The Chromosome Centric Human Proteome Project (C-HPP) Annual Report 2019-2020 prepared for the 2020 HUPO Council

Presented on September 3, 2020 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, 2022
Member-at-Large: Pengyuan Yang	China	to Dec 31, 2021
Member-at-Large: Fernando Corrales	Spain	to Dec 31, 2022
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:	OPEN	
Chromosome 22:	OPEN	
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 19,773 (neXtProt 2020-01). Those proteins confidently identified by protein existence (PE) information number 17,874, an increase of 180 from 17,694 PE1 proteins in 2019, with 1,899 PE2 – 4 proteins "missing proteins" remaining to be detected at the protein level. At the Santiago C-HPP-2018 workshop, the neXt-CP50 Challenge was launched to functionalize proteins in the "Dark Proteome" with no known function, whether predicted or described. 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 2018 PE1 – 4 proteins these numbered 1,937 and in 2020, the numbers were now 1,254.

Phase 2 will focus on identifying the remaining ~10% of the human protein, and to reduce the number of PE1 proteins that so far lack MS evidence (nonMS-PE1 proteins) for their existence. In addition, the neXt-CP2000 (to functionalize 2,000 uPE1s), ~5 PTMs / PE1 protein, and their splice forms.

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

The C-HPP is currently seeking new teams and leaders to adopt Chromosomes 21 and 22 to renew these teams. In total the C-HPP and B/D-HPP Chromosome teams, sometimes in collaboration, published 20 papers in the 2019 Journal of Proteome Research HPP Special Issue, with 15 submitted to the 2020 Journal of Proteome Research HPP Special Issue. A total of 61 other papers addressing different aspects of the human proteome were published in other various journals.

A. neXt-MP50: 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. To address this, 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 and the most recent report from August 20, 2020 is available here.

Summary of neXt-MP50 progress since HUPO-2019 Adelaide:

A. How many MPs (PE2–4) identified as PE1 since 2017	**481**
B. How many MPs (PE2–4) found in 2019 now listed in neXtprot as PE1	**228**
C. How many candidate MPs found in 2019, but not meeting Guidelines?	**Hundreds**
D. Significant findings and conclusions to highlight: • An issue arising from analysis of the team reports is that a large number of missing proteins that were reported "found" and discussed in their papers were either not captured by Peptide Atlas for analysis or failed reanalysis by Peptide Atlas and so were not promoted PE1 status • Considerable evidence was found for MPs, but failed to satisfy the HPP Guidelines 3.0 and so remain as candidate 'found' MPs. • Recommendation: Information regarding these candidate MPs should not be lost but compiled (to be determined where, how and format) so as to be accessible to guide and inform ongoing and future proteomics studies by the community, directed data analysis of similar tissue/cells in Proteome Exchange and current literature to generate evidence sufficient to meet the HPP Guidelines. • Several chromosome teams (e.g. Chr 5, 12) are active in the Cancer Moonshot and CPTAC projects and successfully analysing this data for MPs. • Chromosome 6 inititated a directed search for PE1 proteins lacking MS evidence (termed non-MS PE1 proteins), with several identified by MS (in human bone) that met the HPP Guidelines for PE1 identification by MS. • A precision medicine molecular corrector drug was developed by Chr6 team members that was shown to restore functional levels of a mutant protein isoform of MALT1. Untreated, this mutant protein led to a rare immunodeficiency disease. The disease was phenotyped in a previous paper by proteomics and TAILS that led to this discovery and then drug candidate. • The Chromosome 12 (South and SE Asia) team has recruited Radislaw Sobota, Singapore as a new member of the team.	

