidate was confirmed by finding a matching stranded peptide in PRIDE. Two ranked among the top 10 MS candidates have been reported in this Special Issue, RNF112 by Macron et al.[2] in CSF and C17orf64 by He et al.[17] in testis, adding further comfort to their confident identification. These 3 found MPs will move Chr 17 to at least 46 MPs found when the neXtProt 2019-01 version is released. The discussion notes the importance of evaluating the use of recombinant baits in yeast two-hybrid assays that drive the verification of some of the proteins classified as PE1 based on protein–protein interactions.

PROTEOME ANNOTATION AND STUDIES OF uPE1 PROTEINS LACKING FUNCTIONAL ANNOTATION IN neXtProt

As a new avenue toward the completion and annotation of the human protein parts list, in March 2018, the HUPO C-HPP consortium launched neXt-CP50, in which CP stands for "characterization of protein" and uPE1 refers to the uncharacterized PE1 proteins in neXtProt. neXtProt utilizes both experimental and informatics evidence of functional annotation. In March 2018, neXtProt provided a list of 1260 proteins identified with high confidence at the protein level (PE1) but considered as functionally uncharacterized, that is, without any functional annotation or solely annotated with generic GO terms that are not linked to any specific function, such as protein binding, calcium binding, zinc binding, metal ion binding, identical protein binding, protein homo-oligomerization, or signal transduction. The neXt-CP50 project aims to characterize an initial 50 of these 1260 uPE1 proteins in 3 years. Paik et al.[20] discuss the current status of the human proteome in terms of the functional characterization of individual proteins: 1937 proteins have an unknown or nonspecific function. These are divided into two groups, 1260 of which are PE1 proteins and 677 missing proteins (PE2,3,4) that also have no function (uMPs). Together, these are termed "dark proteins", a phrase first used in the field of structural genomics (see also HPP definitions: https://hupo.org/resources/Documents/HPP%20Scientific%20Terms%20Definitions%20and%20Abbreviations_20180830.pdf). Next, Paik et al. present a detailed experimental strategy and various workflows applied to the functionalization of the targeted uPE1 proteins. Starting points depend on individual research teams' specific expertise and resources and may include protein–protein interactions with inferences from the known functions of interacting partners of the uPE1 protein or overexpression or knockdown/CRISPR Cas9 editing of the target uPE1s. At present, 15 C-HPP teams, including the newly joined Chromosome 14 group (led by Charles Pineau, France),

have initiated plans to characterize three to five uPE1 from their team's chromosome. Completion of this will demonstrate the feasibility and common goal of eventually annotating all ∼2000 dark proteins. As always, this will include contributions from the entire proteomics community, including the B/D-HPP and the Pathology Pillar of the HPP. The authors discuss how the larger characterization initiative will enhance the understanding of the human proteome and integrated cellular networks for the discovery of new mechanisms of pathology, mechanistically informative biomarkers, and rational drug targets (Figure 1).

As the first report on neXt-CP50, Zhang et al.[21] from Chromosome 17 present an elegant bioinformatics approach in which a hybrid pipeline using I-TASSER and COFACTOR is used to predict structures and likely functions of dark proteins. It is interesting that there are experimental 3D structures for 6188 of the 17 470 PE1 proteins but for only 32 of the 1260 uPE1 proteins. Benchmark testing on a control set of 100 well-characterized PE1 proteins randomly selected from Chromosome 17 showed high Gene Ontology (GO) term prediction accuracies of 0.69, 0.56, and 0.67 for molecular function (MF), biological process (BP), and cellular component (CC), respectively. They applied this tool to all 66 uPE1 proteins encoded by human Chr 17 as of 2017-08 and utilized GO to categorize their molecular function for 13, biological process for 33, and cellular components for 49 based on thresholds from the benchmarking results. Similar analyses can now be applied across all of the chromosomes to help guide experimental studies. This hybrid tool is a novel computational approach to throw light on yet-to-be-uncovered dark proteins, with broad applicability to the ∼2000 dark proteins as well as undefined proteoforms.

Representative of laboratory experimental approaches to function prediction for uPE1 proteins, Melaine et al.[22] from the French Chromosome 14 team performed LC−MS/MS analyses of testis protein extracts in combination with public mRNA and protein expression data. Of 30 identified uPE1 proteins and 15 highly enriched in testis, 2 Chromosome X proteins (A0AOU1RQG5 and PNMA6E) were subjected to further analysis. A0AOU1RQG5 contains CT47 domains and is a putative cancer/testis antigen (CTA) localized and expressed in the cytoplasm of elongated spermatids, an immune-privileged site in normal tissue. CTAs appear to be involved in gametogenesis, sperm metabolism, or motility. PNMA6E is enriched in the germ cell nuclei during most stages of spermatogenesis but not in elongated spermatids and not in Leydig cells. The splice isoform PNMA6E-201 was identified by 11 non-nested peptides of at least 9 aa, of which 9 are unique to the PNA6E gene and 4 are specific to this isoform. It is also detected at low levels in the fallopian tube and ovary. The other 14 members of the PNMA family are predominantly expressed in the brain. This study demonstrates that, as in MP studies, testis is a rich source for screening and characterizing uPE1 dark proteins as well as bringing enrichment fractionation to detect low-abundance missing proteins, some of which are currently uMPs. In all, 238 proteins identified in the human testis are uPE1, 19% of the ∼1200 uPE1 proteins in neXtProt. This paper provides a rich discussion of the approaches to molecular studies of the testis that may be applicable to many other uPE1 projects.

