



## Российская база данных устойчивости ВИЧ к антиретровирусным препаратам

Киреев Д.Е.<sup>✉</sup>, Кириченко А.А., Лопатухин А.Э., Шлыкова А.В., Галкин Н.Ю., Савельер Е.В., Глазов М.Б., Покровский В.В., Акимкин В.Г.

Центральный научно-исследовательский институт эпидемиологии Роспотребнадзора, Москва, Россия

### Аннотация

Развитие технологий секвенирования и биоинформатического анализа дало возможность проведения молекулярно-эпидемиологических исследований, в которых нуклеотидные последовательности вируса иммунодефицита человека (ВИЧ) используются в качестве дополнительной характеристики пациента. При этом наиболее значимым с практической точки зрения направлением работ является изучение распространения лекарственной устойчивости (ЛУ) ВИЧ. В различных странах для организации таких исследований применяются базы данных, являющиеся хранилищами генетической и эпидемиологической информации. Российская база данных устойчивости ВИЧ к антиретровирусным препаратам (<https://hivresist.ru/>) была создана в 2009 г. Тем не менее длительное время её применение оставалось ограниченным. С 2021 г. после обновления нормативных документов внесение результатов исследований ЛУ ВИЧ в российскую базу данных устойчивости ВИЧ к антиретровирусным препаратам стало обязательным. В связи с этим были проведены работы по усовершенствованию базы данных и увеличению её функциональных возможностей. Были разработаны различные способы внесения клиничко-эпидемиологических и генетических данных. На момент написания публикации российская база данных ЛУ ВИЧ содержала 10 626 уникальных записей о пациентах и 13 126 нуклеотидных последовательностей, загруженных 10 учреждениями. Для анализа данных были разработаны следующие функции: контроль качества эпидемиологической и клинической информации о пациенте, контроль качества нуклеотидных последовательностей, проверка на контаминацию, субтипирование, выявление мутаций ЛУ, определение вирусного тропизма и генерация стандартизированных отчётов. В планах по дальнейшему развитию российской базы данных ЛУ ВИЧ — разработка инструмента для выявления и анализа молекулярных кластеров и адаптация для рутинного использования в рамках эпидемиологического надзора за ВИЧ-инфекцией.

**Ключевые слова:** ВИЧ, лекарственная устойчивость, база данных, нуклеотидная последовательность, контроль качества, субтипирование, молекулярные кластеры, молекулярная эпидемиология

**Благодарности.** Авторы выражают благодарность Сергею Владимировичу Штреку, заведующему лабораторией Омского НИИ природно-очаговых инфекций Сибирского федерального округа окружного центра по профилактике и борьбе со СПИД; Наталье Николаевне Зайцевой, директору, и Ольге Юрьевне Пекшевой, заведующей лабораторией молекулярно-генетических и серологических методов исследования Нижегородского НИИ эпидемиологии и микробиологии им. академика И.Н. Блохиной; Михаилу Валерьевичу Питерскому, руководителю Уральского окружного центра по профилактике и борьбе со СПИД Екатеринбургского НИИ вирусных инфекций; Светлане Владимировне Яценко, заведующей лабораторией бактериологических и молекулярно-генетических исследований Красноярского краевого центра профилактики и борьбы со СПИД; Наталье Владимировне Филонюк, заведующей клиничко-иммунологической лабораторией Липецкого областного центра по профилактике и борьбе со СПИД и инфекционными заболеваниями; Андрею Борисовичу Шемшуре, врачу клинической лабораторной диагностики Клинического центра профилактики и борьбы со СПИД Министерства здравоохранения Краснодарского края; Дмитрию Сергеевичу Колпакову, заведующему клиничко-диагностической лабораторией Ростовского НИИ микробиологии и паразитологии; Марине Ридовне Бобковой, заведующей лабораторией вирусов лейкозов НИЦЭМ им. Н.Ф. Гамалеи; Валерии Олеговне Котовой, заведующей лабораторией эпидемиологии и профилактики вирусных гепатитов и СПИДа Хабаровского НИИ эпидемиологии и микробиологии за внесение молекулярной, эпидемиологической и клинической информации в Российскую базу данных устойчивости ВИЧ к антиретровирусным препаратам.

**Источник финансирования.** Работа выполнена в рамках внутреннего гранта ЦНИИ Эпидемиологии Роспотребнадзора на фундаментальное исследование «Развитие Российской базы данных лекарственной устойчивости ВИЧ-1 к антиретровирусным препаратам».

**Конфликт интересов.** Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

**Для цитирования:** Киреев Д.Е., Кириченко А.А., Лопатухин А.Э., Шлыкова А.В., Галкин Н.Ю., Савельер Е.В., Глазов М.Б., Покровский В.В., Акимкин В.Г. Российская база данных устойчивости ВИЧ к антиретровирусным препаратам. *Журнал микробиологии, эпидемиологии и иммунобиологии*. 2023;100(2):219–227. DOI: <https://doi.org/10.36233/0372-9311-345> EDN: <https://www.elibrary.ru/maebdg>

Review

<https://doi.org/10.36233/0372-9311-345>

## The Russian database of HIV antiretroviral drug resistance

Dmitry E. Kireev<sup>✉</sup>, Alina A. Kirichenko, Alexey E. Lopatukhin, Anastasia V. Shlykova, Nikolai Yu. Galkin, Evgenii V. Saveler, Maxim B. Glazov, Vadim V. Pokrovsky, Vasily G. Akimkin

