

**PART I: neXt-MP50**

**Executive Summary of neXt-MP50 Reports**

A. Papers submitted to the JPR 2020 HPP SI	<b>15</b>
B. Papers published in the JPR 2019 HPP SI	<b>20</b>
C. Papers relevant to the HPP published elsewhere 2019/2020	<b>61</b>
D. How many MPs (PE2–4) identified as PE1 since 2017	<b>481</b>
E. How many MPs (PE2–4) found in 2019 now listed in neXtprot as PE1	<b>228</b>
F. How many candidate MPs found in 2019, but not meeting Guidelines?	<b>Many hundreds</b>
<p>G. Significant findings to highlight:</p> <ul style="list-style-type: none"> <li>• An issue arising from analysis of the team reports is that a large number of missing proteins that were reported “found” and discussed in their papers were either not captured by Peptide Atlas for analysis or failed reanalysis by Peptide Atlas and so were not promoted PE1 status</li> <li>• Considerable evidence was found for MPs, but failed to satisfy the HPP Guidelines 3.0 and so remain as candidate ‘found’ MPs.</li> <li>• Recommendation: Information regarding these candidate MPs should not be lost but compiled (to be determined where, how and format) so as to be accessible to guide and inform ongoing and future proteomics studies by the community, directed data analysis of similar tissue/cells in Proteome Exchange and current literature to generate evidence sufficient to meet the HPP Guidelines.</li> <li>• Several chromosome teams (e.g. Chr 5, 12) are active in the Cancer Moonshot and CPTAC projects and successfully analysing this data for MPs.</li> <li>• Chromosome 6 initiated a directed search for PE1 proteins lacking MS evidence (termed non-MS PE1 proteins), with several identified by MS (in human bone) that met the HPP Guidelines for PE1 identification by MS.</li> <li>• A precision medicine molecular corrector drug was developed by Chr6 team members that was shown to restore functional levels of a mutant protein isoform of MALT1. Untreated, this mutant protein led to a rare immunodeficiency disease. The disease was phenotyped in a previous paper by proteomics and TAILS that led to this discovery and then drug candidate.</li> <li>• Chromosome 10 has assembled a comprehensive and one of the world’s largest collections of full-length Gateway plasmids representing 90% of all human protein-coding genes and are distributing the collection through their repository and distribution web portal DNASU (dnasu.org). <b>Currently, Chr10 has full-length plasmids for 175 of 804 missing proteins, which are available to the entire C-HPP team.</b> Chr10 (with Chr 5, 15, 16, and 19), have been providing the IVTT-compatible plasmids for missing proteins to other members for IVTT-assisted SRM and continue to generate more plasmids.</li> <li>• The Chromosome 12 (South and SE Asia) team has recruited Radislaw Sobota, Singapore as a new member of the team.</li> <li>• Chr 17 has met the MP50 Challenge: the number of PE2,3,4 missing proteins coded on Chr 17 has been reduced from 148 to 87, meaning that 61 MPs have been detected and incorporated into neXtProt PE1.</li> <li>• Chr X (Japan) also have enjoyed great success in identifying MPs, with 35 now PE1 proteins in neXt-Prot.</li> </ul>	

<ul style="list-style-type: none"> <li>• MT (Italy) some years ago were the first to complete the human proteome encoded by a chromosome, in this case, the smallest in humans with 15 proteins, all now PE1 proteins. Progress is now on the non-MT encoded proteins in the mitochondria.</li> <li>• COVID-19 significantly impaired C-HPP progress on the HPP</li> </ul>	
H. Chromosome Teams reporting in despite COVID-19 shutdowns	<b>21</b>

**Chromosome Number: 1**

**PIC Leaders:** Ping Xu

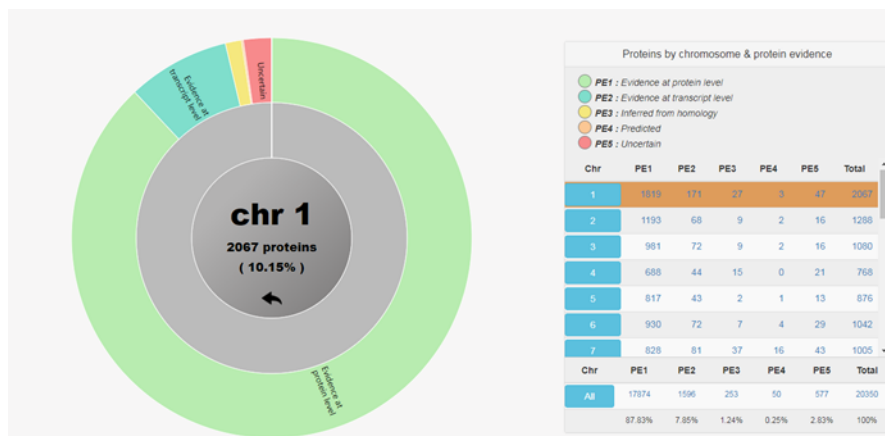
**Part I: Missing Proteins: neXt-MP50 Challenge**

**Major lab members or partners contributing to the neXt-MP50 Challenge**

- Ping Xu (Beijing Proteome Research Center)
- Fuchu He (Beijing Proteome Research Center)
- Yao Zhang (Beijing Proteome Research Center)
- Shujia Wu (Wuhan University)
- Jinshuai Sun (Hebei University)
- Yihao Wang (Beijing Proteome Research Center)
- Jiahui Shi (Hebei University)

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

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



**A) Titles and authors of papers submitted to the 2020 JPR SI or planned.**

Open-pFind Verified Two Missing Proteins from Multi-tissues. Shujia Wu#, Jinshuai Sun#, Feng Xu, Yanchang Li, Bowen Zhong, Yuping Xie, Zhonghua Yan, Lei Chang, Dongxue Wang, Fuchu He\*, Junzhu Wu\*, Yao Zhang\*, Ping Xu\* (pr-2020-00370p, Revised)

**B) Titles and authors of papers published in the 2019 JPR SI.**

Open-pFind Enhances the Identification of Missing Proteins from Human Testis Tissue. Jinshuai Sun#, Jiahui Shi#, Yihao Wang#, Shujia Wu, Liping Zhao, Yanchang Li, Hong Wang, Lei Chang, Zhitang Lyu, Junzhu Wu, Fengsong Liu\*, Wenjun Li\*, Fuchu He\*, Yao Zhang\*, Ping Xu\*. 2019, 18(12): 4189-4196.

**C) Titles and authors of other HPP relevant papers submitted elsewhere in 2019/2020.**

No

**D) How many PE1-found MPs since HUPO-2017 has your chromosome group reported in papers?**

In Wang et al. 2017, we reported the validation of 3 PE2 proteins (Q8N688, P0DMU9, and P0C5Z0) by multi-protease strategy.

In Sun et al. 2018, we reported the validation of 14 PE2 proteins by multiproteases combined with high-pH reverse-phase separation strategy, including Q5TID7, A6H8M9, Q96KW2, Q2VIQ3, Q96JM4, Q8N9V7, Q6NUN7, Q6T311, Q9H4I0, Q8IXR5, Q86WR6, B5MCY1, A0A087WXM9, and Q5VXU9.

In He et al. 2018, we reported the validation of 2 PE2 proteins (Q8N688 and Q86WR6) by low-molecular-weight protein enrichment and a “mirror protease” strategy.

In Sun et al. 2019, we reported the validation of 5 PE2 proteins by multiproteases combined with high-pH reverse-phase separation strategy, including Q8TAA1, Q3ZLR7, A0A1B0GVM6, Q5T1N1, and Q8WW27.

**E) How many PE1-found MPs since HUPO-2019 are now in NeXt-Prot as PE1 proteins? Please check each of your MPs that you reported in the JPR SI.**

Since HUPO 2019, 2 verified MPs (Q8TAA1 and Q5T1N1) from Sun et al. 2019 are now in NeXt-Prot as PE1 proteins.

**F) How many candidate MPs found, but not meeting the guidelines 3.0? (Please state number of peptides identified, their length, and biological replicates found in).**

Other 101 candidate MPs were identified in Wu et al. 2020 work (pr-2020-00370p, Submitted).

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

No

**Chromosome Number: 2**

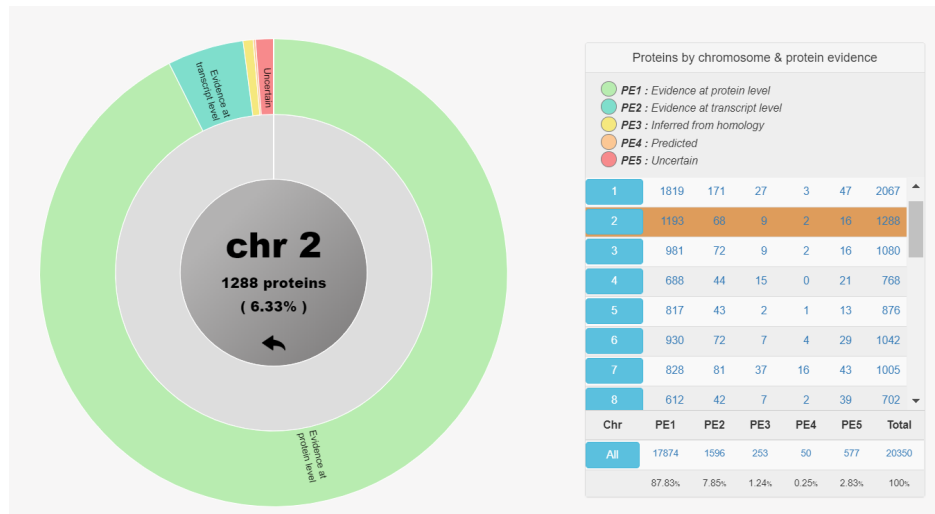
**PIC Leader:** Lydie Lane

**Part I: Missing Proteins: neXt-MP50 Challenge****Major lab members or partners contributing to the neXt-MP50 Challenge**

Paula Duek (SIB/University of Geneva)  
Alain Gateau (SIB/University of Geneva)  
Frédérique Lisacek (SIB/University of Geneva)  
Amos Bairoch (SIB/University of Geneva)  
Charlotte Macron (Nestlé Institute of Health Sciences)  
Antonio Nunez-Galindo (Nestlé Institute of Health Sciences)  
Loïc Dayon (Nestlé Institute of Health Sciences)

**Status of the Chromosome “parts list” for Chromosome 2:**

There are still 79 MPs (68 PE2, 9 PE3 and 2 PE4).



**A) Titles and authors of papers submitted to the 2020 JPR SI or planned.**

*The HUPO Human Proteome Project Reaches a Major Milestone: >90% of Predicted Human Proteins Now Credibly Detected.* Omenn GS, [Lane L](#), Overall CM, Cristea IM, Corrales FJ, Lindskog Bergström C, Paik Y-K, Van Eyk JE, Liu S, Pennington S, Snyder MP, Baker MS, Bandeira N, Aebersold R, Moritz RL, Deutsch EW

*The uncharacterized proteome of human male tissues: a shared resource to uncover new protein functions associated with reproductive biology.* Vandenbrouck Y, Pineau C, [Lane L](#).

**B) Titles and authors of papers published in the 2019 JPR SI.**

*Worming into the Uncharacterized Human Proteome.* [Duek P](#) and [Lane L](#)

*Blinded Testing of Function Annotation for uPE1 Proteins by I-TASSER/COFACTOR Pipeline Using the 2018–2019 Additions to neXtProt and the CAFA3 Challenge.* Zhang C, [Lane L](#), Omenn GS, Zhang Y

*Human Proteome Project Mass Spectrometry Data Interpretation Guidelines 3.0.* Deutsch EW, [Lane L](#), Overall CM, Bandeira N, Baker MS, Pineau C, Moritz RL, Corrales F, Orchard S, Van Eyk JE, Paik Y-K, Weintraub ST, Vandenbrouck Y, and Omenn GS

**C) Titles and authors of other HPP relevant papers submitted elsewhere in 2019/2020.**

*Proteomics Standards Initiative Extended FASTA Format.* (2019) Binz P-A, Shofstahl J, Vizcaíno JA, Barsnes H, Chalkley RJ, Menschaert G, Alpi E, Clauser K, Eng JK, [Lane L](#), Seymour SL, Hernández Sánchez LF, Mayer G, Eisenacher M, Perez-Riverol Y, Kapp EA, Mendoza L, Baker PR, Collins A, Van Den Bossche T, Deutsch EW, JPR

*The neXtProt knowledgebase in 2020: data, tools and usability improvements.* (2020) Zahn-Zabal M, Michel PA, [Gateau A](#), Nikitin F, Schaeffer M, Audot E, Gaudet P, [Duek PD](#), Teixeira D, Rech de Laval V, Samarasinghe K, [Bairoch A](#), [Lane L](#). *Nucleic Acids Res.* 48(D1):D328-D334.

*What will neXtProt help us achieve in 2020 and beyond?* (2020) Zahn-Zabal M, [Lane L](#). *Expert Rev Proteomics.* 17(2):95-98.

*The Human Proteome Project High-Stringency Blueprint* (2020) Adhikari S, Nice E, Deutsch EW, [Lane L](#), Omenn GS, Pennington SR, Overall CM, Paik Y-K, Corrales FJ, Cristea IM, Van Eyk JE, Uhlén M, Lindskog C, Chan DW, [Bairoch A](#), Waddington J, Justice JL, Labaer J, Rodriguez H, He F, Kostrzewa M, Ping P, Gundry RL,

Stewart P, Srivastava S, Srivastava S, Nogueira FCS, Domont GB, Vandembrouck Y, Lam MPY, Wennersten S, Vizcaino JA, Wilkins M, Schwenk JM, Lundberg E, Bandeira N, Marko-Varga G, Weintraub ST, Pineau C, Kusebauch U, Moritz RL, Ahn SB, Palmblad M, Snyder MP, Aebersold R, and Baker MS. *Nat Commun.*, in revision.

**D) How many PE1-found MPs since HUPO-2017 has your chromosome group reported in papers?**

In Carapito et al. 2017, we reported the validation of 12 PE2 proteins by SRM and IHC.

In Robin et al. 2018, we reported the validation of 1 PE2 protein (FRAT2) by reanalysing MS/MS data on 41 HeLa cell datasets.

In the two articles by Macron et al. 2018, we reported the validation of 14 PE2 proteins and 1 PE5 protein (SHISA8) by analysing CSF by MS/MS

We did not report any validation of MP in 2019

**D) How many PE1-found MPs since HUPO-2019 are now in NeXt-Prot as PE1 proteins? Please check each of your MPs that you reported in the JPR SI.**

Due to an incomplete processing of SRM data by PeptideAtlas, only 3/12 proteins validated by Carapito et al. 2017 are now PE1 in neXtProt.

15/16 of the other MPs validated by shotgun approaches are now PE1 in neXtProt.

**→ A total of 18 proteins found by our team since 2017 are now PE1 in neXtProt (5 on chr 2)**

Article	Protein AC	chr	Feb 2017	Feb 2018	Feb 2019	Feb 2020
Carapito 2017 (SRM + Ab-based)	A0AVI2	2	PE2	PE1	PE2	PE2
	Q56UN5	2	PE2	PE1	PE1	PE1
	Q8N7S2	2	PE2	PE2	PE2	PE2
	Q8N6M5	2	PE2	PE2	PE2	PE2
	A6NCI8	2	PE2	PE2	PE2	PE2
	Q03828	2	PE2	PE2	PE2	PE2
	Q9N2J8	2	PE2	PE2	PE2	PE2
	O43506	14	PE2	PE1	PE2	PE2
	Q6ZRR7	14	PE2	PE1	PE2	PE2
	Q8N9Y4	14	PE2	PE2	PE1	PE1
	A4IF30	14	PE2	PE2	PE2	PE2
	C9J3V5	14	PE2	PE2	PE1	PE1
Robin 2018 (reanalysis of HeLa cell data)	O75474	10	PE2	PE2	PE1	PE1
Macron 2018 (MS/MS analysis of CSF)	Q9H1Z8	2	PE2	PE2	PE1	PE1
	Q8NBL3	2	PE2	PE2	PE2	PE1
	Q4G0M1	2	PE2	PE2	PE1	PE1
	B5MCY1	2	PE2	PE2	PE1	PE1
	Q9P2K9	1	PE2	PE2	PE1	PE1
	Q9Y5E5	5	PE2	PE2	PE2	PE1

O60330	5	PE2	PE2	PE2	PE1
Q5BIV9	10	PE2	PE2	PE2	PE2
Q86VR8	11	PE2	PE2	PE1	PE1
Q8N6Y1	13	PE2	PE2	PE1	PE1
A6BM72	15	PE2	PE2	PE1	PE1
Q9ULX5	17	PE2	PE2	PE1	PE1
Q9NUR3	20	PE2	PE2	PE1	PE1
B8ZZ34	22	PE5	PE5	PE1	PE1
Q8IUX8	X	PE2	PE2	PE1	PE1

F) How many candidate MPs found, but not meeting the guidelines 3.0? (Please state number of peptides identified, their length, and biological replicates found in).

Too many to be listed here

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

N/A

**Chromosome Number: 4**

**PIC Leaders:**

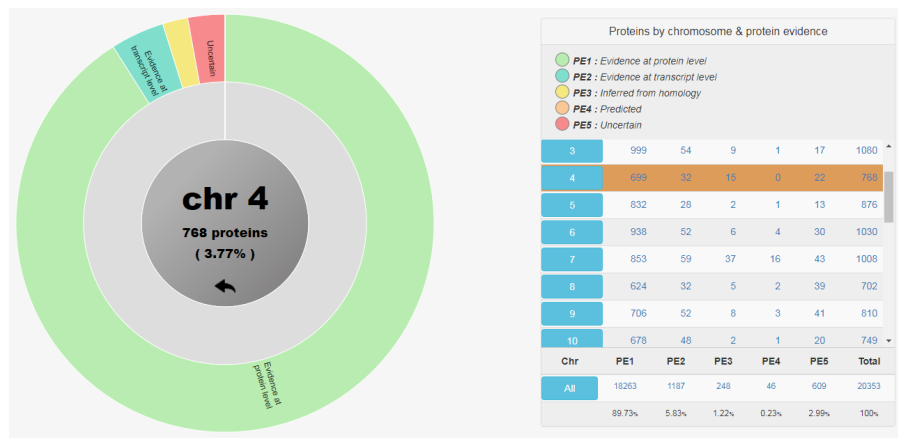
Yu-Ju Chen, Chia-Li Han, Ting-Yi Sung, Sung-Liang Yu

**Part I: Missing Proteins: neXt-MP50 Challenge**

**Major lab members or partners contributing to the neXt-MP50 Challenge**

Reta Birhanu Kitata, Yen-Chen Liao, Wai-Kok Choong Yu-Chang Tyan

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



A) Titles and authors of papers submitted to the 2020 JPR SI or planned.

“Mining Missing Proteins in Tissue Proteome” (data mining ongoing)

**B) Titles and authors of papers published in the 2019 JPR SI.**

Choong, W. K.; Chen, C. T.; Wang, J. H.; Sung, T. Y. iHPDM: In Silico Human Proteome Digestion Map with Proteolytic Peptide Analysis and Graphical Visualizations. *J Proteome Res* 2019, 18, 4124-4132.

**C) Titles and authors of other HPP relevant papers submitted elsewhere in 2019/2020.**

NA

**D) How many PE1-found MPs since HUPO-2017 has your chromosome group reported in papers?**

26 PE1-found MPs were reported in 2018. Based on our recent tissue proteomics profiling on non-smoking lung adenocarcinoma patients (Cell, 182, 226–244, 2020), we identified 125 MPs (115 PE2, 9 PE3 and 1 PE4) and verification is under investigation according to guideline 3.0.

**E) How many PE1-found MPs since HUPO-2019 are now in NeXt-Prot as PE1 proteins? Please check each of your MPs that you reported in the JPR SI.**

23 PE1-found MPs were promoted to PE1 level in neXtProt July 17, 2020 version

**F) How many candidate MPs found, but not meeting the guidelines 3.0? (Please state number of peptides identified, their length, and biological replicates found in).**

NA

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

NA

**Chromosome Number: 5**

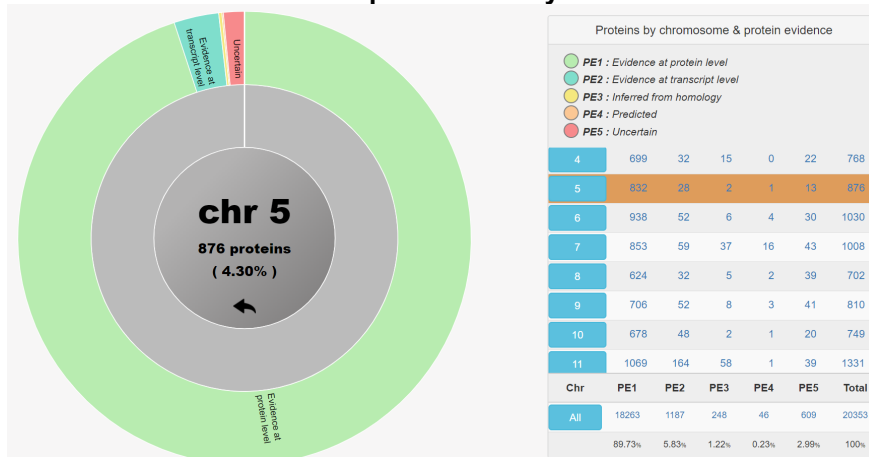
**PIC Leaders:** Peter Horvatovich

**Part I: Missing Proteins: neXt-MP50 Challenge**

**Major lab members or partners contributing to the neXt-MP50 Challenge**

Laboratory of Gyorgy Marko Varga  
Prof. Dr. Rainer Bischoff

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



**A) Titles and authors of papers submitted to the 2020 JPR SI or planned.**

None

**B) Titles and authors of papers published in the 2019 JPR SI.**

None

**C) Titles and authors of other HPP relevant papers submitted elsewhere in 2019/2020. (only papers related to the human proteome)**

- Sanchez A, Kuras M, Murillo JR, Pla I, Pawlowski K, Szasz AM, Gil J, Nogueira FCS, Perez-Riverol Y, Eriksson J, Appelqvist R, Miliotis T, Kim Y, Baldetorp B, Ingvar C, Olsson H, Lundgren L, Ekedahl H, Horvatovich P, Sugihara Y, Welinder C, Wieslander E, Kwon HJ, Domont GB, Malm J, Rezeli M, Betancourt LH, Marko-Varga G., Novel functional proteins coded by the human genome discovered in metastases of melanoma patients, *Cell Biol Toxicol.* 2020, 36(3):261-272.
- K Yang, B Mesquita, P Horvatovich, A Salvati, Tuning liposome composition to modulate corona formation in human serum and cellular uptake, *Acta Biomaterialia* 106, 314-327
- Brandsma CA, Guryev V, Timens W, Ciconelle A, Postma DS, Bischoff R, Johansson M, Ovchinnikova ES, Malm J, Marko-Varga G, Fehniger TE, van den Berge M, Horvatovich P., Integrated proteogenomic approach identifying a protein signature of COPD and a new splice variant of SORBS1, *Thorax* 75 (2), 2020, 180-183.
- Betancourt LH, Szasz AM, Kuras M, Rodriguez Murillo J, Sugihara Y, Pla I, Horvath Z, Pawlowski K, Rezeli M, Miharada K, Gil J, Eriksson J, Appelqvist R, Miliotis T, Baldetorp B, Ingvar C, Olsson H, Lundgren L, Horvatovich P, Welinder C, Wieslander E, Kwon HJ, Malm J, Nemeth IB, Jönsson G, Fenyő D, Sanchez A, Marko-Varga G., The hidden story of heterogeneous B-raf V600E mutation quantitative protein expression in metastatic melanoma—Association with clinical outcome and tumor phenotypes, *Cancers* 11 (12), 1981
- Betancourt LH, Pawlowski K, Eriksson J, Szasz AM, Mitra S, Pla I, Welinder C, Ekedahl H, Broberg P, Appelqvist R, Yakovleva M, Sugihara Y, Miharada K, Ingvar C, Lundgren L, Baldetorp B, Olsson H, Rezeli M, Wieslander E, Horvatovich P, Malm J, Jönsson G, Marko-Varga G., Improved survival prognostication of node-positive malignant melanoma patients utilizing shotgun proteomics guided by histopathological characterization and genomic data, *Scientific reports* 9 (1), 1-14

**D) How many PE1-found MPs since HUPO-2017 has your chromosome group reported in papers?**

**E) How many PE1-found MPs since HUPO-2019 are now in NeXt-Prot as PE1 proteins? Please check each of your MPs that you reported in the JPR SI.**

9 MP (in PMID: 31599373)

**F) How many candidate MPs found, but not meeting the guidelines 3.0? (Please state number of peptides identified, their length, and biological replicates found in).**

24 (in PMID: 31599373)

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

We are currently summarising all mass spectrometry data collected in Cancer Moonshot Melanoma studies with several hundreds of analyzed tumor (primary and metastatic samples). Most of these data sets are or will be uploaded in Proteomic Data Commons (National Cancer Institute).

