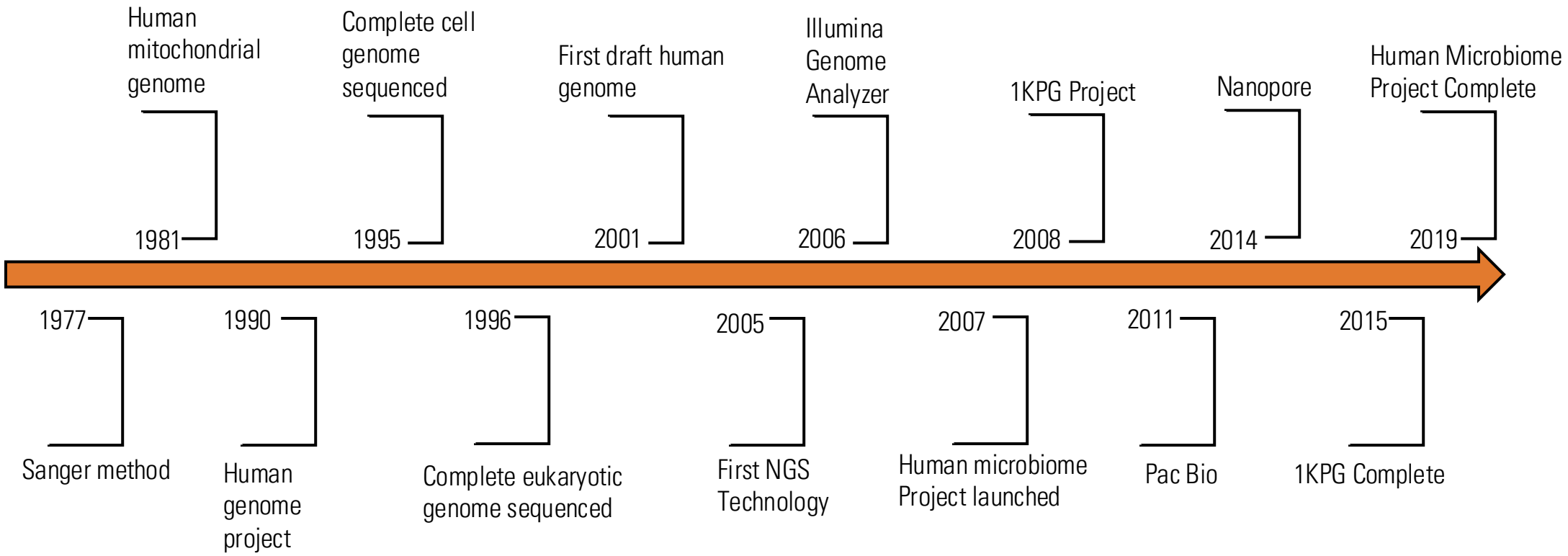
"KNOWLEDGE OF SEQUENCES COULD CONTRIBUTE MUCH TO OUR UNDERSTANDING OF LIVING MATTER"
– FREDRICK SANGER

NEXT GENERATION SEQUENCING

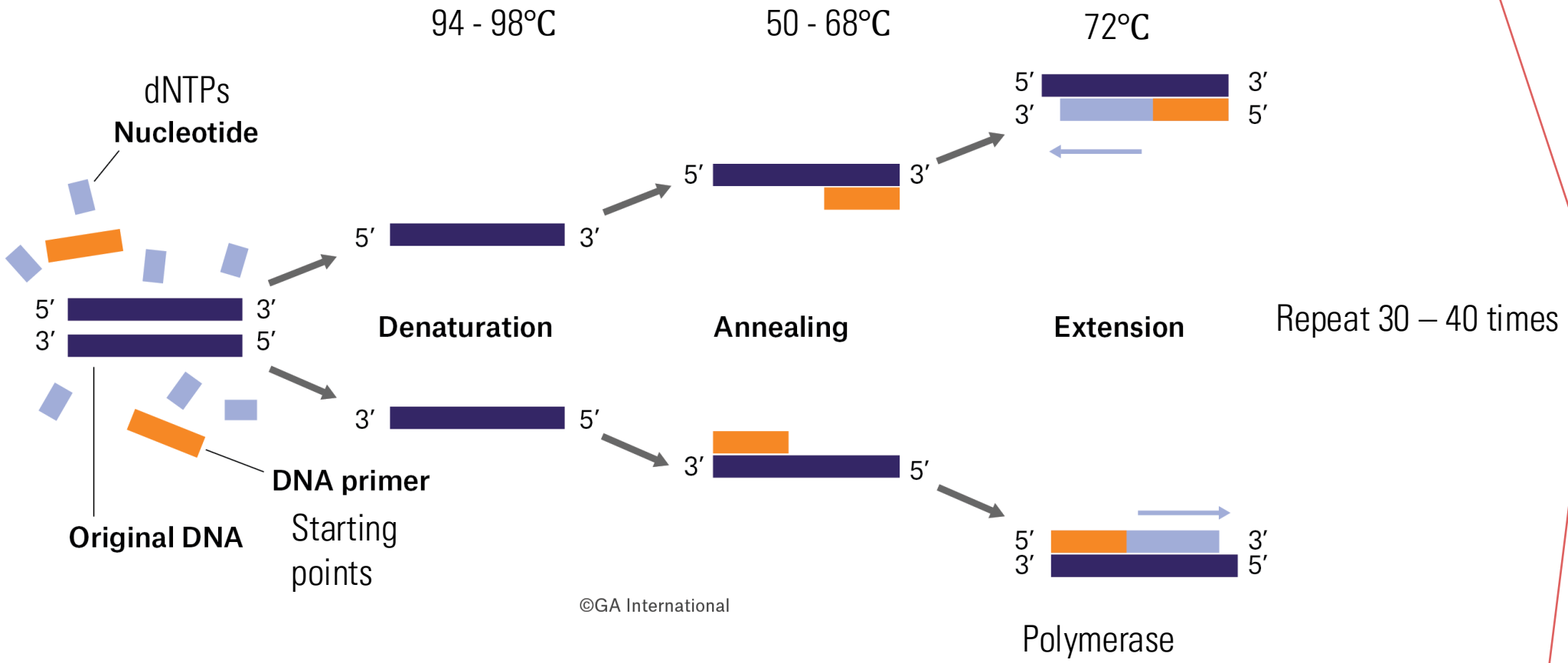
FROM DNA TO SEQUENCING



POLYMERASE CHAIN REACTION

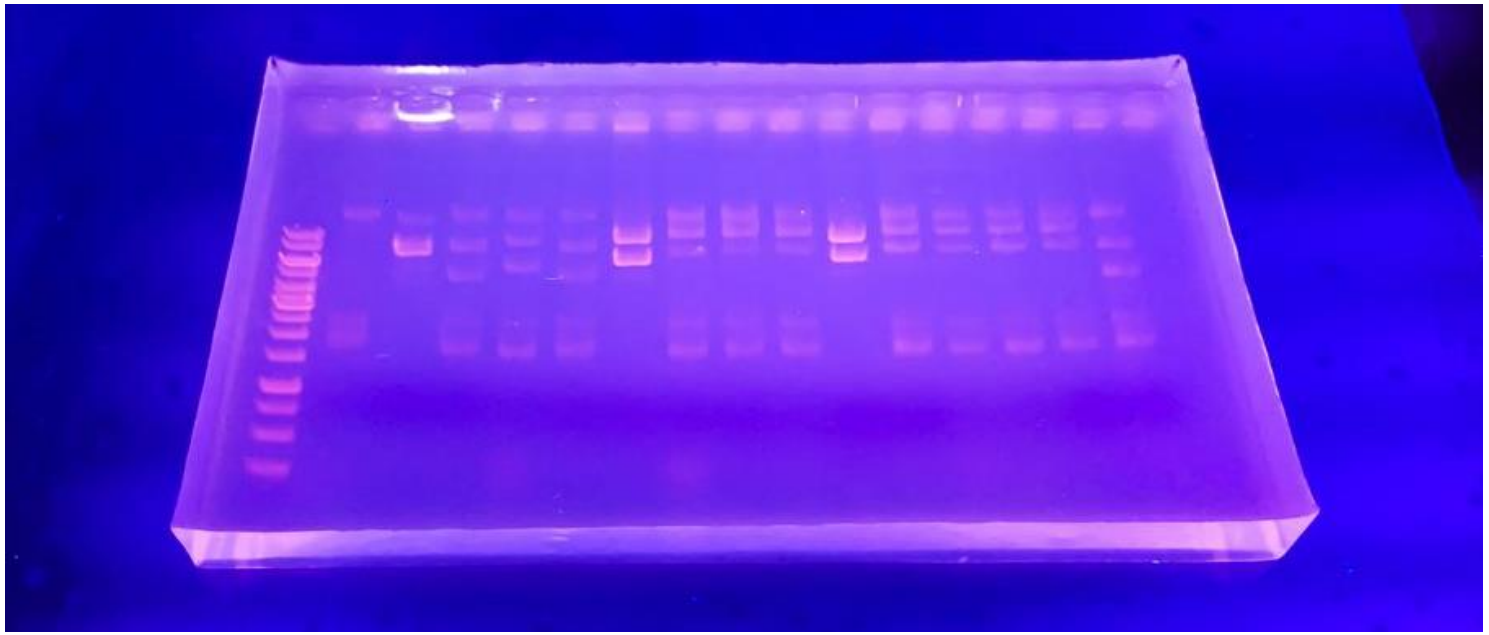
AMPLIFYING DNA SO WE HAVE ENOUGH!

