



The International Consortium of the Chromosome-Centric Human Proteome Project

10th C-HPP Workshop

Proteome Data Management and Identification of Missing Proteins

9 August 2014

**SiMR Building, Siriraj Hospital, Mahidol
University,
2 Wanglang Road, Bangkoknoi, Bangkok,
Thailand**



1. Purpose and Goals of 10th C-HPP Workshop

A. Why do we need this workshop?

Given the recent publication of a large dataset for the HPP in Nature (Pandey lab and Kuster lab: Congratulations to both teams!), data management (verification, validation, peptide level confidence with FDR <1% etc.) became the central issue of the community. As to be expected claims based such large data sets generate issues, such as what information is deposited in ProteomeXchange, the amount of meta-data available for each sample and the availability of MS/MS spectra in the case of novel protein identifications. To keep momentum going after the Busan workshop, which laid out the future direction of the C-HPP development, we are now in a very crucial position to tackle all those pending issues associated with data managements in the course of missing protein identification that should be done in more standardized and coordinated manner. An important discussion will be on the validation process in which a protein identification goes from a laboratory result to an accepted publication to acceptance by PeptideAtlas and GPMDB and is finally included in neXtProt. The organizing committee members propose that this meeting would have a regional flavor but get international chromosome teams together to work on data management which deals with data deposition and curation.

We hope this meeting will bring all those people who are working large databases together and share a common issue as to how each group can provide up-to-date information on each chromosome. For example, the dataset A is the list of missing proteins, from which we have already lot of information but the goal would be to look at the other datasets and see if it is possible to get information in the same way for the missing proteins. The best thing would be to organize a roundtable discussion to address these questions and discuss which information is possible to get from the currently available resources and what is need to be done in the future to get those information. It is quite obvious for us to resolve these issues through this workshop.

B. What things to be achieved through this workshop?

1) Proteomic Dataset: How can we make those deposited data through ProteomeXchange or published data available for the consortium members? Even though many groups wish to share the deposited data, but the process is not so easy. For example, some of data deposited to PDX is not accessible to those annotators (e.g., Eric Deutsch at the PeptideAtlas, Lydie Lane at neXtProt, or Ron Beavis at GPMDB).

-What can we do for such inaccessible datasets in the public DB?

-Do we need to ask people to place a link with PXD identifier on the Wiki in order to see which chromosome team placed which datasets online (for sharing)?

2) Proteogenomic Dataset: We have even more difficulties to get access or deposit proteogenomic dataset which include RNAseq and other types of genetic data. Where are other types of data deposited and how they can be linked together?

We definitely need to establish some versatile programs to handle these datasets.

3) Enhanced Connections: Although a local upload site is fine, connection from outside to the uploading site seems problematic which may be due to the internet speed going through the national firewall for sending large dataset. Therefore our goal is to figure out how make this connection faster and accessible. For example, we should make the connections faster between a local server and ProteomeXchange server.

Thus, we would like to welcome all of you to sign up this meeting and join us in the active discussion of all data management issues.

2. Outline of the Workshop

- **Title:** 10th C-HPP Workshop in Bangkok, Thailand
- **Date & Time:** August 9, 2014 (08:30 - 18:30)
- **Venue:** SiMR Building, Siriraj Hospital, Mahidol University, 2 Wanglang Road (formerly Prannok Road), Bangkoknoi, Bangkok, Thailand

- **Organizing Committee:**

- Convener: **William S. Hancock**
- Program Coordinator: **Young-Ki Paik**
- Administrator: **Peter Horvatovich**
- Local Host: **Visith Thongboonkerd**

- **Theme: “Proteome Data Management and Identification of Missing Proteins”**

This workshop is open to all C-HPP PIs and their co-workers and will serve as a link between the two to keep the C-HPP activities more interactive.

