

Особенности формирования клинической аннотации биообразцов

Копылова О. В., Ершова А. И., Борисова А. Л., Метельская В. А., Драпкина О. М.

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

Одной из технологий, которая способствует преодолению проблемы низкой воспроизводимости научных исследований, является биобанкирование, которое подразумевает соблюдение строгих стандартов качества на всех этапах работы с биообразцами. Помимо информации о биообразце (подробного документирования всех особенностей его получения, транспортировки, пробоподготовки и последующего хранения), одним из ключевых моментов является наличие ассоциированной с биообразцом информации о доноре (пациенте). Целью настоящей статьи был анализ подходов к формированию клинической аннотации биообразцов, объединению данных из различных биобанков и оценка возможностей использования для этого электронных медицинских карт и других современных технологий. Поиск публикаций проводился в базах данных PUBMED, eLIBRARY.RU, РИНЦ. Одним из подходов к созданию клинической аннотации является целенаправленный сбор информации специально обученным сотрудником — первичная информация чаще всего берется из индивидуальных карт участника исследования, которые разрабатываются и утверждаются при планировании научной работы. Альтернативным методом является использование электронных историй болезни и других документов, в которые естественным образом собирается информация в процессе обследования и лечения пациентов. Существуют также смешанные формы сбора клинической информации — ярким примером является биобанк Великобритании. Полнота, структурированность и стандартизация являются важнейшими характеристиками клинической аннотации, ассоциированной с биообразцами. В настоя-

щее время разрабатываются различные стандарты, позволяющие унифицировать представление клинической аннотации, сделать биобанки и коллекции более "видимыми" для сторонних исследователей и организаций, что необходимо для сотрудничества и более эффективного и интенсивного использования хранящихся биообразцов. Гармонизация методологии клинического аннотирования между различными биобанками открывает широкие горизонты для проведения крупномасштабных исследований в сфере персонализированной и трансляционной медицины.

Ключевые слова: клиническая аннотация, биообразцы, биобанк, биобанкирование, электронные истории болезни, электронные медицинские карты.

Отношения и деятельность: нет.

Поступила 13/11-2023

Рецензия получена 21/11-2023

Принята к публикации 21/11-2023



Для цитирования: Копылова О. В., Ершова А. И., Борисова А. Л., Метельская В. А., Драпкина О. М. Особенности формирования клинической аннотации биообразцов. *Кардиоваскулярная терапия и профилактика*. 2023;22(11):3855. doi:10.15829/1728-8800-2023-3855. EDN RKVUZC

Specifics of creating clinical abstract of biospecimens

Kopylova O. V., Ershova A. I., Borisova A. L., Metelskaya V. A., Drapkina O. M.
National Medical Research Center for Therapy and Preventive Medicine. Moscow, Russia

One technology that helps overcome the problem of low research reproducibility is biobanking, which involves maintaining strict quality standards at all stages. In addition to data on the biosample (detailed documentation on sampling, transportation, preparation and subsequent storage), one of the key points is the availability of information about the donor (patient). The aim of this article was to analyze creating clinical abstract of biospecimens, combining data from various biobanks and assessing the possibilities of electronic medical records and other modern technologies for this. The search for publications was carried out in the PUBMED, eLIBRARY.RU, RSCI databases. One approach to creating a clinical description is the targeted collection of information by a specially trained employee.

Primary information is most often taken from the individual records of the study participant, which are developed and approved when planning work. An alternative method is the use of electronic medical records and other documents that collect information during the assessment and treatment of patients. There are also mixed types of clinical data collection, a prime example of which is the UK Biobank. Completeness, structure, and standardization are essential characteristics of clinical description associated with biospecimens. Various standards are currently being developed to unify clinical description, making biobanks and collections more available to external researchers and organizations, which is necessary for collaboration and more efficient use of stored biospecimens. Harmonization of clinical description

*Автор, ответственный за переписку (Corresponding author):

e-mail: sivoksana@yandex.ru

[Копылова О. В.* — к.м.н., с.н.с. лаборатории клиномики, ORCID: 0000-0001-5397-5387, Ершова А. И. — д.м.н., зам. директора по фундаментальной науке, руководитель лаборатории клиномики, ORCID: 0000-0001-7989-0760, Борисова А. Л. — ведущий инженер лаборатории "Банк биологического материала", ORCID: 0000-0003-4020-6647, Метельская В. А. — д.б.н., профессор, г.н.с. лаборатории биохимических маркеров риска хронических неинфекционных заболеваний, ORCID: 0000-0001-8665-9129, Драпкина О. М. — д.м.н., профессор, академик РАН, директор, ORCID: 0000-0002-4453-8430].

methodology between different biobanks open up broad boundaries for large-scale research within personalized and translational medicine.

Keywords: clinical abstract, biospecimens, biobank, biobanking, electronic medical records.

Relationships and Activities: none.

Kopylova O.V.* ORCID: 0000-0001-5397-5387, Ershova A.I. ORCID: 0000-0001-7989-0760, Borisova A.L. ORCID: 0000-0003-4020-6647, Metelskaya V.A. ORCID: 0000-0001-8665-9129, Drapkina O.M. ORCID: 0000-0002-4453-8430.

