Journal of proteome research

Editorial

pubs.acs.org/jpr

Genome-wide Proteomics, Chromosome-centric Human Proteome Project (C-HPP), Part II

With vibrant scientific meetings and collaborative efforts among the research teams responsible for mapping and characterizing the proteins of individual chromosomes, the Chromosome-Centric Human Proteome Project (C-HPP)[1,2] consortium of the Human Proteome Project (HPP)[3] has entered its second productive year. In this Special Issue, we report substantial progress in networking, journal publications, information dissemination, and cooperative proteome analysis among chromosome teams. The annual Human Proteome Organization (HUPO) World Congress of Proteomics in Yokohama, Japan in September 2013 was a scientific showcase for the HPP, which is composed of two engines that include C-HPP and B/D-HPP and the resource pillars. The result was an ideal working mode for translational proteome biology, which embraced basic molecular dynamics from genome to proteome and comprehensive approaches to disease mechanisms from cancers to preeclampsia and neural disease (Figure 1). The baseline metrics of proteomic mapping in a chromosome scaffold[4] were updated[5] by reshaping the analysis of missing proteins and reinforcing the data sharing system through ProteomeXchange. We are grateful to all bioinformatics leaders such as Amos Bairoch and Lydie Lane at the Swiss Institute for Bioinformatics (neXtProt), Eric Deutsch at the Institute for Systems Biology (PeptideAtlas, SRM Atlas), Ron Beavis at the University of British Columbia (GPMDB), and Emma Lundberg at the Karolinska Institute (Human Protein Atlas) as well as the entire C-HPP bioinformatics unity group.

The C-HPP welcomed three new leaders of chromosome teams: Lydie Lane (SIB, Switzerland) for Chr. 2, Daniel Figeys (University of Ottawa, Canada) for Chr. 21, and Akhilesh Pandey (Johns Hopkins University, United States) for Chr. 22. Juan Pablo Albar (Chr. 16, Spain) succeeded Charles Lee (United States) on the C-HPP Executive Committee. Our team is charged with new energy and an excellent information pipeline (led by neXtProt). In our networking, C-HPP held scientific meetings in 2013 that have contributed to the global dissemination of this project, including the Proteomic Forum in Berlin, Germany (March 17–21), the 38th FEBS Congress in St. Petersburg, Russia (July 10–11), and several national proteomics meetings (KHUPO, USHUPO, CNHUPO, and Taiwanese HUPO). Throughout these meetings, the leaders of HPP and C-HPP explored key issues such as consolidation of transcriptomic data, production, submission, and sharing of proteomic data in chromosome-centric format, clinical interests, and technologies for protein identification and quantitation (e.g., multiple reaction monitoring (MRM) standardization) in the context of the C-HPP and HPP knowledge base.

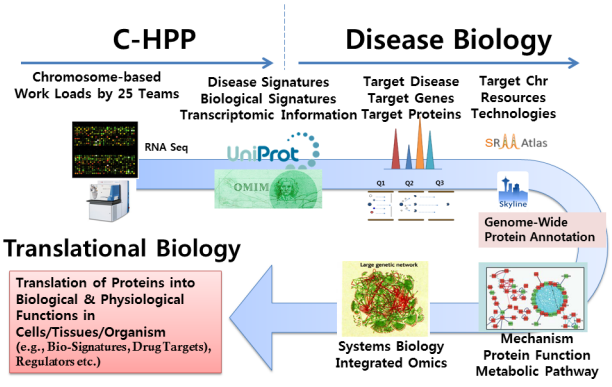

Figure 1. Future direction of the C-HPP through a synergistic collaboration with the disease biology group to produce a list of proteins with biological and physiological relevance.

