Journal of proteome research

Communication

pubs.acs.org/jpr

Use of ENCODE Resources to Characterize Novel Proteoforms and Missing Proteins in the Human Proteome

Carol L. Nilsson,[†] Ekaterina Mostovenko,[†] Cheryl F. Lichti,[†] Kelly Ruggles,[&] David Fenyö,[&] Kate R. Rosenbloom,[§] William S. Hancock,[‖] Young-Ki Paik,[⊥] Gilbert S. Omenn,[#] Joshua LaBaer,[▽] Roger A. Kroes,[○] Matthias Uhlén,[◆] Sophia Hober,[¶] Ákos Végvári,[+] Per E. Andrén,[□] Erik P. Sulman,[●] Frederick F. Lang,[△] Manuel Fuentes,[▲] Elisabet Carlsohn,[▼] Mark R. Emmett,[‡] Joseph R. Moskal,[○] Frode S. Berven,[⬢] Thomas E. Fehniger,[+] and György Marko-Varga*[,+]

[†]Department of Pharmacology and Toxicology and [‡]Biochemistry and Molecular Biology, UTMB Cancer Center, University of Texas Medical Branch, Galveston, Texas 77555, United States

[&]Center for Biomolecular Science and Engineering, School of Engineering, University of California Santa Cruz (UCSC), 1156 High Street, New York, Santa Cruz California 95064, United States

[§]Center for Biomolecular Science and Engineering, University of California, 1156 High St, Mail Stop CBSE, Santa Cruz, California 95064, United States

[‖]College of Science, Northeastern University, 360 Huntington Avenue, Boston, Massachusetts 02115, United States

[⊥]Department of Biochemistry, Yonsei Proteome Research Center, 50 Yonsei-Ro, Seodaemun-gu, Seoul 120-749, South Korea

[#]Center for Computational Medicine and Bioinformatics, University of Michican Medical School, 100 Washtenaw Avenue, Ann Arbor, Michigan 48109, United States

[▽]Biodesign Institute, Arizona State University, 1001 South McAllister Avenue, Tempe, Arizona 85287, United States

[○]The Falk Center for Molecular Therapeutics, McCormick School of Engineering and Applied Sciences, Northwestern University, 1801 Maple Street, Evanston, Illinois 60201, United States

[◆]Biotechnology, AlbaNova University Center, Royal Institute of Technology, Roslagstullsbacken 21, 106 91 Stockholm, Sweden

[¶]School of Biotechnology, Department of Proteomics, Royal Institute of Technology, 106 91 Stockholm, Sweden

[+]Clinical Protein Science & Imaging, Biomedical Center, Department of Biomedical Engineering, Lund University, 221 84 Lund, Sweden

[□]Department of Pharmaceutical Biosciences, Uppsala University, Husargatan 3, 752 37 Uppsala, Sweden

[●]Department of Radiation Oncology, The University of Texas M.D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, Texas 77030, United States

[△]Department of Neurosurgery, The University of Texas M.D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, Texas 77030, United States

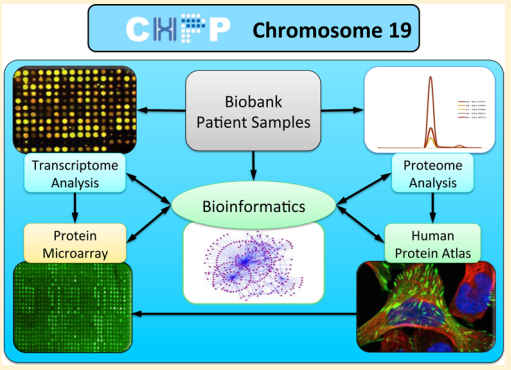
[▲]Centro de Investigacion del Cancer, Medicine-Immunology, CSIC-University of Salamanca, Salamanca 37007, Spain

[▼]Proteomics Core Facility, Sahlgrenska Academy, University of Gothenburg, Medicinaregatan 7A, Gothenburg 413 90, Sweden

[⬢]Department of Biomedicine, University of Bergen, Jonas Lies vei 91, 5009 Bergen, Norway

Supporting Information

ABSTRACT: We describe the utility of integrated strategies that employ both translation of ENCODE data and major proteomic technology pillars to improve the identification of the "missing proteins", novel proteoforms, and PTMs. On one hand, databases in combination with bioinformatic tools are efficiently utilized to establish microarray-based transcript analysis and supply rapid protein identifications in clinical samples. On the other hand, sequence libraries are the foundation of targeted protein identification and quantification using mass spectrometric and immunoaffinity techniques. The results from combining proteoENCODEdb searches with experimental mass spectral data indicate that some alternative splicing forms detected at the transcript level are in fact translated to proteins. Our results provide a step toward the directives of the C-HPP initiative and related biomedical research.



