Journal of proteome research

Editorial

pubs.acs.org/jpr

Progress and Future Direction of Chromosome-Centric Human Proteome Project

ABSTRACT: This special issue of *JPR* celebrates the fifth anniversary of the Chromosome-Centric Human Proteome Project (C-HPP). We present 27 manuscripts in four categories: (i) Metrics of Progress and Resources, (ii) Missing Protein Detection and Validation, (iii) Analytical Methods and Quality Assessment, and (iv) Protein Functions and Disease. We briefly introduce key messages from each paper, mostly from C-HPP teams and some from the Biology and Disease-driven HPP. From the first few months of the C-HPP NeXt-MP50 Missing Proteins Challenge, authors report 73 missing protein detections that meet the HPP guidelines using several novel approaches. Finally, we discuss future directions.

INTRODUCTION

This annual special issue (SI) is the fifth since publication of the first C-HPP edition in January 2013; 164 C-HPP-related papers have now appeared. In this issue, we present many papers that address the HPP goals of (1) progressively completing the human protein parts list and (2) advancing the usefulness of proteomics for the broader community.

METRICS OF PROGRESS AND RESOURCES FOR THE HUPO HUMAN PROTEOME PROJECT

Omenn et al.[1] updated the distribution of the proteins in each category of protein existence (PE), a degree of evidence for protein existence used in UniProtKB and neXtProt in which different number indicates evidence at the level of protein (PE1), transcript (PE2), sequence homology (PE3), prediction (PE4), and dubious sequences (PE5), respectively. Authors used the neXtProt version 2017-01-23 as a baseline for new findings of missing proteins (MPs) which correspond to PE2−4 proteins this year. They started with 17 008 PE1 proteins with reliable protein-level evidence of expression, strikingly up from 13 664 in 2013, even with more stringent HPP Guidelines v2.1 for Interpretation of Mass Spectrometry (MS) data (the HPP Guidelines).[2] That means that 2579 "MPs" remained to be detected. Of the 17 008 PE1 proteins, PeptideAtlas (2017-01) had 15 173 canonical proteins by mass spectrometry. Additionally, Human Protein Atlas v16 (2017) has 10 492 highly confident protein entries based on immunohistochemistry and immunofluorescence, with tissue, cell, and pathology atlases. Omenn et al.[1] discuss the application of the HPP Guidelines[2] and introduce a plan for adding guidelines that will cover data-independent acquisition (DIA) analyses, including SWATH-MS, inviting community inputs. They also assess the presently uncertain status of reports of translation products from long noncoding RNAs (lncRNAs) and small open reading frames (smORFs).

Building on these numbers, this issue reports the detection of 73 proteins that were listed as MPs as of 2017-01. Of course, the supporting evidence from these studies deposited via ProteomeXchange will be reanalyzed by PeptideAtlas; only then can the qualified proteins be submitted to neXtProt for their release in January 2018, when we will know how many new PE1 proteins were determined to be highly credible and have the new baseline for 2018 investigations.

Guided by the importance of open community standards and proteomics software tools, Deutsch et al.[3] shed light on the history of the HUPO Proteomics Standards Initiative (PSI), which has coordinated the development of many tools and standards (mzML, etc.) during the past 15 years (www.psidev.info). The authors envision synergistic outputs in combination with the ProteomeXchange Consortium. Using tools and standard metrics published by the PSI, two papers report updates on proteome database (DB) construction and the development of standard procedures. Alberio et al.[4] address standardization of the mitochondrial proteome DB, with an enrichment protocol for mitochondrial proteins and a proteomic analytical method for producing data sets using model cell lines. They created protein lists from 63 samples that can be used to map a network that represents the functional mitochondrial proteome and the protein−protein interactions within it.

Schwenk et al.[5] report a major update of the human plasma proteome with the Plasma PeptideAtlas 2017-04. They recall the early HUPO plasma proteome results that were very sensitive to criteria for protein matches, ranging from 9000 proteins with a single peptide to 1500 proteins with 3 peptides to 700−900 proteins with more stringent criteria. The current human plasma proteome has 3509 highly confident canonical plasma proteins that meet the HPP guidelines. They note an additional 1337 ambiguous matches and 1740 redundant matches. The authors apply the current guidelines to evaluate the stepwise progress of plasma proteome annotation in PeptideAtlas since 2006.[6] This article also addresses complex questions about correlation of RNA and protein abundances across tissues and in relation to plasma proteins and the detection of proteins in various subcellular fractions (intracellular, membrane, secreted). They then assess the progress in targeted MS in a variety of affinity assays, comparing detection with MS and with five commercial affinity approaches.

