



Scientific Program and Timetable *(Penultimate Program)*

This program will enjoy some minor modifications before the symposium

Train arrivals at Saint-Malo from Paris city center (Friday and Saturday) and Paris CDG Airport (Friday only):

10:58 14:14 16:54

Saturday 11 May:

16:00 – 18:00	Registration opens for early arrivals
17:00 – 19:00	PIC Meeting (For everyone else, enjoy the sea view with aperos)
20:00 –	Dinner at “Le 5” (F & C Hotel Penthouse – private event for C-HPP)

Sunday 12 May:

Posters from the Chromosome teams will be on display for discussion during the whole symposium. During coffee, ~10 min focus poster talks will be given by presenters with discussion led by the moderator.

08:00	Welcome by CP and CO, and INSERM National Representative	
Session 1	Update Session	Mod. Mark Baker (HPP Chair)
08:15	<ul style="list-style-type: none"> ● Peptide Atlas Update and MP Guidelines v3.0 (Eric Deutsch) ● NeXtProt Update (Lydie Lane) ● neXt-MP50 Update (Chris Overall) ● neXt-CP50 Update (Young-Ki Paik) 	
09:15	General Discussion. Mods. Eric Deutsch & Gill Omenn	
10:00	Coffee Break & Ch 1 – 6 Poster Presentations (60 mins) (Mod. Gill Omenn)	
11:00	ProteinExplorer: Integration of community-scale big data for assessment of protein existence	Nuno Bandeira UCSD, San Diego, USA
11:45	Rob Moritz Ch6	Abst 26
12:05	Enrique Santamaria Ch16	Abst 15
12:25	Hongjiu Zhang Ch17	Abst 18
12:45	Keiko Yamamoto Ch3+X	Abst 4
13:05	Lunch & Socialize	

Session 2	CP50 Protein Functions and Approaches	Mod. Young-Ki Paik
14:15	The Dark Proteome: Shedding light on localization and interaction of uncharacterized proteins	Anne-Claude Gingras <i>Lunenfeld-Tanenbaum Research Institute, Toronto, Canada</i>
15:00	Lydie Lane Ch2	Abst 0
15:20	Young Ki Paik Ch13	Abst 8
15:40	Heeyoung Hwang Ch11	Abst 7
16:00	Luis Felipe Clemente Velarde Ch16	Abst 13
16:20	General Discussion on CP50 Strategies. Mod. Young-Ki Paik	
16:45	Coffee Break & Ch 7 – 12 Poster Presentations (50 mins) (Mod. Gill Omenn)	
17:35	Alexander Archakov Ch18	Abst 19
17:55	Ravi Sirdeshmukh Ch12	Abst 25
18:15	The virtuous cycle of human genetics and mouse (and rat) models	Marie-Christine Birling – <i>IMPC, Strasbourg, France</i>
19:00	<i>Informal exchanges between participants and apero</i>	
20:00 –	<i>Dinner at La Trinquette (Saint Malo Yacht Club)</i>	

Monday 13 May:

Session 3	Protein Function: <i>in silico</i> Approaches	Mod. Lydie Lane
09:00	Exploring the dark side of the human proteome using the ProteoRE platform (Abst 11)	Yves Vandembrouck CEA Grenoble, France
09:45	Cecilia Lindskog Ch14	Abst 10
10:05	General Discussion on <i>in silico</i> Approaches	
10:20	Coffee Break & Ch 13 – 17 Poster Presentations (35 mins) (Mod. Chris Overall)	
10:55	Cristina Ruiz-Romero Ch16	Abst 14
11:15	Jin Young Cho Ch13	Abst 9
11:35	Chengxin Zhang Ch17	Abst 17
11:55	Ekaterina Ilgisonis Ch18	Abst 20
Session 4	Protein Function: Structural Biology & Crystallography	Mod. Charles Pineau
12:15	How structural biology contributes to uncover biochemical functions of putative proteins	Mijram Czjzek Roscoff Marine Station, Roscoff, France
13:00	Lunch & Socialize	
14:00 – 17:30	Free afternoon	
18:45	<i>Meet at Seaport – Boat departs to Petit Bé (high tide). Be very very punctual!</i>	
19:00 –	<i>Dinner on Petit Bé Return to mainland by walking over the strand (low tide, so no strandings)</i>	

Tuesday 14 May:

Session 5	Strategy Session	Mod. Eric Deutsch
09:15	Bridging C and B/D HPP to define the biological context of human proteins PTM (Abst 12)	Fernando Corrales Centro Nacional de Biotecnología , Madrid, Spain
10:00	Olfactory Receptors and Membrane Proteins	Mark Baker
10:20	General Discussion on Membrane Protein Challenges (Mod. Eric Deutsch & Mark Baker)	
10:40	Coffee Break & "Telomere" Ch Poster Presentations (50 mins) (Mod. Gill Omenn)	
11:30	Camille Mary Ch2	Abst 2
11:50	Ekaterina Poverennaya Ch18	Abst 21
12:10	Peter Horvatovich Ch5	Abst 23
12:30	General Discussion and Meeting Wrap Up	Chris Overall
13:00	Lunch & Departure	

Train departures to Paris city center (Direct): 13:35 16:39

Train departure to Paris CDG airport (via Rennes): 13:35 -> arrival 18:06 at Terminal 2



translating
the code of life

ABSTRACT # 0

Exploring the Uncharacterized Human Proteome Using neXtProt

Paula Duek Roggli¹, Alain Gateau¹, Amos Bairoch¹, Lydie Lane^{1,2}

¹ CALIPHO group, SIB-Swiss Institute of Bioinformatics. CMU, Michel Servet 1, 1211 Geneva 4, Switzerland

² Department of microbiology and molecular medicine, Faculty of medicine, Geneva University. CMU, Michel Servet 1, 1211 Geneva 4, Switzerland

✉: lydie.lane@sib.swiss

20,230 protein-coding genes have been predicted from the analysis of the human genome (neXtProt release 2018-01-17), and about 10% of them are still lacking functional annotation, either predicted by bioinformatics tools or captured from experimental reports. A systematic exploration of the available literature on uncharacterized human genes/proteins led to proposal of functional annotations for 113 proteins and to consolidation of a list of 1,862 uncharacterized human proteins. The advanced search functionality of neXtProt was used extensively in order to examine the landscape of the uncharacterized human proteome in terms of subcellular locations, protein-protein interactions, tissue expression, association with diseases, and 3D structure. Finally, a deep data mining in various publicly available resources allowed building functional hypotheses for 26 uncharacterized human proteins validated at protein level (uPE1). These hypotheses cover the fields of cilia biology, male reproduction, metabolism, nervous system, immunity, inflammation, RNA metabolism, and chromatin biology. They will require experimental validation before they can be considered for annotation. Despite technological progresses, the pace of human protein characterization studies is still slow. It could be accelerated by a better integration of existing knowledge resources and by initiating large collaborative projects involving specialists of different biology fields. We hope that our analysis will contribute to set up the ground for such collaborative approaches and will be exploited by the HUPO Human Proteome Project teams committed to characterize uPE1 proteins.

Keywords: SPARQL; biocuration; cilium biology; data mining; functional annotation; human protein; knowledge base; neXtProt; systems biology

Duek et al., J Proteome Res 2018; 17(12):4211-4226. PMID: 30191714 doi:10.1021/acs.jproteome.8b00537.



