Genomics: A beginner's guide to Sequencing

Icipe MBBU

15th Oct 2021

EANBiT Bioinformatics internship and incubation program

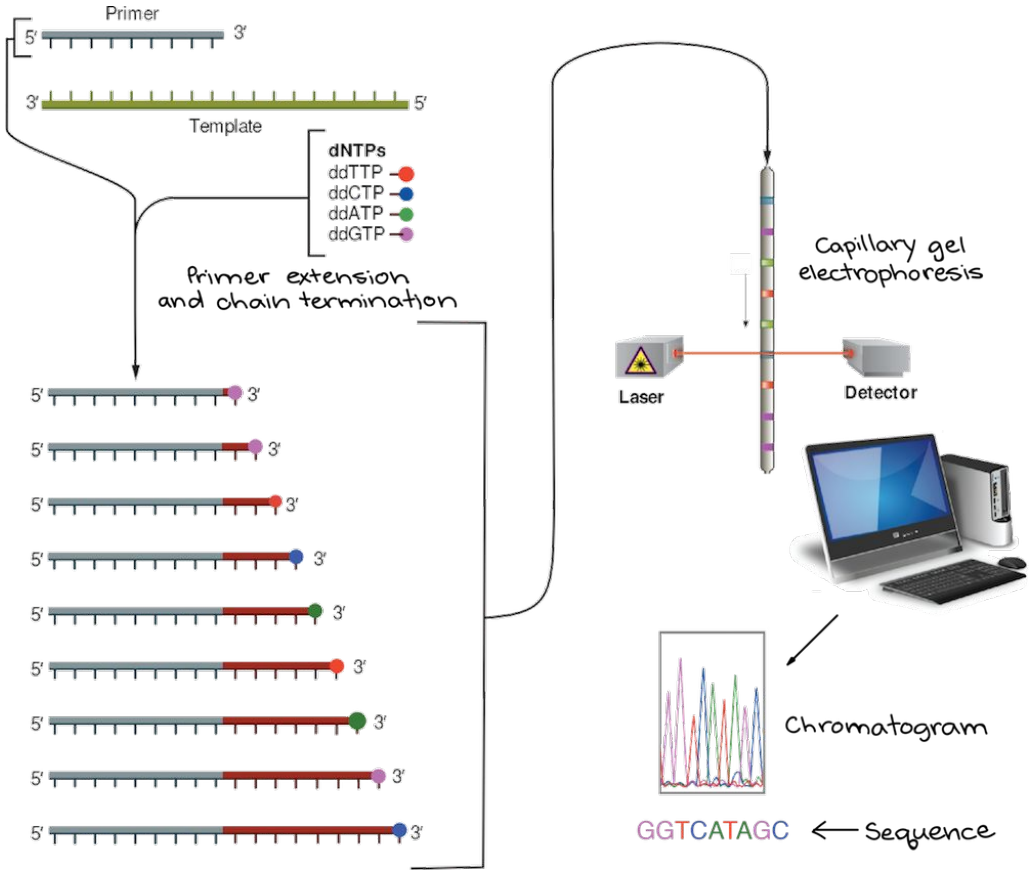
E Kariuki

Sanger sequencing/Chain termination sequencing

Developed by Fredrick Sanger in the 1970's.

I. Primer binds at 3' end
II. DNA polymerase enables primer extension
III. Termination of extension using dideoxynucleotides.
IV. Fluorescence is used to identify the chain terminating nucleotide
V. Capillary electrophoresis to separate the extension products and determines the bases.
VI. Light sensor at the end of the capillary can determine the different signal and interprets it as a base-call.
VII. End product is the chromatogram shown on the right and a .txt/fasta file.