The current state of the detectable proteome is plagued by a "dark matter problem" of proteins that the current MS-based protein analysis approaches have difficulty detecting due to some combination of low abundance, lack of detectable tryptic peptides, high hydrophobicity in membranes, and protein families with high sequence homology. The result is the 2579 MPs. Mohamedali et al.[7] use the cochlear protein Prestin to illustrate these categories of undetected proteins and expression

Special Issue: Chromosome-Centric Human Proteome Project 2017

Published: December 1, 2017

DOI: 10.1021/acs.jproteome.7b00734
J. Proteome Res. 2017, 16, 4253−4258

in unusual or inaccessible cell types. They call for robust criteria for annotation of proteins that are not compatible with reliable detection by MS. We rely on the curation process of neXtProt, which evaluates many non-MS types of protein evidence; there are 1088 PE1 proteins in neXtProt not dependent upon MS data.[1]

MISSING PROTEIN DETECTION AND VALIDATION

In 2016 the C-HPP launched a two-year neXt-MP50 Challenge, calling upon each C-HPP team to develop a strategy for accelerating the detection of current MPs. The prototype for such analyses is reported by Duek et al.,[8] who deal with the 227 Chromosome (Chr) 2 + Chr 14 MPs, from which they identified 40 targets across the two Chr's for SRM targeted analysis. Reflecting the current importance of detection and validation of MPs in the work of the C-HPP, seven papers focused on new strategic workflows that use understudied specimens (kidney, bladder, spermatozoa, testis), enhanced sample preparation, data mining for stranded peptides combined with validation by comparison with SRMAtlas spectral libraries, or non-MS techniques to detect MP candidates.

Li et al.[9] from Chr 20 in China use Triton X-100/TBS solubilization plus ProteoMiner hexapeptide-covered beads as an enrichment/equalization strategy for low-abundance proteins by reducing the concentration of high-abundance proteins in four human cancer tissues (kidney, bladder, liver, colorectal). Among 20 MP candidates with two non-nested uniquely mapping peptides of ≥9 aa, 15 MP matched well the MS fragmentation patterns and chromatographic features of their synthetic peptides by PRM.[9] Also, they find that a high percentage of the total proteins identified with 1% FDR (protein level) have ≥2 uniquely mapping peptides of ≥9 aa, reinforcing this feature of the HPP Guidelines.

Carapito et al.,[10] from the combined Chr 2 and Chr 14 teams in Switzerland and France, carried forward their extensive work that yielded 253 missing protein detections from sperm in 2016,[11] applied the specific strategy published by Duek et al.,[8] identified 38 MP candidates from these two Chr's in sperm as part of the neXt-MP50 campaign, and demonstrated very good matches to synthetic peptides in PRM assays for 12 MP candidates. Further characterization by immunohistochemistry on human testes tissue sections and cytochemistry on sperm smears was performed for eight of these MPs with antibodies available from the Human Protein Atlas. They plan to seek lower abundance missing proteins in subcellular fractions of sperm. To assist the global neXt-MP50 Challenge, they will select new sets of MPs from other Chr's based on preferential expression in testis and perform targeted-MS searches in sperm extracts.

Wang et al.[12] from Chr 1 in China pursued a multiprotease strategy for the detection of MP candidates from their previously analyzed testis samples.[13] Multiprotease mixtures (trypsin, lysargiNase, GluC) improved both the peptide diversity and the sequence coverage, compared with use of trypsin alone. Indeed, the highest number of MP candidates was from the use of LysargiNase, which generated unique peptides and b-ion-rich spectra for 25 of the total of 30 candidate MPs. However, this paper illustrates the difficulty in finding MPs. From 7838 proteins the authors were not able to identify any MP candidates that fully complied with the HPP Guidelines.[2] Three did come close. The authors and the editors propose that beta-defensin 123 (Q8N688, chr 20q) and cancer/testis antigen family 45 member A10 (P0DMU9, chr Xq) be treated as instances of an "exceptional case", as stated in the HPP Guidelines.[2] Each has one peptide of ≥9 aa with seemingly high confidence after a spectrum quality check, isobaric PTM and single amino acid variant (SAAV) filtering, and verification of the spectrum with a synthesized peptide spectrum. The special feature here is overlapping peptides from different proteases. Golgi subfamily A member 6-like protein 1 (Q8N7Z2) also seems to qualify for an exception, based on overlapping 13 aa and 11 aa peptides from the combined use of LysargiNase with GluC, producing the sequence KLQAQVE-ENELWNRLNQQQEE (overlap underlined). The combined length of 21 aa exceeds the $9 \times 2 = 18$ for two peptides. Such an overlap is good for protein identification but must be distinguished from "nesting" of one peptide completely within another longer peptide, which does not necessarily increase confidence in protein identification. Hence, such nested peptides only count as "one" peptide with a maximum length of the encompassing larger peptide.

