



HAL
open science

Human Proteinpedia enables sharing of human protein data

Suresh Mathivanan, Mukhtar Ahmed, Natalie Ahn, Hainard Alexandre, Ramars Amanchy, Philip Andrews, Joel Bader, Brian Balgley, Marcus Bantscheff, Keiryn Bennett, et al.

► **To cite this version:**

Suresh Mathivanan, Mukhtar Ahmed, Natalie Ahn, Hainard Alexandre, Ramars Amanchy, et al.. Human Proteinpedia enables sharing of human protein data. *Nature Biotechnology*, 2008, 26 (2), pp.164-167. 10.1038/nbt0208-164 . hal-02883742

HAL Id: hal-02883742

<https://hal.inrae.fr/hal-02883742v1>

Submitted on 29 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

viewed directly in the MMCD interface or downloaded as a tab-delimited file for viewing with spreadsheet software.

We have also built a 'Miscellanea' search engine that allows users to filter results by the biological species, the type of database to be searched or other criteria. These options allow users to rapidly locate particular sets of records or limit their queries to a preferred subset of records.

In summary, MMCD is a practical tool for expediting the time-consuming steps of identifying and researching small molecules. This freely available resource is compatible with both NMR and MS data (singly or in combination) and facilitates high-throughput metabolomics investigations. Ongoing MMCD support is provided by the National Magnetic Resonance Facility at Madison. Users are encouraged to submit data to the BMRB (supported by the National Library of Medicine, Bethesda, MD, USA), which maintains one of the growing data archives that the MMCD relies upon¹⁰.

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health grants R21 DK070297 (MRS, PI) and P41

RR02301 (J.L.M.). We are indebted to W.W. Cleland, H. Lardy and L. Anderson for access to their chemical compound collections. We thank Joe Porwall from Sigma-Aldrich for assistance in locating compounds, Zsolt Zolnai for development of the Sesame LIMS and Eldon L. Ulrich for help in coordinating MMC activities with the BMRB.

Qiu Cui, Ian A Lewis, Adrian D Hegeman, Mark E Anderson, Jing Li, Christopher F Schulte, William M Westler, Hamid R Eghbalnia, Michael R Sussman, & John L Markley

Department of Biochemistry, University of Wisconsin-Madison, 433 Babcock Drive, Madison, Wisconsin 53706, USA.
e-mail: markley@nmrfam.wisc.edu

- Mendes, P. *Brief Bioinform.* **3**, 134–145 (2002).
- Nicholson, J.K., Lindon, J.C. & Holmes, E. *Xenobiotica* **29**, 1181–1189 (1999).
- Pauling, L., Robinson, A.B., Teranishi, R. & Cary, P. *Proc. Natl Acad. Sci. USA* **68**, 2374–2376 (1971).
- Viant, M.R., Rosenblum, E.S. & Tieerdema, R.S. *Environ. Sci Technol.* **37**, 4982–4989 (2003).
- Dettmer, K., Aronov, P.A. & Hammock, B.D. *Mass Spectrom. Rev.* **26**, 51–78 (2006).
- Markley, J.L. *et al. Pac. Symp. Biocomput.* **12**, 157–168 (2007).
- Wishart, D.S. *et al. Nucleic Acids Res.* **35**, D521–D526 (2007).
- Weininger, D., Weininger, A. & Weininger, J.L. *J. Chem. Inf. Comput. Sci.* **29**, 97–101 (1989).
- Lewis, I.A. *et al. Anal. Chem.* **79**, 9385–9390 (2007).
- Ulrich, E.L. *et al. Nucl. Acids Res.* **36**, D402–D408 (2008).

using modified protein DAS (distributed annotation system) protocols developed by us (DAS protocols were originally developed for sharing mRNA and DNA data) permit contributing laboratories to maintain protein annotations locally. All protein annotations are visualized in the context of corresponding proteins in the Human Protein Reference Database (HPRD)³. **Figure 1** shows tissue expression data for alpha-2-HS glycoprotein derived from three different types of experiments.