In another in silico approach to uPE1 proteins, Duek et al.[23] (Chromosome 2/14 Switzerland/France) utilized the SPARQL advanced search functionality of neXtProt to identify

subcellular locations, protein−protein interactions, tissue expression, disease associations, and 3D structure. A PubMed search identified characterization papers for 113 of the 2323 uncharacterized proteins; these were considered to already be engaged in the curation process, so they were excluded. Deep data mining in various publicly available resources facilitated functional hypotheses for 26 uPE1 proteins from the fields of cilia biology, male reproduction, metabolism, nervous system, immunity, inflammation, RNA metabolism, and chromatin biology; 13 of the 26 were involved in cilia biology. This paper shows the power and many applications of neXtProt data and query tools as well as the desirability of large collaborative projects involving specialists in different fields of biology.

Functional annotation of the N- and C-termini of all human proteins in the human proteome is an essential task of the HPP. Using protein-level dimethylation to label the N-termini, Marshall et al.[24] (Chromosome 6, new U.S./Canadian Team) were able to purify the N-terminome of both the mitochondria and corresponding cytosol. Normal cells and those undergoing early stages of apoptosis prior to caspase-3 activation were also analyzed by stable isotope labeling by amino acids in cell culture (SILAC) and terminal amine isotopic labeling of substrates (TAILS),[25] which enabled the largest mitochondrial N-terminome analysis to date. Together, shotgun and TAILS identified mature and protease-cleaved neo-amino (N)-termini of 26% of all mitochondrial proteins and revealed 97 new sites of proteolysis and 135 novel and 101 known mitochondrial target sequence cleavage sites that displayed a characteristic cleavage motif for mitochondrial processing peptidase. Furthermore, by contributing long ArgC peptides of these proteins to proteome exchange, TAILS analyses extend MS coverage of these mitochondrial proteins. These results implicated specific mitochondrial pathways including protein import, fission, and iron homeostasis in the initiation of apoptosis.

Monti et al.[26] from the Mitochondria B/D-HPP team (mt-HPP) update the mitochondrial proteome interactome database (DB) using neXtProt because it incorporates more kinds of data than MitoCarta 2.0 or the Integrated Mitochondrial Protein Index. The aim is to obtain an informative functional network with mitochondrial proteins as nodes and high-quality binary interactions as edges. Mitochondria are complex organelles; different functions are associated with distinct subcompartments (outer membrane, intermembrane space, inner membrane, matrix). They demonstrate that this DB may be useful to predict the function or submitochondrial localization of proteins from the network and to identify disease proteins of mitochondrial origin. Specifically, they mapped proteins lacking functional annotation to enhance speculation on their role through knowledge of their interactors and investigated mitochondrial proteins in Parkinson's disease, in which mitochondrial dysfunction is thought to be important.

■ PROGRESS ON THE DEVELOPMENT OF BIOINFORMATICS TOOLS AND MS DATA ANALYSIS

The development of bioinformatics tools and MS analysis is essential to promote proteomic studies (Figure 1). Jeong et al.[27] have designed an ASV-ID workflow for predicting candidate protein alternative splice isoforms that are usually not easily distinguishable between proteins with highly homologous peptide sequences. They demonstrate that this

tool enables the identification of ASVs using a cell-type-specific protein sequence DB that is supported by RNA-Seq data. They have identified 1935 distinct proteins under stringent conditions, which resulted in 19 differently expressed isoforms present across several cell lines (A549, HEK293, HeLa, HepFG2, MCF7); IDH2L and IDH2S offer an interesting example of different isoform expression and different N-terminal mitochondrial transit function across the cell lines. These small numbers may be attributed to the fact that this tool counts only isoforms exhibiting cell-type-specific expression supported by RNA-sequencing data sets, as the transcripts can distinguish which of identical sequences may be expressed. This new workflow may have the potential to map difficult protein isoforms in a robust way without turning to the use of multiple proteases to reveal splice junctions.

