STARS Mini-Symposium 9/12/2016

DNA Sequencing: The Past, the Present and the Future

Ralf Kittler, Ph.D.
McDermott Center for Human Growth and Development

ralf.kittler@utsouthwestern.edu

Outline

DNA sequencing is a biochemical method to determine the sequence of the nucleotide bases that make up the DNA

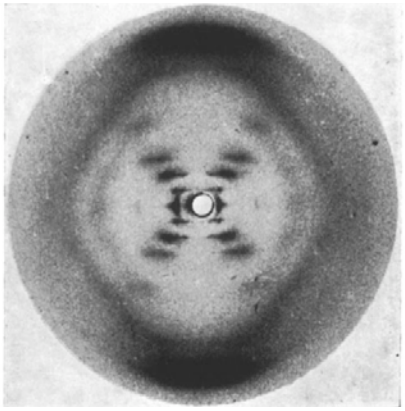
1. **The Past – Sanger Sequencing**
2. **The Present – Illumina Sequencing**
3. **The Future – Single Molecule Sequencing**

Part 1

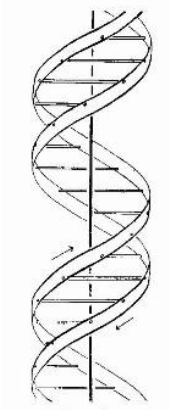
The past
Sanger Sequencing
(First-Generation Sequencing)

The Discovery of the DNA Structure

1953 WATSON and CRICK describe the structure of DNA based on the X-ray analyses of FRANKLIN and WILKINS



"Photograph 51"
(Franklin R and Gosling R, 1953)

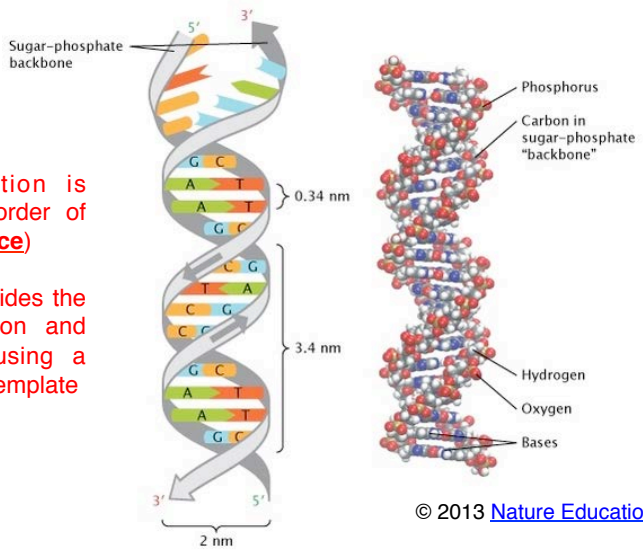


Watson-Crick Model
(Watson J and Crick F, 1953)

DNA Structure

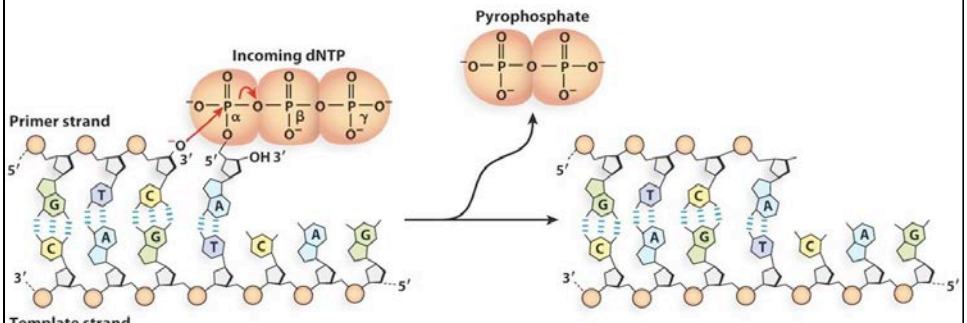
Genetic information is contained in the order of the bases (**Sequence**)

DNA structure provides the basis for replication and transcription by using a single strand as a template (**Base pairing**)



