

BIOL 343

Applied Bioinformatics I

Sequencing-by-synthesis (SBS)

Dr. Nic Wheeler

History of sequencing

A (relatively) new technology

- 1972 - first sequence of a complete gene elucidated (bacteriophage MS2 RNA)
 - 1977 - Sanger sequencing method invented
 - 1987 - Sanger sequencing automated by ABI
-
- 1996 - "Sequencing-by-synthesis" conception invented with pyrosequencing
 - 1998 - Sequencing-by-synthesis using fluorescent dyes by Solexa (later Illumina)
 - 2005 - Pyrosequencing automated by Roche into the 454 system
-
- 2010 - PacBio implements single-molecule, real-time sequencing
 - 2012 - Oxford Nanopore Technologies miniaturizes single-molecule sequencing



History of sequencing

A (relatively) new technology

Sanger sequencing

- Single fragment
- Cheap, fast turnaround
- Low throughput

Sequencing-by-synthesis (short)

- Millions to billions of fragments
- Complex library prep
- Cheap-ish
- Very high-throughput

Long-read sequencing

- Millions of long fragments
- Complex library prep
- High-throughput, expensive

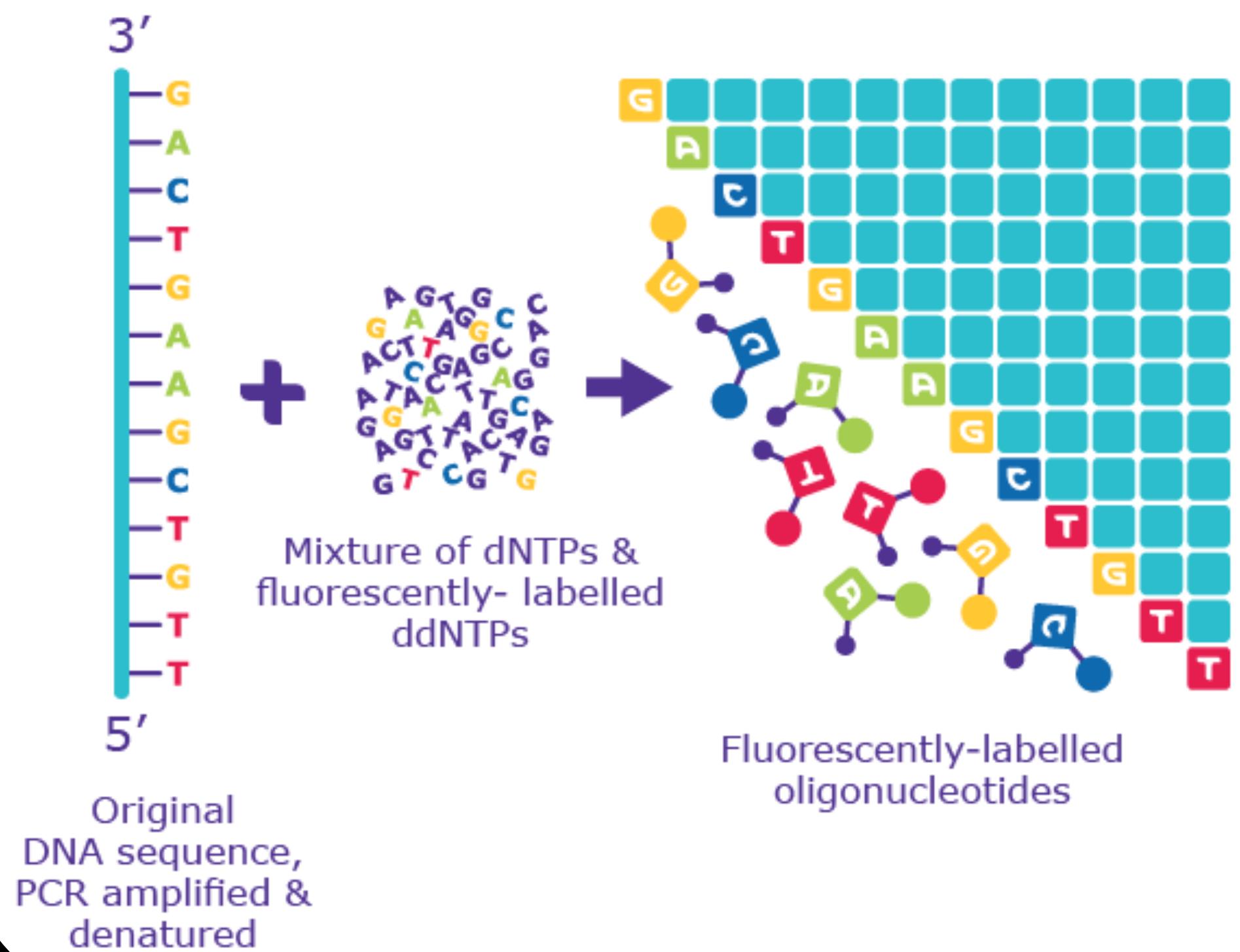


Principles of sequencing technologies

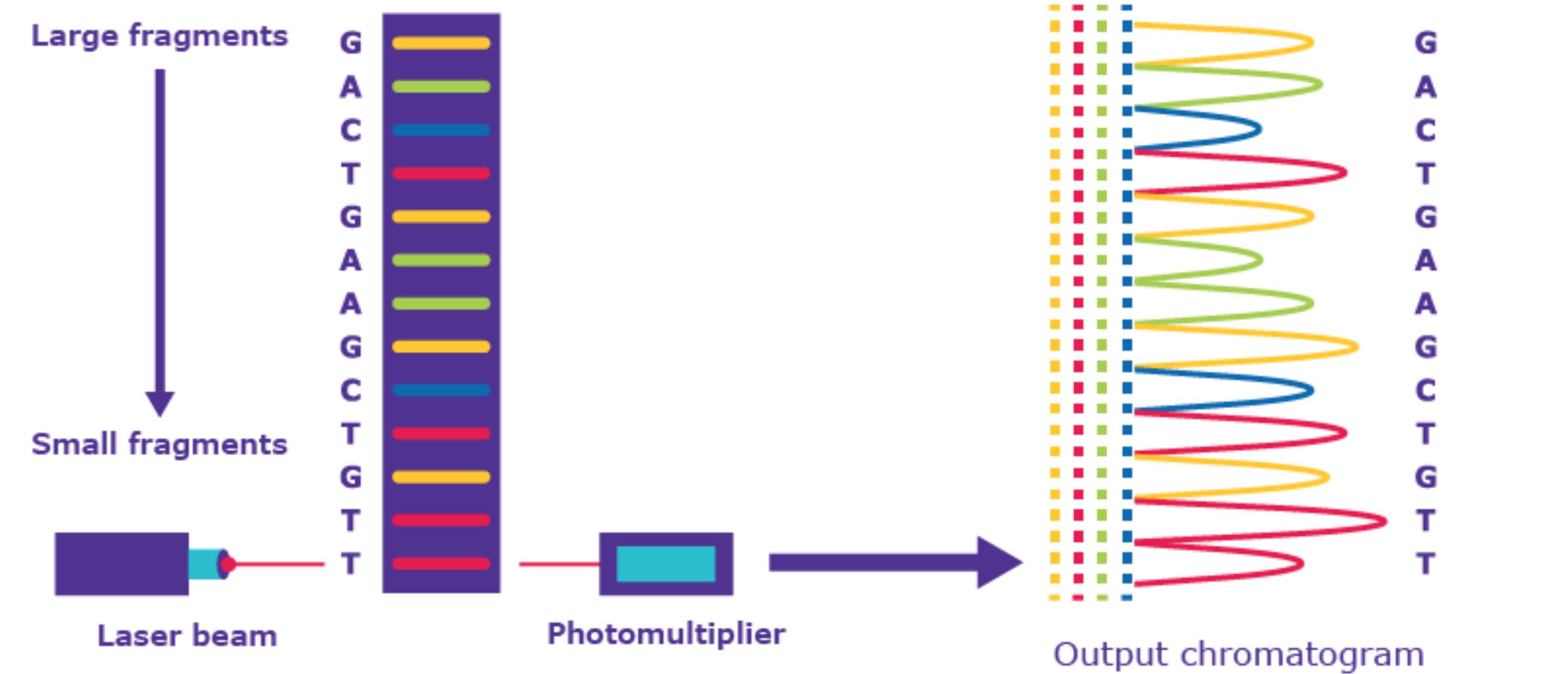
Principles of sequencing

First-generation (Sanger) sequencing

1 PCR with fluorescent, chain-terminating ddNTPs



2 Size separation by capillary gel electrophoresis



3 Laser excitation & detection by sequencing machine

Principles of sequencing

First-generation (Sanger) sequencing

Advantages

- Cheap
- Accurate
- Short preparation time
- Fast turnaround time
- No assembly required

Disadvantages

- Very low-throughput
- Max length of <1000 bp

Principles of sequencing

Next-generation: sequencing-by-synthesis

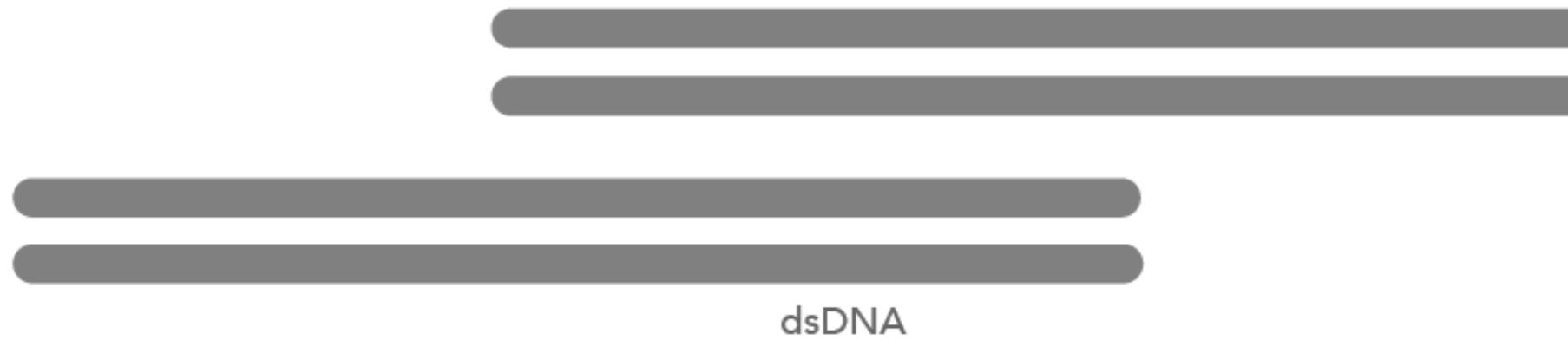
Four steps:

- Library preparation
- Cluster generation
- Sequencing-by-synthesis
- Mapping/analysis

