
DTU

DTU Health Technology
Bioinformatics

Introduction to NGS technology

Gabriel Renaud
Associate Professor
Section of Bioinformatics
Technical University of Denmark
gabriel.reno@gmail.com

Menu

- 2nd generation NGS
- Illumina movietime!
- Your turn to basecall
- 3rd generation NGS

2 main types of approaches

1) Amplify and sequence one base at a time

```
1:A   2:G   3:G  4:T   =     AGGT
```

2) Amplify and count how many of the same base

```
1:1A  2:2G  3:1T       =     AGGT
```

Second generation sequencing

- Illumina sits on 75% of the market

Illumina

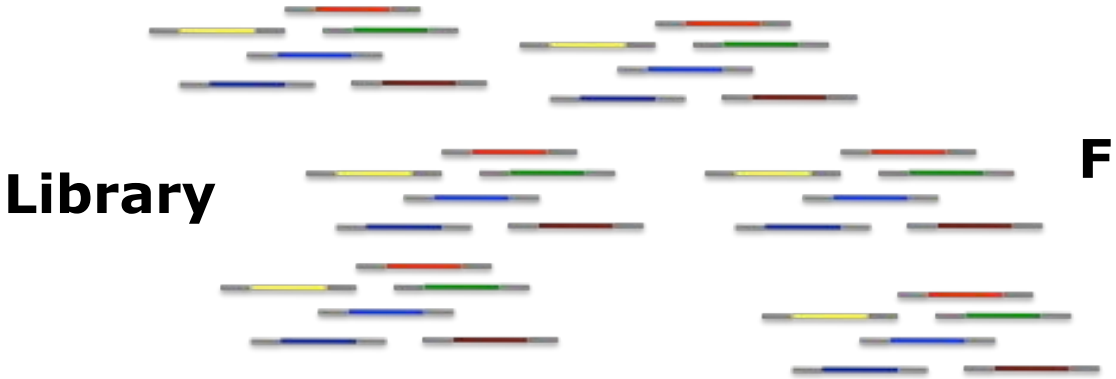
BGISEQ

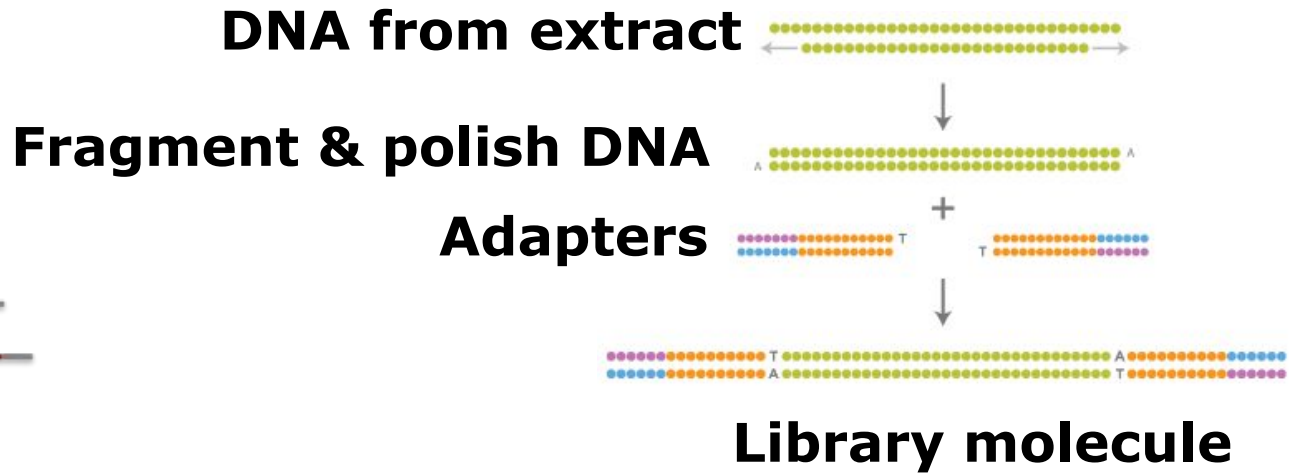
454

Ion Torrent

General library preparation steps

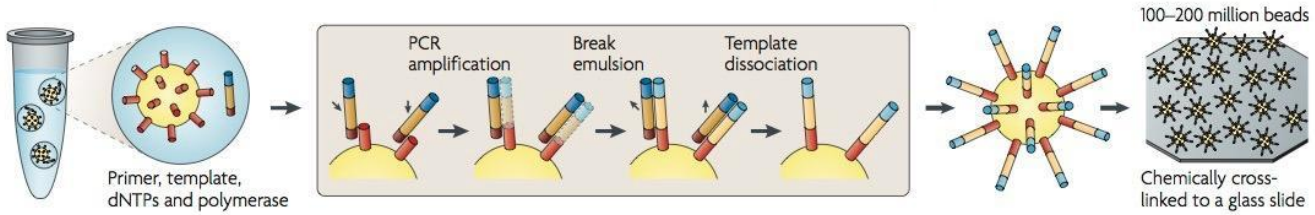
1. Create library molecules
2. Amplification (PCR)
3. Massive parallel sequencing (strength over Sanger)





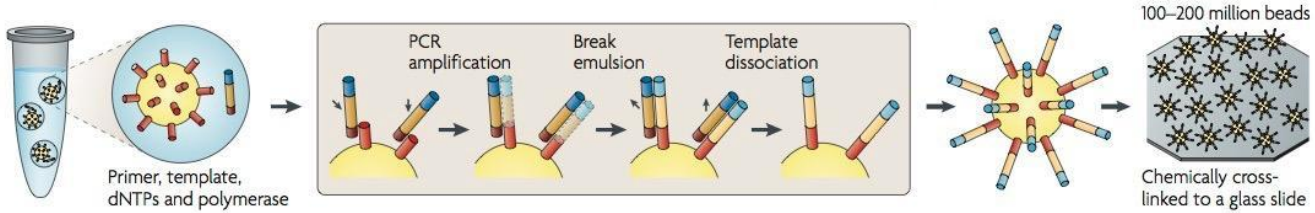
Amplification and immobilization

- Emulsion PCR (454, SOLiD, IonTorrent): Water, oil, beads, one DNA template/droplet

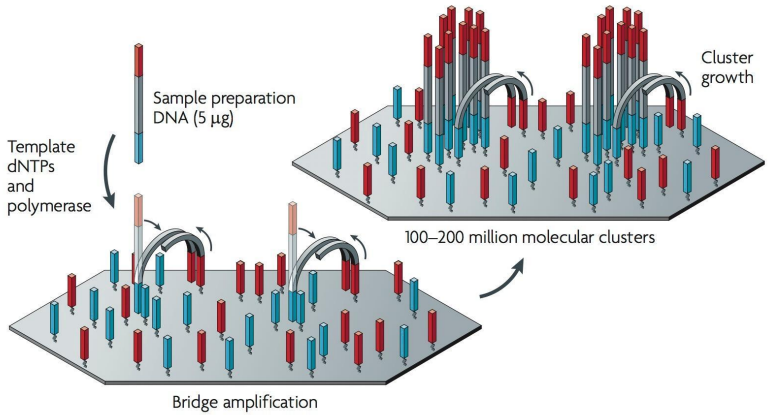


Amplification and immobilization

- Emulsion PCR (454, SOLiD, IonTorrent): Water, oil, beads, one DNA template/droplet



Bridge PCR (Illumina): One DNA template/cluster, primers on surface, grow by bridging primers



2 main types of approaches

1) Amplify and sequence one base at a time

```
1:A   2:G   3:G   4:T   =   AGGT
```

2) Amplify and count how many of the same base

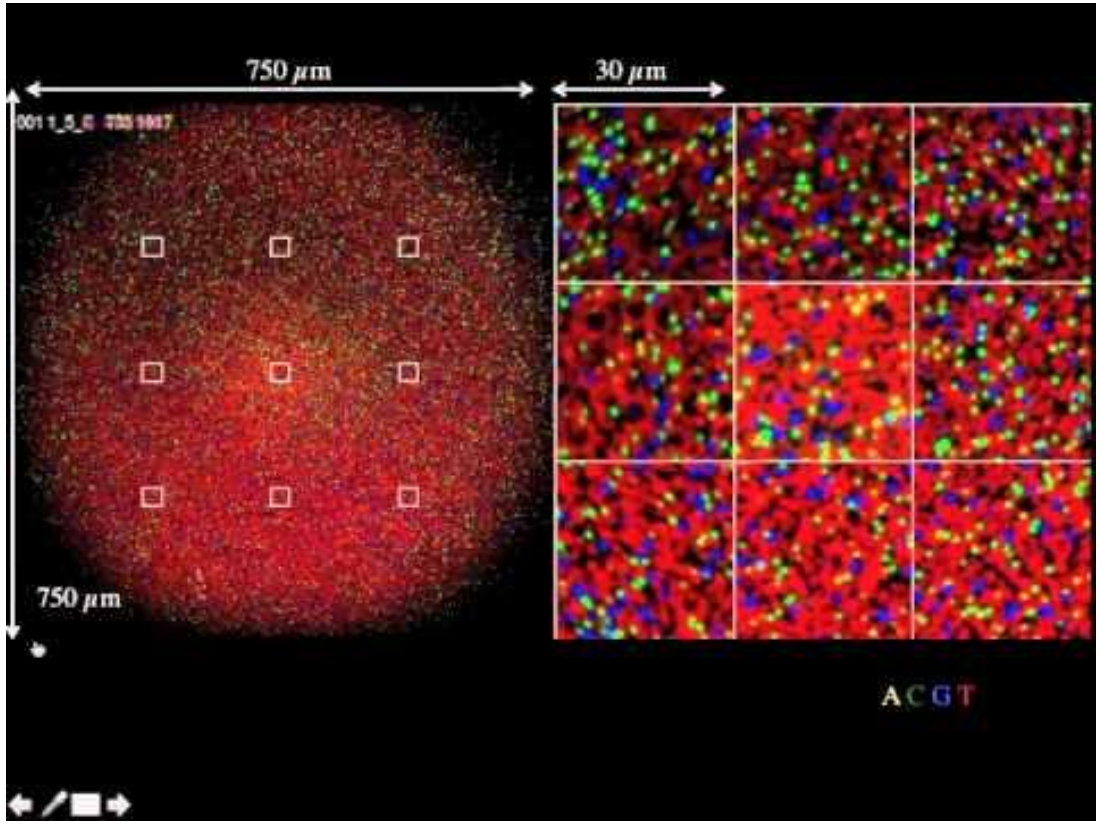
```
1:1A   2:2G   3:1T   =   AGGT
```

Illumina sequencing

corporate propaganda:
https://www.youtube.com/watch?v=HMyCqWhwB8E

Amplicon sequencing on Illumina

- Why can’t you just fill your Illumina flow cell with amplicon libraries?



