



КРАТКАЯ ИСТОРИЯ МЕТОДОВ ВЫСОКОПРОИЗВОДИТЕЛЬНОГО СЕКВЕНИРОВАНИЯ НУКЛЕИНОВЫХ КИСЛОТ

Зубов В.В.^{1*}, Чемерис Д.А.², Василов Р.Г.³, Курочкин В.Е.⁴, Алексеев Я.И.^{4,5}

¹Институт теоретической и экспериментальной биофизики Российской академии наук,
Россия, Пущино, 142290, Институтская ул., 3

²ООО «Максим Медикал», Россия, 123423, Москва, ул. Народного Ополчения, д. 34, стр. 1

³НИЦ «Курчатовский институт», Россия, 123182, Москва, пл. Академика Курчатова, д. 1

⁴Институт аналитического приборостроения Российской академии наук,
Россия, 198095, Санкт-Петербург, ул. Ивана Черных, д. 31/33

⁵ООО «Синтол», Россия, 127550, Москва, ул. Тимирязевская, д. 42

*E-mail: genseq@mail.ru

Резюме

Кратко рассмотрены процессы, происходящие при ферментативном росте цепи ДНК, в виде удлинения этой молекулы, выделения пирофосфата, протона, тепловой энергии и увеличении электрического импеданса, находящих применение в разных методах высокопроизводительного секвенирования ДНК путем синтеза. При этом детекция роста цепи ДНК контролируется с помощью высоковольтного гель-электрофореза и имеет ограниченную масштабируемость. Что касается упомянутых выше прочих побочных продуктов полимеризации цепи ДНК, то их детекция может быть легко масштабируема, что привело к появлению методов полногеномного секвенирования ДНК новых поколений, получивших широко используемую аббревиатуру NGS – Next Generation Sequencing. Однако отнесение того или иного нового метода секвенирования к какому-либо конкретному поколению подчас вызывает затруднения в связи с тем, что используемый в нем принцип зародился раньше, чем был воплощен другой, оказавшийся в итоге менее производительным. Следует разграничивать методы нового секвенирования ДНК на две группы, в которых происходит массовое мультимолекулярное секвенирование одинаковых матриц, или имеет место секвенирование единичных молекул ДНК, получившее обозначение мономолекулярное секвенирование. В данном обзоре, наряду с классическим методом секвенирования ДНК по Сэнгеру, являющимся до сих пор «золотым стандартом», рассмотрены из первой группы пиросеквенирование, полупроводниковое секвенирование, термосеквенирование, электронное секвенирование, флуоресцентные мостиковое секвенирование и секвенирование с помощью наночастиц, а также мономолекулярные методы – tSMS секвенирование, SMRT секвенирование и нанопоровое секвенирование. Уделено внимание стоимостным вопросам секвенирования ДНК и перспективам его развития.

Ключевые слова: секвенирование ДНК, высокопроизводительное секвенирование, пиросеквенирование, полупроводниковое секвенирование, термосеквенирование, электронное секвенирование, флуоресцентные мостиковое секвенирование, флуоресцентное секвенирование наночастицами, мультимолекулярное секвенирование, мономолекулярное секвенирование, нанопоровое секвенирование, NGS

Цитирование: Зубов В.В., Чемерис Д.А., Василов Р.Г., Курочкин В.Е., Алексеев Я.И. Краткая история методов высокопроизводительного секвенирования нуклеиновых кислот // *Biomics*. 2021. Т.13(1). С. 27- 46.

DOI: 10.31301/2221-6197.bmcs.2021-4

© Авторы

BRIEF HISTORY OF HIGH-THROUGHPUT NUCLEIC ACID SEQUENCING METHODS

Zubov V.V.^{1*}, Chemeris D.A.², Vasilov R.G.³, Kurochkin V.E.⁴, Alekseev Ya.I.^{4,5}

¹Institute of Theoretical and Experimental Biophysics RAS, 3 Institutskaya str., Pushchino-na-Oke, 142290, Russia

²Maxim Medical LLC, 34-1 Narodnogo Opolcheniya str., Moscow, 123423, Russia

³National Research Center «Kurchatov Institute», Kurchatov Square 1, Moscow 123182, Russia

⁴Institute for Analytical Instrumentation RAS, 31/33 Ivana Chernyh str., 198095, St. Petersburg, Russia

⁵Syntol Ltd, 42 Timiryazevskaya str., 127550, Moscow, Russia

Resume

The processes occurring during the enzymatic growth of the DNA chain in the form of elongation of the molecules, the release of pyrophosphate, proton, thermal energy, and an increase in electrical impedance, which are used in various methods of high-throughput DNA sequencing by synthesis, are briefly considered. The detection of DNA chain growth is controlled by high-voltage gel electrophoresis and has limited scalability. As for mentioned above other by-products of DNA chain polymerization, their detection can be easily scalable, which has led to the emergence of methods for whole genome sequencing of new generations of DNA, which have received the widely used abbreviation NGS - Next Generation Sequencing. However, the attribution of any new sequencing method to a particular generation is sometimes difficult due to the fact that the principle used in it was born earlier than the other one was implemented, which turned out to be less productive in the end. In addition, it is more important to distinguish the methods of new DNA sequencing into two groups in which the massive parallel sequencing of identical matrices takes place or the sequencing of single DNA molecules takes place and last one have received the designation monomolecular sequencing. In this review, along with the classical Sanger method of DNA sequencing, which is still the "gold standard", pyrosequencing, semiconductor sequencing, thermosequencing, electronic sequencing, fluorescent bridge sequencing and sequencing using nanoballs from the first group, as well as monomolecular methods – tSMS sequencing, SMRT sequencing and nanopore sequencing are considered. Attention is paid to the costs of DNA sequencing and the prospects for its development.

Keywords: DNA sequencing, high-throughput sequencing, pyrosequencing, semiconductor sequencing, thermosequencing, electronic sequencing, fluorescent bridge sequencing, nanoball fluorescence sequencing, massive parallel sequencing, monomolecular sequencing, nanopore sequencing, NGS

Citation: Zubov V.V., Chemeris D.A., Vasilov R.G., Kurochkin V.E., Alekseev Ya.I. Brief history of high-throughput nucleic acid sequencing methods. *Biomcs*. 2021. V.13(1). P. 27-46. DOI: 10.31301/2221-6197.bmcs.2021-4

© The Authors

Введение

Важность определения нуклеотидных последовательностей не вызывала сомнений еще до открытия в 1953 году двойной спирали ДНК [Watson, Crick, 1953]. На то было несколько причин. Например, уже в конце 1930 годов стала очевидной несостоятельность выдвинутой Ф.А. Левиным в начале XX-го века тетрадной теории организации тимонуклеиновой кислоты (ДНК) [Astbury, Bell, 1938; Signer et al., 1938]. Кроме того, в середине 1940-х годов было установлено, что ДНК – молекула наследственности [Avery et al., 1944]. Но подступиться к расшифровке последовательности

нуклеотидов, даже после установления общих принципов её структурной организации, тогда не представлялось возможным.

К середине 1960-х годов определить нуклеотидную последовательность можно было только для коротких олигонуклеотидов, да и то с большим трудом [Василенко и др., 1965]. В 1968 году известный учёный Э. Чаргафф, заметно поспособствовавший разгадке организации ДНК в виде двойной спирали, высказал мысль, что «... чтение последовательности ДНК может стать задачей XXI века ...» [Chargaff, 1968]. Конечно, можно считать, что он сильно ошибся, поскольку меньше чем

через 10 лет после этих его слов появились публикации с описанием быстрых и эффективных методов секвенирования ДНК путем химической дегградации и терминирования роста цепи ДНК [Sanger, Coulson, 1975; Maxam, Gilbert, 1977; Sanger et al., 1977]. Однако все же следует считать, что Чаргафф оказался прав. Именно наступившее столетие дало в руки экспериментаторов небывалые возможности по секвенированию за короткий срок полных геномов практически любых организмов всех уровней генетической сложности и весьма крупного размера, достигающего двух десятков млрд.п.н. Но начинать рассмотрение методов секвенирования ДНК новых поколений лучше всего с ферментативного метода секвенирования ДНК, построенного на использовании терминирующих дидезоксинуклеозидтрифосфатов (ддНТФ) [Sanger et al., 1977]. Обычно этот подход называют секвенированием по Сэнгеру.

Удивительно, но на протяжении четверти века, с момента появления в 1977 году секвенирования ДНК по Сэнгеру, велась лишь очень немногочисленные разработки принципиально иных методов определения последовательности нуклеотидов. Практически все работы были нацелены на улучшение электрофоретических методов секвенирования ДНК, тогда как пиросеквенирование представлялось мало перспективным по причине низкой производительности и некоторых ограничений, из-за которых оно весьма сильно уступало методу Сэнгера. Хотя именно пиросеквенирование стало первой из технологий секвенирования ДНК нового поколения, обозначаемых сейчас аббревиатурой NGS (Next Generation Sequencing).

Весьма показательна датированная 2008-м годом фраза известного специалиста в области секвенирования ДНК J.Shendure (США). Он отметил, что все исследователи увлеклись совершенствованием метода секвенирования ДНК по Сэнгеру и просто забыли, что могут быть и какие-то иные способы определять последовательность нуклеотидов в ДНК, которые необходимо активно разрабатывать [Shendure, 2008]. Данная тенденция прослеживается и в отечественной монографии «Секвенирование ДНК» [Чемерис и др., 1999], в которой наибольшее внимание уделено методу секвенирования ДНК по Сэнгеру, а альтернативные методы упомянуты относительно кратко - пропорционально их доле в общем потоке методологических работ тех лет по секвенированию ДНК.

Однако после завершения проекта «Геном человека» в умах исследователей словно произошел прорыв, и началась активная разработка

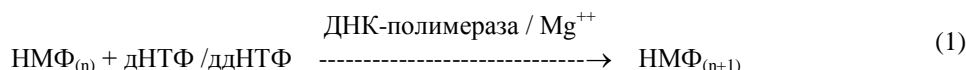
многочисленных новых методов секвенирования ДНК. Даже не очень производительных, но позволяющих за счет массивного параллелизма обогнать классический метод секвенирования ДНК с дидезокситерминаторами. При этом практически единственной страной, оценившей перспективность разработки новых методов секвенирования ДНК и то, какие они сулят перспективы в будущем, в тот момент оказались США, до сих пор удерживающие мировое лидерство в этих технологиях, но в последние годы их стали теснить Китай и Великобритания.

В настоящее время методов секвенирования ДНК стало так много, что их невозможно рассмотреть даже в очень большом обзоре. Тем более что многие такие методы оказались неконкурентоспособными, и их разработки были прекращены. Поэтому даже в монографии Д.В.Ребрикова и соавт. [2014], посвященной секвенированию ДНК методами новых поколений, рассмотрены только основные технологии геномного секвенирования.

Все методы секвенирования нуклеиновых кислот можно разделить на две большие группы – мультимолекулярные и мономолекулярные. В первых используется массовый параллелизм, благодаря которому секвенируются одновременно клоны одинаковых участков ДНК, насчитывающие от нескольких тысяч до нескольких миллионов идентичных молекул. Вторые имеют дело с определением последовательности оснований в каждой отдельной взятой молекуле (фрагменте) ДНК или РНК. При этом и в первой, и во второй группе есть подходы, основанные на секвенировании ДНК посредством синтеза. С них и начнем краткое рассмотрение отдельных методов.

