

УДК 575.174.015.3

ИСПОЛЬЗОВАНИЕ МАСШТАБНОГО ПАРАЛЛЕЛЬНОГО СЕКВЕНИРОВАНИЯ В КРИМИНАЛИСТИКЕ: СРАВНИТЕЛЬНЫЙ АНАЛИЗ ПЛАТФОРМ ДЛЯ СЕКВЕНИРОВАНИЯ

© 2021 г. Т. В. Тяжелова^{1, *}, И. Л. Кузнецова¹, Т. В. Андреева^{1, 2},
С. С. Кунижева^{1, 2}, Е. И. Рогаев^{1, 2, 3}

¹Институт общей генетики им. Н.И. Вавилова Российской академии наук, Москва, 119991 Россия

²Московский государственный университет им. М.В. Ломоносова, Москва, 119991 Россия

³Научно-технологический университет “Сириус”, Краснодарский край, Сочи, 354340 Россия

*e-mail: tatvltz@gmail.com

Поступила в редакцию 10.06.2021 г.

После доработки 18.06.2021 г.

Принята к публикации 20.07.2021 г.

Последнее десятилетие характеризуется значительным прорывом в развитии новых технологий и инструментов для анализа ДНК. Новая технология масштабного параллельного секвенирования (МПС), иначе называемая “секвенирование следующего поколения”, к настоящему времени уже широко используется во многих направлениях биологических и биомедицинских исследований. Внедрение новых технологий параллельного секвенирования в область судебной экспертизы происходит медленнее по сравнению с научными и клиническими исследованиями, в основном в связи со сложностями совмещения с уже имеющимися базами данных ДНК-профилей, а также необходимостью проведения строгих валидационных экспериментов для аккредитации приборов и наборов реагентов. В обзоре представлены существующие в мире решения по использованию масштабного параллельного секвенирования в криминалистике, технические аспекты экспериментов, разработанные коммерческие решения и перспективы применения, а также рассмотрены преимущества и недостатки новых технологий секвенирования по сравнению с методами ДНК-анализа, традиционно используемыми в судебной экспертизе.

Ключевые слова: масштабное параллельное секвенирование (МПС), секвенирование следующего поколения (NGS), короткий тандемный повтор (STR), однонуклеотидный полиморфизм (SNP), криминалистика.

DOI: 10.31857/S0016675821120122

Перспектива одновременного анализа большого количества полиморфных маркеров делает технологию масштабного параллельного секвенирования (МПС) очень мощным и относительно легко применимым инструментом в лабораториях судебной экспертизы. В настоящее время наиболее активно МПС используется в диагностических лабораториях для анализа индивидуальных геномов человека или целевого (таргетного) секвенирования выборочных участков ДНК для выявления генетических вариантов, связанных со здоровьем или с развитием заболеваний. В криминалистике с использованием МПС станет возможным одновременный анализ и стандартных аутосомных ДНК-маркеров (STR) и SNP, и митохондриальной ДНК, и маркеров половых хромосом [1, 2]. Прежде всего технологии МПС будут иметь решающее значение для анализа ДНК человека в таких случаях, как массовые бед-

ствия или другие события, когда криминалистические образцы ДНК подвергаются негативному воздействию и их качество ухудшается.

О возможностях и перспективах использования МПС в криминалистической работе было впервые объявлено в 2015 г. в годовом отчете Нидерландского института судебной экспертизы (NFI) [3]. Первым институтом, аккредитованным на использование технологии МПС в судебном анализе, стала Лаборатория судебно-медицинской экспертизы ДНК Университета Лейдена. Первыми заявленными результатами были уменьшение ложноположительных совпадений при сравнении разных образцов ДНК и успешное разделение различных профилей ДНК в сложной смеси [4]. В 2015 г. использование технологий NGS в криминалистике ограничивалось лишь отдельными пилотными исследованиями. Однако уже к 2017 г. опубликовано большое количество работ по тестирова-

нию различных приборов для МПС и различных панелей криминалистических маркеров. Проведенный в 2017 г. опрос 33 европейских лабораторий, работающих в области судебно-медицинской экспертизы, показал, что 52% из них уже приобрели один или более приборов для МПС [5].

МЕТОДОЛОГИЯ МАСШТАБНОГО ПАРАЛЛЕЛЬНОГО СЕКВЕНИРОВАНИЯ

В основе масштабного параллельного секвенирования лежит несколько различных методологических подходов: секвенирование путем синтеза после амплификации ДНК [6], секвенирование отдельной молекулы в реальном времени на ячейке [7] или с использованием нанопор [8]. Большинство платформ секвенирования второго поколения используют метод секвенирования путем синтеза. Эта технология основана на клональной амплификации ДНК ферментом полимеразой. При мономолекулярном секвенировании этап предварительной амплификации отсутствует.

К основным существенным характеристикам различных методов МПС относятся длина прочтений, производительность и частота ошибок секвенирования. Обзор данных характеристик для трех поколений методов секвенирования приведен в табл. 1. Длина прочтения, производительность и время рабочего процесса каждой из платформ в настоящее время изменяются быстрыми темпами в связи с постоянным совершенствованием методологий, поэтому данные, представленные в табл. 1, показывают только общее представление о каждом из поколений методов и платформах для секвенирования. Несмотря на то что уровень ошибок, получаемых при использовании всех существующих платформ секвенирования нового поколения, выше по сравнению с секвенированием методом Сэнгера, этот недостаток преодолевается в технологиях МПС путем многократного прочтения каждого локуса, т.е. высоким покрытием [2].

В настоящее время в сфере судебной генетики наибольшее применение нашли настольные варианты секвенаторов – это прибор MiSeq (и его модификация ForenSeq), разработанный компанией Illumina, а также секвенаторы от компании ThermoFisher Scientific – Ion Torrent PGM и Ion S5 [5].

Платформа Illumina реализует технологию циклического секвенирования в сочетании со стратегией обратной терминации флуоресцентно меченных модифицированных дезоксирибонуклеотидтрифосфатов, концептуально аналогичной секвенированию методом Сэнгера. Замечены оснований – это наиболее частая ошибка,

возникающая во время секвенирования на машинах Illumina.

Платформы Ion Torrent (PGM и S5) представляют собой полупроводниковый секвенатор, который измеряет изменения рН, как следствие выделения ионов водорода при синтезе ДНК [9]. Для платформы Ion Torrent наиболее частыми ошибками являются инсерции и делеции гомополимерных участков длиной более 6 пн, поскольку корреляция между включенными нуклеотидами и изменением детектируемого напряжения не пропорциональна размеру гомополимера [8, 10]. Количество ошибок, генерируемых при секвенировании на платформе Ion Torrent ($\geq 1\%$), выше, чем на платформе Illumina [11]. Однако отсутствие оптического сканирования и циклического секвенирования на платформе Ion Torrent значительно снижает время секвенирования ДНК по сравнению с платформой Illumina [12]. Протоколы ручной подготовки библиотеки ДНК для платформ Ion Torrent PGM и Ion S5 более трудоемки и громоздки по сравнению с рабочим процессом Illumina, но наличие автоматизированной станции подготовки библиотек Ion Torrent (Ion Chef System) значительно упрощает и стандартизирует рабочий процесс. Недавно аналогичные решения по автоматизации пробоподготовки были предложены и для платформы Illumina [13].

МПС КАК ИНСТРУМЕНТ ДЛЯ ГЕНОТИПИРОВАНИЯ STR-И SNP-МАРКЕРОВ

Первыми маркерами, которые стали использоваться для идентификации личности, стали маркеры ДНК, на основе семейств так называемых минисателлитных или гипервариабельных повторов ДНК. Такой тип повторов ДНК в геноме характеризуется высокой аллельной вариабельностью числа tandemных повторов (VNTR, variable number of tandem repeats). На первых этапах внедрения методов ДНК-идентификации в судебную практику для получения индивидуального профиля ДНК человека (ДНК-фингерпринт) использовали достаточно трудоемкую технологию, которая включала расщепление геномной ДНК специальными ферментами на короткие фрагменты, которые разделяли электрофоретически по размеру и гибридовали с радиоактивно мечеными пробами, соответствующими отдельным VNTR-маркерам [14, 15].