Read Lengths are ~500-800bps



Credit: Khan Academy

Illumina sequencing

1. **Sample preparation:**

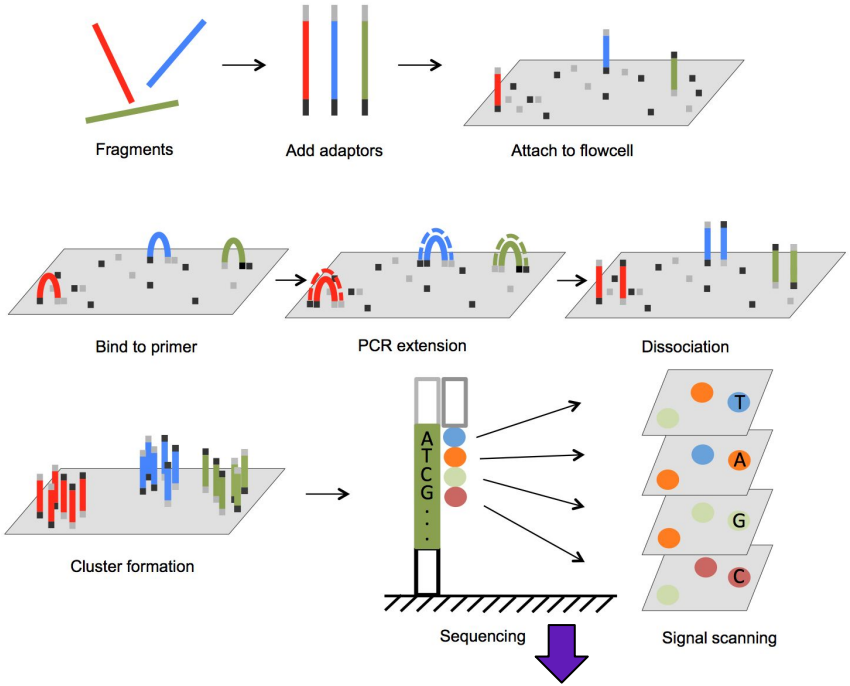
Nucleic acids are extracted and fragmented during library preparation before adapters can be added on.

2. **Cluster generation:**

- Each fragment is amplified by oligonucleotide hybridization (bridge amplification) on a flow cell.
- Reverse strands are cleaved, and washed off leaving the forward strands.
- 3’ ends are blocked to prevent unwanted priming.

3. **Sequencing:**

- Fluorescently tagged nucleotides that match template sequence are added creating a growing chain.
- After adding a nucleotide, a fluorescent signal is emitted.
- This process is known as sequencing-by-synthesis. The no. of cycles determine the length of the read.
- Emission wavelength and signal intensity determine the basecall. Length of the reads is 50 - 300 bp.



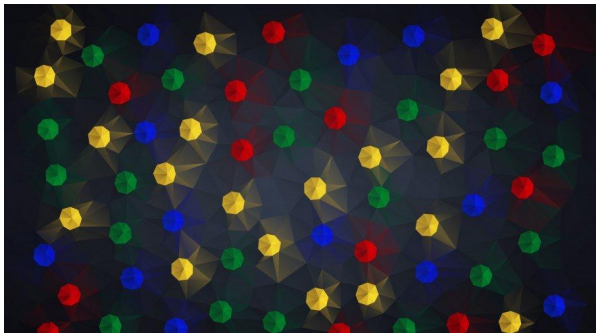
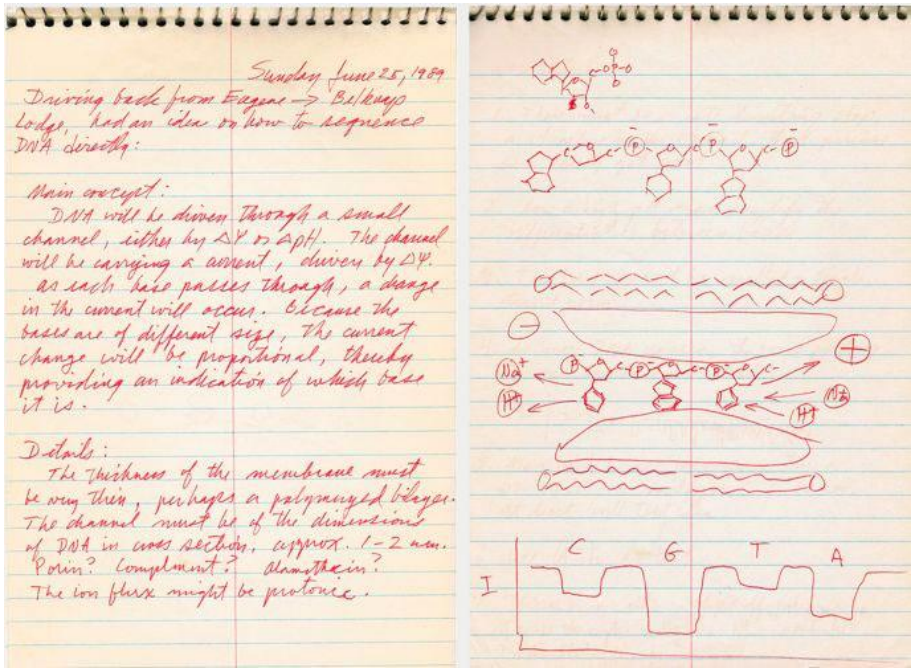


Image credit: yourgrnome.org

The conception to the launch



1989

Idea conception

by prof. David Deamer of UC. Davis.
Shelves the ideas for two years.

1991

Idea development

Professor Dan Branton visits UC Davis and they jointly decide to pursue research into the concept of nanopore sensing, and in 1992 it was agreed that Harvard would be the lead institution for the IP.

