HUPO Highlights

Highlights of B/D-HPP and HPP Resource Pillar Workshops at 12th Annual HUPO World Congress of Proteomics

September 14–18, 2013, Yokohama, Japan

Ruedi Aebersold[1], Gary D. Bader[2], Aled M. Edwards[2], Jennifer van Eyk[3], Martin Kussman[4], Jun Qin[5] and Gilbert S. Omenn[6]

[1] Institute of Molecular Systems Biology, Seattle, WA, USA
[2] Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada
[3] Johns Hopkins Bayview Proteomics Center, Johns Hopkins University, Baltimore, MD, USA
[4] Molecular Biomarkers, Nestlé Institute of Health Sciences, Lausanne, Switzerland
[5] Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, TX, USA
[6] Center for Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA

At the 12th Annual HUPO World Congress of Proteomics in Japan, the Human Proteome Project (HPP) presented 16 scientific workshop sessions. Here we summarize highlights of ten workshops from the Biology and Disease-driven HPP (B/D-HPP) teams and three from the HPP Resource Pillars. Highlights of the three Chromosome-centric HPP sessions appeared in the many articles of the 2014 C-HPP special issue of the Journal of Proteome Research [1].

1 Introduction

The HUPO Human Proteome Project (HPP) is a very active global program, with nearly 50 teams engaged in experimental research, extensive data mining, and resource development [2–4]. The goals of the HPP are (i) major progress in making proteomics technologies and knowledge practical and accessible for wide use across the life sciences and biomedical research communities and (ii) completing a gene-centric parts list for proteins and their isoforms. The Grand Challenge is to use proteomics to bridge major gaps between evidence of genomic variation and diverse phenotypes [5]. Much of that progress lies in understanding and visualizing the biochemical and signaling pathways that capture environmental and

Correspondence: Dr. Gilbert S. Omenn, University of Michigan - Center for Computational Medicine and Bioinformatics, 100 Washtenaw Avenue, Room 2065B, Palmer Commons, Ann Arbor, Michigan 48109-2218, USA
E-mail: gomenn@umich.edu
Fax: +1-734-615-6553

Abbreviations: AH, aqueous humor; **B/D-HPP**, Biology and Disease-driven Human Proteome Project; **Ca-HPP**, Cancer Human Proteome Project; **CPTAC**, Clinical Proteomic Tumor Analysis Consortium; **HDPP**, Human diabetes proteome project; **HEPP**, human EyeOME; **HGPI**, Human Glycoproteomics Initiative; **iMOP**, Initiative on Model Organism Proteomes; **PCT**, pressure cycling technology; **PE**, protein evidence; **PX**, proteomeXchange

behavioral interactions with genetic predispositions, in identifying and quantifying the protein complexes that serve as functional modules to catalyze most biological functions, in determining the roles of protein isoforms and their protein–protein interactions and on identifying the roles of noncoding RNAs and other regulatory features identified with ENCODE, the Encyclopedia of DNA Elements.

Here we present highlights of the biology and disease-driven component of the HPP and the knowledge base, antibody capture, and MS resource pillars, as reflected in the well-attended and quite lively scientific workshops at the 12th Annual HUPO Congress, in Yokohama, Japan, September 14–18, 2013.

2 Workshop summaries

2.1 Knowledge base resource pillar (KBpC) and proteomics standards initiative (PSI)

(Chaired by Lydie Lane, SIB, Geneva, Switzerland, and Henning Hermjakob, EBI, UK.)

For the Proteomics Standards Initiative, Eric Deutsch (Institute for Systems Biology, Seattle, USA) reported that, during the past year, the mzQuantML standard for quantitative information and the corresponding MIAPE Quant guidelines were finalized and published. mzML (MS data), mzIdentML (peptide/protein identifications), TraML (SRM transition lists), and PSI-MI (molecular interactions) are all stable standards. Under development are mzTab (tabular format for identification and quantification results) and PEFF (extended FASTA format for protein sequence databases).

Henning Hermjakob reported that the ProteomeXchange (PX) submission pipeline has been in production since summer 2012, involving at present PRIDE (for MS/MS data) and PeptideAtlas/PASSEL (for SRM data). ProteomeXchange is working closely together with the PSI, with many overlapping staff. Other repositories could join in the near future. The idea is to have a mutual backup to ensure long-term sustainability. The system is heavily used, with more than 400 datasets submitted in the first year [and >650 at the end of 2013]; many uploaded datasets are not publicly released until the relevant manuscript is accepted. Their availability during peer review is valuable to authors and reviewers alike. The HPP requires that investigators submit their experimental datasets to PX and cite the PXD identifier at the end of the Abstract and in the Methods. Datasets can be centrally searched at ProteomeCentral (proteomecentral.proteomexchange.org).

Integration of Proteomics Data in neXtProt was described by Lydie Lane. Protein and peptide/PTM identifications from PeptideAtlas and 20 additional curated large-scale proteomics datasets have been integrated in neXtProt. All peptides from PeptideAtlas human builds were integrated as "gold" (protein evidence level 1). In total, 420 330 peptides were mapped to 15 509 proteins and 45 430 PTM annotations were integrated. Information is integrated with RNA expression data from different sources and protein expression data from the Human Protein Atlas. There are five levels of "Protein Existence" (PE) in neXtProt: level 1 (evidence at the protein level) is given for 15 509 proteins (out of 20 128 entries in SwissProt); among them 1,216 proteins that have PE level 1 were validated by methods other than MS. At the transcriptome level (PE level 2) 17.7% of the remaining proteins have evidence of expression. [6] (See Lane et al., J Proteome Res 2014, for detailed assessment of evidence and for metrics of the Human Proteome Project.)

