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 B Presentation

Outline

- Project Recap
- Thesis B
  - Changes in method
- Current Progress
  - Results
- Timeline

Project Recap

From Thesis A

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

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.

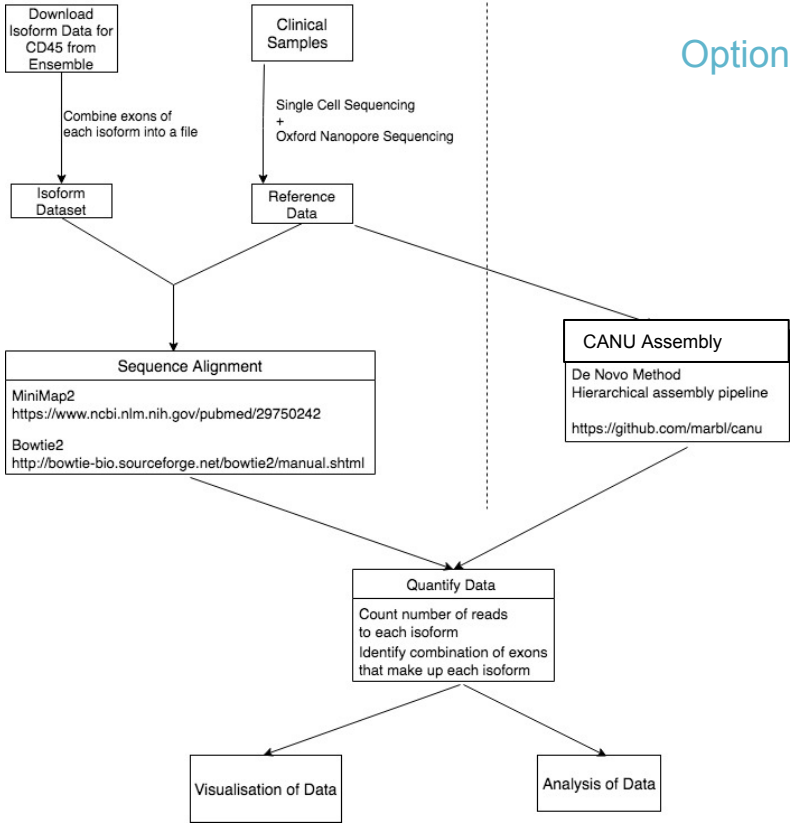
Thesis B Aims

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

Previous method

Option 1

Option 2



Thesis B

Problems with previous method

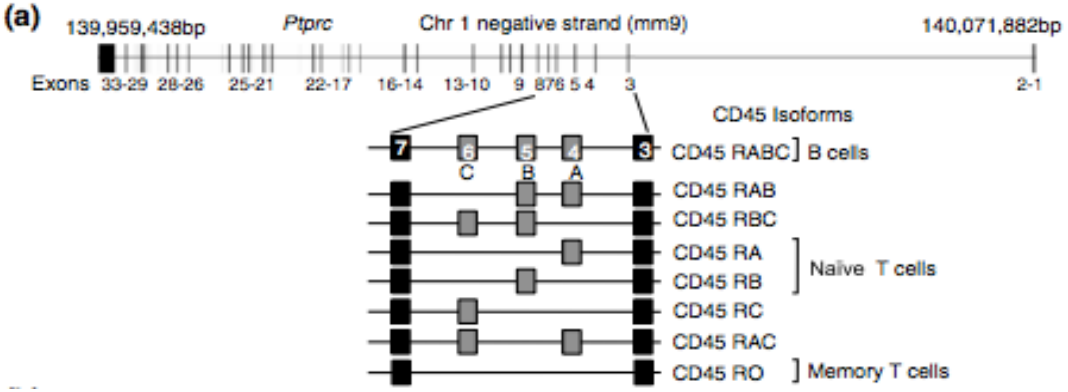
- When downloaded the isoform data, most sequences were not complete
- Not able to form a complete isoform reference dataset

