



К вопросу о базах данных белковых биомаркеров воздействия вредных факторов окружающей среды на организм человека (обзор литературы)

Гиззатуллина О.И., Чемезов А.И.

ФБУН «Екатеринбургский медицинский – научный центр профилактики и охраны здоровья рабочих промышленных предприятий» Роспотребнадзора, ул. Попова, д. 30, г. Екатеринбург, 620014, Российская Федерация

Резюме

Введение. Воздействие вредных химических веществ на организм, в том числе профессионально обусловленное, представляет опасность для здоровья населения. Оно способно привести к оксидативному стрессу, перекисному окислению липидов и модификациям белков. Созданию эффективных профилактических и терапевтических подходов на основании белковых биомаркеров будет способствовать формирование базы данных, связывающей факторы профессиональной экспозиции и влияние вредных веществ на организм.

Цель исследования: провести системный обзор литературных источников для выявления баз данных, содержащих информацию о белковых биомаркерах в организме людей, подверженных влиянию вредных факторов окружающей среды.

Материалы и методы. Осуществлен научный обзор исследований на русском и английском языках с использованием информационных источников eLIBRARY.ru, PubMed, Google Scholar, Scopus, Research Gate за период 2003–2023 гг. Поиск осуществлялся по ключевым словам: база данных, вредное воздействие, биомаркеры, белки, производственная среда, в ходе него была включена доступная информация о существующих базах данных белков – биомаркеров вредного воздействия. Первоначально выявлено 300 статей, из которых после анализа отобрано 40 публикаций, содержащих систематизированные данные исследований, направленных на выявление биомаркеров отрицательного влияния производственной среды на рабочих.

Результаты. Найдено 8 информационных источников, соответствующих критериям поиска. Базы данных разделены на 2 типа по доступности и наличию необходимых сведений. Тип I – включающие результаты исследований влияния производственной среды на содержание биомаркеров (белков) в организме, однако не имеющие свободного доступа. Тип II – с открытым доступом для пользователей, но не содержащие прямой информации о белковых биомаркерах, связанных с профессиональной экспозицией.

Заключение. Найденные базы данных исследований воздействия веществ, имеющих негативное влияние на организм человека, включают информацию о белках или в несистематизированном виде, или в закрытом доступе. Исходя из этого, задача формирования подобных общедоступных информационных источников представляется актуальной.

Ключевые слова: база данных, вредное воздействие, биомаркеры, производственная среда, белки, обзор литературы.

Для цитирования: Гиззатуллина О.И., Чемезов А.И. К вопросу о базах данных белковых биомаркеров воздействия вредных факторов окружающей среды на организм человека (обзор литературы) // Здоровье населения и среда обитания. 2024. Т. 32. № 2. С. 75–81. doi: 10.35627/2219-5238/2024-32-2-75-81

On the Databases of Protein Biomarkers of Human Exposure to Environmental Hazards: A Literature Review

Olga I. Gizzatullina, Aleksei I. Chemezov

Yekaterinburg Medical Research Center for Prophylaxis and Health Protection in Industrial Workers,
30 Popov Street, Yekaterinburg, 620014, Russian Federation

Summary

Introduction: Both environmental and occupational exposure to hazardous chemicals is a public health challenge since it can induce oxidative stress, lipid peroxidation, and protein modifications. Creation of a database linking work-related risk factors and adverse human health outcomes based on protein biomarkers will contribute to the development of effective preventive and therapeutic approaches.

Objective: To conduct a systematic review of literature to identify databases containing information about human protein biomarkers of exposure to environmental risk factors.

Materials and methods: We examined Russian and English-language publications containing information about existing protein biomarker databases issued in 2003–2023 and found in Elibary.ru, PubMed, Google Scholar, Scopus, and Research Gate using the following keywords: database, adverse effect, biomarkers, proteins, and occupational environment. Forty of 300 papers initially selected contained systematized data of research aimed at identifying biomarkers of occupational exposures and were therefore chosen for the review.

Results: Eight information sources have been found to match our search criteria. We have established that the databases are divided into two types by availability and accessibility of the information of interest. Type I databases contain research findings related to the impact of occupational exposures on the content of biomarkers (proteins) but have limited access. Type II databases are open to access, but they do not contain direct information about protein biomarkers associated with occupational exposures.

Conclusion: The existing databases either contain unsystematized data on protein biomarkers of adverse human health effects or are closed to access. Thus, the task of creating such publicly available information sources seems relevant.

Keywords: database, adverse effect, biomarkers, occupational environment, proteins, literature review.

Cite as: Gizzatullina OI, Chemezov AI. On the databases of protein biomarkers of human exposure to environmental hazards: A literature review. *Zdorov'e Naseleniya i Sreda Obitaniya*. 2024;32(2):75–81. (In Russ.) doi: 10.35627/2219-5238/2024-32-2-75-81

Введение. Развитие отраслей промышленности неразрывно связано с влиянием вредных факторов на состояние здоровья рабочих. Задачи обеспечения благоприятных условий окружающей среды и поддержания здоровья населения являются одними из самых актуальных. Улучшение условий труда и предотвращение заболеваний, связанных с производственными факторами, имеют ключевое значение в продлении работоспособного возраста населения, так как число смертей и инвалидностей, наступивших вследствие профзаболеваний, выше, чем из-за получения производственных травм [1]. Так, выявление биомаркеров, отражающих факт экспозиции, уровень восприимчивости организма, а также возникновение ответных реакций на воздействие химических веществ (таких как оксидативный стресс, перекисное окисление липидов и модификации белков) вносят значительный вклад в обнаружении заболеваний, связанных с трудовой деятельностью [2]. Их использование позволяет идентифицировать механизм действия химических веществ, а также этиологию и этапы развития заболевания, в том числе при низких дозах [3]. В этом случае создание и использование базы данных (далее – БД), включающей характер деятельности рабочих, наличие или отсутствие профессиональных заболеваний и результаты исследований тканей организма на наличие биомаркеров (таких, как белки), позволят разработать новые подходы к предотвращению патологий.

