



The neXt-50 Challenge

September 16, 2017 status report of next-50 Challenge by individual Chromosome groups.

	PI	MPs	Silver MPs	JPR SI	JPR SI	JPR SI	Other Papers
Ch				Submitted	In press	In revision	
1	Ping Xu			2	1	1	0
2	Lydie Lane	12	3	2	2	0	2
3	Takeshi Kawamura	0	4	0	0	0	1
4	Yu-Ju Chen	26 in progress	86	1	0	1	0
5							
6							
7	Ed Nice	0		0			1
8							
9							
10	Jin Park	0	1	0	0	0	0
11	Jong Shin	2 in progress	2	1	0	1	0
12	Ravi Sirdeshmukh	0	0	0	0	0	0
13	Young-Ki Paik	0	5	2	1	1	0
14	Charles Pineau	12	3				
15	Gilberto Domont	0	2	1	0	1	0
16	Fernando Corrales	2 (+ 9)	165	2	0	2	5
17	Gil Omenn	0	0	3	1	2	8
18	Andrey Archakov	0	0	1	0	1	3
19							
20	Siqi Liu			1	0	1	
21	Albert Sickmann	0		0	0	0	0
22							
X	Tadashi Yamamoto			1	0	1	0
Y	Hosseini Salekdeh	1		2	1	1	0
Mt							
TOTAL		15	268	19	6	13	20
	(in progress)	37					

Chromosome 1

(Ping Xu)

PIC Leaders: Ping Xu, Fuchu He

Contributing labs:

Ping Xu, Beijing Proteome Research Center

Fuchu He, Beijing Proteome Research Center

Dong Yang, Beijing Proteome Research Center

Wantao Ying, Beijing Proteome Research Center

Pengyuan Yang, Fudan University

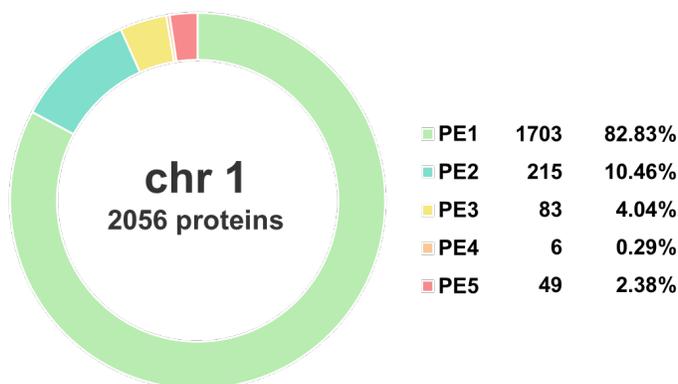
Siqi Liu, Beijing Genome Institute

Qinyu He, Jinan University

Major lab members or partners contributing to the neXt50:

Yao Zhang (Beijing Proteome Research Center), Yihao Wang (Beijing Proteome Research Center), Cuitong He (Beijing Proteome Research Center), Wei Wei (Beijing Proteome Research Center), Yanchang Li (Beijing Proteome Research Center), Feng Xu (Beijing Proteome Research Center), Xuehui Peng (Beijing Proteome Research Center).

Status of the Chromosome “parts list”:



Step by step milestone plan to find, identify and validate MPs

1. We try to identify more missing proteins through high coverage proteomics technology with testis samples. We will finish this project before May.
2. We tested and confirmed that PTM approach could identify significant number of missing proteins. We will finalize the data sets before May.

HUPO 2017 UPDATE

A) Titles and authors of papers submitted to the JPR SI

1. Identification of Missing Proteins in the Phosphoproteome of Kidney Cancer

Xuehui Peng, Feng Xu, Shu Liu, Suzhen Li, Qingbo Huang, Lei Chang, Lei Wang, Xin Ma, Fuchu He, Ping Xu

2. Multi-protease digesting strategy identifies more membrane-associated missing proteins in human testis tissue

Yihao Wang, Yang Chen, Yao Zhang, Wei Wei, Yanchang Li, Yue Gao, Ping Xu

B) Accepted or in revision

Accepted:

1. **Identification of Missing Proteins in the Phosphoproteome of Kidney Cancer.** Xuehui Peng, Feng Xu, Shu Liu, Suzhen Li, Qingbo Huang, Lei Chang, Lei Wang, Xin Ma, Fuchu He, Ping Xu

In revision:

2. **Multi-protease digesting strategy identifies more membrane-associated missing proteins in human testis tissue.** Yihao Wang, Yang Chen, Yao Zhang, Wei Wei, Yanchang Li, Yue Gao, Ping Xu

C) Titles and authors of other HPP relevant papers submitted elsewhere

None.

D) Accepted or in progress

E) How many PE1-found MPs since HUPO-2016 has your chromosome group reported?

F) How many candidate MPs found, but not meeting the guidelines 2.1? (Please state if one or two peptides identified and their length, biological replicates found in, and spectral quality).

G) Any significant change in your Ch Team plans for neXt-50? If so please update the word document form last time and forward. Ill compile this with the other changes as a new summary for the EC.

None.

H) any significant clinical or other successes re a MP that you wish us to consider highlighting in the report.

None.

Chromosome 2

(Lydie Lane)

See combined response and work plan with Chromosome 14.

HUPO 2017 UPDATE

A/B) Accepted papers submitted to the JPR SI

Accepted: Carapito C, **Duek P**, Macron C, Seffals M, Rondel K, Delalande F, Lindskog C, Freour T, Vandembrouck Y, **Lane L**, Pineau C.

Validating missing proteins in human sperm cells by targeted mass spectrometry- and antibody-based methods.

Omenn GS, **Lane L**, Lundberg EK, Overall CM, Deutsch EW.

Progress on the HUPO Draft Human Proteome: 2017 Metrics of the Human Proteome Project.

C/D) Accepted HPP relevant papers submitted elsewhere

Schaeffer M, Gateau A, Teixeira D, Michel PA, Zahn-Zabal M, Lane L. *The neXtProt peptide uniqueness checker: a tool for the proteomics community.* Bioinformatics. 2017 May 17. doi: 10.1093/bioinformatics/btx318.

Gaudet P, Michel PA, Zahn-Zabal M, Britan A, Cusin I, Domagalski M, Duek PD, Gateau A, Gleizes A, Hinard V,

Rech de Laval V, Lin J, Nikitin F, Schaeffer M, Teixeira D, Lane L, Bairoch A. *The neXtProt knowledgebase on human proteins: 2017 update*. Nucleic Acids Res. 2017 Jan 4;45(D1):D177-D182. doi: 10.1093/nar/gkw1062.

E) How many PE1-found MPs since HUPO-2016 has your chromosome group reported?

12

F) How many candidate MPs found, but not meeting the guidelines 2.1? (Please state if one or two peptides identified and their length, biological replicates found in, and spectral quality).

3

G) Any significant change in your Ch Team plans for neXt-50? If so please update the word document form last time and forward. Ill compile this with the other changes as a new summary for the EC.

None.

H) any significant clinical or other successes re a MP that you wish us to consider highlighting in the report.

None.

Chromosome 3

(Takeshi Kawamura)

PIC Leaders: Takeshi Kawamura (since January 1)

Contributing labs: Major lab members or partners contributing to the neXt50:

Takeshi Kawamura: Associate Professor, Proteomics Laboratory, Isotope Science Center, The University of Tokyo, Tokyo, Japan.

Toshihide Nishimura: Professor, Department of Translational Medicine Informatics, St. Marianna University School of Medicine, Kanagawa, Japan.

Hiromasa Tojo: Associate Professor, Department of Biophysics and Biochemistry, Osaka University; Graduate School of Medicine, Osaka, Japan.

Tadashi Imanishi: National Institute of Advanced industrial Science and Technology (AIST), Tokyo, Japan.

Thomas P. Conrads: Adjunct Associate Professor, Chief Scientific Officer, Gynecologic Cancer Center of Excellence Women's Health Integrated Research Center at Inova Health System, USA.

Siu Kwan Sze: Assistant Professor, Director of Proteomics Core of Bioscience Research Centre, Singapore.

Status of the Chromosome "parts list":

In preparation

Confirmation that PIC and C-HPP lab members have read:

Yes

i) Deutsch et al 2016: Human Proteome Project Mass Spectrometry Data Interpretation Guidelines 2.1.

ii) Duek et al 2016: Missing Protein Landscape of Human Chromosomes 2 and 14: Progress and Current Status.

iii) Omenn et al 2016: Metrics for the Human Proteome Project 2016: Progress on Identifying and Characterizing the Human Proteome, Including Post-Translational Modifications.

iv) Vandenbrouck et al 2016: Looking for Missing Proteins in the Proteome of Human Spermatozoa: An Update.

Step by step milestone plan to find, identify and validate MPs:

Milestone dates:

PIC is changing from Prof. Nishimura to Kawamura and the project is under taking over. Therefore, we cannot

set milestones at the present time. Members will gather at the February and plan to consider. I think that milestones can be decided before the JPR special issue in May.

HUPO 2017 UPDATE

A) Titles and authors of papers submitted to the JPR SI

None.

B) Accepted or in revision

Accepted: None.

In revision: None.

C/D) Titles and authors of other HPP relevant papers submitted elsewhere

A simple method for selecting a representative protein with better protein evidence from a set of non-unique proteins in outputs of protein inference tools. *In revision*

E) How many PE1-found MPs since HUPO-2016 has your chromosome group reported?

None.

F) How many candidate MPs found, but not meeting the guidelines 2.1? (Please state if one or two peptides identified and their length, biological replicates found in, and spectral quality).

4.

candidates	Peptides identified	Length	Biological replicates	PE	Chr	Spectral quality
protein1	1	8	1	4	17	good, similar to a synthetic peptide
protein2	1	21	1	5	1	good, similar to a synthetic peptide
protein3	1	10	2	4	20	chimera
protein4	1	13	1	5	9	chimera

G) Any significant change in your Ch Team plans for neXt-50? If so please update the word document form last time and forward. Ill compile this with the other changes as a new summary for the EC.

There isn't significant plan change, but a new project is under planning.

H) any significant clinical or other successes re a MP that you wish us to consider highlighting in the report.

None.

Chromosome 4

(Yu-Ju Chen)

Leader	Yu-Ju Chen
Contributing labs	1. 2. 3. 4. Institute of Chemistry, Academia Sinica PI: Yu-Ju Chen Members: Reta Birhanu Kitata, Mehari Muuz, C Institute of Information Science, Academia Sinica PI: Ting-Yi Sung PI: Wen-Lian Hsu Members: Wai-K Chen Master Program for Clinical Pharmacogenomics and Pharmacoproteomics, Taipei Medical University PI: Chia-Li Han Department of Biochemistry and Molecular Biology, Chang Gung University PI: Jau-Song Yu PI: Chia-J
Address	Institute of Chemistry, Academia Sinica 128 Academia Road Sec. 2, Nankang Taipei 115 Taiwan, ROC
Contact E-mail	yujuchen@gate.sinica.edu.tw
Telephone and Fax (with country code)	Phone: +886-2-27898660; Fax: +886-2-27831237

1. **Status of the Chromosome “parts list”**
2. **Milestone plan to find, identify and validate MPs**

2-1. Investigating the possibility of identifying 82 missing proteins via informatics approach

The unknown abundance of the 82 missing proteins in the biological sample present a great challenge for identifying these missing proteins. To aid experimental design of using proper material, we will develop an informatics ranking system for different cell/tissue types based on their gene expression data and mass spectrometry- based post-translational modification database. We will also perform prediction on subcellular localization and biochemical properties for each missing protein. Finally, the detectivity and properties of the enzymatic peptides of each missing proteins, such as peptide uniqueness, peptide length, and genome-based variants frequency (i.e., the amino acid length per sequence variant, LSV) will be investigated.

Expected milestone: complete analysis in 2017

2-2. Optimized Multi-protease Shotgun Strategy for Identifying 82 MPs

Chromosome 4 contains 761 protein entries of which 659 have protein level evidence while 23 at transcript level, 2 evidence from homologous species and 20 uncertain. At the updated version, there are 82 missing proteins (PE2-PE4) awaiting protein-level identification and validation (neXtProt 2017-01-23).

Based on the above informatics analysis, we will develop and apply an optimized shotgun strategy that integrates multiple enzymatic digestions to increase protein and peptide coverage, peptide fractionation to maximize identified protein number, LC- MS/MS analysis and protein identification by multiple search engines for comprehensive identification of missing proteins. Different sample types (organs/tissue/cells) will be analyzed for facilitating identification of tissue-specific missing proteins.

Expected milestone: complete analysis in 2018

2-3. Cancer Moonshot as a Resource toward Comprehensive Human Proteome Library

To meet urgent need of finding new solution for cancer detection and therapy, Academia Sinica, Chang Gang University have signed Memoranda of Understanding (MOUs) for international collaboration under the *Cancer Moonshot* project in 2016. Specifically, the global efforts will jointly develop state-of-art proteogenomic technology for large-scale analysis of cancer patients and make available an unprecedented international dataset to advance the cancer research and care. Under the global efforts, we expect that the in-depth tissue proteomic and phosphoproteomic dataset for different cancer types will provide a rich resource for searching missing proteins.

Expected milestone: 100 pairs of cancerous and adjacent normal tissue from lung cancer will be expected to be finished in 2017.

2-4. MRM-MS for Validation of Missing Proteins

Multiple reaction monitoring (MRM) mass spectrometry using synthetic heavy isotope labeled peptide will be performed to validate missing proteins identified by the discovery method. The multiprotease approach is expected to provide additional unique peptides for those proteins with few or none detectable tryptic peptides. Informatics analysis of similarity score of spectral feature will be designed along with rigorous analytical assay evaluation to reduce the false positive result during validation.

HUPO 2017 UPDATE

A) Titles and authors of papers submitted to the JPR SI

Wai-Kok Choong; Tung-Shing Lih; Yu-Ju Chen; Ting-Yi Sung. Decoding the Effect of Isobaric Substitutions on Identifying Missing Proteins and Variant Peptides in Human Proteome.

B) Accepted or in revision

Accepted:

In revision: Wai-Kok Choong; Tung-Shing Lih; Yu-Ju Chen; Ting-Yi Sung. Decoding the Effect of Isobaric Substitutions on Identifying Missing Proteins and Variant Peptides in Human Proteome.

C) Titles and authors of other HPP relevant papers submitted elsewhere

None.

D) Accepted or in progress

None.

