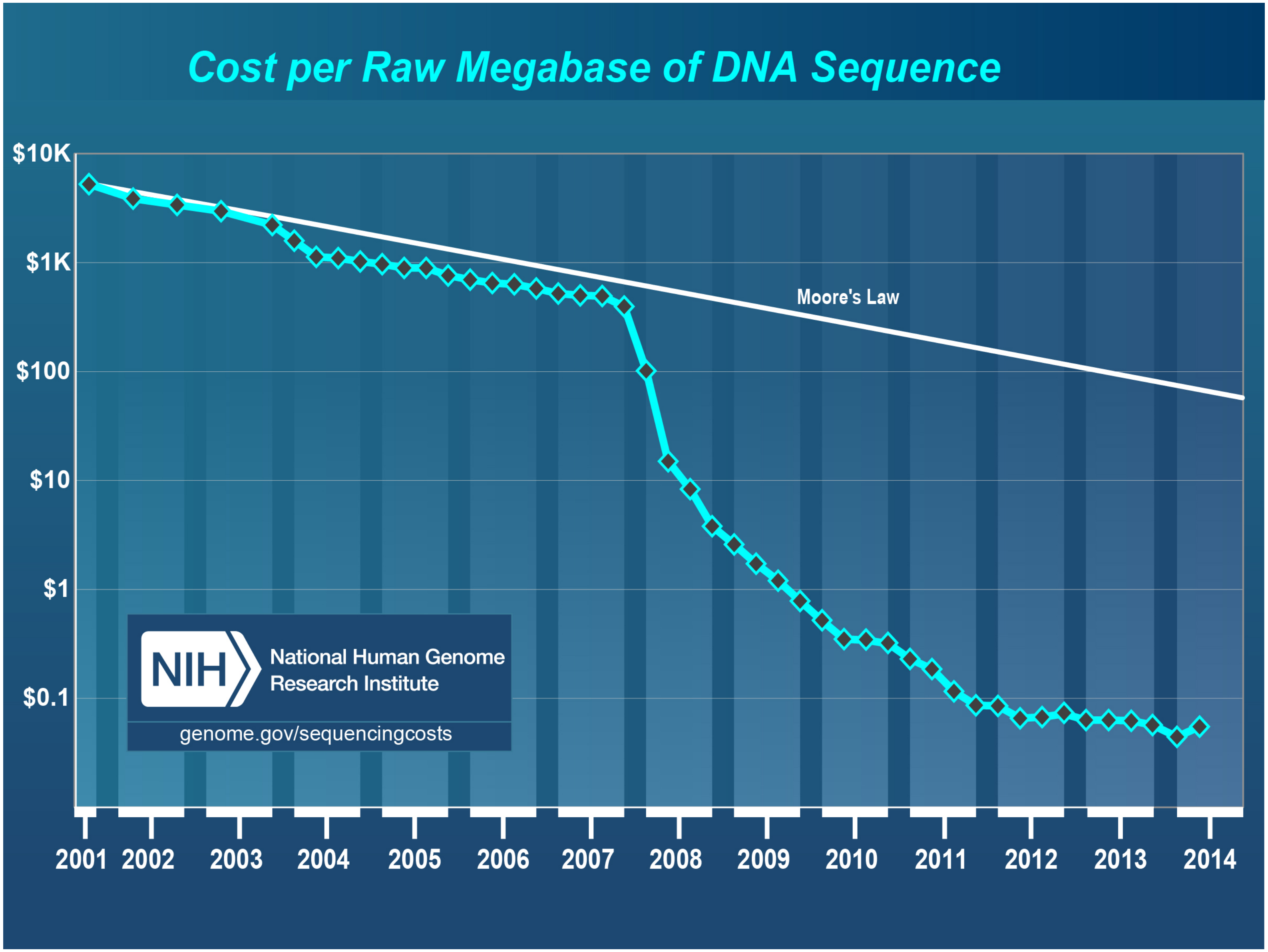
Next Generation Sequencing

October 8, 2014
Eric Chow
Center for Advanced Technology (CAT)
cat.ucsf.edu

Sequencing costs have dropped dramatically



Talk outline

- Traditional sequencing
- Next-generation sequencing
- Illumina sequencing
- NGS applications

Traditional Sequencing

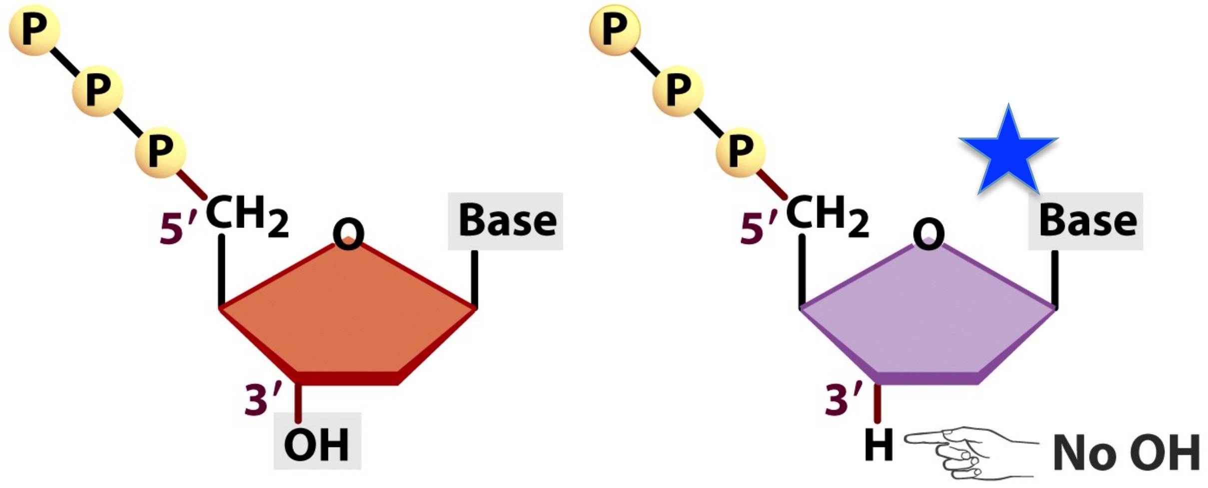
- Primer extension with labeled terminators
- 700 base read length

ACTAGCTGGACTCGTCACACT

Traditional Sequencing

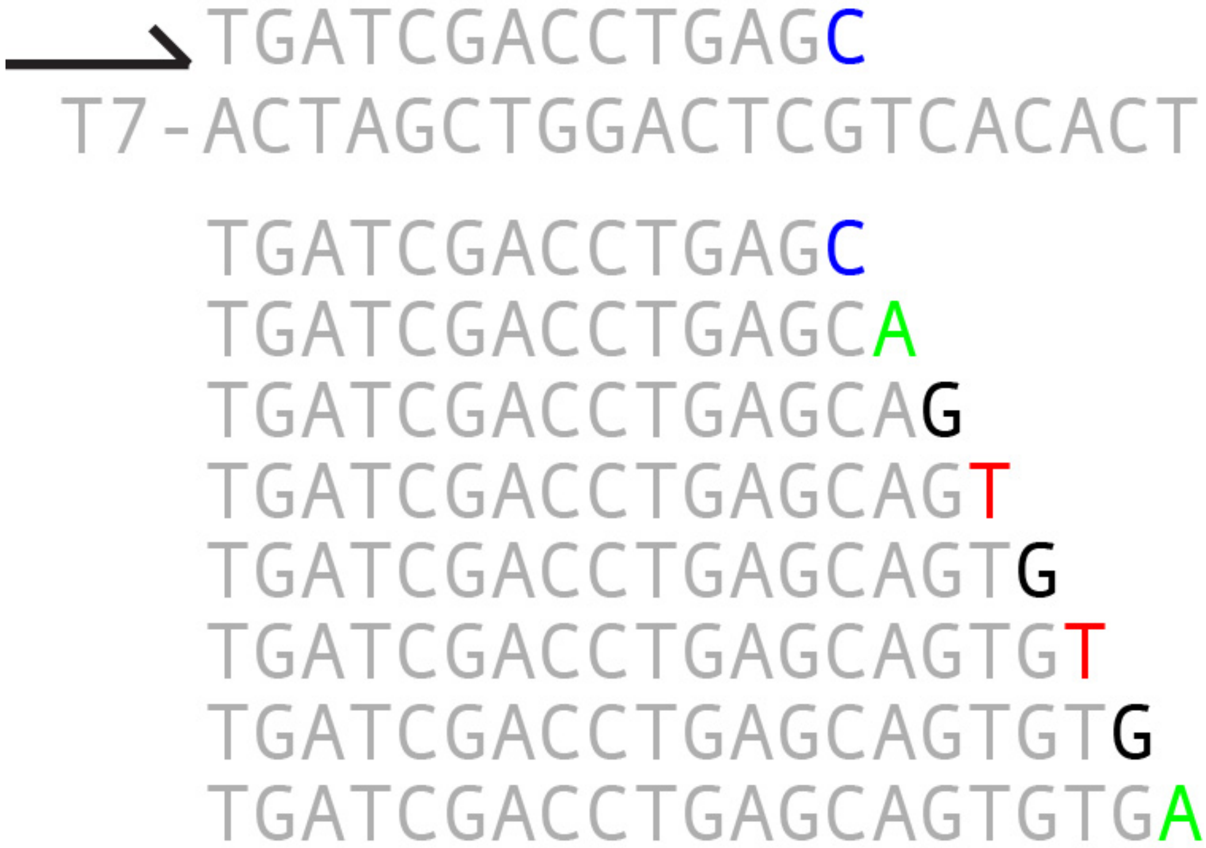
- Primer extension with labeled terminators
- 700 base read length


TGATCGACCTGAGC
T7-ACTAGCTGGACTCGTCACACT



Traditional Sequencing

- Primer extension with labeled terminators
- 700 base read length

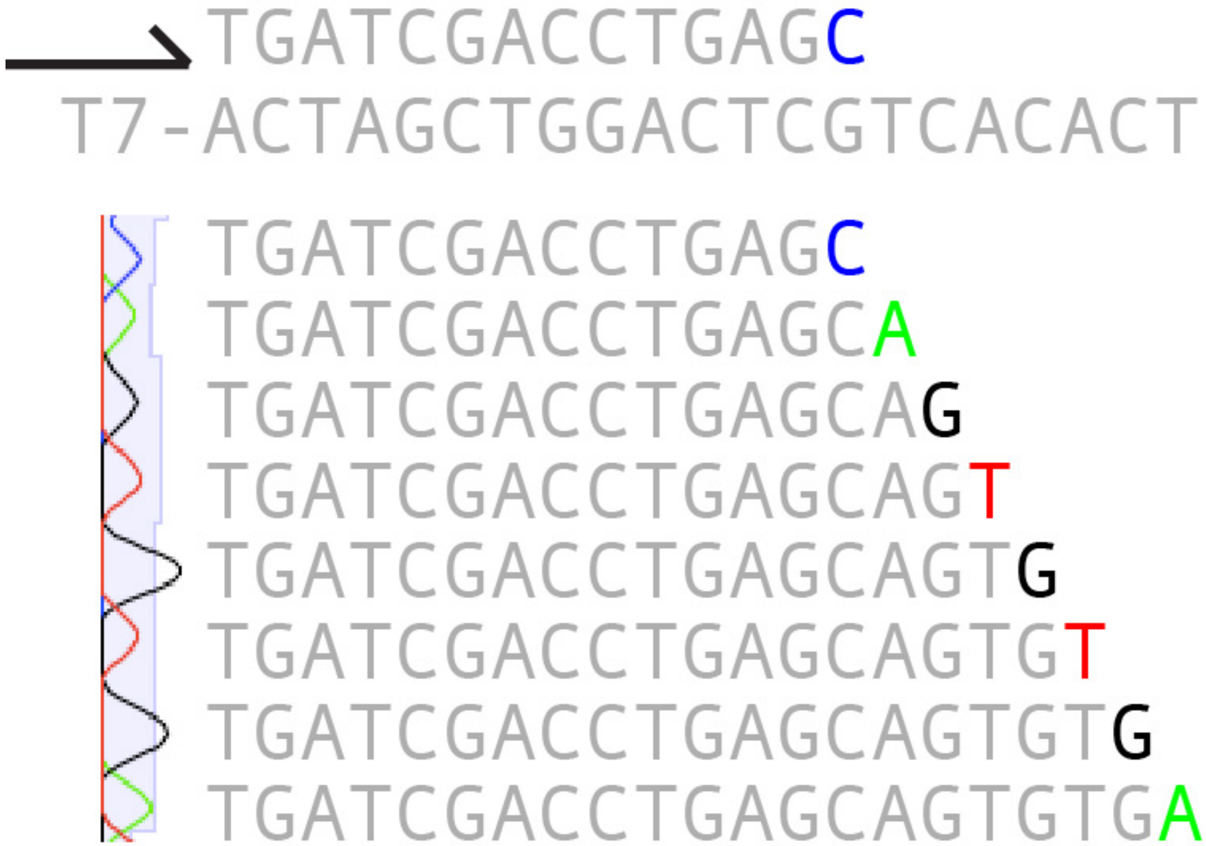


TGATCGACCTGAGC
T7-ACTAGCTGGACTCGTCACACT

TGATCGACCTGAGC
TGATCGACCTGAGCA
TGATCGACCTGAGCAG
TGATCGACCTGAGCAGT
TGATCGACCTGAGCAGTG
TGATCGACCTGAGCAGTGT
TGATCGACCTGAGCAGTGTG
TGATCGACCTGAGCAGTGTGA

Traditional Sequencing

- Primer extension with labeled terminators
- 700 base read length

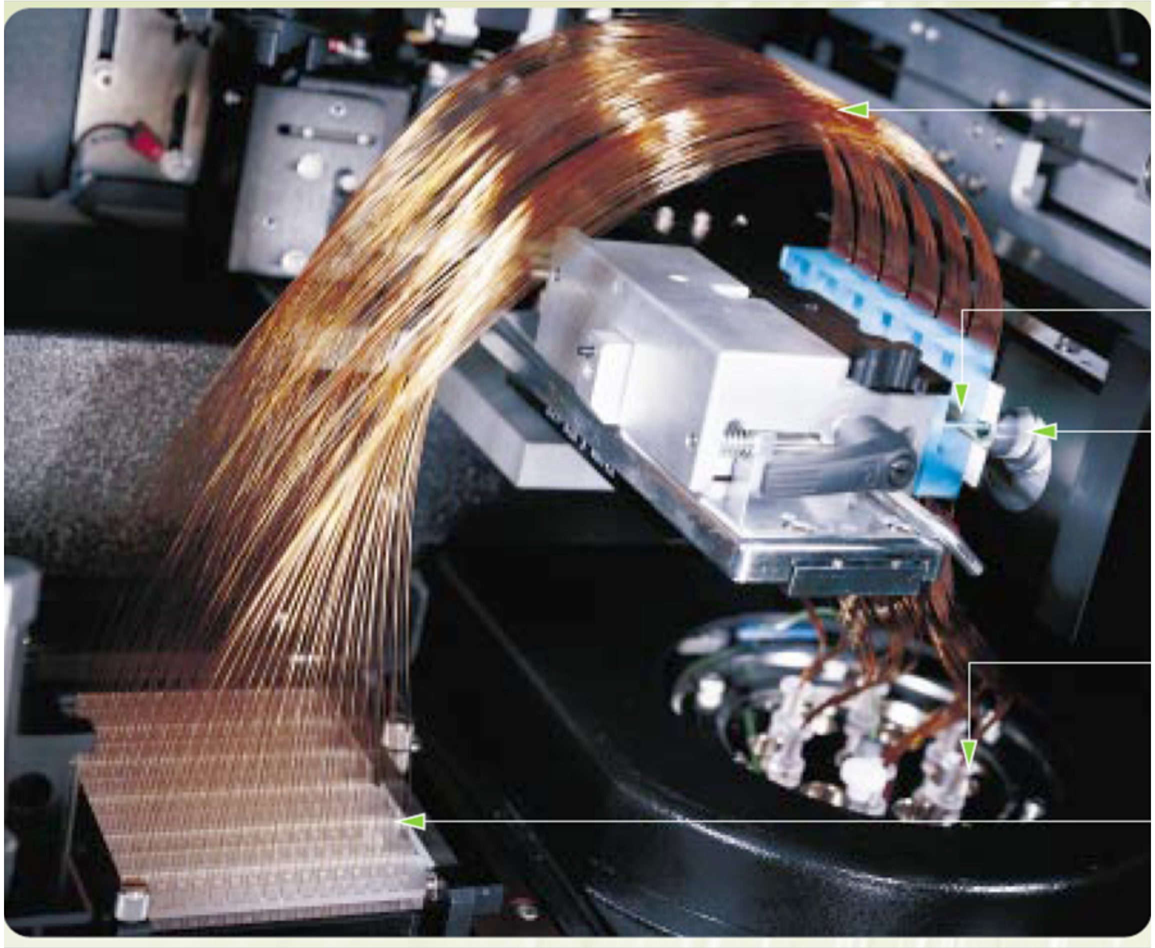


TGATCGACCTGAGC
T7-ACTAGCTGGACTCGTCACACT

TGATCGACCTGAGC
TGATCGACCTGAGCA
TGATCGACCTGAGCAG
TGATCGACCTGAGCAGT
TGATCGACCTGAGCAGTG
TGATCGACCTGAGCAGTGT
TGATCGACCTGAGCAGTGTG
TGATCGACCTGAGCAGTGTGA

