

CHAPTER 3

Recombinant DNA Technology and DNA Cloning

PART 4 of 4

DNA Sequencing

Chain termination sequencing (Sanger method)

Reaction mixture contains:

1. Target DNA to be sequenced
2. Nucleotides (dATP, dCTP, dGTP, dTTP)
3. Nucleotides (ddATP and/OR ddCTP, ddGTP, ddTTP)
4. Buffer
5. DNA polymerase
6. Primer

DNA Sequencing

Original Sanger method

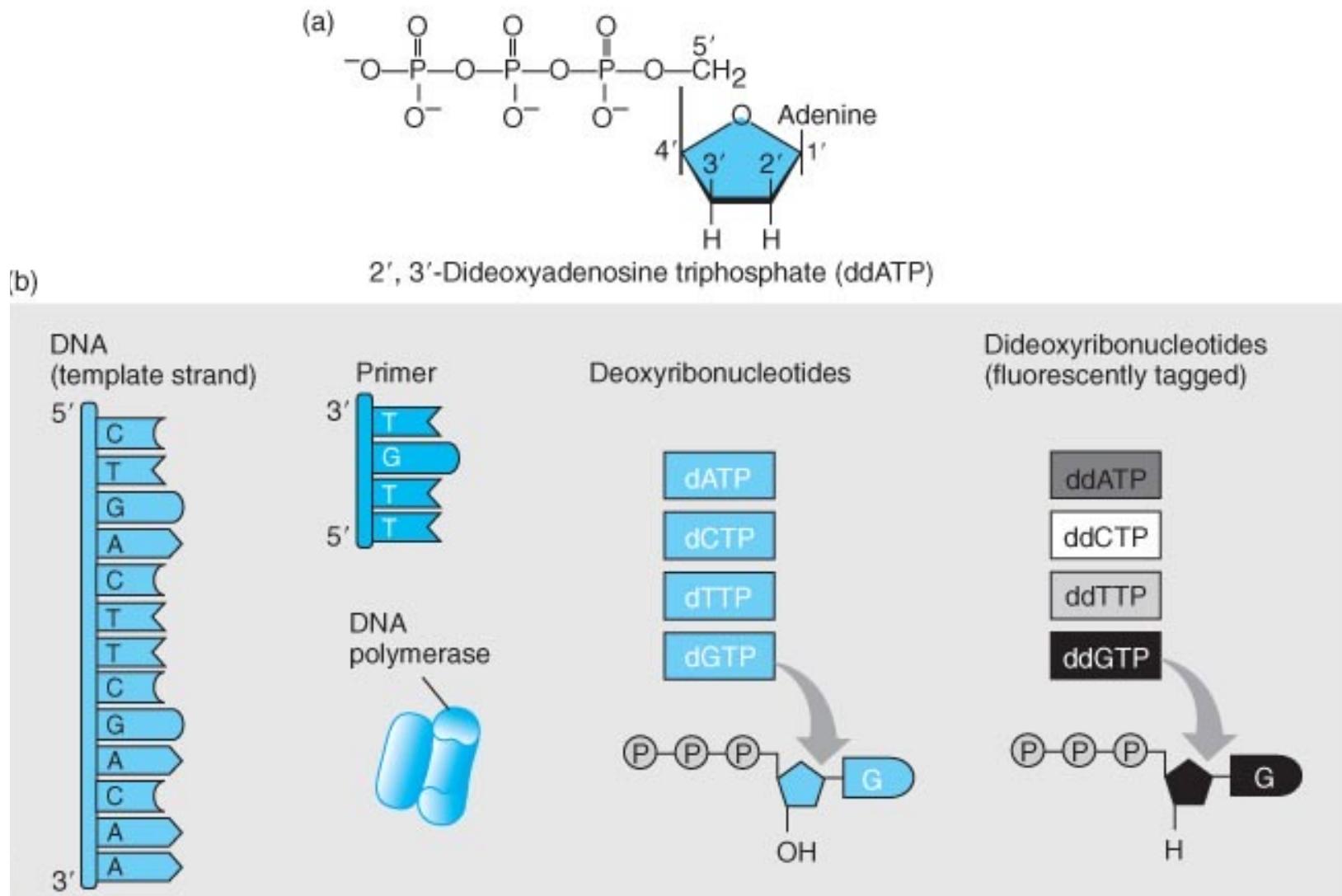
Four separate reaction tubes and each contained the template; primer; dNTPs in which one dNTP is radioactively labeled; and a different small amount of ddNTP; DNA Pol

Over time a ddNTP will be incorporated into all the positions in the newly synthesized strands creating fragments of varying lengths that are terminated at the ddNTP

Then the fragments are separated on polyacryamide gel

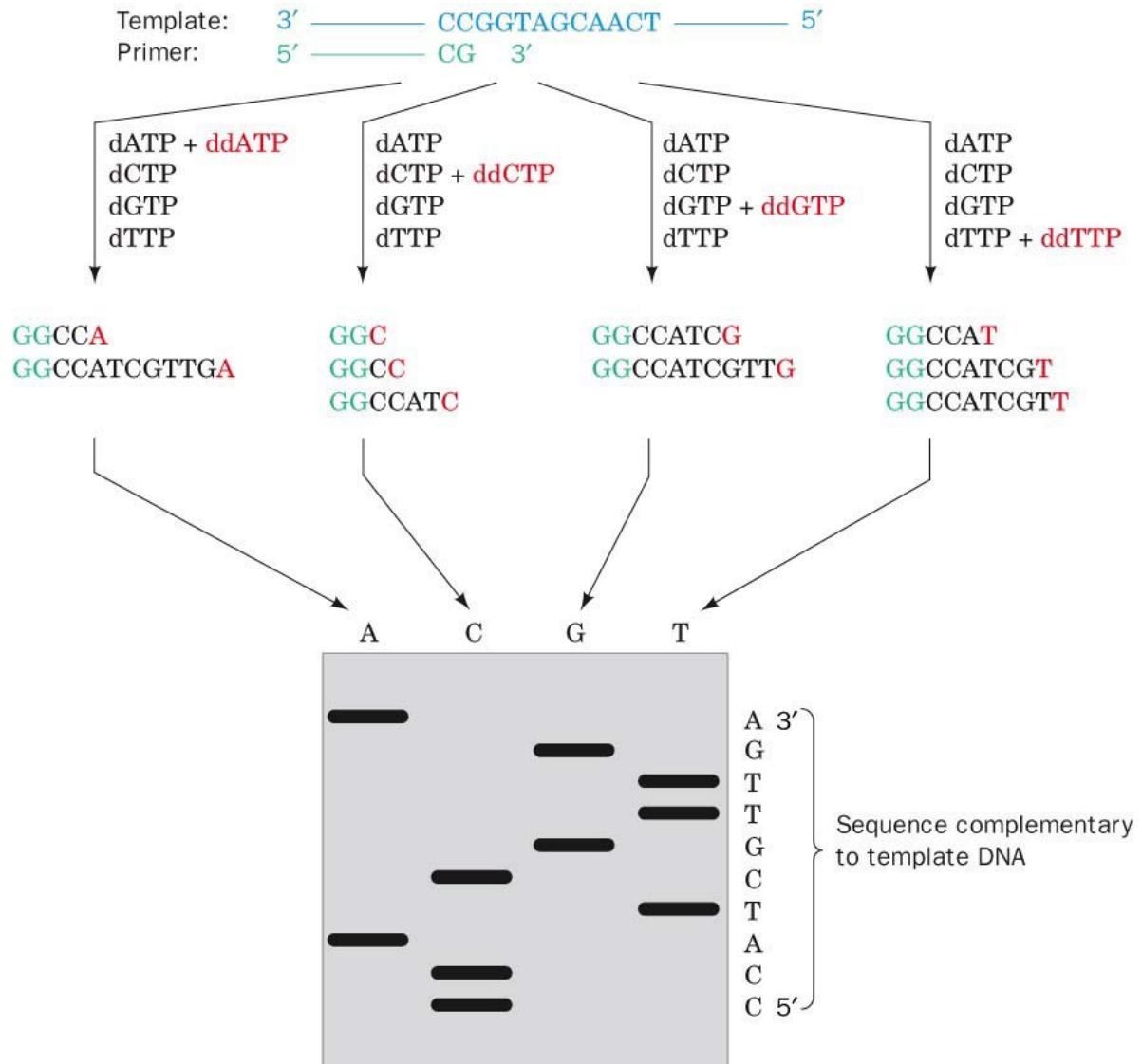
Autoradiography used to then identify radioactive fragments

DNA Sequencing



DNA Sequencing

Original Sanger method



Modern Sanger DNA Sequencing

High throughput computer automated sequencing using Sanger method

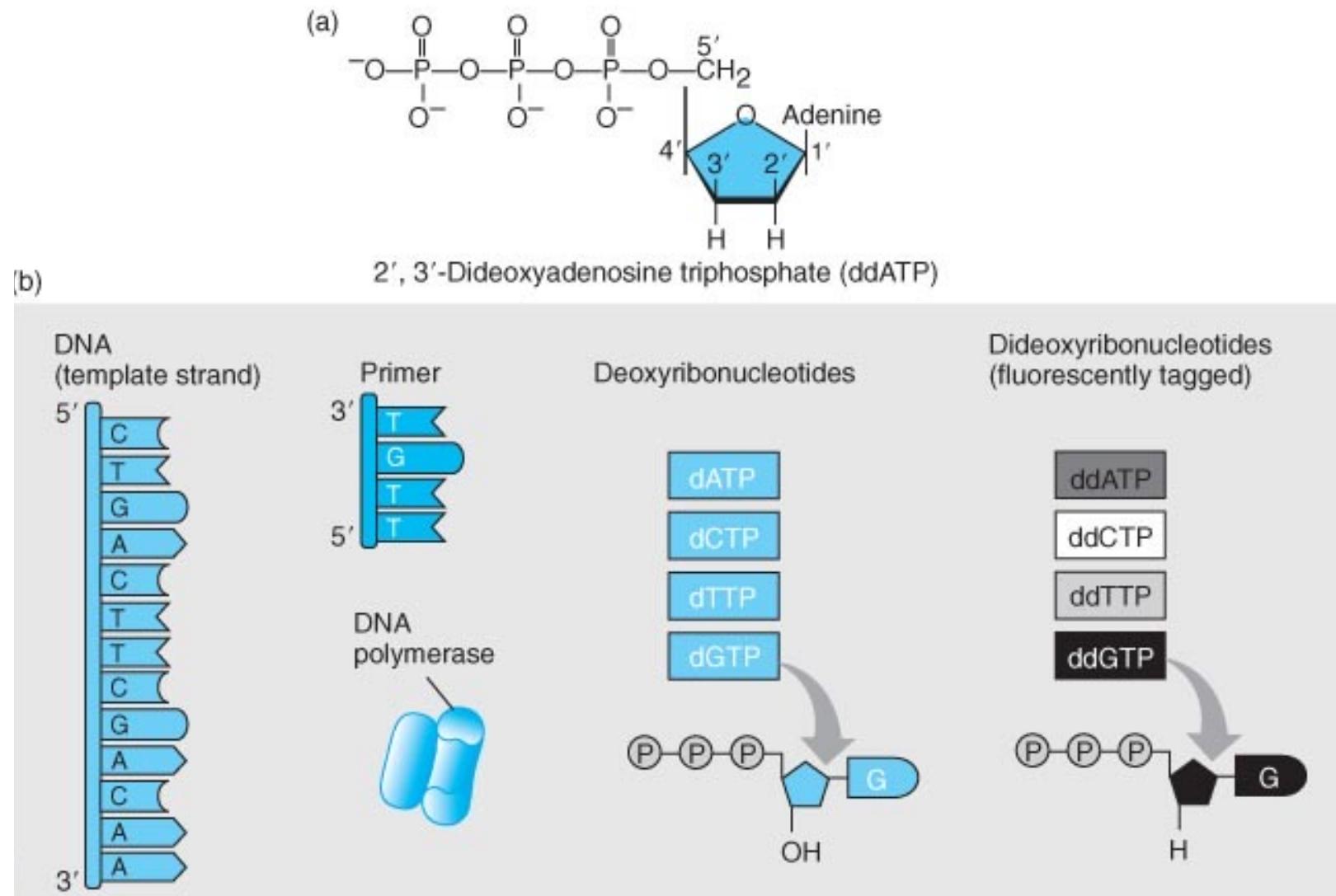
Was very helpful for completing human genome project

Samples are separated on a single-lane capillary gel that is scanned with a laser beam

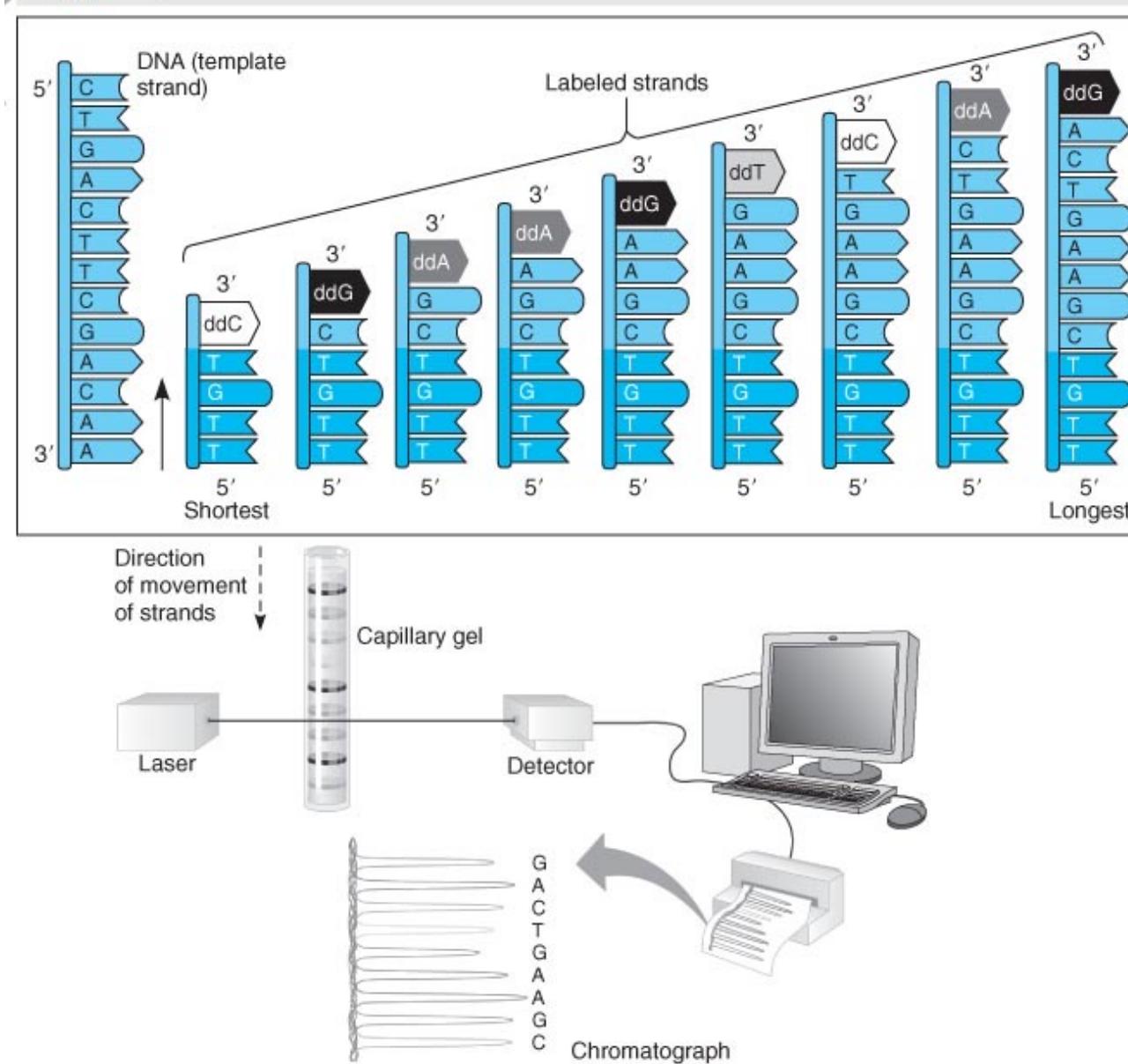
Laser stimulates fluorescent dye on each DNA fragment which emits a different wavelength of light for each different colored ddNTP

Emitted light is collected by a detector that amplifies and feeds this info. to a computer that can run multiple capillary gels at one time = 900 bp sequence

DNA Sequencing

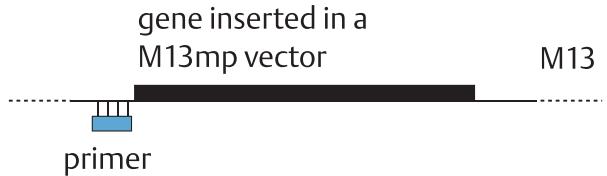


DNA Sequencing



DNA sequencing after Sanger and Coulson

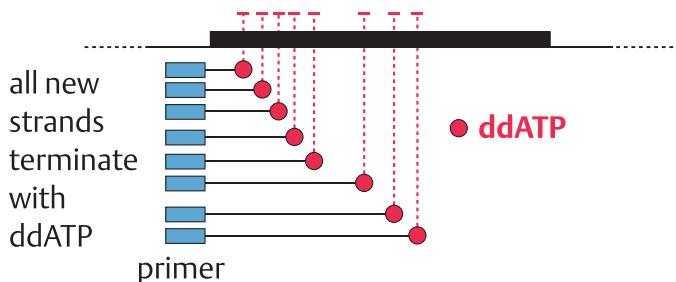
a annealing of the primer



DNA polymerase
dATP, dTTP, dGTP, dCTP
 ^{32}P - or ^{35}S for auto-radiography

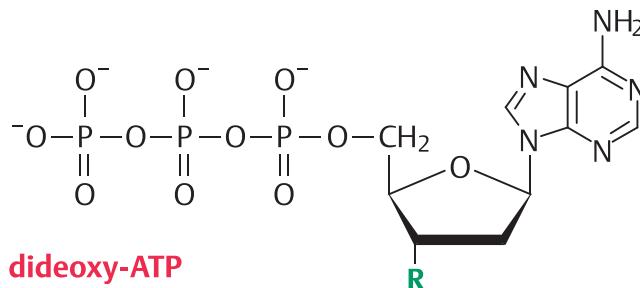
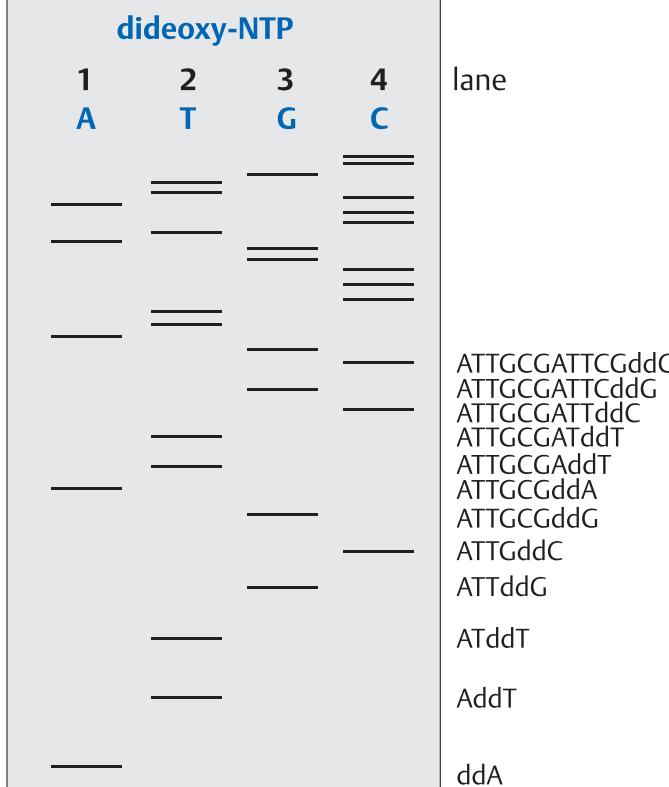
dideoxy-ATP (ddATP)

b example: synthesis of strands in lane 1 addition of dideoxy-ATP



ddTTP has been added in lane 2,
ddGTP in lane 3, and ddCTP in lane 4.

c gel electrophoresis and autoradiography



dideoxy-ATP

R = OH deoxy-ATP (dATP), normal nucleotide, strand synthesis

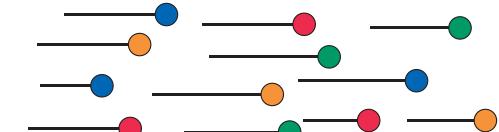
R = H dideoxy-ATP (ddATP), termination of strand elongation