1993 - 2010

Then comes the endless research ...

These were the formative years of the research towards understanding the technology better, acquiring and getting the appropriate sequencing chemistry, forging partnerships and acquiring patents for the technology

2012-

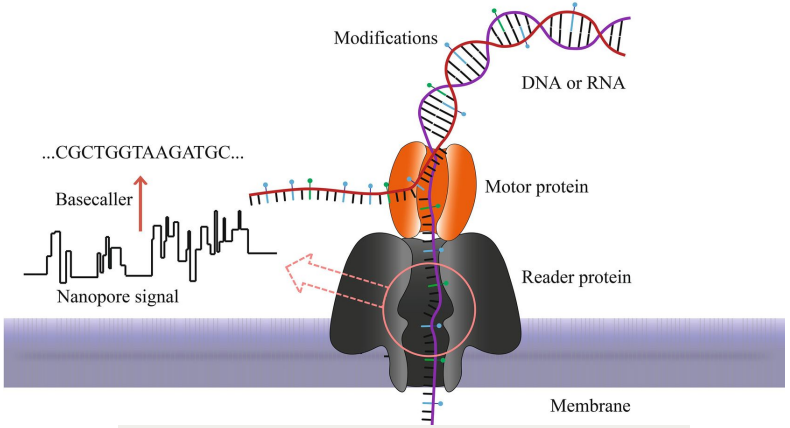
Commercialization

The MinION followed by the GridION and the PromethION platforms were launched. The story continues.

Oxford Nanopore Technologies (ONT) sequencing

- Protein nanopores are embedded in a synthetic membrane bathed in an electrophysiological solution.
- An ionic current is passed through the nanopores. As the nucleic acid molecules pass through the pore they disrupt the current.
- The signal can then be analyzed in realtime to determine the base passing through.
- The nanopore analyzes the entire fragment of DNA or RNA hence the read length is proportional to the DNA or RNA in the sample prepared.
- From the many ONT preparation kits, users can influence the read lengths by choosing the right preparation kit to achieve the desired results.
- Read lengths range from ~1Kb to 2Mb.

https://genomebiology.biomedcentral.com/articles/10.1186/s13059-016-1103-0



Credits: Oxford Nanopore Technologies

ONT: Applications

***Microbial Resistance*:** Identification and antimicrobial resistance (AMR) prediction using EPI2ME Workflows.

Cancer genomics: Identify mutations in complex and repetitive genomic regions

Rapid assay sequencing experiments: Achieve rapid turnaround with immediate access to results e.g. genotyping

Minimum sequencing and computational infrastructure: Making it portable and convenient for off-lab/field activities.

Cost effective

Costing just $90 per flow cell, Flongle is the quickest, most accessible solution for smaller tests and experiments.

Real-time data

Immediate access to data for rapid, actionable results.

On demand

Run single samples on demand instead of multiplexing.

Quality assessment

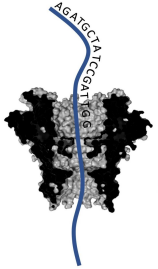
Rapidly assess sample quality prior to starting a larger sequencing experiment.

The $90 Flongle flow cell from ONT capable of generating about ~2.8 GB data

Image Credits: ONT

Sequencing: The Good, bad and the Ugly

Technology	Advantages	Limitations
Sanger sequencing	• Long reads • High accuracy	• Low throughput • Reads <1Kb
Second-generation sequencing i.e. Illumina	• High throughput • High accuracy	• Short reads (50 bp-500 bp) • Amplification biases
Third generation sequencing i.e. Oxford Nanopore Technologies Sequencing	• Long reads (~1kb -2Mb) • High throughput • PCR Amplification not necessary • Can identify and differentiate between base modifications • Simpler library preparation methods • Cost effective and convenient • Enables unambiguous de novo assembly and reference mapping. • Accuracy greatly improved using error-correction algorithms	• High error rates (~5-11%) • Biased towards long fragments



Oxford Nanopore Technologies:

downstream analyses

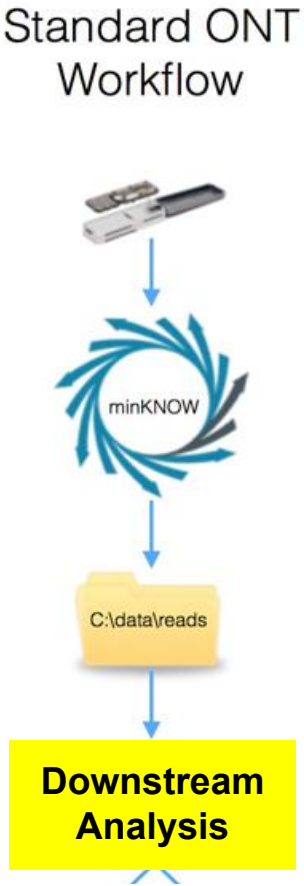
(Jean-Baka Domelevo Entfellner, Peter Emmrich, Oluwaseyi Shorinola, 2020)