Traditional Sequencing

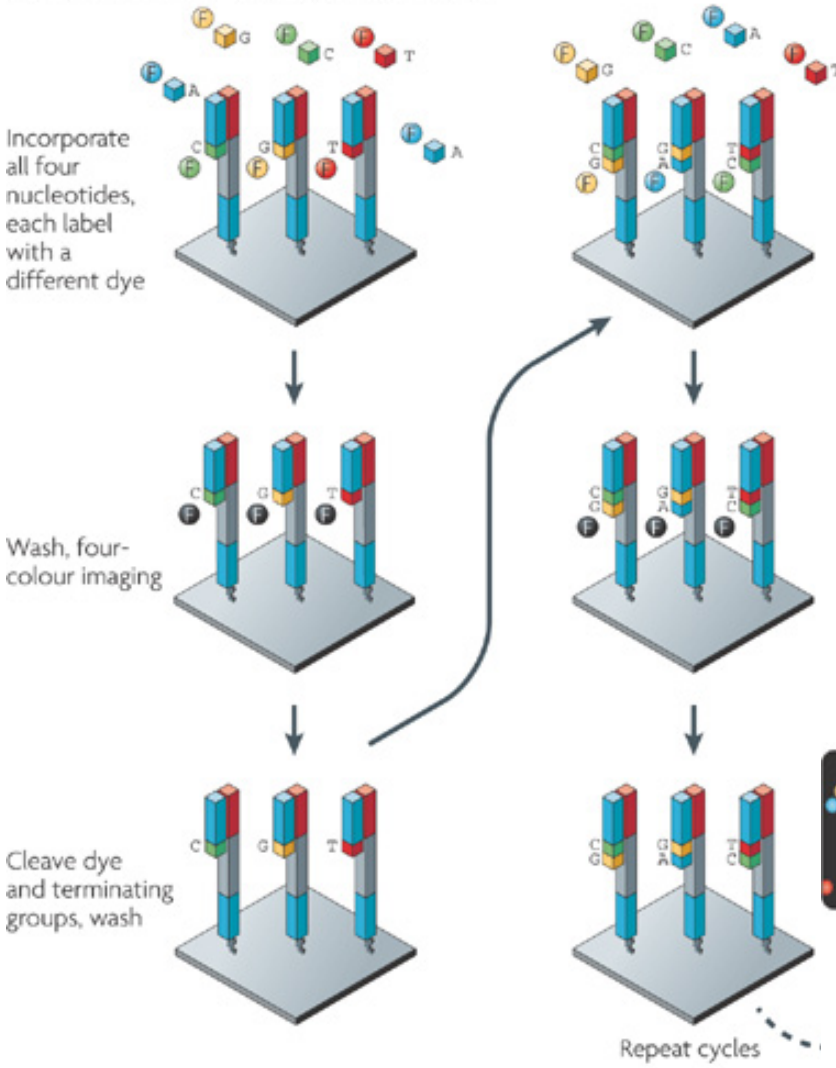
- Primer extension with labeled terminators
- 700 base read length



Next generation Massively Parallel Sequencing Technologies

- Many reads
- Next generation systems
  - Short read
    - Illumina – Sequencing by synthesis (dye)
    - Ion Torrent – sequencing by synthesis (pH)
  - Long read
    - Pacific Bioscience – single molecule (dye)
    - Oxford Nanopore – single molecule (current)

Illumina sequencing

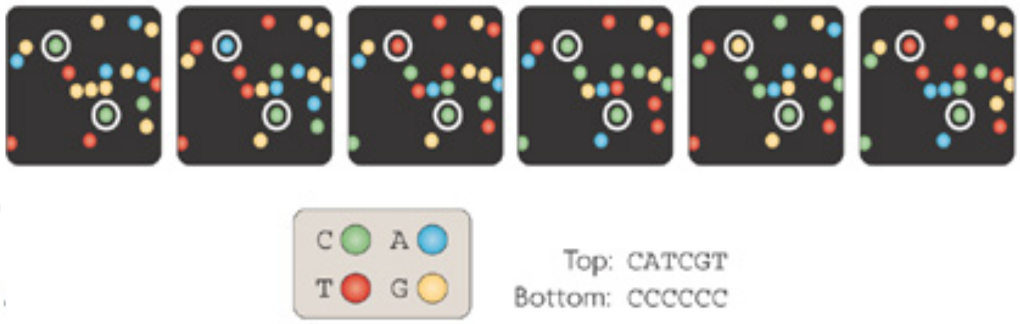


Leads the field

*Not single molecule sequencing!

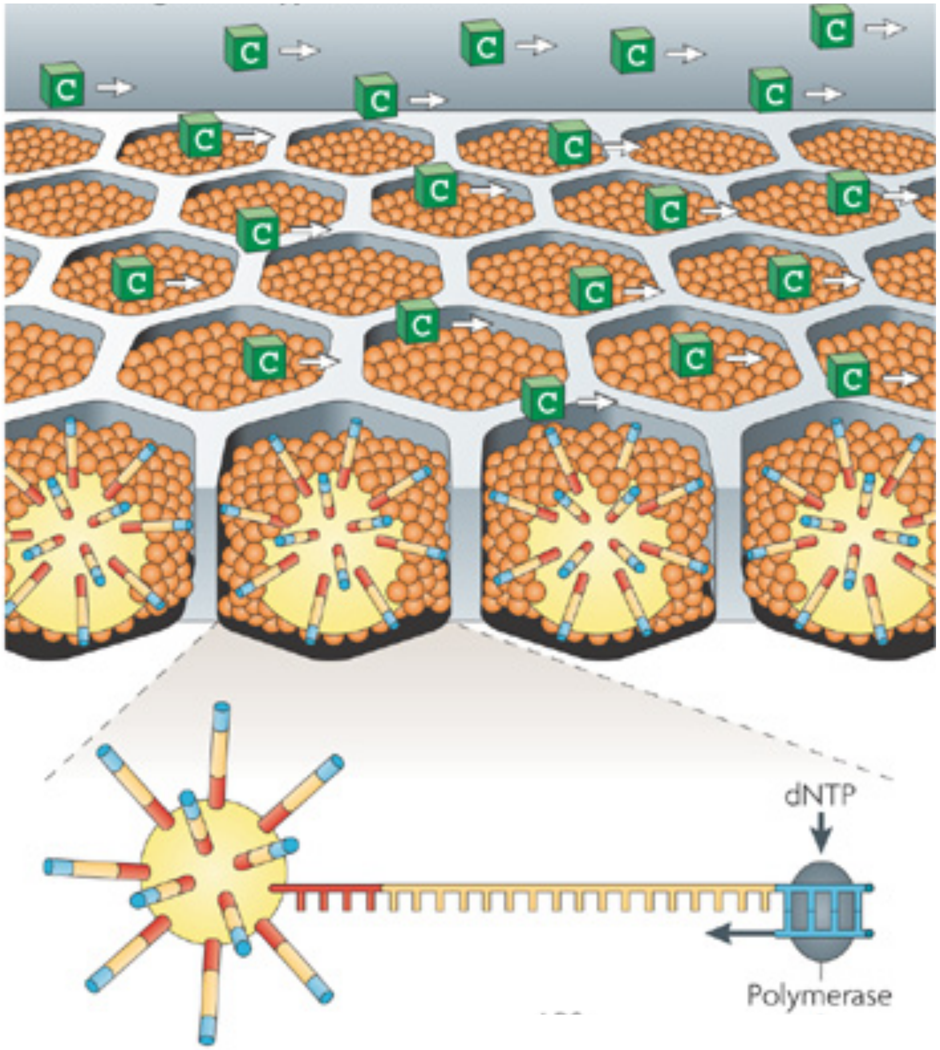
Reversible fluorescent terminators

Substitution errors



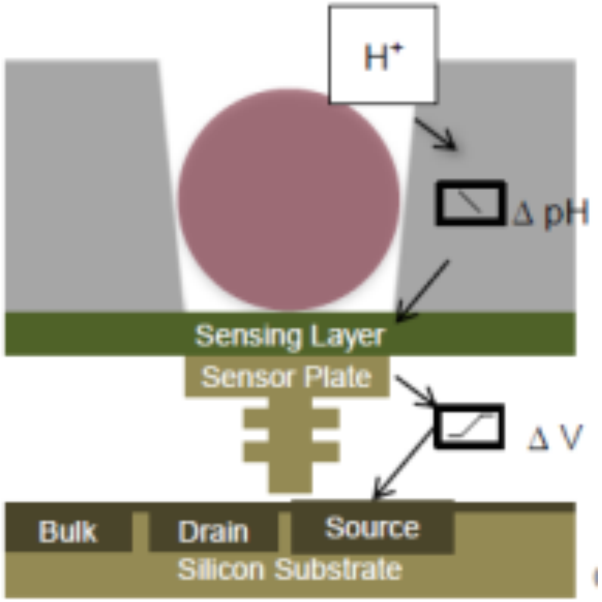
Nature Rev Gen

Ion Torrent



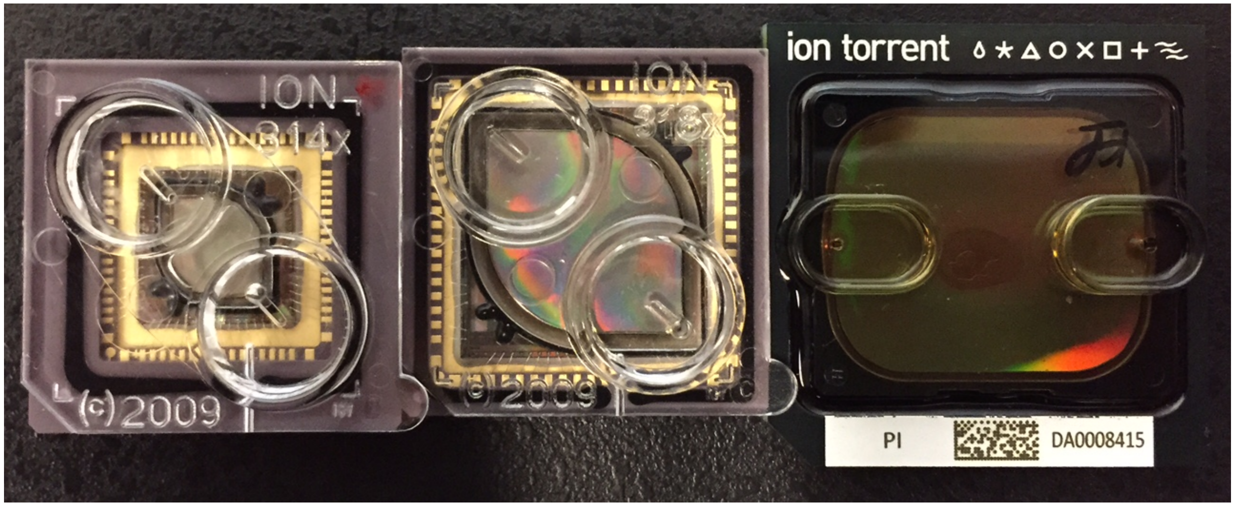
Nature Rev Gen

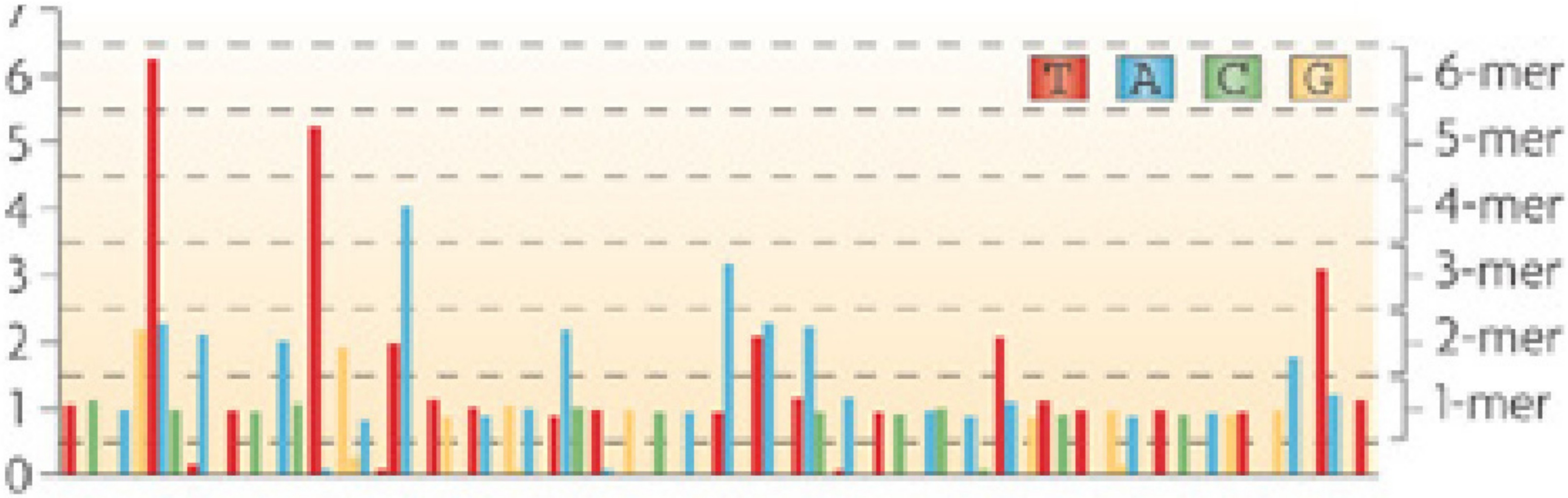
DNA(n) + dNTP → DNA(n+1) + PPi + H+



PGM		Proton	
314	0.6M	PI	82M
316	3M	PII	330M
318	5.5M	PIII	660M

Ion Torrent





Nature Rev Gen

Ion Torrent

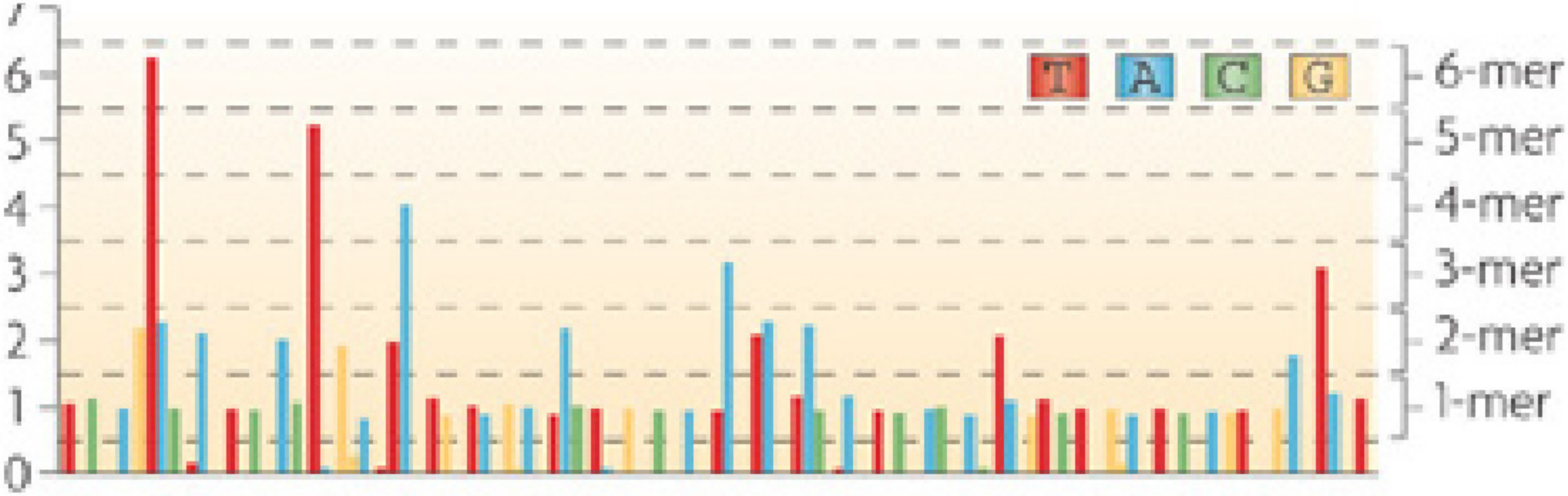
TCAGTACATCACTGATATAT
11126212112152124111

