



Современные аппаратные и программные решения для полногеномного секвенирования, перспективы их внедрения в практику войск радиационной, химической и биологической защиты Вооруженных Сил Российской Федерации

А.А. Петров, А.В. Казанцев, Е.А. Ковальчук, М.Ю. Павлюков, А.В. Сапкулов, Д.А. Кутаев, С.В. Борисевич✉

Федеральное государственное бюджетное учреждение
«48 Центральный научно-исследовательский институт»
Министерства обороны Российской Федерации
141306, Российская Федерация, г. Сергиев Посад-6, ул. Октябрьская, д. 11
✉ e-mail: 48cnii@mil.ru

Одним из современных направлений совершенствования лабораторной диагностики инфекционных болезней является внедрение методов секвенирования геномов опасных патогенов. Среди них наибольшее распространение в настоящее время приобрел метод секвенирования следующего поколения (Next Generation Sequencing, NGS).

Цель работы – определить перспективные направления и проблемы внедрения методов NGS в практику войск РХБ защиты Вооруженных Сил Российской Федерации (войск РХБ защиты ВС РФ).

Источниковая база исследования – данные, опубликованных в отечественных и англоязычных научных изданиях, доступных через сеть Интернет.

Метод исследования – аналитический.

Результаты. Представлены материалы, обобщающие данные о трех поколениях технологий секвенирования (по Сэнгеру, NGS, секвенирования на основе нанопор), рассмотрены их преимущества и недостатки применительно к использованию в практике войск РХБ защиты ВС РФ. Показано, что последние тенденции в инструментах биоинформатики для анализа геномных данных сосредоточены на машинном обучении, облачных приложениях, удобных интерфейсах и программном обеспечении с открытым исходным кодом. Использование параллельных вычислений и оптимизация биоинформатических алгоритмов позволяют повысить эффективность обработки данных NGS секвенирования при неизменности технических характеристик используемого оборудования. Но поскольку большинство биоинформатических программ имеют открытый код, их сертификация органами военного управления представляется маловероятной.

Вывод. Необходимым условием успешного внедрения NGS в практику подразделений войск РХБ защиты ВС РФ является решение двух основных проблем: 1) разработка отечественной аппаратной части (компьютеров достаточной мощности), сертифицированных для защиты информации в установленном порядке и при этом имеющих полный функционал; 2) разработка соответствующих биоинформатических инструментов и баз данных, сертифицированных для защиты информации в установленном порядке и функционирующих в автономном режиме без подключения к сети интернет.

Ключевые слова: биоинформационные инструменты, геном, ДНК, патоген, Сенгер, секвенирование следующего поколения; секвенаторы

Для цитирования: Петров А.А., Казанцев А.В., Ковальчук Е.А., М.Ю. Павлюков, Сапкулов А.В., Кутаев Д.А., Борисевич С.В. Современные аппаратные и программные решения для полногеномного секвенирования, перспективы их внедрения в практику войск радиационной, химической и биологической защиты Вооруженных Сил Российской Федерации. Вестник войск РХБ защиты. 2024;8(2):164–175. EDN:obanjc.
<https://doi.org/10.35825/2587-5728-2024-8-2-164-175>

Прозрачность финансовой деятельности: авторы не имеют финансовой заинтересованности в представленных материалах или методах.

Конфликт интересов: нет.

Финансирование: Государственное бюджетное учреждение «48 Центральный научно-исследовательский институт» Министерства обороны Российской Федерации

Поступила 08.07.2023 г. После доработки 05.06.2024 г. Принята к публикации 28.06.2024 г.

Modern Hardware and Software Solutions for Whole-Genome Sequencing, Prospects of Their Implementation in the Practice of Nuclear, Chemical and Biological Protection Troops of the Armed Forces of the Russian Federation

Aleksandr A. Petrov, Aleksey V. Kazantsev, Elena A. Kovalchuk, Mikhail Yu. Pavlyukov, Aleksey V. Sapkulov, Dmitriy A. Kutaev, Sergey V. Borisevich✉

48 Central Scientific Research Institute
of the Ministry of Defense of the Russian Federation,
Oktyabrskaya St., 11, Sergiev Posad-6 141306, Russian Federation
✉ e-mail: 48cnii@mil.ru

One of the promising areas for improving laboratory diagnostics is the active introduction of next-generation sequencing (NGS), which retains the key advantages of molecular tests and allows the detection of a wide range of diverse microorganisms. Despite the significant advantages of NGS in the laboratory diagnosis of infectious diseases, there are also a number of objective obstacles to its widespread use.

The purpose of the work is to consider existing hardware and software solutions for whole-genome sequencing, as well as the prospects and problems of their implementation in the practice of the RCB protection troops of the Armed Forces of the Russian Federation for the identification of pathogens of various etiologies.

The source base of the research includes an analysis of data published in domestic and English-language scientific publications, as well as posted on the Internet.

The research method is analytical.

Results. Materials are presented that summarize data on three generations of sequencing technologies, their advantages and disadvantages. It is shown that with the development of NGS technologies, many technical problems arise that need to be deeply analyzed and solved. Despite its great potential, NGS has not yet been widely adopted in clinical microbiology laboratories, partly due to the lack of standardized workflows.

Conclusions. For the successful use of NGS by units of the RCB protection troops of the Armed Forces of the Russian Federation, it is necessary to solve the main problem, which is the development of domestic hardware, appropriate bioinformatics tools and databases certified in accordance with the established procedure for functioning offline without connecting to the Internet.

