УДК 681.7.08, 681.7.069.32, 57.621.383, 535.421, 535.37, 535.249

**В.П. Бессмельцев¹, В.С. Терентьев¹, В.В. Вилейко¹,
С.А. Бабин¹, А.М. Шалагин¹, А.В. Латышев², Д.А. Насимов²,
Л.И. Федина², Д.В. Пышный³, П.Е. Воробьев³,
В.В. Анненков⁴, Е.Н. Даниловцева⁴, С.Н. Зелинский⁴,
О.Н. Верховина⁴, М.А. Грачев⁴, Ю.П. Галачьянц⁴**

¹ Институт автоматики и электрометрии СО РАН,
Новосибирск, Россия

² Институт физики полупроводников СО РАН,
Новосибирск, Россия

³ Институт химической биологии и фундаментальной медицины СО РАН,
Новосибирск, Россия

⁴ Лимнологический институт СО РАН, Иркутск, Россия

О СОЗДАНИИ ПЛАТФОРМЫ ДЛЯ ИССЛЕДОВАНИЯ НУКЛЕИНОВЫХ КИСЛОТ (ДНК-СЕКВЕНАТОРА)

Представлен анализ современного положения дел в области методов секвенирования. Описываются достижения и проблемы, касающиеся разработки отечественной платформы оптического секвенирования ДНК по идеологии, предложенной компанией Pacific Biosciences. Продемонстрированы новые технические решения, предложенные в ходе разработки платформы.

Ключевые слова: флуоресценция, дифракция, фотометрия, фотолитография, секвенирование генома.

**V.P. Bessmel'tsev¹, V.S. Terent'ev¹, V.V. Vileiko¹, S.A. Babin¹,
A.M. Shalagin¹, A.V. Latyshev², D.A. Nasimov², L.I. Fedina²,
D.V. Pyshnyi³, P.E. Vorob'ev³, V.V. Annenkov⁴, E.N. Danilovtseva⁴,
S.N. Zelinskii⁴, O.N. Verkhovina⁴, M.A. Grachev⁴, Iu.P. Galach'iants⁴**

¹ Institute of Automation and Electrometry
of the Siberian Branch of the Russian Academy of Sciences,
Novosibirsk, Russian Federation

² Institute of Semiconductor Physics
of the Siberian Branch of the Russian Academy of Sciences,
Novosibirsk, Russian Federation

³ Institute of Chemical Biology and Fundamental Medicine
of the Siberian Branch of the Russian Academy of Sciences,
Novosibirsk, Russian Federation

⁴ Limnological Institute of the Siberian Branch
of the Russian Academy of Sciences, Irkutsk, Russian Federation

DEVELOPMENT OF THE PLATFORM FOR NUCLEIC ACID ANALYSIS (DNA SEQUENCER)

The paper analyzes the state-of-the-art in the field of sequencing techniques. The progress and problems related to the development of the platform of optical DNA sequencing in terms of ideology proposed by Pacific Biosciences is described. New technical solutions proposed during the development of the platform are demonstrated.

Keywords: fluorescence, diffraction, photometry, photolithography, genome sequencing.

Введение

Методы секвенирования ДНК развиваются с 1970-х годов. Классическим и до сих пор не потерявшим своей актуальности является метод секвенирования Сэнгера [1], который основывается на анализе амплифицированных индивидуальных фрагментов ДНК. В исходном варианте метода для исследуемого фрагмента проводят фрагментивный синтез комплементарной нити в четырех независимых реакциях с вероятностным обрывом цепи по позициям одного из четырех нуклеотидов и анализируют распределение длин получившихся фрагментов методом гель-электрофореза. По сравнению с созданными позднее методами такая технология не обладает достаточным быстродействием и требует больших расходов на реагенты.

В начале нового тысячелетия происходило развитие новых технологий секвенирования ДНК, позволяющих получать все большие объемы информации о первичной структуре нуклеиновых кислот. Во многом это развитие было продиктовано становлением и развитием персонализированной медицины, а также лозунгами «Геном человека за тысячу долларов» и «Геном тысячи человек». В течение последних 15 лет было разработано несколько новых принципов массового параллельного секвенирования.

Существующие технологии массового параллельного секвенирования (next-generation sequencing, NGS) [2] можно разделить на две группы. К методам второго поколения (second-generation sequencing, SGS) относятся технологии, в которых для увеличения уровня сигнала при секвенировании используется предварительная амплификация (размножение) единичных молекул ДНК. К этой группе методов относятся пиросеквенирование (454), секвенирование с помощью лигирования (SOLiD), секвенирование кластеров ДНК, иммобилизованных на твердой подложке, с применением модифицированных нуклеотидов в качестве обратимых терминаторов (Illumina) и полупроводниковое секвенирование (Ion Torrent). SGS-технологии являются значительным шагом вперед в секвенировании по сравнению с методом Сэнгера, поскольку позволяют получать большие объемы данных о первичной структуре молекул ДНК за сравнительно небольшое время при низкой стоимости в пересчете на нуклеотид. Например, в рамках одного запуска секвенатора Illumina HiSeq 2500 можно провести секвенирование четырех миллиардов фрагментов ДНК с получением прочтений по 125 нуклеотидным основаниям (н.о.) с двух сторон каждого фрагмента. При этом стоимость секвенирования составляет несколько десятков долларов за миллиард н.о. при однократном покрытии.

Значительными недостатками технологий SGS являются небольшая длина получаемых прочтений и необходимость амплификации анализируемых фрагментов ДНК. Они ограничивают технологии SGS при решении определенного ряда сложных биологических проблем, в частности при прямом секвенировании геномов с протяженными повторами, обнаружении изоформ транскрибирующихся генов, анализе гаплотипов, полногеномном эпигенетическом анализе, анализе фрагментных и ампликонных библиотек сообществ микроорганизмов и т.д. Высокопроизводительная технология секвенирования, позволяющая

получать длинные прочтения, уже внесла существенный вклад в развитие областей науки, связанных с геномными исследованиями не только модельных организмов, но и немодельных объектов и их сообществ.

