



Chromosome-centric Human Proteome Project

NEWSLETTER

No. 4
April 1, 2015



Milan, a Place for 12th C-HPP Workshop, 2015 EuPA Congress and World EXPO

INSIDE

- Editorial P1
- New C-HPP Leadership P5
- Highlights of C-HPP Consortium Activities in 2014 P6
- Individual Teams' Activity Reports P23
- JPR Call for Papers P32
- Wiki updates P33
- Data Submission by C-HPP Teams P34
- 12th C-HPP Workshop in Milan, Italy P35
- Vancouver HPP Meeting Announcement P36
- Obituary p37
- Appendix P38

Editorial : Contribution of the Human Protein Atlas to HPP

Dr. Tove Alm

Project manager Antibodypedia
Science for Life Laboratory (KTH), Sweden



Dr. Emma Lundberg

Director of Subcellular Protein Atlas
Science for Life Laboratory (KTH), Sweden



Dr. Cecilia Lindskog Bergström

Group Leader of the Tissue Protein Atlas
Uppsala University, Sweden



What is the Human Protein Atlas?

The Human Protein Atlas (HPA) is a large-scale project that generates well-validated antibodies towards the human proteome. These antibodies are used for protein profiling in 44 different tissues, 46 cell lines, as well as in 20 of the most common types of cancer. The analysis is complemented with transcriptomics data, based on RNA-seq from 32 different tissues. The HPA also provides subcellular location data from immunofluorescence-based confocal microscopy in cell lines. All data and images generated within the HPA are publicly available through Protein Atlas database, www.proteinatlas.org. The HPA has four sub-atlases: The Tissue Atlas, The Subcellular Atlas, The Cell Line Atlas, and The Cancer Atlas, that in total contains over 13 million manually annotated images.

How is the HPA supporting the C-HPP?

The Human Antibody Initiative (HAI, www.hupo.org/initiatives/human-antibody-initiative) was initiated by HUPO in 2005 with the aim to promote and facilitate the use of antibodies in proteomic research. Two resources are part of HAI; the HPA (www.proteinatlas.org) (1,2) and Antibodypedia (www.antibodypedia.com) (3). The knowledge and data generated in the projects are freely available in public databases and supports the HPP (C-HPP and B/D-HPP) as knowledgebase resources.

HPA recently presented the “tissue-based map of the human proteome”, what did it show?

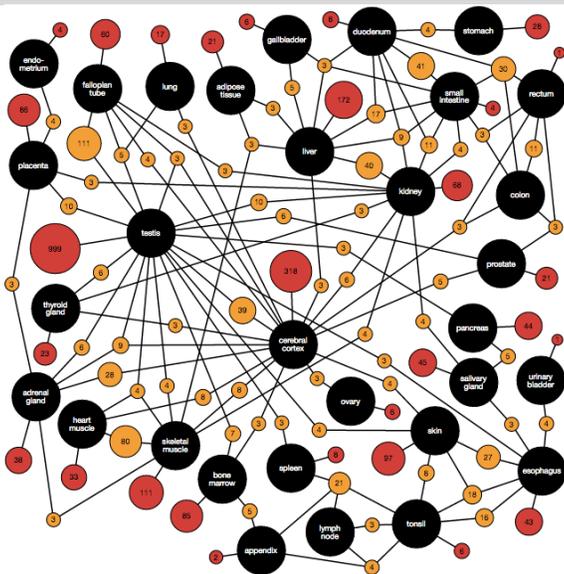


Figure 1: Network plot of tissue enriched (red) and group enriched (orange) genes connected to respective tissue. Click here for an interactive version.

Recently a genome-wide analysis of the human proteome was published to describe a map of the human tissue proteome based on an integrated omics approach which involved quantitative transcriptomics at tissue level, combined with immunohistochemistry to obtain spatial localization of proteins at a single-cell level (1). Almost half of all human genes appear to be housekeeping genes with detectable levels of transcripts in all analyzed

tissues, while approximately 34% show some level of elevated expression in one of the analyzed tissues. Testis and brain show the highest number of tissue specific genes (Fig 1). We further used this approach to explore the human secretome, the membrane proteome, the druggable proteome, the cancer proteome and the metabolic functions in 32 different tissues and organs. This data is integrated into a novel interactive version (v13) of the Protein Atlas database that allows exploration of individual proteins, as well as navigation of global expression patterns, in all major tissues and organs in the human body.

What is the most challenging aspect to complete the Human Protein Atlas?

The major challenge is to find the yet missing proteins. As much as 10 percent of the human protein-coding genes were not detected in any of the 32 analyzed tissues at the RNA level. However, many of these are expected to be expressed in more specific regions of the human body not covered by current tissues in the HPA, such as different regions of brain, retina, taste buds, sebaceous gland, or during short developmental stages. Another major challenge is the interpretation and validation of the obtained staining patterns for previously uncharacterized proteins. There are in many cases no available literature suggesting tissue distribution or subcellular location of the expression; in some cases it is difficult to identify the true protein expression and distinguish specific binding from off-target binding.

What is the strength of antibody-based protein profiling?

It is the ability to study proteins in the context of surrounding cells with retained spatial information. Human tissues are heterogeneous mixtures of different cell types and antibody-based profiling gives the

opportunity to analyze the protein expression with single-cell resolution (Fig 2A). Confocal microscopy gives the opportunity to study the subcellular distribution of the protein and resolve cell-to-cell variations (Fig 2B).

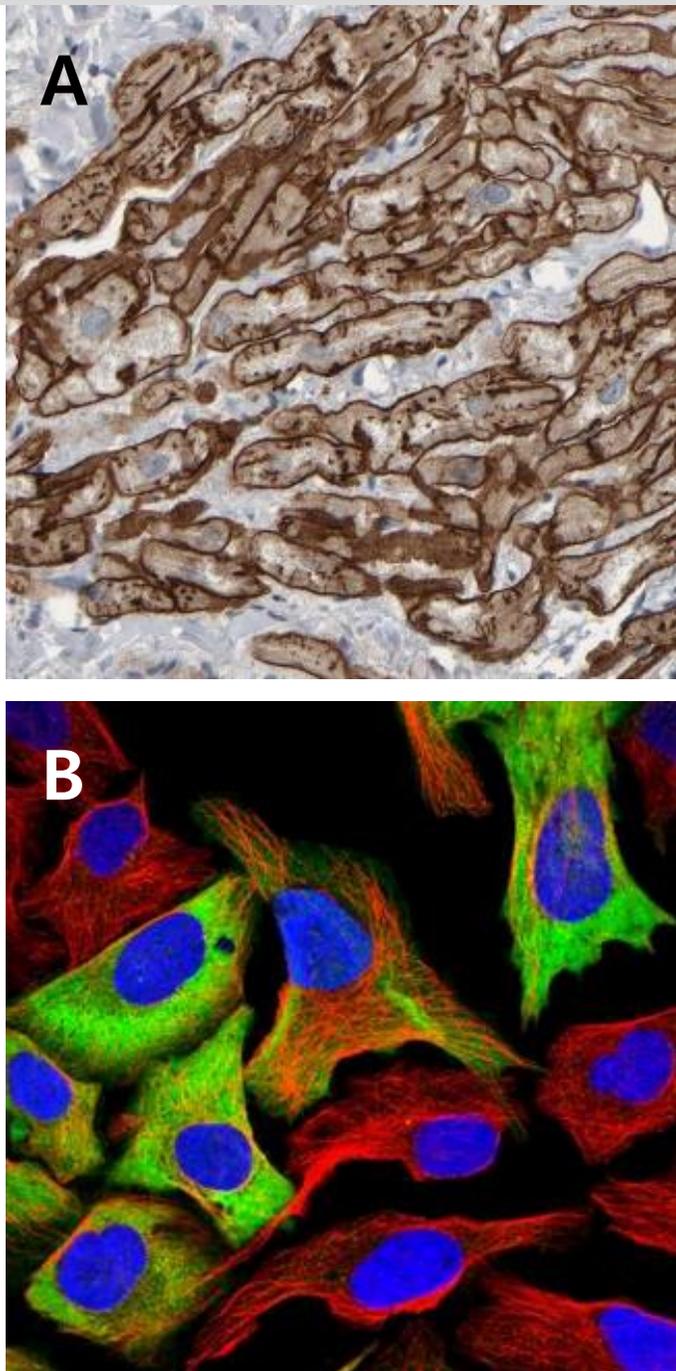


Figure 2. A) Immunohistochemical staining of Purkinje cells in cerebellum, stained with antibody towards protein CLEC2L (brown). B) Immunofluorescence staining of the protein PKM (green) that shows selective expression in the cytoplasm of a subset of U-2 OS cells.

How can data and reagents from the Human Protein Atlas be used as a complement to mass spectrometry studies?

While mass spectrometry provides the standard for quantifying a certain set of proteins in a sample, antibody-based protein profiling provides spatial information on protein expression patterns. Data from the HPA can thus be used to complement most MS based studies. It can both serve as an independent source of protein expression data and provide spatial information on protein distribution on a cell or subcellular level. The antigens generated in the HPA can also be used as spike-in reagents for quantitative mass spectrometry. The antigens, called Protein Epitope Signature Tags (PrEST), are 50-100 amino acids long fused to an Albumin Binding Tag (ABP). Thanks to the ABP tag the PrESTs are ideal for use as isotope labelled quantitative standards for absolute quantification of the corresponding endogenous protein in Mass Spectrometry (5).

How are the HPA antibodies quality assured?

Each antibody is tested according to HPA-standard quality assurance procedures, i.e. on Western blots, Protein Arrays, a specially designed tissue microarray, and on cells using immunofluorescence. The staining pattern is compared with previously published gene/protein characterization data and internally generated RNA-seq data. Recently gene silencing with siRNA has been introduced to the HPA quality assurance procedure (4). Currently nearly 300 antibodies have been validated in this manner.

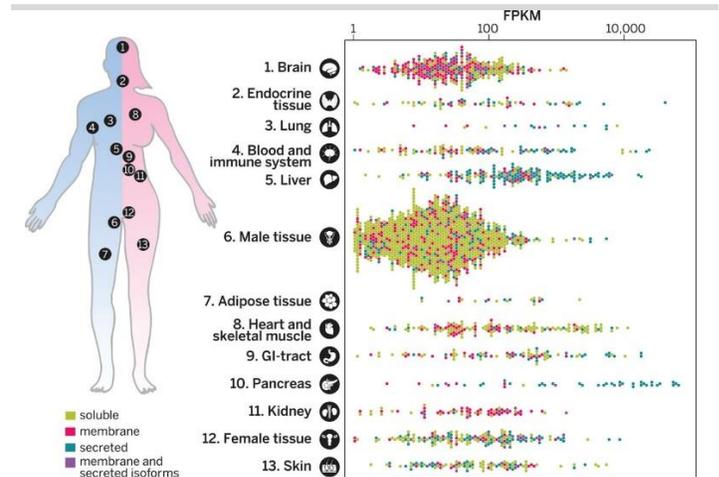
What important aspects should you consider when working with antibodies?

First of all, gather as much information as possible about the antibody, before purchase. You may use a database to

help you find the right antibody, e.g. Antibodypedia (www.antibodypedia.com), where validated and scored antibodies are listed by commercial or academic providers. Then, when you have the antibody at your bench, make sure to validate it in the assay you want to use it in, as the performance of most antibodies is context dependent.

Now that you have established the map of the human tissue-proteome, what is next?

The aim is to publish a rodent brain atlas this year, and to follow with a complete map of the human subcellular-proteome, the human cancer proteome, and also a human plasma proteome. The Protein Atlas is a continuous work; new data will be validated and curated and the results continuously added to the database.



The human tissue-enriched proteins.

All tissue-enriched proteins are shown for 13 representative tissues or groups of tissues, stratified according to their predicted subcellular localization. Enriched proteins are mainly intracellular in testis, mainly membrane bound in brain and kidney, and mainly secreted in pancreas and liver.

Uhlén M *et al*, Science. 2015 Jan 23;347(6220):1260419

Tove Alm
Emma Lundberg
Cecilia Lindskog Bergström

References

1. Uhlén M, Fagerberg L, Hallström BM, et al. Proteomics. Tissue-based map of the human proteome. Science. 2015;347(6220):1260419.
2. Uhlen M, Oksvold P, Fagerberg L, et al. Towards a knowledge-based Human Protein Atlas. Nat Biotechnol. 2010;28(12):1248-50.
3. Alm T, Von feilitzen K, Lundberg E, Sivertsson Å, Uhlén M. A chromosome-centric analysis of antibodies directed toward the human proteome using Antibodypedia. J Proteome Res. 2014;13(3):1669-76.
4. Stadler C, Hjelmare M, Neumann B, et al. Systematic validation of antibody binding and protein subcellular localization using siRNA and confocal microscopy. J Proteomics. 2012;75(7):2236-51.
5. Edfors F, Boström T, Forsström B, et al. Immunoproteomics using polyclonal antibodies and stable isotope-labeled affinity-purified recombinant proteins. Mol Cell Proteomics. 2014;13(6):1611-24.

New C-HPP Leadership

C-HPP Executive Committee (EC) Composition

During 2014, a few changes were made in our C-HPP leadership composition. As György Marko-Varga, co-chair (2012–2014), took the HUPO secretary general job as of January 1, 2015, members of the C-HPP Principal Investigator Council (PIC) recently elected Lydie Lane at SIB, Geneva, Switzerland, as a new co-chair representing the EU region. Two newly elected EC members also joined the team: Chris Overall (America) at UBC, Canada Vancouver, and Fernando Corrales (Europe) at CIMA, University of Navarra, Pamplona, Spain. The current EC composition is listed as follows:

						
Young-Ki Paik	William S. Hancock	Lydie Lane	Peter Horvatovich	Chris Overall	Fuchu He	Fernando Corrales
Chair	Co-Chair	Co-Chair	Secretary General	MAL	MAL	MAL

Position	Name	Affiliation	Term Ends	Remarks
Chair	Young-Ki Paik (Asia Oceania)	Yonsei Univ., Seoul, Korea	Dec 31, 2016	PI, Chr 13
Co-Chairs	William S. Hancock (America)	Northeastern Univ., Boston, MA, US	Dec 31, 2015	PI, Chr 17
	Lydie Lane (Europe)	SIB, Univ. of Geneva, Switzerland	Dec 31, 2017	PI, Chr 2
Secretary General	Peter Horvatovich (Europe)	Univ. Groningen, Groningen, Netherlands	Dec 31, 2016	Chr 5 Group WiKi Manager
Members -at-Large	Fuchu He (Asia Oceania)	Beijing Proteome Research Center (BPRC), China	Dec 31, 2015	PI, Chr 1
	Fernando Corrales (Europe)	CIMA, University of Navarra, Spain	Dec 31, 2016	PI, Chr 16*
	Chris Overall (America)	UBC, Vancouver, Canada.	Dec 31, 2017	Chr 6 Group

*filling up the remaining term left by the late Juan Pablo Albar.

Highlights of C-HPP Consortium Activities in 2014

The following is a summary of C-HPP activities during 2014. Thanks to all of those colleagues who supported and encouraged our project as it has moved forward.

1. JPR Special Issue was published: Part II, January 5, 2014 (total 35 papers)
2. The 9th C-HPP Busan Workshop was successfully carried out (with 20 teams, >60 attendees)
3. Distribution of C-HPP Member Certificates (metal plaques) to all authorized, established PIs in Busan (3/26/2014)
4. An inaugural workshop on the 'Missing Proteins' was successfully carried out in Sydney with great enthusiasm (7/30/2014); it was hosted by Mark Baker (chromosome 7 PI) and his colleagues in the Australasian community
5. The 10th C-HPP Bangkok Workshop was carried out as well; Visith Thongboonkerd and his colleagues did a superb job in this workshop (8/9/2014), where special metal souvenirs (with each PI's name) were presented
6. The 11th C-HPP Workshops were held in Madrid during the Madrid HUPO Congress (10/5–8/2014); the Symposium on New Technology, poster session, PIC meeting, and plenary session were also carried out
7. A joint workshop of the HPP (B/D-HPP and C-HPP) was held in Segovia with more than 100 participants; local hosts, Fernando Corrales and his chromosome 16 team, arranged all the logistics for the welcome party; many interesting topics on joint HPP efforts were discussed with very fruitful conclusions
8. Newsletter (No. 3) was published (3/2/2014); it can be downloaded at either www.c-hpp.org or wiki site <http://c-hpp.webhosting.rug.nl/tiki-index.php?page=C-HPP%20Newsletters>

A Sketch of the Annual C-HPP Council Meeting in Madrid

The HUPO Congress in Madrid began on October 5, 2014, where more than 1600 international participants gathered together to enjoy a full 4 days of proteomics research (see pictures).