Keywords: bioinformatics tools, next-generation sequencing, sequencers

For citation: Petrov A.A., Kazantsev A.V., Kovalchuk E.A., Pavlyukov M.Yu., Sapkulov A.V., Kutaev D.A., Borisevich S.V. Modern Hardware and Software Solutions for Whole-Genome Sequencing, Prospects of Their Implementation in the Practice of Nuclear, Chemical and Biological Protection Troops of the Armed Forces of the Russian Federation. *Journal of NBC Protection Corps*. 2024;8(2):164–175. EDN:obanjc.
<https://doi.org/10.35825/2587-5728-2024-8-2-164-175>

Financial disclosure: The authors have no financial interests in the submitted materials or methods.

Conflict of interest statement: The authors declare no conflict of interest.

Funding: 48 Central Scientific Research Institute of the Ministry of Defence of the Russian Federation.

Received July 27, 2023. Revised June 10, 2024. Accepted June 27, 2024

позволит реконструировать более длинный участок генома из двух и более прочтений методом наложения с максимальным соответствием. Данный подход в конечном итоге позволяет реконструировать геном (частично или полностью) *de novo* [2].

Анализ полученных таким образом данных позволяет выявить новые варианты патогенов, информация о геномах которых до момента проведения исследования отсутствует.

Компания Ion Torrent (США) продвигает на рынке одноименный полупроводниковый прибор, в основе которого по факту лежит высокочувствительный рН-метр, позволяющий осуществлять одновременное независимое измерение рН в большом количестве реакционных ячеек [5]. Прибор детектирует и документирует изменения рН, которые выделяются во время полимеризации ДНК, напрямую оцифровывая биологическую молекулу ДНК. К преимуществам таких секвенаторов следует отнести идеологически простые и предсказуемые реакционные взаимодействия (стандартные реактивы, высокие требования только к чистоте дезоксинуклеотидтрифосфатов). Основным недостатком является эффект недостоверного прочтения гомополимерных участков длиной более 3–4 одинаковых нуклеотидов, что ведет к появлению в рядах различных вариантов так называемых инделов (однонуклеотидных вставок и делеций).

Технологию флуоресцентного секвенирования с помощью мостиковой ПЦР (клональной амплификации) с формированием кластеров ДНК, или нанопор ДНК, наработанных посредством линейной репликации по типу катящегося кольца, развивает компания Illumina (США) и BGI Group (КНР) [6]. Данная технология основана на использовании так называемых обратимых терминаторов, по факту представляющих собой дезоксинуклеотидтрифосфаты, несущие различные флуорофоры ковалентно связанные с гетероциклами нуклеотидных оснований с помощью метилазидных линкеров. На каждом шаге синтеза цепи ДНК происходит встраивание комплементарного матрице дезоксинуклеотидтрифосфата, что приводит к остановке синтеза, поскольку 3'-конец растущей молекулы блокирован и недоступен для присоединения следующего основания. После регистрации флуоресценции в конце шага синтеза происходит специфическое удаление линкера вместе с флуорофором и 3'-конец растущей молекулы снова готов к следующему шагу синтеза. Российским аналогом, эквивалентным по характеристикам

прибору MiSeq компании Illumina является секвенатор ДНК «Нанофор СПС», созданный коллективом авторов Института аналитического приборостроения РАН и компании «Синтол».

Ключевыми преимуществами таких технологий являются высокая точность и производительность, основными недостатками – высокая себестоимость, длительный рабочий цикл и сложная подготовка пробы к анализу.

Огромный массив данных, получаемый исследователем на выходе из полногеномного секвенатора, обуславливает потребность в соответствующих вычислительных мощностях. Общепринятым высокоэффективным подходом к решению данной задачи является аренда вычислительной мощности облачных серверов сети интернет. Очевидно, что данный подход недоступен для российской военной науки по ряду соображений, наиболее значимыми из которых являются режимные. Компромиссным решением может являться использование компьютеров достаточной мощности, позволяющих обрабатывать информацию без доступа к сети интернет и ее вычислительным ресурсам.

Секвенирование третьего поколения

Компания Oxford Nanopore Technologies (Великобритания) выпустила на рынок технологию секвенирования третьего поколения на основе так называемых нанопор [7]. Нанопора представляет собой отверстие нанометрового размера (ионный канал) сквозь который может одновременно проходить только одна молекула ДНК и только определенным образом. При приложении электрического поля ДНК стремится пройти сквозь нанопору, меняя ее ионную проводимость. Изменение ионной проводимости характерно и специфично для каждого из нуклеотидов и именно оно является пошагово регистрируемым прибором аналитическим эффектом. Нанопоровый секвенатор MinION представлен на *рисунке 1*.

Ключевые преимущества данной технологии: возможность сверхдлинных прочтений ДНК; высокая производительность; отсутствие сложных этапов подготовки проб; и относительно низкая себестоимость. Недостатки – низкая точность; высокие требования к качеству пробы, в частности, несмотря на заявления производителя, анализ проб объектов окружающей среды не представляется возможным по причине преимущественного занятия нанопор доминирующими в смеси ДНК, в то время как емкость чипа достаточно ограничена. Следовательно, в условиях ограничения количества пор и



Рисунок 1 – Нанопоровый секвенатор MinION.
 Управление и получение данных осуществляется с помощью смартфона (URL: https://www.skygen.com/katalog/oborudovanie/oxford_nanopore_technologies/nanoporovyy_sekvenator_minion/;
 дата обращения: 07.07.2023)

отсутствия амплификации, вероятность выявить незначительные количества ДНК патогенов на фоне полного доминирования ДНК банальной микрофлоры крайне мала.
 Таким образом, в настоящее время доступно несколько платформ секвенирования, каждая со своими уникальными функциями, которые предоставляют альтернативные варианты механизмов секвенирования, длины считывания, времени выполнения, простоты использования и масштабируемости (таблица 1) [8].