KEYWORDS: *Chromosome-centric Human Protein Project, ENCODE, glioma stem cell, protein sequence mass spectrometry, microassays, missing proteins*

Received: June 8, 2014

dx.doi.org/10.1021/pr500564q | J. Proteome Res. XXXX, XXX, XXX−XXX

The mission of the Chromosome-Centric Human Proteome Project (C-HPP) is to catalogue proteins expressed by ca. 20 300 human genes including the uncharacterized products for known protein-coding genes.[1] It has been estimated at present that so-called "missing proteins" account for roughly 20% of total coding proteins.[2] According to the recently published first drafts of the human proteome, the proteins identified at the expression level represent 84% (17 294 genes[3]) and 92% (19 629 genes[4]) of all of the human genes, as annotated in SwissProt. A closer look at the ProteomicsDB (www.proteomicsdb.org) indicated 1352 chromosome 19 genes and 1304 identified proteins (96.4%), referring to 3911 entries originating from both SwissProt (2301 entries) and TrEMBL (1610 entries).[4] The official C-HPP/HPP metric for "missing or inadequately documented proteins" is 3844 of a total of 19 490 protein-coding genes (identification levels 1, 2, 3, and 4 in neXtProt) according to Lane et al.[2] Therefore, it is vital to the success of this project to integrate and share data derived from proteomic studies using a variety of cell lines, tissues, and biofluids to produce a complete and definitive human proteome parts list that includes all protein forms. Furthermore, results from the ENCODE Consortium[1,5] point to the existence of as yet unknown proteins listed as uncertain, predicted, or homology-derived in the human proteome, necessitating utilization of ENCODE data for this endeavor.[1] Thus, with the high priority given to the identification of missing proteins and proteoforms, the combination of transcriptomic resources, ENCODE,[6] and proteomic technology pillars (mass spectrometry, antibodies, and bioinformatics) enables a powerful strategy to explore the expression patterns, post-translational modifications (PTMs), and functions of missing and known proteins by an integrated technology platform approach. (See Figure 1A.)

The Chromosome 19 Team has defined an improved strategy to mine the human proteome that utilizes the combined resources of the Human Genome Project, ENCODE, neXtProt, and well-curated biobanks[7−9] to leverage the discovery and