Секвенирование ДНК с дидезокситерминаторами по Сэнгеру

Секвенирование ДНК по Сэнгеру, называемое также дидезоксисеквенированием, основано на ферментативном синтезе комплементарной цепи ДНК по существующей одноцепочечной матрице и на происходящем в разных местах цепи ДНК запрограммированном, но в тоже время носящем случайный характер, прерывании ее дальнейшего роста. Терминация роста цепи ДНК происходит из-за того, что, наряду с обычными дНТФ, в реакционной смеси присутствуют в меньшей концентрации их дидезокси-аналоги (ддНТФ), не имеющие ОН-группы в 3'-положении дезоксирибозы. Их включение в ДНК приводит к прерыванию её дальнейшего синтеза. Упрощённо это можно описать следующим уравнением:



В данном уравнении $\text{НМФ}_{(n)}$ обозначает цепь ДНК, состоящую из некоторого количества (n) дезоксинуклеотидов, а $\text{НМФ}_{(n+1)}$ – ее удлиненный вариант. Такое удлинение цепи ДНК происходит под действием подходящей ДНК полимеразы в присутствии солей магния. Для секвенирования ДНК этим методом требуется проведение четырех параллельных реакций в отдельных пробирках, в каждой из которых в нужной пропорции присутствует только один из терминирующих реакцию ддНТФ. Затем проводится разделение продуктов реакции по размеру с помощью высоковольтного электрофореза в полиакриламидном геле высокого разрешения, позволяющего в широком диапазоне разделять молекулы ДНК, отличающиеся между собой по длине всего на один нуклеотид. Причём первоначально использовалась радиоактивная метка, проявляемая при помощи радиоавтографии.

Все составляющие процесса секвенирования ДНК этим методом за годы, прошедшие с момента его разработки, подверглись сильной модификации. Главным новшеством было, конечно же, применение автоматических секвенаторов ДНК, рассчитанных на детекцию флуоресцентной метки и непосредственное занесение получаемых результатов в компьютер. Первое время это были секвенаторы с разделением фрагментов ДНК в пластинах геля. Несколько позже появились более производительные и более экономичные капиллярные секвенаторы, число капилляров в которых довольно быстро выросло с одного до 96. Что касается флуоресцентных меток, то они могут находиться как в самом праймере, так и включаться в новую цепь ДНК при ее построении, в том числе в составе терминирующих ддНМФ.

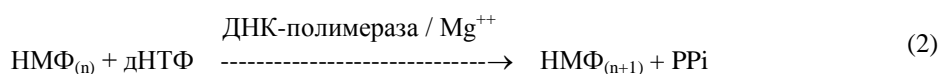
Несмотря на огромные усилия, направленные на совершенствование как непосредственно самих секвенирующих/терминирующих реакций, так и электрофоретического разделения получающихся продуктов, количество уверенно «читаемых» нуклеотидов при таком подходе почти за четверть века увеличилось не более чем на порядок – примерно со ста до тысячи нуклеотидов. Причина очень проста. Был достигнут предел разрешающих возможностей

гель-электрофореза нуклеиновых кислот, обеспечивающих детекцию различий в один нуклеотид. После чего уже никакие дополнительные усовершенствования не могли ощутимо повысить эффективность этой процедуры. Ну, разве что еще на 10–15%, тогда как повышать производительность секвенирования ДНК необходимо было не на проценты, а на порядки. Причём с одновременным значительным снижением себестоимости процесса. Поэтому исчерпавшие свой потенциал методы секвенирования, построенные на электрофоретическом разделении фрагментов ДНК, быстро уступили свои позиции секвенаторам нового поколения. Тем не менее, работающие по технологии Сэнгера капиллярные секвенаторы и сейчас весьма востребованы. Они широко применяются для решения многих рутинных задач секвенирования, для финишного завершения проектов по полногеномному секвенированию, для картирования геномов и т. п. При этом секвенирование по Сэнгеру, благодаря своей точности, до сих пор считается «золотым стандартом».

В настоящее время выпускается несколько моделей подобных приборов, одна из которых производится в России. Восьмикапиллярный секвенатор «Нанофор-5», разработанный Институтом аналитического приборостроения РАН совместно с компанией «Синтол», имеет увеличенное количество каналов детекции флуоресценции, и по своим основным рабочим характеристикам не уступает импортным аналогам.

Секвенирование ДНК с применением массового параллелизма

Пиросеквенирование. Еще одним методом секвенирования ДНК посредством синтеза, появившимся еще в «дополногеномную» эпоху, было пиросеквенирование [Human, 1988]. Для его описания уравнение (1) следует несколько усложнить, поскольку в процессе присоединения очередного нуклеотида к растущей цепи ДНК от дНТФ отщепляется пирофосфат (PPi), который как раз и детектируется.



В пиросеквенировании не используются дидезоксинуклеозидтрифосфаты, терминирующие рост цепи ДНК. Для определения того, какой

нуклеотид (и их количество) оказался включенным в растущую цепь ДНК, используется биоллюминесценция, возникающая за счет целого

каскада ферментативных реакций. В реакционную смесь, находящуюся в проточной ячейке, поочередно добавляют различные дНТФ, регистрируют сигнал и промывают ячейку перед добавлением следующего дНТФ. После присоединения очередного нуклеотида выделяющийся пирофосфат под действием АТФ-сульфуриказы превращает находящийся в реакционной смеси аденозинфосфосульфат в АТФ. Образующаяся АТФ вызывает окисление люциферина люциферазой с выделением регистрируемых секвенатором квантов света. При этом из-за перекрестной реакции люциферазы с дАТФ приходится использовать его синтетический аналог с тио-группой в альфа-положении (дАТФ α S), что заметно удорожает реагенты.

Из-за отсутствия терминации в растущую цепь ДНК может включиться сразу несколько одинаковых нуклеотидов. Количество таких нуклеотидов можно определить по уровню сигнала биолюминесценции, но при увеличении длины гомополимерных повторов точность определения их количества быстро снижается, что приводит к ошибкам секвенирования таких повторов в виде пропусков или вставок нуклеотидов. Безусловно, данный вариант пиросеквенирования по производительности значительно уступал секвенированию по Сэнгеру. Но благодаря тому, что он не зависел от гель-электрофореза, его оказалось гораздо легче масштабировать. И в 1997 году в Швеции была учреждена компания «Pyrosequencing» (с 2003 года – «Biotage AB»), занимающаяся разработкой, производством и продажей приборов, наборов реагентов и тест-систем медицинского назначения, в которых используется пиросеквенирование.

Новую жизнь пиросеквенирование обрело после организации в США компании 454 Life Science Inc. Её основатель, J.M.Rothberg, приобрёл у шведских патентообладателей эксклюзивную лицензию на использование пиросеквенирования, которую применил для параллельного анализа большого количества проб ДНК (>10 тысяч). В разработанной этой компанией технологии матричные клоны молекул ДНК (идентичные копии одного фрагмента ДНК) получают с помощью эмульсионной ПЦР на носителе, которым служат агарозные микросферы. Множество таких микросфер помещают в лунки, расположенные на дне проточной ячейки, а затем следуют те же стадии, что и при «старом» методе. Усовершенствованная таким образом технология пиросеквенирования получила название «454 Sequencing».

В 2005 году на рынке появился первый работающий по этой технологии геномный секвенатор (GS20), который был способен за один рабочий цикл

прочитать от 10 до 20 миллионов нуклеотидов. Это позволило за 4 часа¹ ресеквенировать небольшой (чуть больше 580 тысяч пар нуклеотидов) геном бактерии *Mycoplasma genitalium* [Margulies et al., 2005], на что в 1995 г. ушло 4 месяца [Fraser et al., 1995]. В мае 2007 года с помощью нового пиросеквенирования было завершено прочтение первого индивидуального генома человека, которым стал Нобелевский лауреат Джеймс Уотсон [Wheeler et al., 2008], после чего в июне того же года привлеченные к этой разработке инвесторы продали компанию 454 Life Science Inc. корпорации Roche за 120 миллионов долларов.

Позднее появились более производительные модели пиросеквенаторов (GS FLX и GS FLX⁺), имеющие большее число работоспособных лунок. Их общее количество возросло до 400 тысяч. При этом размер гелевых микросфер был уменьшен с 30 до 10 мкм (рис. 1), но продолжительность рабочего цикла увеличилась с 4 до 7,5 часов.

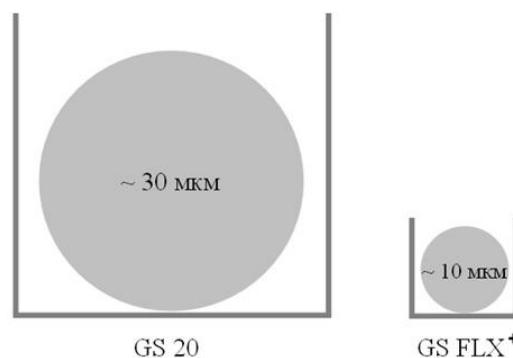


Рис. 1. Размеры гелевых микросфер из агарозы и микроячеек в пикотитраторных планшетах, используемых в пиросеквенировании (масштаб не соблюден)

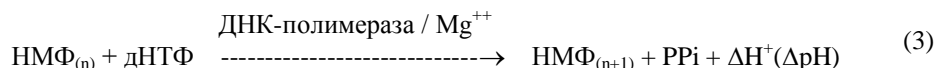
Fig. 1. Dimensions of agarose gel microspheres and microcells in picotitrator plates used in pyrosequencing (scale are not kept)

Оптимизация рабочих режимов, улучшение программного обеспечения и повышение чистоты используемых реагентов позволили увеличить длину читаемых последовательностей сначала до 200–300, а затем до 400, и даже до 800 нуклеотидов. Общая производительность секвенирования увеличилась в 4–5 раз (до 100 млн. нуклеотидов за один рабочий цикл). Правда, возрос и расход реагентов (примерно в 2 раза). В результате расчётная стоимость секвенирования генома человека с использованием

¹ На самом деле четыре часа шло само секвенирование в виде построения новых цепей ДНК полимеразой и детекции протекания этого процесса, а подготовительная работа занимала несколько дней.

данной технологии уменьшилась примерно в 2 раза - с двух до одного миллиона долларов [Droege, Hill, 2008]. Однако в 2016 году производство секвенаторов этого типа было прекращено в связи с тем, что пиросеквенирование не смогло конкурировать с нарождающимися новыми подходами. В первую очередь - с полупроводниковым секвенированием.

Полупроводниковое секвенирование. Для более подробного описания процесса полимеризации ДНК уравнение (2) требует уточнения, поскольку присоединение очередного нуклеотида к растущей цепи ДНК приводит к выделению не только пирофосфата. Выделяется ещё и протон, изменяющий pH раствора. В уравнении (3) $\Delta H^+(\Delta pH)$ обозначает изменение концентрации протонов в реакционной смеси, приводящее к сдвигу pH.



В 2001 году С. Toumazou и S. Purushothaman подали заявку на патент, в котором предлагалось для регистрации синтеза ДНК добавлять в реакционную смесь пирофосфатазу. И определять снижение pH, происходящее в результате гидролиза выделяющегося пирофосфата [Toumazou, Purushothaman, 2001]. Эту реакцию они предложили использовать для секвенирования ДНК [Purushothaman et al., 2002], для определения однонуклеотидных полиморфизмов [Purushothaman et al., 2006] и для проведения ПЦР в реальном времени [Toumazou et al., 2013]. В итоге несколько патентных заявок и патентов, имеющих отношение к секвенированию на основе регистрации изменения pH при синтезе ДНК, стали принадлежать основанной С. Toumazou английской компании DNA Electronics Ltd. Но работоспособную (и конкурентоспособную) технологию секвенирования ДНК, построенную на регистрации изменения pH, разработала небольшая американская компания Ion Torrent. Её основателем в 2008 году стал всё тот же J.M.Rothberg, ушедший из 454 Life Science.

В сентябре 2010 г., незадолго до начала производства полупроводниковых секвенаторов, компания Ion Torrent заключила неэксклюзивное лицензионное соглашение с DNA Electronics Ltd., позволившее устранить возможные коммерческие риски. После чего её приобрела корпорация Life Technologies за 725 миллионов долларов. А в 2012 году уже сама Life Technologies была приобретена корпорацией Thermo Fisher Scientific за 13,6 миллиардов долларов. Незадолго до этого полупроводниковая технология была успешно использована для секвенирования индивидуального генома Г.Мура [Rothberg et al., 2011] – одного из основателей компании Intel и автора широко известного закона развития микроэлектроники [Moore, 1965].