Central Research Institute for Epidemiology, Moscow, Russia

### Abstract

The development of sequencing technologies and bioinformatic analysis made it possible to conduct molecular and epidemiological studies, in which nucleotide sequences of the human immunodeficiency virus (HIV) are used as information added to the patient profile. From a practical perspective, studies of prevalence of HIV drug resistance (HIVDR) are of the highest significance. To promote such studies, different countries use databases that serve as repositories of genetic and epidemiological information. The Russian HIVDR database (<https://hivresist.ru/>) was created in 2009. Nevertheless, it was characterized by limited applicability for a long time. Since 2021, after the regulatory documents had been revised and updated, the entry of HIVDR research results into the Russian HIVDR database has been mandatory. Therefore, the priority attention has been given to upgrading the database and improving its functional capabilities. Different methods have been developed to enter clinical, epidemiological and genetic data. At the time of this study, the Russian database HIVDR contained 10,626 unique records about patients and 13,126 nucleotide sequences deposited by 10 institutions. The following functions have been provided for data analysis: quality control of the epidemiological and clinical information about a patient, quality control of nucleotide sequences, contamination check, subtyping, detection of DR mutations, identification of viral tropism and generation of standardized reports. The efforts toward further development of the Russian HIVDR database will be focused on designing tools for detection and analysis of molecular clusters, adaptation to routine application for epidemiological surveillance of HIV infection.

**Keywords:** HIV, drug resistance, database, nucleotide sequence, quality control, subtyping, molecular clusters, molecular epidemiology

**Acknowledgements.** The authors express their gratitude to Sergey Vladimirovich Shtrek, head of the laboratory at the Omsk Research Institute of Natural and Focal Infections of the Siberian Federal District Center for AIDS Prevention and Control; Natalia Nikolaevna Zaitseva, director, and Olga Yurievna Peksheva, head of the laboratory of molecular, genetic, and serological research methods at the Blokhnina Nizhny Novgorod Research Institute of Epidemiology and Microbiology; Mikhail Valerievich Pitersky, head of the Ural District Center for AIDS Prevention and Control at the Ekaterinburg Research Institute of Viral Infections; Svetlana Vladimirovna Yaschenko, head of the laboratory of bacteriological, molecular, and genetic studies at the Krasnoyarsk Regional Center for AIDS Prevention and Control; Natalia Vladimirovna Filonyuk, head of the laboratory of clinical immunology at the Lipetsk Regional Center for Prevention and Control of AIDS and Infectious Diseases; Andrei Borisovich Shemshura, a doctor of clinical laboratory diagnostics at the Clinical Center for AIDS Prevention and Control at the Ministry of Health of the Krasnodar Territory; Dmitry Sergeyevich Kolpakov, head of the clinical and diagnostic laboratory at the Rostov Research Institute of Microbiology and Parasitology; Marina Ridovna Bobkova, head of the laboratory of leukemia viruses at the Gamaleya National Research Center of Epidemiology and Microbiology, Valeria Olegovna Kotova, head of the laboratory of epidemiology and prevention of viral hepatitis and AIDS at the Khabarovsk Research Institute of Epidemiology and Microbiology for depositing molecular, epidemiological and clinical information to the Russian HIVDR database.

**Funding source.** The study was funded through the inhouse grant of the Central Research Institute of Epidemiology of the Federal Service for Surveillance on Consumer Rights Protection and Human Wellbeing (Rosпотребнадзор) for the fundamental research "Development of the Russian HIV-1 antiretroviral drug resistance database".

**Conflict of interest.** The authors declare no apparent or potential conflicts of interest related to the publication of this article.

**For citation:** Kireev D.E., Kirichenko A.A., Lopatukhin A.E., Shlykova A.V., Galkin N.Yu., Saveler E.V., Glazov M.B., Pokrovsky V.V., Akimkin V.G. The Russian database of HIV antiretroviral drug resistance. *Journal of microbiology, epidemiology and immunobiology = Zhurnal mikrobiologii, epidemiologii i immunobiologii*. 2023;100(2):219–227. DOI: <https://doi.org/10.36233/0372-9311-345> EDN: <https://www.elibrary.ru/maebdg>

### Введение

Молекулярно-генетические технологии стали незаменимым инструментом в диагностике, эпидемиологии и лечении инфекционных заболеваний. Методы определения нуклеотидных последовательностей (НП) вируса иммунодефицита человека (ВИЧ) более 20 лет используются для определения

лекарственной устойчивости (ЛУ) вируса с целью оценки эффективности антиретровирусных (АРВ) препаратов. В связи с увеличением количества применяемых препаратов и количества пациентов, принимающих АРВ-терапию, значимость этого анализа только увеличивается и количество проводимых исследований возрастает. Однако промежуточный

результат исследования ЛУ — НП вируса — может быть использован также в эпидемиологии с целью изучения особенностей распространения вируса. В данном случае информация о вирусе представляет собой дополнительный параметр, который анализируется совместно с эпидемиологическими, клиническими и демографическими показателями ВИЧ-инфицированного человека. Являясь объективной характеристикой, НП особенно полезна при изучении заболеваний, характеризующихся сложностью диагностирования, хроническим течением и высоким уровнем стигматизации.