Цель исследования: провести системный обзор литературных источников для выявления баз данных, содержащих информацию о белковых биомаркерах в организме людей, подверженных влиянию вредных факторов окружающей среды.

Материалы и методы. Осуществлен научный обзор исследований на русском и английском языках с использованием информационных источников eLIBRARY.ru, PubMed, GoogleScholar, Scopus, ResearchGate за период 2003–2023 гг. Поиск осуществлялся по ключевым словам: база данных, вредное воздействие, биомаркеры, белки, производственная среда, в ходе него была включена доступная информация о существующих базах данных, отражающих наличие биомаркеров вредного воздействия. После этого она была разделена на два типа: содержащая сведения о влиянии производственной среды в зависимости от профессии рабочих (тип I); включающая изменения в организме под воздействием химических веществ (тип II). Из первоначально выявленных 300 статей после анализа отобрано 40 публикаций, содержащих систематизированные данные исследований, направленных на выявление биомаркеров отрицательного влияния производственной среды на рабочих. Преимущественно использовались источники, имеющие информацию о химических факторах воздействия, но также учитывались физические.

Результаты. В данном обзоре рассматриваются базы данных, отражающие связь заболеваний с воздействием вредных химических факторов. Критерием отбора выступало наличие информации об отрицательном воздействии и содержании белков

в организме. Найденные базы данных разделены на 2 типа.

В ходе исследования найдено 3 базы данных I типа, систематически отражающие результаты исследований о влиянии производственной среды на содержание биомаркеров (белков) в организмах работников. Также стоит отметить, что в обзор включены БД, относящиеся к различным отраслям промышленности и, следовательно, с различными типами воздействия, в том числе и химическим.

Базы данных I типа

- База данных обследований работников предприятий горнодобывающей промышленности, использованная для построения модели управления профессиональными рисками горнорабочих, предложенной Федеральным научным центром гигиены им. Ф.Ф. Эрисмана, для выявления ранних биомаркеров формирования профессиональной и общей патологии и определения критериев риска развития заболеваний [4].

- Персонифицированная база данных работников предприятий порошковой металлургии, созданная в результате реализации программы исследований по оценке факторов риска для рабочих метало-порошковой промышленности, – идея ее создания была выдвинута Федеральным научным центром медико-профилактических технологий управления рисками для здоровья населения г. Перми. Цель создания – обнаружение профессиональных заболеваний, ограничивающих продолжение трудовой деятельности, в условиях влияния вредных факторов производства [5].

- База данных Национального центра гигиены труда и профессиональных заболеваний (Казахстан), использованная для создания математической модели развития антракосиликоза у горнорабочих угольной промышленности. В ее основу легли данные истории болезни больных антракосиликозом [6].

В рассмотрение также были включены базы данных с классификацией на основании изолированного влияния химических веществ (тип II), поскольку их воздействие может происходить в результате осуществления производственных процессов.

- CTD – база данных сравнительной токсикогеномики, которая объединяет информацию о связи воздействия химических веществ с изменениями на генетическом и протеомном уровнях [7].

- TPDB – база данных белков – мишеней реактивных метаболитов. Она разработана с целью объединения информации о белках, о том, какие они могут образовывать аддукты и могут ли они быть мишенями для токсинов [8].

- Проект Exposome – это подход к оценке воздействия высокоприоритетных загрязнителей окружающей среды и поиск взаимосвязей между внешним воздействием и изменениями в организмах людей на молекулярном уровне [9].

- PubChem – база данных, содержащая сведения о химических веществах и их биологической активности [10].

- LINCS – база данных, в основе которой лежит создание сетевого понимания механизмов изменений в клетках живых организмов, происходящих при воздействии различных химических веществ [11].

Из всех выявленных баз данных в открытом доступе находятся CTD, Exposome, PubChem, LINCS. Отечественные базы данных находятся в закрытом доступе.

Базы данных I типа

Описанные ниже базы по большей части находятся в стадии разработки и не имеют свободного доступа. В найденных источниках имеются сведения о внесении информации в базы данных, о биомаркерах, в том числе и белковых, с целью ранней диагностики профессиональных заболеваний.

База данных обследований работников предприятий горнодобывающей промышленности

В целях выявления ранних биомаркеров формирования профессиональной и общей патологий, а также определения критериев риска развития заболеваний у рабочих горнодобывающей промышленности была предложена модель управления профессиональными рисками [4]. Апробация проводилась в исследовании при участии 1067 рабочих, результаты содержат данные о показателях: липидного спектра, периферической гемодинамики, ферментативного звена антиоксидантной защиты (супероксиддисмутаза, каталаза, церулоплазмин, эндотелин), проводящей функции нервов. В результате проведения работ выявлены преобладающие факторы, которые обуславливают степень развития заболеваний периферической нервной системы, к ним относятся химические, физические, психологические и физиологические. Определена категория профессионального риска для различных групп рабочих, что требует индивидуального подхода при создании профилактических мероприятий, направленных на совершенствование условий труда.

