TARGETED PROTEOMICS FOR HEALTH ANALYTICS: OPPORTUNITIES AND CHALLENGES

Elena A. Ponomarenko[1]*, Svetlana E. Novikova[1], Ekaterina V. Ilgisonis[1], Olga I. Kiseleva[1], Svetlana N. Tarbeeva[1], Anna A. Kluchnikova[1], Mikhail A. Pyatnitskiy[1], Andrey V. Lisitsa[1], Tatyana E. Farafonova[1], Ekaterina V. Poverennaya[1], Arthur T. Kopylov[1], Young-Ki Paik[2], Ghasem Hosseini Salekdeh[3], Andrea Urbani[4], Victor G. Zgoda[1] and Alexander I. Archakov[1]

1 – Institute of Biomedical Chemistry
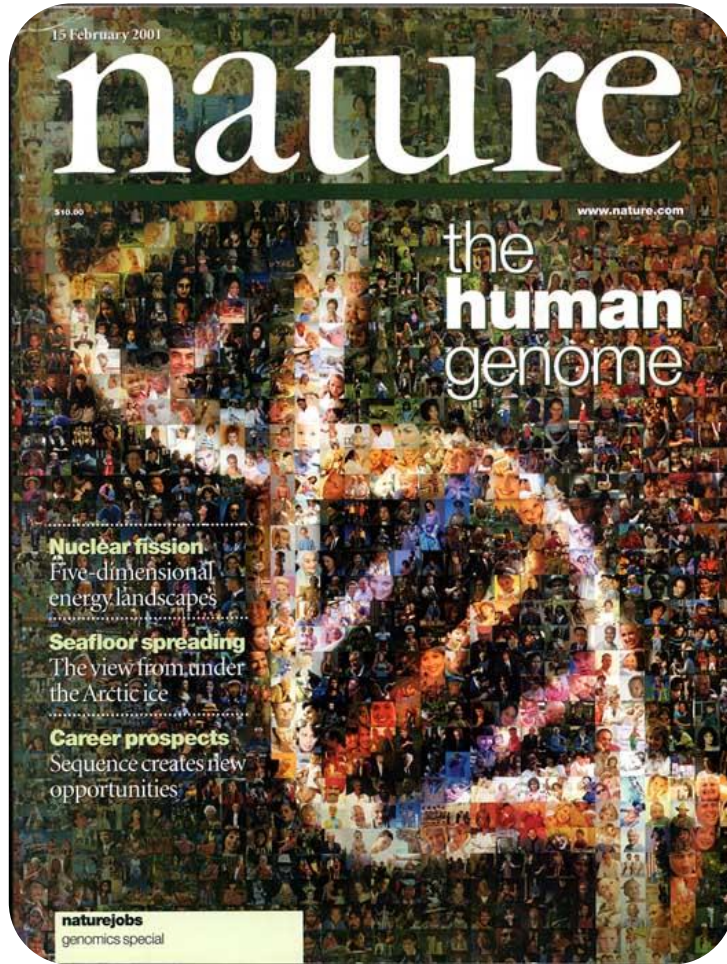2 - Yonsei Proteome Research Center, Yonsei University, Seoul 03722, Korea
3 - Department of Molecular Systems Biology, Royan Institute for Stem Cell Biology and Technology, ACECR, Tehran, Iran
4 - Area of Diagnostic Laboratories, Fondazione Policlinico Gemelli—IRCCS, Rome 00168, Italy
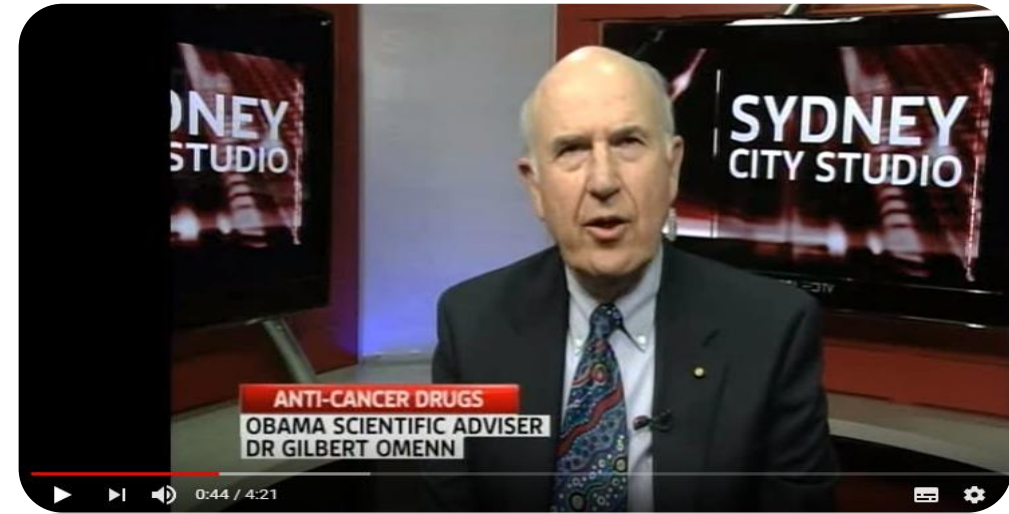
Moscow

24.05.2022

2001: Human genome

2010: International Human Proteome Project was launched

2014: Drafts of Human Proteome

*Legrain P. et al., **The human proteome project: current state and future direction.** Mol Cell Proteomics. 2011 Jul;10(7):M111.009993.*
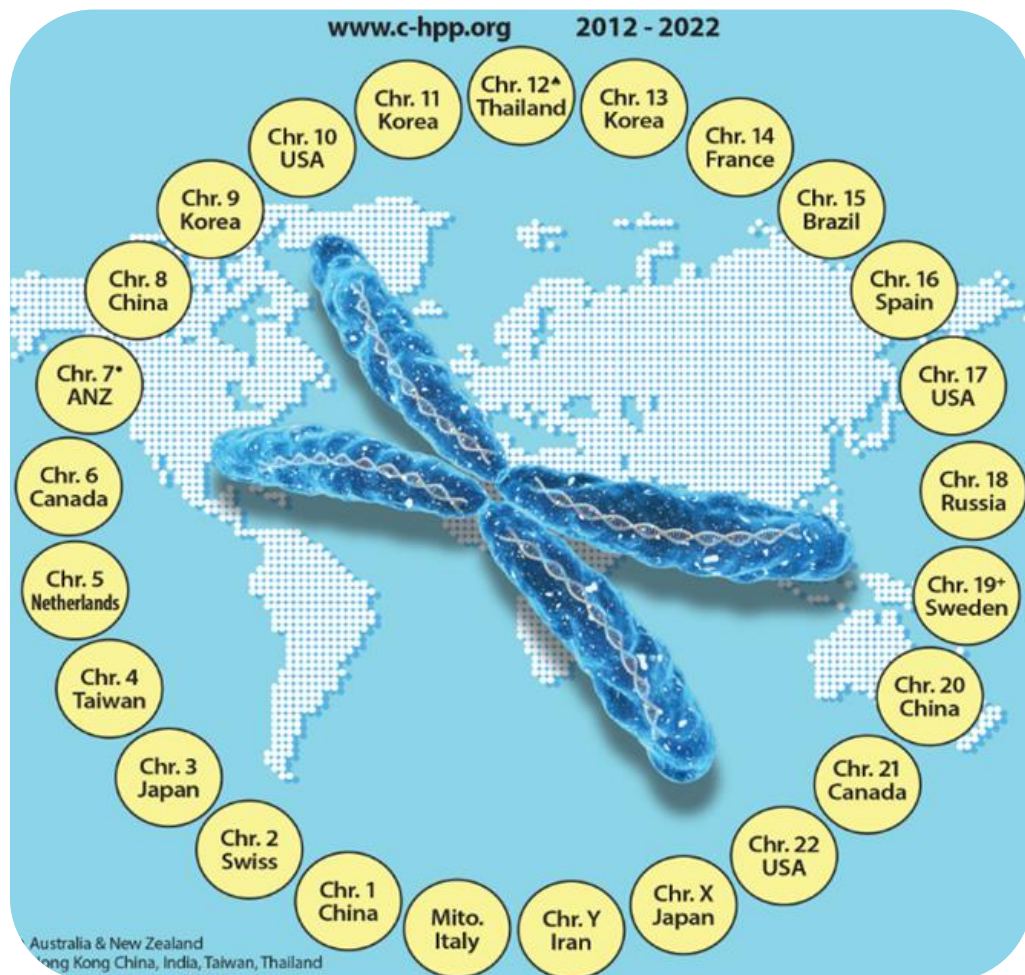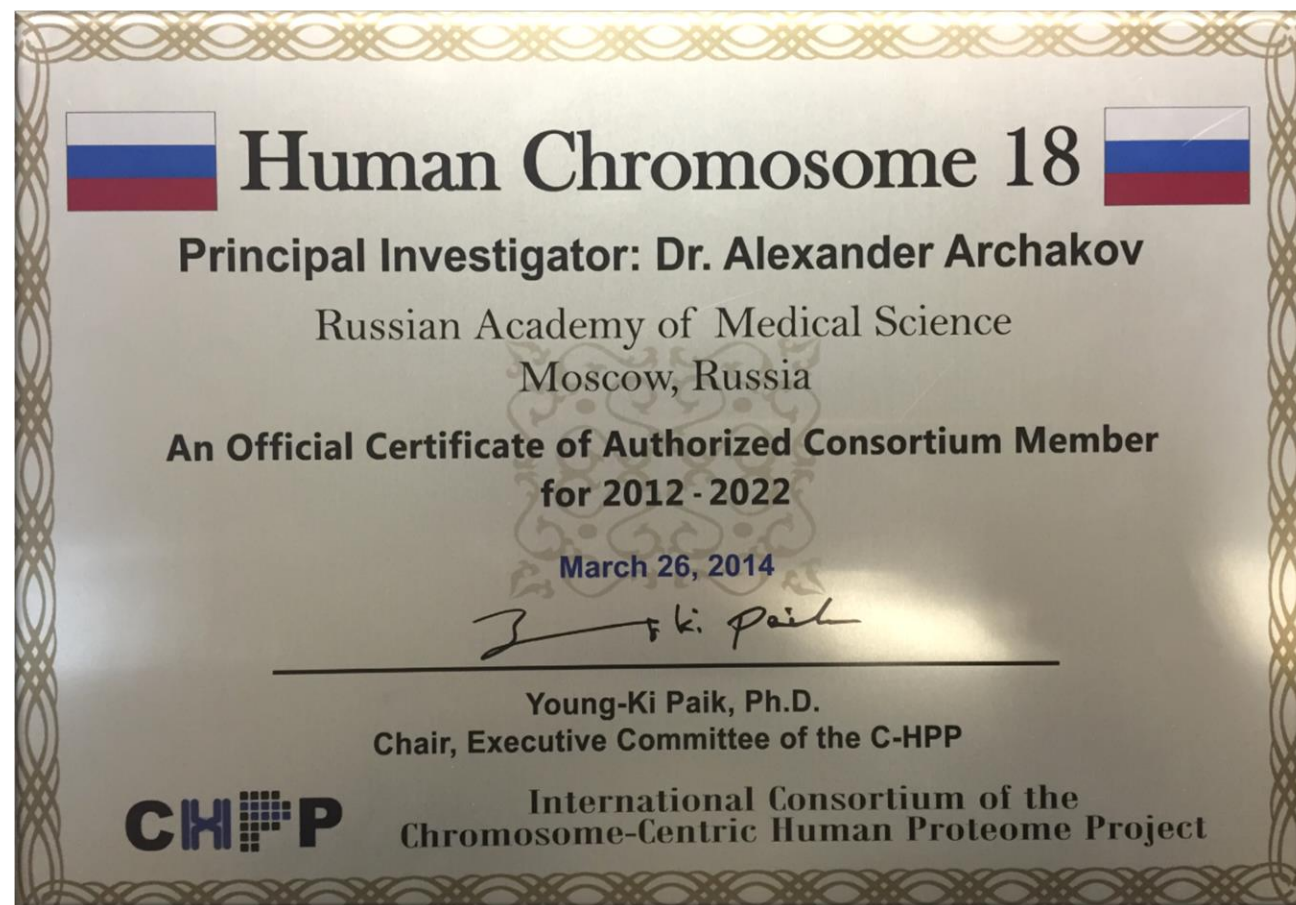
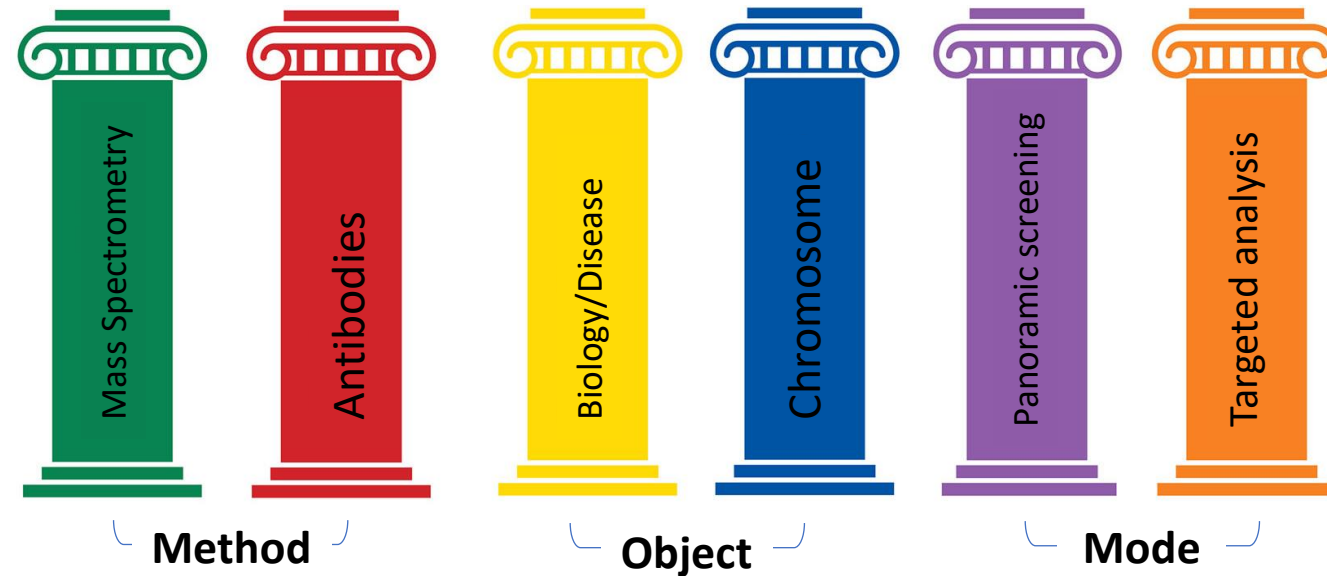# The Chromosome-Centric Human Proteome Project for Cataloging Proteins Encoded in the Genome

