

### **Human Proteome Project: Russian Roadmap for Chromosome 18**

Archakov A., Bykov V., Grigoriev A., Govorun V., Ivanov V., Khlunov A., Lisitsa A., Makarov A., Sagdeev R., Skryabin K.

The Human Proteome Project (HPP) has been proposed to determine and characterize the proteins encoded by the human genome [Pearson, 2008]. This is expected to be a greater challenge than mapping the human genome. First, the human genome is constant over time and generally similar in each cell. In contrast to the genome, the proteome differs essentially in different cell types and biological fluids and varies over time [Frenkel-Morgenstern et al., 2010]. Secondly, translational and post-translational modifications cause the wide diversity of various protein forms accounted for a single gene. Finally, due to the absence of a PCR analog for proteins, a technological problem exists in determining ultralow protein concentrations [Archakov et al., 2009].

The scope of the HPP requires the distribution of work among participants. In the same way as for the Human Genome Project, it has been suggested to distribute the work based on a gene-centric approach [HUPO, 2010]; *i.e.* to identify protein products of genes according to their distribution over chromosomes. A gene-centric approach, unlike the formerly used organ-based and disease-based approaches, enables the proteome study to be organized as a “roadmap”. With such roadmap, the scientific problems that are resolved, along with the advances in analytical technologies, will create the basis for the development of translational medicine.

The roadmap section that will be established by Russian scientists is identification of the proteins coded by genes of chromosome 18. This comparatively tiny chromosome consists of 76 Mbases and contains about 500 genes [Nisbaum et al., 2005]; however, only around 300 of them encode potentially expressed proteins. According to UniProt annotations, 89% of expressed genes were deciphered from cDNA/EST databases.

Almost 80% of the proteins translated from genes on chromosome 18 have already been identified with mass-spectrometric methods and introduced into PRIDE. According to ProteinAtlas, 107 proteins were revealed by immuno-histological methods in tissue samples. Having analyzed the available data on proteins of this chromosome, identification of 46 proteins associated with chromosome 18 is currently lacking.

Identifying the proteins of chromosome 18 will have a pilot and a master phase. During the 3-year pilot phase, proteins of chromosome 18 will be identified in three types of biological material: blood plasma, cell culture HepG2 and liver tissue. The goal of the pilot phase is to identify, at least one protein for each gene, determine the level of its expression and predominant modifications. The results of the pilot phase will include data on individual variability of the proteome in blood plasma and liver tissue.

Tissue/organ-based dimension of the roadmap provides the links between gene centric HPP and pre-existing HUPO initiatives. Among HUPO initiatives liver proteome project (HLPP) and plasma proteome project (HPPP) are most profoundly developed. The expertise of these initiatives if inherited by the gene centric HPP establishes a steady springboard for the pilot phase. Other tissues would complement the master phase in case some chromosome-specific gene products still will be missing in plasma and liver even at the lowermost detection limit.

The master phase of the roadmap elaboration of chromosome 18 will be completed in 5 years. It will include experimental revelation of the modifications for all proteins of chromosome 18. Modifications will include single amino acid polymorphisms, the products of alternative splicing and post-translational modifications. Based on a rough estimate of 100 modified variants for one gene, it is expected that about 30,000 variants of the proteins will be identified upon execution of the master phase.

It is expected that the roadmap elaboration process will encounter several tasks, which will span both phases of the project. These tasks fall into the following categories:

- genome/transcriptome analysis using single-molecular DNA/RNA readers [Tsutsui et al, 2010] to perform deep sequencing of putative coding regions of chromosome 18 and to elucidate the alternatively spliced transcripts;
- detection of medium- and low-copied proteins by utilizing multi-dimensional separation with MRM technology; proteotyping and proteogenomic profiling [Armengaud, 2010]; protein affinity capturing [Buneeva et al., 2010] to gather the chromosome-centered portion of interactome;
- advanced technology for detection of ultralow-copied proteins using nanowire detectors and atomic force microscopy in combination with mass-spectrometry [Kaisheva et al., 2010].

The advance of analytical technologies is the basic focus of the Russian roadmap efforts. In our opinion, the top target in this direction is the usage of nanotechnological approaches, which are already gaining strength in genomics [Tsutsui et al, 2010]. Provided that high-speed single molecule detection is applicable for proteins [Shibata et al., 2010], then the problem of concentration sensitivity in proteomics will be resolved [Archakov et al., 2009]. There will emerge the opportunity to analyze cells and biological fluids at the resolution of tens of protein molecule copies per sample. Perhaps, using nanotechnologies, there will be success in resolving the second problem related to situational variability of the proteome.