- **Funding Support:**

- This workshop is kindly supported by:
- Faculty of Medicine Siriraj Hospital, Mahidol University
 - The Thailand Research Fund (TRF)

- **Transportation:**

A shuttle service (kindly supported by Faculty of Medicine Siriraj Hospital, Mahidol University), to and from the workshop venue, will be provided only at Miracle Grand Convention Hotel.

Please wait at the hotel lobby on August 9th, and the shuttle will depart at 7:00 am sharp (Breakfast at the hotel starts at 6:00 am).

- **Administration:**

All matters related to letter of invitation, workshop venue, and accommodation will be handled by the local host, Prof. Visith Thongboonkerd (vtongbo@yahoo.com).

Scientific Program

(Note: This program has been updated from time to time and can be downloaded at www.c-hpp.org, www.thehpp.org, www.hupo.org, and C-HPP Wiki site <http://c-hpp.webhosting.rug.nl>)

08:30 - 08:50 Welcome & Opening Remarks

Moderator: Visith Thongboonkerd (Mahidol University, Bangkok, Thailand)

Siriraj Hospital VDO Presentation
Dean, Faculty of Medicine Siriraj Hospital, Mahidol University
Chair, HUPO C-HPP Consortium (Young-Ki Paik)

08:50 - 09:40 Session 1: Introduction and Opening Invited Talk

Chair: Visith Thongboonkerd (Mahidol University, Bangkok, Thailand)

Part A: Briefings on the C-HPP Progress (10 min each)

Update & Future Plans on the C-HPP Publications in 2014 and 2015

William Hancock

(Co-Chair of C-HPP EC & Editor-in-Chief of JPR) (Northeastern University, Boston, MA, USA)

Update on the Bio-banks within the HPP

Peter Horvatovich

(Secretary General of C-HPP Consortium) (University of Groningen, The Netherlands) (for György Marko-Varga)

Part B: Opening Invited Talk (30 min)

Integrative Proteo-Genomic Analysis of Early Onset Gastric Cancer

Sanghyuk Lee

(Ewha Womans University, Seoul, Korea)

09:40 - 10:25 Session 2: Invited Talks on

“Progress on the Proteogenomics/Transcriptomic Data Productions”

Part I (15 min each)

Chair: William Hancock (Co-Chair of C-HPP EC & Editor-in-Chief of JPR)
(Northeastern University, Boston, MA, USA)

Looking for missing proteins: an enlightenment from the analysis of free-mRNA and RNC-mRNA data

Siqi Liu

(Beijing Institute of Genomics and BGI-Shenzhen, China)

Integration of ENCODE, Human Body Map and Proteomics Data in a Devoted Dashboard.

Victor Segura

(CIMA, University of Navarra, Pamplona, Spain)

Progress of high coverage proteomics study on C-HPP in China

Ping Xu

(BPRC, Beijing, China)

10:25 - 10:45 Coffee Break

**10:45 - 11:30 Session 3: Invited Talks on
“Progress on the Proteogenomics/Transcriptomic Data Productions”
Part II (15 min each)**

*Chair: William Hancock (Co-Chair of C-HPP EC & Editor-in-Chief of JPR)
(Northeastern University, Boston, MA, USA)*

Proteogenomic analysis of the human chromosome 9-encoded genes

Je-Yoel Cho

(Seoul National University, Seoul, Korea)

***Comprehensive characterization of a liver tissue and HepG2 cells
transcriptoproteome for human chromosome 18***

Andrey Lisitsa

(Institute of Biomedical Chemistry RAMS, Moscow, Russia)

Human Y Chromosome Proteome Project: 2014 update

Ghasem Hosseini Salekdeh

Royan Institute for Stem Cell Biology and Technology, Tehran, Iran

11:30 - 12:30 Session 4: Strategy for Identifying Missing Proteins (I)

Part I (15 min each)

*Co-Chairs: Mark Baker (HUPO President-Elect)
(Macquarie University, Sydney, Australia)
Ping Xu (BPRC, Beijing, China)*