In discussion, participants emphasized that the amount of proteomics data submitted to public resources keeps increasing. HPP datasets have to be submitted to ProteomeXchange. Reviewers should reinforce this requirement. The wiki of the C-HPP is a useful resource, but information needs to be extended and improved. A workshop during the HUPO conference aimed to improve the wiki and its use.

2.2 Antibody pillar committee and human antibody initiative

(Chaired by Matthias Uhlen, Sweden, and Michael Snyder, USA.)

The chairs highlighted the contribution of antibodies to the human proteome project, the challenge of demonstrating antibody performance and specificity in different kinds of assays, the huge and increasing number of antibodies, multiple protein capture technologies, and the marriage of antibody and MS approaches.

Ed Nice (Monash University, Australia) described the Victoria antibody initiative, a one-stop shop for antibodies. They have produced with robotic infrastructure >1000 hybridoma cell lines from 500 to 1000 projects per year of monoclonal Ab production in two mice per antigen. Validation is performed on SDS-PAGE, Western blot, surface plasmon resonance, cell sorting, and immunohistochemistry. Phage display is utilized in parallel. They are engaged in chromosome 7 and chromosome 17 projects, with a pilot project using 94 PrESTs, including olfactory receptors as targets. They collaborate with SciLifeLab and the Human Protein Atlas in Stockholm. Their antibodies are available on favorable terms for HPP investigators.

Peter Nilsson (Sweden) presented the European Union Affinomics consortium, led by Michael Taussig. Established in 2010, this AffinityProteome binders project comprises 19 partners and seven working parties, from generation of binders, antigens, and databases, to molecular tools and validation by immunohistochemistry, immunofluorescence, Western blot, and protein microarrays. Antigen production for the targets specified by the program has been completed; binder production is 75% complete. For example, KTH has delivered 60 PrEST antigens and 800 binders. They have focused on renewable binders of different scaffolds for 1000 targets on kinases, SH2, phosphatases, and cancer related proteins (www.proteinbinders.org).

From the United States, Salvatore Sechi of NIH (Bethesda, MD) reviewed the parameters of the program of renewable antibody reagents. The pilot at seven centers is focused on transcription factors; two centers dealing with antigens are using recombinant antibody network immunoprecipitation and protein arrays with monospecific monoclonal antibodies; the other centers are improving the pipeline from target selection to reagent distribution with diverse platforms. There is a central website to access these binders. The centers interact with Human Protein Atlas and Affinomics.

A major annotated database for antibodies, mostly commercial, is Antibodypedia. Tove Alm (Sweden) described the combined efforts of academia and companies. Interested researchers can track information about available antibodies for various applications, with some annotation from the shared experience of the broader community. Features include a gene-centric structure, search and compare functions, advanced filtering options (e.g. application), and direct link to provider. Nearly 1 000 000 nonunique antibodies are entered in the portal, covering 90% of human genome, with information organized by chromosome. About 72% of all genes (15 000) have >10 antibody entries, 19% (3900) have 1–9, and 9% (1900) have no antibody. For example, there are 2500 antibodies listed for EGFR. In all there are >100 000 publication references, providing useful information about antibody functionality.

There was extensive discussion. Should HUPO offer a service of quality-assured antibodies? No, the data should speak for themselves, reflecting data labs, different applications, but different results for the same listed antibody. General endorsements are not appropriate. Antibodypedia welcomes conflicting reports; however, there have actually been relatively few contributions. Users extract information far more than contributing information. UniProt similarly has had low participation from the community for its Wikipedia approach. Experience shared via UCSC Genome Browser indicates that about 50% of polyclonal antibodies and 30% of monoclonal antibodies "work," a very rough summary of extremely heterogeneous

applications. Off-target binding remains common to all binding reagents. The overall focus remains on relatively few targets with many binders. Overall 70% of Ab binders are polyclonal and 30% monoclonal. There are multiple other binders, including SOMAmers with pM affinity reported. Methods for high throughput production of recombinant binders require further development. Clones of mAb remain available, for example, in the hybridoma bank at the University of Iowa.

2.3 The human EyeOME (HEPP)

(Organized by Richard Semba, Johns Hopkins, USA, the newest B/D-HPP initiative participating was the HEPP.)

Jennifer van Eyk (Johns Hopkins) noted that investigative proteomic tools can transform vision science and enhance understanding of physiology and disease processes that affect sight. An estimated 285 million people worldwide have major visual impairment; 39 million are blind. A literature review of 4842 total proteins reported for eight different eye compartments was published by the HEPP [7, 8]. HEPP goals are to standardize tissue collection methods, establish a repository, develop a reference data set, exchange trainees, and develop an Eyeome community.

The human tear proteome was introduced by Zhou Lei (Singapore Eye Research Institute and Duke-National University of Singapore). Tears are an easily accessible body fluid covering the ocular surface; ~1500 proteins have been identified from normal human tears. One dimensional-RPLC-MS/MS identified 500–700 tear proteins using 4 mg of total tear proteins (0.5 mL tears), and MS/MS ALL with SWATH acquisition enabled MRM-based quantitation of 500 proteins.

Roger Beuerman (Singapore) discussed biomarker development using tears for several inflammatory disorders. Clinical severity of the dry eye condition was correlated with a four-protein biomarker panel. Similar biomarkers for MGD, pterygium, thyroid eye disease, and lacrimal gland tumors correlated with disease, and may outperform current clinical signs and testing. Tear biomarkers even distinguished between surgical wounds created by different instruments. An ongoing clinical trial using tear proteomic biomarkers may outperform lengthy, often inconclusive clinical observations. The audience discussed specific methods for optimizing tear protein collection, how quantitative biomarker measurements can correlate with disease severity, and international tissue sharing and collaboration.

