Next-generation sequencing
Высокопроизводительное секвенирование

Лаборатория для исследования древних ДНК открыта в СО РАН

«Агентство Химэксперт»

ХИМЭКСПЕРТ

Next-generation sequencing
Высокопроизводительное секвенирование

Лаборатория для исследования древних ДНК открыта в СО РАН

«Агентство Химэксперт»

ХИМЭКСПЕРТ

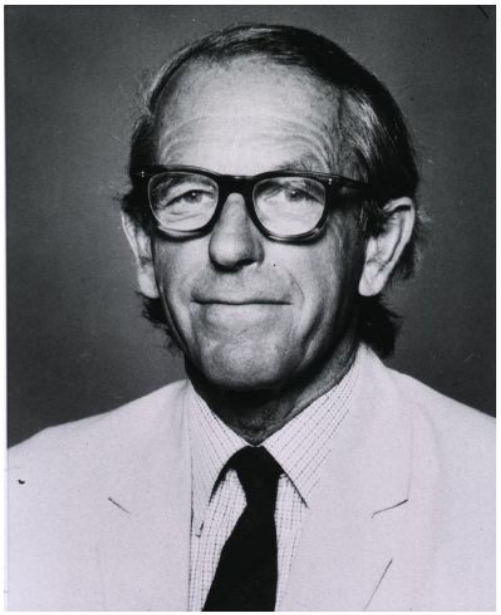
Введение в высокопроизводительное секвенирование

- Эволюция секвенирования: от Сэнгера к NGS
- Ключевые понятия NGS
- Этапы подготовки и проведения высокопроизводительного секвенирования

«Агентство Химэксперт»

ХИМЭКСПЕРТ

Поколения секвенирования



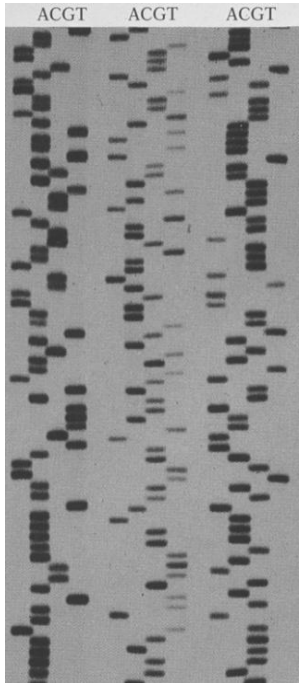
1918 - 2013

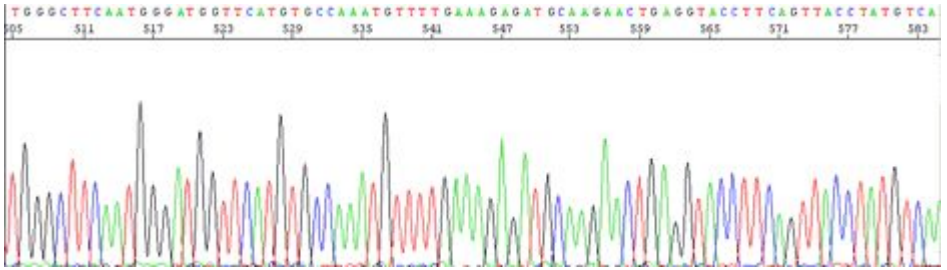
Фредерик Сэнгер

ХИМЭКСПЕРТ

Поколения секвенирования

I поколение
метод Сэнгера





ХИМЭКСПЕРТ

Поколения секвенирования

II поколение

Next-generation sequencing

Позволяет определять нуклеотидную последовательность принципиально большей общей протяженности за один рабочий цикл по сравнению с методами секвенирования предыдущего поколения (методы Сэнгера и Максама-Гилберта)

ХИМЭКСПЕРТ

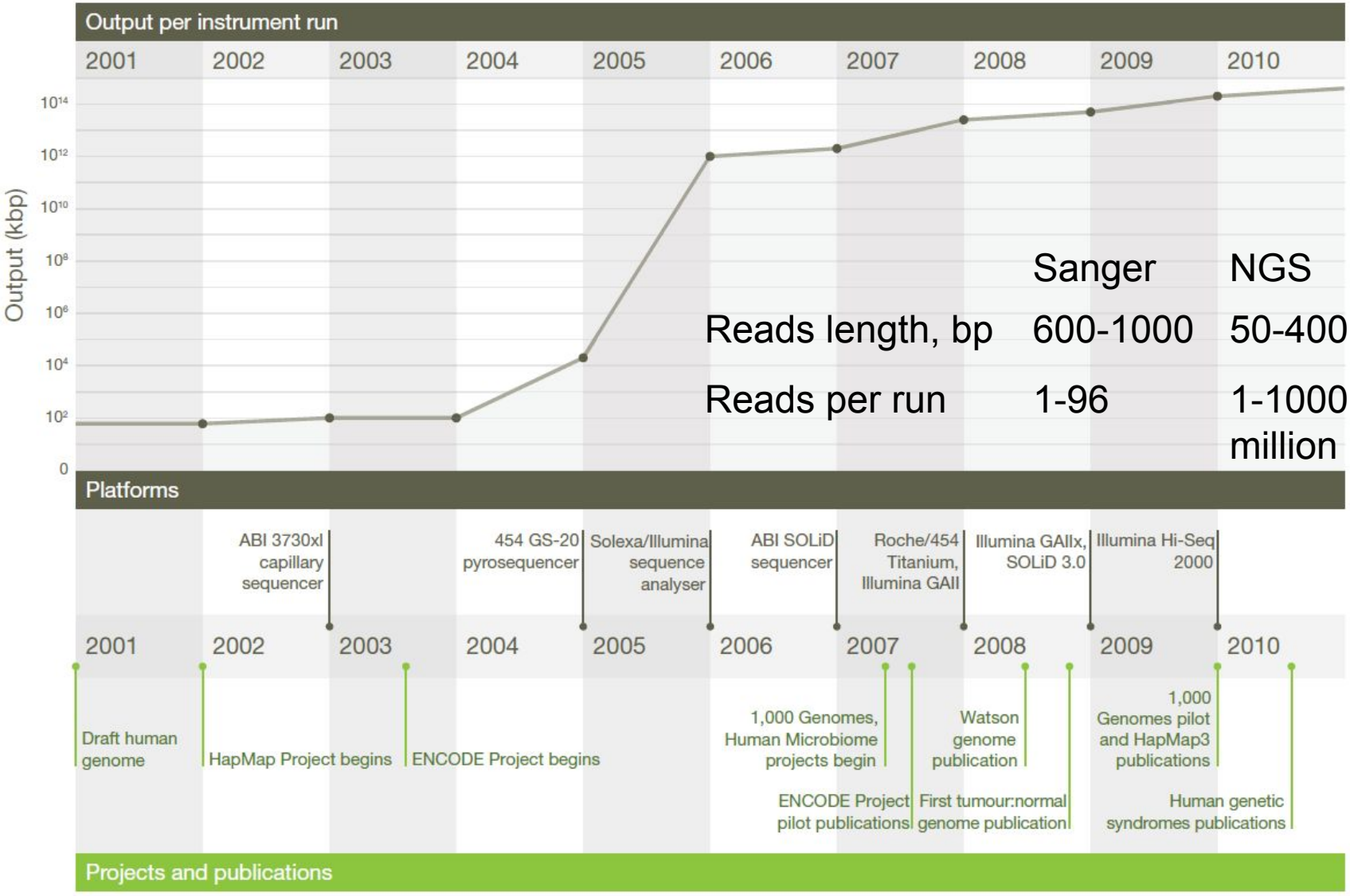
Next-generation sequencing

секвенирование нового поколения
секвенирование следующего поколения
высокопроизводительное секвенирование

- определение нуклеотидной последовательности ДНК, основанная на параллельном «чтении» перекрывающихся фрагментов исходной молекулы.

