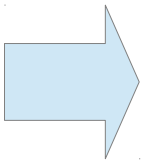
Биоинформатика для высокопроизводительного секвенирования

Павел Мазин
iaa.aka@gmail.com



Все материалы курса лежат тут – shorturl.at/hqK17 (google drive)

ВШЭ-2019/20

Доступ на вычислительный кластер

Для выполнения части ДЗ будет нужен вычислительный кластер. Для доступа к нему вам необходимо вписать свою почту и желаемый логин в табличку «Список студентов» - она лежит в папке google drive с матерьялами курса

Оценка

Вычисляется как средняя оценка за все ДЗ

План курса

когда	что	кто
09/09	Методы массового секвенирования	П. Мазин
16/09	Транскриптомика: введение	П. Мазин
23/09	Транскриптомика: сборка транскриптов, подсчет каунтов, нормализация, проверка самосогласованности	П. Мазин
30/09	Транскриптомика: Линейные модели, дифференциальная экспрессия, альтернативный сплайсинг	П. Мазин
7/10	Single-cell transcriptomics	И. Курочкин
14/10	RiboSeq	И. Кулаковский
28/10	Эпигенетика	А. Миронов
11/11	ChipSeq	И. Кулаковский
18/11	Метилирование ДНК, открытый хроматин, модификации гистонов	А. Кононкова
25/11	3D-структура хроматина	А. Кононкова
2/12	de novo сборка геномов и транскриптомов	А. Касьянов
9/12	Экзомное секвенирование с фокусом на биомедицинские исследования	Е. Набиева
16/12	Метагеномика	П. Шелякин

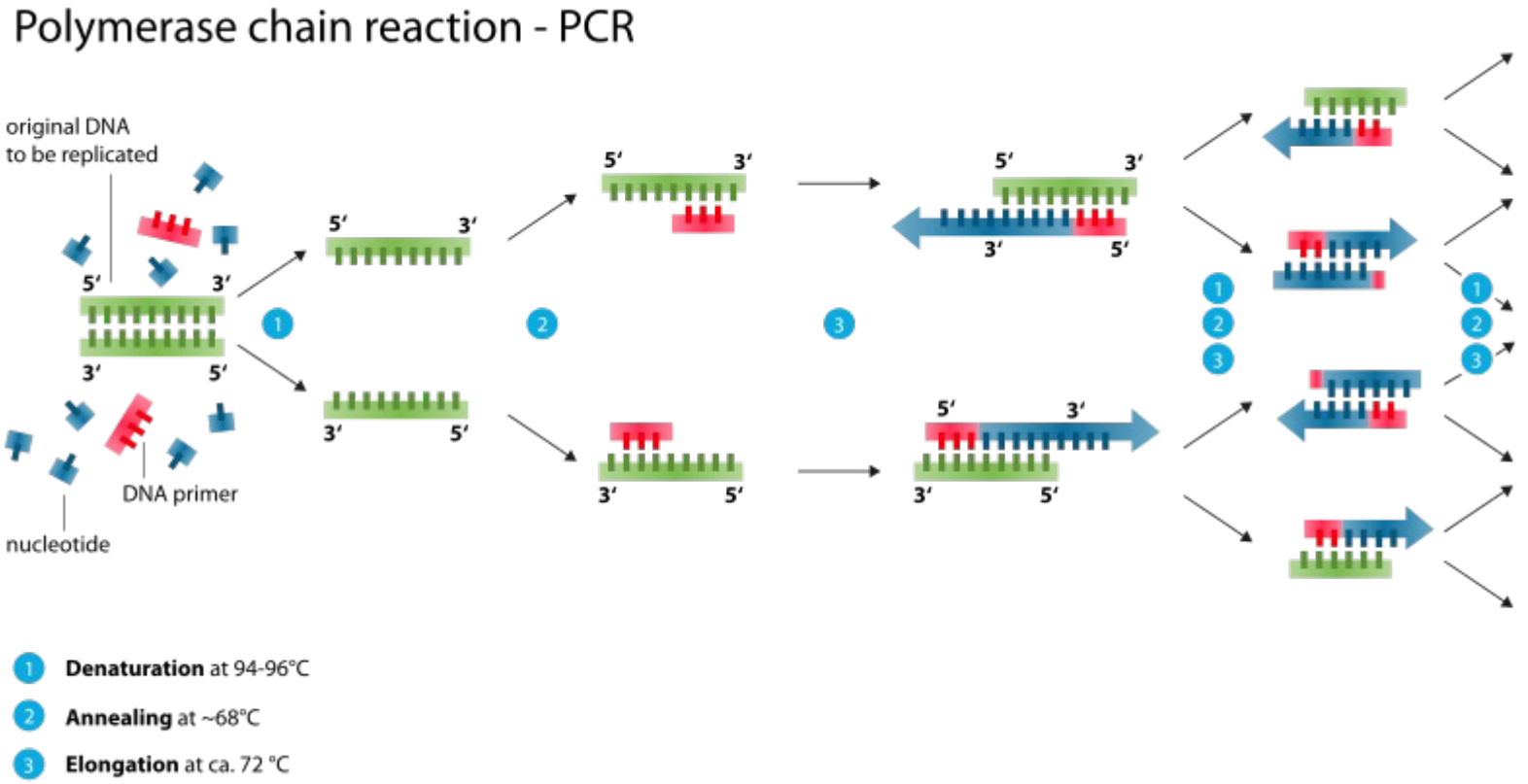
First generation
- automated Sanger method

Next generation
- Roche/454
- IonTorrent
- Illumina/Solexa
- Life/APG - SOLiD

Last-generation
- Pacific Biosciences
- Oxford Nanopore Technologies

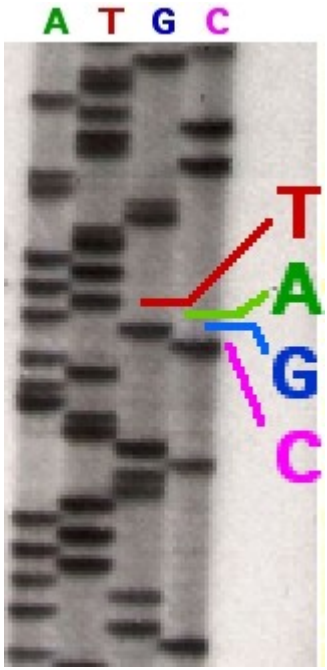
Most figures were taken from Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet. 2010 Jan;11(1):31-46. doi: 10.1038/nrg2626. Epub 2009 Dec 8.

