MEDICAL AND BIOLOGICAL ASPECTS RELATED TO ASSESSMENT OF IMPACTS EXERTED BY RISK FACTORS

UDC 57.044; 616.092
DOI: 10.21668/health.risk/2022.2.16.eng

Read online

Research article

CREATING BIOINFORMATICS MATRIX OF MOLECULAR MARKERS TO PREDICT RISK-ASSOCIATED HEALTH DISORDERS

M.A. Zemlyanova[1,2,3], N.V. Zaitseva[1], Yu.V. Koldibekova[1], E.V. Peskova[1,2], N.I. Bulatova[1]

[1]Federal Scientific Center for Medical and Preventive Health Risk Management Technologies, 82 Monastyrskaya Str., Perm, 614045, Russian Federation
[2]Perm State University, 15 Bukireva Str., Perm, 614990, Russian Federation
[3]Perm National Research Polytechnic University, 29 Komsomolsky Ave., Perm, 614990, Russian Federation

Long-term or permanent chemical ambient air pollution in residential areas is among priority factors that cause medical and demographic losses. It is necessary to achieve greater precision when assessing risks of changes in homeostasis at their early reversible stage (molecular level). These changes are highly likely to transform into pathological processes at an older age in case the exposure persists.

Our research goal was to create a bioinformatics matrix of molecular markers to predict risk-associated health disorders (exemplified by a marker of exposure). We introduced a stepwise research algorithm that involved using the proteome technology to identify expressed proteins and cause-effect relations between them and influencing factors; revealing molecular-cellular and functional relationships within the "exposure factor – gene – protein – negative outcome" system to predict risk-associated health disorders. The algorithm was implemented to examine the proteomic blood plasma profile of children aged 3–6 years living under long-term aerogenic exposure to fluoride-containing compounds.

We established certain changes in the proteomic profiles of the exposed children in comparison with non-exposed ones as per 27 identified proteins. A bioinformatics matrix was created on the example of cathepsin L1; we established that changes in the level of this protein had a cause-effect relationship with fluoride ion concentrations in urine. Qualitative synthesis of molecular-cellular localization, functional and tissue belonging showed that cathepsin L1 expression caused by elevated fluoride ion levels in urine could affect extracellular matrix remodeling, degradation and post-translation modification of proteins in cells of the lungs, large intestine, and pancreas, in cardiomyocytes and in glomerular podocytes. It also mediated proteolysis of the subunits of the SARS-CoV-2 S1 protein necessary for the virus penetration into a cell and its replication. This created bioinformatics matrix exemplified by cathepsin L1 made it possible to predict risk-associated negative outcomes in exposed people including cardiomyopathy, colitis, glomerulonephritis, diabetes mellitus, atherosclerosis, and coronavirus infection. These predictive estimates raise effectiveness of early detection and development of preventive measures aimed at minimizing possible negative outcomes.

Keywords: *proteomic profile, children, cellular-molecular and tissue belonging, molecular markers, gene expression, negative effects, fluoride ion in urine, cathepsin L1, bioinformatics resources.*

Marina A. Zemlyanova – Doctor of Medical Sciences, Chief Researcher acting as the Head of the Department of Biochemical and Cytogenetic Methods of Diagnostics; Associate Professor at the Department of Microbiology and Virology (e-mail: zem@fcrisk.ru; tel.: +7 (342) 236-39-30; ORCID: http://orcid.org/0000-0002-8013-9613).

Nina V. Zaitseva – Academician of the Russian Academy of Sciences, Doctor of Medical Sciences, Professor, Scientific Director (e-mail: znv@fcrisk.ru; tel.: +7 (342) 237-25-34; ORCID: http://orcid.org/0000-0003-2356-1145).

Yulia V. Koldibekova – Candidate of Biological Sciences, Senior Researcher, the Head of the Laboratory for Metabolism and Pharmacokinetics at the Department for Biochemical and Cytogenetic Diagnostic Techniques (e-mail: koldibekova@fcrisk.ru; tel.: +7 (342) 237-18-15; ORCID: http://orcid.org/0000-0002-3924-4526).

Ekaterina V. Peskova – Junior Researcher at the Laboratory of Biochemical and Nanosensory Diagnostics at the Department for Biochemical and Cytogenetic Diagnostic Techniques (e-mail: peskova@fcrisk.ru; tel.: +7 (342) 237-18-15; ORCID: https://orcid.org/0000-0002-8050-3059).

Natalya I. Bulatova – Researcher at the Laboratory of Biochemical and Nanosensory Diagnostics (e-mail: 1179815@mail.ru; tel: +7 (342) 236-80-18; ORCID: https://orcid.org/0000-0003-3392-9097).

Health preservation and improvement, a growth in population life expectancy, and a decrease in mortality caused by variable diseases, including diseases of the cardiovascular system, malignant neoplasms and chronic respiratory diseases, are established as national priorities within the state policy in the sphere of socio-economic and demographic development in the Russian Federation for the period up to 2024.

According to the WHO experts, long-term or permanent exposure to chemical pollution in ambient air, water and soils is among leading risk factors causing additional adverse health outcomes[1] [1]. Overall, the burden of diseases caused by complex chemical loads is estimated to account for 15–35 %. If we want to reduce health losses among people who are exposed to chemical pollution (first of all, the most sensitive population groups such as children, teenagers, and young people), we should achieve greater precision when trying to predict risks of changes in homeostasis. It is especially vital when it comes down to early reversible stages in such changes (at the molecular level) since they are very likely to transform into pathological processes at an older age if exogenous introduction of chemicals persists.

