



# **Семинар**

## **Медицинская генетика**

**Фармация Курс 3 ЦИОП «Медицина  
будущего»**

**Технологии высокопроизводительного  
параллельного секвенирования (NGS)**

# NGS И ЕГО ПРИМЕНЕНИЕ В РАЗЛИЧНЫХ ОБЛАСТЯХ БИОЛОГИЧЕСКОЙ НАУКИ

Мария Логачева



**ИЛС**  
ИнтерЛабСервис



# Центр по секвенированию геномов животных по методу Сэнгера

Одна машина может анализировать 96 образцов одновременно (96 капилляров), 750 п.н. за один прогон, 6 прогонов в день.

Итого:  
одна машина может определять **~345 600 п.н.** в один день



# Развитие технологий секвенирования генома

**1953**

Расшифрована структура ДНК  
Д. Уотсон,  
Ф. Крик

**1987**

Первый автоматический секвенатор  
Applied Biosystems



**2005**

Технология пиро-секвенирования  
Д. Ротберг

**2007**

Технология Ion Torrent  
Д. Ротберг

**1977**

Секвенирование методом обрыва цепи  
Ф. Сэнгер

**1997**

Разработка теор основ технологии Solexa Ш. Баласурбаниан  
Д. Кленерман

**2001**

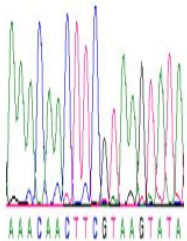
Первый геном человека

**2006**

Технологии Solexa и SOLiD

**2009**

Технология секвенирования Pacific Biosciences



# Технологии секвенирования

## 1-е поколение



Sanger sequencing

## 2-е поколение

454 Life Sciences



Solexa sequencing



Полупроводниковое секвенирование



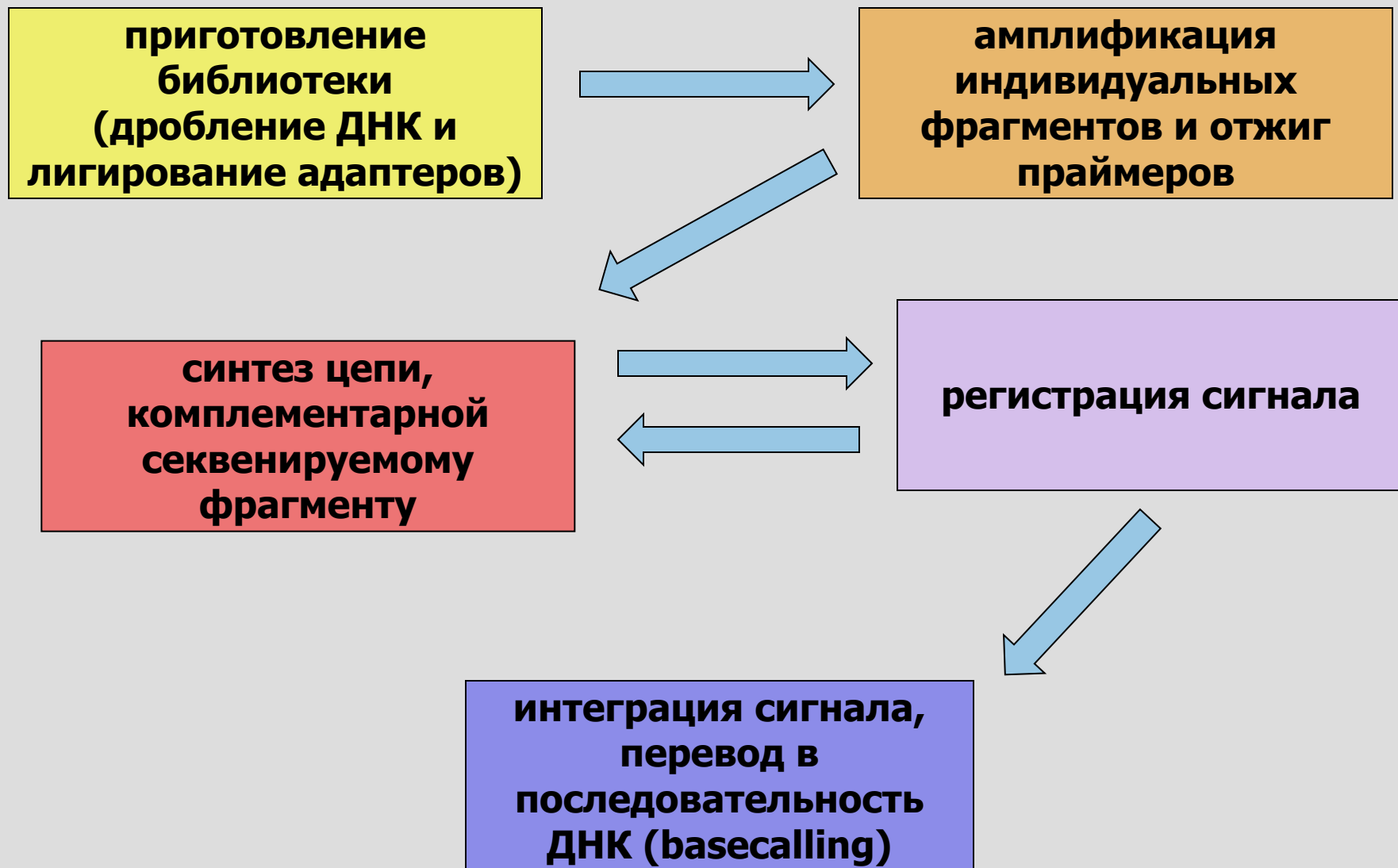
SOLiD



## 3-е поколение

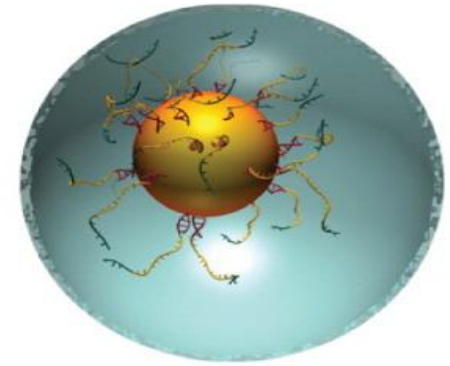
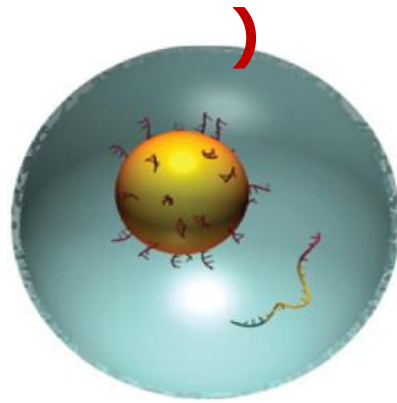
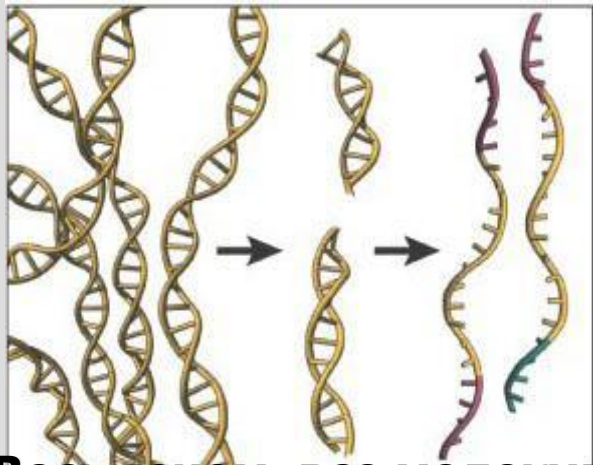


# Общая схема работы NGS: от исходной ДНК до «букв» на экране



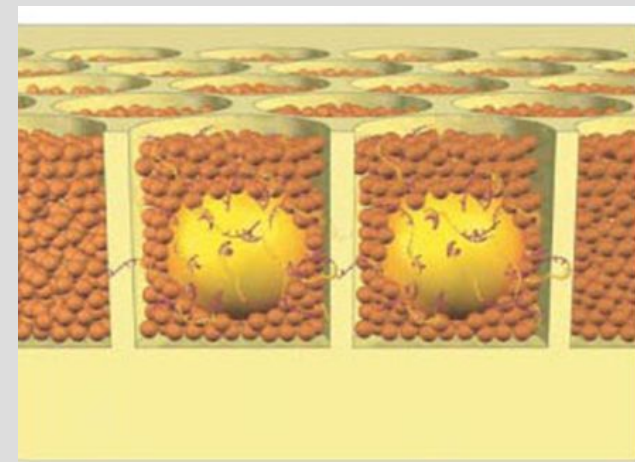
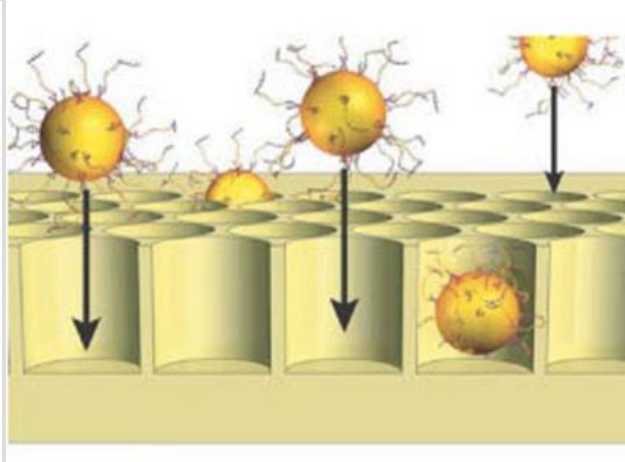
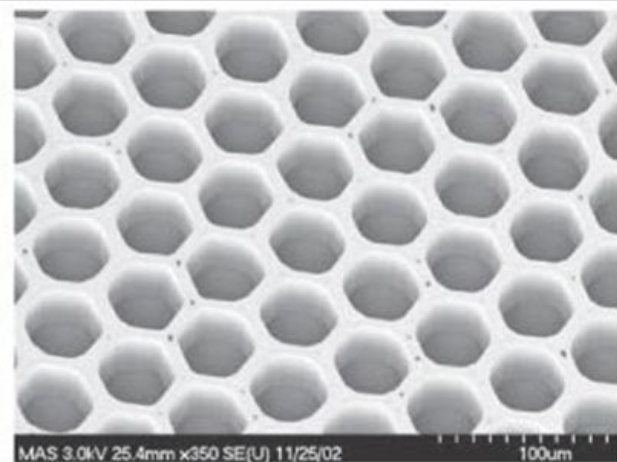


# Высокопроизводительное пиросеквенирование (454Life Sciences)



Весь геном, все молекулы ДНК, случайным образом фрагментируются на фрагменты по 300–500 пар оснований. Комплементарные цепи фрагмента разделяются, к каждой цепи фрагментов пришивается одинаковый для всех олигонуклеотид - «адаптер». Адаптер позволяет отдельным цепям налипать на пластиковые бусинки, которые также имеют одноцепочечные адаптерные последовательности. При этом смесь разъединённых на комплементарные цепи фрагментов разбавляют таким образом, что каждая бусинка получает лишь по одной (!) индивидуальной цепи. Каждая бусинка оказывается заключённой в капельку, окружённую маслом и содержащую смесь для осуществления полимеразной цепной реакции (ПЦР), которая и проходит отдельно в каждой капельке эмульсии (так называемая эмульсионная ПЦР, эПЦР). Это приводит к «клональной амплификации» цепей ДНК, в результате чего на поверхности бусинки удерживается уже не одна, а около 10 млн. копий («клонов») уникальной ДНК-матрицы.



