# Nanopore Sequencing Technology and Tools:

## Computational Analysis of the Current State, Bottlenecks and Future Directions

Damla Senol[1], Jeremie Kim[1,3], Saugata Ghose[1], Can Alkan[2] and Onur Mutlu[3,1]

[1] Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, USA
[2] Department of Computer Engineering, Bilkent University, Bilkent, Ankara, Turkey
[3] Department of Computer Science, Systems Group, ETH Zürich, Zürich, Switzerland

Carnegie Mellon ETH zürich  Bilkent University SAFARI

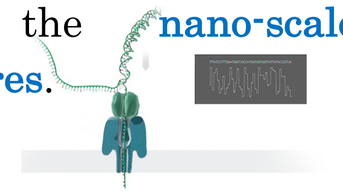
## Nanopore Sequencing

### Nanopore Sequencing

- a **single molecule DNA sequencing technology** that could potentially surpass current sequencing technologies
- promises
  - **higher throughput**
  - **lower cost**
  - **increased read length**
  - no prior amplification step.

It has one major drawback: **high error rates**.

Nanopore sequencers rely solely on the electrochemical structure of the different nucleotides for identification and measure **the change in the ionic current** as long strands of DNA (ssDNA) pass through the **nano-scale protein pores**.



### Biological Nanopores for DNA Sequencing

- first proposed in the 1990s,
- recently made commercially available in May 2014 by **Oxford Nanopore Technologies (ONT).**

### MinION

- first commercial nanopore sequencing device
- high-throughput sequencing apparatus
- produces real-time data
- inexpensive
- pocket-sized / portable

## Pipeline and Current Tools

### Step 1. Basecalling

- Translates raw signal output of MinION to generate DNA sequences.
- *Metrichor [1]* (extracted with *poretools [2]*), *nanonet [3]*, *nanocall [4]*

### Step 2. Genome Assembly for noisy long reads

- Using only the basecalled DNA reads, generates longer contiguous fragments called draft assemblies.
- *canu [5]*, *miniasm [6]*

### Step 3. Polishing (Optional)

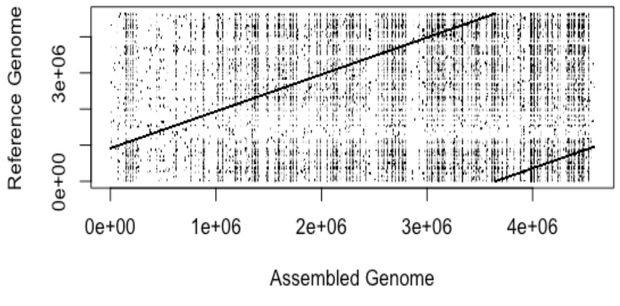
- Generates an improved consensus sequence from the draft assembly
- *nanopolish [7]*, *racon [8]*

[1] Metrichor. [https://nanoporetech.com/products/metrichor]
[2] Loman, Nicholas J., and Aaron R. Quinlan. "Poretools: a toolkit for analyzing nanopore sequence data." *Bioinformatics* 30.23 (2014): 3399-3401.
[3] Nanonet. [https://github.com/nanoporetech/nanonet]
[4] David, Matei, et al. "Nanocall: An Open Source Basecaller for Oxford Nanopore Sequencing Data." bioRxiv (2016): 046086.
[5] Berlin, Konstantin, et al. "Assembling large genomes with single-molecule sequencing and locality-sensitive hashing." Nature biotechnology 33.6 (2015): 623-630.
[6] Li, Heng. "Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences." Bioinformatics (2016): btw152.
[7] Loman, Nicholas J., Joshua Quick, and Jared T. Simpson. "A complete bacterial genome assembled de novo using only nanopore sequencing data." Nature methods (2015).
[8] Vaser, Robert, et al. "Fast and accurate de novo genome assembly from long uncorrected reads." *bioRxiv* (2016): 068122.

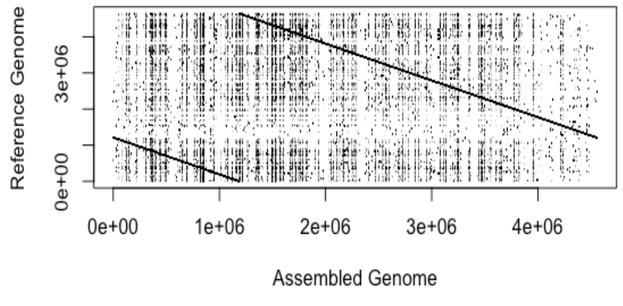
## Our Goal

- In order to take advantage of nanopore sequencing, it is important to **increase the accuracy and the speed of the whole pipeline.**
- Although new nanopore chemistry R9 improves the data accuracy, the tools used for nanopore sequence analysis are of critical importance as they should **overcome the high error rates of the technology.**
- **Our goal** in this work is to comprehensively analyze tools for nanopore sequence analysis, with a focus on understanding **the advantages, disadvantages, and bottlenecks of the various tools.**

## Results




| Basecaller | Execution Time (Basecalling) | Assembler | Execution Time (Assembly) | # contigs | # bp | # bp / genome length | % Accuracy |
|---|---|---|---|---|---|---|---|
| Metrichor | - | Canu | 2195m | 1 | 4,586,878 | 0.9882 | 99.7769 |
| | | Miniasm | 372sec | 1 | 4,477,194 | 0.9646 | 89.9817 |
| Nanonet | 2517m | Canu | 124m | 2 | 4,559,230 | 0.9822 | 99.7809 |
| | | Miniasm | 60sec | 6 | 3,549,677 | 0.9491 | 90.2123 |
| Nanocall (Fast mode) | 5359m | Canu | 154m | 0 | - | - | - |
| | | Miniasm | 607sec | 0 | - | - | - |




- ONT's **cloud-based basecaller**, *Metrichor* and **local basecaller**, *nanonet* perform similarly with **high accuracy**. However, another **local basecaller** *nanocall* is **not suitable for R9 data**.
- *Canu*, **the assembler with error correction**, produces **high-quality assemblies** but is **relatively slow** compared to *Miniasm*, **the assembler without error correction**.
- *Miniasm* is **not as accurate as** *canu*, but it is suitable for **fast initial analysis** and the quality of the assembly can be increased with **an additional polishing step**.