DNA Polymerase

1957 KORNBERG discovers DNA polymerase as enzyme for DNA replication

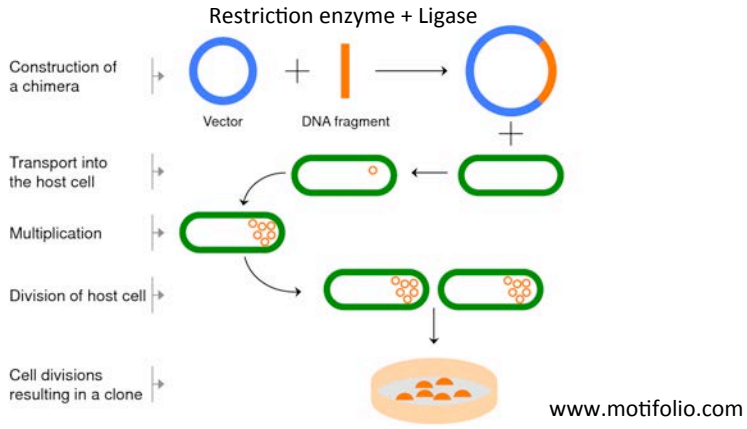


(Molecular Biology: Principles and Practice, 2012)

Complementary nucleotide is incorporated by forming a phosphoester bond with the 3'OH of the deoxyribose of the preceding nucleotide (strand extension)

Molecular Cloning

1970 BERG, BOYER, and COHEN develop Molecular Cloning



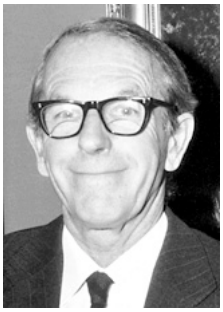
Molecular cloning allows isolation, amplification and manipulation of specific DNA fragments

First-Generation Sequencing

1977 SANGER develops chain-termination sequencing (Sanger sequencing)

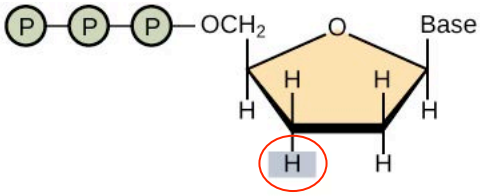
GILBERT and MAXAM develop sequencing by base-specific chemical fragmentation


Walter Gilbert

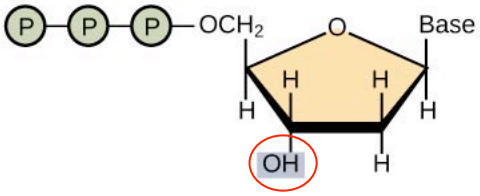

Frederick Sanger

https://www.nobelprize.org/nobel_prizes/chemistry/laureates/1980/

Sanger Sequencing: Dideoxynucleotides (ddNTPs)

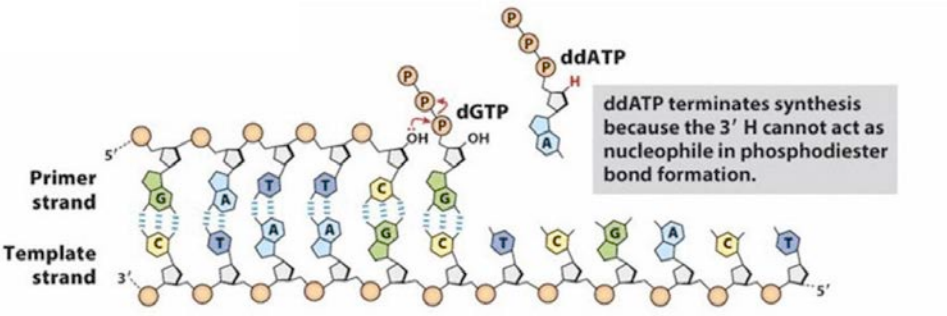


Dideoxynucleotide (ddNTP)



Deoxynucleotide (dNTP)