PUBLICATIONS FROM THE C-HPP CONSORTIUM DURING THE FIRST YEAR

The first *Journal of Proteome Research* C-HPP Special Issue in 2013 presented major developments from the C-HPP. In the January 2013 issue, there were 22 papers from C-HPP teams on the following subjects: 13 articles from chromosome teams 1, 4, 7, 8, 11, 13, 16, 17, 18, 19, 20, X, and Y; database reports (including neXtProt, PeptideAtlas, and CAPER); technology papers; and cross-cutting articles (from the C-HPP and B/D-HPP leadership and the HUPO Industrial Advisory Board). There were 11 additional articles stimulated by the announcement of the Special Issue, which were related to the C-HPP. The virtual issue of the C-HPP published 15 papers in June 2013 on the following subjects: major papers on transcripts and immunohistochemistry from the Human Protein Atlas;[6] the placental proteome;[7] functional annotation of chromosome 7 missing proteins;[8] ERBB2 and EGFR pathways in inflammatory breast-cancer cell lines from chromosome 17;[9] MaxQuant analyses of 11 cell lines and colon cancer specimens;[10] Peppy1.0 proteogenomic analyses;[11] and the effect of alternative splicing on protein–protein interaction networks.[12] Therefore, there were 48 publications from or related to the C-HPP during the first year. These articles constitute the 2013 virtual C-HPP Special Issue (see www.thehpp.org and http://pubs.acs.org/page/jprobs/vi/c-hhp.html), which demonstrates the great scientific productivity of our consortium.

UPDATED GOALS AND DELIVERABLES FOR THE C-HPP

Because we have made significant progress on the consortium projects, it is timely to refresh and update our scientific goals in each phase.[1,2] For example, for phase I (9/2012–9/2018), we set goals to map the entire missing proteins. There is some concern that our previously scheduled deliverables and timetable might not reflect the rate of progress to date. Reassessing and updating the scientific goals, deliverables, and timetable will be a top-priority activity, along with other

Special Issue: Chromosome-centric Human Proteome Project

Published: December 16, 2013

   dx.doi.org/10.1021/pr4011958 | *J. Proteome Res.* 2014, 13, 1–4

important issues to be discussed in the upcoming C-HPP workshop in Busan, Korea on March 26, 2014. On the basis of the standard metrics updated by Lane et al.,[5] the number of total coding genes is now 19,490 and the neXtProt gold total of genes with protein expression level 1 evidence, 15,646. Constitutes 80% of the total and reflects significant progress from new large studies and from reanalyses of major databases. New findings published in this special issue will drive the 2014 update of the C-HPP metrics. This progress encourages us to reset the goals and scope of our timeline and will guide the continuing direction of the project. Considering the current speed of progress and advances in detection techniques, we are optimistic about shortening the period for complete identification of all missing proteins, the first major target of the project. Thus, we may have to reach a new consensus for more realistic target numbers for missing proteins and other targets (e.g., number of ASTs, PTMs, and SNP-derived protein characterization).

DATA SHARING

At the 2013 Berlin Workshop, we set out our policy on research freedom, mutual data sharing, and study guidelines for individual proteins of interest by consortium members, which stimulated in-depth research and functional analysis of specific proteins of interest involved in specific diseases or biological systems by individual chromosome teams. Although each team enjoys the freedom to study any protein encoded on any chromosome, we encourage each member to honor the spirit of mutual collaboration between teams (e.g., collaboration between the individual team and the team responsible for the chromosome encoding the protein of interest). This approach benefits both teams and the wider consortium.

Data submission to ProteomeXchange should be a direct barometer of success because each team will benefit from this obligatory data submission, sharing, and retrieval process. The accumulating data for individual chromosomes will facilitate the completion of protein mapping for all chromosomes (Figure 2). Led by Juan Antonio Vizcaino and his international colleagues, the ProteomeXchange site now meets the needs of investigators and connects all of the individual data resources, including PRIDE, PeptideAtlas, neXtProt, and GPMDB. As of November 22, 2013, there were 561 data sets uploaded to ProteomeXchange. It accepts raw data sets and metadata from all consortium teams.