Благодаря быстрому развитию методов молекулярной биологии, в частности появлению метода полимеразной цепной реакции (ПЦР) анализа ДНК, а также разработкам платформ Applied Biosystems для капиллярного электрофореза, используемых для прямого определения последовательности ДНК (секвенирования ДНК) или фрагментного анализа ПЦР-продуктов, ДНК-типиро-

Таблица 1. Основные характеристики для трех поколений методов секвенирования

Метод	Год	Принцип действия	Длина чтения, число нуклеотидов (нт)	Производительность и время работы	Ошибка секвенирования, %
I поколение секвенирования					
Сэнгера	1977	Клонирование/обрыв цепи	25–1200	96.84×10^3 , 2 ч	
II поколение секвенирования					
454	2005	эм-ПЦР/SBS/пиросеквенирование	100–1000	10^5 нт, 0.7 Гб, 24 ч	0.003
Solexa/HiSeqR_/MiSeqR_	2006	Мостиковая ПЦР/SBS/обратная терминация	2×100 (HiSeq); 2×300 (MiSeq)	6×10^9 нт, 1.8 Тб, несколько дней (в зависимости от длины чтения)	<0.01
SOLiD	2007	эм-ПЦР/лигирование/зонды	35–75	6×10^9 нт, 320 Гб, 1–2 недели	0.001
Ion Torrent	2010	эм-ПЦР/ионно-чувствительный SBS/изменение pH	200–400	$60–80 \times 10^6$ нт, 50 Гб, 2 ч	<1
III поколение секвенирования					
PacBio R	2010	Одномолекулярное секвенирование в реальном времени на ячейке (SMRT/ZMW wells)	8000–20000	35×10^4 нт, 7 Гб, 0.5–6 ч	<0.001
Oxford Nanopore	2014	Измерение электрического тока при прохождении молекулы нуклеиновой кислоты через нанопору	500–200000	10^5 нт, 2–4 Тб, до 48 ч	до 15

Примечание. эм-ПЦР – эмульсионная ПЦР, Гб – гигабайт, Тб – терабайт.

вание стало широко использоваться в практике судебных экспертиз. Одним из первых стандартизированных наборов, получившим международное применение в области криминалистики, стал набор базовых STR (short tandem repeats)-локусов CODIS (Combined DNA Index System). Этот набор был разработан по заказу ФБР при совместном усилии корпорации Promega и Applied Bio systems для создания стандартов ДНК-идентификации для национальной базы ДНК США, с целью увеличения эффективности идентификации прежде всего военных лиц, преступников, а также лиц, пропавших без вести. В 1997 г. был утвержден набор CODIS, состоящий из 13 базовых STR-локусов: CSF1PO, FGA, TH01, TPOX, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11 и амелогенинового маркера пола. Также был создан набор локусов STR-маркеров для территории Евросоюза (ESS): FGA, TH01, D1S1656, D2S441, D3S1358, D8S1179, D10S1248, D12S391, D18S51, D21S11, D22S1045. Локусы, входящие в данные наборы, стали мировым стандар-

том и были рекомендованы для использования во всех криминалистических лабораториях мира [16].

Традиционный метод STR-анализа при помощи геле- или капиллярного электрофореза (КЭ) основан на выявлении количества повторяющихся “единиц” STR-маркера по размеру ПЦР-ампликона. Анализ таких маркеров с использованием платформ МПС позволяет получить больше информации – определить не только длину повторяющегося участка ДНК, но и полную последовательность как самой области повторов, так и фланкирующих ее регионов, которые часто содержат дополнительные полиморфные варианты. Многочисленные исследования продемонстрировали, что анализ STR-маркеров на основе полной последовательности локуса приводит к значительному увеличению дискриминационного потенциала маркеров по сравнению с анализом, основанным только на длине этих маркеров [17–21]. Также было отмечено, что выявляемые при этом дополнительные полиморфные варианты могут быть популяционно-специфическими. Например, в ис-

следовании четырех основных популяционных групп США было показано, что все локусы стандартных STR-маркеров, за исключением маркера TPOX, содержат дополнительные варианты последовательности [20]. Для каталогизации новых данных вариабельных последовательностей STR-маркеров создан проект STRSeq, который суммирует все наблюдаемые “дополнительные” аллели последовательностей основных локусов STR-маркеров, используемых в судебной экспертизе [21].

Возможность выявить дополнительные вариации в последовательности ДНК при анализе STR-маркеров с использованием методологии масштабного параллельного секвенирования имеет и ряд других преимуществ. Во-первых, увеличивается точность предсказания при использовании этих маркеров для прямого сопоставления или вычислений родственных взаимосвязей. Во-вторых, более удачно решается проблема так называемых “статтер-аллелей” (или “статтер-пигов”) – наиболее распространенного артефакта, наблюдаемого при анализе STR-маркеров и вызванного проскальзыванием ДНК во время амплификации. Образование статтер-аллеля может встречаться в 6–10% продуктов амплификации. Если в копируемой нити наблюдается “проскальзывание”, то некоторые из амплифицированных цепей получаются короче, чем должен быть продукт ПЦР. С помощью масштабного параллельного секвенирования было показано, что статтер-аллели маркера зависят от нуклеотидной последовательности основного аллеля и, таким образом, уровень статтер-аллелей может быть различным для аллелей одинаковой длины, имеющих различные нуклеотидные составы. Это дает дополнительное преимущество МПС по сравнению с методом фрагментного анализа STR-маркеров, в частности позволяет дифференцировать аллели в смешанных образцах [22]. В-третьих, при работе со смесью образцов ДНК выявление дополнительных аллелей, которые ранее были бы замаскированы из-за одинаковой длины различных аллелей STR-маркера, снижает вероятность случайного совпадения профилей ДНК разных индивидов, присутствующих в смешанном образце. Использование дополнительных вариаций в локусе STR-маркера было успешно продемонстрировано для анализа смеси образцов в работе Phillips и др. [23]. Кроме того, было показано, что число прочтений, полученных при секвенировании, позволяет точно определить пропорции, в которых смешаны образцы [19].

Разделение фрагментов ДНК по размеру при проведении исследования методом МПС не является необходимым условием анализа STR-маркеров, как при использовании фрагментного анализа. Все маркеры могут быть амплифицированы с получением фрагментов наименьшей возможной длины, что важно для анализа деградированной ДНК,

а количество одновременно исследуемых маркеров не ограничивается количеством флуоресцентных красителей, которые могут быть детектированы электрофоретическими системами капиллярного электрофореза. Это открывает возможность одновременного использования большего числа STR-маркеров, что, например, было продемонстрировано при использовании большой панели Y-хромосомных и аутомсомных STR, применение которых значительно повышает эффективность анализа смешанных образцов мужской и женской ДНК в образцах, полученных после сексуального насилия [24].

Современная судебно-медицинская экспертиза анализа образцов ДНК не ограничивается STR-маркерами. Однонуклеотидные полиморфные варианты (SNP или SNV) имеют ряд преимуществ при анализе деградированных криминалистических образцов ДНК по сравнению с STR-маркерами, потому что для их определения можно использовать более короткие фрагменты ДНК [25]. Например, была показана возможность определения полного профиля из 20 однонуклеотидных маркеров на ДНК из пятен крови после длительного хранения образцов (243 дня), тогда как только 9% STR-маркеров было возможно типировать на этих же образцах. В случае образцов ДНК, выделенной из слюны после длительного хранения (147 дней), были определены 16 однонуклеотидных маркеров (81%) и 18% STR-маркеров [26].

Также на основании анализа криминалистических образцов различных тканей разной степени деградации была продемонстрирована возможность получения качественного профиля на основании SNP-панелей, тогда как генотипирование микросателлитных маркеров почти не дало положительных результатов [27]. Из 36 различных криминалистических образцов, для которых не смогли получить STR-профили или они были низкого качества, для 16 образцов (45%) были получены полные SNP-профили [28].

За последние несколько лет был опубликован ряд работ, посвященных разработке высокоинформативных SNP-маркеров, эффективных для ДНК-идентификации индивида [29, 30]. В них была показана возможность предсказаний вероятного цвета глаз, волос, кожи, а также этнического происхождения на основании анализа сочетаний гаплотипов SNP-маркеров. Наборы ДНК-маркеров для определения внешности HirisPlex [29], этногеографического происхождения AISNP [30] и идентификации личности [31] в настоящее время лежат в основе большинства коммерческих панелей для ДНК-идентификации методами масштабного параллельного секвенирования (табл. 2).

Таблица 2. Характеристики наиболее распространенных коммерческих наборов реагентов для ДНК-идентификации с помощью МПС