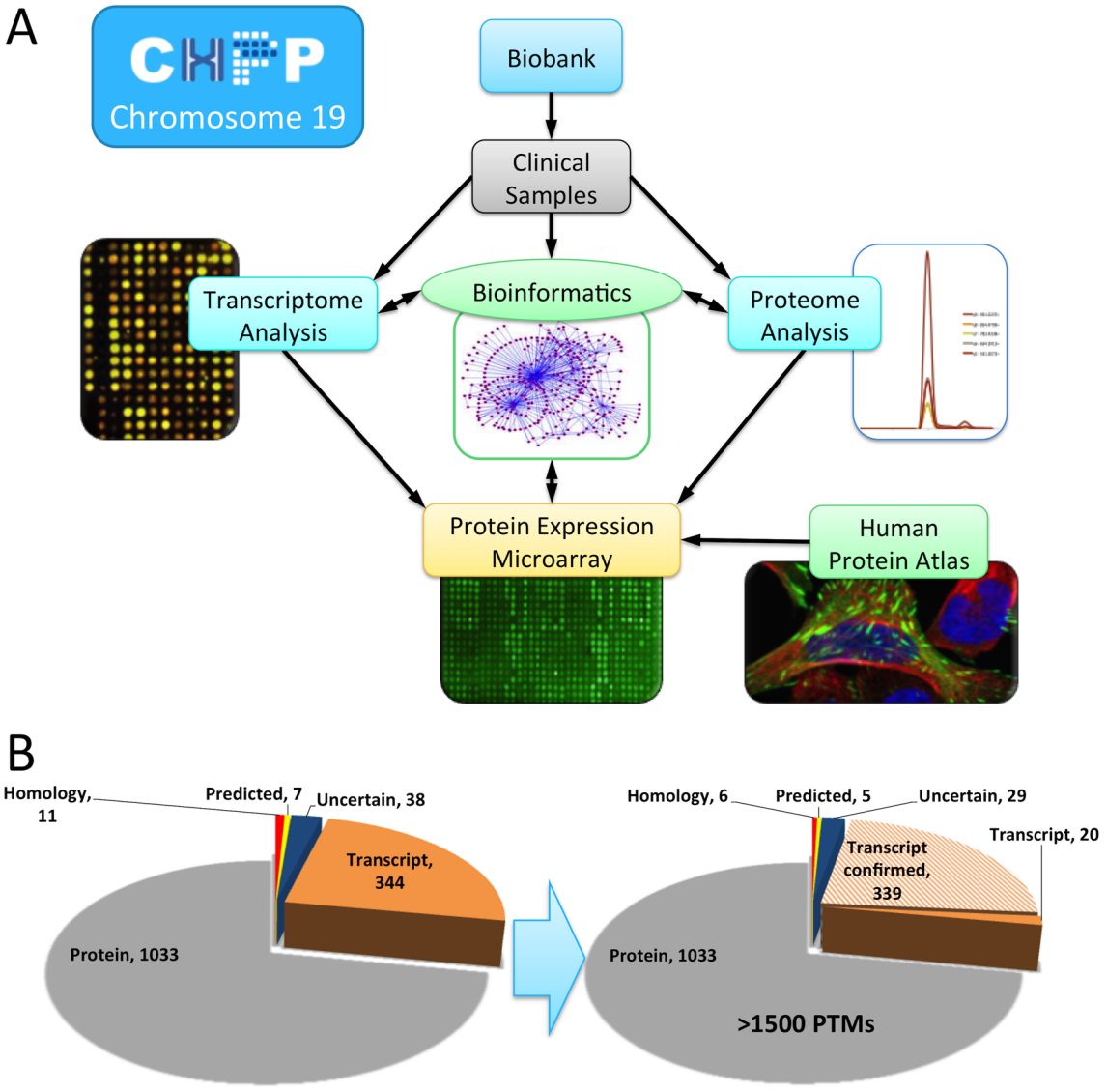


Figure 1. Schematic of the integrated strategies for identification of missing proteins and novel proteoforms. (A) Chromosome 19 strategy includes the integration of gene activity and proteomic data, chip-based assays for rapid screening of targets across many samples, and tissue imaging. (B) In comparison with the latest neXtProt data (left-side panel), the application of the chromosome 19 strategy on 36 GSCs has resulted in 15 new transcript level identification along with additional 339 confirmations at transcript level and more than 1500 PTM identifications (right-side panel).

dx.doi.org/10.1021/pr500564q | J. Proteome Res. XXXX, XXX, XXX−XXX

identification of missing proteins and novel proteoforms.[10,11] Our workflow uses new orthogonal bioinformatic strategies to mine proteomic data with custom databases as well as protein microarrays and selected reaction monitoring (SRM) quantification and verification of novel proteins. Our goal is to provide new knowledge about the molecular constellations of gene and protein expression in relationship to chromosome specific location, genome-wide interactions, and further definition to the clinical metrics used to diagnose and evaluate treatment efficacy. Those data would be of great importance to deliver future drug targets and protein diagnostics.

At the cellular level, the germline DNA that defines the entire human genome is only fractionally expressed in any of the 210 differentiated human cell types. Examples include the silencing of germ cell genes during cell differentiation by epigenetic mechanisms (histone conformation, methylation), DNA elimination (human erythroblasts), tissue-specific gene expression (the keratin protein family exclusive to different epithelial types), and various loss-of-function mutations and variants.[12] The study of chromosome-centric protein expression is thus complicated by our incomplete understanding of which genetic loci are commonly maintained in all or any one cell type. We have chosen in this study to investigate in detail a glioma stem cell (GSC) model, in which detailed gene, mRNA, and protein expression data (available at http://proteomeXchange.org: PXD000563) have been annotated.[11] GSCs are a multipotent subset of migratory cells derived from the most common adult primary brain tumor, glioblastoma multiforme. GSCs are an important clinical focus because of their invasiveness and radioresistance.

