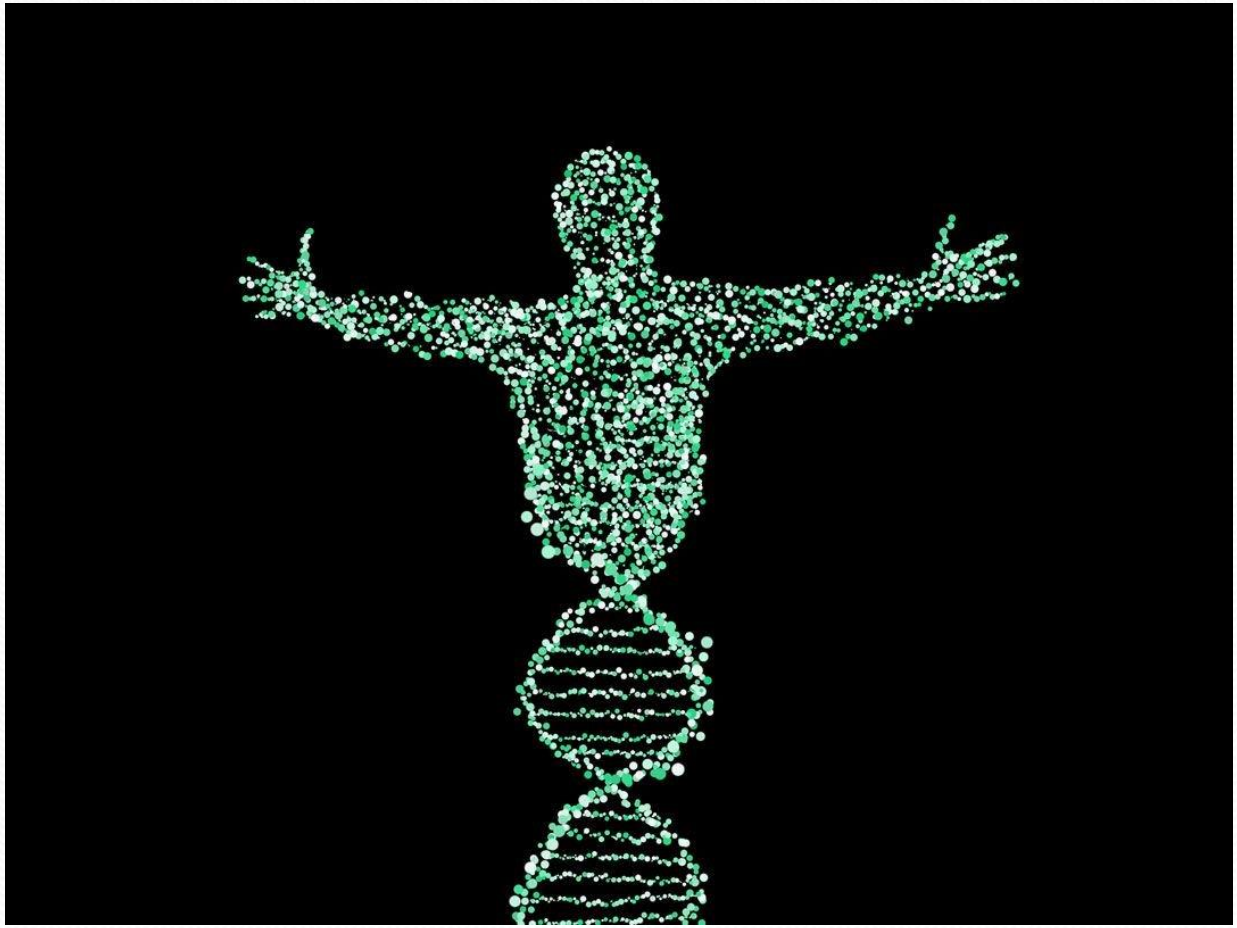
The Chemistry of Advanced Molecular Detection

Rebecca Kramer & Kevin Rodeman
Michigan Department of Health and Human Services
Bureau of Laboratories

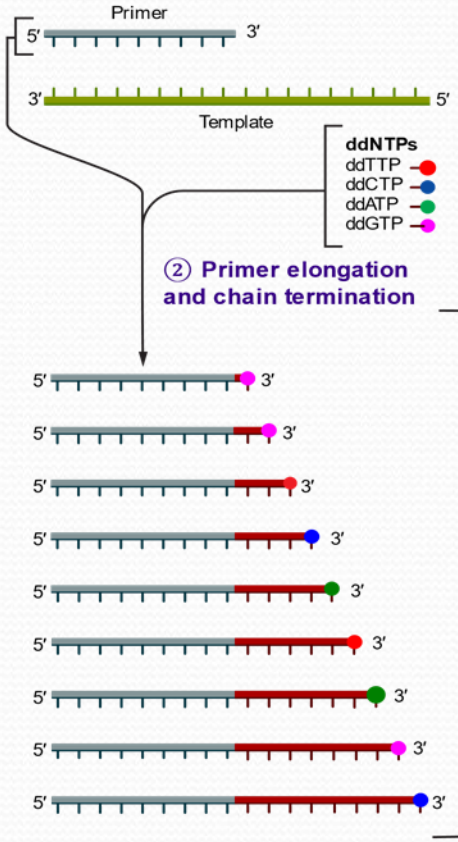
Prevent Disease – Promote Wellness – Improve Quality of Life

What is Sequencing?

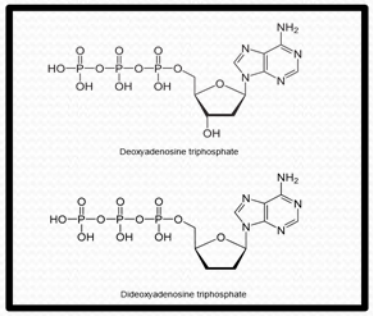


Sanger Sequencing

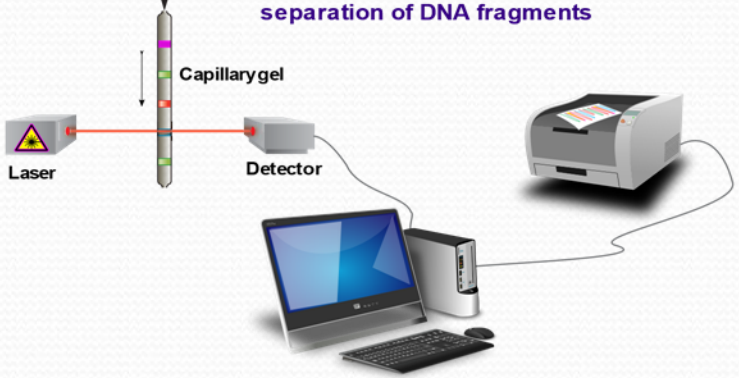
① Reaction mixture
- Primer and DNA template
- DNA polymerase
- ddNTPs with flourochromes
- dNTPs (dATP, dCTP, dGTP, and dTTP)