*Corresponding author:
sivoksana@yandex.ru

Received: 13/11-2023

Revision Received: 21/11-2023

Accepted: 21/11-2023

For citation: Kopylova O.V., Ershova A.I., Borisova A.L., Metelskaya V.A., Drapkina O.M. Specifics of creating clinical abstract of biospecimens. *Cardiovascular Therapy and Prevention*. 2023;22(11): 3855. doi:10.15829/1728-8800-2023-3855. EDN RKVUZC

СГХС — семейная гиперхолестеринемия, МИС — медицинская информационная система, МКБ — международная классификация болезней, ЭМК — электронные медицинские амбулаторные карты, ЭИБ — электронные истории болезни, ВСБА — Brain Cancer Biobanking Australia (Биобанкирование материала пациентов с онкологическими заболеваниями головного мозга), BRISQ — Biospecimen Reporting for Improved Study Quality (Правила описания биообразцов для улучшения качества научных исследований), eMERGE Network — The Electronic Medical Records and Genomics Network (Сеть электронных медицинских карт и геномики), FHIR — Fast Health Interoperability Resources (Ресурсы быстрого взаимодействия в сфере здравоохранения), FSN — fully specified name (полностью определенное имя), GWAS — Genome-wide association studies (полногеномные исследования ассоциаций), LOINC — Logical Observation Identifiers Names and Codes (Имена и коды идентификаторов логических наблюдений), MIABIS — The Minimum Information About Biobank data Sharing (Минимальная информация об обмене данными биобанка), PheWAS — Phenome-wide association studies (полногеномные исследования ассоциаций), SNOMED-CT — Systemized Nomenclature of Medicine-Clinical Terms (Систематизированная номенклатура в медицине — клинические термины).

Ключевые моменты

Что известно о предмете исследования?

- Ценность биообразцов, хранящихся в биобанке, возрастает при увеличении объема качественных стандартизированных данных, ассоциированных с биообразцами.
- Обширная и качественно собранная клиническая аннотация биообразцов является одним из ключевых моментов при проведении исследований, направленных на поиск новых методов профилактики, диагностики и лечения различных заболеваний.

Что добавляют результаты исследования?

- Существуют различные методы формирования клинической аннотации биообразцов.
- Использование стандартов при создании клинической аннотации биообразцов открывает возможности для сотрудничества между биобанками на национальном и международном уровне с целью проведения крупномасштабных исследований в сфере трансляционной и персонализированной медицины.

Key messages

What is already known about the subject?

- The value of biospecimens stored in a biobank increases with the volume of quality standardized biosample-related data.
- Extensive and high-quality collected clinical description of biospecimens is one of the key points in research aimed at finding novel methods for the prevention, diagnosis and treatment of various diseases.

What might this study add?

- There are various methods for creating clinical data on biospecimens.
- The use of standards in the clinical description of biospecimens opens up opportunities for collaboration between biobanks to conduct large-scale translational and personalized medicine research.

Введение

Количество биомедицинских исследований, проводимых с использованием биообразцов, с каждым годом возрастает [1]. При этом одной из важных проблем является недостаточная воспроизводимость результатов подобных исследований. Так, в 2011г исследователям из компании Amgen, несмотря на полное соблюдение описанной в оригинальных статьях методики, удалось воспроизвести только 6 из 53 (т.е. всего 11%) крупных исследований, результаты которых были опубликованы в высокорейтинговых журналах, что чрезвычайно мало [2]. Это и подобные ему исследования способствовали интенсификации поиска путей повышения воспроиз-

водимости научно-медицинских исследований с использованием биообразцов.

Одной из технологий, которая способствует преодолению проблемы низкой воспроизводимости, является технология биобанкирования. Согласно российскому глоссарию биобанков Российской Федерации, 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 и включает в себя полный комплекс мероприятий, связанных



Рис. 1 Ключевая роль наличия обширной и качественной клинической аннотации биообразцов в разработке новых профилактических, диагностических и терапевтических подходов.

с его функционированием" [3]. Биобанкирование подразумевает соблюдение стандартов качества на всех этапах работы с биообразцами, а также подробное их описание [4].

Помимо информации о биообразце (подробного документирования всех особенностей преаналитического этапа, включая забор биообразца, транспортировку, пробоподготовку, последующее хранение и извлечение), одним из ключевых моментов является наличие ассоциированной с биообразцом информации о доноре (пациенте) [5]. Подобная информация может включать социально-демографические, антропометрические данные, результаты физикального осмотра, лабораторных и инструментальных исследований, различных опросников и т.п. При прочих равных условиях ценность биообразца возрастает по мере увеличения количества ассоциированных с ним данных, поскольку это позволяет использовать его в большем количестве разнообразных исследований, в т.ч. в сфере трансляционной и персонализированной медицины [5]. Обширная и качественно собранная клиническая аннотация биообразцов является одной из ключевых составляющих при проведении исследований, направленных на поиск новых методов профилактики, диагностики и лечения различных заболеваний (рисунок 1).

К примеру, накопление большого количества качественных данных позволяет проводить масштабные исследования с целью поиска генетической основы различных заболеваний, в т.ч. полногеномные исследования ассоциаций (GWAS — Genome-wide association studies), в которых изучается связь большого количества вариантов нуклеотидной последовательности с определенным состоянием или признаком, также известным как фенотип [6, 7]. В последнее время появилось новое направление — полнофеномные исследования ассоциаций (PheWAS — Phenome-wide association studies), в котором идея GWAS зеркально отражена — в таких исследованиях проводится поиск фе-

нотипов, связанных с конкретным вариантом нуклеотидной последовательности, в диапазоне тысяч фенотипов человека, или "фенома". Ключевым для PheWAS является значительное количество данных, включающих широкий спектр фенотипов, выявленных у большого числа пациентов [7, 8].

Наличие больших объемов биообразцов с обширной качественной клинической аннотацией, включающей не только стандартные клинические данные, но и, например, результаты визуализирующих, генетических исследований открывает возможности для получения новых научных знаний с применением технологий искусственного интеллекта [9, 10].

Настоящая статья представляет собой обзор литературы по вопросам особенностей формирования клинической аннотации биообразцов, возможностям использования с этой целью электронных медицинских карт (ЭМК) и других современных подходов.