**Chromosome Number: 6**

**PIC Ch 6 Leaders:**

Robert L. Moritz (USA)

Christopher M. Overall (Canada)



## Part I: Missing Proteins: neXt-MP50 Challenge

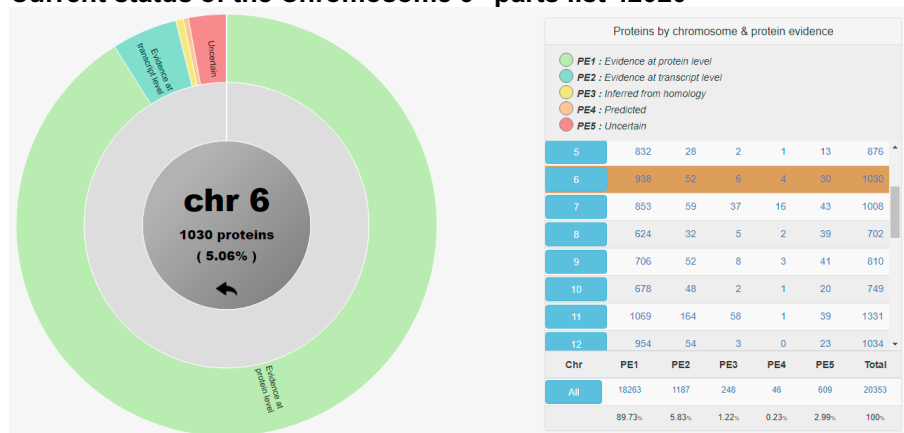
### Major lab members or partners contributing to the neXt-MP50 Challenge

Frank Schmid (Qatar)

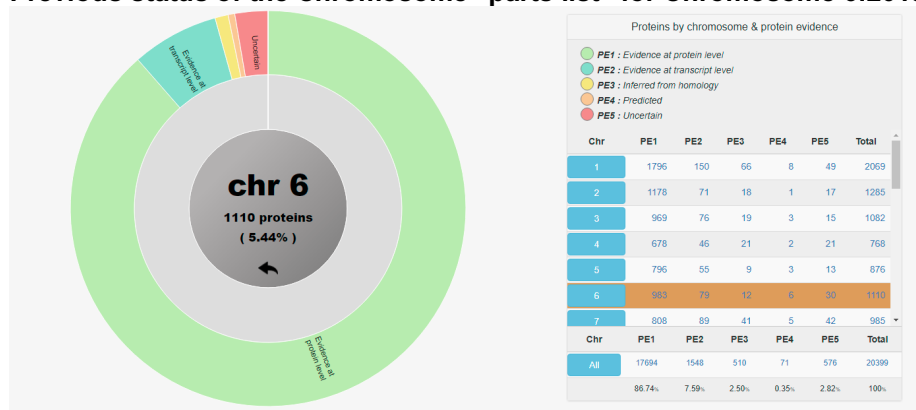
John Wilson (USA)

Eric Deutsch (USA)

### Current status of the Chromosome 6 “parts list”:2020



### Previous status of the Chromosome “parts list” for Chromosome 6:2019



In 2020, the proteome classification has changed and the expected numbers of proteins on Ch6 has reduced from 1110 to 1030 in total. This has reduced the number of PE1 proteins in 2020 to 938 from 983 and PE2-4 proteins have been reduced to 52, 6, and 4 from 79, 12, and 6 respectively. The PE5 proteins remain unchanged.

#### A) Titles and authors of papers submitted to the 2020 JPR SI or planned.

The HUPO Human Proteome Project Reaches a Major Milestone: >90% of Predicted Human Proteins Now Credibly Detected. Gilbert S Omenn, Lydie Lane, Christopher M Overall, Ileana M. Cristea, Fernando J Corrales, Cecilia Lindskog, Young-Ki Paik, Jennifer E Van Eyk, Siqi Liu, Michael P Snyder, Mark S Baker, Nuno Bandeira, Ruedi Aebersold, Robert L. Moritz, and Eric W Deutsch. Submitted to JPR HUPO2020 SI.

#### B) Titles and authors of papers published in the 2019 JPR SI.

Bell, P., Solis, N., Kizhakkedathu, J., Matthew, I., and Overall, C.M. 2019. Proteomic and N-Terminomic TAILS Analyses of Human Alveolar Bone Proteins: Improved Protein Extraction Methodology and LysargiNase Digestion Strategies Increase Proteome Coverage and Missing Protein Identification. Journal of Proteome

Research 18, 4,167 – 4,179, doi: 10.1021/acs.jproteome.9b00445.

Paik, Y-K., Overall, C.M., Corrales, F., Deutsch, E.W., Lane, L., Omenn, G.S. 2019. Advances in Identifying and Characterizing the Human Proteome. *Journal of Proteome Research* 18, 4,079–4,084. doi.org/10.1021/acs.jproteome.9b00745.

Omenn, G.S., Lane, L., Overall, C.M., Corrales, F.J., Schwen, J.M., Paik, Y-K., Van Eyk, J.E., Liu, S., Snyder, M., Baker, M.S., Eric W. Deutsch, E.W. 2019. Progress on Identifying and Characterizing the Human Proteome: 2019 Metrics from the HUPO Human Proteome Project. *Journal of Proteome Research* 18, 4,098 – 4,107. doi: 10.1021/acs.jproteome.9b00434.

Deutsch, E.W, Lane, L., Overall, C.M., Bandeira, N., Baker, M.S., Pineau, C., Moritz, R.L., Corrales, F., Orchard, S., Van Eyk, J.E., Paik, Y-K., Weintraub, S.T., Vandenbrouck, Y., and Omenn, G.S. 2019. Human Proteome Project Mass Spectrometry Data Interpretation Guidelines 3.0. *Journal of Proteome Research* 18, 4,108 – 4,116. doi.org/10.1021/acs.jproteome.9b00542.

Yang, H., Li, Y., Zhao, M., Wu, F., Wang, X., Xiao, W., Wang, Y., Zhang, J., Wang, F., Xu, F., Zeng, W.F., Overall C.M., He, S.M., Chi, H., Xu, P. 2019. Precision and *de novo* Peptide Sequencing Using Mirror Proteases of Ac-LysargiNase and Trypsin for Large-scale Proteomics. *Molecular Cellular Proteomics* 18, 773 – 785

**C) Titles and authors of other HPP relevant papers submitted elsewhere in 2019/2020.**

Quancard, J., Klein, T., Fung, S-Y., Renatus, M., Hughes, N., Israël, L., Priatel, J.J., Kang, S., Blank, M.A., Viner, R.I., Blank, J., Schlapbach, A., Erbel, P., Kizhakkedathu, J., Villard, F., Hersperger, R., Turvey, S.E., Eder, J., Bornancin, F., and Overall, C.M. 2019. An Allosteric MALT1 Inhibitor is a Molecular Corrector Rescuing Function in an Immunodeficient Patient. *Nature Chemical Biology* 15, 304 – 313. doi 10.1038/s41589-018-0222-1

**D) How many PE1-found MPs since HUPO-2017 has your chromosome group reported in papers?**

3

**E) How many PE1-found MPs since HUPO-2019 are now in NeXt-Prot as PE1 proteins? Please check each of your MPs that you reported in the JPR SI.**

3

**F) How many candidate MPs found, but not meeting the guidelines 3.0? (Please state number of peptides identified, their length, and biological replicates found in).**

Many PE1 proteins lacking MS evidence (termed non-MS PE1 proteins) were identified by MS that met the HPP Guidelines for PE1 identification by MS.

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

Developed a precision medicine molecular corrector drug that was proven to restore a mutant protein isoform of MALT1 that untreated led to a rare immunodeficiency disease. The disease was phenotyped in a previous paper by proteomics and TAILS that led to this discovery and then treatment.

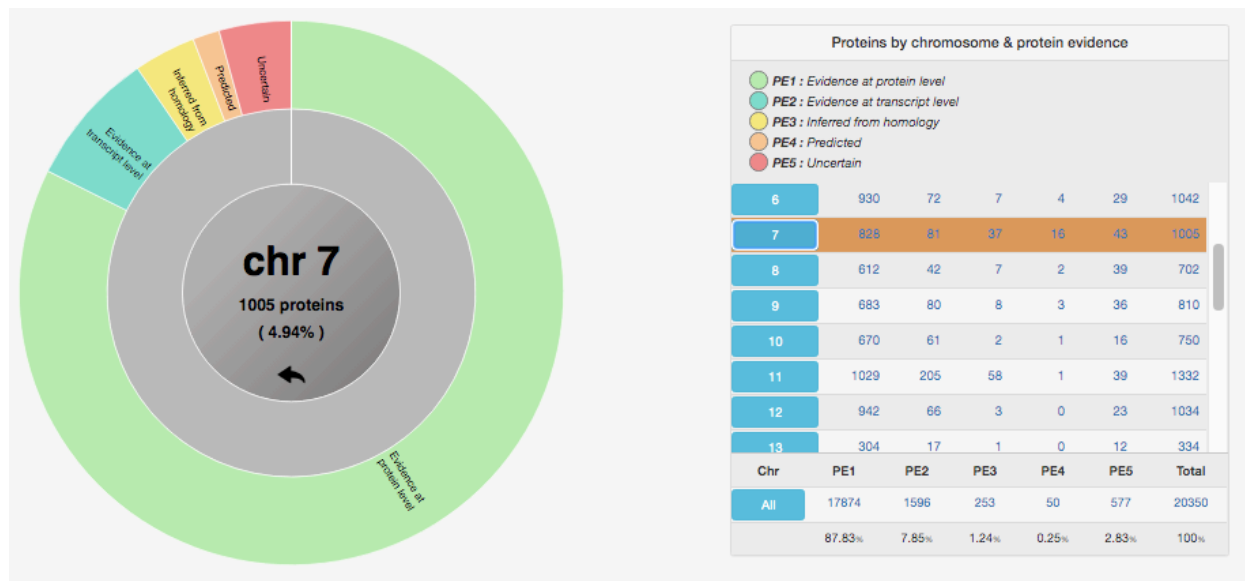
**Chromosome Number: 7**

**PIC Leaders:** Prof Ed Nice

**Major lab members or partners contributing to the neXt-MP50 Challenge:**

At present we are not directly addressing the neXt-MP50 Challenge, although we do have activities addressing missing proteins (e.g. olfactory receptors).

Status of the Chromosome “parts list” for your Chromosome:



**Increase 2019 – 2020: 808 to 828 (see above).** A point of interest: how should we assess exactly where these discoveries have come from. Should we encourage people to mark their PXDs as from a particular C-HPP teams?

B) Titles and authors of papers submitted to JPR.

[Human Proteome Project Mass Spectrometry Data Interpretation Guidelines 3.0.](#)

Deutsch EW, Lane L, Overall CM, Bandeira N, **Baker MS**, Pineau C, Moritz RL, Corrales F, Orchard S, Van Eyk JE, Paik YK, Weintraub ST, Vandenbrouck Y, Omenn GS. *J Proteome Res.* 2019 Dec 6;18(12):4108-4116

[Mass Spectrometry-Based Plasma Proteomics: Considerations from Sample Collection to Achieving Translational Data.](#)

Ignjatovic V, Geyer PE, Palaniappan KK, Chaaban JE, Omenn GS, **Baker MS**, Deutsch EW, Schwenk JM. *J Proteome Res.* 2019 Dec 6;18(12):4085-4097

[Progress on Identifying and Characterizing the Human Proteome: 2019 Metrics from the HUPO Human Proteome Project.](#)

Omenn GS, Lane L, Overall CM, Corrales FJ, Schwenk JM, Paik YK, Van Eyk JE, Liu S, Pennington S, Snyder MP, **Baker MS**, Deutsch EW. *J Proteome Res.* 2019 Dec 6;18(12):4098-4107

[In Silico Peptide Repertoire of Human Olfactory Receptor Proteomes on High-Stringency Mass Spectrometry.](#)

Adhikari S, Sharma S, Ahn SB, **Baker MS**. *J Proteome Res.* 2019 Dec 6;18(12):4117-4123

C) Titles and authors of other HPP relevant papers submitted elsewhere in 2019/2020.

[The omics revolution: beyond genomics. A meeting report.](#)

**Nice EC**. *Clin Proteomics.* 2020 Jan 24;17:1

[Potential early clinical stage colorectal cancer diagnosis using a proteomics blood test panel.](#)

Ahn SB, Sharma S, Mohamedali A, Mahboob S, Redmond WJ, Pascovici D, Wu JX, Zaw T, Adhikari S, Vaibhav V, **Nice EC**, **Baker MS**

[Proteomics Reveals Cell-Surface Urokinase Plasminogen Activator Receptor Expression Impacts Most Hallmarks of Cancer.](#)

Ahn SB, Mohamedali A, Pascovici D, Adhikari S, Sharma S, **Nice EC**, Baker MS. *Proteomics*. 2019 Nov;19(21-22):e1900026.

[Mass spectrometry-based protein identification in proteomics-a review.](#)

Noor Z, Ahn SB, **Baker MS**, Ranganathan S, Mohamedali A. *Brief Bioinform*. 2020 Feb 11:bbz163

[Oncoproteomics: Current status and future opportunities.](#)

He Y, Mohamedali A, Huang C, **Baker MS**, Nice EC. *Clin Chim Acta*. 2019 Aug;495:611-624

[Proteomics and the microbiome: pitfalls and potential.](#)

Lin H, He QY, Shi L, Sleeman M, **Baker MS**, Nice EC. *Expert Rev Proteomics*. 2019 Jun;16(6):501-511

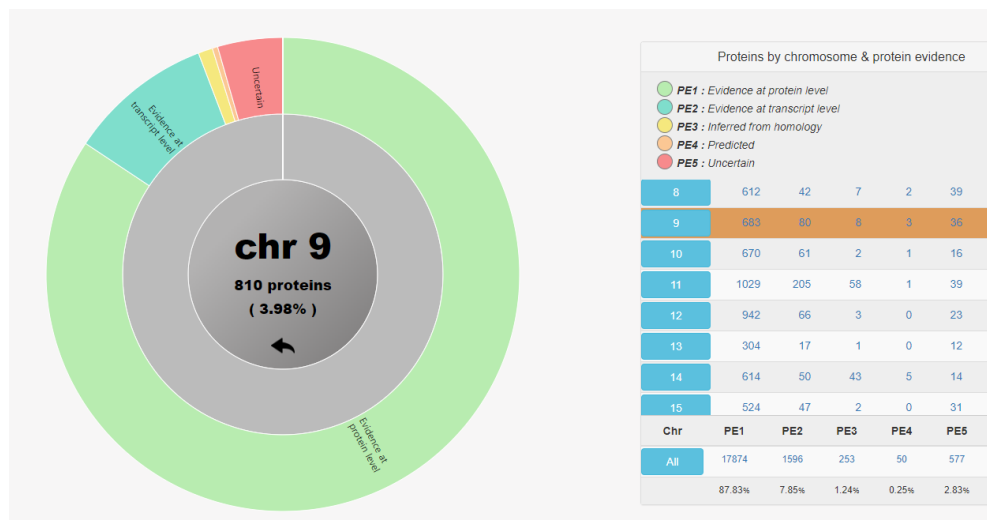
## Chromosome Number: 9

**PIC Leaders:** Je-Yoel Cho

**Major lab members or partners contributing to the neXt-MP50 Challenge**

Soo-Youn Lee, Yong-In Kim, Dong Wook Kim, HuiSu Kim, Hyoung-Min Park, Jinwhan Eugene Lee

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



A) Titles and authors of papers submitted to the 2020 JPR SI or planned.

MS evidence acquisitions of missing proteins in chromosome 9 by using halo tag purification system and identification of their cellular roles. Kim HS, Kim YI, Park HM, Kim DW, and Cho JY

→ Not ready yet

B) Titles and authors of papers published in the 2019 JPR SI.

Common Repository of FBS Proteins (cRFP) To Be Added to a Search Database for Mass Spectrometric Analysis of Cell Secretome, Jihye Shin, Yumi Kwon, Seonjeong Lee, Seungjin Na, Eun Young Hong, Shinyeong Ju, Hyun-Gyo Jung, Prashant Kaushal, Sungho Shin, Ji Hyun Back, Seon Young Choi, Eun Hee Kim, Su Jin Lee, Yae Eun Park, Hee-Sung Ahn, Younghee Ahn, Mohammad Humayun Kabir, Seong-Jun Park, Won Suk Yang, Jeonghun Yeom, Oh Young Bang, Chul-Won Ha, Jin-Won Lee, Un-Beom Kang, Hye-Jung Kim, Kang-Sik Park, J. Eugene Lee, Ji Eun Lee, Jin Young Kim, Kwang Pyo Kim, Youngsoo Kim, Hisashi Hirano, Eugene C. Yi, Je-Yoel Cho, Eunok Paek, and Cheolju Lee

**C) Titles and authors of other HPP relevant papers submitted elsewhere in 2019/2020.**

Gel-based proteomics in disease research: Is it still valuable?

Kim YI, Cho JY.

Biochim Biophys Acta Proteins Proteom. 2019 Jan;1867(1):9-16. doi: 10.1016/j.bbapap.2018.08.001. Epub 2018 Aug 15. Review.

Exploring the key communicator role of exosomes in cancer microenvironment through proteomics

Kim HS, Kim DW, Cho JY

Proteome Sci. Oct 29;17:5. doi: 10.1186/s12953-019-0154-z. eCollection 2019.

**D) How many PE1-found MPs since HUPO-2017 has your chromosome group reported in papers?**

Not ready yet

**E) How many PE1-found MPs since HUPO-2019 are now in NeXt-Prot as PE1 proteins? Please check each of your MPs that you reported in the JPR SI.**

Not identified yet.

**F) How many candidate MPs found, but not meeting the guidelines 3.0? (Please state number of peptides identified, their length, and biological replicates found in).**

We have found 2 MPs which are within the same gene family. Each peptide length identified is 30 and 35 AA long. Due to the sample's rareness, biological replicates were unable to perform. Yet 6 technical replicates were processed by 3 different fraction methods.

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

We are trying to reveal biological function of MPs using human cell line models that express MPs and IP-MS. This strategy is not only useful for neXt-CP50 uPE1 functional characterization project, but also next-MP50 MPs identification and validation project. Five MPs (FOXD4, ARID3C, OR1J1, ANKRD18A, ZNF510) have been turned out its subcellular localization. Two MPs (FOXD4, ARID3C) binding partner proteins were identified via our IP-MS strategy.

**Chromosome Number: 10**

**PIC Leaders:**

PI: Josh LaBaer

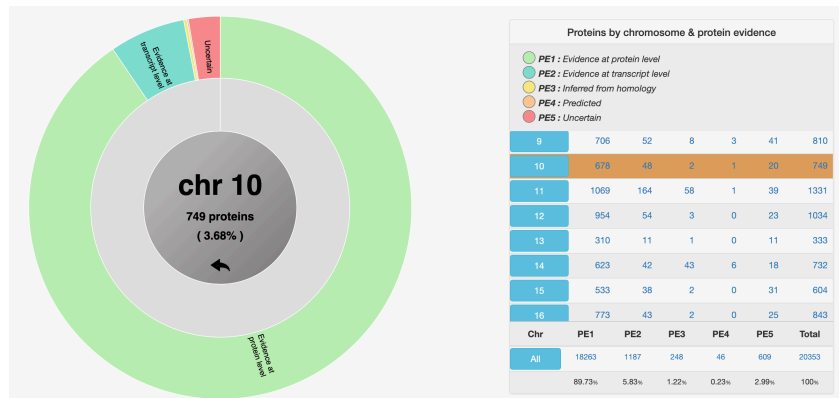
Co-I: Jin Park

**Part I: Missing Proteins: neXt-MP50 Challenge**

**Major lab members or partners contributing to the neXt-MP50 Challenge**

Vel Murugan, Joe Miceli

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



A) Titles and authors of papers submitted to the **2020 JPR SI** or planned.

NA

B) Titles and authors of papers published in the **2019 JPR SI**.

NA

C) Titles and authors of other HPP relevant papers submitted elsewhere in 2019/2020.

NA

D) How many PE1-found MPs since HUPO-2017 has your chromosome group reported in papers?

0

E) How many PE1-found MPs since HUPO-2019 are now in NeXt-Prot as PE1 proteins? Please check each of your MPs that you reported in the JPR SI.