Paik YK et al., The Chromosome-Centric Human Proteome Project for cataloging proteins encoded in the genome. Nat Biotechnol. 2012 Mar 7;30(3):221-3. doi: 10.1038/nbt.2152.

Archakov et al. Gene-centric view on the human proteome project: the example of the Russian roadmap for chromosome 18. Proteomi V. 11(10). P. 1853-6

Pillars of Human Proteome Project

Mass Spectrometry | Antibodies | Biology/Disease | Chromosome | Panoramic screening | Targeted analysis

Method | **Object** | **Mode**

Russian part of the

Our aim: to find a powerful tool *for detection and quantification* all major splice forms of all proteins encoded on the Chromosome 18 (n = 275) of the human genome using targeted proteomics

IBMC
Institute of Biomedical Chemistry
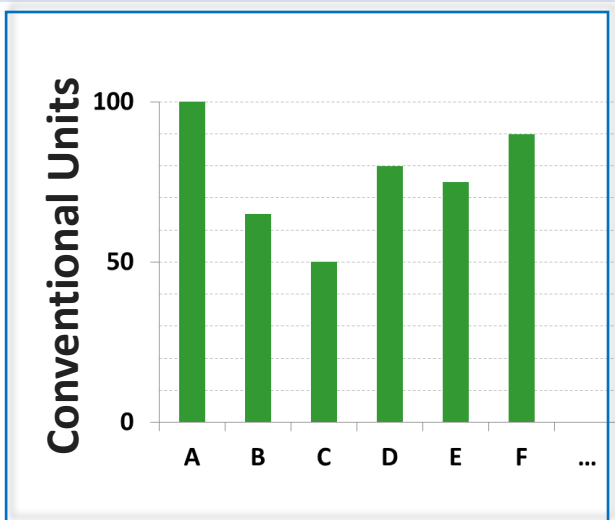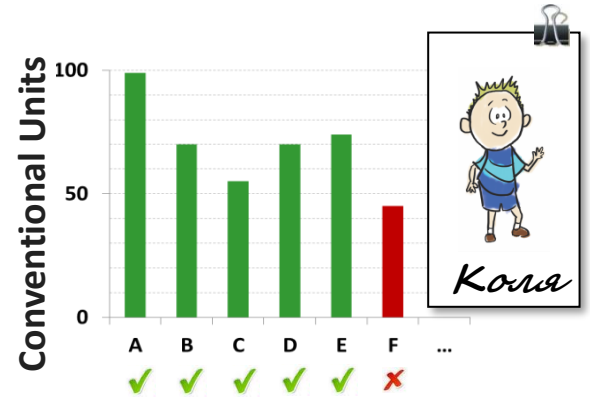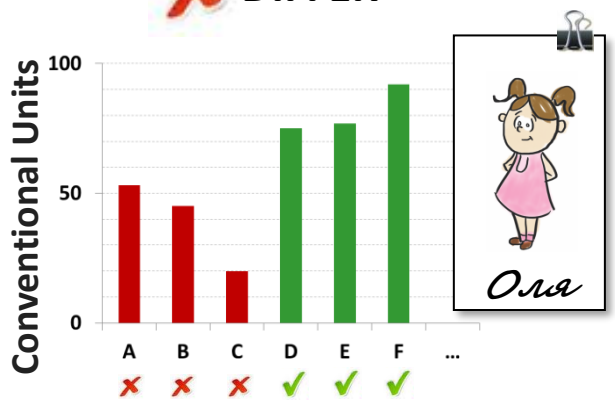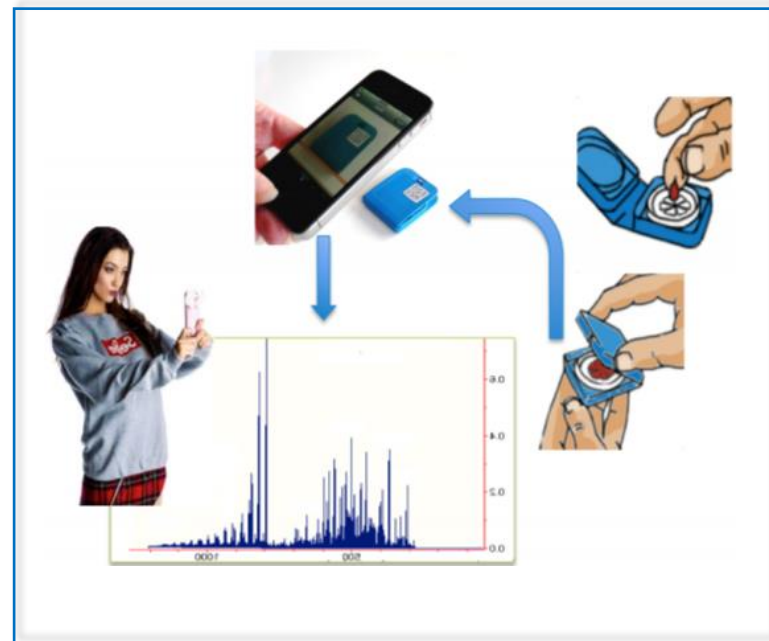
Selected Reaction Monitoring with Stable Isotope-labeled peptide Standards (SRM SIS)
were applied for the quantitative measurement of proteins in human blood plasma

Bioinformatic part

LIST OF PROTEOTYPIC PEPTIDES:

	aa Sequence		
Protein A1*	AAAAAA	XYYXXY	AAA**X**AA
Protein A2*	AAAAAA	XYYXXY	AAA**Y**AA
Protein B	AAAAAA	XXXXXX	AAAAAA

A1 and A2 coded by the same gene (proteoforms)

Proteotypic peptides are specific peptides, for search and measure of which MS is set up:

AAA**X**AA (PROTEIN A1)

AAA**Y**AA (PROTEIN A2)

XXXXXX (PROTEIN B)

Experiment (Targeted MS analysis)

MEASUREMENTS OF PROTEIN CONTAINING THE PROTEOTYPIC PEPTIDE IN THE SAMPLE

AAA**X**AA

AAA**Y**AA

XXXXXX

PROTEIN A1 DETECTED IN THE SAMPLE

PROTEIN A2 DETECTED IN THE SAMPLE

PROTEIN B NOT DETECTED IN THE SAMPLE
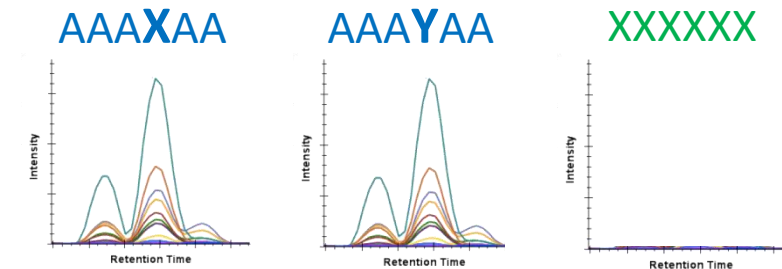
The Size of Human Chromosome 18 Master* Proteome (Human blood Plasma, Liver and HepG2 cells)