Metrics and strategy for identifying missing proteins

Gilbert Omenn

(Chair of HPP EC) (University of Michigan, Ann Arbor, MI, USA)

A Combined strategy for Identifying missing proteins

Young-Ki Paik

(Chair, C-HPP Consortium, Yonsei Univ., Seoul, Korea)

***Identification of missing proteins by profiling of specific tissues and
cell lines, lessons learnt in the hunt for olfactory receptors located on
chromosome 17 and identification of potential co-expression events***

William Hancock

(Co-Chair of C-HPP EC & Editor-in-Chief of JPR) (Northeastern
University, Boston, MA, USA)

12:30 - 13:30 Lunch

13:30 - 15:00 Session 5: Strategy for Identifying Missing Proteins (II)

Part 2 (15 min each)

*Co-Chairs: Mark Baker (HUPO President-Elect)
(Macquarie University, Sydney, Australia)
Ping Xu (BPRC, Beijing, China)*

Sydney Inaugural HPP Missing Proteins Workshop Report

Mark Baker

(HUPO President-Elect) (Macquarie University, Sydney, Australia)

Identification of missing proteins in aggregated proteins

Qing-Yu He

(Jinan University, Guangzhou, China)

Missing proteins hidden in the membrane proteome

Yu-Ju Chen

(Institute of Chemistry, Academia Sinica, Taipei, Taiwan).

To conquer the last hard-core of missing proteome

Pengyuan Yang

(Fudan University, Shanghai, China)

Size of master proteome expressed by single chromosome in different tissues and cells if there are missing master proteins in proteome?

Alexander Archakov

(Institute of Biomedical Chemistry RAMS, Moscow, Russia)

15:00 - 16:00 Session 6: Practical Aspects of Individual Data Management (60 min)

*Co-Chairs: Peter Horvatovich (Secretary General of C-HPP Consortium)
(University of Groningen, The Netherlands)*

*Andrey Lisitsa (Institute of Biomedical Chemistry RAMS,
Moscow, Russia)*

Part A: Proteomic Dataset with PXD:

Special Invited Speaker (Skype)

Juan Antonio Vizcaino

EMBL-EBI, Hinxton, Cambridge, UK

Q1. How to make deposited through ProteomeXchange or published data available for the consortium members as well as public DB managers (GPMDB, neXtProt, PeptideAtlas and ProteinAtlas)?

Q2. What can we do for such inaccessible datasets in the public DB?

Q3. Do we need to ask people to place a link with PXD identifier on the Wiki in order to see which chromosome team placed which datasets online (for sharing)?

Q4. Sharing any experience with public databases (PeptideAtlas?, ProteinAtlas? ProteomeXchange? GPMDB? neXtProt?)

Part B: Proteogenomic Dataset:

Q1. How to make easy access to or deposit proteogenomic dataset including RNAseq and other types of genetic data?

Q2. Where are these types of data deposited and how can they be linked together?

Part C: Difficulties in Connections:

Q1. How to make the connections between local server and central DBs much faster and accessible (e.g. local server and ProteomeXchange)?

Q2. Strategy for data sharing within HPP: What are the bottlenecks or obstacles?

Q3. Partnering with B/D-HPP teams: What is the full list of active teams with their research interests and resources?

Q4. Integration with the overall HPP plan: How?

Q5. Sharing the information about sample sets analyzed: How?

Q6. Sharing reagents such as antibodies, expression vector clones, cell lines, etc.: How?

Q7. Disease/biology collaborations based on pathways, gene sets or amplicons: How?

