



Leiden University
Medical Center

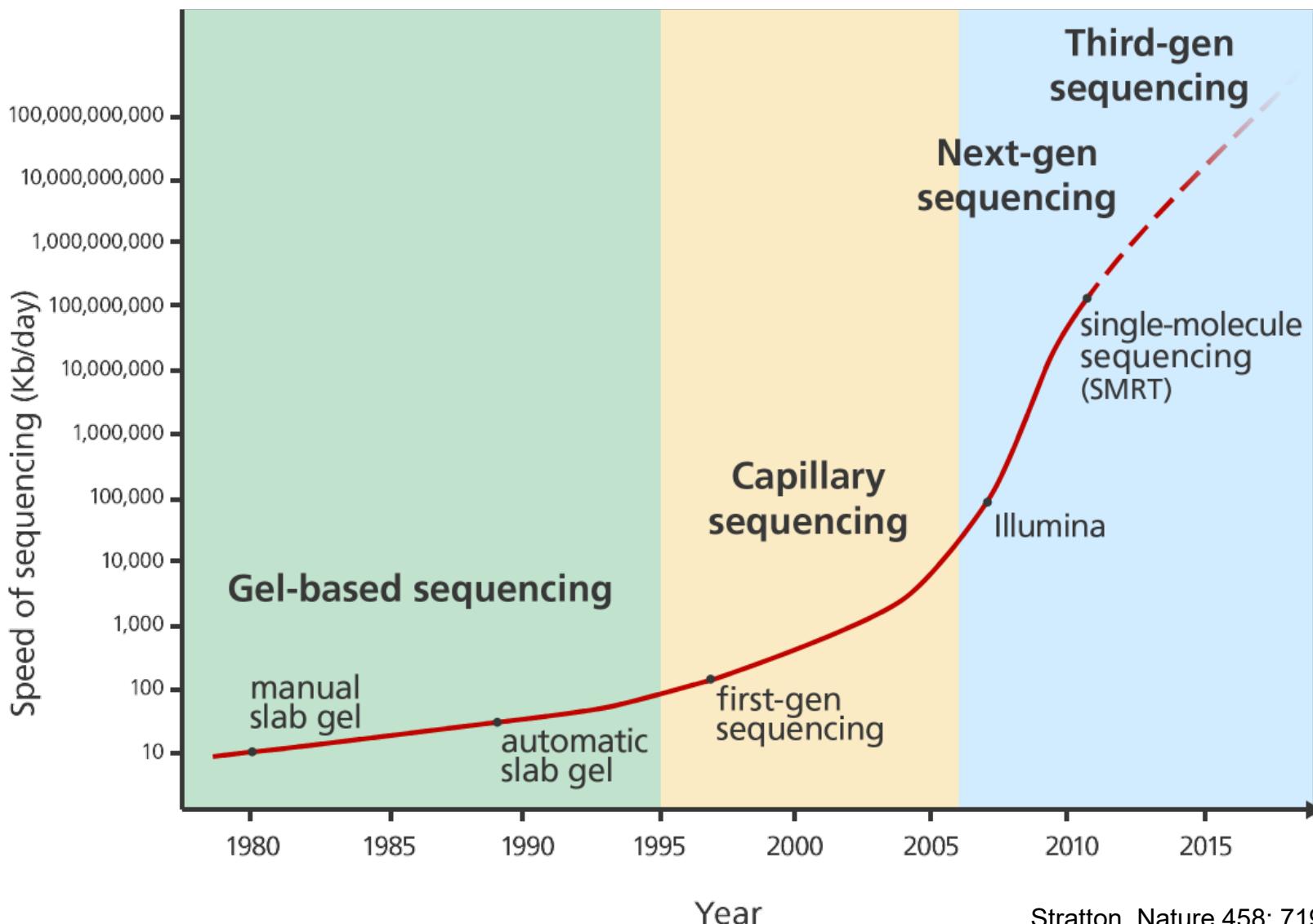
Next Generation Sequencing Technologies

Susan Kloet, PhD

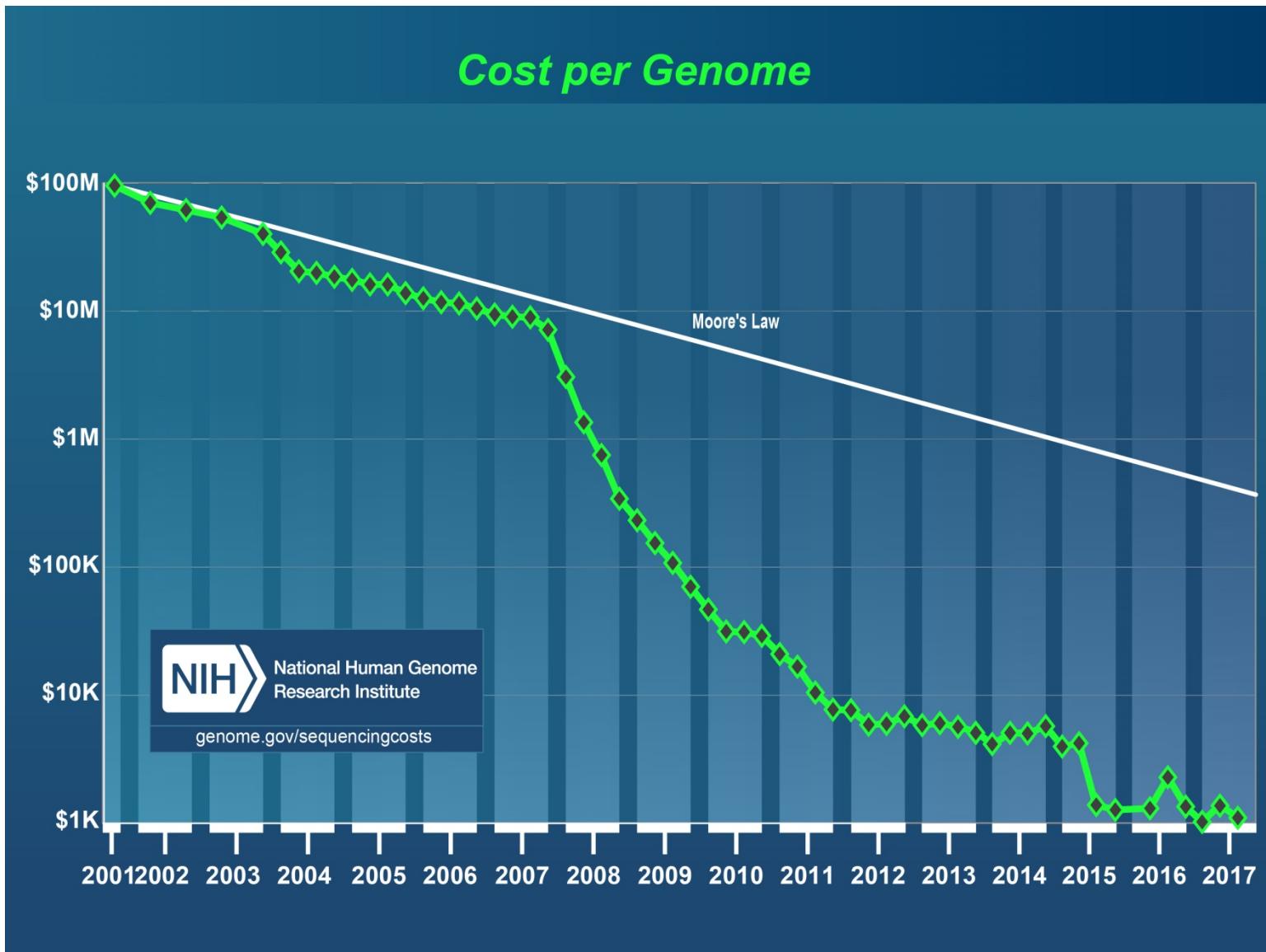
Leiden University Medical Center
Department of Human Genetics
Leiden Genome Technology Center



