

[www.leedsomics.org](http://www.leedsomics.org)  
@leedsomics  
omics@leeds.ac.uk

## Dealing with NGS data: Aligning/Mapping reads

**Club Moderators:** Elton Vasconcelos, Euan McDonell, Chew Cheng, and Dapeng Wang

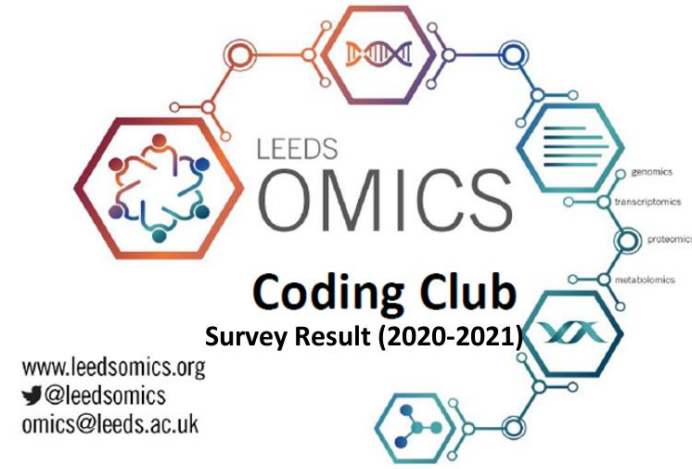
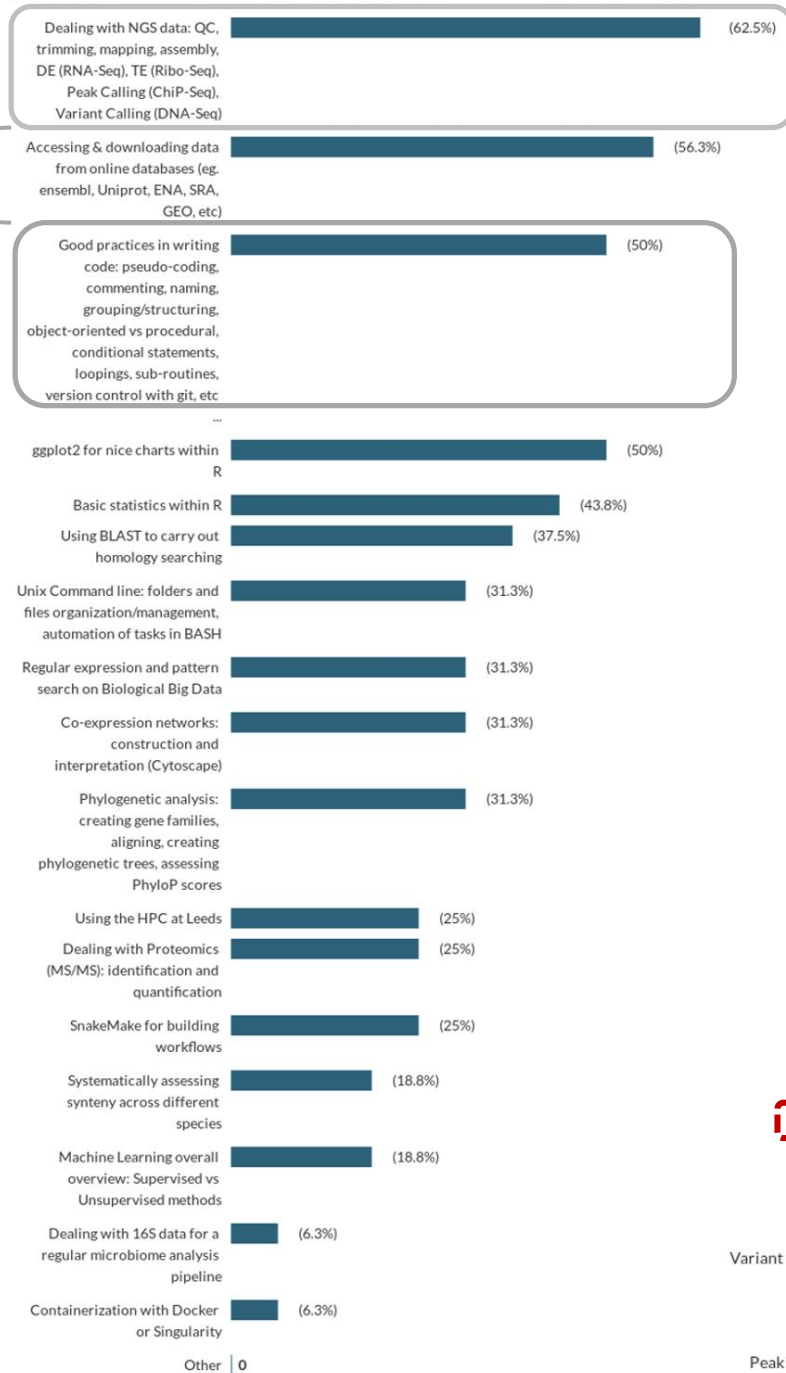
# Topics to be addressed on the 2020-21 season - Survey Result

