РОССИЙСКИЕ НАНОТЕХНОЛОГИИ, 2023, том 18, № 3, с. 292–301

ОБЗОРЫ

УДК 575.113

ЦИФРОВИЗАЦИЯ ГЕНЕТИЧЕСКОЙ ИНФОРМАЦИИ: ПЕРСПЕКТИВЫ И ВЫЗОВЫ

© 2023 г. З. Б. Намсараев[1,*], А. А. Корженков[1], Д. Ю. Федосов[1], М. В. Патрушев[1]

[1]Национальный исследовательский центр "Курчатовский институт", Москва, Россия

*E-mail: zorigto@gmail.com

Поступила в редакцию 28.04.2023 г.
После доработки 28.04.2023 г.
Принята к публикации 05.05.2023 г.

В настоящее время процесс цифровизации биологической информации находится на ранних этапах развития, характеризующихся ускоренным экспоненциальным ростом баз данных, что открывает широкие возможности для развития персонализированной медицины, сохранения биоразнообразия, развития биотехнологий и сельского хозяйства, но в то же время дает дополнительные возможности для разработки технологий двойного назначения и биологического терроризма. Проведен анализ текущей ситуации в области цифровизации генетической информации и глобальных вызовов, стоящих перед человечеством при дальнейшем увеличении объемов генетической информации. Насущной является разработка механизмов государственного и межгосударственного контроля в этой области, а также поиск оптимального баланса между требованиями конфиденциальности персональных данных населения, соблюдением правовых и этических норм, необходимостью проведения научных исследований, развитием технологий персонализированной медицины и разработкой бизнес-моделей и организационно-правовых форм, способных сопровождать в дальнейшем рост объемов генетических данных.

DOI: 10.56304/S199272232303007X

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ

Генетическая и, более широко, биологическая информация находятся в основе важнейших областей, таких как производство продуктов питания, медицина, экология, а также ключевых вопросов развития человечества, включая разработку средств борьбы с пандемиями, обеспечение долгосрочной устойчивости общества во время климатических изменений и глобального сокращения биоразнообразия. Развитие технологий анализа ДНК привело к резкому снижению стоимости секвенирования и увеличению объема оцифрованной генетической информации. Согласно данным Национального института исследования генома человека (США) стоимость секвенирования одного генома человека упала со 100 млн долл. в 2001 г. до 1 тыс. долл. в 2017 г. Начиная с 2007 г. темпы снижения стоимости секвенирования ДНК даже превышают скорость увеличения производительности компьютерных микросхем согласно закону Мура [1]. В результате по состоянию на 2023 г. заявленная стоимость анализа генома человека с 30-кратным покрытием может достигать 299 долл. США и со 100-кратным покрытием – 999 долл. [2].

Рост производительности секвенирования сопровождается ростом объема баз данных генетической информации. Одна из ведущих мировых баз данных генетической информации Genbank, поддерживаемая Национальным центром биотехнологической информации США, по данным 2022 г. увеличивается со скоростью 61% в год [3]. Это значительно превышает скорость увеличения объема оцифрованной информации в мире, прогнозируемый совокупный среднегодовой темп роста которой составляет 23% на промежутке между 2020 и 2025 г. [4]. При сохранении таких темпов цифровизации генетической информа-

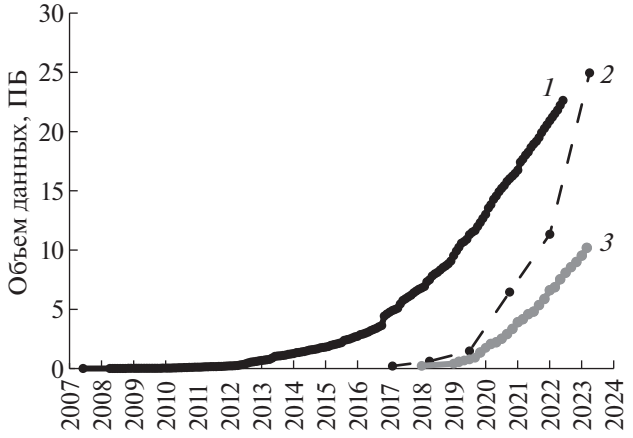

Рис. 1. Рост объема баз данных генетической информации в 2007–2023 гг. в Петабайтах: *1* – NCBI SRA, *2* – NGDC GSA, *3* – CNGBdb.

ции на временном промежутке от нескольких десятилетий до 110 лет ДНК всех живых организмов на Земле может быть проанализирована и переведена в цифровой вид [5]. Развитие методов анализа полученной информации и ее применения для изменения свойств живых организмов приведет к глобальному внедрению генетических технологий как в сельское хозяйство, так и в повседневную жизнь человека – через технологии медицины, обеспечения качества и продолжительности жизни. Тем не менее генетические технологии могут быть использованы во вред человечеству отдельными государствами, криминальными и террористическими группами. Таким образом, человечество находится на том этапе развития, когда увеличение объема знаний о биологических основах жизни способно привести как к улучшению качества жизни, так и к глобальным вызовам самому существованию человечества. В этом ключе генетические технологии, как ничто иное, максимально близки к ядерным технологиям по возможности своего двойного применения.

Цель настоящей работы – анализ текущей ситуации в области цифровизации генетической информации и глобальных вызовов, стоящих перед человечеством при появлении в широком доступе генетической информации и инструментов реализации генетических технологий.

