

БИОИНЖЕНЕРИЯ

№



Кактусы – не самый подходящий объект биоинженерных изысканий.

Секвенирование путем синтеза с обратным терминированием (Illumina)

Метод основан на детекции флюорофора при присоединении каждого нового нуклеотида к растущей цепи ДНК в ходе ПЦР.

Основной отличительной чертой технологии является применение так называемого метода молекулярных колоний.

Длина одного ряда: 100-300 нуклеотидов.

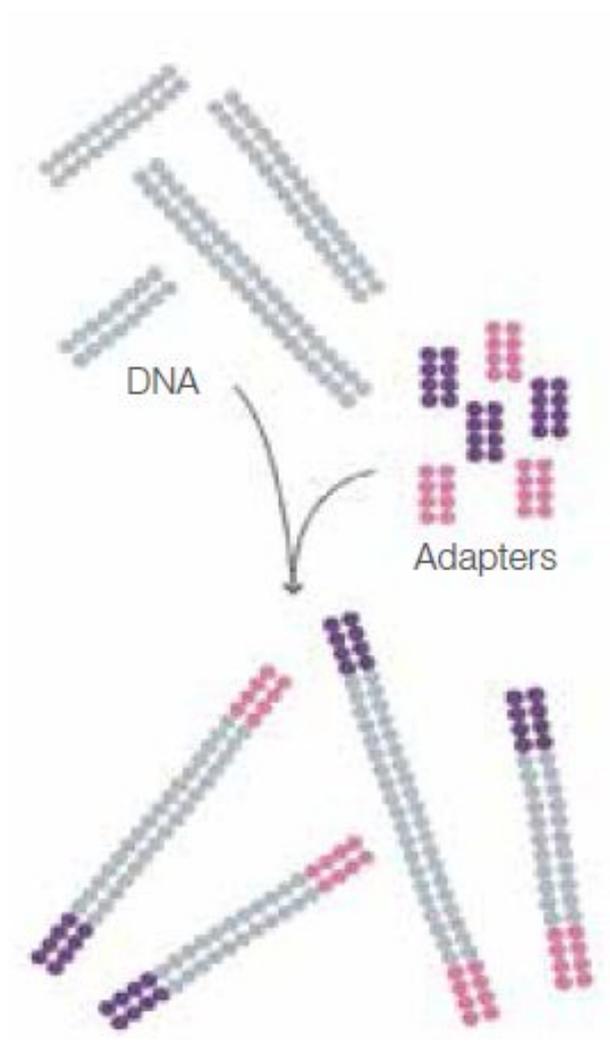
В настоящее время **до 80% геномных данных** генерируется посредством метода Illumina.

Основное достоинство: всё.

Основной недостаток: в общем-то, нет.

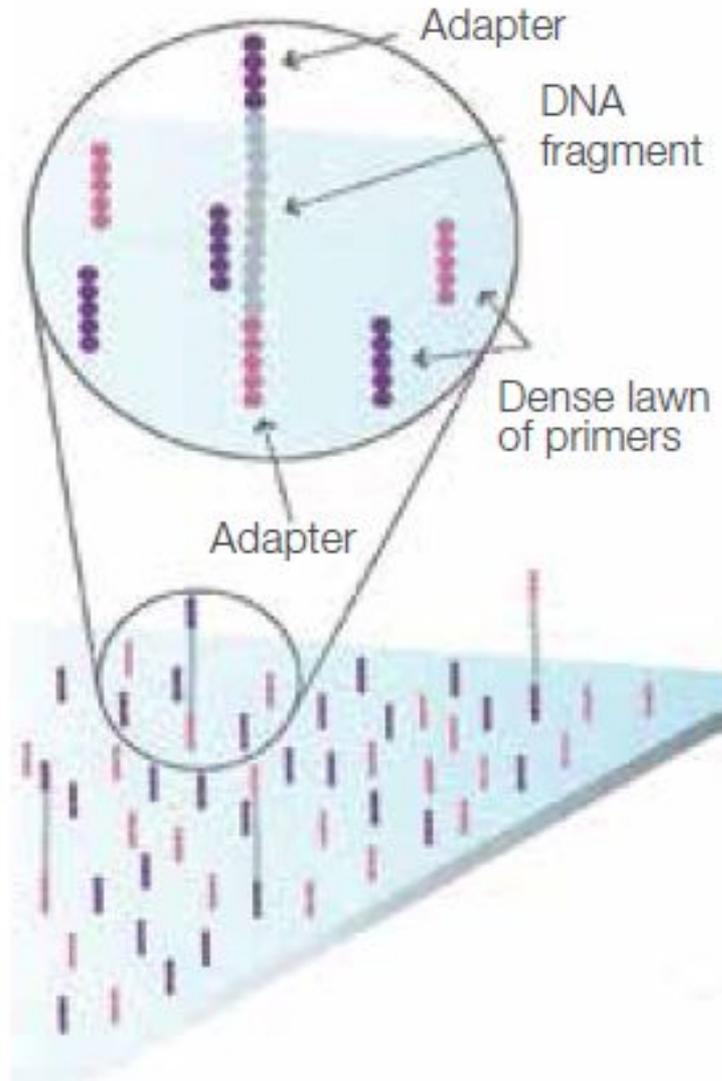
Первый прибор: компания Solexa, 2006 год.