Peng et al.[14] investigate the phosphoproteome in kidney. Using kidney cancer and adjacent nontumor tissue, the authors claim 75 MP candidate detections among 8962 proteins and 6415 phosphoproteins identified. After rigorous screening and manual checking, 9 of these 75 MP candidate detections were subjected to validation against synthetic peptides and filtering for matches with SAAVs; unfortunately, only one MP, the phospho-protein transcription factor SIM 1, single-minded homologue 1 (P81133, Chr 6q), was confirmed according to the Guidelines v 2.1.

Using an integrated search engine strategy, Guruceaga et al.[15] from Chr 16 in Spain report 5 MP candidates (FREM3, Chr 4; LAMB4, Chr 7; MYEOV, Chr 11; RAD21L1, Chr 20; and TLDC2 Chr 20) from NCI 60 cell lines, based on a new workflow that integrated results coverage. Cell biology experiments revealed the involvement of these MP candidates in various cellular processes in the tissues where they are expressed—nervous, immune, muscle, secretory, reproductive, and internal system. At this time the confirmation is incomplete.

In another biological approach to characterize a missing protein, Meyfour et al.[16] use non-MS-based detection for the MP F-box-like/WD repeat-containing protein, TBL1Y, by measuring the differential expression of genes in the male-specific region of the Y Chr (MSY) and their X Chr counterparts during cardiac differentiation of human embryonic stem cells (hESCs). Whereas the two homologous proteins differ only in one tryptic peptide of ≥9 aa, they were distinguished with antibodies. TBL1Y expression showed an increase during differentiation, whereas TBL1X was decreased. Validation was achieved at the cellular level using gene knockdown assays during cardiogenesis. Thus TBL1Y is proposed for neXtProt consideration as PE1.

The final article on MP detection is a pan-proteome "data mining" approach by El Guoshy et al.[17] of the Chr X team in Japan. They accessed the MS-based PeptideAtlas and GPM databases to search for stranded peptides mapping to protein-coding genes across all of the chromosomes, with Chr 19 (27%) and Chr X (10%) among the most prominent. Most of the pairs of proteotypic non-nested peptides of ≥9 aa were found in GPMdb, although some were a combination of single peptides found in PeptideAtlas and in GPMdb. They then address the HPP Guideline for use of synthetic peptides by using reference spectra available in the SRMAtlas[18] to generate

DOI: 10.1021/acs.jproteome.7b00734
J. Proteome Res. 2017, 16, 4253−4258

402 MP candidates. The SRM matching exercise so far resulted in identifying 41 missing protein detections with ≥2 proteotypic peptides of ≥9 aa. Very good spectral matches to SRM Atlas synthetic peptides are illustrated with two proteotypic peptides each matching Krueppel-like factor 7 (O75840) and zinc finger protein 182 (P17025). This approach has the potential to identify many more MPs. A process for submitting the native and synthetic peptide data to ProteomeXchange is needed to facilitate review in PeptideAtlas and neXtProt. Meanwhile, their analysis of RNA expression data for PE1 versus PE2,3,4 proteins found the latter to have lower expression, shorter length, and fewer isoforms.

Overall, these different MP strategies and the initial findings from several groups illustrate the progress in detecting and validating MP candidates yet also show that the yield of MP detections is quite limited. We can summarize that there are 15 from Li et al, 12 from Carapito et al.,[10] 3 from Wang et al.,[12] 1 from Meyfour et al.,[16] 1 from Peng et al.,[14] and 41 from El Guoshy et al.,[17] for a total of 73. Of course, new findings come from throughout the proteomics community and are captured by PeptideAtlas and neXtProt; for example, at the HPP Workshop after the Dublin HUPO Congress it was pointed out that the total of PE1 proteins coded on Chr 17 has increased by 34 from the neXtProt 2016-01 baseline for the neXt-MP50 Challenge, as MPs have decreased from 148 in 2016-01 to 123 in 2017-01 to 114 in 2017-08.