Методы секвенирования третьего поколения (third-generation sequencing, TGS) получают информацию о первичной структуре фрагментов *индивидуальных* молекул ДНК. Технология компании Pacific Biosciences (PacBio) является технологией одномолекулярного секвенирования в режиме реального времени (single-molecule real-time sequencing, SMRT) [3] и относится к группе методов TGS наряду с технологией Oxford Nanopore [4]. Значительным преимуществом технологии PacBio является большая протяженность получаемых прочтений. Средняя длина прочтений для секвенаторов первого поколения (PacBio RS) составляла $\approx 1,5$ тыс. н.о. В настоящее время для секвенаторов PacBio RSII средняя длина прочтений превышает 10 тыс. н.о., а N50 составляет ≈ 20 тыс. н.о., т.е. более половины объема секвенирования составляют прочтения длиннее 20 тыс. н.о. Важно отметить, что технология PacBio свободна от ошибок, связанных с амплификацией, и требует на много порядков меньшего количества реагентов, в отличие от технологий SGS. У технологии PacBio, как и у Oxford Nanopore, имеются свои недостатки: это высокая вероятность ошибок и связанная с необходимостью разработки нанореакторов высокая добавленная стоимость секвенирования в пересчете на нуклеотид. Технология PacBio в сравнении с другими методами секвенирования, а также основные стратегии ее применения хорошо освещены в обзоре Rhoads и Au [5]. При планировании использования приборов технологии третьего поколения приходится учитывать, что подобные приборы относятся к группе приборов двойного применения и их импорт в Россию затруднен по ряду причин.

В технологии PacBio секвенирование фрагментов индивидуальных молекул ДНК происходит в нанореакторах. Комплексы, состоящие из одной молекулы ДНК, связанной с одной молекулой ДНК полимеразы, размещают на кварцевом дне специальных наноколодцев с линейным размером порядка 80 нм, выполненных в слое алюминия. Чип SMRT состоит из нескольких тысяч наноколодцев, подсвечиваемых снизу сфокусированными лазерными лучами с тремя длинами волн. Наноколодцы представляют собой световоды нулевой моды

(zero-mode waveguide, ZMW), и поэтому при их освещении лазерами со стороны дна свет проникает в наноколодец на высоту около 20 нм. Считывание информации о последовательности нуклеотидов происходит в ходе синтеза комплементарной цепи нуклеиновой кислоты. В качестве субстратов используются производные дезоксинуклеозидтрифосфатов, несущие флуорофоры с различающимся спектром флуоресценции, прикрепленные к их гамма-концевой фосфатной группе. В ходе синтеза комплементарной цепи очередной флуоресцентный нуклеозидтрифосфат образует комплекс со связанной с полимеразой матричной нуклеиновой кислотой. Этот комплекс находится в освещаемой зоне наноколодца и живет в течение порядка 10–100 мс – времени, необходимого для синтеза фосфодиэфирной связи и отщепления модифицированного пирофосфата, испуская свет соответствующей длины волны. После этого растущая цепь нуклеиновой кислоты удлиняется на одно звено, а флуоресцентная группа израсходованного субстрата отрывается от комплекса и покидает освещенную зону вследствие молекулярной диффузии. Затем происходят формирование нового фермент-субстратного комплекса, синтез новой фосфодиэфирной связи и т.д. Испускаемая во время жизни фермент-субстратного комплекса флуоресценция регистрируется спектрозональной оптико-электронной системой на основе сверхвысококочувствительной фотоприемной матрицы с электронным умножением (ЕМ CCD), установленной конгруэнтно с матрицей освещающих пучков после матрицы конфокальных диафрагм. Свет флуоресценции от каждого наноколодца разлагается в видимый спектр, который воспринимается несколькими элементами фотоприемника. Сигналы всех элементов фотоприемников запоминаются, идентифицируются и анализируются с помощью мощных вычислительных средств и программного обеспечения.

В основу разрабатываемого нами секвенатора был положен принцип массового параллельного оптического секвенирования фрагментов индивидуальных молекул ДНК в режиме реального времени, аналогичный разработке компании Pacific Biosciences. В этой работе представлены основные результаты, которых удалось достичь при разработке отечественной платформы для одномолекулярного секвенирования ДНК, а также опыт использования платформы PacBio на коммерческой основе.

1. Результаты анализа ядерного генома байкальской диатомеи *Synedra acus* на основе данных секвенирования, полученных на коммерческой основе с использованием оборудования иностранных фирм

Чтобы проиллюстрировать возможности технологии, мы приведем результаты секвенирования PacBio, полученные нами при анализе генома диатомовой водоросли *Synedra acus*. С целью получить набор фрагментов ДНК надлежащего качества после нескольких неудачных попыток мы пришли к выводу, что наилучшим методом подготовки ДНК *S. acus* к анализу является ее очистка путем ультрацентрифугирования в градиенте плотности CsCl. Препарат геномной ДНК выделяли из 1 г клеток аксеничной культуры *S. acus*, штамм A280 ($\approx 2 \cdot 10^8$ клеток) с помощью методики, описанной ранее в работе [6]. Дополнительную очистку препарата ДНК проводили с применением ультрацентрифугирования в градиенте плотности CsCl; препарат ДНК в растворе CsCl с плотностью 1,7 г/мл центрифугировали 28 ч при ускорении 187 тыс. g (g – ускорение свободного падения, $g = 9,8 \text{ м/с}^2$). Приготовление библиотеки SMRTbell с размером фрагментов ДНК ≈ 10 тыс. н.о. и секвенирование на приборе PacBio RSII с применением реагентов P6-C4 проводили в Центре функциональной геномики (Цюрих, Швейцария). Для получения покрытия генома $\approx 40\times$, необходимого для проведения сборки, библиотеку анализировали на пяти ячейках SMRT. Суммарный объем секвенирования составил 5,9 млрд н.о., количество полученных прочтений – ≈ 450 тыс., $N50 = 18,2$ тыс. н.о. (рис. 1).

На рис. 1 видно, что запуск проводился в условиях, близких к оптимальным, о чем свидетельствуют высокие значения количества ячеек (ZMW loading) с единичной продуктивностью, а также итоговые показатели объема секвенирования, качества и протяженности прочтений.