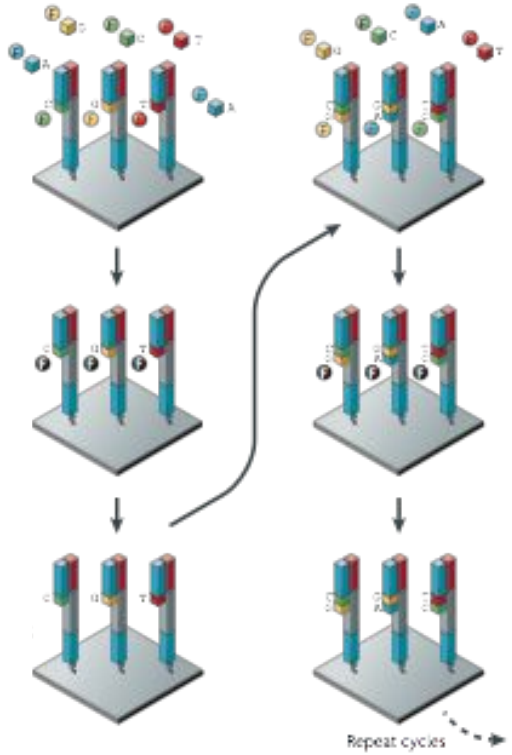
Fluorescence detection

Illumina - Cyclic reversible termination

Add all dNTPs labelled w. diff dye

Create four-color image

Cleave dye and repeat next cycle



2G: Imaging handout

Illumina 1:_____

Illumina 2:_____

2G: Imaging handout Answers!

