

# ПЕРЕДОВЫЕ ТЕХНОЛОГИИ В ДИАГНОСТИКЕ ВИРУСНЫХ ЗАБОЛЕВАНИЙ НЕЯСНОЙ ЭТИОЛОГИИ

К.Ф. Хафизов<sup>1,4</sup>, А.С. Сперанская<sup>1,2</sup>, А.Д. Мацвай<sup>1,3</sup>, Г.А. Шипулин<sup>4</sup>, В.Г. Дедков<sup>5,6</sup>

<sup>1</sup> ФБУН ЦНИИ эпидемиологии Роспотребнадзора, Москва, Россия

<sup>2</sup> Московский государственный университет имени М.В. Ломоносова, Москва, Россия

<sup>3</sup> Московский физико-технический институт, г. Долгопрудный, Московская область, Россия

<sup>4</sup> ФГБУ Центр стратегического планирования и управления медико-биологическими рисками здоровью МЗ РФ, Москва, Россия

<sup>5</sup> ФБУН НИИ эпидемиологии и микробиологии имени Пастера, Санкт Петербург, Россия

<sup>6</sup> Институт медицинской паразитологии, тропических и трансмиссивных заболеваний им. Е.И. Марциновского, Москва, Россия

**Резюме.** Расшифровка инфекционных заболеваний неясной этиологии является одной из актуальных проблем современной медицины, потому как получение лабораторно подтвержденного диагноза, к сожалению, удастся осуществить лишь в весьма небольшой доле случаев таких заболеваний. Так как большая часть часто встречающихся в средних широтах инфекционных заболеваний имеет характерную выраженную клиническую картину, до недавнего времени этой проблеме не уделялось должного внимания. Возрастание числа случаев инфекционных заболеваний, не характеризующихся идентифицируемым набором клинических признаков, наблюдаемое в последнее время, заставляет рассматривать проблему более пристально. Считается, что такая тенденция обусловлена рядом обстоятельств, включая ослабление санитарного контроля территорий, усиление миграционных потоков, как внутренних, так и внешних, отказ от вакцинации на фоне длительного периода эпидемического благополучия, возникновение атипичных штаммов бактерий как следствие нерациональной антибиотикотерапии, и другие. Вирусы являются наиболее распространенными организмами на нашей планете, что обуславливает ведущую роль инфекционных агентов вирусной природы в структуре инфекционных заболеваний неясной этиологии. По некоторым оценкам, полученным методами математического моделирования, существует не менее 320 000 видов вирусов, способных к инфицированию млекопитающих, большая часть которых еще не описана. Поэтому мониторинг циркуляции известных вирусных патогенов, отслеживание путей их распространения, эволюции и изменений нуклеотидной последовательности их геномов, а также выявление новых видов вирусов становятся жизненно важными аспектами эпидемиологического надзора, необходимыми для своевременного реагирования на возникающие угрозы, прогнозирования и раннего выявления вспышек вирусных заболеваний человека и животных. В представленном обзоре рассматриваются как традиционные молекулярно-генетические методы выявления вирусных патогенов, такие как методы ПЦР, ПЦР в режиме реального времени, секвенирование по Сенгеру с предварительным клонированием, так и методы, основанные на применении секвенирования второго и третьего поколений. В связи с тем, что исследование вирусных инфекционных агентов с помощью технологий высокопроизводительного секвенирования NGS (англ. Next Generation Sequencing) на сегодняшний день приобретает все большее практическое значение для диагностики, бор-

---

**Адрес для переписки:**

Хафизов Камилль Фаридович  
111123, Россия, Москва, ул. Новогиреевская, 3а,  
ФБУН ЦНИИ эпидемиологии Роспотребнадзора.  
Тел.: 8 917 597-20-85 (моб.).  
E-mail: kkhafizov@gmail.com

**Contacts:**

Kamil F. Khafizov  
111123, Russian Federation, Moscow, Novogireevskaya str. 3a,  
Central Research Institute of Epidemiology.  
Phone: +7 917 597-20-85 (mobile).  
E-mail: kkhafizov@gmail.com

**Библиографическое описание:**

Хафизов К.Ф., Сперанская А.С., Мацвай А.Д., Шипулин Г.А., Дедков В.Г.  
Передовые технологии в диагностике вирусных заболеваний неясной  
этиологии // Инфекция и иммунитет. 2020. Т. 10, № 1. С. 9–25.  
doi: 10.15789/2220-7619-ATI-824

**Citation:**

Khafizov K.F., Speranskaya A.S., Matsvay A.D., Shipulin G.A., Dedkov V.G.  
Advanced technologies in diagnostics of viral diseases of unknown etiology //  
Russian Journal of Infection and Immunity = Infektsiya i immunitet, 2020,  
vol. 10, no. 1, pp. 9–25. doi: 10.15789/2220-7619-ATI-824

Работа поддержана грантом РФФИ №17-74-20096.

бы с болезнями, молекулярной эпидемиологии и инфекционного контроля, более подробное рассмотрение получили различные методы, основанные на применении этих технологий. Особое место было также отведено подходам, применяющимся для обогащения вирусного генетического материала в образцах с низким содержанием нуклеиновых кислот патогена.

**Ключевые слова:** вирусы, диагностика, неизвестная этиология, инфекционные заболевания, молекулярная генетика, NGS, секвенирование, ПЦР, обогащение.

## ADVANCED TECHNOLOGIES IN DIAGNOSTICS OF VIRAL DISEASES OF UNKNOWN ETIOLOGY

Khafizov K.F.<sup>a,d</sup>, Speranskaya A.S.<sup>a,b</sup>, Matsvay A.D.<sup>a,c</sup>, Shipulin G.A.<sup>d</sup>, Dedkov V.G.<sup>e,f</sup>

<sup>a</sup> Central Research Institute of Epidemiology, Moscow, Russian Federation

<sup>b</sup> Lomonosov Moscow State University, Moscow, Russian Federation

<sup>c</sup> Moscow Institute of Physics and Technology, Dolgoprudny, Moscow Region, Russian Federation

<sup>d</sup> Center of Strategic Planning of the Ministry of Health, Moscow, Russian Federation

<sup>e</sup> St. Petersburg Pasteur Institute, St. Petersburg, Russian Federation

<sup>f</sup> Martsinovskiy Institute of Medical Parasitology, Tropical and Vector-Borne Diseases, Moscow, Russian Federation

**Abstract.** Unveiling origin of infectious diseases with unknown etiology is one of the major issues in contemporary medicine, since a laboratory-confirmed diagnosis may, unfortunately, be obtained solely in very few cases. Because the majority of the most common mid-latitude infections display a typical overt clinical picture, this problem has not been paid a proper attention until recently. Recent rise in incidence rate of infectious diseases lacking typical clinical signs observed lately makes it extremely important to consider the problem more closely. It is believed that such trend is due to a whole body of reasons, including impaired sanitary control, increased both internal and external migration flows, refusal of vaccination in case of long-lasting epidemic wellbeing, emergence of atypical bacterial strains of bacteria resulting from irrational antibiotic therapy etc. Viruses constitute the largest group of organisms on our planet accounting for them as the most common causative agent of infectious diseases of unknown etiology. Some estimates obtained by mathematical modeling propose that at least 320,000 types of viruses capable of infecting mammals may exist, most of which have not been described yet. Hence, monitoring circulation of the known viral pathogens, tracing down their spreading and changes in genome nucleotide sequence as well as revealing new types of viruses become important aspects of epidemiological surveillance necessary for timely response to emerging threats, prediction and early detection of outbreaks both in humans and animals. This review summarizes traditional molecular genetics methods for detection of viral pathogens, such as PCR, real-time PCR, Sanger sequencing with pre-cloning, and methods based on the second and third generation sequencing. Therefore, a more detailed overview was provided to diverse methods based on using such technologies because viral infectious agents investigated with high-throughput sequencing (or NGS — Next Generation Sequencing) has been increasingly appreciated as feasible for diagnostics, disease control, molecular epidemiology and infection control. Finally, a special attention was also paid to the approaches used to enrich the viral genetic material in samples containing low amount of pathogen nucleic acids.

**Key words:** viruses, diagnostics, unknown etiology, infectious diseases, molecular genetics, NGS, sequencing, PCR, enrichment.

## Введение

Расшифровка случаев инфекционных заболеваний неясной этиологии является актуальной проблемой современной медицины. К сожалению, лишь в весьма небольшой доле случаев таких заболеваний удается получить лабораторно подтвержденный диагноз. До недавнего времени этой проблеме не уделялось достаточного внимания в связи с тем, что большая часть наиболее часто встречающихся в средних широтах инфекционных заболеваний имеет выраженную клиническую картину с хорошо различимой патогномоничной симптоматикой. Однако в последние годы наблюдается рост случаев инфекционных заболеваний, не имеющих характерной клинической картины. В частности, большинство трансмиссивных заболеваний вирусной этиологии в большинстве своем имеют сходное клиническое течение, особенно

на начальном этапе заболевания. В качестве примера можно привести лихорадку Зика, сходную по клиническим проявлениям с лихорадкой денге и лихорадкой Чикунгунья. Первые два заболевания обусловлены флавивирусами, тогда как лихорадка Чикунгунья вызывается представителем рода *Alphavirus*. При этом природные очаги циркуляции, а также переносчики для всех трех вирусов одинаковы [50]. Такая тенденция обусловлена рядом обстоятельств: ослаблением санитарного контроля территорий, усилением миграционных потоков, как внутренних, так и внешних, отказом от вакцинации на фоне длительного периода эпидемического благополучия, возникновением атипичных штаммов бактерий вследствие нерациональной антибиотикотерапии и пр. В этих условиях качество диагностики инфекционных заболеваний и организация надзора за ними находятся в зависимости от уровня развития

лабораторной службы и способности проводить диагностику с помощью современных молекулярных методов.

В структуре инфекционных заболеваний неясной этиологии ведущую роль играют инфекционные агенты вирусной природы. Это обстоятельство связано с тем, что вирусы являются наиболее распространенным источником генетического материала на Земле, вероятно инфицируя все клеточные организмы. Высокая изменчивость геномов вирусов и способность адаптироваться к новым хозяевам, в сочетании с климатическими изменениями, а также изменяющийся характер хозяйственной деятельности человека и процессы глобализации способствуют возникновению эмерджентных вспышек вирусных инфекций, обусловленных как вновь возникающими, так и редкими либо нехарактерными для данной территории видами вирусов [48]. Многие специалисты считают, что мы знаем о вирусах меньше, чем о любой другой группе организмов. Вероятно на сегодняшний день охарактеризована всего лишь доля одного процента от всех вирусов [34]. Согласно данным математического моделирования существует не менее 320 000 видов вирусов млекопитающих, большая часть которых еще не описана [18]. Все они потенциально являются возбудителями инфекционных заболеваний человека и животных. В связи с этим выявление новых видов вирусов, мониторинг циркуляции известных вирусов и оценка путей их распространения являются одним из жизненно важных аспектов дозорного эпидемиологического надзора, проводимого с целью своевременного реагирования на вновь возникающие угрозы, прогнозирования и раннего выявления вспышек вирусных заболеваний человека и животных.

## Традиционные методы диагностики инфекционных агентов

До наступления эры молекулярных исследований выявление новых возбудителей базировалось на использовании классических методов микробиологии и вирусологии, в том числе накоплении бактерий и вирусов на культуральных средах, клеточных линиях и лабораторных животных с последующим определением их антигенных и патогенных свойств, особенностей метаболизма, а также фенотипических особенностей на основе данных микроскопических исследований.

Появление приборов для автоматического капиллярного секвенирования в сочетании с развитием методик молекулярного клонирования расширили возможности по определению новых патогенных микроорганизмов. Суть данных методов состояла в амплифика-

ции тотальной ДНК (либо кДНК в случае РНК-содержащих вирусов) с последующим молекулярным клонированием полученных библиотек и анализе большого количества индивидуальных клонов методом капиллярного секвенирования в надежде получить последовательности нового микроорганизма. В других случаях для исследований использовался метод гибридизационного анализа с помощью ДНК-чипов, на которых иммобилизованы олигонуклеотиды, соответствующие консервативным участкам большого количества видов вирусов. При этом исследователи сталкивались с двумя основными проблемами: во-первых, относительно низкое содержание генетического материала вируса в исследуемом материале и, во-вторых, высокий процент нуклеиновых кислот человека и нормофлоры в клиническом материале. В тех случаях, когда не удавалось увеличить количество вирусных частиц за счет культивирования *in vitro*, эти задачи решались с использованием физических и химико-ферментативных методов. С этой целью на начальном этапе исследуемый образец фильтровали и концентрировали методом ультрацентрифугирования. Далее следовала стадия обработки экзонуклеазами, что позволяло избавиться от высокомолекулярной геномной ДНК и РНК хозяина. Считалось, что генетический материал, заключенный в вирусную оболочку, при этом остается интактным. После этапа предварительной обработки следовал этап выделения нуклеиновых кислот, лигирование с олигонуклеотидными адаптерами и амплификация материала, обогащенного вирусной ДНК (или РНК) с помощью ПЦР (или ОТ-ПЦР) для получения библиотек.

