

МЕДИКО-БИОЛОГИЧЕСКИЕ АСПЕКТЫ ОЦЕНКИ ВОЗДЕЙСТВИЯ ФАКТОРОВ РИСКА

УДК 57.044; 616.092
DOI: 10.21668/health.risk/2022.2.16



Научная статья

СОЗДАНИЕ БИОИНФОРМАЦИОННОЙ МАТРИЦЫ МОЛЕКУЛЯРНЫХ МАРКЕРОВ ДЛЯ ПРОГНОЗИРОВАНИЯ РИСК-АССОЦИИРОВАННЫХ НАРУШЕНИЙ ЗДОРОВЬЯ

М.А. Землянова^{1,2,3}, Н.В. Зайцева¹, Ю.В. Кольдибекова¹, Е.В. Пескова^{1,2}, Н.И. Булатова¹

¹Федеральный научный центр медико-профилактических технологий управления рисками здоровью населения, Россия, 6140045, г. Пермь, ул. Монастырская, 82

²Пермский государственный национальный исследовательский университет, Россия, 614990, г. Пермь, ул. Букирева, 15

³Пермский национальный исследовательский политехнический университет, Россия, 614990, г. Пермь, Комсомольский проспект, 29

Одним из приоритетных факторов, формирующих медико-демографические потери, является долговременное непрерывное загрязнение атмосферного воздуха селитебных территорий химическими компонентами. Актуальным является повышение точности прогнозных оценок риска изменений гомеостаза на ранних обратимых стадиях (молекулярный уровень), которые с достаточно высокой степенью вероятности могут трансформироваться в патологические процессы в зрелом возрасте при сохраняющихся условиях экспозиции.

Исследование ставило целью создание биоинформационной матрицы молекулярных маркеров для прогнозирования риск-ассоциированных нарушений здоровья (на примере маркера экспозиции). Предложен поэтапный алгоритм исследований, включающий применение протеомной технологии для выделения экспрессированных белков, выявление: причинно-следственных связей белков с факторами; молекулярно-клеточных и функциональных взаимоотношений в системе «фактор экспозиции – ген – белок – негативный исход» для прогнозирования риск-ассоциированных нарушений здоровья. Реализация алгоритма выполнена на примере исследования белкового профиля плазмы детей в возрасте 3–6 лет, подвергавшихся длительной аэрогенной экспозиции фторсодержащих соединений.

Показано изменение белковой композиции плазмы крови экспонированных детей относительно неэкспонированных по 27 идентифицированным белкам. На примере белка катепсина L1, с изменением уровня которого установлена причинно-следственная связь концентрации фторид-иона в моче, построена биоинформационная матрица. Качественный синтез молекулярно-клеточной локализации, функциональной и тканевой принадлежности показал, что экспрессия катепсина L1, обусловленная повышенным содержанием фторид-иона в моче, может инициировать нарушение ремоделирования внеклеточного матрикса, деградацию и посттрансляционную модификацию белков в клетках легких, толстого кишечника, поджелудочной железы, в кардиомиоцитах, подоцитах почечных клубочков, опосредует процессинг расщепления субъединицы спайкового белка S1 SARS-CoV-2, необходимого для проникновения

© Землянова М.А., Зайцева Н.В., Кольдибекова Ю.В., Пескова Е.В., Булатова Н.И., 2022

Землянова Марина Александровна – доктор медицинских наук, главный научный сотрудник, заведующий отделом биохимических и цитогенетических методов диагностики; доцент кафедры микробиологии и иммунологии (e-mail: zem@fcrisk.ru; тел.: 8 (342) 236-39-30; ORCID: <http://orcid.org/0000-0002-8013-9613>).

Зайцева Нина Владимировна – академик РАН, доктор медицинских наук, профессор, научный руководитель (e-mail: znv@fcrisk.ru; тел.: 8 (342) 237-25-34; ORCID: <http://orcid.org/0000-0003-2356-1145>).

Кольдибекова Юлия Вячеславовна – кандидат биологических наук, старший научный сотрудник, заведующий лабораторией метаболизма и фармакокинетики отдела биохимических и цитогенетических методов диагностики (e-mail: koldibekova@fcrisk.ru; тел.: 8 (342) 237-18-15; ORCID: <http://orcid.org/0000-0002-3924-4526>).

Пескова Екатерина Владимировна – младший научный сотрудник лаборатории биохимической и наносенсорной диагностики отдела биохимических и цитогенетических методов диагностики (e-mail: peskova@fcrisk.ru; тел.: 8 (342) 237-18-15; ORCID: <https://orcid.org/0000-0002-8050-3059>).

Булатова Наталья Ивановна – научный сотрудник лаборатории биохимической и наносенсорной диагностики (e-mail: 1179815@mail.ru; тел.: 8 (342) 236-80-18; ORCID: <https://orcid.org/0000-0003-3392-9097>).

в клетку и репликации вируса. Созданная биоинформационная матрица на примере белка катепсина L1, позволила спрогнозировать развитие риск-ассоциированных негативных эффектов у экспонированных лиц в виде кардиомиопатии, колита, гломерулонефрита, сахарного диабета, атеросклероза, коронавирусной инфекции. Полученные прогнозные оценки позволяют повышать эффективность раннего выявления и разработки профилактических превентивных мер, направленных на минимизацию негативных последствий.

Ключевые слова: протеомный профиль, детское население, клеточно-молекулярная и тканевая принадлежность, молекулярные маркеры, экспрессия гена, негативные эффекты, фторид-ион в моче, катепсин L1, биоинформационные ресурсы.

Сохранение и укрепление здоровья, повышение продолжительности жизни населения, снижение смертности от болезней, в том числе сердечно-сосудистой системы, злокачественных и хронических респираторных заболеваний и др., обозначены в качестве национальных государственных приоритетов в области социально-экономического и демографического развития Российской Федерации на период до 2024 г.

