

Электронные истории болезни и биобанкирование

Копылова О. В., Ершова А. И., Ефимова И. А., Блохина А. В., Лимонова А. С.,
Борисова А. Л., Покровская М. С., Драпкина О. М.

ФГБУ "Национальный медицинский исследовательский центр терапии и профилактической медицины" Минздрава России.
Москва, Россия

Сохранение биообразцов для последующих исследований является фундаментальным компонентом трансляционной медицины. При этом ценность хранимых биообразцов во многом определяется наличием сопряженной с ними клинической и других видов информации. Электронные истории болезни представляют собой уникальный источник большого количества информации, получаемой на протяжении длительного времени. В связи с этим генетические и другие виды данных, получаемых при анализе биообразцов, могут быть сопряжены с фенотипической и другими видами информации, хранящейся в электронных историях болезни, что открывает новые горизонты как для проведения широкомасштабных генетических исследований, так и для совершенствования оказания медицинской помощи. Целью настоящего обзора является анализ литературы, посвященной возможностям объединения электронных историй болезни и баз данных биобанков в научной и клинической практике.

Ключевые слова: биобанкирование, электронная история болезни, медицинская информационная система, геномная медицина, трансляционные исследования.

Отношения и деятельность: нет.

Поступила 20/09-2022

Рецензия получена 05/10-2022

Принята к публикации 13/10-2022



Для цитирования: Копылова О. В., Ершова А. И., Ефимова И. А., Блохина А. В., Лимонова А. С., Борисова А. Л., Покровская М. С., Драпкина О. М. Электронные истории болезни и биобанкирование. *Кардиоваскулярная терапия и профилактика*. 2022;21(11):3425. doi:10.15829/1728-8800-2022-3425. EDN OGIGOO

Electronic medical records and biobanking

Kopylova O. V., Ershova A. I., Efimova I. A., Blokhina A. V., Limonova A. S., Borisova A. L., Pokrovskaya M. S., Drapkina O. M.
National Medical Research Center for Therapy and Preventive Medicine. Moscow, Russia

Biosample preservation for future research is a fundamental component of translational medicine. At the same time, the value of stored biosamples is largely determined by the presence of related clinical data and other information. Electronic medical records are a unique source of a large amount of information received over a long period of time. In this regard, genetic and other types of data obtained from the biosample analysis can be associated with phenotypic and other types of information stored in electronic medical records, which pushes the boundaries in large-scale genetic research and improves healthcare. The aim of this review was to analyze the literature on the potential of combining electronic medical records and biobank databases in research and clinical practice.

Keywords: biobanking, electronic medical records, medical information system, genomic medicine, translational research.

Relationships and Activities: none.

Kopylova O. V.* ORCID: 0000-0001-5397-5387, Ershova A. I. ORCID: 0000-0001-7989-0760, Efimova I. A. ORCID: 0000-0002-3081-8415, Blokhina A. V. ORCID: 0000-0002-3019-3961, Limonova A. S. ORCID: 0000-0003-1500-3696, Borisova A. L. ORCID: 0000-0003-4020-6647, Pokrovskaya M. S. ORCID: 0000-0001-6985-7131, Drapkina O. M. ORCID: 0000-0002-4453-8430.

*Corresponding author: sivoksana@yandex.ru

Received: 20/09-2022

Revision Received: 05/10-2022

Accepted: 13/10-2022

For citation: Kopylova O. V., Ershova A. I., Efimova I. A., Blokhina A. V., Limonova A. S., Borisova A. L., Pokrovskaya M. S., Drapkina O. M. Electronic medical records and biobanking. *Cardiovascular Therapy and Prevention*. 2022;21(11):3425. doi:10.15829/1728-8800-2022-3425. EDN OGIGOO

ВНП — вариант нуклеотидной последовательности, ЕГИСЗ — единая государственная информационная система в сфере здравоохранения, СГХС — семейная гиперхолестеринемия, ХС ЛНП — холестерин липопротеинов низкой плотности, ЭИБ — электронная история болезни, ЭПМЗ — электронная персональная медицинская запись, EHR4CR — The Electronic Health Records for Clinical Research (электронные истории болезни для клинических исследований), eMERGE Network — The Electronic Medical Records and Genomics Network (сеть электронных медицинских карт и геномики).

*Автор, ответственный за переписку (Corresponding author):

e-mail: sivoksana@yandex.ru

[Копылова О. В.* — к.м.н., н.с. лаборатории клиномики, ORCID: 0000-0001-5397-5387, Ершова А. И. — д.м.н., руководитель лаборатории клиномики, зам. директора по фундаментальной науке, ORCID: 0000-0001-7989-0760, Ефимова И. А. — ведущий эксперт лаборатории "Банк биологического материала", ORCID: 0000-0002-3081-8415, Блохина А. В. — м.н.с. лаборатории клиномики, ORCID: 0000-0002-3019-3961, Лимонова А. С. — м.н.с. лаборатории клиномики, ORCID: 0000-0003-1500-3696, Борисова А. Л. — ведущий инженер лаборатории "Банк биологического материала", ORCID: 0000-0003-4020-6647, Покровская М. С. — к.б.н., руководитель лаборатории "Банк биологического материала", ORCID: 0000-0001-6985-7131, Драпкина О. М. — д.м.н., профессор, академик РАН, руководитель отдела фундаментальных прикладных аспектов ожирения, директор, ORCID: 0000-0002-4453-8430].

Ключевые моменты

Что известно о предмете исследования?