At the opening session, Fernando Corrales (left), Chair of the Madrid Congress Committee declared an opening of the meeting. Mathias Uhlen, Sweden (middle), and Fuchu He (right), China presented the plenary lectures on the major accomplishments of the HUPO HPP

The C-HPP Principal Investigator Council Meeting (PIC) was held in Madrid on the morning of October 8, 2014, where council members made a few key decisions on the terms of EC members, which was followed by an open discussion on several pending issues regarding the consortium's future directions. The agenda was presented to the PIC, which was previously selected at the C-HPP EC meeting held on October 6. In his opening remarks, Prof. Young-Ki Paik, Yonsei University, Seoul, Korea, chair, eulogized Dr. Juan Pablo Albar, former C-HPP EC member and co-chair of the HUPO Congress Organizing Committee, who died suddenly in July 2014.



C-HPP PIC meeting on Oct. 8 (top) and EC meeting on Oct. 6 (bottom) during HUPO Congress in Madrid, Spain.

In Part 1, Fernando Corrales of the University of Navarra, Pamplona, Spain, provided members with some updates on the chromosome 16 project with respect to its new structure and research direction (see “Minutes of 2014 PIC meeting” at <http://c-hpp.webhosting.rug.nl/tiki-index.php?page=General%20meeting>). Pengyuan Yang of Fudan University, Shanghai, China, highlighted the recent progress of the projects on chromosomes 1, 8, and 20, which have been well supported by the Chinese government (see details on the Chinese Grant News in the “National Teams’ activity” section). He emphasized the endeavors of the high-tech programs’ support of the joint proteomics projects for “Big Data” analysis and studies on missing protein identification with biological research and epigenetics using MRM technology. Lastly, on behalf of Joshua Labaer, Arizona State University, USA (chromosome

10), Manuel Fuentes briefly presented the current state of the human cDNA expression system, which can be used for SRM assays and various biological validation studies of missing proteins in the C-HPP. This system will also be useful for identifying the right human cells and tissues where missing proteins may be enriched. This resource will be shared with the C-HPP teams.

In Part 2, Bill Hancock, co-chair, led the main discussion session, where adjustments of C-HPP EC members’ terms and related topics were reviewed and discussed with a few important decisions. (1) Officers’ terms have been harmonized with each calendar year (Jan. 1–Dec. 31), effective immediately. (2) The resignation of György Marko-Varga was approved, leaving the vacant position of co-chair in the EC membership. (3) A proposal for a 1-year extension of one of the co-chairs’ terms was also approved to stabilize the organization, allowing Young-Ki Paik to serve as chair until the end of 2016. This adjustment also makes each co-chair’s term end at the end of different years (as opposed to their terms all ending at the same time; see EC composition table).

In Part 3, Young-Ki Paik led the election process of new EC members via the voting of the PIC. As a result, Fernando Corrales and Chris Overall were unanimously elected as new EC members. All PIC members thanked Carol and György for their dedicated service as the inaugural EC members during the past 3 years (2012–2014). In the middle of January 2015, PIC members also elected Lydie Lane as co-chair, representing the EU region. Lydie Lane filled the position that György Marko-Varga vacated at the end of 2014. The roles and terms of all EC members are listed in the table (see above section).

In Part 4, there was an open discussion about how to further stimulate C-HPP activities, which included an evaluation of each team’s performance by SSAB members (based on the publications, attendance at major C-HPP meetings, and responsiveness to C-HPP activities), more

support from the chromosome zero group (e.g., resources pillar of the HPP), the high quality of all bioinformatics datasets, and active interactions between C-HPP teams and B/D teams. There were many good suggestions and action items from this session (see the minutes of the PIC at Wiki site <http://c-hpp.webhosting.rug.nl/tiki-index.php>).

In Part 5, a few plans for 2015 were presented and discussed by PIC members: (1) a Pre-Milan Workshop and

a Milan C-HPP Showcase as part of World EXPO 2015 (June 23, 2015) (to be hosted by Paola Roncada and Andrea Urbani); (2) a JPR special issue to be published on September 2015; and (3) the 13th C-HPP Workshop at the Vancouver HUPO Congress, September 27–30, 2015. At the end of the PIC meeting, Young-Ki Paik thanked all PIC members and their associates for the successful meeting with fruitful outcomes. The meeting was adjourned at 9:15 am.

Summary of 11th C-HPP Workshop in Madrid, Spain

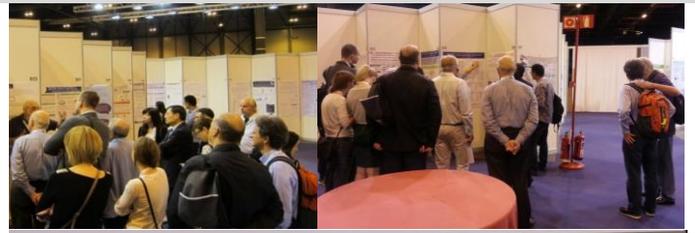
This report covers C-HPP scientific activities during the Madrid Congress (Oct. 5–8, 2014) in Spain, where two meetings were jointly hosted by both the C-HPP and B/D-HPP (e.g., General Investigators meeting on Sunday, Oct. 5, and the plenary session on Wed., Oct. 8). On Sunday morning, October 5, all HPP groups (chaired by Gil Omenn, University of Michigan, Ann Arbor, USA) participated in the “HPP General Investigators meeting” from 10:00–16:00, where newly emergent issues were discussed extensively with some slide presentations. Following a brief overview of the key agenda, including such topics as progress, metrics, and publications by Gil Omenn, the chair of each consortium gave a brief summary of the highlights of the activities of the B/D-HPP (by Ruedi Aebersold) and the C-HPP (by Young-Ki Paik). In the discussion sections at the working lunch, HPP colleagues exchanged their views on the bioinformatics pillar (Amos Bairoch, Lydie Lane, Eric Deutsch), the activities of the Ab and MS resource pillars (Emma Lundberg, Bruno Domon), the large datasets available from several sources, including two Nature papers on the HPP (Pandey and Kuster Labs), PeptideAtlas, PXD, and neXtProt. Special attention was paid to the bioinformatics approaches across the interconsortium teams. A few SSAB

members also provided their comments on the direction of the HPP (Mike Snyder, Cathy Costello, John Yates). Key issues included HPP metrics based on the PeptideAtlas released in August 2014, neXtProt released in October 2014, and the basis of HPP papers to be submitted in the coming months. In addition, the participants touched on some strategies for the formation of a working group to address the false-discovery rate (FDR) problem, rigorous quality thresholds, and the approaches for the identification of missing proteins. In a town hall-type meeting, both the B/D-HPP and C-HPP groups tried to find a cooperative way to enhance the data quality and communication on common interests (e.g., disease biology, missing proteins, technology development, etc.).

On Monday, October 6, the C-HPP consortium held a scientific session on “New Technology” led by Prof. Bill Hancock, co-chair. In this session, leaders from different working groups presented their views on the SRM measurements of missing proteins (Christoph Borchers, Rob Moritz, Ulrike Yassene), integration of RNA-seq data into proteomics (Siqi Liu, György Marko-Varga, Ravi Sirdeshmukh), ASV matters (Gil Omenn, Akhilesh Pandey), and the integrated HPP browsers (Dong Li and Ed Nice).

Stimulatory discussions followed each presentation (visit Wiki for details).

On Tuesday, October 7, a poster session was held in the exhibition hall, which was chaired by Prof. György Marko-Varga, co-chair (picture). There were nearly 50 posters presented by the C-HPP group, and Young Investigators Awards were given to the outstanding presenters: Seul-Ki Jeong (Yonsei Proteome Research Center, Seoul, Korea), Heeyoun Hwang (Korea Basic Science Institute, Chungbuk, Korea), Hongdong Li (Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA), Ekaterina Poverennaya (Institute Of Biomedical Chemistry, RAMS), Carmen González (Centro Nacional De Biotecnología [CNB-CSIC], Madrid - España), Alessio Soggiu (Department of Veterinary Science And Public Health, University Of Milan), and Hsin-Yi Wu (Academia Sinica, Taipei, Taiwan). Congratulations to all of the awardees.



(top) C-HPP poster session on Oct. 7 (bottom) The C-HPP HUPO Young Investigators' Awardees: from left – Pierre Legrain (HUPO president), Alessio Soggiu, Ekaterina Poverennaya, Carmen González, Hsin-Yi Wu, Heeyoun Hwang, Seul-Ki Jeong, Gil Omenn (on behalf of Hongdong Li) and György Marko-Varga (Co-chair of C-HPP)

On Wednesday morning, October 8, the PIC meeting was held from 8:00–9:20 AM. Details on the discussion topics are summarized in the "PIC meeting report" session.



On Wed afternoon, Oct 8, 2014, HPP Plenary session (Co-Chaired by Gil Omenn and Young-Ki Paik) was held in the main auditorium where the leaders of HPP and participants discussed and shared their views on the major milestones, emerging issues and future directions of HUPO HPP.

Summary of the Joint C-HPP-B/D-HPP Workshop in Segovia, Spain

Time and Venue: October 9, 2014, Escuela Universitaria de Magisterio de la Universidad de Valladolid, Campus María Zambrano, Plaza Colmenares nº 1, Segovia, Spain



This joint workshop was designed to stimulate various ongoing efforts under the umbrella of the HUPO HPP, which is composed of two consortia (C-HPP and B/D-HPP) and the resources and technology pillars. Aided by the bioinformatics developments, the HPP community has focused on a few areas, including but not limited to (1) proteogenomics data integration to reveal the effects of genetic variability on the proteome; (2) development of a new generation of antibodies for protein trace analysis to explore subcellular localization; and (3) application of new technologies to the identification of missing proteins in biological samples. The recent publication of two large datasets for the HPP in *Nature* by the Akhilesh Pandey (Johns Hopkins University School of Medicine) and Bernhard Küster (Technische Universität München) labs showed the importance of data management and quality assessment for the proteomics community. This workshop also had the common goal of bringing together the applied proteomics group at the B/D-HPP and the technology-oriented consortium of the C-HPP. It should be noted here that Fernando Corrales and his colleagues in Madrid made a great effort to make this meeting successful and fruitful. This meeting was jointly organized by the leaders of the HPP and PROTEORED: Fernando Corrales, Concha Gil, Madrid, Gil Omenn, Tadashi

Yamamoto, Young-Ki Paik, Bill Hancock, Ruedi Aebersold, and Jennifer van Eyk.

Session 1: Common Reagents and Resources for HPP Networks

Ulrika Qundos from SciLifeLab (Sweden) provided a comprehensive presentation on both a tissue-based map of the human proteome explored by the Human Protein Atlas (HPA) and the cross-validation of HPA antibody data with MS measurements, RNA-seq, and other publicly available data. HPA is now providing images at a single-cell resolution of protein distribution based on antibodies; their collections include 13 million annotated images. Results mentioned in this presentation were included in a recent *Science* article, entitled, "Tissue-based map of the human proteome" (<http://dx.doi.org/10.1126/science.1260419>) authored by Mathias Uhlén and his colleagues. Eduard Nice from the Monash University (Australia) showed a strategy using high-throughput antibodies for the detection of missing proteins in collaboration with SciLifeLab. György Marko-Varga (Lund University, Sweden) and Henry Rodriguez (NCI, USA) gave an overview of the strategy to develop and access samples in biobanks.



In Segovia, Oct 9, 2014, more than 110 HPP scientists shared their views on some key matters arising from proteomics data management and standards in both C-HPP and B/D-HPP, and more efficient joint efforts between two consortia within HPP.

Session 2: Defining Highly Informative Sets of Proteins for Studies on Specific Diseases: Moving from an HPP Focus on Missing Proteins to the 80% of Known Proteins

Peter Bergsten from Uppsala University (Sweden) presented a proteomics study on the analysis of consecutive islet protein datasets for understanding obesity-associated type 2 diabetes mellitus. Eric Deutsch (Institute for Systems Biology, USA) provided an update on the bioinformatics resources of the PeptideAtlas project components, including PeptideAtlas, Trans Proteomics Pipeline, SRMATlas, PASSEL, SRMQuantAtlas, and SWATHAtlas. PeptideAtlas contains peptides with MS/MS spectra, representing 50% of human proteins, and SRMATlas provides support for efficient SRM assay design for 99% of the human proteome based on synthetic peptides. Currently, neXtProt includes data from SRMATlas to represent the available MS/MS spectra for SRM-type peptide measurements. SWATHAtlas includes a spectral library for the identification of 10,000 proteins. Eric Deutsch highlighted that peptides corresponding to two olfactory receptors were present in the high quality spectral library of PeptideAtlas: one is low quality and the other corresponds to a frequent variant of a high-abundant protein, with no solid protein evidence for an olfactory receptor. Christoph Borchers (University of Victoria, Canada) presented the results of an

interlaboratory assessment of SRM/MRM assays, whereas George Rosenberg (ETH, Zurich) showed the PAN-human repository, consisting of the SWATH spectra library that was built from 331 MS measurements of various tissues and cell types, and can now be applied to the quantification of 10,000 human proteins.

To facilitate the collaboration between the C-HPP consortium and the B/D-HPP group, short talks on the ongoing studies of heart and vascular diseases (Peipei Ping), liver diseases (Fuchu He), cancer (Ruth Huttenhain), and kidney diseases (Tadashi Yamamoto) were presented. In addition, Bruno Domon talked about the ProteomeAnalyzer project (B/D-HPP) with the aim of developing a robust, low-cost, high-throughput triple-quadrupole instrument to better link the proteomics research community to clinical application and research.

Session 3: Combined C-HPP + B/D-HPP Studies of Proteome Biology and Disease

Ongoing joint C-HPP and B/D-HPP studies were presented in this session by showing the integration of disease- and technology-driven programs of HUPO. Joint efforts on the MS-based automated PTM analysis and biological characterization of brain proteins (done by Jong Shin Yoo, PI of chromosome 11 group, KBSI, Korea) and glioblastoma were presented by Young-Mok Park (HBPP, KBSI, Korea). The cancer biomarker discovery program using the combined resources and technologies of the B/D-HPP and C-HPP were presented by György Marko-Varga (Lund University) and Hui Zhang (Johns Hopkins, US). Ravi Sirdeshmukh (Bangalore, India) reported a joint study for mapping protein isoforms (chromosome 12 of the C-HPP) and characterizing biological functions (with the B/D-HPP) based on RNA-seq proteogenomics data integration. Takeshi Tomonaga (National Institute of Biomedical Innovation: B/D-HPP) and Yu-Ju Chen (Taipei, Taiwan: chromosome 4 PI) developed a combined approach for characterizing missing membrane-

embedded proteins and applications to studies of proteome changes in various diseases as a part of the AOHUPO Membrane Proteome Initiative (chaired by Bill Jordan, New Zealand).



Luncheon time chat on the Milan workshop in 2015 (Top) Panel discussion on the PTM, glycol-biology, protein interaction and ASV in Session 4 (Bottom).

Session 4: Data Resources, Data Standards, and Common Informatics Tools

This session began with a panel discussion on the update of HPP metrics for high-confidence protein identification and missing proteins. With the integration of major datasets available in PeptideAtlas (August 2014), GPMdb, neXtProt (Oct. 2014), and Human Protein Atlas resources, Lydie Lane (SIB, Switzerland), Eric Deutsch (ISB, USA), Ulrika Qundos (SciLife Lab, Sweden), Alberto Pascual-Montano (Spain), Fernando Corrales (Spain), and the

representative of the Integrated Human Proteome Browser (IHPB) Working Group shared their views on a better way to maintain high standards for data production. The discussion topics also included (1) the reinforcement of policies on dataset deposition at proteomeXchange, (2) protocols to apply stringent protein FDR (false-detection rate) thresholds, and (3) the reanalysis of datasets using standard PeptideAtlas and GPMdb pipelines. In this session, the creation of the IHPB with major groups developing tools to browse the human proteome, such as Caper (China: chromosomes 1, 8, and 20), TPB (Australia: chromosome 7), genome-wide PDB (Korea: chromosomes 9, 11, and 13), was announced. Intensive discussion focused on the importance of the identification of missing proteins (e.g., proteins having restricted tissue expression, low abundance, high-hydrophobicity, low molecular weight, and members of highly homologous protein families). Lydie Lane and Amos Bairoch provided updated information on neXtProt. Alberto Pascual-Montano introduced the development of a proteogenomics dashboard tool to identify cell lines and tissues, which may express missing proteins using mRNA data. The computational mass spectrometry initiative of HUPO to enhance bioinformatics development was also presented by Shoba Ranganathan. In the second part of the panel discussion led by Cathy Costello (Boston University, USA), topics such as post-translational modifications, alternative splice variants, protein interactions, and protein and glycan chemistry were discussed by William Hancock (Northeastern University, USA), Hisashi Narimatsu (Research Center for Medical Glycoscience, AIST, Japan), Naoyuki Taniguchi (RIKEN Advanced Science Institute, Japan), and Gil Omenn (Univ. Michigan, USA) (picture). The meeting was adjourned after final remarks, conclusions, and perspectives on the C-HPP project presented by Cathy Costello, Pierre Legrain, Mark Baker, and Gil Omenn.

