

Smartphone apps for portable DNA sequencing and data analysis

Ben Zwart, Carson Zhang, Huan Yi Shen, Julia Danieli, Kamalesh Reddy, Prakriti Chhabra

January 8, 2021

Abstract

The following entails recent advancements in portable DNA sequencing and aims to answer some questions about the same.

Introduction

First, we will look at some basic biology concepts and answer some questions to better understand DNA sequencing:

What is DNA sequencing?

DNA determines what proteins are synthesized and so, they play an important role in determining how an organism behaves and looks like (determines its characteristics and also characteristics of its offspring). How are these instructions stored? They are stored as a long sequence of 4 chemicals/letters (Adenine \rightarrow A, Thymine \rightarrow T, Cytosine \rightarrow C, and Guanine \rightarrow G). DNA sequencing is the process of determining the order of these 4 chemicals, given a strand of DNA.

How is DNA sequenced? ("Shotgun Sequencing")

DNA is a very long sequence of A's, T's, C's, and G's. So, the strand is first broken (how?) into smaller strands. The smaller strands are then sequenced individually (If you can sequence a shorter strand why not a longer one; and how long of a strand CAN you sequence?). Then, we create thousands of copies of each small strand (to easily identify them?). These small strands are then individually sequenced using enzymes and the four molecules; A, C, G, and T (look into mechanism). Now, somehow using the aforementioned enzymes and molecules we attain the sequences of these individual strands, these small sequences are stitched together using computers (is this where data analysis plays a major role?) to give us *the sequence of the entire genome*.

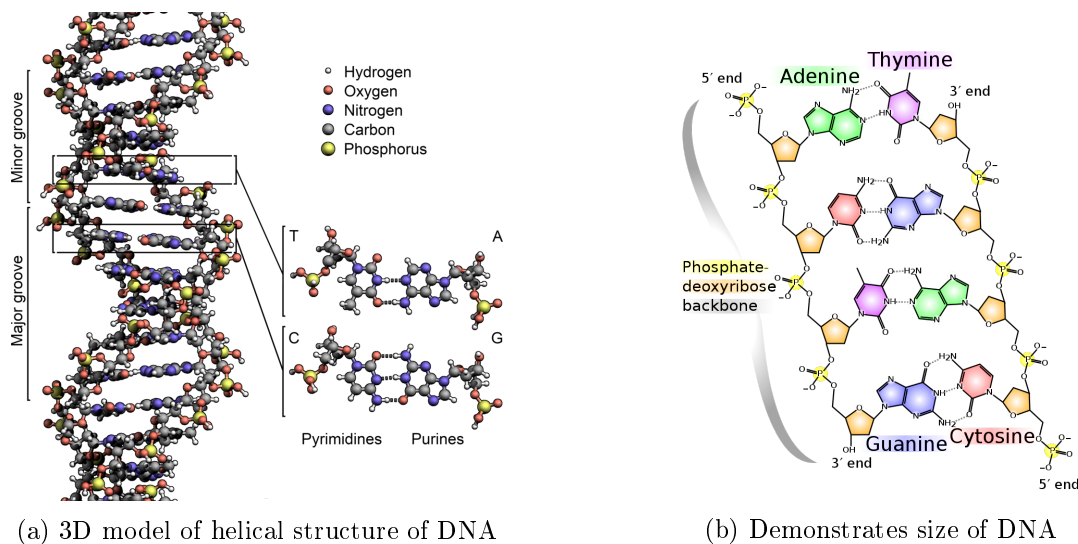


Figure 1: Structure of double-stranded DNA for reference

Portable DNA Sequencing

There are a variety of sequencing methods, we will be looking into "Nanopore Sequencing". This method is, by far, the most portable. There are other sequencers (Illumina HiSeq 2500 sequencer and BGI MGISEQ-2000RS sequencer for instance) that use fairly small machines (possibility of miniturization?) and other sequencing methods, but our current state of technology does not allow miniturization (yet!).