1st, 4th, 6th, 8th, ... sessions

2nd session

3rd, 5th, 7th, 9th, ... sessions

1 Which of the following topics would you like to attend in our Coding Club sessions?



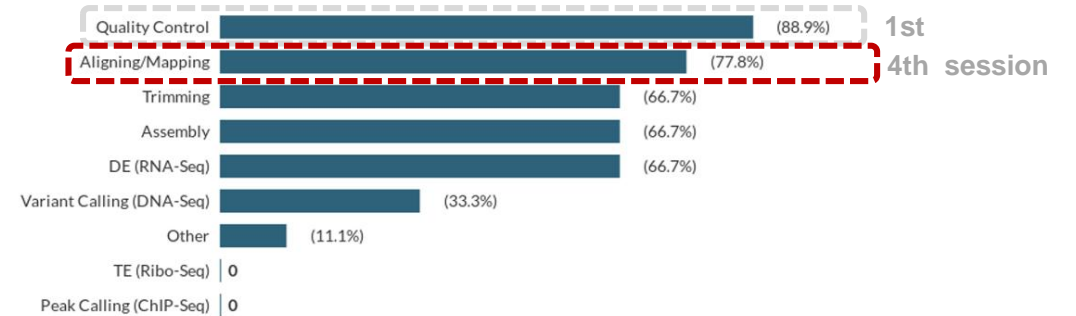
1.b Do you think we should address "Good practices in writing code" topic in more sessions?



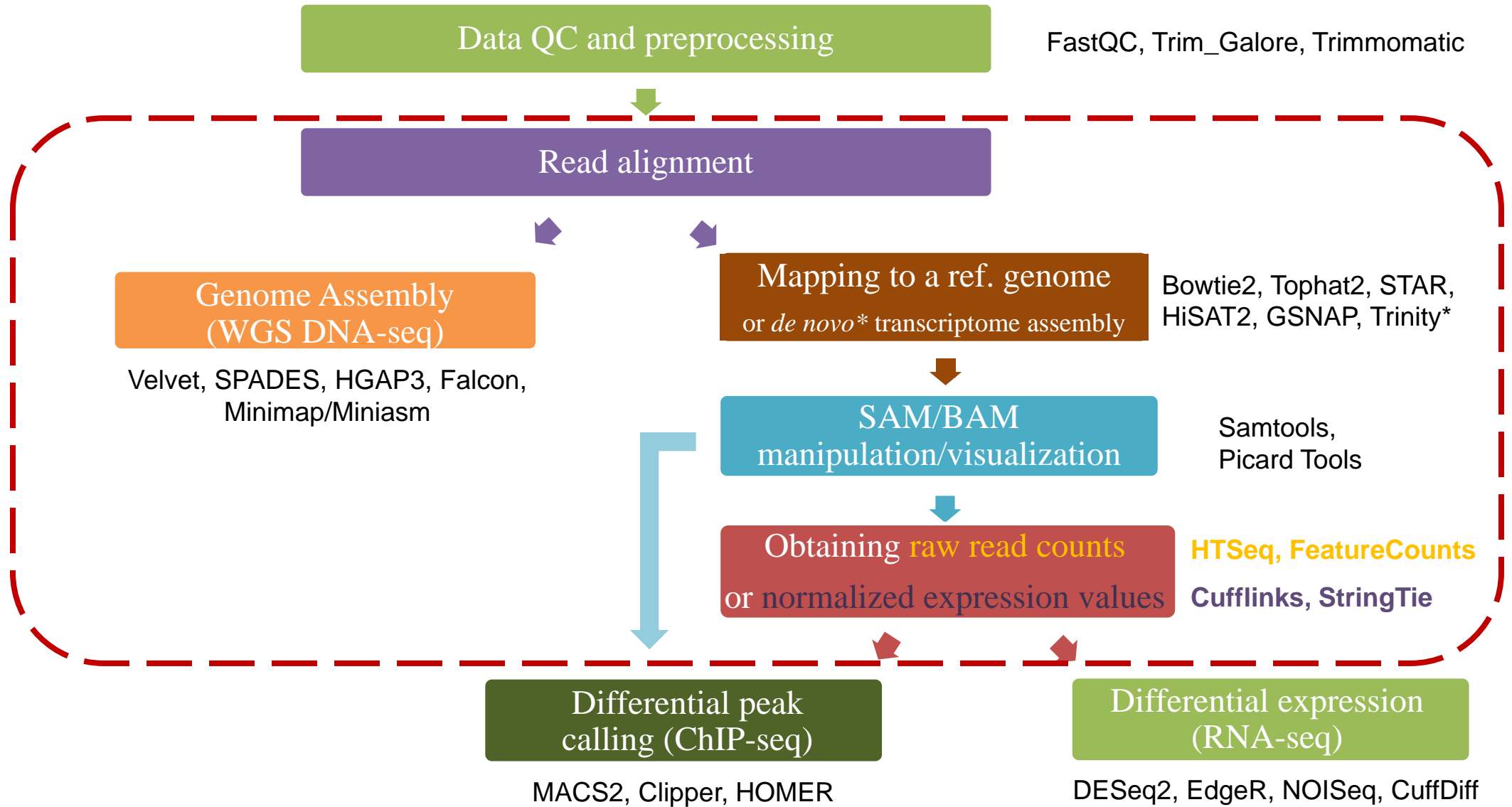
1.c Do you think we should address "Dealing with NGS data" topic in more sessions?