Мы самостоятельно провели обработку данных с использованием двух сборщиков. Программу Canu v.1.03 [7] использовали для сборки с исходным набором данных, в который входили только прочтения PacBio. Программу MaSuRCA v.3.2.1 [8] применяли для «гибридной» сборки, входными данными которой, помимо прочтений PacBio, служили результаты анализа геномных библиотек *S. acus* с различной длиной вставки на секвенаторах Illumina Miseq и Hiseq. Расчеты проводили на вычислительном кластере «Академик В.М. Матросов» [9]. В нашем случае использование данных Illumina позволило увеличить

Рис. 1. Статистика результатов секвенирования PacBio для геномной ДНК *S. asus*

протяженность сборки более чем в полтора раза и лишь незначительно улучшить ее смежность (рис. 2). Мы предполагаем, что второе обстоятельство связано со сложной структурой генома *S. acuis*, в котором имеются протяженные повторяющиеся участки. Результаты сборки показывают, что при использовании данных PacBio с достаточным покрытием (от 40–50 крат и более) возможно получить геномную сборку, сравнимую по смежности и протяженности с «гибридной» сборкой с более высоким покрытием (таблица). Столь большая длина сборки является преимуществом технологии PacBio.

Рис. 2. График зависимости относительной доли покрытия сборки от протяженности контигов

Статистика результатов сборки генома *S. acuis*

Параметр	Сборка PacBio	«Гибридная» сборка
Сборщик	Canu v.1.3	MaSuRCa v.3.2.1
Среднее покрытие ¹	41×	260×
Исходные данные для сборки, млрд н.о.	PacBio – 6	PacBio – 6, Miseq PE 2×250 – 3, Miseq MP 2×250 – 3, Hiseq PE 2×125 – 18
Общая протяженность сборки, млн н.о.	104	167
Количество контигов/скэффолдов ²	1807/1807	4095/3834
N50, тыс. н.о.	87	81
L50 для контигов/скэффолдов	274/274	344/348
E-size, тыс. н.о. ³	160	243
N's/100, тыс. н.о.	0	14
GC-состав, %	38,46	39,30

Примечание.

¹ При протяженности генома ≈150 млн н.о.

² Количество контигов/скэффолдов длиннее N50.

³ E-size – ожидаемая протяженность контига при случайном выборе координаты сборки (рассчитывается по формуле $\sum_i (L_i^2)/G$, где L_i – длина контига; G – общая протяженность сборки).

Технология РасВю быстро развивается. Можно предположить, что в ближайшем будущем будут продемонстрированы ее более совершенные варианты.

2. Оптическая схема

Основная задача оптической схемы (рис. 3) прибора, разрабатываемого в рамках настоящей работы, состояла в одновременном освещении каждого наноклодда чипа (см. вставку на рис. 3) сфокусированными лазерными пучками трех длин волн (488, 532, 638 нм) и фиксации флуоресцентного света из наноклодцев на фотоматрице [10]. Принцип функционирования оптической схемы заключается в мультиплицировании исходного луча каждого лазерного источника ($ЛИ_{1,2,3}$) в $N \times N$ матрицу лучей и построении изображения этой матрицы в плоскости объекта (чип) с помощью проекционной системы. На рис. 3 показано положение специально разработанных дифракционно-оптических элементов ($ДОЭ_{1,2,3}$) для трех длин волн с целью создания матрицы 25×25 лучей, которые при этом фокусировались на расстоянии 150 мм в плоскости матрицы конфокальных диафрагм ($МД_1$),

Рис. 3. Принципиальная оптическая схема прибора для секвенирования ДНК: $ЛИ_{1,2,3}$ – лазерный источник; $ДОЭ_{1,2,3}$ – дифракционные оптические элементы; $Ф_{1,2}$ – зеркала-фильтры; $МД_{1,2}$ – матрицы диафрагм; $ТЛ_{1,2}$ – тубусные линзы; ДФ, ПФ – дихроичные светофильтры; О – объектив; чип – объект; $Л_{1,2}$ – линзы; П – призма; ВМ – видеоматрица

которая пространственно фильтрует матрицу лучей от паразитных порядков и спеклов. Далее с помощью тубусной линзы (ТЛ₁) и объектива (О) матрица лучей фокусируется с коэффициентом 1:60 в объектную плоскость (чип), где размещается матрица наноколодцев.

Свет, возникающий в наноколодцах в результате флуоресценции, направляется с помощью дихроичных светофильтров на три лазерные длины волны: 45° Semrock Di01-T488/532/638-25 (ДФ) и 90° полосового фильтра (ПФ) на фоточувствительную ЕМ CCD видеоматрицу (ВМ). При этом светоделитель ДФ выбран так, чтобы лазерное излучение трех длин волн не ослаблялось в осветительном канале и частично отраженное от металлической пленки чипа не проходило на видеоматрицу. Отраженный от светоделителя сигнальный свет направляется оптической системой на видеоматрицу через соответствующие (конгруентные) конфокальные отверстия (микродиафрагмы МД₂), чтобы уменьшить количество шумового рассеянного света.

Сигнальный флуоресцентный свет проходит призму (П), расположенную между линзами (Л_{1,2}). Призма разделяет свет, принятый от каждого наноколодца, на спектральные составляющие и пространственно распределяет их на светочувствительных элементах видеоматрицы (см. рис. 3). По распределению можно определять тип флуорофора и, следовательно, основание молекулы ДНК.

Наиболее важными параметрами ДОЭ являются его эффективность, т.е. отношение суммарной интенсивности света в матрице лучей к падающей на ДОЭ, а также равномерность интенсивности лучей внутри матрицы.

Для каждой лазерной длины волны использовался соответствующий ДОЭ для длин волн 488, 532 и 638 нм. Они были рассчитаны и изготовлены в амплитудном бинарном исполнении с использованием круговой лазерной записывающей системы CLWS-300C/M (технология прямой лазерной записи на пленках хрома), однако их относительная интенсивность была низка. В связи с этим были разработаны три ДОЭ – в бинарном фазовом исполнении на подложках из кварца, имеющие значительно лучшее распределение. На рис. 4 показано сравнение распределения интенсивностей для амплитудного и фазового элементов бинарного типа в плоскости фокусировки.