In our first application of the integrated strategy, gene activity in 36 GSC lines was measured through untargeted deep sequencing and proteins identified in the same cell lines.[11] Furthermore, we assigned approximately 1500 sites of PTMs in the untargeted, nonenriched proteomic data set (Figure 1B). All of the detailed data of the post-translationally modified peptides identified in 36 GSC lines are presented ase Supporting Information , clarifying the type of PTMs observed. Within this large-scale study, we developed a targeted transcriptomic approach, which allows us to rapidly measure chromosome 19 gene activity with high specificity.[10] Through the targeted transcriptomic approach, we measured differential regulation of roughly 200 chromosome 19 genes between subtypes of GSCs. By comparing the targeted transcriptomic data to the level of evidence in neXtProt (version 2013_12), we identified 15 proteins classified as uncertain or predicted. These 15 missing proteins are now a top priority to target at the protein level. Furthermore, 339 transcript entries, listed as transcript only in neXtProt, will be targeted in a similar manner. We identified 41 gene loci-encoding zinc finger protein transcripts and all 8 of the olfactory receptor genes. The zinc finger proteins play important roles in cancer pathophysiology, whereas the olfactory receptor proteins are mostly uninvestigated. It is of direct interest to determine whether any of these transcripts actually encode translated, functional proteins. This set of "missing proteins" is being highly scrutinized to be quantified at the protein level in all cell lines by SRM.

With the recent integration of the ENCODE data into our workflow, we are positioned to identify both unknown human proteins and new proteoforms at an accelerated pace. We have developed the expertise to build custom protein databases by translating RNA-Seq data, generate nonsynonymous single nucleotide polymorphism (nsSNP) databases, and create spectral libraries from large-scale mass spectrometric data sets derived from the GSCs. By using the databases that contain predicted sequences of known and novel proteoforms, we have identified nsSNPs translated in proteins as single amino acid variants (SAVs), derived from GSCs that may be associated with metabolic phenotypes and invasiveness. Bioinformatic searches of ENCODE data translated into a searchable database for proteomics (proteoENCODEdb) yielded 80 previously unpredicted proteins, three from chromosome 19; one was the result of a fusion between a known chromosome 19 exon and an unknown exon, and two were the result of fusions between two unknown exons. In addition, novel fusion proteins derived from chromosomal (somatic) rearrangements and novel alternative splice forms derived from known and previously unknown open reading frames (ORFs) are also being identified. For example, we have identified a SAV in a protein not previously associated with glioma, branched chain aminotransferase 2 (BCAT2). A newly identified SAV in Xaa-Pro dipeptidase (PEPD), a secreted protease, will be studied to determine how the proteoforms may affect modulation of extracellular matrix. The integration of RNA-Seq and proteomic data has allowed us to study the somatic-proteomic landscape of GSCs, thereby contributing new knowledge of novel fusion proteins in GSC pathobiology.

The identification of novel proteins and proteoforms in GSCs will allow further studies of their role in pathogenesis, tumor recurrence, and resistance to chemotherapy and radiation. The next step is to quantify their expression by SRM across 36 GSC lines under baseline conditions and following standard-of-care treatments. Furthermore, their expression in glioma tumors will be quantified.

Cell-free protein arrays, such as nucleic-acid-programmable protein arrays (NAPPA),[13] were customized for our studies. The NAPPA arrays were produced by printing cDNAs encoding ORFs with a tag at C-termini, followed by in situ protein translation at the time of the functional experiment using human in vitro transcription−translation systems (IVTT) (Figure 2A, left). The validity of these arrays as a tool for functional proteomics is already well-established. Protein microarrays allow detection of proteins in a high-density array, yielding a large amount of data from a single sample. This characteristic is particularly useful in the identification of low abundance proteins in complex samples (i.e., clinical specimens).

Identified ORFs are cloned and subcloned as cDNAs, then captured into a recombinational cloning platform to allow rapid transfer to a wide variety of experimental vectors and sharing with the research community through DNASU (http://DNASU.org). Our approach avoids reliance on prediction algorithms, costly peptide synthesis, or the purchase of expensive recombinant proteins. Instead, full-length proteins are rapidly produced as in-vitro-synthesized samples, which can be followed by tryptic digestion and mass spectrometric analysis to develop SRM assays.[10]

In addition, the antigen arrays, such as protein epitope signature tags (PrESTs), and the antibody arrays originating from the Human Protein Atlas project are compatible with NAPPA arrays and prove to be powerful technologies in disease proteomics. The combination of these multiarray platforms allows a novel comprehensive approach in identification of signature protein profiles in clinical proteomics, useful for biomarker discovery and diagnosis.[14]