The highly informative proteome technology can be a useful tool for searching for potential molecular markers indicating the dynamic balance has been impaired. This technology involves examining protein compositions and thereby makes it possible to identify changes resulting from expression of genes that code target proteins (DNA-RNA-protein relations) under exposure to harmful factors, including chemical ones [2–5]. It is most advisable to use this approach since proteins bear the major functional loads when a living organism interacts with the environment (either directly or through their enzymatic activities). Identification of molecular protein and peptide targets, determination of their structure, functions, tissue belonging, and their involvement into pathogenesis of functional disorders should be accomplished by using up-to-date bioinformatics and toxicogenetic resources [6–10]. When done, this guarantees effective predictions of expected negative health outcomes. Identifying associative pathogenetic relations between effects produced by exposure factors and expression of protein molecular markers makes it possible to predict risk-associated negative outcomes. It is especially vital for achieving greater effectiveness of early (cellular and molecular) detection of health disorders in exposed population.

Given all the above-stated, it seems advisable to elaborate on approaches to performing studies with their focus on identifying expressed proteins associated with exposure to chemical factors. It is also important to perform biological and biochemical interpretation of their cellular and functional belonging since it helps examine molecular-cellular mechanisms in detail and achieve more precision in estimating probable negative health outcomes participating in pathogenesis of risk-associated non-communicable diseases. All this gave grounds for determining the goal of the present research.

Our research goal was to create a bioinformatics matrix of molecular markers to predict risk-associated health disorders (exemplified by a marker of exposure).

Materials and methods. It is possible to create a bioinformatics matrix of molecular markers to predict risk-associated health disorders by using the following stepwise research algorithm:

– creating samplings made of exposed and non-exposed people from population groups that are the most sensitive to negative effects produced by chemical factors (a test group and a control group);

– establishing actual exposure by determining elevated contents of chemicals in biological media that are proven to be associated with this exposure (identifying markers of exposure);

[1] O sostoyanii sanitarno-epidemiologicheskogo blagopoluchiya naseleniya v Rossiiskoi Federatsii v 2020 godu: Gosudarstvennyi doklad [On sanitary-epidemiological welfare of the population in the Russian Federation in 2020: The State Report]. Moscow, The Federal Service for Surveillance over Consumer Rights Protection and Human Wellbeing, 2021, 256 p. (in Russian).

– creating a proteomic profile, identification, comparative analysis and spotting out protein strains that differ in their intensity in a test group against a control one and these differences are statistically significant;

– establishing cause-effect relations between identified expressed proteins and exposure factors (as per contents of markers of exposure in biological media);

– qualitative synthesis and creation of a phylogenetic tree that reflects molecular-cellular and functional relations within the "exposure factor – gene – protein – negative outcome" system to predict risk-associated health disorders.

The suggested algorithm was implemented in a study that focused on proteomic blood plasma profiles of children aged 3–6 years who lived under long-term aerogenic exposure to gaseous and solid fluoride compounds. This exposure created non-carcinogenic health risks that were by 1.5–2 times higher than their acceptable levels. We made our test group by including children with elevated fluoride-ion contents in urine that were by 1.5 or more times higher than the reference value[2]. This indicator was selected due to being the marker of exposure to fluoride compounds according to the previously established cause-effect relations. Children from the control group had the examined chemical in their urine in contents not exceeding the reference value or being close to it. The children were examined in full conformity with the ethical principles stated in the WMA Declaration of Helsinki (Ethical Principles for Medical Research Involving Human Subjects, 2013). The study was approved by the Committee on Biomedical Ethics of the Federal Scientific Center for Medical and Preventive Health Risk Management Technologies; it was mandatory to obtain an informed voluntary consent from legal representatives of all the examined children prior to any examination.

Proteomic profiles were examined by using two-dimensional gel electrophoresis in polyacrylamide gel. The resulting blood plasma electrophoretograms were visualized by dyeing with silver and documented with a gel documentation system (BioRad, USA). The obtained proteomic maps were comparatively analyzed with PDQuest software package (BioRad, USA); we spotted out significant protein stains as per their intensity and performed subsequent analysis with liquid chromatography together with mass-spectral analysis (with UltiMate 3000 chromatograph (Germany) and AB Sciex 4000 QTRAP tandem mass spectrometer with Nanospray 3 ionization source (Canada)). Data produced by tandem examinations were analyzed with ProteinPilot, verison 4.5 (AB Sciex) and identified according to UniProt_sprot_fasta database (as of November 24, 2017), with sampling as per Homo Sapience taxon (peptide fingerprint). Major part of the information about the obtained protein was extracted from two databases, Gene Ontology Resource and UniProt[3]. We established genes that coded expression of the identified proteins by using HGNC database of human gene name[4].

All the obtained data were statistically analyzed with Statistica 10 software package. We comparatively assessed relative volumes of protein stains in the test group against the same parameters in the control one. The study results are given as the mean value ($\bar{X}$) and the standard error of the mean (*SEM*). Statistical significance of differences in variables between the groups was determined as per Mann – Whitney test ($U \leq Ucr$) with the significance level set at $p \leq 0.05$.

We detected cause-effect relations between changes in statistically significantly different protein strains and fluoride-ion concentrations and assessed them by building a mathematical logistic regression model. The built models were estimated to test their au-

[2] Tits N.U. Klinicheskoe rukovodstvo po laboratornym testam [The clinical guide on laboratory tests]. Moscow, YuNIMED-press, 2003, 960 p. (in Russian).

[3] Gene Ontology Resource. Available at: http://geneontology.org/ (April 08, 2022); UniProt. Available at: http://www.uniprot.org (April 08, 2022).