Технология, разработанная компанией Ion Torrent, построена на регистрации изменения pH при синтезе ДНК под действием ДНК-полимеразы. Её основные стадии мало отличаются от стадий

пиросеквенирования компании 454 Life Science, поэтому на разработку полупроводникового секвенатора ушло всего три года. В обеих технологиях считывание информации происходит при циклической поочередной подаче в проточные ячейки четырёх растворов дНТФ высокой чистоты. Но в полупроводниковом способе встраивание (или отсутствие встраивания) нуклеотидов определяется по изменению pH. При этом сенсорные лунки в проточных ячейках полупроводниковых секвенаторов на порядок меньше лунок в ячейках пиросеквенаторов (рис. 2), что значительно снижает расход реагентов и увеличивает производительность приборов такого типа.



Рис. 2. Размеры гелевых микросфер из полиакриламида и ячеек на чипах, используемых в полупроводниковом секвенировании (масштаб не соблюден)

Fig. 2. Dimensions of polyacrylamide gel microspheres and cells on chips used in semiconductor sequencing (scale are not kept)

Замена биолюминесцентной регистрации полимеризации ДНК прямым определением pH значительно удешевило как секвенаторы, так и расходимые ими реагенты, поскольку главными их компонентами стали обычные дНТФ, тогда как для регистрации выделения пирофосфата нужны дополнительные ферменты и реагенты, преобразующие пирофосфат через каскад реакций в кванты света.

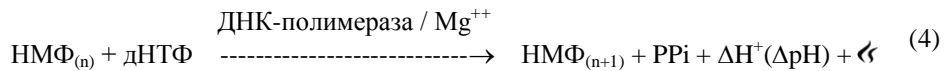
Встраивание в ДНК одного нуклеотида (точнее миллионов одинаковых нуклеотидов в такое же количество идентичных матриц, располагающихся на микросферах) приводит к сдвигу pH примерно на 0,02 единицы, что регистрируется специальными сенсорными чипами. По сути, полупроводниковый

секвенатор представляет собой высокочувствительный pH-метр с множеством измерительных ячеек. Перспективы, достоинства и недостатки полупроводниковой технологии секвенирования подробно разобраны в обзоре 2013 года [Зубов (Zubov), 2013].

Первый полупроводниковый секвенатор PGM (Personal Genome Machine), разработанный компанией Ion Torrent, появился в продаже в начале 2011 года. Его начальные характеристики были сравнительно скромными (10 - 15 млн. нуклеотидов за рабочий цикл), однако к 2013 году производительность выросла примерно на два порядка и превысила 1 млрд. нуклеотидов [Rothberg et al., 2011]. Продажи следующей модели (Ion Proton) начались в сентябре 2012 года. Производительность у нее значительно повысилась (~10 млрд. нуклеотидов), но качество секвенирования (длина ридов и точность чтения) несколько ухудшилось. Многие недоработки

этих моделей были устранены в следующем поколении приборов этого типа - S5 System и S5 XL System, появившихся в продаже в конце 2015 – начале 2016 годов. А в 2018 году появились Ion GeneStudio™ S5 System, Ion GeneStudio™ S5 Plus System и Ion GeneStudio™ S5 Prime System. Производительность этих систем зависит от типа используемых полупроводниковых pH-сенсорных чипов и может достигать 25 млрд. нуклеотидов за рабочий цикл.

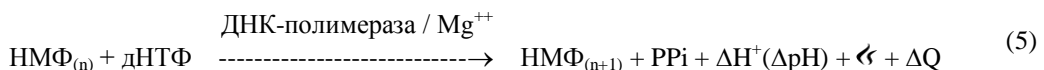
Термосеквенирование. Процесс ферментативного роста цепи ДНК не ограничивается появлением в реакционной смеси пирофосфата и протонов. При разрыве макроэргических связей в молекулах дНТФ выделяется тепло, которое также можно регистрировать. Поэтому уравнение (3) следует продолжить, вставив в неё иконку « \llcorner », обозначающую выделение тепла.



Разработке подобной технологии термосеквенирования ДНК был посвящен небольшой цикл работ американских авторов, также решивших регистрировать определенные изменения, происходящие в результате присоединения дезоксирибонуклеотидмонофосфата (расщепления дезоксирибонуклеотидтрифосфата) [Esfandyarpour, Davis, 2007; Esfandyarpour et al., 2008]. Поскольку при этом рвется макроэргическая связь, то в окружающую среду выделяется некоторое количество тепла и происходит локальный нагрев реакционных ячеек, что можно контролировать. Так, ими было обнаружено, что при присоединении дезоксирибонуклеотидмонофосфатов к миллионам одинаковых цепей ДНК, пришитых к микрошарику, помещенному в микроячейку, происходит нагрев последней на несколько сот микроградусов, что способны улавливать высокочувствительные термосенсоры. При этом измерительная ячейка прибора должна быть надежно термостатирована, чтобы полностью исключить влияние внешней среды,

в которой неизбежно происходят флуктуации температур, в том числе за счет конвекции. Данная технология не была доведена до работоспособного состояния, хотя на начальных стадиях разработки выглядела довольно перспективной.

Электронное секвенирование. Позднее этими же авторами [Esfandyarpour et al., 2020] был разработан иной подход к секвенированию ДНК, основанный регистрации изменения общего заряда ДНК. Дело в том, что каждый встраиваемый в ДНК нуклеотид содержит один остаток фосфорной кислоты, который в нормальных условиях заряжен отрицательно. Поэтому в уравнение (4), описывающее процесс роста цепи ДНК под действием ДНК-полимеразы, следует добавить еще один параметр, подающийся регистрации - ΔQ , обозначающий изменение импеданса сенсорной ячейки при удлинении ДНК.



Способ определения последовательностей нуклеотидов, основанный на регистрации импеданса, получил название электронного секвенирования ДНК. Его основные стадии мало отличаются от стадий полупроводникового секвенирования. В нем также происходит поочередная подача разных дНТФ, и для

регистрации произошедшего (или не произошедшего) встраивания дНМФ используются CMOS-сенсорные чипы. В 2010 году в США была основана фирма GenapSys Inc., которая поначалу должна была вывести на рынок новый геномный секвенатор для упоминавшегося выше термосеквенирования ДНК, но

затем она переключилась на электронное секвенирование ДНК. Относительно недавно данной фирмой на эту технологию был получен очередной патент США за номером US 10,093,975 B2 от 28 января 2020 года.

В настоящее время GenapSys производит довольно компактный геномный секвенатор (4,3 кг), который комплектуется двумя чипами – G2-1M и G3-16M, содержащими, соответственно, 1 или 16 миллионов сенсоров. 16M чип позволяет определять до 2 миллиардов нуклеотидов за рабочий цикл, длящийся около 24 часов. Средняя длина чтения составляет 150 нуклеотидов при общей заявленной точности Q20 (1% ошибок). При этом 85% последовательностей определяются с точностью Q30 (<0,1% ошибок), а 65% - Q40 (<0,01% ошибок). В 2021 году предполагается появление более производительного чипа G4-50M со 144 миллионами сенсоров, который позволит «читать» до 20 миллиардов нуклеотидов за запуск.

Все описанные выше варианты технологий NGS, рассчитанные на детекцию побочных продуктов реакции полимеризации ДНК, имеют общий недостаток в виде проблемного прочтения протяженных гомополимерных участков. Конечно, этот недостаток может быть преодолен за счет использования дНТФ с удаляемыми терминаторными группами, но в таком случае длительность секвенирования сильно возрастет, а стоимость одного шага заметно повысится. Однако именно возможность временной терминации роста цепи ДНК реализована в рассматриваемых ниже вариантах флуоресцентного секвенирования ДНК.

Флуоресцентное секвенирование с помощью мостиковой ПЦР. Идея использования флуоресцентно меченых дНТФ с обратимым блокированием полимеразного синтеза для секвенирования ДНК, применяемая во всех (за редким исключением) современных флуоресцентных секвенаторах, зародилась в середине 90-х годов прошлого века в Кембридже. Там же в 1998 году была организована компания Solexa, задачей которой стала реализация этой идеи. Привлечение крупных инвестиций позволило ей в 2004 году приобрести швейцарскую компанию Manteia, а с ней – технологию синтеза молекулярных кластеров ДНК и меченых дНТФ с флуорофорами, присоединяемыми к основаниям нуклеотидов легко расщепляемыми дисульфидными линкерами. В 2005 году Solexa объединилась с компанией Linx Therapeutics Inc., разработавшей довольно сложную технологию секвенирования коротких последовательностей нуклеотидов - MPSS (Massively Parallel Signature Sequencing). В этой технологии использовалось получение клонов ДНК на

микросферных (5 мкм) носителях при помощи эмульсионной ПЦР, и циклически повторяющееся отщепление концевых участков секвенируемых фрагментов ДНК ПS-рестрикционной эндонуклеазой с последующей лигазной пришивкой к образовавшимся липким концам меченых флуорофорами двунитевых адапторов, содержащих необходимые для повторения циклов сайты рестрикции [Brenner et al., 2000]. Но наиболее ценным приобретением в данном случае стала приборная база, позволяющая сканировать большое количество микросфер с молекулярными клонами фрагментов ДНК, меченых различными флуорофорами.

Благодаря этому объединению, напряжённой работе, а также преобразованию в акционерное общество и привлечению значительных инвестиций, в 2006 году была завершена разработка первой модели флуоресцентного секвенатора GA1 с производительностью рабочего цикла до 1 млрд. нуклеотидов. В начале 2007 года Solexa была поглощена компанией Illumina, которая приложила немало усилий для развития флуоресцентной технологии секвенирования и совершенствования модельного ряда своих секвенаторов, и теперь является бесспорным лидером рынка NGS.

В основе технологии, развиваемой компанией Illumina, лежит получение мультимолекулярных клонов ДНК на поверхности стекла при помощи твердофазной мостиковой ПЦР. Нуклеотидная последовательность полученных молекулярных клонов определяется при помощи включения ДНК-полимеразой флуоресцентно меченых дезоксинуклеозидтрифосфатов, служащих временными терминаторами дальнейшего роста цепи ДНК. Терминирующие группы и флуоресцентные метки после регистрации сигнала удаляются, и цикл повторяется - добавляются новые порции аналогичных дезоксинуклеотидтрифосфатов. Здесь следует заметить, что авторами мостиковой ПЦР являются основатели фирмы Mosaic Technologies, Inc., получившие на эту технологию патент США [Adams, Kron, 1997] и не упомянувших ее возможное применение для секвенирования ДНК.

Основной частью секвенатора является проточная ячейка, смонтированная на позиционирующем столике флуоресцентного микроскопа. Количество сканируемых в одной проточной ячейке молекулярных клонов (кластеров молекул диаметром 1 - 2 мкм) при их оптимальной плотности достигает 40 миллионов. Продолжительность рабочего цикла обычно составляет 72 часа. В настоящее время выпускается целая линейка таких приборов с производительностью от 1,2 миллиарда до 6 триллионов нуклеотидов за рабочий цикл.