Методологический подход

Поиск публикаций проводился в базах данных PUBMED, eLIBRARY.RU, РИНЦ с использованием ключевых слов: клиническая аннотация биообразцов (biospecimen clinical annotation), биобанк и клиническая аннотация (biobank and clinical annotation), электронные медицинские карты (electronic medical records), электронные истории болезни (electronic health records), технологии клинического аннотирования (clinical annotation technologies), биобанк и индивидуальная регистрационная карта (biobank and case report form), стандарты клинического аннотирования (clinical annotation standards), обмен информации между биобанками (exchange of information between biobanks), интеграция информации коллекций биообразцов (integration of information across sample collections). Из 173 отобранных источников в анализ вошло 40 публикаций, включая литературные обзоры, оригинальные статьи, метаанализы. Глубина поиска составила 10 лет: 2013–2023 гг. В обзоре так-

же представлены источники, опубликованные ранее 2013г, если в них содержится ценная информация, касающаяся данной темы.

Результаты и обсуждение

Подходы к созданию клинической аннотации биообразцов, используемых для научных исследований

Одним из классических подходов к созданию клинической аннотации биообразцов, используемых для научных исследований, является целенаправленный сбор клинической и других видов информации специально обученным сотрудником — первичная информация чаще всего берется из индивидуальных регистрационных карт участника исследования [11]. Такие карты разрабатываются и утверждаются при планировании исследования. Сбор данных может осуществляться как на бумажный носитель (анкеты, опросники, различные виды медицинских карт), так и в электронном виде (чаще всего с использованием специально созданных форм сбора информации, в т.ч. в случае многоцентрового исследования — с использованием сети Интернет и внесением данных в деперсонализированном виде) [11]. При любом варианте в последующем информация обрабатывается таким образом, чтобы была сформирована база данных, пригодная для статистического анализа. Такой метод широко распространен, но сопряжен с рядом ограничений: нередко ввод данных осуществляется сотрудником, не принимающим непосредственное участие в ведении пациента, могут возникать ошибки при переносе данных [5, 12]. Примером широкого использования подобного подхода к сбору информации может служить биобанк Люксембурга, в сферу услуг которого помимо сбора биообразцов из крупнейших клиник страны входит создание так называемых "индивидуальных карт пациента" (Case Report Form, CRF)¹. Подобный метод сбора информации часто используется при проведении различных популяционных исследований, когда изучается отобранная по специальной методике представительная выборка населения. Примером российских исследований, проводимых по данной технологии, являются три этапа крупного эпидемиологического проекта ЭССЕ-РФ (Эпидемиология сердечно-сосудистых заболеваний и их факторов риска в регионах Российской Федерации) [13].

Manders P, et al. разработали 3-ступенчатый алгоритм выбора клинических данных, необходимых для регистрации при биобанкировании [14, 15].

В качестве первого этапа предлагается определение сферы планируемого научного исследования. При этом выделяют пять основных сфер. Первой из них является изучение терапевтических подхо-

дов, в т.ч. подбор персонифицированных вариантов лечения в зависимости от выявляемых у пациента уровней биологических маркеров. Второй сферой является профилактика — поиск и совершенствование превентивных подходов. Третья сфера — это диагностика — изучение новых биомаркеров, позволяющих улучшить выявляемость, способствующих более точной постановке диагноза, уточняющих существующие классификации болезней. Четвертая сфера подразумевает изучение прогноза — поиск путей его улучшения и оценка прогностической значимости различных биомаркеров. Пятая сфера — это изучение этиологии того или иного заболевания — постижение механизмов развития и течения патологии [14, 15].

Вторым этапом является выделение основных категорий и подкатегорий информации, необходимой для проведения данного конкретного исследования. В частности, авторы выделяют такие категории, как "идентификация донора" (номер и демографическая информация), "характеристика заболевания" (диагноз, анамнез, результаты обследований и т.п.), "факторы риска" (образ жизни, факторы окружающей среды, семейный анамнез и т.п.), "терапия" (информация о лечении основного заболевания и сопутствующей терапии) и "течение и исход заболевания" (информация о степени тяжести, эффективности лечения и т.п.).

Третий этап — это создание эффективной структуры данных биобанка, которое подразумевает критическую оценку выбранных параметров с целью их оптимизации и при необходимости сокращения [14, 15].

Еще одним вариантом получения клинических и других видов данных является использование ЭМК, электронных историй болезни (ЭИБ) и других документов, в которые естественным образом собирается информация в процессе обследования и лечения пациентов [5]. Данный вариант стал возможен в последние десятилетия благодаря широкому распространению компьютерной техники и массовому переходу медицинских учреждений к электронным вариантам создания и хранения записей при оказании помощи пациенту в рамках так называемых медицинских информационных систем (МИС). Преимуществами такого подхода является широкий спектр получаемых данных (обычно в ЭМК и ЭИБ поступает вся основная клинически значимая информация о пациенте), проспективный характер имеющейся информации (данные о повторных визитах или госпитализациях, в т.ч. на протяжении длительного времени), большая точность информации (обычно ЭМК и ЭИБ заполняются медперсоналом, непосредственно участвующим в лечении пациента), отсутствие ошибок, связанных с переносом данных, более низкая стоимость создания клинической аннотации и др. В то же время существует ряд огра-

¹ Integrated Biobank of Luxembourg (IBBL). <https://www.lih.lu/en/translational-medicine/translational-medicine-operations-hub/integrated-biobank-of-luxembourg-ibbl/>.



