

SNV Analysis

4 Main Stages

1

Biological
Question

2

Design &
Conduct
Experiment

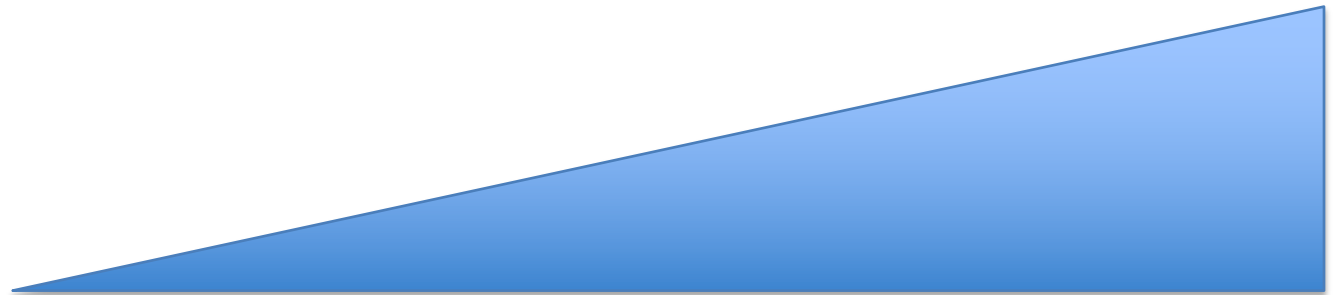
3

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 SNV

- 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 with high accuracy that it is not mistaken for an artifact or error.'

Preparing files for SNV discovery

1. Create an index

- Similar to the first step in read mapping, and for the same principles.
- This must be done on the same reference file that was done.

2. Convert sam files to bam files

- While sam files are nice and readable to us as humans, bam files are binary and much more readable to computers.

3. Sort bam files

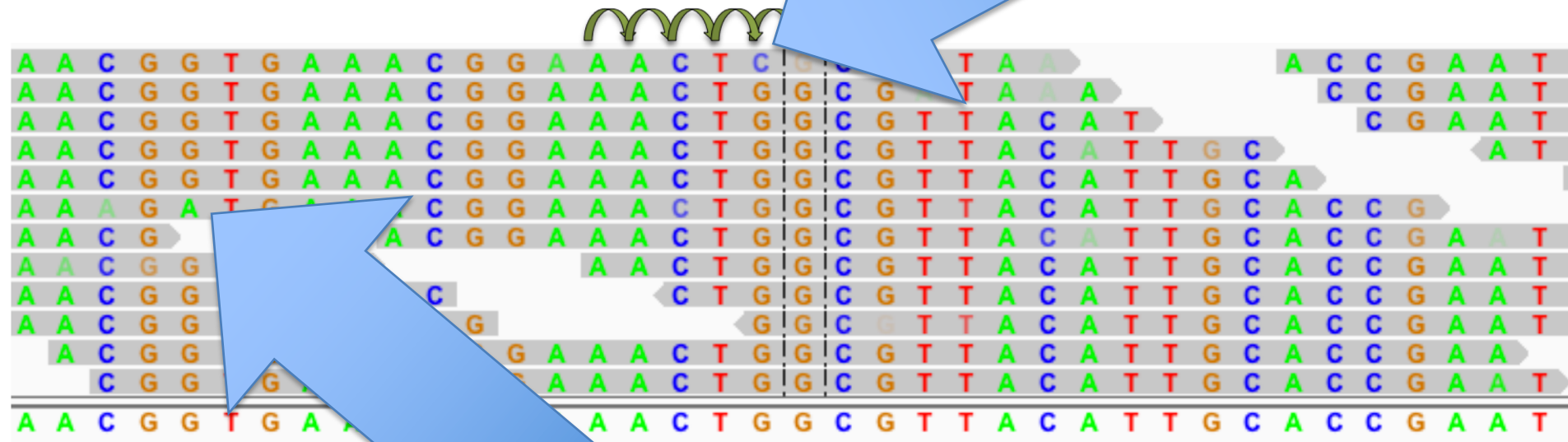
- Sorting allows program to look easily look for all reads overlapping a given region of the genome without having to search each line. (remember Hamlet example)

4. Index sorted bam files

- Indexes make things faster.

Errors?

- samtools mpileup
 - Look at each base 1 at a time to determine how many reads support which base.



SNV Tutorials

- Using the data YOU created from the read mapping tutorial, identify SNV in the SNV tutorial.
- There is a secondary objective to give you some information about version control and how the module system and BioTeam don't always work together.
 - PAY attention to the warning boxes in the tutorial to make sure you get the output that is listed or your tutorial will not work.
- There are multiple 'extra' or 'optional' examples at the end of the tutorial.

BUT FIRST ... STRUCTURAL VARIANTS