ХИМЭКСПЕРТ

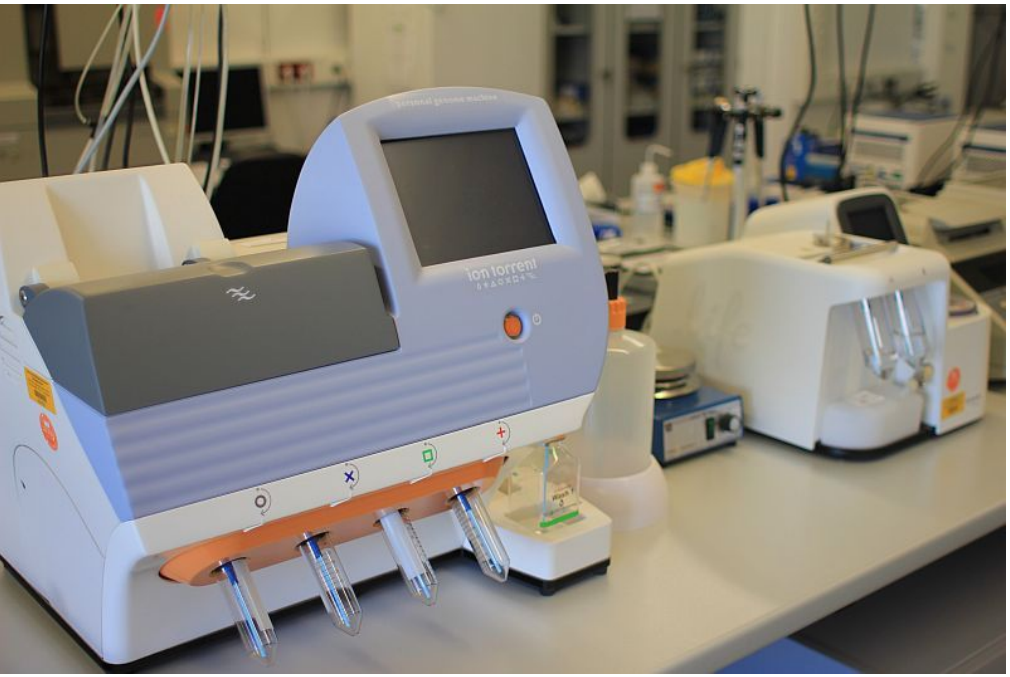
Эволюция секвенирования



A decade's perspective on DNA sequencing technology. ER Mardis - Nature, 2011

ХИМЭКСПЕРТ

Технология NGS позволяет проводить:



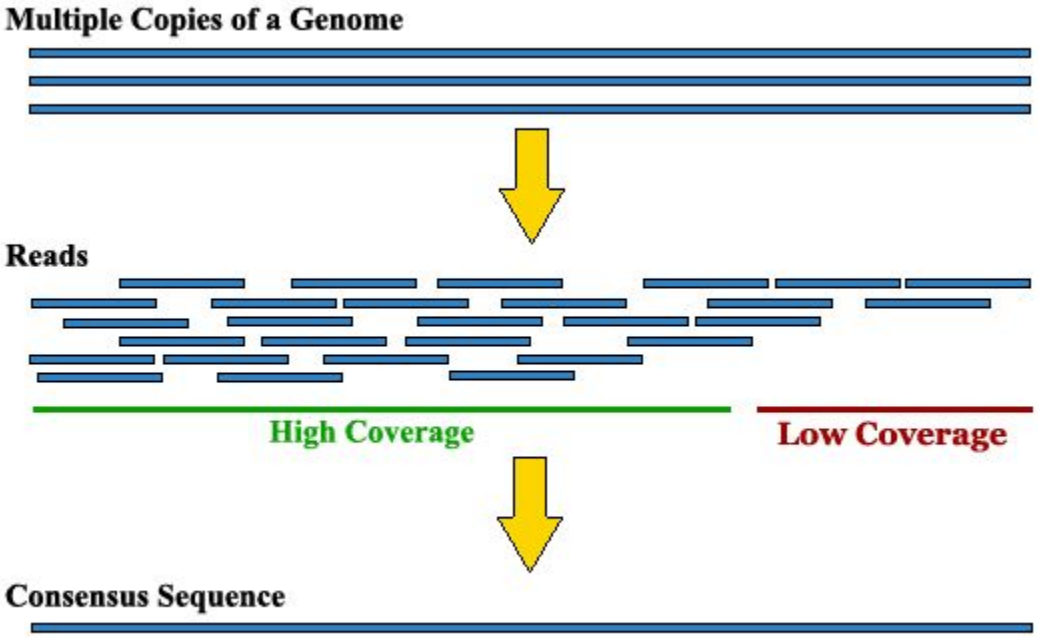
- Секвенирование отдельных генов или наборов генов
- Массовое секвенирование ампликонов
- Секвенирование транскриптомов
- Секвенирование экзомов
- Полногеномное секвенирование

ХИМЭКСПЕРТ

NGS. Общий принцип

- ДНК нарезается на фрагменты определенной длины
- К ним лигируются адаптеры
- Амплификация каждого отдельного фрагмента в изолированных от других условиях
- Анализ последовательности амплифицированных клонов ДНК

