Published in final edited form as:
J Proteome Res. 2016 November 4; 15(11): 3945–3950. doi:10.1021/acs.jproteome.6b00803.

Progress in the Chromosome-Centric Human Proteome Project as Highlighted in the Annual Special Issue IV

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

William S. Hancock, and
Department of Chemical Biology, Northeastern University

Gilbert S. Omenn[*]
Departments of Computational Medicine & Bioinformatics, Internal Medicine, and Human Genetics and School of Public Health, University of Michigan

The Chromosome-Centric Human Proteome Project (C-HPP) aims to identify and characterize as many as possible of the 19 467 proteins predicted from analysis of the human genome.[1–3] The number of predicted proteins with high-quality protein evidence or existence (PE) level of expression in human specimens has steadily risen each year for the past 4 years from 13 664 to 15 646, then 16 491, and now 16 537 confident PE1 protein identifications in neXtProt version 2016-08. The latest number is based on the stringent HPP Mass Spectrometry (MS) Data Interpretation Guidelines 2.1[4] adopted by the HPP and by the *Journal of Proteome Research* for this fourth annual Special Issue led by the C-HPP Consortium and as documented in the 2016 HPP Metrics paper by Omenn et al.[5] in this issue. Thus 85% of predicted human proteins have now been detected with high confidence using mass spectrometry or multiple other experimental methods, of which 75% have been detected via mass spectrometry alone. The HPP in 2013 decided to exclude PE5 "dubious" or "uncertain" predicted proteins from the denominator of predicted proteins so that the HPP "missing proteins" are those scored as PE2 or 3 or 4 and the percentage of PE1 proteins is PE1/(PE1+2+3+4) × 100. Before the work for this Special Issue the count of missing proteins was 2930. A notable highlight from these papers is the detection and confirmation of 267 of these 2930 missing proteins. Hence the number of missing proteins now stands at 2663, subject to standardized reanalysis by PeptideAtlas.

[*] Corresponding Authors: Y.-K.P.: paikyk@yonsei.ac.kr; Fax: +82-2-393-6589. [*] Corresponding Authors: G.S.O : gomenn@med.umich.edu; Fax: 734-615-6553.

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

The HPP issued revised Guidelines for Mass Spectrometry Data Interpretation (v2) in November 2015 (www.thehpp.org/guidelines); they were included in the Call for Papers for this Special Issue, along with a checklist adopted by HPP investigators at the Vancouver HUPO Congress in September 2015. Here Deutsch et al. (2016)[4] provide a detailed explanation for the 15 specific guidelines, now v2.1. Although the higher stringency implemented by neXtProt, PeptideAtlas, and HPP has inevitably raised the bar for identification of missing proteins, these quality criteria will lead to a more accurate and more reliable human proteome knowledgebase. We are pleased that the authors of the papers in this Special Issue have adopted these guidelines despite the extensive work involved in reanalysis of data that was required in many of the studies. The standardized reanalysis by PeptideAtlas will occur in the next few months in preparation for the 2017 cycle of research. As discussed below we highlight selected papers from the 17 published in the Special Issue that we consider of high relevance to the C-HPP.

Guidelines for Mass Spectrometry Data Interpretation

A major responsibility of the HPP has been the development of guidelines and resources to enhance the quality of mass spectrometry data in proteomics. The HUPO Proteomics Standards Initiative[6] was one of the very first HUPO activities and continues to be very active, working in conjunction with the HPP Bioinformatics Group and Knowledgebase Resource Pillar.[1] In 2012 the HPP issued guidelines for more confident identification of proteins, including insistence on a protein-level false discovery rate (FDR) of ≤1%. By 2015 it was clear from several independent reanalyzes of major published data sets that thousands of false-positive protein identifications resulted from use of lax filters.[7,8] Therefore, the HPP Bioinformatics group proposed and the HPP Investigators and Executive Committee decided to issue more stringent Guidelines for MS Data Interpretation. These guidelines are now fully implemented by neXtProt and by PeptideAtlas. The key elements are the requirement for protein-level FDR ≤1% and two uniquely mapping (proteotypic) peptides of at least nine amino acids in length for claims of finding missing proteins (neXtProt PE2+3+4) or novel proteins (from pseudogenes, lncRNAs, or PE5 dubious predicted proteins), careful scrutiny of the mass spectra, and search for matches to known abundant proteins with a single amino acid variant (SAAV) in the sequence or an isobaric PTM.[4] These considerations greatly strengthened the confirmation of many “missing protein candidates”. An assessment of the experience of authors in this Special Issue will drive the further evolution of the Guidelines.