Используемые в первых секвенаторах модифицированные дНТФ метили четырьмя различными флуорофорами, но с появлением модели NextSeq начался переход на двухцветные дНТФ и приборы с компактными оптическими модулями, сканирующими только два флуорофора. А в 2018 году появился iSeq 100 – первый «моноканальный» секвенатор, в котором применяется “One-Channel SBS Chemistry” («одноканальная химия»). iSeq 100 отличается лёгкостью (15,9 кг), компактностью (42,5x30,5x33 см) и сравнительно невысокой стоимостью (\$20000). За один рабочий цикл продолжительностью 17,5 часов он способен генерировать до 1,2 млрд. нуклеотидов. Этого вполне достаточно для решения большинства задач клинического секвенирования, но для оцифровки генома человека с высокой точностью желательно иметь его 30-кратное покрытие ридями. Отсюда следует, что производительность полногеномного секвенатора должна быть, как минимум, на пару порядков выше (~100 млрд. нуклеотидов). Этим требованиям удовлетворяет последняя разработка компании Illumina - NovaSeq 6000. Стоимость этого секвенатора может показаться чрезмерной (\$985000), но за один рабочий цикл этот прибор оцифровывает до 6 триллионов оснований. Появление подобных высокопроизводительных секвенаторов сделало возможным поточное секвенирование индивидуальных геномов человека. Причём с недавних пор такие приборы производятся не только в США, но и в Китае.

В России Институтом аналитического приборостроения РАН совместно с ООО «Синтол» ведется разработка полногеномного секвенатора ДНК Нанофор СПС, аналогичного по рабочим характеристикам прибору MiSeq производства компании Illumina. Опытный образец первого отечественного полногеномного секвенатора ДНК успешно прошел государственные испытания. Серийный выпуск прибора запланирован на конец 2021 года.

Флуоресцентное секвенирование с применением наночастиц. В Китае предпринималось несколько попыток наладить сборку, скопировать или разработать аналоги американских пиросеквенаторов, а также полупроводниковых и флуоресцентных приборов, но наблюдаемый прорыв в области секвенаторостроения там наметился только после приобретения китайской компанией BGI американской компании Complete Genomics, добившейся в этой области заметных успехов.

Первые китайские полногеномные секвенаторы (BGISEQ-50 и BGISEQ-500) немного уступали американским аналогам по максимальной длине чтения (100 или 100x2 нуклеотидов вместо 150

или 150x2 нуклеотидов), но точность чтения ДНК у них была вполне удовлетворительной. После доработки этих моделей на рынке NGS (только в Китае) в 2018 году появились новые флуоресцентные секвенаторы - MGISEQ-200 и MGISEQ-2000 [Mak et al., 2017]. Их заявленная производительность составляет 60 миллиардов нуклеотидов для MGISEQ-200 и 1080 миллиардов нуклеотидов для MGISEQ-2000. В конце 2019 года в продаже появился секвенатор MGISEQ-T7 с производительностью 6 триллионов нуклеотидов, позволяющей ему конкурировать с американским NovaSeq 6000. А в марте 2020 года на конференции AGBT состоялась презентация новой китайской разработки - роботизированной геномной платформы, предназначенной для использования в нацеленных на массовое секвенирование геномов человека крупных национальных проектах.

Особенностью используемой в этих секвенаторах технологии является применение ДНК-наночастиц (DNB – DNA nanoballs) – клубков одноцепочной ДНК, формирующихся путем амплификации катящимся кольцом и содержащих сотни или даже тысячи tandemно повторяющихся секвенируемых участков ДНК, а также упорядоченное расположение ДНК-наночастиц на сканируемой поверхности проточных рабочих ячеек. Важным преимуществом данной технологии является то, что в каждом цикле (обороте ДНК полимеразы) происходит копирование исходной матрицы ДНК, что уменьшает число возможных ошибок, не вызывая их «наслоения».

У китайской и американской технологий флуоресцентного секвенирования есть и другие отличия, но они не столь существенны. Объединяет их то, что в обоих случаях секвенирование проводится в результате синтеза полимеразой комплементарной цепи ДНК (SBS – sequencing by synthesis), а также то, что субстратами, включаемыми в растущую цепь ДНК, служат модифицированные дНТФ, содержащие 3'-концевые метилазидные группы и флуорофоры, связанные с основаниями нуклеотидов метилазидными линкерами. Важной особенностью азидометильных производных дНТФ, синтез которых был разработан российскими учёными (ИБХ РАН) в начале 90-х годов прошлого века [Chen et al., 2013], является их сравнительно высокая стабильность, сочетающаяся с простотой и быстротой деблокирования при обработке ДНК-наночастиц раствором трис(2-карбоксиил)фосфина (TCEP) [Liu et al., 2010]. Именно сложность синтеза таких реагентов до недавних пор защищала компанию Illumina от конкурентов, а освоение их производства позволило Китаю догнать (или почти догнать) Америку.

В сентябре 2017 года компания BGI объявила о начале приёма заявок на секвенирование геномов по \$600, что сразу же сделало Китай одним из лидеров геномных гонок. Значительное снижение стоимости секвенирования стало возможным благодаря созданию CNGB (China National GeneBank) — крупного центра, располагающего 150 флуоресцентными китайскими секвенаторами BGISEQ-500.

Составить конкуренцию описанным выше флуоресцентным технологиям NGS (технологиям второго поколения) способны только технологии секвенирования третьего поколения, основанные на чтении нуклеотидных последовательностей отдельных молекул ДНК (или РНК).

Мономолекулярные технологии геномного секвенирования

Истинное мономолекулярное секвенирование (tSMS). Впервые реализовать секвенирование единичных молекул ДНК в виде работающей технологии и готового прибора удалось компании Helicos BioScience, которая в 2009 году выпустила на рынок мономолекулярный флуоресцентный секвенатор HeliScope™. В данном приборе, рассчитанном на технологию tSMS™ (true Single Molecule Sequencing), проводилась временная терминация роста цепей ДНК, пришитых к стеклянной подложке. Для этого ДНК фрагментировалась на короткие участки, к концам которых добавлялся гомополимер из примерно десяти азотистых оснований (например, AAAAAAAAAA). Назначение этой олигонуклеотидной последовательности — закрепиться на одной из миллиардов «опор» — столбиков с комплементарным набором нуклеотидов (в данном случае - TTTTTTTTTT), заранее нанесенных на стеклянную подложку. После такой фиксации (гибридизации гомополимерных участков) все фрагменты исходной молекулы ДНК оказываются расположены на стекле, которое подвергают многочисленным циклам поочередного добавления/удаления различных флуоресцентно-меченных дНТФ, содержащих удаляемые 3'-концевые группы.

Для удешевления данной технологии все азотистые основания метятся одним и тем же флуорофором. После освещения подложки лазером на ней возникают светящиеся точки, но лишь в тех местах, где произошла достройка очередного нуклеотида. С помощью CCD-камеры высокого разрешения и сканирования всей поверхности проточной ячейки производится панорамный «фотоснимок», на котором видно, где присоединился очередной нуклеотид. После удаления ультрафиолетом защитной группы и промывки подложки 3'-концы всех «столбиков» ДНК готовы к

присоединению очередного комплементарного нуклеотида. И так шаг за шагом. Компьютер записывает положение миллионов вспышек после каждой реакции, фактически работая в режиме цейтраферной съемки. Этот метод позволяет читать миллиарды нуклеотидов в день. К сожалению, длина читаемых таким образом фрагментов ДНК не превышает в среднем 32 звеньев (разброс от 24 до 70 нуклеотидов), что достаточно лишь для ресеквенирования ДНК, а не для секвенирования *de novo*. Тем не менее, с помощью этой технологии всего тремя (причем один из них – россиянин) участниками выполнения проекта завершено секвенирование генома одного из авторов (руководителя работы) за очень короткий срок и с рекордно низкими для того времени затратами, составившими всего около 50 тыс. долларов [Pushkarev et al., 2009]. Однако прибор оказался очень дорогим (\$900000), громоздким (приблизительно одну тонну весом), да и чтение ДНК было очень неточным, поэтому в 2010 году эта компания обанкротилась.

SMRT секвенирование ДНК. В 2009 году появилась ещё одна технология мономолекулярного секвенирования ДНК, также основанная на использовании ДНК-полимеразы. В отличие от технологии tSMS, в которой фиксируются нуклеиновые кислоты, в этой к подложке пришивается ДНК-полимераза [Eid et al., 2009]. И секвенирование ведется в режиме реального времени, без смены реагентов. Данный подход, названный Single Molecule Real Time (SMRT) DNA Sequencing, был реализован американской компанией Pacific Biosciences, Inc.

Технология SMRT построена на использовании ячеек с лунками, которые не пропускают свет из-за слишком маленького диаметра, но при интенсивном освещении этих лунок в их придонной части происходит возбуждение флуорофоров. Ко дну таких лунок прикрепляют одиночные молекулы ДНК-полимеразы, попадающие в небольшие (20 цептолитров) области возбуждения флуорофоров, создаваемые с помощью набора специальных лазеров. В ячейку добавляют дезоксинуклеотидтрифосфаты всех четырех типов, меченые разными флуорофорами, причём эти флуорофоры присоединены не к основанию нуклеотида, а к остатку фосфорной кислоты в γ -положении. Присоединение нуклеотида (дезоксинуклеотидмонофосфата) к цепи ДНК приводит к отщеплению пирофосфата, а с ним и флуоресцентной метки.

Благодаря броуновскому движению свободно «плавающие» меченые молекулы дНТФ не задерживаются надолго в каждой точке и не оставляют заметного сигнала, тогда как во время

присоединения очередного комплементарного нуклеотида он на некоторое время (порядка миллисекунды против 2 – 10 микросекунд при обычной диффузии) удерживается ДНК полимеразой. После присоединения нуклеотида светящаяся метка теряет с ним связь и быстро удаляется диффузией. Благодаря этому в сканируемой области уровень фоновой флуоресценции сравнительно невелик, но при встраивании нуклеотидов в ДНК удается зарегистрировать достаточно чёткие пики флуоресценции четырёх меток, порядок чередования которых соответствует последовательности встраиваемых нуклеотидов.

Скорость чтения ДНК в каждой лунке такого прибора равна скорости работы ДНК-полимеразы. А общая производительность определяется количеством таких лунок в ячейке секвенатора. При этом основания синтезированной ДНК не содержат характерных для других флуоресцентных технологий линкерных «хвостов», которые могут тормозить удлинение ДНК. Поэтому длина достраиваемых полимеразой (и секвенируемых) участков ДНК измеряется тысячами, а иногда и десятками тысяч нуклеотидов.

Технология SMRT подходит и для эпигеномного секвенирования, т.е. для определения модифицированных оснований [Flusberg et al., 2010; Fang et al., 2012]. Было показано, что продолжительность процесса присоединения тимина заметно зависит от того, находятся ли в соответствующем положении матрицы обычный аденин или его метилированный аналог (доли секунды против полутора секунд соответственно). Подобные задержки синтеза характерны и для метилированного цитозина, определяемого обычно при помощи сложного бисульфитного секвенирования.

В 2010 году компания Pacific Biosciences выпустила на рынок NGS секвенатор PacBio RS. Прибор оказался довольно дорогим (\$695000) и громоздким (около одной тонны), но вскоре появилась его доработанная версия PacBio RSII, а в начале 2017 года в продаже появился секвенатор Sequel, отличающийся от своих предшественников повышенной производительностью (до 20 миллиардов нуклеотидов на 500 тысяч лунок), сниженной стоимостью (\$350000) и улучшенной транспортабельностью (350 кг). А в 2019 ему на смену пришёл Sequel II, производительность которого была увеличена примерно в 8 раз.

Однако мономолекулярная технология секвенирования ДНК SMRT характеризуется относительно низкой точностью чтения, частично компенсируемой большой длиной секвенируемых фрагментов ДНК (15 - 20 тыс. нуклеотидов). В результате Sequel используется в основном не для

выявления редких нуклеотидных замен, а для обнаружения крупных делеций, инсерций и многочисленных тандемных повторов ДНК, затрудняющих завершение геномного секвенирования при использовании только коротких ридов [Chin et al., 2016; Yoshinaga et al., 2018]. Но сейчас компания Pacific Biosciences быстро теряет этот сектор рынка NGS. И уступает его компании Oxford Nanopore Technologies (ONT), совершенствующей нанопоровую технологию секвенирования нуклеиновых кислот².

