



pRESTO FLAIRR Report

Contents

Overview of Data 1

Quality Scores 2

Read Lengths 3

Primer Identification 4

    Count of primer matches . . . . . 4

    Primer match error rates . . . . . 5

Multiple Alignment of UMI Read Groups 7

    Reads per UMI . . . . . 7

## Warning: package 'ggplot2' was built under R version 4.2.3

## Warning: package 'scales' was built under R version 4.2.3

## Warning: package 'bibtex' was built under R version 4.2.3

## Warning: package 'rlang' was built under R version 4.2.3

## Warning: package 'configr' was built under R version 4.2.3

Overview of Data

Date: Date

Author: Author

Description: Description

Run: Run

Sample: Sample

pRESTO Version: Version

## Quality Scores

Quality filtering is an essential step in most sequencing workflows. pRESTO's FilterSeq tool remove reads with low mean Phred quality scores. Phred quality scores are assigned to each nucleotide base call in automated sequencer traces. The quality score (Q) of a base call is logarithmically related to the probability that a base call is incorrect (P):  $Q = -10\log_{10}P$ . For example, a base call with Q=30 is incorrectly assigned 1 in 1000 times. The most commonly used approach is to remove read with average Q below 20.

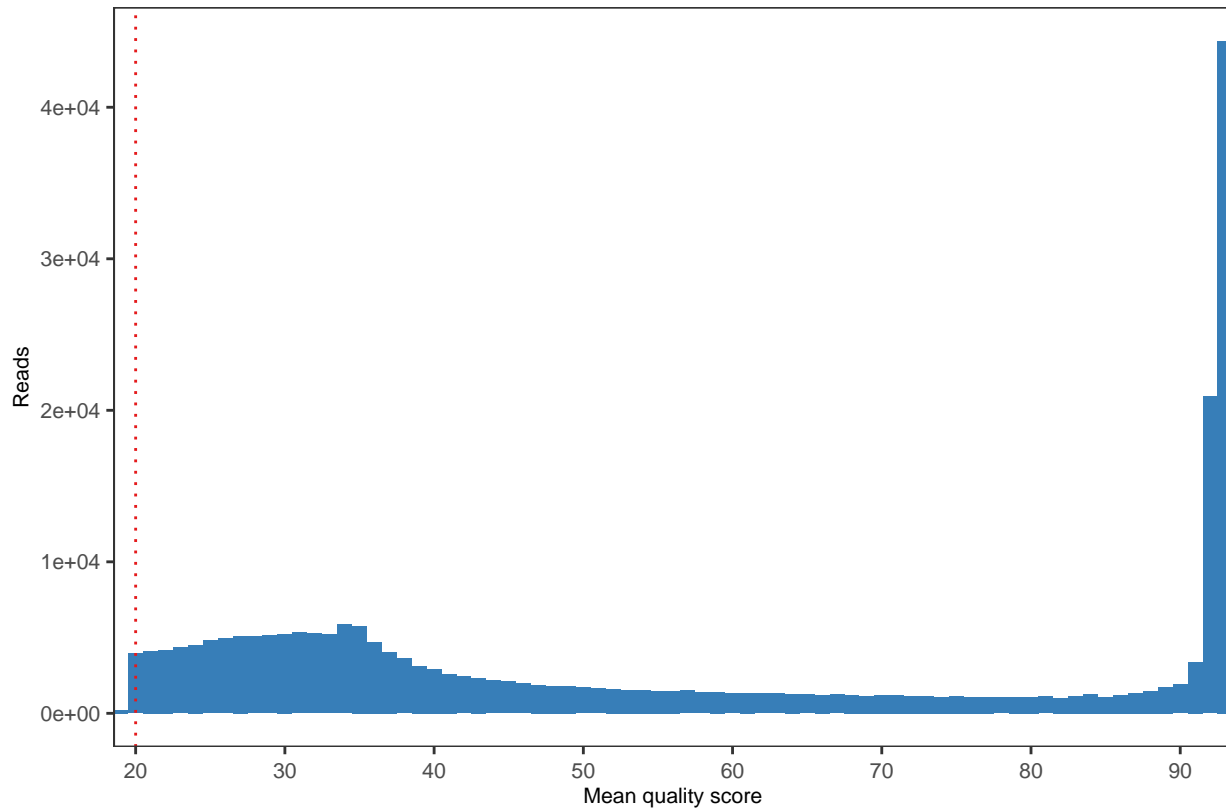


Figure 1: Mean Phred quality scores. The dotted line indicates the average quality score of 20 under which reads were removed.

