Homework 11

Chandrima Bhattacharya

30 March 2019

Question 1

The paper I selected was from an open-peer reviewed journal F1000 research. The paper is based on forming a consortium for setting a world-wide standards for analysing of data using MinIon. It also looks at reproducibility of MinIon along with describing the technology. Also, I went through another paper on Oxford Nanopore based on MinIon technologies.

`Ip, Camilla LC, et al. "MinION Analysis and Reference Consortium: Phase 1 data release and analysis." F`

`Lu, Hengyun, Francesca Giordano, and Zemin Ning. "Oxford Nanopore MinION sequencing and genome assembly`

a

Nanopore sequencing is a third-gen sequencing technique where a single DNA/RNA molecules traverse a nanosized pore when an electric field is applied, and the subsequent change in field potential corresponds to specific base translocation, permitting sequencing by translocation without the need for amplification by PCR.

b

- 1a: DNA double helix [The nanopore processes the length of DNA or RNA presented to it. The user can control this through the library preparation protocol utilised]
- 1b: DNA base
- 2: Unwinding Enzyme or Motor Enzyme [An enzyme motor controls the translocation of the DNA or RNA strand through the nanopore. Once the DNA or RNA has passed through, the motor protein detaches and the nanopore is ready to accept the next fragment.]
- 3: Membrane (Lipid bilayer) [An electrically resistant membrane means all current must pass through the nanopore, ensuring a clean signal]
- 4: Protein Pore [DNA or RNA passes through a nano-scale hole. The fluctuations in current during translocation are used to determine the DNA or RNA sequence.]

c

Characteristics of protein pore (refered as porin) includes:
* The pore is only large enough for a single strand to fit through, one nucleotide at a time * Uniform Pore strructure
* Can control the precise control of sample translocation through pore channels
* Can detect individual nucleotides in samples can be facilitated by unique proteins from a variety of organism types
* Can tolerate high voltage range Common examples includes alpha-heamolysin and MspA porin.

d

The ionic current passes through nanopores and measures the changes in current as biological molecules pass through the nanopore or near it as a voltage is applied across the membrane and measured through the sensor. The membrane is resistant to the electric current, meaning the only way for current to flow across the membrane is to travel through the pore.

e

SecA is an example of narurally occuring motor protein which can be used for the following. SecA protein is a cell membrane associated subunit of the eubacterial Sec or Type II secretory pathway, a system which is responsible for the secretion of proteins through the cell membrane.
Also, as motor enzyme is responsible for translocation, translocase would be a great choice of proteins. Helicase would also be a nice option as it unwinds the DNA.

f

The adapters used in nanopore sequence includes:
* The leader adapter which is denoted as the "Y adapter" as it has a "Y" shaped structure. Sequencing begins at the single-stranded 5' end of the Y adapter, followed by the "template" strand, then the HP adapter, and the "complementary" strand.
* The adapter with hairpin structure is called the "HP adapter", which attaches to both strands of dsDNA, thus allowing both to be sequenced.
* There is also a tether which is attached to the hairpin end of the sequence which brings the adapters closer thus increasing the DNA capture rate.
The adapters increase the probability that the complement strand will immediately follow the template strand

Question 2

Advantages compared to Illumina-seq techniques:
* Ability to analyze much longer reads as there is no need to significantly fragment the DNA and incorporate bases into the strands over time.
* Allows for detection of other molecules such as RNA or possible epigenetic markers due to measuring current change not base incorporation.

Disadvantages compared to Illumina-seq techniques:
* Less high-throughput compared to other sequencing set-ups such as Illumina.
* Data is noisier due to electric signal on single molecules being detected instead of fluorescence. The error rate can be as high as 15% compared to .1% error in Illumina-based NGS techniques. (Even though the current rate is approx 8%)

Question 3

a

```
spack load fastqc
fastqc /home/frd2007/ANGSD_2019/nanopore_seq/guppy_out/*.fastq.gz -o angsd_hw/hw11/
ls angsd_hw/hw11/
```

```
scp chb4004@pascal.med.cornell.edu:angsd_hw/hw11/*.html SingleCell/
```

b

- All per base quality sequence, per base sequence content, per sequence GC content has failed!
- Many of them have overpresented sequences.
  The following is expected as I am running it as default Illumina way, instead for nanopore option. I tried running it using nanopore parameter, which gave error in file formating. Hence, probably adapters, etc., are probably detected incorrectly.