Нанопоровое секвенирование нуклеиновых кислот. Появление идеи использования для секвенирования ДНК регистрации изменений проводимости ионных каналов, формируемых некоторыми мембранными белками и токсинами в бислойной липидной мембране, датируется серединой 90-х годов прошлого века. Описание связанных с этим событий довольно подробно изложено в историческом обзоре и комментариях к нему, посвященных трем десятилетиям разработки нанопорового секвенирования нуклеиновых кислот [Deamer et al., 2016; Kasianowicz, Bezrukov, 2016]. Но для реализации этой технологии потребовались длительные прикладные и фундаментальные исследования. Поэтому нанопоровое секвенирование можно считать ярким примером реализации на практике красивой, но надолго отложенной идеи.

Первая публикация, посвященная использованию нанопоры гептамерного белка α -гемолизина для установления последовательности нуклеотидов, вышла еще в 1996 году [Kasianowicz et al., 1996]. Но перед этой публикацией 17 марта 1995 года G. Church, D.W. Deamer, D. Branton, R. Baldarelli и J. Kasianowicz подали заявку на патент США (Pat. US 5,795,782). За долгие годы, ушедшие на совершенствование нанопорового секвенирования, было опубликовано множество экспериментальных работ и обзорных публикаций, посвященных разным вопросам развития этой технологии [Deamer, Akeson, 2000; Maitra et al., 2012; Ying et al., 2013; Wang et al., 2015; Lu et al., 2016; Rang et al., 2018; Runtuweni et al., 2019; Senol et al., 2019; Makalowski, Shabardina, 2020; Xu et al., 2020]. В одном из обзоров [Feng et al., 2015] нанопоровое секвенирование сочли четвертым поколением методов секвенирования ДНК. Если считать методы Максама-Гильберта [Maxam, Gilbert, 1977] и Сэнгера [Sanger et al., 1977] первым поколением, метод массового параллелизма – вторым, а мономолекулярное секвенирование с использованием процессов полимеризации ДНК третьим, то нанопоровое секвенирование вполне

² PacBio vs. Oxford Nanopore sequencing. – URL: <https://blog.genohub.com/2017/06/16/pacbio-vs-oxford-nanopore-sequencing/>

соответствует четвертому поколению, хотя идея его появилась гораздо раньше, однако воплощение сильно затянулось.

В основе нанопорового секвенирования лежит «протяжка» полимерной цепочки нуклеотидов через отверстие нанометрового размера (ионный канал) под действием приложенного электрического поля. В ходе этого процесса производится регистрация изменений ионной проводимости такого канала, позволяющая определять последовательность нуклеотидов.

В 2005 году в Великобритании была основана компания Oxford Nanopore Technologies Ltd., занимающаяся исключительно разработкой нанопоровой технологии секвенирования. Выпуск первых секвенаторов, работающих по этой технологии, несколько раз откладывался (начиная с 2012 года [Eisenstein, 2016]) и стартовал только в 2014 году. Да и то по программе «раннего доступа», позволившей компании снять с себя ответственность за получаемые результаты и привлечь к доведению технологии до работоспособного состояния широкий круг сторонних специалистов [Lu et al., 2016]. Но только в 2016 году, благодаря нескольким значительным усовершенствованиям данной технологии, а также тщательной оптимизации всех её составляющих, точность нанопорового секвенирования приблизилась к 90% - к показателю, обеспечившему возможность её широкого применения в научных и медицинских учреждениях [Ameur et al., 2018].

Ключевым элементом технологии нанопорового секвенирования является белковая пора – ионный канал в бислоидной липидной мембране. Через этот канал под действием приложенного к мембране электрического потенциала движется отрицательно заряженная однонитевая цепь ДНК, причём при прохождении через самую узкую часть поры разные нуклеотиды по-разному влияют на величину проходящего через пору ионного тока. После усиления и оцифровки модуляций ионного тока получается диаграмма, расшифровка которой позволяет определить последовательность прошедших через пору нуклеотидов.

Теоретически всё выглядит очень просто, но на практике эта технология стала работоспособной только после решения нескольких довольно сложных задач. Во-первых, сама по себе однонитевая ДНК обычно движется через пору со скоростью около миллиона нуклеотидов в секунду, тогда как электронная система секвенатора не может считывать за секунду больше тысячи дискретных сигналов. Для торможения ДНК пытались использовать различные приёмы, но самым лучшим вариантом оказалось ферментативное торможение хеликазами – ферментами, расплетающими двунитевую ДНК с использованием энергии АТФ. Скорость работы

хеликазы, используемой в последних наборах реагентов ONT, составляет 450 нуклеотидов в секунду, что почти оптимально для оцифровки считываемых сигналов.

Ещё одной проблемой является нестабильность обычных мембран, из-за чего пришлось разрабатывать технологию получения их более стабильных синтетических аналогов на основе триблок-сополимеров. И оптимизировать состав мембран для снижения фоновой проводимости. Тщательно оптимизировался также состав рабочих растворов, от которых зависит величина ионных токов, характер их модуляции, скорость работы хеликазы, стабильность белковой нанопоры, фоновая проводимость мембраны и т.п.

Особые требования нанопоровая технология секвенирования ДНК предъявляет к алгоритмам расшифровки считываемых сигналов, имеющих крайне запутанный профиль. Это связано с тем, что величина ионного тока зависит не от отдельных нуклеотидов, а от их сочетаний, включающих одновременно по 4 - 5 нуклеотидов. Извлечь из модулируемых ими сигналов полезную информацию удаётся только при помощи сложного программного обеспечения, построенного на применении искусственного интеллекта,

С некоторыми нововведениями, используемыми в технологии нанопорового секвенирования, можно ознакомиться по патентам, принадлежащим компании ONT³, но технология быстро развивается, и о некоторых её секретах можно только догадываться. Одним из лучших нововведений стал переход на поры R9.4 и R9.5, ионные каналы которых отличаются широкими входным и выходным вестибулями, а также минимальной длиной сенсорного отверстия, оптимизированного методами белковой инженерии⁴.

Применение девятого поколения нанопор (R9), кодируемых модифицированными производными гена CsgG, позволило к концу 2017 года увеличить точность однократного чтения ДНК до 87 - 90%, к октябрю 2018 года до 92 - 93%, а к концу 2019 года – до 95 - 97% (с порами R9.4.1). И повысить точность многократного чтения до Q40 (99,99%) [Jain et al., 2018]. У появившегося недавно следующего поколения пор (R10) в наиболее узкой части имеется две перетяжки, что повышает точность чтения гомополимерных повторов. Сочетание пор девятого поколения (R9.4.1) с порами десятого поколения (R10.3) позволяет увеличить среднюю точность многократного чтения ДНК до Q50 (99,999%). Из

³ Oxford Nanopore Technologies. Patents. – URL: <https://nanoporetech.com/patents>

⁴ Mutant CsgG pores. – Pat. Appl. US 2017/0283470 A1

сравнительно недавних нововведений можно отметить также переход на чипы (“Rev D”), позволяющие менять полярность потенциала в сенсорных лунках и восстанавливать работоспособность заблокированных пор. Это позволило довести производительность секвенатора MinION до 30 миллиардов нуклеотидов за один рабочий цикл.

Широкому применению секвенатора MinION как в научных исследованиях, так и в лабораторной диагностике мешает высокая стоимость его расходных ячеек (\$500 - \$900, в России - 150 тыс. руб.). Но с 2020 года этот прибор поставляется с адапторной вставкой (*Flongle*), позволяющей работать с менее дорогими (хотя и с менее производительными) одноразовыми ячейками (~\$100). С этими же ячейками работает и MinION Mk1C – вариант MinION, встроенный в миникомпьютер с операционной системой Linux. Это повысило стоимость секвенатора, но позволило устранить проблемы с операционной системой Windows 10, у которой очень сложно отключить автообновления, автоархивацию данных, антивирусные проверки и некоторые другие функции, непредсказуемое включение которых может вызывать сбои при секвенировании ДНК.

Фирма ONT производит целую линейку нанопоровых секвенаторов, представленных на их сайте <https://nanoporetech.com/>. Так, рабочие ячейки секвенатора MinION используются и в секвенаторе GridION – «пятыячеечном» варианте MinION, оснащённом высокопроизводительным сервером для обработки всей получаемой информации. После недавнего апгрейда вычислительной системы он получил название GridION Mk1. Для полногеномного секвенирования в ONT разработали секвенаторы PromethION 24 и PromethION 48, позволяющие одновременно работать с 24 или 48 ячейками, отличающимися повышенной производительностью (130 – 200 миллиардов нуклеотидов на ячейку). Но такие ячейки и стоят по 2000 долларов. Правда, при оптовых закупках (2880 шт. - \$1 800 000) их стоимость может уменьшиться до 625 долларов.

Применение нанопоровой технологии секвенирования ДНК способно значительно уменьшить стоимость и таргетного, и геномного секвенирования, но в полной мере проявиться эта способность сможет только после появления на рынке конкурентов. Например, компании Axbio Inc. (США) [Ivanov et al., 2018], планирующей начать бета-тестирование нанопоровых секвенаторов нового поколения [Ivanov I., pers. commun.].

Себестоимость полногеномного секвенирования

Чтобы проследить динамику удешевления стоимости секвенирования полных геномов за последние два десятилетия, необходимо вернуться в 2001 г., когда было объявлено о завершении секвенирования в

черновых вариантах двух геномов человека, выполненных международным консорциумом и Институтом геномных исследований под руководством К. Вентера. Первый проект выполнялся 13 лет (вместо планировавшихся 15 и изначально даже 17 лет) и было потрачено приблизительно 3 млрд. долларов, из которых только около 300 млн. пошло непосредственно на проведение секвенирующих процедур. Что касается проекта Вентера, то при его выполнении был применен передовой для того времени так называемый «shot-gun» или иначе «случайный» подход к секвенированию больших геномов, что позволило справиться с секвенированием генома человека всего за несколько лет, затратив всего около 100 млн. долларов. Так, с помощью данного подхода к концу 2003 г. стоимость секвенирования генома млекопитающего, приблизительно равного по размеру человеческому, составила около 25 млн. долларов. В этот период единственным методом секвенирования полных геномов был метод Сэнгера. Дальнейшее снижение стоимости полногеномного секвенирования уже продолжалось с помощью разных методов новых поколений. Так, в 2008 г. при секвенировании генома Дж. Уотсона с помощью метода пиросеквенирования ДНК достигнут исторический рубеж в один миллион долларов.

Как следует из информации на сайте National Human Genome Research Institute, на протяжении многих лет отслеживающего стоимости ресеквенирования генома человека, видно, что темпы снижения стоимости с 2001 г. до конца 2007 г. были практически сопоставимы с тем, что происходит в микроэлектронике, для которой в середине 60-х годов прошлого столетия Г.Мур высказал предположение (получившее известность как «закон Мура»), что число транзисторов на кристалле будет удваиваться каждые два года, после того, как он подметил, что это происходит на протяжении нескольких лет [Moore, 1965]. И действительно эта тенденция в микроэлектронике сохраняется уже на протяжении более полувека. Но с начала 2008 г. кривая стоимости секвенирования пошла резко вниз и к концу 2015 г. вместо того, чтобы согласно закону Мура уменьшиться за этот период приблизительно в 8 раз, она снизилась на четыре (!) порядка и составила немногим более одной тыс. долларов для генома человека (рис. 3). А если сравнить с 2001 годом, то окажется, что стоимость секвенирования генома человека упала уже на пять порядков. Справедливости ради следует отметить, что в последние годы темпы снижения стоимости секвенирования серьезно замедлились. Не менее впечатляющим выглядит снижение стоимости секвенирования одного миллиона нуклеотидов (рис. 4) - с 10 миллионов долларов в 2001 году до 8 центов к концу лета 2020 года или в 12 миллионов раз (7 порядков) [Wetterstrand, 2021].

