Introduction to Illumina Core NGS Technology

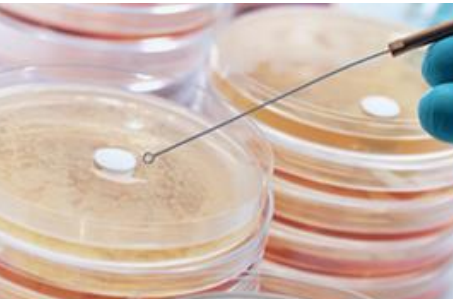
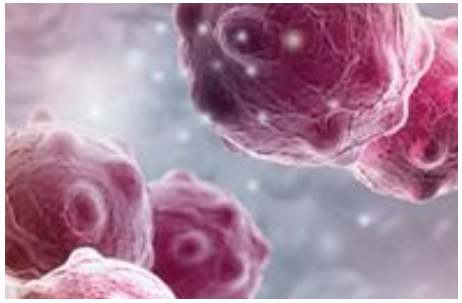
Thomas Yang |
Sr. Territory Account Manager, Taiwan



COMPANY CONFIDENTIAL – FOR INTERNAL USE ONLY © 2016 Illumina, Inc. All rights reserved.
Illumina, 24sure, BaseSpace, BeadArray, BlueFish, BlueFuse, BlueGnome, cBot, CSPro, CytoChip, DesignStudio, Epicentre, ForenSeq, Genetic Energy, GenomeStudio, GoldenGate, HiScan, HiSeq, HiSeq X, Infinium, iScan, iSelect, MiniSeq, MiSeq, MiSeqDx, MiSeq FGx, NeoPrep, NextBio, Nextera, NextSeq, Powered by Illumina, SureMDA, TruGenome, TruSeq, TruSight, Understand Your Genome, UYG, VeraCode, verifi, VeriSeq, the pumpkin orange color, and the streaming bases design are trademarks of Illumina, Inc. and/or its affiliate(s) in the US and/or other countries. All other names, logos, and other trademarks are the property of their respective owners.

illumina®

Our Mission



To improve human health by unlocking the power of the genome







For Research Use Only. Not for use in diagnostic procedures.

illumina®

Our Background

Company

April, 1998
Founded

July 27, 2000
IPO

Francis deSouza
President and CEO

~7,000
Employees

San Diego, CA
Headquarters

22
Offices Globally

90% DNA sequenced on ILMN instruments

Awards

Top 10 Most Innovative Companies in Biotech
2017 Fast Company

10 Breakthrough Technologies
2016 MIT Technology Review

Top 3 Smartest Companies
2016 MIT Technology Review

Top 10 Innovations
2015 The Scientist Magazine

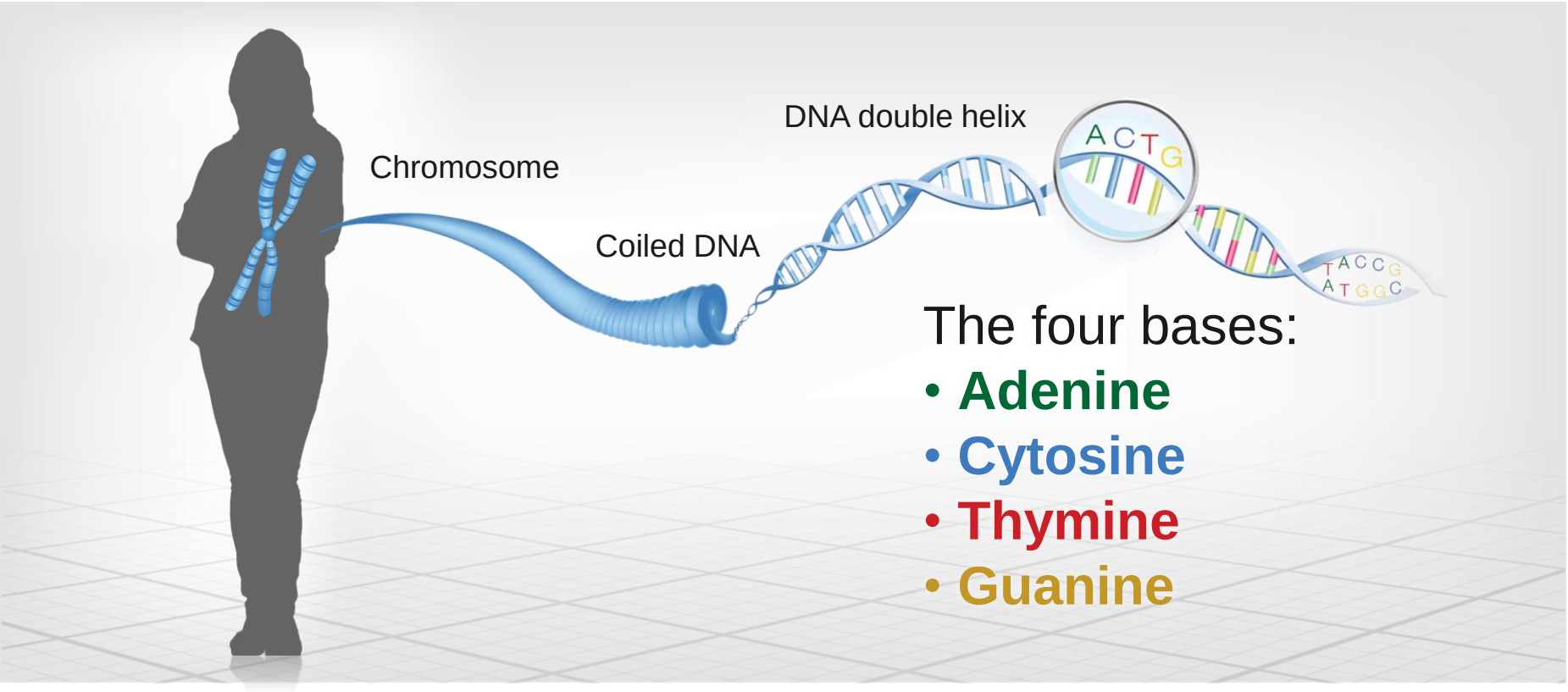
12 Most Disruptive Names In Business
2013 Forbes

Top 10 Innovations
2012 The Scientist Magazine

Fastest Growing Tech Companies
2010 Forbes

illumina®

Understanding Genomics



illumina®

Gaining a Global Sequencing Presence

Breadth of technology enabling markets worldwide

>90% of all the sequencing performed on earth has been with Illumina systems and chemistry[1]

>40.6K
Publications based on Illumina technology

>9K
Installs Worldwide[2]

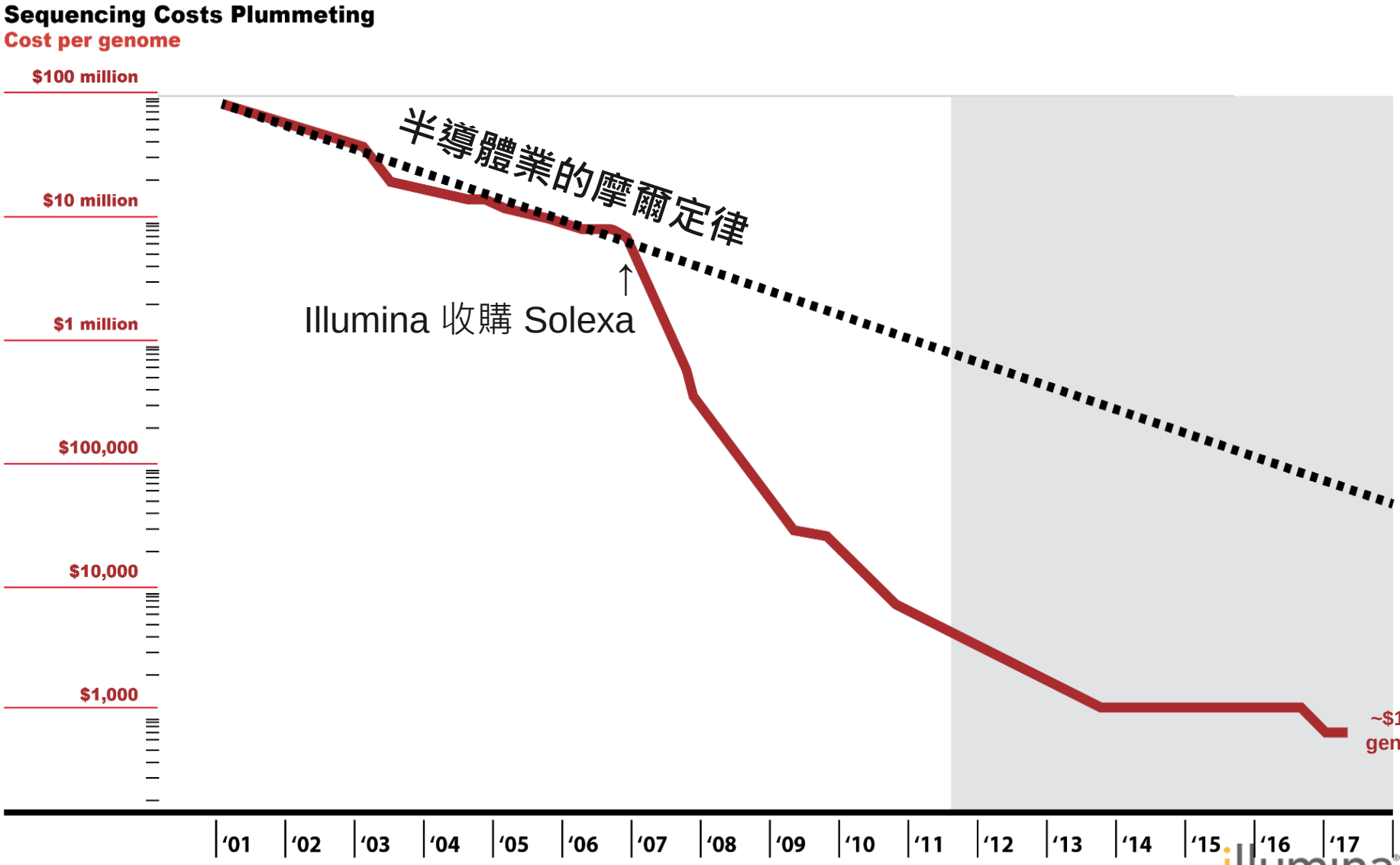
Sequencing Power
for virtually every scale