Было разработано несколько альтернативных подходов для идентификации генетического материала инфекционных агентов и определения последовательностей их геномов. К числу таких подходов следует отнести метод SISPA (Sequence-Independent Single Primer Amplification), а также различные его модификации [5, 6, 7, 8, 9]. Еще один метод, оптимизированный для задач виromики и названный VIDISCA, основан на принципе AFLP [4, 106].

Другим направлением развития методик выявления неизвестных патогенов стала универсальная ПЦР (broad-range PCR), основанная на амплификации небольших консервативных фрагментов вирусных геномов с помощью праймеров, специфических для всех представителей данного рода (реже семейства) вирусов. Полученный продукт амплификации анализировался методом капиллярного секвенирования напрямую, либо после этапа молекулярного клонирования. Универсальная ПЦР была успешно использована для изучения биоразнообразия [5, 109] и идентификации новых эмер-

джентных патогенов [27], но этот подход все же характеризуется низкой чувствительностью, сложностью и длительностью выполнения, отсутствием универсальности и невозможностью стандартизации. Кроме того интерпретация таких результатов в клинической лаборатории может быть заметно затруднена, потому описанные подходы выявления неизвестных вирусных агентов не получили широкого применения в практической лабораторной диагностике.

Не вызывает сомнения тот факт, что наиболее удобным, дешевым и результативным методом лабораторной диагностики инфекционных заболеваний является метод ПЦР в реальном времени (Real Time PCR) с использованием специфических праймеров и ДНК-зондов. Этот подход уже довольно долго и успешно [104] используется как для исследования представителей отдельных вирусных родов, так и при анализе разнообразия вирусов в различных типах биологического материала [11, 12, 13]. На настоящее время описано значительное количество праймеров, предназначенных для обнаружения определенных вирусных видов [15, 16, 17]. Однако их невозможно использовать в мультиплексных реакциях из-за различных температур отжига праймеров, неспецифической амплификации и потенциальной самокомплементарности. Таким образом, для видового определения вирусов разных родов необходимо проводить ряд экспериментов ПЦР. Во многих случаях также бывает сложно очертить возможный круг инфекционных агентов, которые могли послужить причиной данного заболевания. Таким образом, для проведения исследований методом специфической ПЦР требуется предварительная гипотеза о наличии конкретных вирусов в образце, в отсутствие которой процесс идентификации патогенов может занять значительное время, что накладывает ограничение на выбор тактики лечения и снижает эффективность противоэпидемических мероприятий. Кроме того, большинство диагностических лабораторий имеет довольно ограниченный спектр наборов, позволяющих выявлять лишь очень небольшое количество инфекционных агентов, наиболее часто встречающихся на данной территории.

## Технологии массового параллельного секвенирования для изучения вирусных патогенов

Появление технологий массового параллельного секвенирования (или как их часто называют — NGS, Next Generation Sequencing) вызвало настоящую революцию в самых различных областях медицины и биологии. Вирусология

не осталась в стороне и также получила существенный толчок в развитии благодаря возможности обнаруживать новые патогены и прочитывать вирусные геномы целиком. Так, полногеномное секвенирование (WGS, Whole Genome Sequencing) вирусов становится все более важным не только для фундаментальных исследований, но и для клинической науки и, в перспективе, медицинской диагностики. В вирусологии этот подход уже получил серьезную роль в разработке новых методов лечения и вакцин, а также для расширения возможностей молекулярной эпидемиологии и эволюционной геномики [41].

В то время как большинство экспериментов по секвенированию бактериальных геномов в настоящее время осуществляется на клинических изолятах, которые проходят этап предварительного культивирования, и таким образом их пробоподготовка для NGS сравнительно проста, изучение вирусных нуклеиновых кислот, будь то выделенных из культуральных сред или непосредственно из клинических образцов, практически всегда осложняется наличием контаминационной (загрязняющей) ДНК-хозяина [65]. Чтобы решить эту проблему, изучение геномов вирусов может проводиться как путем ультраглубокого секвенирования, так и обогащением нуклеиновых кислот вирусов перед секвенированием. Последнее может осуществляться либо напрямую, либо же путем концентрации вирусных частиц. При этом все эти подходы имеют свои собственные издержки и сложности.

Так, широко распространен метагеномный подход, позволяющий обнаруживать различные вирусные патогены с использованием так называемого «шот-ган» (англ. «shot-gun») секвенирования [20, 21, 22], который все более активно применяется для выявления патогенов и для характеристики микробного разнообразия в экологических и клинических образцах [72, 105]. Однако, несмотря на постоянное снижение стоимости прочтения ДНК ([https://www.genome.gov/images/content/costpermb\\_2017.jpg](https://www.genome.gov/images/content/costpermb_2017.jpg)), этот способ все еще достаточно дорог и нецелесообразен для скрининга большого количества образцов. Метагеномный подход отличается тем, что для подготовки библиотек используют препараты тотальной ДНК/РНК, выделенные из образцов. Такие препараты содержат нуклеиновые кислоты не только целевого организма, но также хозяина и контаминантных примесей (бактерий, грибов и т.п.). Прямое «шот-ган» секвенирование таких образцов (без предварительной деплеции нежелательных примесей) приводит к генерации данных, преимущественно состоящих из последовательностей, относящихся к геномам хозяина и/или других организмов

[65], и лишь с крайне незначительной долей последовательностей искомым вирусных патогенов, что заметно снижает чувствительность эксперимента. В связи с тем, что до начала эксперимента часто даже приблизительная доля содержимого патогена неизвестна, практически невозможно оценить, сколько всего прочтений потребуется на один образец, чтобы гарантированно обнаружить патоген в файле с «сырыми» данными секвенирования. Так, доля прочтений, которые соответствуют геному целевого вируса в результатах метагеномного секвенирования, часто весьма низка; например, 0,008% для вируса Эпштейна—Барр (ВЭБ) в крови здорового взрослого [2], 0,0003% для вируса Ласса в клинических образцах [66] и 0,3% для вируса Зика в образце, который был обогащен посредством фильтрации и центрифугирования [14]. Наши собственные эксперименты по секвенированию транскриптомных библиотек, полученных от птиц и клещей, также во многом подтверждают эти данные. Например, содержание вирусных прочтений в образцах с калифорнийским энцефалитом обычно в пределах 0,03–0,06%, в образце с вирусом Вад-Медани — 0,7%, авиаденовирусом — 0,1%, авастровирусом — 0,02% и коронавирусом — 0,0001%. При этом в ряде случаев мы вообще не смогли найти вирусные прочтения в результатах секвенирования биоматериала, в котором наличие вируса было точно подтверждено с использованием специализированных родоспецифичных праймеров, а общее число прочтений в «fastq»-файлах достигало одного миллиона. Отдельно стоит упомянуть и про различные наборы для выделения вирусной ДНК/РНК. Наши эксперименты (неопубликованные данные) по сравнению различных наборов и методик показали, что количество получаемых вирусных прочтений может отличаться на порядки для одних и тех же образцов, вновь показывая, что результаты, получаемые в разных лабораториях, могут очень сильно различаться.

Глубина покрытия при такого рода экспериментах также обычно совершенно недостаточна для обнаружения мутаций резистентности к лекарственным препаратам [99], а стоимость подобных исследований высока. Таким образом, метагеномное секвенирование обычно проводится только на небольшом количестве образцов для исследовательских целей [2, 14]. Широко известны такие подходы как концентрация вирусных частиц из клинических образцов путем опосредованного сдерживания антителами, фильтрация, ультрацентрифугирование и деплеция свободных нуклеиновых кислот, принадлежащих в основном хозяину [54, 60, 77, 91], и хотя они могут повысить процент вирусных последовательностей, однако они также значи-

тельно уменьшают общее количество вирусных нуклеиновых кислот, что часто приводит к ситуациям, когда количество ДНК патогена будет недостаточно для подготовки библиотеки для секвенирования. Кроме того, все эти методы заметно увеличивают стоимость и длительность исследования. Неспецифические методы амплификации, такие как амплификация с множественным вытеснением цепи (MDA, Multiple Displacement Amplification), использующая случайные праймеры и полимеразу Ф29, могут увеличивать выход ДНК, но эти подходы требуют дополнительного времени, финансовых затрат и приводят к неравномерной представленности фрагментов, ошибкам и загрязнениям, не обязательно улучшая чувствительность [46, 88]. Более того, доля прочтений хозяина часто все еще остается высокой [21].

Когда метагеномные подходы используются для обнаружения или диагностики патогенов, крайне важно использовать соответствующие биоинформационные инструменты и базы данных, которые помогут оценить являются ли обнаруженные патогенные последовательности причинами инфекции, случайными находками или контаминацией. Сопутствующий биоинформационный анализ объемных метагеномных данных обычно требует значительных вычислительных ресурсов, а также определенных знаний и опыта по работе со средствами анализа. Как правило, сам анализ заключается в попарном сравнении полученных последовательностей с теми что хранятся в базах данных, для чего используются различные программные средства, среди которых наиболее популярен BLAST [3], хотя существуют и другие программы [6].

Что касается опубликованных биоинформатических пайплайнов для анализа данных с целью поиска вирусных патогенов, то они, как правило, делятся на две категории: 1) те, что первоначально удаляют прочтения «хозяина» путем точного выравнивания на соответствующий хозяйский геном, и в дальнейшем анализ идет уже только для оставшихся «нехозяйских» прочтений; 2) те, что производят сборку прочтений до более длинных контигов, а затем последние уже анализируются на предмет принадлежности к какому-либо организму, включая вирусные. Однако у первого подхода имеется один существенный недостаток — он подходит только для тех случаев, когда хозяином выступают хорошо изученные организмы с хорошо охарактеризованными геномами, такие как, например, человек или мышь. Впрочем, с развитием технологий секвенирования все больше полных геномов становятся доступными публично, и эта проблема постепенно уходит в прошлое. У второго подхода тоже есть свой не-

достаток — для сборки длинных контигов требуется высокое покрытие, что существенно увеличивает как время анализа, так и стоимость эксперимента. При недостаточном количестве данных возможны пробелы между прочтениями, и контиги часто не могут быть построены, в конечном счете, мешая идентификации вирусов. Впрочем, общее у обоих подходов — все прочтения-кандидаты («выжившие» после фильтрации хозяйской ДНК в подходе первого типа, или «удлиненные» прочтения при втором подходе) сравниваются с известными вирусными последовательностями. Среди опубликованных пайплайнов первого типа можно перечислить подход Petty и соавт. [55], в котором необработанные прочтения сначала сопоставляются с человеческим референсным геномом, а оставшиеся затем сравниваются с вирусной базой данных. Удаление хозяйских прочтений является важным шагом и было реализовано в VirusFinder [107], VirusHunter [113], VirusSeq [16] и Vy-PER [103]. Отдельно стоит упомянуть, что если хозяйские прочтения остаются в анализируемых данных, то могут возникнуть ложноположительные результаты за счет возможного высокого сходства с вирусными геномами. VirusFinder [107] сначала выполняет этап предварительной обработки, в котором сырые прочтения выравниваются на человеческий геном с использованием Bowtie2 [58, 59]. Затем оставшиеся прочтения выравниваются на вирусную базу данных с использованием BLAT [51]. Наконец, предположительно вирусные прочтения собираются до более длинных контигов с использованием Trinity [37]. VirusFinder предполагает высокую глубину секвенирования, чтобы можно было получать хорошие результаты сборки, и в основном разрабатывался для обнаружения сайтов интеграции вирусов в геноме человека. Примеры, представленные в статье VirusFinder, имеют глубину секвенирования от 31X до 121X. Собранные контиги затем используются для построения филогенетических деревьев и оценки взаимосвязи друг с другом. VirusHunter [113] использует BLASTn для первоначальной фильтрации прочтений, принадлежащих хосту, после некоторой оценки качества. Оставшиеся прочтения затем классифицируются с использованием BLASTn и BLASTx в таксономические группы. Таким образом, VirusHunter нуждается в исходном геноме хозяина хорошего качества, чтобы разделить прочтения на хозяйские и нехозяйские. Кроме того, повторный запуск BLAST для обработки данных потребует времени. VirusSeq [16] сфокусирован на идентификации вирусных штаммов в раковой ткани человека. Сначала все прочтения человеческой ДНК удаляются путем сопоставления с человеческим геномом,

и оставшиеся прочтения затем выравниваются на вирусную базу данных, с использованием программы MOSAIK на обоих этапах [61]. VirusSeq использует подсчет общего количества прочтений для идентификации вирусного штамма. Тем не менее, пороговое значение установлено как 1000 прочтений на вирус в отношении 30-кратного покрытия вирусного генома. Этот порог может быть изменен, но VirusSeq разработан именно для образцов с высоким значением глубины прочтения, поэтому он не может использоваться для анализа наборов данных с низким процентным содержанием вирусного материала. Vy-PER [103] использует на первом этапе человеческий геном для удаления прочтений хозяйской ДНК. Прочтения, которые не выравниваются на человеческий геном, затем сравниваются с базой данных вирусных геномов NCBI с использованием BLAT [51]. Описанный пример на образцах лейкемии выполняется с довольно высокой глубиной (80X для случаев и 40X для контролей), что не является обязательным требованием, но важной предпосылкой для устранения ложных результатов.