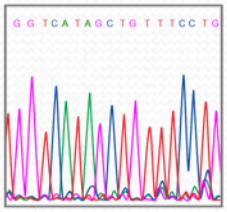


② Primer elongation and chain termination



③ Capillary gel electrophoresis separation of DNA fragments



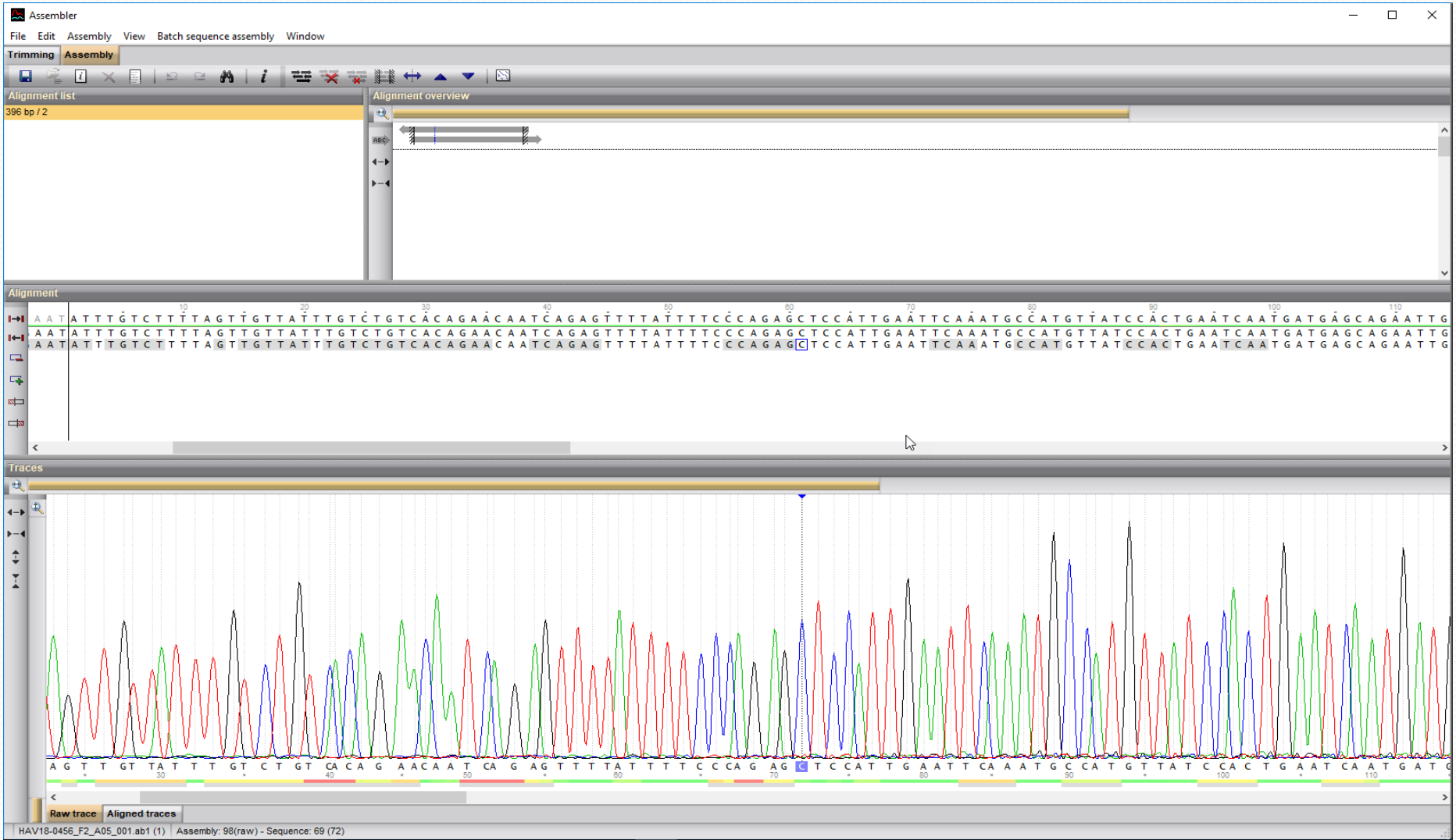


Chromatograph

④ Laser detection of flourochromes and computational sequence analysis

MDHHS Bureau of Laboratories

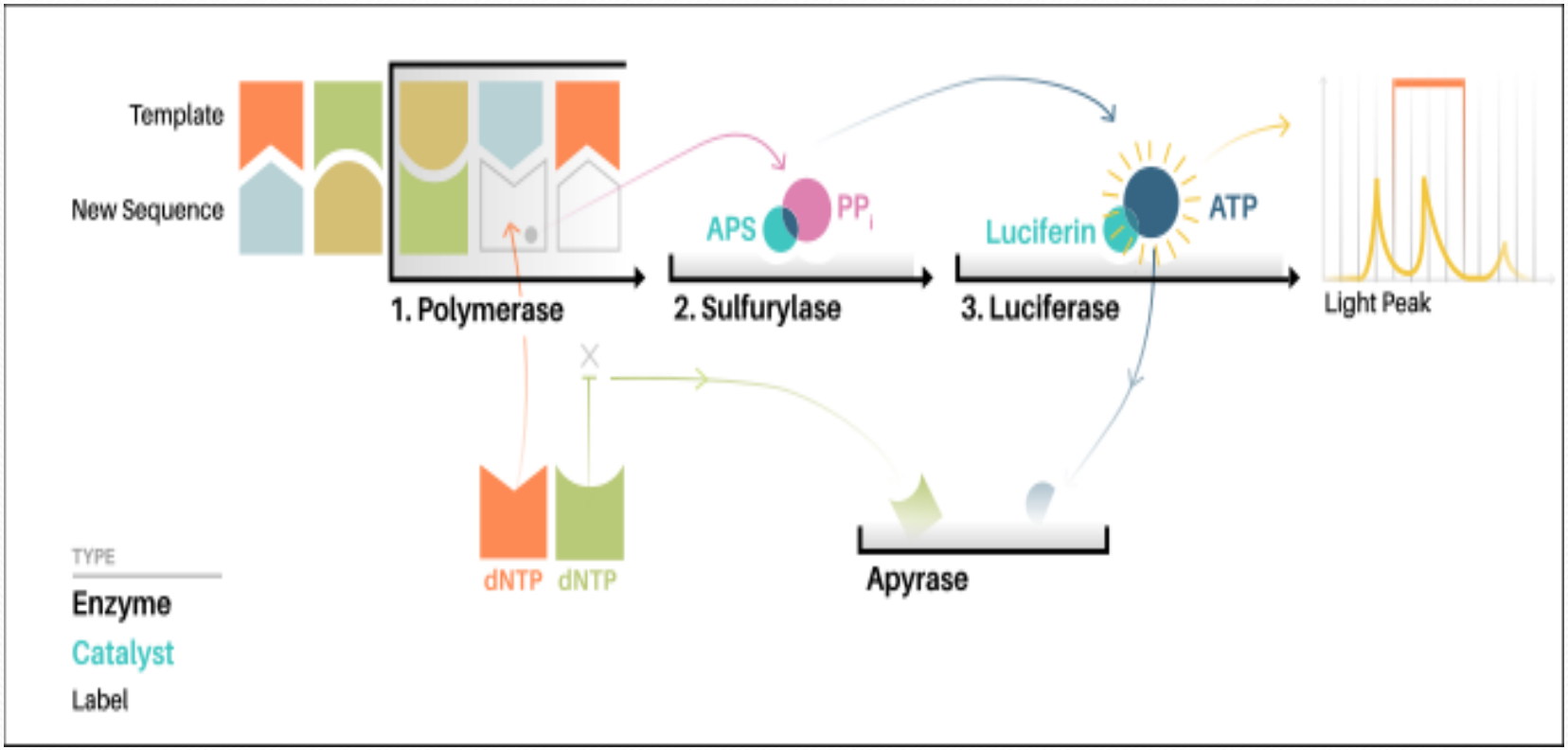
Sequence Analysis



Pyrosequencing - the Start of NGS

- Sequencing method based on the “sequencing by synthesis” principle.

Pyrosequencing Schematic



Roche 454 (Life Sciences): First Next Generation Sequencer

- Alternative approach by attaching DNA for sequencing on a microfabricated array.
- This allowed for high-through put DNA sequencing.
- Automated instrument allowed for a new era of genomics research.
- Limitations

Second Wave of NGS Devices

Illumina : MiSeq and NextSeq

- Attach indices to each DNA specimen.
- Specimens are pooled, diluted and denatured to create a library.
- The Library is added to the Flow Cell. Bridge amplification creates clusters of clones for each index set (specimen).
- Sequencing by Synthesis is done on the forward strand then on the reverse strand. Similar to pyrosequencing but chemistry is proprietary.
- Run time 4-48 hours; error rate 0.1%; 300bp read length

MDHHS Bureau of Laboratories

Ion Torrent

Ion Torrent™ next-generation sequencing

Discover the technology ›

Second Wave of NGS Devices

Ion Torrent Ion Proton

- Double stranded DNA is fragmented and double stranded adapter is added.
- The double stranded DNA is denatured and a sequencing primer is added.
- Emulsion PCR creates many clones and the amplicons attached to an acrylamide ball at 5' end. Only the forward strands are sequenced.
- The ball is added to a microchip with thousands of wells with each well containing an ion-sensitive field effect transistor.
- As dNTPs are added in waves, when a complimentary base is added a proton is detected as an electrical signal.
- Run time 2-3 hours; 200bp read length; 2% error rate.

