Combined Summary Report

April 1, 2016

1 Pre-Processing Data

This section contains the input read data before processing.

1.1 Base Quality Graph

This graph shows an overview of the range of quality values across all bases at each position in the FastQ file.

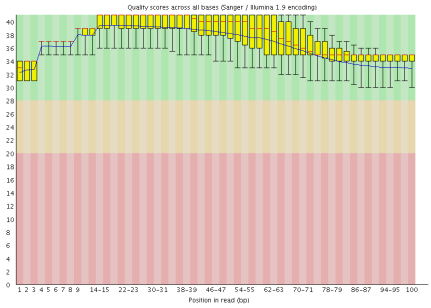

Figure 1: read1

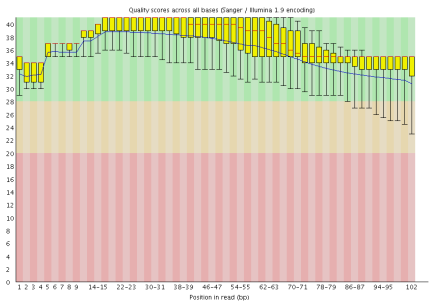

Figure 2: read2

1.2 Data Summary

- read1
  Basic Statistics:

  Total Sequences 12147911
  Sequences flagged as poor quality 0
  Sequence length 100
  %GC 46

Other Statistics Summary:

Metrics ResultBasic Statistics PASS
Per base sequence quality PASS
Per tile sequence quality PASS
Per sequence quality scores PASS
Per base sequence content FAIL
This could've failed due to fragmentation bias in the first 13bp. Checking...
Only fragment bias in the first 13bp found.Per sequence GC content PASS
Per base N content PASS
Sequence Length Distribution PASS
Sequence Duplication Levels FAIL
Overrepresented sequences PASS
Adapter Content PASS
Kmer Content FAIL

- read2
  Basic Statistics:

  Total Sequences 12147911
  Sequences flagged as poor quality 0
  Sequence length 102
  %GC 46

  Other Statistics Summary:

  Metrics ResultBasic Statistics PASS
  Per base sequence quality PASS
  Per tile sequence quality PASS
  Per sequence quality scores PASS
  Per base sequence content FAIL
  This could've failed due to fragmentation bias in the first 13bp. Checking...
  Only fragment bias in the first 13bp found.Per sequence GC content PASS
  Per base N content PASS
  Sequence Length Distribution PASS
  Sequence Duplication Levels FAIL
  Overrepresented sequences PASS
  Adapter Content PASS
  Kmer Content FAIL

2 Post-Processing Data

2.1 Base Quality Graph

This graph shows an overview of the range of quality values across all bases at each position in the FastQ file.

2.2 Data Summary

- read1
  Basic Statistics:

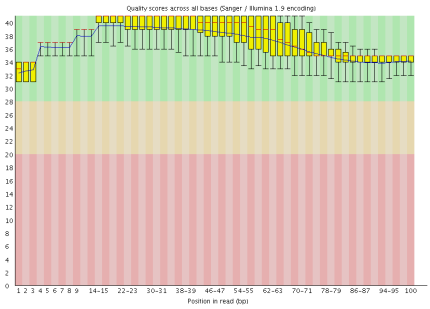

Figure 3: read1

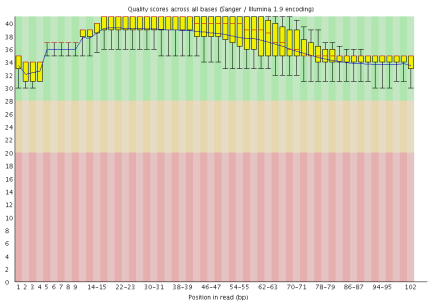

Figure 4: read2

Total Sequences 11940844
Sequences flagged as poor quality 0
Sequence length 2-100
%GC 46

Other Statistics Summary:

Metrics ResultBasic Statistics PASS
Per base sequence quality PASS
Per tile sequence quality PASS
Per sequence quality scores PASS
Per base sequence content FAIL
This could've failed due to fragmentation bias in the first 13bp. Checking...
Only fragment bias in the first 13bp found.Per sequence GC content WARN
Per base N content PASS
Sequence Length Distribution WARN
Sequence Duplication Levels FAIL
Overrepresented sequences PASS
Adapter Content PASS
Kmer Content FAIL

- read2
  Basic Statistics:

  Total Sequences 11940844
  Sequences flagged as poor quality 0
  Sequence length 2-102
  %GC 46

Other Statistics Summary:

Metrics ResultBasic Statistics PASS
Per base sequence quality PASS
Per tile sequence quality PASS
Per sequence quality scores PASS
Per base sequence content FAIL
This could've failed due to fragmentation bias in the first 13bp. Checking...
Only fragment bias in the first 13bp found.Per sequence GC content WARN
Per base N content PASS
Sequence Length Distribution WARN
Sequence Duplication Levels WARN
Overrepresented sequences PASS
Adapter Content PASS
Kmer Content FAIL

3 Post Mapping Data

3.1 Data Summary

Left reads:
Input : 11940844
Mapped : 11552387 (96.7% of input)
of these: 959345 (8.3%) have multiple alignments (1997 have >20)
Right reads:
Input : 11940844
Mapped : 11543943 (96.7% of input)
of these: 955444 (8.3%) have multiple alignments (2001 have >20)
96.7% overall read mapping rate.

Aligned pairs: 11284623
of these: 936771 (8.3%) have multiple alignments
33823 (0.3%) are discordant alignments
94.2% concordant pair alignment rate.

#==========
#All numbers are READ count
#==========

Total records: 25684066

QC failed: 0
Optical/PCR duplicate: 0
Non primary hits 2587736
Unmapped reads: 0
mapq < mapq_cut (non-unique): 1914789

mapq >= mapq_cut (unique): 21181541
Read-1: 10593042
Read-2: 10588499
Reads map to '+': 10590111

Reads map to '-': 10591430
Non-splice reads: 15278135
Splice reads: 5903406
Reads mapped in proper pairs: 17927342
Proper-paired reads map to different chrom:0

3.2 Visual Summary

This graph shows an overview of the range of quality values across all bases at each position in the alignment file.

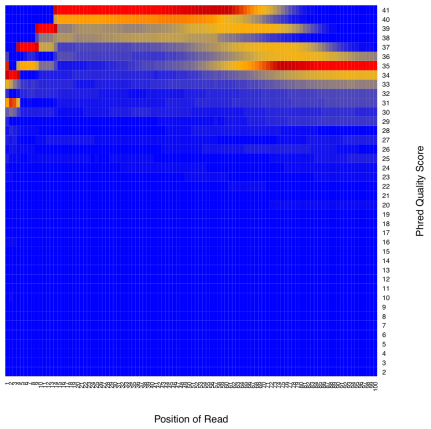

Figure 5: Alignment Quality

This graph shows read duplication rate.

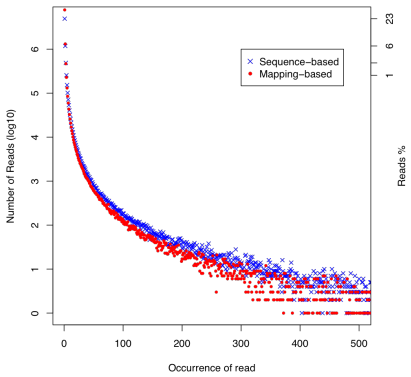

Figure 6: Duplication Rate

- *Sequence based: reads with identical sequence are regarded as duplicated reads.*
- *Mapping based: reads mapped to the exactly same genomic location are regarded as duplicated reads.*