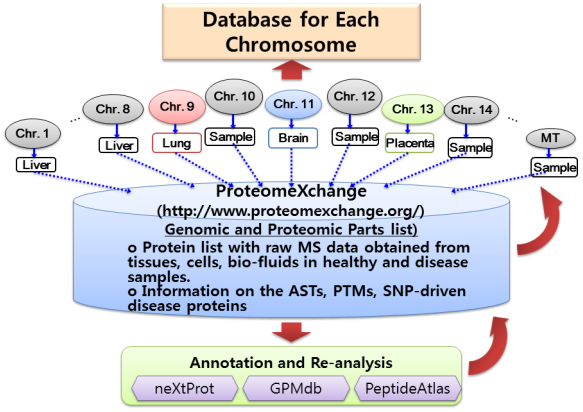

Figure 2. Data exchange and utilization through ProteomeXchange by the C-HPP consortium members.

RNA Seq should now become routine practice for all chromosome teams to improve data quality. On the basis of extensive results from the Chr. 17 and 7 teams[9,13] (see also Menon et al., this issue) and the CCPD Chr. 1, 8, and 20 teams (this issue), collaborations between the providers and investigators are already very productive and highly recommended. Regarding transcriptomic data, we have reached the point where we need to set out our strategy regarding utilization of ENCODE[14] through some interfaces between ENCODE and C-HPP. ENCODE has been featured at the 2012 and 2013 HUPO World Congresses. Once the latest ENCODE data sets are uploaded into ENSEMBL on a regular basis, we will start working on the proteomics to generate evidence of the activity of novel gene sequences and transcript analysis of long reads.[8,11] GPMDB has been updated in various aspects by Ron Beavis (e.g., the phosphorylation site information in human proteins, annotating ENSEMBL v. 70 protein sequences at ftp://ftp.proteomecentral.org/modifications/phosphoryl/). The C-HPP Wiki site, managed by Peter Horvatovich (Netherlands, Chr. 5 group), provides an excellent communication forum for all consortium members; they can report their progress and freely obtain important resources for their research (http://c-hpp.webhosting.rug.nl/tiki-index.php).

COLLABORATIONS OF C-HPP WITH B/D-HPP

It is necessary for the C-HPP to establish a firm collaboration with the B/D-HPP for the identification of missing proteins because this joint collaboration will enhance the synergistic cooperation between two consortia.[15] We believe that the ultimate goal of the HPP to construct a comprehensive knowledge base for proteome science that can be used in all biomedical fields. Given that the project of each chromosome team is coupled to studies of individual diseases (e.g., cancer, metabolic disease, male infertility, female preeclampsia, and auto-immune disorder), extensive biological studies along with protein mapping and characterization are required. Once our project identify missing proteins for each chromosome approaches a saturation point within 2 to 3 years, exploration of the functions of those newly identified missing proteins and determination of their status as disease regulators, markers, or drug targets, will require close collaboration with B/D-HPP teams within the context of translational biology (Figure 3). Complementary progress in the B/D-HPP component of the HPP was presented in 12 workshops and additional meetings at

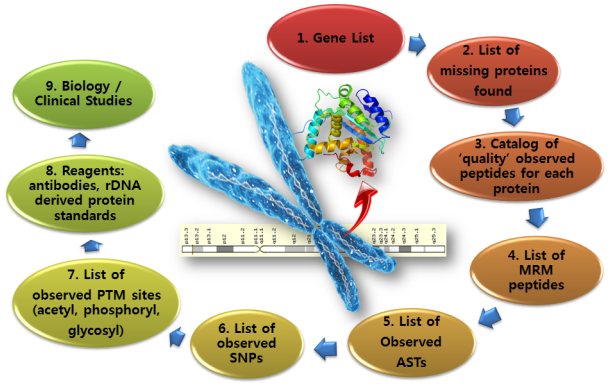

Figure 3. Potential deliverables from the synergistic collaboration between the C-HPP and B/D-HPP groups.