E) How many PE1-found MPs since HUPO-2016 has your chromosome group reported?

For the report of missing proteins (using gold standard), we are still working on the functional evaluation. The following information (not published yet) can be for your information. 26 proteins meet C-HPP criteria (guidelines 2.1), 2 of them are in ch 4.

F) How many candidate MPs found, but not meeting the guidelines 2.1? (Please state if one or two peptides identified and their length, biological replicates found in, and spectral quality).

By previous criteria using identification with single unambiguous peptide, we obtained a total of 86 potential candidates of MPs and 26 among them have two unique peptides. Validation of biological replicate experiments are still ongoing.

G) Any significant change in your Ch Team plans for neXt-50? If so please update the word document form last time and forward. Ill compile this with the other changes as a new summary for the EC.

None.

H) any significant clinical or other successes re a MP that you wish us to consider highlighting in the report.

None.

Chromosome 5

(Peter Horatovich)

PIC Leaders: Peter Horvatovich and Rainer Bischoff

Contributing labs (Lab Heads named with affiliation University/Institute/Company):

- Rainer Bischoff and Peter Horvatovich, University of Groningen, Department of Pharmacy, Analytical Biochemistry
- Gyorgy Marko-Varga and Peter Horvatovich, University of Lund, The Center of Excellence in Biological and Medical Mass Spectrometry

Status of the Chromosome “parts list”:

According to NextProt (release: 2016-12-02):

PE1	PE2	PE3	PE4	PE5	Total
738	99	18	5	10	870

Confirmation that PIC and C-HPP lab members have read:

Peter Horvatovich read all papers and lab members will do it until end of January 2017.

Step-by-step milestone plan to find, identify and validate MPs:

- Participation in IVTT cluster, which target the 50 missing proteins in chromosomes 5, 10, 15, 16, and 19 that has the highest probability for identification in COV318, PANC1, DFC1024, D341MED, ST486, KLE cell line determined using mRNA array expression.
- Close collaboration with chromosome 19 team in Moonshot project performing deep profiling of 4000 melanoma samples and sharing bioinformatics expertise. In this project missing proteins can be obtained from deep profiling of large number of primary and metastatic melanoma samples.

- European transcan-2 project on ovarian cancer will allow to obtain primary and metastatic protein profiles on large number of ovarian cancer samples, where missing proteins is expected to be identified.
- Local project on proteogenomics profiling of human lung tissue in context of COPD. Human lung tissue is highly complex in term of cell types, which can be the source of multiple missing proteins.

Milestone dates (corresponding to the project listed above):

- We are currently analyzing COV318 cell line using MRM approach. We expect to analyze additional cell lines and human tissue (human lung, tissues from oral cavities) and report it until next HUPO in Ireland.
- This project is just started, but we expect to have the first results in the end of 2017.
- This project is just started, but we expect to have the first results in the end of 2017.
- This project already provided protein profile and is currently under bioinformatics evaluation. Identification of missing proteins is expected to be performed within three months.

HUPO 2017 UPDATE

A) Titles and authors of papers submitted to the JPR SI

B) Accepted or in revision

Accepted:

In revision:

C) Titles and authors of other HPP relevant papers submitted elsewhere

D) Accepted or in progress

E) How many PE1-found MPs since HUPO-2016 has your chromosome group reported?

F) How many candidate MPs found, but not meeting the guidelines 2.1? (Please state if one or two peptides identified and their length, biological replicates found in, and spectral quality).

G) Any significant change in your Ch Team plans for neXt-50? If so please update the word document form last time and forward. Ill compile this with the other changes as a new summary for the EC.

H) Any significant clinical or other successes re a MP that you wish us to consider highlighting in the report.

Chromosome 6

(Christoph Borchers)

*PIC Leaders: *Christoph Borchers is new PIC as of January 20.*

Co-PI:

Yassene Mohammed, Leiden University, Leiden, Netherlands, University of Victoria, Victoria, Canada

Team:

Leonard Forster, Christopher M. Overall, Paul Keown, Rob MacMaster, Michael Siu, Juergen Kast

Research Plan:

Mining for human missing proteins in dissected mouse tissues

Mice are the most commonly used human model organism for biology and medicine. In addition to accessibility, around 90% of their genes are closely related to those in humans, making them a perfect tool for studying biology and diseases. In one project our goal is to develop mouse-specific MRM assays for 20 different tissues.

Besides using these assays to answer various biological and disease-related questions, these assays also allows us to mine the proteome for missing human proteins. Especially because it is ethically inappropriate to collect and use human tissues for discovery analyses, using mouse tissue to detect homologs of the missing human proteins is a very promising approach.

In the first discovery phase, we performed analyses using high-resolution mass spectrometry on 40 different mouse tissues, with the goal of developing MRM assays using peptides that are shared between mouse and human. We have already identified 23 mouse proteins in various tissues corresponding to missing human proteins from chromosome 6 using this high-resolution data. The majority of the proteins were found in only 2 out of the 40 different tissues (Table 1). Expression of these missing chromosome 6 proteins seems to be concentrated in bone tissue and marrow, as well as in ovaries and the nervous system.

Chris Overall is therefore initiated high mass accuracy shotgun and TAILS proteomics analyses (for N and C termini) of human bone and bone marrow (Q2-4, 2017) and ovary/fallopian tube/placenta (Q3-4 2017), and monocytes/macrophages/B cells for several other MPs (Q4 2017).

While this high-resolution data provides the first indication of the presence of these proteins, it is our goal to verify these results with targeted MRM-based proteomics, using synthetic heavy labeled peptides which is the "gold standard" for the quantitation. We will also develop MRM assays for any predicted chromosome-6 proteins which could not be detected by high resolution MS, because these targeted assays are more sensitive than shotgun approaches, particularly if 2D LC-MRM/MS is use. For both groups of proteins, the MRM approach will provide the best information on tissue expression levels and in which human tissue the subsequent verification should be performed. Additionally, while our current results are related to Chromosome 6, this method is applicable to all missing human proteins as long as there is a mouse homolog.

Table 1. Some of human chromosome 6 missing proteins and their corresponding mouse proteins expressed in various mouse tissues.

Human chr6 missing proteins	Gene	Corresponding mouse proteins	Number of mouse tissues with protein expression	Mouse tissues with protein expression
Q8N7X0	ADGB	G3UZ78	15	bone marrow, brain, cecum, colon, epididymis, femur, gallbladder, liver c-r, lung, ovaries, seminal vesicles, small intestine, spinal cord, stomach, thymus
Q8IZF7	ADGRF2	E9Q4J9	7	femur, gallbladder, mammary gland, ovaries, plasma, salivary glands, sciatic nerve
Q96KW2	POM121L2	Q5SW25	5	kidneys, mln, sciatic nerve, skin, spinal cord
Q7Z6R9	TFAP2D	Q91ZK0	5	bone marrow, small intestine, spinal cord, stomach, thymus
Q5SW24	DACT2	Q7TN08	4	bone marrow, plasma, spinal cord, tongue
Q8IXS0	FAM217A	Q9D9W6	4	cecum, eye left, sciatic nerve, small intestine
Q8TC41	RNF217	D3YYI7	4	femur, eye lens right, seminal vesicles, urinary bladder
Q14916	SLC17A1	Q61983	4	bone marrow, spleen, wbc, white fat
Q8NFK1	GJC3	Q921C1	3	heart, tongue, urinary bladder
Q9BZJ6	GPR63	Q9EQQ3	3	liver c-r, rbc, white fat
Q9NQE7	PRSS16	Q9QXE5	3	cecum, thymus, trachea
Q9BWV7	TLL2	A4Q9E4	3	brain, epididymis, testes
Q3MIW9	DPCR1	Q3TNW5	2	femur, sciatic nerve
P24046	GABRR1	P56475	2	skin, wbc

Q5JXM2	METTL24	Q8CCB5	2	Adrenals glands, gallbladder
Q6ZW05	PTCHD4	B9EKX1	2	eye left, salivary glands
Q5SSQ6	SAPCD1	Q9CY86	2	femur, salivary glands
Q8IW70	TMEM151B	Q68FE7	2	epididymis, trachea
Q5TFG8	ZC2HC1B	Q9D534	2	ovaries, thymus
Q8NEN0	ARMC2	Q3URY6	1	mln
Q8N4B5	PRR18	Q6PAN7	1	Ovaries
Q5TGI4	SAMD5	Q3V1H9	1	White fat
Q9Y2C5	SLC17A4	Q5NCM1	1	wbc

HUPO 2017 UPDATE

A) Titles and authors of papers submitted to the JPR SI

B) Accepted or in revision

Accepted:

In revision:

C) Titles and authors of other HPP relevant papers submitted elsewhere

D) Accepted or in progress

E) How many PE1-found MPs since HUPO-2016 has your chromosome group reported?

F) How many candidate MPs found, but not meeting the guidelines 2.1? (Please state if one or two peptides identified and their length, biological replicates found in, and spectral quality).

G) Any significant change in your Ch Team plans for neXt-50? If so please update the word document form last time and forward. Ill compile this with the other changes as a new summary for the EC.

H) any significant clinical or other successes re a MP that you wish us to consider highlighting in the report.

Chromosome 7

(Ed Nice)

The Chr 7 NeXT 50 challenge will continue along the lines presented for the Missing Proteins at the recent C-HPP workshops. The Top 50 proteins to be pursued will be identified from the current NeXt Prot release and will be driven around perceived feasibility. A Biology/Disease driven approach will be used to populate the project. This could result in missing proteins from other chromosomes also being identified.

In brief the following steps will be used:

1. Check the Chr 7 missing proteins against MissingProteinPedia
2. Identify proteins/protein families where credible non-MS based evidence exists for their existence (we are already aware this is the case for more than 50 proteins including a number of olfactory receptors)
3. Identify "champions" for these proteins/protein families who can be recruited into HUPO if not already members. This will include biochemists, physicians, pathologists.

4. Develop SOPs and build up a Biobank meeting best practice for available tissues or cell lines.

5. Use state of the art MS technologies for deep mining of the associated proteomes.

It is anticipated that steps 1-3 will be complete for presentation at the Dublin meeting.

HUPO 2017 UPDATE

A) Titles and authors of papers submitted to the JPR SI.

B) Accepted or in revision

Accepted: None.

In revision: None.

C) Titles and authors of other HPP relevant papers submitted elsewhere

Baker MS1, Ahn SB, Mohamedali A, Islam MT, Cantor D, Verhaert PD, Fanayan S, Sharma S, Nice EC, Connor M, Ranganathan S. Accelerating the search for the missing proteins in the human proteome. Nat Commun. 2017 8:14271. doi: 10.1038/ncomms14271

D) Accepted or in progress

None

E) How many PE1-found MPs since HUPO-2016 has your chromosome group reported?

None, but other groups have used MPP and reported prestin.

F) How many candidate MPs found, but not meeting the guidelines 2.1? (Please state if one or two peptides identified and their length, biological replicates found in, and spectral quality).

None.

G) Any significant change in your Ch Team plans for neXt-50? If so please update the word document form last time and forward. Ill compile this with the other changes as a new summary for the EC.

Strategy remains the same: olfactory receptors using MPP.

H) Any significant clinical or other successes re a MP that you wish us to consider highlighting in the report.

None.

Chromosome 8

(Pengyuan Yang)

***PIC Leaders:** Pengyuan Yang

***Contributing labs (Lab Heads named with affiliation):**

Pengyuan Yang, Fudan University

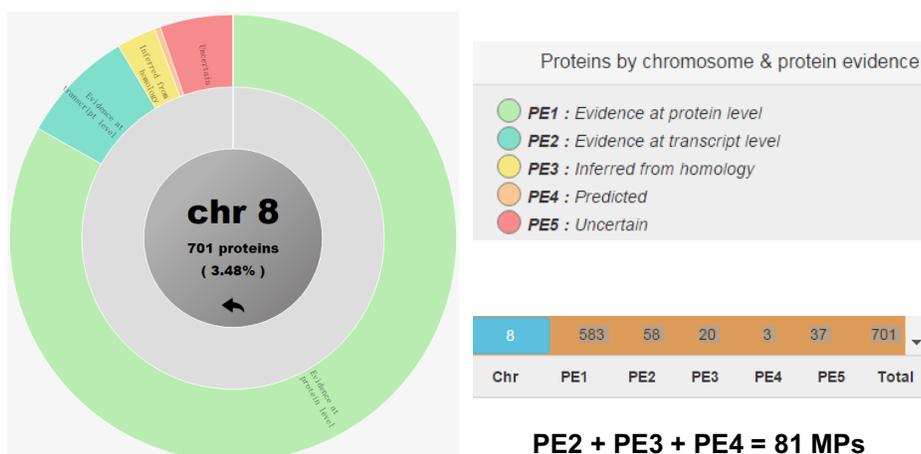
Fan Zhong, Fudan University

Fei Lan, Fudan University

Major lab members or partners contributing to the neXt50:

Yang Liu (Fudan University), Zhennan Shi (Fudan University), Yang Zhang (Fudan University), Huali Shen (Fudan University), Xiaohui Liu (Fudan University), Guoquan Yan (Fudan University), Lei Zhang (Fudan University), Weimin Zheng (Fudan University), Zhaoyu Qin (Fudan University)

***Status of the Chromosome “parts list”:**



***Confirmation that PIC and C-HPP lab members have read:**

OK, we have checked.

***Step by step milestone plan to find, identify and validate MPs:**

We would like to inspect all the MPs (not limited to Chr.8) according to the latest NeXtProt release, to find if there is any epigenetic obstacle of hindering gene expression. ENCODE and other epigenetic profiling data will be used. After that, we are planning to remove these obstacles of Chr. 8 MPs by first epigenetic and then gene-editing approaches in cell lines. We also plan to conquer the last hardcore of Chr.8 MPs by transfecting lentiviral vectors.

***Milestone dates:**

- We plan to figure out the activation candidates before May.
- We expect to get some epigenetic and then gene-editing results by the end of 2017.
- We hope to start to explore the last hardcore in 2018.

HUPO 2017 UPDATE

A) Titles and authors of papers submitted to the JPR SI

B) Accepted or in revision

Accepted:

In revision:

C) Titles and authors of other HPP relevant papers submitted elsewhere

D) Accepted or in progress

E) How many PE1-found MPs since HUPO-2016 has your chromosome group reported?

F) How many candidate MPs found, but not meeting the guidelines 2.1? (Please state if one or two peptides identified and their length, biological replicates found in, and spectral quality).