В связи с этим хранение молекулярных данных и сопутствующей информации о пациенте в сочетании с их последующим анализом является важным направлением работ в области эпидемиологии ВИЧ-инфекции. Существует ряд непрерывно пополняемых национальных и региональных баз данных НП ВИЧ. Наиболее полные национальные базы данных Швейцарии [1] и Великобритании [2] содержат информацию о более чем 50% ВИЧ-инфицированных граждан этих стран. Основным источником пополнения этих баз генетическими данными о вирусе являются выполняемые в рамках рутинного оказания медицинской помощи ВИЧ-инфицированным исследования ЛУ ВИЧ.

Темпы развития эпидемии ВИЧ-инфекции в России значительно превосходят таковые в развитых странах. В 2021 г. показатели заболеваемости в стране составили 48,7 на 100 тыс. населения, а пораженности — 782,0 на 100 тыс. населения. К концу 2021 г. в России проживало 1 137 596 человек с ВИЧ, а более 660 000 принимали АРВ-препараты. При этом ежегодно более 15% пациентов, принимающих АРВ-препараты, не достигают недетектируемой вирусной нагрузки<sup>1</sup>. А согласно стандарту оказания медицинской помощи лицам, больным ВИЧ-инфекцией, тест на резистентность ВИЧ может быть проведен каждому десятому пациенту 2 раза в год<sup>2</sup>.

Такой масштаб эпидемии и нормативные документы определяют ежегодную потребность более чем в 100 тыс. исследований на ЛУ ВИЧ в нашей стране. Однако вследствие дороговизны, трудоёмкости исследования и недостаточной оснащённости лабораторий центров по профилактике и борьбе со СПИДом и инфекционными заболеваниями (центры СПИД) выполняется менее 10% требуемого количества. В 2020 и 2021 гг. было закуплено набо-

ров реагентов для проведения анализа на ЛУ ВИЧ в количестве 7438 и 7232 штуки соответственно<sup>3</sup>. Таким образом, в силу объективных причин при сохранении финансового обеспечения центров СПИД и тех же технологий секвенирования в России невозможно добиться необходимого охвата ВИЧ-позитивных лиц исследованиями на ЛУ. Однако даже немногочисленные результаты анализов резистентности ВИЧ часто остаются недоступными для дальнейших эпидемиологических исследований, поскольку отсутствует инструмент для сбора, хранения и анализа такой информации.

Для минимизации потерь генетических данных о вирусе и осуществления мониторинга ЛУ ВИЧ в 2009 г. в ЦНИИ Эпидемиологии Роспотребнадзора была создана Российская база данных устойчивости ВИЧ к АРВ-препаратам (РБД<sup>4</sup>). Использование РБД регламентировалось сначала методическими рекомендациями от 2013 г.<sup>5</sup>, а затем методическими указаниями от 2016 г.<sup>6</sup> С 2021 г. согласно обновлённым Санитарным правилам и нормам 3.3686-21<sup>7</sup> всем центрам СПИД следует загружать НП, полученные в результате выполнения исследований ЛУ ВИЧ, и сопутствующую деперсонифицированную информацию о пациентах в РБД.

Сбор генетической информации о вирусе и последующий биоинформатический анализ собранной информации могут повысить точность эпидемиологического надзора за ВИЧ-инфекцией, улучшить качество оказания медицинской помощи и в конечном итоге снизить количество новых случаев заболевания. Развитие методов анализа потребовало совершенствования РБД.

**Цель работы** — описание функциональных возможностей Российской базы данных устойчивости ВИЧ к АРВ-препаратам в отношении анализа НП и сопутствующей информации о пациентах.

## Языки программирования и биоинформатические методы

Внутренняя часть РБД, доступная только зарегистрированным пользователям, реализована на

<sup>1</sup> ВИЧ-инфекция в Российской Федерации на 31.12.2021. Справка СНИОП по профилактике и борьбе со СПИДом ФБУН Центральный НИИ Эпидемиологии Роспотребнадзора. URL: <http://www.hivruussia.info/wp-content/uploads/2022/03/Spravka-VICH-v-Rossii-na-31.12.2021-g..pdf>

<sup>2</sup> Приказ Министерства здравоохранения РФ № 438н от 23.06.2022 «Об утверждении стандарта первичной медико-санитарной помощи взрослым при ВИЧ-инфекции (диагностика, лечение и диспансерное наблюдение)».

<sup>3</sup> Международная коалиция по готовности к лечению, Восточная Европа и Центральная Азия. Анализ закупок средств диагностики для лечения ВИЧ в России в 2020–2021 годах. 2022. URL: <https://itpc-eeca.org/wp-content/uploads/2022/07/monitoring-testov-vich-2020-21-gg-1.pdf>

<sup>4</sup> URL: <https://www.hivresist.ru/>

<sup>5</sup> Методические рекомендации «Надзор за распространением штаммов ВИЧ, резистентных к антиретровирусным препаратам». М.; 2013.

<sup>6</sup> Эпидемиологический надзор за ВИЧ-инфекцией. Методические указания. М.; 2016.

<sup>7</sup> Постановление Главного государственного санитарного врача РФ № 4 от 28.01.2021 «Об утверждении санитарных правил и норм СанПиН 3.3686-21 "Санитарно-эпидемиологические требования по профилактике инфекционных болезней" (с изменениями на 25 мая 2022 года)».

языках PHP, JS, HTML, CSS. Внешняя часть приложения, доступная для любого пользователя сети Интернет, реализована на платформе Bitrix.