Highlights of 10th C-HPP Bangkok Workshop

Time and Venue: August 9, 2014, Siriraj Hospital, Mahidol University, 2 Wanglang Road, Bangkok, Thailand

The 10th Chromosome-Centric Human Proteome Project (C-HPP) Workshop focused on the management of large proteome datasets and strategies to identify missing proteins. The recent publication of two large datasets for the HPP in Nature by the groups of Akhilesh Pandey (PI of chromosome 22, Johns Hopkins University School of Medicine) and Bernhard Küster (Technische Universität München) not only added very useful resources, but also demonstrated the importance of data management and quality assessment for the proteomics community. Such large datasets trigger community issues, such as the meta-information deposited in ProteomeXchange using raw data, the types of metadata available for each sample, and the availability of MS/MS spectra for the identification of missing proteins. Also discussed was the validation process, in which protein identification goes from a laboratory result to a publication and its inclusion in PeptideAtlas and GPMDDB, followed by processing in neXtProt. The 10th C-HPP Workshop also provided a great opportunity to demonstrate the very valuable and important contributions of the AOHUPO countries to the C-HPP program.

In summary, the presentations of the 10th C-HPP Workshop in Bangkok showed the increasing momentum of the C-HPP activities. The characterization of missing proteins by the chromosome teams is at an advanced stage and shows that missing proteins are either small-size or hydrophobic proteins. Some of the missing proteins may belong to GPCR and olfactory receptors. Additionally, the assessment of current results since the publication of the two Nature papers enables the prediction of the remaining missing proteins and of a strategy to locate them. It is also clear that proteogenomics studies and bioinformatics workflows are more and more popular with the C-HPP teams, which not only supports the identification of missing proteins in samples with high mRNA expression, but also contributes

significantly to the identification of new protein forms caused by genetic variability. ProteomeXchange has now reached a stage where it can be used routinely, and the uses of the Aspera protocol enable a considerably shorter upload time.

Biobank

Biobanking is a very important activity for the teams of the C-HPP consortium, as sample quality, protocols, storage conditions, and relevant metadata of clinical samples are key prerequisites for conducting high quality proteomics experiments and the accurate annotation of produced proteomics data. György Marko-Varga (a talk presented by Peter Horvatovich) updated the C-HPP members with the most important current biobank-related issues, such as sample exchange between laboratories. The HUPO EC has recently decided to establish collaboration with the International Society for Biological and Repository (ISBER), an international forum that addresses the technical, legal, ethical, and managerial issues regarding the handling of repositories of biological and environmental specimens. Their experience would be very valuable for both the C-HPP and HUPO.



Group photo after 10th C-HPP workshop held at Siriraj Hospital, Mahidol University, Bangkok, Thailand (Local organizer: Dr. Visith Thongboonkerd)

ProteomeXchange

Juan Antonio Vizcaíno from EBI provided an update on the development and current functionality of ProteomeXchange, which is the central raw data repository for MS-based proteomics and has a central role in the activities of the C-HPP. Since June 2014, the MassIVE initiative from UCSD (US) joined ProteomeXchange, with a mission to store large quantities of MS/MS data similarly to PRIDE. Currently, LC-MS/MS data can be submitted to ProteomeXchange via PRIDE in addition to direct submissions. Datasets can now be labeled with HPP, C-HPP, and B/D-HPP tags. ProteomeXchange allows complete data submissions, which contain raw data and identification results in the mzIdentML format, as well as partial submissions, which can contain any type of data. Presentation slides from Juan Antonio Vizcaíno are now available at the C-HPP Wiki site and contain information on the numerous practical issues regarding data submission and ProteomeXchange.

Data management & bioinformatics infrastructure in Sweden/Europe

Peter Horvatovich presented a talk on behalf of György Marko-Varga on the Bioinformatics Infrastructure for Life Sciences (BILS), which is a Swedish initiative that aims to provide bioinformatics support for life science research, and the European Initiative Elixir, which has similar goals. He proposed to connect BILS and the Elixir initiative with the C-HPP, which shall be reported at the C-HPP PIC meeting during the HUPO 2014 Congress in Madrid.

Missing Proteins

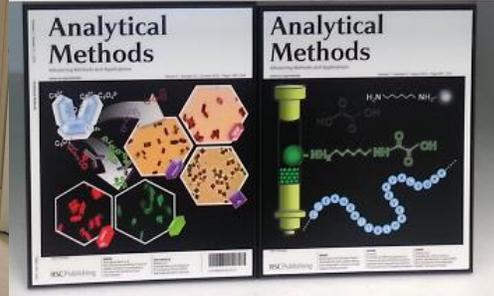
• Advanced in proteogenomics approach

Sanghyuk Lee from Ewha Woman's University, Seoul, Korea, presented a proteogenomics study on early onset gastric cancer using a multilayered integrative approach

derived from genomics, transcriptomics, proteomics, and epigenetics. Gastric tumor tissue was analyzed using the high-low pH reverse-phase 2D-LC-MS/MS, and peptides and proteins were identified using a proteogenomics approach. In this study, Exome sequencing data, RNA-seq data, and proteomics data were combined to identify single amino acid variants and somatic mutations in intestinal and diffuse cancers. Ping Xu (chromosome 1, co-PI, BPRC, Beijing, China) presented a proteogenomics study that attained a high coverage of genes by analyzing cell lines and tissues originating from the liver, the stomach, and the colon. In all the samples, proteomics and transcriptomics analyses were performed to identify proteins from 12,101 genes, reaching 59.8% protein coverage in Swissprot. The Ping Xu group (at Fuchu He Lab) concluded that protein abundance plays an important role in the identification of missing proteins and enrichment of protein subclasses as TF enhances the detectability of low-abundant proteins.

Comprehensive proteogenomics work was presented by Siqi Liu (chromosome 20, PI, BGI, China) on the analysis of three hepatocellular carcinoma cell lines, Hep3B, HCCLM3, and MHCC97H, for mRNA and ribosome-bound mRNA (translatome or RNC-mRNA) analyses. A high correlation between mRNA and RNC-mRNA was found in this study, demonstrating that hydrophobicity (28%) and small molecular mass play (<30 kDa, 75%) predict the unsuccessful detection of a protein. In contrast, the isoelectric point and half-life do not seem to play important roles in protein detectability. This study narrowed the number of missing proteins to 1534, 50% have properties that make detection difficult.

Victor Segura (Spain) presented work on the Spanish chromosome 16 using transcript data from his own experiments and publicly available data (GEO) obtained with cancer samples, cell lines, and healthy tissues to identify missing proteins. The study found that 2861 genes encoding missing proteins were expressed at the



Group photo taken at the entrance of Dr. Visit Thongboonkerd's Research Center (Main site of Chromosome 12 research) at Mahidol University, Bangkok, Thailand.

mRNA level in at least one sample, and that the majority of the genes show sample specificity. The scalable dasHPPboard web platform, developed by the Spanish C-HPP consortium, could serve to visualize public proteomics, genomics, and proteogenomics data, and is accessible online at <http://sphppdashboard.cnb.csic.es/>.

Andrey Lisitsa (chromosome 18, Moscow, Russia) presented proteogenomics work on chromosome 18 using liver tissue and the HepG2 cell line, combining RNA-seq mRNA data and MRM-based targeted proteomics quantification. Quantitative peptide signatures from missing proteins showed that the majority of missing proteins can be detected in these different samples but with a low correlation between them. The quantitative data between transcriptomics and missing proteins did not correlate.

• **Strategy for Identifying Missing Proteins.**

Ghasem Hosseini Salekdeh from the Iranian chromosome Y team presented a detailed overview of the structure, function, and evolution of the Y chromosome. For example, the Y chromosome functions in nucleic-acid

binding, transcription, and translation. Proteogenomics was performed using normal testis tissue (NT), testis tissue in maturation arrest (MA), and testis tissue with Sertoli cell-only syndrome (SCOS), and transcript and protein concentrations were measured in each sample. Several other studies of different cell types (e.g., osteoblasts, eye cells, cardiomyocytes) from various types of stem cells are ongoing.

Gilbert Omenn presented the main goals of the C-HPP, listed the main data sources for the C-HPP project (Ensemble, neXtProt, PeptideAtlas, GPMdb, and Human Protein Atlas), and described the strategies for searching for "missing or inadequately documented proteins," such as the identification of samples using mRNA data, the analysis of a broad range of samples (e.g., samples from early life stages or proteins induced by stress or disease), prediction of proteins undetectable by tryptic digestion and MS, and approaches for highly homologous proteins. The presentation critically assessed the quality of information provided in the two recent Nature papers on the Human Proteome based on several properties (e.g., the identification of Y chromosome genes in fetal tissue,

the discussion of FDR calculations, and the detection of olfactory receptors in non-neuronal tissues) and encouraged all chromosome teams to perform manual assessments of their spectra in the two MS/MS databases from the two studies. Data should be checked against single amino acid changes, as such variability (e.g., nsSNPs, RNA editing where DNA and mRNA information does not match) could cause misleading protein identification.

Young-Ki Paik began his presentation with a eulogy for Juan Pablo Albar and provided an update on the recent C-HPP activities, such as the publication of newsletters and the Busan meeting results. Not detected low-abundant transcripts of low-abundant proteins are due to MS sensitivity issues. Workflows were developed to enhance the sensitivity of MS detection using specific antibody-based enrichment of missing proteins with low-abundant transcripts. Additionally protein forms due to genetic variability such as alternative splicing, SNP detection, and gene fusion can be predicted from mRNA data. His second topic was a strategy to identify the high-quality MS/MS spectra in cases of errors in reference genome annotation or mRNA sequencing data. Both 6 and 3 frames of translation can predict additional protein and peptide forms that can be predicted from the DNA and mRNA data, and can be used to match unidentified, high-quality MS/MS spectra. The presented strategy described the logic of completing the current canonical protein database with information containing genetic variability and reanalyzing unassigned, high-quality MS/MS spectra within a complex bioinformatics workflow. Inclusion of a high-quality peptide spectral library in the peptide spectrum-matching workflow provides an additional improvement, because a spectral library search considers the intensity of the MS/MS spectra, which is left out in current database search approaches.

Yu-Ju Chen's presentation covered a strategy to enhance the identification of hydrophobic membrane proteins

using high and low pH, 2D-LC-MS/MS separation. A total of 11 lung cancer cell lines were analyzed with stage Tip-based, off-line fractionation using high pH, reverse-phase chromatography followed by low pH LC-MS/MS analysis. A total of 7149 proteins (35% from the human Uniprot database) were identified in this study, including 4147 membrane proteins (46% of all human membrane proteins). The majority (76%) of the identified membrane proteins was plasma membrane protein, and the remaining proteins were from other cellular fractions. The study completed a phosphoprotein analysis, identifying 498 missing proteins, 96% at the transcriptomic level. Chromosomes 1 and 19 showed the highest enrichment of the identified missing proteins (11% each). Of the identified missing proteins, 44% were predicted to be integral membrane proteins, whereas the others included peripheral and anchored membrane proteins. Of the 498 missing proteins identified in this study, only 82 were not identified in the two Nature papers. An example of a small size transmembrane protein, Q9NUH8 14B, was shown, which was identified uniquely in this study as two peptides in multiple lung cancer cell lines.

Alexander Archakov from the Russian chromosome 18 team presented an MRM-based approach to identify missing "master" proteins encoded by genes on chromosome 18 from blood using the irreversible binding of proteins on BrCN-sepharose to increase detection sensitivity. The master proteome was defined as follows: "The master protein is the primary translation product of the coding sequence. It is one of the known protein forms, coded by the gene sequence, containing in a prevalent concentration in the tissue." (PMID 21563312). This strategy enabled the identification of 221 out of 265 proteins and quantified 153 out of 260 missing "master" proteins. In summary, the team identified 267 missing proteins (97%) in three types of samples and successfully quantified 260 (94%) of the 276 missing proteins from genes found on chromosome 18, which is a much higher

number than proteins identified in a shotgun proteomics study applying a multidimensional separation technology.

Qing-Yu He presented two strategies to identify missing proteins from aggregated (detergent-insoluble) complexes, combining epigenetics, translome (RNC-RNA), and proteomics analyses. Aggregated protein complexes form after unbalanced protein homeostasis following a regulated pathway in cells from bacteria to humans (PMID 20944667). Steps of the sample preparation method with quality control steps was presented, resulting in 27 missing proteins identified from lung cells and 8 from liver cells, with 3 proteins shared between the tissues. In summary, 32 missing proteins were identified in the studied cells. This method appears promising to enrich missing proteins by aggregation.

Mark Baker from the chromosome 7 team summarized the successful C-HPP Workshop on "Inaugural C-HPP Human Proteome Missing Protein" held in Sydney on July 30, 2014, and provided a summary on the metrics of missing proteins. Using the example of the prestin protein, he showed how to incorporate protein evidence from non-HUPO data sources. He noted that olfactory and taste receptors belong to the GPCR family with seven highly conserved transmembrane domains and are enriched among missing proteins, particularly those encoded by genes on chromosomes 1, 6, and 11. Olfactory receptors (ORs) are expressed in neurons for smell, with one type of receptor expressed per cell. There

are 855 ORs in humans and 391 have an open reading frame. ORs are not only expressed in the nose but are also present in other tissue types (e.g., pancreas, thyroid tissue, sperm) as shown by immunohistochemistry and mRNA expression analyses. Many ORs are orphans, but some have known functions responsible for sensing a particular type of chemical compound; some have other types of functions, such as metabolism of developed drugs, such as orexin receptors. The presentation provided a list of strategies to identify missing proteins and homologous proteins in other species.

Pengyuan Yang gave a talk on the strategies to conquer the remaining missing proteins after the publication of the two Nature papers on the draft of the human proteome. The presentation compared the identified proteins of these two studies and established the remaining proteins without MS evidence. The missing proteins show enrichment in melanoma, olfactory tissues, non-small cell lung adenocarcinoma, and clear-cell adenocarcinoma by Ingenuity Pathway Analysis. Transcriptomics analysis showed that rarely analyzed tissues are enriched in these missing proteins, including tissues from embryonic development, salivary gland, and cerebellum. GPCR, ORs, and membrane proteins form an important component of these missing proteins. This analysis confirmed that hydrophobicity and low abundance are the key characteristics of these "hardcore" missing proteins, which are mainly enriched in genes from chromosomes 11 and 1.

Highlights of C-HPP "Missing Proteins" Workshop

Time and Venue: July 30th, 2014 Australian School of Advanced Medicine, Faculty of Medicine & Health Sciences, Macquarie University, Sydney, NSW, Australia

The inaugural Chromosome-Centric Human Proteome Project (C-HPP) "Missing Proteins" workshop focused on a broad ranging scientific discussion about the metrics and

outcomes from the first analyses of the proteins found missing from the Human Proteome when the metrics outlined in Lane et al., Metrics for the Human Proteome

Project 2013-2014 and strategies for finding missing proteins. *J Proteome Res.* 2014 Jan 3;13(1):15-20.

The workshop aimed to bring together (for the first time) groups of scientists from outside proteomics to raise issues that puzzled them about the HPP and its focus on the "missing proteins". The audience (~60 participants in total) included pharmacologists, neuroscientists, oncologists and chemical biologists, synthetic chemists and researchers from a broad human biomedical research background, as well as APAF mass spectrometrists and proteomic specialists. The workshop also recognized the recent publication of two relatively large and unverified HPP datasets in the journal *Nature* by the groups of Küster and Pandey.



Gil Omenn from the University of Michigan and Chair of the HPP introduced the workshop's first session

(sponsored by the 13th International Conference on Bioinformatics & Korean C-HPP Initiative) and carefully walked the audience through the process of establishing the metrics HUPPO and the HPP had endorsed in the Lane et al., publication and strategies for identifying the HPP "Missing Proteins". Young-Ki Paik from Yonsei University and Chair of the C-HPP Consortium followed up with a general analysis of the strategies that have been proposed to functionally validate the HPP missing proteins. The morning session concluded with a breakdown of major missing protein families by Mark Baker including olfactory and taste receptors. He challenged HPP to consider all forms of scientific evidence for the functionality of a protein and to consider how to capture data streams from non-MS and immunological sources. Examples of proteins where purported reasonable scientific evidence exists from non-HUPPO sources were also presented. Mark urged participants to "speak the same language" with regards to use of the endorsed MS metrics for all HPP studies.