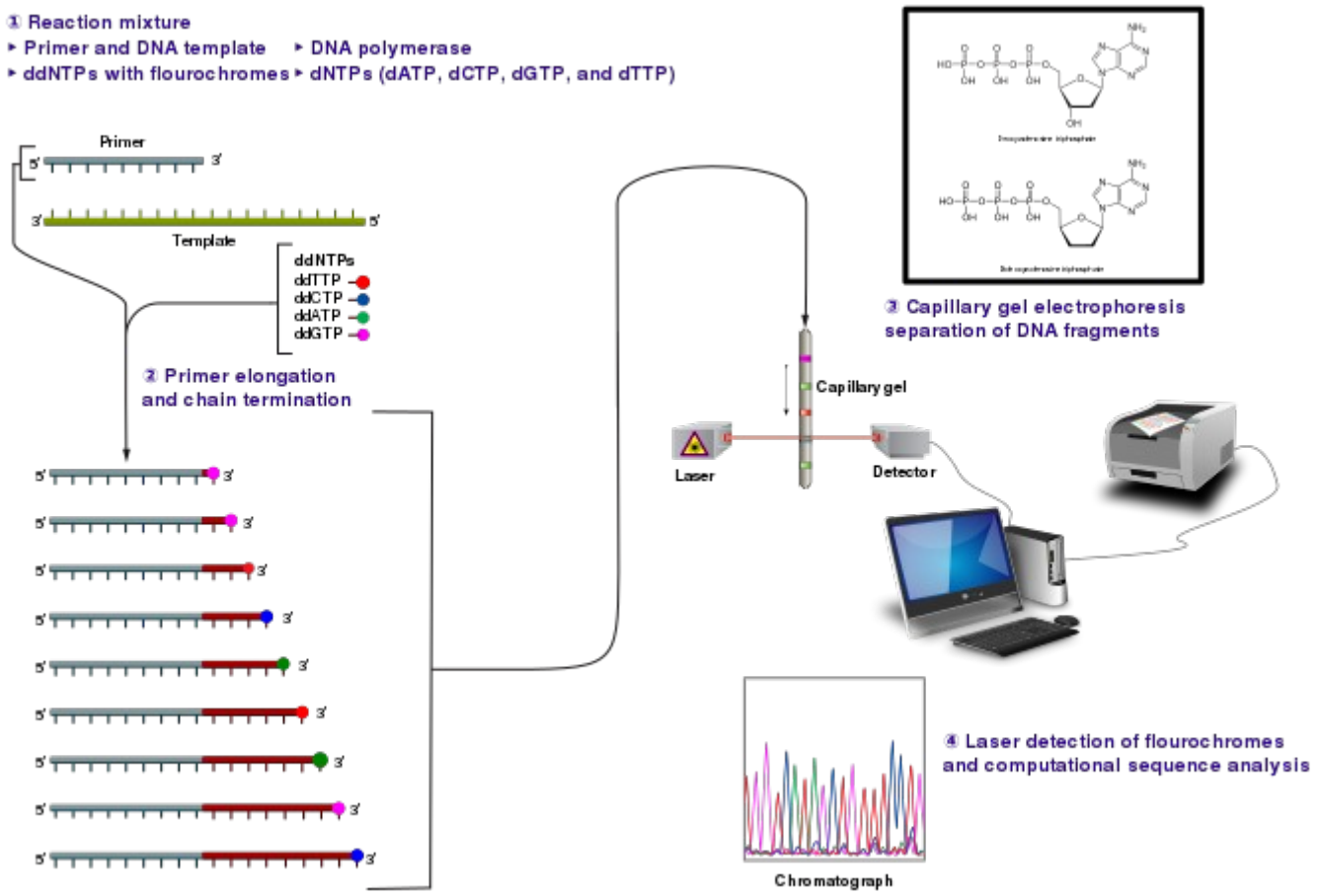
Polymerase chain reaction (PCR)



wikipedia.org

Sanger sequencing





wikipedia.org

Methods

Template preparation:

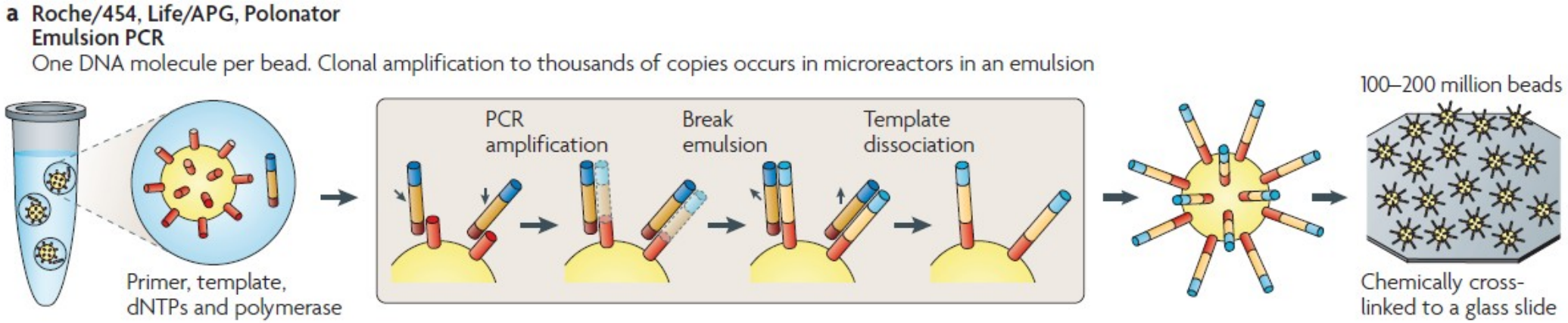
- Clonally amplified templates
  - Emulsion PCR
  - Solid phase PCR
- Single DNAmolecule templates
- Direct RNA-sequencing (Oxford Nanopore)
- Barcodes, unique molecular identifiers

Sequencing

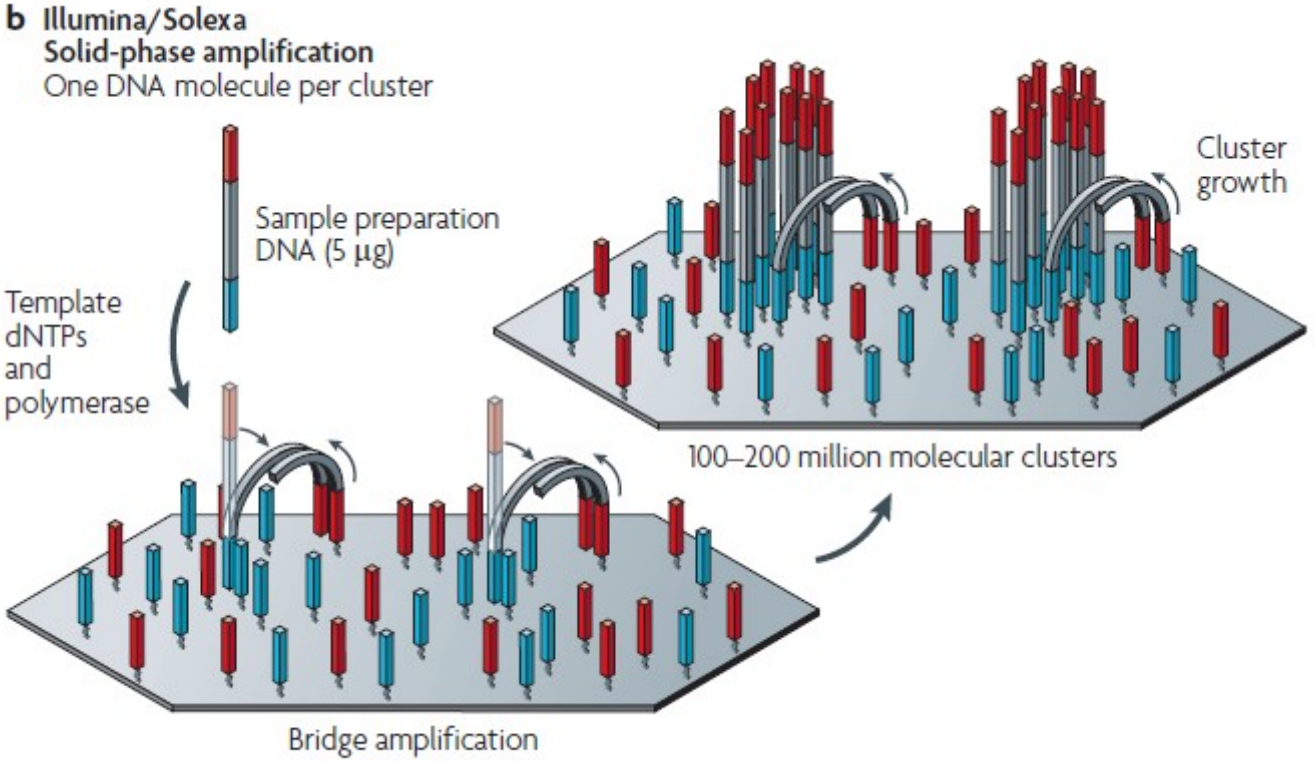
- Cyclic reversible termination
- Single-nucleotide addition
- Sequencing by ligation
- Real-time sequencing

