



Секвенирование нового поколения (NGS)

Васильев Геннадий Владимирович

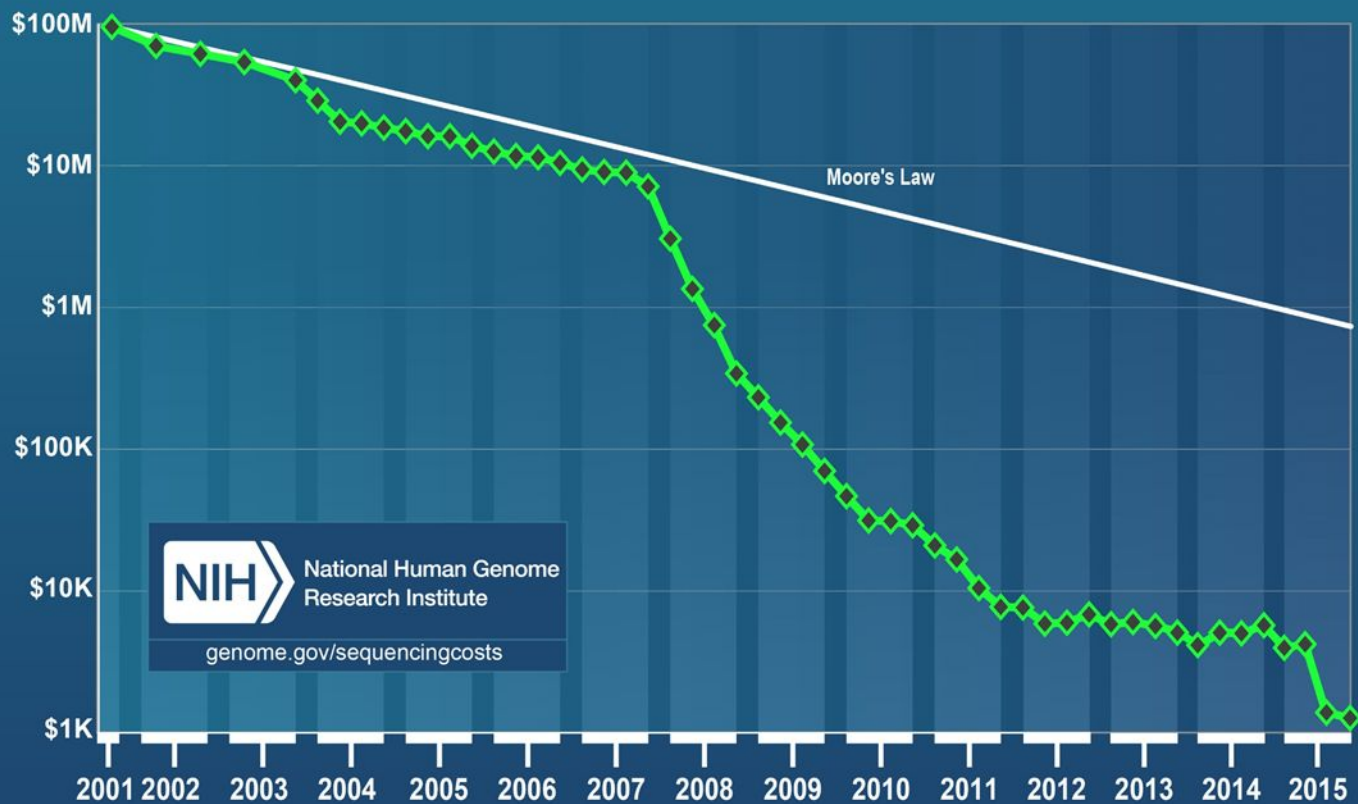
Институт цитологии и генетики СО РАН



Революция цены в геномных исследованиях



Cost per Genome



454 GS20

1000 ГЕНОМОВ

SOLID 3.0
Illumina GA2



Второе поколение секвенаторов – NGS – создатели геномики

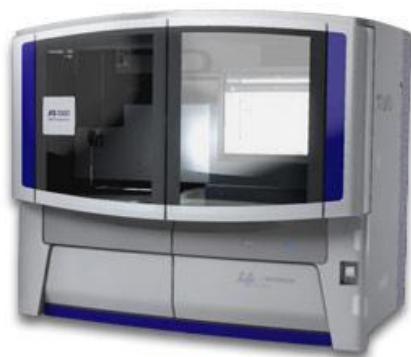


Roche FLX Titanium



©2010, Illumina Inc. All rights reserved.

Illumina HiSeq



SOLiD 5500



Ion Proton



Принцип пиросеквенирования

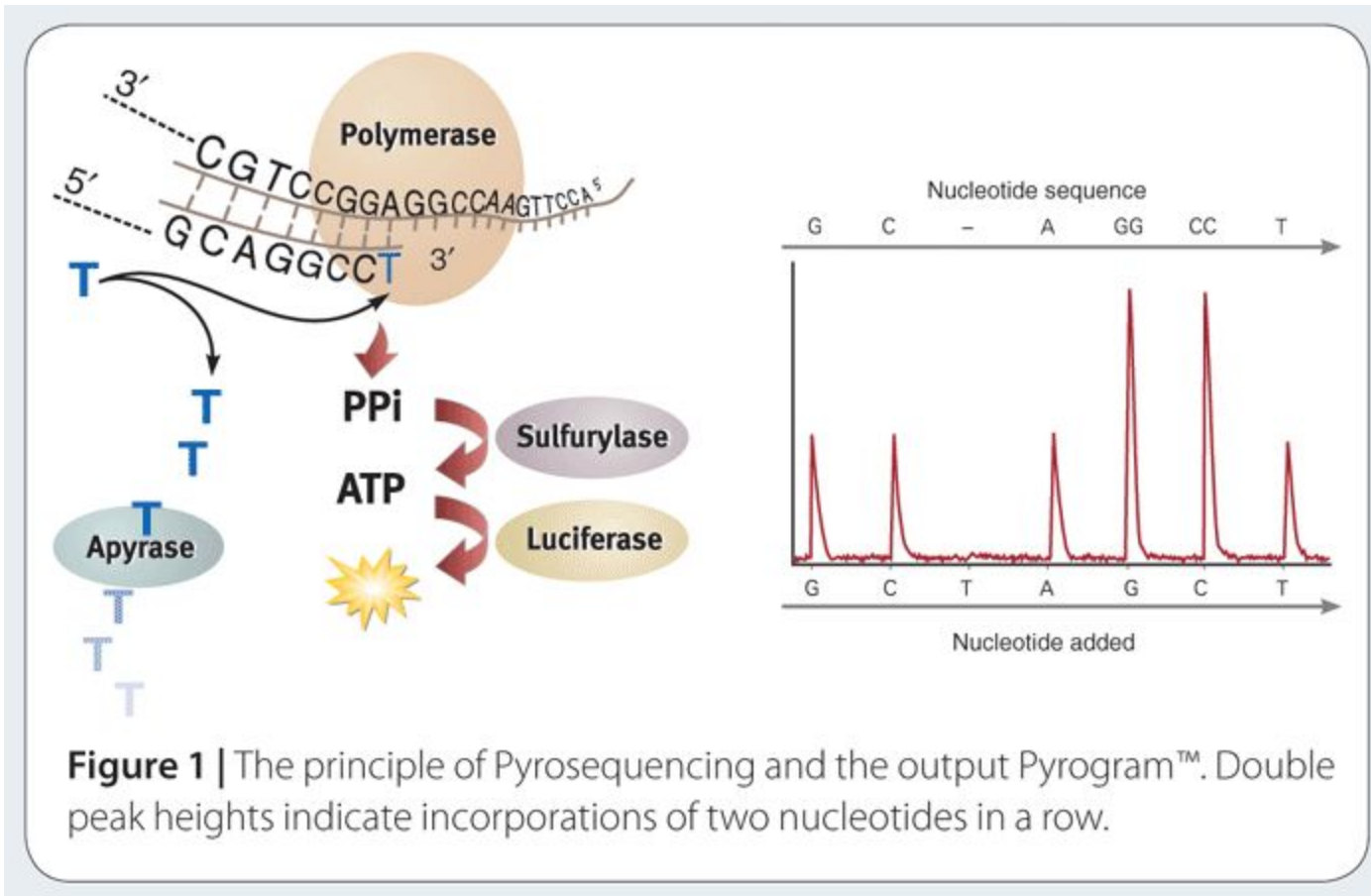


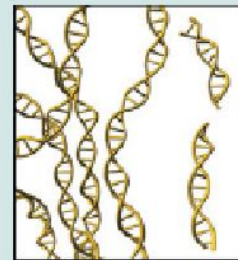
Figure 1 | The principle of Pyrosequencing and the output Pyrogram™. Double peak heights indicate incorporations of two nucleotides in a row.



Пиросеквенирование: приготовление библиотек

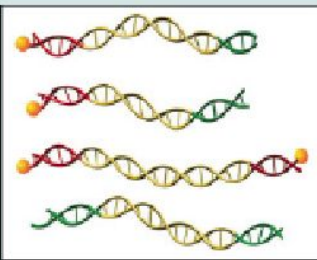
DNA Library Preparation and Titration

4.5 hours

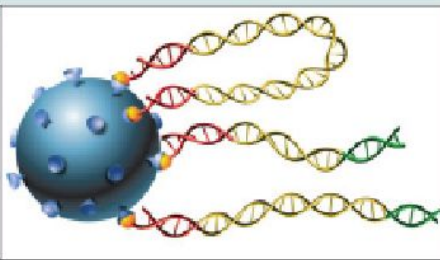


emPCR

10.5 hours

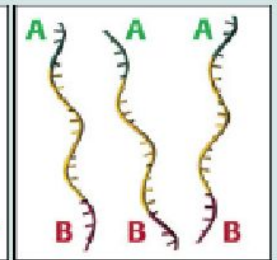
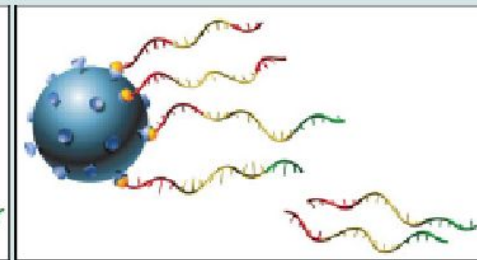