Секвенирование Illumina



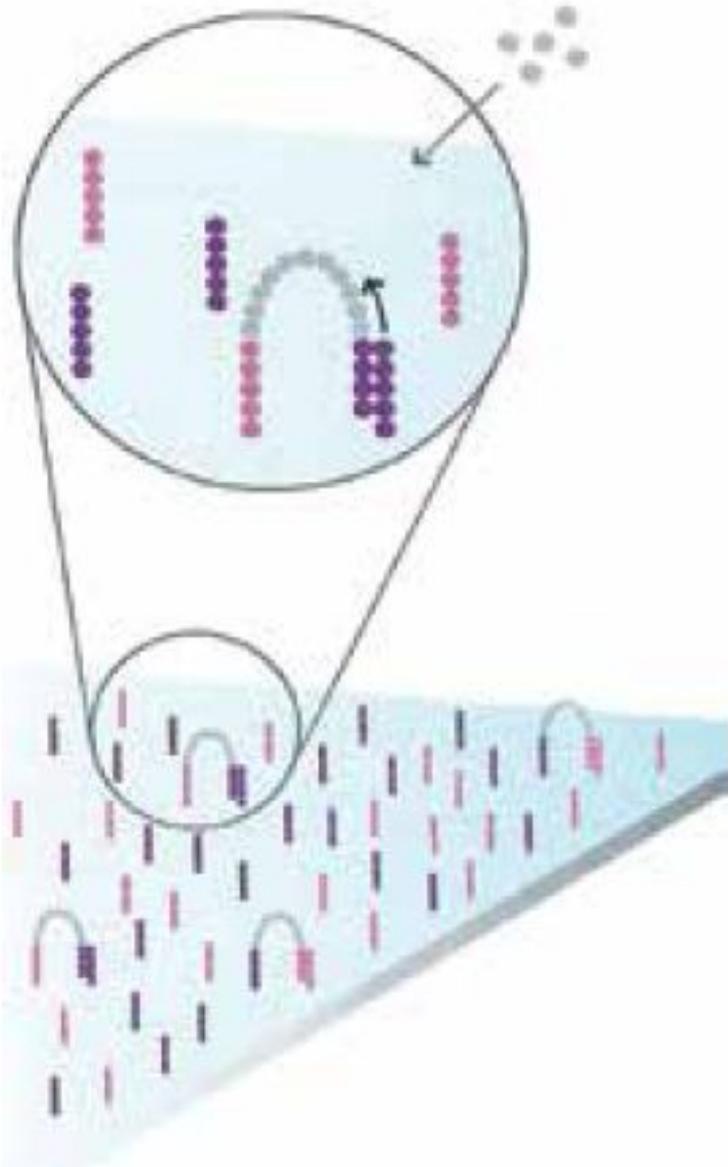
1. Фрагментируйте двуцепочечную ДНК и пришейте к концам фрагментов адаптеры, причем разные.
2. Амплифицируйте фрагменты при помощи ПЦР.

Секвенирование Illumina



3. Подготовьте ячейку, к поверхности которой пришиты олигонуклеотиды, соответствующие адаптерам. В одном из нуклеотидов должен быть редкий сайт рестрикции.
4. Денатурируйте опытную ДНК и закрепите ее в ячейке. Фрагменты ДНК окажутся где-то между олигонуклеотидами.

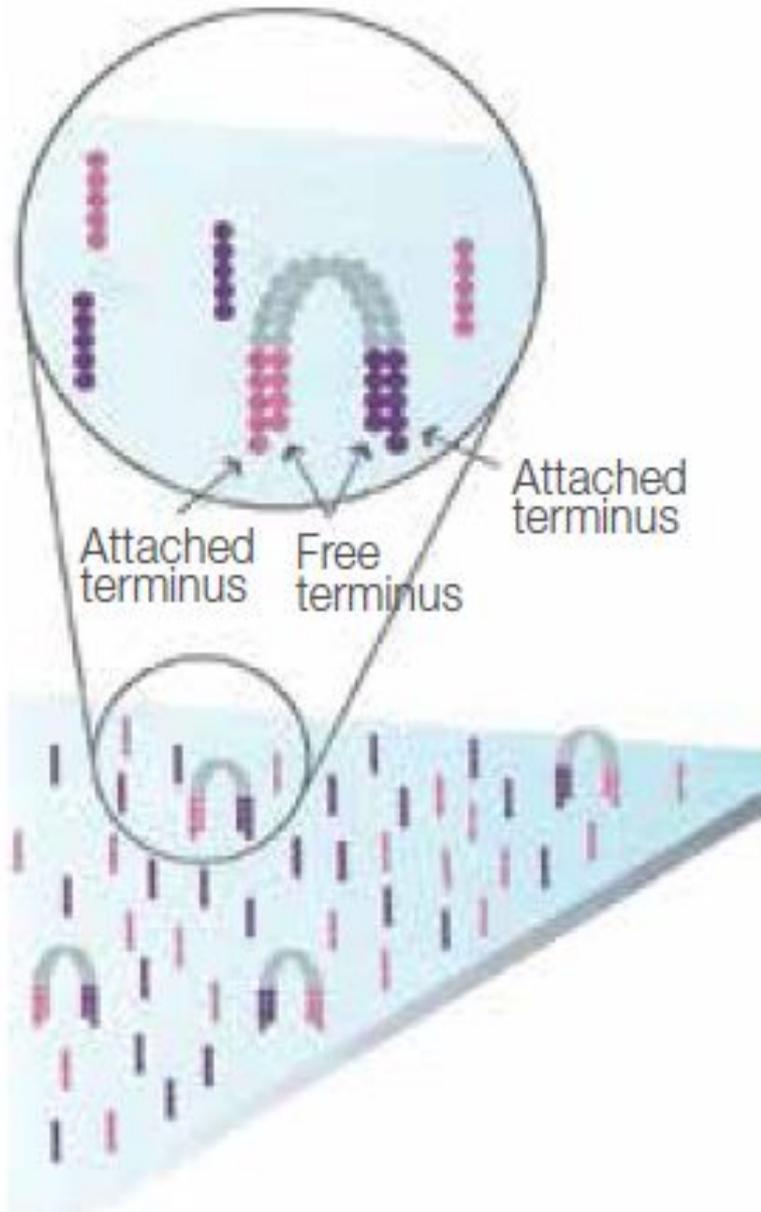
Секвенирование Illumina



5. Добавьте все для ПЦР, кроме праймеров. В этом случае что-либо может случиться, только если фрагмент ДНК изогнулся, как показано на рисунке, и его свободный конец отжегся на свободный олигонуклеотид, закрепленный в ячейке.