■ ANALYTICAL METHODS AND QUALITY ASSESSMENT

The C-HPP faces several long-standing questions about the development of analytical methods to resolve MP issues. Four papers address these two questions: (i) How best can we improve methodological approaches to detect low-abundance MPs and (ii) how can we combine genomic resources (e.g., GENCODE, H-DBAS, DB-AT etc.) with proteome DBs to identify rare proteoforms as components of the human protein parts list?

Choong et al.[19] conducted an extensive comparative analysis of data sets in PeptideAtlas Tiered Human Integrated Search Proteome (THISP, 2017-03 release),[20] including the neXtProt (2017-01 release), to systematically investigate the possibility of unique peptides in MPs, unique peptides in dubious proteins, and variant peptides affected by 11 types of isobaric substitutions, causing doubtful identification results. They conclude that <5% of the unique peptides of missing proteins and <6% of variant peptides became shared with peptides of PE1 proteins after isobaric substitutions.

The use of ENCODE or GENCODE in proteome mapping can enable a better understanding of gene function.[21] Hwang et al.[22] used a streamlined pipeline composed of a customized neXtProt, in-house-built neXtPP, and GENCODE to identify MPs and ASVs with stringent FDR filtering (0.1% at the peptide level and 1% at the protein level) by screening previously reported data sets deposited from a rarely studied tissue, the human hippocampus (MSV000081385), in ProteomeXchange (PXD007166). Using these combined resources, they claimed two MP candidates (ARC, Chr 8; GRIK5, Chr 19) plus seven ASVs. Interestingly, the identified peptides of the GENCODE ASVs were mapped onto novel exon insertions, alternative translations at 5′-untranslated regions, or novel protein-coding sequences. When this scheme was applied to data sets of testis (PXD000561 and PXD002179) and spermatozoa (PXD003947), 52 GENCODE ASVs were identified. The combined use of genomic and proteomic DBs may help discovery of low-abundance proteoforms.

Toward development of bioinformatic tools for better detection of MP candidates, Cho et al.[23] report Epsilon-Q, an automated analyzer interface for mass-spectral library searching and label-free protein quantification. This tool is an upgraded version of Combo-Spec Search[24] and may strengthen the power of both the Combo-Spec Search method and label-free protein quantification. Epsilon-Q searches multiple spectral libraries, with control of class-specific false discovery rate and integration of results from identification and label-free quantification of proteins.

Meanwhile, Zhao et al.[25] assessed the performance of multiple combinations of seven search engines (MaxQuant, Mascot, OMSSA, X!Tandem, pFind, InsPecT, and ProVerB) for accurate identification of MP candidates, using the result of translating mRNA sequences from RNC-Seq as a standard. They searched their previously reported label-free MS data sets of human Hep3B, MHCCLM3, and MHCC97H liver and liver cancer cell lines. Compared with the use of all seven search engines, selective use of just two of the seven not only increased true-positive identifications but also reduced false-positives, with improved peptide coverage. They argued that the various algorithms contribute to identification of different peptides, which increases sequence coverage.

■ PROTEIN FUNCTIONS AND DISEASE

During the functional characterization of any protein (including PE1 proteins with at least one well-known function and especially PE1 proteins with no known function or inferred function, termed uPE1), many potential pitfalls and mistakes can be anticipated. However, with the continuing progress of the C-HPP and B/D-HPP, functional characterization of proteins has become productive.

Na et al.[26] present a proteogenomic approach to explore potential additional functions of a well-known PE1 protein, Na(+)/H(+) exchange regulatory cofactor 1 (NHERF1/SLC9A3R1; hereafter NHERF1). Using in vitro cell lines and an in vivo model in *Caenorhabditis elegans*, they found a new function of NHERF1 in reproduction, in connection with human preeclampsia (hPE). This study also provides a lesson about the importance of a careful comparative survey of multiple proteome DBs, especially over multiple versions, during a long-running experimental program. Zhang et al.[27] describe proteomic and metabolomics studies of hPE, seeking evidence for MPs in the detergent-insoluble fraction of placenta and noting increased HDL-C levels in metabolomic profiles of lipid pathways.

Turning our attention to developmental biology and disease outcomes, Meyfour et al.[28] of the Chr Y team explore the impact of Chr Y in organ development using human pluripotent stem cells. MSY genes show coexpression with their X homologues and play a key role in gonad formation, spermatogenesis, and heart and kidney development. Velásquez et al.[29] performed a neurobiological study of the synaptosomal proteome. Representing the Chr 15 team and B/D-HPP Brain Proteome Project in Brazil, they quantitatively analyzed synaptic proteins using label-free methods and iTRAQ labeling and identified a total of 67 dysregulated proteins, some of which are likely involved in regulation of human behavior through calcium signaling pathways and limbic-system-associated membrane proteins. This work provides a link

DOI: 10.1021/acs.jproteome.7b00734
J. Proteome Res. 2017, 16, 4253−4258

between protein expression patterns and social behavior in patients with schizophrenia. Disease biomarker discovery is an important potential deliverable of both the C-HPP and B/D-HPP work on molecular signatures and MP candidates.