Рис. 4. Изображение (а) и сечение распределения интенсивностей (б) для амплитудного и фазового ДОЭ, изготовленного для работы с лазерным пучком 638 нм

3. Матрица наноканальных реакторов (наноколодцев)

Важную роль в данной технологии играет матрица наноканальных реакторов (наноколодцев) (см. рис. 3, чип). В нашем случае чип представляет собой эквидистантно расположенные на расстоянии 5 мкм по поперечным координатам наноколодцы диаметром 100 нм в пленке алюминия толщиной 100 нм. Малые габариты наноколодцев, в которых протекает биологическая реакция, обеспечивают возможность наблюдения одномолекулярных событий и, в частности, позволяют осуществлять ДНК-секвенирование.

Создание матрицы чипов, состоящей из тысяч указанных наноколодцев в слое алюминия, нанесенном на кварцевую подложку, базировалось на использовании высокоразрешающей электронной литографии. Данный метод при высокой разрешающей способности (до 5 нм) позволяет быстро менять в широких пределах дизайн (топографию) создаваемых структур, обладает большой вариативностью и относительно дешев при мелкосерийном производстве матриц небольшой площади. Однако создание матриц большой площади требует существенных временных затрат и удорожания. В качестве примера можно

указать, что при отработанной нами технологии нанесения резиста для экспонирования площади размером 400×400 мкм, состоящей из 9 чипов, содержащих матрицу из 25×25 наноколодцев в каждом чипе, требуется около двух часов, включая проявление резиста и юстировку электронной системы.

Для разработки технологии создания матрицы наноколодцев реакторов необходимо было решить две главные проблемы. Первая связана с необходимостью разработки метода создания маски с размером элементов 100 нм и менее на диэлектрической подложке, которая под действием электронного пучка подвергается зарядке. Важнейшей задачей при решении этой проблемы является планаризация резиста на поверхности подложки. Вторая проблема связана с перенесением рисунка маски на поверхность толстой (100 нм) металлической пленки, с сохранением всех структурных элементов маски. При такой толщине поликристаллической металлической пленки и диаметре канала 100 нм и менее (аспектное соотношение больше единицы) создание наноколодцев в металлической пленке является достаточно сложной проблемой.

Основным способом создания металлических наноструктур является метод «взрыва» (Lift-OFF-литография) [11], поэтому следующим шагом было решение вопросов по созданию масок из резистов, подходящих для взрывной литографии. В этом случае создаваемая путем электронной засветки и последующего химического травления резиста маска состоит из отдельных столбиков резиста, на которые затем напыляется слой металла необходимой толщины. Последующее удаление столбиков резиста из пленки алюминия методом «взрыва» резиста и частей металла над столбиками приводит к созданию наноколодцев в данной пленке. Следует также отметить, что отдельной большой проблемой при электронной засветке резиста на кварце, диэлектрическом материале, является зарядка резиста, которая влияет на точность позиционирования электронного луча и, соответственно, на точность изготовления маски резиста и наноколодцев. Указанные проблемы решались последовательно в процессе разработки технологии создания наноколодцев матрицы.

Исследования проводили на электронном литографе Raith 150, который представляет собой модернизированный сканирующий электронный микроскоп (СЭМ) фирмы Zeiss. Он оснащен большой вакуумной камерой, позволяющей загружать образцы размером до 150 мм.

Внутри камеры смонтирован роботизированный лазерный стол, позволяющий позиционировать образец размером до 150 мм с точностью до 20 нм.

Было изготовлено несколько прототипов чипов, один из которых показан на рис. 5.

Рис. 5. Изображение матрицы наноколодцев и маркеров внутри поля одного чипа

Для прецизионного позиционирования прототипа чипа относительно матрицы освещающих пучков в оптической установке были сделаны отверстия диаметром 800 нм, расположенные на диагоналях матрицы, и дополнительные треугольные маркеры.

При изготовлении матрицы наноколодцев возникали трудности, связанные со смещением рядов наноколодцев относительно их идеального эквидистантного положения, что приводило к сравнительно большому проценту брака. Предположительно причиной дислокаций была неравномерность стекания заряда в алюминиевом слое.

4. Изображение чипа наноколодцев

Чип наноколодцев располагался на столе с четырехкоординатной x - y - z - ϕ -подвижкой. На рис. 6, *а* представлен снимок на просвет без призмы и фильтрации в регистрирующем канале; на рис. 6, *б* – на отражение в белом свете; на рис. 6, *в* чип освещался дополнительно тремя лазерами. Таким образом, совмещение матриц лучей и наноколодцев можно было осуществлять в режиме визуализации на видеоматрице.

Яркость лазерных пятен была неравномерной: отличие интенсивности в центре и с края достигало 50 %, что объясняется неидеальностью используемых на данный момент ДОЭ. Кроме того, на равномерность могло оказывать влияние совмещение отверстий матрицы диафрагм и матрицы лучей, а также качество отверстий и эквидистантность их положения в матрице диафрагм. Усовершенствование характеристик ДОЭ мультипликатора является предметом дальнейших исследований в рамках темы данного проекта.

Рис. 6. Изображение чипа нанокколодцев: *а* – без призмы в белом свете на просвет; *б* – на отражение; *в* – на отражение с совмещенными лазерными пучками трех длин волн; *г* – с призмой на просвет в белом свете; *д* – увеличенная область

На рис. 6, *г*, *д* показано изображение на цветной фотоматрице фотоаппарата Canon 450D, проходящего через чип белого света, разложенного на спектр призмой. Каждый флуорофор имеет свою уникальную форму спектра, по которому его можно идентифицировать с помощью нескольких пикселей монохромной матрицы с большой долей вероятности.

5. Выбор приемника излучения

Приемник излучения также является ключевым элементом оптической системы, так как он должен быть способен детектировать сверхмалый сигнал за короткий промежуток времени. Из анализа публикаций, посвященных устройству-прототипу [3, 10], были определены оценочные требования к фотоприемному элементу:

1. Тип фотоприемника – матрица элементов с разрешением, определяемым конфигурацией применяемой подложки: в прототипе на подложке расположены 3069 наноколодцев, организованных в матрицу 93×33 элемента. В устройстве-прототипе используется глубоко охлаждаемая ПЗС-матрица с электронным умножением (EMCCD) с разрешением 512×512 пикселей с размером элементарного приемника 16×16 мкм² и геометрическими размерами фоточувствительной области около 8×8 мм².