1. УНИВЕРСАЛЬНЫЕ БАЗЫ ДАННЫХ ГЕНЕТИЧЕСКОЙ ИНФОРМАЦИИ

По состоянию на конец 2021 г. по данным ежегодного обзора Molecular Biology Database Collection журнала Nucleic acid research в мире насчитывалось 1645 баз данных молекулярно-биологической информации [6]. Крупнейшей базой данных являются консорциум International Nucleotide Sequence Database Collaboration (INSDC), включающий в себя DNA Data Bank of Japan (DDBJ), European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI) и National Center for Biotechnology Information (NCBI). В рамках этого консорциума в 2005 г. была создана объединенная база данных International Nucleotide Sequence Database (INSD), но фактически сотрудничество между тремя организациями, входящими в консорциум, началось в начале 1980-х гг. Самый крупный раздел базы данных включает в себя Sequence Read Archive (SRA), размер публично доступной базы которой по состоянию на июнь 2022 г., составил 14 петабайт (**ПБ**) [7].

Близкой по масштабу базой данных является National Genomics Data Center (NGDC), входящая в China National Center for Bioinformation (CNCB), который официально был основан в 2019 г. CNCB-NGDC основан на базе трех институтов Chinese Academy of Sciences: Beijing Institute of Genomics, Institute of Biophysics и Shanghai Institute of Nutrition and Health. По состоянию на май 2023 г. размер базы данных Genome Sequence Archive – 25 ПБ [8]. В партнерстве с европейским и американским центрами EBI и NCBI был реализован проект BIG Search, масштабируемая универсальная поисковая система с кросс-базами данных. Кроме того, NGDC был разработан проект Database Commons, в котором каталогизируются открытые базы биологических данных по всему миру и обеспечивается доступ к ним [9]. По состоянию на август 2022 г. в каталоге находились 5832 базы данных из 72 стран. Наибольшее количество баз зарегистрировано в США (1433 базы), на втором месте Китай (1110 баз), затем Индия (431 база), Великобритания (425 баз) и другие страны.

В НИЦ “Курчатовский институт” в соответствии с Федеральным законом от 29.12.2022 № 643-ФЗ “О внесении изменений в Федеральный закон "О государственном регулировании в области генно-инженерной деятельности” создан прототип Национальной базы генетической информации (**НБГИ**), создаваемой для обеспечения национальной безопасности, охраны жизни и здоровья граждан, суверенитета в сфере хранения и использования генетических данных, а также обеспечения обмена содержащейся в ней информацией между государственными органами, органами местного самоуправления и обладателями генетических данных при их взаимодействии в рамках осуществления генно-инженерной деятельности. С 1 сентября 2024 г. в НБГИ будет депонироваться вся генетическая информация, получаемая в Российской Федерации.

2. БАЗЫ ДАННЫХ ГЕНЕТИЧЕСКОЙ ИНФОРМАЦИИ НАСЕЛЕНИЯ

В 1990-е г. развитие генетических технологий подтолкнуло национальные правительства к созданию генетических баз данных населения, основными стимулами к созданию которых стали вопросы безопасности, а именно быстрого опознания опасных преступников, а также решения медицинских задач.

В настоящее время крупнейшей базой данных населения обладает Китай, в состав базы данных Chinese Institute of Forensic Science Ministry of Public Security входит информация о генетических профилях не менее 68 млн чел. по состоянию на сентябрь 2018 г., при этом размер базы данных вырос с 55 до 68 млн менее чем за год [10]. Согласно сторонним оценкам, сделанным Australian Strategic Policy Institute, по состоянию на 2020 г. в базе данных находилась информация о 105–140 млн чел. [11].

Второй по размеру базой данных о генетических профилях населения обладает США, где функционирует база Combined DNA Index System (CODIS), принадлежащая Federal Bureau of Investigation. База данных была основана в 1998 г. и включала в себя более 17 млн профилей по состоянию на 2018 г. [10]. В основном база данных включает в себя информацию о правонарушителях, также собираются образцы ДНК родственников людей, пропавших без вести, и военных. Особенность системы – ее структура, включающая в себя уровни местный, штата и федеральный, и автономная информационная сеть Criminal Justice Information Systems Wide Area Network (CJIS WAN). В США действуют и другие проекты геномного анализа населения. В 2018 г. в США под эгидой National Institute of Health с медицинскими целями была запущена программа массового анализа геномов населения “All of US”. В планах программы проанализировать к 2022 г. геномы не менее 1 млн чел., по состоянию на 2021 г. проанализировано не менее 270 тыс. геномов [12].

Третья по размеру национальная база данных принадлежит Великобритании. С 1995 г. по программе NDNAD (Национальная база данных ДНК), находящейся в ведении Home Office, собрано по состоянию на 30 июня 2022 г. 6.9 млн генетических профилей лиц, так или иначе привлекавшихся к уголовной ответственности. В базе данных находится информация о тандемных повторах, а не полногеномная информация. Тем не менее образцы биологических материалов находятся в хранилище, что делает их доступными для более глубокого анализа. Кроме базы данных NDNAD, решающей криминалистические задачи, в Великобритании с 2006 г. функционирует программа UK Biobank. В рамках программы к середине 2023 г. планируется сделать доступными полногеномные последовательности ДНК 500 тыс. чел. Программа реализуется в рамках частно-государственного партнерства при участии Wellcome Sanger Institute и компании deCODE Genetics, а также фармацевтических компаний Amgen, AstraZeneca, GlaxoSmithKline и Johnson & Johnson [13].