The PrESTs can be produced as heavy-isotope-labeled sequences and used for quantification of targeted proteins.[15]

dx.doi.org/10.1021/pr500564q | J. Proteome Res. XXXX, XXX, XXX−XXX

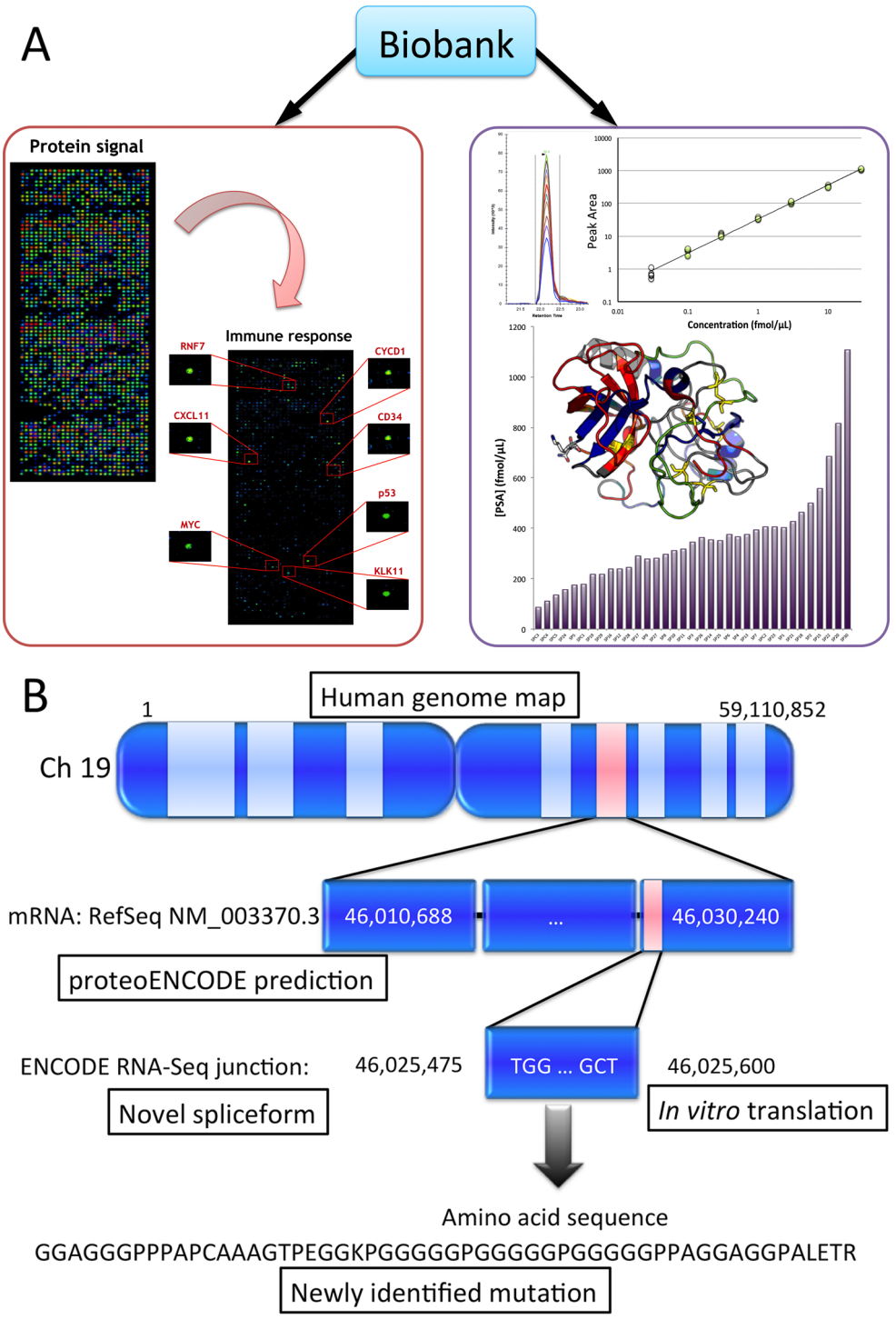


Figure 2. Use of an integrated strategy yields the identities of nsSNPs and novel proteoforms. (A) Self-assembled protein arrays containing 6000 human proteins (of which 100 encoded in chromosome 19) were incubated with serum from colorectal cancer patient, identifying a tumor specific antigen profile and tested to measure protein expressions (left-side panel). Quantitative SRM assay applied to biobank samples with improved quantitations made on prostate specific antigen (PSA) that took account also for the novel mutant (*L132I*) proteoform (right-side panel). (B) ENCODE data provides information on the exon−exon splice junctions and is described as a set of coordinates for start and end position on the genome. The region on chromosome 19 corresponds to a junction (46 025 475−46 025 600 nucleotides), located within known gene transcript (RefSeq NM_003370.3). A sequence in ENCODE RNA-Seq junction file defines a set of coordinates that corresponds to a protein sequence identified by MS in GSCs.

