Фонд интеллект

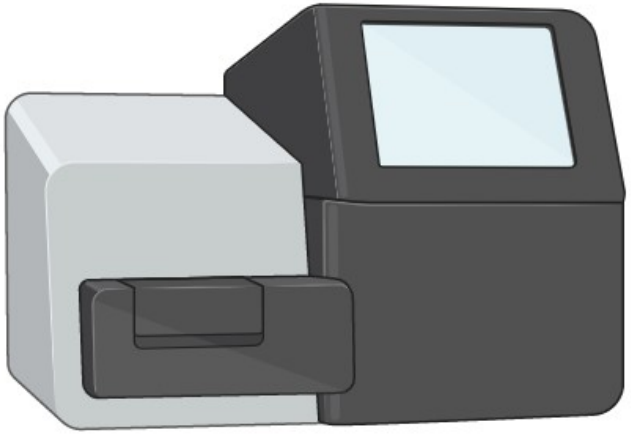
Машинное обучение в иммунологии и медицине

Получение данных в молекулярной биологии и их структура, часть 2. Методы и задачи секвенирования. Скрининг мутаций по данным WES

Ключевые различия между существующими методами секвенирования

Первое поколение

Читаем один фрагмент, представленный большим количеством копий



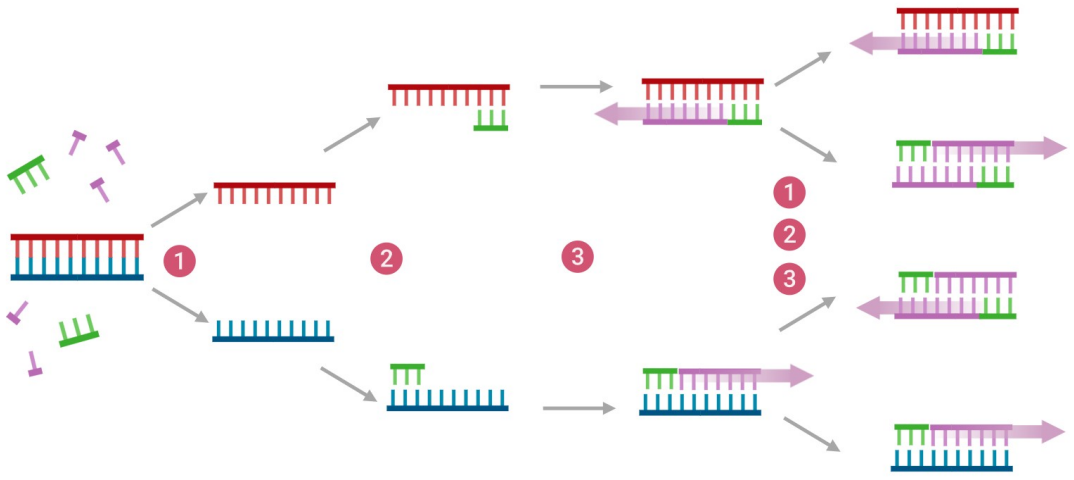
Второе поколение

Читаем много фрагментов одновременно

Третье поколение

Читаем исходную цепь:
а) меньше амплификаци,
б) видим нуклеотидные модификации

Компоненты и этапы ПЦР



В состав смеси входят:

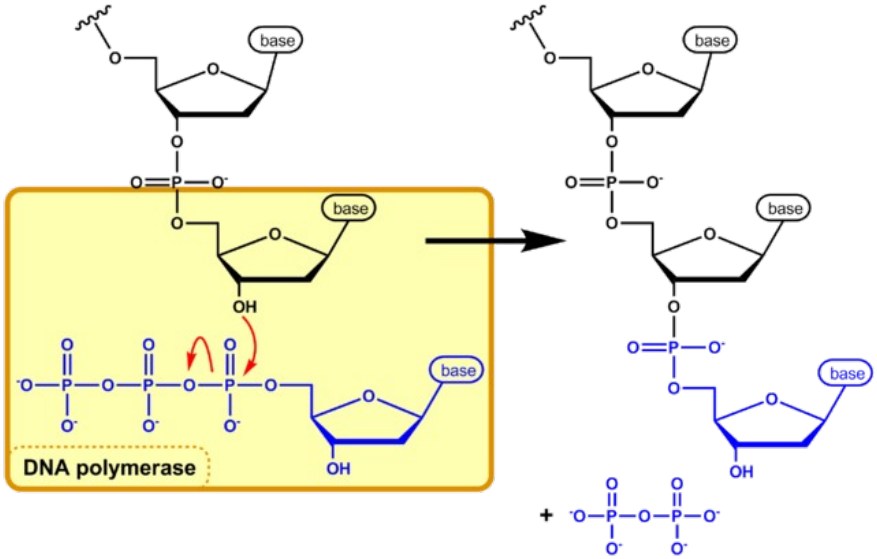
- Матричная ДНК
- ДНК-зависимая-ДНК-полимераза
- Дезоксирибонуклеотидтрифосфаты
- Праймеры
- Буфер

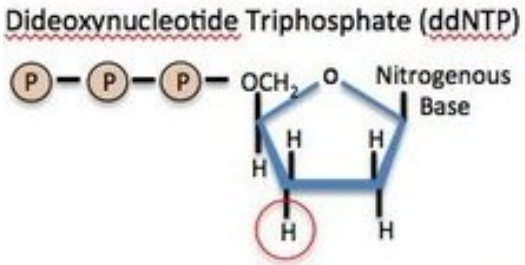
Денатурация — за счет нагрева до 95С происходит разрушение водородных связей между нуклеотидами, двуцепочечная ДНК расплетается.

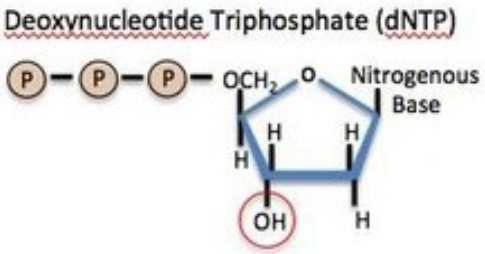
Отжиг праймеров — Температура зависит от последовательности праймеров. Понижение -> неспецифика; Повышение -> низкая эффективность связывания праймеров с ДНК.

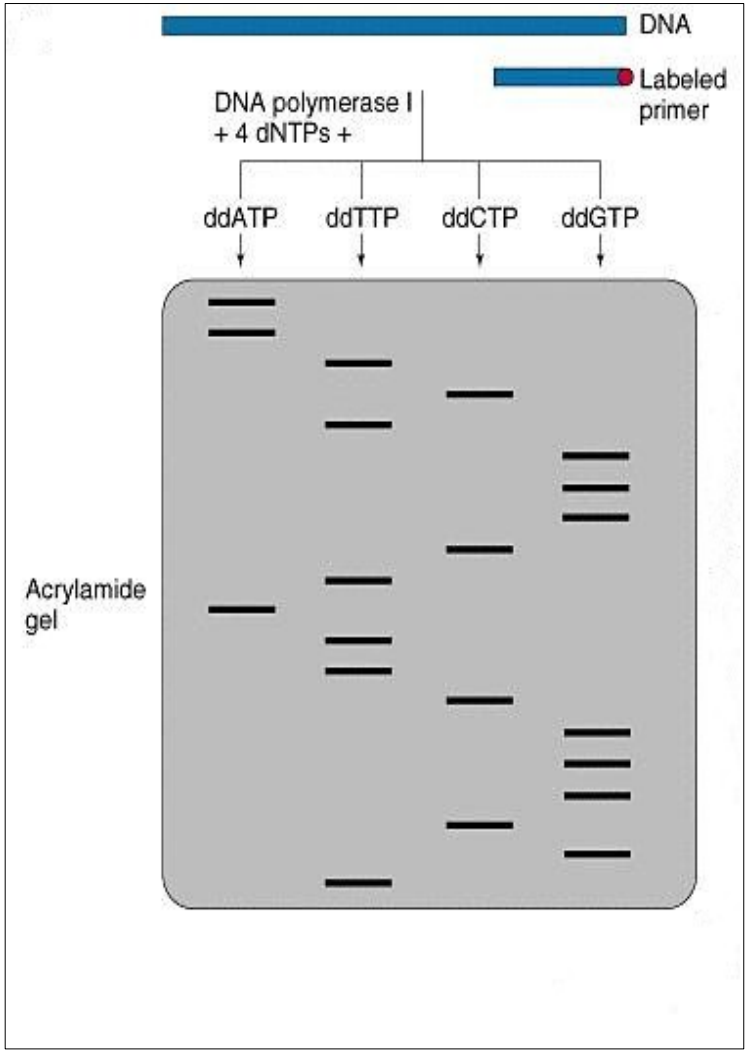
Элонгация — удлинение комплементарной цепи. Температура и время зависят от типа и скорости работы полимеразы.