Наименование панели	Состав маркеров (по состоянию на 01.05.2021)	Рекомендации по секвенированию
Фирма Verogene (США)		
ForenSeq DNA Signature Prep Kit	A: 27 маркеров аутосомных STR, 24 Y-STR, 7 X-STR, 94 SNP-маркера для идентификации личности; B: 27 маркеров аутосомных STR, 24 Y-STR, 7 X-STR, 172 SNP-маркера (94 для идентификации личности, 22 для определения фенотипа, 56 для определения биогеографического происхождения)	Количество ДНК 1 нг, число библиотек за прогон 8–96, ~28 ч MiSeq FGx Reagent Kit, ~22 ч MiSeq FGx Reagent Micro Kit
ForenSeq MainstAY Kit	27 маркеров аутосомных STR, 25 Y-STRs-маркеров (минимальный набор для определения Y-гаплотипа по базе YHRD*), средний размер ампликонов 235 пн (максимальный 481 пн)	Количество ДНК 1 нг, число библиотек за прогон 8–96, ~28 ч MiSeq FGx Reagent Kit, ~22 ч MiSeq FGx Reagent Micro Kit
ForenSeq mtDNA Whole Genome Kit	245 ампликонов, восстанавливающих полную последовательность мтДНК генома (16569 пн)	Количество геномной ДНК 100 пг, число библиотек за прогон 16, ~28 ч MiSeq FGx Reagent Kit
ForenSeq mtDNA Control Region Kit	18 ампликонов, восстанавливающих последовательность контрольного региона мтДНК (1200 пн)	Количество геномной ДНК 100 пг, число библиотек за прогон 3–48, ~28 ч MiSeq FGx Reagent Kit, ~22 ч MiSeq FGx Reagent Micro Kit
ForenSeq Kintelligence Kit	10230 SNP-маркеров для судебно-медицинской экспертизы (94 для идентификации личности, 22 для определения фенотипа, 56 для определения биогеографического происхождения, 9867 для определения родства, 106 X-SNP и 85 Y-SNP, средний размер ампликонов 150 пн)	Количество ДНК 1 нг, число библиотек за прогон 3, ~28 ч MiSeq FGx Reagent Kit
Фирма Promega (США)		
PowerSeq 46GY System	22 аутосомных STR-маркера, 23 Y-STR-маркера, амелогенин, размер ампликонов 140–300 пн	Количество ДНК 1 нг, число библиотек за прогон 96, ~28 ч MiSeq FGx Reagent Kit
PowerSeq CRM Nested System	HVI, HVII и HVIII контрольного региона МТ, размер ампликонов 144–237 пн	Количество ДНК 1 нг, число библиотек за прогон до 16, ~28 ч MiSeq FGx Reagent Kit
Фирма NimaGen (Нидерланды)		
EasySeq™ Human ID and Sample Tracking Kit	34 SNP-маркера для идентификации личности и амелогенин	Количество ДНК 1 нг, число библиотек за прогон до 16, ~28 ч MiSeq FGx Reagent Kit
Фирма QIAGENE		
QIAseq Investigator Missing Persons SNP Panel	1448 ампликонов (1200 аутосомных SNP-маркеров для идентификации личности и родственных связей, 33 X-SNP-маркера, 55 биогеографического происхождения). Панель включает 46 микрогаплотипов	Количество ДНК 10–40 нг, число библиотек за прогон до 60. В случае панели QIAseq Investigator Missing Persons, SNP panel – до 24, ~28 ч MiSeq FGx Reagent Kit
QIAseq Investigator ID SNP Panel	139 SNP-маркеров для идентификации личности	возможно использовать платформы и Ion Torrent

Таблица 2. Окончание

Наименование панели	Состав маркеров (по состоянию на 01.05.2021)	Рекомендации по секвенированию
QIAseq Investigator Ancestry SNP Panel	102 SNP-маркера для определения биогеографического происхождения (Global ancestry), 75 SNP-маркеров для определения биогеографического происхождения (Middle East)	То же
QIAseq Human Mitochondria Panel	111 ампликонов, восстанавливающих полную последовательность МТ-генома (16411 пн)	
QIAseq Investigator Human Mitochondria Control Region Panel	13 ампликонов, восстанавливающих последовательность контрольного региона мтДНК (1200 пн)	
Фирма BGI (Китай)		
MGIEasy Signature Identification Library Prep Kit	52 аутосомных STR-маркера, 27 STR-маркеров X-хромосомы, 48 STR-маркеров Y-хромосомы, 227 SNP-маркеров (145 для идентификации личности, 29 для определения фенотипа, 53 для определения биогеографического происхождения), HVR 1–HVR 3 МТ, амелогенин	Количество ДНК 10–40 нг, число библиотек за прогон до 60, 28 ч MGISEQ-2000
Фирма Thermo Fisher Scientific		
Precision ID GlobalFiler NGS STR Panel	30 аутосомных STR-маркеров, 1 Y-STR-маркер, 3 маркера определения пола (амелогенин, SRY, SNP rs2032678), размер ампликонов 140–300 пн	Количество ДНК 1 нг, число библиотек за прогон 8, 2 ч Ion S5 System
Precision ID Identity Panel	124 SNP-маркера для идентификации личности, средний размер ампликонов 138 пн	
Precision ID Ancestry Panel	165 SNP-маркеров для определения биогеографического происхождения, средний размер ампликонов 127 пн	
Precision ID mtDNA Whole Genome Panel	81 ампликон, восстанавливающий полную последовательность МТ генома (16569 пн), средний размер ампликонов 163 пн	Количество ДНК 100 пг, число библиотек за прогон 8, 2 ч Ion S5 System
Precision ID mtDNA Control Region Panel	7 ампликонов, восстанавливающих последовательность контрольного региона мтДНК, средний размер ампликонов 153 пн	

* YHRD – Y-STR Haplotype Reference Database.

Преимущества анализа SNP-маркеров для судебных целей с помощью МПС, по сравнению с альтернативными технологиями, такие как SNaPshot и системы чипов, заключаются в первую очередь в том, что можно одновременно определять большее количество маркеров на малых количествах ДНК [32–35]. Кроме того, технология МПС позволяет дополнительно к выявлению индивидуальных генетических вариантов определять также микрогаплотипы, т.е. наборы SNP-маркеров, расположенных в непосредственной близости друг от друга (менее 100 пн). Такие микрогаплотипы демонстрируют большой потенциал в криминалистических исследованиях как для целей идентификации неизвестного ин-

дивида, так и для определения вероятного этнического происхождения [36–38].

Анализ полной последовательности ампликонов ПЦР при типировании SNP-маркеров с помощью технологии МПС дает также возможность использовать любой наблюдаемый полиморфизм в исследуемом участке ДНК для увеличения дискриминационного потенциала целевого SNP-маркера. Так, при анализе фланкирующих областей локусов, входящих в панель ForenSeq™ DNA Signature Prep Kit, в одной из популяций коренных американцев было обнаружено, что из 94 ампликонов, полученных с помощью набора ForenSeq DNA Signature Prep Kit, 22 содержали дополнительные полиморфные варианты, из кото-

Таблица 3. Список маркеров, отмеченных в валидационных исследованиях различными недостатками при генотипировании с помощью МПС

MiSeq FGx	Ion Torrent	BGI	MiniON
STR-маркеры			
DY389II, DY448, DX10148	D3S1358, D7S820, D8S1179	DYS389II, DYS612	—
SNP-маркеры			
Rs459929*, Rs1029047*, Rs2399332*, Rs7251928, Rs7722456*, Rs110488710*	Rs321198, Rs576261, Rs917118, Rs4530059, Rs10318258*	rs1736442, rs1736442, rs914165, rs735480	Rs733164, Rs873196, Rs1029047, Rs1493232, Rs1031825*

* Маркеры удалены из последних версий панелей.

рых 14 оказались информативными, т.е. не были полностью сцеплены с аллелем основного маркера [36].

КОММЕРЧЕСКИЕ НАБОРЫ ДЛЯ КРИМИНАЛИСТИКИ

Одним из первых коммерчески доступных наборов для криминалистического анализа с использованием платформы для масштабного параллельного секвенирования Illumina MiSeq FGx стал набор реагентов ForenSeq DNA Signature Prep Kit (Verogen, CA), который позволяет одновременно анализировать до 231 маркера. Этот набор реагентов содержит два набора праймеров (набор А и набор В), в состав каждого из которых входят праймеры для анализа используемых в криминалистике 27 аутосомных STR-маркеров, 24 STR-маркера Y-хромосомы, 7 STR-маркеров X-хромосомы, маркеры пола и 94 SNP-маркера для индивидуальной идентификации (набор праймеров А). Набор праймеров В дополнительно включает праймеры для 22 фенотипических маркеров и 56 маркеров этногеографического происхождения [39].

Первая версия этого набора была протестирована Churchill et al. в 2016 г. с помощью секвенатора MiSeq (Illumina) [40]. В работе использовали образцы стандартной ДНК: 2800M (Promega® Corporation, Madison, WI) и AmpFISTR® Control DNA 9947A (Life Technologies Corporation), а также 10 образцов ДНК индивидов из разных мировых популяций. Чувствительность набора ForenSeq DNA Signature Prep Kit оценивали путем изменения количества ДНК, взятой в анализ: от 1 нг до 50 пг в двух контрольных образцах. При использовании 1 нг были получены полные профили по всем SNP- и STR-маркерам, совпавшие с профилями контрольных образцов ДНК. С уменьшением количества ДНК (500, 250 и 100 пг) наблюдалось “выпадение” некоторых аллелей (0.5, 1.7 и 5.6% от всех аллелей соответственно). При использовании 50 пг ДНК были успешно генотипированы только 69.4% от всех маркеров. Таким образом

была показана относительная нечувствительность к колебаниям количества ДНК в диапазоне от 1 нг до 100 пг. При проведении сравнительного анализа 10 образцов с помощью ForenSeq DNA Signature Prep Kit и наборов Identifiler Plus и Yfiler (Thermo Fisher Scientific) было показано полное совпадение аллелей, полученных для каждого из образцов двумя методами, за исключением двух локусов: при анализе методом капиллярного электрофореза наблюдалось выпадение аллелей в локусах DYS392 и DYS635, а при использовании МПС – в локусе DYS392 [40].