(A) Proteome Width

Blood plasma (n=84)
Liver tissue (n=73)
HepG2 cell line (n=66)
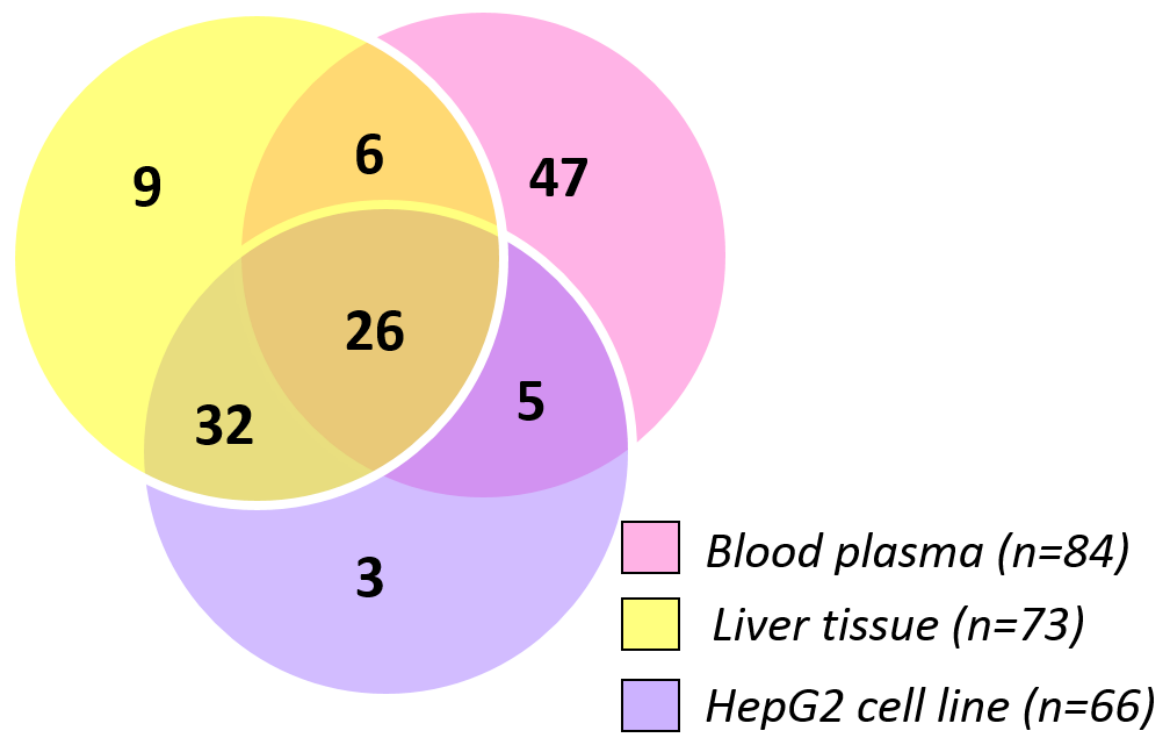
*at least one of the proteoforms encoded by the gene

(B) Proteome Depth: plasma proteins are on average more abundant in comparison with the proteins found in the cells

Plasma
Liver tissue
HepG2 cell line
Proteins detected at $S=10^{-18}$M
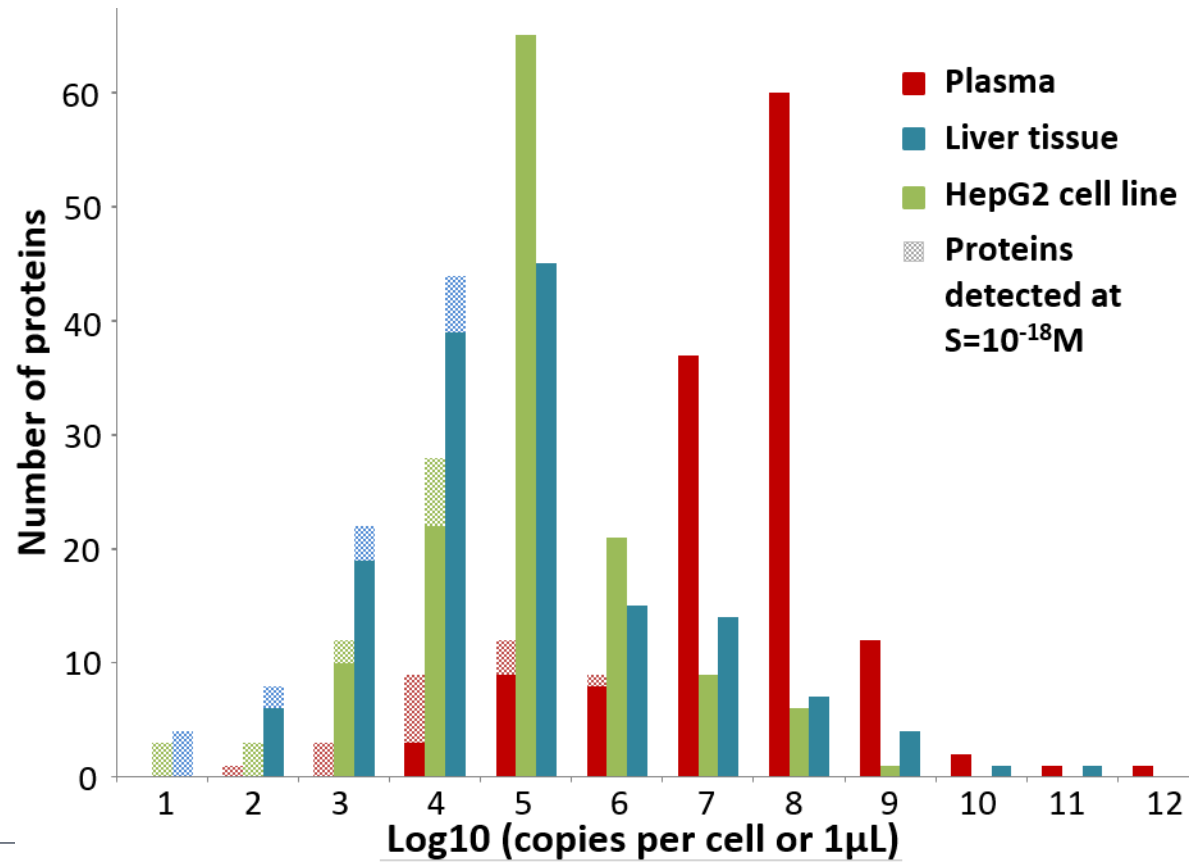
1. Zgoda et al., Journal of Proteome Research. – 2013. – V. 12(1). – P. 123–134.
2. Ponomarenko et al., Journal of Proteome Research. – 2014. – V. 13(1). – P. 183–190.
3. Poverennaya et al., Journal of Proteome Research. – 2016. – V. 15(11). – P. 4030–4038.

IBMC
Institute of Biomedical Chemistry

Commercial solutions for targeted proteomics

Pros:
- Robust, reproducible results
- Up to 500 proteins per run
- Quantitative analysis
- Standardized sample preparation protocol

MRM Proteomics
Vancouver, Canada
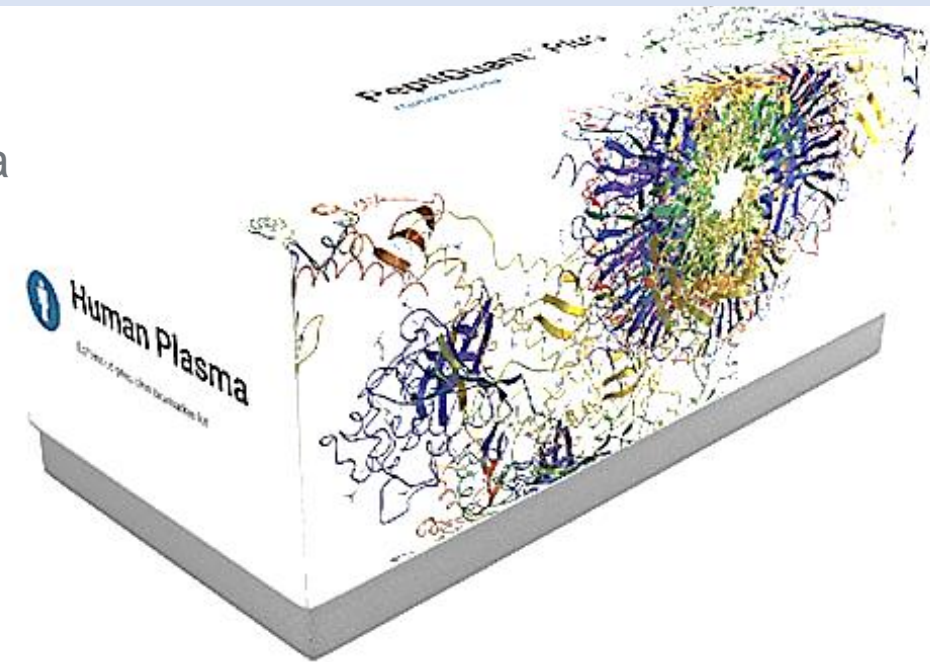
Cons:
- Proprietary software
- Price
- Unknown peptides
- No reference concentrations

Biognosys AG
Schlieren, Switzerland

Absolute concentration instead of Relative concentration open the way to precision molecular health monitoring

What is the normal range of protein concentrations in the blood plasma of healthy people?

Part 1. Application of targeted quantitative SRM SIS for quantification of proteins approved by the FDA for clinical use *(111 target proteins*)* in human blood plasma

- ✓ Compare results with data obtained by other methods
- ✓ Measure the proteomic profiles and decipher the normal range of protein concentrations in the blood plasma of healthy people, taking into account the interindividual variability