Метод Сэнгера

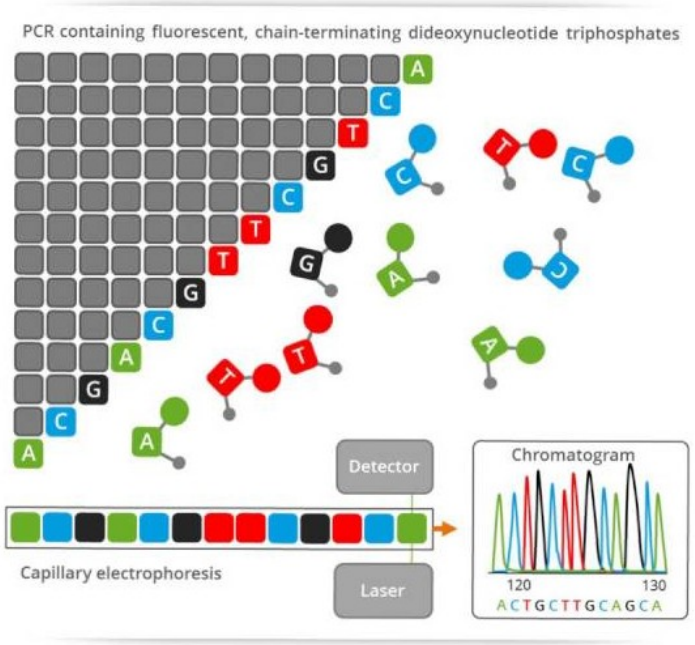


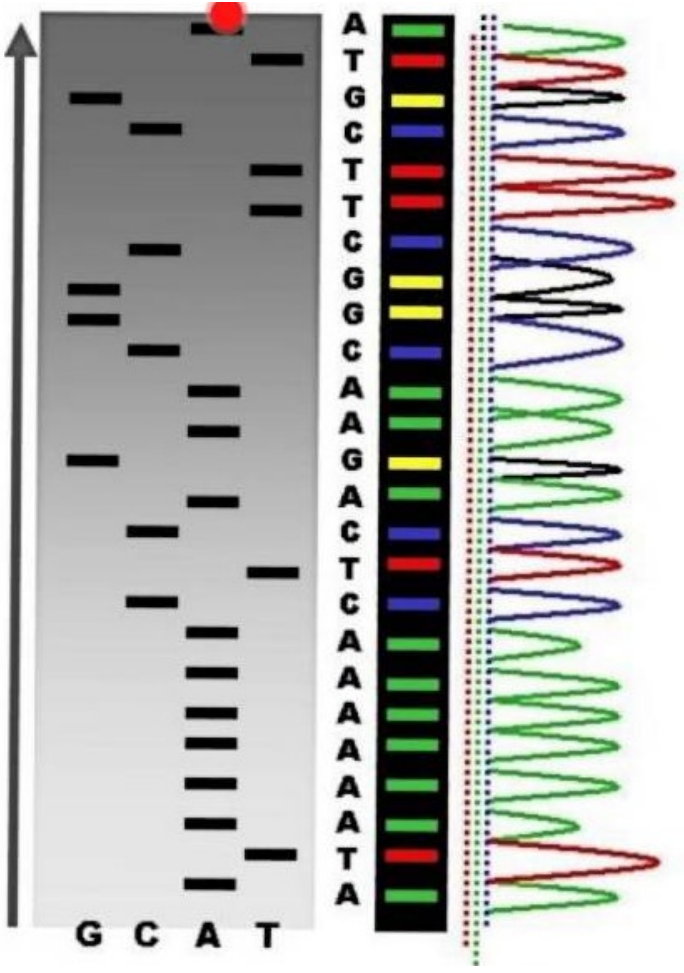




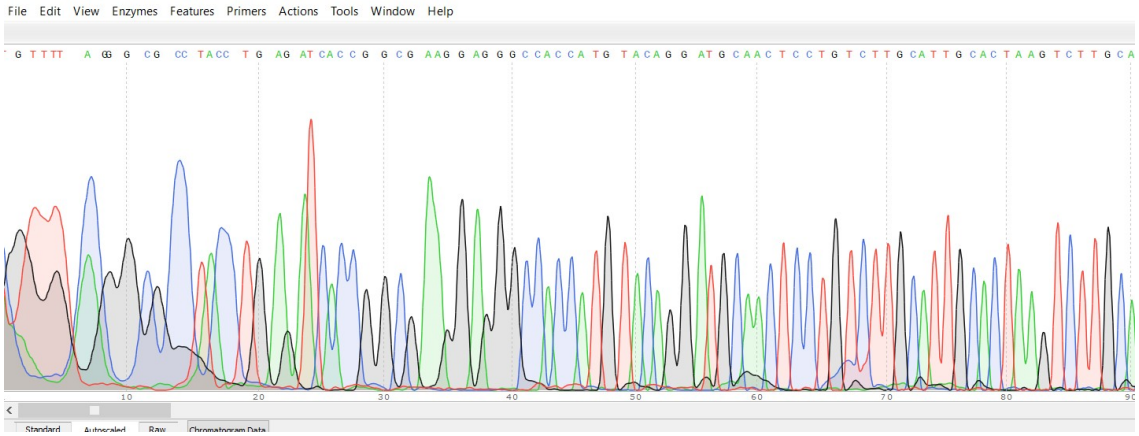


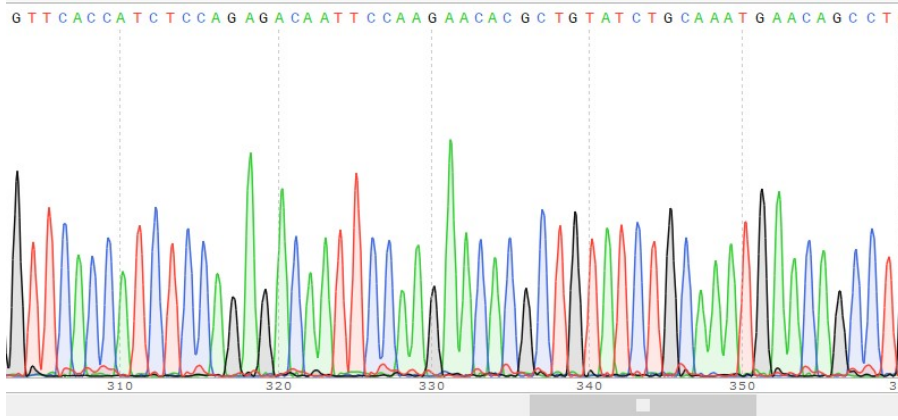
Метод Сэнгера

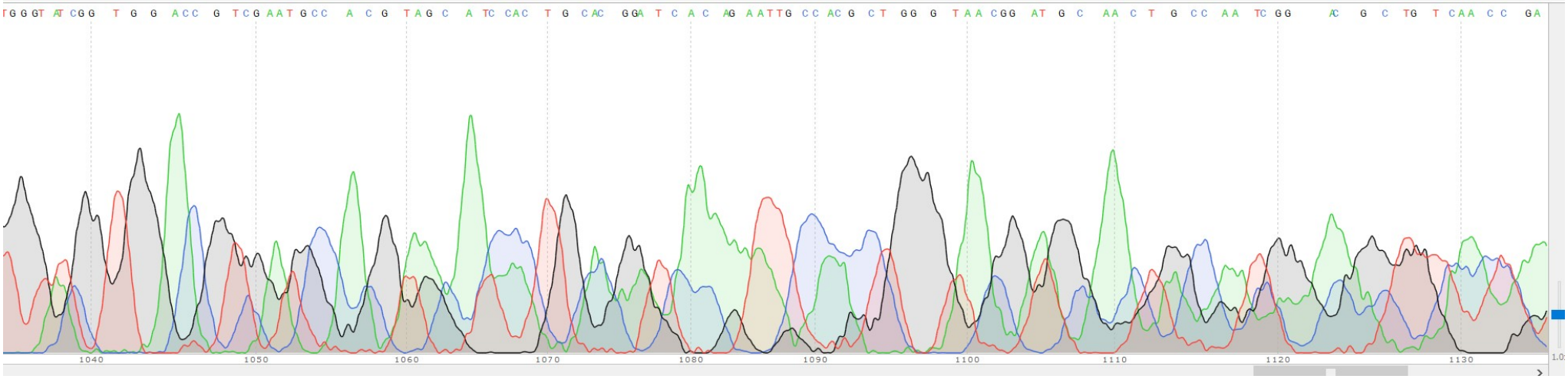




Метод Сэнгера







Методы второго поколения

Roche

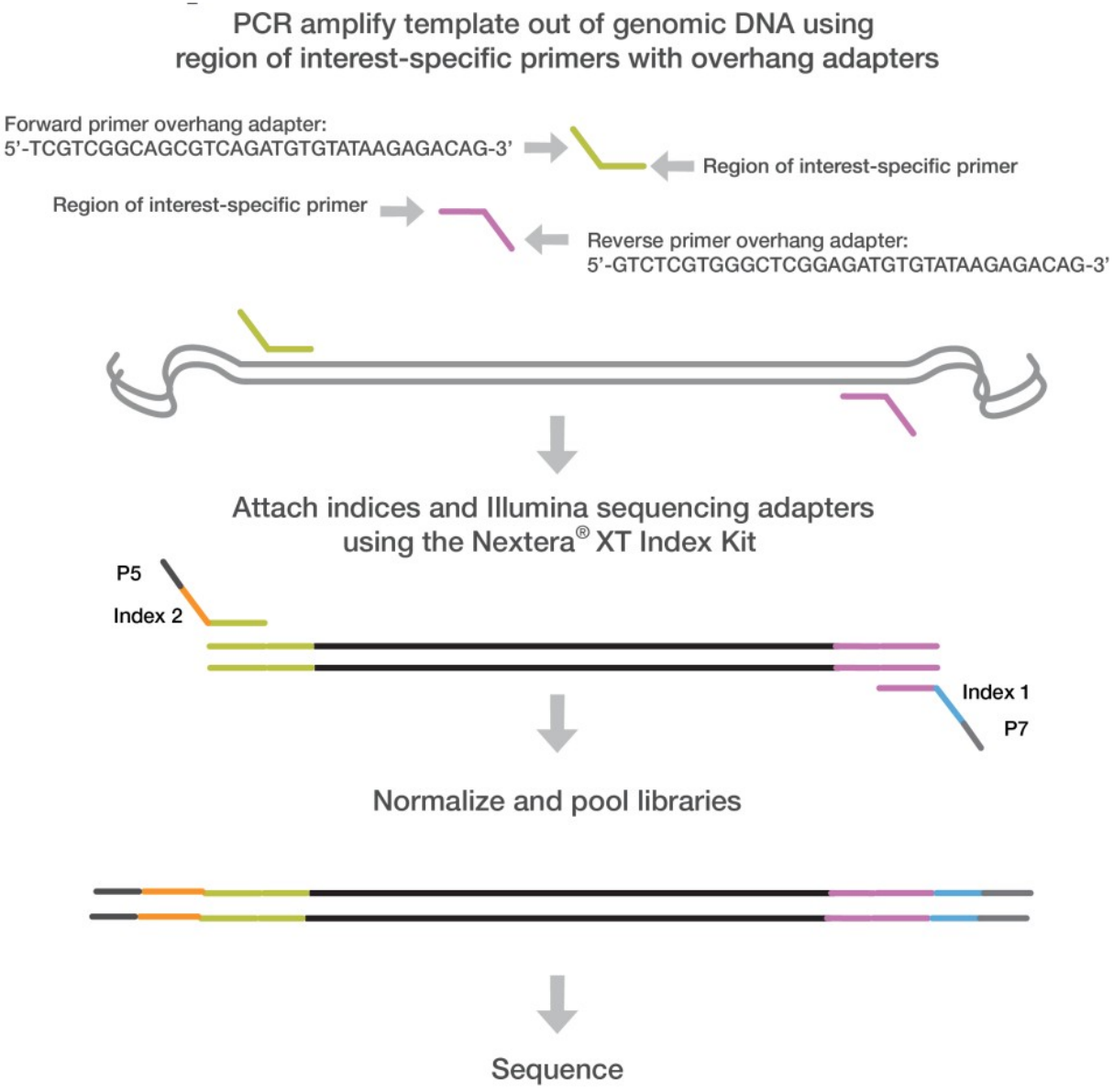
LifeTechnologies

IonSystem

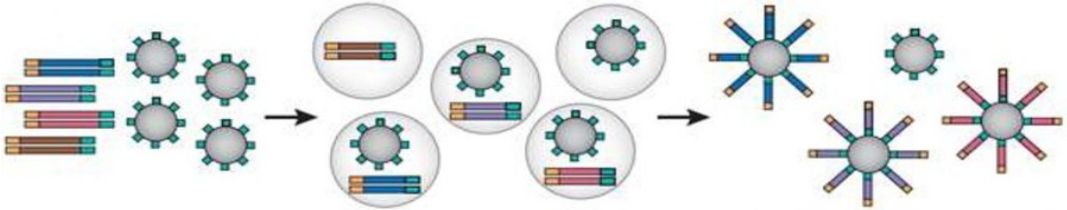
BGI

Illumina

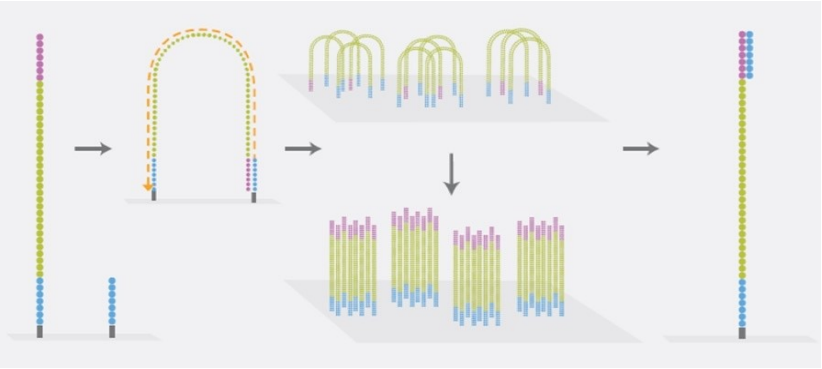
SOLEXA/ Illumina



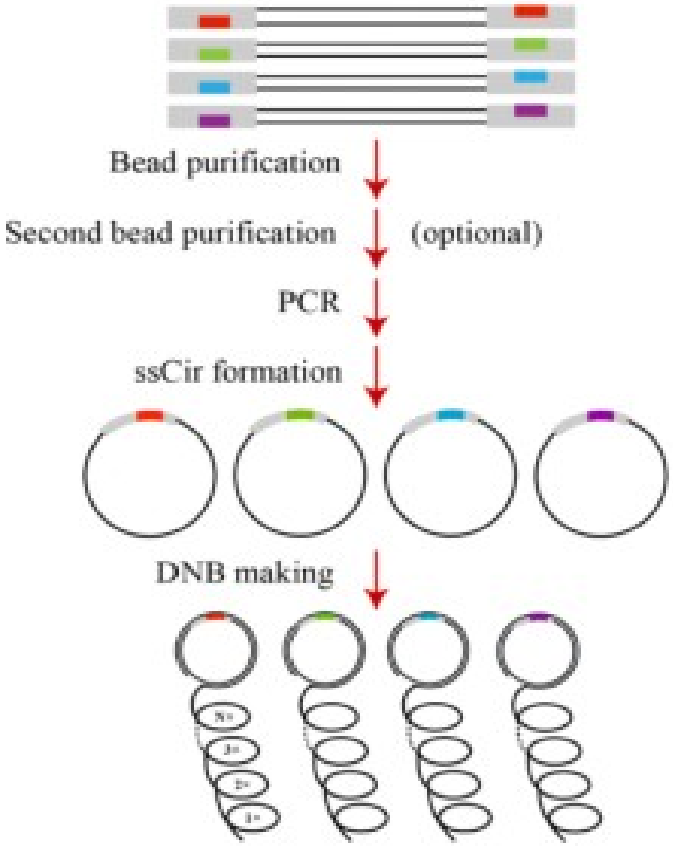
Методы амплификации сигнала



Эмульсионная ПЦР

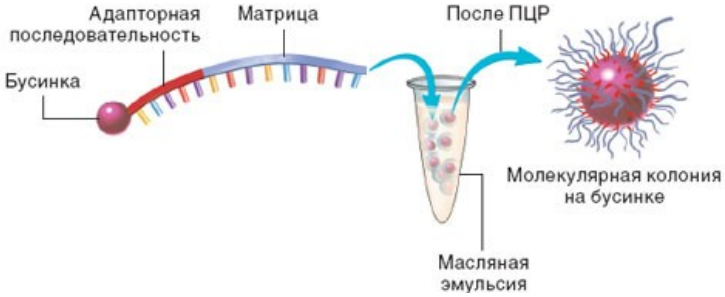


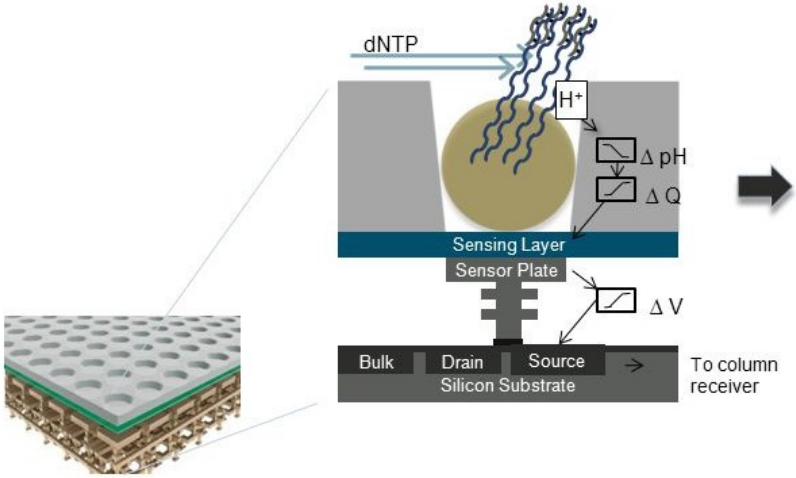
Мостиковая ПЦР

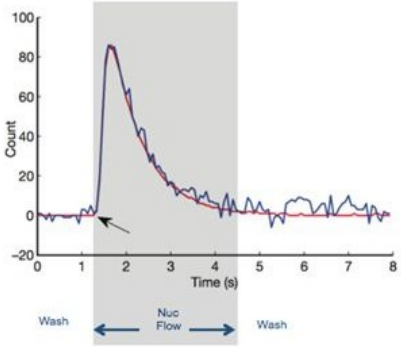


Rolling Circle Replication (DNA Nanoballs)