Четвертая по размеру национальная база данных находится во Франции. С 1998 г. работает сеть FNAEG (Fichier National Automatisé des Empreintes Génétiques), в которой по состоянию на 31 декабря 2021 г. собрано более 6 млн генетических паспортов, в основном правонарушителей [14].

В настоящее время в Европейском союзе идет работа по объединению национальных баз данных ДНК населения в единую сеть, которая включит в себя по состоянию на конец 2021 г. не менее 16.8 млн профилей. Сеть будет включать в себя национальные базы стран-членов Евросоюза и Великобритании [15].

Одной из крупнейших международных баз генетической информации обладает Interpol. Сообщается, что в базе данных находятся 247 тыс. профилей, предоставленных 84 странами-членами этой организации [16]. Информация, предоставляемая странами в Interpol, не включает в себя данные о личности человека, а содержит в себе буквенно-цифровой код. Страны-члены сохраняют право собственности на информацию и могут выбирать, с какими другими странами они делятся своими данными.

Размер генетических баз данных коммерческих компаний, предоставляющих услуги анализа ДНК населению, сравним с размером крупнейших государственных баз данных. По оценкам на начало 2019 г. более 26 млн человек предоставили свои образцы четырем крупнейшим компаниям. Наибольшее число анализов было проведено компанией Ancestry (14 млн), за ней следуют 23andMe (9 млн), MyHeritage (2.5 млн) и Gene By Gene (2 млн) [17]. Отметим, что генетическое тестирование может быть достаточно прибыльным бизнесом. В августе 2020 г. компания Ancestry была куплена фондом Blackstone за 4.7 млрд долл., при этом сообщается, что годовой доход Ancestry составляет более 1 млрд долларов [18].

Повышение производительности секвенирования и снижение его стоимости привели к вовлечению в программы генетической паспортизации населения более широкого круга стран, а также переход от анализа отдельных участков генома к полногеномному секвенированию. Это позволило перейти от задач, нацеленных в первую очередь на идентификацию человека, к задачам выявления генов, кодирующих редкие болезни, разработке целевых фармацевтических средств и т.д. По состоянию на начало 2021 г. национальные геномные программы действуют в 41 стране мира. В основном

программы нацелены на изучение вариабельности геномов у населения (90%), определение вариантов, приводящих к заболеваниям (71%), развитие геномной инфраструктуры (59%) и развитие персонализированной медицины (37%) [19].

В России действует ряд программ генетического анализа населения. С 2006 г. ведется неонатальный скрининг всех новорожденных на наличие пяти заболеваний (адреногенитальный синдром, галактоземию, врожденный гипотиреоз, муковисцидоз, фенилкетонурию), с 2022 г. к обязательному списку добавлены еще три теста на спинальную мышечную атрофию и первичные нарушения иммунитета [20]. Начиная с 2021 г. Министерство здравоохранения РФ проводит пилотный проект по анализу 2000 генов экзома новорожденных [21]. Указом Президента Российской Федерации от 11 марта 2019 г. № 97 были установлены основы государственной политики Российской Федерации в области обеспечения химической и биологической безопасности на период до 2025 г., которые подразумевают “осуществление генетической паспортизации населения с учетом правовых основ защиты данных о персональном геноме человека и формирование генетического профиля населения”.

В Белоруссии также начата выдача генетических паспортов, работает специальная программа. Ведущей организацией является Институт генетики и цитологии Национальной академии наук. Проводится тестирование на генетическую предрасположенность к сердечно-сосудистым заболеваниям, диабету, остеопорозу, метаболическому синдрому, тестируются спортсмены для выявления благоприятных и неблагоприятных для спорта генетических особенностей, женщины с проблемами невынашивания беременности и т.д. Протестированы свыше 15 тыс. чел. [22].

На сегодня ближе всего практически к полной генетической паспортизации населения подошли в малых развитых странах. К таковым относится, в частности, Исландия, где по информации компании deCODE genetics были получены полные геномы более чем 160 тыс. жителей (из 365 тыс. общего населения) [23]. Островное государство представляет большой интерес для изучения геномики человеческих популяций, наследственных болезней и т.д. ввиду того, что предки-первопоселенцы большинства исландцев известны документально и в течение ряда веков наблюдалась фактически нулевая миграция на остров. Эти особенности позволили создать генеалогическую базу данных населения Iclendingabok (“Книга исландцев”) [24]. Проект был открыт в 2003 г., на нем выложены профили 904000 чел. Как считается, это половина людей, живших в Исландии с ее колонизации в IX–X веках. Доступ к базе данных для граждан Исландии возможен по индивидуальному ID, регистрацию на сайте прошло около двух третей населения. Одним из востребованных сервисов данной базы данных является возможность установить степень своего родства с любым из жителей страны.

Пользуясь тем, что большая часть исландского населения происходит от сравнительно небольшого круга общих предков, компания deCODE genetics успешно ведет исследования по наследственности сердечных заболеваний, диабета 2-го типа, болезни Альцгеймера, шизофрении и др. Сотрудничая с национальной медицинской системой, компания анализирует наследственные данные вместе с поведенческими и другими, влияющими на развитие врожденных склонностей к заболеваниям. Пользуясь широким доступом к вариативным данным, компания с 2002 г. активно участвовала в расшифровке генома человека, опубликовав, в частности, 5000 микросателлитных маркеров в привязке к хромосомам [25].