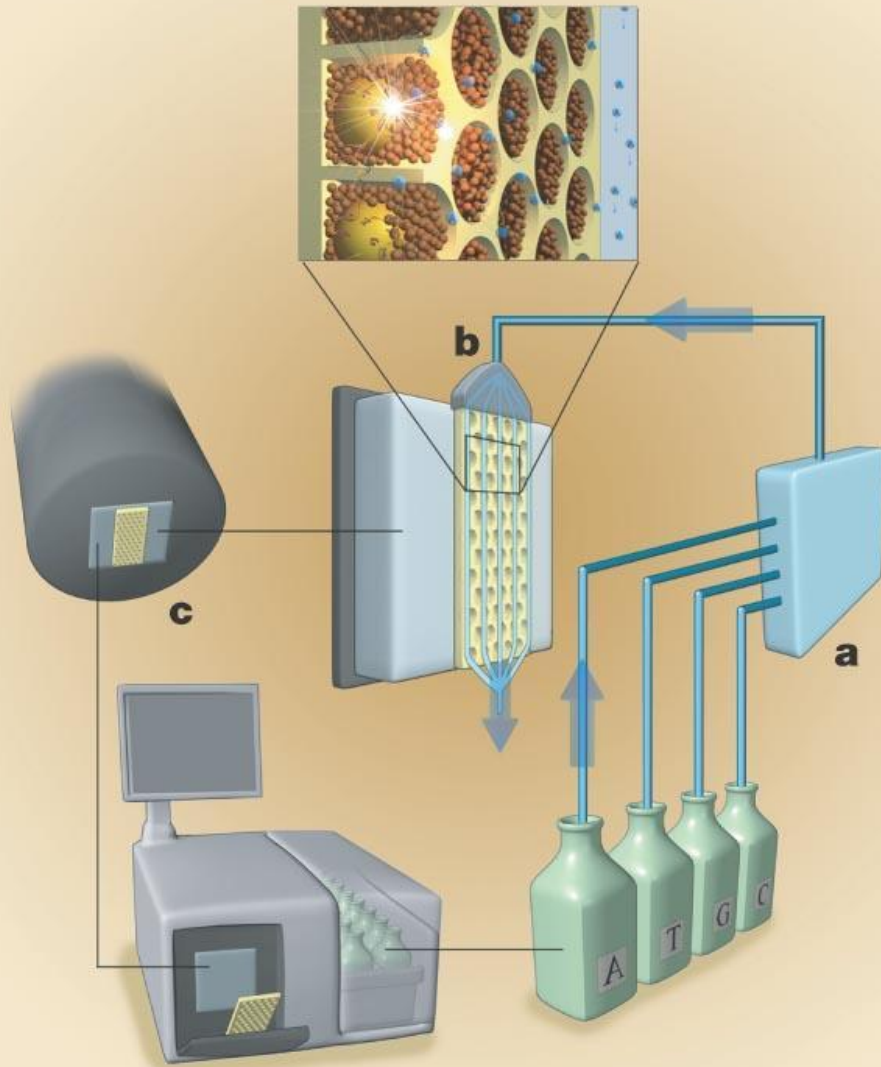


**Далее эмульсия разрушается, и бусинки, несущие одноцепочечные копии ДНК-матрицы, помещаются в лунки «предметного стекла» — слайда особой конструкции. Каждая лунка такого слайда образует отдельный пиколитровый «реактор», в котором и будет происходить реакция секвенирования.**

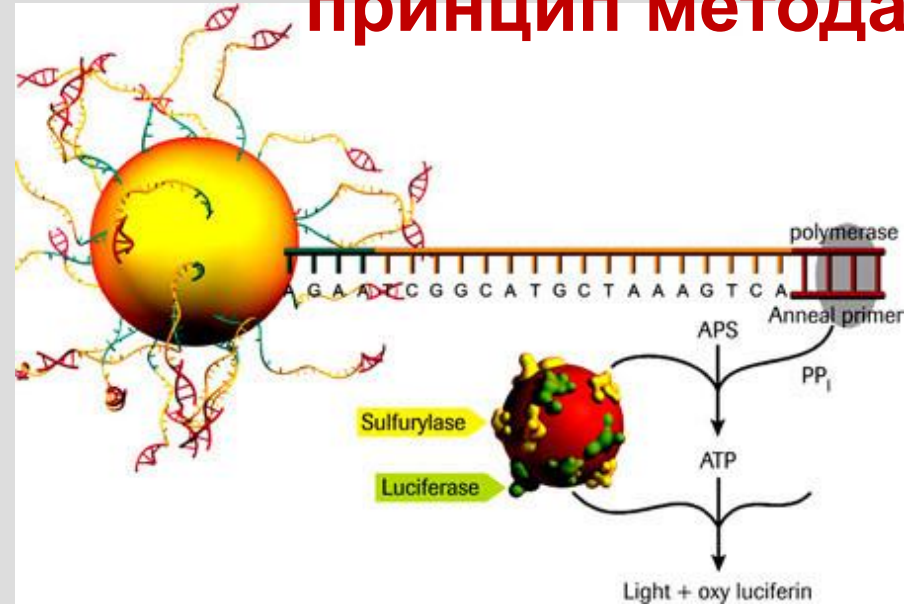
**Каждый слайд несёт около 1,6 миллионов лунок, в каждую из которых попадает одна (!) бусинка с ДНК-матрицей. Слайд помещается в проточную камеру таким образом, что над отверстиями лунок создаётся канал высотой 300 мкм, по которому в лунки поступают необходимые реактивы.**

**Доставляемые в проточную камеру реактивы текут в слое, перпендикулярном оси лунок. Такая конфигурация позволяет одновременно осуществлять реакции на бусинках, несущих ДНК-матрицы, внутри отдельных лунок. Добавление и удаление реагентов и продуктов реакции происходит автоматически, примерно раз в 10 секунд и зависят от высоты проточной камеры и глубины лунок.**

# 454 Life Sciences принцип метода



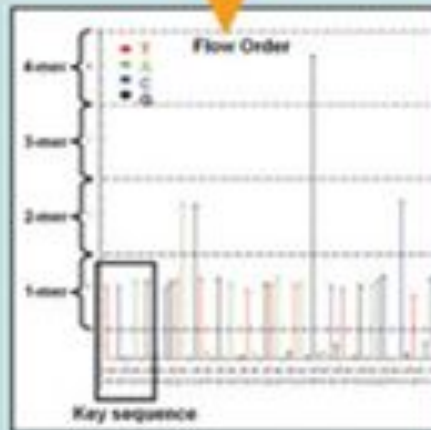
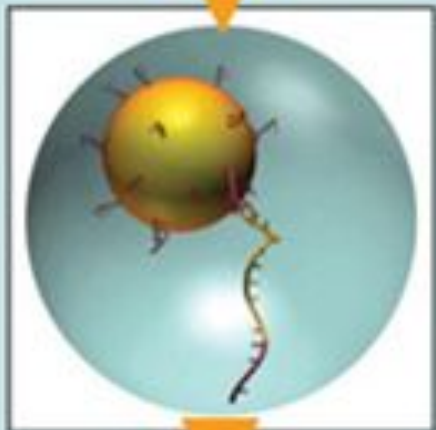
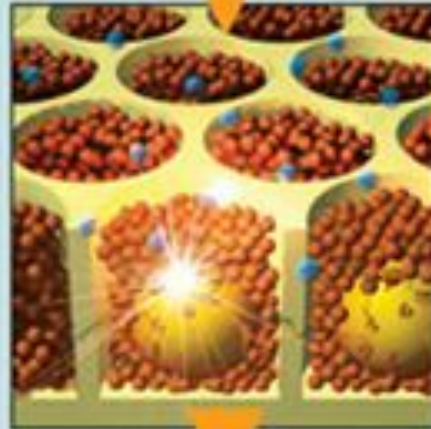
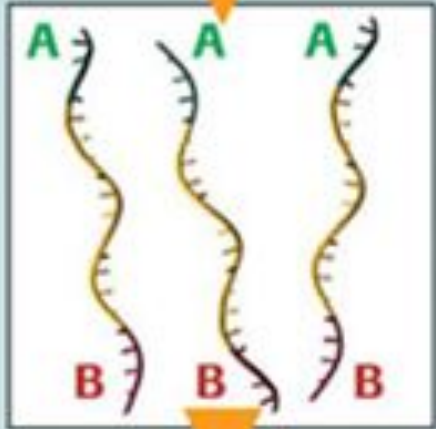
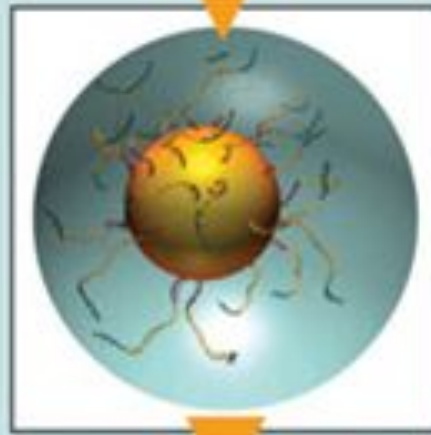
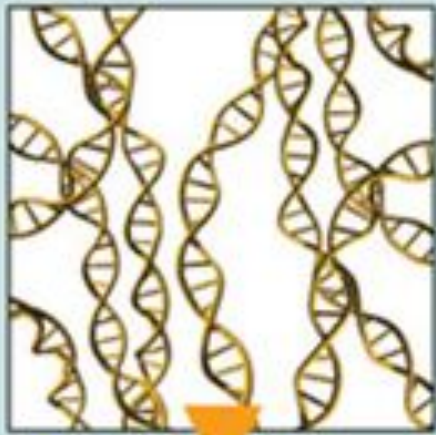
Через лунки в заданном порядке пропускают реагенты (dNTP, люциферин, аденозинфосфосульфат)



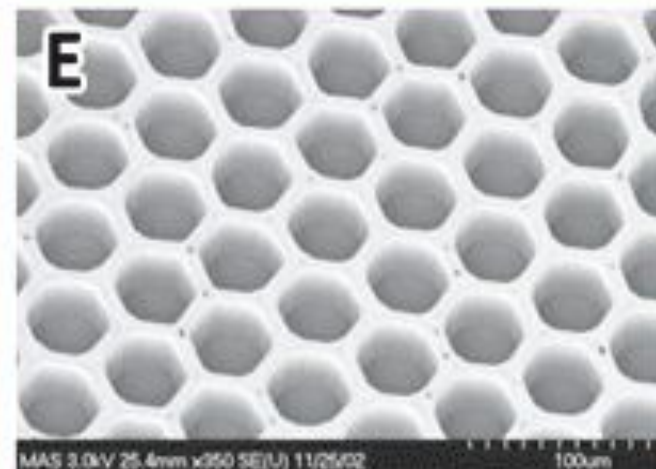
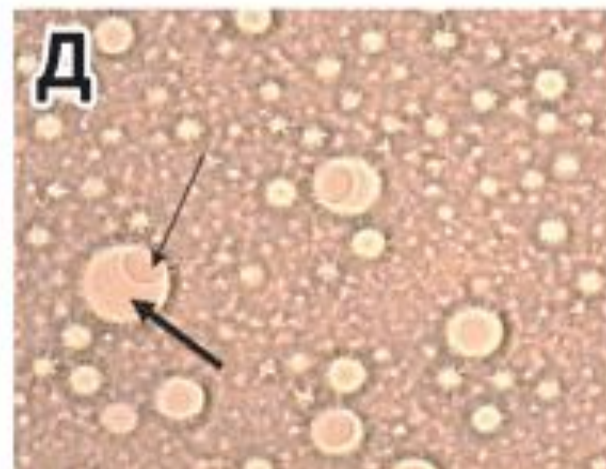
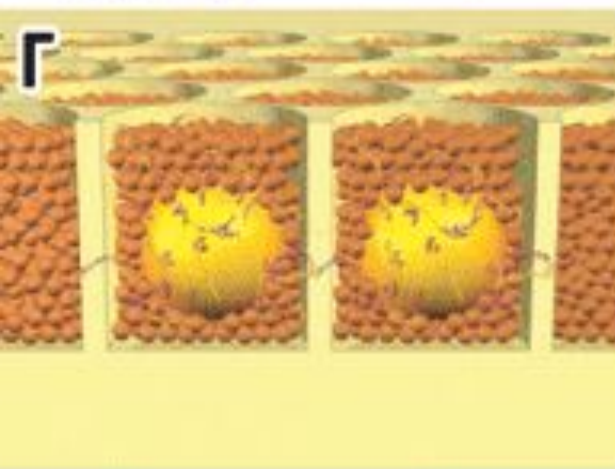
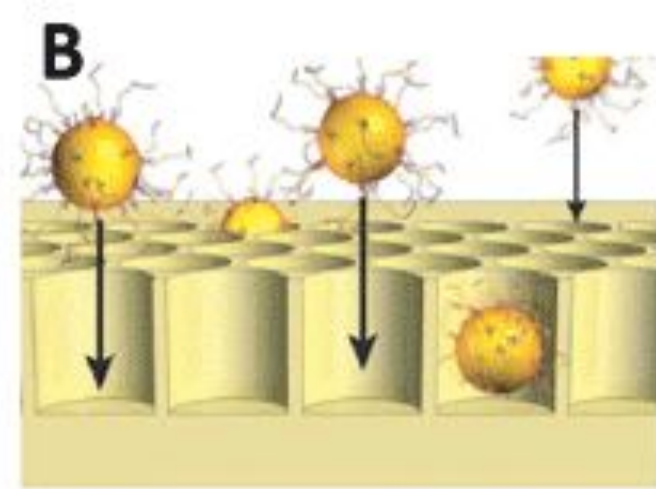
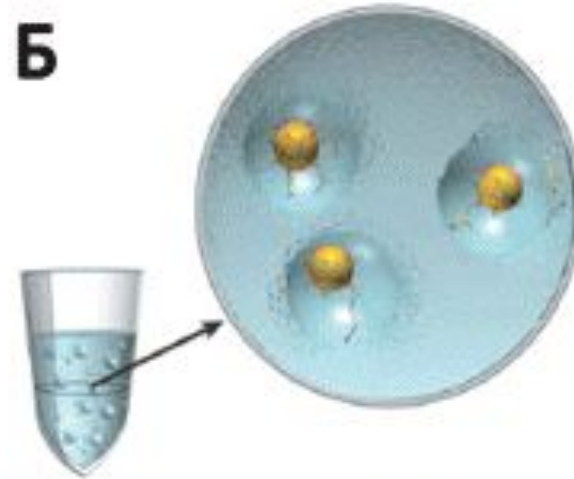
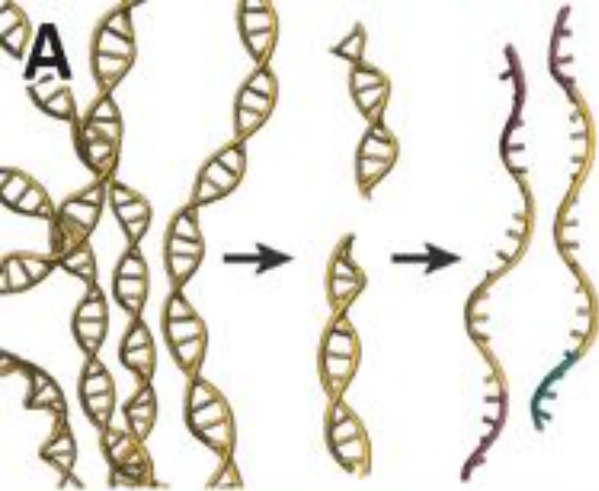
При присоединении dNTP выделяется пирофосфат. Сульфурилаза преобразует пирофосфат + аденозинфосфосульфат в АТФ. АТФ используется для окисления люциферина люциферазой. Световой сигнал регистрируется фотокамерой.