Our unique effort differs significantly from existing repositories, such as PeptideAtlas⁴ and PRIDE⁵ in several respects. First, most proteomic repositories are restricted to one or two experimental platforms, whereas Human Proteinpedia can accommodate data from diverse platforms, including yeast two-hybrid screens, MS, peptide/protein arrays, immunohistochemistry, western blots, co-immunoprecipitation and fluorescence microscopy-type experiments.

Second, Human Proteinpedia allows contributing laboratories to annotate data pertaining to six features of proteins (post-translational modifications, tissue expression, cell line expression, subcellular localization, enzyme substrates and protein-protein interactions; **Supplementary Fig. 2** online). No existing repository currently permits annotation of all these features in proteins.

Third, all data submitted to Human Proteinpedia are viewable through HPRD in the context of other features of the corresponding proteins. To aid comparison and interpretation, meta-annotations pertaining to samples, method of isolation and experimental platform-specific information are provided (e.g., labeling method, protease used, ionization method, details of primary antibody used).

And fourth, in spite of accommodating multiple data types, the data submission is simplified. This means that a biologist with no technical expertise can login and contribute data.

Thus far, a considerable body of data has been contributed to Human Proteinpedia by the community (see **Table 1**), with a total of >1.8 million peptides and >4 million MS/MS-spectra deposited. The above-mentioned data were derived from 2,695 individual experiments (single experiments are defined as immunohistochemistry performed with a specific antibody, a single MS run or a yeast two-hybrid screen). We have imported MS data from two Human Proteome Organization initiatives, including the human plasma proteome project (HPPP)⁶ and the human liver

Human Proteinpedia enables sharing of human protein data

To the editor:

Proteomic technologies, such as yeast two-hybrid, mass spectrometry (MS), protein/peptide arrays and fluorescence microscopy, yield multi-dimensional data sets, which are often quite large and either not published or published as supplementary information that is not easily searchable. Without a system in place for standardizing and sharing data, it is not fruitful for the biomedical community to contribute these types of data to centralized repositories. Even more difficult is the annotation and display of pertinent information in the context of the corresponding proteins. Wikipedia, an online encyclopedia that anyone can edit, has already proven quite successful¹ and can be used as a model for sharing biological data. However, the need for experimental evidence, data standardization and ownership of data creates scientific obstacles.

Here, we describe Human Proteinpedia (<http://www.humanproteinpedia.org/>) as a portal that overcomes many of these obstacles to provide an integrated view of the human proteome. Human Proteinpedia also allows users to contribute and edit proteomic data with two significant differences from Wikipedia: first, the contributor is expected to provide experimental evidence for the data annotated; and second, only the original contributor can edit their data.

Human Proteinpedia's annotation system provides investigators with multiple options for contributing data including web forms and annotation servers (**Supplementary Fig. 1** online). Although registration is required to contribute data, anyone can freely access the data in the repository. The web forms simplify submission through the use of pull-down menus for certain data fields and pop-up menus for standardized vocabulary terms. Distributed annotation servers²