The second session sponsored by the Australian Peptide Association was chaired by Ed Nice and included talks by pharmacologist Mark Connor from Macquarie University who provided non-proteomics evidence for many GPCRs that had been deorphanized by big pharma or pharmacological studies but were still considered HUPPO as "missing proteins". Jens Coorsen from the University



Photo taken from the meeting room after scientific session was well done with good productive discussions.

of Western Sydney, then questioned a major technical issue many report around 2DE and proposed that the method could be useful in exposing certain classes of HPP missing protein species. Benjamin Parker from the Garvan Institute then addressed the challenge of undertaking a detailed discovery and targeted phosphoproteomics study in under a week – what a challenge. His data indicated the tremendous capabilities phosphoproteomics now delivers on. The final speaker of the session was Peter Hoffman from University of Adelaide who addressed the issue of tissue MALDI and opportunities/challenges associated with driving this technique forward under the HPP.

During lunch a session run by Ignatius Pang from Marc Wilkins' UNSW group addressed a series of useful PG Nexus Tools that participants may find useful for their studies.

The afternoon session was chaired by Nikki Verrills from the University of Newcastle who addressed identification of oncogenic signalling pathways in myeloid leukaemia using phosphoproteomics. Shoba Ranganathan from Macquarie University the specifically spoke to how we need to revisit FDR rates and bioinformatics analysis of large datasets. In particular, Shoba suggested the two Nature papers need to be put under a similar microscope to existing data sources and using the same metrics suggested by HUPO. Ed Nice from Monash University

then entertained the crowd with his discussion of the challenges of relying upon affinity reagents in a timely paper entitled "The Good, Bad & Ugly Ab-based Missing Protein Strategies". Finally, in a presentation supported by a number of Juan Pablo Albar's PowerPoint slides, Macquarie PhD Student David Cantor outlined the processes necessary for all scientists to deposit raw MS data through the PRIDE/ProteomeXchange process into the HPP.

The final session of the workshop (sponsored by HUPO2010 Association) was chaired by Tony Purcell from Monash University who discussed advances in Immunoproteomics. This was followed by the official launching of the new HPP PediOme Project co-chaired by Vera Ignjatovic from the University of Melbourne. In conclusion, chemical biologist Peter Karuso from Macquarie University reviewed current progress towards establishing small molecule approaches for protein discovery as part of the HPP Missing Proteins effort.

A magnificent "truffle proteome" celebration dinner for 35 participants was held at the Gourmet Mediterranean Pizza co-sponsored by the following whom we sincerely thank for a wonderful experience:

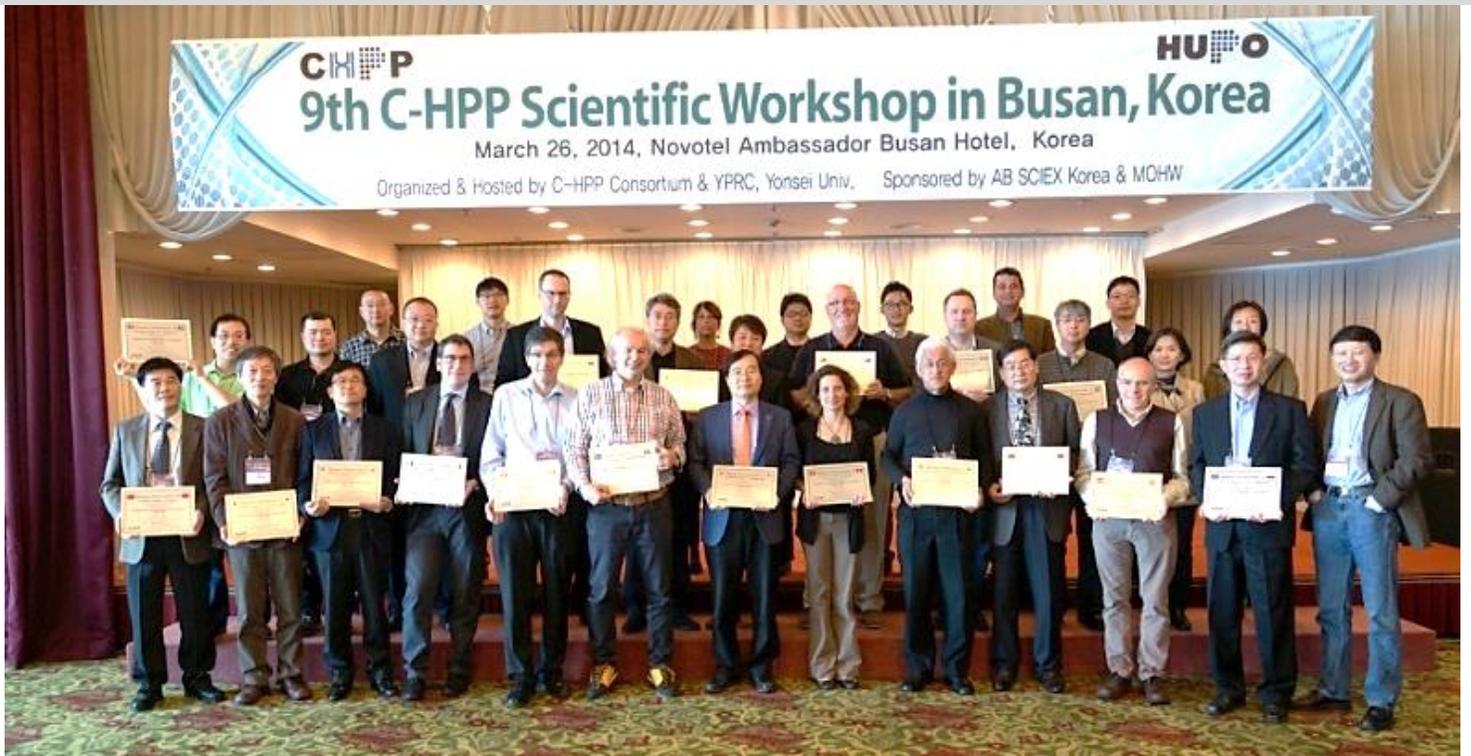
Sigma Aldrich Australia, Terra Preta Truffles, De Iulius Wines, and Diasorin Bioplatforms Australia

Highlights of the 9th C-HPP Busan Workshop

Time and Venue: March 26, 2014, Novotel in Haewundae Beach, Busan, Korea

With the scientific slogan, "Integration of the proteomic parts list with transcriptomic information," this workshop was designed to brainstorm various ideas on the integration of a genome-wide proteomics parts list into transcriptomics data, making the C-HPP dataset more

informative and user friendly to those working in biomedical sciences. It was also intended to develop a new paradigm for an interdisciplinary collaboration and discuss how synergistic outputs can be realized by putting these individual chromosome initiatives together



A group picture taken after presenting a metal plaque of chromosome assignment to PIs of the C-HPP consortium

for better biomedical research. We also hoped to focus the C-HPP goals and develop strategies for identifying and characterizing missing proteins using new technologies and rapid changes in scientific trends in the multi-omics era. This meeting was organized by the C-HPP leaders (chair: Young-Ki Paik, co-chairs: Bill Hancock and György Marko-Varga) and sponsored by the Korean Ministry of Health and Welfare and ABSCIEX Korea. This meeting took place at the Novotel Ambassador Hotel, located in beautiful, sunny Haeundae Ocean Beach, Busan, Korea, on March 26, 2014. A total of 68 people, including PIs, their designated representatives, co-PIs, and co-workers, attended this meeting, which is fully packed with many exciting talks and key sessions under the specific themes. During the meeting, an official metal plaque of chromosome assignment was presented to the PIs or their designate (see picture).

Session 1: Inter-lab Standardization

Some of the highlights of this meeting are outlined here. In session 1, which was moderated by Young-Ki Paik, the long-term goals and plans of the C-HPP were discussed and the proposed revised version of the 10-year plan (2012–2022) was reviewed following a brief overview of past activities. After exchanging some ideas about amendments to the original goals so as to comply with new technology and updated resources, it was suggested that each participant provide a feedback after the meeting. In fact, a few colleagues sent their opinions and comments on the revised plans, which will be further updated and presented at the 10th C-HPP Workshop in Bangkok, Thailand.

Session 2: MRM Analysis of Missing Proteins

In session 2, led by Mark Baker, MRM standardization was discussed right after an energetic talk given by Christoph Borchers (chromosome 6 group), University of Victoria,

Canada. He reported updates on the progress of the large-scale Inter-lab MRM Initiative for the C-HPP, which focuses on standardization kit developments and applications to quantitative plasma proteomics studies. Many issues centered on team cooperation for data production and sharing, publications, and quality control of submitted datasets.

Session 3: Plenary Lectures

In session 3, chaired by Christoph Borchers, Jong Shin Yoo at KBSI (PI of the chromosome 11 group) in Ochang, Korea, presented a very stimulating talk on MRM-based analyses of protein isoforms and automatic analyses of site-specific N-linked glycoproteins. His lecture was impressive with very innovative platforms that allow one to identify, characterize, and quantify disease-specific isoforms and missing proteins in real samples.

Session 4: Long-Term Plans Revisited

In session 4, moderated by György Marko-Varga (PI of the chromosome 19 group, Sweden), C-HPP members reevaluated the time line and milestones of work on the search for missing proteins and a genome-wide proteomics parts list. In particular, the outstanding supportive roles of those involved in the management of major public databases (i.e., neXtProt, GPMDB, PeptideAtlas, and HPA) were highly appreciated for their provisions of up-to-date information on the status of missing proteins and protein lists identified from the literature. The cooperation of the B/D-HPP group was also discussed. To this end, Gil Omenn, chair of the HPP, promised to provide an updated list of B/D-HPP teams where C-HPP scientists can make more interactions.

Regarding the publication of the next consortium papers, due to the nature of some uncertainty about the data production schedule (e.g., missing protein identification and characterization), this plan was delayed until the long-term plans are fixed.

Luncheon Session: New Tools and Strategy for the C-HPP

In the luncheon seminar, co-led by Lydie Lane, SIB, Geneva, and Toshihide Nishimura, Tokyo Medical University, Dr. Jason Neo from ABSCEX presented a talk on "New LCMS Data Acquisition and Processing Strategies for Quantitative Proteomics." This stimulating talk fit well with the overall direction of the consortium, in terms of data collection and sharing that were further discussed at the 10th C-HPP Workshop in Bangkok, on August 9, 2014. This talk was followed by three special presentations made by teams from Korea (on chromosomes 9, 11, and 13 by Young-Ki Paik), China (on chromosomes 1, 8, and 20 by Pengyuan Yang), and Spain (on chromosome 16 by Juan Pablo Alba). These presentations gave some information on each country's grant size, goals, and deliverables, along with some tips for grant applications for other members who are in the stage of grant preparation for the project.

Session 5: Interaction with Genomic Groups

In session 5, led by György Marko-Varga, co-chair of the C-HPP consortium, two plenary talks were presented: 'Discovery of druggable targets from liver cancer genome' (by Dr. Hyun Goo Woo, Ajou University School of Medicine, Korea) and 'Introduction of ENCODE data for potential linkage with proteomics research' (by Dr. Sung-Min Ahn, University of Ulsan College of Medicine). These two talks addressed key points of the C-HPP strategies with respect to the pursuit of disease studies and applications of ENCODE DB for the identification of missing proteins. Insightful discussions and comments were offered in these plenary talks.

Session 6: Transcriptomics and ENCODE

In session 6, led by Tadashi Yamamoto, Niigata University, two invited talks on transcriptomics and ENCODE were presented: "Integrated chromosome 19 transcriptomic and proteomic data sets derived from glioma cancer stem-cell lines" (by György Marko-Varga on behalf of Carol Nilsson, UTMB, US) and "Transcriptomic maps for all chromosomes using RNA-seq data from ENCODE" (by Alberto Pascual-Montano [chromosome 16 group], Madrid, Spain). These two talks revealed new information and practical aspects on the current use of ENCODE and RNA-seq data for the C-HPP consortium.

Session 7: Progress Reports from Individual Research Teams

In sessions 7 and 8, led by Juan Pablo Albar, Visith Thongboonkerd, and Daniel Figey, seven talks were presented: "Updates on the neXtProt" (by Lydie Lane, SIB), "Caper 2.0 covering chromosomes 1, 8, and 20" (by Siqi Liu on behalf of Ping Xu, BPRC, China), "Chromosome 8 missing proteins" (by Pengyuan Yang, Fudan University, Shanghai, China), "Bioinformatics pipelines and methodology for the discovery of missing proteins" (by Lydie Lane for Jerome Garin, CEA, France), "Chromosome 18 TranscriptoProteome: Update 2013" (by Andrey Lisitsa, Moscow, Russia), "Translating evidence and chromosome-centric human proteome investigations" (by Qingyu He, Jinan University, Guangzhou, China), and "Progress on the Studies of the Subunit 4 of Complex (MT-ND4L)" (by Andrea Urbani, University of Rome, Italy). This series of talks mostly focused on potential solutions for missing protein identification, characterization, and transcriptomic assessment of each project.

Session 8: Future Plans and Perspectives

In the final session, co-chaired by Young-Ki Paik and György Marko-Varga, the future plans of C-HPP workshops in 2014/2015 were discussed. Following an introduction about the Madrid Congress by Juan Pablo Alba and the Bangkok Workshop by Visith Thongboonkerd, Paola Roncada, University of Milano, presented an overview of the Milano EXPO in conjunction with EuPA during which the 12th C-HPP Workshop will be held (June 23, 2015). In conclusion, throughout this C-HPP held in Busan, it was thought that the C-HPP group had a golden opportunity to assess the current state of the project that delivers their results on a genome-wide protein parts lists and to outline the future plans to continue C-HPP development. The workshop concluded around 18:30 with a summary of action plans given by Young-Ki Paik. Briefly, the following topics were discussed: 1. Long-Term Goals: To revise according to survey responses from individual teams based on the current pace of research; 2. Large-Scale MRM: Moving towards data collection deadline April 18; 3. Milestone Papers: Go with the revised long-term plans; 4. Funding Strategy: Use all available resources and tips given by teams from China, Korea, and Spain; 5. ENCODE: Revisit non-long coding RNAs with translation potential; 6. RNA-seq: More efforts on data acquisition and integration; 7. Update neXtProt, Caper, TranscriptoProteome, etc.; 8. New Strategy for DB Construction of Nontranslating RNAs; and 9. Future Plans on Bangkok, Madrid, and Milano 2015: Well presented to move this project forward.

A farewell dinner was sponsored by ABSCIEX Korea, during which everyone enjoyed a traditional Korean cuisine with a special performance on the Bamboo flute by Prof. Hui-Soo Kim of Pusan University. This workshop is supported in part by a grant from the Korean Ministry of Health and Welfare (to YKP, The International Consortium Project, HI13C2098).



Introduction of the C-HPP session by Young-Ki Paik, Chair of the consortium (left) and discussion session (right)

Individual Teams' Activity Reports



Chinese Team:
Chromosomes 1, 8, and 20

Fuchu He & Ping Xu; Pengyuan Yang; Siqui Liu & Qing-Yu He

Grants: The Chinese teams (chromosomes 1, 8, and 20) have one major grant approved by the central government, which provides \$4M USD for the C-HPP study for 3 years from 2014–2016. In addition, another grant is being submitted to support this study in 2015 (Fuchu He, Pengyuan Yang, Siqui Liu, Ping Xu, and Qing-Yu He). As the top leader of the core facility and center for proteomics in China, Fuchu He played a key role in this series of studies together with Pengyuan Yang (chromosome 8), Siqui Liu (chromosome 20), Qing-Yu He (chromosome 20, co-PI), and Ping Xu (chromosome 1 co-PI) in these cooperative studies.

Workshops and Meetings: The Chinese team had 10 workshops on the C-HPP in China and related studies during the annual CNHUPO conference during the past 2 years. In 2015, four to five C-HPP workshops will be held in China.

Progress on missing proteins: More than 10 manuscripts are now under preparation for submission to the JPR special issue in 2015 from which more than 300 missing proteins are likely to be identified.



Japanese Team:
Chromosomes 3 and X

Chr3: Toshihide Nishimura, X: Tadashi Yamamoto

• Chromosome 3

Chromosome 3 (Toshihide Nishimura, Hiromasa Tojo, Gyorgy Marko-Varga, Neuman Sze): There is a good effort to identify missing proteins from clinical tissue specimens of early lung cancer using laser-microdissection and high-performance shotgun mass spectrometry.