The use of another chromosome 19 resource, a peptide library representing the entire human proteome (constituting 100 000 peptide sequences), will aid in the validation of protein identities discovered by in vitro translation.[16] An example is given (Figure 2A, right), in which p.L132I, a novel proteoform of prostate-specific antigen (PSA), was included in an improved quantitation of the protein in samples from a biobank.

dx.doi.org/10.1021/pr500564q | *J. Proteome Res.* XXXX, XXX, XXX−XXX

On the basis of the sequences of transcripts, we design reagents for targeted SRM and protein chips to screen biological/clinical samples and quantify expression of novel proteins in a high-throughput manner. Furthermore, the increasing availability of high-quality, validated antibodies from the Human Protein Atlas[17] allows us to create custom reversed protein arrays, enable sensitive quantification with mass spectrometry, and probe protein expression in tissues by immunohistochemistry. The chimeric fusion proteins identified in the proteoENCODEdb will be given high priority for antibody development to accelerate our understanding of their biological roles. Finally, we have the capability to map proteins and peptides within histological compartments by use of mass spectrometric imaging (MSI) and confocal microscopy.

In conclusion, our newly developed multilayered strategies to refine the identification of missing proteins, along with proteoENCODEdb searches with experimental mass spectral data, demonstrate that novel spliceforms detected at the level of transcript are in fact translated to proteins. As an example of Figure 1 workflow we present a subsequent process whereby an initial annotation on the chromosome 19 mRNA microarray chip is followed by RNA-Seq confirmation, leading to the protein sequence prediction by proteoENCODEdb and the subsequent verification by MS/MS sequence of the translated proteins (Figure 2B).

The unidentified proteins are likely important candidate targets for pharmaceutical intervention into disease processes as well as important biomarkers of the more than 5000 known human diseases. In addition, the high number of newly identified PTM sites opens up an opportunity to align protein modifications with functional links to disease onset and progression. It is encouraged that the development of new tools exemplified by the activities of the Chromosome 19 Team will help others to further complete the process of defining the entire human proteome.

■ ASSOCIATED CONTENT

Supporting Information

Table of post-translationally modified peptides identified in 36 GSC lines, as identified by PEAKS DB and Mascot (phosphorylation, methionine oxidation) and PEAKS DB (all other PTMs). Mascot and PEAKS searches were combined using inChorus; scores represent peptide probabilities as calculated in PEAKS using the Peptide Prophet algorithm. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Tel: +46-46-222 3402. Fax: +46-46-222 4521. E-mail: Gyorgy.Marko-Varga@bme.lth.se.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We kindly acknowledge the support from Berta Kamprad Foundation, The Swedish Academy of Pharmaceutical Sciences, a grant from the Korean Ministry of Health and Welfare (to Y.-K.P., HI13C2098), and a grant from Instituto Salud Carlos III-Fondos FEDER (ISCIII-PI11/02114) to M.F. The Proteomics Unit at CIC belongs to ProteoRed, PRB2-ISCIII, supported by grant PT13/0001, the Cancer Prevention Research Institute of Texas (to C.L.N., R1122), the University of Texas Medical Branch (C.L.N. and M.R.E., and The Ralph and Marian Falk Medical Research Trust Chicago, IL (to J.R.M.).

■ 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) Lane, 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 Res.* **2014**, *13* (1), 15−20.