Персонализированная база данных работников предприятий порошковой металлургии

Федеральным научным центром медико-профилактических технологий управления рисками для здоровья населения г. Перми предложен алгоритм программы исследований, цель которого – обнаружение профессиональных болезней у рабочих металло-порошковой промышленности. Он включает: проведение клинических осмотров, лабораторные и функциональные исследования, исследования на факт возможных ответов организма на отрицательные воздействия (донозологические сдвиги) [5]. Алгоритм был протестирован на предприятии порошковой металлургии. В процессе проведения испытаний создана персонализированная база, в которую внесена информация по 281 работающему, включающая индивидуальные факторы риска и результаты обследований. Данные распределены по категориям: профессиональные и социальные факторы, характеристика питания, результаты клинического осмотра и лабораторных исследований (общий, биохимический и иммунологический анализ крови, биохимический анализ мочи, анализ биосред на содержание химических соединений). При этом учитывалось то, что на различных производственных стадиях, включающих получение порошка, формирование, спекание и окончательная обработка заготовок – присутствует уникальный набор вредных факторов. Как следствие, рабочие подвержены разному воздействию, харак-

терному для их специализации [12]. Рассмотрены изменения в анализах крови, связанные с влиянием вредных факторов производственной среды, одной из профессий данной отрасли (паяльщик). Установлено, что воздействие толуола и ксилола приводило к снижению уровня гемоглобина в крови; действие ацетона сопровождалось снижением уровня альбуминов и повышением уровня α -, β - и γ -глобулинов в сыворотке крови [5].

База данных Национального центра гигиены труда и профессиональных заболеваний

Для составления математической модели развития антракосиликоза проведен анализ 696 историй болезни горнорабочих угольной промышленности, проходивших лечение в условиях стационара терапевтического отделения Национального центра гигиены труда и профессиональных заболеваний с 1994 по 2008 год [6]. Данное исследование можно считать актуальным, поскольку частота хронических заболеваний легких является показателем состояния окружающей среды и социально-экономического развития [13]. В исследовании принимали участие три группы больных антракосиликозом, в том числе осложненным бронхитом, в возрасте от 30 до 60 лет, со стажем работы во вредных условиях труда от 4 до 35 лет. Модель включала следующие параметры: дата установления диагноза; стадия заболевания; наличие эмфиземы; дыхательная недостаточность первой степени; форсированная жизненная емкость легких; данные лабораторных исследований, включающие белковые показатели [6].

Базы данных II типа

В свою очередь базы данных CTD, exposome, PubChem, LINCS имеют открытый доступ для пользователей, но не содержат прямой информации о белковых биомаркерах, связанных с профессиональной экспозицией. Сильная сторона этих хранилищ состоит в предоставляемых ссылках на первоисточники, в которых имеется информация об исследованиях, в том числе о влиянии производственной среды на рабочих и о белках в их организмах. Недостатком данного типа баз данных, с точки зрения имеющих критериев отбора, является отсутствие акцента на производственных факторах вредного воздействия.

База данных сравнительной токсикогеномики (CTD)

База данных сравнительной токсикогеномики объединяет информацию о взаимодействии между химическими веществами и изменениями экспрессии генов белков, связывая ее с функциональными данными о патогенезе, которые впоследствии становятся основой гипотез механизмов возникновения патологий, вызванных негативным воздействием окружающей среды [14–16]. Информация для составления БД отбирается из научной литературы, анализируется и систематизируется. Стоит отметить, что пользователям доступны первоисточники оригинальных публикаций, благодаря чему не возникает препятствий к получению исходных данных [17].

Научную литературу, с помощью которой составлена CTD, разделяют на группы, посвященные:

1) химическим взаимодействиям между генами и белками;

2) взаимосвязи химических веществ и заболеваний;

3) взаимосвязи генов, белков и заболеваний [18].

Данные систематизируются с использованием словарей и онтологий для химических веществ, генов или белков, заболеваний, молекулярных взаимодействий и организмов [19]. В CTD входят: токсикогеномные данные позвоночных и беспозвоночных, включая 124 000 химических веществ, 2,6 миллиона последовательностей генов и белков и связанную с ними генную онтологию, 128 000 таксонов и 6300 болезней человека [7, 20]. Таким образом, CTD способствует пониманию значимости консервативных последовательностей ДНК – генетической основы переменной чувствительности к агентам окружающей среды [21].

База данных белков-мишеней реактивных метаболитов (TPDB)

К данной базе в настоящее время нет доступа. В 2020 году отмечалось снижение ее востребованности. По имеющимся литературным данным, есть сведения о том, что она была создана для понимания того, какие белки образуют аддукты и могут ли некоторые из них выступать общими мишенями нескольких токсинов [22]. Негативные эффекты органических соединений могут быть обусловлены их биотрансформацией в химически активные метаболиты, которые ковалентно связываются с клеточными белками, что нарушает их структуру и функции [8]. TPDB содержала информацию о: токсикантах и тканях, подверженных поражению ими; идентифицированных целевых белках; примечаниях об используемых методах разделения и идентификации [23].

Проект Exposome

Главной целью проекта Exposome является разработка базы данных, посвященной биомаркерам воздействия факторов риска заболеваний, связанных с влиянием окружающей среды [9]. Также база данных создана для поиска взаимосвязей между внешним антропогенным воздействием и профилями молекулярных особенностей у одного человека [24]. Данная работа проводится для нахождения взаимосвязей между заболеванием и вредными воздействиями окружающей среды на человека и оценки влияния на следующие поколения [25]. Ее цель – предоставить данные по всем известным биомаркерам воздействия: загрязняющим веществам и контаминантам, измеренным в популяционных исследованиях, которые представлены в научных публикациях [26, 27].