dx.doi.org/10.1021/pr4011958 | J. Proteome Res. 2014, 13, 1−4

the Yokohama Congress. The B/D-HPP is particularly focused on generating resources, reagents, spectral libraries, and databases that will facilitate widespread use of proteomics throughout the life sciences/biomedical community. Major achievements are the completion of the SRM Atlas, SRM spectral library, and SRM proteotypic peptide reagents for proteins from all of the 20 000 protein-coding genes. This work was led by Ruedi Aebersold at ETH-Zurich and Rob Moritz at Institute for Systems Biology.[16,17] The SRMAtlas and PASSEL, the database for submission of SRM/MRM data sets, can be accessed at www.thehpp.org and www.peptideatlas.org. The B/D-HPP, the mass spectrometry resource pillar of the HPP, and the HUPO Technology Committee created an initiative with instrument manufacturers to generate a Proteome Analyzer that would be robust, high-throughput, and moderate cost for use in clinical laboratories and epidemiological studies. A collaboration team with Chr. 2 and 14 has reportedly utilized PepPSy, a user-friendly peptide/protein prioritization system, to support initiatives linking C-HPP and B/D-HPP projects. According to their annual report (www.c-hpp.org), this tool should facilitate the arduous task of prioritizing protein candidates for dedicated follow-up studies within the context of a specific biological process. This will give us good motivation to assess and share some research progress across all C-HPP and B/D-HPP teams and their collaborations in biological studies, defined data sets, clinical specimens, and analytical methods. These joint efforts will achieve translational biology in practice (Figure 1), which delivers more products, tools, and resources for the discovery of key biological mechanisms involved in human diseases (Figure 3).

PROSPECTIVES OF THE 2014 SPECIAL ISSUE AND THE COMING YEAR

Given the outstanding achievements in the first year (Sept 10, 2012-Sept 9, 2013), it is exciting to anticipate what will be coming in the second year. In this 2014 Special Issue, we and our colleagues present approximately three dozen papers from individual chromosome teams, database developments, and multiple disease-oriented studies. For some examples, the Chinese consortium for chromosomes 1, 8, and 20 presents cogent information about the translatome, which is poly A-rich mRNA found in the nascent chain complexes of the first step of translation on polyribosomes. The results suggest that 3−7% of identified mRNAs are not actually translated in various specimens (Zhong et al. and Chang et al., this issue). They also identify clusters of functionally related genes and proteins, the family of 38 beta-defensins, of which 30 occur in clusters on chromosomes 6, 8, and 20. These genes and proteins may be activated (and detectable) only under stress states resulting from inflammation or microbial infections (Liu et al., this issue). The Brazilian chromosome 15 team presents interesting results for gastric cancers, including highly tissue-specific expression of gastrokine 2 (GKN2) (Aquino et al., this issue) and for contrasting regions of the brain, the anterior temporal lobe (gray matter), and the corpus callosum (white matter) (Martins-de-Souza et al., this issue). The chromosome 17 group reports extensive analyses with RNA-seq and proteomics of splice variants associated with specific pathways in three breast-cancer cell lines (Menon et al., this issue). This issue also contains wide coverage of the following topics related to C-HPP: lung cancer profiling on Chr. 9, Ahn et al., this issue; SNVs transmissions on Chr. 20 in liver-cancer cell lines, Wang Q. et al., this issue; RNA- and antibody-based profiling; and the state of the human proteome as viewed through a coorporative analysis of kidney, urine, and plasma PeptideAtlases (Farrah et al., this issue).

A new development in our C-HPP community is that Christoph Borchers, the Canadian Chr. 6 team has just initiated an MRM-standardization project as part of the C-HPP initiatives. This idea surfaced at the C-HPP scientific meeting (July 10−11) in St. Petersburg. This joint project between the C-HPP consortium and HUPO New Technologies and Resources Committee will be carried out in a medium-scale interlaboratory study, seeking to demonstrate that MRM for quantitative proteomics can be very reproducible across several types of instruments used in the studies, provided that standard operating procedures and internal standards are used.[18] We envision that this development will stimulate research on technology and application and create a standard for team work within the consortium.