Togayachi et al.[30] used a glycoproteomics approach to identify glycobiomarker candidates in small-cell lung carcinomas. They performed comparative immunohistochemical analysis and secretion profiling and found a fucosylated low-molecular-weight SgIII (short-form SgIII) secreted into the culture medium. This form of SgIII was also detected in the sera of patients with squamous cell lung cancers, suggesting its potential role as a noninvasive biomarker for this cancer. Of course, rigorous clinical validation is needed to determine the target stage of the cancer and the specificity, sensitivity, and positive predictive value of the biomarker candidate. Mora et al.[31] provide a good interface between their C-HPP Chr 16 team and B/D-Human Liver Proteome team. They apply the B/D-HPP strategy of building lists of organ-specific "popular proteins" from bibliometric analysis of the literature combined with SRM targeted proteomics analyses.[32] From a set of hepatocellular cancer cell (HCC)-related popular proteins, they selected four enzymes involved in metabolic remodeling of liver cancer cells. Using SRM, they describe adaptation of one-carbon metabolism in carcinogenesis (from fibrosis to cirrhosis to HCC) and during liver regeneration following exposure to CCl_4 in the mouse. This approach could prove useful in the surveillance of patients with chronic liver disease at risk of HCC.

Another case of metabolic peptide profiling with clinical application is the report by Zhang et al.[33] of the B/D-HPP Cardiovascular Disease team on development of an analytical system for peptidomes produced by cleavage of B-type natriuretic peptide (BNP) present in human plasma; BNPs are used routinely in the diagnosis and monitoring of heart failure. They developed a capillary electrophoresis separation system (CESI 8000) coupled to MS that can monitor the dynamic generation and breakdown of five BNP peptidoforms formed in plasma within an hour. Finally, Tholey et al.[34] propose that iMOP (the initiative on Model Organism Proteomes) can be used as a resource for the B/D-HPP.

CONCLUSIONS AND FUTURE DIRECTIONS

On the basis of the reports highlighted here, there are several strategies that can enhance the prospects of detecting and validating neXtProt defined MPs. Subject to reanalysis by PeptideAtlas and curation by neXtProt, the number of MPs may decline from 2579 to 2506. There was only a half-year between the announcement of the neXt-MP50 Challenge and the due date for submissions for this SI. We hope the yield of MP detections will be greater next year, enhanced by the methods presented here. Meanwhile, DIA/SWATH−MS is gaining popularity compared with SRM−MS. We will refine the HPP Guidelines to address DIA/SWATH−MS and may give more consideration to lncRNAs. In the B/D-HPP, the organ-specific popular proteins approach is proving useful in liver and cardiovascular studies and has high potential. A multiyear program lies ahead for characterization of multiple proteoforms of each expressed protein, with special attention to the >1200 uPE1 proteins lacking any known function.

Young-Ki Paik*
Yonsei Proteome Research Center and Department of Biochemistry, 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
Eric W. Deutsch
Institute for Systems Biology
Jennifer E. Van Eyk
Advanced Clinical BioSystems Research Institute, Department of Medicine, Cedars-Sinai Medical Centre
Gilbert S. Omenn*
Institute for Systems Biology
Departments of Computational Medicine & Bioinformatics, Internal Medicine, and Human Genetics and 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
Eric W. Deutsch: 0000-0001-8732-0928
Gilbert S. Omenn: 0000-0002-8976-6074

Notes

Views expressed in this editorial are those of the authors and not necessarily the views of the ACS.

ACKNOWLEDGMENTS

This work was supported by grants from the Korean Ministry of Health and Welfare (HI13C2098-International Consortium Project and HI16C0257 to Y.-K.P.) and from the U.S. National Institutes of Health (P30ES017885-01A1 and U24CA210967 to G.S.O.)