COMPANY CONFIDENTIAL – FOR INTERNAL USE ONLY

illumina®

Illumina的技術突破大幅降低基因定序成本



COMPANY CONFIDENTIAL – FOR INTERNAL USE ONLY

Source: U.S. National Human Genome Research Institute, *Nature*

Who We Serve

Innovation drives expanding market opportunities

illumina®



Reproductive Health



Oncology



Population Sequencing



Research



Complex Disease



Consumer



Infectious Disease



Forensics



Agriculture



Genetic Health

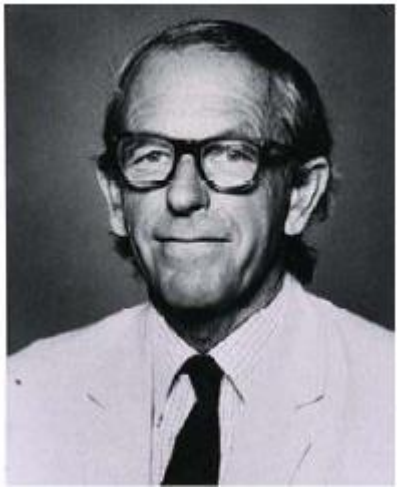


BioPharm

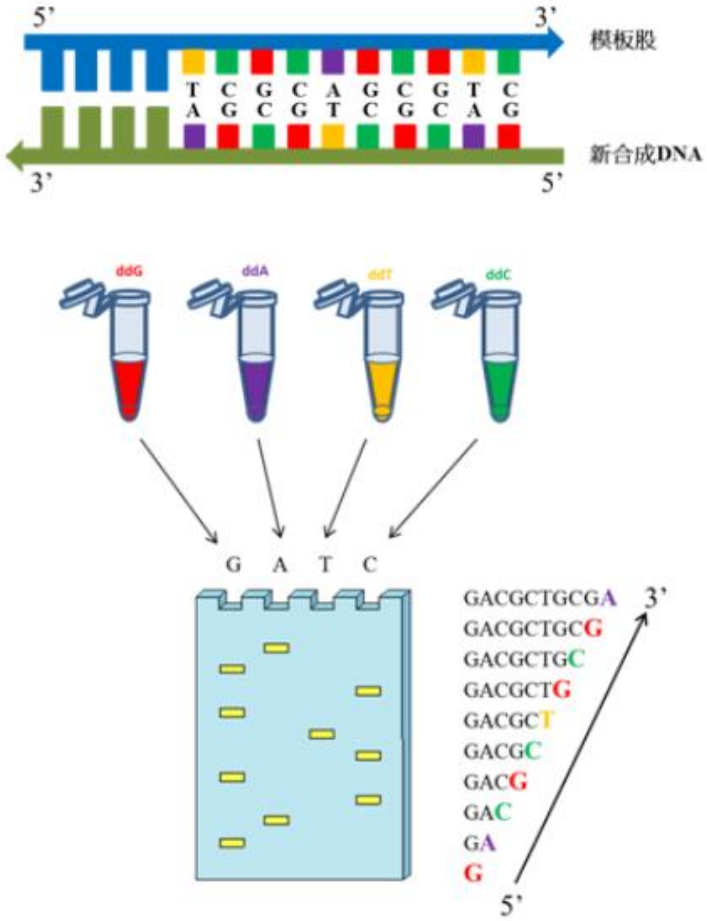
illumina®

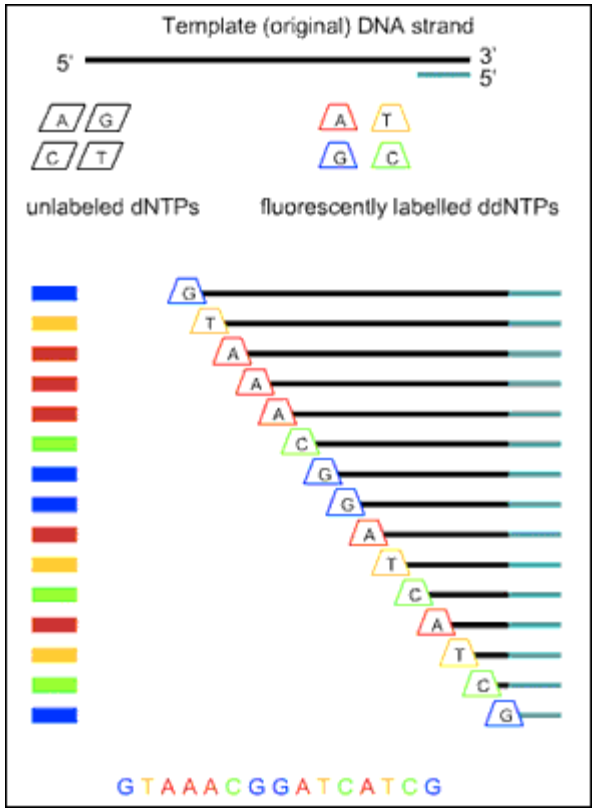
How to read DNA? 1st Gen. or Sanger Seq

Traditional, chain-termination, dideoxynucleotide (dideoxy), Sanger, CE (capillary electrophoresis) sequencing. The sequencing gold standard for the past 30 yrs



Prof. Fed Sanger
Twice awarded
Nobel prize

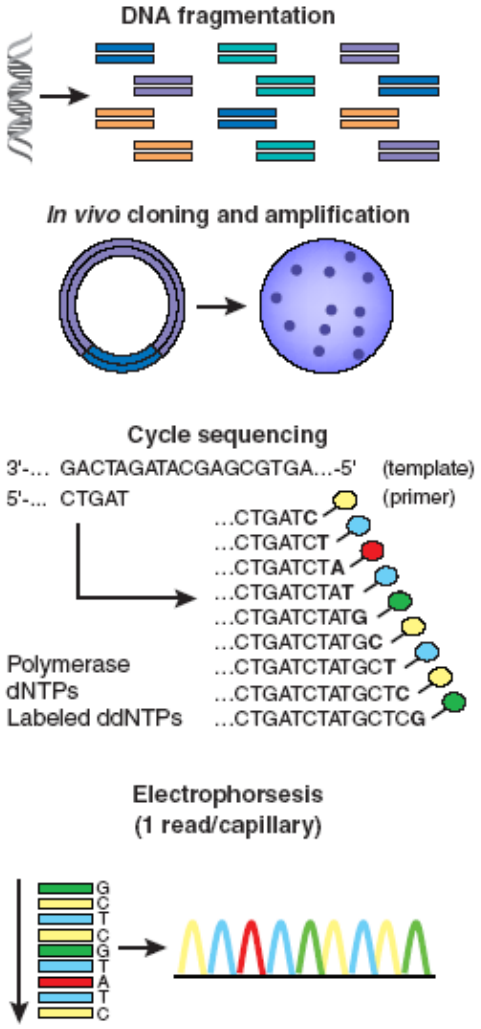




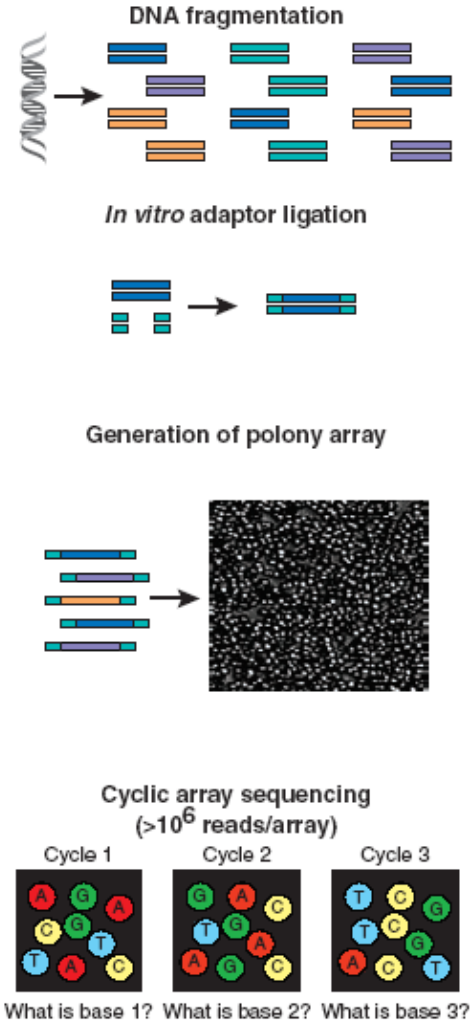
http://www.genomebc.ca/index.php?cID=603

Differences between Sanger and NGS

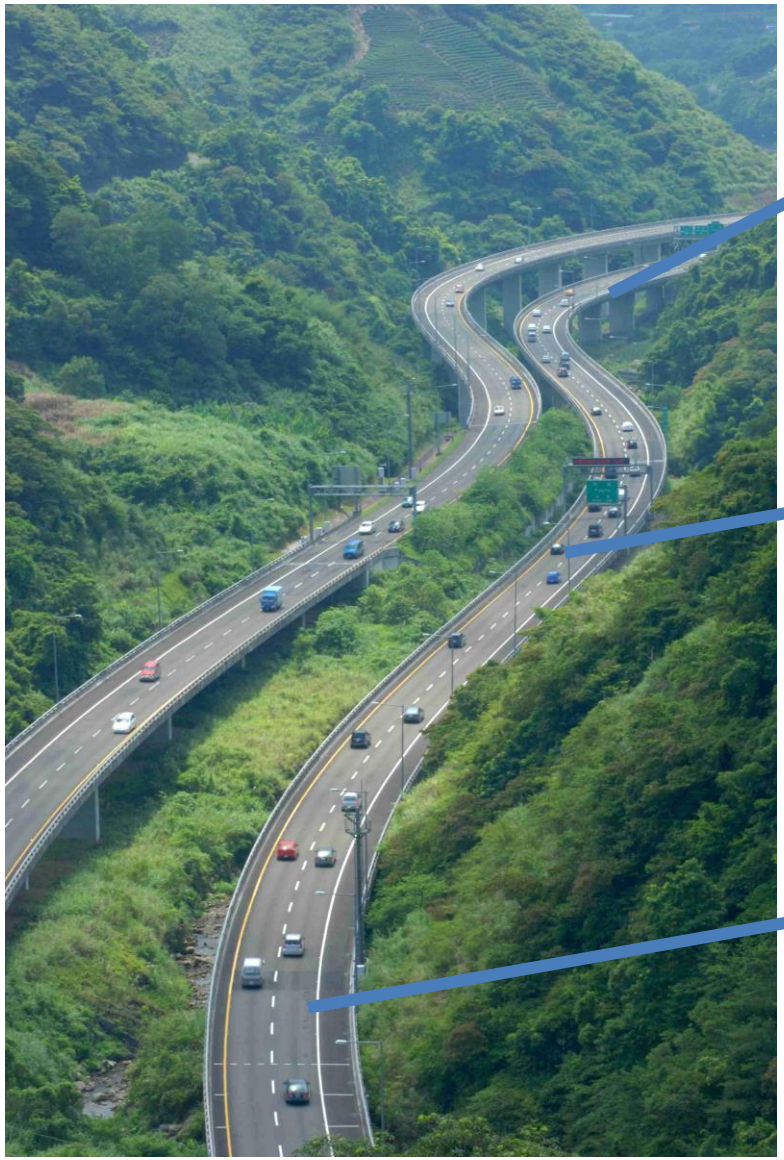
Sanger: DNA synthesis and detection are separate



NGS: DNA synthesis + detection + multiple DNA pieces sequenced



Differences between Sanger and NGS







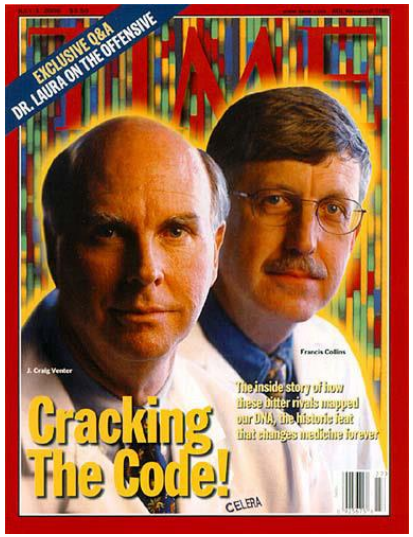


Why is NGS more powerful than Sanger sequencing?

