

NGS секвенирование. Основные технологии и платформы.

Доцент кафедры
молекулярной биологии и генетики,
к.м.н. Замарина Т.В.

Технологии секвенирования

- 1-е поколение



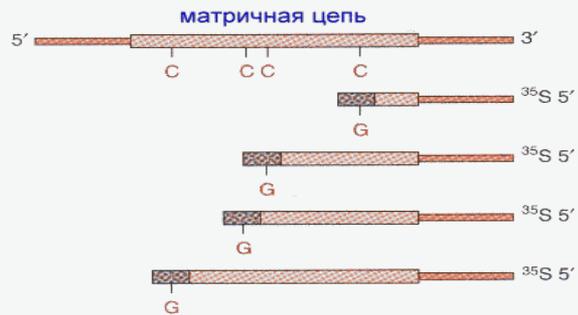
Sanger sequencing

- 2-е поколение

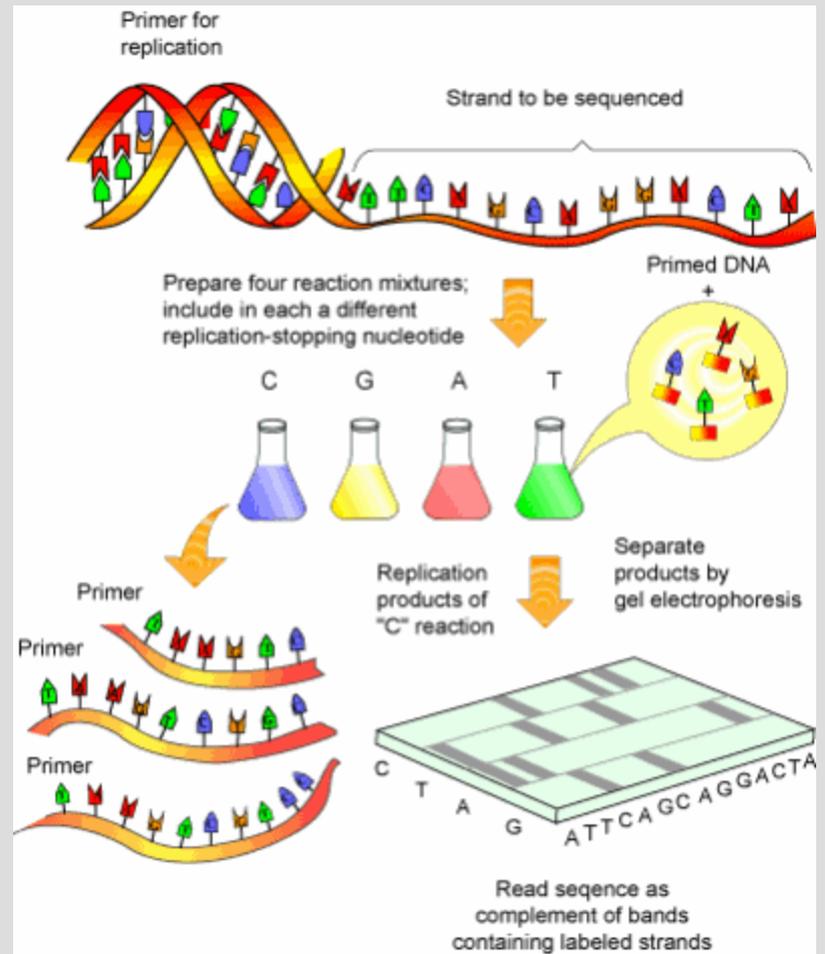
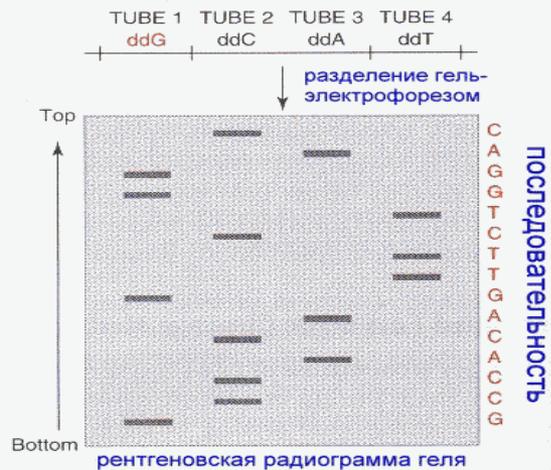