Illumina 1:_____

Illumina 2:_____

TOP: **CATCGT**

BOTTOM: **CCCCCC**

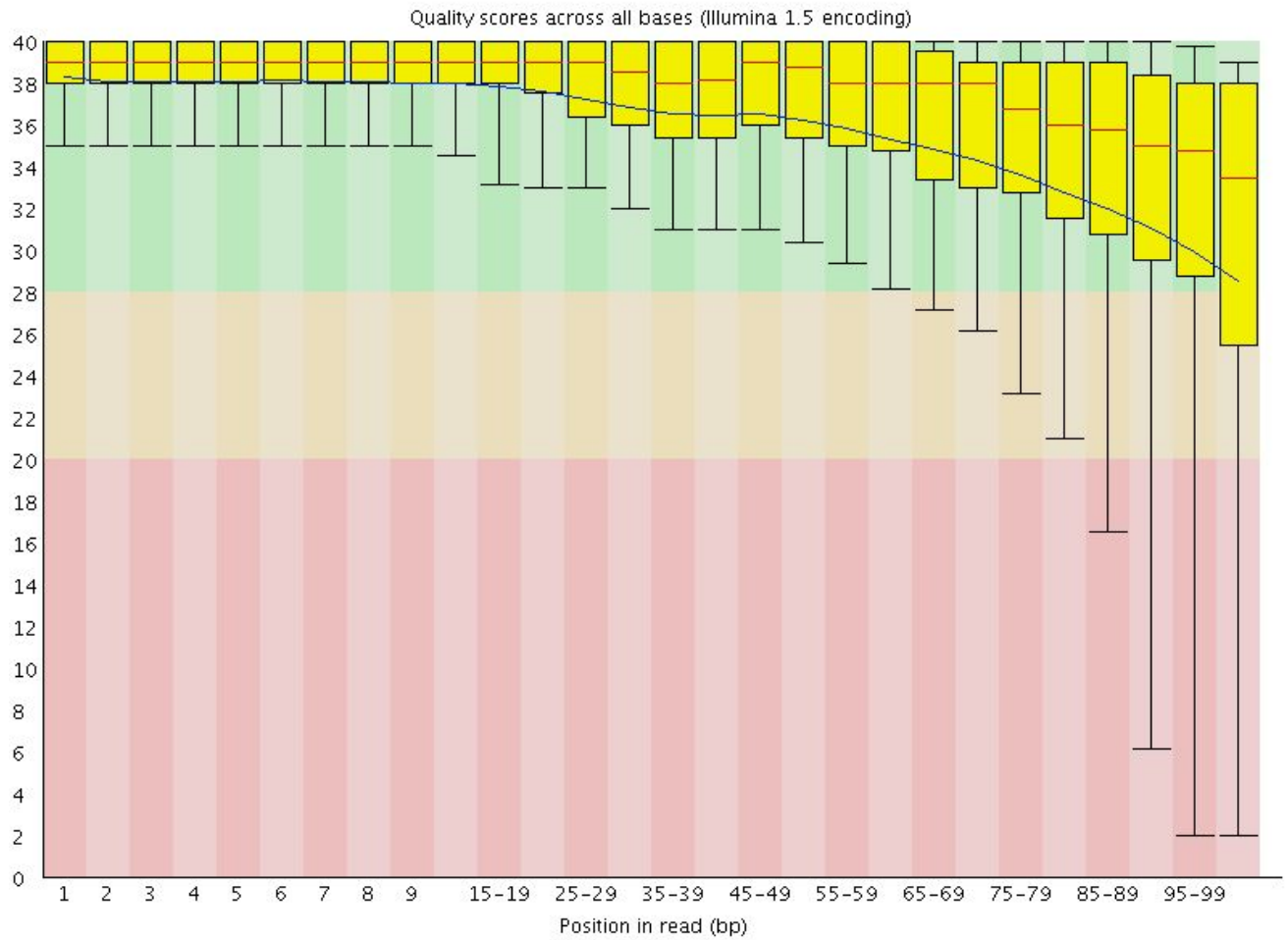
___

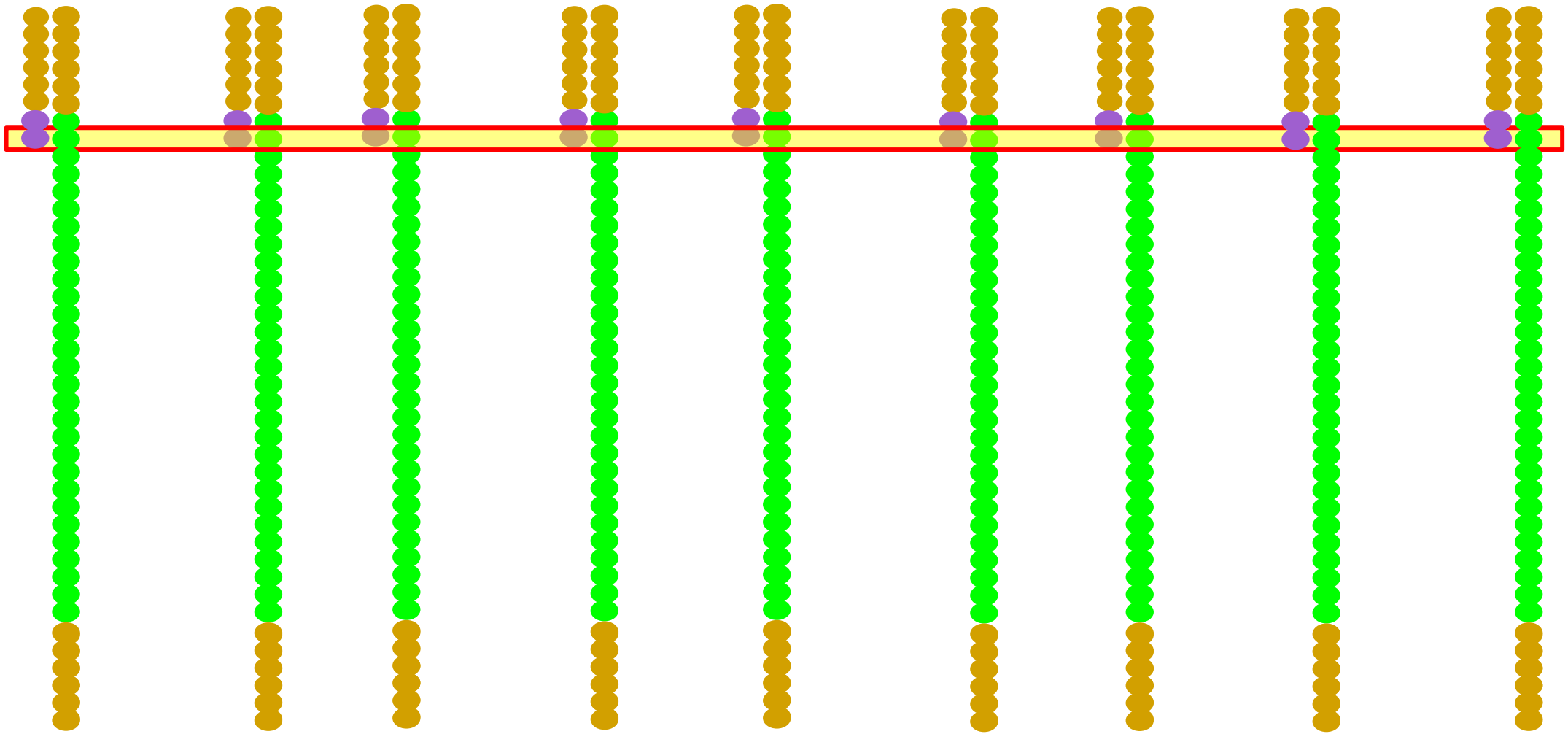
__

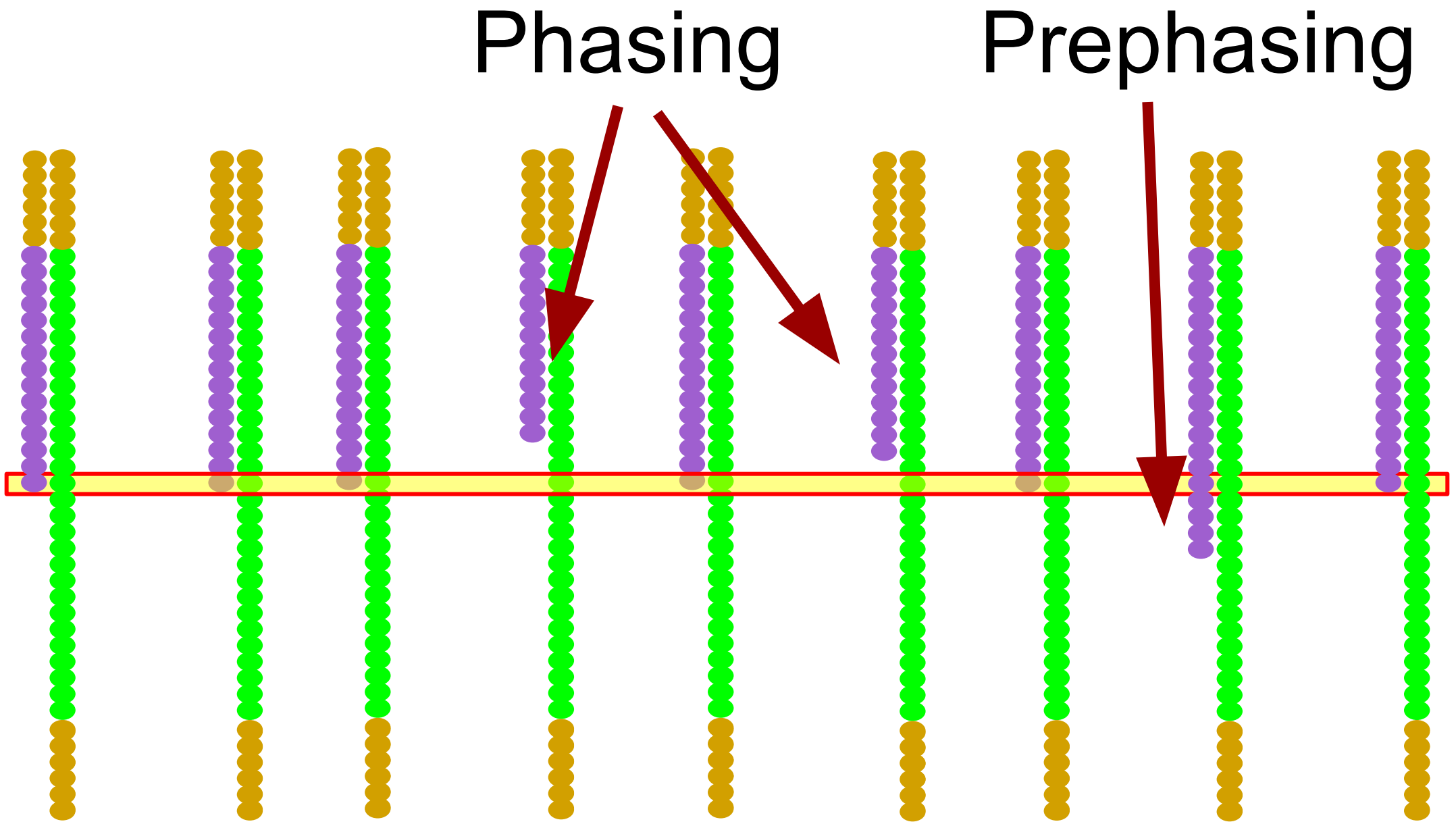
__

Illumina: Quality deterioration

- Quality goes down
- Especially 2nd read
- Can you think of why?

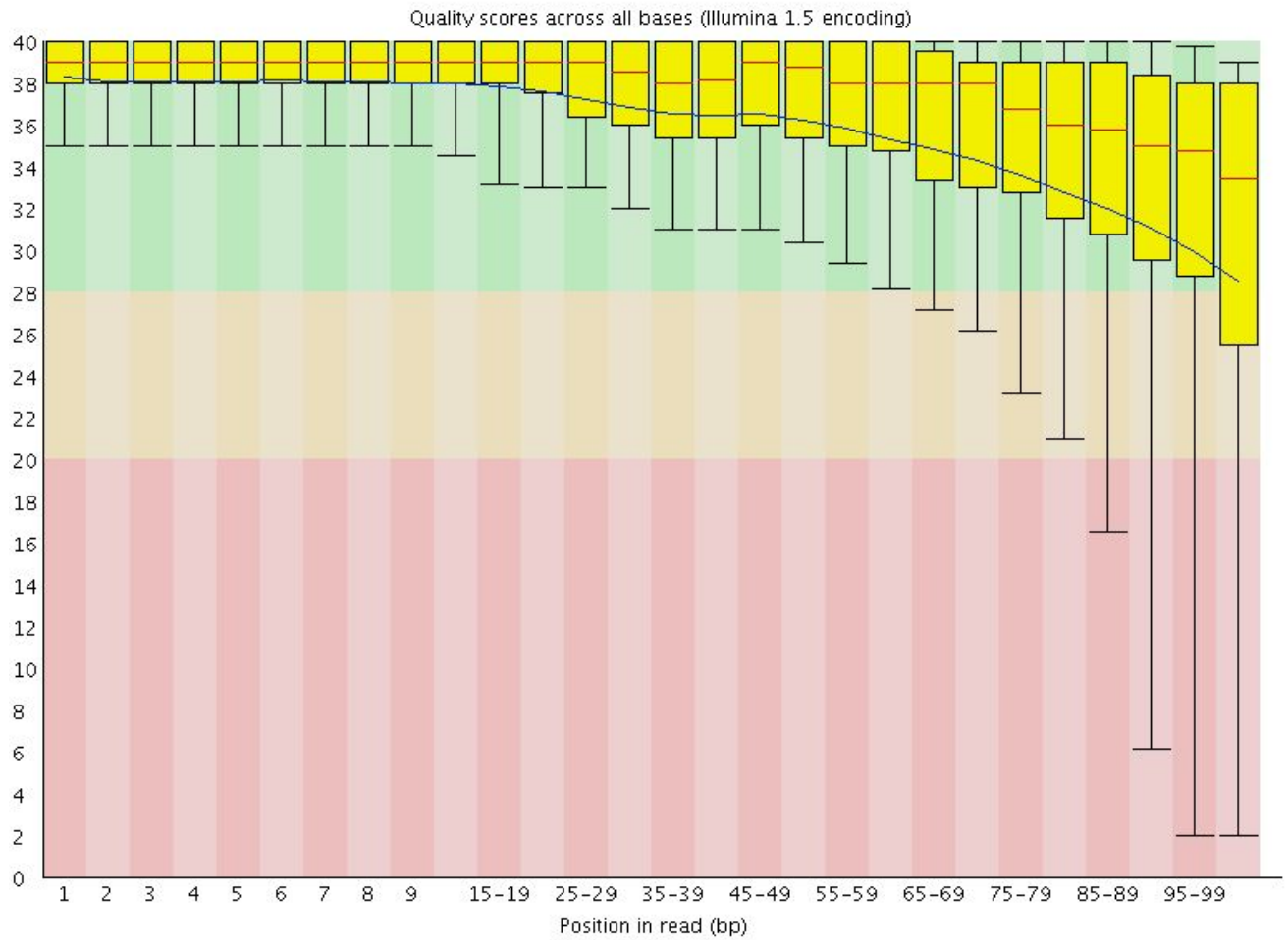





Phasing
Prephasing

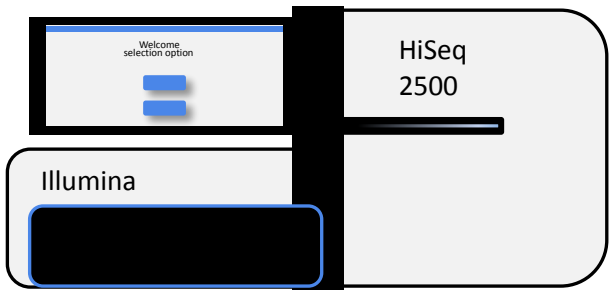
Illumina: Quality deterioration

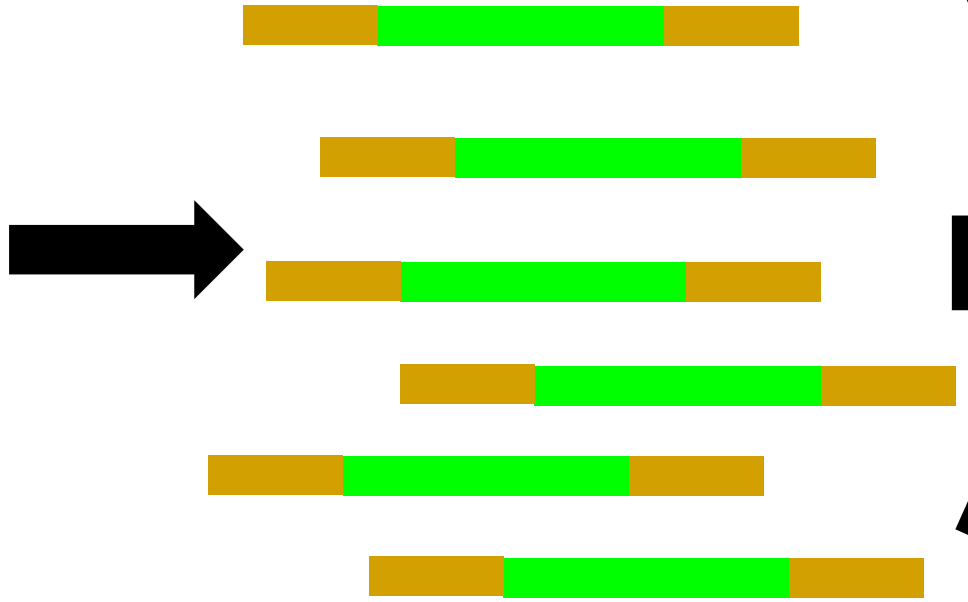
- Quality goes down
- Especially 2nd read
- Can you think of why?

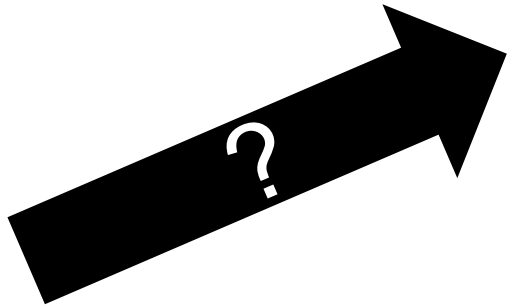


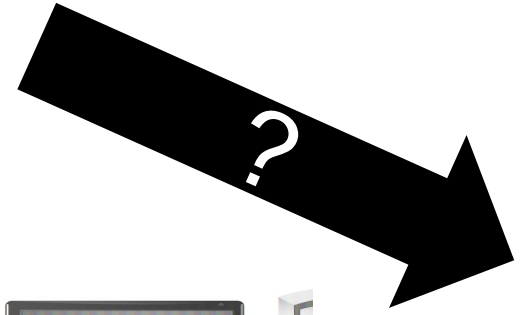
- Efficiency of incorporation
- Phasing
- Prephasing