New Approach

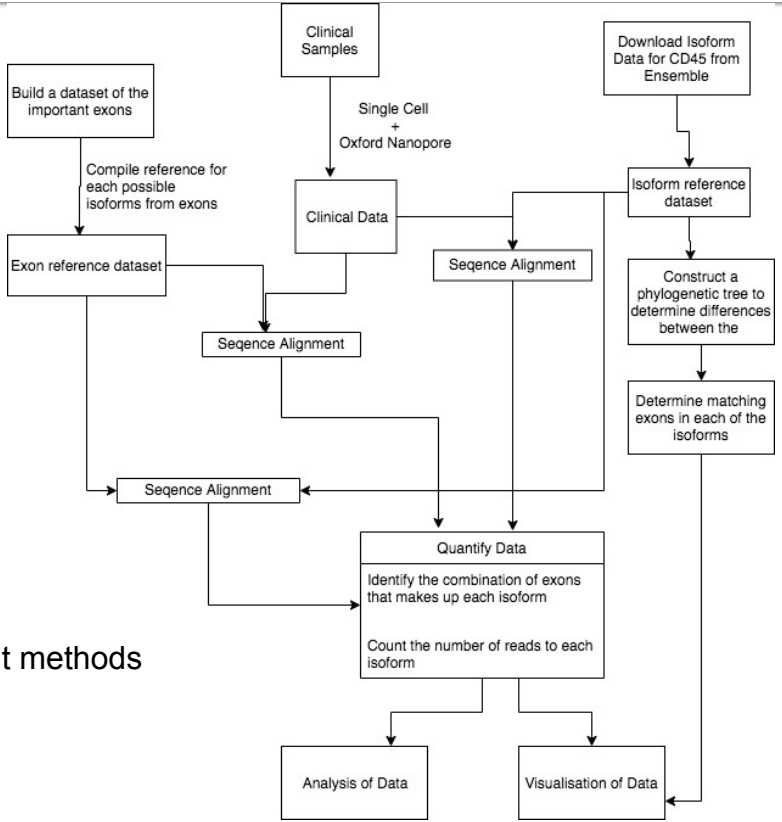
- Focus on the exons themselves in detail
- Notice the differences between the isoforms from the exons
- Build a reference data set from exons only

Exons in more detail

- Based on the paper (*Vicky Cho, Genome Biology 2014, 15:R26*)