The anterior segment proteome is important for glaucoma studies, as described by Deepak Edward (Johns Hopkins and King Khalid Eye Hospital, Saudi Arabia). Glaucoma is a heterogeneous group of diseases leading to optic neuropathy, a leading cause of blindness globally. Glaucoma involves impairment of the aqueous humor (AH) outflow from the anterior chamber. Proteomic analyses of the AH, ciliary body, iris, and trabecular meshwork are relevant to disease pathogenesis. The normal iris proteome included 550 proteins. The AH proteome of controls and primary congenital glaucoma (PCG) patients and biomarkers associated with other types of glaucoma are under study in the AH, ciliary body, and trabecular meshwork.

Vinit Mahajan (University of Iowa, USA) presented proteomic studies of patients with genetically identical inflammatory vitreoretinal disease. Disease stage-specific cytokines were identified in surgical samples from patients with inherited CAPN5-associated vitreoretinal disease. Proteomic profiles differentiated drug-responsive and nonresponsive pathways. The absence of TNF-alpha and B-cell markers correlated with prior treatment failure using infliximab and B-cell immunosuppression. Biomarkers showed anti-VEGF therapy was highly effective in controlling neovascularization as was anti-IL-17 therapy for vitreous inflammation. Failure to control pathologic fibrosis was associated with persistent expression of IL-6; IL-6 therapy improved a case of tractional retinal detachment. The audience

discussed the limits of genomics for personalized medicine and the advantages of proteomics to identify disease stages and severity.

2.4 Cancer human proteome project

(Chairs, Hui Zhang, Johns Hopkins, USA; Christopher Kinsinger, NCI, USA; Juan Pablo Albar, Spain.)

As stated by Hui Zhang, the goals of the cancer human proteome project (Ca-HPP) are (1) to identify and prioritize proteins related to specific cancer pathways and cancer types, and (2) to generate and disseminate assays and resources to support the analysis of biological networks underlying biological processes or clinical investigations of cancer, including antibodies, affinity-based arrays, targeted MS, and immuno-MRM.

Christopher Kinsinger described the Proteogenomics Network of the Cancer Clinical Proteomic Tumor Analysis Consortium (CPTAC), National Cancer Institute, USA. The CPTAC is currently analyzing approximately 100 tumors from each of three tumor types: colorectal, breast, and ovarian cancers. Both proteomic and genomic data and antibodies are publicly available. In addition, the network is developing a portal for MRM and immuno-MRM assays.

Analysis of tissue biopsies by pressure cycling technology (PCT) SWATH was the theme of Guo Tiannan (ETH Zurich, Institute of Molecular Systems Biology, Switzerland). A PCT to lyse tissue from biopsies for proteomic analysis by SWATH reproducibly yields 50 μg of peptides from 1 mg wet weight tissue in a process that lasts 12 h, across a variety of samples. Applied to 18 kidney tumor biopsies, the CV was 20% for the raw SWATH signals, 10% for MS2 spectra, and 20% for the MS1 spectra. SWATH identified some 3000 proteins in common across the samples, while data-dependent analysis (DDA) detected many more proteins in only one sample but fewer proteins consistently across the cohort. Using the OPEN-Swath software tool to cluster the PCT-SWATH data, known subtypes of kidney cancer (papillary and clear cell) were clearly distinguished.

Peter James (Lund University, Sweden) presented research on DNA repair pathways in breast cancer. An elaborate workflow was applied to 15 000 fresh frozen breast cancer tissue samples and breast cancer cell lines to identify and classify proteins by organelle, select proteins associated with DNA damage repair, and develop SRM assays for those proteins. The heterogeneity of tumor tissue is clearly revealed when cells are FACS sorted prior to analysis. Another issue is different proteomes between cell model and human tumor. Of 13 241 proteins identified across the sample cohort, 3000 were unique to tumor, 3000 unique to cell line, and 7400 in common; 250 proteins are involved with DNA damage repair. For breast cancer subtypes treated with various chemotherapeutic agents, the assay confirms chemosensitivity in Luminal A samples and chemoresistance in Basal tumors.

Integrative Analysis: Bridging the Gaps in Genetic Information Flow between Genomic and Proteomic Data was the topic of Zhen Zhang (Johns Hopkins, USA). Challenges of cancer proteomics include intratumor heterogeneity, dearth of shared genomic aberrations within the same histopathological tumor subtype, many data gaps, lack of early-stage disease-specific signals, low prevalence, limited statistical power, and overfitting of data. The solution space at a given false discovery rate is too large for cost-effective follow-up studies. One solution is to reduce the solution space by triangulating multiple data sets of different types and/or sources. For example, proteomic and genomic data for CompRef, a xenograft-based sample used by CPTAC, served as a training set for differential analysis, enrichment analysis, network signature identification, and then validation by network-based classification. The result separated basal and luminal subtypes from a test set of samples.

Edouard Nice (Monash University, Australia) spoke about proteomics-based biomarker studies of colorectal cancer. Their three-stage approach to identify candidate biomarkers comprises (1) an inventory of proteins in 10 colon cancer-associated human cell lines, with 1D-SDS nanoLC-MS/MS of secretome and exosome of each cell line; (2) search for the 2500 proteins so identified in fecal material from the APC min mouse model of colon cancer, using 1D and multidimensional purification steps, yielding 91 murine proteins of interest; and (3) collection and planned SRM analysis of fecal samples from ten human colon cancer subjects and five normal subjects.