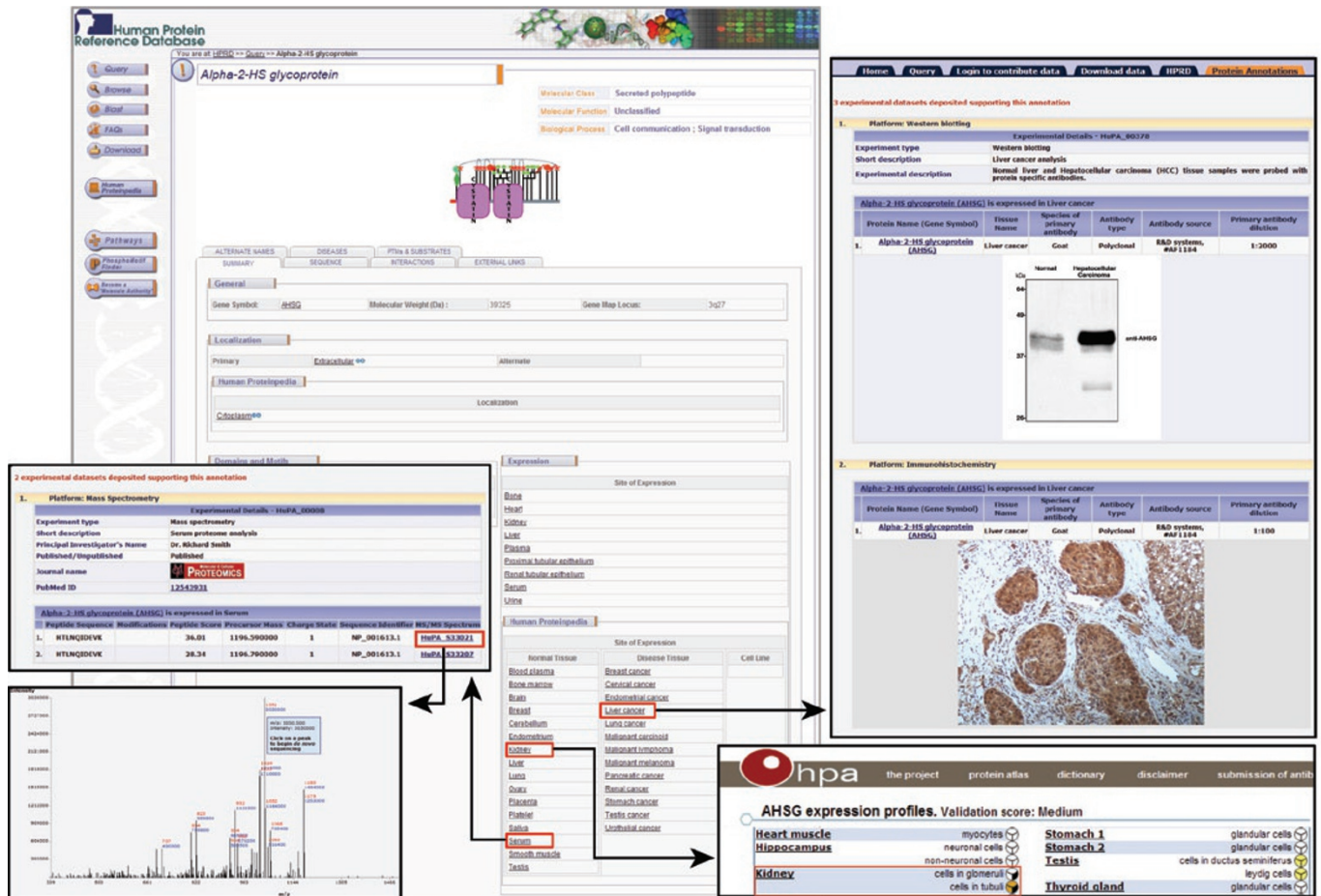


Figure 1 Display of tissue expression data obtained from immunohistochemistry, mass spectrometry and western blot analysis that were submitted to Human Proteinpedia. The molecule page for α -2-HS glycoprotein in Human Protein Reference Database is shown. Annotations pertaining to sites of tissue expression from three types of experimental platforms contributed to Human Proteinpedia are shown. The meta-annotation of a mass spectrometry experiment confirming expression in serum is shown (lower left) along with the corresponding MS/MS spectrum for a peptide displayed using a spectrum viewer obtained from PRIDE. Expression of α -2-HS glycoprotein in liver cancer is provided by two entries: one from western blotting (top right) and the other from immunohistochemical labeling, each detailing the antibody used. Expression in kidney is hyperlinked to the corresponding page in the Human Protein Atlas, which contains data from immunohistochemical labeling experiments (lower right).

proteome project (HLPP)⁷. Data from other initiatives like the Human Protein Atlas⁸, Human Unidentified Gene-Encoded (HUGE) project⁹ and LIFEdb¹⁰ are also included in this effort. The vocabulary follows accepted community standards,

such as eVOC¹¹, Gene Ontology¹², RESID¹³, PSI-MI¹⁴ and PSI-MS.