Полупроводниковое секвенирование







Rothberg J.M. *et al Nature doi:10.1038/nature10242*

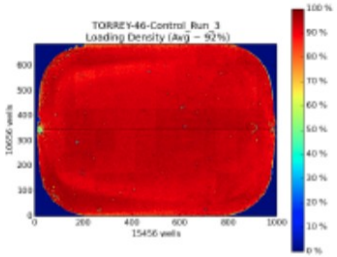
Полупроводниковое секвенирование

Run Summary

17.3 G Total Bases | **83** Key Signal

92% ISP Loading

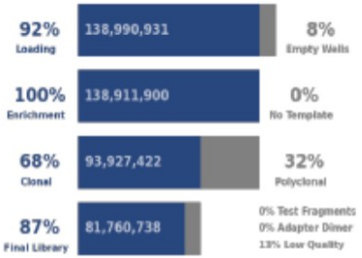
ISP Density



81,760,738 Total Reads

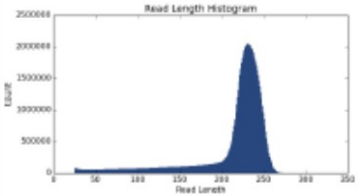
59% Usable Reads

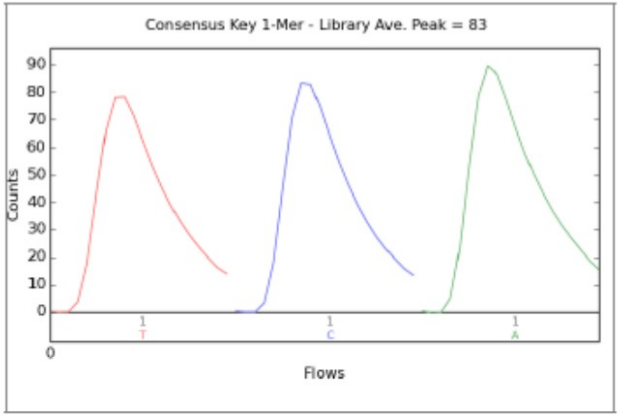
ISP Summary



211 bp Mean | **228 bp** Median | **230 bp** Mode

Read Length

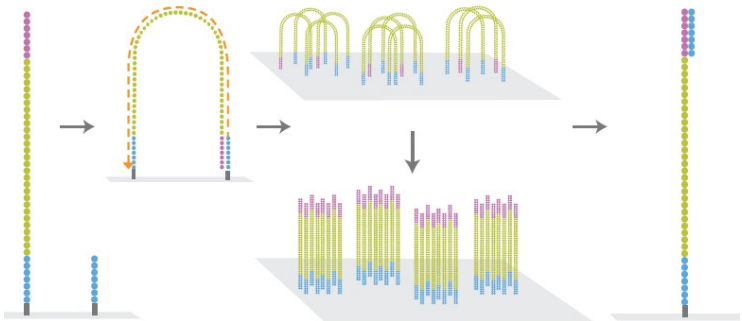


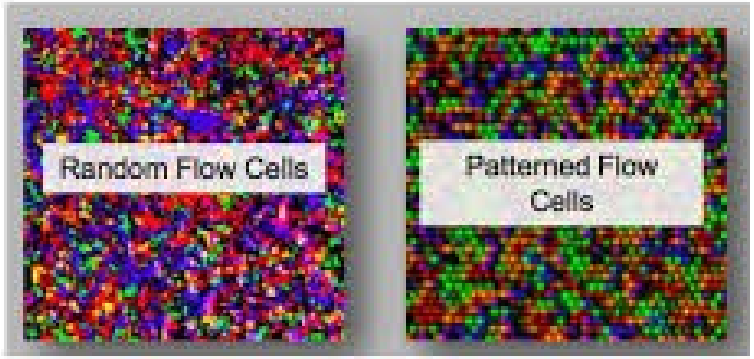


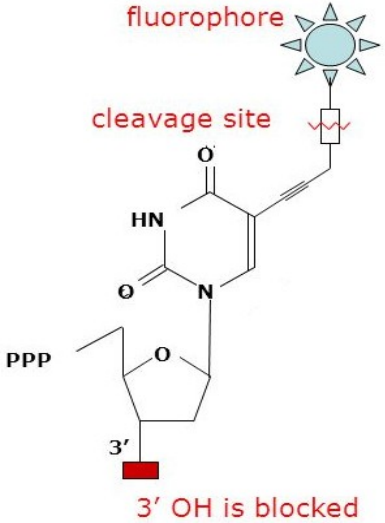
Addressable Wells	**151,539,288**	
With ISPs	138,990,931	91.7%
Live	138,911,900	99.9%
Test Fragment	459,014	00.3%
Library	138,452,886	99.7%