Nature Rev Gen

Ion Torrent

TCAGGTTTTTTAACAATCAACTTTTTGGATTAAAATGT

TCAGTACATCACTGATATAT
11126212112152124111

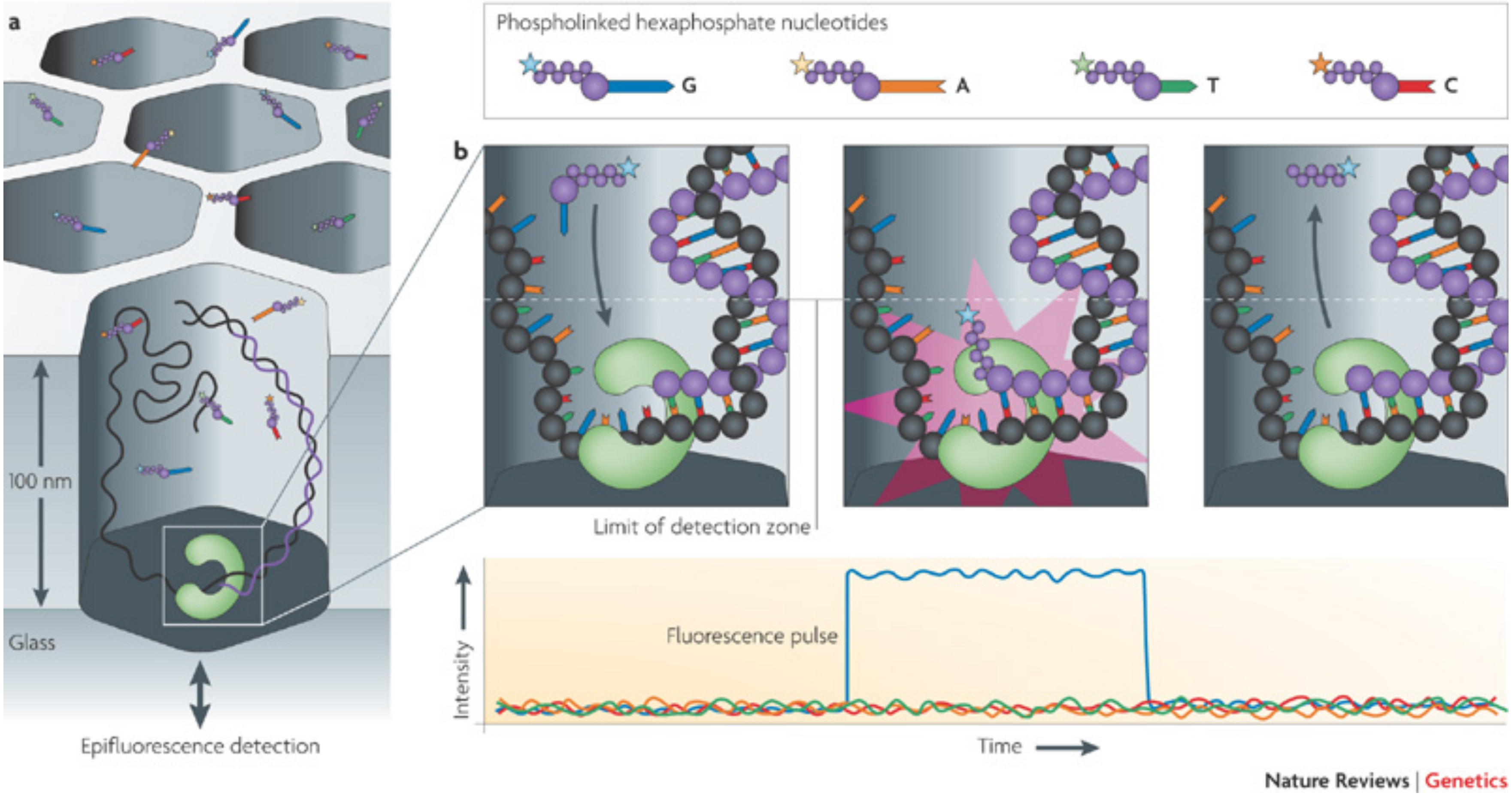


Insertion/deletion errors

Problems with scaling up

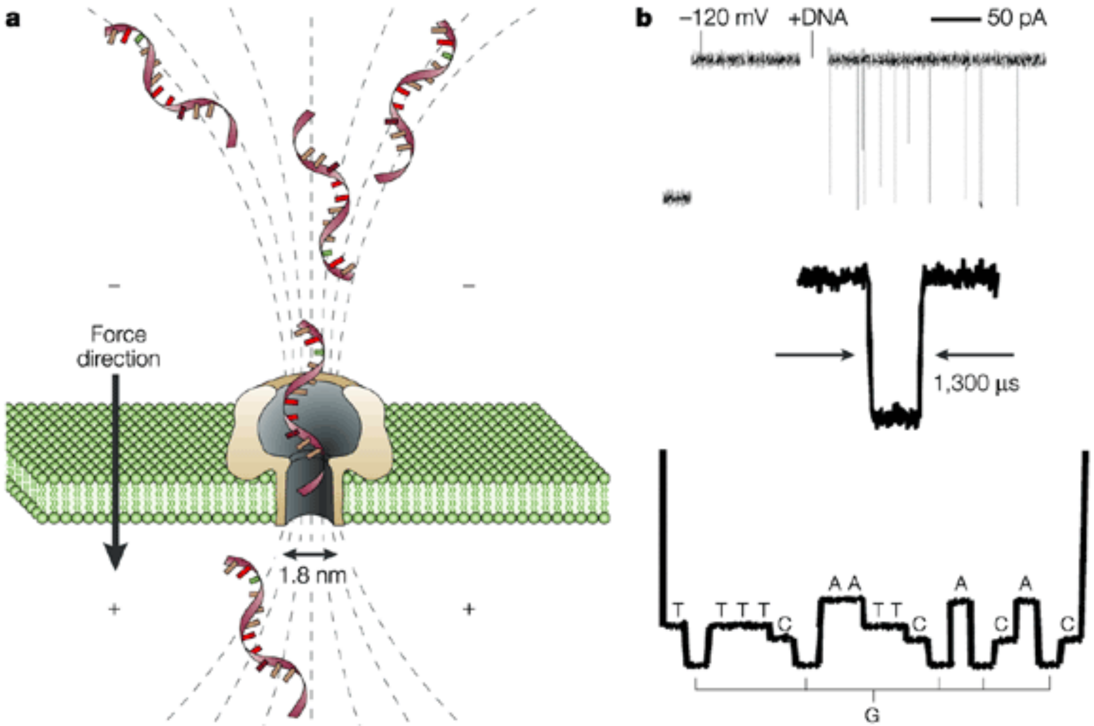
Nature Rev Gen

Pacific Biosciences



Single molecule sequencing
20 kB read lengths
15% error rate

Oxford Nanopore



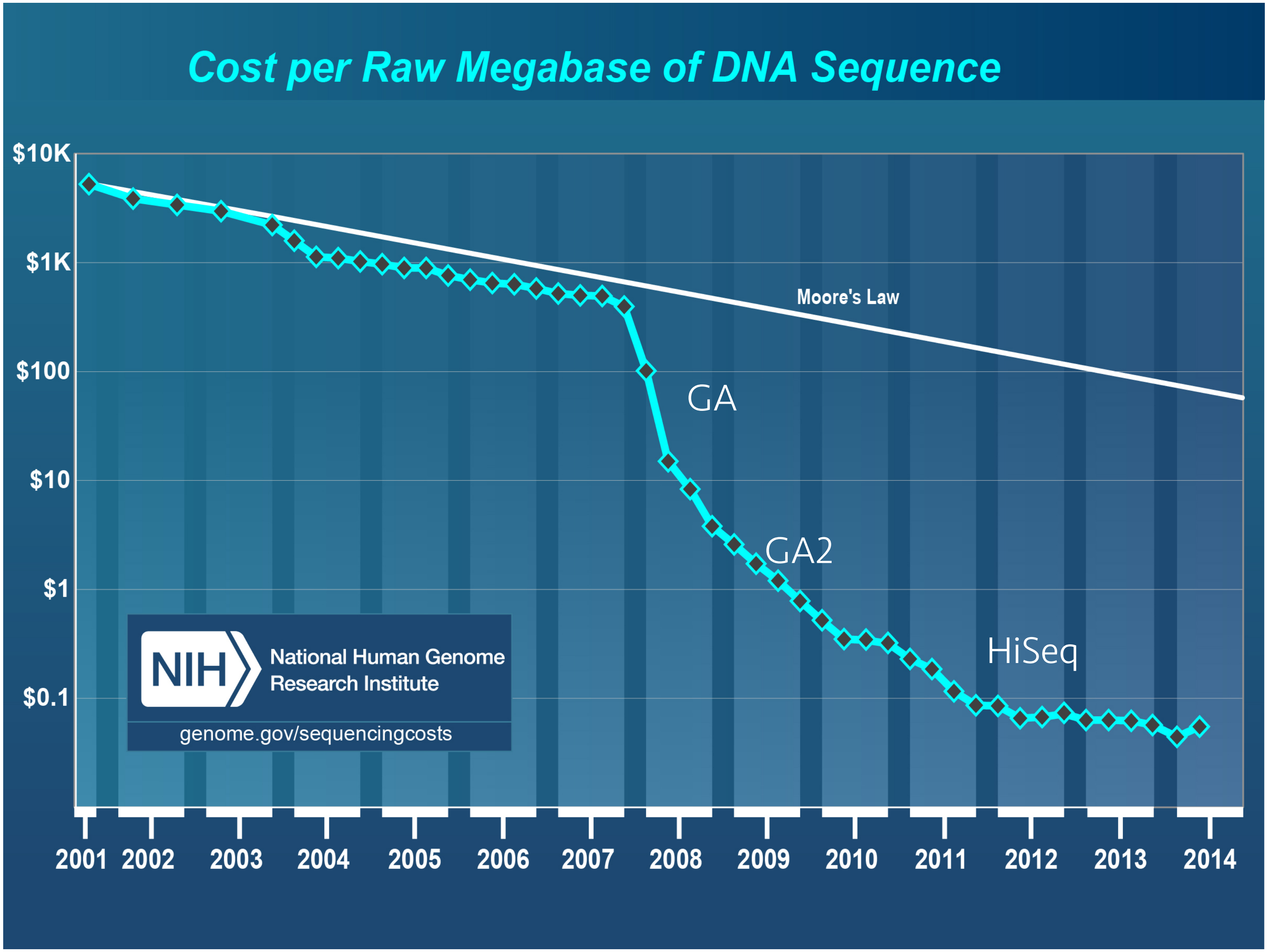
Nature Reviews | Drug Discovery

Single molecule sequencing
100 kB read lengths
15% error rate
Can detect DNA modifications

Which system to use?

- Depends on the application
  - Counting experiments (RNA/ChIP-Seq)
  - Genome assembly
  - Structural rearrangements
- Read numbers and length
  - Numbers– Illumina, Ion Torrent
  - Length – Pac Bio, Oxford Nanopore
- Error rates, types, bias
  - Illumina → substitution
  - Ion Torrent → Indel
  - PacBio/Oxford → high error rates (10-15%)

Sequencing costs have dropped dramatically



Instrument throughput has increased exponentially

Platform	Bases/ read	Reads/ run	Bases/run	Run/ day	Bases/day	cost/ MB
Sanger	700	96	6.7x10^5	24	1.6x10^6	$500
Illumina GAII	150	10^8	1.5x10^10	1/8	1.8x10^9	$0.520
Illumina HiSeq	250	4x10^9	1x10^12	1/6	1.6x10^11	$0.029
Illumina XTen	300	6x10^9	1.8x10^12	1/3	6x10^11	$0.007

You can sequence anything

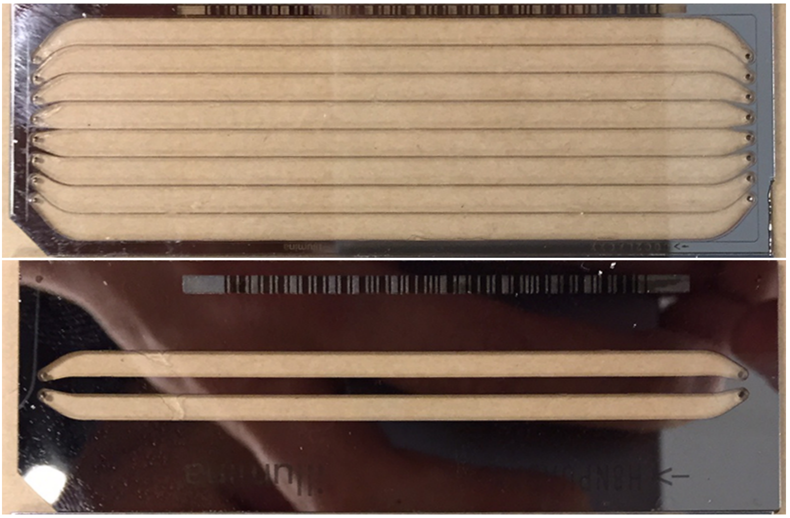
- Well, almost
  - Make DNA
  - Add adapters
  - Total size <1kb
- Many ways to add adapters
  - PCR
  - Ligation
  - Reverse transcription
  - Primer extension

You can sequence anything

- Counting applications
  - RNA-Seq
  - ChIP-Seq
  - Footprinting (ribosome, transcription,…)
- Genome
  - Whole-genome (was $1000s/genome)
  - Exome – pull out coding sequence
  - DNA methylation (bisulfite conversion)
  - metagenomics

Sequencing is really, really cheap

- $1000 human genome
- Non-invasive prenatal testing
- Large population studies
  - Pick out SNPs and mutations responsible for disease
  - Tumor/normal sequencing
  - Whole-genome sequencing will be universal
- Mostly due to Illumina (but watch out for others)