1.c.i Which sub-topics would be of most interest to you?



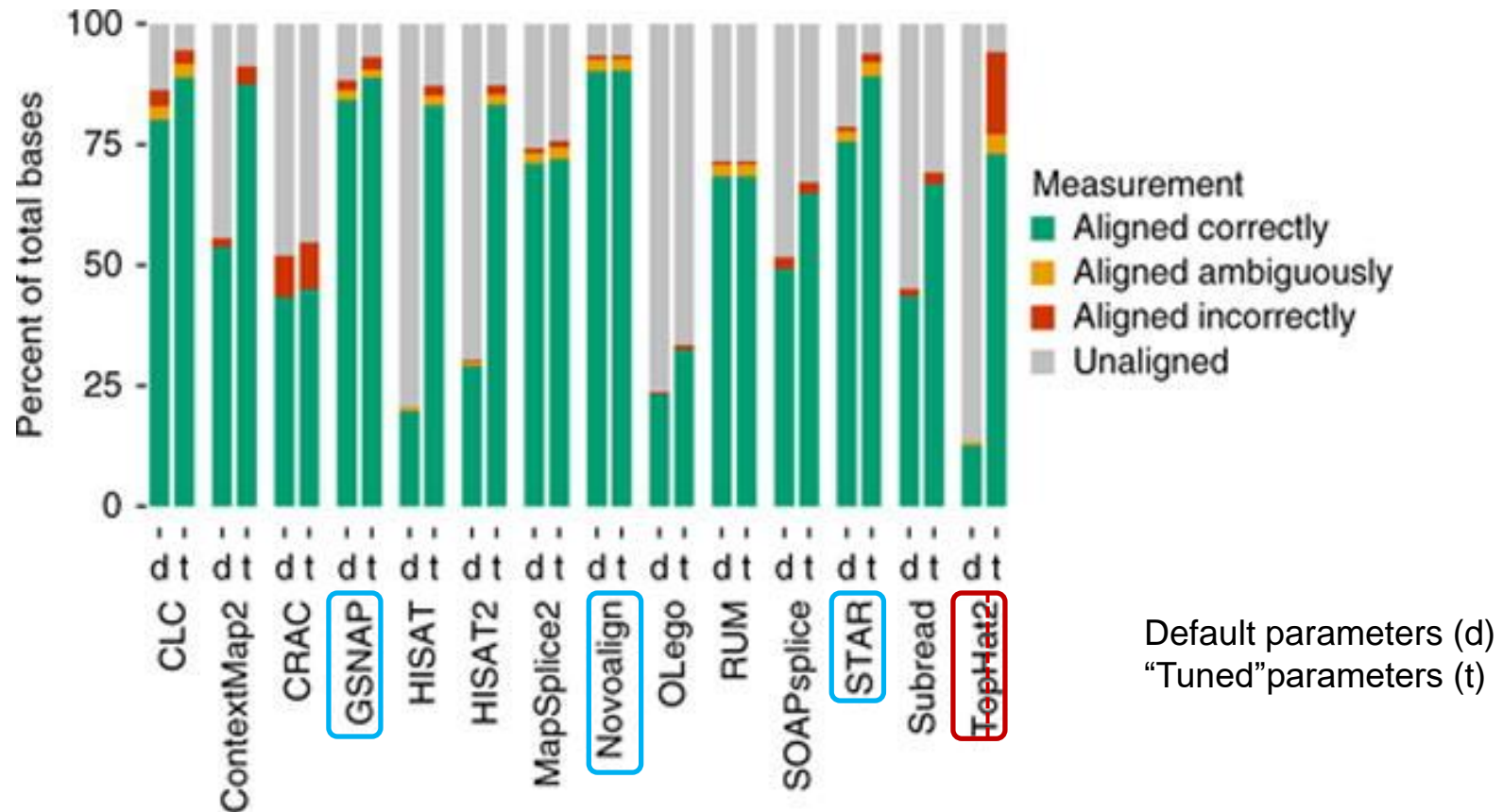
# Important steps on NGS data analysis workflow