8 hours



Sequencing

5.5 hours



gDNA

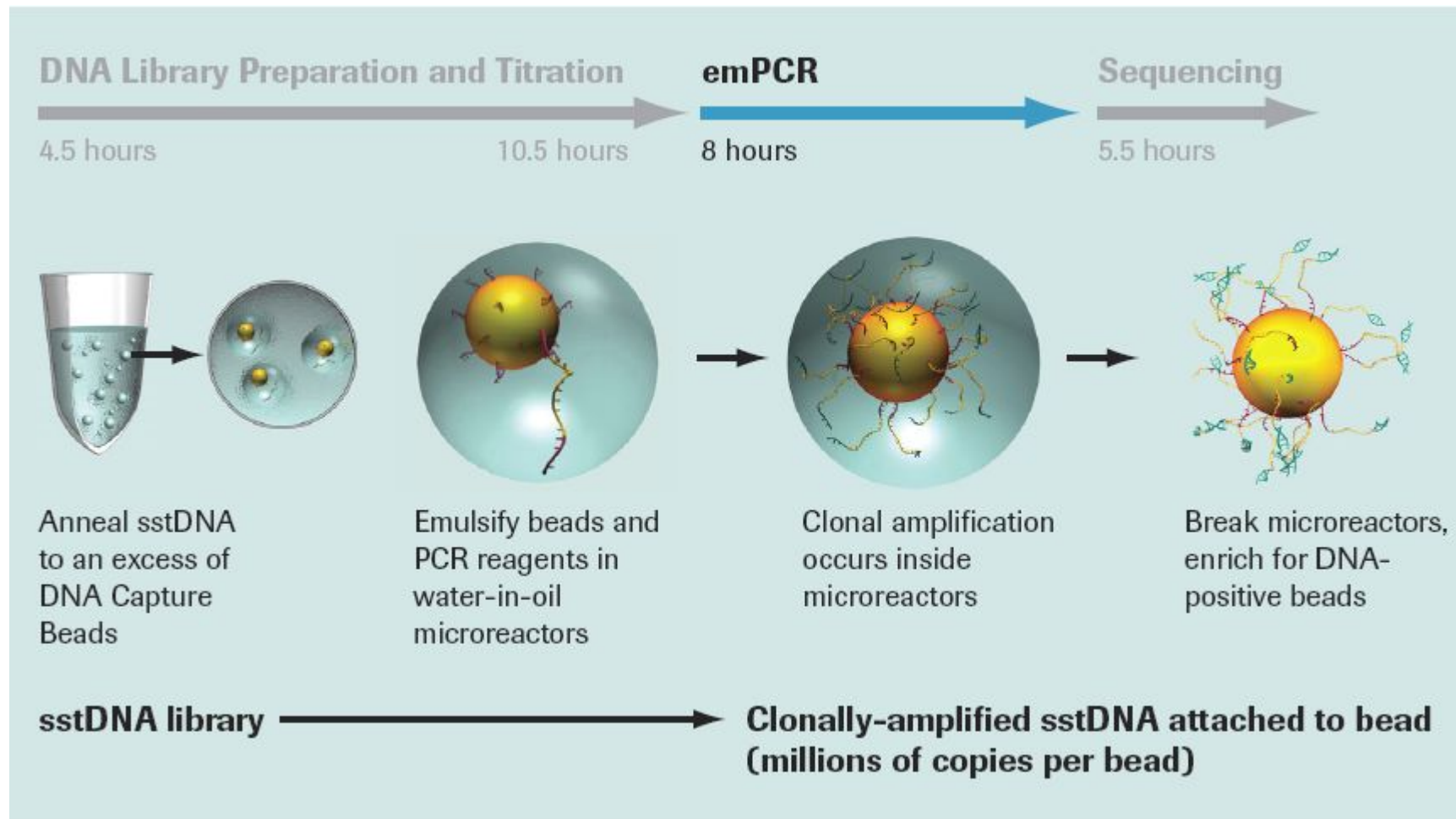
→ sstDNA library

- Genome fragmented by nebulization
- No cloning; no colony picking
- sstDNA library created with adaptors. The adaptors are used as primers, and for binding to beads.
- A/B fragments selected using streptavidin-biotin purification





Пиросеквенирование: ПЦР в эмульсии





Пиросеквенирование: секвенирование



DNA Library Preparation and Titration

4.5 hours

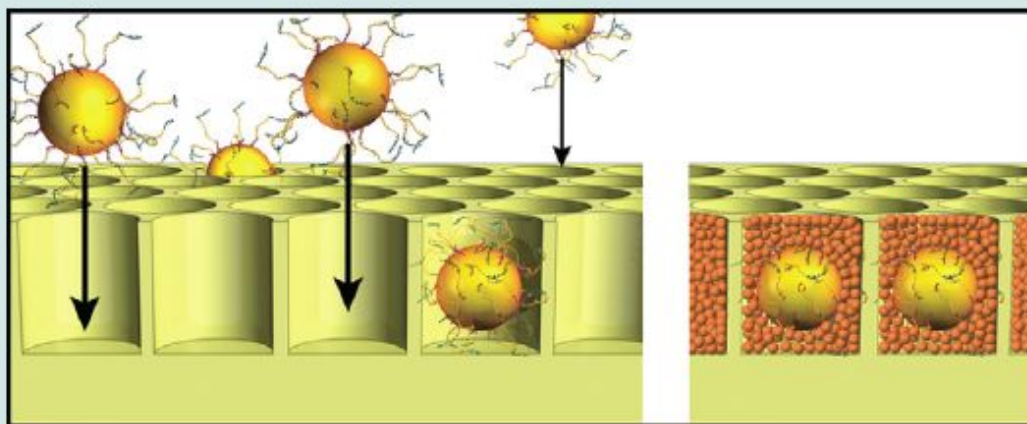
10.5 hours

emPCR

8 hours

Sequencing

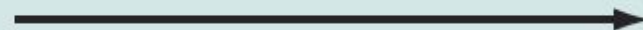
5.5 hours



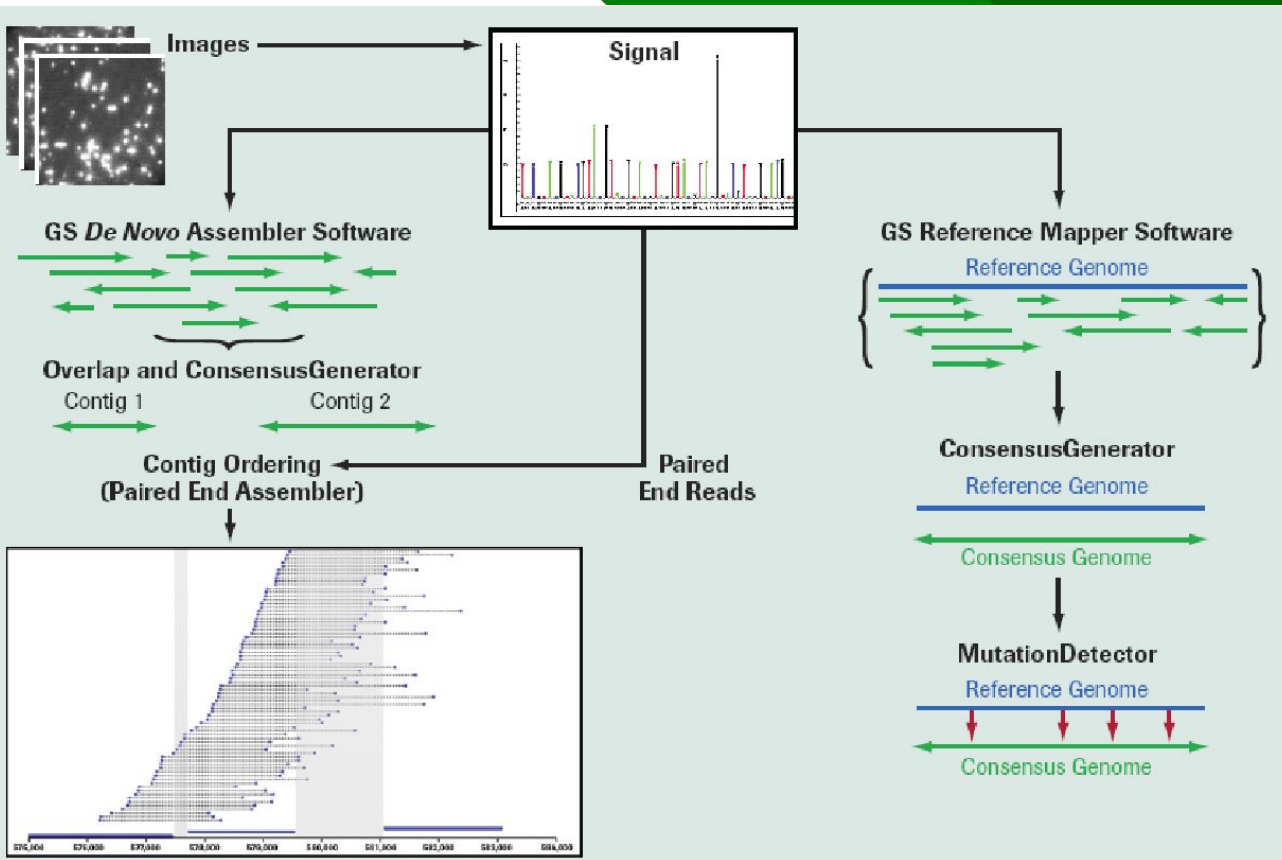
- Well diameter: average of 44 μm
- A single clonally amplified sstDNA bead is deposited per well
- 200,000 reads obtained in parallel on large-format PicoTiterPlate device

Amplified sstDNA library beads

Quality reads



Пиросеквенирование: анализ результатов



Особенности:

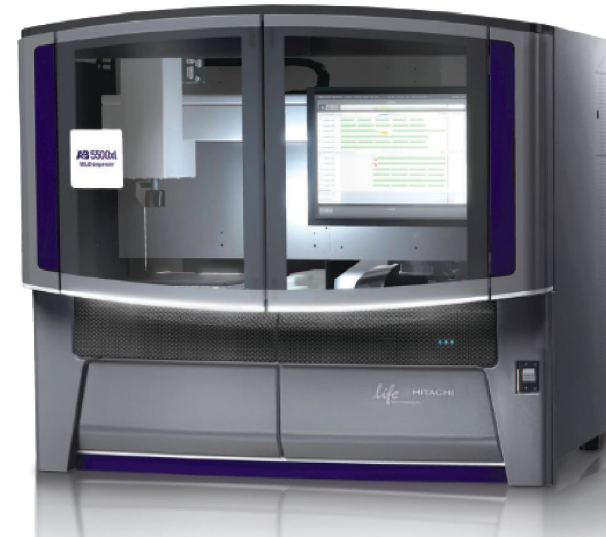
- 1) Лучшее качество и высокая цена
- 2) Невозможность разрешения политрактов длиннее 5-7-9 осн.
- 3) Длинные фрагменты (600-800 п.н.) без использования mate-pair библиотек
- 4) Активно применялся для секвенирования *de novo* и метагеномов

Ресеквенирование генома *S. cerevisiae* - 12,070,820 (12 Mb) при 20x покрытии
Total Genomic Coverage **95.14%**, Number of Contigs – **694**, Average Contig Size **16.547** bases

SOLiD 4



SOLiD 5500

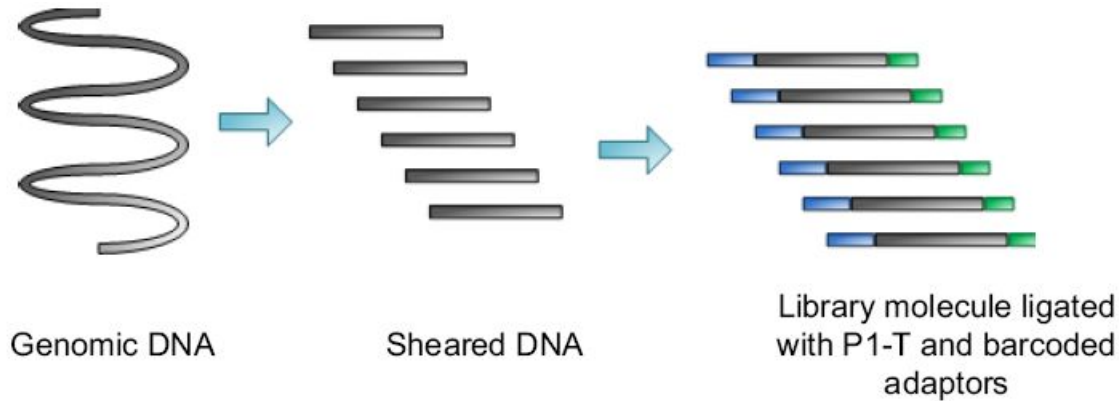


Особенности:

- 1) Самые короткие чтения 50 или 75 п.н., используется при ресеквенировании и анализе транскриптомов
- 2) Есть чтение mate-pair библиотек
- 3) Достаточно высокая точность при использовании специализированного ПО