При имеющихся особенностях в технологических процессах предварительной подготовки образцов нуклеиновых кислот NGS-платформ, общим для них требованием остается создание набора (пула, библиотеки) фрагментов, которые будут в дальнейшем служить матрицей для синтеза комплементарной цепи ДНК, в процессе которого будет происходить детекция последовательно присоединяющихся нуклеотидов [9, 10]. Общая для всех процедура проведения NGS-секвенирования представлена на рисунке 2.
 Основными этапами подготовки ДНК для NGS-секвенирования являются: 1) фрагментация целевых последовательностей до желаемой длины, 2) присоединение синтетических олигонуклеотидных адаптеров к концам фрагментов и 3) анализ библиотек для секвенирования.
 Для фрагментации ДНК доступны три подхода: физический, ферментативный и химический. Фрагментация ДНК обычно выполняется физическими (обработка ультразвуком) или ферментативными методами (например, неспецифическая обработка специальным ферментом тагментазой) [11]. Сравнения результатов секвенирования библиотек, фрагментированных при помощи обработки ультразвуком по сравнению с

Таблица 1 – Сравнение различных NGS платформ [8]

Производитель / наименование оборудования	Общая продолжительность анализа, ч	Длина прочтения (рида), нуклеотидов	Технология секвенирования	Преимущества	Недостатки
Платформы секвенирования второго поколения					
Illumina (NovaSeq 6000) (США) Нанофор СПС (Россия)	44	500	Секвенирование путем синтеза	Высокая пропускная способность, более низкая стоимость исследования по сравнению с другими платформами	Короткая длина ридов, сложная пробоподготовка, необходимость ПЦР-амплификации, длительное время работы
Thermo Fisher Scientific (Ion GeneStudio S5 Prime) (США)	12	600	Полупроводниковое секвенирование		
BGI/Complete Genomics (Китай)	40	200	Формирование наношариков ДНК и секвенирование путем синтеза		
Платформы секвенирования третьего поколения					
PacBio (HiFi Reads) (США)	30	20 000	Одномолекулярное секвенирование	Отсутствие ПЦР-амплификации, требуется меньше исходного материала, низкий процент ошибок при подготовке библиотек	Высокая частота ошибок секвенирования, для обеспечения качества сверхдлинных ридов требуется ДНК высокого качества
Nanopore (PromethION) (Великобритания)	72	25 000	Нанопоровое секвенирование		

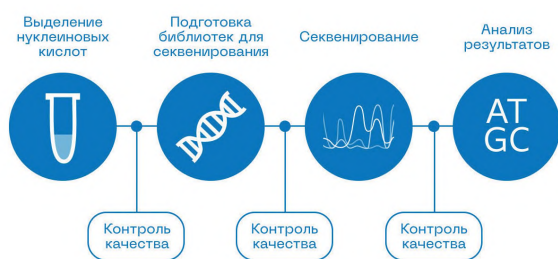


Рисунок 2 – Этапы проведения NGS-секвенирования
(URL: <https://www.skygen.com/podderzhka/obzory/3-etapa-kontrolya-kachestva-kotorye-dolzhen-znat-lyuboy-ngs-issledovatel/>;
дата обращения: 24.01.2023)

приготовленных с помощью трансфозазы, показывают, что ультразвуковой метод является более надежным и воспроизводимым, однако требует соответствующего оборудования [12].

Для возможности одновременного анализа нескольких образцов ДНК используется подход мультиплексирования библиотек целью которого является присвоение каждой из библиотек специальных коротких нуклеотидных кодов (индексов). Для этого используются адаптеры с набором олигонуклеотидных последовательностей, введенных в их структуру (8–18 нуклеотидными индексами), позволяющими после этапа секвенирования дифференцировать массивы данных различных образцов. Индексы присутствуют непосредственно в структуре адаптеров или могут быть введены на этапе амплификации ПЦР с использованием различных индекс-содержащих праймеров для ПЦР для амплификации различных образцов в рамках процесса пробоподготовки [10].

Исследование геномов изолированных и культивированных патогенов является традиционным подходом, позволяющим осуществить обогащение пробы нуклеиновыми кислотами изучаемого микроорганизма. Применительно к бактериям обогащенная проба в идеальном случае представляет собой колонию, снятую с плотной питательной среды. Препарат ДНК, полученный из подобной пробы содержит только бактериальную ДНК и пригоден для анализа любым из методов секвенирования.

Применительно к вирусам обогащенная проба представляет собой препарат, содержащий нативные и дефектные вирионы, а также элементы среды накопления (культуры клеток, материал животных и т.д.). Препарат, полученный из подобной пробы, содержит нуклеиновые кислоты вируса на фоне колоссального количества генетиче-

ского материала среды накопления и в исходном виде пригоден для анализа только после амплификации либо с использованием высокопроизводительных секвенаторов. Для решения данной проблемы применяют методы избирательного удаления нуклеиновых кислот среды накопления из пробы, например, обработка ДНКазой для удаления ДНК с последующим избирательным удалением рибосомальных РНК с использованием гибридизации на зондах.