# 454 Life Sciences

Самая первая из технологий NGS (Margulies et al. Nature 2005; 441.7089)







капли реакционной смеси, окруженные маслом. Количество молекул на бусинке увеличивается в миллионы раз в результате эмульсионной полимеразной цепной реакции (эПЦР). В — Эмульсия разбивается, и цепи ДНК-фрагментов, образовавшиеся в результате эПЦР, разделяются. Бусинки, несущие на своей поверхности миллионы одноцепочечных копий первоначального фрагмента ДНК, помещаются в лунки оптико-волоконного слайда, по одной в каждую лунку. Г — В каждую лунку добавляются бусинки поменьше, несущие на своей поверхности ферменты, необходимые для пиросеквенирования. Д — Микрофотография эмульсии, изображающая «пустые» капли и капли, содержащие бусинки с ДНК-матрицей. Толстая стрелка указывает на 100-мкм каплю, тонкая — на 28-мкм бусинку. Е — Микрофотография фрагмента оптико-волоконного слайда, полученная при помощи сканирующего электронного микроскопа.

**Высокая точность расшифровки последовательности достигается тем, что система осуществляет многочисленное прочтение одного и того же фрагмента, что позволяет построить единую обобщённую (так называемую консенсусную) последовательность. Отдельные прочтения одного и того же участка ДНК выравниваются относительно друг друга исходя из интенсивности сигналов в момент протекания через камеру того или иного нуклеотида, а не на основе последовательности этих прочтений. Затем соответствующие сигналы усредняют, и только тогда записывают полученную последовательность. Такой подход значительно улучшает качество расшифровки последовательности и предоставляет возможность оценки её качества**

# Технологии секвенирования 2 поколения: 454

платформа	GS Junior	GS FLX
длина чтения (п.н.)	400	800
число чтений, М	0.1	1
объем данных	35 Мб	700 Мб
цена за запуск/ цена за Мб (в \$)	1 100/22	6 200/7
время работы	10 часов	24 часа
частота и тип ошибок	1%	1%

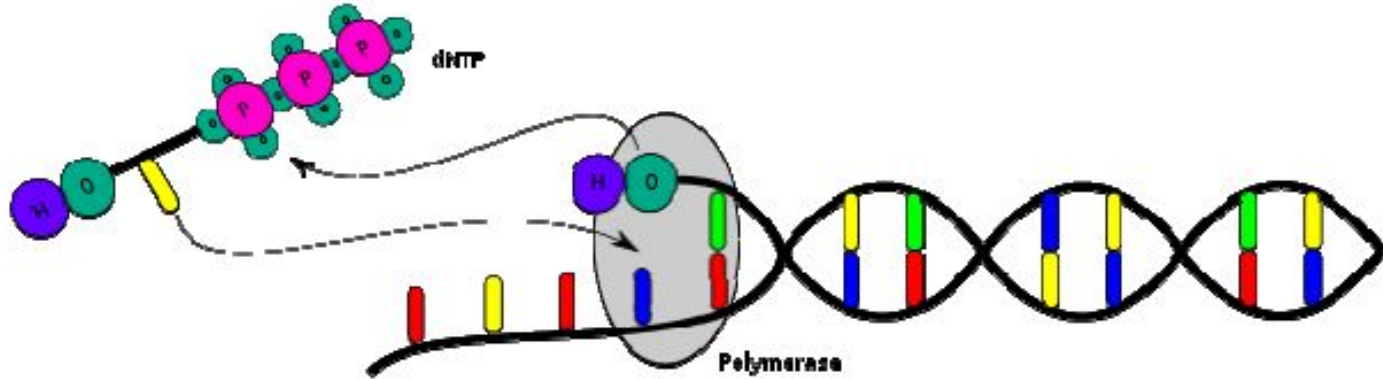
## Преимущества:

- большая длина чтения (сравнимо с секвенированием по Сэнгеру)
- короткое время работы
- наименее чувствительна к GC-составу

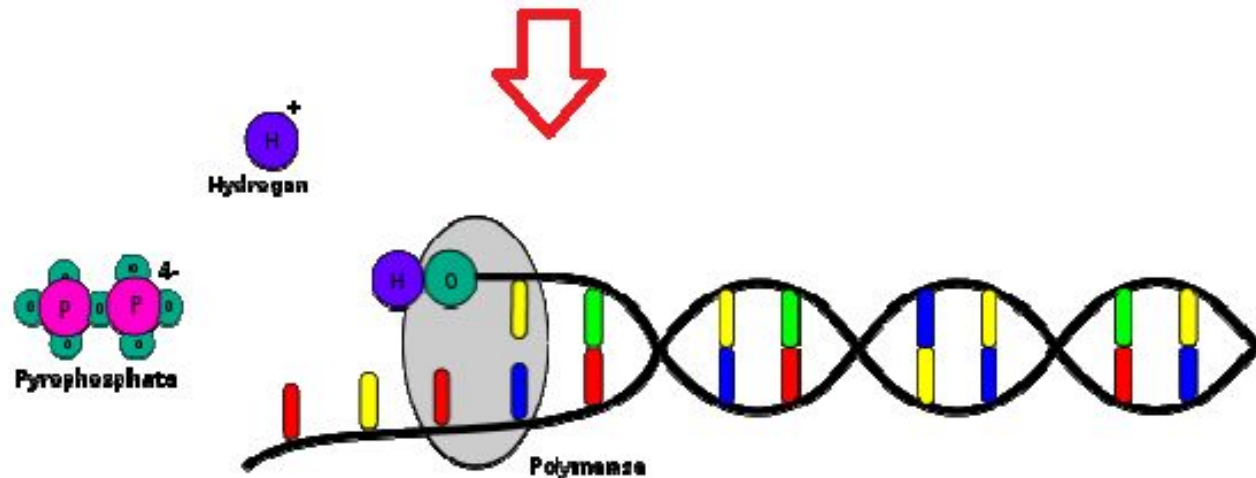
## Недостатки

- неточность прочтения гомополимерных участков
- высокая цена в расчете на нуклеотид
- в 2016 году прекращается поддержка

# Полупроводниковое секвенирование Ion Torrent



**Полимераза встраивает нуклеотид**



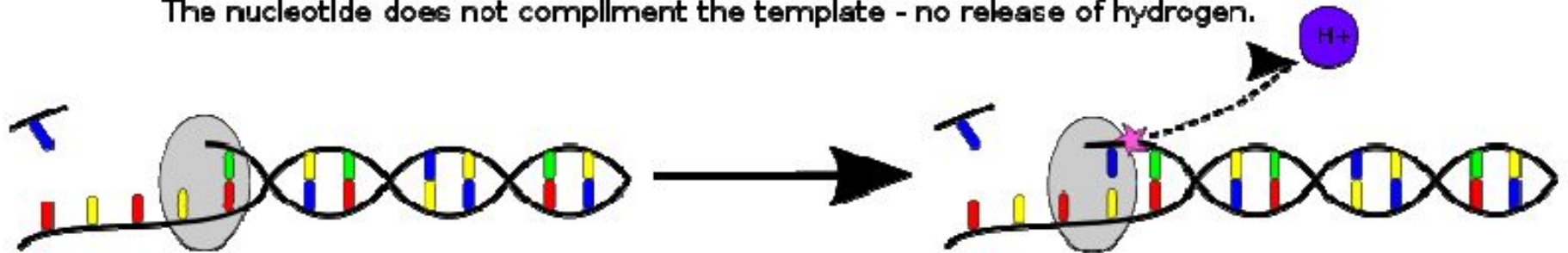
**Включение дезоксирибонуклеотида в растущую цепь ДНК вызывает высвобождение водорода и пирофосфата**



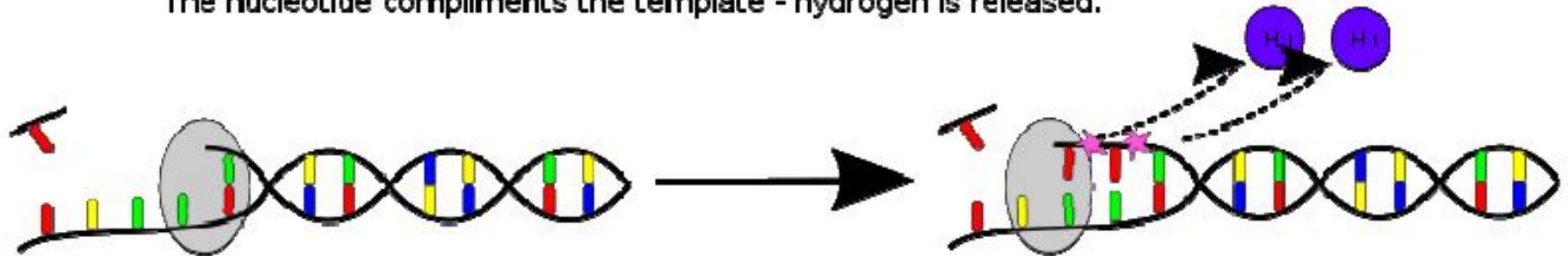
# Полупроводниковое секвенирование Ion Torrent



The nucleotide does not complement the template - no release of hydrogen.