[4] HGNC database of human gene name. Available at: https://www.genenames.org (April 12, 2022).

thenticity and relevance; the estimation was based on dispersion analysis with using Fischer's F-test, determination coefficient (*R2*), and Student's t-test at the statistical significance level being $p \leq 0.05$.

Proteins were classified as per classes, biological and molecular functions by using Panther Classification System and Gene Ontology and GO Annotations databases[5]. To establish expression of proteins in various tissues in a body, we used data provided by Tissue expression database and The Human Protein Atlas[6]. Relationships within the "exposure factor (marker of exposure) – gene – protein – disease" system were described with the use of Comparative Toxicogenomics as a source of information[7].

Results and discussion. We comparatively analyzed the results produced by densitometry of proteomic blood plasma profiles of the examined children. The analysis revealed authentic differences in relative volumes of 27 protein stains between the children from the test and control groups (Table 1).

Approximately 75 % of the detected protein stains were by 1.4–37.6 greater in volumes in children from the test group than in their counterparts from the control one; the remaining 25 % had by 1.5–36.6 times smaller volumes ($p = 0.0001–0.009$).

Figure 1 shows a fragment of a two-dimensional electrophoretogram used for detecting quantitative differences in cathepsin L1 in blood plasma of the examined children.

We performed bioinformatics analysis to detect how the identified proteins were localized; to do that, we used Panther classification. The analysis established that most of them (57.7 %) were localized in cellular structures (GO:0110165) and 15.4 % of the proteins were localized in protein complexes (GO:00032991). The proteins were classified as follows depending on their participation in the life cycle: 22.7 % proteins were signal transfers/carriers (PC00219), 13.6 % metabolite interconversion enzymes (PC00262), and also three groups each containing 9.1 % of the proteins were protein modifying enzymes (PC00260), cytoskeletal proteins (PC00085) and adaptor proteins (PC00226) (Table 2 and Figure 2).

We analyzed class properties as per molecular processes and established that most proteins were responsible for binding with other molecules (GO:0005488; 44.8 %) and catalytic activity (GO:0003824; 31.3 %) (Table 3 and Figure 3).

We examined and classified biological functions performed by the identified proteins. The results showed that most of them were responsible for cellular (GO:0009987; 30 %) and metabolic processes (GO:0008152; 15 %) as well as biological regulation (GO:0065007; 12 %) (Table 4 and Figure 4).

The established class properties (localization, participation in the life cycle, and biological functions) of the identified proteins were identical in the children from the both groups, the test and control.

We analyzed cause-effect relations between changes in contents of 27 expressed proteins and fluoride-ion concentrations in urine. The analysis established an authentic direct correlation only for cathepsin L1 ($R^2 = 0.45$; $b_0 = 764.23$; $b_1 = 51.47$; $p = 0.016$) and this was in line with experimental data indicating that exposure to fluoride compounds led to elevated cathepsin L1 expression [11].

We took cathepsin L1 as an example protein to create a bioinformatics matrix showing functional belonging (Figure 5) using Gene Ontology and GO Annotations bioinformatics platform[8].

---

[5] Panther Classification System. Available at: http://www.pantherdb.org (March 23, 2022); Gene Ontology and GO Annotations. Available at: https://www.ebi.ac.uk/QuickGO (March 23, 2022).

[6] Tissue expression database. Available at: https://tissues.jensenlab.org/Search (March 18, 2022); The Human Protein Atlas. Available at: https://www.proteinatlas.org/ (March 18, 2022).

[7] Comparative Toxicogenomics. Available at: http://ctdbase.org/ (April 12, 2022).

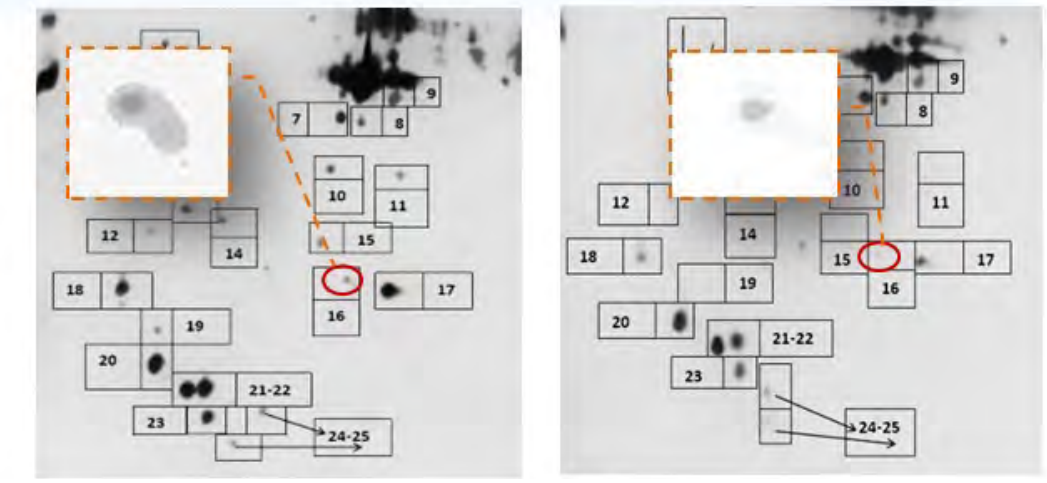
[8] Gene Ontology and GO Annotations. Available at: https://www.ebi.ac.uk/QuickGO/ (March 19, 2022).