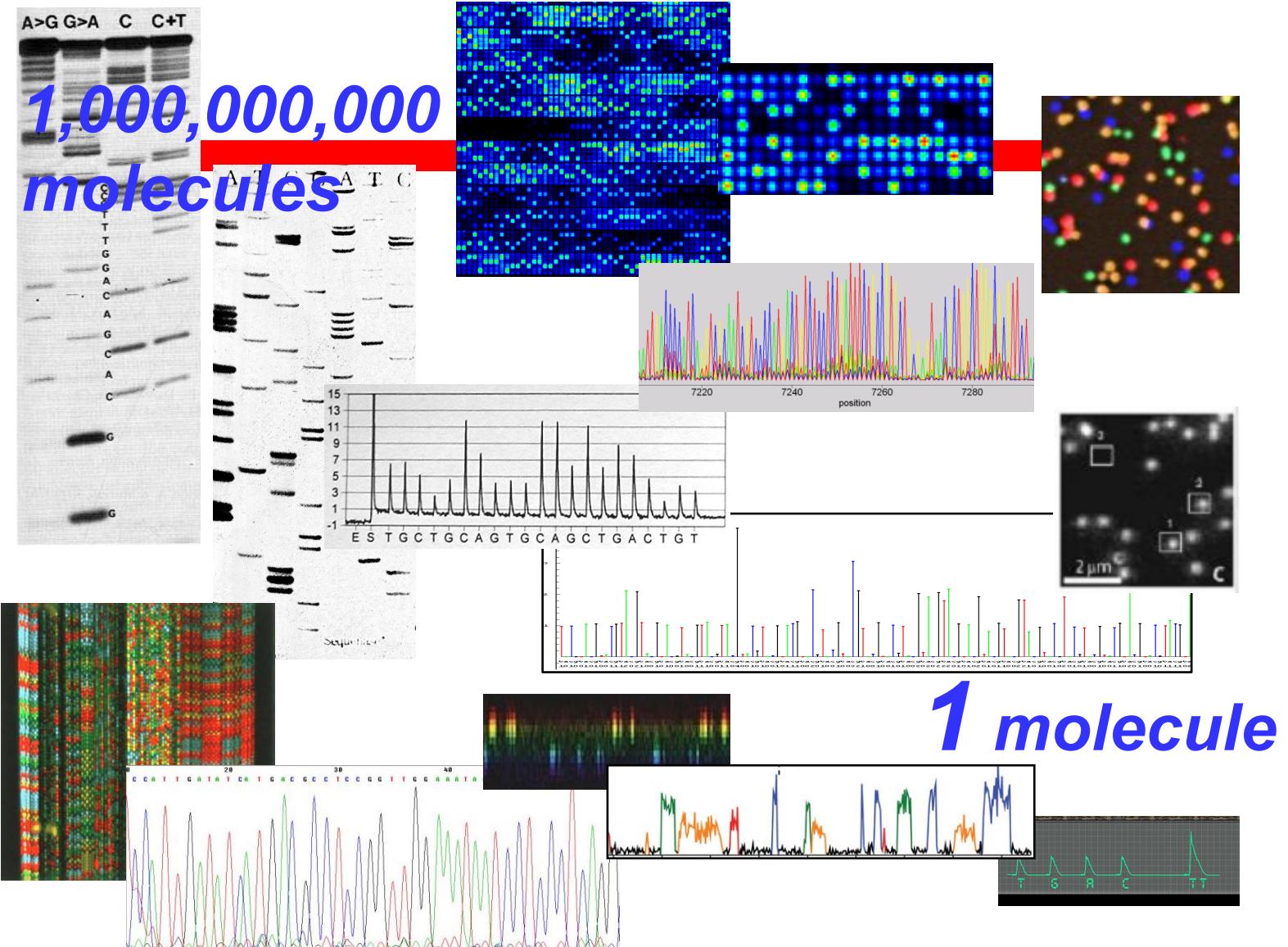
Sequencing Technology: Output Revolution



Sequencing Technology: Output Revolution



Sequencing Technology: Input Revolution



What is Next Generation Sequencing?

Relatively novel (~12 years old) sequencing technology development

Analyzing complex mixtures of nucleic acid strands

Short sequencing reads

Millions, now billions of reads in parallel in a single run

Giga - Tera base scale

Next generation sequencing – short read technologies

Illumina

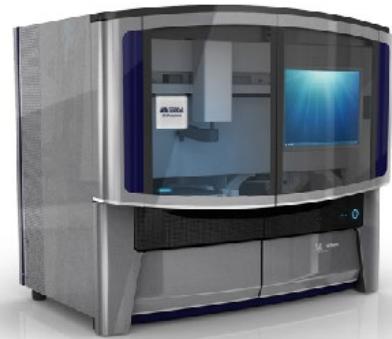
>85% of bases sequenced are with Illumina technologies

Sequencing systems for every lab

	*	*	*	*	
					
iSeq	MiniSeq	MiSeq	NextSeq	HiSeq	HiSeq X
					
€600 4M	€1200 25M	€3000 400M			€30k 10B

Next generation sequencing – short read technologies

Other



LifeTechnologies
SOLiD



LifeTechnologies
PGM*



LifeTechnologies
Ion Proton*



Roche 454*



Helicos
Heliscope* → <http://seqll.com/>

Illumina HiSeq technology

2007: Solexa
Bought by Illumina
Upgrades to improve output and quality
→ GAI

2011 → HiSeq 2000
2013 → HiSeq 2500
2015 → HiSeq 4000 (X10)
2017 → NovaSeq 6000



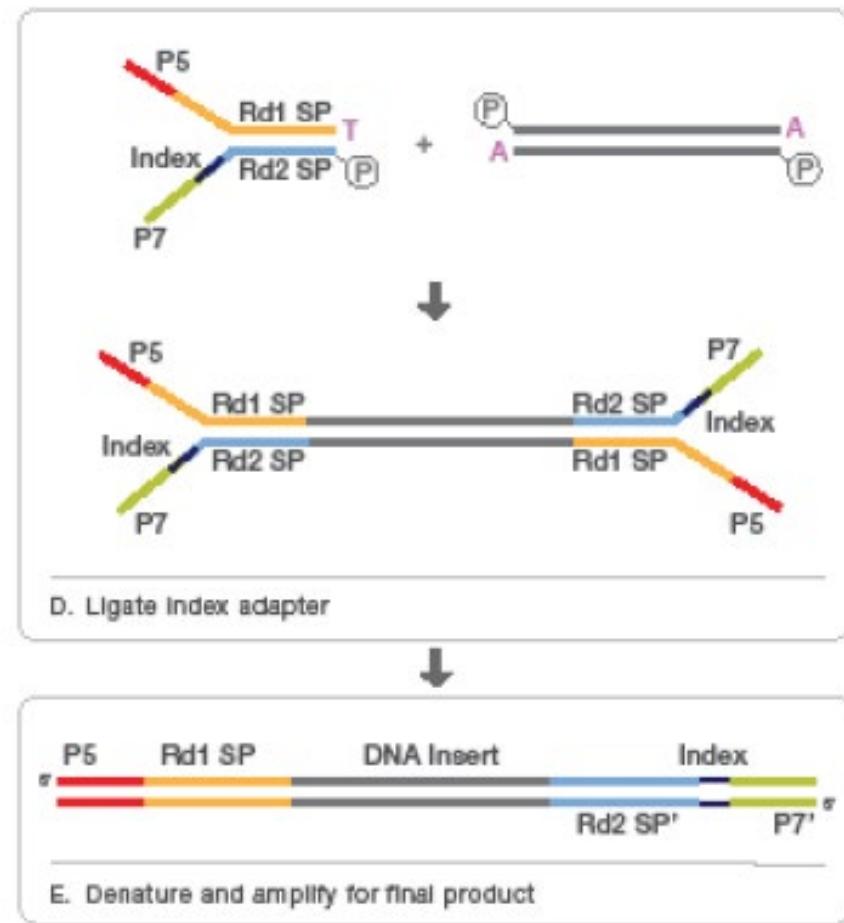
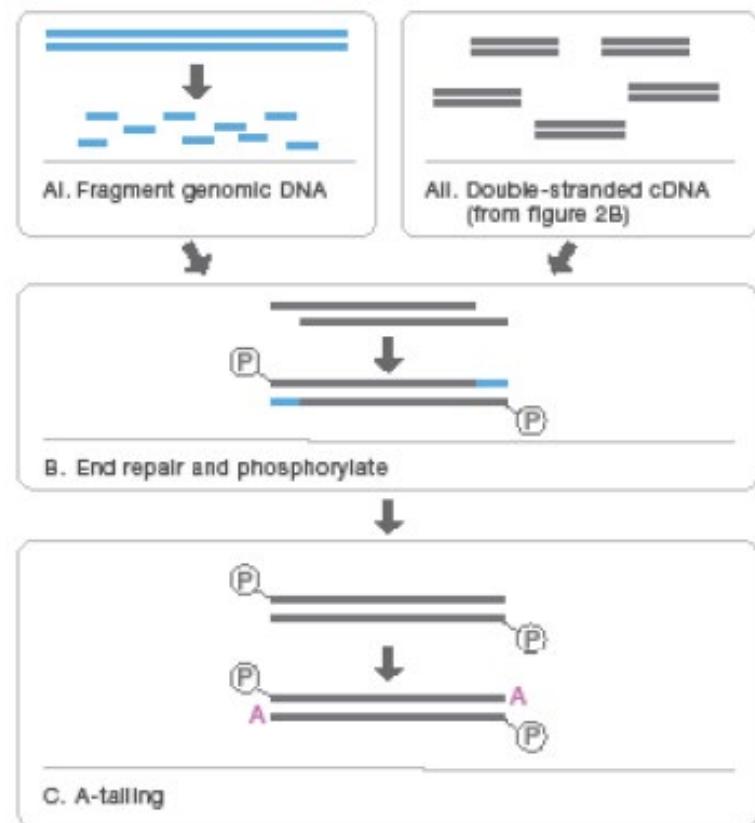
Latest reagents/flowcell versions

