Analysing the genomic complexity at the single cell level using Oxford Nanopore Technology

BY: VARSHA BALRAJ (z5113220)

Supervisors: A/Prof. Fabio Luciani ,
Dr Preston Leung

Thesis A Presentation

Outline

- Background
  - Genome complexity: isoforms
  - Sequencing technologies
  - Single Cell Analysis
  - Isoforms of CD45 gene
- Project Proposal
  - Aim & Hypothesis
  - Methods
  - Validation
  - Timeline

Genomic Complexity

BACKGROUND

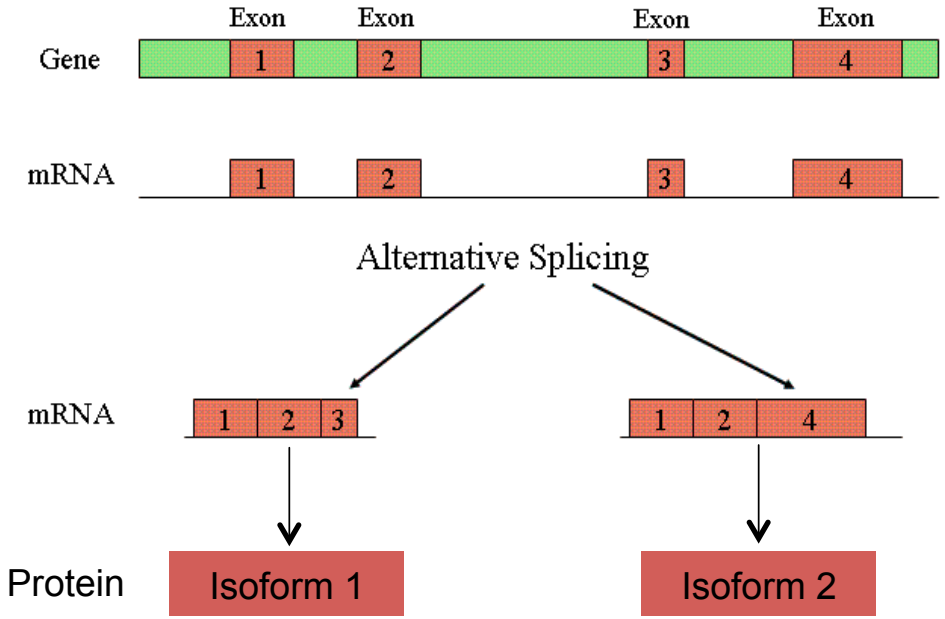
Complexity of Genomes

The genome can be regarded as a complex structure of coding sequence (exons) separated by noncoding sequences (introns)

https://designmatrix.wordpress.com/2010/03/07/alternative-splicing/

Alternative Splicing

- A single gene can code for multiple proteins via different isoforms
- Involves the inclusion and exclusion of particular exons

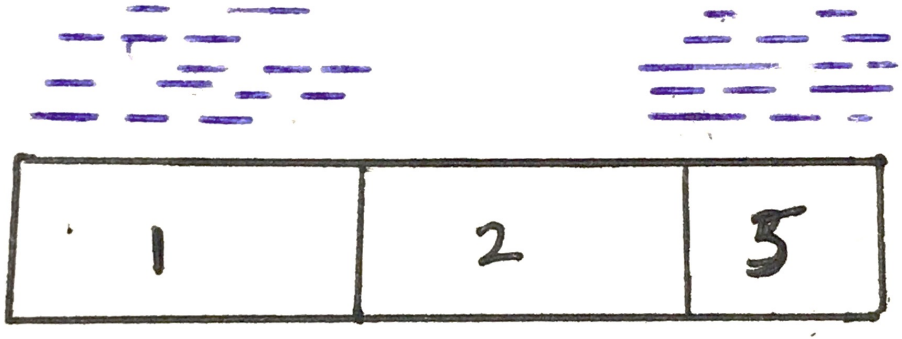


https://designmatrix.wordpress.com/2010/03/07/alternative-splicing/

Splicing and isoforms

GENE EXPRESSION IS CHARACTERISED BY MULTIPLE ISOFORMS FOR EACH GENE

- Isoform detection is difficult to achieve due to multiple exons undergoing splicing events
- Current methods are laborious and based on RNAseq analysis of bulk populations
  - This approach add noise and different cells have different isoforms