High-throughput sequencing

DNA single strand by PCR

primer

DNA polymerase,
dATP, dTTP, dGTP, dCTP
ddATP ddTTP ddGTP ddCTP*



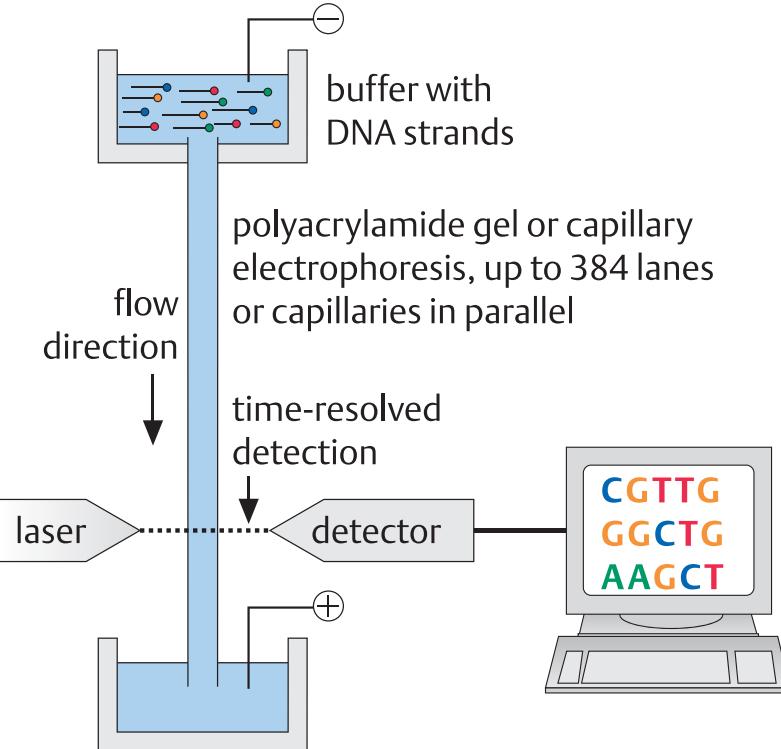
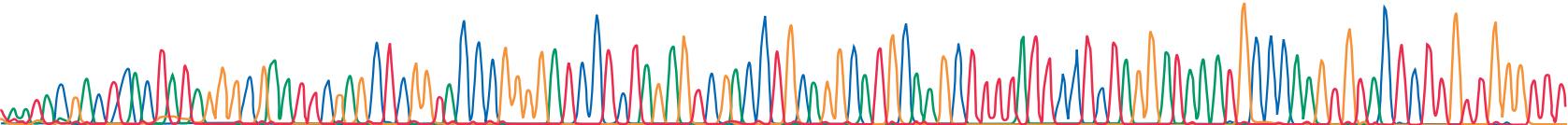
DNA strands of various chain lengths
with terminal fluorescence marker

separation by electrophoresis,
time-resolved detection

* using dideoxynucleotides with base-specific fluorescence marker for chain termination

time-resolved detector signal

AATAACGACTCAC TATAAGGGCGAATT CGAGCTCGGTAC CC GGGGAT CCT CTAGAGT CCAC CT CCAGGCAT GCAA GC TTTTATT CCTCTAGG ATAAATGC CCA AGTGTACT CTTGTT GG GT TT
10 20 30 40 50 60 70 80 90 100 110 120



Sequencing Techniques by principle

1. Chain Termination

- Example: Sanger sequencing

2. Sequencing by Synthesis

A) ***Single Nucleotide Addition (SNA)***

- Example: Roche 454 and Ion Torrent

B) ***Cyclic Reversible Termination (CRT)***

- Example: Illumina (Solexa) and Qiagen (GeneReader)

3. Sequencing by Ligation

- Example: Applied Biosystem (SOLiD) and Complete Genomics (BGI)

4. Single-Molecule Real-Time (SMRT)

- Example: Pacific Biosciences (PacBIO) and Oxford Nanopore

Sequencing by Synthesis (SBS): Roche 454

Produces highly accurate and long stretches of DNA sequence (greater than 1 gigabase) of DNA per reaction

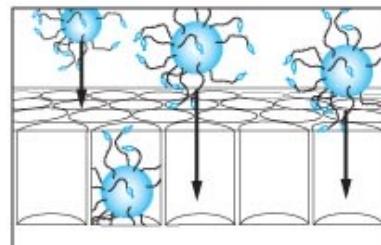
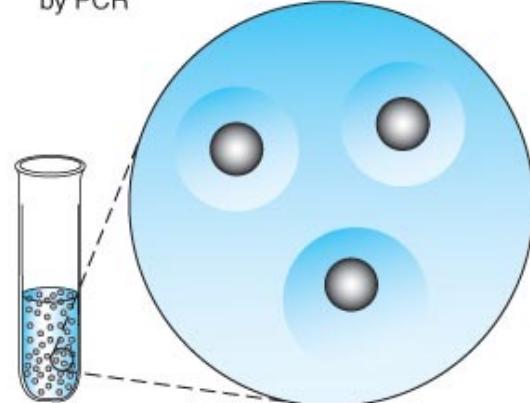
Use parallel formats that use fluorescence imaging techniques

Roche 454 commercial system using pyrosequencing

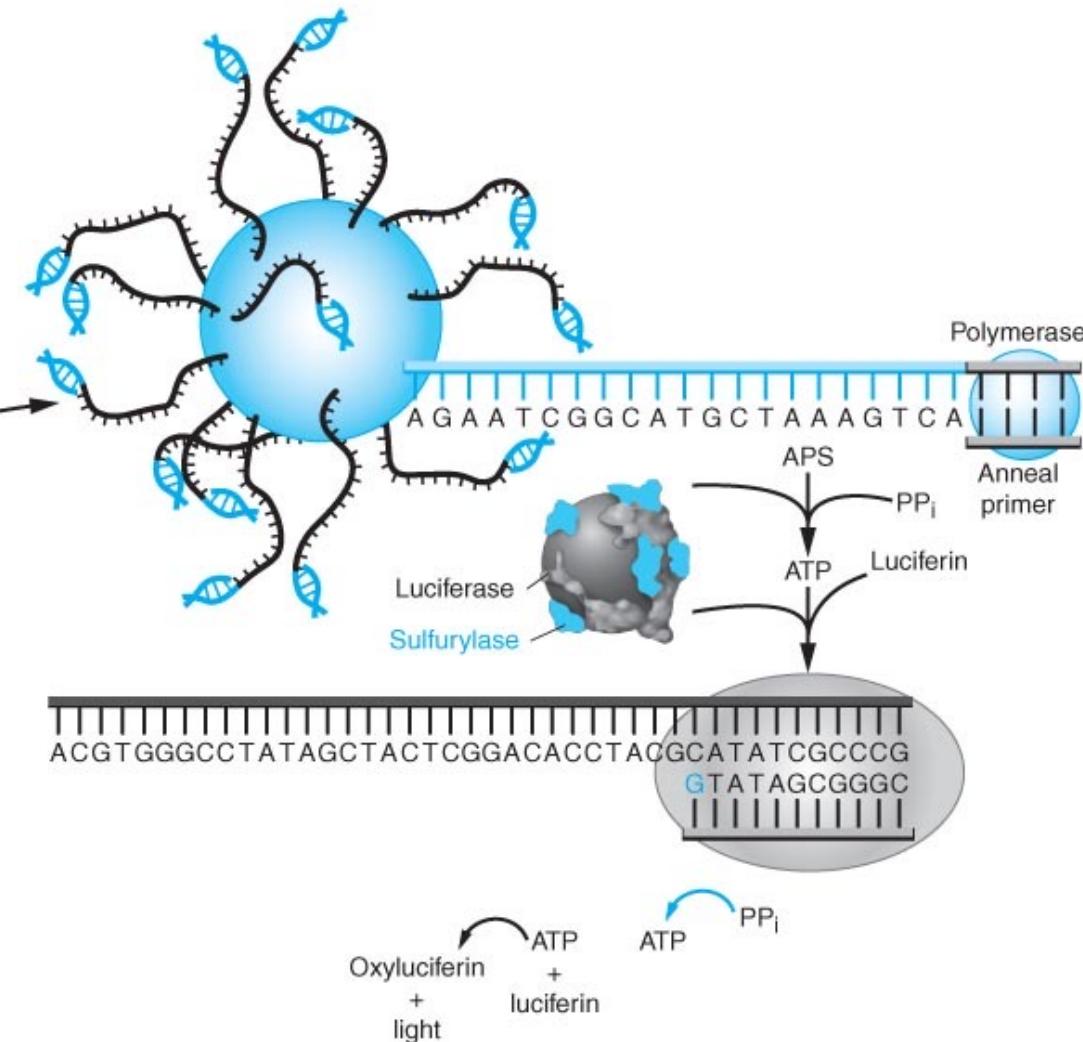
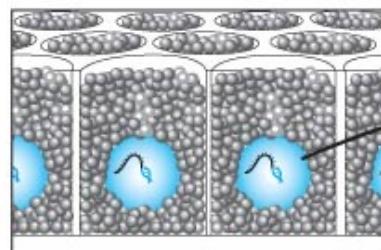
Pyrosequencing: beads are attached to fragmented genomic DNA which is PCR amplified in separate water droplets in oil for each bead then loaded into multi-well plates and mixed with DNA polymerase; then actual pyrosequencing

Roche 454 pyrosequencing

1. DNA fragments on beads amplified by PCR

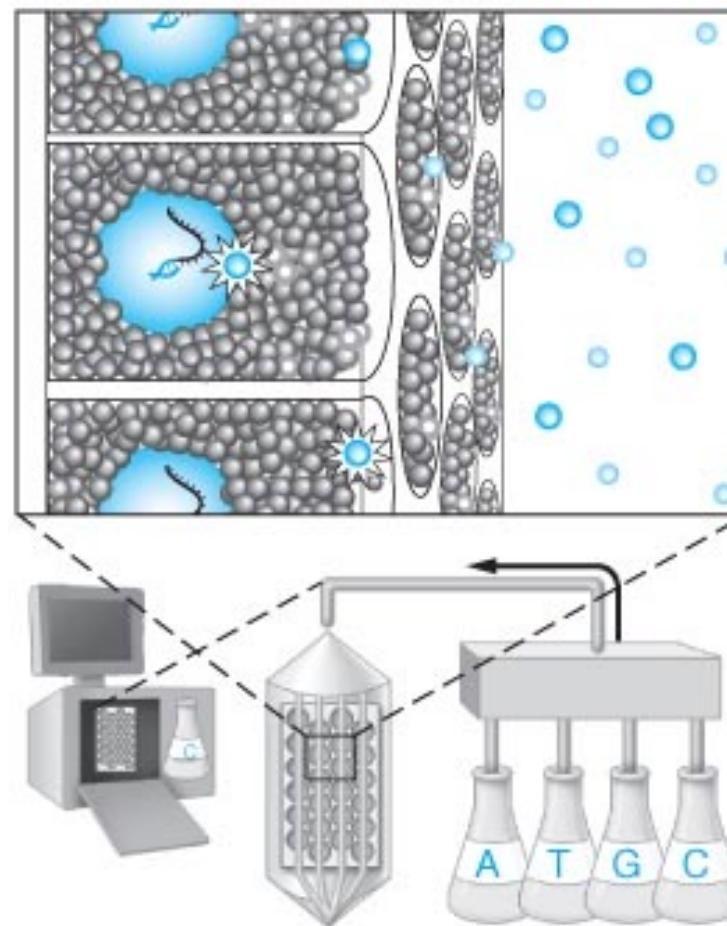


2. DNA on beads subjected to pyrosequencing reaction



Roche 454 pyrosequencing

3. Light emitted by pyrosequencing reactions captured by detector



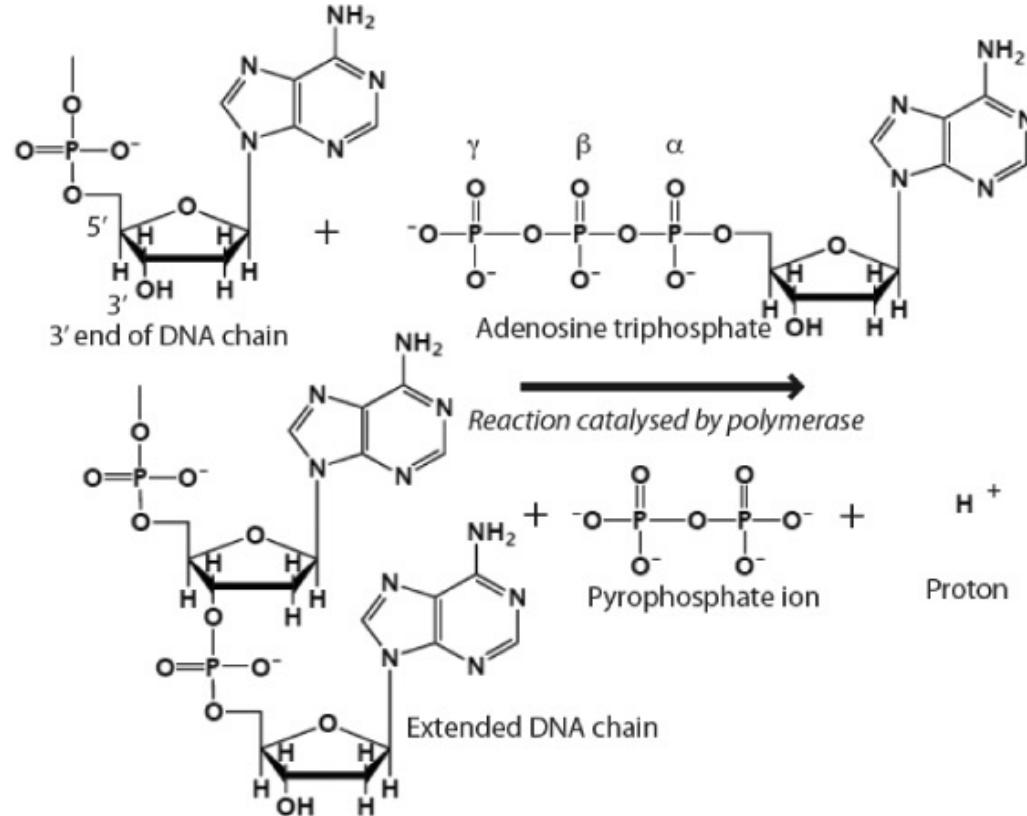
Sequencing by Synthesis (SNA): Ion semiconductor sequencing

This method is based on the detection of hydrogen ions that are released during the polymerization of DNA

A microwell containing a template DNA strand to be sequenced is flooded with a single type of nucleotide. If the introduced nucleotide is complementary to the leading template nucleotide it is incorporated into the growing complementary strand.

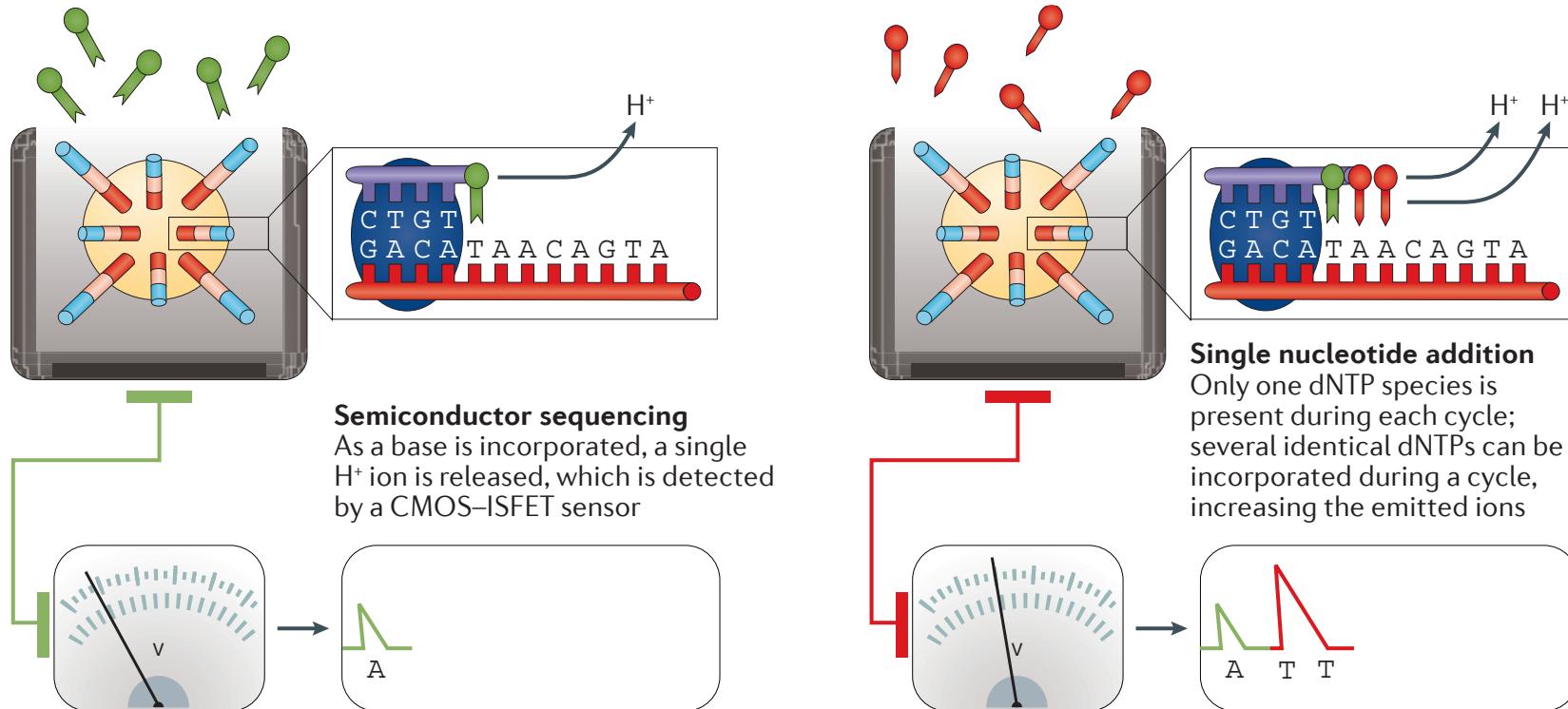
This causes the release of a hydrogen ion that triggers a hypersensitive ion sensor, which indicates that a reaction has occurred. If homopolymer repeats are present in the template sequence multiple nucleotides will be incorporated in a single cycle.

Ion semiconductor sequencing

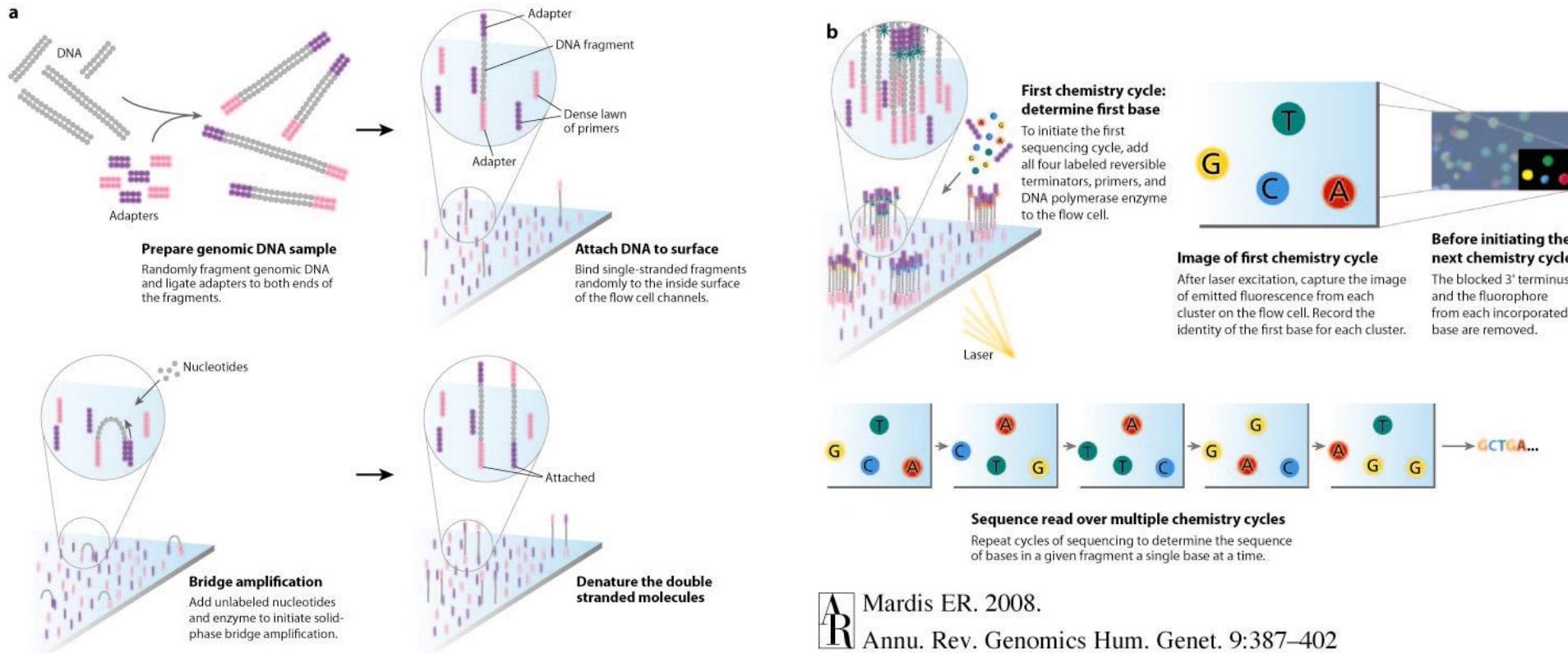


In nature, when a nucleotide is incorporated into a strand of DNA by a polymerase, a hydrogen ion is released as a byproduct.

Ion semiconductor sequencing



Sequencing by Synthesis (CRT): Illumina Solexa



Mardis ER. 2008.
Annu. Rev. Genomics Hum. Genet. 9:387–402

Sequencing by Synthesis (CRT): Illumina Solexa

This sequencing technology is based on reversible dye-terminators.

DNA molecules are first attached to primers on a slide and amplified so that local clonal colonies are formed (isothermal bridge amplification).