- 3-е поколение

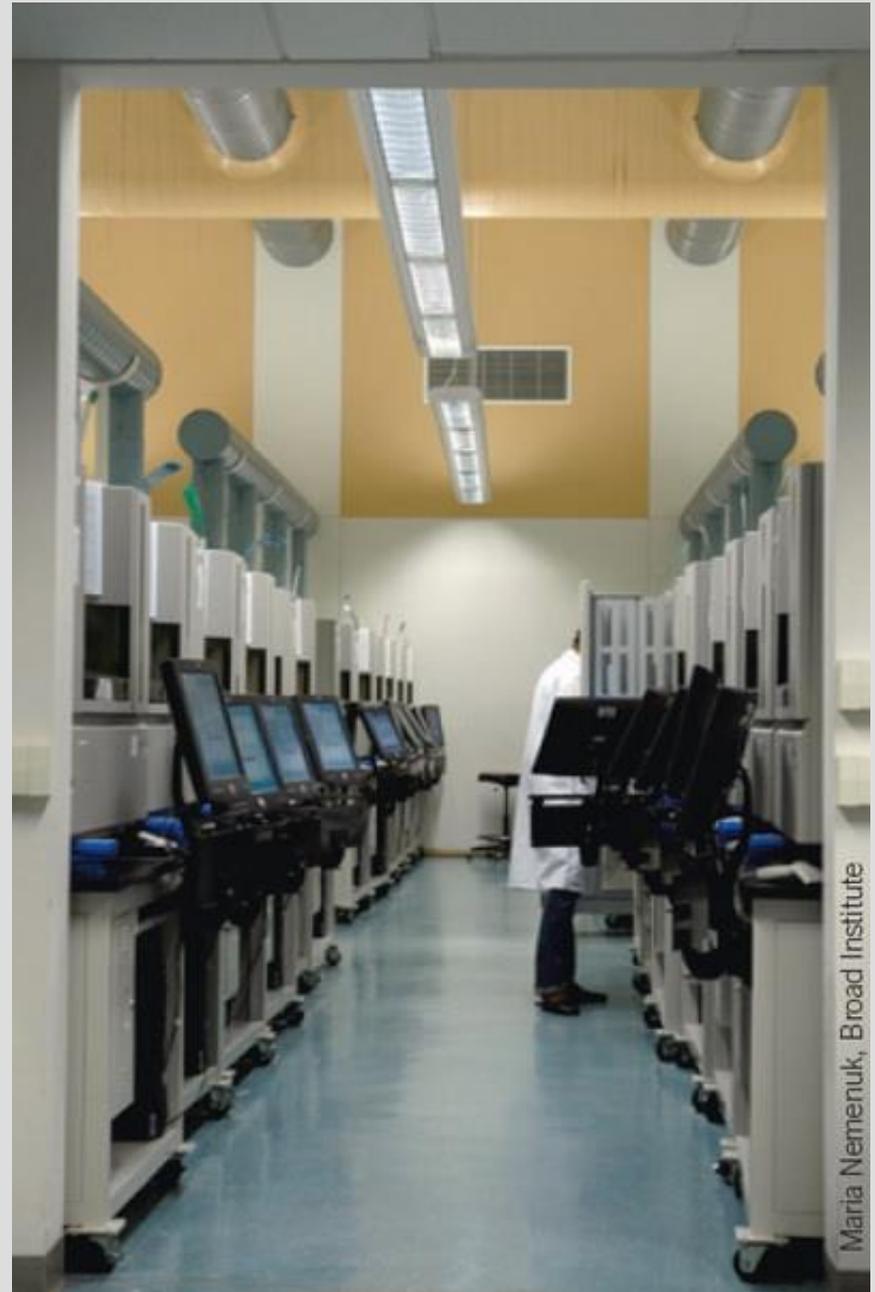


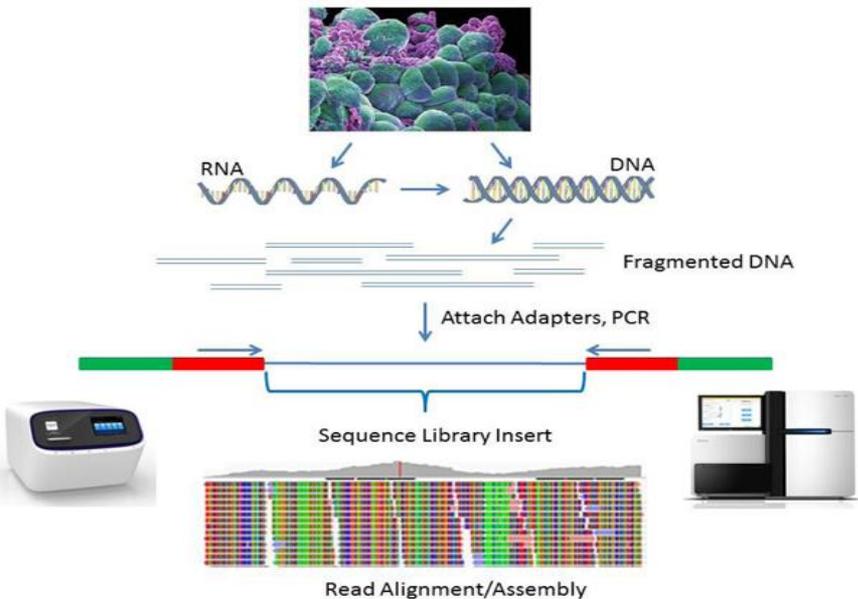


синтезированные фрагменты ДНК
(синтез прерывается в местах вставок
2',3'-дидеоксигуанозинтрифосфата (ddGTP))



Центры, осуществляющие крупные проекты по расшифровке геномов млекопитающих методом Сэнгера, напоминают фабрики своими размерами и количеством обслуживающего персонала. Фото Nature Methods.