In 2017, the HPP Bioinformatics Resource Pillar team led by Eric Deutsch addressed a longstanding need for a tool for automated detailed comparison of spectra, both observed versus predicted and observed natural versus observed synthetic peptides.[28] This year, the Proteomics Standards Initiative reports from the October 2017 Dagstuhl Seminar on Computational Proteomics and its April 2018 Heidelberg Workshop a framework for organizing the desired metadata for spectra using a controlled vocabulary in a spectral library.[29] This review addresses four levels of granularity, the library (collection), the individual peptide ion, the peak fragment ion, and the peak annotation level, as well as strategies for comparing spectra, generating representative spectra for a library, selecting optimal signature ions for targeted workflows, and merging two or more libraries. The rapid development of data-independent acquisition (DIA) workflows has spurred new interest in spectral libraries to analyze extracted ion chromatograms for peptide ions, whereas targeted workflows (SRM/PRM) increasingly rely on large-scale spectral libraries to determine which proteotypic peptides and fragment ions to monitor. The libraries have lagged far behind the massive accumulation of data, so these advances are timely. A "gold standard test data set" is proposed: one spectral library with 10 000 entries from synthetic peptides and one mzML file with 10 000 spectra, in which 5000 peptide ions (but not exact spectra) are in common.

Sajulga et al.[30] describe a versatile bioinformatics tool that enables a functional connection between C-HPP and B/D-HPP proteogenomics data sets. This tool is composed of Galaxy-P and CRAVAT and seems to be quite useful for assessing data sets obtained from the HPP. The tool uses "proBed" from the HUPO Proteomics Standards Initiative, making this open software useful for improved interpretation of the function and pathological effects of protein sequence variants detected via proteogenomics. Mendoza et al.[31] introduce a very interesting open-source software tool, ProteoMapper, which enables ultrafast mapping of peptides to a reference proteome by creating an index for an input proteome from a FASTA or PEFF file. ProMaST, one of the key components of this tool, performs fast mapping of input peptides against the index. Using a fuzzy search function, ProteoMapper can identify not only sequence variation but also highly similar peptides (e.g., isobaric or near-isobaric substitution). It is useful for checking peptide uniqueness, distinguishing between catalogued and uncatalogued sequence variations in a protein and across the proteome. Although 90% of peptides in the Human PeptideAtlas map uniquely to one gene, complications arise when one gene yields multiple splice isoforms (several of which may be present simultaneously in a sample), when peptides map to multiple proteins (in the same protein family or in repeated motifs in unrelated proteins), and due to sequence variation (mutations/polymorphisms). ProteoMapper will enhance compliance with the HPP MS Data Interpretation Guidelines v2.1.0, which are required for reports of finding MPs, and facilitate the incorporation of additional very large data sets into the reanalysis pipeline of PeptideAtlas.

A particularly notable contribution to the Special Issue is by Pullman et al.,[32] who report the development of "ProteinExplorer", a repository-scale resource for exploration at MassIVE-KB at UC San Diego. This tool facilitates not only the systematic reanalysis of public data sets but also the confirmation of detected novel proteins. Furthermore, it helps end-users build customized peptide spectral libraries to inspect novel proteins by crosschecking spectra of synthetic peptides. A critical aim is to overcome the serious problem that true matches coincide across most data sets but false matches are likely to be unique to each data set; therefore, the naïve union of discovered MPs would result in an uncontrolled increase in FDRs. This resource tool will enhance MassIVE-KB, which is connected to ProteomeXchange, and thereby will link in a new collaboration to PeptideAtlas and neXtProt. This paper forwards 107 PE2,3,4 MPs for review at PeptideAtlas for the incorporation into neXtProt. MassIVE-KB currently holds 31 TB of high or higher energy C-trap dissociation (HCD) data for >1 million unique peptide sequences mapped to >16 000 proteins, expanded further by BioPlex and ProteomeTools data.

Finally, in the area of methods development for MS analysis, Ilgisonis et al.[33] (Chromosome 18, Russia) report a study aimed at demonstrating the enhanced sensitivity of targeted and shotgun LC−MS/MS analysis due to alkaline reverse-phase (RP) prefractionation of the sample, using results from a UPS2 standard spiked into *E. coli* or *P. pastoris* protein extracts.

■ PROGRESS ON PROTEOMIC STUDIES OF BIOLOGY AND DISEASE

Glycoproteomics is a very significant subfield of proteomics, combining complex glycan structures with post-translational N- and O-glycosylated modifications of the proteins. The Japanese colleagues have been pioneers in glycoproteomics and leaders of the HPP efforts in this domain. Narimatsu et al.[34] extensively review current technologies, including improved sensitivity and speed of mass spectrometry, new software for complex spectrum assignment, a unique approach to identify intact glycopeptides using MS1-based accurate masses, and spatial distribution within cells using MS imaging and lectin microarray. There are many applications to biology and disease. An important observation is that glycan structures are cell-specific and hence are good candidates for biomarkers.

