**Департамент образования города Москвы**

**Государственное бюджетное общеобразовательное учреждение города Москвы «Школа № 1505 «Преображенская»**

**ДИПЛОМНОЕ ИССЛЕДОВАНИЕ**

на тему

**ИССЛЕДОВАНИЕ МЕТОДОВ СЕКВЕНИРОВАНИЯ ДНК**

Выполнила:

Пашинцева Анастасия Валерьевна, 10 класс

Руководитель:

Воробьева Екатерина Андреевна, учитель биологии

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ (подпись руководителя)

Рецензент:

Ноздрачева Анна Николаевна

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ (подпись рецензента)

**Москва,**

**2018/2019 уч.г.**

|  |  |
| --- | --- |
| ОГЛАВЛЕНИЕ |  |
| Введение……………………………………… | 4 |
| Глава 1. Обзор методов определения последовательности нуклеиновых кислот…………………………………… | 11 |
| 1.1 Методы, основанные на детекции сигнала от множества одинаковых молекул ДНК……………… | 12 |
| 1.1.1 Метод Максама-Гилберта, или метод химической деградации……………………………………  | 12 |
| 1.1.2 Метод Сенгера…………………………………………… | 14 |
| 1.1.3 Гибридизация на твердой фазе………………………………………………… | 17 |
| 1.1.4 Секвенирование лигированием……………… | 18 |
| 1.1.5 Пиросеквенирование……………………………........ | 19 |
| 1.1.6 Обратимые терминирующие нуклеотиды…..... | 20 |
| 1.1.7 Полупроводниковое секвенирование…………………… | 21 |
| 1.2 Методы, основанные на детекции сигнала от одной молекулы ДНК………………………………………………………………….. | 23 |
| 1.2.1 Секвенирование при помощи электронного микроскопа……………………………………………………….. | 23 |
| 1.2.2 Секвенирование путем протаскивания ДНК через нанопоры………………………………………………………….. | 25 |
| 1.2.3 Секвенирование методом спектроскопии комбинационного рассеяния…………………………………………………………… | 25 |
| 1.2.4 Использование обратимых терминирующих нуклеотидов на одиночных молекулах нуклеиновых кислот………………………………………………………………. | 26 |
| 1.2.5 Другие методы секвенирования………………………………… | 27 |
| Глава 2 Исследование коммерческих технологий секвенирования ДНК……………………………………………………………………… | 28 |
| Материалы и методы………………………………………………….. | 29 |
| Результаты……………………………………………………………. | 39 |
| Выводы…………………………………….…………………………. | 40 |
| Список литературных источников…………….................................... | 42 |
| Приложение …………………………………………………………. | 44 |

**Введение**

Во всех живых организмах есть органические вещества: белки, липиды, углеводы и нуклеиновые кислоты. Все они выполняют разные функции.

Главная функция белков — строительная. Строение организма определяется белками. Белки отвечают за транспорт веществ в высших многоклеточных организмах, регулируют обмен веществ, защищают организмы от чужеродных агентов. У многоклеточных животных есть сократительные белки актин и миозин, обеспечивающие сокращение мышц. Молекулы белков, встроенные в поверхностные мембраны клетки, обеспечивают сигнальную функцию.

Белки могут выступать и в качестве запасного вещества. Однако в организме животных белки, как правило, не запасаются. Есть и исключения: альбумин яиц, казеин молока. Другая сторона этой функции белков заключается в том, что благодаря белкам в организме могут откладываться в запас некоторые вещества. Например, при распаде гемоглобина железо не выводится из организма, а сохраняется, образуя комплекс с белком — ферритином.

Белки — ферменты ускоряют в клетках биохимические реакции. Когда другие источники энергии, то есть жиры и углеводы, израсходованы в качестве источника энергии организмом могут использоваться белки. При распаде 1 грамма белка до воды, углекислого газа и аммиака выделяется 17,6 кДж.

Выполнение этих функций обеспечивается разнообразным строением белков. Разнообразие в строении белков обеспечивается аминокислотным составом первичной структуры белка. В белках встречаются 20 аминокислот, их последовательность определяет структуру и свойства белков. Последовательность аминокислот в белке записана в ДНК (дезоксирибонуклеиновой кислоте).

Нуклеиновые кислоты отвечают за хранение и передачу наследственной информации, ее реализацию. В ДНК также записана информация о строении нуклеиновых кислот — РНК (рибонуклеиновых кислот).

Соединение из азотистого основания, пентозы и остатка фосфорной кислоты называется нуклеотидом. В качестве пентозы в ДНК выступает дезоксирибоза, в РНК – рибоза. Азотистое основание – гетероцикл, циклическая структура, в которой присутствуют атомы углерода, водорода, азота и функциональные группы. В состав нуклеиновых кислот входит пять оснований. В состав ДНК входят аденин, тимин, гуанин и цитозин. В состав РНК входят аденин, урацил, гуанин и цитозин.

Ген — это единица наследственной информации. Кроме того, ген — это участок ДНК, кодирующий информацию о первичной структуре одной полипептидной цепи. Также ген — это участок ДНК, кодирующий активную молекулу РНК (рибонуклеиновой кислоты) и отвечающий за синтез этой активной молекулы РНК.

Совокупность наследственного материала организма, называется геномом.

Правила перевода нуклеотидной последовательности в нуклеиновой кислоте в аминокислотную последовательность белка называют генетическим кодом. Он был расшифрован в 60-х годах XX века в результате ряда экспериментов и математических расчетов. Сначала в 50–60х годах XX века Георгий Гамов предположил, что кодоны состоят из трех нуклеотидов. Таким образом, можно получить 64 кодона, этого достаточно для 20 аминокислот, встречающихся в природе.

Позже, в 1961 году удалось подтвердить гипотезу Гамова экспериментально. Маршалл Ниренберг и Дж. Генрих Маттеи определили, что триплет UUU кодирует аминокислоту фенилаланин. После Хар Гобинд Корана доказал, что последовательность UCUCUCUCUCUC кодирует набор аминокислот серин-лейцин-серин-лейцин. К 1965 году генетический код был разгадан полностью. Если прочитать этот код, можно понять индивидуальные особенности организма. Если прочитать его у разных организмов, можно понять насколько близки разные этнические группы, как люди расселялись по планете, причины наследственных заболеваний. [10]

Секвенирование — это общее название методов, которые позволяют установить последовательность нуклеотидов в молекуле ДНК. Первое секвенирование человеческого генома продолжалась в течении 10 лет, в 2003 году эта задача была решена.

**Актуальность моего исследования:** с генетической точки зрения все люди одинаковы более чем на 99%, разница в менее чем 1% делает людей уникальными. Этим числом определяются наши физические возможности, предрасположенность к болезням и реакции на лекарства и т.д. Из-за этой разницы одно лекарство подходит одним пациентам, а другим – нет. Поэтому в будущем будет развиваться персональная медицина, основанная на данных генома пациента. [3]

Нобелевский лауреат Дан Шехтман в интервью заявил: «Еслибышкольник или совсем юный студент, избравший стезю учёного, спросил меня, какой наукой заниматься, я бы посоветовал молекулярную биологию. Именно её методы помогут решить большую часть наших проблем, избавить от самых тяжёлых заболеваний. Лекарства от рака — это то, что действительно нужно. Как и персонализированная медицина — лекарственные препараты, подобранные для каждого конкретного человека. Я думаю, что в этой области неизбежно случится взрыв технологий». [7]

Однако, чтобы узнать, как работает весь организм в целом знаний о геноме недостаточно. Геном у всех клеток одинаковый, есть небольшая разница из-за мутаций, также у половых клеток 23 хромосомы, но все клетки организма разные. Они различаются, потому что в каждой клетке активны, то есть экспрессируются, разные гены. Чтобы понять какие именно работают в том или ином органе или ткани, нужно изучить транскриптом – набор присутствующих там РНК. Экспрессия генов зависит от набора РНК в органе или ткани.

Обычно исследуют транскриптом именно органа или клеток определенного типа, потому что транскриптом характеризует протекание конкретных процессов и меняется со временем, в отличие от генома, чьи данные относительно постоянная характеристика. Однако совокупности знаний о геноме и транскриптоме также недостаточно. Более полную картину происходящего в нашем организме дает изучение протеома — набора белков, синтезируемых тканью, органом или отдельной клеткой.

Синтез белков — результат работы кодирующих их генов, но дальше судьба каждого белка складывается по-разному: кто-то распадется, даже не выполнив свои функции, кто-то накопится в клетке, кто-то модифицируется и т.д. Поэтому, чтобы понять, что происходит в нашем организме в целом изучают протеом.