Все упомянутые пайплайны второй категории для выравниваний используют программные средства, такие как Blast или Bowtie2. Более полный обзор доступных программ приведен в работе Fonseca и соавт. [32]. В случае обнаружения вируса определенную роль играют специфические проблемы: 1) высокая гетерогенность геномов; 2) скорость мутаций; 3) вставки целых геномных областей и 4) заражение новых хозяев адаптацией вирусного генома. Следовательно, возникают проблемы при обращении с образцами из большого разнообразия потенциально инфицированными вирусами видами. Во-первых, полностью собранный референсный геном доступен только для небольшого количества животных, хотя эта ситуация постепенно меняется. Кроме того, качество референсных геномов может различаться. Во-вторых, количество вирусной ДНК, читаемой в биологическом образце, зависит от производственного цикла вируса, временной точки заражения и выбора правильного типа ткани, чтобы получить большую часть вируса из образца. Следовательно, количество возможных детектируемых вирусных последовательностей может быть низким. Для построения контигов должно быть доступно множество прочтений вирусной последовательности, то есть должен быть обеспечен хороший охват, и эти чтения не должны быть загрязнены последовательностями из организма-хозяина и других микроорганизмов.

В любом случае, при анализе отдельно стоит вопрос наличия качественных баз данных аннотированных последовательностей. Существуют

курируемые вручную специализированные БД вирусных последовательностей, такие как, например, ViPR [81], RVDB [36], viruSITE [98] и другие (обширный список может быть найден по адресу [44]), однако они в основном содержат весьма ограниченное число информации, и не могут использоваться для поиска и идентификации патогенов, особенно новых и неизвестных. Напротив, наиболее обширная база NCBI GenBank содержит очень большое количество вирусных последовательностей, в том числе геномных фрагментов, и при этом постоянно пополняется, однако качество хранимой информации вызывает серьезные вопросы. Впрочем, даже эта база данных очень далека от того, чтобы считаться исчерпывающей в части охвата вирусного генетического материала.

## Проблемы метагеномного NGS подхода для исследования вирусов

Метагеномика не требует предварительной гипотезы о наличии конкретного вирусного патогена, и это является значительным преимуществом подхода [99], поскольку позволяет секвенировать новые вирусы без необходимости разработки и синтеза специализированных праймеров или ДНК-зондов. Это особенно актуально для быстрого реагирования на возникающие угрозы, такие как вирус Зика [70]. Для вирусов, ассоциированных с развитием рака, метагеномика может помочь при оказании клинической помощи, предоставлять информацию о развитии заболевания и генерировать данные с высоким охватом интегрированных вирусных геномов [23]. Однако случайные находки как в хозяйской, так и в микробиологических последовательностях могут также представлять определенные этические и даже диагностические дилеммы для клинической метагеномики [39]. Имеет место и тот факт, что на сегодняшний день основное влияние метагеномики было на расширение разнообразия существующих семей, а не на определение новых, в связи с чем наблюдаются определенные ограничения такого подхода для изучения неизвестных ранее патогенов. На практике очень вероятно ситуация, при которой по результатам метагеномного секвенирования прочтения вирусной ДНК могут находиться в файле с данными, но не обнаруживаться путем сравнения с записями в базе данных из-за отсутствия сколь либо заметной гомологии с ними. Это особенно вероятно для вирусов, присутствующих в хостах, которые ранее не подвергались скринингу на вирусные патогены, или те, которые сильно разошлись генетически. Как следствие, для лучшего понимания «виросферы» требуется со-

четание более экспансивной выборки и серьезных улучшений в алгоритмах вычислительного анализа. Парадоксально, но чем больше мы вникаем в «виросферу», тем более очевидным становится тот факт, что мы охарактеризовали лишь крайне незначительную часть вирусов, с систематическим уклоном против выявления наиболее расходящихся геномов [95, 112].

Наконец, существует и другая распространенная проблема большинства тестов на основе технологий NGS, как метагеномных так и прочих, заключающаяся в том, что сложные многоступенчатые процессы могут создавать серьезные проблемы для воспроизводимости результатов, что также мешает внедрению этих технологий в диагностическую практику, где цена ошибки особенно высока.

## Получение полных геномов вирусов

Для небольших вирусов, таких как вирус иммунодефицита человека (ВИЧ, англ. HIV), вирус гриппа, вирус гепатита В (HBV) и вирус гепатита С (HCV), широко используется секвенирование фрагментов геномов для научных исследований, что, впрочем, имеет и важное клиническое применение. Таргетное секвенирование небольшого числа генов, кодирующих мишени противовирусных агентов, таких как ВИЧ, уже стало нормой в клинической практике. Для обнаружения ограниченного числа антивирусоустойчивых вариантов полногеномное секвенирование до недавнего времени было слишком дорогостоящей и трудоемкой процедурой для использования по сравнению с прочтением отдельных генов, на которые нацелены лекарственные препараты. Тем не менее, постепенно мы получаем новую информацию о все большем числе генов устойчивости, что вкупе с уменьшением затрат на секвенирование приводит к переосмыслению потребности в WGS при исследованиях. Рутинное использование WGS патогенов для диагностических целей [1], вероятно, в скором будущем будет иметь более широкие клинические и исследовательские преимущества. Например, последовательности вируса Зика, которые были обнаружены в эпидемиологических целях, влияют на принятия решений в общественном здравоохранении [31]. Геномы ВИЧ, секвенированные для идентификации противовирусных резистентных вариантов, также использовались для изучения эволюции вируса [68] и вирусной генетической ассоциации с заболеванием, включая исследования ассоциаций генотип–фенотип, а также связи генетических вариантов хозяина и результатов инфицирования, включая установление вирусной нагрузки при ВИЧ-инфекции [8, 82].

В настоящее время стоимость секвенирования полных вирусных геномов, несмотря на их небольшой размер, все же существенно выше, чем стоимость секвенирования отдельных генов устойчивости к лекарственным препаратам. Разница в затратах между секвенированием целевой области и целого генома вируса во многом определяется размером генома по сравнению с размером и количеством целевых локусов. Тем не менее, информация о целом геноме может предоставить важную дополнительную информацию, как описано выше.

### Обогащение вирусного генетического материала ПЦР-амплификацией с использованием специфичных праймеров

Альтернативой метагеномным подходам является целевое обогащение определенного вирусного генома(-ов) или их фрагментов перед секвенированием. Так, эффективность исследования может быть значительно увеличена за счет комбинирования родоспецифичной ПЦР и технологий NGS. ПЦР-амплификация вирусного генетического материала с использованием праймеров, которые являются комплементарными известной нуклеотидной последовательности, является наиболее распространенным подходом для обогащения небольших вирусных геномов, таких как ВИЧ и вирус гриппа. Недавние примеры обогащения ПЦР-амплификацией, за которым следует секвенирование, включают филогенетический анализ вспышки вируса кори на зимних Олимпийских играх 2010 г. [33] и отслеживание недавних эпидемий вирусов Эболы и Зика. С помощью секвенирования амплифицированных длинных (2,5–3 т.п.о.) фрагментов была исследована вариабельность генома норовируса и возможность его нозокомиальной передачи среди пациентов нескольких госпиталей во Вьетнаме [19, 57], что выявило как независимые интродукции возбудителя в больницу, так и внутрибольничную передачу, несмотря на предпринимаемые меры по борьбе с инфекцией. Аналогичный подход был использован в работе по определению чувствительности и специфичности платформы Illumina при идентификации минорных полиморфизмов в смешанных популяциях ВИЧ [53]. В других исследованиях глубокого секвенирования на основе ПЦР были получены полные геномы вируса гриппа [108] (13,5 кб), вируса денге [79] (11 кб) и HCV [75] (9,6 кб). Это было осуществимо на практике, поскольку эти вирусы имеют сравнительно небольшие геномы, для покрытия и сборки полногеномных последовательностей которых требуется всего несколь-

ко ПЦР-ампликонов. Однако, гетерогенность РНК содержащих вирусов, таких как HCV27, норовирусов [19] и вируса бешенства [69], может потребовать использования множественных перекрывающихся сетов праймеров для обеспечения амплификации всех известных геновариантов вируса. Известно, что ПЦР-амплификация более подходит для WGS образцов с низкой концентрацией вирусов, чем метагеномные методы [99], хотя и другие методы, такие как целевое обогащение вирусных последовательностей, могут тоже приводить к хорошим результатам в такого рода образцах. Например, это было показано для норовируса [13].

Перекрывающиеся ПЦР-ампликоны в сочетании с NGS использовались для секвенирования целых геномов больших вирусов, таких как, например, HCMV [84], но этот метод имеет ограниченную масштабируемость, так как требуется синтез множества праймеров и относительно большое количество исходной ДНК. Этот факт ограничивает число подходящих образцов, а соответственно и геномов, которые могут быть получены с использованием такого подхода. Например, для амплификации генома вируса Эболы потребовалось 8–19 продуктов ПЦР [83], а в двух исследованиях норовирусов было необходимо 14 и 22 продуктов ПЦР соответственно [19, 57]. Для получения генома вируса Парамушир понадобилось 60 ПЦР продуктов, а также дополнительные эксперименты по секвенированию с применением прямого метода Сенгера для покрытия недосеквенированных участков [89]. Однако, для клинической диагностики такой подход является проблематичным из-за значительного объема лабораторной нагрузки, связанной с многочисленными ПЦР постановками, необходимости индивидуальной нормализации концентраций различных ПЦР-продуктов до смешивания, высокой вероятности отказа ПЦР-реакции из-за несоответствия праймера матрице (особенно для быстро мутирующих вирусов [89]), а также высокой стоимости расходных материалов и трудозатрат [10]. Поэтому, несмотря на техническую возможность секвенирования вирусов размером вплоть до 250 кб, пропорциональная взаимосвязь между длиной генома и технической сложностью делает секвенирование вирусных геномов более чем 50 кб на основе ПЦР нецелесообразным с использованием современных технологий, в частности для крупных исследований с большим числом образцов или рутинной диагностики. Другая проблема заключается в том, что повышение количества ПЦР-реакций требует соответствующего увеличения количества материала, что не всегда возможно, так как объем клинического образца обычно ограничен. Развитие микрофлюидных

технологий может помочь преодолеть некоторые из этих барьеров. ПЦР на основе микрофлюидики и пулирование большого числа ампликонов были успешно использованы для секвенирования нескольких локусов устойчивости к противомикробным препаратам (например, из микробиома свиней) [47], и могут также применяться для вирусных геномов, потенциально вплоть до уровня единичных геномов [76].

Высоковариабельные патогены, особенно те, которые имеют широко расходящиеся генетические линии и генотипы, такие как HCV [11], вирус гриппа и норовирусы, вызывают серьезные проблемы для ПЦР-амплификации, особенно из-за несоответствия праймеров матрице [13, 19, 99]. Тщательный подход при дизайне структур праймеров может помочь смягчить эти проблемы, но процесс изучения новых вариантов все же остается проблематичным. Стоит также отметить, что в ряде случаев различия между матрицей и праймерами настолько велики, что могут выпадать целые вирусные сегменты, как было показано для вируса Парамушир [89].