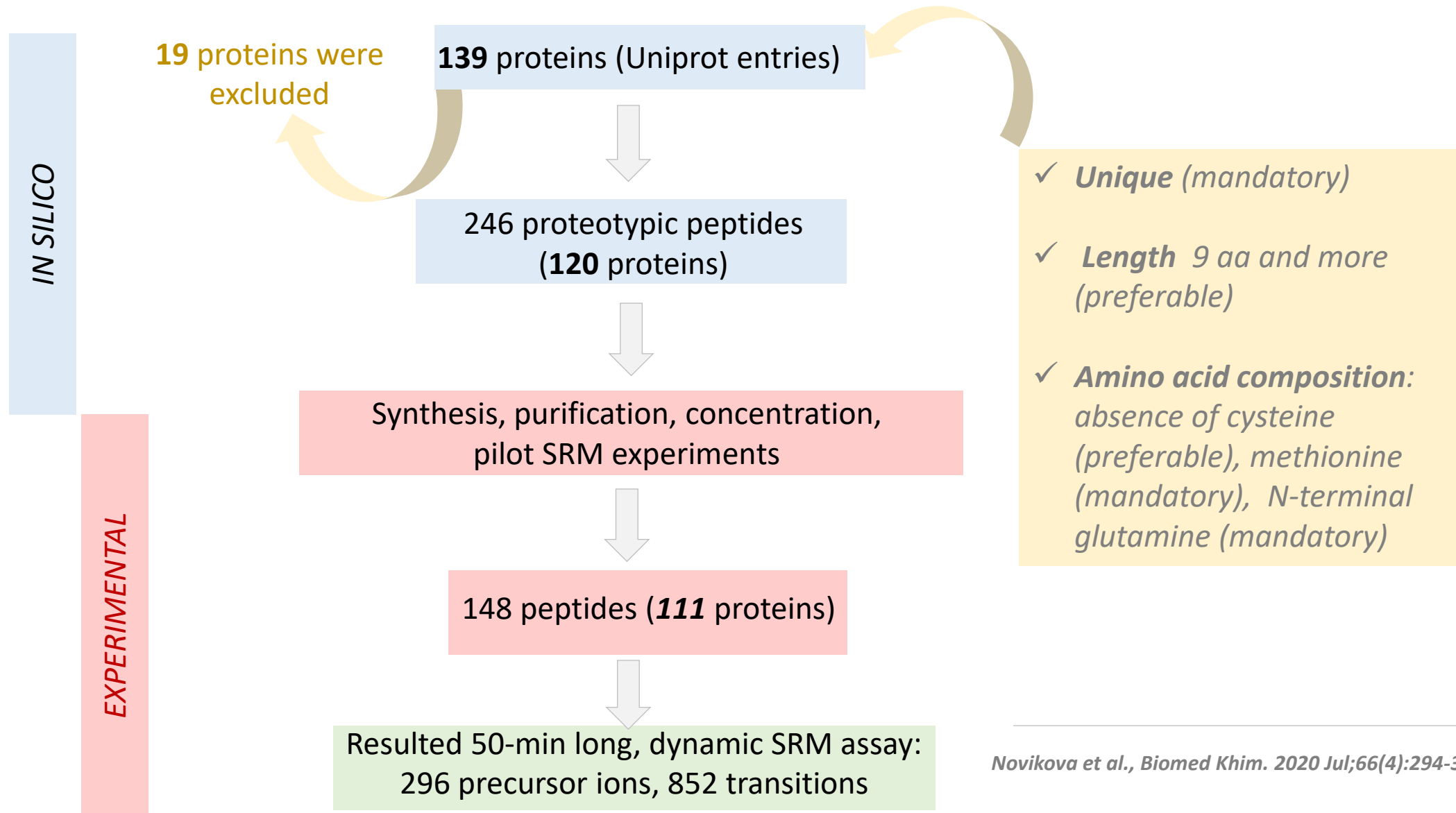**Part 2.** Targeted Quantitative SRM SIS Screening of Chr 18, 13, Y and the Mt Encoded Proteome **(600+ target proteins)**

- ✓ Expand the concentration range of detected proteins
- ✓ Propose new methods and approaches to increase the number of detected proteins
- ✓ Define a set of proteins for multiplex quantitative protein analysis

*Anderson, Clin Chem. 2010 Feb;56(2):177-85.

IBMC
Institute of Biomedical Chemistry

The SRM assay development scheme: FDA-approved proteins

IN SILICO

EXPERIMENTAL

19 proteins were excluded

139 proteins (Uniprot entries)

246 proteotypic peptides (**120** proteins)

Synthesis, purification, concentration, pilot SRM experiments

148 peptides (*111* proteins)

Resulted 50-min long, dynamic SRM assay: 296 precursor ions, 852 transitions

- ✓ **Unique** (mandatory)
- ✓ **Length** 9 aa and more (preferable)
- ✓ **Amino acid composition**: absence of cysteine (preferable), methionine (mandatory), N-terminal glutamine (mandatory)

Novikova et al., Biomed Khim. 2020 Jul;66(4):294-316.

IBMC
Institute of Biomedical Chemistry

Proteomic profiling of blood plasma samples derived from healthy volunteers

1. Healthy volunteers
(N=31, 18 males, 12 females)

**2. Medical surveillance
for two days in Clinic**
(Institute of Nutrition, Moscow, Russia)

**3. Samples
collection**

4. Tryptic digestion

LC-MRM analysis

QQQ Quantiva

**5. Samples were spiked with
isotopically labeled peptide standards**

Novikova et al., Biomed Khim. 2020 Jul;66(4):294-316.

IBMC
Institute of Biomedical Chemistry
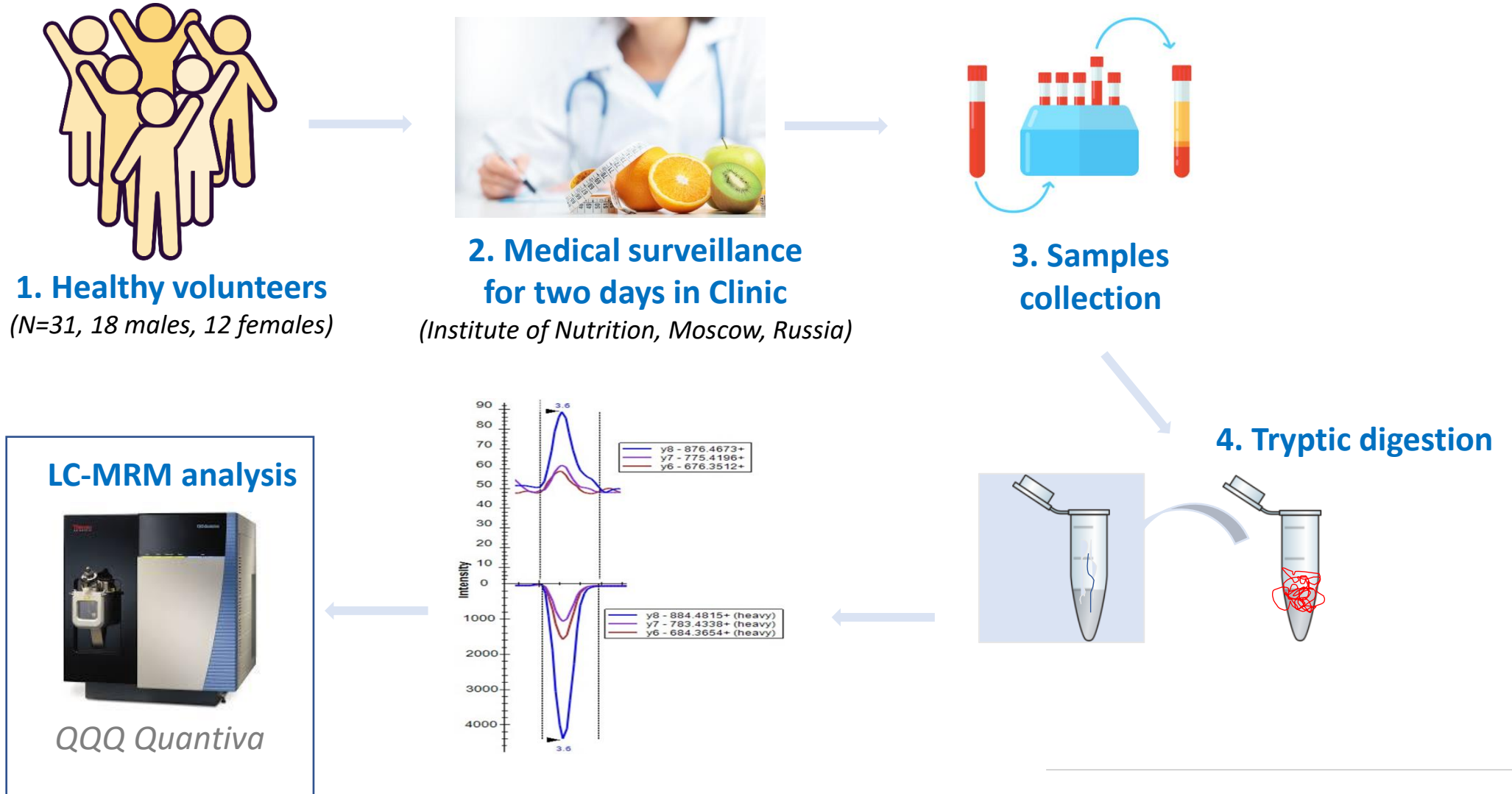
Results: In all experimental samples (n = 31), 42 out of 111 target proteins (38%) were registered by using proteotypic peptides. This set of proteins can be considered as the «human plasma proteome core of a healthy person», taking into account items with InterV<40%.

Technical variability (TechV): CV for each peptide, measured in three technical replicates (<20%)
*Biological variability (InterV)*: CV between measurements performed in 31 individual samples (<40%)