## Read Lengths

pRESTO's FilterSeq tool is used here to remove reads which are shorter than the nominated length.

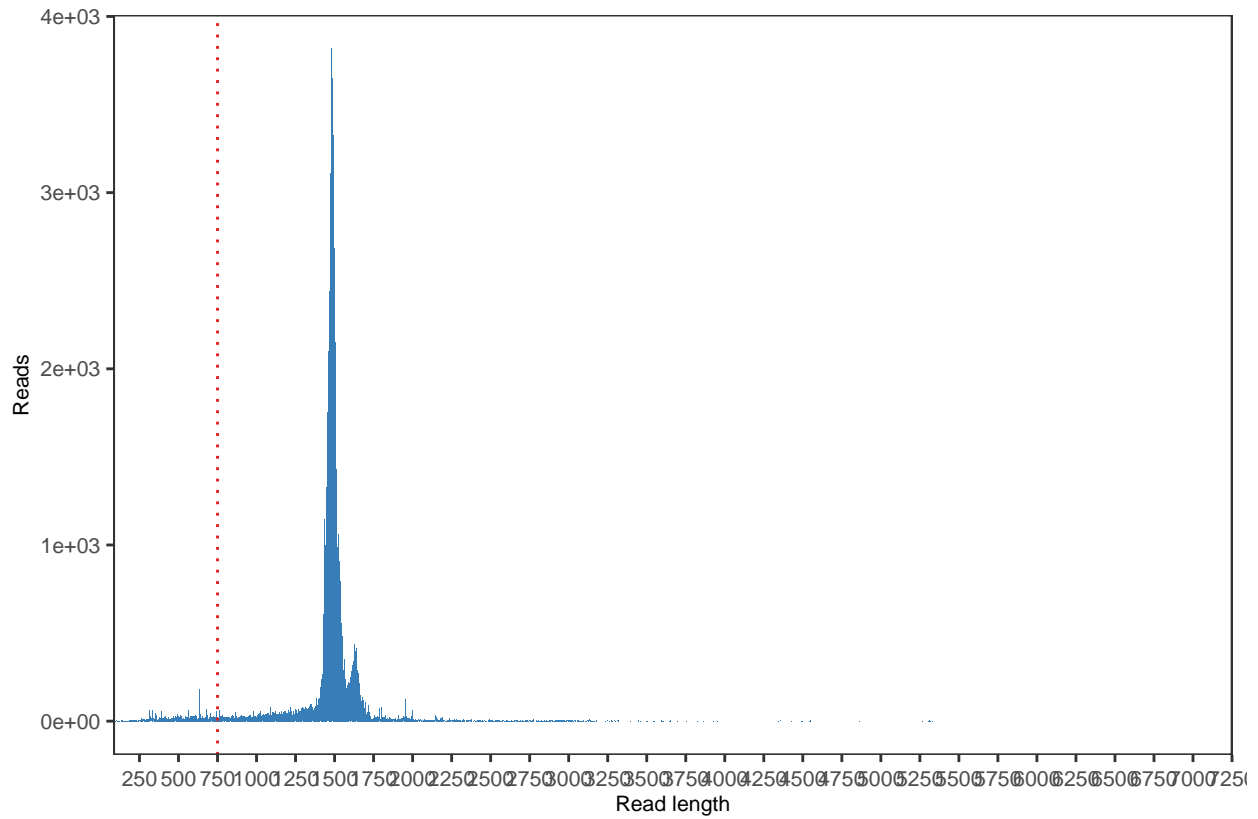


Figure 2: Read length distribution. The dotted line indicates the length of 750 under which reads were removed. Reads shorter than this are unlikely to cover the full V and C region.

## Primer Identification

The MaskPrimers tool supports identification of multiplexed primers and UMIs. Identified primer regions may be masked (with Ns) or cut to mitigate downstream SHM analysis artifacts due to errors in the primer region. An annotation is added to each sequences that indicates the UMI and best matching primer. In the case of the constant region primer, the primer annotation may also be used for isotype assignment.