Сейчас протеомика, как наука, находится в начальной фазе развития. Однако уже существует проект «Протеом человека». Инициаторы проекта Россия, США, Южная Корея, Швеция, Канада и Иран. По масштабам это проект можно сравнить с проектом «Геном человека». Цель проекта — систематизация всех производимых нашим организмом белков[3].

Зато благодаря секвенированию можно выяснить происхождение новых и уже известных видов организмов. Это помогает систематизировать знания о живых организмах, населяющих землю и проследить их происхождение.

В 2015 году были открыты локиархеи. Все их свойства, родственные связи и само их существование удалось вычислить на основе маленьких фрагментов ДНК, выделенных из донных осадков глубоководного гидротермального поля «замок Локи» в Северной Атлантике. Из проб воды, взятых у курильщика, донных остатков были выделены геномы.

Новые методы метагеномики позволили собрать из этих фрагментов большие участки геномов и даже почти полные геномы. Отсеквенировав, их геном, выяснили, что у этой группы архей много белков, ранее считавшимися характерными только для эукариот. Геномные данные указывают на более сложную организацию их клеток по сравнению с обычными прокариотами. Предки эукариот скорее всего были асгардархеямивозможно, близкими к хеймдалльархеям. Открытие показало, что движение в сторону усложнения организации было сделано предками эукариот задолго до приобретения митохондриального симбионта[6].

Метагеномика — это раздел молекулярной генетики, изучающий генетический материал, полученный из образцов окружающей среды. Метагеномика изучает набор генов всех микроорганизмов, находящихся в образце среды, — метагеном.

С помощью методов метагеномики можно определить видовое разнообразие исследуемого образца без необходимости выделения и культивирования микроорганизмов. Это позволяет исследовать и некультивируемые организмы. Развитие метагеномики обусловлено развитием методов секвенирования нового поколения. Они позволяют получить последовательности практически всех генов каждого микроорганизма сообщества. [11]

Сдав ДНК-тест и получив анализ результатов секвенирования своего экзома от врача-генетика можно выявить риски наследственных заболеваний, выбрать безопасные и эффективные для себя лекарства, разработать систему питания и тренировок. Генеалогический ДНК-тест позволяет выяснить происхождение ваших далеких предков и определить, какие национальности оставили след в вашем геноме. Можно отследить как расселялись по планете предки ваших родителей. [12]

Самки комаров Aedes aegypti заражают более 400 миллионов человек опасными вирусными инфекциями, такими как лихорадка денге, желтая лихорадка, Зика и чикунгунья. Изучение биологии комаров существенно затруднено из-за недостаточных знаний об их геноме и отсутствия его полной сборки. В Квинслендском институте медицинских исследований имени Клайва Бергофера под руководством специалиста по популяционной генетике Горданы Рашич недавно создали наиболее полную на текущий момент карту генома этих комаров AaegL5. Она позволит разработать новые биологические методы и стратегии вмешательства для борьбы с этим переносчиком смертельных заболеваний.

Сейчас в России работает проект «Российские геномы». Миссия проекта: «создание открытой компьютерной базы данных, содержащей анонимную информацию о полногеномных последовательностях по меньшей мере от 3000 мужчин и женщин из разных регионов России, чьи предки являются коренными жителями данного региона в нескольких поколениях, а также описание вариаций в геноме у этих групп, определение особенностей, влияющих на распространение заболеваний и создание информационной базы медицински-значимых геномных вариантов, характерных для населения России, что станет основой для разработки принципов медицины будущего»[8].

По инициативе биологов из разных стран был запущен проект «Earth BioGenome». Проект нацелен отсеквенировать геномы всех живых организмов, найти и классифицировать новых животных организмов. На данный момент, классификационная биология объединяет 0,2% всех эукариот, населяющих Землю. Считается, что новых организмов 15 миллионов видов. [9, 13]

В настоящее время на рынке появилось много разных коммерческих технологий секвенирования, однако все эти технологии дорогие, и в основном выбор исследовательских лабораторий останавливается на одной-двух технологиях. Сравнительный анализ технологий поможет лабораториям и людям, заинтересованным в данной теме, выбрать оптимальную технологию для себя.

**Проблема моего исследования:** коммерческие технологии секвенирования дороги, поэтому каждая лаборатория выбирает пару наиболее выгодных технологий. Однако в настоящее время существует большое количество приборов, и каждые несколько лет появляются новые. Это сильно затрудняет выбор.

**Цель моего исследования:** сравнить наиболее известные коммерческие технологии секвенирования и выявить оптимальный спектр применения.

**Задачи моего исследования:**

* выбор параметров для сравнения коммерческих технологий секвенирования ДНК;
* сравнительный анализ наиболее известных секвенаторов;
* выявление приборов, оптимальных для решения следующих задач:
* возможность секвенировать экзом человека по доступной цене, что нужно для клинических лабораторий;
* исследования в полевых условиях;
* секвенирования геномов *de novo*.
* выявление приборов со следующими характеристиками:
* большая длина прочтения (рида);
* высокая скорость работы;
* возможность чтения гомополимеров.

**Глава 1.**

**Обзор методов определения последовательности нуклеиновых кислот**

На сегодняшний день выделяют три поколения технологий секвенирования. К первому поколению относят метод Максама-Гилберта и метод Сенгера. Вторым поколением считают коммерческие технологии высокопроизводительного секвенирования, которые разработали в середине 1990-х. Все они требуют предварительного увеличения количества молекул ДНК.

На сегодняшний день появляются технологии секвенирования, которые позволяют проводить секвенирование одной молекулы нуклеиновой кислоты. Такие подходы называют секвенированием третьего поколения. Однако, некоторые авторы второе и третье поколения секвенирования объединяют под терминами «NGS» и «высокопроизводительное секвенирование» как равнозначными.

«Методы определения нуклеотидной последовательности в РНК пока недостаточно эффективны – технологии секвенирования одиночных молекул, позволяющие работать непосредственно с РНК, только начали появляться».
Несмотря на это, секвенирование РНК тоже проводится. Розер Венто-Тормо из Института Сенгера с коллегами, интересуясь иммунологическим парадоксом беременности провели секвенирование РНК одиночных клеток, выделенных из плаценты и децидуальной ткани, сопоставив данные с клетками крови матери. На основе полученных данных был составлен молекулярный атлас, являющийся впечатляющим ресурсом для будущих исследований беременности и её осложнений.

«В тоже время сейчас для определения последовательности РНК исследователи чаще используют секвенирование комплементарной ДНК, потому что процесс обратной транскрипции достаточно стандартен и консервативен.» [1]

**1.1 Методы, основанные на детекции сигнала от множества одинаковых молекул ДНК (методы с предварительной амплификацией1 фрагментов ДНК)**

Первое и второе поколения технологий определения нуклеотидной последовательности требуют использования множества идентичных молекул ДНК для усиления сигнала. [1]

**1.1.1 Метод Максама-Гилберта, или метод химической деградации**

В 1977 году Аллан Максам и Уолтер Гилберт разработали метод секвенирования ДНК, в основе которого лежит специфичность химической деградации нуклеотидов при обработке нуклеиновой кислоты различными химическими агентами3. Сначала, у фрагмента ДНК, длиной в 100-1000 н. п. один конец цепи помечают радиоактивной меткой.

На втором этапе образец разделяют еще на четыре фрагмента. Каждая часть обрабатывается определенным реагентом, который приводит ДНК к гидролизу по конкретным основаниям или их сочетаниям. Параметры каждой из четырех реакций подбираются таким образом, что изменения проходят по некоторым позициям в каждой молекуле ДНК. Получается смесь расщепленных фрагментов ДНК, которые соответствуют по длине локализации определяемых оснований. [14] Смесь гидролизируют, после этого с четырьмя частями нуклеиновой кислоты проводят электрофорез таким образом, чтобы в результате получить фрагменты, отличающиеся на одно основание. На следующем этапе получают электрофореграмму. По ней устанавливают последовательность оснований в исследуемом фрагменте ДНК. Этот метод позволяет определить максимально 200 нуклеотидов за один рид (прочтение).