- **Lower sequencing/instrument cost** ($$$$- 0.$)
- **Faster turn around time** (years vs. hrs)
- **Higher throughput (**1Mb vs. 1Tb)
- **Higher sensitivity** (samples mixtures)
- **More automation compatible**

Example: Human Genome Project (1990~2003)

	1990 ~ 2003	2018
Tech	Sanger	NGS
Cost	500$ million USD*	1,000 $ USD
Time	13 yrs	1 hr*
Man power	100+	1 person
Sequencer(s)	1000+	1 unit

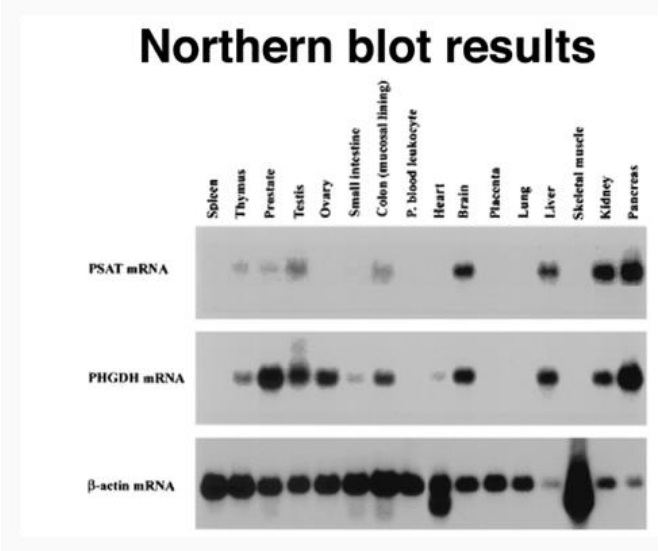


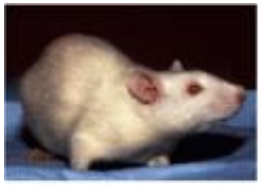
Time cover, July 2000





NGS revolutionized genome research





Rat
(*Rattus norvegicus*)



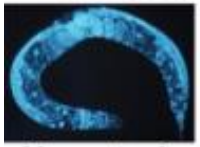
Mouse
(*Mus musculus*)



Fruitfly
(*Drosophila melanogaster*)



Plant
(*Arabidopsis thaliana*)



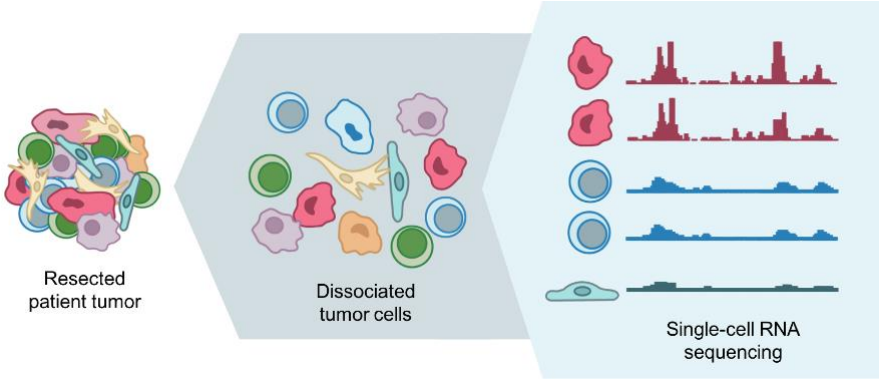
Nematode
(*Caenorhabditis elegans*)



Sea Urchin
(*Strongylocentrotus purpuratus*)



Frog
(*Xenopus laevis*)

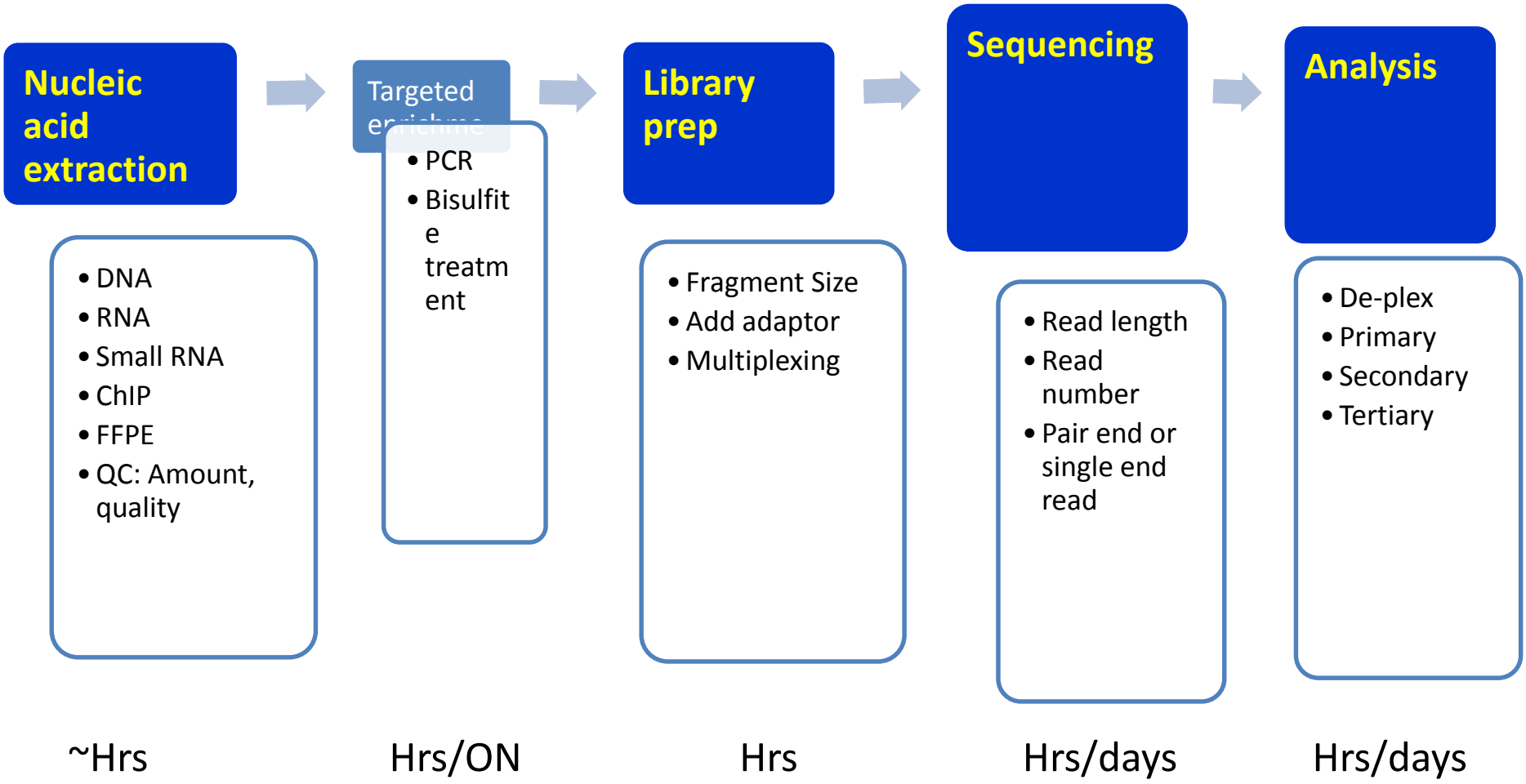




Arabica Coffee Genome Sequenced

Coincides With Birth of California-Grown Specialty Coffee Industry

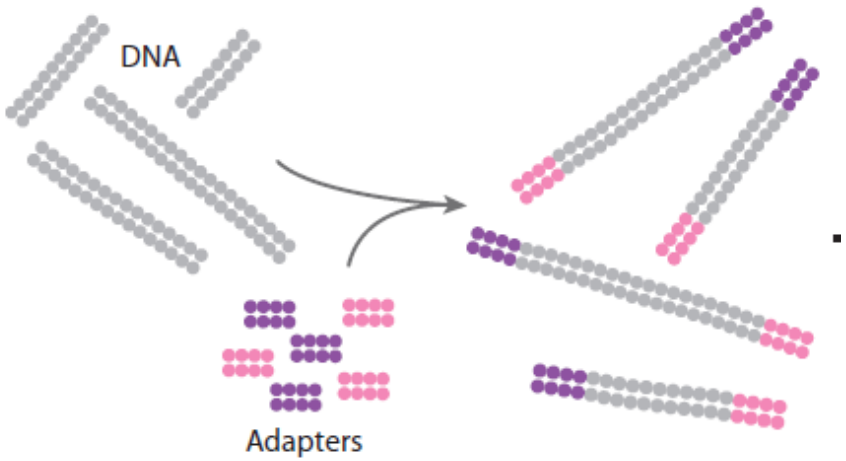
NGS workflow: from DNA to data



Illumina NGS technology summary

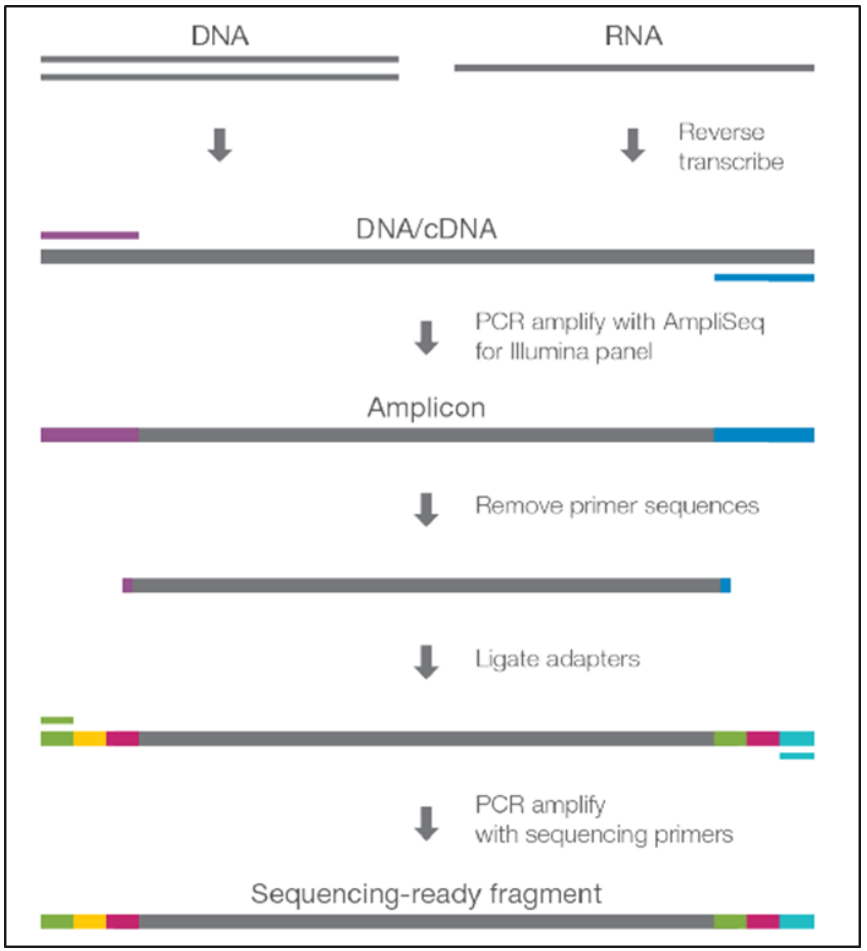
1. Sequencing By Synthesis (SBS)
2. Cluster or bridge amplification on solid surface (flow cell)
3. Colored nucleotides (reversible terminators) to represent 4 different bases (A, T, C, G)

Illumina NGS – Library prep

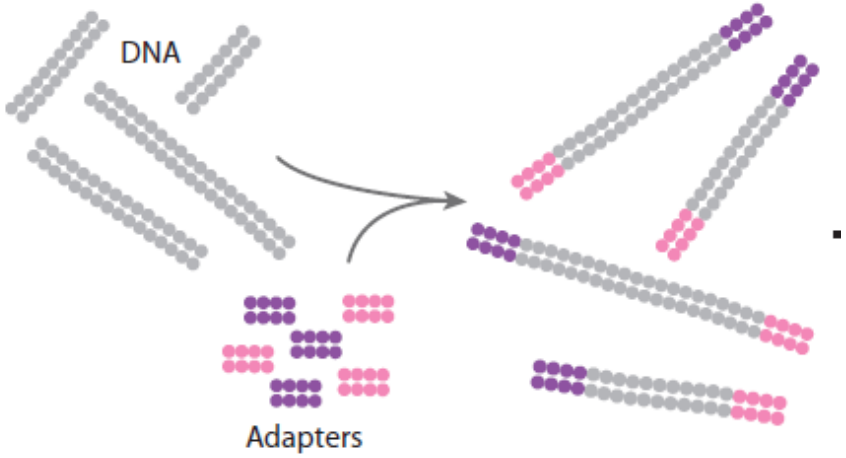


Prepare genomic DNA sample

Randomly fragment genomic DNA and ligate adapters to both ends of the fragments.