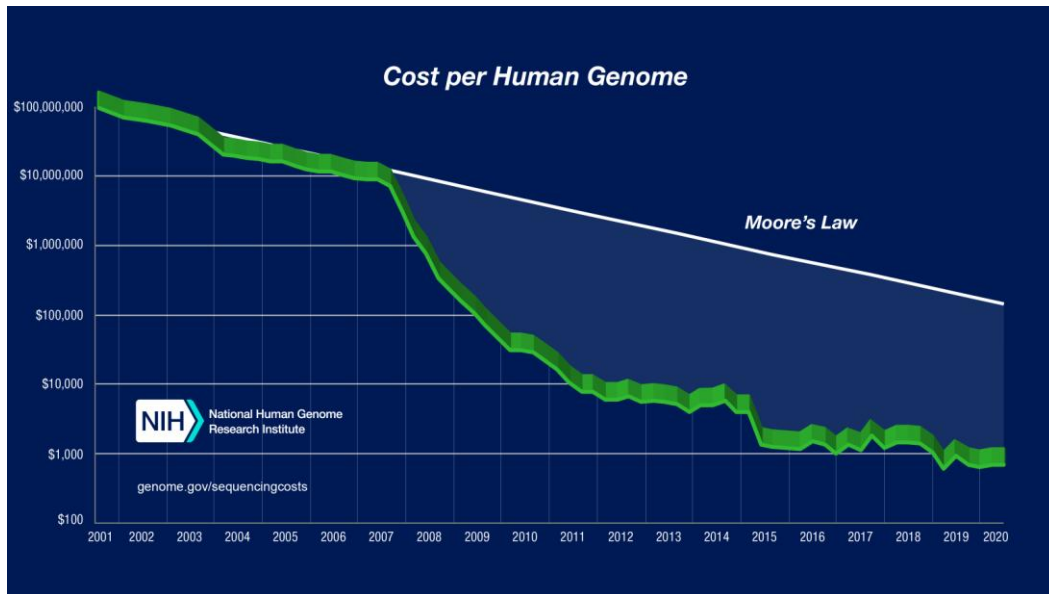


Рис. 3. Сопоставление снижения стоимости секвенирования генома человека с законом Мура
<https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data>
Fig. 3. Comparison of the reduction in the cost of sequencing the human genome with Moore's law
<https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data>

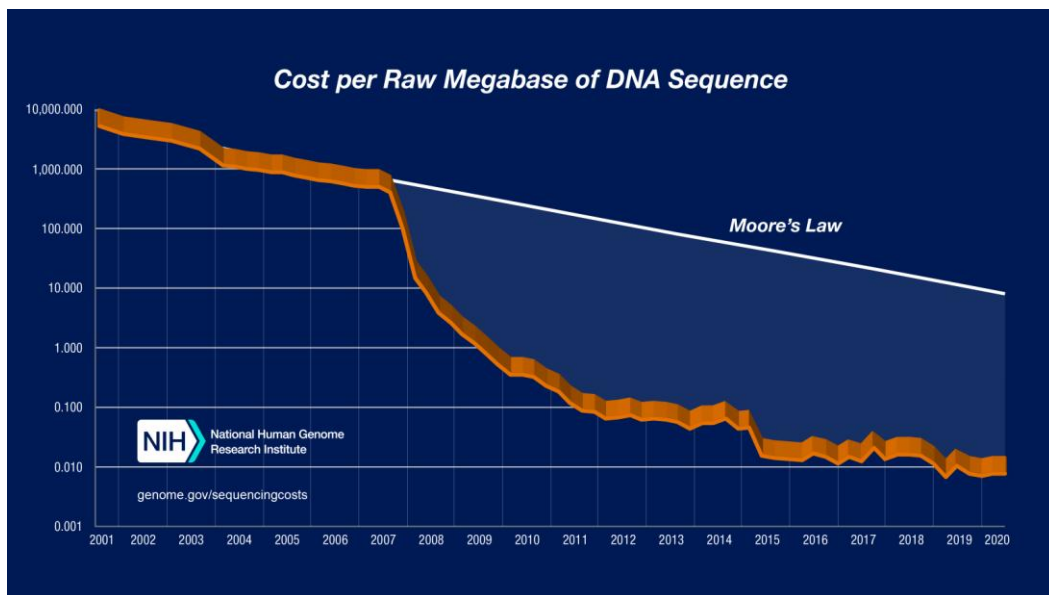


Рис. 4. Сопоставление снижения стоимости секвенирования одного миллиона нуклеотидов с законом Мура
<https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data>
Fig. 4. Comparison of the reduction in the cost of sequencing one million nucleotides with Moore's law
<https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data>

При этих подсчетах принималось, что за счет более высокой точности секвенирования по Сэнгеру со средней длиной чтения 500 - 600 нуклеотидов требуется 6-ти кратное покрытие генома, для пиросеквенирования со средней длиной чтения 300 – 400 нуклеотидов кратность покрытия уже была равна 10, а для флуоресцентного секвенирования, включая неупомянутую в данном обзоре технологию SOLiD с их 75 – 150 нуклеотидами чтения принималось даже 30-ти кратное покрытие. Необходимо также заметить, что эти суммы (усредненные по нескольким секвенирующим центрам) включают в себя помимо стоимости пробоподготовки и расходных материалов на само секвенирование, еще амортизацию секвенаторов и другого оборудования со сроком амортизации свыше трех лет, биоинформатическую составляющую, труд работников, административные расходы и прочие издержки в виде некоторых непрямых расходов.

Перспективы полногеномного секвенирования

Как уже говорилось выше, в 2003 г. на секвенаторщиков нашло некое прозрение и один за другим стали появляться многочисленные проекты нового секвенирования ДНК. В США, начиная с 2004 г. через NHGRI - National Human Genome Research Institute в рамках Программы развития технологий секвенирования (Sequencing Technology Development Program), им оказывалась, в том числе, и государственная поддержка. Тогда все проекты были первоначально поделены на две основные категории. Проекты первой должны были обеспечить секвенирование генома человека за 100 тыс. долларов, а второй - за 1 тыс. долларов. Как можно видеть, по прошествии полутора десятилетий этот рубеж оказался достигнут. Налицо огромный прогресс в очень важном деле секвенирования ДНК. Но возможно ли дальнейшее снижение стоимости секвенирования? Или теперь на первый план выходит повышение достоверности чтения? Ведь проблема коротких прочтений отчасти решена благодаря нанопоровому секвенированию, фактически не имеющему верхней границы длины ридов.

Ажиотажный спрос инвесторов и крупных корпораций на технологии NGS рождает соответствующие предложения, поэтому количество предлагаемых способов секвенирования нуклеиновых кислот, их вариантов и патентов на эту тему измеряется сотнями. Трудно перечислить и начинающие зарубежные компании, которые в последние годы пытались (или до сих пор пытаются) разработать новые технологии секвенирования. В подавляющем большинстве случаев они ориентируются на разработку мономолекулярных секвенаторов. При этом очень важным моментом для

массового использования нового удобного метода секвенирования ДНК будущих поколений станет легкая возможность секвенирования как ДНК, так и абсолютно всего пула молекул РНК, присутствующих в одной клетке (или в группе одинаковых клеток одного типа ткани без этапов амплификации), что позволит во многом приблизиться к пониманию функционирования всего геномного ансамбля [Picelli et al., 2017].

Быстро развивающаяся нанопоровая технология секвенирования ДНК способна потеснить всех конкурентов на рынке NGS. Точнее – почти всех, за исключением высокоточных технологий флуоресцентного секвенирования, разрабатываемых компаниями Illumina, Inc. в США и MGI Tech Co., Ltd. в Китае. При этом обострение конкурентной борьбы между разработчиками флуоресцентных технологий может привести к снижению стоимости полногеномного секвенирования. Нужно отметить, что потенциал совершенствования подобных технологий далеко не исчерпан. Значительно уменьшить расход реагентов может замена проточной системы их прокачки через рабочую ячейку секвенатора возвратной схемой («схема Ларина»), позволяющей многократно использовать дорогостоящие меченые дНТФ⁵. Существуют и другие потенциальные возможности значительного снижения стоимости флуоресцентного секвенирования ДНК. Например, корейские учёные прорабатывают возможность использования меток, присоединяемых не к основаниям нуклеотидов, а к 3'-ОН группе их дезоксирибозного кольца [Shin et al., 2011; Kim et al., 2014]. Такие модифицированные дНТФ могут быть намного дешевле, хотя возможность их использования зависит от наличия ДНК-полимеразы (или ревертазы), способной встраивать подобные нуклеотиды в синтезируемую цепь ДНК.

Все это может привести к появлению новых моделей флуоресцентных секвенаторов, отличающихся простотой конструкции (монокромных?) и высокоэкономичных – с возможностью многократного использования проточных ячеек и предельно минимизированным расходом реагентов. В результате прогнозы снижения стоимости полногеномного секвенирования до \$100 в течение 3 - 10 лет, вызывавшие большие сомнения пару лет назад, сейчас уже не кажутся столь фантастическими /79, 80/.

Нельзя также сбрасывать со счетов и возможное появление принципиально новых

⁵ Sequencing for all. – URL: <http://molbiol.ru/forums/index.php?act=Attach&type=post&id=272881>

технологий секвенирования нуклеиновых кислот очередных поколений, которые окажутся еще более производительными и точными.