**Рис.1.1.** Принцип метода Максама-Гилберта. Расщепление одинаковых, помеченных с одного из концов, фрагментов ДНК по разным позициям дает фрагменты разной длины, затем фрагменты могут быть разделены при помощи гель-электрофореза

До последнего времени этот метод использовали, когда фермент ДНК-полимераза, который используется в методе Сенгера, не мог пройти через вторичную структуру биомолекулы. Однако, в настоящее время метод практически не используется ввиду сложности подготовки образцов ДНК и работы с вредными для здоровья химическими реагентами. Преимуществами метода Максама-Гилберта, по сравнению с методом Сенгера, являются полная его независимость от вторичных структур и отсутствия необходимости знания участка последовательности, интересующей ДНК, что позволяет пропустить стадию клонирования. [1]

**1.1.2 Метод Сенгера (остановка синтеза ДНК ферментом на дидезоксинуклеотидах)**

В 1975 году Фредерик Сенгер и Алан Кулзон из лаборатории молекулярной биологии в Кембридже предложили метод определения последовательности нуклеотидов в ДНК, основанный на использовании радиоактивно меченных нуклеотидов и ДНК-полимеразы, который получил название «плюс-минус секвенирование». Через два года Сенгер усовершенствовал этот метод и создал новый – метод дидезокситерминаторов, который впоследствии получил название «метод Сенгера». Через три года, в 1980 году Фредерик Сенгер за эту работу был удостоен Нобелевской премии по химии, которую разделил с Уолтером Гилбертом. [1]

Основа метода – использование модифицированных «нуклеотидов» - дидезоксинуклеозидтрифосфатов (ддНТФ). Отличие ддНТФ от обычных дезоксинуклеозидтрифосфатов (дНТФ) в том, что они не несут OH-группу в 3’-положении дезоксирибозы.



**Рис.1.2.** Структурные формулы дезоксирибозы и дидезоксирибозы

Вследствие этих особенностей в строении ддНТФ, они не способны к присоединению полимеразой следующего нуклеотида. Исследуемый участок ДНК добавляют в реакцию, добавляется в реакцию, по условиям схожую с обычной полимеразной цепной реакцией (ПЦР) в пробирке находятся: термостабильная ДНК-полимераза, дНТФ всех четырех типов, и олигонуклеотид, выступающий в качестве затравки для синтеза новой цепи ДНК. Помимо этих компонентов, в концентрации примерно в двадцать раз меньшей чем концентрация дНТФ, присутствуют четыре соответствующих ддНТФ, меченых каждый своим флуоресцентным красителем (до применения флуорофоров в качестве меток использовали изотопы, обычно P32).

В ходе реакции мечения - ферментативного синтеза ДНК в каком-то положении случайным образом происходит включение в строящуюся цепь вместо дНТФ меченого ддНТФ, что приводит к остановке синтеза. [2] Реакцию проводят циклически многократно, повторяя синтез ферментом новых одноцепочечных фрагментов. Поскольку ддНТФ составляет 1/20, или 5% от дНТФ, а мечение включает 40-50 циклов реакции, то в конце такой линейной амплификации получается набор одиночных цепей ДНК, отличающихся по длине и заканчивающихся меченым нуклеотидом. После проведения реакции мечения проводят разделение полученных одноцепочечных фрагментов методом электрофореза в геле.

Чтобы сделать процесс проведения электрофореза удобнее, были созданы автоматические секвенаторы. Они проводят разделение флуоресцентно-меченых фрагментов ДНК в тонком капилляре, заполненном гелем.



**Рис.1.3.** Современный капиллярный автоматический секвенатор для определения последовательности нуклеиновых кислот методом Сенгера (ABI 3130xl, Applied Biosystems)

Детекция разделенных фрагментов происходит на дальнем конце капилляра за счет регистрации флуоресценции ддНТФ у проходящих через детектор молекул. В зависимости от типа дидезоксинуклеотидов прибор регистрирует свой спектр флуоресценции прибор регистрирует свой спектр флуоресценции. Анализ данных капиллярного секвенирования сводится к «прочтению» последовательных пиков флуоресценции. В настоящее время с использованием современных автоматических секвенаторов длина одного прочтения – рида – по методу Сенгера составляет 800 –1000 нуклеотидов. [1]

Изобретение автоматических секвенаторов упростило и удешевило секвенирование ДНК настолько, что к середине 1980-х годов стало возможно говорить о возможности определения полной последовательности генома человека. Это вылилось в крупное исследование – Human Genome Project (HGP), в которое были вовлечены США, Великобритания, Франция, Япония, Китай и Германия.

Целями проекта являлись определение последовательности нуклеотидов в человеческом геноме и, соответственно, идентификация всех генов человека. Исследования начались в 1990 году и финишировали в 2001 году публикацией в журнале Nature. Однако более полный анализ полученных данных был закончен только в 2003 году. [15]

Чтобы определить полную последовательность генома человека весь геном разделили на фрагменты по 150 000 пар нуклеотидов, которые вставили в искусственные бактериальные хромосомы (BAC), для каждой такой вставки7 было определено его расположение на хромосоме.

Вставку каждой BAC секвенировали методом дробовика (shotgun sequencing), для чего фрагмент генома из BAC расщепляли на более короткие фрагменты – 2000-3000 пар нуклеотидов, каждый из которых субклонировали8, то есть переносили часть уже клонированной молекулы ДНК в другой клонирующий вектор9, а именно в бактериальный и определяли последовательность фрагмента методом Сенгера.

Ввиду случайности фрагментации получаемые последовательности частично перекрывались, что позволяло при помощи специального программного обеспечения выбирать из отдельных прочтений BAC-вставки, а затем и целые хромосомы. [1]

Параллельно с HGP в 1998 году секвенированием генома человека занялась компания Celera под управлением Крейга Вентера. Последовательность в 3 млрд пар нуклеотидов ее сотрудниками была получена в течение 9 месяцев в 20 раз быстрее, чем участниками HGP. Это связано с тем, что Celera стартовала на автоматических секвенаторах последнего поколения, выигрывающих по производительности.

Кроме того, компания не применяла клонирование генома в BAC, а использовала метод дробовика для всего генома. Бюджет международного проекта составлял более 3 миллиардов долларов, а бюджет проекта от «Celera Genomics» всего 300 миллионов. Кроме того, в отличии от международного проекта, Крейг Вентер и «Celera Genomics» не собирались открывать доступ к полученным данным своего исследования. [15]

**1.1.3 Гибридизация на твердой фазе (принцип комплементарности цепей ДНК)**

В конце 1980-х годов был предложен подход к определению последовательности ДНК, получивший название секвенирование путем гибридизации (sequencing by hybridization, SBH), или секвенирование на чипе. Основа метода – гибридизация меченой одноцепочечной ДНК, разрушенной до коротких фрагментов, с синтетическими олигонуклеотидами известной структуры и определенной длины, точечно расположенными на подложке. На подложке присутствуют все возможные варианты последовательности олигонуклеотида данной длины.

Условия гибридизации подбираются так, что только полностью комплементарные фрагменты ДНК вступают в взаимодействие с олигонуклеотидом на подложке. Таким образом, после удаления не связавшихся молекул ДНК можно зарегистрировать сигнал в тех позициях чипа, где находится олигонуклеотид, последовательность которого есть в секвенируемом образце ДНК. Полученный гибридизацией паттерн можно использовать для восстановления исходной последовательности путем сборки перекрывающихся участков сработавших проб. [2]

Из-за низкой дискриминирующей способности гибридизационного подхода – невозможно подобрать условия, в которых к олигонуклеотидам «прилипали» бы только полностью комплементарные фрагменты, всегда найдутся GC-богатые участки, которые будут гибридизоваться и при наличии одного или даже нескольких неспаренных оснований – метод SBH пока не нашел практического применения в секвенировании ДНК.

Однако алгоритмы, разработанные авторами этого метода для сборки коротких прочтений в более длинные, стали основой для следующих алгоритмов высокоскоростной сборки и выравнивания. [1]

**1.1.4 Секвенирование лигированием**

Современные методы секвенирования лигированием основаны на использовании коллекции коротких, как правило, 8-10 оснований, флуоресцентно-меченых с помощью четырех красителей, вырожденных олигонуклеотидов, так, что каждому флуорофору соответствует определенный нуклеотид, или два нуклеотида, в определенной позиции.

На первом этапе создают иммобилизованную на твердой фазе клональную библиотеку одноцепочечных фрагментов ДНК. Секвенирование начинается с отжига праймера, комплементарного адаптеру на одном из концов библиотеки ДНК. После чего к библиотеке добавляют флуоресцентно-меченные вырожденные олигонуклеотиды и проводят реакцию лигирования, что приводит к фиксированию олигонуклеотида на фрагменте в случае его полного соответствия. Затем считывают флуоресценцию, определяя тем самым, какой нуклеотид, или пара нуклеотидов находятся в определенной позиции. Флуорофор удаляют и лигируют следующий олигонуклеотид. Всего проводят 10-15 последовательных лигирований. Затем выполняют «перезагрузку», отсоединяя праймер лигированными мечеными олигонуклеотидами и повторяют цикл с другим праймером со сдвигом на букву.