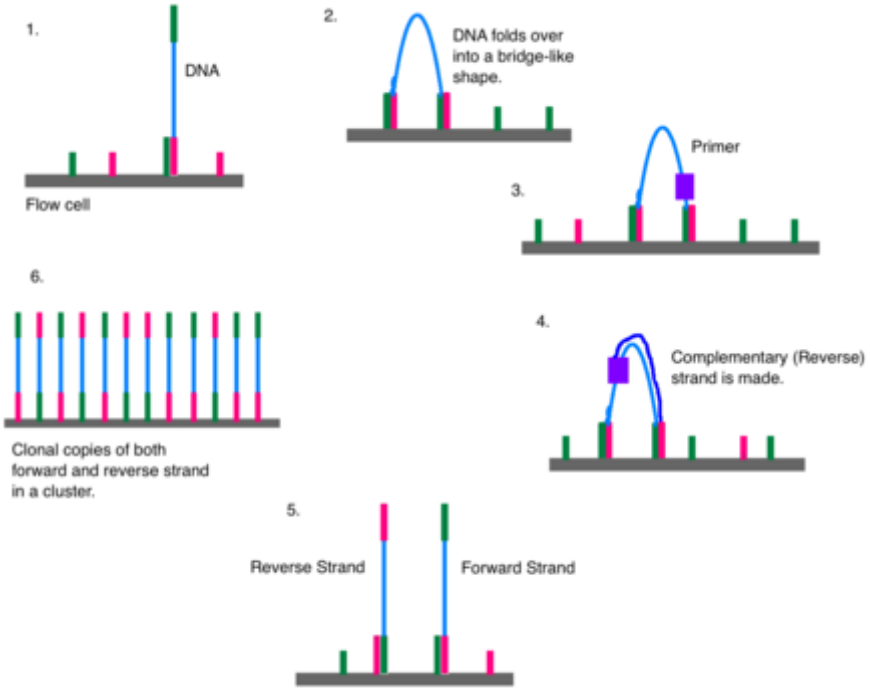
Sequencing Technology

BACKGROUND

Illumina Sequencing

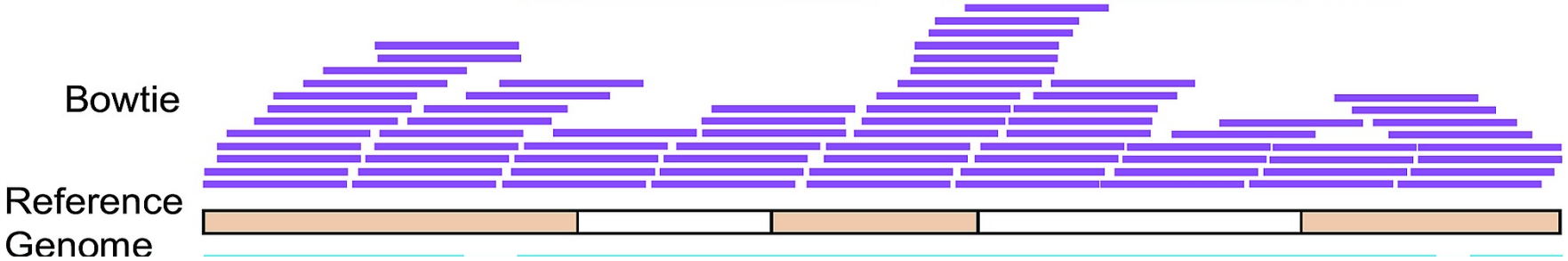
NEXT GEN SEQUENCING

- Also known as Next- Gen Sequencing or High throughput method of sequencing
- Most widely used method is Illumina sequencing
- A synthesis approach is used for this method