База данных PubChem

База данных содержит сведения о химических веществах и их биологической активности [28–31]. По большинству включенных в базу соединений предоставляется вся известная на данный момент информация об их химических и физических свойствах, токсичности и воздействии на организм человека, а также о молекулярных мишенях [10]. Данные сведения позволяют находить связь между вредным веществом и его влиянием на организм [32]. Источниками информации являются: государственные организации; научно-исследовательские

лаборатории; фармацевтические компании; издательства в сфере научной публикации; химико-биологические ресурсы, а также уже существующие базы данных [33, 34].

База данных LINCS

Направленность БД LINCS – изучение взаимосвязи процессов в клетках живых организмов, расширение информационной базы о клеточных реакциях тканей на химические, генетические изменения и изменения микросреды [35–38]. Она включает сведения как о различных типах клеток в целом, так и в частности о малых молекулах, генах, белках, пептидах, антителах [11]. Консорциум LINCS собирает данные с помощью различных способов, включающих мультиомиксные подходы (протеомика, транскриптомика, эпигеномика), визуализацию и анализ микрочипов микроокружения (MEMA), платформу для оценки эффекта комбинирования факторов микроокружения (например, факторы роста, белки внеклеточного матрикса) на определенные биологические конечные точки (пролиферация, дифференцировка, повреждение ДНК и старение) [39]. Веб-сайт и портал данных LINCS содержат ссылки на источники информации и программное обеспечение, которое может быть использовано для анализа данных [40].

Обсуждение. Изучение доступной информации показало, что основой для классификации существующих баз данных является или единичный фактор, или их совокупность, относящаяся к профессиональной деятельности. Оба подхода имеют свои преимущества и недостатки.

Для БД, аккумулирующих информацию об эффектах воздействия единичных веществ, преимуществом является большой объем и доступность информации, источником которой являются оригинальные исследования. Однако данный подход не совсем точно отражает критерии, в соответствии с которыми проводилось исследование, так как целью было рассмотрение совокупности факторов химического воздействия.

Оценка изменений показателей состояния здоровья на основании изучения влияния профессионального воздействия учитывает возможные эффекты комбинированного воздействия, однако при оценке условий труда учитывается лишь часть наиболее значимых факторов. Однако вариации совокупности реальных факторов вследствие совершенствования технологических процессов либо по другим возможным причинам приведут к изменениям характеристического паттерна факторов профессии и отклика организма на него.

Важнейший критерий ценности БД – доступность сведений, содержащихся в ней. Большинство информационных источников, систематизирующих эффекты воздействия на основании единичных веществ, находятся в открытом доступе, а также содержат ссылки на первоисточники, что дает возможность извлечь дополнительные сведения, полученные в оригинальных исследованиях (CTD, TPDB, Exposome, PubChem, LINCS) [7–11]. Найденные закрытые БД (База данных обследований работников предприятий горнодобывающей промышленности,

Персонализированная база данных работников предприятий порошковой металлургии, База данных Национального центра гигиены труда и профессиональных заболеваний) содержат информацию, полученную в результате исследований авторов баз, причем большинство таких агрегаторов проводят систематизацию данных на основании профессии, что затрудняет исследования по оценке влияния факторов таковой на здоровье рабочих разными научными группами [4–6].

В найденных базах данных сведения о веществах-биомаркерах белковой природы присутствуют в ограниченном объеме. По-видимому, это обусловлено трудоемкостью и дороговизной проведения подобных исследований в скрининговом формате. Показатели, включающие исследование содержания веществ белковой природы, представляли собой общепринятые: общий белок, гемоглобин, фибриноген, альбумины, α -, β - и γ -глобулины [5]. Базы данных исключительно белковых биомаркеров на момент написания статьи обнаружить не удалось.

Заключение. Таким образом, данный анализ показывает, что существующие базы данных отражают информацию о воздействии вредных факторов производственной среды на здоровье людей, однако сведения о белковых биомаркерах такого воздействия представлены в ограниченном объеме. Лишь три из восьми описанных выше БД представляют сведения о связи профессионального воздействия с содержанием белков, находящиеся в ограниченном доступе. Оставшиеся пять баз собирают информацию из научных публикаций по единичным факторам окружающей среды. Вследствие этого на сегодня отсутствует удобный общедоступный информационный канал, предоставляющий доступ к материалам исследований профессиональных рисков в зависимости от рода деятельности рабочих, с подробными описаниями историй болезни и результатами анализов по выявлению биомаркеров, в качестве которых будут выступать белки.

