The Chromosome Centric Human Proteome Project (C-HPP) Annual Report 2018-2019 prepared for the 2019 HUPO Council

Submitted July 31, 2019 by:

Christopher M. Overall Chair, Young-Ki Paik Co-Chair, Lydie Lane, Co-Chair
On behalf of the C-HPP Executive Committee

1. Name of Initiative: Chromosome-Centric Human Proteome Project (C-HPP)

2. Name of Committee Chair: Chair: **Christopher M. Overall,**
Co-Chairs: **Young-Ki Paik, Lydie Lane**

3. Names of Committee Members:

C-HPP Executive Committee (EC):

Chair: Christopher M. Overall	Canada	to Dec 31, 2021
Co-Chair: Lydie Lane	Switzerland	to Dec 31, 2020
Co-Chair: Young-Ki Paik	Korea	to Dec 31, 2021
Secretary General: Peter Horatovich	The Netherlands	to Dec 31, 2019
Member-at-Large: Pengyuan Yang	China	to Dec 31, 2021
Member-at-Large: Fernando Corrales	Spain	to Dec 31, 2019
Member-at-Large: Gilberto Domont	Brazil	to Dec 31, 2021

Principal Investigators Council (PIC):

Chromosome 1:	Ping Xu	China
Chromosome 2:	Lydie Lane	Switzerland
Chromosome 3:	Takeshi Kawamura	Japan
Chromosome 4:	Yu Ju Chen	Taiwan
Chromosome 5:	Peter Horvatovich	The Netherlands
Chromosome 6:	Rob Moritz	USA/Canada
Chromosome 7:	Edouard Nice	Australia
Chromosome 8:	Pengyuan Yang	China
Chromosome 9:	Je-Yoel Cho	Korea
Chromosome 10:	Josh Labaer	USA
Chromosome 11:	Jong Shin Yoo	Korea
Chromosome 12:	Ravi Siredeshmukh	India
Chromosome 13:	Young-Ki Paik	Korea
Chromosome 14:	Charles Pineau	France
Chromosome 15:	Gilberto Domont	Brazil
Chromosome 16:	Fernando Corrales	Spain
Chromosome 17:	Gilbert S. Omenn	USA
Chromosome 18:	Alexander Archakov	Russia
Chromosome 19:	Sergio Encarnacion	Mexico
Chromosome 20:	Siqi Liu	China
Chromosome 21:	Albert Sickmann	Germany
Chromosome 22:	Akhilesh Pandey	USA
Chromosome X:	Yasushi Ishihama	Japan
Chromosome Y:	Ghasem Hoeissini Salekedeh	Iran
Mitochondrial:	Andrea Urbani	Italy

4. C-HPP Mission and Objectives

The mission of the C-HPP is to map and annotate the entire human proteome comprising the individual proteins encoded by each chromosome, their major splice forms, mature N- and C-termini, and their major protein post-translational modifications (PTMs) (*see* HUPO.org). In the C-HPP this is accomplished by directed studies initiated by the 25-international chromosome + mitochondrial DNA teams. Effective collaborations exist between the chromosome teams and other members of HUPO within the 19 B/D-HPP initiatives and the 4 HPP Pillars.

Phase 1 of the HPP project is focused on identifying by mass spectrometry all human proteins, presently estimated in the human genome to be 20,399 (neXtProt 2019-01-11). Those proteins identified by protein existence (PE) information number some 17,694 (PE1), with PE2 – 4 proteins remaining to be detected at the protein level—the so called "missing proteins" (MPs). At present, there remain 2,129 MPs (PE 2 – 4) yet to be identified. In Santiago C-HPP-2018, the neXt-CP50 Challenge was launched to functionalize proteins in the "Dark Proteome" with no known function, whether predicted or described. In 2018 PE1 – 4 proteins these numbered 1,937.