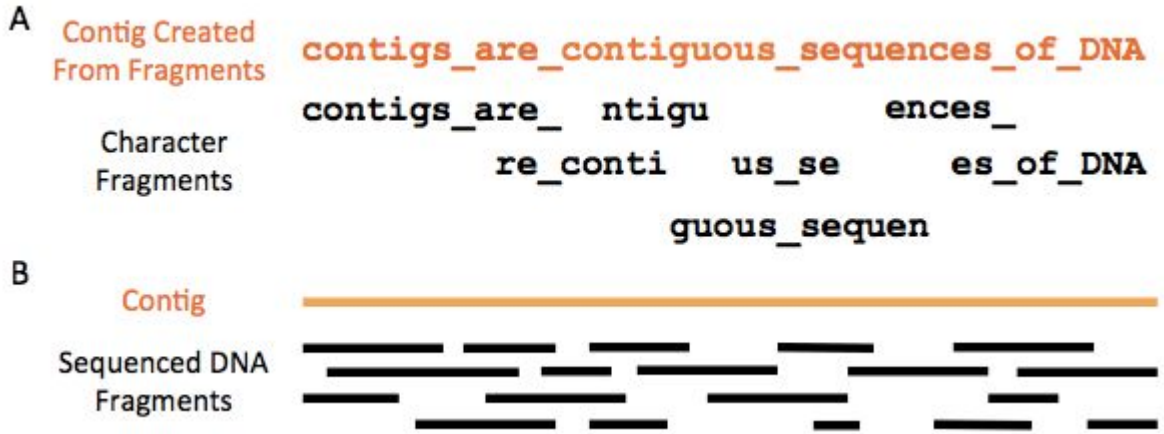
NGS



http://gcat.davidson.edu/phast/

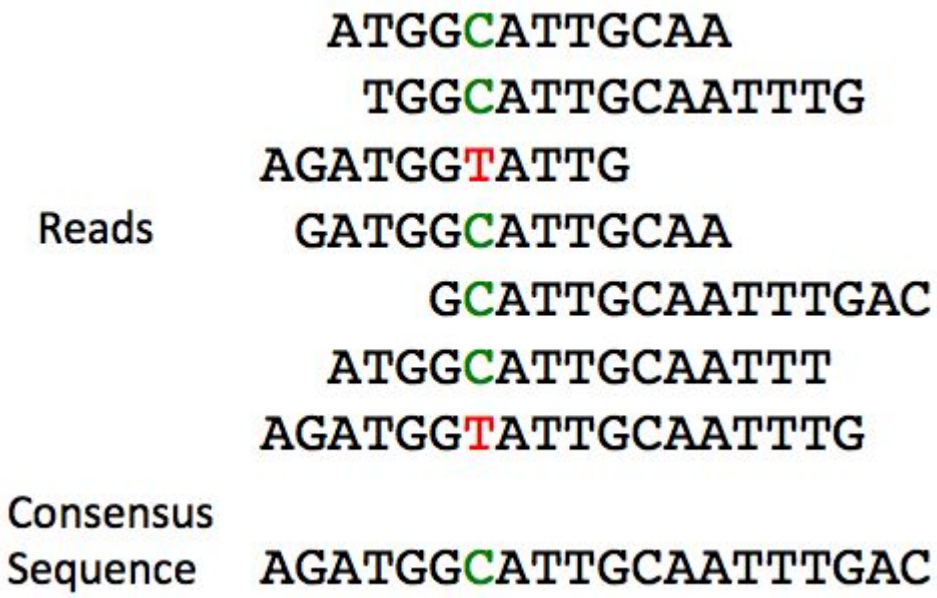
ХИМЭКСПЕРТ

NGS



http://gcat.davidson.edu/phast/

NGS

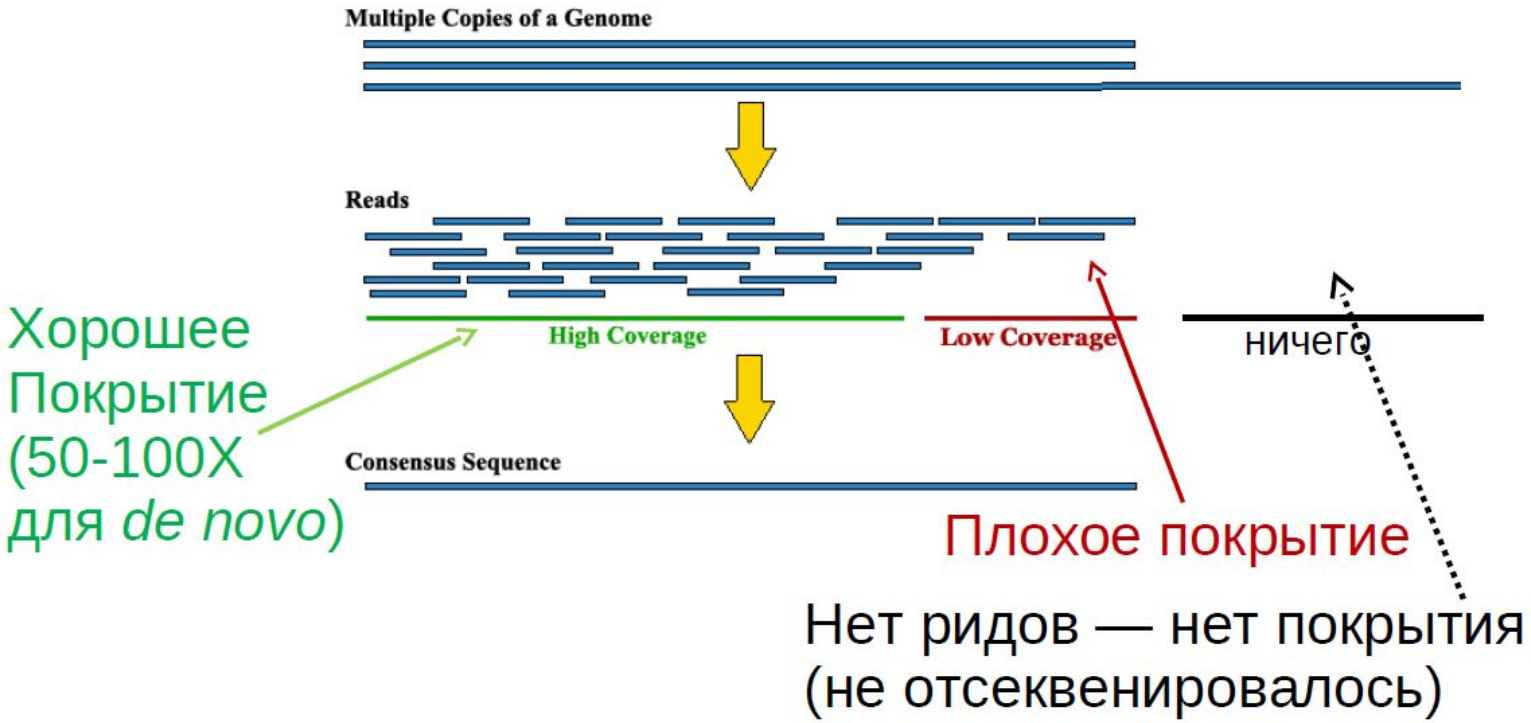


http://gcat.davidson.edu/phast/

ХИМЭКСПЕРТ

Что такое покрытие (coverage)

- Это сколько раз в среднем нуклеотид генома покрыт ридами



Как собрать геном *de novo* из коротких чтений? Практические советы. Науменко С.А.
http://gcat.davidson.edu/phast/

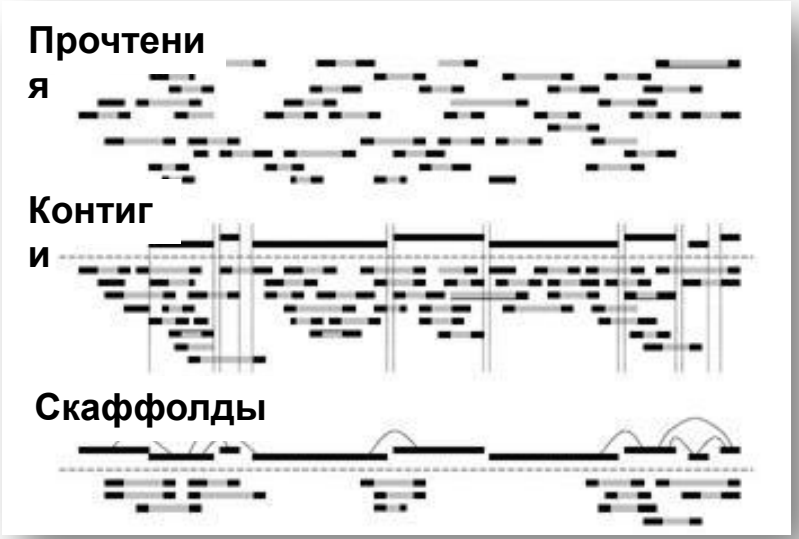
ХИМЭКСПЕРТ

de novo сборка генома

- **Сборка последовательности** - выравнивание и слияние фрагментов более длинной последовательности ДНК для восстановления исходной последовательности

- *de novo* сборка выполняется в случае, если отсутствует референсная последовательность для сравнения