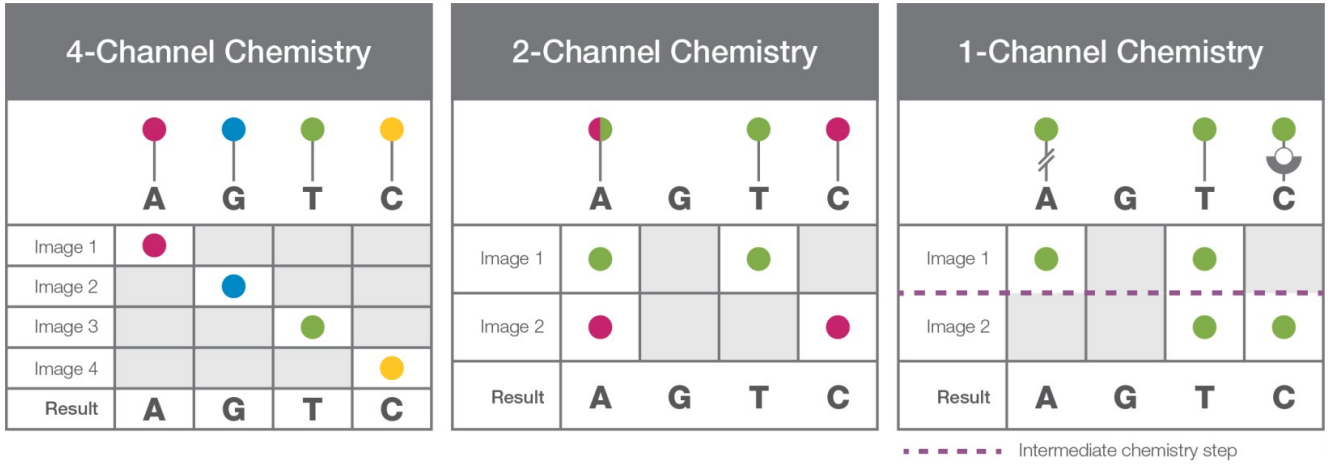
Library ISPs	**138,452,886**	
Filtered: Polyclonal	44,984,478	32.5%
Filtered: Low Quality	11,705,299	08.5%
Filtered: Adapter Dimer	2,371	00.0%
Final Library ISPs	**81,760,738**	**59.1%**

