SequencingMetrics

Catherine Hogg

15/06/2021

The following data is taken from the *SequencingMetrics.txt* file located on ARC3/4 in the following location: */nobackup/medlste/data/RNAseq/PvR_consolidated/SequencingMetrics.txt*. From the original set of sequencing metrics (183 samples across 92* patients) I removed the following samples:

- Those which are absent from the metadata file (10 samples across 5 patients). Agreed in previous meeting that these should not be included in analysis.
- Those which we previously agreed to filter out based on metadata values (55 samples across 28 patients).
- This left us with 118 samples across 59 patients. This is the data we have used in the below sequencing metric analysis.
  - Patient R104 only has sequencing metrics available for their primary tumour, but based on the metadata, their recurrent tumour should be included in the analysis. Need to follow-up on this missing information.

Figure 2 shows that some of the Kim samples have a low sequencing depth of ~10m reads. However, Figure 1 shows that the Kim samples also have longer read lengths (~275) when compared with other sample sources (aside from the Stead samples). Based on this

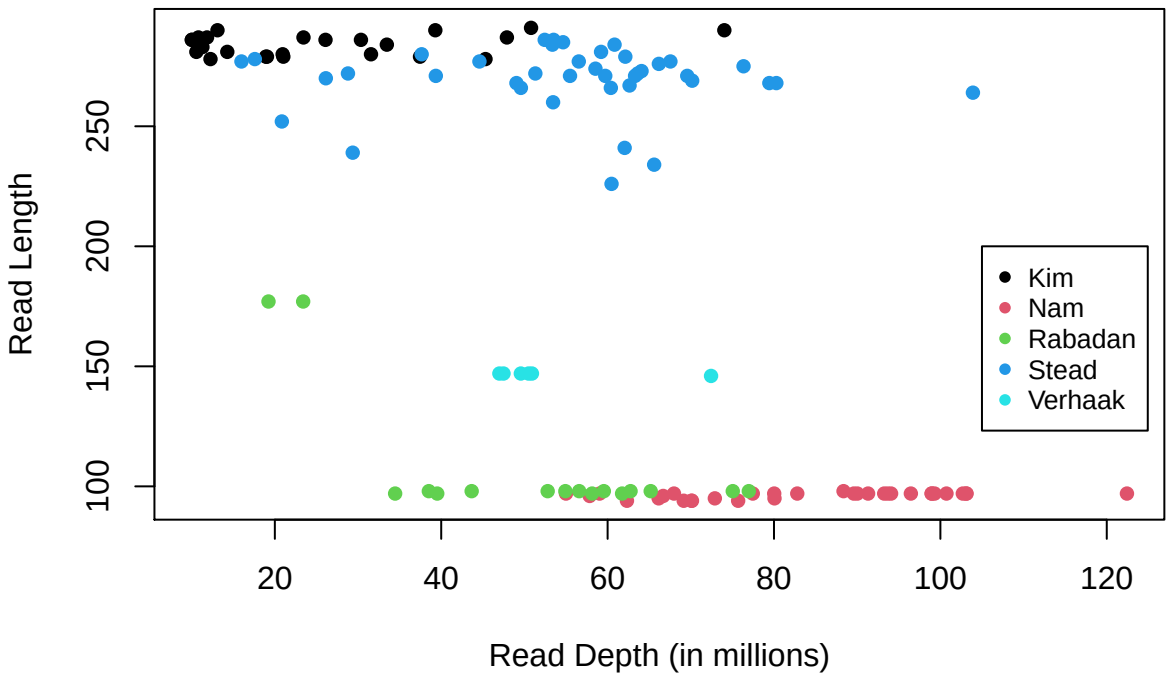


Figure 1: Read depth and length by sample source

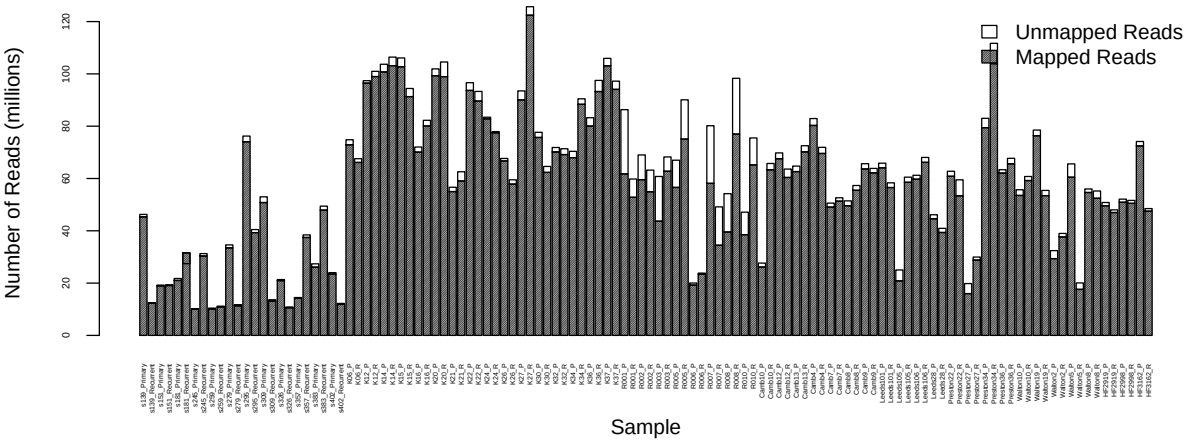