В большинстве случаев этап культивирования требует дополнительного времени и специализированных условий, если вообще возможен (методы культивирования многих патогенов не разработаны), поэтому NGS с момента своего появления применяется для исследований смешанных образцов в рамках метагеномного подхода. Метагеномный подход на основе технологии NGS позволяет идентифицировать и проанализировать гены и геномы микроорганизмов, присутствующих в различных гетерогенных образцах окружающей среды или биологических материалов, без необходимости культивирования этих микроорганизмов [13–16]. При исследовании произвольного смешанного образца из него выделяется препарат тотальных ДНК/РНК который в дальнейшем используется для подготовки библиотеки, секвенируется и анализируется на предмет выявления в полученном массиве данных прочтений, соответствующих ранее исследованным аннотированным геномам [17–19] или, в случае ранее не исследованных, сборка и аннотация *de novo* [20, 21]. Результатом анализа смешанных образцов в большинстве случаев является отчет, содержащий данные об относительных долях нуклеиновых кислот микроорганизмов (рисунок 3).

Технология NGS имеет ряд ограничений, препятствующих ее широкому применению в диагностической практике. Наиболее очевидным ограничением технологии NGS является скорость проведения анализа, так как весь процесс является многоэтапным и занимает несколько дней, отсутствие стандартизированных рабочих процессов, проблемы, связанные с анализом сложных геномных данных, а также высокая стоимость используемых материалов. Кроме того, аналитические подходы к реконструкции и классификации геномов из смешанных образцов остаются ограниченными по своей производительности и удобству использования [22]. Необходимым условием получения хоть каких-либо практически применимых результатов является высокая квалификация персонала одновременно в нескольких научных

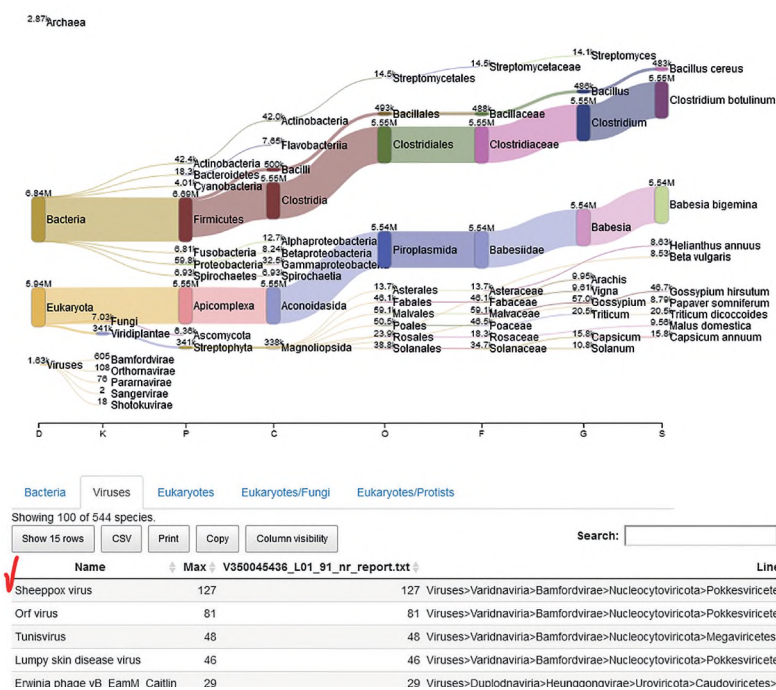


Рисунок 3 – Результаты NGS-секвенирования смешанного образца
(Представлены результаты практического исследования с использованием платформы DNBSEQ-G50, BGI, Китай). Данные авторов

областях – молекулярной биологии, биоинформатики, анализа данных. Наконец, для надежной отчетности необходимы сертифицированные в установленном порядке базы данных патогенов и биоинформатические инструменты анализа последовательностей, позволяющие минимизировать влияние человеческого фактора на результат анализа.

Для получения высококачественных результатов секвенирования оптимизация пробоподготовки имеет решающее значение и требует системной интеграции различных технологий и методов. Например, обнаружение патогенов с низким титром в клинически значимых образцах затруднено из-за абсолютного доминирования в образце генетического материала человека [17]. Образцы окружающей среды часто содержат разнообразные микробные популяции и компоненты, которые мешают обработке образцов и подготовке библиотеки для секвенирования.

Разработка универсального метода обнаружения патогенов от образца до получения результата требует выбора оптимального метода выделения бактерий и вирусов, очистки ДНК/РНК, уменьшения количества балластных нуклеиновых кислот (например, рибосомальной РНК прокариот и эукариот) и концентрирование патогенов с низкими концентрациями в среде культивирования.

В работе К. Parker с соавт. [23] представлено стандартизированное решение для обнаружения возбудителей инфекционных заболеваний. Рабочий процесс PanGIA (Pan-Genomics for Infectious Agents) был оптимизирован для использования конечными пользователями при минимизации времени ручной обработки, чтобы обеспечить 24-часовой цикл от образца до ответа (рисунок 4). Стандартизированные операционные процедуры обнаружения патогенов включают в себя этапы выделения нуклеиновых кислот из исследуемых проб, удаления ДНК человека из клинических проб, подготовки библиотек и одновременного секвенирования РНК- и ДНК-мишеней с использованием платформы секвенирования второго поколения компании Illumina (США) с последующим биоинформационным анализом полученных данных.

В настоящее время полногеномное секвенирование позволяет получать большие массивы данных, которые необходимы обрабатывать, поэтому одним из важных элементов внедрения метода NGS в практику лабораторной диагностики является проработка контроля качества исследования, в том числе на уровне биоинформатической обработки данных. Инструменты биоинформатики являются неотъемлемой частью секвенирования и анализа геномных данных.