Одним из ведущих факторов риска формирования дополнительных ущербообразующих последствий здоровью, по мнению экспертов ВОЗ, является длительное и постоянное воздействие загрязнений атмосферного воздуха, питьевых вод и почв селитебных территорий химическими компонентами¹ [1]. В целом бремя болезней, обусловленных комплексной химической нагрузкой, оценивается на уровне 15–35 %. Для снижения потерь здоровья у подверженного химическому воздействию населения (в первую очередь, наиболее чувствительные контингенты – дети, подростки, молодые люди) актуальным является повышение точности прогнозных оценок риска формирования изменений гомеостаза на ранних обратимых стадиях (на молекулярном уровне), которые с достаточно высокой степенью вероятности могут трансформироваться в патологические процессы в зрелом возрасте при сохраняющихся условиях экзогенного поступления химических веществ.

Одним из инструментов поиска потенциальных молекулярных маркеров нарушения динамического равновесия может являться применение высокоинформативной протеомической технологии, позволяющей на основании исследований белковой композиции идентифицировать изменения, происходящие в результате экспрессии генов, кодирующих целевые белки (взаимоотношения по типу «ДНК – РНК – белок»), в условиях негативного воздействия факторов, в том числе химического генеза [2–5]. Использование данного подхода является предпочтительным, поскольку основную функциональную нагрузку при взаимодействии любого организма с внешней средой несут белки (либо непосредственно, либо посредством своих ферментативных активностей). Выявление молекулярных белковых и пептидных мишеней, определение их структуры и функций, тканевой принадлежности, вовлеченности в патогенез функциональных нарушений

с применением существующих биоинформационных и токсикогенетических ресурсов [6–10] обеспечивает эффективное предположение ожидаемых негативных исходов со стороны здоровья. Установление при этом ассоциативных патогенетических связей между воздействием факторов экспозиции и экспрессией белковых молекулярных маркеров позволяет прогнозировать риск-ассоциированные негативные эффекты, что приобретает особую значимость для решения задач повышения эффективности раннего (клеточно-молекулярного) выявления нарушений состояния здоровья экспонированного населения.

На основании вышеизложенного представляется целесообразным уточнение подходов к проведению исследований по выявлению экспрессированных белков, ассоциированных с экспозицией химических факторов, биологическая и биохимическая интерпретация клеточно-функциональной принадлежности для детализации молекулярно-клеточных механизмов и повышения точности оценок возможного развития негативных эффектов, участвующих в патогенезе риск-ассоциированных неинфекционных заболеваний, что явилось основанием для определения цели настоящего исследования.

Цель исследования – создание биоинформационной матрицы молекулярных маркеров для прогнозирования риск-ассоциированных нарушений здоровья (на примере маркера экспозиции).

Материалы и методы. Создание биоинформационной матрицы молекулярных маркеров для прогнозирования риск-ассоциированных нарушений здоровья возможно посредством поэтапного алгоритма исследований:

- формирование выборок из числа экспонированных и неэкспонированных лиц, относящихся к наиболее чувствительным к негативному воздействию химических факторов (группа наблюдения и группа контроля);
- установление факта экспозиции на основании определения повышенного содержания в биосредах химических веществ, доказанно связанных с экспозицией (идентификация маркеров экспозиции);
- построение протеомного профиля, идентификация, сравнительный анализ и выделение статистически различающихся по интенсивности белковых пятен в группе наблюдения относительно контроля;

¹ О состоянии санитарно-эпидемиологического благополучия населения в Российской Федерации в 2020 году: Государственный доклад. – М.: Федеральная служба по надзору в сфере защиты прав потребителей и благополучия человека, 2021. – 256 с.

– выявление причинно-следственных связей идентифицированных экспрессированных белков с факторами экспозиции (по содержанию маркеров экспозиции в биосредах);

– качественный синтез и построение филогенетического древа, отражающего молекулярно-клеточные и функциональные взаимоотношения в системе «фактор экспозиции – ген – белок – негативный исход» для прогнозирования риск-ассоциированных нарушений здоровья.

Реализация предложенного алгоритма выполнена на примере исследования протеомного профиля плазмы детей в возрасте 3–6 лет, подвергающихся длительной аэрогенной экспозиции фтористых газобразных и твердых соединений, формирующей неканцерогенный риск здоровью, в 1,5–2 раза превышающий приемлемый уровень. Критерием формирования выборок детей группы наблюдения явилось повышенное в 1,5 раза и более относительно референтного значения² содержание фторид-иона в моче (маркера экспозиции фтористых соединений в соответствии с установленными причинно-следственными связями). У детей группы контроля уровень изучаемого химического вещества соответствовал референтной величине или являлся близким ему. Обследование детей выполнено с соблюдением этических принципов Хельсинкской декларации (WMA Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects, 2013) и одобрено Комитетом по биомедицинской этике ФБУН «ФНЦ медико-профилактических технологий управления рисками здоровью населения» с обязательным получением информированного добровольного согласия законного представителя.

Исследование протеомного профиля выполнено по технологии двухмерного электрофореза в полиакриламидном геле. Полученные электрофореограммы плазмы крови визуализировали методом окраски серебром и документировали с помощью системы для гель-документирования (BioRad, США). Сравнительный анализ полученных протеомных карт проводили с помощью программного комплекса PDQuest (BioRad, США), выделяли значимые белковые пятна по их интенсивности и осуществляли последующий анализ методом жидкостной хроматографии в сочетании с масс-спектрометрическим анализом (на хроматографе UltiMate 3000 (Германия) и тандемном масс-спектрометре AB Sciex 4000 QTRAP с источником ионизации Nanospray 3 (Кана-

да)). Данные тандемных исследований обрабатывали программой ProteinPilot, версия 4.5 (AB Sciex), с идентификацией по базе данных UniProt_sprot_fasta (от 24.11.2017), с выборкой по таксону Homo Sapiens (пептидный фингерпринт). Основную часть информации о полученных белках экстрагировали из баз данных Gene Ontology Resource и UniProt³. Установление генов, кодирующих экспрессию идентифицированных белков, выполнено с помощью базы данных HGNC database of human gene name⁴.

Статистическую обработку полученных данных осуществляли с помощью пакета программ Statistica 10. Сравнительную оценку полученных значений относительного объема белковых пятен у детей группы наблюдения выполняли относительно аналогичных показателей у детей группы контроля. Результаты проведенных исследований представлены в виде среднего значения (\bar{X}), ошибки среднего (SEM). Статистическую значимость различий переменных между группами определяли по критерию Манна – Уитни ($U \leq U_{кр}$) при заданном уровне значимости $p \leq 0,05$.