CONCLUDING REMARKS

The C-HPP has an ambitious meetings schedule for 2014, starting with the ninth C-HPP workshop in Haewoondae Beach, Busan, Korea, on March 26, in conjunction with the KHUPO annual meeting, followed by the 10th C-HPP workshop in Bangkok, Thailand, on August 6, and the 11th C-HPP workshop in Madrid, Spain, during the 13th HUPO Congress, on October 5. The success of these meetings will depend on the participation of our members and the quality of our programs. Throughout these meetings and workshops, we will prepare some new strategies and reach consensus for several key items on the agenda. These include revising our goal for real-time updating of the missing proteins on the Wiki and c-hpp.org, efficient routine data submission and sharing, dissemination of our project results through publications and the web, reaching out to the wider scientific society with the deliverables e.g., a YouTube tutorial for major informatics resources (GPM DB, neXtProt, PeptideAtlas, ProteomeXchange), supply of reagents throughout the community, annual performance review systems (online), and stimulation of interchromosomal team collaboration. We are approaching the stage where more information derived from the study of cellular and biochemical processes can be compiled with detailed protein characterization.

It is anticipated that more teams will receive national funding in 2014. In the first year of the C-HPP consortium, we have observed visible funding seed money increases in several countries including China (Chr. 1, 8, 20), Iran (Chr. Y), Canada (Chr. 6), Australia (Chr. 7), Brazil (Chr. 15), Spain (Chr. 16), United States (Chr. 17), Russia (Chr. 18), and Sweden (Chr. 19). Several teams are investing efforts to obtain new grants for their projects. We will maintain our collaborative efforts to share information and resources regarding available grant applications and will continuously provide support for the global initiative. One year after establishing our Headquarters office at Yonsei Proteome Research Center, Yonsei University, Seoul, Korea, the Korean government (Ministry of Health and Welfare) decided to fund our HQ operations through a long-term grant. This grant money supports the operational costs of the HQ and the direct research costs of three Korean chromosome teams (Chr. 9, 11, 13). This grant support started on Dec. 1, 2013, will cover the first 5 years to Nov. 30, 2018 and is extendable for another 4 years to Nov. 30, 2022, matching the entire period of the consortium project. With this support, our consortium will move steadily forward to achieve

dx.doi.org/10.1021/pr4011958 | J. Proteome Res. 2014, 13, 1−4

its planned goals throughout the entire period of the project (2012−2022). Thus, we believe the C-HPP is now right on track!

Young-Ki Paik*

Yonsei Proteome Research Center, Departments of Integrated Omics for Biomedical Science and Biochemistry, Yonsei University, 50 Yonsei-ro, Sudaemoon-ku, Seoul 120-749, Korea

Gilbert S. Omenn

Center for Computational Medicine and Bioinformatics, University of Michigan, 2017 Palmer Commons Building, 100 Washtenaw Avenue, Ann Arbor, Michigan 48109, United States

Visith Thongboonkerd

Faculty of Medicine Siriraj Hospital, Mahidol University, 2 Prannok Road, Siriraj, Bangkoknoi, Bangkok 10700, Thailand

Gyorgy Marko-Varga

Institutionen för Elektrisk Mätteknik, Lund University, Ole Römers Väg 3, Lund 22100, Sweden

William S. Hancock*

The Barnett Institute of Chemical and Biological Analysis, Northeastern University, 140 The Fenway, Boston 02115, Massachusetts, United States

AUTHOR INFORMATION

Corresponding Authors

*Y.-K.P.: E-mail: paikyk@yonsei.ac.kr. Tel: 82-2-2123-4242. Fax: 82-2-393-6589.
*W.S.H.: E-mail: wi.hancock@neu.edu. Tel: 1-617-869-8458. Fax: 617-373-2855.

Notes

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

ACKNOWLEDGMENTS

We thank Hyoung-Joo Lee and Seul-Ki Jeong for their assistance on the graphic works. Part of this work was supported by a grant from the Korean Ministry of Health and Welfare (to Y.K.P., HI13C2098).