G) Any significant change in your Ch Team plans for neXt-50? If so please update the word document form last time and forward. Ill compile this with the other changes as a new summary for the EC.

H) Any significant clinical or other successes re a MP that you wish us to consider highlighting in the report.

Chromosome 9

(Je-Yoel Cho)

PIC leaders: Je-Yoel Cho^{1*}, Soo-Yeon Lee², Jin-Hwan Lee³

Contributing Members: Yong-in Kim¹, Hyung-Min Park¹, Roa Yoon¹, Ji-Sook Park²,

1. Department of Biochemistry, BK21 PLUS Program for Creative Veterinary Science Research and Research Institute for Veterinary Science, College of Veterinary Medicine, Seoul National University, South Korea

2. Samsung Medical Center, Seoul, South Korea

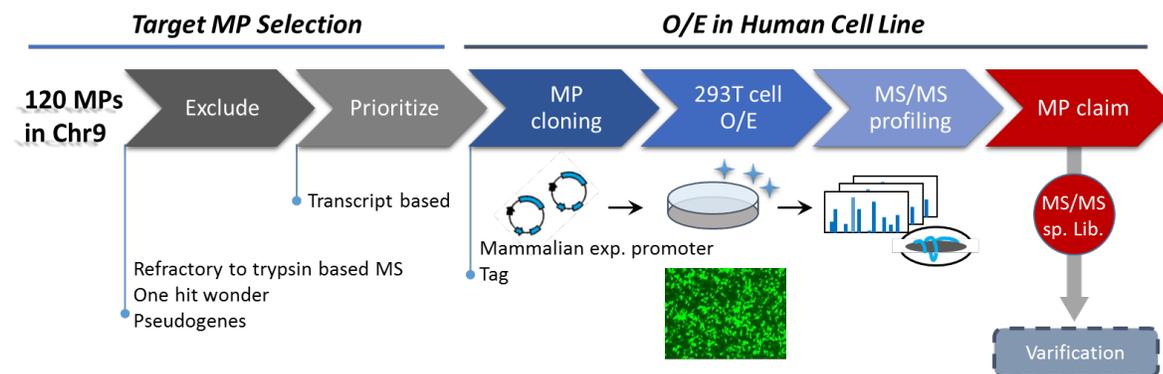
3. Korea Research Institute of Standards and Science, KRISS, Daejeon, South Korea

Milestone Dates (Expected):

Step 1: March 2017

Step 2: March 2018

Step 3: November 2018



STEP 1. TARGET MP SELECTION

Targeting to missing proteins (MPs), which are theoretically easy to detect.

1. Refractory to trypsin based MS/MS → exclude
2. One hit to wonder proteins → exclude
3. Transcript expression database (e.g. HPA, GTEx, BGee...) → prioritize

*Chr9 status: 120 proteins are remained missing (neXtProt rel. 2016-12-02)

4. Select top 60 lists

STEP 2. OVER-EXPRESSION of TOP 50 MP IN HUMAN CELL LINE

1. CDS of top 60 missing proteins insert to plasmid vector that contain mammalian cell expressed promoter and affinity/fluorescence tag.
2. Transfection to 293T cell.
3. If available, utilize tag for transfection check, subcellular localization observation and affinity-based MP enrich.
4. MS/MS profiling

5. For top 50, MP claim and MS/MS spectra library construction for further verification step.

STEP 3. VERIFICATION: NATURALLY EXPRESSED MPs IN HUMAN CELL/TISSUE

1. Select top 50 target proteins from Step 2 that are show clear MS spectrum and expressed in target tissues/cells samples such as testis using transcript expression database (e.g. HPA, GTEx, BGee etc).
2. Targeted analysis by MRM or PSM (MS inclusion list) with synthetic peptides spikes.

HUPO 2017 UPDATE

A) Titles and authors of papers submitted to the JPR SI

B) Accepted or in revision

Accepted:

In revision:

C) Titles and authors of other HPP relevant papers submitted elsewhere

D) Accepted or in progress

E) How many PE1-found MPs since HUPO-2016 has your chromosome group reported?

F) How many candidate MPs found, but not meeting the guidelines 2.1? (Please state if one or two peptides identified and their length, biological replicates found in, and spectral quality).

G) Any significant change in your Ch Team plans for neXt-50? If so please update the word document form last time and forward. Ill compile this with the other changes as a new summary for the EC.

H) Any significant clinical or other successes re a MP that you wish us to consider highlighting in the report.

Chromosome 10

(Jin Park)

Updating by Feb 7 after grants submitted.

Reminder sent 19 Feb. Report promised soon.

Several months ago my lab sent out the lysates from several selected cell lines – chosen based on mRNA expression of the missing proteins – to the members of our mini-consortium. They have probed these lysates looking for evidence of the missing proteins using the optimized spectra that they determined were ideal for these individual proteins – learned by producing them in cell free lysates and testing them on MS.

Reminder sent March 16.

HUPO 2017 UPDATE

A) Titles and authors of papers submitted to the JPR SI

None.

B) Accepted or in revision

Accepted: None.

In revision: None.

C) Titles and authors of other HPP relevant papers submitted elsewhere

None.

D) Accepted or in progress

None.

E) How many PE1-found MPs since HUPO-2016 has your chromosome group reported?

0.

F) How many candidate MPs found, but not meeting the guidelines 2.1? (Please state if one or two peptides identified and their length, biological replicates found in, and spectral quality).

1 for Q8WW14 or C10orf82: KPLQEPPIPGWAGYLPR and DFLEITER were found in sperm lysate. This is done in collaboration with Chr16 team.

G) Any significant change in your Ch Team plans for neXt-50? If so please update the word document form last time and forward. Ill compile this with the other changes as a new summary for the EC.

None.

H) Any significant clinical or other successes re a MP that you wish us to consider highlighting in the report.

We continue to add more full-length ORF expression clones for missing protein identification by IVTT-MS method, and have recently synthesized ~1,000 more ORF clones. As of Sep 1, 2017, we have 35 clones for Chr10 PE2 missing proteins (according to the most recent NextProt data), and 27 more clones are under construction. We are also providing the IVTT clones to the consortium of Chr 5, 15, 16, and 19 teams.

Chromosome 11

(Jong Shin)

PIC Leaders: Jong Shin Yoo, Ph.D.

Contributing labs: Jong Shin Yoo/Biomedical Omics Research Group/Korea Basic Science Institute

Major lab members or partners contributing to the neXt50: Jin Young Kim/KBSI, Bong Hee Lee/Gacheon Univ.

Status of the Chromosome “parts list”:

PE2- 213,

PE3- 98,

PE4- 6,

Total 317 are missing proteins in chromosome 11.

(from <https://www.nextprot.org/about/protein-existence>)

Confirmation that PIC and C-HPP lab members have read: *OK, we have checked.*

Step by step milestone plan to find, identify and validate MPs:

Step. 1 20 missing proteins in all Chr. mapped

Step. 2 20 more missing proteins in all Chr. mapped

Step. 3 10 more missing proteins in all Chr. mapped

Total missing proteins: 50 in all Chr. Mapped

Milestone dates:

Step. 1 Apr. 2017

Step. 2 Dec. 2017

Step. 3 Jun. 2018

HUPO 2017 UPDATE

A) Titles and authors of papers submitted to the JPR SI

Next Generation Proteomic Pipeline for Chromosome-based Proteomic Research Using GENCODE and neXtProt Databases

B) Accepted or in revision

Accepted: None.

In revision: Next Generation Proteomic Pipeline for Chromosome-based Proteomic Research Using GENCODE and neXtProt Databases

C) Titles and authors of other HPP relevant papers submitted elsewhere

None.

D) Accepted or in progress

None.

E) How many PE1-found MPs since HUPO-2016 has your chromosome group reported?

In 2016 Park et al. JPR SI 15:4082-4090, total 10 missing proteins, Q16478(Chr#19), P22794(Chr#17), Q5JVV73(Chr#X), Q7Z407(Chr#8), O43300(Chr#5), Q9ULB4(Chr#9), P0CW23(Chr#18), Q8IYE0(Chr#7), C4AMC7(Chr#15), and Q9Y5F6(Chr#5), were reported.

F) How many candidate MPs found, but not meeting the guidelines 2.1? (Please state if one or two peptides identified and their length, biological replicates found in, and spectral quality).

In 2017 Hwang et al. JPR SI, in progress, two missing proteins (Q7LC44 and Q16478) having two or more distinct peptides were identified.

As a silver missing protein, we found a candidate (Q6UXN7) with one distinct peptide (LQELFLQEVN) and one short (8 mer) peptide (MGELWLSR), but without biological replication.

G) Any significant change in your Ch Team plans for neXt-50? If so please update the word document from last time and forward. Ill compile this with the other changes as a new summary for the EC.

We have tried to find the missing proteins from Chr 11, but no many because there are many missing proteins with tansmembrane domains in Chr 11.

H) Any significant clinical or other successes re a MP that you wish us to consider highlighting in the report.

None.

Chromosome 12

(Ravi Sirdeshmukh)

No response.

Reminder sent February 19, March 16.

Terence Poon and two of the new Indian groups have MS platform ready for targeted validations and have agreed to contribute on this front.

HUPO 2017 UPDATE

A) Titles and authors of papers submitted to the JPR SI

B) Accepted or in revision

Accepted:

In revision:

C) Titles and authors of other HPP relevant papers submitted elsewhere

D) Accepted or in progress

E) How many PE1-found MPs since HUPO-2016 has your chromosome group reported?

F) How many candidate MPs found, but not meeting the guidelines 2.1? (Please state if one or two peptides identified and their length, biological replicates found in, and spectral quality).

G) Any significant change in your Ch Team plans for neXt-50? If so please update the word document form last time and forward. Ill compile this with the other changes as a new summary for the EC.

The challenge is everyone is preoccupied with other responsibilities and we still do not have funding to keep the working targets and time lines. The result is significant progress is still much awaited. Thus, we are NOT pursuing MPs in a mission-mode although we keep looking for them in other lines of work.

We decided to focus on Transcriptomics and Proteomics data integration.

i. Proteogenomics and splice variants - identification, biology, and clinical aspect. Experimental targeted validations. This effort we may be going beyond Chr 12.

In an initial effort, we have identified a novel differentially expressed splice variant of NCAM (Chr 11) from the GBM data (not present in neXtProt). The identification part has been published in a recently published book chapter (Springer) and the detailed ms is ready to go except for one small exptl result.

ii. In continuation to the 2 Dimensional molecular maps we published in the earlier JPR SI, linking differentially expressed miRNA and mRNA Protein target regulatory cascades in various functional contexts (Two dimensions being regulatory and functional), we have some progress. In one effort, we have integrated membrane and nuclear protein data from GBM to construct integrated ppi network for the transcription and invasion process. Interestingly the network revealed two transcription factors interacting with key hub entities in the invasion process. We believe such an exercise would yield important regulatory factors that may have clinical significance (B/D perspective). This ms, which carries even other observations, is also in the publication pipeline.

H) Any significant clinical or other successes re a MP that you wish us to consider highlighting in the report.

Chromosome 13

(Young Ki Paik)

PIC Leaders: Young-Ki Paik, Ph.D.

Young-Ki Paik/Yonsei Proteome Research Center/Yonsei University

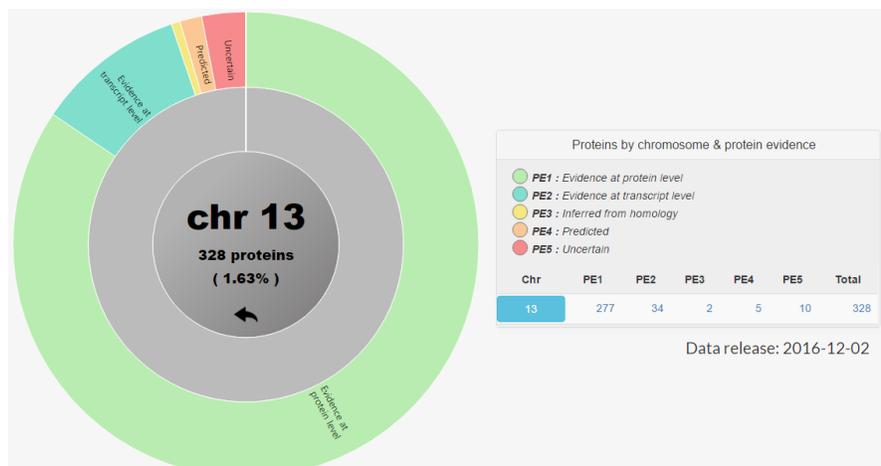
Major lab members or partners contributing to the neXt50:

Seul-Ki Jeong/YPRC/Yonsei University.

Jin-Young Cho/YPRC/Yonsei University.

Chae-Yeon Kim/YPRC/ Integrated Omics for Biomedical Sciences, Yonsei University.

Status of the Chromosome “parts list”:



(from <https://www.nextprot.org/about/protein-existence>)

Confirmation that PIC and C-HPP lab members have read:

OK, we have checked.

Step by step milestone plan to find, identify and validate MPs:

Top-Down Approaches

- Step. 1 20~30 missing proteins from liver (normal-like, tumor, and/or cell lines) in all Chr. mapped
- Step. 2 20~30 missing proteins from pancreas (normal-like, tumor, and/or cell lines) in all Chr. mapped
- Step. 3 10~20 missing proteins from placenta (normal, pre-eclampcia) in all Chr. Mapped

Total missing proteins: 50 in all Chr. mapped

Milestone dates:

- Step. 1 Jul. 2017
- Step. 2 Feb. 2018
- Step. 3 Aug. 2018

Step by step milestone plan to find, identify and validate MPs:

Bottom-Up Approaches

- Step 1 Select target missing proteins under criteria (available/target sample, transcript expression level, number of proteotypic peptides, etc.)
- Step 2 Makes synthetic peptides and MS spectra of target proteins.
- Step 3 Discovering and validating missing proteins.

Milestone dates:

- Step. 1 Feb. 2017
- Step. 2 Feb. 2018
- Step. 3 Aug. 2018

HUPO 2017 UPDATE

A) Titles and authors of papers submitted to the JPR SI

(1) A systematic proteogenomic approach to exploring a novel function for NHERF1 in human reproductive disorder: lessons for exploring missing proteins.

(2) Epsilon-Q: An Automated Analyzer Interface for Mass Spectral Library Search and Label-Free Protein Quantification (in press)

B) Accepted or in revision

Accepted: Epsilon-Q: An Automated Analyzer Interface for Mass Spectral Library Search and Label-Free Protein Quantification (in press).