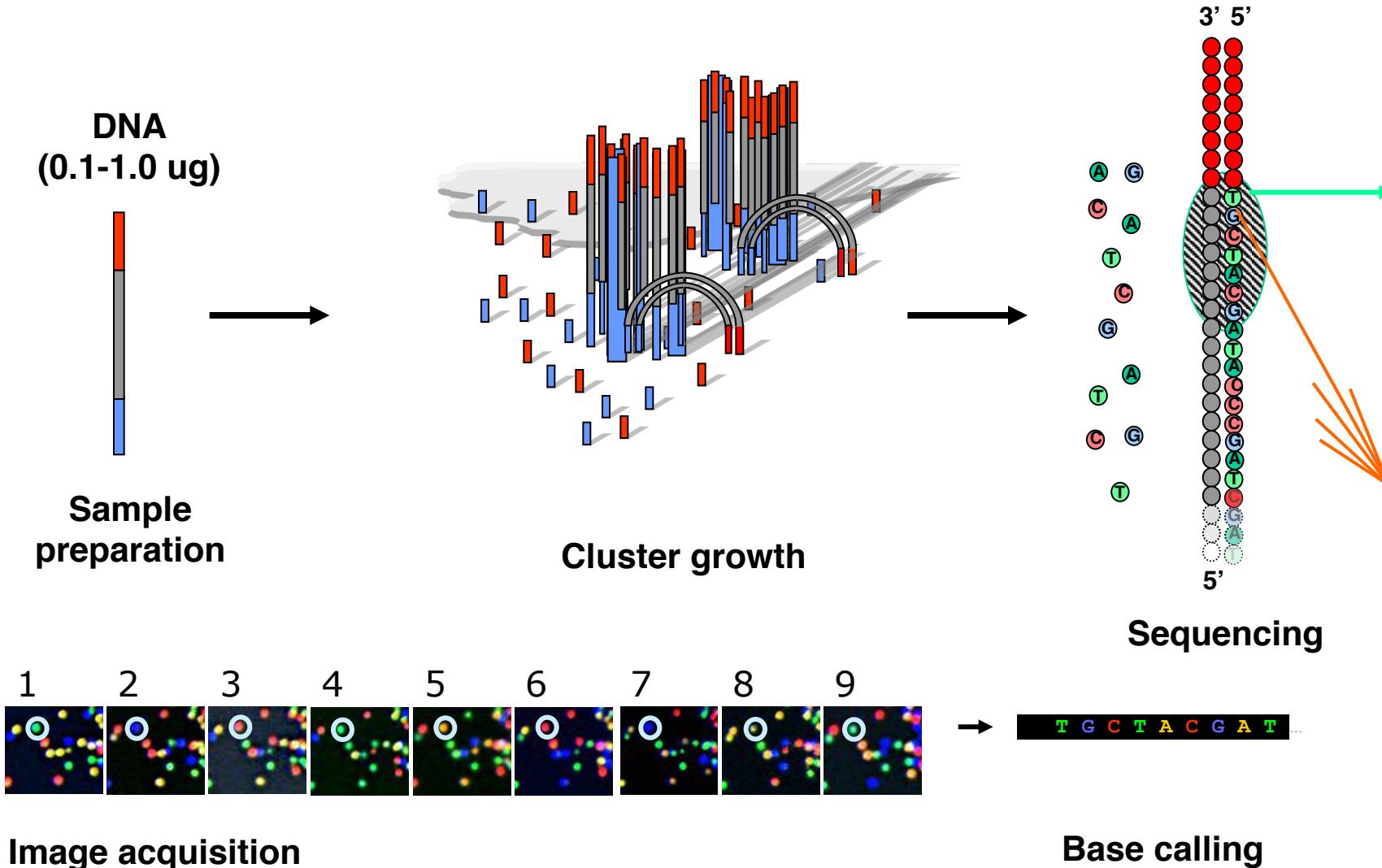
Four types of reversible terminator bases (RT-bases) are added, and non-incorporated nucleotides are washed away.

Unlike pyrosequencing, the DNA can only be extended one nucleotide at a time.

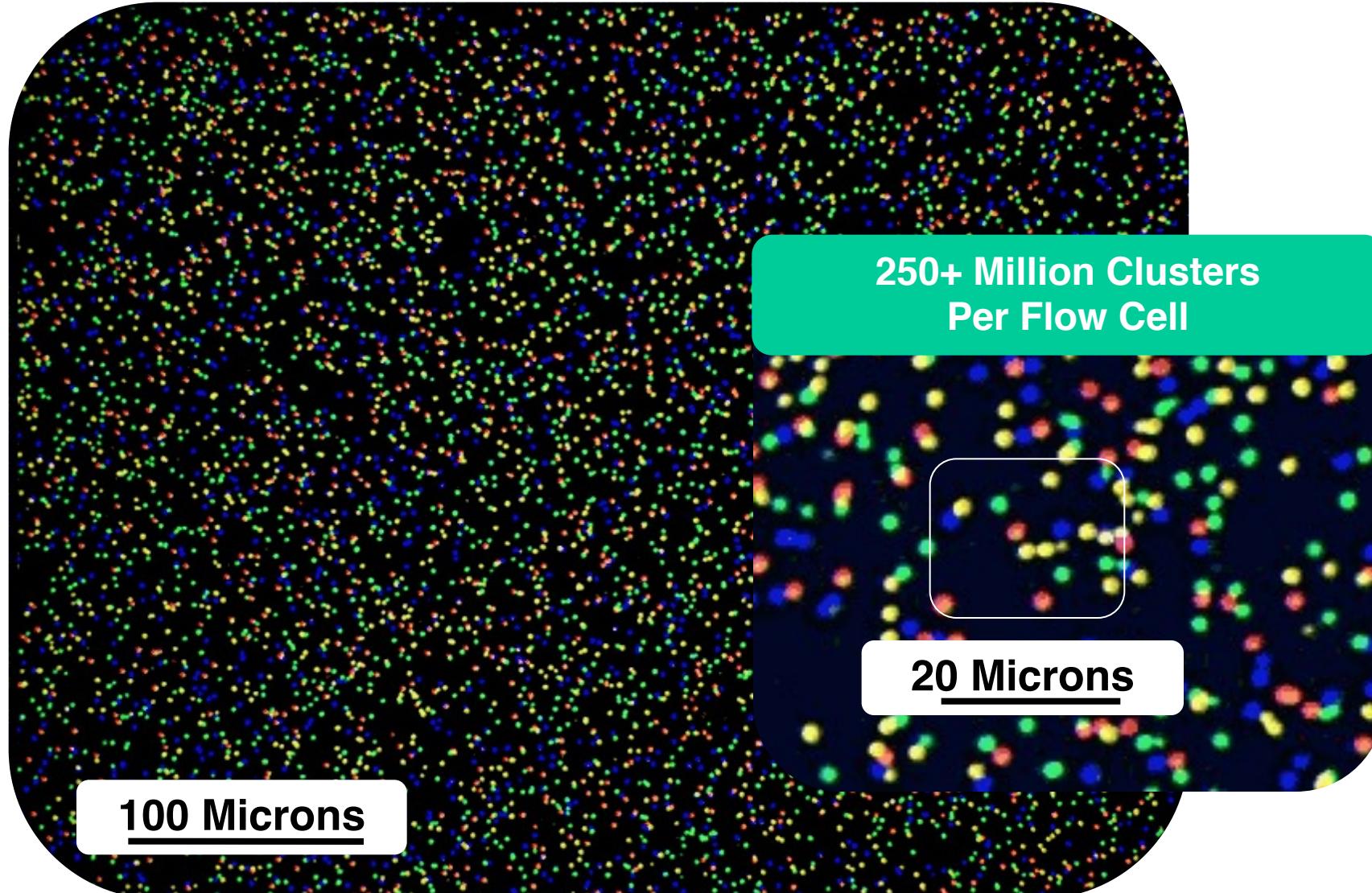
A camera takes images of the fluorescently labelled nucleotides, then the dye along with the terminal 3' blocker is chemically removed from the DNA, allowing the next cycle.

Illumina Sequencing Technology

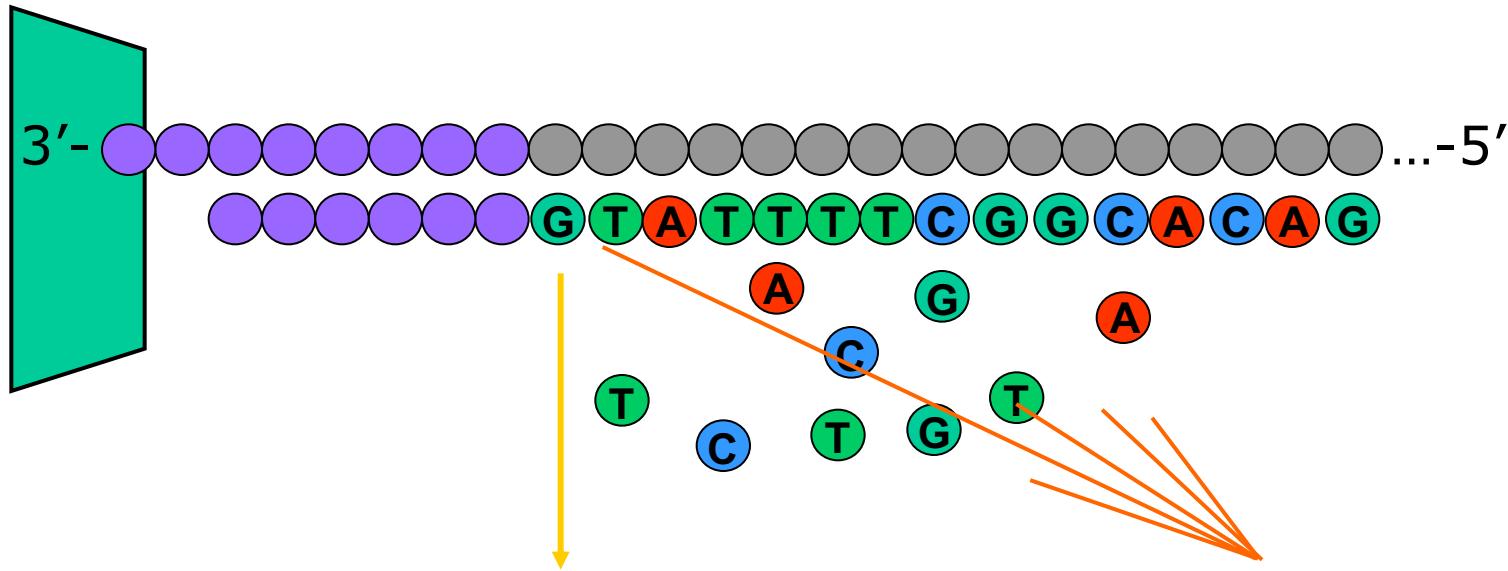
Robust Reversible Terminator Chemistry Foundation



Sequencing



Sequencing by Synthesis, One Base at a Time



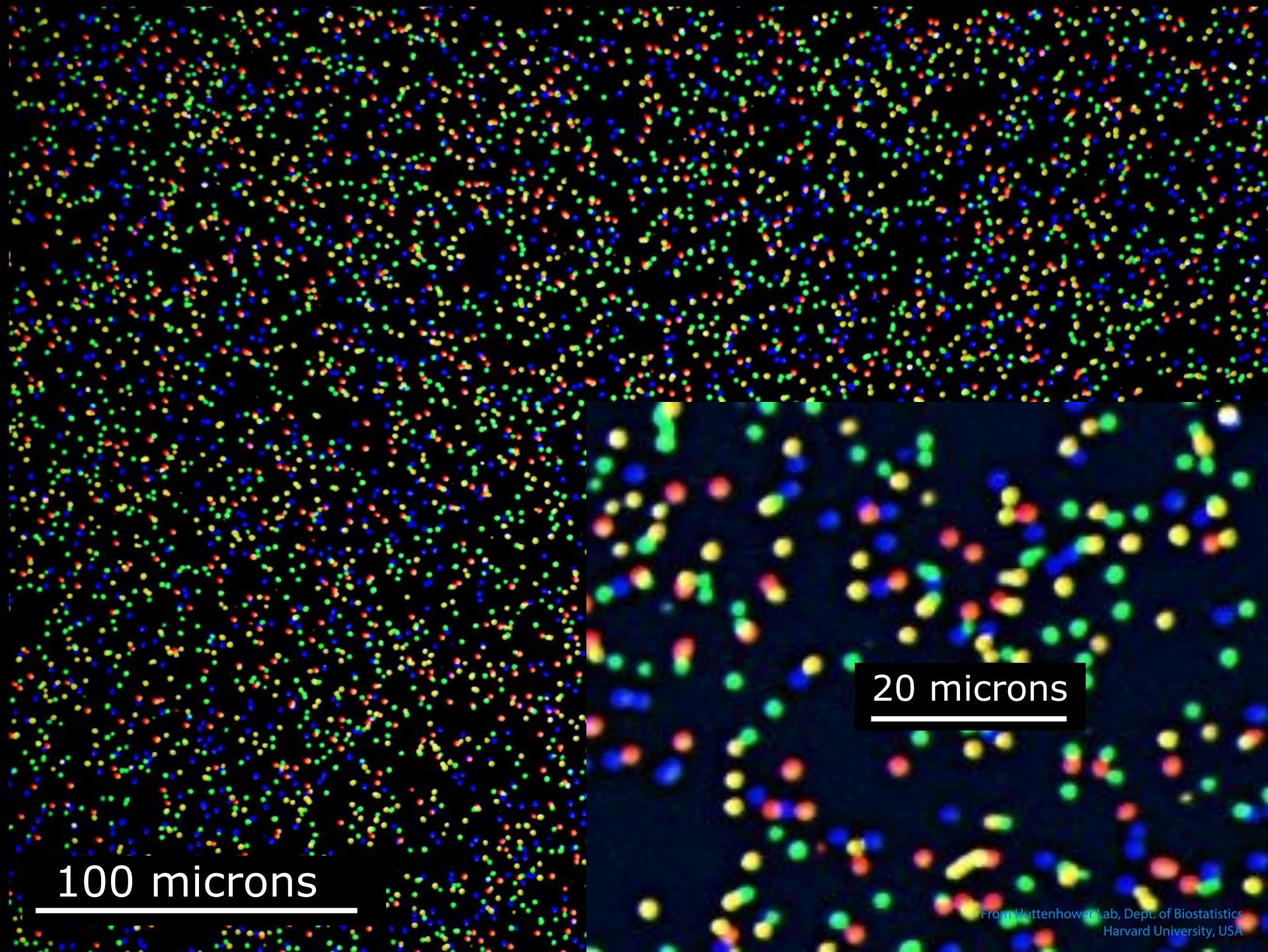
Cycle 1: Add sequencing reagents

First base incorporated

Remove unincorporated bases

Detect signal

Cycle 2-n: Add sequencing reagents and repeat



100 microns

20 microns

Flow Cells and Imaging

Solexa: single 8-channel

Reagents flowed in here

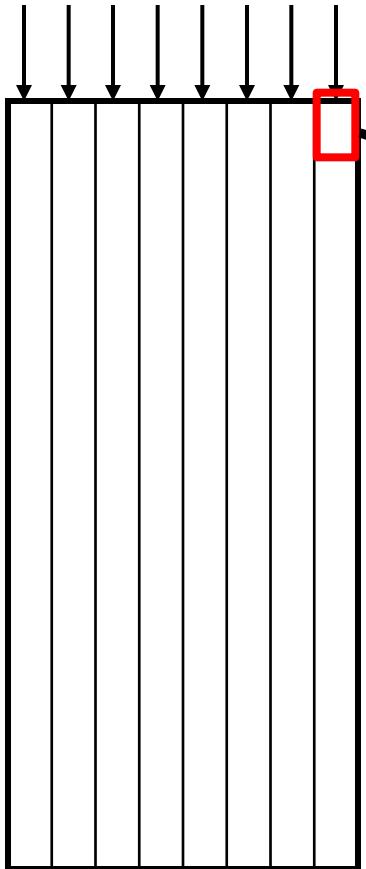
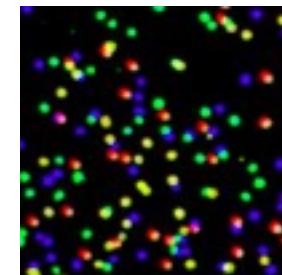
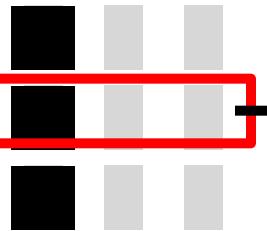


Image for each
channel is
actually tiled (four
images per panel,
one for each
color)



Sequencing by Synthesis (CRT): Qiagen GeneReader

Similar to 454 approach in which DNA fragments are linked to beads and amplified but different sequencing technologies used

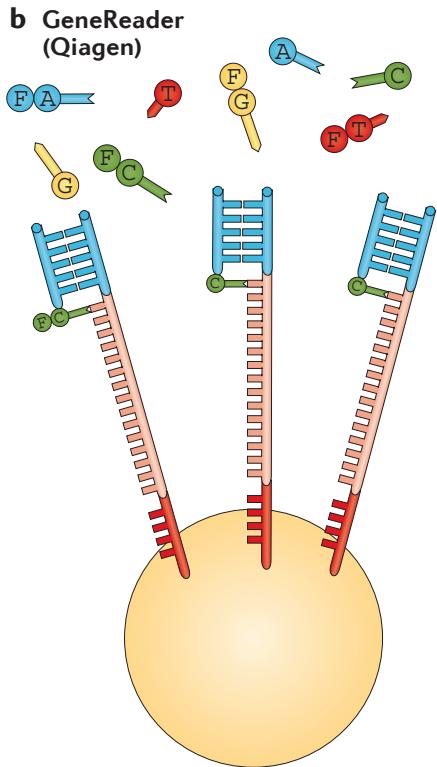
After bead-based template enrichment, a mixture of primers, DNA polymerase and modified nucleotides are added to the flow cell.

Each nucleotide is blocked by a 3'-O-allyl group and some of the bases are labelled with a base-specific, cleavable fluorophore.

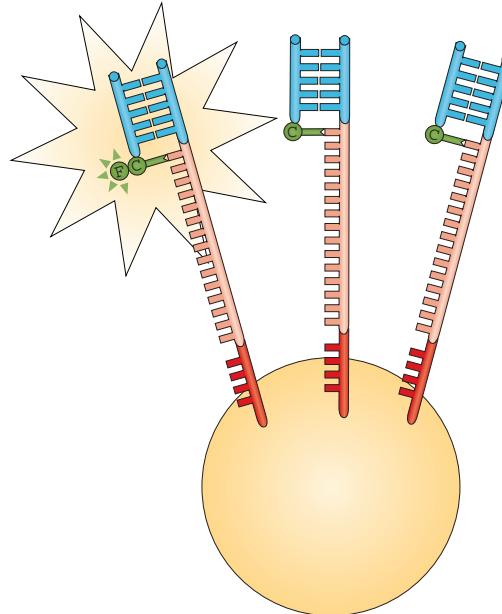
After base incorporation, unincorporated bases are washed away and the slide is imaged

The dye is then cleaved and the 3'-OH is regenerated

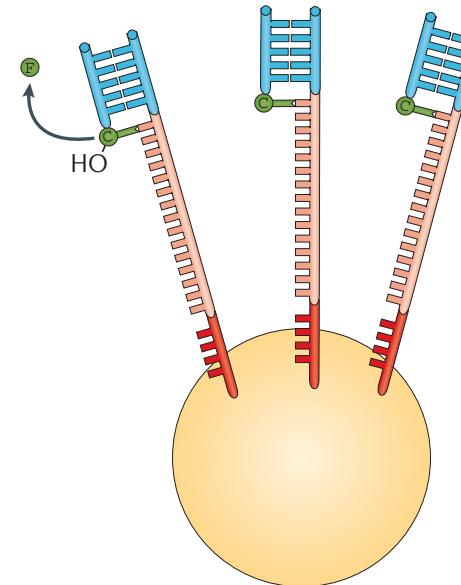
Sequencing by Synthesis (CRT): Qiagen GeneReader



Nucleotide addition
A mixture of fluorophore-labelled, terminally blocked nucleotides and unlabelled, blocked nucleotides hybridize to complementary bases. Each bead on a slide can incorporate a different base.



Imaging
Slides are imaged with four laser channels. Each bead emits a colour corresponding to the base incorporated during this cycle, but only labelled bases emit a signal.



Cleavage
Fluorophores are cleaved and washed from flow cells and the 3'-OH group is regenerated. A new cycle begins with the addition of new nucleotides.

Sequencing by Ligation: ABI SOLiD (Supported Oligonucleotide Ligation and Detection)

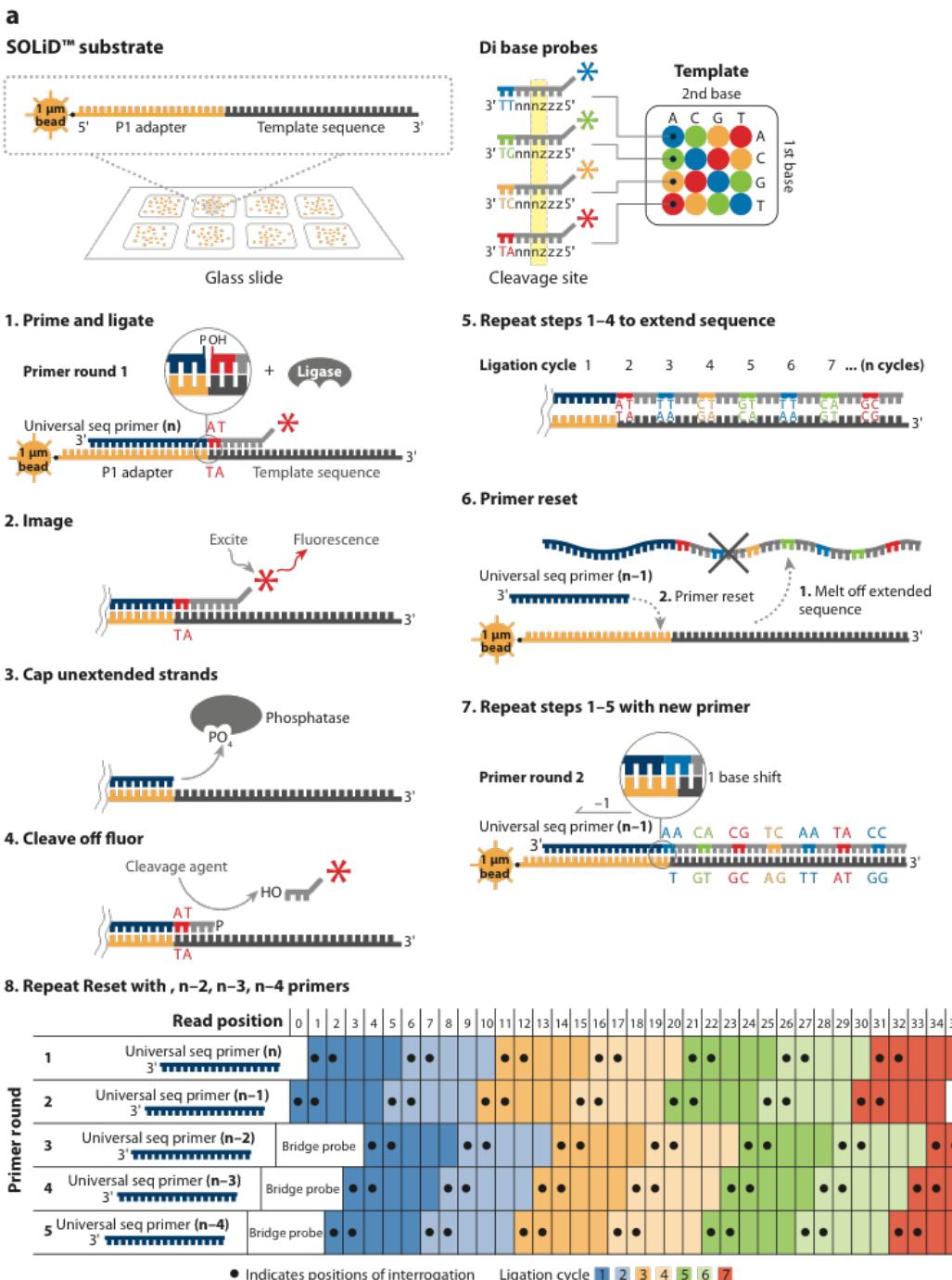
Similar to 454 approach in which DNA fragments are linked to beads and amplified but different sequencing technologies used

SOLiD technology employs sequencing by ligation. Here, a pool of all possible oligonucleotides of a fixed length are labeled according to the sequenced position.