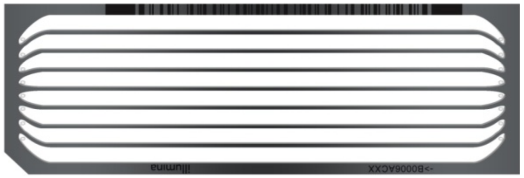
Flow cells, where the magic happens

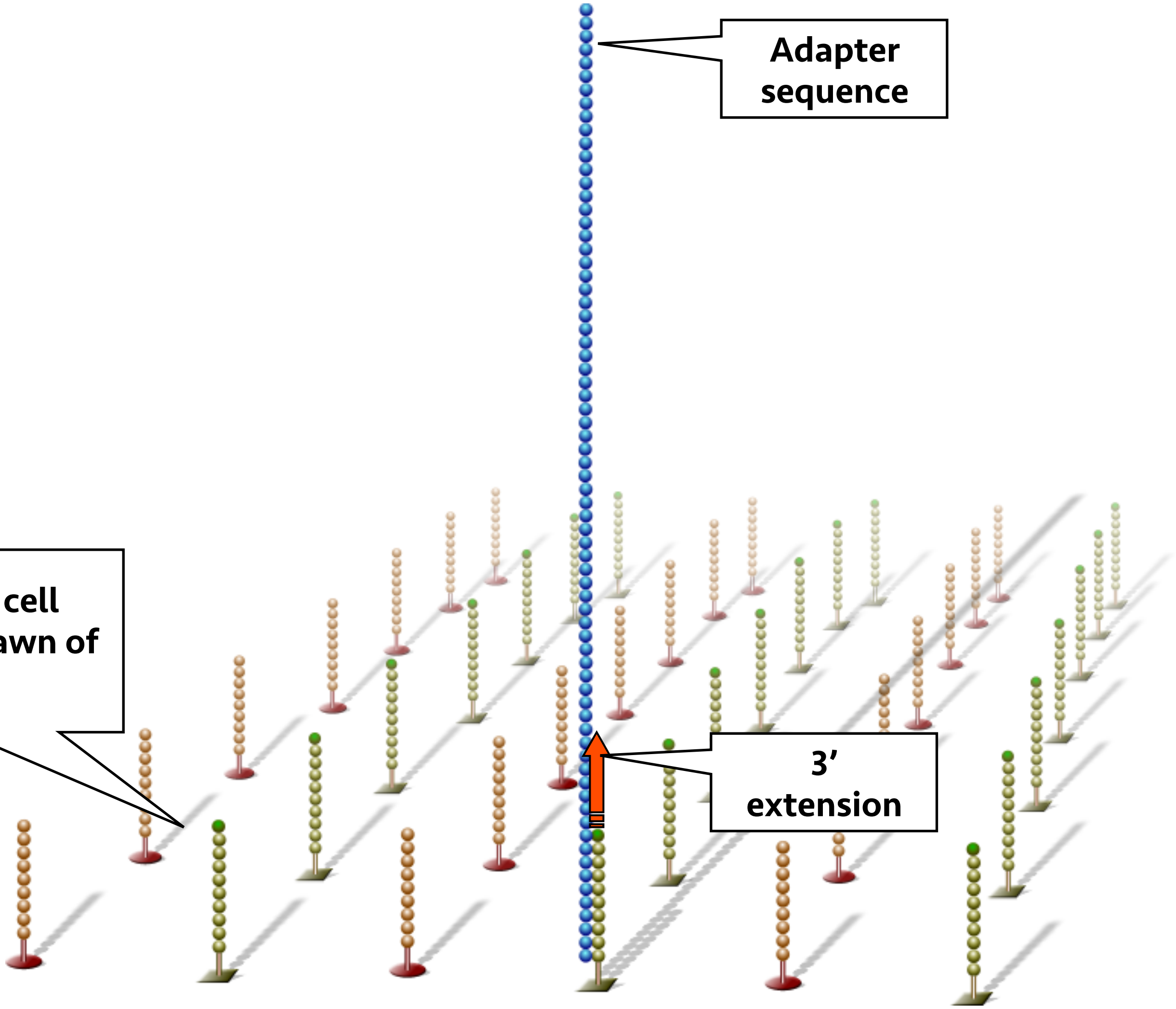
Illumina library

Illumina library

- Adapter sequence includes primer binding sites and capture sequences.
- Ion Torrent library adapters are very similar.

Library binding to flow cell



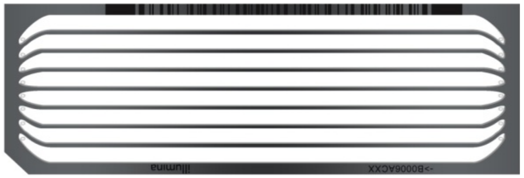


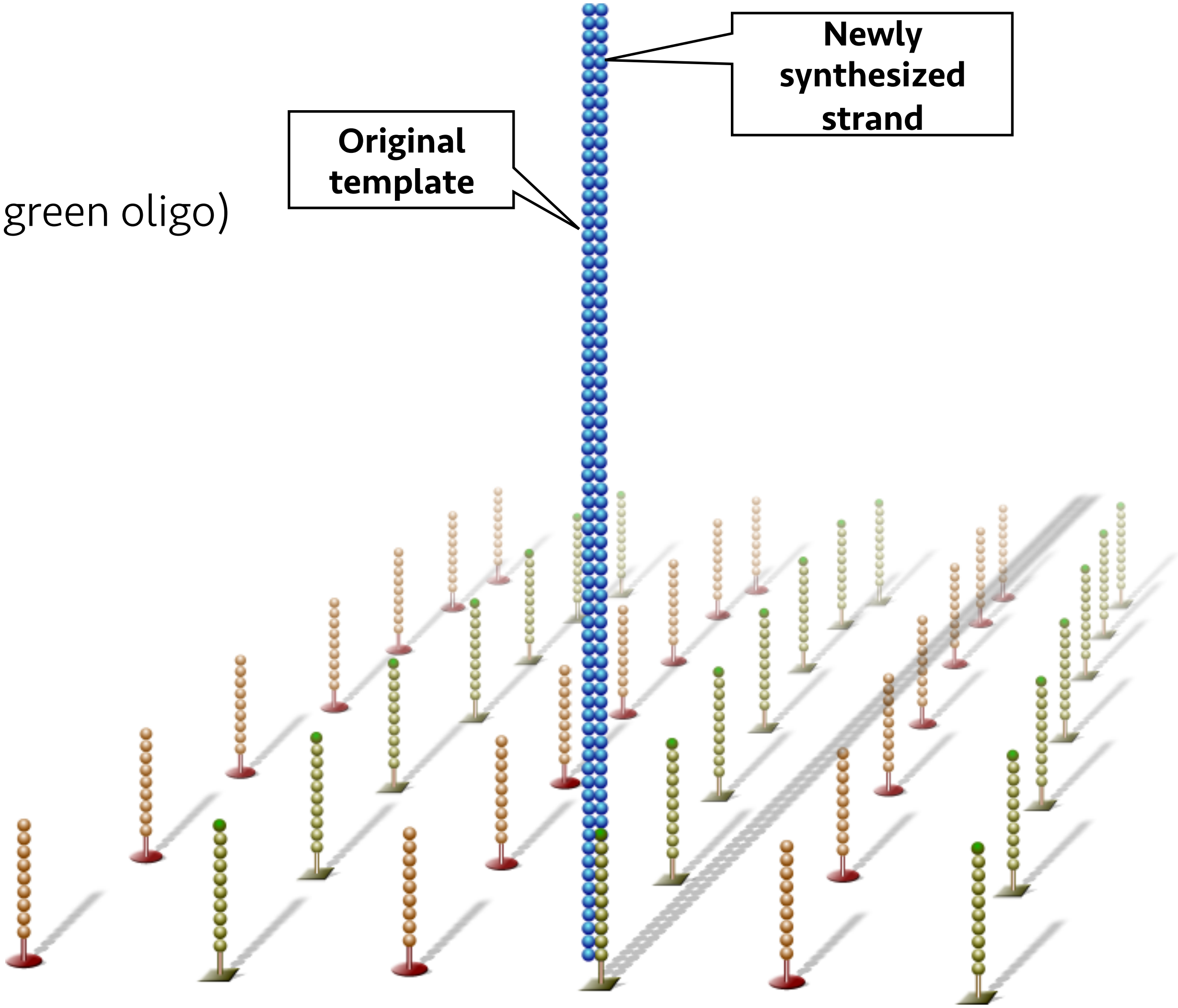
Adapter sequence

Surface of flow cell coated with a lawn of oligo pairs

3’ extension

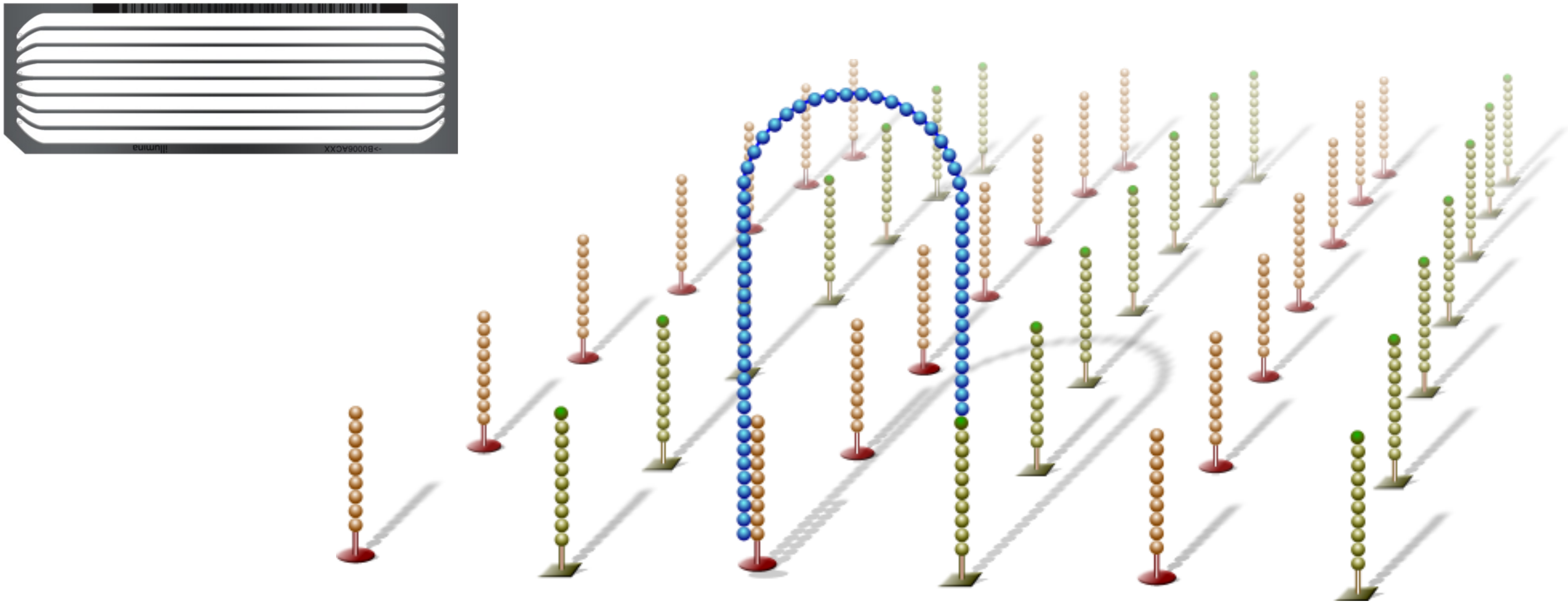
Reverse strand is created (green oligo)