(3) Kim, M. S.; Pinto, S. M.; Getnet, D.; Nirujogi, R. S.; Manda, S. S.; Chaerkady, R.; Madugundu, A. K.; Kelkar, D. S.; Isserlin, R.; Jain, S.; Thomas, J. K.; Muthusamy, B.; Leal-Rojas, P.; Kumar, P.; Sahasrabuddhe, N. A.; Balakrishnan, L.; Advani, J.; George, B.; Renuse, S.; Selvan, L. D.; Patil, A. H.; Nanjappa, V.; Radhakrishnan, A.; Prasad, S.; Subbannayya, T.; Raju, R.; Kumar, M.; Sreenivasamurthy, S. K.; Marimuthu, A.; Sathe, G. J.; Chavan, S.; Datta, K. K.; Subbannayya, Y.; Sahu, A.; Yelamanchi, S. D.; Jayaram, S.; Rajagopalan, P.; Sharma, J.; Murthy, K. R.; Syed, N.; Goel, R.; Khan, A. A.; Ahmad, S.; Dey, G.; Mudgal, K.; Chatterjee, A.; Huang, T. C.; Zhong, J.; Wu, X.; Shaw, P. G.; Freed, D.; Zahari, M. S.; Mukherjee, K. K.; Shankar, S.; Mahadevan, A.; Lam, H.; Mitchell, C. J.; Shankar, S. K.; Satishchandra, P.; Schroeder, J. T.; Sirdeshmukh, R.; Maitra, A.; Leach, S. D.; Drake, C. G.; Halushka, M. K.; Prasad, T. S.; Hruban, R. H.; Kerr, C. L.; Bader, G. D.; Iacobuzio-Donahue, C. A.; Gowda, H.; Pandey, A. A draft map of the human proteome. *Nature* **2014**, *509* (7502), 575−581.

(4) Wilhelm, M.; Schlegl, J.; Hahne, H.; Moghaddas Gholami, A.; Lieberenz, M.; Savitski, M. M.; Ziegler, E.; Butzmann, L.; Gessulat, S.; Marx, H.; Mathieson, T.; Lemeer, S.; Schnatbaum, K.; Reimer, U.; Wenschuh, H.; Mollenhauer, M.; Slotta-Huspenina, J.; Boese, J. H.; Bantscheff, M.; Gerstmair, A.; Faerber, F.; Kuster, B. Mass-spectrometry-based draft of the human proteome. *Nature* **2014**, *509* (7502), 582−587.

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

(6) Khatun, J.; Yu, Y.; Wrobel, J. A.; Risk, B. A.; Gunawardena, H. P.; Secrest, A.; Spitzer, W. J.; Xie, L.; Wang, L.; Chen, X.; Giddings, M. C. Whole human genome proteogenomic mapping for ENCODE cell line data: identifying protein-coding regions. *BMC Genomics* **2013**, *14*, 141.

(7) Malm, J.; Danmyr, P.; Nilsson, R.; Appelqvist, R.; Végvári, Á.; Marko-Varga, G. Blood Plasma Reference Material: A Global Resource for Proteomic Research. *J. Proteome Res.* **2013**, *12* (7), 3087−3092.

(8) Malm, J.; Fehniger, T. E.; Danmyr, P.; Végvári, Á.; Welinder, C.; Lindberg, H.; Appelqvist, R.; Sjödin, K.; Wieslander, E.; Laurell, T.; Hober, S.; Berven, F.; Fenyö, D.; Wang, X. D.; Andrén, P. E.; Edula, G.; Carlsohn, E.; Fuentes, M.; Nilsson, C. L.; Dahlbäck, M.; Rezeli, M.; Erlinge, D.; Marko-Varga, G. Developments in Biobanking Workflow Standardization Providing Sample Integrity and Stability. *J. Proteomics* **2013**, *95*, 38−45.

(9) Malm, J.; Fehniger, T. E.; Danmyr, P.; Végvári, Á.; Welinder, C.; Lindberg, H.; Upton, P.; Carter, S.; Appelqvist, R.; Sjödin, K.; Wieslander, E.; Dahlbäck, M.; Rezeli, M.; Erlinge, D.; Marko-Varga, G. Standardization Developments for Large Scale Biobanks in Respiratory Diseases − A Model System for Blood Sample Processing and Storage. *Transl. Resp. Med.* **2013**, *1* (14), 10.1186/2213-0802-1-14.

(10) Nilsson, C. L.; Berven, F.; Selheim, F.; Liu, H.; Moskal, J. R.; Kroes, R. A.; Sulman, E. P.; Conrad, C. A.; Lang, F. F.; Andrén, P. E.; Nilsson, A.; Carlsohn, E.; Lilja, H.; Malm, J.; Fenyö, D.; Subramaniyam, D.; Wang, X.; Gonzales-Gonzales, M.; Dasilva, N.;

Diez, P.; Fuentes, M.; Végvári, Á.; Sjödin, K.; Welinder, C.; Laurell, T.; Fehniger, T. E.; Lindberg, H.; Rezeli, M.; Edula, G.; Hober, S.; Marko-Varga, G. Chromosome 19 Annotations with Disease Speciation: A First Report from the Global Research Consortium. *J. Proteome Res.* **2013**, *12* (1), 135−150.