10 billion reads per flowcell (4 lanes)
Paired-end 2x150 bp reads
→ 3000 Gbase per flowcell (750 Gbase per lane)
→ 6000 Gbase per run = 6 Tbase

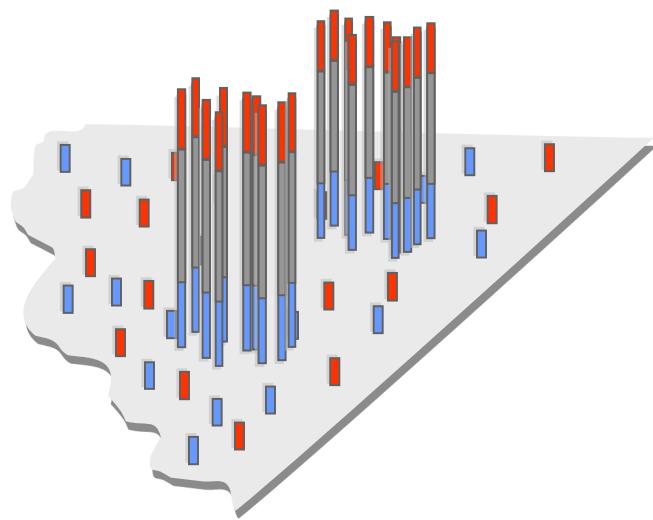
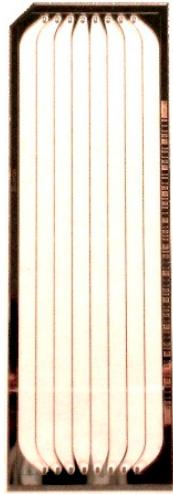
~48 whole human genomes @ 30x coverage

Basic Illumina library preparation

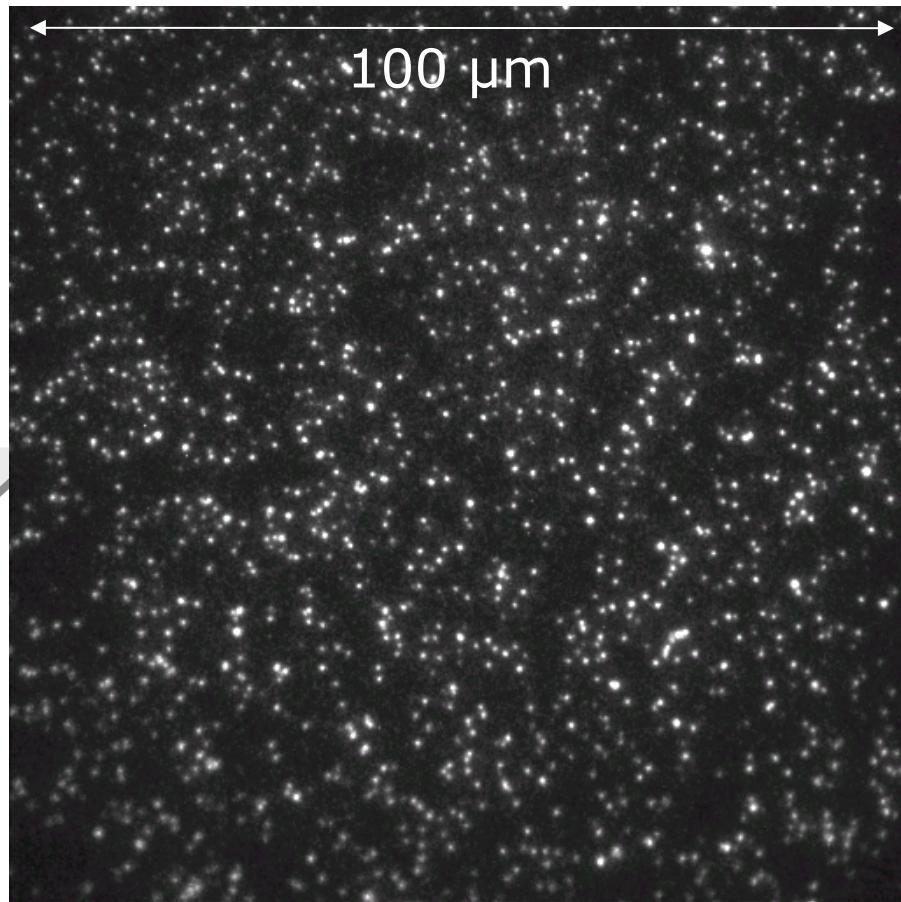
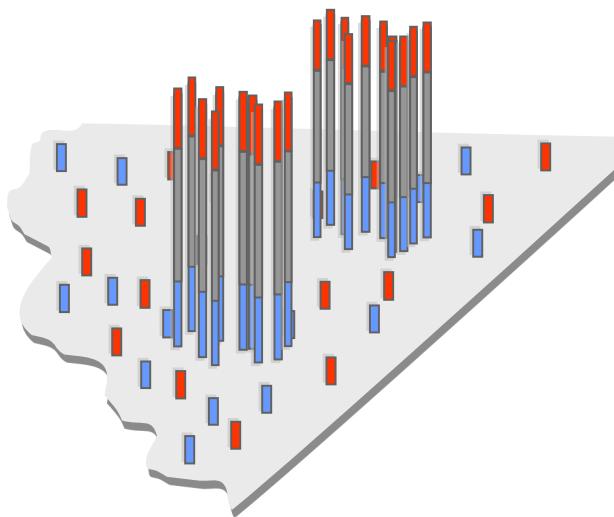
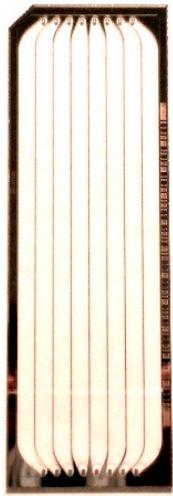
Figure 4: Adapter Ligation Results in Sequence-Ready Constructs without PCR