POLYMERASE CHAIN REACTION



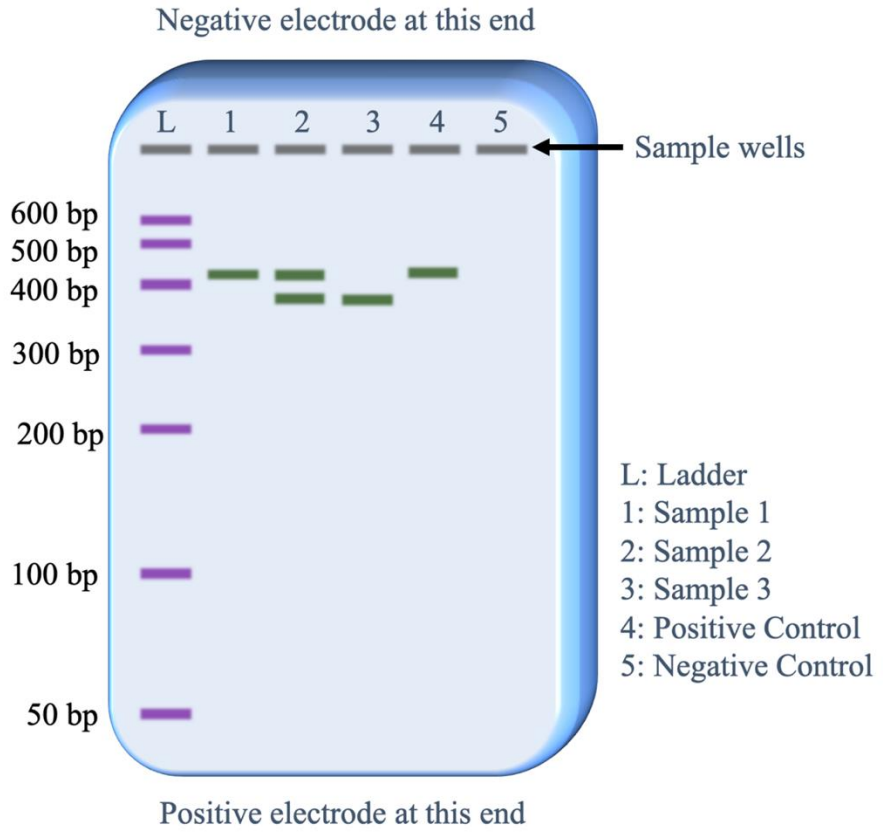
POLYMERASE CHAIN REACTION

POLYMERASE CHAIN REACTION



GEL ELECTROPHORESIS

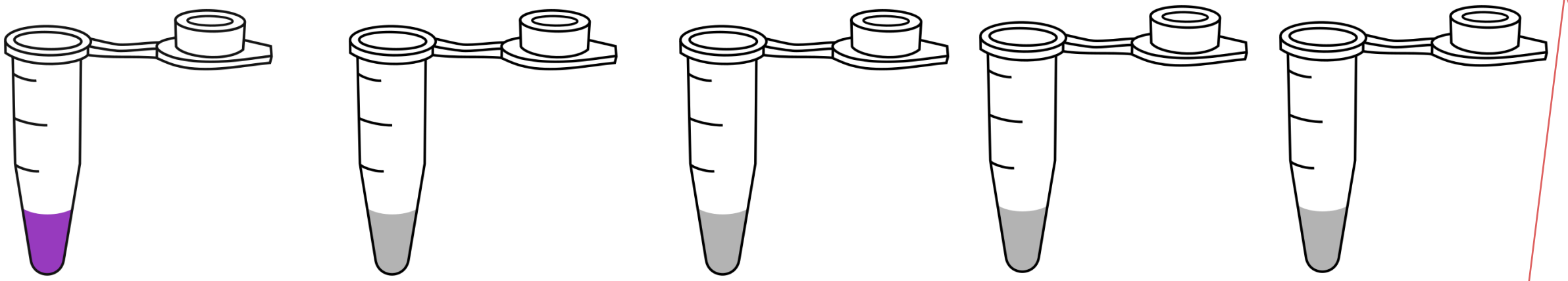
- Load samples into wells with dye in agarose gel
- Load ladder for a size metric
- Phosphate backbone of DNA is attracted to positive node
- Shorter strands will “run” faster along the gel
- Longer strands will “run” much slower



SANGER SEQUENCING METHODS

SANGER SEQUENCING

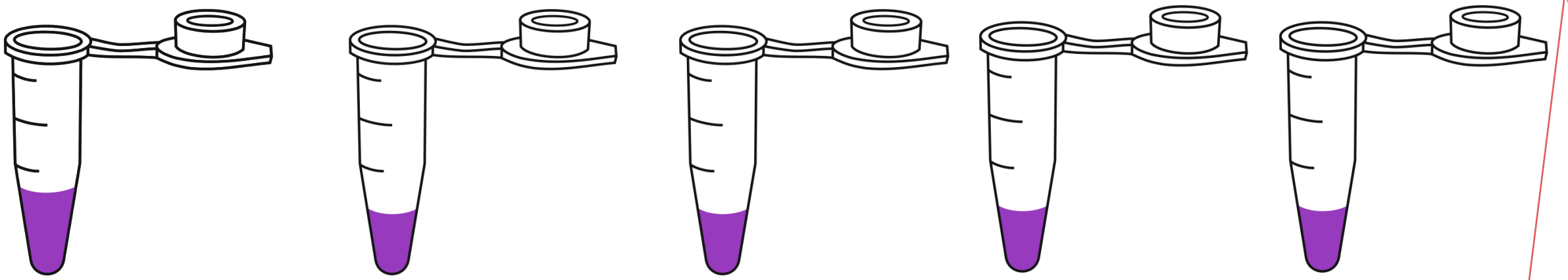
1. Primer



Samples 1 - 4

SANGER SEQUENCING

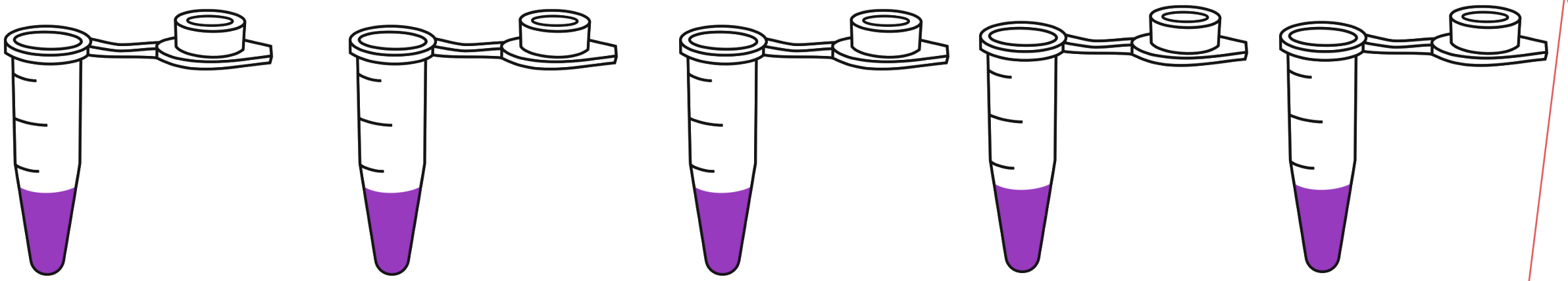
1. Primer
2. DNA Polymerase



Samples 1 - 4

SANGER SEQUENCING

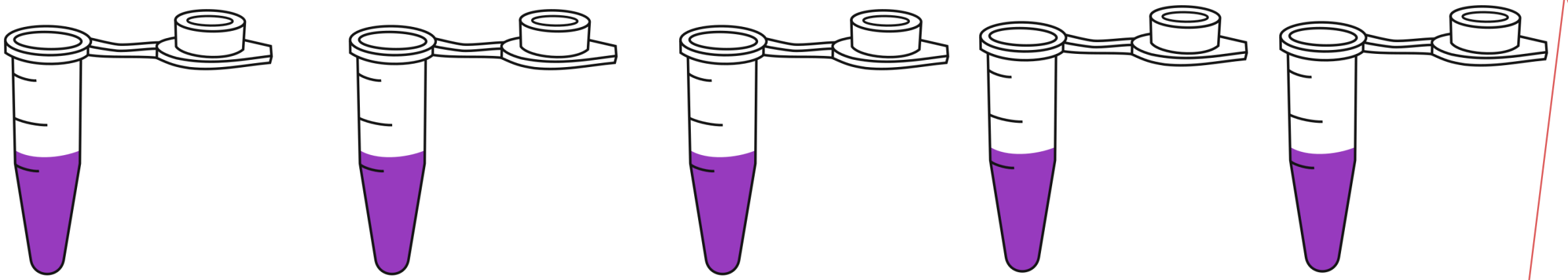
1. Primer
2. DNA Polymerase



Samples 1 - 4

SANGER SEQUENCING

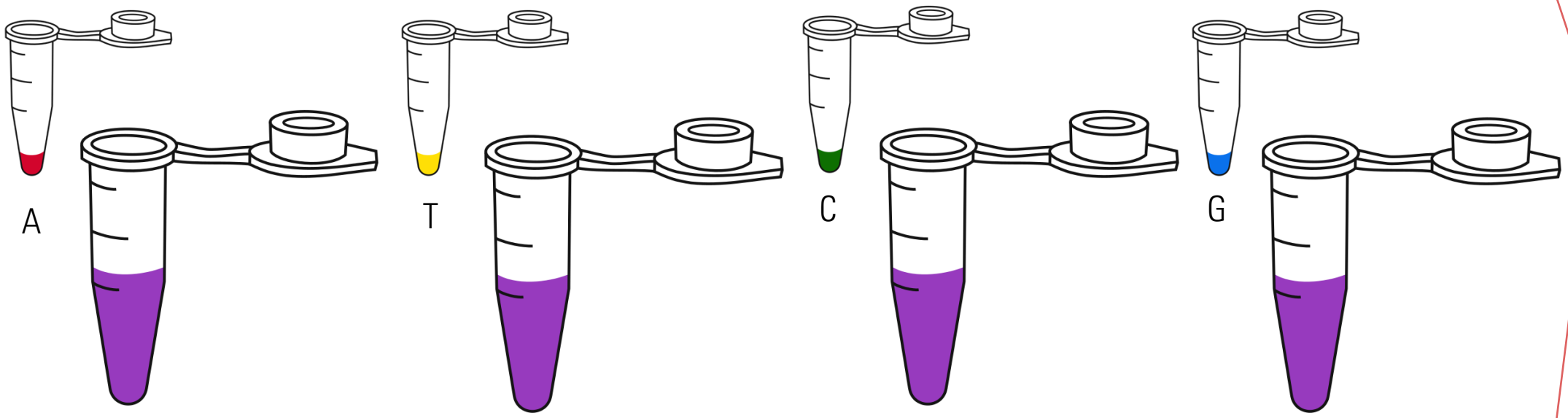
1. Primer
2. DNA Polymerase
3. dNTPS (A, C, G, T)