This team tentatively captured 119 missing proteins (PE=2–5) throughout chromosomes 1–X, in which seven and eight missing proteins (PE=2, 3, and 5), respectively, on chromosomes 3 and 19 were included. They are currently re-evaluating the validity of their identification. Those data will be submitted for publication in the HUPO official journal. The article would describe the above missing proteins obtained from clinical tissues of lung cancer subtypes, large-cell neuroendocrine lung carcinoma (LCNEC), other large-cell carcinoma (LCC), and small cell lung cancer (SCLC).

In addition to missing proteins, they have performed a chromosome-centric, protein-protein interaction (PPI)

network analysis using 1998 proteins identified from three lung cancer subtypes to unveil how expressed proteins are associated with key pathways; this work makes it possible to compare the network associations among disease subtypes, and helps to better understand the molecular mechanisms involved in each lung cancer subtype.

• Chromosome X

Missing Proteins: Physicochemical properties of missing proteins have been analyzed and will be submitted to the journal for publication.

Collection of antibodies for proteins encoded on chromosome X is also ongoing. Only eight antibodies for chromosome X proteins, which are not established in The Human Protein Atlas, have been collected.

Grants: Dr. Yasushi Ishihama has been chairing a team to apply a JST grant on the Database Integration Coordination Program, which was submitted in February 2015.

Meeting: The chromosome X meeting was held during the Japanese Proteomics Society annual meeting in Tsukuba (July 17 and 18, 2014).



Taiwanese Team: Chromosome 4

Yu-Ju Chen

Grants, Workshops, and Meetings: In 2014, the chromosome 4 team successfully invited more team members to join the working group. They welcome scientists from Kaohsiung Medicine University, led by Vice President Yi-Ming Chen and his colleagues, including Drs. Pin Huang, Yaw-Syan Fu, Yu-Chang Tyan, Chi-Yu Lu, and Wan-Chi Tsai, as well as from Taipei Medical University (Drs. Wei-Chiao Chang and Chia-Li Han). Based on the idea of collecting seed budgets from individual institutions, a total of 5M (NT\$) (2.5-fold increase over the

2014 budget) will be jointly supported by six institutions for their activities in 2015. To coordinate the activities and brainstorming for the project, a round-table discussion and two workshops were held in May, June and November in 2014.

Progress on Missing Proteins: Progress on the chromosome 4 project includes the completion of the first version of the knowledge base of chromosome 4 protein-coding genes. In hunting for missing proteins, we performed a systematic analysis of the membrane proteome and phosphoproteome in lung cancer cells and tissues. Based on the annotated missing proteins from UniProt and HuMemProtDB, we observed 498 annotated missing proteins, including 423 annotated missing proteins from membrane fractions and 75 missing proteins from the enriched phosphoproteome of lung cancer specimens. All the data have been submitted to ProteomeXchange. These exciting results were also recognized by the HUPO, with an award for our young scientist, Dr. Hsin-Yi Wu of Academia Sinica during the C-HPP poster presentation at the 2014 HUPO congress in Madrid. In summary, our approach to target the membrane sub-proteome has enabled the identification of under-explored missing proteins with hydrophobic transmembrane helical peptides.



Netherlands' Team: Chromosome 5

Rainer Bischoff

Grants: The chromosome 5 team has prepared several grant applications that include 1) Pediatric Obesity Weight Reduction (POWER) (EU Horizon 2020), 2) Analytics for Biotechnology (A4B) Innovative Training Networks (ITN) (EU Horizon 2020), 3) Common Mechanisms in Type-2 Diabetes Mellitus and COPD (CoMeDiaCO) (EU Horizon 2020), 4) the Biomarker Development Center (COST) (EU Horizon 2020), 5) validation of new and existing

biomarker candidates for clinical use in early diagnoses of lung carcinoma (Dutch Cancer Society, KWF), 6) high-resolution mass spectrometry for meta-proteogenomics in healthy aging research (Dutch Research Foundation, NWO, investment grant), 7) high-resolution ICP-MS in the bioanalysis of proteins, metal-based pharmaceuticals, and inorganic nanoparticles (Dutch Technology Foundation, STW), 8) a proteogenomics approach to identify causal networks involved in COPD development; (Longfonds clinical application), 9) a proteogenomics study of primary fibroblasts from COPD patients, (internal funding), and 10) head and neck cancer proteomics (internal funding).

Progress on missing proteins: Participation in a joint collaboration with chromosome teams 10, 16, and 19 to express missing proteins with an in vitro expression kit and develop a sensitive SRM method to identify missing proteins with transcription evidence in human cell cultures and tissues. An ongoing effort is also in the proteogenomics study of primary fibroblasts from COPD patients.

Others: Péter Horvatovich of the chromosome 5 team is the managing editor of the C-HPP Wiki at <http://c-hpp.webhosting.rug.nl>.



**Canadian Teams:
Chromosomes 6 and 21**

Chr6: Paul Keown, Chr21: Daniel Figeys

Grants: No grant updates are related to this project. They are using our own resources.

Member or leadership changes: The chromosome 21 group is expecting a new bioinformatics postdoc to join next month, and he will be involved in this project

Publications

- **Chromosome 6**

- **A new protease discovered for improved coverage of the human proteome.**

Huesgen, P.F., Lange, P.F., Rogers, L.D., Solis, N., Eckhardt, U., Kleifeld, O., Goulas, T., Gomis-Rüth, F.X., and Overall, C.M. 2015. LysargiNase Mirrors Trypsin for Protein C-Termini and Methylation Sites Identification. *Nature Methods* 12, 55-58

- **New bioinformatics analysis tools launched on the TopFIND v3.0 Knowledge-base of protein N termini.**

Fortelny, N., Yang, Sh., Pavlidis, P., Lange, P.F., and Overall, C.M. 2015. Proteome TopFIND 3.0 and TopFINDER: Database and Analysis Tools for the Association of Protein Termini to Pre- and Post-translational Events. *Nucleic Acids Research (Database Issue)*, D290-297.

- **The human erythrocyte proteome and N-terminome published.**

Lange, P., Huesgen, P.F., Nguyen, K., and Overall, C.M. 2014. Annotating N termini for the Human Proteome Project: N termini and N -acetylation Status Differentiate Stable Cleaved Protein Species from Degradation Remnants in the Human Erythrocyte Proteome. *Journal of Proteomics Research* 13, 2028–2044.

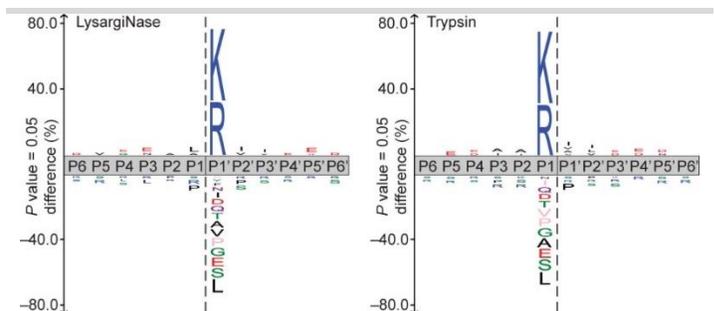
- **The human platelet proteome and N-terminome published.**

Prudova, A., Serrano, K., Eckhard, U., Fortelny, N., Devine, D., and Overall, C.M. 2014. TAILS N-Terminomics of Human Platelets Reveals Pervasive Metalloproteinase-Dependent Proteolytic Processing in Storage. *e-Blood*, 124 (26), e49-e60

- **Editorial**

Overall, C.M. 2014. Can Proteomics Fill the Gap Between Genomics and Phenotypes? *Journal of Proteomics* 100, 1-2

☞ **Additional reports of Chr 6 teams are summary in Appendix.**



Frequency distribution plots of cleavage sites identified in human MDA-MB-231 breast cancer cell lysates digested with LysargiNase (n=1,917) and trypsin (n=3,235). Sequences were aligned at cleavage sites between P1 and P1' and normalized to the natural human amino acid abundances (Pitter F. *et al.*, 2014, *Nat. Methods*)

Workshops and Meetings

- **Chromosome 21:** On June 3, 2014, they organized the third China-Canada Systems Biology Symposium in Shanghai.

Progress on missing proteins

- **Chromosome 21:** A few months ago, they had found five new proteins; however, they are redoing the analysis using a much larger human dataset.



Korean Team:

Chromosomes 9, 11 and 13

Chr9: Je-Yoel Cho/Soo Youn Lee, Chr11: Jong Shin Yoo, and Chr13: Young-Ki Paik

Grants: In 2013, the Korean team successfully secured a 9-year grant (5+4) supported by the Ministry of Health and Welfare, Korea, for the C-HPP (PI: Young-Ki Paik, Yonsei University), which covers three chromosome projects (chromosomes 9, 11, and 13) and some expenses for the operation of the C-HPP HQ office at the YPRC. Recently, the chromosome 9 team (PI: Je-Yoel Cho, SNU/Soo Youn Lee, SKKU) received a grant of \$200k USD/year (3 years) from the Korean Ministry of Science, ICT, and Future Planning (MISP) for the proteogenomics analysis of metabolic diseases. This grant will help to promote the progress of the C-HPP project on Chr 9.

Workshops and Meetings: The Korean C-HPP (KC-HPP) team held five research workshops during 2014 where they set up the overall strategy for deciphering missing proteins and disease-related proteins (lung cancer: chromosome 9, neurological disease: chromosome 11, and preeclampsia/liver cancer: chromosome 13). The workshops took place at five different places at different times (Jan. 21; Feb. 10; March 27; May 23; and July 4 in Korea).



The 5th KC-HPP Workshop held in Kangneung Resort on July 4, 2014 where 26 people attended.

Progress on missing proteins : The Korean teams (chromosomes 9, 11, and 13) have successfully identified 97 missing proteins that contain 37 disease-related proteins and 26 alternatively spliced proteins. In particular, the chromosome 9 group identified six new UTR-encoded peptides that were further subjected to validation using synthetic peptides. As recommended by the C-HPP consortium, the Korean teams are working towards MRM/SRM-based protein quantification and characterization for missing protein identification using particular tissue samples (e.g., lung, brain, and placenta). The Korean teams have recently launched an inter-lab standardization of glycoprotein analysis, which will be used as an SOP for future identification and characterization of disease-related proteins in the three chromosomes studied by them.



Group Picture of Chr. 11 team in front of a metal plaque: from left, Bonghee Lee, Jin-Young Kim, Kyung-Hoon Kwon and Jong Shin Yoo.

**Thailand Team:
Chromosome 12**

Visith Thongboonkerd

The draft maps of the human proteome were published in two independent articles and provide a catalog of proteins encoded by the human genome, including missing proteins, based on large-scale mass spectrometric analysis. Six months later, a tissue-based human proteome map was published, which includes an expression and distribution study of human proteins across tissues and cells using specific antibodies. Independently, the ongoing global effort, the Chromosome-centric, and Biology and Disease centric Human Proteome Project (C-HPP and B/D-HPP) initiated by HUPO, aims to study human proteins and their variants in terms of biology and disease in a chromosome-centric manner. Under this initiative, a consortium formed among five Asian research teams, including India, to study chromosome 12-encoded proteins and their disease context. Together, the group has recently published the first series of three papers giving the overall vision and the initial contributions. While one of the draft maps of the human proteome was largely contributed by an Indian team, Indian researchers play significant roles in the other two initiatives as well. These efforts will be pursued further as more teams join and more disease and biology components are incorporated.

**French Team:
Chromosome 14**

Jérôme Garin

Grants: Currently, the French team does not have any funding to support the C-HPP project. However, due to the high interest, we have been participating in this consortium, ProFI (the national proteomics infrastructure in France; coordinator J. Garin) and have managed to allocate some permanent people to the C-HPP. Charles

Pineau (IRSET) also allocated some permanent people from his group to the C-HPP and launched a priority project to identify missing proteins in the human male genital sphere (organs and isolated cells).

Member or leadership changes: Jérôme GARIN became PI of the chromosome 14 group, seconded by Yves VANDENBROUCK who is organizing a "computational café" at HUPO Vancouver 2015 (under discussion with Henning HERMJAKOB and Lennart MARTENS). They are ready to share resources (MS-MS datasets, SRM facilities, etc.) for C-HPP teams and bioinformatics staff, particularly for the chromosome annotation workshop discussed in Segovia with neXtProt.

Progress on Missing Proteins (collaboration with the Switzerland team): Franco-Swiss consortium (Chr 14 and 2) recently performed a study on missing proteins. The French proteomics infrastructure has collected high-quality datasets from 40 human samples, including a series of rarely studied cell lines, tissue types, and sample preparations. They designed a strategy that combined the use of bioinformatics screening and subsequent MS-based validation to identify what were previously missing proteins in these datasets. A screening database search (85,288.dat files) identified 85 missing proteins with 126 unique peptides. These peptides were evaluated by applying two different criteria: peptide-level false discovery rate (FDR) calculations and expert MS/MS spectral quality assessments. Synthetic peptides were then produced for 76 proteins (103 peptides) and used to generate reference MS/MS spectra. Spectral similarity scores were calculated for each pair of reference-endogenous spectra. Finally, LC-SRM assays were developed to target proteotypic peptides from four missing proteins detected in tissue/cell samples that were still available and for which sample preparation could be reproduced. These LC-SRM assays unambiguously detected the endogenous unique peptide for three proteins.

The Franco-Swiss consortium (chromosomes 14 and 2) also performed a study demonstrating the value of spermatozoa in the search for missing proteins. Validation criteria for mass spectrometry data was established by ProFI (the national Proteomics infrastructure in France; Coordinator is J. Garin). It is known that the testis is a promising site to search for missing proteins. Interestingly, a large number of protein-coding genes are specifically expressed in the germ cell lineage after meiosis. Unfortunately, it is extremely difficult to isolate post-meiotic germ cells from human testes. They demonstrated that ejaculated spermatozoa are an alternative potential source that can be obtained noninvasively of proteins whose expression is restricted and/or specific to the germ cell lineage. A transchromosome-based data analysis was performed to catalog missing proteins in total protein extracts from isolated human spermatozoa. They have identified and manually validated unambiguous peptide matches for 109 missing proteins in human spermatozoa. In addition, we carefully validated three proteins that were scored as dubious by neXtProt.

Ongoing work on the characterization of missing proteins in the human testis (isolated cells) and epididymis is a collaborative project between the teams of C. Pineau and Lydie Lane.



Brazilian Team: Chromosome 15

Gilberto B. Domont

Workshops: Past or future plans in 2015:

The Brazilian team is planning on hosting an HPP meeting and workshop in Rio de Janeiro, Brazil around November, 2015. It would be very interesting to see progress on the project carried out by South American teams.



Spanish Team: Chromosome 16

Fernando J. Corrales

Fernando J. Corrales (CIMA, University of Navarra, Pamplona, Spain) assumed the PI leadership of the chromosome 16 team, whose activity has been officially granted for a period of 4 years (2014–2017) by the Spanish Institutes of Health (ISCIII). The Spanish consortium hosted the XIII HUPO Annual Congress in Madrid and the Workshop on “Efforts to accelerate the HUPO Human Proteome Project” in Segovia, where cutting-edge advances in proteomics were presented, and current challenges and future research avenues for the HPP were discussed. Following our the foundational objectives, they observed 486 proteins encoded by chromosome 16 using shotgun MS experiments in different cell lines and subcellular fractions, including three missing proteins. Preliminary SRM methods were developed for 97 chromosome 16 proteins; recombinant forms for 40 missing proteins were produced in cell-free systems, which are currently under analysis for the optimization of SRM methods that will be applied to their detection in biological matrices. A method to predict the probability of missing protein detection in human cell lines and tissues was recently developed [Journal of Proteome Research 2015 Mar 6;14(3):1350-60] using more than 3,400 transcriptomic experiments. It might facilitate the quest for missing proteins. They have developed a web platform to integrate and distribute proteomics, genomics, and proteogenomics data (in preparation for the JPR special issue), including different technologies (NGS, shotgun, MRM), different platforms (RNA-seq, microarrays), different sample types (cell lines, normal and diseased tissues), different file formats (fastq, FASTA, txt), and fully traceable databases (GEO, ProteomeXchange) (<http://sphppdashboard.cnb.csic.es>). Finally, B/D and C-HPP connections were established by defining the protein priority lists in cancer, obesity, and liver diseases,

infectious diseases, and rheumatoid diseases, whose detection and quantification will be optimized using SRM methods.