Chain Termination by ddNTP Incorporation



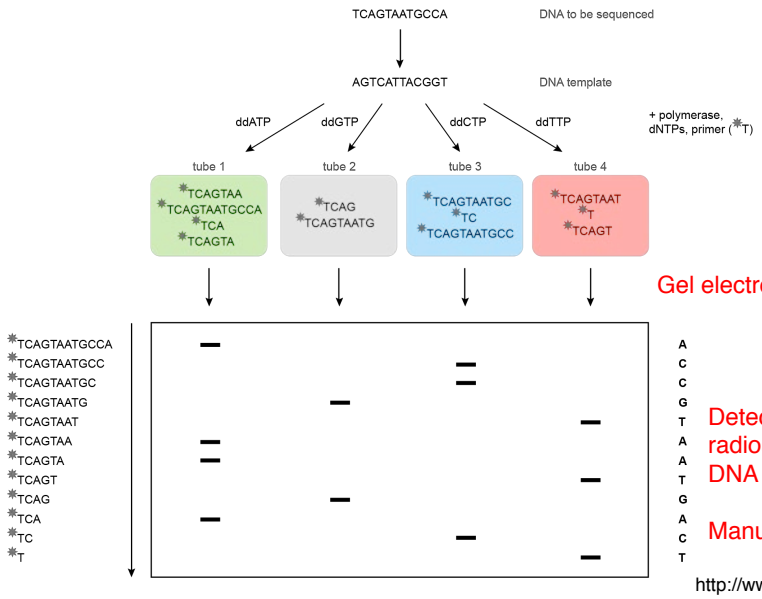
(Molecular Biology: Principles and Practice, 2012)

For Sanger sequencing ratio of dNTP:ddNTP ≈ 100:1

➔Mixture of terminated strands is produced

➔The identity of the last incorporated base can be identified by gel electrophoresis and radioactive, or fluorescent detection

Sanger Sequencing 1.0 (1977-1980s)

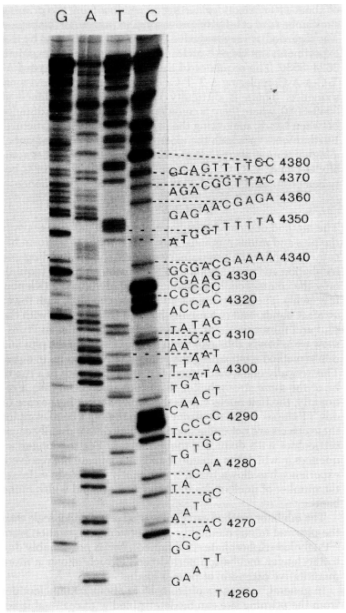


Detection of radioactive DNA strands

Manual process

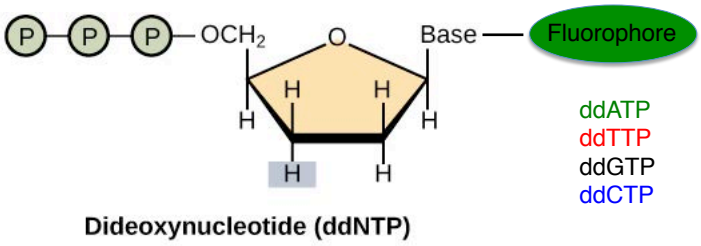
http://www.atdbio.com

Sanger Sequencing 1.0 (1977-1980s)



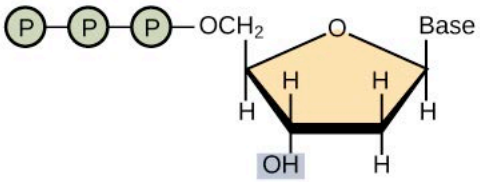
(Sanger et al., 1977, *PNAS*)