In revision: A systematic proteogenomic approach to exploring a novel function for NHERF1 in human reproductive disorder: lessons for exploring missing proteins (2nd Revision being submitted by 9/10).

C) Titles and authors of other HPP relevant papers submitted elsewhere

None.

D) Accepted or in progress

None.

E) How many PE1-found MPs since HUPO-2016 has your chromosome group reported?

0

F) How many candidate MPs found, but not meeting the guidelines 2.1? (Please state if one or two peptides identified and their length, biological replicates found in, and spectral quality).

5 candidates that have been identified by only one peptide. (FDR 1%, Peptide length \geq 9).

G) Any significant change in your Ch Team plans for neXt-50? If so please update the word document form last time and forward. Ill compile this with the other changes as a new summary for the EC.

We are currently working on the testis samples after being disappointed from earlier search works on the liver, pancreas and placenta tissues.

H) Any significant clinical or other successes re a MP that you wish us to consider highlighting in the report.

Chromosome 14

(Charles Pineau)

PIC Leaders:

PI: Charles Pineau; Co-PI: Yves Vandenbrouck

Contributing labs:

- Charles Pineau

PROTIM - Inserm U1085 - Irset, Campus de Beaulieu, 35042 Rennes cedex France

- Jérôme Garin

Proteomics French Infrastructure (ProFi). <http://www.profiroteomics.fr> - MS/MS platform members:

- o Virginie Brun

EDyP - CEA, DRF, BIG, Laboratoire de Biologie a Grande Echelle, Inserm U1038, 17 rue des martyrs, Grenoble F-38054, France

- o Sarah Cianferani

Laboratoire de Spectrométrie de Masse BioOrganique (LSMBO), IPHC, Université de Strasbourg, CNRS UMR7178, 25 Rue Becquerel, 67087 Strasbourg, France

- o Odile Burlet-Schiltz

Institut de Pharmacologie et de Biologie Structurale, Université de Toulouse, CNRS, UPS, 31062 Toulouse, France

- Thomas Fréour

Service de Médecine de la Reproduction, Inserm U1064, CHU de Nantes, 38 boulevard Jean Monnet, 44093 Nantes cedex, France

Major lab members or partners contributing to the neXt50:

- **Major lab members (*alphabetical order*)**

Christine Carapito (*Targeted mass spectrometry*)

Anne Gonzalez de Peredo (*Discovery mass spectrometry*)

Thomas Fréour (*Clinical male infertility*)

Charles Pineau (*Male Reproduction, clinical proteomics*)

Yves Vandenbrouck (*Bioinformatics*)

- **Partners:**

- Lydie Lane (CALIPHO, SIB, Geneva, Switzerland)
- Cecilia Lindskog (HPA, Uppsala, Sweden)

Status of the Chromosome “parts list”:

Current statistics in neXtProt indicate that 93 entries are still not experimentally validated and considered as missing proteins (PE2-4) on chromosome 14. There are also 17 PE5 remaining to be eventually identified as a product of a protein-coding gene and validated.

Chromosome 14 report in neXtProt release 2016-12-02:

PE1	PE2	PE3	PE4	PE5	Total
531	65	23	5	17	641

Two years ago, the Chromosome 14 consortium has selected the testis as a major source for searching Missing Proteins. We consider our work as a trans-chromosome initiative based on a biological function that is spermatogenesis and sperm maturation. Mass spectrometry analyses are performed by the French Proteomics Infrastructure (ProFI) for which the C-HPP chromosome 14 (France) and 2 (Switzerland) initiatives represent a flagship project. Of note is that no dedicated funding could be obtained in France to tackle C-HPP objectives.

According to our calculation and based on our in house RNAseq transcriptomes datasets, only 18 missing proteins correspond to genes on chromosome 14 that are either preferentially or exclusively expressed in the testis. Interestingly, most of Missing Proteins that are testis-specific are expected to be present in meiotic, postmeiotic germ cells and spermatozoa. These correspond to genes that are statistically more represented on the longer chromosomes (*e.g.*, chromosomes 1, 2). As a consequence, we consider our work as a trans-chromosome initiative and will indeed contribute significantly to other chromosomes with newly found MPs.

Based on mRNA distribution (ArrayExpress, RNAseq Atlas and in house RNAseq datasets), the estimated number of remaining Missing Proteins to be found in reproductive organs/cells is close to 800. As much as 650 could be found in meiotic and postmeiotic germ cells. Approximately 100 could be found in the epididymis, most of which correspond to small peptides (*e.g.*, defensins...) thus calling for specific mass spectrometry validation guidelines. Smaller numbers of missing proteins could be found in the ovary, placenta and embryos. Thus, members of the C-HPP Reproductive cluster will contribute, on the basis of their savoir-faire, to the following steps of our global project by providing news rare and valuable biological materials in which to search for additional MPs (*e.g.*, purified ovarian cells).

PIC and major lab members are aware and strictly follow the HPP metrics and most up to date HPP guidelines. The Vandenbrouck et al., 2016 was produced by our consortium in collaboration with L. Lane (Chr 2) and C. Lindskog (HPA), and all members are co-authors. In addition, Y. Vandenbrouck has participated to the Duek et al., 2016 paper and has been involved in the definition of the latest version of HPP guidelines (Deutsch *et al.*, 2016).

Step by step milestone plan to find, identify and validate MPs:

In a first ongoing step, we have decided to chose the spermatozoa as a source for Missing Proteins. Ongoing projects are:

- 1) The search and identification of **the top 40 MPs** proposed in Duek *et al.*, 2016.
- 2) The identification, characterization and validation of novel protein-coding genes in the human testis using a proteogenomics approach.
- 3) The search, identification and validation of **rare MPs** in the human spermatozoa. That specific project involves prefractionation protocols of the human sperm cell into enriched head, intermediate pieces and flagellar fractions, then preparation of membranes and solubilization of lipophilic transmembrane proteins.

In parallel, an ongoing project aims at demonstrating that some MPs identified in the human spermatozoa proteome (Vandenbrouck *et al.*, 2016) originate from other organs. The corresponding genes are not expressed in the germ cell lineage but in the epididymis head or seminal vesicles. These proteins will integrate the spermatozoa during its transit through the epididymis. This is an interesting new concept that will help understanding the post-testicular maturation of the male gamete.

Milestone dates:

Q2-2017

Identification of the top 40 MPs proposed in Duek *et al.*, 2016.
To be submitted to the 2017 Special Issue C-HPP of J. Prot Res.

Q4-2017

Identification, characterization and validation of novel protein-coding genes in the human testis using a proteogenomics approach. This ongoing work contributes to the validation of additional proteins that are not considered to date as part of the human proteome. *Work will be submitted to a specialized journal in the field of Reproductive Biology*

Q2_2018

Identification, characterization and validation of rare MPs in the human spermatozoa. *To be submitted to the 2018 Special Issue C-HPP of J. Prot Res.*

Identification, characterization and validation of rare MPs in male and female reproductive tracts.

To be submitted to the 2018 Special Issue C-HPP of J. Prot Res.

HUPO 2017 UPDATE

A) Titles and authors of papers submitted to the JPR SI

B) Accepted or in revision

Accepted:

In revision:

C) Titles and authors of other HPP relevant papers submitted elsewhere

D) Accepted or in progress

E) How many PE1-found MPs since HUPO-2016 has your chromosome group reported?

F) How many candidate MPs found, but not meeting the guidelines 2.1? (Please state if one or two peptides identified and their length, biological replicates found in, and spectral quality).

38 targeted

12 identified with >2 peptides

3 identified with 1 peptide

Antibody validation for 7 MPs and for one for which MS evidence was weak

Identified with >=2 peptides

Chro 14: ADAM20, FAM181A, SLC35F4, TEX22, LRRC9 = 5

Chro 2: DNAJC5G, ALLC, MAP3K19, FER1L5, C2orf78, EVX2, HERV-H_2q24.1 = 7

1 peptide only:

Chro 14:

Chro 2: KIAA2012, ANKRD30BL, HERV-H_2q24.3

G) Any significant change in your Ch Team plans for neXt-50? If so please update the word document form last time and forward. Ill compile this with the other changes as a new summary for the EC.

No.

H) Any significant clinical or other successes re a MP that you wish us to consider highlighting in the report.

Our plans are to continue digging MPs from the unseen. We will continue to team up with Chro 2

- Presently working on the preparation of highly enriched Sperm parts (Flagella, Nucleus and intermediate piece) to facilitate the search for new MPs. Ongoing optimization of cell fractionation.

- Stimulating idea from Lydie Lane: Perform a targeted search in relevant cell lines of those MPs we did not succeed in identifying this year. Agenda and workplace to be discussed Q4_2017.

Chromosome 15

(Gilberto Domont)

PIC Leaders:

Gilberto B Domont, Fabio CS Nogueira, Paulo C Carvalho, Laboratory for Proteomics and Protein Engineering, Carlos Chagas Institute, Fiocruz, Paraná, Brazil

Contributing labs:

Gilberto B Domont, Lab Head, Proteomics Unit, Federal University of Rio de Janeiro, RJ, Brazil,
gilbertodomont@gmail.com

Fabio Cesar Sousa Nogueira, Proteomics Unit, Lab Head, Proteomics Unit, Federal University of Rio de Janeiro, RJ, Brazil, fabiocsn@gmail.com

Paulo C Carvalho, Laboratory for Proteomics and Protein Engineering, Carlos Chagas Institute, Fiocruz, Paraná, Brazil

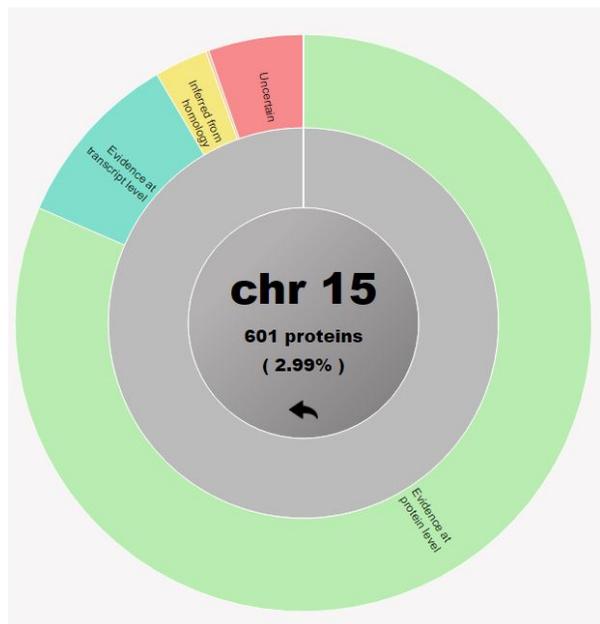
Rafael Melani, PhD, Proteomics Unit, Federal University of Rio de Janeiro, RJ, Brazil

Erika Velasquez, PhD student, Proteomics Unit, Federal University of Rio de Janeiro, RJ, Brazil

Clarissa Ferolla, PhD student, Proteomics Unit, Federal University of Rio de Janeiro, RJ, Brazil

Jimmy Murillo, PhD student, Proteomics Unit, Federal University of Rio de Janeiro, RJ, Brazil

Status of the Chromosome “parts list”:



Proteins by chromosome & protein evidence

● PE1 : Evidence at protein level
● PE2 : Evidence at transcript level
● PE3 : Inferred from homology
● PE4 : Predicted
● PE5 : Uncertain

14	531	65	23	5	17	641
15	490	60	18	1	32	601
16	713	76	21	3	23	836
17	997	122	18	6	23	1166
18	244	18	2	2	9	275
19	1138	222	29	7	32	1428
20	456	62	15	3	13	549
21	179	29	20	1	24	253
Chr	PE1	PE2	PE3	PE4	PE5	Total
All	16618	2247	573	96	574	20108
	82.65%	11.17%	2.85%	0.48%	2.85%	100%

Confirmation that PIC and C-HPP lab members have read: *Confirmed*

Step by step milestone plan to find, identify and validate MPs:

STEP 1.

Part of the IVTT cluster, which targets 50 missing proteins in C5 10, 15, 16, and 19 that have high probability of identification in cell lines COV318, PANC1, DFCI024, D341MED, ST486, KLE. COV318 and KLE cell lines are under analysis. MRM approach.

STEP 2:

PRM of 65 missing proteins. First classical PRM using 65 synthetic peptides that map for missing proteins. Almost all initial adjusting experiments were done. Literature search showed that thyroid, brain and testis express the proteins to be assayed. Also we have already collected the thyroid normal and cancerous lobes as the samples to be tested first. Next we will use sample from fetuses brain. This project just started and we expect to present results during Dublin HUPO.

Milestone dates:

STEP 1 - Apr. 2017

STEP 2 - Sept 2017

HUPO 2017 UPDATE

A) Titles and authors of papers submitted to the JPR SI

Erika Velásquez, Fabio CS Nogueira, Ingrid Velásquez, Andrea Schmitt, Peter Falkai, Gilberto B Domont and Daniel Martins-de-Souza. Synaptosomal proteome of the orbitofrontal cortex from schizophrenia patients using quantitative label-free and iTRAQ-based shotgun proteomics.

B) Accepted or in revision

Accepted:

In revision: Erika Velásquez, Fabio CS Nogueira, Ingrid Velásquez, Andrea Schmitt, Peter Falkai, Gilberto B

Domont and Daniel Martins-de-Souza. Synaptosomal proteome of the orbitofrontal cortex from schizophrenia patients using quantitative label-free and iTRAQ-based shotgun proteomics.

C) Titles and authors of other HPP relevant papers submitted elsewhere

None.

D) Accepted or in progress

None.

E) How many PE1-found MPs since HUPO-2016 has your chromosome group reported?

None. Five before 2016.

F) How many candidate MPs found, but not meeting the guidelines 2.1? (Please state if one or two peptides identified and their length, biological replicates found in, and spectral quality).

In the paper submitted to the JPR SI we report two missing proteins candidates:

1) Protein unc-79 homolog isoform Iso 3 (14q32.12): identified with two unique peptides, with eleven and ten aa lengths respectively. They were found in at least two technical replicates in seven schizophrenia patients and in one replicate in a pool of control (n=8). The spectra quality reported by Proteome discoverer 2.1 software was high confidence.