- Ценность хранимых в биобанке биообразцов во многом определяется наличием сопряженной с ними клинической и других видов информации.
- Электронные истории болезни представляют собой уникальный источник большого количества информации, получаемой на протяжении длительного времени.

Что добавляют результаты исследования?

- Объединение баз данных биобанков и электронных историй болезни открывает новые перспективы в проведении широкомасштабных биомедицинских исследований, развитии и внедрении геномной медицины.
- Использование информации из электронных историй болезни имеет ряд особенностей и требует их учета при планировании исследований и анализе данных.

Key messages

What is already known about the subject?

- The value of biosamples stored in a biobank is largely determined by the presence of related clinical data and other information.
- Electronic medical records are a unique source of a large amount of information obtained over a long time period.

What might this study add?

- Combining databases of biobanks and electronic medical records opens up new opportunities for large-scale biomedical research, development and implementation of genomic medicine.
- The use of data from electronic medical records has a number of features and requires their consideration in research planning and data analysis.

Введение

В последние годы в России и мире происходит стремительное развитие сферы биобанкирования. Под биобанком понимается "организация или подразделение организации, которая может принимать, обрабатывать, хранить и распространять биологические образцы и ассоциированные с ними данные для текущих и будущих исследований, диагностики и терапии в соответствии со стандартными операционными процедурами (СОП) и включает в себя полный комплекс мероприятий, связанных с его функционированием" [1]. Биобанки являются одной из важнейших инфраструктур, необходимых для проведения научных исследований с использованием биообразцов на современном уровне. В частности, они позволяют обеспечить высокий уровень стандартизации во время пробоподготовки, четкое отслеживание всех параметров транспортировки, долгосрочное ответственное хранение биообразцов, сопровождающееся мониторингом температуры и других параметров [2]. В связи с этими особенностями в настоящее время биобанки становятся неотъемлемой частью при планировании и проведении крупномасштабных научных исследований (эпидемиологических, генетических и др.), нередко включающих данные десятков и даже сотен тысяч человек. В России примером реализации такого подхода может служить работа биобанка Национального медицинского исследовательского центра терапии и профилактической медицины Минздрава России [2, 3].

Важно отметить, что определение термина "биобанк" включает в себя две равноправные части — хранение образцов и хранение ассоциированных с ними данных. Действительно, даже самый ценный биообразец, полученный с большим трудом, например, от пациента с очень редким заболеванием, теряет смысл, если данные о нем (диагноз, возраст больного и др.) неизвестны. В последнее время даже появилось название "биодататбанк" ("biodatabank"), которым подчеркивается важность хранения как биообразцов, так и ассоциированных с ними данных¹. Существуют разнообразные виды данных (например, информация непосредственно о биообразце — дата и время забора, время и условия транспортировки, особенности хранения; информация о пациенте — социально-демографическая, клиническая и т.п.; информация, получаемая по результатам научных исследований), а также рекомендованные формы их описания [4]. Сбор клинической информации представляет собой отдельную важную задачу ввиду необходимости привлечения значительных ресурсов (привлечение квалифицированного медицинского персонала, использование специализированного медицинского оборудования и др.). Интеграция информации, собираемой в электронные истории болезни (ЭИБ) в рамках рутинной клинической работы, с базами данных биобанка может значительно облегчить за-

¹ The CENTOGENE Biodatabank. <https://www.centogene.com/about-us/biodatabank>. (21.08.2022).

дачу сбора и хранения клинической и других видов информации, а также предоставить новые возможности для планирования и проведения биомедицинских исследований [5]. Целью настоящего обзора является анализ литературы, посвященной возможностям объединения электронных историй болезни и баз данных биобанков в научной и клинической практике.

Методология

Проведен поиск в электронных научных библиотеках и базах данных за период 2002-2022гг:

1. "elibrary.ru" по словам, словосочетаниям и их комбинациям (в названиях и ключевых словах публикаций): "Электронная история болезни", "Медицинская информационная система", "Биобанкирование";

2. "pubmed.gov" по словам, словосочетаниям и их комбинациям (в названиях и ключевых словах публикаций): "Electronic health record", "medical information system", "Biobanking".

ЭИБ в современной медицинской практике

С появлением и повсеместным распространением компьютерной техники документооборот значительного большинства производственных процессов стал переходить в электронную форму. Данный процесс затронул и медицинские учреждения. При работе с пациентом одними из основных документов являются история болезни (в стационаре) и амбулаторная карта (в амбулаторном звене). Именно в этих документах фиксируются основные клинические данные того или иного пациента: жалобы, анамнез, результаты физикального осмотра, лабораторных и инструментальных исследований, диагноз, назначаемая медикаментозная и немедикаментозная терапия, протоколы оперативных вмешательств и другие данные. Переход к электронному введению медицинских данных при грамотной организации имеет целый ряд преимуществ: позволяет более четко формализовать записи; уменьшает количество рукописной работы и устраняет проблемы, связанные с неразборчивым почерком специалистов; сокращает время введения информации; обеспечивает медперсоналу оперативный и удобный доступ к необходимой информации и уже имеющимся в системе записям, в т.ч. при длительном наблюдении; повышает преемственность оказания медицинской помощи [6-8]. Кроме того, введение электронного документооборота в медицинском учреждении позволяет решить целый ряд организационных задач: объединить различные службы и подразделения медицинского учреждения в единое информационное пространство; эффективно управлять потоками пациентов; четко отслеживать перемещения материально-технических средств (например, лекарственных препаратов и расходных материалов в стационаре); анализировать результаты работы подразделений за необходимые периоды времени и др. [6].