СПИСОК ЛИТЕРАТУРЫ

1. Мясегутова Л.М., Абдрахманова Е.Р., Габдулвалеева Э.Ф., Перминова В.А. Риск формирования профессиональной, производственно-обусловленной и общесоматической патологии у работников металлургических производств // Вестник Авиценны. 2021. Т. 23. № 2. С. 280–290. doi: 10.25005/2074-0581-2021-23-2-280-290
2. Зайцева Н.В., Май И.В., Шур П.З., Алексеев В.Б., Лебедева-Несевря Н.А. Практика оценки и управления рисками здоровью на базе новых методов и подходов // Анализ риска здоровью в стратегии государственного социально-экономического развития. Пермь: Изд-во Пермск. нац. исслед. политехн. ун-та, 2014. С. 625–711.
3. Schulte PA, Hauser JE. The use of biomarkers in occupational health research, practice, and policy. *Toxicol Lett.* 2012;213(1):91–99. doi: 10.1016/j.toxlet.2011.03.027
4. Жеглова А.В. Персонализированный профессиональный риск у работников горнодобывающих предприятий. Здоровье и безопасность на рабочем месте: Материалы III Международной научно-практической конференции, Новополюцк-Полоцк, Беларусь, 15–17 мая 2019 года / Под редакцией Бухтияров И.В., Рыбина Т.М. Новополюцк-Полоцк, Беларусь: Федеральное бюджетное научное учреждение «научно-исследовательский институт медицины труда имени академика Н.Ф. Измерова», 2019. С. 113–117.
5. Зайцева Н.В., Шур П.З., Костарев В.Г., Клименко А.Р., Пузиков Ю.Г., Болотова Е.И. Обоснование программы исследований по оценке факторов риска на предприятии порошковой металлургии // Вестник ПГУ. Биология. 2010. № 2. С. 50–56.
6. Ажиметова Г.Н., Койгельдинова Ш.С. Прогнозирование сроков развития антракосиликоза у горнорабочих угольной промышленности // Медицина и экология. 2010. № 2 (55). С. 46–49.
7. Davis AP, Murphy CG, Rosenstein MC, Wiegiers TC, Mattingly CJ. The Comparative Toxicogenomics Database facilitates identification and understanding of chemical–gene–disease associations: Arsenic as a case study. *BMC Med Genomics.* 2008;1:48. doi: 10.1186/1755-8794-1-48
8. Hanzlik RP, Koen YM, Theertham B, Dong Y, Fang J. The reactive metabolite target protein database (TPDB) – a web-accessible resource. *BMC Bioinformatics.* 2007;8:95. doi: 10.1186/1471-2105-8-95
9. Lamurias A, Jesus S, Neveu V, Salek RM, Couto FM. Information retrieval using machine learning for biomarker curation in the Exposome–Explorer. *Front Res Metr Anal.* 2021;6:689264. doi: 10.3389/frma.2021.689264
10. Kim S. Public chemical databases. In: Ranganathan S, Gribskov M, Nakai K, Schönbach C. *Encyclopedia of Bioinformatics and Computational Biology.* Academic Press; 2019;2:628–639. doi: 10.1016/B978-0-12-809633-8.20192-1
11. Pawar G, Madden JC, Ebbrell D, Firman JW, Cronin MTD. *In silico* toxicology data resources to support read-across and (Q)SAR. *Front Pharmacol.* 2019;10:561. doi: 10.3389/fphar.2019.00561
12. Зайцева Н.В., Шур П.З., Клименко А.Р., Устинова О.Ю., Лебедева-Несевря Н.А., Костарев В.Г. Гигиеническая оценка факторов риска на производствах порошковой металлургии // Медицина труда и промышленная экология. 2011. № 11. С. 16–20.
13. Кыткова О.Ю., Гвозденко Т.А., Антонюк М.В. Современные аспекты распространенности хронических бронхолегочных заболеваний // Бюллетень физиологии и патологии дыхания. 2017. № 64. С. 94–100.
14. Davis AP, Grondin CJ, Johnson RJ, et al. Comparative Toxicogenomics Database (CTD): Update 2021. *Nucleic Acids Res.* 2021;49(D1):D1138–D1143. doi: 10.1093/nar/gkaa891
15. Mattingly CJ. Chemical databases for environmental health and clinical research. *Toxicol Lett.* 2009;186(1):62–65. doi: 10.1016/j.toxlet.2008.10.003
16. Cai Y, Rosen Vollmar AK, Johnson CH. Analyzing metabolomics data for environmental health and exposome research. *Methods Mol Biol.* 2020;2104:447–467. doi: 10.1007/978-1-0716-0239-3_22
17. Davis AP, Wiegiers TC, Johnson RJ, Sciaky D, Wiegiers J, Mattingly CJ. Comparative Toxicogenomics Database (CTD): Update 2023. *Nucleic Acids Res.* 2023;51(D1):D1257–D1262. doi: 10.1093/nar/gkac833
18. Mattingly CJ, Colby GT, Rosenstein MC, Forrest JN Jr, Boyer JL. Promoting comparative molecular studies in environmental health research: An overview of the Comparative Toxicogenomics Database (CTD). *Pharmacogenomics J.* 2004;4(1):5–8. doi: 10.1038/sj.tpj.6500225
19. Mattingly CJ, Rosenstein MC, Colby GT, Forrest JN Jr, Boyer JL. The Comparative Toxicogenomics Database (CTD): A resource for comparative toxicological studies. *J Exp Zool A Comp Exp Biol.* 2006;305(9):689–692. doi: 10.1002/jez.a.307
20. Wiegiers TC, Davis AP, Cohen KB, Hirschman L, Mattingly CJ. Text mining and manual curation of chemical–gene–disease networks for the Comparative Toxicogenomics Database (CTD). *BMC Bioinformatics.* 2009;10:326. doi: 10.1186/1471-2105-10-326
21. Mattingly CJ, Colby GT, Forrest JN, Boyer JL. The Comparative Toxicogenomics Database (CTD). *Environ Health Perspect.* 2003;111(6):793–795. doi: 10.1289/ehp.6028