Выявление и оценку причинно-следственных связей между изменением статистически отличающихся белковых пятен и концентрацией в моче фторид-иона проводили с помощью построения математической модели линейной регрессии. Достоверность и адекватность полученных моделей оценивали на основе дисперсионного анализа с использованием F -критерия Фишера, коэффициента детерминации (R^2), t -критерию Стьюдента при статистической значимости $p \leq 0,05$.

Классификация белков по классам, биологическим и молекулярным функциям выполнена с помощью баз данных Panther Classification System и Gene Ontology and GO Annotations⁵. Для установления экспрессии белков в тканях организма применяли информацию, содержащуюся на платформах Tissue expression database и The Human Protein Atlas⁶. Описание отношений в системе «фактор экспозиции (маркер экспозиции) – ген – белок – заболевание» производили с использованием информационного ресурса Comparative Toxicogenomics⁷.

Результаты и их обсуждение. Сравнительный анализ результатов денситометрического измерения протеомных карт плазмы крови обследуемых детей позволил выявить наличие достоверных различий относительного объема 27 белковых пятен у детей группы наблюдения и группы сравнения (табл. 1).

² Тиц Н.У. Клиническое руководство по лабораторным тестам. – М.: ЮНИМЕД-пресс, 2003. – 960 с.

³ Gene Ontology Resource [Электронный ресурс]. – URL: <http://geneontology.org/> (дата обращения: 08.03.2022); UniProt [Электронный ресурс]. – URL: <http://www.uniprot.org> (дата обращения: 08.03.2022).

⁴ HGNC database of human gene name [Электронный ресурс]. – URL: <https://www.genenames.org> (дата обращения: 12.04.2022).

⁵ Panther Classification System [Электронный ресурс]. – URL: <http://www.pantherdb.org> (дата обращения: 23.03.2022); Gene Ontology and GO Annotations [Электронный ресурс]. – URL: <https://www.ebi.ac.uk/QuickGO> (дата обращения: 23.03.2022).

⁶ Tissue expression database [Электронный ресурс]. – URL: <https://tissues.jensenlab.org/Search> (дата обращения: 18.03.2022); The Human Protein Atlas [Электронный ресурс]. – URL: <https://www.proteinatlas.org/> (дата обращения: 18.03.2022).

⁷ Comparative Toxicogenomics [Электронный ресурс]. – URL: <http://ctdbase.org/> (дата обращения: 12.03.2022).

Белки, достоверно отличающиеся от таковых группы сравнения, выделенные в протеомном профиле плазмы крови детей

№ п/п	Наименование белка	Идентификатор UniProt	Кодирующий ген	Средний объем белкового пятна, $\bar{X} \pm SEM$	
				Группа наблюдения	Группа сравнения
1	Кератин 9	P35527	KRT9	1406,3 ± 148,7*	2232,5 ± 272,2
2	Vand4.1-подобный белок 3	Q9Y2J2	EPB41L3	1890,6 ± 56,2*	2833,8 ± 135,7
3	Инозитол 1,4,5-трисфосфатный рецептор взаимодействующий белокоподобный 1	Q6GPH6	ITPRIPL1	1736,9 ± 56,2*	2743,4 ± 304,4
4	Бета субъединица гемоглобина	P68871	HBV	2350,7 ± 161,6*	1447,9 ± 253,0
5	Субъединица ламинина альфа-3	Q16787	LAMA3	2435,4 ± 148,4*	66,5 ± 61,5
6	Аполипопротеин А-I	P02647	APOA1	66,5 ± 12,5*	2435,4 ± 148,4
7	Фактор высвобождения пептидной цепи эукариот GTP-связывающая субъединица ERF3B	Q8IYD1	GSPT2	2547,5 ± 118,0*	94,0 ± 81,0
8	J-цепь иммуноглобулина	P01591	JCHAIN	1785,0 ± 86,0*	101,0 ± 34,0
9	ТНО комплексная субъединица 2	Q8NI27	THOC2	3157,56 ± 65,4*	2983,4 ± 61,1
10	Nck-ассоциированный белок 5	Q9HCH0	NCKAP5L	2586,6 ± 105,1*	2840,3 ± 92,5
11	GRB10-взаимодействующий GYF-белок 1	Q75420	GIGYF1	2340,7 ± 183,3*	95,4 ± 61,4
12	Кельч-подобный белок 4	Q9C0H6	KLHL4	1806,6 ± 78,0*	100,5 ± 76,9
13	Протромбин	P00734	F2	1893,7 ± 117,3*	1724,6 ± 47,4
14	Нуклеофосмин	P06748	NPM1	469,0 ± 120,0*	94,0 ± 26,0
15	Дополнение C4-B	P0C0L5	C4B	2749,8 ± 56,0**	1821,2 ± 259,7
16	Катепсин L1	P07711	CTSL	1712,0 ± 123,7*	518,5 ± 125,5
17	Анкирин-1	P16157	ANK1	1765,6 ± 91,0*	454,2 ± 505,1
18	Транспортный холиноподобный белок 3	Q8N4M1	SLC44A3	1715,2 ± 123,0*	72,7 ± 71,0
19	Обонятельный рецептор 8A1	Q8NGG7	OR8A1	1712,1 ± 123,7*	518,5 ± 248,0
20	Трансиретин	P02766	TTR	3966,9 ± 189,3*	2344,4 ± 300,2
21	Витронектин	P04004	VTN	3568,8 ± 233,2*	1844,1 ± 142,6
22	Тирозин-протеинфосфатаза нерцепторного типа 14	Q15678	PTPN14	1639,1 ± 156,8*	99,5 ± 95,63
23	Фактор терминации транскрипции 1	Q15361	TTF1	3109,5 ± 75,2*	3342,2 ± 118,7
24	Аполипопротеин С-III	P02656	APOC3	4249,6 ± 250,1*	3027,4 ± 250,1
25	Миотубулярин	Q13496	MTM1	4408,1 ± 263,9*	3228,0 ± 335,8
26	Аполипопротеин С-II	P02655	APOC2	3802,7 ± 272,0*	2326,7 ± 295,1
27	Сывороточный амилоидный белок А-1	P0DJ18	SAA1	2326,3 ± 154,1*	61,8 ± 5,9

Примечание: * – достоверность межгруппового различия по средним для $p = 0,0001$; ** – достоверность межгруппового различия по средним для $p = 0,009$.

