Report

The 2006 Human Liver Proteome Project (HLPP) Workshops

José M. Mato[1], *Fuchu He*[2], *Laura Beretta*[3]

[1]CIC bioGUNE, CIBER HEPAD, Parque Tecnológico de Bizkaia, Derio, Bizkaia, Spain; [2]Beijing Proteome Research Center, Beijing, China; [3]Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA

Corresponding Author: Laura Beretta

Fred Hutchinson Cancer Research Center (M5-A864)

1100 Fairview Avenue North

Seattle, Washington 98109

Voice: 206-667-7080

Fax: 206-667-2537

E-mail: lberetta@fhcrc.org

Key words: Human Liver Proteome Project, Human Proteome Organisation, Workshop report

Abstract

In 2006, scientists participating to the Human Liver Proteome Project (HLPP) launched by the Human Proteome Organisation (HUPO) convened on two occasions to present and discuss their progress. A workshop was held over two days in May in Bilbao, Spain, and a brief 3-hour meeting was held in October in conjunction with the 5th HUPO World Congress in Long Beach, California. Highlights included progress on the construction of the human normal liver proteome expression profile and of subcellular proteomes, establishment of a liver ORFeome bank and of a liver antibody bank, identifications of protein-protein interaction maps in the liver, application of a robust strategy for quantitative proteomics and the characterization of fatty liver diseases using mouse models.

The scientific objectives of HLPP are to characterize the proteome of the healthy human liver (expression profiling, cellular localization, post-translational modifications, protein-protein interactions) and to develop tools for its study (sample banking, ORFeome, antibodies). The information and tools that are generated by this international effort will be available to the scientific community with the hope that they will help understand, diagnose and treat liver diseases more effectively.

At the 2005 HLPP workshop held at the Airlie Center, USA, initial targets for HLPP were identified, as follows: 1- identification of 5,000 individual proteins in the normal human liver; 2- availability of antibodies against these 5,000 proteins; 3- determination of the presence of these 5,000 proteins in plasma. The aim of these targets is to generate sufficient knowledge of the normal liver and resources to allow the study of liver diseases and biomarker discovery. As learned during the 2006 workshops, major progress has been made toward these goals. The 2006 annual HLPP Workshop was held May 4-5 in Bilbao, Spain. The workshop was hosted by Dr. José Mato (CIC bioGUNE) and sponsored by the Banco Bilbao Vizcaya Argentaria (BBVA) Foundation. The meeting took place in a restored historic building in downtown Bilbao, the 19th-century home of BBVA: an elegant setting for a valuable exchange of ideas. The Fall 2006 HLPP Workshop was held October 29 at the Long Beach Convention Center, California, immediately prior to the HUPO 5th Annual World Congress. The workshop was hosted by Dr. Laura Beretta, Fred Hutchinson Cancer Research Center, current chair of the HLPP. This report summarizes the presentations and subsequent discussion.

Drs. Fuchu He, Xiaohong Qian and Wantao Ying from the Beijing Proteome Research Center are leading the effort on the HLPP subproject on expression profiling in China and presented in both workshops important progress in the construction of the human normal liver

proteome expression profile. The achievements presented during the workshops included: 1/ standardization of tissue acquisition, sample processing, protein extraction, protein/peptide separation and identification and data submission and management; 2/ the semi-integration of data from multiple platforms, the semi-quantitative analysis of shotgun data, the comparison between SDS-PAGE, 2-DE, shotgun and LC-shotgun, preliminary annotation of key physiological and pathological pathways in the liver and GO annotation. For the next phase, these groups plan on focusing on the profiling of subcellular organelles including quantitative analysis using high accuracy mass spectrometry; isolation of hepatocytes and other liver cell types and their profiling analysis; characterization of post-translational modifications of liver proteins; standardization for data quality control at both the peptide and the protein levels; and intra- and inter-organelle quantification of proteins. To date, the expression profiling of the human fetal liver has been finalized with 2,495 unique proteins identified and reported [1]. The data are also publicly available in the PRIDE database at the European Bioinformatics Institute (www.ebi.ac.uk/pride/startBrowse.do) by selecting the heading: "HUPO HLPP: Beijing Proteome Research Center" (accession numbers 1779 through 1853). The expression profiling of the first adult liver reference sample (referred to as the French liver sample) is also finalized with 4,998 unique proteins identified. The expression profiling of the second adult liver reference sample (referred to as the Chinese liver sample) is currently under analysis with 6,788 proteins identified to date.