В 2008г вступил в силу Национальный стандарт РФ ГОСТ Р 52636-2006 "Электронная история болезни. Общие положения"². В данном стандарте даны определения ряду терминов:

— электронная история болезни — это "информационная система, предназначенная для ведения, хранения на электронных носителях, поиска и выдачи по информационным запросам (в т.ч. по электронным каналам связи) персональных медицинских записей";

— электронная персональная медицинская запись — "любая персональная медицинская запись, сохраненная на электронном носителе". Понятие "электронная персональная медицинская запись" (ЭПМЗ) соответствует международному термину EHR — Electronic Health Record".

В указанном стандарте подчеркивается важность обеспечения безопасности и конфиденциальности данных при их хранении и перемещении. Особенностью ЭПМЗ является целая система содержащихся в ней идентификаторов, а именно: идентификатор пациента; идентификатор данной ЭПМЗ; идентификатор типа ЭПМЗ; дата и время поступления данных; номер истории болезни или амбулаторной карты; идентификатор лица, создавшего запись и др. Важными моментами являются права и организация доступа к ЭПМЗ².

Наряду с этим в настоящее время широко используется термин "медицинская информационная система" (МИС) — автоматизированная система документооборота для лечебно-профилактических учреждений, которая содержит в цифровой форме специальные категории персональных данных о состоянии здоровья: медицинские карты пациентов, данные медицинских исследований, данные мониторинга состояния пациента с медицинских приборов, средства общения между сотрудниками, финансовую и административную информацию [9].

В последние годы в РФ ведется масштабная работа по совершенствованию информатизации здравоохранения. В 2011г опубликован приказ Министерства здравоохранения и социального развития № 364 "Об утверждении концепции создания единой государственной информационной системы в сфере здравоохранения"³. Под единой государственной информационной системой в сфере здравоохранения (ЕГИСЗ) подразумевается "совокуп-

² Национальный Стандарт РФ ГОСТ Р 52636-2006 "Электронная история болезни. Общие положения" Утвержден Приказом федерального агентства по техническому регулированию и метрологии РФ от 27 декабря 2006г № 407-Ст. <https://dokipedia.ru/document/5324420>. (21.08.2022).

³ Приказ Министерства здравоохранения и социального развития РФ от 28 апреля 2011г № 364 "Об утверждении концепции создания единой государственной информационной системы в сфере здравоохранения." <https://www.garant.ru/products/ipo/prime/doc/4092541/>. (21.08.2022).

ность информационно-технологических и технических средств, обеспечивающих информационную поддержку методического и организационного обеспечения деятельности участников системы здравоохранения". Предпосылками создания данной системы явился целый ряд задач (в области управления здравоохранением, непосредственного оказания медицинской помощи, взаимодействия медицинских служб с населением), решение которых затруднено или невозможно в случае отсутствия комплексного подхода к внедрению и развитию современных информационно-коммуникационных технологий в медицинской сфере. Основной целью создания ЕГИСЗ является обеспечение эффективной информационной поддержки процесса управления системой медицинской помощи, а также процесса оказания медицинской помощи³.

ЭИБ в научной практике

Развитие электронных информационных систем для ведения и хранения персональных медицинских записей имеет большой потенциал и для сферы медицинских научных исследований, поскольку позволяет собирать, хранить и анализировать большие объемы медицинской информации.

ЭИБ представляют собой базы данных, в которые систематически поступает информация о пациенте, получаемая в процессе его обследования и лечения в амбулаторных и стационарных условиях. Использование ЭИБ в качестве источника демографической, клинической и других видов информации имеет несколько преимуществ по сравнению с отдельной регистрацией этих данных в форме отчета о клиническом случае, традиционно используемой в клинических исследованиях: могут быть уменьшены ошибки, возникающие при переносе данных, поскольку в ЭИБ информация вводится медицинским персоналом, непосредственно участвующим в лечении пациента; снижаются стоимость и сроки проведения исследования; кроме того, применение ЭИБ способно облегчить внедрение удаленного мониторинга данных, что потенциально может значительно снизить затраты на проведение клинических испытаний [10].

Одним из наиболее распространенных видов научных работ с использованием ЭИБ являются национальные и другие виды регистров пациентов, которые представляют собой наблюдательные исследования, имеющие заранее оговоренную цель, и в которых используются единообразные методы сбора научных данных [10, 11]. Однако использование ЭИБ для клинических исследований, особенно проспективных, все еще ограничено, несмотря на тот факт, что информация, получаемая в рамках рутинной медицинской помощи, в значительной степени совпадает с данными, необходимыми для проведения исследований. Наиболее простым и общепринятым применением ЭИБ в настоящее

время является оценка критериев включения и не-включения пациентов в научное исследование, что облегчает набор пациентов [10]. Доказано, что использование ЭИБ в клинических испытаниях увеличивает скорость набора пациентов [12-14]. Более сложной задачей является использование информации, собранной в ЭИБ в рамках рутинной клинической помощи, в качестве исходных данных для исследований. В то время как преимущества ЭИБ в непосредственном уходе за пациентами широко признаны, получение преимуществ от использования данных ЭИБ в исследовательских целях по-прежнему встречается редко: недооценивается или упускается из вида [12].

В настоящее время научно-медицинским сообществом активно изучаются и разрабатываются возможности использования данных из ЭИБ при проведении научных исследований, в т.ч. с применением биообразцов. Проект "Электронные истории болезни для клинических исследований" (The Electronic Health Records for Clinical Research, EHR4CR) был создан с целью изучения возможности использования данных из ЭИБ для клинических исследований в Европе. Подобный подход предоставляет уникальные возможности для проведения научных работ, в т.ч. благодаря объединению ресурсов различных исследовательских и клинических центров⁴.