22. Hanzlik RP, Koen YM, Garrett MJ. The reactive metabolite target protein database archive. *Medicinal Chemistry Scholarly Works*. 2020. Accessed February 21, 2024. <https://kuscholarworks.ku.edu/handle/1808/30592>
23. Hanzlik RP, Fang J, Koen YM. Filling and mining the reactive metabolite target protein database. *Chem Biol Interact*. 2009;179(1):38-44. doi: 10.1016/j.cbi.2008.08.016
24. Zhao F, Li L, Chen Y, et al. Risk-based chemical ranking and generating a prioritized human exposome database. *Environ Health Perspect*. 2021;129(4):47014. doi: 10.1289/EHP7722
25. International Agency for Research on Cancer/World Health Organization. What is Exposome-Explorer? Accessed June 2, 2023. <http://exposome-explorer.iarc.fr/about>
26. Neveu V, Nicolas G, Salek RM, Wishart DS, Scalbert A. Exposome-Explorer 2.0: An update incorporating candidate dietary biomarkers and dietary associations with cancer risk. *Nucleic Acids Res*. 2020;48(D1):D908-D912. doi: 10.1093/nar/gkz1009
27. Neveu V, Moussy A, Rouaix H, et al. Exposome-Explorer: A manually-curated database on biomarkers of exposure to dietary and environmental factors. *Nucleic Acids Res*. 2017;45(D1):D979-D984. doi: 10.1093/nar/gkw980
28. Wang Y, Xiao J, Suzek TO, Zhang J, Wang J, Bryant SH. PubChem: A public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res*. 2009;37(Web Server issue):W623-33. doi: 10.1093/nar/gkp456
29. Wang Y, Bolton E, Dracheva S, et al. An overview of the PubChem BioAssay resource. *Nucleic Acids Res*. 2010;38(Database issue):D255-66. doi: 10.1093/nar/gkp965
30. Kim S, Chen J, Cheng T, et al. PubChem 2023 update. *Nucleic Acids Res*. 2023;51(D1):D1373-D1380. doi: 10.1093/nar/gkac956
31. Li Q, Cheng T, Wang Y, Bryant SH. PubChem as a public resource for drug discovery. *Drug Discov Today*. 2010;15(23-24):1052-1057. doi: 10.1016/j.drudis.2010.10.003
32. National Library of Medicine. About PubChem. Accessed June 2, 2023. <https://pubchem.ncbi.nlm.nih.gov/docs/about>
33. Kim S, Thiessen PA, Bolton EE, et al. PubChem substance and compound databases. *Nucleic Acids Res*. 2016;44(D1):D1202-D1213. doi: 10.1093/nar/gkv951
34. Kim S, Chen J, Cheng T, et al. PubChem in 2021: New data content and improved web interfaces. *Nucleic Acids Res*. 2021;49(D1):D1388-D1395. doi: 10.1093/nar/gkaa971
35. Musa A, Tripathi S, Dehmer M, Emmert-Streib F. L1000 Viewer: A search engine and web interface for the LINCS data repository. *Front Genet*. 2019;10:557. doi: 10.3389/fgene.2019.00557
36. Xie Z, Kropiwnicki E, Wojciechowski ML, et al. Getting started with LINCS datasets and tools. *Curr Protoc*. 2022;2(7):e487. doi: 10.1002/cpz1.487
37. Cheng L, Li L. Systematic quality control analysis of LINCS data. *CPT Pharmacometrics Syst Pharmacol*. 2016;5(11):588-598. doi: 10.1002/psp4.12107
38. Stathias V, Turner J, Koletti A, et al. LINCS Data Portal 2.0: Next generation access point for perturbation-response signatures. *Nucleic Acids Res*. 2020;48(D1):D431-D439. doi: 10.1093/nar/gkz1023
39. Koletti A, Terryn R, Stathias V, et al. Data portal for the Library of Integrated Network-based Cellular Signatures (LINCS) program: Integrated access to diverse large-scale cellular perturbation response data. *Nucleic Acids Res*. 2018;46(D1):D558-D566. doi: 10.1093/nar/gkx1063
40. Vempati UD, Chung C, Mader C, et al. Metadata standard and data exchange specifications to describe, model, and integrate complex and diverse high-throughput screening data from the Library of Integrated Network-based Cellular Signatures (LINCS). *J Biomol Screen*. 2014;19(5):803-816. doi: 10.1177/1087057114522514