Metrics for the Human Proteome: HPP 2016

In each annual C-HPP-led Special Issue there has been a Metrics paper summarizing the progress in confidently identifying the predicted proteins, with neXtProt as the primary curated resource, drawing upon PeptideAtlas for its standardized reanalysis of all available MS data sets. The HPP also monitors MS findings in the GPMDB. The HPP draws upon its Affinity Capture Resource Pillar led by the Human Protein Atlas for extensive data on tissue and subcellular expression of predicted proteins from immunohistochemistry and immunofluorescence. The HPA has recently added extensive transcriptomic analyses pointing the way to tissue-specific searches for missing proteins, as exploited for spermatozoa and testis in this issue. The Metrics paper also provides a substantial discussion about topical themes. In 2015 the theme was quality of interpretation of large data sets from proteomics and proteogenomics publications. In 2016 the theme is the growing information and annotation of proteoforms arising from mutations (sequence variants), alternative splicing, and multiple types of post-translational modifications (PTMs).

Over the past 4 years, the Metrics paper has documented the number of predicted proteins with high-quality protein-level evidence of expression in human specimens corresponding to PE1 in neXtProt as the baseline for the C-HPP analyses, which, as stated above, was 16 518 in 2015 (neXtProt version 2016-02). Notably, the adoption by neXtProt of the more stringent HPP Guidelines reduced the PE1 total by 485 from what the total would have been with the 2014 criteria of two peptides of at least seven amino acid residues or one peptide of at least nine amino acid residues and without the scrutiny for sequence variants and isobaric PTMs that make many abundant proteins match with newly identified peptides.[9] The paper provides extensive comments about the resources utilized by HPP. For example, a quality initiative by the Human Protein Atlas parallels the more stringent guidelines for mass spectrometry described above.

Not all predicted proteins are detectable, as Duek et al.[10] thoroughly discuss in their Perspective on the prospects for identifying all 1231 predicted proteins from chromosome 2 and 624 predicted proteins from chromosome 14. Of these, 134 and 93 entries, respectively, are not experimentally validated as of now and are considered missing proteins (they excluded 18 and 17 PE5 entries, respectively). These 227 represent 7.7% of the 2949 missing proteins in neXtProt 2016-08. Of these 227, some may never be detected because they have no significant levels of their cognate transcripts in any tissue specimens so far studied; in some cases, DNase hypersensitivity assays show that the genes or chromosomal segments themselves are inaccessible. Maybe these genes are expressed only in early fetal development or in unusual tissue types not yet studied or under environmental or pathological perturbations that induce expression. Other limitations are due to current mass spectrometry methods, incompatible biochemical features like insolubility within membranes, hydrophobicity, abundance below the limits of detection, or failure to yield tryptic peptides of 7 to 30 or perhaps even 50 amino acid residues upon digestion with trypsin or similar peptides from other preparative proteases that could be utilized. In addition, Duek et al.[10] find that there are 34 missing proteins on Chr 2 and 11 on Chr 14 that have mass spectrometry information but do not meet the HPP guidelines and so are potentially ambiguous but are included as *missing protein candidates* (for HPP accepted definitions, see Deutsch et al.[4]) to stimulate others to follow up on these protein candidates in other tissues or by orthogonal techniques. The workflow of Duek et al. identified 99 proteins that remain missing but are amenable for further investigation. Duek et al. put aside olfactory receptors and pseudogenes after noting that the higher percentage of missing proteins on Chr 14 compared with Chr 2 was mostly due to a large cluster of 27 olfactory receptor genes at 14q11.2 compared with only two olfactory receptor genes on Chr 2. On the basis of information from the literature and from the Human Protein Atlas (immunohistochemistry), Peptide Atlas, and neXtProt, they then identified a subset of 40 missing proteins (25 on Chr 2 and 15 on Chr 14) that were considered theoretically detectable; 38 have transcripts in testis or spermatozoa and 2 others have single peptide hits in spermatozoa. Duek et al. propose these 40 as a priority roadmap for the Swiss and French teams to use in designing targeted proteomics analyses of spermatozoa and testis in the near future. We highly recommend that all C-HPP and Biology/Disease-driven (B/D)-HPP investigators carefully study the Duek et al. paper as a guide to organizing their priorities for the C-HPP Top 50 Marathon Challenge for finding missing proteins on their chromosome or highly informative proteins for biological and disease-oriented studies.