A. Distribution of InterV for 42 FDA-verified proteins
The largest scatter was observed for haptoglobin (68%), immunoglobulin heavy constant delta IGHD (90%), angiotensin (72%), sex hormone-binding globulin SHBG (100%), and lipoprotein-(a) (136%).
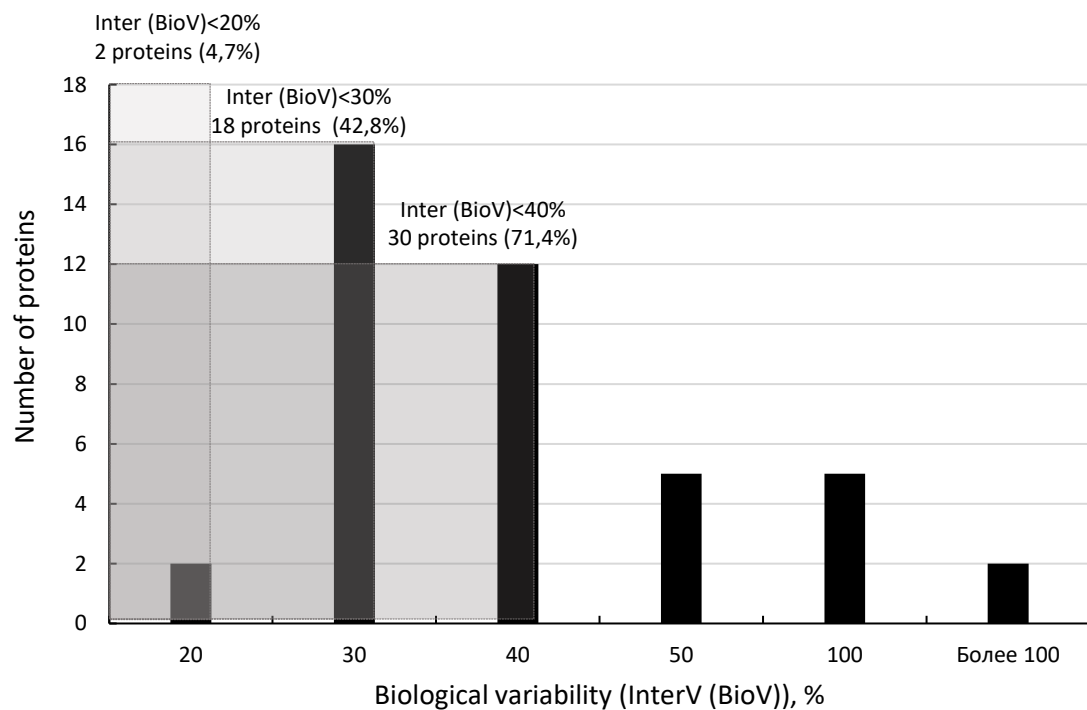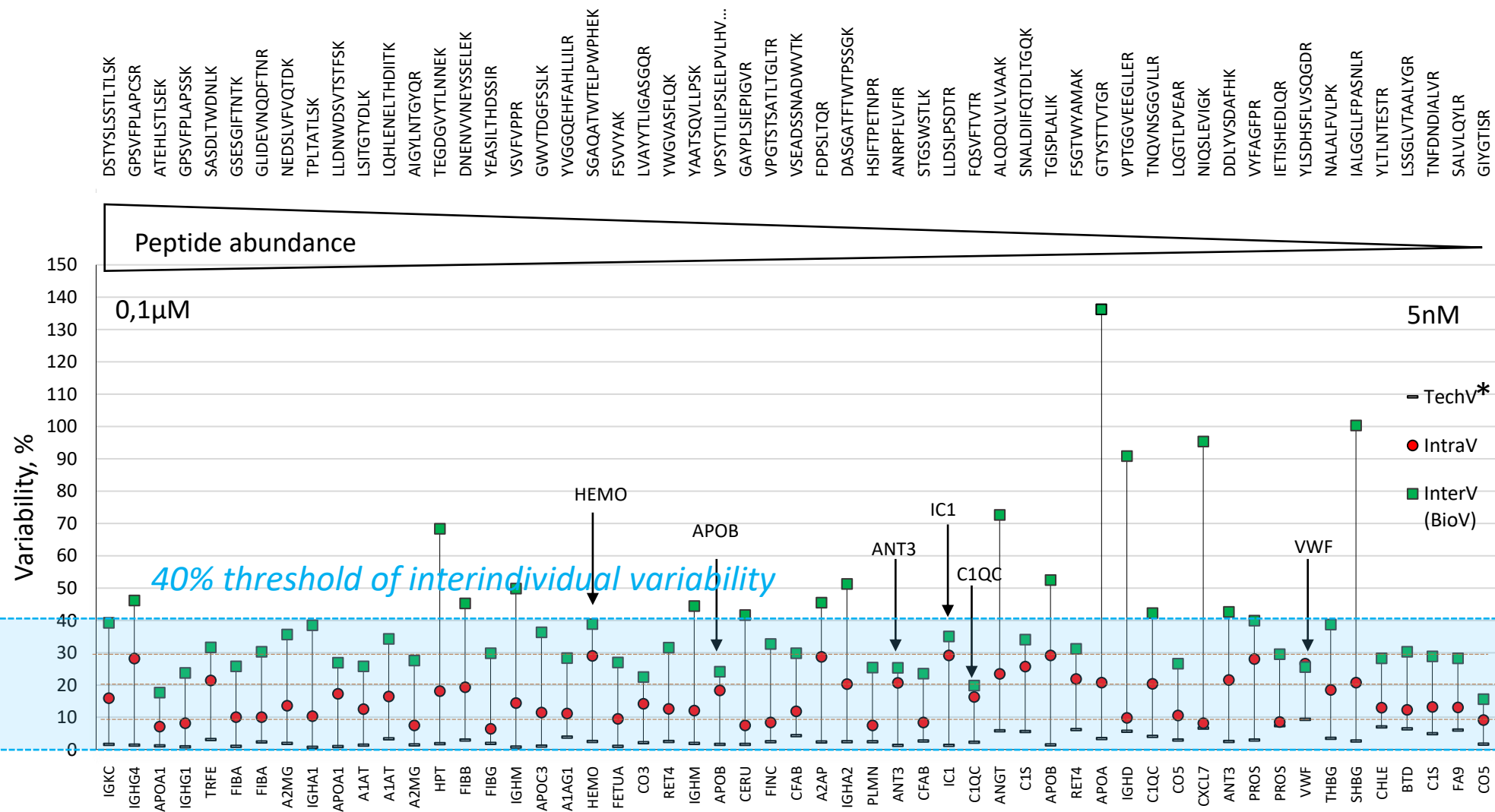
B. Boxplot for the content of 42 proteins of the plasma proteome core
The MIN and MAX concentrations were determined for coagulation factor IX (FA9) (12± 3.4*10^{-9}M) and the immunoglobulin kappa light chain constant region (IGKC) (0.1 ± 0.04*10^{-6}M), respectively.
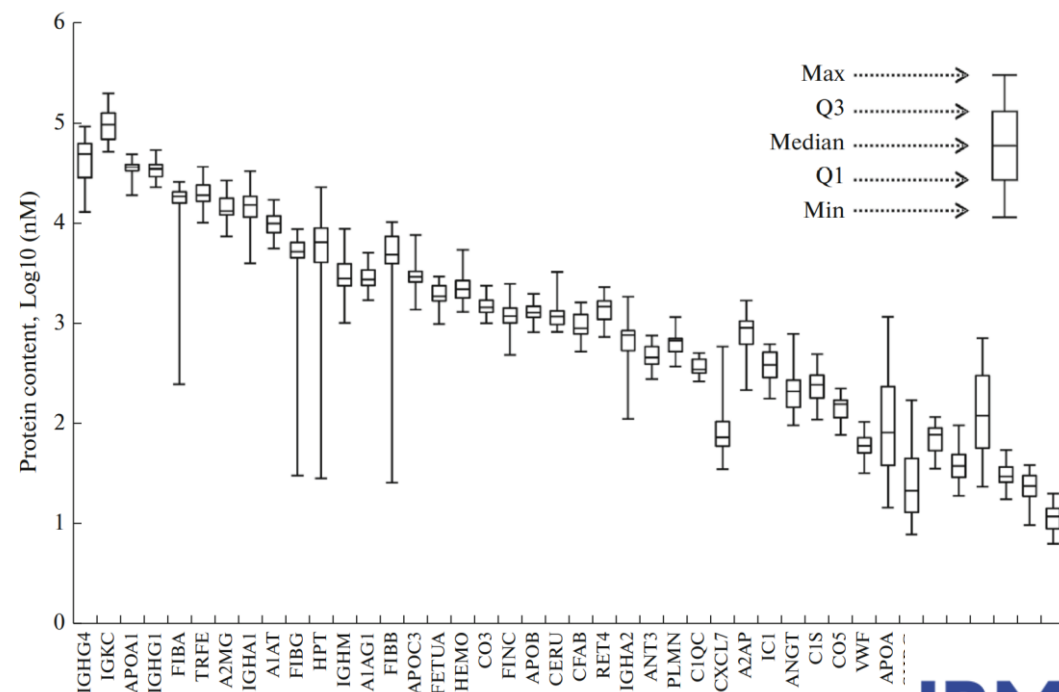
Technical (TechV) and interindividual (InterV) variability for 55 peptides mapped to 42 proteins detected in all experimental samples (n = 31)
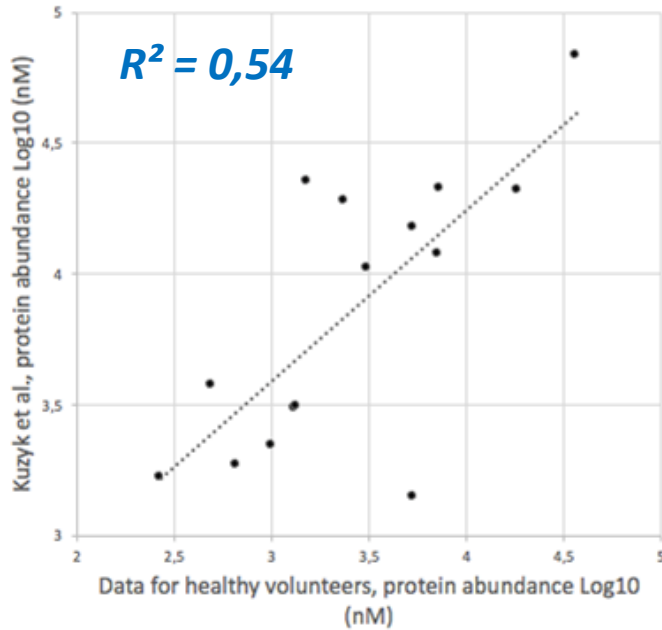
Novikova et al., Biomed Khim. 2020 Jul;66(4):294-316.

Comparison of quantitative data on healthy volunteers blood plasma proteome with literature data

The obtained concentration of proteins compared with published data

(a) Validation of quantitative data obtained using MRM

(b) Validation of quantitative data by other methods

$R^2 = 0,54$

$R^2 = 0,66$

$R^2 = 0,84$

Kuzyk et al., Mol Cell Proteomics, 2009 (n=15)
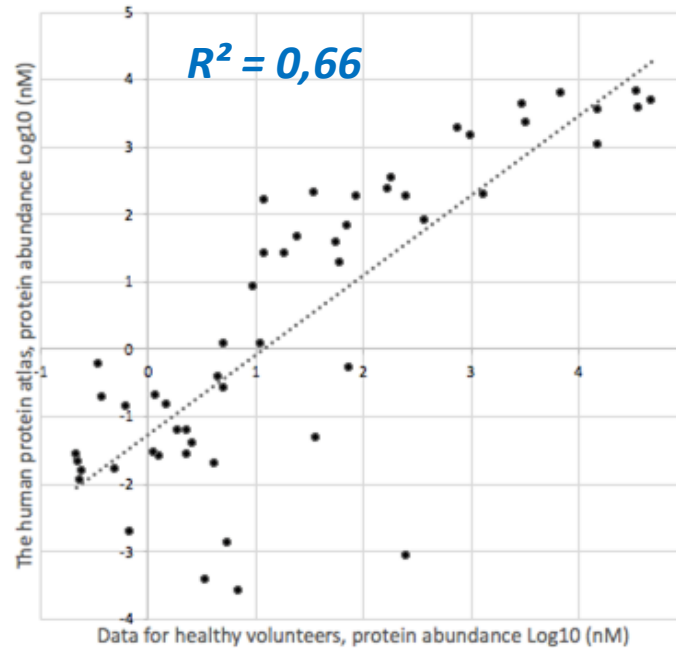
The Human Plasma Protein Atlas Data (n=53)
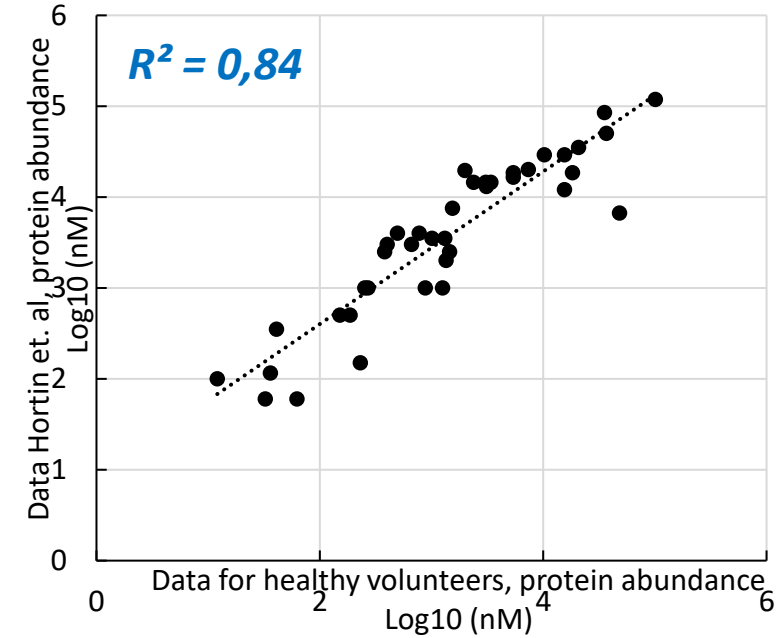
Hortin et al., Clinical Chemistry, 2008 (n=38)

Novikova et al., Biomed Khim. 2020 Jul;66(4):294-316.