Table 1

Proteins detected in children proteomic blood plasma profiles that are authentically different from the control group

No.	Protein	UniProt identifier	Coding gene	An average protein stain volume, $\bar{X} \pm SEM$	
				Test	Control
1	Keratin, type I cytoskeletal 9	P35527	KRT9	1406.3 ± 148.7*	2232.5 ± 272.2
2	Band 4.1-like protein 3	Q9Y2J2	EPB41L3	1890.6 ± 56.2*	2833.8 ± 135.7
3	Inositol 1,4,5-triphosphate receptor interacting protein-like 1	Q6GPH6	ITPRIPL1	1736.9 ± 56.2*	2743.4 ± 304.4
4	Hemoglobin subunit beta	P68871	HBB	2350.7 ± 161.6*	1447.9 ± 253.0
5	Laminin subunit alpha-3	Q16787	LAMA3	2435.4 ± 148.4*	66.5 ± 61.5
6	Apolipoprotein A-I	P02647	APOA1	66.5 ± 12.5*	2435.4 ± 148.4
7	Eukaryotic peptide chain release factor GTP-binding subunit ERF3B	Q8IYD1	GSPT2	2547.5 ± 118.0*	94.0 ± 81.0
8	Immunoglobulin J chain	P01591	JCHAIN	1785.0 ± 86.0*	101.0±34.0
9	THO complex subunit 2	Q8NI27	THOC2	3157.56 ± 65.4*	2983.4 ± 61.1
10	NCK-associated protein 5-like	Q9HCH0	NCKAP5L	2586.6 ± 105.1*	2840.3 ± 92.5
11	GRB10-interacting GYF protein 1	O75420	GIGYF1	2340.7 ± 183.3*	95.4 ± 61.4
12	Kelch-like protein 4	Q9C0H6	KLHL4	1806.6 ± 78.0*	100.5 ± 76.9
13	Prothrombin	P00734	F2	1893.7 ± 117.3*	1724.6 ± 47.4
14	Nucleophosmin	P06748	NPM1	469.0 ± 120.0*	94.0 ± 26.0
15	Complement C4-B	P0C0L5	C4B	2749.8 ± 56.0**	1821.2 ± 259.7
16	Cathepsin L1	P07711	CTSL	1712. 0 ± 123.7*	518.5 ± 125.5
17	Ankyrin-1	P16157	ANK1	1765.6 ± 91.0*	454.2 ± 505.1
18	Choline transporter-like protein 3	Q8N4M1	SLC44A3	1715.2 ± 123.0*	72.7 ± 71.0
19	Olfactory receptor 8A1	Q8NGG7	OR8A1	1712.1 ± 123.7*	518.5 ± 248.0
20	Transthyretin	P02766	TTR	3966.9 ± 189.3*	2344.4 ± 300.2
21	Vitronectin	P04004	VTN	3568.8 ± 233.2*	1844.1 ± 142.6
22	Tyrosine-protein phosphatase non-receptor type 14	Q15678	PTPN14	1639.1 ± 156.8*	99.5 ± 95.63
23	Transcription termination factor 1	Q15361	TTF1	3109.5 ± 75.2*	3342.2 ± 118.7
24	Apolipoprotein C-III	P02656	APOC3	4249.6 ± 250.1*	3027.4 ± 250.1
25	Myotubularin	Q13496	MTM1	4408.1 ± 263.9*	3228.0 ± 335.8
26	Apolipoprotein C-II	P02655	APOC2	3802.7 ± 272.0*	2326.7 ± 295.1
27	Serum amyloid protein A-1	P0DJI8	SAA1	2326.3 ± 154.1*	61.8 ± 5.9

N o t e : * means differences in mean values are authentic between the groups at $p = 0.0001$; ** means authentic differences in mean values between the groups at $p = 0.009$.



a *b*

Figure 1. A fragment of a two-dimensional children blood plasma electrophoretogram: a, a child from the test group; b, a child from the control group

Table 2

Classes of proteins identified in children blood plasma depending on their participation in the life cycle

Protein class*	Protein
Protein modifying enzyme	Tyrosine-protein phosphatase non-receptor type 14, Prothrombin, Cathepsin L1
Adaptor	Ankyrin-1, Kelch-like protein 4
Transfer/carrier	Apolipoprotein A-I, Apolipoprotein C-II, Apolipoprotein C-III, Hemoglobin subunit beta
Cytoskeletal protein	Band4.1-like protein 3, Keratin 9
Metabolite interconversion enzyme	Transthyretin, Inositol 1,4,5-triphosphate receptor interacting protein-like 1, Myotubularin
Transporter	Choline transporter-like protein 3
Chaperone	Nucleophosmin
Cell adhesion molecule	Laminin subunit alpha-3
Protein-binding activity modulator	Complement C4-B
Transmembrane signaling receptor	Olfactory receptor 8A1
Defense protein	Immunoglobulin J chain
Gene-specific transcriptional regulator	THO complex subunit 2
Translation protein	Eukaryotic peptide chain release factor GTP-binding subunit ERF3B

Note: * according to The PANTHER Classification System[9].