Этот принцип в настоящее время используют коммерческие технологии Polonator (Dover/Harvard) и SOLiD (Life Technologies Thermo Fisher Scientific). [1]

**1.1.5 Пиросеквенирование (регистрация акта присоединения нуклеотида по образующемуся пирофосфату)**

В 1996 году специалисты Королевского технологического института в Стокгольме Мустафа Рональди и Пол Нирен опубликовали подход к секвенированию ДНК, в основе которого лежит принцип регистрации пирофосфата, возникающего в результате присоединения очередного нуклеотида ДНК-полимеразой. Для детекции выделяющегося в процессе образования фосфодиэфирной связи пирофосфата авторы предложили использовать каскад последовательных химических реакций, заканчивающийся высвечиванием кванта света. Этот метод позволяет определять нуклеотидную последовательность фрагментов геномной ДНК размером 300–500 пар оснований. [16, 1]

Сначала любым методом создается иммобилизованная клональная библиотека10 одноцепочечных фрагментов ДНК на твердой фазе. Предварительно ко всем фрагментам нуклеиновой кислоты присоединяют адаптер11, на который будет гибридизоваться праймер12. Праймер будет служить затравкой для синтеза ДНК-полимеразой комплементарной цепи. [1]

Дальнейшая реакция состоит из последовательных циклов, в процессе которых к исходной ДНК по очереди добавляют дНТФ всех четырех типов. В случае если на секвенируемой цепи ДНК имеется комплементарный данному нуклеотид, в процессе формирования ДНК-полимеразой фосфодиэфирной связи побочным продуктом реакции станет пирофосфат.

Он активирует каскад химических реакций, в результате которых возникает световой сигнал, интенсивность которого прямо пропорциональна числу включенных в цепь нуклеотидов. Если подряд идут несколько одинаковых нуклеотидов, сигнал будет ярче.

В ячейке присутствуют АТФ-сульфурилаза, люцифераза и апираза, осуществляющие ферментативные реакции, а также вместе с ними присутствуют аденозинсульфофосфат (APS) и люциферин (стехиометрически выделившийся пирофосфат вместе с АСФ при помощи АТФ-сульфурилазы образует АТФ, являющийся источником энергии для проведения люциферазой реакции окисления люциферина в оксилюциферин, в процессе которой генерируется свет в видимом спектре в количестве, пропорциональном количеству включенных нуклеотидов.

Световой сигнал обычно детектируется ПЗС-матрицей, аналогичной матрице, встроенной в фотоаппарат, и анализируется с помощью програмного обеспечения, преобразующего пирограмму в последовательность нуклеотидов. Нуклеотиды, которые не вовлечены в синтез новой цепи, и АТФ деградируются с помощью апиразы. После этого можно начинать следующий цикл – добавлять другой тип нуклеотида.

Позднее авторы этого метода усовершенствовали технологию, предложив сочетание пиросеквенирования с эмульсионной полимеразной цепной реакцией (эПЦР). эПЦР проводится в капельках масла, содержащих смесь для осуществления ПЦР, которая и проходит отдельно в каждой капельке эмульсии. Капелька эмульсии с ДНК называется микросфера. На принципе пиросеквенирования основана коммерческая технология 454 Life Sciences компании Roche. [1]

**1.1.6 Обратимые терминирующие нуклеотиды**

Идею определения нуклеотидной последовательности в ДНК синтезом предложили сотрудники химического факультета Кембриджского университета Баласубрамян и Кленерман. Этот метод похож на предыдущий, но, в отличие от пиросеквенирования, регистрация присоединенного нуклеотида происходит непосредственно, по сигналу от присоединенного основания. [16]

В этом методе за один цикл реакции может быть добавлен только один нуклеотид, для этого в реакционную смесь добавляют 3’-блокированные нуклеозидтрифосфаты, с возможностью его снятия. Блок обычно обеспечивается метилазидным блокатором (AZM). Второе условие для использования этого метода: метка нуклеотида должна быть отщепляемой.[1]

Сначала создается иммобилизованная клональная библиотека с одноцепочечными фрагментами ДНК. Процесс секвенирования начинается с отжига праймера, который комплементарен адаптеру на одном конце библиотеки.

На втором этапе к библиотеке добавляют четыре типа флуоресцентно-меченых обратимых терминирующих нуклеозидтрифосфатов, или RT-оснований. ДНК-полимераза присоединяет подходящий нуклеотид к фрагменту ДНК из библиотеки, на этом моменте синтез временно останавливается, так как трифосфаты блокированы.

Нуклеотиды, которые не включились в синтезируемую цепь ДНК, смывают. После этого оптическая система считывает флуоресценцию от каждой ДНК-колонии библиотеки, каждая из которых флуоресцирует соответственно нуклеотиду, который включился в эту колонию. На следующем этапе флуофор и блокатор химическим способом удаляют из синтезируемой цепи, это дает возможность повторить цикл сначала. [16]

В этом методе, в отличие от пиросеквенирования, возможно регистрировать сигнал от присоединенного нуклеотида в течение длительного времени.

На таком принципе основаны коммерческие технологии компаний Illumina и Pacific Bioscience. [1]

**1.1.7 Полупроводниковое секвенирование**

Полупроводниковое секвенирование – это метод определения последовательности ДНК, основанный на регистрации ионов водорода, выделяющихся при синтезе ДНК ферментом.

Сначала любым методом создается иммобилизованная клональная библиотека одноцепочечных фрагментов ДНК на твердой фазе, таким образом, чтобы метод создания позволял отделить каждую колонию ДНК от других так, чтобы выравнивание pH, в случае его изменения в районе колонии, происходило достаточно медленно. При использовании эПЦР это обеспечивается закатыванием микросфер в микрореакторы, в каждом реакторе помещается только одна частица. Реакторы сообщаются только одной поверхностью на проточном чипе. [1]

Секвенирование начинают с отжига праймера, который комплементарен адаптеру на одном из концов библиотеки ДНК. Затем к микрореакторам с микросферами по очереди добавляют обычные нуклеотиды. Если добавленный нуклеотид комплементарен матрице, ДНК-полимераза встраивает его в синтезируемую цепь. В результате этой реакции выделяется пирофосфат и протон, вызывающий локальное изменение водородного показателя раствора в микрореакторе, которое детектируется сенсором, подключенным к каждому микрореактору.

Если нуклеотид не комплементарен матрице, то сигнал отсутствует. После каждого добавленного нуклеотида прибор промывает систему от остатков, не включившихся дНТФ данного типа. У полупроводникового секвенирования имеются трудности определением гомополимерных участков – в случае протяженного мононуклеотида, например, ТТТТТТТ, сигнал теряет дискретность, определение количества нуклеотидов становится трудным. [16]

Важным отличием полупроводникового секвенирования от пиросеквенирования и других методов является отсутствие оптического детектора сигнала. Это значительно упрощает и удешевляет конструкцию прибора.

Оптическая детекция осуществляется встроенным в секвенатор микроскопом и имеет ограничения, которые связаны с полем зрения этого микроскопа и его разрешающей способностью.

Детекция на полупроводниковом чипе такого ограничения практически не имеет. [1]

На этом методе основана коммерческая технология Ion Torrent от Life Technologies Thermo Fisher Scientific. [16]

**1.2 Методы, основанные на детекции сигнала от одной молекулы ДНК (секвенирование одиночных молекул ДНК)**

Эти методы отличаются отсутствием этапа клональной амплификации библиотеки или фрагмента ДНК. Это упрощает методику, а главное, убирает один из этапов искажения исходного материала – в перечисленных выше методах определяется последовательность не исходной ДНК, а примерно ее 20-ой копии. С другой стороны, регистрация сигнала от одной молекулы нуклеиновой кислоты накладывает высокие требования к соответствующим детекторам.

Ряд авторов называют методы, описанные ниже, технологиями третьего поколения. [1]

**1.2.1 Секвенирование при помощи электронного микроскопа**

В конце 50-х годов ХХ века Ричард Фейнман предложил метод секвенирования, предполагающий использование электронного микроскопа. В электронной микроскопии можно выделить три технологических разновидности: сканирующая электронная микроскопия (СЭМ, SEM), просвечивающая электронная микроскопия (ПЭМ, TEM), и сканирующая просвечивающая электронная микроскопия (СПЭМ, STEM).