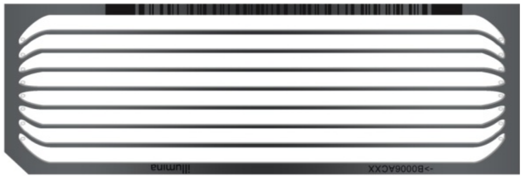


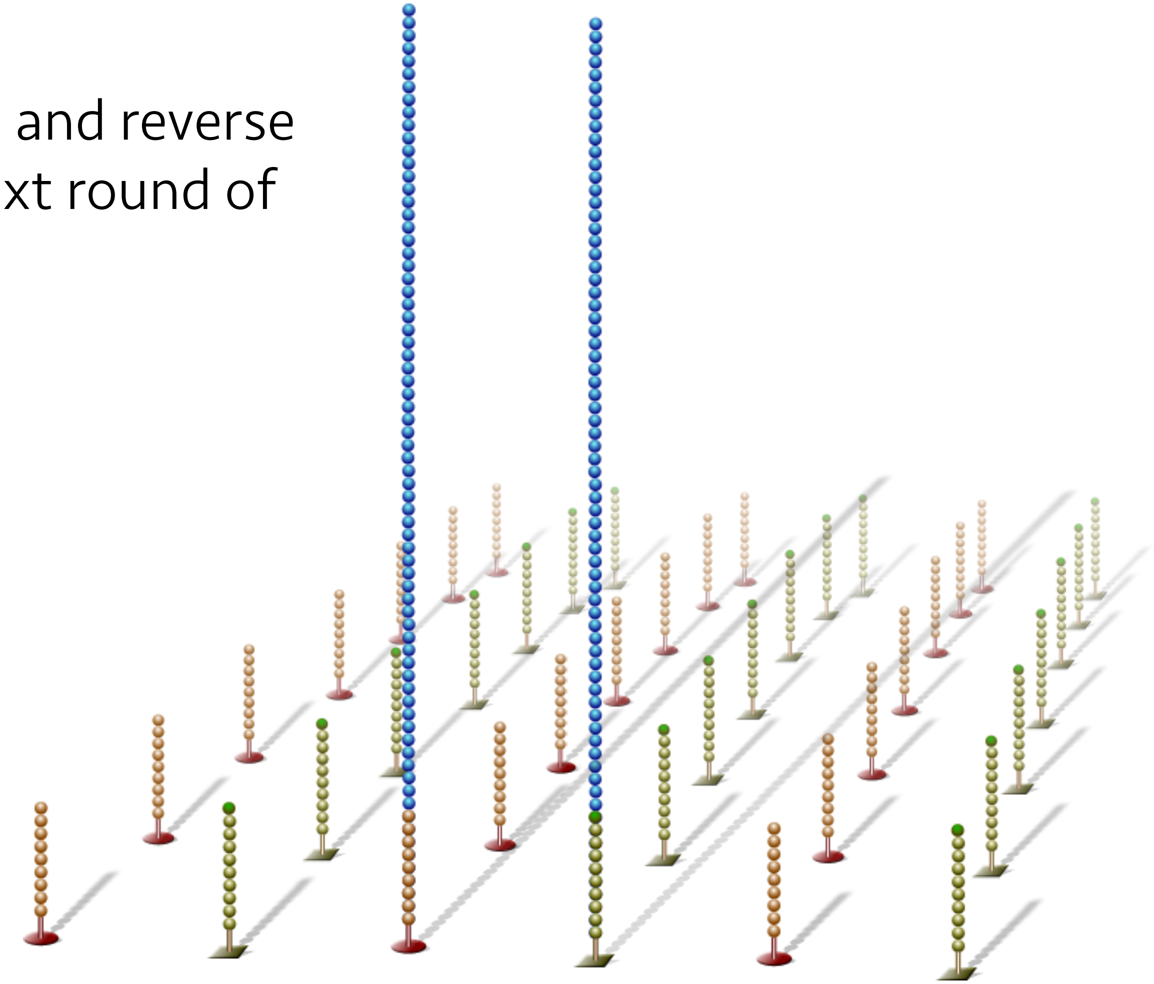
Newly synthesized strand

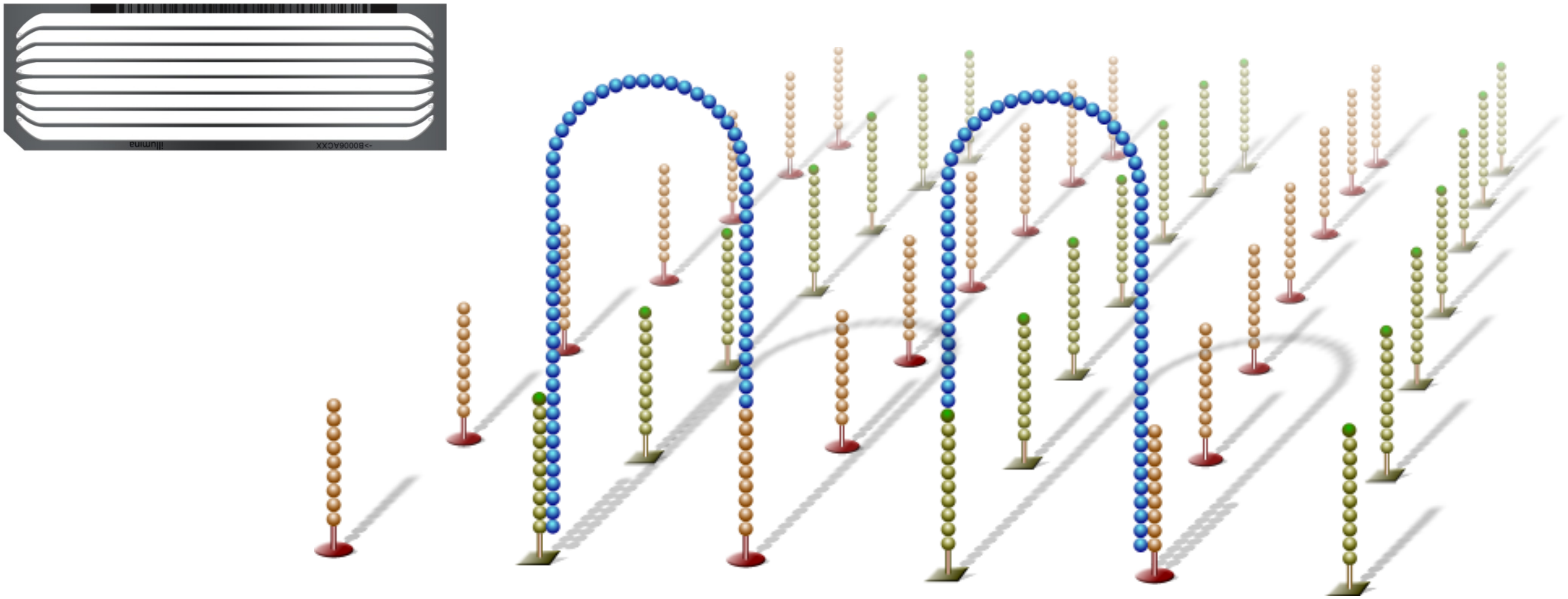
Original template



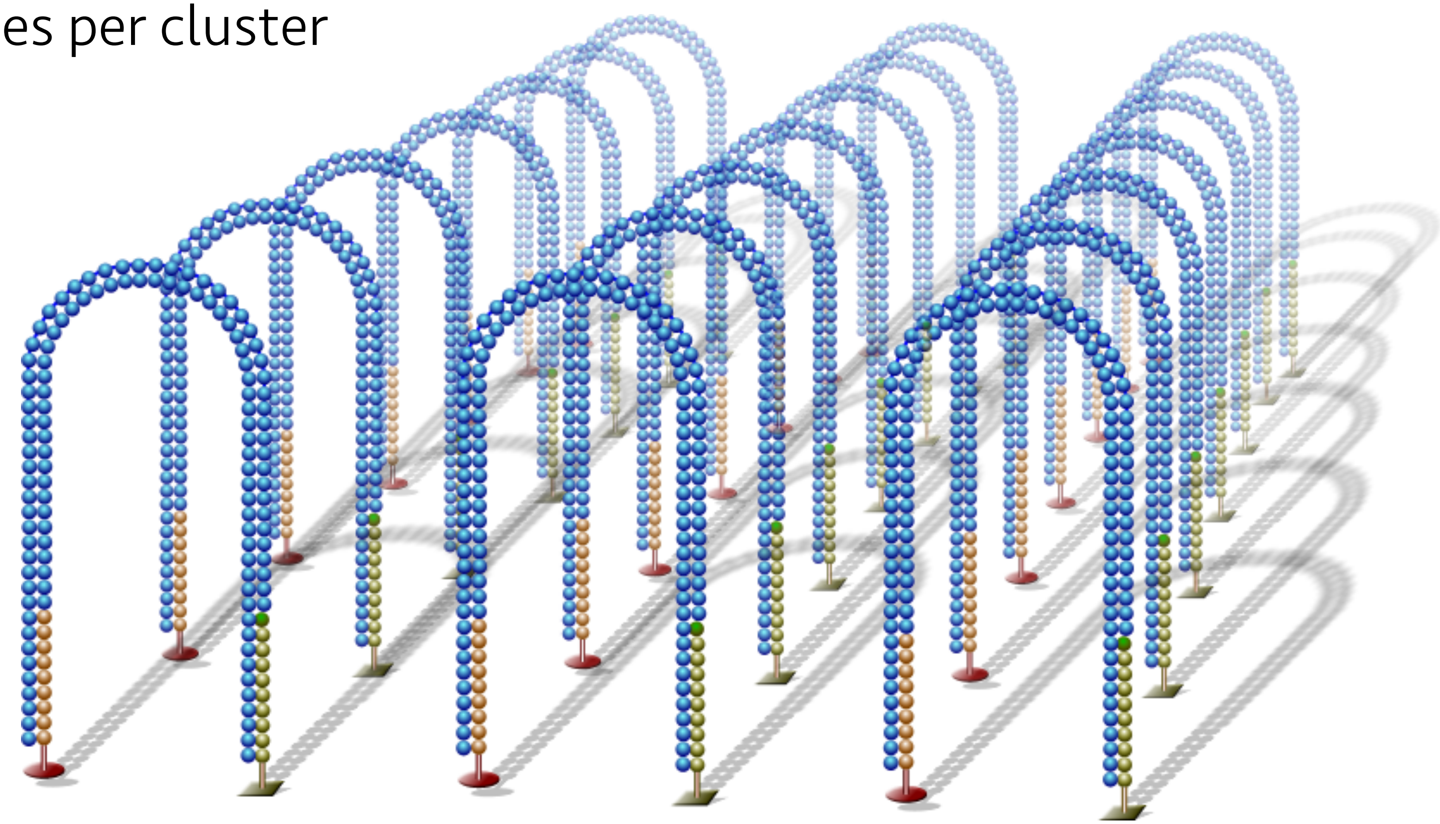
Original (orange oligo) and reverse strand are ready for next round of PCR .



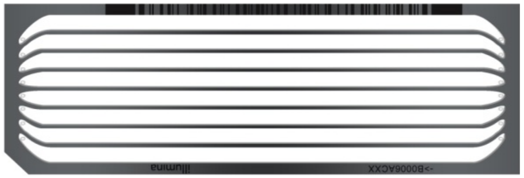


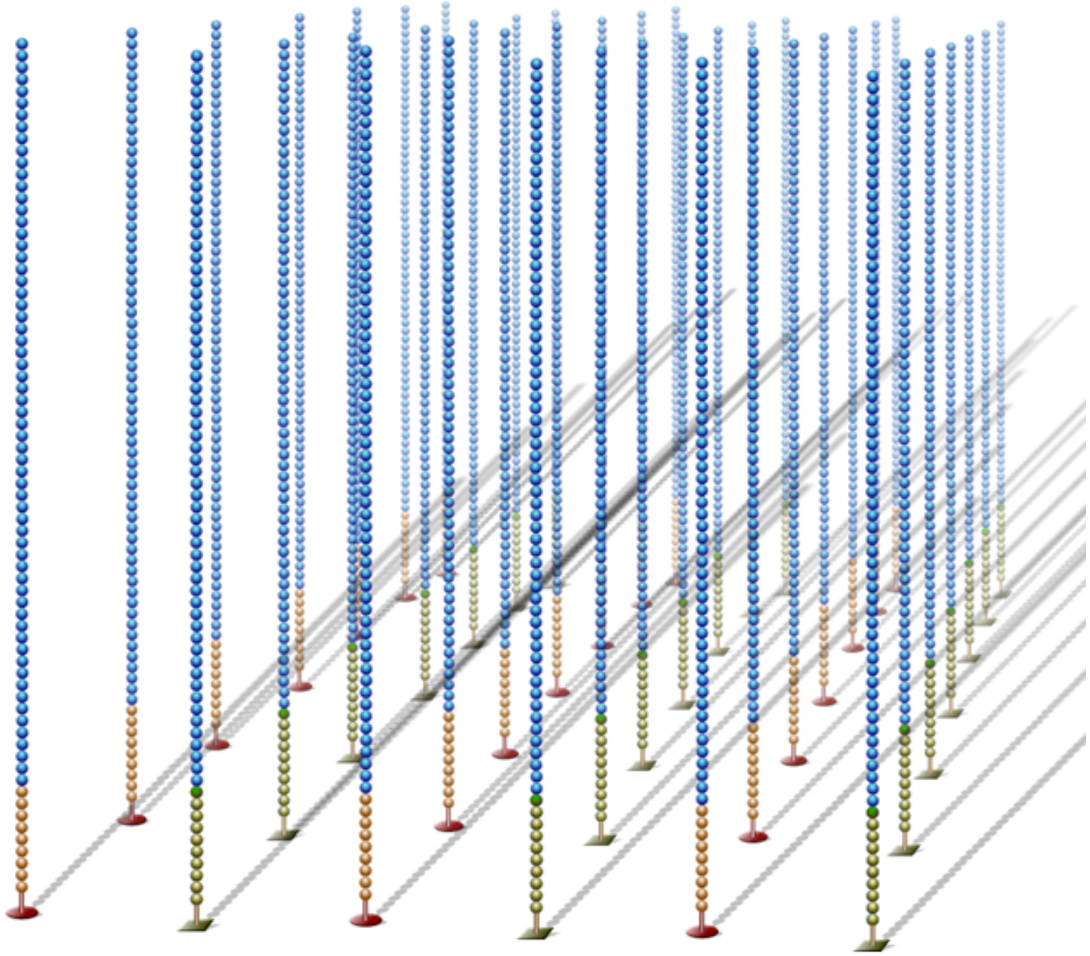


At the end of clustering
~1000 copies per cluster



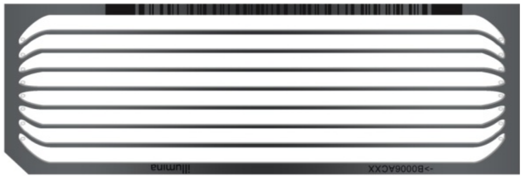
Molecules are linearized

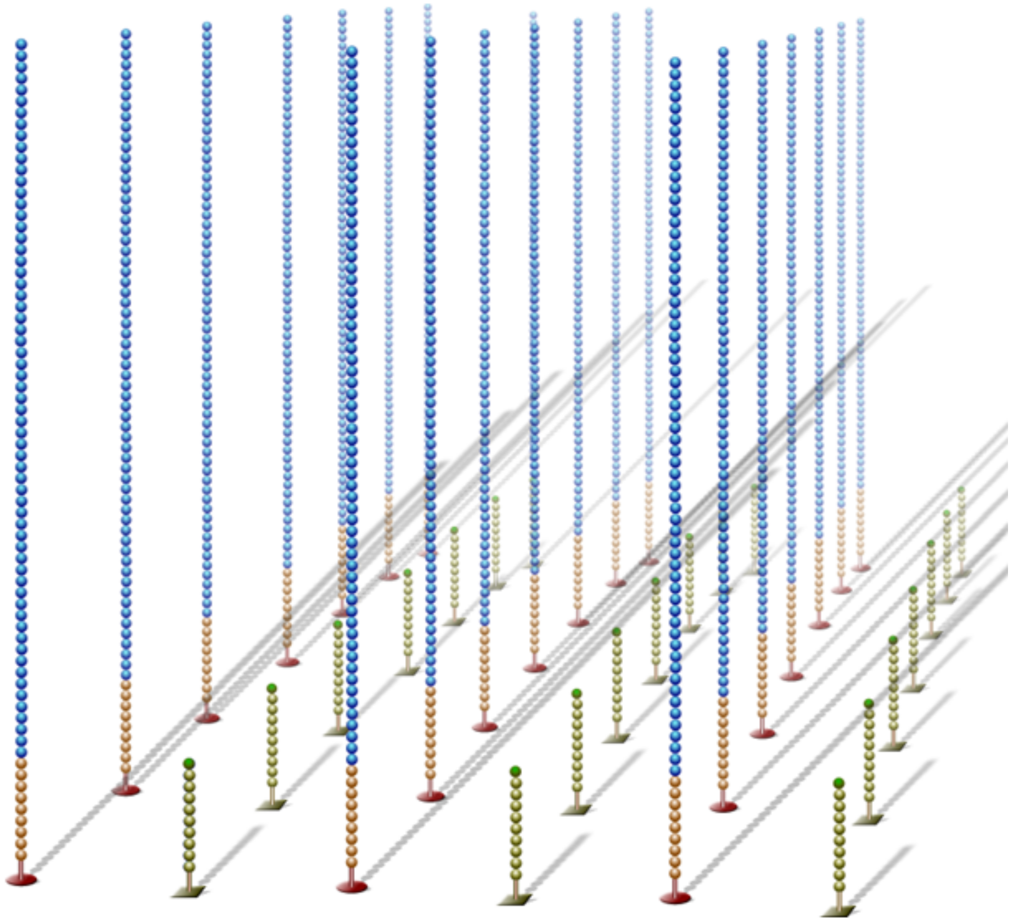




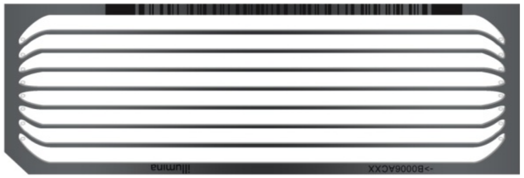
Reverse Strands are cleaved

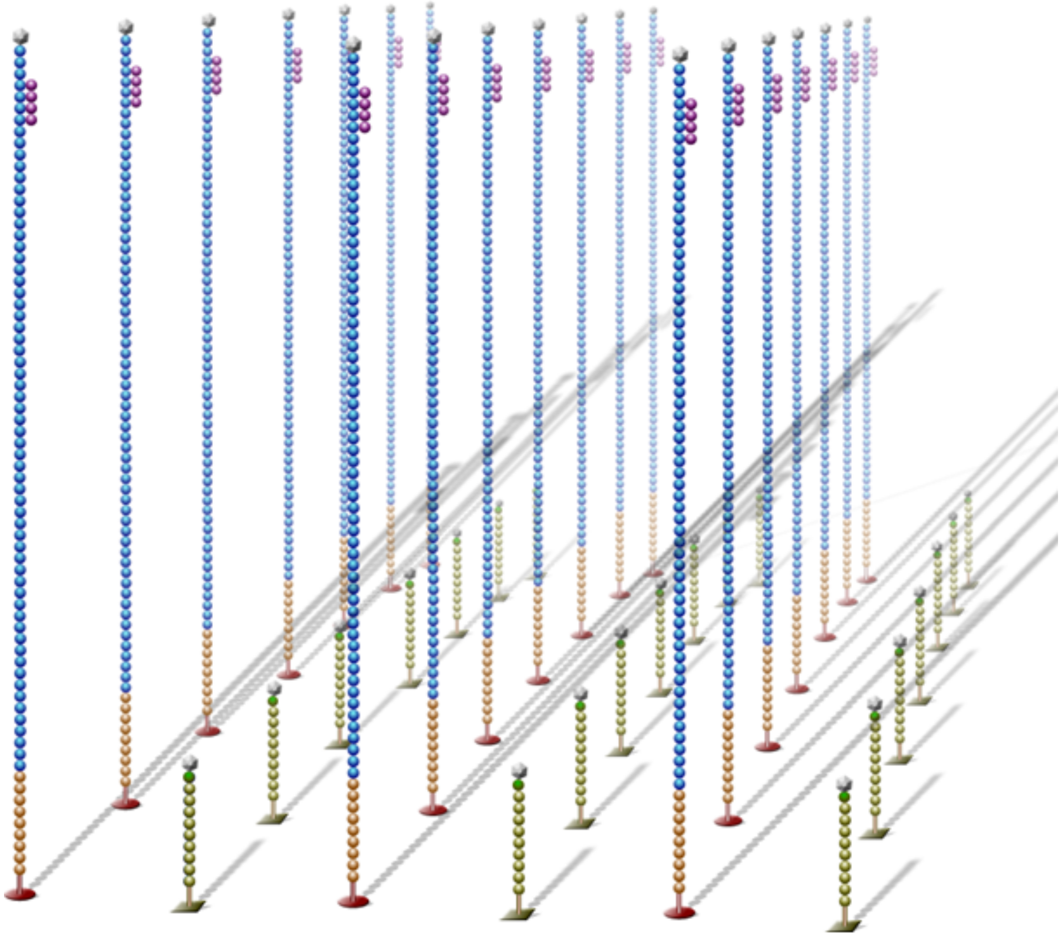
Original template strand remains (orange oligo)