Fluorescent ddNTPs



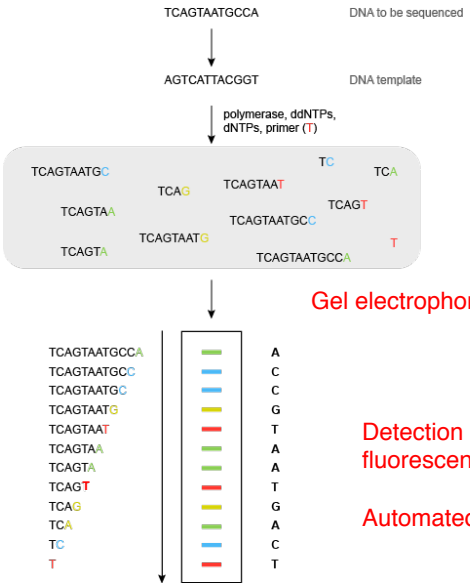
ddATP
ddTTP
ddGTP
ddCTP

Dideoxynucleotide (ddNTP)



Deoxynucleotide (dNTP)

Sanger Sequencing 2.0 (1990s-present)



Fluorescently labeled ddNTPs

Gel electrophoresis

Detection of fluorescent products

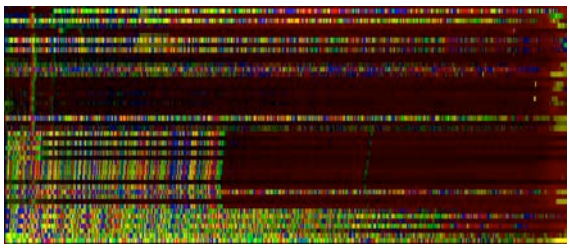
Automated process

http://www.atdbio.com

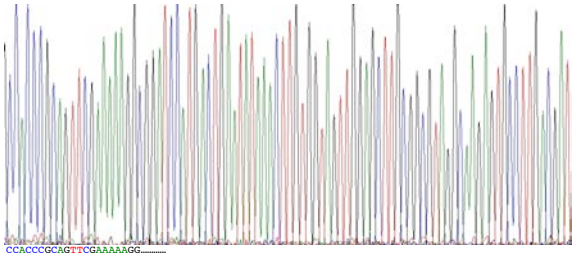
Sanger Sequencing at UT Southwestern

ABI 3730xl DNA Analyzer (Capillary Sequencer)
96 DNA samples with ~700 nucleotide reads (~ 70,000 bases) in 2.5 hours

Sanger Sequencing at UT Southwestern



Detected fluorescent DNA products



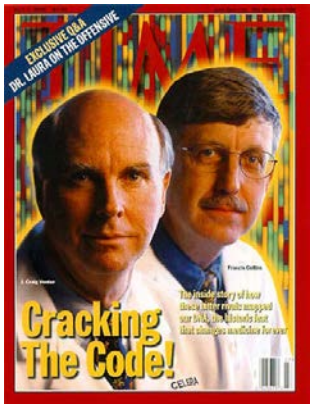
A single trace (electropherogram)

Sequence

Sanger Sequencing Enabled Genome Sequencing

1995 First bacterial genome (*Haemophilus influenzae*) sequenced

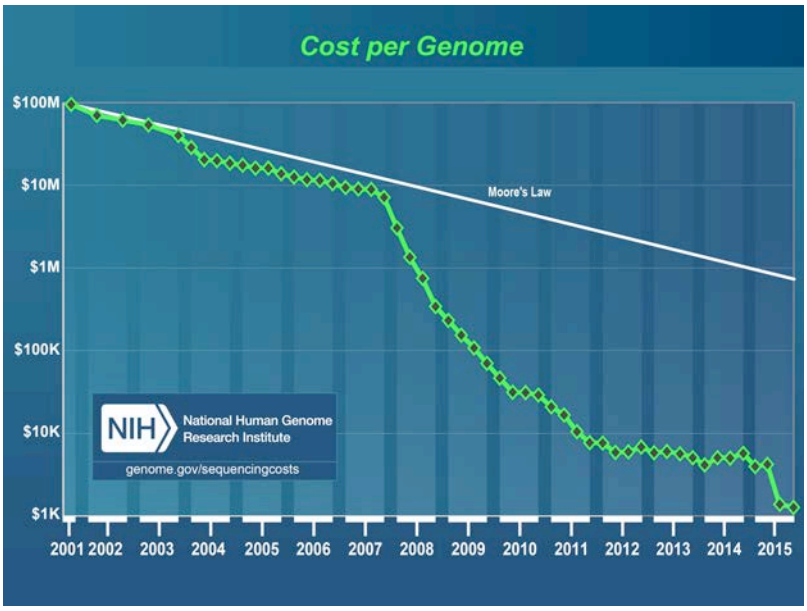
2001 First draft of the human genome published