В двух последующих десятилетиях методы ПЭМ активно разрабатывались, тогда же разрабатывали и подходы для определения последовательности в ДНК. В 1970 году Альберт Крю предложил метод визуализации в сканирующем электронном микроскопе (high-angle annular dark-field imaging, HAADF).

Использование этой техники дает возможность обнаружить отдельные тяжелые атомы на тонких пленках аморфного углерода. Визуализировать отдельные основания в ДНК можно, пометив их атомами тяжелых металлов. Методом не пользовались и не пользуются на практике, ввиду быстрого разрушения молекулы ДНК пучком электронов.

В 1990-х годах получила широкое распространение технология секвенирования с помощью метода сканирующей туннельной микроскопии, однако возможность воспроизвести результаты секвенирования была достаточно низкой, в какой-то момент принимать к публикации исследования перестали.

В 2009 году японские исследователи сообщили, что им удалось на отдельной одноцепочечной молекуле ДНК визуализировать гуанин по геометрическим характеристикам.



**Рис.1.4** Визуализация гуаниновых нуклеотидов на цепи ДНК посредством возможностей сканирующей туннельной микроскопии

В 2010 году Криванек с коллегами предложили усовершенствование метода ПЭМ, это позволило видеть одиночные замены атомов в монослое нитрида бора.

Теоретически электронная микроскопия может обеспечить чрезвычайно длинные прочтения, это важно для сборки эукариотических геномов, в основном состоящих из повторяющихся последовательностей. В 2012 году коллектив ученых из Гарвардского университета, Университета штата Нью-Гемпшир и компании ZS Genetics продемонстрировал возможность прочтения длинных последовательностей при помощи электронной микроскопии, однако подобные технологии пока далеки от широкого коммерческого применения. [1]

**1.2.2 Секвенирование путем протаскивания ДНК через нанопоры**

Идея этого метода была предложена Касьяновичем в 1996 году. [1] Принцип метода заключается в регистрации изменений ионного тока, вызванных прохождением одноцепочечной ДНК через нанопору в тонкой пленке под действием электрического поля.

Поры могут быть биологическими, например, могут использовать клеточную мембрану с порой, а могут быть искусственными. В качестве искусственных используют сенсоры, фиксирующие изменения какой-либо характеристики, например, туннельного тока, или ионного тока. При проходе через нанопору каждый тип азотистых оснований по-своему «закрывает» пору и влияет на ток. [4]

Предварительные результаты показали возможность различать длинные гомополимерные отрезки, например, 30 аденинов за ними 70 цитозинов. Однако короткое время прохождения основания сквозь пору и тепловые колебания пока не позволяют определять количество стоящих подряд однотипных нуклеотидов. [1]

Дальше всех в разработке технологий секвенирования путем протаскивания ДНК через нанопоры продвинулась компания Nanopore, заявившая выпуск секвенатора в 2013 году, но на начало 2014 года коммерческого варианта технологии на рынке не появилось. Разработкой занимаются также и другие компании.

Основная проблема – высокая чувствительность к факторам внешней среды, в результате чего происходит разрыв билипидного слоя, в котором находятся поры.[1]

**1.2.3 Секвенирование методом спектроскопии комбинационного рассеяния**

Комбинационное рассеяние (эффект Рамана) – неупругое рассеяние оптического излучения на молекулах вещества, сопровождающееся заметным изменением частоты излучения. В случае эффекта Рамана в спектре рассеянного излучения начинают появляться спектральные линии, которых нет в спектре первичного, возбуждающего света. Число и расположение этих линий определяется молекулярным строением вещества. Спектроскопия комбинационного рассеяния света, или рамановская спектроскопия – эффективный метод химического анализа, изучения состава и строения веществ.

В 2007 году немецкий исследователь Волкер Дэкерт показал, что с помощью этой методики, или методики TERS (tip-enhanced raman spectroscopy) можно, пройдя по цепочке РНК, определить по спектрам, какие нуклеотиды находятся в цепи. Золото и серебро усиливают сигнал комбинационного рассеяния, и в результате получается спектр, который достаточно легко можно детектировать. По усиленному таким образом спектру удается расшифровать последовательность нуклеотидов со всеми модификациями, так как спектр содержит необходимую информацию. Через 5 лет после своей первой работы на одной из конференций Дэкерт показал работу, где он смог различить отдельные модификации на молекуле ДНК.

Сейчас работы в этом направлении ведутся крупнейшими компаниями, такими как Intel, IBM, HP. Intel пытается секвенировать ДНК с помощью разбиения ее на отдельные нуклеотиды и последующим получением сигнала с наночастиц серебра. [1]

**1.2.4 Использование обратимых терминирующих нуклеотидов на одиночных молекулах нуклеиновых кислот (регистрация каждого присоединенного нуклеотида по отщепляемой метке)**

Один из вариантов был предложен исследователями Корнелльского университета в 2003 году, они разработали технологию оптической детекции включения флуоресцентно-меченых нуклеотидов в строящуюся цепь одной молекулы ДНК. Используя принципы ближнего поля с апертурой в несколько нанометров, становится возможно регистрировать флуоресценцию единственного флуорофора.

Для секвенирования ДНК-полимеразу прикрепляют к твердой фазе, затем добавляют в реакцию матрицу и специальные флуоресцентно-меченые нуклеотиды, каждый из которых помечен своим светом. В процессе синтеза комплементарной цепи ДНК при образовании фосфодиэфирной связи происходит отщепление флуорофора от вновь присоединенного основания. Прибор фиксирует флуоресценцию и длительность.

Преимущества этого метода в возможности читать очень длинные фрагменты ДНК – десятки тысяч нуклеотидов – и теоретическая возможность определять модифицированные азотистые основания.

Компания Helicos Biosciences предложила другой метод, но он по сути отличается лишь принципом регистрации сигнала: компания использует четыре лазера и систему, сходную с конфокальным микроскопом.

На этом методе секвенирования основаны технологии PacBio (Pacific Bioscience) и HeliScope (Helicos Biosciences). [1]

**1.2.5 Другие методы секвенирования**

Кроме выше перечисленных подходов есть множество менее технически и идейно проработанных методов: метод колебаний, секвенирование при помощи вращающегося поля и т.д.

**Глава 2.**

**Исследование коммерческих технологий секвенирования ДНК**

**Материалы и методы**

Для сравнения были выдвинуты признаки, важные для исследовательских лабораторий:

* вид ПЦР, либо ее отсутствие;
* метод секвенирования, лежащий в основе;
* длина прочтения;
* продолжительность одного запуска;
* цена за миллион нуклеотидных пар;
* цена оборудования.

Объектами исследования являлись секвенаторы компании Roche 454 — Genome Sequencer и 454 Genome Sequencer Junior, компании SOLiD — SOLiD 5500, компании Illumina — HiSeq 2500 и MiSeq, компании Ion Torrent — Ion PGM, компании Pacific Biosciences — PacBio RSII, компании Helicos Biosciences — HeliScope и компании Oxford Nanopore Technologies — MinION, GridION X5, PromethION и SmidgION.

**Технология 454 Life Sciences компании Roche**

В 2005 году вышел первый широкодоступный автоматический секвенатор второго поколения, базировавшийся технологии 454 Life Sciences. Данная технология основана на принципе пиросеквенирования. Технология была разработана подразделением компании CuraGen Corporation, которое в 2007 году приобрела компания Roche Diagnostics.

На сегодняшний день Roche Diagnostics предоставляет две модели секвенаторов, базирующиеся на этой технологии: Genome Sequencer FLX и Genome Sequencer Junior.

С помощью первого прибора можно получать до 1 млн прочтений с длиной e до 800­­­ –1000 п. н. (пар нуклеотидов), второй – до 100 000 прочтений с длиной 600 – 800 п. н. Длина прочтений важное преимущество 454 Life Sciences в сравнении с другими технологиями, за исключением технологии PacBio RS).

Технология дает возможность достаточно просто решать некоторые специфические задачи, например, определять гаплотипы HLA при скрининговых исследованиях или метагеномные исследования микробиомов на основе анализа 16S рибосомальной РНК. [1]

Подготовка образцов проводится стандартным образом – исследуемый образец нуклеиновой кислоты расщепляют. Затем к фрагментам длиной около 1000 п. н. присоединяют разные по последовательности адаптеры и проводят амплификацию фрагментов, методом эПЦР. Полученные на второй стадии сферы с однотипными молекулами ДНК помещают в матрицу с микрореакторами таким образом, что в каждую ячейку попадает только одна микросфера. Матрицу размещают в секвенаторе, который осуществляет последовательную подачу всех типов нуклеотидов и детекцию уровня ослабления света при его распространении в пространстве от каждой ячейки. Свет проходит по световоду на ПЗС-матрицу.