Phase 2 will focus on the neXt-CP2000 (to functionalize 2,000 uPE1s), ~5 PTMs / PE1 protein, and their splice forms. Also targeted are nonconventional-encoded small open reading frame translation products (smORFs), fusion proteoforms, and translatable products of long non-coding (lnc) RNAs.

5. Summary of Recent Accomplishments, Current Activities, and Tasks

A. neXt-MP50: The neXt-MP50 Challenge was launched at Sun Moon Lake C-HPP-2015 to encourage the Chr teams to identify 50 new MPs each from 2,949 MPs (2016) and to devise and employ innovative approaches to uncover MPs. This challenge has been extended past the original two-year window. Semi-annual reports from each chromosome team are posted on the C-HPP Wiki.

The number of MPs declined from 2,168 (neXtProt 2018-01-17) to 2,129 now (neXtProt 2019-01-11) with a number of Chromosome teams having completed the neXt-MP50: Chromosome 1, 2, 5, 17, 19, and X. However, the number of protein entries also increased from 20,230 (neXtProt 2018-01-17) to 20,399 (neXtProt 2019-01-11). Thus, the decreasing numbers of MPs found each year reflects both the increasing difficulty in devising and executing deep discovery of MPs in the human proteome as well as some realignment of protein encoding gene numbers and PE identifications occurring from time to time by database curators. At Orlando HUPO-2018 the chromosome teams reported around 401 MPs that were identified in 2018 according to the C-HPP guidelines v2.1.

B. neXt-CP50: With the official launching of the neXt-CP50 challenge, the goal is to characterize 50 uPE1 proteins within 3 years by 15 Chromosome teams, to date. To start this challenge with realistically attainable goals, only those uncharacterized (u) proteins that have already been positively identified at the protein level (PE1) are being analyzed. In March 2018 there were 1,937 proteins with no known function, of which 1,260 were PE1 proteins, now termed uPE1 proteins (10.1021/acs.jproteome.8b00383). In 2019 there are 1,254 uPE1, out of 2,222 uPE1 – 4 entries in neXtProt according to sparkle query NXQ_00022. These proteins have mass spectrometry identified and archetypic peptides reported from some tissues and cells, providing uPE1 project start points.

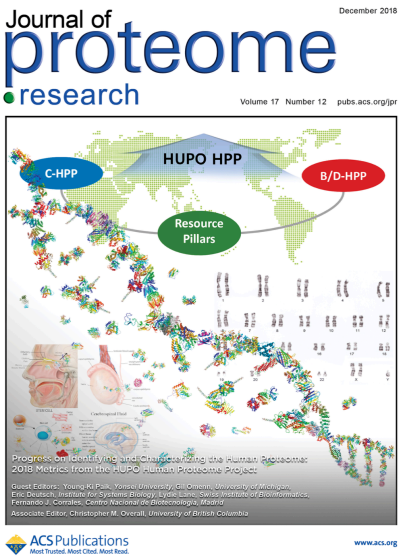
C. 21st C-HPP Symposium, St Malo, France: Between May 12 – 14, 2019 the C-HPP held its Semi-Annual Workshop, the 21st in a series of highly successful updates on the chromosome teams progress for completion of the HPP, with a focus on the newly initiated 'Dark Proteins' neXt-CP50 challenge. Charles Pineau was the local organizer, *par excellence*—French cuisine, wine and *la belle vie* flowed through the meeting attended by 43 registrants. The Journal of Proteome Research (JPR) partnered with the C-HPP and heavily publicized the workshop by JPR splash page "sliders" with updates on the C-HPP Wiki and the HUPO web sites.

On May 11, a one-day session on Mass Spectrometry Data Interpretation Guidelines 3.0, was led by Eric Deutsch (ISB, Seattle, HPP Bioinformatics Working Group) to update the HPP guidelines v2.1 for identification of MPs. Vigorous debate and discussion successfully resolved 25 open questions (*see* HPP Chair's report). The 2019 Special Issue (SI) of JPR will present a paper on the new v3.0 Guidelines.