Dong Li (Beijing Proteome Research Center, China) described proteome-wide prediction of the substrate for human ubiquitin-protein ligase. Several methods are being employed to predict and identify substrates for the E3 component of the ubiquitin-proteasome system.

Panel discussion was led by Hisashi Narimatsu, Tadashi Kondo, and Ignacio Casal. Each panel member raised the issue of intratumor heterogeneity. Laser-capture microdissection was recommended. Another suggestion was to consider 2D-PAGE rather than MS as the primary analysis method. The panel was highly supportive of the use of animal models to understand cancer proteomes, but clearly recognized the importance of answering the requirements of clinicians, not just scientific curiosity.

Ruedi Aebersold encouraged the group to pool resources as a means to enable access to results and reagents to as many researchers as possible. One way to do this for SRM assays would be to set up a cancer-specific section of SRMAtlas. Another solution would be to expand the CPTAC portals for antibodies and SRM assays to allow deposition of reagents and assays from outside of CPTAC. The CPTAC guidelines would be maintained for all depositors. Kinsinger asked how to define a Ca-HPP; for instance, how many tumors of a given cancer type would constitute a sufficient set of characterized samples? Moreover, the source of the samples poses no small challenge, particularly if proteomic analysis continues to require fresh frozen, resected tumors with minimal ischemic exposure. There was general interest in new approaches to proteogenomic analysis, moving beyond from protein inventory experiments toward biological insights.

Finally, discussions following the formal session revealed substantial interest in coordination of a formal Ca-HPP. Perhaps a full-day workshop at the 2014 HUPO World Congress is in order. Moreover, the United States has a well-coordinated effort in this area with CPTAC, now achieving connection with the TCGA genomics/epigeomics program. It was not clear from the workshop that other regions had cancer proteomics projects that extended beyond a single lab. Perhaps this is an opportunity for the HUPO Ca-HPP.

2.5　HUPO brain proteome project (HBPP)

(Chairs Lea T. Grinberg, Sao Paulo, Brazil; Helmut E. Meyer, Bochum, Germany; Young Mok Park, Daejeon, Korea.)

The 20th HBPP workshop of the HUPO Brain Proteome Project Initiative (HUPO BPP) engaged 40 participants to discuss new developments in the proteomics and pathology of neurodegenerative diseases. The aim of the HUPO Brain Proteome Project Initiative is to understand human diseases which affect the central nervous system, especially the biogenesis of Alzheimer's disease (AD), Parkinson's disease (PD), and fronto-temporal dementia (FTD).

Lea T. Grinberg highlighted the Contributions of Neuropathology to Elucidating Neurodegenerative Diseases: Insights from a Revamped Science. She summarized the common features of neurodegenerative diseases and reported C9orf72 mutations as the most common causes of familial FTD and FTD-ALS.

Helmut E. Meyer presented Biomarker Discovery for Alzheimer and Parkinson disease, and Andreas Schrötter (Bochum, Germany) spoke on Differential Proteomic Analysis of Human Hippocampal Regions of Interest (CA1, CA2, CA3, fascia dentata) Relevance for Autoimmunity in Alzheimer`s disease? They both propose the hypothesis that AD arises as a result of pathology of the immune system. The generation of autoantibodies against neuronal structures could drive the loss of neurones leading to AD.

The HUPO brain proteome project initiative is focused on the detailed biochemical mechanism of the biogenesis of the neurological diseases and on development of biomarkers to detect the disorders early, when changes of the biochemical equilibrium might be compensated by specific treatment before there is irreversible damage.

2.6 Joint human kidney/urine proteome project (HKUPP) and human plasma proteome project (HPPP) workshop

(Chaired by Tadashi Yamamoto, Niigata, Japan, and Mark Baker, Sydney, Australia.)

Tadashi Yamamoto presented Proteomic Analysis of Nephron Segments of Formalin-Fixed Paraffin-Embedded Human Kidney Tissues. The HKUPP has an on-going experimental program to show the origin of urine proteins from kidney and from other urinary tract tissues, using MS analysis of nephron segments and other tissue components from formalin-fixed paraffin-embedded kidney specimens. The proportion of urinary proteins derived from the kidney and blood was discussed, referring to the KHUPP database.

A Collective Analysis of Three Human Subproteomes using PeptideAtlas was summarized by Eric Deutsch of the Institute of Systems Biology (Seattle, USA). Comparative analysis of kidney, urine, and plasma subproteome PeptideAtlases showed which proteins are unique to, or highly enriched in, each of the subproteomes relative to the others; the same analysis revealed which urine proteins were likely derived from kidney, which likely derived from plasma, and which likely not derived from either. The detailed findings have been published in Farrah et al., [9]. The significance of standardization for such comparisons was discussed.

Mark Baker then led a discussion of the status and future directions for the Human Plasma Proteome Project. He invited HUPO colleagues interested in plasma proteomics to join an expanded mailing list and to contact him with regard to EuPA and PanAm HUPO leadership representatives for the HPPP going forward. Proteome analysis of the cellular components of blood cells was a specific recommendation.

Different Levels of Variability in the human plasma proteome was discussed by Yansheng Liu (ETH Zurich). He introduced the newly developed data independent acquisition (DIA) method SWATH MS with quantitative analyses of the plasma proteome. The principle of the method and individual difference in plasma proteome were discussed.