Время прохождения цикла – добавление нуклеотида, каскад химических реакций, детекция света, удаление остатков дНТФ – определяется скоростью доставки реагентов до каждой ячейки. С учетом количества циклов получается 10 часов работы.

Технология 454 Life Sciences имеет проблемы с чтением гомополимеров – от 8 до 10 повторяющихся нуклеотидов приводят к ошибке в подсчете точного числа оснований в повторе.

По производительности и стоимости определения одного нуклеотида эта технология отлично подходит в случаях, когда важна длина одного прочтения.

Однако, с 2016 года компания Roche снимает с производства секвенаторы на основе этой технологии. [1]

**Технология SOLiD компании Life Technologies Thermo Fisher Scientific.**

Технология SOLiD (sequencing by oligonucleotide ligation and detection) основана на методе, разработанном исследователями Гарвардского университета. На данный момент эта технология принадлежит компании Thermo Fisher Scientific. Сейчас компания предлагает платформу сразу в двух вариантах: SOLiD 5500 (на один чип) и SOLiD 5500xl (на два чипа). Существует вариант SOLiD 5500 Wildfire, в которой вместо эмульсионной ПЦР применяется мостиковая, но этот вариант платформы не получил широкого распространения.

В отличие от большинства других платформ технология SOLiD использует принцип секвенирования с помощью лигирования, без применения ДНК-полимеразы. Это дает возможность прочтения цепи ДНК в любом направлении, в отличие от методов на основе ДНК-полимераз, которые могут «читать» только в направлении растущего 3’-конца затравки.

На первом этапе работы методом эПЦР создают иммобилизованную библиотеку фрагментов ДНК, таким образом, что каждая частица несет множество одинаковых фрагментов ДНК. Микрочастицы впрыскивают в плоскую прозрачную камеру. [1]

Камера похожа на расположенные на расстоянии нескольких микрометров два предметных стекла. Камеру называют проточной, на английском языке flow cell или flow chip, поскольку с одной стороны в нее впрыскивать реагент, с другой стороны он вытекает. Микрочастицы должны распределиться одним слоем между стеклами.

Метод секвенирования по технологии SOLiD основывается на использовании флуоресцентно-меченных олигонуклеотидов (проб) с двумя специфичными основаниями с 3’-конца, остальные позиции вырождены. Всего существует 16 типов проб, по 4 каждого цвета.

Процесс секвенирования начинают с отжига праймера, комплементарного адаптеру на одном из концов библиотеки. После в проточную камеру добавляют все 16 типов проб, которые присоединяются на фрагменты ДНК по принципу комплементарности. Затем проводится отжиг проб, а лигаза зашивает разрыв между пробой и праймером, если проба отожглась в плотную к праймеру. Пробы, которые не смогли связаться, смывают. У связавшихся проб прибор детектирует флуоресценцию.

После три 5’-концевых нуклеотида пробы вместе с флуорофорами химически отсоединяют и удаляют из проточной камеры. В результате в камере остается только часть с известным динуклеотидом и тремя случайными основаниями, а на 5’-конце появляется место для присоединения новой пробы.

По прошествии 10–15 последовательных лигирований производится перезагрузка праймера – новосинтезированную цепь отплавляют от матрицы, на ней отжигают праймер, отличающийся по длине от первого на один нуклеотид. Перезагрузки проводятся 4 раза. В результате каждое основание исходной цепи гарантировано прочитывается дважды в разных пробах. Эта технология достаточно производительна, но сильно уступает по длине прочтения Illumina: 75 нуклеотидов у SOLiD против 250 у Illumina.

Секвенирование по технологии SOLiD довольно дорогостоящее, сложное и длительное, поэтому можно спрогнозировать ее скорый уход с рынка. [1]

**Технология Illumina Genome Analyser компании Illumina**

Эта технология, базируется на методе секвенирования синтезом, разработанном в 1998 году в Кембриджском университете Баласубраманяном и Кленерманом. [16] Одновременно с разработкой метода они основали компанию Solexa, задавшись целью коммерциализировать изобретенный метод. В 2004 году компания Solexa приобрела компанию Manteia Predictive Medicine и получила технологию параллельного секвенирования на основе ДНК-колоний, которые получают, используя мостиковую ПЦР. В этом методе праймеры прикрепляются к подложке, отжигающиеся на них молекулы ДНК нарабатывает ДНК-полимераза. [1]

В 2006 году компания Solexa выпустила свой первый коммерческий секвенатор – Genome Analyzer. За четыре дня работы – один запуск – прибор давал 1 млрд п. н. Через год компанию Solexa приобрела компания Illumina, и через четыре года был выпущен новый вариант секвенатора — HiSeq 2500, дающий по 25 Гбайт данных в день, или 200 Гбайт данных за один день. Этот прибор дает возможность сделать как парные прочтения ДНК, так и чтение, с одной стороны.

В 2012 году компания выпустила новый секвенатор MiSeq. [1] У этого секвенатора пониженная частота ошибок, что позволяет выполнять более длинные прочтения. Компания позиционирует прибор для небольших клинических лабораторий, так как позволяет за относительно небольшие деньги (имея ввиду стоимость запуска) секвенировать экзом человека. В 2014 году вышел средний по производительности NextSeq 500. Процент ошибок в прочтении на приборах компании Illumina в среднем менее 1.

Библиотека ДНК для всех секвенаторов подготавливается методом мостиковой ПЦР. Секвенирование осуществляется с использованием обратимых терминаторов, что ограничивает действия ДНК-полимеразы, не давая ей включать более одного основания за цикл. Каждый цикл работы аппарата можно разделить на 4 этапа.

На первом этапе все типы нуклеотидов вводят в поточный чип вместе с ПЦР-колониями (кластерами) на твердой фазе. На втором этапе полимераза достраивает цепь ДНК, включая в нее один меченный нуклеотид. На третьем этапе прибор детектирует флуоресценцию от каждой колонии. На четвертом этапе химическими реагентами снимают 3’-блок вместе с флуофорами. Последовательность действий у всех вышеописанных секвенаторов данной компании одинакова. [1]

**Платформы Ion PGM и Ion Proton компании Life Technologies Thermo Fisher Scientific**

В 2010 году компания Ion Torrent Inc. Представила первый полупроводниковый секвенатор – Ion Personal Genome Machine, или Ion PGM. В настоящее время технология принадлежит компании Thermo Fisher Scientific, которая в феврале 2014 года приобрела эту технологию и компанию Life Technologies. В 2012 году вышел более мощный Ion Proton, работающий на том же принципе регистрации изменения проводимости среды за счет выделившихся в результате акта присоединения нуклеотидов протонов.

На этапе подготовки образец ДНК пробанда расщепляют любым способом до фрагментов длиной 100–500 п. н. Длина фрагментов зависит от типа реагентов, используемого в процессе. Далее к фрагментам присоединяют разные по последовательности адаптеры и проводят клональную наработку фрагментов ДНК методом эмульсионной ПЦР: фрагменты амплифицируют на ISP-сферах (с англ. ion sphere particles), на которых закрепляют комплементарные одному из адаптеров праймеры. [1]

Приборы серии OneTouch дают возможность проводить эПЦР для платформ Ion PGM и Ion Proton в автоматическом режиме. Эти приборы сконструированы так, что реактивы в них помещают на несколько запусков одного типа. Возможен переход с одного типа реагентов на другой, но это приводит к потере реагентов и, следовательно, потери количества возможных реакций. Поэтому более логично выстраивать работу с этим прибором, таким образом, чтобы последовательно проводилась эПЦР одного типа.

В 2013 году появилась полностью автоматизированная система подготовки – Ion Chef. Система выполняет эПЦР, обращение и загрузку чипа. Однако покупка такой системы, может быть оправдана только в высокоспециализированных лабораториях, где секвенирование проводится непрерывно.

Для Ion PGM поставляются реагенты, обеспечивающие длину прочтения 200 или 400 п. н. Регулировать длину прочтения можно вручную, на тех же реагентах можно добиться прочтений 250 и 440 п. н. с достаточно качественным прочтением последних 50 п. н. К секвенатору предлагаются три варианта чипов: 314, 316 и 318, имеющие соответственно 1,2, 6,2 и 11,1 млн пикселей – микрореакторов для микросфер. Они увеличивают увеличение масштаба получаемых данных в 3–5 раз. Так, например, на чипе 318 с количеством реагентов, рассчитанных на 400 п. н., можно получить 3 млрд п. н. «сырых» данных.