REFERENCES

1. Masyagutova LM, Abdrakhmanova ER, Gabdulvaleeva EF, Perminova VA. Risk of occupational, work-related, and somatic morbidity among metallurgical industries workers. *Vestnik Avitsenny*. 2021;23(2):280-290. doi: 10.25005/2074-0581-2021-23-2-280-290
2. Zaitseva NV, May IV, Shur PZ, Alekseev VB, Lebedeva-Nesevrya NA. [Practice of health risk assessment and management based on new methods and approaches.] In: *Health Risk Analysis in the Strategy of State Socio-Economic Development*. Perm: Perm National Research Polytechnic University; 2014: 625-711. (In Russ.)
3. Schulte PA, Hauser JE. The use of biomarkers in occupational health research, practice, and policy. *Toxicol Lett*. 2012;213(1):91-99. doi: 10.1016/j.toxlet.2011.03.027
4. Zheglova AV. Personified occupational risk for employees of mining enterprises. In: Bukhtiyarov IV, Rybina TM, eds. *Health and Safety at the Workplace: Proceedings of the III International Scientific Forum, Novopolotsk-Polotsk, Belarus, May 15-17, 2019*. Novopolotsk-Polotsk, Belarus: Polycraft LLC; 2019;1(3):113-117. (In Russ.) doi: 10.31089/978-985-7153-76-3-2019-1-3-113-117
5. Zaitseva NV, Shur PZ, Kostarev VG, Klimenko AR, Puzikov YuG, Bolotova EI. Substantiation of a research programme for powder metallurgy risk factor assessment. *Vestnik Permskogo Universiteta. Series: Biology*. 2010;(2):50-56. (In Russ.)
6. Azhimetova GN, Koigeldinova SS. Prediction of anthraco-silicosis development terms at miners of coal industry. *Meditsina i Ekologiya*. 2010;(2(55)):46-49. (In Russ.)
7. Davis AP, Murphy CG, Rosenstein MC, Wiegerts TC, Mattingly CJ. The Comparative Toxicogenomics Database facilitates identification and understanding of chemical-gene-disease associations: Arsenic as a case study. *BMC Med Genomics*. 2008;1:48. doi: 10.1186/1755-8794-1-48
8. Hanzlik RP, Koen YM, Theertham B, Dong Y, Fang J. The reactive metabolite target protein database (TPDB) – a web-accessible resource. *BMC Bioinformatics*. 2007;8:95. doi: 10.1186/1471-2105-8-95
9. Lamurias A, Jesus S, Neveu V, Salek RM, Couto FM. Information retrieval using machine learning for biomarker curation in the Exposome-Explorer. *Front Res Metr Anal*. 2021;6:689264. doi: 10.3389/frma.2021.689264
10. Kim S. Public chemical databases. In: Ranganathan S, Gribskov M, Nakai K, Schönbach C. *Encyclopedia of Bioinformatics and Computational Biology*. Academic Press; 2019;2:628-639. doi: 10.1016/B978-0-12-809633-8.20192-1
11. Pawar G, Madden JC, Ebbrell D, Firman JW, Cronin MTD. *In silico* toxicology data resources to support read-across and (Q)SAR. *Front Pharmacol*. 2019;10:561. doi: 10.3389/fphar.2019.00561
12. Zaitseva NV, Shur PZ, Klimenko AR, Ustinova OYu, Lebedeva-Nesevria NA, Kostarev VG. Hygienic evaluation of risk factors on powder metallurgy production. *Meditsina Truda i Promyshlennaya Ekologiya*. 2011;(11):16-20. (In Russ.)
13. Kytikova OYu, Gvozdenko TA, Antonyuk MV. Modern aspects of prevalence of chronic bronchopulmonary diseases. *Byulleten' Fiziologii i Patologii Dykhaniya*. 2017;(64):94-100. (In Russ.) doi: 10.12737/article_5936346dfc1f3.32482903
14. Davis AP, Grondin CJ, Johnson RJ, et al. Comparative Toxicogenomics Database (CTD): Update 2021. *Nucleic Acids Res*. 2021;49(D1):D1138-D1143. doi: 10.1093/nar/gkaa891
15. Mattingly CJ. Chemical databases for environmental health and clinical research. *Toxicol Lett*. 2009;186(1):62-65. doi: 10.1016/j.toxlet.2008.10.003
16. Cai Y, Rosen Vollmar AK, Johnson CH. Analyzing metabolomics data for environmental health and exposome research. *Methods Mol Biol*. 2020;2104:447-467. doi: 10.1007/978-1-0716-0239-3_22