The scientific program included 5 invited and 23 talks selected from the abstracts on various aspects of the human proteome and the dark proteome, including structural biology,

a summary of the changes to the HPP Data Interpretation Guidelines by Eric Deutsch, updates on neXtProt by Lydie Lane, neXt-MP50 by Chris Overall, and the neXt-CP50 by Young Ki Paik. Anne-Claude Gingras (Toronto, Canada) presented proximity-dependent biotinylation (BioID) for revealing the localization of human cellular proteins that could be forward/reverse exploited for MP identification. Nuno Bandeira presented ProteinExplorer for integrating community-scale big data for assessing protein existence. Yves Vandenbrouck (Chr 14) explored the dark side of the human proteome using ProteoRE, and Fernando Corrales (Chair B/D-HPP, C-HPP EC, Head Chr 16) discussed activities and ideas for collaboration between the B/D-HPP and chromosome teams.

D. Publication of the Special Issue of the HPP in the Journal of Proteome Research: On December 7, 2018, the sixth annual special issue (SI) of the HPP was published in the Journal of Proteome Research, Volume 17, Issue 12, Pages 4,023 – 4,358: Associate Editor: Christopher M. Overall, Guest Editors: Young-Ki Paik, Eric Deutsch, Fernando Corrales, Lydie Lane, and Gil Omenn. Formerly this SI was dedicated to the C-HPP, but in 2017 we expanded its scope to all the HPP. In this issue, 32 papers covered 4 major research topics: (i) missing proteins (MPs), (ii) uPE1



proteins, (iii) bioinformatics tool development and (iv) biology/disease proteomes. The launch of the neXt-CP50 was discussed in Young-Ki Paik *et al* (Chr 13). Deutsch *et al* presented the use of spectral library searches in proteomics workflows. The work of Macron *et al* (Chr 2) described identification of 12 MP candidates in human cerebrospinal fluid following immunodepletion and TMT labeling, from which 8 MP were found. Sun *et al* (Chr 1) presented a study on human testis, using multiple proteases and high and low pH deep proteomics analysis and identified 14 MPs. The study by He *et al* (Chr 1) used LysargiNase to identify and validate 2 MPs from 7 MP candidates. Pullman *et al* presented ProteinExplorer, to explore the large amount of reanalyzed public proteomics data available in MASSIVE, which allowed validation of HPP-compliant evidence for 107 MPs (PE2, PE3, and PE4) and 23 dubious (PE5) proteins. Fernando Corrales (Chr 16) used the new letter format for the identification of the long-sought hyaluronan synthetase 1, surprisingly a MP till then. The C-HPP HQ office in Korea has freely distributed one copy of the printed version to all C-HPP PIs and HPP leaders. The **2019 HPP Special Issue** to date has 23 submissions in progress or accepted, with a number of papers rejected. Over the past 3 years, SI submissions have been steadily increasing and track consistently well 2 years after publication where they maintain higher, or not significantly different, citation rates versus to the standard JPR Issues and maintain a consistent average download rate.

E. C-HPP Newsletter No. 8 **(August 1, 2019)** is posted of the C-HPP wiki **https://c-hpp.web.rug.nl/**.

F. C-HPP 2.0 Organization. At the 19th C-HPP Workshop, Santiago de Compostela, Spain (June 16 – 17, 2018) and the HPP workshop @ HUPO-2018 Orlando a new organizational plan, C-HPP 2.0, was presented to the PIC and C-HPP membership designed to streamline the Chromosome