### Count of primer matches

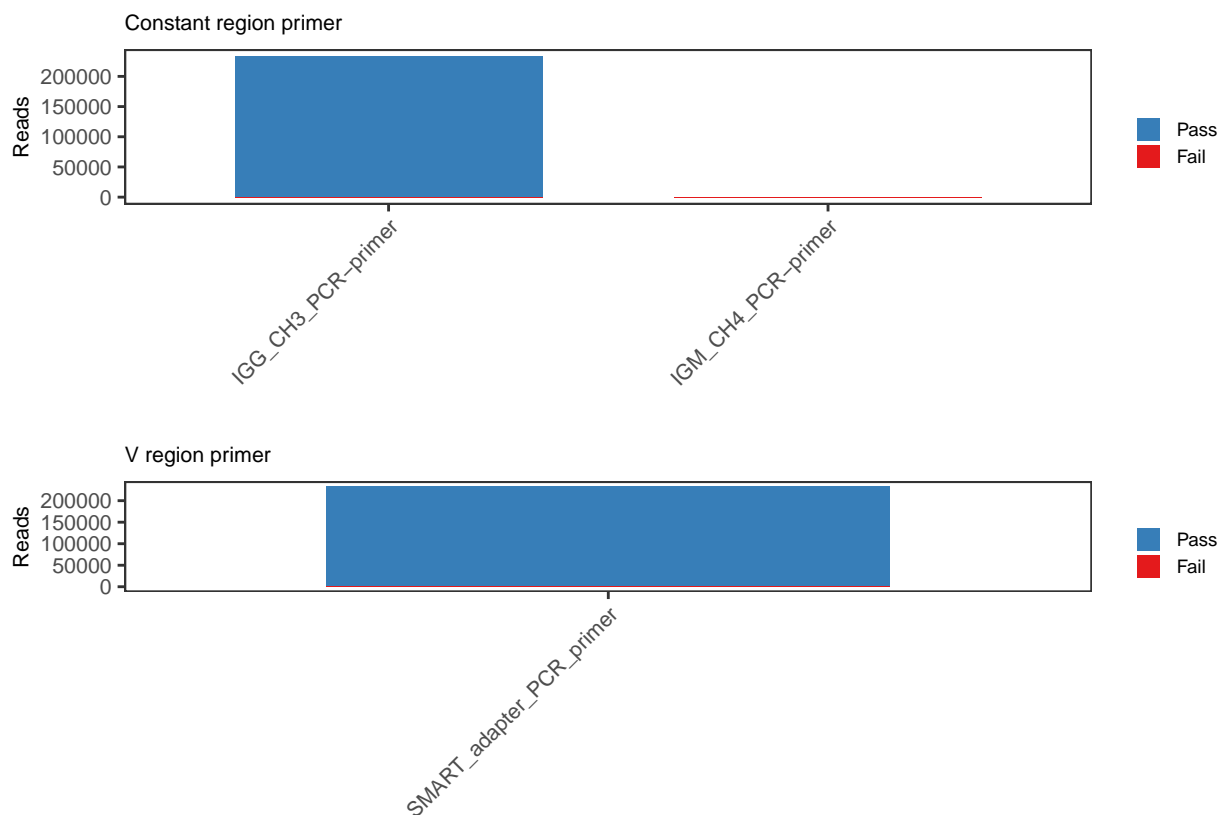


Figure 3: Count of assigned primers. The bar height indicates the total reads assigned to the given primer, stacked for those under the error rate threshold (Pass) and over the threshold (Fail).