Рис. 2 Развитие трансляционной геномики опирается на исследования всего генома и фенома. Прогресс зависит от наличия ресурсов и от аналитических методов и инструментов, позволяющих извлечь пользу из этих ресурсов. Крупномасштабные исследования с использованием новых технологий, основанных на объединении информации из ЭМК и ЭИБ с геномными данными могут привести к совершенствованию оказания медицинской помощи (адаптировано из [8]).

Примечание: GWAS — полногеномный поиск генетических ассоциаций, PheWAS — полнофеномный поиск генетических ассоциаций, е-фенотипирование — фенотипирование на основе данных, хранящихся в электронном виде, ЭМК — электронные медицинские карты, ЭИБ — электронные истории болезни.

ничений, которые необходимо учитывать при использовании данного метода: объем информации для разных пациентов может существенно различаться, длительность наблюдения также может значительно варьировать, неизбежным является смещение выборки — за медицинской помощью более тяжелые больные обращаются чаще, обычно отсутствует четкая стандартизация диагностических и лечебных процедур, проводимых различными операторами и врачами, отсутствует рандомизация и др. [5].

Интересным является опыт создания специальных некоммерческих ресурсов, в т.ч. на государственном уровне, призванных оптимизировать и структурировать деперсонифицированные биомедицинские данные, поступающие из ЭМК и ЭИБ медицинских учреждений страны, с возможностью предоставления их различным исследовательским группам страны для ускорения научных исследований, что, к примеру, уже реализовано в Австралии².

² Connecting health information. <https://www.biogrid.org.au/>.

Массив данных, доступных в ЭМК и ЭИБ, обеспечивает более обширную и потенциально более детальную информацию о состоянии здоровья пациентов по сравнению с той, которую получают в большинстве когортных клинических исследований [8]. Например, результаты лабораторных и инструментальных исследований, консультации врачей разного профиля, эпикризы предоставляют объективные данные, позволяющие определить обоснование диагноза, стадию заболевания и другие показатели. Кроме того, доступность длительного наблюдения дает возможность изучения взаимосвязи факторов риска и заболеваний на протяжении жизни человека.

Недавние достижения в области генетических технологий и широкое использование данных ЭМК и ЭИБ ускорили развитие геномной медицины, привнеся многообещающие результаты исследований генома в клиническую практику. Данные генома и фенома, связанные с ЭМК и накопленные для больших популяций благодаря технологиям биобанкирования,

позволяют изучать генетические вариации в масштабе всего фенома (рисунок 2). Так, совмещение данных ЭМК и ЭИБ с результатами секвенирования соответствующих биообразцов от 50726 пациентов, хранящихся в биобанке, дало возможность изучить влияние генетических вариантов, связанных с семейной гиперхолестеринемией (СГХС), на вероятность развития ишемической болезни сердца в молодом возрасте, которая оказалась в 3,7 раза выше у носителей вариантов, ассоциированных с СГХС [16]. Примечательно, что лишь 25% носителей данных вариантов соответствовали клиническим критериям постановки диагноза СГХС. Это исследование стало возможным только благодаря наличию многолетних данных, хранящихся в ЭМК, и генетическому анализу биообразцов, хранящихся в биобанке [16].

Проект "eMERGE Network" (The Electronic Medical Records and Genomics Network, сеть электронных медицинских карт и геномики), инициированный в 2007г, сфокусирован на изучении возможностей, которые дает объединение данных МИС медицинских учреждений с ресурсами биобанков и биорепопозиториев, предоставляющих биообразцы для проведения генетических исследований [17]. С момента основания проект прошел несколько этапов, в рамках которых были проведены работы по таким направлениям, как клиническое аннотирование генетических вариантов, фармакогеномика, интеграция медицинских и генетических результатов в МИС в клинических центрах, изучение распространенных и редких генетических вариантов, развитие научно-обоснованных методов фенотипирования с использованием данных из МИС с целью создания выборок для исследований, изучение влияния получаемых в рамках проекта генетических данных на различные аспекты оказания медицинской помощи пациенту, исследование юридических, этических и социальных особенностей интеграции генетических данных с ЭМК и ЭИБ [5, 17]. В частности, именно в рамках проекта eMERGE впервые был разработан метод PheWAS, а в 2013г был создан каталог PheWAS, процитированный к 2020г >300 раз [18, 19]. Анализ данных и идентификация обобщенных признаков (фенотипа) того или иного заболевания могут занимать довольно продолжительное время. Всю информацию, содержащуюся в МИС, подразделяют на дискретные данные (представленные заранее предопределенными структурированными значениями, например, дата рождения, пол, код диагноза согласно международной классификации болезней, МКБ), а также большое количество другой, менее структурированной информации, часто в виде текста в свободном формате. Сеть eMERGE разработала и внедрила алгоритмы обработки информации на основе ЭМК и ЭИБ для различных фенотипов [20, 21]. Для извлечения клинической информации из текстового формата и об-

легчения идентификации фенотипа применяются различные стандарты представления данных (часть из них описаны ниже): SNOMED-CT (Systemized Nomenclature of Medicine — Clinical Terms, систематизированная номенклатура в медицине — клинические термины), CPT (Current Procedural Terminology, современная терминология процедур), LOINC (Logical Observation Identifiers Names and Codes, имена и коды идентификаторов логических наблюдений), FHIR (Fast Health Interoperability Resources, ресурсы быстрого взаимодействия в сфере здравоохранения) и др., а также системы обработки естественного языка, например, "MedLEE", "MetaMap", "KnowledgeMap" и "cTAKES" [22-24]. Подобный подход позволяет успешно проводить анализ данных как в рамках одного клинического центра, так и в многоцентровых исследованиях, даже при наличии различий в клинической практике между разными учреждениями [20, 24].