При анализе смешанных ДНК образцов изменения в гомо- и гетерозиготности маркеров обнаруживались, если второстепенная (минорная) ДНК присутствовала в образце даже на уровне 5%. Авторы [41] считают, что при числе прочтений менее 85000 на образец возможно увеличение количества неоднозначно определенных генотипов, что является достоверным наблюдением и может использоваться в качестве руководства в судебно-медицинской практике. Независимая проверка набора ForenSeq DNA Signature Prep Kit была проведена в работах Xavier et al. (набор праймеров А) [41] и Silvia et al. (набор праймеров В) [42]. В обоих исследованиях для всех стандартных контрольных образцов, использованных в работах, были получены полные профили STR-маркеров, за исключением двух случаев “выпадения” аллелей маркера Penta E при использовании смеси праймеров В [41]. Также в обеих работах профили всех STR-маркеров совпали с таковыми, полученными методом капиллярного электрофореза, за исключением ранее упомянутого маркера DXS10148 [20]. Заметим, что этот маркер позже был удален из панели.

При анализе чувствительности набора реагентов с использованием смеси праймеров А было выявлено “выпадение” аллеля для маркера DX10103 при использовании 250 пг ДНК. При использовании 50 пг для 93% STR-маркеров генотип образцов совпал с контрольным вариантом [42]. При этом были выявлены случаи отсутствия ампликонов (2.8%), низкого покрытия (0.6%), некорректное

определение генотипа было получено для 0.6% маркеров. Для маркера D22S1045 было отмечено уменьшение среднего числа прочтений по мере увеличения размера аллеля, что наблюдалось и в других исследованиях [21, 43, 44]. Предполагают, что причиной этому мотив повтора — последовательность АТТ [45]. Также сообщалось о проблемах с покрытием маркера DYS392, имеющего аналогичный мотив повтора — АТА [43, 45]. Результаты генотипирования аутосомных SNP-маркеров совпадали в трех повторностях с покрытием от 1072 до 21 чтений на локус, с правильным и полным генотипированием при использовании более 100 пг матрицы ДНК [46]. В целом эти исследования показали, что некоторые аутосомные SNP-маркеры в этом наборе, включая rs6955448 [46, 47], rs7041158 [45] и rs2920816 [45], некорректно генотипируются и в дальнейшем они были исключены из набора реагентов.

С помощью набора ForenSeq DNA Signature Prep Kit также исследовали смеси мужской и женской ДНК, выделенной из буккальных соскобов [41]. Были исследованы смеси образцов буккального эпителия мужчины и женщины. SNP-маркеры мужчины были выявлены в количестве 3.3% (смесь мужского и женского образца 1 : 20), 4.4% (смесь 1 : 10) и 34.8% (смесь 1 : 1). Недопредставленность мужских SNP-маркеров в смесях связана с совпадением некоторых аллелей мужчины с женскими аллелями [41]. Silvia и др. проанализировали образцы ДНК, выделенные из пятен крови, слюны или спермы, с выстиранной или ношенной одежды из хлопка, нейлона и джинсовой ткани [42]. Для испытаний набора использовали 1 нг ДНК. При использовании прибора ForenSeq, 13 из 15 образцов дали полные STR-профили, совпадающие с контрольным типированием (AmpFISTR Identifier PCR Amplification Kit Life Technologies/3130xl Genetic Analyzer). В шести образцах наблюдалось место выпадения маркера Penta E. Надо отметить, что длина ампликона Penta E составляет от 362 до 467 пн, он является одним из самых длинных STR-маркеров. В образце ДНК, полученной с ношенной футболки, также три маркера (CSF1PO, D16S539, D2S1338) не были типированы корректно ни одной системой. В образце ДНК с выстиранной футболки уже девять маркеров не были генотипированы (CSF1PO, D13S317, D16S539, D18S51, D2S1338, TH01, TPOX, D7S820a, vWAa). Вероятная причина этого — деградация ДНК, полученной из образцов исследуемого материала. Для 16 образцов были получены полные профили SNP-маркеров, соответствующие профилю ДНК донора. Один локус (rs1426654) в образце ДНК с ношенных джинс был по-разному генотипирован системой MiSeqFGx (по сравнению с контрольным типированием в фирме 23andMe), вероятно из-за того, что порог гетерозиготности в MiSeq FGx был установлен на уровне 25% [42].

Вторая лидирующая на современном рынке компания, выпускающая приборы МПС для криминалистических лабораторий, — это ThermoFisher Scientific. Эта компания разработала и выпустила целый ряд панелей Precision ID для использования на платформе Ion Torrent: панель идентификационных STR-маркеров, панель идентификационных SNP-маркеров, панель SNP-маркеров для определения этногеографического происхождения, панель SNP-маркеров для определения цвета радужной оболочки глаз и цвета волос, а также для определения последовательности митохондриальной ДНК, включая отдельную панель для анализа гипервариабельных регионов митохондриальной ДНК. За последние годы большое число публикаций было посвящено тестированию этих панелей [48–51]. Например, при работе с панелью маркеров, включающей 165 маркеров для определения происхождения индивида (Precision ID Ancestry Panel), все группы исследователей отмечали, что рекомендованное фирмой-производителем число циклов амплификации (21 цикл) недостаточно для получения полных профилей по включенным в состав набора маркерам в случае использовании малых количеств (менее 1 нг) ДНК — в таких условиях 30% локусов имеют покрытие менее 100 [48–51]. При этом увеличение циклов амплификации до 25 приводит к значительному улучшению результатов, хотя и повышает аллельный дисбаланс за счет предпочтительной амплификации более коротких фрагментов ДНК. Во всех исследованиях было отмечено низкое покрытие маркера rs1296819. Pereira и др. [48] пришли к выводу, что три SNP-маркера (rs7722456, rs459929 и rs7251928) работали настолько плохо, что их следует исключить из панели (см. табл. 3). По результатам описанных выше работ как в дизайн панелей Precision ID, так и в программное обеспечение был внесен ряд изменений [52–54].

В настоящее время основной набор для анализа STR-маркеров (панель Precision ID GlobalFiler NGS STR) включает все локусы расширенной панели CODIS, а также девять дополнительных аутосомных STR-локусов, четыре маркера пола, включая DYS391. Этот набор был протестирован на приборах Ion PGM [55] и Ion S5 [56]. Значительный дисбаланс в покрытии между разными маркерами все равно сохраняется. При этом Müller et al. [55] продемонстрировали, что дисбаланс воспроизводимо изменялся, когда панель использовалась в разных лабораториях. Wang et al. [57] предложили некоторые модификации, чтобы уменьшить разницу в уровнях покрытия маркеров.

При работе с панелями STR-маркеров наблюдаемые коэффициенты статтер-аллелей различались между исследованиями [55, 57]. Wang et al. сообщают о 8%-ном проскальзывании для всех (кроме трех) маркеров. В то время как Müller et al.

сообщают о более высоких коэффициентах – до 15–21% для некоторых маркеров (в обеих работах использовали 1 нг ДНК). Причина такой большой разницы между исследованиями не очевидна, хотя использование разных типов инструментов – одна из возможных причин. При этом в обоих исследованиях набор продемонстрировал хорошую чувствительность, и большинство аллелей детектировались и корректно определялись при количествах ДНК-матрицы ниже 100 пг [55, 57].

В качестве сравнительного анализа использования платформ IonTorrent и MiSeq в криминалистике можно рассматривать исследования панелей маркеров, разработанных различными фирмами. Так, панель идентификационных маркеров фирмы Qiagen, содержащая 140 аутосомных SNP, была протестирована с использованием платформ MiSeq [33] и Ion Torrent [58]. Было обнаружено, что из 140 маркеров один (rs1058083) дает стабильно низкое покрытие на обеих платформах во всех исследованиях, и в целом одни и те же SNP работают плохо во всех лабораториях, что указывает на низкую эффективность амплификации в начальной ПЦР. Покрытие, получаемое для некоторых маркеров, вероятно может зависеть от применяемой для секвенирования платформы. Так, покрытие маркера rs9951171 при секвенировании на платформе Ion Torrent в работе [58] было в 4 раза ниже, чем при использовании платформы MiSeq [33]. Типирование маркера rs1029047 оказалось проблематичным во всех исследованиях как с гомозиготными, так и с гетерозиготными генотипами; для него был выявлен значительный аллельный дисбаланс. Примечательно, что проблемы с генотипированием данного маркера также были описаны при его анализе с помощью панели от ThermoFisher [57]. Это, вероятно, связано с расположением вблизи маркера rs1029047 полиА тракта.

Аллельный дисбаланс также наблюдался во всех работах при анализе маркера rs2399332. Авторы интерпретируют этот результат по-разному: наличием полиморфизма в сайте связывания праймера [33], включением лишнего основания [47, 58], соседством с полиТ трактом, секвенирование которого приводит к ошибочной гетероплазмии в этом участке [57].

Некоторые SNP-маркеры также характеризуются менее выраженным аллельным дисбалансом (преимущественно в случае использования платформы MiSeq) или высокими показателями неправильного включения. К сожалению, на сегодняшний день нет единого стандарта параметров для анализа полученных данных. Один из примеров – порог гомозиготного генотипа, рекомендованный в трех различных исследованиях: 95% в работе [47], 90% [22] или 85% [49].