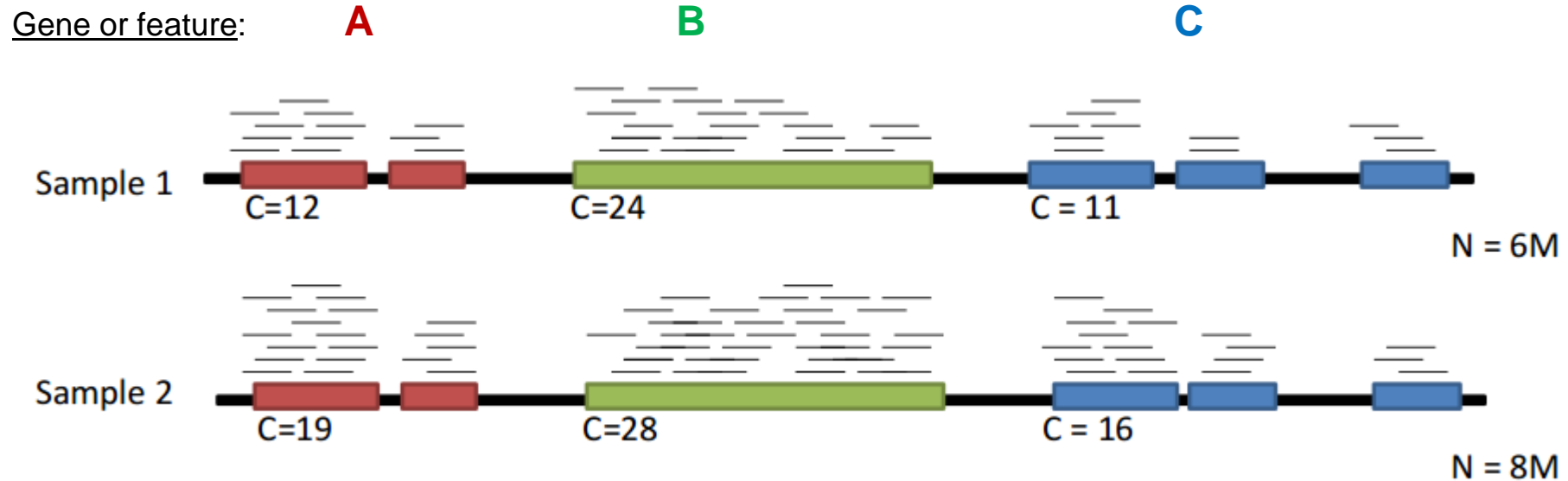
## Which aligner should I run on my **RNA-seq samples**?

