Final Circular (last updated 7/6/2014)

CHPP

The International Consortium of the Chromosome-Centric Human Proteome Project

10th C-HPP Workshop in Bangkok, Thailand

August 9, 2014, Siriraj Hospital, Mahidol University,

2 Wanglang Road, Bangkok, Thailand

Proteome Data Management and Identification of Missing Proteins

SiMR Building

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 NexProt, 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 Meeting

●**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), Bangkok, Thailand

●**Organizing Committee:**
-Convener: William S. Hancock
-Program Coordinator: Young-Ki Paik,
-Administrator: Peter Horvatovich
-Visith Thongboonkerd (Local Host)

●**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 colleagues in more interactive.

●**Administration:** All matters related to 'a letter of invitation' will be handled by local host, Dr. Visith Thongboonkerd (Visith Thongboonkerd vthongbo@yahoo.com.

3. Scientific Program (final version, updated 7/6/2014)

*(Note: This program will be updated from time to time at www.c-hpp.org, http://thehpp.org, hupo.org, wiki site at http://c-hpp.webhosting.rug.nl/tiki-index.php***)**

08:30-08:40 **Welcome & Opening Remarks**

Moderator: Visith Thongboonkerd, Mahidol Univ., Bangkok, Thailand

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

08:40-09:30 **Session 1: Introduction and Opening Talk**

Chair: Visith Thongboonkerd, Mahidol Univ., Bangkok, Thailand

Part A: Briefings on the C-HPP Progress

Update & Future Plans on the C-HPP Publications in 2014 and 2015
Bill Hancock, Co-Chair of the C-HPP Consortium & Editor-in-Chief, JPR

Update on the Bio-banks within the HPP
Peter Horvatovich, Secretary General, C-HPP
(for Gyorgy Marko-Varga)

Part B: Opening Invited Talks (30 min)

Integrative Proteo-Genomic Analysis of Early Onset Gastric Cancer
Sanghyuk Lee
Dept. of Life Science, Ewha Womans Univ.
Seoul, Korea

09:30-11:10 **Session 2: Progress on the Proteogenomics/Transcriptomic Data Productions (Invited Talks, 15 min each)**

Chair: Bill Hancock
Co-Chair, C-HPP Consortium, Editor-in-Chief of JPR

Looking for missing proteins: an enlightenment from the analysis of free-mRNA and RNC-mRNA data
Siqi Liu
Beijing Genome Institute, 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

Proteogenomic analysis of the human chromosome 9-encoded genes
Je-Yoel Cho
Seoul Natl Univ., Seoul, Korea

Comprehensive characterization of a liver tissue and HepG2 cells transcriptoproteome for human chromosome 18
Andrey Lisitsa
Institute of Biomedical Chemistry of Rus. Acad. Med. Sci., Russia

Human Y Chromosome Proteome Project: 2014 update
Ghasem Hosseini Salakdeh
Royan Institute for Stem Cell Biology and Technology, Tehran, Iran

11:10-11:30 **Coffee Breaks**

11:30-13:00 **Session 3: Round Table Discussion**

"Problems and Solutions for the Large Dataset Validations"

Co-Chairs: Bill Hancock, Co-Chair, C-HPP EC, Northeastern Univ

Gil Omenn, Chair, HPP EC, Univ. of Michigan

Panelists (Country Representatives):

Mark Baker (Australia)
Siqi Liu (China)
Ghasem Hosseini Salakdeh (Iran)
Tadashi Yamamoto (Japan)
Jong Shin Yoo (Korea)
Peter Horvatovich (Netherlands)
Andrey Lisitsa (Russia)
Victor Segura (Spain)
Visith Thongboonkerd (Thailand)
Yu-Ju Chen (Taiwan)

A: What are the most burning issues?

Q1. Proof of validity of a large dataset available in the public DB:
Production, Repository, Exchange, Quality, 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

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

13:00-14:00 **Lunch**

14:00-15:00 **Session 4: Practical aspects of individual data management**

Co-Chairs: Peter Hovartovich, Secretary General of the C-HPP Consortium
University of Groningen, Netherlands

Andrey Lisitsa, Russian Academy of Medical Science, Russia

A. Proteomic Dataset with PXD

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?)

B. Proteogenomic Dataset:

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

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

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. Share information about sample sets analyzed

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

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

15:00-17:00 **Session 5: Strategy for Identification and Characterization of Missing Proteins (Invited Talks, 15 min each)**

Co-Chairs: Mark Baker
HUPO President-elect, Macquarie Univ., Sydney, Australia

Fuchu He, President of AOHUPO
BPRC, Beijing, China

Metrics and strategy for identifying missing proteins
Gil Omenn, Chair, HPP EC, Univ. of Michigan, Ann Arbor, USA

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
Bill Hancock, Co-Chair, C-HPP Consortium

Sydney Inaugural HPP Missing Proteins Workshop Report
Mark Baker, President-elect HUPO (Chr 7 PI)

Identification of Missing Proteins in Aggregated Proteins
Qing-Yu He
Jinan University, Guangzhou, China

To conquer the last hard-core of missing proteome
Pengyuan Yang
Fudan Univ., 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 of Rus. Acad. Med. Sci., Russia

17:00-17:20 **Coffee Breaks**

17:20-18:20 **Session 6: Amended Version of Long Term Plans and Deliverables**

Co-Chairs: Young-Ki Paik
YPRC, Yonsei Univ., Seoul, Korea

Bill Hancock
Northeastern Univ., Boston, 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 Alba (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)

18:20-18:30 **Conclusions:** Bill Hancock

18:20-20:00 **Dinner (Sponsored by Local Host)**

--The end of workshop---

Excursion: August 10, 2014, a Half Day Tour:

Grand Palace & Temple of the Emerald Buddha (Wat Phra Kaew)
All participants will join this a half day tour sponsored by the local organizers.