Refined Method

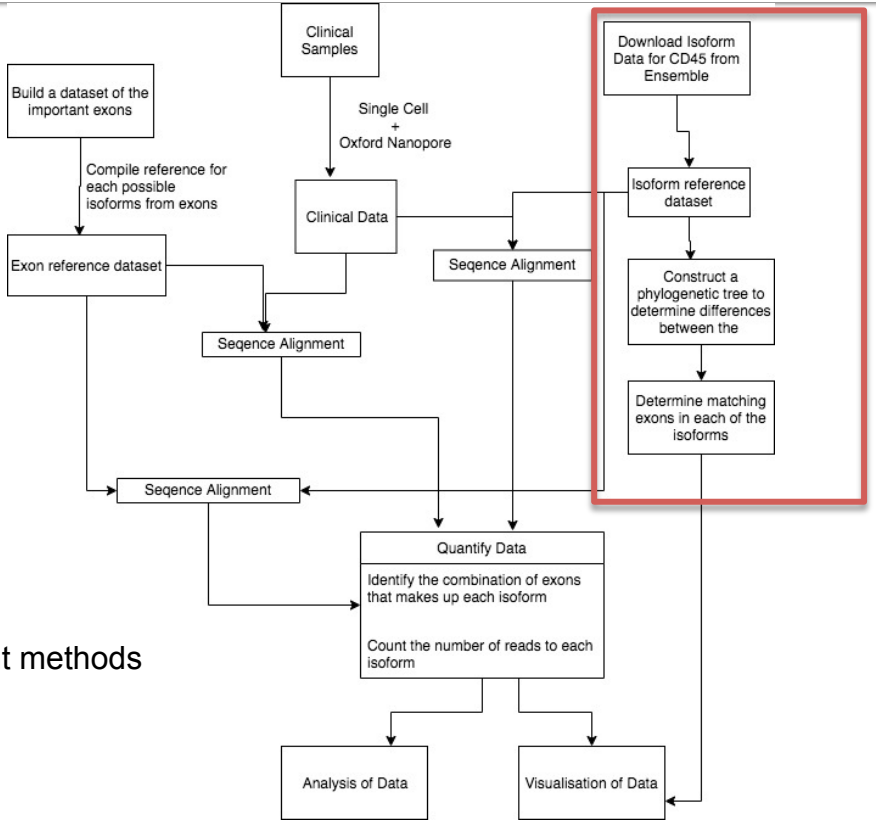


Note: Sequence alignment methods
-Minimap2
-Bowtie2

Current Progress

Results

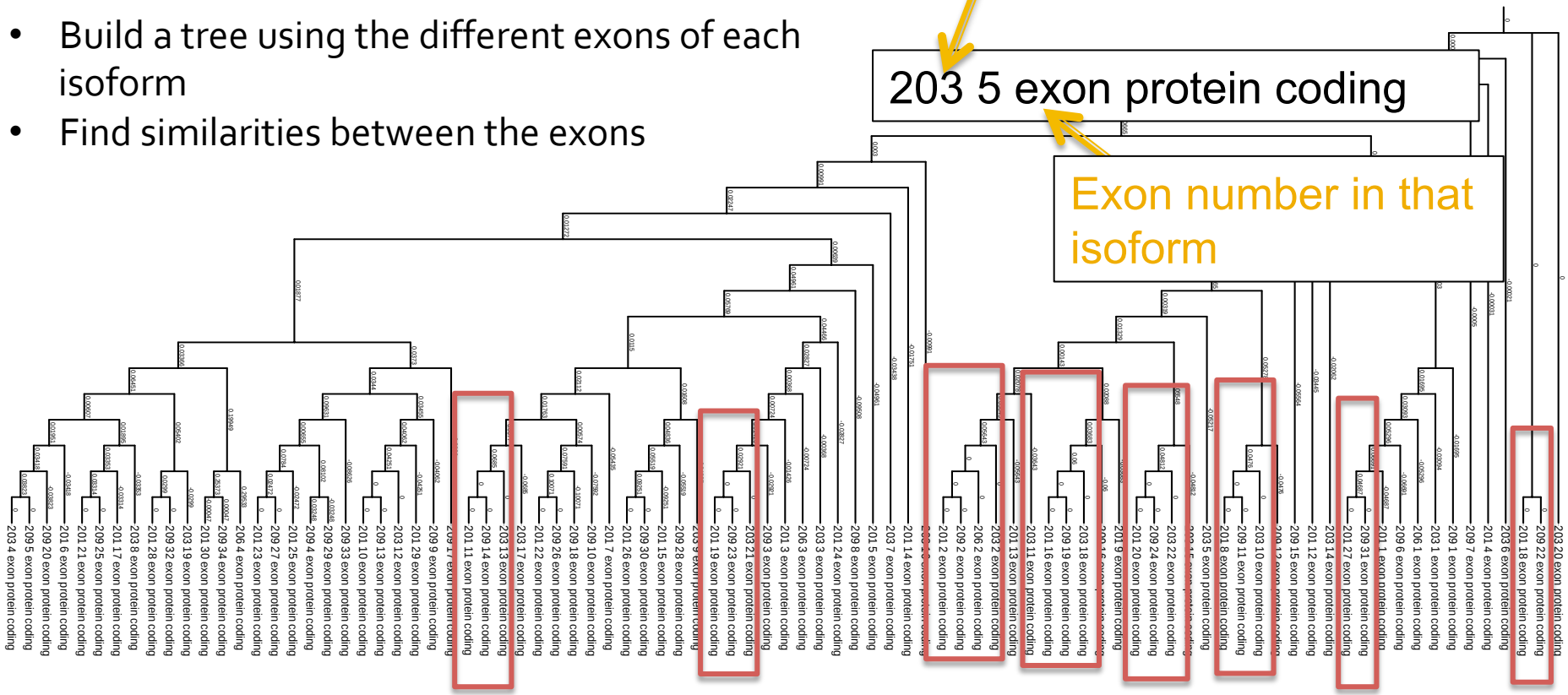
Refined Method



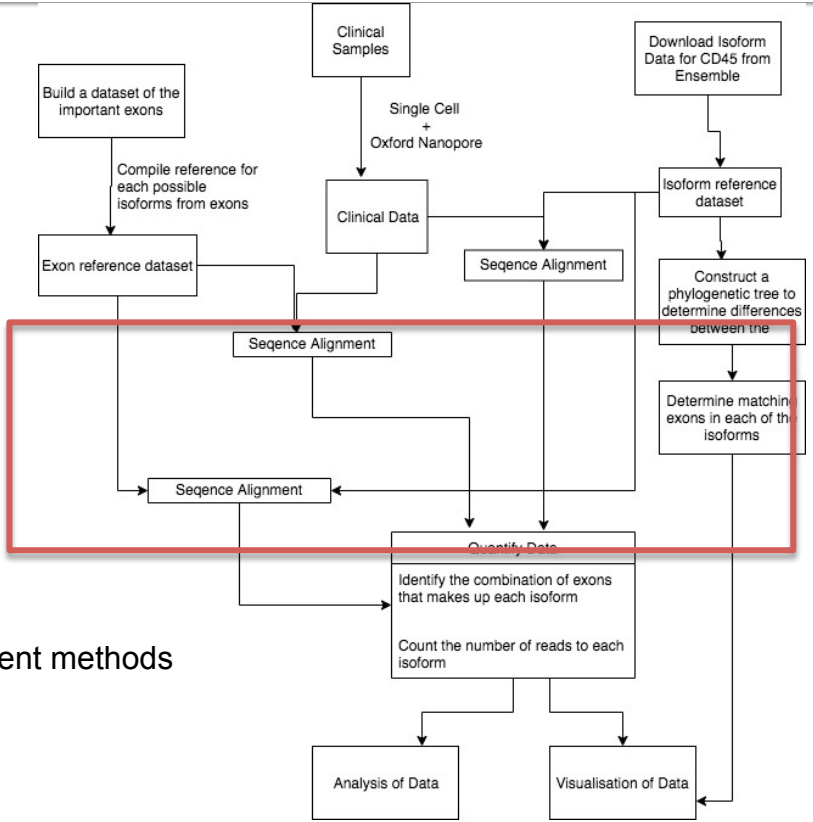
Note: Sequence alignment methods
-Minimap2
-Bowtie2