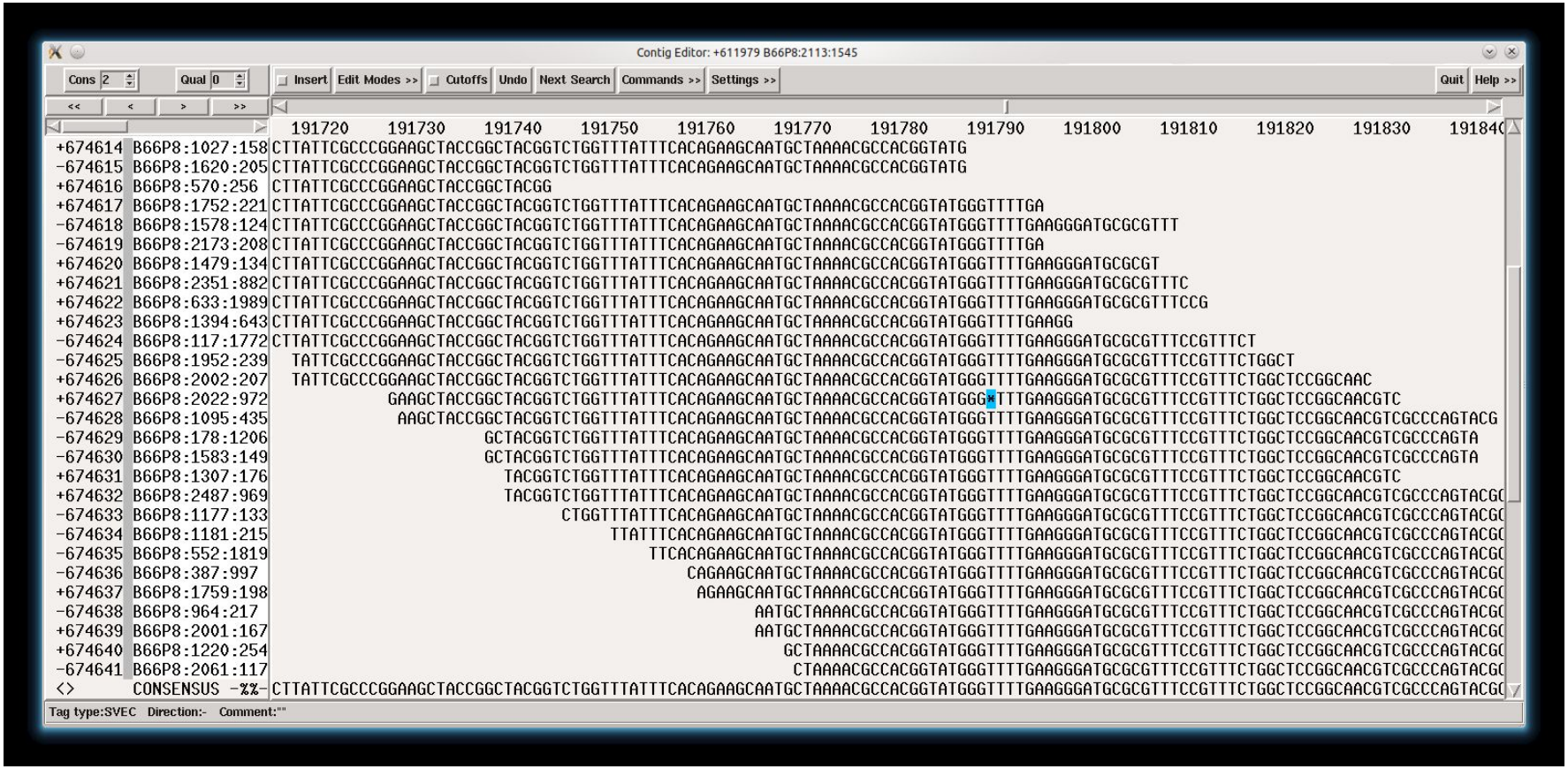
- При гибридной сборке можно использовать элементы близкородственной или частично референсной последовательности



Circos plot of a bacterial genome assembly. Comparison between a german *E. coli* outbreak strain and reference strain *E. Coli 559889*

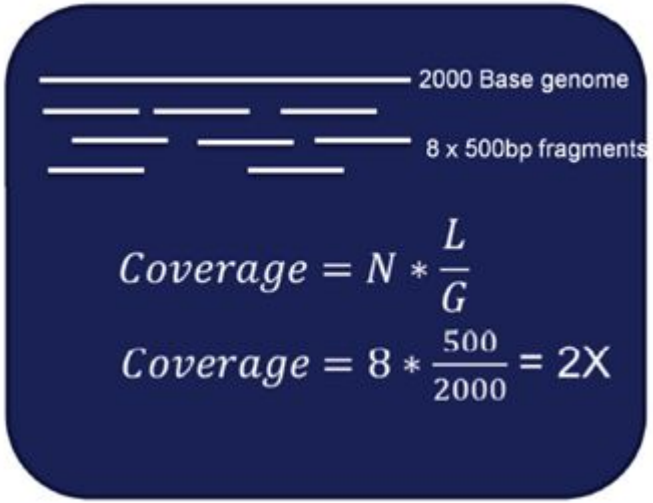
ХИМЭКСПЕРТ

Пример контига, собранного по данным Torrent Data в программе MIRA



Что такое покрытие?

- Среднее количество прочтений, содержащих данный нуклеотид в реконструированной последовательности
- Позволяет оценить % покрытия генома
- Высокое покрытие исправляет ошибки при сборке



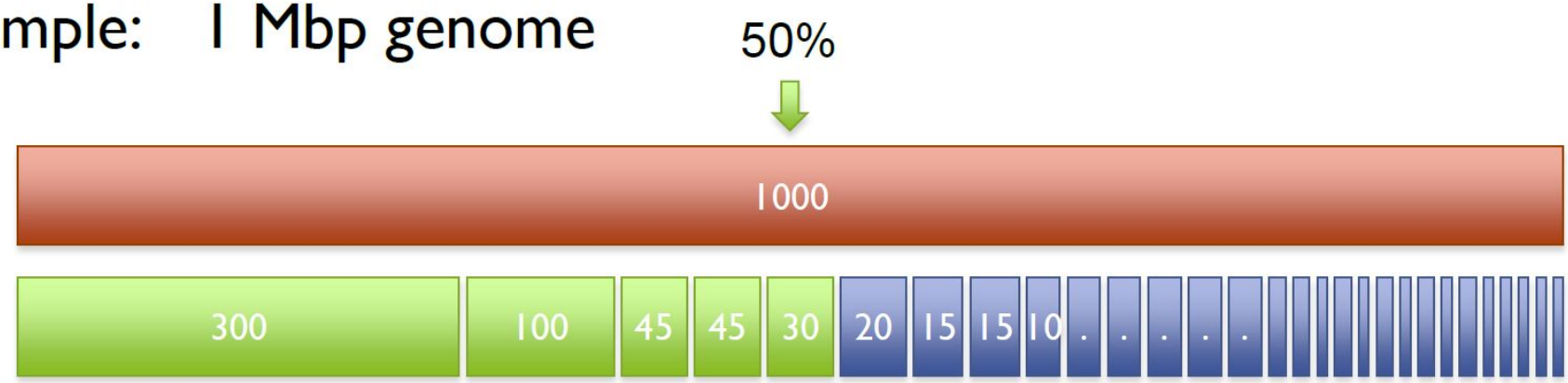
N = Кол-во прочтений
L = средняя длина прочтений
G = Длина прочитываемого участка (генома)

Типичное желательное покрытие генома 30x

ХИМЭКСПЕРТ

Что такое N50

Example: 1 Mbp genome

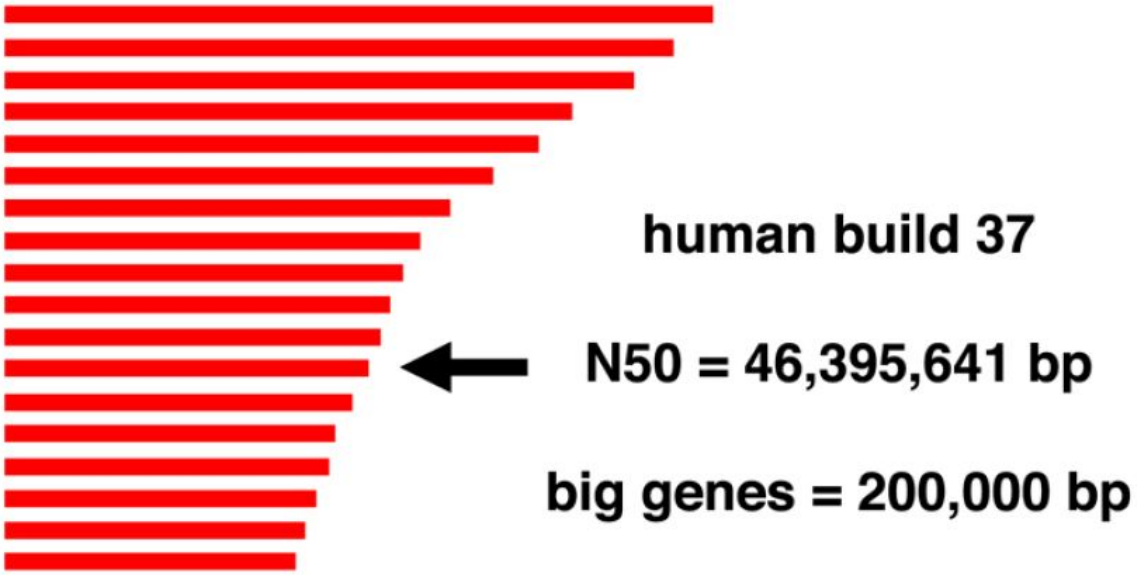


N50 size = 30 kbp

(300k+100k+45k+45k+30k = 520k >= 500kbp)