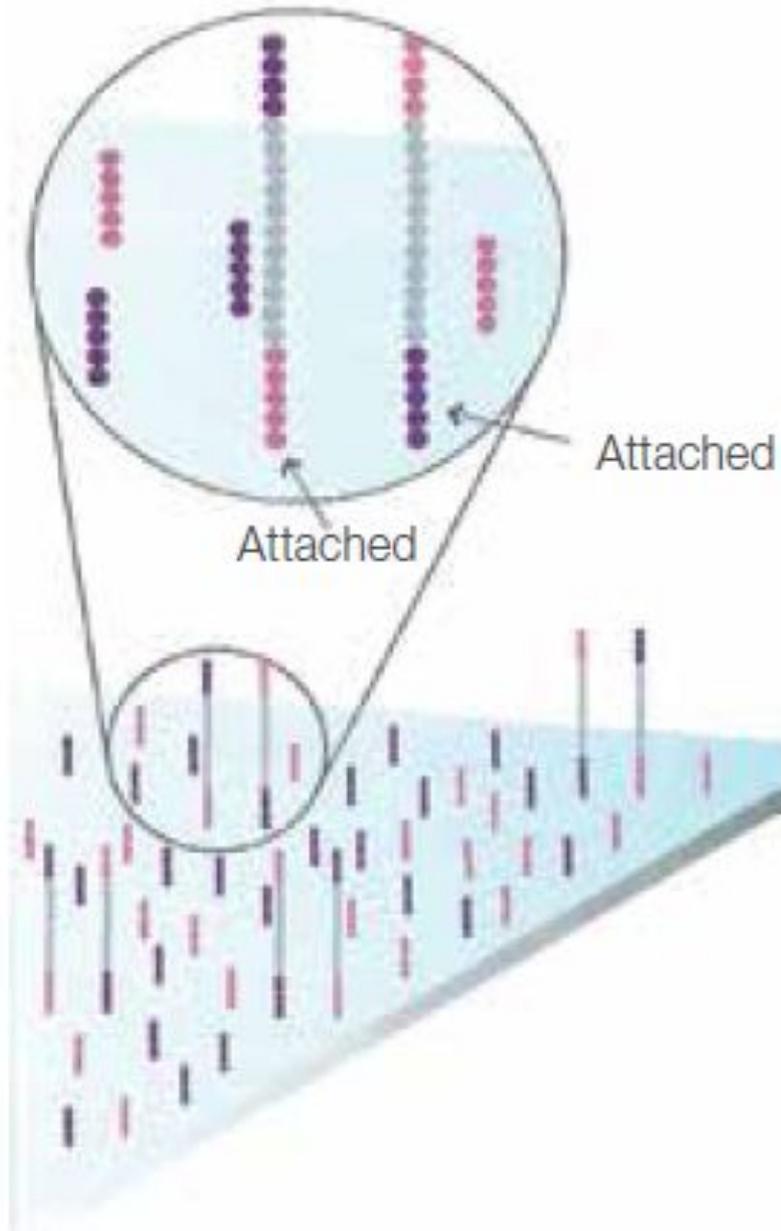
Это называется «мостиковая амплификация».

Секвенирование Illumina



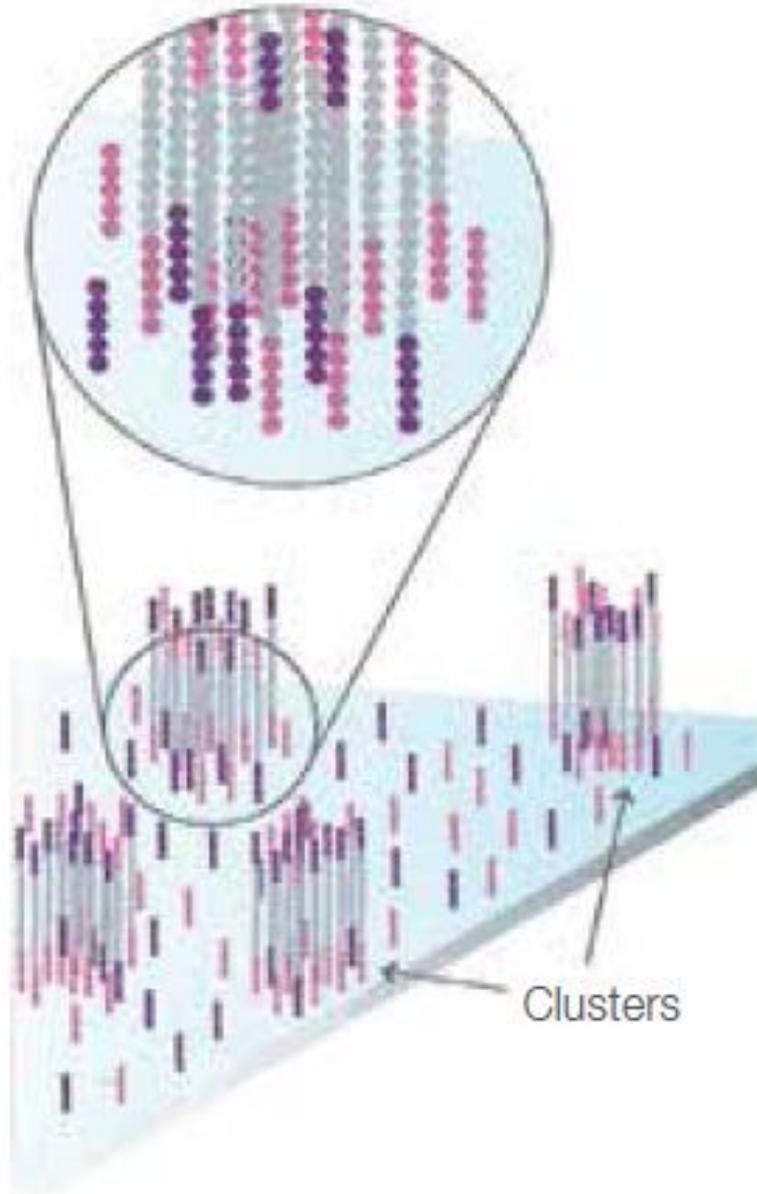
6. В результате первого раунда амплификации образуются двуцепочечные молекулы ДНК, каждая цепочка которых одним концом ковалентно закреплена на ячейке, а другой конец вовлечен только в комплементарное взаимодействие.