MDHHS Bureau of Laboratories

3rd Generation Sequencing
aka
Long Read Sequencing

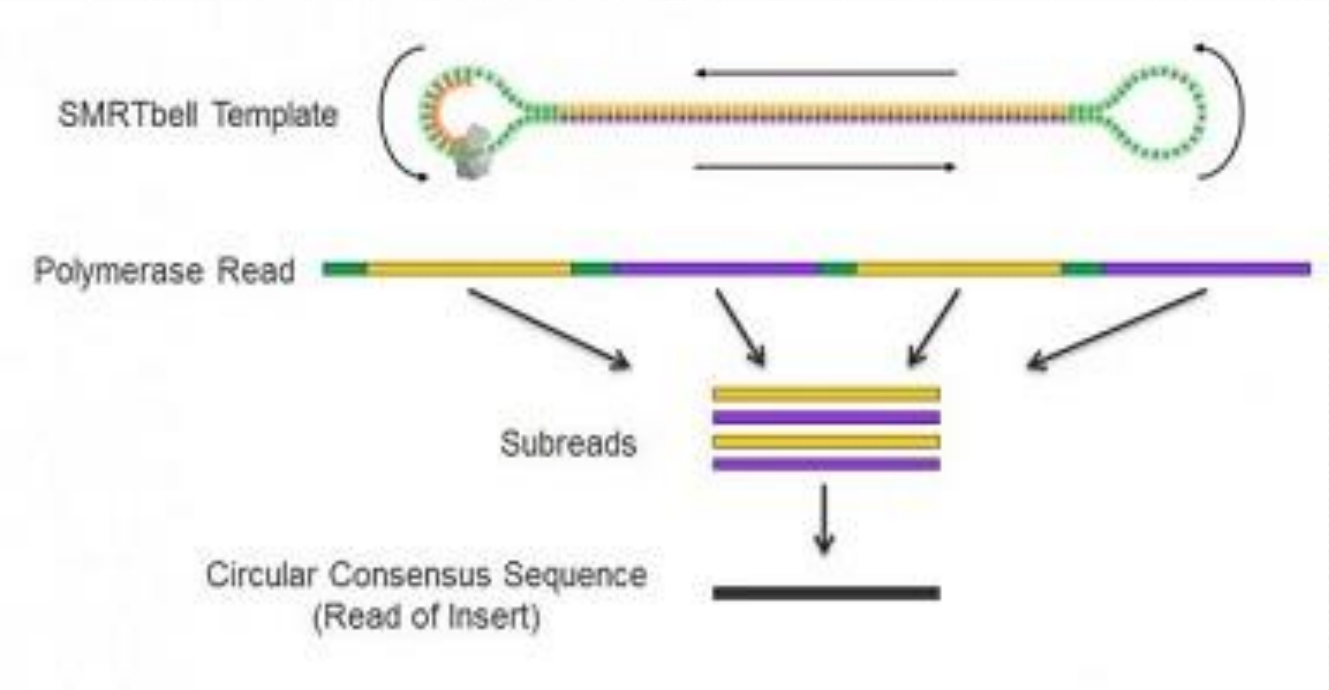
MDHHS Bureau of Laboratories

Third Generation Sequencing: PAC BIO

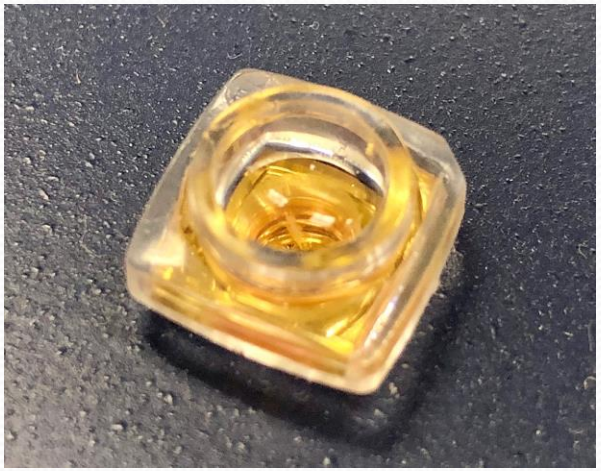
- SMRTbell™, stream lined protocol to create libraries of varying insert length 250bp to >20,000bp.
- DNA is sheared into fragments of desired size.
- Resulting fragments are repaired by treating the sample with DNA damage repair mix to repair nicks, abasic sites, and oxidation damage.
- Blunt ends are created on each end and the hairpin adapters are ligated to each end.
- In the final step, sequencing polymerase is bound to the template to sequence.