Секвенирование нового поколения — техника определения нуклеотидной последовательности ДНК и РНК для получения формального описания её первичной структуры. Технология позволяет «прочитать» одновременно сразу несколько участков генома, что является главным отличием от более ранних методов секвенирования. СНП осуществляется с помощью повторяющихся циклов удлинения цепи, индуцированного полимеразой, или многократного лигирования олигонуклеотидов.

Все основные принципы работы технологий СНП базируются на **секвенировании ДНК-чипов**, используя интерактивные циклические ферментативные реакции с дальнейшим сбором полученной информации в виде иллюстраций.

Полученные данные используются для восстановления нуклеотидной последовательности.

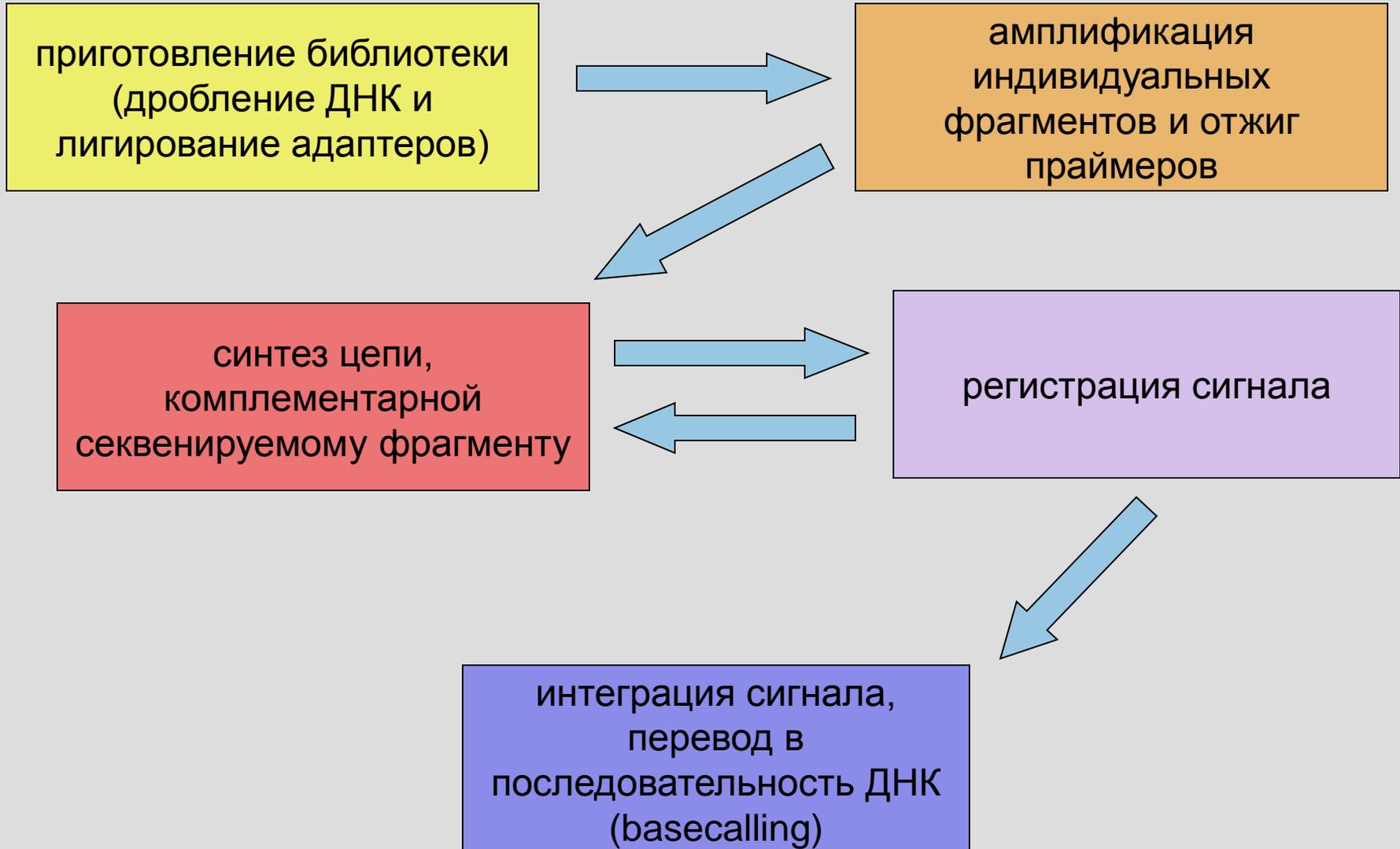
Несмотря на разные методы получения копий (амплификация) участков генома и на техническую разницу дифференциации различных нуклеотидов в прочтённых последовательностях, общая схема работы для всех секвенаторов одна.