Секвенирование Illumina



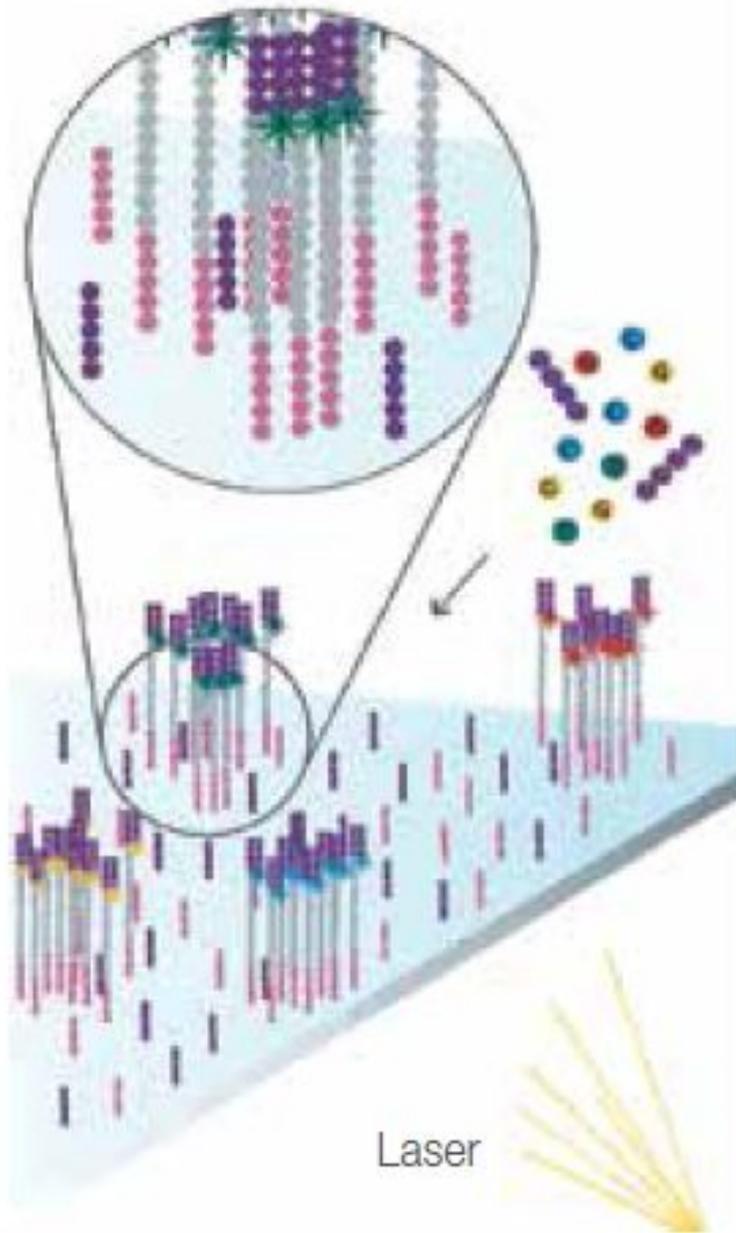
7. По окончании первого раунда ПЦР денатурируйте ДНК. Каждая цепочка будет одним концом ковалентно прикреплена к поверхности ячейки. Одинаковые молекулы (две разделившиеся цепи двуцепочечного продукта) будут всегда находиться рядом друг с другом.

Секвенирование Illumina



8. Повторите процедуру многожды (типа ПЦР). У вас получатся кластеры одинаковых молекул, ориентированных двумя различными способами.
9. Обработайте ячейку рестриктазой, сайт для которой был внесен в один из олигонуклеотидов. Молекулы ориентации А оторвутся от ячейки. Промойте ячейку. Теперь на ней остались только молекулы ориентации Б.

Секвенирование Illumina



10. Добавьте праймеры и ДНК-полимеразу. Добавьте также 3'-О-азидометил 2'-деоксинуклеозидтрифосфаты (А, С, G и Т), каждый с отделяемой флуоресцентной меткой своего цвета. Наличие 3'-О-азидометила не позволяет ДНК-полимеразе присоединить больше одного нуклеотида. Полимераза присоединяет один модифицированный нуклеотид, оставшиеся нуклеотиды вымываются.

11. Индуцируйте флуоресценцию. Присоединенный флюорофор светится своим цветом, и этот сигнал от кластера одинаковых молекул будет сильным. Зарегистрируйте сигнал.

12. Добавьте трис(2-карбоксиэтил)фосфин, из-за которого флюорофор и азидометил отделяются и вымываются. 3'-гидроксильная группа становится доступной для присоединения ещё одного нуклеотида.

13. Повторите пп.10-12.

Нерешенные до конца проблемы NGS

Необходима точная оценка концентрации библиотеки

- qPCR – необходима стадия контроля качества!

Аmplification вызывает «перекос» в сторону тех или иных фрагментов – неравномерность покрытия

- более совершенные полимеразы
- уменьшение числа циклов ПЦР или PCR-free библиотеки

Дублицированные чтения

- уменьшение числа циклов ПЦР
- повторности амплификации

Продукты самолигирования (димеры) адаптеров

- более совершенные ферменты
- дополнительная очистка

Области применения NGS

- **секвенирование геномов и транскриптомов *de novo***

отправная точка большинства молекулярно-биологических и генетических исследований на немодельных объектах

- **полногеномное ресеквенирование**

поиск мутаций, ассоциированных с болезнями, картирование генов

- **направленное ресеквенирование**

биомедицина: скрининг мутаций с известной ролью в развитии болезней и поиск новых мутаций

- **анализ транскриптома (RNA-seq)**

сравнение уровней экспрессии, поиск новых генов и изоформ, аннотация секвенированных *de novo* геномов

- **ДНК-белковые и ДНК-ДНКовые взаимодействия (ChIP-Seq)**

поиск сайтов связывания транскрипционных факторов, изучение пространственной организации хроматина

- **метагеномика**

анализ разнообразия микробных сообществ

Секвенирование *de novo*

Особенности:

- диапазон размеров геномов эукариот – от 2.9 миллионов до 150 миллиардов
- повторы
- гетерозиготность, полиплоидия