Чип – это матрица из микрореакторов, в каждый из которых может поместиться только одна ISP-сфера с прикрепленными к ней одинаковыми молекулами ДНК. Каждый микрореактор расположен над ISFET-сенсором (ISFET – ion sensitive field-effect transistor, или канальный транзистор измерения концентрации ионов). [1]

Сенсор передает сигнал от возникающих ионов водорода в программу, переводящую его в последовательность нуклеотидов. Такие чипы обладают большим разрешением по сравнению с оптическими системами детекции. Детекция сигнала происходит сразу в виде цифрового значения, а это значит, что не требуется анализировать изображение, полученное с микроскопа.

Важное преимущество полупроводникового секвенирования наряду с достаточно простыми и дешевыми реагентами – это скорость. Один цикл реакции от введения раствора нуклеотида до очистки чипа занимает несколько секунд, а время всего процесса секвенирования – несколько часов.

Для Ion Proton пока выпускаются только чипы в 150 мегапикселей, реагенты доступны для прочтения 200 п. н. Таким образом, производительность этого прибора сейчас достигает 10–12 млрд п. н. за один запуск. Подготовка для Ion Proton такая же, как и для Ion PGM.

Проблемой полупроводникового секвенирования остается чтение гомополимеров: при повторе из более чем пяти одинаковых азотистых оснований, выделяемое число протонов не позволяет точно определить число нуклеотидов, и эти результаты чтения отбраковываются программным обеспечением прибора.

Кроме того технология очень чувствительна к качеству воды, из которой готовятся рабочие буферы. Поэтому в близости от прибора должен находиться источник деионизированной воды (mQ-воды). [1]

Система подачи реактивов в секвенаторах компании Thermo Fisher Scientific работает на инертных газах, поэтому рядом с секвенатором должен располагаться баллон высокого давления, по правилам техники безопасности, хранящийся в специальном металлическом шкафу. Оператор, работающий с секвенатором, должен иметь сертификаты, удостоверяющие его навыки работы с газобаллоным оборудованием.

**Платформа PacBio компании Pacific Biosciences**

Эта технология базируется на методе секвенирования синтезом одиночных молекул. Компания Pacific Biosciences была создана в 2004 году. А в 2011 году выпустила свой первый коммерчески доступный секвенатор PacBio RS. В 2013 году вышла вторая версия прибора – PacBio RSII.

Метод, лежащий в основе действия прибора, называется SMRT (single molecule real time), или секвенирование молекул «в реальном времени». В основе SMRT лежит технология оптической детекции включения отдельных флуоресцентно-меченых нуклеотидов в строящуюся цепь молекулы ДНК. В чипе для секвенирования находятся рабочие камеры, на дне которых закреплена молекула phi29 ДНК-полимеразы. Каждый нуклеотид мечен своим флуофором, отсоединяющимся при образовании фосфодиэфирной связи. Прибор фиксирует флуоресценцию единичной молекулы флуофора и длительность испускания света.[16]

Основное преимущество данной технологии – длина прочтений [16], в среднем сейчас она составляет 8500 п. н. Проанализировав такой массив данных, становится возможным прочитать проблемные для других платформ участки генома, например, тандемные повторы. Анализ кинетики включения нуклеотидов в цепь, позволяет проводить прямой анализ уровня метилирования геномов.

Однако у PacBio низкая точность прочтения отдельных нуклеотидов, поэтому обычно этот прибор используется в качестве дополнения к приборам Illumina или Ion Torrent. Также прибор достаточно громоздкий – он весит больше тонны. [1]

**Платформа HeliScope компании Helicos Biosciences**

Компания была основана в 2003 году, однако сейчас не ведет деятельности. В 2008 году компания выпустила секвенатор HeliScope, который на то время был первым прибором, проводящим секвенирование одиночных молекул по технологии tSMS (true single molecule sequencing) – истинное секвенирование отдельных молекул.

Технология HeliScope похожа на технологию Illumina, с отличием в том, что на чипе HelicoScope не требуется амплификация. Фрагменты ДНК расперделяются так, что в один пиксель матрицы попадает только одна молекула, чтобы прибор мог зарегистрировать единичный сигнал. [1]

Для секвенирования фрагментированную ДНК модифицируют с 3’-конца, добавляя участок полиаденин, чтобы матрица могла отжечься на иммобилизованных на подложке праймерах (олиго-dT). Матрица представляет собой одноцепочечные фрагменты. В процессе секвенирования ДНК-полимераза последовательно добавляет в синтезируемую цепь один из четырех флуоресцентно-меченных нуклеотидов в течение одного цикла.

Детекция сигнала флуоресценции от отдельных нуклеотидов в конце каждого цикла осуществляют четыре лазера и система, сходная с конфокальным микроскопом. Ввиду того, что молекулы ДНК неподвижны на чипе, в процессе секвенирования получается набор изображений, на которых, в случае присоединения нуклеотида, видны яркие точки. Эти точки – это флуорофоры, излучающие свет под воздействием лазера.

Программное обеспечение Helicos сопоставляет сигнал на изображении с координатами точек и определяет последовательность отсеквенированных нуклеотидов у каждого фрагмента матрицы. Максимальная длина прочтения у Helicos составляет 55 нуклеотидов. [1]

**Платформы MinION, GridION X5, PromethION и SmidgION компании Oxford Nanopore Technologies**

В 2009 году сотрудники Оксфордского университета научились различать все четыре типа нуклеотидов в одной цепи ДНК по мере их прохождения через ионный канал диаметром 2,6 нанометра, находящийся в бислойной липидной мембране. В 2014 году на рынок вышел первый секвенатор компании Oxford Nanopore Technologies.

Метод нанопорового секвенирования, описанный в первой главе исследования, реализуется в приборах MinION, GridION X5, PromethION и SmidgION. [4]

Методика позволяет достигать длины прочтений 100 тысяч п. н. Длина прочтения, которую предлагает нанопоровое секвенирование, позволяет значительно упростить и ускорить сборку ранее не опубликованных геномов, а также улучшить уже существующие сборки. Кроме того, приборы, базирующиеся на методе нанопорового секвенирования, не имеют проблем с определением нуклеотидов в гомополимерных участках генома. [1]

Однако частота ошибок на таких приборах достаточно велика. Компания решает эту проблему: так, в 2015 году предварительно сообщали о сорока процентах ошибок в прочтениях секвенатора MinION, но к настоящему времени оно сократилось до десяти. Размер нанопоровых секвенаторов позволяет проводить исследования в любых местах.

Преимуществом нанопорового секвенирования является также анализ полученных данных в режиме реального времени. Это дает возможность остановки процесса в любой момент при достаточном накоплении данных, при очевидных ошибках пробоподготовки или других сбоях в работе.

Компания Oxford Nanopore Technologies выпустила специальный набор для пробоподготовки 1D2, который дает возможность считывать комплементарные цепи ДНК, что увеличивает точность секвенирования почти до 99%. [4]

Прибор MinION представляет собой устройство карманного формата, которое можно подключить к USB-порту компьютера. В основе прибора — ячейка с 512 нанопорами, позволяющая провести одновременное секвенирование до 512 молекул ДНК. Процесс пробоподготовки длится около 10 минут (с помощью набора Rapid Sequencing Kit). Сейчас на рынке доступны три секвенатора разной производительности:

GridION — настольный секвенатор, в основе которого 5 проточных ячеек (2560 каналов с нанопорами). Каждая ячейка работает независимо. В устройство встроен компьютер.
PromethION — высокопроизводительный секвенатор, в основе 48 проточных ячеек (144 000 каналов с нанопорами). Каждая ячейка работает независимо. Применяется для анализа больших массивов данных.
SmidgION — разрабатываемый в данный момент прибор, с вожмостью подключения к смартфону или другому мобильному устройству. Предназначается для работы в полевых условиях. [5]