Choosing the Tissues Most Likely to be Expressing Missing Proteins: Diving Deep into Testis and Spermatozoa Proteomes

Each chromosome team of the C-HPP is working toward identifying the missing proteins. One of the major strategies others and we have championed is to take a deep dive into proteomic analyses of tissues reported to have good evidence of transcript expression of the genes of interest (i.e., neXtProt PE2; Human Protein Atlas). The Chromosome 2 and Chromosome 14 teams of the C-HPP have exploited this approach admirably with their focus on testis and spermatozoa proteomes. Last year the Human Protein Atlas reported from transcriptome analyses that, among all of the tissues and organs of the body, testis has by far the largest number of tissue-specific transcripts (50×) or highly enriched transcripts (5×) more than any other tissue type, with an initial total of 999 predicted proteins,[11] adjusted this year to 879 as described by Omenn et al.[5] in this Special Issue.

In the 2015 C-HPP Special Issue, Zhang et al.[12] reported 166 previously missing proteins in testis, and Jumeau et al.[13] reported 89 more in spermatozoa. Together they contributed to a total of 354 in PeptideAtlas 2016–01, the most recent update for this 2016 C-HPP Special Issue. This left 525 still missing proteins in testis and spermatozoa. The same teams have continued their analyses and reported further identifications this year, with 47 more in testis from Wei et al.[14] and 206 more in spermatozoa from Vandenbrouck et al.[15] These teams used SRM and immunohistochemistry to strengthen evidence of their identifications. Wei et al.[14] used a fast-scanning, high-resolution Q Exactive HF instrument to detect many low-abundance missing proteins from a total of 8526 proteins that are supported by transcript expression in testis. They identified 81 missing protein candidates, including 15 membrane proteins, 14 testis-specific proteins, and 9 spermatogenesis-related proteins; of these, 47 candidates survived scrutiny of the spectra, the requirements for two uniquely mapping (proteotypic) peptides of at least nine amino acid residues in length, and a search for potential matches to more abundant proteins with single amino acid variants or isobaric PTM modifications.[4] This also emphasizes the difficulty in closing the human proteome gap with high-quality unambiguous MS data and the necessity for further targeted orthogonal studies and strategies to reveal shy proteins. Vandenbrouck et al.[15] started with 235 missing protein candidates in spermatozoa, of which 206 met the HPP guidelines v2.1. Some were confirmed with immunohistochemistry, such as orphan proteins CXorf58 and C19orf81. Developmentally expressed proteins in spermatozoa included CXorf58, C20orf85, CFAP46, FAM1876B, and AXDND1. Interestingly this group also reported four matches to PE5 neXtProt predicted proteins, but we caution that further curation must be performed on these four PE5 proteins at Swiss-Prot.

Finding Better Ways to Solubilize Membrane-Embedded Proteins

Membrane proteins are estimated to represent 20–30% of the total encoded human proteome and are notoriously difficult to identify by MS. Zhao et al.[16] utilized an MCF7 ER+/PR+ breast cancer cell line lysate to evaluate alternative methods for membrane protein solubilization and enrichment. The authors compared ultracentrifugation and detergent-based extraction with in-solution digestion with detergents and enhanced filter-aided sample preparation with detergents and then with in-gel digestion plus SDS detergent for a total of five digestion methods. Among the reagents, sodium deoxycholate (SDC) and RapiGest showed good solubility and enzyme compatibility as well as easy removal in later steps prior to MS analysis. In-solution digestion with RapiGest and eFASP with SDC methods were time-saving and consistent in identification of membrane proteins; they produced the most identifications and the most potential missing proteins. The numbers of membrane proteins identified ranged from 1069 to 1125 in different samples and by different methods for a combined total of 1345 having two or more uniquely mapping peptides consistent with the HPP Guidelines. This number is notable as it is larger than all previously reported membrane protein data sets; indeed the largest overlap was with Muraoka et al.[17] from the first C-HPP Special Issue. Surprisingly, they found only 13 unique peptides matching to 8 missing proteins across the proteome; spectral matches to synthetic peptides validated eight such peptides. In the end, the authors confirmed two as missing proteins (Q6UWH6 and Q8IZD6) and noted two others as missing protein candidates (O75474 and Q3SY17/Q9H1U9, highly homologous solute carrier proteins).