### Специфическое обогащение вирусного генетического материала путем гибридизации

Целевое обогащение на зондах (также известное как гибридизация или специфическое обогащение) может быть использовано для секвенирования полных вирусных геномов непосредственно из клинических образцов без необходимости предварительного культивирования или ПЦР [24, 101, 111]. Эти методы обычно включают небольшие РНК- или ДНК-зонды, которые являются комплементарными к референсной последовательности патогенов (или панели эталонных последовательностей). В отличие от методов на основе ПЦР-амплификации, в данном случае реакцию можно проводить в одной пробирке, содержащей перекрывающиеся зонды, которые покрывают весь геном. В реакции гибридизации зонды, связанные с твердой подложкой (например, мечеными стрептавидином магнитными частицами), захватывают или «вытягивают» комплементарные последовательности ДНК из общего пула нуклеиновых кислот, присутствующих в образце. Захват сопровождается последующим лигированием адаптеров, специфичных для секвенатора, и небольшим числом циклов ПЦР для обогащения успешно лигированных фрагментов. Этот подход ранее был успешно использован для характеристики клинически значимых вирусов разного размера, таких как вирус гепатита С [99], HSV-1 [28], вирус ветряной оспы (VZV) [24], вирус Эпштейна–Барр (EBV) [78], вирус герпеса че-

ловека 5 (HCMV) [23], вирус герпеса человека 6 (HHV6) [102] и вирус герпеса человека 7 (HHV7) [26]. Реакция проводится в одной пробирке и, подобно ПЦР на основе микрофлюидики, поддается высокой степени автоматизации [78]. Отсутствие стадии культивирования означает, что полученные последовательности более характерны для исходного вируса, а не для культивируемых вирусных изолятов, соответственно получается меньшее число мутаций, чем в ПЦР-амплифицированных образцах [23, 24]. Успех этого метода очень сильно зависит от доступных эталонных (референсных) вирусных последовательностей. Специфичность увеличивается, когда дизайн ДНК-зондов произведен на основе большого набора референсных последовательностей, поскольку это приводит к лучшему охвату как внутри одного, так и между разными образцами. Целевое обогащение возможно даже когда имеется небольшое несоответствие между матрицей и зондом; однако, в то время как ПЦР-амплификация требует только знания фланкирующих областей целевой области, для целевого обогащения путем гибридизации необходима информация о всей внутренней последовательности для дизайна зондов. Впрочем, даже если какой-то зонд не «срабатывает», его область может по-прежнему захватываться другими (перекрывающимися или соседними) зондами [23, 24]. Недавно, Briese и соавт. [12] разработали платформу для секвенирования вирусных позвоночных (VirCapSeq-VERT), состоящую из ~2 млн биотинилированных ДНК-зондов для целевого обогащения вирусных нуклеиновых кислот, направленную на повышение чувствительности обнаружения вирусных патогенов. Описанный метод позволяет идентифицировать (и в ряде случаев даже осуществлять сборку целых геномов обнаруженных вирусов) большое число вирусов, включая новые (имеющие не менее 60% идентичности с известными геномами) вирусы. Впрочем, общая стоимость секвенирования на образец остается все еще достаточно высокой, а сам метод пока не был широко востребован другими исследовательскими группами, и его эффективность еще предстоит оценить. Однако стоит отметить, что целевое обогащение не подходит для характеристики совершенно новых вирусов, которые имеют низкую гомологию с известными вирусами, и для которых метагеномика, а также в некоторых случаях и ПЦР с использованием вырожденных (смесь аналогичных, но вариативных) праймеров, могут быть более подходящими подходами, хотя и приводящими к очень «шумному» результату. Впрочем, проблему анализа данных, а именно идентификации новых вирусных агентов при отсутствии гомологичных геномов в базах данных, это не решает.

## ДНК-метабаркодинг как способ детекции и идентификации патогенов в биологических образцах

ДНК-баркодинг — метод, который использует короткую часть генома организма (так называемый штрих- или бар-код), чтобы определить принадлежит ли он к определенному семейству, роду или даже виду, используя последующее прочтение фрагмента генома. Этот метод был первоначально разработан для изучения бактериальных сообществ (например, исследования микробиоты кишечника), но сегодня он широко используется для решения самых различных задач, включая даже такие необычные применения как выявление фальсификации в пищевой промышленности [97], изучение диет морских сообществ [9] и анализ биотоплива [45]. Однако, в отличие от других таксонов, вирусы не имеют универсально разделенного филогенетического маркера (как, например, 16S у бактерий, оксидазы цитохрома С для птиц и млекопитающих, *rbcL* и *matK* для растений и внутреннего транскрибируемого спейсера ITS для грибов и растений), что делает невозможным разработку универсальных пар праймеров для амплификации и дифференциации различных вирусных последовательностей. Кроме того, вирусная таксономия зачастую основана на признаках вызываемых заболеваний, а не на генетическом сходстве патогенов. Этот факт усложняет выбор штрих-кода (короткой стандартизированной нуклеотидной последовательности ДНК организма), даже для одного рода (например, *Mammarenavirus* может быть серологически, филогенетически и географически разделен на две основные группы: комплекс «Старого Света», распространенный в Африке, Европе и Азии, а также комплекс «Новый мир», обнаруженный в Северной и Южной Америках [38]), не говоря уже о более высоких таксонах.

В ДНК-метабаркодинге вирусов ПЦР продукты разных постановок (на каждой отдельной паре праймеров) для каждого образца собирают в одну пробирку, очищают, элюируют в минимальном объеме и готовят для последующего NGS секвенирования [20]. Однако этот подход не исключает требования выполнения множества ПЦР реакций на каждый образец, что является проблемой при одновременном изучении большого количества биологического материала. Кроме того, существуют ограничения на использование разных протоколов для подготовки библиотеки, особенно когда протокол включает стадию эмульсионной ПЦР, требующей ПЦР-продуктов с длинами в довольно узких границах. Потому в нашем недавнем исследовании [7] мы внедрили способ дизайна

олигонуклеотидных панелей для целевого обогащения вирусных нуклеиновых кислот, где основной целью являлось использование минимального набора олигонуклеотидов праймеров для покрытия максимального числа различных вирусных таксонов в рамках единой реакции ПЦР. Мы применили этот подход для разработки панели родоспецифических праймеров для целевого обогащения кДНК от зоонозных РНК-вирусов и оценили ее с использованием технологий NGS на образцах перелетных птиц, продемонстрировав значительное обогащение вирусными последовательностями отдельных родов. Так, например, мы сумели добиться того, что до половины всех полученных прочтений в конечном «fastq»-файле соответствовали регионам вирусных геномов, в то время как при секвенировании транскриптомных библиотек, полученных от тех же самых образцов, в ряде случаев мы вообще не смогли обнаружить вирусных фрагментов, либо их доля была крайне мала (тысячные доли процента).

## Технологии секвенирования третьего поколения

Современные технологии NGS второго поколения, такие как Illumina или Ion Torrent, генерируют данные с короткими прочтениями (обычно 200–500 п.н.), что среди прочего создает проблемы при гаплотипировании, то есть определении находятся ли генетические варианты на едином локусе (один вирусный геном или клонал) или на родственных, очень сходных, но все же разных генетических локусах в одной и той же популяции. Более того, с использованием коротких прочтений сложнее различать повторяющиеся и рекомбинантные регионы из-за неоднозначностей при картировании. Клинические последствия понимания того, встречаются ли, например, варианты с множественной лекарственной устойчивостью на одном вирусном геноме или распределяются между смешанной популяцией вирусов, каждая с разными профилями лекарственной устойчивости, в настоящее время неясны.

Хотя для решения этих проблем есть вычислительные инструменты [40], новые технологии могут генерировать более длинные прочтения. Новые, одномолекулярные секвенаторы, такие как PacBio (Pacific Biosciences) и MinION (Oxford Nanopore) — так называемые «секвенаторы третьего поколения», способны производить чрезвычайно длинные прочтения, включая целые вирусные геномы, которые теоретически можно было бы получить из единичных прочтений. Преимущество секвенирования на приборе MinION, помимо его невысокой стоимости

(относительно приборов от компаний Illumina и Thermo Fisher Scientific) и, в то же время, высокой мобильности, заключается в том, что это достаточно быстрый процесс — зачастую требуется менее одного дня, чтобы перейти от полученного образца к отчету по анализируемым данным [92]. Наши собственные эксперименты с прибором MinION показали, что, действительно, применяя эту технологию можно прочитывать полные геномы ВИЧ с точностью, которая, как минимум, не отличается от точности современных полупроводниковых секвенаторов Ion S5 [96].

Недостатком всех NGS методов, в том числе одномолекулярного секвенирования, является потребность в высоком покрытии для минимизации влияния ошибок секвенирования. Это особенно проблематично для исследований резистентности к лекарственным средствам, поскольку резистентность к лекарствам чаще всего проявляется в результате однонуклеотидных мутаций или небольших делеций (1–3 оснований), особенно в некоторых РНК-вирусах [52]. Достижение высокого покрытия, необходимого для обеспечения точного варианта типирования, является сложной технической задачей, особенно когда присутствует много ДНК хозяина по сравнению с вирусными последовательностями, и когда профиль ошибок технологии делает точечные мутации особенно трудными для обнаружения [92]. На момент написания этого обзора секвенирование на MinION имеет «сырое» качество (так называемые «2D-чтения») с частотой ошибок одного случайного чтения примерно 10% (наши собственные наблюдения), что заметно уступает частотам ошибок других технологий (Illumina (< 0,1%), Ion Torrent (~1%), хотя точность и может быть улучшена с использованием циклического консенсусного считывания [62, 100]. Стоит также отметить, что увеличение покрытия с последующим усреднением для устранения случайных ошибок секвенирования также приводит к невозможности обнаружения так называемых минорных вариантов, присутствующих лишь в очень ограниченной доле патогенов, находящихся в образце.

Впрочем, объединение технологий длинных прочтений с целевым обогащением дает потенциальный путь к дальнейшему развитию [29, 49], поскольку часть сложностей может быть разрешена, особенно если для целевого патогена достигнута достаточная глубина секвенирования, а частота ошибок для всех методологий может быть уменьшена за счет дальнейшего технологического и аналитического улучшения. Деплеция нуклеиновых кислот хозяина является другим альтернативным (и зачастую дополнительным) решением, так как более вы-

сокая доля вирусных прочтений будет восстанавливаться при каждом цикле секвенирования. Несмотря на то что для достижения этой цели для бактериального секвенирования уже имеются решения (например, деплеция рибосомальной РНК человека или митохондрий, а также селективная деплеция ДНК с определенной моделью метилирования), для секвенирования вирусов до сих пор подобных методов просто не существует.

## Проблемы, которые предстоит решить в ближайшем будущем

Вопросы чувствительности и контаминации особенно актуальны в секвенировании вирусов из-за риска как ложноотрицательного, так и ложноположительного обнаружения патогенов. Высокочувствительное секвенирование (будь то метагеномное, основанное на ПЦР, или целевом обогащении с помощью зондов) помогает обнаруживать даже небольшие количества загрязняющих вирусных нуклеиновых кислот [42, 73]. Например, вирус мышинной лейкемии [30, 43] и парвовирус-подобные последовательности [74, 87] представляют собой всего лишь два классических примера среди многих контаминационных агентов, которые могут быть получены даже из обычных лабораторных реагентов, таких как колонки для выделения нуклеиновых кислот [90]. Как и в случае с другими высокочувствительными технологиями, для минимизации загрязнения необходимы надежные лабораторные методы и протоколы, во многом схожие, например, с применяемыми при работах с древней ДНК [67], хотя обычно и не настолько жесткие. Также важно помнить, что обнаружение вирусных нуклеиновых кислот не обязательно определяет причину заболевания, и при использовании методов NGS для диагностики вирусных инфекций важно подтверждать результаты с помощью альтернативных независимых методов, которые не зависят от тестирования на присутствие фрагментов ДНК. Например, в случаях энцефалита неизвестного происхождения положительные результаты NGS могут быть подтверждены иммуногистохимическим анализом пораженной ткани [64, 71] или идентификацией вируса с помощью электронной микроскопии или тканевой культуры [39].