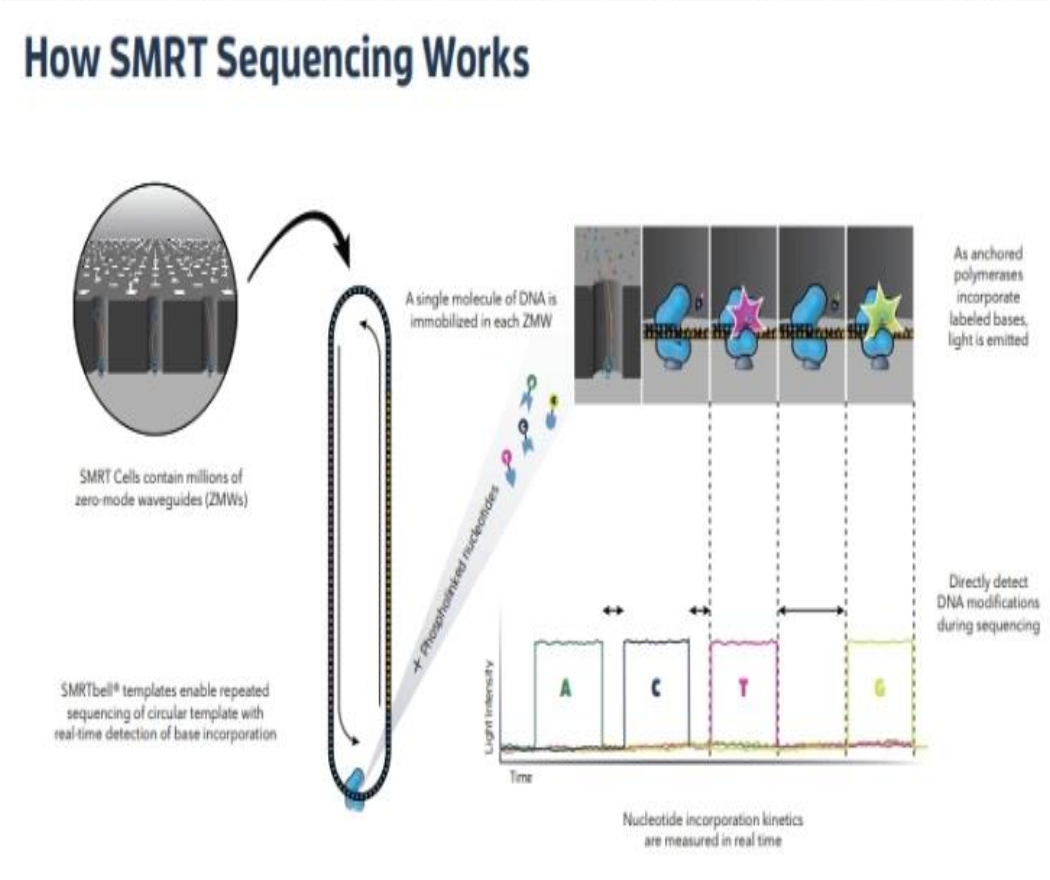
MDHHS Bureau of Laboratories

PAC BIO Sequencing



MDHHS
Bureau of Laboratories



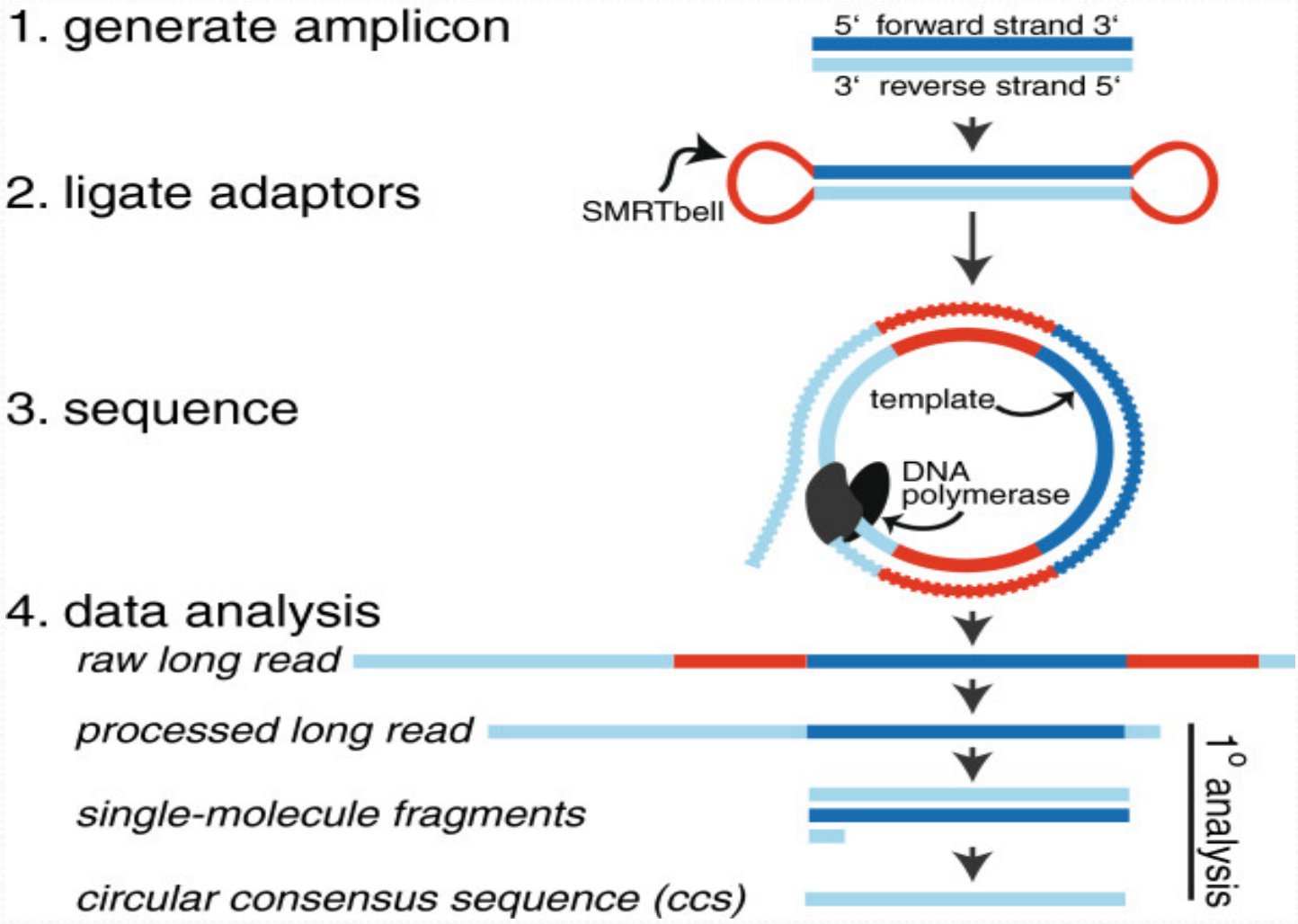

How SMRT Sequencing Works
SMRT Cells contain millions of zero-mode waveguides (ZMWs)
A single molecule of DNA is immobilized in each ZMW
SMRTbell® templates enable repeated sequencing of circular template with real-time detection of base incorporation
Phospholinked nucleotides
As anchored polymerases incorporate labeled bases, light is emitted
Directly detect DNA modifications during sequencing
Light intensity
A
C
T
G
Time
Nucleotide incorporation kinetics are measured in real time

1. generate amplicon

2. ligate adaptors

3. sequence

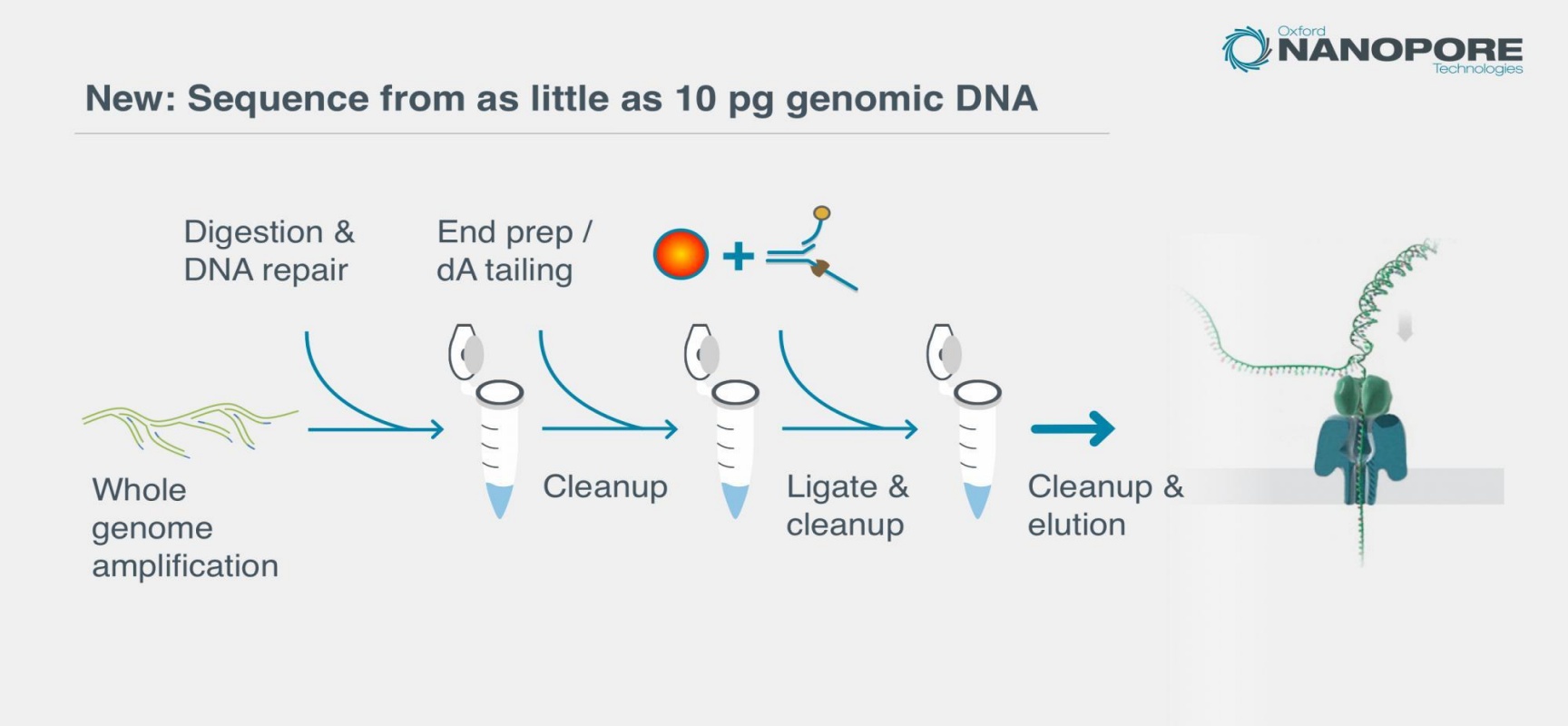
4. data analysis



MDHHS Bureau of Laboratories

Oxford Nanopore MinION

Oxford Nanopore MinION



MDHHS Bureau of Laboratories

Extractions

Manual versus Robotics

Instrument Extraction

- Roche: MagNA Pure Compact, MagNA Pure 24 and 96 System
- BIOMERIEUX NUCLISENS® easyMAG® and eMAG™
- QIAGEN: EZ1 Advanced XL, MagAttract 48/96 kits, QIAcube, and QIAsymphony