Samples 1 - 4

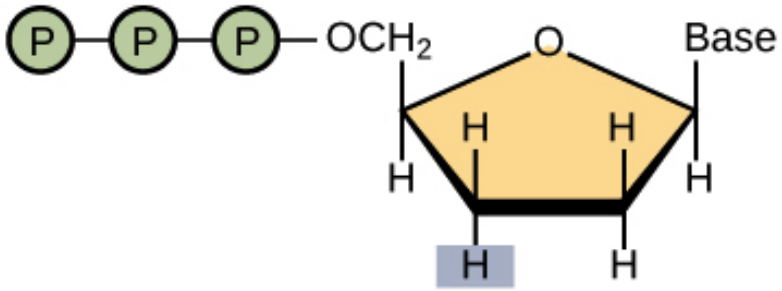
SANGER SEQUENCING

1. Primer
2. DNA Polymerase
3. dNTPS (A, C, G, T)
4. **ddNTPs**

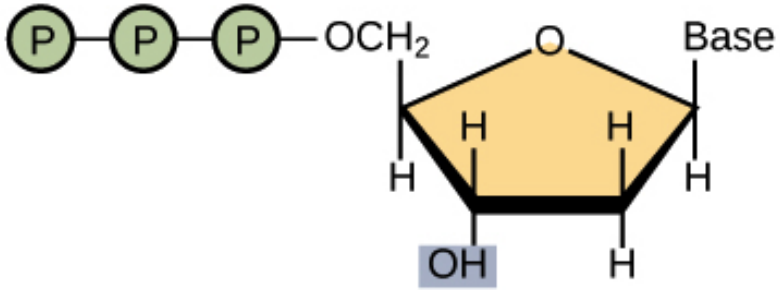


Samples 1 - 4

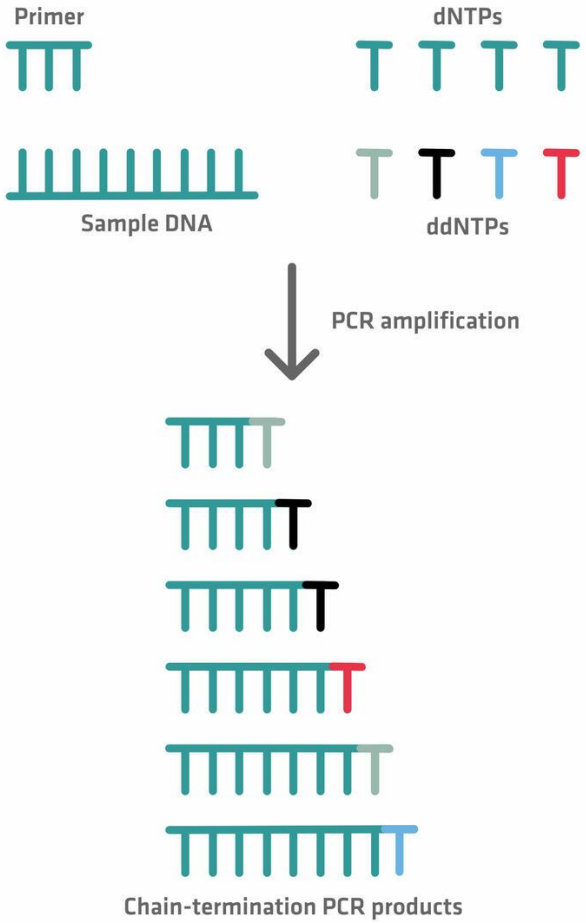
SANGER SEQUENCING



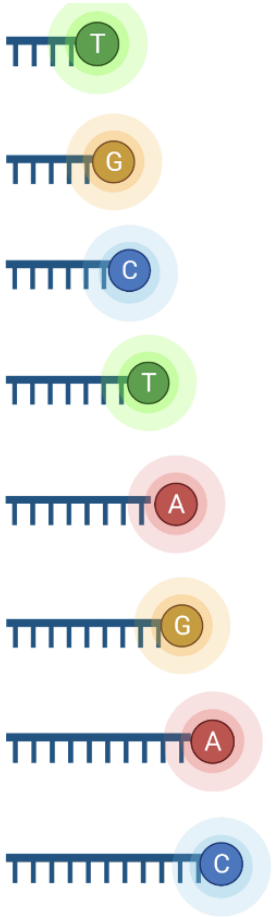
Dideoxynucleotide (ddNTP)

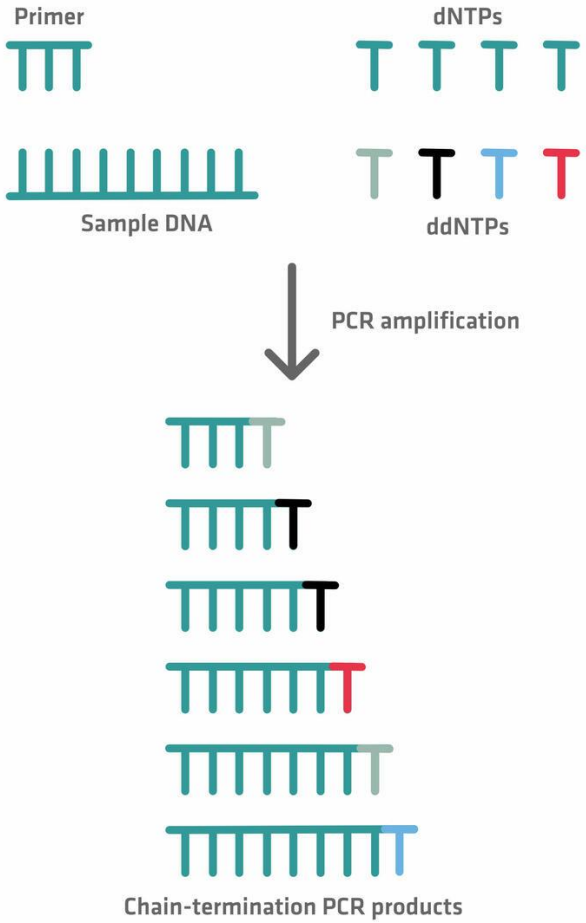


Deoxynucleotide (dNTP)