Selected Reaction Monitoring Mass Spectrometry for Targeted Identification of Missing Proteins

The advancement of selected reaction monitoring (SRM) methodology has been an achievement of the B/D-HPP and mass spectrometry leaders in the Human Proteome Project, building on the original scheme of Anderson.[18] SRM offers efficient assays for a wide variety of life science research and should help integrate proteomics into broader omics research, a specific goal of the HPP. Indeed, in recent years the C-HPP has recommended that teams utilize SRM assays to confirm identifications of missing proteins and to target protein mixtures for evidence of specific proteins, both known and missing.[19] As published in the July issue of *Cell*, Kusebauch et al.[20] presented an exhaustive resource of targeted assays to quantify the complete human proteome, the Human SRMAtlas. This mammoth resource enables accurate detection and quantification of any known or predicted human protein from complex biological specimens. The SRMAtlas contains 166 000 proteotypic peptides, verified high-resolution spectra, multiplexed SRM assays, and a web database with unlimited free access (www.srmatlas.org). Reported are assays for splice variants, nonsynonymous mutations (single amino acid variants), and many PTMs. The article includes examples of investigating the network response to inhibition of cholesterol synthesis in liver cells and to docetaxel treatment of prostate cancer cell lines.

As noted above, SRMs were used very effectively by Vandenbrouck et al.[15] to confirm missing protein identifications in testis, and they are at the heart of the targeted approach for detecting the 40 missing proteins proposed by Duek et al.[10] SRMs were also used by Zhao et al.[16] to confirm missing proteins identified in the membrane enrichment studies. Poverennaya et al.[21] from the Chromosome 18 team used stable isotope-labeled standards for SRM (SRM/SIS) analyses, as well as shotgun LC–MS/MS, of normal liver and HepG2 liver cells. According to neXtProt 2016-02, there were only 24 (PE2+3+4) missing proteins (plus 9 PE5 dubious proteins) encoded on Chr 18. However, none of these 33 missing and dubious proteins was detected by either shotgun MS/MS or SRM in this study. A complementary study of plasma samples from 54 healthy male volunteers who were astronaut candidates (age 20–47) used SRM assays with 267 stable isotope-labeled peptide standards to estimate the levels of 84 of the 276 Chr 18-encoded proteins.[22] Thus, this study nicely highlights the desired result of the stricter HPP guidelines that resulted in a more accurate and higher confidence map of the chromosome 18 proteins from one year to the next.

Biological Studies of Colon Cancer Cell Lines

Multiple C-HPP teams have organized biological studies of cancers. In previous years there have been many papers about liver cancers from the Chinese HPP consortium addressing chromosomes 1, 8, and 20 and Korean consortium for Chr 9, 11, 13; glioblastoma from the Indian consortium for Chr 12; gastric cancers from the Brazilian Chr 15 team; and breast cancers from the U.S. Chr 17 team. This year there are two articles investigating colon cancer cell lines. Guo et al. from the Wang Lab[23] characterized the phosphoproteome of SW620 cell-derived exosomes. They reported 313 phosphoproteins with 1091 phosphosites, including 202 new phosphosites using highly optimized MS analysis of exosomal and cellular proteins. Genes for exosomal phosphoproteins were enriched on Chr 11. The high percentage of P-tyrosyl proteins (6.4%), possibly involved in cytoskeletal remodeling, including the ephrin signaling pathway in cell–cell communication, was notable. Guo et al. from the Liu Lab[24] studied drug resistance in relation to aneuploidy and gene expression profiles in oxiliplatin- and irinotecan-resistant colorectal cancer cell lines. Comparative proteogenomic analysis indicated a link between an increase in gene copies in Chr 14 and aneuploidy in LoVo cells (trisomy of 5, 7, 12, and 15, with proportional increases in mRNA transcripts and proteins coded on those chromosomes). Comparison of near-diploid HT116 cells with aneuploidy in LoVo cells suggested a possible, but rather uncertain, contribution of aneuploidy to drug resistance.