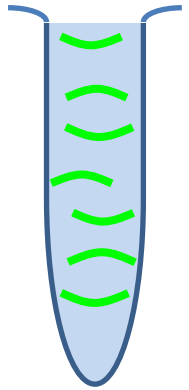
Demultiplexing

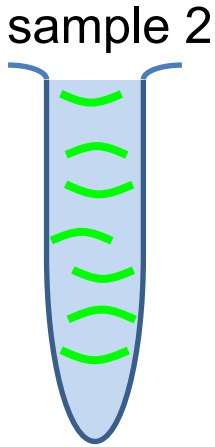


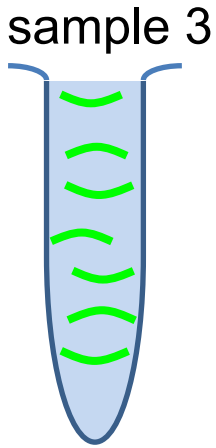






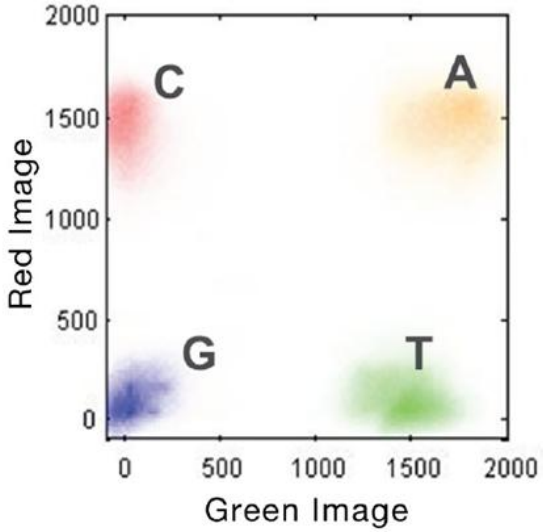






NextSeq/NovaSeq

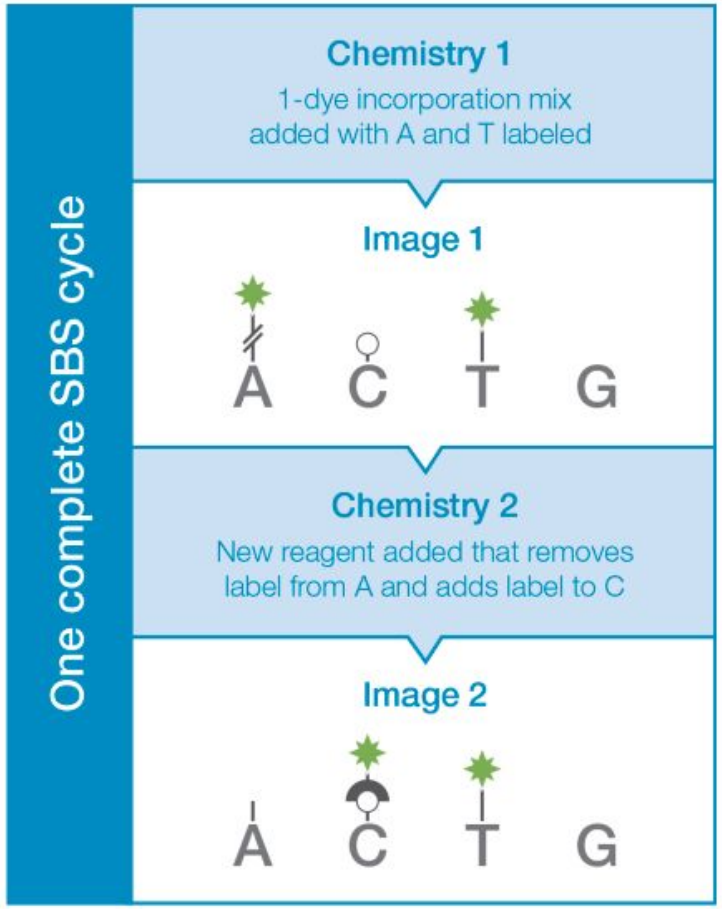
- Chemistry is not based 4 dyes (as before) but 2 dyes
  - T (red), C (green), A (both) and G (none = “dark”)
  - Faster processing rate and cheaper reagents
  - Slightly increases error rate
  - Problem with G stretches because G is not dyed



source: Illumina

1 dye, 2 images

A.



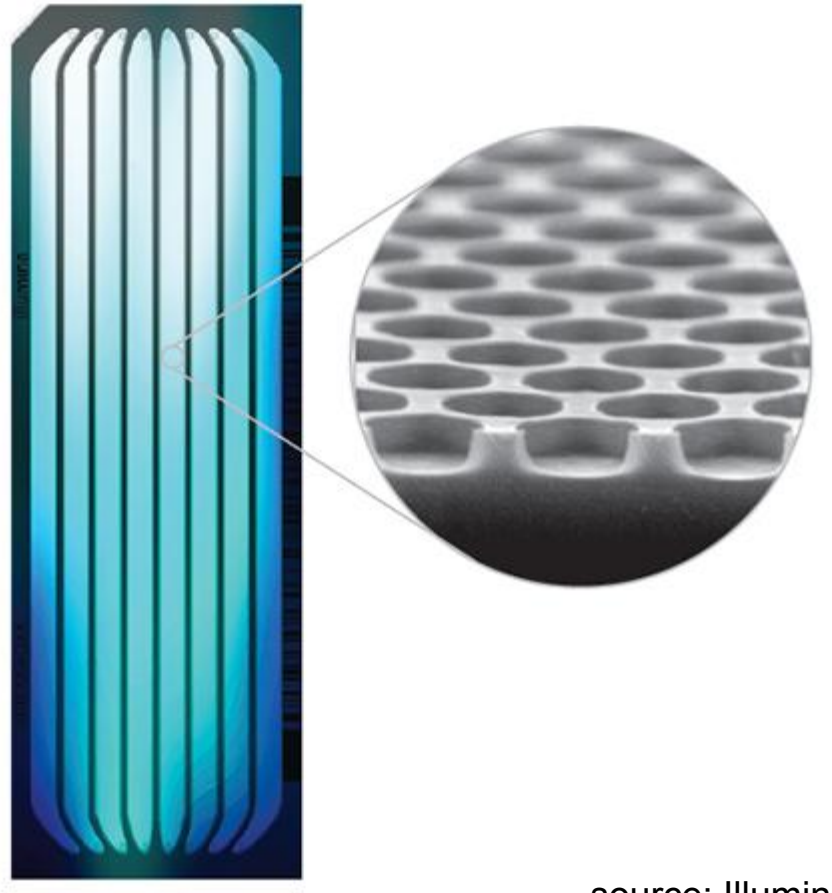
B.

Image 1	Image 2	Result
ON	OFF	A
OFF	ON	C
ON	ON	T
OFF	OFF	G

source: Illumina

Patterned flowcell

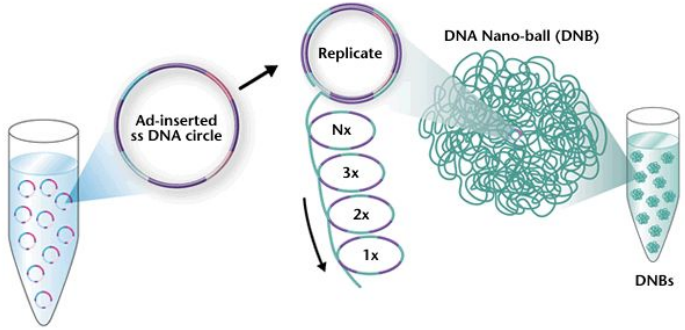
- Patterned wells
- Solves overloading flowcell
- More duplicates



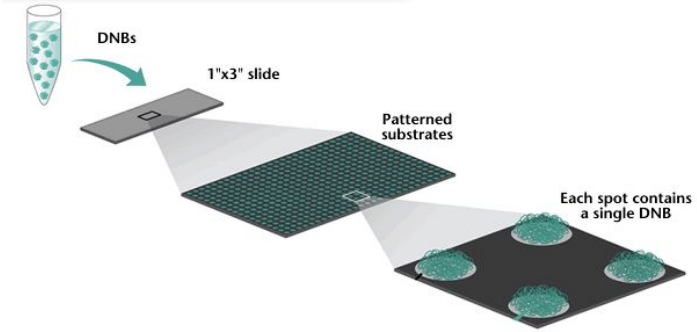
source: Illumina

BGI-Seq

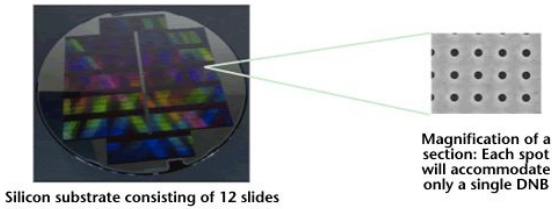
ssDNA -> DNA nanoballs



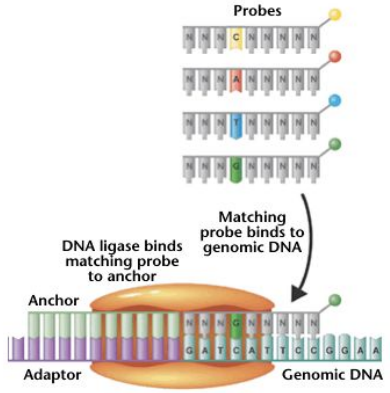
Place DNBs into each spot



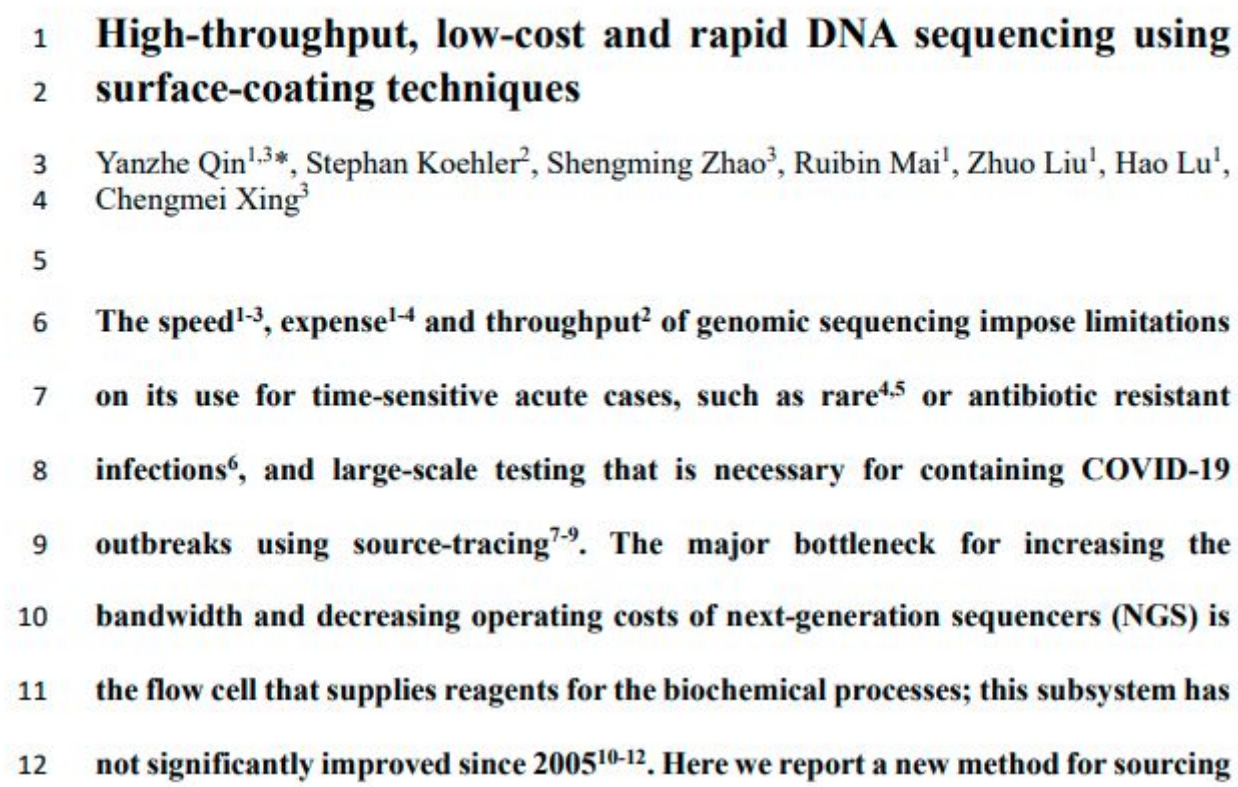
Use silicon chips with sticky spots



Sequence using ligase and fluorescent labeled probes



BGI-Seq



High-throughput, low-cost and rapid DNA sequencing using surface-coating techniques

Yanzhe Qin[1,3]*, Stephan Koehler[2], Shengming Zhao[3], Ruibin Mai[1], Zhuo Liu[1], Hao Lu[1], Chengmei Xing[3]

The speed[1-3], expense[1-4] and throughput[2] of genomic sequencing impose limitations on its use for time-sensitive acute cases, such as rare[4,5] or antibiotic resistant infections[6], and large-scale testing that is necessary for containing COVID-19 outbreaks using source-tracing[7-9]. The major bottleneck for increasing the bandwidth and decreasing operating costs of next-generation sequencers (NGS) is the flow cell that supplies reagents for the biochemical processes; this subsystem has not significantly improved since 2005[10-12]. Here we report a new method for sourcing

"These improvements drop the turn-around time from days to twelve hours and the cost for whole genome sequencing (WGS) from about $1000 to $15, as well as increase data production by several orders of magnitude."