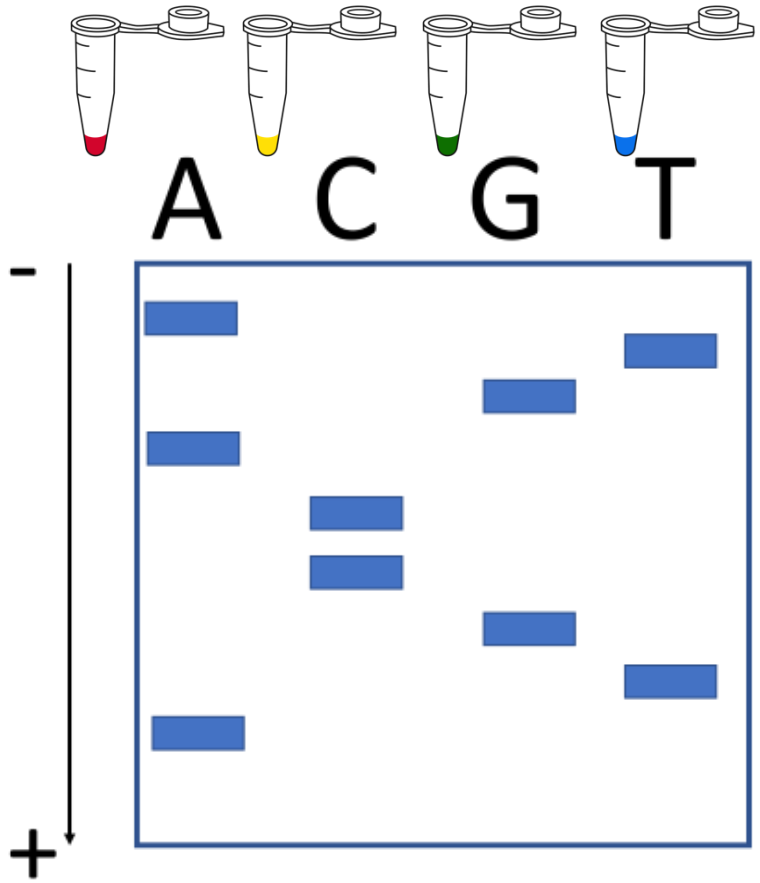


SANGER SEQUENCING

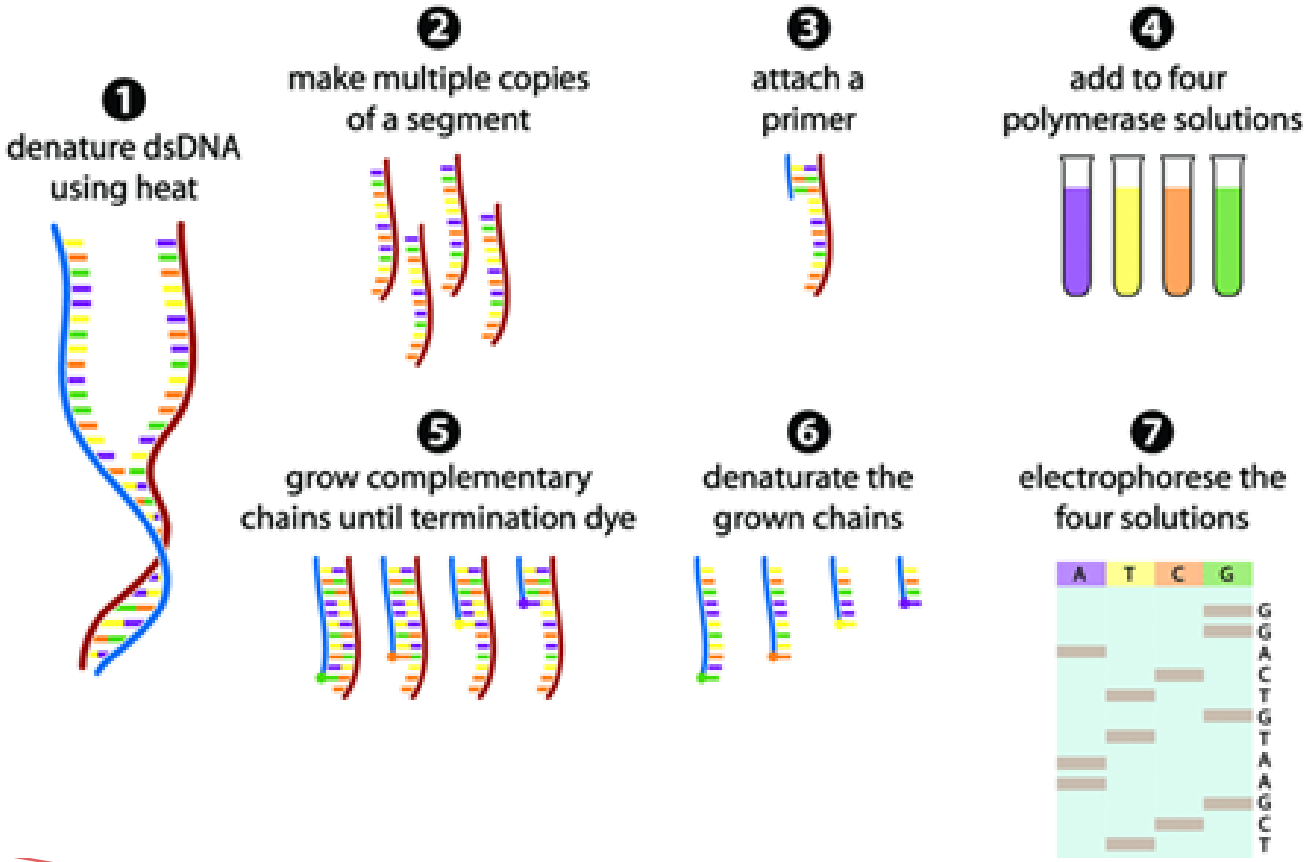




SANGER SEQUENCING



SANGER SEQUENCING

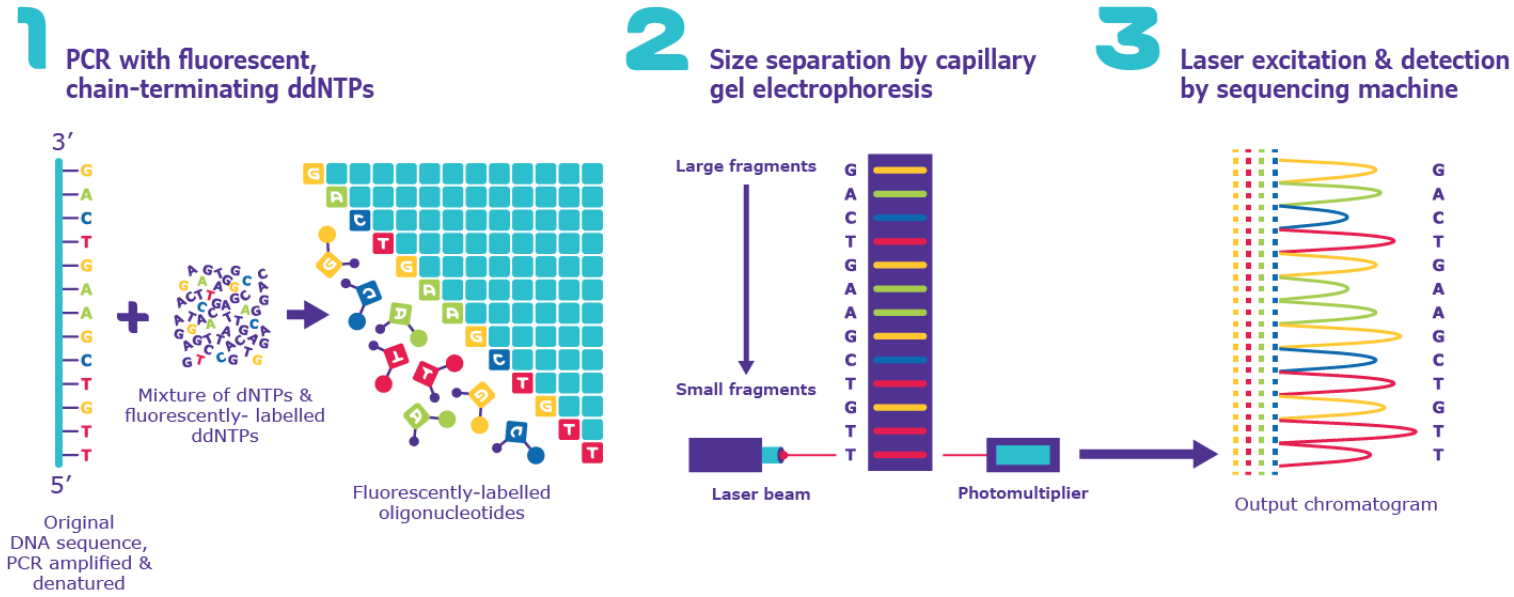


- Time consuming
- No automation

SANGER SEQUENCING - 1987

Leroy Hood and Michael Hunkapiller Improve Sanger Method

1. DNA labeled with fluorescent dyes instead of radioactive molecules
2. Data acquisition/analysis made easier with computers



SANGER SEQUENCING - 1987

Leroy Ho ger Method

1. DNA labe bactive molecules
2. Da computers

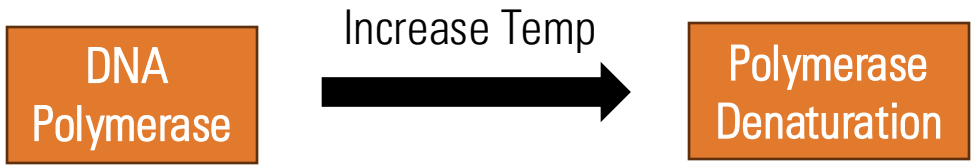
CYCLE SEQUENCING - 1989

Sanger Sequencing

1. DNA Polymerase
2. Much more DNA than Polymerase
3. A bunch of wasted DNA fragments

Cycle Sequencing

1. **Taq Polymerase** – Vincent Murray
2. Taq can withstand high temps
3. Can use our normal PCR process!



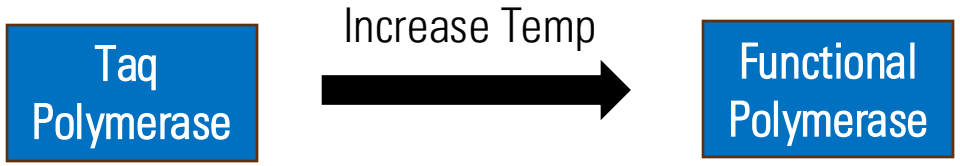
CYCLE SEQUENCING - 1989

Sanger Sequencing

1. DNA Polymerase
2. Much more DNA than Polymerase
3. A bunch of wasted DNA fragments

Cycle Sequencing

1. **Taq Polymerase** – Vincent Murray
2. Taq can withstand high temps
3. Can use our normal PCR process!



CYCLE SEQUENCING - 1989

Sanger Sequencing

1. DNA Polymerase
2. Much more DNA than Polymerase
3. A bunch of wasted DNA fragments

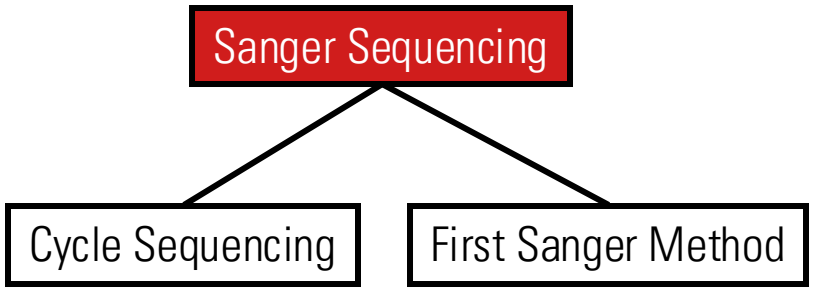
Cycle Sequencing

1. **Taq Polymerase** – Vincent Murray
2. Taq can withstand high temps
3. Can use our normal PCR process!