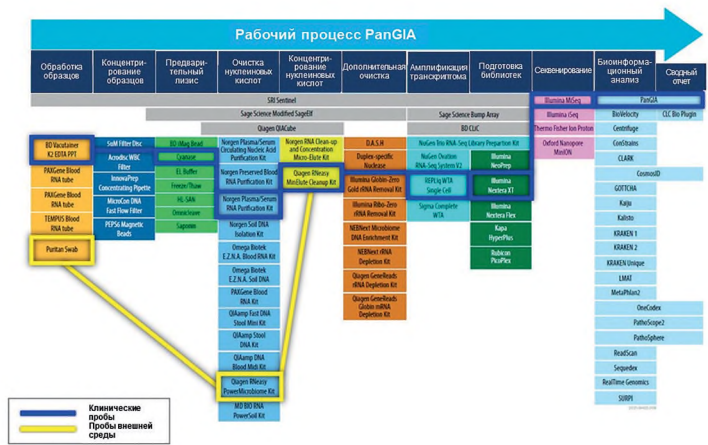


Рисунок 4 – Стандартизированное решение для обнаружения возбудителей инфекционных заболеваний (рабочий процесс PanGIA) [23]

Для контроля качества и предварительной обработки первичных данных, полученных из секвенатора, фильтрации нецелевых последовательностей, сборки и контроля качества прочтений, а также таксономического профилирования последовательностей можно использовать широкий спектр аналитических инструментов. До сих пор не существует единого мнения, какой из подходов следует использовать для решения конкретных задач анализа данных.

Первым этапом анализа является контроль качества полученных из секвенатора данных, который осуществляется путем устранения технических ошибок. Основной целью данного этапа является предварительная обработка данных для удаления последовательностей адаптера, чрезмерно коротких и некачественных прочтений (ридов) или нуклеотидов, которые могут присутствовать в образце, с использованием определенных программ в зависимости от анализируемых данных. После выявления технических ошибок в первичных данных секвенирования, с помощью биоинформатических инструментов осуществляется их удаление, чтобы избежать получения ложноположительных и ложноотрицательных результатов. Наиболее часто применяемые с этой целью биоинформатические инструменты приведены в таблице 2 [24].

Вторым этапом является фильтрация, т.е. удаление балластных последовательностей с использованием биоинформатических инструментов для картирования последовательностей и для дальнейшей сборки генома. После того, как геном собран, следует определить качество сборки. Инструменты для этой цели можно разделить на две основные категории: те, которые требуют эта-

лонных геномов (референсов) для контроля качества, и те, которые не нуждаются в них (таблица 2) [24].

Важным шагом для идентификации вирусов является выполнение таксономического профилирования (установление видов патогенов, последовательности геномов которых были получены в ходе секвенирования). Существует два основных метода решения этой задачи: первый заключается в классификации прочтений в соответствии с таксономиями, а второй – в установлении таксономических групп по набору перекрывающихся последовательностей фрагментов ДНК (так называемым контигам).

Для идентификации вирусных геномов программы анализа используют справочные базы данных на основе различных алгоритмов и программ (на основе k-меров, например, kraken2 [25], bracken [26], CLARK [27]; на основе генов-маркеров, например, MetaPhlAn4 [28], IGGsearch [29]; на архитектуре искусственных нейронных сетей, например, VirusTaxo [30], Metavir2 [31]), отличающихся по времени проведения анализа и объему вычислительных ресурсов. Существуют также биоинформационными инструменты для идентификации вирусных последовательностей без использования референсов (такие как VirSorter [32] и VirFinder [33]).

Ключевую роль при обработке данных высокопроизводительного секвенирования играют разработка и внедрение алгоритмов биоинформатики, выполняемых в определенной последовательности. Разработка биоинформационного алгоритма контроля качества образцов, включающего в себя инструменты, базирующиеся на различных подходах, позволит значительно

Таблица 2 – Программные решения, наиболее часто применяемые для обработки данных полногеномного секвенирования

Функция	Программа	Тип данных	Источник
Контроль качества первичных данных из секвенатора	FastQC	Короткие риды	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
	MinionQC	Длинные риды	https://github.com/roblanf/minion_qc
Редактирование первичных данных из секвенатора, удаление технических ошибок и служебных последовательностей (адаптеров)	Trimomatic	Короткие риды	http://www.usadellab.org/cms/?page=trimmomatic
	Fastp	Короткие риды	https://github.com/OpenGene/fastp
	NanoPack	Длинные риды	https://github.com/wdecoster/nanopack
	SequelTools	Длинные риды	https://github.com/ISUgenomics/SequelTools
Картирование последовательностей	BWA	Короткие риды, длинные риды	https://bio-bwa.sourceforge.net
	Minimap2	Длинные риды	https://github.com/lh3/minimap2
Сборка генома	MEGAHIT	Короткие риды	https://github.com/voutcn/megahit
	metaSPADES	Короткие риды	https://github.com/ablab/spades
	NECAT	Длинные риды (Nanopore)	https://github.com/xiaochuanle/NECAT
Контроль качества сборки последовательностей	MetaQuast	На основе референса	http://bioinf.spbau.ru/metaquast
	DeepMAseD	Без референса	https://github.com/leylabmpi/DeepMAseD

упростить как оценку качества образцов, так и выявление нестандартных данных, требующих особого внимания на стадии интерпретации.

Последние тенденции в инструментах биоинформатики для анализа геномных данных сосредоточены на машинном обучении, облачных приложениях, удобных интерфейсах и программном обеспечении с открытым исходным кодом. Использование параллельных вычислений и оптимизация биоинформатических алгоритмов позволяют повысить эффективность обработки данных NGS секвенирования при неизменности технических характеристик используемого оборудования.