All data in Human Proteinpedia are freely available to the community for downloading. Storage and dissemination of the raw and processed MS data sets is through the

Tranche file-sharing network supporting ProteomeCommons.org¹⁵ (<http://www.proteomecommons.org/dev/dfs/>) whereas the remainder of the data is available directly from Proteinpedia website. The MS data sets are currently hosted by over 16 servers (~50 TB of aggregate capacity) in triplicate including three servers (in India, Japan and the United States) set up especially for this initiative (**Supplementary Fig. 1**).

Four years of continued manual curation of the scientific literature has yielded >228,800 protein annotations in HPRD. The information content in Proteinpedia collected over the course of the past year alone has resulted in the addition of roughly an equal number of entries. Although large collaborations for genome projects are now commonplace, this collaborative initiative of 71 participating laboratories is unprecedented for proteomic data. We are now beginning to contact authors of

Table 1 Statistics of submitted data

Data type	Number
Individual laboratories submitting data	71
Experiments submitted	2,695
Protein annotations	203,293
MS/MS spectra	4,567,235
Peptide sequences deposited	1,851,124
Protein expression	138,487
Post-translational modifications	17,108
Protein-protein interactions	31,476
Subcellular localizations	2,906
Phosphorylation	16,062

published proteomic studies in a systematic fashion requesting their participation. We eventually hope to capture most of the published data sets, which could be facilitated if depositing experimental data to a public repository is made mandatory for publication, as is already the case for nucleotide sequences, gene expression profiles and protein structures. In this regard, we have instituted a mechanism that allows referees to access data sets submitted by authors before publication in an authorized and anonymous fashion for evaluation purposes.

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

This project has been funded in part by a grant from the National Institutes of Health Roadmap initiative U54 RR020839 (J.S.B., P.A.C., R.J.C. and A.P.) and a contract N01-HV-28180 from the National Heart Lung and Blood Institute (J.V.E., R.J.C. and A.P.). The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organization imply endorsement by the United States Government.

- Giles, J. *Nature* **438**, 900–901 (2005).
- Dowell, R.D., Jokerst, R.M., Day, A., Eddy, S.R. & Stein, L. *BMC Bioinformatics* **2**, 7 (2001).

- Peri, S. *et al. Genome Res.* **13**, 2363–2371 (2003).
- Desiere, F. *et al. Genome Biol.* **6**, R9 (2005).
- Martens, L. *et al. Proteomics* **5**, 3537–3545 (2005).
- States, D.J. *et al. Nat. Biotechnol.* **24**, 333–338 (2006).
- He, F. *Mol Cell Proteomics* **4**, 1841–1848 (2005).
- Uhlen, M. *et al. Mol. Cell Proteomics* **4**, 1920–1932 (2005).
- Nakajima, D. *et al. DNA Res* **12**, 257–267 (2005).
- Mehrle, A. *et al. Nucleic Acids Res.* **34**, D415–D418 (2006).
- Kelso, J. *et al. Genome Res.* **13**, 1222–1230 (2003).
- Ashburner, M. *et al. Nat. Genet.* **25**, 25–29 (2000).
- Garavelli, J.S. *Nucleic Acids Res.* **31**, 499–501 (2003).
- Hermjakob, H. *et al. Nat. Biotechnol.* **22**, 177–183 (2004).
- Falkner, J.A., Falkner, J.W. & Andrews, P.C. *Bioinformatics* **23**, 262–263 (2007).