У детей группы наблюдения относительно соответствующих данных группы сравнения установлено достоверное увеличение в 1,4–37,6 раза объема порядка 75 % выявленных белковых пятен от общего количества, а также уменьшение в 1,5–36,6 раза объема 25,0 % белков ($p = 0,0001-0,009$).

На рис. 1 представлен фрагмент двумерной электрофореграммы для установления количественных различий белка катепсина L1 в плазме крови обследованных детей.

Биоинформационный анализ локализации идентифицированных белков, выполненный с помощью системы классификации Panther, позволил установить, что большинство из них (57,7 %) входят в состав клеточных структур (GO: 0110165), и 15,4 % белков локализованы в протеиновых комплексах (GO: 00032991). В зависимости от участия в жизненном цикле белки классифицированы следующим образом: 22,7 % белков относятся к протеинам, участвующим в передаче сигналов (PC00219), 13,6 % –

к метаболитам взаимосвязи ферментов (PC00262), по 9,1 % – к ферментативным модуляторам (PC00260), белкам цитоскелета (PC00085) и белкам-адаптерам (PC00226) (табл. 2, рис. 2).

Анализ классификационных признаков по молекулярным процессам показал, что большинство белков осуществляют функции связывания с другими молекулами (GO: 0005488; 44,8 %) и каталитической активности (GO: 0003824; 31,3 %) (табл. 3, рис. 3).

Исследование и оценка классификации биологических функций показали, что основная часть белков обеспечивает клеточные (GO: 0009987; 30 %) и метаболические процессы (GO: 0008152; 15 %), а также биологическое регулирование (GO: 0065007; 12 %) (табл. 4, рис. 4).

Установленные классификационные признаки (локализация, участие в жизненном цикле, биологические функции) идентифицированных белков у детей группы наблюдения аналогичны таковым белков группы контроля.

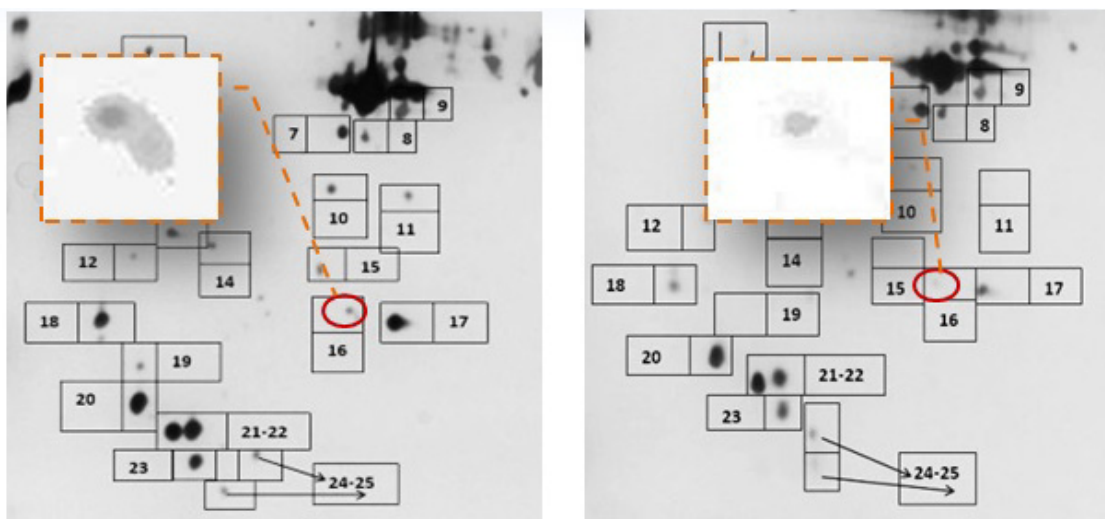


Рис. 1. Фрагмент двумерной электрофореграммы плазмы крови детей:
 а – индивид группы наблюдения; б – индивид группы сравнения

Таблица 2

Классы выделенных белков плазмы крови детей в зависимости от участия в жизненном цикле

Класс белков*	Наименование белка
Ферментативные модуляторы	Тирозин-протеинфосфатаза нерцепторного типа 14, протромбин, катепсин L1
Белок адаптер	Анкирин-1, Кельч-подобный белок 4
Белок передачи	Аполипопротеин А-I, АПОЛИПОПРОТЕИН С-II, аполипопротеин С-III, бета субъединица гемоглобина
Белок цитоскелета	Band4.1-подобный белок 3, Кератин 9
Метаболит взаимосвязи ферментов	Транстиретин, инозитол 1,4,5-трисфосфатный рецептор взаимодействующий белкоподобный 1, миотубулярин
Транспортер	Транспортный холиноподобный белок 3
Шаперон	Нуклеофосмин
Молекула клеточной адгезии	Субъединица ламинина альфа-3
Модулятор активности белкового связывания	Дополнение С4-В
Трансмембранный сигнальный рецептор	Обонятельный рецептор 8A1
Белок защиты	J-цепь иммуноглобулина
Регулятор специфичной геной транскрипции	ТНО-комплексная субъединица 2
Белок трансляции	Фактор высвобождения пептидной цепи эукариот GTP-связывающая субъединица ERF3B

Примечание: * – согласно данным The PANTHER Classification System⁸.