https://en.wikipedia.org/wiki/Illumina_dye_sequencing

Problems with Illumina



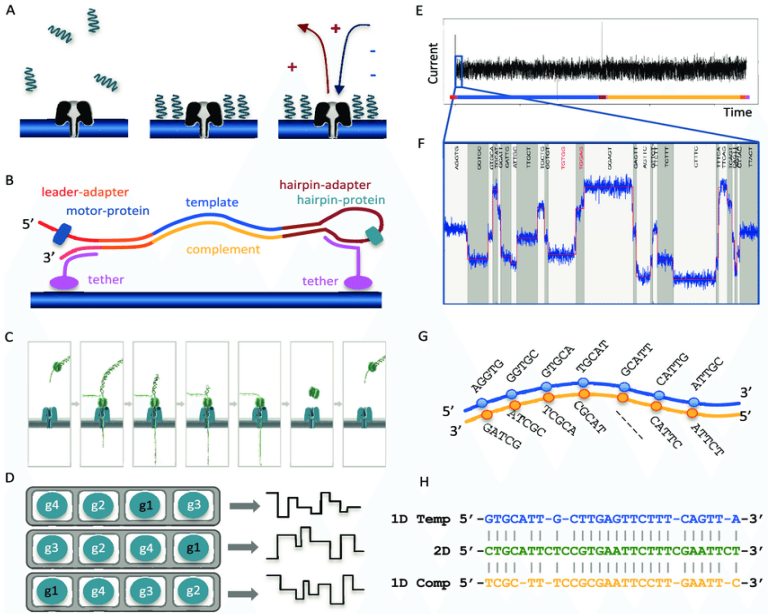
https://plos.figshare.com/articles/Comparison_of_Illumina_HiSeq_to_Oxford_Nanopore_MinION_read_data_/4980248

- The relatively short reads made genome assembly more difficult.
- Does not facilitate the assembly of repetitive structures of interest that extend beyond the maximum read length generated

New Technology

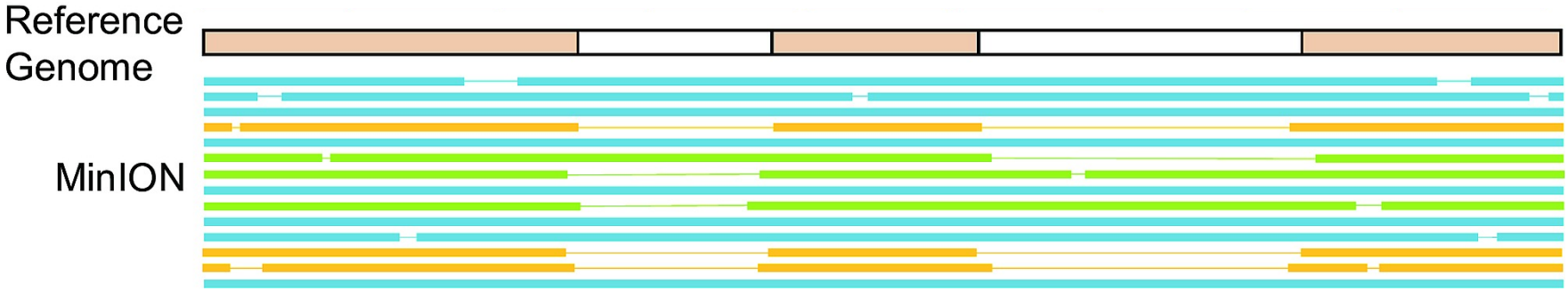
THIRD GEN SEQUENCING

- Oxford Nanopore technology (ONT) will be the sequencing method used for this project
- Uses electrophoresis to transport an unknown sample through an orifice
- The composition of the sample DNA can be identified by the magnitude of the electric current density across a nanopore surface



MinION Analysis and Reference Consortium: Phase 1 data release and analysis [version 1; referees: 2 approved]. F1000 Research. 4. 10.12688/f1000research.7201.1.