Современные программные решения для полногеномного секвенирования представляют несомненный интерес для подразделений войск РХБ защиты ВС РФ, выполняющих задачи по биологической защите войск и населения Российской Федерации от биологических угроз, в том числе и вызванных ранее неизвестными патогенами.

Общепринятым высокоэффективным подходом к использованию рассмотренных выше программных продуктов является их скачивание и постоянное обновление с помощью сети Интернет. Очевидно, что данный подход недоступен для военной науки в связи с невозможностью использования несертифицированного программного обеспечения и ряду других причин. Поскольку большин-

ство биоинформатических программ имеют открытый код, их сертификация органами военного управления представляется маловероятной.

Компромиссным решением может являться разработка отечественной промышленностью соответствующих биоинформатических инструментов, сертифицированных органами военного управления в установленном порядке и способных функционировать в автономном режиме без подключения к сети интернет. В противном случае внедрение современных методов генетического анализа в практику войск не представляется возможным.

Заключение

Основные направления внедрения NGS в практику подразделений войск РХБ защиты ВС РФ следующие: идентификация патогенных биологических агентов путем расшифровки их генома (генотипирование), обнаружение новых биологических угроз и осуществление молекулярного эпидемиологического надзора за вспышками опасных и особо опасных инфекционных заболеваний (геномный мониторинг) на территориях дислокации ВС РФ.

Принципиально необходимым условием успешного внедрения полногеномного секвенирования в практику подразделений войск РХБ защиты ВС РФ является решение двух основных проблем, имеющих общий источник:

- разработка отечественной аппаратной части (компьютеров достаточной мощности), сертифицированных для защиты информации в установленном порядке и при этом имеющих полный функционал;

- разработка соответствующих биоинформатических инструментов и баз данных, сертифицированных для защиты информации в установленном порядке и функционирующих в автономном режиме без подключения к сети Интернет.

Список источников/References

1. Sozhamannan S, Holland MY, Hall AT, Negrón DA, Ivancich M, Koehler JW, et al. Evaluation of Signature Erosion in Ebola Virus Due to Genomic Drift and Its Impact on the Performance of Diagnostic Assays. *Viruses*. 2015;7(6):3130–54.
<https://doi.org/10.3390/v7062763>
2. Sohn JI, Nam JW. The present and future of de novo whole-genome assembly. *Briefings in Bioinformatics*. 2018;19(1):23–40.
<https://doi.org/10.1093/bib/bbw096>
3. Бородин АГ, Манойлов ВВ, Заруцкий ИВ, Петров АИ, Курочкин ВЕ. Поколения методов секвенирования ДНК (обзор). *Научное приборостроение*. 2020;30(4):3–20.
Borodinov AG, Manoilov VV, Zarutsky IV, Petrov AI, Kurochkin VE. Generations of DNA sequencing methods (review). *Nauchnoe priborostroenie*. 2020;30(4):3–20 (in Russian).
4. Алексеев ЯИ, Белов ЮВ, Малюченко ОП, Монахова ЮА, Натыров АН, Орехов ВА и др. Генетический анализатор для фрагментного анализа ДНК. *Научное приборостроение*. 2012;22(4):86–92.
Alekseev YaI, Belov YuV, Malyuchenko OP, Monakhova YuA, Natyrov AN, Orekhov VA, et al. Genetic analyzer for DNA fragment analysis. *Nauchnoe priborostroenie*. 2012;22(4):86–92 (in Russian).
5. Зубов ВВ. Секвенирование по Ротбергу (потенциал полупроводникового секвенирования). *Биомика*. 2013;5(1–2):48–61.
Zubov VV. Rotberg sequencing (potential of semiconductor sequencing). *Biomica*. 2013;5(1–2):48–61 (in Russian).
6. Зубов ВВ, Чемерис ДА, Васильев РГ, Курочкин ВЕ, Алексеев ЯИ. Краткая история методов высокопроизводительного секвенирования нуклеиновых кислот. *Биомика*. 2021;13(1):27–46.
<https://doi.org/10.31301/2221-6197.bmcs.2021-4>
Zubov VV, Chemeris DA, Vasilov RG, Kurochkin VE, Alekseev YaI. Brief history of high-throughput nucleic acid sequencing methods. *Biomics*. 2021;13(1):27–46 (in Russian).
<https://doi.org/10.31301/2221-6197.bmcs.2021-4>
7. MacKenzie M, Argyropoulos C. An introduction to nanopore sequencing: past, present, and future considerations. *Micromachines*. 2023;14(2):459.
<https://doi.org/10.3390/mi14020459>
8. Pervez M, Hasnain MJ, Abbas SH, Moustafa MF, Aslam N, Shah SSM. A Comprehensive Review of Performance of Next-Generation Sequencing Platforms. *Biomed Res Int*. 2022; 2022:3457806.
<https://doi.org/10.1155/2022/3457806>
9. Head SR, Komori HK, LaMere SA, Whisenant T, Van Nieuwerburgh F, Salomon DR, Ordoukhanian P. Library construction for next-generation sequencing: overviews and challenges. *Biotechniques*. 2014;56(2):61–77.
<https://doi.org/10.2144/000114133>
10. Marine R, Polson SW, Ravel J, Hatfull G, Russell D, Sullivan M, et al. Evaluation of a transposase protocol for rapid generation of shotgun high-throughput sequencing libraries from nanogram quantities of DNA. *Appl Environ Microbiol*. 2011;77(22):8071–9.
<https://doi.org/10.1128/AEM.05610-11>
11. Knierim E, Lucke B, Schwarz JM. Systematic comparison of three methods for fragmentation of long-range PCR products for next generation sequencing. *PLoS ONE*. 2011;6(11):e28240.
<https://doi.org/10.1371/journal.pone.0028240>
12. Podnar J, Deiderick H, Huerta G, Hunicke-Smith S. Next-Generation Sequencing RNA-Seq Library Construction. *Curr Protoc Mol Biol*. 2014;106(1):4.21.1–19.
<https://doi.org/10.1002/0471142727.mb0421s106>
13. Barzon L, Lavezzo E, Militello V, Toppo S, Palù G. Applications of next-generation sequencing technologies to diagnostic virology. *Int J Mol Sci*. 2011;12(11):7861–84.
<https://doi.org/10.3390/ijms12117861>
14. Wong K, Fong TT, Bibby K, Molina M. Application of enteric viruses for fecal pollution source tracking in environmental waters. *Environ Int*. 2012;45:151–64.
<https://doi.org/10.1016/j.envint.2012.02.009>