Two Major Improvements

1. Higher Fluorescent Signal
2. Less DNA needed per reaction

SEQUENCING METHODS

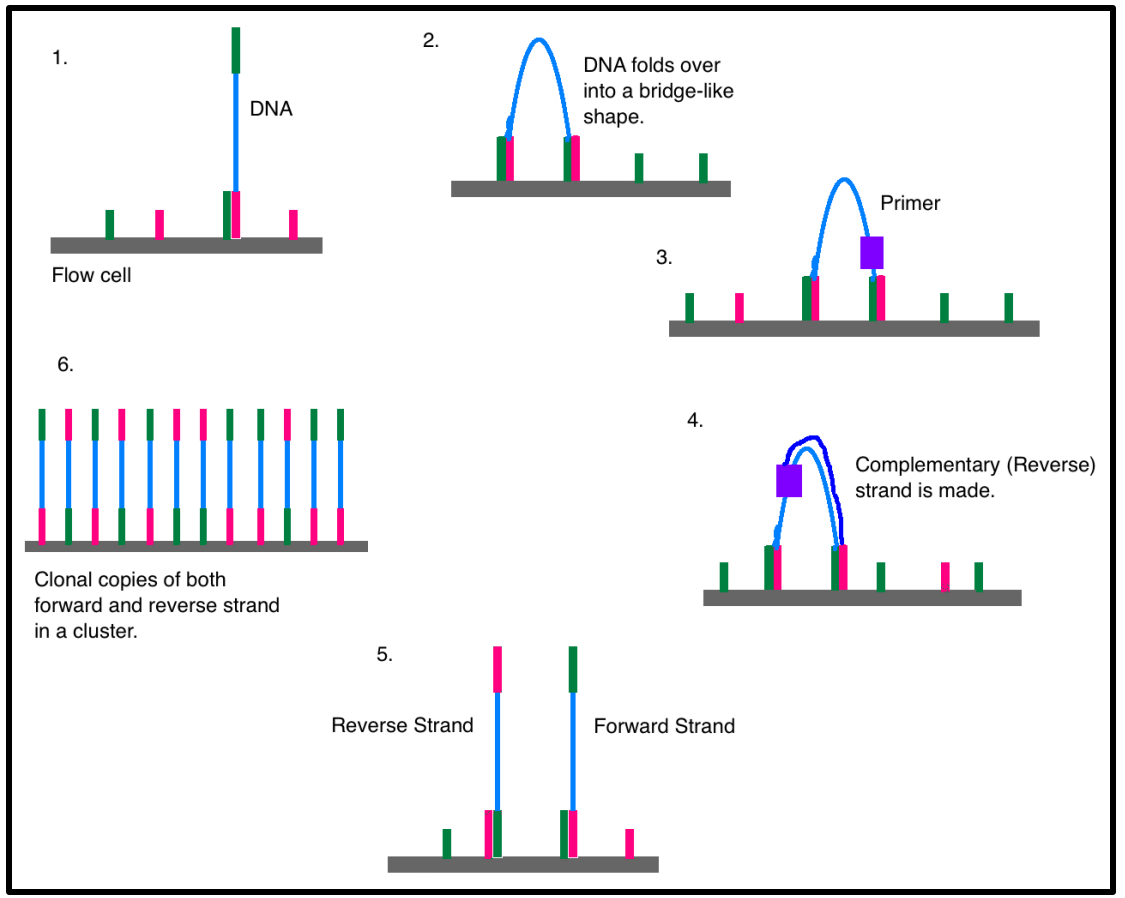


2ND GENERATION

ILLUMINA

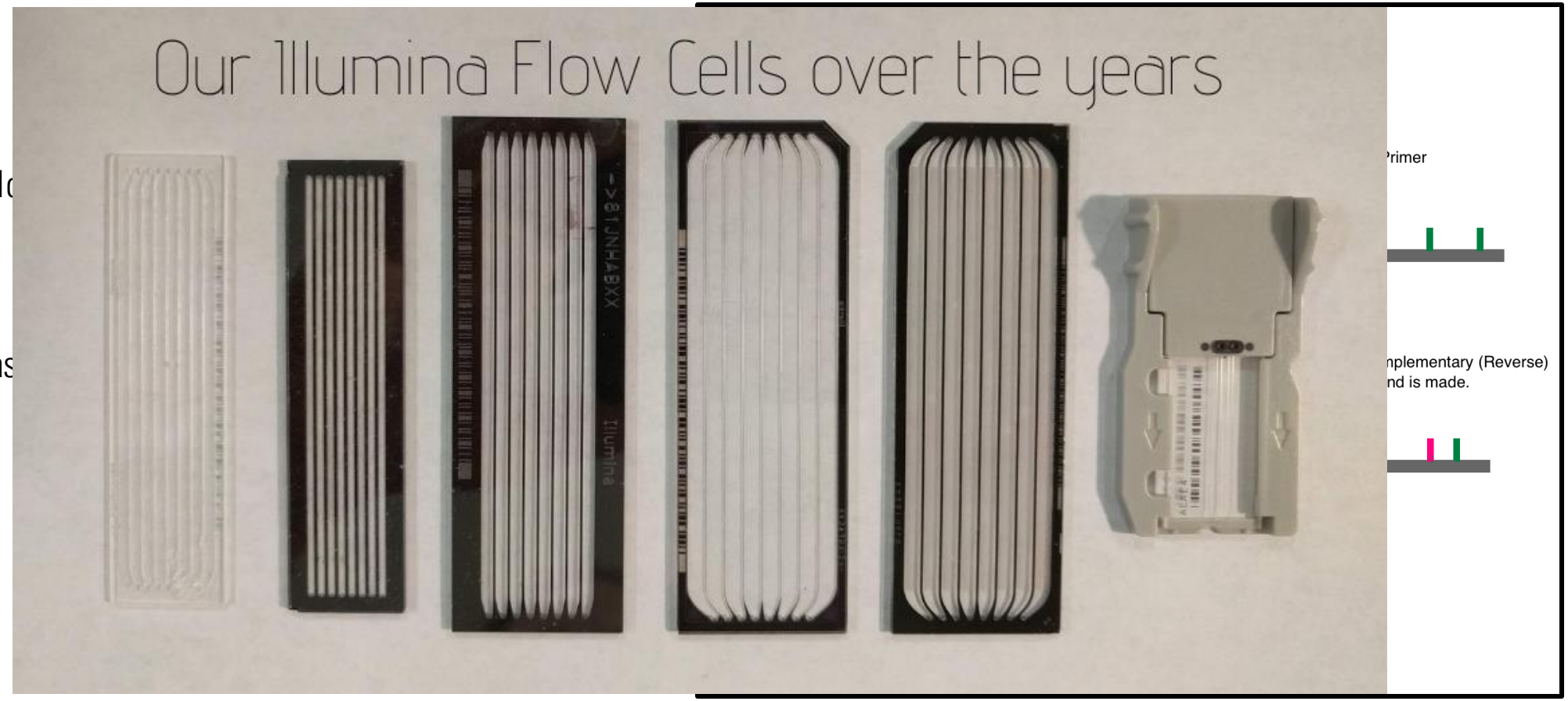
DYE SEQUENCING

- Follows core principles of Sanger Methods
  - Amplify > Sequence > Analyze
  - Sequence by Synthesis (SBS)
- Massive Parallel Sequencing



