# The good, the bad, the aberrant: the role of prevailing splice-forms in proteomic studies

E. Poverennaya\*, O. Kiseleva, M. Pyatnitskiy, E. Ilgisonis, E. Ponomarenko, A. Lisitsa
*Institute of Biomedical Chemistry, Moscow, Russia*
*\* e-mail: k.poverennaya@gmail.com*

*Motivation and Aim*: Thousands of genes expressed in a particular cell determine the functionality of the cell. Each step in the flow from DNA via RNA and finally to protein supplies the cell with a potential control point for self-regulation of its functions. This control is implemented by alternation of the amount and type of proteoforms the cell generates by "fine adjustment" of different aberrations. One of the fundamental resources of such aberrations is alternative splicing. The difference between splice-forms is defined by a combination of exons. While the majority of proteomic studies are devoted to examination of canonical sequences, we focused on more sophisticated task of analysis of aberrant forms. It is non trivial because of high homology between the canonical and splice-forms: the selection of proteotypic peptide which uniquely characterizes the splice-form and does not map on the canonic form is often challenging.

*Methods and Algorithms*: RNAseq data of normal liver tissue, lung, bladder, kidney, stomach and esophagus, publicly available in SRA database, were analyzed to identify the prevailing form (canonical or aberrant) by level of expression for each gene. The transcriptomic data were processed by PPLine software. Then transcripts were translated into amino acid sequences, which were *in silico* cleaved by trypsin. The frequency of detection of proteotypic peptides in proteomic experiments was calculated according to the GPMdb.

*Results*: According to the results of the transcriptomic analysis of six tissues, splice-forms prevail over canonical forms in more than 2.7 thousand cases. Notably, that in 1.6 thousand cases gene products were presented only by splice-forms. We analyzed the prevalence of splice-forms on the example of human chromosome 18, key chromosome of the Russian part of Chromosome-Centric Human Proteome Project. For 52 of 275 genes of chr18 we revealed, that 54 splice-forms prevail over canonical forms in 6 examined types of tissues: 24 forms were found in all tissues, two were detected only in esophagus, one – in stomach, bladder, and lung each, while in kidney tissue there was no specific splice-form.

According to the UniProt database, there are ca. 550 thousand of proteotypic peptides in human proteome: 50 % of them correspond both to canonical and splice-forms, 40 % (223 thousand) are specific only for canonical form, and 4 % are specific exclusively for splice-forms.

*Conclusion*: Knowledge about prevailing form is crucial on every step of a proteomic study – from experiment design to data interpretation.