Обновление данных статистики происходит автоматически, для чего на сервере установлены временные интервалы для обновления. Процесс получения данных для главной страницы реализуется посредством API (Application Programming Interface — программный интерфейс приложения) в формате json. Графики, представленные на главной странице сайта, создаются с помощью библиотеки ChartJS.

Анализ НП на предмет определения субтипов, наличия мутаций резистентности и ЛУ к АРВ-препаратам осуществлялся с помощью ресурсов базы данных Стэнфордского университета<sup>8</sup>.

Анализ НП, кодирующих V3-петлю, на предмет определения вирусного тропизма осуществлялся с помощью ресурса Института информатики Макса Планка<sup>9</sup>.

При анализе загружаемых НП на контаминацию был применён инструмент BLAST (Basic Local Alignment Search). В качестве пороговых значений были выбраны уровни 98% и 99% абсолютного и относительного генетического сходства НП друг с другом.

### Внесение информации в базу данных

Внесение информации осуществляется двумя способами. В первом случае информация вносится вручную по каждому пациенту отдельно. Этот способ удобен в случае регулярного пополнения базы данных.

Для удобства внесения информации таким способом были применены различные механизмы заполнения полей карточки пациента. Первый механизм реализован в качестве выпадающего списка с возможностью выбора из предложенных вариантов и, таким образом, исключает ошибки ручного ввода. Для этого разработаны справочники с информацией о населённых пунктах, АРВ-препаратах, кодах причин заражения ВИЧ и обследования на ВИЧ. Второй механизм позволяет на основании города проживания автоматически определить и подставить соответствующие регион проживания и федеральный округ. Третий механизм позволяет выбирать множественные позиции лекарственных препаратов. Все механизмы имеют подсказки и фильтрацию. Фильтрация была реализована возможностью выбора как по полному, так и по короткому названию АРВ-препарата из выпадающего списка.

В карточке пациента предусмотрена возможность многократного добавления любых динамических показателей, например, таких как концен-

трация РНК ВИЧ или CD4-клеток. При сохранении записи реализована проверка корректности заполнения, в частности, например, дата начала терапии не может быть позже даты её окончания. Все выявляемые ошибки подсвечиваются красным цветом, обращая на себя внимание пользователя. Разработаны функции автозаполнения и быстрого заполнения ячеек. В случае необходимости изменения перечней, например, появления новых АРВ-препаратов, существующие справочники могут быть изменены администраторами РБД.

Второй способ внесения информации подразумевает внесение данных о большом количестве пациентов путём загрузки электронной таблицы. Шаблон таблицы доступен для скачивания каждому пользователю в личном кабинете РБД.

В любой момент пользователи могут возвращаться к уже добавленным пациентам с целью обновления или дополнения уже внесённой информации.

### Контроль качества информации, внесённой в базу данных

#### *Контроль качества эпидемиологической и клинической информации о пациенте в случае загрузки электронной таблицы*

В момент массовой загрузки путём использования электронной таблицы базой данных проводятся следующие проверки:

- проверка корректности заполнения шаблона;
- проверка наличия одинаковых номеров карт пациентов;
- сопоставление номеров карт загружаемых пациентов с номерами, имеющимися в РБД;
- сверка загружаемой сопутствующей информации о пациенте с уже имеющейся.

Во всех случаях при обнаружении дубликатов, т.е. загрузке записей о пациенте, который уже присутствует в РБД, система предлагает пользователю подтвердить, что загружаемая информация не содержит ошибок, и после подтверждения дополняет новой информацией существующую запись. Система осуществляет поиск дубликатов по совпадающим номерам карт, а также по совпадающей комбинации даты рождения пациента, даты первого положительного иммунного блота и пола пациента.

#### *Контроль качества нуклеотидных последовательностей*

Для оценки качества загружаемых НП нами был разработан алгоритм, учитывающий следующие параметры: номер аминокислоты гена, с которой НП начинается; номер аминокислоты гена, которой НП заканчивается; количество сдвигов рамки считывания, инсерций/делеций, стоп-кодонов, вырожденных позиций, АРОВЕС и нетипичных мутаций. Загружа-

<sup>8</sup> URL: <https://hivdb.stanford.edu/>

<sup>9</sup> URL: <https://coreceptor.genopheno.org/>

емые НП анализируются с помощью разработанного алгоритма и, в случае обнаружения НП недостаточного качества, помечаются соответствующими флагами. Пользователю предлагается несколько вариантов действия: исключить помеченные НП из загружаемого пула, отредактировать и затем загрузить НП повторно либо загрузить их в том виде, который есть. Для исключения искажения результатов последующих анализов в случае непрохождения контроля качества НП маркируется как сомнительная и впоследствии не используется при анализе данных и формировании стандартизированного отчёта.

#### *Проверка на контаминацию*

Для выявления случаев контаминации образцов, произошедшей во время проведения исследования, нами была разработана программа, которая осуществляет расчёт генетической близости НП, загружаемых в РБД. После конвертации fasta-файлов, загружаемых в базу НП и уже имеющихся в ней, в бинарный формат для сравнения с НП проводится расчёт генетической близости для различных групп НП (внутри одной группы, загруженной в базу одновременно; внутри группы, загруженной в базу за последние 3 мес; внутри всех НП, загруженных данным учреждением). В случае обнаружения аномально высокого генетического сходства между проанализированными НП программа помечает их и предупреждает пользователя о потенциальной контаминации.