Шаг 1: Подготовка библиотек



Fragment Library

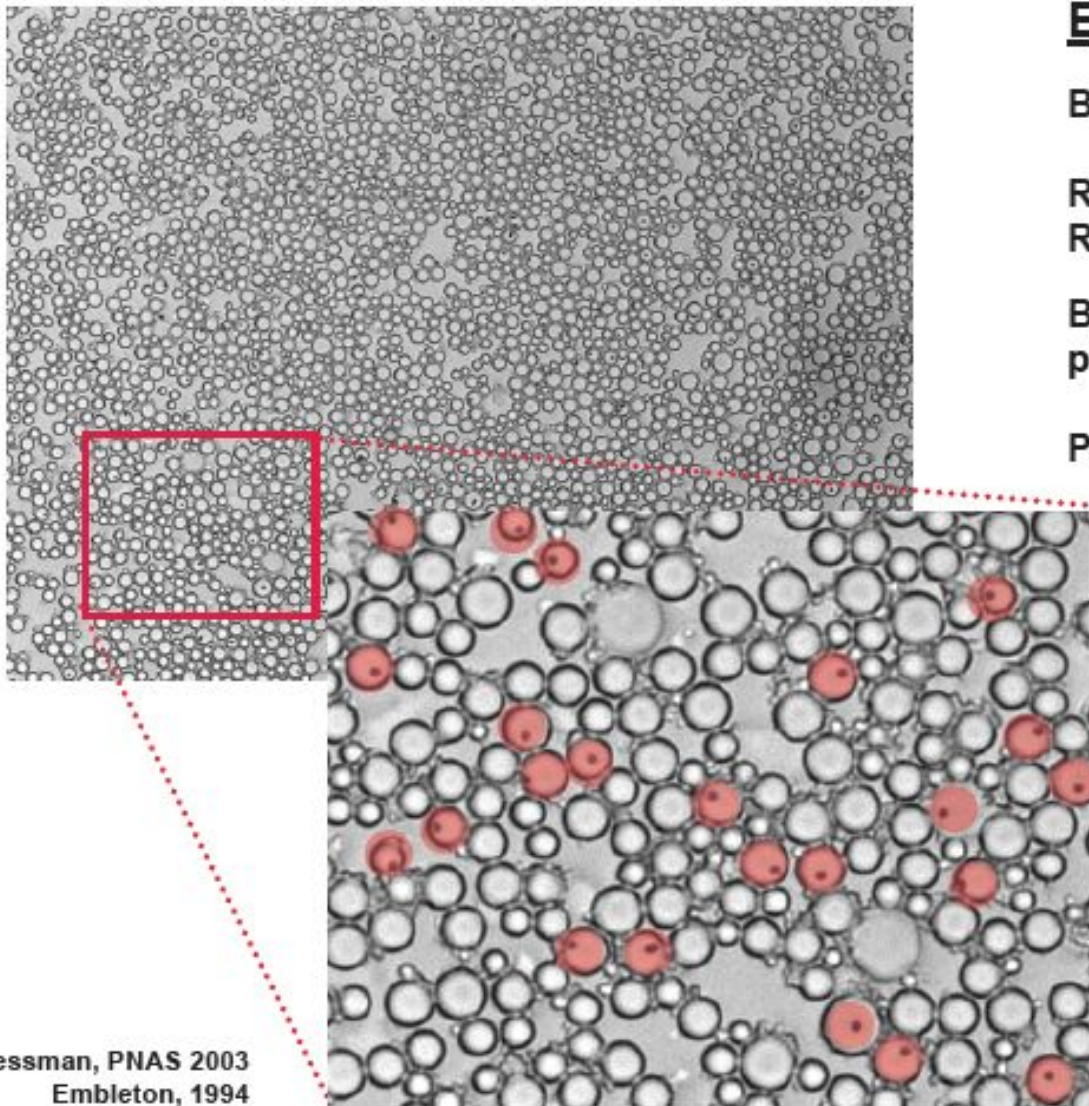


Mate-Paired Library





SOLiD: результат ПЦР в эмульсии



Emulsion Metrics

Bead size: 1 μm

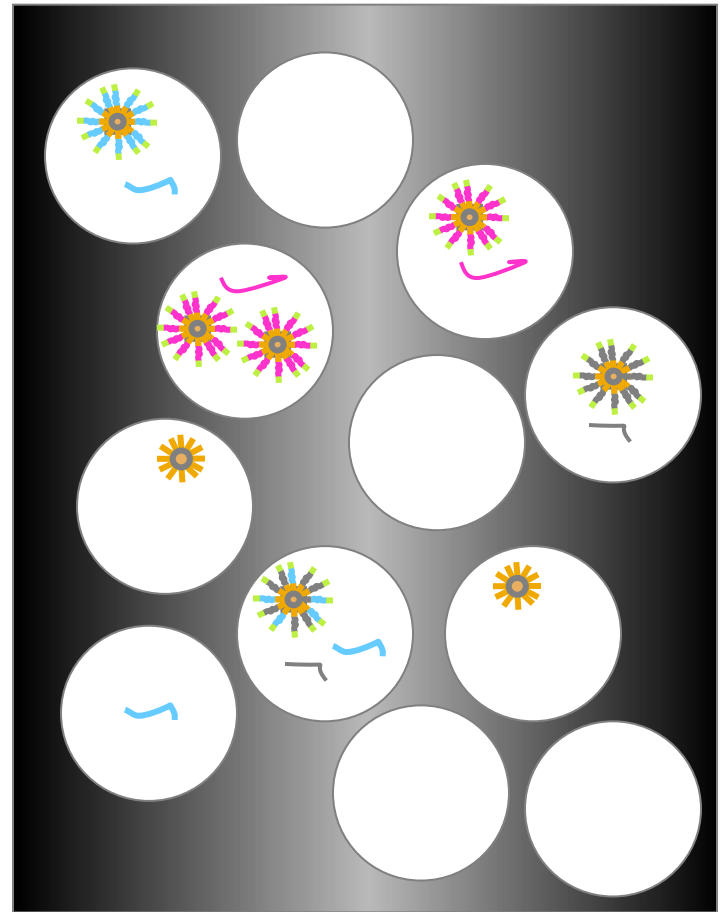
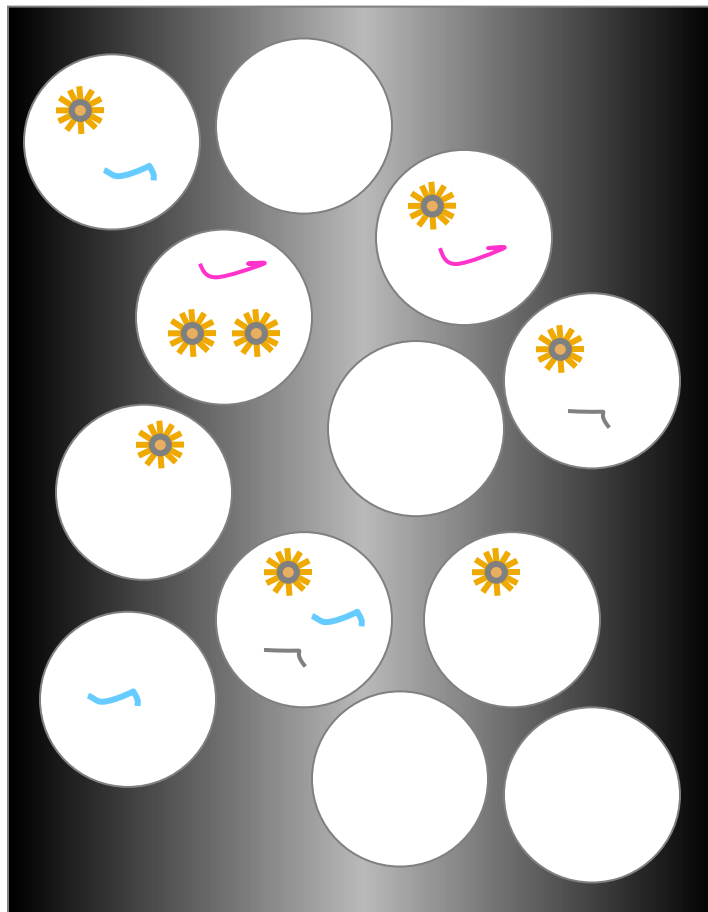
Reactor size: 4 μm

Reactor volume: 34 fL

Beads / emulsion plate (96-well): 1.6×10^9

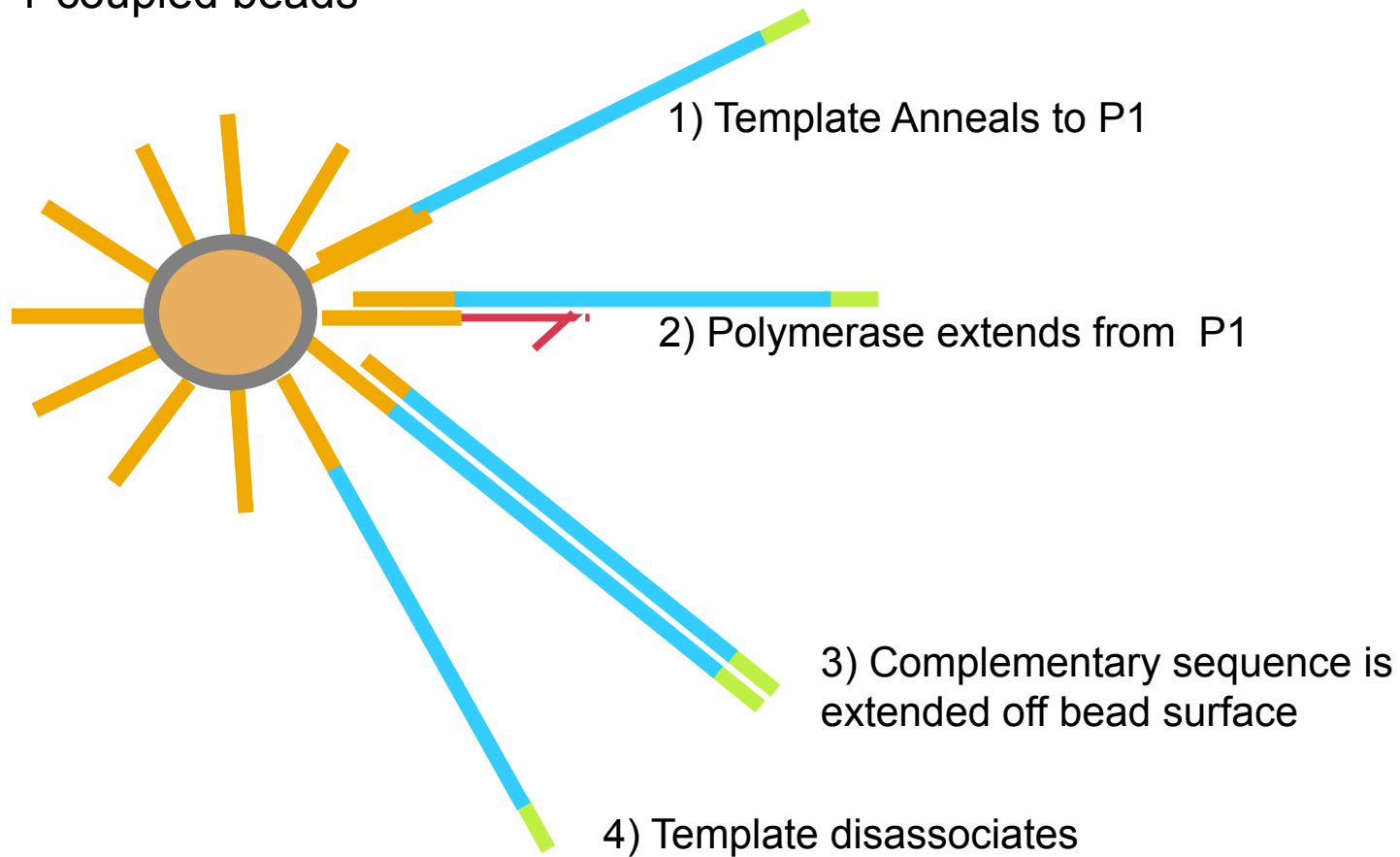
Post Enrichment: 150 – 300 $\times 10^9$