Стандартизация методов, включая этап биоинформатического анализа, скорее всего, будет являться основным ключом к успеху NGS секвенирования в клинической вирусологии. Программные пакеты, которые будут использовать графический интерфейс пользователя, предпочтительнее инструментов, требующих экспертных знаний в командной строке Linux.

Кроме того, разработка и создание достоверных баз данных, показывающих какие варианты действительно свидетельствуют о резистентности к лекарственным средствам, будет иметь решающее значение для точной клинической интерпретации. Такие базы данных уже созданы для ВИЧ [93], HBV [35, 86] и HCV [56].

Несмотря на то что существуют серьезные причины для перехода от субгеномного к полногеномному секвенированию и в целом для использования NGS в клинической вирусологии, специалистов диагностических или прибольничных лабораторий необходимо убедить в важности такого перехода, продемонстрировав преимущества наличия дополнительной информации для ведения больных и практической осуществимости полногеномного секвенирования вирусов. Последнее должно обеспечиваться масштабируемым и автоматизируемым рабочим процессом, адекватной нормативной базой и стоимостью, сопоставимой с секвенированием фрагментов.

## Заключение

Секвенирование вирусов с помощью технологий NGS приобретает все большее клиническое значение для диагностики, борьбы с болезнями, молекулярной эпидемиологии и инфекционного контроля. Как уже перечис-

лено выше, существует несколько подходов, доступных для достижения прочтения генетического материала вирусов из клинических образцов, ампликонов, целевого обогащения или метагеномики, и выбор метода специфичен как для вируса, так и для задачи. Метагеномное секвенирование наиболее подходит для диагностического секвенирования неизвестных или плохо охарактеризованных вирусов, секвенирование ПЦР-ампликонов хорошо работает для коротких вирусных геномов и низкого генетического разнообразия в сайтах связывания праймеров, а целевое обогащение подходит для патогенов всех размеров, в особенности для крупных вирусов, а также тех, которые имеют разнообразные, но хорошо охарактеризованные геномы. В настоящее время существуют две очевидные области инноваций: методы, которые могут эффективно устранять ДНК хозяина и бактериальных микроорганизмов, не влияя на вирусную ДНК/РНК, и дальнейшее развитие технологий длинных прочтений для достижения гибкого и конкурентного ценообразования по сравнению с технологиями коротких прочтений. Новые технологии необходимы для объединения сильных сторон этих разных методов, и позволят поставщикам медицинских услуг инвестировать в единую технологию, подходящую для всех вирусологических приложений.

## Список литературы/References

1. Aanensen D.M., Feil E.J., Holden M.T.G., Dordel J., Yeats C.A., Fedosejev A., Goater R., Castillo-Ramírez S., Corander J., Colijn C., Chlebowicz M.A., Schouls L., Heck M., Pluister G., Ruimy R., Kahlmeter G., Åhman J., Matuschek E., Friedrich A.W., Parkhill J., Bentley S.D., Spratt B.G., Grundmann H. Whole-genome sequencing for routine pathogen surveillance in public health: a population snapshot of invasive *Staphylococcus aureus* in Europe. *mBio*, 2016, vol. 7, no. 3. doi: 10.1128/mBio.00444-16
2. Allen U.D., Hu P., Pereira S.L., Robinson J.L., Paton T.A., Beyene J., Khodai-Booran N., Dipchand A., Hébert D., Ng V., Nalpathamkalam T., Read S. The genetic diversity of Epstein–Barr virus in the setting of transplantation relative to non-transplant settings: a feasibility study. *Pediatr. Transplant.*, 2016, vol. 20, no. 1, pp. 124–129. doi: 10.1111/ptr.12610
3. Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. Basic local alignment search tool. *J. Mol. Biol.*, 1990, vol. 215, no. 3, pp. 403–410. doi: 10.1016/S0022-2836(05)80360-2
4. Ambrose H.E., Clewley J.P. Virus discovery by sequence-independent genome amplification. *Rev. Med. Virol.*, 2006, vol. 16, no. 6, pp. 365–383. doi: 10.1002/rmv.515
5. Anthony S.J., Epstein J.H., Murray K.A., Navarrete-Macias I., Zambrana-Torrel C.M., Solovyov A., Ojeda-Flores R., Arrigo N.C., Islam A., Ali Khan S., Hosseini P., Bogich T.L., Olival K.J., Sanchez-Leon M.D., Karesh W.B., Goldstein T., Luby S.P., Morse S.S., Mazet J.A.K., Daszak P., Lipkin W.I. A Strategy to estimate unknown viral diversity in mammals. *mBio*, 2013, vol. 4, no. 5. doi: 10.1128/mBio.00598-13
6. Artimo P., Jonnalagedda M., Arnold K., Baratin D., Csardi G., de Castro E., Duvaud S., Flegel V., Fortier A., Gasteiger E., Grosdidier A., Hernandez C., Ioannidis V., Kuznetsov D., Liechti R., Moretti S., Mostaguir K., Redaschi N., Rossier G., Xenarios I., Stockinger H. ExPASy: SIB bioinformatics resource portal. *Nucl. Acids Res.*, 2012, vol. 40, no. W1, pp. W597–W603. doi: 10.1093/nar/gks400
7. Ayginin A.A., Pimkina E.V., Matsvay A.D., Speranskaya A.S., Safonova M.V., Blinova E.A., Artyushin I.V., Dedkov V.G., Shipulin G.A., Khafizov K. The Study of viral RNA diversity in bird samples using de novo designed multiplex genus-specific primer panels. *Adv. Virol.*, 2018, vol. 2018, pp. 1–10. doi: 10.1155/2018/3248285
8. Bartha I., Carlson J.M., Brumme C.J., McLaren P.J., Brumme Z.L., John M., Haas D.W., Martinez-Picado J., Dalmau J., López-Galíndez C., Casado C., Rauch A., Günthard H.F., Bernasconi E., Vernazza P., Klimkait T., Yerly S., O'Brien S.J., Listgarten J., Pfeifer N., Lippert C., Fusi N., Kotalik Z., Allen T.M., Müller V., Harrigan P.R., Heckerman D., Telenti A., Fellay J., for the HIV Genome-to-Genome Study and the Swiss HIV Cohort Study. A genome-to-genome analysis of associations between human genetic variation, HIV-1 sequence diversity, and viral control. *eLife*, 2013, vol. 2. doi: 10.7554/eLife.01123
9. Berry T.E., Osterrieder S.K., Murray D.C., Coghlan M.L., Richardson A.J., Grealy A.K., Stat M., Bejder L., Bunce M. DNA metabarcoding for diet analysis and biodiversity: a case study using the endangered Australian sea lion (*Neophoca cinerea*). *Ecol. Evol.*, 2017, vol. 7, no. 14, pp. 5435–5453. doi: 10.1002/ece3.3123

10. Bialasiewicz S., McVernon J., Nolan T., Lambert S.B., Zhao G., Wang D., Nissen M.D., Sloots T.P. Detection of a divergent Parainfluenza 4 virus in an adult patient with influenza like illness using next-generation sequencing. *BMC Infect. Dis.*, 2014, vol. 14, no. 1. doi: 10.1186/1471-2334-14-275
11. Bonsall D., Ansari M.A., Ip C., Trebes A., Brown A., Klenerman P., Buck D., STOP-HCV Consortium, Piazza P., Barnes E., Bowden R. ve-SEQ: robust, unbiased enrichment for streamlined detection and whole-genome sequencing of HCV and other highly diverse pathogens. *F1000Research*, 2015, vol. 4, p. 1062. doi: 10.12688/f1000research.7111.1
12. Briese T., Kapoor A., Mishra N., Jain K., Kumar A., Jabado O.J., Lipkin W.I. Virome capture sequencing enables sensitive viral diagnosis and comprehensive virome analysis. *mBio*, 2015, vol. 6, no. 5. doi: 10.1128/mBio.01491-15
13. Brown J.R., Roy S., Ruis C., Yara Romero E., Shah D., Williams R., Breuer J. Norovirus whole-genome sequencing by sureselect target enrichment: a robust and sensitive method. *J. Clin. Microbiol.*, 2016, vol. 54, no. 10, pp. 2530–2537. doi: 10.1128/JCM.01052-16
14. Calvet G., Aguiar R.S., Melo A.S.O., Sampaio S.A., de Filippis I., Fabri A., Araujo E.S.M., de Sequeira P.C., de Mendonça M.C.L., de Oliveira L., Tschoeke D.A., Schrago C.G., Thompson F.L., Brasil P., dos Santos F.B., Nogueira R.M.R., Tanuri A., de Filippis A.M.B. Detection and sequencing of Zika virus from amniotic fluid of fetuses with microcephaly in Brazil: a case study. *Lancet Infect. Dis.*, 2016, vol. 16, no. 6, pp. 653–660. doi: 10.1016/S1473-3099(16)00095-5
15. Castrignano S.B., Nagasse-Sugahara T.K., Kisielius J.J., Ueda-Ito M., Brandão P.E., Curti S.P. Two novel circo-like viruses detected in human feces: complete genome sequencing and electron microscopy analysis. *Virus Res.*, 2013, vol. 178, no. 2, pp. 364–373. doi: 10.1016/j.virusres.2013.09.018
16. Chen Y., Yao H., Thompson E.J., Tannir N.M., Weinstein J.N., Su X. VirusSeq: software to identify viruses and their integration sites using next-generation sequencing of human cancer tissue. *Bioinformatics*, 2013, vol. 29, no. 2, pp. 266–267. doi: 10.1093/bioinformatics/bts665
17. Choi S.K., Choi J.K., Park W.M., Ryu K.H. RT-PCR detection and identification of three species of cucumoviruses with a genus-specific single pair of primers. *J. Virol. Methods*, 1999, vol. 83, no. 1–2, pp. 67–73. doi: 10.1016/S0166-0934(99)00106-8
18. Clarke S., Innocenti G.M. Organization of immature intrahemispheric connections. *J. Comp. Neurol.*, 1986, vol. 251, no. 1, pp. 1–22. doi: 10.1002/cne.902510102
19. Cotten M., Petrova V., Phan M.V.T., Rabaa M.A., Watson S.J., Ong S.H., Kellam P., Baker S. Deep sequencing of Norovirus genomes defines evolutionary patterns in an urban tropical setting. *J. Virol.*, 2014, vol. 88, no. 19, pp. 11056–11069. doi: 10.1128/JVI.01333-14
20. De Vries M., Deijs M., Canuti M., van Schaik B.D.C., Faria N.R., van de Garde M.D.B., Jachimowski L.C.M., Jebbink M.F., Jakobs M., Luyf A.C.M., Coenjaerts F.E.J., Claas E.C.J., Molenkamp R., Koekkoek S.M., Lammens C., Leus F., Goossens H., Ieven M., Baas F., van der Hoek L. A sensitive assay for virus discovery in respiratory clinical samples. *PLoS One*, 2011, vol. 6, no. 1: e16118. doi: 10.1371/journal.pone.0016118
21. Dedkov V.G., Lukashev A.N., Deviatkin A.A., Kuleshov K.V., Safonova M.V., Poleshchuk E.M., Drexler J.F., Shipulin G.A. Retrospective diagnosis of two rabies cases in humans by high throughput sequencing. *J. Clin. Virol.*, 2016, vol. 78, pp. 74–81. doi: 10.1016/j.jcv.2016.03.012
22. Denesvre C., Dumarest M., Rémy S., Gourichon D., Eloit M. Chicken skin virome analyzed by high-throughput sequencing shows a composition highly different from human skin. *Virus Genes*, 2015, vol. 51, no. 2, pp. 209–216. doi: 10.1007/s11262-015-1231-8
23. DePew J., Zhou B., McCarrison J.M., Wentworth D.E., Purushe J., Koroleva G., Fouts D.E. Sequencing viral genomes from a single isolated plaque. *Virol. J.*, 2013, vol. 10, no. 1: 181. doi: 10.1186/1743-422X-10-181
24. Depledge D.P., Kundu S., Jensen N.J., Gray E.R., Jones M., Steinberg S., Gershon A., Kinchington P.R., Schmid D.S., Balloux F., Nichols R.A., Breuer J. Deep sequencing of viral genomes provides insight into the evolution and pathogenesis of Varicella Zoster virus and its vaccine in humans. *Mol. Biol. Evol.*, 2014, vol. 31, no. 2, pp. 397–409. doi: 10.1093/molbev/mst210
25. Depledge D.P., Palser A.L., Watson S.J., Lai I.Y.-C., Gray E.R., Grant P., Kanda R.K., Leproust E., Kellam P., Breuer J. Specific capture and whole-genome sequencing of viruses from clinical samples. *PLoS One*, 2011, vol. 6, no. 11: e27805. doi: 10.1371/journal.pone.0027805
26. Djikeng A., Halpin R., Kuzmickas R., DePasse J., Feldblyum J., Sengamalay N., Afonso C., Zhang X., Anderson N.G., Ghedin E., Spiro D.J. Viral genome sequencing by random priming methods. *BMC Genomics*, 2008, vol. 9, no. 1: 5. doi: 10.1186/1471-2164-9-5
27. Donaldson C.D., Clark D.A., Kidd I.M., Breuer J., Depledge D.D. Genome sequence of Human Herpesvirus 7 strain UCL-1. *Genome Announc.*, 2013, vol. 1, no. 5. doi: 10.1128/genomeA.00830-13
28. Drosten C., Günther S., Preiser W., van der Werf S., Brodt H.-R., Becker S., Rabenau H., Panning M., Kolesnikova L., Fouchier R.A.M., Berger A., Burguière A.-M., Cinatl J., Eickmann M., Escriou N., Grywna K., Kramme S., Manuguerra J.-C., Müller S., Rickerts V., Stürmer M., Vieth S., Klenk H.-D., Osterhaus A.D.M.E., Schmitz H., Doerr H.W. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N. Engl. J. Med.*, 2003, vol. 348, no. 20, pp. 1967–1976. doi: 10.1056/NEJMoa030747
29. Ebert K., Depledge D.P., Breuer J., Harman L., Elliott G. Mode of virus rescue determines the acquisition of VHS mutations in VP22-negative Herpes Simplex Virus 1. *J. Virol.*, 2013, vol. 87, no. 18, pp. 10389–10393. doi: 10.1128/JVI.01654-13
30. Eckert S.E., Chan J.Z.-M., Houniet D., Breuer J., Speight G. Enrichment of long DNA fragments from mixed samples for Nanopore sequencing. *bioRxiv: The preprint server for biology*, 2016. doi: 10.1101/048850
31. Erlwein O., Robinson M.J., Dustan S., Weber J., Kaye S., McClure M.O. DNA extraction columns contaminated with murine sequences. *PLoS One*, 2011, vol. 6, no. 8: e23484. doi: 10.1371/journal.pone.0023484
32. Faria N.R., Azevedo R.D.S.D.S., Kraemer M.U.G., Souza R., Cunha M.S., Hill S.C., Theze J., Bonsall M.B., Bowden T.A., Rissanen I., Rocco I.M., Nogueira J.S., Maeda A.Y., Vasami F.G.D.S., Macedo F.L.D.L., Suzuki A., Rodrigues S.G., Cruz A.C.R., Nunes B.T., Medeiros D.B.D.A., Rodrigues D.S.G., Nunes Queiroz A.L., Silva E.V.P.D., Henriques D.F., Travassos da Rosa E.S., de Oliveira C.S., Martins L.C., Vasconcelos H.B., Casseb L.M.N., Simith D.D.B., Messina J.P., Abade L., Lourenco J., Alcantara L.C.J., Lima M.M.D., Giovanetti M., Hay S.I., de Oliveira R.S., Lemos P.D.S., Oliveira L.F.D., de Lima C.P.S., da Silva S.P., Vasconcelos J.M.D., Franco L., Cardoso J.F., Vianez-Junior J.L.D.S.G., Mir D., Bello G., Delatorre E., Khan K., Creatore M., Coelho G.E., de Oliveira W.K., Tesh R., Pybus O.G., Nunes M.R.T., Vasconcelos P.F.C. Zika virus in the Americas: Early epidemiological and genetic findings. *Science*, 2016, vol. 352, no. 6283, pp. 345–349. doi: 10.1126/science.aaf5036