Первый этап секвенирования — создание библиотеки случайных последовательностей ДНК, которые можно будет сшить с общедоступными адаптерными последовательностями.

Второй этап — создание ампликонов с помощью ПЦР, которые будут использованы как образцы.

Третий этап — определение первичной структуры всех фрагментов.

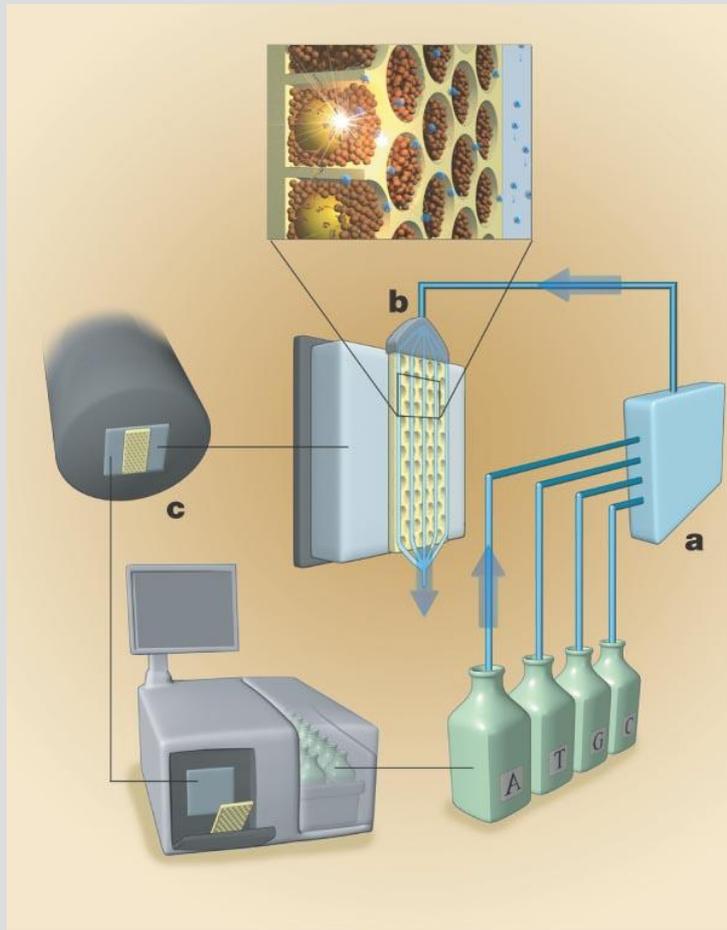
Общая схема работы NGS: от исходной ДНК до «букв» на экране



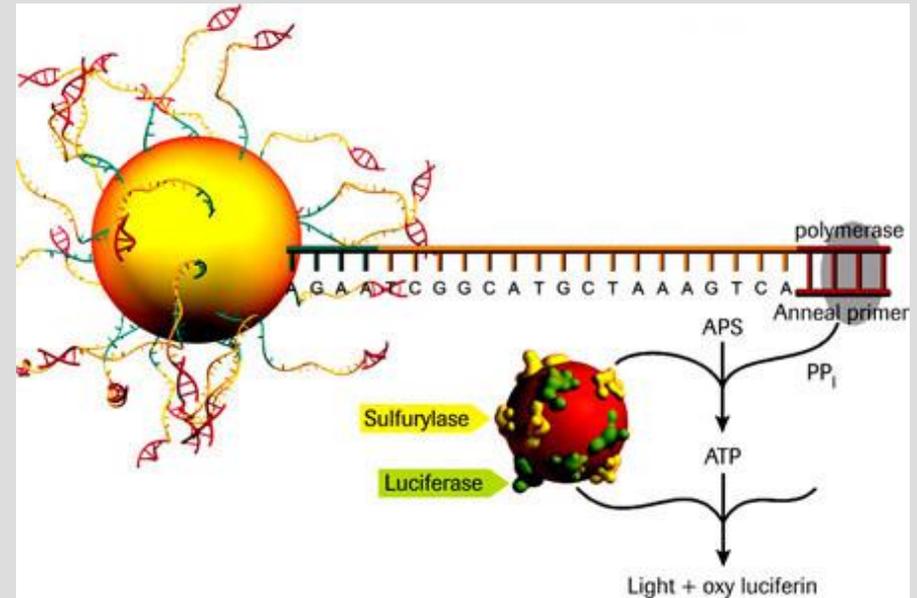
Скорость является одним из главных преимуществ нового метода секвенирования. Не потому ли название метода отсылает нас к легендарному Chevrolet Chevelle SS 454 1970-го года с двигателем мощностью 360 лошадиных сил?



Высокопроизводительное пиросеквенирование (454): 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.