Oligonucleotides are annealed and ligated; the preferential ligation by DNA ligase for matching sequences results in a signal informative of the nucleotide at that position.

Before sequencing, the DNA is amplified by emulsion PCR. The resulting bead, each containing only copies of the same DNA molecule, are deposited on a glass slide. The result is sequences of quantities and lengths comparable to Illumina sequencing.

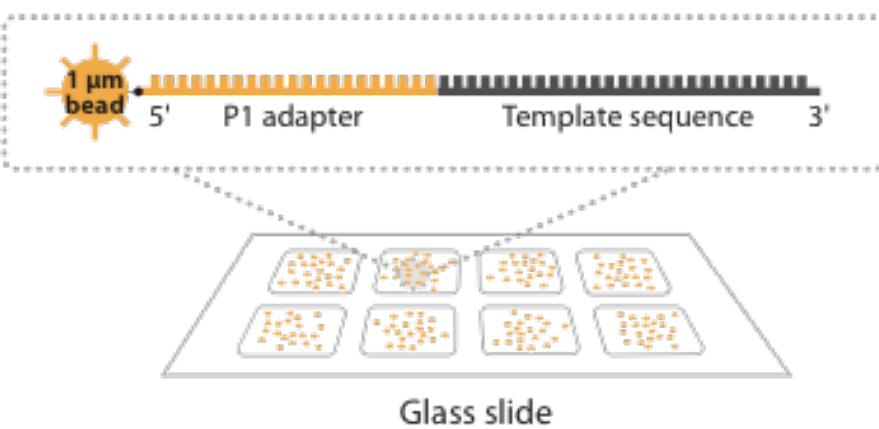
ABI SOLiD



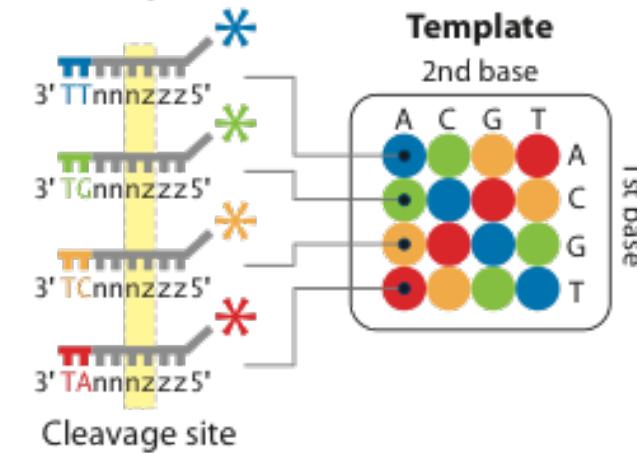
ABI SOLID

a

SOLiD™ substrate

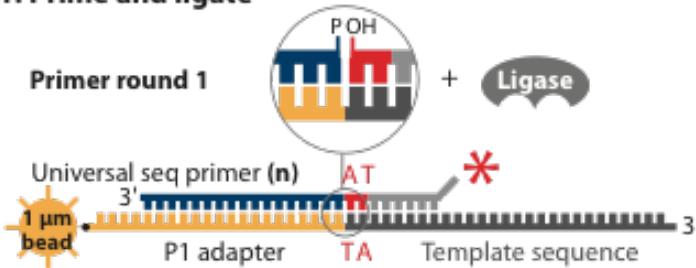


Di base probes

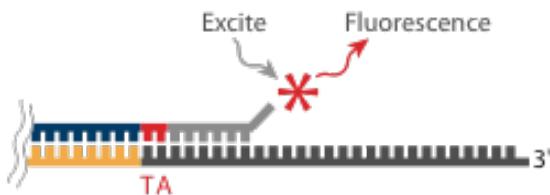


ABI SOLID

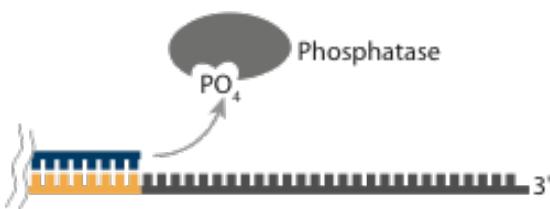
1. Prime and ligate



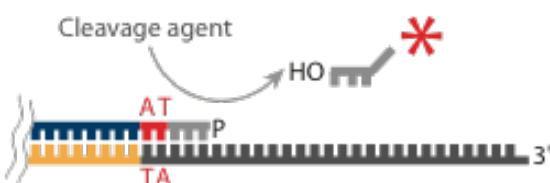
2. Image



3. Cap unextended strands



4. Cleave off fluor



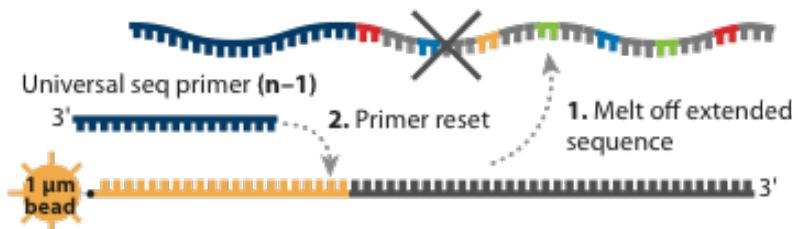
5. Repeat steps 1–4 to extend sequence

5. Repeat steps 1–4 to extend sequence

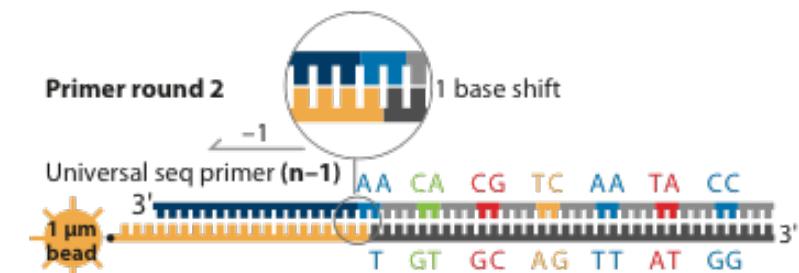
Ligation cycle 1 2 3 4 5 6 7 ... (n cycles)



6. Primer reset

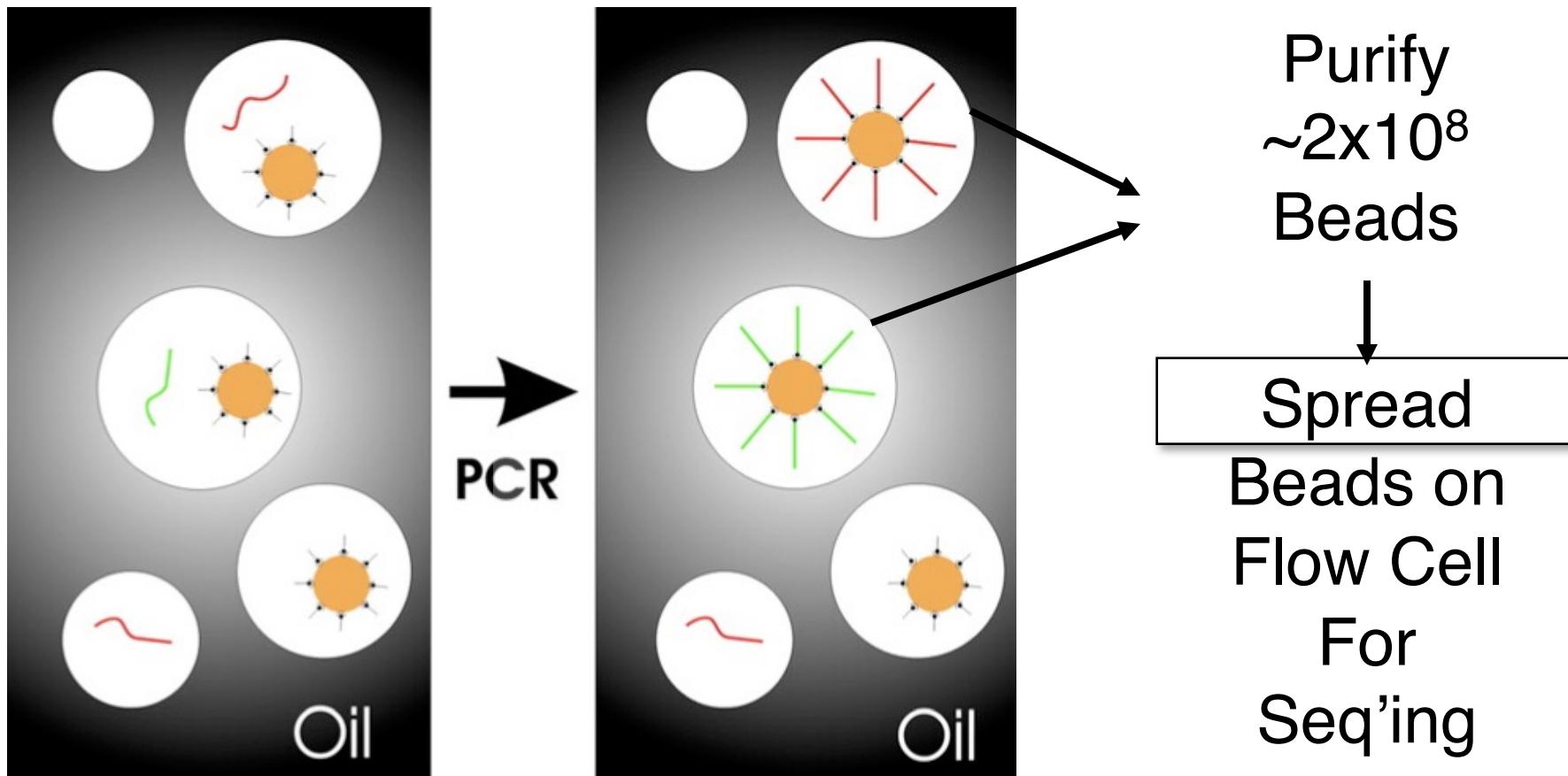


7. Repeat steps 1–5 with new primer



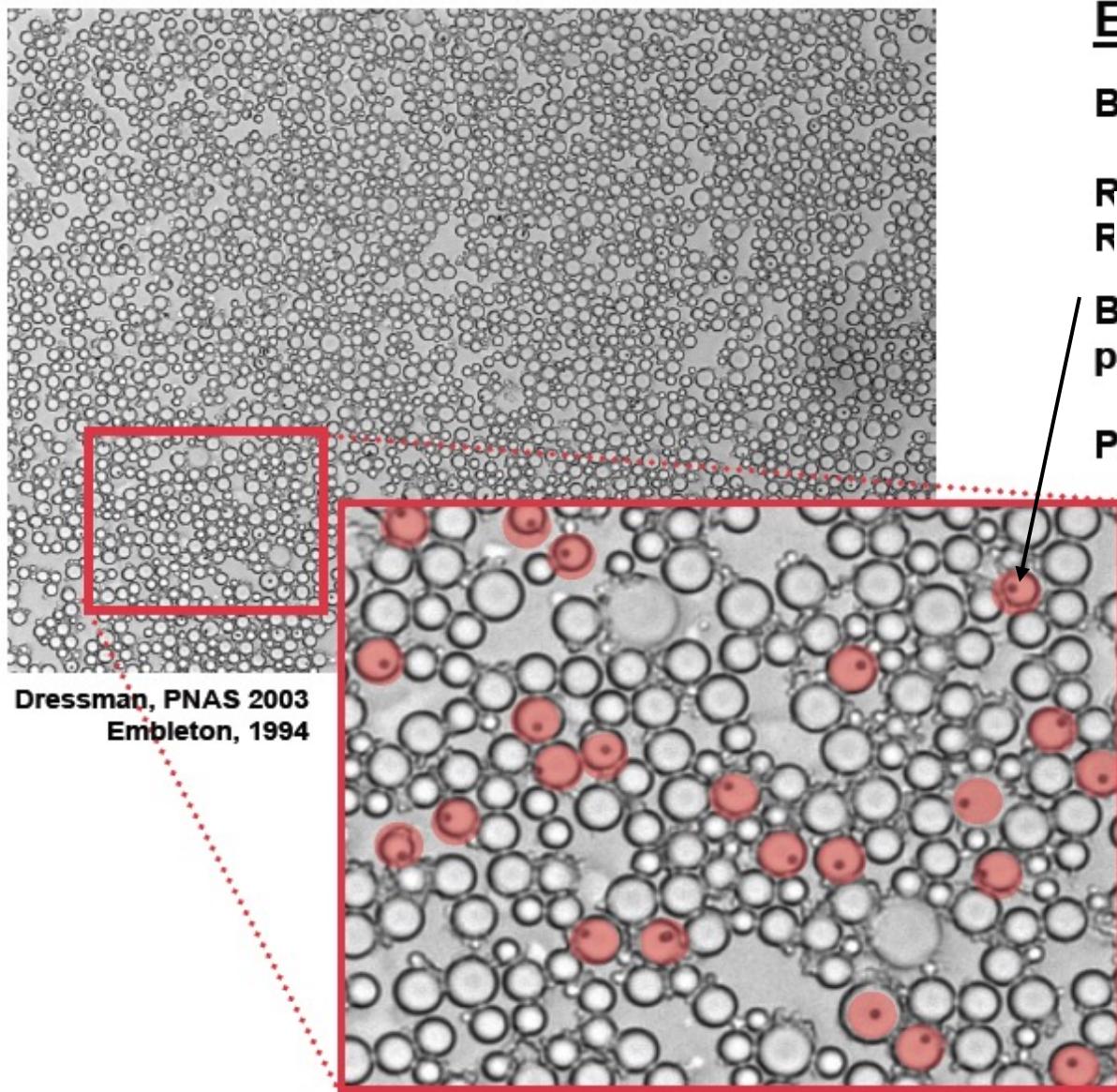
8. Repeat Reset with , n-2, n-3, n-4 primers

SOLiD Emulsion PCR on Beads



From Dressman et al. (2003) Proc. Natl.
Acad. Sci. USA 100, 8817-8822

Emulsions with 1 M Beads (SOLiD)

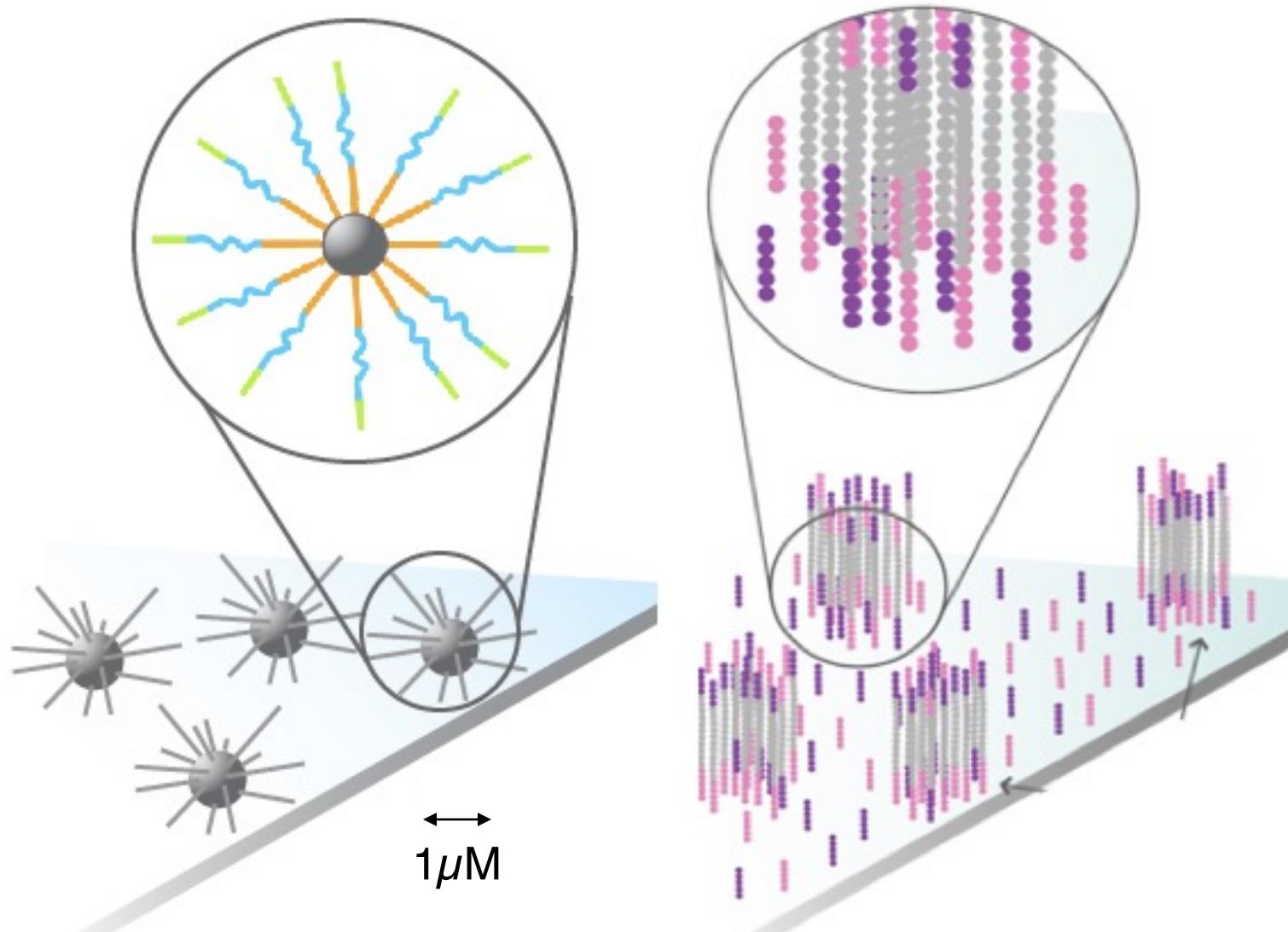


Ei
Be
R-
R
Be
pla
Po

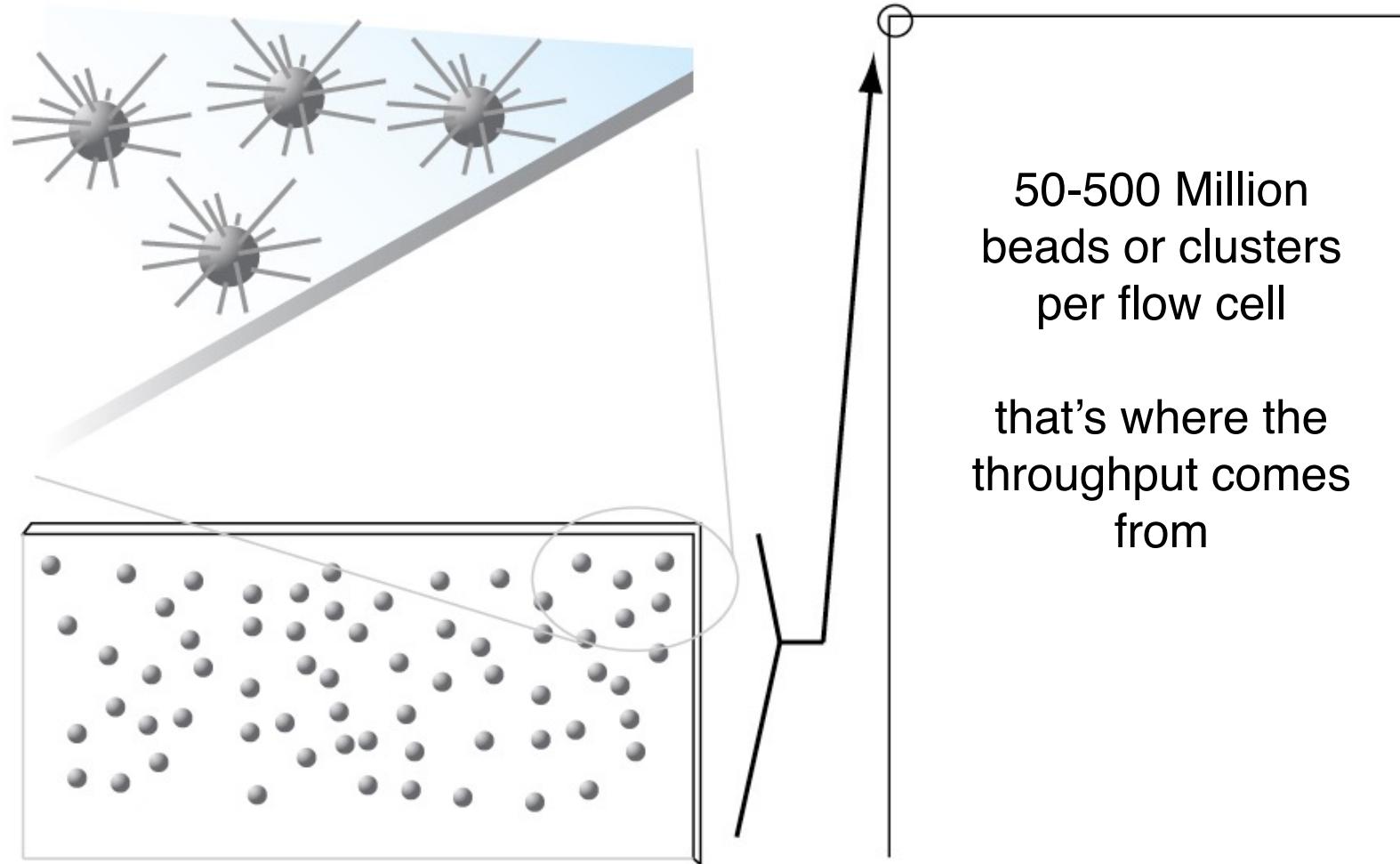
reactors containing a
bead are false-colored