Boersema et al.[35] from the B/D-HPP Protein Aggregation Diseases initiative discuss recent advances in our knowledge of protein aggregation diseases (PADs). PADs are characterized by the intracellular and extracellular accumulation of aggregated proteins, as commonly seen in Parkinson's, Alzheimer's, and Huntington's diseases, ALS, amyloidosis, and frontotemporal dementia. Each disease has distinctive protein aggregates. Their team has been active in developing MS-based proteomics approaches, which would be useful for clinical diagnosis through the quantification of aggregation-prone proteins to support both basic and clinical research.

Proteins are known to undergo aberrant conformational transitions, leading to misfolding and aggregation. The group has developed protein extraction protocols and MS methods to detect and quantify proteins involved in systemic and localized amyloidosis and probing of the conformation transitions in cells and tissue extracts. Assay coordinates are shared through public repositories and databases. There is a good network between clinicians and proteomics investigators for improving methods for early diagnosis and subtyping these devastating diseases.

Wang et al.[36] provide a deep analysis of the cysteine oxidative proteome in cardiac hypertrophy, a key feature of hypertensive heart disease, congestive heart failure, and sudden cardiac death. Reactive oxygen and nitrogen species (RS/NS) mediate hypertrophy through multiple types of oxidation of protein cysteine residues, triggering molecular switches. The major cysteine O-PTMs are reversible S-nitrosylation, S-glutationylation, sulfenic acid, disulfide bonds, and S-sulfhydration, plus irreversible sulfinic acid ($CysSO_2H$) or sulfonic acid ($CysSO_3H$). The Van Eyk and Ping laboratories have developed assays specific for these chemistries. Here they combine a customized biotin switch-based quantitative proteomics pipeline with an analytic workflow using cubic spline-based temporal clustering to characterize the dynamic hypertrophy landscape in an isoproterenol-induced mouse model. They identified 2505 protein containing 6818 cysteine sites, with 1655 proteins with 3324 sites modified by at least one of three cysteine O-PTM types. They also discuss mechanisms of cardio protection.

Naryzhny et al.[37] report an updated version of their virtual 2DE proteome map based on the analysis of several cell lines (HepG2, glioblastoma, LEH, HEK), normal liver, and plasma, which was coupled to shotgun mass spectrometry using LC–ESI–MS/MS. They claim confident identification of information (minimum of two significant sequences) about the proteoforms of 117 isoforms coded by 104 genes of Chromosome 18, awaiting additional study for validation.

During the past 2 to 3 years, two groups have exploited bibliometric analyses to identify "popular proteins" for the B/D-HPP teams and for the global community.[38,39] These are proteins that have been studied extensively and reported frequently, for which multiplex-targeted proteomics could provide a potent approach to quantitation and biological insights. Advances in DIA, such as MS-SWATH, could accelerate progress. Many B/D-HPP teams are utilizing the organ-specific popular proteins lists. Yu et al.[40] now extend this approach to targeted metabolomics and biochemical studies. Text mining for metabolites and chemicals is hindered by numerous synonyms and nonstandard names. Here they use cloud-based literature mining to connect copublication trends of metabolites/chemicals and B/D-HPP topics using Protein Universal Reference Publication-Originated Search Engine (PURPOSE) scores. Note the use of PURPOSE in the report by Boeresma et al. from the B/D-HPP PAD team.[35] Lau et al.[41] identify high-priority proteins across the human diseaseome using semantic similarity between a protein and a query term in the biomedical literature. They adjust for impact and recency of the articles across cell types, subanatomical regions, and standardized vocabularies with 20 000 disease terms. Reverse protein-to-disease analyses might be useful for the neXt-CP50 Challenge to predict and experimentally demonstrate the functions of uPE1 proteins.

CONCLUSIONS

The HUPO HPP stimulates and coordinates analyses of high-quality data for the whole field of proteomics. The resource pillars of the HPP, including PeptideAtlas, neXtProt, ProteomeXchange/PRIDE, and Human Protein Atlas provide a unifying foundation for the field and for proteogenomic and multiomics analyses in all disease categories.

This sixth annual special issue of the *Journal of Proteome Research* documents the remarkable progress in detecting and validating the expression of predicted human proteins, with 17 470 of the neXtProt PE1,2,3,4 predicted proteins now PE1 (89%), signifying compelling evidence at the protein level. Progress on the PE2,3,4 MP Challenge is documented here, and the new challenge to characterize the functions of 1260 presently unannotated PE1 proteins is introduced in substantial depth. Additional papers address bioinformatics tools and then a wide range of applications to studies of biological processes and particular diseases.