15. Zhang L, Chen F, Zeng Z, Xu M, Sun F, Yang L, et al. Advances in Metagenomics and Its Application in Environmental Microorganisms. *Front Microbiol.* 2021;12:e766364.
<https://doi.org/10.3389/fmicb.2021.766364>
16. Sen P, Oresic M. Metabolic Modeling of Human Gut Microbiota on a Genome Scale: An Overview. *Metabolites.* 2019;9(2):22.
<https://doi.org/10.3390/metabo9020022>
17. Schlager R, Chiu CY, Miller S, Procop GW, Weinstock G. Validation of Metagenomic Next-Generation Sequencing Tests for Universal Pathogen Detection. *Arch Pathol Lab Med.* 2017;141(6):776–86.
<https://doi.org/10.5858/arpa.2016-0539-RA>
18. Seol D, Jhang SY, Kim H, Kim SY, Kwak HS, Kim SH, et al. Accurate and Strict Identification of Probiotic Species Based on Coverage of Whole-Metagenome Shotgun Sequencing Data. *Front Microbiol.* 2019;10:1683.
<https://doi.org/10.3389/fmicb.2019.01683>
19. Afridi OK, Ali J, Chang JH. Next-Generation Sequencing Based Gut Resistome Profiling of Broiler Chickens Infected with Multidrug-Resistant *Escherichia coli*. *Animals (Basel).* 2020;10(12):2350.
<https://doi.org/10.3390/ani10122350>
20. Del Chierico F, Ancora M, Marcacci M, Cammà C, Putignani L, Conti S. Choice of next-generation sequencing pipelines. *Methods Mol Biol.* 2015;1231:31–47.
https://doi.org/10.1007/978-1-4939-1720-4_3
21. McCorrison J, Venepally P, Sing I, Fouts DE, Lasken RS, Methé BA. NeatFreq: reference-free data reduction and coverage normalization for De Novo sequence assembly. *BMC Bioinformatics.* 2014;15(1):357.
<https://doi.org/10.1186/s12859-014-0357-3>
22. Jia X, Hu L, Wu M, Ling Y, Wang W, Lu H, et al. A streamlined clinical metagenomic sequencing protocol for rapid pathogen identification. *Sci Rep.* 2021;11(1):4405.
<https://doi.org/10.1038/s41598-021-83812-x>
23. Parker K, Wood H, Russell JA, Yarmosh D, Shteyman A, Bagnoli J, et al. Development and optimization of an unbiased, metagenomics-based pathogen detection workflow for infectious disease and biosurveillance applications. *Trop Med Infect Dis.* 2023;8(2):121.
<https://doi.org/10.3390/tropicalmed8020121>
24. Ibañez-Lligoña M, Colomer-Castell S, González-Sánchez A, Gregori J, Campos C, Garcia-Cehic D, et al. Bioinformatic tools for NGS-based metagenomics to improve the clinical diagnosis of emerging, re-emerging and new viruses. *Viruses.* 2023;15(2):587.
<https://doi.org/10.3390/v15020587>
25. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* 2019;20(1).
<https://doi.org/10.1186/s13059-019-1891-0>
26. Lu J, Breitwieser FP, Thielen P, Salzberg SL. Bracken: Estimating species abundance in metagenomics data. *PeerJ Comput Sci.* 2017;3:e104.
<https://doi.org/10.7717/peerj-cs.104>
27. Ounit R, Wanamaker S, Close TJ, Lonardi S. CLARK: fast and accurate classification of metagenomic and genomic sequences using discriminative k-mers. *BMC Genomics.* 2015;16: 236.
<https://doi.org/10.1186/s12864-015-1419-2>
28. Blanco-Míguez A, Beghini F, Cumbo F, McIver LJ, Thompson KN, Zolfo M, et al. Extending and Improving Metagenomic Taxonomic Profiling with Uncharacterized Species with using MetaPhlAn 4. *Nat Biotechnol.* 2023;41(11):1633–4.
<https://doi.org/10.1038/s41587-023-01688-w>
29. Nayfach S, Shi ZJ, Seshadri R, Pollard KS, Kyrpides NC. New insights from uncultivated genomes of the global human gut microbiome. *Nature.* 2019;568(7753):505–10.
<https://doi.org/10.1038/s41586-019-1058-x>
30. Raju RS, Al Nahid A, Chondrow Dev P, Islam R. VirusTaxo: Taxonomic classification of viruses from the genome sequence using k-mer enrichment. *Genomics.* 2022;114:10414.
<https://doi.org/10.1016/j.ygeno.2022.110414>
31. Roux S, Tournayre J, Mahul A, Debroas D, Enault F. Metavir 2: New tools for viral metagenome comparison and assembled virome analysis. *BMC Bioinformatics.* 2014;15:76.
<https://doi.org/10.1186/1471-2105-15-76>
32. Roux S, Enault F, Hurwitz BL, Sullivan MB. VirSorter: mining viral signal from microbial genomic data. *PeerJ.* 2015;3:e985.
<https://doi.org/10.7717/peerj.985>
33. Ren J, Ahlgren NA, Lu YY, Fuhrman JA, Sun F. VirFinder: A novel k-mer based tool for identifying viral sequences from assembled metagenomic data. *Microbiome.* 2017;5(1):69.
<https://doi.org/10.1186/s40168-017-0283-5>