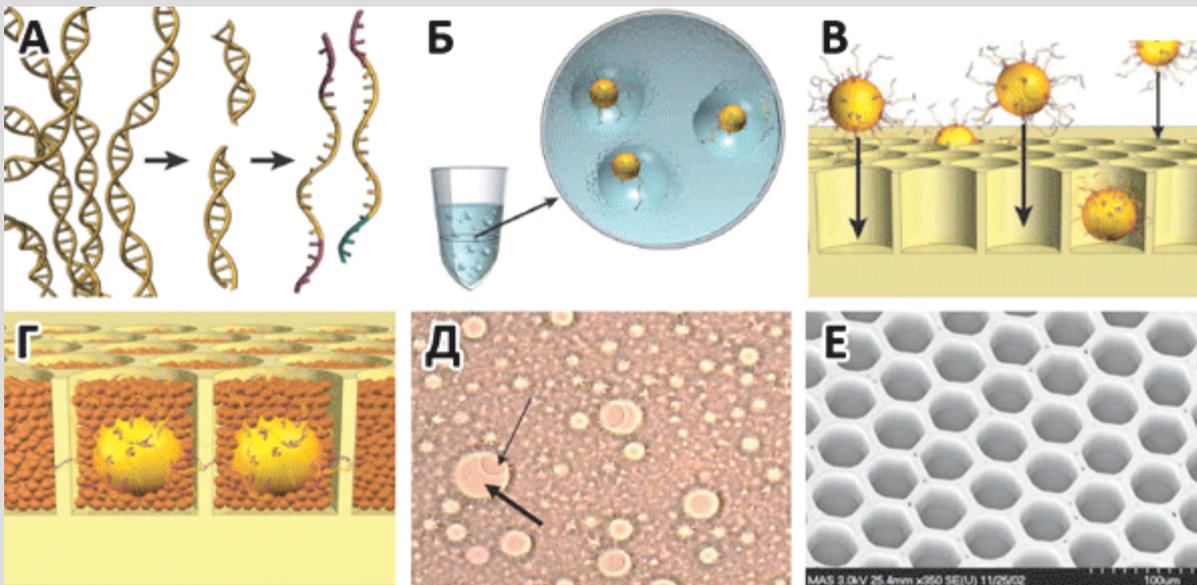


Схема пиросеквенирования

А — ДНК фрагментируется, к фрагментам пришиваются олигонуклеотиды-«адаптеры»; полученные двуцепочечные молекулы ДНК разделяются на две комплементарные цепи. Б — Одноцепочечные молекулы ДНК прикрепляются к бусинкам в условиях, стимулирующих попадание лишь одной молекулы на бусинку. Отдельные бусинки заключаются в 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. Видна плакировка оптических волокон и пустые лунки.

Технологии секвенирования 2 поколения: 454

платформа	GS Junior	GS FLX
длина чтения	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% (индели)	1% (индели)

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



Одного запуска инструмента Roche (454) GS20 было достаточно для анализа 13 млн. пар оснований последовательности генома 28 000-летнего мамонта

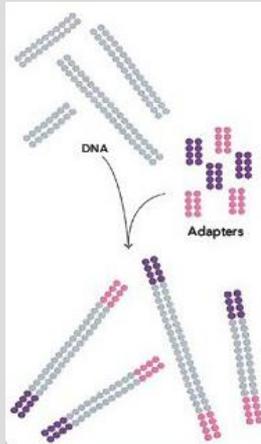


Джонатан Ротберг

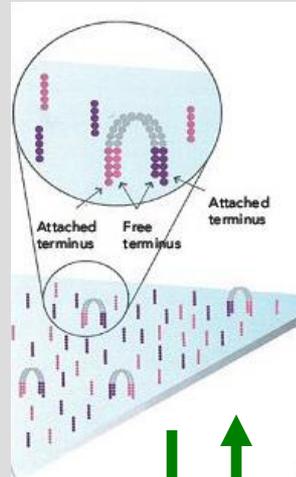


Секвенирование Illumina: принцип метода

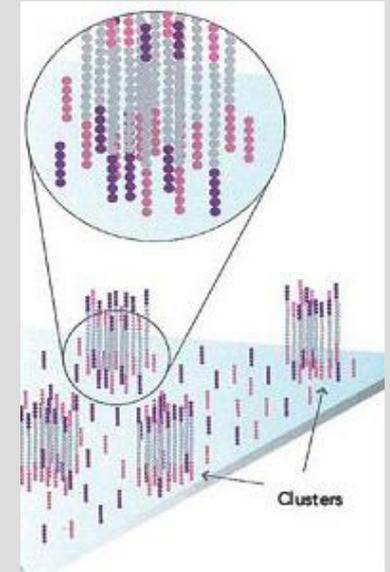
1. ДНК фрагментируют и присоединяют к фрагментам адаптеры