Bruno Domon (Luxembourg) presented Quantification of Peptides in Clinical Samples based on high-resolution parallel reaction monitoring (PRM) mass measurements. This new method for quantification of peptides extracts postacquisition ion traces of specific fragment ions and applies the results to the analysis of complex clinical samples (plasma and urine). He demonstrated benefits of the gain in selectivity and the identification of the fragments through accurate mass to increase confidence in the measurements. The participants agreed that the cross-initiative collaboration and the joint initiative workshop was quite useful and stimulating.

2.7 Human liver proteome project workshop

(Chaired by Pumin Zhang, Beijing, China; Felix Elortza, Bilbao, Spain; Fernando Corrales, Pamplona, Spain.)

Felix Elortza described Proteomics Analyses of Urine as a Source of Liver Disease Biomarkers. Liver proteomic research continues to be a major focus for CIC bioGUNE's proteomics platform. For multiple liver pathologies, they employ peptidome profiling and exosome proteomics with standardized protocols for urine specimens from patients. In addition, they collaborate with Carolina Armengol to perform differential proteomics analyses on human hepatoblastoma samples. Hepatoblastoma is a malignant embryonal tumor of the liver usually diagnosed in children <3 years of age. They have combined DIGE with the nLC MS/MS label-free quantitation searching for new prognostic biomarkers. The results obtained from both techniques are complementary.

Fernando Corrales (Pamplona, Spain) presented Identification of Proteins Driving the Progression of Liver Injury and Potential Biomarkers. They have identified mitochondrial Prohibitin 1 as a central protein to preserve liver homeostasis; its downregulation correlates with development of nonalcoholic steatohepatitis. Defective prohibitin 1 induces apoptosis in human hepatoma cells; its partial deletion in mice induces liver damage and an exacerbated inflammatory response. Meanwhile, proteomic approaches led them to define a collection of proteins that might be relevant in the progression of human hepatocellular carcinoma (HCC). Vasodilator-stimulated phosphoprotein has emerged as a biomarker candidate for the early detection of HCC as its plasma levels increase significantly. This change can be detected in cirrhosis, which predisposes to HCC. A group of HCC patients subjected to radioembolization is under study.

Toward the System Medicine of NonAlcoholic Fatty Liver Disease (NAFLD) was presented by Tommy Nilsson (McGill University, Canada). NAFLD comprises a disease spectrum from benign nonalcoholic fatty liver to nonalcoholic steatohepatitis (NASH) to cirrhosis and risk of hepatocellular carcinomas. Between 3 and 19% of general populations have some form of NASH (inflammation) depending on their body mass index. The McGill liver biobank performs sample processing, subcellular fractionation with organelles, and advanced bioinformatics. They reported identification of 6053 proteins (FDR below 0.1%, per Protein Prophet/ Scaffold).

Finally, Pumin Zhang described progress made by Chinese Human Liver Proteome Project collaborating groups. The summary of results is available in Liverbase2.0 with >12 000 proteins identified from human liver; more than half were assigned cellular locations. Liverbase3.0 will be constructed to show profiles of the four major cell types.

2.8 Proteome biology of stem cells HUPO

(Chaired by AJR Heck, Utrecht, The Netherlands.)

Myo Kiwasaki (Kyoto University, Japan) described the use of meter scale monolithic LC for in-depth analysis of induced pluripotent stem cells (iPSC) and of fibroblasts, identifying 5000 protein groups within a single LC-MS analysis. Subcellular fractionation enhanced detection of lower abundance proteins. In all, 46% of 8700 proteins were identified in both samples, while 29 and 25% were found only in iPSC and in fibroblasts, respectively, with distinctive functional annotations.

Albert Heck explored the Proteome Biology of Adult Stem Cells, which have enormous potential for therapeutic applications. Because of their low abundance, unique markers are needed for their isolation and characterization. For example, Lgr5 was identified as one of the surface markers uniquely expressed by a

crypt stem cell subtype (cycling CBC cells) from the intestinal epithelium. Lgr5 associates with Wnt receptor and has R-spondin as ligand. FACS sorted Lgr5+/− cells were characterized by proteomics (4000 proteins quantitated) and transcriptome analysis. Biological replicates had a Pearson correlation of 0.7 for the proteomics measurement; technical replicates had $r = 0.85$. They found $r = 0.7$ between the proteomics and transcriptomics measurements.

Akhilesh Pandey (Johns Hopkins, USA) presented a study of embryoid bodies from embryonic stem cell (ESC) differentiation and selected for neural progenitor cells. Two differentiation protocols generated astrocytes and motor neurons, respectively. Proteomics analysis using SCX prefractionation, LC-MS/MS, and iTRAQ-8plex quantification of 1200 proteins found stage-specific dynamic markers in ES cells differentiated into neurons or astrocytes. From distinct expression patterns, seven targets were selected for validation by WB and immunocytochemistry. A similar approach was useful to monitor differentiation into oligodendrocytes; iTRAQ 4-plex quantitated a total of 3000 proteins. Functional bioinformatics analyses of these data are ongoing.

2.9 Human diabetes proteome project (HDPP)

(Chaired by Jean-Charles Sanchez, Geneva, Switzerland; Martin Kussman, Lausanne, Switzerland; Peter Bergsten, Uppsala, Sweden.)