HiSeq: Single molecule → Cluster



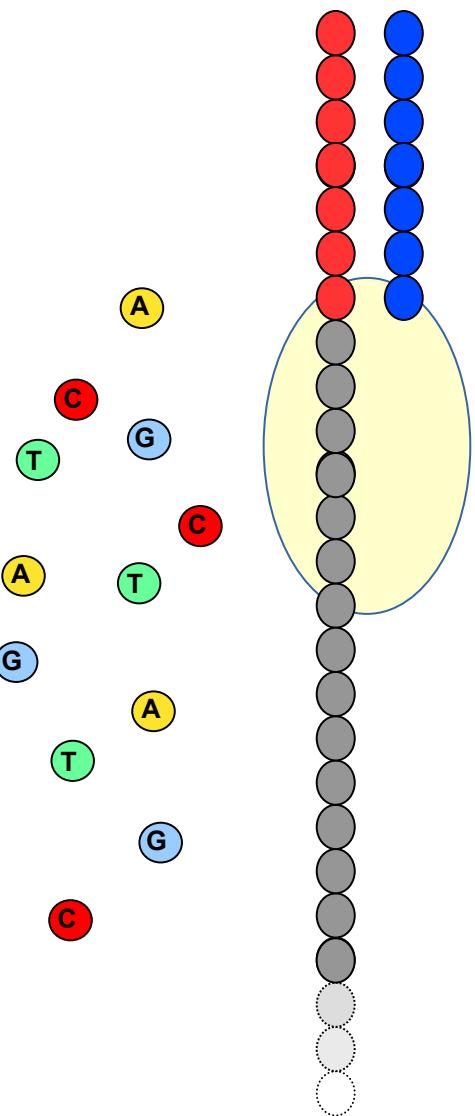
HiSeq: Single molecule → Cluster



HiSeq: Cluster → Sequence

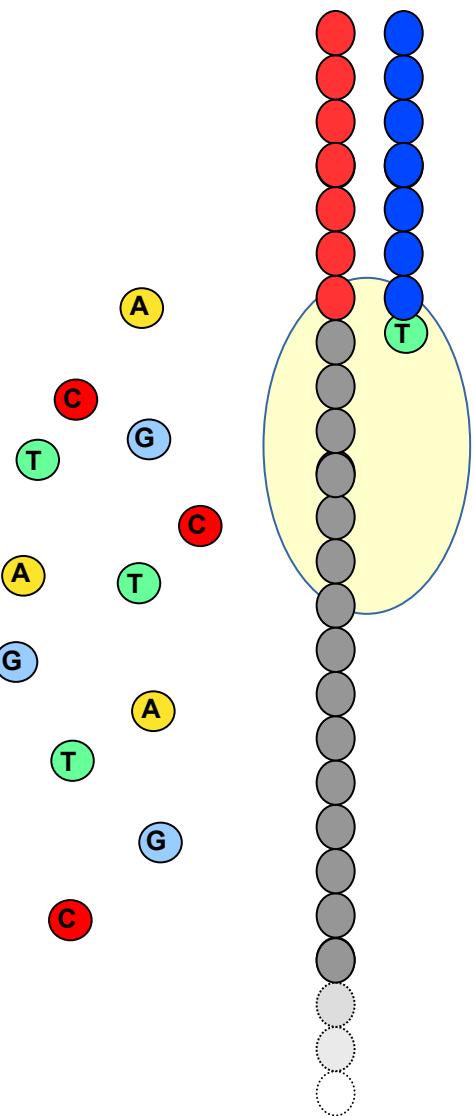


HiSeq: Cluster → Sequence



Cycle 1: Add sequencing reagents

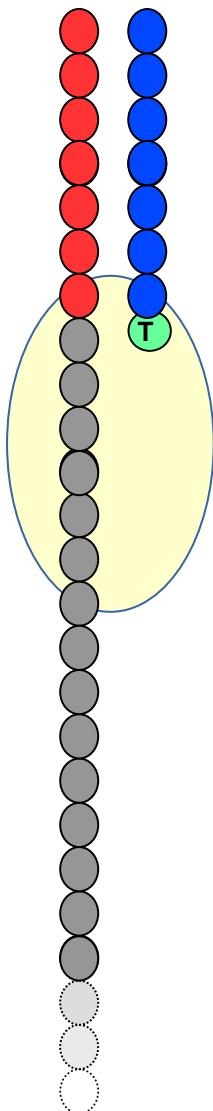
HiSeq: Cluster → Sequence



Cycle 1: Add sequencing reagents

First base incorporated

HiSeq: Cluster → Sequence

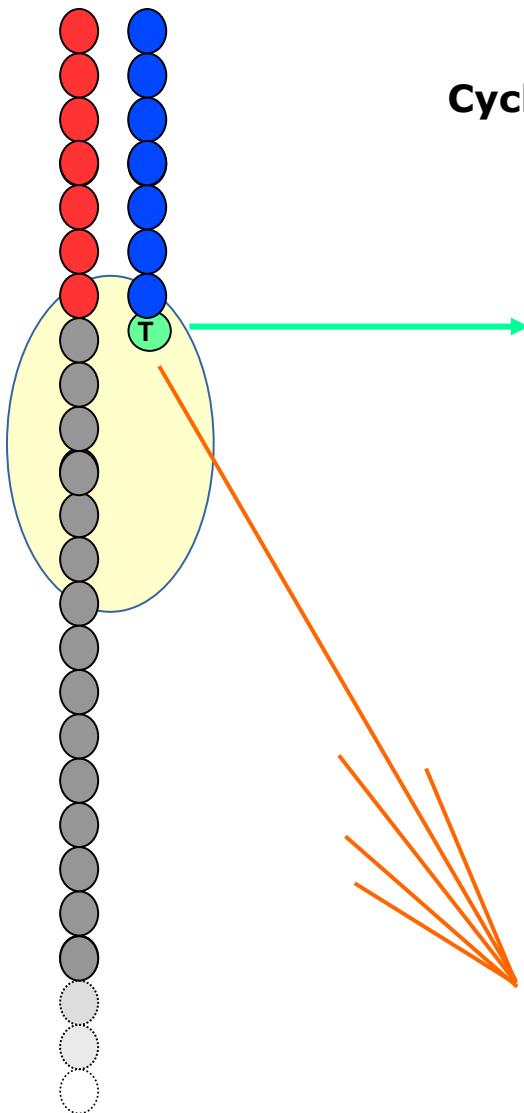


Cycle 1: Add sequencing reagents

First base incorporated

Remove unincorporated bases

HiSeq: Cluster → Sequence



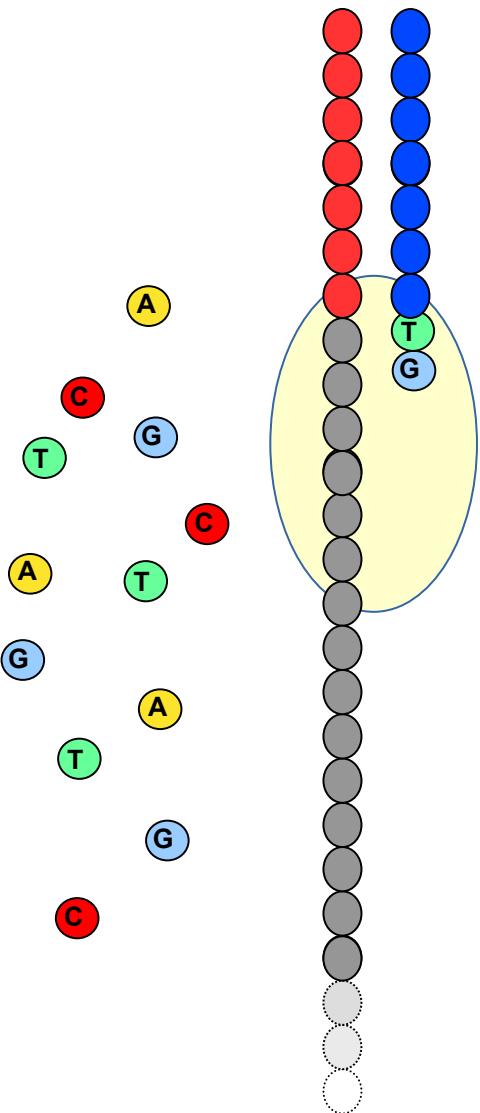
Cycle 1: Add sequencing reagents

First base incorporated

Remove unincorporated bases

Detect signal

HiSeq: Cluster → Sequence



Cycle 1: Add sequencing reagents

First base incorporated

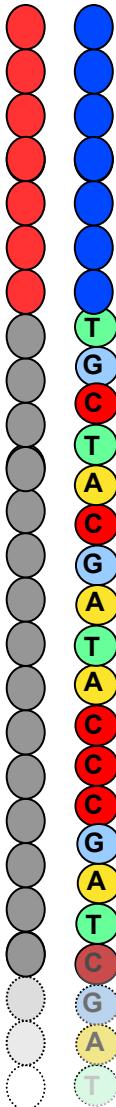
Remove unincorporated bases

Detect signal

Cycle 2-n: Add sequencing reagents and repeat

One cycle = ~15 minutes (HiSeq4000)