Помимо фирмы Qiagen, многие коммерческие компании, например такие как Promega (США) и NimaGene (Нидерланды), также выпустили панели для ДНК-идентификации с использованием технологии МПС.

Набор PowerSeq Auto Kit (Promega) включает 23 стандартных аутосомных STR-маркеров из системы CODIS, один Y-STR и амелогениновый маркер пола. При валидационных исследованиях PowerSeq на платформе MiSeq [59, 60], маркер D22S1045 не вызвал замечаний, в отличие от ранее упомянутых проблем в наборе ForenSeq DNA Signature Prep Kit (Verogen). Исследования чувствительности с помощью системы PowerSeq Auto демонстрируют “выпадение” аллелей только при использовании менее 62 пг ДНК [60]. Parson и др. провели эксперименты по оптимизации системы PowerSeq Auto/Y и указали на целый ряд усовершенствований лабораторного процесса путем удаления, модификации или автоматизация различных шагов [61].

Для использования дополнительных информативных полиморфных вариантов в STR-маркерах также требуется коррекция номенклатуры этих маркеров. Были предложены различные системы номенклатуры STR-аллелей, содержащих дополнительные SNP [17, 61, 62]; кроме того, комиссия ДНК Международного общества судебной генетики опубликовала ряд своих рекомендаций [63]. Тем не менее на сегодняшний день общепризнанных рекомендаций для номенклатуры STR еще не выпущено. Одна из причин этой задержки – недостаточный объем популяционных данных для создания единой номенклатуры. Эта проблема частично решается в последние годы – уже представлены результаты крупных популяционных исследований [20, 21, 64, 65] и разработана база STRseq для каталогизации вариантов STR-маркеров [20].

ТЕХНОЛОГИЧЕСКИЕ ПЕРСПЕКТИВЫ СЕКВЕНИРОВАНИЯ НА ДРУГИХ ПЛАТФОРМАХ

Компания MGI Tech Co., Ltd. (Шэньчжэнь, Китай) недавно запустила серию новых машин МПС, основанных на технологии комбинаторного синтеза зонд-якорь (cPAS, combinatorial Probe-Anchor Synthesis) “наношариков” ДНК (DNB, DNA Nanoballs) [66]. При подготовке библиотек для секвенирования ДНК расщепляют до фрагментов длиной около 500 пн, к ней лигируют специальные адаптеры, обеспечивающие последующую циркуляризацию фрагментов ДНК и амплификацию их по типу катящегося колеса с укладкой длинной цепи ДНК в компактные “наношарики”. Использование такого типа клональной амплификации исключает накопление ошибок ДНК-полимеразы в ходе ПЦР, так как матрицей в каждом цикле служит

только исходный фрагмент ДНК. Дальнейшее секвенирование происходит с использованием подхода комбинаторного синтеза зонд-якорь, при этом генерируются считывания парных концов длиной 50–150 нуклеотидов.

На сегодняшний день опубликованы лишь единичные работы, представляющие использование платформы MGI при проведении судебно-медицинской экспертизы. Первая работа была посвящена определению отцовства с использованием SNP-маркеров в исследовании Chang и др. [67].

Совсем недавно компания MGIEasy (MGI Tech, Шэньчжэнь, Китай) разработала криминалистическую панель для использования на платформе MGI [68]. Указанная панель позволяет проводить одновременный анализ 362 генетических маркеров, включая маркеры ядерного и митохондриального геномов. Панель включает маркеры пола, 52 аутосомных STR-маркера, 27 STR-маркеров X-хромосомы, 48 STR-маркеров Y-хромосомы и 227 SNP-маркеров для идентификации личности, определения вероятного этногеографического происхождения и вероятной внешности, а также позволяет анализировать все гипервариабельные регионы мтДНК. В качестве преимущества системы MGIEasy разработчики выделяют малый размер ампликонов для маркеров, входящих в панель (около 150 пн), при этом для 210 SNP-маркеров он составляет меньше 90 пн. В работе Ran Li представлены первые результаты валидации этого набора и самой платформы MGI для криминалистики [68].

Хотя секвенирование с помощью технологии синтеза (SBS) в настоящее время все же доминирует в области МПС, предпринимается множество попыток ввести другие подходы и преодолеть недостатки секвенирования 2-го поколения [2].

В настоящее время наиболее перспективной технологией является нанопоровое секвенирование, основанное на детекции изменения ионного тока при прохождении молекулы ДНК через белковую пору. Эта технология представляет собой секвенирование в реальном времени с возможностью определять последовательность ДНК длиной несколько сотен тысяч нуклеотидов.

Недавно был представлен портативный секвениратор на основе нанопор MinION (Oxford Nanopore Technologies). Cornelis et al. впервые исследовали возможность его использования в криминалистике для анализа SNP-маркеров [69]. Авторы проанализировали контрольный образец 9947 A (Promega, Madison, US) по 52 SNP-маркерам, разработанным консорциумом SNPforID [27], и сопоставили данные с генотипами, полученными с помощью секвенирования на платформе Illumina. Для анализа использовали 2.5 нг ДНК, полученные ампликоны лигировали друг с другом случайным образом для создания более длинных

фрагментов ДНК, необходимых для нанопорового секвенирования. Среднее полученное в результате секвенирования покрытие локуса составило 17933, что достаточно для генотипирования SNP-маркеров. Для маркера rs1029047 было получено наименьшее количество чтений, вероятно из-за расположения в участке поли-А тракта. Об аналогичных трудностях секвенирования гомополимерных областей также сообщали в работе Loman et al. [70]. Для двух локусов SNP (rs143232 и rs1031825), один из которых соседствует, а второй находится внутри гомополимерного участка (см. табл. 3), наблюдался значительный аллельный дисбаланс. В результате анализа при помощи пакета программ Metrichor service, разработанного Oxford Nanopore Technologies, генотип для маркера rs1031825 в гомозиготном образце был неправильно определен как гетерозиготный. Очевидно, при разработке панелей маркеров, ориентированных на систему MinION, следует особенно строго избегать гомополимерных участков.

В настоящий момент к основным недостаткам системы MinION относятся высокая частота ошибок при секвенировании (до 15%, по сравнению с 0.5% у Illumina) и несовместимость с существующими криминалистическими базами данных, содержащими в основном STR-профили индивидов [71, 72], что однако относится и к SNP-панелям для других платформ. Тем не менее недавние исследования [72, 73], в которых проверялась применимость этой технологии для идентификации ДНК, демонстрируют хорошие перспективы для использования системы MinION в судебно-медицинских исследованиях. Так, Zaaijer et al. [72] разработали специальный модуль для биоинформационной обработки данных, полученных с помощью MinION, и показали, что этот подход позволяет достичь совпадение профилей ДНК для 91–195 SNP с точностью 99.9%, при этом сам процесс секвенирования занимает всего 3 минуты. Хотя подход, использованный в этом исследовании, пока не совместим с существующими методологиями судебной экспертизы и базами данных, он демонстрирует прекрасный потенциал для оперативной ДНК-идентификации в полевых условиях, когда доступ к стандартным технологиям профилирования ДНК ограничен, а также в критически ограниченных по времени случаях, когда надо повторно идентифицировать неизвестного индивида или сделать анализ близкого родства у жертв катастроф непосредственно на месте происшедшего [73].

ПЕРСПЕКТИВЫ И НАПРАВЛЕНИЯ РАЗВИТИЯ

К преимуществам использования МПС можно отнести возможность анализировать и другие типы нуклеотидных последовательностей. На-

пример, анализ статуса метилирования определенных регионов ДНК для определения возраста индивида [74], использование РНК-маркеров для идентификации жидкостей организма [75]. Интересно, что уже апробируются для криминалистических целей недавно разработанные протоколы МПС, которые позволяют использовать образцы с малым количеством геномного материала для одновременного анализа РНК и ДНК. Использование новых видов маркеров актуально для особых случаев идентификации, например дифференцирования одного из монозиготных близнецов [76].

В связи с тем, что современные генетические технологии уже широко применяются в криминалистике, возникла необходимость создания на основе результатов генетических экспертиз судебных баз данных. На данный момент уже более 60 стран мира создали национальные базы данных судебно-медицинской экспертизы ДНК на основе STR-маркеров, среди которых, например, общепринятая база данных CODIS и Национальная база данных ДНК Великобритании. Количество и объемы информации, содержащиеся в этих базах данных, продолжают быстро расти. Например, национальная база данных судебной экспертизы Китая насчитывает более 27 миллионов записей об индивидуальных STR-профилях. Совместить уже имеющиеся базы данных и профили ДНК, полученные с использованием МПС, — задача, которую необходимо решить. Разработаны программы и веб-приложения, чтобы помочь судебным экспертам просто и эффективно работать с данными МПС, особенно при типировании STR маркеров [77, 78]. Разрабатываются алгоритмы, которые генерируют имена для секвенированных аллелей STR на основе старой номенклатуры, но с учетом новых SNP, в соответствии с рекомендациями Международного общества судебной генетики (ISFG) [61], например программа STRNaming [79].