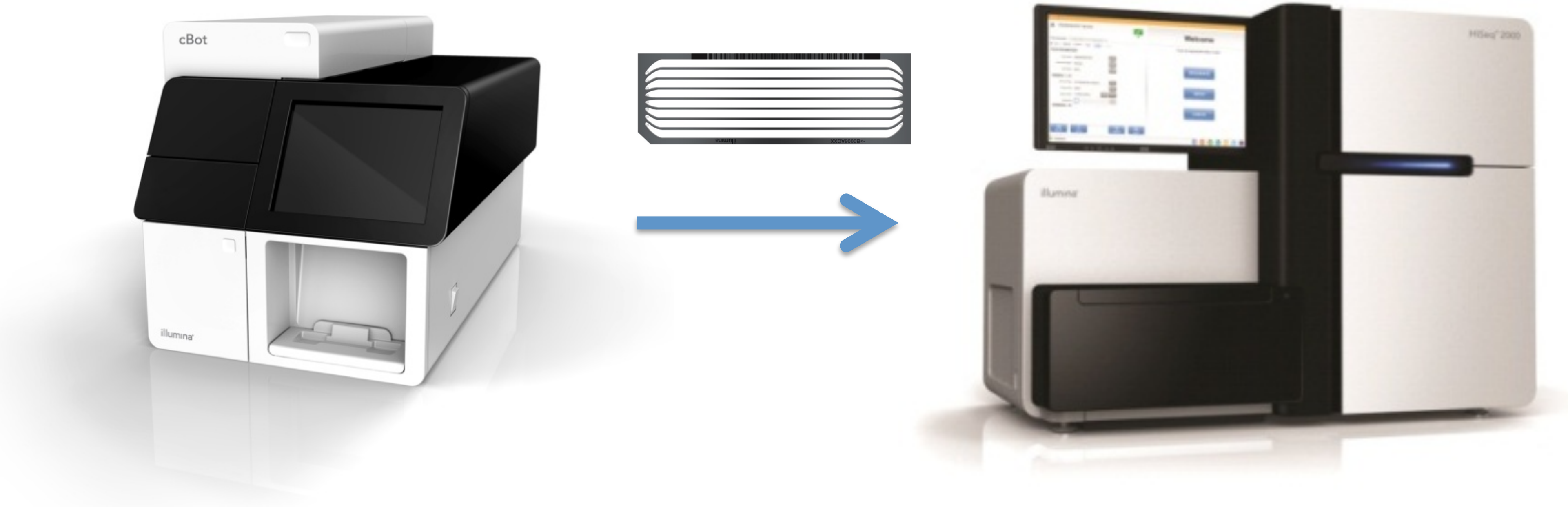


Ends are blocked and sequencing primers hybridized

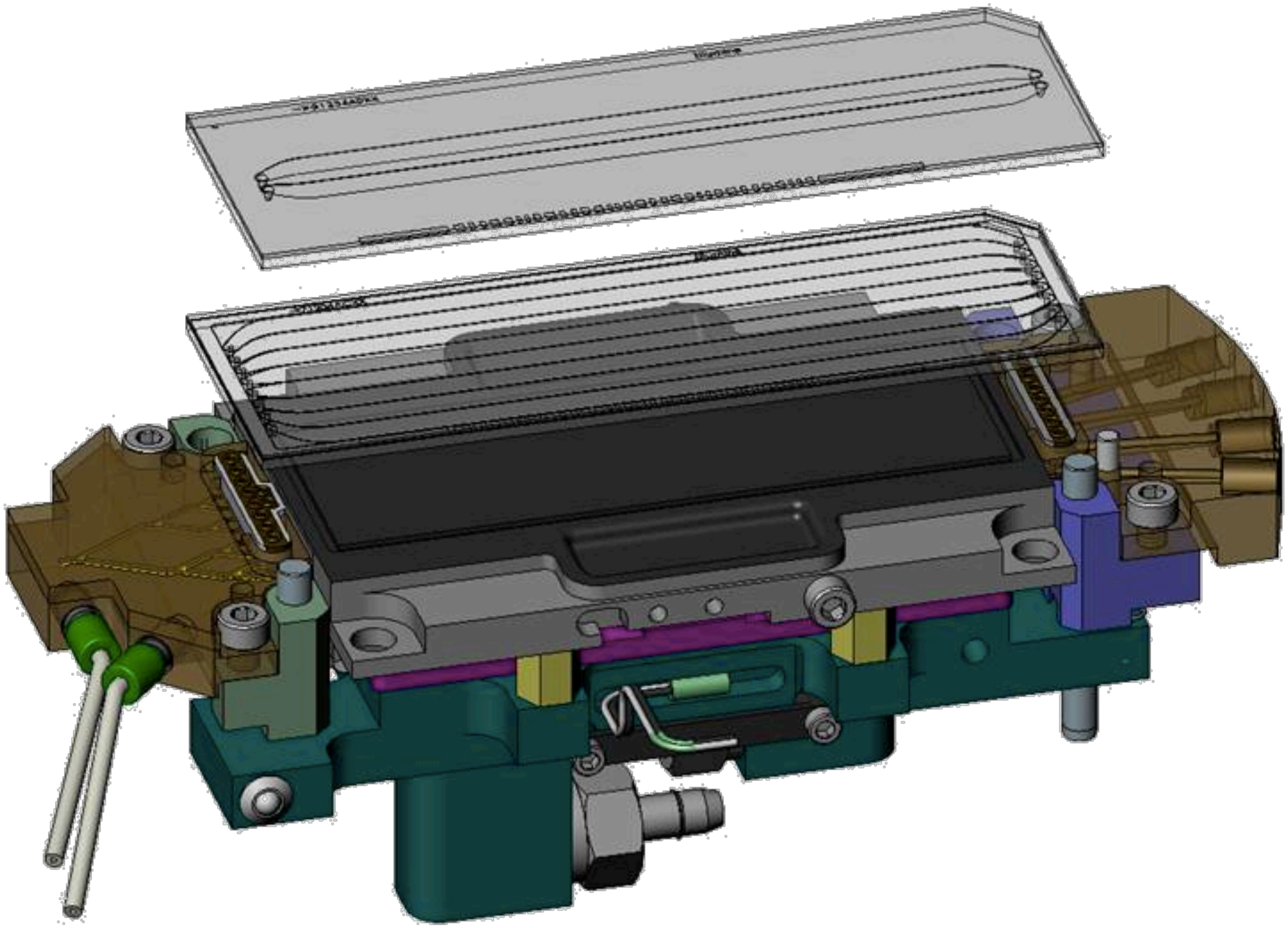




After clustering flow cell moves to the HiSeq

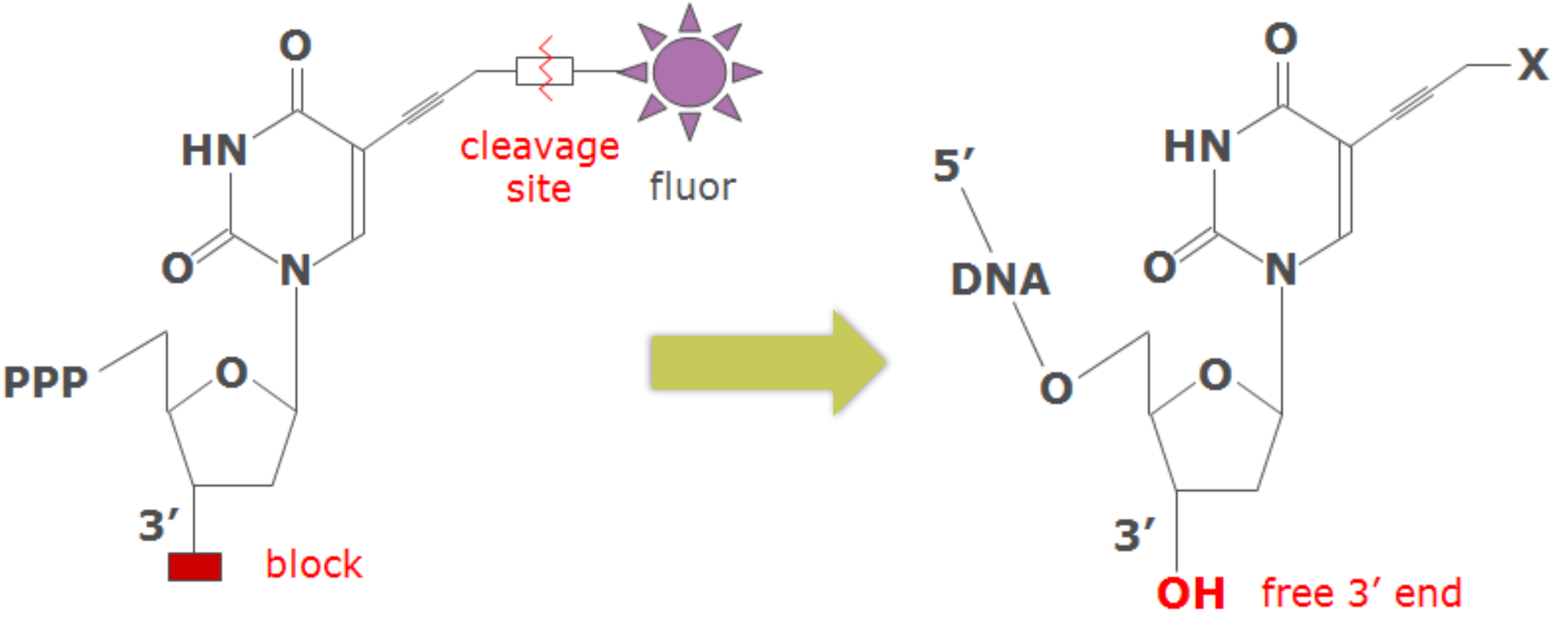


Sequencer is a microscope with fluidic channels

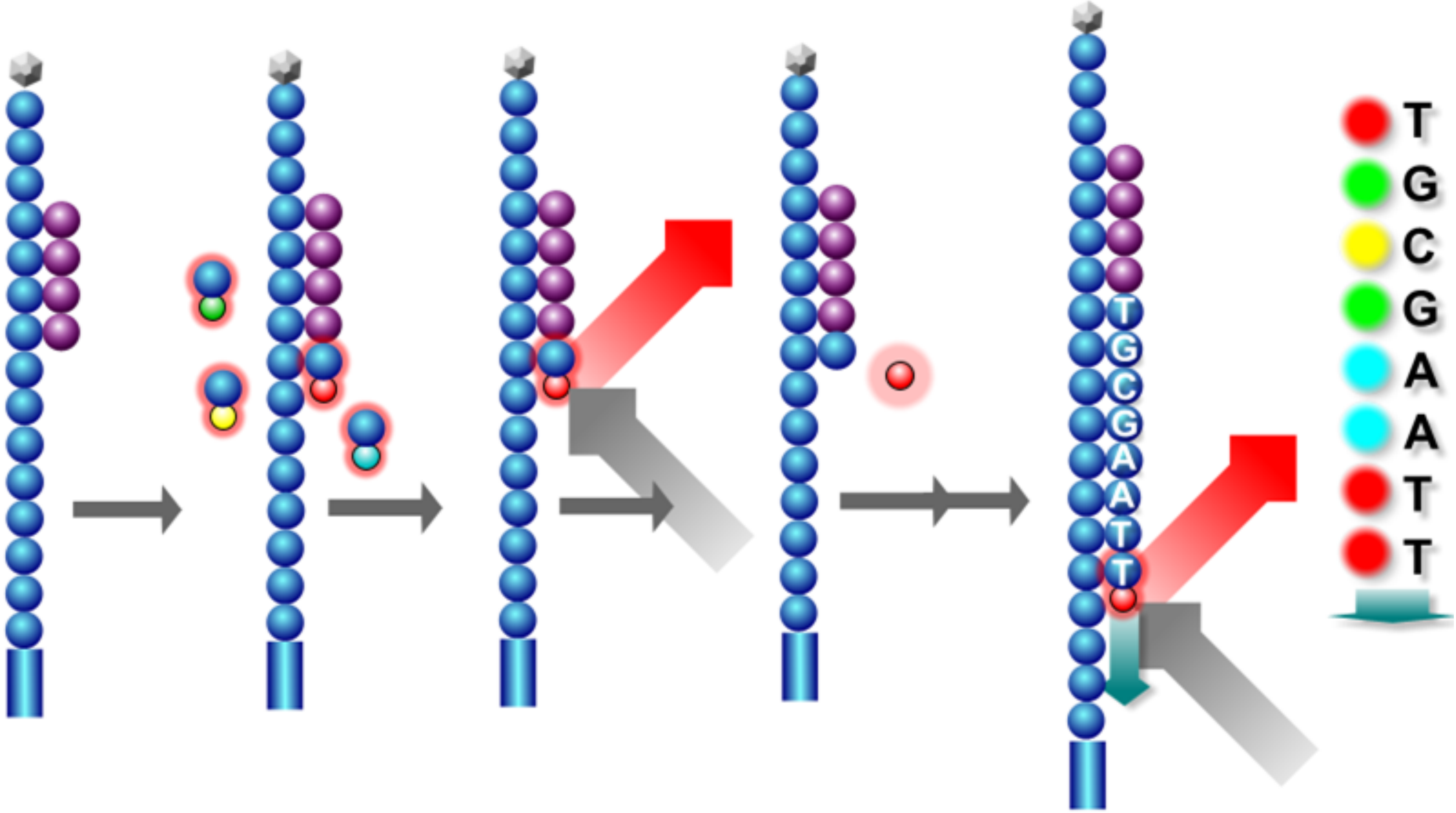


Reversible Terminator Chemistry

🔴 A
🟢 C
🟡 G
🔵 T

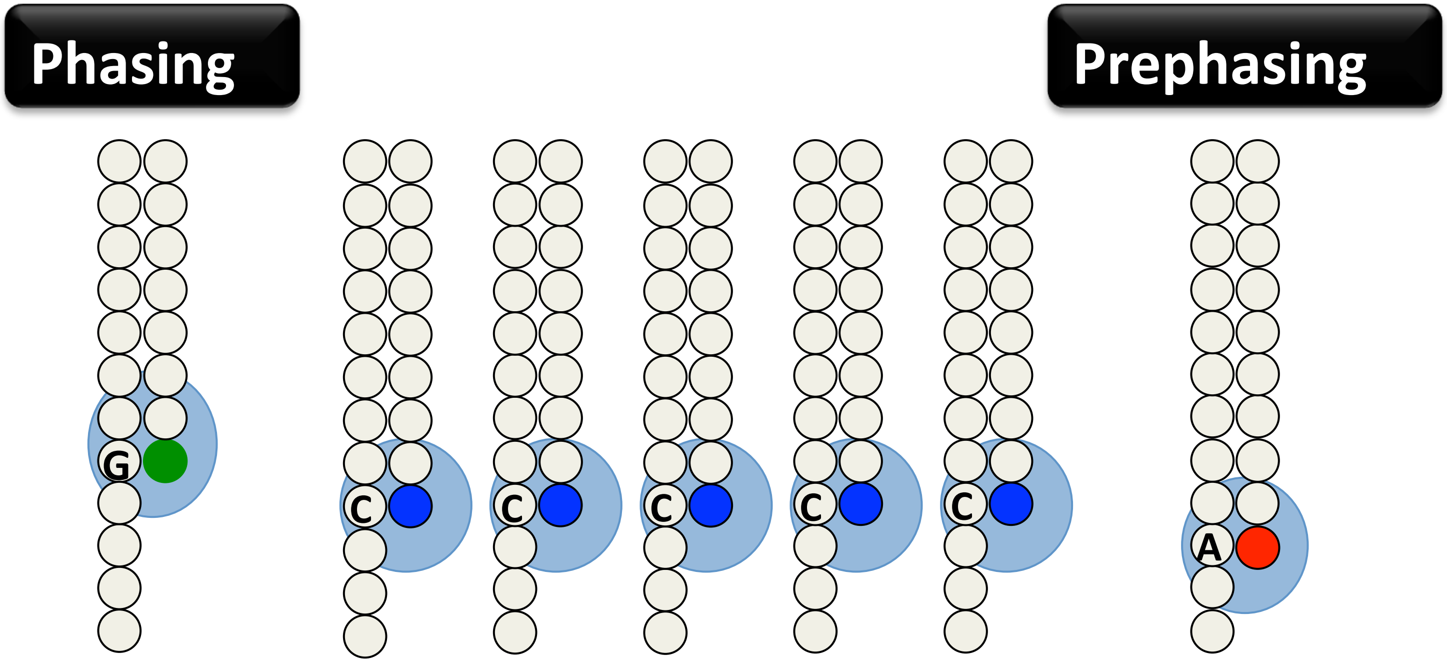


Illumina SBS technology



Limitations of SBS

Each clonal cluster contains ~1,000 copies
Imperfect chemistry → some strands will lag and others will jump ahead