HiSeq: Cluster → Sequence



Cycle 1: Add sequencing reagents

First base incorporated

Remove unincorporated bases

Detect signal

Cycle 2-n: Add sequencing reagents and repeat

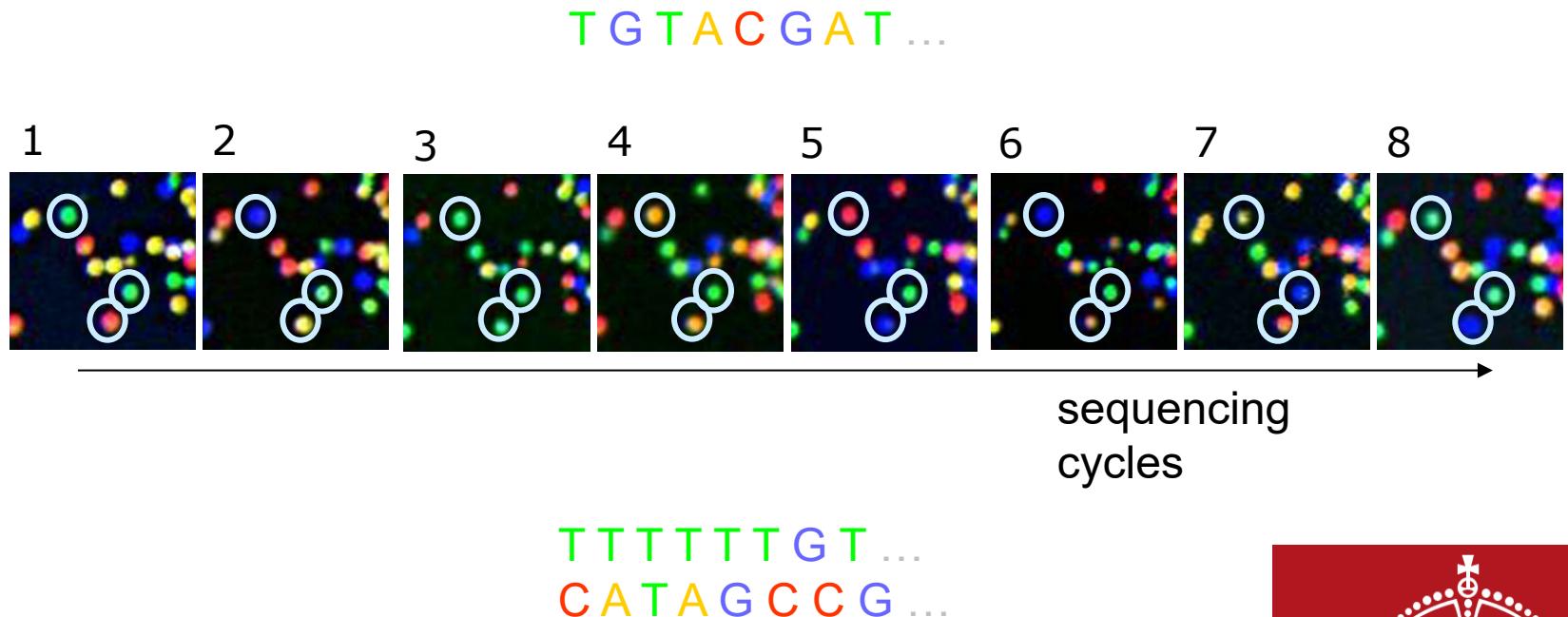
One cycle = ~15 minutes (HiSeq4000)

- All four labelled nucleotides in one reaction
- High accuracy
- Base-by-base sequencing

HiSeq: Cluster → Sequence



HiSeq: Cluster → Sequence



KEEP CALM
and
let software
take care of this

Biases due to PCR

Not all DNA can be amplified by PCR
some sequences resist amplification

- priming may fail
- low quality DNA

Polymerases are *NOT* perfect! → Errors introduced

All modifications lost, or copied erroneously

Relative abundance disturbed
Variable amplification efficiency
DNA quality, %GC

Contamination issues
Preferential amplification

The Illumina system also uses a polymerase during sequencing !

Illumina: Pros and Cons

Pros

- Market leader
- Broad range of applications (Google:For all you seq...)
- High output → High sample throughput
- Easy workflow for sample prep
- Automation protocols available
- Enables dual barcoding
- “One-tube” reactions for sample prep
- Sample input: single cell to 1000ng

Cons

- Short reads
- Drop in quality during run (out of phase molecules)
- Problems with structural rearrangements and context
- Sequencing is on a synthetic template

→ Avoid the PCR amplification
→ Single molecule sequencing

3rd Generation Sequencing: PacBio Single Molecule, Real-Time (SMRT) Sequencing

Long Read Lengths

Sequence inserts >20,000 bases

Some reads >200 kb

Highest Consensus Accuracy

Achieves >99.999% (QV50)

Lack of systematic sequencing errors

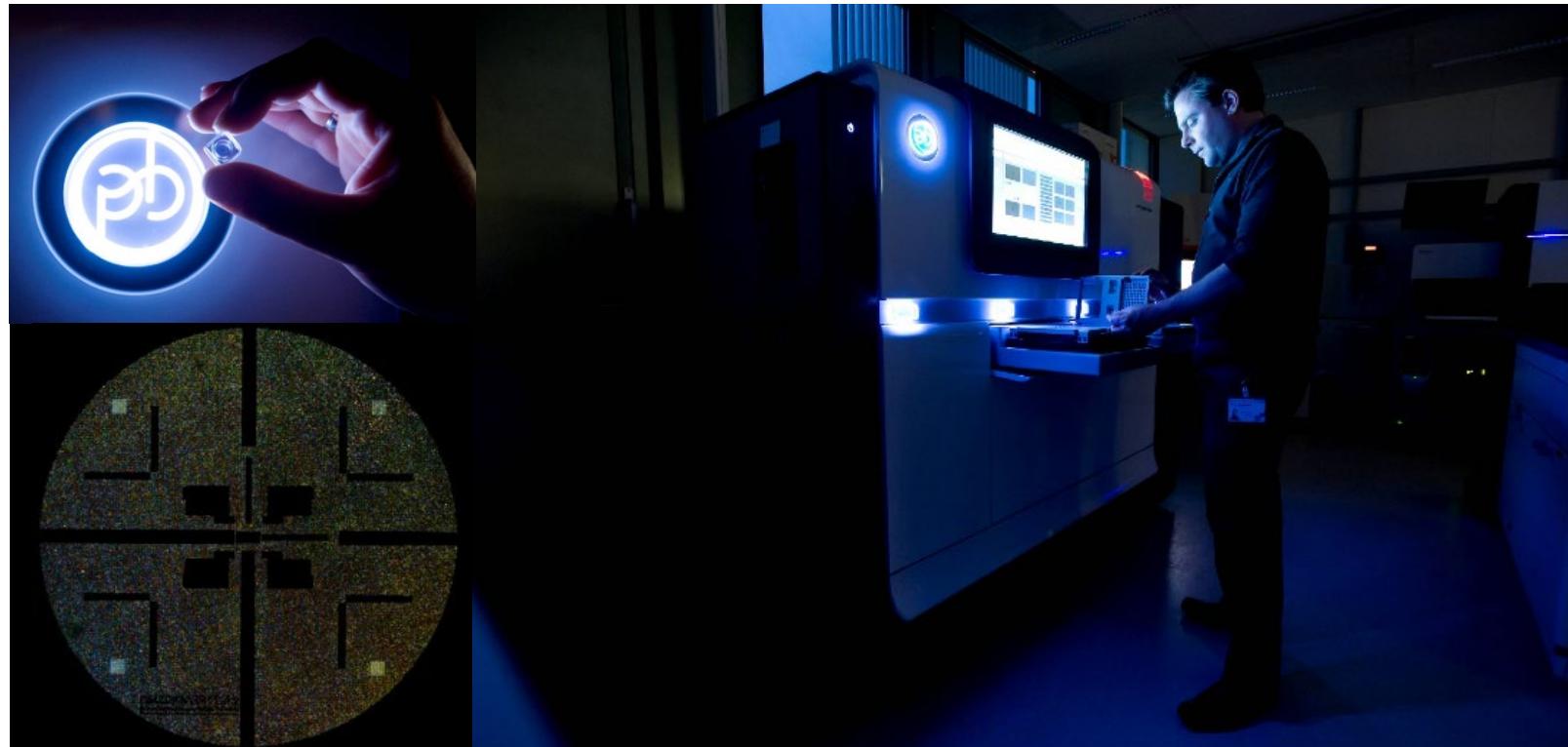
Greatest Uniformity

Lack of GC content or sequence complexity bias

Can Sequence Native DNA

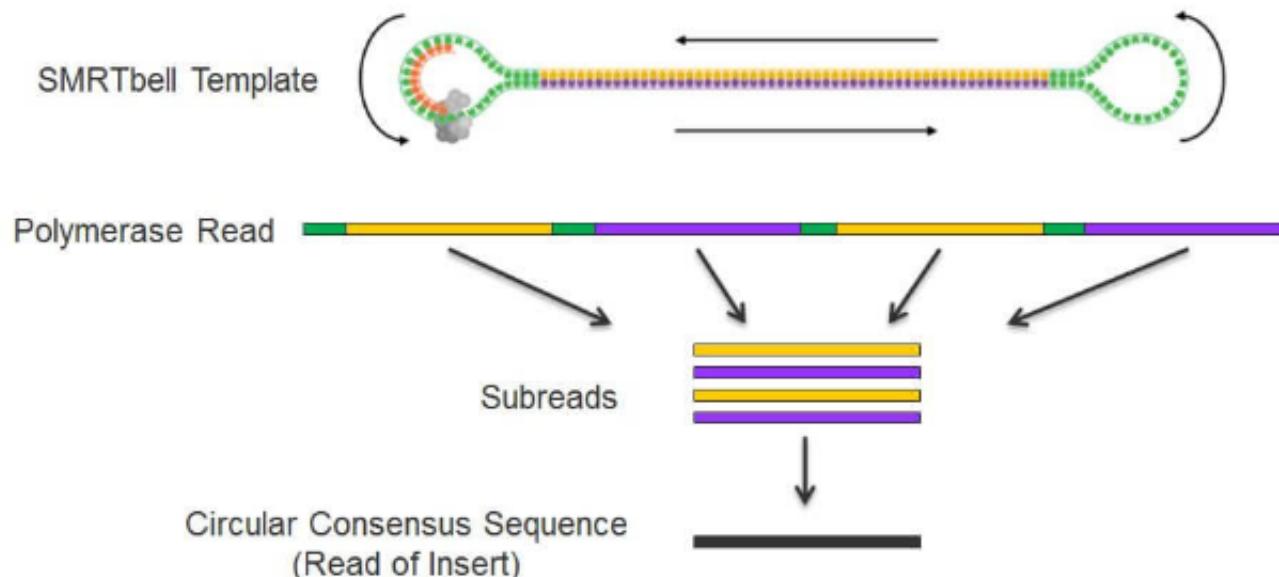
No DNA amplification

Epigenome characterization



SMRT sequencing library prep

PacBio circular consensus sequencing method provides long reads with high accuracy at the single-molecule level