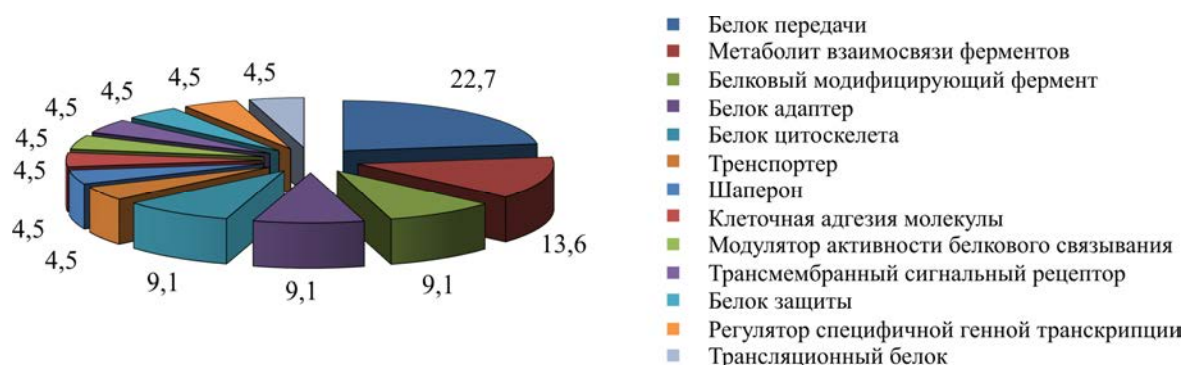


Рис. 2. Классы выделенных белков плазмы крови детей в зависимости от участия в жизненном цикле, %

⁸ The PANTHER Classification System [Электронный ресурс]. – URL: <http://www.pantherdb.org/> (дата обращения: 19.03.2022).

Таблица 3

Молекулярные функции выделенных белков плазмы крови детей

Молекулярная функция*	Название белка
Активность молекулярного преобразователя	Обонятельный рецептор 8A1
Активность молекулярного адаптера	Анкирин-1
Активность транспорта	Транспортный холиноподобный белок 3
Регуляторная активность	Фактор высвобождения пептидной цепи эукариот GTP-связывающая субъединица ERF3B
Связи с другими молекулами	Обонятельный рецептор 8A1; ТНО-комплексная субъединица 2; бета-субъединица гемоглобина; фактор высвобождения пептидной цепи эукариот GTP-связывающая субъединица ERF3B; нуклеофосмин; анкирин-1; витронектин
Каталитическая активность	Фактор высвобождения пептидной цепи эукариот GTP-связывающая субъединица ERF3B; тирозин-протеинфосфатаза нерецепторного типа 14; протромбин; бета-субъединица гемоглобина; миотубулярин; катепсин L1

Примечание * – согласно данным The PANTHER Classification System⁸.

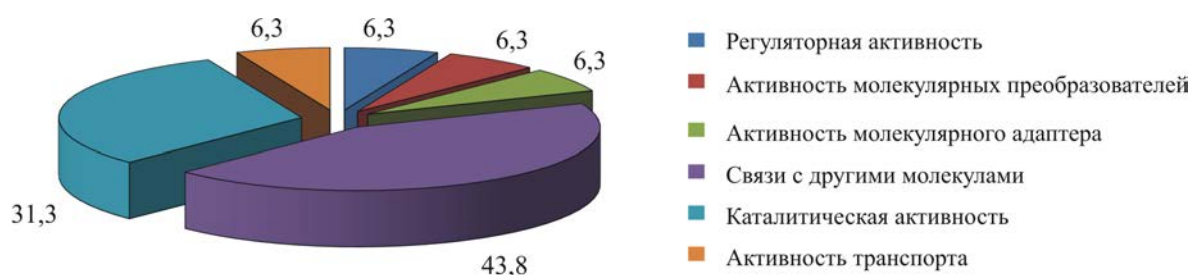


Рис. 3. Молекулярные функции выделенных белков плазмы крови детей, %

Таблица 4

Биологические функции выделенных белков плазмы крови детей

Биологическая функция	Название белка
Биологическая адгезия	Витронектин
Процессы роста	Протромбин
Процессы иммунной системы	Дополнение С4-В; катепсин L
Биологическое регулирование	Протромбин; обонятельный рецептор 8A1; нуклеофосмин; дополнение С4-В; GRB10-взаимодействующий GYF-белок 1; миотубулярин
Клеточные процессы	Nck-ассоциированный белок 5; протромбин; ТНО-комплексная субъединица 2; обонятельный рецептор 8A1; бета-субъединица гемоглобина; транспортный холиноподобный белок 3; фактор высвобождения пептидной цепи эукариот GTP-связывающая субъединица ERF3B; нуклеофосмин; анкирин-1; транстеритин; миотубулярин; витронектин; GRB10-взаимодействующий GYF-белок 1; катепсин L1
Процессы развития	Тирозин-протеинфосфатаза нерецепторного типа 14; протромбин; миотубулярин;
Локализация	ТНО-комплексная субъединица 2; транспортный холиноподобный белок 3; нуклеофосмин; анкирин-1; миотубулярин
Метаболический процесс	ТНО-комплексная субъединица 2; бета-субъединица гемоглобина; фактор высвобождения пептидной цепи эукариот GTP-связывающая субъединица ERF3B; нуклеофосмин; транстеритин; дополнение С4-В; миотубулярин; катепсин L1
Многоклеточный процесс организма	Тирозин-протеинфосфатаза нерецепторного типа 14; протромбин; обонятельный рецептор 8A1
Ответ на стимул	Протромбин; обонятельный рецептор 8A1; дополнение С4-В; GRB10-взаимодействующий GYF-белок 1; катепсин L1
Сигнальная функция	Обонятельный рецептор 8A1; GRB10-взаимодействующий GYF-белок 1

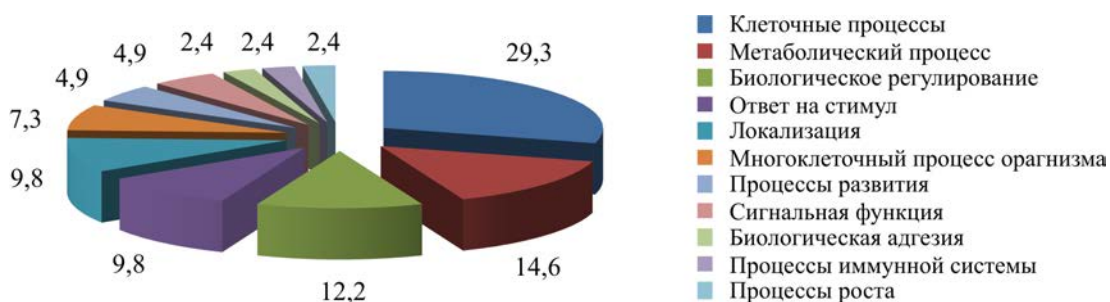


Рис. 4. Биологические процессы выделенных белков плазмы крови детей, %

Анализ причинно-следственных связей между изменением уровня экспрессированных 27 белков и концентрацией фтор-иона в моче позволил выявить достоверную прямую зависимость только для белка катепсина L1 ($R^2 = 0,45$; $b_0 = 764,23$; $b_1 = 51,47$; $p = 0,016$), что согласуется с данными экспериментальных исследований, свидетельствующих, что

экспозиция соединений фтора приводит к увеличению экспрессии катепсина L1 [11].