Courtesy of Gina
Costa (AB)

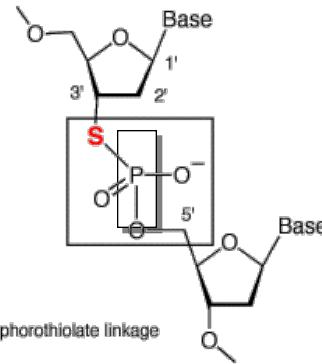
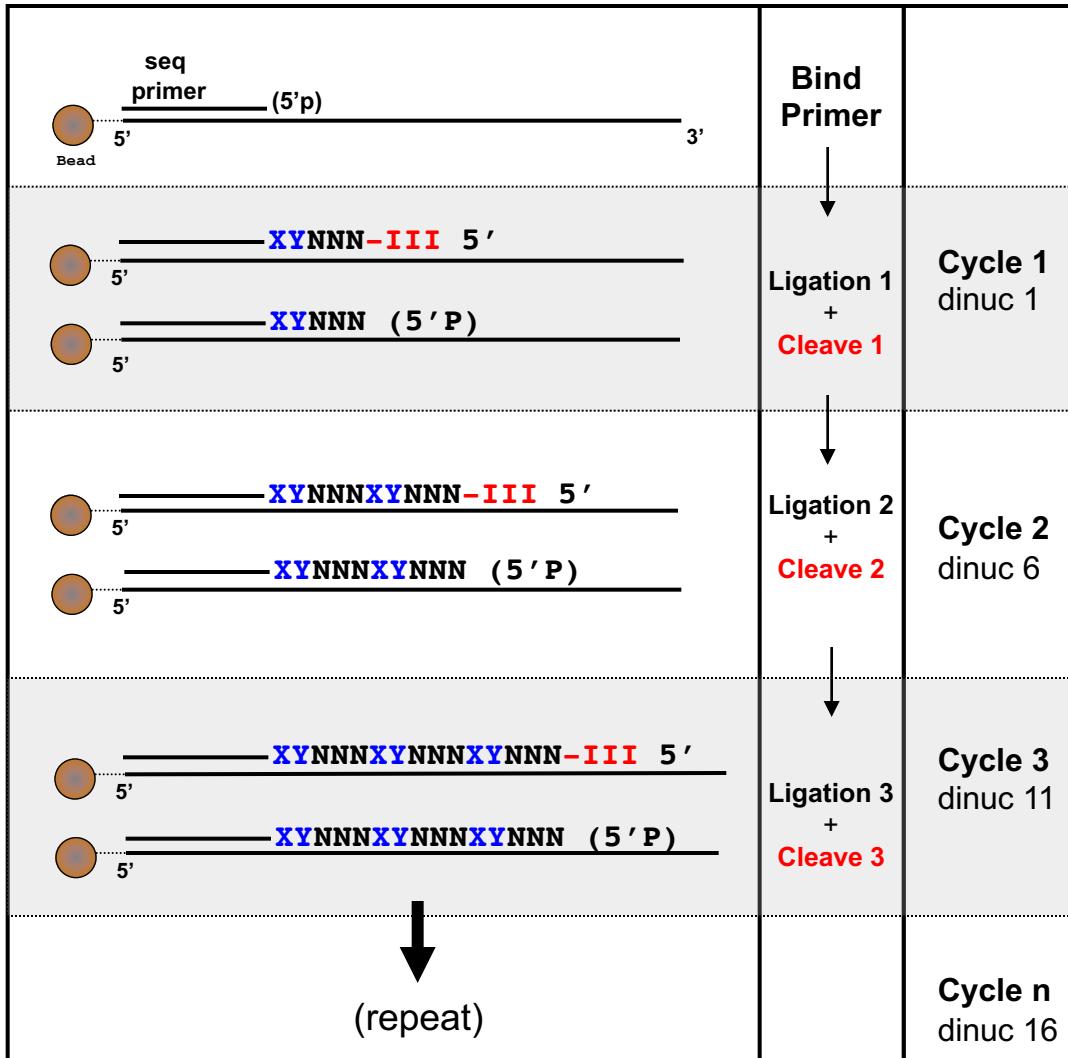
Clonally Amplified Molecules on Flow Cell



Scale



SOLiD: Multiple Cycles and “Dinucleotide Color Space”

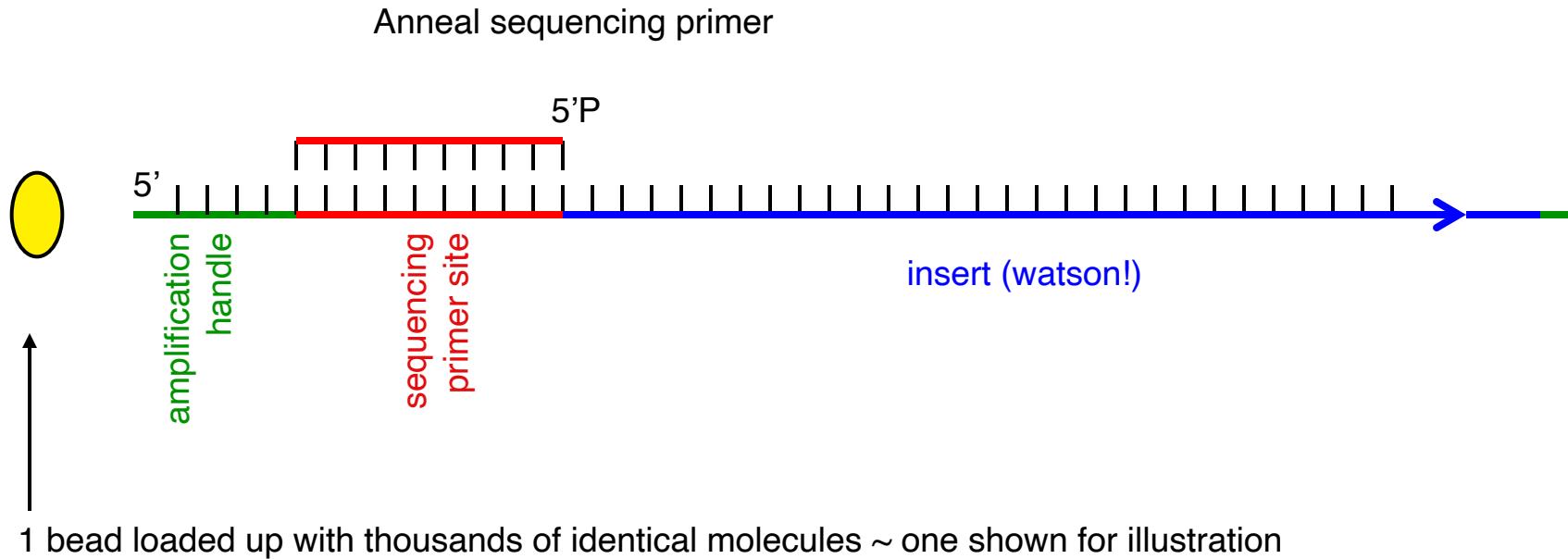


Phosphorothiolate Linkage

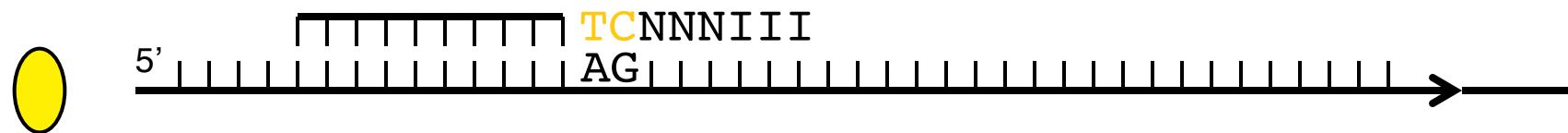
- Cleaves with neutral pH, aqueous conditions in seconds
- Removes dye, cleaves off Inosines, and produces 5' Phosphate (native DNA end) in a single step

Dinuc	FAM	Cy3	TXR	Cy5
XY ₁	AA	AC	AG	AT
XY ₂	CC	CA	GA	TA
XY ₃	GG	GT	CT	CG
XY ₄	TT	TG	TC	GC

SOLiD Sequencing



SOLiD Sequencing



CCNNNIII	ACNNNIII	AGNNNIII	ATNNNIII
TTNNNIII	CAANNIII	GAANNIII	CGNNNIII
AAANNIII	TGNNNIII	TCNNNIII	GCNNNIII
GGNNNIII	GTNNNIII	CTNNNIII	TAANNIII

Sixteen probes, but only four colors

SOLiD Sequencing



CCNNNIII	ACNNNIII	AGNNNIII	ATNNNIII
TTNNNIII	CANNNIII	GANNNIII	CGNNNIII
AAANNNIII	TGNNNIII	TCNNNIII	GCNNNIII
GGNNNIII	GTNNNIII	CTNNNIII	TAANNNIII

SOLiD Sequencing

After five cycles



CCNNNIII

ACNNNIII

AGNNNIII

ATNNNIII

TTNNNIII

CA~~NNN~~III

GA~~NNN~~III

CGNNNIII

AA~~NNN~~III

TGNNNIII

TCNNNIII

GCNNNIII

GGNNNIII

GTNNNIII

CTNNNIII

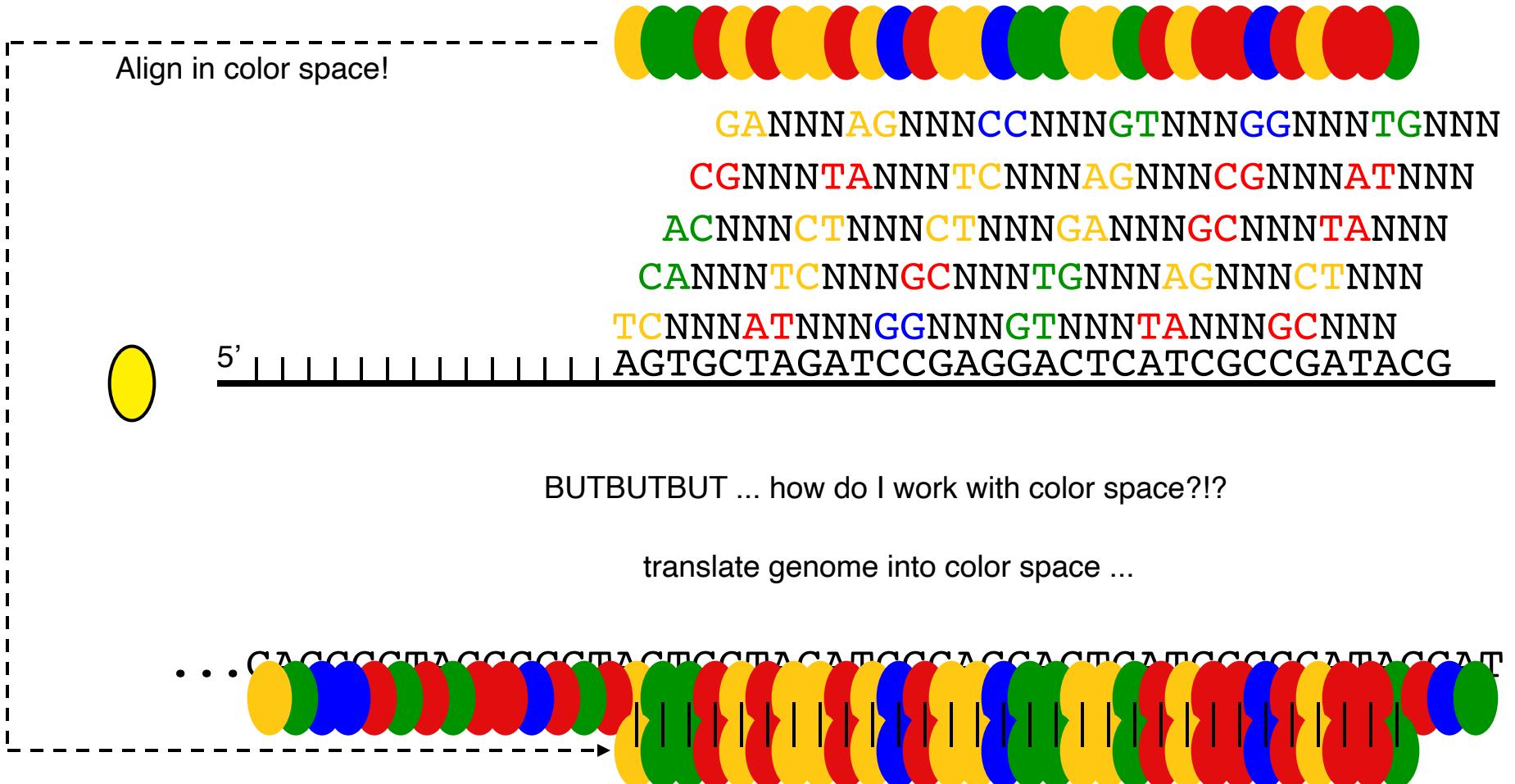
TA~~NNN~~III

SOLiD Sequencing

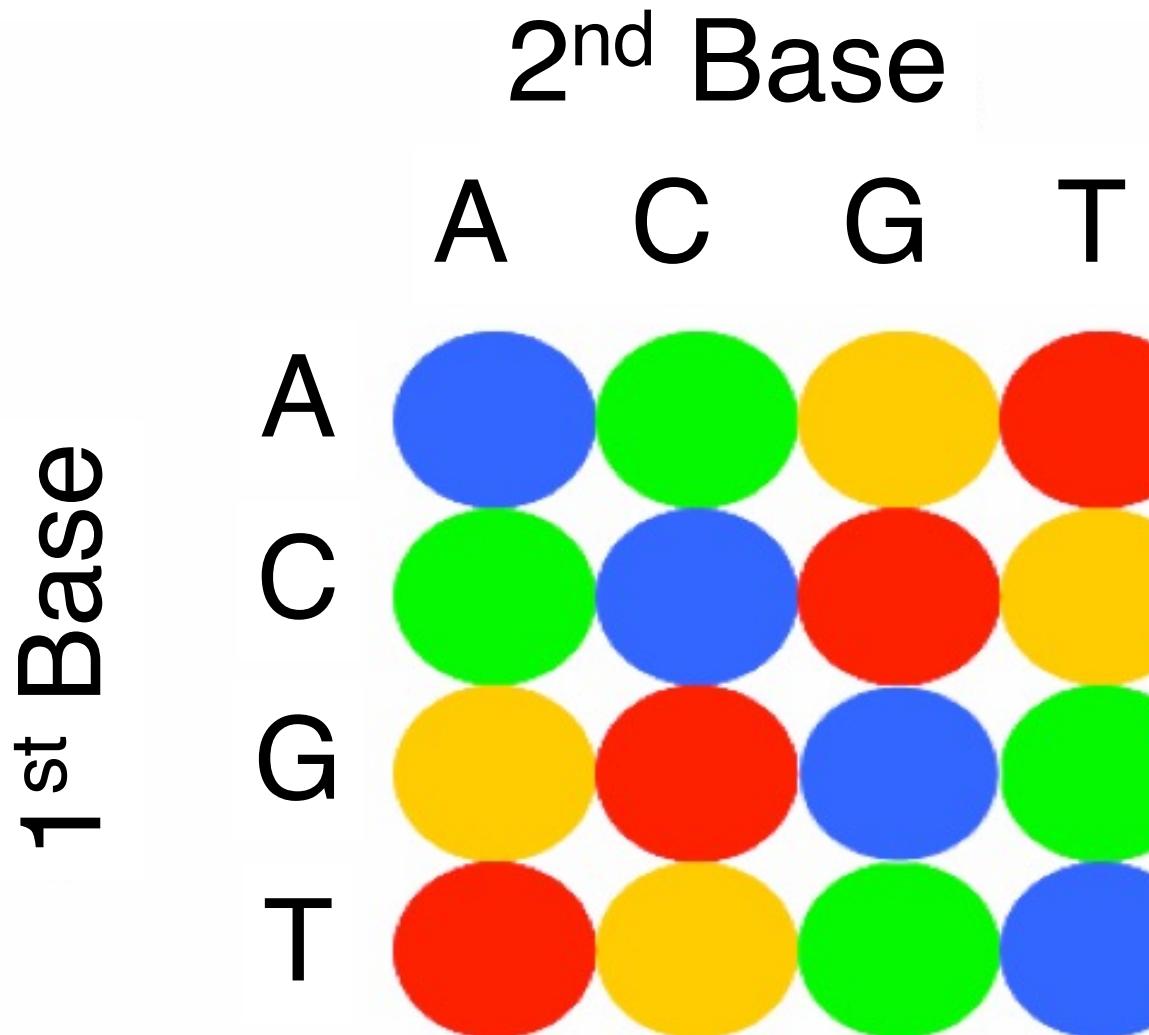
“Reset”: Melt off previous extension products
and repeat five cycles with offset of one base



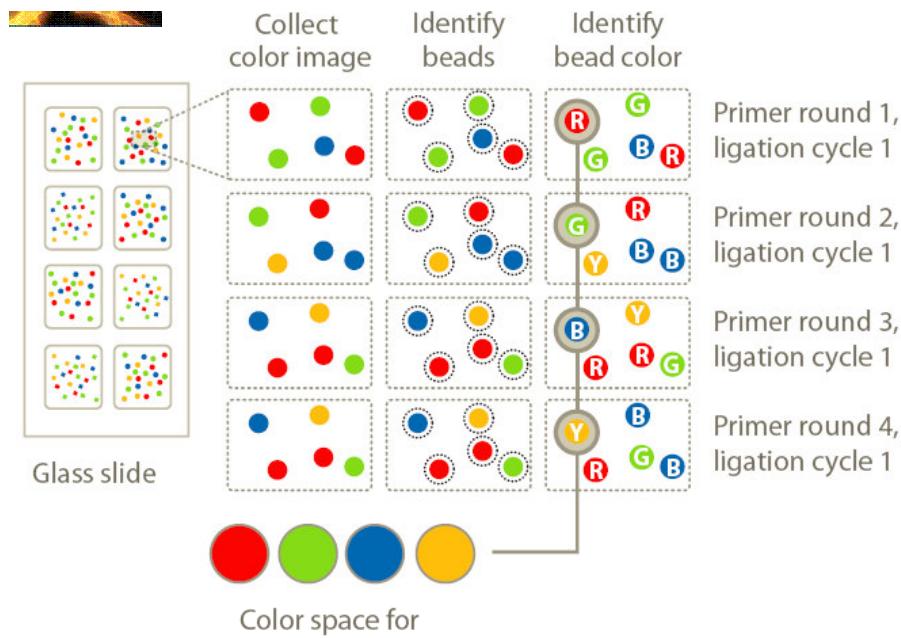
SOLiD Sequencing



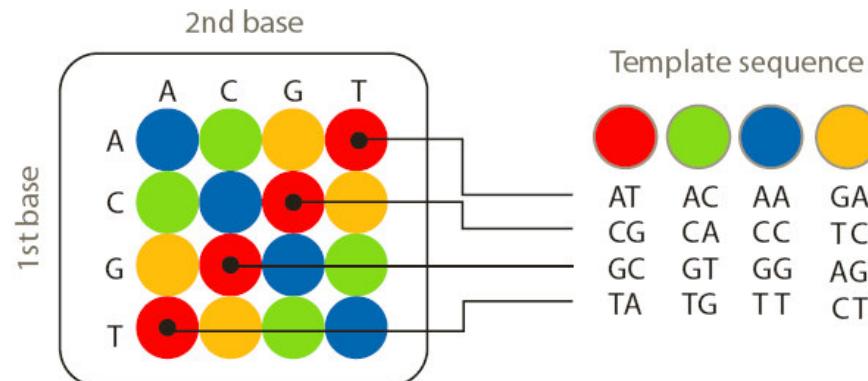
Translating color space back to base space



Translating color space back to base space

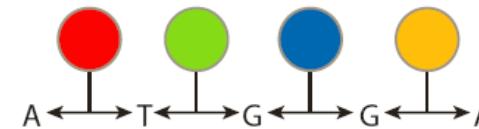


Possible dinucleotides encoded by each color

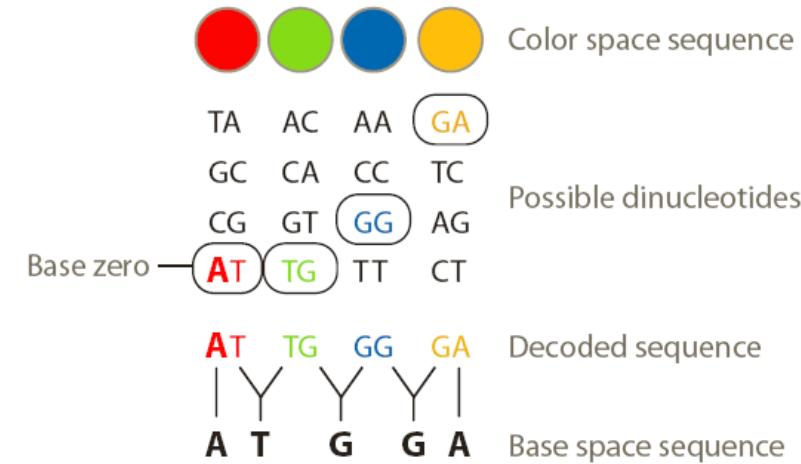


Double interrogation

With 2 base encoding each base is defined twice



Decoding



Sequencing by Ligation: Complete Genomics (BGI)

BGI is the only approach that enriches and amplifies DNA in solution.