IBMC
Institute of Biomedical Chemistry

Is it possible to apply this approach to any set of proteins?

Part 2. Targeted Quantitative SRM SIS Screening of Chr 18, 13, Y and the Mt Encoded Proteome **(600+ target proteins)**

Baseline metrics and meta-analysis data

Team	Chr	Number of protein-coding genes	Number of detected proteins		Number of measured proteins
			Human PeptideAtlas	Human Plasma PeptideAtlas	Plasma Proteome DB (PPDB*-2019)
Russia	18	275	236	62	22
Korea	13	333	283	58	16
Iran	Y	47	12	1	0
Italy	MT	15	13	1	0
	Total	**670**	**544**	**122**	**38**

Number of measured in blood plasma proteins encoded by the selected chromosomes was limited to only 38 proteins, according to the meta-analysis presented in PPDB.

IBMC
Institute of Biomedical Chemistry

Protein Concentrations in Healthy Human Blood Plasma:
Targeted Quantitative SRM SIS Screening revealed 205 proteins (30.7% out of 667 genes) were measured at least in one of 54 blood plasma samples of the volunteers

Kopylov et al., J Proteome Res. 2019, 18, 120-129.

Targeted quantitative screening: **The distribution of the concentrations of the proteins** encoded by chromosomes 18, 13, Y and MT in the whole (*n = 147*) and depleted (*n = 142*) blood plasma of a healthy person (*n=54*)

(A) The distribution of the concentrations of the proteins encoded by selected chromosomes
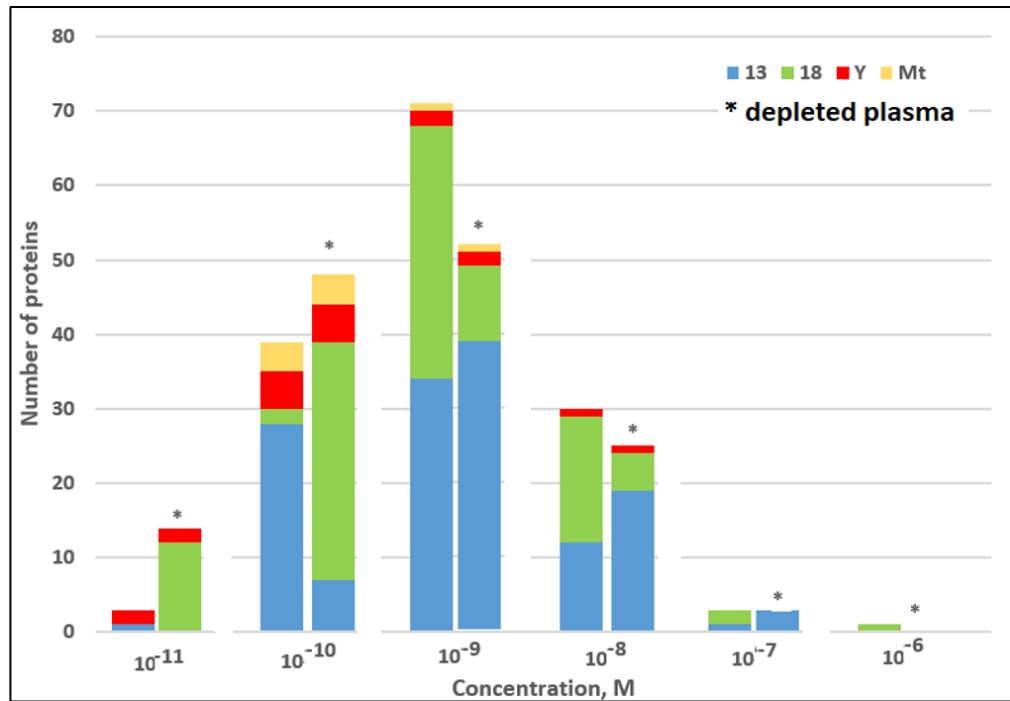
(B) The maximum difference in the concentration for the same protein is about 1.5 orders of magnitude
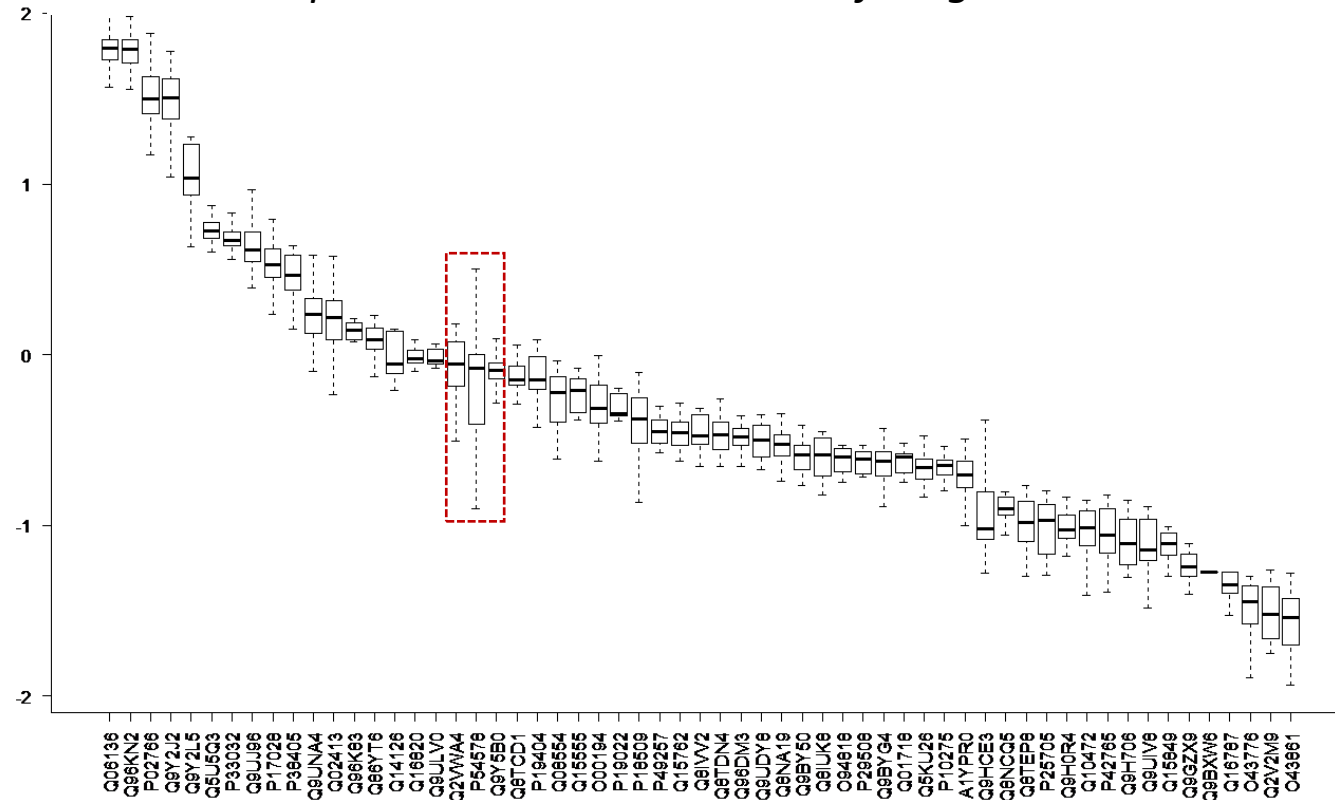
54 males (age 20-47) were examined and approved for space-related simulations and experiments (IMBP RAS, Moscow, Russia)

Kopylov et al., J Proteome Res. 2019, 18, 120-129.

IBMC
Institute of Biomedical Chemistry

Healthy Human Blood Plasma:
Inter-individual coefficient of variance (CV)

Set of targeted proteins for health analytics

✓ Presented in 70%+ «healthy» samples;
✓ Inter-individual CV ≤40%
✓ Obtained ranges of concentration for healthy people!!!

Kopylov et al., J Proteome Res. 2019, 18, 120-129.

■ Whole ■ Depleted

Number of Proteins

Inter-individual CV
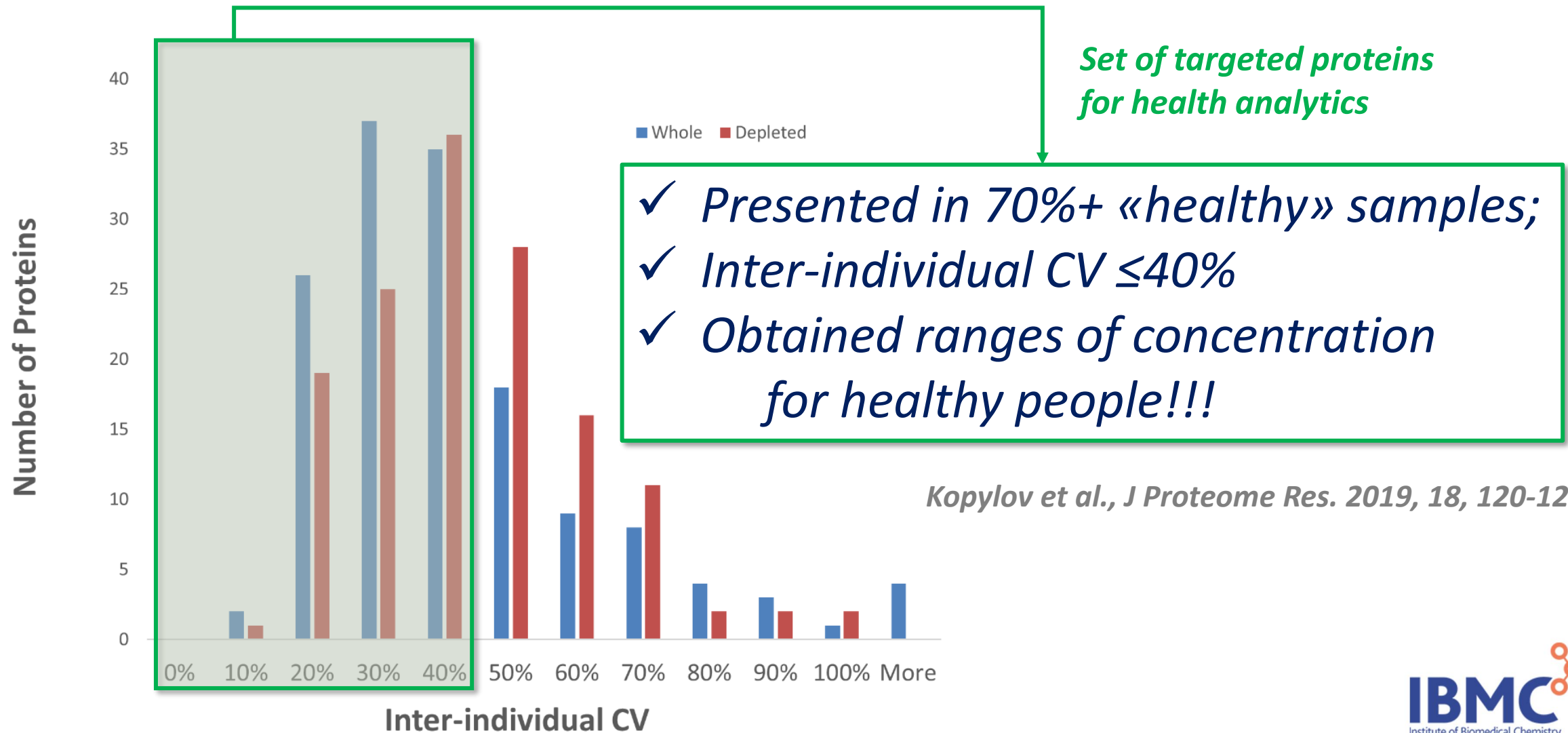
Average Concentration of Frequently Detected Proteins