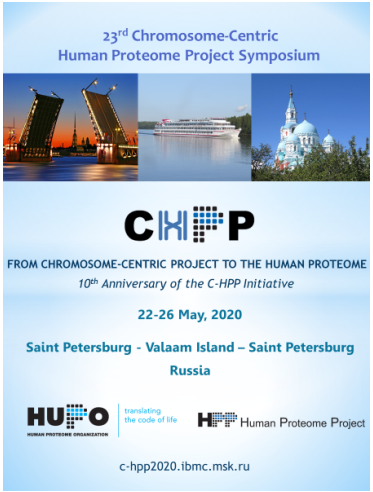
teams and to mold the C-HPP according to interest in: Protein families; Rare Tissues and Cells; Chromosome Biology; Geography/National Groups; Proteoforms; uPE1 Functionalization Developmental Biology; Technology and New Strategies; Interactomics. However, this was not at all supported by the PIC and members who voted unanimously to maintain the current chromosome-based structure. The National Chromosome team structure was viewed as a strong positive yet still provides opportunities for interdisciplinary projects. It was recognized that a core strength and success of the C-HPP has always been, and should continue to be, the annotation of the human proteome led by Lydie Lane (Head, neXtProt, Chr 2) and Eric Deutsch (Head, Peptide Atlas, Chr 6). The HPP cannot rely upon *ad hoc* community data uploads to complete the HPP by most 'outside' groups that do not recognize the stringency required for MP identification. Nonetheless, several elements of the 2.0 plan have been globally implemented over the C-HPP and in the coming years. Moreover, the C-HPP aims to collaborate with B/D-HPP teams to utilise diseased tissue samples to seek MP "responder proteins" at selected stages of injury/disease/infection/stress and resolution in all the different human tissues likely to harbor MPs that may be key for regeneration or repair of these specific tissues. Similarly, the Pathology Pillar has great potential to confirm distribution of MPs and uPE1 proteins in health and disease at the cell and tissue level, particularly by tissue microarrays (TMAs) for targeted identification searches and clues for functionalization.

6. Future Activities

A. HUPO-2019 Adelaide: C-HPP Poster Session will be held on Monday, September 16, 2019 during the HUPO Congress. The discussion will be led by Gilbert Omenn at each poster where authors will start with a lightning presentation. We thank ProtiFi, LLC (Dr. John P Wilson) for their generous support of USD600 for the Annual C-HPP Poster Awards. Dr. Sean O'Donoghue was invited by Chris Overall to present at the Post Congress HPP Day on the Structural Dark Proteome. Updates on the C-HPP activities will be presented on both of days of the HPP Workshop in Adelaide.

B. neXt-MP50 2019-2020: With the much of the "low hanging" proteome fruits having now been harvested, targeted proteome analyses of specific tissues and cells are now desperately needed, particularly of rare or rarely analyzed human cells and tissues including developing tissues in the human embryo. This is the focus of the "Rare Cells and Tissues Proteomes" concept, now a recognized strategy in collaboration with Neil Keller by Top Down analyses. To identify these temporally or spatially rare proteins will take dedicated searches of specific cells and tissues in adult tissues *eg.* olfactory epithelium below the cribriform plate, hard connective tissues *eg.* membranous and cartilaginous bone, dental cementum, dentine and enamel, or during embryo and fetal growth at precise developmental windows—an ethical and practical challenge for many countries. Cerebrospinal fluid, and both male and female reproductive tissues have proven to be rich sources of MPs that require extensive reanalysis using a variety of new tissue sample and proteomic techniques. Specific strategies to identify MPs can be beta-tested using, for example, recombinant MPs from full-length plasmids for ~62% of the missing proteins available from the Chr 10 team, led by Josh Labaer at Arizona State University.

C. neXt-CP50 2019-2020: Progress will be accelerated based on availability of resources useful for their characterization including antibodies through the Human Protein Atlas and expression clones for ~70% uPE1s available from the Chr 10 team led by Josh Labaer at Arizona State University.



D. 23rd C-HPP Symposium St. Petersburg *to* Valaam Island, Russia: In 2020 we look forward to Russia hosting the 23rd C-HPP Annual Workshop "*From Chromosome-Centric Project to the Human Proteome*", on a river-class cruise ship, traveling along the rivers and lakes between St Petersburg and Valaam Island. This Symposium format provides an opportunity for many informal and fruitful discussions between participants, combining the high level of the scientific program at the same time visiting Russian cultural and historic places between sessions.