Bioinformatics Studies

Park et al.[25] of the Chr 11 team present an Integrated Proteomic Pipeline (IPP) using multiple search engines (SEQUEST, MASCOT, MS-GF+) for LC–MS/MS analyses of brain tissues with controlled FDR ≤ 1% at the protein level. They compared IPP to a conventional proteomic pipeline. In hippocampal tissue, IPP yielded 5756 proteins, including 477 alternative splice variants (ASVs) versus 4453 proteins and 182 ASVs using CPP. They reported 12 missing proteins validated by MS and SRM with synthetic peptides. Deutsch et al.[26] compared components of Tiered Human Integrated Sequence Search Databases for Shotgun Proteomics (THISP), with emphasis on the completeness and efficiency of the alternative searches. The results of analysis of shotgun proteomics studies can be greatly affected by the selection of the reference protein sequence database against which the spectra are matched. A common flaw is the use of outdated or unsuited reference databases. Deutsch et al.[26] compiled a tiered set of four sequence databases of varying sizes, from a small database consisting of only the ~20 000 primary isoforms plus contaminants to a very large database that includes almost all nonredundant protein sequences from several sources. They compared the performance with two data sets, from HeLa cells and from normal liver tissue. Modest percentage increments of additional peptides (0.8, 1.1, and 1.5% for Tiers Levels 2, 3, and 4, respectively) were obtained but at substantially increasing computational cost. Even when a complex database is not used for direct processing of mass spectra, a complex database can be used to ensure that peptides that seem to be uniquely mapping in the smaller databases do not have additional mappings to other proteins when potential variants are considered. These resources are automatically updated monthly at PeptideAtlas.

Similarly, it is important to always use the most recent version of neXtProt to compare data from any one study before publication. Although this was not done for Garin et al.,[27] they did report which of the missing proteins that emerged from their new pipeline were no longer missing in the most recent version of neXtProt in their tables. This team designed a pipeline to reanalyze data sets from the PRIDE Archive to examine the many features that influence the detectability of missing proteins, applied to spermatozoa and the HEK293 human embryonic kidney cell line. This paper builds on the missing protein map of Guruceaga et al.[28] After functional analysis of missing proteins and quantitative confirmation with SRM assays, they reported expression of two missing proteins in spermatozoa (e.g., DNAH3 and TEPP) and two missing proteins in HEK293 cells (e.g., UNCX and ATAD3C) among those 97 potential missing proteins. Kim et al.[29] report on the development of gFinder, a web-based bioinformatics tool for the analysis of N-glycopeptides, which can be used for analyzing mixtures of native N-glycopeptides. They demonstrate that the simultaneous integration of collision-induced dissociation (CID) and high-energy collisional dissociation (HCD) fragmentation facilitates rapid identification of both glycans and N-glycopeptide backbones in tandem MS data, resulting in the identification of missing proteins having glycans (e.g., Q8N9B8) present in liver cancer (http://gFinder.proteomix.org/). Panwar et al.[30] from the Chr 17 team presented a proteome-wide analysis of predicted splice isoform-level functional networks. The accumulation of a large amount of RNA-seq data in the public domain greatly increases our ability to examine the functional annotation of genes at isoform level. Panwar et al.[30] used a multiple instance learning (MIL)-based approach to predict the function of protein-coding splice variants with transcript-level expression values and gene-level functional associations from the Gene Ontology database. A support vector machine (SVM)-based 5-fold cross-validation technique was applied. Comparatively, genes with splice variants performed better than single isoform genes. Predictions were illustrated using literature evidence of ADAM15, LMNA/C, and DMXL2 genes. All predictions and functional networks are available to the community in “IsoFunc” at http://guanlab.ccmb.med.umich.edu.proxy.lib.umich.edu/isofunc.