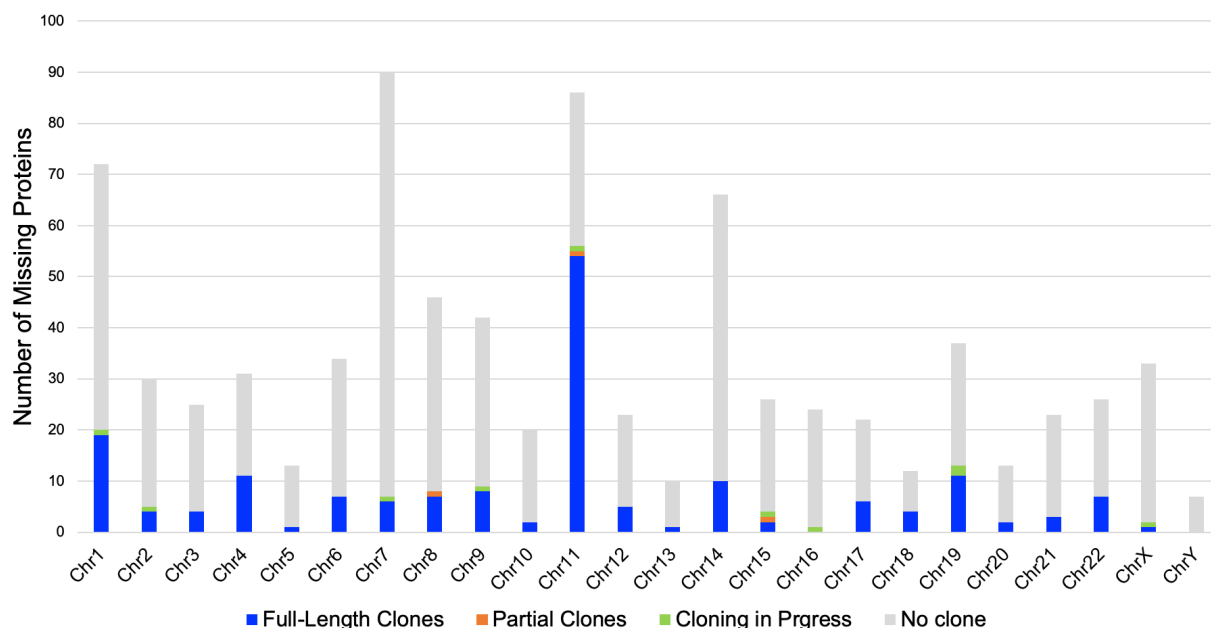
13

F) How many candidate MPs found, but not meeting the guidelines 3.0? (Please state number of peptides identified, their length, and biological replicates found in).

NA

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

As a member of the 5-chromosome consortium of Chr 5, 10, 15, 16, and 19, we have been providing the IVTT-compatible plasmids for missing proteins to other members for IVTT-assisted SRM and continue to generate more plasmids. We have assembled a comprehensive and one of the world's largest collections of full-length Gateway plasmids representing 90% of all human protein-coding genes and are distributing the collection through our repository and distribution web portal DNASU (dnasu.org). Currently, we have full-length plasmids for 175 of 804 missing proteins (shown below), which is available to the entire C-HPP team.



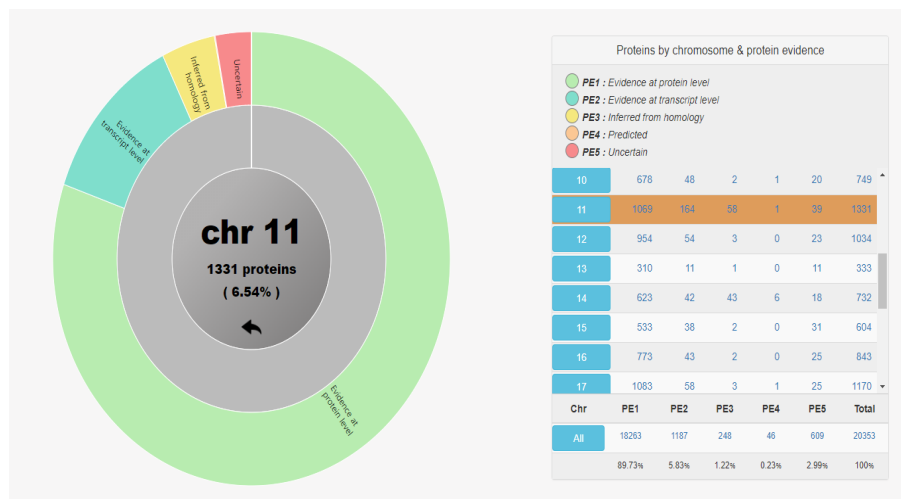
**Chromosome Number: 11**

**PIC Leaders:** Jong Shin Yoo (KBSI)

**Major lab members or partners contributing to the neXt-MP50 Challenge**

- Jin Young Kim (KBSI)
- Heeyoun Hwang (KBSI)
- Kyung Hoon Kwon (KBSI)
- Sang Myung Woo (NCC)
- Yun-Hee Kim (NCC)

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



**A) Titles and authors of papers submitted to the 2020 JPR SI or planned.**

- Bioinformatic Prediction of Gene Ontology Terms of Uncharacterized Proteins from Chromosome 11,

Heeyoun Hwang<sup>1,\*</sup>, Ji Eun Im<sup>2,\*</sup>, Yeji Yang<sup>1</sup>, Hyejin Kim<sup>1,3</sup>, Kyung-Hoon Kwon<sup>1</sup>, Yun-Hee Kim<sup>2,4,\*\*</sup>, Jin Young Kim<sup>1,\*\*</sup>, and Jong Shin Yoo<sup>1,3,\*\*</sup>

**B) Titles and authors of papers published in the 2019 JPR SI.**

1. SAAVpedia: identification, functional annotation, and retrieval of single amino acid variants for proteogenomic interpretation

Soo Youn Lee<sup>1,\*</sup>, Heeyoun Hwang<sup>1,\*</sup>, Young-Mook Kang<sup>2</sup>, Hyejin Kim<sup>1,3</sup>, Dong Geun Kim<sup>1,3</sup>, Ji Eun Jeong<sup>1,3</sup>, Jin Young Kim<sup>1,\*\*</sup> and Jong Shin Yoo<sup>1,3,\*\*</sup>

**C) Titles and authors of other HPP relevant papers submitted elsewhere in 2019/2020.**

1. Classification of Mucin-Type O-Glycopeptides Using Higher-Energy Collisional Dissociation in Mass Spectrometry

Gun Wook Park, Ji Won Lee, Hyun Kyoung Lee, Jong Hwan Shin, Jin Young Kim,\* and Jong Shin Yoo\*

2. Machine Learning Classifies Core and Outer Fucosylation of N-Glycoproteins Using Mass Spectrometry

Heeyoun Hwang, Hoi Keun Jeong, Hyun Kyoung Lee, Gun Wook Park, Ju Yeon Lee, Soo Youn Lee, Young-Mook Kang, Hyun Joo An, Jeong Gu Kang, Jeong-Heon Ko, Jin Young Kim\* & Jong Shin Yoo\*

**D) How many PE1-found MPs since HUPO-2017 has your chromosome group reported in papers?**

A total of 7 MPs was reported in our Chr. Group.

Protein Acc. No.	Peptides	2017-PE	Now-PE	Chr.	Reporting Ref.
P0C7M7	NFNFAADVLDQWSQKEK	PE2	PE1	12	Hwang et al. 2018
	TGERPANPALWWWNGKGDEVK				
	HCLTGGEPLNPEVLEQWR				
P46721	STVLKDDELKTKL	PE2	PE1	12	Hwang et al. 2018
	YGITKDFLPMK				
P59826	IDKDELGKAIQNSLVGEPILQNVLGSVTAVNR	PE2	PE1	20	Hwang et al. 2018
	AIQNSLVGEPILQNVLGSVTAVNR				
	GTPESLFELNSVMTVR				
Q658L1	VTQNALFEGSTEFRESFQPWEIPPEVK	PE2	PE1	15	Hwang et al. 2018
	SSVPFDDVTMYSVEYTPK				
Q8N434	KLSLGAEPQVKEPK	PE2	PE1	7	Hwang et al. 2018
	ALGMGTSGSLCR				
Q16478	LYSAGAGGDAGSAHGGPQR	PE2	PE1	19	Hwang et al. 2017
	SFNYPASLICK				
Q7LC44	QGEPLDQFLWR	PE2	PE1	8	Hwang et al. 2017
	EFLQYSEGTLR				

**E) How many PE1-found MPs since HUPO-2019 are now in NeXt-Prot as PE1 proteins? Please check each of your MPs that you reported in the JPR SI.**

All 7 MPs were changed their status to PE1 level in now in neXtProt.



F) How many candidate MPs found, but not meeting the guidelines 3.0? (Please state number of peptides identified, their length, and biological replicates found in).

N/A

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

For the breakthrough to find out the MP 50, we need to analyse special human samples (e.g. Olfactory Epithelial Tissues) and develop an analytical method (e.g. membrane protein extracting method).

**Chromosome Number: 12**

**PIC Leader:** Ravi Sirdeshmukh

**Part I: Missing Proteins: neXt-MP50 Challenge**

**Major lab members or partners contributing to the neXt-MP50 Challenge**

Ravi Sirdeshmukh (Hari PS, Manoj K Gupta, Mahesh Kulkarni, Srikanth Rapole)

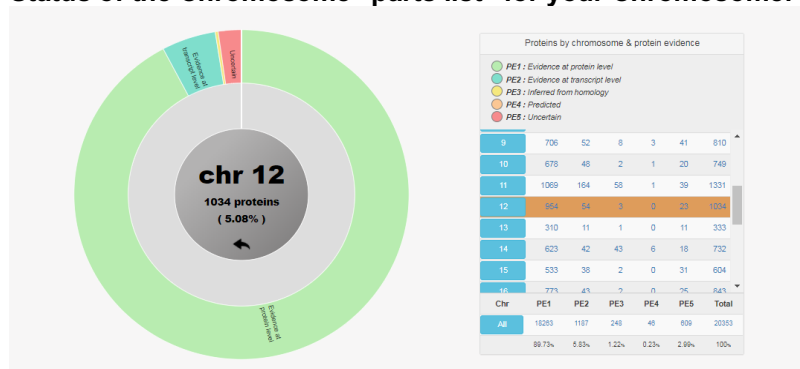
Yuju Chen

Terence Poon

Radislaw Sobota

(Maxey Chung, one of the initial members of the team, informed formally in 2019 that he was unable to continue in the effort; Radislaw Sobota, Singapore is a new member of the team).

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



A) Titles and authors of papers submitted to the 2020 JPR SI or planned.  
No. Work in Progress.

B) Titles and authors of papers published in the 2019 JPR SI.  
No.

C) Titles and authors of other HPP relevant papers submitted elsewhere in 2019/2020.  
No

D) How many PE1-found MPs since HUPO-2017 has your chromosome group reported in papers?

No new MPs published after 2017 (actually after 2014).

We provided first MS evidence for 89 MPs in 2014. Out of which 71 have been entered in neXtProt, as per the recent version (possibly through quality evaluation and support from other relevant publications). 18 are still listed among the MPs. Out of 71, 11 have been added between Jan 2019 version (earlier report) and July 2020 version. This list is given below.

**List of 11 that got added between 2019-2020** →

NX_A8MV81
NX_Q9H1C0
NX_Q6ZN79
NX_Q6IE36
NX_Q5BKT4
NX_Q95626
NX_Q96RD1
NX_O15218
NX_Q7Z769
NX_Q32M45
NX_A6NCE7

Further, we now have some new **unpublished** identifications, which are being reported here. **PI see column F below.** Only overall findings and plan are given here not specifics.

**E) How many PE1-found MPs since HUPO-2019 are now in NeXt-Prot as PE1 proteins? Please check each of your MPs that you reported in the JPR SI.**

20 MPs have got added since Jan 2019 version of neXtProt, out of which 11 were reported by our team in JPR SI (2014).

**F) How many candidate MPs found, but not meeting the guidelines 3.0? (Please state number of peptides identified, their length, and biological replicates found in).**

As indicated in the earlier report, it has been our plan to search for peptides for Chr 12 MPs in some of the public domain cancer proteomic datasets and the deep RNA and proteome datasets for Glioblastoma recently generated in our own lab. Accordingly, we downloaded and re-searched the RNA and protein datasets for multiple cancer types from CPTAC Resource and the Glioblastoma dataset from our lab. The CPTAC datasets particularly one of them yielded peptides for several MPs encoded by Chr 12. The detailed **methodology** is described below (*Italics*).

*MzML files of samples analyzed were downloaded from CPTAC resource (<https://cptac-data-portal.georgetown.edu/cptac/public?scope=Phase+II+%2525>) and converted to mgf files with peakPicking MS2 spectra option using msconvert (3.0.9393). The latest nextprot human database was searched against mgf files of corresponding experiments using three search engines present in SearchGul (v3.3.15) including X!tandem, MSGF+ and Tide. Carbamidomethylation of C was chosen as static modification, while oxidation of M, acetylation of protein N terminal and deamidation of N were chosen as dynamic modifications. Precursor ion tolerance and fragment ion tolerance were set as 10 ppm and 0.05 da, respectively. PeptideShaker (v1.16.40) was used to apply QC filter and proteins corresponding to each MS spectrum. Peptides with 9 amino acids each were selected and FDR of 1% at protein level was used to select peptides. Peptides obtained were matched against protein sequences of missing proteins of chromosome 12. Spectra of individual peptides were extracted from corresponding mgf files and the exported spectra were loaded to PDV (v1.5.1) and evaluated manually. Only MS spectra that qualified with respect to sequence coverage, S/N ratio and ion spread were accepted.*

We identified 15 MP specific peptides, out of which 7 peptides passed the criteria applied. The specific MP peptides identified and their details are not provided here, as they are still being further studied. The MPs mapped included two uncharacterized proteins and others studied but not well characterized.

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

Not directly with regard to MPs.

But we have initiated work on alternatively spliced variants (ASVs) of proteins that include known variants (using public domain database such as SpliceSeq for Glioblastoma) as well as novel variants (using in-house

generated RNA and protein data) using Proteogenomics pipeline validated in-house. Selected ASVs mapping to Chr 12 are planned to be investigated in depth in Glioblastoma.

Out of already known ASVs differentially altered in Glioblastoma (TCGA Data; Splice Seq Resource), we identified 55 ASV events corresponding to 53 genes mapping to Chr 12. Survival analysis carried out revealed 3 candidates (X,Y,Z) to be clinically significant. These variants will be studied in detail. One of them is involved in mitotic activity and is of particular interest.

Using Proteogenomics approach applied to in-house generated deep transcriptomics and proteomics data, we identified over 300 number of Novel ASV peptides across all chromosomes, which also included some new ORFs, LncRNAs and other variations. Of the differentially expressed Novel peptides identified, 19 mapped to Chr 12 encoded genes. From the survival analysis of the novel peptide data, based on significance score, we have selected a membrane protein with growth regulating activity mapping to Chr 12 and a splicing factor mapping to Chr 19, for further detailed investigation (Specific details to be reported in future Reports).

**Chromosome Number: 13**

**PIC Leaders:** Young-Ki Paik

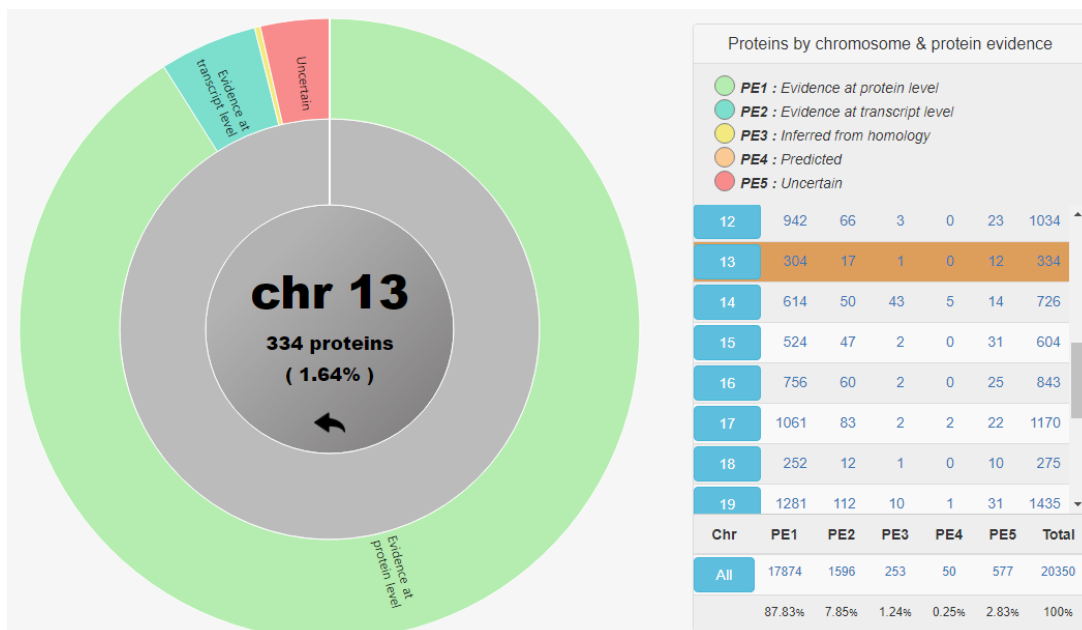
**Part I: Missing Proteins: neXt-MP50 Challenge**

**Major lab members or partners contributing to the neXt-MP50 Challenge**

- Keun-Na (YPRC, Yosei University, Korea)
- Ju-Wan Kim (YPRC, Yosei University, Korea)
- Jin-Young Cho (YPRC, Yosei University, Korea)
- Chae-Yeon Kim (YPRC, Yosei University, Korea)
- Jun-Young Park (YPRC, Yosei University, Korea)

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

*There are still 18 MPs (17 PE2, 1 PE3 and 0 PE4)*



**A) Titles and authors of papers submitted to the 2020 JPR SI or planned.**

- Ju-Wan Kim, Keun-Na, Jin-Young Cho and Young-Ki Paik. Advanced missing protein detection strategy by using sample fractionation and multiple data search (now working)

**B) Titles and authors of papers published in the 2019 JPR SI.**

- Young-Ki Paik, Christopher M Overall, Fernando Corrales, Eric W Deutsch, Lydie Lane, Gilbert S Omenn. Advances in Identifying and Characterizing the Human Proteome. J Proteome Res. 2019 Dec 6;18(12):4079-4084.

**C) Titles and authors of other HPP relevant papers submitted elsewhere in 2019/2020.**

- Chae-Yeon Kim, Keun Na, Saeram Park, Seul-Ki Jeong, Jin-Young Cho, Heon Shin, Min Jung Lee, Gyoonee Han, Young-Ki Paik. FusionPro, a Versatile Proteogenomic Tool for Identification of Novel Fusion Transcripts and Their Potential Translation Products in Cancer Cells. Mol Cell Proteomics. 2019. Aug;18(8):1651-1668

**D) How many PE1-found MPs since HUPO-2017 has your chromosome group reported in papers?**

- 0 protein (see F entry below)

**E) How many PE1-found MPs since HUPO-2019(2019-01-11) are now in NeXt-Prot as PE1 proteins? Please check each of your MPs that you reported in the JPR SI.**

- According to the current version of neXtProt DB (2020-01-17), our Chr13-encoded 10 MPs were promoted to PE1 but we did not claim them as candidate MPs because they were identified as one-hit wonders in our studies (see F entry below).

**F) How many candidate MPs found, but not meeting the guidelines 3.0? (Please state number of peptides identified, their length, and biological replicates found in).**

neXtProt Acc.	Guideline 3.0	PE (2020.1.17)	Comment
NX_Q9BYX7	Not satisfied, One-hit wonder	5	One of two peptides (SSVEKSYELPDGQVITIGNER) has additional mappings with known variants of PE1
NX_O75949	Not satisfied, One-hit wonder	2	
NX_Q32Q52	Not satisfied, One-hit wonder	2	
NX_Q70EL3	Not satisfied, One-hit wonder and nested peptides	2	<u>IIIFHLK</u> and ASISKAPK <u>IIIFHLK</u>
NX_Q8N878	Not satisfied, One-hit wonder	2	
NX_Q8NGC7	Not satisfied, One-hit wonder	2	
NX_Q9NZQ8	Not satisfied, One-hit wonder	2	

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

- None

**Chromosome Number: 14**

**PIC Leaders:** Dr. Charles Pineau (Chair; [charles.pineau@univ-rennes1.fr](mailto:charles.pineau@univ-rennes1.fr))

Dr. Yves Vandenbrouck (co-Chair; [yves.vandenbrouck@cea.fr](mailto:yves.vandenbrouck@cea.fr))

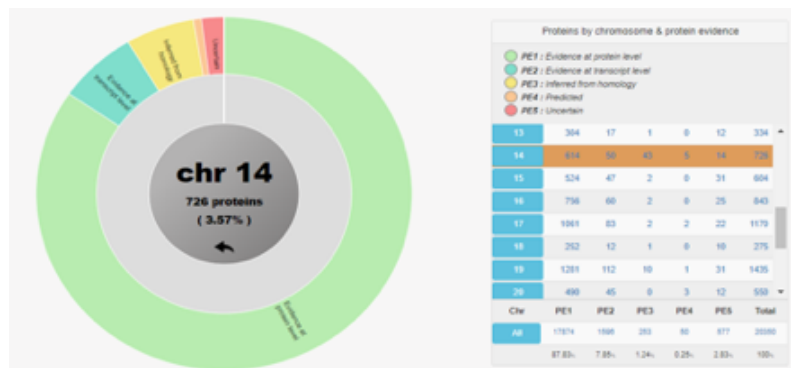
**Part I: Missing Proteins: neXt-MP50 Challenge**

**Major lab members or partners contributing to the neXt-MP50 Challenge**

Univ. Rennes, Inserm, EHESP, Irset (Institut de recherche en santé, environnement et travail) - UMR\_S 1085, F-35042 Rennes cedex, France

Univ. Grenoble Alpes, INSERM, CEA, IRIG-BGE, Health Department, U1038, 38000, Grenoble, France

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



**A) Titles and authors of papers submitted to the 2020 JPR SI or planned.**

Vandenbrouck Y, Pineau C, Lane L. The uncharacterized proteome of human male tissues: a shared resource to uncover new protein functions associated with reproductive biology. *Submitted*

**B) Titles and authors of papers published in the 2019 JPR SI.**

Pineau C, Hikmet F, Zhang C, et al. Cell Type-Specific Expression of Testis Elevated Genes Based on Transcriptomics and Antibody-Based Proteomics. *J Proteome Res.* 2019;18(12):4215-4230. doi:10.1021/acs.jproteome.9b00351

Deutsch EW, Lane L, Overall CM, et al. Human Proteome Project Mass Spectrometry Data Interpretation Guidelines 3.0. *J Proteome Res.* 2019;18(12):4108-4116. doi:10.1021/acs.jproteome.9b00542

**C) Titles and authors of other HPP relevant papers submitted elsewhere in 2019/2020.**

2020: Macron C, Lavigne R, Nunez Galindo A, Affolter M, Pineau C, Dayon L. Exploration of human cerebrospinal fluid: A large proteome dataset revealed by trapped ion mobility time-of-flight mass spectrometry. *Data Brief.* 2020 May 16;31:105704. doi: 10.1016/j.dib.2020.105704

**D) How many PE1-found MPs since HUPO-2017 has your chromosome group reported in papers?**

2017: Carapito C, Duek P, Macron C, et al. Validating Missing Proteins in Human Sperm Cells by Targeted Mass-Spectrometry- and Antibody-based Methods. *J Proteome Res.* 2017;16(12):4340-4351. doi:10.1021/acs.jproteome.7b00374

*12 were identified with two or more peptides and 3 with one peptide after extensive SDS-PAGE fractionation of the two samples and with overall low-intensity signals.*

2018: Melaine N, Com E, Bellaud P, et al. Deciphering the Dark Proteome: Use of the Testis and Characterization of Two Dark Proteins. *J Proteome Res.* 2018;17(12):4197-4210. doi:10.1021/acs.jproteome.8b00387

*Only nine missing proteins were potentially identified by single peptide that pass through the automatic validation, among which 5 MPs were evidenced with one unique peptide (Supplemental Table S2)*

**E) How many PE1-found MPs since HUPO-2019 are now in NeXt-Prot as PE1 proteins? Please check each of your MPs that you reported in the JPR SI.**

N/A

**F) How many candidate MPs found, but not meeting the guidelines 3.0? (Please state number of peptides identified, their length, and biological replicates found in).**

MPs validated by LC-SRM/PRM from Carapito et al. 2017, Vandenbrouck et al., 2016 (see previous report)

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

We would like to highlight the case of CATIP. In 2014, we participated to a study on this protein lead by Lydie Lane and collaborators (Chr2 team). CATIP is a former uPE1 initially called C2orf62 produced in large amounts in the testes of zebrafish, rat and human, that was shown to regulate actin polymerization and ciliogenesis and proposed to be involved in the formation of mature spermatozoa (Bontems et al., PLoS One 2014). Recently a recessive mutation of CATIP was described in humans that may contribute to asthenozoospermia by impairing actin polymerisation and the actin cytoskeleton in sperm (Arafat et al., J Med Genet 2020). It confirmed the study published six years before and the relevance of the approach used to look for function of a uPE1 protein.