Высокая степень изученности населения позволяет выявить роль генотипа в возникновении или наличии предрасположенности к ряду заболеваний, но поднимает вопросы этического характера, к которым относится обнародование конфиденциальных вопросов родства, наследственной заболеваемости и т.п. [26]. Необходимо ли сообщать гражданам и их родственникам о риске заболевания неизлечимыми или тяжелыми заболеваниями (рак, болезнь Альцгеймера и т.д.)? Каковы могут быть механизм и правовые основы для такого информирования? Каким должен быть механизм доступа к информации о наследственных заболеваниях в базе данных и т.д. [27]?

3. ВЛИЯНИЕ НА СИСТЕМУ ЗДРАВООХРАНЕНИЯ

Накопление информации о геномах населения имеет важнейшее значение для развития персонализированной медицины, что может стать основным драйвером цифровизации генетической информации в ближайшие годы, а также приведет к более интенсивному развитию точной медицины и фармакогеномики, особенно в части влияния лекарственных препаратов на различные расы и этносы [28]. Исследования, опубликованные в последние годы, свидетельствовали, что, например, белые и афроамериканские популяции в США имеют различающиеся аллели, регулирующие биохимию олеиновой кислоты, которые по-разному реагируют с атенололом [29]. В этой связи стоит отметить ряд вызовов, стоящих перед глобальной системой здравоохранения, на настоящий момент ориентированной во многом на использование массовых лекарств, рынок которых может обвалиться с развитием точной медицины. Соответственно, прогнозируема малая заинтересованность текущих игроков в развитии “точной фармацевтики” в связи с отсут-

ствием быстрой экономической эффективности. При том, что в 2013 г. ряд фармацевтических компаний запустил исследования в области фармакогенетики [30], есть основания полагать, что гиганты фармацевтики по-прежнему настроены скептически по отношению к прибыльности массового производства “точечных препаратов”: их себестоимость трудно уменьшить ввиду малых производимых партий, особенно если в их рекомендациях пациентам будут учтены расовые, этнические, а также наследственные факторы. Исследователи сравнивают существующий интерес к фармакогенетике с первыми работами в области генной терапии в 1990-е гг.: надежды, возникшие во время ее зарождения, привели к буму инвесторов, но невозможность быстрого коммерческого успеха ввиду недостатка научных разработок привела к краху многих малых биотехнологических компаний, а также к выходу крупных фармацевтических холдингов из подобных проектов [31].

Необходима разработка новых подходов в медицинском страховании. Ключевой функцией точной медицины будет выявление тех пациентов, которые могут получить максимальную пользу от геномного тестирования – людей с редкими заболеваниями или различными формами рака, а затем предоставление отчета для страховой компании, учитывающего наиболее эффективную медикаментозную модель лечения. Специалисты называют это переходом “от диагностической одиссеи к диагностике”, рассчитывая на то, что правильная и быстрая постановка диагноза станет максимально возможной именно с учетом геномных данных [32]. Необходимы также изменения в подготовке медицинских кадров [33]. В учебные курсы требуется включить расширенные знания по эпигеномике, транскриптомике, протеомике и метаболомике. По мере быстрого развития технологий именно профессиональная подготовка врачей будет иметь первостепенное значение, возникнет необходимость в инновационных методах поддержки обучения медработников.

4. ПРИНЯТИЕ НАСЕЛЕНИЕМ ПРОГРАММ ГЕНЕТИЧЕСКОЙ ПАСПОРТИЗАЦИИ

За исключением программ генетической паспортизации преступников ключевым фактором дальнейшего роста генетических баз данных населения является согласие населения на проведение генетического анализа и сохранение полученной информации в базах данных. Тем не менее существуют факторы, которые могут затормозить дальнейший рост объемов генетической информации о населении [34]. Примером этого могут быть изменения на рынке генетического тестирования населения в сегменте ditrect-to-consumer. В 2016–2019 гг. число клиентов компаний, предлагающих генетическое тестирование в основном с целью выявления этнического происхождения, росло экспоненциальными темпами. Например, между 2018 и 2019 г. количество клиентов компаний Ancestry, 23andMe и других выросло на 117% [35]. Тем не менее по состоянию на февраль 2020 г. отмечалось замедление роста числа клиентов коммерческих компаний, предлагающих генетическое тестирование населения [36]. Предполагаемыми причинами этого могут быть насыщение рынка, сомнения населения в конфиденциальности тестирования и несоответствие ожидаемых результатов. Насыщение рынка связано с тем, что желающие определить свое происхождение уже получили результаты, тогда как их ближайшие родственники не заинтересованы в тестировании, так как уже получили результат. Также необходимо учитывать, что точность генетического тестирования зависит от базы, с которой проводится сравнение результатов. На ранних этапах накопления информации это приводит к более низкой точности тестирования, что в свою очередь может привести к разочарованию от исследования. Наиболее острой проблемой являются сомнения населения и государственных организаций в конфиденциальности результатов тестирования. Ряд случаев показывает, что повышение уровня конфиденциальности генетических данных должно стать одним из важнейших приоритетов в развитии генетических технологий.