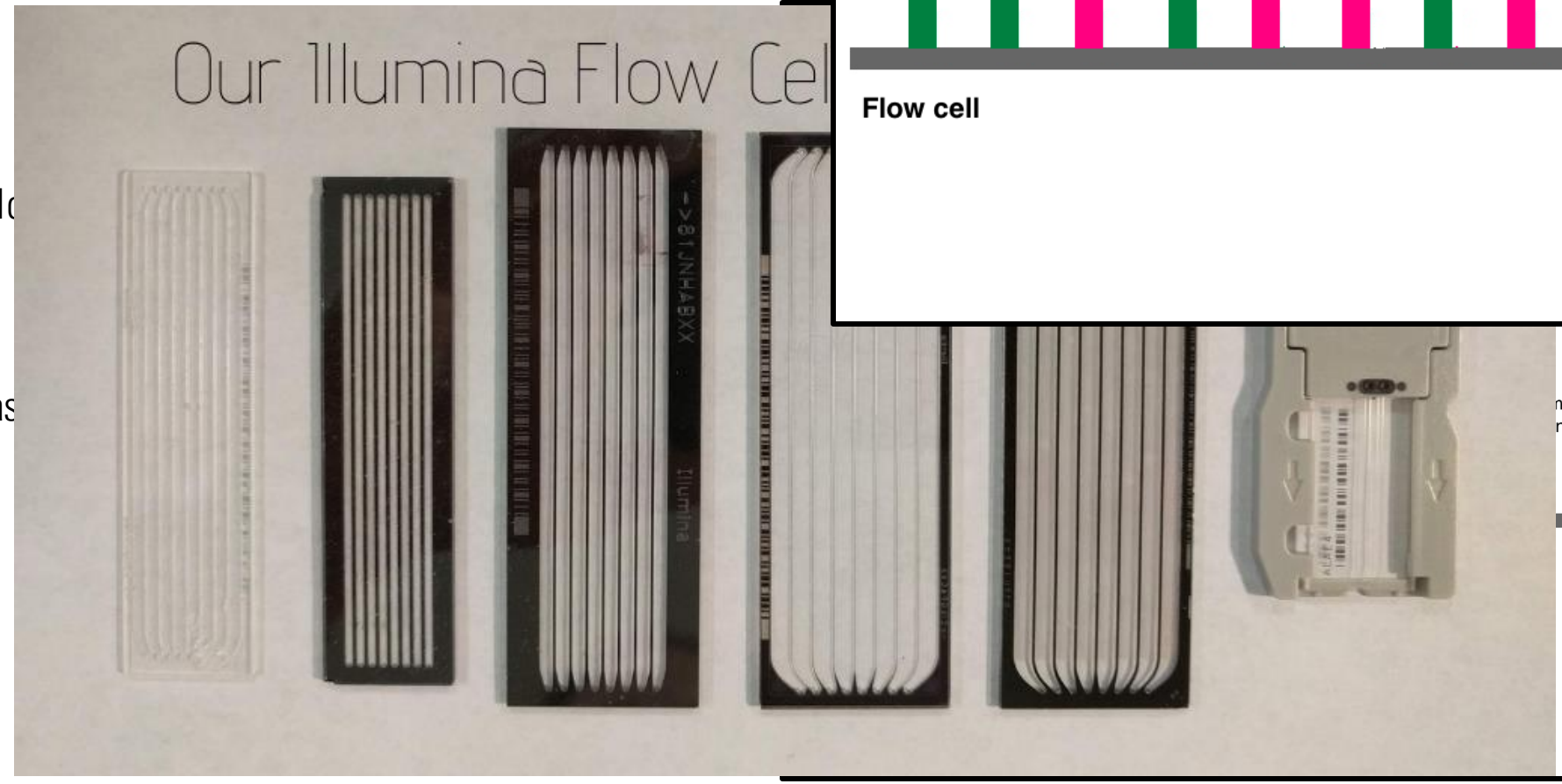
DYE SEQUENCING

- Follo
  -
  -
- Mas



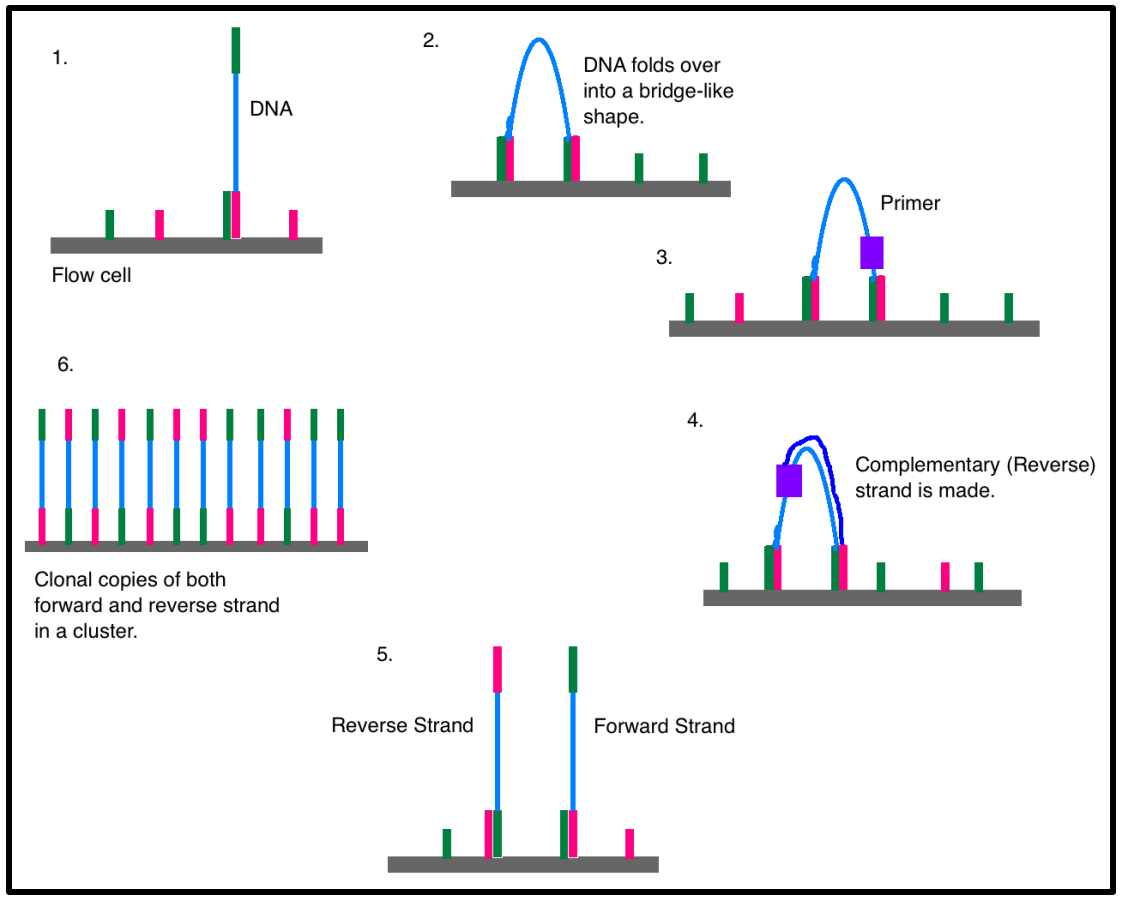
DYE SEQUENCING

- Follo
  -
  -
- Mas



DYE SEQUENCING

- Follows core principles of Sanger Methods
  - Amplify > Sequence > Analyze
  - Sequence by Synthesis (SBS)
- Massive Parallel Sequencing
- HUGE SPEED IMPROVENT



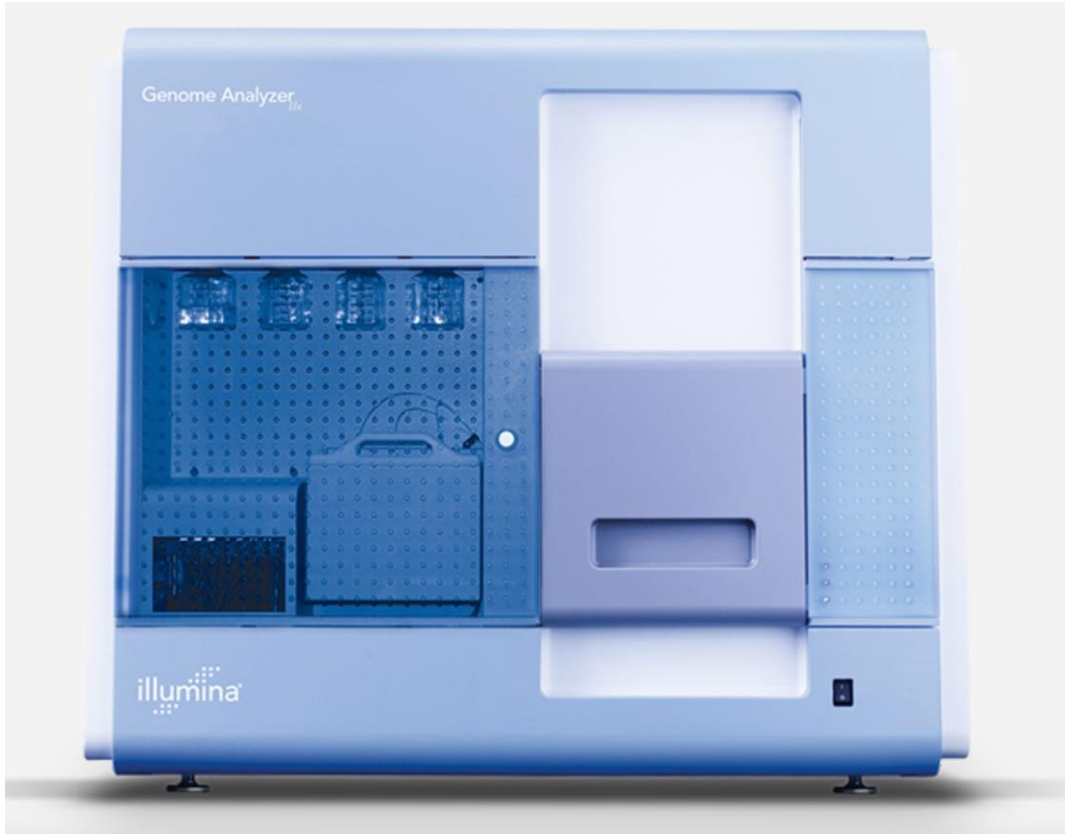
DYE SEQUENCING

Solexa

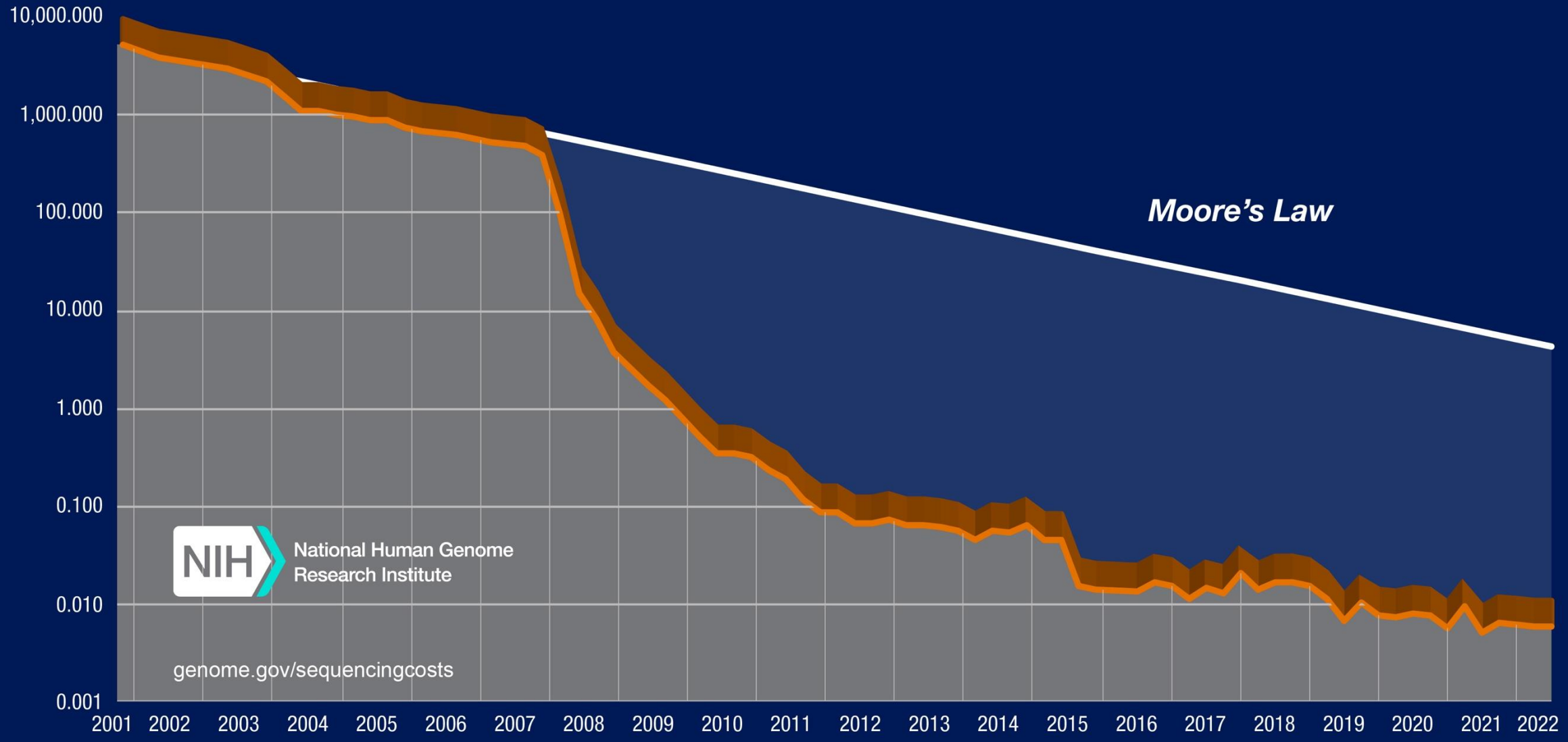
Solexa 1G Genome Analyzer

November 2006

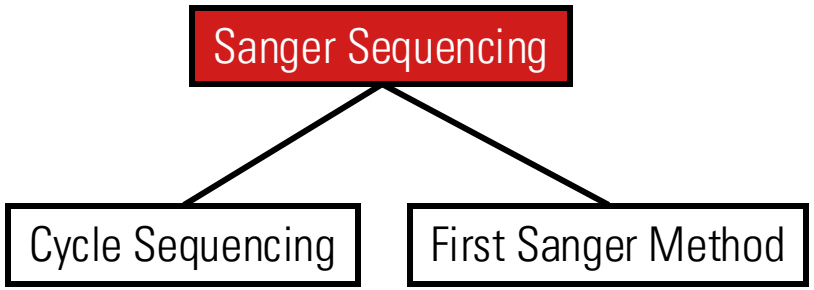
4

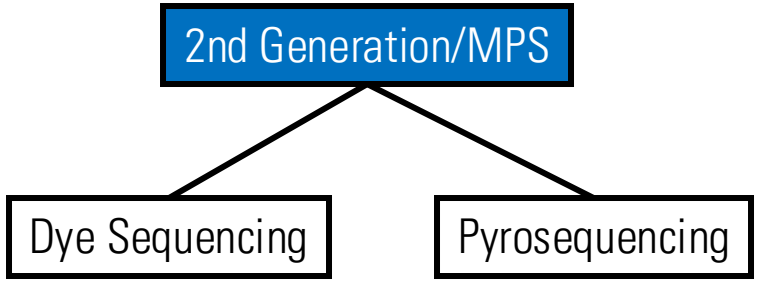


Cost per Raw Megabase of DNA Sequence



SEQUENCING METHODS





3RD GENERATION AND ONWARDS

PACBIO AND OXFORD NANOPORE

WHAT MAKES 3RD GEN 3RD GEN?