Nanopore Sequencing

Since we have established that: right now, nanopore sequencing is the most portable form of sequencing; we will first look at what nanopore sequencing is. It is best explained by (as we will see later) the current pioneer in nanopore technology, Oxford Nanopore Technologies:

“Nanopore sequencing is a unique, scalable technology that enables direct, real-time analysis of long DNA or RNA fragments. It works by monitoring changes to an electrical current as nucleic acids are passed through a protein nanopore. The resulting signal is decoded to provide the specific DNA or RNA sequence.” [7]

So, nanopore sequencing involves passing a single-stranded DNA molecule through a "protein nanopore" and measuring changes in electric current to determine the base pair (by "base calling" using *Guppy*^[17]). This makes sense because we know that atoms are around 10^{-10}m in diameter, so for a single strand of DNA (which is around 8-10 atoms wide) we have a width of $\approx 10^{-9}\text{m} = 1\text{nm}$; hence nanopore! The devices used are surprisingly small, they are **really** small. [8]



(a) Fongle



(b) MinION

Figure 2: The Fongle and MinION can be used for portable analysis



(a) GridION



(b) PromethION

Figure 3: GridION and PromethION can be used for flexible, high-throughput benchtop sequencing



(a) SmidgION



(b) Plongle

Figure 4: SmidgION and Plongle are under development and are excitingly small

Smartphone apps

DNA sequencing	DNA data analysis	
MinKNOW	GenoPo	iGenomics

DNA sequencing

What mobile apps are available in this space?

Oxford Nanopore Technologies has their very own app: **MinKNOW**. As far as nanopore DNA sequencing (portable) is concerned, there are no other prominent apps (explained later).

When were they developed?

MinKNOW is a general application that existed for a while (since at least 2015) but the mobile app was formally released on **November 25th, 2020**.^[9]

What features do they offer?

As stated in ONT's website:

“It carries out several core tasks, including:

- **Data acquisition**
- **Real-time analysis and feedback**
- **Local basecalling**
- **Data streaming**
- **Progressive unblocking**
 - Increased data yield of a flow cell
 - Active voltage control → improved signal consistency during a run.
- whilst providing device control including **selecting the run parameters, sample identification and tracking**, and ensuring that the platform chemistry is performing correctly to run the samples.”^{[9] [10]}

It also enables users to **run group experiments** and **view/export metrics to a PDF**. Essentially, the app helps manage their devices, we shall also discuss complementary apps, iGenomics and GenoPo.

Which apps are used most frequently by consumers?

MinKNOW basically dominates every other app (if any) as far as nanopore sequencing goes. This is mostly due to the availability of their own native devices, enabling them to modify their app as per device upgrades. Consumers are likely to use MinKNOW if they own a, say, MinION. Similar to how Android users likely use Google Chrome. Besides, sources indicate that they have patented the technology \implies nanopore sequencing is intellectual property of ONT (for now?).^[11]

Are there any advantages and limitations of these tools?

Advantages:

- With new device rollouts, ONT can simply update MinKNOW to accommodate new devices.

Limitations:

- As of now, ONT does not compete with anyone when it comes to nanopore sequencing (a little with Illumina^[12]). This indicates that there is not as much a desire to release new technology.

DNA data analysis

What mobile apps are available in this space?

There are two apps that ONT recognizes ^{[13][14]} as complements to its devices and MinKNOW: **iGenomics** and **GenoPo**. They were developed without ONT's direct involvement.

When were they developed?

- *iGenomics* was developed by Aspyn Palatnick, Bin Zhou, Elodie Ghedin, and Michael Schatz ^[15] and was published in **April 2020** (As shown in Apple's App Store).
- *GenoPo* (formerly known as F5N ^[16]) was developed by students from the *University of Peradiniya* ^[14], it was released on **August 12th, 2020** (As shown in Google's Play Store).

What features do they offer?

iGenomics:

- **Alignment reading** (?)
- **Coverage profile** (?)
- **Call variants** (?)
- **Visualize results on any IOS device** (?)
- **AirDrop sequencing data from one device to another** (?)

GenoPo:

- **Real-world applications**: GenoPo analysis happens very quickly, averaging about 27 minutes per SARS-CoV-2 sample with 5-10 mutations detected, and a consensus sequence generated. This is a small fraction of the normal turnaround time for your average SARS-CoV-2 nanopore sequencing.
- **Compatibility**: Designed for compatibility with many bioinformatics analyses including sequence alignment, variant detection and DNA methylation profiling.
- **Real time sequence analysis**: Allows real-time methylation calling on nanopore sequencing datasets which can also accommodate both small and large genomes.

- **Control:** Features auto configured pipelines available for Arctic C Pipeline and Consensus Generation Pipeline, but also allows for manual configuration of pipelines; including a terminal environment allowing users to enter command line arguments. The terminal mode allows users to run any sub tool as well.

Which apps are used most frequently by consumers?

- Apple developers can choose to hide app statistics, which is unfortunately the case for iGenomics (messed developer (Aspyn), waiting for an unlikely response).
- GenoPo has 1000+ installs on the Google Play Store (as shown in the Play Store).

Are there any advantages and limitations of these tools?

iGenomics:

Advantages:

- AirDrop - DNA analysis in remote location, with/without internet
- User Reviews - on the Apple store, the reviews are mostly positive (limited samples)
- “The novelty of this application is not in the algorithms used but rather how they have been implemented in a mobile environment.”