Dr. Laura Beretta, Fred Hutchinson Cancer Research Center in Seattle, presented the analysis performed by her group on the profiling of the adult liver reference sample. The work was described in three sections: methods used for protein extraction, separation and identification; comparison between the transcriptomic and proteomic profiles of the reference

sample and comparisons with the human fetal liver and the human plasma proteome. Intact proteins extracted from the reference sample were extensively separated using a three-dimensional approach (a combination of 2D HPLC and SDS-PAGE). Following in-gel digestion, the resulting peptides were analyzed by LC-MS/MS. The generated data were submitted to CPAS (Computational Proteomics Analysis System) developed at the Fred Hutchinson Cancer Research Center, using the X!Tandem search engine and ProteinProphet. Using a cut-off of 0.9 ProteinProphet score, 8149 individual proteins and 472 protein groups were identified by at least 3 peptides. A relative protein abundance score was calculated for each identified protein. Annotation of the identified proteins with properties such as MW, Gene Ontology classification and post-translational modifications is underway. In parallel, the transcriptomic profile of the same sample was established using Affymetrix chips containing 54,675 probe sets. This analysis identified 10,982 genes expressed in the reference liver sample. Comparative analysis of the transcriptomic and the proteomic data sets was presented. The results indicated that protein products for over 50% of the genes expressed were identified using this proteomic approach, including protein products of low-abundance genes. Comparison with the human fetal liver proteome recently published [1] identified 945 (out of 1807) proteins commonly expressed in both fetal and adult livers. Comparison with the HPPP human plasma proteome [2] identified 792 (out of 3017) proteins detected in both plasma and liver tissue. Future directions for Dr. Beretta's group are to refine the criteria for selection of candidate disease markers, to prioritize proteins for production of antibodies, to further compare liver and plasma proteomes, to evaluate and validate methods for label-free protein quantification.

Drs. John Bergeron and Alexander Bell from McGill University, Montreal presented their quantitative proteomic study of detoxification in the hepatic endoplasmic reticulum (ER). ER

organelles in the liver play an essential role not only in protein synthesis and folding but also in liver detoxification. The proteomes of highly enriched organellar ER rough and smooth membranes and Golgi fractions from rat liver were characterized. Organellar proteins were resolved by 1D-SDS PAGE, the gels were sliced and the proteins were in-gel digested with trypsin. Peptides were extracted and subjected to LC-QTOF MS analysis. Identified proteins were annotated into 23 functional categories and the detoxification category was further curated into 13 sub-functions. The Cytochrome P450, Glucuronosyltransferase and Carboxylesterase families of proteins, with 34, 11 and 9 isoforms identified, respectively, represented about 67% of the total ER protein as determined by redundant peptide counting although these highly sequence related families represented only 54 of the 122 detoxification proteins. Peptide heatmaps, revealing frequency, identification and location of peptides (assigned at >95% confidence) in these families were emphasized. KDEL- and KKxx/KxKxx-related trafficking motifs were identified in the Carboxylesterase and Glucuronosyltransferase proteins whereas several Cytochrome P450s revealed an exposed DxD motif (as deduced from their x-ray crystallography) that may be involved in trafficking. Peptide sequencing information was presented that distinguishes Cytochrome P450 allelic isoforms and database DNA sequencing errors were identified. Proteomics readily distinguished multiple members of the Cytochrome P450, Glucuronosyltransferase and Carboxylesterase families of proteins, distinguished allelic isoforms differing by a single amino acid and corrected DNA-related sequencing errors in the protein database.

Dr. Ying Jiang from the Beijing Proteome Research Center has been developing sample preparation strategies for the analysis of subcellular proteomes using livers from C57BL/6J strain mice. This study has been recently published [3]. Dr. Siqi Liu from the Beijing Genomics

Institute presented data mining analysis on cellular organelles (specifically on mitochondria with over 600 proteins involved in all metabolic pathways in this organelle) and validation by other proteomic approaches.

Dr. Pengyuan Yang from the Institute of Biomedical Sciences and the Fudan University Proteomics Research Center in Shanghai, reported the identification of 15,765 genes expressed in the human liver, of which 13,211 correspond to protein-coding genes. Approximately 5,500 open reading frames (ORFs) of genes expressed in the human liver have been cloned. Of these 5,500 ORFs, approximately 4,000 have been constructed into vectors for use in yeast two-hybrid assays. The majority of these ORF clones are in the range of 500-1500 base pairs in length. A web interface has been created in order to access the current ORF bank data set (http://202.127.18.238/hlpp/).

Dr. Xiaoming Yang from the Beijing Proteome Research Center presented the undergoing effort to establish a liver protein-protein interaction (PPI) map. To date, 1249 baits have been used for PPI studies. Approximately 30 baits per day can be processed in yeast two-hybrid assays. Antibody chips containing over 1000 probes are also used for this subproject. The resulting data are further validated by co-immunoprecipitation or by affinity purification. The largest network identified to date includes 1035 PPIs between 810 proteins. A primary map was presented and the functions of specific interactions, such as those involved in the Keap1-Nrf2-ARE signaling pathway, were examined in detail.