2) Putative cytochrome b-c1 complex subunit Rieske-like protein 1 isoform Iso 1(22q13.1): identified with one unique peptide, with fourteen aa length. It was found in at least two technical replicates in all schizophrenia patients and pool of control (n=8). The spectra quality reported by Proteome discoverer 2.1 software was high confidence.

G) Any significant change in your Ch Team plans for neXt-50? If so please update the word document form last time and forward. Ill compile this with the other changes as a new summary for the EC.

H) Any significant clinical or other successes re a MP that you wish us to consider highlighting in the report.

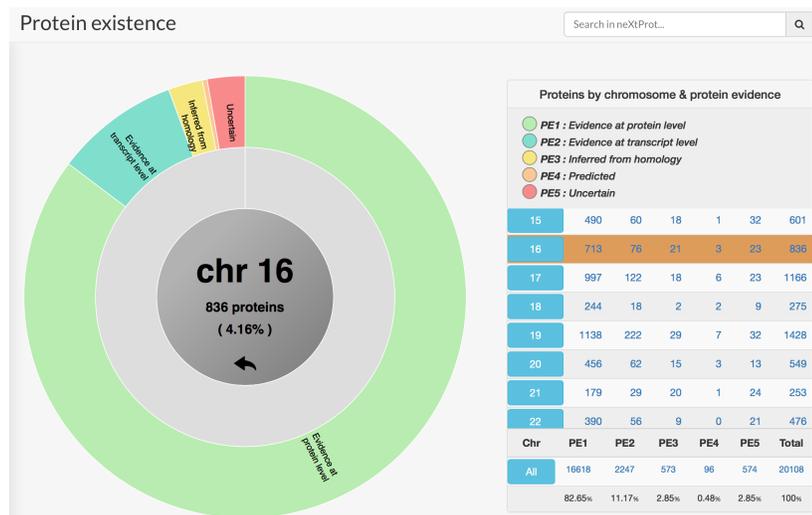
Chromosome 16

(Fernando Corrales)

PIC Leaders: Fernando J. Corrales, Concha Gil

Contributing labs:

Major lab members or partners contributing to the neXt50: ProteoRed (Coordinator Fernando J. Corrales; Centro Nacional de Biotecnología (CNB-CSIC), Madrid, Spain).



**The numbers in this schema are not the same in the downloadable table.

Confirmation Reading key papers:
Done.

Step by step milestone plan to find, identify and validate MPs:

- Prioritization of the 50 chr16 MPs to be targeted. Annotation with Missing ProteinPedia, HPA, Human MRM Atlas

and Chr16 already generated information. February 2017.

- List of chr16 MPs that would require non standard procedures (non-trypsin digestion sites, lack of unique peptides...). March 2017.
- Preparation of SRM methods and identification of the biological sources with highest probability of finding the selected chr16 MPs. June 2017
- SRM analysis, 10 MPS/lab. Starting on June-July 2017).
- IVTT initiative (5 chromosomes participating). 5-10 proteins/year. (2 MPs already identified by SRM and 2 additional are pending on validation)

IVTT initiative (5 chromosomes participating). 5-10 proteins/year.

Targeted search of Ch.16 MPs identified in our extensive shotgun datasets (compliant with HUPO guidelines) in sperm and HEK293 cells. 10-15 proteins/year.

Identification of those MPs that would require nonstandard procedures (non-trypsin digestion sites, lack of unique peptides...). March 2017.

HUPO 2017 UPDATE

A) Titles and authors of papers submitted to the JPR SI

1. Elizabeth Guruceaga, Alba Garin-Muga, Bartolomé Bejarano, Juan Vizcaíno Fernando J. Corrales, Victor Segura. Enhanced missing proteins detection in NCI60 cell lines using an integrative search engine approach.
2. María Isabel Mora, Manuela Molina, Leticia Odriozola, Félix Elortza, José María Mato, Barbara Sitek, Pumin Zhang, Fuchu He, María Uxue Latasa, Matías Antonio Ávila and Fernando José Corrales* Prioritizing Popular Proteins in Liver Cancer: Remodelling One-Carbon Metabolism

B) Accepted or in revision

Accepted:

In revision:

1. Elizabeth Guruceaga, Alba Garin-Muga, Bartolomé Bejarano, Juan Vizcaíno Fernando J. Corrales, Victor Segura. Enhanced missing proteins detection in NCI60 cell lines using an integrative search engine approach.
2. María Isabel Mora, Manuela Molina, Leticia Odriozola, Félix Elortza, José María Mato, Barbara Sitek, Pumin Zhang, Fuchu He, María Uxue Latasa, Matías Antonio Ávila and Fernando José Corrales* Prioritizing Popular Proteins in Liver Cancer: Remodelling One-Carbon Metabolism

C) Titles and authors of other HPP relevant papers submitted elsewhere

1. Alvarez-Sola G, Uriarte I, Latasa MU, Urtasun R, Bárcena-Varela M, Elizalde M, Jiménez M, Rodríguez-Ortigosa CM, Corrales FJ, Fernández-Barrena MG, Berasain C, Avila MA. Fibroblast Growth Factor 15/19 in Hepatocarcinogenesis.
2. Martínez-Fernández L, González-Muniesa P, Laiglesia LM, Sáinz N, Prieto-Hontoria PL, Escoté X, Odriozola L, Corrales FJ, Arbones-Mainar JM, Martínez JA, Moreno-Aliaga MJ. Maresin 1 improves insulin sensitivity and attenuates adipose tissue inflammation in ob/ob and diet-induced obese mice.
3. Alvarez-Sola G, Uriarte I, Latasa MU, Fernandez-Barrena MG, Urtasun R, Elizalde M, Barcena-Varela M, Jiménez M, Chang HC, Barbero R, Catalán V, Rodríguez A, Frühbeck G, Gallego-Escuredo JM, Gavaldà-Navarro A, Villarroya F, Rodríguez-Ortigosa CM, Corrales FJ, Prieto J, Berraondo P, Berasain C, Avila MA. Fibroblast growth factor 15/19 (FGF15/19) protects from diet-induced hepatic steatosis: development of an FG F19-based chimeric molecule to promote fatty liver regeneration.
4. Vialas V, Colomé-Calls N, Abian J, Aloria K, Alvarez-Llamas G, Antúnez O, Arizmendi JM, Azkargorta M, Barceló-Batllori S, Barderas MG, Blanco F, Casal JI, Casas V, de la Torre C, Chicano-Gálvez E, Elortza F, Espadas G, Estanyol JM, Fernandez-Irigoyen J, Fernandez-Puente P, Fidalgo MJ, Fuentes M, Gay M, Gil C, Hainard A, Hernaez ML, Ibarrola N, Kopylov AT, Lario A, Lopez JA, López-Lucendo M, Marcilla M, Marina-Ramírez A, Marko-Varga G, Martín L, Mora MI, Morato-López E, Muñoz J, Odena MA, de Oliveira E, Orea I, Ortea I, Pasquarello C, Ray KB, Rezeli M, Ruppen I, Sabidó E, Del Pino MM, Sancho J, Santamaría E, Vazquez

J, Vilaseca M, Vivanco F, Walters JJ, Zgoda VG, Corrales FJ, Canals F, Paradela A. A multicentric study to evaluate the use of relative retention times in targeted proteomics.

5. Segura V, Garin-Muga A, Guruceaga E, Corrales FJ. Progress and pitfalls in finding the 'missing proteins' from the human proteome map.

D) Accepted or in progress

All accepted.

E) How many PE1-found MPs since HUPO-2016 has your chromosome group reported?

IVTT initiative: 2 confirmed in sperm. The Swiss-French team already annotated as PE1 in the 2017 neXtprot release.

IVTT initiative: 9 pending for validation with standards.

F) How many candidate MPs found, but not meeting the guidelines 2.1? (Please state if one or two peptides identified and their length, biological replicates found in, and spectral quality).

165

G) Any significant change in your Ch Team plans for neXt-50? If so please update the word document form last time and forward. Ill compile this with the other changes as a new summary for the EC.

None.

H) Any significant clinical or other successes re a MP that you wish us to consider highlighting in the report.

None.

Chromosome 17

(Gil Omenn)

PIC Leaders: New Gil Omenn, Mike Snyder, and Ron Beavis

Contributing labs: Guan Lab, University of Michigan; Eric Deutsch, ISB.

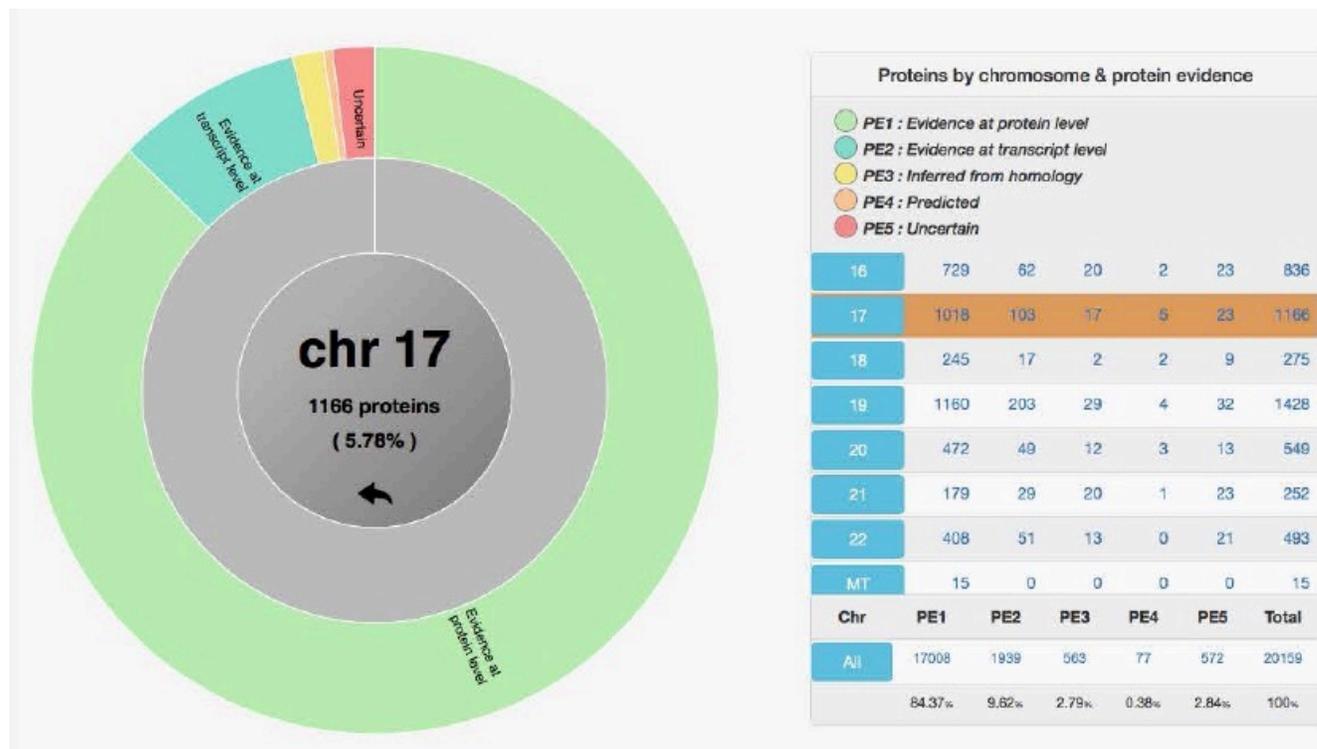
Major lab members or partners contributing to the neXt50: Bharat Panwar, Hongjiu Zhang, Yuanfang Guan

Status of the Chromosome “parts list”:

We have expression profiles from GTEx and from Cancer Genome Cloud.

We have initiated analysis of the Chr 17 full set of 1167 predicted proteins following the Duek et al Chr 2/Chr 14 pipeline, as reported to C-HPP leadership and Chr 17 colleagues on 23 Dec 2016.

Step by step milestone plan to find, identify and validate MPs:



(a) 23 December 2016 email to C-HPP (and B/D-HPP) leaders and Chr 17 team:

“Of the 1167 protein-coding genes on Chr 17, there are 148 “missing proteins”, which are neXtProt 2016-02 PE2,3,4 (excluding PE5). The C-HPP goal is to find credible evidence for at least 50 from each chromosome. I propose that we carefully annotate the missing proteins to identify those most likely to be detectable by mass spectrometry and to identify those not (sufficiently) expressed as transcripts or not suitable for solubilization and tryptic digestion. We need guidance on the use of HPA immunohistochemistry findings (illustrated by SMIM6 and SMIM5, below); at present, I believe neXtProt uses HPA findings for corroboration but not for primary PE1 identification.

We also need to know which of these 148 moves into the neXtProt to SwissProt curation pipeline with pending approval as PE1.

I will make inquiries about securing SRM analyses for our top candidates if appropriate tissues can be obtained; sharing of key specimens across C-HPP teams would be highly desirable.

Ron and Mike: we welcome your guidance and participation, especially with algorithms for characterizing these protein sets. My colleague, Bharat Panwar, has linked his MI-Proteome Visualization Tool for quantitative tissue expression of transcripts and proteins to searches of neXtProt, GPMdb, GTEEx, Genomic Data Commons, PeptideAtlas, and Protein Atlas.

Fernando and Lydie: We are keen to interact with the Spanish Chr 16 group and the Chr 2/14 consortium on applying each other’s methods and tools.

Mike, Chris, et al: I think it would be desirable to have a de minimus median level for transcript expression in each tissue (probably 1.0 RPKM) for the whole project, to avoid vast confusion across the C-HPP teams.

Jenny and Fernando: Hopefully, the B/D-HPP teams will contribute actively, too, with tissue-based orientation.

1. neXtProt 485 nearly-PE1 proteins excluded under Guidelines v2.1 (from Metrics 2016 Omenn et al, Supplementary table): 33 are on Chr 17, including 28 PE2, 4 PE3, 1PE4.

- We will determine where is the deficiency in qualifying for PE1 status in each case.
- We can determine tissues with transcript expression across multiple data resources (neXtProt/PeptideAtlas,

GPMdb/green, Pandey/human proteome map, GTEx, Genomic Data Commons, and Protein Atlas) using MI-Proteome Visualization Tool MI-PVT (Panwar, JPR 2015).

- Last month I inquired of Lydie whether they had analyzed needed data for the 485; she responded “I would suggest to wait for our new release in February, since many of those proteins will probably be upgraded to PE1 due to the integration of all the HPP data from 2016 JPR SI.”