2. Требуемый спектральный диапазон фотоприемного элемента при использовании флуорофоров ряда Alexa 488, 532, 555, 568, 647 или Cy 3,5 находится в диапазоне 490–730 нм.

3. Детектируемые события являются асинхронными, частота их возникновения оценивается как единицы в секунду [3]. Оценочное значение длительности световых импульсов – 10–100 мс. Учитывая асинхронность возникновения событий и малый уровень сигнала, фотоприемник должен иметь, возможно, меньшие временные потери на считывание накопленного сигнала, и, так как информативным параметром является не форма импульса, а его наличие, требуемая частота считывания кадра может быть установлена в диапазоне 30–100 кадров в секунду.

4. Оценка требуемой чувствительности фотоприемной системы может быть выполнена на основании данных, приведенных в работах [3, 10]. В прототипе используются три линии возбуждения (488, 532, 643 нм) и микрообъектив с $NA = 0,9$. Диаметр наноколодца – 70 нм. Плотность мощности на образце – около 5 мкВт/мкм² в каждой линии.

Сделаем оценку уровня полезного сигнала на примере флуорофора Alexa 555 на основании данных, приведенных в работах [3, 12], и общих соображений.

Для Alexa 555 максимум характеристики возбуждения расположен на длине волны 555 нм, максимум характеристики испускания –

565 нм. Коэффициент экстинкции ($150\,000 \text{ моль}^{-1} \cdot \text{см}^{-1}$) пересчитаем в эффективное оптическое сечение (molecular cross-section): $\sigma = 1000 (\text{см}^3/\text{л}) \cdot 150\,000 (\text{моль}^{-1} \cdot \text{см}^{-1}) \cdot \ln(10)/6,023 \cdot 10^{23} (\text{молекул} \times \text{моль}^{-1}) = 5,7 \cdot 10^{-16} (\text{см}^2/\text{молекулу})$.

Квантовая эффективность флуоресценции (оценочно) составляет $Q_e = 0,2$. Энергия фотона на длине волны возбуждения вычисляется по формуле $E_{ph} = h \cdot \nu = h \cdot c/\lambda$, где h – постоянная Планка, λ – длина волны возбуждения; ν – частота возбуждающего света. $E_{ph} = 6,65 \cdot 10^{-34} (\text{Дж} \cdot \text{с}) \times 3 \cdot 10^8 (\text{м/с})/550 \cdot 10^{-9} (\text{м}) = 3,63 \cdot 10^{-19} (\text{Дж})$.

Плотность мощности возбуждающего света на входе наноклода составит $P_{ex} = 1,38 \cdot 10^{13} \text{ фотонов} \cdot \text{с}^{-1} \cdot \text{мкм}^{-2}$. Интенсивность сигнала флуоресценции равна $I = P_{ex} \cdot \sigma \cdot Q_e = 1,38 \cdot 10^{13} \cdot 10^8 (\text{фотонов} \cdot \text{с}^{-1} \cdot \text{см}^{-2}) \times 5,7 \cdot 10^{-16} (\text{см}^2/\text{молекулу}) \cdot 0,2 = 1,57 \cdot 10^5 \text{ фотонов} \cdot \text{с}^{-1}$.

Принимая во внимание, что эффективность возбуждения меньше единицы (0,8) из-за несогласованности длины волны возбуждения с максимумом спектральной характеристики возбуждения флуорофора и значение эффективности сбора флуоресцентного света, равное 0,15 и обусловленное конечной апертурой объектива и потерями в оптической схеме, получаем интенсивность сигнала $I = 1,88 \cdot 10^4 \text{ фотонов} \cdot \text{с}^{-1}$ на входе фотоприемника. Данная оценка уровня сигнала достаточно груба, но по порядку величины совпадает с приведенными в работе [10] экспериментальными данными $I = 0,5 - 1 \cdot 10^4 \text{ фотонов} \cdot \text{с}^{-1}$.

Следует отметить, что при обсуждаемых уровнях плотности мощности возбуждающего излучения параметры рабочего процесса находятся далеко от ограничений по уровню возбуждения как из-за ограничений, обусловленных конечным временем жизни возбужденного состояния (интенсивность света флуоресценции $\approx 2 \cdot 10^8 \text{ фотонов} \cdot \text{с}^{-1}$), так и из-за эффекта фотодеградации флуорофора (10^4 – 10^5 квантов излучения для единичной молекулы флуорофора). Учитывая принятую оценку длительности световых импульсов сигнала 10–100 мс, регистрируемые уровни сигнала составляют 10^2 – 10^3 фотонов. Поскольку для определения спектрального состава сигнала используется распределение принятого света на несколько элементарных приемников (пикселей матрицы), для отдельного элементарного приемника значение интенсивности сигнала еще меньше.

Возможные варианты матричных фотоприемных устройств – традиционные неохлаждаемые ПЗС-матрицы, охлаждаемые ПЗС-матри-

цы, ПЗС-матрицы с электронным умножением (EMCCD), системы с использованием микроканальных пластин (например, сборка типа «усилитель яркости на микроканальной пластине – ПЗС-матрица»).