### **Анализ внесённой информации, выгрузка данных и формирование отчётов**

#### *Анализ нуклеотидных последовательностей*

Разработаны инструменты для автоматического анализа НП. В настоящее время функциональные возможности РБД включают в себя набор наиболее важных с практической точки зрения инструментов анализа генетических данных. Анализ осуществляется в момент загрузки НП ВИЧ в РБД.

Определение генетического варианта осуществляется с помощью алгоритма университета Стэнфорда. РБД отправляет НП на сайт университета и затем возвращает информацию об определённом генетическом варианте.

Анализ НП на предмет наличия мутаций резистентности к АРВ-препаратам проводится с использованием базы данных Стэнфорда. Осуществляются экспорт последовательности, её дистанционный анализ и последующий импорт с автоматическим переводом на русский язык результатов анализа. После завершения анализа ЛУ ВИЧ информация сохраняется в индивидуальной карточке пациента и может быть экспортирована в виде электронной таблицы в Excel или файла в формате \*.pdf.

Определение вирусного тропизма осуществляется РБД в случае загрузки НП ВИЧ региона V3

петли гена белка оболочки gp120. Анализ осуществляется с помощью алгоритма geno2pheno, разработанного Институтом биоинформатики Макса Планка [3]. После окончания анализа информация о тропизме вируса также добавляется в индивидуальную карточку пациента.

#### *Выгрузка данных и формирование отчётов*

Разработаны функции выгрузки как всех данных центра, так и части загруженной информации. Для подготовки необходимой к выгрузке информации разработаны фильтры по всем основным параметрам, в том числе по дате загрузки, дате забора крови, дате диагноза, опыту приёма АРВ-препаратов и др. Также доступна функция выгрузки индивидуальной карточки пациента в формате \*.pdf.

Дополнительно был разработан стандартизированный отчёт, который формируется на основании анализа загруженных данных. В этом случае механизм собирает необходимые данные в РБД, проводит анализ этих данных и генерирует стандартизированный годовой отчёт-справку. Отчёт содержит текстовые поля, которые пользователи могут редактировать и дополнять по мере необходимости, а также таблицы с данными о пациентах и распространённости мутаций ЛУ ВИЧ в регионе. Также в отчёте генерируются статистические данные о поле, причинах заражения пациентов, ЛУ ВИЧ. Сгенерированные отчёты можно сохранять для последующего использования, а также выгружать в формате \*.pdf. Кроме того, реализована возможность выгрузки отдельных статистических графиков для использования их в докладах или презентациях.

### **Количество пользователей и объём загруженных данных**

На 01.07.2022 в РБД было зарегистрировано 10 пользователей. Общее количество уникальных записей о пациентах на эту дату составило 10 626, а НП — 13 126. Описание количества записей о пациентах и НП, загруженных зарегистрированными пользователями, представлено в **табл. 1**.

Поскольку проведение молекулярно-эпидемиологических исследований возможно только при наличии клинико-эпидемиологической информации, особое внимание было уделено сбору данных о поле, возрасте, регионе проживания, предполагаемом пути инфицирования и дате постановки диагноза пациента, а также об опыте приёма АРВ-препаратов и применяемых схемах АРВ-терапии. Информация о наличии таких данных представлена в **табл. 2**.

### **Авторское право и обмен данными между пользователями**

Каждое учреждение, имеющее личный кабинет, имеет доступ ко всем описанным инструментам базы данных и информации, которую загрузили

**Таблица 1.** Информация о пользователях РБД и объёме загруженной информации**Table 1.** Information about users of the Russian HIV-1 antiviral drug resistance database and the amount of uploaded information

Учреждение Institution	Количество пациентов (записей) Number of patients (records)	Количество НП (область генома) Number of sequences (genome region)				
		pro-rev	int	pro-rev-int	env	full
Сибирский федеральный окружной центр по профилактике и борьбе со СПИД, Омский научно-исследовательский институт природно-очаговых инфекций Роспотребнадзора Siberian Federal District Center for the Prevention and Control of AIDS, Omsk Research Institute of Natural Focal Infections	1548	1548	0	0	0	0
Нижегородский научно-исследовательский институт эпидемиологии и микробиологии им. академика И.Н. Блохиной Роспотребнадзора Nizhny Novgorod Research Institute of Epidemiology and Microbiology named after academician I.N. Blokhina	82	82	0	0	0	0
Екатеринбургский научно-исследовательский институт вирусных инфекций Роспотребнадзора Yekaterinburg Research Institute of Viral Infections	265	275	0	0	0	0
Красноярский краевой центр профилактики и борьбы со СПИД Krasnoyarsk Regional Center for Prevention and Control of AIDS	62	62	0	0	0	0
Липецкий областной центр по профилактике и борьбе со СПИД и инфекционными заболеваниями Lipetsk Regional Center for the Prevention and Control of AIDS and Infectious Diseases	57	57	0	0	0	0
Клинический центр профилактики и борьбы со СПИД Министерства здравоохранения Краснодарского края Clinical Center for the Prevention and Control of AIDS of the Ministry of Health of the Krasnodar Territory	10	10	0	0	0	0
Ростовский научно-исследовательский институт микробиологии и паразитологии Роспотребнадзора Rostov Research Institute of Microbiology and Parasitology	175	175	0	0	0	0
Научный исследовательский центр эпидемиологии и микробиологии им. Н.Ф. Гамалеи Минздрава России N.F. Gamaleya Research Center of Epidemiology and Microbiology	3544	3544	24	0	93	0
Центральный научно-исследовательский институт эпидемиологии Роспотребнадзора Central Research Institute for Epidemiology	4801	4729	899	293	857	396
Хабаровский научно-исследовательский институт эпидемиологии и микробиологии Роспотребнадзора Khabarovsk Research Institute of Epidemiology and Microbiology	82	82	0	0	0	0
Общее количество   Total	10 626	10 564	923	293	950	396