В декабре 2019 г. Министерство обороны США разослало американским военным рекомендации не проходить генетическое тестирование в коммерческих компаниях и предупреждение, что генетическое тестирование увеличивает риски для военнослужащих: “Exposing sensitive genetic information to outside parties poses personal and operational risks to Service members” [37]. В Великобритании в 2007 г. были осуждены пятеро сотрудников организации Forensic Science Service, поддерживающей национальную британскую базу данных ДНК населения, которые украли данные для того, чтобы создать конкурирующую компанию по генетической идентификации населения [38]. Юрист из Великобритании потеряла работу из-за того, что информация о нахождении ее ДНК в базе данных преступников стала доступной ее работодателю. При этом сама она провела под арестом 24 ч и была признана невиновной [39]. Эффективность британской базы данных также вызывает вопросы исследователей. Обходясь более чем в 2.5 млн фунтов стерлингов в год, данная база позволяет отслеживать лишь серийных убийц или насильников, которые при эффективной работе правоохранительных органов и без генетической информации находятся под постоянным контролем или в местах заключения. Кроме того, сконцентрировавшись лишь на гражданах Великобритании, причем с определенными криминальными наклонностями, база становится бессильна в слу-

чае совершения преступлений гражданами других стран и тем более нелегальных мигрантов [40]. Компания GEDmatch, изначально предоставляющая услуги поиска родственников или биологических родителей усыновленных детей, в 2018 г. после объявления об обнаружении серийного убийцы через базу данных компании столкнулась с массовым закрытием профилей пользователей от возможности их использования для поиска преступников. В связи с этим компания была вынуждена изменить условия предоставления информации правоохранительным органам и впоследствии была куплена компанией, специализирующейся на ДНК-криминалистике [41].

Компания 23andMe в 2018 г. была вынуждена ужесточить доступ к данным для сторонних компаний, к которым относятся разработчики программного обеспечения для мониторинга здоровья, снижения веса и т.д. из-за опасений населения в конфиденциальности полученных результатов. Утверждается, что сейчас доступ для сторонних организаций предоставляется по более сложной процедуре [42]. При этом вопрос генотипирования человека часто являлся предметом законодательных и судебных разбирательств [43]. Кроме того, исследователи подвергают тесты компании критике за провоцирование у клиентов фобий на основе недоказанной расположенности к тем или иным смертельным заболеваниям (например, выводы делаются на основе пары мутаций гена, а не взаимодействий генной сети) [44]. Официально компания не предоставляла данных тестирования спецслужбам и правоохранительным органам, однако подобное положение содержится в пользовательском соглашении “если такое раскрытие является разумно необходимым”. На 2019 г. компания собрала генетическую базу 9 млн чел., утверждая, что дальнейшему продвижению мешает законодательное регулирование со стороны FDA [45].

Негативное восприятие генетической паспортизации широкими массами населения может стать препятствием на пути развития генетических баз данных. По-видимому, ближайшим аналогом подобного поведения может являться реакция жителей различных стран на вакцинацию в период пандемии COVID-2019. Этот пример становится максимально очевидным как ввиду широкого распространения заболевания, так и ввиду того, что необходимость вакцинации активно подчеркивается правительственными и общественными институтами. Нерешительность в отношении вакцинации (Vaccine hesitancy), определяемая WHO как “delay in acceptance or refusal of vaccination despite availability of vaccination services”, была объявлена WHO в 2019 г. как одна из десяти важнейших угроз здоровью населения [46]. По состоянию на 2021 г. готовность вакцинироваться в среднем составила 75.2% в 23 исследованных странах, что на 3.7% выше, чем в 2020 г. [47]. Тем не менее значительная доля населения проявляет нерешительность в отношении вакцинации и может проявлять аналогичное отношение к генетической паспортизации. Введение сертификатов о вакцинации от COVID-19 и различная степень свободы передвижения в зависимости от наличия вакцинации также поднимают вопрос об этических аспектах введения генетических паспортов. Одним из потенциальных вопросов может быть теоретическое наличие у человека генетически обусловленного иммунитета к какому-либо заболеванию, что может привести к большей свободе передвижения и большим возможностям у человека с наличием таких генов [48]. Доступность информации о наличии у человека “плохих” и “хороших” генов справедливо вызывает обеспокоенность у населения, так как это может проявиться в создании неравноценных условий при приеме на работу, тарифов на медицинское страхование и т.д. [49]. Поэтому принятие любых решений об использовании генетических технологий применительно к человеку должно принимать во внимание его интересы, результаты научных исследований и этические нормы [50]. Необходим поиск оптимального баланса между требованиями конфиденциальности персональных данных, необходимости проведения научных исследований, развития технологий персонализированной медицины и разработки бизнес-модели, способной поддержать в будущем экспоненциальный рост генетических данных [51, 52].

5. АНАЛИЗ ГЕНЕТИЧЕСКИХ ДАННЫХ

Необходимость обработки больших объемов генетических данных с помощью современных цифровых технологий стала стимулом к сотрудничеству информационных гигантов с генетическими программами и компаниями. Так, Институт Броуда (Broad Institute of MIT and Harvard) еще в 2015 г. разработал совместно с Google Genomics целую программу взаимодействия по облачному доступу к генетической информации, в том числе для сторонних исследователей. В свою очередь, дочерняя компания Google, получившая более широкое наименование Cloud Life Sciences, предлагает сегодня широкие возможности своих серверов для биоинформатиков в области хранения и обработки данных. Не отстают и конкуренты: компания Oracle Life Science создала широкие возможности для передачи и хранения такой дискретной информации, как результаты клинических исследований, включая хранение данных пациентов; к числу клиентов относятся Oyster Point Pharma и Pfizer. Соответственно, есть подобный сервис и у Microsoft Azure, предлагающий “ускорение обработки данных геномных анали-

зов, точной медицины, клинических испытаний" путем предоставления широкоформатных хранилищ, скоростной обработки данных и поддержки биоинформатиков. Происходит формирование цифровых экосистем генетической информации, что может привести к монополизации доступа к подобной информации по примеру текущей монополизации социальных сетей и других цифровых платформ.