**USA Team:
Chromosomes 10, 17 and 22**

Chr10: Joshua LaBaer, Chr17: William S. Hancock and Chr22: Akhilesh Pandey

Chromosome 17 team members are William Hancock, Boston (PI), Gilbert Omenn, Michigan, Michael Snyder, Stanford, and Ronald Beavis, Manitoba, Canada

Chr 17 is focused on the ERBB2 (Her2/neu) amplicon and its involvement in one type of human breast cancers. They examine co-expression among the 23 genes of the amplicon and the role of alternative splicing in key pathways for breast cancers, notably involving ERBB2 (EGFR). They use proteomic analysis extensively.

Grants: The Chr 17 team has recently received two grants

- 1) Proteomics of neuroblastoma and medulloblastoma, Northeastern University and Children’s Hospital Seattle. 2) NIEHS P30 1U54 ES017885 (P.I. Rita Loch-Carusio, Omenn, Co-I) (4/30/11-4/29/16) Environmental Health Center Core Grant: Life-stage Exposures and Adult Disease.

Workshops and Meetings: Chr 17 works was held during US HUPO meeting, Tempe, Arizona, USA, 15-18 March



**Russia Team:
Chromosome 18**

Alexander Archakov

Grants: The chromosome 18 team was supported in 2014 by the following grants: Chromosome-centric targeted proteomics for human blood plasma proteins investigation and Development of MS-test for mining plasma proteins with high clinical perspective, Scientific

Program of the Federal Agency of Scientific Organizations («Human Proteome»).

Member or leadership changes: Principal Investigator: Alexander Archakov; co-PIs: Elena Ponomarenko (bioinformatics) and Andrey Lisitsa (standardization).

Participating members: Institute of Biomedical Chemistry (core facility), Moscow; Institute of Physical-Chemical Medicine of the FMBA, Moscow; Centre "Bioengineering," Moscow; Institute of Bioorganic Chemistry, Moscow; Institute of Molecular Biology, Moscow; Institute of Biomedical Problems, Moscow; International Tomography Center, Novosibirsk.

Progress on missing proteins: In a broad sense, the scope of “missing” master proteins is restricted to unidentified proteins; detection does not meet the gold threshold of neXtProt, the green threshold of GPMdb, an FDR of 1% of the protein level in PeptideAtlas, or confident detection in Human Protein Atlas. Per our understanding, “missing” master proteins do not exist in separate tissues and cells, so these master proteins are not expressed in the human body and cannot be detected due to the limitations of method sensitivity (Zgoda et al., JPR 2013).

The list of ‘missing’ master proteins for chromosome 18 is in the results of our SRM proteome (Table).



**Swedish Team:
Chromosome 19**

GyÖrgy Marko-Varga

Grants: The Chr 19 team has received several grants: The Cancer Prevention and Research Institute of Texas, USA; Research Council of Norway; Fudan University; “Center for Clinical Bioinformatics,” China: \$4.5M USD; Kamprad Foundation, Lund, Sweden; Swedish Pharmaceutical Society, National Grant, Sweden; Lund University Hospital,

Sweden; Lund University Infrastructure Grant, Sweden; and Knut & Alice Wallenberg Foundation, Sweden.

Member or leadership changes: György Marko-Varga, Lund, Sweden (PI); Carol L. Nilsson, Galveston, USA; Elisabet Carlsohn, Gothenburg, Sweden; Manuel Fuentes, Salamanca, Spain; Sophia Hober, Stockholm, Sweden; Frode Berven, Bergen, Norway; Per Andren, Uppsala, Sweden; Goutham Edula, Hyderabad, India; and Xiangdong Wang, Shanghai, China.

Workshops and Meetings: The chromosome 19 team has hosted four face-to-face meetings recently: Berlin, Germany, March 2013; Galveston, USA, November 2013; Uppsala, Sweden, August 2014; and Salamanca, Spain, September 2014.

At these meetings, they discussed the scientific progress, organization, and research activities. They have been funded by UTMB, Uppsala University, and Salamanca University for these meetings. For the Salamanca summit, we also invited Josh LaBaer from the chromosome 10 team, Peter Horvatovich from the chromosome 5 team, and a representative from the Spanish chromosome 16 team. This initiative resulted in the joint action of the four chromosome teams to share resources and create a common bioinformatics and analysis platform strategy to study the missing proteins of these chromosomes. Currently, the number of proteins with no unambiguous evidence at the protein level is 2949 (PE2, PE3, and PE4) with the addition of 616 (PE5) dubious proteins whose existence is questioned, and the number of genes considered to encode a protein is 20,655 according to neXtProt on September 19, 2014, which leaves 17.77% that are missing proteins of the complete set of human protein-coding genes. There are three main reasons for the missing proteins: (1) current mainstream proteomics technologies are unable to identify the proteins, for example, due to low abundance of the protein; the sequence does not contain tryptic peptides that could

uniquely identify the proteins; and the proteins or peptides are lost during sample preparation and analysis; (2) current missing proteins are expressed in rare tissues, cell types, or are expressed transiently; or (3) the proteins are not expressed at all and provide silent information about the human genome. The consortium for chromosomes 5 (the Netherlands), 10 (US), 16 (Spain), and 19 (Sweden, US) presents a unique strategy to identify expressed human proteins, which currently have no evidence at the protein level, representing 863 missing proteins in chromosomes 5, 10, 16, and 19.

In China, Fudan University just created the "Fudan University Center for Clinical Bioinformatics" as a platform to integrate medical informatics, data mining, bioinformatics, a tissue biobank, and others (preliminary information can be found at www.fucb.org) and received \$4.5M USD coordinated by Prof. Wang, by joint efforts with the chromosome 19 consortia.

They have published 12 original research papers, perspectives, and book chapters during this period, and have uploaded the data to Proteome Exchange: "PXD" with ID codes.

Plans for upcoming activities: We have planned to host the following chromosome face-to-face meetings in 2015: Tucson, Arizona, USA, March 13 and 14; Lund, Sweden, August 21 and 22; Galveston, Texas, USA, November.

Presentations at other international congresses:

Advances of Proteomics in Personalized Medicine and Translational Research, 15 and 16 November, 2014, Doha, Qatar, Lecture; ESPP Biobanking Symposium, 23–25 October, 2014, Leipzig, Germany, Lecture; CECE Symposium, 22 and 23 October, 2014, Brno, Czech Republic, Lecture; the 13th Human Proteome Organization World Congress, 5–8 October, 2014, Madrid, Spain, Lecture; JASIS 2014, 3–6 September, 2014, Keynote Lecture, Tokyo, Japan; ISBER 2014 (International Society for Biological and Environmental Repositories), 20–25 May,

Lecture; the 30th International Symposium on MicroScale Bioseparations, Univ. Pécs, Hungary, 20 April–1 May, 2014.;



**Iranian Team:
Chromosome Y**

Ghasem Hosseini Salekdeh

Grants: The Human Y Chromosome Proteome Project has become a key project of the Royan Institute. This guarantees a secure annual budget for Y-HPP projects, thanks to the support from both research and clinics. The second Human Y Chromosome Proteome Project will be held on May 7, 2015.

Workshops and Meetings: The results of following three projects will be submitted to a special issue of JPR:

1. The role of DDX3Y in the differentiation of NT2 cells to dopaminergic neurons
2. Comprehensive expression analysis of Y chromosome in diseased and healthy testes samples
3. The role of lysine-specific demethylase 5d, a Y chromosome-located gene, in prostate cancer cells

Ongoing projects are:

1. The role of Y chromosome genes in human ESC differentiation to myelin-producing cells and in myelination
2. The role of Y chromosome genes in human ESC differentiation into cardiomyocytes



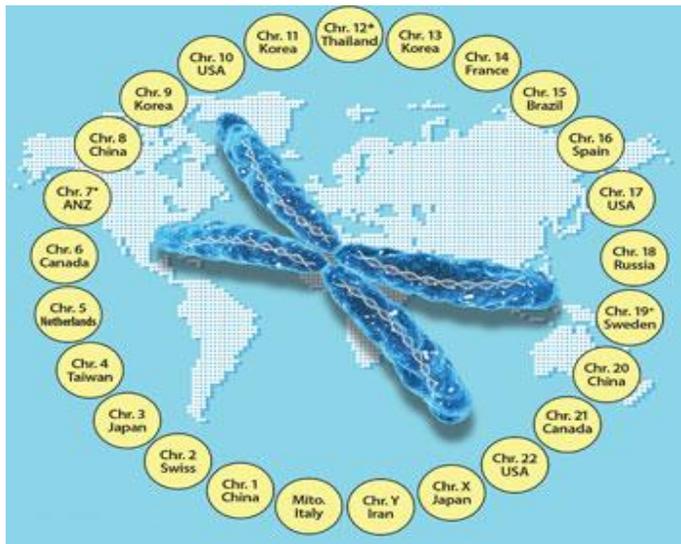
JPR Call for Papers

Journal of Proteome Research Special Issue 2015: The Chromosome-Centric Human Proteome Project

Submission Deadline: May 28th, 2015

The Journal is dedicating a third annual Special Issue to the Chromosome-Centric Human Proteome Project (C-HPP). Previous special issues appeared in January 2013 and 2014, with additional articles in June of each year [see www.thehpp.org/documents]. Manuscripts are now being accepted for the third C-HPP special issue, which will be published in September 2015 in conjunction with the 2015 HUPO World Congress 27 Sept-1 Oct in Vancouver, Canada.

Original, previously unpublished research, as well as review papers, is sought in all relevant areas. These include but are not limited to:



GUEST EDITORS

Young-Ki Paik (Yonsei University)

Gilbert S. Omenn (University of Michigan)

CO-EDITORS

Chris Overall (University of British Columbia)

William S. Hancock (Northeastern University)

- Reports on the protein parts list of individual chromosomes and groups of chromosomes, annotating known proteins and their isoforms and/or credibly identifying "missing proteins"
- Use of targeted proteomics, especially SRM and MS-SWATH, to validate chromosome-based protein findings
- Disease studies that utilize chromosome information, characterizing amplicons, identifying cis-regulated pathways or networks
- Integration of genomic, transcriptomic, epigenomic, or metabolomic data with proteomics, using advanced bioinformatics tools
- Proteomic studies of variants produced by alternative splicing (ASV) or coding non-synonymous single nucleotide polymorphisms (cSNPs).
- Generation and characterization of post translational modifications (PTM) libraries for chromosome-based protein sets
- Proteomic variation induced by chromosome abnormalities
- Novel findings on currently-unstudied reference cell lines and tissues
- Combined analyses with mass spectrometry and polyclonal or monoclonal antibodies for targeted studies of tissue samples
- Biomarker discoveries based on the identification of novel ASVs, PTMs or cSNPs in proteomic studies.

SUBMISSION PROCEDURE

Manuscripts must be submitted by May 28th, 2015 to be considered for this Special Issue. Manuscripts must be submitted electronically through the ACS Paragon Plus Environment online submission system. Prospective authors should prepare manuscripts according to the JPR guidelines for submission. Authors should indicate in the Comments to the Editor that the manuscript is submitted for publication in the Special Issue: The Chromosome-Centric Human Proteome Project. Database versions 09-19-2014 for neXtProt and 08-2014 for PeptideAtlas and false discovery rates of 1% at the protein level must be used. Raw data and metadata must be deposited via ProteomeXchange, with the PXD identifier in the final sentence of the Abstract.

REVIEW AND PUBLICATION PROCESS

All submitted manuscripts will be subject to initial editorial review to determine whether they are appropriate for the Special Issue. Those not meeting this criterion may be considered for a regular issue of the journal or returned. All relevant papers will go through full peer review. Due to the publication schedule, only papers requiring no or minor revision will be accepted for this Special Issue. Papers requiring major revision may be considered for later regular issues.

Wiki updates: Good Slide Materials for the C-HPP Workshops in 2014

- Full program and most of the presentations of Madrid and Segovia 10th C-HPP workshop is available at the C-HPP Wiki at <http://c-hpp.webhosting.rug.nl/tiki-index.php?page=11th+C-HPP+Workshop%2C+Appendix+5>
- Program and minutes of joint collaboration between chromosome 5, 10 and 19 can be find as example for join collaboration of multiple chromosome teams at <http://c-hpp.webhosting.rug.nl/tiki-index.php?page=Chromosome+19+Team+meeting%2C+Salamanca%2C+Spain%2C+October+2-4%2C+2014>
- Few groups are active to provide comprehensive information of their activities such as Chromosome 18 (Russia) [<http://c-hpp.webhosting.rug.nl/tiki-index.php?page=Chromosome+18>] Chromosome 13 (South-Korea) [<http://c-hpp.webhosting.rug.nl/tiki-index.php?page=Chromosome+13>].
- The full program of 9th C-HPP Workshop in Busan (South-Korea, March 26, 2014) including presentation is available at <http://c-hpp.webhosting.rug.nl/tiki-index.php?page=9th+C-HPP+Workshop%2C+March+26%2C+2014%2C+Busan%2C+South-Korea>
- The full program of 10th C-HPP Workshop in Bangkok (Thailand, August 9, 2014) including presentation is available at <http://c-hpp.webhosting.rug.nl/tiki-index.php?page=C-HPP+Workshop+Bioinformatics%2C+August+9%2C+2014>
- The full program of Inaugural C-HPP Human Proteome Missing Proteins Workshop in Sydney (Australia, 30th July 2014) including presentation is available at <http://c-hpp.webhosting.rug.nl/tiki-index.php?page=Inaugural+C-HPP+Human+Proteome+Missing+Proteins+Workshop%2C+30th+July%2C+2014>

Data Submission by the C-HPP Teams

Teams	Dataset ID or Bioinformatics Resources	Remarks
China (Chr 1, 8, 20)	The Chinese team submitted three data to PDX and has built iProx and CAPER. It is anticipated that more data sets will be deposited through PDX, making more useful resources for the community. They are also working with Prof. Ed nice for an integrating webserver for the C-HPP study.	
Netherland (Chr 5)	This team has submitted their data sets to PDX. PXD001034 & PXD001055 & PeptideAtlas http://www.peptideatlas.org/PASS/PASS00513	
Canada (Chr 6, 21)	<p>Human Erythrocytes</p> <p>In total, This team identified 1,234 proteins and proteoforms, including both cytosolic and membrane proteins at a protein level FDR of $\leq 0.7\%$ or at a peptide level FDR $\leq 1\%$ with at least two independent supporting spectra peptides. Using the N-terminomics procedure TAILS we identified 1,369 human erythrocyte natural and neo-N-termini. Multiple semi-tryptic N-terminal peptides exhibited improved mass spectrometric identification properties vs. the intact tryptic peptide enabling identification of 281 novel erythrocyte proteins and six missing proteins identified for the first time in the human proteome.</p> <p>The ProteomeXchange dataset identifier is PXD000434.</p> <p>Human Platelet Proteome</p> <p>This team published the identification of 2,938 proteins and 7,503 unique peptides from the human platelet proteome. They also identified for the first time in platelets 10 missing proteins. Most of N-termini (77%) were internal neo-N-termini: 105 were novel potential alternative translation start sites; and 2,180 represented stable proteolytic products.</p> <p>The ProteomeXchange dataset identifier is PXD000906.</p>	
France (Chr 14)	<p>Two MS/MS datasets were submitted: PXD000889 and PXD001178: "missing proteins mapping human chromosome 2 and 14 detected in different samples" (Carapito et al, submitted)</p> <p>Project name: "The human spermatozoa proteome"</p> <p>Project accession PXD001535. They have updated the C-HPP wiki http://c-hpp.webhosting.rug.nl/tiki-index.php?page=Chromosome%2014</p> <p>Another two MS/MS datasets were submitted by C. Pineau's group:</p> <p>PXD001535 - The human spermatozoa proteome - missing proteins mapping all human chromosomes detected in the spermatozoa (Jumeau et al., Submitted)</p> <p>PXD001575 - The Human testis proteome : embargo until manuscript submission</p> <p>We have updated the C-HPP wiki http://c-hpp.webhosting.rug.nl/tiki-index.php?page=Chromosome%2014</p>	In collaboration with Chr 2 team.
Spain (Chr 16)	Overall 270 files were sent to the ProteomeXchange repository through the recently incorporated ASPERA protocol. The files were grouped according to the analyzed cell lines generating four data sets, whose ProteomeXchange accession numbers are PXD000442, PXD000443, PXD000447, and PXD000449 for MCF7, Jurkat, Ramos, and CCD18 cell lines.	
Russia (Chr 18)	PASS00091 . PASS00092 . PASS00093 . PASS00094 . PASS00276 . PASS00277 . PASS00278 . http://kb18.ru/protein/matrix/212020 http://c-hpp.webhosting.rug.nl/tiki-index.php?page=Chromosome%2018 www.kb18.ru Master proteome of Chr18: http://kb18.ru/protein/matrix/212020	