Principles of sequencing

Library preparation

Fragmentation



End repair and A-tailing



Ligation



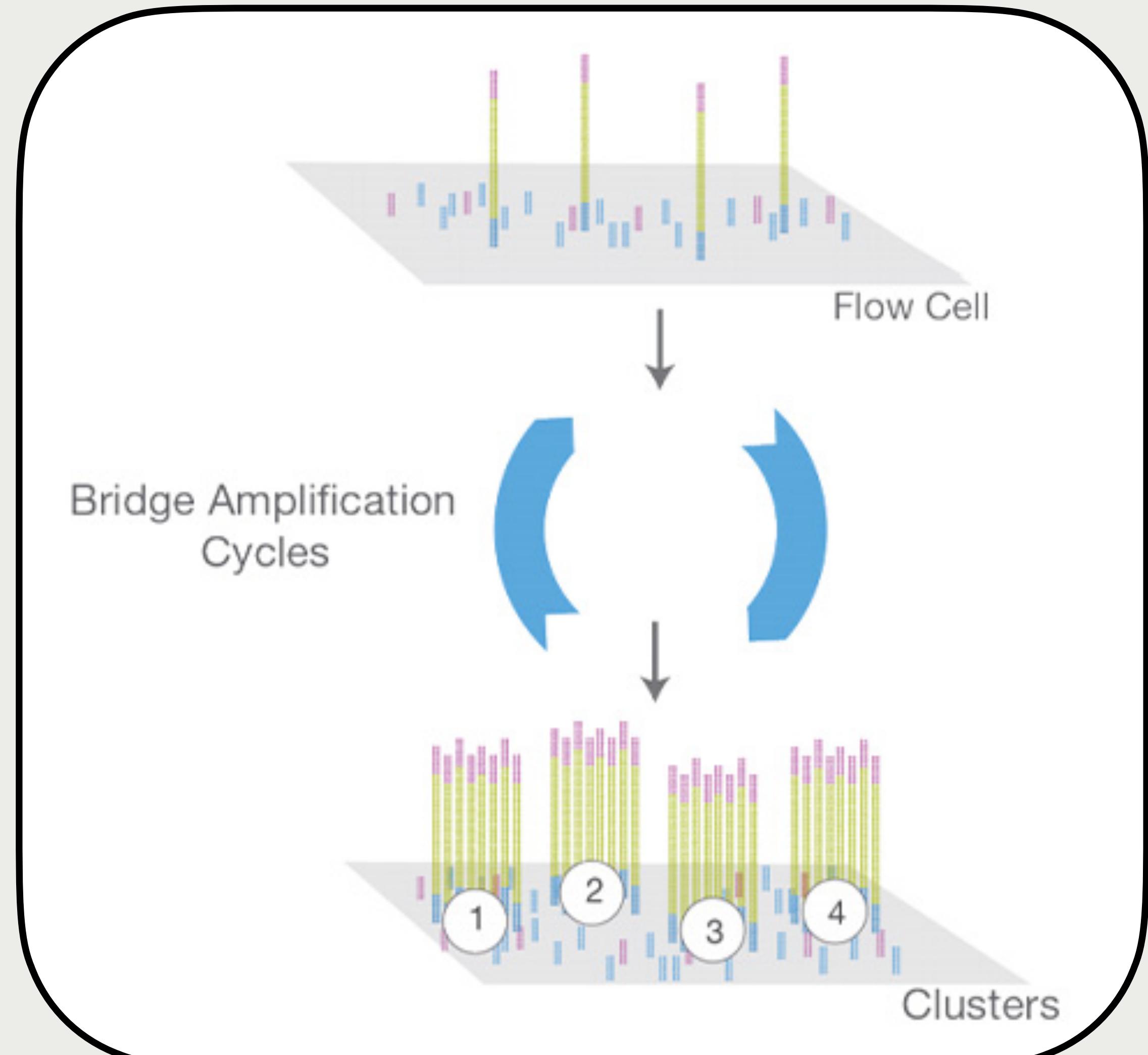
PCR amplification



Principles of sequencing

Cluster generation

- Major problem: fluorescent nucleotides aren't very bright
- Solution: add a fluorescent nucleotide at the same time to many identical fragments that are arranged in a cluster (~1000 clones)