REFERENCES

(1) Omenn, G. S.; Lane, L.; Lundberg, E. K.; Overall, C. M.; Deutsch, E. W. Progress on the HUPO Draft Human Proteome: 2017 Metrics of the Human Proteome Project. *J. Proteome Res.* **2017**, DOI: 10.1021/acs.jproteome.7b00375.
(2) 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.
(3) 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**, DOI: 10.1021/acs.jproteome.7b00370.
(4) Alberio, T.; Pieroni, L.; Ronci, M.; Banfi, C.; Bongarzone, I.; Bottoni, P.; Brioschi, M.; Caterino, M.; Chinello, C.; Cormio, A.; Cozzolino, F.; Cunsolo, V.; Fontana, S.; Garavaglia, B.; Giusti, L.; Greco, V.; Lucacchini, A.; Maffioli, E.; Magni, F.; Monteleone, F.; Monti, M.; Monti, V.; Musicco, C.; Petrosillo, G.; Porcelli, V.; Saletti, R.; Scatena, R.; Soggiu, A.; Tedeschi, G.; Zilocchi, M.; Roncada, P.; Urbani, A.; Fasano, M. Toward the Standardization of Mitochondrial Proteomics: The Italian Mitochondrial Human Proteome Project Initiative. *J. Proteome Res.* **2017**, DOI: 10.1021/acs.jproteome.7b00350.
(5) Schwenk, J. M.; Omenn, G. S.; Sun, Z.; Campbell, D. S.; Baker, M. S.; Overall, C. M.; Aebersold, R.; Moritz, R. L.; Deutsch, E. W. The

Human Plasma Proteome Draft of 2017: Building on the Human Plasma PeptideAtlas from Mass Spectrometry and Complementary Assays. *J. Proteome Res.* **2017**, DOI: 10.1021/acs.jproteome.7b00467.

(6) Omenn, G. S.; States, D. J.; Adamski, M.; Blackwell, T. W.; Menon, R.; Hermjakob, H.; Apweiler, R.; Haab, B. B.; Simpson, R. J.; Eddes, J. S.; Kapp, E. A.; Moritz, R. L.; Chan, D. W.; Rai, A. J.; Admon, A.; Aebersold, R.; Eng, J.; Hancock, W. S.; Hefta, S. A.; Meyer, H.; Paik, Y. K.; Yoo, J. S.; Ping, P.; Pounds, J.; Adkins, J.; Qian, X.; Wang, R.; Wasinger, V.; Wu, C. Y.; Zhao, X.; Zeng, R.; Archakov, A.; Tsugita, A.; Beer, I.; Pandey, A.; Pisano, M.; Andrews, P.; Tammen, H.; Speicher, D. W.; Hanash, S. M. Overview of the HUPO Plasma Proteome Project: Results from the Pilot Phase with 35 Collaborating Laboratories and Multiple Analytical Groups, Generating a Core Dataset of 3020 Proteins and a Publicly-Available Database. *Proteomics* **2005**, *5* (13), 3226−3245.

(7) Mohamedali, A.; Ahn, S. B.; Sreenivasan, V. K. A.; Ranganathan, S.; Baker, M. S. Human Prestin: A Candidate PE1 Protein Lacking Stringent Mass Spectrometric Evidence? *J. Proteome Res.* **2017**, DOI: 10.1021/acs.jproteome.7b00354.

(8) Duek, P.; Bairoch, A.; Gateau, A.; Vandenbrouck, Y.; Lane, L. Missing Protein Landscape of Human Chromosomes 2 and 14: Progress and Current Status. *J. Proteome Res.* **2016**, *15* (11), 3971−3978.

(9) Li, S.; He, Y.; Lin, Z.; Xu, S.; Zhou, R.; Liang, F.; Wang, J.; Yang, H.; Liu, S.; Ren, Y. Digging More Missing Proteins Using an Enrichment Approach with ProteoMiner. *J. Proteome Res.* **2017**, DOI: 10.1021/acs.jproteome.7b00353.

(10) 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**, DOI: 10.1021/acs.jproteome.7b00374.

(11) Vandenbrouck, Y.; Lane, L.; Carapito, C.; Duek, P.; Rondel, K.; Bruley, C.; Macron, C.; Gonzalez de Peredo, A.; Couté, Y.; Chaoui, K.; Com, E.; Gateau, A.; Hesse, A. M.; Marcellin, M.; Méar, L.; Mouton-Barbosa, E.; Robin, T.; Burlet-Schiltz, O.; Cianferani, S.; Ferro, M.; Fréour, 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.

(12) 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**, DOI: 10.1021/acs.jproteome.7b00340.

(13) Wei, W.; Luo, W.; Wu, F.; Peng, X.; Zhang, Y.; Zhang, M.; Zhao, Y.; Su, N.; Qi, Y.; Chen, L.; Zhang, Y.; Wen, B.; He, F.; Xu, P. Deep Coverage Proteomics Identifies More Low-Abundance Missing Proteins in Human Testis Tissue with Q-Exactive HF Mass Spectrometer. *J. Proteome Res.* **2016**, *15* (11), 3988−3997.