Young-Ki Paik* ⓘ

Yonsei Proteome Research Center, College of Life Science and Technology, Yonsei University

Christopher M. Overall ⓘ

Centre for Blood Research, Departments of Oral Biological & Medical Sciences and Biochemistry & Molecular Biology, Faculty of Dentistry, University of British Columbia

Fernando Corrales

Functional Proteomics Laboratory National Center of Biotechnology, CSIC

Eric W. Deutsch ⓘ

Institute for Systems Biology

Lydie Lane ⓘ

CALIPHO Group, SIB Swiss Institute of Bioinformatics and Department of Microbiology and Molecular Medicine, Faculty of Medicine, CMU, University of Geneva

Gilbert S. Omenn* ⓘ

Institute for Systems Biology, Departments of Computational Medicine & Bioinformatics, Internal Medicine, and Human Genetics & School of Public Health, University of Michigan

AUTHOR INFORMATION

Corresponding Authors

*Y.-K.P.: E-mail: paikyk@gmail.com.
*G.S.O.: E-mail: gomenn@med.umich.edu.

ORCID ⓘ

Young-Ki Paik: 0000-0002-8146-1751
Christopher M. Overall: 0000-0001-5844-2731
Eric W. Deutsch: 0000-0001-8732-0928
Lydie Lane: 0000-0002-9818-3030
Gilbert S. Omenn: 0000-0002-8976-6074

Notes

ACKNOWLEDGMENTS

■ REFERENCES

(1) Omenn, G. S.; Lane, L.; Overall, C. M.; Corrales, F. J.; Schwenk, J. M.; Paik, Y. K.; Van Eyk, J. E.; Liu, S.; Snyder, M.; Baker, M. S.; Deutsch, E. W. Progress on Identifying and Characterizing the Human Proteome: 2018 Metrics from the HUPO Human Proteome Project. *J. Proteome Res.* **2018**, DOI: 10.1021/acs.jproteome.8b00441.

(2) Macron, C.; Lane, L.; Nunez Galindo, A.; Dayon, L. Identification of Missing Proteins in Normal Human Cerebrospinal Fluid. *J. Proteome Res.* **2018**, DOI: 10.1021/acs.jproteome.8b00194.

(3) Deutsch, E. W.; Overall, C. M.; Van Eyk, J. E.; Baker, M. S.; Paik, Y. K.; Weintraub, S. T.; Lane, L.; Martens, L.; Vandenbrouck, Y.; Kusebauch, U.; Hancock, W. S.; Hermjakob, H.; Aebersold, R.; Moritz, R. L.; Omenn, G. S. Human Proteome Project Mass Spectrometry Data Interpretation Guidelines 2.1. *J. Proteome Res.* **2016**, *15* (11), 3961–3970.

(4) Macron, C.; Lane, L.; Nunez Galindo, A.; Dayon, L. Deep Dive on the Proteome of Human Cerebrospinal Fluid: A Valuable Data Resource for Biomarker Discovery and Missing Protein Identification. *J. Proteome Res.* **2018**, DOI: 10.1021/acs.jproteome.8b00300.

(5) Hwang, H.; Jeong, J. E.; Lee, H. K.; Yun, K. N.; An, H. J.; Lee, B.; Paik, Y. K.; Jeong, T. S.; Yee, G. T.; Kim, J. Y.; Yoo, J. S. Identification of Missing Proteins in Human Olfactory Epithelial Tissue by Liquid Chromatography–Tandem Mass Spectrometry. *J. Proteome Res.* **2018**, DOI: 10.1021/acs.jproteome.8b00408.

(6) Sjöstedt, E.; Sivertsson, A.; Hikmet Noraddin, F.; Katona, B.; Nasstrom, A.; Vuu, J.; Kesti, D.; Oksvold, P.; Edqvist, P. H.; Olsson, I.; Uhlen, M.; Lindskog, C. Integration of Transcriptomics and Antibody-Based Proteomics for Exploration of Proteins Expressed in Specialized Tissues. *J. Proteome Res.* **2018**, DOI: 10.1021/acs.jproteome.8b00406.

(7) Clemente, L. F.; Hernáez, M. L.; Ramos-Fernandez, A.; Ligero, G.; Gil, C.; Corrales, F. J.; Marcilla, M. Identification of the Missing Protein Hyaluronan Synthase 1 in Human Mesenchymal Stem Cells Derived from Adipose Tissue or Umbilical Cord. *J. Proteome Res.* **2018**, DOI: 10.1021/acs.jproteome.8b00384.

(8) Weldemariam, M. M.; Han, C. L.; Shekari, F.; Kitata, R. B.; Chuang, C. Y.; Hsu, W. T.; Kuo, H. C.; Choong, W. K.; Sung, T. Y.; He, F. C.; Chung, M. C. M.; Salekdeh, G. H.; Chen, Y. J. Subcellular Proteome Landscape of Human Embryonic Stem Cells Revealed Missing Membrane Proteins. *J. Proteome Res.* **2018**, DOI: 10.1021/acs.jproteome.8b00407.