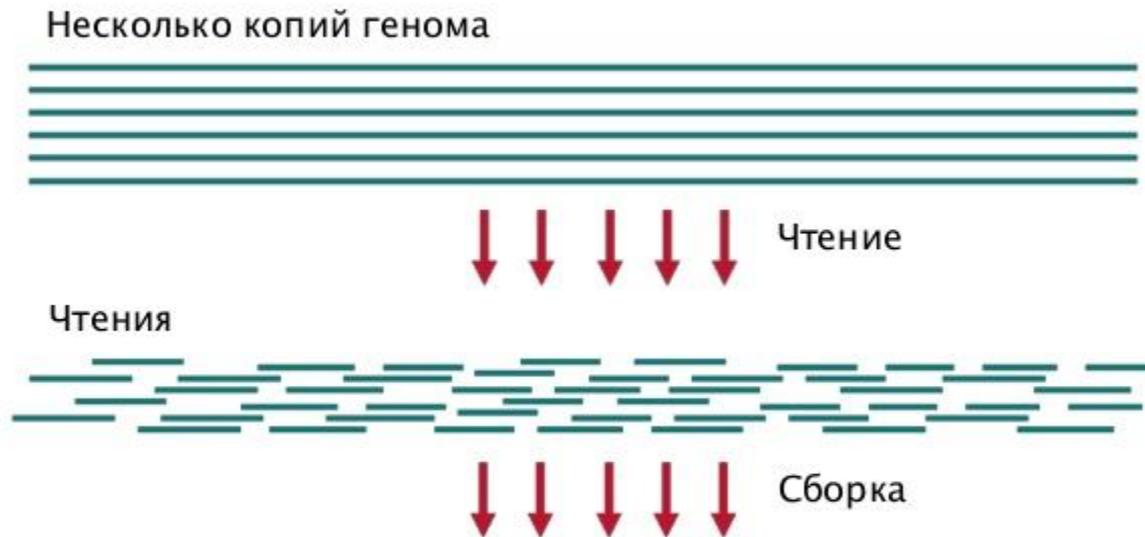
Требования:

- покрытие не менее 50x (каждую букву генома нужно прочесть не менее 50 раз)
- длина ридов – чем больше, тем лучше
- обязательные парные чтения (независимое параллельное секвенирование минимум двух одинаковых образцов)
- точность не критична (поскольку у нас высокое покрытие), но большое количество ошибок ухудшает последующую сборку
- образец для секвенирования должен быть **ОЧЕНЬ** чистым, без какой бы то ни было сторонней ДНК

Секвенирование *de novo*

Самая большая проблема собственно к секвенированию не относится.

СБОРКА ГЕНОМА ИЗ РИДОВ – вот самая большая проблема.



О том, как она решается, вам расскажут на биоинформатике



Сравнение технологий NGS

платформа/задача	454	Illumina	Ion Torrent/Proton	Solid
секвенирование de novo	+++	+++	++ для небольших геномов	-
полногеномное ресеквенирование	-	+++	++ для небольших геномов	++
направленное ресеквенирование	+ ампликоны	+++	+++	++
анализ экспрессии	-	+++	++ для небольших задач	++
ДНК-белковые взаимодействия (ChIP-seq) и т.д.	-	+++	++ для небольших задач	++
метагеномика анализ разнообразия микробных сообществ	+++ особенно 16S	+++ Miseq – 16S, Hiseq - полногеномное	+	-

Сравнение технологий NGS

(2012 год, когда они еще реально конкурировали друг с другом)

	Цена прибора (\$)	Цена одного запуска (\$)	Время одного запуска (ч)	Данные в рез-те одного запуска	Цена одного мега-байта (\$)	Частота ошибок (%)
Illumina	125000	750	27	2 Gb	0,5	0,8
Пиросеквенир	110000	1100	8	1 Gb	31	1,9
Ion Torrent	80000	400	3	1 Gb	4,25	1,7

Данные для «настольных» секвенаторов (малые размеры, малая мощность)

Секвенирование третьего поколения (TGS)

Технологии TGS начали развиваться совсем недавно. Первые научные работы на эту тему были опубликованы в 2008-2009 годах, а первый прибор появился на рынке в 2011 году.

Основные отличия TGS от NGS:

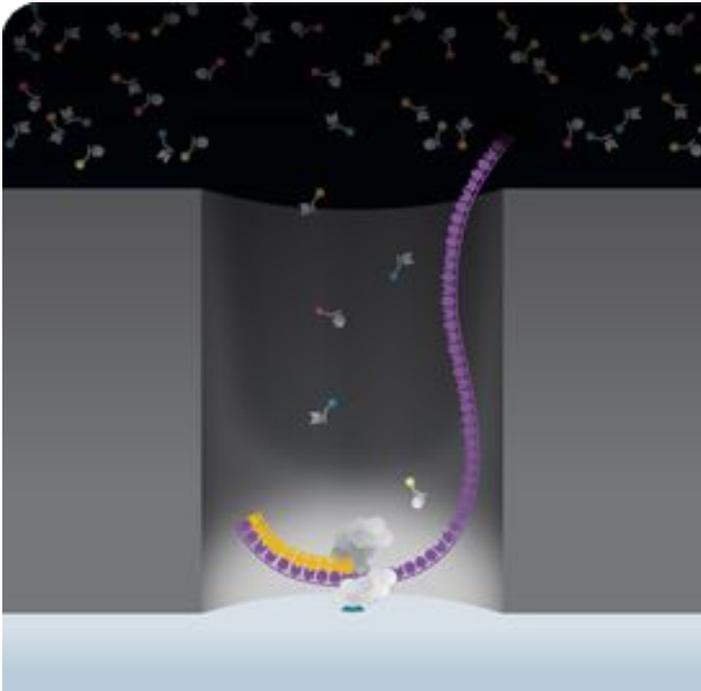
- Значительно более длинные риды (до сотен тысяч нуклеотидов!)
- Отсутствие необходимости амплификации ДНК перед секвенированием

В настоящее время коммерциализованы две технологии TGS, но на подходе еще несколько.

Single Molecule Real-Time Sequencing (SMRT)

Прибор производит компания Pacific Biosystems.