The theme of the HDPP is "From Network Biology to Targets for Therapies and Prevention." A 10-year roadmap was defined in November 2012 to focus on beta-cells, islets, stem cells, and blood-based biomarkers with integration of all omics platforms. Sanchez et al. launched the portal www.hdpp.info with news, partners, the HDPP-1000 list, and information about diabetes and the broader HPP. In all, 1379 proteins were defined as of special interest in the context of diabetes, posted with the UniProt and neXtProt accession numbers and Peptide Atlas and Human Protein Atlas cross-references. This draft will evolve and serve as a starting point for sublists, such as the 25 Diagnostic-HDPP and the 118 plasma proteins and 449 islet proteins quantifiable by isobaric tagging. In all 5300 human islet proteins were identified and integrated into the Human Islet Database and into neXtProt. The first publication [10] appeared in Translational Proteomics, the voice of HDPP, the inaugural article of the journal.

Domitille Schvartz (Geneva) introduced SMARTER as an alternative to SRM for biomarker discovery of islet-specific proteins in plasma. >700 islet proteins were quantified in plasma, of which 449 were among the HDPP-1000 proteins of great interest for diabetes. Also, 81 proteins were quantified for the first time in plasma (not present in the most recent Plasma Proteome Database build) and were contributed to the HPP parts list via neXtProt.

Michael Snyder (Stanford, USA) described his Personal Omics Profiling approach, a longitudinal self-study with multiple omics platforms [11]. The aim is prediction and early recognition of disease risks. He outlined its application to patients with diabetes or prediabetes.

Peter Bergsten presented Beta-JUDO, the European Project on Beta-cell Function in Juvenile Diabetes and Obesity, which was integrated into the HDPP. The goal is to generate knowledge of cell dysfunction that might inform innovative therapies for Type 1 or Type 2 diabetes by halting destruction and/or facilitating recovery of functionally impaired metabolic tissues, particularly beta-cells. Palmitate-induced dysfunction is an experimental model.

Martin Kussman focused on signal transduction in beta-cell mitochondria, an essential organelle to understand insulin secretion. Phosphorylation and acetylation of mitochondrial proteins are under investigation.

Young Ah Goo (U. Maryland, USA) presented plans to determine proteomic profiles of individuals at high genetic risk for development of T1 diabetes and matched

controls and to integrate the data with parallel transcriptomics and metabolomics measurements. Preliminary data suggest signatures associated with T1D onset and progression. Salvatore Sechi (NIDDK, NIH, USA) outlined the range of NIDDK-funded research projects.

Finally, the rotating chairs named as deliverables for 2014 analytical results for the HDPP-1000 proteins in plasma by MS, integration of the human and rodent islet of Langerhans proteome databases and the human blood glycated proteome database into the HDPP website, and close coordination of HDPP with the B/D-HPP. Research projects will include mitochondrial role in beta-cell dysfunction, dysfunction of insulin-producing cell lines induced by high glucose, effects of antiplatelet therapy on cardiovascular diseases and diabetes, and advances in bioinformatics for network-based biology.

2.10 Cardiovascular proteome initiative

(Co-chaired by Peipei Ping, UCLA, Jenny van Eyk, Johns Hopkins, & Cathy Costello, Boston University, USA.)

The cardiovascular workshop was focused on overcoming technical and experimental challenges in the application of proteomics-based methods to the heart and vascular system. Peipei Ping opened the session with an overview of the role that proteomics can play in areas of cardiovascular disease. She outlined what are the large clinical questions that our field should or could address. Anthony Gramolini (University of Toronto, Canada) focused on his work on the difficult but important plasma membrane proteome. His seminal work compared this proteome across a number of heart and vascular system cell types, including primary mouse neonatal and human fetal ventricular cardiomyocytes, endothelium, and smooth muscle cells, as well as adult human coronary artery smooth muscle and endothelial cells, and human adult cardiac muscle-derived cardiomyocytes. Albert Sickmann (Leibniz-Institut für Analytik, Germany) discussed how to handle the really difficult and challenging cardiovascular proteomes. His laboratory is a world-leader in the systematic characterization of human platelets in arterial vascular disorders. Naoto Minamino (National Cerebral and Cardiovascular Center Research Institute, Japan) focused on challenges in the translation from discovery to biomarkers. He described work on dilated cardiomyopathy, seeking to understand the underlying pathogenic mechanism and generate ideas for targeted therapy. He also addressed the proteome analysis of disease stage classification of aortic aneurysms. Finally, Cathy Costello and Jenny Van Eyk discussed potential routes forward for the application of proteomics to cardiovascular disorders.

2.11 Initiative on model organism proteomes (iMOP)

(Chaired by Michael Hentgartner, Zurich, Switzerland.)

Three speakers contributed to the iMOP session, following an introduction by Sabine Schrimpf (University of Zurich). Marc Wilkins (University of New South Wales, Australia) addressed the methylproteome network of *Saccharomyces cerevisiae.* Paola Roncada (Italian Experimental Institute, Milan, Italy) reported on secretome protein profiling of the growth interaction between *Listeria monocytogenes* and *Lactobacillus lactis subsp. Lactis.* Brian Tate Weinert (University of Copenhagen, Denmark) spoke about how acetyl-phosphate links metabolism to global acetylation dynamics in *E. coli.*

An extensive discussion followed. The group has held excellent workshops at HUPO annual congresses for the past 4 years and is considered an integral part of the HPP due to the importance of comparative proteomics. iMOP would like to increase its visibility at future meetings, partly by placing iMOP presentations in

other scientific sessions or tracks, including sessions on clinical proteomics. They would like an option for researchers to select iMOP during the abstract submission procedure and would like their session organizers to have better access to all the submitted abstracts. They pointed out that the participants in the discussion represented the diverse model organism communities: bacteria, plants, animal models, food production, and bioinformatics. Chair rotation and deliverables were also discussed. We all congratulate Michael Hengartner on his election as president of the University of Zurich. His successor as iMOP chair is Emoke Bendixen of Aarhus University in Denmark.