**Примечание.** Pro-rev — НП гена протеазы и 2/3 гена обратной транскриптазы; int — НП гена интегразы; pro-rev-int — НП генов протеазы, обратной транскриптазы и интегразы; env — НП V3 петли гена белка gp120; full — НП всей кодирующей области генома ВИЧ.  
**Note.** Pro-rev — protease gene NS and 2/3 reverse transcriptase gene NS; int — integrase gene NS; pro-rev-int — protease, reverse transcriptase and integrase gene NS; env — gp120 V3 loop protein NS; full — NS of the entire coding region of HIV genome.

сотрудники этого учреждения. Информация, загруженная сотрудниками других учреждений, также может быть доступна, если другое учреждение предоставит такую возможность. Таким образом, с одной стороны РБД обеспечивает конфиденциальность загруженной информации. А с другой стороны, создаёт возможность быстрого обмена и совместного анализа объединённых данных в случае проведения несколькими учреждениями совместных научных и клинических исследований.

### Заключение

Сбор и хранение генетической информации о патогене и сопутствующей информации о пациенте

чрезвычайно важны, и на современном этапе централизованный анализ этих данных позволяет эффективно решать эпидемиологические задачи. В частности, Российская платформа агрегации информации о геномах вирусов VGARus [4], созданная во время развития пандемии COVID-19, показала свою незаменимость при мониторинге появления новых вирусных вариантов и анализе их контагиозности.

В случае ВИЧ-инфекции также существуют соответствующие базы данных, использующиеся для различных целей. Одни исследователи используют их для оценки общих тенденций развития эпидемии [5, 6], другие применяют для решения конкретных противоэпидемических мероприятий [7]. Наиболее



**Таблица 2.** Полнота клинической и эпидемиологической информации в записях пациентов, присутствующих в РБД  
**Table 2.** Completeness of clinical and epidemiological information in patient records available in RDB

Характеристика пациентов Patient profile		Записи о пациентах Patient records	
		n	%
Пол Sex	известен   available	10 490	98,7
	нет данных   no data	136	1,3
Год рождения Year of birth	известен   available	10 281	96,8
	нет данных   no data	345	3,2
Год первого положительного иммуноблота Year of the first positive immunoblot	известен   available	10 344	97,3
	нет данных   no data	282	2,7
Предполагаемый путь инфицирования Alleged route of infection	известен   available	7998	75,3
	нет данных   no data	2628	24,7
Информация об опыте приёма АРВ-препаратов Information about ARV treatment history	известен   available	10 295	96,9
	нет данных   no data	331	3,1
Регион проживания Region of residence	известен   available	10 143	95,5
	нет данных   no data	483	4,5
Информация о схемах АРВ-терапии Information about ART regimens	известен   available	4342	81,2
	нет данных   no data	1005	18,8

активно изучаемые направления работ по ВИЧ-инфекции в России, где база может быть с пользой применена, это изучение ЛУ [8] и анализ особенностей распространения различных генетических вариантов вируса [9].

Наиболее важными характеристиками любой базы данных являются объём информации, который она содержит, и функционал, позволяющий решать задачи, встающие перед специалистами. Это особенно актуально в случае проведения молекулярно-эпидемиологических работ, поскольку отсутствие необходимых инструментов анализа делает базу данных обычным хранилищем, не отличающим её от электронных таблиц, а малое количество информации значительно сокращает достоверность результатов исследований.

Необходимость загрузки НП ВИЧ и сопутствующей информации о пациентах в Российскую базу данных устойчивости ВИЧ к АРВ-препаратам согласно СанПиН 3.3686-21 позволяет рассчитывать на повышение объёма собираемой информации в будущем. В связи с этим возросла актуальность совершенствования функциональных возможностей РБД. Благодаря изменениям, позволившим быстрее и удобнее вносить информацию, количество записей в РБД значительно выросло за последние несколько лет. К 01.11.2022 количество наиболее актуальных для анализа последовательностей генов протеазы и обратной транскриптазы выросло до более чем 10 тыс. Это уже превышает количество общедоступных российских последовательностей ВИЧ, представленных во Всемирной базе данных о вирусах ВИЧ Национальной лаборатории

Лос-Аламос<sup>10</sup>. В ней на текущий момент депонировано всего 8875 последовательностей из России. Не менее важно отметить, что в РБД для каждой записи известно гораздо большее количество эпидемиологической и клинической информации. Во Всемирной базе данных о вирусах ВИЧ информация о ряде значимых характеристик, таких как дата постановки диагноза, регион проживания и схемы АРВ-терапии, отсутствует, а для других, например, таких как предполагаемый путь передачи (54,2%), пол пациента (43,9%), возраст (16,9%), она известна далеко не для всех записей.

Разработанный в РБД удобный способ заполнения характеристик пациентов позволил добыть более высокой собираемости этой информации; суммарная заполняемость данных о поле, возрасте, дате диагноза, предполагаемом пути инфицирования, регионе проживания и информации об опыте приёма АРВ-препаратов составляет 93,4%. Таким образом, согласно данным Роспотребнадзора о ВИЧ-инфекции в России на 31.12.2021, доля ВИЧ-инфицированных граждан России, для которых известна НП ВИЧ, составила 0,66% от общего количества выявленных случаев инфекции (1 562 570 случаев) и 0,90% от лиц, живущих с ВИЧ (1 137 596 человек).