Кроме того, уже сейчас возникает необходимость внесения изменений в международные и национальные законодательства с целью хранения и использования в судебной практике генетических данных, определенных с помощью как традиционных методов судебно-генетического анализа, так и новых видов геномных данных, полученных с помощью методов и платформ МПС [80, 81]. В Российской Федерации в настоящее время рассматриваются изменения закона “О государственной геномной регистрации в Российской Федерации”, с целью формирования федеральной базы данных геномной информации (ФБДГИ) для более эффективного использования данных полного генома для раскрытия и предотвращения преступления, в том числе особо тяжких (террористических актов, убийств и изнасилований) [82].

Еще одна немаловажная проблема, стоящая на пути массового внедрения МПС в область судебной медицины и криминалистики, — необходимость подготовки большого количества профильных специалистов высокого уровня.

ЗАКЛЮЧЕНИЕ

На сегодняшний день среди платформ для масштабного параллельного секвенирования приборы Illumina имеют значительные преимущества перед другими платформами благодаря высокой точности анализа, большой пропускной способности, а также широкого диапазона доступных коммерческих панелей. Тем не менее высокая степень кросс-платформенной совместимости позволяет предположить, что в будущем возможна успешная адаптация разработанных криминалистических панелей для использования на любых платформах для масштабного параллельного секвенирования.

Технологии секвенирования ДНК совершенствуются очень быстрыми темпами: модернизируются новые методы МПС (секвенирование 3-го поколения), разрабатываются полностью автоматизированные системы (конвейеры), в которых весь процесс подготовки образцов и секвенирования происходит без вмешательства человека. Очевидно, что уже в ближайшем будущем платформы МПС будут использоваться при проведении рутинных судебно-медицинских исследований ДНК.

Работа поддержана Государственным контрактом № 011-17 от 26.09.2017 г.

Настоящая статья не содержит каких-либо исследований с использованием в качестве объекта животных.

Настоящая статья не содержит каких-либо исследований с участием в качестве объекта людей.

Авторы заявляют, что у них нет конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

1. Margulies M., Egholm M., Altman W.E. et al. Genome sequencing in microfabricated high-density picolitre reactors // *Nature*. 2005. V. 437(7057). P. 376–380. <https://doi.org/10.1038/nature03959>
2. Goodwin S., McPherson J.D., McCombie W.R. Coming of age: Ten years of next-generation sequencing technologies // *Nat. Rev. Genet.* 2016. V. 17. № 6. P. 333–351. <https://doi.org/10.1038/nrg.2016.49>
3. https://dnadatabank.forensischinstituut.nl/binaries/dna-jaarverslag-2015_tcm37-87649.pdf
4. Iozzi S., Carboni I., Contini E. et al. Forensic genetics in NGS era: New frontiers for massively parallel typing // *Forensic Sci. Int. Genet. Suppl. Ser.* 2015. V. 5. P. e418–e419. <https://doi.org/10.1016/j.fsigs.2015.09.166>

5. *Alonso A., Müller P., Roewer L. et al.* European survey on forensic applications of massively parallel sequencing // *Forensic Sci. Int. Genet.* 2017. V. 29. P. 23–25. <https://doi.org/10.1016/j.fsigen.2017.04.017>
6. *Fuller C.W., Middendorf L.R., Benner S.A. et al.* The challenges of sequencing by synthesis // *Nat. Biotechnol.* 2009. V. 27(11). P. 1013–1023. <https://doi.org/10.1038/nbt.1585>
7. *Rhoads A., Au K.F.* PacBio sequencing and its applications // *Genomics Proteomics Bioinformatics.* 2015. V. 13(5). P. 278–289. <https://doi.org/10.1016/j.gpb.2015.08.002>
8. *Jain M., Fiddes I.T., Miga K.H. et al.* Improved data analysis for the MinIONnanopore sequencer // *Nat. Methods.* 2015. V. 12(4). P. 351–356. <https://doi.org/10.1038/nmeth.3290>
9. *Rothberg J.M., Hinz W., Rearick T.M. et al.* An integrated semiconductor device enabling non-optical genome sequencing // *Nature.* 2011. V. 475(7356). P. 34852. <https://doi.org/10.1038/nature10242>
10. *Bentley D.R., Balasubramanian S., Swerdlow H.P. et al.* Accurate whole human genome sequencing using reversible terminator chemistry // *Nature.* 2008. V. 456. P. 53–59. <https://doi.org/10.1038/nature07517>
11. *Zhao X., Li H., Wang Z. et al.* Massively parallel sequencing of 10 autosomal STRs in Chinese using the ion torrent personal genome machine (PGM) // *Forensic Sci. Int. Genet.* 2016. V. 25. P. 3438. <https://doi.org/10.1016/j.fsigen.2016.07.014>
12. *Quail M.A., Smith M., Coupland P. et al.* A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers // *BMC Genomics.* 2012. V. 13. P. 341. <https://doi.org/10.1186/1471-2164-13-341>
13. *Hollard C., Ausset L., Chantrel Y. et al.* Automation and developmental validation of the ForenSeq™ DNA Signature Preparation kit for high-throughput analysis in forensic laboratories // *Forensic Sci. Int. Genet.* 2019. V. 40. P. 37–45. <https://doi.org/10.1016/j.fsigen.2019.01.010>
14. *Jeffreys A.J., Wilson V., Thein S.L.* Hypervariable “minisatellite” regions in human DNA // *Nature.* 1985. V. 314. P. 67–73. <https://doi.org/10.1038/314067a0>
15. *Rogaev E.I.* Two novel human DNA tandem repeat families from the hypervariable DNA probe homologous to human apolipoprotein CII-gene intron and *D. virilis* satellite // *Nucl. Acids. Res.* 1989. V. 17. P. 1246. <https://doi.org/10.1093/nar/17.3.1246>
16. *Перепечина И.О.* Обмен информацией о результатах криминалистического исследования ДНК: международный опыт // Сб. матер. 54-х криминалистических чтений “Судебная 88 экспертиза в парадигме российской науки (к 85-летию Ю.Г. Корухова)”: В 2-х ч. М.: Акад. управления МВД России, 2013. P. 140–146.
17. *Gelardi C., Rockenbauer E., Dalsgaard S. et al.* Second generation sequencing of three STRs D3S1358, D12S391 and D21S11 in Danes and a new nomenclature for sequenced STR alleles // *Forensic Sci. Int. Genet.* 2014. V. 12. P. 38–41. <https://doi.org/10.1016/j.fsigen.2014.04.016>
18. *Gettings K.B., Aponte R.A., Vallone P.M. et al.* STR allele sequence variation: Current knowledge and future issues // *Forensic Sci. Int. Genet.* 2015. V. 18. P. 118–130. <https://doi.org/10.1016/j.fsigen.2015.06.005>
19. *Devesse L., Ballard D., Davenport L. et al.* Concordance of the ForenSeq system and characterisation of sequence-specific autosomal STR alleles across two major population groups // *Forensic Sci. Int. Genet.* 2017. V. 34. P. 57–61. <https://doi.org/10.1016/j.fsigen.2017.10.012>
20. *Phillips C., Gettings K.B., King J.L. et al.* “The devil’s in the detail”: release of an expanded, enhanced and dynamically revised forensic STR sequence guide // *Forensic Sci. Int. Genet.* 2018. V. 34. P. 162–169. <https://doi.org/10.1016/j.fsigen.2018.02.017>
21. *Gettings K.B., Borsuk L.A., Ballard D. et al.* STRSeq: A catalog of sequence diversity at human identification short tandem repeat loci // *Forensic Sci. Int. Genet.* 2017. V. 31. P. 111–117. <https://doi.org/10.1016/j.fsigen.2017.08.017>
22. *van der Gaag K.J., de Leeuw R.H., Hoogenboom J. et al.* Massively parallel sequencing of short tandem repeats – population data and mixture analysis results for the PowerSeq system // *Forensic Sci. Int. Genet.* 2016. V. 24. P. 86–96. <https://doi.org/10.1016/j.fsigen.2016.05.016>
23. *Phillips C., Fang R., Ballard D. et al.* Evaluation of the Genplex SNP typing system and a 49plex forensic marker panel // *Forensic Sci. Int. Genet.* 2007. V. 1(2). P. 180–185. <https://doi.org/10.1016/j.fsigen.2007.02.007>
24. *Tozzo P., Gabbin A., Politi C. et al.* Combined statistical analyses of forensic evidence in sexual assault: A case report and brief review of the literature // *J. Forensic Sci.* 2020. V. 65. P. 1767–1773. <https://doi.org/10.1111/1556-4029.14487>
25. *Amorim A., Pereira L.* Pros and cons in the use of SNPs in forensic kinship investigation: A comparative analysis with STRs // *Forensic Sci. Int.* 2005. V. 150. № 1. P. 17–21. <https://doi.org/10.1016/j.forsciint.2004.06.018>
26. *Dixon L.A., Murray C.M., Archer E.A. et al.* Validation of a 21-locus autosomal SNP multiplex for forensic identification purposes // *Forensic Sci. Int.* 2005. V. 154. № 1. P. 62–77. <https://doi.org/10.1016/j.forsciint.2004.12.011>
27. *Sanchez J., Phillips C., Borsting C. et al.* A multiplex assay with 52 single nucleotide polymorphisms for human identification // *Electrophoresis.* 2006. V. 27. № 9. P. 1713–1724. <https://doi.org/10.1002/elps.200500671>
28. *Dario P., Oliveira A.R., Ribeiro T. et al.* SNPforID 52-plex in casework samples: “Cracking” bones and other difficult samples // *Forensic Sci. Int. Genet. Suppl. Ser.* 2015. V. 5. P. e118–e120. <https://doi.org/10.1016/j.fsigs.2015.09.048>
29. *Kidd K.K., Speed W.C., Pakstis A.J. et al.* Progress toward an efficient panel of SNPs for ancestry inference // *Forensic Sci. Int. Genet.* 2014. V. 12. P. 38–41. <https://doi.org/10.1016/j.fsigen.2014.04.016>