33. Fonseca N.A., Rung J., Brazma A., Marioni J.C. Tools for mapping high-throughput sequencing data. *Bioinformatics*, 2012, vol. 28, no. 24, pp. 3169–3177. doi: 10.1093/bioinformatics/bts605
34. Gardy J.L., Naus M., Amlani A., Chung W., Kim H., Tan M., Severini A., Krajden M., Puddicombe D., Sahni V., Hayden A.S., Gustafson R., Henry B., Tang P. Whole-genome sequencing of measles virus genotypes H1 and D8 during outbreaks of infection following the 2010 Olympic Winter Games reveals viral transmission routes. *J. Infect. Dis.*, 2015, vol. 212, no. 10, pp. 1574–1578. doi: 10.1093/infdis/jiv271
35. Geoghegan J.L., Holmes E.C. Predicting virus emergence amid evolutionary noise. *Open Biol.*, 2017, vol. 7, no. 10. doi: 10.1098/rsob.170189
36. Gnanesan S., Ijaz S., Moran J., Ramsay M., Green J. HepSEQ: international public health repository for hepatitis B. *Nucleic Acids Res.*, 2007, vol. 35, pp. D367–D370. doi: 10.1093/nar/gkl874
37. Goodacre N., Aljanahi A., Nandakumar S., Mikailov M., Khan A.S. A Reference Viral Database (RVDB) to enhance bioinformatics analysis of high-throughput sequencing for novel virus detection. *mSphere*, 2018, vol. 3, no. 2. doi: 10.1128/mSphereDirect.00069-18
38. Grabherr M.G., Haas B.J., Yassour M., Levin J.Z., Thompson D.A., Amit I., Adiconis X., Fan L., Raychowdhury R., Zeng Q., Chen Z., Mauceli E., Hacohen N., Gnirke A., Rhind N., di Palma F., Birren B.W., Nusbaum C., Lindblad-Toh K., Friedman N., Regev A. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.*, 2011, vol. 29, no. 7, pp. 644–652. doi: 10.1038/nbt.1883
39. Günther S., Lenz O. Lassa virus. *Crit. Rev. Clin. Lab. Sci.*, 2004, vol. 41, no. 4, pp. 339–390. doi: 10.1080/10408360490497456
40. Hall R.J., Draper J.L., Nielsen F.G.G., Dutilh B.E. Beyond research: a primer for considerations on using viral metagenomics in the field and clinic. *Front. Microbiol.*, 2015, vol. 6. doi: 10.3389/fmicb.2015.00224
41. Hong L.Z., Hong S., Wong H., Aw P., Yan C., Wilm A., de Sessions P.F., Lim S., Nagarajan N., Hibberd M.L., Quake S.R., Burkholder W.F. BAsE-Seq: a method for obtaining long viral haplotypes from short sequence reads. *Genome Biol.*, 2014, vol. 15, no. 11: 517. doi: 10.1186/PREACCEPT-6768001251451949
42. Houldcroft C.J., Beale M.A., Breuer J. Clinical and biological insights from viral genome sequencing. *Nat. Rev. Microbiol.*, 2017, vol. 15, no. 3, pp. 183–192. doi: 10.1038/nrmicro.2016.182
43. Houldcroft C.J., Breuer J. Tales from the crypt and coral reef: the successes and challenges of identifying new herpesviruses using metagenomics. *Front. Microbiol.*, 2015, vol. 6. doi: 10.3389/fmicb.2015.00188
44. Hue S., Gray E.R., Gall A., Katzourakis A., Tan C.P., Houldcroft C.J., McLaren S., Pillay D., Futreal A., Garson J.A., Pybus O.G., Kellam P., Towers G.J. Disease-associated XMRV sequences are consistent with laboratory contamination. *Retrovirology*, 2010, vol. 7, no. 1: 111. doi: 10.1186/1742-4690-7-111
45. Hulo C., de Castro E., Masson P., Bougueleret L., Bairoch A., Xenarios I., Le Mercier P. ViralZone: a knowledge resource to understand virus diversity. *Nucleic Acids Res.*, 2011, vol. 39, suppl. 1, pp. D576–D582. doi: 10.1093/nar/gkq901
46. Jaenicke S., Ander C., Bekel T., Bisdorf R., Dröge M., Gartemann K.-H., Jünemann S., Kaiser O., Krause L., Tille F., Zakrzewski M., Pühler A., Schlüter A., Goesmann A. Comparative and joint analysis of two metagenomic datasets from a biogas fermenter obtained by 454-pyrosequencing. *PLoS One*, 2011, vol. 6, no. 1: e14519. doi: 10.1371/journal.pone.0014519
47. Jensen R.H., Møllerup S., Mourier T., Hansen T.A., Fridholm H., Nielsen L.P., Willerslev E., Hansen A.J., Vinner L. Target-dependent enrichment of virions determines the reduction of high-throughput sequencing in virus discovery. *PLoS One*, 2015, vol. 10, no. 4: e0122636. doi: 10.1371/journal.pone.0122636
48. Johnson T.A., Stedtfeld R.D., Wang Q., Cole J.R., Hashsham S.A., Looft T., Zhu Y.-G., Tiedje J.M. Clusters of antibiotic resistance genes enriched together stay together in swine agriculture. *mBio*, 2016, vol. 7, no. 2. doi: 10.1128/mBio.02214-15
49. Jones B.A., Grace D., Kock R., Alonso S., Rushton J., Said M.Y., McKeever D., Mutua F., Young J., McDermott J., Pfeiffer D.U. Zoonosis emergence linked to agricultural intensification and environmental change. *Proc. Natl. Acad. Sci.*, 2013, vol. 110, no. 21, pp. 8399–8404. doi: 10.1073/pnas.1208059110
50. Karamitros T., Magiorkinis G. A novel method for the multiplexed target enrichment of MinION next generation sequencing libraries using PCR-generated baits. *Nucleic Acids Res.*, 2015, vol. 43, no. 22, pp. e152–e152. doi: 10.1093/nar/gkv773
51. Karkhah A., Nouri H.R., Javanian M., Koppolu V., Masrouh-Roudsari J., Kazemi S., Ebrahimpour S. Zika virus: epidemiology, clinical aspects, diagnosis, and control of infection. *Eur. J. Clin. Microbiol. Infect. Dis.*, 2018, vol. 37, no. 11, pp. 2035–2043. doi: 10.1007/s10096-018-3354-z
52. Kent W.J. BLAT — the BLAST-like alignment tool. *Genome Res.*, 2002, vol. 12, no. 4, pp. 656–664. doi: 10.1101/gr.229202
53. Kimberlin D.W., Whitley R.J. Antiviral resistance: mechanisms, clinical significance, and future implications. *J. Antimicrob. Chemother.*, 1996, vol. 37, no. 3, pp. 403–421.
54. Kireev D.E., Lopatukhin A.E., Murzakova A.V., Pimkina E.V., Speranskaya A.S., Neverov A.D., Fedonin G.G., Fantin Y.S., Shipulin G.A. Evaluating the accuracy and sensitivity of detecting minority HIV-1 populations by Illumina next-generation sequencing. *J. Virol. Methods*, 2018, vol. 261, pp. 40–45. doi: 10.1016/j.jviromet.2018.08.001
55. Kohl C., Brinkmann A., Dabrowski P.W., Radonić A., Nitsche A., Kurth A. Protocol for metagenomic virus detection in clinical specimens. *Emerg. Infect. Dis.*, 2015, vol. 21, no. 1, pp. 48–57. doi: 10.3201/eid2101.140766
56. Kruppa J., Jo W.K., van der Vries E., Ludlow M., Osterhaus A., Baumgaertner W., Jung K. Virus detection in high-throughput sequencing data without a reference genome of the host. *Infect. Genet. Evol.*, 2018, vol. 66, pp. 180–187. doi: 10.1016/j.meegid.2018.09.026
57. Kuiken C., Yusim K., Boykin L., Richardson R. The Los Alamos hepatitis C sequence database. *Bioinformatics*, 2005, vol. 21, no. 3, pp. 379–384. doi: 10.1093/bioinformatics/bth485
58. Kundu S., Lockwood J., Depledge D.P., Chaudhry Y., Aston A., Rao K., Hartley J.C., Goodfellow I., Breuer J. Next-generation whole genome sequencing identifies the direction of norovirus transmission in linked patients. *Clin. Infect. Dis.*, 2013, vol. 57, no. 3, pp. 407–414. doi: 10.1093/cid/cit287
59. Langmead B., Salzberg S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, 2012, vol. 9, no. 4, pp. 357–359. doi: 10.1038/nmeth.1923
60. Langmead B., Trapnell C., Pop M., Salzberg S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, 2009, vol. 10, no. 3: R25. doi: 10.1186/gb-2009-10-3-r25