SOLEXA/Illumina

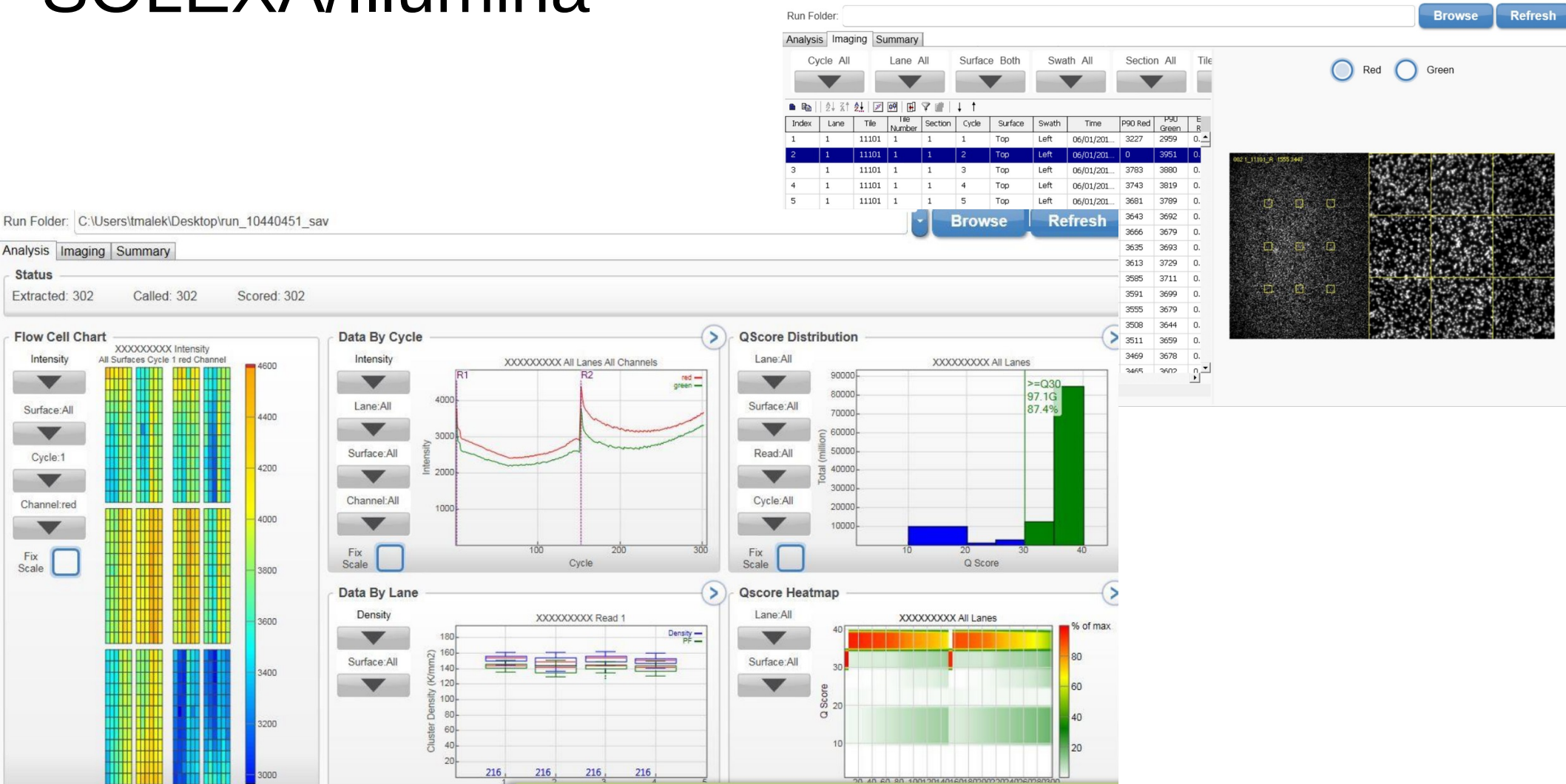




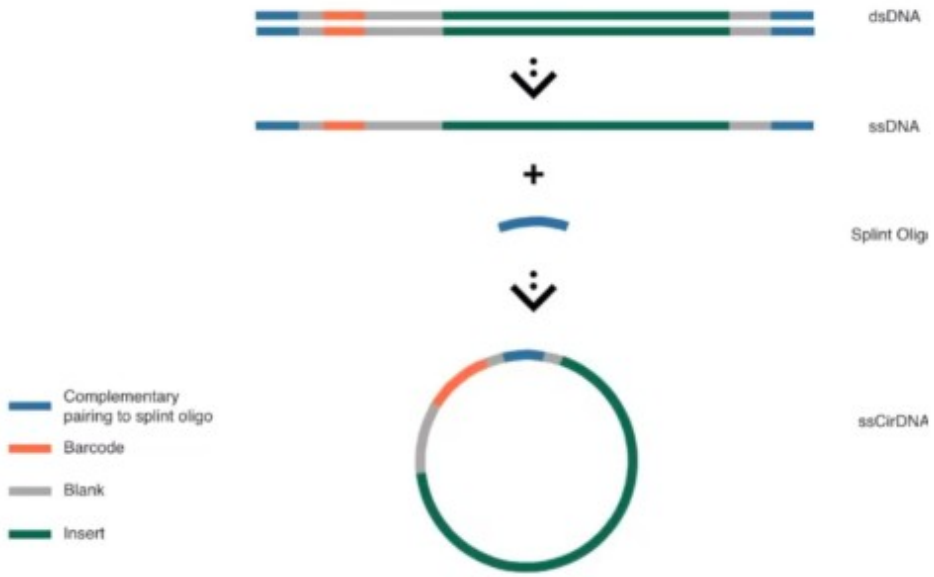


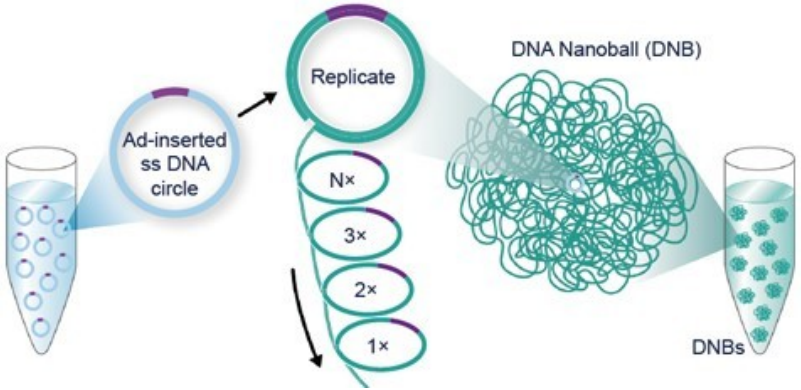


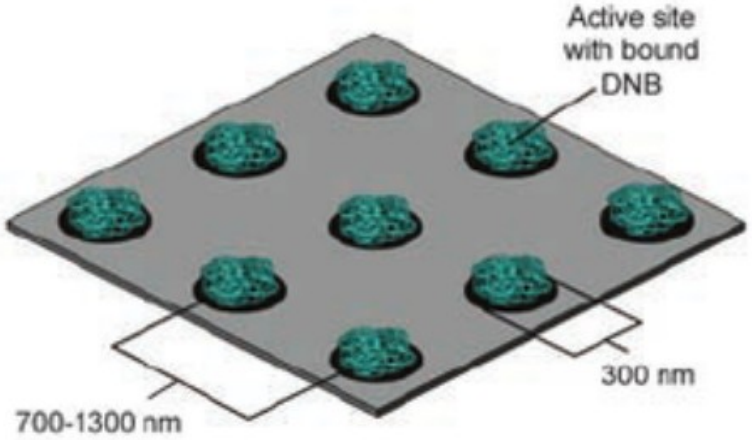
SOLEXA/Illumina



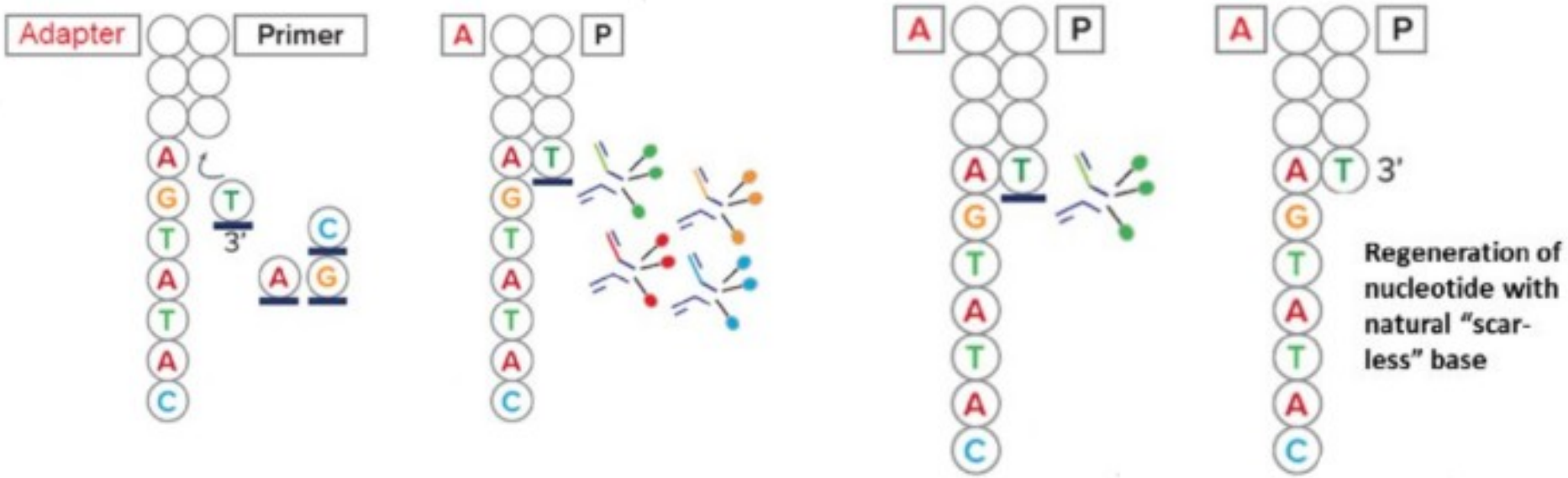
MGI
(Beijing Genomics Institute)



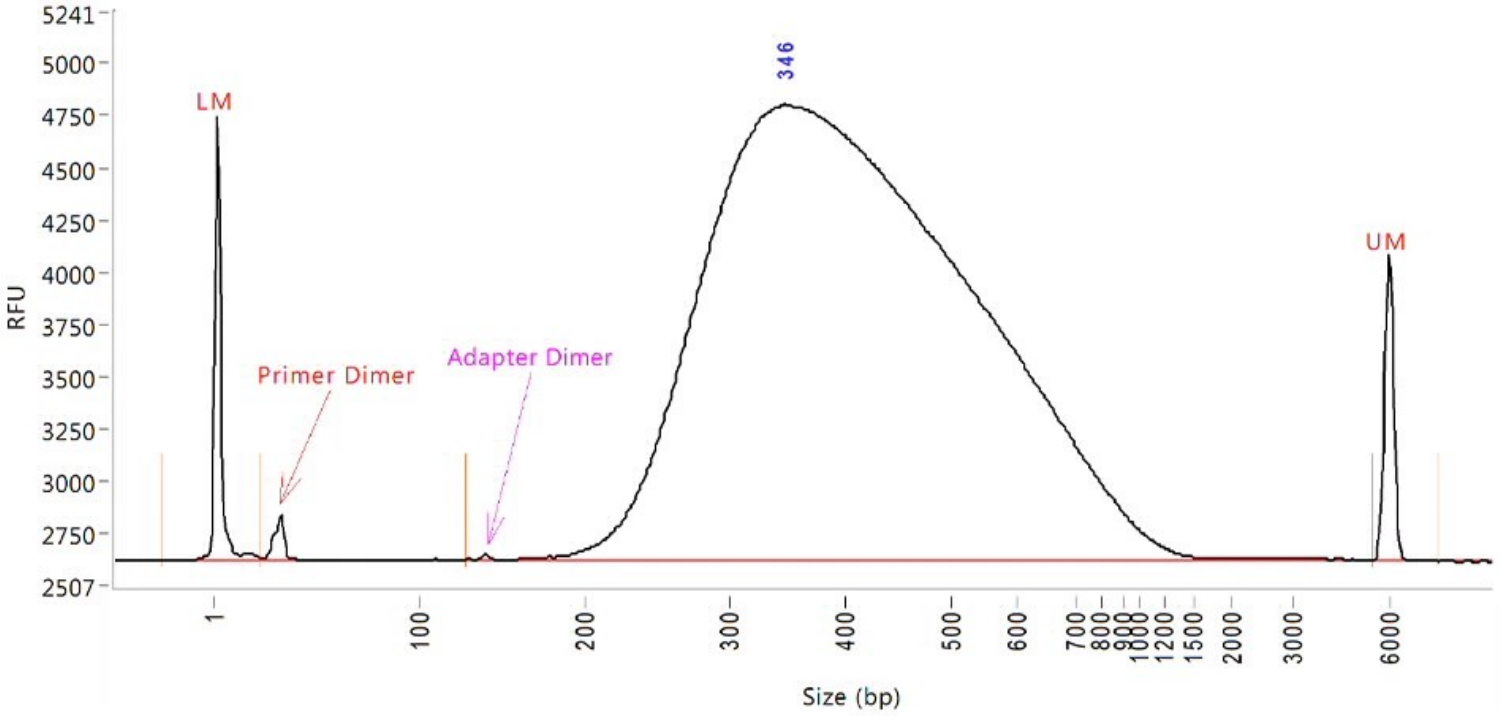




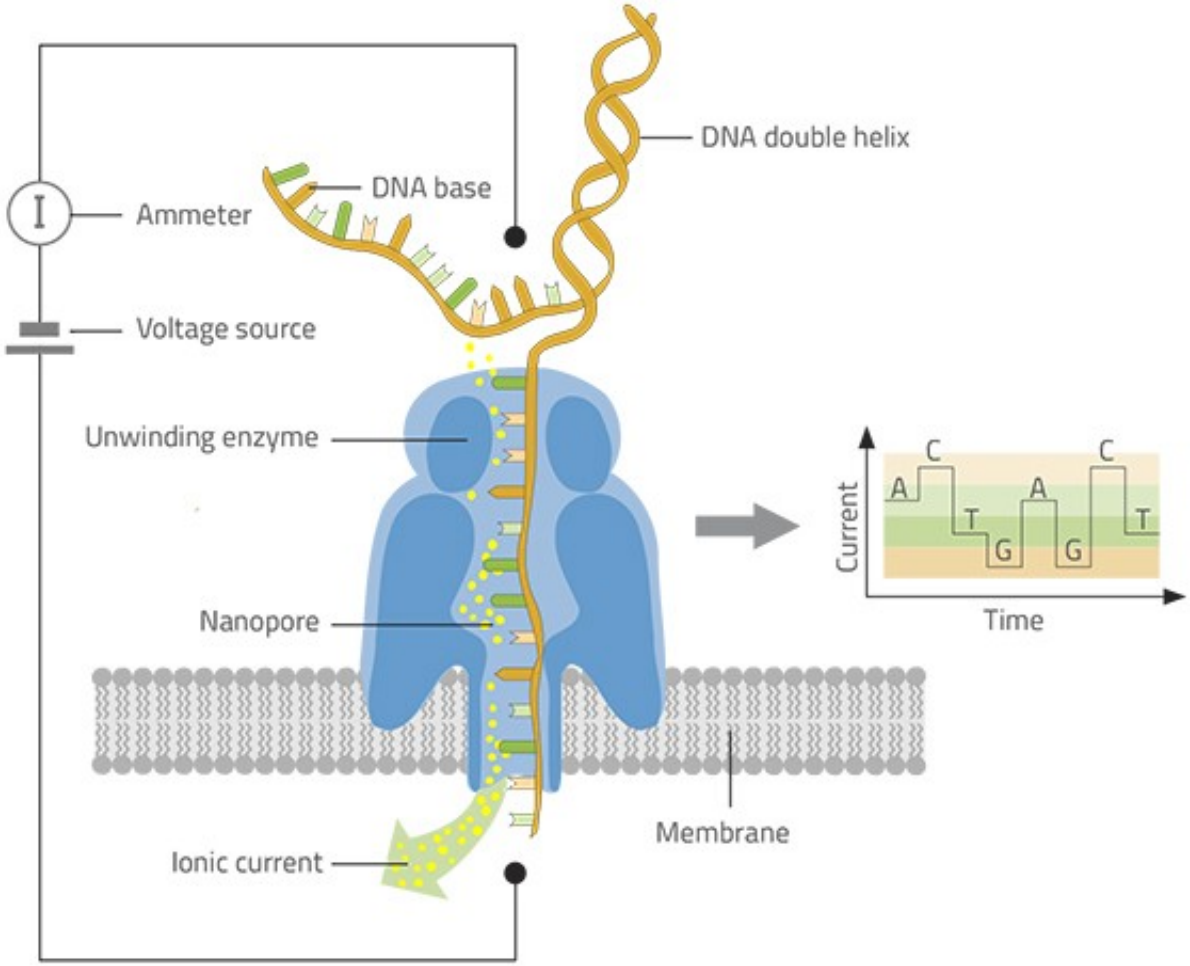
MGI
(Beijing Genomics Institute)