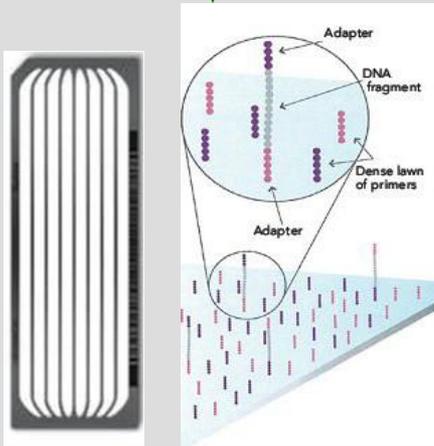
3. Через ячейку пропускают реагенты для достраивания второй цепи ДНК



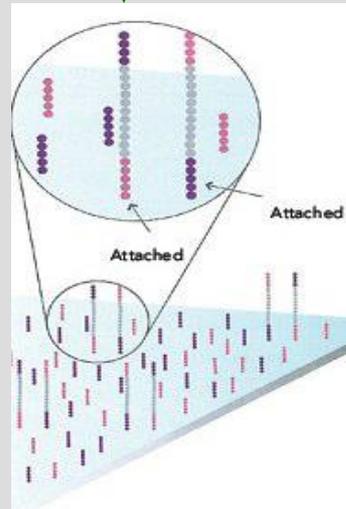
Стадии 3-4 повторяются 30-35 раз



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



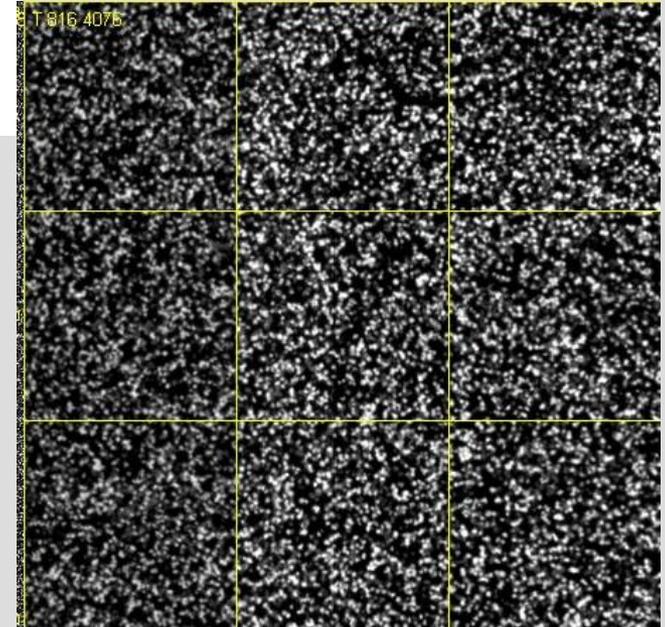
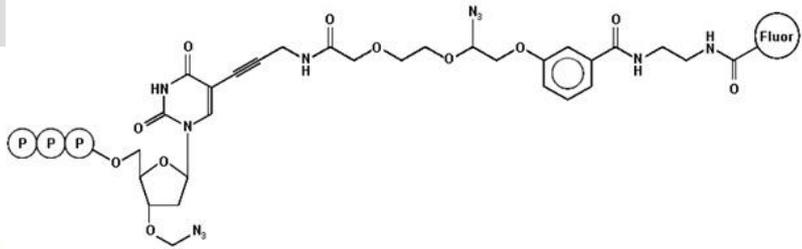
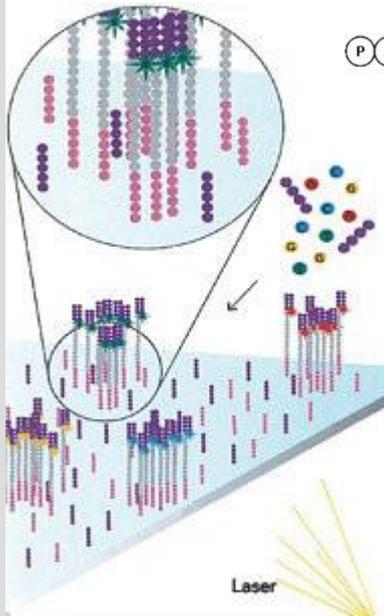
4. Двухцепочечные фрагменты денатурируют



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



Секвенирование Illumina - принцип метода



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

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

8. На ячейку светят лазером
и проводят съемку.