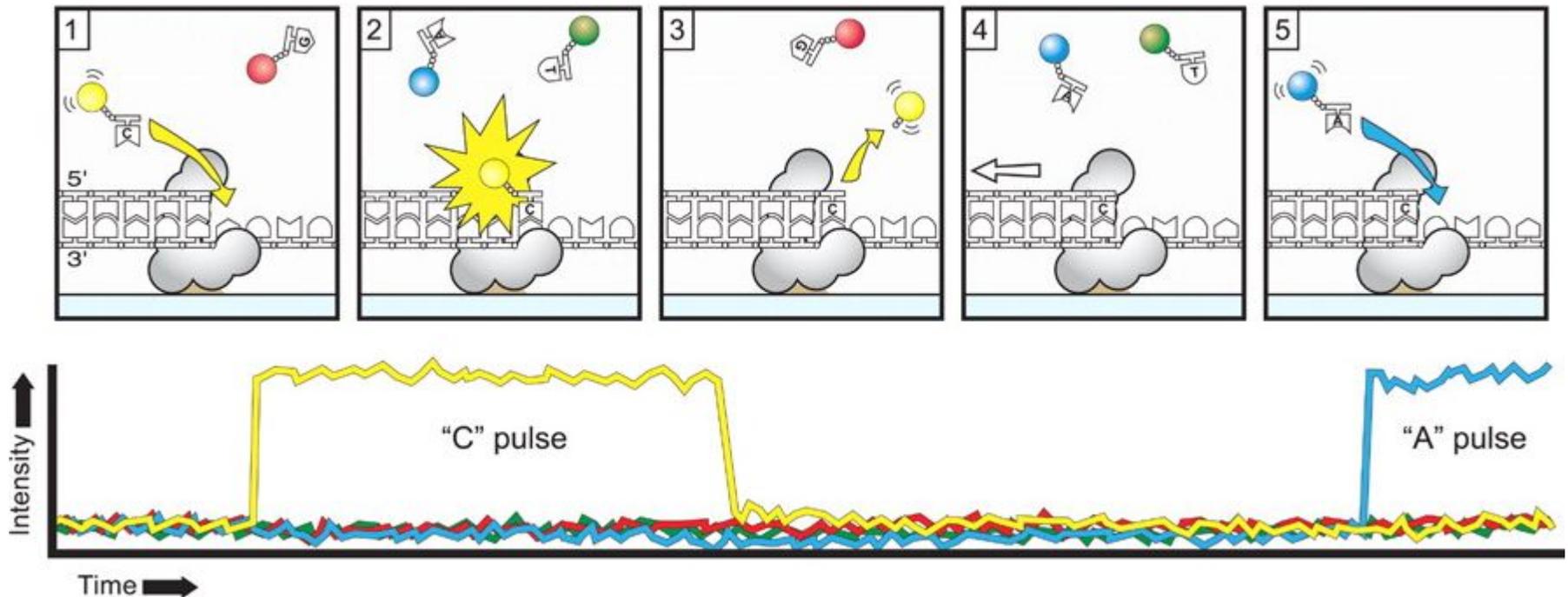
Технология основана на жестком закреплении ДНК-полимеразы в некоей ячейке, а также на том факте, что при помощи адски чувствительного детектора, работающего в очень маленьком участке пространства, можно достоверно поймать разницу в длительности свечения флюорофора, который присоединяется к растущей цепи ДНК (миллисекунды) и флюорофора, который безыдейно диффундирует в растворе (микросекунды).



Zero-mode waveguide (ZMW) — углубление диаметром несколько десятков нм, ко дну которого прикреплена ОДНА молекула ДНК-полимеразы. Сквозь дно в ZMW-ячейку подаётся свет. Особенность конструкции ZMW-ячейки не даёт распространяться световой волне и оставляет освещённым только объём порядка 10^{-20} литров около дна ячейки. Это позволяет регистрировать флуоресценцию от одной молекулы флюорофора.

Single Molecule Real-Time Sequencing (SMRT)

Флюорофоры пришиты к терминальным фосфатным группам нуклеотидов. Такая метка меньше влияет на работу ДНК-полимеразы, что крайне важно при секвенировании в реальном времени. После встраивания нуклеотида флуорофор диффундирует из наблюдаемого объёма и больше не влияет на регистрируемый сигнал. Таким образом, измеряя длительное (миллисекунды) свечение одного цвета при присоединении полимеразой меченного нуклеотида на фоне быстро диффундирующих (микросекунды) четырёх, можно определить последовательность матричной цепи ДНК.

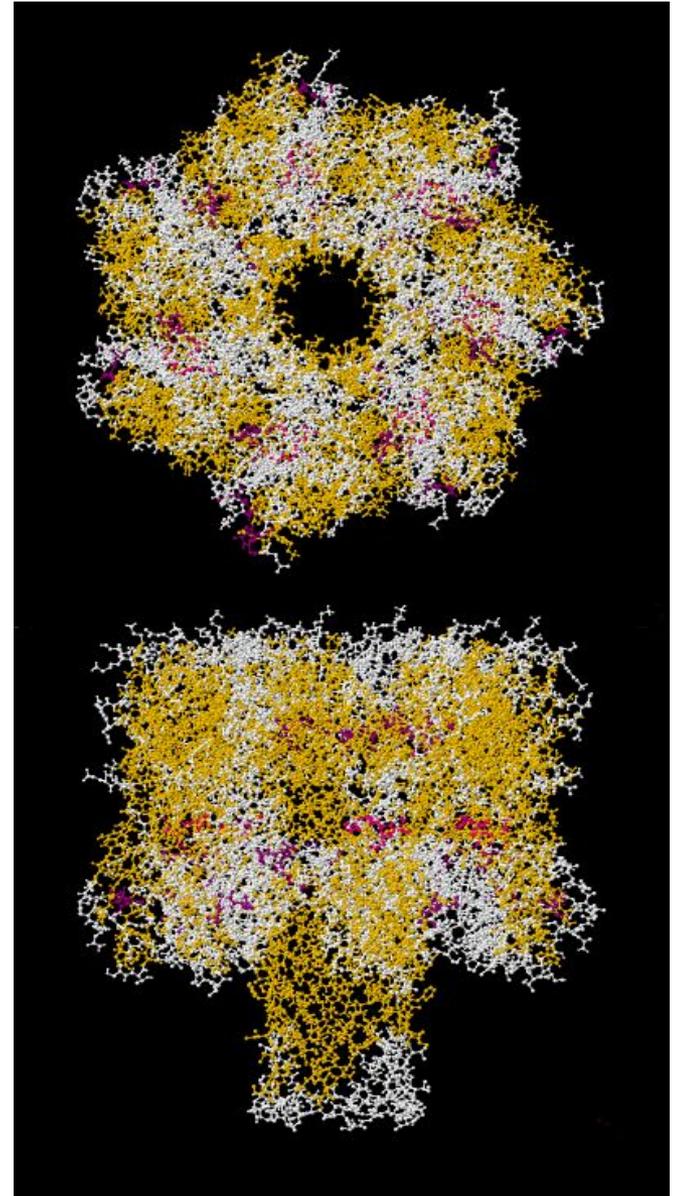


Нанопоровое секвенирование

Прибор производит компания Oxford Nanopore.