Template preparation. With amplification

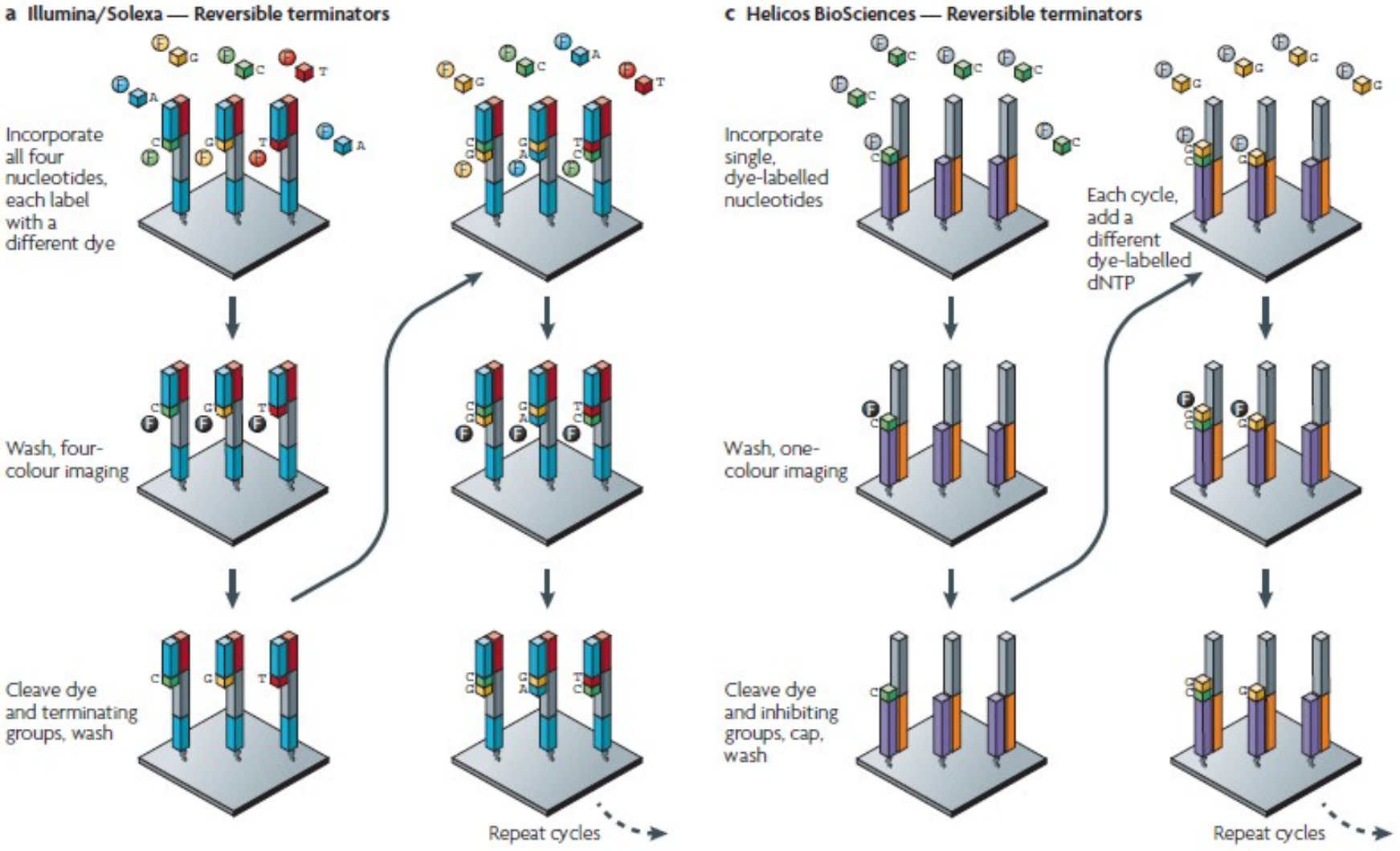
emulsion PCR (emPCR), ~1 million reads



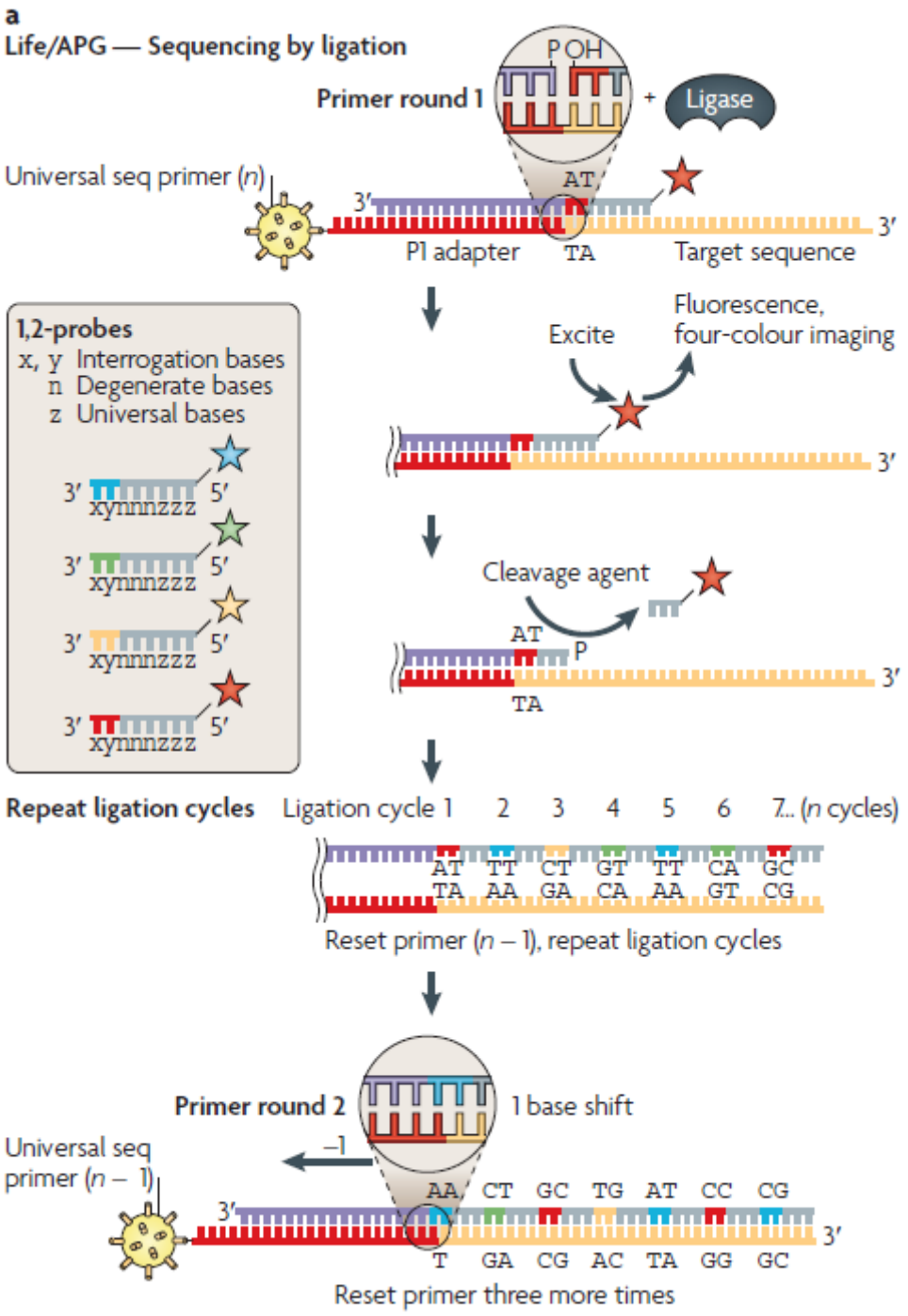
solid-phase amplification, ~ 100 million reads