16:00 - 16:20 Coffee Breaks

16:20 - 17:50 Session 7: Roundtable Discussion on

“Problems and Solutions for the Large Dataset Validations” (90 min)

*Co-Chairs: William Hancock (Co-Chair of C-HPP EC & Editor-in-Chief of JPR)
(Northeastern University, Boston, MA, USA)*

*Gilbert Omenn (Chair of HPP EC)
(University of Michigan, Ann Arbor, MI, USA)*

Panelists (Country Representatives):

Mark Baker	(Australia)
Siqi Liu	(China)
Ghasem Hosseini Salakdeh	(Iran)
Tadashi Yamamoto	(Japan)
Jong Shin Yoo	(Korea)
Peter Horvatovich	(The Netherlands)
Andrey Lisitsa	(Russia)
Victor Segura	(Spain)
Visith Thongboonkerd	(Thailand)
Yu-Ju Chen	(Taiwan)
Ravi Sirdeshmukh	(India)
Terence Poon	(Hong Kong)
Maxey Chung	(Singapore)

Part A: What are the most burning issues?:

Q1. Proof of validity of a large dataset available in the public DB: Production, repository, exchange, quality, and reproducibility

(e.g., the protein identifications with higher rate of false discovery than the C-HPP consensus methods.)

Q2. How to make major analytical assessments of the large datasets to be incorporated into well annotated data resources (GPMDB, PeptideAtlas, neXtProt, ProteinAtlas).

Q3. How to make corrections if the spectra do not support the assignments?

Q4. How well the definition of datasets are suited to store and provide information from already data resources/infrastructure (GPMDB, Nextprot, ProteomeXchange etc) and if information is not present, what resources are available to implement these needs. Role of Wiki and individual groups in Data and Information management/sharing

Part B: Potential Solutions-Standard Protocols Adopting a Multi-level Data Validation (proposed by Bill Hancock):

1. Individual scientists' laboratory generates data using accepted practices and submit data to ProteomeXchange
2. Work is reviewed and accepted (or not) by a high impact journal and published (work now has published status).
3. MS data is reviewed by PeptideAtlas and/or GPMDB if OK the identification is labeled as provisional (correct word?)
4. Final step where neXtProt accepts the identification along with other data such as sample curation
5. The same process is followed for tissue expression studies where the final arbitrator is ProteinAtlas

17:50 - 18:20 Session 8: Amended Version of Long-Term Plans and Deliverables

Co-Chairs: Young-Ki Paik (Chair of C-HPP EC)
(YPRC, Yonsei University, Seoul, Korea)
William Hancock (Co-Chair of C-HPP EC & Editor-in-Chief of JPR)
(Northeastern University, Boston, MA, USA)

Drafted version of updated long-term plans and milestones along with HPP community: Free Discussion following the presentation of plan.

(Based on the feedbacks given by KC-HPP (Chr 9, 11, 13), Juan Pablo Albar (Chr 16), Mark Baker (Chr 7), Alex Archakov (Chr 18), Jerome Garin & Yves Vandenbrouck (Chr 14) and Lydie Lane & A. Bairoch (Chr 2); Reflection of the recent Nature papers on the milestones

Future Plans and Perspectives for the C-HPP

- 11th C-HPP Workshop Plans in Madrid 2014 (Oct 5-8, Oct 9 Segovia)
- 12th C-HPP Workshop Plans in Milano 2015 (June 23-24, 2015)
- 13th C-HPP Workshop Plans in Vancouver (Sept 26-30, 2015)
- 14th C-HPP Workshop Plans in Shanghai (Spring, 2016) (TBA)
- 15th C-HPP Workshop Plans in Taipei (Sept 18-22, 2016)
- 16th C-HPP Workshop Plans in Tehran, Iran (Spring, 2017) (TBA)
- 17th C-HPP Workshop Plans in Glasgow, UK (Sept 2018) (TBA)

18:20 - 18:30 Conclusions

William Hancock

(Co-Chair of C-HPP EC & Editor-in-Chief of JPR)

(Northeastern University, Boston, MA, USA)

18:30 - 21:30 PI/Speaker Dinner

(Sponsored by Faculty of Medicine Siriraj Hospital, Mahidol University)

-----**The end of workshop**-----

4. Excursion: A Half-Day Tour (August 10, 2014)

Complimentary to all participants.