21st Chromosome-Centric Human Proteome Project Symposium
"Illuminating the dark proteome" - May 12-14, 2019, Saint-Malo, France

ABSTRACT # 1

Using neXtProt and other bioinformatics resources to identify human uncharacterized proteins potentially involved in male reproduction

Paula Duek¹, Gustavo L. Verón², Mónica H. Vazquez-Levin² and Lydie Lane^{1,3}

¹ CALIPHO group, SIB Swiss Institute of Bioinformatics, CMU, Michel Servet 1, 1211 Geneva 4, Switzerland

² IBYME Instituto de Biología y Medicina Experimental (CONICET-FIBYME), Vuelta de Obligado 2490, C1428ADN-Buenos Aires, Argentina

³ Department of microbiology and molecular medicine, Faculty of Medicine, University of Geneva, CMU, Michel Servet 1, 1211 Geneva 4, Switzerland

✉: lydie.lane@sib.swiss

Infertility affects up to 15% of couples in reproductive age. Half of the cases are due to defects in the characteristics of the spermatozoa that affect their fertilizing potential. The molecular entities involved in sperm fertilizing potential are still poorly characterized or unknown. Some of them might reside in the substantial fraction (>10%) of the 20,230 protein-coding genes predicted from the analysis of the human genome that still has to be experimentally validated and characterized. Indeed, recent shotgun and targeted mass spectrometry analyses allowed to validate the existence of more than 300 previously undetected proteins in sperm. Deep data mining across various resources allowed to identify dozens of uncharacterized proteins potentially involved in sperm motility or gamete fusion. We selected eight of those proteins that are either secreted or transmembrane and amenable to functional analysis using antibodies targeting their extracellular domain. The presence of these proteins in sperm was confirmed by Western immunoblotting. Their subcellular localization was determined using fluorescence immunocytochemistry under conditions that mimic fertilization-related events. Sperm preincubation with antibodies against these proteins significantly altered sperm movement parameters analyzed using a computerized system. Incubation with antibodies against five of the selected proteins significantly decreased the number of sperm penetrating the cumulus and/or inhibited sperm-oocyte fusion in heterologous models. Further studies using animal models will be required to characterize their molecular mechanism of action.



ABSTRACT # 2

Characterization of two newly identified nuclear encoded mitochondrial proteins

Camille Mary¹, Paula Duek Roggli ², Insaf Fkih M'Hamed¹, Amos Bairoch¹, Lydie Lane^{1,2}

¹ Department of microbiology and molecular medicine, Faculty of medicine, Geneva University.

² SIB-Swiss Institute of Bioinformatics. CMU, Michel Servet 1, 1211 Geneva 4, Switzerland

: camille.mary@sib.swiss

Analysis of the human genome has led to the prediction of 20,230 protein coding genes. Currently, about 10% of these genes are still lacking functional annotation. The CALIPHO group aims to characterize some of these gene products, starting with those that have been experimentally validated by proteomics in human samples (uPE1). We describe here the preliminary characterization of two uPE1 proteins: C12orf73 and C15orf61. Both proteins have no functional domain and no characterized orthologs or paralogs. Based on RNAseq data, C12orf73 and C15orf61 are both ubiquitously expressed in human tissues. C15orf61 presents a strong expression in muscles, testis and brain. In databases, both proteins are annotated as secreted due to the presence of a predicted N-terminal signal sequence. However, co-expression studies followed by GO enrichment analysis suggest an involvement in mitochondrial related functions. In accordance with this analysis, both C12orf73 and C15orf61 proteins were found by mass spectrometry in mitochondrial fractions. By immunofluorescence, we confirmed that C12orf73 and C15orf61 were not secreted but localized at the mitochondria. Mitochondria sub-fraction analysis showed that both proteins are anchored in the mitochondrial inner membrane. By downregulating or overexpressing the proteins in human cell lines, we next explored the potential involvement of C12orf73 and C15orf61 in mitochondrial functions. C12orf73 protein level decreases rapidly upon depolarization of the mitochondrial membrane. C15orf61 over-expressed protein was found to co-immunoprecipitate with respiratory complexes 1 and 4 implying a potential involvement in OXPHOS system assembly or activity.

ABSTRACT # 3

Can *Caenorhabditis elegans* data help us to illuminate the human dark proteome?

Paula Duek¹ & Lydie Lane^{1,2}

¹ CALIPHO group, SIB-Swiss Institute of Bioinformatics. CMU, Michel Servet 1, 1211 Geneva 4, Switzerland

² Department of microbiology and molecular medicine, Faculty of medicine, Geneva University. CMU, Michel Servet 1, 1211 Geneva 4, Switzerland

✉: lydie.lane@sib.swiss

The neXtProt query NXQ_00022 applied to the release 2019-01-11 results in a list of 2222 human entries without functional annotation. Out of these entries, 330 are probably not coding and 37 are characterized and await annotation. In order to collect information on the remaining 1855 uncharacterized proteins, we sought if they had homologs in *Caenorhabditis elegans*. This model organism has proven to be a useful experimental system to address biological questions relevant to human development and physiology. By querying OrthoList 2 (<http://ortholist.shaye-lab.org/>), we retrieved worm homologs for 283 uncharacterized human genes. Out of those, only 24% have one-to-one homology relationships. The other relationships are more complex and difficult to interpret: 29% are many-to-many, 44% are 1 worm-to-many humans and 3% are 1 human-to-many worms. We mined WormBase (<https://www.wormbase.org/#012-34-5>) with the 69 worm genes having one-to-one homology to human and found that there are available phenotypic data based on RNAi or allele variants for 34 worm genes. For the remaining 35 worm homologs there are available alleles without reported phenotype or RNAi experiments with negative phenotypic observations. In addition, there are physical or genetic interaction data for 23 worm homologs. A total of 15 homologs have both phenotypic and interaction data and 42 have either phenotypic or interaction data. We are now performing a detailed analysis of the available phenotypic or interaction data for these 42 worm genes in order to formulate new functional hypotheses for their human counterparts. Experimental efforts should be encouraged to collect data for the 27 *C. elegans* genes with one-to-one homology to humans but no phenotypic or interaction data in WormBase.



ABSTRACT # 4

Proteomics and Peptidomics Workflow of Urine Biomarker Discovery and Validation for “All-in-One Urine Test”: Biomarkers for Kidney-site unique injuries and more sensitive ones than albuminuria

Tadashi Yamamoto^{1,2}, Amr Elguoshy¹, Keiko Yamamoto¹, Bo Xu¹, Yoshitoshi Hiro¹ and Japan Chromosome X and 3 teams³.

1 Biofluid Biomarker Center, Niigata University, Niigata, JAPAN

2 Shinrakuen Hospital, Niigata, JAPAN

3 Japan Chromosome X and 3 teams

: tadashiy-bbc@ccr.niigata-u.ac.jp

According to the neXtProt (release 2018-9-3) human protein database, 17,487 entries were annotated as PE1 proteins, 1,728 as PE2, 515 as PE3, 67 as PE4 and 571 as PE5. Previously we searched peptides signatures of the neXtProt-annotated missing entries (PE2-4) in public protein repository databases (Global Proteome Machine [GPM], PeptideAtlas, and SRMATlas) and found several peptides, which had peptide fragment profiles comparable to those found in the SRMATlas for synthetic peptides. These missing protein peptides, which had been identified with almost identical profiles in both GPM/ PeptideAtlas, and SRMATlas were called “Stranded peptides”. We were able to find 41 missing proteins, in which more than 2 distinct “Stranded peptides” had been identified in GPM / PeptideAtlas databases. In the current study, we extended the search for the “Stranded peptides” in against the protein list in human SRMATlas synthetic peptides peptide dataset and mapped them uniquely to the neXtProt missing protein entries. Then, we used the uniqueness checker tool to filter out the peptides that count as variant forms for other proteins. Thereafter, we searched these filtered synthetic peptides against the GPM peptide dataset to get their natural counterparts. Subsequently, these GPM peptides were annotated manually, considering only missing proteins, which can be detected in a certain study by at least 2 peptides. And the unmodified peptides were taken as the priority over the modified ones. Then, the GPM and SRMATlas spectra retrieved were manually filtered and matched each other using our Matching algorithm, in order to get the final list of validated peptides and their corresponding missing proteins. At the end, 25 new missing proteins were identified and validated by 60 peptides (each protein has at least 2 peptides).