Благодарности / Acknowledgements

Авторы выражают свою признательность Квону Дмитрию Аркадьевичу, кандидату биологических наук, за предоставленные обзорные материалы в области современных методов секвенирования / The authors express their gratitude to Dmitry Arkadyevich Kwon, Candidate of Biological Sciences, for providing review materials in the field of modern sequencing methods.

Вклад авторов / Authors' contributions

Все авторы подтверждают соответствие своего авторства критериям ICMJE. Наибольший вклад распределен следующим образом: **Петров А.А.** – формирование концепции статьи, написание текста рукописи; **Казанцев А.В.** – критический пересмотр и коррекция текста рукописи; **Ковальчук Е.А.** – сбор и анализ данных научной литературы, переработка текста рукописи; **Павлюков М.Ю.** – подготовка материалов для статьи по методам секвенирования; **Сапкулов А.В.** – подготовка материалов для статьи по программному обеспечению; **Кутаев Д.А.** – редактирование текста рукописи; **Борисевич С.В.** – окончательное утверждение рукописи для публикации / All the authors confirm that they meet the International Committee of Medical Journal Editors (ICMJE) criteria for authorship. The most significant contributions were as follows. **Petrov A.A.** conceptualised the study, drafted the manuscript. **Kazantsev A.V.** critically reviewed and corrected the manuscript. **Kovalchuk E.A.** collected and analysed scientific literature and revised the manuscript. **Pavlyukov M.Yu.** preparation of the article section on sequencing methods. **Sapkulov A.V.** preparation of the article section on software. **Kutaev D.A.** edited the manuscript. **Borisevich S.V.** approved the final version of the manuscript for publication.

Сведения о рецензировании / Peer review information

Статья прошла двустороннее анонимное «слепое» рецензирование двумя рецензентами, специалистами в данной области. Рецензии находятся в редакции журнала и в РИНЦе / The article has been doubleblind peer reviewed by two experts in the respective field. Peer reviews are available from the Editorial Board and from Russian Science Citation Index database.

Об авторах/ Authors

Федеральное государственное бюджетное учреждение «48 Центральный научно-исследовательский институт» Министерства обороны Российской Федерации, 141306, Российская Федерация, г. Сергиев Посад-6, ул. Октябрьская, д. 11.

Петров Александр Анатольевич. Начальник научно-исследовательского управления, д-р мед. наук.

ORCID: <https://orcid.org/0000-0002-9714-2085>

Казанцев Алексей Васильевич. Старший научный сотрудник, канд. биол. наук.

Ковальчук Елена Анатольевна. Научный сотрудник, канд. мед. наук.

Павлюков Михаил Юрьевич. Научный сотрудник.

Сапкулов Алексей Владимирович. Помощник начальника ФГБУ «48 ЦНИИ» Минобороны России.

Кутаев Дмитрий Анатольевич. Заместитель начальника ФГБУ «48 ЦНИИ» Минобороны России по научной работе, канд. биол. наук.

Борисевич Сергей Владимирович. Начальник ФГБУ «48 ЦНИИ» Минобороны России, д-р биол. наук, профессор, академик РАН.

ORCID: <https://orcid.org/0000-0002-6742-3919>

Контактная информация для всех авторов: 48cnii@mil.ru

Контактное лицо: Борисевич Сергей Владимирович; 48cnii@mil.ru

48 Central Scientific Research Institute of the Ministry of Defense of the Russian Federation, Oktyabrskaya St. 11, Sergiev Posad 141306, Russian Federation.

Aleksandr A. Petrov. Head of Research Department, Dr Sci. (Med.).

ORCID: <https://orcid.org/0000-0002-9714-2085>

Aleksey V. Kazantsev. Leading researcher, Cand. Sci. (Biol.).

Elena A. Kovalchuk. Researcher, Cand. Sci. (Biol.).

Mikhail Yu. Pavlyukov. Researcher.

Aleksey V. Sapkulov. Assistant of Head of Federal State Budgetary Establishment “48 Central Scientific Research Institute” of the Ministry of Defense of the Russian Federation.

Dmitriy A. Kutaev. Deputy of Head of Federal State Budgetary Establishment “48 Central Scientific Research Institute” of the Ministry of Defense of the Russian Federation on scientific research, Cand. Sci. (Biol.).

Sergey V. Borisevich. Head of Federal State Budgetary Establishment “48 Central Scientific Research Institute” of the Ministry of Defense of the Russian Federation, Academician of Russian Academy of Sciences, Dr Sci. (Biol.), Professor.

ORCID: <https://orcid.org/0000-0002-6742-3919>

Contact information for all authors: 48cnii@mil.ru

Contact person: Sergey V. Borisevich; 48cnii@mil.ru