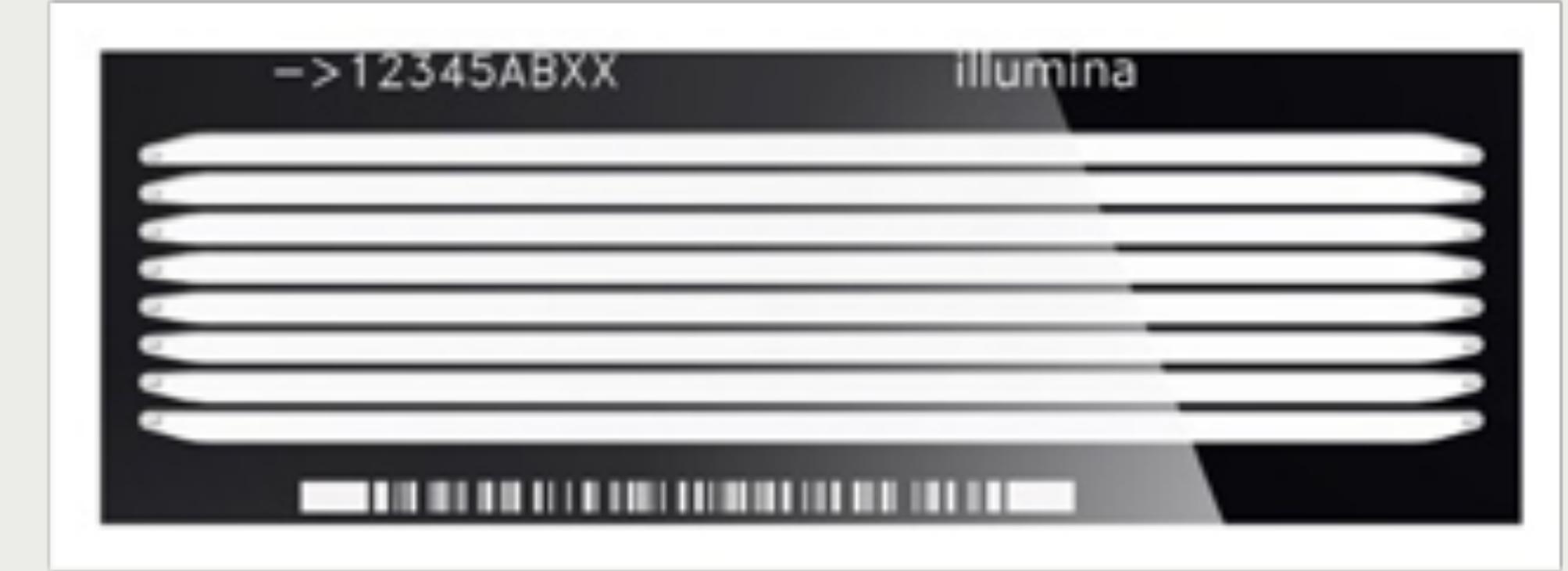


Principles of sequencing

Cluster generation

Cluster generation

1. Attach ssDNA to a solid support (flow cell)
 - ssDNA hybridizes to flow cell primers via P5/P7
 - Flow cell doesn't bind dNTPs
 - Ideally 1 ssDNA/nanowell
2. Isothermal bridge amplification to generate cluster
3. Cleave reverse direction strands

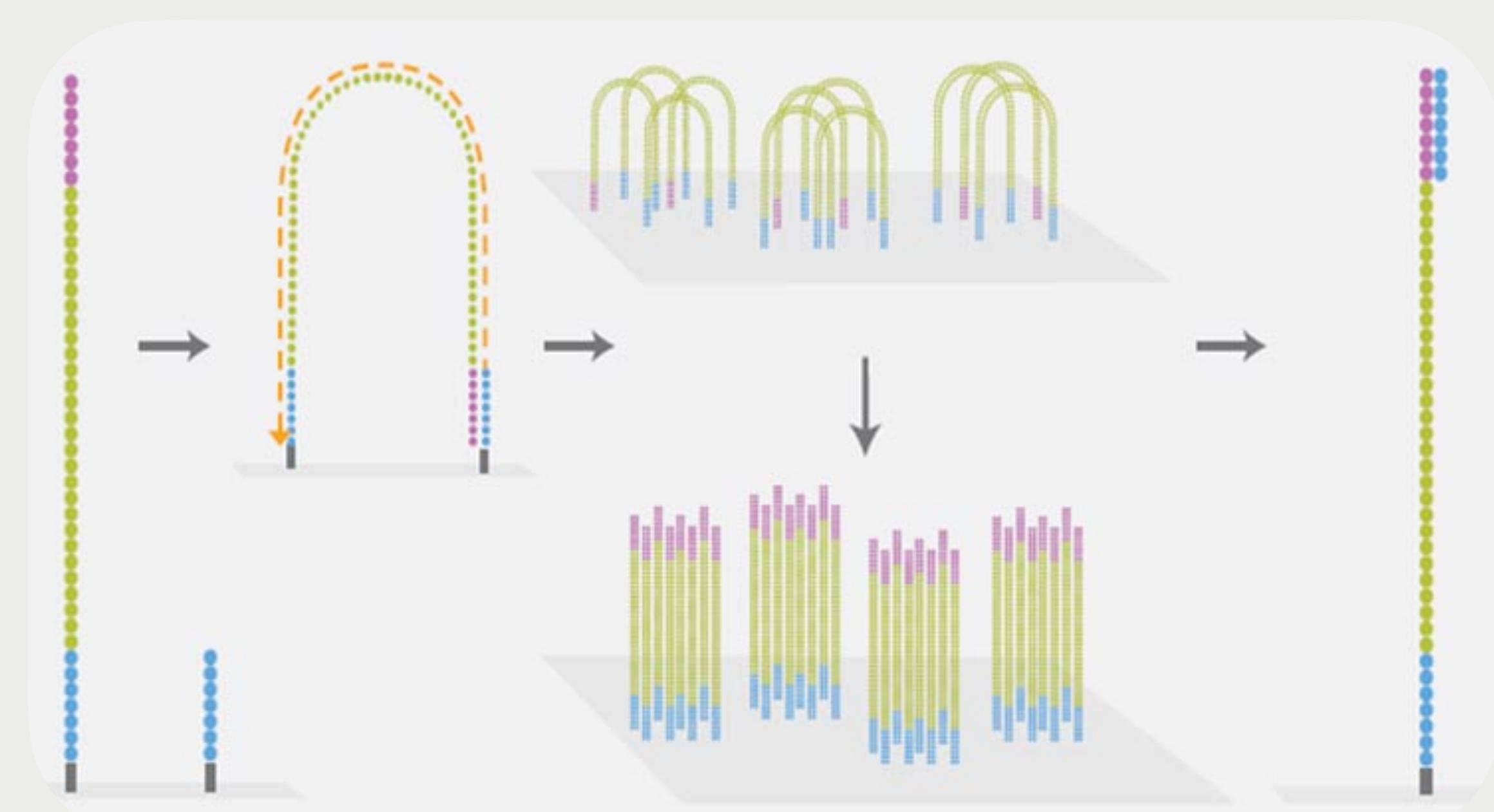


Principles of sequencing

Cluster generation

Cluster generation

1. Attach ssDNA to a solid support (flow cell)
 - ssDNA hybridizes to flow cell primers via P5/P7
 - Flow cell doesn't bind dNTPs
 - Ideally 1 ssDNA/nanowell
2. Isothermal bridge amplification to generate cluster
3. Cleave reverse direction strands