Несмотря на впечатляющие темпы роста, процесс цифровизации генетической информации находится на начальном этапе своего развития. В настоящее время происходит ускоренный рост баз данных, характерный для этапа накопления информации. Тогда как развитие средств анализа информации в базах данных и принятие решений на ее основе развиваются с меньшей скоростью [53]. Основным ограничителем является недостаток сопутствующей информации об исследуемых организмах, например наличие заболеваний, биохимические параметры крови и т.д. В итоге системы анализа данных на основе искусственного интеллекта (**ИИ**) могут успешно анализировать генетическую информацию, но правильно интерпретировать ее могут только в отдельных случаях [54, 55]. Вторым ограничивающим фактором является необходимость стандартизации и проверки рекомендаций, сделанных ИИ. Например, в области медицины экспертам необходимо понимать, каким образом ИИ пришел к решению задачи (explainable AI) и не является ли это решение ошибкой, которая может привести к гибели пациента [56]. В ряде стран разрабатываются процедуры, которые регулировали бы применение методов ИИ в медицине. Например, Управлением по санитарному надзору за качеством пищевых продуктов и медикаментов США в январе 2021 г. были опубликованы план действий и требования к разработчикам подобных технологий [57]. При дальнейшем развитии методов ИИ и расширении баз данных биологической информации можно ожидать ускоренного развития исследований в этой области и выходу их на принципиально новый уровень.

6. БЕЗОПАСНОСТЬ ДАННЫХ

Широкий доступ к базам данных генетической информации и результатам научных исследований приводит к массовому развитию любительских исследований в области молекулярной биологии ("биохакинг"). Примерами таких исследований являются разработки любительских вакцин от SARS-CoV-2 [58], лекарств от старения [59] и т.д. Такое развитие событий вызывает обеспокоенность у экспертов, так как при негативном сценарии это может привести к созданию более совершенных видов биологического оружия и появлению нового типа "кибербиологических атак", в ходе которых может быть взломана инфраструктура синтеза нуклеотидных последовательностей, и компоненты для синтеза возбудителей болезней могут попасть в руки террористов [60]. Не могут не вызывать тревогу населения и научного сообщества публикуемые работы об экспериментальной модификации опасных вирусов в поисках контагиозных форм для отслеживания вероятных природных мутаций в этом направлении [61].

В результате параллельно с увеличением объема цифровых генетических данных растет объем закрытых данных, когда генетические данные в публикациях больше не выкладываются в открытый доступ. И речь идет не только о вирусах и иных патогенах, но и о сельхозкультурах. Так, представляющая большой научный интерес публикация международного коллектива авторов, финансируемых из фондов Китая, не содержит никаких открытых ссылок на геномные данные [62].

Встает вопрос об организационной форме крупных баз данных генетической информации, содержащих чувствительную информацию. Например, во Франции подобная структура (France Genomique) инициирована государством на базе национальных институтов здравоохранения и медицинских исследований, сельскохозяйственных исследований, научных исследований. Головной и объединяющей организацией с января 2019 г. является CEA – Комиссариат атомной и альтернативной энергий, созданный Ш. де Голлем еще в 1945 г. Существование подобного консорциума под руководством организации, отвечавшей десятилетиями за французский атомный проект, обусловлено ее управленческими компетенциями, вычислительными ресурсами, а главное, необходимостью концентрации такого важного национального проекта, как геномные исследования, в одном центре [63].

7. ПРИРОДНЫЕ БИОЛОГИЧЕСКИЕ РЕСУРСЫ И ЦИФРОВИЗАЦИЯ

Цифровизация генетической информации природных экосистем и организмов сыграет важную роль в разработке мер по сохранению исчезающих видов организмов, изучении природного биоразнообразия, выявлении новых патогенных организмов и послужит источником информации о новых генетических последовательностях с потенциалом использования в биотехнологии, сельском хозяйстве и медицине. В этом направлении важным аспектом является детальная характеризация существующих биоресурсных коллекций микроорганизмов, растений и животных. Цифровизация биоресурсных коллекций уже является существенным драйвером роста генетической информации в мировом масштабе. За ис-

ключением работы с патогенными микроорганизмами, подобные исследования не вызывают противодействия широких слоев населения. Напротив, речь идет о сохранении биоразнообразия, сохранении исчезающих видов, автохтонных пород животных или сортов сельскохозяйственных культур, что гарантирует этому процессу широкую общественную и государственную поддержку.

Также детальная характеризация биоресурсных коллекций играет важную роль для экономики и повышения устойчивости местных сообществ в условиях климатических изменений. Показателен пример работы Института биоразнообразия Эфиопии (речь идет об одной из беднейших стран мира) [64]. Проект финансируется Глобальным экологическим фондом (GEF) ООН и Всемирного банка, он стартовал в 1994 г. и сфокусирован на разнообразии местных культур (зерновых, кофе, фруктов, лекарственных растений), возделываемых традиционным способом малыми фермерским хозяйствами. Сохраняемые в коллекции культуры генотипируются, сохраняясь при этом в виде живых культивируемых коллекций, локализованных в шести агроэкологических регионах. Для каждого региона были сформированы ассоциации фермеров-питомниководов, изучены и задокументированы местные знания фермеров о своих культурных сортах, методах их культивации и переработки, народной селекции.