(9) Zhang, Y.; Lin, Z.; Hao, P.; Hou, K.; Sui, Y.; Zhang, K.; He, Y.; Li, H.; Yang, H.; Liu, S.; Ren, Y. Improvement of Peptide Separation for Exploring the Missing Proteins Localized on Membranes. *J. Proteome Res.* **2018**, DOI: 10.1021/acs.jproteome.8b00409.

(10) Robin, T.; Bairoch, A.; Muller, M.; Lisacek, F.; Lane, L. Large-Scale Reanalysis of Publicly Available HeLa Cell Proteomics Data in the Context of the Human Proteome Project. *J. Proteome Res.* **2018**, DOI: 10.1021/acs.jproteome.8b00392.

(11) Ronci, M.; Pieroni, L.; Greco, V.; Scotti, L.; Marini, F.; Carregari, V. C.; Cunsolo, V.; Foti, S.; Aceto, A.; Urbani, A. Sequential Fractionation Strategy Identifies Three Missing Proteins in the Mitochondrial Proteome of Commonly Used Cell Lines. *J. Proteome Res.* **2018**, DOI: 10.1021/acs.jproteome.8b00422.

(12) Vandenbrouck, Y.; Lane, L.; Carapito, C.; Duek, P.; Rondel, K.; Bruley, C.; Macron, C.; Gonzalez de Peredo, A.; Coute, Y.; Chaoui, K.; Com, E.; Gateau, A.; Hesse, A. M.; Marcellin, M.; Mear, L.;

Mouton-Barbosa, E.; Robin, T.; Burlet-Schiltz, O.; Cianferani, S.; Ferro, M.; Freour, T.; Lindskog, C.; Garin, J.; Pineau, C. Looking for Missing Proteins in the Proteome of Human Spermatozoa: An Update. *J. Proteome Res.* **2016**, *15* (11), 3998–4019.

(13) Carapito, C.; Duek, P.; Macron, C.; Seffals, M.; Rondel, K.; Delalande, F.; Lindskog, C.; Freour, T.; Vandenbrouck, Y.; Lane, L.; Pineau, C. Validating Missing Proteins in Human Sperm Cells by Targeted Mass-Spectrometry- and Antibody-based Methods. *J. Proteome Res.* **2017**, *16* (12), 4340–4351.

(14) Zhang, Y.; Li, Q.; Wu, F.; Zhou, R.; Qi, Y.; Su, N.; Chen, L.; Xu, S.; Jiang, T.; Zhang, C.; Cheng, G.; Chen, X.; Kong, D.; Wang, Y.; Zhang, T.; Zi, J.; Wei, W.; Gao, Y.; Zhen, B.; Xiong, Z.; Wu, S.; Yang, P.; Wang, Q.; Wen, B.; He, F.; Xu, P.; Liu, S. Tissue-Based Proteogenomics Reveals that Human Testis Endows Plentiful Missing Proteins. *J. Proteome Res.* **2015**, *14* (9), 3583–94.

(15) Wang, Y.; Chen, Y.; Zhang, Y.; Wei, W.; Li, Y.; Zhang, T.; He, F.; Gao, Y.; Xu, P. Multi-Protease Strategy Identifies Three PE2Missing Proteins in Human Testis Tissue. *J. Proteome Res.* **2017**, *16* (12), 4352–4363.

(16) Sun, J.; Shi, J.; Wang, Y.; Chen, Y.; Li, Y.; Kong, D.; Chang, L.; Liu, F.; Lv, Z.; Zhou, Y.; He, F.; Zhang, Y.; Xu, P. Multiproteases Combined with High-pH Reverse-Phase Separation Strategy Verified Fourteen Missing Proteins in Human Testis Tissue. *J. Proteome Res.* **2018**, DOI: 10.1021/acs.jproteome.8b00397.

(17) He, C.; Sun, J.; Shi, J.; Wang, Y.; Zhao, J.; Wu, S.; Chang, L.; Gao, H.; Liu, F.; Lv, Z.; He, F.; Zhang, Y.; Xu, P. Digging for Missing Proteins Using Low-Molecular-Weight Protein Enrichment and a "Mirror Protease" Strategy. *J. Proteome Res.* **2018**, DOI: 10.1021/acs.jproteome.8b00398.

(18) Huesgen, P. F.; Lange, P. F.; Rogers, L. D.; Solis, N.; Eckhard, U.; Kleifeld, O.; Goulas, T.; Gomis-Rüth, F. X.; Overall, C. M. LysargiNase Mirrors Trypsin for Protein C-Terminal and Methylation-Site Identification. *Nat. Methods* **2015**, *12* (1), 55–8.