На примере белка катепсина L1 построена биоинформационная матрица функциональной принадлежности (рис. 5), полученная на биоинформационной платформе Gene Ontology and GO Annotations⁹.

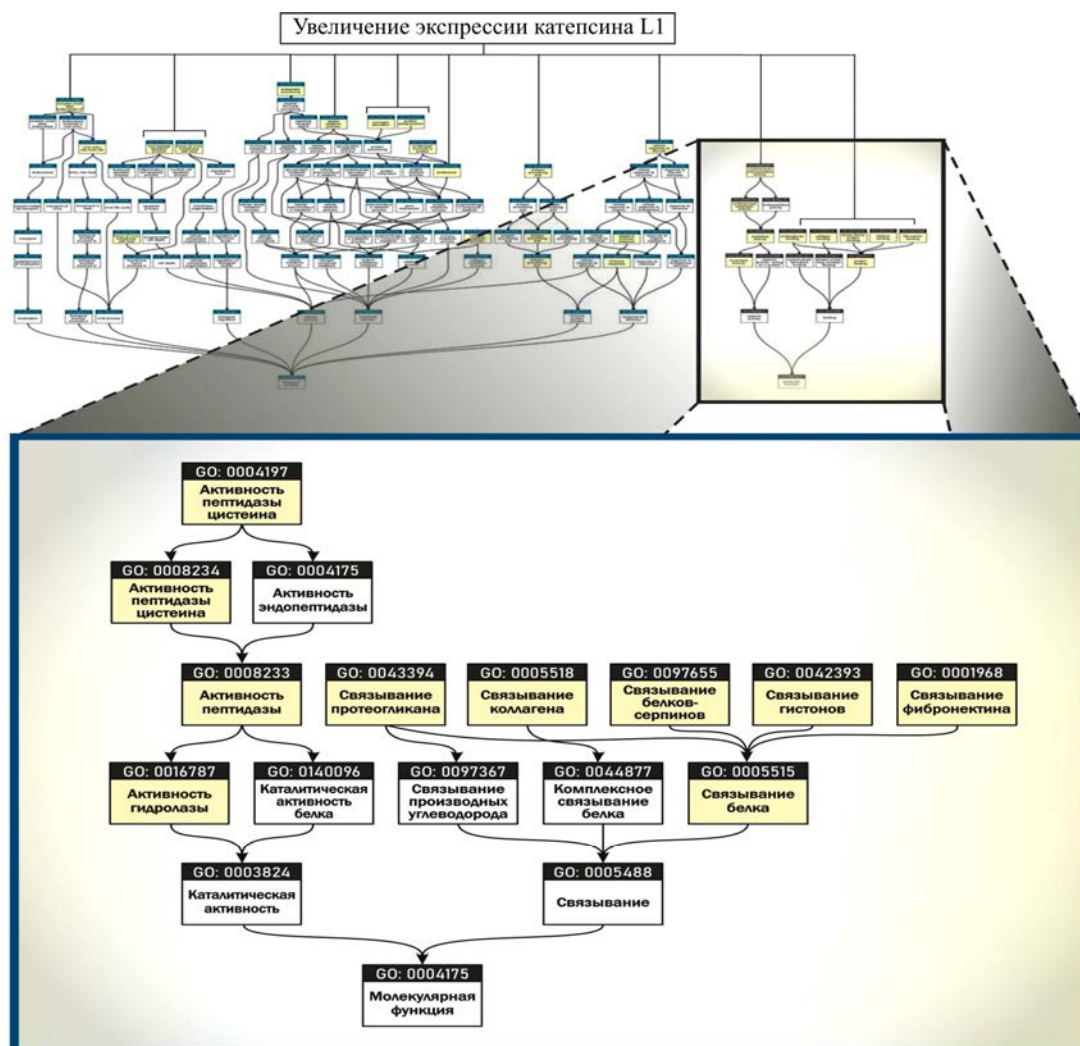


Рис. 5. Схема функциональной принадлежности белка катепсина L1 (база данных Gene Ontology and GO Annotations, 2022)

⁹ Gene Ontology and GO Annotations [Электронный ресурс]. – URL: <https://www.ebi.ac.uk/QuickGO/> (дата обращения: 19.03.2022).