High-throughput, low-cost and rapid DNA sequencing using surface-coating techniques
Yanzhe Qin, Stephan Koehler, Shengming Zhao, Ruibin Mai, Zhuo Liu, Hao Lu, Chengmei Xing
bioRxiv 2020.12.10.418962; doi: https://doi.org/10.1101/2020.12.10.418962

2 main types of approaches

1) Amplify and sequence one base at a time

```
1:A   2:G   3:G   4:T   =   AGGT
```

2) Amplify and count how many of the same base

```
1:1A   2:2G   3:1T   =   AGGT
```

454: Pyrosequencing

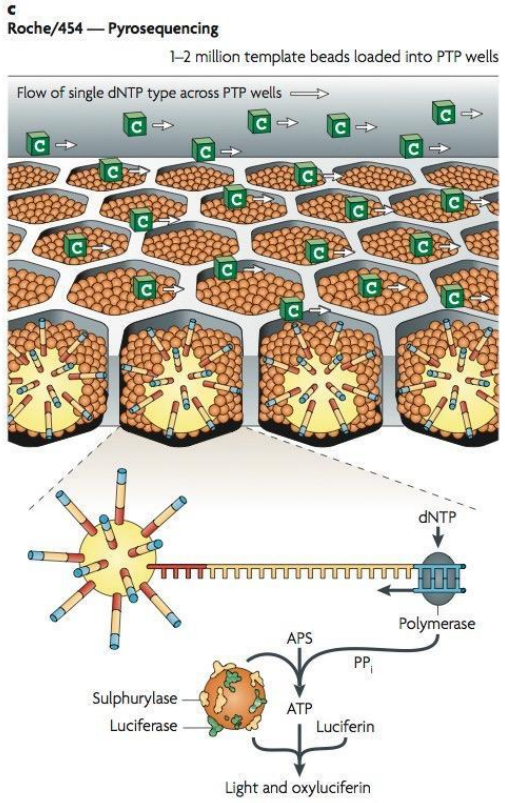
Load template beads into wells

Flow one dNTP across wells

Polymerase incorporates nucleotide

Release of PPi leads to light

Imaging, next dNTP



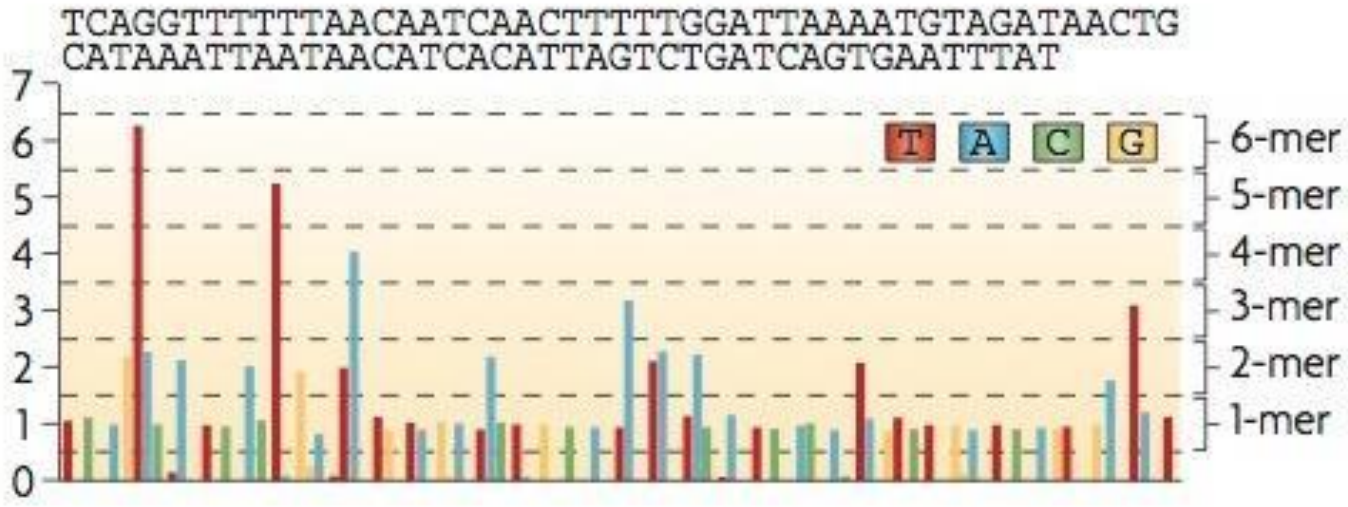
2G: Imaging handout

454: _____

_____

_____

2G: Imaging handout Answers!



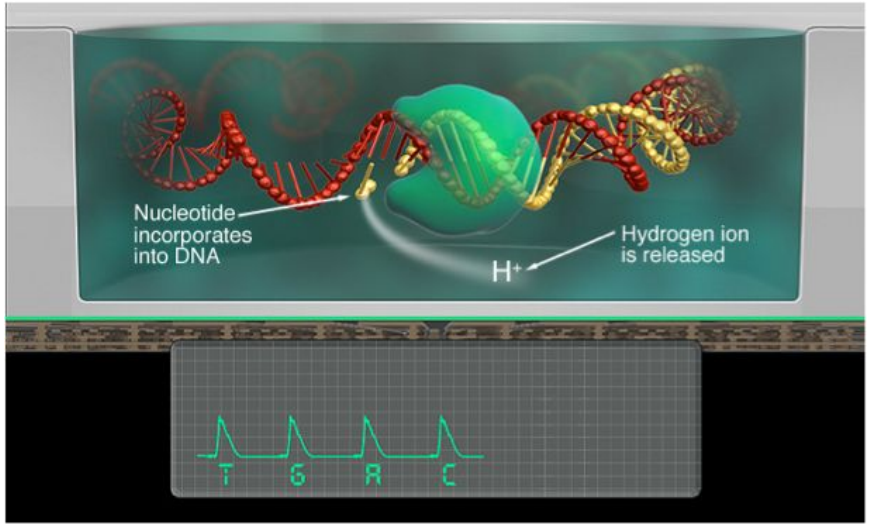
454: _____

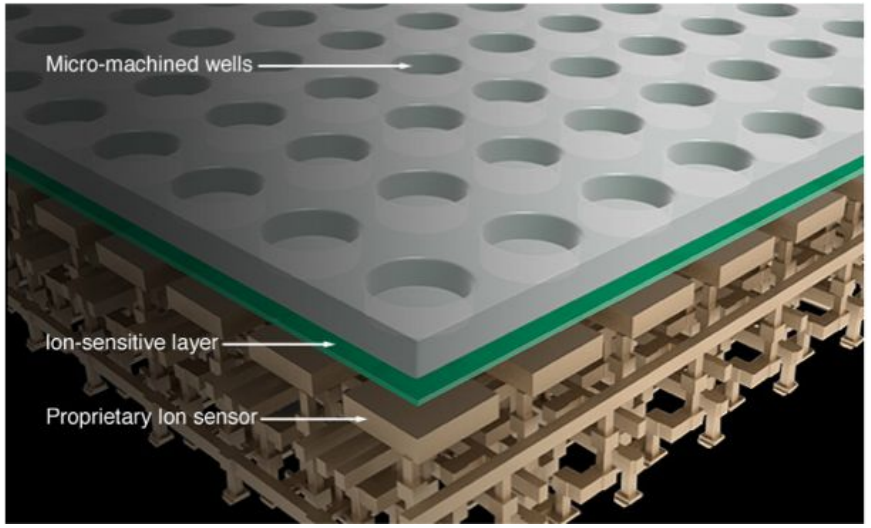
_____

_____

Ion Torrent

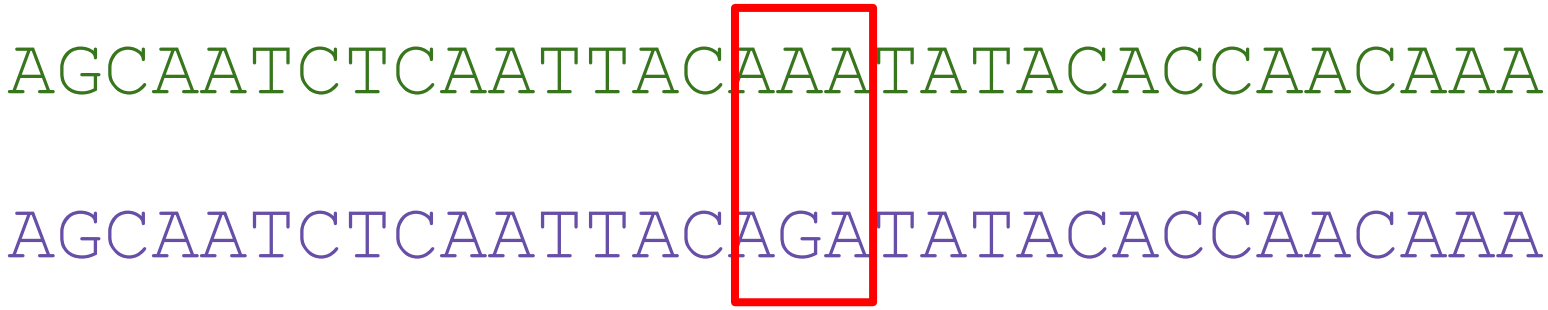
- Corporate propaganda: https://www.youtube.com/watch?v=zBPKj0mMcD
- Similar principle to 454
- Library: Emulsion PCR
- Based on semiconductors
- Detection is based on H ions (pH) changes