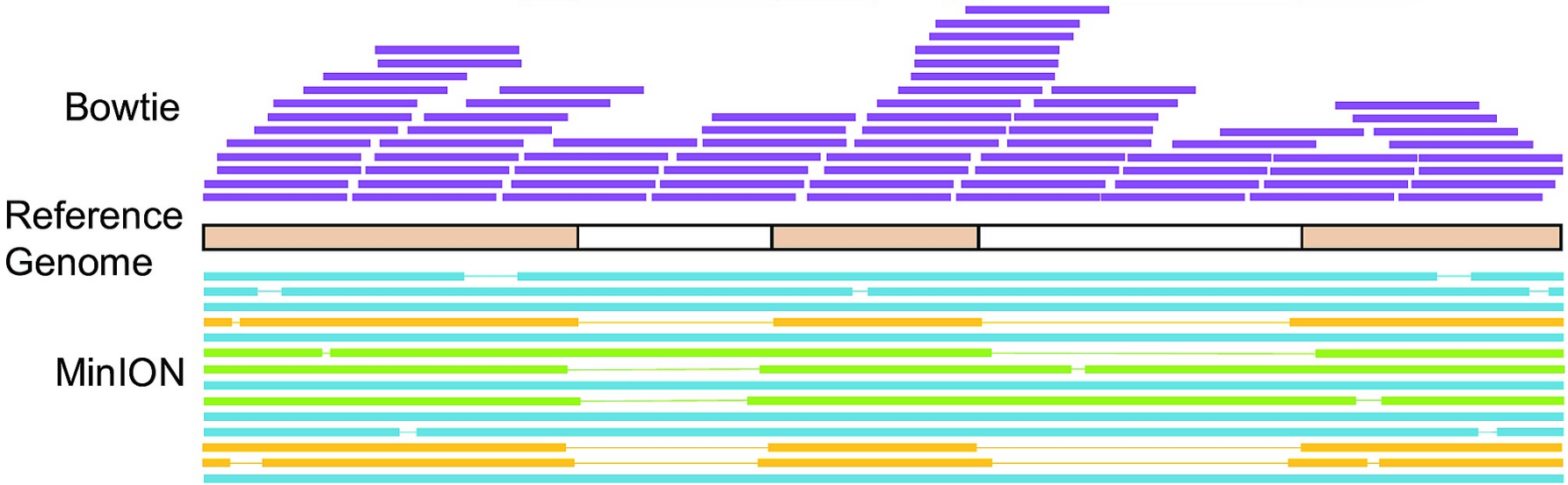
So why choose ONT?



https://plos.figshare.com/articles/Comparison_of_Illumina_HiSeq_to_Oxford_Nanopore_MinION_read_data_/4980248

- Overcomes the limitations of Illumina by generating sequencing median read lengths of 8-10kb and as long as 100kb.
- Low-med sequencing cost and easy sample preparation without the need PCR amplification

Overview



https://plos.figshare.com/articles/Comparison_of_Illumina_HiSeq_to_Oxford_Nanopore_MinION_read_data_/4980248