- rensic Sci. Int. Genet. 2014. V. 10. P. 23–32.
<https://doi.org/10.1016/j.fsigen.2014.01.002>
30. Walsh S., Liu F., Wollstein A. et al. The HiRiPlex system for simultaneous prediction of hair and eye colour from DNA // Forensic Sci. Int. Genet. 2013. V. 7. № 1. P. 98–115.
<https://doi.org/10.1016/j.fsigen.2012.07.005>
31. Kidd K.K., Pakstis A.J., Speed W.C. et al. Developing a SNP panel for forensic identification of individuals // Forensic Sci. Int. 2006. V. 164. № 1. P. 20–32.
<https://doi.org/10.1016/j.forsciint.2005.11.017>
32. Daniel R., Santos C., Phillips C. et al. A SNaPshot of next generation sequencing for forensic SNP analysis // Forensic Sci. Int. Genet. 2015. V. 14. P. 50–60.
<https://doi.org/10.1016/j.fsigen.2014.08.013>
33. Grandell I., Samara R., Tillmar A.O. A SNP panel for identity and kinship testing using massive parallel sequencing // Int. J. Legal. Med. 2016. V. 130. № 4. P. 90514.
<https://doi.org/10.1007/s00414-016-1341-4>
34. Kidd K.K., Pakstis A.J., Speed W.C. et al. Current sequencing technology makes microhaplotypes a powerful new type of genetic marker for forensics // Forensic Sci. Int. Genet. 2014. V. 12. P. 215–224.
<https://doi.org/10.1016/j.fsigen.2014.06.014>
35. Pakstis A.J., Fang R., Furtado M.R. et al. Mini-haplotypes as lineage informative SNPs and ancestry inference SNPs // Europ. J. Hum. Genet. 2012. V. 20. № 11. P. 1148–1154.
<https://doi.org/10.1038/ejhg.2012.69>
36. Wendt F.R., King J.L., Novroski N.M. et al. Flanking region variation of ForenSeq DNA signature Prep Kit STR and SNP loci in Yavapai native Americans // Forensic Sci. Int. Genet. 2017. V. 28. P. 146–154.
<https://doi.org/10.1016/j.fsigen.2017.02.014>
37. Pang J.B., Rao M., Chen Q.F. et al. A 124-plex microhaplotype panel based on next-generation sequencing developed for forensic applications // Sci. Reports. 2020. V. 10. 1945.
<https://doi.org/10.1038/s41598-020-58980-x>
38. Bulbul O., Pakstis J., Soundararajan U. et al. Ancestry inference of 96 population samples using microhaplotypes // Int. J. Legal Med. 2018. V. 132. № 3. P. 703–711.
<https://doi.org/10.1007/s00414-017-1748-6>
39. Jager A.C., Alvarez M.L., Davis C.P. et al. Developmental validation of the MiSeqFGx forensic genomics system for targeted next generation sequencing in forensic DNA casework and database laboratories // Forensic Sci. Int. Genet. 2017. V. 28. P. 52–70.
<https://doi.org/10.1016/j.fsigen.2017.01.011>
40. Churchill J.D., Schmedes S.E., King J.L., Budowle B. Evaluation of the Illumina(®) Beta version ForenSeq DNA Signature Prep Kit for use in genetic profiling // Forensic Sci. Int. Genet. 2016. V. 20. P. 20–29.
<https://doi.org/10.1016/j.fsigen.2015.09.009>
41. Xavier C., Parson W. Evaluation of the Illumina ForenSeq DNA Signature Prep kit – MPS forensic application for the MiSeqFGx benchtop sequencer // Forensic Sci. Int. Genet. 2017. V. 28. P. 188–194.
<https://doi.org/10.1016/j.fsigen.2017.02.018>
42. Silvia A.L., Shugarts N., Smith J. A preliminary assessment of the ForenSeqFGx system: Next generation sequencing of an STR and SNP multiplex // Int. J. Legal Med. 2017. V. 131. № 1. P. 73–86.
<https://doi.org/10.1007/s00414-016-1457-6>
43. Eduardoff M., Gross T.E., Santos C. et al. Inter-laboratory evaluation of the EUROFORGEN Global ancestry-informative SNP panel by massively parallel sequencing using the Ion PGM // Forensic Sci. Int. Genet. 2016. V. 23. P. 178–189.
<https://doi.org/10.1016/j.fsigen.2016.04.008>
44. Parson W., Strobl C., Huber G. et al. Evaluation of next generation mtGenome sequencing using the ion torrent personal genome machine (PGM) // Forensic Sci. Int. Genet. 2013. V. 7. № 5. P. 543–549.
<https://doi.org/10.1016/j.fsigen.2013.06.003>
45. Almalki N., Chow H.Y., Sharma V. et al. Systematic assessment of the performance of Illumina’s MiSeqFGx forensic genomics system // Electrophoresis. 2017. V. 38. № 6. P. 846–854.
<https://doi.org/10.1002/elps.201600511>
46. Churchill J.D., Novroski N.M.M., King J.L. et al. Population and performance analyses of four major populations with Illumina’s FGx forensic genomics system // Forensic Sci. Int. Genet. 2017. V. 30. P. 81–92.
<https://doi.org/10.1016/j.fsigen.2017.06.004>
47. Just R.S., Moreno L.I., Smerick J.B., Irwin J.A. Performance and concordance of the ForenSeq system for autosomal and Y chromosome short tandem repeat sequencing of reference-type specimens // Forensic Sci. Int. Genet. 2017. V. 28. P. 1–9.
<https://doi.org/10.1016/j.fsigen.2017.01.001>
48. Pereira V., Mogensen H.S., Borsting C., Morling N. Evaluation of the precision ID ancestry panel for crime case work: A SNP typing assay developed for typing of 165 ancestral informative markers // Forensic Sci. Int. Genet. 2017. V. 28. P. 138–145.
<https://doi.org/10.1016/j.fsigen.2017.02.013>
49. Al-Asfi M., McNevin D., Mehta B. et al. Assessment of the precision ID ancestry panel // Int. J. Legal Med. 2018. V. 132(6). P. 1581–1594.
<https://doi.org/10.1007/s00414-018-1785-9>
50. Jin S., Chase M., Henry M. et al. Implementing a biogeographic ancestry inference service for forensic casework // Electrophoresis. 2018. V. 39. № 21. P. 2757–2765.
<https://doi.org/10.1002/elps.201800171>
51. Fordyce S.L., Mogensen H.S., Borsting C. et al. Second-generation sequencing of forensic STRs using the Ion Torrent HID STR 10-plex and the Ion PGM // Forensic Sci. Int. Genet. 2015. V. 14. P. 132–140.
<https://doi.org/10.1016/j.fsigen.2014.09.020>
52. Guo F., Zhou Y., Liu F. et al. Evaluation of the early access STR kit v1 on the Ion Torrent PGM platform // Forensic Sci. Int. Genet. 2016. V. 23. P. 111–120.
<https://doi.org/10.1016/j.fsigen.2016.04.004>
53. Kulstein G., Pably P., Furst A. et al. “The acid test” – validation of the ParaDNA® Body Fluid ID Test for routine forensic casework // Int. J. Legal Med. 2019. V. 133. № 3. P. 751–757.
<https://doi.org/10.1016/j.fsigen.2014.09.020>