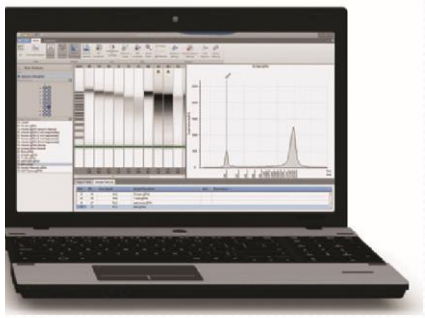
Manual Extractions

- QIAGEN QIAamp kits: DNA Mini and DNA Blood Mini Kits, and Viral RNA Mini Kits
- Roche: High Pure Kits
- Phenol/Chloroform Extraction
- Nuclease Treatment before extraction and DNase Treatment On-column for RNA Viruses

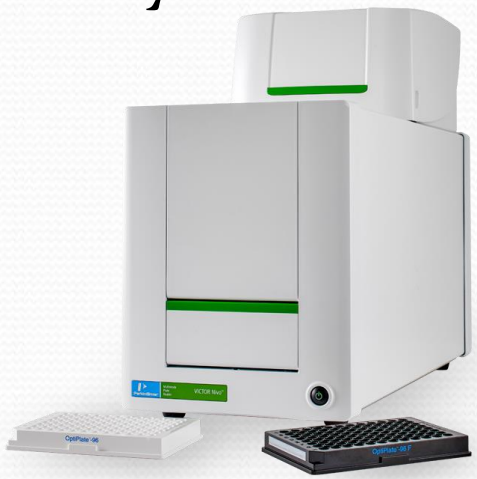
MDHHS Bureau of Laboratories

QC Checks

- Quality



- Quantity



MDHHS Bureau of Laboratories

METHOD	EXAMPLE	BRIEF DESCRIPTION	BENEFITS/LIMITATIONS
Spectrophotometry (260/280)	NanoDrop™	This method detects the absorption of UV light by the macromolecules in the sample.	✓ Low cost, as most laboratories already have access to UV/vis spectrophotometers ✗ Not specific for DNA ✗ Results can be skewed by RNA or protein contamination ✗ Cannot determine fragment size
Fluorimetry	Qubit®	This method measures the enhanced fluorescence of a dye upon binding to DNA/ macromolecules.	✓ Low cost, as most laboratories already have access to fluorimeters ✓ Can quantitate specifically dsDNA, ssDNA, RNA or protein ✗ Quantitates all nucleic acid present in sample, not just molecules to be sequenced ✗ Cannot determine fragment sizes
Electrophoretic	Bioanalyzer®, TapeStation®, Fragment Analyzer™	This method relies on capillary electrophoresis of DNA fragments for size estimation, as well as intercalating dyes for quantity determination.	✓ Accurate determination of fragment size distribution ✗ Less reliable quantitation ✗ Requires expensive equipment

MDHHS Bureau of Laboratories

Manual Pipetting

vs.