Emerging C-HPP Initiatives to Stimulate Collaboration

To expedite the further discovery and annotation of missing proteins, the C-HPP leadership has proposed two new initiatives for the C-HPP. The first is a collaborative effort, introduced at the 2015 HUPO Congress, called “clusters” of C-HPP teams to accelerate work on cancers (Chr 1, 8, 20, 9, 11, 13, 7, 12, 17), reproductive biology (Chr 2, 14, X, Y), and the broad application of the In Vitro Transcription/Translation platform (IVTT) (Chr 5, 10, 15, 16, 19). Chinese and Korean teams are exploring sharing their specialty database on RNA-Seq and MS profiling of SAAVs; other topics might be roles of amplicons, cancer stem cells, single-cell analyses using Cy-Tof, and differential expression of splice isoforms of key proteins in cancer pathways. In 2016 Taipei HUPO Congress, additional groups on membrane proteins (Chr 4, 18, 21) and neuro-degenerative disorders (Chr 1, 3, 6, 11, 12) have been organized. A special attribute of the clusters is expected to be collaboration also with the corresponding B/D-HPP groups (cancers and CPTAC, reproductive biology, and brain). This collaborative strategy will be applied to all areas of the HPP, from missing protein and proteoform discovery and validation to data integration, disease mechanisms, and biological studies. The second accelerator is an effort to build “Top 50 Missing Protein Marathon Challenge”. This new campaign is initiated based on the Chr 16 analyses and Chr 2/Chr 14 collaboration to identify the most likely specimens and methods to detect specific missing proteins. Although several chromosomes have fewer than 50 missing proteins (see chromosome-by-chromosome tables and figure in the Metrics paper),[6] whereas at the other extreme, several have more than 200 missing proteins, “Top 50 Missing Protein Marathon Challenge” will encourage each chromosome team to specifically discover 50 missing proteins from their chromosome in the next 2 years to drive the completion of a nearly complete draft of the human proteome.

Conclusions and Perspectives

It is certain that not all 20 055 PE1-5 or 19 467 PE1-4 predicted proteins will be detectable with typical tissue specimens and technologies. We may be fast approaching saturation; the best test is to design targeted studies using SRM-MS or SWATH-MS in the most likely specimens to show expression of the protein along with the transcript. This challenge was refined and activated at the HUPO-2016 Congress. Because most proteomics analyses are chromosome agnostic, it may be wise for the B/D-HPP and C-HPP teams to organize a major collaborative effort on jointly studying the most likely detectable remaining missing proteins rather than relying on each chromosome team to work in isolation on the missing proteins encoded by its chromosome. The B/D-HPP was formed several years ago on the foundation of the original HUPO proteome projects: plasma, brain, liver, cardiovascular, kidney/urine, diabetes, and glycoproteomics. Now there are 22 B/D-HPP teams, including eight formed since 2013 (Van Eyk et al).[31] Cross-B/D initiatives include the development of SRM strategies, reagents, spectral libraries, and assays[20] and the popular protein sets for targeted proteomics studies based on bibliometric studies of Lam et al.[32] for the cardiovascular, cerebral, hepatic, renal, pulmonary, and intestinal organ systems. When the C-HPP was officially launched in 2012,[2,3] the first phase of identifying proteins and proteoforms was projected to require 6 years, followed by a second phase of extensive biological studies for 4 years. Together with the broad proteomics community, we are making good progress.

Acknowledgments

Part of this work was supported by a grant from the Korean Ministry of Health and Welfare for the Global C-HPP Research (to Y.-K.P., H1132C2098) and from the National Institutes of Health ES017885 (to G.S.O).