12th C-HPP Workshop in Milan, Italy, June 23, 2015

(in conjunction with EuPA Congress and EXPO 2015)

General Info on the Workshop (Web sites: <http://www.eupa2015.org/>)

- Title: 12th C-HPP Workshop in Milan, Italy, June 23, 2015
- Theme: **Deciphering Proteome Complexity: Mitochondria and Olfactory Receptors (Proposed)**
- Organized by: 12th C-HPP Workshop Organizing Committee
Members: Andrea Urbani, Paola Roncada, Young-Ki Paik, Peter Horvatovich and Chris Overall and Lydie Lane (6 members)
- Local Hosts: Andrea Urbani and Paola Roncada
- Venue: EXPO2015 Site and Univ. of Milan
- Logistics: Best Western Antares Hotel CONCORDE
- Program Coordinators:
 Young-Ki Paik (Chair, C-HPP Side) and Andrea Urbani (President, EuPA side)
- Major Milestones to be announced to the Public:
 - Theme for EXPO: **"Feeding the Planet, Energy for Life"**
 - Media Vehicles: TV, Scientific Journals, Social Media, Business Journals and Newspaper

We believe this C-HPP Workshop in Milan would represent a unique showcase to communicate with the large general public as well as the EU Proteomics pioneering community, laying out the fundamental milestones of this important international project. In fact, during the proposed dates, the Italian Proteomics Association (previously I-HUPO) society will host the European Proteomics Association annual congress (EuPA2015). Moreover, all PI of C-HPP consortium have been invited to present the key milestones of the HPP in a dedicated event within the world exposition on 2015 (EXPO2015). In this light, the latest achievements and the actual deliverables made by the HPP group (C-HPP and B/D-HPP) during the past 4 years will make a landmark showcase to the large public through the mass-media communication. Expo 2015 will be an extraordinary universal event displaying tradition, creativity and innovation with a special focus on food and health. This event will bring together many themes that have already been handled in the past, and set them out anew in light of new global possibilities. EXPO2015 key themes will cover the following reference topic for the HPP:

Padiglione Italia EXPO 2015	UNIVERSITA' DEGLI STUDI DI MILANO, RETTORATE 24 - 26 JUNE 2015			Università degli Studi di Milano RECTORATE		
	23 JUNE	24 JUNE	25 JUNE	26 JUNE	27 JUNE	28 JUNE
AUDITORIUM CHPP WORKSHOP	10.30 am Opening M. UHLEN	Breakfast Seminar - 08.00 am Thermo Fisher Scientific	Breakfast Seminar - 08.00 am to be confirmed	Breakfast Seminar - 08.00 am to be confirmed	EuPA COUNCIL	MAIN AUDITORIUM and related rooms MORNING Protein crosslinking track SOCIAL PROGRAM Walking Tour in Leonardo da Vinci's footsteps AFTERNOON Protein HDX / covalent labeling track Experimental design in proteomics research POSTER PRIZE
	M. Fasano MT-TEAM L. Lane CHR 2 TEAM P. Horvatovich CHR 5 TEAM C. Overall CHR 6 TEAM M. Baker CHR 7 TEAM P. Yang CHR 8 TEAM L. LaBaer CHR 10 TEAM J.S. Yoo CHR 11 TEAM Y.K. Paik CHR 13 TEAM C. Pineau CHR 14 TEAM F. Corrales CHR 16 TEAM G. Omen CHR 17 TEAM A. Archacov CHR 18 TEAM G. MarkoVarga CHR 19 TEAM	PLENARY SESSION - MAIN AUDITORIUM - 09.00 am			EDUCATIONAL PROGRAM	
		MICROBIAL and VIRAL A. Aguzzi, C. Overall	NON HUMAN PROTEOMICS L. Foster, A. Gonzales-Bulnes	DIAGNOSTICS and TRANSLATIONAL G. Merlini, J.C. Sanchez	MAIN AUDITORIUM	
		Coffee Break and POSTER SESSIONS - 10.15 am				
		PARALLEL SESSIONS - 10.45 am				
		Integrated OMICS O. Kravkin	Proteomics of Metabolic and Cardiovascular disorders F.J. Corrales Izquierdo	Neuroproteomics P. Picotti		
		Proteomics of Microbiota and Ecosystems-Nutrition to be confirmed	iMOP A. Tholey	Food Allergy K. Hoffmann-Sommergruber C. Huber		
		Plant J. Jorrin Novo	Mitochondria & Metabolism M. Babu	Native MS and Biopharmaceutical COST BM 1403 - F. Sabbat, H. Urlaub, P. Barran, E. de Pauw		
		Luncheon Seminar - 1.00 pm WATERS	Luncheon Seminar - 1.00 pm ABSCIEX	Luncheon Seminar - 1.00 pm BRUKER		
		Coffee&Cake Seminar - 2.00 pm to be confirmed	Coffee&Cake Seminar - 2.00 pm Thermo Fisher Scientific	Coffee&Cake Seminar - 2.00 pm to be confirmed		
	PLENARY SESSION - MAIN AUDITORIUM - 03.00 pm					
	BIOINFORMATICS A. Nesvizhski, L. Martens	HUMAN PROTEOMICS G. Superti-Furga, M. Baker	CELLULAR and SYSTEMS BIOLOGY K. Lilley, H. Westerhoff			
	Coffee Break and POSTER SESSIONS - 04.15 pm					
	PARALLEL SESSIONS - 04.45 pm					
	Biobanking and Standardization G. Marko Varga	Cancer Y. K. Paik	Imaging MS L. Mc Donnell COST BM 1104 B. Bodenmiller			
	Proteoinformatics J. Cox, M. Vaudeil	PTMs Phospho G. Corrales	Structural Proteomics COST BM 1403 - A. Ashcraft, L. Jones, S. Cianferani, K. Rand			
	PTMs Glyco H. Clausen	EuPA INITIATIVES & COMPANY CLUB	EuPA Young Investigators POSTER PRIZE			
	PRESIDENT DINNER	SOCIAL DINNER	Concluding Remarks			

The workshop program will include scientific presentations, discussions on the most challenging topics in the C-HPP and exchange on the idea for the future direction based on the results so far achieved by the different Chromosome teams.

Vancouver HPP Meeting Announcement

(<http://hupo2015.com/program/hpp-sessions/>)

The Human Proteome Project (HPP) is an international project organized by the Human Proteome Organization (HUPO) that aims to revolutionize our understanding of the human proteome via a coordinated effort by many research laboratories around the world. It is designed to map the entire human proteome in a systematic effort using currently available and emerging techniques. Completion of this project will enhance understanding of human biology at the cellular level and lay a foundation for development of diagnostic, prognostic, therapeutic, and preventive medical applications. **Visit the HPP website for detailed information.**

The HUPO Annual World Congress provides an opportunity to convene face-to-face meetings and sessions across the full-range of HPP activities. Detailed information regarding all activities will be posted here once available.

Pre-Congress: HPP General Leadership Meeting

Sunday, September 27, 09:00 – 15:30, Vancouver Convention Centre

An opportunity to review all aspects of the HPP prior to the congress start. Participation in this meeting is included in registration.

13th C-HPP Workshop at a Glance

(Vancouver Convention Center)

SUN, Sept 27 09:00-15:30	MON, Sept 28 07:30-9:00	TUES, Sept 29 07:30-9:00 (Exhibit Hall)	WED, Sept 30 07:30-9:00	THR, Oct 1 07:30-9:00
HPP General Investigators Meeting (Chaired by Gil Omenn)	C-HPP (Part 1) • New Technology (Chaired by Bill Hancock)	C-HPP (Part 2) • Poster Session (Chaired by Lydie Lane) • Poster Awards Planned	C-HPP (Part 3) • PIC Meeting (Chaired by Young-Ki Paik) HPP Plenary Session (TBA)	Post-Congress HPP Workshop (TBA)

Post-Congress in Vancouver

Thursday, October 1, all day, Vancouver (venue TBA)

Co-Chairs: Gil Omenn, Jennifer van Eyk, Young-Ki Paik

Topics for the full day HPP Session on Thursday, October 1 will include but are not limited to:

- Highlights from the Vancouver Conference for the next stage of the HPP
- Collaborative initiatives of C-HPP, B/D-HPP, and the two consortia together
- Special insights from the HPP Resource Pillars: Antibodies, Mass Spectrometry, Knowledgebase
- Updated Metrics for the HPP
- Cross-analyses with antibodies and MS for the same specimens
- Deliverables for 2015-2016

Obituary

In the last summer we lost our friend, Juan Pablo Albar, PI of Chromosome 16 and a newest EC member of C-HPP. All of our colleagues and friends have been deeply saddened by this shocking news. Thus, we would like to remember and thank him for his great service and contribution to our C-HPP consortium as well as proteomics community. The HUPO2014 Congress in Madrid was dedicated to the memory of Juan Pablo Albar.

Juan Pablo Albar (1953-2014)

Lover of great challenges, tireless explorer, Juan Pablo Albar has started his last expedition, that in which none of us can follow him. The climber when miles climbed up, the marathon man when miles were ahead, the adventurer of the human proteome, in any case, the best known biochemist of Antarctica, passed away the night from Friday to Saturday. Between proteins and incredible adventures he left us quietly, like the whisper of a soft wind in the lonely mountains. In Juan Pablo, science, sport and adventure merged into their innermost essence, the nobility, the tackling of new challenges, the generation of new knowledge and, ultimately, the extension of the dimension of the human being. There were tiny boundaries between his scientific and explorer sides and this special nature made him to go through the ice of the Antarctica reaching the South Pole to send us all a "warm" invitation to work with him on one of the biggest challenges of Biology, the Human Proteome.



Degree in Chemistry by the Complutense University of Madrid, where he obtained the Ph.D. in 1981. After several years working in the private environment, Juan Pablo started his scientific activity at the National Biotechnology Center (CSIC), becoming one of the pioneers of proteomics in Spain. He was a CSIC Senior Scientist, Head of the Proteomics facility, Biomolecular and Bioinformatics Resources Platform Coordinator, member of the Human Proteome Organization (HUPO) Council, member of the Chromosome-HPP Executive Committee, member of the New Technologies Committee of the European Proteomics Association and member of the Spanish Proteomics Society Board. His work has been published in more than 160 scientific papers and was a source of inspiration for many scientists in the field.

Our dear friend has left behind, beyond the sadness of the moment, a legacy of positivity and strength that will help us to go ahead with all the enthusiasm that Juan Pablo used to have. We should then face the future with optimism and work together to make real all projects and dreams he already started.

Appendix

Special Report by Chr 6 Teams

| Missing proteins found

The Overall Lab has embarked on a Missing Protein Hunt utilizing two distinct strategies. First, the extent and rate of proteoform expression depends on cell type, location, stimulus and often human developmental stage. This can be exploited to identify missing proteins and new isoforms. Second, monitoring in vivo proteolytic cleavage of a protein can not only be exploited to observe protein proteoform PTMs but also to obtain proof for the expression of missing proteins. Proteolysis alters the protein sequence and results in neo-N termini and hence novel semi-tryptic N-terminal peptides upon tryptic digestion in bottom-up proteomics. We therefore hypothesized that some of these N terminal peptides would exhibit beneficial m/z, ionization and fragmentation properties over their fully tryptic counterparts, rendering these peptides and cognate proteins more amenable to mass spectrometric identification. Thus, we aim to analyze less commonly studied or accessible cells and tissues by high throughput N terminomics using terminal amino isotopic labeling of substrates (TAILS) in order to identify rare cell-restricted or developmental-stage restricted expression of proteins and their speciation. In addition, TAILS has the advantage of being able to capture all naturally blocked and unblocked N termini.

| Human Erythrocytes

In total, we identified 1,234 proteins and proteoforms, including both cytosolic and membrane proteins at a protein level FDR of $\leq 0.7\%$ or at a peptide level FDR $\leq 1\%$ with at least two independent supporting spectra peptides. Using the N-terminomics procedure TAILS we identified 1,369 human erythrocyte natural and neo-N-termini. Multiple semi-tryptic N-terminal peptides

exhibited improved mass spectrometric identification properties vs. the intact tryptic peptide enabling identification of 281 novel erythrocyte proteins and six missing proteins identified for the first time in the human proteome.

The ProteomeXchange dataset identifier is <PXD000434>.

We obtained evidence for six of the missing proteins that had previously escaped mass spectrometric detection (neXtProt evidence level PE2). Two of these were encoded on chromosome 2 and one each on chromosomes 3, 9, 19 and 21. Five of these proteins, the cancer/testis antigen 75 (CT75, Q6PK30, 2q36.1), the uncharacterized protein C9orf84 (CI084, Q5VXU9,9q31,3), protein FAM43A (FA43A, Q8N2R8, 3q29), the PR domain zinc finger protein 15 (PRD15, P57071, 21q22.3) and the leucine-rich repeat transmembrane protein 1 (LRRT1, Q86UE6, 2p12) were found due to our TAILS workflow and database search strategy. These proteins were identified by peptides representing N- or C-termini of proteolytically processed protein species, which therefore matched the specificity of the protease used for proteome digest, i.e. trypsin or GluC, only on one side. We specifically enriched and searched such "semi-specific" peptides that would be missed in conventional "specific" database searches. The identification of FAM43A and LRRT1 furthermore benefitted from the use of GluC as an alternative digestive protease, since those termini were not found in tryptic digests. Notably, LRRT1 was identified by 58 high confidence PSMs across all 4 biological replicates. Identification of the sixth of the missing proteins, intercellular adhesion molecule 4 (ICAM4, Q14773, 19p13.2), was enabled by cell type specific analysis of erythrocytes.

We identified 1,369 N-termini at <1% FDR, of which 45% had a free alpha-amine at their N-terminus, 53% were acetylated and 2% had an N-terminal pyroGlu derived by conversion of Glu. Notably, only 32% of the observed N-termini could be explained by known, annotated protein starts. These included termini starting with intact or removed initiator-Met (original N-termini) or at sites consistent with maturation by signal peptide or propeptide removal. An additional 4% (51 N-termini) started at or directly after a genome-encoded internal Met and had the sequence properties favoring initiator Met retention or removal, respectively, indicating alternative translation initiation sites. 84% (43 N-termini) of those alternative translation initiation sites were acetylated with similar sequence determinants and in similar proportion as N-termini starting at position 1 and 2 of the genome-encoded protein sequence, indicating co-translational acetylation. Remarkably, the majority (64%) of all identified N-termini mapped to positions within the genome-encoded protein sequences due to endoproteolytic processing.

| Human Platelet Proteome

We published the identification of 2,938 proteins and 7,503 unique peptides from the human platelet proteome. We also identified for the first time in platelets 10 missing proteins. Most of N-termini (77%) were internal neo-N-termini: 105 were novel potential alternative translation start sites; and 2,180 represented stable proteolytic products.

The ProteomeXchange dataset identifier is PXD000906.

Combining preTAILS and TAILS data we identified with high confidence (FDR 1%) 10 peptides from 10 proteins that were previously classified as missing with little or no evidence at protein level. These included 2 "uncertain" and 8 detected-only-as-transcripts proteins, expression of

two of which was validated by western blotting. Among the 10 peptides 7 were N-termini, with 5 detected by TAILS. Such proteins, previously missed by shotgun proteomics, could be identified by TAILS due to its 5 orders of magnitude dynamic range and enrichment for different semi-tryptic peptides that have altered m/z, ionization and fragmentation properties compared to the tryptic peptides analyzed by traditional proteomics.

Among the 2,960 annotated N-termini identified 676 (23%) represented natural protein N-termini, including naturally acetylated (397 or 59%) and free N-termini (279 or 41%). Acetylated peptides were mostly represented by natural N-termini (397 or 59%), whereas internal neo-N-termini prevailed among the free N-termini (2,010 or 88%). The majority of natural N-termini started after methionine excision (125 or 56%; 296 or 74% for free and acetylated natural N-termini, respectively). Noteworthy, many new instances of N-terminal PTMs that were not previously observed were identified. Internal N-termini that mostly are products of proteolysis accounted for 77% (2,284) of all N-termini.