ABSTRACT # 5

Proteomics and Peptidomics Workflow of Urine Biomarker Discovery and Validation for “All-in-One Urine Test”: Biomarkers for Kidney-site unique injuries and more sensitive ones than albuminuria

Tadashi Yamamoto^{1, 2}, Keiko Yamamoto¹, Amr Elguoshy¹, Bo Xu¹, Yoshitoshi Hiro¹ and HKUPP³

¹ Biofluid Biomarker Center, Niigata University, Niigata, JAPAN

² Shinrakuen Hospital, Niigata, JAPAN

³ Human Kidney and Urine Proteome Project (HKUPP) Team

✉: tadashiy-bbc@ccr.niigata-u.ac.jp

Background: Various sites in the kidney, such as glomerulus and other sites may be injured more or less in kidney diseases resulting in functional impairments and developing symptoms according to the injured sites and their degrees. Kidney tissue biopsy has been regarded as an only significant examination to evaluate the kidney site-unique injuries. In the current study, we aimed to discover and validate urine biomarkers for the kidney site-unique injuries by proteomic strategy. In addition, we also searched for urine biomarkers, which are more sensitive than albuminuria or proteinuria for glomerular injury.

Methods: By proteomics analysis of human kidney sites (glomerulus, proximal tubule, distal tubule, collecting duct, cortex and medulla), urine and plasma, about one hundred proteins were selected as urine proteins derived uniquely from kidney sites and not from plasma and as biomarker candidates of the kidney sites. For validation of the biomarker candidates, we set up antibody-based assay using a surface plasmon resonance (SPR) device (ProteOn, Bio-Rad) to quantitate these in many urine samples from healthy volunteers and kidney disease patients.

Results: So far, we have established 3 glomerulus-unique injury biomarkers, 2 proximal tubule-unique, 1 distal tubule-unique, 1 collecting duct-unique, 1 intersitium-unique, and 2 urine biomarkers, which are more sensitive than albumin. Concentrations of these proteins increased significantly in urine samples from chronic kidney disease in parallel with clinical stages

Conclusions: Our strategy for urine biomarker discovery and validation seems to work well and expects to complete discovery and validation of biomarkers for all diseases as “All-in-One Urine Test” in the future.



ABSTRACT # 6

MS evidence acquisitions of missing proteins in chromosome 9 by using halo-tag purification system and identification of their cellular roles

HuiSu Kim¹, Yong-In Kim^{1,2}, Hyoung-Min Park¹, Dong Wook Kim¹, and Je-Yoel Cho¹

¹BK21 PLUS Program for Creative Veterinary Science Research and Research Institute for Veterinary Science, College of Veterinary Medicine, Seoul National University, Seoul, South Korea. ²Center for Bioanalysis, Korea Research Institute of Standards and Science, 267 Gajeong-ro, Yuseong-gu, Daejeon, South Korea

✉: jeycho@snu.ac.kr

The NeXtProt data, released on 2018-09-17, reports 2,890 proteins remaining on the missing protein level, also known as PE 2,3,4 level (14.2%). Our goal is to find MS evidence of missing proteins and to identify their biological roles in cells. Obstacles in acquiring MS evidence for the missing proteins arise due to the very low expression level or specific expression pattern in certain cell types or time, and drove us to take advantage of the overexpression system. First, we tried to find missing proteins with no MS evidences in Chromosome 9 using the NeXtProt database. We chose 5 missing proteins: FOXD4, ARID3C, OR1J1, ANKRD18A, and ZNF510. Then, we constructed 5 plasmid DNAs harboring cDNA of the missing proteins fused to Halo-tags and artificially transfected into HEK293T cells. Overexpressed missing proteins were purified with Halo tag purification system and subjected to LC-MS/MS analysis. This strategy allowed great advancement in the detection of MS evidence for missing proteins. Using this method, we obtained MS evidence of 2 missing proteins (FOXD4 and ARID3C) under the rule of Human Proteome Project Data Interpretation Guidelines. In short, we used the following search parameter; peptide FDR < 0.01, protein FDR < 0.01, and 2 or more unique peptides with 7 amino acids. Additional cellular image analysis identified all of the 5 proteins' subcellular localizations. In conclusion, purification and overexpression system we used in the study could be an excellent method to identify MS evidence of missing proteins and characterization of the missing proteins.

Keywords: Human Proteome Project (HPP), Chromosome-centric HPP (C-HPP), Missing protein (MP), Halo tag, Interactome, Characterization of proteins (CP)



ABSTRACT # 7

Analysis of Single Amino Acid Variants from Breast Cancer Using Next-Generation Proteomic Pipeline

Heeyoun Hwang¹, Hyejin Kim^{1,2}, SooYoun Lee¹, Jin Young Kim¹, and Jong Shin Yoo^{1,2}

¹Biomedical OMICS Group, Korea Basic Science Institute, Cheongju 28119, Rep. of Korea

²Graduate School of Analytical Science and Technology, Chungnam National University, Daejeon 34134, Rep. of Korea

✉: heeyounh@kbsi.re.kr

In order to expand the understanding of all human protein coding genes, the chromosome-centric human proteome project (C-HPP) has looked for 'Dark Proteins' of missing proteins (MP) and uncharacterized proteins with unknown function, but also 'Protein Variants' with single amino acid variants (SAAV). Herein, we (Chr #11 team of C-HPP groups) update our customized database of next-generation proteomic pipeline (nextPP) using the latest version of neXtProt (2019-01-11) and GENCODE (V.29, 05-2018). Using the nextPP with customized databases including 13,879,977 number of in-silico tryptic SAAV peptides, we re-analyzed proteomic data from three breast cancer studies with < 1.0% FDR at peptide and proteins level, where 28 out of 31 genes having SAAVs were breast cancer susceptibility genes or had a primary relationship with breast cancer-related genes. The result contains nine genes such as COL1A2 (Collagen Type I Alpha 2 Chain), VCAN (Versican) and MYH9 (Myosin Heavy Chain 9), which are related to metastasis, invasion and migration of human breast cancer. Our customized protein sequence database would be useful to find variant proteins.



ABSTRACT # 8

Exploring the Function of Dark Proteome for Biomedical Applications

Keun Na¹, Chae-Yon Kim¹, Lydie Lane², Christopher M. Overall³ and Young-Ki Paik¹

¹Yonsei Proteome Research Center, Yonsei University, 50 Yonsei-ro, Seodaemun-ky, Seoul, 03722, Korea. Swiss ²Institute of Bioinformatics, Geneva, Switzerland and ³Univ. of British Columbia, Vancouver, Canada

✉: paiky@gmail.com

One of the important goals of the HUPO Chromosome-centric Human Proteome Project (C-HPP) is to correctly define the function of less annotated proteins encoded by their cognate open reading frames on each chromosome in the human genome (1). This can be achieved by collaboration with various research groups and resource pillars within the HUPO community (2). For example, with collaborative efforts by the proteomics community, the number of missing proteins (MPs) has been reduced from ~6000 to 2,129 out of 19,823 human canonical proteins (neXtProt 1-11-2019 release). In 2018, HUPO C-HPP launched a new pilot project over 3 years, neXt-CP50, that aims to systematically explore the biological functions of ~50 uncharacterized proteins. These comprise some MPs and/or annotated proteins (uPE1s), termed Dark Proteins, and performed in a close collaboration by 15-international C-HPP teams (3). The Dark Proteins are rising to form one of the major targets in proteome biology due to the nature of their unknown functionality and low abundance properties in general, which render them as promising drug target candidates, biomarkers, cellular regulators and structural components. We present the status of our work-in-progress on the study of a few selected Dark Proteins that are located in **chromosome 13** by using the combined multi-omics approaches, which employed cell-based assays, CRISPR/cas9 genome editing, Protein-Protein Interaction (PPI) tool (e.g., COFACTOR, I-TASSER) (4) and proteogenomic analyses. We will also discuss some phenotypic changes as well as functional and structural aspects observed from a loss of function mutant of YPRC-DP1. Finally, we address some common technical issues and annotation problems associated with functional characterization of Dark Proteins (Supported by HI13C2098, a grant from the Ministry of Health and Welfare to Y.K.P.).