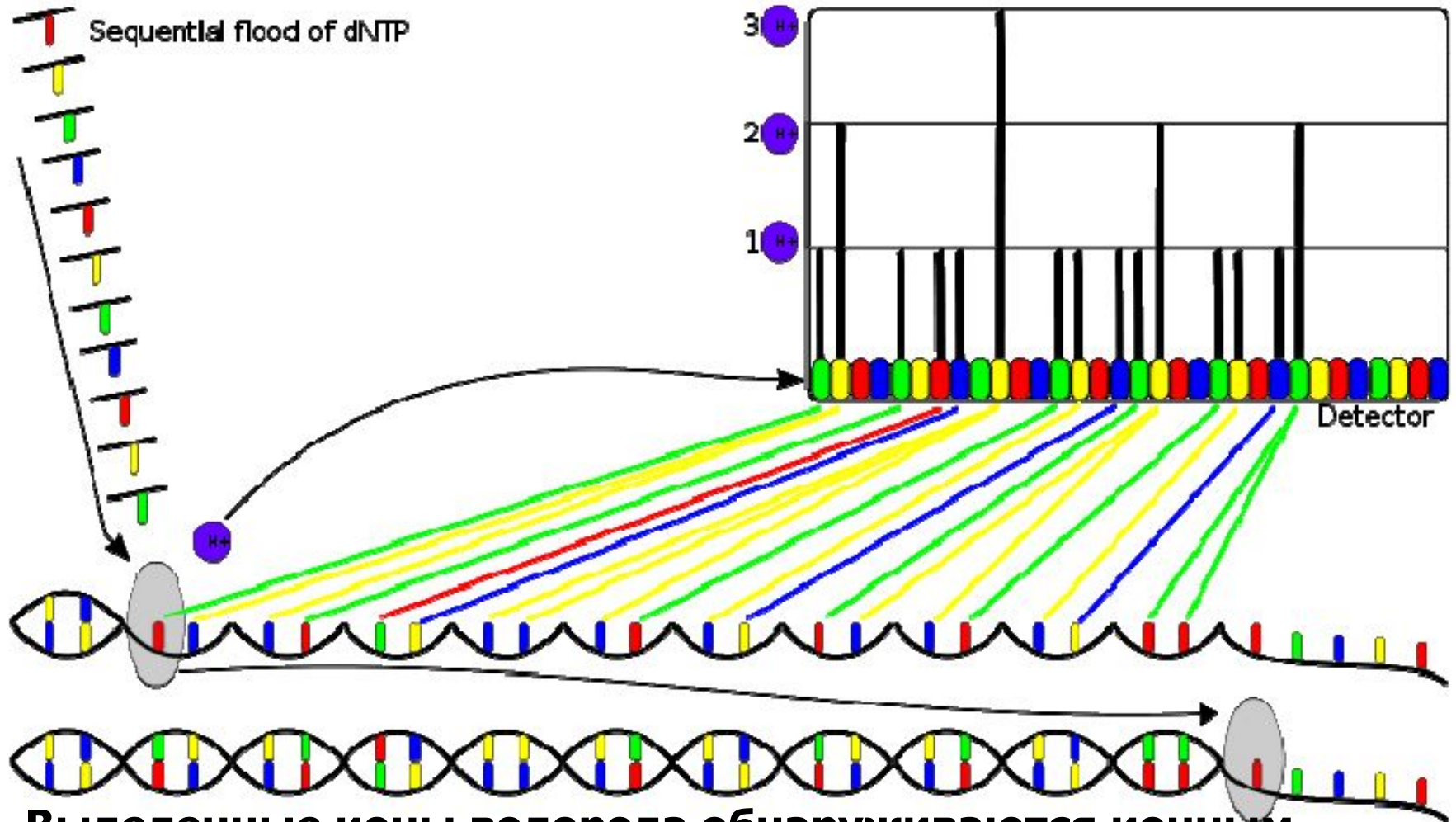


The nucleotide complements the template - hydrogen is released.



The nucleotide complements several bases in a row - multiple hydrogen ions are released.

# Полупроводниковое секвенирование Ion Torrent



**Выделенные ионы водорода обнаруживаются ионным датчиком. Множественное включение приводит к выделению нескольких атомов водорода, что приводит к повышению интенсивности сигнала.**

# Полупроводниковое секвенирование



Ion 314™



Ion 316™



Ion 318™

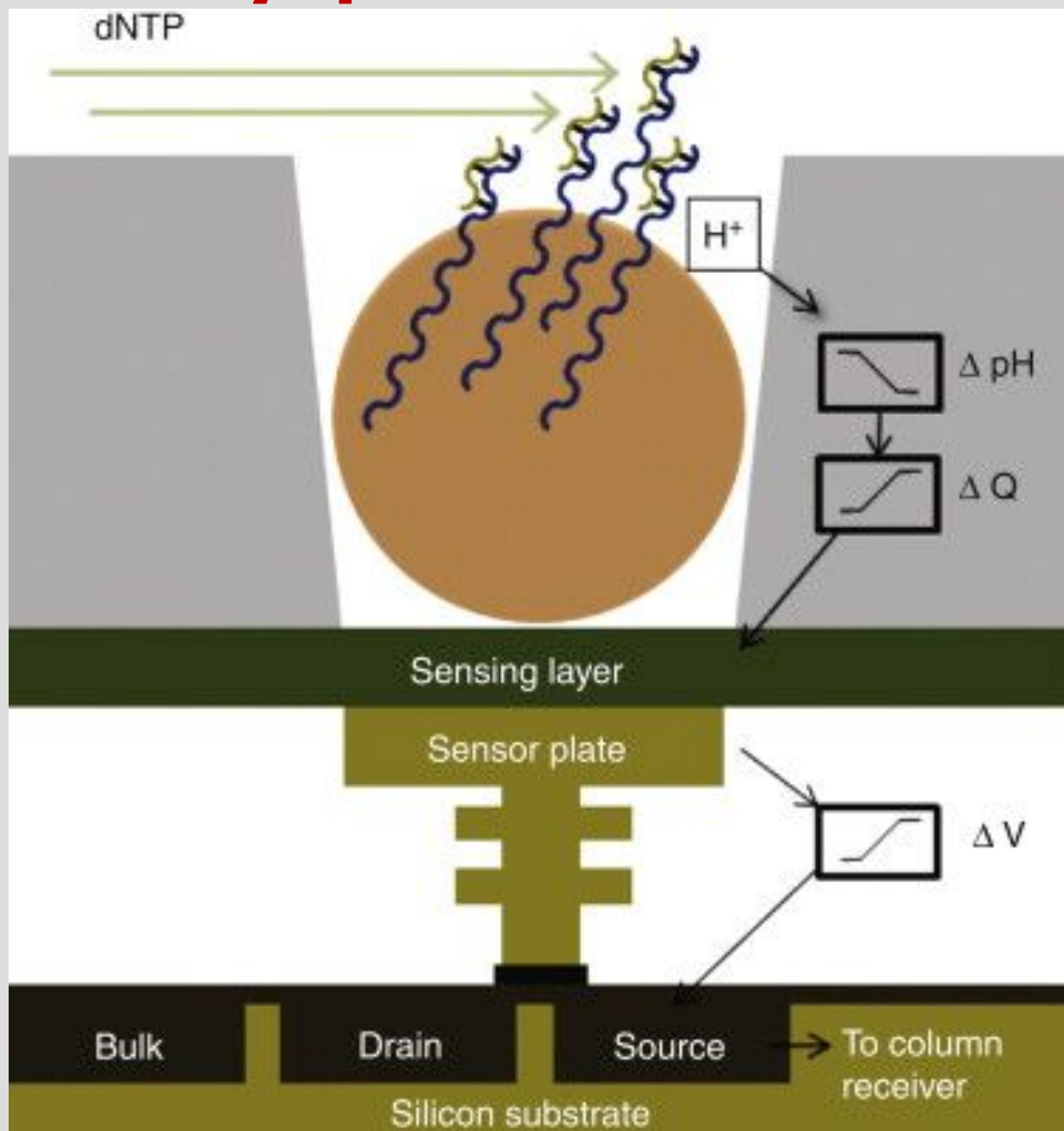


Ion PI™

Первый полупроводниковый секвенатор PGM (Personal Genome Machine, 49,5 тыс.\$), разработанный компанией Ion Torrent, появился в продаже в начале 2011 года. Его начальные характеристики были сравнительно скромными (10...15 Mb), но к 2013 году производительность выросла примерно на два порядка.

В рабочей зоне основания проточной ячейки у сенсорного чипа Ion 314 содержится 1,2 млн. (Мр) таких ячеек (пикселей), у Ion 316 – 6,3 Мр, а у Ion 318 – 11,3 Мр. Благодаря дальнейшему совершенствованию чипов и оптимизации состава реагентов длину ридов удалось увеличить до 400 п.о. и к 2014 году повысить производительность PGM до 2 Gb (с чипом Ion 318 v2).

# Полупроводниковое секвенирование Ion Torrent



# Полупроводниковое секвенирование

Самая новая из технологий секвенирования 2 поколения

Сходно с 454-секвенированием, но регистрируется не свет, а pH

платформа	Ion Torrent	Ion Proton
длина чтения	до 400	200
число чтений, М	4-5.5	60-80
объем данных	2 Гб	12-16 Гб
цена за запуск/цена за Мб (в \$)	939/0.60	1 000/0.02
частота и тип ошибок	0.5-2.5%	0.5-2.5%
время работы	7 часов при длине 400	4 часа

## Преимущества:

- относительно низкая цена за запуск
- быстрота

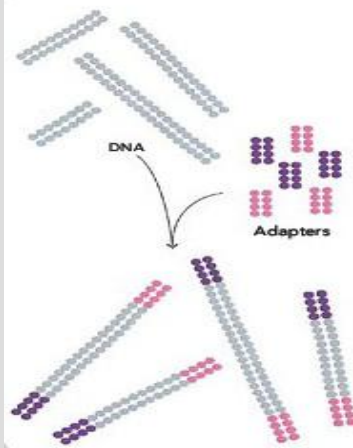
## Недостатки

- невысокая точность прочтения гомополимерных участков
- низкая производительность

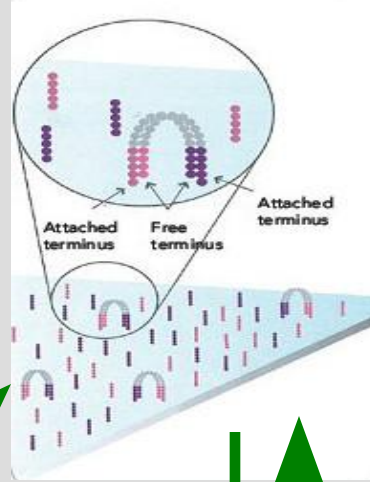


# Секвенирование Illumina: принцип метода

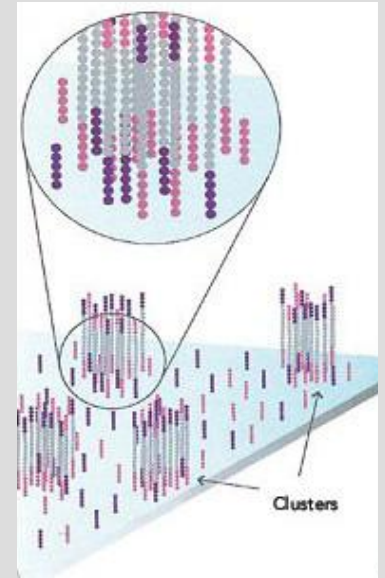
1. ДНК фрагментируют и присоединяют к фрагментам адаптеры



3. Через ячейку пропускают реагенты для достраивания второй цепи ДНК

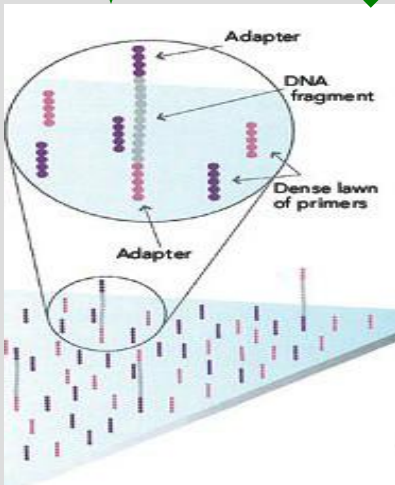


bridge amplification

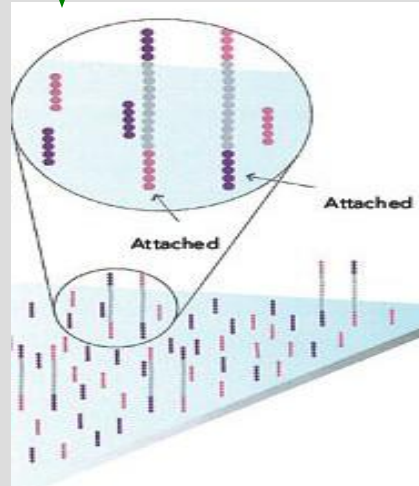


Стадии 3-4 повторяются 30-35 раз

2. ДНК пропускают через каналы ячейки, покрытые праймерами, комплементарными концам адаптеров