(19) Siddiqui, O.; Zhang, H.; Guan, Y.; Omenn, G. S. Chromosome 17 Missing Proteins: Recent Progress and Future Directions as Part of the neXt-MP50 Challenge. *J. Proteome Res.* **2018**, DOI: 10.1021/acs.jproteome.8b00442.

(20) Paik, Y. K.; Lane, L.; Kawamura, T.; Chen, Y. J.; Cho, J. Y.; LaBaer, J.; Yoo, J. S.; Domont, G. B.; Corrales, F.; Omenn, G. S.; Archakov, A.; Encarnacion-Guevara, S.; Liu, S.; Salekdeh, G. H.; Cho, J. Y.; Kim, C. Y.; Overall, C. M. Launching the C-HPP Pilot Project for Functional Characterization of Identified Proteins with No Known Function. *J. Proteome Res.* **2018**, DOI: 10.1021/acs.jproteome.8b00383.

(21) Zhang, C.; Wei, X.; Omenn, G. S.; Zhang, Y. Structure and Protein Interaction-based Gene Ontology Annotations Reveal Likely Functions of Uncharacterized Proteins on Human Chromosome 17. *J. Proteome Res.* **2018**, DOI: 10.1021/acs.jproteome.8b00453.

(22) Melaine, N.; Com, E.; Bellaud, P.; Guillot, L.; Lagarrigue, M.; Morrice, N. A.; Guevel, B.; Lavigne, R.; Velez de la Calle, J. F.; Dojahn, J.; Pineau, C. Deciphering the Dark Proteome: Use of the Testis and Characterization of Two Dark Proteins. *J. Proteome Res.* **2018**, DOI: 10.1021/acs.jproteome.8b00387.

(23) Duek, P.; Gateau, A.; Bairoch, A.; Lane, L. Exploring the Uncharacterized Human Proteome Using neXtProt. *J. Proteome Res.* **2018**, DOI: 10.1021/acs.jproteome.8b00537.

(24) Marshall, N. C.; Klein, T.; Thejoe, M.; von Krosigk, N.; Kizhakkedathu, J.; Finlay, B. B.; Overall, C. M. Global Profiling of Proteolysis from the Mitochondrial Amino Terminome during Early Intrinsic Apoptosis Prior to Caspase-3 Activation. *J. Proteome Res.* **2018**, DOI: 10.1021/acs.jproteome.8b00675.

(25) Abbey, S. R.; Eckhard, U.; Solis, N.; Marino, G.; Matthew, I.; Overall, C. M. The Human Odontoblast Cell Layer and Dental Pulp Proteomes and N-Terminomes. *J. Dent. Res.* **2018**, *97* (3), 338–346.

(26) Monti, C.; Lane, L.; Fasano, M.; Alberio, T. Update of the Functional Mitochondrial Human Proteome Network. *J. Proteome Res.* **2018**, DOI: 10.1021/acs.jproteome.8b00447.

(27) Jeong, S. K.; Kim, C. Y.; Paik, Y.-K. ASV-ID, a Proteogenomic Workflow To Predict Candidate Protein Isoforms on the Basis of

Transcript Evidence. *J. Proteome Res.* **2018**, DOI: 10.1021/acs.jproteome.8b00548.

(28) Deutsch, E. W.; Orchard, S.; Binz, P. A.; Bittremieux, W.; Eisenacher, M.; Hermjakob, H.; Kawano, S.; Lam, H.; Mayer, G.; Menschaert, G.; Perez-Riverol, Y.; Salek, R. M.; Tabb, D. L.; Tenzer, S.; Vizcaino, J. A.; Walzer, M.; Jones, A. R. Proteomics Standards Initiative: Fifteen Years of Progress and Future Work. *J. Proteome Res.* **2017**, *16* (12), 4288−4298.

(29) Deutsch, E. W.; Perez-Riverol, Y.; Chalkley, R. J.; Wilhelm, M.; Tate, S.; Sachsenberg, T.; Walzer, M.; Kall, L.; Delanghe, B.; Bocker, S.; Schymanski, E. L.; Wilmes, P.; Dorfer, V.; Kuster, B.; Volders, P. J.; Jehmlich, N.; Vissers, J. P.C.; Wolan, D. W.; Wang, A. Y.; Mendoza, L.; Shofstahl, J.; Dowsey, A. W.; Griss, J.; Salek, R. M.; Neumann, S.; Binz, P. A.; Lam, H.; Vizcaino, J. A.; Bandeira, N.; Rost, H. Expanding the Use of Spectral Libraries in Proteomics. *J. Proteome Res.* **2018**, DOI: 10.1021/acs.jproteome.8b00485.