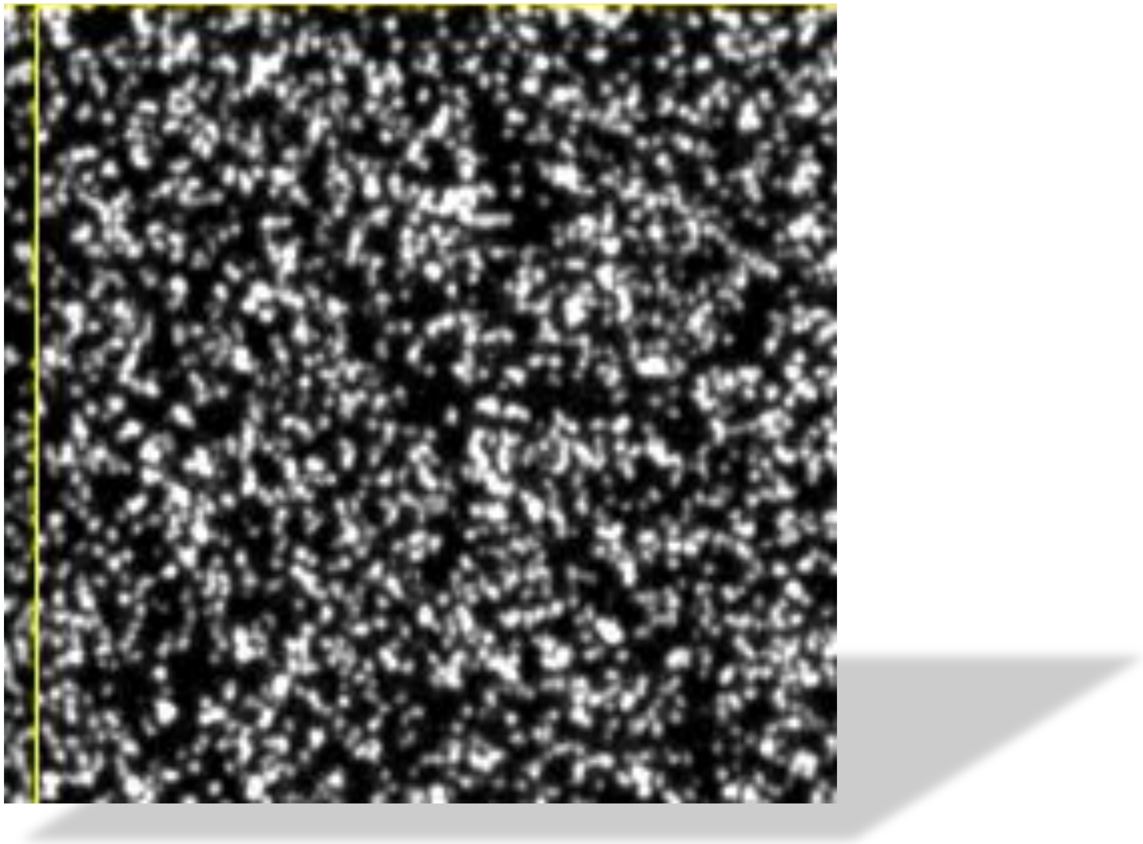


This limits the length of runs.
HiSeq 2x150
MiSeq 2x300

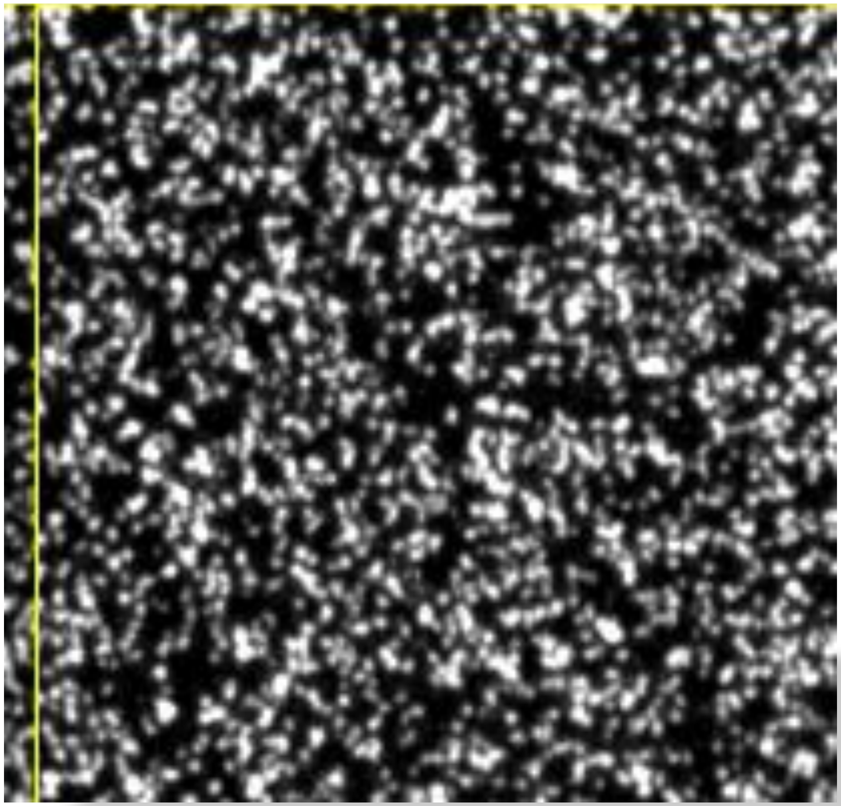
Going from images to sequence

- Find clusters
- Calculate intensities
- Make basecalls

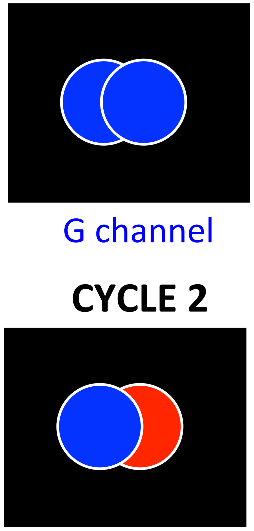
Sequence diversity is critical for template generation



Sequence diversity is critical for template generation



CYCLE 1



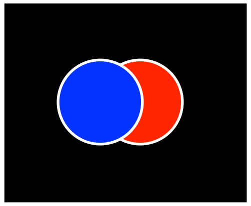
G channel

CYCLE 2

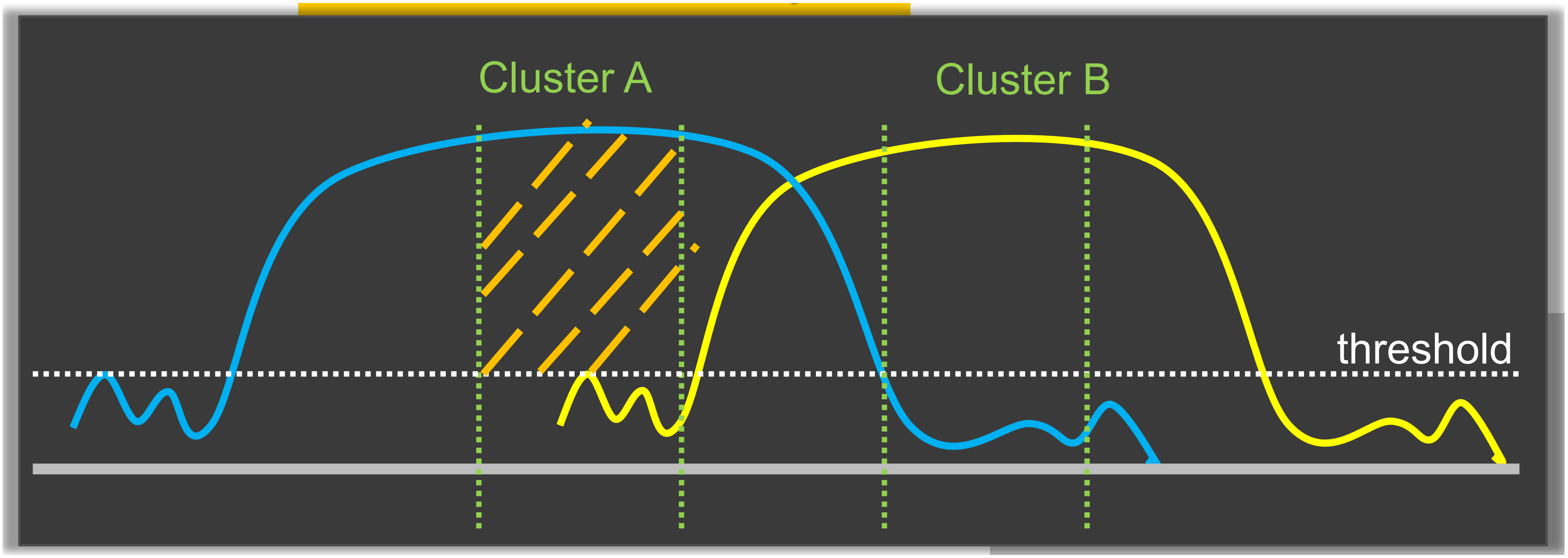
C channel
G channel

Sequence diversity is critical for template generation

CYCLE 2

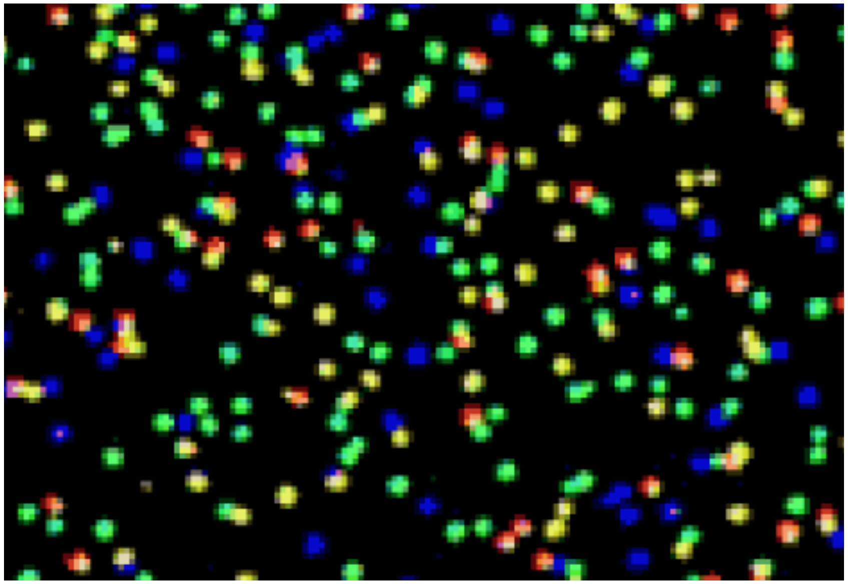


C channel
G channel



Raw images

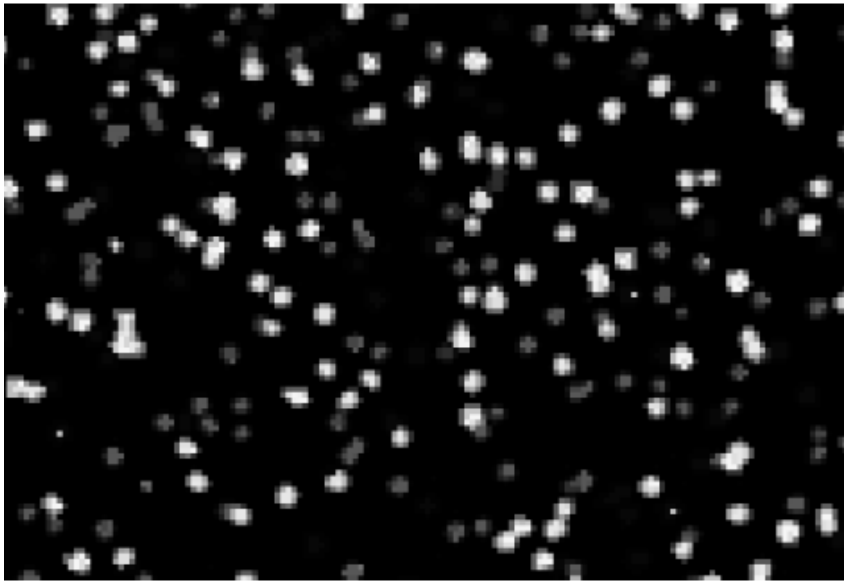
System takes 4 images each cycle, 1 per nucleotide



Raw images

System takes 4 images each cycle, 1 per nucleotide

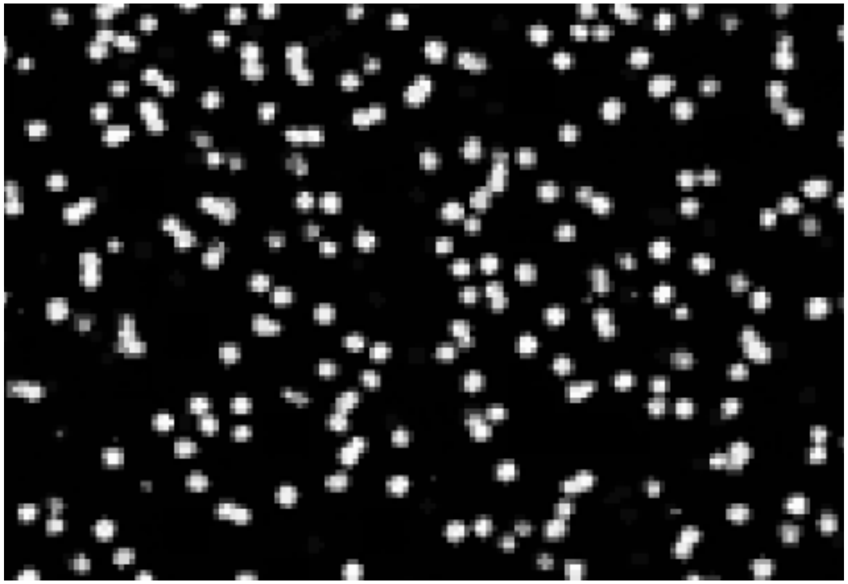
Red



Raw images

System takes 4 images each cycle, 1 per nucleotide

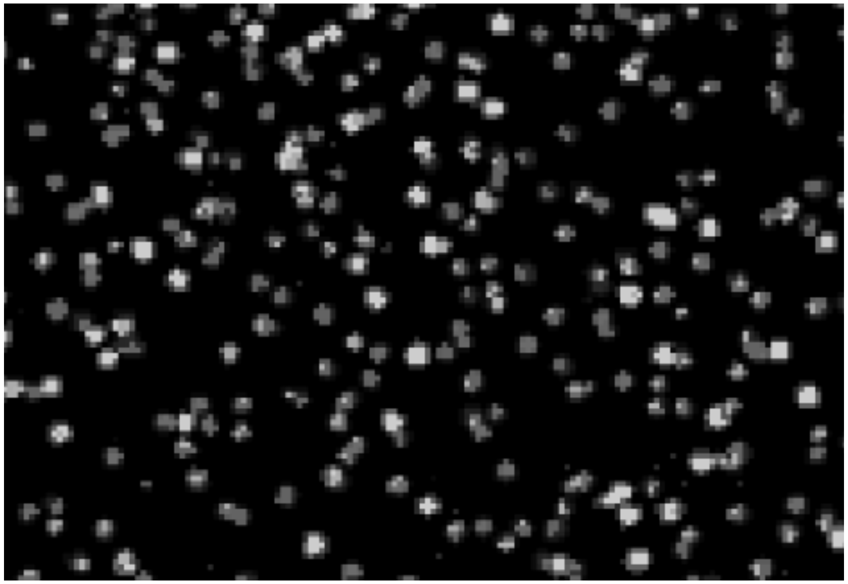
Green



Raw images

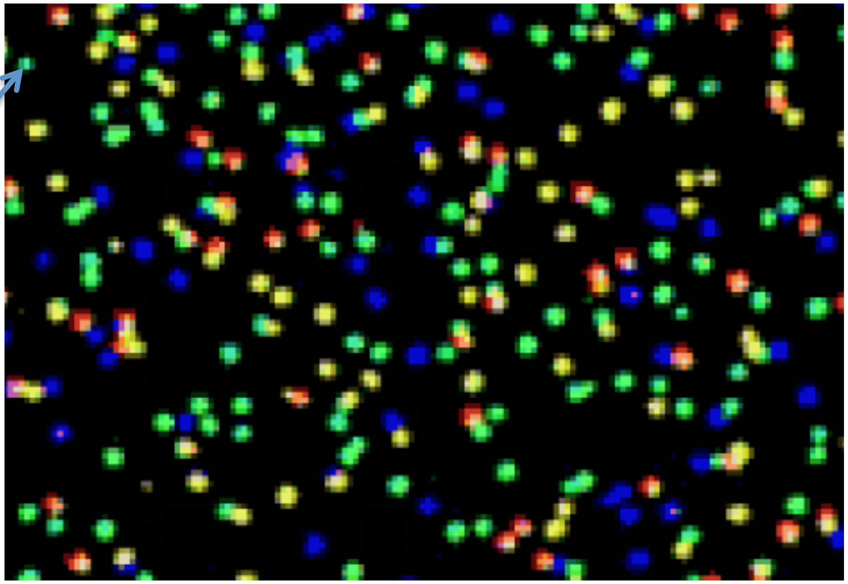
System takes 4 images each cycle, 1 per nucleotide

