

SV Analysis

4 Main Stages

1

Biological
Question

2

Design &
Conduct
Experiment

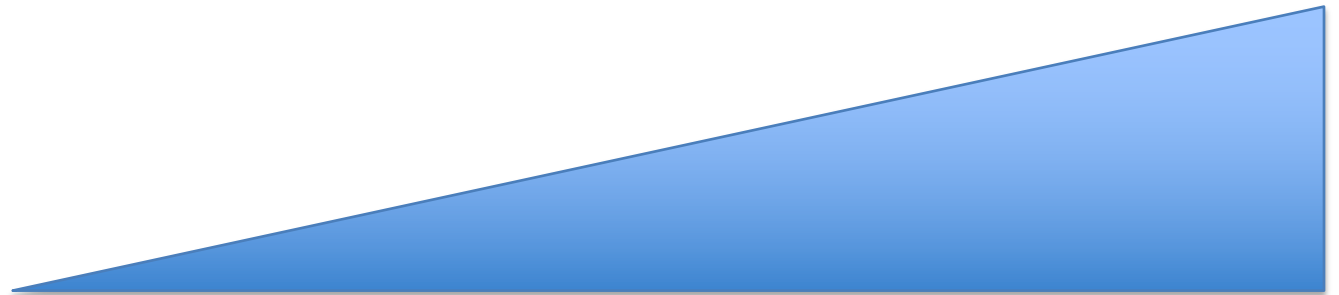
3

Prepare NGS
Library &
Sequence

4

Sequencing
Analysis

Class time



4 Typical Stages of Variant Analysis

1

Read Quality
Control



2

Map Reads



3

Identify Variants



4

Visualize Variants

What is a SV

- SNV = Single Nucleotide Variant
- In reality SNV is broader than a single nucleotide.
- Better definition would be 'variant small enough that majority of reads will span the variant site(s) and map correctly on both sides of it.
- SV = Structural Variant
- Anything that is not a SNV.

Why do we care about SVs?

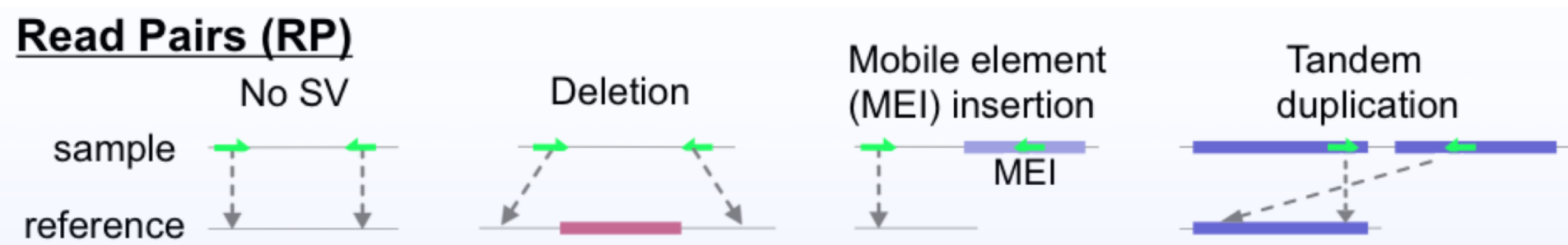
- ~25% of of total variation may be structural in nature ([Deatherage et al 2015 Frontiers in genetics](#)).
- Many programs/pipelines/papers don't bother looking for them because they are classically more difficult to find.

How do we find SVs?

- Larger insert sizes (mate pair sequencing) or read lengths (Nanopore) can significantly increase the ability to identify large rearrangements such as inversions and translocations.
- Make use of paired end information and split read information.

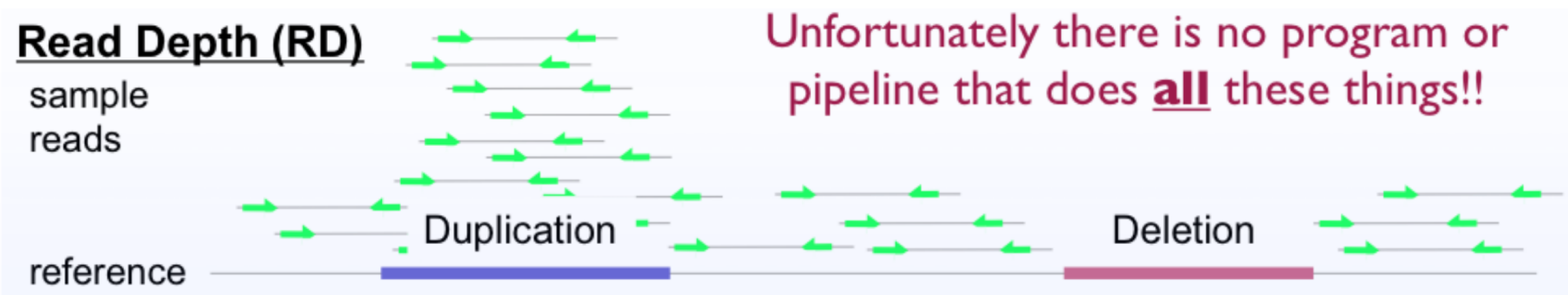
Paired end mapping

- Structural variants can cause different discordant mapping locations between pairs of reads.



Coverage Depth

- Deletions and Amplifications are associated with decrease and increase in coverage depth respectively.
 - Coverage depth alone does not identify specific edges of deletion or amplification.
 - Particularly well hidden in mixed population data.



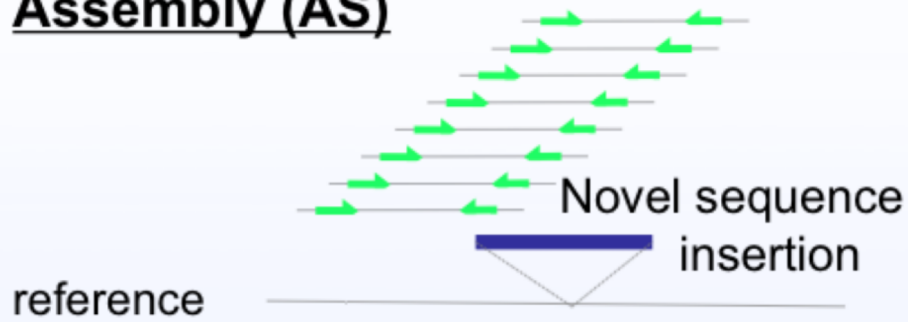
Split read alignment

- Similar to RNA seq identifying splice variants.
- Often more powerful when leveraged with other data such as coverage data.

Split Reads (SR)



Assembly (AS)



Tutorials

- Using a small data set use the SVdetect program to identify structural variants.
- This will include a quick read mapping step as it is a new data set, but try not to get bogged down on that part.
- Together with the data from the SNV tutorial tomorrow we will go through some more advanced visualization tutorials to actually see the variants in a genomic context.