Кроме увеличения количества записей в РБД повысилась и качество внесённой информации. Благодаря автоматической проверке снижается вероятность ошибки в клинико-эпидемиологических характеристиках пациентов, а также загрузки лож-

<sup>10</sup> URL: <https://hiv.lanl.gov>

ных данных о НП. В частности, нами обнаружено, что до 5% загружаемых НП идентичны друг другу, что невозможно, согласно эпидемиологическим данным пациентов. Такая ситуация может возникнуть либо вследствие ошибочного копирования в процессе заполнения данных о пациенте, либо в результате контаминации. Под контаминацией в данном случае подразумевается загрязнение одного клинического образца другим образцом в процессе проведения исследования в лаборатории. Какова бы ни была причина появления идентичных НП ВИЧ у различных пациентов, это может иметь катастрофический эффект. В таком случае пациент получит результаты анализа ЛУ не того вируса, которым он инфицирован. Разработанный инструмент проверки на контаминацию полностью исключает возникновение таких ошибок [10].

Другими важными дополнениями явились возможности автоматического субтипирования, выявления мутаций ЛУ и определения тропизма вируса. Например, данные о ЛУ могут быть применены не только на индивидуальном, но и на популяционном уровне. Информацию о структуре и распространённости ЛУ следует использовать при выборе схем АРВ-терапии на основании национальных рекомендаций или в ходе подготовки к закупкам АРВ-препаратов.

В рамках дальнейшей работы нами запланирована разработка наиболее сложного и современного инструмента для проведения биоинформатического анализа. Для проведения многих работ в области молекулярной эпидемиологии выявление молекулярных кластеров абсолютно необходимо [11, 12]. Такая функция станет доступна пользователям РБД в ближайшее время.

Обновление РБД позволит улучшить сохранность генетической информации о вирусе, получаемой в результате проведения анализов ЛУ ВИЧ в нашей стране, а также повысить качество исследований по теме молекулярной эпидемиологии ВИЧ-инфекции. Увеличение количества анализов и непрерывное пополнение РБД этой информацией позволит применять её в рутинной эпидемиологической работе. Выявленные молекулярные кластеры будут указывать сотрудникам специализированных учреждений на уязвимые группы и администра-

тивные районы, в которых происходит ускоренная передача вируса, что в свою очередь позволит проводить противоэпидемические мероприятия с максимальной эффективностью. Таким образом, в обозримом будущем РБД может перестать быть исключительно научным инструментом и стать значимым компонентом эпидемиологического надзора.

#### СПИСОК ИСТОЧНИКОВ | REFERENCES

- Scherrer A.U., Traytel A., Braun D.L., et al. Cohort profile update: The Swiss HIV Cohort Study (SHCS). *Int. J. Epidemiol.* 2022;51(1):33–4j. DOI: <https://doi.org/10.1093/ije/dyab141>
- Dunn D., Pillay D. UK HIV drug resistance database: background and recent outputs. *J. HIV Ther.* 2007;12(4):97–8.
- Lengauer T., Sander O., Sierra S. et al. Bioinformatics prediction of HIV coreceptor usage. *Nat. Biotechnol.* 2007;25(12):1407–10. DOI: <https://doi.org/10.1038/nbt1371>
- Акимкин В.Г., Семенов Т.А., Углева С.В. и др. COVID-19 в России: эпидемиология и молекулярно-генетический мониторинг. *Вестник Российской академии медицинских наук.* 2022; 77(4): 254–60. Akimkin V.G., Semenenko T.A., Ugleva S.V., et al. COVID-19 in Russia: epidemiology and molecular genetic monitoring. *Vestnik Rossiyskoy akademii meditsinskikh nauk.* 2022;77(4):254–60. DOI: <https://doi.org/10.15690/vramn2121>
- Kusejko K., Salazar-Vizcaya L., Shah C., et al. Sustained effect on hepatitis c elimination among men who have sex with men in the Swiss HIV cohort study: a systematic re-screening for hepatitis C RNA two years following a nation-wide elimination program. *Clin. Infect. Dis.* 2022;75(10):1723–31. DOI: <https://doi.org/10.1093/cid/ciac273>
- Kusejko K., Marzel A., Nguyen H., et al. Differences in social and mental well-being of long-term survivors among people who inject drugs and other participants in the Swiss HIV Cohort Study: 1980–2018. *Antivir. Ther.* 2020;25(1):43–54. DOI: <https://doi.org/10.3851/IMP3347>
- Poon A.F., Gustafson R., Daly P., et al. Near real-time monitoring of HIV transmission hotspots from routine HIV genotyping: an implementation case study. *Lancet HIV.* 2016;3(5):e231–8. DOI: [https://doi.org/10.1016/S2352-3018\(16\)00046-1](https://doi.org/10.1016/S2352-3018(16)00046-1)
- Ebbert M.T., Mallory M.A., Wilson A.R., et al. Application of a new informatics tool for contamination screening in the HIV sequencing laboratory. *J. Clin. Virol.* 2013;57(3):249–53. DOI: <https://doi.org/10.1016/j.jcv.2013.03.013>
- Hassan A.S., Pybus O.G., Sanders E.J., Albert J., Esbjörnsson J. Defining HIV-1 transmission clusters based on sequence data. *AIDS.* 2017;31(9):1211–22. DOI: <https://doi.org/10.1097/QAD.0000000000001470>
- Kosakovsky S.L., Weaver S., et al. HIV-TRACE (TRANsmiission Cluster Engine): a tool for large scale molecular epidemiology of HIV-1 and other rapidly evolving pathogens. *Mol. Biol. Evol.* 2018;35(7):1812–9. DOI: <https://doi.org/10.1093/molbev/msy016>