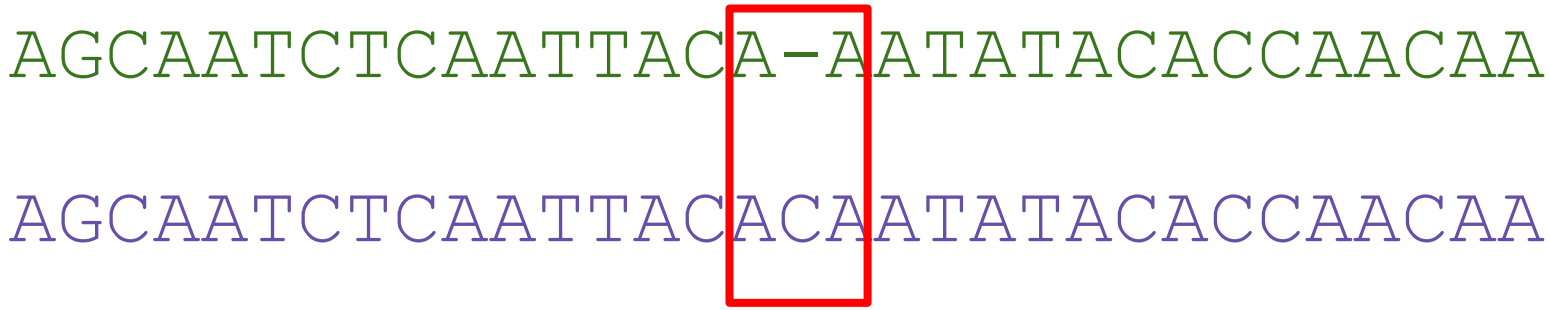


Let's remember the types of errors

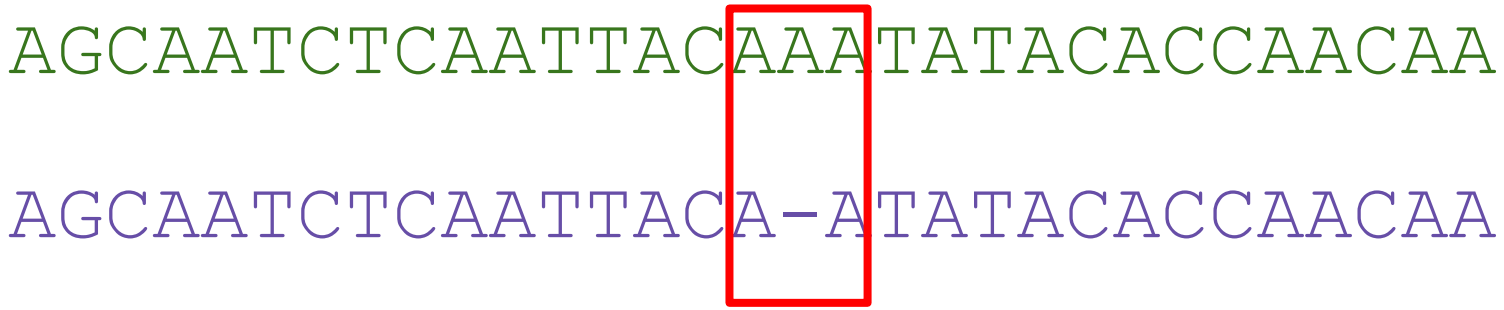
mismatch

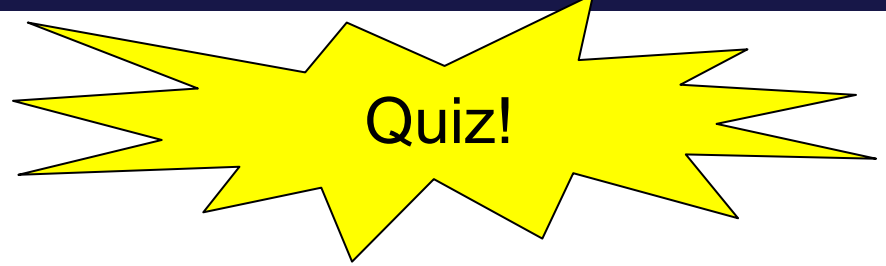


insert



deletion





Which of the the 2 main types of approaches would be more prone to indels?

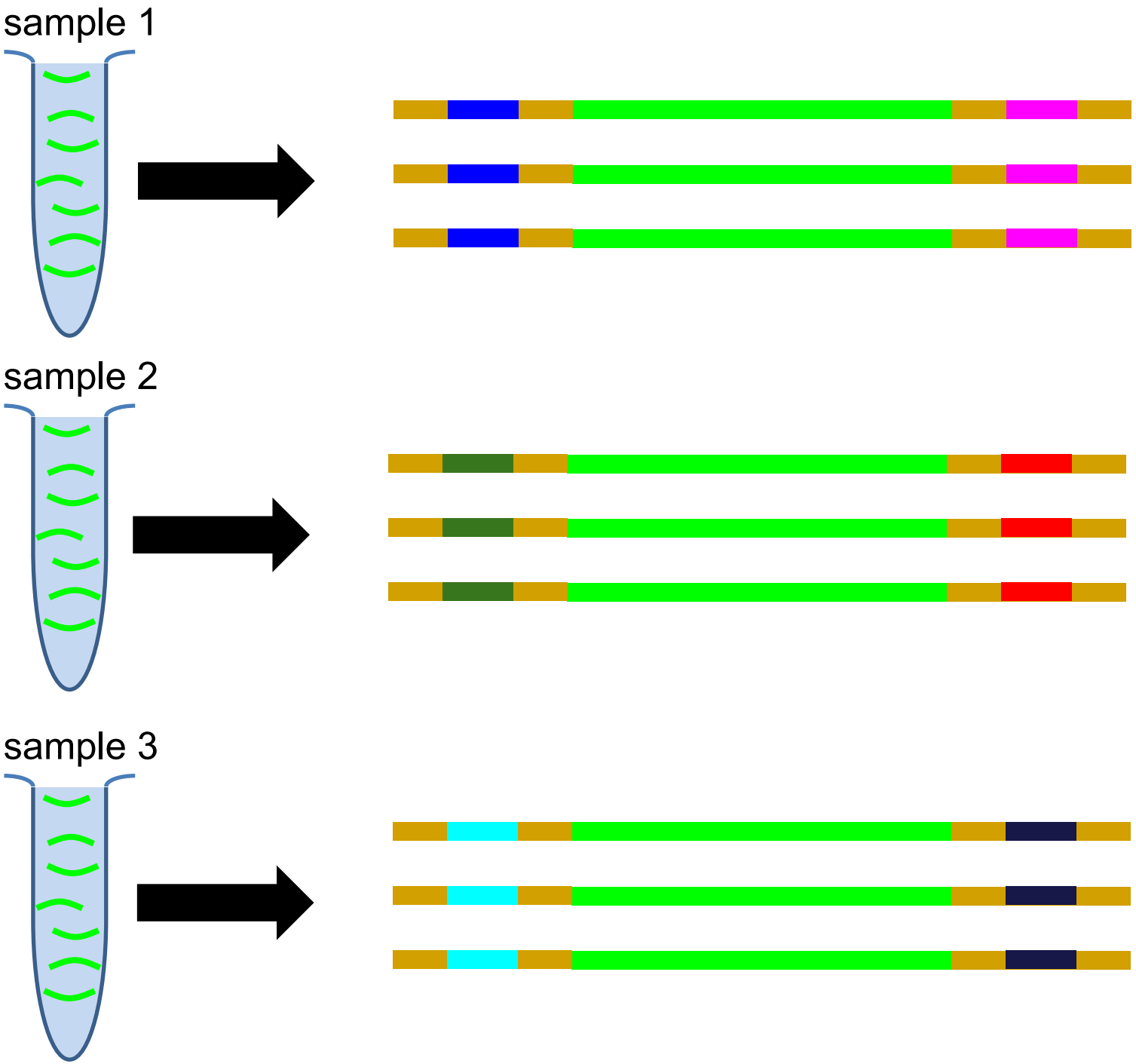
1) Amplify and sequence one base at a time

```
1:A   2:G   3:G  4:T   =     AGGT
```

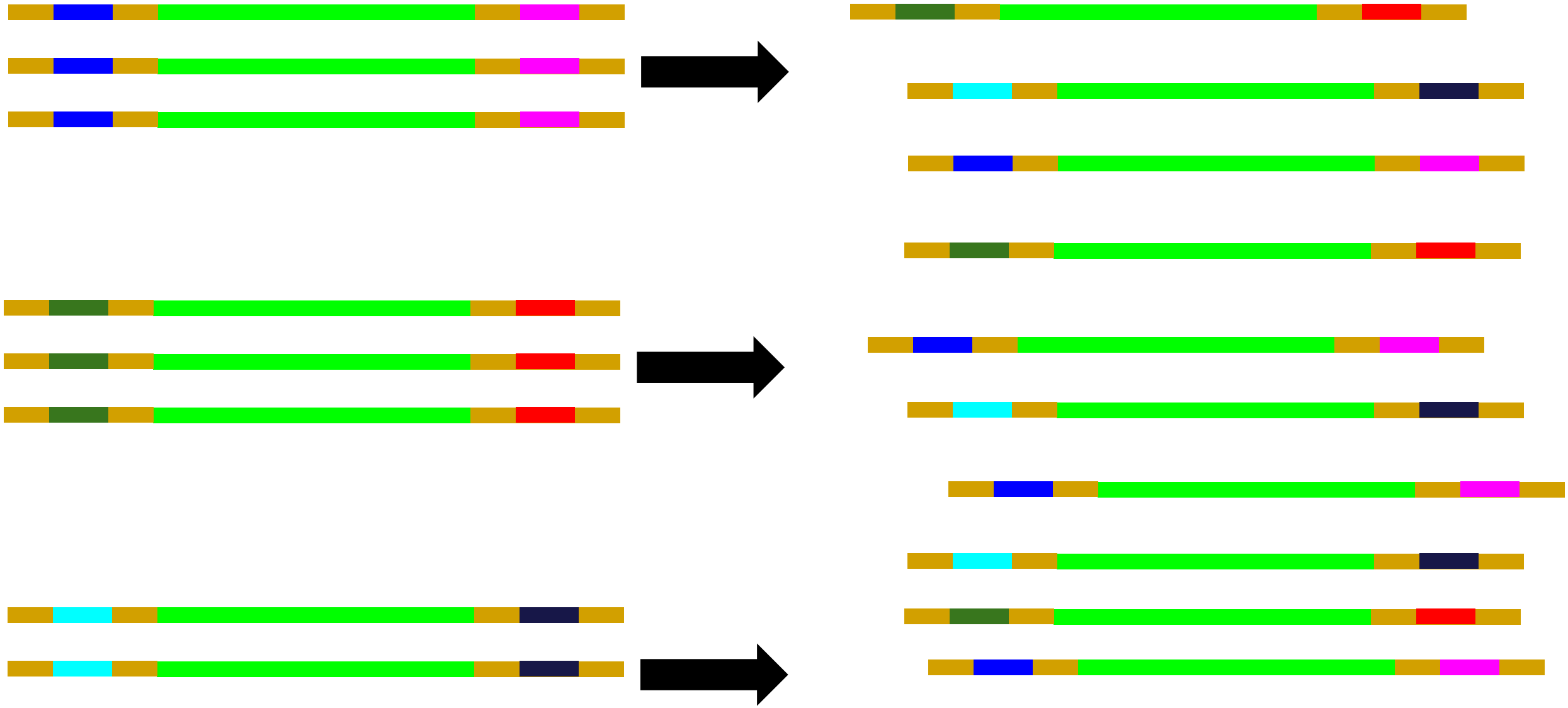
2) Amplify and count how many of the same base