The DNA is prepared as circular template with 4 adaptors, each separated by a template sequence.

DNA is sequenced using the combinational probe-anchor ligation. An anchor complementary to one of four adapter sequences and a fluorophore-labelled probe are bound to each nanoball.

The probe is degenerate at all but the first position. The anchor and probe are then ligated into position and imaged to identify the first base on either the 3' or the 5' side of the anchor.

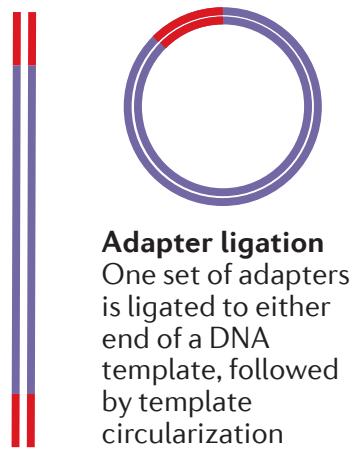
Next, the probe–anchor complex is removed and the process begins again with the same anchor but a different probe with the known base at the n + 1 position.

This is repeated until five bases from the 3' end of the anchor and five bases from the 5' end of the anchor are identified.

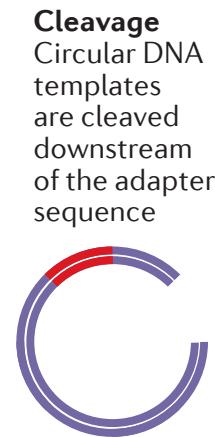
Another round of hybridization occurs, this time using anchors with a five-base offset identifying an additional five bases on either side of the anchor.

Sequencing by Ligation: Complete Genomics (BGI)

d In-solution DNA nanoball generation (Complete Genomics (BGI))



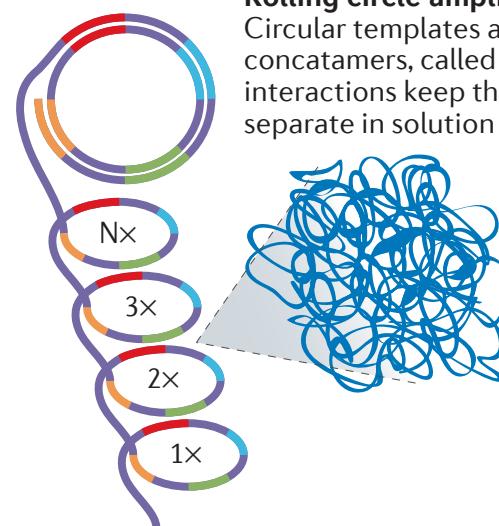
Adapter ligation
One set of adapters is ligated to either end of a DNA template, followed by template circularization



Cleavage
Circular DNA templates are cleaved downstream of the adapter sequence

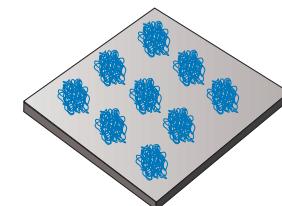


Iterative ligation
Three additional rounds of ligation, circularization and cleavage generate a circular template with four different adapters



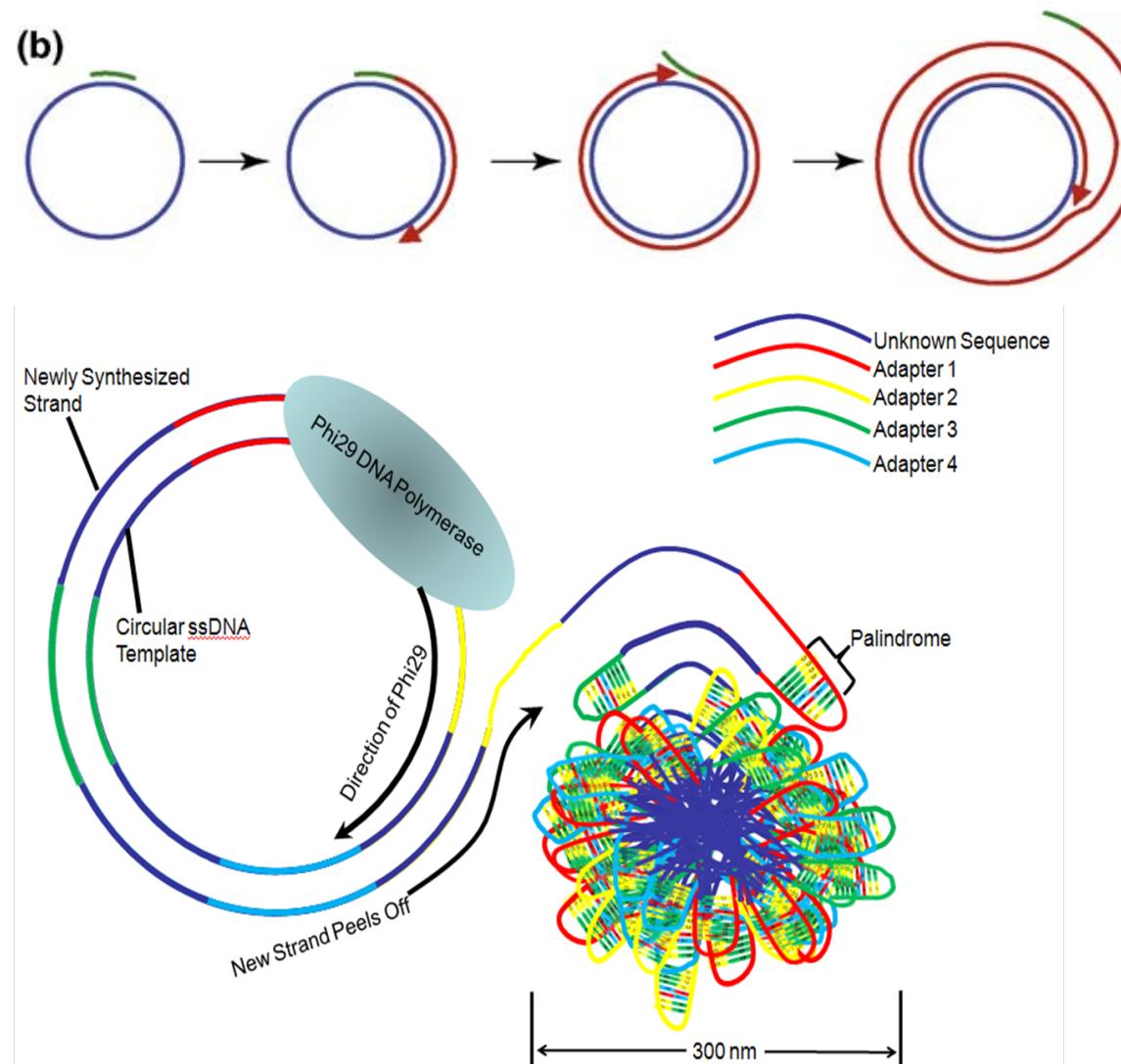
Rolling circle amplification

Circular templates are amplified to generate long concatamers, called DNA nanoballs; intermolecular interactions keep the nanoballs cohesive and separate in solution

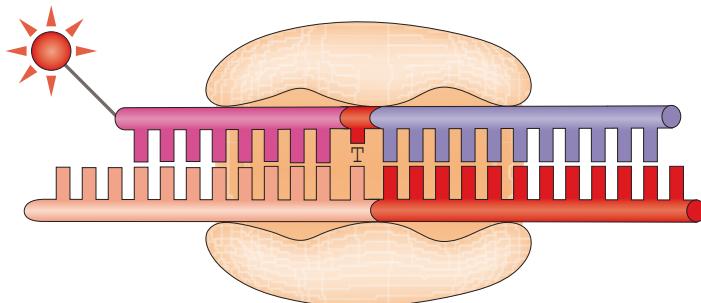


Hybridization
DNA nanoballs are immobilized on a patterned flow cell

Sequencing by Ligation: Complete Genomics (BGI)

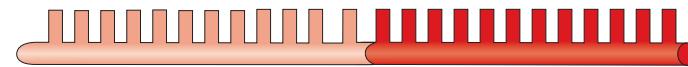


Sequencing by Ligation: Complete Genomics (BGI)



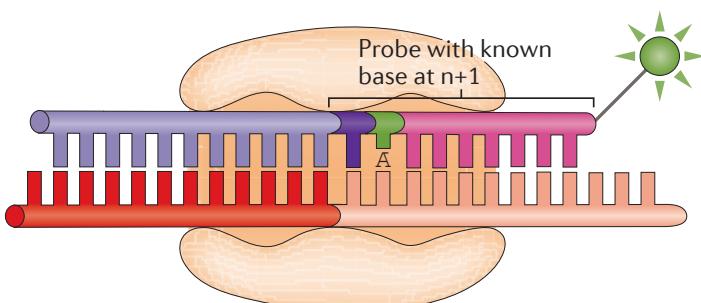
Single-base-encoded probes

A probe with a single known base and degenerate bases hybridizes to a template and is imaged



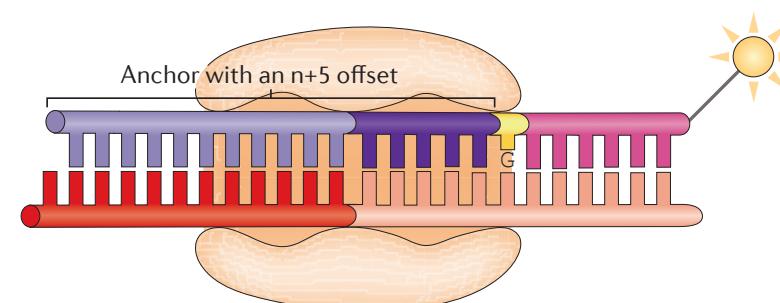
Reset

After each imaging step, both the probe and anchor are removed



Paired-end sequencing

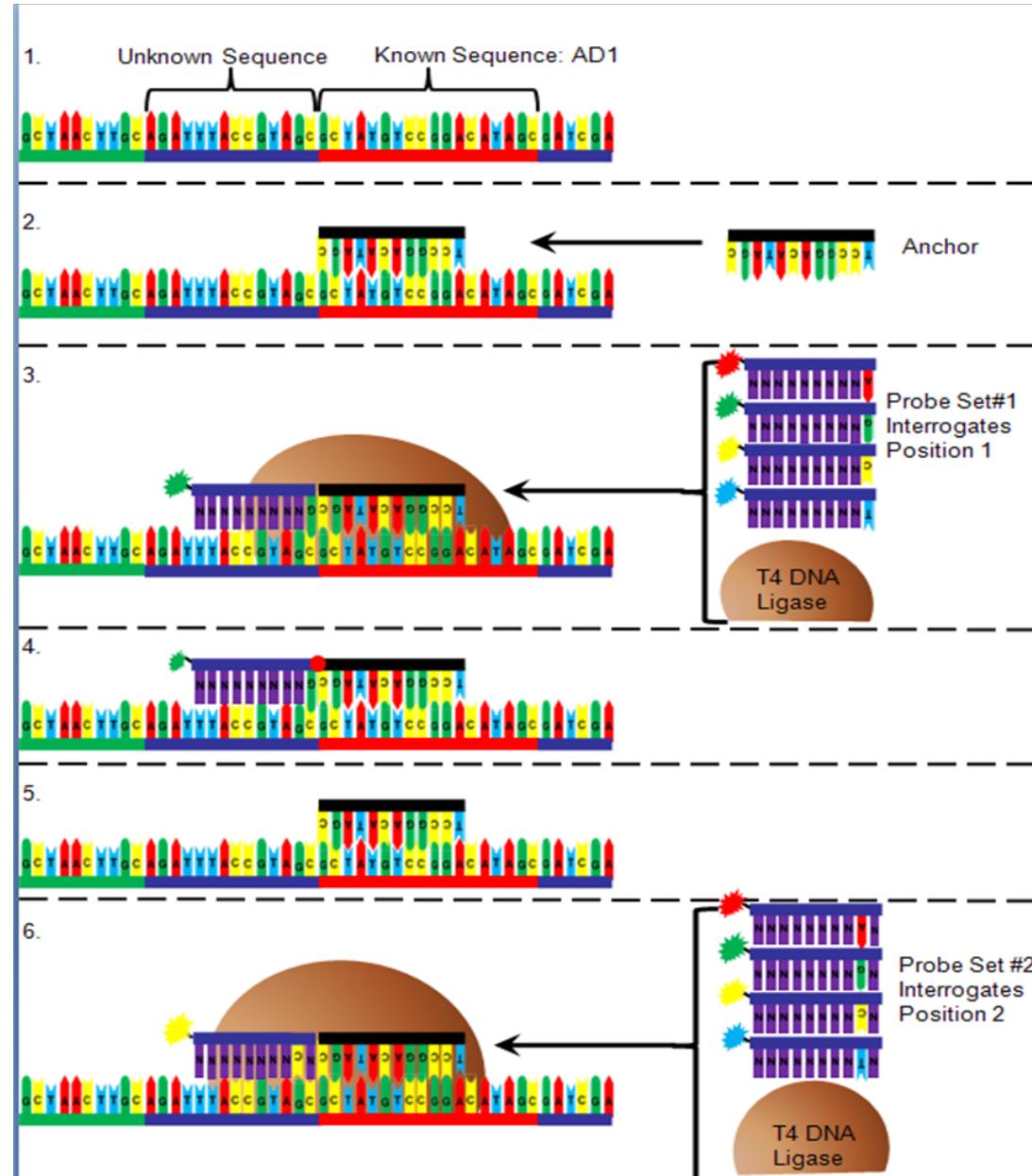
Sequencing is performed for both the left and right sides of the adapter



Offset anchors

Subsequent rounds of hybridization and ligation use offset anchors to sequence more-distant bases

Sequencing by Ligation: Complete Genomics (BGI)



Single-Molecule Real-Time Sequencing

BacBio

SMRT sequencing is based on the sequencing by synthesis approach.

The DNA is synthesized in small well-like containers with the capturing tools located at the bottom of the well.

The sequencing is performed with use of unmodified polymerase (attached to the ZMW bottom) and fluorescently labelled nucleotides flowing freely in the solution.

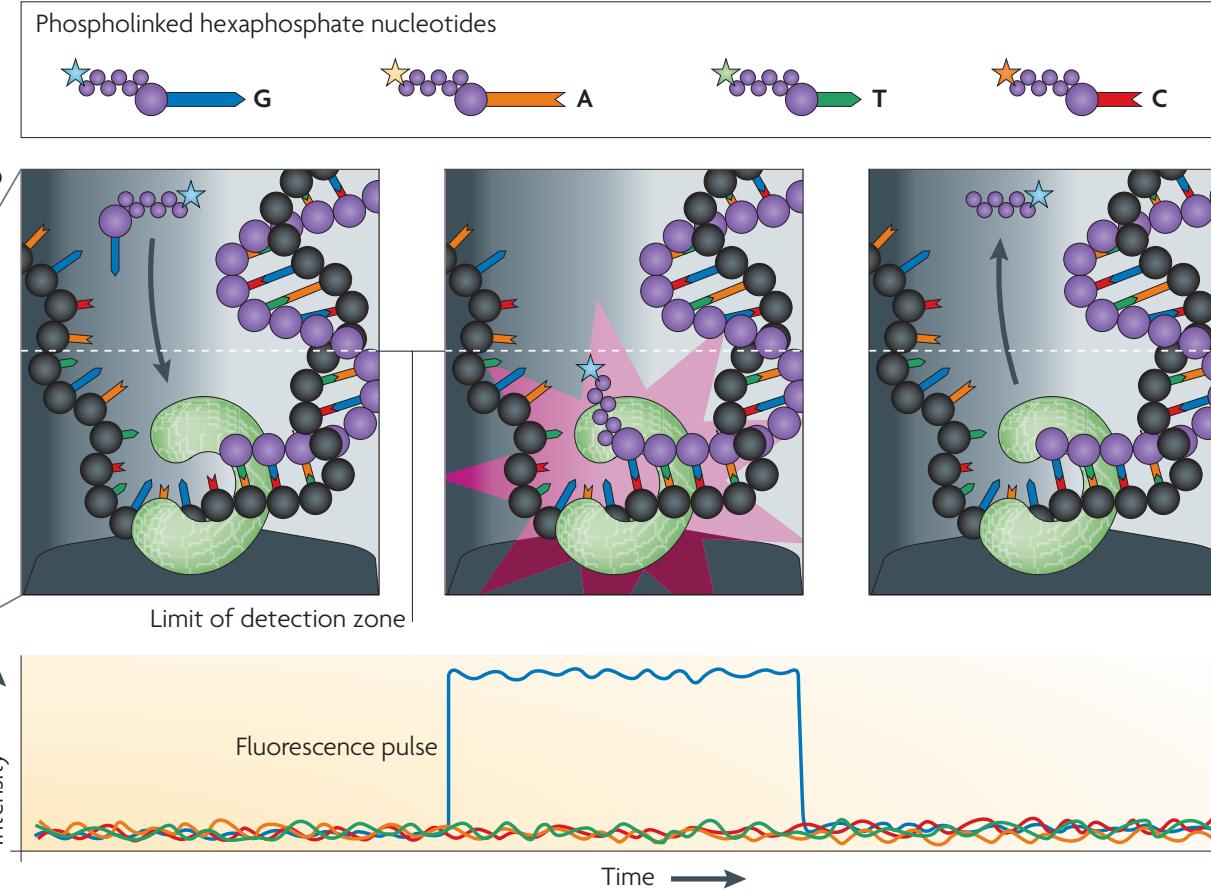
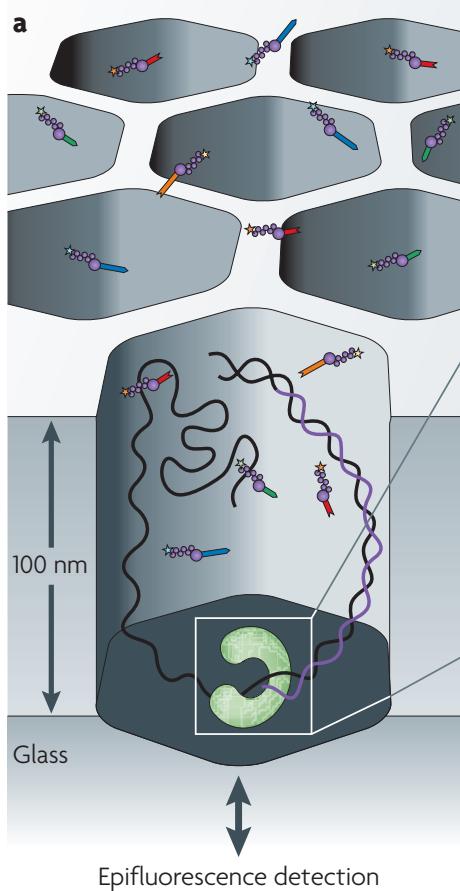
The wells are constructed in a way that only the fluorescence occurring by the bottom of the well is detected.

The fluorescent label is detached from the nucleotide at its incorporation into the DNA strand, leaving an unmodified DNA strand.

Single-Molecule Real-Time Sequencing

BacBio

Pacific Biosciences — Real-time sequencing

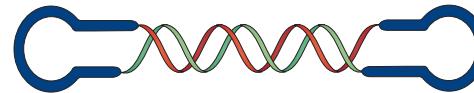


Single-Molecule Real-Time Sequencing

BacBio

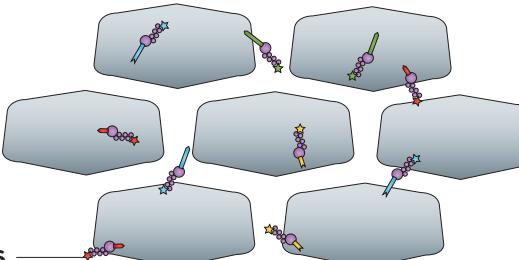
SMRTbell template

Two hairpin adapters allow continuous circular sequencing



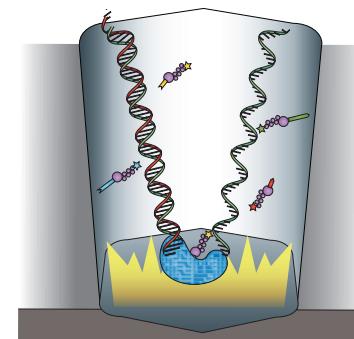
ZMW wells

Sites where sequencing takes place



Labelled nucleotides

All four dNTPs are labelled and available for incorporation



Modified polymerase

As a nucleotide is incorporated by the polymerase, a camera records the emitted light

PacBio output

A camera records the changing colours from all ZMWs; each colour change corresponds to one base



Single-Molecule Real-Time Sequencing

Oxford Nanopore

This method is based on the readout of electrical signal occurring at nucleotides passing by a nanopores

The DNA passing through the nanopore changes its ion current.