Large-scale Sequencing and Assembly of Cereal Genomes Using Blacklight
Philip Blood

Что такое N50



- N50 показывает качество сборки
- Скаффолды располагают по убыванию длины
- Суммируют длину, начиная с самого большого скаффолда

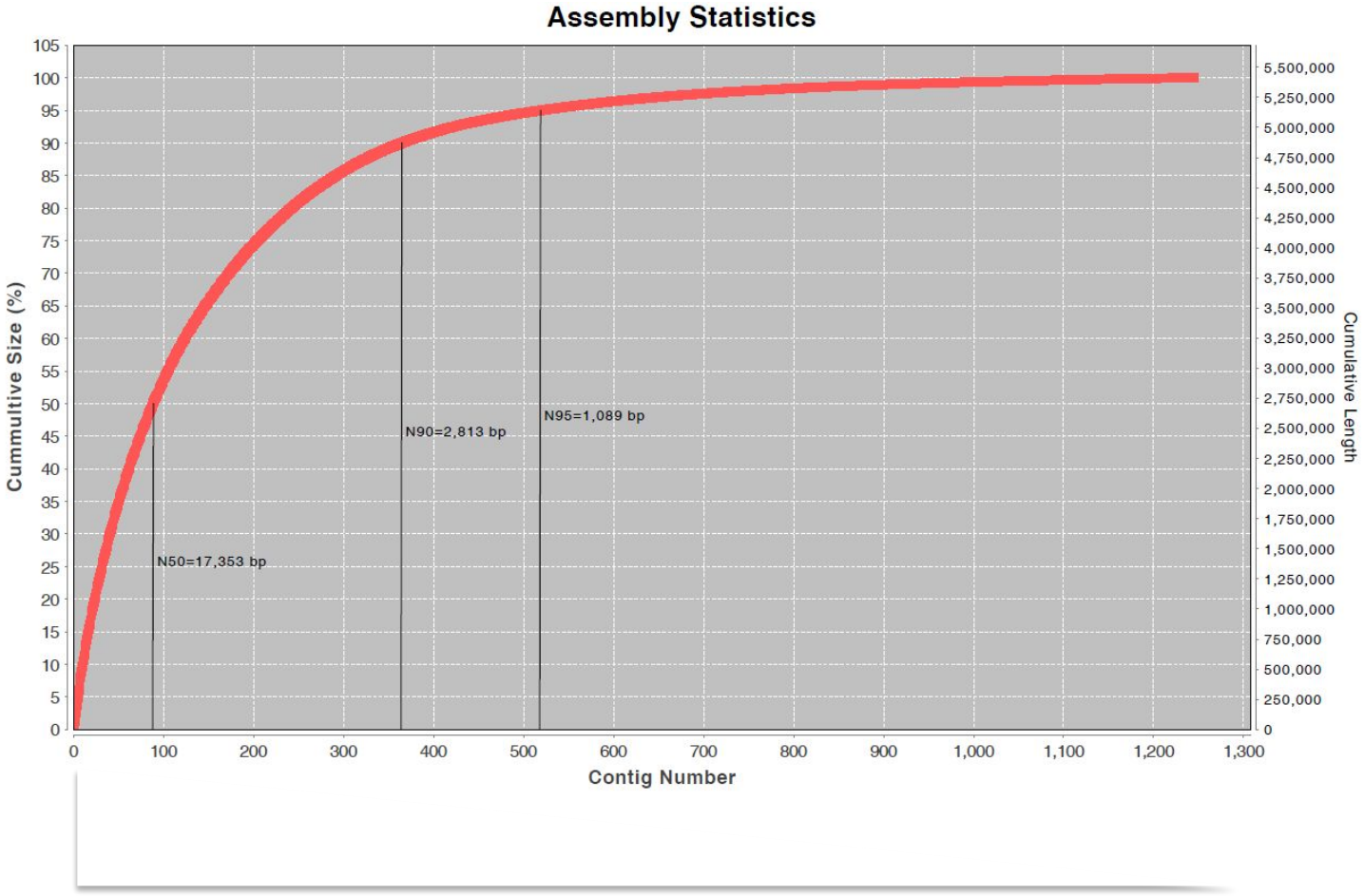
❑ На каком скаффолде покроем половину генома?

✔ Длина этого скаффолда называется N50.

Как собрать геном *de novo* из коротких чтений? Практические советы. Науменко С.А.
http://www.discoveryandinnovation.com/BIOL202/notes/lecture25.html

ХИМЭКСПЕРТ

N50 – это длина контига, при которой контиги такой же или большей длины составляют 50% всех нуклеотидов в сборке



Контиги, отсортированные по длине

ХИМЭКСПЕРТ

Ключевые параметры *de novo* сборки

Параметр	Описание	Пример
N50	«Чем больше, тем лучше» N50 это длина контига, при которой контиги такой же или большей длины составляют 50% всех нуклеотидов в сборке.	N50 = 98.6Kb *E. coli* 400bp PGM
Число контигов **Контиг**: набор перекрывающихся фрагментов ДНК, которые в совокупности представляют собой консенсус области ДНК	«Чем меньше, тем лучше» По мере увеличения числа контигов, которые укладываются в сборку, это число уменьшается. В идеале полностью собранный геном без пробелов будет состоять из 1 контига.	# Contigs = 158 *E. coli* 400bp PGM
% покрытия референса	«Чем ближе к 100%, тем лучше» % покрытия референса означает % всех оснований в геноме, которые были покрыты по крайней мере одним прочтением.	% Ref Coverage = 98.15% *E. coli* 400bp PGM

ХИМЭКСПЕРТ

Рабочий процесс NGS

ХИМЭКСПЕРТ

Основные этапы секвенирования

Приготовление библиотеки

Подготовка матрицы

Секвенирование

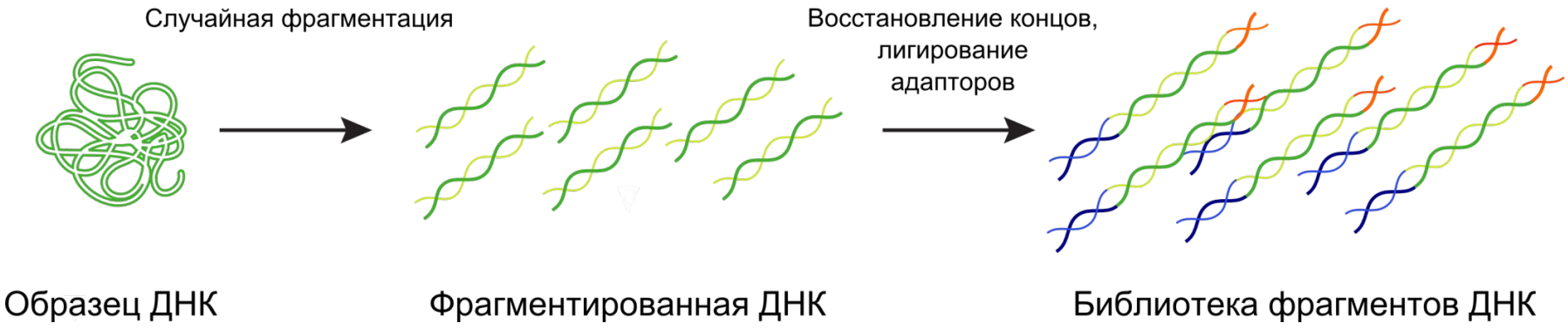
Анализ данных

ХИМЭКСПЕРТ

Приготовление библиотеки

Источники ДНК для приготовления библиотек:

- геномная ДНК (Whole genome);
- часть геномной ДНК (Targeted Enrichment);
- набор ПЦР-продуктов (ампликонов);
- кДНК (RNA-Seq);
- иммунопреципитированный хроматин (ChIP-Seq) и т. д.