Для современных неохлаждаемых ПЗС-матриц требуемого разрешения (S10811, S9978 производства HAMAMATSU) характерные значения темнового тока составляют величины порядка 10^3 электрон $\cdot$ с $^{-1}$ $\times$ $\times$ пиксель $^{-1}$, что неприемлемо для данной задачи. Лучшие по шумовым характеристикам приборы (S10140-1009 HAMAMATSU) дают значения темнового тока 50–100 электрон $\cdot$ с $^{-1}$ $\cdot$ пиксель $^{-1}$ при температуре 0 °С. При использовании охлаждаемых ПЗС-приборов темновой ток может иметь предельно малые значения – доли или единицы электронов в секунду на пиксель, но при этом доминирующий вклад в общие шумы прибора начинает давать шум считывания, который существенно растет при увеличении быстродействия прибора. Типичные значения шума считывания 20–100 электронов rms (среднеквадратичное значение) для частоты считывания составляют порядка десятков кГц. Этот шум быстро нарастает при повышении частоты считывания и становится неприемлемым при требуемых для данной задачи значениях частоты считывания (11 МГц для матрицы размерностью 512 $\times$ 512 элементов и частоте вывода кадров порядка 30–100 Гц). Шумовые параметры для недавно появившихся на рынке охлаждаемых EMCCD-матриц на примере камеры C9100-13 (HAMAMATSU) следующие: темновой ток – 0,01 электрон $\cdot$ с $^{-1}$ $\cdot$ пиксель $^{-1}$ при температуре –65 °С, шум считывания – 25 электронов rms при частоте 11 МГц и коэффициенте умножения 4 и 1 электрон rms при частоте 11 МГц и коэффициенте умножения 1200. Недостатком этих приборов является высокая стоимость.

Из неохлаждаемых приемников следует выделить ПЗС-камеры с усилителями яркости изображения, ЭОПами 2 и 3-го поколения на основе микроканальных пластин. Для того чтобы сравнить характеристики чувствительности EMCCD-приборов и приборов на основе микроканальных пластин, рассмотрим интегральный параметр чувствительности – эквивалентную шумовую освещенность на единицу площади. Таким образом, мы исключим из рассмотрения характеристики элементарных приемников (пикселей в ПЗС-матрице и системы наноклодцев и люминесцирующего экрана в ЭОП) и получим более общую характеристику потенциально достижимой чувствительности вне зависимости от вариантов построения системы с применением микро-

канальных пластин. Примем размерность сравниваемой величины – эквивалентной шумовой освещенности, фотон $\cdot \text{с}^{-1} \cdot \text{см}^{-2}$.

Рассмотрим пример такого прибора – V8070U (HAMAMATSU) – по данным производителя [3]. Для модификации с максимальной квантовой эффективностью в требуемом спектральном диапазоне $Q_e \approx 50\%$ EBI (Equivalent Background Input) $\approx 8 \cdot 10^{-15}$ Вт $\cdot \text{см}^{-2}$, что соответствует значению EBI MCPs = $2,3 \cdot 10^4$ фотон $\cdot \text{с}^{-1} \cdot \text{см}^{-2}$ в пике характеристики спектральной чувствительности (500 нм). Отметим также, что в длинноволновой части спектра спектральная характеристика квантовой эффективности имеет значительное снижение, $Q_e \approx 20\%$ (750 нм).

Параметры для оценки чувствительности EMCCD C9100-13 (HAMAMATSU) следующие: темновой ток – 0,001 электрон $\cdot \text{с}^{-1} \times \times \text{пиксель}^{-1}$ при температуре -85°C , шум считывания – <1 электрона rms при частоте 11 МГц и коэффициенте умножения 1200. Квантовая эффективность равна 95 % в пике спектральной чувствительности и 85 % на краях требуемого спектрального диапазона. Из приведенных данных следует: $\text{EBIEMCCD} = 512^2 \cdot 0,001$ (электрон $\cdot \text{с}^{-1} \cdot \text{пиксель}^{-1}$) $\times \times 1,1/0,8^2$ (см^2) = $4,51 \cdot 10^2$ фотон $\cdot \text{с}^{-1} \cdot \text{см}^{-2}$.

Согласно указанным оценкам чувствительности матричных фотоприемников оптимальным представляется выбор фотоприемника, выполненного по EMCCD-технологии.

6. Синтез новых флуоресцентных субстратов

Для расширения номенклатуры и оптических характеристик используемых аналогов дезоксирибонуклеозидтрифосфатов, в которых в качестве флуоресцентной группы выступают производные веществ из ряда триарилметановых (ксантеновых), цианиновых и нитробензофурановых красителей: флуоресцеин, родамин Б, родамин 101, цианиновый краситель нео-Су5, 7-нитробензофуразан (рис. 7), нами был предпринят синтез таких соединений. Полученные результаты изложены нами в патенте¹.

¹ Аналоги природных дезоксирибонуклеозидтрифосфатов и рибонуклеозидтрифосфатов, содержащие репортерные флуоресцентные группы, для использования в аналитической биоорганической химии: пат. № 2582198 Рос. Федерация / В.В. Анненков, Д.В. Пышный, Е.Н. Даниловцева, П.Е. Воробьев, С.Н. Зелинский, О.Н. Верховзина, М.А. Грачев. Заявл. 30.03.2016.

Нами была исследована эффективность включения нуклеотидов в растущую цепь ДНК Taq полимеразой при использовании модифицированных флуорофорами по терминальной фосфатной группе три- и тетрафосфатов нуклеозидов. Наибольшая эффективность полимеризации наблюдается при использовании следующих модифицированных нуклеозидтрифосфатов: dATP-RhodB, dATP-Cy5, dATP-NBD2, TTP-RhodB, TTP-Cy5, TTP-Flu, dCTP-RhodB, dCTP-Rhod101, dGTP-RhodB. Показана возможность эффективного удлинения праймера до полно-размерного продукта полимеризации 7250 пн при использовании четырех перечисленных модифицированных нуклеозидтрифосфатов: dCTP-Rhod101, dATP-Cy5, TTP-Flu, dGTP-RhodB.

Рис. 7. Флуоресцентные группы модифицированных дезоксинуклеозидтрифосфатов

Заключение

В настоящей статье мы предприняли попытку описать состояние дел, касающихся разработки отечественной платформы оптического секвенирования ДНК по идеологии компании Pacific Biosciences. Были представлены новые технические решения, предложенные нами в ходе разработки.

Для приобретения практического опыта использования системы оптического секвенирования была проведена обработка «сырых» данных секвенирования полного генома байкальской диатомеи *Synedra acus*, полученных на нескольких известных платформах на коммерческой основе. Преимуществом платформы PacBio RSII является то, что она дает наиболее длинные прочтения протяженностью по крайней мере более 10 тыс. н.о. Ее недостаток – относительно высокая вероятность ошибок.