Phylogenetic Tree

- Build a tree using the different exons of each isoform
- Find similarities between the exons



Refined Method



Note: Sequence alignment methods
-Minimap2
-Bowtie2

Sequence Alignment

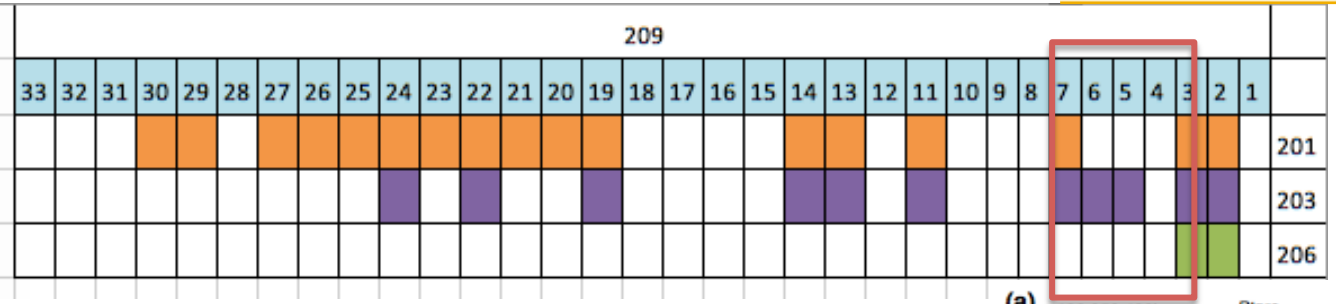
- Align exons to the four isoforms using minimap2

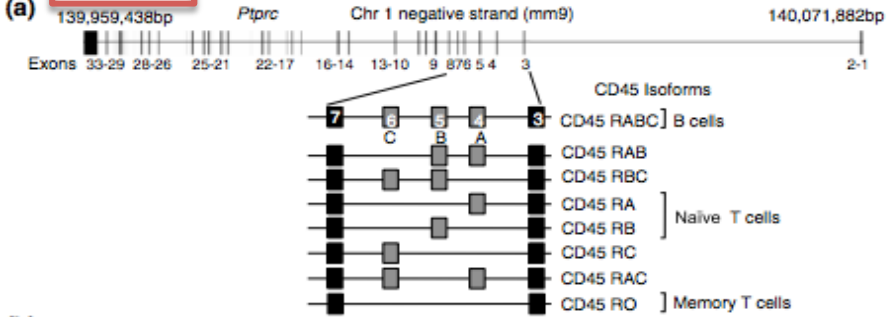
	Exons					
	2	3	4	5	6	7
Isoforms						
201	Y	Y	N	N	N	Y
203	Y	Y	N	Y	Y	Y
206	Y	Y	N	N	N	N
209	Y	Y	Y	Y	Y	Y

Mapping exons in each Isoform

- Construct a map for isoforms and corresponding exons
- Narrowing into genomic exons 4, 5, 6

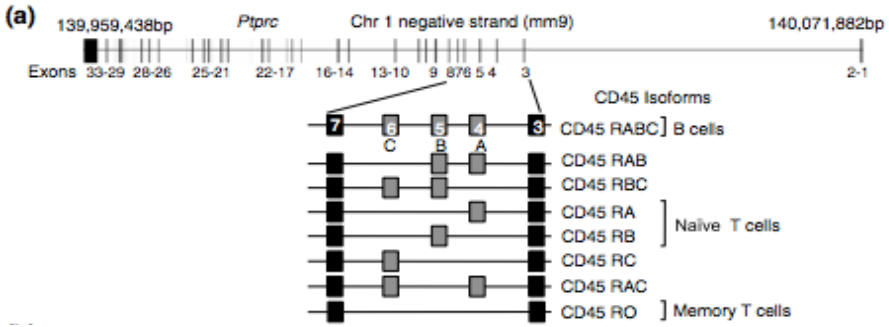
Isoform	201	203	206	209
Total no. of Exons	30	22	4	33