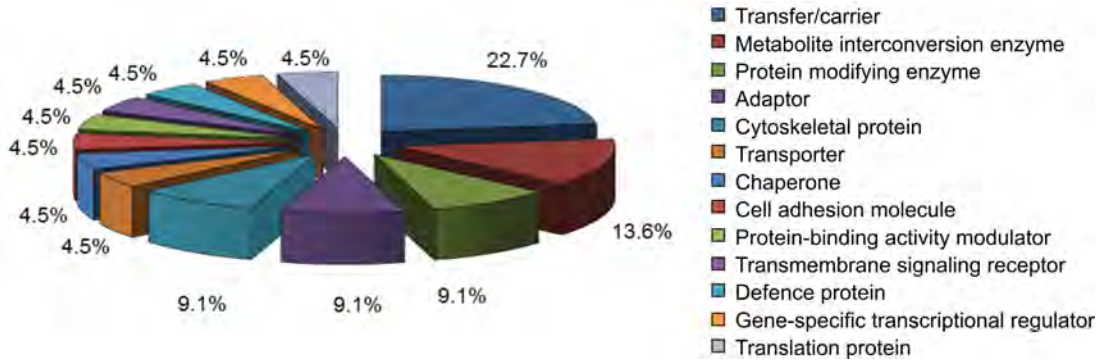


Figure 2. Classes of proteins identified in children blood plasma depending on their participation in the life cycle

Table 3

Molecular functions performed by the identified proteins in children blood plasma

Molecular function*	Protein
Molecular transducer activity	Olfactory receptor 8A1
Molecular adaptor activity	Ankyrin-1
Transporter activity	Choline transporter-like protein 3
Translation regulator activity	Eukaryotic peptide chain release factor GTP-binding subunit ERF3B
Binding	Olfactory receptor 8A1; THO complex subunit 2; Hemoglobin subunit beta; Eukaryotic peptide chain release factor GTP-binding subunit ERF3B; Nucleophosmin; Ankyrin-1; Vitornectin
Catalytic activity	Eukaryotic peptide chain release factor GTP-binding subunit ERF3B; Tyrosine-protein phosphatase non-receptor type 14; Prothrombin; Hemoglobin subunit beta; Myotubularin; Cathepsin L1

Note: * according to The PANTHER Classification System[9].

[9] The PANTHER Classification System. Available at: http://www.pantherdb.org/ (March 19, 2022).

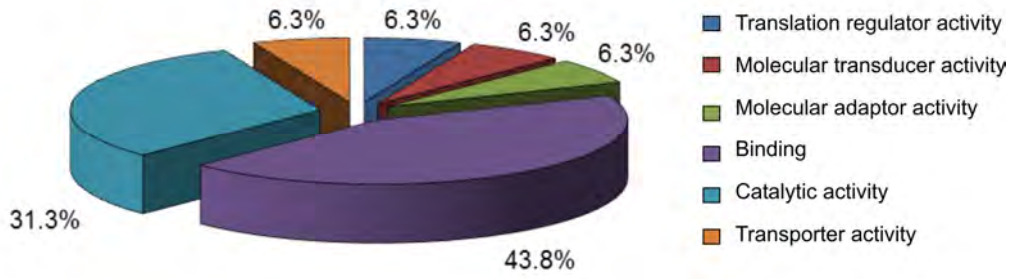


Figure 3. Molecular functions performed by the identified proteins in children blood plasma

Table 4

Biological functions performed by the identified proteins in children blood plasma

Biological function	Protein
Biological adhesion	Vitronectin
Growth	Prothrombin
Immune processes	Complement C4-B; Cathepsin L1
Biological regulation	Prothrombin; Olfactory receptor 8A1; Nucleophosmin; Complement C4-B; GRB10-interacting GYF protein 1; Myotubularin
Cellular processes	NCK-associated protein 5-like; Prothrombin; THO complex subunit 2; Olfactory receptor 8A1; Hemoglobin subunit beta; Choline transporter-like protein 3; Eukaryotic peptide chain release factor GTP-binding subunit ERF3B; Nucleophosmin; Ankyrin-1; Transthyretin; Myotubularin; Vitronectin; GRB10-interacting GYF protein 1; Cathepsin L1
Development processes	Tyrosine-protein phosphatase non-receptor type 14; Prothrombin; Myotubularin;
Localization	THO complex subunit 2; Choline transporter-like protein 3; Nucleophosmin; Ankyrin -1; Myotubularin
Metabolic processes	THO complex subunit 2; Hemoglobin subunit beta; Eukaryotic peptide chain release factor GTP-binding subunit ERF3B; Nucleophosmin; Transthyretin; Complement C4-B; Myotubularin; Cathepsin L1
Multicellular organismal process	Tyrosine-protein phosphatase non-receptor type 14; Prothrombin; Olfactory receptor 8A1
Response to stimulus	Prothrombin; Olfactory receptor 8A1; Complement C4-B; GRB10-interacting GYF protein 1; Cathepsin L1
Signaling	Olfactory receptor 8A1; GRB10-interacting GYF protein 1

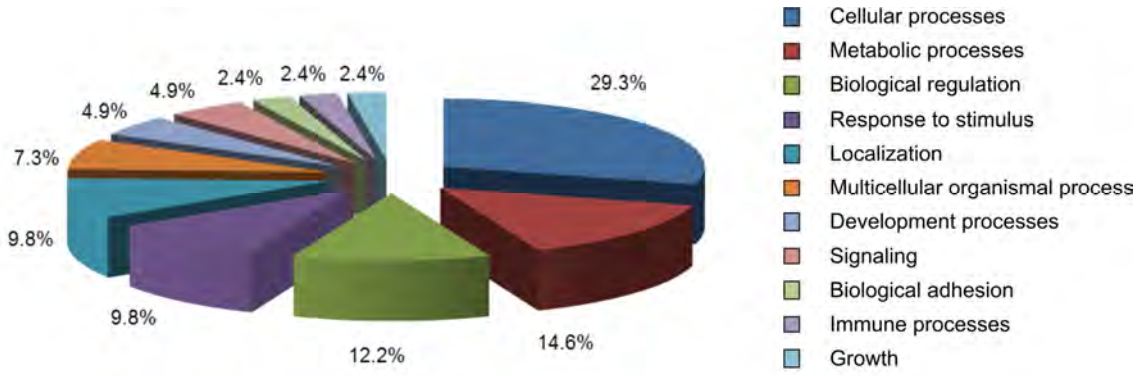


Figure 4. Biological processes provided by the identified proteins in children blood plasma