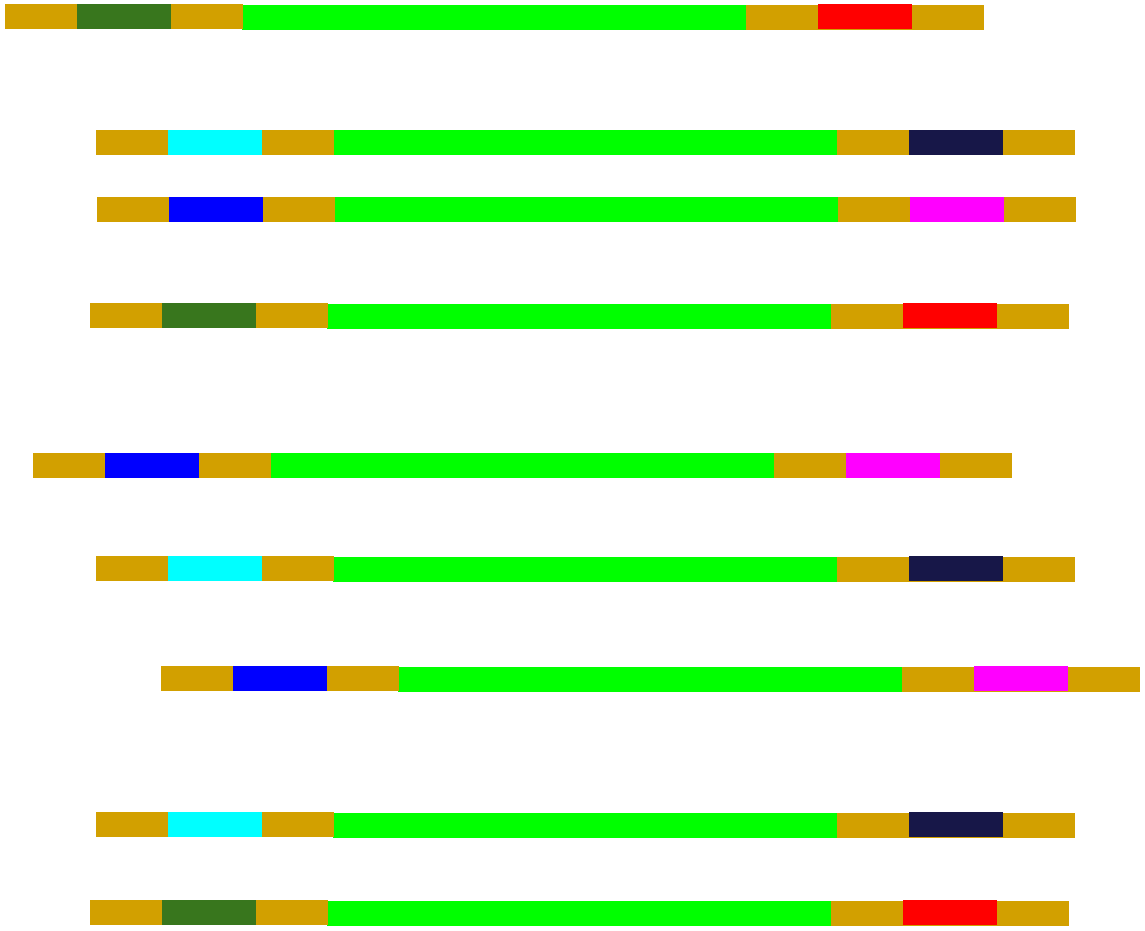
```
1:1A  2:2G  3:1T       =     AGGT
```

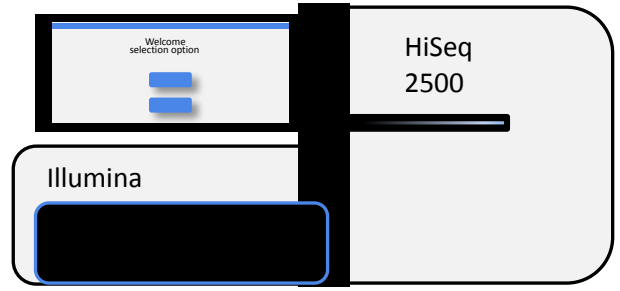
Technology	read length	amount reads	errors?
Sanger	400 to 900 bp	96	mm 0.01%
Illumina MiSeq	2x 200-300bp	20-30 M per flow cell	mm 0.1-0.2%
Illumina HiSeq	2x 100-150bp	~2G per flow cell	mm 0.1-0.2%
Illumina NovaSeq	2x 100-150bp	~6-10G per flow cell	mm 0.1%?
Ion Torrent	~200-400 bp	5-50M reads	indel 0.46 to 2.4%



Multiplexing

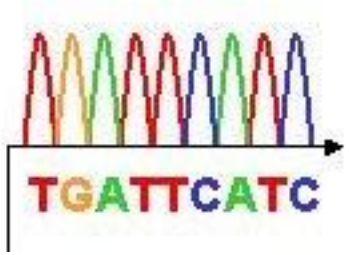





Welcome
selection option
HiSeq
2500
Illumina

3rd generation

1977 1980 1983 1986 1989 1992 1995 1998 2001 2004 2007 2010 2013 2016 2019



Sanger

454

PacBio

Ion Torrent

Oxford Nanopore

SOLiD

BGI

Illumina

3rd generation

- Single-molecule sequencing
- No amplification -> less bias -> observations are more independent

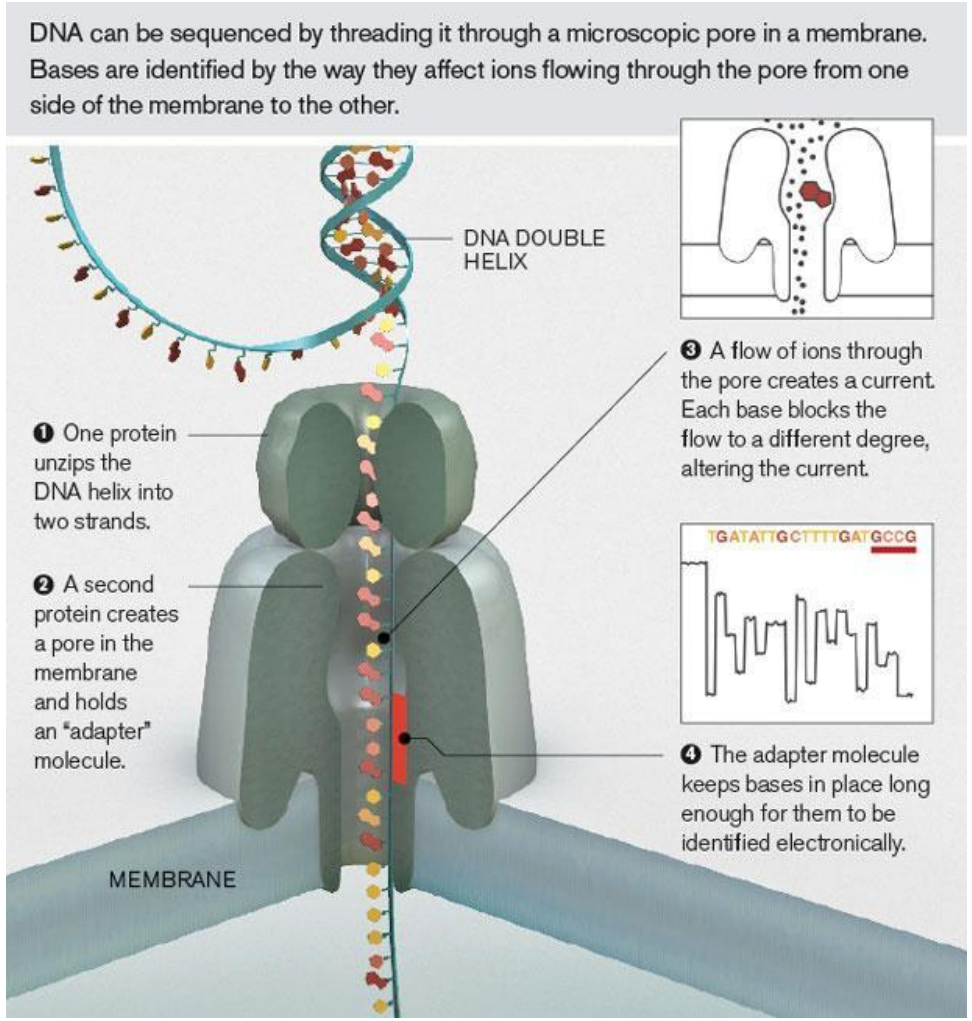
Helicos

PacBio

Oxford Nanopore

Oxford Nanopore

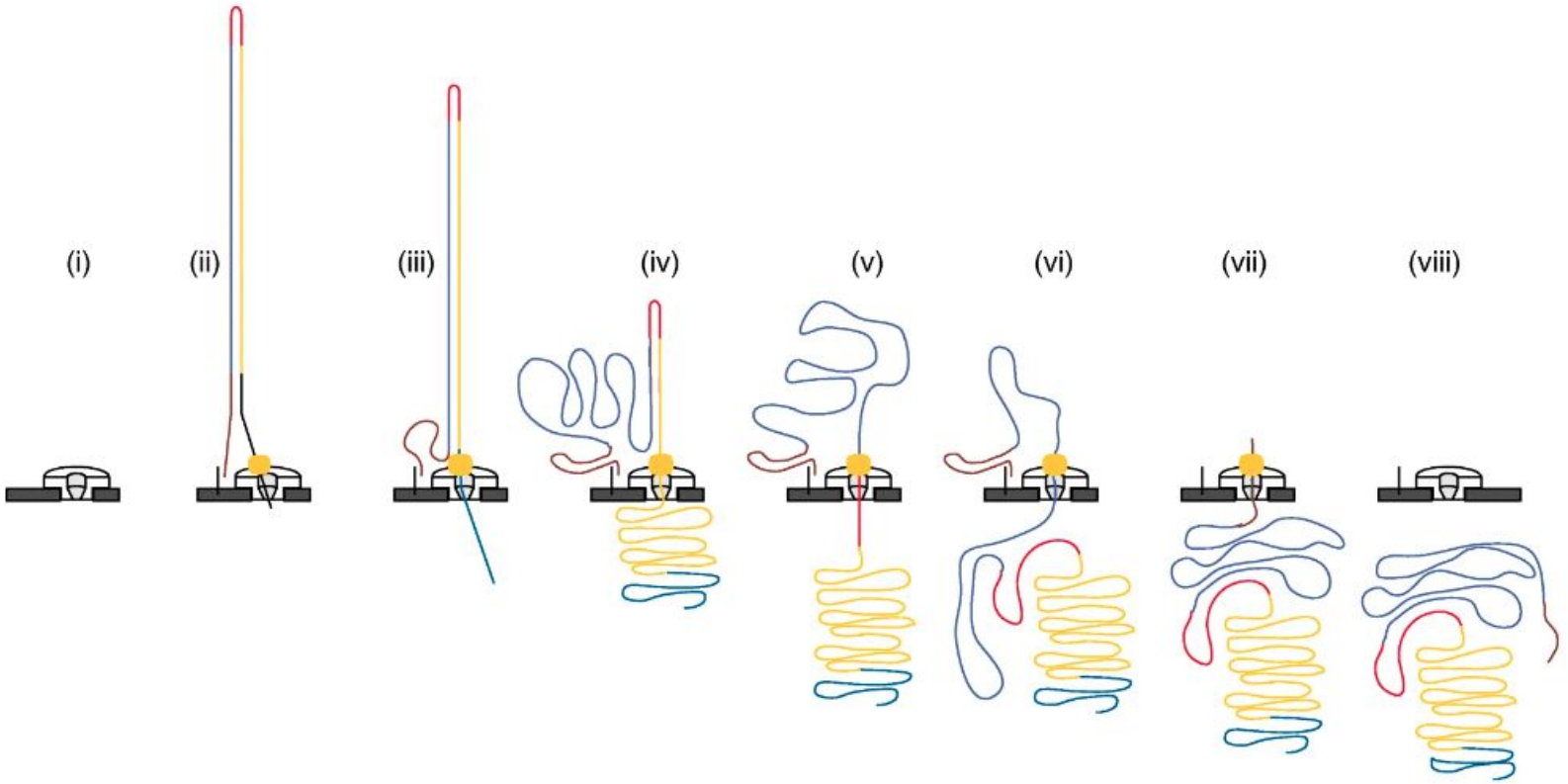
- Literal nanopores
- Current per base
- Non-random errors
- https://www.youtube.com/watch?v=RcP85JHLmnI
- Very high error rate