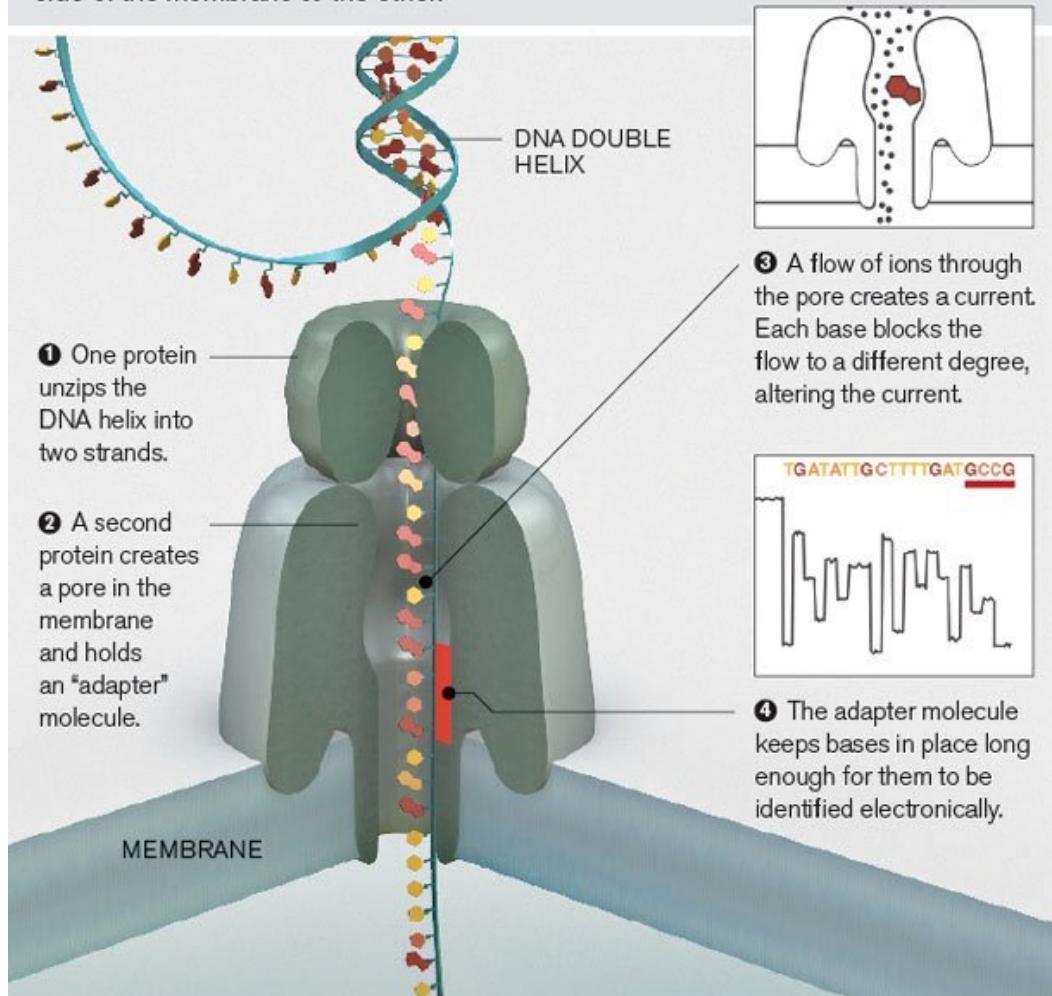
This change is dependent on the shape, size and length of the DNA sequence.

Each type of the nucleotide blocks the ion flow through the pore for a different period of time.

Single-Molecule Real-Time Sequencing

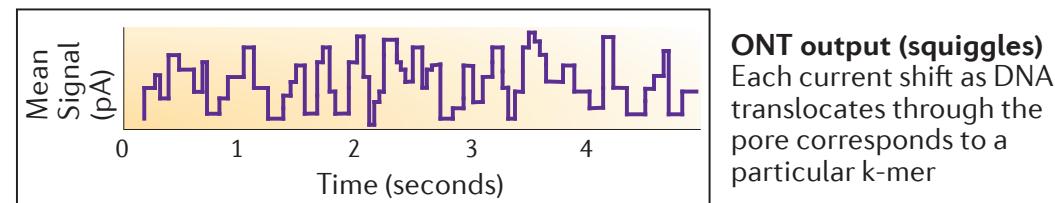
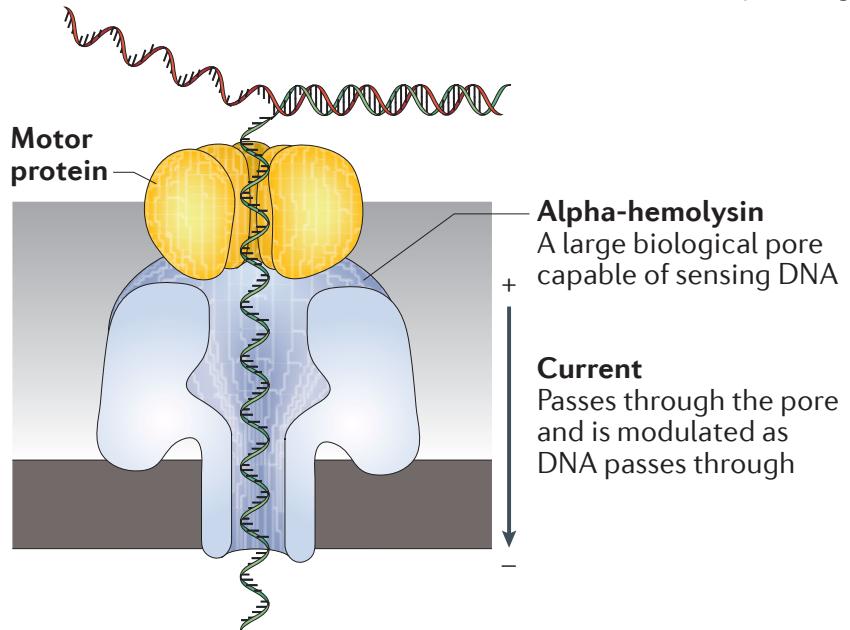
Oxford Nanopore

DNA can be sequenced by threading it through a microscopic pore in a membrane. Bases are identified by the way they affect ions flowing through the pore from one side of the membrane to the other.



Single-Molecule Real-Time Sequencing

Oxford Nanopore



Comparison between the different sequencing platforms

Table 2. Comparison of the different sequencing platforms. The data shown refer to the most favorable conditions for each platform.

Platform	Read Length (bp)	Accuracy (%)	Run Time	Bases Per Run (Gb)	Cost/Gb
454 Roche	1000	99	24 h	0.54	\$10,000
SOLiD	75	99.9	7 d	520	\$10
Illumina	300	99.9	3 d	1800	\$10
Ion Torrent	400	99	2 h	15	\$100
Pacific Bioscence	20,000	90	3 h	12,000	\$600
Oxford Nanopore	10,000	90	2 d	42	\$1000

bp, base pairs; Gb, gigabase pairs; h, hours; d, days.

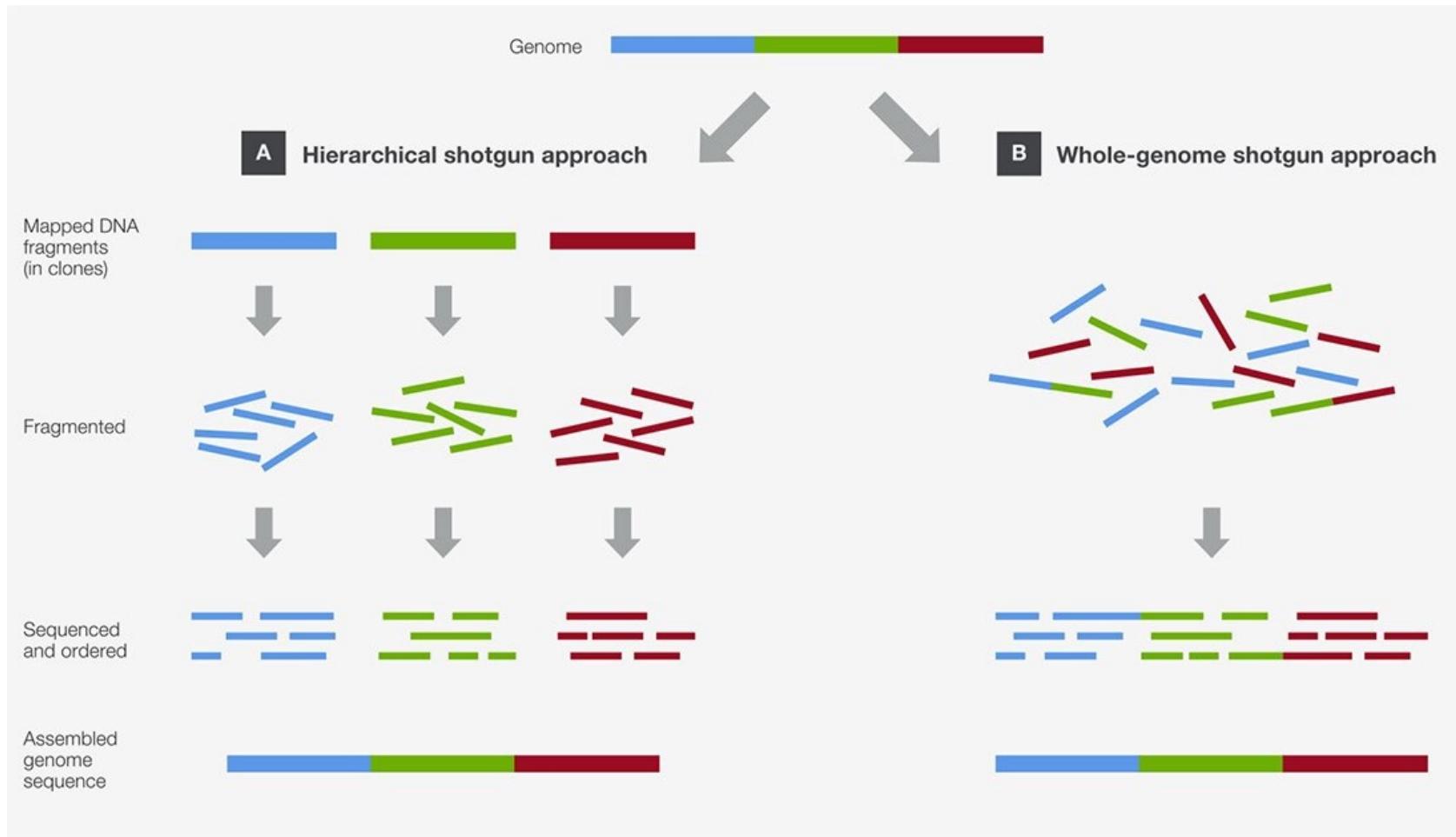
Comparison between the different sequencing platforms

Table 3. Pros and cons of each platform.

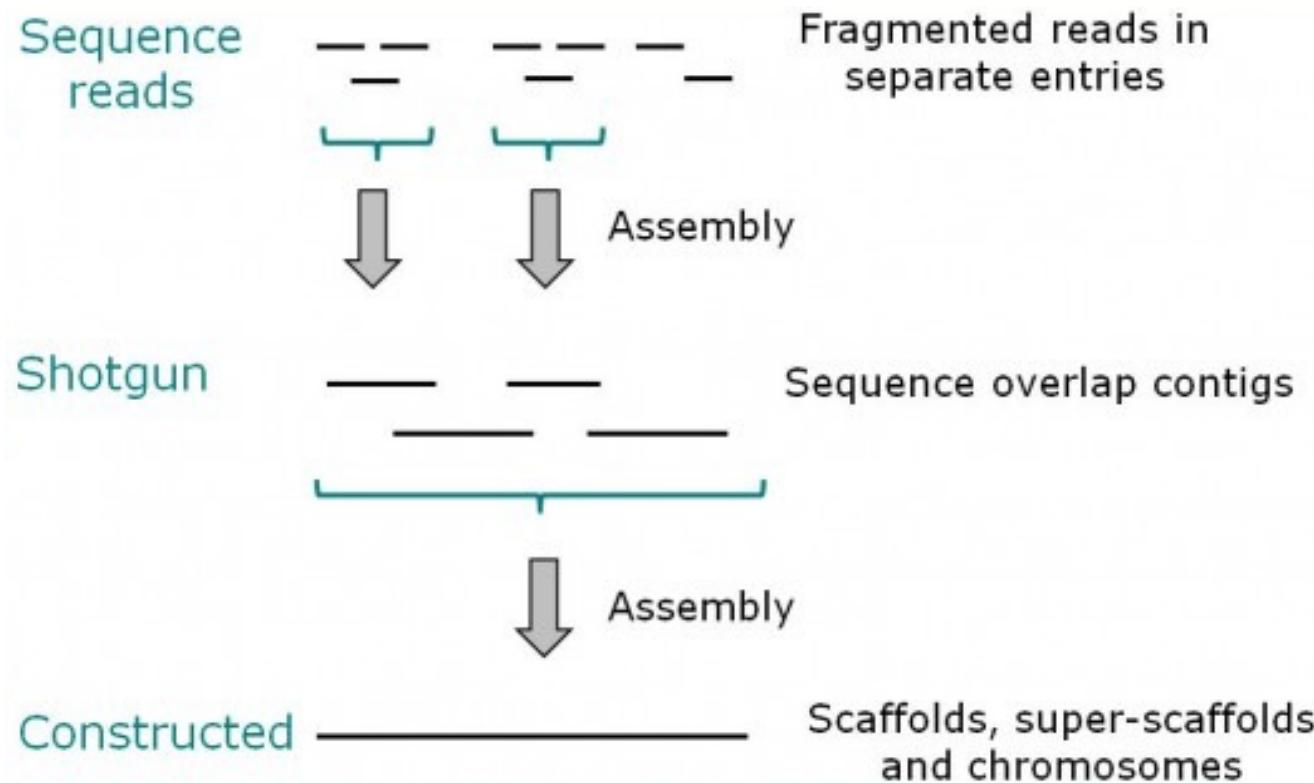
Platform	Pros	Cons
454 Roche	Long reading length. Low analysis time. Low cost for small studies	High error rate in homopolymers. Low performance. High instrumental cost. High cost per Gb data
SOLiD	High throughput. Low cost per Gb data. High accuracy	Short reading length. High instrumental cost
Illumina	High throughput. Low cost per Gb data. High accuracy	Short reading length. High instrumental cost
Ion Torrent	Low instrumental and operational cost. Short execution time. Very simple machine	Error rate not very good. Intermediate cost per Gb data. More hands-on time
Pacific Bioscience	Longest reading length available. Short instrument execution time	High error rate. High cost per Gb data. Many methods are still under development
Oxford Nanopore	Small, portable, and low cost instrument	High error rate. Biased errors. High cost per Reading

Gb, gigabase pairs.

Shotgun sequencing



Genome Assembly



Contig = a contiguous stretch of nucleotides resulting from the assembly of several reads

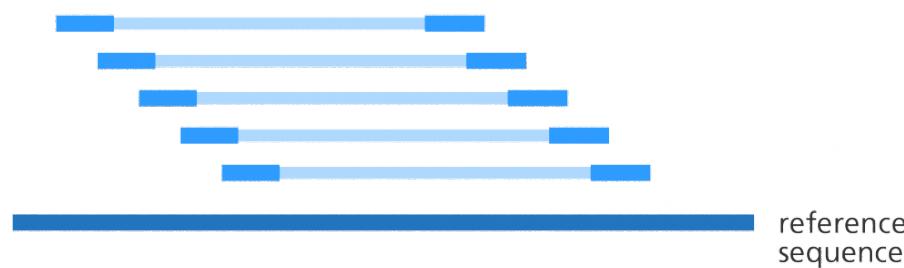
Scaffold = several contigs stitched together with NNNs in between

Types of reads

Single-end reads



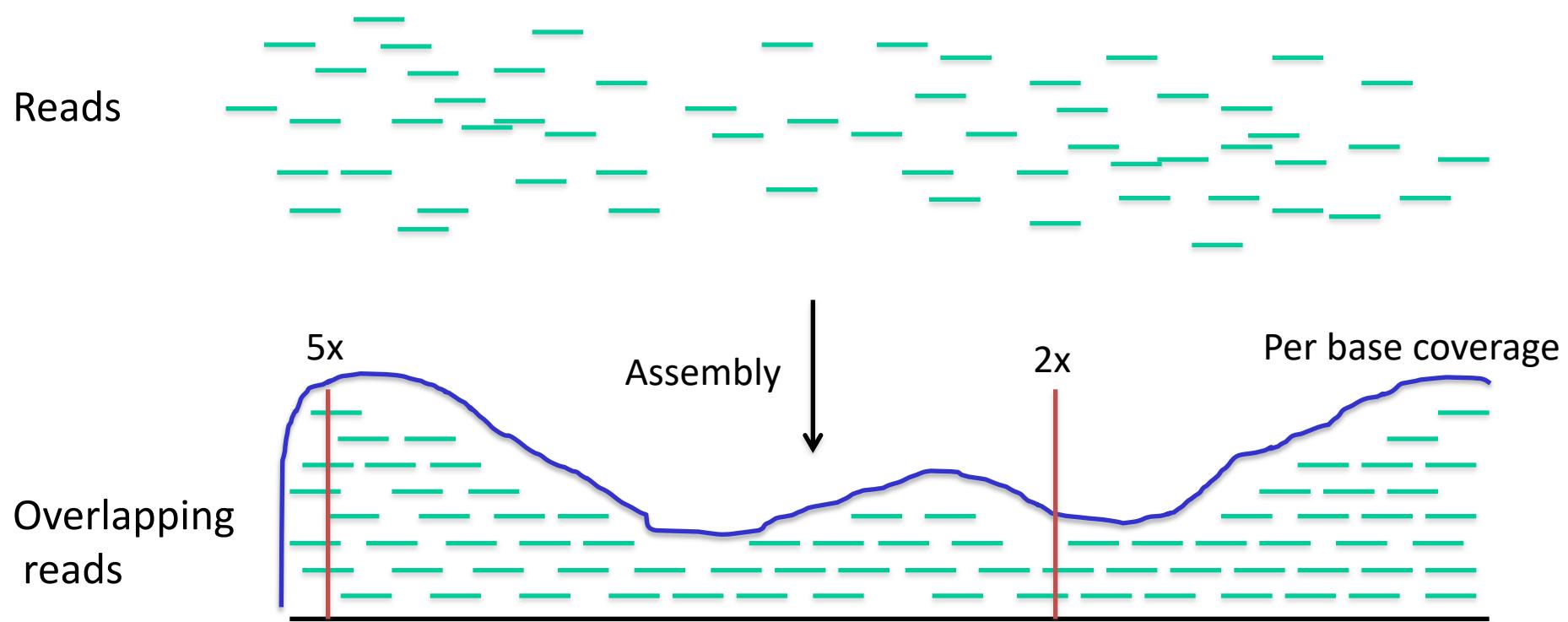
Paired-end reads



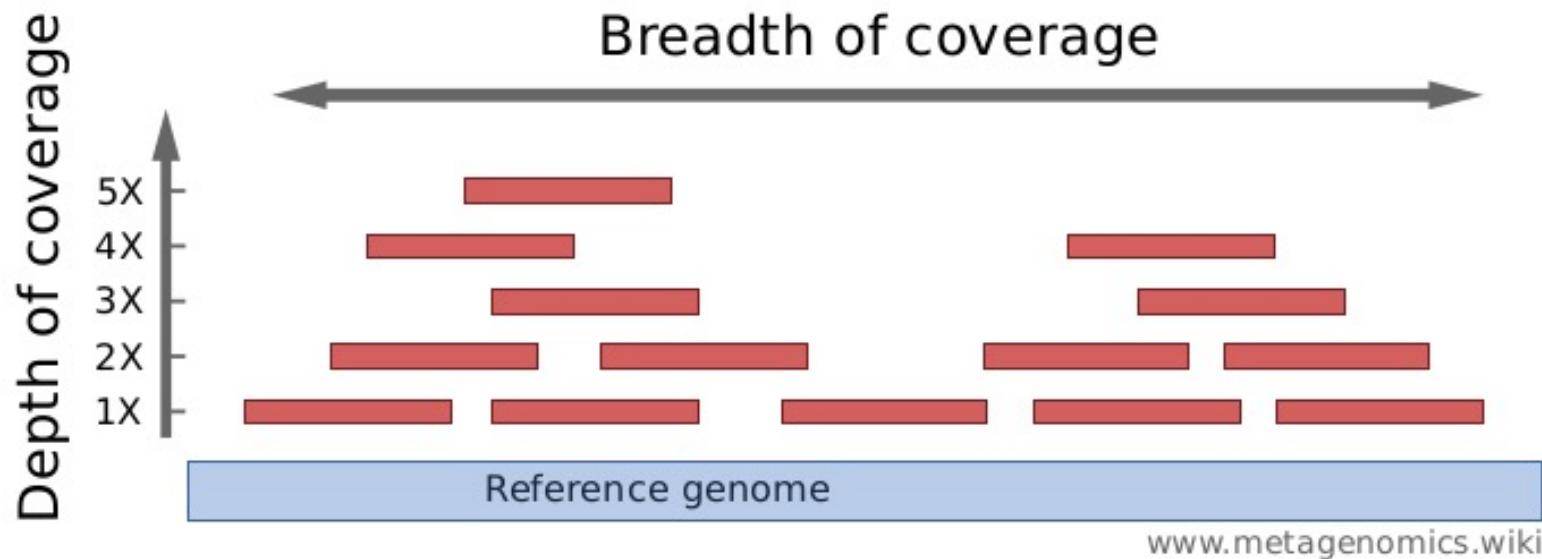
sequenced fragment unknown sequence sequenced fragment



Coverage



Coverage



Depth of coverage

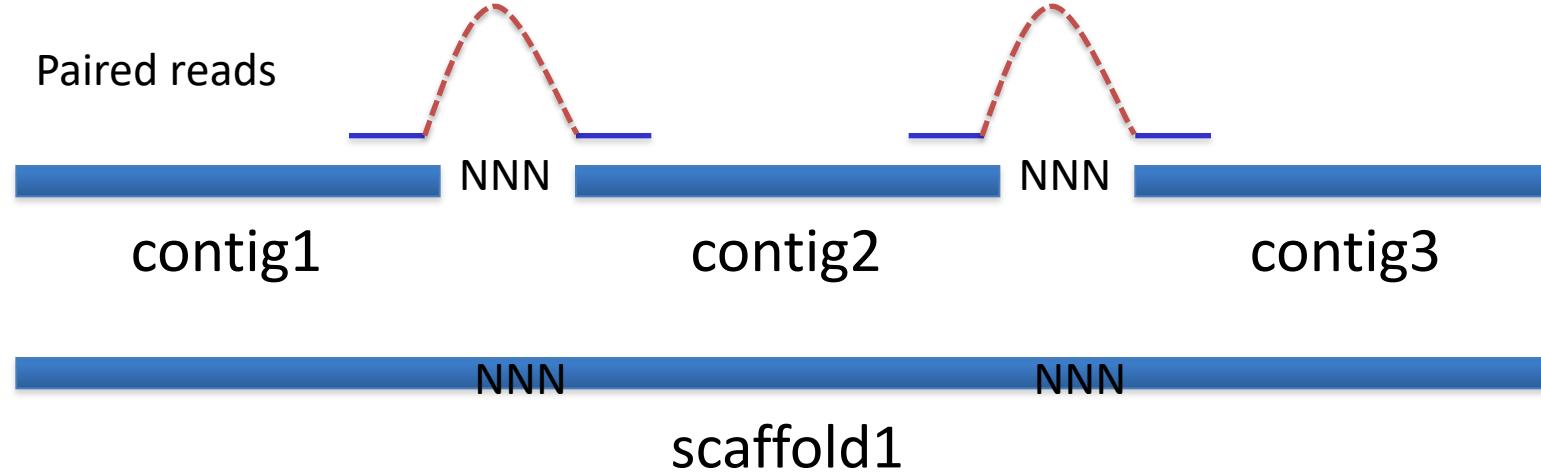
How strong is a genome "covered" by sequenced fragments (short reads)?