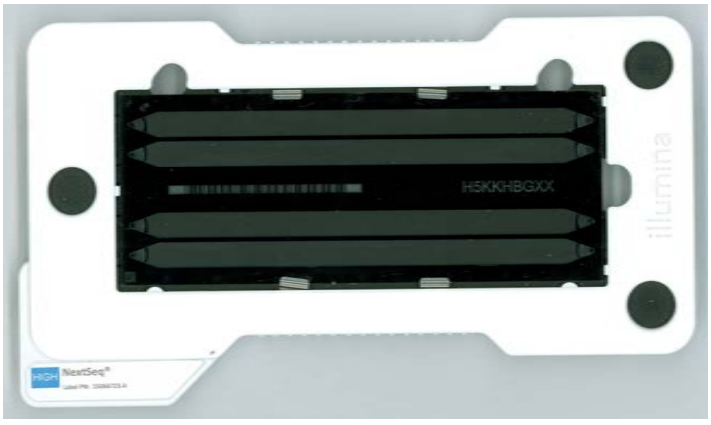
Part 2

The Present
Illumina Sequencing
(Next-Generation Sequencing)

Impact of Next-Generation Sequencing



Illumina NextSeq Flow Cell



400,000,000 DNA fragments X 300 nucleotides
=
120,000,000,000 nucleotides in 29 hours

Illumina Sequencing Benchmarks

NextSeq 500

vs.

The Human Genome Project

1 Technician
29 hours
$4000
120,000,000,000 nucleotides
Human genome X 40

>1000 Scientists
13 years (1990-2003)
$3,000,000,000
24,000,000,000 nucleotides
Human genome X 8

Illumina Sequencing is Scalable

HiSeq X Ten: Single device has 15X capacity of the NextSeq 500

Cost for sequencing of a human genome (at 40X): ~$1000

Next-Generation Sequencing

Technological advances that enabled massively parallel sequencing:

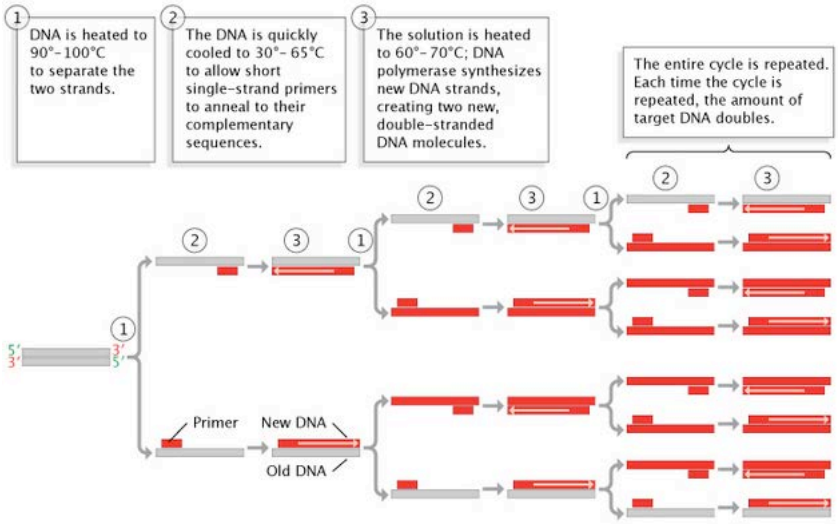
1. Miniaturization of clonal amplification of millions of individual DNA fragments on a solid support

   Illumina sequencing: Bridge Polymerase Chain Reaction (PCR)