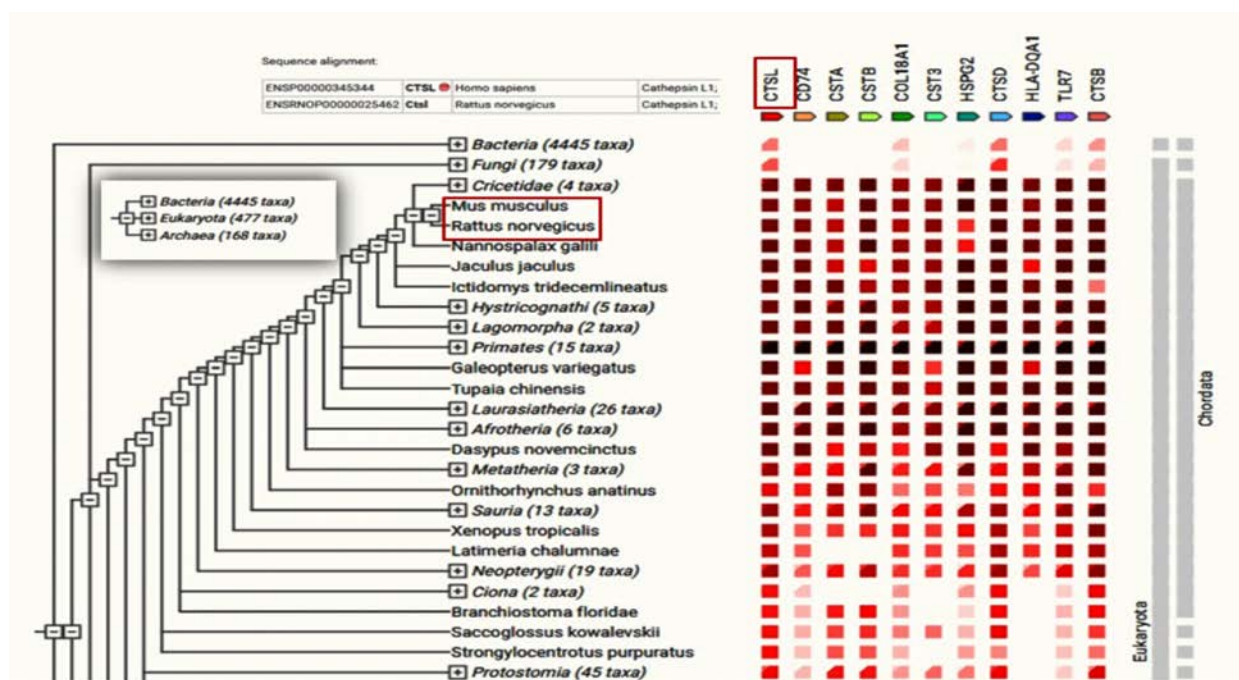


Рис. 6. Фрагмент филогенетического дерева ортологичных генов катепсина L1 (база данных STRING Consortium¹⁰, 2022)

Экспрессия белка катепсина L1 осуществляется в тканях респираторного тракта, эндокринной системы, желудочно-кишечного тракта, поджелудочной железы, почках и мышечной ткани. К этому подсемейству белков, согласно филогенетическому дереву, относится 477 ортологичных генов, принадлежащих к различным таксонам, в том числе *Rattus norvegicus* (крысы) и *Mus musculus* (мыши) (рис. 6).

Биоинформационный анализ локализации, функциональной и тканевой принадлежности показал, что белки семейства катепсинов являются лизосомальными цистеиновыми протеиназами, участвующими в процессах, связанных с гибелью клеток, деградацией и посттрансляционными модификациями белков, ремоделированием внеклеточного матрикса, аутофагией и передачей иммунных сигналов. Конкретно с увеличением уровня катепсина L1 связаны деградация и нарушение ремоделирования внеклеточного матрикса в макрофагах легких и толстого кишечника; нарушение баланса «трипсиноген – трипсин» с развитием воспаления в паренхиме поджелудочной железы; нарушение передачи сигналов пути митоген-активируемых протеинкиназ (МАРК) в кардиомиоцитах, повреждение подоцитов почечных клубочков [12–17]. Кроме этого, катепсин L1 опосредует процессинг расщепления субъединицы спайкового белка S1 SARS-CoV-2, необходимого для проникновения в клетку и последующей репликации вируса [18–20]. При сохраняющихся условиях экспозиции

фторсодержащими соединениями увеличивается вероятность развития таких риск-ассоциированных негативных исходов со стороны здоровья, как прогрессирующая дилатационная кардиомиопатия, атеросклероз, панкреатит и сахарный диабет, гломерулонефрит, колит, коронавирусная инфекция.

Выводы. В результате реализации предложенного алгоритма и сравнительного анализа протеомных карт плазмы крови детей, экспонированных и не экспонированных фторсодержащими соединениями, идентифицировано 27 экспрессированных белков. На примере белка катепсина L1, экспрессия которого достоверно связана с повышенным содержанием фтор-иона в моче (маркера экспозиции), создана биоинформационная матрица, позволившая спрогнозировать развитие риск-ассоциированных негативных эффектов у экспонированных лиц, в том числе в виде кардиомиопатии, колита, гломерулонефрита, сахарного диабета, атеросклероза, коронавирусной инфекции. Полученные прогнозные оценки позволяют повышать эффективность раннего выявления и разработки профилактических превентивных мер, направленных на минимизацию негативных последствий.

Финансирование. Исследование не имело спонсорской поддержки.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

¹⁰ STRING Consortium [Электронный ресурс]. – URL: <https://string-db.org/> (дата обращения: 02.03.2022).