References:

1. Paik YK et al., A Chromosome-Centric Human Proteome Project to Characterize the Sets of Proteins Encoded in the Genome *Nature Biotechnol.*, March 7, 2012, **30**, 221-223
2. Paik YK, et al. Advances in the Chromosome-Centric Human Proteome Project: Looking to the Future. *Expert Rev Proteomics.*, 2017 Dec; **14**(12):1059-1071
3. Paik YK et al., Launching the C-HPP pilot project neXt-CP50 for functional characterization of identified proteins with no known function. *J Proteome Res.* 2018, **17**, 4242-4050
4. Zhang C, Wei X, Omenn GS, Zhang Y. Structure and Protein Interaction-based Gene Ontology Annotations Reveal Likely Functions of Uncharacterized Proteins on Human Chromosome 17. *J Proteome Res.* 2018, **17**, 4186–4196



ABSTRACT # 9

FusionPro, a versatile proteogenomic tool for identification of novel fusion transcripts and their potential translation products in cells

Chae-Yeon Kim^{1,2}, Keun Na², Saeram Park², Seul-Ki Jeong², Jin-Young Cho², Heon Shin², Min Jung Lee², Gyoonee Han³, and Young-Ki Paik^{2*}

Yonsei Proteome Research Centre, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea

✉: jinyoung@kbsi.re.kr

Fusion proteoforms are translation products derived from gene fusion. Although very rare, the fusion proteoforms play important roles in biomedical science. For example, fusion proteoforms influence the development of tumors by serving as cancer markers or cell cycle regulators. Although numerous studies have reported bioinformatics tools that can predict fusion transcripts, few proteogenomic tools are available that can predict and identify proteoforms. In this study, we develop a versatile proteogenomic tool “FusionPro”, which facilitates the identification of fusion transcripts and their potential translatable peptides. FusionPro provides an independent gene fusion prediction module and can build sequence databases for annotated fusion proteoforms. FusionPro shows greater sensitivity than the available fusion finders when analyzing simulated or real RNA sequencing datasets. We use FusionPro to identify 18 fusion junction peptides and three potential fusion-derived peptides by MS/MS-based analysis of leukemia cell lines (Jurkat and K562) and ovarian cancer tissues from the Clinical Proteomic Tumor Analysis Consortium. Among the identified fusion proteins, we molecularly validate two fusion junction isoforms and a translation product of *FAM133B:CDK6*. Moreover, sequence analysis suggests that the fusion protein participates in the cell cycle progression. In addition, our prediction results indicate that fusion transcripts often have multiple fusion junctions and that these fusion junctions tend to be distributed in a non-random pattern at both the chromosome and gene levels. Thus, FusionPro allows users to detect various types of fusion translation products using a transcriptome-informed approach and to gain a comprehensive understanding of the formation and biological roles of fusion proteoforms.

[This study was supported by grants from the Korean Ministry of Health and Welfare: [HI13C2098] International Consortium Project and [HI16C0257] (to Y.-K. Paik). The data used in this study were generated by the Clinical Proteomic Tumor Analysis Consortium (NCI/NIH)].



21st Chromosome-Centric Human Proteome Project Symposium
“Illuminating the dark proteome” - May 12-14, 2019, Saint-Malo, France

ABSTRACT # 10

Cell type-specific expression of testis elevated genes based on transcriptomics and antibody-based proteomics

Charles Pineau^{1,2}, FERIA Hikmet Noraddin³, Nathalie Melaine^{1,2}, Per Oksvold⁴, Cheng Zhang⁴, Shuqi Chen⁴, Linn Fagerberg⁴, Mathias Uhlén⁴, Cecilia Lindskog³

¹ Univ Rennes, Inserm, EHESP, Irset (Institut de recherche en santé, environnement et travail) - UMR_S 1085, F-35042 Rennes cedex, France

² Protim, Univ Rennes, F-35042 Rennes cedex, France

³ Uppsala University, Department of Immunology, Genetics and Pathology, Rudbeck Laboratory, 75185 Uppsala, Sweden

⁴ Science for Life Laboratory, School of engineering Sciences in Chemistry, Biotechnology and health, KTH - Royal Institute of Technology, 17121 Stockholm, Sweden

✉: cecilia.lindskog@igp.uu.se

Spermatogenesis takes place in the testis that is undoubtedly the most complex organ in the human body. This physiological process involves thousands of genes and proteins that are activated and repressed, making testis the organ with the highest number of tissue-specific genes. However, the function of a large proportion of the corresponding proteins remains unknown. In this context, the testis has also been shown to harbor many missing proteins, defined as products of protein-coding genes that lack experimental mass spectrometry evidence.

Here, an integrated omics approach was used for exploring the cell type-specific protein expression of genes with an elevated expression in the testis. By combining RNA-Seq with immunohistochemistry, 516 proteins with distinct testicular protein expression patterns were identified and selected for a more detailed characterization of their in situ expression in eight different testicular cell types. Eighty-nine of the proteins have an unknown function, and 66 proteins have previously been classified as missing proteins. Testis elevated expression levels based on RNA-Seq suggest that these proteins could be related to testis-specific functions, and analysis of their cell type-specific expression patterns is a first step towards further characterization of their function.

Results of our in-depth characterization confirmed previously well-known protein expression patterns related to the spermatogenetic process, and six distinct clusters of cell type-specific expression at different stages of spermatogenesis could be identified. The analysis highlighted numerous poorly characterized proteins in each of these clusters whose expression overlapped with that of known proteins involved in spermatogenesis. Furthermore, we characterized the in situ distribution of several proteins that previously lacked information on spatial resolution and cell type-specific expression within the testis.

The strategy of combining transcriptomics with antibody-based proteomics is an attractive approach for further characterization of missing proteins. By determining the cell type-specific expression, clues about the potential protein functions were provided, and the analyzed proteins constitute important targets for further testis-specific research in male reproductive disorders.

How the ProteoRE platform can contribute to the exploration of the dark side of the Human proteome

Florence Combes¹, David Christiany^{1,2}, Nguyen Lien^{1,2}, Virginie Brun¹, Valentin Loux², Charles Pineau^{3,4}, Yves Vandenbrouck¹

1. Univ Grenoble Alpes, CEA, INSERM, BGE U1038, F-38000, Grenoble, France
2. Univ. Paris-Saclay, INRA, MAIAGE Unit, F-78352, Jouy-en-Josas, France
3. Univ Rennes, Inserm, EHESP, IRSET, UMR_S 1085, F-35042 Rennes, France
4. Protim, Univ Rennes, F-35042 Rennes, France

✉: yves.vandenbrouck@cea.fr

With the increased simplicity associated with producing MS-based proteomics data, the bottleneck has now shifted to the functional analysis and exploration of large lists of expressed proteins to extract meaningful biological knowledge. Bioinformatics resources are often spread and disseminated under different forms (program/libraries/software/web tools and databases) and their access is rather limited for researchers without programming experience or no in-house bioinformatics support. As a consequence, interpretation of their data by experts remains a tedious and time-consuming process, and potentially error-prone (e.g., due to manual handling or input error, use of outdated resources). The ProteoRE (Proteomics Research Environment) aims at fulfilling this need by centrally providing an online research service to assist biologists/clinicians in the interpretation of their proteomics data in a unified framework. Built upon the Galaxy environment, this web-based platform for computational biomedical research, allows researchers to apply a large range of dedicated bioinformatics tools and data analysis workflows on their data, share their analyses with others, and enable tiers to repeat the same analysis while keeping tracks of the overall process. Currently, ProteoRE implements 19 tools organized into four subsections for: i) data manipulation; ii) human and mouse species annotation; iii) functional analysis; and iv) pathway analysis along with graphical representations. The ProteoRE platform has been designed in collaboration with biomedical researchers and has been implemented for the functional analysis of a human MS/MS proteomics sample [1] and the selection of the selection of candidate proteomics biomarkers of human disease [2]. ProteoRE also provides online support, tutorials and is in free access: <http://www.proteore.org>.