Suresh Mathivanan^{1–4}, Mukhtar Ahmed¹, Natalie G Ahn⁵, Hainard Alexandre⁶, Ramars Amanchy^{2,3}, Philip C Andrews⁷, Joel S Bader^{8,9}, Brian M Balgley¹⁰, Marcus Bantscheff¹, Keiryn L Bennett¹², Erik Björling¹³, Blagoy Blagoev¹⁴, Ron Bose¹⁵, Samir K Brahmachari¹⁶, Alma S Burlingame¹⁷, Xosé R Bustelo¹⁸, Gerard Cagney¹⁹, Greg T Cantin²⁰, Helene L Cardasis²¹, Julio E Celis²², Raghobama Chaerkady^{1,2}, Feixia Chu¹⁷, Philip A Cole¹⁵, Catherine E Costello²³, Robert J Cotter²⁴, David Crockett²⁵, James P DeLany²⁶, Angelo M De Marzo²⁷, Leroi V DeSouza²⁸, Eric W Deutsch²⁹, Eric Dransfield³⁰, Gerard Drewes³¹, Arnaud Droit³¹, Michael J Dunn¹⁹, Kojo Elenitoba-Johnson²⁵, Rob M Ewing^{32,33}, Jennifer Van Eyk³⁴, Vitor Facc³⁵, Jayson Falkner⁷, Xiangming Fang³⁶, Catherine Fenselau³⁷, Daniel Figgeys³⁸, Pierre Gagné³¹, Cecilia Gelfi³⁹, Kris Gevaert⁴⁰, Jeffrey M Gimble^{41,42}, Florian Gnäd⁴³, Renu Goel¹, Pavel Gromov²², Samir M Hanash³⁵, William S Hancock⁴⁴, HC Harsha^{1–3}, Gerald Hart³, Faith Hays³⁷, Fuchu He^{45,46}, Prashantha Hebbar¹, Kenny Helsen⁴⁰, Heiko Hermeking⁴⁷, Winston Hide⁴⁸, Karin Hjerno¹⁴, Denis F Hochstrasser^{6,49}, Oliver Hofmann⁴⁸, David M Horn⁵⁰, Ralph H Hruban⁵¹, Nieves Ibarrola¹⁸, Peter James⁵², Ole N Jensen¹⁴, Pia Hønnerup Jensen¹⁴, Peter Jung⁴⁷, Kumaran Kandasamy^{1–4}, Indu Kheterpal⁵³, Reiko F Kikuno⁵⁴, Ulrike Korf⁵⁵, Roman Körner⁵⁶, Bernhard Kuster¹¹, Min-Seok Kwon⁵⁷, Hyoung-Joo Lee⁵⁷, Young-Jin Lee⁵⁸, Michael Lefevre^{41,59}, Minna Lehtvaslaihio⁴⁸, Pierre Lescuyer⁶, Fredrik Levander⁵², Megan S Lim²⁵, Christian Löbke⁵⁵, Joseph A Loo⁶⁰, Matthias Mann^{14,43}, Lennart Martens^{40,61}, Juan Martinez-Heredia⁶², Mark McComb²³, James McRedmond¹⁹, Alexander Mehrle⁵⁵, Rajasree Menon⁶³, Christine A Miller⁵⁰, Harald Mischak⁶⁴, S Sujatha Mohan¹, Riaz Mohmood⁴, Henrik Molina^{2,3}, Michael F Moran^{65,66}, James D Morgan²⁷, Robert Moritz⁶⁷, Martine Morzel⁶⁸, David C Muddiman⁶⁹, Anuradha Nalli^{1–3}, J Daniel Navarro¹, Thomas A Neubert²¹, Osamu Ohara^{54,70}, Rafael Oliva⁶², Gilbert S Omenn⁷¹, Masaaki Oyama⁷², Young-Ki Paik⁵⁷, Kyla Pennington¹⁹, Rainer Pepperkok⁷³, Balamurugan Periaswamy^{1–4}, Emanuel F Petricoin⁷⁴, Guy G Poirier³¹, T S Keshava Prasad¹, Samuel O Purvine⁷⁵, B Abdul Rahiman⁴, Prasanna Ramachandran⁶⁰, Y L Ramachandra⁴, Robert H Rice⁵⁸, Jens Rick¹¹, Ragna H Ronnholm⁷⁶, Johanna Salonen⁷⁷, Jean-Charles Sanchez⁶, Thierry Sayd³⁰, Beerelli Seshi⁷⁸, Kripa Shankari¹, Shi Jun Sheng³⁴, Vivekananda Shetty²¹, K Shivakumar^{1,4}, Richard J Simpson⁶⁷, Ravi Sirdeshmukh⁷⁹, K W Michael Siu²⁸, Jeffrey C Smith³⁸, Richard D Smith⁷⁵, David J States⁶³, Sumio Sugano⁷², Matthew Sullivan¹⁹, Giulio Superti-Furga¹², Maarit Takatalo⁷⁶, Visith Thongboonkerd⁸⁰, Jonathan C Trinidad¹⁷, Mathias Uhlen¹³, Joël Vandekerckhove⁴⁰, Julian Vasilescu³⁸, Timothy D Veenstra⁸¹, José-Manuel Vidal-Taboada⁶², Mauno Vihinen⁷⁷, Robin Wait⁸², Xiaoyue Wang³, Stefan Wiemann⁵⁵, Billy Wu⁴⁴, Tao Xu²⁰, John R Yates²⁰, Jun Zhong^{2,3}, Ming Zhou⁸¹, Yunping Zhu⁴⁵, Petra Zurbig⁶⁴ and Akhilesh Pandey^{2,3}