- 1st and 2nd generation sequencing methods typically use the same framework

Amplify Sequence Analyze

- 2nd generation sequencing breaks DNA into small segments (50 – 400 bp) for sequencing
- Smaller chunks are easier to manage and result in fewer errors but take longer to sequence and analyze

WHAT MAKES 3RD GEN 3RD GEN?

- 3rd generation makes two improvements over general 2nd generation methods
  1. Sequence without breaking DNA into smaller parts (short-reads) and sequence longer length of DNA (long reads, **1,500 – 100k bp**)
  2. Skipping the amplification step in typical 1st and 2nd gen sequencing frameworks

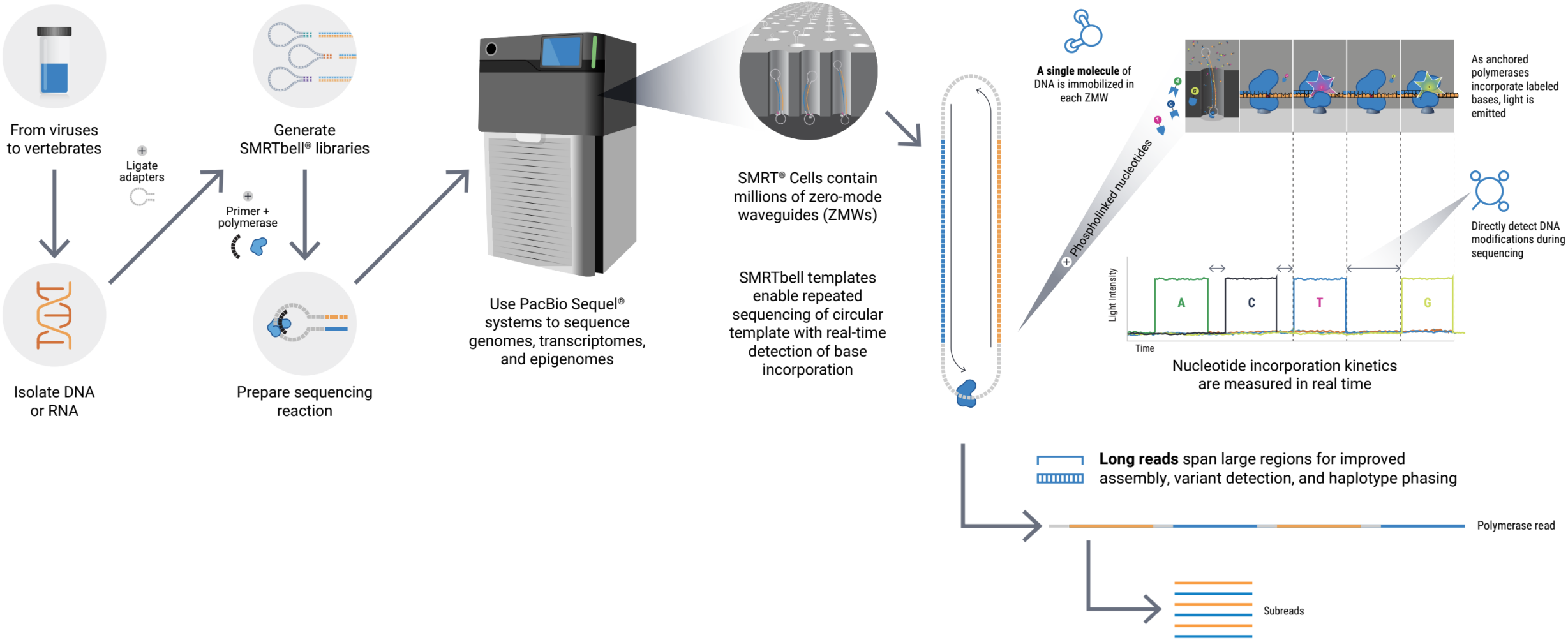
- 3rd generation is **not strictly better** than 2nd generation but is just **different**
  - Long reads can tell us different information but is overall **less accurate** than 2nd generation

PacBio

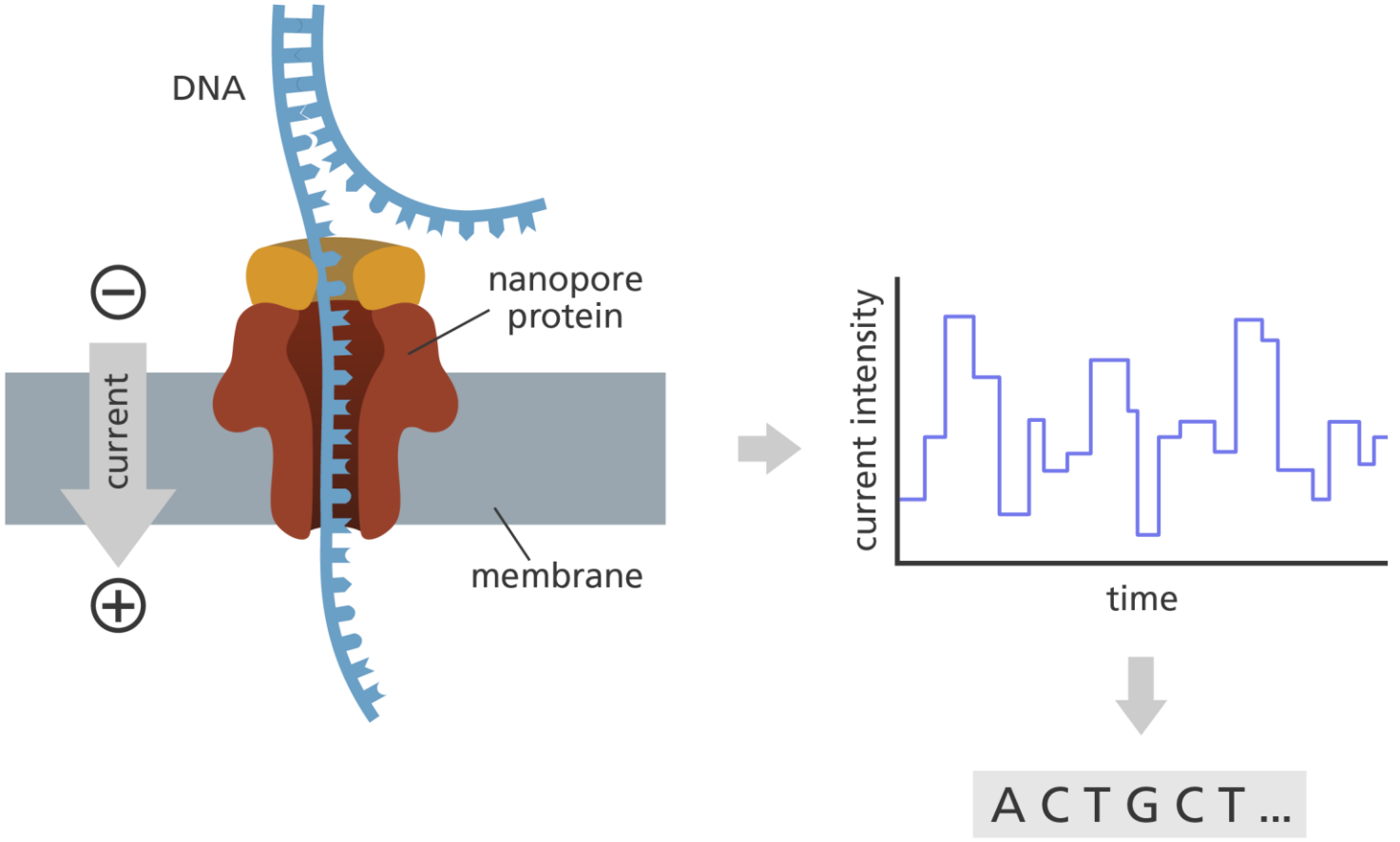
Oxford NANOPORE Technologies

PACBIO SMRT SEQUENCING - 2011

Single Molecule, Real-Time (SMRT) Sequencing



NANOPORE


DNA
nanopore protein
current
membrane
current intensity
time
A C T G C T ...