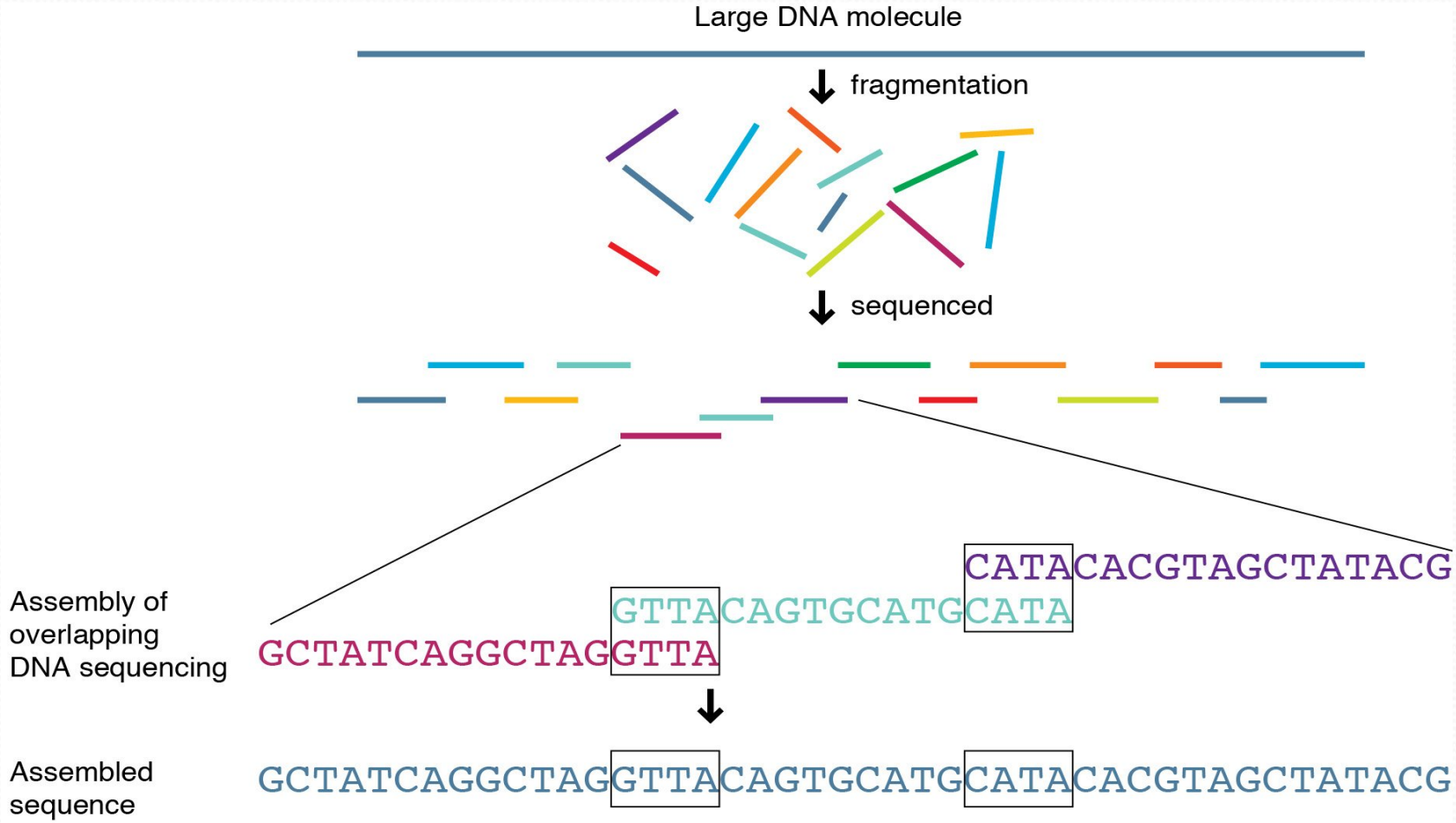
Liquid Handler

MDHHS Bureau of Laboratories

Basics of NGS Chemistry

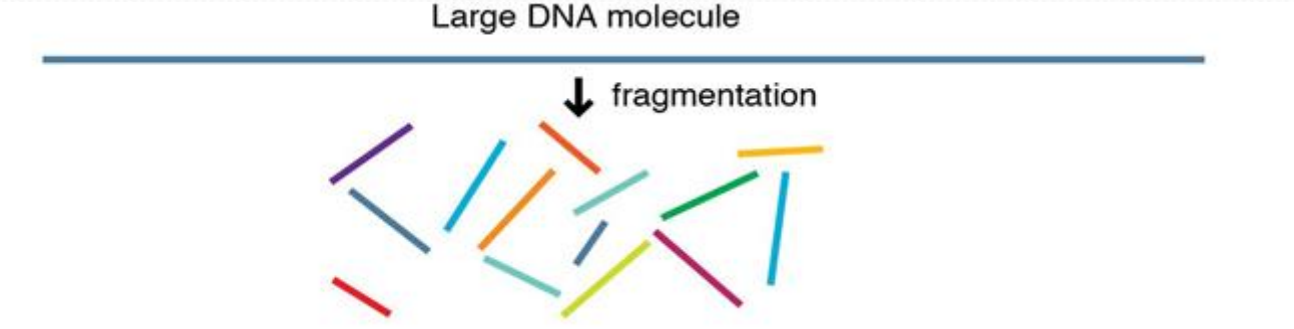
1. Library Prep
2. Cluster Generation
3. Sequencing
4. Data Analysis

MDHHS Bureau of Laboratories


Large DNA molecule
fragmentation
sequenced
Assembly of overlapping DNA sequencing
CATACACGTAGCTATACG
GTTACAGTGCATGCATA
GCTATCAGGCTAGGTTA
Assembled sequence
GCTATCAGGCTAGGTTACAGTGCATGCATACACGTAGCTATACG

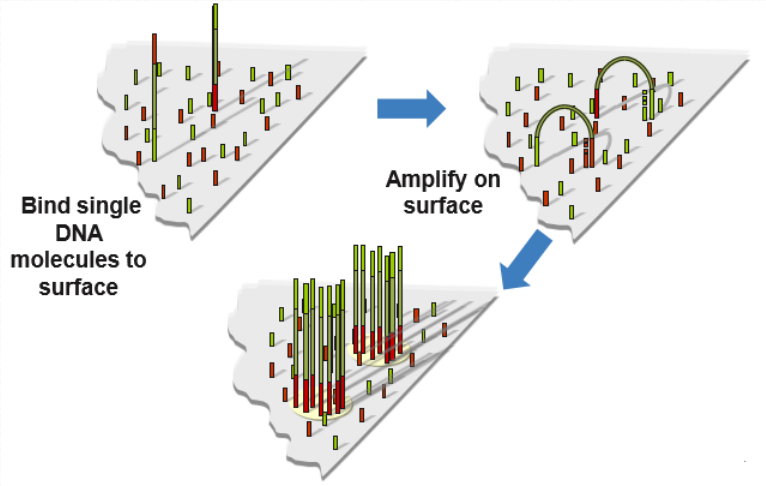
1. Library Prep

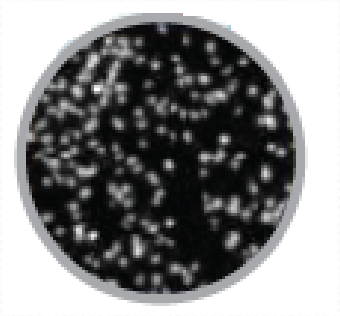
- Random Fragmentation of DNA
- 5' and 3' Adapter Ligation
- PCR Amplification
- Purification
- Pooling



2. Cluster Generation

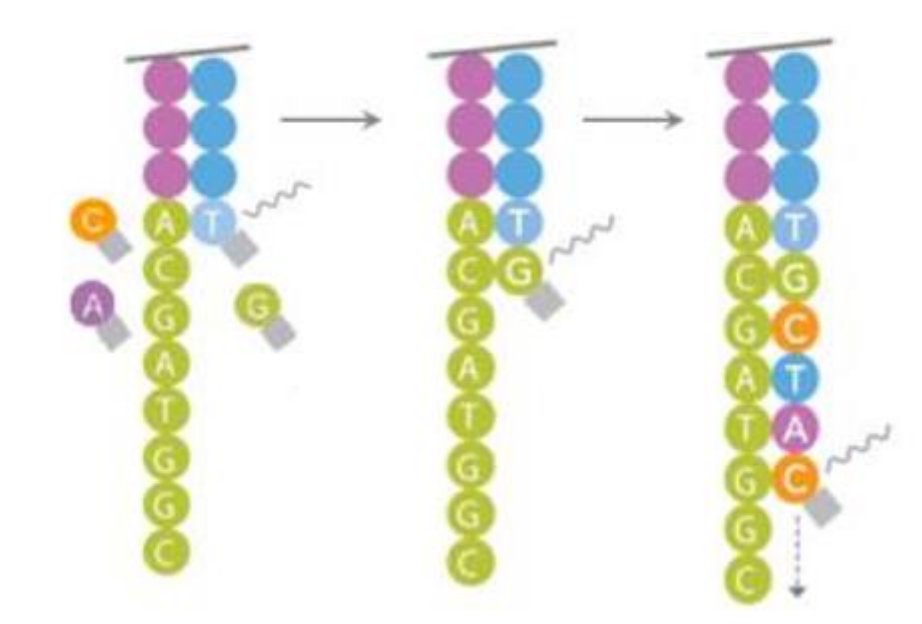
- Library loaded onto flow cell
- Fragments bind to surface bound oligos
- Amplified into clusters through bridge amplification
- Clusters templates ready for sequencing