GenoPo:

Advantages:

- An advantage of GenoPo is that it is able to methylate genomes of all sizes, and has been used in similar research to that we are interested in with SARS-CoV-2.
- It contains manual pipeline configuration and a terminal environment which makes it very easily customizable to each Bioinformatician’s needs. GenoPo is also very quickly able to complete it’s entire workflow.

Limitations:

- Although GenoPo has been popular on the Google Play store, it is unfortunately unavailable for iOS. It is safe to assume GenoPo would be used by more people if it were available for iOS devices as well.
- Limited reviews of the app revealed significant flaws in the user-interface design. 11 product reviews suggest that the UI is difficult to navigate and interpret.

DNA data analysis - comprehensive (not portable)

What mobile apps are available in this space?

Search results for "Dna analysis", "Dna sequencing", and "Genomics":

(Legend: **Active and Relevant**,
Active /Inactive and Interesting/Buggy,
 Gimmick/Irrelevant)

iOS

Apps	Date	Installs
23andMe	2020	at least 10997
Ancestry	2020	at least 8067
MyHeritage	2020	at least 1642
AncestryDNA	2020	at least 376
iGenomics	2020	n
Seq Analysis	2020	n
Genewall	2020	n
DNA2App	2015	n
DNAApp	2014	n
MySequence	2013	at least 2
Dynamic News Analysis	N/A	N/A
Data Navigation and Analysis	N/A	N/A
ERHA.DNA	N/A	N/A
My Skineaud	N/A	N/A

Android

Apps	Date	Installs
Ancestry	2011	5000000+
MyHeritage	2011	5000000+
23andMe	2018	1000000+
Genomapp	2015	50000+
Genetica	2020	1000+
DNA test	2020	1000+
GenoPo	2020	1000+
Face2Gene	2016	50000+
CircleDNA	2019	10000+
dietgene	2020	10000+
DNA+	2019	1000+
My DNA	N/A	N/A

References

- [1] Genetic Engineering Will Change Everything Forever - CRISPR, Kurzgesagt - In a Nutshell, <https://www.youtube.com/watch?v=jAhjPd4uNFY>
- [2] How to sequence the human genome - Mark J. Kiel, TED-Ed, <https://www.youtube.com/watch?v=MvuYATH7Y74>
- [3] DNA Sequencing - 3D, yourgenome, <https://www.youtube.com/watch?v=ONGdehkB8jU>
- [4] DNA, *Wikipedia*, <https://en.wikipedia.org/wiki/DNA>
- [5] DNA sequencing, *Wikipedia*, https://en.wikipedia.org/wiki/DNA_sequencing
- [6] Shotgun sequencing, *Wikipedia*, https://en.wikipedia.org/wiki/Shotgun_sequencing

- [7] Nanopore DNA sequencing, *Oxford Nanopore Technologies*, <https://nanoporetech.com/applications/dna-nanopore-sequencing>
- [8] Portable DNA sequencing devices, *Oxford Nanopore Technologies*, <https://nanoporetech.com/products>

- [9] Introducing the new MinKNOW App, *Oxford Nanopore Technologies*, <https://nanoporetech.com/about-us/news/introducing-new-minknow-app>
- [10] MinKNOW 2.0, a software update for nanopore sequencing, *Oxford Nanopore Technologies*, <https://www.youtube.com/watch?v=JuBIxHaONrQ>
- [11] Intellectual property, *Oxford Nanopore Technologies*, <https://nanoporetech.com/about-us/intellectual-property>
- [12] Illumina and Oxford Nanopore Settle Patent Infringement Lawsuit, *genomeweb*, <https://www.genomeweb.com/sequencing/illumina-and-oxford-nanopore-settle-patent-infringement-lawsuit>
- [13] iGenomics: Comprehensive DNA sequence analysis on your Smartphone, *Oxford Nanopore Technologies*, <https://academic.oup.com/gigascience/article/9/12/giaa138/6025149>
- [14] Genopo: a nanopore sequencing analysis toolkit for portable Android devices, *Oxford Nanopore Technologies*, <https://nanoporetech.com/resource-centre/genopo-nanopore-sequencing-analysis-toolkit-portable-android-devices>
- [15] iGenomics, *Schatz Lab*, <http://schatz-lab.org/iGenomics/>
- [16] f5n, *GitHub*, <https://github.com/SanojPunchihewa/f5n>
- [17] Basecalling using Guppy, *Guppy*, https://timkahlke.github.io/LongRead_tutorials/BS_G.html