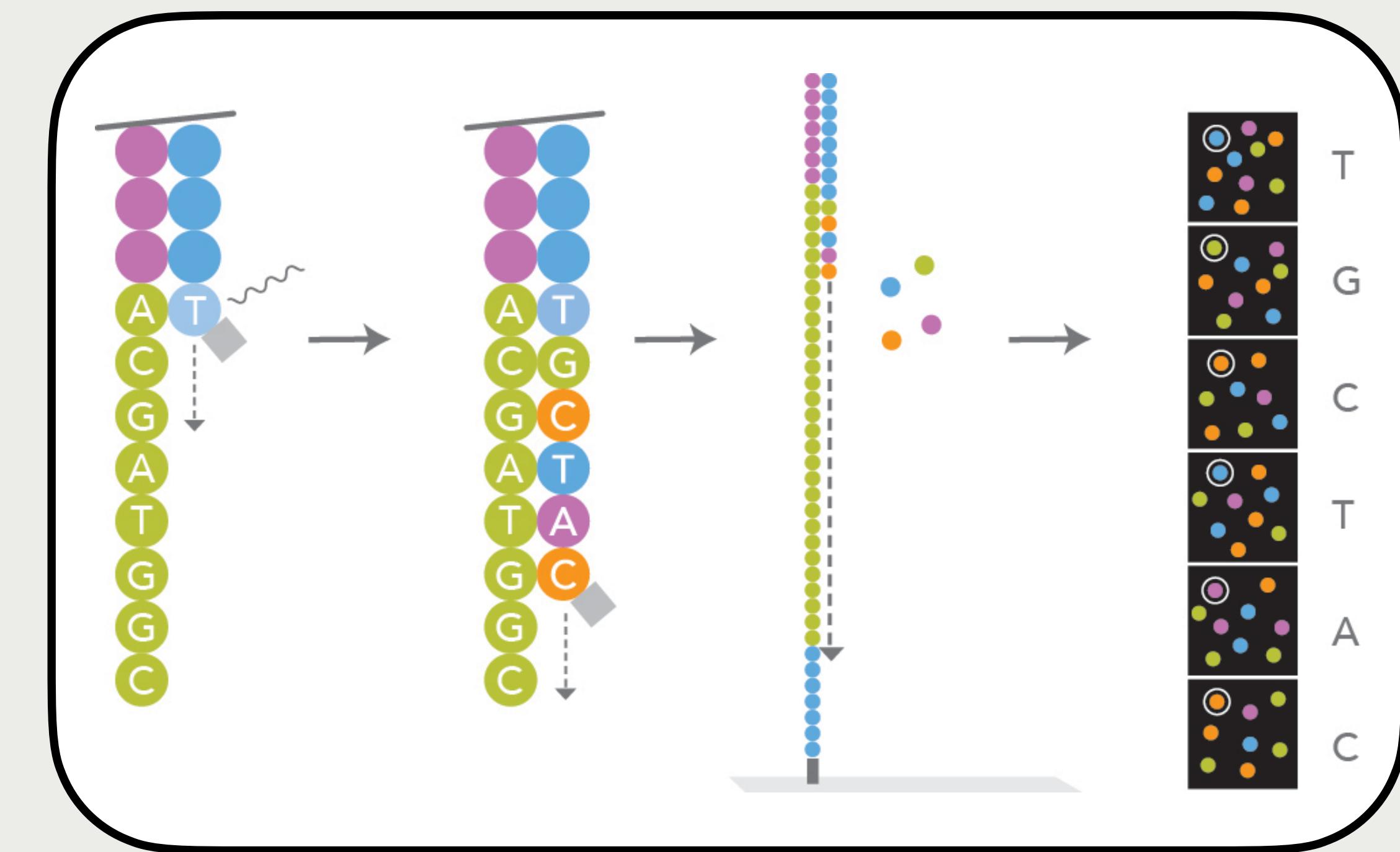


Principles of sequencing

Sequencing-by-synthesis

SBS

1. Add fl-NTPs
 1. 4, 2, or 1 color
 2. Reversible terminators
2. Run one “cycle” (polymerize one nucleotide)
3. Pause
4. Take 4 images (one laser, 4 filters)
5. Cleave the fluorescent label from the NTP
6. Repeat once per bp in read