MinKNOW software

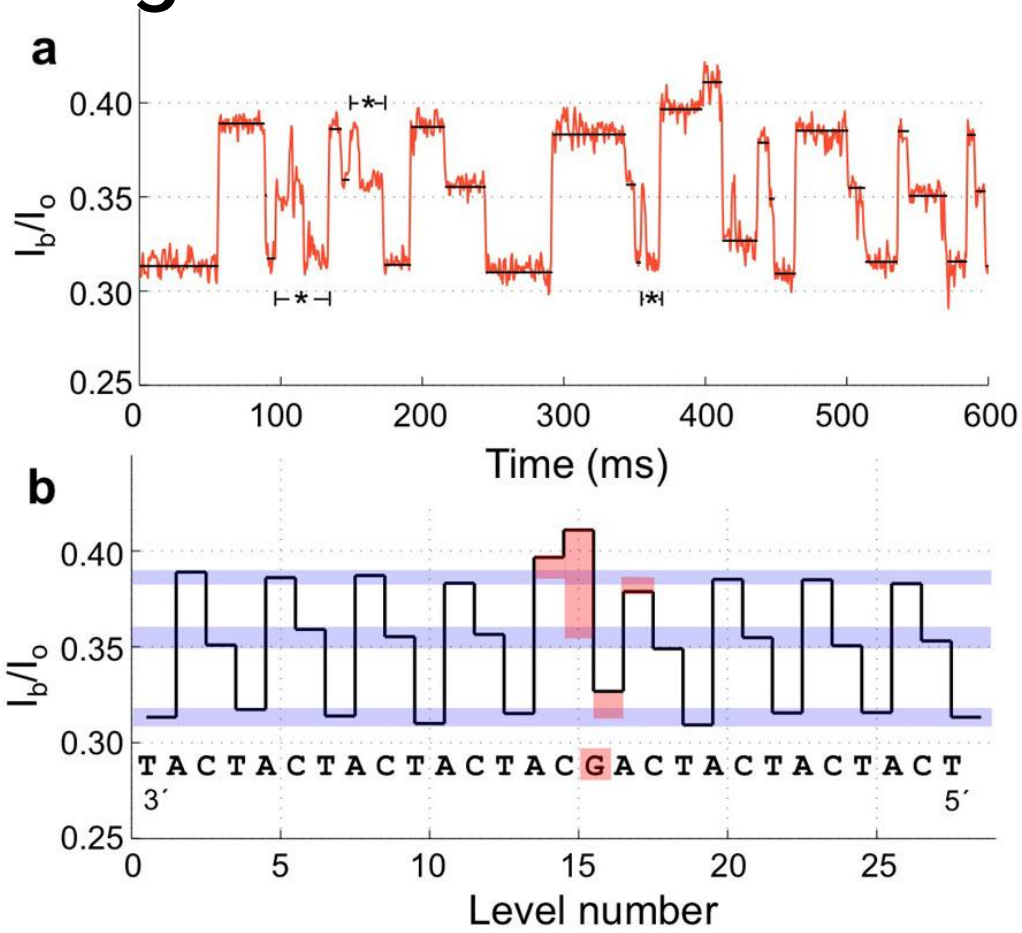
- MinKNOW is the proprietary operating software from ONT that drives nanopore sequencing devices.

It carries out several core tasks such as:

- Data acquisition,
- Real-time analysis and feedback
- Local basecalling
- Data streaming – whilst providing device control including selecting the run parameters,
- Sample identification and tracking,
- Ensuring that the platform chemistry is performing correctly to run the samples.



ONT Basecalling



The figure above shows how the fast5 signals are interpreted.

Sequencing: The Good, bad and the Ugly

Technology	Advantages	Limitations
Sanger sequencing	• Long reads • High accuracy	• Low throughput • Reads <1Kb
Second-generation sequencing i.e. Illumina	• High throughput • High accuracy	• Short reads (~100-500bp) • Amplification biases
Third generation sequencing i.e. Oxford Nanopore Technologies Sequencing	• Long reads (~1kb -2Mb) • High throughput • PCR Amplification not necessary • Can identify and differentiate between base modifications • Simpler library preparation methods • Cost effective and convenient • Enables unambiguous de novo assembly and reference mapping. • Accuracy greatly improved using error-correction algorithms	• High error rates (~5-11%) • Biased towards long fragments

Additional Resources

https://www.atdbio.com/content/58/Next-generation-sequencing