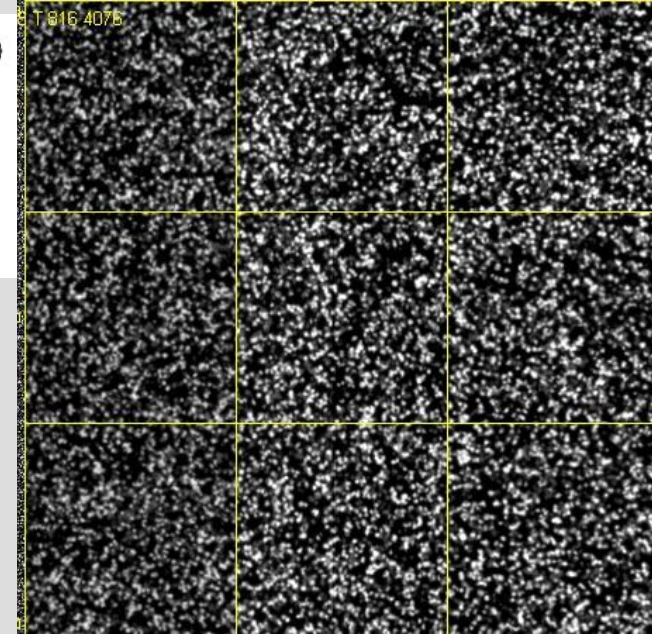
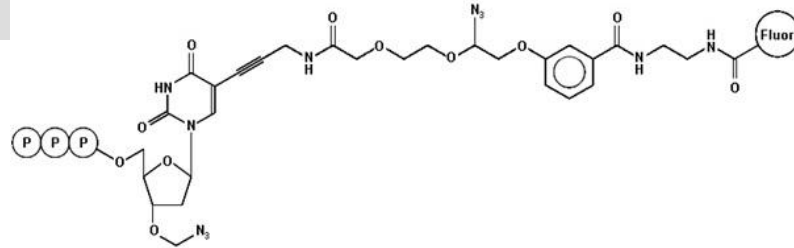
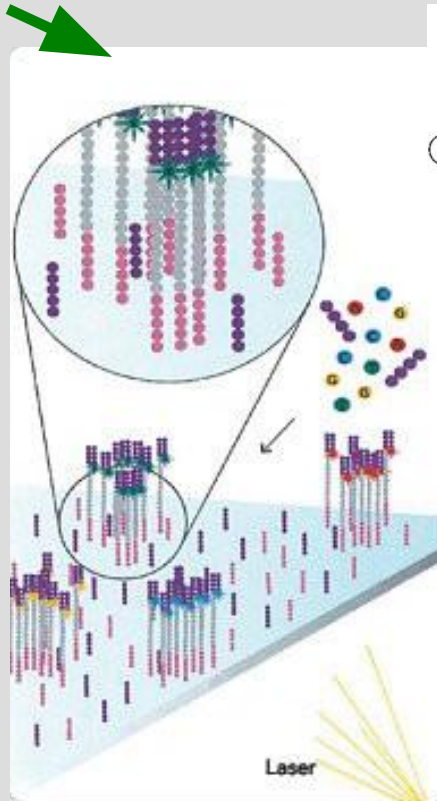


4. Двухцепочечные фрагменты денатурируют



5. Каждый фрагмент оказывается окружен группой идентичных молекул («кластеры»).

# Секвенирование Illumina - принцип метода



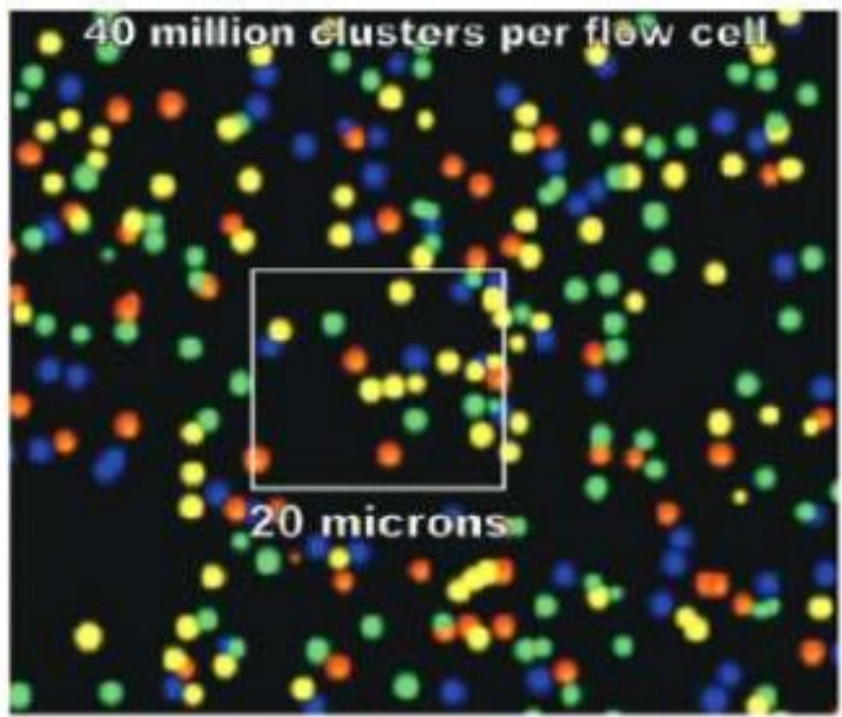
**6. Через ячейку пропускают реагенты (флуоресцентно меченые терминированные dNTP и полимеразу)**

**9. Повторение 7-9  
нужное число раз  
(50-300). Число  
циклов соответствует  
длине чтения.**

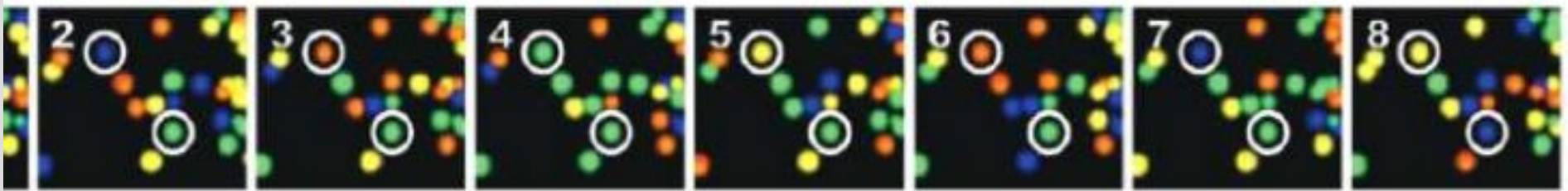
**7. На ячейку светят  
лазером и проводят  
съемку.**

**8. Через ячейку пропускают  
реагенты, отщепляющие  
флуорофор и терминатор**





TGCTACGAT...



TTTTTTTGT...

# Illumina – приборы

Самая распространенная из технологий NGS (~ 80% всех данных)

платформа	HiSeq2000	HiSeq2500	MiSeq
длина чтения	100+100	150+150	300+300
число чтений, М	4 000	600	15-25
объем данных, Гб	1 000	180	15
цена за запуск/цена за Мб (в \$)	23 470/0.04	6 145/0.05	1600/0.14
частота и тип ошибок	0.1% (замены)	0.1% (замены)	0.1-0.5% (замены)
время работы	6 дней	40 часов	65 часов

**HiSeq2000 и HiSeq2500 – модификации одного и того же прибора. MiSeq существует также в варианте MiSeqDx – первый NGS-прибор, разрешенный для использования в диагностике. В начале 2014 г. появились два новых прибора – NextSeq500 и HiSeqX 10**

# Illumina – преимущества и недостатки

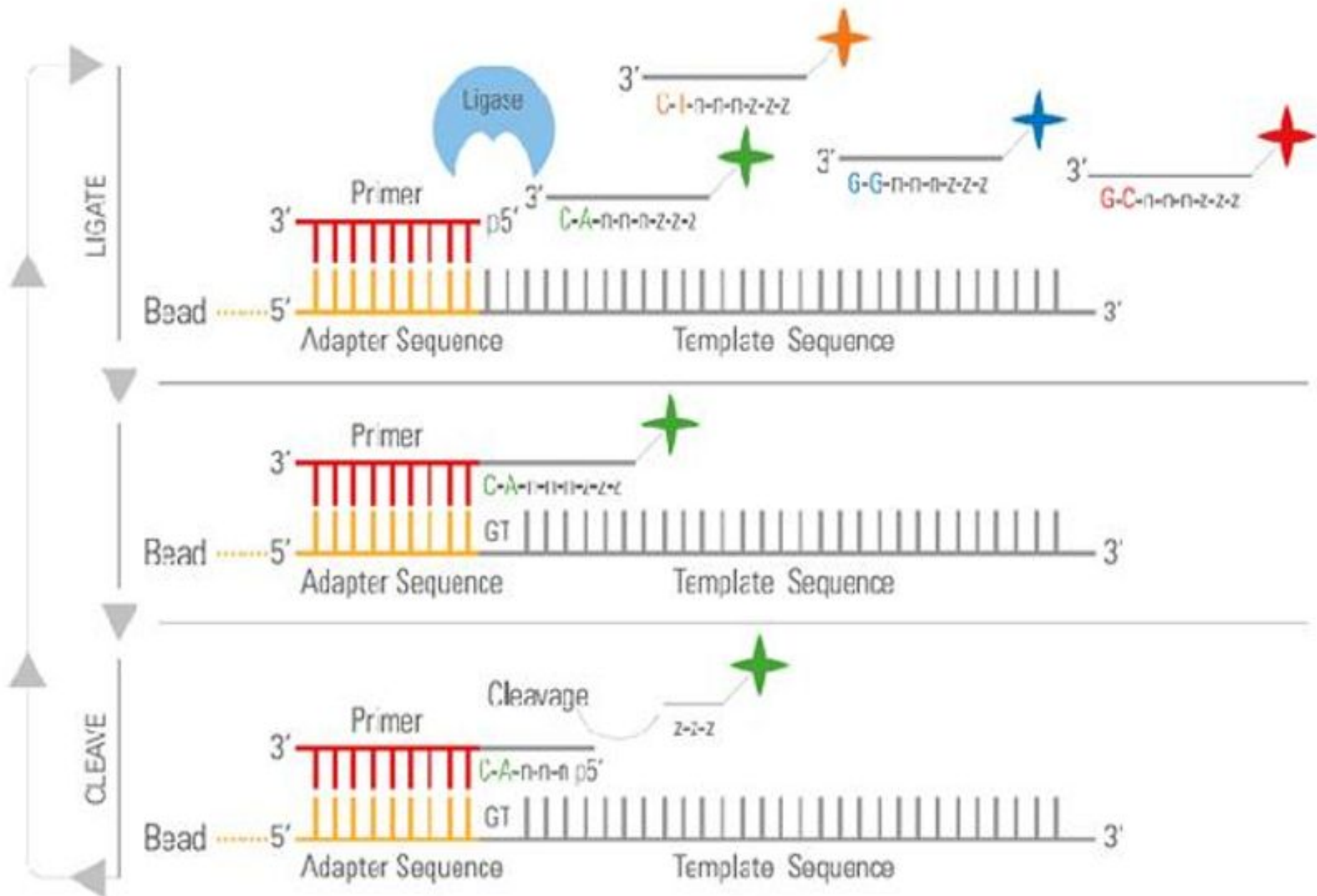
## Преимущества:

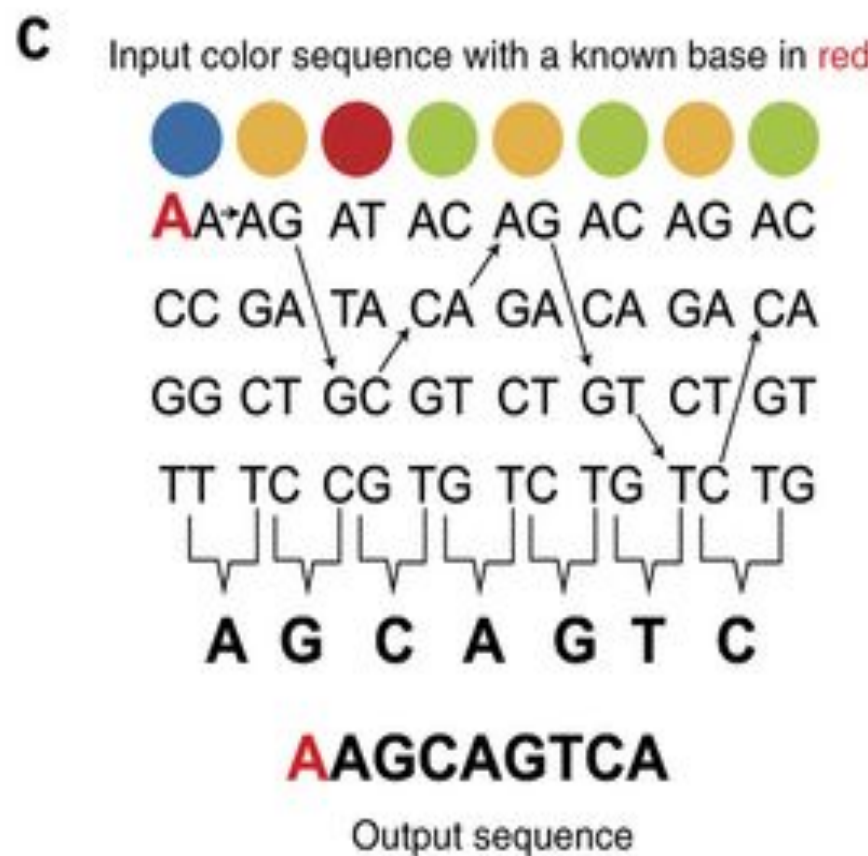
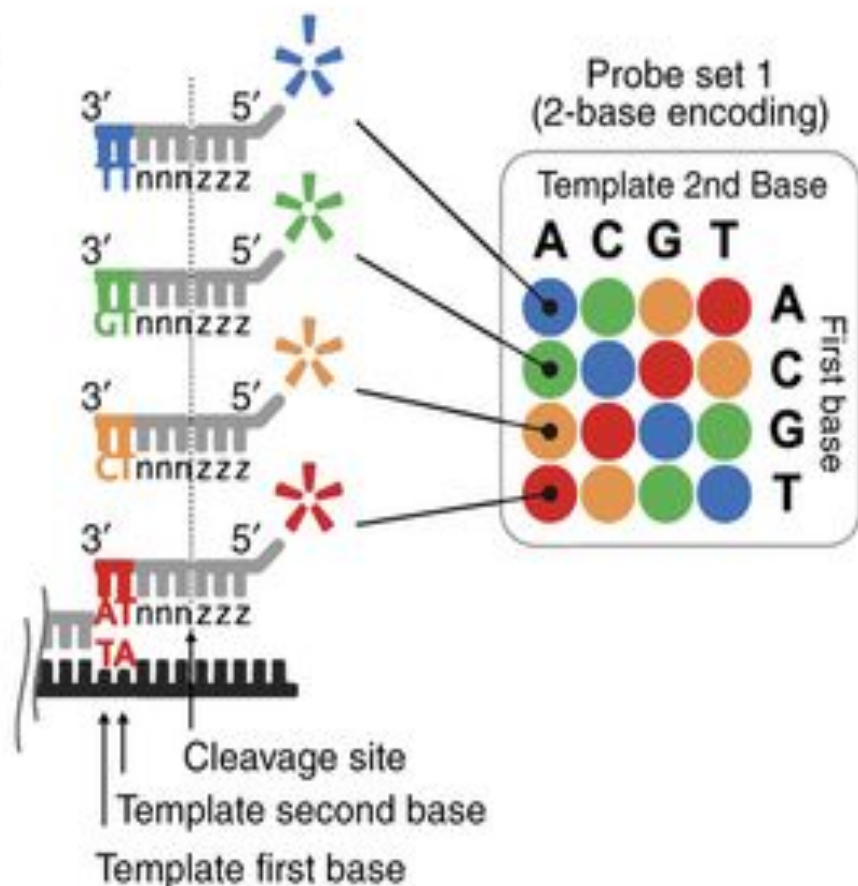
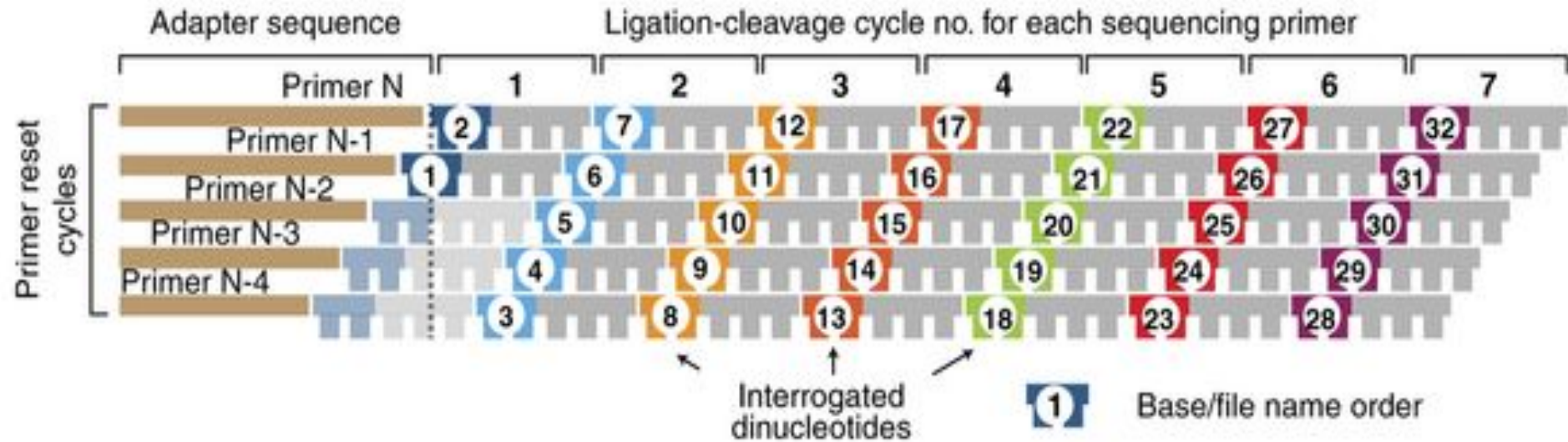
- высокая точность
- универсальность
- доступность ПО для обработки и анализа результатов
- наименьшая цена получаемых данных (в расчете на нуклеотид)

## Недостатки

- высокая цена реагентов
- проблемы с секвенированием матриц с низкой сложностью
- большая длительность прогона
- ошибки в GC-богатых участках

# SOLiD







# Секвенирование путем лигирования (SOLiD)

платформа	Solid 5500
длина чтения	75+35, 60+60
число чтений, М	>1 400
объем данных	150 Гб
цена за запуск/цена за Мб (в \$)	10 503/0.07
время работы	8 дней

## Преимущества:

- высокая точность
- возможность использовать часть дорожек на ячейке

## Недостатки

- очень короткие чтения
- длительность работы
- относительно малая доступность свободного ПО

**Покрытие** – количество прочтений определенного нуклеотида из секвенируемой области нуклеиновой кислоты

**Среднее покрытие** – усредненное количество прочтений определенного нуклеотида из секвенируемой области. Часто термин «покрытие» употребляют в значении «среднее покрытие». Например, «Мы отсеквенировали геном с покрытием 30X».

**Выход секвенирования** – суммарный размер произведенных сиквенс-данных. Измеряется в млн. или млрд п.н. (англ. Mb, Gb) • Выход зависит от используемой технологии и методики секвенирования • Выход определяет диагностические возможности прибора и производительность (кол-во образцов за запуск)



# Что такое покрытие (coverage)

Это сколько раз в среднем нуклеотид генома покрыт ридами

Multiple Copies of a Genome



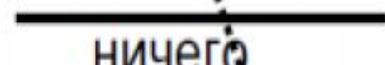
Reads



High Coverage



Low Coverage



ничего



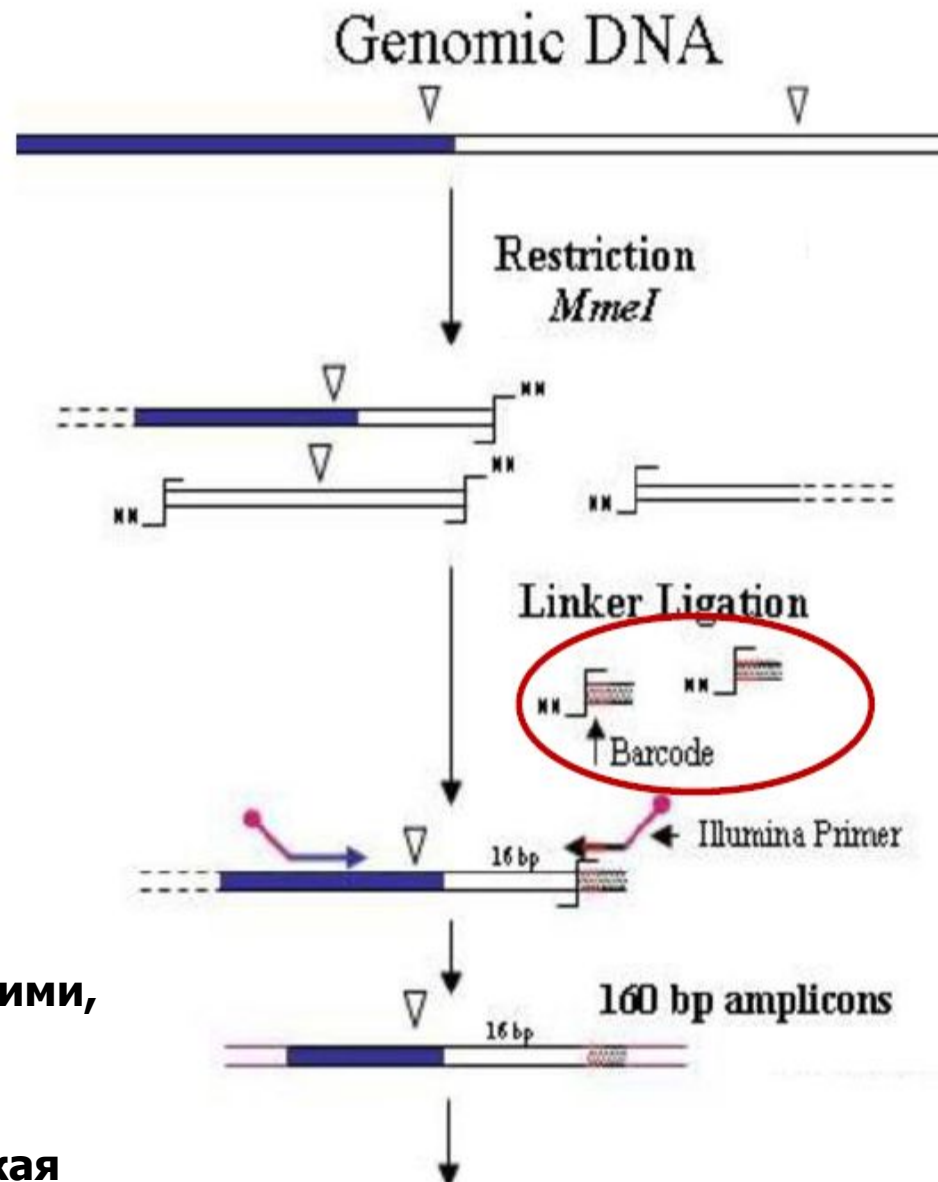
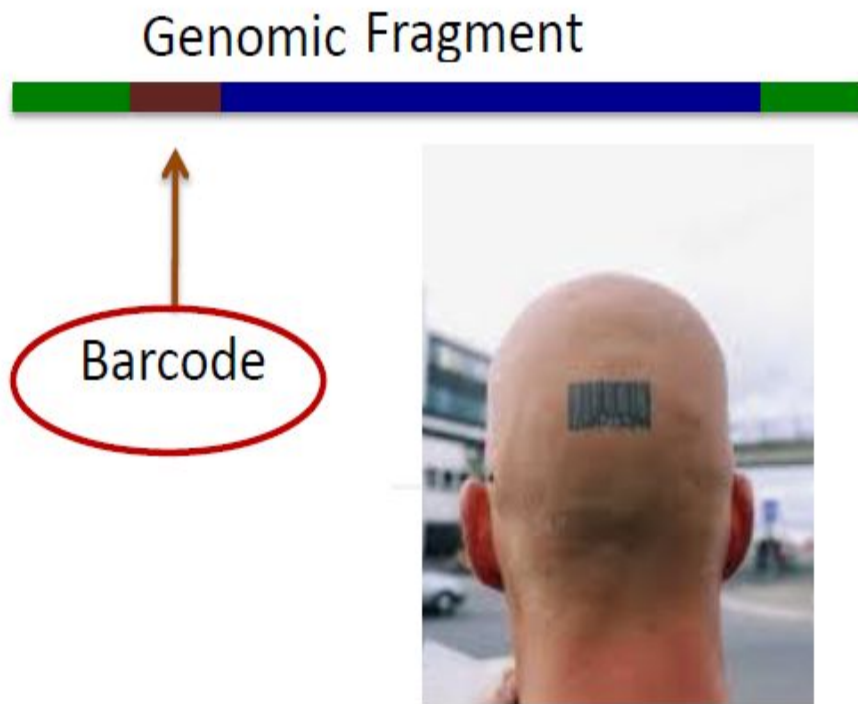
Consensus Sequence



Плохое покрытие

Нет ридов — нет покрытия  
(не отсеквенировалось)

Хорошее  
Покрытие  
(50-100X  
для *de novo*)



Баркод — это своего рода штрих код, только в штрих коде информация записана в одном измерении за счет толщины полос и расстояния между ними, а в баркоде в двух измерениях, как в матрице.