entry	protein names	gene name	whole plasma		depleted plasma	
			number of samples	average concentration, M	number of samples	average concentration, M
P02766	transthyretin	TTR	54	7.1×10^{-6}	54	6.2×10^{-8}
P00742	coagulation factor X	F10	54	1.7×10^{-7}	54	1.8×10^{-7}
P08709	coagulation factor VII	F7	54	1.9×10^{-8}	54	1.8×10^{-8}
P11279	lysosome-associated membrane glycoprotein 1	LAMP1	54	9.5×10^{-9}	154	9.7×10^{-9}
Q14126	desmoglein-2	DSG2	48	1.1×10^{-8}	53	8.2×10^{-10}
Q9Y2J2	band 4.1-like protein 3	EPB41L3	47	2.5×10^{-7}	53	3.4×10^{-8}
Q96IY4	carboxypeptidase B2	CPB2	54	7.5×10^{-11}	53	2.3×10^{-9}
Q06136	3-ketodihydrosphingosine reductase	KDSR	52	8.9×10^{-8}	52	6.0×10^{-8}

Kopylov et al., J Proteome Res. 2019, 18, 120-129.

There was no correlation between protein abundances and corresponding number of samples in which this protein was detected.
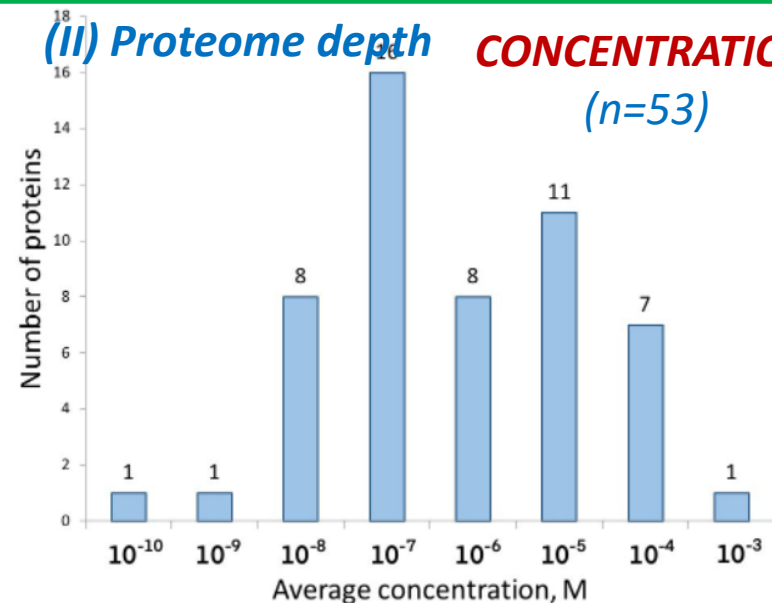
IBMC
Institute of Biomedical Chemistry

Set of proteins, selected as targeted for health analytics according to SRM-screening 700+ proteins (FDA and Chrs) in healthy human samples

(I) Proteome Width

53 PROTEINS
(CV<40%)
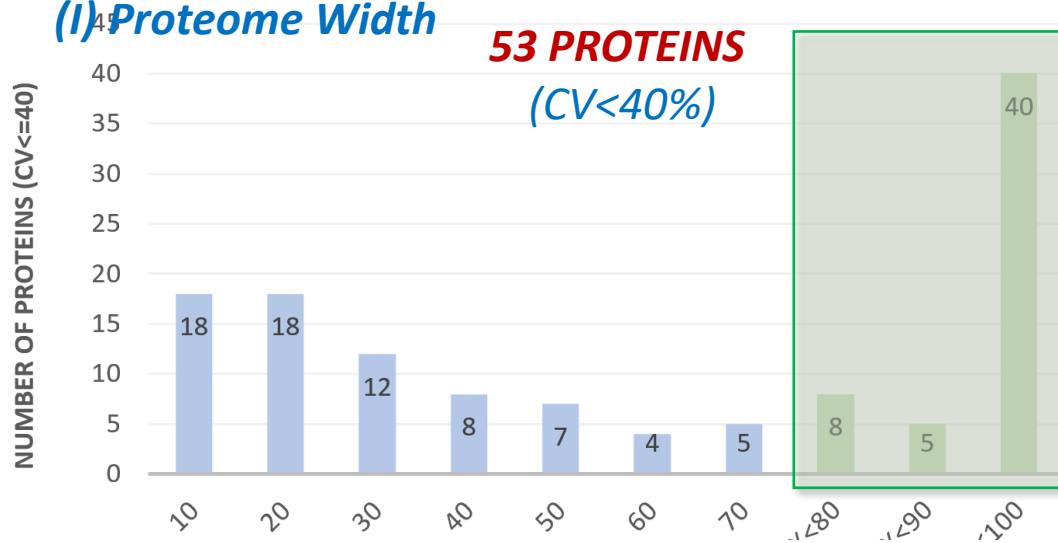
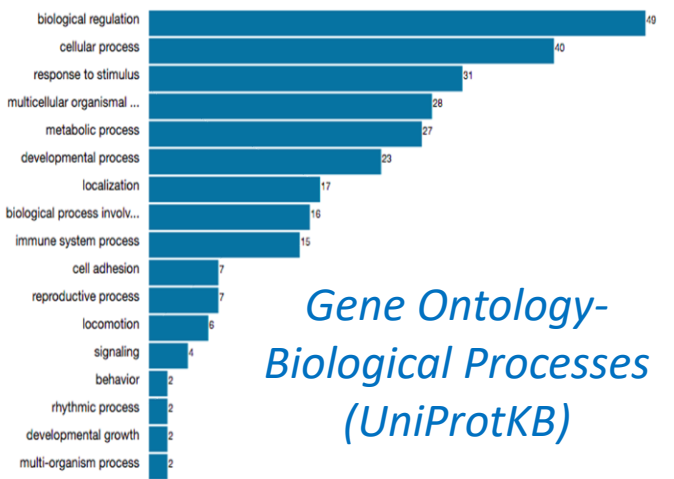*Part of the samples in which the protein was detected, %*

(II) Proteome depth

CONCENTRATION
(n=53)

(III) Functional annotations

Protein- Diseases Associations (UniProtKB)

Gene Ontology- Biological Processes (UniProtKB)
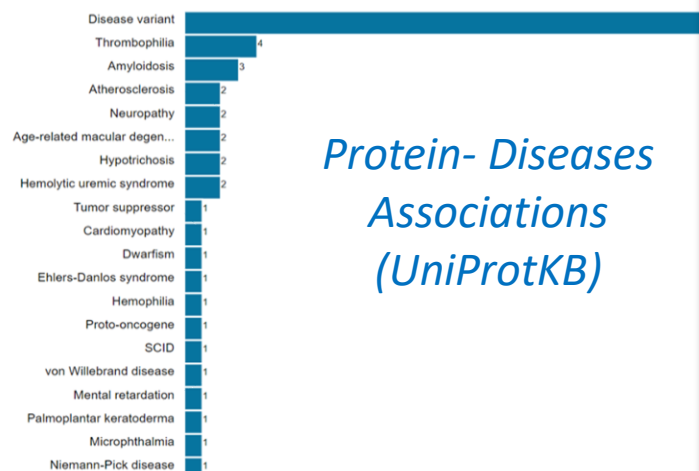
- *Gene-Diseases Association (DisGeNet)*

- *Text-mining (ScanBious)*

- *Any other ANNOTATIONS....*

IBMC
Institute of Biomedical Chemistry

Targeted mass-spectrometry provides unified quantitative data that could be analyzed and visualized in a same way