References

1. Legrain P, Aebersold R, Archakov A, Bairoch A, Bala K, Beretta L, Bergeron J, Borchers CH, Corthals GL, Costello CE. The human proteome project: current state and future direction. Mol Cell Proteomics. 2011; 10(7) M111.009993.
2. Paik YK, Jeong SK, Omenn GS, Uhlen M, Hanash S. The Chromosome-Centric Human Proteome Project for cataloging proteins encoded in the genome. Nat Biotechnol. 2012; 30:221–3. [PubMed: 22398612]
3. Paik YK, Omenn GS, Uhlen M, Hanash S, Marko-Varga G. Standard guidelines for the chromosome-centric human proteome project. J Proteome Res. 2012; 11(4):2005–13. [PubMed: 22443261]
4. Deutsch EW, Overall CM, Van Eyk JE, Baker MS, Paik YK, Weintraub ST, Lane L, Martens L, Vandenbrouck Y, Kusebauch U, Hancock WS, Hermjakob H, Aebersold R, Moritz RL, Omenn GS. Human Proteome Project Mass Spectrometry Data Interpretation Guidelines 2.1. J Proteome Res. 2016; doi: 10.1021/acs.jproteome.6b00392
5. Omenn GS, Lane L, Lundberg EK, Beavis RC, Overall CM, Deutsch EW. Metrics for the Human Proteome Project 2016: Progress on Identifying and Characterizing the Human Proteome, Including Post-Translational Modifications. J Proteome Res. 2016; doi: 10.1021/acs.jproteome.6b00511
6. Deutsch EW, Albar JP, Binz PA, Eisenacher M, Jones AR. Development of data representation standards by the human proteome organization proteomics standards initiative. J Am Med Inform Assoc. 2015; 22(3):495–506. [PubMed: 25726569]
7. Omenn GS, Lane L, Lundberg EK, Beavis RC, Nesvizhskii AI, Deutsch EW. Metrics for the Human Proteome Project 2015: Progress on the Human Proteome and Guidelines for High-Confidence Protein Identification. J Proteome Res. 2015; 14(9):3452–60. [PubMed: 26155816]
8. Savitski MM, Wilhelm M, Hahne H, Kuster B, Bantscheff M. A Scalable Approach for Protein False Discovery Rate Estimation in Large Proteomic Data Sets. Mol Cell Proteomics. 2015; 14(9): 2394–404. [PubMed: 25987413]
9. Lane L, Bairoch A, Beavis RC, Deutsch EW, Gaudet P, Lundberg E, Omenn GS. Metrics for the Human Proteome Project 2013-2014 and strategies for finding missing proteins. J Proteome Res. 2014; 13(1):15–20. [PubMed: 24364385]
10. 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; doi: 10.1021/acs.jproteome.6b00443
11. Uhlén M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P. Tissue-based map of the human proteome. Science. 2015; 347(6220):1260419. [PubMed: 25613900]
12. Zhang Y, Li Q, Wu F, Zhou R, Qi Y, Su N. Tissue-Based Proteogenomics Reveals that Human Testis Endows Plentiful Missing Proteins. J Proteome Res. 2015; 14(9):3583–94. [PubMed: 26282447]
13. Jumeau F, Com E, Lane L, Duek P, Lagarrigue M, Lavigne R. Human Spermatozoa as a Model for Detecting Missing Proteins in the Context of the Chromosome-Centric Human Proteome Project. J Proteome Res. 2015; 14(9):3606–20. [PubMed: 26168773]
14. Wei W, Luo W, Wu F, Peng X, Zhang Y, Zhang M, Zhao Y, Su N, Qi Y, Chen L. Deep Coverage Proteomics Identifies More Low-Abundance Missing Proteins in Human Testis Tissue with Q-Exactive HF Mass Spectrometer. J Proteome Res. 2016; doi: 10.1021/acs.jproteome.6b00390
15. Vandenbrouck Y, Lane L, Carapito C, Duek P, Rondel K, Bruley C, Macron C, Gonzalez de Peredo A, Coute Y, Chaoui K. Looking for Missing Proteins in the Proteome of Human Spermatozoa: An Update. J Proteome Res. 2016; doi: 10.1021/acs.jproteome.6b00400

16. Zhao M, Wei W, Cheng L, Zhang Y, Wu F, He F, Xu P. Searching Missing Proteins Based on the Optimization of Membrane Protein Enrichment and Digestion Process. J Proteome Res. 2016; doi: 10.1021/acs.jproteome.6b00389

17. Muraoka S, Kume H, Adachi J, Shiromizu T, Watanabe S, Masuda T, Ishihama Y, Tomonaga T. In-depth membrane proteomic study of breast cancer tissues for the generation of a chromosome-based protein list. J Proteome Res. 2013; 12(1):208–13. [PubMed: 23153008]

18. 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). J Proteome Res. 2004; 3:235–244. [PubMed: 15113099]

19. Jeong SK, Hancock WS, Paik YK. Genomewide PDB 2.0: A Newly Upgraded Versatile Proteogenomic Database for the Chromosome-Centric Human Proteome Project. J Proteome Res. 2015; 14(9):3710–3719. [PubMed: 26272709]

20. Kusebauch U, Campbell DS, Deutsch EW, Chu CS, Spicer DA, Brusniak MY. Human SRMAtlas: A resource of targeted assays to quantify the complete human proteome. Cell. 2016; 166:766–778. [PubMed: 27453469]