Несмотря на многочисленные преимущества использования ЭИБ в рамках биомедицинских исследований, существует ряд осложняющих работу моментов, которые необходимо учитывать при организации подобных работ. В частности, в большинстве случаев данные в ЭИБ не организованы согласно исследовательским целям [15]. Часто информация хранится в больших реляционных базах данных, которые могут содержать тысячи таблиц, каждая из которых может состоять из множества столбцов. Эта сложная форма предназначена для обеспечения детализации данных, однако затрудняет работу ученых и статистиков с исходными данными. Обычно необходимо сотрудничество с программистом, который имеет глубокое понимание внутреннего устройства базы данных. При определении необходимых для клинического исследования элементов ЭИБ, совместно с клинической исследовательской командой, программист может написать код для извлечения необходимых данных [15]. Анализ различных вариантов ЭИБ, проведенный немецкими исследователями, показал, что ни в одном из них не было специальных функций для организации и ведения научных работ [16].

⁴ Kalra D. Case Report from the EHR4CR Project — A European Survey on Electronic Health Records Systems for Clinical Research. https://www.academia.edu/13410374/Case_Report_from_the_EHR4CR_Project_A_European_Survey_on_Electronic_Health_Records_Systems_for_Clinical_Research. (14.09.2022).

Сложности возникают и при определении имеющихся у пациента диагнозов. Часто инструмент для ввода диагноза представляет собой неструктурированное рукописное поле, которое может быть заполнено произвольным образом. Подобная ситуация приводит к тому, что формулировка одних и тех же диагнозов может орфографически существенно различаться и автоматизированный поиск по тому или иному термину не дает необходимых результатов (выдает не все реально имеющиеся значения). Поиск по части слова (обычно корень) также затруднен, поскольку приводит к большому количеству "шумовых" данных — включает пациентов, имеющих другие диагнозы, но с подобным сочетанием букв. Задача осложняется тем, что обычно в рамках исследований необходимо отобрать пациентов по сочетанию нескольких диагнозов или их отсутствию — в таком случае все неточности поиска возрастают на порядок. Кодирование диагнозов тоже может быть использовано, но в большинстве случаев проставляется лишь код основного заболевания, что также недостаточно для отбора и анализа данных в рамках научных исследований. Вместо этого исследователям приходится использовать так называемые вычисляемые фенотипы, которые представляют собой логические определения, обычно основанные на комбинации диагностических кодов, результатов тестирования и/или лекарств, которые используются для указания на то, есть ли у пациента интересующее заболевание. Например, вычисляемый фенотип сахарного диабета может включать наличие диагностического кода диабета, или прием лекарств для его лечения, или уровень гликированного гемоглобина $>6,5\%$ [15, 17]. Одним из часто используемых способов решения проблемы поиска диагнозов является использование общих моделей данных (common data models, CDM). Общие модели данных представляют собой набор "правил", указывающих, какие элементы данных необходимо извлечь из медицинской карты и как с ними следует обращаться [15].

Одним из преимуществ использования данных ЭИБ является возможность их анализа на протяжении длительного времени. Однако это также означает, что за один и тот же период одни и те же данные (например, результаты лабораторных анализов) у некоторых пациентов могут быть представлены многократно, у некоторых в небольшом количестве, а у некоторых — полностью отсутствовать. При этом обычно существует корреляция между состоянием здоровья пациента и количеством имеющихся данных: например, для лабораторных показателей у пациентов с одним диагнозом типично проведение нескольких определений, а с другим — ни одного. Таким образом, пропуск данных может ограничивать использование ряда статистических методов, что должно быть обязательно учтено при анализе [15, 18].

Сложностью при использовании ЭИБ в рамках биомедицинских исследований может являться полное отсутствие ряда необходимых данных. В большинстве ЭИБ мало внимания уделяется социально-бытовым показателям: часто отсутствует информация о социально-экономическом статусе, уровне употребления алкоголя, особенностях питания и др. В зависимости от исследования может возникнуть необходимость собирать эти элементы данных отдельно или фиксировать их с помощью косвенных переменных. Например, косвенный показатель индивидуального социально-экономического статуса может быть получен на основе знания адреса человека и характеристики социально-экономического статуса его района [19]. Более того, иногда может понадобиться связать данные ЭИБ с внешними данными, такими как информация об окружающей среде или даже геолокационные данные [20].

Еще одной особенностью работы с ЭИБ в рамках биомедицинских исследований является то, что имеющиеся данные обладают свойствами как длительных (лонгитудинальных) наблюдений, так и одномоментных (поперечных) срезов. В частности, глубина наблюдения нередко ограничена той датой, когда был осуществлен переход медицинской организации на данную конкретную ЭИБ. Кроме того, разные пациенты, даже с одинаковым диагнозом имеют различную длительность наблюдения и разное количество клинических данных, они могут появляться у врача, а затем надолго исчезать, при этом неизвестно, отсутствие повторных визитов связано с хорошим состоянием здоровья или, например, с тем, что человек проживает в отдаленной местности. Для того чтобы нивелировать подобные ограничения, рекомендуется очень четкое формулирование критериев включения и невключения в исследование с учетом не только клинических характеристик, но и наличия данных в ЭИБ. Также важно учитывать, что использование дизайна исследований "случай-контроль" с применением данных ЭИБ может быть ограничено в связи с разной длительностью наблюдения и разным количеством данных у разных пациентов — неверный отбор пациентов в группу контроля может приводить к системным ошибкам [15, 21].