Dressman, PNAS 2003
Embleton, 1994



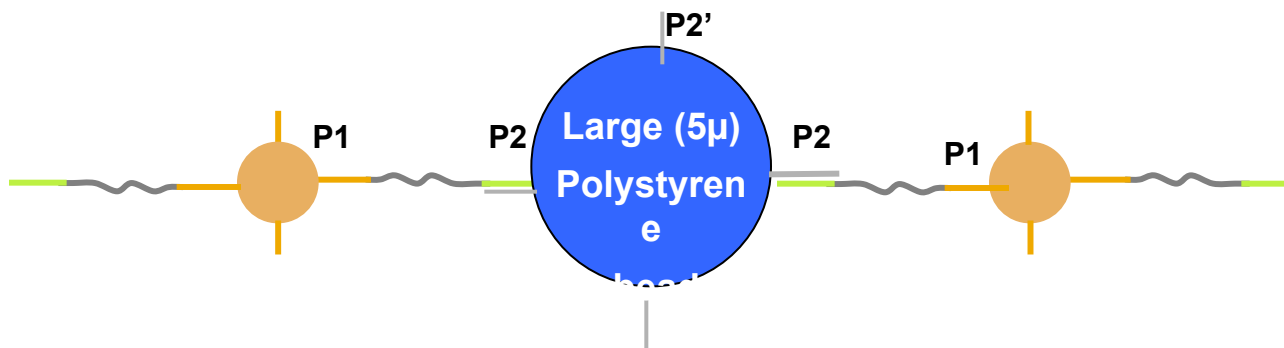


P1-coupled beads

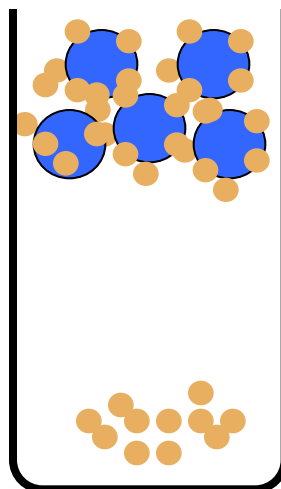




Шаг 3: обогащение полученных бидсов

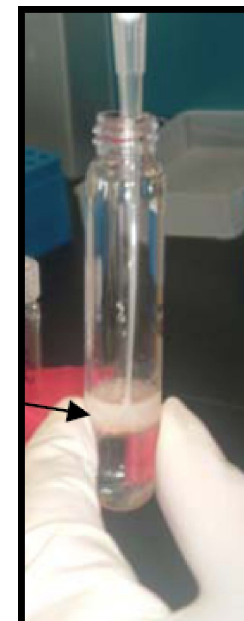


**Centrifuge in
glycerol gradient**



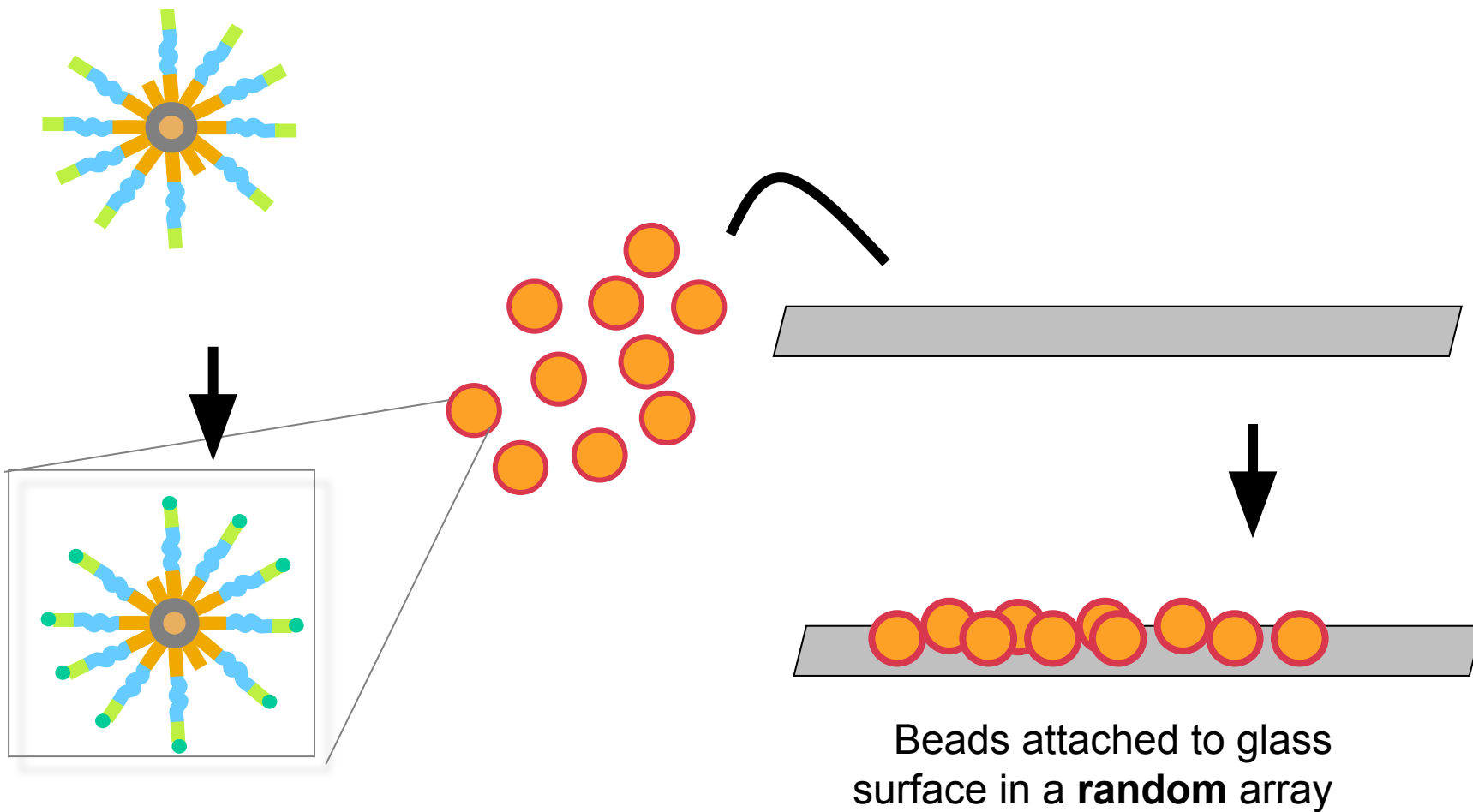
Supernatant
*Captured beads with
templates*

Pellet
Beads with no template





Шаг 4: 3'-модификация концов и нанесение бидсов

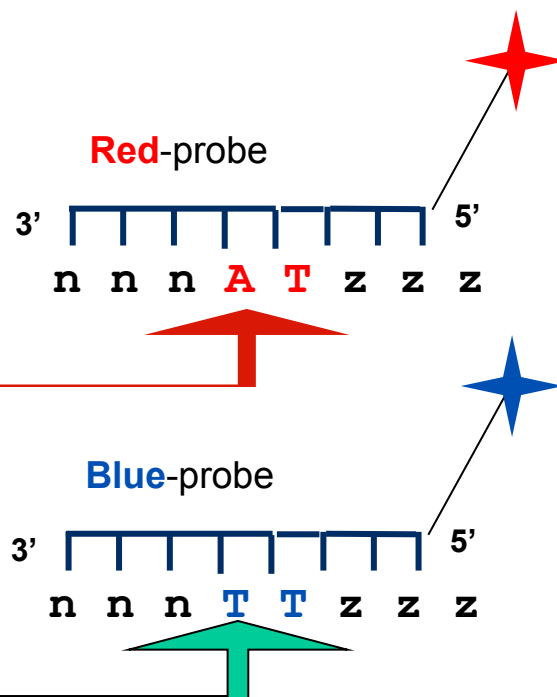
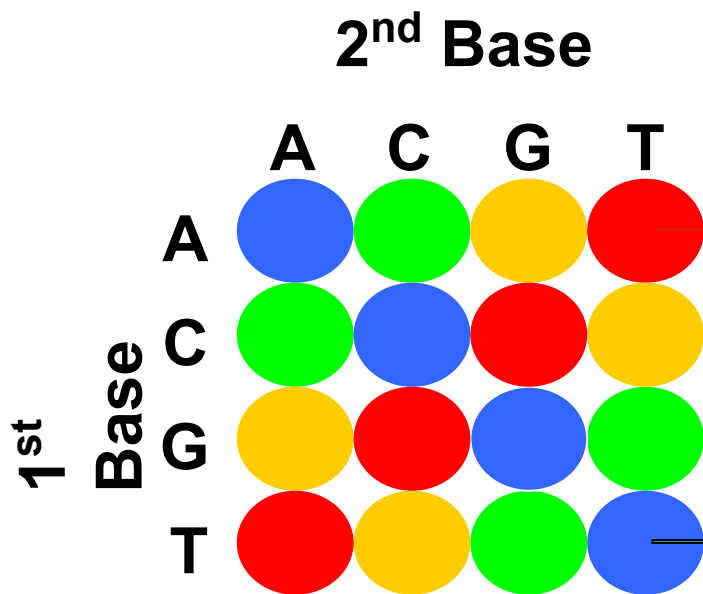




1,024 Octamer Probes (4^5)

4 Dyes, 4 dinucleotides, 256 probes per dye

2 Base Pair Encoding Using 4 Dyes



On our probes the 1st base encoded is position 4
the 2nd base encoded is position 5

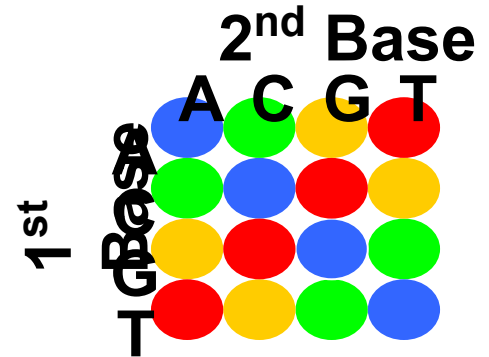


SOLiD read

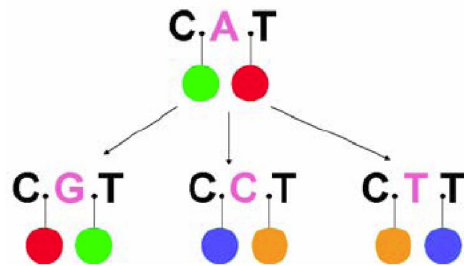
A C G G T C G T C G T G T G C G T



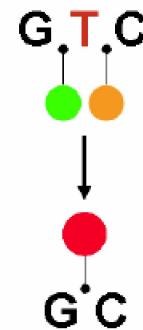
A · C · G · G · T · C · G · T · C · G · T · G · T · G · C · G · T