Dr. Qihong Sun from the Beijing Proteome Research Center is working towards the development of a liver antibody bank. Dr. Sun and colleagues have reported a strategy to establish the antibody bank of murine monoclonal antibodies for human liver proteins using fractionated native liver proteins as immunogens. Synthesized peptides from liver-specific

proteins and purified recombinant protein antigens have been used as well. To date, 1907 hybridoma cell lines have been established. Of these cell lines, 1400 have been characterized and produced monoclonal antibodies for approximately 150 different proteins. Rabbit polyclonal antibodies have been produced for 56 different proteins. Approximately 44% of the hybridoma cell lines have been assigned to immune response functions and approximately 41% of the monoclonal antibodies have been assigned to metabolic functions. The produced antibodies are currently used for depletion of highly abundant proteins expressed in liver tissues or plasma, for immunohistochemistry and for isolation of protein complexes.

Dr. José Mato from CIC bioGUNE in Bilbao, Spain, introduced different approaches to study the early stages of steatosis and steatohepatitis using genomics, metabolomics and proteomics. Dr. Mato introduced mouse models of non-alcoholic fatty liver disease (NAFLD) leading to hepatocellular carcinoma. These models target the metabolism of methionine and include MAT1A knockout mice. The MAT1A KO mice spontaneously develop steatohepatitis and later hepatocellular carcinoma. Another key player in the metabolism of methionine is GNMT. Interestingly, a GNMT polymorphism has been reported to be associated with HCC in humans. GNMT KO mice have been generated. These mice spontaneously develop steatohepatitis and fibrosis. Liver disease progression in older mice is under investigation. The integrative gene expression profiling of liver biopsies of NASH patients with liver samples of MAT1A KO mice identified genes and gene pathways that are associated with NASH. Based on this information, Dr. Mato's group generated a novel mouse model and, remarkably, the mice spontaneously develop steatohepatitis. Proteomic analysis of liver tissues from patients with steatosis or with NASH was performed using 2D gel electrophoresis (DIGE) to identify proteins differentially-expressed in NASH and in steatosis. Proteomics and metabolomics studies of sera

from controls, patients with steatosis and patients with NASH were also performed. The generated data were presented.

During the workshop in Bilbao, a session was organized around interactions with other HUPO initiatives. Drs. Rolf Apweiler and Henning Hermjakob from the European Bioinformatics Institute, reported on the HUPO Proteomics Standards Initiative (PSI). The PSI was founded in 2002 to develop data representation formats in proteomics. Released products so far are community standards for the representation of molecular interactions and mass spectra. The joint workshop of the PSI and the HUPO Publications Committee in San Francisco (April 2006) attracted an attendance of approximately 100 bioinformaticians, experimentalists, and representatives from journals and standardization initiatives from related domains. Representatives of the PSI are currently collaborating with the HLPP to generate a representation of the HLPP data in PSI format. They emphasized the need for comparing the data collected by the individual HUPO Initiatives, including HLPP, the brain proteome initiative (HBPP) and the plasma proteome initiative (HPPP). Dr. Gil Omenn from the University of Michigan, chair of the HPPP, urged that comparisons of liver and plasma profiles across the two HUPO initiatives and in specimens from the same people or animals studied in the HLPP should be a standard feature. The HPPP will share its protocols and would be pleased to participate in such cross-analyses.

During this workshop in Bilbao, special guests were invited to discuss potential collaborations with HLPP. Dr. Ron Appel from the Swiss Institute of Bioinformatics at the University of Geneva introduced activities on proteome imaging (Melanie / ImageMaster and MSight software for 2-D gel, respectively LC-MS data analysis) and protein identification and characterization (Aldente PMF identification) to identify mixtures of proteins, Popitam for

MS/MS open-search identification, and GeneBio's Phenyx comprehensive MS/MS identification platform. He also showed the importance for correct validation of identification results. Finally, swissPIT (Protein Identification Toolbox), the new pipeline designed for knowledge extraction from MS data developed within SwissBioGrid, was presented. Dr. Juan-Pablo Albar from the Centro Nacional de Biotecnología, Bilbao introduced the ProteoRed initiative. ProteoRed (Spanish National Network of Proteomics Facilities; www.proteored.org) aims to coordinate and integrate the activities of the proteomics facilities and services in Spain and to evaluate the quality of the services offered. ProteoRed was formed in June 2005 and integrates 15 laboratories and 5 associated laboratories. Six working groups (WG) have been created to reach ProteoRed objectives: standardization of protocols and technologies, collection and handling of samples, bioinformatics support, functional organization of proteomics facilities and price policy, education, training and diffusion, and international networking.

Additional information on HLPP structure and activities can be obtained at the HUPO web page and in the newly launched HLPP Newsletters (http://www.hupo.org/research/hlpp/).

Acknowledgements

We offer special thanks to Paul Farley, Miryam Asunción and Junjie Zheng for their help in the preparation of the workshops.

References

[1] Ying, W., Jiang, Y., Guo, L., Hao, Y., *et al.*, *Mol Cell Proteomics* 2006, *5*, 1703-1707.

[2] Omenn, G. S., States, D. J., Adamski, M., Blackwell, T. W., *et al.*, *Proteomics* 2005, *5*, 3226-3245.

[3] Song, Y., Hao, Y., Sun, A., Li, T., *et al.*, *Proteomics* 2006, *6*, 5269-5277.