2. Sequencing reaction chemistry that can be performed on clonally amplified DNA fragments on a solid support

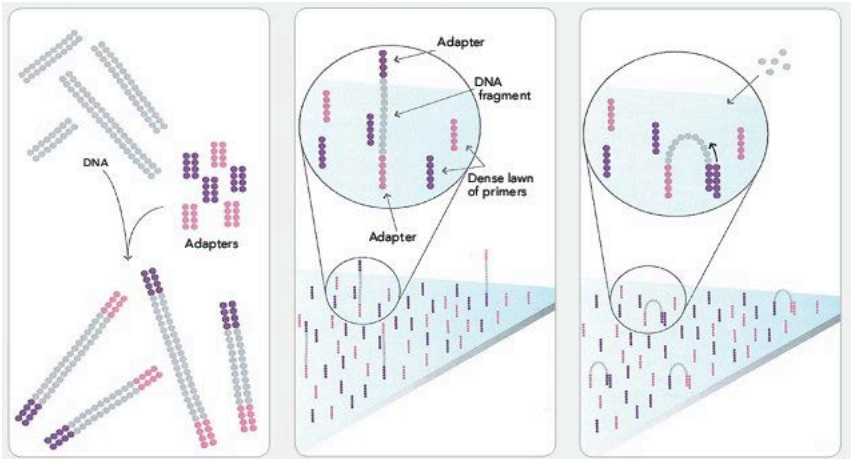
   Illumina sequencing: Cyclic-reversible termination sequencing

DNA Amplification: Polymerase Chain Reaction (PCR)



Illumina Sequencing: Bridge PCR

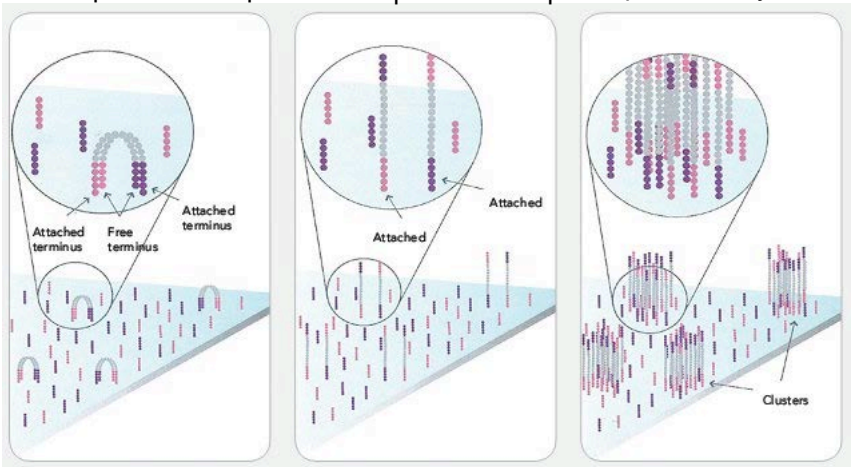
Library generation | Attach DNA to surface | Bridge PCR



Illumina, Inc.

Illumina Sequencing: Bridge PCR

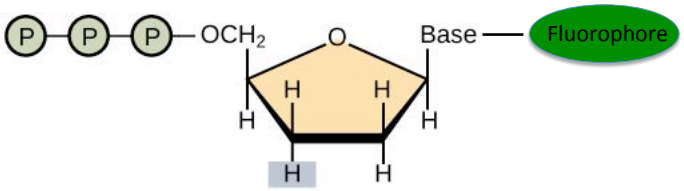
DNA fragment after 1st amplification step | Denaturation and next amplification step | Clusters after multiple amplification cycles



Illumina, Inc.

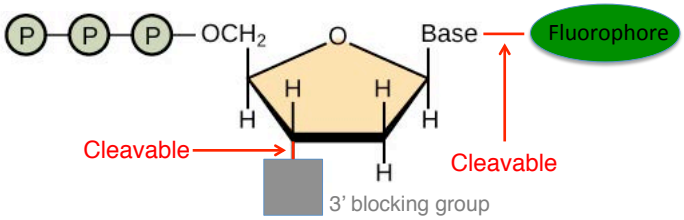
Reversible Terminator Nucleotides

Sanger Sequencing



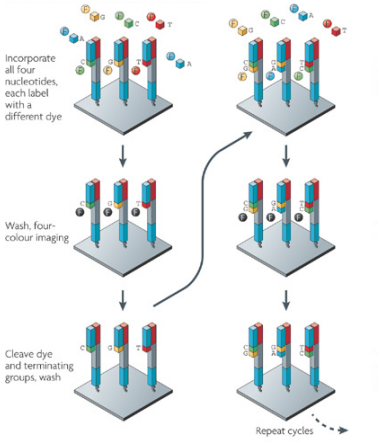
Dideoxynucleotide (ddNTP)