Kindly supported by Faculty of Medicine Siriraj Hospital, Mahidol University.

Tentative Program:

8:30 - 10:00 Grand Palace & Temple of the Emerald Buddha (Wat Phra Kaew)

10:30-12:00 Vimanmek Mansion

The tour should be ended no later than 13:00 and the shuttle will send all back to Miracle Grand Convention Hotel. A lunch box will be served on the shuttle bus.

Please wait at the hotel lobby on August 10th, and the shuttle will depart at 7:15 am sharp.

Those who plan to travel after the tour can check-out on August 10th and leave luggage on the shuttle bus during the tour.

After the tour, there will be another complimentary shuttle to go directly to Suvarnabhumi International Airport. Please register for the shuttle on Aug 9th during the workshop.

5. Acknowledgements

The organizing committee would like to thank for kind contributions (congress materials and other supports) from:

- AB SCIEX
- Bruker Daltonics
- Bio-Rad Laboratories, Inc.

6. Confirmed Participants

Alessio Soggiu	University of Milan, Milan, Italy
Alexander Archakov	Institute of Biomedical Chemistry RAMS, Moscow, Russia
Andrey Lisitsa	Institute of Biomedical Chemistry RAMS, Moscow, Russia
Anil K. Madugundu	Institute of Bioinformatics, Bangalore/Karnataka, India
Fuchu He	Beijing Proteome Research Center, Beijing, China
Ghasem Hosseini Salekdeh	Royan Institute, Tehran, Iran
Gilbert S. Omenn	University of Michigan, MI, USA
Haidan Sun	Beijing Institute of Genomics, Beijing, China
Je-Yoel Cho	Seoul National University, Seoul, Korea
Jin-Young Cho	Yonsei Proteome Research Center, Seoul, Korea
Jong-Shin Yoo	Korea Basic Science Institute, Daejeon, Korea
Ju-Wan Kim	Yonsei Proteome Research Center, Seoul, Korea
Juan Antonio Vizcaino	EMBL-EBI, Hinxton, Cambridge, UK
Kyung-Hoon Kwon	Korea Basic Science Institute, Cheongju, Chungbuk, Korea
Mark S. Baker	Macquarie University, NSW, Australia
Maxey Chung	National University of Singapore, Singapore
Pengyuan Yang	Fudan University, Shanghai, China
Peter Horvatovich	University of Groningen, Groningen, The Netherlands
Ping Xu	Beijing Proteome Research Center, Beijing, China
Qing-Yu He	Jinan University, Guangzhou, China
Quanhui Wang	Beijing Institute of Genomics, Beijing, China
Ravi Sirdeshmukh	Institute of Bioinformatics, Bangalore, India
Sanghyuk Lee	Ewha Womans University, Seoul, Korea
Savita Jayaram	Institute of Bioinformatics, Bangalore/Karnataka, India
Seul-Ki Jeong	Yonsei Proteome Research Center, Seoul, Korea
Siqi Liu	Beijing Institute of Genomics and BGI-Shenzhen, China
Soo-Youn Lee	Samsung Medical Center, Seoul, Korea
Tadashi Yamamoto	Niigata University, Niigata, Japan
Terence C.W. Poon	Chinese University of Hong Kong, Shatin, Hong Kong
Tong Wang	Jinan University, Ghangzhou, China
Victor Segura	CIMA (Center for Applied Medical Research), Pamplona, Spain
Visith Thongboonkerd	Siriraj Hospital, Mahidol University, Bangkok, Thailand
William S. Hancock	Northeastern University, Boston, MA, USA
Young-Ki Paik	Yonsei University, Seoul, Korea
Yu-Ju Chen	Institute of Chemistry, Academia Sinica, Taipei, Taiwan