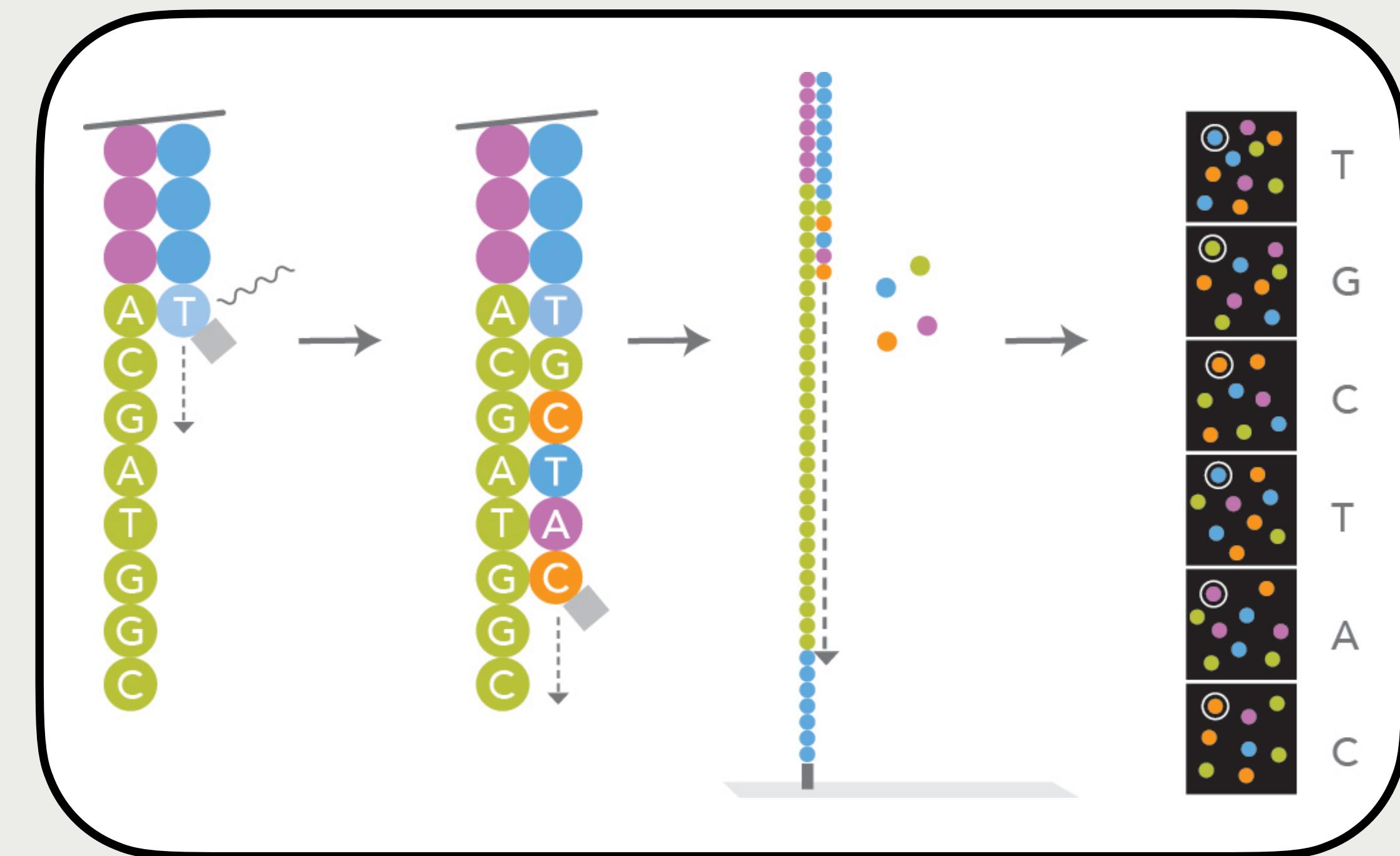
Incorporate, excite, image, and cleave

Principles of sequencing

Sequencing-by-synthesis

Paired-end sequencing

1. Run through all cycles for one end
2. Denature/remove everything
3. Bridge amplify (x1)
4. Sequence read 2, which will be the reverse-complement of read 1