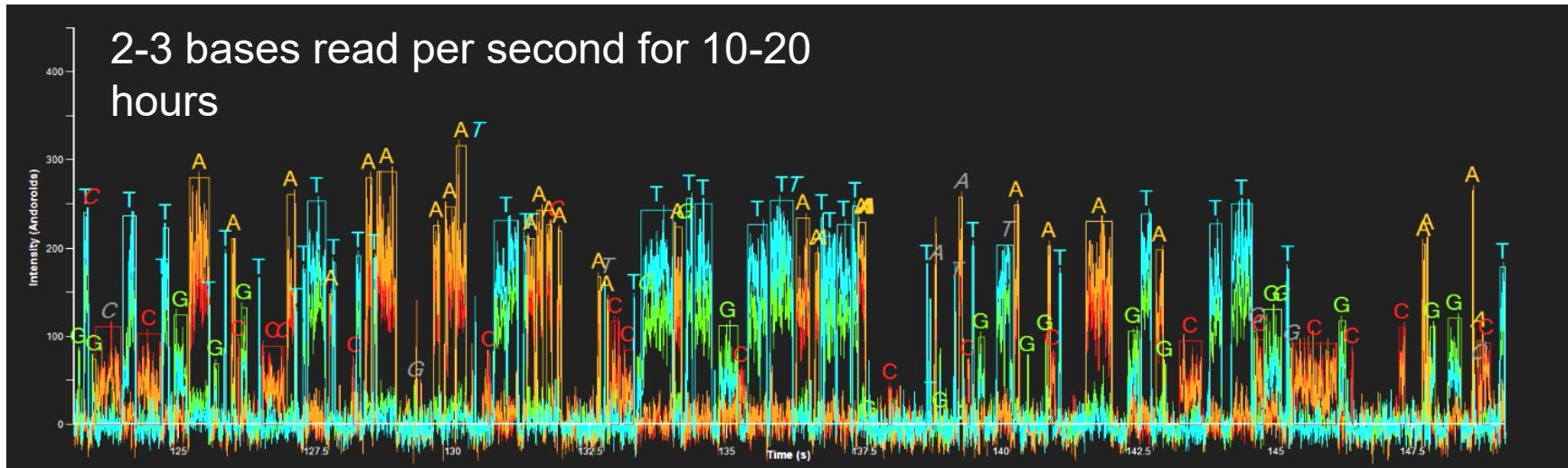


The circular nature of the SMRTbell DNA template allows polymerase to sequence the same DNA molecule multiple times with multiple passes. This produces high intra-molecular consensus accuracy.

PacBio: How it works:

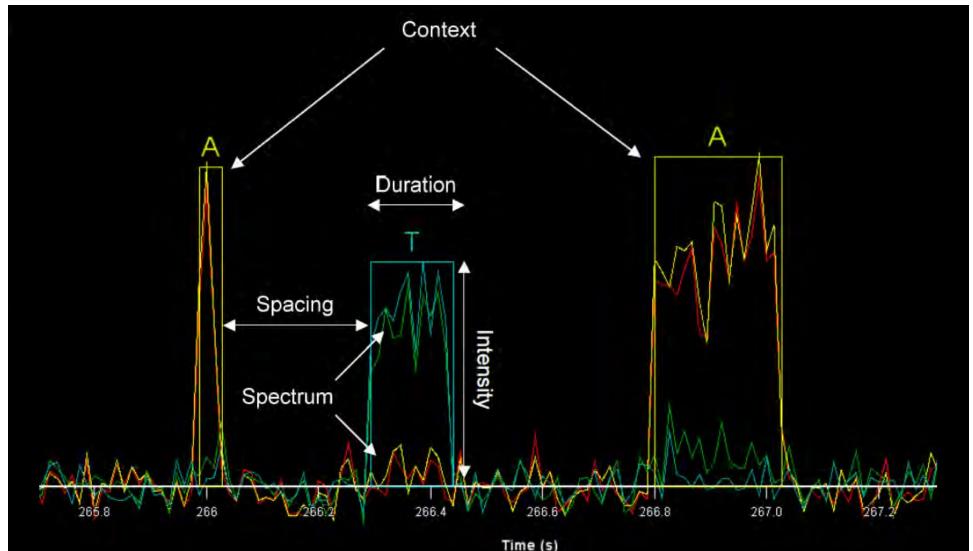
<https://www.youtube.com/watch?v=WMZmG00uhwU>

PacBio Sequel: sequencing and optics



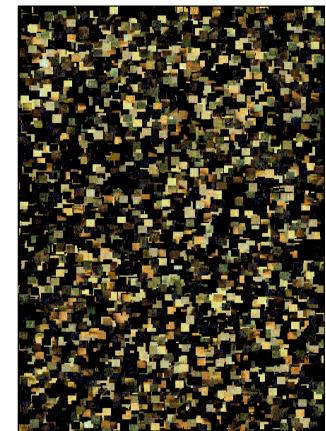
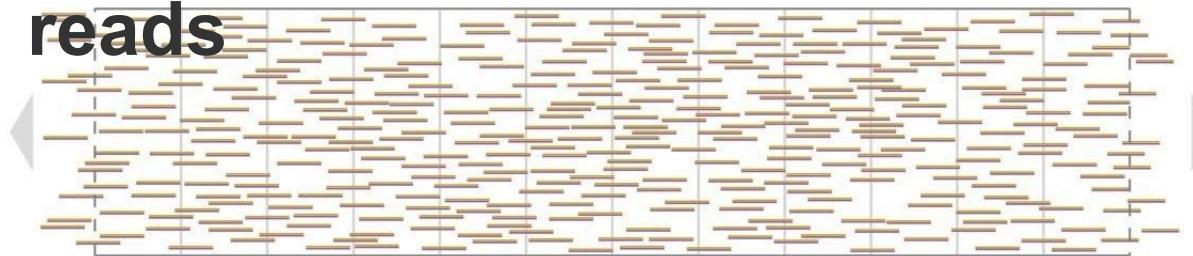
Measurements for base calling:

- Duration
- Spectrum
- Intensity
- Spacing relative to neighbors
- Local context

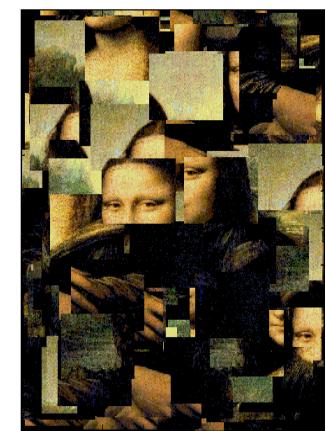
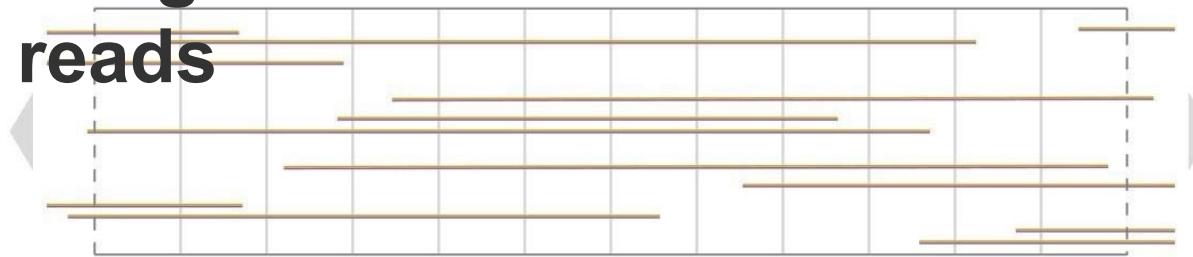