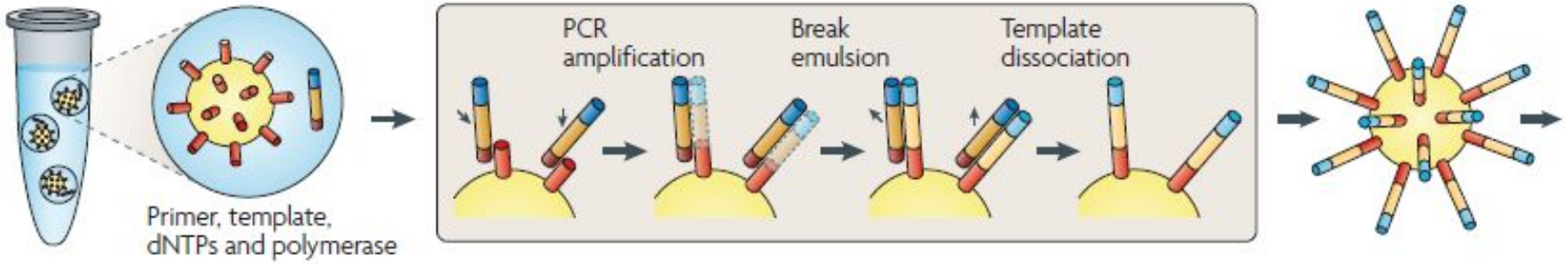


ХИМЭКСПЕРТ

Подготовка матрицы

Эмульсионная ПЦР и автоматизированная система обогащения на магнитных частицах

Система **Ion OneTouch 2**

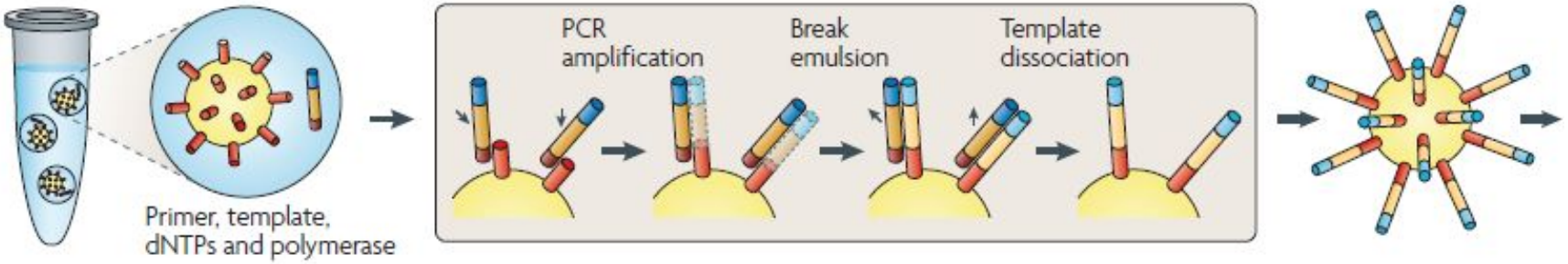


ХИМЭКСПЕРТ

Подготовка матрицы

- Амплификация библиотеки
- Восстановление и обогащение сфер
- Загрузка чипа

Станция **Ion Chef**



ХИМЭКСПЕРТ

Секвенирование

– регистрация локального изменения pH на полупроводниковом микрочипе при последовательном удлинении олигонуклеотидной затравки ДНК-полимеразой