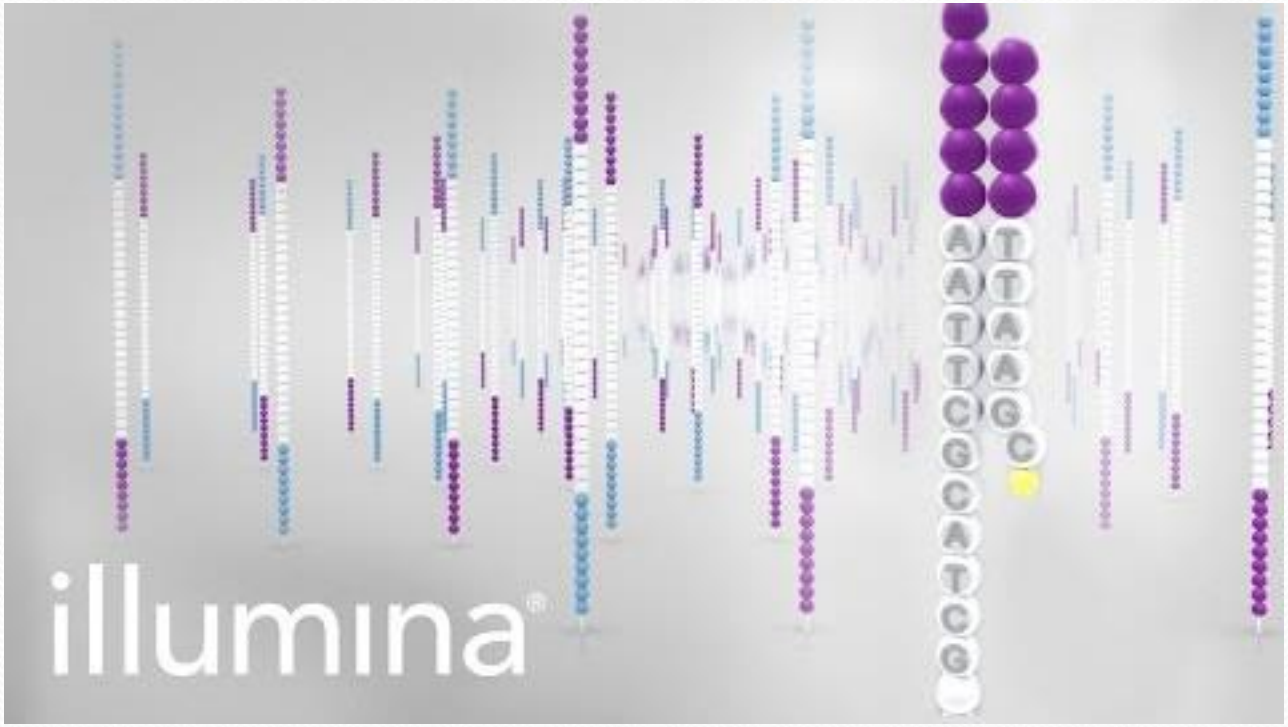




3. Sequencing

- Sequencing by Synthesis

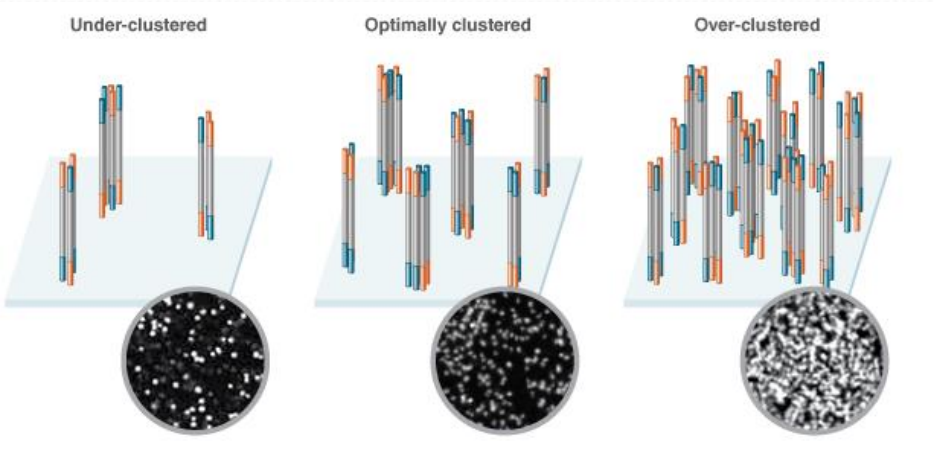




MDHHS
Bureau of Laboratories

Run Monitoring

- Cluster Density - under clustering can lead to insufficient data, while over clustering can lead to diminished data quality



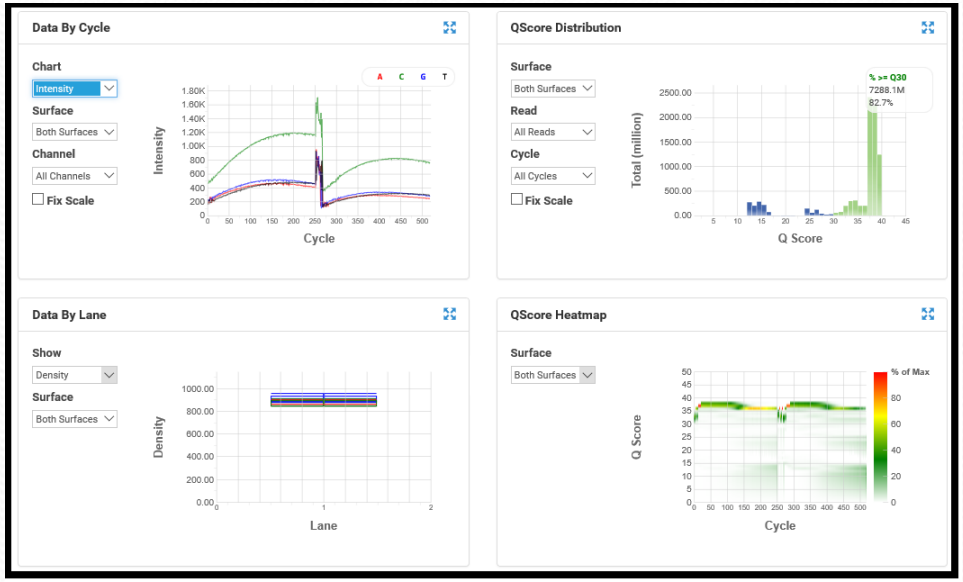
Run Monitoring (continued)

- Clusters Passing Filter
- Q30 Score
- Estimated Yield

MDHHS Bureau of Laboratories

Run Analysis

- Sequence Analysis Viewer
  - 1. Q-score
  - 2. Data by Lane
  - 3. Intensity
  - 4. Full Width/ Half Max
  - 5. Summary Tab



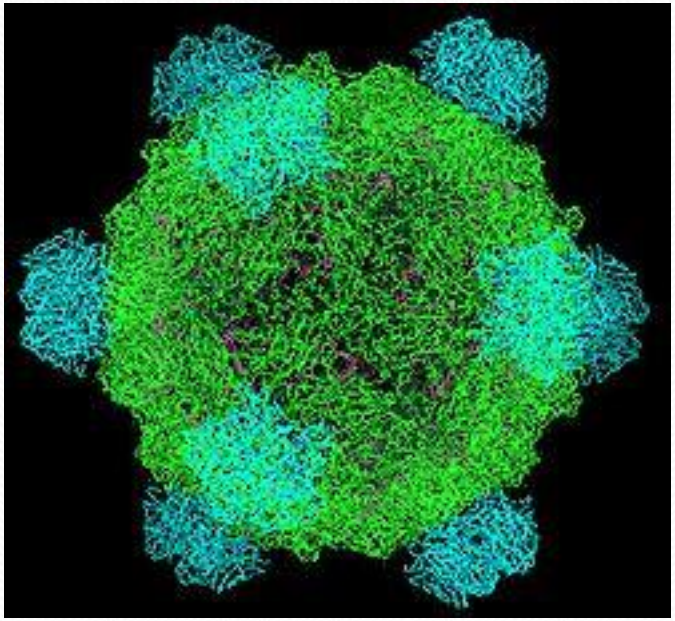
GHOST MiSeq Run Analysis

GHOST Run No.	Cluster Density (K/mm^2)	Clusters Passing Filter	Estimated Yield (MB)	Q Score >=Q30	MiSeq Reagent Kit
MI-GHOST04	855	90.3	11547.6	78.4	600 cycles v3
MI-GHOST06	981	93.9	13471	9.7G 87.7%	600 cycles v3
MI-GHOST07	957	92.4	12893.8	9.4G 89.0%	600 cycles v3
MI-GHOST08	726	93.2	6624.5	5.9G 90.4%	500 cycles v2
MI-GHOST09	883	91.5	561.8	0.5G 86.1%	500 cycles v2 Nano
MI-GHOST10	840	94.6	552.7	0.5G 88.8%	500 cycles v2 Nano
MI-GHOST11	844	89.9	523.9	0.5G 90.7%	500 cycles v2 Nano
MI-GHOST12	870	91.9	7697.1	6.8G 90.1%	500 cycles v2