Single Cell Sequencing

BACKGROUND

"Bulk" Sequencing

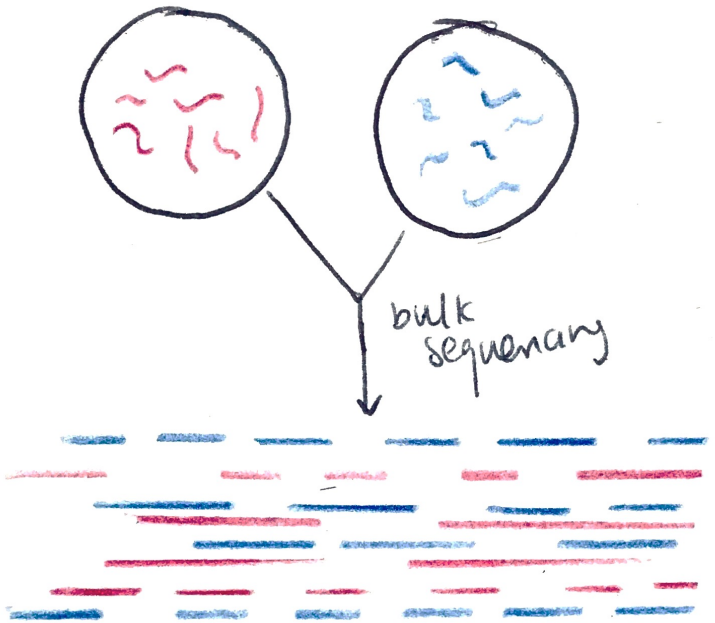
- Sequencing has typically performed in "bulk" RNA-Seq and the data represents an average of gene expression patterns across thousands to millions of cells
- Isoform detection can be possible in "bulk".

(https://pdfs.semanticscholar.org/4e19/bd479da2e4781f2eca67a2ef489ce77c385f.pdf)

- How do we know which isoform corresponds to which cell?

Problems with "Bulk" Sequencing

- It can be hard to identify which sequence is from what cell
- Also creates lots of 'noise' which affects the accuracy of the data
- Unable to reveal the cellular heterogeneity that drives complexity.

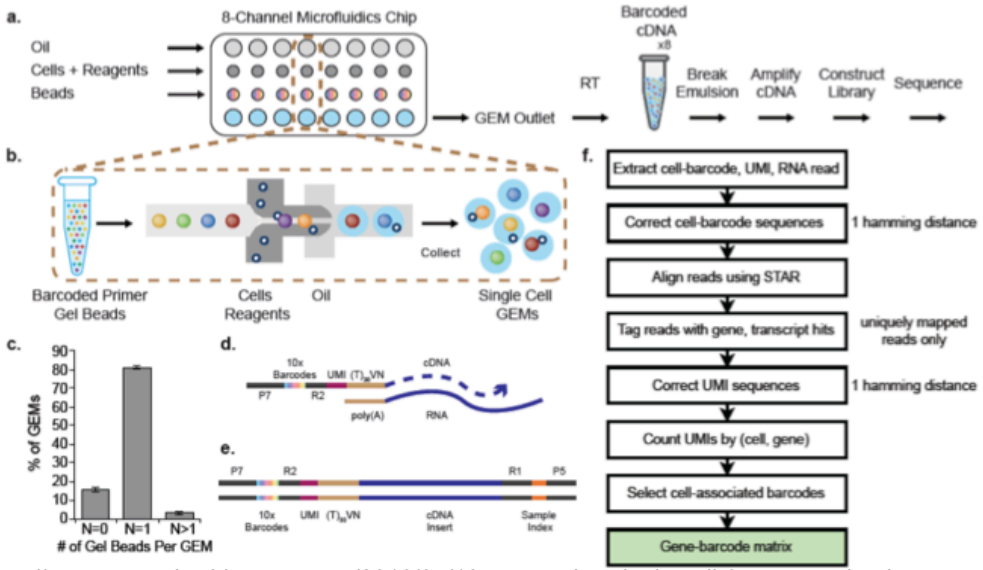


Single Cell Sequencing

- Explore the complex gene expression profile of individual cells in complex tissues and understand cellular subpopulation responses to environmental cues.
- High resolution analysis enables discovery of cellular differences usually masked by bulk sampling methods
- Robust transcriptome analysis down to single-cell input levels for high-quality samples

Single Cell Sequencing

Integrated protocol proceeds directly from whole cells and preserves integrity through barcoding



http://core-genomics.blogspot.com/2016/07/10x-genomics-single-cell-3mrna-seq.html