Ion PGM™ Sequencer

Ion Proton™ Sequencer

ХИМЭКСПЕРТ

Секвенирование

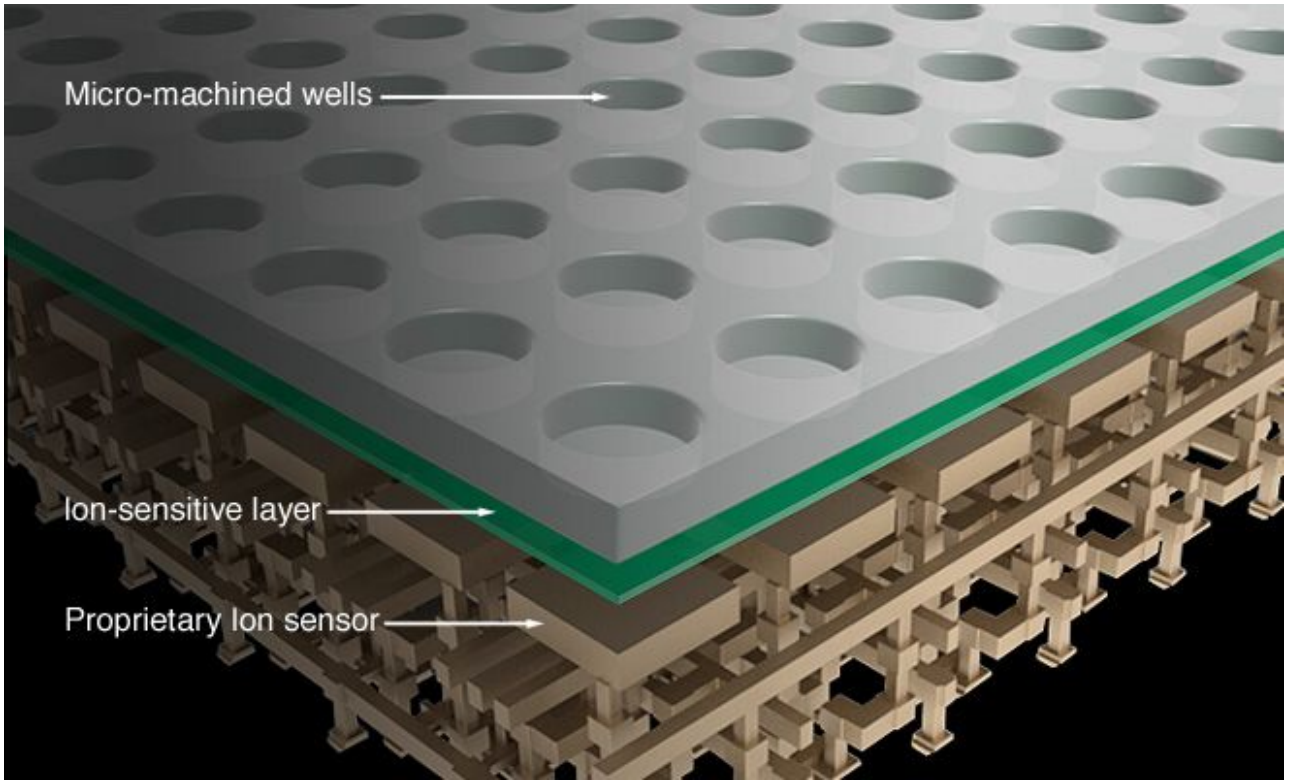
Ion 314™

Ion 316™

Ion 318™

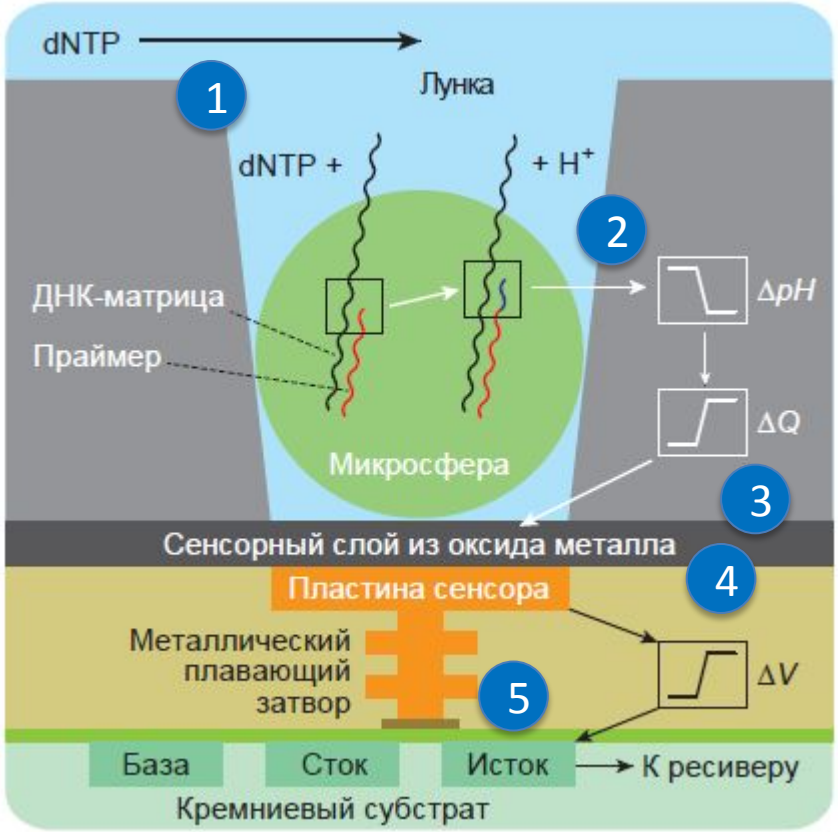
PI ™

Ion чип



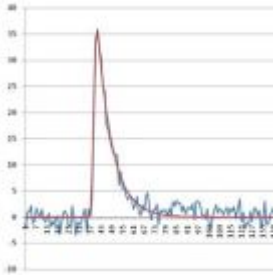
ХИМЭКСПЕРТ

Детекция на чипе в реальном времени



Схематическое изображение поперечного сечения одной ячейки на чипе

1. дНТФ последовательно подаются один за другим на чип
2. После встраивания нуклеотида в цепь происходит отщепление протона и изменение pH в ячейке
3. Чувствительный слой регистрирует изменение pH
4. Сенсор переводит химический сигнал в цифровой
5. Изменение напряжения пропорционально количеству встроенных нуклеотидов

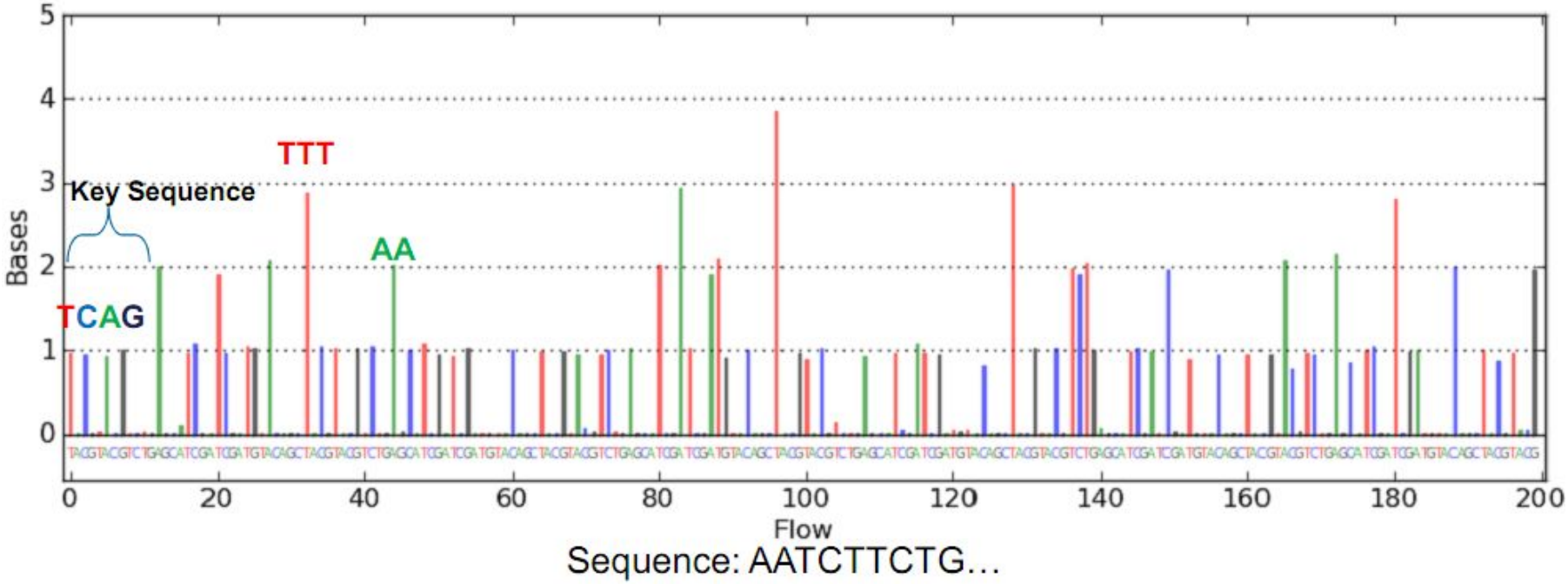


Измерение напряжения

https://www.invitrogen.com/site/us/en/home/Products-and-Services/Applications/Sequencing/Semiconductor-Sequencing/Semiconductor-Sequencing-Technology/Ion-Torrent-Technology-How-Does-It-Work.html

ХИМЭКСПЕРТ

Иониограмма



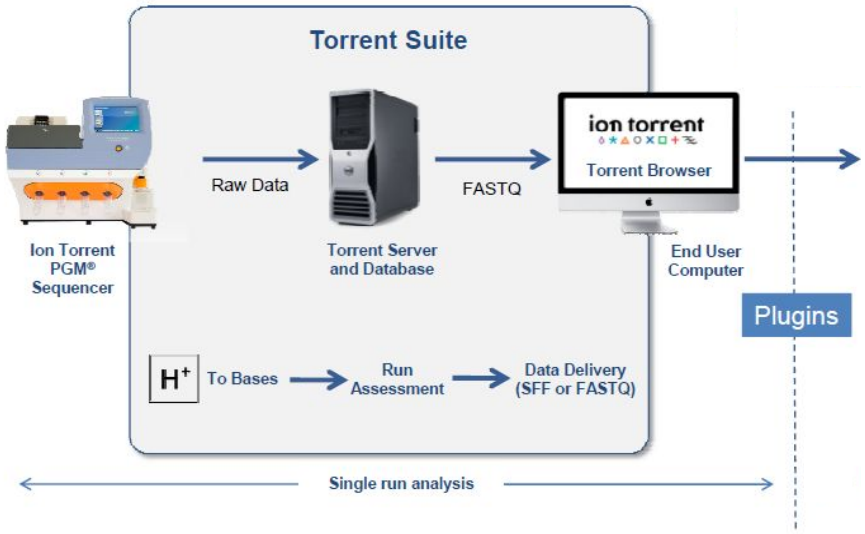
- Читается слева – направо и по высоте
- Высота говорит о том, сколько нуклеотидов инкорпорирует во время одной подачи нуклеотида
- Отсутствие пика при подачи нуклеотида говорит об отсутствии данного нуклеотида в матрице

ХИМЭКСПЕРТ

Анализ данных

Программное обеспечение

- **Torrent Suite**
  - Программа для запуска и настройки хода секвенирования.
  - Простые опции для показа метрики анализа, доступные для понимания тех, кто не является экспертом в области биоинформатики.
  - Точность при определении гомополимеров.
  - Различные приложения.
  - Отправка данных через «облако» по сети.