REFERENCES

(1) Paik, Y. K.; Jeong, S. K.; Omenn, G. S.; Uhlen, M.; Hanash, S.; Cho, S. Y.; Lee, H. J.; Na, K.; Choi, E. Y.; Yan, F.; Zhang, F.; Zhang, Y.; Snyder, M.; Cheng, Y.; Chen, R.; Marko-Varga, G.; Deutsch, E. W.; Kim, H.; Kwon, J. Y.; Aebersold, R.; Bairoch, A.; Taylor, A. D.; Kim, K. Y.; Lee, E. Y.; Hochstrasser, D.; Legrain, P.; Hancock, W. S. The Chromosome-Centric Human Proteome Project for cataloging proteins encoded in the genome. *Nat. Biotechnol.* **2012**, *30* (3), 221−223.

(2) Paik, Y. K.; Omenn, G. S.; Uhlen, M.; Hanash, S.; Marko-Varga, G.; Aebersold, R.; Bairoch, A.; Yamamoto, T.; Legrain, P.; Lee, H. J.; Na, K.; Jeong, S. K.; He, F.; Binz, P. A.; Nishimura, T.; Keown, P.; Baker, M. S.; Yoo, J. S.; Garin, J.; Archakov, A.; Bergeron, J.; Salekdeh, G. H.; Hancock, W. S. Standard guidelines for the chromosome-centric human proteome project. *J Proteome Res.* **2012**, *11* (4), 2005−2013.

(3) Legrain, P.; Aebersold, R.; Archakov, A.; Bairoch, A.; Bala, K.; Beretta, L.; Bergeron, J.; Borchers, C. H.; Corthals, G. L.; Costello, C. E.; Deutsch, E. W.; Domon, B.; Hancock, W.; He, F.; Hochstrasser, D.; Marko-Varga, G.; Salekdeh, G. H.; Sechi, S.; Snyder, M.; Srivastava, S.; Uhlén, M.; Wu, C. H.; Yamamoto, T.; Paik, Y. K.; Omenn, G. S. The human proteome project: current state and future direction. *Mol. Cell. Proteomics.* **2011**, *10* (7), M111.009993.

(4) Marko-Varga, G.; Omenn, G. S.; Paik, Y. K.; Hancock, W. S. A first step toward completion of a genome-wide characterization of the human proteome. *J. Proteome Res.* **2013**, *12* (1), 1−5.

(5) Lane, L. L.; Bairoch, A.; Beavis, R. C.; Deutsch, E. W.; Gaudet, P.; Lundberg, E.; Omenn, G. S. Metrics for the Human Proteome Project—2013−2014 and strategies for finding missing proteins. *J. Proteome Re.s* **2014**, DOI: 10.1021/pr401144x.

(6) Fagerberg, L.; Oksvold, P.; Skogs, M.; Algenäs, C.; Lundberg, E.; Pontén, F.; Sivertsson, A.; Odeberg, J.; Klevebring, D.; Kampf, C.; Asplund, A.; Sjöstedt, E.; Al-Khalili Szigyarto, C.; Edqvist, P. H.; Olsson, I.; Rydberg, U.; Hudson, P.; Ottosson Takanen, J.; Berling, H.; Björling, L.; Tegel, H.; Rockberg, J.; Nilsson, P.; Navani, S.; Jirström, K.; Mulder, J.; Schwenk, J. M.; Zwahlen, M.; Hober, S.; Forsberg, M.; von Feilitzen, K.; Uhlén, M. Contribution of Antibody-based Protein Profiling to the Human Chromosome-centric Proteome Project (C-HPP). *J. Proteome Res.* **2013**, *12* (6), 2439−2448.

(7) Lee, H. J.; Jeong, S. K.; Na, K.; Lee, M. J.; Lee, S. H.; Lim, J. S.; Cha, H. J.; Cho, J. Y.; Kwon, J. Y.; Kim, H.; Song, S. Y.; Yoo, J. S.; Park, Y. M.; Kim, H.; Hancock, W. S.; Paik, Y. K. Comprehensive genome-wide proteomic analysis of human placental tissue for the chromosome-centric human proteome project. *J. Proteome Res.* **2013**, *12* (6), 2458−2466.