The authors are listed alphabetically by their last name except for the first (S.M.) and the last (A.P.) authors.

¹Institute of Bioinformatics, International Tech Park, Bangalore 560 066, India. ²McKusick-Nathans Institute of Genetic Medicine and the Departments of Pathology and Oncology, Johns Hopkins University, Baltimore, Maryland 21205, USA. ³Department of Biological Chemistry, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. ⁴Department of Biotechnology, Kuvempu University, Shankaraghatta 577451, Karnataka, India. ⁵Department of Chemistry and Biochemistry, Howard Hughes Medical Institute, University of Colorado, Boulder, Colorado 80309, USA. ⁶Biomedical Proteomics Research Group, Department of Structural Biology and Bioinformatics, Geneva University, Faculty of Medicine, CH-1211 Geneva 4, Switzerland. ⁷Department of Biological Chemistry, University of Colorado, Ann Arbor, Colorado 48109-0674, USA. ⁸The High Throughput Biology Center, Johns Hopkins University, Baltimore, Maryland 21205, USA. ⁹Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland 21205, USA. ¹⁰Calibrant Biosystems, 910 Clopper Road, Suite 220N, Gaithersburg, Maryland 20878, USA. ¹¹Cellzome AG, Meyerhofstrasse 1, 69117 Heidelberg, Germany. ¹²Research Center for Molecular Medicine (CeMM), Lazarettgasse 19/3, 1090 Vienna, Austria. ¹³Royal Institute of Technology, School of Biotechnology, AlbaNova University Center, SE-106, 91 Stockholm, Sweden. ¹⁴Department of Biochemistry and Molecular Biology, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark. ¹⁵Departments of Pharmacology and Oncology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. ¹⁶GN Ramachandran Knowledge Center for Genome Informatics, Institute of Genomics and Integrative Biology, CSIR, Mall Road, Delhi 110 007, India. ¹⁷Department of Pharmaceutical Chemistry, University of California-San Francisco, San Francisco, California 94143-0446, USA. ¹⁸Centro de Investigación del Cáncer and Instituto de Biología Molecular y Celular del Cáncer. CSIC-University of Salamanca, Campus Unamuno, E-37007 Salamanca, Spain. ¹⁹Proteome Research Centre, UCD Conway Institute of Biomolecular and Biomedical Research, Belfield, Dublin 4, Ireland. ²⁰Department of Cell Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA. ²¹Skirball Institute of Biomolecular Medicine and Department of Pharmacology, New York University School of Medicine, New York, New York 10016, USA. ²²Danish Centre for Translational Breast Cancer Research (DCTB) and Department of Proteomics in Cancer, Institute of Cancer Biology, Danish Cancer Society, Strandboulevarden 49, DK-2100, Copenhagen, Denmark. ²³Cardiovascular Proteomics Center, Boston University School of Medicine, Boston, Massachusetts 02118, USA. ²⁴Department of Pharmacology and Molecular Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. ²⁵Department of Pathology, University of Michigan Medical School in Ann Arbor, 1301 Catherine 5240 Medical Science I, Ann Arbor, Michigan 48109-0602, USA. ²⁶Endocrinology and Metabolism Division, Department of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania 15213, USA. ²⁷Departments of Urology and Pathology, The Johns Hopkins University School of Medicine, 600 North Wolfe Street, Baltimore, Maryland 21287, USA. ²⁸Centre for Research in Mass Spectrometry and Department of Chemistry, York University, 4700 Keele Street, Toronto, Ontario M3J 1P3,