## Primer match error rates

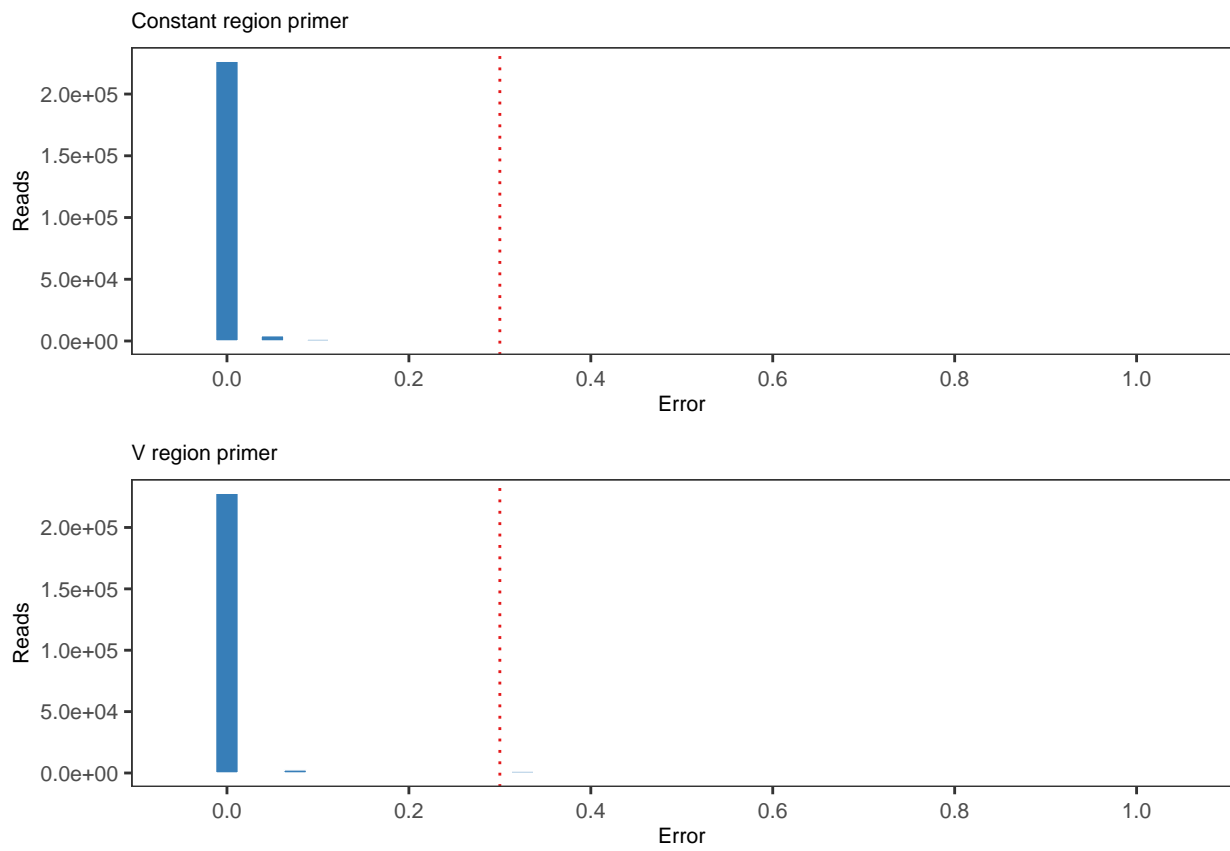


Figure 4: Distribution of primer match error rates. The error rate is the percentage of mismatches between the primer sequence and the read for the best matching primer. The dotted line indicates the error threshold used.