Illumina Sequencing

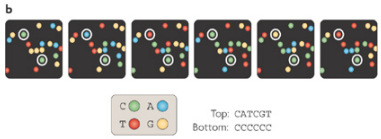


Reversible terminator nucleotide

Illumina Sequencing Reaction



Sequencing by synthesis with reversible terminators



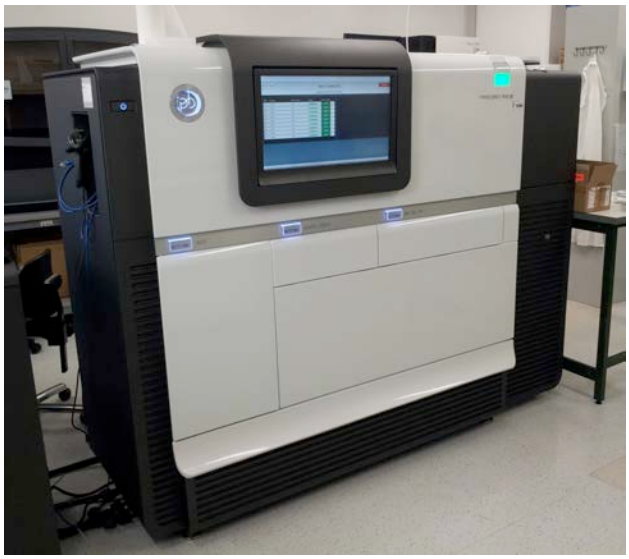
Imaging after each cycle

(Metzker M, 2010)

Part 3

The Future
Single-Molecule Sequencing
(Third-Generation Sequencing)