• Chr 17 has met the MP50 Challenge: the number of PE2,3,4 missing proteins coded on Chr 17 has been reduced from 148 to 87, meaning that 61 MPs have been detected and incorporated into neXtProt PE1. • Chr X (Japan) also have enjoyed great success in identifying MPs, with 35 now PE1 proteins in neXt-Prot. • MT (Italy) some years ago were the first to complete the human proteome encoded by a chromosome, in this case, the smallest in humans with 15 proteins, all now PE1 proteins. Progress is now on the non-MT encoded proteins in the mitochondria. • COVID-19 significantly impaired C-HPP progress on the HPP	

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. Numbers of uPE1 Dark Proteins under investigation: Chr2, 1 (jointly published); Chr 10, several; Chr11, 3; Chr13, 2; Chr 16, Tool Development mainly; Chr 18, 4; Chr 19, 4; Chr X, 4. During the past year there was no net decrease in the current 1,254 uPE1 proteins. The HPP needs a much-concerted effort to significantly address this Challenge. Chr 17 made a major contribution to the neXt-CP50 Challenge by creating the I-TASSER/COFACTOR function prediction pipeline which neXtProt has adopted as a community service. This pipeline predicted Gene Ontology terms for all 66 Chr 17 uPE1 protein (Zhang C, et al, JPR 2018). Its predictions were put to a blinded test with to-be-released results from neXtProt and from CAFA3 (Zhang C, et al, JPR 2019). neXtProt added a link to facilitate submission of uPE proteins for a report of predicted functions from the group at the University of Michigan. As of 15 May 2020, documentation of requests for C-I-TASSER function predictions showed a total of 561 proteins from 181 users from 35 countries, including 201 neXtProt proteins [https://zhanglab.ccmb.med.umich.edu/C-I-TASSER/bin/stat.cgi].

C. C-HPP Workshop: The 22nd C-HPP Symposium, St Petersburg, Russia was cancelled due to COVID-19. The next C-HPP workshop will be held in Busan, South Korea, June 28 –30, 2021, immediately prior to AO-HUPO meeting.

D. Publication of the Special Issue of the HPP in the Journal of Proteome Research: In December, 6, 2019, the seventh annual special issue (SI) of the HPP was published in the Journal of Proteome Research, Volume 18, Issue 12, Pages 4,079 – 4,282: Associate Editor: Christopher M. Overall, Guest Editors: Young-Ki Paik, Eric Deutsch, Fernando Corrales, Lydie Lane, and Gil Omenn. In this issue, 21 papers covered 4 major research topics: (i) missing proteins (MPs), (ii) uPE1 proteins, (iii) bioinformatics tool development and (iv) biology/disease proteomes. Except for this year due to COVID-19, 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. A virtual issue of HPP papers from the past 5 years will be published around the time of HUPO-Connect to celebrate the identification of 90% of the human proteome.

E. C-HPP Newsletter No. 8 is posted of the C-HPP wiki https://c-hpp.web.rug.nl/.

F. HUPO-2019 Adelaide: C-HPP Poster Session was held on Monday, September 16, 2019 during the HUPO Congress. The discussion was be led by Dr. Gilbert Omenn at each poster where authors started 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 in 2020 will be presented at the HPP Workshop, HUPO CONNECT.

G. Expresion Plasmid Library for MPs and uPE1s: The Chromosome 10 team, led by Josh Labaer at Arizona State University, has assembled a comprehensive and one of the world's largest collections of full-length Gateway plasmids representing 90% of all human protein-coding genes and are distributing the collection through their repository and distribution web portal DNASU (dnasu.org). Currently, Chr10 has full-length plasmids for 175 of 804 missing proteins, which are available to the entire HPP. Chr10 is also producing more full-length plasmid clones for uPE1 proteins for functional studies. Currently, they have full-length plasmids for around 80% of 1,646 uPE1 proteins and aim to reach >90% by the end of 2020, which is available to the entire C-HPP team. These full-length plasmids for the majority of dark proteins in multiple vectors can be applied to many types of experiments for functional characterization of the Dark Proteins. IVTT-produced proteins (GST-tagged) can be used for targeted MS or antibody validation, and the Lenti-based plasmid can be used for cell-based assays or screening. All these are available to the C-HPP team via our web portal DNASU.org, and Chr10 is always open to collaboration.