Tool	Functionality	Organisms
Bowtie2	Non-splice-aware local alignment against a ref. genome	lacking introns
Tophat2	Splice-aware local alignment against a ref. genome	Any
GSNAP	Splice-aware local alignment against a ref. genome	Any
STAR	Splice-aware local alignment against a ref. genome	Any
HISAT2	Splice-aware local alignment against a ref. genome	Any
Trinity	DBG <i>de novo</i> assembly	lacking a ref. genome
Salmon	Pseudoalignment against a ref. transcriptome	with a robust/reliable transcript isoforms annotation
Kallisto	Pseudoalignment against a ref. transcriptome	with a robust/reliable transcript isoforms annotation

# Aligners performance on RNA-seq data



# Read Counts and Normalization Metrics



- **Reads or Fragments per kilobase per million (RPKM or FPKM)**

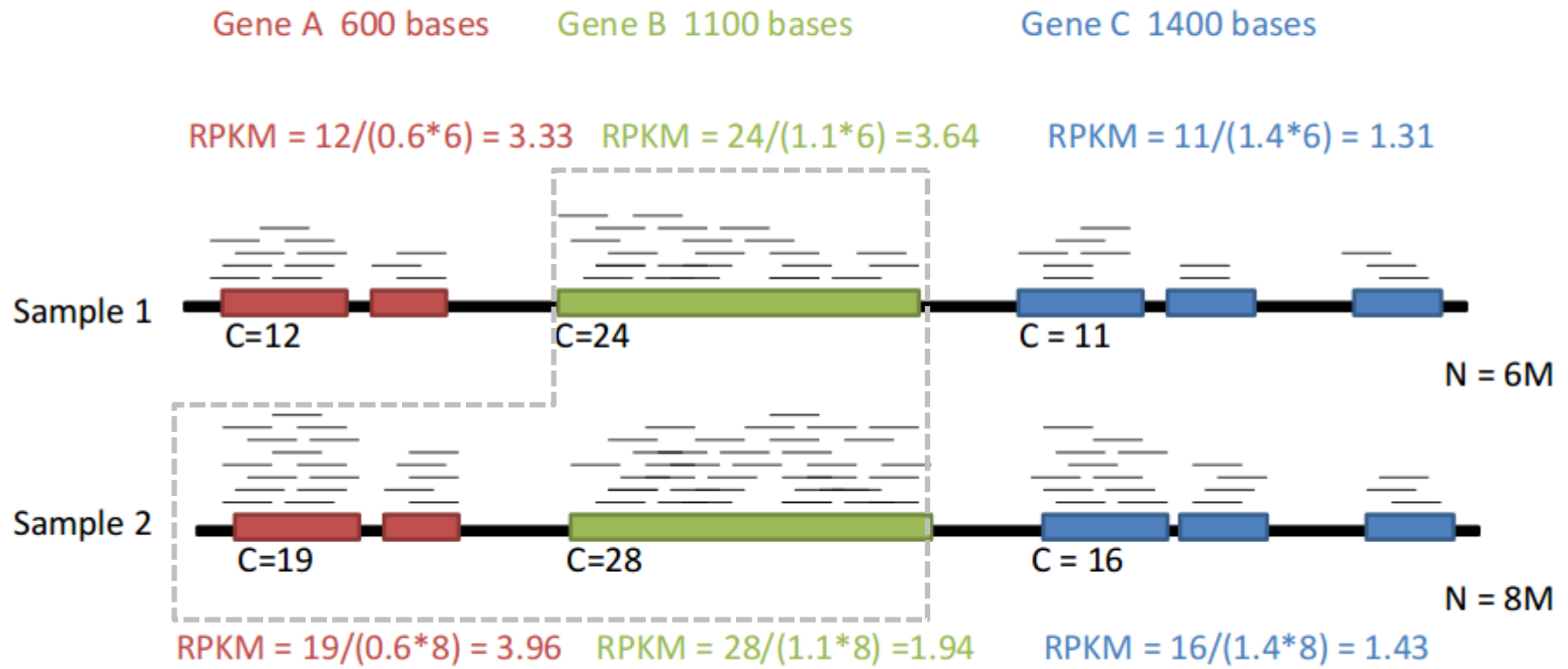
$$\text{RPKM} = \left( \frac{\text{\#aligned reads onto feature (C)}}{\text{feature length in kb} \times \text{\#total reads on sample (N)}} \right) \times 1,000,000$$

- **Transcripts per million**

$$\text{TPM} = (\text{RPK} / \text{sum of all RPKs on sample}) \times 1,000,000$$

→ where **RPK** = #aligned reads onto feature (C) / feature length in kb

# RPKM Example



Bring your issues on!