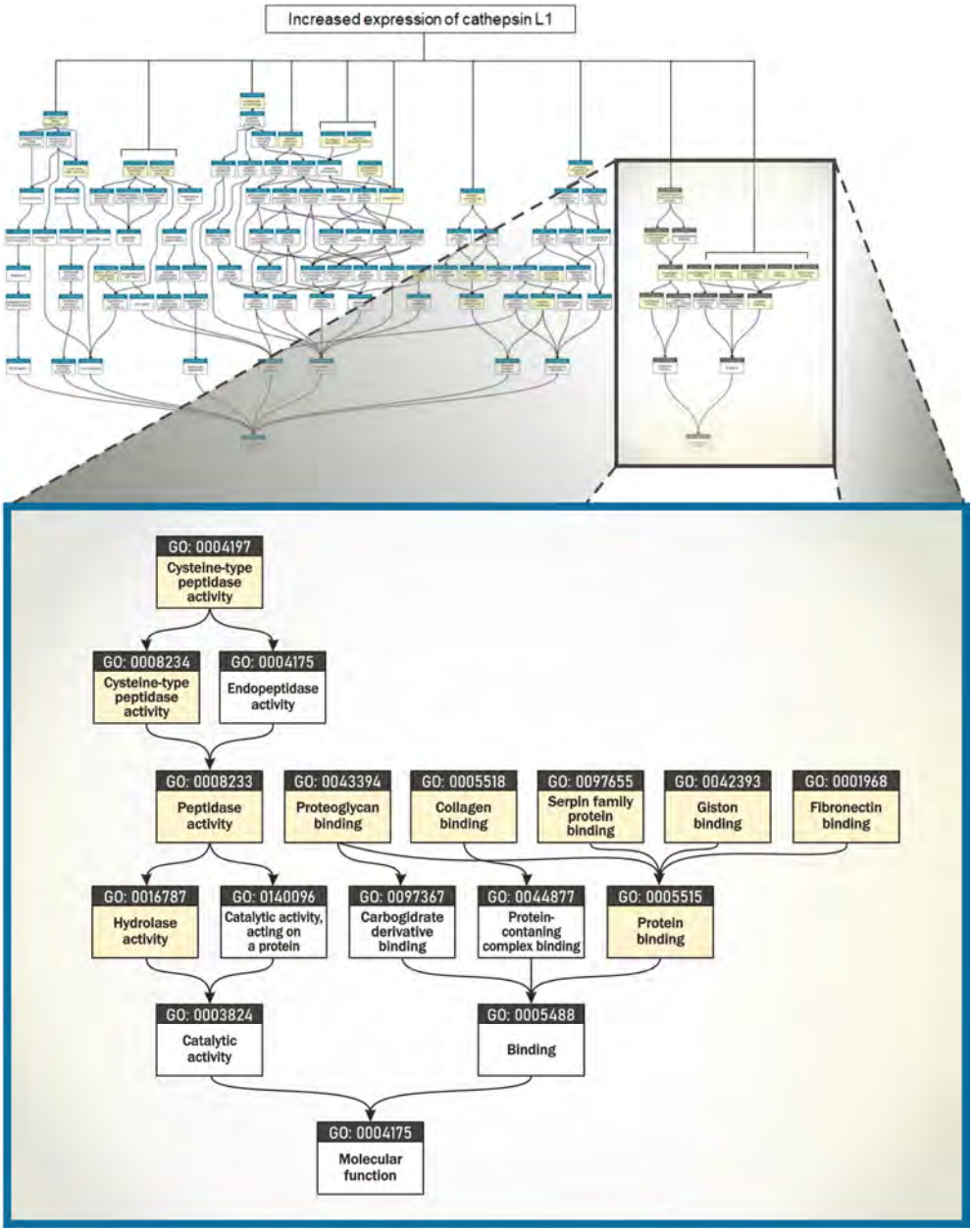


Figure 5. A scheme showing functional belonging of Cathepsin L1
(Gene Ontology and GO Annotations database, 2022)

Cathepsin L1 is expressed in tissues of the respiratory tract, endocrine system, gastrointestinal tract, pancreas, kidneys and muscular tissues. According to the phylogenetic tree, 477 orthologous proteins belong to this protein subfamily. They are from different taxons including Rattus norvegicus (rats) and Mus musculus (mice) (Figure 6).

Bioinformatics analysis of localization, functional and tissue belonging has revealed that proteins from cathepsin family are lysosomal cysteine proteases. They participate in processes involved in cell death, protein degradation and post-translation modification, extracellular matrix remodeling, autophagy and immune signals transfer. Some processes are specifically associated with elevated cathepsin L1 levels. They include degradation and impaired extracellular matrix remodeling in lung and large intestine macrophages; trypsinogen-trypsin imbalance with developing inflammation in pancreatic parenchyma; disrupted transfer of signals through mitogen-activated protein kinase (MAPK) pathways in cardiomyocytes;