2.12 Human glycoproteomics initiative (HGPI)

(Chaired by Hisashi Narimatsu, National Institute of Advanced Science & Technology, Japan.)

Glycoproteomics was a dominant program theme of the Yokohama 12th Annual HUPO Congress, reflecting the high activity in this research area in Japan, as well as globally. This 12th workshop of the HGPI comprised three major presentations, with about 80 participants, one of the larger workshops.

Hiromi Ito (Fukushima Medical University, Japan) reported on completed and ongoing pilot studies performed as activities of the HGPI. Two collaborative pilot studies on evaluation and standardization of technologies for glycomics used purified glycoproteins as common samples. It was concluded that the protein-based glycomic technologies were established and can provide consistent results across several methods. However, there are still challenges for standardization and consistency on the glycome analysis for the purified glycoproteins. This problem was magnified in a pilot study using complex mixtures such as cell extracts. The major reason is inconsistency in the preanalytical stage of extraction of glycoproteins, especially from small amounts of crude, complex samples. More studies are planned.

The HGPI Chair, Hisashi Narimatsu, proposed evolution of the HGPI to a new international collaboration under the HPP, the "Biology/Disease-driven Glycoproteome Project." The aim is to expand the current theme of glycomics to glycoproteomics, with technical development to merge glycomics and glycoproteomics; they are currently rather dissociated. The next aim is construction of a common database to capture large-scale datasets generated with various advanced technologies as those data emerge. The discussion recognized that this step is quite ambitious, possibly even premature. However, all agreed that technologies are needed to bring together peptide sequences and glycan structures. There was enthusiasm to plan the direction of such future activity of this initiative.

Finally, Katrine ter-Borch Gram Schjoldager (Copenhaven University, Denmark) introduced a powerful and novel technology to analyze the *O*-glycoproteome. The technology was applied to reveal target proteins of several specific *O*-glycosylation enzymes and polypeptide *N*-acetyl-galactosamine transferase (pp-GalNAcT), among many isoenzymes. This technology could contribute strongly to the proposed B/D-GPP.

2.13 Proteome analyzer MS initiative

(Chaired by Bruno Domon, Luxembourg and HUPO Industrial Advisory Board)

While genomics has progressed from Northern blots to microarrays to next-gen sequencing across a great many research groups, protein studies by nonspecialists in proteomics remain rooted in Western blots. Biologists need proteomics tools that are readily adapted to a wide variety of laboratory applications, robust in use, moderate in cost, and capable of high throughput analyses for clinical or

epidemiologic studies. An analogy was presented of automobiles in the 1920s–1930s (exclusive and cost-prohibitive, except for the wealthy) and the emergence in the late 1930s of the Citroen 2 CV for farmers and the Volkswagen Beetle for families.

This group is considering engineering requirements for Proteome Analyzer I, a replacement for the Western blot, and a Proteome Analyzer II, an analogy to RNA-Seq. Most common workflows today involve complicated chromatography and MS strategies, which severely limit sample throughput.

Desired features for a Western blot replacement are the following: reliable identification and quantitation of specific proteins; insight into PTMs were needed; well-developed, highly reliable assays with integrated quality control; dynamic range of 6–7 orders of magnitude; multiplexing capability to address 10s or 100s of targets per experiment; easy to use and learn to use; <1 h per sample; cost in the range of a Western blot; available in a reagent-rental model, avoiding capital expense.

Ideal attributes of a genomics-like, RNA-Seq equivalent Proteome Analyzer II would be comprehensive, quantitative, reproducible, fast, and sensitive. In more detail, we seek identification and (semi)-quantitation of every protein in a cell with, say, 50% sequence coverage, including PTMs; reproducible measurements of differences between samples of cells; standardized workflow, including sample preparation and data conversion; and wide accessibility for the biology community from technical and cost perspectives, with cost points such as $1/protein or $10 000 per cellular proteome. Biologists currently use RNA-Seq as a proxy for protein expression and abundance, but cannot demonstrate translation of transcripts to proteins, let alone critical PTMs and differential expression of splice variant proteins.

The next steps for this initiative include a HUPO web discussion forum, a survey targeting a wide range of biologists, close engagement with the instrument manufacturers through the HUPO Industrial Advisory Board, and a workshop at the next HUPO Congress, 6–8 October, 2014, in Madrid.

3   Goals of the B/D-HPP

The B/D-HPP has been created on the foundation of pre-existing HUPO initiatives on organ and biofluid proteomes (Plasma, Liver, Brain, Kidney-Urine, Cardiovascular, Stem Cells, Model Organisms), and the Proteomics Standards Initiative, the Human Antibody Initiative, and the Human GlycoProteomics Initiative. New initiatives on cancers, diabetes, and the eye are now quite active, while groups working on mitochondria, epigenetics, pediatric disorders, infectious diseases, and autoimmune disorders are in early stages of development.

The B/D-HPP aims to explore whether access to reliable and reproducible measurement techniques and quality-assured resources to support these techniques will facilitate research on a much larger fraction of the proteome by a broad range of life scientists, beyond those representing the core proteomics community. As described by Aebersold et al. [12], the HPP is pursuing a three-prong strategy by providing: (i) specific, validated mass spectrometric methods to quantify the abundance of all proteins within a focus area of biology, (ii) specific and selective protein affinity reagents for each of these proteins and their key isoforms, and (iii) the knowledge how to use these SRM assays and reagents in a format readily accessible for the scientific community and supported by annotations based on protein networks and pathways. In short, the B/D-HPP attempts to provide a spectrum of research tools for studying relevant proteins in specific areas of biology and specific diseases. Biological processes and disease mechanisms are expected to be analyzed increasingly in the full complexity of the living cell.
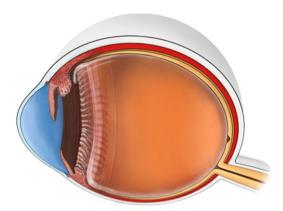
Figure 1. The EyeOME, the latest addition to the B/D-HPP, addresses eight compartments of the eye [7,8].