9. Через ячейку пропускают
реагенты, отщепляющие
флуорофор и терминатор

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

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 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

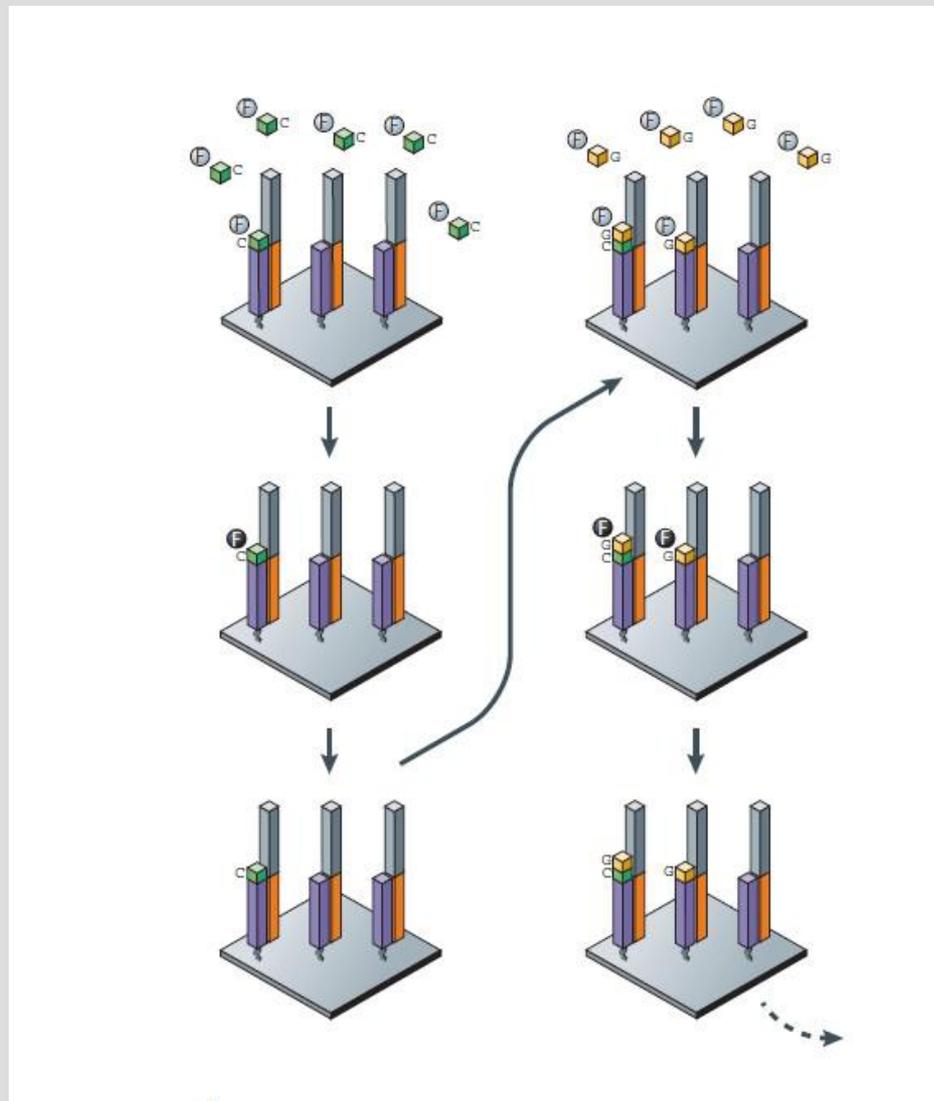
- очень короткие чтения
- длительность работы
- относительно малая доступность свободного ПО

Секвенирование единичных молекул

Helicos

- 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%)

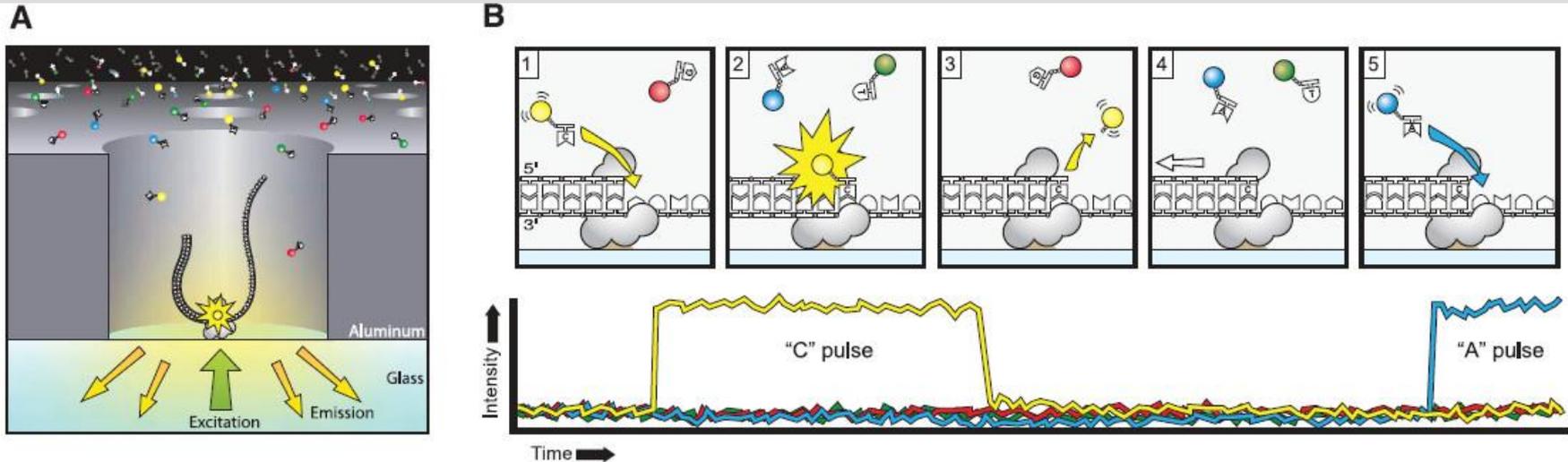
Область применения – высокоточный анализ экспрессии, оценка копийности участков генома



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.