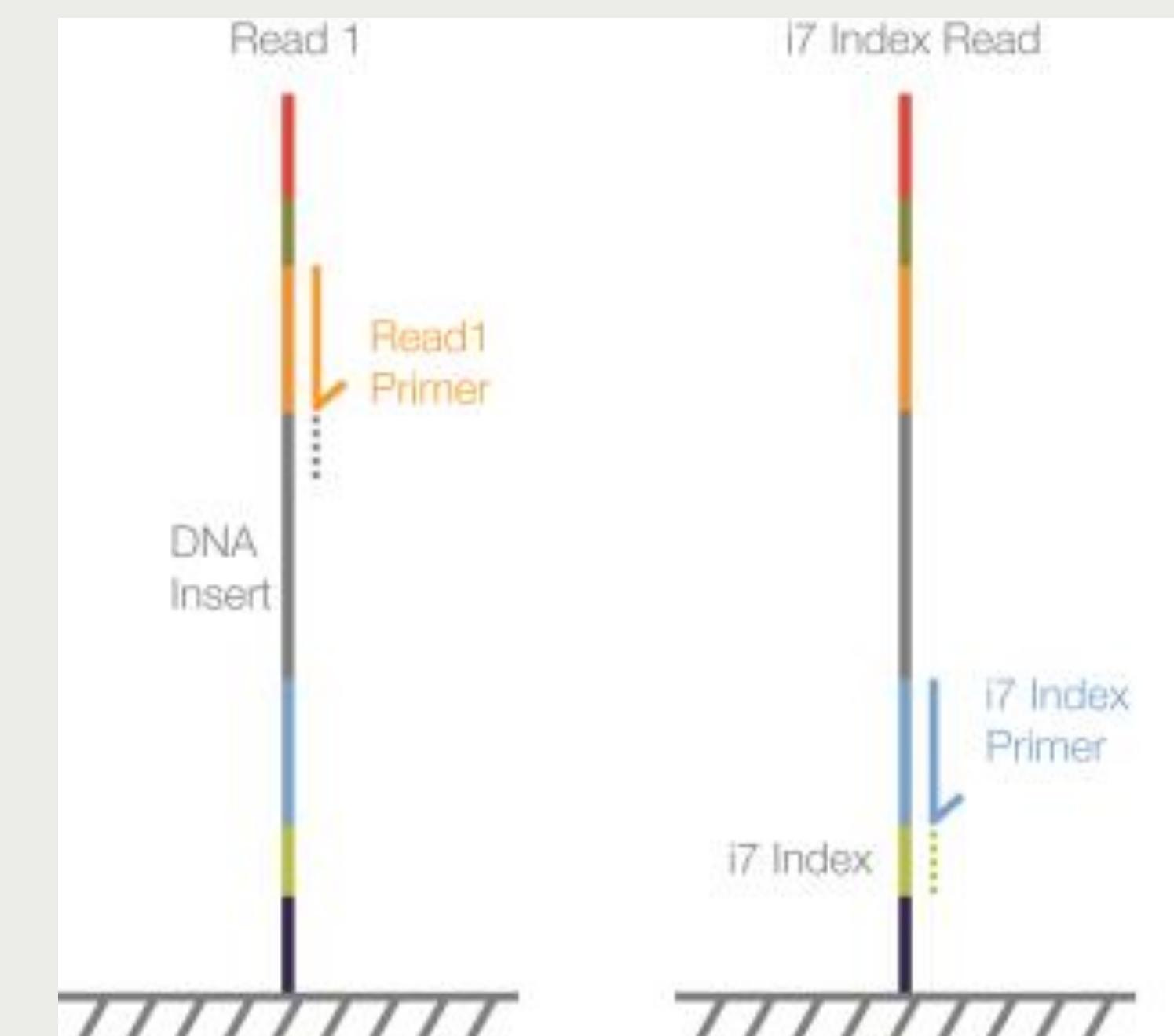


Principles of sequencing

Sequencing-by-synthesis

Index sequencing

1. After completing read 1, add the index primer
 - This i7 primer binds to the adaptor sequence
2. SBS of the index (barcode)
 - Cycle number depends on the length of the index (3-8 nt)



Principles of sequencing

Next-generation sequencing

Advantages

- Cheap
- Very accurate
- Very high-throughput
- Many software options
for analysis

Disadvantages

- Short read length (max 300 bp)
- Long library preparation
protocol
- Analysis can be difficult

illumina

Principles of sequencing

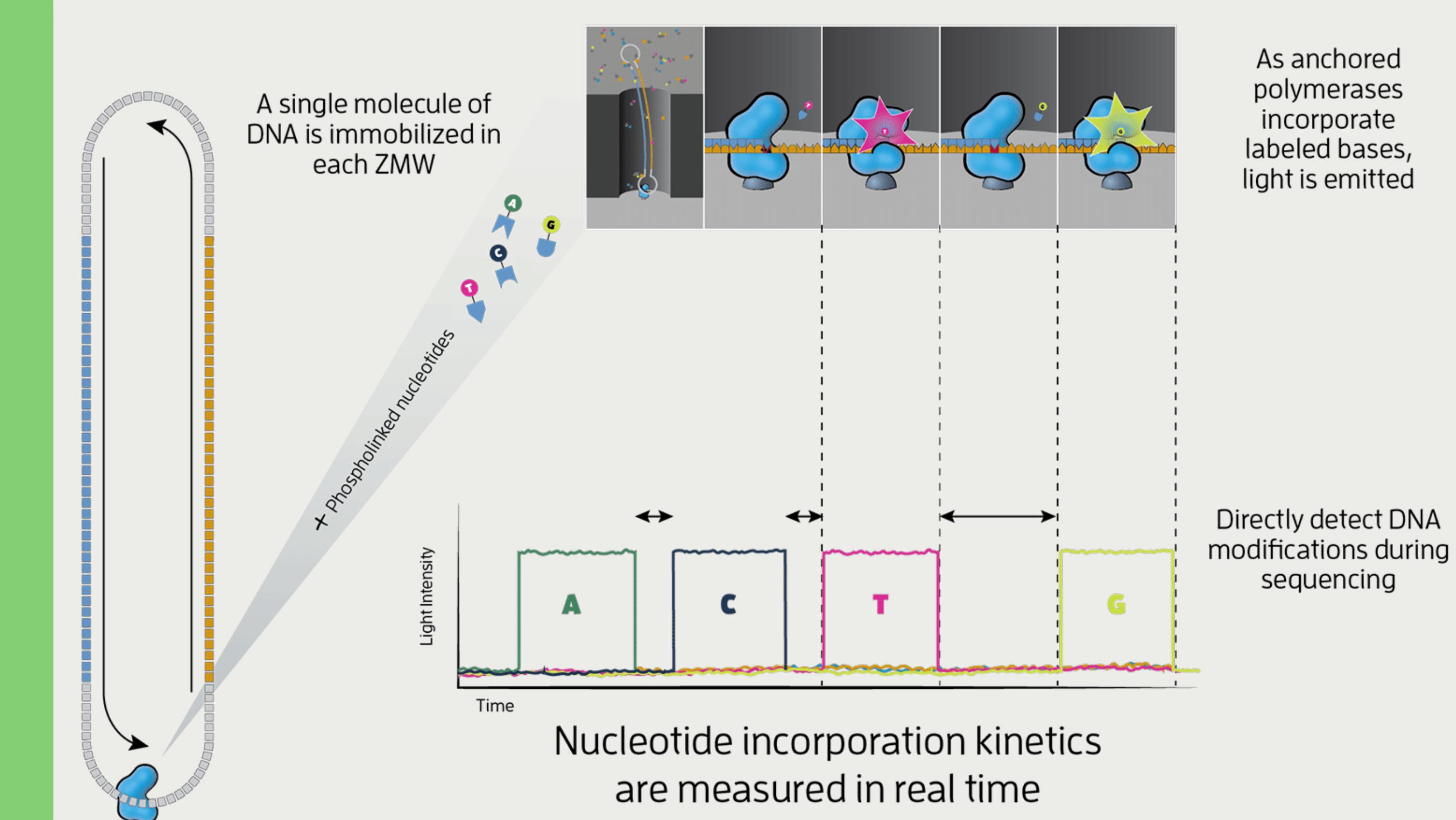
Third-generation: single-molecule, real-time