Литература

1. Василенко С.К., Демушкин .П., Будовский Э.И., Кнорре Д.Г. Определение нуклеотидной последовательности в олигонуклеотидах // Докл. АН СССР. 1965. Т.162. С. 694-697.
2. Зубов В.В. Секвенирование по Ротбергу (потенциал полупроводникового секвенирования) // Биомика. - 2013. Т.5, №1-2. С.48-61.
3. Ребриков Д.В., Коростин Д.О., Шубина Е.С., Ильинский В.В. NGS: высокопроизводительное секвенирование. М.: Бином. Лаборатория знаний. 2014. 232 С.
4. Чемерис А.В., Ахунов Э.Д., Вахитов В.А. Секвенирование ДНК. М., Наука, 1999. 429 с.
5. Astbury W.T., Bell F.O. X-Ray study of thymonucleic acid. *Nature*. 1938. V. 141. P.747-748. doi:10.1038/141747b0
6. Avery O.T., MacLeod C.M., McCarty M. Studies of the chemical nature of the substance inducing transformation of pneumococcal types. Induction of transformation by a desoxyribonucleic acid fraction isolated from *Pneumococcus* Type III. *J. Exp. Med.* 1944. V.79(2). P.137-158. DOI: 10.1084/jem.79.2.137
7. Brenner S., Johnson M., Bridgham J., Golda G., Lloyd D.H., Johnson D., Luo S., McCurdy S., Foy M., Ewan M., Roth R., George D., Eletr S., Albrecht G., Vermaas E., Williams S.R., Moon K., Burcham T., Pallas M., DuBridge R.B., Kirchner J., Fearon K., Mao J-i., Corcoran K. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbeads arrays. *Nat. Biotechnol.* 2000. V.18(6). P.630-634. doi: 10.1038/76469
8. Chargaff E. What really is DNA? Remarks on the changing aspects of a scientific concept // *Prog. Nucl. Acid Res. Mol. Biol.* 1968. V.8. P.297-333. 10.1016/s0079-6603(08)60549-8
9. Chen F., Dong M., Ge M., Zhu L., Ren L., Liu G., Mu R. The history and advances of reversible terminators used in new generations of sequencing technology // *Genomics, Proteomics, Bioinformatics.* 2013. V.11(1). P.34-40. doi: 10.1016/j.gpb.2013.01.003
10. Chin C-S., Peluso P., Sedlazeck F.J., Nattestad M., Concepcion G.T., Clum A., Dunn C., O'Malley R., Figueroa-Balderas R., Morales-Cruz A., Cramer G.R., Delledonne M., Luo C., Ecker J.R., Cantu D., Rank D.R., Schatz M.C. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods.* 2016. V.13(12). P.1050-1054. doi: 10.1038/nmeth.4035
11. Deamer D., Akeson M., Branton D. Three decades of nanopore sequencing // *Nature Biotechnol.* 2016. V.34(5). P.518-524. doi: 10.1038/nbt.3423.
12. Droege M, Hill B. The Genome Sequencer FLX System--longer reads, more applications, straight forward bioinformatics and more complete data sets. *J Biotechnol.* 2008. V.136(1-2). P.3-10. doi: 10.1016/j.jbiotec.2008.03.021
13. Eid J., Fehr A., Gray J., Luong K., Lyle J., Otto G., Peluso P., Rank D., Baybayan P., Bettman B., Bibillo A., Bjornson K., Chaudhuri B., Christians F., Cicero R., Clark S., Dalal R., Dewinter A., Dixon J., Foquet M., Gaertner A., Hardenbol P., Heiner C., Hester K., Holden D., Kearns G., Kong X., Kuse R., Lacroix Y., Lin S., Lundquist P., Ma C., Marks P., Maxham M., Murphy D., Park I., Pham T., Phillips M., Roy J., Sebra R., Shen G., Sorenson J., Tomaney A., Travers K., Trulson M., Vieceli J., Wegener J., Wu D., Yang A., Zaccarin D., Zhao P., Zhong F., Korlach J., Turner S. Real-time DNA sequencing from single polymerase molecules // *Science.* 2009. V.323. P.133-138. doi: 10.1126/science.1162986
14. Esfandyarpour H., Davis R.W. An integrated differential nanocalimeter with on-chip microfluidic multiplexing for high throughput genomics and proteomics // 14th International Conference on Miniaturized Systems for Chemistry and Life Sciences, 3 - 7 October 2010, Groningen. P.1349-1351.
15. Esfandyarpour H., Zheng B., Pease R.F., Davis R.W. Structural optimization for heat detection of DNA thermosequencing platform using finite element analysis // *Biomicrofluidics.* 2008. V.2. P.24102. doi: 10.1063/1.2901138
16. Esfandyarpour H., Parizi K.B., Barmi M.R. <...> Dong B., Witney F.R., Davis R.W. High accuracy DNA sequencing on a small, scalable platform via electrical detection of single base incorporations. *bioRxiv preprint.* 2020. doi: 10.1101/604553
17. Fang G, Munera D, Friedman DI, Mandlik A, Chao MC, Banerjee O, Feng Z, Losic B, Mahajan MC, Jabado OJ, Deikus G, Clark TA, Luong K, Murray IA, Davis BM, Keren-Paz A, Chess A, Roberts RJ, Korlach J, Turner SW, Kumar V, Waldor MK, Schadt EE. Genome-wide mapping of methylated adenine residues in pathogenic *Escherichia coli* using single-molecule real-time sequencing // *Nat. Biotechnol.* 2012. V.30. P.1232-1239. doi: 10.1038/nbt.2432
18. Flusberg B.A., Webster D.R., Lee J.H., Travers K.J., Olivares E.C., Clark T.A., Korlach J., Turner S.W. Direct detection of DNA methylation during single-molecule, real-time sequencing // *Nat. Methods.* 2010. V.7. P.461-465. doi: 10.1038/nmeth.1459
19. Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A., Fleischmann R.D., Bult C.J., Kerlavage

- A.R., Sutton G., Kelley J.M., Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J., Nguyen D., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M., Tomb J-F., Dougherty B.A., Bott K.F., Hu P-C., Lucier T.S., Peterson S.N., Smith H.O., Hutchinson C.A., III, Venter J.C. The minimal gene complement of *Mycoplasma genitalium* // *Science*. 1995. V.270. P.397-403. DOI: 10.1126/science.270.5235.397
20. Hyman E.D. A new method of sequencing DNA // *Anal. Biochem.* 1988. V.174. P.423-436. DOI: 10.1016/0003-2697(88)90041-3
21. Jain M., Koren S., Miga K.H., Quick J., Rand A.C., Sasani T.A., Tyson J.R., Beggs A.D., Dilthey A.T., Fiddes I.T., Malla S., Marriott H., Nieto T., O'Grady J., Olsen H.E., Pedersen B.S., Rhie A., Richardson H., Quinlan A.R., Snutch T.P., Tee L., Paten B., Phillippy A.M., Simpson J.T., Loman N.J., Loose M. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat. Biotechnol.* 2018. V. 36(4). P.338-345. doi: 10.1038/nbt.4060
22. Ivanov I., Niu L., Chueh A., Tian H., Deng S. Apparatus and methods for continuous diagnostics of macromolecules. *Pat. Appl. US 2018/0230531 A1*.
23. Kasianovich J.J., Bezrukov S.M. On "three decades of nanopore sequencing" // *Nature Biotechnol.* 2016. V.34(5). P.481-482. doi: 10.1038/nbt.3570
24. Kim D-R., Kim T-S., Kim E., Min S-J., Shin D., Ahn D-R. Synthesis of 3'-O-fluorescently mono-modified reversible terminators and their uses in sequencing-by-synthesis. *Bioorg Med Chem Lett.* 2014. V.24(1). P.209-213. doi: 10.1016/j.bmcl.2013.11.040
25. Liu X., Milton J., Ruediger S. Labelled nucleotides. *Pat. US 7,795,424 B2.* (Sep 14, 2010)
26. Mak S.S.T., Gopalakrishnan S., Carøe C., Geng C., Liu S., Sinding M-H.S., Kuderna L.F.K., Zhang W., Fu S., Vieira F.G., Germonpré M., Bocherens H., Fedorov S., Petersen B., Sicheritz-Pontén T., Marques-Bonet T., Zhang G., Jiang H., Gilbert M.T.P. Comparative performance of the BGISEQ-500 vs Illumina HiSeq2500 sequencing platforms for palaeogenomic sequencing. *Gigascience.* 2017. V.6(8). P.1-13. doi: 10.1093/gigascience/gix049
27. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bembem LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM. Genome sequencing in microfabricated high-density picolitre reactors // *Nature*. 2005. V.437. P.376-380. doi: 10.1038/nature03959
28. Maxam A.M., Gilbert W. A new method for sequencing DNA // *Proc. Natl. Acad. Sci. USA*. 1977. V.74. P.560-564. doi: 10.1073/pnas.74.2.560
29. Moore G.E. Cramming more components onto integrated circuits // *Electronics*. 1965. V. 38. P. 114-117.
30. Purushothaman S., Toumazou C., Georgiou J. Towards fast solid state DNA sequencing // *Circuits and Systems*. 2002. V. 4. P.169-172. DOI: 10.1109/ISCAS.2002.1010416
31. Purushothaman S., Toumazou C., Ou C.P. Protons and single nucleotide polymorphism detection: A simple use for the ion sensitive field effect transistor // *Sensors and Actuators*. 2006. V.114(2). P.964-968. DOI: 10.1016/j.snb.2005.06.069
32. Pushkarev D., Neff N.F., Quake S.R. Single-molecule sequencing of an individual human genome. *Nat. Biotechnol.* 2009. V.27(9). P.847-850. doi: 10.1038/nbt.1561
33. Ronaghi M, Karamohamed S, Pettersson B, Uhlén M, Nyren P. Real-time DNA sequencing using detection of pyrophosphate release. *Anal Biochem.* 1996 Nov 1;242(1):84-9. doi: 10.1006/abio.1996.0432.
34. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, Milgrew MJ, Edwards M, Hoon J, Simons JF, Marran D, Myers JW, Davidson JF, Branting A, Nobile JR, Puc BP, Light D, Clark TA, Huber M, Branciforte JT, Stoner IB, Cawley SE, Lyons M, Fu Y, Homer N, Sedova M, Miao X, Reed B, Sabina J, Feierterstein E, Schorn M, Alanjary M, Dimalanta E, Dressman D, Kasinskas R, Sokolsky T, Fidanza JA, Namsaraev E, McKernan KJ, Williams A, Roth GT, Bustillo J. An integrated semiconductor device enabling non-optical genome sequencing // *Nature*. 2011. V.475. P.348-352. doi: 10.1038/nature10242
35. Sanger F., Nicklen S., Coulson A.R. DNA sequencing with chain-terminating inhibitors // *Proc. Natl. Acad. Sci. USA*. 1977. V.74. P.5463-5467. doi: 10.1073/pnas.74.12.5463.
36. Shin D., Ahn D-R., Ahn H-C. 3'-O-fluorescently mono-modified nucleotides and uses thereof / *Pat. US 8,030,466.* Oct., 4, 2011
37. Signer R., Caspersson T., Hammarsten E. Molecular shape of thymonucleic acid. *Nature*. 1938. V. 141. P.122. doi:10.1038/141122a0
38. Toumazou C., Purushothaman S. Sensing apparatus and method / *GB Pat. Appl. № 0105831.2* (09.02.2001).
39. Toumazou C., Shepherd L.M., Reed S.C., Chen G.I., Patel A., Garner D.M., Wang C-J.A., Ou C-P., Amin-Desai K., Athanasiou P., Bai H., Brizido I.M.Q.,

- Caldwell B., Coomber-Alford D., Georgiou P., Jordan K.S., Joyce J.C., La Mura M., Morley D., Sathyavvruthan S., Temelso S., Thomas R.E., Zhang L. Simultaneous DNA amplification and detection using a pH-sensing semiconductor system. *Nat. Methods*. 2013. V.10(7). P.641-646. doi: 10.1038/nmeth.2520
40. Watson J.D., Crick F.H.C. A structure for deoxyribose nucleic acid. *Nature*. 1953. V. 171(4356). P. 737-738. doi:10.1038/171737a0
41. Wetterstrand K.A. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP)
42. <https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data> [Accessed 02.02.2021].
43. Wheeler D.A., Srinivasan M., Egholm M., Shen Y., Chen L., McGuire A., He W., Chen Y.-J., Makhijani V., Roth G.T., Gomes X., Tartaro K., Niazi F., Turcotte C.L., Irzyk G.P., Lupski J.R., Chinault C., Song X.-z., Liu Y., Yuan Y., Nazareth L., Qin X., Muzny D.M., Margulies M., Weinstock G.M., Gibbs R.A., Rothberg J.M. The complete genome of an individual by massively parallel DNA sequencing. *Nature*. 2008. V.452(7189). P.872-876. doi: 10.1038/nature06884
44. Yoshinaga Y., Daum C., He G., O'Malley R. Genome Sequencing. *Methods Mol. Biol.* 2018. V.1775. P.37-52. doi: 10.1007/978-1-4939-7804-5_4