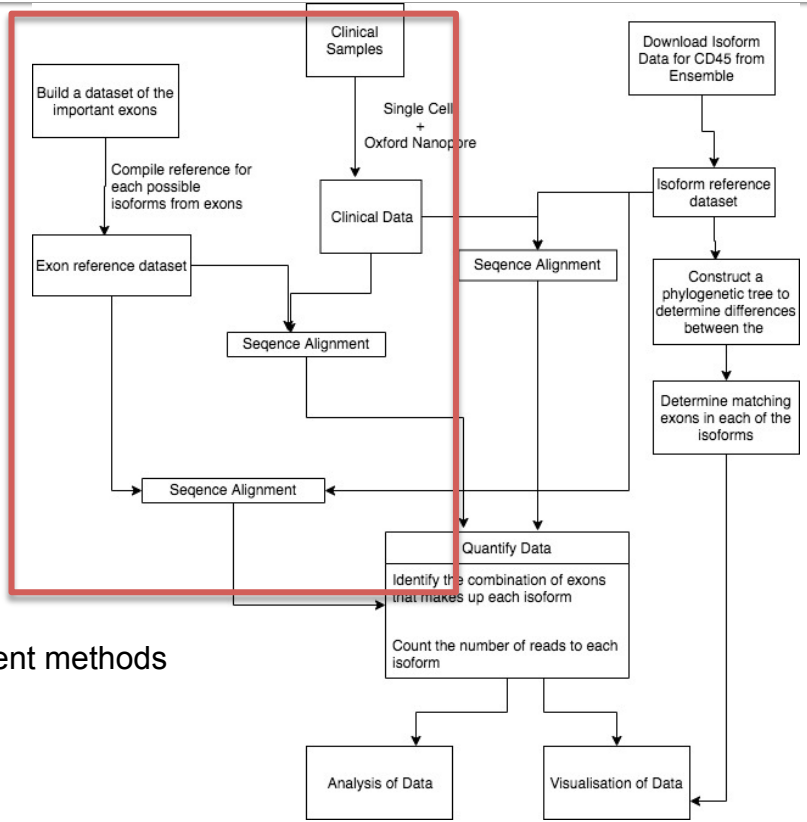


Candidate Isoforms

209	206	203	201
CDRABC	Very little- almost no alignments	CDRBC	CDRO



Refined Method

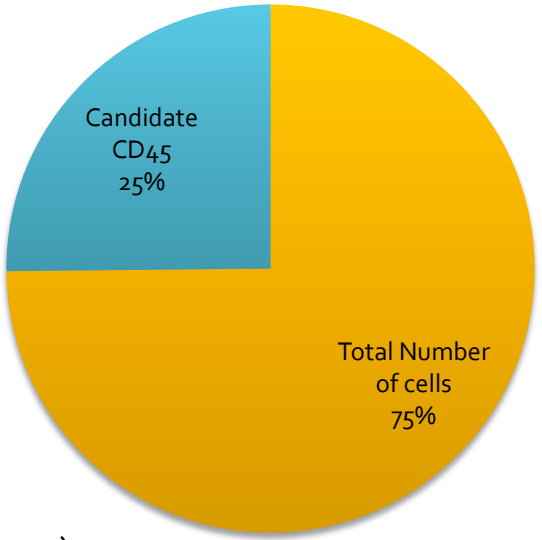


Note: Sequence alignment methods
-Minimap2
-Bowtie2

Experimental measurement of CD45

- Results of the alignment of N=476 for which experimental measurements of CD45 protein expression was available
- N=160 (25%) had contigs that aligned to ANY reference isoform .
- These candidate cells aligned to either 209 (all the exons) or exon 2& 3 &7.