Баркод в секвенировании — это короткая последовательность, которая позволяет разделять секвенируемые фрагменты разных индивидуумов

	<b>Ion Torrent</b>	<b>454 Sequencing</b>	<b>Illumina</b>	<b>SOLiD</b>
Sequencing Chemistry	Ion semiconductor sequencing	Pyrosequencing	Polymerase-based sequence-by-synthesis	Ligation-based sequencing
Amplification approach	Emulsion PCR	Emulsion PCR	Bridge amplification	Emulsion PCR
Mb per run	100	100	600,000	170,000
Time per run	1.5 hours	7 hours	9 days	9 days
Read length	200 bp	400 bp	2x100 bp	35x75 bp
Cost per run	\$ 350 USD	\$ 8,438 USD	\$ 20,000 USD	\$ 4,000 USD
Cost per Mb	\$ 5.00 USD	\$ 84.39 USD	\$ 0.03 USD	\$ 0.04 USD
Cost per instrument	\$ 50,000 USD	\$ 500,000 USD	\$ 600,000 USD	\$ 595,000 USD

# Первое правило геномного секвенирования



*Trash In - Trash Out*

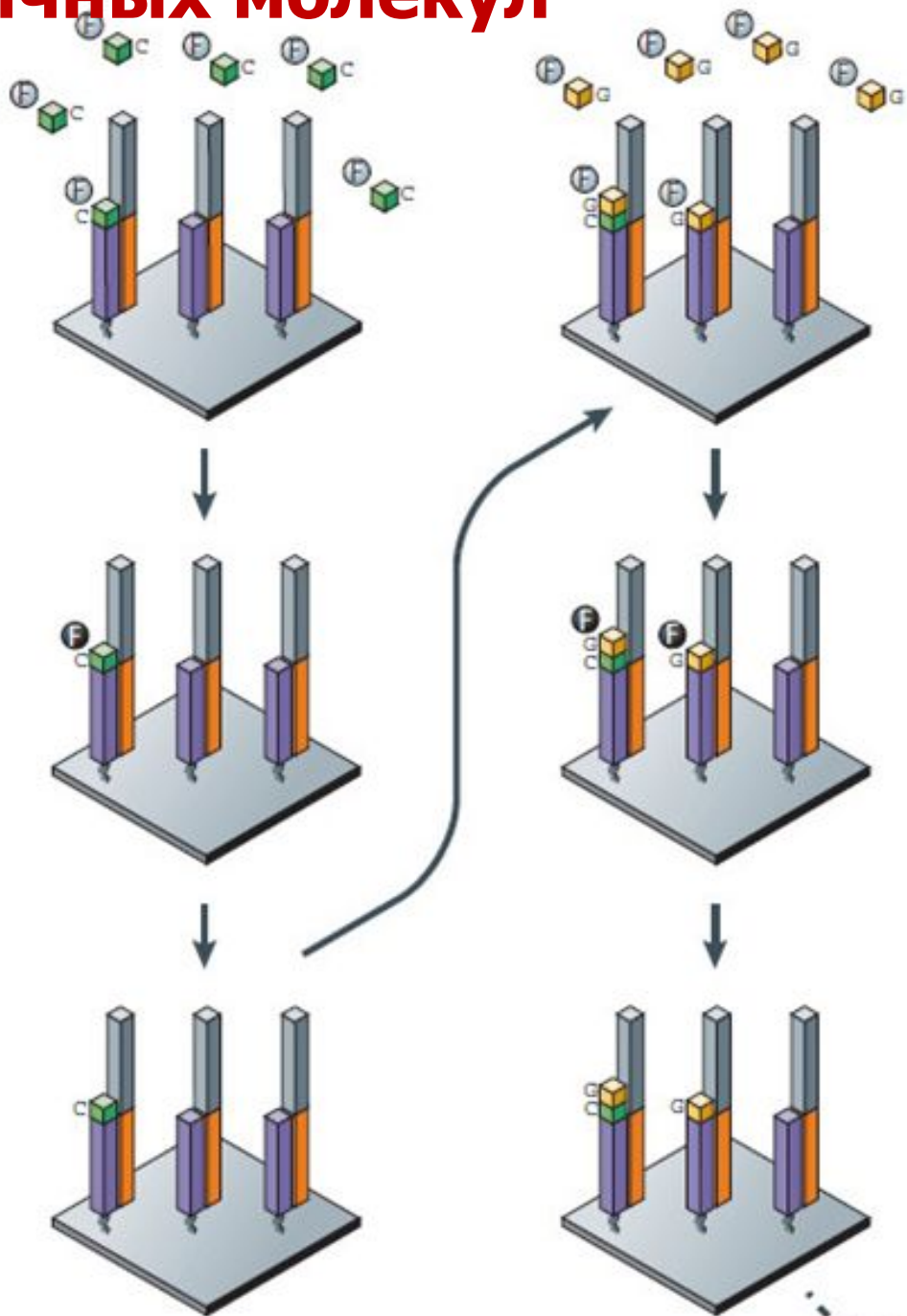




# Секвенирование единичных молекул

- простая пробоподготовка - фрагментация ДНК и аденилирование фрагментов. Затем фрагменты закрепляются на ячейке с олиго-dT
- принцип секвенирования сходен с Illumina – используются флуоресцентно меченые обратимо терминированные нуклеотиды (один тип нуклеотидов за цикл)
- небольшая (до 50 пн) длина чтения, много ошибок (3-5%)

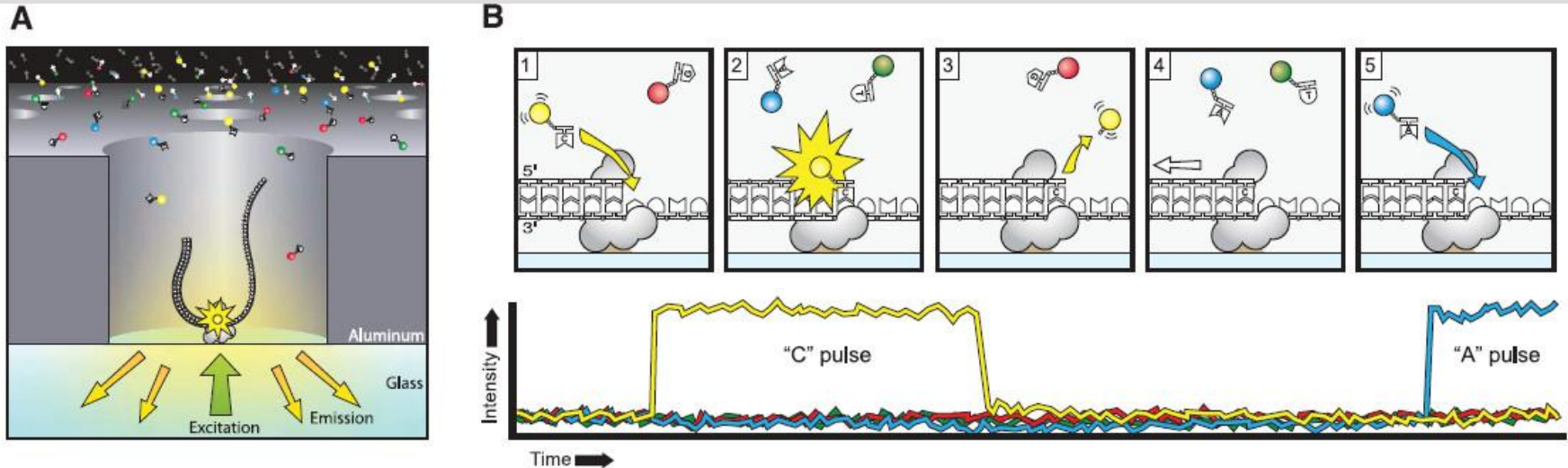
Область применения – высокоточный анализ экспрессии, оценка копийности участков генома



# Секвенирование единичных молекул.

## Pacific Biosciences (SMRT sequencing)

Используется полимераза, иммобилизованная в 100-нм лунках, и флуоресцентно меченые dNTP. Возможны очень длинные чтения (> 10 000), но высокая частота ошибок (до 10%). Возможно прямое секвенирование метилированной ДНК, ведется работа над разработкой прямого секвенирования РНК.



Области применения – де novo сборка (в сочетании с Illumina для коррекции ошибок), анализ изоформ, поиск модифицированных оснований

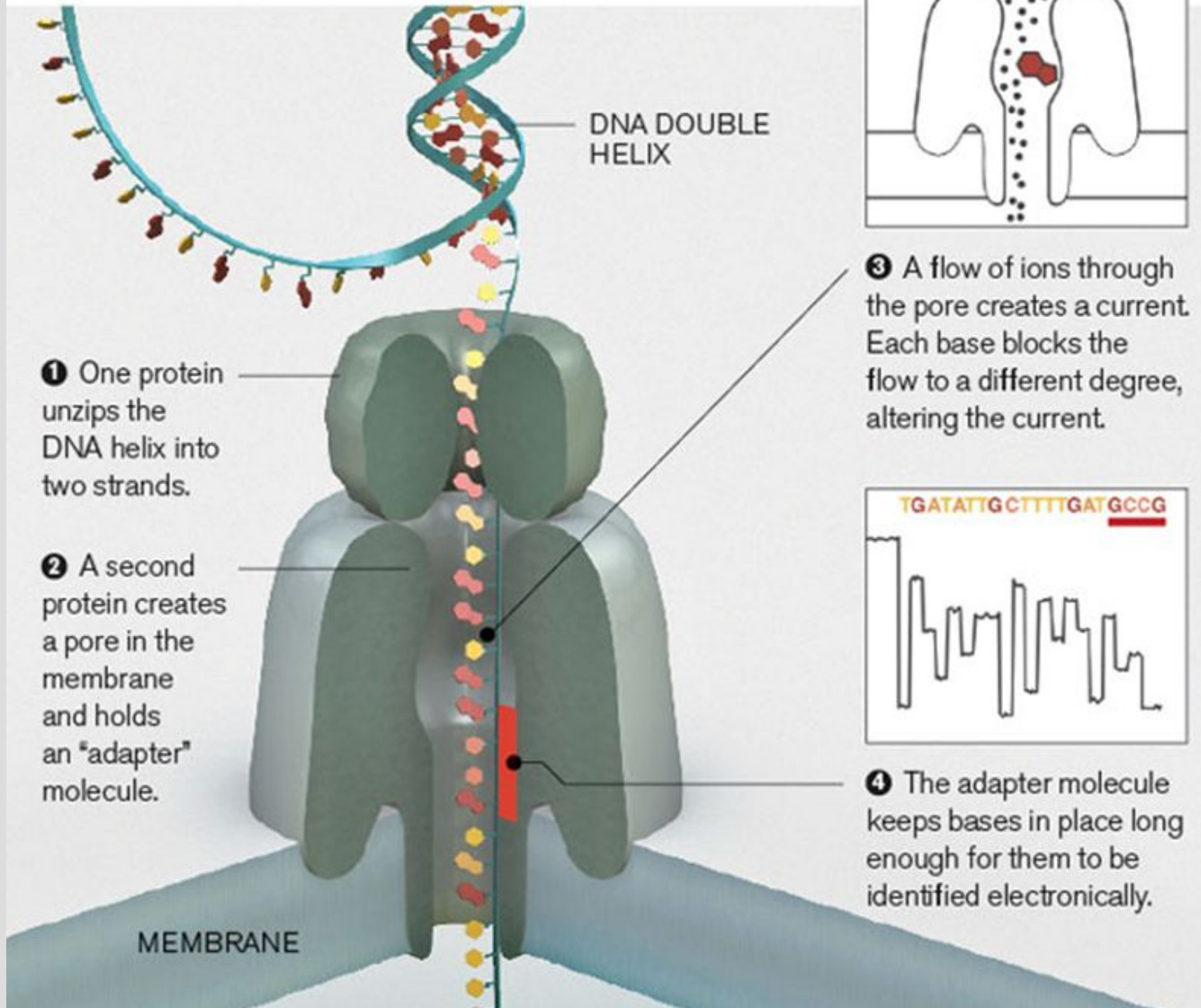
# Секвенирование единичных молекул.