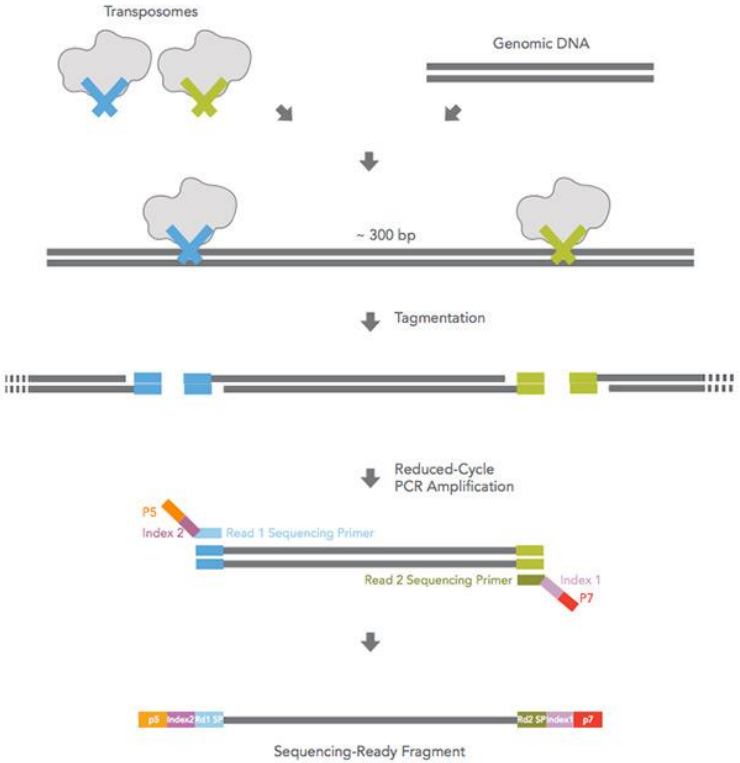


Illumina NGS – Library prep

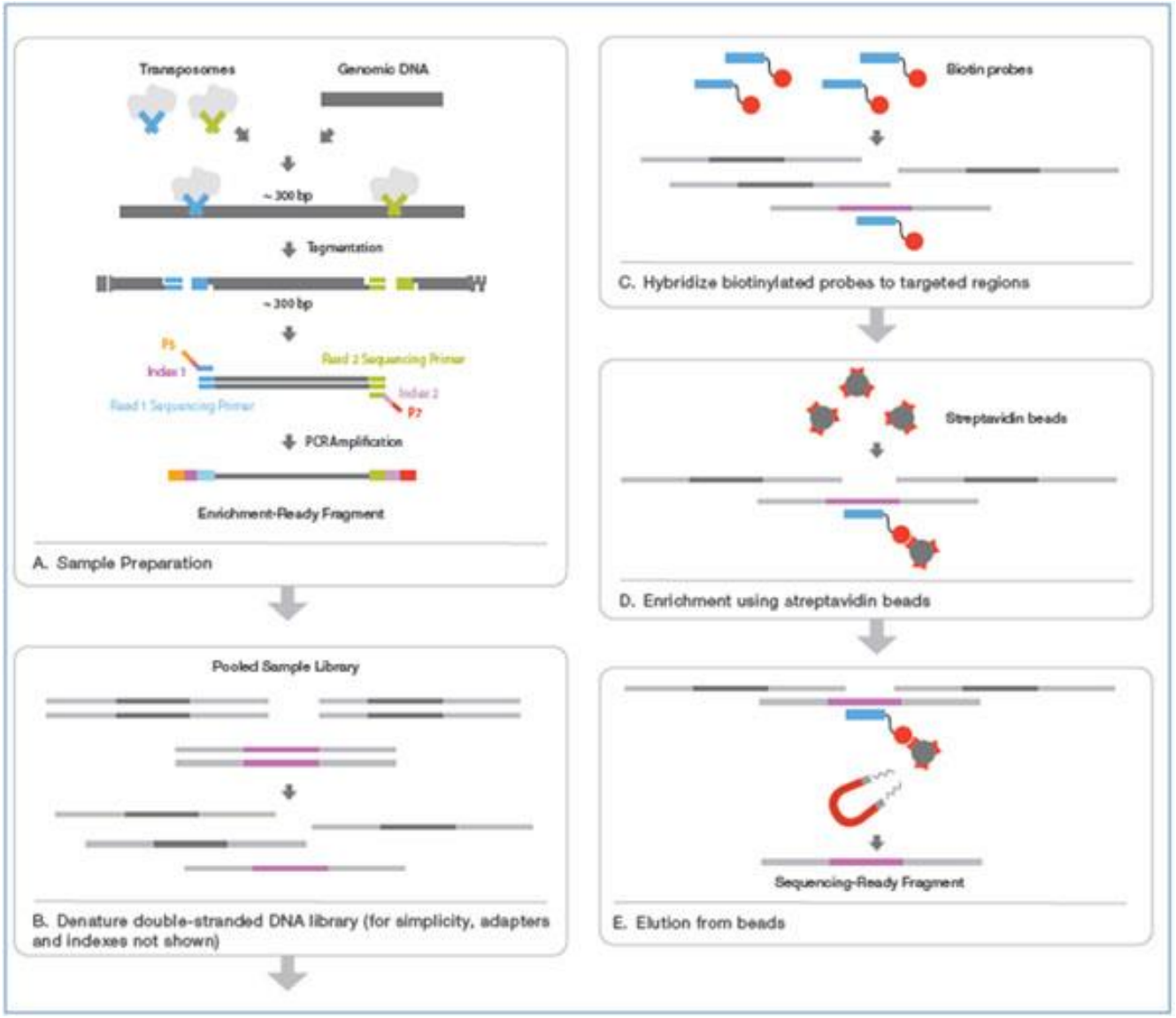


Prepare genomic DNA sample

Randomly fragment genomic DNA and ligate adapters to both ends of the fragments.



Illumina NGS – Library prep



Illumina's kitted Library Prep Solutions

RNA & Methylation

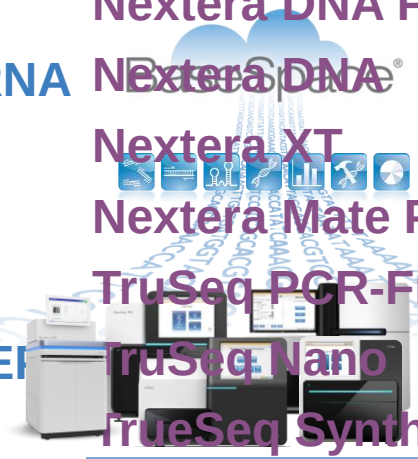
TruSeq Stranded mRNA
TruSeq Stranded Total RNA
TruSeq Small RNA
TruSeq Targeted RNA
TruSeq RNA Exome
TruSeq Methyl Capture EP



Library Prep

DNA

Nextera DNA Flex
Nextera DNA
Nextera XT
Nextera Mate Pair
TruSeq PCR-Free
TruSeq Nano
TrueSeq Synthetic Long Read



Sequence

Targeted DNA

AmpliSeq for Illumina
TruSeq Custom Amplicon
Nextera Rapid Capture Custom
TruSeq DNA Exome
Nextera DNA Exome
TruSight Panels



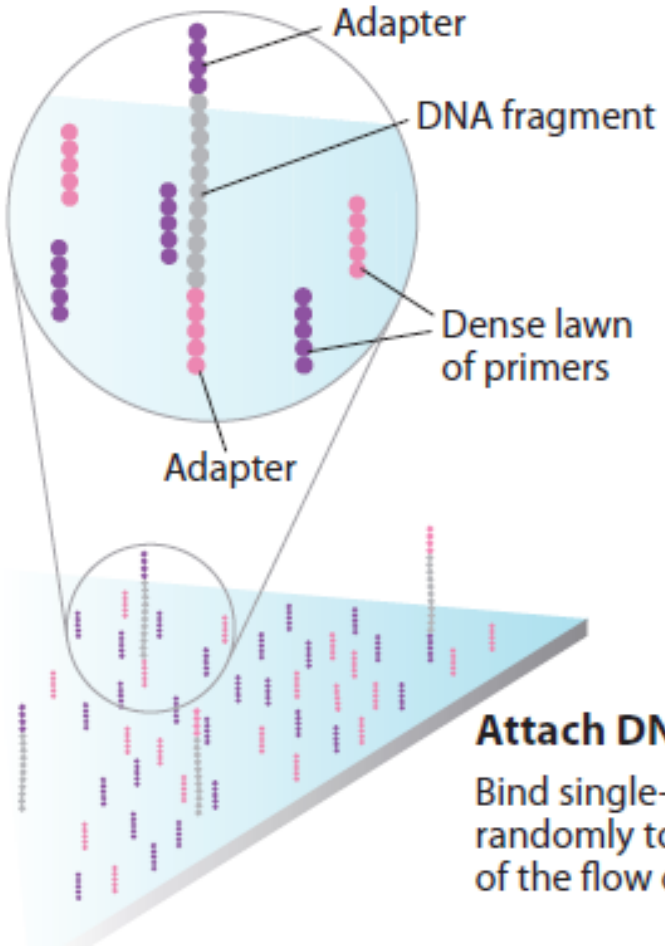
Answer

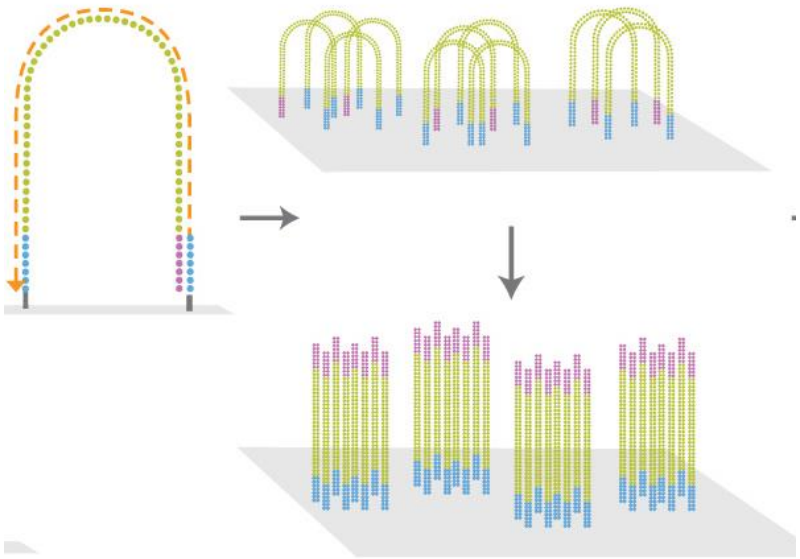




Automation protocols from Agilent, Beckman, Eppendorf, Hamilton, Perkin Elmer, Tecan