Нами была разработана оптическая схема ДНК-секвенатора с многоканальным осветителем на основе дифракционного мультипликатора лазерных лучей, полученных с помощью как стандартных полупроводниковых лазеров, так и с помощью волоконного иттербиевого лазера с удвоением частоты, который отличается возможностью перестройки длины волны излучения в широком диапазоне (530–560 нм) [14], и конфокальной схемы параллельного приема излучения от системы наноклодцев, параллельно освещаемых лазерными лучами. Были проведены эксперименты по инициированию и детектированию флуоресценции с применением выбранных лазеров и флуорофоров. С помощью разработанных наноканальных матриц были разработаны методы настройки матрицы наноклодцев на лазерные пучки, проведены эксперименты с растворами флуорофоров по инициированию и детектированию флуоресценции с разложением на спектр. Также была спроектирована экспериментальная система для увеличенного числа каналов считывания.

Разработка и изготовление отдельных узлов многоканальной системы освещения на основе оригинального дифракционного элемента, совмещающего функции мультипликации лазерных лучей и их фокусировки, создание методов прецизионной юстировки и автоматизированного управления положением оптических и оптико-механических элементов с точностью 100 нм и высокой долговременной стабильностью, исследование предельных возможностей детектирования светового излучения с помощью новейших ЕМССД-камер имеют самостоятельное значение для проведения научно-исследовательских и опытно-конструкторских работ. В ходе разработки были получены новые флуоресцентные аналоги дезоксирибонуклеозидтрифосфатов. Опыт взаимодействия членов межинститутского коллектива Института автоматики и электрометрии СО РАН, Института физики полупроводников СО РАН, Лимнологического института СО РАН и Института хи-

мической биологии и фундаментальной медицины СО РАН в рамках указанного проекта показал техническую возможность создания современных приборов для секвенирования ДНК в России.

Впервые была разработана технология получения матрицы наноколодцев реакторов в толстой поликристаллической пленке алюминия на кварцевой подложке с помощью высокоразрешающей электронной литографии. Был оптимизирован метод подготовки поверхности кварцевых подложек для нанесения негативного электронного резиста *maN-2403*, обеспечивающего механическую устойчивость элементов маски с размерами менее 50 нм. Были определены параметры электронной литографии для получения на негативном резисте *maN-2400* маски с размером элемента менее 50 нм. Был оптимизирован процесс нанесения пленки *Al* для взрывной литографии на резисте *maN-2403*. Однако крупная зернистость толстой металлической пленки (50–70 нм) даже после оптимизации процесса нанесения алюминия затрудняет получение наноколодцев с размером менее 70 нм. Были получены и апробированы наноканальные матрицы с диаметром каналов 70–100 нм различного дизайна в пленках алюминия толщиной 100 нм на кварце. В качестве оценки возможности для мелкосерийного производства матриц отметим, что при отработанной нами технологии нанесения резиста для экспонирования площади размером 400×400 мкм, состоящей из 9 чипов, содержащих матрицу из 25×25 наноколодцев в каждом чипе, требуется около двух часов, включая проявление резиста и его подчистку, что с учетом последующих операций по нанесению металлической пленки, проведения взрыва, при одновременной загрузке трех подложек позволяет производить не более шести платформ в день.

Эксперименты показали, что для успешного продолжения работ необходимы совершенствование и упрощение технологии получения чипа. Нужно также провести дополнительные диагностические исследования возможности вхождения фрагментов ДНК в наноколодцы, их плотности и распределения. Кроме того, требуется полностью автоматизировать подачу растворов на чип при запуске реакции, поскольку чувствительность конфокальных схем к перемещению чипа по координате *Z* (что неизбежно возникает в настоящей конструкции чипа) составляет 150–200 нм и еще меньше по координатам *XY*. Таким образом, по-видимому, необходима не только автоматическая настройка совмещения лазерных лучей и конфокальной оптики на чип перед про-

ведением секвенирования, но и соответствующая автоподстройка во время данного процесса.

Представленный подход по созданию нового секвенатора характеризуется довольно высокой стоимостью и сложностью, однако такая ситуация стандартна при создании макетов устройств массового параллельного секвенирования. Важно, что значительную часть стоимости оборудования, реагентов и расходных материалов для NGS составляют затраты на НИР и ОКР, которые несет производитель при разработке и улучшении платформы. На данный момент среди платформ TGS безусловным лидером является технология PacBio [15].

В настоящее время оформляются патенты на новые варианты опико-электронных модулей для одномолекулярного секвенирования и методы быстрой автоматической настройки приборов, упрощающие и удешевляющие конструкцию секвенаторов. Разрабатываемая аналитическая платформа может быть использована в качестве инструмента исследования широкого класса биомолекулярных реакций и процессов. Создаются также новые флуоресцентные аналоги десоксирибонуклеозидтрифосфатов.

Работа выполнена по проекту комплексной программы фундаментальных исследований СО РАН № II.2П/II.10-8.

Список литературы

1. Sanger F., Coulson A.R. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase // J. Mol. Biol. – 1975. – Vol. 94. – P. 444–448.
2. Goodwin S., McPherson J.D., McCombie W.R. Coming of age: ten years of next-generation sequencing technologies // Nature Reviews Genetics. – 2016. – Vol. 17, № 6. – P. 333–351.
3. Real-time DNA sequencing from single polymerase molecules / J. Eid, A. Fehr, J. Gray, K. Luong, J. Lyle, G. Otto [et al.] // Science. – 2009. – Vol. 323, iss. 5910. – P. 133–138. DOI: 10.1126/science.1162986
4. Eisenstein M. Oxford nanopore announcement sets sequencing sector abuzz // Nature Biotechnology. – 2012. – Vol. 30. – P. 295–296. DOI: 10.1038/nbt0412-295
5. Rhoads A., Au K.F. PacBio sequencing and its applications // Genomics, Proteomics & Bioinformatics. – 2015. – Vol. 13, iss. 5. – P. 278–289.