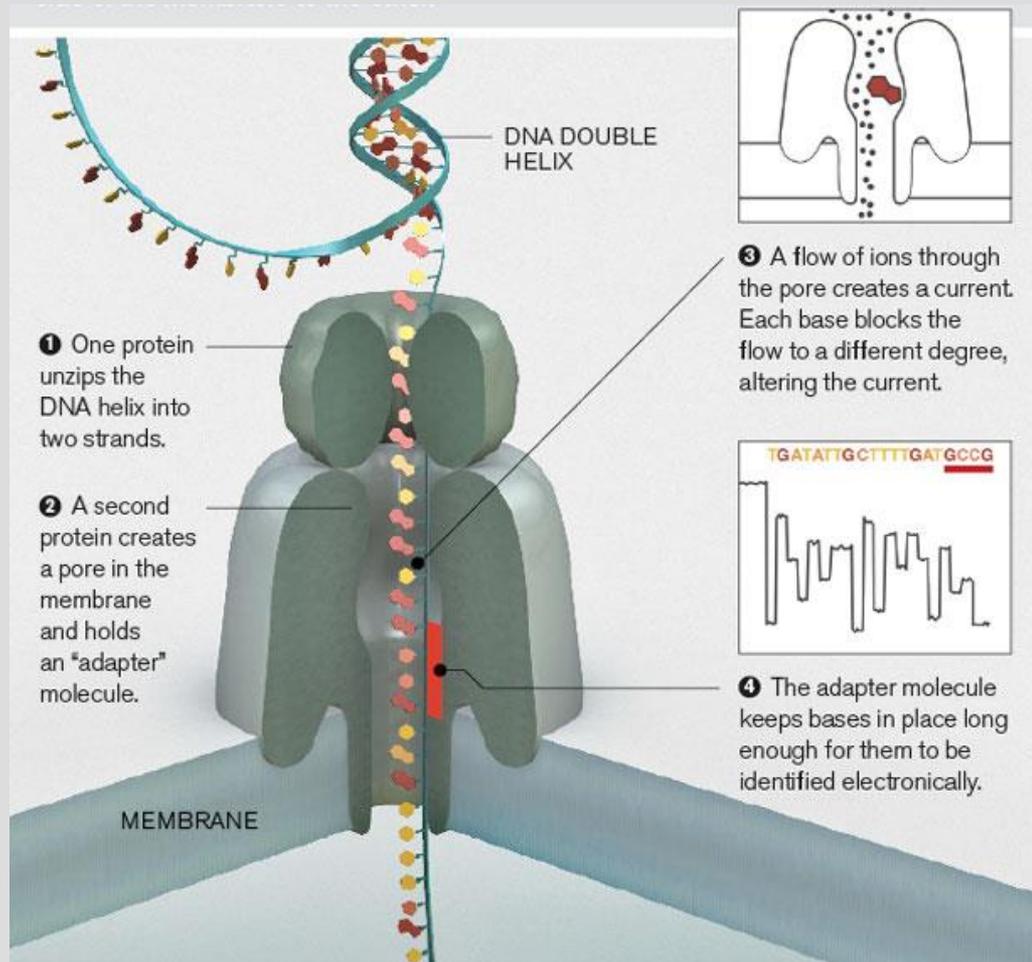


Области применения – де ново сборка (в сочетании с Illumina для коррекции ошибок), анализ изоформ, поиск модифицированных оснований

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!)



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

Сравнение различных методов СНП

метод	принцип	длина одного прочтения, пар оснований	стоимость секвенирования 1 млн пар оснований	стоимость секвенатора	время работы за цикл	количество прочтений за цикл	преимущества	недостатки
454 Life Sciences	пиросеквенирование	400	10\$	500 000\$	7 часов	1 000 000	длина прочтённых геномных участков; скорость	стоимость; погрешность
Illumina-SOLEXA	SBS (sequencing-by-synthesis)	300	0,05—0,15\$	600 000\$	9 дней	до 3 000 000 000	эффективность, стоимость	скорость
IonTorrent	ионный полупроводник	200	1\$	50 000\$	1,5 часа	до 5 000 000	стоимость; скорость	погрешность
SOLID	секвенирование на основе лигирования	35—50	0,13\$	595 000\$	9 дней	1 300 000 000	стоимость	скорость
Helicos	HeliScope	2900	2\$		1 час	35 000—75 000	длина прочтённых геномных участков; скорость	низкая производительность при желаемой малой погрешности; стоимость