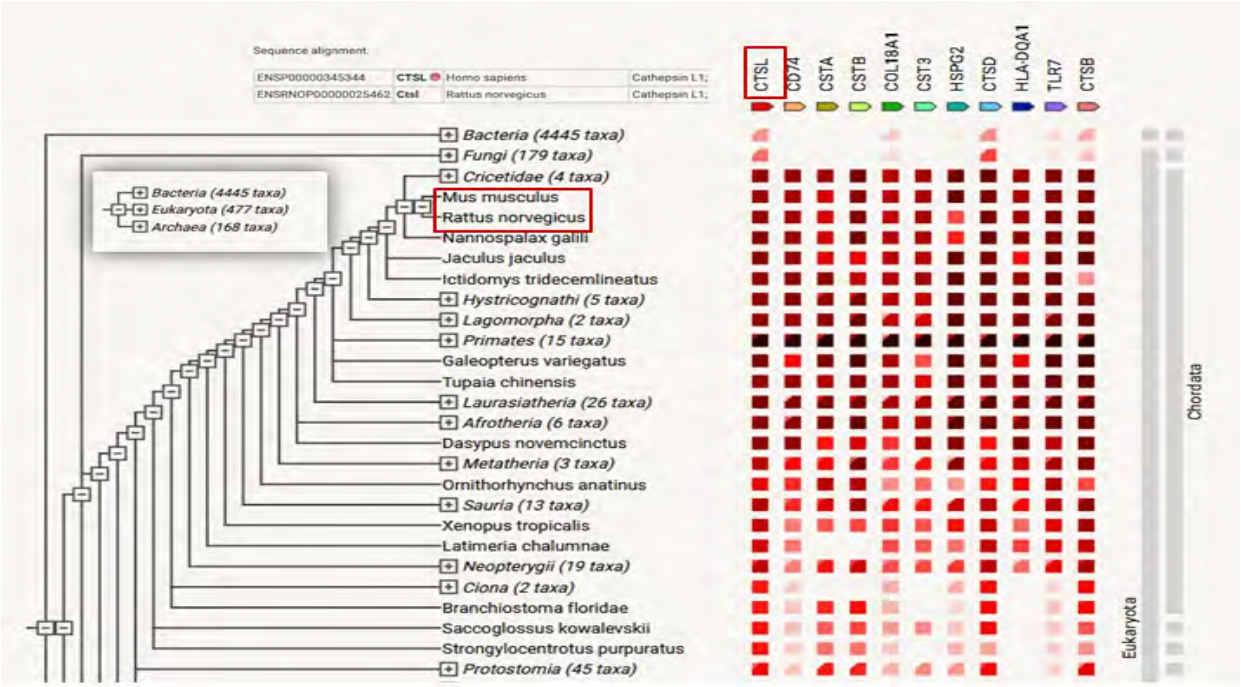


Figure 6. A fragment of the phylogenetic tree showing orthologous genes of Cathepsin L1 (STRING Consortium database[10], 2022)

damage to glomerular podocytes [12–17]. In addition, cathepsin L1 mediates proteolysis of the subunits of the SARS-CoV-2 S1 protein necessary for the virus penetration into a cell and its subsequent replication [18–20]. If exposure to fluoride-containing compounds persists, this results in greater likelihood of risk-associated negative health outcomes including progressing dilated cardiomyopathy, atherosclerosis, pancreatitis and diabetes mellitus, glomerulonephritis, colitis, and coronavirus infection.

Conclusion. We implemented the suggested algorithm and comparatively analyzed proteomic blood plasma profiles of non-exposed children and children who were exposed to fluoride-containing compounds. As a result, 27 expressed proteins were identified. We took cathepsin L1 as an example since its expression correlated authentically with elevated fluoride-ion concentrations in urine (marker of exposure) and created a bioinformatics matrix. The matrix gave a possibility to predict likely risk-associated negative health outcomes in exposed children. These outcomes included cardiomyopathy, colitis, glomerulonephritis, diabetes mellitus, atherosclerosis, and coronavirus infection. The resulting predictive estimates raise effectiveness of early detection and allow developing preventive activities that are more effective for minimizing negative health outcomes.

Funding. The research was not granted any financial support.

Competing interests. The authors declare no competing interests.

[10] STRING Consortium. Available at: https://string-db.org/ (June 02, 2022).

References

1. Action plan for the prevention and control of noncommunicable diseases in the WHO European Region. *WHO*, 2016. Available at: https://apps.who.int/iris/bitstream/handle/10665/341522/WHO-EURO-2016-2582-42338-58618-eng.pdf?sequence=1&isAllowed=y (02.03.2022).

2. Human biomonitoring: facts and figures. Copenhagen, WHO Regional Office for Europe, 2015. Available at: https://apps.who.int/iris/bitstream/handle/10665/164588/WHO-EURO-2015-3209-42967-60040-eng.pdf?sequence=3&isAllowed=y (02.03.2022).

3. Miroshnichenko I.I., Ptitsina S.N. Biomarkers in the modern medical and biologic practice. *Biomeditsinskaya khimiya*, 2009, vol. 55, no. 4, pp. 425–440 (in Russian).