Важной особенностью при планировании исследований с использованием информации из ЭИБ является учет ряда данных, не всегда являющихся клиническими, но способными оказать влияние на результат. Очевидно, что за медицинской помощью, в основном, обращаются люди, страдающие тем или иным заболеванием, в связи с чем в ЭИБ накапливается гораздо больше информации о популяциях больных людей, но не о здоровой части населения, что необходимо иметь ввиду при разработке дизайна исследования [22]. Вид ме-

дицинской помощи, оказываемой пациенту, и вид медицинского учреждения также могут иметь значение: например, данные одного и того же типа от одного и того же человека, получаемые в отделении реанимации и в условиях амбулаторного поликлинического приема, могут значительно различаться [23]. Существенно могут различаться пациенты из различных местностей. В частности, те, кто проживает в районе локализации медицинского центра, чаще будут обращаться за медицинской помощью в плановом порядке, тогда как пациенты, чье постоянное место проживания находится в другом регионе (например, в данный регион они приехали на отдых), более склонны обращаться за медицинской помощью в связи с серьезными и острыми заболеваниями [23]. Если пациент обращается за медицинской помощью только в период обострения заболевания, его анализы и другие данные в ЭИБ могут отличаться от данных в его обычном стабильном состоянии, что также может вызвать ошибки при анализе результатов исследования [24].

В процессе реализации европейской инициативы EHR4CR исследователи выявили ряд дополнительных сложностей, возникающих при необходимости объединения данных ЭИБ из разных медицинских центров; основными из них являются защита конфиденциальности, вопросы совместимости систем, приемлемость для пациентов и этические моменты⁵.

Таким образом, при организации исследований с использованием ЭИБ необходимо учитывать различные факторы формирования данных в ЭИБ для предотвращения неверных и невоспроизводимых результатов.

Место ЭИБ в биобанкировании

Объединение данных ЭИБ и информации, получаемой о биообразцах, хранящихся в биобанках, весьма перспективно как в научном, так и в клиническом плане [5]. Поскольку, как уже было сказано, биобанки по факту являются биодататанками, ценность хранимых биообразцов во многом определяется наличием сопряженной с ними клинической и других видов информации, а стандартизация видов хранения информации в рамках ЕГИСЗ позволит интегрировать усилия множества биобанков одновременно.

Примером объединения баз данных биобанков и ЭИБ является проект "eMERGE Network" (The Electronic Medical Records and Genomics Network, Сеть электронных медицинских карт и геномики), который был основан в США в 2007г. Основными целями проекта являются объединение информации из биорепопозитория, являющихся источниками дезоксирибонуклеиновой кислоты (ДНК), с системой ЭИБ для крупномасштабных генетических исследований высокопроизводительными методами, создание шкал генетического риска с целью опти-

мизации стратегий управления рисками распространенных заболеваний (сердечно-сосудистых, неврологических, эндокринологических и др.) и снижения заболеваемости и смертности от этих заболеваний^{5,6}. Сеть объединяет исследователей с широким спектром специализации и большим опытом экспертизы в области геномики, статистики, этики, информатики и клинической медицины из ведущих медицинских исследовательских институтов для проведения генетических исследований с использованием клинических данных из электронных медицинских записей, в т.ч. с целью более широкого использования геномной медицины. В рамках данного проекта продемонстрировано, что клинические данные, полученные из ЭИБ, могут быть успешно использованы в сочетании с генетическими данными для комплексного изучения предрасположенности к заболеваниям среди различных групп пациентов^{5,6} [25]. При этом, с одной стороны, ЭИБ используются как источник информации для изучения ассоциаций фенотипа и генотипа, а с другой — изучаются особенности и эффекты интеграции получаемой генетической информации с ЭИБ, возможности ее использования для совершенствования системы оказания помощи пациентам. Для достижения целей проекта проведено генотипирование >100 тыс. участников^{5,6} [25].

В 2015-2020гг были созданы 7 рабочих групп, фокусирующихся на решении различных задач: 1) клиническое аннотирование — создание согласованных подходов к клинической интерпретации генов и вариантов нуклеотидной последовательности (ВНП); 2) интеграция в ЭИБ — интеграция медицинских и генетических результатов в ЭИБ в клинических центрах; 3) геномика — изучение передового опыта и совершенствование анализа данных о распространенных и редких генетических вариантах; 4) фармакогеномика — координация и продвижение работы по изучению и внедрению в клиническую практику фармакогеномики; 5) фенотипирование — развитие научно-обоснованных методов фенотипирования с использованием данных из ЭИБ, в т.ч. выделение выборок для исследований типа "случай-контроль" и др.; 6) оценка влияния интеграции генетических данных в ЭИБ на различные аспекты оказания медицинской помощи; 7) изучение этических, юридических и социальных последствий интеграции в ЭИБ генетических данных, получаемых в рамках проекта [25].

В рамках проекта eMERGE Denny JC, et al. (2010) был разработан метод "PheWAS" (phenome-wide association scan), который подразумевает поиск ассоциаций между фенотипическими и генотипи-

⁵ <https://www.genome.gov/Funded-Programs-Projects/Electronic-Medical-Records-and-Genomics-Network-eMERGE>. (14.09.2022).

⁶ <https://emerge-network.org/>. (14.09.2022).