## Oxford Nanopore

Принцип основан на использовании мембран с белковыми нанопорами, через которые протягивается молекула ДНК. Секвенатор размером с USB-диск. Высокая скорость, очень высокая частота ошибок, низкая производительность (пока!)







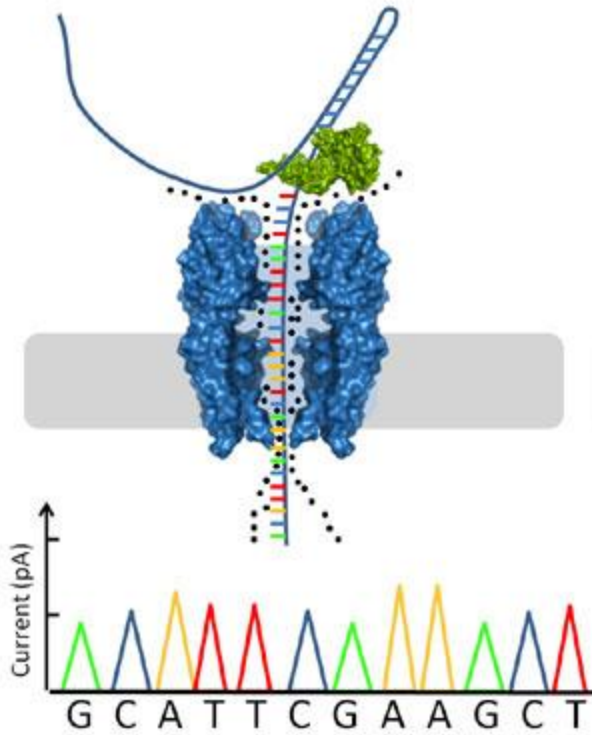
Нанопоровые системы представляют собой реакционную камеру, внутри которой находится **раствор электролита**. Камера разделена на две части липидной мембранной или иной тонкой непроводящей поверхностью, в которую внедрена **единичная нанопора**.

К частям камеры прикладывают напряжение, из-за чего возникает **ток ионов через пору**. Когда исследуемые молекулы проходят через пору по направлению поля, они уменьшают сечение, доступное для ионов, и сила тока падает. Анализируя изменение силы тока, можно определить свойства молекулы, проходящей через пору.

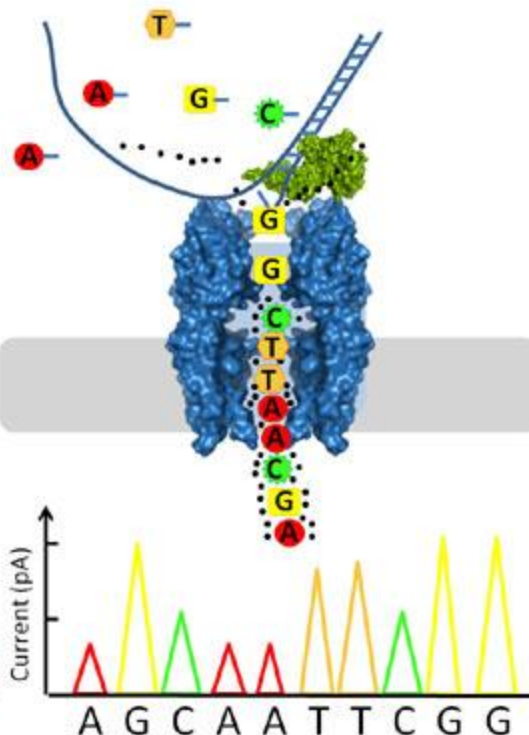
В случае нуклеиновых кислот, диаметр используемых нанопор составляет несколько нанометров, из-за чего ДНК и РНК способны проходить сквозь пору только в одноцепочечной форме, но не в двухцепочечной. При прохождении молекулы нуклеиновой кислоты через пору отдельные нуклеотиды задерживаются в определенных сайтах внутри поры, в результате чего происходит измеримое падение силы тока.

# Nanopore Sequencing Technologies

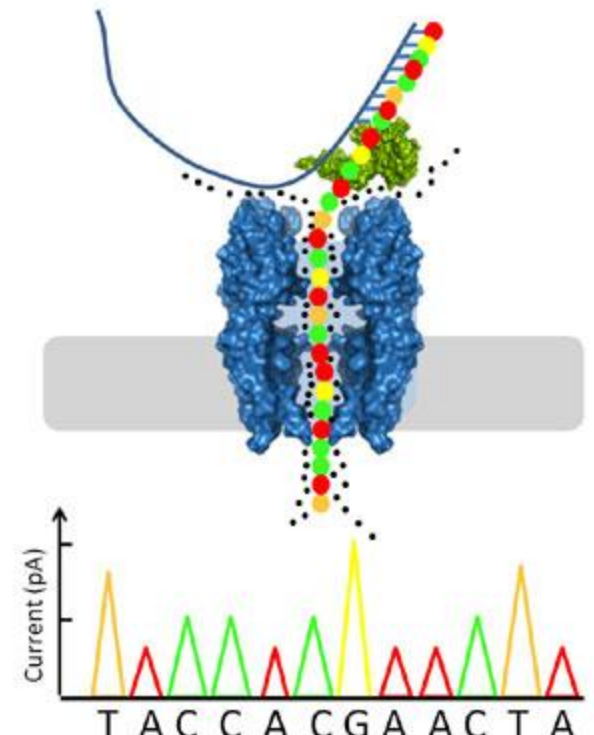
## A) Oxford Nanopore Technologies



## B) Genia Technology



## C) Stratos Genomics



Semiconductivity based measurements of current flow changes are translated into DNA sequence information

# Области применения NGS

## What can next generation sequencing do for you?

- **секвенирование геномов и транскриптомов de novo**

отправная точка большинства молекулярно-биологических и генетических исследований на немодельных объектах, поиск крупных геномных перестроек

- **полногеномное ресеквенирование**

поиск мутаций, ассоциированных с болезнями, картирование генов, геномы отдельных типов клеток, анализ древней ДНК

- **направленное ресеквенирование**

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

- **анализ транскриптома**

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

- **ДНК-белковые и ДНК-ДНКовые взаимодействия**

поиск сайтов связывания транскрипционных факторов, изучение пространственной организации хроматина

- **метагеномика**

анализ разнообразия микробных сообществ

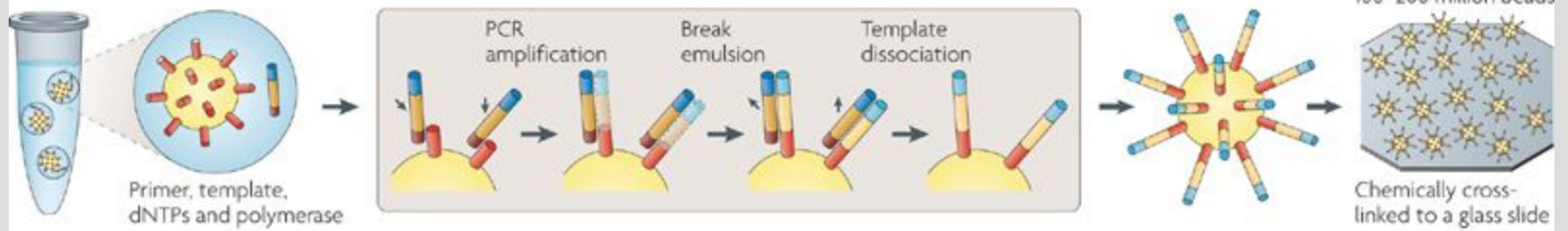
# Секвенирование 2 поколения: области применения

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	+	-



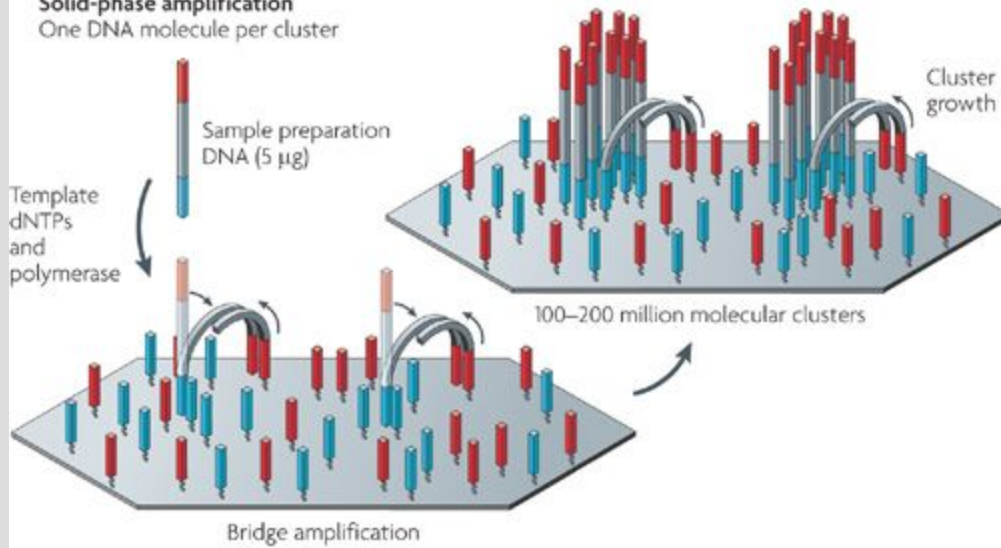
**a Roche/454, Life/APG, Polonator**  
**Emulsion PCR**

One DNA molecule per bead. Clonal amplification to thousands of copies occurs in microreactors in an emulsion

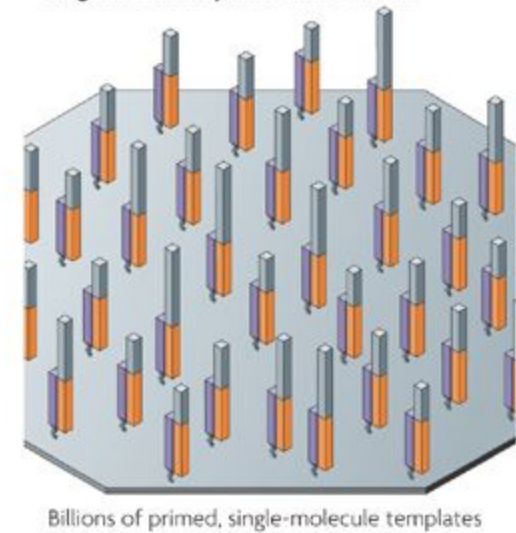


**b Illumina/Solexa**  
**Solid-phase amplification**

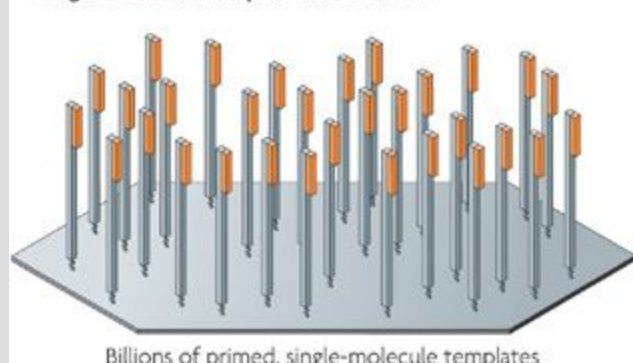
One DNA molecule per cluster



**c Helicos BioSciences: one-pass sequencing**  
**Single molecule: primer immobilized**



**d Helicos BioSciences: two-pass sequencing**  
**Single molecule: template immobilized**



**e Pacific Biosciences, Life/Visigen, LI-COR Biosciences**  
**Single molecule: polymerase immobilized**