4. Anderson N.L., Anderson N.G. The human plasma proteome: history, character, and diagnostic prospects. *Mol. Cell. Proteomics*, 2002, vol. 1, no. 11, pp. 845–867. DOI: 10.1074/mcp.r200007-mcp200

5. Baer B., Millar A.H. Proteomics in evolutionary ecology. *J. Proteomics*, 2016, vol. 135, pp. 4–11. DOI: 10.1016/j.jprot.2015.09.031

6. Frenkel-Morgenstern M., Cohen A.A., Geva-Zatorsky N., Eden E., Prilusky J., Issaeva I., Sigal A., Cohen-Saidon C. [et al.]. Dynamic Proteomics: a database for dynamics and localizations of endogenous fluorescently-tagged proteins in living human cells. *Nucleic Acids Res.*, 2010, vol. 38, suppl. 1, pp. D508–D512. DOI: 10.1093/nar/gkp808

7. Mi H., Muruganujan A., Thomas P.D. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res.*, 2003, vol. 41, pp. D377–D386. DOI: 10.1093/nar/gks1118

8. Baumgartner C., Osl M., Netzer M. Baumgartner D. Bioinformatic-driven search for metabolic biomarkers in disease. *J. Clin. Bioinforma*, 2011, vol. 1, no. 1, pp. 2. DOI: 10.1186/2043-9113-1-2

9. Palasca O., Santos A., Stolte Ch., Gorodkin J., Jensen L.J. TISSUES 2.0: an integrative web resource on mammalian tissue expression. *Database*, 2018, vol. 2018, pp. 1–12. DOI: 10.1093/database/bay003

10. Mi H., Muruganujan A., Casagrande J.T., Thomas P.D. Large-scale gene function analysis with the PANTHER classification system. *Nat. Protoc.*, 2013, vol. 8, no. 8, pp. 1551–1566. DOI: 10.1038/nprot.2013.092

11. Sun Z., Li S., Yu Y., Chen H., Ommati M.M., Manthari R.K., Niu R., Wang J. Alterations in epididymal proteomics and antioxidant activity of mice exposed to fluoride. *Arch. Toxicol.*, 2018, vol. 92, no. 1, pp. 169–180. DOI: 10.1007/s00204-017-2054-2

12. Jormsjö S., Wuttge D.M., Sirsjö A., Whatling C., Hamsten A., Stemme S., Eriksson P. Differential expression of cysteine and aspartic proteases during progression of atherosclerosis in apolipoprotein E-deficient mice. *Am. J. Pathol.*, 2002, vol. 161, no. 3, pp. 939–945. DOI: 10.1016/S0002-9440(10)64254-X

13. Tang Q., Cai J., Shen D., Bian Z., Yan L., Wang Y.-X., Lan J., Zhuang G.-Q. [et al.]. Lysosomal cysteine peptidase cathepsin L protects against cardiac hypertrophy through blocking AKT/GSK3β signaling. *J. Mol. Med. (Berl.),* 2008, vol. 87, no. 3, pp. 249–260. DOI: 10.1007/s00109-008-0423-2

14. Hsing L.C., Kirk E.A., McMillen T.S., Hsiao S.-H., Caldwell M., Houston B., Rudensky A.Y., LeBoeuf R.C. Roles for cathepsins S, L, and B in insulitis and diabetes in the NOD mouse. *J. Autoimmun.*, 2010, vol. 34, no. 2, pp. 96–104. DOI: 10.1016/j.jaut.2009.07.003

15. Reiser J., Oh J., Shirato I., Asanuma K., Hug A., Mundel T.M., Honey K., Ishidoh K. [et al.]. Podocyte migration during nephrotic syndrome requires a coordinated interplay between cathepsin L and alpha3 integrin. *J. Biol. Chem.*, 2004, vol. 279, no. 33, pp. 34827–34832. DOI: 10.1074/jbc.m401973200

16. Liao M.C., Miyata K.N., Chang S.Y., Zhao X.P., Lo C.S., El-Mortada M.A., Peng J., Chenier I. [et al.]. Angiotensin II type-2-receptor stimulation ameliorates focal and segmental glomerulosclerosis in mice. *Clin. Sci*. (*Lond.*), 2022, vol. 136, no. 10, pp. 715–731. DOI: 10.1042/CS20220188

17. Cattaruzza F., Lyo V., Jones E., Pham D., Hawkins J., Kirkwood K., Valdez–Morales E., Ibeakanma C.H. [et al.]. Cathepsin S is activated during colitis and causes visceral hyperalgesia by a par2-dependent mechanism in mice. *Gastroenterology*, 2011, vol. 141, no. 5, pp. 1864–1874.e1-3. DOI: 10.1053/j.gastro.2011.07.035

18. Zhou N., Pan T., Zhang J., Li Q., Zhang X., Bai C., Huang F., Peng T. [et al.]. Glycopeptide Antibiotics Potently Inhibit Cathepsin L in the Late Endosome/Lysosome and Block the Entry of

Ebola Virus, Middle East Respiratory Syndrome Coronavirus (MERS-CoV), and Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV). *J. Biol. Chem.*, 2016, vol. 291, no. 17, pp. 9218–9232. DOI: 10.1074/jbc.M116.716100

19. Xiong Y., Liu Y., Cao L., Wang D., Guo M., Jiang A., Guo D., Hu W. [et al.]. Transcriptomic characteristics of bronchoalveolar lavage fluid and peripheral blood mononuclear cells in COVID-19 patients. *Emerg. Microbes Infect.*, 2020, vol. 9, no. 1, pp. 761–770. DOI: 10.1080/22221751.2020.1747363

20. Adedeji A.O., Severson W., Jonsson C., Singh K., Weiss S.R., Sarafianos S.G. Novel inhibitors of severe acute respiratory syndrome coronavirus entry that act by three distinct mechanisms. *J. Virol.*, 2013, vol. 87, no. 14, pp. 8017–8028. DOI: 10.1128/JVI.00998-13

Zemlyanova M.A., Zaitseva N.V., Koldibekova Yu.V., Peskova E.V., Bulatova N.I. Creating bioinformatics matrix of molecular markers to predict risk-associated health disorders. Health Risk Analysis, 2022, no. 2, pp. 174–184. DOI: 10.21668/health.risk/2022.2.16.eng

Received: 07.04.2022
Approved: 10.06.2022
Accepted for publication: 21.06.2022