ческими данными с использованием ЭИБ [26]. На основе данного метода в 2013г был создан каталог PheWAS, процитированный к 2020г >300 раз [25]. В настоящее время также разрабатываются возможности использования фенотипических шкал риска (phenotype risk score, PheRS) [5]. Отдельным направлением явилось создание методов изучения фармакогенетических взаимодействий в рамках сети eMERGE, что привело к разработке веб-портала SPHINX (Sequence and Phenotype Integration Exchange), нацеленного на изучение влияния генетических факторов на ответ пациента на медикаментозную терапию [25].

В рамках сети eMERGE разработаны технические и нормативные решения для интеграции геномной информации в ЭИБ, а также изучаются подходы к обучению врачей и информированию пациентов при включении геномных данных в процесс диагностики и лечения^{5,6}. Данная работа позволила приступить к формированию и реализации стратегии для переноса, хранения и предоставления геномных данных в ЭИБ с целью обеспечения поддержки принятия клинических решений. Изучается воздействие геномного секвенирования на систему здравоохранения, пациентов и их семей. Получаемые результаты демонстрируют ценность полученных при длительном наблюдении клинических данных, зафиксированных в ЭИБ, в плане выявления новых генетических вариантов, ассоциированных с заболеваниями [25].

Учитывая, что сеть eMERGE включает большое количество разноплановых клинических и научных организаций, большое внимание уделяется конфиденциальности и безопасности данных, которые имеют решающее значение в сферах геномных исследований и медицинской помощи. При зарождении данного проекта были установлены специальные протоколы для сбора и обмена деидентифицированными данными в сети, чтобы облегчить исследования и обеспечить равный доступ всем организациям. Эти рекомендации стали важнейшим элементом успеха всех последующих исследований в рамках сети eMERGE [25].

В настоящее время проект продолжается и сейчас сфокусирован на разработке и внедрении полигенных шкал риска, интеграции их в ЭИБ и изучении информативности их применения, в т.ч. с учетом возраста, образа жизни, лекарственной терапии, этнических особенностей^{5,6} [25].

Glazer AM, et al. (2022) провели многоцентровое исследование на когорте из 21846 участников, не имевших ранее показаний для генетического кардиологического скрининга⁵. Было секвенировано 109 генов, ассоциированных с моногенными наследственными заболеваниями, в т.ч. 10 генов, связанных с развитием аритмий. Фенотипы носителей ВНП оценивали с помощью данных, имеющихся

в ЭИБ, и последующего клинического обследования. У 3% участников были выявлены патогенные или вероятно патогенные ВНП в каких-либо из 109 исследуемых генов. Из них у 120 (0,6%) участников были обнаружены варианты, связанные с развитием наследственных жизнеугрожающих аритмий. В рамках работы была возможность сообщить результат 54 участникам, при этом важно отметить, что у 12 из них диагноз наследственной аритмии был поставлен впервые, только после получения результатов генетического исследования и размещения полученных данных в ЭИБ (0,05%). Постановка диагноза таким пациентам позволяет вовремя начать лечение и проводить профилактику внезапной сердечной смерти. Исследование наглядно продемонстрировало значение интеграции биобанкирования, генетического скрининга и ЭИБ в научной и клинической практике⁵.

В исследовании с использованием генетических результатов анализа биообразцов, хранящихся в биобанке, и данных ЭИБ Palmer MR, et al. (2021) выявили гены, ассоциированные с развитием каротидного стеноза⁶; Rosenthal EA, et al. (2021) — с гипертриглицеридемией [27]; Rouby NE, et al. (2018) — с резистентной гипертонией [28]. Zhang X, et al. (2019), объединяя данные ЭИБ и генетических исследований, разработали новый подход к выявлению потенциального плеiotропного влияния генов на развитие сердечно-сосудистых и неврологических заболеваний [29]. В работе Holzinger ER, et al. (2017), синтез биобанкирования и анализа ЭИБ позволил получить новые данные о взаимодействиях между различными ВНП и липидным профилем [30].

Abul-Husn NS, et al. (2016) на основе анализа данных длительного наблюдения, имеющихся в ЭИБ 50726 человек и результатов генетического тестирования, выявили, что вероятность возникновения и преждевременного развития ишемической болезни сердца у носителей патогенных ВНП, ассоциированных с семейной гиперхолестеринемией (СГХС) выше, соответственно, в 2,6 и 3,7 раза; а в случае наличия ВНП в гене *LDLR*, вызывающих потерю функции белка, — в 5,5 и 10,3 раза, соответственно. Интересным оказался тот факт, что ВНП, связанные с СГХС, были обнаружены лишь у 2,5% пациентов с выраженным повышением уровня холестерина липопротеинов низкой плотности (ХС ЛНП). В то же время, у 45% пациентов, имеющих патогенные ВНП, уровень ХС ЛНП был ниже диагностического уровня (190 мг/дл). Расчет вероятности СГХС, согласно голландским критериям (по данным ЭИБ), указал на наличие вероятного или определенного диагноза лишь у 24% носителей патогенных ВНП, что свидетельствует об ограничениях скрининга на СГХС с помощью лишь анализа данных ЭИБ. К моменту анализа данных 14 носи-

телей патогенных ВМП умерли, при этом ни у одного из них не был клинически установлен диагноз СГХС [31]. Данное исследование имеет практическое значение в отношении организации диагностики и лечения СГХС.

Получаемые в результате подобных исследований данные представляют большую ценность и отдельным направлением исследований является изучение интеграции генетических данных в ЭИБ, чему, в частности, посвящен проект IGNITE (Implementing Genomics in Practice, внедрение геномики в практику) [32]. В рамках данного проекта выявлен ряд моментов, которые важно учитывать при таком внедрении: 1) техническая интеграция геномных данных в ЭИБ пациентов; 2) укрепление знаний и представлений клиницистов о геномной медицине; 3) вовлечение пациентов в проекты геномной медицины.