PacBio Sequel: advantages of long reads

**Short
reads**



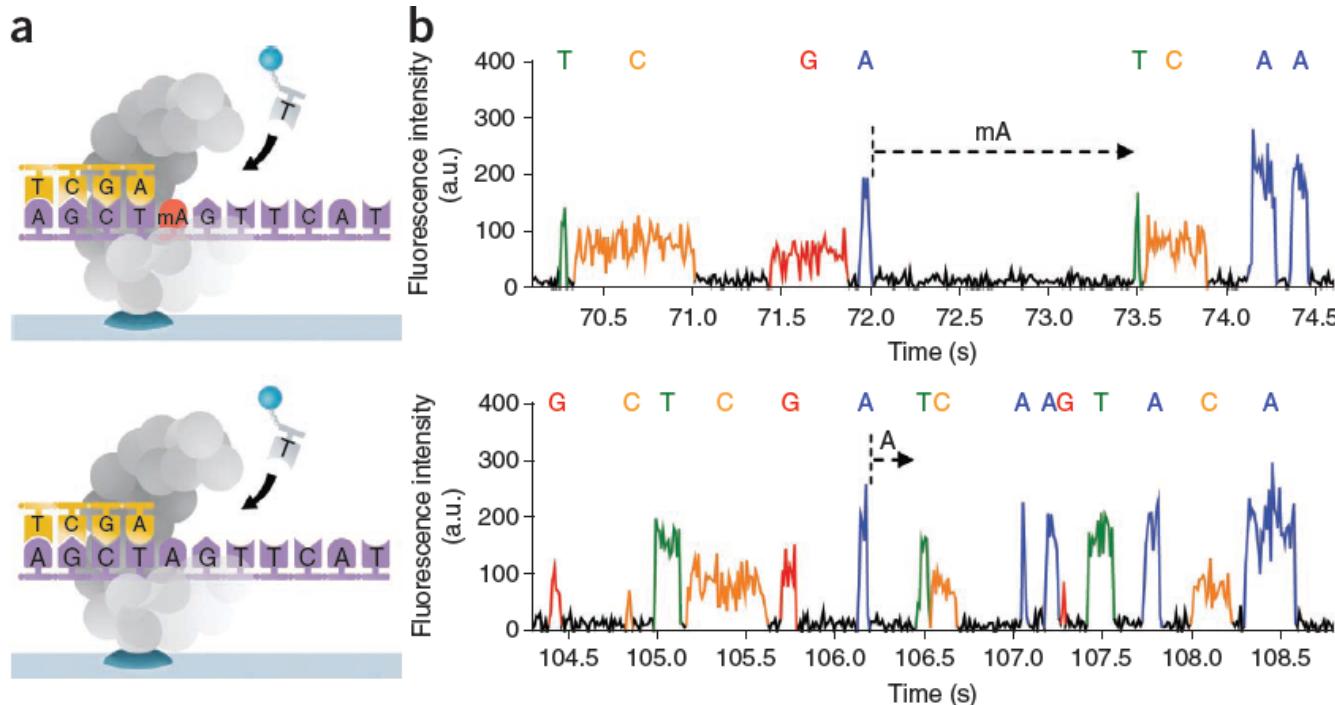
**Long
reads**



PacBio Sequel: Epigenetics

Potential epigenetic applications

Elongation of a “mA” template takes longer than an “A”



3rd Generation Sequencing Continued: Oxford Nanopore

SEQUENCE

Nanopore devices perform DNA/RNA sequencing directly and in real time.

The technology is scalable from miniature devices to high-throughput installations.

Which device is best for you?



SmidgION



Flongle



MinION



GridION



PromethION

50(?) channels

126

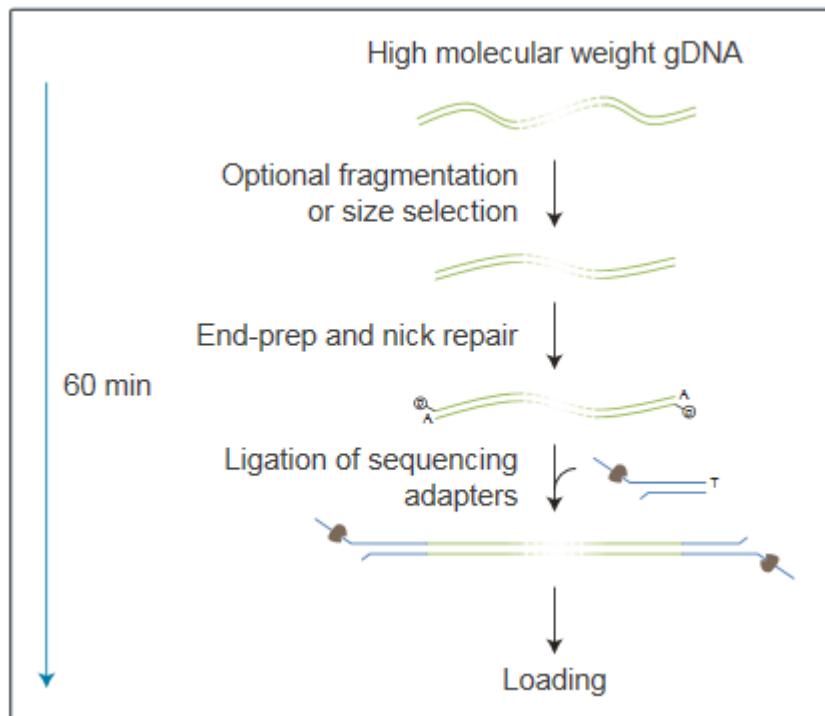
512

512 x 5

3000 x 48

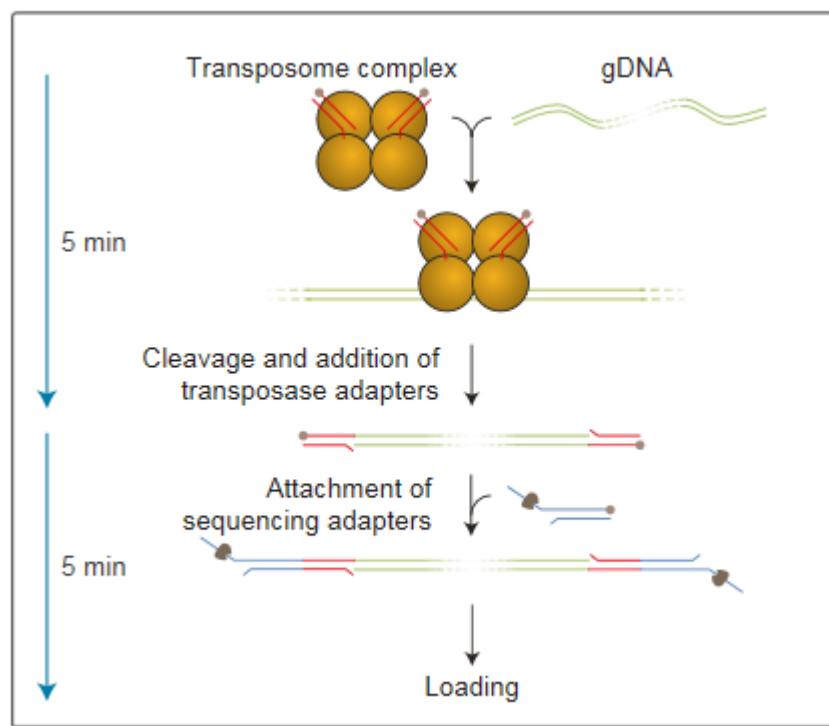
Oxford Nanopore library prep

Starting material: gDNA,
cDNA, amplicons



Ligation-based

Starting material:
gDNA only



Transposon-based

ONP - how it works: <https://nanoporetech.com/resource-centre/nanopore-dna-sequencing>

Oxford Nanopore MinION



System is commercially available with a growing community

Source:
<http://www.nanoporetech.com>

PacBio (SMRT) sequencing - applications

High-quality sequencing for genomes, transcriptomes and epigenomes



Whole Genome Sequencing

For humans, plants, animals and microbes including *de novo* sequencing and structural variant detection



Complex Populations

Understand variants among bacterial, viral and cancer cell populations



RNA Sequencing

In-depth analysis of cDNA sequences across the entire transcriptome or targeted genes



Epigenetics

Detect DNA modifications in your samples while you sequence on the PacBio platform



Targeted Sequencing

Study relevant genome targets across any regions of interest

Oxford nanopore sequencing - applications

Whole genome sequencing

De novo assembly

Scaffolding and finishing

Variant analysis: structural variation

Variant analysis: SNVs, phasing

Resequencing

Targeted sequencing

Panels – amplicons, sequence capture, exome

Variant analysis: structural variation

Variant analysis: SNVs, phasing

16S rRNA analysis

RNA sequencing

Splice variant analysis

Transcriptome / gene expression

Fusion transcript analysis

Metagenomics

Real-time, unbiased analysis of mixed samples

Epigenetics

Methylation

Histone modification

Non-coding RNA activity

3rd generation technologies: Pros and Cons

Pros

- Long reads
- De novo assembly → platinum genomes (PacBio)
- Detection of epigenetic signature
- No problems with structural rearrangements and context
- Sequencing native DNA → no amplification
- Iso-Seq → Full-length transcript sequencing (PacBio)
- Direct RNA sequencing (ONP)

Cons

- Low to medium output
- Sample prep (PacBio)
- Homopolymers (especially ONP)
- High error rate (ONP)

Sequel vs PromethION

	Sequel	PromethION
Read length	100 kb	up to 2 Mb
Error rate	10-15% (no corr.)	2-5%
Output	10-20 Gb	< 75 Gb
# of reads	500k	up to 6M
Instrument price/access fee	€375k	None
Run price	€ 1000	€ 2000

But wait, there's more! 10x Genomics



Single Cell Gene Expression

3' gene expression profiling at scale with single cell resolution.