6. Sequencing of the complete genome of an araphidpennate diatom *Synedraacus* subsp. *radians* from Lake Baikal / Y.P. Galachyants, Y.R. Zakharova, D.P. Petrova [et al.] // *Dokl. Biochem. Biophys.* – 2015. – Vol. 461. – P. 84. DOI: 10.1134/S1607672915020064

7. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation / S. Koren, B.P. Walenz, K. Berlin, J.R. Miller, A.M. Phillippy // *BioRxiv.* – 2016. DOI: <http://dx.doi.org/10.1101/071282>

8. Hybrid assembly of the large and highly repetitive genome of *aegilops tauschii*, a progenitor of bread wheat, with the mega-reads algorithm / A.V. Zimin, D. Puiu, M.-C. Luo, T. Zhu, S. Koren, J.A. Yorke, J. Dvorak, S. Salzberg // *BioRxiv.* – 2016. DOI: <http://dx.doi.org/10.1101/066100>

9. Иркутский суперкомпьютерный центр СО РАН [Электронный ресурс]. – URL: <http://hpc.icc.ru> (дата обращения: 17.10.2016).

10. Parallel confocal detection of single molecules in real time / P.M. Lundquist, C.Z. Zhong, P. Zhao, A.B. Tomaney, P.S. Peluso, J. Dixon, B. Bettman, Y. Lacroix, D.P. Kwo, E. McCullough, M. Maxham, K. Hester, P. McNitt, D.M. Grey, C. Henriquez, M. Foquet, S.W. Turner, D. Zaccarin // *Optics Lett.* – 2008. – № 33.

11. Bouchiat V., Esteve D. Lift-off lithography using an atomic force microscope // *Appl. Phys. Lett.* – 1996. – № 69. – P. 3098. DOI: 10.1063/1.117317

12. Pawley J.B. Handbook of biological confocal microscopy. – 2nd ed., ch. 16. – Springer US, 2006. – 985 p.

13. Quantitative comparison of long-wavelength Alexa Fluor dyes to Cy Dyes: fluorescence of the dyes and their bioconjugates / J.E. Berlier [et al.] // *J. Histochemistry & Cytochemistry.* – 2003. – № 51. – P. 1699–1712.

14. Акулов В.А., Каблуков С.И., Бабин С.А. Удвоение частоты излучения перестраиваемого иттербиевого волоконного лазера в кристаллах КТР с синхронизмом в плоскостях *XY* и *YZ* // *Квантовая электроника.* – 2012. – Т. 42, № 2. – С. 120–124.

15. All Seg. The sequencing marketplace. – URL: <http://allseq.com/kb-category/sequencing-platforms> (дата обращения: 10.11.2016).

References

1. Sanger F., Coulson A.R. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology*, 1975, vol. 94, pp. 444-448.
2. Goodwin S., McPherson J.D., McCombie W.R. Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 2016, vol. 17, no. 6, pp. 333-351.
3. Eid J., Fehr A., Gray J., Luong K., Lyle J., Otto G. [et al.]. Real-time DNA sequencing from single polymerase molecules. *Science*, 2009, vol. 323, iss. 5910, pp. 133-138. DOI: 10.1126/science.1162986
4. Eisenstein M. Oxford nanopore announcement sets sequencing sector abuzz. *Nature Biotechnology*, 2012, vol. 30, pp. 295-296. DOI: 10.1038/nbt0412-295
5. Rhoads A., Au K.F. PacBio Sequencing and its applications. *Genomics, Proteomics & Bioinformatics*, 2015, vol. 13, iss. 5, pp. 278-289.
6. Galachyants Y.P., Zakharova Y.R., Petrova D.P. [et al.]. Sequencing of the complete genome of an araphidpennate diatom *Synedraacus* subsp. *radians* from Lake Baikal. *Doklady Biochemistry and Biophysics*, 2015, vol. 461, p. 84. DOI: 10.1134/S1607672915020064
7. Koren S., Walenz B.P., Berlin K., Miller J.R., Phillippy A.M. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *bioRxiv*, 2016. DOI: <http://dx.doi.org/10.1101/071282>
8. Zimin A.V., Puiu D., Luo M.-C., Zhu T., Koren S., Yorke J.A., Dvorak J., Salzberg S. Hybrid assembly of the large and highly repetitive genome of *aegilops tauschii*, a progenitor of bread wheat, with the mega-reads algorithm. *bioRxiv*, 2016. DOI: <http://dx.doi.org/10.1101/066100>
9. Irkutskii superkomp'iuternyi tsentr Sibirskogo otdeleniia Rossiiskoi akademii nauk [Irkutsk Supercomputer Center of the Siberian Branch of the Russian Academy of Sciences], available at: <http://hpc.icc.ru> (accessed 17 October 2016).
10. Lundquist P.M., Zhong C.Z., Zhao P., Tomaney A.B., Peluso P.S., Dixon J., Bettman B., Lacroix Y., Kwo D.P., McCullough E., Maxham M., Hester K., McNitt P., Grey D.M., Henriquez C., Foquet M., Turner S.W., Zaccarin D. Parallel confocal detection of single molecules in real time. *Optics Letters*, 2008, no. 33.

11. Bouchiat V., Esteve D. Lift-off lithography using an atomic force microscope. *Applied Physics Letters*, 1996, no. 69, p. 3098. DOI: 10.1063/1.117317

12. Pawley J.B. Handbook of biological confocal microscopy. Springer US, 2006. 985 p.

13. Berlier J.E. [et al.]. Quantitative comparison of long-wavelength Alexa Fluor dyes to Cy Dyes: fluorescence of the dyes and their bioconjugates. *Journal of Histochemistry & Cytochemistry*, 2003, no. 51, pp. 1699-1712.

14. Akulov V.A., Kablukov S.I., Babin S.A. Udvoenie chastoty izlucheniia perestraivaemogo itterbievogo volokonnoho lazera v kristallakh KTP s sinkhronizmom v ploskostiakh XY i YZ [Doubling the frequency of the radiation of a tunable fiber laser in KTP crystals with synchronism in the XY plane and YZ]. *Kvantovaya elektronika*, 2012, vol. 42, no. 2, pp. 120-124.

15. AllSeq. The sequencing marketplace, available at: <http://allseq.com/kb-category/sequencing-platforms> (accessed 10 November 2016).

Получено 06.12.2016