Canada. ²⁹Institute for Systems Biology, Seattle, Washington 98103, USA. ³⁰Wageningen Centre for Food Sciences, Diedenweg 20, PO Box 557, 6700 AN Wageningen, The Netherlands. ³¹Health and Environment Unit, Laval University Medical Research Center (CHUL), Faculty of Medicine, 2705 Boul. Laurier, Quebec City, QC, G1V 4G2, Canada. ³²Protana (now Transition Therapeutics), 101 College Street, Suite 220, Toronto, Ontario, M5G 1L7, Canada. ³³Infochromics, MaRS Discovery District, 101 College St., Toronto, Ontario M5G 1L7, Canada. ³⁴Department of Medicine, Johns Hopkins Bayview Proteomics Center, Johns Hopkins University, Baltimore, Maryland 21224. ³⁵Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA. ³⁶GenWay Biotech, Inc., 6777 Nancy Ridge Drive, San Diego, California 92121, USA. ³⁷Department of Chemistry and Biochemistry, University of Maryland, College Park, Maryland 20742, USA. ³⁸The Ottawa Institute of Systems Biology, University of Ottawa, 451 Smyth Road, Ottawa, Ontario, K1H 8M5 Canada. ³⁹Department of Biomedical Sciences and Technologies, Faculty of Medicine, University of Milano, Via Fratelli Cervi 93, Segrate Milano, Italy. ⁴⁰Department of Biochemistry and Medical Protein Research, Faculty of Medicine and Health Sciences, Ghent University, A. Baertsoenkaai 3, B9000 Ghent, Belgium. ⁴¹Stem Cell Laboratory, Pennington Biomedical Research Center, Louisiana State University System, Baton Rouge, Louisiana 70808, USA. ⁴²Cell Biology Core Facility, Pennington Biomedical Research Center, Louisiana State University System, Baton Rouge, Louisiana 70808, USA. ⁴³Department of Proteomics and Signal Transduction, Max-Planck-Institute for Biochemistry, D-82152 Martinsried, Germany. ⁴⁴Barnett Institute and Department of Chemistry and Chemical Biology, Northeastern University, Boston, Massachusetts 02115, USA. ⁴⁵Beijing Proteome Research Center, Beijing Institute of Radiation Medicine, Beijing, 100850, China. ⁴⁶Institutes of Biomedical Sciences, Fudan University, Shanghai, 200032, China. ⁴⁷Molecular Oncology, Independent Max-Planck-Research Group, Max-Planck-Institute of Biochemistry, Am Klopferspitz 18, D-82152 Martinsried, Germany. ⁴⁸South African National Bioinformatics Institute, University of the Western Cape Bellville 7535, South Africa. ⁴⁹Central Clinical Chemistry Laboratory, Geneva University Hospital, Geneva, Switzerland. ⁵⁰Agilent Technologies, 5301 Stevens Creek Blvd., Santa Clara, California 95051, USA. ⁵¹Department of Pathology and Oncology, The Sol Goldman Pancreatic Cancer Research Center, Johns Hopkins Medical Institutions, 401 North Broadway, Weinberg 2242, Baltimore, Maryland 21231-2410, USA. ⁵²Protein Technology, Lund University, BMC D13, Lund, Sweden. ⁵³Proteomic Core Facility, Pennington Biomedical Research Center, Louisiana State University System, Baton Rouge, Louisiana 70808, USA. ⁵⁴Department of Human Genome Technology, Kazusa DNA Research Institute, 2-6-7 Kazusa-kamatari Kisarazu Chiba 292-081, 8 Japan. ⁵⁵Division of Molecular Genome Analysis, German Cancer Research Center—DKFZ Heidelberg, Im Neuenheimer Feld 580, 69120 Heidelberg, Germany. ⁵⁶Department of Cell Biology, Max Planck Institute of Biochemistry, D-82152 Martinsried, Germany. ⁵⁷Department of Biochemistry, Yonsei Proteome Research Center and Biomedical Proteome Research Center, Yonsei University, Seoul 120-749, South Korea. ⁵⁸Department