Blue



Raw images

System takes 4 images each cycle, 1 per nucleotide

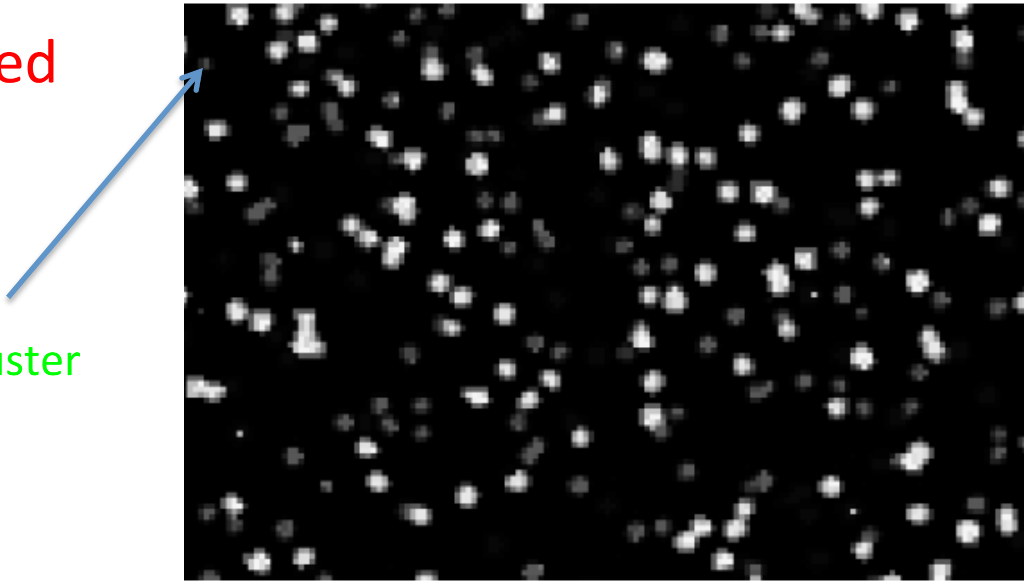


Merge

Green cluster

Fluorescence crosstalk

Fluorescent label spectra overlap

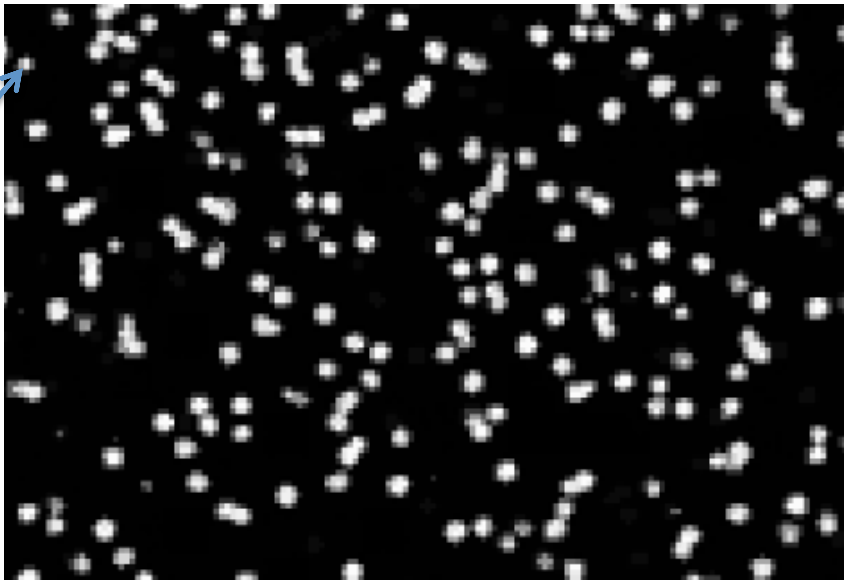


Red

Green cluster

Fluorescence crosstalk

Fluorescent label spectra overlap

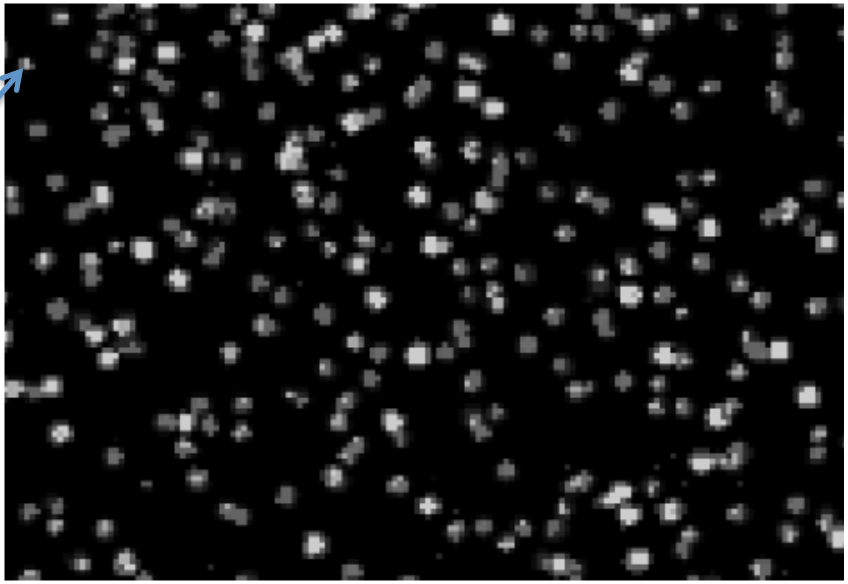


Green

Green cluster

Fluorescence crosstalk

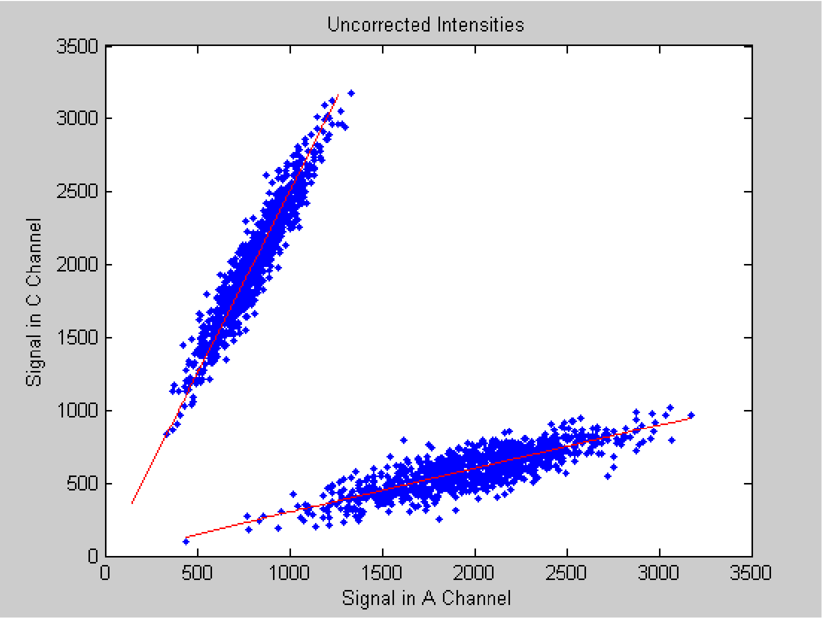
Fluorescent label spectra overlap

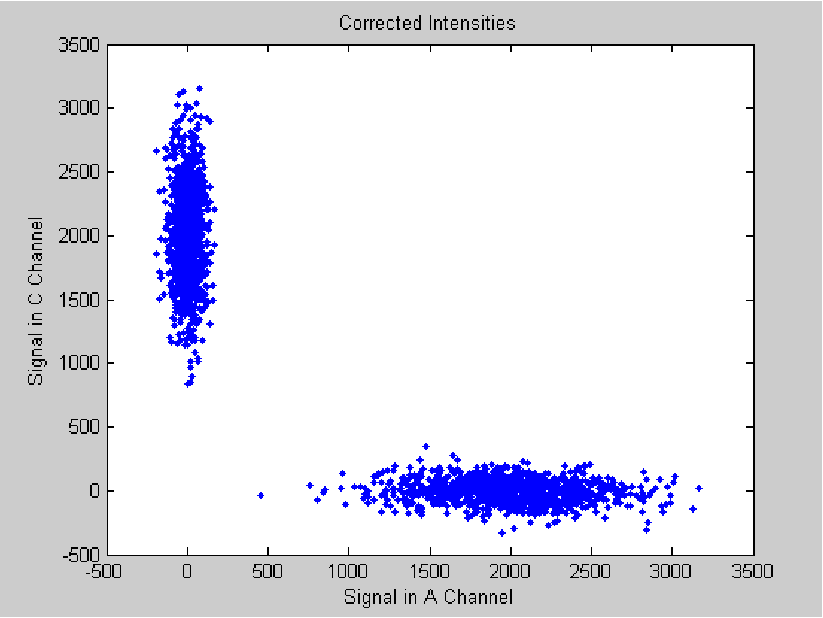


Blue

Green cluster

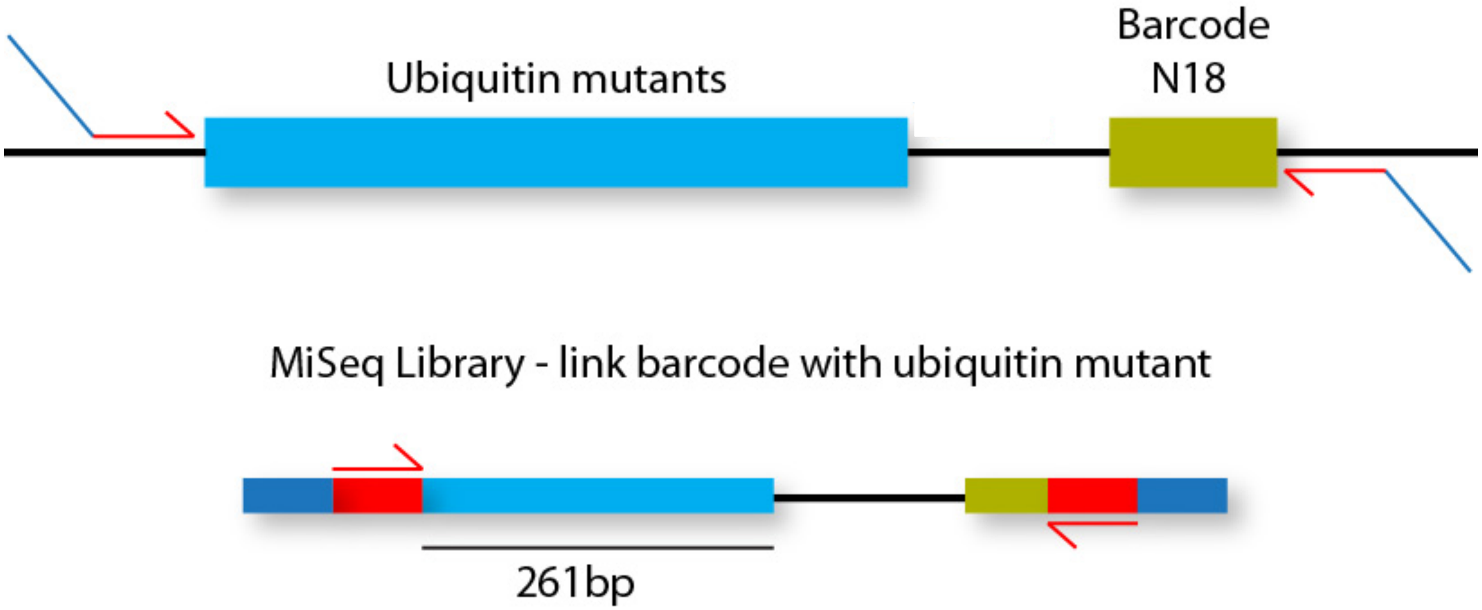
Signal cross talk correction



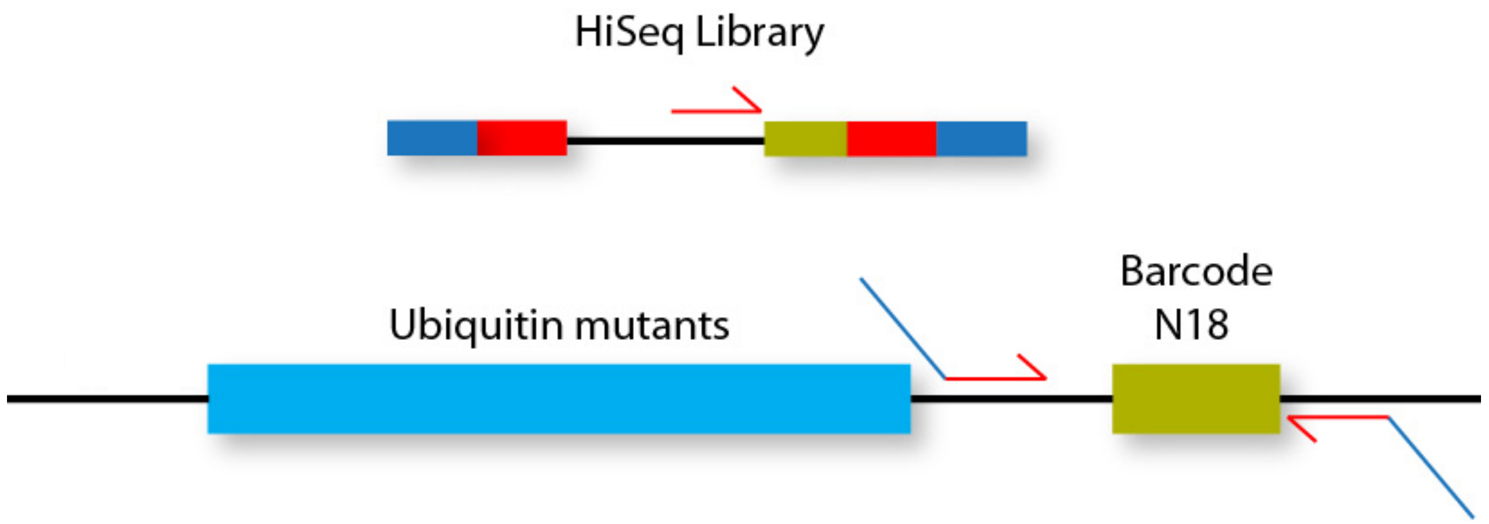


Ubiquitin library

Ubiquitin library



Ubiquitin library



- HiSeq is much cheaper than the MiSeq
    - 10-20x more reads
    - 20% more expensive
    - Save MiSeq for longer reads