**Chromosome Number: 15**

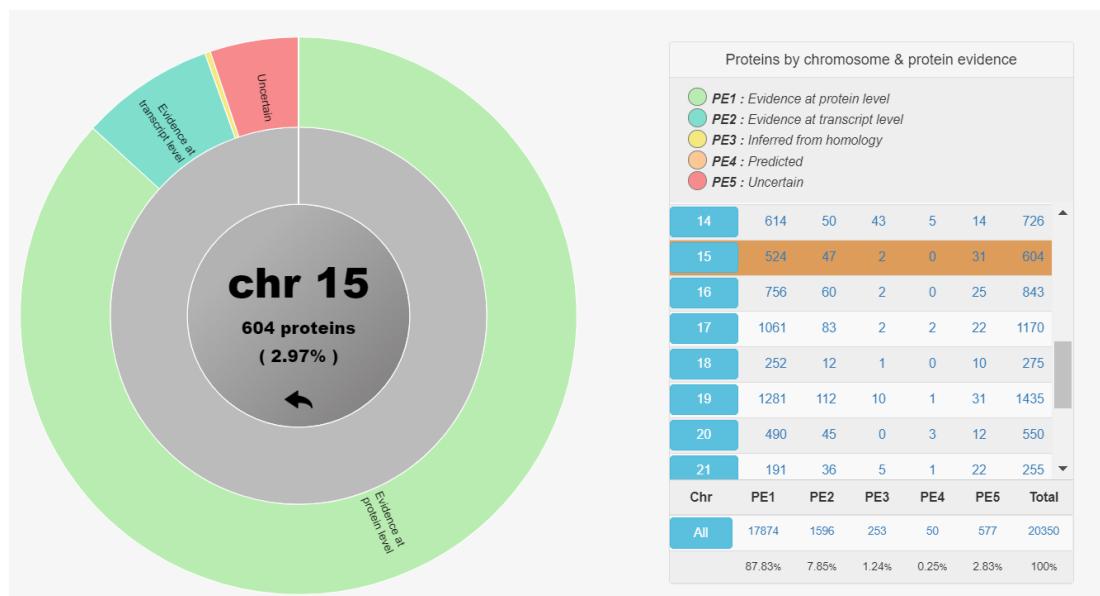
**PIC Gilberto B Domont**

Co-chair: Fabio CS Nogueira

**Major lab members or partners contributing to the neXt-MP50 Challenge**

- Natália P Almeida, UFRJ
- Maurício Quiñones, UFRJ
- Patrícia S Acosta, UFRJ
- Jéssica de S Guedes, UFRJ
- Gustavo Monnerat, UFRJ
- Aniel Sanchez, Lund University
- Gyorgy Marko-Varga, Lund University

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



**A) Titles and authors of papers submitted to the 2020 JPR SI or planned.**

None

**B) Titles and authors of papers published in the 2019 JPR SI.**

1. Quantitative Subcellular Proteomics of the Orbitofrontal Cortex of Schizophrenia Patients.

Velásquez E, Martins-de-Souza D, Velásquez I, Carneiro GRA, Schmitt A, Falkai P, **Domont GB**, Nogueira FCS. *J Proteome Res.* 2019 Dec 6;18(12):4240-4253. doi: 10.1021/acs.jproteome.9b00398.

**C) Titles and authors of other HPP relevant papers submitted elsewhere in 2019/2020.**

2. Novel functional proteins coded by the human genome discovered in metastases of melanoma patients. Sanchez A, Kuras M, Murillo JR, Pla I, Pawlowski K, Szasz AM, Gil J, Nogueira FCS, Perez-Riverol Y, Eriksson J, Appelqvist R, Miliotis T, Kim Y, Baldetorp B, Ingvar C, Olsson H, Lundgren L, Ekedahl H, Horvatovich P, Sugihara Y, Welinder C, Wieslander E, Kwon HJ, **Domont GB**, Malm J, Rezeli M, Betancourt LH, Marko-Varga G. *Cell Biol Toxicol.* 2020 Jun;36(3):261-272. doi: 10.1007/s10565-019-09494-4.

3. Proteomic signatures of brain regions affected by tau pathology in early and late stages of Alzheimer's disease.

Mendonça CF, Kuras M, Nogueira FCS, Plá I, Hortobágyi T, Csiba L, Palkovits M, Renner É, Döme P, Marko-Varga G, **Domont GB**, Rezeli M. *Neurobiol Dis.* 2019 Oct;130:104509. doi: 10.1016/j.nbd.2019.104509.

**D) How many PE1-found MPs since HUPO-2017 has your chromosome group reported in papers?**

Nine. Reference 3 above

**E) How many PE1-found MPs since HUPO-2019 are now in NeXt-Prot as PE1 proteins? Please check each of your MPs that you reported in the JPR SI.**

None. They were reported in *Cell Biol Toxicol.* 2020 Jun;36(3):261-272. doi: 10.1007/s10565-019-09494-4

**F) How many candidate MPs found, but not meeting the guidelines 3.0? (Please state number of peptides identified, their length, and biological replicates found in).**

Nine. All of them meet guideline 3.0 requirements except the validation by S/PRM. The attached Table contains all the information requested.

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

No

**Chromosome Number: 16**

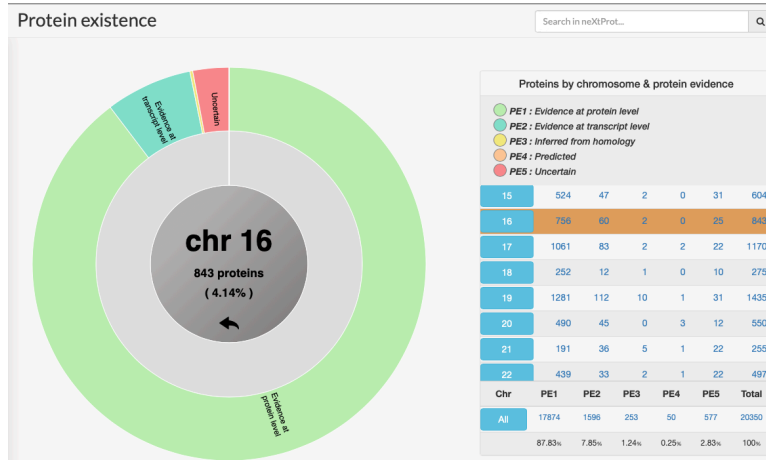
**PIC Leaders:** Fernando J. Corrales, Concha Gil, Francisco Blanco

**Part I: Missing Proteins: neXt-MP50 Challenge**

**Major lab members or partners contributing to the neXt-MP50 Challenge**

CNB-CSIC (FJ Corrales, A Paradela), UCM (Concha Gil), INIBIC (F. Blanco, C Ruiz), CIB-CSIC (I Casal), CBMSO-CSIC (A Marina), CNIO (J. Muñoz), CNIC (J Vázquez), IIBB (J Abian, M Carrscal), PCB (E Oliveira), IRB (M Vilaseca), CIMA (V. Segura), CRG (E Sabido), VHIO (F Canals), UV (MM Sánchez del Pino), CIC-USAL (M Fuentes), CIC bioGUNE (F Elortza), UPV (JM Arizmendi), Navarrabiomed (J Fernández, E Santamaría), IJC (C de la Torre), HNP (ME González Barderas), FJD (G Álvarez Llamas), IACS (I Orera), IPBLN-CSIC (J Sancho).

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



**A) Titles and authors of papers submitted to the 2020 JPR SI or planned.**

UPEFinder: a bioinformatic tool for the study of uncharacterized proteins based on the PageRank algorithm  
 Guillermo Serrano, Jose Gonzalez-Gomariz, Elizabeth Guruceaga, Carlos M. Tilve-Alvarez, Fernando J. Corrales, and Victor Segura

Smelling the dark proteome: Functional Characterization of PITH domain2 containing protein 1 (C1orf128) in olfactory metabolism

Mercedes Lach.n-Montes, Naroa Mendizuri, Karina Ausin, Alberto Perez- Mediavilla, Mikel Azkargorta, Ibon Iloro, Felix Elortza, Isidre Ferrer, Rafael de la Torre, Patricia Robledo, Joaquin Fernandez-Irigoyen, Enrique Santamaria

**B) Titles and authors of papers published in the 2019 JPR SI.**

[Mining the Proteome Associated with Rheumatic and Autoimmune Diseases.](#)

Ruiz-Romero C, Lam MPY, Nilsson P, Önnérjörd P, Utz PJ, Van Eyk JE, Venkatraman V, Fert-Bober J, Watt FE, **Bianco FJ.** J Proteome Res. 2019 Dec 6;18(12):4231-4239. doi: 10.1021/acs.jproteome.9b00360. Epub 2019 Oct 23.

**C) Titles and authors of other HPP relevant papers submitted elsewhere in 2019/2020.**

These are papers more related to B/D initiatives that we coordinate re Liver, Infectious and Rheumatic-autoimmune. Feel free to consider them or not since they are not directly related to C-HPP.

[Serum Exosome Isolation by Size-Exclusion Chromatography for the Discovery and Validation of Preeclampsia-Associated Biomarkers.](#)

Navajas R, **Corrales FJ,** Paradela A. Methods Mol Biol. 2019;1959:39-50. doi: 10.1007/978-1-4939-9164-8\_3.

[Getting insights into hepatocellular carcinoma tumour heterogeneity by multiomics dissection.](#)

Sarobe P, **Corrales F.** Gut. 2019 Nov;68(11):1913-1914. doi: 10.1136/gutjnl-2019-319410. Epub 2019 Aug 2.

[The Human Brain Proteome Project: Biological and Technological Challenges.](#)

Fernández-Irigoyen J, **Corrales F,** Santamaria E. Methods Mol Biol. 2019;2044:3-23. doi: 10.1007/978-1-4939-9706-0\_1.

[Which Low-Abundance Proteins are Present in the Human Milieu of Gamete/Embryo Maternal Interaction?](#)

Canha-Gouveia A, Paradela A, Ramos-Fernández A, Prieto-Sánchez MT, Sánchez-Ferrer ML, **Corrales F,** Coy P. Int J Mol Sci. 2019 Oct 24;20(21):5305. doi: 10.3390/ijms20215305.

[Blockade of the trans-sulfuration pathway in acute pancreatitis due to nitration of cystathionine β-synthase.](#)

Rius-Pérez S, Pérez S, Torres-Cuevas I, Martí-Andrés P, Taléns-Visconti R, Paradela A, Guerrero L, Franco L,



López-Rodas G, Torres L, **Corrales F**, Sastre J. *Redox Biol.* 2020 Jan;28:101324. doi: 10.1016/j.redox.2019.101324. Epub 2019 Sep 8.

[Enrichment of ATP Binding Proteins Unveils Proteomic Alterations in Human Macrophage Cell Death, Inflammatory Response, and Protein Synthesis after Interaction with \*Candida albicans\*.](#)

Vaz C, Reales-Calderon JA, Pitarch A, Vellosillo P, Trevisan M, Hernández ML, Monteoliva L, **Gil C**. *J Proteome Res.* 2019 May 3;18(5):2139-2159. doi: 10.1021/acs.jproteome.9b00032. Epub 2019 Apr 23.

[Trk1-mediated potassium uptake contributes to cell-surface properties and virulence of \*Candida glabrata\*.](#)

Llopis-Torregrosa V, Vaz C, Monteoliva L, Ryman K, Engstrom Y, Gacser A, **Gil C**, Ljungdahl PO, Sychrová H. *Sci Rep.* 2019 May 17;9(1):7529. doi: 10.1038/s41598-019-43912-1.

[Multiomics Assessment of Gene Expression in a Clinical Strain of CTX-M-15-Producing ST131 \*Escherichia coli\*.](#)

Pinto L, Torres C, **Gil C**, Nunes-Miranda JD, Santos HM, Borges V, Gomes JP, Silva C, Vieira L, Pereira JE, Poeta P, Igrejas G. *Front Microbiol.* 2019 May 3;10:831. doi: 10.3389/fmicb.2019.00831. eCollection 2019.

[Analysis of Endogenous Peptides Released from Osteoarthritic Cartilage Unravels Novel Pathogenic Markers.](#)

Fernández-Puente P, González-Rodríguez L, Calamia V, Picchi F, Lourido L, Camacho-Encina M, Oreiro N, Rocha B, Paz-González R, Marina A, García C, **Blanco FJ**, Ruiz-Romero C. *Mol Cell Proteomics.* 2019 Oct;18(10):2018-2028. doi: 10.1074/mcp.RA119.001554. Epub 2019 Jul 27.

[Discovery of an autoantibody signature for the early diagnosis of knee osteoarthritis: data from the Osteoarthritis Initiative.](#)

Camacho-Encina M, Balboa-Barreiro V, Rego-Perez I, Picchi F, VanDuin J, Qiu J, Fuentes M, Oreiro N, LaBaer J, Ruiz-Romero C, **Blanco FJ**. *Ann Rheum Dis.* 2019 Dec;78(12):1699-1705. doi: 10.1136/annrheumdis-2019-215325. Epub 2019 Aug 30.

[Predictive modeling of therapeutic response to chondroitin sulfate/glucosamine hydrochloride in knee osteoarthritis.](#)

**Blanco FJ**, Camacho-Encina M, González-Rodríguez L, Rego-Pérez I, Mateos J, Fernández-Puente P, Lourido L, Rocha B, Picchi F, Silva-Díaz MT, Herrero M, Martínez H, Verges J, Ruiz-Romero C, Calamia V. *Ther Adv Chronic Dis.* 2019 Aug 24;10:2040622319870013. doi: 10.1177/2040622319870013. eCollection 2019.

[Profile of Matrix-Remodeling Proteinases in Osteoarthritis: Impact of Fibronectin.](#)

Pérez-García S, Carrión M, Gutiérrez-Cañas I, Villanueva-Romero R, Castro D, Martínez C, González-Álvaro I, **Blanco FJ**, Juarranz Y, Gomariz RP. *Cells.* 2019 Dec 22;9(1):40. doi: 10.3390/cells9010040.

D) How many PE1-found MPs since HUPO-2017 has your chromosome group reported in papers?

None

E) How many PE1-found MPs since HUPO-2019 are now in NeXt-Prot as PE1 proteins? Please check each of your MPs that you reported in the JPR SI.

F) How many candidate MPs found, but not meeting the guidelines 3.0? (Please state number of peptides identified, their length, and biological replicates found in).

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

**Chromosome Number: 17**

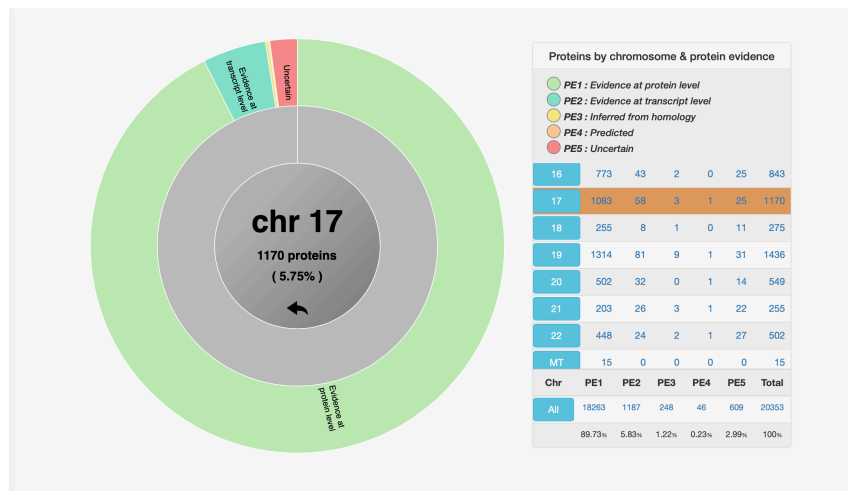
**PIC Leaders:**

Gil Omenn

**Part I: Missing Proteins: neXt-MP50 Challenge**

**Major lab members or partners contributing to the neXt-MP50 Challenge**

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



**A) Titles and authors of papers submitted to the 2020 JPR SI or planned.**

Omenn GS, Lydie L, Overall CM, Cristea IM, Corrales FJ, Lindskog C, Paik YK, van Eyk J, Liu S, Snyder MP, Baker MS, Bandeira N, Aebersold R, Moritz RL, Deutsch EW. HUPO Human Proteome Project reaches major milestone: >90% of predicted human proteins now credibly detected.

**B) Titles and authors of papers published in the 2019 JPR SI.**

Omenn GS, Lydie L, Overall CM, Cristea IM, Corrales FJ, Schwenk JM, Paik YK, van Eyk J, Liu S, Pennington S, Snyder MP, Baker MS, Deutsch EW. J Proteome Res 2019

**C) Titles and authors of other HPP relevant papers submitted elsewhere in 2019/2020.**

**D) How many PE1-found MPs since HUPO-2017 has your chromosome group reported in papers?**

No direct experimental discovery papers.

**E) How many PE1-found MPs since HUPO-2019 are now in NeXt-Prot as PE1 proteins? Please check each of your MPs that you reported in the JPR SI.**

Chr 17 has met the MP50 Challenge: the number of PE2,3,4 missing proteins coded on Chr 17 has been reduced from 148 to 87, meaning that 61 MPs have been detected and incorporated into neXtProt PE1.

**F) How many candidate MPs found, but not meeting the guidelines 3.0? (Please state number of peptides identified, their length, and biological replicates found in).**

N/A

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

N/A

**Chromosome 18**

**PIC Leaders:**

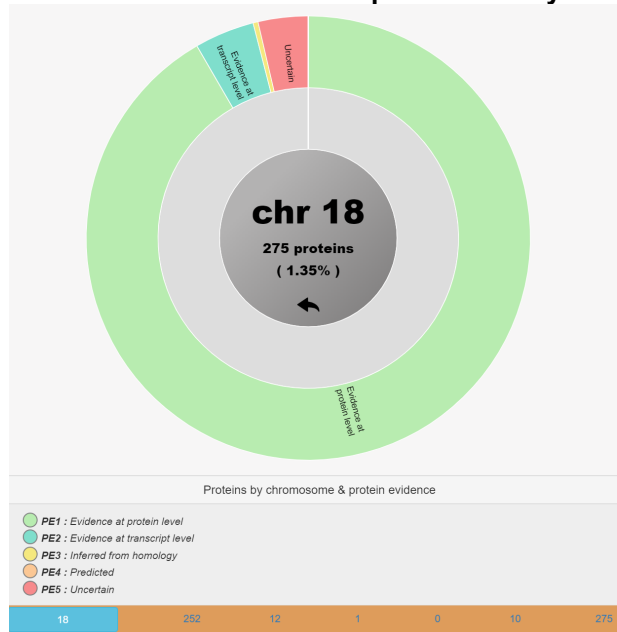
[Alexander Archakov](#)

[Elena Ponomarenko](#) (bioinformatics), [Andrey Lisitsa](#) (standardization)

**Major lab members or partners contributing to the neXt-MP50 Challenge**

Katerina V. Poverennaya, Ekaterina V. Ilgisonis, Olga I. Kiseleva, Sergey P. Radko, Mikhail V. Gorshkov, Stanislav N. Naryzhny, Elena S. Zorina, Svetlana S. Novikova, Arthur T. Kopylov, Olga V. Tikhonova, Eugene N. Nikolaev, Victor G. Zgoda

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



**A) Titles and authors of papers submitted to the 2019 JPR SI and their current status.**

- 1) Manuscript ID: pr-2020-00368c Title: "Human Chr18 Transcriptome using RT-qPCR, Illumina HiSeq, and ONT MinION Technologies Applied to the Same Set of HepG2 Cells and Liver Samples" Author(s): Krasnov, George; Radko, Sergey; Ptitsyn, Konstantin; Shapovalova, Valeriya; Timoshenko, Olga; Khmeleva, Svetlana; Kurbatov, Leonid; Kiseleva, Yana; Ilgisonis, Ekaterina; Pyatnitskiy, Mikhail; Poverennaya, Ekaterina; Kiseleva, Olga; Vakhrushev, Igor; Tsvetkova, Anastasia; Buromski, Ivan; Markin, Sergey; Zgoda, Victor; Archakov, Alexander; Lisitsa, Andrey; Ponomarenko, Elena Manuscript Status: Under revision
- 2) Manuscript ID: pr-2020-003755 Title: "In-depth proteomic analysis of Chr 18 proteins using 2D fractionation" Authors: Vavilov, Nikita; Zgoda, Victor; Tikhonova, Olga; Farafonova, Tatiana; Shushkova, Natalya; Novikova, Svetlana; Yarygin, Konstantin; Ilgisonis, Ekaterina; Ponomarenko, Elena; Lisitsa, Andrey; Archakov, Alexander. Manuscript Status: Submitted

**B) Titles and authors of papers published in the 2019 JPR SI.**

- 1) Challenges of the Human Proteome Project: 10-year Experience of the Russian Consortium. Archakov A.I., Aseev A.L., Bykov V.A., Grigoriev A.I., Govorun V.M., Ilgisonis E.V., Ivanov Y.D., Ivanov V.T., Kiseleva O.I., Kopylov A.T., Lisitsa A.V., Mazurenko S.N., Makarov A.A., Naryzhny S.N., Pleshakova T.O., Ponomarenko E.A., Poverennaya E.V., Pyatnitskiy M.A., Sagdeev R.Z., Skryabin K.G., Zgoda V.G.
- 2) The "Missing" Proteome: Undetected Proteins, Not-Translated Transcripts, and Untranscribed Genes. Radko S.P., Poverennaya E.V., Kurbatov L.K., Ponomarenko E.A., Lisitsa A.V., Archakov A.I.