Существуют и смешанные варианты создания клинической аннотации биообразцов, объединяющие оба указанных выше подхода. Ярким примером реализации такого смешанного варианта является Биобанк Великобритании (UK Biobank), который включает биоматериал и информацию от ~500 тыс. доноров³. При включении в исследование заполняется целый пул опросников и форм на каждого участника, при этом клиническая аннотация постоянно пополняется благодаря сопряжению с ЭМК национальной системы здравоохранения Великобритании. Кроме того, к клинической аннотации добавляются данные, получаемые в рамках проводимых научных работ — генетическая информация, данные, полученные с помощью высокотехнологичных методов исследования (магнитно-резонансная и мультиспиральная компьютерная томография) и многое другое. Подобный подход позволяет анализировать и проводить исследования в совершенно различных направлениях поиска новых ассоциаций биомаркеров, факторов окружающей среды, образа жизни с развитием того или иного заболевания, выявления предикторов заболеваний и т.п. [25-27].

Стандарты, затрагивающие вопросы клинического аннотирования биообразцов

Учитывая наличие проблемы недостаточной воспроизводимости результатов научных исследований, ряд организаций и консорциумов опубликовали различные стандарты и рекомендации, посвященные аннотированию биообразцов и правилам изложения при публикации результатов исследований.

В 2011г были опубликованы рекомендации BRISQ (Biospecimen Reporting for Improved Study Quality, Правила описания биообразцов для улучшения качества научных исследований) [28, 29]. В этих рекомендациях, помимо описания основных харак-

³ UK Biobank. <https://www.ukbiobank.ac.uk/>.

теристик получения и хранения самого биообразца в соответствии с его жизненным циклом, содержится часть, посвященная клинической аннотации биообразцов. Все элементы по степени важности распределены на 3 группы — обязательные к описанию, рекомендуемые (полезные) и дополнительные. Диагноз, витальный статус и клинические данные пациентов относятся к характеристикам, обязательным к описанию. К рекомендуемым характеристикам исследователи относят демографическую информацию, а также время между постановкой диагноза и взятием биообразца. Дополнительными параметрами клинической аннотации, согласно рекомендациям BRISQ, являются стадия заболевания (например, острая или хроническая), время относительно приема лекарственной терапии, различные виды терапии, которые могут повлиять на биообразцы (например, химио- или лучевая терапия, внутривенные инфузии), статус курения, репродуктивный статус [28].

В литературе представлен ряд рекомендаций, разработанных для конкретных патологий или видов исследований. Например, рекомендации REMARK (Reporting Recommendations for Tumor Marker Prognostic Studies, Рекомендации по представлению данных исследований прогностических опухолевых маркеров) среди прочего содержат такие характеристики, как стадия заболевания, клинические показатели, специфичные для изучаемого заболевания и влияющие на прогноз (в т.ч. в сравнении с изучаемым маркером) и сопутствующие диагнозы [30]. Рекомендации TRIPOD (Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis, четкое представление многомерной модели прогнозирования для индивидуального прогнозирования или диагностики) направлены на стандартизацию при проведении исследований в сфере прогнозирования и содержат 22 элемента; в частности, рекомендуется тщательное описание клинических данных каждого пациента и методов его лечения [31].

Отдельной задачей является объединение усилий различных биобанков при проведении крупномасштабных исследований по различным направлениям. Фактически многие биобанки обладают огромным количеством образцов, которые используются недостаточно, но потенциально представляют собой "золотую жилу" для самых разнообразных научных проектов и коллабораций, при условии, что информация об их наличии станет доступной и видимой для различных исследовательских групп [32, 33]. Предпосылкой для сотрудничества является возможность комплексного поиска и доступа к фенотипической, клинической и другой информации об образцах, которые в настоящее время хранятся в биобанках. Однако проблемы конфиденциальности, наряду с отсутствием согласованных словарей и гетерогенными информационными

системами, приводят к значительным сложностям при поиске образцов в разных биобанках [34].

С целью унификации информации при обмене данными между биобанками в 2012г был инициирован проект MIABIS (The Minimum Information About BioBank data Sharing, Минимальная информация об обмене данными биобанка), в рамках которого предложены 52 характеристики, необходимые для описания как биобанка, так и конкретной коллекции [32]. При описании коллекции часть характеристик посвящена особенностям исследования, в рамках которого данная коллекция собрана (например, "случай-контроль", когортное или популяционное исследование и др.); относительно клинической аннотации выделены такие характеристики, как минимальный, средний и максимальный возраст доноров коллекции, основной диагноз и сопутствующие заболевания. Также рекомендуется предоставление информации о дополнительных медицинских данных, доступных для данной коллекции. Кроме того, в случае уже имеющихся результатов научных исследований, проведенных на биообразцах из конкретной коллекции, рекомендуется предоставить информацию о том, какие "омиксные" данные доступны (например, данные исследования генома, транскриптома, протеома, метаболома и т.п.) [32]. В расширенной обновленной версии MIABIS представлены не только рекомендуемые правила описания на уровне биобанков и коллекций, но также и на уровне биообразцов и доноров биоматериала [35]. В частности, относительно доноров биообразцов составлен предустановленный список вариантов описания, ассоциированных с ним данных и биообразцов: вид биообразцов, данные опросников, изображений, пол, национальность, данные из медицинских карт, национальных регистров, генеалогические данные, информация из патологоанатомических подразделений, физиологические или биохимические измерения, психологические данные и др.; рекомендован формат отображения даты рождения, как гггг-мм-дд [35]. Подчеркивается важность не только ассоциированного с каждым биообразцом диагноза с указанием кода из МКБ, но также временной шкалы событий (даты, возраста пациента в момент развития события) [35].