(8) Ranganathan, S.; Khan, J. M.; Garg, G.; Baker, M. S. Functional annotation of the human chromosome 7 "missing" proteins: a bioinformatics approach. *J. Proteome Res.* **2013**, *12* (6), 2504−2510.

(9) Zhang, E. Y.; Cristofanilli, M.; Robertson, F.; Reuben, J. M.; Mu, Z.; Beavis, R. C.; Im, H.; Snyder, M.; Hofree, M.; Ideker, T.; Omenn, G. S.; Fanayan, S.; Jeong, S. K.; Paik, Y. K.; Zhang, A. F.; Wu, S. L.; Hancock, W. S. Genome Wide Proteomics of ERBB2 and EGFR and Other Oncogenic Pathways in Inflammatory Breast Cancer. *J. Proteome Res.* **2013**, *12* (6), 2805−2817.

(10) Neuhauser, N.; Nagaraj, N.; McHardy, P.; Zanivan, S.; Scheltema, R.; Cox, J.; Mann, M. High Performance Computational Analysis of Large-scale Proteome Data Sets to Assess Incremental Contribution to Coverage of the Human Genome. *J. Proteome Res.* **2013**, *12*, 2858−2868.

(11) Risk, B. A.; Spitzer, W. J.; Giddings, M. C. Peppy: proteogenomic search software. *J. Proteome Res.* **2013**, *12*, 3019−3025.

(12) Sinha, A.; Nagarajaram, H. A. Effect of Alternative Splicing on the Degree Centrality of Nodes in Protein-Protein Interaction Networks of Homo sapiens. *J. Proteome Res.* **2013**, *12*, 1980−1988.

(13) Liu, S.; Im, H.; Bairoch, A.; Cristofanilli, M.; Chen, R.; Deutsch, E. W.; Dalton, S.; Fenyo, D.; Fanayan, S.; Gates, C.; Gaudet, P.; Hincapie, M.; Hanash, S.; Kim, H.; Jeong, S. K.; Lundberg, E.; Mias, G.; Menon, R.; Mu, Z.; Nice, E.; Paik, Y. K.; Uhlen, M.; Wells, L.; Wu, S. L.; Yan, F.; Zhang, F.; Zhang, Y.; Snyder, M.; Omenn, G. S.; Beavis, R. C.; Hancock, W. S. A chromosome-centric human proteome project (C-HPP) to characterize the sets of proteins encoded in chromosome 17. *J. Proteome Res.* **2013**, *12* (1), 45−57.

(14) Paik, Y. K.; Hancock, W. S. Uniting ENCODE with genome-wide proteomics. *Nat. Biotechnol.* **2012**, *30* (11), 1065−1067.

(15) Omenn, G. S. The strategy, organization, and progress of the HUPO Human Proteome Project. *J. Proteomics* **2013**, DOI: 10.1016/j.jprot.2013.10.012.

(16) Picotti, P.; Aebersold, R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat. Methods* **2012**, *9* (6), 555−566.

(17) Farrah, T.; Deutsch, E. W.; Kreisberg, R.; Sun, Z.; Campbell, D. S.; Mendoza, L.; Kusebauch, U.; Brusniak, M. Y.; Hüttenhain, R.; Schiess, R.; Selevsek, N.; Aebersold, R.; Moritz, R. L. PASSEL: the PeptideAtlas SRM experiment library. *Proteomics* **2012**, *12* (8), 1170−1175.

(18) Percy, A. J.; Chambers, A. G.; Yang, J.; Jackson, A. M.; Domanski, D.; Burkhart, J.; Sickmann, A.; Borchers, C. H. Method and platform standardization in MRM-based quantitative plasma proteomics. *J. Proteomics* **2013**, *95*, 66−76.