Total Number of cells	Candidate CD45
476	160

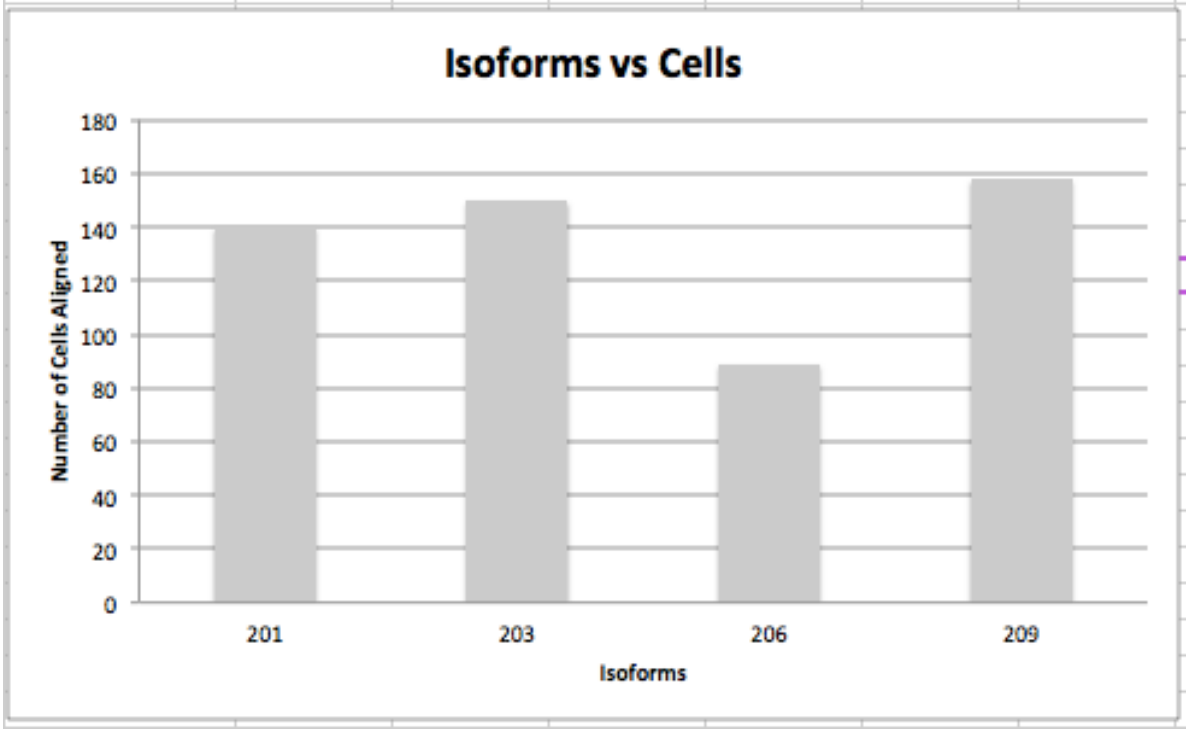


RESULTS shown are from alignment with references (using minimap2 aligner)

Applying to Cell Data

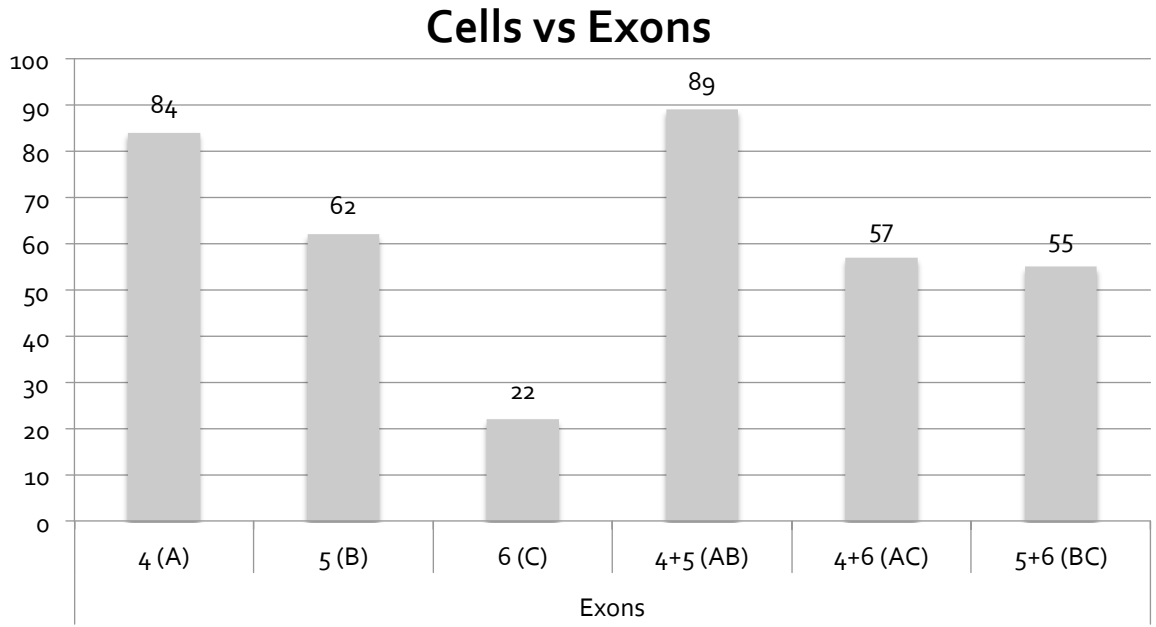
- Total number of candidate cells aligned = 160

Isoforms	201	203	206	209
No. Cells Aligned	140	150	89	158



Mapping of contigs within Exon & Cell Alignment

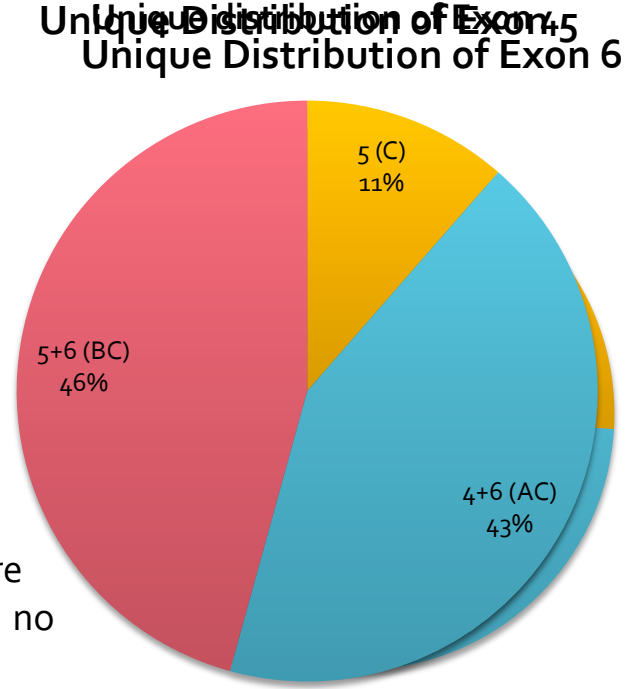
	Exons					
	4 (A)	5 (B)	6 (C)	4+5 (AB)	4+6 (AC)	5+6 (BC)
Num Cells Aligned	84	62	22	89	57	55