21. Poverennaya EV, Kopylov AT, Ponomarenko EA, Ilgisonis EV, Zgoda VG, Tikhonova OV, Novikova SE, Farafonova TE, Kiseleva YY, Radko SP, Vakhrushev IV. State of the Art of Chromosome 18-Centric HPP in 2016: Transcriptome and Proteome Profiling of Liver Tissue and HepG2 Cells. J Proteome Res. 2016; doi: 10.1021/acs.jproteome.6b00380

22. Kopylov AT, Ilgisonis EV, Moysa AA, Tikhonova OV, Zavialova MG, Novikova SE, Lisitsa AV, Ponomarenko EA, Moshkovskii SA, Markin AA. Targeted Quantitative Screening of Chromosome 18 Encoded Proteome in Plasma Samples of Astronaut Candidates. J Proteome Res. 2016; doi: 10.1021/acs.jproteome.6b00384

23. Guo W, Jiang C, Yang L, Li T, Liu X, Jin M, Qu K, Chen H, Jin X, Liu H. Quantitative Metabolomic Profiling of Plasma, Urine and Liver Extracts by 1H NMR Spectroscopy Characterizes Different Stages of Atherosclerosis in Hamsters. J Proteome Res. 2016; doi: 10.1021/acs.jproteome.6b00179

24. Guo J, Xu S, Huang X, Li L, Zhang C, Pan Q, Ren Z, Zhou R, Ren Y, Zi J. Drug Resistance in Colorectal Cancer Cell Lines is Partially Associated with Aneuploidy Status in Light of Profiling Gene Expression. J Proteome Res. 2016; doi: 10.1021/acs.jproteome.6b00387

25. Park GW, Hwang H, Kim KH, Lee JY, Lee HK, Park JY, Ji ES, Park SKR, Yates JR, Kwon KH. Integrated Proteomic Pipeline Using Multiple Search Engines for a Proteogenomic Study with a Controlled Protein False Discovery Rate. J Proteome Res. 2016; doi: 10.1021/acs.jproteome.6b00376

26. Deutsch EW, Sun Z, Campbell DS, Binz PA, Farrah T, Shteynberg D, Mendoza L, Omenn GS, Moritz RL. TieredHuman Integrated Sequence Search Databases for Shotgun Proteomics. J Proteome Res. 2016; doi: 10.1021/acs.jproteome.6b00445

27. Garin A, Odriozola L, Martinez-Val A, del Toro N, Martínez R, Molina M, Cantero L, Rivera R, Garrido N, Dominguez F, Vizcaino JA, Corrales F, Segura V. Detection of missing proteins using the PRIDE database as a source of mass-spectrometry evidence. J Proteome Res. 2016; doi: 10.1021/acs.jproteome.6b00437

28. Guruceaga E, Sanchez del Pino MM, Corrales FJ, Segura V. Prediction of a missing protein expression map in the context of the human proteome project. J Proteome Res. 2015; 14(3):1350–60. [PubMed: 25612097]

29. Kim JW, Hwang H, Lim JS, Lee HJ, Jeong SK, Yoo JS, Paik YK. gFinder: a web-based bioinformatics tool for the analysis of N-glycopeptides. J Proteome Res. 2016; doi: 10.1021/acs.jproteome.6b00772

30. Panwar B, Menon R, Eksi R, Li HD, Omenn GS, Guan Y. Genome-Wide Functional Annotation of Human Protein-Coding Splice Variants Using Multiple Instance Learning. J Proteome Res. 2016; 15(6):1747–53. [PubMed: 27142340]

31. Van Eyk JE, Corrales FJ, Aebersold R, Cerciello F, Deutsch EW, Roncada P, Sanchez JC, Yamamoto T, Yang P, Zhang H. Highlights of the Biology and Disease-driven Human Proteome Project, 2015-2016. J Proteome Res. 2016; doi: 10.1021/acs.jproteome.6b00444

32. Lam MP, Venkatraman V, Xing Y, Lau E, Cao Q, Ng DC, Su AI, Ge J, Van Eyk JE, Ping P. Data-Driven Approach To Determine Popular Proteins for Targeted Proteomics Translation of Six Organ Systems. J Proteome Res. 2016; doi: 10.1021/acs.jproteome.6b00095