Toward this end, the B/D-HPP defined the following specific goals for its component initiatives:

(i) Select biological focus areas. As listed above, the areas might derive from self-organizing groups of scientists who study specific biological processes or diseases, who recognize the essential role of proteins and their isoforms, and who want to support the application of proteomics technologies to their field. Optimally the specific focus areas will be discussed in workshops at future HUPO meetings.

(ii) Generate target lists of proteins. Once the areas of biology are defined, the area experts should deliver a list of proteins known to be, or hypothesized to be, especially relevant for that field. For example, a biological theme on cellular signaling might comprise all human protein kinases and phosphatases, or at least those likely to be involved in the disease of interest. It is well understood that these lists are subject to change and adaptation. However, even early lists of proteins relevant for a particular disease or biological process and the associated resources to precisely and reliably measure these proteins would constitute significant progress toward the routine use of proteomics for the study of the respective disease or process.

(iii) Define and generate relevant assays and reagents for these targets. The collaborative team of biologists, physicians, and proteomics experts should then define the assays and reagents of most relevance to the problem, and generate a scientific plan to generate and characterize those assays and reagents.

(iv) Disseminate the knowledge and reagents. The impact of the B/D-HPP will depend on community uptake, and thus the B/D-HPP proposes that all MS and affinity based assays and reagents would be made publicly accessible, with no restriction on use, and full datasets and useful metadata would be made available via ProteomeXchange and interconnected databases.

(v) Introduce clinician-scientists to the developments in proteomics that would be valuable for their disease-oriented research. At the Yokohama Congress, Shiying Shao, MD, PhD, endocrinologist from Tongji Hospital, Huazhong University of Science and Technology, Wuhan, China; Kuo-Chen Wei, MD, chair of Neurosurgery and vice-chair Chang Gung Cancer Center, in Taiwan; and Leslie Chen PhD, Assistant Professor in the Medical Research Department, Chang Gung Memorial Hospital, Taiwan, received travel fellowships from HUPO to participate in the Congress and specifically with their respective B/D-HPP groups—Diabetes, Cancers, and Brain. Due to scheduling challenges, others who were accepted from Japan, Australia, and

China were unable to attend. This year the announcement of the fellowships was issued in March for the October 2014 HUPO Congress in Madrid.

The authors have declared no conflict of interest.

4   References

[1] Paik, Y. K., Omenn, G. S. (Eds.), From the genome to the parts list of the human proteome: Chromosome-centric Human Proteome Project (C-HPP), Part II. *J. Proteome Res.* 2014, *13*, 1–324.

[2] Legrain, P., Aebersold, R., Archakov, A., Bairoch, A. et al., The human proteome project: current state and future direction. *Mol. Cell. Proteomics* 2011, *10*, M111 009993.

[3] Paik, Y. K., Jeong, S. K., Omenn, G. S., Uhlen, M. et al., The Chromosome-Centric Human Proteome Project for cataloging proteins encoded in the genome. *Nat. Biotechnol.* 2012, *30*, 221–223.

[4] Omenn, G. S., Menon, R., Zhang, Y., Innovations in proteomic profiling of cancers: alternative splice variants as a new class of cancer biomarker candidates and bridging of proteomics with structural biology. *J. Proteomics* 2013, *90*, 28–37.

[5] Hood, L. E., Omenn, G. S., Moritz, R. L., Aebersold, R. et al., New and improved proteomics technologies for understanding complex biological systems: addressing a grand challenge in the life sciences. *Proteomics* 2012, *12*, 2773–2783.

[6] Lane, L., Bairoch, A., Beavis, R. C., Deutsch, E. W. et al., Metrics for the human proteome project 2013–2014 and strategies for finding missing proteins. *J. Proteome Res.* 2013, *13*, 15–20.

[7] Semba, R. D., Enghild, J. J., Venkatraman, V., Dyrlund, T. F. et al., The human eye proteome project: Perspectives on an emerging proteome. *Proteomics* 2013, *13*, 2500–2511.

[8] Omenn, G. S., The human eye proteome project. *Proteomics* 2013, *13*, 2375–2376.

[9] Farrah, T., Deutsch, E. W., Omenn, G. S., Sun, Z. et al., The state of the human proteome in 2013 as viewed through peptideatlas: comparing the kidney, urine, and plasma proteomes for the biology and disease-driven human proteome project. *J. Proteome Res.* 2013, *13*, 60–75.

[10] Topf, F., Schvartz, D., Gaudet, P., Priego-Capote, F. et al., The Human Diabetes Proteome Project (HDPP): From network biology to targets for therapies and prevention. *Transl. Proteomics* 2013, *1*, 3–11.

[11] Chen, R., Mias, G. I., Li-Pook-Than, J., Jiang, L. et al., Personal omics profiling reveals dynamic molecular and medical phenotypes. *Cell* 2012, *148*, 1293–1307.

[12] Aebersold, R., Bader, G. D., Edwards, A. M., van Eyk, J. E. et al., The biology/disease-driven human proteome project (B/D-HPP): enabling protein research for the life sciences community. *J. Proteome Res.* 2013, *12*, 23–27.