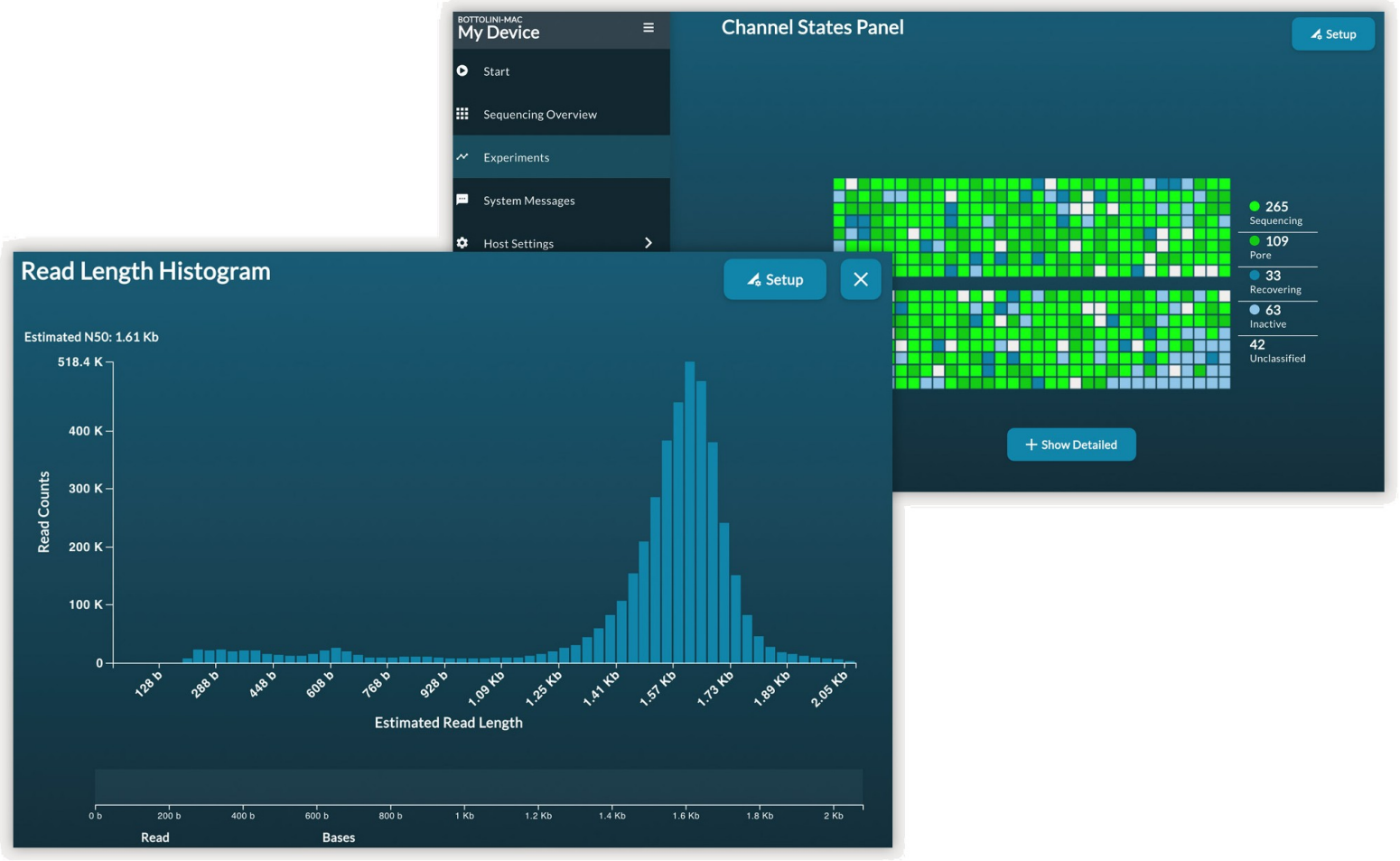
Контроль качества библиотек



Oxford Nanopore Technologies



Oxford Nanopore Technologies



Как выглядят данные

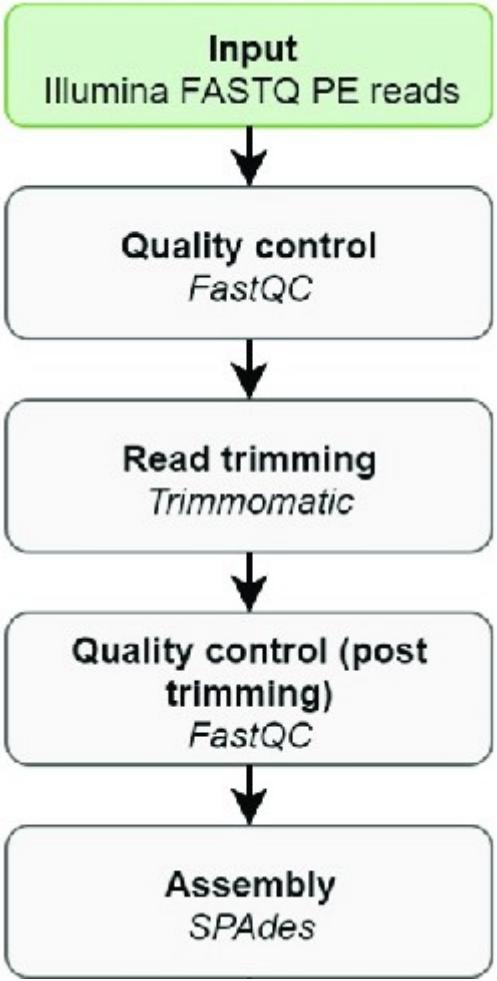
FASTQ file:

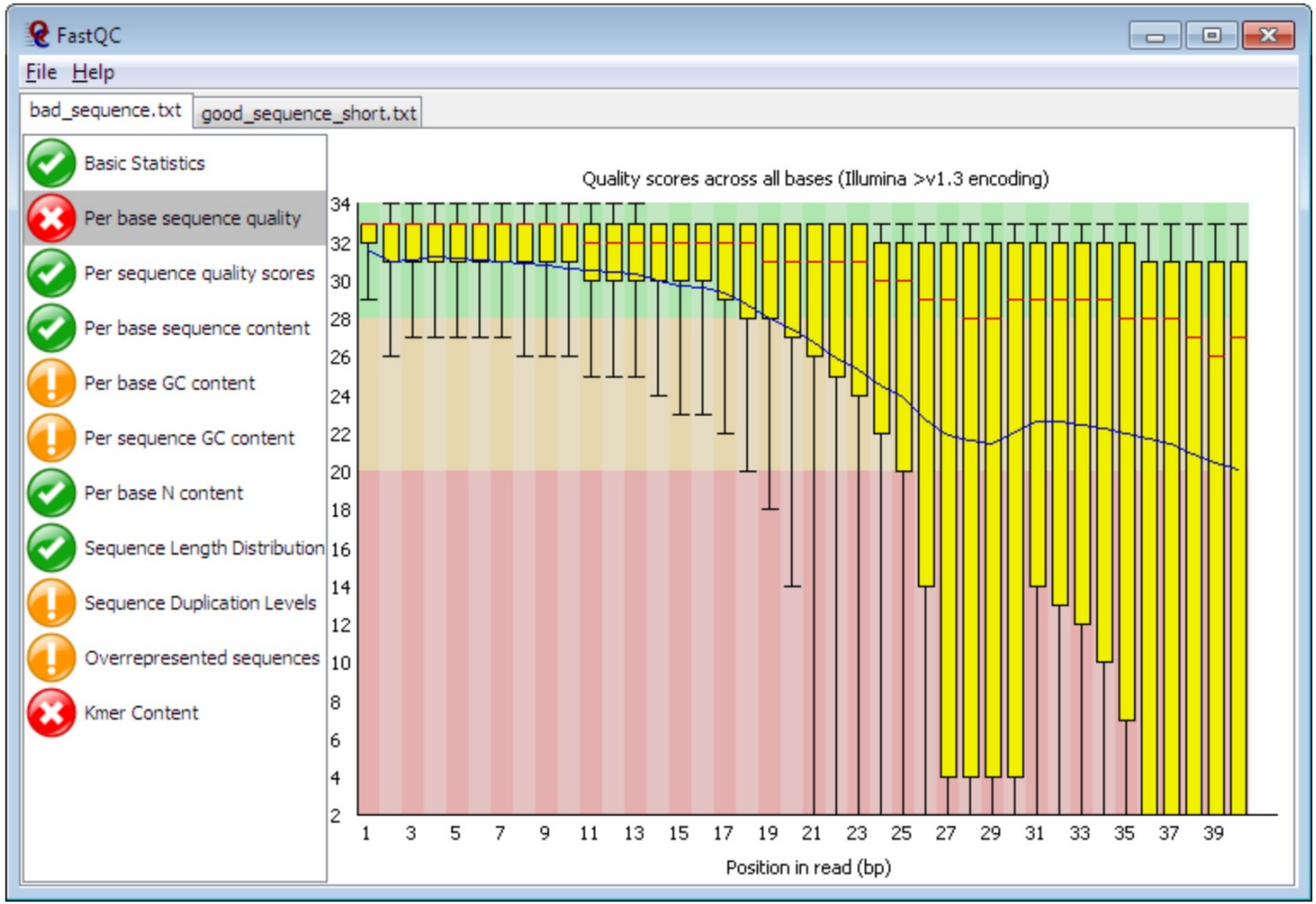
@M02459:189:000000000-CMKP4:1:1101:9104:1573 1:N:0:9
CAGCGGGAAGACCTTGAACCGCTAGGCGAGGGGCTCAAATAAACCCTTTCTTCCTCCTCCTTCCTCTTCTTCTTCTCTTCTCT
CTTCTTCCTCTCCTTCTCCCTTCCTCTCTCCTCTCTCTTTTTTTTTTTTCTTTTCCTTCCTCTTTCCTCTTCTCTCCTTTCCCTTTT
CTTCCTTTCCCTTTCCTTCTACTTTTCCTTTTTCTTCTCTTTTTCTTTCTTCTCTTTTTTCCCTCTTTCCCTCCTTTTTCTATTCCTT
CCTTTCCCCCTCCCTCCTCCCCTCCTCCTCCCCTTTCCTCTTCTTTTCTTTTC
+
--8AC7CDGG ?FGFGCA@FF:@@7F8F7FF7@+@@C,C9,,,,6966,,666,,6,6,,6,6,,,6,6,,5,,55,,4,,,,,5,,5++44++
+,,5,,,3,3,,,,3,,,,,3,,,,,,,,335,,,,3,4,,,3,,,,,,33,,,,,,,,,,2,2,2,,,2,2,,+,+++,5,**111++4++++++1/))**+)*0+1+*+.)0+)**/.+)++).))()+/)))((()(,,((((,
((.,((((((,(,(()))-)))))))))).)))
,,4:,,,*1*)2)))10<*0/2000*****0*1**1*0*17)0)75).//*)))).*)**((-(-,*(.)..-)*).--)(,(((,-(()),)))-).)-)-.),).64),(()))

FASTA file:

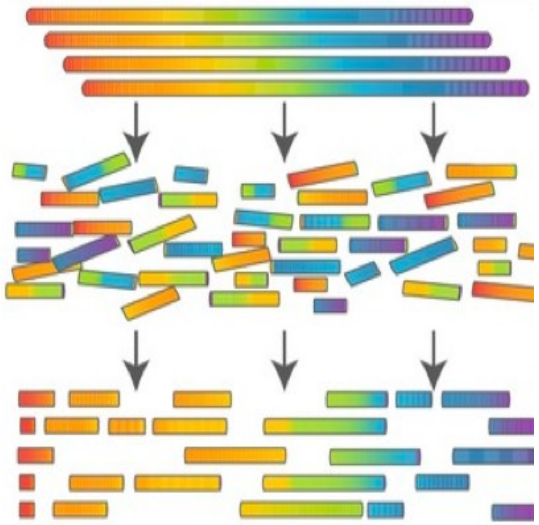
>M02165:19:000000000-AJF39:1:1101:14444:2258 2:N:0:16
GACTACAGGGGTATCTAATCCTGTTTGCTCCCCACGCTTTCGAGCCTCAGCGTCAGTTACAAGCCAGAGAGCCGCTTTCGCC
ACCGGTGTTCCTCCATATATCTACGCATTTCACCGCTACACATGGAATTCCACTCTCCCCTCTTGCA
>M02165:19:000000000-AJF39:1:1101:20355:2309 2:N:0:16
GACTACACGGGTATCTAATCCTGTTCGCTCCCCTGGCTTTCGCTCCTCAGCGTCAGTTTTCGTCCAGCAAGCCGCCTTCGCCA
CTGGTGTTCTTCCTAATCTCTACGCATTTCACCGCTACACCTGGAATTCCACCCTCCTCTCCGATA
…..

Анализ данных

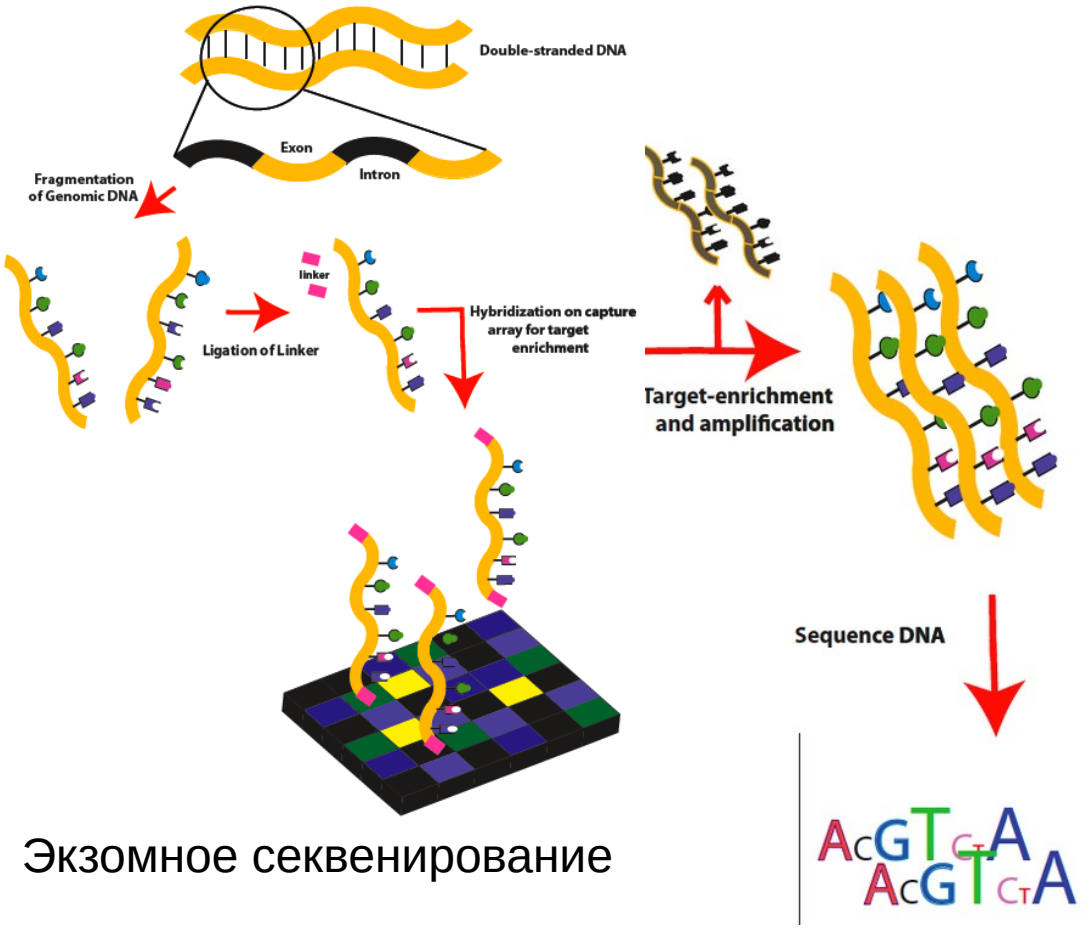




Типы библиотек

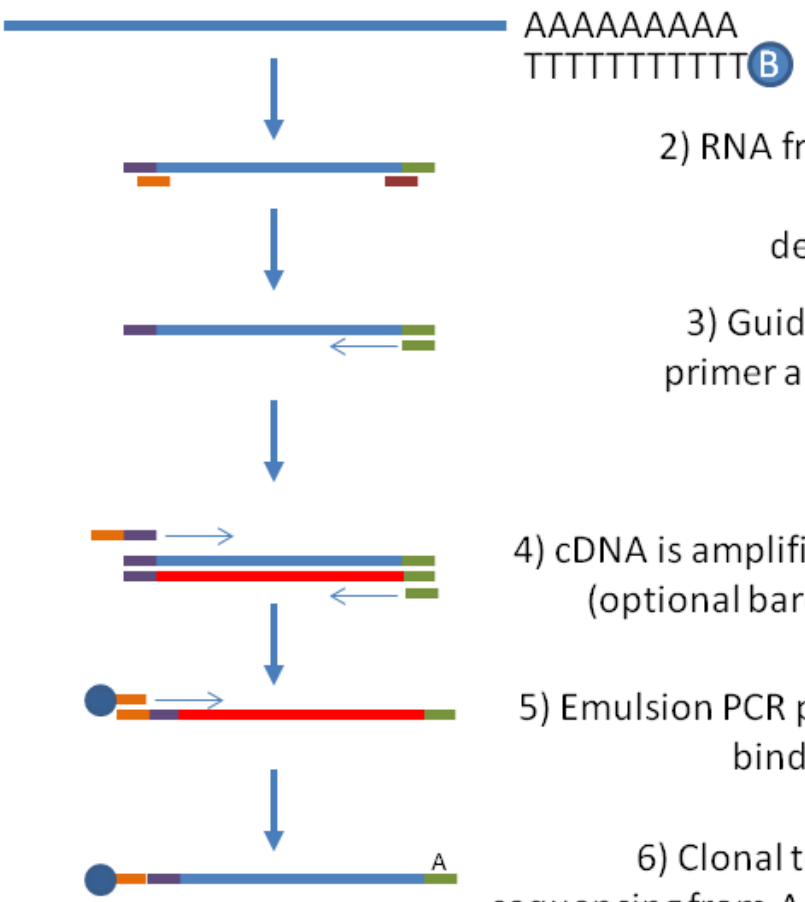


Фрагментные библиотеки



Экзомное секвенирование

Типы библиотек - транскриптом



1) PolyA+ RNA captured

2) RNA fragmented and 5’ and 3’ adapters ligated using degenerate guide adapter

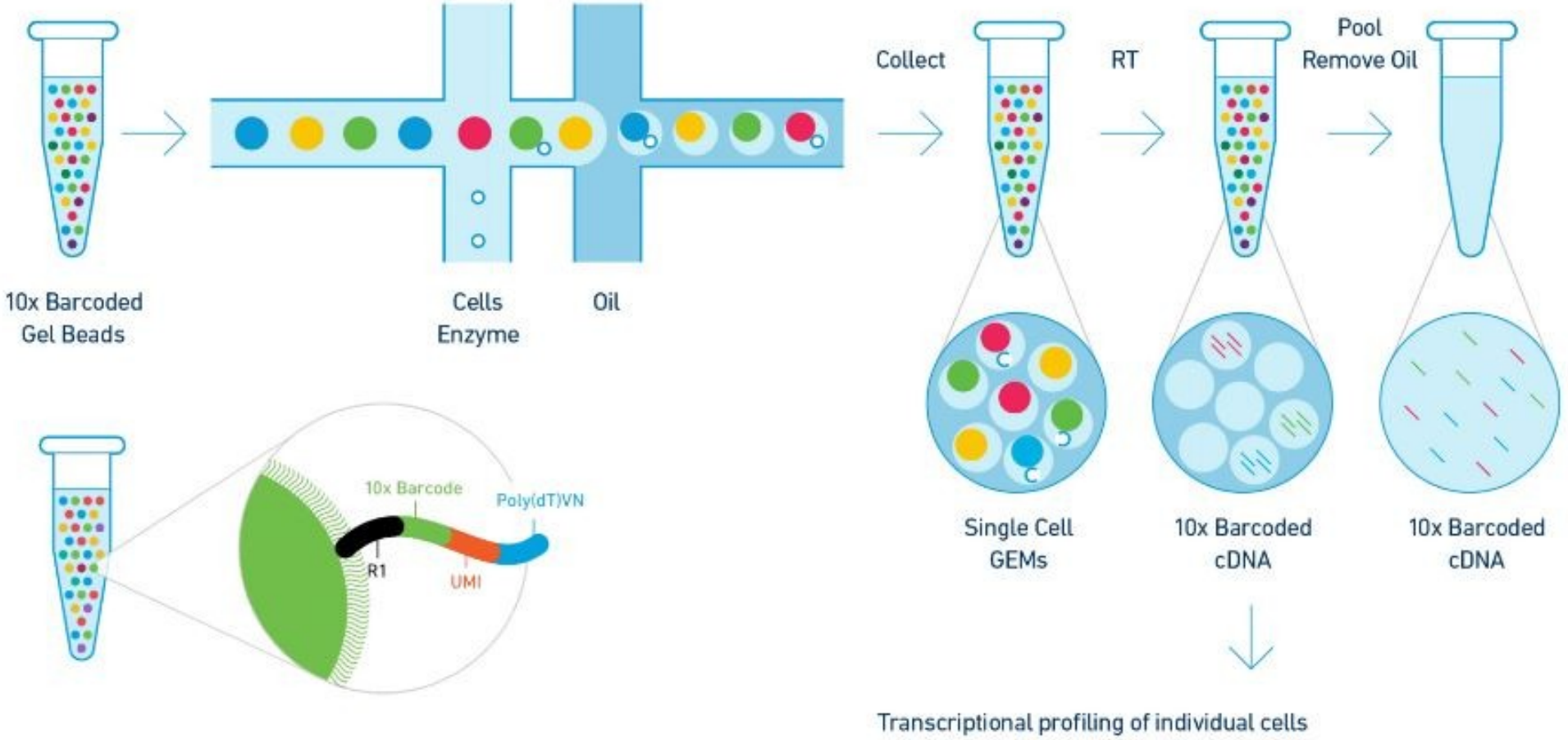
3) Guide adapters removed, RT primer annealed and first strand cDNA synthesized (red)