Полезным представляется опыт консорциума "Биобанкирование материала пациентов с онкологическими заболеваниями головного мозга" (Brain Cancer Biobanking Australia, BCBA), созданного с целью объединения усилий биобанков по всей Австралии, специализирующихся на онкологических, в т.ч. редких патологиях головного мозга у взрослых и детей. Для унификации клинических данных, ассоциированных с биообразцами в разных биобанках страны, был создан специальный "словарь", содержащий стандартизованное описание, структуру и формат для каждого элемента данных и ограничи-

вающий свободный текст во всех возможных случаях, при этом рекомендации по представлению клинических данных сравнивались с другими стандартами, такими как BRISQ и MIABIS [36].

Рабочая группа консорциума BCVA выделила 5 категорий клинических данных: демографическая информация, клинические и радиологические диагнозы, информация о выполненных хирургических вмешательствах, установленный при нейропатологическом исследовании диагноз, методы лечения пациента и данные, получаемые в ходе дальнейшего наблюдения пациента [36]. В каждой из категорий выделены наиболее важные обязательные данные (включают информацию, которая может быть собрана большинством биобанков), рекомендуемые к описанию (включают информацию, которая может повысить ценность потенциальных научных исследований) и расширенные (включают практически всю клиническую информацию, полученную при обследовании и лечения пациента и потенциально важную для проведения исследований в сфере персонализированной медицины). Например, обязательными общими данными о больном являются идентификаторы пациента, дата рождения, пол, возраст при постановке диагноза, рекомендуемыми — фамилия лечащего врача, расширенными — национальность пациента. В отношении диагноза обязательными являются: дата постановки диагноза, первичный источник опухоли, обоснование диагноза, дата и тип радиологического исследования, наличие наследственных синдромов, гистологический тип опухоли и др.; рекомендуемыми к описанию являются радиологический диагноз и заключение, а расширенными данными считаются все имеющиеся клинические данные, в т.ч. оригинальные изображения компьютерной и магнитно-резонансной томографии в формате DICOM (Digital Imaging and Communications in Medicine, стандарт цифровых медицинских изображений и документов) и т.п. В отношении описания лечения в качестве обязательной выбрана информация о том, получал ли пациент лечение, рекомендуемыми — даты терапии, сведения о том, проводилась ли она до или после операционного лечения, проводилась ли химио- или радиолучевая терапия, осуществлялась ли она в рамках того или иного исследования; в качестве расширенных данных приведены точные даты начала и окончания терапии, режимы дозирования, была ли проведена в полном объеме и, если нет, то по какой причине и т.п. [36].

Динамичное развитие индустрии биобанкирования привело к увеличению количества систем с использованием информационных технологий для управления клиническими и другими данными, связанными с биообразцами. Широкодоступные программные инструменты относятся, в основном, к конкретному биобанку или биоколлекции. Сотрудничество между учреждениями, оснащенными различными системами информационных технологий,

затруднено. Эффективные и универсальные протоколы, включающие подробную информацию о доноре и образце, все еще очень ограничены. В связи с этим Jarczak J, et al. создали и предлагают к использованию специальный размещенный в открытом доступе протокол обмена данными (авторское название BioSCOOP), целью которого является гармонизация и стандартизация правил общения между биобанками на уровне информации о доноре вместе с информацией об образце при решении двух видов задач: обмена информацией между различными биобанками и обеспечения поиска и представления наборов образцов и данных при проведении совместных исследований [37]. Импорт данных в формате BioSCOOP реализован в связанном с ним проекте "Bioface". "Bioface" имеет распределенную архитектуру и представлен как браузер для членов сети биобанков Польши и других биобанков, используемый с целью поиска образцов из различных биобанков и биохранилищ [37, 38]. Кроме того, авторы предложили достаточно подробное стандартизированное описание фенотипических и антропометрических данных (например, индекс массы тела, окружность талии, размеры головы, туловища, конечностей и т.п.). Для описания клинических данных авторы предлагают использование МКБ, а также коллекции медицинских терминов SNOMED CT⁴.

"SNOMED International" — некоммерческая организация, которая владеет, управляет и развивает SNOMED CT — базу стандартизированных медицинских терминов, включая разработку стандартов кодифицированного языка, который представляет группы клинических терминов и поддерживает представление подробной клинической информации таким образом, чтобы ее можно было обрабатывать автоматически. В настоящее время база доступна на английском и испанском языках и используется во многих странах мира⁴. Основными типами компонентов в данной системе являются "концепты", "описания" и "отношения". Логическая модель определяет способ связи и представления каждого типа компонентов. Каждый "концепт" представляет собой уникальное клиническое понятие, которое обозначается с помощью числового идентификатора, обеспечивающего ссылку на это понятие. "Отношения" представляют собой связь между двумя понятиями. "Отношения" используются для логического определения значения понятия таким образом, чтобы его мог обработать компьютер. "Описания" — это удобочитаемые термины, связанные с клиническими понятиями. "Описания" включают такие разновидности, как "полностью определенное имя" (FSN, fully specified name) и "синонимы". FSN — это тип описания, которое однозначно и полностью отражает смысл кли-

⁴ SNOMED CT, the world's most comprehensive clinical terminology. <https://www.snomed.org/>.

нического понятия. Выделяют также "синонимы", которые подразумевают описания, позволяющие выразить одно и то же понятие разными способами, каждый из которых связан с одним и тем же идентификатором понятия⁴. Наиболее частым применением данной системы является стандартизация данных в ЭМК и ЭИБ, необходимая для повышения преемственности медицинской помощи, проведения научно-исследовательских работ и др. [39].

