HIGHLIGHTS FROM HPP WORKSHOP 21 SEPTEMBER 2017 in DUBLIN

About 70 HPP investigators and interested HUPO Congress attendees participated in the post-Congress Workshop at University College Dublin on 21 September 2017.

Highlights from the Congress and HPP meetings and sessions

Mike Snyder initiated and moderated the discussion. He highlighted the importance of replicable, quantitative analyses throughout biology, the emergence of single-cell analyses with CyTof and MS, contrasts between proteomics and metabolomics in terms of depth of analysis (favorable for proteomics) and time and cost of analyses, and growing emphasis on clinical applications and wellness projects. Jenny van Eyk pointed to dynamics of proteomics studies and new work with cross-linking and protein structures. Eric Deutsch emphasized statistical validation and localization of proteins with PTMs (mzIdentML now covers PTMs). Nikki Parker mentioned glycoproteins in the secretome.

Young-Ki Paik introduced a C-HPP initiative complementary to the Next50 Missing Proteins Challenge (neXtProt PE2,3,4 predicted proteins), the uPE1 proteins lacking any functional annotation (estimated at 1232 in neXtProt version 2017-08-08). He introduced the 5 new leaders of C-HPP teams, the schedule of C-HPP workshops beginning with Santiago, Spain, in June 2018, and referred us to the C-HPP website and wiki. Lydie Lane emphasized new advanced search tools in neXtProt. There was general discussion about linking C-HPP and B/D-HPP teams; Jenny van Eyk challenged us to clarify the process for declaring a new project within the HPP, not just in the C-HPP. Likewise, the Early Career Researchers initiative of the B/D-HPP might be better placed under the whole HPP. There are groups in both arms pursuing PTMs, splice variants, and functional characterization. Several people called for HPP to strengthen governance.

Progress on the Missing Proteins Next 50 Challenge

Chris Overall reported that there were 216 previously PE2,3,4 predicted proteins confidently identified in 2016. Comparing neXtProt PE2,3,4 MPs from 2016-01 to 2017-01, we find the total was reduced from 2949 to 2579,

a gain of 370 altogether, and the total of PE1 proteins rose from 16,518 to 17,008. Most of these came from the Swiss-French Chr 2+14 study of sperm plus the Chinese consortium study of testis. The assessment of newly identified proteins from the manuscripts submitted to the JPR 2017 special issue was still under review, with many candidates, but a relatively modest number fully confirmed; now that the manuscripts are all online, the best estimate is 73. See the Paik et al Editorial for the December special issue, which provides a good introduction to the multiple strategies for detecting MPs and their yield to date. It is hoped that more C-HPP teams will become active in this Next50 Challenge. There were many comments.

Progress on the B/D-HPP Popular Proteins Strategy

The proteomics strategy of building a bridge to organ-focused biomedical researchers with lists of most studied proteins and suitable SRM assays for quantitative has gained some momentum. Fernando Corrales, from the B/D-Liver Proteome Project and Chr 16, used this approach to focus on the enzymes of one-carbon metabolism pathways in progression of non-alcoholic fatty liver disease/non-alcoholic steatohepatitis. Jenny van Eyk from the B/D Cardiovascular Proteome team focused on cysteine S-nitrosylation sites involved in some switch functions in human, mouse, and rat. Their database has 2955 unique SNO-proteins with 5826 sites, including myofilaments, myosin H chain, and other proteins of interest. Tadashi Yamamoto described proteomics findings in multiple compartments of the kidney and in urine studies.

A major advance is the work of Kun-Hsing Yu, now at Harvard, after a post-doc with Mike Snyder at Stanford. He performed a complementary literature search for each of the 22 B/D-HPP groups, including 11 organ systems, using PubMedQuery and PubtatorQuery. Mike summarized the approach and noted that the work will be prepared for publication in the coming year, hopefully with information about uses by B/D teams.

Comments from HPP Senior Scientific Review Board

In the Sunday HPP investigators meeting, SSRB members Lee Hood, Mathias Uhlen, Cathy Costello, and Mike Snyder provided many useful comments. At the Thursday workshop, Naoyuki Taniguchi and Mike Snyder carried forward these discussions. Taniguchi recommended a roadmap to enhance communication, attention to additional organs (like

lung, ovary, and prostate), clinical applications, and the renewed glycoproteomics activities. Snyder called the role of the HPP EC critical, supported the thrust on PTMs, noted the deliverables from the B/D, acknowledged the interest in lncRNAs and smORFs addressed critically in the annual Metrics paper, and called for more attention to deliverables from each of the B/D and C-HPP teams. He agreed that the ECR initiative should be adopted by the whole of HUPO, perhaps linked with the Education Committee chaired by Gary Corthals.

There ensued a discussion about the proposed new Pathology Proteomics Initiative, whether it should be under B/D, or as a new Resource Pillar. Ed Nice, Dan Chan, Mark Baker, and others are actively engaged in the planning, including an upcoming conference in Australia. The tentative advice from the Workshop was to place Pathology under the B/D and connect it with multiple organ-specific and biofluid teams. We will welcome a well-drafted assessment of alternatives following the Australia conference.

Relationship with the Journal of Proteome Research

Chris Overall announced that the Editor of JPR, John Yates, and the editorial board have committed to a 6th annual issue on the same schedule as worked pretty well this year: Call for Papers in December 2017, using PeptideAtlas 2018-01 and neXtProt 2018-01 as the baselines for analyses and discussion; deadline for submission of manuscripts 30 April 2018; completion of review/revision/re-review process by 1 September 2018; submission of finalized manuscript to the production process by October 1 2018; and publication in December 2018. Each accepted manuscript will be published online when accepted.