61. Lecuit M., Eloit M. The diagnosis of infectious diseases by whole genome next generation sequencing: a new era is opening. *Front. Cell. Infect. Microbiol.*, 2014, vol. 4. doi: 10.3389/fcimb.2014.00025
62. Lee W.-P., Stromberg M.P., Ward A., Stewart C., Garrison E.P., Marth G.T. MOSAIK: a hash-based algorithm for accurate next-generation sequencing short-read mapping. *PLoS One*, 2014, vol. 9, no. 3: e90581. doi: 10.1371/journal.pone.0090581
63. Li C.-X., Shi M., Tian J.-H., Lin X.-D., Kang Y.-J., Chen L.-J., Qin X.-C., Xu J., Holmes E.C., Zhang Y.-Z. Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. *eLife*, 2015, vol. 4. doi: 10.7554/eLife.05378
64. Li C., Chng K.R., Boey E.J.H., Ng A.H.Q., Wilm A., Nagarajan N. INC-Seq: accurate single molecule reads using nanopore sequencing. *GigaScience*, 2016, vol. 5, no. 1. doi: 10.1186/s13742-016-0140-7
65. Lipkin W.I. A Vision for investigating the microbiology of health and disease. *J. Infect. Dis.*, 2015, vol. 212, suppl. 1, pp. S26–S30. doi: 10.1093/infdis/jiu649
66. Liu P., Fang X., Feng Z., Guo Y.-M., Peng R.-J., Liu T., Huang Z., Feng Y., Sun X., Xiong Z., Guo X., Pang S.-S., Wang B., Lv X., Feng F.-T., Li D.-J., Chen L.-Z., Feng Q.-S., Huang W.-L., Zeng M.-S., Bei J.-X., Zhang Y., Zeng Y.-X. Direct sequencing and characterization of a clinical isolate of Epstein–Barr virus from nasopharyngeal carcinoma tissue by using next-generation sequencing technology. *J. Virol.*, 2011, vol. 85, no. 21, pp. 11291–11299. doi: 10.1128/JVI.00823-11
67. Matranga C.B., Andersen K.G., Winnicki S., Busby M., Gladden A.D., Tewhey R., Stremlau M., Berlin A., Gire S.K., England E., Moses L.M., Mikkelsen T.S., Odiya I., Ehiane P.E., Folarin O., Goba A., Kahn S.H., Grant D.S., Honko A., Hensley L., Happi C., Garry R.F., Malboeuf C.M., Birren B.W., Gnirke A., Levin J.Z., Sabeti P.C. Enhanced methods for unbiased deep sequencing of Lassa and Ebola RNA viruses from clinical and biological samples. *Genome Biol.*, 2014, vol. 15, no. 11. doi: 10.1186/s13059-014-0519-7
68. Matsvay A.D., Alborova I.E., Pimkina E.V., Markelov M.L., Khafizov K., Mustafin K.K. Experimental approaches for ancient DNA extraction and sample preparation for next generation sequencing in ultra-clean conditions. *Conserv. Genet. Resour.*, 2018. doi: 10.1007/s12686-018-1016-1
69. Mbisa J.L., Fearnhill E., Dunn D.T., Pillay D., Asboe D., Cane P.A. Evidence of self-sustaining drug resistant HIV-1 lineages among untreated patients in the United Kingdom. *Clin. Infect. Dis.*, 2015, vol. 61, no. 5, pp. 829–836. doi: 10.1093/cid/civ393
70. Miia J.-V., Tiina N., Tarja S., Olli V., Liisa S., Anita H. Evolutionary trends of European bat lyssavirus type 2 including genetic characterization of Finnish strains of human and bat origin 24 years apart. *Arch. Virol.*, 2015, vol. 160, no. 6, pp. 1489–1498. doi: 10.1007/s00705-015-2424-0
71. Mlakar J., Korva M., Tul N., Popović M., Poljšak-Prijatelj M., Mraz J., Kolenc M., Resman Rus K., Vesnaver Vipotnik T., Fabjan Vodusek V., Vizjak A., Pizem J., Petrovec M., Avšič Županc T. Zika virus associated with microcephaly. *N. Engl. J. Med.*, 2016, vol. 374, no. 10, pp. 951–958. doi: 10.1056/NEJMoa1600651
72. Morfopoulou S., Brown J.R., Davies E.G., Anderson G., Virasami A., Qasim W., Chong W.K., Hubank M., Plagnol V., Desforges M., Jacques T.S., Talbot P.J., Breuer J. Human Coronavirus OC43 associated with fatal encephalitis. *N. Engl. J. Med.*, 2016, vol. 375, no. 5, pp. 497–498. doi: 10.1056/NEJMc1509458
73. Mulcahy-O’Grady H., Workentine M.L. The challenge and potential of metagenomics in the clinic. *Front. Immunol.*, 2016, vol. 7. doi: 10.3389/fimmu.2016.00029
74. Munro A.C., Houldcroft C. Human cancers and mammalian retroviruses: should we worry about bovine leukemia virus? *Future Virol.*, 2016, vol. 11, no. 3, pp. 163–166. doi: 10.2217/fvl.16.5
75. Naccache S.N., Greninger A.L., Lee D., Coffey L.L., Phan T., Rein-Weston A., Aronsohn A., Hackett J., Delwart E.L., Chiu C.Y. The perils of pathogen discovery: origin of a novel parvovirus-like hybrid genome traced to nucleic acid extraction spin columns. *J. Virol.*, 2013, vol. 87, no. 22, pp. 11966–11977. doi: 10.1128/JVI.02323-13
76. Newman R.M., Kuntzen T., Weiner B., Berical A., Charlebois P., Kuiken C., Murphy D.G., Simmonds P., Bennett P., Lennon N.J., Birren B.W., Zody M.C., Allen T.M., Henn M.R. Whole genome pyrosequencing of rare hepatitis C virus genotypes enhances subtype classification and identification of naturally occurring drug resistance variants. *J. Infect. Dis.*, 2013, vol. 208, no. 1, pp. 17–31. doi: 10.1093/infdis/jis679
77. Ocwieja K.E., Sherrill-Mix S., Mukherjee R., Custers-Allen R., David P., Brown M., Wang S., Link D.R., Olson J., Travers K., Schadt E., Bushman F.D. Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing. *Nucleic Acids Res.*, 2012, vol. 40, no. 20, pp. 10345–10355. doi: 10.1093/nar/gks753
78. Oude Munnink B.B., Jazaeri Farsani S.M., Deijns M., Jonkers J., Verhoeven J.T.P., Ieven M., Goossens H., de Jong M.D., Berkhout B., Loens K., Kellam P., Bakker M., Canuti M., Cotten M., van der Hoek L. Autologous antibody capture to enrich immunogenic viruses for viral discovery. *PLoS One*, 2013, vol. 8, no. 11: e78454. doi: 10.1371/journal.pone.0078454
79. Palser A.L., Grayson N.E., White R.E., Corton C., Correia S., Baabdullah M.M., Watson S.J., Cotten M., Arrand J.R., Murray P.G., Allday M.J., Rickinson A.B., Young L.S., Farrell P.J., Kellam P. Genome diversity of Epstein–Barr Virus from multiple tumor types and normal infection. *J. Virol.*, 2015, vol. 89, no. 10, pp. 5222–5237. doi: 10.1128/JVI.03614-14
80. Parameswaran P., Charlebois P., Tellez Y., Nunez A., Ryan E.M., Malboeuf C.M., Levin J.Z., Lennon N.J., Balmaseda A., Harris E., Henn M.R. Genome-wide patterns of intrahuman dengue virus diversity reveal associations with viral phylogenetic clade and interhost diversity. *J. Virol.*, 2012, vol. 86, no. 16, pp. 8546–8558. doi: 10.1128/JVI.00736-12
81. Pfeiffer M., Proebster B., Kinney R.M., Kaaden O.R. Genus-specific detection of alphaviruses by a semi-nested reverse transcription-polymerase chain reaction. *Am. J. Trop. Med. Hyg.*, 1997, vol. 57, no. 6, pp. 709–718.
82. Pickett B.E., Sadat E.L., Zhang Y., Noronha J.M., Squires R.B., Hunt V., Liu M., Kumar S., Zaremba S., Gu Z., Zhou L., Larson C.N., Dietrich J., Klem E.B., Scheuermann R.H. ViPR: an open bioinformatics database and analysis resource for virology research. *Nucleic Acids Res.*, 2012, vol. 40, no. D1, pp. D593–D598. doi: 10.1093/nar/gkr859
83. Power R.A., Davaniah S., Derache A., Wilkinson E., Tanser F., Gupta R.K., Pillay D., de Oliveira T. Genome-Wide Association Study of HIV Whole Genome Sequences Validated using Drug Resistance. *PLoS One*, 2016, vol. 11, no. 9: e0163746. doi: 10.1371/journal.pone.0163746
84. Quick J., Loman N.J., Duraffour S., Simpson J.T., Severi E., Cowley L., Bore J.A., Koundouno R., Dudas G., Mikhail A., Ouédraogo N., Afrough B., Bah A., Baum J.H.J., Becker-Ziaja B., Boettcher J.P., Cabeza-Cabrerizo M., Camino-Sánchez Á., Carter L.L., Doerrbecker J., Enkirch T., Dorival I.G., Hetzelt N., Hinzmann J., Holm T., Kafetzopoulou L.E., Koropogui M., Kosgey A., Kuisma E., Logue C.H., Mazzarelli A., Meisel S., Mertens M., Michel J., Ngabo D., Nitzsche K., Pallasch E.,