MDHHS Bureau of Laboratories

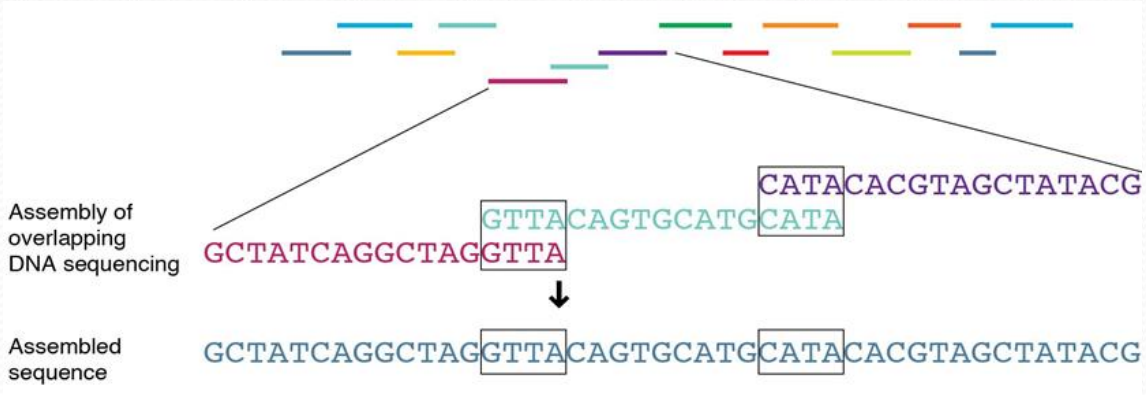
Troubleshooting

- PhiX – acts as positive control for instrument and reagent performance



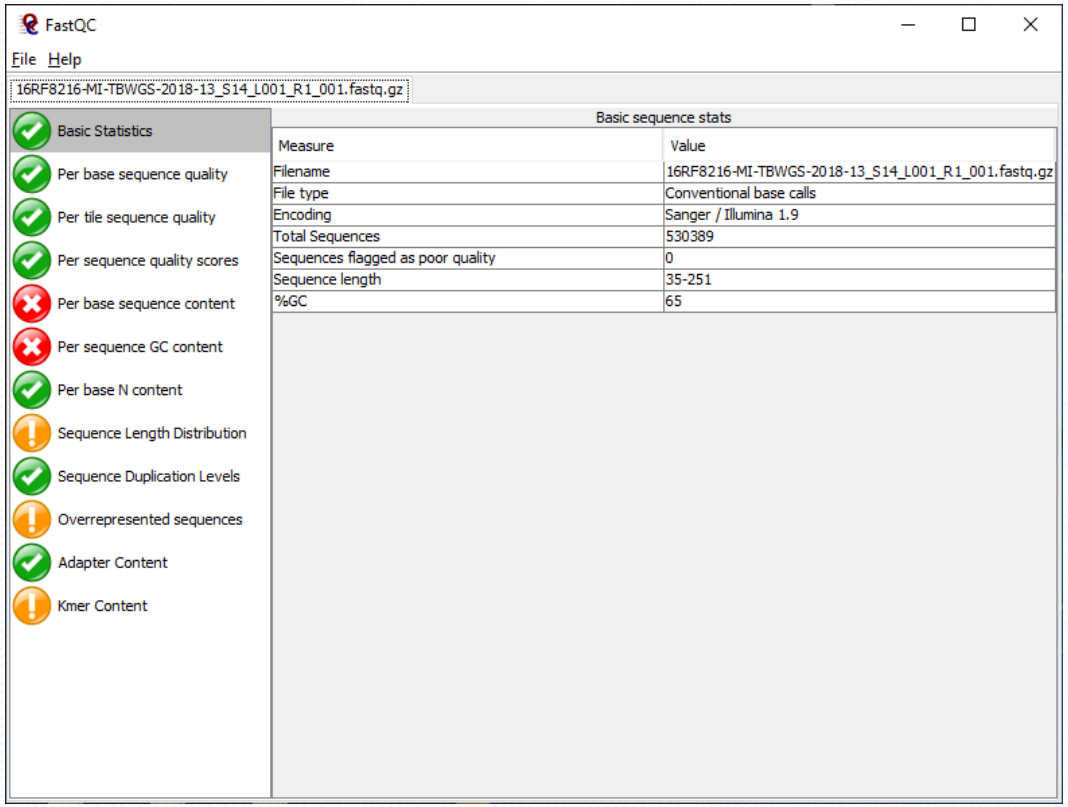
4. Data Analysis

- Newly identified reads are aligned to reference genome
- After alignment, many analysis options: SNP (single nucleotide polymorphism), Indel (Insertion/Deletion), Read Counting for RNA, Phylogenetic Analysis, Metagenomic Analysis



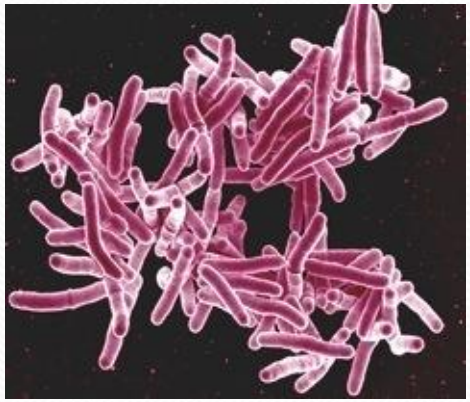
Determining Coverage

- Fast QC



Determining Coverage

- Example:



- 530389 sequences
- 2x250 cycle kit
- TB genome – 4.5 million base pairs

- (530389 sequences*500 cycles)/4500000 base pairs
  = 58.9x theoretical coverage

Sequencing Platforms

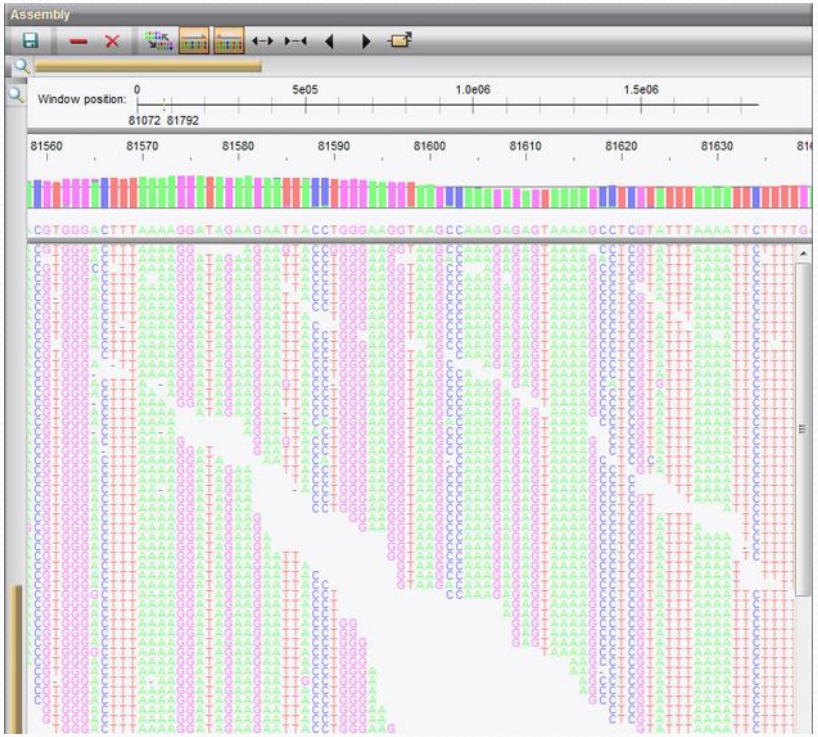



Method	Generation	Read length (bp)	Single pass error rate (%)	No. of reads per run	Time per run	Cost per million bases (USD)
Sanger ABI 3730×1	1st	600–1000	0.001	96	0.5–3 h	500
Ion Torrent	2nd	200	1	8.2×10^7	2–4 h	0.1
454 (Roche) GS FLX +	2nd	700	1	1×10^6	23 h	8.57
Illumina HiSeq 2500 (High Output)	2nd	2 × 125	0.1	8×10^9 (paired)	7–60 h	0.03
Illumina HiSeq 2500 (Rapid Run)	2nd	2 × 250	0.1	1.2×10^9 (paired)	1–6 days	0.04
PacBio RS II: P6-C4	3rd	$1.0–1.5 \times 10^4$ on average	13	$3.5–7.5 \times 10^4$	0.5–4 h	0.40–0.80
Oxford Nanopore MinION	3rd	$2–5 \times 10^3$ on average	38	$1.1–4.7 \times 10^4$	50 h	6.44–17.90



MDHHS Bureau of Laboratories

Software Analysis



Best Alignment Software?

There is no best read aligner! It depends on each laboratories goals and testing needs. What is the application? What sequencing technology? What is the species? What are the computational constraints, etc.? One must take into account these questions, each software's features, and determine the suitable read mapper software.

MDHHS Bureau of Laboratories

Any Questions?

MDHHS
Bureau of Laboratories