suspicious samples
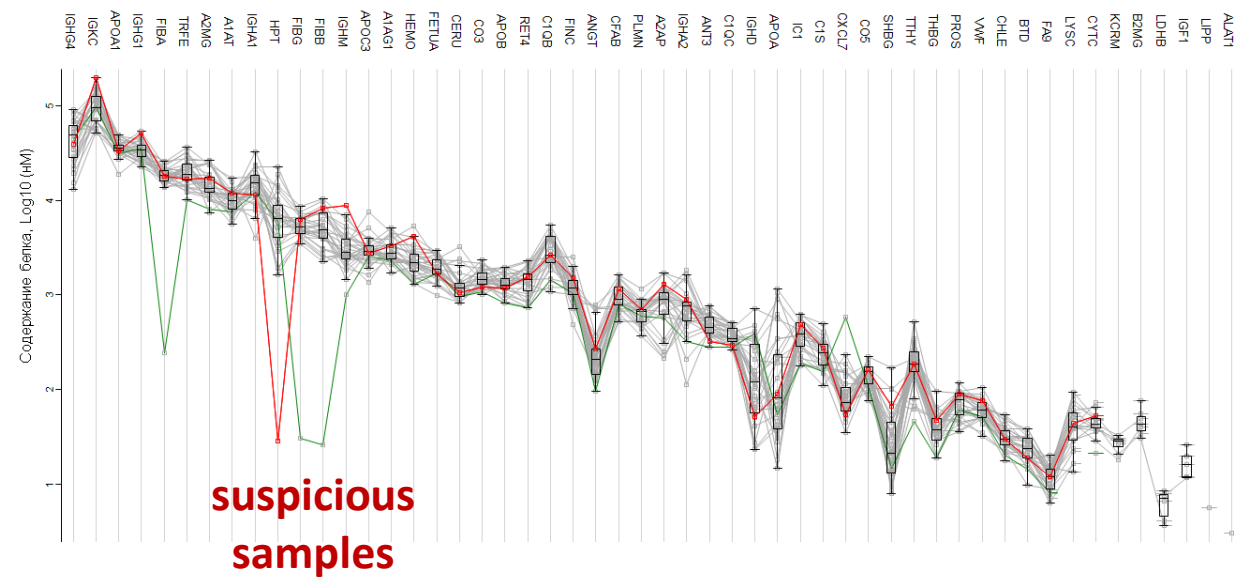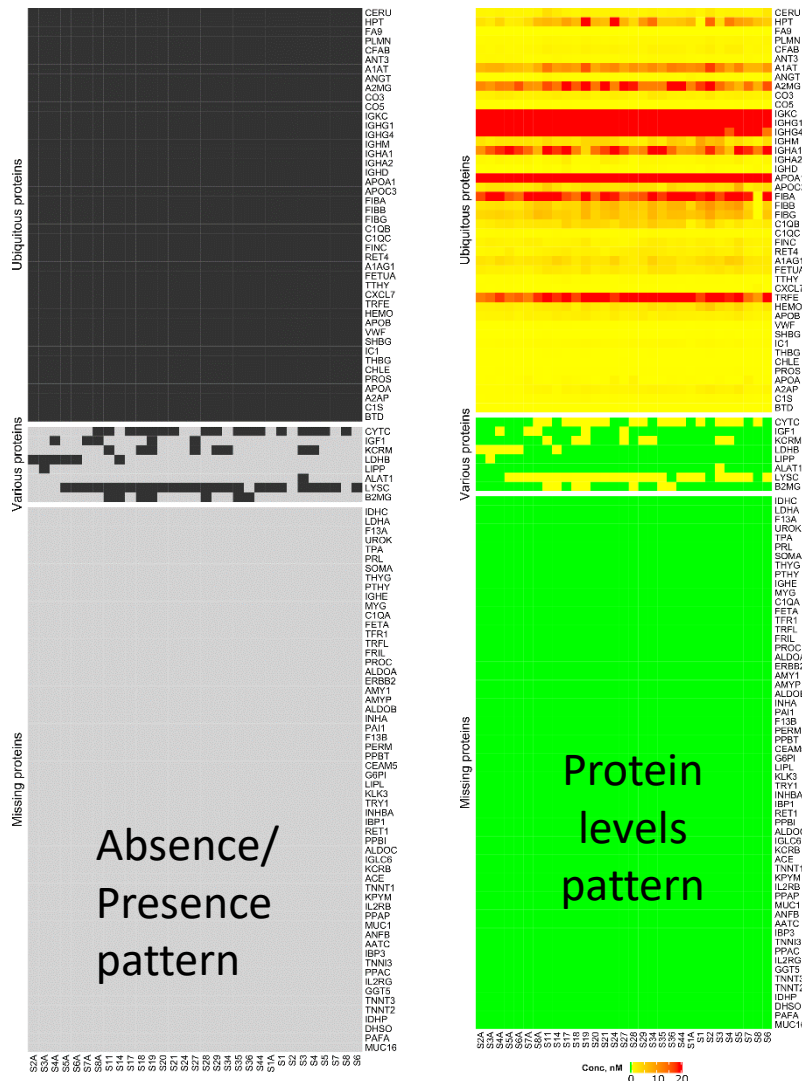
It will be of help to shift the paradigm of clinical diagnostics from numerous antibody-based tests, which could be combined in bundles, toward truly multiplex tests.

Such an approach will underlie the development of the digital image of the human plasma proteome.

Absence/ Presence pattern

Protein levels pattern

Conc, nM
0   10   20

Comparison of protein concentrations in normal and pathological conditions:
protein levels could differ by several orders of magnitude

Protein name	Gene name	Log 10 (average concentration, fM)		
		Healthy volunteers (n=85)	Neurodegenerative disease (n=19, Kiseleva et al., Clin Trans Med, 2015)	Lung adenocarcinoma (n=102, Wu et al., Proteomics Clin Appl, 2020)
Alpha-1-antitrypsin	A1AT	10,0	9,5	
Hemopexin	HEMO	9,4	8,25	
Alpha-2-macroglobulin	A2MG	10,2	8,5	
Serotransferrin	TRFE	10,3	9,3	
Haptoglobin	HPT	9,9	9	
Apolipoprotein A-I	APOA1	10,6	9,1	
Fibrinogen alpha chain	FIBA	10,3	8,9	
Transthyretin	TTR	9,9	8,5	
Complement C3	CO3	9,2	8,4	
Plasma protease C1 inhibitor	IC1	8,6	7,5	
von Willebrand factor	VWF	7,8	7	
Platelet basic protein	CXCL7	8,0	7,8	
Ceruloplasmin	CERU	9,1	6,5	
Lambda-crystallin homolog	CRYL1	6,6		6,2
Desmoglein-2	DSG2	7,0		6,4

The maximum difference in the concentration for the same protein among healthy samples is about 1.5 orders of magnitude.

Future directions:
(1) *Longitudinal studies of healthy volunteers*
(2) *Dependence on gender, age, lifestyle*
(3) *Proteomic profiles for pathological conditions*

Review > Brain Res Brain Res Rev. 2005 Nov;49(3):633-40. doi: 10.1016/j.brainresrev.2005.03.003.

Ceruloplasmin in neurodegenerative diseases

Ceruloplasmin Deamidation in Neurodegeneration: From Loss to Gain of Function

Alan Zanardi [1], Massimo Alessio [1]

Affiliations + expand

PMID: 33440850   PMCID: PMC7827708   DOI: 10.3390/ijms22020663

Free PMC article

Abstract

Neurodegenerative disorders can induce modifications of several proteins; one of which is ceruloplasmin (Cp), a ferroxidase enzyme found modified in the cerebrospinal fluid (CSF) of neurodegenerative diseases patients. Cp modifications are caused by the oxidation induced by the pathological environment and are usually associated with activity loss. Together with oxidation,

IBMC
Institute of Biomedical Chemistry

Pipeline for biomarker panel development

BACKGROUNG

ELABORATION #1

ELABORATION #2

Meta-list of "stable" proteins detected and quantified in healthy volunteers' samples
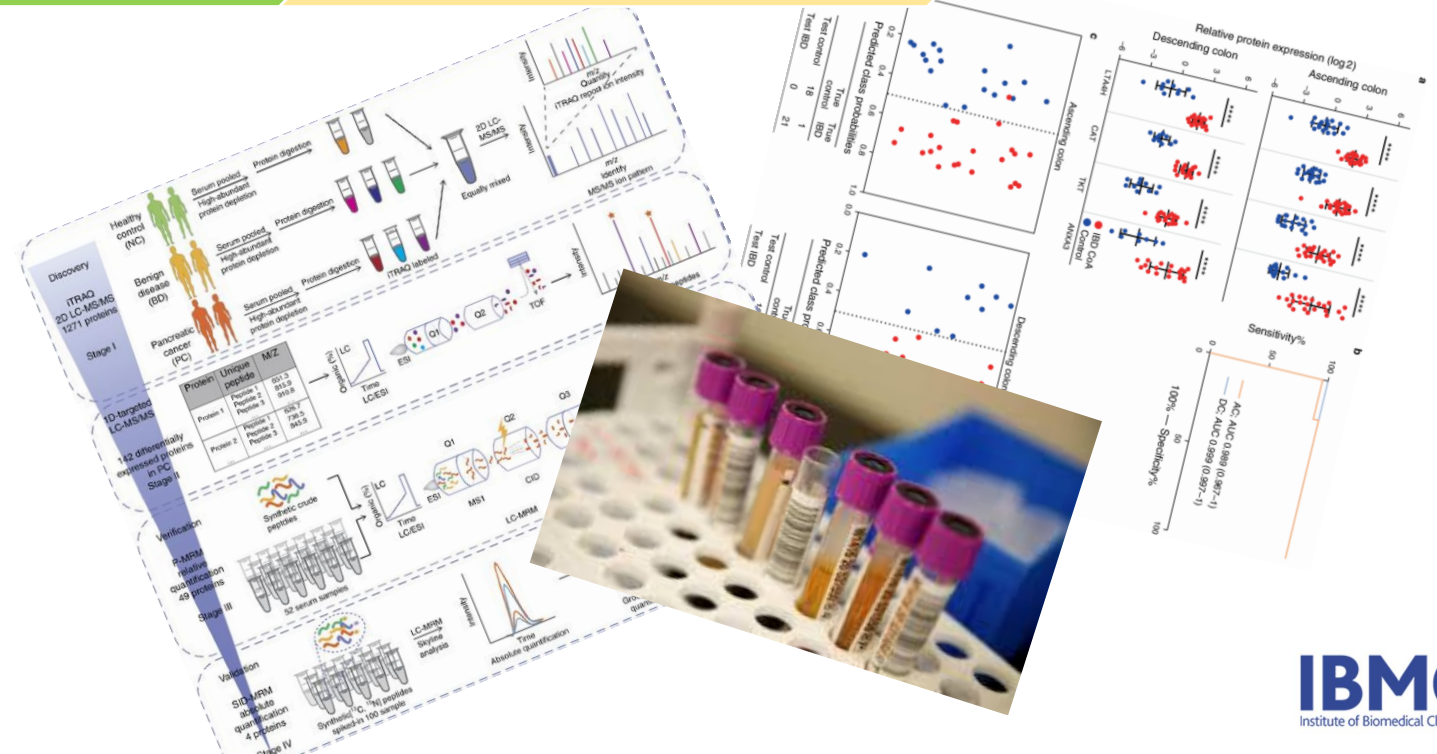
Public and custom proteomic data on the certain organ or tissue

Extract of text-mining focused on the certain pathology

"Test piece" of biomarker panel for lab validation

IBMC
Institute of Biomedical Chemistry

Home-message

- The concentration for 248 proteins was accurately measured with SRM SIS assay in healthy humans (Chr 18, 13, Y, Mt and FDA-approved proteins).
- The concentration range covered by the SRM SIS technology was six orders of magnitude (from 10^{-6} to 10^{-11} M); there was no correlation between protein abundances and corresponding number of samples in which this protein was detected.
- Among 700+ targeted proteins only 53 could be used as a pillar for creation SRM-assays for personal health analytics (inter-individual CV ≤40%, technical variability <20%, detected in more that 70% samples of healthy persons).
- Targeted proteomics provides opportunities for standardized multiplexed measurements of the absolute concentration of targeted proteins.
- Further work in this area includes expanding the concentration range and determining the proteins concentration ranges in various human physiological states, as well as standardizing the procedures for sample collection, preparation and data-analysis.

Acknowledgements

Alexander Archakov

Institute of Biomedical Chemistry
Institute of Medico-Biological Problems
Institute of Nutrition

HUMAN PROTEOME PROJECT

TRANSLATING THE CODE OF LIFE   HPP   twitter | linkedin

- *Chr 13 (Korea)*
- *Chr Y (Iran)*
- *Chr MT (Italy)*
- *Chr 18 (Russia)*

IBMC "Human Proteome" Core Facility and "Avogadro" large-scale research facilities were used for the generation of mass-spectrometry data.

Mail to: 2463731@gmail.com