Illumina NGS – Library prep



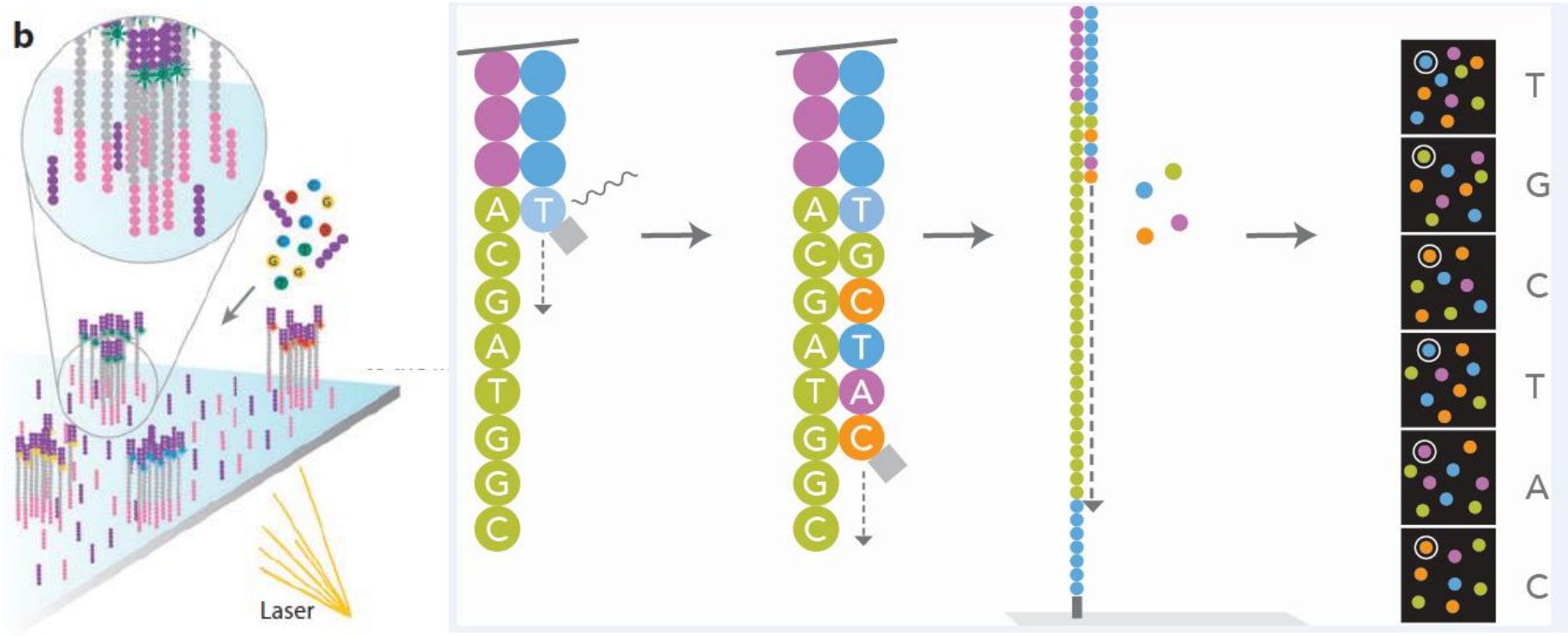


Attach DNA to surface

Bind single-stranded fragments randomly to the inside surface of the flow cell channels.

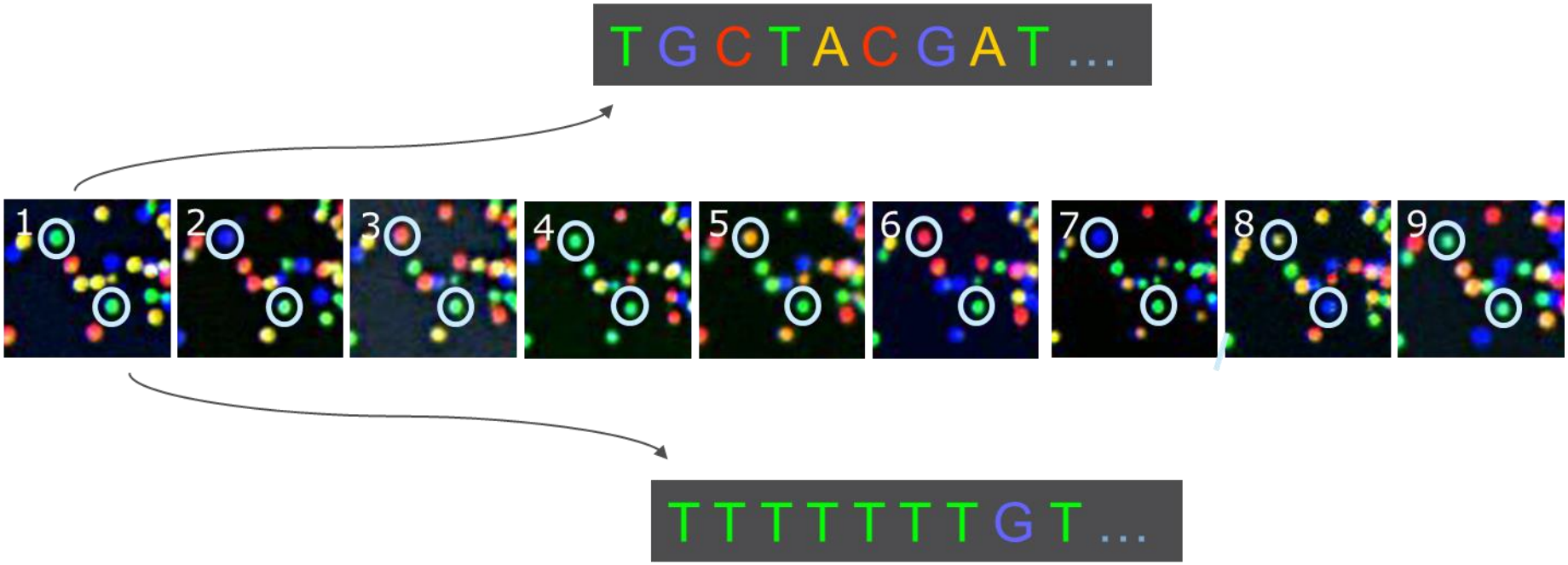
21

NGS technologies – Illumina

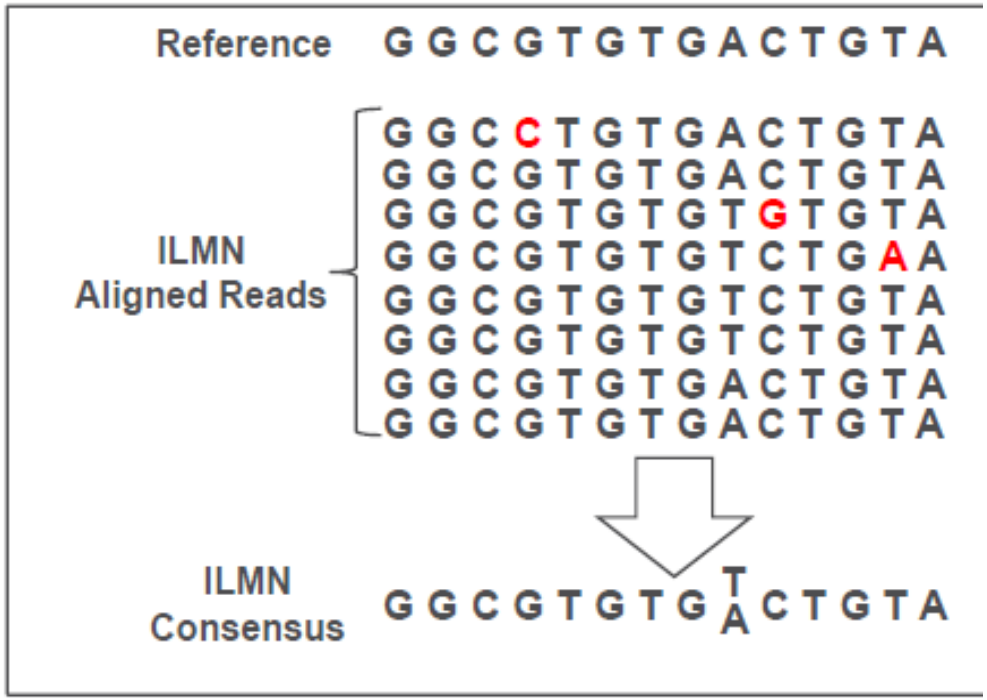


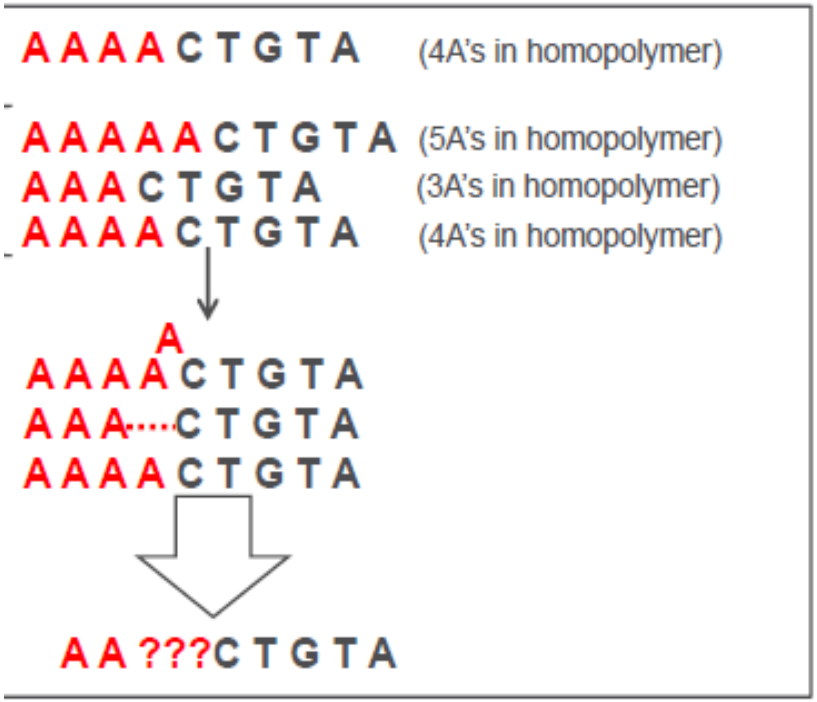
Automated Base Calling from Raw Data:

Reading each base of a cluster from sequential images



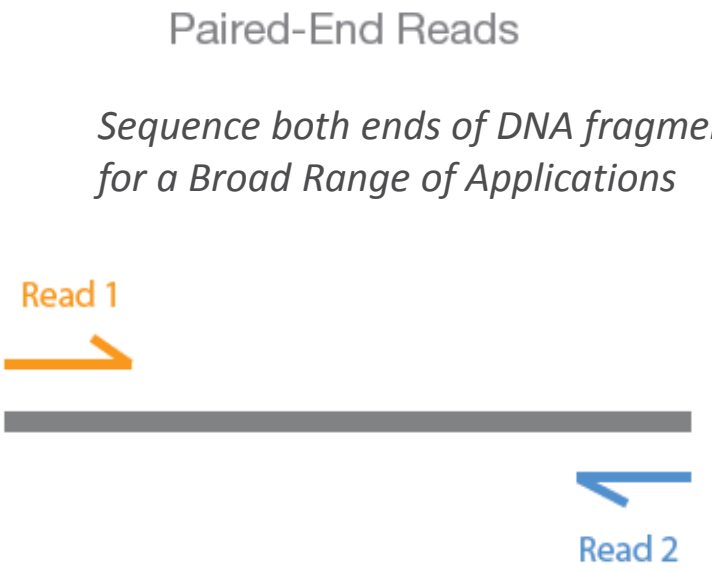
Error caused by homopolymer region

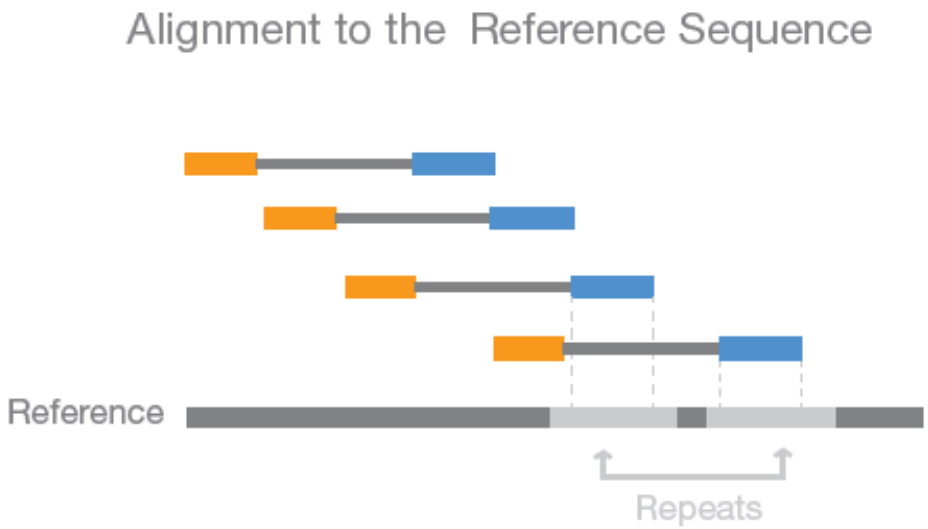




Paired-End Sequencing Advantage:

Unique placement of one end resolve ambiguous placement



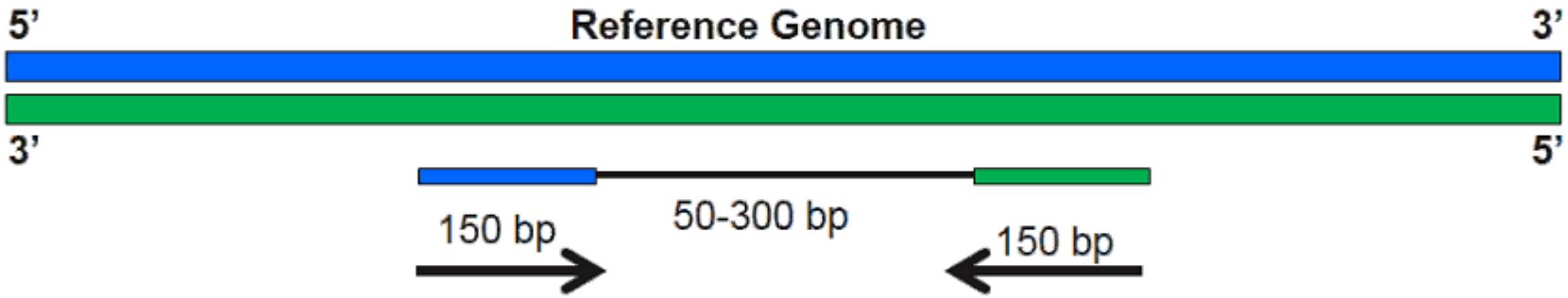


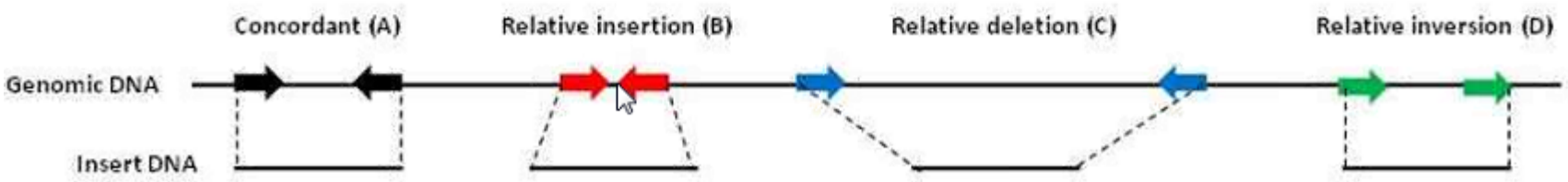
De novo* assembly of *E.coli

- Pair info **not** used : largest assembled fragment = **128 Kb**
- Pair info used : largest assembled fragment = **326 Kb**

illumina®

Structural variants can be determined by evaluating the orientation and the insert size of the aligned read

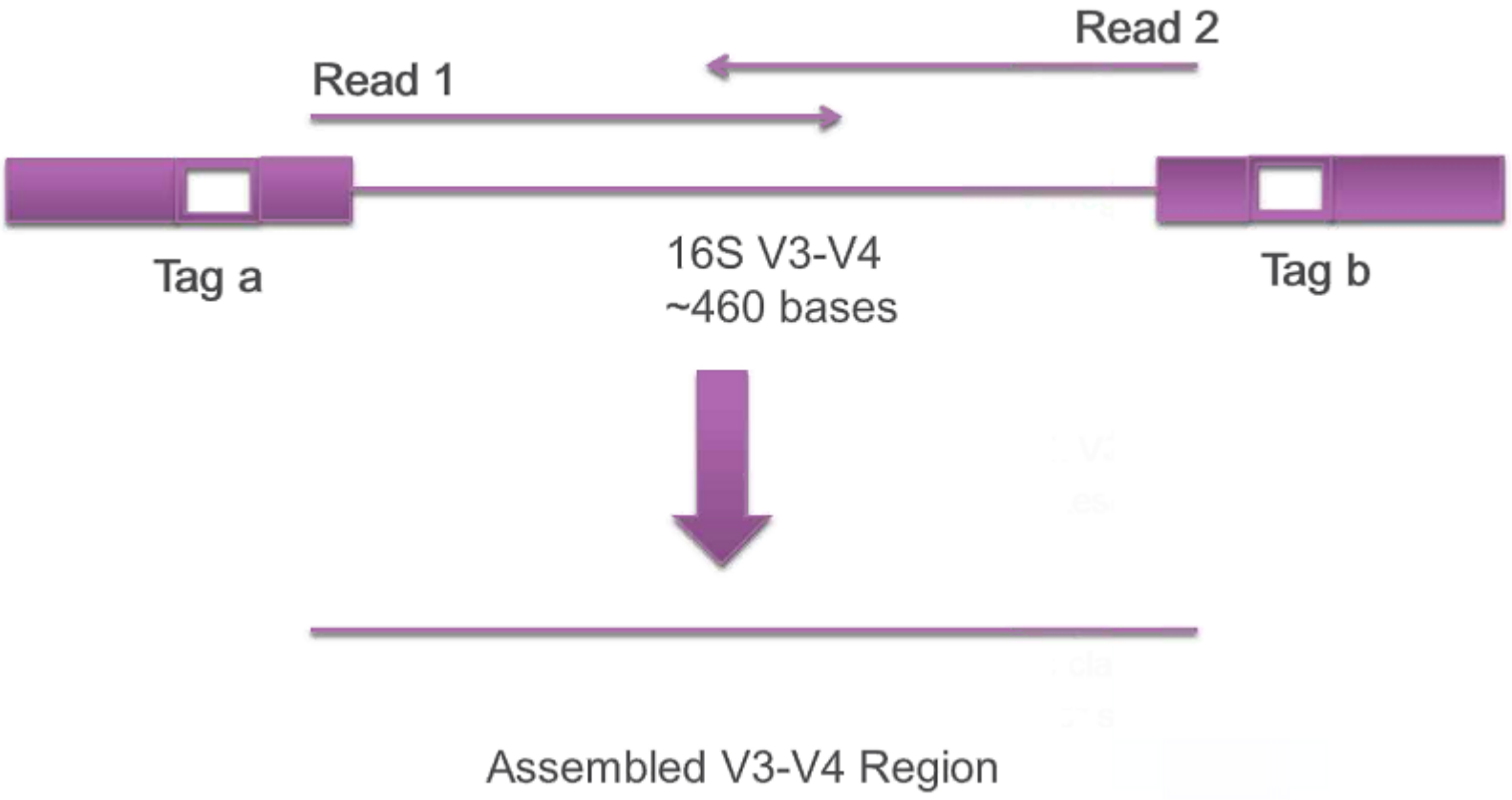




illumina

PE gives overlapping reads in amplicon sequencing

High quality sequencing across the whole insert



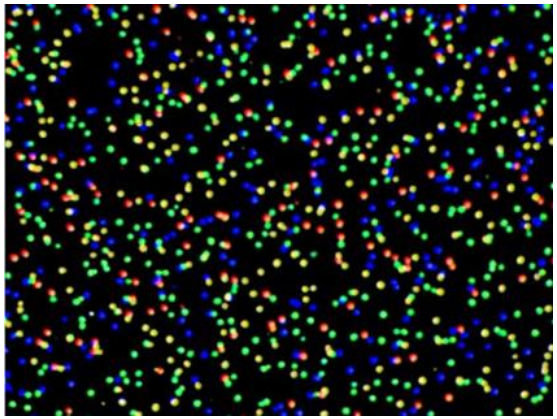
illumina®

How to calculate NGD data output?

From spec. sheet:

Read Length	Data Output
1 x 36 bp	~600 Mb
2 x 100 bp	~3 Gb
2 x 300 bp	~15 Gb



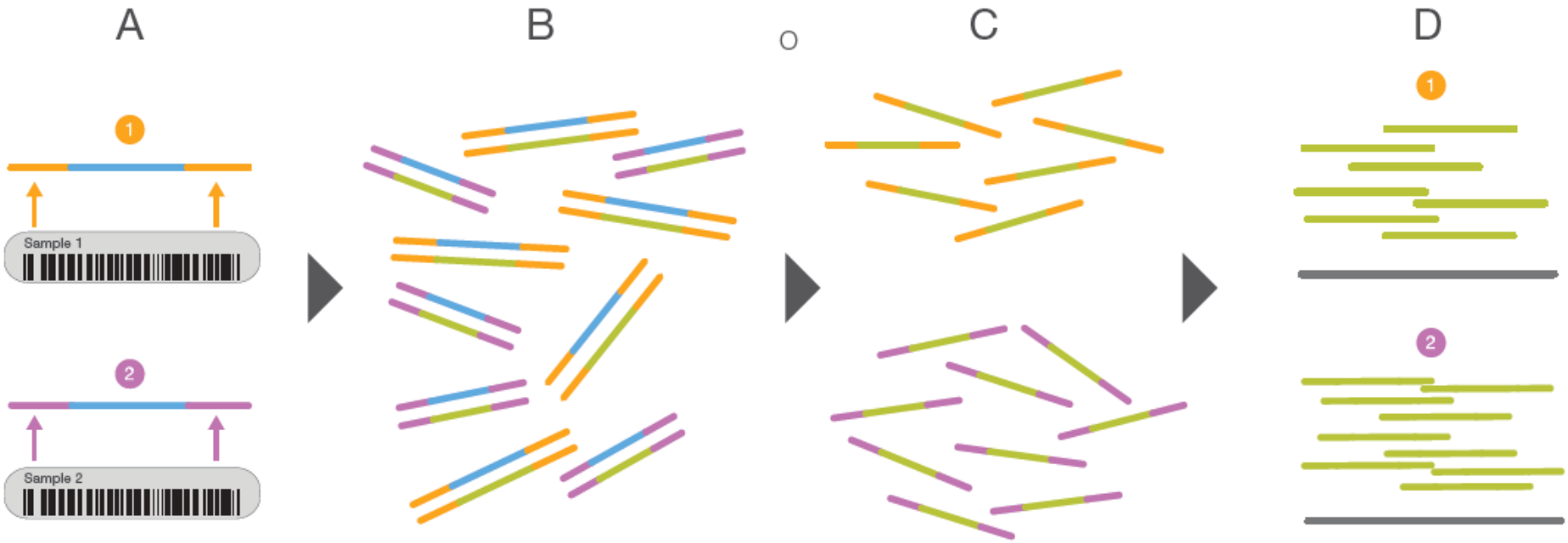


- **Read: Each dot is a DNA strand**
- **Read length: is number of bases sequenced**
- **Output depend on number of reads and length sequenced**
- **MiSeq can sequence 25m DNA fragments per run**