NANOPORE MINION

$1,999

LONG READS VS SHORT READS

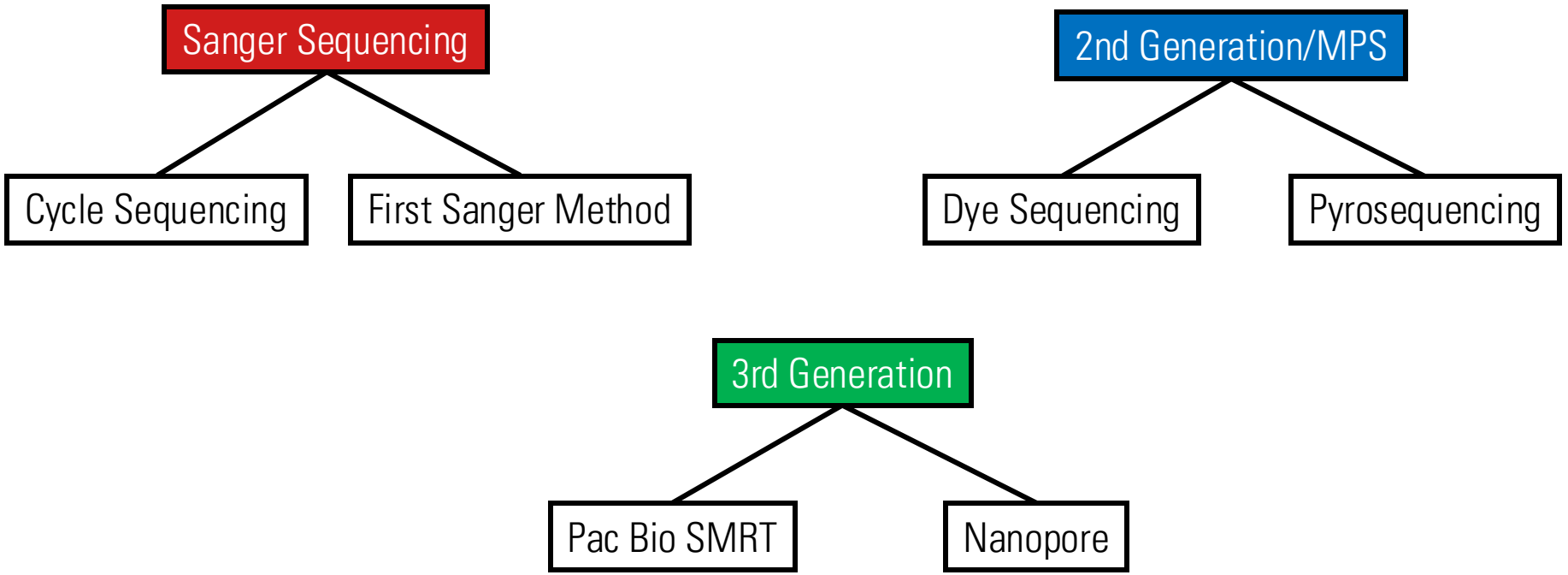
SHORT READS

- Shorter read length can't capture complex genomic structures
- Struggles to reconstruct entire genomes
- Very accurate
- Low DNA input requirement
- Inexpensive per base but can scale fast

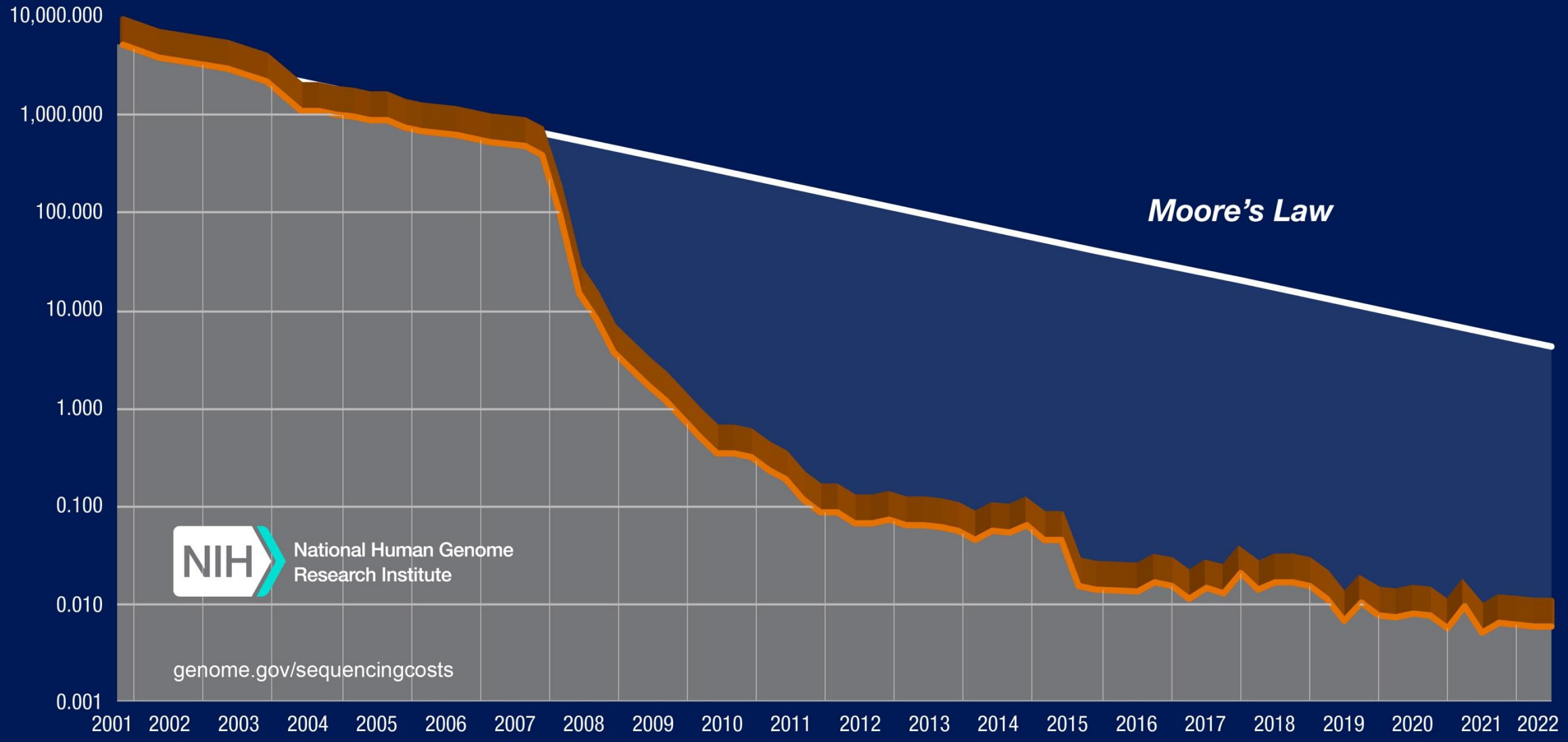
LONG READS

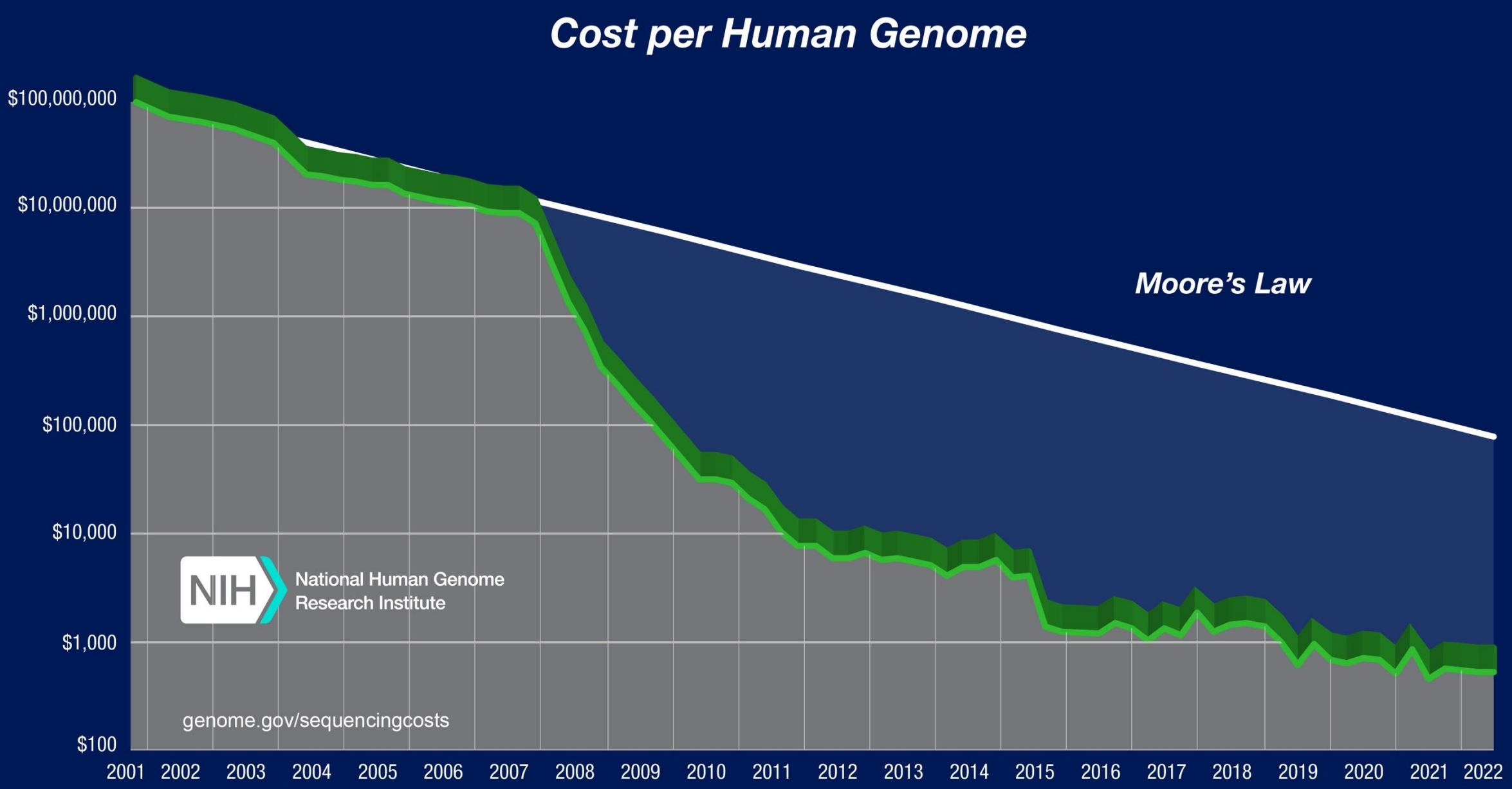
- Long reads capture complex genomic structures
- Facilitates complete genome assembly
- Less accurate but improving
- Need more DNA input
- Higher operating cost

SEQUENCING METHODS



Cost per Raw Megabase of DNA Sequence




Cost per Human Genome
$100,000,000
$10,000,000
$1,000,000
$100,000
$10,000
$1,000
$100
Moore's Law
NIH
National Human Genome
Research Institute
genome.gov/sequencingcosts
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022