Single Cell Sequencing & Long Reads

- Previously single cell sequencing has only been performed along with Illumina short reads
- The advancement of long reads and single cell sequencing together can provide a data with better validity and at higher-resolution

CD45 Gene Complexity

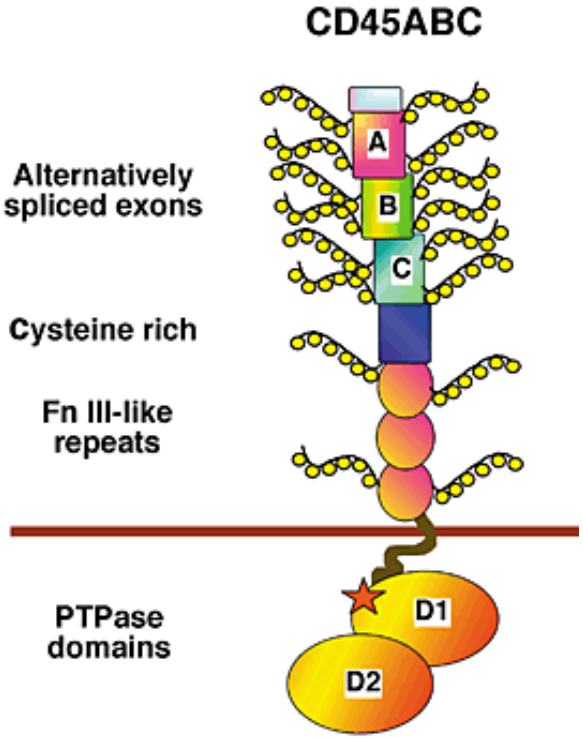
BACKGROUND

CD45

- Codes for the PTPRC protein
- Receptor-linked protein plays a crucial role in the function of leucocytes
- Regulates a variety of cellular processes including cell growth, differentiation, mitotic cycle, and oncogenic transformation

CD45

- On T cells the extracellular domain of CD45 is expressed in several different isoforms
- Essential for the activation of T cells via the TCR, and that different CD45 isoforms display a different ability to support T cell activation.



https://designmatrix.wordpress.com/2010/03/07/alternative-splicing/

Project Proposal

Problem?

- There is no research to study isoforms occurring in T cell data.
- Previous technology made study of isoforms almost impossible due to the estimation involved
- Currently there are only 17 known isoforms of CD45, now with ONT possibility of uncovering more?

Aim

To study single cell data with Oxford Nanopore sequencing to understand the different isoforms ofCD45 in T cells

Hypothesis

Analysis of single cells data with long reads can be used to identify multiple isoforms present in the CD45 gene.

Data

- Nanopore sequences form scRNAseq of T cells from healthy donors N=2
- Clinical samples from Chimeric Antigen receptor t cells (CAR T cells) engineered to treat leukemia. N=4