PacBio Sequencing

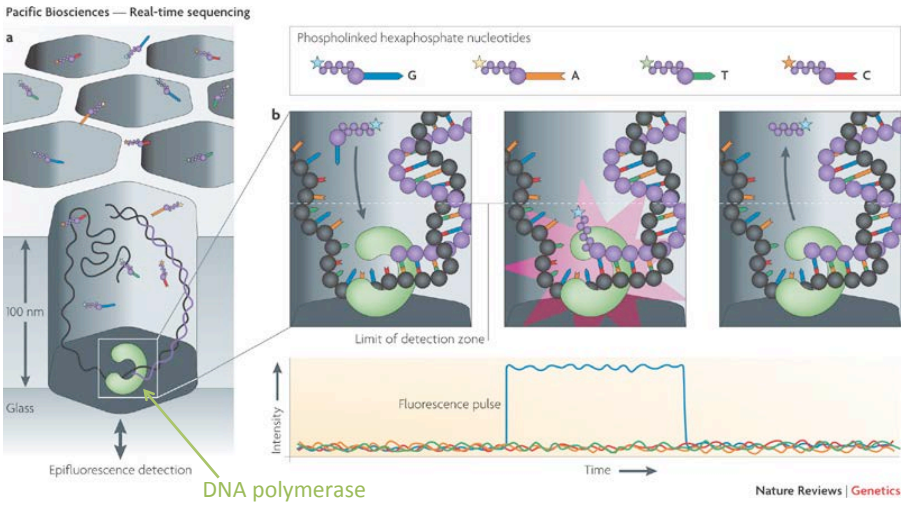


PacBio RSII

www.pacb.com

PacBio Sequencing

Single Molecule, Real-Time (*SMRT*) *Sequencing*



PacBio Sequencing

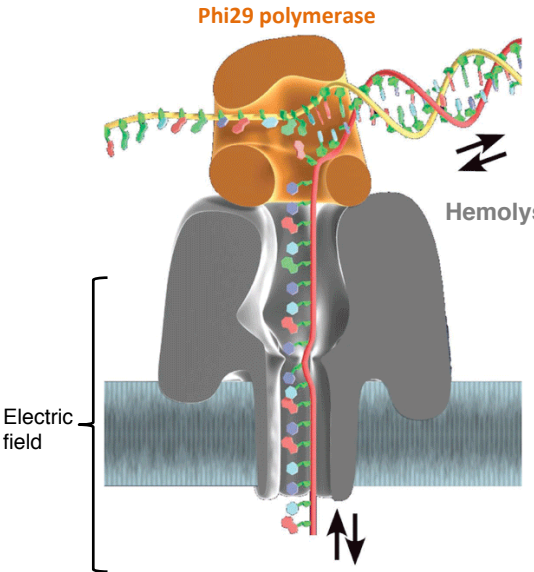
Advantages

- No amplification required
- Fast (> 1 nucleotide per second)
- Long reads (10,000-15,000 bases vs. 300-500 bases for Illumina)

Disadvantages

- Lower throughput (35,000-70,000 reads per run)
- Lower sequence yield (~5% of Illumina)
- High sequencing error rate (single pass 13% vs. 0.1%)

Nanopore Sequencing



(*Nature Biotechnology* 30, 326–328, 2012)

Electric field because of differences in ion concentration across the membrane

DNA moving through pore affects ion current in a base-specific manner ('electric signature')

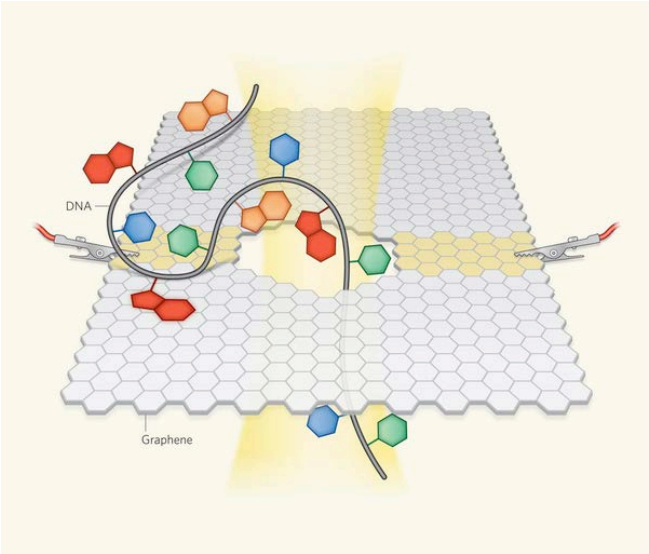
Changes in current are measured and recorded to determine the DNA sequence

Oxford Nanopore Technologies: MinIon

Real-time, long reads, but low through-put and high error rate

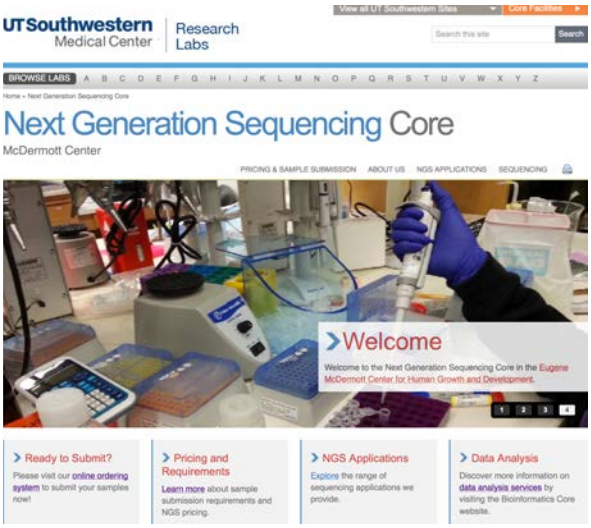
Forbes.com

The Future: Solid-state Nanopore Sequencing



(*Nature* 467, 164–165, 2010)

The McDermott NGS Core



http://www.utsouthwestern.edu/labs/next-generation-sequencing-core/