Individual Exons

- Total number of Candidate CD45 cells
  N = 160

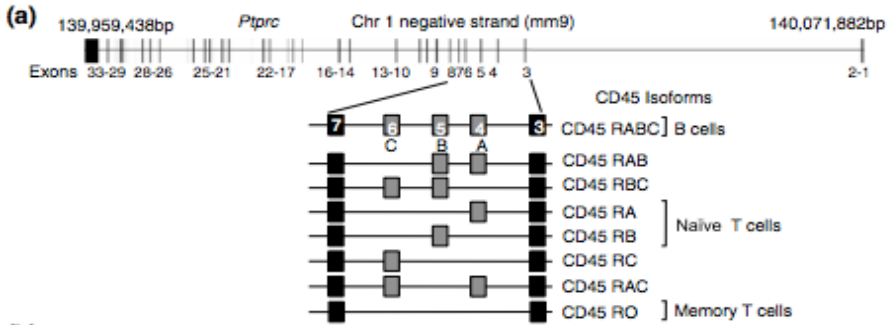
	5 (C) 5 (B)	4+6 (AC) 4+5 (AB)	5+6 (BC) 5+6 (BC)
Num Cells Aligned	82 62	87 89	55 55
Uniquely	27 4	63 58	15 16



- Unique – exon that was present but no other exons of interest were
- Unique Combination – only exons the present 2 exons aligned, no traces of 3rd

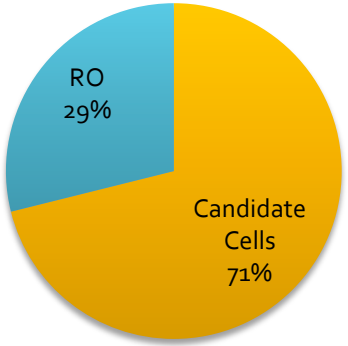
Finding RO

- Known CD45 cells that did not align to exons 4, 5, 6



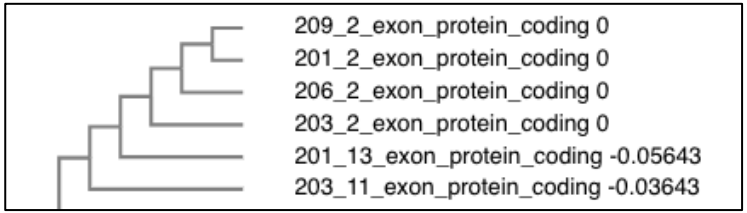
Candidate Cells	RO
160	65

Candidate RO Cells

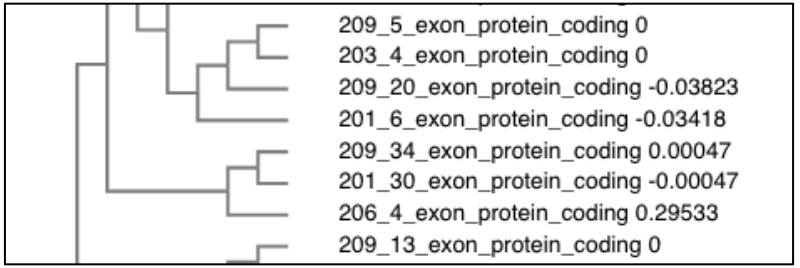


Other interesting findings

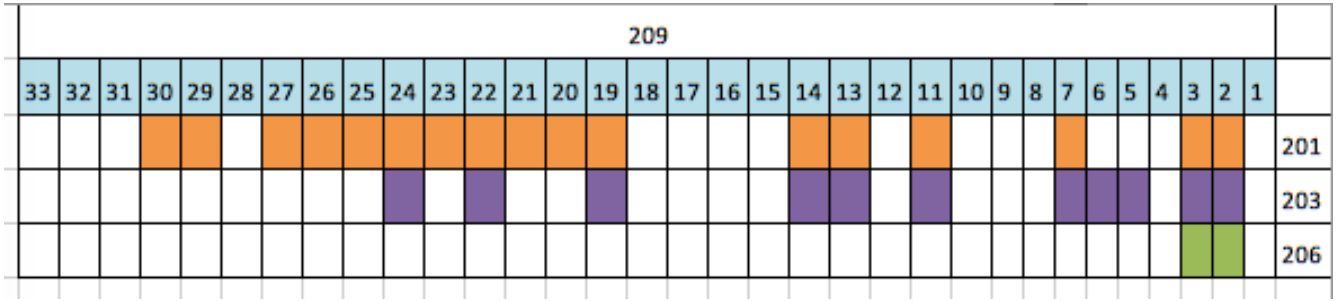
- Some exons in the same isoforms are very similar



- Potential repeating segments?