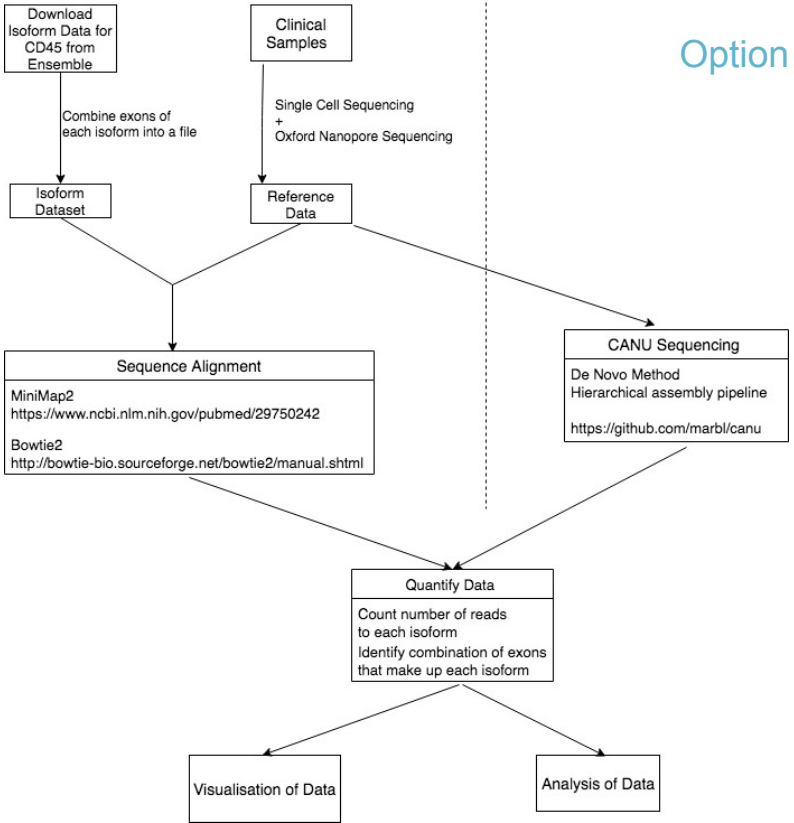
Sequence Alignment

- Minimap2 & Bowtie will be used as they can align long ONT reads
- Sequence alignment program that aligns DNA or mRNA sequences against a large database
- Used to find overlaps between long reads, splice-aware alignment and assembly-assembly alignment