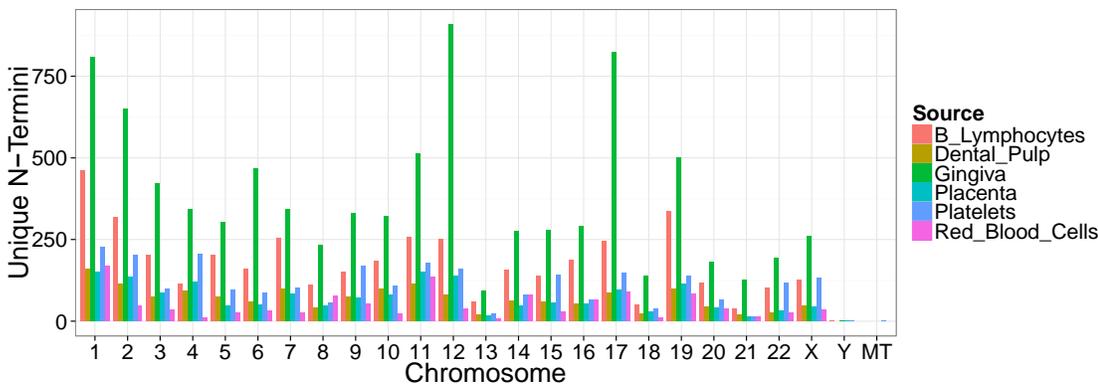
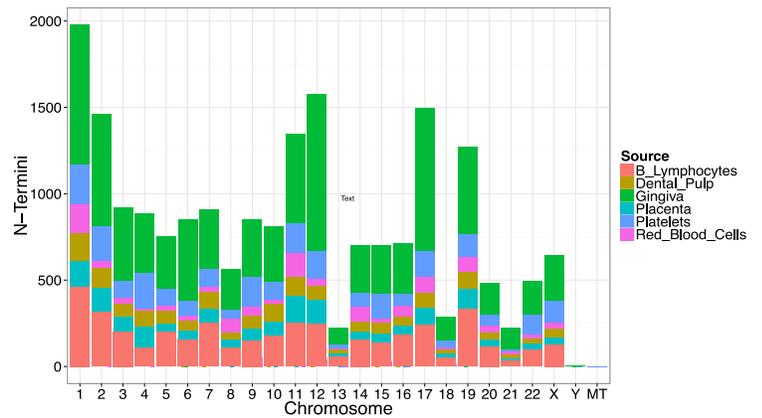
As shown in the accompanying charts and graphs we have extended our studies to other human tissues including unusual tissues including dental pulp (the "nerve" of teeth), dental gingiva (gums), B lymphocytes, placenta, and heart valves. Numbers of missing proteins are presented in the accompanying tables. Notably **TAILS identified 346 missing proteins in total.**

Together, these results validate our strategy of applying N-terminomics on specific cells and tissues that are not often analyzed by proteomics for the detection of missing proteins within the C-HPP.

Total Proteins Statistics

Total Proteins and N-Termini Discovered by TAILS and PreTAILS in: Placenta, Platelets, B Cells, Erythrocytes, Dental Pulp and Gingiva

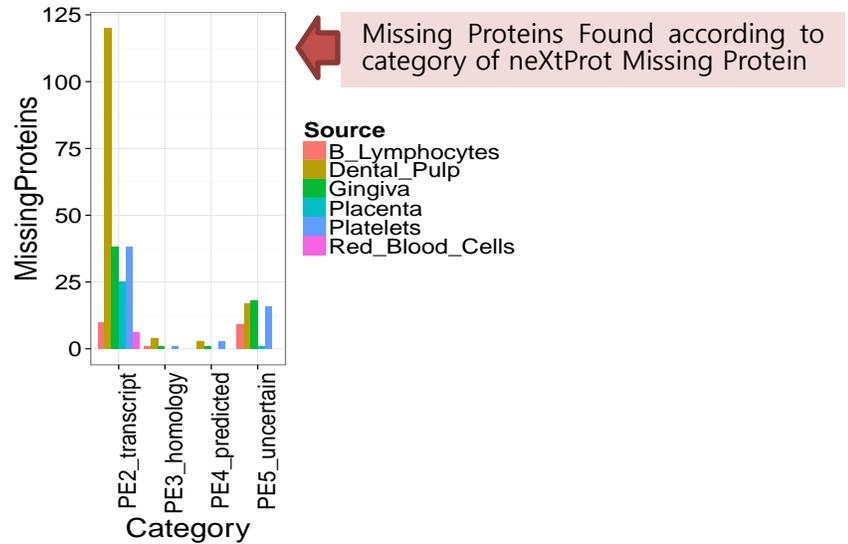
Chr	Unique Proteins	Unique Protein N-Termini	Proteins (Identified by TAILS)	Proteins (Identified by PRETAILS)
1	957	1661	648	806
2	608	1241	425	483
3	492	776	336	410
4	328	704	207	270
5	417	642	279	345
6	388	759	258	306
7	385	740	275	318
8	284	482	192	232
9	353	742	243	284
10	358	666	220	303
11	568	1079	389	469
12	480	1395	331	397
13	151	188	92	128
14	310	594	206	267
15	293	593	197	242
16	406	593	260	342
17	535	1324	368	460
18	138	239	103	110
19	599	1087	414	501
20	266	404	194	210
21	86	192	54	71
22	229	421	152	200
MT	1	1	1	1
X	305	549	202	245
Y	6	5	3	5
TOTAL	8943	17077	6049	7405



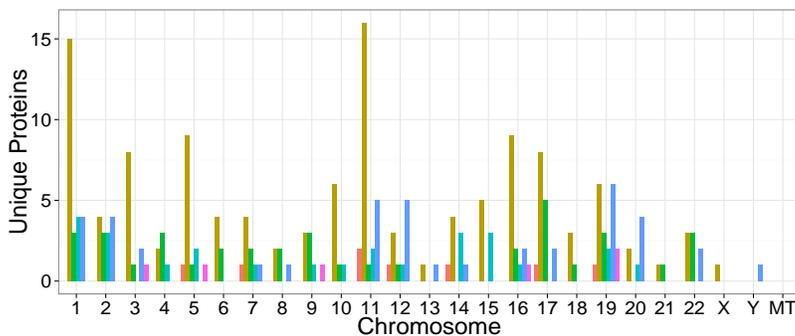
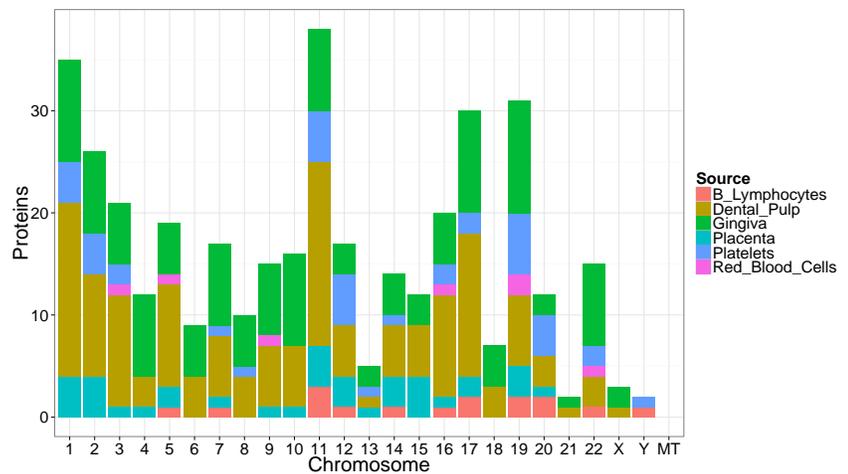
Missing Proteins Statistics

Human Proteome-Wide Missing Proteins Discovered by TAILS and PreTAILS in: Placenta, Platelets, B Cells, Erythrocytes, Dental Pulp and Gingiva (Overall Lab)

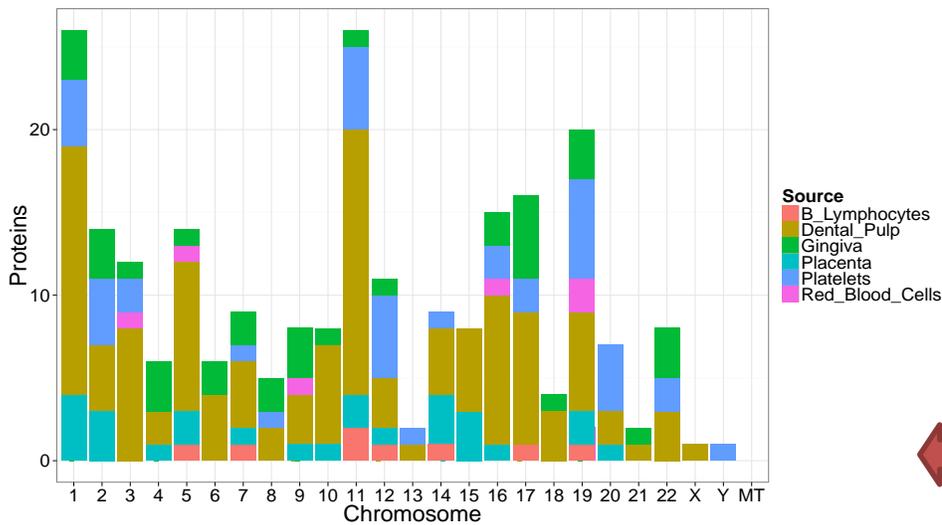
Chr	Unique Proteins	Protein N-termini of Missing Proteins	TAILS	Pre-TAILS
1	28	14	13	24
2	22	16	13	13
3	19	11	11	11
4	11	7	7	6
5	17	7	5	12
6	9	3	3	6
7	16	9	9	9
8	9	6	6	5
9	15	7	7	8
10	14	8	8	7
11	36	17	16	26
12	12	10	9	8
13	4	4	3	2
14	13	6	6	9
15	11	4	4	8
16	18	7	5	15
17	26	19	16	16
18	7	3	3	4
19	30	19	19	19
20	12	9	9	7
21	2	0	0	2
22	11	12	8	6
X	3	2	2	1
Y	1	2	1	1
TOTAL	346	202	183	225



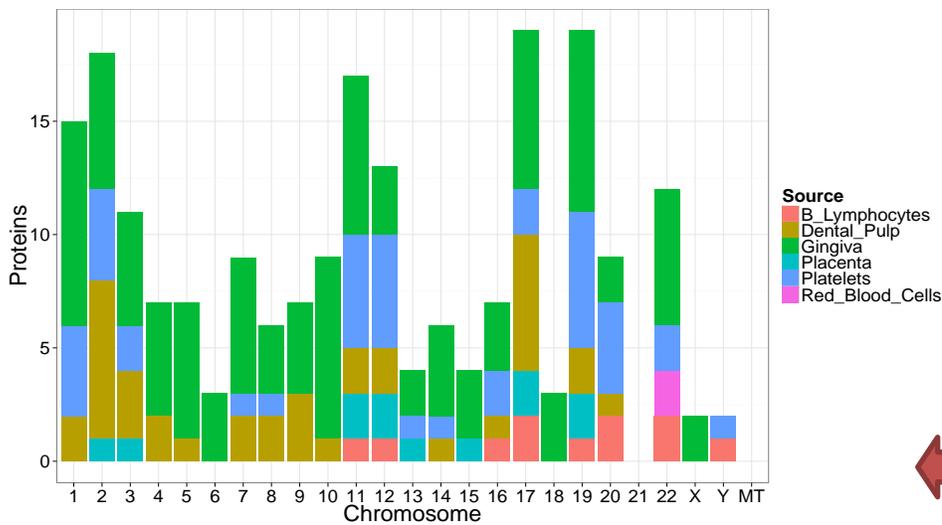
All Missing Proteins found by TAILS + PreTAILS according to tissue



All Unique Missing Proteins found by PreTAILS according to tissue



Missing Proteins found by PreTAILS in Unusual Tissue Sources



Missing Proteins found by TAILS in Unusual Tissue Sources

Special Report by Chr 18 Teams

Roadmap: Roadmap available at http://www.ibmc.msk.ru/content/intelligence/RHUPO_Roadmap.pdf

- The size of the human proteome requires estimates of number of protein species (proteome width) and number of copies of the same protein molecules in a biosample (proteome depth)
- Master proteome of a single chromosome is the results of the identification and measurement of all master proteins encoded by the selected chromosome and expressed in the selected type of biological

material, there master protein is the primary translation of the coding sequence resembling at least one of the known protein forms, coded by the gene.

- Complete human proteome in this context is the result of (1) constellation of master proteomes of different chromosomes and types of biological material and (2) data about modified (AS, SAP, PTM) protein species.

- In case PTM and SAP can occur both in the proteins encoded by gene with canonical sequence and splice variants, the number of human protein species (proteome width) was estimated based on the NeXtProt (v2013) data as 1.1 million. Using the assumption that PTMs appear exclusively in sequences of master proteins, but not in splice variants, the width of the human proteome contains approximately 650 000 protein species. If all types of modifications (AS, SAP and PTM) are independent events, the potential proteome width is approximately 8.5 million.
- One of the most sensitive SRM technology available detection with sensitivity limit of 10⁻¹⁸ M for BSA standard, but additional sample treatment with irreversible binding onto BrCN-Sepharose beads allowed a sensitivity of 10⁻¹⁸ M to be achieved (the proteins CYP102 and BSA was used for calibration), which corresponds to one protein copy/μl of blood plasma.
- Operating within an ultra-low concentration range, it is convenient to refer to protein copies rather than concentrations because it allows comparison of results of transcriptomic and proteomic experiments

Special expertise

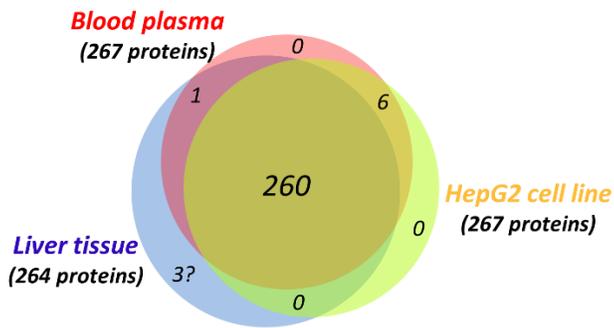
- Comparative ranking of chromosomes based on post-genomic data shows that all chromosomes are not much different from each other except ChrY and ChrMt due to their shortest length. Chr18 for the Russian portion of the HPP was selected based on the combination of proposed criteria. (Ponomarenko et al., 2012)
- Meta-analysis of proteomics data from publication and databases for Chr18 shows that only 37%, 43% and 7% of total number of master proteins coding on chromosome detected in blood plasma, liver tissue

and HepG2 cells, respectively, and the very few of them measured.

- Detection of ultra-low copied protein species in concentrations as low as 10⁻¹⁸ M using BSA standard (e.g. 1 molecule per 1 μL of plasma). These low-copied proteins are enriched from the excessive sample volume by the irreversible binding to the beads. See Kopylov et al., 2013 for details.
- Quantitative correlations of transcriptome-to-proteome data to evaluate the quality of measurements obtained by RNAseq, survey and targeted proteomics. (Ponomarenko et al., 2014)
- Development of technology for direct molecular fishing on paramagnetic particles for protein interactomics (Ivanov et al., 2014)
- Cataloguing the non-canonical protein species (proteoforms) by in-depth analysis of transcriptome data (Shargunov et al., 2013) and prediction of SAP and AS (Lisitsa et. al., 2014).

Major achievements

- Transcriptome profiling for chr18 in liver tissue and HepG2 cell line using RNA-Seq (Illumina and SOLiD) and RT-PCR protocols.
- Proteome profiling for 276 master proteins coding on Chr18 in blood plasma, liver tissue and HepG2 cell line using SRM.



Venn diagram representing the 269 proteins of chromosome 18 detected by SRM in three types of biomaterial.

- Construction Gene-Centric Knowledgebase(kb18) and SRM Registry (pikb18.ru) database for post-genomic data of Chr18.

LIST OF 'MISSING' MASTER PROTEINS FOR CHR18 IN RESULTS OF OUR SRM PROTEOME PROFILING

GENE	AC	Protein name	Our results of transcriptome profiling	
			LIVER	HepG2 cell line
C18ORF64	J3KSC0	Putative uncharacterized protein encoded by LINC01387	Not expression	Not expression
LDLRAD4	O15165	Low-density lipoprotein receptor class A domain-containing protein 4	Expression	Expression
PPP4R1	Q8TF05	Serine/threonine-protein phosphatase 4 regulatory subunit 1	Expression	Expression
ST8SIA5	O15466	Alpha-2,8-sialyltransferase 8E	Below thresholds	Below thresholds
SERPINB10	P48595	Serpin B10	Below thresholds	Not expression
TTC39C	Q8N584	Tetratricopeptide repeat protein 39C	Expression	Expression



Chromosome-centric
Human Proteome Project

| HQ Office:

Yonsei Proteome Research Center, Yonsei University
50 Yonsei-ro Sudaemoon-ku, Seoul, 120-740, Korea (www.c-hpp.org)

| Chair: Young-Ki Paik

| Co-chairs: William S. Hancock, György Marko-Varga (2012-2014), Lydie Lane

| Newsletter Editor: Young-Ki Paik

| Editorial Staff: Jin-Young Cho and Peter Horvatovich