Mass Spectrometry Resource Pillar Initiative on Post-Transational Modifications

Sue Weintraub, leader of the rejuvenated MS Resource Pillar, introduced this initiative on Sunday and again at the Workshop. She announced the availability of a valuable resource prepared specifically for HUPO Congress attendees and HPP investigators, a standard sample with 96 phospho-peptides either in solution or in a yeast protein matrix by SynPeptides Ltd of Shanghai and aliquoted by the Moritz Lab at the Institute for Systems Biology. If you did not sign up at the Exhibit Hall and take home the vials

with these standard samples, you can still obtain them and perform your analyses and share your results.

Yingming Zhao described extensive work on a family of metabolism-sensitive PTMs: acetyl, propionyl, butyl, crotonyl, hydroxyl-butyryl, malonyl, succinyl...lysyl acylations; histone acylations and HDACs sirtuins, and acyl transferases; and the different specificity in different pathways of P300/CBP. Justyna Fert-Bober reported studies of citrillunation of arginine residues. Removing positive charges can be impactful and create neoantigens. Increased hydrophobicity increases retention time. DIA SWATH workflows have been developed. Databases in heart and brain have been launched. Hongxiu Yu of Fudan University presented on succinylation, a dynamic PTM important in HDAC classes I, II, III, IV, as well as deacetylase SIRT5. NADP+ IDH mutations cause excess succinylation, while SIRTS5 de-succinylates and activates PK2. Such changes affect immunoregulation in tumors and tumor cells.

Rob Moritz summarized advances in tools and approaches, including inclusion lists and SWATH and SRM. There are many unidentified peptides; TPP now offers respect for chimeric spectra and Alexey Nesvizhskii and Andy Kong have developed MSFragger. D-amino acids, especially of Ala and Ser, may be significant in aging brain and eye. Of course, searching for modified peptides increases the computational burden. Pengyuan Yang emphasized the expression patterns, splice variants, and PTMs related to regulation of protein-coding genes, epigenetics, metabolites, and cell biology.

Long non-coding RNAs

Tong Wang from Guangzhou described the translatome based on full-length RNC-mRNA from initiation fractions on polysomes, as described in 2013. Some 47% of missing proteins have positive evidence in the translatome in liver. Epigenetic induction may enhance expression. Testis may be a particularly rich organ for study. From his presentations at the Tehran Workshop and the Guangzhou Conference this year, we invited him to present to HPP investigators for better recognition of findings to date and critical evaluation of claims of protein translation products.

From studies of 9 cell lines, 2969 lncRNAs were identified. Applying the HPP Guidelines v2.1 of two non-nested proteotypic peptides of at least 9aa

in length, and FDR <1% at protein level, they identified substantial numbers of protein candidates using search engines and have performed MRM on many peptides and immunofluorescence on a small subset. The preliminary impression at the Workshop was that much more detailed assessment of the spectra and the comparisons with spectra of synthetic peptides will be necessary to evaluate these extraordinary claims. Rob Moritz and Eric Deutsch agreed to scrutinize these findings.

Gong Zhang described an unpublished de novo sequencing approach,with 13,000 human proteins and 20,000 proteoforms identified in a single experiment on three hepatocellular cancer cell lines. No FDR method and no reference database was used. They performed MRM on a Sciex 6500 Q-TRAP instrument and claimed 424 PE2,3,4 MPs detected with 2 peptides of 9aa. A Chinese biotech offers RNA-Seq translatome analyses at no charge for C-HPP teams.

Bioinformatics Hub and HPP Guidelines for Interpretation of Mass Spec Data v2.1

Eric Deutsch reported that the Guidelines are being widely used and were kept stable for a second year. There are several new developments:

1. A Universal Spectrum Identifier, for claims of findings of missing proteins PDF screenshots in manuscripts are simply not satisfactory. For example, mzspec:PxD002145//HeLa45_20_160423/scan14321. The identifier will be required for the 2018 JPR C-HPP Special Issue. Will be finalized by April 2018.
2. On the HPP Guidelines Checklist for manuscripts submitted to JPR, we will now require page number and line number for each checkmark. Staff will check whether “not applicable” checks are explained; only some authors used that box. Item #10, if spectral quality is not great, the experiment should be repeated, not ignored.
3. We are preparing to clarify the guidelines for DIA-SWATH datasets. If these findings are analyzed like SRM with trace extraction via spectral libaries, SRM guidelines apply, with display of traces and heavy-labeled reference peptides to demonstrate co-elution and match of intensity patterns. If like shotgun (DIA-Umpire), apply shotgun guidelines, display spectra, show labeled peptides spectra. Need to talk with companies as methods of acquisition are quite different.

4. Continue to allow “MP candidate detections” for non-compliant evidence. Comment: call them “unverified detections”.
5. Take notice of exceptions to the two non-nested unique 9 aa peptides, as revealed in the 2017 manuscripts; e.g., beta-defensin 123 (Q8N688) has only one 9 aa uniquely-mapping tryptic peptide; it was a PE1 protein when the criterion was 7aa. Should we establish an Exceptions Committee? Need to work together with neXtProt curators.
6. Should we have guidelines for PTMs and their unmodified peptides? Will work with the MS Pillar team.

In closing, the many attendees still present, expressed enthusiasm for keeping the same format of pre-Congress (Sunday) HPP Investigators meeting, scientific track, Bioinformatics Hub, and post-Congress HPP Workshop (Thursday) to capitalize on learnings during the Congress and build the aims and deliverables for the future.