- Sequences one molecule of DNA at a time
- Requires the isolation of a single DNA polymerase with a single template strand of DNA
- Highly sensitive fluorescent or current detectors

Principles of sequencing

Third-generation: single-molecule, real-time

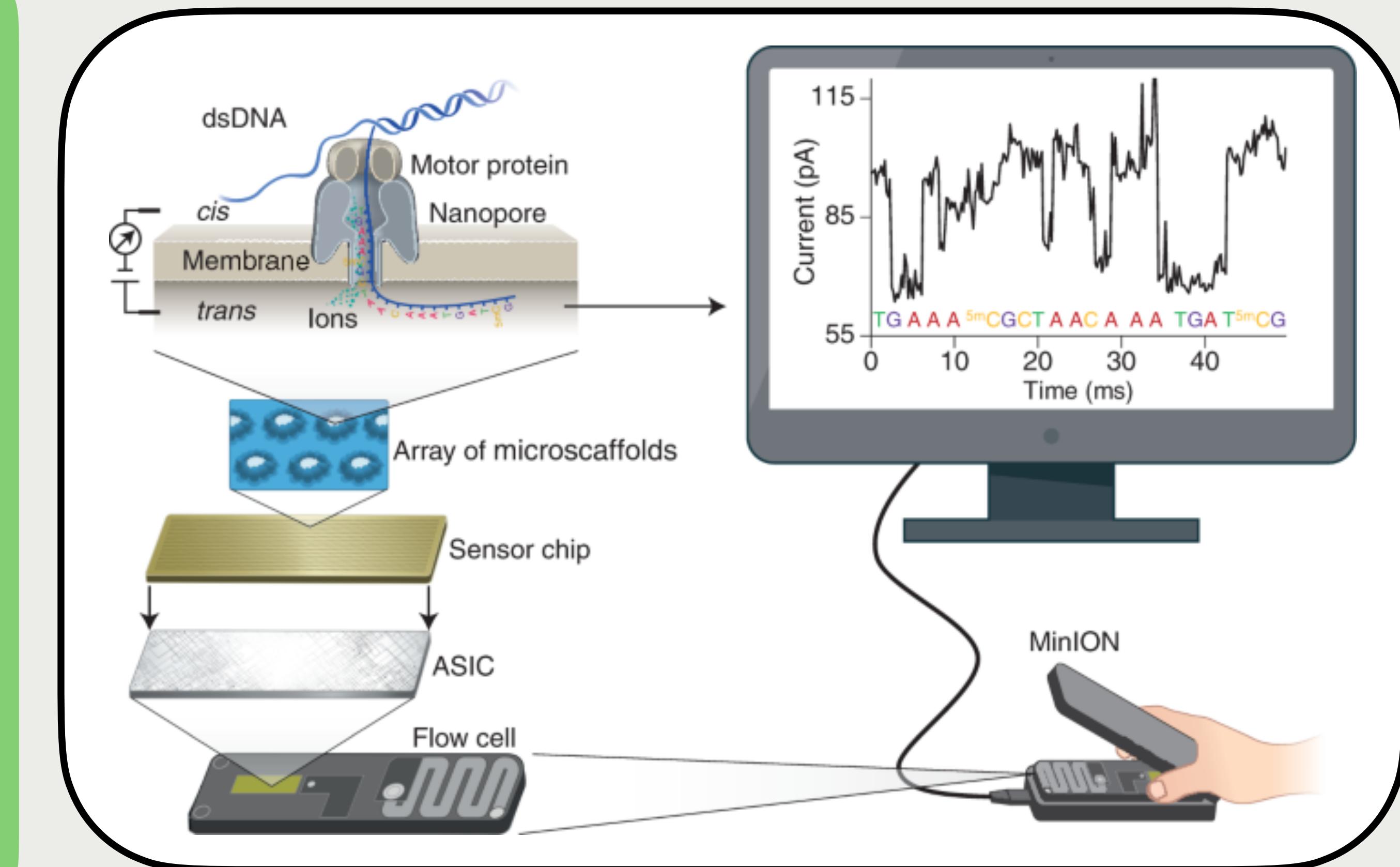
- Two primary technologies:
 - Pacific Biosciences (PacBio) - isolate DNA + polymerase in “zero-mode waveguide” (very small hole)
 - Oxford Nanopore Technologies (ONT) - push DNA through nanopores and detects ionic currents



Principles of sequencing

Third-generation: single-molecule, real-time

- Two primary technologies:
 - Pacific Biosciences (PacBio) - isolate DNA + polymerase in "zero-mode waveguide" (very small hole)
 - Oxford Nanopore Technologies (ONT) - push DNA through nanopores and detects ionic currents



Principles of sequencing

Third-generation: single-molecule, real-time

Advantages

- Single-molecule
- Very long read length

Disadvantages

- More expensive
- Long library prep protocol
- Requires high-quality template
- More error-prone
- Fewer tools for analysis