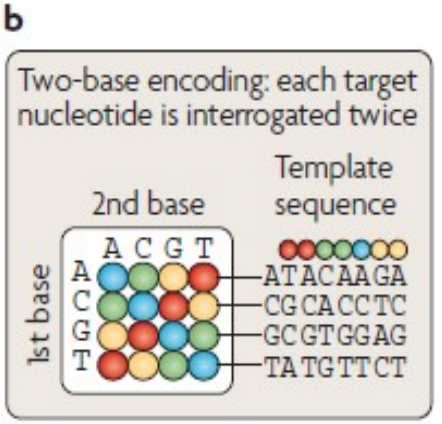


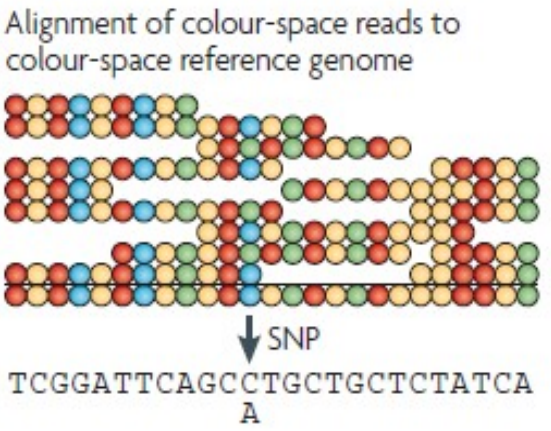
by cyclic reversible termination (Illumina, BioScience)



Sequencing by ligation (SOLiD)







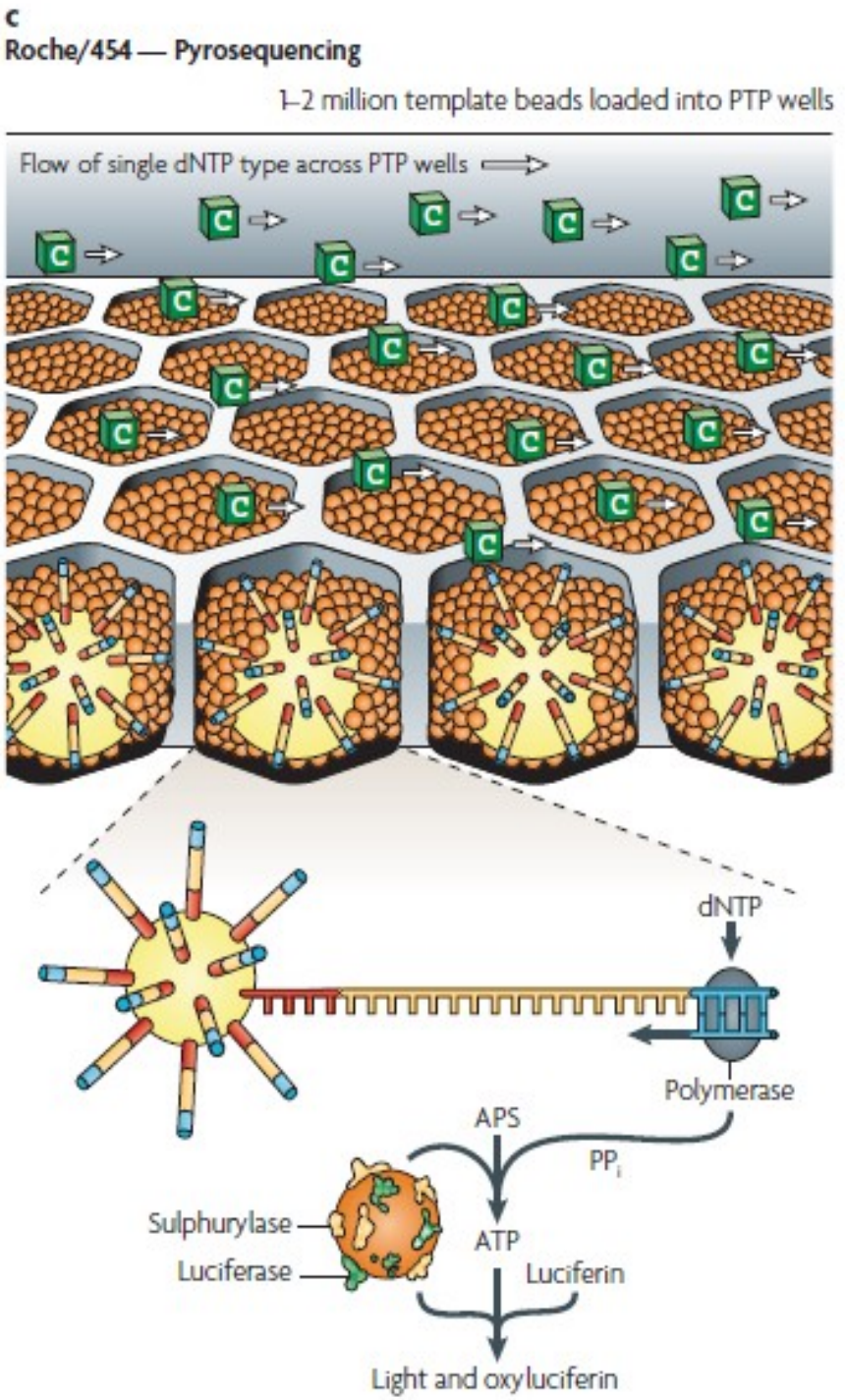
8nt probes are hybridized and ligated:

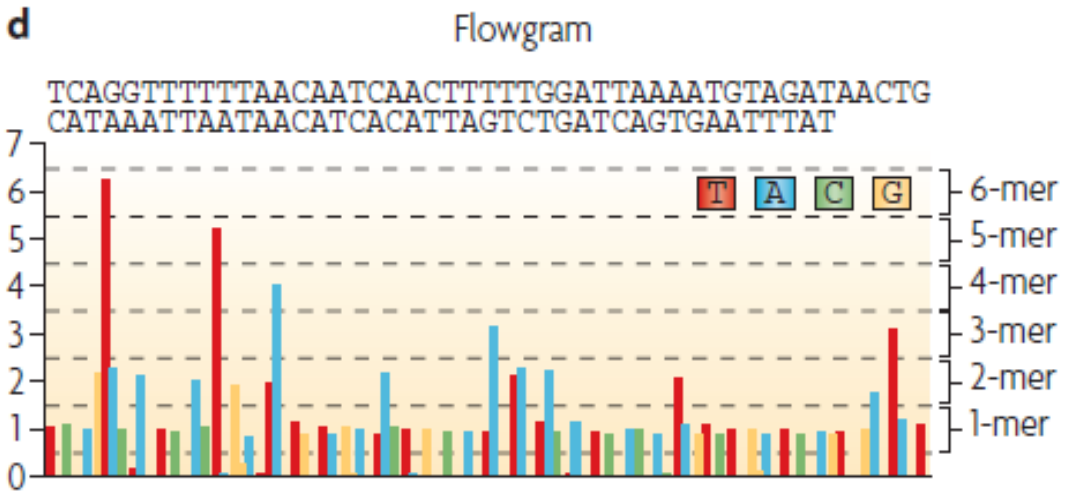
3-8nt are inosine.

5-6 is cleavage site (phosphorothiolate bound is cleaved by silver ions)

- 5 rounds (start position is shifted by 1nt each round) with 10 ligation cycles are performed.
- Each nt is interrogated twice
- Result is read in color space: sequence of four colors. Color denotes set of nt pairs (overlapping).
- To decode it to nt sequence one need to know at least one nt. But it could be mapped.

Single-nucleotide addition: pyrosequencing (454)





- Nts aren't modified
- Washing step is absent (Unincorporated nucleotides and ATP are degraded by the apyrase)
- Number of nt in homopolymeric repeats estimated by peak height