**C) Titles and authors of other HPP relevant papers submitted elsewhere in 2019/2020.**

- 1) Challenges of the Human Proteome Project: 10-year Experience of the Russian Consortium. Archakov A.I., Aseev A.L., Bykov V.A., Grigoriev A.I., Govorun V.M., Ilgisonis E.V., Ivanov Y.D., Ivanov V.T., Kiseleva O.I., Kopylov A.T., Lisitsa A.V., Mazurenko S.N., Makarov A.A., Naryzhny

S.N., Pleshakova T.O., Ponomarenko E.A., Poverennaya E.V., Pyatnitskiy M.A., Sagdeev R.Z., Skryabin K.G., Zgoda V.G.

- 3) The "Missing" Proteome: Undetected Proteins, Not-Translated Transcripts, and Untranscribed Genes. Radko S.P., Poverennaya E.V., Kurbatov L.K., Ponomarenko E.A., Lisitsa A.V., Archakov A.I.
- 4) Is It Possible to Find Needles in a Haystack? Meta-Analysis of 1000+ MS/MS Files Provided by the Russian Proteomic Consortium for Mining Missing Proteins. Poverennaya E., Kiseleva O., Ilgisonis E., Novikova S., Kopylov A., Ivanov Y., Kononikhin A., Gorshkov M., Kushlinskii N., Archakov A., Ponomarenko E. *Proteomes*, 2020, 8(2), 12. doi:[10.3390/proteomes8020012](https://doi.org/10.3390/proteomes8020012)
- 5) Method for Identification of Threonine Isoforms in Peptides by Ultraviolet Photofragmentation of Cold Ions. Solovyeva E.M., Kopysov V.N., Pereverzev A.Y., Lobas A.A., Moshkovskii S.A., Gorshkov M.V., Boyarkin O.V. *Analytical Chemistry*, 2019, 91(10), 6709-6715. doi:[10.1021/acs.analchem.9b00770](https://doi.org/10.1021/acs.analchem.9b00770)
- 6) Profiling modifications for glioblastoma proteome using ultra-tolerant database search: Are the peptide mass shifts biologically relevant or chemically induced? Tarasova I.A., Chumakov P.M., Moshkovskii S.A., Gorshkov M.V. *Journal of Proteomics*, 2019, 191, 16-21. doi:[10.1016/j.jprot.2018.05.010](https://doi.org/10.1016/j.jprot.2018.05.010)
- 7) Detectability of Plasma Proteins in SRM Measurements. Kiseleva O.I., Ponomarenko E.A., Romashova Yu.A., Poverennaya E.V., Lisitsa A.V. *Current Proteomics*, 2019, 16(1), 74-8. doi:[10.2174/1570164615666180718151135](https://doi.org/10.2174/1570164615666180718151135)
- 8) Inventory of proteoforms as a current challenge of proteomics: Some technical aspects. Naryzhny S. *Journal of Proteomics*, 2019, 191, 22-28. doi: [10.1016/j.jprot.2018.05.008](https://doi.org/10.1016/j.jprot.2018.05.008)

**D) How many PE1-found MPs since HUPO-2017 has your chromosome group reported in papers?**

No PE1-found MPs since HUPO-2017 were detected by Chr18 team.

**E) How many PE1-found MPs since HUPO-2019 are now in NeXt-Prot as PE1 proteins? Please check each of your MPs that you reported in the JPR SI.**

Since HUPO-2019 two (B2RU33 and Q9BXX2) MPs were found at PE1. We confirm.

**F) How many candidate MPs found, but not meeting the guidelines 3.0? (Please state number of peptides identified, their length, and biological replicates found in).**

HSBP1L1 protein (C9JCN9) detection by SRM in progress. Based on RNAseq data we observed a mutation in the sequence corresponding to a unique peptide.

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

We developed the approach to epitranscriptome analysis for prediction of detection on proteome level "missing" proteins. The experimental protocol of 'missing' protein isolation using gene editing is also in progress.

**Chromosome Number: 19**

**PIC Leaders:**

Sergio Encarnación-Guevara

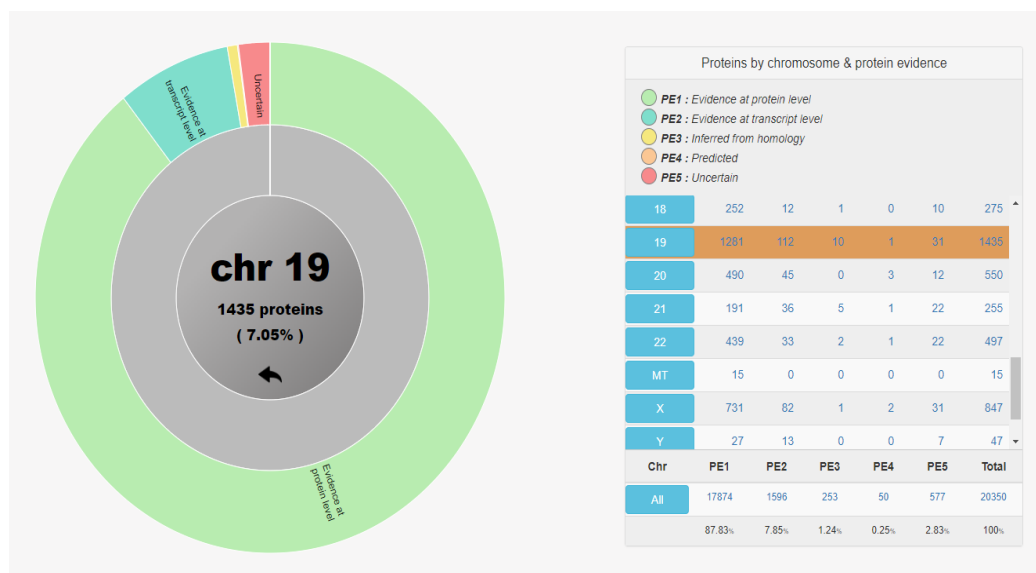
**Major lab members or partners contributing to the neXt-MP50 Challenge**

Orlando Morales-Tarré, Magdalena Hernández-Ortiz, Ramiro Alonso, María del Carmen Vargas-Lagunas.

Proteomics Laboratory at Center for Genomic Sciences UNAM.  
 Jeovanis Gil-Valdes. Lund University, Department of Clinical Sciences  
 Julio Collado-Vides, Program of Computational Biology at Center for Genomic Sciences UNAM.  
 Alejandro García Carranca, National Cancer Institute  
 Osbaldo Resendis Antonio, National Institute of Genomic Medicine.  
 Guadalupe Ayala, National Institute of Public Health

Partners:  
 Fernando J. Corrales. ProteoRed, Spain.

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



A) Titles and authors of papers submitted to the 2020 JPR SI or planned.

None

B) Titles and authors of papers published in the 2019 JPR SI.

None

C) Titles and authors of other HPP relevant papers submitted elsewhere in 2019/2020.

García-Venzor A, Mandujano-Tinoco EA, Lizarraga F, Zampedri C, Krötzsch E, Salgado RM, Dávila-Borja VM, Encarnación-Guevara S, Melendez-Zajgla J, Maldonado V. Microenvironment-regulated lncRNA-HAL is able to promote stemness in breast cancer cells. *Biochim Biophys Acta Mol Cell Res.* 2019 Dec;1866(12):118523. doi: 10.1016/j.bbamcr.2019.118523. Epub 2019 Aug 8.

Gómez-Caudillo L, Ortega-Lozano AJ, Martínez-Batallar ÁG, Rosas-Vargas H, Minauro-Sanmiguel F, Encarnación-Guevara S. Principal component analysis on LC-MS/MS and 2DE-MALDI-TOF in glioblastoma cell lines reveals that mitochondria act as organelle sensors of the metabolic state in glioblastoma. *Oncol Rep.* 2020 Aug;44(2):661-673. doi: 10.3892/or.2020.7625. Epub 2020 May 27.

Sanchez A, Kuras M, Murillo JR, Pla I, Pawlowski K, Szasz AM, Gil J, et al., Novel functional proteins coded by the human genome discovered in metastases of melanoma patients. *Cell Biol Toxicol.* 2020 Jun;36(3):261-272. doi: 10.1007/s10565-019-09494-4. Epub 2019 Oct 10.

D) How many PE1-found MPs since HUPO-2017 has your chromosome group reported in papers?

Three

E) How many PE1-found MPs since HUPO-2019 are now in NeXt-Prot as PE1 proteins? Please check each of your MPs that you reported in the JPR SI.

One reported in Cell Biol Toxicol, none in JPR SI.

F) How many candidate MPs found, but not meeting the guidelines 3.0? (Please state number of peptides identified, their length, and biological replicates found in).

We are still working on the validation through the use of synthetic peptides of the 4 proteins listed in the following table, these proteins were identified in the indicated cell lines, under conditions of inhibition of sirtuin 1 by compound EX527.

ID	Gene name	Peptide	Sample Identified
Q8IVC4	ZNF584	KDALVLHQR	SiHa EX527
		VIQHQDTHSEGKPR	SiHa EX527
P52742	ZNF135	IHTGEKPYKCTQCGR	CaLo CTRL
		IHTGEKPYECNQCGR	SiHa EX527
		IHTGEKPYKCNQCER	SiHa EX527
Q8WXB4	ZNF606	SALTKHER	SiHa EX527
		THTGEKPYR	HaCaT CTRL
		EKPYECGECGKAFR	SiHa EX527
Q9Y2G7	ZFP30	IKSCGLEEQESPHEVCFR	SiHa EX527

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

In a study in progress that analyses acetylated proteins present in cell membranes of breast cancer cell lines, we identified a significant number of missing proteins belonging to different chromosomes, which confirms that by extracting the acylated proteins from the membranes, we access a group of proteins that are normally lost during the extraction process in proteomic studies. These 12 proteins enlist in the next table, were identified using the corresponding peptides noted in the same table, where we can observe the evidence obtained so far. This suggests that an additional step of fractionating the samples could increase the number of missing proteins identified and increase the number of unique peptides in each one. we will continue this approach in the future in order to identify missing proteins belonging to chromosome 19.

Protein name	Unique peptides	Shared peptides	Chromosome	NextProt classification
Kinesin-likeproteinKIF28P	1	0	1	PE3
Ovostatin homolog 2	1	0	12	PE2
Otolin-1	1	0	3	PE2
G-protein coupledreceptor 22	1	0	7	PE2
PutativeRas-relatedproteinRab-1C	1	8	9	PE5
Disrupted in renal carcinomaprotein1	1	0	2	PE2
Spermatogenesis-associatedprotein1	1	0	1	PE2
Putative heat shock protein HSP90-alphaA4	1	4	4	PE5
Putative uncharacterized protein FLJ37218	1	0	9	PE5
Ankyrin repeat domain-containing protein20B	1	1	2	PE2
Putative heat shock protein HSP90-beta4	2	0	15	PE5
Single-pass membrane and coiled-coildomain-containing protein2	1	0	12	PE2

**Chromosome Number: 20**

**PIC Chr 20 Leaders:**

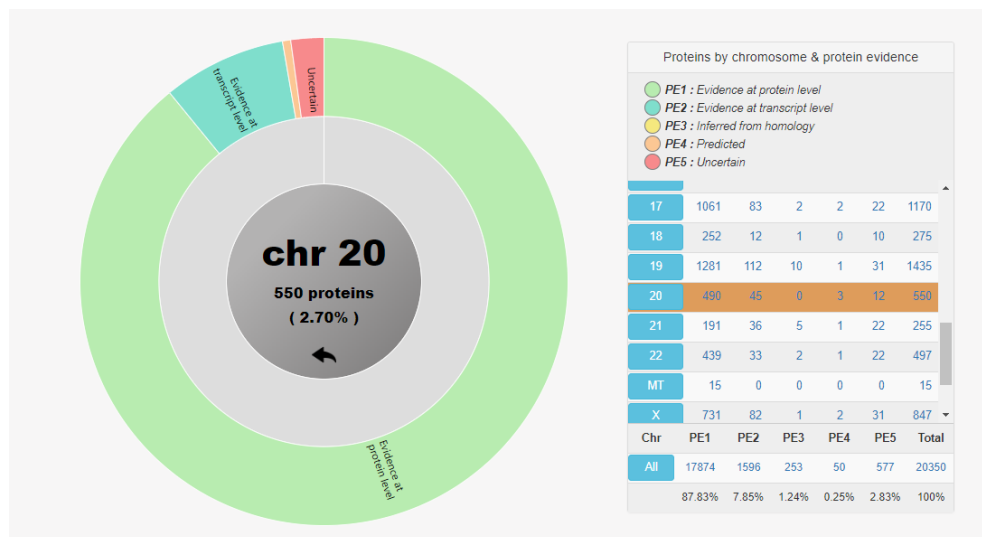
Siqi Liu

**Part I: Missing Proteins: neXt-MP50 Challenge**

**Major lab members or partners contributing to the neXt-MP50 Challenge**

Siqi Liu (BGI-Shenzhen)  
Yan Ren (BGI-Shenzhen)  
Yuanliang Zhang (BGI-Shenzhen)  
Keren Zhang (BGI-Shenzhen)  
Fanyu Bu (BGI-Shenzhen)  
Yifan Tan (BGI-Shenzhen)

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



**A) Titles and authors of papers submitted to the 2020 JPR SI or planned.**

A medulloblastoma stem cell line characterized with metastasis: D283 med is a good choice for MP discovery

**B) Titles and authors of papers published in the 2019 JPR SI.**

Alternative Strategy To Explore Missing Proteins with Low Molecular Weight

Zhilong Lin, Yuanliang Zhang, Huozhen Pan, Piliang Hao, Siqi Li, Yanbin He, Huanming Yang, Siqi Liu\*, and Yan Ren\*

**C) Titles and authors of other HPP relevant papers submitted elsewhere in 2019/2020.**

Exploration of Missing Proteins by a Combination Approach to Enrich the Low-Abundance Hydrophobic Proteins from Four Cancer Cell Lines

Yuanliang Zhang, Zhilong Lin, Yifan Tan, Fanyu Bu, Piliang Hao, Keren Zhang, Huanming Yang, Siqi Liu\*, and Yan Ren\*

**D) How many PE1-found MPs since HUPO-2017 has your chromosome group reported in papers?**

There are 55 MPs were found and checked by PRM with synthesis peptides since 2017.

**E) How many PE1-found MPs since HUPO-2019 are now in NeXt-Prot as PE1 proteins? Please check each of your MPs that you reported in the JPR SI.**

Last year we reported 23 MPs in papers and 20 of them have been PE1-found MPs.

F) How many candidate MPs found, but not meeting the guidelines 3.0? (Please state number of peptides identified, their length, and biological replicates found in).

Too many peptides to be listed here and most of them only have 1 peptide but 2 or more than 2 peptides. we believe the Guidelines V3.0 is essential to filter the peptides with our massive potential peptides in mass spectrum data.

G) Any significant clinical or other successes re a MP that you wish us to consider highlighting in the report.  
None

**Chromosome Number: X**

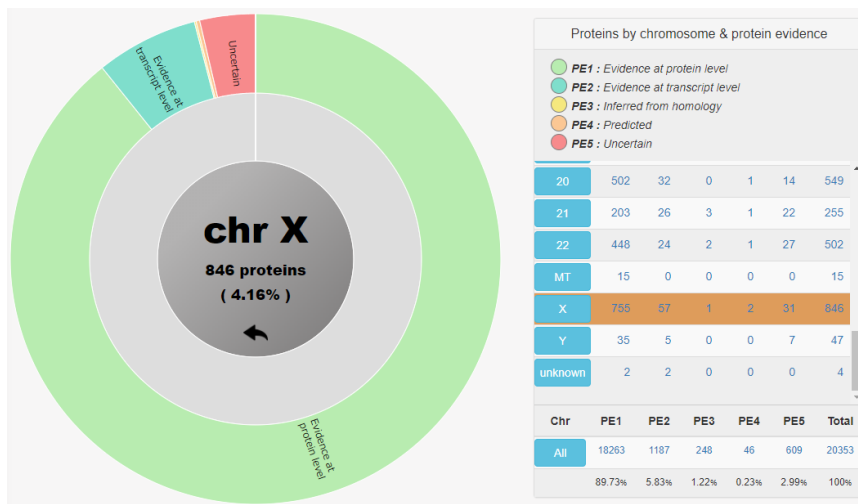
**PIC Leaders:**

Yasushi Ishihama (PI)  
Tadashi Yamamoto (co-PI)

**Major lab members or partners contributing to the neXt-MP50 Challenge**

- **TEAM A:** Yasushi Ishihama, Norie Araki, Susumu Goto, Shin Kawano, Masaki Matsumoto, Shujiro Okuda, Naoyuki Sugiyama, Akiyasu C Yoshizawa, Mio Iwasaki, Tsuyoshi Tabata, Yoshinori Yamanouchi, Yuki Moriya, Daiki Kobayashi, Atsushi Hatano, Yu Watanabe, Tomoyo Takami
- **Team B:** Tadashi Yamamoto, Yoshitoshi Hirao, Tomohiro Uchimoto, Keiko Yamamoto, Yanagita Kengo

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



A) Titles and authors of papers submitted to the 2020 JPR SI or planned.  
No

B) Titles and authors of papers published in the 2019 JPR SI.

Utilization of the Proteome Data Deposited in SRMAtlas for Validating the Existence of the Human Missing Proteins in GPM, Elguoshy A, **Hirao Y**, Yamamoto K, Xu B, Kinoshita N, Mitsui T, **Yamamoto T**. J Proteome Res. 2019 Dec 6;18(12):4197-4205. doi: 10.1021/acs.jproteome.9b00355. Epub 2019

C) Titles and authors of other HPP relevant papers submitted elsewhere in 2019/2020.

Molecular Network Analysis of the Urinary Proteome of Alzheimer's Disease Patients. Watanabe Y, **Hirao Y**, Kasuga K, Tokutake T, Semizu Y, Kitamura K, Ikeuchi T, Nakamura K, **Yamamoto T**. Dement Geriatr Cogn Dis Extra. 2019 Feb 8;9(1):53-65. doi: 10.1159/000496100. eCollection 2019 Jan-Apr. PMID: 31043964



Proteomics Analysis of Urine to Examine Physical Effects of Warm Nano Mist Sauna Bathing. Hirao Y, Kinoshita N, Xu B, Saito S, Quadery AF, Elguoshy A, Yamamoto K, **Yamamoto T.** *Healthcare (Basel)*. 2019 May 19;7(2):71. doi: 10.3390/healthcare7020071.PMID: 31109144

The ProteomeXchange consortium in 2020: enabling 'big data' approaches in proteomics. Deutsch, E. W.; Bandeira, N.; Sharma, V.; Perez-Riverol, Y.; Carver, J. J.; Kundu, D. J.; Garcia-Seisdedos, D.; Jarnuczak, A. F.; Hewapathirana, S.; Pullman, B. S.; Wertz, J.; Sun, Z.; **Kawano, S.; Okuda, S.; Watanabe, Y.**; Hermjakob, H.; MacLean, B.; MacCoss, M. J.; Zhu, Y.; **Ishihama, Y.**; Vizcaino, J. A., *Nucleic Acids Res* **2020**, *48* (D1), D1145-D1152.

D) How many PE1-found MPs since HUPO-2017 has your chromosome group reported in papers?

29 Proteins out of 41 MPs, we found in GPM, are now PE1 in NeXtProt 2019.

E) How many PE1-found MPs since HUPO-2019 are now in NeXt-Prot as PE1 proteins? Please check each of your MPs that you reported in the JPR SI.

35

F) How many candidate MPs found, but not meeting the guidelines 3.0? (Please state number of peptides identified, their length, and biological replicates found in).

sp|Q5T1N1|AKND1\_HUMAN, 2 pepts (length=16, 31), n=1

sp|Q6IC83|CV042\_HUMAN, 1 pept (length=11), n=1

sp|Q7Z570|Z804A\_HUMAN, 2 pepts (length=20, 25), n=1

sp|Q8IVF6|AN18A\_HUMAN, 2 pepts (length=11, 13), n=1

sp|Q8IZA3|H1FOO\_HUMAN, 2 pepts (length=9, 13), n=1

sp|Q96KH6|CR012\_HUMAN, 1 pept (length=16), n=1

sp|Q96LU7|MRFL\_HUMAN, 1 pept (length=11), n=1

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

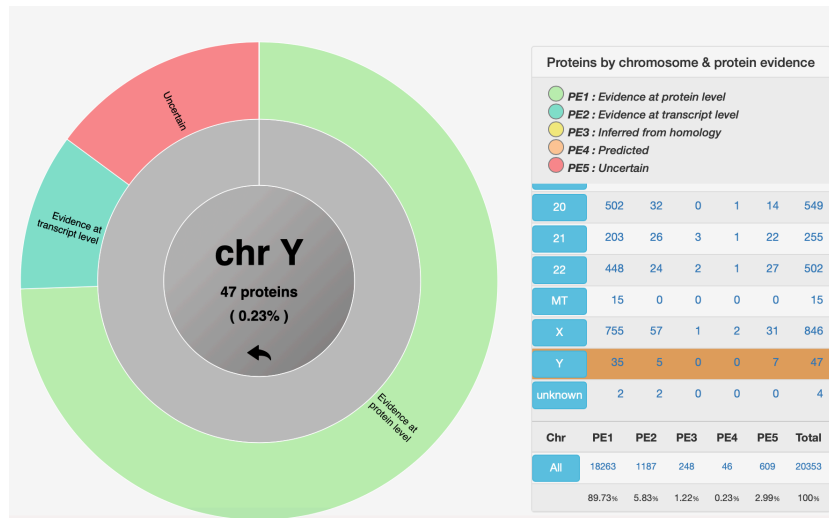
**Chromosome Number: Y**

**PIC Leaders:** Ghasem Hosseini Salekdeh

**Part I: Missing Proteins: neXt-MP50 Challenge**

**Major lab members or partners contributing to the neXt-MP50 Challenge**

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



A) Titles and authors of papers submitted to the 2020 JPR SI or planned.