Национальные генетические банки сельскохозяйственных и диких культур существуют и в ряде других стран, в Германии, Канаде, Бразилии, действует и пополняется Национальный банк генетических ресурсов хозяйственно-полезных растений Республики Беларусь. Подобная программа появляется и в России: 30 марта 2023 г. в Государственную Думу внесен законопроект “О биоресурсных центрах и биологических (биоресурсных) коллекциях”, вводящий, в частности, понятия генетических ресурсов, их национальных каталогов и коллекций, в том числе, в целях регулирования получаемой цифровой генетической информации [65]. Учитывая, что речь идет об одних из крупнейших в мире биоресурсных коллекций, таких как Всероссийская коллекция промышленных микророганизмов НИЦ “Курчатовский институт” (более 20000 образцов) и Коллекции генетических ресурсов растений Всероссийского института генетических ресурсов растений им. Н.И. Вавилова (более 320000 образцов) [66], Россию ожидает взрывной рост цифровизации генетической информации, что потенциально должно сделать НБГИ в перспективе до 2030 г. динамично развивающейся базой генетической информации.

ЗАКЛЮЧЕНИЕ

В настоящее время процесс цифровизации биологической информации находится на ранних этапах развития, характеризующихся ускоренным экспоненциальным ростом баз данных, что открывает широкие возможности для развития персонализированной медицины, сохранения биоразнообразия, развития биотехнологий и сельского хозяйства. Насущным вопросом является подключение технологий ИИ к анализу биологических данных, хотя его эффективность доказана только в отдельных направлениях, где наблюдается хорошее структурирование данных. Наличие генетической и биологической информации в цифровом виде и широкий доступ к ней открывают новые горизонты развития синтетической биологии, но в то же время дают возможности для разработки технологий двойного назначения и биологического терроризма. Ввиду этого необходима разработка механизмов контроля и регламентации развития цифровых генетических и биологических баз данных с широким привлечением научного сообщества и бизнеса. Насущным является разработка стандартов безопасного использования цифровых генетических данных, а также разработка механизмов государственного и межгосударственного контроля в этой области, возможно, аналогичного Международному агентству по атомной энергии (МАГАТЭ). В этой связи важнейшей задачей становится поиск оптимального баланса между требованиями конфиденциальности персональных данных населения, соблюдением правовых и этических норм, необходимостью проведения научных исследований, развитием технологий персонализированной медицины и разработкой бизнес-моделей и организационно-правовых форм, способных сопровождать в дальнейшем рост объемов генетических данных.

Работа выполнена при поддержке НИЦ “Курчатовский институт” (приказ № 91 от 20 января 2023 г.).

СПИСОК ЛИТЕРАТУРЫ

1. https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data
2. https://nebula.org/whole-genome-sequencing/
3. *Sayers E.W., Cavanaugh M., Clark K. et al.* // Nucleic Acids Res. 2021. V. 49. № D1. P. D92. https://doi.org/10.1093/nar/gkaa1023
4. https://www.idc.com/getdoc.jsp?containerId=prUS47560321
5. *Gillings M.R., Hilbert M., Kemp D.J.* // Trends Ecol. Evol. 2016. V. 31. № 3. P. 180. https://doi.org/10.1016/j.tree.2015.12.013
6. *Rigden D.J., Fernández X.M.* // Nucleic Acids Res. 2022. V. 50. № D1. P. D. https://doi.org/10.1093/nar/gkab1195