Other interesting findings



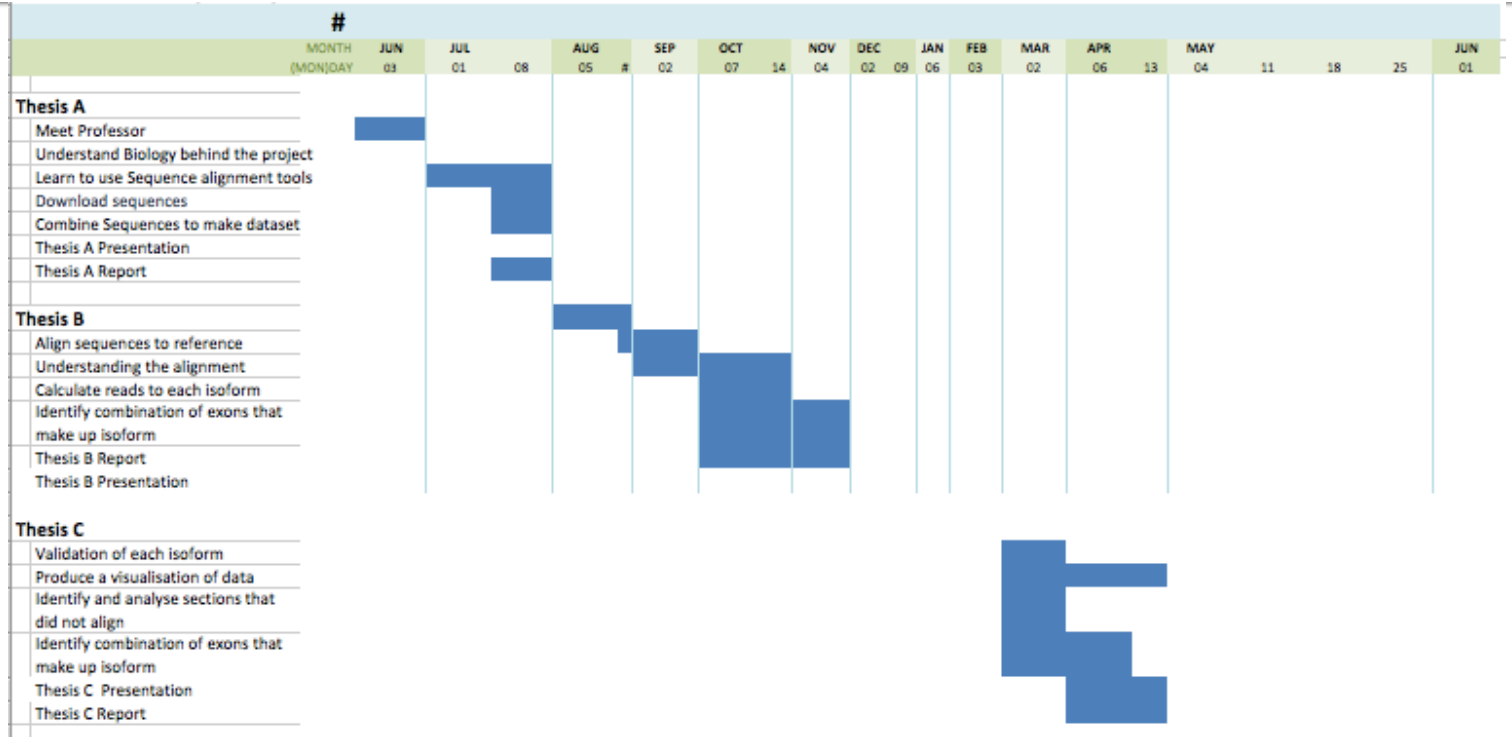
- What lies beyond the exons of interest (exon 4, 5, 6)
- 33 known exons – functions of the other exons and if they contribute to isoforms as well?

Limitations and future directions

- Isoform alignments produced more alignments, overlaps of other 30 exons also playing a part as variables
- Used only protein coding sequences
- Results will need to be validated by comparing protein expression of RA & RO

Project Timeline

Project Timeline



Thesis C

- Validation of each isoform
- Produce a visualisation of data
- Investigate other 30 exons
- Investigate similarities between exons

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].