Human Proteome Project and Human Pluripotent Stem cell: odd bedfellows or a perfect match

B) Titles and authors of papers published in the 2019 JPR SI.

1. Taleahmad S, Alikhani M, Mollamohammadi S, Taei A, Yousefi M, Hassani SN, Baharvand H, Salekdeh GH. Inhibition of Human Y Chromosome Gene, SRY, Promotes Naïve State of Human Pluripotent Stem Cells. J Proteome Res. 2019 18(12):4254-4261
2. Meyfour A, Pahlavan S, Ansari H, Baharvand H, Salekdeh GH. Down-Regulation of a Male-Specific H3K4 Demethylase, KDM5D, Impairs Cardiomyocyte Differentiation. J Proteome Res. 2019 18(12):4277-4282.

C) Titles and authors of other HPP relevant papers submitted elsewhere in 2019/2020.

None

D) How many PE1-found MPs since HUPO-2017 has your chromosome group reported in papers?

One (TBL1Y) + 21 with YuJu Chen's groups in 2018 SI

E) How many PE1-found MPs since HUPO-2019 are now in NeXt-Prot as PE1 proteins? Please check each of your MPs that you reported in the JPR SI.

TBL1Y has been included

F) How many candidate MPs found, but not meeting the guidelines 3.0? (Please state number of peptides identified, their length, and biological replicates found in).

None

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

Although there are a small number of Y chromosome genes, their adequate expression is required to regulate transcription, translation, and protein stability in males, beyond sex-determination. We found that Y chromosome genes are involved in not only maintaining the pluripotency of ESC but also in its differentiation to all three lineages. We have also knocked several Y chromosome genes in ESCs and hope that under better circumstances, will differentiate them to all three developmental lineages and characterize them.

**Chromosome Number: MT**

**PIC Leaders:**

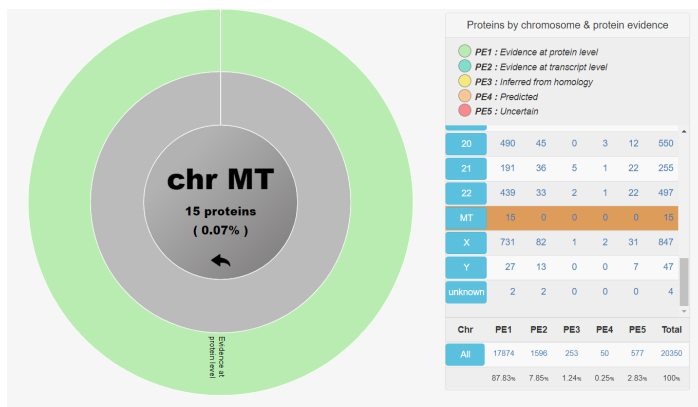
Andrea Urbani, Mauro Fasano, Paola Roncada

**Part I: Missing Proteins: neXt-MP50 Challenge**

**Major lab members or partners contributing to the neXt-MP50 Challenge**

Andrea Urbani

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



A) Titles and authors of papers submitted to the 2020 JPR SI or planned.

NONE

B) Titles and authors of papers published in the 2019 JPR SI.

NONE

C) Titles and authors of other HPP relevant papers submitted elsewhere in 2019/2020.

Mitochondrial Proteins in the Development of Parkinson's Disease.

Zilocchi M, Fasano M, Alberio T.

Adv Exp Med Biol. 2019;1158:17-44. doi: 10.1007/978-981-13-8367-0\_2. PMID: 31452133

Role of Mitochondria in Host-Pathogen Interaction.

Soggiu A, Roncada P, Bonizzi L, Piras C.

Adv Exp Med Biol. 2019;1158:45-57. doi: 10.1007/978-981-13-8367-0\_3. PMID: 31452134

A Tag-Based Affinity Purification Mass Spectrometry Workflow for Systematic Isolation of the Human Mitochondrial Protein Complexes.

Wu Z, Maly R, Moutaoufik MT, Zhang Q, Jessulat M, Babu M.

Adv Exp Med Biol. 2019;1158:83-100. doi: 10.1007/978-981-13-8367-0\_6.

PMID: 31452137

Misconnecting the dots: altered mitochondrial protein-protein interactions and their role in neurodegenerative disorders.

Zilocchi M, Moutaoufik MT, Jessulat M, Phanse S, Aly KA, Babu M.  
Expert Rev Proteomics. 2020 Feb;17(2):119-136. doi: 10.1080/14789450.2020.1723419. Epub 2020 Feb 6.  
PMID: 31986926

Exploring the Impact of PARK2 Mutations on the Total and Mitochondrial Proteome of Human Skin Fibroblasts.  
Zilocchi M, Colugnat I, Lualdi M, Meduri M, Marini F, Corasolla Carregari V, Moutaoufik MT, Phanse S, Pieroni L,  
Babu M, Garavaglia B, Fasano M, Alberio T.  
Front Cell Dev Biol. 2020 Jun 11;8:423. doi: 10.3389/fcell.2020.00423. eCollection 2020.  
PMID: 32596240

Investigating the Role of Mitochondria in Type 2 Diabetes - Lessons from Lipidomics and Proteomics Studies of  
Skeletal Muscle and Liver.  
Kappler L, Kollipara L, Lehmann R, Sickmann A.  
Adv Exp Med Biol. 2019;1158:143-182. doi: 10.1007/978-981-13-8367-0\_9. PMID: 31452140

Exploring the HeLa Dark Mitochondrial Proteome.  
Marini F, Carregari VC, Greco V, Ronci M, Iavarone F, Persichilli S, Castagnola M, Urbani A, Pieroni L.  
Front Cell Dev Biol. 2020 Mar 5;8:137. doi: 10.3389/fcell.2020.00137. eCollection 2020. PMID: 32195257

Crosstalk Between Oxidative Stress and Mitochondrial Damage: Focus on Amyotrophic Lateral Sclerosis.  
Greco V, Longone P, Spalloni A, Pieroni L, Urbani A.  
Adv Exp Med Biol. 2019;1158:71-82. doi: 10.1007/978-981-13-8367-0\_5. PMID: 31452136

Brain mitochondrial proteome alteration driven by creatine deficiency suggests novel therapeutic venues for  
creatine deficiency syndromes.  
Giusti L, Molinaro A, Alessandri MG, Boldrini C, Ciregia F, Lacerenza S, Ronci M, Urbani A, Cioni G, Mazzoni  
MR, Pizzorusso T, Lucacchini A, Baroncelli L.  
Neuroscience. 2019 Jun 15;409:276-289. doi: 10.1016/j.neuroscience.2019.03.030. Epub 2019 Apr 25.  
PMID: 31029731

D) How many PE1-found MPs since HUPO-2017 has your chromosome group reported in papers?  
3

E) How many PE1-found MPs since HUPO-2019 are now in NeXt-Prot as PE1 proteins? Please check each of  
your MPs that you reported in the JPR SI.  
3

F) How many candidate MPs found, but not meeting the guidelines 3.0? (Please state number of  
peptides identified, their length, and biological replicates found in).

Protein name	Accession (Gene name)	Identified in	Chr #	Matching proteotypic Peptides	RNA cell category (HPA)	Main location (HPA)
RAS p21 protein activator 4B	NX_C9J798, RASA4B	U-2 OS, HepG2, SH-SY5Y	7	ELSGGAEAGTVPTSPGK, VVQEEGWFR, DITGSSDPYCIK, VSINNTGLLGSYHPGVFR, AHLGALLSALSR	Cell line enhanced (TPM U-2 OS = 16.6, TPM SH-SY5Y = 14.3)	Localized to the Cell Junctions (uncertain) In addition localized to the Vesicles (uncertain)

RAS p21 protein activator 4	NX_O4 3374, RASA4	U-2 OS, HepG 2, SH- SY5Y	7	ELSGGAEAGTVPTSPGK, VVQEEGWFR, DITGSSDPYCIVK, VSINNTGLLGSYHPGVFR, AHLGALLSALSR	Cell line enhanced (TPM U-2 OS = 23.3, TPM SH- SY5Y = 34.2)	Localized to the Vesicles (uncertain)
60S acidic ribosomal protein P0-like	NX_Q8 NHW5, RPLP0 P6	Hek29 3	2	AFLADPSAFVAAAPVAADT TAAPAAAAAPAK, FLADPSAFVAAAPVAADTT AAPAAAAAPAK	N/A	N/A
Putative keratin-87 protein	NX_A6 NCN2, KRT87P	HepG 2	12	KSDLEANVEALTQEIDFLR, SDLEANVEALTQEIDFLR, KSDLEANVEALTQEIDFLR R, SDLEANVEALTQEIDFLRR, LEANVEALTQEIDFLR, LASELNHVQEVLEGYK	N/A	N/A
PR domain zinc finger protein 13	NX_Q9 H4Q3, PRDM1 3	HeLa	6	LDSGTLPPAVAAAGGTGG GGSGGSGAGKPK, AAGGTGGGGSGGSGAGK PK, VAAAGGTGGGGSGGSGA GKPK	Not detected	Not available
Protein CXorf40B	NX_Q9 6DE9, CXorf40 B	SH- SY5Y	X	LGMTPAQIQALLR, YLTVISNPR, WLEPIPR	Expressed in all (TPM SH- SY5Y = 19.4)	Not available

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

none

**PART II: neXt-CP50**

**Executive Summary of neXt-CP50 Reports**

No	Items	Response
1	Numbers of uPE1 Dark Proteins Under Investigation (in cell)	Chr2: 1 (jointly published), Chr 10: several Chr11: 3, Chr13: 2, Chr 16: Tool Dev, Chr 18: 4, Chr 19: 4, Chr X*: 4 (cellular level)
2	Extension of pilot phase to 2023**	Agreed: 14 (Chr 2, 4, 5, 6, 10, 11, 13, 14, 16, 18, 19, 20, Y, Mt) No response: 3 (Chr 7, 15, X)
3	Suggestions/Comments**	<p><b>1. Joint C- and B/D Works (E. Nice):</b> I assume that this is being discussed by the HPP executive? Much of the data appears to now be generated from B/D-HPP efforts, often from other Chr teams. Since we are moving towards protein <u>function</u>, is it more suited to be a <u>joint</u> C- and B/D-HPP project (Ed Nice).</p> <p><b>RESPONSE:</b> In contrast, to date the uPE1 projects have been initiated by the Chromosome teams within the C-HPP under the neXt-CP50 project lead by Young Ki Paik. We would welcome more input and collaboration with interested B/D-HPP teams.</p> <p><b>2. Resources (J. LaBaer):</b> We are also producing more full-length plasmid clones for uPE1 proteins for functional studies, and the current clone coverage is shown below. Currently, we have full-length plasmids for around 80% of 1,646 uPE1 proteins (shown below) and aim to reach &gt;90% by the end of 2020, which is available to the entire C-HPP team. As shown above in the Question A, we have a full-length plasmid collection for the majority of dark proteins in multiple vectors, which can be applied to many types of experiments for functional characterization of the Dark Proteins. IVTT-produced proteins (GST-tagged) can be used for targeted MS or antibody validation, and the Lenti-based plasmid can be used for cell-based assays or screening. All these are available to the C-HPP team via our web portal DNASU.org, and we are always open to collaboration.</p> <p><b>3. Network (YK Paik):</b> We start considering establishment of a sort of resource network web through which each PI can communicate with others on certain issues and materials. Perhaps, our secretary general can add this extension to the current C-HPP wiki web.</p> <p><b>4. Cell Bank (A. Archakov).</b> The access to a collection of 'clear' cell lines from a cell bank would be an excellent tool for a comprehensive check of protein function.</p> <p><b>5. Collaboration (A Urbani):</b> Development of a joint NDA under the HUPO leadership for data sharing before publication of collected experimental and in silico evidences (e.g., MS data, protein-protein interaction matrices, pQTR/eQTR etc. etc.).</p> <p><b>6. Funding (A Urbani):</b> Lobbying for funding!</p> <p><b>7. (R. Moritz)</b> Develop an international fund for neXt-CP50. A goal for HUPO HEDI to undertake.</p> <p><b>8. (H. Salekdeh)</b> We suggest C-HPP to use pluripotent stem cells (PSC) for identification of MPS and characterisation of uPE1 due to their ability to differentiate into three embryonic germ layers including endoderm, mesoderm, and ectoderm. The importance of this ability is that there are considerable number of proteins, the expression of which are limited to embryonic developmental stages. iPSCs have been widely used to generate patient-specific disease models. The enthusiasm rose higher with the rapid advances in precise DNA editing and CRISPR–Cas9 technology in particular, owing to its simplicity in design and ease of use. Moreover, the generation of organoids that possess part of characteristics of the corresponding in vivo tissue, provides an ideal opportunity for functional analysis of proteins. Integrating CRISPR engineering, hiPSC-derived disease modelling systems, and organoid technologies provides unique platform for C-HPP for</p>

identification of MPs and functional characterization of proteins especially uPE1. C-HPP can collaborate with groups which can generate cell lines and differentiate them.

**9. (G. Omenn)** During the past year there was no net decrease in the 1254 uPE1 proteins. The HPP needs a much-concerted effort to significantly address this Challenge. Chr 17 made a major contribution to the neXt-CP50 Challenge by creating the I-TASSER/COFACTOR function prediction pipeline which neXtProt has adopted as a community service. This pipeline predicted Gene Ontology terms for all 66 Chr 17 uPE1 protein (Zhang C, et al, JPR 2018). Its predictions were put to a blinded test with to-be-released results from neXtProt and from CAFA3 (Zhang C, et al, JPR 2019). neXtProt added a link to facilitate submission of uPE proteins for a report of predicted functions from the group at the University of Michigan. As of 15 May 2020, documentation of requests for C-I-TASSER function predictions showed a total of 561 proteins from 181 users from 35 countries, including 201 neXtProt proteins [<https://zhanglab.ccmb.med.umich.edu/C-I-TASSER/bin/stat.cgi>].

**10. (C. Pineau)** Organize close collaboration with the International Mouse Phenotyping Consortium (IMPC: <https://www.mousephenotype.org>) so as to get real time information on your gene/protein of interest.

**11. (P. Horatovich)** I think we should have initiative to complete the human proteome with more accurate information on proteoforms, protein variants, PTMs and potentially translated new human proteins/peptides.

**12. (Siqi Liu)** We found that a good sample resource is very important for the Dark Protein studies. Recently, we dig more than 10 MPs from the D283 med cell line sourced from brain but derived from metastatic peritoneum. The cell proteins were simply digested and separated into 20 fractions and thus nearly 20 MPs were dig out.

**13. (L. Lane, Y. Vandenbrouck, C. Pineau)** A recessive mutation of CATIP was described in humans that may contribute to asthenozoospermia by impairing actin polymerisation and the actin cytoskeleton in sperm (Arafat et al., J Med Genet 2020). It confirmed a study published six years before and the relevance of the approach by Dr Lydie Lane and collaborators (Chr2 team) used to look for function of a uPE1 protein. There, a former uPE1 initially called C2orf62 produced in large amounts in the testes of zebrafish, rat and human, was shown to regulate actin polymerization and ciliogenesis and proposed to be involved in the formation of mature spermatozoa (Bontems et al., PLoS One 2014).

\*41 uPE1 dark proteins for SRM validation in collaboration with Chr 4 team. \*\*So far, **14 out of 17** teams have agreed on extension of our pilot phase to 2023, while **12** teams have submitted their comments and suggestions on this pilot neXt-CP50 project.

## Chromosome Number: 2

### Part 2: uPE1 Proteins (Dark Proteins): neXt-CP50 Challenge

#### Major lab members or partners contributing to the neXt-CP50 Challenge:

Paula Duek (SIB/University of Geneva)  
 Alain Gateau (SIB/University of Geneva)  
 Camille Mary (University of Geneva)  
 Amos Bairoch (SIB/University of Geneva)

**A) Please list the neXt-CP50 Challenge Proteins that your team is characterising and briefly describe your teams progress made to date including any publications or planned papers in 2020.**

In the C-HPP meeting last year in Saint-Malo, Camille Mary presented her results on **C12orf73**. Because the abstract was made public, we were contacted by Dr. Lena Ho from Singapore who was working on the exact same protein. Finally this led to a joint publication this year!

*Mitochondrial peptide BRAWNIN is essential for vertebrate respiratory complex III assembly.* (2020) Zhang S, Reljić B, Liang C, Kerouanton B, Francisco JC, Peh JH, Mary C, Jagannathan NS, Olexiuk V, Tang C, Fidelito G, Nama S, Cheng RK, Wee CL, Wang LC, Duek Roggli P, Sampath P, Lane L, Petretto E, Sobota RM, Jesuthasan S, Tucker-Kellogg L, Reversade B, Menschaert G, Sun L, Stroud DA, Ho L. *Nat Commun.* 11(1):1312.

Camille has also been working on **C15orf61**, another mitochondrial protein, and on **THEM6**, an uncharacterized enzyme. She is currently preparing the publications.

Sadly, after 10 years of hard work on uPE protein experimental characterization, we have to close our lab in September due to lack of funding. We will continue to assist teams by providing in silico tools and expertise. In particular, we initiated a collaboration with Y-K Paik's team to use *C. elegans* as a model to characterize a selected set of human proteins (grant from ETH Zurich).

Here is the list of uPE proteins that were characterized by our lab before 2020:

*Functional identification of **APIP** as human *mtnB*, a key enzyme in the methionine salvage pathway* (2012) Mary C, Duek P, Salleron L, Tienz P, Bumann D, Bairoch A, Lane L, *PLoS One* 7(12):e52877.

**DERA** is the human deoxyribose phosphate aldolase and is involved in stress response (2014) Salleron L, Magistrelli G, Mary C, Fischer N, Bairoch A, Lane L. *Biochim Biophys Acta.* 1843(12):2913-25.

**C2orf62** and **TTC17** are involved in actin organization and ciliogenesis in zebrafish and human (2014) Bontems E, Fish RJ, Borlat I, Lembo F, Chocu S, Chalmel F, Borg J-P, Pineau C, Neerman-Arbez M, Bairoch A, Lane L, *PLoS One* 9(1):e86476.

**C11orf83**, *a mitochondrial cardiolipin-binding protein involved in bc1 complex assembly and supercomplex stabilization.* (2015) Desmurs M, Foti M, Raemy E, Vaz FM, Martinou JC, Bairoch A, Lane L. *Mol Cell Biol.* 35(7):1139-56.

**B) Your opinion on the extension of next-CP50 (2018-2021).**

Due to the nature of protein characterization work, which requires a lot more extra efforts in combination of in vitro, in vivo and in silico approaches, and COVID-19, we consider that the 3 year pilot phase needs to be extended. For instance, we can extend this to 2023 instead of 2021.

If you agree with this suggestion, please respond by checking "Yes" below. If not, you can add a few comments



as you wish.

Yes ( X ) or/and Your comments: indeed, experimental characterization work takes a lot of time!

C) So far, we have received only two manuscripts on Dark Protein studies. We are wondering when we expect to see your 1<sup>st</sup> submission to the journal (JPR or any). Would it be possible for you to expect approximate date?

D) Any suggestions on the Dark Protein Studies? e.g., resources, reagents, cell lines, ab, ...others.

## Chromosome Number: 4

### Part 2: uPE1 Proteins (Dark Proteins): neXt-CP50 Challenge

#### Major lab members or partners contributing to the neXt-CP50 Challenge:

Reta Birhanu Kitata, Yen-Chen Liao, Yu-Wen Liao, Wen-Hsin Chang Sung-Liang Yu

A) Please list the neXt-CP50 Challenge Proteins that your team is characterising and briefly describe your teams progress made to date including any publications or planned papers in 2020.

#### Functional characterization of missing protein, Ch4-DP1, in lung adenocarcinoma: CH4-DP1 serves as a tumor suppressor

The information about the two uPE1 proteins selected for functional analysis, which were identified by our own from chromosome 4, were further compared with the expression data from paired tumor and adjacent normal tissues from 96 lung adenocarcinoma (LUAD) patients. In 96 lung adenocarcinoma patients, the expression level of CH4-DP1 was down-regulated in tumor tissues compared with adjacent normal tissues, particularly in late stage patients. (early vs late stage,  $P < 0.05$ ), (N0 vs N1-2,  $P = 0.0845$ ). In addition, CH4-DP1 was highly expressed in TP53 mutant patients.  $P = 0.0034$ .

To investigate the role of CH4-DP1 in lung cancer, we analyzed the clinical correlation of CH4-DP1 based on The Cancer Genome Atlas (TCGA) dataset in a cohort of 1,144 LUAD patients. We found that higher CH4-DP1 expression is associated with the favorable overall survival of lung adenocarcinoma patients (long rank  $p = 3.5e-09$ ). Next, to further characterize the functional role of CH4-DP1, we established stable cell lines with CH4-DP1 overexpression. The gene expression profiles revealed 294 differentially expressed genes between CH4-DP1 overexpression and vector control cells by RNA sequencing. Pathway enrichment analysis shows the most significantly ranking pathways in cell adhesion and extracellular matrix remodeling. Several genes such as MMP-A, MMP-B and MMP-C were downregulated and therefore affecting cellular functions such as collagen mediated cell adhesion and motility. In addition, decreased LAMA had effect on cell proliferation, migration and invasion.

Given several invasion/migration pathways enriched in CH4-DP1-overexpressed transfectants and CH4-DP1 associated with better patients' outcome, we suggest that CH4-DP1 might be a metastasis suppressor. The cell mobility was performed by wound healing assay. The results showed that stably expressed CH4-DP1 in CL1-5, PC9 and H1650 cells could inhibit the cell migration ability 20%, 36%, 30%, respectively (mean  $\pm$  SD,  $n = 3$ ). The cell proliferation was performed by MTT assay. Overexpression of CH4-DP1 decreased cell proliferation. For the anchorage-dependent colony formation, control group cells and stable overexpressed CH4-DP1 cells were incubated for 2 weeks, then stained with methylene blue. We observed 40% decrease of growth-suppressive effect on CH4-DP1 overexpression cells .

Taken together, CH4-DP1 expression is associated with favorable survival, inhibition of cell migration and colony formation in vitro and expression analysis also indicates CH4-DP1 significantly alters invasion/migration related pathways. These preliminary results implied that CH4-DP1 acts as a tumor suppressor in lung cancer.

B) Your opinion on the extension of next-CP50 (2018-2021).

Due to the challenge of protein functional characterization and the COVID-19 outbreak, we hope that the 3 year pilot phase can be further extended to another 1-2 year.