- Patrono L.V., Portmann J., Repits J.G., Rickett N.Y., Sachse A., Singethan K., Vitoriano I., Yemanaberhan R.L., Zekeng E.G., Racine T., Bello A., Sall A.A., Faye O., Faye O., Magassouba N., Williams C.V., Amburgey V., Winona L., Davis E., Gerlach J., Washington F., Monteil V., Jourdain M., Bererd M., Camara A., Somlare H., Camara A., Gerard M., Bado G., Baillet B., Delaune D., Nebie K.Y., Diarra A., Savane Y., Pallawo R.B., Gutierrez G.J., Milhano N., Roger I., Williams C.J., Yattara F., Lewandowski K., Taylor J., Rachwal P., J. Turner D., Pollakis G., Hiscox J.A., Matthews D.A., Shea M.K.O., Johnston A.M., Wilson D., Hutley E., Smit E., Di Caro A., Wölfel R., Stoecker K., Fleischmann E., Gabriel M., Weller S.A., Koivogui L., Diallo B., Keita S., Rambaut A., Formenty P., Günther S., Carroll M.W. Real-time, portable genome sequencing for Ebola surveillance. *Nature*, 2016, vol. 530, no. 7589, pp. 228–232. doi: 10.1038/nature16996
85. Renzette N., Bhattacharjee B., Jensen J.D., Gibson L., Kowalik T.F. Extensive genome-wide variability of Human cytomegalovirus in congenitally infected infants. *PLoS Pathog.*, 2011, vol. 7, no. 5: e1001344. doi: 10.1371/journal.ppat.1001344
86. Reyes G.R., Kim J.P. Sequence-independent, single-primer amplification (SISPA) of complex DNA populations. *Mol. Cell. Probes*, 1991, vol. 5, no. 6, pp. 473–481. doi: 10.1016/S0890-8508(05)80020-9
87. Rhee S.-Y., Margeridon-Thermet S., Nguyen M.H., Liu T.F., Kagan R.M., Beggel B., Verheyen J., Kaiser R., Shafer R.W. Hepatitis B virus reverse transcriptase sequence variant database for sequence analysis and mutation discovery. *Antiviral Res.*, 2010, vol. 88, no. 3, pp. 269–275. doi: 10.1016/j.antiviral.2010.09.012
88. Rosseel T., Pardon B., De Clercq K., Ozhelvaci O., Van Borm S. False-positive results in metagenomic virus discovery: a strong case for follow-up diagnosis. *Transbound. Emerg. Dis.*, 2014, vol. 61, no. 4, pp. 293–299. doi: 10.1111/tbed.12251
89. Sabina J., Leamon J.H. Bias in whole genome amplification: causes and considerations. In: Whole genome amplification; ed. Kroneis T. *New York, NY: Springer New York*, 2015, vol. 1347, pp. 15–41. doi: 10.1007/978-1-4939-2990-0\_2
90. Safonova M.V., Shchelkanov M.Y., Khafizov K., Matsvay A.D., Ayginin A.A., Dolgova A.S., Shchelkanov E.M., Pimkina E.V., Speranskaya A.S., Galkina I.V., Dedkov V.G. Sequencing and genetic characterization of two strains Paramushir virus obtained from the Tyuleniy Island in the Okhotsk Sea (2015). *Ticks Tick-Borne Dis.*, 2018. doi: 10.1016/j.tbd.2018.11.004
91. Salter S.J., Cox M.J., Turek E.M., Calus S.T., Cookson W.O., Moffatt M.F., Turner P., Parkhill J., Loman N.J., Walker A.W. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.*, 2014, vol. 12, no. 1. doi: 10.1186/s12915-014-0087-z
92. Sauvage V., Eloit M. Viral metagenomics and blood safety. *Transfus. Clin. Biol.*, 2016, vol. 23, no. 1, pp. 28–38. doi: 10.1016/j.traci.2015.12.002
93. Schmidt K., Mwaigwisya S., Crossman L.C., Doumith M., Munroe D., Pires C., Khan A.M., Woodford N., Saunders N.J., Wain J., O'Grady J., Livermore D.M. Identification of bacterial pathogens and antimicrobial resistance directly from clinical urines by nanopore-based metagenomic sequencing. *J. Antimicrob. Chemother.*, 2017, vol. 72, no. 1, pp. 104–114. doi: 10.1093/jac/dkw397
94. Shafer R.W. Rationale and uses of a public HIV drug-resistance database. *J. Infect. Dis.*, 2006, vol. 194, suppl. 1, pp. S51–S58. doi: 10.1086/505356
95. Shi M., Lin X.-D., Chen X., Tian J.-H., Chen L.-J., Li K., Wang W., Eden J.-S., Shen J.-J., Liu L., Holmes E.C., Zhang Y.-Z. The evolutionary history of vertebrate RNA viruses. *Nature*, 2018, vol. 556, no. 7700, pp. 197–202. doi: 10.1038/s41586-018-0012-7
96. Shi M., Lin X.-D., Tian J.-H., Chen L.-J., Chen X., Li C.-X., Qin X.-C., Li J., Cao J.-P., Eden J.-S., Buchmann J., Wang W., Xu J., Holmes E.C., Zhang Y.-Z. Redefining the invertebrate RNA virosphere. *Nature*, 2016, vol. 540, no. 7634, pp. 539–543. doi: 10.1038/nature20167
97. Speranskaya A.S., Khafizov K., Ayginin A.A., Krinitsina A.A., Omelchenko D.O., Nilova M.V., Severova E.E., Samokhina E.N., Shipulin G.A., Logacheva M.D. Comparative analysis of Illumina and Ion Torrent high-throughput sequencing platforms for identification of plant components in herbal teas. *Food Control*, 2018, vol. 93, pp. 315–324. doi: 10.1016/j.foodcont.2018.04.040
98. Speranskaya A.S., Lopatukhin A.E., Khafizov K., Ayginin A.A., Korneenko E.V., Kireev D.E., Shipulin G.A. Evaluation of MinION nanopore platform for HIV whole coding regions sequencing. *Bioinformatics of genome regulation and structure/systems biology: The Eleventh International Conference*, 2018. doi: 10.18699/BGRSSB-2018-059
99. Stano M., Beke G., Klucar L. viruSITE — integrated database for viral genomics. *Database*, 2016, vol. 2016, p. baw162. doi: 10.1093/database/baw162
100. Thomson E., Ip C.L.C., Badhan A., Christiansen M.T., Adamson W., Ansari M.A., Bibby D., Breuer J., Brown A., Bowden R., Bryant J., Bonsall D., Da Silva Filipe A., Hinds C., Hudson E., Klenerman P., Lythgow K., Mbisa J.L., McLauchlan J., Myers R., Piazza P., Roy S., Trebes A., Sreenu V.B., Witteveldt J., STOP-HCV Consortium, Barnes E., Simmonds P. Comparison of next-generation sequencing technologies for comprehensive assessment of full-length hepatitis C viral genomes. *J. Clin. Microbiol.*, 2016, vol. 54, no. 10, pp. 2470–2484. doi: 10.1128/JCM.00330-16
101. Travers K.J., Chin C.-S., Rank D.R., Eid J.S., Turner S.W. A flexible and efficient template format for circular consensus sequencing and SNP detection. *Nucleic Acids Res.*, 2010, vol. 38, no. 15, pp. e159–e159. doi: 10.1093/nar/gkq543
102. Tsangaras K., Wales N., Sicheritz-Pontén T., Rasmussen S., Michaux J., Ishida Y., Morand S., Kampmann M.-L., Gilbert M.T.P., Greenwood A.D. Hybridization capture using short PCR products enriches small genomes by capturing flanking sequences (CapFlank). *PLoS One*, 2014, vol. 9, no. 10: e109101. doi: 10.1371/journal.pone.0109101
103. Tweedy J., Spyrou M.A., Donaldson C.D., Depledge D., Breuer J., Gompels U.A. Complete genome sequence of the Human Herpesvirus 6A strain AJ from Africa resembles strain GS from North America. *Genome Announc.*, 2015, vol. 3, no. 1. doi: 10.1128/genomeA.01498-14
104. UFO Sequencing Consortium within the I-BFM Study Group, Forster M., Szymczak S., Ellinghaus D., Hemmrich G., Rühlemann M., Kraemer L., Mucha S., Wienbrandt L., Stanulla M., Franke A. Vy-PER: eliminating false positive detection of virus integration events in next generation sequencing data. *Sci. Rep.*, 2015, vol. 5, no. 1. doi: 10.1038/srep11534
105. VanDevanter D.R., Warren P., Bennett L., Schultz E.R., Coulter S., Garber R.L., Rose T.M. Detection and analysis of diverse herpesviral species by consensus primer PCR. *J. Clin. Microbiol.*, 1996, vol. 34, no. 7, pp. 1666–1671.
106. Venter J.C. Environmental genome shotgun sequencing of the sargasso sea. *Science*, 2004, vol. 304, no. 5667, pp. 66–74. doi: 10.1126/science.1093857
107. Wang Q., Jia P., Zhao Z. VirusFinder: Software for efficient and accurate detection of viruses and their integration sites in host genomes through next generation sequencing data. *PLoS One*, 2013, vol. 8, no. 5: e64465. doi: 10.1371/journal.pone.0064465

108. Watson S.J., Langat P., Reid S.M., Lam T.T.-Y., Cotten M., Kelly M., Van Reeth K., Qiu Y., Simon G., Bonin E., Foni E., Chiapponi C., Larsen L., Hjulsgaard C., Markowska-Daniel I., Urbaniak K., Dürrwald R., Schlegel M., Huovilainen A., Davidson I., Dán Á., Loeffen W., Edwards S., Bublot M., Vila T., Maldonado J., Valls L., ESNIP3 Consortium, Brown I.H., Pybus O.G., Kellam P. Molecular epidemiology and evolution of influenza viruses circulating within European swine between 2009 and 2013. *J. Virol.*, 2015, vol. 89, no. 19, pp. 9920–9931. doi: 10.1128/JVI.00840-15
109. Weiss S., Witkowski P.T., Auste B., Nowak K., Weber N., Fahr J., Mombouli J.-V., Wolfe N.D., Drexler J.F., Drosten C., Klempa B., Leendertz F.H., Kruger D.H. Hantavirus in bat, Sierra Leone. *Emerg. Infect. Dis.*, 2012, vol. 18, no. 1, pp. 159–161. doi: 10.3201/eid1801.111026
110. Worthey E.A., Mayer A.N., Syverson G.D., Helbling D., Bonacci B.B., Decker B., Serpe J.M., Dasu T., Tschannen M.R., Veith R.L., Basehore M.J., Broeckel U., Tomita-Mitchell A., Arca M.J., Casper J.T., Margolis D.A., Bick D.P., Hessner M.J., Routes J.M., Verbsky J.W., Jacob H.J., Dimmock D.P. Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. *Genet. Med.*, 2011, vol. 13, no. 3, pp. 255–262. doi: 10.1097/GIM.0b013e3182088158
111. Wylie T.N., Wylie K.M., Herter B.N., Storch G.A. Enhanced virome sequencing using targeted sequence capture. *Genome Res.*, 2015, vol. 25, no. 12, pp. 1910–1920. doi: 10.1101/gr.191049.115
112. Zhang Y.-Z., Shi M., Holmes E.C. Using metagenomics to characterize an expanding virosphere. *Cell*, 2018, vol. 172, no. 6, pp. 1168–1172. doi: 10.1016/j.cell.2018.02.043
113. Zhao G., Krishnamurthy S., Cai Z., Popov V.L., Travassos da Rosa A.P., Guzman H., Cao S., Virgin H.W., Tesh R.B., Wang D. Identification of novel viruses using ViruHunter — an automated data analysis pipeline. *PLoS One*, 2013, vol. 8, no. 10: e78470. doi: 10.1371/journal.pone.0078470
114. Zheng L., Tang J., Clover G.R.G., Spackman M.E., Freeman A.J., Rodoni B.C. Novel genus-specific broad range primers for the detection of furoviruses, hordeiviruses and rymoviruses and their application in field surveys in South-east Australia. *J. Virol. Methods*, 2015, vol. 214, pp. 1–9. doi: 10.1016/j.jviromet.2014.11.022

**Авторы:**

**Хафизов К.Ф.**, к.б.н., руководитель научной группы, ФБУН ЦНИИ эпидемиологии Роспотребнадзора, Москва, Россия; зав. лабораторией разработки новых методов геномики ФГБУ Центр стратегического планирования и управления медико-биологическими рисками здоровью МЗ РФ, Москва, Россия;

**Сперанская А.С.**, к.б.н., научный сотрудник отдела молекулярной диагностики и эпидемиологии ФБУН ЦНИИ эпидемиологии Роспотребнадзора, Москва, Россия; научный сотрудник кафедры высших растений Московского государственного университета имени М.В. Ломоносова, Москва, Россия;

**Мацвай А.Д.**, младший научный сотрудник отдела молекулярной диагностики, научная группа разработки новых методов диагностики на основе секвенирования следующего поколения ФБУН ЦНИИ эпидемиологии Роспотребнадзора, Москва, Россия; лаборант лаборатории исторической генетики, радиоуглеродного анализа и прикладной физики Московского физико-технического института, г. Долгопрудный, Московская область, Россия;

**Шипулин Г.А.**, к.м.н., зам. директора по научно-производственной работе ФГБУ Центр стратегического планирования и управления медико-биологическими рисками здоровью МЗ РФ, Москва, Россия;

**Дедков В.Г.**, к.м.н., зам. директора по науке ФБУН НИИ эпидемиологии и микробиологии имени Пастера, Санкт-Петербург, Россия; ведущий научный сотрудник Института медицинской паразитологии, тропических и трансмиссивных заболеваний им. Е.И. Марциновского, Москва, Россия.

**Authors:**

**Khafizov K.F.**, PhD (Biology), Scientific Group Leader, Central Research Institute of Epidemiology, Moscow, Russian Federation; Head of the Laboratory for the Development of New Genomics Methods, Center of Strategic Planning, Moscow, Russian Federation;

**Speranskaya A.S.**, PhD, Researcher, Department of Molecular Diagnostics and Epidemiology, Central Research Institute of Epidemiology, Moscow, Russia; Researcher, Department of Higher Plants, Lomonosov Moscow State University, Moscow, Russian Federation;

**Matsvay A.D.**, Junior Researcher, Department of Molecular Diagnostics, Group for the Development of New Diagnostic Methods Based on Next-Generation Sequencing, Central Research Institute of Epidemiology, Moscow, Russian Federation; Laboratory Assistant, Laboratory of Historical Genetics, Radiocarbon Analysis and Applied Physics, Moscow Institute of Physics and Technology, Dolgoprudny, Moscow Region, Russian Federation;

**Shipulin G.A.**, PhD (Medicine), Deputy Director on Science and Production, Center of Strategic Planning, Moscow, Russian Federation;

**Dedkov V.G.**, PhD (Medicine), Deputy Director on Science, St. Petersburg Pasteur Institute, St. Petersburg, Russian Federation; Leading Researcher, Martinsonsky Institute of Medical Parasitology, Tropical and Vector-Borne Diseases, Moscow, Russian Federation.

Поступила в редакцию 14.12.2018  
Отправлена на доработку 27.05.2019  
Принята к печати 04.06.2019

Received 14.12.2018  
Revision received 27.05.2019  
Accepted 04.06.2019