IonTorrent

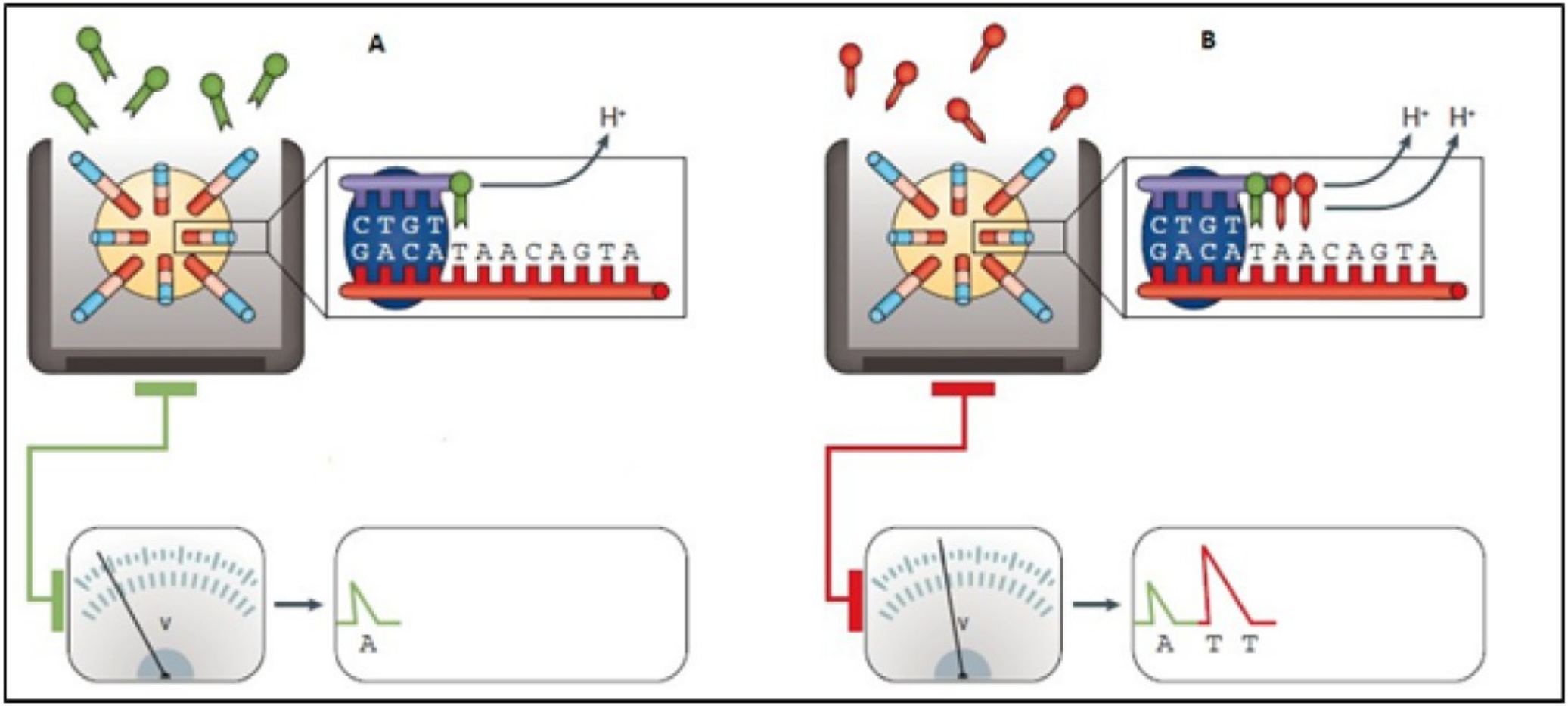
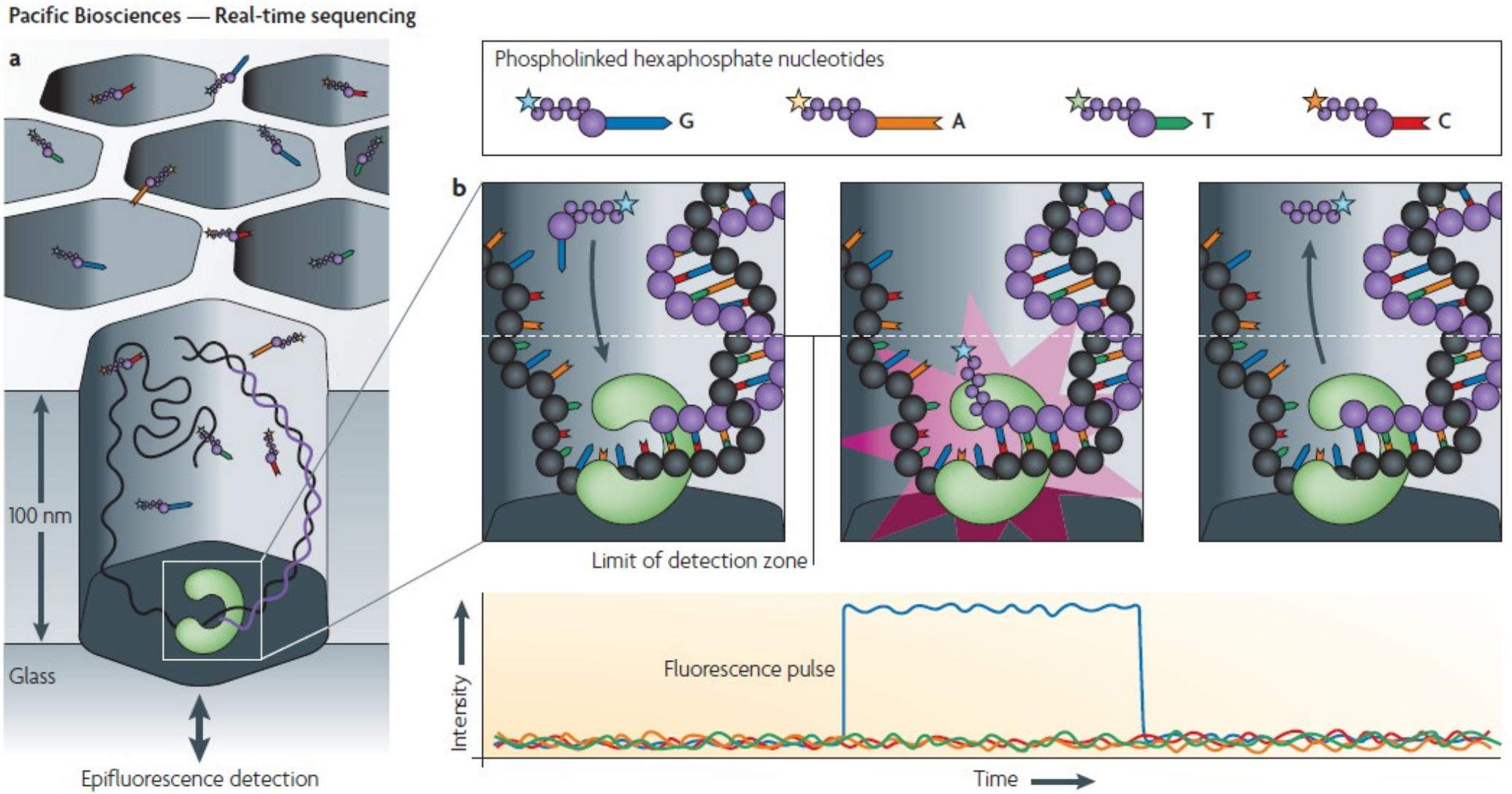