(11) Lichti, C. F.; Liu, H.; Shavkunov, A. S.; Mostovenko, E.; Sulman, E. P.; Ezhilarasan, R.; Wang, Q.; Kroes, R. A.; Moskal, J. R.; Fenyö, D.; Oksuz, B. A.; Conrad, C. A.; Lang, F. F.; Berven, F. S.; Végvári, Á.; Rezeli, M.; Marko-Varga, G.; Hober, S.; Nilsson, C. L. Integrated Chromosome 19 Transcriptomic and Proteomic Data Sets Derived from Glioma Cancer Stem-Cell Lines. *J. Proteome Res.* **2014**, *13* (1), 191−199.

(12) MacArthur, D. G.; Balasubramanian, S.; Frankish, A.; Huang, N.; Morris, J.; Walter, K.; Jostins, L.; Habegger, L.; Pickrell, J. K.; Montgomery, S. B.; Albers, C. A.; Zhang, Z. D.; Conrad, D. F.; Lunter, G.; Zheng, H.; Ayub, Q.; DePristo, M. A.; Banks, E.; Hu, M.; Handsaker, R. E.; Rosenfeld, J. A.; Fromer, M.; Jin, M.; Mu, X. J.; Khurana, E.; Ye, K.; Kay, M.; Saunders, G. I.; Suner, M. M.; Hunt, T.; Barnes, I. H.; Amid, C.; Carvalho-Silva, D. R.; Bignell, A. H.; Snow, C.; Yngvadottir, B.; Bumpstead, S.; Cooper, D. N.; Xue, Y.; Romero, I. G.; Wang, J.; Li, Y.; Gibbs, R. A.; McCarroll, S. A.; Dermitzakis, E. T.; Pritchard, J. K.; Barrett, J. C.; Harrow, J.; Hurles, M. E.; Gerstein, M. B.; Tyler-Smith, C. A systematic survey of loss-of-function variants in human protein-coding genes. *Science (New York, N.Y.)* **2012**, *335* (6070), 823−828.

(13) Ramachandran, N.; Raphael, J. V.; Hainsworth, E.; Demirkan, G.; Fuentes, M. G.; Rolfs, A.; Hu, Y.; LaBaer, J. Next-generation high-density self-assembling functional protein arrays. *Nat. Methods* **2008**, *5* (6), 535−538.

(14) Henjes, F.; Lourido, L. M.; Ruiz-Romero, C.; Fernandez-Tajes, J.; Schwenk, J. M.; Gonzalez-Gonzalez, M.; Blanco, F. J.; Nilsson, P.; Fuentes, M. Analysis of autoantibody profiles in osteoarthritis using comprehensive protein array concepts. *J. Proteome Res.* **2014**, DOI: 10.1021/pr500775a.

(15) Zeiler, M.; Straube, W. L.; Lundberg, E.; Uhlen, M.; Mann, M. A Protein Epitope Signature Tag (PrEST) library allows SILAC-based absolute quantification and multiplexed determination of protein copy numbers in cell lines. *Mol. Cell. Proteomics* **2012**, *11* (3), O111.009613.

(16) Anderson, N. L. Libraries of specific assays covering whole proteomes: from yeast to man. *Clin. Chem.* **2010**, *56* (10), 1521−1522.

(17) Fagerberg, L.; Oksvold, P.; Skogs, M.; Algenas, C.; Lundberg, E.; Ponten, F.; Sivertsson, A.; Odeberg, J.; Klevebring, D.; Kampf, C.; Asplund, A.; Sjostedt, E.; Al-Khalili Szigyarto, C.; Edqvist, P. H.; Olsson, I.; Rydberg, U.; Hudson, P.; Ottosson Takanen, J.; Berling, H.; Bjorling, L.; Tegel, H.; Rockberg, J.; Nilsson, P.; Navani, S.; Jirstrom, K.; Mulder, J.; Schwenk, J. M.; Zwahlen, M.; Hober, S.; Forsberg, M.; von Feilitzen, K.; Uhlen, M. Contribution of antibody-based protein profiling to the human Chromosome-centric Proteome Project (C-HPP). *J. Proteome Res.* **2013**, *12* (6), 2439−2448.