Представляет интерес идентификатор LOINC, который создан для унификации представления данных в ЭИБ, в т.ч. результатов лабораторных и других методов обследования⁵. Идентификатор размещен в открытом доступе и широко используется различными лабораториями. Любой термин в кодификаторе LOINC определяется как комбинация кода LOINC и так называемого "полностью определенного имени" (FSN). FSN состоит из пяти или шести основных частей: названия измеряемого компонента или аналита (например, глюкоза, пропранолол), наблюдаемого свойства (например, концентрация вещества, масса, объем), временного аспекта измерения (например, является ли оно протяженным во времени или мгновенным), типа системы или образца (например, моча, сыворотка), шкала измерения (например, качественный или количественный) и, где уместно, метод измерения (например, радиоиммуноанализ, иммуноферментный анализ). FSN представляет собой комбинацию основных частей и символа двоеточия ":", который действует как разделитель, например: <Аналит/компонент>:<вид свойства наблюдения или измерения>:<временной аспект>:<система (образец)>:<масштаб>:<метод>⁶. Системы SNOMED-CT и LOINC были успешно применены для объединения данных 15 региональных биобанков Кореи, а также широко используются в рамках исследований сети eMERGE и других проектов [23, 40].

⁵ <https://loinc.org/>.

⁶ <https://loinc.org/>.

Литература/References

1. Astrin JJ, Betsou F. Trends in Biobanking: A Bibliometric Overview. *Biopreserv Biobank*. 2016;14(1):65-74. doi:10.1089/bio.2015.0019.
2. 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.
3. Pokrovskaya MS, Sivakova OV, Efimova IA, et al. Biobanking as a necessary tool for research in the field of personalized medicine in the scientific medical center. *Per Med*. 2019;16(6):501-9. doi:10.2217/pme-2019-0049.
4. Begley CG, Ellis LM. Raise standards for preclinical cancer research. *Nature*. 2012;483(7391):531-3. doi:10.1038/483531a.
5. Kopylova OV, Ershova AI, Efimova IA, et al. Electronic medical records and biobanking. *Cardiovascular Therapy and Prevention*. 2022;21(11):3425. (In Russ.) Копылова О. В., Ершова А. И., Ефимова И. А. и др. Электронные истории болезни и биобанкирование. *Кардиоваскулярная терапия и профилактика*. 2022;21(11):3425. doi:10.15829/1728-8800-2022-3425.
6. Manolio TA. Genomewide Association Studies and Assessment of the Risk of Disease. Feero WG, Gutmacher AE, eds. *N Engl J Med*. 2010;363(2):166-76. doi:10.1056/NEJMra0905980.
7. Bastarache L, Denny JC, Roden DM. Phenome-Wide Association Studies. *JAMA*. 2022;327(1):75. doi:10.1001/jama.2021.20356.
8. Linder JE, Bastarache L, Hughey JJ, Peterson JF. 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.

В 1987г была основана международная некоммерческая организация по разработке стандартов для обмена, интеграции, совместного использования и поиска электронной медицинской информации (Health Level Seven International, HL7)⁷. "Level Seven (седьмой уровень)" относится к 7-уровневой модели связи Международной организации по стандартизации (ISO) для взаимодействия открытых систем — уровень применения и внедрения. Среди прочего организацией разработан стандарт FHIR, который предназначен для обмена информацией, связанной со здравоохранением (клинические и административные данные, информация из научных исследований и др.). HL7 поддерживают >1600 участников из >50 стран⁷.

Заключение

Таким образом полнота, структурированность и стандартизация относятся к важнейшим характеристикам клинической аннотации, ассоциированной с биообразцами. Основными методами составления клинической аннотации являются использование индивидуальных регистрационных карт пациента и объединение данных МИС с информационной системой биобанка. В настоящее время разрабатываются различные стандарты, позволяющие унифицировать представление клинической аннотации, сделать биобанки и коллекции более "видимыми" для сторонних исследователей и организаций, что необходимо для сотрудничества и более эффективного и интенсивного использования хранящихся биообразцов. Гармонизация методологии клинического аннотирования между различными биобанками открывает широкие горизонты для проведения крупномасштабных исследований в сфере персонализированной и трансляционной медицины.

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

⁷ HL7 International. <https://www.hl7.org/about/index.cfm?ref=nav>.