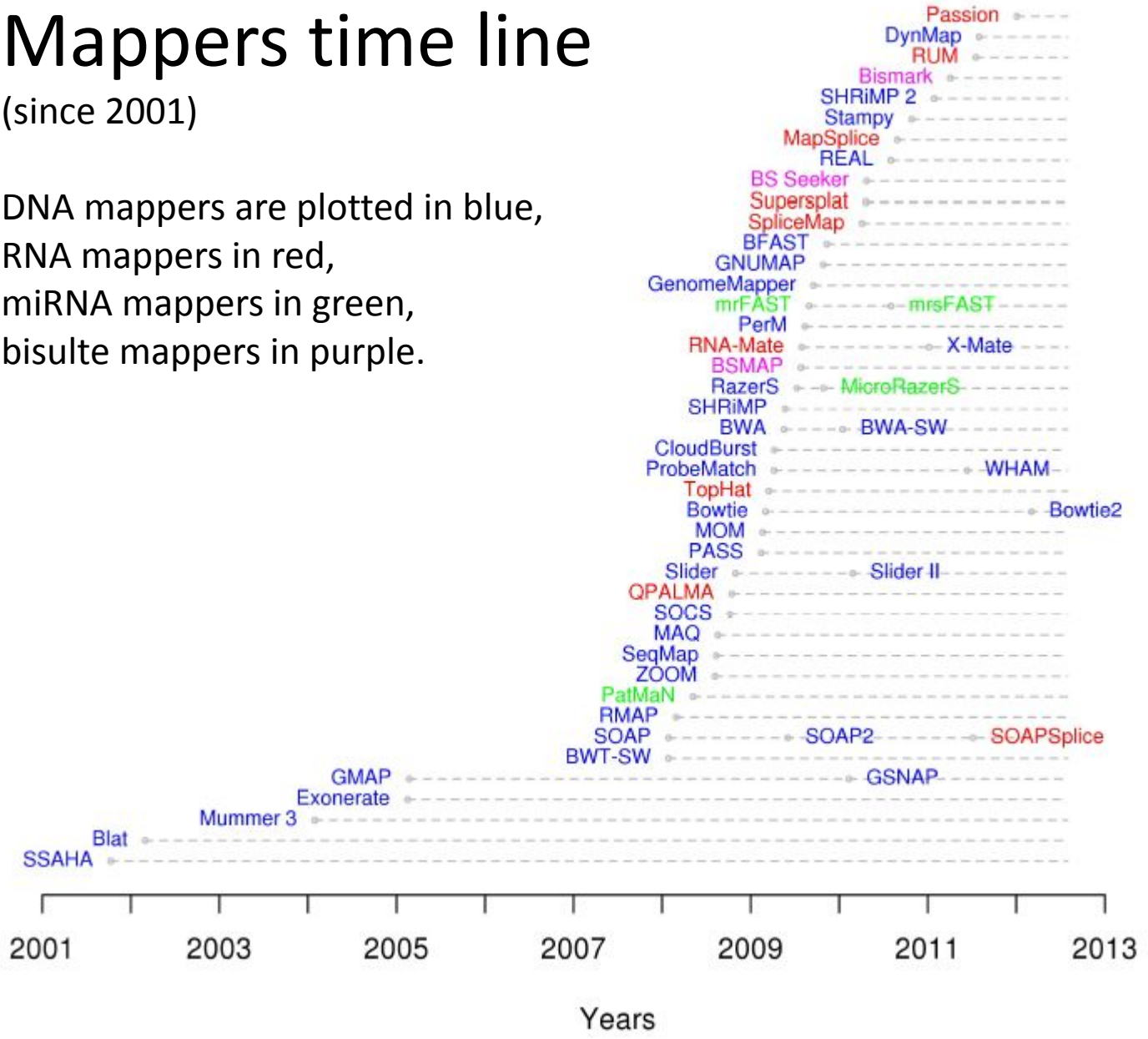
- **Ion Reporter**
  - Анализ данных через «облако», аннотация и составление отчёта о результате.

ХИМЭКСПЕРТ

Mappers time line

(since 2001)

DNA mappers are plotted in blue,
RNA mappers in red,
miRNA mappers in green,
bisulte mappers in purple.



Tools for mapping high-throughput sequencing data
Nuno A. Fonseca, Johan Rung, Alvis Brazma, John C. Marioni. 2012

Масштабное применение NGS

- 1000 Genomes Project: поиск геномных вариаций в ~20 популяциях
- PGP: personal genome project, изучение вклада наследственности и среды в проявление различных признаков; секвенирование геномов добровольцев, согласных предоставить личную информацию
- GEUVADIS Genetic European Variation in Disease: медицинскя интерпретация сиквенсовых данных для разных моно- и полигенных заболеваний; на основе RNA-Seq и ExonSeq
- ESGI: European Sequencing and Genotyping Infrastructure: организация NGS инфраструктуры для европейского научного сообщества
- IHEC: International Human Epigenome Consortium: характеристика эпигенома человека, ~1000 образцов

- **Раковые программы**
  выявление диагностичеких маркеров, анализ и выбор лечения
- ICGC: International Cancer Genome Consortium: подпрограммы по разным видам рака (напрмер, рак простаты, детские опухоли мозга)
- Treat 1000: разные виды рака
- OncoTrack: рак прямой кишки
- CAGEKID: рак почки
- Treat 20: меланома

Геном

Сиквенс и ресиквенс

- микроорганизмы / малые геномы: самый эффективный подход
- большие геномы: ресиквенс с анализом небольших перестроек был хорош всегда, сиквенс de novo и анализ больших перестроек совершенствуются на глазах
- классификация организмов по молекулярным признакам
- сиквенс определенных областей генома: обогащение: функциональный экзом (~5%) и произвольно выбранных областей

Геномная архитектура

- позиция нуклеосом
- трёхмерная структура генома: 3C (hypothesis-free) и 5C (высокопроизводительный анализ)

Эпигенетика

- метилирование: сиквенс после бисульфидной обработки или ChIP-Seq
- возможно, машины третьего поколения смогут определять модификации напрямую

Взаимодействие нуклеиновых кислот с белками

- ChIP-Seq
- полногеномный неселективный анализ мест посадки белков
- гиперчувствительность к нуклеазам, FAIRE-Seq
- анализ специфичности взаимодействия (напрямую), поиск консенсусных последовательностей

Транскриптом

Полнота анализа

- все типы РНК: кодирующие последовательности, малые некодирующие РНК, антисенс-транскрипция и т.д.
- чувствительность: при адекватной организации эксперимента, можно быть уверенным, что замечены все действующие РНК

RNA-Seq позволяет

- определять как относительный, так и абсолютный уровни экспрессии
- аннотировать новые и уточнять аннотацию известных генов
- смотреть отдельно ядерную, цитоплазматическую, ассоциированную с рибосомами (в цитоплазме и на мембранах) и т.п. РНК
- выявлять присутствие в образце микроорганизмов и вирусов

Чувствительность

Сравнительный функциональный анализ

Сравнение геномов разных видов позволяет

- выявить функциональные элементы генома
- связать эволюционные нововведения с появлением в геноме новых функциональных участков, типа проект «Origins of Multicellularity»
- выявить механизмы и движущие факторы эволюции для разных эпох
- разобраться, что отличает нас от ближайших и дальних родичей

Внутривидовое сравнение геномов

- причина тонких различий в реакции на факторы окружающей среды (болезни) и поведении
- факторы, определяющие сложные полигенные признаки (рост, вес, предрасположенность к распространенным болезням)
- механизмы рекомбинации
- факторы отбора в «историческое» время
- история расселения человечества

Медицина

Микробиология

- выявление патогенов: холера на Гавайях, токсичная E.coli в Германии
- анализ симбиотических организмов: микробиом

Предрасположенность к болезни → генотип

- GWAS, вовлеченные гены, молекулярные причины болезни, выбор мишени
- редкие болезни, семейный анализ: «быстрый» тест сиквенсом кодирующих последовательностей
- иммунный репертуар B- и T-клеточных рецепторов: подверженность болезням, ответ на вакцинацию
- неинвазивная и пренатальная диагностика по ДНК в крови

Болезнь → профиль экспрессии

- классификация опухолей, тонкий и/или ранний диагноз
- спектр мутагенеза – причина опухолеобразования

Медицина

Проекты по сиквенсу разных типов опухолей

- спектр мутагенеза, «завершить» список онкогенов
- тестирование лекарств на клеточных линиях
- тонкая диагностика и выработка «стандартных» способов лечения
- неинвазивная диагностика (посттреатмент): опухолевые ДНК и клетки в крови

Медицина

Превентивная медицина

- Болезнь Альцгеймера
- Диабет II типа

Персональная медицина

- Генетически опосредованная чувствительность к лекарствам (varfarin)