In the frame of the Human proteome Project (HPP), the “dark proteome” is defined as the protein parts list of the human proteome having no function either annotated from experimental characterization or predicted by homology to other proteins [3, 4]. The detection, the functional annotation of yet function orphan proteins (designated as uPE1) as well as their mechanistic implications in biology and disease are of major interest and recently lead to the launch of the neXt-CP50 pilot project [3]. We will illustrate how ProteoRE can assist experts in this quest by focusing on a subset of “dark proteins” specifically enriched in the human reproductive system [4].

1. Lacombe M et al. Proteomic characterization of human exhaled breath condensate. *J Breath Res.* 2018. 12(2):021001.
2. Nguyen L. et al., Designing an In Silico Strategy to Select Tissue-Leakage Biomarkers Using the Galaxy Framework. *Methods Mol Biol.* 2019;1959:275-289.
3. Paik YK et al., Launching the C-HPP neXt-CP50 Pilot Project for Functional Characterization of Identified Proteins with No Known Function. *J Proteome Res.* 2018. doi: 10.1021/acs.jproteome.8b00383.
4. Melaine N et al. Deciphering the Dark Proteome: Use of the Testis and Characterization of Two Dark Proteins. *J Proteome Res.* 2018 Aug 31. doi: 10.1021/acs.jproteome.8b00387.

12 - LECTURE ABSTRACT

Bridging C and B/D HPP to define the biological context of human proteins

Fernando J. Corrales and Miguel Marcilla

Functional Proteomics Laboratory. Centro Nacional de Biotecnología, CNB-CSIC. Madrid, Spain

✉: fcorrales@cnb.csic.es

The Human Proteome Project (HPP) aims to provide an accurate and reliable description of the human proteome as well as standardized detection and quantitation analytical resources, to enhance our understanding of human biology and disease. Pursuing these goals the HPP is organized in four technological pillars: pathology, antibody-based, mass spectrometry and knowledgebase and two main initiatives: the Chromosome Centric and the Biology and Disease driven HPP (C- and B/D-HPP). Close interaction between these initiatives ensures cross-fertilization and efficient progress in the functional annotation of the human proteome. Many tissues and cell types must be carefully screened to complete the list of the about 20,000 gene products and their major proteoforms. In this regard it is especially relevant the search for missing proteins (about 15% of the human proteins with scarce or non-mass spectrometry data supporting their expression) that is currently being accomplished by the 50 MPs Challenge. Besides detection, the functional annotation of yet function orphan proteins (uPE1) as well as their mechanistic implications in biology and disease are items of major interest (CP50 Challenge). This C-HPP-launched projects can greatly benefit from the activity of the 19 B/D teams that have generated extensive proteomic data in valuable/rare biological samples and count on expertise to dig deep into the functional implications of the dark proteome in health and disease. As an example, a preliminary study on APIP will be presented. APIP was a former uPE1 protein that has been identified as a methionine salvage pathway enzyme. Since the maintenance of this pathway is central to preserve quiescent and differentiated hepatocytes, the Liver HPP and Chromosome 2 teams have started a collaboration to characterize the phenotype associated to the deregulation of APIP in human liver cells.

Strategy for MPs identification

Clemente LF¹, Marcilla M², Hernandez ML¹, Ramos-Fernandez A³, Liger G⁴, Gil C^{1,5}, Corrales FJ².

1 Proteomics Unit, Faculty of Pharmacy, Complutense University of Madrid (UCM), Plaza Ramon y Cajal s/n, 28040 Madrid, Spain.

2 Proteomics Unit, Spanish National Biotechnology Centre (CNB-CISC), Darwin 3, 28049 Madrid, Spain.

3 Proteobotics SL, Darwin 3, 28049 Madrid, Spain.

4 Andalusian Public Health System Biobank, Avenida Del Conocimiento s/n, 18016 Granada, Spain.

5 Department of Microbiology & Parasitology, Faculty of Pharmacy, Complutense University of Madrid (UCM) and Ramon y Cajal Institute of Health Research (IRYCIS), Plaza Ramon y Cajal s/n, 28040 Madrid, Spain.

: luisfecl@ucm.es

On the scope of the C-HPP project, we study the proteomic profiling of human Mesenchymal Stem Cells (hMSCs) derived from umbilical cord or adipose tissue. The hMSC are undifferentiated cells that can turn into some kinds of specific cells, because of this, their proteomic profile will be different from that of differentiated cells and could express some proteins that will be not present in normal adult tissues. In this way, we found evidence of the expression of HAS1 (hyaluronan synthase 1, NX_Q92839-1) in both kind of cells, and confirmed its identification by MS/MS spectra, and the comparison of four heavy-labelled peptides precursor/transition signals and retention times with those of the endogenous peptide signals and times. In this analysis, we provide suggestive evidence of the presence of some additional missing proteins that deserve further analysis for their confirmation. Currently, taking into account the new NeXtProt release (02/13/2019), 4 of that missing proteins remain in this status, DNAJB7 (Q7Z6W7-1), C8orf58 (Q8NAV2-1), OR1M1 (Q8NGA1-1) and MEIS3P2 (A8K0S8-1). Our goal is to get evidence of these proteins in the hMSCs using synthetic peptides selected among those of the predicted highly observable peptides list available in PeptideAtlas for each protein. The synthetic peptides will be used to validate the spectra, retention time and transition relative intensity of the endogenous peptides.

ABSTRACT # 14

Exploration of unique samples to characterize “missing proteins”: a RAD-HPP example

Cristina Ruiz-Romero, Patricia Fernandez-Puente, Valentina Calamia and Francisco J. Blanco

Unidad de Proteómica, Grupo de Investigación de Reumatología (UP-GIR), Instituto de Investigación Biomédica de A Coruña – Hospital Universitario A Coruña, Spain

✉: Cristina.Ruiz.Romero@sergas.es

The Rheumatic and Autoimmune Diseases Initiative of the Human Proteome Project (RAD-HPP) was launched in 2017. Currently, 6 research groups from 5 countries participate in this initiative. Its primary scientific goals were to define the proteome of human joint tissues (cartilage, synovial membrane, subchondral bone or meniscus), and to assemble prioritized lists of proteins clinically relevant in RAD using the ‘popular proteins’ strategy and text mining tools.

In order to address the first of these two main objectives, RAD-HPP groups have performed a number of proteomic studies on a plethora of different samples that can be obtained from the human joint and disease-related biological fluids. Among these, we have recently analyzed conditioned media from osteoarthritic and healthy human articular cartilage explants, in an attempt to discover novel neo-peptides released as a consequence of the dynamic turnover (either physiological or pathological) of its extracellular matrix (ECM). This study allowed the identification of specific peptides derived from human proteins, with a significant enrichment in those ECM-related. Interestingly, our approach enabled the detection of “missing proteins” such as C2orf8/SNORC (PE2), which was not previously detected at the protein level in humans. Intriguingly, SNORC has a known homologue in mouse that seems to play a role in the regulation of chondrocyte maturation and postnatal endochondral ossification, a fact that turns this protein into an interesting target in cartilage biology studies.

In addition, we have also been able to deepen the functional characterization of other recently annotated proteins. This is the case of Augurin (C2orf40/AUGN), a protein that was first reported to be expressed in rat brain and to have a putative hormonal activity in the central nervous system. In our study, we found 15 different endogenous peptides of this small (17 kDa) protein that were released from human articular cartilage. This strongly suggests a role of Augurin in the turnover processes of this particular tissue.

Altogether, this work provides an example of how the analysis of unique samples promoted from the different B/D-HPP initiatives may aid in the identification and characterization of “missing proteins” carried out by the C-HPP.