2. While awaiting nX 2017-02, we searched nX2016-02 from the Supplementary Table of 2949 missing proteins in the Metrics 2016 paper: There are 148 Chr 17 missing proteins = 123 PE2, 19 PE3, 6 PE4. All of the 33/485 above are included in the 148.

We have started to perform Duek et al-like analysis for Chr 17.

There are 14 olfactory receptor coding genes, with 11 in a cluster, 10 PE2 and 4 PE3. Specifically:

- OR1A1 is listed as PE2, but in GTEx shows expression only in endocervix (N=5), with max RPKM value 0.04 and median value 0.00!

- OR1A2 listed as PE2, but GTEx shows expression only in testis (N=172) with median 0.00, 75th percentile 0.05, and a handful of points higher, with highest outlier 0.23 RPKM.

- OR1D4, PE3 in neXtProt apparently based on Zebrafish! However, nX gives a citation to Ben Arie, Lancet, et al, Hum Mol Genetics 3:229, 1994 (a top group in Israel’s Weizmann Institute; developers of Gene Cards and many studies of olfactory receptor genes. Meanwhile, GTEx shows only negligible activity for one isoform in human testis; median RPKM 0.00 (N=172). Glusman/Lancet in Genomics 2000;63:227 described extremely complex sequences in the 17p13.3 cluster of 17 OR coding regions (6 pseudogenes; various interspersed repetitive elements/LINE repeats; mapped to two families and seven subfamilies in seemingly random orientations; sounds like non-functional sequences!).

- We have long discussed in HPP potential strategies for analyses of gene families; those sequences that are highly homologous will be hard to distinguish from a few peptide sequences, thereby contributing to non-detectable protein products. The Lancet Lab published DEFOG: A practical scheme for deciphering families of genes; Genomics 2002;80:295. Using degenerate primers for PCR, DEFOG tripled the initial repertoire of human ORs in a single experiment. There were a billion combinations of distinct PCR primer pairs.

- I think we can exclude the OR genes from our analysis, as Duek et al did, given the failure of essentially all previous OR protein expression claims to survive independent reanalyses. This example shows a limitation of the PE2 strategy, namely reliance on extremely low RPKM reports.

- **Mark:** Are you proceeding with OR analyses across all chromosomes? We are interested to collaborate. I highly recommend the large body of work from Doron Lancet (with refs above).

- There are also 14 keratin-associated proteins at 17q21.2 For example, KRTAP2-2 (PE2) in GTEx shows expression only in skin, with a few very high values (RPKM 500-1000) but most very close to zero. KRTAP9-1 (PE3): also limited to skin, with only a few values above 0.5 RPKM.

- We will create a table showing not just PE2 or PE3, but highest expressed tissue(s) and median RPKM for the whole 148 predicted proteins. It is important to use multiple sources.

- For example, SMIM6 (P0D180) is “small integral membrane protein 6”. It is PE4 in nX. But in GTEx, there is quite high expression in testis (median 20 RPKM) and in stomach (median 7 RPKM) and Kidney cortex (median 6) across 172 samples; this would not qualify as testis-enriched, since the testis level is <5x other tissues. We even have a project using Drop-Seq for single cell analyses of spermatogenesis stages in mouse testis; that could yield some interesting findings for homology to human (PE3). For MS/MS, SMIM6 has only 62 amino acids with a single K at position 7 and no R residue! No tryptic peptides of 9-30 aa in length can be generated; ProteomeMap is blank by MS for all tissues (even though Pandey’s Nature paper used a single peptide of 6aa as sufficient; we checked that paper. Protein Atlas shows highest IHC in testis (along with RPKM 22, great agreement with GTEx).

- SMIM5 (77aa): ProteomeMap shows highest activity in platelets; none in testis; low in placenta, frontal cortex, spinal cord. This protein sequence has 6 R’s and 2 K’s, but they are clustered such that only the N-terminal and C-terminal tryptic peptides might be useful. In fact, Pandey reported only the N-terminal (10 aa/16 aa with a

missed cleavage) and two modified N-terminal peptides (acetylated/oxidized). This is expected for an integral membrane protein. GTEx shows transcript expression medians >5 for several tissues, not including testis. In Protein Atlas, SMIM5 is detected as highly expressed in essentially all tissues by immunohistochemistry, but is highly variable and not well correlated in RNA expression. Might be a good one to subject to validation in the Antibody Validation Work Group project.

Response from Ron Beavis:

I've attached my analysis of the list.

The spreadsheet has 5 categories, broken out as separate tabs.

As of 1/16/17, we have added the numbers in each category and searched Vandenbrouck et al for their missing protein findings of Chr 17 proteins (see (c), below):

1. high priority (n=44): proteins that should be observable and for which there is some reasonably good evidence at the moment. I have indicated the tissues/cell types where they should be present. The (11) KRTAP proteins have considerable sequence homology, but I think there are enough differences to tease them out of good data.
2. peptide overlap (n=25): proteins that are observable, but very difficult (some impossible) to distinguish from other proteins at the level of tryptic peptides because of sequence identity. In these cases the problem is biology, not technique.
3. low priority (n=69): not enough information to know where to look for these proteins, if they exist at all.
4. revised genes (n=3): proteins associated with genes that have been significantly revised over the last few years. The extensive revisions may indicate a problem with the DNA sequences.
5. retired+non-coding (n=28): genes that have either been retired from the current genome assembly (GRCh38) or are now considered either lincRNA or pseudogenes.

(c) Additional analyses as of 16 January 2017

Eric and Lydie:

Compared the 14 Chr 17 missing proteins reported by Vandenbrouck et al with the latest PeptideAtlas 148 proteins still missing (link below).

Eric, 10 of the 14 failed to pass your filters: FAM1887A (0 distinct pep); LRRC37A2 (representative); C17orf105 (0 pep); C17orf 74 (1 pep, subsumed); TMEM95 (0 pep); FBXW10 (2 pep, but marginally distinguished); FBX039 (1 pep); PROCA1 (0 pep); C17orf50 (1 pep); SLC25A39 (2 pep, marginally distinguished). That leaves only LRRC37A1, C17orf987, SPEMC17orf83, and EFCAB3. Vandenbrouck et al report PRM assays for the one-hit wonders, including FBX039 (positive PRM); C17orf50 (PRM weak); TMEM95 (PRM missing, I think).

Ron, of these 14 you had LRRC37A2, LRRC37A1, C17orf98, C17orf105, C17orf74, PROCA1, C17orf50, and SLC25A39 as "high priority" (of your 44 Chr 17 missing proteins with high priority), with the rest ranked lower.

Another striking finding involves the 11 KRTAPs. Several have 30-60 peptides, but all end up as subsumed or indistinguishable in PeptideAtlas. Among the 25 from families with peptide overlap, there are 10 of the TBC1D3 family; there are several of these on Eric's list for Chr 17.

An interesting individual low-priority protein is testis-expressed sequence 19 protein (Q8NA77) which generates only 6 peptides, two of which are <9 aa, 1 C-terminal, 1 extremely long, and 2 that seemed well-suited to MS; however, this protein has not been detected. PeptideAtlas shows this protein as negative by proteomics, but positive by antibodies.

MP Status:

Ron, of these 14 you had LRRC37A2, LRRC37A1, C17orf98, C17orf105, C17orf74, PROCA1, C17orf50, and SLC25A39 as "high priority" (of your 44 Chr 17 missing proteins with high priority), with the rest ranked lower.

Another striking finding involves the 11 KRTAPs; Ron, you suggested these might be differentiable. Several have 30-60 peptides, but all end up as subsumed or indistinguishable in PeptideAtlas. Among the 25 from families with peptide overlap, there are 10 of the TBC1D3 family; and several of these on Eric's list for Chr 17.

An interesting individual low-priority protein is testis-expressed sequence 19 protein (Q8NA77) which generates only 6 peptides, two of which are <9 aa, 1 C-terminal, 1 extremely long, and 2 that seemed well-suited to MS; however, this protein has not been detected. PeptideAtlas shows this protein as negative by proteomics, but positive by antibodies.

HUPO 2017 UPDATE

A) Titles and authors of papers submitted to the JPR SI

Accepted

1) Progress on the HUPO Draft Human Proteome: 2017 Metrics of the Human Proteome Project.

Omenn GS, Lane L, Lundberg EK, Overall CM, Deutsch EW.

J Proteome Res. 2017 Aug 30. doi: 10.1021/acs.jproteome.7b00375. [Epub ahead of print]

PMID: 28853897

In revision

2) A systematic proteogenomic approach to exploring a novel function for NHERF1 in human reproductive disorder: lessons for exploring missing proteins. Na, Keun; Shin, Heon; Cho, Jin-Young ; Jung, Sang Hee ; Lim, Jaeseung; Lim, Jong-Sun; Kim, Eun Ah ; Kim, Hye Sun ; Kang, Ah Reum ; Kim, Ji Hye ; Shin, Jeong Min ; Jeong, Seul-Ki; Kim, Chae-Yeon ; Park, Jun Young; Chung, Hyung-Min ; Omenn, Gilbert; Hancock, William; Paik, Young-Ki, Manuscript ID: pr-2017-00146s.R2.

3) Schwenk, Jochen; Omenn, Gilbert; Sun, Zhi; Baker, Mark; Overall, Christopher; Aebersold, Ruedi; Moritz, Robert; Deutsch, Eric The Human Plasma Proteome draft of 2017: Building on the Human Plasma PeptideAtlas and Alternative Assays. ID: pr-2017-004678.R1

C) Titles and authors of other HPP relevant papers submitted elsewhere

Accepted

A wellness study of 108 individuals using personal, dense, dynamic data clouds. Price ND, Magis AT, Earls JC, Glusman G, Levy R, Lausted C, McDonald DT, Kusebauch U, Moss CL, Zhou Y, Qin S, Moritz RL, Brogaard K, Omenn GS, Lovejoy JC, Hood L. Nat Biotechnol. 2017 Aug;35(8):747-756. doi: 10.1038/nbt.3870. Epub 2017 Jul 17. PMID: 28714965

Functional proteogenomics reveals biomarkers and therapeutic targets in lymphomas. Rolland DCM, Basrur V, Jeon YK, McNeil-Schwalm C, Fermin D, Conlon KP, Zhou Y, Ng SY, Tsou CC, Brown NA, Thomas DG, Bailey NG, Omenn GS, Nesvizhskii AI, Root DE, Weinstock DM, Faryabi RB, Lim MS, Elenitoba-Johnson KSJ. Proc Natl Acad Sci U S A. 2017 Jun 20;114(25):6581-6586. doi: 10.1073/pnas.1701263114. Epub 2017 Jun 12. PMID: 28607076

Nucleome Analysis Reveals Structure-Function Relationships for Colon Cancer. Seaman L, Chen H, Brown M, Wangsa D, Patterson G, Camps J, Omenn GS, Ried T, Rajapakse I. Mol Cancer Res. 2017 Jul;15(7):821-830. doi: 10.1158/1541-7786.MCR-16-0374. Epub 2017 Mar 3. PMID: 28258094

Annotation of Alternatively Spliced Proteins and Transcripts with Protein-Folding Algorithms and Isoform-Level Functional Networks. Li H, Zhang Y, Guan Y, Menon R, Omenn GS. Methods Mol Biol. 2017;1558:415-436. doi: 10.1007/978-1-4939-6783-4_20. PMID: 28150250

miRmine: a database of human miRNA expression profiles. Panwar B, Omenn GS, Guan Y. Bioinformatics. 2017 May 15;33(10):1554-1560. doi: 10.1093/bioinformatics/btx019. PMID: 28108447.

Advances of the HUPO Human Proteome Project with broad applications for life sciences research. Omenn GS. Expert Rev Proteomics. 2017 Feb;14(2):109-111. doi: 10.1080/14789450.2017.1270763. Epub 2016 Dec 22. PMID: 27935328.

The proteomes of the human eye, a highly compartmentalized organ. Omenn GS. Proteomics. 2017 Jan;17(1-2). doi: 10.1002/pmic.201600340. Epub 2016 Dec 21. PMID: 27860232.

Progress in the Chromosome-Centric Human Proteome Project as Highlighted in the Annual Special Issue IV. Paik YK, Overall CM, Deutsch EW, Hancock WS, Omenn GS. J Proteome Res. 2016 Nov 4;15(11):3945-3950. PMID: 27809547.

E) How many PE1-found MPs since HUPO-2016 has your chromosome group reported?

F) How many candidate MPs found, but not meeting the guidelines 2.1? (Please state if one or two peptides identified and their length, biological replicates found in, and spectral quality).

G) Any significant change in your Ch Team plans for neXt-50? If so please update the word document form last time and forward. Ill compile this with the other changes as a new summary for the EC.

H) Any significant clinical or other successes re a MP that you wish us to consider highlighting in the report.

Progress on the Confident identification of neXtProt PE2,3,4 Missing Proteins from the Human Proteome Project Omenn GS, Beavis R, Zhang H, Panwar B, Guan YF, Snyder MP.

From the inception to 2018-08, the PE2,3,4 Missing Proteins have been reduced from 148 to 114, a gain of 34 toward the C-HPP goal of 50.

Starting from the official announcement of the Next50 Challenge in September 2016 in Sun Moon Lake, based on the neXtProt version 2016-01, we have documented progress for Chromosome 17 as follows:

<u>neXtProt version</u>	<u>PE2+3+4</u>	<u>PE2</u>	<u>PE3</u>	<u>PE4</u>	<u>[PE5]</u>
2016-01	148	123	19	6	23
2017-01	125	103	17	5	23
2018-01	114	98	12	4	23

JPR SI finalized 10/1/17 Pending

Following the model of the Chr 2 and Chr 14 consortium (Duek et al, JPR 2016), we examined all the PE2,3,4 entries for Chromosome 17. The UM team and Ron Beavis performed parallel independent assessments. We provided a detailed report to the Next50 Challenge in December 2016. Most of the upgrades in neXtProt were foreseen in the Beavis and Omenn et al high priority candidate lists for Chr 17. Attached is a tabulation with Excel sheets for PE levels 1,2,3,4,5 according to the neXtProt status in version 2017-08. Rightward columns show the PE level for each protein entry as of 2016-01 (the baseline for the 2016 C-HPP SI) and 2017-01 (the baseline for the 2017 C-HPP SI). The individual entries document the progression from PE 3 or PE 2 to PE 1, and occasionally from PE1 downward.