If you agree with this suggestion, please respond by checking “Yes” below. If not, you can add a few comments as you wish.

Yes (  ) or/and Your comments:

C) So far, we have received only two manuscripts on Dark Protein studies. We are wondering when we expect to see your 1<sup>st</sup> submission to the journal (JPR or any). Would it be possible for you to expect approximate date? We hope to complete our work and submit a manuscript in 2021

D) Any suggestions on the Dark Protein Studies? e.g., resources, reagents, cell lines, ab, ...others.

### Chromosome Number: 5

#### Part 2: uPE1 Proteins (Dark Proteins): neXt-CP50 Challenge

##### Major lab members or partners contributing to the neXt-CP50 Challenge:

Gyorgy Marko Varga,  
Victor Guryev

A) Please list the neXt-CP50 Challenge Proteins that your team is characterising and briefly describe your teams progress made to date including any publications or planned papers in 2020.

My team do not characterise uPEX proteins but contribute to identification of novel proteins from the human genome both as part of variants or novel translated proteins and peptides by using proteogenomics data integration. Example is peptides mapping to a novel exon of SORBS1 identified only in COPD patient with proteomics and other SAAVs enriched in control or COPD patients. Main new variants were validated with synthetic peptides in this study. We have multiple other proteogenomics projects in progress and we aim to publish and make it public our pipeline as well. In addition to the proteogenomics pipeline, we have an MS1 LC-MS/MS workflow, which can deliver both identified and non-identified peptides (features) matched in samples in a dataset, which facilitate the discovery of new PTMs and peptide/protein variants.

B) Your opinion on the extension of next-CP50 (2018-2021).

Due to the nature of protein characterization work, which requires a lot more extra efforts in combination of in vitro, in vivo and in silico approaches, and COVID-19, we consider that the 3 year pilot phase needs to be extended. For instance, we can extend this to 2023 instead of 2021.

If you agree with this suggestion, please respond by checking “Yes” below. If not, you can add a few comments as you wish.

Yes

C) So far, we have received only two manuscripts on Dark Protein studies. We are wondering when we expect to see your 1<sup>st</sup> submission to the journal (JPR or any). Would it be possible for you to expect approximate date?

We consider our Thorax paper (see in the paper list) containing proteogenomics analysis of human lung tissue to detect novel protein variants and translated proteins as part of the dark proteomics contribution.

D) Any suggestions on the Dark Protein Studies? e.g., resources, reagents, cell lines, ab, ...others.

I think we should have initiative to complete the human proteome with more accurate information on proteoforms, protein variants, PTMs and potentially translated new human proteins/peptides.

### Chromosome Number: 6

#### Part 2: uPE1 Proteins (Dark Proteins): neXt-CP50 Challenge

**Major lab members or partners contributing to the neXt-CP50 Challenge:**

Robert L. Moritz, Christopher M. Overall, Frank Schmid, John Wilson, Eric Deutsch, Cecilia Lindskog, Ulrike Kusebauch

A) Please list the neXt-CP50 Challenge Proteins that your team is characterising and briefly describe your teams progress made to date including any publications or planned papers in 2020.

acc. code	protein name	gene name(s)
<a href="#">NX_O76002</a>	Olfactory receptor 2J2	OR2J2
<a href="#">NX_O95047</a>	Olfactory receptor 2A4	OR2A4
<a href="#">NX_O95918</a>	Olfactory receptor 2H2	OR2H2
<a href="#">NX_P35410</a>	Mas-related G-protein coupled receptor MRG	MAS1L
<a href="#">NX_P58173</a>	Olfactory receptor 2B6	OR2B6
<a href="#">NX_P58182</a>	Olfactory receptor 12D2	OR12D2
<a href="#">NX_Q96KK4</a>	Olfactory receptor 10C1	OR10C1
<a href="#">NX_Q9GZK3</a>	Olfactory receptor 2B2	OR2B2
<a href="#">NX_Q9GZK7</a>	Olfactory receptor 11A1	OR11A1
<a href="#">NX_Q9UGF5</a>	Olfactory receptor 14J1	OR14J1
<a href="#">NX_Q9UGF6</a>	Olfactory receptor 5V1	OR5V1
<a href="#">NX_Q9UGF7</a>	Olfactory receptor 12D3	OR12D3
<a href="#">NX_Q9Y3N9</a>	Olfactory receptor 2W1	OR2W1

We have made extensive searches of disparate datasets and have yet to find evidence of the olfactory receptors as described in this table. We continue to search new extensive datasets utilizing different purification schemes and design experiments

B) Your opinion on the extension of next-CP50 (2018-2021).

The difficulty in protein characterization work requires many different approaches and additional efforts in combination to try and define many of the missing proteins. We believe that the 3-year pilot phase needs to be extended for at least another 3 years to allow for novel approaches and newer technologies to be employed to make an impact on next-CP50.

If you agree with this suggestion, please respond by checking “Yes” below. If not, you can add a few comments as you wish.

Yes (  ) or/and Your comments:

There is a significant lack of funding for these activities

C) So far, we have received only two manuscripts on Dark Protein studies. We are wondering when we expect to see your 1<sup>st</sup> submission to the journal (JPR or any). Would it be possible for you to expect approximate date?

We plan on describing efforts to date with the hope of at least a single identification however, we have not been successful in obtaining a single identification to date for Ch6 missing proteins

D) Any suggestions on the Dark Protein Studies? e.g., resources, reagents, cell lines, ab, ...others.

- 1) Develop an international fund for next-CP50. A goal for HUPO HEDI to undertake.

**Chromosome Number: 7**

**Part 2: uPE1 Proteins (Dark Proteins): neXt-CP50 Challenge**

**Major lab members or partners contributing to the neXt-CP50 Challenge:**

N/A

**B) Your opinion on the extension of next-CP50 (2018-2021).**

I assume that this is being discussed by the HPP executive? Much of the data appears to now be generated from B/D-HPP efforts, often from other Chr teams. Since we are moving towards protein function, is it more suited to be a joint C- and B/D-HPP project

**Chromosome Number: 9**

**Part 2: uPE1 Proteins (Dark Proteins): neXt-CP50 Challenge**

**Major lab members or partners contributing to the neXt-CP50 Challenge:**

Soo-Youn Lee, Yong-In Kim, Dong Wook Kim, HuiSu Kim, Hyoung-Min Park, Jinwhan Eugene Lee

**A) Please list the CP50 Challenge Proteins that your team is characterising and briefly describe your teams progress made to date including any publications or planned papers in 2020.**

We have been trying to characterize CP50 proteins (NIPSNAP3A and TSTD2) using plasmids and an IMPC mouse model. Using plasmids, we will identify cellular roles in human cell lines and the IMPC mouse model, validate cellular roles *in vivo* level and also identify CPs' roles in the mouse model. Now, we are trying to construct the plasmids.

**Chromosome Number: 10**

**Part 2: uPE1 Proteins (Dark Proteins): neXt-CP50 Challenge**

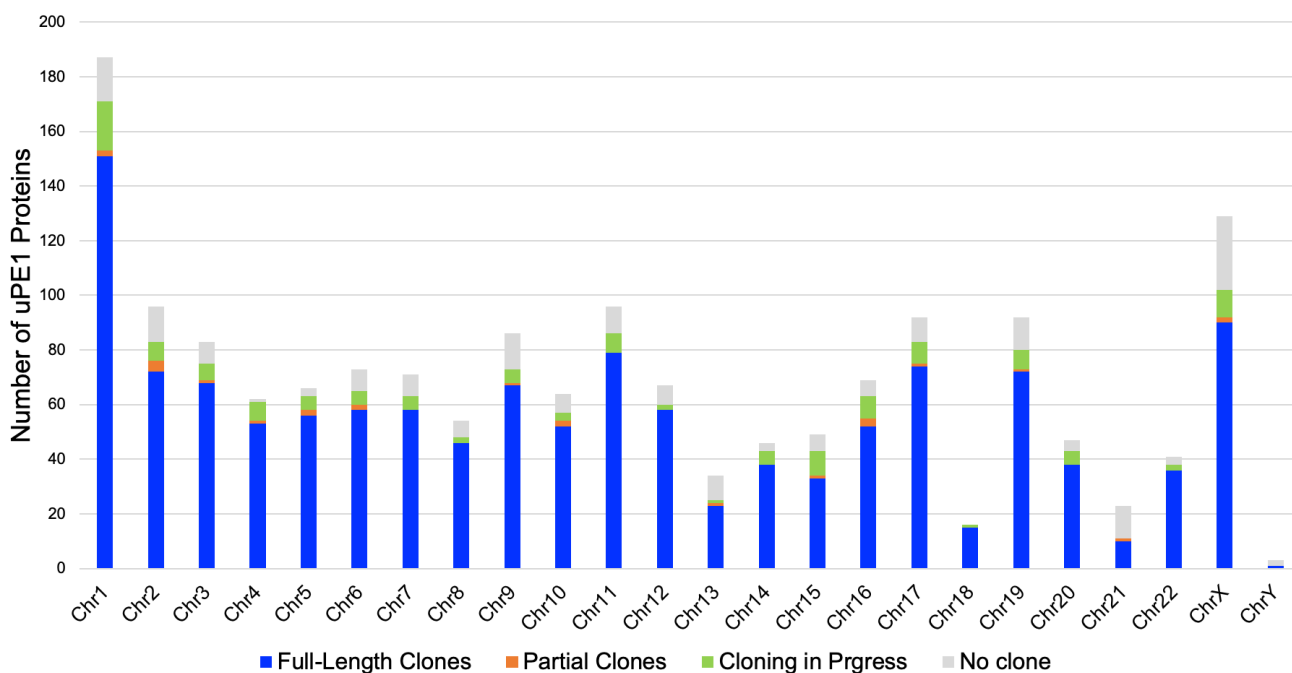
**Major lab members or partners contributing to the neXt-CP50 Challenge:**

Anasuya Pal, Chenxi Xu

**A) Please list the neXt-CP50 Challenge Proteins that your team is characterising and briefly describe your teams progress made to date including any publications or planned papers in 2020.**

We performed genome-wide CRISPR-based function genomics screen to identify mutations that can promote cancer progression, especially invasion, in breast epithelial cells expressing different mutant p53 proteins. From the *in vitro* cell-based screens, a few hundred hits were identified for 2 different p53 mutants, and we are currently down-selecting the top candidates, including several uPE1 proteins, for individual validation. In addition, we performed *in vivo* mouse-based CRISPR screen and identified tumor-initiating mutations. We are aiming to submit the manuscript describing the screening results in combination with RNA-Seq and CHIP-Seq data in 2020.

We are also producing more full-length plasmid clones for uPE1 proteins for functional studies, and the current clone coverage is shown below. Currently, we have full-length plasmids for around 80% of 1,646 uPE1 proteins (shown below) and aim to reach >90% by the end of 2020, which is available to the entire C-HPP team.



**B) Your opinion on the extension of next-CP50 (2018-2021).**

Due to the nature of protein characterization work, which requires a lot more extra efforts in combination of in vitro, in vivo and in silico approaches, and COVID-19, we consider that the 3 year pilot phase needs to be extended. For instance, we can extend this to 2023 instead of 2021.

If you agree with this suggestion, please respond by checking “Yes” below. If not, you can add a few comments as you wish.

Yes ( X ) or/and Your comments:

**C) So far, we have received only two manuscripts on Dark Protein studies. We are wondering when we expect to see your 1<sup>st</sup> submission to the journal (JPR or any). Would it be possible for you to expect approximate date?**  
NA

**D) Any suggestions on the Dark Protein Studies? e.g., resources, reagents, cell lines, ab, ...others.**

As shown above in the Question A, we have a full-length plasmid collection for the majority of dark proteins in multiple vectors, which can be applied to many types of experiments for functional characterization of the Dark Proteins. IVTT-produced proteins (GST-tagged) can be used for targeted MS or antibody validation, and the Lenti-based plasmid can be used for cell-based assays or screening. All these are available to the C-HPP team via our web portal DNASU.org, and we are always open to collaboration.

**Chromosome Number: 11**

**Part 2: uPE1 Proteins (Dark Proteins): neXt-CP50 Challenge**

**Major lab members or partners contributing to the neXt-CP50 Challenge:**

- Jong Shin Yoo (KBSI)
- Jin Young Kim (KBSI)
- Yun-Hee Kim (NCC)
- Heeyoun Hwang (KBSI)

Sang Myung Woo (NCC)

A) Please list the neXt-CP50 Challenge Proteins that your team is characterising and briefly describe your teams progress made to date including any publications or planned papers in 2020.

We have discovered 44 uPE1 proteins from all human chromosomes, which were shown different expression pattern between three subclass of cholangiocarcinoma study. Particularly, three out of five candidates coded in chromosome 11 have used for functional study on progress.

B) Your opinion on the extension of next-CP50 (2018-2021).

Due to the nature of protein characterization work, which requires a lot more extra efforts in combination of in vitro, in vivo and in silico approaches, and COVID-19, we consider that the 3 year pilot phase needs to be extended. For instance, we can extend this to 2023 instead of 2021.

If you agree with this suggestion, please respond by checking “Yes” below. If not, you can add a few comments as you wish.

Yes (  ) or/and Your comments:

C) So far, we have received only two manuscripts on Dark Protein studies. We are wondering when we expect to see your 1<sup>st</sup> submission to the journal (JPR or any). Would it be possible for you to expect approximate date?

We have submitted a paper in this issue in which we suggest a new method for selection of GO terms from the iTASSER/COFACTOR result and we validated the cellular composition of the three uPE1 from chromosome 11 using cellular expression system.

D) Any suggestions on the Dark Protein Studies? e.g., resources, reagents, cell lines, ab, ...others.

I think it is time to encourage for paper submission of uPE1 studies.

## Chromosome Number: 12

### Part 2: uPE1 Proteins (Dark Proteins): neXt-CP50 Challenge

#### Major lab members or partners contributing to the neXt-CP50 Challenge:

Ravi Sirdeshmukh (Sanjeev Shukla)

A) Please list the neXt-CP50 Challenge Proteins that your team is characterising and briefly describe your teams progress made to date including any publications or planned papers in 2020.

Work likely to be initiated in the succeeding period.

B) Your opinion on the extension of next-CP50 (2018-2021).

Due to the nature of protein characterization work, which requires a lot more extra efforts in combination of in vitro, in vivo and in silico approaches, and COVID-19, we consider that the 3 year pilot phase needs to be extended. For instance, we can extend this to 2023 instead of 2021.

If you agree with this suggestion, please respond by checking “Yes” below. If not, you can add a few comments as you wish.

Yes (  )

C) So far, we have received only two manuscripts on Dark Protein studies. We are wondering when we

expect to see your 1<sup>st</sup> submission to the journal (JPR or any). Would it be possible for you to expect approximate date?

The work is yet to be initiated.

D) Any suggestions on the Dark Protein Studies? e.g., resources, reagents, cell lines, ab, ...others.

For any Dark Protein(S) identified on the basis peptides and selected for characterization, we want to first confirm its cellular existence in the full form and then subcellular localization. Other functional and structural studies will follow only after this confirmation.

### Chromosome Number: 13

#### Part 2: uPE1 Proteins (Dark Proteins): neXt-CP50 Challenge

##### Major lab members or partners contributing to the neXt-CP50 Challenge:

Keun-Na (YPRC, Yosei University, Korea)  
 Ju-Wan Kim (YPRC, Yosei University, Korea)  
 Jin-Young Cho (YPRC, Yosei University, Korea)  
 Chae-Yeon Kim (YPRC, Yosei University, Korea)  
 Jun-Young Park (YPRC, Yosei University, Korea)

A) Please list the CP50 Challenge Proteins that your team is characterising and briefly describe your teams progress made to date including any publications or planned papers in 2020.

Table 1

Nick name of uPE1s	Status	Nick name of uPE1	Status
13DP1	CRISPR/cas9 mutant was successfully constructed for in vivo study in model animals <i>C. elegans</i> .	13DP6	Cancelled due to publication by others during preparation of mutant construct.
13DP2	Lower priority	13DP7	Lower priority
13DP3	Lower priority	13DP8	Lower priority
13DP4	Failure on the CRISPR/cas9 mutant construction due to multiple isoforms	13DP9	Cancelled due to publication by others during preparation of mutant construct.
13DP5	Lower priority	13DP10	Lower priority

B) Your opinion on the extension of next-CP50 (2018-2021).

Due to the nature of protein characterization work, which requires a lot more extra efforts in combination of in vitro, in vivo and in silico approaches, we recently realized that 3 year pilot phase needs to be extended. For instance, we can extend this to 2023 instead of 2021. In particular, since COVID-19 has influenced many labs and scientists around world, this action is really needed to reflect such unexpected circumstances.

What do you think? If you agree with this suggestion, please respond by checking “Yes” below. If not, you can add a few comments as you wish.

Your Response: Yes (  ) or/and Your comments:

- Covid-19 appears to negatively influence our work in many areas among which the purchase of reagents (antibody, assay kit, etc.) turned out to be the most problematic. Due to the restriction of travel and transportation

in each country, we need to wait for much longer period of time to get the reagents (e.g., 6-8 weeks instead of 1 or 2 weeks) for our work.

- Given that our functional study of uPE1 usually requires many established cell lines, transgenic and knock-out animals, we should consider extension of our pilot study, neXt-CP50. Even more, the pandemic staycation really affects our research negatively in many areas. Thus, I would like to suggest extend the original term, 3 years, to 6 years (i.e., 2018-2024).

C) So far, we have received only two manuscripts on Dark Protein studies. We are wondering when we expect to see your 1<sup>st</sup> submission to the journal (JPR or any). Would it be possible for you to expect approximate time? (e.g., July 2021).

Your Response: ~June 2022

D) Any suggestions on the Dark Protein Studies? e.g., resources, reagents, cell lines, ab, ...others.

We start considering establishment of a sort of resource network web through which each PI can communicate with others on certain issues and materials. Perhaps, our secretary general can add this extension to the current C-HPP wiki web.

## Chromosome Number: 14

### Part 2: uPE1 Proteins (Dark Proteins): neXt-CP50 Challenge

#### Major lab members or partners contributing to the neXt-CP50 Challenge:

Nathalie Melaine (Protim, Inserm, University of Rennes)  
Emmanuelle Com (Protim, Inserm, University of Rennes)  
Thomas Fréour (Reproductive Medicine unit, Nantes Hospital)

#### Major Partners :

Lydie Lane (SIB/University of Geneva)  
Cecilia Lindskog-Bergström (Human Protein Atlas, Uppsala, Sweden)

A) Please list the neXt-CP50 Challenge Proteins that your team is characterising and briefly describe your teams progress made to date including any publications or planned papers in 2020.

14DP1

B) Your opinion on the extension of next-CP50 (2018-2021).

Due to the nature of protein characterization work, which requires a lot more extra efforts in combination of in vitro, in vivo and in silico approaches, and COVID-19, we consider that the 3 year pilot phase needs to be extended. For instance, we can extend this to 2023 instead of 2021.

If you agree with this suggestion, please respond by checking “Yes” below. If not, you can add a few comments as you wish.

**Yes**

C) So far, we have received only two manuscripts on Dark Protein studies. We are wondering when we expect to see your 1<sup>st</sup> submission to the journal (JPR or any). Would it be possible for you to expect approximate date?

Summer 2021

D) Any suggestions on the Dark Protein Studies? e.g., resources, reagents, cell lines, ab, ...others.



Organize close collaboration with the International Mouse Phenotyping Consortium (IMPC): <https://www.mousephenotype.org> so as to get real time information on your gene/protein of interest.

### Chromosome Number: 15

#### Part 2: uPE1 Proteins (Dark Proteins): neXt-CP50 Challenge

##### Major lab members or partners contributing to the neXt-CP50 Challenge:

###### PIC Gilberto B Domont

Co-chair: Fabio CS Nogueira

Natália P Almeida, UFRJ

Maurício Quiñones, UFRJ

Patrícia S Acosta, UFRJ

Jéssica de S Guedes, UFRJ

Gustavo Monnerat, UFRJ

Aniel Sanchez, Lund University

Gyorgy Marko-Varga, Lund University

A) Please list the CP50 Challenge Proteins that your team is characterising and briefly describe your teams progress made to date including any publications or planned papers in 2020.

None

### Chromosome Number: 16

#### Part 2: uPE1 Proteins (Dark Proteins): neXt-CP50 Challenge

##### Major lab members or partners contributing to the neXt-CP50 Challenge

Víctor Segura (CIMA), Fernando Corrales (CNB-CSIC).

A) Please list the neXt-CP50 Challenge Proteins that your team is characterising and briefly describe your teams progress made to date including any publications or planned papers in 2020.

UPEFinder: a bioinformatic tool for the study of uncharacterized proteins based on the PageRank algorithm

Paper submitted to the 2020 JPR SI.

B) Your opinion on the extension of next-CP50 (2018-2021).

Due to the nature of protein characterization work, which requires a lot more extra efforts in combination of in vitro, in vivo and in silico approaches, and COVID-19, we consider that the 3 year pilot phase needs to be extended. For instance, we can extend this to 2023 instead of 2021.

If you agree with this suggestion, please respond by checking “Yes” below. If not, you can add a few comments as you wish.

Yes (  ) or/and Your comments:

C) So far, we have received only two manuscripts on Dark Protein studies. We are wondering when we expect to see your 1<sup>st</sup> submission to the journal (JPR or any). Would it be possible for you to expect approximate date?

Already submitted

D) Any suggestions on the Dark Protein Studies? e.g., resources, reagents, cell lines, ab, ...others.

## Chromosome Number: 17

### Part 2: uPE1 Proteins (Dark Proteins): neXt-CP50 Challenge

#### Major lab members or partners contributing to the neXt-CP50 Challenge:

Chr 17 made a major contribution to the neXt-CP50 Challenge by creating the I-TASSER/COFACTOR function prediction pipeline which neXtProt has adopted as a community service. This pipeline predicted Gene Ontology terms for all 66 Chr 17 uPE1 protein (Zhang C, et al, JPR 2018). Its predictions were put to a blinded test with to-be-released results from neXtProt and from CAFA3 (Zhang C, et al, JPR 2019). neXtProt added a link to facilitate submission of uPE proteins for a report of predicted functions from the group at the University of Michigan. As of 15 May 2020, documentation of requests for C-I-TASSER function predictions showed a total of 561 proteins from 181 users from 35 countries, including 201 neXtProt proteins [<https://zhanglab.ccmb.med.umich.edu/C-I-TASSER/bin/stat.cgi>].

**A) Please list the neXt-CP50 Challenge Proteins that your team is characterising and briefly describe your teams progress made to date including any publications or planned papers in 2020.**

To protect your IP you may wish to not disclose the protein ID. In this case please use this abbreviation to designate your target uPE1: Ch-DPx (Ch add number for Ch team, For Chr X, Y and Mt, = XDP1, YDP1 or MtDP1; DP, dark protein; x 1,2,3 etc)

Examples:

ATXN8 (Chr 13), CCDC70 (Chr 13)...