Smelling the dark proteome

Enrique Santamaría

Proteored-ISCI, Clinical Neuroproteomics Unit, Navarrabiomed, Complejo Hospitalario de Navarra (CHN), Universidad Pública de Navarra (UPNA), IdiSNA. Irunlarrea, 3 31008 Pamplona, Spain

: enrique.santamaria.martinez@navarra.es

Olfactory dysfunction is a prodromal event in neurodegenerative diseases, being considered as an early biomarker for diagnosis. However, the molecular architecture associated with decreased smell function are not completely understood. LC-MS/MS analysis have allowed us to characterize the protein content of 3 consecutive areas involved in olfactory transduction: i) the olfactory epithelium, ii) the olfactory bulb, and iii) the olfactory tract. 178 out of 7,100 identified proteins are uPE1, corresponding to approximately 14% of the uPE1 referenced in neXtProt-2019 database. Public RNA-seq data indicate that 6% of the olfactory uPE1 proteins are considered highly enriched genes for specific cell types in the brain. A massive olfactory proteotyping across 7 neurological syndromes has pointed out commonalities and differences between human tauopathies and synucleinopathies, revealing 12 deregulated uPE1 proteins across different types of dementia. In summary, primary olfactory areas appear to be a relevant system to analyse the dark proteome. Functional studies will help us to decipher the role of the differential dark proteins in the presence or absence of neuropathological substrates in olfactory cells.

Progress on Identifying and Characterizing the Human Proteome: 2018-2019 Metrics from the HUPO Human Proteome Project

Gilbert S. Omenn, Lydie Lane, Christopher M. Overall, Fernando J. Corrales, Jochen M. Schwenk, Young-Ki Paik, Jennifer E. Van Eyk, Siqi Liu, Michael Snyder, Mark S. Baker, Stephen Pennington, and Eric W. Deutsch

Summary: The Human Proteome Project (HPP) annually reports on progress throughout the field in credibly identifying and characterizing the human protein parts list and making proteomics an integral part of multi-omics studies in medicine and the life sciences. NeXtProt release 2019-01 contains 17,694 PE1 proteins, 89.3% of all neXtProt predicted PE1-4 proteins, up from 17,470 in release 2018-01 (Table 1, Figure 1). Conversely, the number of neXtProt PE2,3,4 missing proteins has been reduced from 2949 to 2579 to 2186 to 2129 over the past three years. Figure 2 shows the distribution of proteins by nature of the scientific evidence. PeptideAtlas is the primary source of mass spectrometry data for neXtProt; PeptideAtlas gained 625 canonical proteins as of 2018 and 394 more from 2018 to 2019 (Table 1). The largest reason for missing proteins is low abundance. Meanwhile, the Human Proteome Project utilizes the Human Protein Cell, Pathology, and Tissue Atlases and applies SRM multiplex targeted proteomics for organ-specific popular proteins for various disease categories.

Table 1. neXtProt Protein Existence Evidence Levels from 2012 to 2019: Progress in Identifying PE1 Proteins and PeptideAtlas Canonical Proteins (more stringent guidelines imposed in 2016)

PE Level	2012-02	2013-09	2014-10	2016-04	2017-01	2018-01	2019-01
1: Evidence at protein level	13,975	15,646	16,491	16,518	17,008	17,470	17,694
2: Evidence at transcript level	5205	3570	2647	2290	1939	1660	1548
3: Inferred from homology	218	187	214	565	563	452	510
4: Predicted	88	87	87	94	77	74	71
5: <i>Uncertain or dubious</i>	622	638	616	588	572	574	576
Human PeptideAtlas canonical proteins	12,509	13,377	14,928	14,629	15,173	15,798	16,192

^a PE1/PE1+2+3+4 = 17,694/19,823 = 89.3%; ^b PE 2+3+4 = 2129 "missing proteins"

HPP Initiatives to characterize the “Dark Proteome” of missing (PE2,3,4) and functionally unannotated uPE1 proteins (see JPR 2018):

- Missing Proteins identified in the 26 papers of the 6th annual HPP special issue of J Proteome Research: 13 from cerebrospinal fluid (Macron); 1, mesenchymal stem cells (Clemente); 2 using mirror proteases (He); 5, olfactory epithelium (Hwang); 1, HeLa cells (Robin); 3, mitochondria (Ronci); 14 with multiple proteases (Sun); 26, embryonic stem cells (Weldemariam); 30, membrane proteins (Zhang); 9 from unusual tissues (Sjostedt), totaling 108, plus 107 proposed from MassIVE database (Pullman).
- A new initiative uses experimental and informatics approaches to functionally annotate 1260 uPE1 proteins. This is a companion to the Missing Proteins Challenge.
- HPP annual special issue #7: deadline for manuscript to JPR is 31 May 2019.

Fig 1. These flowcharts depict the changes in neXtProt PE1-5 categories from release 2017-01 to 2018-01 (left) to 2019-01 (right). First, 431 missing proteins were promoted to PE1, 44 new SwissProt proteins added as PE1, 3 PE1 proteins demoted to PE2,3,4 MPs, 10 PE1 deleted. Then 218 promoted to PE1, 78 new, 65 demoted, 7 deleted = 224 net (2019).

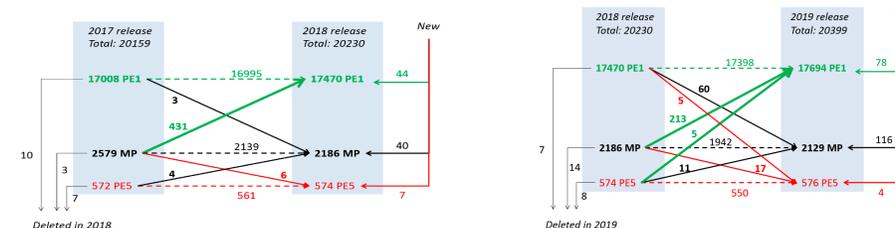
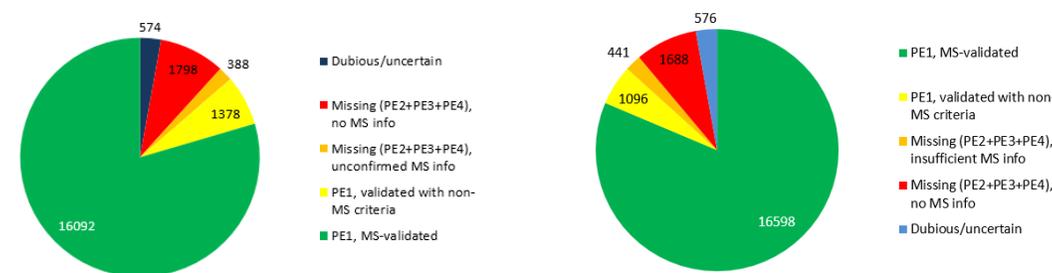


Fig. 2. Pie charts show type of evidence data for PE1 proteins, as well as the numbers in other categories as of neXtProt releases 2018-01 and 2019-01. There are 16,598 PE1 proteins identified with MS data compliant with HPP Guidelines, of which 98% are canonical in PeptideAtlas. There were 1096 additional PE1 proteins, identified by protein-protein interactions, disease mutations, PTMs or proteolytic processing, Edman sequencing, 3D structures, or Ab-based techniques. PE2,3,4 missing proteins comprise those with no MS data (1688) plus with insufficient or unconfirmed MS data (441).