The table also shows the annotations generated in late 2016 by Beavis as high priority, low priority, gene revised or retired in the GRCh38 genome assembly, high homology presenting likely peptide overlaps (similar to indistinguishable in PeptideAtlas). It is important to note, as we did in December 2016, the major groups of homologous proteins, including 8 + 3 olfactory receptor genes, 8 TBC1D genes, 11 KRTAP keratin-associated protein genes, and _ Zinc finger protein genes.

There are many interesting individual observations.

Chromosome 18

(Andrey Archakov / Elena Ponomarenko)

At our page of Chr-18 we do not have so much "completely undiscovered" proteins, but we will be looking forward to verify the most unclear ones.

Reminder sent February 19 for a more detailed plan to complete Ch18.

Response:

In version neXtProt v.2.3.0. Chr18 contains 23 missing proteins (PE2-19; PE3-2 and PE4-2) from 275 entries. During our previous work we did transcriptome and proteome analysis of liver tissue and HepG2 cell line and

found 4 missing proteins on transcriptomic level (1 for Liver and 3 for HepG2) and none at the protein level in these types of biomaterial. Now we are using the same transcripto-proteomic approach (RNASeq and SRM) for the study of sperm, as has been shown (Vandenbrouck et al., 2016) that this type of biomaterial contains the greatest diversity of expressed proteins.

As for the missing proteins there are some questions about status of proteins coded Chr18 according neXtProt. For example, GALR1 (https://www.nextprot.org/entry/NX_P47211/proteomics) has protein existence PE1 however there are no references to any experiments PeptideAtlas or Human Protein Atlas. And opposite situation for some PE2 proteins, which have proteomic experimental data from PeptideAtlas and HPA (https://www.nextprot.org/entry/NX_Q8TB69/proteomics). We wonder how it can be explanation of?

HUPO 2017 UPDATE

A) Titles and authors of papers submitted to the JPR SI

1. Ekaterina V. Poverennaya, Ekaterina, V. Ilgisonis, Elena A. Ponomarenko, Arthur T. Kopylov, Victor G., Zgoda, Sergey P. Radko, Andrey V. Lisitsa, Alexander I. Archakov. Why is the correlations between mRNA and protein levels so low among the 275 predicted protein-coding genes on human?

B) Accepted or in revision

Accepted:

In revision:

1. Ekaterina V. Poverennaya, Ekaterina, V. Ilgisonis, Elena A. Ponomarenko, Arthur T. Kopylov, Victor G., Zgoda, Sergey P. Radko, Andrey V. Lisitsa, Alexander I. Archakov. Why is the correlations between mRNA and protein levels so low among the 275 predicted protein-coding genes on human?

C/D) Titles and authors of other HPP relevant papers submitted elsewhere

1) Kopylov A. T., Ilgisonis E. V., Tikhonova O. V., Farafonova T.E., Novikova S. E., Zgoda V. G., Ponomarenko E. A., Lisitsa A.V., Markin A. A., Grigoriev A. I., Seul-Ki Jeong, Young-Ki Paik, Archakov A.I. Quantitative targeted proteomics of proteins encoded by Chromosome 13 genes in human blood plasma., Submitted.

2) PCR analysis of the absolute number of copies of human chromosome 18 transcripts in liver and HepG2 cells. Kiseleva YY, Ptitsyn KG, Tikhonova OV, Radko SP, Kurbatov LK, Vakhrushev IV, Zgoda VG, Ponomarenko EA, Lisitsa AV, Archakov AI. Biomed Khim. 2017, published in Russian.

3) Dataset of protein species from human liver, Naryzhny S., Maynskova, M., Zgoda V., Archakov A., Data in Brief, 2017, V.12, P.584-588.

E) How many PE1-found MPs since HUPO-2016 has your chromosome group reported?

0

F) How many candidate MPs found, but not meeting the guidelines 2.1? (Please state if one or two peptides identified and their length, biological replicates found in, and spectral quality).

No candidate MPs were found either. After preliminary experiments conducted on sperm, it was found that the amount of mRNA in this type of biomaterial is not sufficient for carrying out Next-Generation Sequencing (NGS). At this stage, the Chr 18 team's efforts are focused on finding the optimal biological material for detection of MP using all public transcriptomic and proteomic data. Besides of using the gene-centric transcripto-proteomic approach (RNASeq+PCR and Shotgun LC-MS/MS+SRM) for the study of different types of biomaterial, we are going to use alternative proteases Glu-C, Lys-C, AspN for improving of MP coverage. Bioinformatic approach for MP includes de novo sequencing (Use of de novo sequencing for proteins identification., Skvortsov VS et al. Biomed Khim.(2017)).

G) Any significant change in your Ch Team plans for neXt-50? If so please update the word document form last time and forward. Ill compile this with the other changes as a new summary for the EC.

None.

H) Any significant clinical or other successes re a MP that you wish us to consider highlighting in the report.

None.

Chromosome 19

(Sergio Encarnación-Guevara)

PIC Leader: Sergio Encarnación-Guevara

Contributing labs:

Julio Collado-Vides. Program of Computational Biology at Center for Genomic Sciences UNAM.

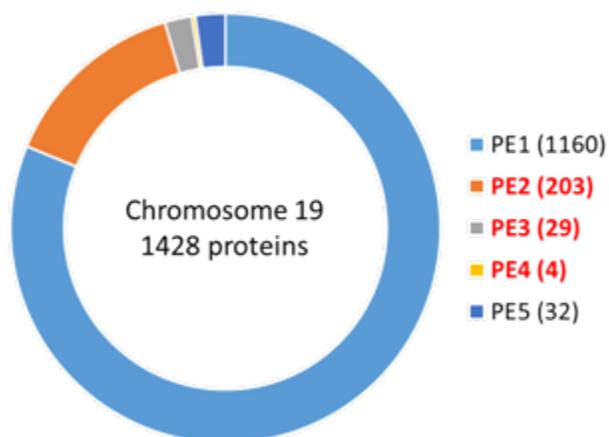
Rafael Palacios de la Lama. International Laboratory for Human Genome Research UNAM

Jeovanis Gil. laboratory of Proteomics Center for Genomic Sciences UNAM.

Partners:

Fernando J. Corrales. ProteoRed, Spain.

Status of the Chromosome “parts list”:



236 proteins from Chromosome 19 remain to be confirmed their existence at protein level.

Confirmation that PIC and HPP lab members have read:

Confirmed

Step by step milestone plan to find identified and validated MPs:

1. Bioinformatics analysis to characterize the 236 MP in the Chromosome 19 according to their feasibility to be identified in proteomics experiments (peptides generated enzymatic digestion, m/z range of peptides, unique peptides, etc.).
2. Transcripts analysis to determine cell lines expressing the mRNA of the MP in different experimental

conditions.

3. In a deep proteomics analysis in cell lines expressing mRNAs of MP, in different experimental conditions.

Large scale lysine acetylation analysis to go deeper in the characterization of this PTM.

4. MRM analysis to validate MP, using synthetic peptides labeled with heavy isotopes.

Milestone dates:

1. March 2017
2. End 2017
3. May 2018
4. October 2018

HUPO 2017 UPDATE

A) Titles and authors of papers submitted to the JPR SI

B) Accepted or in revision

Accepted:

In revision:

C) Titles and authors of other HPP relevant papers submitted elsewhere

D) Accepted or in progress

E) How many PE1-found MPs since HUPO-2016 has your chromosome group reported?

F) How many candidate MPs found, but not meeting the guidelines 2.1? (Please state if one or two peptides identified and their length, biological replicates found in, and spectral quality).

G) Any significant change in your Ch Team plans for neXt-50? If so please update the word document form last time and forward. Ill compile this with the other changes as a new summary for the EC.

H) Any significant clinical or other successes re a MP that you wish us to consider highlighting in the report.

Chromosome 20

(Siqi Liu)

PIC Leaders: Siqi Liu, Qingyu He, Gong Zhang

Contributing labs:

Center of Proteomics Analysis, BGI, China

Institute of Biotechnology, Jinan University, China

Major lab members or partners contributing to the neXt50:

Yan Ren, Qidan Li, Shaohang Xu, Quanhui Wang (BGI)

Tong Wang, Gong Zhang (Jinan University)

Status of the Chromosome “parts list”:

In neXtprot, data release: 2016-12-02, application release: v2.3.0

Chr	PE1	PE2	PE3	PE4	PE5	Total
20	<u>456</u>	<u>62</u>	<u>15</u>	<u>3</u>	<u>13</u>	<u>549</u>

According to CHPP definition to missing proteins, the proteins in PE2+PE3+PE4 are defined as missing proteins. Therefore, the missing proteins in chromosome 20 in the neXtprot database are 80 (62+15+3).

Confirmation that PIC and C-HPP lab members have read:

We have read the articles and understood the core concepts of missing proteins.

Step by step milestone plan to find, identify and validate MPs:

In the project of “Cancer Proteogenomics of the Esophageal Squamous Cell Carcinoma tissues”, we are going to identify more than 40 missing proteins in chr. 20.

In the project of “Cancer Proteogenomics of the colorectal cancer”, we are going to find more than 15 missing

proteins in chr. 20.

We are able to search for missing proteins for all chromosomes, not only for just one chromosome. This may be even easier for us and can contribute more to the community.

To tackle the NeXt-50 challenge, we developed a new method, "high-throughput de novo proteome identification aided by translome sequencing" to maximize the utilization of mass spectra. Using this method, we have identified altogether 840 missing proteins (755 PE2, 58 PE3, 7 PE4 and 26 PE5). Among them, we have identified 519 missing proteins with ≥ 2 unique peptides. We have verified some of them using MRM and antibody. We presented this strategy in AOHUPO 2016 as an oral presentation.

Milestone dates:

We will try the best to find over 50 missing proteins in ch 20 within 2017.

HUPO 2017 UPDATE

A) Titles and authors of papers submitted to the JPR SI

Digging more missing proteins using an enrichment approach with ProteoMiner. Siqi Li, Yanbin He, Zhilong Lin, Shaohang Xu, Ruo Zhou, Feng Liang, Jian Wang, Huanming Yang, Siqi Liu, Yan Ren

B) Accepted or in revision

Accepted:

In revision: Digging more missing proteins using an enrichment approach with ProteoMiner. Siqi Li, Yanbin He, Zhilong Lin, Shaohang Xu, Ruo Zhou, Feng Liang, Jian Wang, Huanming Yang, Siqi Liu, Yan Ren

C) Titles and authors of other HPP relevant papers submitted elsewhere

D) Accepted or in progress

E) How many PE1-found MPs since HUPO-2016 has your chromosome group reported?

F) How many candidate MPs found, but not meeting the guidelines 2.1? (Please state if one or two peptides identified and their length, biological replicates found in, and spectral quality).

G) Any significant change in your Ch Team plans for neXt-50? If so please update the word document from last time and forward. Ill compile this with the other changes as a new summary for the EC.

H) Any significant clinical or other successes re a MP that you wish us to consider highlighting in the report.

Chromosome 21

(Albert Sickmann)

Newly appointed PIC for Ch21 in February 2017.

PIC leaders:

The Co-PI's are: Dr. Robert Ahrends (ISAS)

Dr. Andreas Roos (John Walton Muscular Dystrophy Research Centre, Newcastle upon Tyne)

Contributing Labs:

Andreas Roos, The John Walton Muscular Dystrophy Research Centre, Newcastle University, United Kingdom

Robert Ahrends, Leibniz-Institut für Analytische Wissenschaften - ISAS - e.V. Dortmund, Germany

About chromosome 21:

- Small chromosome
- Sequenced as 2nd chromosome in 2000
- About 34 million bases (1.5%) – 50 million bases expected
- 14 monogenic human diseases
- 225 genes, 59 pseudo genes (as published in 2000)

Chromosome 21 and disease genes:

Selected monogenic diseases:

- Homocystinuria (CBS)
- Alzheimer's disease (APP)
- Leukaemia (AML1)
- Amyotrophic lateral sclerosis (SOD1)
- Autoimmune polyglandular disease (AIRE)
- Progressive myoclonus epilepsy (CSTB)

Selected syndromes:

- Down Syndrome (4 types)

(free) **trisomy 21** (~ 95 %)

translocation trisomy 21 (2-4%)

(attached to other chromosomes, mostly to chromosomes 13, 14, 15 and 22)

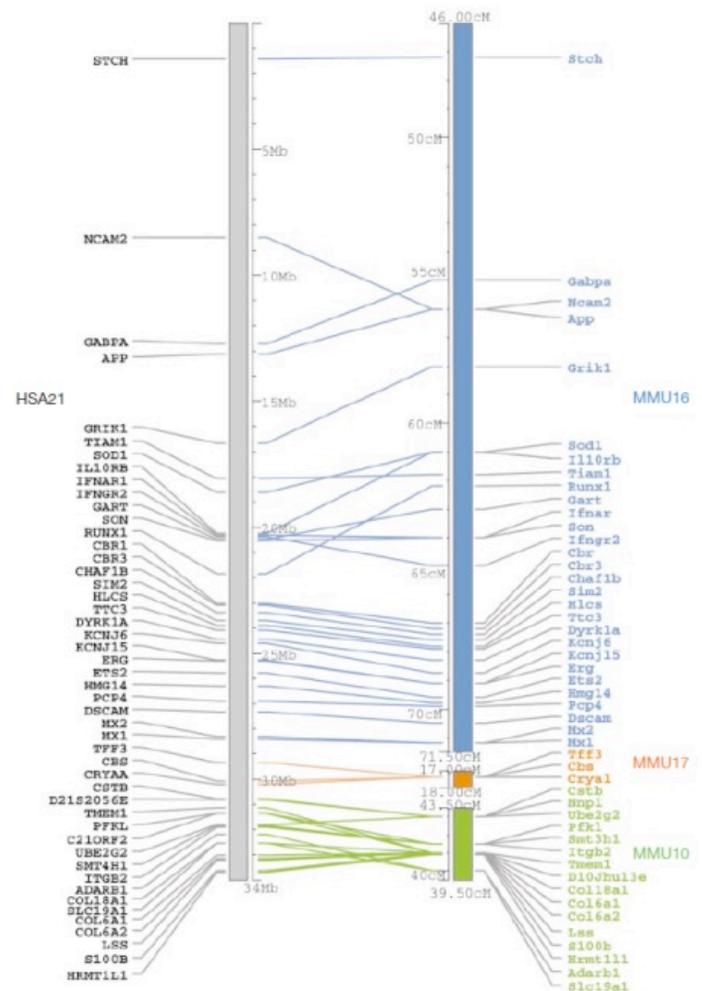
Mosaik trisomy 21 (1-2%)