Заключение

Таким образом, объединение данных ЭИБ с информацией, получаемой при анализе биообразцов, хранящихся в биобанках, предоставляет уникальную возможность для проведения целого ряда научных исследований, в т.ч. широкомасштабных генетических и клинических проектов. Использование ЭИБ может быть полезно при определении дизайна исследований, поиска и набора пациен-

тов, формировании высоко репрезентативных выборок, определении конечных точек исследований, а также в качестве непосредственного источника клинических, демографических и других видов данных. При этом существует целый ряд особенностей, которые необходимо учитывать при организации исследований, объединяющих информацию из ЭИБ и биобанков. Во многом они связаны с тем, что методология поступления и хранения данных в ЭИБ отличается от таковой при проведении научных исследований, что следует учитывать как при создании алгоритмов работы с ЭИБ, так и при планировании и реализации научных проектов с использованием ЭИБ. Разработка унифицированных подходов к ведению электронной медицинской документации в рамках проекта ЕГИСЗ и подобных национальных инициатив имеет большой потенциал не только в плане повышения преемственности оказания медицинской помощи, но и в плане объединения ресурсов различных научных центров, и в перспективе может внести значительный вклад в развитие широко-масштабных биомедицинских, в т.ч. генетических исследований.

Отношения и деятельность: все авторы заявляют об отсутствии потенциального конфликта интересов, требующего раскрытия в данной статье.

Литература/References

- Mikhailova AA, Nasykhova YA, Muravyov AI, et al. Towards the creation of a unified glossary of Russian biobanks. *Cardiovascular Therapy and Prevention*. 2020;19(6):2710. (In Russ.) Михайлова А.А., Насыхова Ю.А., Муравьев А.И. и др. На пути к созданию общего глоссария биобанков Российской Федерации. *Кардиоваскулярная терапия и профилактика*. 2020;19(6):2710. doi:10.15829/1728-8800-2020-2710.
- Kopylova OV, Ershova AI, Pokrovskaya MS, et al. Population-nosological research biobank of the National Medical Research Center for Therapy and Preventive Medicine: analysis of biosamples, principles of collecting and storing information. *Cardiovascular Therapy and Prevention*. 2021;20(8):3119. (In Russ.) Копылова О.В., Ершова А.И., Покровская М.С. и др. Популяционно-нозологический исследовательский биобанк "НМИЦ ТПМ": анализ коллекций биообразцов, принципы сбора и хранения информации. *Кардиоваскулярная терапия и профилактика*. 2021;20(8):3119. doi:10.15829/1728-8800-2021-3119.
- Pokrovskaya MS, Borisova AL, Metelskaya VA, et al. Role of biobanking in managing large-scale epidemiological studies. *Cardiovascular Therapy and Prevention*. 2021;20(5):2958. (In Russ.) Покровская М.С., Борисова А.Л., Метельская В.А. и др. Роль биобанкирования в организации крупномасштабных эпидемиологических исследований. *Кардиоваскулярная терапия и профилактика*. 2021;20(5):2958. doi:10.15829/1728-8800-2021-2958.
- Sivakova OV, Pokrovskaya MS, Metelskaya VA, et al. International rules for description of biospecimens are an important factor in improving the quality of researches. *Profilakticheskaya meditsina*. 2019;22(6):95-9. (In Russ.) Сивакова О.В., Покровская М.С., Метельская В.А. и др. Международные правила описания биообразцов — важный фактор повышения качества научных исследований. *Профилактическая медицина*. 2019;22(6):95-9. doi:10.17116/profmed20192206295.
- Linder JE, Bastarache L, Hughey JJ, et al. The Role of Electronic Health Records in Advancing Genomic Medicine. *Annu Rev Genomics Hum Genet*. 2021;22:219-38. doi:10.1146/annurev-genom-121120-125204.
- Grushevskaya VV. Electronic medical history is the most important link in medical information systems. Accounting, analysis and audit problems of theory and practice. 2012;(9):37-8. (In Russ.) Грушевская В.В. Электронная история болезни — важнейшее звено медицинских информационных систем. Учет, анализ и аудит проблемы теории и практики. 2012;(9):37-8.
- Guliyev YaI, Komarov SI. Medical integrated information systems: theory and practice. *Physician and information technology*. 2005;(3):29-32. (In Russ.) Гулиев Я.И., Комаров С.И. Медицинские интегрированные информационные системы: теория и практика. *Врач и информационные технологии*. 2005;(3):29-32.
- Rinner C, Sauter SK, Endel G, et al. Improving the informational continuity of care in diabetes mellitus treatment with a nationwide Shared EHR system: Estimates from Austrian claims data. *Int J Med Inform*. 2016;92:44-53. doi:10.1016/J.IJMEDINF.2016.05.001.
- Dorofeev YaA, Abramova EA. Ensuring the protection of personal data in medical information systems. *Collection of scientific*