Therefore the data output per run for a 2x300 run is:
25m reads (dots or clusters) X (300x2)bp = **15Gb**

illumina®

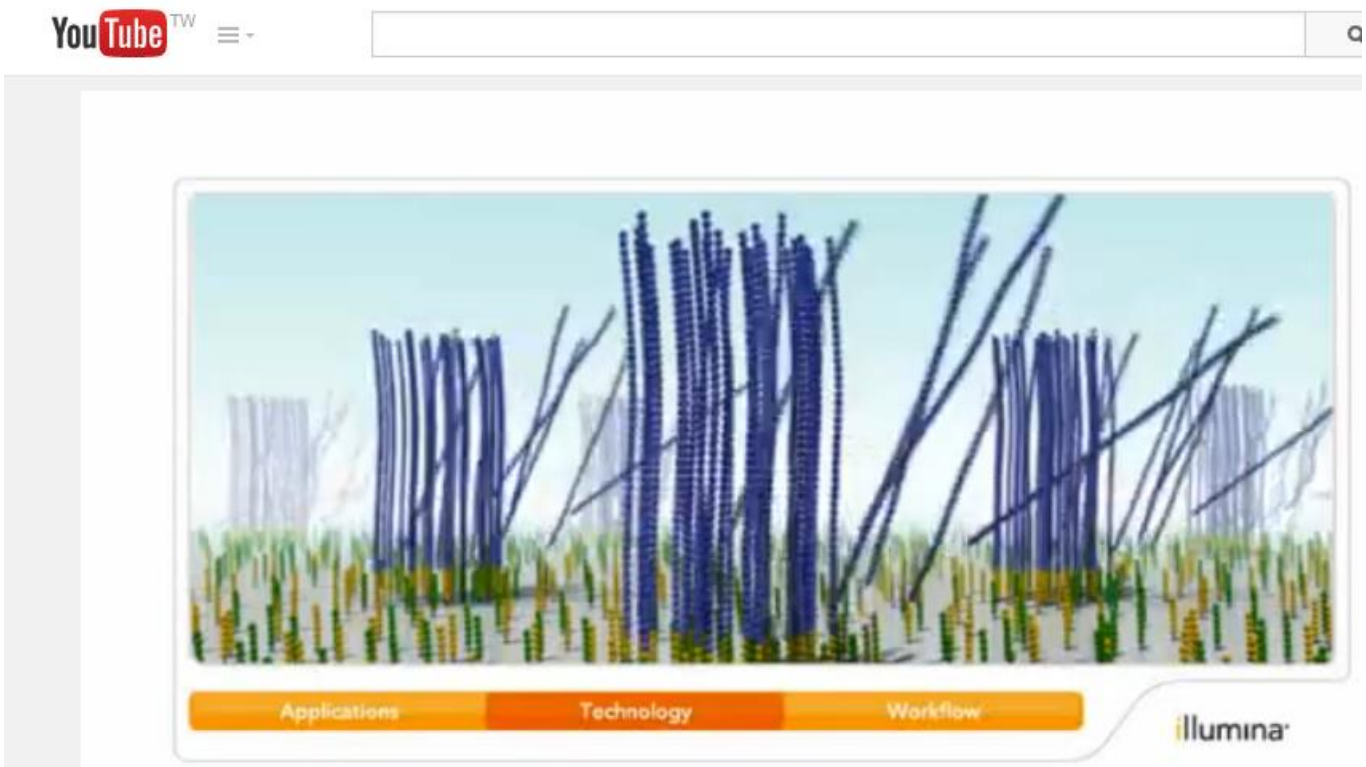
NGS technologies — multiplexing



29

illumina®

Watch illumina's SBS in real time on YouTube



illumina

NGS technologies – Illumina





MiSeq

First all in one desktop sequencer

MiSeqDX First FDA cleared IVD NGS sequencer

NextSeq

First high throughput desktop sequencer

HiSeq

First 1k$ Human genome

First one day Human genome

NovaSeq

launched 2017 Aim to achieve 100$ 1 human genome

Illumina's Sequencing-by-Synthesis (SBS) NGS technology

- 8億
- 16億
- 40億
- 200億

- 1百萬
- 4百萬
- 1千5百萬
- 2千5百萬

4百萬

- 1億3千萬
- 4億



iSeq™ 100	MiSeq™, MiSeq™Dx*	NextSeq™ 550, NextSeq™ 550Dx*	NovaSeq™ 6000

illumina®

A large sequencer is like an A380

passengers are like DNA sequences









A smaller sequencer is like a Gulfstream

passengers are like DNA sequences









Specific market need: clinical samples









NGS data analysis



From raw signal to base-pairs

- Acquire reads FASTQ
- Generate bases with quality score
- De-multiplexing

Application specific

- Assembly (de novo)
- Annotation: Match to refs or database(16S)
- Alignment to reference to find variants (SNVs, CNVs, SVs)

Advanced, application specific

- Data visualization
- Variant filtering/annotation
- Diff. exp. analysis
- Pathway analysis
- Statistical/data mining

Automatic, on-instrument

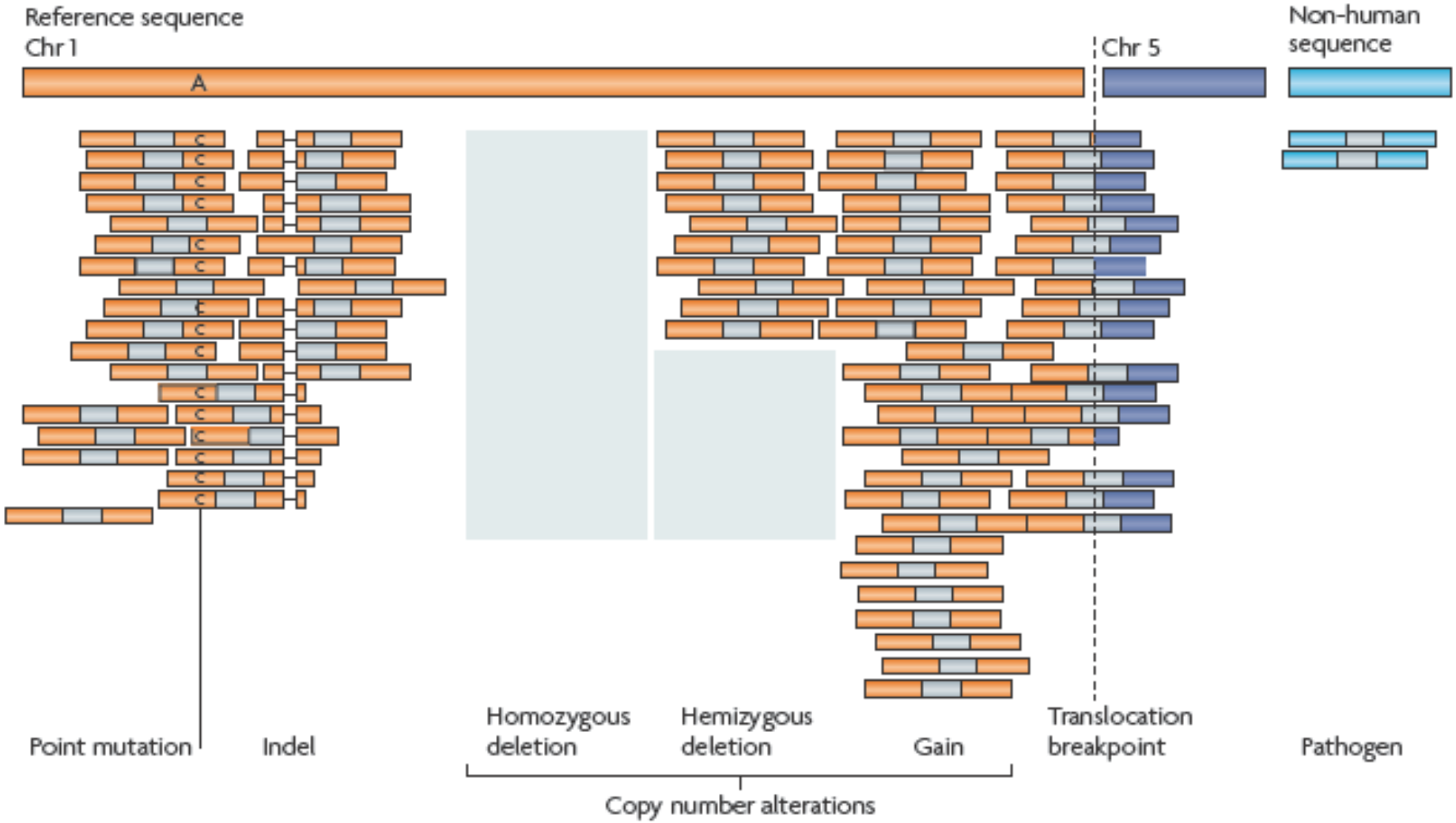
Automatic, manual or Cloud-based

Manual, Cloud-based

Standardized

In development

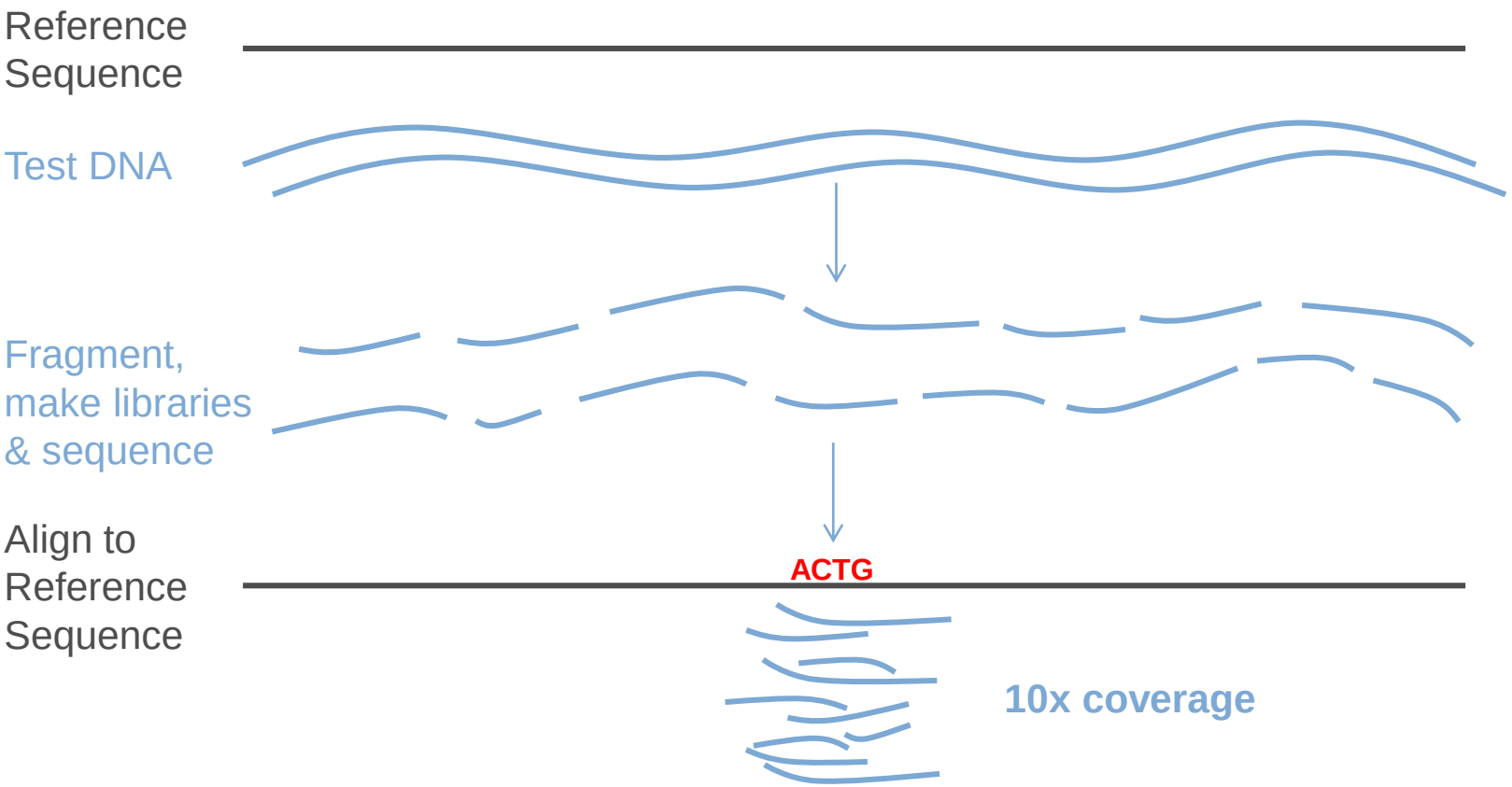
Re-sequencing: How to find alterations on the Human genome



Source: Meyerson M, Gabriel S, Getz G. Nat Rev Genet. 2010 Oct ;11(10):685-96. Review.

What is Sequencing Depth or Coverage?