9. Zhu T, Wang W, Chen Y, et al. Machine Learning of Functional Connectivity to Biotype Alcohol and Nicotine Use Disorders. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 2023;S2451-9022(23)00222-7. doi:10.1016/j.bpsc.2023.08.010.
10. Wang X, Khurshid S, Choi SH, et al. Genetic Susceptibility to Atrial Fibrillation Identified via Deep Learning of 12-Lead Electrocardiograms. *Circ Genomic Precis Med*. 2023;16(4):340-9. doi:10.1161/CIRCGEN.122.003808.
11. Bellary S, Krishnankutty B, Latha M. Basics of case report form designing in clinical research. *Perspect Clin Res*. 2014;5(4):159. doi:10.4103/2229-3485.140555.
12. Cowie MR, Blomster JI, 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.
13. Shalnova SA, Drapkina OM. Contribution of the ESSE-RF study to preventive healthcare in Russia. *Cardiovascular Therapy and Prevention*. 2020;19(3):2602. (In Russ.) Шальнова С.А., Драпкина О.М. Значение исследования ЭССЕ-РФ для развития профилактики в России. *Кардиоваскулярная терапия и профилактика*. 2020;19(3):2602. doi:10.15829/1728-8800-2020-2602.
14. Manders P, Peters TM, Siezen AE et al. A Stepwise Procedure to Define a Data Collection Framework for a Clinical Biobank. *Biopreserv Biobank*. 2018;16(2):138-47. doi:10.1089/BIO.2017.0084.
15. Meshkov AN, Glotov AS, Anisimov SV, et al. Biobanking: national guidelines. prepared by experts from the National Association of Biobanks and Biobanking Professionals (NASBIO). М.: Triumph, 2022. р. 308. (In Russ.) Мешков А.Н., Глотов А.С., Анисимов С.В. и др. Биобанкирование: национальное руководство. подготовлено экспертами Национальной ассоциации биобанков и специалистов по биобанкированию (НАСБИО). М.: Триумф, 2022. 308 с. ISBN: 978-5-93673-322-2.
16. 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.
17. Gottesman O, Kuivaniemi H, Tromp G, et al. The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. *Genet Med*. 2013;15(10):761-71. doi:10.1038/gim.2013.72.
18. 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.
19. 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.
20. Dumitrescu L, Goodloe R, Bradford Y, et al. The effects of electronic medical record phenotyping details on genetic association studies: HDL-C as a case study. *BioData Min*. 2015;8(1):15. doi:10.1186/s13040-015-0048-2.
21. Newton KM, Peissig PL, Kho AN, et al. Validation of electronic medical record-based phenotyping algorithms: results and lessons learned from the eMERGE network. *J Am Med Informatics Assoc*. 2013;20(e1):e147-54. doi:10.1136/amiajnl-2012-000896.
22. Duan R, Cao M, Wu Y, et al. An Empirical Study for Impacts of Measurement Errors on EHR based Association Studies. *AMIA. Annu Symp Proc AMIA Symp*. 2016;2016:1764-73.
23. Krishnamoorthy P, Gupta D, Chatterjee S, et al. A Review of the Role of Electronic Health Record in Genomic Research. *J Cardiovasc Transl Res*. 2014;7(8):692-700. doi:10.1007/s12265-014-9586-0.
24. Pacheco JA, Rasmussen LV, Wiley K, et al. Evaluation of the portability of computable phenotypes with natural language processing in the eMERGE network. *Sci Rep*. 2023;13(1):1971. doi:10.1038/s41598-023-27481-y.
25. Khera A V., Chaffin M, Wade KH, et al. Polygenic Prediction of Weight and Obesity Trajectories from Birth to Adulthood. *Cell*. 2019;177(3):587-96.e9. doi:10.1016/j.cell.2019.03.028.
26. Bešević J, Lacey B, Conroy M, et al. New Horizons: the value of UK Biobank to research on endocrine and metabolic disorders. *J Clin Endocrinol Metab*. 2022;107(9):2403-10. doi:10.1210/clinem/dgac407.
27. Conroy MC, Lacey B, Bešević J, et al. UK Biobank: a globally important resource for cancer research. *Br J Cancer*. 2023;128(4):519-27. doi:10.1038/s41416-022-02053-5.
28. Moore HM, Kelly A, Jewell SD, et al. Biospecimen Reporting for Improved Study Quality. *Biopreserv Biobank*. 2011;9(1):57-70. doi:10.1089/BIO.2010.0036.
29. 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. (In Russ.) Сивакова О.В., Покровская М.С., Метельская В.А. и др. Международные правила описания биобразцов — важный фактор повышения качества научных исследований. *Профилактическая медицина*. 2019;22(6):95-9. doi:10.17116/profmed20192206295.
30. Sauerbrei W, Taube SE, McShane LM, et al. Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK): An Abridged Explanation and Elaboration. *JNCI J Natl Cancer Inst*. 2018;110(8):803-11. doi:10.1093/jnci/djy088.
31. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *Eur J Clin Invest*. 2015;45(2):204-14. doi:10.1111/eci.12376.
32. Norlin L, Fransson MN, Eriksson M, et al. A Minimum Data Set for Sharing Biobank Samples, Information, and Data: MIABIS. *Biopreserv Biobank*. 2012;10(4):343-8. doi:10.1089/bio.2012.0003.
33. Metzler I, Ferent LM, Felt U. On samples, data, and their mobility in biobanking: How imagined travels help to relate samples and data. *Big Data Soc*. 2023;10(1):205395172311586. doi:10.1177/20539517231158635.
34. Spjuth O, Krestyaninova M, Hastings J, et al. Harmonising and linking biomedical and clinical data across disparate data archives to enable integrative cross-biobank research. *Eur J Hum Genet*. 2016;24(4):521-8. doi:10.1038/ejhg.2015.165.
35. Eklund N, Andrianarisoa NH, van Enckevort E, et al. Extending the Minimum Information About Biobank Data Sharing Terminology to Describe Samples, Sample Donors, and Events. *Biopreserv Biobank*. 2020;18(3):155-64. doi:10.1089/bio.2019.0129.
36. Gedye C, Sachchithanathan M, Leonard R, et al. Driving innovation through collaboration: development of clinical annotation datasets for brain cancer biobanking. *Neuro-Oncology Pract*. 2020;7(1):31-7. doi:10.1093/nop/npz036.
37. Jarczak J, Lach J, Borówka P, et al. BioSCOOP — Biobank Sample Communication Protocol. New approach for the transfer of information between biobanks. *Database*. 2019;2019. doi:10.1093/database/baz105.
38. Proynova R, Alexandre D, Lablans M, et al. A Decentralized IT Architecture for Locating and Negotiating Access to Biobank Samples. *Stud Health Technol Inform*. 2017;243:75-9.
39. Vuokko R, Vakkuri A, Palojoki S. Systematized Nomenclature of Medicine-Clinical Terminology (SNOMED CT) Clinical Use Cases in the Context of Electronic Health Record Systems: Systematic Literature Review. *JMIR Med Informatics*. 2023;11:e43750. doi:10.2196/43750.
40. Park HS, Cho H, Kim HS. Development of an Integrated Biospecimen Database among the Regional Biobanks in Korea. *Health Inform Res*. 2016;22(2):129. doi:10.4258/hir.2016.22.2.129.