- papers of Russian universities "Problems of economics, finance and production management." 2020;(46):36-41. (In Russ.) Дорофеев Я.А., Абрамова Е.А. Обеспечение защиты персональных данных в медицинских информационных системах. Сборник научных трудов вузов России "Проблемы экономики, финансов и управления производством." 2020;(46):36-41.
10. Cowie MR, Blomster JL, Curtis LH, et al. Electronic health records to facilitate clinical research. *Clin Res Cardiol.* 2017;106:1-9. doi:10.1007/s00392-016-1025-6.
11. Martsevich SY, Kutishenko NP, Lukina YV, et al. Observational studies and registers. Their quality and role in modern evidence-based medicine. *Cardiovascular Therapy and Prevention.* 2021;20(2):2786. (In Russ.) Марцевич С.Ю., Кутишенко Н.П., Лукина Ю.В. и др. Наблюдательные исследования и регистры. Их качество и роль в современной доказательной медицине. Кардиоваскулярная терапия и профилактика. 2021;20(2):2786. doi:10.15829/1728-8800-2021-2786.
12. De Moor G, Sundgren M, Kalra D, et al. Using electronic health records for clinical research: The case of the EHR4CR project. *J Biomed Inform.* 2015;53:162-73. doi:10.1016/J.JBI.2014.10.006
13. Pisinger C, Døssing M. A systematic review of health effects of electronic cigarettes. *Prev Med (Baltim).* 2014;69:248-60. doi:10.1016/j.ypmed.2014.10.009.
14. Lai YS, Afseth JD. A review of the impact of utilising electronic medical records for clinical research recruitment. *Clin Trials.* 2019;16(2):194-203. doi:10.1177/1740774519829709.
15. Goldstein BA. Five analytic challenges in working with electronic health records data to support clinical trials with some solutions. *Clin Trials.* 2020;17(4):370-6. doi:10.1177/1740774520931211.
16. Schreiweis B, Trinczek B, Köpcke F, et al. Comparison of electronic health record system functionalities to support the patient recruitment process in clinical trials. *Int J Med Inform.* 2014;83(11):860-8. doi:10.1016/J.IJMEDINF.2014.08.005.
17. Richesson RL, Rusincovitch SA, Wixted D, et al. A comparison of phenotype definitions for diabetes mellitus. *J Am Med Inform Assoc.* 2013;20(e2). doi:10.1136/AMIAJNL-2013-001952.
18. Wells BJ, Nowacki AS, Chagin K, et al. Strategies for handling missing data in electronic health record derived data. *EGEMS (Washington, DC).* 2013;1(3):7. doi:10.13063/2327-9214.1035.
19. Bhavsar NA, Gao A, Phelan M, et al. Value of Neighborhood Socioeconomic Status in Predicting Risk of Outcomes in Studies That Use Electronic Health Record Data. *JAMA Netw open.* 2018;1(5):e182716. doi:10.1001/JAMANETWORKOPEN.2018.2716.
20. Xie S, Greenblatt R, Levy MZ, et al. Enhancing Electronic Health Record Data with Geospatial Information. *AMIA Jt Summits Transl Sci proceedings AMIA Jt Summits Transl Sci.* 2017;2017:123-32.
21. Schuemie MJ, Ryan PB, Man KKC, et al. A plea to stop using the case-control design in retrospective database studies. *Stat Med.* 2019;38(22):4199-208. doi:10.1002/SIM.8215.
22. Weiskopf NG, Rusanov A, Weng C. Sick patients have more data: the non-random completeness of electronic health records. *AMIA. Annu Symp proceedings AMIA Symp.* 2013;2013:1472-7.
23. Phelan M, Bhavsar NA, Goldstein BA. Illustrating Informed Presence Bias in Electronic Health Records Data: How Patient Interactions with a Health System Can Impact Inference. *EGEMS (Washington, DC).* 2017;5(1):22. doi:10.5334/egems.243.
24. Goldstein BA, Phelan M, Pagidipati NJ, et al. How and when informative visit processes can bias inference when using electronic health records data for clinical research. *J Am Med Inform Assoc.* 2019;26(12):1609-17. doi:10.1093/JAMIA/OCZ148.
25. eMERGE Consortium. Lessons learned from the eMERGE Network: balancing genomics in discovery and practice. *Hum Genet Genomics Adv.* 2021;2(1):100018. doi:10.1016/J.XHGG.2020.100018.
26. Denny JC, Ritchie MD, Basford MA, et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics.* 2010;26(9):1205-10. doi:10.1093/BIOINFORMATICS/BTQ126.
27. Rosenthal EA, Crosslin DR, Gordon AS, et al. Association between triglycerides, known risk SNVs and conserved rare variation in SLC25A40 in a multi-ancestry cohort. *BMC Med Genomics.* 2021;14(1). doi:10.1186/S12920-020-00854-2.
28. El Rouby N, McDonough CW, Gong Y, et al. Genome-wide association analysis of common genetic variants of resistant hypertension. *Pharmacogenomics J.* 2019;19(3):295-304. doi:10.1038/S41397-018-0049-X.
29. Zhang X, Veturi Y, Verma S, et al. Detecting potential pleiotropy across cardiovascular and neurological diseases using univariate, bivariate, and multivariate methods on 43,870 individuals from the eMERGE network. *Pacific Symp Biocomput.* 2019;24(2019):272-83. doi:10.1142/9789813279827_0025.
30. Holzinger ER, Verma SS, Moore CB, et al. Discovery and replication of SNP-SNP interactions for quantitative lipid traits in over 60,000 individuals. *BioData Min.* 2017;10(1). doi:10.1186/S13040-017-0145-5.
31. Abul-Husn NS, Manickam K, Jones LK, et al. Genetic identification of familial hypercholesterolemia within a single U.S. health care system. *Science.* 2016;354(6319). doi:10.1126/SCIENCE.AAF7000.
32. Sperber NR, Carpenter JS, Cavallari LH, et al. Challenges and strategies for implementing genomic services in diverse settings: experiences from the Implementing GeNomics In pracTice (IGNITE) network. *BMC Med Genomics.* 2017;10(1). doi:10.1186/S12920-017-0273-2.