#### Информация об авторах

Киреев Дмитрий Евгеньевич<sup>✉</sup> — к.б.н., зав. лаб. диагностики и молекулярной эпидемиологии ВИЧ-инфекции ЦНИИ Эпидемиологии, Москва, Россия, [dmitkireev@yandex.ru](mailto:dmitkireev@yandex.ru), <https://orcid.org/0000-0002-7896-2379>

Кириченко Алина Алексеевна — н.с. лаб. диагностики и молекулярной эпидемиологии ВИЧ-инфекции ЦНИИ Эпидемиологии, Москва, Россия, <https://orcid.org/0000-0002-7116-0138>

Лопатухин Алексей Эдуардович — н.с. лаб. диагностики и молекулярной эпидемиологии ВИЧ-инфекции ЦНИИ Эпидемиологии, Москва, Россия, <https://orcid.org/0000-0002-2826-699X>

#### Information about the authors

Dmitry E. Kireev<sup>✉</sup> — Cand. Sci. (Biol.), Head, HIV diagnostic and molecular epidemiology laboratory, Central Research Institute for Epidemiology, Moscow, Russia, [dmitkireev@yandex.ru](mailto:dmitkireev@yandex.ru), <https://orcid.org/0000-0002-7896-2379>

Alina A. Kirichenko — researcher, HIV diagnostic and molecular epidemiology laboratory, Central Research Institute for Epidemiology, Moscow, Russia, <https://orcid.org/0000-0002-7116-0138>

Alexey E. Lopatukhin — researcher, HIV diagnostic and molecular epidemiology laboratory, Central Research Institute for Epidemiology, Moscow, Russia, <https://orcid.org/0000-0002-2826-699X>



ОБЗОРЫ

*Шлыкова Анастасия Вениаминовна* — н.с. лаб. диагностики и молекулярной эпидемиологии ВИЧ-инфекции ЦНИИ Эпидемиологии, Москва, Россия, <https://orcid.org/0000-0002-1390-8021>

*Галкин Николай Юрьевич* — старший программист информационных систем центра по развитию информационных технологий и систем ЦНИИ Эпидемиологии, Москва, Россия, <https://orcid.org/0000-0002-2144-8774>

*Савельев Евгений Викторович* — руководитель группы разработки информационных систем центра по развитию информационных технологий и систем ЦНИИ Эпидемиологии, Москва, Россия, <https://orcid.org/0000-0003-2811-8097>

*Глазов Максим Борисович* — руководитель центра по развитию информационных технологий и систем ЦНИИ Эпидемиологии, Москва, Россия, <https://orcid.org/0000-0002-2195-1580>

*Покровский Вадим Валентинович* — д.м.н., профессор, академик РАН, зав. специализированным отделом по эпидемиологии и профилактике СПИДа ЦНИИ Эпидемиологии, Москва, Россия, <https://orcid.org/0000-0002-9514-7288>

*Акимкин Василий Геннадьевич* — д.м.н., профессор, академик РАН, директор ЦНИИ Эпидемиологии, Москва, Россия, <https://orcid.org/0000-0003-4228-9044>

**Участие авторов.** Все авторы внесли существенный вклад в проведение поисково-аналитической работы и подготовку статьи, прочли и одобрили финальную версию до публикации.

Статья поступила в редакцию 24.11.2022;  
принята к публикации 01.02.2023;  
опубликована 28.03.2023

*Anastasia V. Shlykova* — researcher, HIV diagnostic and molecular epidemiology laboratory, Central Research Institute for Epidemiology, Moscow, Russia, <https://orcid.org/0000-0002-1390-8021>.

*Nikolai Yu. Galkin* — senior programmer, Center for the development of information technologies and systems, Central Research Institute for Epidemiology, Moscow, Russia, <https://orcid.org/0000-0002-2144-8774>

*Evgenii V. Saveler* — Group Head, Center for the development of information technologies and systems, Central Research Institute for Epidemiology, Moscow, Russia, <https://orcid.org/0000-0003-2811-8097>

*Maxim B. Glazov* — Head, Center for the development of information technologies and systems, Central Research Institute for Epidemiology, Moscow, Russia, <https://orcid.org/0000-0002-2195-1580>

*Vadim V. Pokrovsky* — D. Sci. (Med.), Prof., Full Member of RAS, Head, HIV department, Central Research Institute for Epidemiology, Moscow, Russia, <https://orcid.org/0000-0002-9514-7288>

*Vasily G. Akimkin* — D. Sci. (Med.), Prof., Full Member of RAS, Director, Central Research Institute for Epidemiology, Moscow, Russia, <https://orcid.org/0000-0003-4228-9044>

**Author contribution.** All authors made a substantial contribution to the conception of the work, acquisition, analysis, interpretation of data for the work, drafting and revising the work, final approval of the version to be published.

The article was submitted 24.11.2022;  
accepted for publication 01.02.2023;  
published 28.03.2023