Breadth of coverage

How much of a genome is "covered" by short reads? Are there regions that are not covered, even not by a single read?

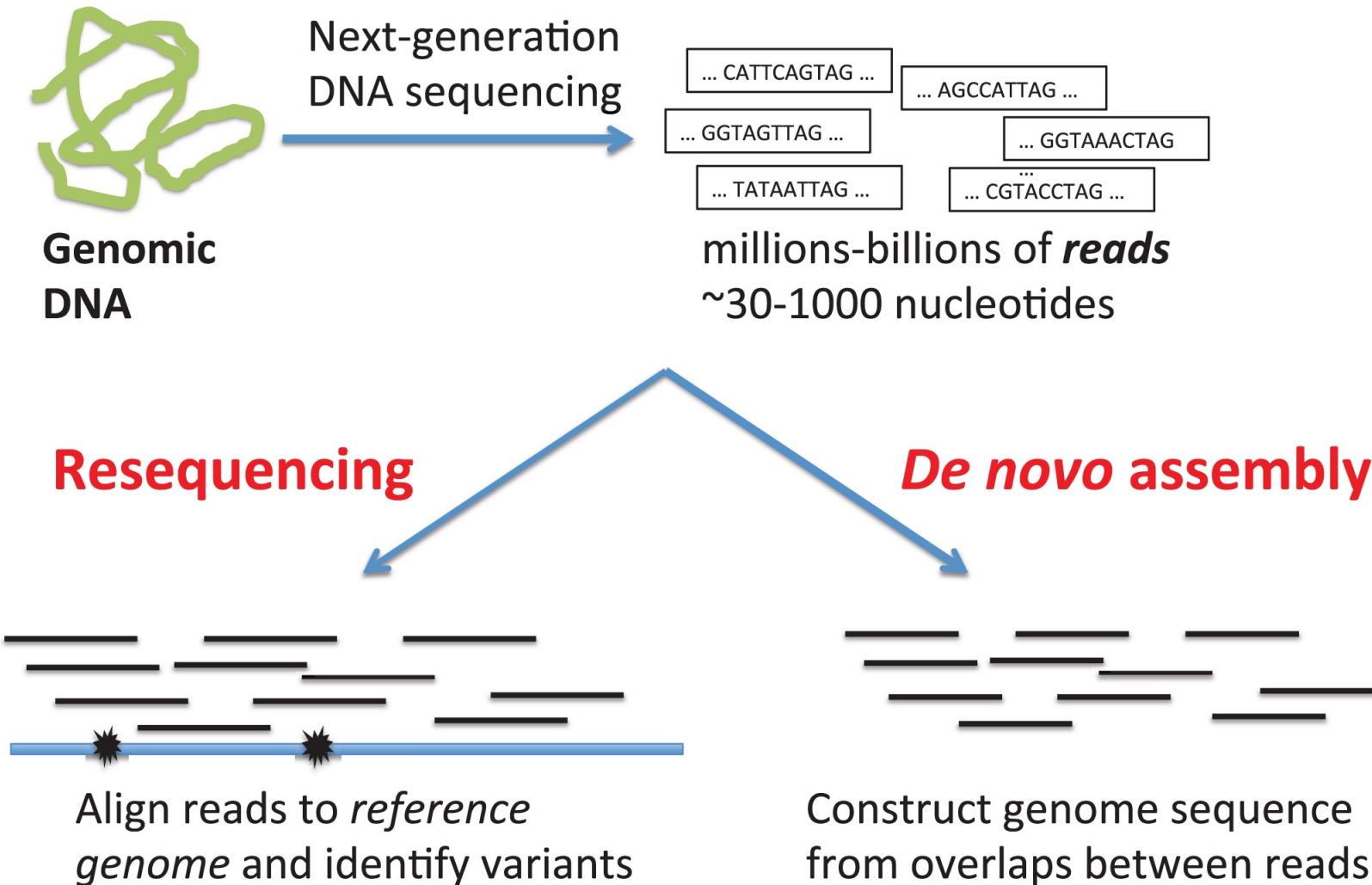
Gaps

(Genome areas not covered by random shotgun)



Contig = a contiguous stretch of nucleotides resulting from the assembly of several reads
Scaffold = several contigs stitched together with NNNs in between

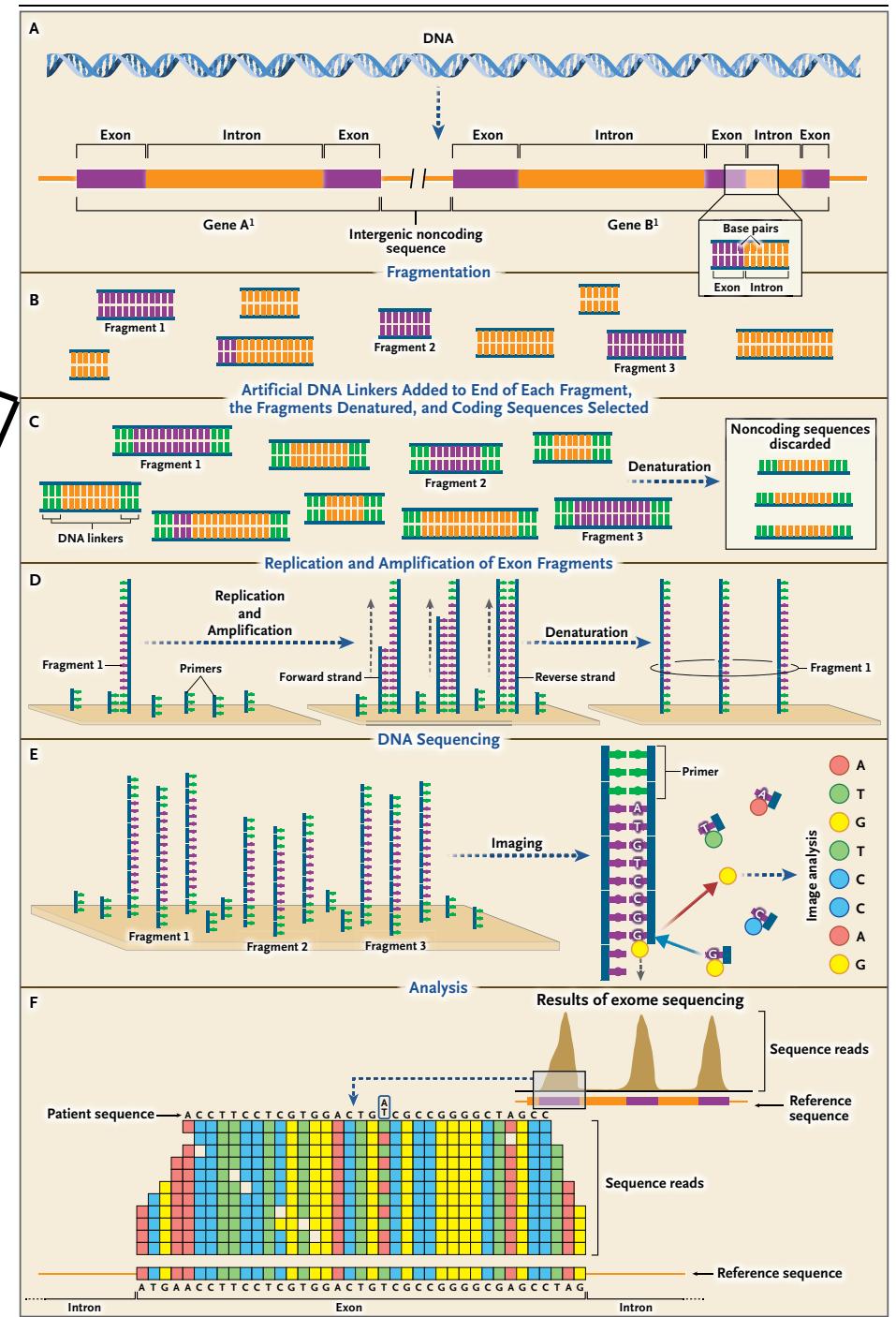
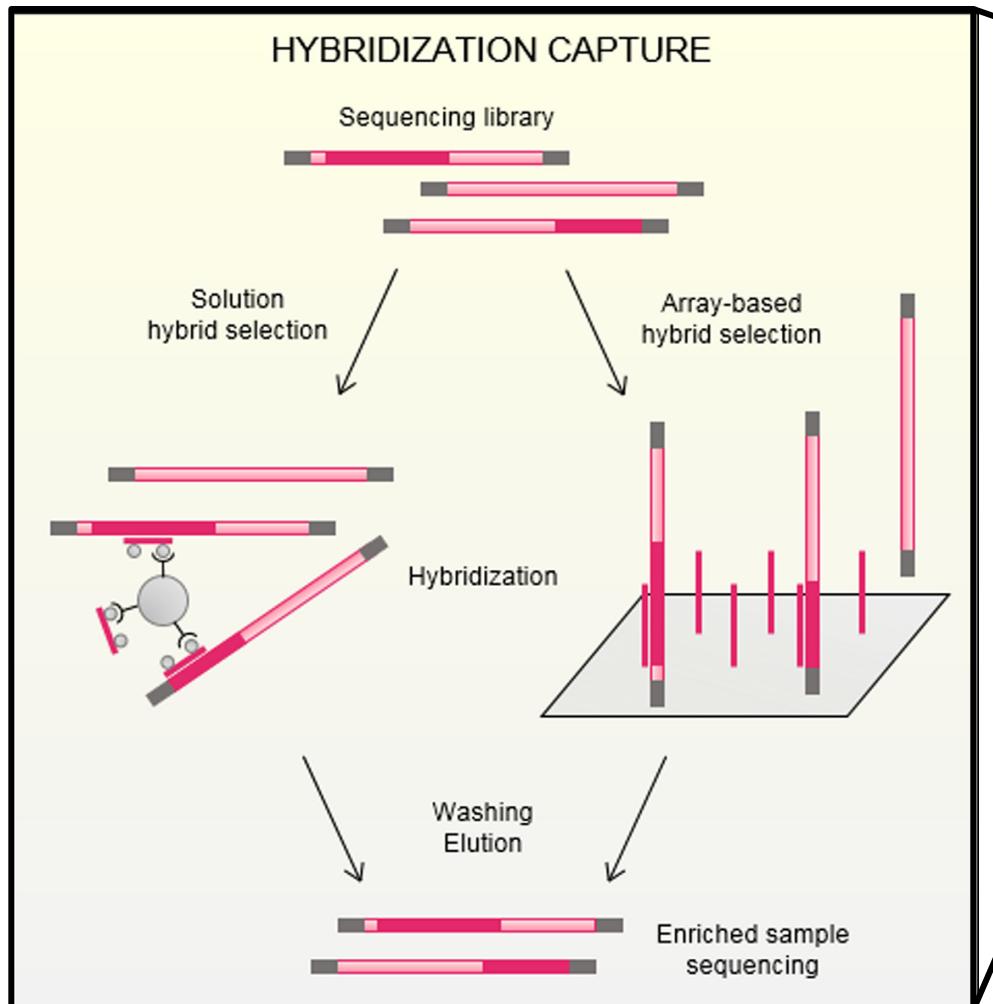
Types of sequencing



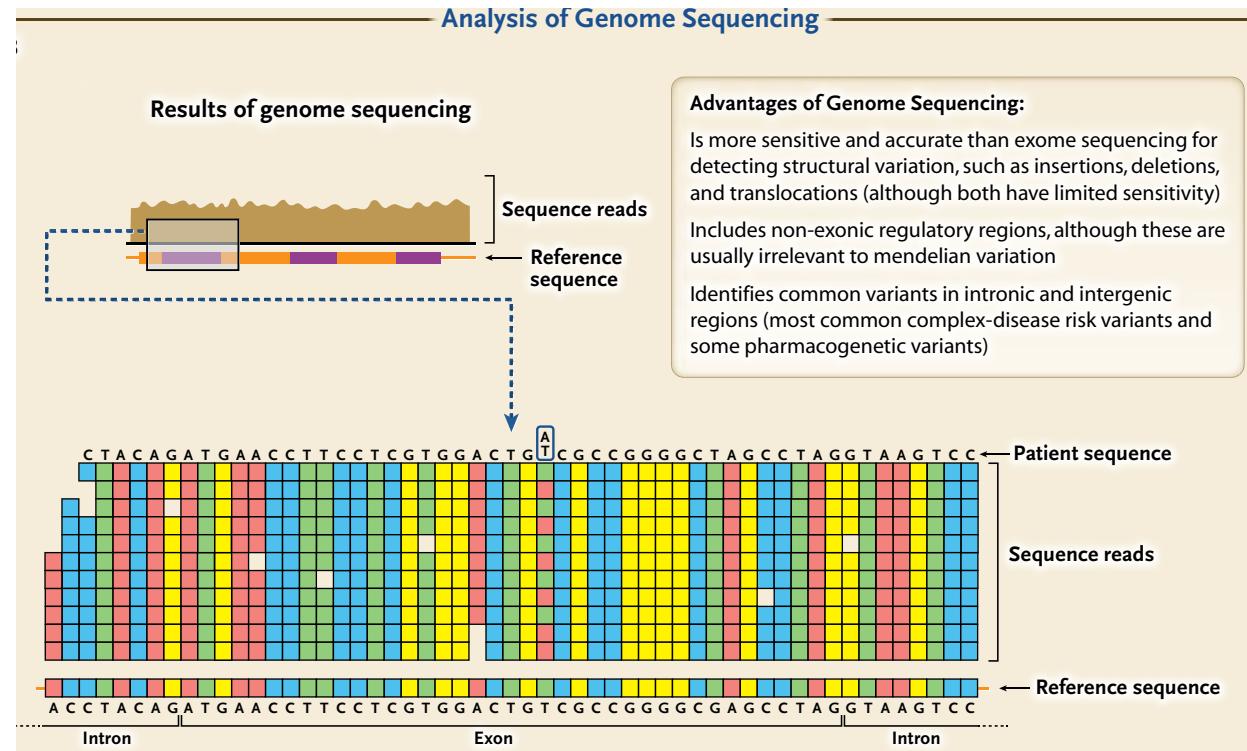
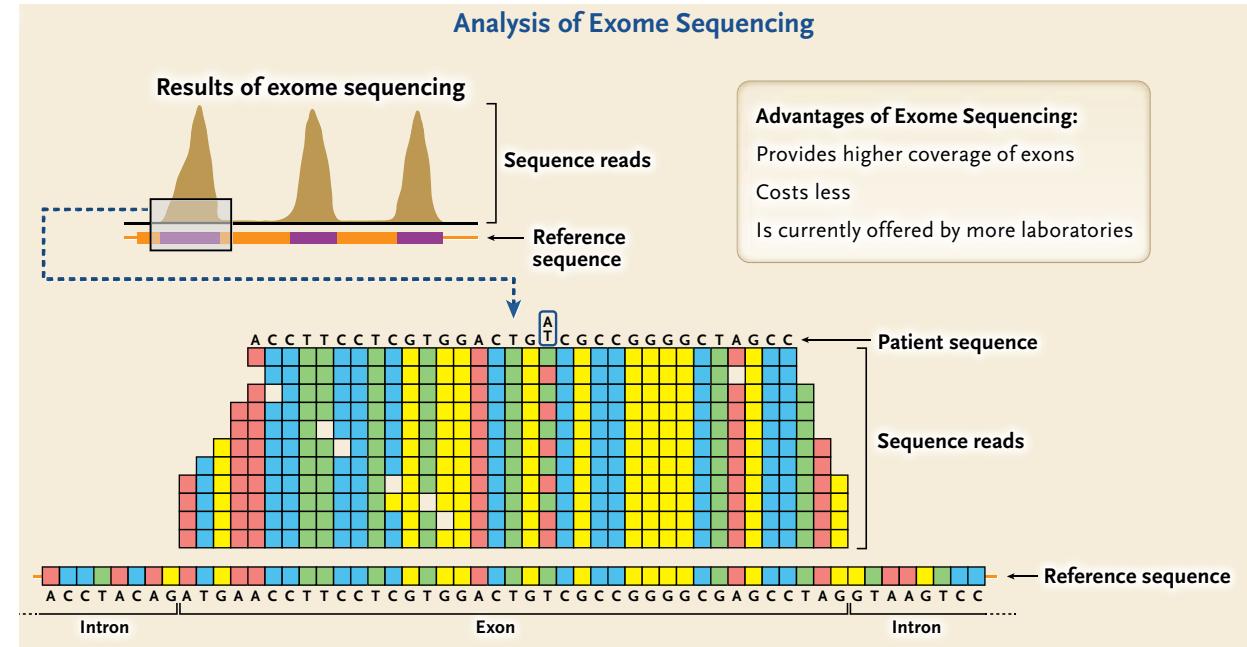
Targeted sequencing: Exome sequencing

- Whole genome sequencing
 - Redundant raw data (6Gb in each human diploid genome)
- Exome sequencing (targeted exome capture)
 - The sequence of all exons of the genome
 - Focuses on the part of the genome we understand best, the exons of the genes
 - ~85% of all known disease causing mutations are located on exons
 - Exome sequencing costs 1/6 of the cost of whole genome sequencing
 - Exons are short and 180,000 exons constitute 1% of the human genome
 - The goal is to identify the functional variation that is responsible for both mendelian and common diseases

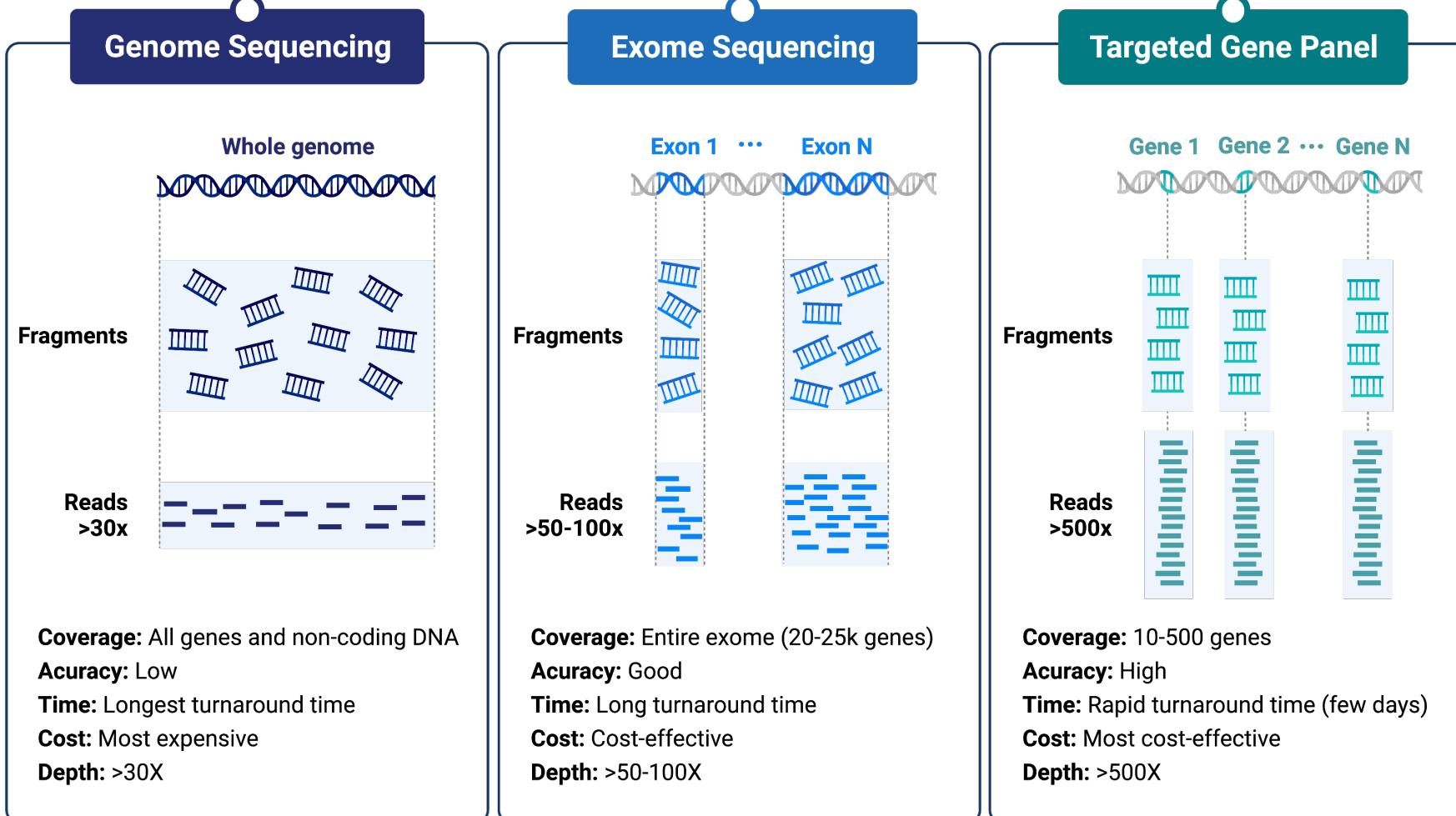
Exome sequencing steps



Analysis: Exome VS WGS



Next Generation Sequencing



*Template adapted from: Dr. Roshini Abraham
Clinical Immunologist at Nationwide Children's Hospital*

Useful Videos

Nanopore ([Video](#))

Illumina ([Video](#))

Pacific Biosciences ([Video](#))

Ion Torrent ([Video](#))

ABI SOLiD ([Video](#))

Roche 454 ([Video](#))

Qiagen GeneReader ([Video](#))

Complete genomics BGI ([Video](#))