17. Davis AP, Wiegiers TC, Johnson RJ, Sciaky D, Wiegiers J, Mattingly CJ. Comparative Toxicogenomics Database (CTD): Update 2023. *Nucleic Acids Res.* 2023;51(D1):D1257–D1262. doi: 10.1093/nar/gkac833
18. Mattingly CJ, Colby GT, Rosenstein MC, Forrest JN Jr, Boyer JL. Promoting comparative molecular studies in environmental health research: An overview of the Comparative Toxicogenomics Database (CTD). *Pharmacogenomics J.* 2004;4(1):5–8. doi: 10.1038/sj.tpj.6500225
19. Mattingly CJ, Rosenstein MC, Colby GT, Forrest JN Jr, Boyer JL. The Comparative Toxicogenomics Database (CTD): A resource for comparative toxicological studies. *J Exp Zool A Comp Exp Biol.* 2006;305(9):689–692. doi: 10.1002/jez.a.307
20. Wiegiers TC, Davis AP, Cohen KB, Hirschman L, Mattingly CJ. Text mining and manual curation of chemical-gene-disease networks for the Comparative Toxicogenomics Database (CTD). *BMC Bioinformatics.* 2009;10:326. doi: 10.1186/1471-2105-10-326
21. Mattingly CJ, Colby GT, Forrest JN, Boyer JL. The Comparative Toxicogenomics Database (CTD). *Environ Health Perspect.* 2003;111(6):793–795. doi: 10.1289/ehp.6028
22. Hanzlik RP, Koen YM, Garrett MJ. The reactive metabolite target protein database archive. *Medicinal Chemistry Scholarly Works.* 2020. Accessed February 21, 2024. <https://kuscholarworks.ku.edu/handle/1808/30592>
23. Hanzlik RP, Fang J, Koen YM. Filling and mining the reactive metabolite target protein database. *Chem Biol Interact.* 2009;179(1):38–44. doi: 10.1016/j.cbi.2008.08.016
24. Zhao F, Li L, Chen Y, et al. Risk-based chemical ranking and generating a prioritized human exposome database. *Environ Health Perspect.* 2021;129(4):47014. doi: 10.1289/EHP7722
25. International Agency for Research on Cancer/World Health Organization. What is Exposome-Explorer? Accessed June 2, 2023. <http://exposome-explorer.iarc.fr/about>
26. Neveu V, Nicolas G, Salek RM, Wishart DS, Scalbert A. Exposome-Explorer 2.0: An update incorporating candidate dietary biomarkers and dietary associations with cancer risk. *Nucleic Acids Res.* 2020;48(D1):D908–D912. doi: 10.1093/nar/gkz1009
27. Neveu V, Moussy A, Rouaix H, et al. Exposome-Explorer: A manually-curated database on biomarkers of exposure to dietary and environmental factors. *Nucleic Acids Res.* 2017;45(D1):D979–D984. doi: 10.1093/nar/gkw980
28. Wang Y, Xiao J, Suzek TO, Zhang J, Wang J, Bryant SH. PubChem: A public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res.* 2009;37(Web Server issue):W623–33. doi: 10.1093/nar/gkp456
29. Wang Y, Bolton E, Dracheva S, et al. An overview of the PubChem BioAssay resource. *Nucleic Acids Res.* 2010;38(Database issue):D255–66. doi: 10.1093/nar/gkp965
30. Kim S, Chen J, Cheng T, et al. PubChem 2023 update. *Nucleic Acids Res.* 2023;51(D1):D1373–D1380. doi: 10.1093/nar/gkac956
31. Li Q, Cheng T, Wang Y, Bryant SH. PubChem as a public resource for drug discovery. *Drug Discov Today.* 2010;15(23–24):1052–1057. doi: 10.1016/j.drudis.2010.10.003
32. National Library of Medicine. About PubChem. Accessed June 2, 2023. <https://pubchem.ncbi.nlm.nih.gov/docs/about>
33. Kim S, Thiessen PA, Bolton EE, et al. PubChem substance and compound databases. *Nucleic Acids Res.* 2016;44(D1):D1202–D1213. doi: 10.1093/nar/gkv951
34. Kim S, Chen J, Cheng T, et al. PubChem in 2021: New data content and improved web interfaces. *Nucleic Acids Res.* 2021;49(D1):D1388–D1395. doi: 10.1093/nar/gkaa971
35. Musa A, Tripathi S, Dehmer M, Emmert-Streib F. L1000 Viewer: A search engine and web interface for the LINCS data repository. *Front Genet.* 2019;10:557. doi: 10.3389/fgene.2019.00557
36. Xie Z, Kropiwnicki E, Wojciechowski ML, et al. Getting started with LINCS datasets and tools. *Curr Protoc.* 2022;2(7):e487. doi: 10.1002/cpz1.487
37. Cheng L, Li L. Systematic quality control analysis of LINCS data. *CPT Pharmacometrics Syst Pharmacol.* 2016;5(11):588–598. doi: 10.1002/psp4.12107
38. Stathias V, Turner J, Koletti A, et al. LINCS Data Portal 2.0: Next generation access point for perturbation–response signatures. *Nucleic Acids Res.* 2020;48(D1):D431–D439. doi: 10.1093/nar/gkz1023
39. Koletti A, Terry R, Stathias V, et al. Data portal for the Library of Integrated Network-based Cellular Signatures (LINCS) program: Integrated access to diverse large-scale cellular perturbation response data. *Nucleic Acids Res.* 2018;46(D1):D558–D566. doi: 10.1093/nar/gkx1063
40. Vempati UD, Chung C, Mader C, et al. Metadata standard and data exchange specifications to describe, model, and integrate complex and diverse high-throughput screening data from the Library of Integrated Network-based Cellular Signatures (LINCS). *J Biomol Screen.* 2014;19(5):803–816. doi: 10.1177/1087057114522514

Сведения об авторах:

Гиззатуллина Ольга Ирековна – лаборант-исследователь отдела молекулярной биологии и электронной микроскопии; e-mail: gizzatullinaoi@ymrc.ru; ORCID: <https://orcid.org/0009-0009-7807-2139>.

☒ Чemezov Алексей Игоревич – научный сотрудник отдела молекулярной биологии и электронной микроскопии; e-mail: chemezov@ymrc.ru; ORCID: <https://orcid.org/0000-0001-6167-7347>.

Информация о вкладе авторов: обзор литературы и обработка данных, подготовка проекта рукописи: Гиззатуллина О.И.; концепция и дизайн исследования, редактирование, утверждение окончательного варианта рукописи: Чemezov А.И. Оба автора ознакомились с результатами работы и одобрили окончательный вариант рукописи.

Соблюдение этических стандартов: данное исследование не требует представления заключения комитета по био-медицинской этике или иных документов.

Финансирование: исследование проведено без спонсорской поддержки.

Конфликт интересов: авторы декларируют отсутствие явных и потенциальных конфликтов интересов в связи с публикацией данной статьи.

Статья получена: 24.08.23 / Принята к публикации: 09.02.24 / Опубликовано: 29.02.24

Author information:

Olga I. Gizzatullina, Research Laboratory Assistant, Department of Molecular Biology and Electron Microscopy; e-mail: gizzatullinaoi@ymrc.ru; ORCID: <https://orcid.org/0009-0009-7807-2139>.

☒ Aleksei I. Chemezov, Researcher, Department of Molecular Biology and Electron Microscopy; e-mail: chemezov@ymrc.ru; ORCID: <https://orcid.org/0000-0001-6167-7347>.

Author contributions: data collection and processing, draft manuscript preparation: Gizzatullina O.I.; study conception and design, critical revision: Chemezov A.I. Both authors reviewed the results and approved the final version of the manuscript.

Compliance with ethical standards: Not applicable.

Funding: This research received no external funding.

Conflict of interest: The authors have no conflicts of interest to declare.

Received: August 24, 2023 / Accepted: February 9, 2024 / Published: February 29, 2024