"If a nanopore was the size of a fist, a 1MB strand of DNA passing through that nanopore would be 2 miles (3.2 km) long"
-Adam Philippy, NHGRI

Oxford Nanopore

- Hairpin allows double sequencing (2D)



Jain, M., Olsen, H.E., Paten, B. et al. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome Biol 17, 239 (2016). https://doi.org/10.1186/s13059-016-1103-0

Cheap & mobile

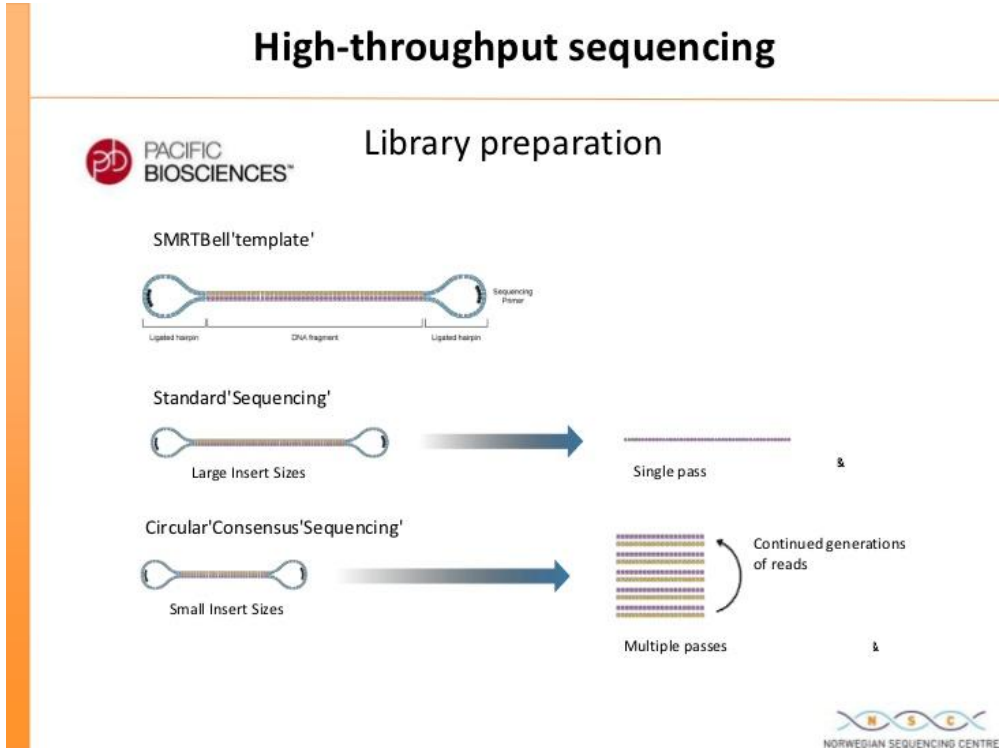
- Long reads, low quality
- Low establishment and maintenance costs
- Portability

PacBio: Single-molecule real-time (SMRT) sequencing

- Expensive machinery
- Not very portable

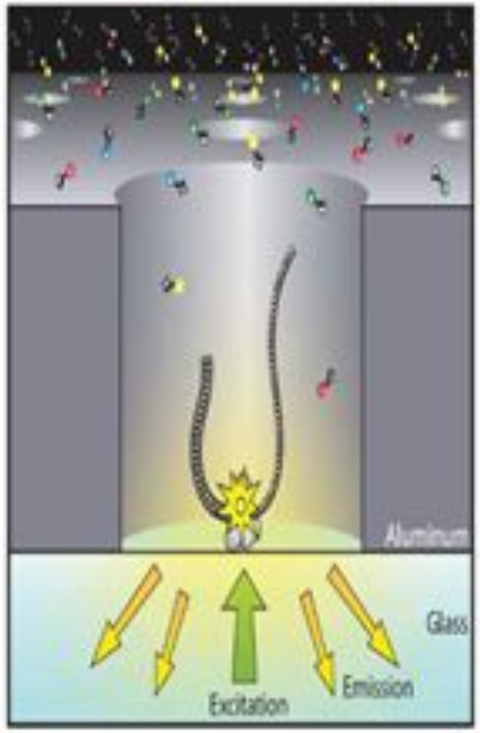
PacBio

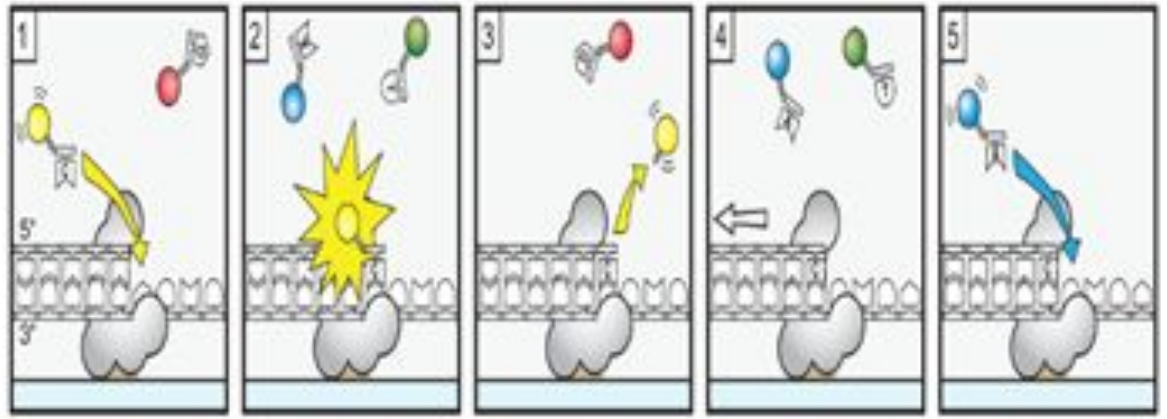
- Flexibility
  - Long but low quality or shorter but better reads
  - Robust
  - https://www.youtube.com/watch?v=_lD8JyAbwEo
  - New 2019: HiFi read same fragment multiple times

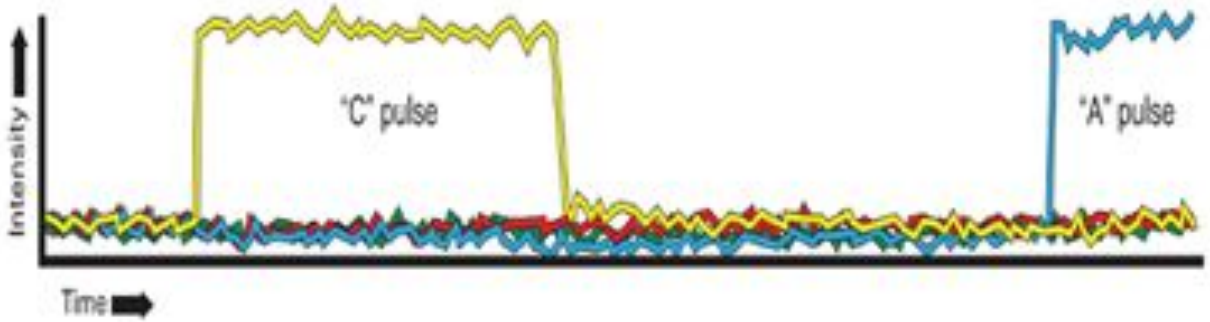


Tiny wells

- 1 million wells per cell
- Hit the lights







Technology	read length	amount reads	errors?
Oxford Nanopore	avg. 2 kbp-10 kbp	~300-500k	2020 update: 3%-7% 1D: indel+mm 20% 2D: indel+mm 7%
PacBio	avg. >10kbp	~55k	indel+mm 13-15% HiFi: indel 5%+mm 0.1%

Summary

- Illumina is the current workhorse
  - Great for many applications
- Long read technology
  - Adding information
  - Resolves difficult regions during genome assembly

Article | Open Access | Published: 14 July 2020

Telomere-to-telomere assembly of a complete human X chromosome

Karen H. Miga, Sergey Koren, Arang Rhie, Mitchell R. Vollger, Ariel Gershman, Andrey Bzikadze, Shelise Brooks, Edmund Howe, David Porubsky, Glennis A. Logsdon, Valerie A. Schneider, Tamara Potapova, Jonathan Wood, William Chow, Joel Armstrong, Jeanne Fredrickson, Evgenia Pak, Kristof Tigyi, Milinn Kremitzki, Christopher Markovic, Valerie Maduro, Amalia Dutra, Gerard G. Bouffard, Alexander M. Chang, Nancy F. Hansen, Amy B. Wilfert, Françoise Thibaud-Nissen, Anthony D. Schmitt, Jon-Matthew Belton, Siddarth Selvaraj, Megan Y. Dennis, Daniela C. Soto, Ruta Sahasrabudhe, Gulhan Kaya, Josh Quick, Nicholas J. Loman, Nadine Holmes, Matthew Loose, Urvashi Surti, Rosa ana Risques, Tina A. Graves Lindsay, Robert Fulton, Ira Hall, Benedict Paten, Kerstin Howe, Winston Timp, Alice Young, James C. Mullikin, Pavel A. Pevzner, Jennifer L. Gerton, Beth A. Sullivan, Evan E. Eichler & Adam M. Phillippy - Show fewer authors

Nature **585**, 79–84(2020) | Cite this article