SNP



Insertion / Deletion





SOLiD: SNP и ошибки чтения

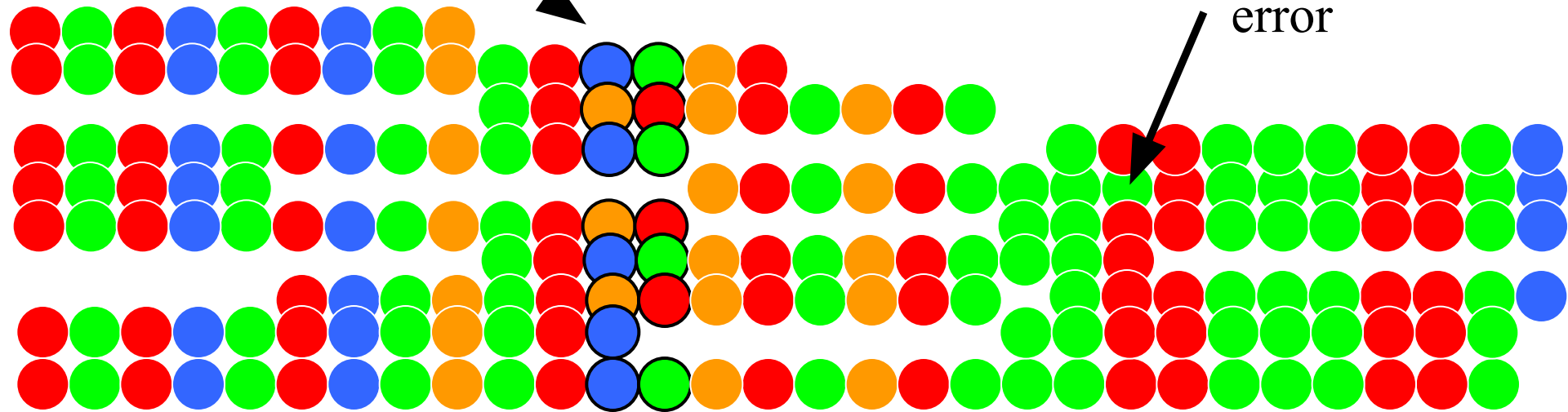


SNP

98% raw base accuracy

99,99% consensus accuracy (~ 20x coverage)

error



Reference

SNP 2 colors change



Технология фирмы Illumina – HiSeq, MySeq, NextSeq, NovaSeq



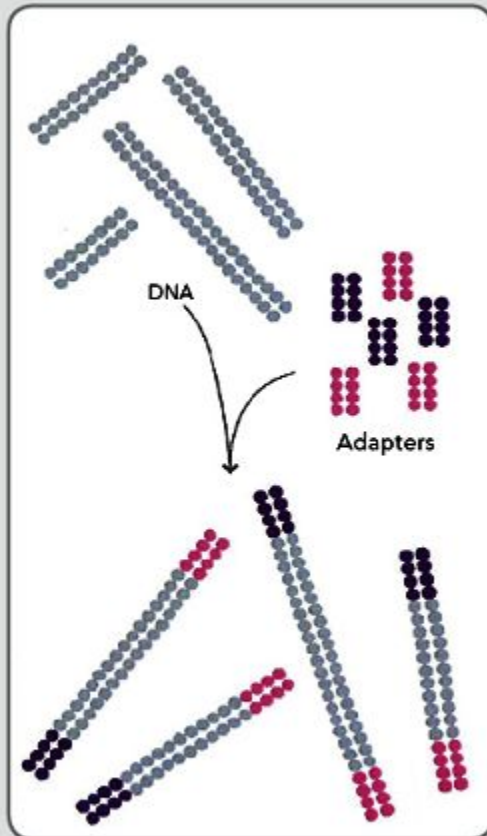
©2010, Illumina Inc. All rights reserved.

Особенности:

- 1) Чтения 50-300 п.н. прямые либо парные**
- 2) Есть чтение mate-pair библиотек**
- 3) Доминирующая на рынке (80-90%), применяется везде.**

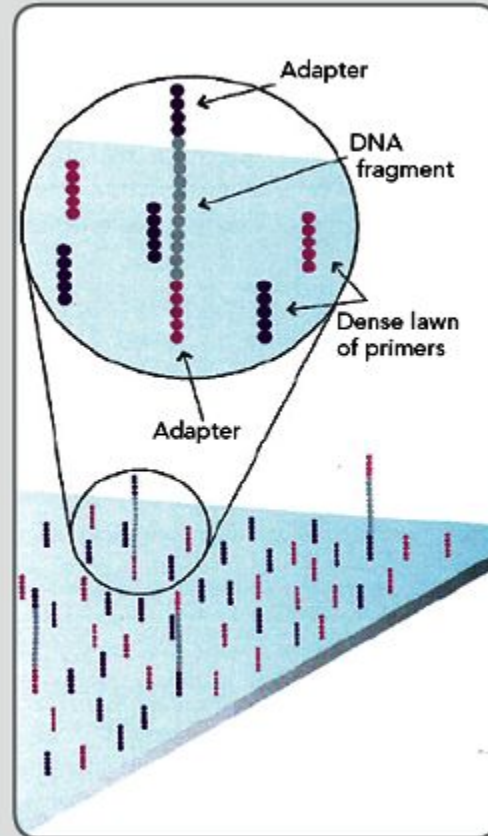


1. PREPARE GENOMIC DNA SAMPLE



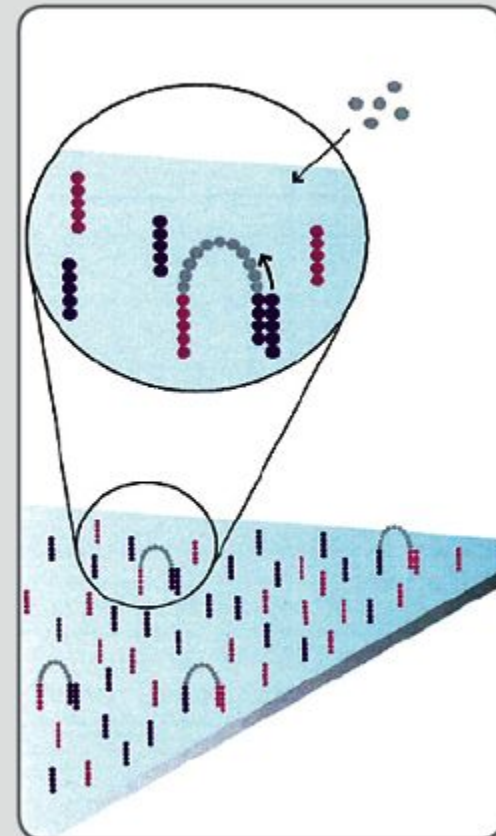
Randomly fragment genomic DNA and ligate adapters to both ends of the fragments.

2. ATTACH DNA TO SURFACE



Bind single-stranded fragments randomly to the inside surface of the flow cell channels.

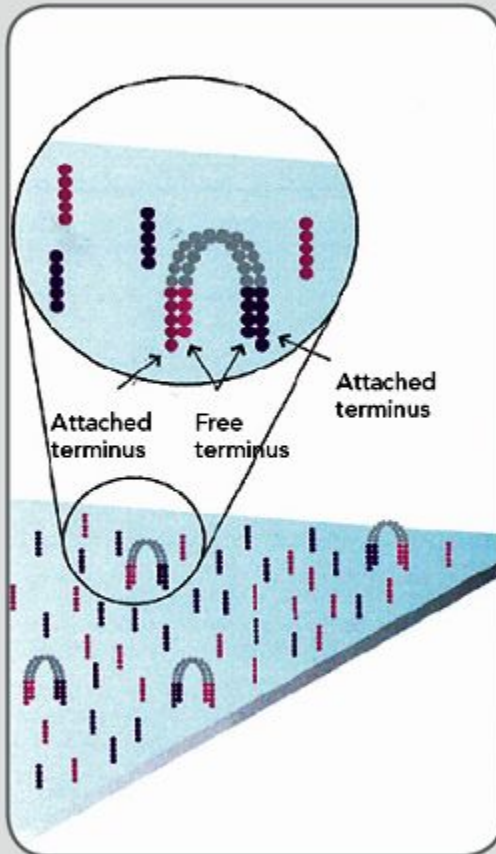
3. BRIDGE AMPLIFICATION



Add unlabeled nucleotides and enzyme to initiate solid-phase bridge amplification.

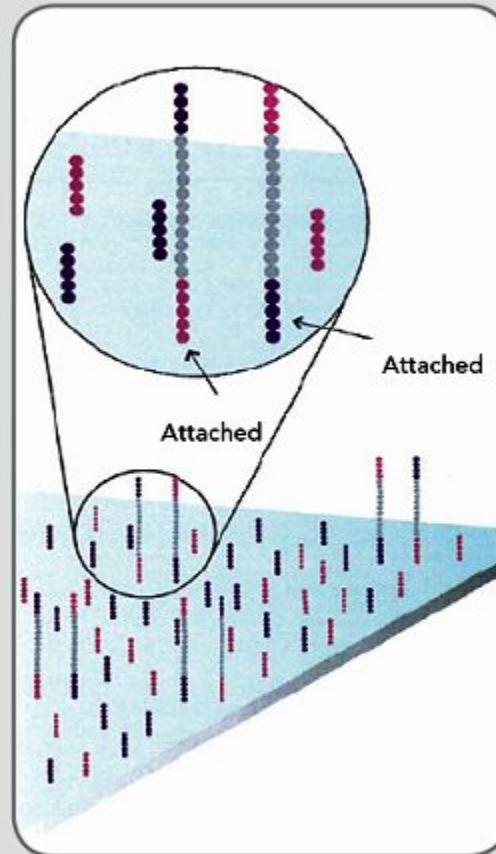


4. FRAGMENTS BECOME DOUBLE STRANDED



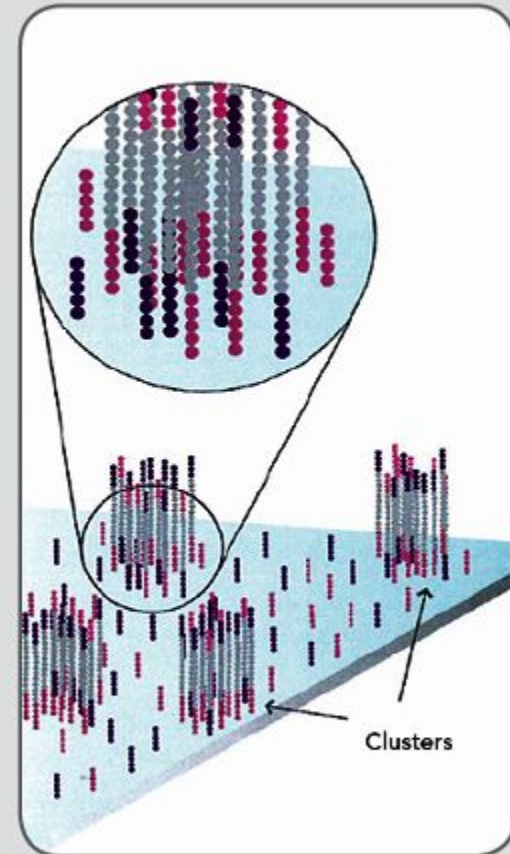
The enzyme incorporates nucleotides to build double-stranded bridges on the solid-phase substrate.

5. DENATURE THE DOUBLE-STRANDED MOLECULES



Denaturation leaves single-stranded templates anchored to the substrate.