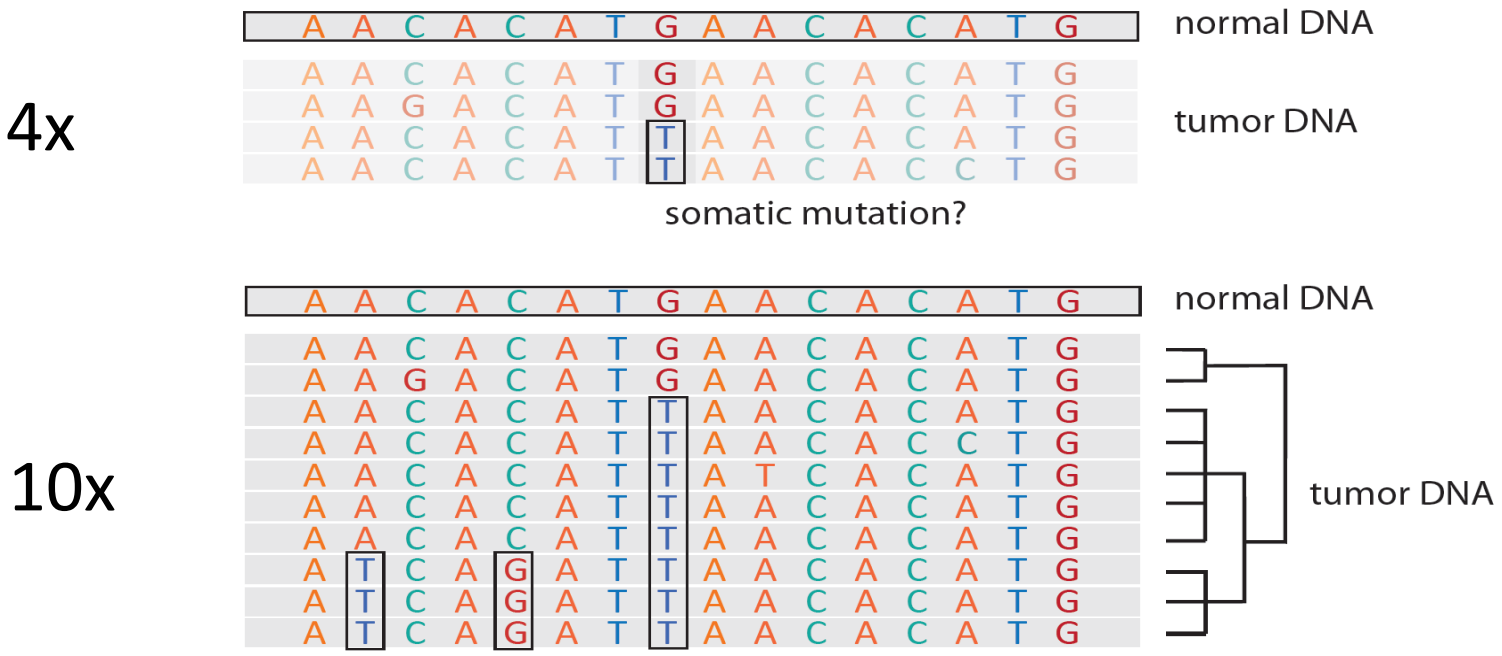
- The number of times the same region of DNA is sequenced independently
- If the same base is covered 10 times, you have 10x sequencing depth



illumina®

Why is sequencing depth important?

- Higher sequencing depth = higher accuracy, higher sensitivity, higher resolution for Mixture Samples
- Depth is important for RNA seq too, but in terms of reads



Illumina Informatics

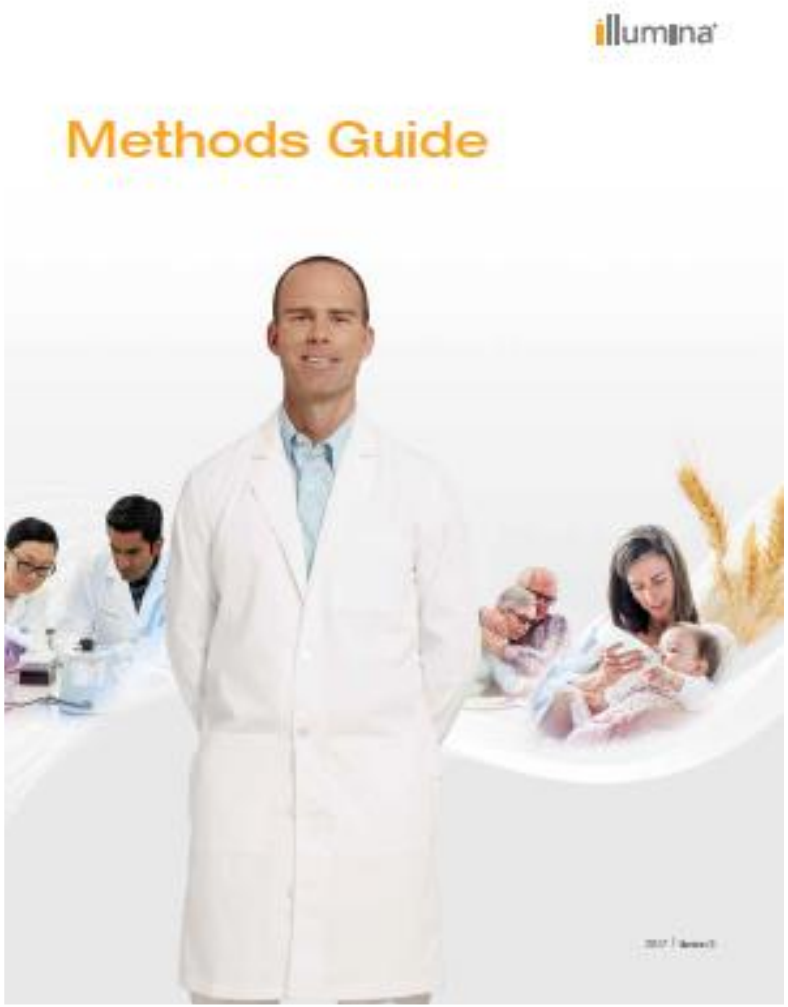
Key enabler of comprehensive workflow solutions

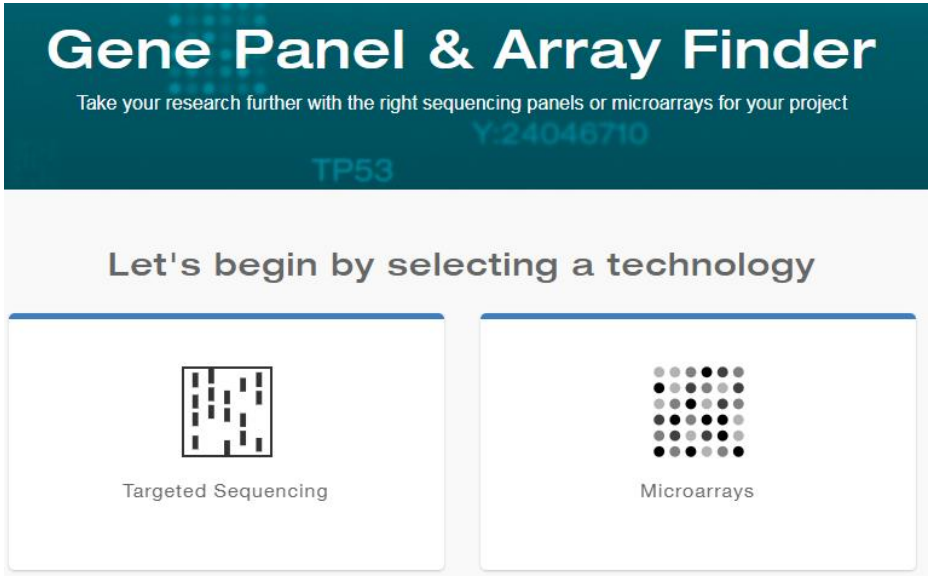


amazon web services™

ISO 27001 CERTIFIED by schellman

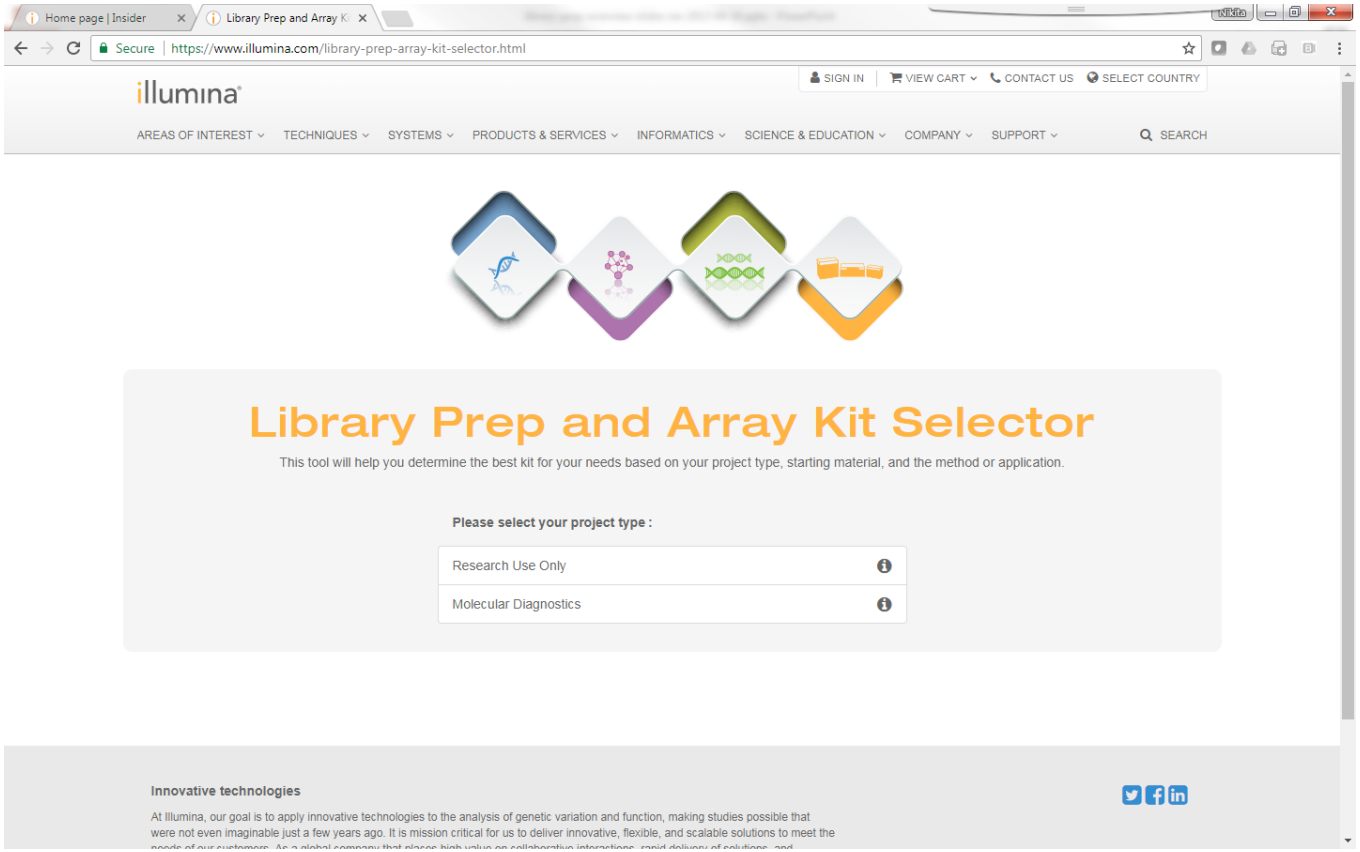
HIPAA





Find if your favorite gene is on a commercial panel; can input a list of genes and find the % covered by ILMN panels

Not sure which kit to use?
Try the Library Prep and Array Kit Selector



https://www.illumina.com/library-prep-array-kit-selector.html

Thank you!

tyang2@illumina.com