(14) Peng, X.; Xu, F.; Liu, S.; Li, S.; Huang, Q.; Chang, L.; Wang, L.; Ma, X.; He, F.; Xu, P. Identification of Missing Proteins in the Phosphoproteome of Kidney Cancer. *J. Proteome Res.* **2017**, DOI: 10.1021/acs.jproteome.7b00332.

(15) Guruceaga, E.; Garin-Muga, A.; Prieto, G.; Bejarano, B.; Marcilla, M.; Marin-Vicente, C.; Perez-Riverol, Y.; Casal, J. I.; Vizcaino, J. A.; Corrales, F.; Segura, V. Enhanced Missing Proteins Detection in NCI60 Cell Lines Using an Integrative Search Engine Approach. *J. Proteome Res.* **2017**, DOI: 10.1021/acs.jproteome.7b00388.

(16) Meyfour, A.; Ansari, H.; Pahlavan, S.; Mirshahvaladi, S.; Rezaei-Tavirani, M.; Gourabi, H.; Baharvand, H.; Salekdeh, G. H. Y Chromosome Missing Protein, TBL1Y, May Play an Important Role in Cardiac Differentiation. *J. Proteome Res.* **2017**, DOI: 10.1021/acs.jproteome.7b00391.

(17) El Guoshy, A.; Hirao, Y.; Xu, B. O.; Saito, S.; Quadery, A. F.; Yamamoto, K.; Mitsui, T.; Yamamoto, T. Identification and Validation of Human Missing Proteins and Peptides in Public Proteome Databases; Data Mining Strategy. *J. Proteome Res.* **2017**, DOI: 10.1021/acs.jproteome.7b00423.

(18) Kusebauch, U.; Campbell, D. S.; Deutsch, E. W.; Chu, C. S.; Spicer, D. A.; Brusniak, M. Y.; Slagel, J.; Sun, Z.; Stevens, J.; Grimes, B.; Shteynberg, D.; Hoopmann, M. R.; Blattmann, P.; Ratushny, A. V.; Rinner, O.; Picotti, P.; Carapito, C.; Huang, C. Y.; Kapousouz, M.; Lam, H.; Tran, T.; Demir, E.; Aitchison, J. D.; Sander, C.; Hood, L.; Aebersold, R.; Moritz, R. L. Human SRMAtlas: A Resource of Targeted Assays to Quantify the Complete Human Proteome. *Cell* **2016**, *166* (3), 766−778.

(19) Choong, W. K.; Lih, T. M.; Chen, Y. J.; Sung, T. Y. Decoding the Effect of Isobaric Substitutions on Identifying Missing Proteins and Variant Peptides in Human Proteome. *J. Proteome Res.* **2017**, DOI: 10.1021/acs.jproteome.7b00342.

(20) Deutsch, E. W.; Sun, Z.; Campbell, D. S.; Binz, P. A.; Farrah, T.; Shteynberg, D.; Mendoza, L.; Omenn, G. S.; Moritz, R. L. Tiered Human Integrated Sequence Search Databases for Shotgun Proteomics. *J. Proteome Res.* **2016**, *15* (11), 4091−4100.

(21) Paik, Y. K.; Hancock, W. S. Uniting ENCODE with Genome-Wide Proteomics. *Nat. Biotechnol.* **2012**, *30* (11), 1065−1067.

(22) Hwang, H.; Park, G. W.; Park, J. Y.; Lee, H. K.; Lee, J. Y.; Jeong, J. E.; Park, S. R.; Yates, J. R.; Kwon, K. H.; Park, Y. M.; Lee, H. J.; Paik, Y. K.; Kim, J. Y.; Yoo, J. S. Next Generation Proteomic Pipeline for Chromosome-Based Proteomic Research Using NeXtProt and GENCODE Databases. *J. Proteome Res.* **2017**, DOI: 10.1021/acs.jproteome.7b00223.

(23) Cho, J. Y.; Lee, H. J.; Jeong, S. K.; Paik, Y. K. Epsilon-Q: An Automated Analyzer Interface for Mass Spectral Library Search and Label-Free Protein Quantification. *J. Proteome Res.* **2017**, DOI: 10.1021/acs.jproteome.6b01019.