Figure was taken from: doi:10.4172/0974-8369.1000395

Real time sequencing. Pacific Biosciences

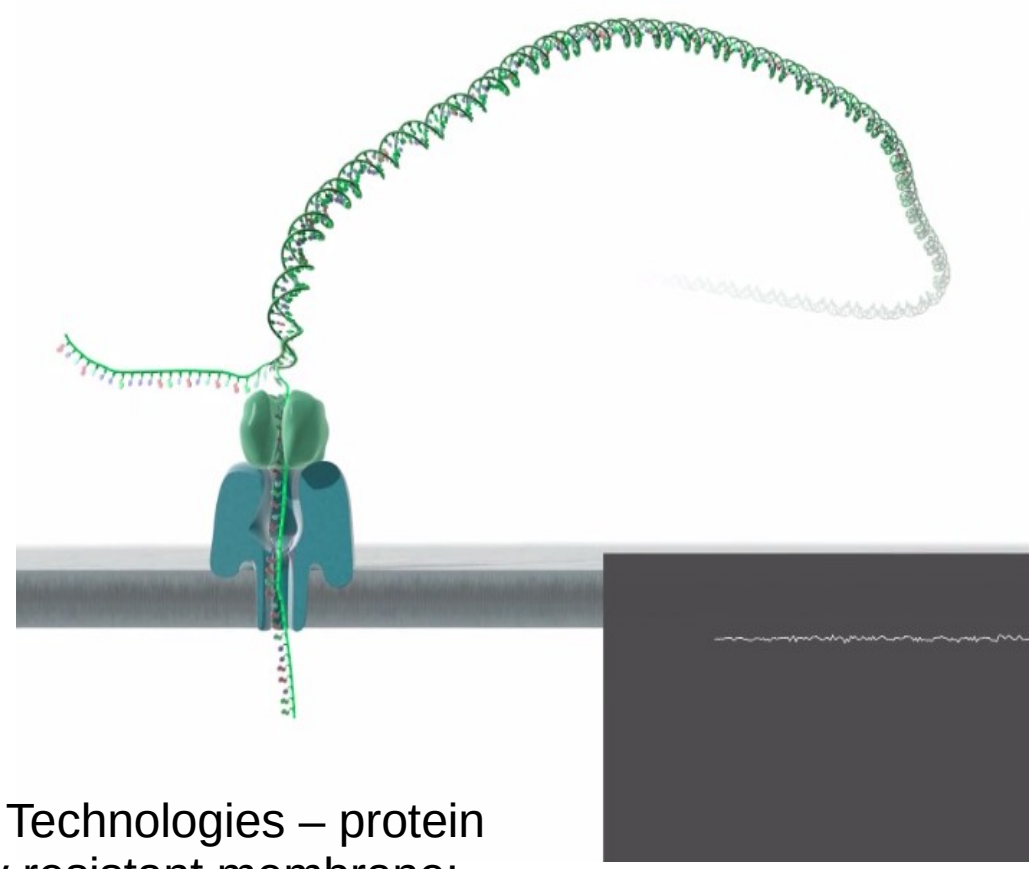


Comprehensive transcriptome analysis using synthetic long-read sequencing reveals molecular co-association of distant splicing events

Hagen Tilgner[1,3], Fereshteh Jahanbani[1,3], Tim Blauwkamp[2], Ali Moshrefi[2], Erich Jaeger[2], Feng Chen[2], Itamar Harel[1], Carlos D Bustamante[1], Morten Rasmussen[1] & Michael P Snyder[1]

Alternative splicing shapes mammalian transcriptomes, with many RNA molecules undergoing multiple distant alternative splicing events. Comprehensive transcriptome analysis, including analysis of exon co-association in the same molecule, requires deep, long-read sequencing. Here we introduce an RNA sequencing method, synthetic long-read RNA sequencing (SLR-RNA-seq), in which small pools (≤1,000 molecules/pool, ≤1 molecule/gene for most genes) of full-length cDNAs are amplified, fragmented and short-read-sequenced. We demonstrate that these RNA sequences reconstructed from the short reads from each of the pools are mostly close to full length and contain few insertion and deletion errors. We report many previously undescribed isoforms (human brain: ~13,800 affected genes, 14.5% of molecules; mouse brain ~8,600 genes, 18% of molecules) and up to 165 human distant molecularly associated exon pairs (dMAPs) and distant molecularly and mutually exclusive pairs (dMEPs). Of 16 associated pairs detected in the mouse brain, 9 are conserved in human. Our results indicate conserved mechanisms that can produce distant but phased features on transcript and proteome isoforms.

Real-time sequencing



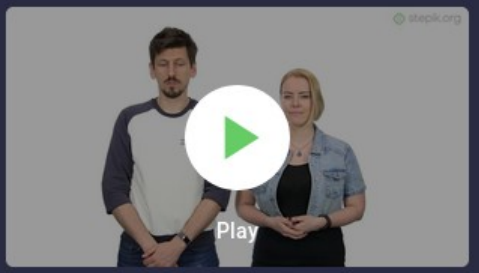
Oxford Nanopore Technologies – protein pore in electrically resistant membrane:

- Strand sequencing
- Exonuclease sequencing

stepik Catalog Teach Search English Log in Register

Секвенирование 3-го поколения на Oxford Nanopore

Основы секвенирования 3-го поколения на платформе Oxford Nanopore. Рассматриваются основные применения, типы предлагаемых приборов, используемые протоколы, и особенности биоинформатической обработки получаемых данных.



2 часа в неделю

Certificate Stepik

★★★★★ 4.8 25 reviews

1428 learners

Bioinformatics Institute

About this course

Секвенирование 3-го поколения (3rd generation NGS) - методы высокоэффективного секвенирования, позволяющие получать длинные прочтения (от 1 тыс. и до миллионов нуклеотидов). Наш курс посвящен секвенированию при помощи приборов от компании Oxford Nanopore. Вы познакомитесь с теоретическими основами работы этих приборов, получите практические советы по запуску и решению основных проблем в работе, а также ознакомитесь с особенностями биоинформатической обработки длинных прочтений.

Данный курс требует понимания основ молекулярной биологии, секвенирования, и биоинформатики, однако не имеет жестких требований по предварительному прохождению других курсов.

Free

Join this course

You can learn right away

This course includes

25 lessons
3 hours of video
24 quizzes

Course content

Last update 07/16/2019

https://stepik.org/course/8092/promo