References

1. Astbury W.T., Bell F.O. X-Ray study of thymonucleic acid. *Nature*. 1938. V. 141. P.747-748. doi:10.1038/141747b0
2. Avery O.T., MacLeod C.M., McCarty M. Studies of the chemical nature of the substance inducing transformation of pneumococcal types. Induction of transformation by a desoxyribonucleic acid fraction isolated from *Pneumococcus* Type III. *J. Exp. Med.* 1944. V.79(2). P.137-158. DOI: 10.1084/jem.79.2.137
3. Brenner S., Johnson M., Bridgham J., Golda G., Lloyd D.H., Johnson D., Luo S., McCurdy S., Foy M., Ewan M., Roth R., George D., Eletr S., Albrecht G., Vermaas E., Williams S.R., Moon K., Burcham T., Pallas M., DuBridge R.B., Kirchner J., Fearon K., Mao J.-i., Corcoran K. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbeads arrays. *Nat. Biotechnol.* 2000. V.18(6). P.630-634. doi: 10.1038/76469
4. Chargaff E. What really is DNA? Remarks on the changing aspects of a scientific concept // *Prog. Nucl. Acid Res. Mol. Biol.* 1968. V.8. P.297-333. 10.1016/s0079-6603(08)60549-8
5. Chemeris A.V., Akhunov E.D., Vakhitov V.A. DNA sequencing. Moscow. Nauka. 1999. 429 p.
6. Chen F., Dong M., Ge M., Zhu L., Ren L., Liu G., Mu R. The history and advances of reversible terminators used in new generations of sequencing technology // *Genomics, Proteomics, Bioinformatics*. 2013. V.11(1). P.34-40. doi: 10.1016/j.gpb.2013.01.003
7. Chin C.-S., Peluso P., Sedlazeck F.J., Nattestad M., Concepcion G.T., Clum A., Dunn C., O'Malley R., Figueroa-Balderas R., Morales-Cruz A., Cramer G.R., Delledonne M., Luo C., Ecker J.R., Cantu D., Rank D.R., Schatz M.C. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods*. 2016. V.13(12). P.1050-1054. doi: 10.1038/nmeth.4035
8. Deamer D., Akeson M., Branton D. Three decades of nanopore sequencing // *Nature Biotechnol.* 2016. V.34(5). P.518-524. doi: 10.1038/nbt.3423.
9. Droege M, Hill B. The Genome Sequencer FLX System--longer reads, more applications, straight forward bioinformatics and more complete data sets. *J Biotechnol.* 2008. V.136(1-2). P.3-10. doi: 10.1016/j.jbiotec.2008.03.021
10. Eid J., Fehr A., Gray J., Luong K., Lyle J., Otto G., Peluso P., Rank D., Baybayan P., Bettman B., Bibillo A., Bjornson K., Chaudhuri B., Christians F., Cicero R., Clark S., Dalal R., Dewinter A., Dixon J., Foquet M., Gaertner A., Hardenbol P., Heiner C., Hester K., Holden D., Kearns G., Kong X., Kuse R., Lacroix Y., Lin S., Lundquist P., Ma C., Marks P., Maxham M., Murphy D., Park I., Pham T., Phillips M., Roy J., Sebra R., Shen G., Sorenson J., Tomaney A., Travers K., Trulson M., Vieceli J., Wegener J., Wu D., Yang A., Zaccarin D., Zhao P., Zhong F., Korklach J., Turner S. Real-time DNA sequencing from single polymerase molecules // *Science*. 2009. V.323. P.133-138. doi: 10.1126/science.1162986
11. Esfandyarpour H., Davis R.W. An integrated differential nanocalimeter with on-chip microfluidic multiplexing for high throughput genomics and proteomics // 14th International Conference on Miniaturized Systems for Chemistry and Life Sciences, 3 - 7 October 2010, Groningen. P.1349-1351.
12. Esfandyarpour H., Zheng B., Pease R.F., Davis R.W. Structural optimization for heat detection of DNA thermosequencing platform using finite element analysis // *Biomicrofluidics*. 2008. V.2. P.24102. doi: 10.1063/1.2901138
13. Esfandyarpour H., Parizi K.B., Barmi M.R. <...> Dong B., Witney F.R., Davis R.W. High accuracy DNA sequencing on a small, scalable platform via electrical detection of single base incorporations. *bioRxiv preprint*. 2020. doi: 10.1101/604553
14. Fang G, Munera D, Friedman DI, Mandlik A, Chao MC, Banerjee O, Feng Z, Losic B, Mahajan MC,

- Jabado OJ, Deikus G, Clark TA, Luong K, Murray IA, Davis BM, Keren-Paz A, Chess A, Roberts RJ, Korlach J, Turner SW, Kumar V, Waldor MK, Schadt EE. Genome-wide mapping of methylated adenine residues in pathogenic *Escherichia coli* using single-molecule real-time sequencing // *Nat. Biotechnol.* 2012. V.30. P.1232-1239. doi: 10.1038/nbt.2432
15. Flusberg B.A., Webster D.R., Lee J.H., Travers K.J., Olivares E.C., Clark T.A., Korlach J., Turner S.W. Direct detection of DNA methylation during single-molecule, real-time sequencing // *Nat. Methods.* 2010. V.7. P.461-465. doi: 10.1038/nmeth.1459
 16. Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A., Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M., Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J., Nguyen D., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M., Tomb J-F., Dougherty B.A., Bott K.F., Hu P-C., Lucier T.S., Peterson S.N., Smith H.O., Hutchinson C.A., III, Venter J.C. The minimal gene complement of *Mycoplasma genitalium* // *Science.* 1995. V.270. P.397-403. DOI: 10.1126/science.270.5235.397
 17. Hyman E.D. A new method of sequencing DNA // *Anal. Biochem.* 1988. V.174. P.423-436. DOI: 10.1016/0003-2697(88)90041-3
 18. Jain M., Koren S., Miga K.H., Quick J., Rand A.C., Sasani T.A., Tyson J.R., Beggs A.D., Dilthey A.T., Fiddes I.T., Malla S., Marriott H., Nieto T., O'Grady J., Olsen H.E., Pedersen B.S., Rhie A., Richardson H., Quinlan A.R., Snutch T.P., Tee L., Paten B., Phillippy A.M., Simpson J.T., Loman N.J., Loose M. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat. Biotechnol.* 2018. V. 36(4). P.338-345. doi: 10.1038/nbt.4060
 19. Ivanov I., Niu L., Chueh A., Tian H., Deng S. Apparatus and methods for continuous diagnostics of macromolecules. *Pat. Appl. US 2018/0230531 A1.*
 20. Kasianovich J.J., Bezrukov S.M. On "three decades of nanopore sequencing" // *Nature Biotechnol.* 2016. V.34(5). P.481-482. doi: 10.1038/nbt.3570
 21. Kim D-R., Kim T-S., Kim E., Min S-J., Shin D., Ahn D-R. Synthesis of 3'-O-fluorescently mono-modified reversible terminators and their uses in sequencing-by-synthesis. *Bioorg Med Chem Lett.* 2014. V.24(1). P.209-213. doi: 10.1016/j.bmcl.2013.11.040
 22. Liu X., Milton J., Ruediger S. Labelled nucleotides. *Pat. US 7,795,424 B2.* (Sep 14, 2010)
 23. Mak S.S.T., Gopalakrishnan S., Carøe C., Geng C., Liu S., Sinding M-H.S., Kuderna L.F.K., Zhang W., Fu S., Vieira F.G., Germonpré M., Bocherens H., Fedorov S., Petersen B., Sicheritz-Pontén T., Marques-Bonet T., Zhang G., Jiang H., Gilbert M.T.P. Comparative performance of the BGISEQ-500 vs Illumina HiSeq2500 sequencing platforms for palaeogenomic sequencing. *Gigascience.* 2017. V.6(8). P.1-13. doi: 10.1093/gigascience/gix049
 24. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM. Genome sequencing in microfabricated high-density picolitre reactors // *Nature.* 2005. V.437. P.376-380. doi: 10.1038/nature03959
 25. Maxam A.M., Gilbert W. A new method for sequencing DNA // *Proc. Natl. Acad. Sci. USA.* 1977. V.74. P.560-564. doi: 10.1073/pnas.74.2.560
 26. Moore G.E. Cramming more components onto integrated circuits // *Electronics.* 1965. V. 38. P. 114-117.
 27. Purushothaman S., Toumazou C., Georgiou J. Towards fast solid state DNA sequencing // *Circuits and Systems.* 2002. V. 4. P.169-172. DOI: 10.1109/ISCAS.2002.1010416
 28. Purushothaman S., Toumazou C., Ou C.P. Protons and single nucleotide polymorphism detection: A simple use for the ion sensitive field effect transistor // *Sensors and Actuators.* 2006. V.114(2). P.964-968. DOI: 10.1016/j.snb.2005.06.069
 29. Pushkarev D., Neff N.F., Quake S.R. Single-molecule sequencing of an individual human genome. *Nat. Biotechnol.* 2009. V.27(9). P.847-850. doi: 10.1038/nbt.1561
 30. Rebrikov D.V., Korostin D.O., Shubina E.S., Il'inskii V.V. NGS: vysokoproizvoditel'noe sekvenirovanie. M.: Binom. Laboratoriya znanii. 2014. 232 S. [NGS: high-performance sequencing] (In Russian)
 31. Ronaghi M, Karamohamed S, Pettersson B, Uhlén M, Nyrén P. Real-time DNA sequencing using detection of pyrophosphate release. *Anal Biochem.* 1996 Nov 1;242(1):84-9. doi: 10.1006/abio.1996.0432.
 32. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, Milgrew MJ, Edwards M, Hoon J, Simons JF, Marran D, Myers JW, Davidson JF, Branting A, Nobile JR, Puc BP, Light D, Clark TA, Huber M, Branciforte JT, Stoner IB, Cawley SE, Lyons M, Fu Y, Homer N, Sedova M, Miao X, Reed B, Sabina J, Feierstein E, Schorn M, Alanjary M, Dimalanta E, Dressman D, Kasinskas R, Sokolsky T, Fianza JA,

- Namsaraev E, McKernan KJ, Williams A, Roth GT, Bustillo J. An integrated semiconductor device enabling non-optical genome sequencing // *Nature*. 2011. V.475. P.348-352. doi: 10.1038/nature10242
33. Sanger F., Nicklen S., Coulson A.R. DNA sequencing with chain-terminating inhibitors // *Proc. Natl. Acad. Sci. USA*. 1977. V.74. P.5463-5467. doi: 10.1073/pnas.74.12.5463.
 34. Shin D., Ahn D-R., Ahn H-C. 3'-O-fluorescently mono-modified nucleotides and uses thereof / *Pat. US 8,030,466*. Oct., 4, 2011
 35. Signer R., Caspersson T., Hammarsten E. Molecular shape of thymonucleic acid. *Nature*. 1938. V. 141. P.122. doi:10.1038/141122a0
 36. Toumazou C., Purushothaman S. Sensing apparatus and method / *GB Pat. Appl. № 0105831.2* (09.02.2001).
 37. Toumazou C., Shepherd L.M., Reed S.C., Chen G.I., Patel A., Garner D.M., Wang C-J.A., Ou C-P., Amin-Desai K., Athanasiou P., Bai H., Brizido I.M.Q., Caldwell B., Coomber-Alford D., Georgiou P., Jordan K.S., Joyce J.C., La Mura M., Morley D., Sathyavruthan S., Temelso S., Thomas R.E., Zhang L. Simultaneous DNA amplification and detection using a pH-sensing semiconductor system. *Nat. Methods*. 2013. V.10(7). P.641-646. doi: 10.1038/nmeth.2520
 38. Vasilenko S.K., Demushkin .P., Budovskii E.I., Knorre D.G. Opredelenie nukleotidnoi posledovatel'nosti v oligonukleotidakh // *Dokl. AN SSSR*. 1965. T.162. S. 694-697 [Determination of the nucleotide sequence in oligonucleotides] (In Russian)
 39. Watson J.D., Crick F.H.C. A structure for deoxyribose nucleic acid. *Nature*. 1953. V. 171(4356). P. 737-738. doi:10.1038/171737a0
 40. Wetterstrand K.A. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP)
 41. <https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data> [Accessed 02.02.2021].
 42. Wheeler D.A., Srinivasan M., Egholm M., Shen Y., Chen L., McGuire A., He W., Chen Y.-J., Makhijani V., Roth G.T., Gomes X., Tartaro K., Niazi F., Turcotte C.L., Irzyk G.P., Lupski J.R., Chinault C., Song X-z., Liu Y., Yuan Y., Nazareth L., Qin X., Muzny D.M., Margulies M., Weinstock G.M., Gibbs R.A., Rothberg J.M. The complete genome of an individual by massively parallel DNA sequencing. *Nature*. 2008. V.452(7189). P.872-876. doi: 10.1038/nature06884
 43. Yoshinaga Y., Daum C., He G., O'Malley R. Genome Sequencing. *Methods Mol. Biol.* 2018. V.1775. P.37-52. doi: 10.1007/978-1-4939-7804-5_4
 44. Zubov V.V. Rothberg sequencing (potentials for semiconductor sequencing). *Biomics*. 2013. V.5(1-2). P. 62-74.