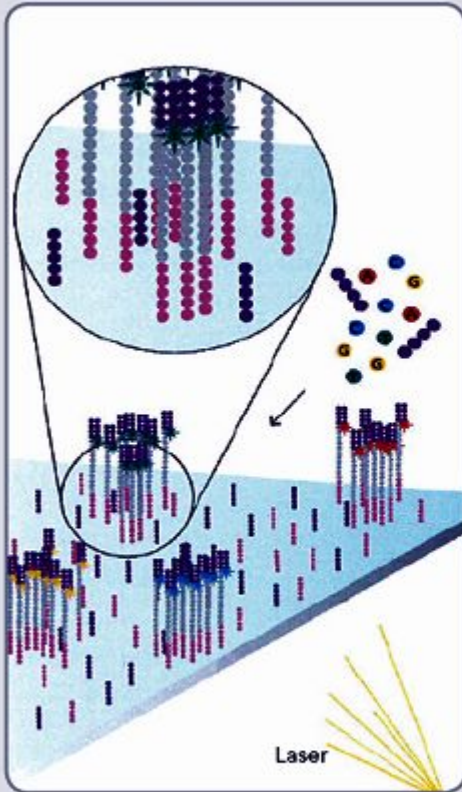
6. COMPLETE AMPLIFICATION



Several million dense clusters of double-stranded DNA are generated in each channel of the flow cell.

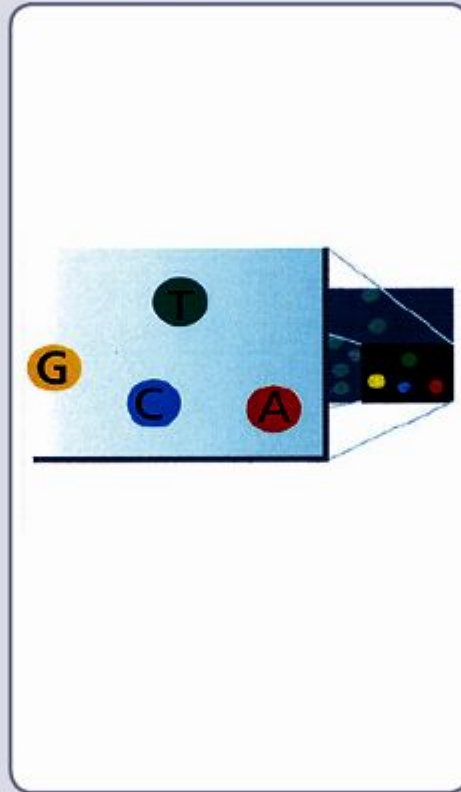


7. DETERMINE FIRST BASE



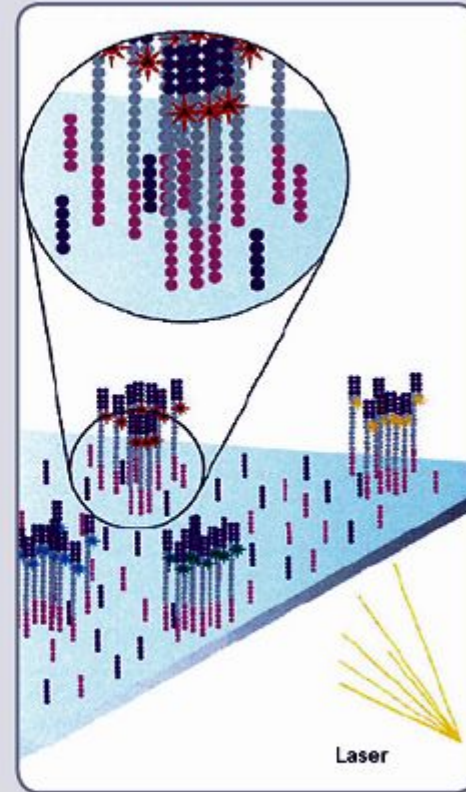
First chemistry cycle: to initiate the first sequencing cycle, add all four labeled reversible terminators, primers and DNA polymerase enzyme to the flow cell.

8. IMAGE FIRST BASE



After laser excitation, capture the image of emitted fluorescence from each cluster on the flow cell. Record the identity of the first base for each cluster.

9. DETERMINE SECOND BASE

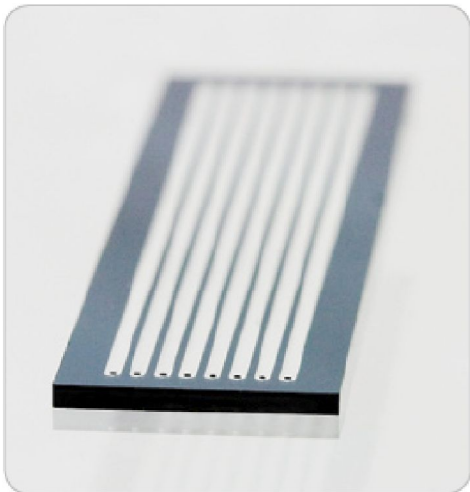


Second chemistry cycle: to initiate the next sequencing cycle, add all four labeled reversible terminators and enzyme to the flow cell.

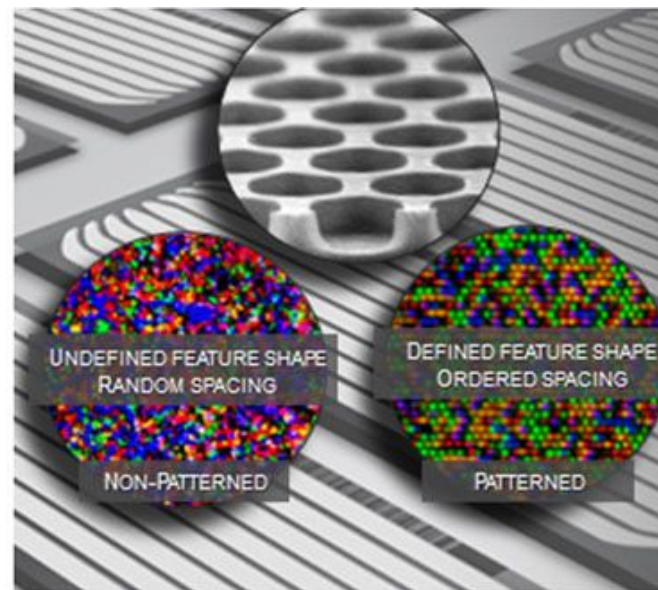
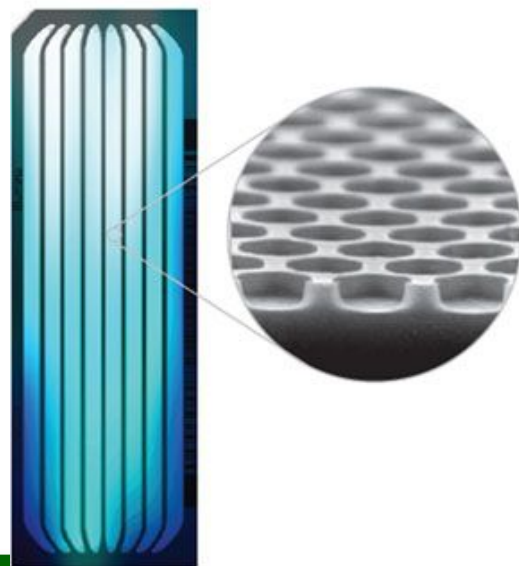
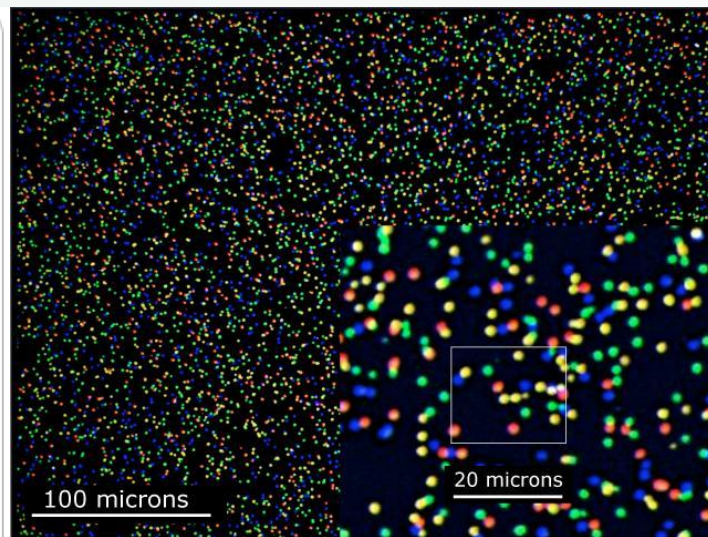


Illumina GA и HiSeq vs NovaSeq

FIGURE 1: ILLUMINA GENOME ANALYZER FLOW CELL

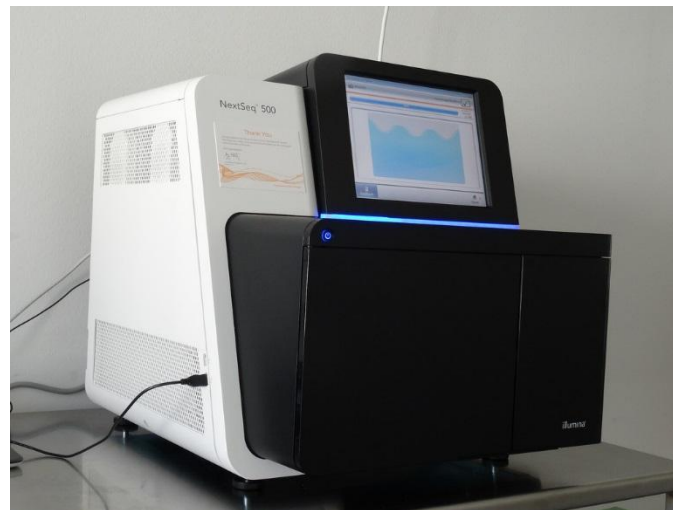
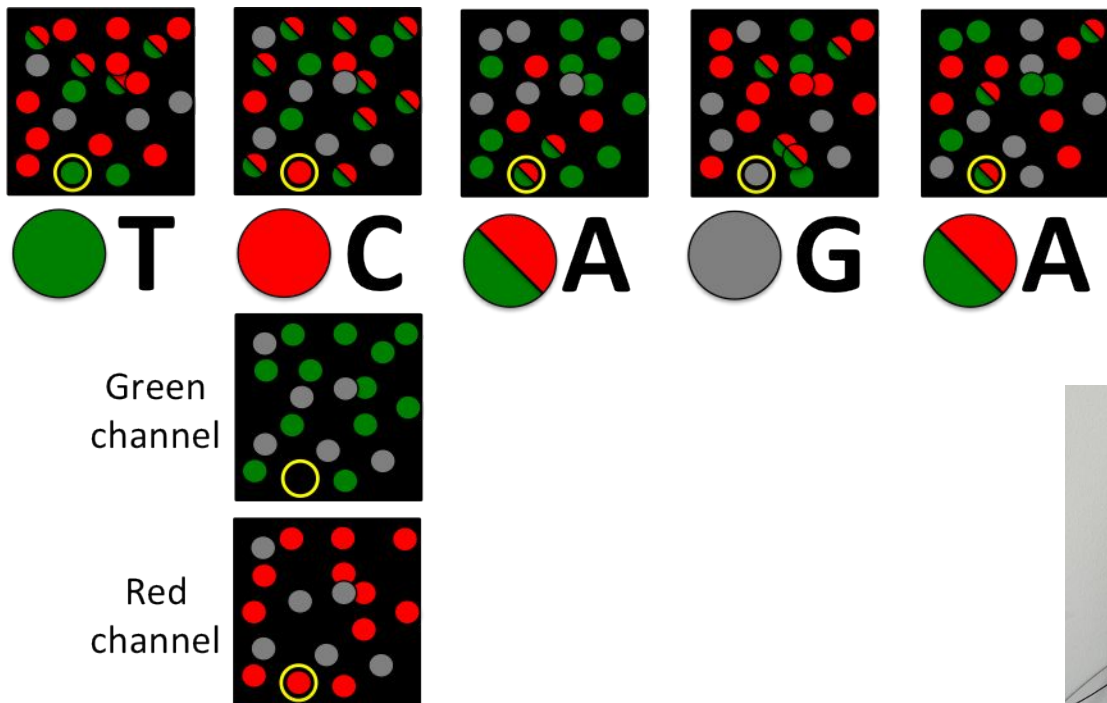


Up to eight samples can be loaded onto the flow cell for simultaneous analysis on the Illumina Genome Analyzer.