(30) Sajulga, R.; Mehta, S.; Kumar, P.; Johnson, J. E.; Guerrero, C. R.; Ryan, M. C.; Karchin, R.; Jagtap, P. D.; Griffin, T. J. Bridging the Chromosome-centric and Biology/Disease-driven Human Proteome Projects: Accessible and Automated Tools for Interpreting the Biological and Pathological Impact of Protein Sequence Variants Detected via Proteogenomics. *J. Proteome Res.* **2018**, DOI: 10.1021/acs.jproteome.8b00404.

(31) Mendoza, L.; Deutsch, E. W.; Sun, Z.; Campbell, D. S.; Shteynberg, D. D.; Moritz, R. L. Flexible and Fast Mapping of Peptides to a Proteome with ProteoMapper. *J. Proteome Res.* **2018**, DOI: 10.1021/acs.jproteome.8b00544.

(32) Pullman, B. S.; Wertz, J.; Carver, J.; Bandeira, N. ProteinExplorer: A Repository-Scale Resource for Exploration of Protein Detection in Public Mass Spectrometry Data Sets. *J. Proteome Res.* **2018**, DOI: 10.1021/acs.jproteome.8b00496.

(33) Ilgisonis, E. V.; Kopylov, A. T.; Ponomarenko, E. A.; Poverennaya, E. V.; Tikhonova, O. V.; Farafonova, T. E.; Novikova, S.; Lisitsa, A. V.; Zgoda, V. G.; Archakov, A. I. Increased Sensitivity of Mass Spectrometry by Alkaline Two-Dimensional Liquid Chromatography: Deep Cover of the Human Proteome in Gene-Centric Mode. *J. Proteome Res.* **2018**, DOI: 10.1021/acs.jproteome.8b00754.

(34) Narimatsu, H.; Kaji, H.; Vakhrushev, S. Y.; Clausen, H.; Zhang, H.; Noro, E.; Togayachi, A.; Nagai-Okatani, C.; Kuno, A.; Zou, X.; Cheng, L.; Tao, S. C.; Sun, Y. Current Technologies for Complex Glycoproteomics and Their Applications to Biology/Disease-Driven Glycoproteomics. *J. Proteome Res.* **2018**, DOI: 10.1021/acs.jproteome.8b00515.

(35) Boersema, P. J.; Melnik, A.; Hazenberg, B. P.C.; Rezeli, M.; Marko-Varga, G.; Kamiie, J.; Portelius, E.; Blennow, K.; Zubarev, R. A.; Polymenidou, M.; Picotti, P. Biology/Disease-Driven Initiative on Protein-Aggregation Diseases of the Human Proteome Project: Goals and Progress to Date. *J. Proteome Res.* **2018**, DOI: 10.1021/acs.jproteome.8b00401.

(36) Wang, J.; Choi, H.; Chung, N. C.; Cao, Q.; Ng, D. C.M.; Mirza, B.; Scruggs, S. B.; Wang, D.; Garlid, A. O.; Ping, P. Integrated Dissection of Cysteine Oxidative Post-translational Modification Proteome During Cardiac Hypertrophy. *J. Proteome Res.* **2018**, DOI: 10.1021/acs.jproteome.8b00372.

(37) Naryzhny, S. N.; Zorina, E. S.; Kopylov, A. T.; Zgoda, V. G.; Kleyst, O. A.; Archakov, A. I. Next Steps on in Silico 2DE Analyses of Chromosome 18 Proteoforms. *J. Proteome Res.* **2018**, DOI: 10.1021/acs.jproteome.8b00386.

(38) Lam, M. P.; Venkatraman, V.; Xing, Y.; Lau, E.; Cao, Q.; Ng, D. C.; Su, A. I.; Ge, J.; Van Eyk, J. E.; Ping, P. Data-Driven Approach To Determine Popular Proteins for Targeted Proteomics Translation of Six Organ Systems. *J. Proteome Res.* **2016**, *15* (11), 4126−4134.

(39) Yu, K. H.; Lee, T. M.; Wang, C. S.; Chen, Y. J.; Re, C.; Kou, S. C.; Chiang, J. H.; Kohane, I. S.; Snyder, M. Systematic Protein Prioritization for Targeted Proteomics Studies through Literature Mining. *J. Proteome Res.* **2018**, *17* (4), 1383−1396.

(40) Yu, K. H.; Lee, T. M.; Chen, Y. J.; Re, C.; Kou, S. C.; Chiang, J. H.; Snyder, M.; Kohane, I. S. A Cloud-Based Metabolite and Chemical Prioritization System for the Biology/Disease-Driven Human Proteome Project. *J. Proteome Res.* **2018**, DOI: 10.1021/acs.jproteome.8b00378.

(41) Lau, E.; Venkatraman, V.; Thomas, C. T.; Wu, J. C.; Van Eyk, J. E.; Lam, M. P. Y. Identifying High-Priority Proteins Across the Human Diseasome Using Semantic Similarity. *J. Proteome Res.* **2018**, DOI: 10.1021/acs.jproteome.8b00393.