As a deliverable from participating in the HPP, Russia is planning to establish technologies for proteomic studies integrating mass-spectrometry with atomic-force microscopy (AFM-MS). With such technologies one may expect to attain sensitivity at the level of  $10^{-18}$  M in blood plasma and 10-100 copies of protein molecules for 1 cell of liver tissue. Application of AFM-MS hybrids will make it possible to reveal a principally new group of early biomarkers directly relating to the onset of the disease.

Roadmap supports the translational medicine by incorporating a plan for the construction of molecular sensors based on nanowire detectors [Lee et al., 2008]. In the foreseeable future, these devices will provide an affordable means for complex, multiparametric molecular analyses. Appropriate reliability of such devices allows point-of-care and even home environment diagnostics. The roadmap master phase will establish a benchmark for personalized medicine by creating prerequisite hardware for fostering individual prophylaxis and treatment of socially significant diseases.

In addition to the diagnostic milestones, the Russian effort in the HPP emphasizes the peptidome as a source of prototypes for new drugs; *i.e.* peptidomimetics with high biological compatibility [Ivanov & Yatskin, 2005].

Finally, we make note of a particular feature of the Russian roadmap effort in the context of HPP data management. It has been suggested that a knowledge-based informational model [Kuchar et al., 2004] may allow access to protein identifications deposited through the ProteomeXchange. In this way, the accumulation of experimental information may be changed to a hypothesis generation mode, which may defrost the true utility of the postgenomic era [Hull et al., 2008].

The roadmap effort described here resulted from intensive discussion of the HPP among leading scientists of Russia. In April 2010, the roadmap effort was presented to the Advisory Council of the Prime-minister of Russia and official endorsement was obtained. A full description of the roadmap effort is available on the HUPO HPP web-page [<http://www.hupo.org/--->] and international discussion is encouraged on the optimal way to couple the country-specific expectations with the general scaffold of HPP.

Pearson H. (2008) Biologists initiate plan to map human proteome.

HUPO (2010). A Gene-centric Human Proteome Project: HUPO-The Human Proteome organization. *Mol Cell Proteomics*, 9, pp. 427-429.

Frenkel-Morgenstern M., Cohen A.A., Geva-Zatorsky N., Eden E., Prilusky J., Issaeva I., Sigal A., Cohen-Saidon C., Liron Y., Cohen L., Danon T., Perzov N., and Alon U. (2010) "Dynamic Proteomics: a database for dynamics and localizations of endogenous fluorescently-tagged proteins in living human cells.," *Nucleic acids research*, vol. 38, pp. D508-12.

Nusbaum C., Zody M.C., Borowsky M.L., et al., and Lander E.S., "DNA sequence and analysis of human chromosome 18.," *Nature*, vol. 437, 2005, pp. 551-5.

Tsutsui M., Taniguchi M., Yokota K., and Kawai T. (2010), "Identifying single nucleotides by tunnelling current.," *Nature nanotechnology*, 5, pp. 286-90.

Armengaud J., "Proteogenomics and systems biology: quest for the ultimate missing parts.," *Expert review of proteomics*, vol. 7, 2010, pp. 65-77.

Ivanov V.T. and Yatskin O.N., "Peptidomics: a logical sequel to proteomics.," *Expert review of proteomics*, vol. 2, 2005, pp. 463-73.

Kaysheva A.L., Ivanov Y.D., Zgoda V.G., Frantsuzov P.A., Pleshakova T.O., Krokhin N.V., Ziborov V.S., and Archakov A.I. (2010), "Visualization and identification of hepatitis C viral particles by atomic force microscopy combined with MS/MS analysis," *Biochemistry (Moscow) Supplement Series B: Biomedical Chemistry*, vol. 4, pp. 15-24.

Archakov A., Ivanov Y., Lisitsa A., and Zgoda V. (2009) "Biospecific irreversible fishing coupled with atomic force microscopy for detection of extremely low-abundant proteins.," *Proteomics*, vol. 9, pp. 1326-43.

Lee M., Lee D., Jung S., Lee K., Park Y.S., and Seong W. (2010) "Measurements of serum C-reactive protein levels in patients with gastric cancer and quantification using silicon nanowire arrays.," *Nanomedicine : nanotechnology, biology, and medicine*, vol. 6, pp. 78-83.