Illumina NextSeq и Nova Seq



Используются только два красителя и длины волн



Ion \ Proton Torrent



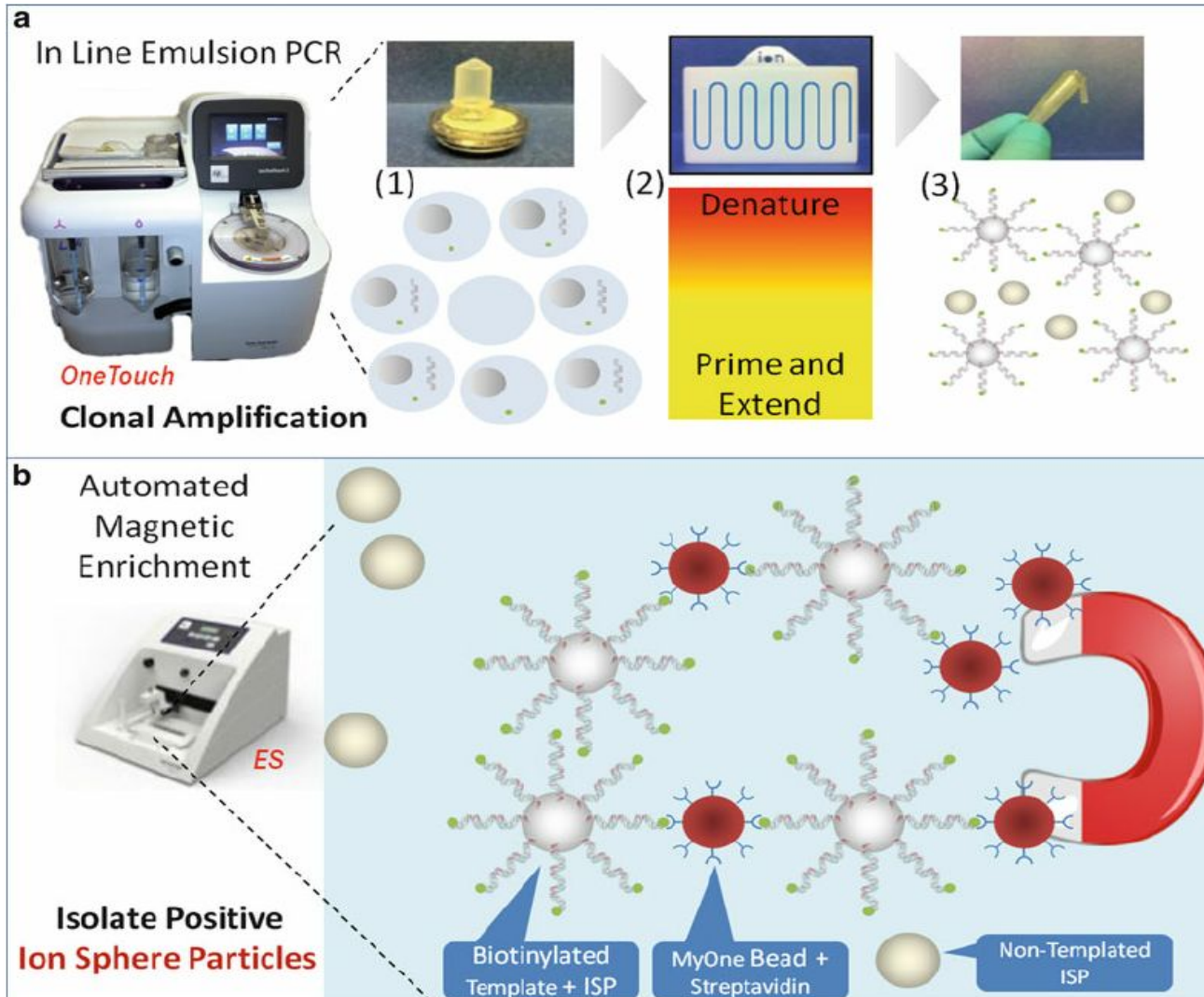
Особенности:

- 1) Не используется высокочувствительная оптика
- 2) Длинное чтение (до 400 п.н. непостоянной длины)
- 3) Самая быстрая скорость работы
- 4) Меньшая точность





Ion \ Proton Torrent - E-PCR

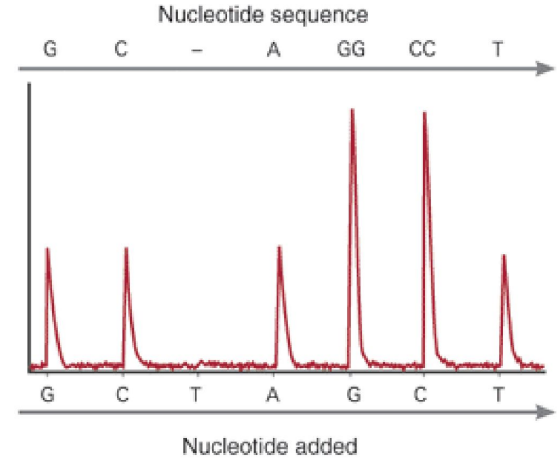
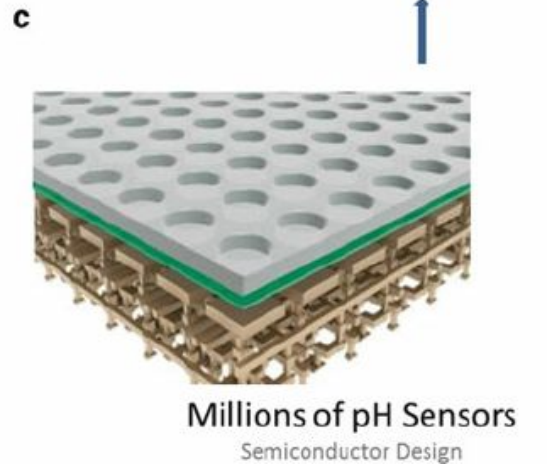
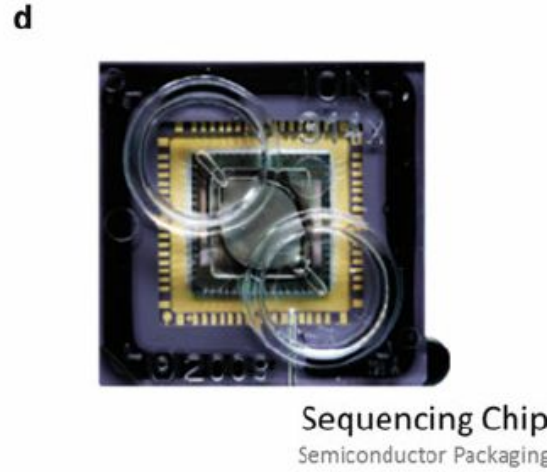
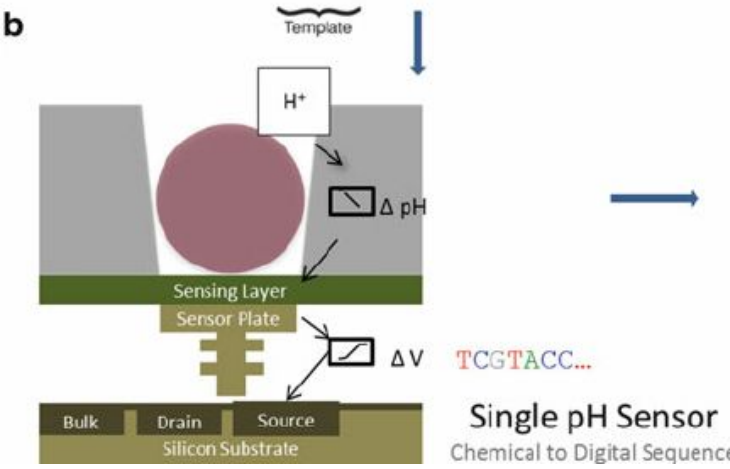
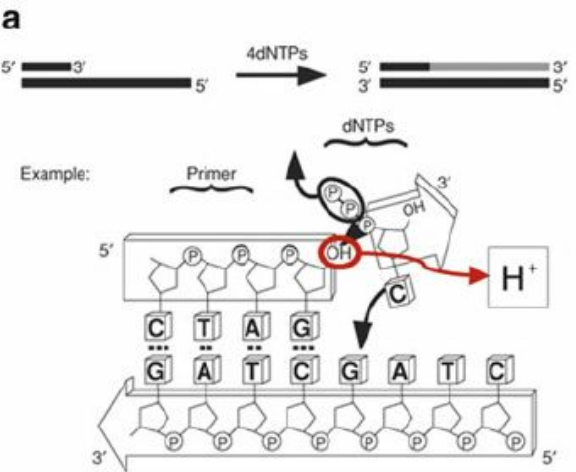




Ion \ Proton Torrent

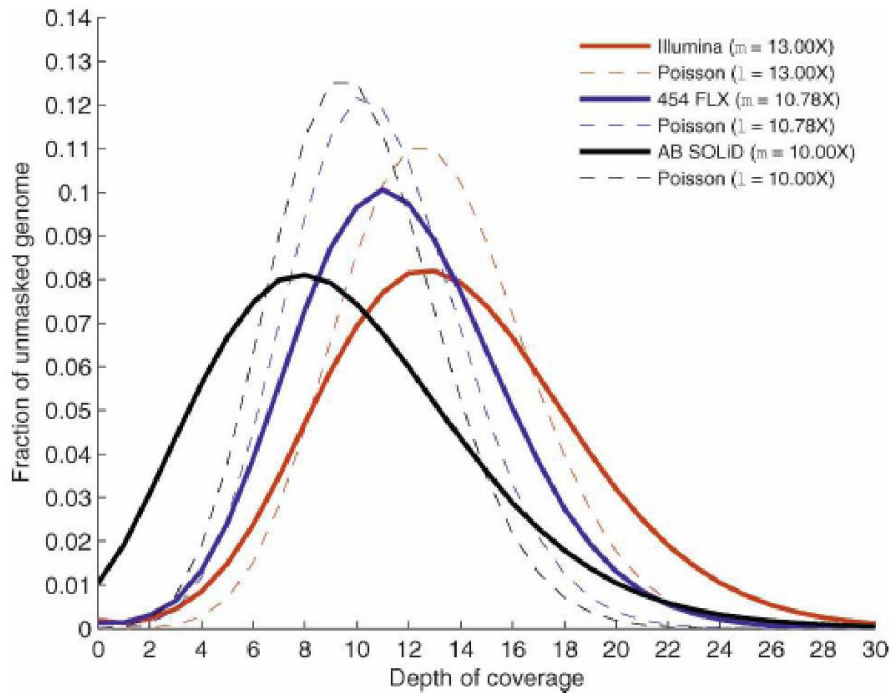
Principle and Elements of Semiconductor Sequencing