Технология основана на использовании бактериальных токсинов (производные стафилококкового альфа-гемолизина). Такие белки умеют образовывать в мембранах поры диаметром несколько нанометров. ДНК в солевом растворе при приложении электродов начинает проникать сквозь такую нанопору. Измеряя малейшие изменения проводимости поры, можно получать информацию о нуклеотидном составе ДНК!

Секвенатор в данном случае – это прибор размером с сотовый телефон, подключаемый к компьютеру через USB-порт.



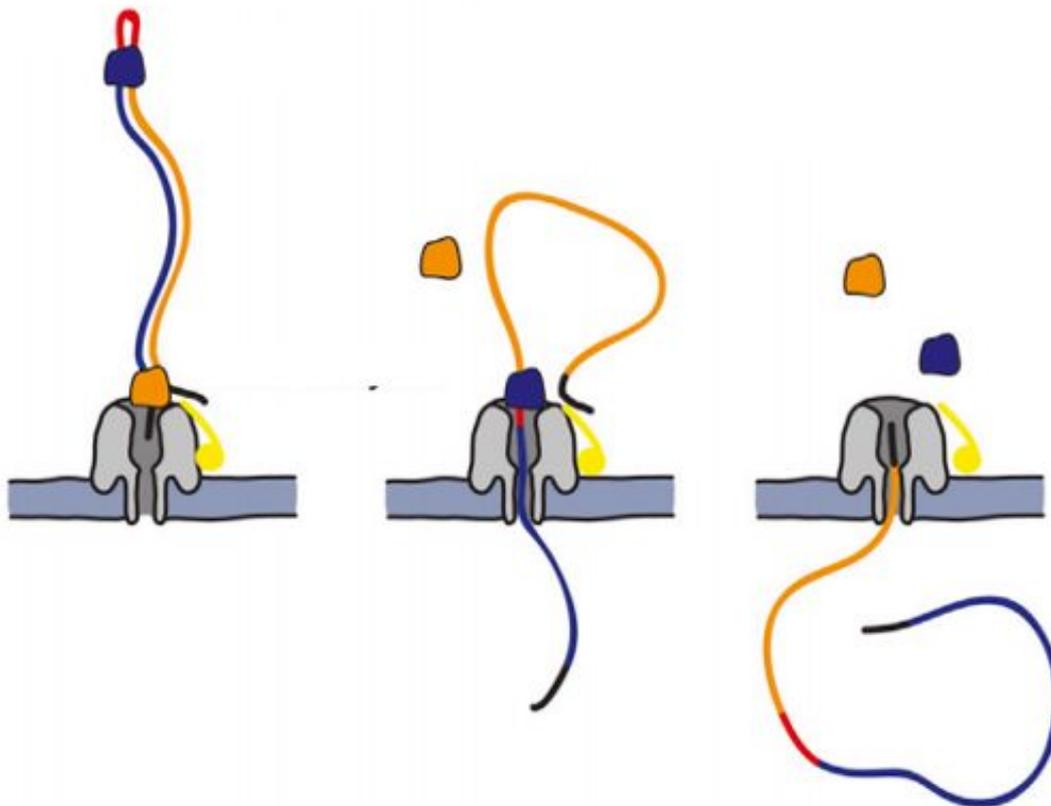
Нанопоровое секвенирование

Секвенатор MinION от Oxford Nanopore стоимостью 1500 долларов!!!!!!11



Нанопоровое секвенирование

Двухцепочечная ДНК сшивается с адаптером, представляющим собой шпилечный олигонуклеотид, замыкающий две цепочки ДНК друг на друга с одного из концов. После того, как одна цепочка ДНК целиком прошла через нанопору, туда начинает лезть другая. И тоже секвенируется!

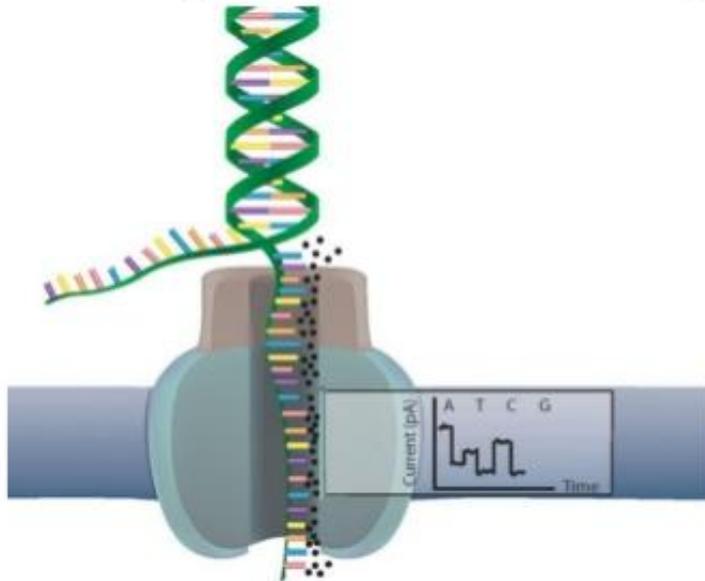


ЗАЧЕМ?

Нанопоровое секвенирование

В каждый момент времени детектируется изменение проводимости от 5-нуклеотидного участка ДНК. В следующий момент времени ДНК продвинулась сквозь пору на 1 нуклеотид, и детектируется изменение проводимости уже от другого 5-членника. Детектор должен уметь распознавать 1024 состояния проводимости! Он это и умеет, но все же с неидеальной точностью. Да еще и модифицированные основания мешают. В общем, точность метода – всего около 80%. Но можно просто повторить секвенирование 5 раз, и точность повышается до 99%!