► 13DP1 for ATXN8, 13DP2 for CCDC70 and so on.

**B) Your opinion on the extension of next-CP50 (2018-2021).**

Due to the nature of protein characterization work, which requires a lot more extra efforts in combination of in vitro, in vivo and in silico approaches, and COVID-19, we consider that the 3 year pilot phase needs to be extended. For instance, we can extend this to 2023 instead of 2021.

If you agree with this suggestion, please respond by checking “Yes” below. If not, you can add a few comments as you wish.

Yes ( ) or/and Your comments:

During the past year there was no net decrease in the 1254 uPE1 proteins. The HPP needs a much concerted effort to significantly address this Challenge.

**C) So far, we have received only two manuscripts on Dark Protein studies. We are wondering when we expect to see your 1<sup>st</sup> submission to the journal (JPR or any). Would it be possible for you to expect approximate date?**

N/A

**D) Any suggestions on the Dark Protein Studies? e.g., resources, reagents, cell lines, ab, ...others.**

## Chromosome Number: 18

### Major lab members or partners contributing to the neXt-CP50 Challenge:

Ekaterina V. Ilgisonis, Ekaterina V.Poverennaya, Mikhail A.Pyatnitskii, Olga I. Kiseleva, Elena A. Ponomarenko

**A) Please list the neXt-CP50 Challenge Proteins that your team is characterising and briefly describe your teams progress made to date including any publications or planned papers in 2020.**

Currently, 10 genes encoding uPE1 proteins are localized on chromosome 18. Among them, for 4 gene we predicted the GO category by analyzing the data obtained within the BioPlex project (see Table 1, *Poverennaya E., Kiseleva O., Romanova A., Pyatnitskiy M., Predicting Functions of Uncharacterized Human Proteins: From Canonical to Proteoforms, Genes, 2020, 1(6), 677*).

Also, the different protocols of function annotation were developed (*Savosina P., Karasev D., Veselovsky A., Miroshnichenko Yu., Sobolev B., Functional and structural features of proteins associated with alternative splicing, International Journal of Biological Macromolecules, 2020, 147, 513-520; Ershov P.V., Mezentsev Yu.V., Yablokov E.O., Kaluzhskiy L.A., Florinskaya A.V., Gnedenko O.V., Zgoda V.G., Vakhrushev I.V., Raeva O.S., Yarygin K.N., Gilep A.A., Usanov S.A., Medvedev A.E., Ivanov A.S., Direct Molecular Fishing of Potential Partners of Protein-Protein Interactions in the HepG2 Cells Lysate That Involve Proteins Encoded by Genes of Human Chromosome 18, Russian Journal of Bioorganic Chemistry, 2018, 44(6), 759-768*).

For five of ten uPE1 proteins the experimental evidence of function annotation in progress.

**Table 1. GO term prediction for uPE1 proteins coded by Chr18.**

#	Gene	AC	GO category		
			biological process	molecular function	cellular component
1	POTEC	B2RU33	-	GO:0019838 - growth factor binding; GO:0019894 - kinesin binding; GO:0030742 - GTP-dependent protein binding; GO:0042169 - SH2 domain binding	GO:0005901 - caveola
2	C18orf21	Q32NC0	-	GO:0005515 - protein binding	
3	CCDC102B	Q68D86	GO:0008284 - positive regulation of cell proliferation; GO:0030335 - positive regulation of cell migration GO:0071222 - cellular response to lipopolysaccharide	GO:0005515 - protein binding	
4	KLHL14	Q9P2G3	GO:0043001 - Golgi to plasma membrane protein transport; GO:0043087 - regulation of GTPase activity; GO:0045786 - negative regulation of cell cycle; GO:0050821 - protein stabilization	GO:0005515 - protein binding; GO:0043022 - ribosome binding	
		Q9P2G3-2	GO:0043001 - Golgi to plasma membrane protein transport; GO:0043087 - regulation of GTPase activity; GO:0045786 - negative regulation of cell cycle; GO:0050821 - protein stabilization		

**B) Your opinion on the extension of next-CP50 (2018-2021).**

*Due to the nature of protein characterization work, which requires a lot more extra efforts in combination of in vitro, in vivo and in silico approaches, and COVID-19, we consider that the 3 year pilot phase needs to be extended. For instance, we can extend this to 2023 instead of 2021.*

If you agree with this suggestion, please respond by checking “Yes” below. If not, you can add a few comments as you wish.

**Yes.** Comments: the function annotation is untrivial task. There aren't protocols so a few part of functionally annotation proteins have experimentally validated GO-terms.

**C) So far, we have received only two manuscripts on Dark Protein studies. We are wondering when we expect to see your 1st submission to the journal (JPR or any). Would it be possible for you to expect approximate date?**

We hope, that there will be no more interruptions in experimental work caused by the COVID19 pandemic and we will be able to complete the experiments and prepare the material for publication by the end of this year.

**D) Any suggestions on the Dark Protein Studies? e.g., resources, reagents, cell lines, ab, ...others.**

The access to a collection of 'clear' cell lines from a cell bank would be an excellent tool for a comprehensive check of protein function.

## Chromosome Number: 19

### Part 2: uPE1 Proteins (Dark Proteins): neXt-CP50 Challenge

#### Leaders:

Sergio Encarnación-Guevara

#### Major lab members or partners contributing to the neXt-CP50 Challenge:

Nohemi Salinas Jazmín-School of Medicine, UNAM

Orlando Morales-Tarré, Emmanuel Osio Becerro, Angelina Herrera Quiterio, Magdalena Hernández-Ortiz, María del Carmen Vargas-Lagunas-Proteomics laboratory at Center for Genomic Sciences UNAM.

Jeovanis Gil-Valdes. Lund University, Department of Clinical Sciences

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

Emmanuel Salazar Bustamante-Universidad Autónoma del Estado de Morelos

Osbaldo Resendis Antonio, National Institute of Genomic Medicine.

Alejandro García Carranca, National Cancer Institute

**A) Please list the neXt-CP50 Challenge Proteins that your team is characterising and briefly describe your teams progress made to date including any publications or planned papers in 2020.**

uPE1s under research: CCDC97, TMEM160, CCDC61, LENG8

#### CCDC97.

-We have obtained evidence that CCDC97 mRNA and protein are expressed in cervical cancer lines (HeLa [HPV-18 positive], SiHa [HPV-16 positive] and C33A [HPV negative]) and HaCaT cells (transformed keratinocyte cells line, as control). Immunofluorescence microscopic analysis of CCDC97 in cervical cancer lines, showed a differential cellular distribution both nuclear and cytoplasmic.

-We are currently conducting immunohistochemistry experiments on cancerous breast, ovarian and cervical tissues in which, according to Protein Atlas, a very high expression of the mRNA of this gene is observed.

- One of the first evidences that we had regarding the role of the CCDC97 protein, was its apparent relationship with viral proteins of HPV18 and HPV16, therefore we found that we carried out cell-based screening of the target gene in cervical cancer cells.

-Knockout (KO) was generated by the CRISPR/Cas9 system without any off-target effect detected. Western blot results showed successful validation of the CCDC97 knockout in the cervical cancer lines (HeLa, SiHa, C33A)

and HaCaT cells.

-The screen also revealed a potential role for CCDC97, in many cellular functions (cytoskeleton arrangement, adhesion, migration or proliferation), since we observe a different morphology in KO cells. However, relevant assays are required to assign a protein function and to identify if CCDC97 loss conferred a selective disadvantage or vantage on cells.

-Identifying the partners of a given protein (the interactome) may provide leads about the protein function and the molecular mechanisms in which it is involved. To identify proteins interacted with CCDC97, we have made an immunoprecipitation with specific antibodies and soon we will do a mass spectrometry assay to characterize protein interactomes obtains of co-immunoprecipitation from each cellular line.

- Currently the CCDC97 protein function search is the research project of a student to obtain a doctor's degree.

#### TMEM160

- Proteomic analysis by liquid chromatography coupled to mass spectrometry (LC-MS / MS) of CC cell lines showed the expression of the TMEM160 protein.

-Analysis of protein-protein interaction networks reveal that the TMEM160 protein interacts directly with several proteins, including KEAP1 (protein 1 associated with ECH, Kelch type). KEAP1 directly interacts with transcriptional factor NRF2, associated with resistance to chemotherapy and tumor growth in CC. Therefore, we propose to describe the role of TMEM160 in CC using CC cell lines, evaluating the effect of silencing and overexpression of the TMEM160 protein in biological processes such as adhesion, proliferation, migration and resistance to drugs, and thereby suggesting the role of TMEM160 in cervical carcinogenesis.

- In addition, we are conducting immunohistochemical experiments with different tissues from different types of cancer to explore the role of this protein in different types of cancer.

- Currently the TMEM160 protein function search is the research project of a student to obtain a doctor's degree.

#### CCDC61, C19orf47 and LENG8

- This group of proteins we have experimental evidence that they are expressed in cervical cancer cell lines.

-CCDC61, TMEM160, C19orf47 and LENG8 mRNA are expressed in cervical cancer lines (HeLa [HPV-18 positive], SiHa [HPV-16 positive] and C33A [HPV negative]) and HaCaT cells (transformed keratinocyte cells line).

-CRISPR-Cas9 will use for the knockout of individual genes in genome-scale functional screens.

#### B) Your opinion on the extension of next-CP50 (2018-2021).

Due to the nature of protein characterization work, which requires a lot more extra efforts in combination of in vitro, in vivo and in silico approaches, and COVID-19, we consider that the 3 year pilot phase needs to be extended. For instance, we can extend this to 2023 instead of 2021.

If you agree with this suggestion, please respond by checking “Yes” below. If not, you can add a few comments as you wish.

Yes (X) or/and Your comments:

#### C) So far, we have received only two manuscripts on Dark Protein studies. We are wondering when we expect to see your 1<sup>st</sup> submission to the journal (JPR or any). Would it be possible for you to expect approximate date?

If in the next months we can re-integrate into the laboratory work, since the pandemic allows it, we hope to have in the next 12 months solid experimental evidence of the function of at least CCDC97 and important advances in TMEM160

#### D) Any suggestions on the Dark Protein Studies? e.g., resources, reagents, cell lines, ab, ...others.

**Chromosome Number: 20**

#### Part 2: uPE1 Proteins (Dark Proteins): neXt-CP50 Challenge

Major lab members or partners contributing to the neXt-CP50 Challenge:

A) Please list the neXt-CP50 Challenge Proteins that your team is characterising and briefly describe your teams progress made to date including any publications or planned papers in 2020.

Chr No.	20	Work plan and progress
uPE1s Under Investigation	MANBAL, FNDC11	<p><b>Work Plan</b></p> <ol style="list-style-type: none"> <li>(1) Construction of stable cell lines with high expression of uPE1s.</li> <li>(2) Knocking down the uPE1s in cells using siRNAs.</li> <li>(3) Quantitative proteomic study of the constructed cells to discover the related pathways.</li> <li>(4) Function validation in cells and animal models.</li> </ol> <p><b>Progress</b></p> <ol style="list-style-type: none"> <li>(1) Got the stable strains of HeLa cell lines with higher expression of MANBAL and FNDC11.</li> <li>(2) The stable strains of HepG2 cell lines with higher expression of MANBAL and FNDC11 are under construction.</li> </ol>

B) Your opinion on the extension of next-CP50 (2018-2021).

Due to the nature of protein characterization work, which requires a lot more extra efforts in combination of in vitro, in vivo and in silico approaches, and COVID-19, we consider that the 3 year pilot phase needs to be extended. For instance, we can extend this to 2023 instead of 2021.

If you agree with this suggestion, please respond by checking “Yes” below. If not, you can add a few comments as you wish.

Yes ( ✓ ) or/and Your comments:

C) So far, we have received only two manuscripts on Dark Protein studies. We are wondering when we expect to see your 1<sup>st</sup> submission to the journal (JPR or any). Would it be possible for you to expect approximate date?

We have identified more than 10 PE2 MPs from a special medulloblastoma stem cell line only by adopting a traditional shotgun proteomic analysis. Now we are validating these proteins by PRM using their synthetic peptides. We plan to submit our manuscript at the middle of August.

D) Any suggestions on the Dark Protein Studies? e.g., resources, reagents, cell lines, ab, ...others.

We found that a good sample resource is very important for the Dark Protein studies. Recently, we dig more than 10 MPs from the D283 med cell line sourced from brain but derived from metastatic peritoneum. The cell proteins were simply digested and separated into 20 fractions and thus nearly 20 MPs were dig out.

**Chromosome Number: X**

**Part 2: uPE1 Proteins (Dark Proteins): neXt-CP50 Challenge**

**PIC Leaders:**

PI: Yasushi Ishihama (Kyoto University)

**Major lab members or partners contributing to the neXt-CP50 Challenge**  
(Members: jPOST team)

Masaki Matsumoto (Niigata University)  
 Shujiro Okuda (Niigata University)  
 Shin Kawano (Database Center for Life Science, DBCLS)  
 Susumu Goto (Database Center for Life Science, DBCLS)  
 Norie Araki (Kumamoto University)

**(Partners)**

Taiwan ICPC-Chr 4 team (PI: Prof Yu-Ju Chen)

A) Please list the CP50 Challenge Proteins that your team is characterising and briefly describe your teams progress made to date including any publications or planned papers in 2020.

Our jPOST team is in charge of validating the existence by SRM assay for the candidate proteins both in Chr-X and 4 selected by the Taiwan Chr-4 team based on their results from large-scale proteome measurements. Currently, 22 and 19 uPE1 protein candidates were selected for further analysis.

Table 41 uPE1 candidates for further validation

Chr 4	P78312	Q0P651	Q56VL3	Q5BJH2	Q5M9N0	Q68CR1
	Q6NW29	Q6ZU35	Q6ZUS6	Q86YA3	Q8IUW5	Q8N1A6
	Q8N8J7	Q8NEC7	Q8WVX3	Q96EY4	Q96QK8	Q9C0D6
	Q9C0F1	Q9P2B7	Q9ULE4	Q9Y605		
Chr X	A2AJT9	A6NJG2	A6ZKI3	O14668	Q14656	Q5JSJ4
	Q5U3C3	Q6P1M9	Q6ZTR5	Q7Z309	Q8N9E0	Q8NFB2
	Q8TB03	Q96A49	Q9BVG4	Q9BWD3	Q9H5V9	Q9HAI6
	Q9Y4X0					

In addition, 4 proteins such as sp|Q6IC83|CV042\_HUMAN, sp|Q7Z570|Z804A\_HUMAN, sp|Q8IVF6|AN18A\_HUMAN and sp|Q8IZA3|H1FOO\_HUMAN in human iPS cells will be analysed.

**Chromosome Number: Y**

**Part 2: uPE1 Proteins (Dark Proteins): neXt-CP50 Challenge**

**Major lab members or partners contributing to the neXt-CP50 Challenge:**

A) Please list the neXt-CP50 Challenge Proteins that your team is characterising and briefly describe your teams progress made to date including any publications or planned papers in 2020.

► PRY: we expressed and purified the proteins and its characterization is in progress.

**B) Your opinion on the extension of next-CP50 (2018-2021).**

Due to the nature of protein characterization work, which requires a lot more extra efforts in combination of in vitro, in vivo and in silico approaches, and COVID-19, we consider that the 3 year pilot phase needs to be extended. For instance, we can extend this to 2023 instead of 2021.

If you agree with this suggestion, please respond by checking “Yes” below. If not, you can add a few comments as you wish.

Yes ( ✓ ) or/and Your comments:

C) So far, we have received only two manuscripts on Dark Protein studies. We are wondering when we expect to see your 1<sup>st</sup> submission to the journal (JPR or any). Would it be possible for you to expect approximate date?

2021. Since there is only one Y chr uPE1, we have no option but working on PRY which has some challenges for functional characterization.

D) Any suggestions on the Dark Protein Studies? e.g., resources, reagents, cell lines, ab, ...others.

We suggest C-HPP to use pluripotent stem cells (PSC) for identification of MPS and characterisation of uPE1 due to their ability to differentiate into three embryonic germ layers including endoderm, mesoderm, and ectoderm. The importance of this ability is that there are considerable number of proteins, the expression of which are limited to embryonic developmental stages. iPSCs have been widely used to generate patient-specific disease models. The enthusiasm rose higher with the rapid advances in precise DNA editing and CRISPR–Cas9 technology in particular, owing to its simplicity in design and ease of use. Moreover, the generation of organoids that possess part of characteristics of the corresponding in vivo tissue, provides an ideal opportunity for functional analysis of proteins. Integrating CRISPR engineering, hiPSC-derived disease modeling systems, and organoid technologies provides unique platform for C-HPP for identification of MPs and functional characterization of proteins especially uPE1. C-HPP can collaborate with groups which can generate cell lines and differentiate them.

## Chromosome Number: MT

### Part 2: uPE1 Proteins (Dark Proteins): neXt-CP50 Challenge

#### Major lab members or partners contributing to the neXt-CP50 Challenge:

Andrea Urbani, Paola Roncada, Mauro Fasano, Alberio T, Pieroni L, Ronci M, Brioschi M, Cunsolo V, Giusti L, Greco V, Lucacchini A, Soggiu A, Zilocchi M, Marini F, Carregari VC, Iavarone F, Persichilli S, Castagnola M, Ciregia F, Sickmann A, Babu M,

A) Please list the neXt-CP50 Challenge Proteins that your team is characterising and briefly describe your teams progress made to date including any publications or planned papers in 2020.

Accession	Description	Mito evidence IMPI	Mito localization IMPI	HPA cell localization	Enzyme	Exp. fraction	Gravy score
NX_O60941-1	Dystrobrevin beta	Predicted	Unknown	Mitochondria (A)	Try; Glu-C	F1, F2	-0.55
NX_Q3SXM5-1	Inactive hydroxysteroid dehydrogenase-like protein 1	Known	OM	Intracellular, Membrane (P)	Chym; Try	F2	+0.15
NX_Q4VC31-1	Coiled-coil domain-containing protein 58	Known	IMS	Nucleoli, Mitochondria (A)	Try; Chym; Glu-C	F2	-0.60
NX_Q56VL3-1	OCIA domain-containing protein 2	Known	Unknown	Mitochondria (E)	Try	F2	-0.26
NX_Q8IYQ7-1	Threonine synthase-like 1	Known	Matrix	Nuclear bodies (A), Mitochondria (A), Cytosol	Try	F2	-0.13



Accession	Description	Mito evidence IMPI	Mito localization IMPI	HPA cell localization	Enzyme	Exp. fraction	Gravy score
NX_Q8NFV4-1	Protein ABHD11	Known	Matrix	(A) Mitochondria (S)	Try; Chym	F2	-0.09
NX_Q96EX1-1	Small integral membrane protein 12	Predicted	Unknown	Mitochondria (A)	Try	F2	-0.53
NX_Q96C01-1	Protein FAM136A	Known	IMS	Mitochondria (A)	Try; Glu-C	F2	-0.43
NX_Q96ER9-1	Coiled-coil domain-containing protein 51	Known	Matrix	Nucleosome (S), Mitochondria (S) Centrosome (A)	Try; Glu-C	F2	-0.38
NX_P56378-1	6.8 kDa mitochondrial proteolipid	Known	IM	Mitochondria (S),Nucleoli (S)	Try	F1,F2	-0.02
NX_Q9GZT6-1	Coiled-coil domain-containing protein 90B	Known	Matrix	Mitochondria (E)	Try; Glu-C	F2	-0.55
NX_A8MTT3-1	Protein CEBPZOS	Known	IMS	Nucleoplasm (A)	Try; Glu-C	F2	-0.27
NX_Q9H4I3-1	TraB domain-containing protein	Known	OM	Nucleus (A), Mitochondria (A)	Try	F2	-0.21
NX_Q9UFN0-1	Protein NipSnap homolog 3A	Known	Matrix	not available	Try; Glu-C	F2	-0.37
NX_Q6P1 × 6-1	UPF0598 protein C8orf82	Known	Matrix	Nucleus (A)	Try; Chym	F2	-0.23
NX_Q8N2U0-1	Transmembrane protein 256	Predicted	Unknown	Vesicles (A)	Try	F2	+0.46
NX_Q8WV10-1	Small integral membrane protein 4	Predicted	Unknown	Nucleoplasm (A), Mitochondria (A)	Try	F2	-0.54
NX_Q8WW59-1	SPRY domain-containing protein 4	Known	Matrix	Nucleoplasm (A)	Try	F1,F2	-0.07
NX_Q96BQ5-1	Coiled-coil domain-containing protein 127	Known	OM/IMS	Nucleus (S), Nucleoli (S)	Try; Chym	F1,F2	-0.72
NX_Q96DB5-1	Regulator of microtubule	Known	OM/IMS	Centrosomes (S),Actin	Try	F1,F2	-0.37

Accession	Description	Mito evidence IMPI	Mito localization IMPI	HPA cell localization	Enzyme	Exp. fraction	Gravy score
	dynamics protein 1			filaments (S)			
NX_Q96KF7-1	Small integral membrane protein 8	Known	OM	Vesicles (A)	Try; Chym	F2	-0.55
NX_Q9NU23-1	LYR motif-containing protein 2	Known	matrix	Cytosol (A)	Try	F2	-0.70

See for reference: Exploring the HeLa Dark Mitochondrial Proteome.

Marini F, Carregari VC, Greco V, Ronci M, Iavarone F, Persichilli S, Castagnola M, Urbani A, Pieroni L. Front Cell Dev Biol. 2020 Mar 5;8:137. doi: 10.3389/fcell.2020.00137. eCollection 2020. PMID: 32195257

**B) Your opinion on the extension of next-CP50 (2018-2021).**

Due to the nature of protein characterization work, which requires a lot more extra efforts in combination of in vitro, in vivo and in silico approaches, and COVID-19, we consider that the 3 year pilot phase needs to be extended. For instance, we can extend this to 2023 instead of 2021.

If you agree with this suggestion, please respond by checking “Yes” below. If not, you can add a few comments as you wish.

Yes (  ) or/and Your comments:

There is a significant lack of funding for these activities

**C) So far, we have received only two manuscripts on Dark Protein studies. We are wondering when we expect to see your 1<sup>st</sup> submission to the journal (JPR or any). Would it be possible for you to expect approximate date?**

The current pandemic crisis is limiting our investigations in this field, most probably we will not submit any contribution

**D) Any suggestions on the Dark Protein Studies? e.g., resources, reagents, cell lines, ab, ...others.**

- 2) Development of a joint NDA under the HUPO leadership for data sharing before publication of collected experimental and in silico evidences (eg. MS data, protein-protein interaction matrices, pQTR/eQTR , etc. etc.)
- 3) Lobbying for funding!