**Результаты**

**Таблица 1. Сравнение производительности и цены разных технологий**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Технология** | **Прибор** | **Клональная ПЦР, либо ее отсутствие** | **Принцип действия** | **Чтение с двух сторон** | **Длина прочтения, п. н.** | **Продолжительность одного запуска, дни** | **Цена за млн п. н., руб** | **Цена оборудования, млн руб.** |
| Roche 454 Life Sciences | 454 Genome Sequencer | эмульсионная | пиросеквенирование | нет | 800 | 1 | 600 | 10 |
| 454 Genome Sequencer Junior | 500 | 0,4 | 1500 | 5 |
| SOLiD | SOLiD 5500 | эмульсионная | меченые октамеры | есть | 60 | 7–14 | 5 | 20 |
| Illumina | HiSeq 2500(2 чипа) | мостиковая | обратимые терминаторы | есть | 100 | 3–10 | 4 | 20 |
| MiSeq | 150 | 0,5 | 5 | 10 |
| Ion Torrent | Ion PGM – 314 чип | эмульсионная | полупроводниковое секвенирование | нет | 200 | 0,1 | 400 | 5 |
| Ion PGM – 318 чип | 200 | 40 | 5 |
| Ion PGM – PI чип | 150 | 7 | 10 |
| Pacific Biosciences | PacBio RSII | нет | обратимые терминаторы | нет | 30 000 | 0,1 | 20 | 40 |
| Helicos Biosciences | HeliScope | нет | обратимые терминаторы | нет | 50 | — | 30 | 30 |
| Oxford Nanopore Technologies | MinION | нет | нанопоровое секвенирование | нет | 100 000 | 2 | — | 0, 067 |
| GridION | — | 3,35 |
| PromethION | — | 11, 055 |
| SmidgION | в разработке |

**Выводы**

1. Приборы, подходящие для клинических лабораторий:

Секвенаторы компании Illumina, так как эти приборы позволяют за относительно небольшие деньги (если иметь ввиду стоимость запуска, а не прочтения миллиона оснований) секвенировать экзом человека.

1. Приборы, подходящие для работы в полевых условиях:

Секвенаторы компании Oxford Nanopore Technologies компактные и производительные. Однако их результаты лучше дополнять данными, полученными с другого прибора, так как на данный момент они в среднем совершают ошибку на каждом десятом основании. Компания продолжает совершенствовать свои технологии.

1. Приборы, подходящие для секвенирования геномов *de novo*:

Секвенатор PacBio RSII отлично подходит для такой цели.

1. Приборы, с большой длиной прочтения:

Секвенаторы, в основе которых лежит технология Roche 454 Life Sciences, секвенатор PacBio RSII и секвенаторы компании Oxford Nanopore Technologies.

1. Приборы с высокой скоростью работы:

Секвенаторы Ion Torrent и PacBio RSII.

1. Приборы, которые могут читать гомополимерные участки генома:

Секвенаторы компании Oxford Nanopore Technologies.

Секвенирование по технологии SOLiD довольно дорогостоящее, сложное и длительное, поэтому можно спрогнозировать ее скорый уход с рынка. Согласно имеющимся данным, технология HeliScope в России не используется, а на начало 2014 года в мире установлено лишь 20 приборов.

Список литературных источников:

1. **Ребриков, Д.В. /NGS. Высокопроизводительное секвенирование** [Текст]**// Д.В. Ребриков Коростин, Шубина – Чехов: Ильинский Издательский дом «БИНОМ. Лаборатория знаний», 2014. – С.13-40, 66-83**
2. Уолкер Ш., [Текст]. Биотехнология без тайн – Москва: Издательство «Эксмо», 2008 – С. 163-166
3. Елена Байтингер, «Секвенируй меня полностью» [Текст]/Елена Байтингер// Научно-популярный журнал «Кот Шредингера». – 2017. – Специальный номер. – С. 47-49
4. Научно-популярный сайт «Биомолекула», статья «Нанопоровое секвенирование: на пороге третьей геномной революции»/Артем Недолужко 2018 [Электронный ресурс] (<https://biomolecula.ru/articles/nanoporovoe-sekvenirovanie-na-poroge-tretei-genomnoi-revoliutsii>) (09.02.19)
5. Сайт компании «Диаэм». Нанопоровое секвенирование [Электронный ресурс] (<https://www.dia-m.ru/news.php?newsid=59771>) (09.02.19)
6. Статья об открытии локиархей [Электронный ресурс] (<https://elementy.ru/novosti_nauki/432910/Opisan_novyy_nadtip_arkhey_k_kotoromu_otnosyatsya_predki_eukariot>) (10.11.18)
7. Статья «Неудача – это нормально», интервью с нобелевским лауреатом Даном Шехтманом [Электронный ресурс] (<https://ology.sh/ponder/neudacha-eto-normalno/>) (10.11.18)
8. Сайт проекта «Российские геномы» [Электронный ресурс] (<http://genomerussia.spbu.ru/>) (10.11.18)
9. Сайт проекта «Earth BioGenome Project» [Электронный ресурс] (<https://www.earthbiogenome.org/>) (10.11.18)
10. Статья «Разгадывая генетический код» [Электронный ресурс] (<https://www.scienceinschool.org/ru/content/%D1%80%D0%B0%D0%B7%D0%B3%D0%B0%D0%B4%D1%8B%D0%B2%D0%B0%D1%8F-%D0%B3%D0%B5%D0%BD%D0%B5%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B8%D0%B9-%D0%BA%D0%BE%D0%B4-%D0%B2%D0%BE%D1%81%D1%81%D0%BE%D0%B7%D0%B4%D0%B0%D0%BD%D0%B8%D0%B5-%D0%BD%D0%B0%D1%83%D1%87%D0%BD%D0%BE%D0%B3%D0%BE-%D0%BE%D1%82%D0%BA%D1%80%D1%8B%D1%82%D0%B8%D1%8F>) (11.11.18)
11. Курс «Введение в биоинформатику: метагеномика» [Электронный ресурс] (<https://stepik.org/course/2557/>) (10.11.18)
12. Офицальный сайт компании «Genotek» [Электронный ресурс] (<https://www.genotek.ru/>) (08.11.18)
13. Научно-популярная статья о проекте «Earth BioGenome Project» [Электронный ресурс] (<https://fainaidea.com/nauka/biologija/zapushhen-unikalnyj-masshtabnyj-proekt-earth-biogenome-142975.html>) (10.11.18)
14. Статья про метод секвенирования по Максаму-Гилберту [Электронный ресурс] (<http://enc.sci-lib.com/article0001457.html>) (16.12.18)
15. Статья про проект «Геном человека» [Электронный ресурс] (<https://elementy.ru/nauchno-populyarnaya_biblioteka/431280/Proekt_Genom_cheloveka_desyat_let_spustya>) (16.12.18)
16. Статья «12 методов в картинках: секвенирование нуклеиновых кислот» [Электронный ресурс] (<https://biomolecula.ru/articles/metody-v-kartinkakh-cekvenirovanie-nukleinovykh-kislot>) (20.11.18)
17. Сайт компании «Диаэм». Секвенирование нуклеиновых кислот [Электронный ресурс] (<https://www.dia-m.ru/page.php?pageid=33696>) (21.02.19)

**Приложение**

Словарь терминов:

«1 — **Амплификация** — процесс образования дополнительных копий участков хромосомной ДНК.

2 — **Хроматография** — метод выделения, разделения и идентификации близких по химическим свойствам веществ.

3 — **Агент** — действующий фактор в каком-либо процессе или явлении.

4 — **Рентгеноструктурный анализ** — один из дифракционных методов исследования структуры вещества.

5 — **Масс-спектрометрия** — метод исследования вещества, основанный на определении отношения массы к заряду ионов.

6 — **Метилирование** – это химическая модификация, катализируемая ферментом, реакция добавления метильных группп на ДНК.

7 — **Вставка** — чужеродный ген, вносимый в клетку.

8 — **Субклонирование** — перенос части уже клонированной молекулы ДНК в другой клонирующий вектор. С. используют для секвенирования длинных фрагментов ДНК, а также для получения различных сложных генно-инженерных конструкций.

9 — **Вектор** — это последовательность ДНК, в которую вшивается вставка.

10— **Библиотека (банк) генов** — это набор фрагментов ДНК, в котором представлены все (или определенная часть) генов организма. Библиотека генов представляет собой совокупность культур микроорганизмов (бактерий, дрожжей), в каждую клетку которых введен вектор, несущий один из фрагментов генома организма.

Библиотеку генов можно длительно хранить (в замороженном состоянии) и, по необходимости, выделять отдельные микроорганизмы, содержащие фрагменты ДНК с нужными генами, и клонировать их.

11— **Адаптер** — синтетический двухцепочечный олигонуклеотид с одним тупым и одним липким концами или с двумя липкими концами; после пришивания адаптера тупым концом к ДНК-мишени последнюю можно затем встраивать в подходящий вектор, используя приобретенный ею от адаптера липкий конец;

12— **Праймер** — это короткий фрагмент нуклеиновой кислоты, комплементарный ДНК- или РНК-мишени, служит затравкой для синтеза комплементарной цепи с помощью ДНК-полимеразы.»