Genome & Exome

Long-range analysis and phasing of SNVs, indels, and structural variants.



Single Cell Immune Profiling

V(D)J repertoires of T and B cells integrated with 5' Gene Expression.



De Novo Assembly

Everyday *de novo* assemblies for reference-free genomic analysis.

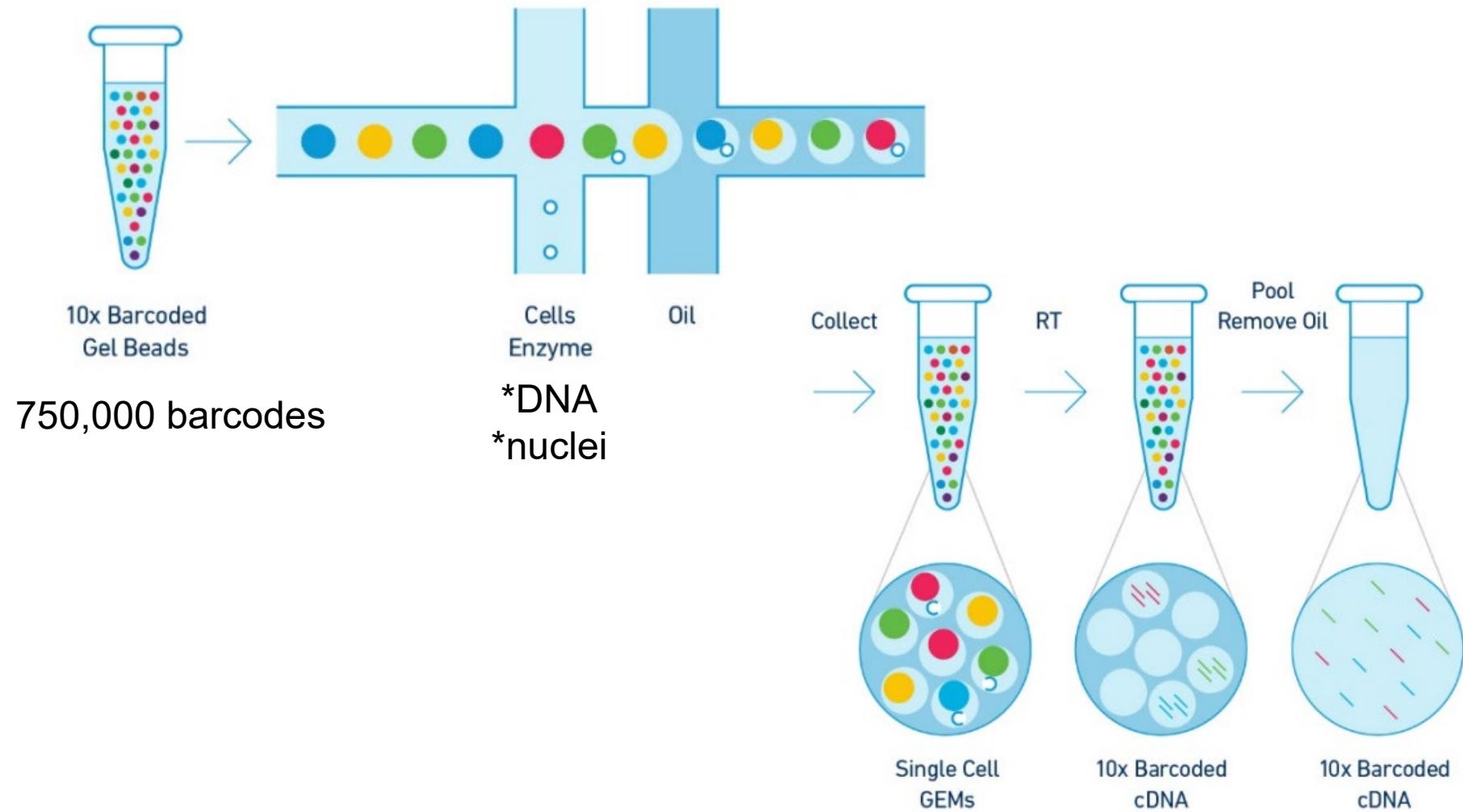


Single Cell CNV

Copy number variation and genomic heterogeneity at single cell resolution.

Single Cell Epigenetics (ATAC-seq) Coming in 2 weeks!

10x Genomics: GEM technology



10x Genomics: Applications

10x Advantages in Cancer Research

1

Unmask Tumor Heterogeneity

Build a comprehensive view of tumor sub-clones and clonal evolution.



3

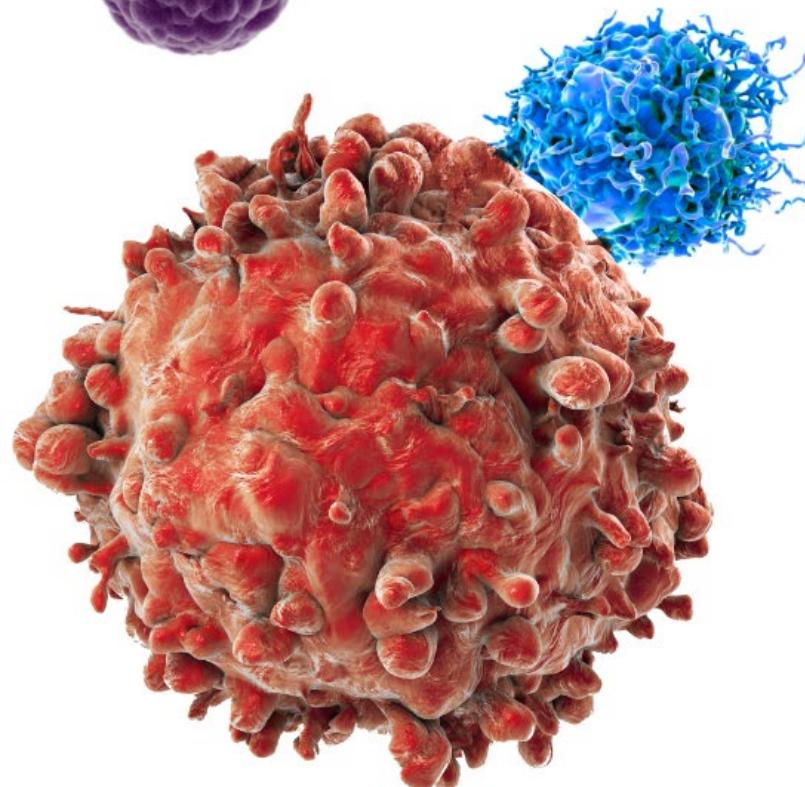
Unravel the Tumor Microenvironment

Determine the identity and heterogeneity of cell types in the tumor microenvironment.

4

Map the Immune Response

Elucidate tumor infiltrating lymphocyte phenotypes and clonal antigen recognition within the tissue microenvironment, and research molecular and cellular responses to novel immunotherapies derived from single cell genomics' discoveries.



Better Characterize Mutations in Cancer Susceptibility Genes

Reveal the molecular genetic basis for cancer by accurately detecting and resolving ambiguous variants, while maintaining haplotype information.

Wish list for the next NGS platforms

Cheaper machines

Cheaper reagents

Faster

Flexible data output per run (as much as is needed)

Longer reads

More accurate data

Less work to generate the data

→ Easy data analysis solutions

Acknowledgements

Dept of Human Genetics / LGTC

Yavuz Ariyurek

Rolf Vossen

Emile de Meijer

Roberta Menafra

Yahya Anvar

Guy Allard

Jeroen Laros

Johan den Dunnen

Contact:

info@lgtc.nl

S.L.Kloet@lumc.nl

