Mapping

Lecture 11
Sept 26, 2016

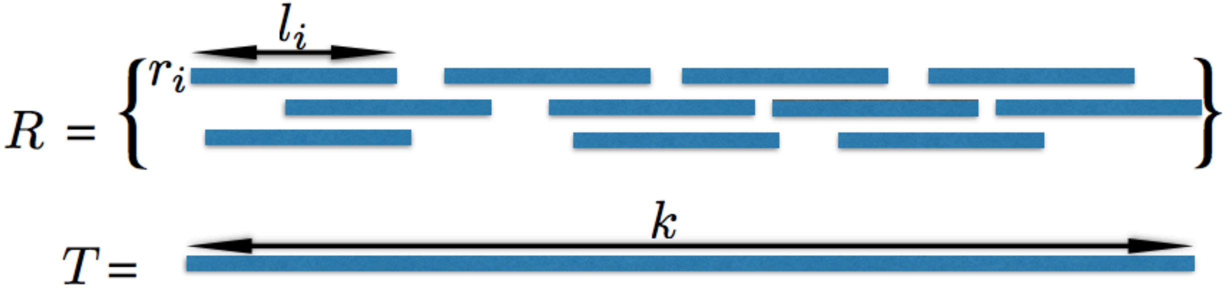
Announcements

- ???

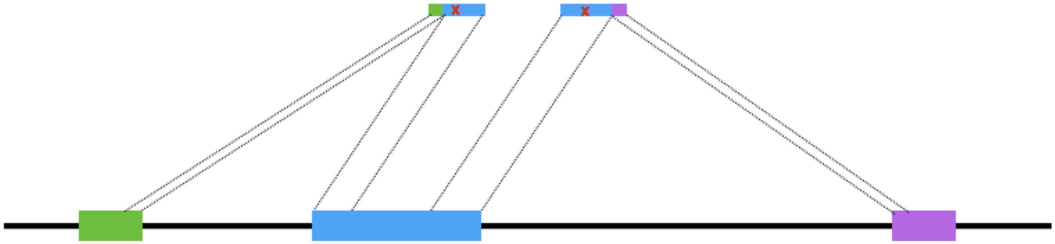
What is the alignment problem?

Given: A collection of sequencing reads, and some target sequence (e.g. a genome)

Find: For each read, all locations where the read is within edit distance **ϵ** of the reference, and the edits that achieve this distance.



Spliced Alignment



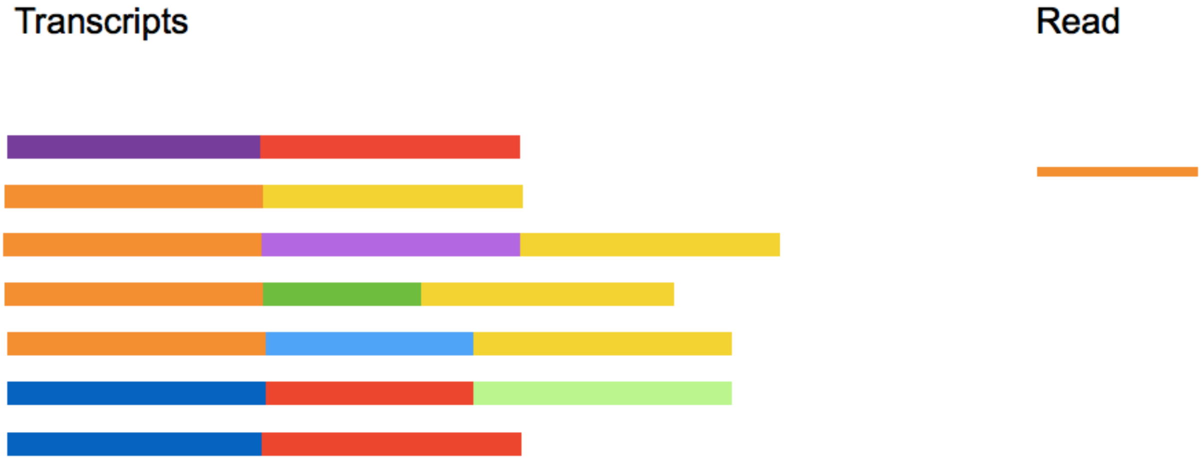
Splice junctions might be known, or *unknown.*

Overlap of read with exon may be *very short*, sequence is ambiguous (e.g. 10 bases).

Sequence of read might be repetitive in the genome.

Aligning reads to a Transcriptome

Consider the following scenario:



Transcripts

Read

Mapping

BWT

Mapping

BWT

https://youtu.be/G7YBi04HOEY?t=1m10s

https://youtu.be/DqdjbK68l3s

https://youtu.be/4n7NPk5lwbI

MAPPING - BWT

A B A A B A

$$\begin{array}{ccccccc}
\$ & a & b & a & a & b & a \\
a & \$ & a & b & a & a & b \\
b & a & \$ & a & b & a & a \\
a & b & a & \$ & a & b & a \\
a & a & b & a & \$ & a & b \\
b & a & a & b & a & \$ & a \\
a & b & a & a & b & a & \$
\end{array}$$

Mapping - BWT

A B A A B A

$$
\begin{array}{ccccccc}
\$ & a & b & a & a & b & a \\
a & \$ & a & b & a & a & b \\
a & a & b & a & \$ & a & b \\
a & b & a & \$ & a & b & a \\
a & b & a & a & b & a & \$ \\
b & a & \$ & a & b & a & a \\
b & a & a & b & a & \$ & a
\end{array}
$$

Mapping - BWT

G A C T C G

Mapping - BWT

Tools Available

Mapping

Format specification: http://samtools.github.io/hts-specs/SAMv1.pdf

MAPPING - BWT

Alignment Fields

- Col1
- Col2
- Col3
- Col4
- Col5
- Col6
- Col7
- Col8
- Col9
- Col10
- Col11

Mapping - BWT

http://broadinstitute.github.io/picard/explain-flags.html