ABSTRACT # 17

A Blinded Comparison of Function Annotation of uPE1 Proteins using the I-TASSER/COFACTOR Pipeline and the 2018-2019 Additions to neXtProt

Chengxin Zhang¹, Lydie Lane², Yang Zhang¹, Gilbert S. Omenn¹

¹ Department of Computational Medicine & Bioinformatics, University of Michigan, USA

² Swiss Institute of Bioinformatics, Geneva, Switzerland

: zcx@umich.edu

Understanding the function of human proteins is essential to decipher the molecular mechanisms of human diseases and phenotypes. Of the 17 470 human protein-coding genes in the neXtProt 2018-01-17 release with unequivocal protein existence evidence (PE1), 1260 proteins did not have specific characterized functions (uPE1). Last year we reported a hybrid pipeline that creates protein structure prediction using I-TASSER and infers functional insights for the target protein from the templates recognized by COFACTOR (C. Zhang et al, JPR 2018; PMIC 30265558). As a case study, the pipeline was applied to all 66 uPE1 proteins from human chromosome 17 as of neXtProt 2017-07-01. Benchmark testing on a control set of 100 randomly-selected well-annotated Chr 17 proteins showed high Gene Ontology (GO) term prediction accuracies of 0.69 and 0.57 for molecular function (MF) and biological process (BP), respectively, respecting neXtProt exclusions of terms not sufficiently specific. For the chromosome 17 uPE1 proteins, the I-TASSER/COFACTOR pipeline confidently assigned MF and/or BP for 13 and 33 proteins, respectively. This year our colleagues at neXtProt informed us that 44 uPE1 proteins were under review for assignment of function; 25 of the 44 were annotated in time for the 2019-01-11 release: 8 and 21 of the 25 acquired new GO terms for MF and BP, respectively. We submitted to neXtProt blinded predictions of MF and BP GO terms and a free-text function interpretation based on predicted GO terms for each target before the release of the curated results for those 25 protein entries. For 10 of the 25, a good to excellent match of free-text function annotation was obtained, as shown in the table below, and quantitated with F-scores. Another 4 targets have good MF and/or BP GO term prediction accuracy demonstrated in GO trees, despite inconsistent free-text interpretation. [neXtProt disregards CC terms.] This ratio is roughly comparable to our benchmark and Chr 17 results. We await the curation results for the remaining 19 of the original 44. This study supports a novel computational approach to systematically annotate protein function in the human proteome which can be applied by or for each C-HPP chromosome team. The predictions provide useful insights to guide experimental design and follow-up validation studies of these uPE1 “dark proteins”.

Comparison of Function Annotation from I-TASSER/COFACTOR Prediction & neXtProt Curation

Accession	Our function annotation	neXtProt annotation (excerpt)
O75363	Myelin	Myelination
O75677	Ubiquitin-protein transferase	Promoting cyclin B1/CCNB1 and CDK1 proteasomal degradation
P0C870	Oxidoreductase for oxidative demethylation	Oxidoreductase activity
Q5VTQ0	Cholesterol metabolism	Regulates HDL cholesterol metabolism
Q6AI39	Regulation of expression at chromosome level	Changes chromatin structure by altering DNA-histone contacts within a nucleosome
Q8IUW5	p38 MAPK cascade regulation	Induces activation of MAPK14/p38 cascade
Q8NDM7	Ciliary structure and movement	Flagellar protein involved in sperm flagellum axoneme organization and function
Q8WTR8	Anatomical structure morphogenesis.	Neurogenesis.
Q96M27	Protein kinase A regulation.	Activation of protein kinase A activity.
Q9C0D6	Binding of cytoskeleton, e.g. actin.	Regulates both actin and microtubule dynamics.



21st Chromosome-Centric Human Proteome Project Symposium
"Illuminating the dark proteome" - May 12-14, 2019, Saint-Malo, France

ABSTRACT # 18

Chromosome 17: neXt-MP50 Challenge Completed!

Hongjiu Zhang, Yuanfang Guan, Gilbert S. Omenn

Department of Computational Medicine & Bioinformatics, University of Michigan, USA

[✉: zcx@umich.edu](mailto:zcx@umich.edu)

Chromosome 17 had 148 PE2+3+4 Missing Proteins (MPs) when C-HPP announced the neXt-MP50 Challenge in Oct 2016. As of neXtProt release 2019-01, Chromosome 17 has 98 remaining MPs, down from 105 in the 2018-01 release and 148 in the baseline. The pathway from 148 to 105 was extensively documented by Siddiqui et al [1]. Table 1 shows the 13 MPs promoted to PE1, the 5 PE1 proteins demoted to PE2 or 3, plus 3 new neXtProt entries directly assigned to PE1 (CD300H, SPEM3, and SMIM36). Among the newly promoted 13 MPs now PE1, RNF222, SLC16A5, and EVPLL represent proteins we highlighted with potential stranded peptide spectra [1]; they are now confirmed with fresh spectral evidence. Among the 11 promoted from PE2 to PE1, ADORA2b, C17orf64, and RNF112 had been identified as priority candidates [1]. In addition to the 5 PE1 proteins demoted from PE1, neXtProt 2019-01 removed a former PE2 entry (NBR2), demoted a former PE2 entry (FAM215A) to PE5, and included 3 new MPs for Chromosome 17: RNF227 (PE2) and ANKRD40CI (PE4) promoted from PE5 and PVALEF, a new PE3 entry. neXtProt is certainly a dynamic database!

Chromosome 17 is the first of the 24 chromosomes (excluding mitochondria) to achieve the neXt-MP50 Challenge goal. Similar analyses by or for the other chromosome teams could accelerate progress across the entire proteome.

Table 1. Changes in the PE1 protein class between 2018-01 and 2019-01 versions of neXtProt database

<i>Proteins that are promoted to PE1 from 2018-01 to 2019-01</i>		
neXtProt ID	Gene name	Update
NX_A6NF36	CCDC182	11 promoted from PE2
NX_A8MZ36	EVPLL	
NX_O15375	SLC16A5	
NX_P29275	ADORA2B	
NX_Q6DHY5	TBC1D3G	
NX_Q6UXU6	TMEM92	

NX_Q86WR6	C17orf64	
NX_Q8IVW1	ARL17B	
NX_Q9BYQ8	KRTAP4-9	
NX_Q9ULX5	RNF112	
NX_Q9UQ05	KCNH4	
NX_A8MTY7	KRTAP9-7	1 promoted from PE3
NX_A6NCQ9	RNF222	1 promoted from PE4
NX_A0A0K2S4Q6	CD300H	3 new entries
NX_A0A1B0GUW6	SPEM3	
NX_A0A1B0GVT2	SMIM36	
<i>Proteins that are demoted from PE1 from 2018-01 to 2019-01</i>		
neXtProt ID	Gene name	Update
NX_O14610	GNGT2	4 demoted to PE2
NX_O14894	TM4SF5	
NX_Q07627	KRTAP1-1	
NX_Q8N9I5	FADS6	
NX_Q9BYP8	KRTAP17-1	1 demoted to PE3

[1] Siddiqui O, Zhang H, Guan Y, Omenn GS. Chromosome 17 Missing Proteins: recent progress and future directions as part of the neXt-MP50 Challenge. *J Proteome Res* 2018;17: 4061–4071.

Proteomics on the way to reverse Avogadro number

Alexander I. Archakov¹, Ekaterina V. Ilgisonis¹, Andrey V. Lisitsa¹, Elena A. Ponomarenko¹, Victor G. Zgoda¹

1 – Institute of Biomedical Chemistry, RAS, Moscow, Russia

✉: alexander.archakov@ibmc.msk.ru

The major challenges of the C-HPP are insufficient analytical sensitivity of proteomic technologies and the complexity of the proteome. The increased analytical sensitivity of proteomic technologies and the combined results from transcriptomic and proteomic analyses of a single chromosome of individual sample has facilitated the study of proteoforms. For example, for the 275 protein-coding genes of human chromosome 18, 85% of the transcripts were discovered using next-generation sequencing (NGS) and polymerase chain reaction (PCR), and 45% of the proteins were detected using shotgun and selected reaction monitoring technology, in both liver tissue and the HepG2 cell line. To increase proteome coverage, combination of shotgun technology and selected reaction monitoring with two-dimensional alkaline fractionation has been recently developed by using UPS 1 and 2 (Sigma Aldrich) sets as the “gold standard”. To detect proteoforms that cannot be identified by such technologies, nanotechnologies such as combined atomic force microscopy with molecular fishing and/or nanowire detection were used. Both technologies provide a powerful tool for single molecule analysis, by analogy with nanopore sequencing during genome analysis. We believe that such approach could be used for detection of the «missing» proteins of the single chromosome during C-HPP.

ABSTRACT #21

Focus on Gene Editing for human Chr18 annotation

Ekaterina Poverennaya¹, Leonid Kurbatov¹, Petr Sergiev², Andrey Lisitsa¹

¹ Orekhovich Institute of biomedical chemistry, 10/8 Pogodinskaya str., Moscow, Russia

² Lomonosov Moscow State University, 1 Leninskie gory, Moscow, Russia

: k.poverennaya@gmail.com

Gene editing based on CRISPR-Cas9 system opens up broad possibilities for the study of the genome and its transfer to phenotype. The changes of gene will be reflected in this expression and translation processes. Therefore, in case of gene editing we really work with three levels of realization of genomic information - genome, transcriptome and proteome which allows us based on some data like different targets or biomaterial not only describe the function of target gene but also to approach understanding of all cell processes' details.