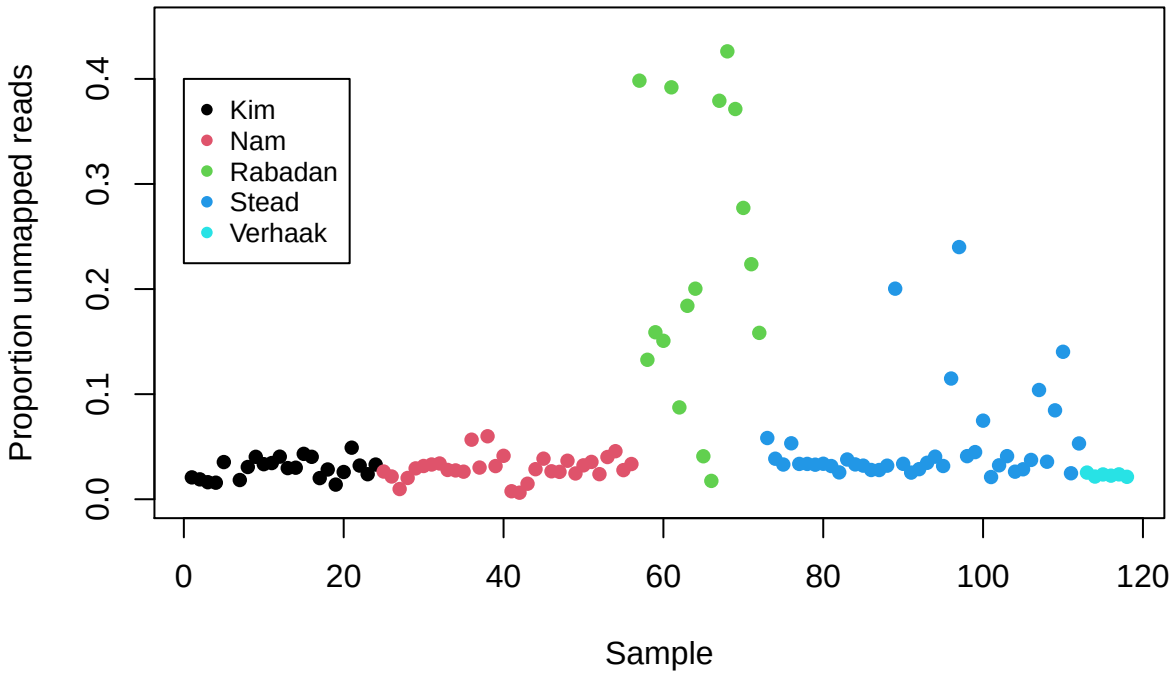
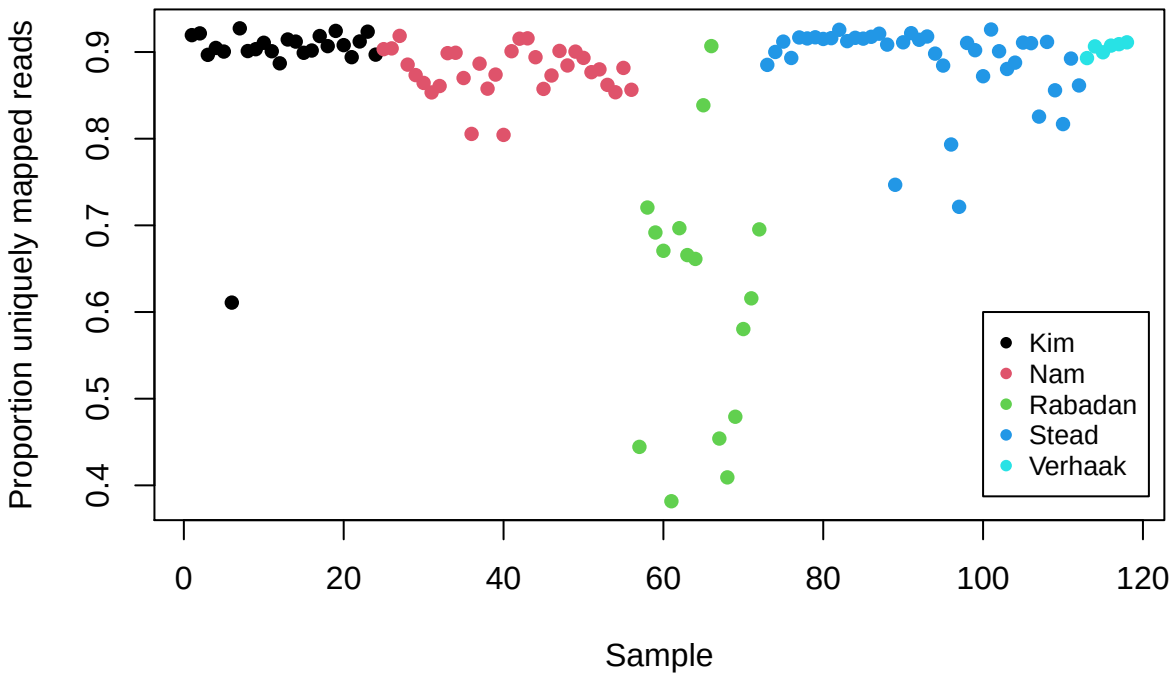
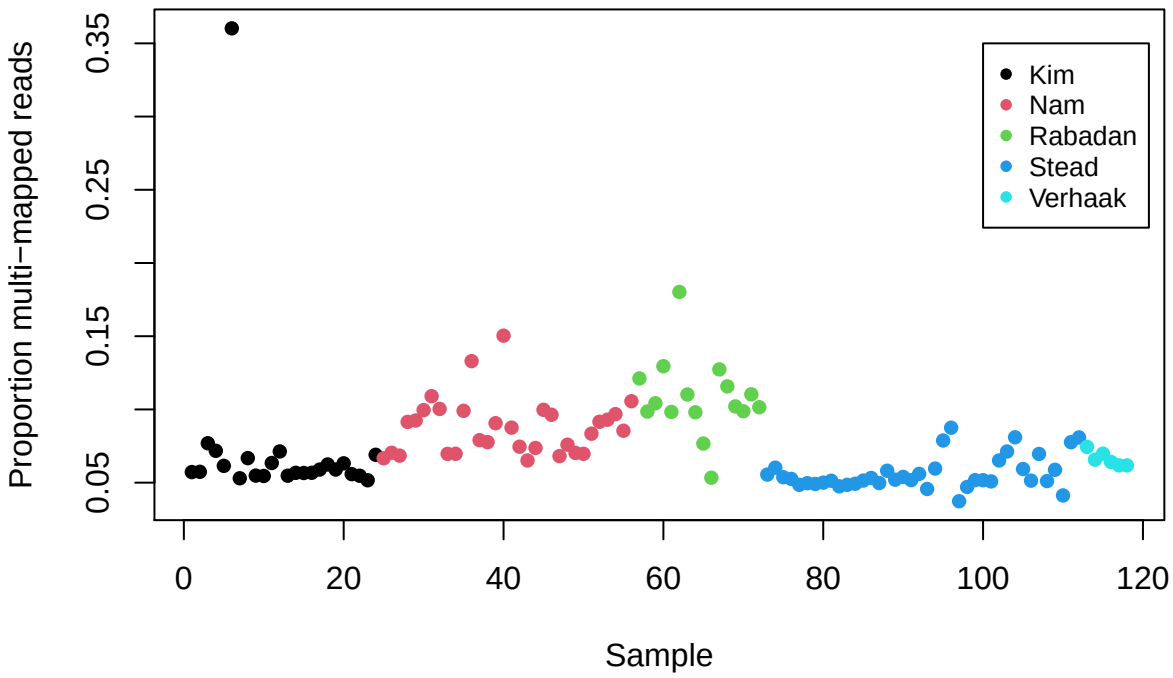


Figure 2: Number of mapped and unmapped reads. Total number of samples = 118 (59 patients). Samples grouped by source in the following order: Kim, Nam, Rabadan, Stead, Verhaak, Verhaak


Proportion unmapped reads
0.0
0.1
0.2
0.3
0.4
Kim
Nam
Rabadan
Stead
Verhaak
0
20
40
60
80
100
120
Sample


Proportion uniquely mapped reads
0.4
0.5
0.6
0.7
0.8
0.9
0
20
40
60
80
100
120
Sample
Kim
Nam
Rabadan
Stead
Verhaak


Proportion multi−mapped reads
0.05
0.15
0.25
0.35
0
20
40
60
80
100
120
Sample
Kim
Nam
Rabadan
Stead
Verhaak