7. *Katz K., Shutov O., Lapoint R. et al.* // Nucleic Acids Res. 2022. V. 50. № D1. P. D387. https://doi.org/10.1093/nar/gkab1053
8. CNCB-NGDC Members and Partners // Nucleic Acids Res. 2022. V. 50. № D1. P. D27. https://academic.oup.com/nar/article/50/D1/D27/6413834
9. https://ngdc.cncb.ac.cn/databasecommons
10. *Bernotaite A.* // Forensic DNA Typing: Principles, Applications and Advancements. Singapore: Springer, 2020. P. 639. https://doi.org/10.1007/978-981-15-6655-4_33
11. https://www.aspi.org.au/report/genomic-surveillance
12. https://theconversation.com/scientists-are-on-a-path-to-sequencing-1-million-human-genomes-and-use-big-data-to-unlock-genetic-secrets-157210
13. https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/news/whole-genome-sequencing-data-on-200-000-uk-biobank-participants-available-now
14. https://www.statewatch.org/news/2022/april/eu-policing-france-proposes-massive-eu-wide-dna-sweep-automated-exchange-of-facial-images/
15. https://www.statewatch.org/news/2022/april/eu-policing-france-proposes-massive-eu-wide-dna-sweep-automated-exchange-of-facial-images/
16. https://www.interpol.int/How-we-work/Forensics/DNA
17. https://www.technologyreview.com/2019/02/11/103446/more-than-26-million-people-have-taken-an-at-home-ancestry-test/
18. https://www.blackstone.com/news/press/blackstone-to-acquire-ancestry-leading-online-family-history-business-for-4-7-billion/
19. *Kovanda A., Zimani A.N., Peterlin B.* // Hum. Genomics. 2021. V. 15. P. 20. https://doi.org/10.1186/s40246-021-00315-6
20. Приказ МЗ РФ № 185 от 22.03.2006 “О массовом обследовании новорожденных детей на наследственные заболевания”
21. https://exome.ncagp.ru/
22. https://genpasport.igc.by/news#price170820
23. https://www.decode.com/research
24. https://www.islendingabok.is/
25. *Kong A., Gudbjartsson D., Sainz J. et al.* // Nat. Genet. 2002. V. 31. P. 241. https://doi.org/10.1038/ng917
26. *Pálsson G.* // Kinship and Beyond: The Genealogical Model Reconsidered. Oxford: Berghahn Books, 2009. P. 84.
27. https://www.seattletimes.com/nation-world/iceland-faces-a-dna-dilemma-whether-to-notify-people-carrying-cancer-genes/
28. *Bloche M.G.* // New Eng. J. Med. V. 351. № 20. P. 2035. https://doi.org/10.1056/NEJMp048271
29. *Wikoff W.R., Frye R.F., Zhu H. et al.* // PLoS One. 2013. V. 8. № 3. P. e57639. https://doi.org/10.1371/journal.pone.0057639
30. *Hedgecoe A., Martin P.* // Soc. Stud. Sci. 2003. V. 33. P. 327. https://doi.org/10.1177/03063127030333002
31. *Corrigan O.P.* // J. Med. Ethics. 2005. V. 31. № 3. P. 144. https://doi.org/10.1136/jme.2004.007229
32. https://pharmaceutical-journal.com/article/feature/genomic-medicine-is-going-mainstream-and-pharmacists-need-to-be-prepared
33. *Hicks J.K., Aquilante C.L., Dunnenberger H.M. et al.* // J. Am. Coll. Clin. Pharm. 2019. V. 2. № 3. P. 303. https://doi.org/10.1002/jac5.1118
34. *Hazel J.W., Hammack-Aviran C., Brelsford K.M. et al.* // PLoS One. 2021. V. 16. № 11. P. e0260340. https://doi.org/10.1371/journal.pone.0260853
35. https://www.technologyreview.com/2019/02/11/103446/more-than-26-million-people-have-taken-an-at-home-ancestry-test/]
36. https://www.vox.com/recode/2020/2/13/21129177/consumer-dna-tests-23andme-ancestry-sales-decline
37. https://www.yahoo.com/news/pentagon-warns-military-members-dna-kits-pose-personal-and-operational-risks-173304318.html
38. https://www.standard.co.uk/hp/front/five-civil-servants-suspended-over-dna-espionage-7179521.html
39. https://www.bbc.co.uk/news/10278173
40. *Amankwaa A.O., McCartney C.* // Forensic Sci. Int.: Synergy. 2019. V. 1. P. 45. https://doi.org/10.1016/j.fsisyn.2019.03.004
41. *Guerrini C.J., Wickenheiser R.A., Bettinger B. et al.* // J. Law. Biosci. 2021. Apr. V. 8. № 1. P. lsab001. https://doi.org/10.1093/jlb/lsab001
42. https://www.cnbc.com/2018/08/23/23andme-is-telling-developers-that-it-plans-to-shut-off-its-api-in-two-weeks.html
43. https://www.nytimes.com/2010/06/12/health/12genome.html
44. https://www.bbc.com/news/health-50069155
45. https://www.technologyreview.com/2019/02/11/103446/more-than-26-million-people-have-taken-an-at-home-ancestry-test/
46. https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019
47. *Lazarus J.V., Wyka K., White T.M. et al.* // Nat. Commun. 2022. V. 13. № 1. P. 3801. https://doi.org/10.1038/s41467-022-31441-x
48. *Gyngell C., Savulescu J.* // J. Med. Ethics. 2022. V. 48. P. 689. https://doi.org/10.1136/medethics-2021-107297
49. *Henneman L., Vermeulen E., Van E.L. et al.* // Eur. J. Hum. Gen. 2013. V. 21. № 8. P. 793.
50. *Lehmann L.S., Sulmasy L.S., Burke W. et al.* // Ann. Intern. Med. 2022. V. 175. P. 1322. https://doi.org/10.7326/M22-0743
51. *Wan Z., Hazel J.W., Clayton E.W. et al.* // Nat. Rev. Genet. 2022. V. 23. P. 429. https://doi.org/10.1038/s41576-022-00455-y
52. *Thiebes S., Toussaint P.A., Ju J. et al.* // J. Med. Internet Res. 2020. V. 22. № 1. P. e14890. https://doi.org/10.2196/14890
53. *Koumakis L.* // Comput. Struct. Biotechnol. J. 2020. V. 18. P. 1466. https://doi.org/10.1016/j.csbj.2020.06.017

54. *Patel N.M., Michelini V.V., Snell J.M. et al.* // Oncologist. 2018. V. 23. № 2. P. 179. https://doi.org/10.1634/theoncologist.2017-0170

55. https://www.technologynetworks.com/informatics/articles/the-hype-of-watson-why-hasnt-ai-taken-over-oncology-333571

56. *Amann J., Blasimme A., Vayena E. et al.* // BMC Med. Inform. Decis. Mak. 2020. V. 20. P. 310. https://doi.org/10.1186/s12911-020-01332-6

57. https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device

58. https://www.bloomberg.com/news/articles/2020-10-10/home-made-covid-vaccine-appeared-to-work-but-questions-remained

59. https://www.technologyreview.com/2015/10/14/165802/a-tale-of-do-it-yourself-gene-therapy/

60. *Puzis R., Farbiash D., Brodt O. et al.* // Nat. Biotechnol. 2020. V. 38. № 12. P. 1379. https://doi.org/10.1038/s41587-020-00761-y

61. *Kaiser J.* // Science. 2022. V. 378. № 6617. https://doi.org/10.1126/science.adf3762

62. *Dong Y., Duan S., Xia Q. et al.* // Science. 2023. V. 379. № 6635. P. 892. https://doi.org/10.1126/science.add8655

63. https://www.france-genomique.org/france-genomique/

64. https://www.ebi.gov.et/biodiversity/conservation/fgbs/

65. https://sozd.duma.gov.ru/bill/325647-8

66. https://regnum.ru/news/innovatio/2020382.html