54. *Novroski N.M.M., King J.L., Churchill J.D. et al.* Characterization of genetic sequence variation of 58 STR loci in four major population groups // *Forensic Sci. Int. Genet.* 2016. V. 25. P. 214–226.
<https://doi.org/10.1016/j.fsigen.2016.09.007>
55. *Müller P., Alonso A., Barrio P.A. et al.* Systematic evaluation of the early access applied biosystems precision ID Globalfiler mixture ID and Globalfiler NGS STR panels for the ion S5 system // *Forensic Sci. Int. Genet.* 2018. V. 36. P. 95–103.
<https://doi.org/10.1016/j.fsigen.2018.06.016>
56. *Gettings K.B., Kiesler K.M., Faith S.A. et al.* Sequence variation of 22 autosomal STR loci detected by next generation sequencing // *Forensic Sci. Int. Genet.* 2016. V. 21. P. 15–20.
<https://doi.org/10.1016/j.fsigen.2015.11.005>
57. *Wang Z., Zhou D., Wang H. et al.* Massively parallel sequencing of 32 forensic markers using the precision ID GlobalFiler NGS STR panel and the ion PGM system // *Forensic Sci. Int. Genet.* 2017. V. 31. P. 126–134.
<https://doi.org/10.1016/j.fsigen.2017.09.004>
58. *de la Puente M., Phillips C., Santos C. et al.* Evaluation of the Qiagen 140-SNP forensic identification multiplex for massively parallel sequencing // *Forensic Sci. Int. Genet.* 2017. V. 28. P. 35–43.
<https://doi.org/10.1016/j.fsigen.2017.01.012>
59. *Zeng X., King J., Hermanson S. et al.* An evaluation of the PowerSeq Auto System: A multiplex short tandem repeat marker kit compatible with massively parallel sequencing // *Forensic Sci. Int. Genet.* 2015. V. 19. P. 172–179.
<https://doi.org/10.1016/j.fsigen.2015.07.015>
60. *Montano E.A., Bush J.M., Garver A.M. et al.* Optimization of the PromegaPowerSeq Auto/Y system for efficient integration within a forensic DNA laboratory // *Forensic Sci. Int. Genet.* 2018. V. 32. P. 26–32.
<https://doi.org/10.1016/j.fsigen.2017.10.002>
61. *Parson W., Ballard D., Budowle B. et al.* Massively parallel sequencing of forensic STRs: Considerations of the DNA commission of the International Society for Forensic Genetics (ISFG) on minimal nomenclature requirements // *Forensic Sci. Int. Genet.* 2016. V. 22. P. 54–63.
<https://doi.org/10.1016/j.fsigen.2016.01.009>
62. *Huszar T.I., Jobling M.A., Wetton J.H.* A phylogenetic framework facilitates Y-STR variant discovery and classification via massively parallel sequencing // *Forensic Sci. Int. Genet.* 2018. V. 35. P. 97–106.
<https://doi.org/10.1016/j.fsigen.2018.03.012>
63. *Gettings K., Ballard D., Bodner M. et al.* Report from the STRAND Working Group on the 2019 STR sequence nomenclature meeting // *Forensic Sci. Int. Genet.* 2019. V. 43. P. 10216.
<https://doi.org/10.1016/j.fsigen.2019.102165>
64. *Gettings K.B., Borsuk L.A., Steffen C.R. et al.* Sequence-based U.S. population data for 27 autosomal STR loci // *Forensic Sci. Int. Genet.* 2018. V. 37. P. 106–115.
<https://doi.org/10.1016/j.fsigen.2018.07.013>
65. *Gunnarsdottir E.D., Li M., Bauchet M.F. et al.* High-throughput sequencing of complete human mtDNA genomes from the Philippines // *Genome Res.* 2011. V. 21. № 1. P. 1–11.
<https://doi.org/10.1101/gr.107615.110>
66. *Drmanac R., Sparks A.B., Callow M.J. et al.* Human genome sequencing using unchained base reads on Self-Assembling DNA nanoarrays // *Science.* 2010. V. 327. iss. 5961. P. 78–81.
<https://doi.org/10.1126/science.1181498>
67. *Chang L., Yu H., Miao X. et al.* Development and comprehensive evaluation of a noninvasive prenatal paternity testing method through a scaled trial // *Forensic Sci. Int. Genet.* 2019. V. 43. P. 102158.
<https://doi.org/10.1016/j.fsigen.2019.102158>
68. *Ran Li., Xuefeng Shen, Hui Chen et al.* Developmental validation of the MGIEasy Signature Identification Library Prep Kit, An all-in-one multiplex system for forensic applications // *Int. J. Legal Med.* 2021. V. 135. P. 739–753.
<https://doi.org/10.1007/s00414-021-02507-0>
69. *Cornelis S., Gansemans Y., Deleye L. et al.* Forensic SNP genotyping using nanopore MinION sequencing // *Sci. Reports.* 2017. V. 7. P. 41759.
<https://doi.org/10.1038/srep41759>
70. *Loman N.J., Quick J., Simpson J.T.* A complete bacterial genome assembled de novo using only nanopore sequencing data // *Nat. Methods.* 2015. V. 12. № 8. P. 733–735.
<https://doi.org/10.1038/nmeth.3444>
71. *Kchouk M., Gibrat J.-F., Elloumi M.* Generations of sequencing technologies: From first to next generation // *Biol. Med.* 2017. V. 9. P. 395.
<https://doi.org/10.4172/0974-8369.1000395>
72. *Zaaijer S., Gordon A., Speyer D. et al.* Rapid re-identification of human samples using portable DNA sequencing // *eLife.* 2017. V. 6. P. e27798.
<https://doi.org/10.7554/eLife.27798>
73. *Zubakov D., Kokmeijer I., Ralf A. et al.* Towards simultaneous individual and tissue identification: A proof-of-principle study on parallel sequencing of STRs, amelogenin, and mRNAs with the Ion Torrent PGM // *Forensic Sci. Int. Genet.* 2015. V. 17. P. 122–128.
<https://doi.org/10.1016/j.fsigen.2015.04.002>
74. *Vidaki A., Daniel B., Syndercombe Court D.* Forensic DNA methylation profiling – potential opportunities and challenges // *Forensic Sci. Int. Genet.* 2013. V. 7. P. 499–507.
<https://doi.org/10.1016/j.fsigen.2013.05.004>
75. *Albani P.P., Fleming R.* Novel messenger RNAs for body fluid identification // *Sci. Justice.* 2018. V. 58. P. 145–152.
<https://doi.org/10.1016/j.scijus.2017.09.002>
76. *Turrina S., Bortoletto E., Giannini G., De Leo D.* Monozygotic twins: Identical or distinguishable for science and law? // *Med Sci Law.* 2021. V. 61. № 1. P. 62–66.
<https://doi.org/10.1177/0025802420922335>
77. *Ganschow S., Silvery J., Kalinowski J., Tiemann C.* toaSTR: A web application for forensic STR genotyping by massively parallel sequencing // *Forensic Sci. Int. Genet.* 2018. V. 37. P. 21–28.
<https://doi.org/10.1016/j.fsigen.2018.07.006>
78. *Hoogenboom J., van der Gaag K., de Leeuw R. et al.* FDSTools: A software package for analysis of massively parallel sequencing data with the ability to recognise and correct STR stutter and other PCR or sequencing

- noise // *Forensic Sci. Int. Genet.* 2017. V. 27. P. 27–40. <https://doi.org/10.1016/j.fsigen.2016.11.007>
79. Hoogenboom J., TitiaSijen T., van der Gaag K. STRNaming: Generating simple, informative names for sequenced STR alleles in a standardised and automated manner // *Forensic Sci. Int. Genet.* 2021. V. 52. P. 102473. <https://doi.org/10.1016/j.fsigen.2021.102473>
80. Чемерис А., Анисимов В., Аминев Ф. Актуальные вопросы правового регулирования, создания и функционирования баз данных генома человека // *Правовое государство: теория и практика.* 2020. Т. 16. № 82. С. 134–144. <https://doi.org/10.33184/pravgos-2020.2.12>
81. Vladimirova D. Analysis of judicial practice in the field of genomic and genetic research including legal positions of the ECHR and practice of Russian courts // *Ross. Pravo. Obraz. Prakt. Nauk.* 2020. № 5. P. 16–25. <https://doi.org/10.34076/2410-2709-2020-5-16-25>
82. <https://sozd.duma.gov.ru/bill/1048800-7>

Using NGS in Forensics: A Comparative Analysis of Sequencing Platforms

T. V. Tyazhelova^{a, *}, I. L. Kuznetsova^a, T. V. Andreeva^{a, b}, S. S. Kunizheva^{a, b}, and E. I. Rogayev^{a, b, c}

^a*Vavilov Institute of General Genetics Russian Academy of Sciences, Moscow, 119991 Russia*

^b*Lomonosov Moscow State University, Moscow, 119991 Russia*

^c*“Sirius” University of Science and Technology, Krasnodarskii kray, Sochi, 354340 Russia*

**e-mail: tatvlytz@gmail.com*

The last decade has been characterized by a significant breakthrough in the development of new technologies and tools for DNA analysis. A new technology of massive parallel sequencing, otherwise known as “next-generation sequencing” (NGS) is by now already widely used in many areas of biological and biomedical research. The introduction of new parallel sequencing technologies into the forensic field is slower compared to scientific and clinical research, mainly due to the difficulties of compatibility with already existing DNA profile databases, as well as the need for rigorous validation experiments for accreditation of instruments and assays sets of reagents. This review presents the worldwide existing solutions for the use of massively parallel sequencing in forensics, technical aspects of experiments, developed commercial solutions and application prospects, as well as the advantages and disadvantages of new sequencing technologies compared to the methods of DNA analysis traditionally used in forensic science.

Keywords: massive parallel sequencing (MPS), next generation sequencing (NGS), short tandem repeat (STR), single nucleotide polymorphism (SNP), forensic science.