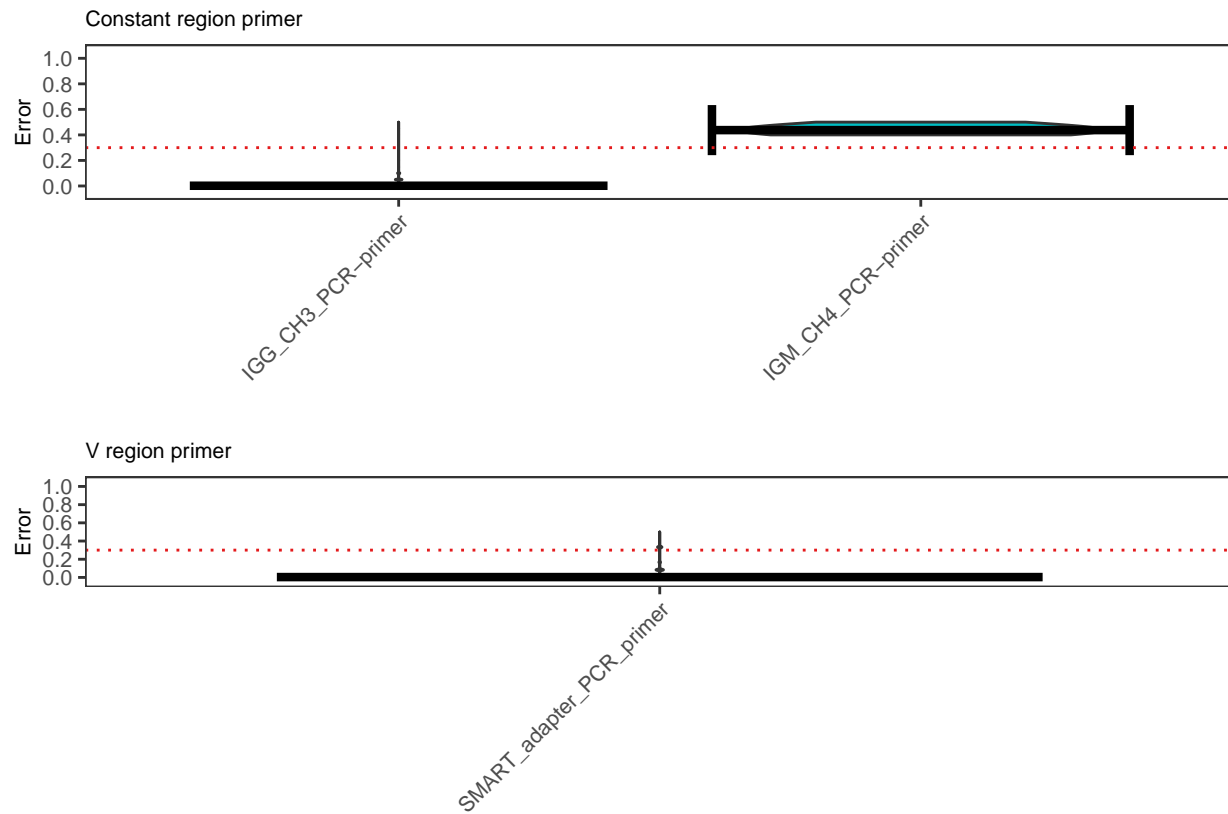


Figure 5: Distribution of primer match error rates, broken down by assigned primer. The error rate is the percentage of mismatches between the primer sequence and the read for the best matching primer. The dotted line indicates the error threshold used.

““

## Multiple Alignment of UMI Read Groups

Reads sharing the same UMI are multiple aligned using the muscle wrapper in the AlignSets tool.

### Reads per UMI

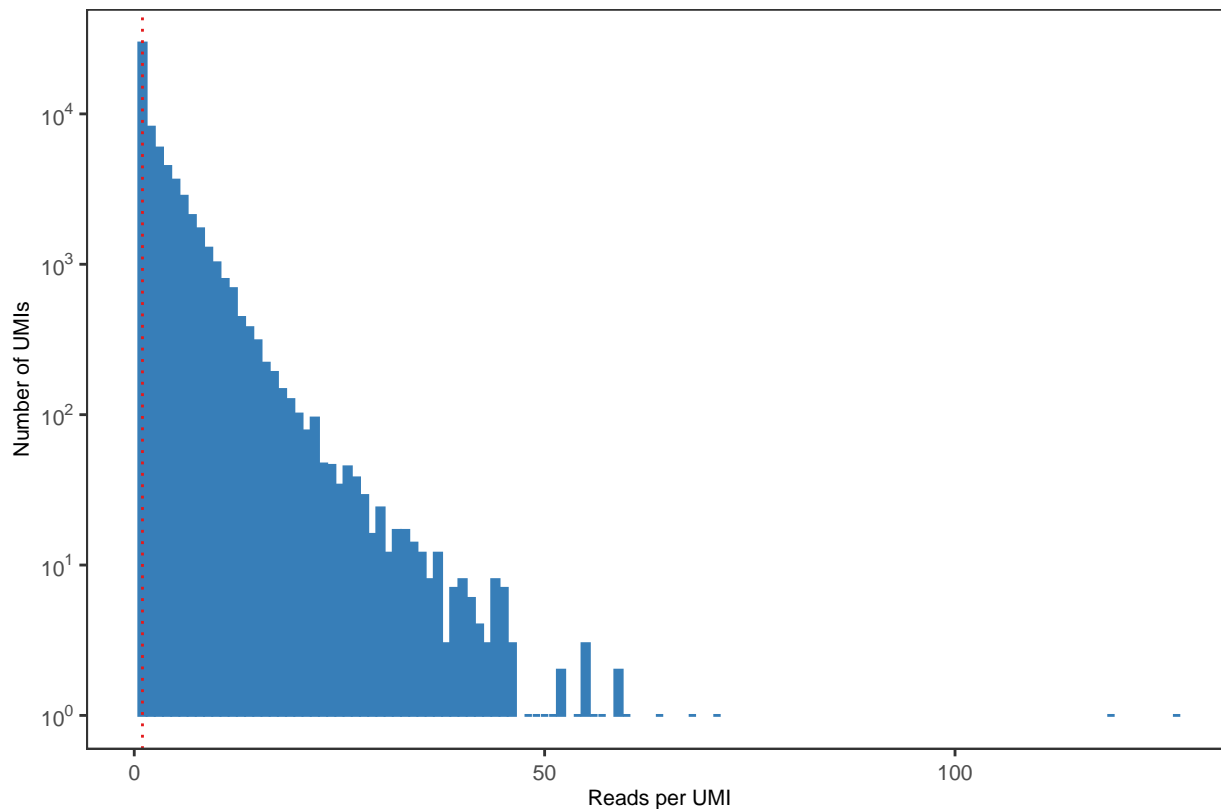


Figure 6: Histogram of UMI read group sizes (reads per UMI). The x-axis indicates the number of reads in a UMI group and the y-axis is the number of UMI groups with that size.