(24) Cho, J. Y.; Lee, H. J.; Jeong, S. K.; Kim, K. Y.; Kwon, K. H.; Yoo, J. S.; Omenn, G. S.; Baker, M. S.; Hancock, W. S.; Paik, Y. K. Combination of Multiple Spectral Libraries Improves the Current Search Methods Used to Identify Missing Proteins in the Chromosome-Centric Human Proteome Project. *J. Proteome Res.* **2015**, *14* (12), 4959−66.

(25) Zhao, P.; Liu, W.; Zhao, J.; Zhong, J.; Zhang, G. Protein-Level Integration Strategy of Multi-Engine MS Spectra Search Results for Higher Confidence and Sequence Coverage. *J. Proteome Res.* **2017**, DOI: 10.1021/acs.jproteome.7b00463.

(26) Na, K.; Shin, H.; Cho, J. Y.; Jung, S. H.; Lim, J.; Lim, J. S.; Kim, E. A.; Kim, H. S.; Kang, A. R.; Kim, J. H.; Shin, J. M.; Jeong, S. K.; Kim, C. Y.; Park, J. Y.; Chung, H. M.; Omenn, G. S.; Hancock, W. S.; Paik, Y. K. a Systematic Proteogenomic Approach to Exploring a Novel Function for NHERF1 in Human Reproductive Disorder: Lessons for Exploring Missing Proteins. *J. Proteome Res.* **2017**, DOI: 10.1021/acs.jproteome.7b00146.

(27) Zhang, W.; Chen, X.; Yan, Z.; Chen, Y.; Cui, Y.; Chen, B.; Huang, C.; Zhang, W.; Yin, X.; He, Q. Y.; He, F.; Wang, T. Detergent-Insoluble Proteome Analysis Revealed Aberrantly Aggregated Proteins in Human Preeclampsia Placentas. *J. Proteome Res.* **2017**, DOI: 10.1021/acs.jproteome.7b00352.

(28) Meyfour, A.; Pooyan, P.; Pahlavan, S.; Rezaei-Tavirani, M.; Gourabi, H.; Baharvand, H.; Salekdeh, G. H. Chromosome-Centric Human Proteome Project Allies with Developmental Biology: a Case Study of the Role of Y chromosome Genes in Organ Development. *J. Proteome Res.* **2017**, DOI: 10.1021/acs.jproteome.7b00446.

(29) Velásquez, E.; Nogueira, F. C. S.; Velasquez, I.; Schmitt, A.; Falkai, P.; Domont, G. B.; Martins-de-Souza, D. Synaptosomal Proteome of the Orbitofrontal Cortex from Schizophrenia Patients Using Quantitative Label-Free and iTRAQ-Based Shotgun Proteomics. *J. Proteome Res.* **2017**, DOI: 10.1021/acs.jproteome.7b00422.

(30) Togayachi, A.; Iwaki, J.; Kaji, H.; Matsuzaki, H.; Kuno, A.; Hirao, Y.; Nomura, M.; Noguchi, M.; Ikehara, Y.; Narimatsu, H. Glycobiomarker, Fucosylated Short-Form Secretogranin III Levels Are Increased in the Serum of Patients with Small Cell Lung Carcinoma. *J. Proteome Res.* **2017**, DOI: 10.1021/acs.jproteome.7b00484.

(31) Mora, M. I.; Molina, M.; Odriozola, L.; Elortza, F.; Mato, J. M.; Sitek, B.; Zhang, P.; He, F.; Latasa, M. U.; Avila, M. A.; Corrales, F. J. Prioritizing Popular Proteins in Liver Cancer: Remodelling One-Carbon Metabolism. *J. Proteome Res.* **2017**, DOI: 10.1021/acs.jproteome.7b00390.

DOI: 10.1021/acs.jproteome.7b00734
J. Proteome Res. 2017, 16, 4253−4258

(32) 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.
(33) Zhang, S.; Raedschelders, K.; Santos, M.; Van Eyk, J. E. Profiling B-Type Natriuretic Peptide Cleavage Peptidoforms in Human Plasma by Capillary Electrophoresis with Electrospray Ionization Mass Spectrometry. *J. Proteome Res.* **2017**, DOI: 10.1021/acs.jproteome.7b00482.
(34) Tholey, A.; Taylor, N. L.; Heazlewood, J. L.; Bendixen, E. We Are Not Alone: The iMOP Initiative and Its Roles in a Biology and Disease Driven Human Proteome Project. *J. Proteome Res.* **2017**, DOI: 10.1021/acs.jproteome.7b00408.

DOI: 10.1021/acs.jproteome.7b00734
J. Proteome Res. 2017, 16, 4253−4258