of Environmental Toxicology, University of California, Davis, California 95616-8588, USA. ⁵⁹Lipoprotein Laboratory, Pennington Biomedical Research Center, Louisiana State University System, Baton Rouge, Louisiana 70808, USA. ⁶⁰Department of Chemistry and Biochemistry and Department of Biological Chemistry, University of California Los Angeles, Los Angeles, California 90095, USA. ⁶¹EMBL Outstation, The European Bioinformatics Institute, Wellcome Trust Genome Campus, Cambridge CB10 1SD, UK. ⁶²Human Genetics Research Group, IDIBAPS, Faculty of Medicine, University of Barcelona, Casanova 143, 08023, Barcelona, Spain. ⁶³Bioinformatics Program, University of Michigan, Ann Arbor, Michigan 48109, USA. ⁶⁴Mosaïques Diagnostics & Therapeutics AG, Hanover, Mellendorfer Strasse 7-9D-30625 Hannover, Germany. ⁶⁵Cancer Program, Hospital For Sick Children and McLaughlin Centre for Molecular Medicine, 101 College Street, East Tower, Toronto, Ontario, M5G 1L7 Canada. ⁶⁶Banting and Best Department of Medical Research and Department of Molecular and Medical Genetics, University of Toronto, 112 College St., Toronto, Ontario M5G 1L6, Canada. ⁶⁷Joint Proteomics Laboratory, Ludwig Institute for Cancer Research (Melbourne Branch) and The Walter and Eliza Hall Institute of Medical Research, 1G Royal Parade, Parkville, 3050 Victoria, Australia. ⁶⁸INRA, UR 370, Proteomics Platform, 63122 Saint Genès Champanelle, France. ⁶⁹W.M. Keck FT-ICR Mass Spectrometry Laboratory, Department of Chemistry, North Carolina State University, Raleigh, North Carolina 27695, USA. ⁷⁰Laboratory for Immunogenomics, RIKEN Research Center for Allergy and Immunology, RIKEN Yokohama Institute, 2nd floor, North Research Building, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama City, Kanagawa, 230-0045, Japan. ⁷¹Department of Internal Medicine, Center for Computational Medicine and Biology and Proteomics Alliance for Cancer Research, University of Michigan, Ann Arbor, Michigan 48109-2218, USA. ⁷²Medical Proteomics Laboratory, Institute of Medical Science, University of Tokyo, Minato-ku, Tokyo 108-8639, Japan. ⁷³European Molecular Biology Laboratory, Cell Biology and Biophysics Programme Meyerhofstrasse 1, D-69117 Heidelberg, Germany. ⁷⁴Center for Applied Proteomics and Molecular Medicine, George Mason University, Manassas, Virginia 20110, USA. ⁷⁵Environmental Molecular Science Laboratory and Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Washington 99352, USA. ⁷⁶Department of the Biological and Environmental Sciences, Division of Biochemistry, University of Helsinki, PO Box 56, Viikinkaari 9, FI-00014, Finland. ⁷⁷Institute of Medical Technology, University of Tampere, Finn-Medi2, 5th floor, Biokatu 8, Tampere, FI-33014, Finland. ⁷⁸Department of Pathology, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, David Geffen School of Medicine at UCLA, Torrance, California 90502-2064, USA. ⁷⁹Centre for Cellular and Molecular Biology, Uppal Road, Hyderabad 500 007, India. ⁸⁰Medical Molecular Biology Unit, Office for Research and Development, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand. ⁸¹Laboratory of Proteomics and Analytical Technologies, SAIC-Frederick Inc., NCI-Frederick, Frederick, Maryland 21702, USA. ⁸²Kennedy Institute of Rheumatology, Faculty of Medicine, Imperial College, South Kensington Campus, London SW7 2AZ, UK. e-mail: pandey@jhmi.edu