Simple Natural Chemistry of Sequencing-by-Synthesis with H⁺ release detection





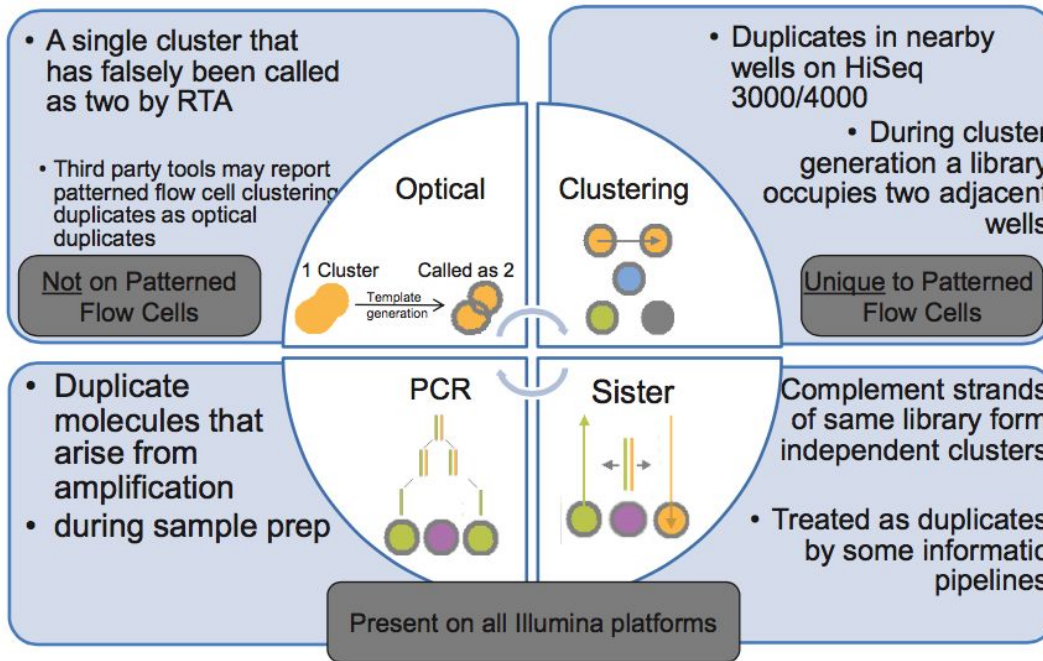
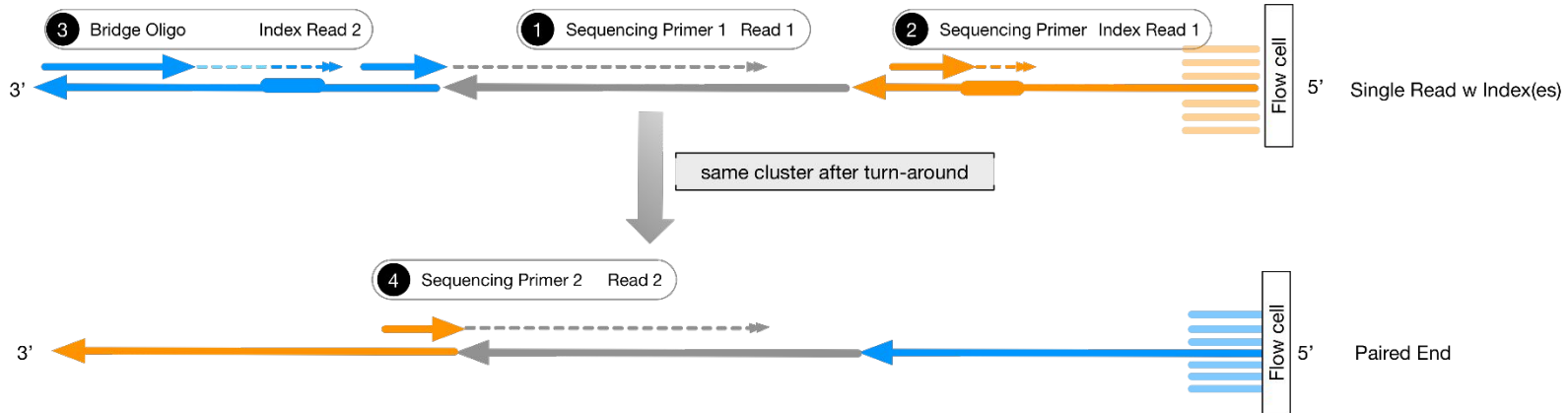
Ошибки и проблемы NGS



- Связанные с пробоподготовкой
- 1) Искажения представленности
 - 2) Химерные последовательности, ошибки баркодирования
 - 3) Фоновый шум

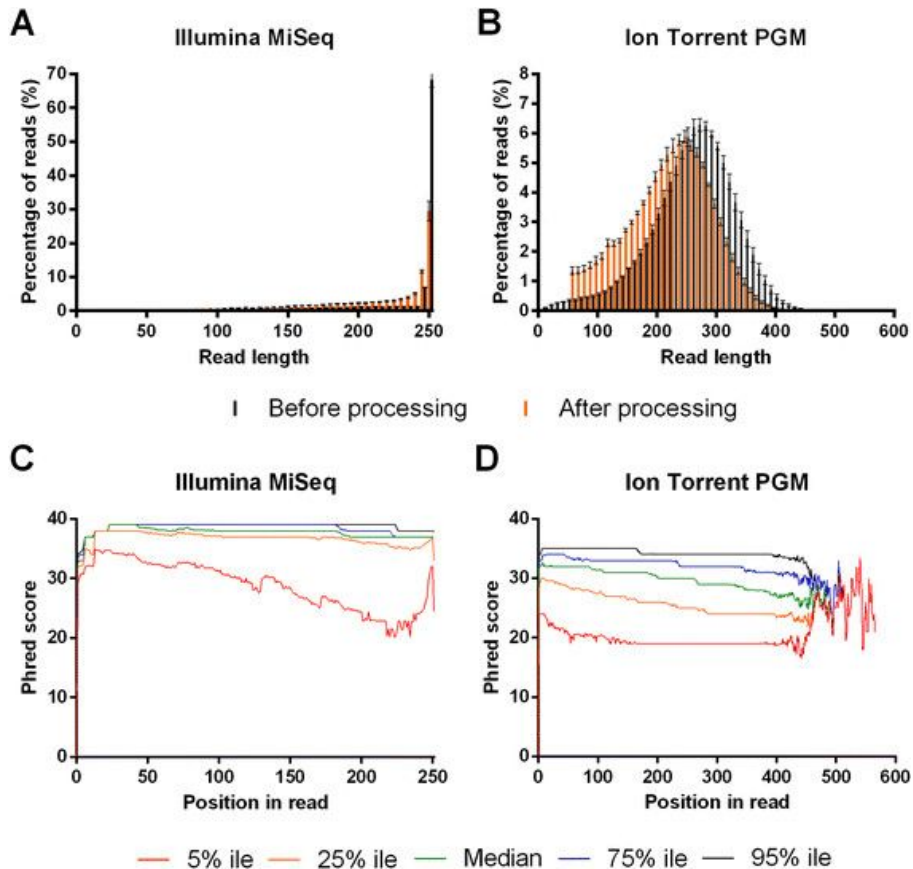


Ошибки кластеризации и баркодирования, поликлональность





Ошибки и проблемы NGS



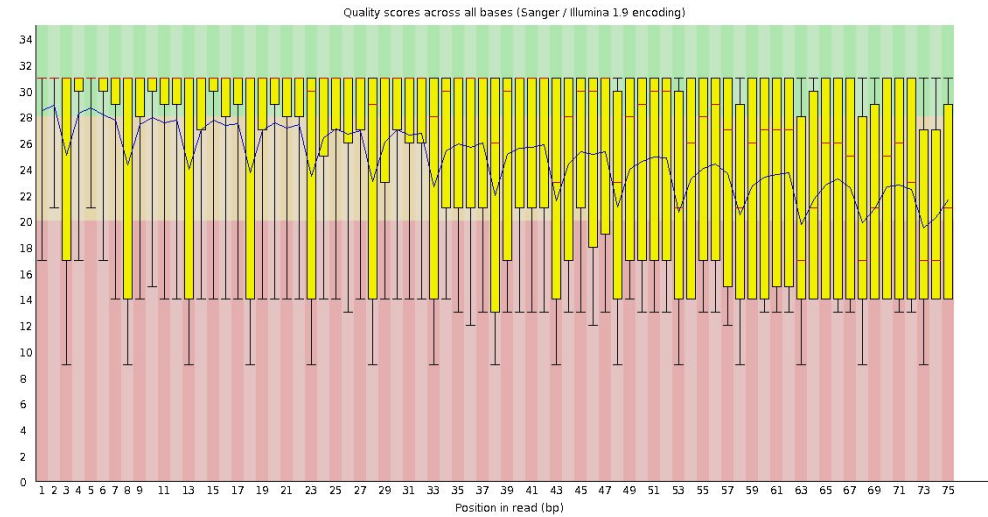
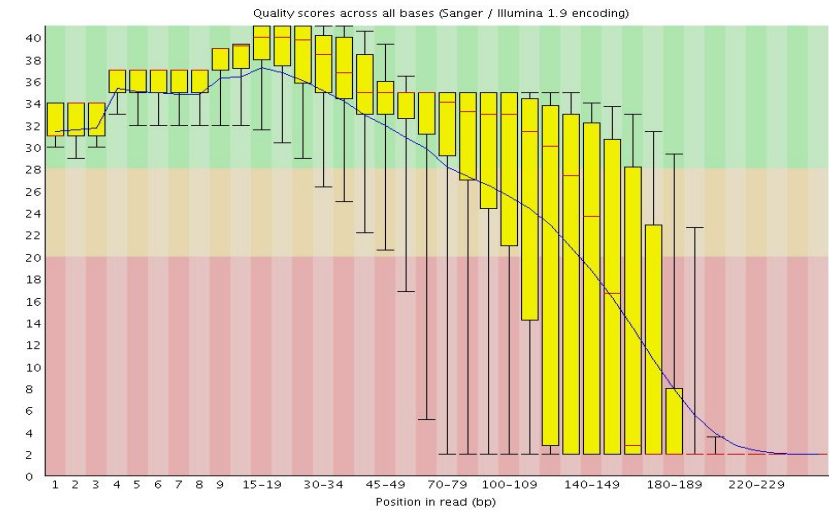
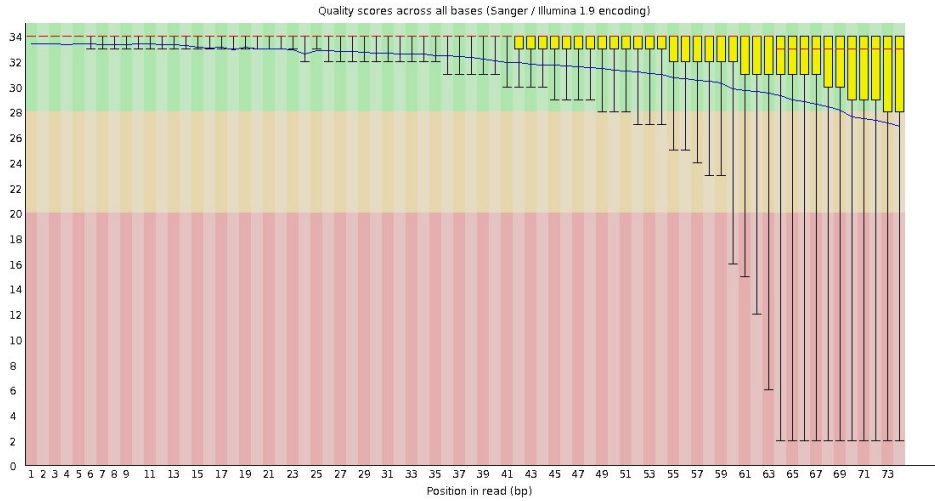
- Связанные с особенностями метода
- 1) Искажения представленности
 - 2) Систематические ошибки
 - 3) Химерные контиги, ошибки анализа данных
 - 4) Неверные алгоритмы анализа



Ошибки и проблемы NGS – анализ Q-scores



+ искажения, вносимые анализом





Спасибо за внимание