(mitotic non disjunction, trisomy not in every cell)

partial trisomy 21 (few hundred)

(long arm of chromosome 21 is fused to another chromosome 21)

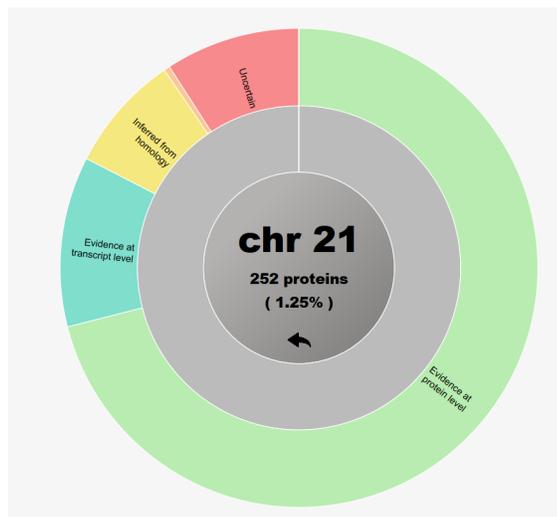
- Usher syndrome 1E
- Knobloch Syndrome
- Leucocyte adhesion deficiency
- Betlehem Myopathy
- Ullrich Myopathy



Homology of Chromosome 21 (human) to different mouse chromosomes: Nature, 405, 311-319

Available gene information:

Figure 1: Overview Chromosome 21 (www.nextprot.org)



NextProt (Nature, 405, 311-319)
 34 million bases, 99,97 % ready
 252 (225) genes
 179 (127) known proteins
 29 (13) with homology to other proteins
 20 (17) with domain homology
 1 (0) predicted
 23 (68) hypothetical
 (59) pseudo genes

Next steps:

1. Detailed look at the data (Check original publication and Nextprot)

2. Proteomics data for 18 genes available (PE2 to PE5)
3. Definition of MRM / PRM assays for missing proteins
4. Definition of sample material
5. Stepwise validation of proteins

Sample material which should be analyzed:

1. Access to specimen from Down Syndrome patients
2. Access to skin biopsy material
3. Access to specimen from Mouse
Homology to proteins from chr. 10, 16 and 17.
Comparison with a dataset of about 12000
proteins from 42 mouse tissues.
4. Probing for missing proteins with SIL based
MRM / PRM assays.
5. Development of a Chromosome 21 assay (Whole chromosome assay)

HUPO 2017 UPDATE

A) Titles and authors of papers submitted to the JPR SI

None.

B) Accepted or in revision

Accepted:

In revision:

C) Titles and authors of other HPP relevant papers submitted elsewhere

None.

D) Accepted or in progress

None.

E) How many PE1-found MPs since HUPO-2016 has your chromosome group reported?

None.

F) How many candidate MPs found, but not meeting the guidelines 2.1? (Please state if one or two peptides identified and their length, biological replicates found in, and spectral quality).

None.

G) Any significant change in your Ch Team plans for neXt-50? If so please update the word document form last time and forward. Ill compile this with the other changes as a new summary for the EC.

We put efforts in ethics to obtain tissues from Trisomie 21 patients which will arrive later this year.

H) Any significant clinical or other successes re a MP that you wish us to consider highlighting in the report.

None.

Chromosome 22

(Akhilesh Pandy)

No response.

Reminder sent February 19, March 16.

HUPO 2017 UPDATE

A) Titles and authors of papers submitted to the JPR SI

B) Accepted or in revision

Accepted:

In revision:

C) Titles and authors of other HPP relevant papers submitted elsewhere

D) Accepted or in progress

E) How many PE1-found MPs since HUPO-2016 has your chromosome group reported?

F) How many candidate MPs found, but not meeting the guidelines 2.1? (Please state if one or two peptides identified and their length, biological replicates found in, and spectral quality).

G) Any significant change in your Ch Team plans for neXt-50? If so please update the word document form last time and forward. Ill compile this with the other changes as a new summary for the EC.

H) Any significant clinical or other successes re a MP that you wish us to consider highlighting in the report.

Chromosome X

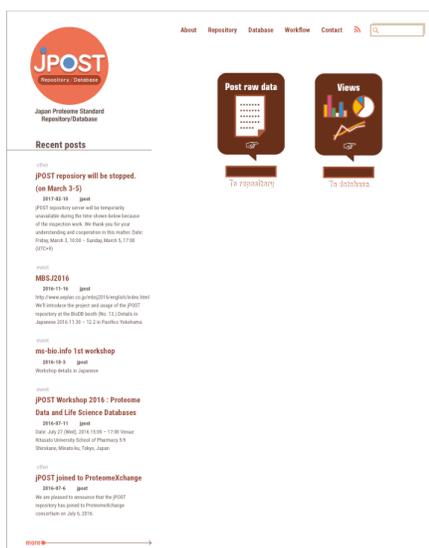
(Tadashi Yamamoto)

Pending ChX meeting on Jan 20 regarding new leadership, team organization and plan.

Reminder sent February 19.

As of 21 Feb new leadership:

PI	Yasushi Ishihama	Kyoto University
Co-PI?	Tadashi Yamamoto	Niigata University



HUPO 2017 UPDATE

A) Titles and authors of papers submitted to the JPR SI

Amr Elguoshy, Y.Hirao, Bo.Xu. S.Saito, Ali.F.Quadery, K.Yamamoto, T.Mitsui, T.Yamamoto. Identification and validation of human missing proteins and peptides in public proteome databases; Data mining strategy.

B) Accepted or in revision

Accepted: None.

In revision: Amr Elguoshy, Y.Hirao, Bo.Xu. S.Saito, Ali.F.Quadery, K.Yamamoto, T.Mitsui, T.Yamamoto. Identification and validation of human missing proteins and peptides in public proteome databases; Data mining strategy.

C) Titles and authors of other HPP relevant papers submitted elsewhere

None.

D) Accepted or in progress

None.

E) How many PE1-found MPs since HUPO-2016 has your chromosome group reported?

None.

F) How many candidate MPs found, but not meeting the guidelines 2.1? (Please state if one or two peptides identified and their length, biological replicates found in, and spectral quality).

None.

G) Any significant change in your Ch Team plans for neXt-50? If so please update the word document form last time and forward. Ill compile this with the other changes as a new summary for the EC.

None.

H) Any significant clinical or other successes re a MP that you wish us to consider highlighting in the report.

None.

Chromosome Y

(Hosseini Salekdeh)

No response.

Reminder sent Feb 19.

Less than 20 unique missing proteins on Ch Y. We are focusing to identify them in various tissues and look for the function of missing proteins. Report expected soon.

Reminder sent March 16.

HUPO 2017 UPDATE

A) Titles and authors of papers submitted to the JPR SI

Meyfour, Anna; Ansari, Hassan; Pahlavan, Sara; Mirshahvaladi, Shahab; Rezaei-Tavirani, Mostafa; Gourabi, Hamid; Baharvand, Hossein; and Salekdeh, Ghasem. Y Chromosome Missing Protein, TBL1Y, May Play an Important Role in Cardiac Differentiation.

Meyfour, Anna; Pooyan, Paria; Pahlavan, Sara; Rezaei-Tavirani, Mostafa ; Gourabi, Hamid; Baharvand, Hossein; and Salekdeh, Ghasem. Chromosome-Centric Human Proteome Project Allies with Developmental Biology: A Case Study of the Role of Y chromosome Genes in Organ Development (Manuscript ID: pr-2017-00446d.R1)

B) Accepted or in revision

Accepted: Meyfour, Anna; Ansari, Hassan; Pahlavan, Sara; Mirshahvaladi, Shahab; Rezaei-Tavirani, Mostafa; Gourabi, Hamid; Baharvand, Hossein; and Salekdeh, Ghasem. Y Chromosome Missing Protein, TBL1Y, May Play an Important Role in Cardiac Differentiation.

In revision: Meyfour, Anna; Pooyan, Paria; Pahlavan, Sara; Rezaei-Tavirani, Mostafa ; Gourabi, Hamid; Baharvand, Hossein; and Salekdeh, Ghasem. Chromosome-Centric Human Proteome Project Allies with Developmental Biology: A Case Study of the Role of Y chromosome Genes in Organ Development.

C) Titles and authors of other HPP relevant papers submitted elsewhere

None.

D) Accepted or in progress

E) How many PE1-found MPs since HUPO-2016 has your chromosome group reported?

F) How many candidate MPs found, but not meeting the guidelines 2.1? (Please state if one or two peptides identified and their length, biological replicates found in, and spectral quality).

One, TBL1Y.

G) Any significant change in your Ch Team plans for neXt-50? If so please update the word document form last time and forward. Ill compile this with the other changes as a new summary for the EC.

We will now also study hESCs and their derivatives to identify missing proteins.

H) any significant clinical or other successes re a MP that you wish us to consider highlighting in the report.

Role of Y Chr genes in differentiation of hESC to cardimyoocyte and probably cardiac cell function.

Chromosome Mt

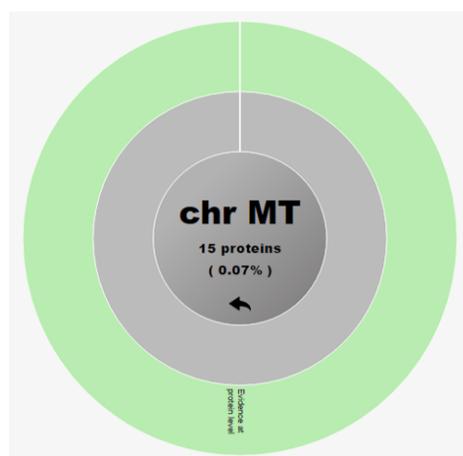
(Andrea Urbani)

Officially closed with all MPs identified.

*Contributing labs:

- Andrea Urbani, Catholic University of Sacred Heart;
- Paola Roncada, Istituto Sperimentale Italiano L. Spallanzani, Milan, Italy;
- Mauro Fasano, University of Insubria, Busto Arsizio, Italy;
- Maurizio Ronci, University G. d'Annunzio of Chieti-Pescara, Chieti, Italy;
- Tiziana Alberio, University of Insubria, Busto Arsizio, Italy;
- Luisa Pieroni, Santa Lucia Foundation, IRCCS, Rome, Italy
- Alessio Soggiu, University of Milan, Milan, Italy;
- Antonio Lucacchini, University of Pisa, Pisa, Italy;
- Italia Bongarzone, IRCCS National Cancer Institute (INT) Milan
- Luigi Palmieri, University of Bari and CNR - IBBE – Bari, Italy;
- Cristina Banfi, Cardiological center, Monzino, Milan;
- Fulvio Magni, University of Milano-Bicocca, Milan, Italy;
- Gabriella Tedeschi, University of Milan, Milan, Italy;
- Salvatore Foti, University of Catania;
- Simona Fontana, University of Palermo, Palermo, Italy;
- Pietro Pucci, University of Naples 'Federico II', Naples, Italy.
- Albert Sickmann, ISAS-Dortmund, DE
- Emma Lundberg, SciLifeLab, KTH Royal Institute of Technology, Stockholm, SW
- Mohan Babu, University of Regina, CA
- Peipei Ping, UCLA, California, USA
- Chris Borchers, University of Victoria Genome British Columbia Proteomics Centre, CA

***Status of the Chromosome “parts list”:**



Chr	PE1	PE2	PE3	PE4	PE5	Total
MT	15	0	0	0	0	15

***Confirmation that PIC and C-HPP lab members have read:**

Yes, we have read them.

***Step by step milestone plan to find, identify and validate MPs:**

For the Mitochondrial Proteome Initiative, Italian consortium purpose is to investigate the protein mitochondrial repertoire in 10 cell lines, which include the following cell lines: HeLa, Hek293, U2OS, BJ, SH-SY5Y, MDA-MB-231, NCI-H28, HUVEC, THP1, HepG2. It has been necessary and indispensable develop a well-standardized method for each cellular model and for the specific biological issue faced.

In brief the following steps have been done:

- Mitochondria have been isolated by three different methods (differential centrifugation, sucrose gradient separation and Mitochondrial Isolation Kit MITOISO2 (Sigma-Aldrich)) in order to evaluate the best preparation;
- Integrity of mitochondrial preparations have been assessed by measuring the oxygen consumption rate and the activity of selected enzymes. They have been evaluated with microscale oxygraphy (Seahorse Bioscience XF96 Analyzer);
- After digestion, mitochondrial peptides have been analyzed using several MS platforms and chromatographic conditions (short list of instrument currently employed: nanoAcquity M Class coupled to Synapt G2 Si (waters); nanoEASY II coupled to Bruker maXis HD; Dionex UltiMate 3000 coupled to Bruker Impact HD; Dionex UltiMate 3000 coupled to Orbitrap Velos ETD; Dionex UltiMate 3000 coupled to Orbitrap Fusion; Eksport nanoLC 400 coupled to TripleTOF 5600+; nanoEASY II coupled to LTQ-Orbitrap-XL);
- Raw data have been processed using Peaks 7.5 (Bioinformatics Solutions Inc., Waterloo, ON Canada) and searched without taxonomy restriction against a custom database containing human proteins, common contaminants and yeast enolase (P00924) downloaded from Uniprot repository (20,442 total entries);
- In the light of the initiative of TAMPA we are checking the presence of missing proteins following the updated Guidelines 2.1.0 evaluating the coding in any chromosome. Open data sharing available before publication following data transfer agreement acceptance.

***Milestone dates:**

- May 2017: special issue JPR, standardization paper on mitochondria preparation
- September 2017: Dublin, HUPO 2017

HUPO 2017 UPDATE

A) Titles and authors of papers submitted to the JPR SI

B) Accepted or in revision

Accepted:

In revision:

C) Titles and authors of other HPP relevant papers submitted elsewhere

D) Accepted or in progress

E) How many PE1-found MPs since HUPO-2016 has your chromosome group reported?

F) How many candidate MPs found, but not meeting the guidelines 2.1? (Please state if one or two peptides identified and their length, biological replicates found in, and spectral quality).

G) Any significant change in your Ch Team plans for neXt-50? If so please update the word document from last time and forward. Ill compile this with the other changes as a new summary for the EC.

H) Any significant clinical or other successes re a MP that you wish us to consider highlighting in the report.