The scheme of CRISPR-Cas9 system is described as delivering to cell the guide RNA-Cas9 complex via plasmid where guide RNA is complementary to target DNA. The DNA target must be adjacent to a short stretch of sequence termed PAM that is met in DNA approximately every 20-30 nucleotides. And Cas9 endonuclease causes a double strand break at the target site. And in the moment of put the native system of DNA repair we have an opportunity to add something new to target gene.

Our idea is to add a tag to the gene to further extract the corresponding protein. Using tagged protein is desirable to fish out interacting partners, and thereby gaining clues to their function that can be followed sequentially by focused functional screens to decipher their biological role and activity.

According to start neXt-CP50 challenge there were 16 uncharacterized proteins for human Chr18. Despite the simplicity of gene editing procedure based on CRISPR-Cas9 system, each protein required the development of a separate protocol. We combine our experience to create common pipeline for add tag (HA-tag) to interesting gene.

ABSTRACT # 22

Y chromosome genes play roles in sex-dependent development outside of sex determination

Ghasem Hosseini Salekdeh, Sara Taleahmad, Anna Meyfour, Hossein Baharvand

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

: hsalekdeh@gmail.com

Despite the small number of Y chromosome genes, their adequate expression is required for regulation of transcription, translation, and protein stability of male individuals beyond sex-determination. In addition to their roles in male infertility, Y chromosome has an inevitable role in the sexual dimorphism of healthy and disease phenotypes and development. We previously reported that the Y chromosome genes, *DDX3Y*, is associated to neural induction in human embryonal carcinoma NTERA-2 cell line. Furthermore, *TBL1Y* may play an important role in cardiac development. We recently observed that another Y chromosome gene, *KDM5D*, may also contribute in cardiac differentiation of human embryonic stem cells (hESCs). We also found that most of the Y chromosome genes like *SRY* showed high expression level in primed versus naïve hESCs. Therefore, we hypothesize that *SRY* prevents WNT/ β -catenin signaling by interaction and inhibition of β -catenin translocation. By the loss of function approach through depletion of *SRY* of the primed cells, we observed that the expression of pluripotency markers and WNT signaling target genes such as *Brachyury*, *GCBM*, *TBX2* and *TBX3* increased in si-*SRY* treated cells. In conclusion, we revealed that inhibition of *SRY* results in nucleus translocation of β -catenin and up-regulation of WNT signaling pathway which important to naïve conversion. Our findings present a potential role for the contribution of Y chromosome genes in sex-dependent development outside of sex determination.

The role of proteogenomics to reveal the molecular mechanisms of COPD and Head and Neck cancer

Corry-Anke Brandsma^{1,2}, Victor Guryev^{2,3}, Frank Klont⁴, Martijn Vochteloo^{3,5}, Wim Timens^{1,2}, Ana Ciconelle⁴, Daria Vedernikova^{2,3}, Rainer Bischoff⁴, Maria Johansson⁶, Ekaterina S. Ovchinnikova^{3,7}, Johan Malm^{6,8}, Linda Bras⁹, Renee Verhoeven⁹, Gyorgy Marko-Varga⁶, Thomas E. Fehniger⁶, Gyorgy Halmos⁹, Maarten van den Berge^{2,10}, Peter Horvatovich⁴

¹ University of Groningen, University Medical Center Groningen, Department of Pathology and Medical Biology, Groningen, the Netherlands

² University of Groningen, University Medical Center Groningen, Groningen Research Institute for Asthma and COPD, Groningen, the Netherlands

³ European Research Institute for the Biology of Ageing, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands

⁴ University of Groningen, Department of Analytical Biochemistry, Groningen Research Institute of Pharmacy, Groningen, the Netherlands

⁵ Hanze University of Applied Sciences, Institute for Life Science & Technology, Groningen, the Netherlands

⁶ Lund University, Center of Excellence in Biological and Medical Mass Spectrometry, Biomedical Center, Lund, Sweden

⁷ University of Groningen, University Medical Center Groningen, Department of Cardiology, Groningen, the Netherlands

⁸ Lund University, Department of Translational Medicine, Malmö, Sweden

⁹ Department of Otorhinolaryngology, Head and Neck Surgery, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands.

¹⁰ University of Groningen, University Medical Center Groningen, Department of Pulmonary Diseases, Groningen, the Netherlands

: peter.horvatovich@gmail.com

Proteogenomics data integration offers the possibility to assess the effect of genome variability on the expressed proteome and subsequently on the contribution of the DNA variants to complex diseases. This talk will present findings of proteogenomics analyses of lung tissue from late-stage COPD patients and head and neck cancer tissue in young and elderly patients. The results include identifications of peptides outside of the reference proteome such as provided by Ensembl, Uniprot and neXtProt databases. We utilized extended proteome reference from potentially translated polyadenylated transcript sequences and investigated level differences and similarities between polyadenylated transcripts and the proteome in human lung tissue and laryngeal squamous cell carcinoma. The proteogenomics analysis revealed several single amino acid variants, new antibody related immune peptides as well as novel variants that contribute to the molecular etiology of these complex diseases. The presentation will discuss the bioinformatic and statistical aspects of proteogenomics data integration as well as the different approaches used to detect novel protein sequence variants using alignment and de novo assembly of polyadenylated transcript data obtained with next generation sequencing.

ABSTRACT # 25

Splice Variants in Glioblastoma – Chromosome 12 and Beyond

Hari P S, Pranali S, Raksha A G, Nameeta Shah, Ravi Sirdeshmukh

Mazumdar Shaw Center for Translational Research and Institute of Bioinformatics, Bangalore, India

[✉: ravisirdeshmukh@gmail.com](mailto:ravisirdeshmukh@gmail.com)

Splice variants contribute significantly to the diversity of proteins and human biology. It is estimated that over 90% of the human genes are alternatively spliced. Alternative splicing in cancer could be an actively regulated process and carries potential of providing specific biomarkers. We have been trying to understand the role of the known splice variants as well as identify novel ones in glioblastoma (GBM, grade IV glioma), a highly malignant brain tumor. We want to study them in following perspectives: High confidence identification, tissue or cell specific expression, if any, structural aspects and clinical significance of the variants. I will describe three of our ongoing efforts in this direction.

1. GPCR 56 – an adhesion GPCR, yet to be understood completely, is known to have short and long variants differing in the 5' region which is also the ligand-binding domain. GPCR 56 is differentially regulated in GBM and the two variants show different expression patterns in foetal and adult brain and sub-types of GBM.

2. Proteogenomic analysis and identification of a novel variant of Neural cell adhesion molecule (NCAM1), important for cell adhesion and signalling. The novel variant appears to constitute significant proportion of NCAM1 in brain, differentially expressed in GBM. Domain analysis indicates that the novel exon contributes extra low-complexity region to the protein that may be important for protein-protein interactions. Novel splice variants of two other proteins identified include a variant of Hepacam – another cell adhesion molecule and a variant of a brain specific protein PDE6B involved in transmission of visual signals.

3. Using RNA SpliceSeq resource, we have mined the splicing events identified for differentially regulated genes in Glioblastoma and other grades of glioma, across all chromosomes. Out of a total 4402 splicing events covering 2504 differentially expressed genes from GBM, Chromosome 12 harbours 312 differentially expressed splice variants mapping to 176 genes that include oncogenes (ERBB3), CDKs, IGF1, and others. Other chromosomes enriched with splice variants are Chromosome 1 (232 genes), 11 (184 genes), 19 (147 genes) 17 (58 genes) and chromosome 7 (130 genes). We would like to pursue an integrated effort for generating biological and clinical knowledge about key splice variants, using experimental and bioinformatics approaches.