Список литературы

1. План действий по профилактике и борьбе с неинфекционными заболеваниями в Европейском регионе ВОЗ [Электронный ресурс] // ВОЗ. – 2017. – URL: https://www.euro.who.int/_data/assets/pdf_file/0019/346330/NCD-ActionPlan-RU.pdf (дата обращения: 02.03.2022).
2. Биомониторинг человека: факты и цифры [Электронный ресурс]. – Копенгаген: Европейское региональное бюро ВОЗ, 2015. – URL: <https://apps.who.int/iris/bitstream/handle/10665/164589/WHO-EURO-2015-3209-42967-60041-rus.pdf?sequence=3&isAllowed=y> (дата обращения: 02.03.2022).
3. Мирошниченко И.И., Птицина С.Н. Биомаркеры в современной медико-биологической практике // Биомедицинская химия. – 2009. – Т. 55, № 4. – С. 425–440.
4. Anderson N.L., Anderson N.G. The human plasma proteome: history, character, and diagnostic prospects // Mol. Cell. Proteomics. – 2002. – Vol. 1, № 11. – P. 845–867. DOI: 10.1074/mcp.r200007-mcp200
5. Baer B., Millar A.H. Proteomics in evolutionary ecology // J. Proteomics. – 2016. – Vol. 135. – P. 4–11. DOI: 10.1016/j.jprot.2015.09.031
6. Dynamic Proteomics: a database for dynamics and localizations of endogenous fluorescently-tagged proteins in living human cells / M. Frenkel-Morgenstern, A.A. Cohen, N. Geva-Zatorsky, E. Eden, J. Prilusky, I. Issaeva, A. Sigal, C. Cohen-Saidon [et al.] // Nucleic Acids Res. – 2010. – Vol. 38, Suppl. 1. – P. 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. – P. D377–D386. DOI: 10.1093/nar/gks1118
8. Bioinformatic-driven search for metabolic biomarkers in disease / C. Baumgartner, M. Osl, M. Netzer, D. Baumgartner // J. Clin. Bioinforma. – 2011. – Vol. 1, № 1. – P. 2. DOI: 10.1186/2043-9113-1-2
9. TISSUES 2.0: an integrative web resource on mammalian tissue expression / O. Palasca, A. Santos, C. Stolte, J. Gorodkin, L.J. Jensen // Database. – 2018. – Vol. 2018. – P. 1–12. DOI: 10.1093/database/bay003
10. Large-scale gene function analysis with the PANTHER classification system / H. Mi, A. Muruganujan, J.T. Casagrande, P.D. Thomas // Nat. Protoc. – 2013. – Vol. 8, № 8. – P. 1551–1566. DOI: 10.1038/nprot.2013.092
11. Alterations in epididymal proteomics and antioxidant activity of mice exposed to fluoride / Z. Sun, S. Li, Y. Yu, H. Chen, M.M. Ommati, R.K. Manthari, R. Niu, J. Wang // Arch. Toxicol. – 2018. – Vol. 92, № 1. – P. 169–180. DOI: 10.1007/s00204-017-2054-2
12. Differential expression of cysteine and aspartic proteases during progression of atherosclerosis in apolipoprotein E-deficient mice / S. Jormsjö, D.M. Wuttge, A. Sirsjö, C. Whatling, A. Hamsten, S. Stemme, P. Eriksson // Am. J. Pathol. – 2002. – Vol. 161, № 3. – P. 939–945. DOI: 10.1016/S0002-9440(10)64254-X
13. Lysosomal cysteine peptidase cathepsin L protects against cardiac hypertrophy through blocking AKT/GSK3 β signaling / Q. Tang, J. Cai, D. Shen, Z. Bian, L. Yan, Y.-X. Wang, J. Lan, G.-Q. Zhuang [et al.] // J. Mol. Med. (Berl.). – 2008. – Vol. 87, № 3. – P. 249–260. DOI: 10.1007/s00109-008-0423-2
14. Roles for cathepsins S, L, and B in insulinitis and diabetes in the NOD mouse / L.C. Hsing, E.A. Kirk, T.S. McMillen, S.-H. Hsiao, M. Caldwell, B. Houston, A.Y. Rudensky, R.C. LeBoeuf // J. Autoimmun. – 2010. – Vol. 34, № 2. – P. 96–104. DOI: 10.1016/j.jaut.2009.07.003
15. Podocyte migration during nephrotic syndrome requires a coordinated interplay between cathepsin L and alpha3 integrin / J. Reiser, J. Oh, I. Shirato, K. Asanuma, A. Hug, T.M. Mundel, K. Honey, K. Ishidoh [et al.] // J. Biol. Chem. – 2004. – Vol. 279, № 33. – P. 34827–34832. DOI: 10.1074/jbc.m401973200
16. Angiotensin II type-2-receptor stimulation ameliorates focal and segmental glomerulosclerosis in mice / M.C. Liao, K.N. Miyata, S.Y. Chang, X.P. Zhao, C.S. Lo, M.A. El-Mortada, J. Peng, I. Chenier [et al.] // Clin. Sci. (Lond.). – 2022. – Vol. 136, № 10. – P. 715–731. DOI: 10.1042/CS20220188
17. Cathepsin S is activated during colitis and causes visceral hyperalgesia by a PAR2-dependent mechanism in mice / F. Cattaruzza, V. Lyo, E. Jones, D. Pham, J. Hawkins, K. Kirkwood, E. Valdez-Morales, C.H. Ibeakanma [et al.] // Gastroenterology. – 2011. – Vol. 141, № 5. – P. 1864–1874.e1-3. DOI: 10.1053/j.gastro.2011.07.035
18. 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) / N. Zhou, T. Pan, J. Zhang, Q. Li, X. Zhang, C. Bai, F. Huang, T. Peng [et al.] // J. Biol. Chem. – 2016. – Vol. 291, № 17. – P. 9218–9232. DOI: 10.1074/jbc.M116.716100
19. Transcriptomic characteristics of bronchoalveolar lavage fluid and peripheral blood mononuclear cells in COVID-19 patients / Y. Xiong, Y. Liu, L. Cao, D. Wang, M. Guo, A. Jiang, D. Guo, W. Hu [et al.] // Emerg. Microbes Infect. – 2020. – Vol. 9, № 1. – P. 761–770. DOI: 10.1080/22221751.2020.1747363
20. Novel inhibitors of severe acute respiratory syndrome coronavirus entry that act by three distinct mechanisms / A.O. Adedeji, W. Severson, C. Jonsson, K. Singh, S.R. Weiss, S.G. Sarafianos // J. Virol. – 2013. – Vol. 87, № 14. – P. 8017–8028. DOI: 10.1128/JVI.00998-13

Создание биоинформационной матрицы молекулярных маркеров для прогнозирования риск-ассоциированных нарушений здоровья / М.А. Землянова, Н.В. Зайцева, Ю.В. Кольдибекова, Е.В. Пескова, Н.И. Булатова // Анализ риска здоровью. – 2022. – № 2. – С. 174–184. DOI: 10.21668/health.risk/2022.2.16



Research article

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

M.A. Zemlyanova^{1,2,3}, N.V. Zaitseva¹, Yu.V. Koldibekova¹, E.V. Peskova^{1,2}, N.I. Bulatova¹¹Federal Scientific Center for Medical and Preventive Health Risk Management Technologies, 82 Monastyrskaya Str., Perm, 614045, Russian Federation²Perm State University, 15 Bukireva Str., Perm, 614990, Russian Federation³Perm 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.

© Zemlyanova M.A., Zaitseva N.V., Koldibekova Yu.V., Peskova E.V., Bulatova N.I., 2022

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 of the Department of Biochemical and Cytogenetic Methods of Diagnostics (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>).

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/GSK3 β 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 insulinitis 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 α 3 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

Получена: 07.04.2022

Одобрена: 10.06.2022

Принята к публикации: 21.06.2022