Nanopore - technology



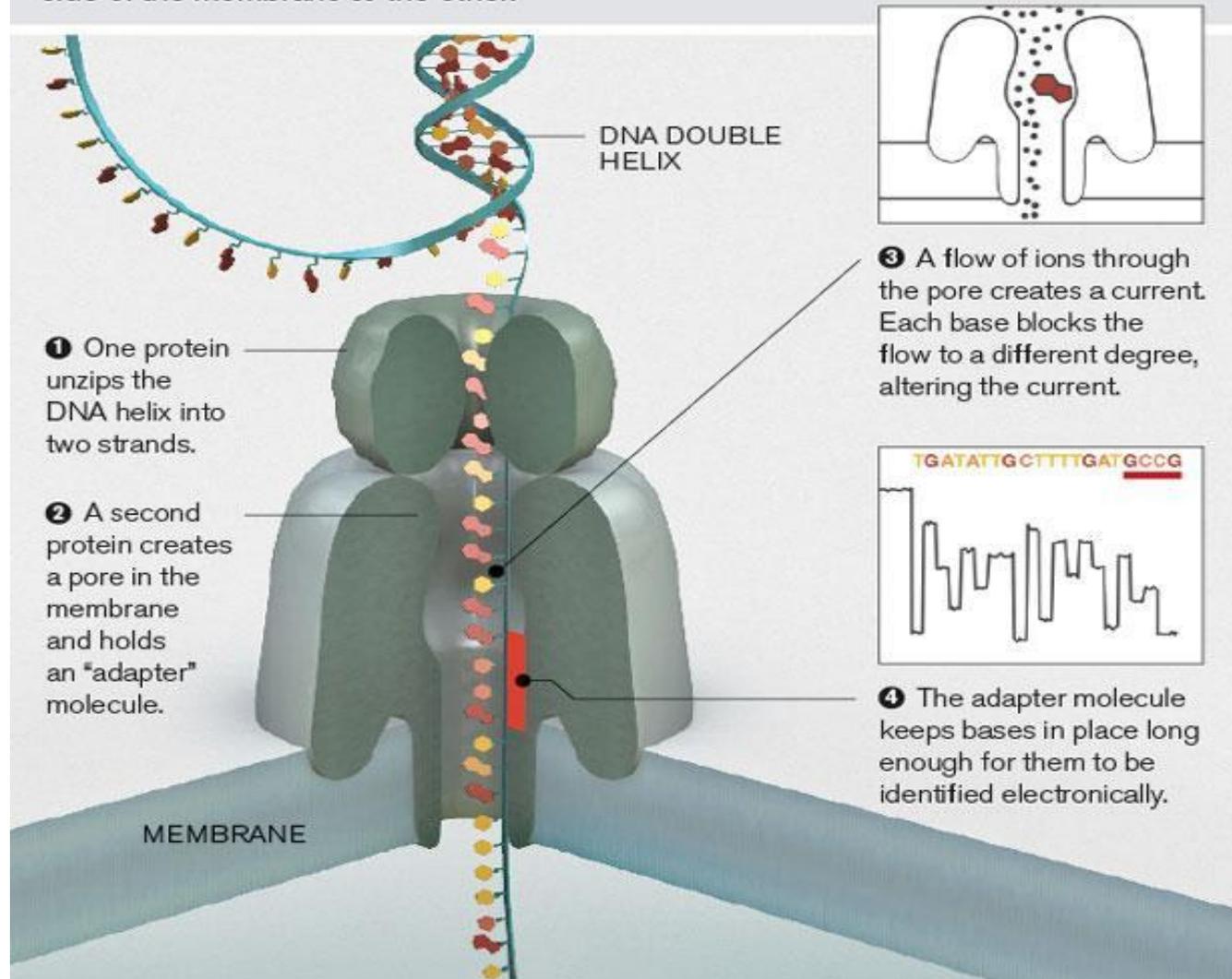
Signal is measured from 5 bases

Timing is irregular

Base modifications do alter the signal

Нанопоровое секвенирование

DNA can be sequenced by threading it through a microscopic pore in a membrane. Bases are identified by the way they affect ions flowing through the pore from one side of the membrane to the other.



Сравнение всех технологий секвенирования

Platform	Sequencer	Costs sequencing platform	Reads per run/lane	Output per run/lane	Maximal read lengths ¹	Average run duration
Sanger	ABI 3730xl	\$100,000	96	100 kbp	1000 bp	2–3 hours
454	GS FLX	\$450,000	1,000,000	700 mpb	1000 bp	24 hours
Illumina	HiSeq 3000	\$750,000	300,000,000 ²	150 gbp ³	250 bp	4 days
Illumina	NextSeq500	\$250,000	400,000,000	120 gbp ³	150 bp	30 hours
Illumina	MiSeq	\$100,000	25,000,000	15 gbp ³	300 bp	24 hours
Ion Torrent	Proton II	\$224,000	330,000,000	66 gbp	200 bp	4 hours
Ion Torrent	PGM 318	\$50,000	5,000,000	2 gbp	400 bp	7 hours
PacBio	RS II	\$700,000	50,000	400 mbp	54 kbp	3 hours
Nanopore	MinION	\$1,000	80,000 ⁴	490 mbp ⁴	150 kbp	n.a. ⁴

Сравнение реально использующихся сегодня технологий секвенирования

	<i>Illumina TruSeq Synthetic Long Reads</i>	<i>Pacific Biosciences</i>	<i>Oxford Nanopore Technologies</i>
Technology	Barcoded & Amplified Synthetic long reads	Single Molecule Real Time Sequencing	Nanopore Sequencing
Mean Length	3-5kbp	10-15kbp	5-10kbp
Raw Error Rate	0.1%	10-15%	10-30%
Costs / GB	~\$2500*	~\$500 [†]	~\$1000 [†]
Time / GB	2-3 days*	2-3 hours	1-2 days

Даже страшно подумать, что будет в этой области лет через 10, глядя на экспоненциальный прогресс технологий секвенирования.



Мы реально живем в
очень интересное
время!