Method

Option 1

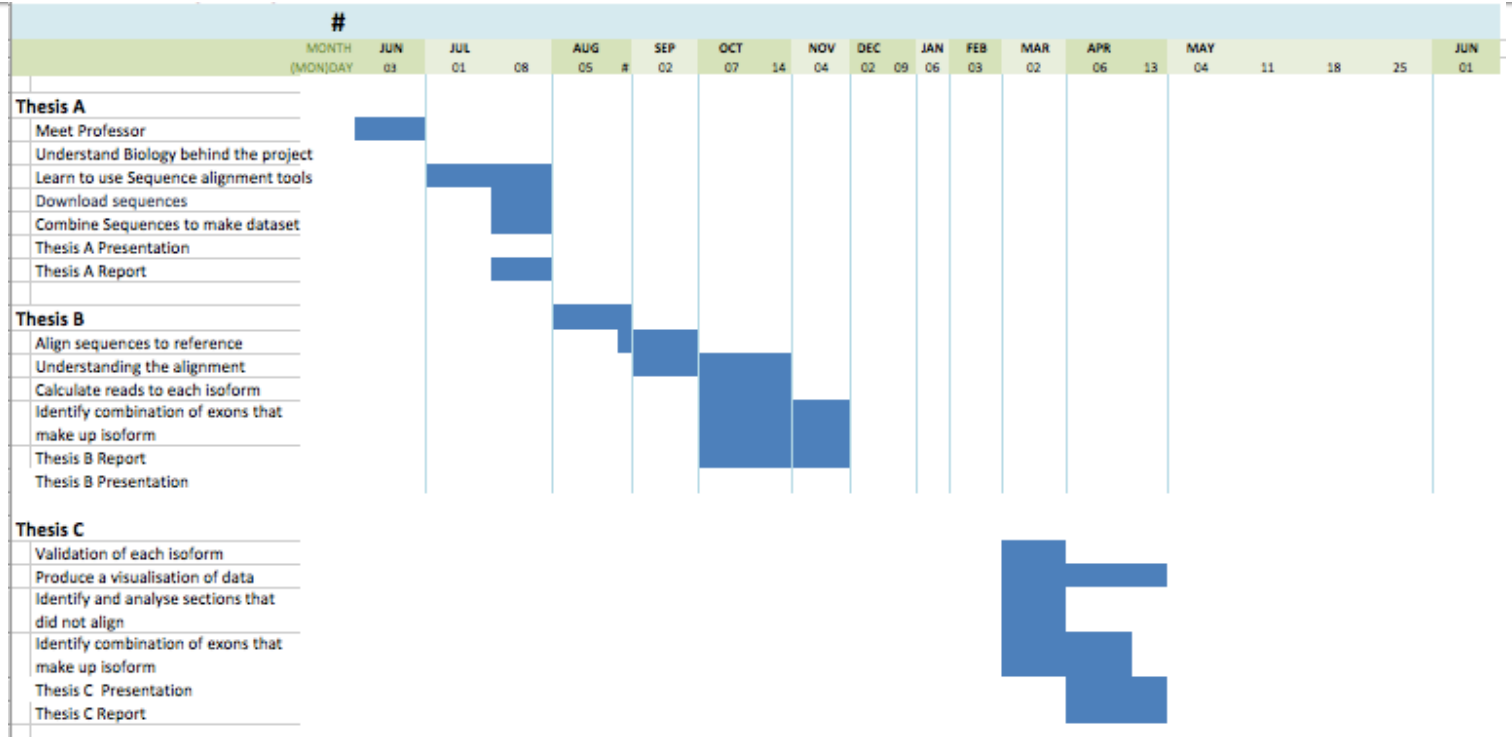
Option 2



Validation

- T cells carrying CD45RA or RO isoforms will be used as ground truth
- ONT reads will be aligned against those two isoforms and compared to the result of the bioinformatics pipeline that will be generated in this study

Project Timeline



Current progress

- Understand Biology
- Learn to use sequence alignment tools
- Download and combine a data set of sequences

Thesis B

- Align sequences to reference
- Understanding the alignment
- Calculate reads to each isoform
- Identify combination of exons that make up each isoform

Thesis C

- Validation of each isoform
- Produce a visualisation of data
- Identify and analyse sections that did not align

References

In-text: (Longdom.org, 2019)
Your Bibliography: Longdom.org. (2019). [online] Available at: https://www.longdom.org/open-access/generations-of-sequencing-technologies-from-first-to-next-generation-0974-8369-1000395.pdf

In-text: (Illumina.com, 2019)
Your Bibliography: Illumina.com. (2019). [online] Available at: https://www.illumina.com/documents/products/illumina_sequencing_introduction.pdf [Accessed 20 Jul. 2019].

In-text: (Biorxiv.org, 2019)
Your Bibliography: Biorxiv.org. (2019). [online] Available at: https://www.biorxiv.org/content/biorxiv/early/2019/01/28/530824.full.pdf [Accessed 20 Jul. 2019].

In-text: (Anon, 2019)
Your Bibliography: Anon, (2019). [online] Available at: https://www.sciencedirect.com/topics/neuroscience/nanopore-sequencing https://bitesizebio.com/36592/introduction-nanopore-sequencing/ [Accessed 20 Jul. 2019].

Your Bibliography: Anon, (2019). [online] Available at: https://www.sciencedirect.com/topics/neuroscience/nanopore-sequencing https://bitesizebio.com/36592/introduction-nanopore-sequencing/ [Accessed 20 Jul. 2019].

DNA SEQUENCING FACT SHEET
In-text: (Genome.gov, 2019)
Your Bibliography: Genome.gov. (2019). *DNA Sequencing Fact Sheet*. [online] Available at: https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Fact-Sheet

SINGLE CELL DNA SEQUENCING | NOVOGENE
In-text: (Novogene, 2019)
Your Bibliography: Novogene. (2019). *Single Cell DNA Sequencing | Novogene*. [online] Available at: https://en.novogene.com/next-generation-sequencing-services/single-cell-sequencing/single-cell-dna-sequencing/

SINGLE-CELL AND LOW-INPUT RNA-SEQ | SINGLE-CELL SEQUENCING BENEFITS
In-text: (Sapac.illumina.com, 2019)
Your Bibliography: Sapac.illumina.com. (2019). *Single-Cell and Low-Input RNA-Seq | Single-cell sequencing benefits*. [online] Available at: https://sapac.illumina.com/techniques/sequencing/rna-sequencing/ultra-low-input-single-cell-rna-seq.html [Accessed 20 Jul. 2019].