4) cDNA is amplified by PCR and adapters (optional barcode in green 3’ primer)

5) Emulsion PCR performed to covalently bind fragment to microbead

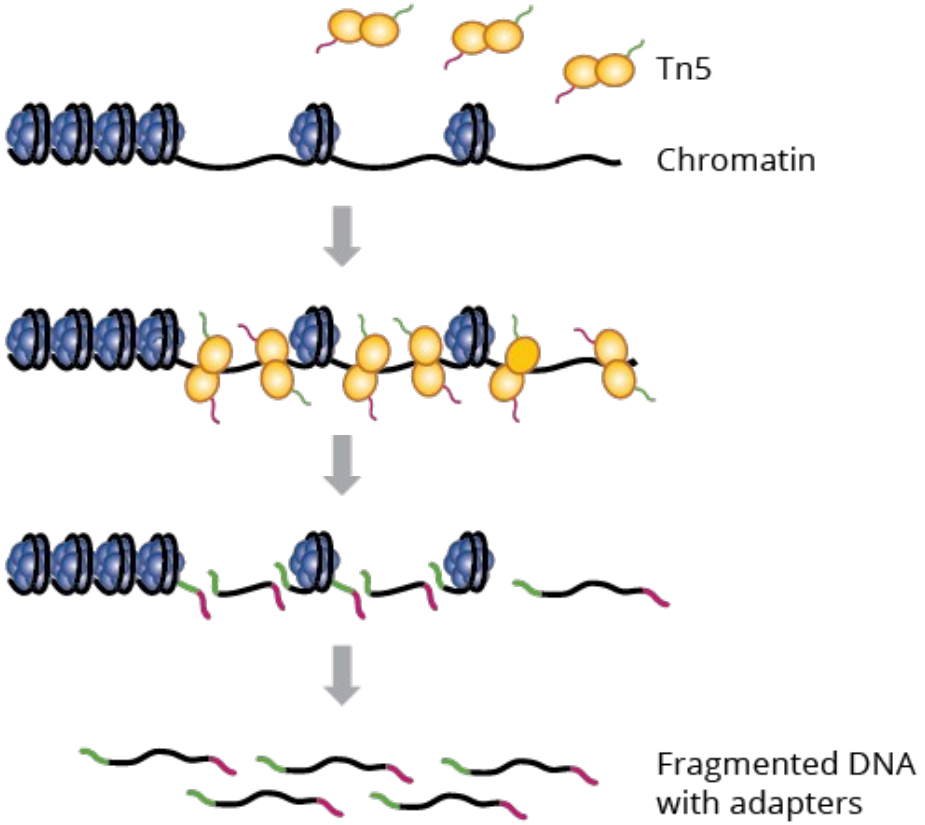
6) Clonal template cDNA ready for sequencing from A towards the microbead

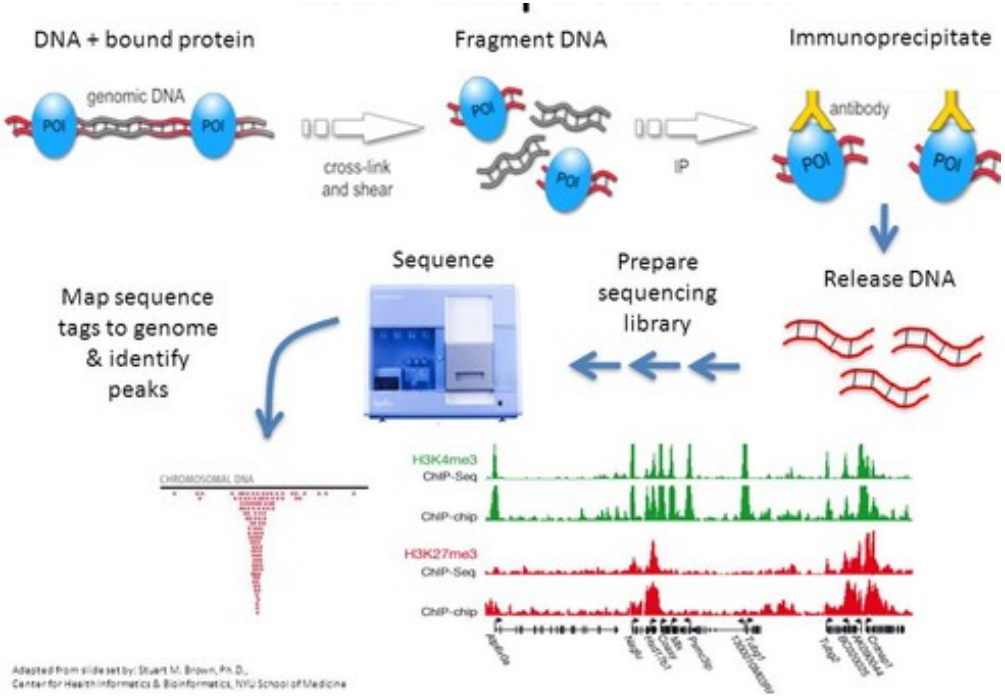
Типы библиотек - Single cell 10x



Методы исследования хроматина

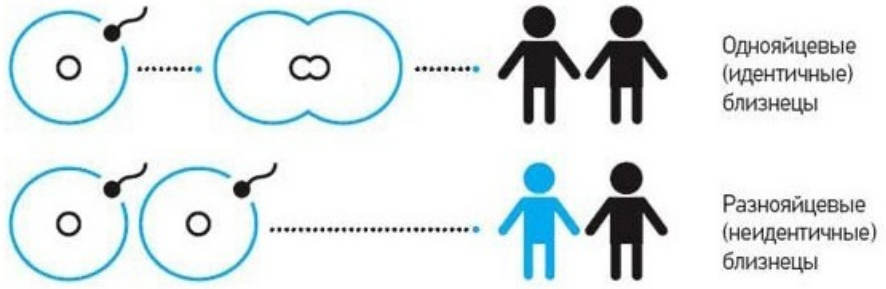
ATAC-seq (Assay for Transposase-Accessible Chromatin using sequencing) - метод исследования доступности хроматина.

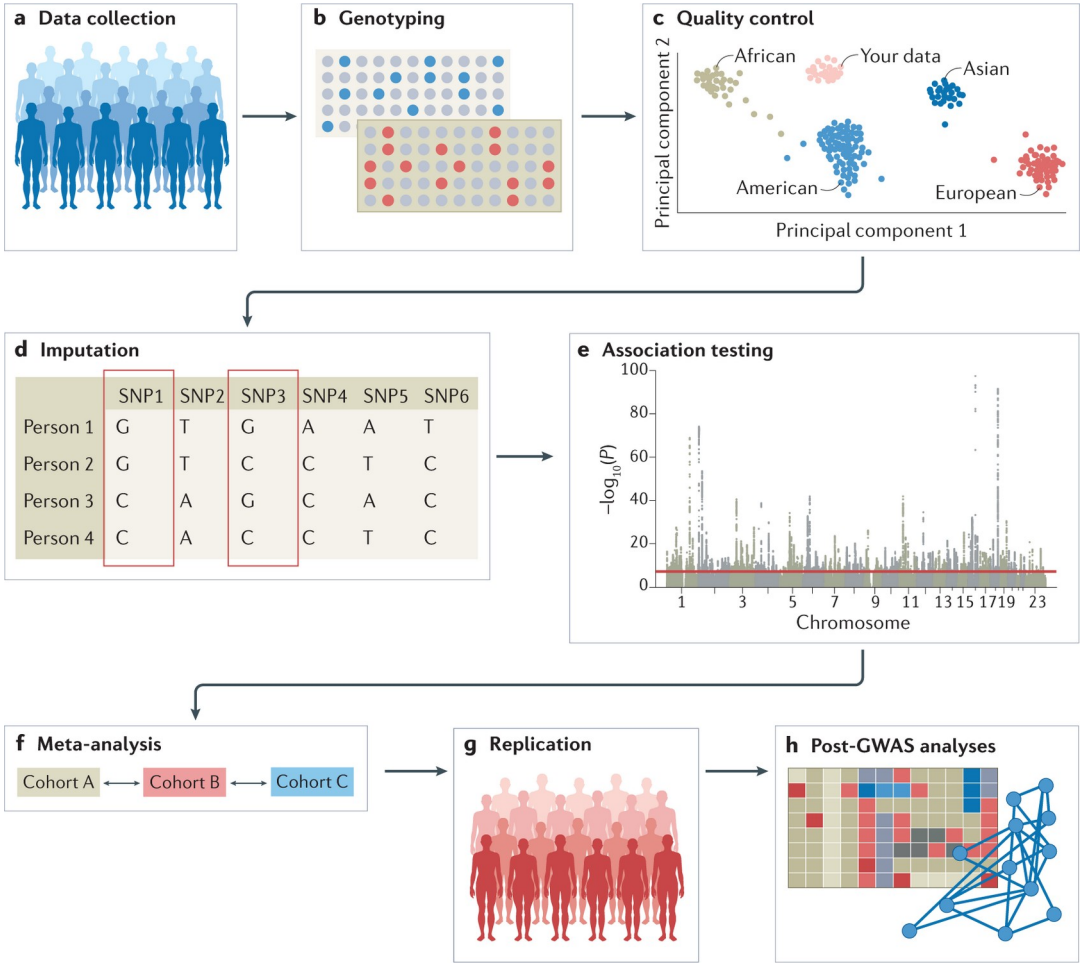




ChIP-seq — метод анализа ДНК-белковых взаимодействий, основанный на иммунопреципитации хроматина (ChIP)

Полногеномный поиск ассоциаций





Обзор GWAS: https://www.nature.com/articles/s43586-021-00056-9

Изучение системного склероза с применением GWAS: https://www.nature.com/articles/s41467-019-12760-y