

CurAHack - Challenge 2 - RNA-Seq

Presented by: The Bioinformagicians

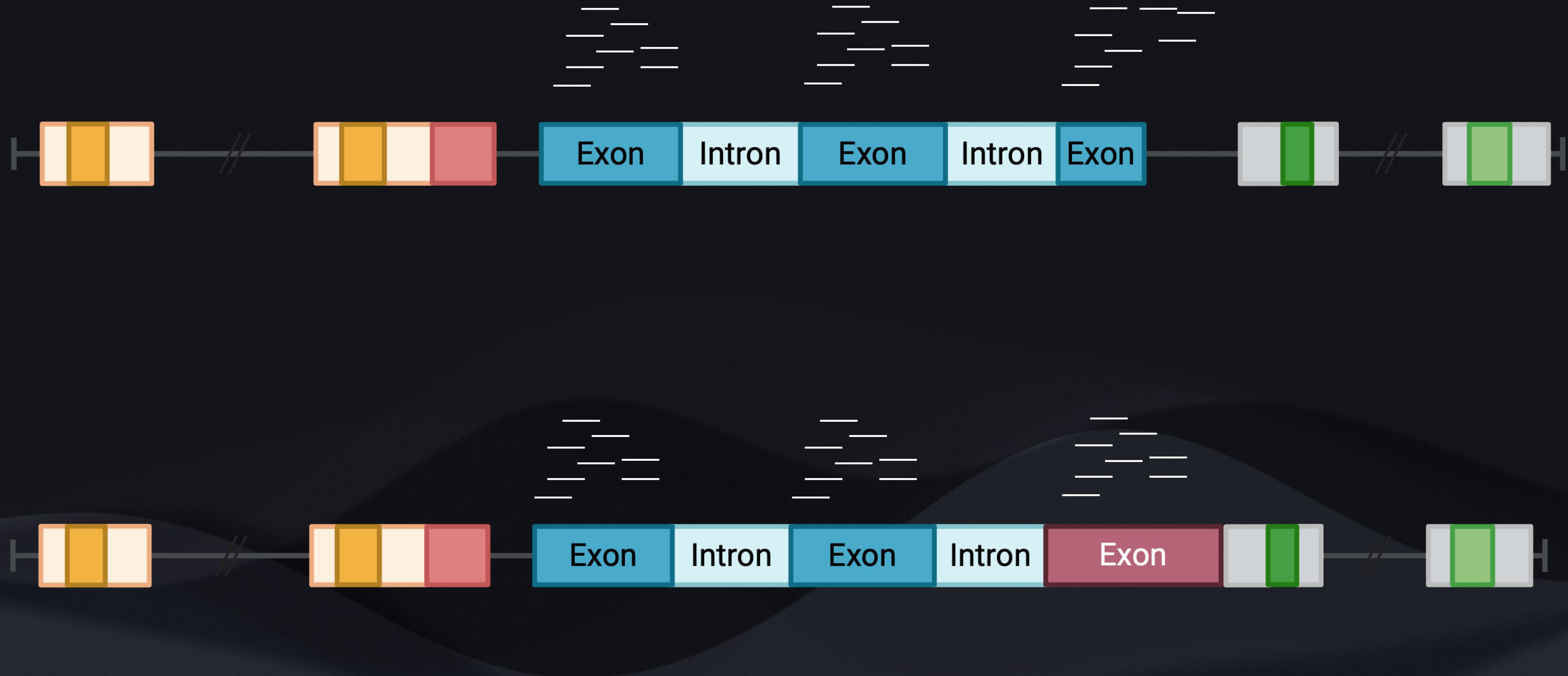
8th-9th of April



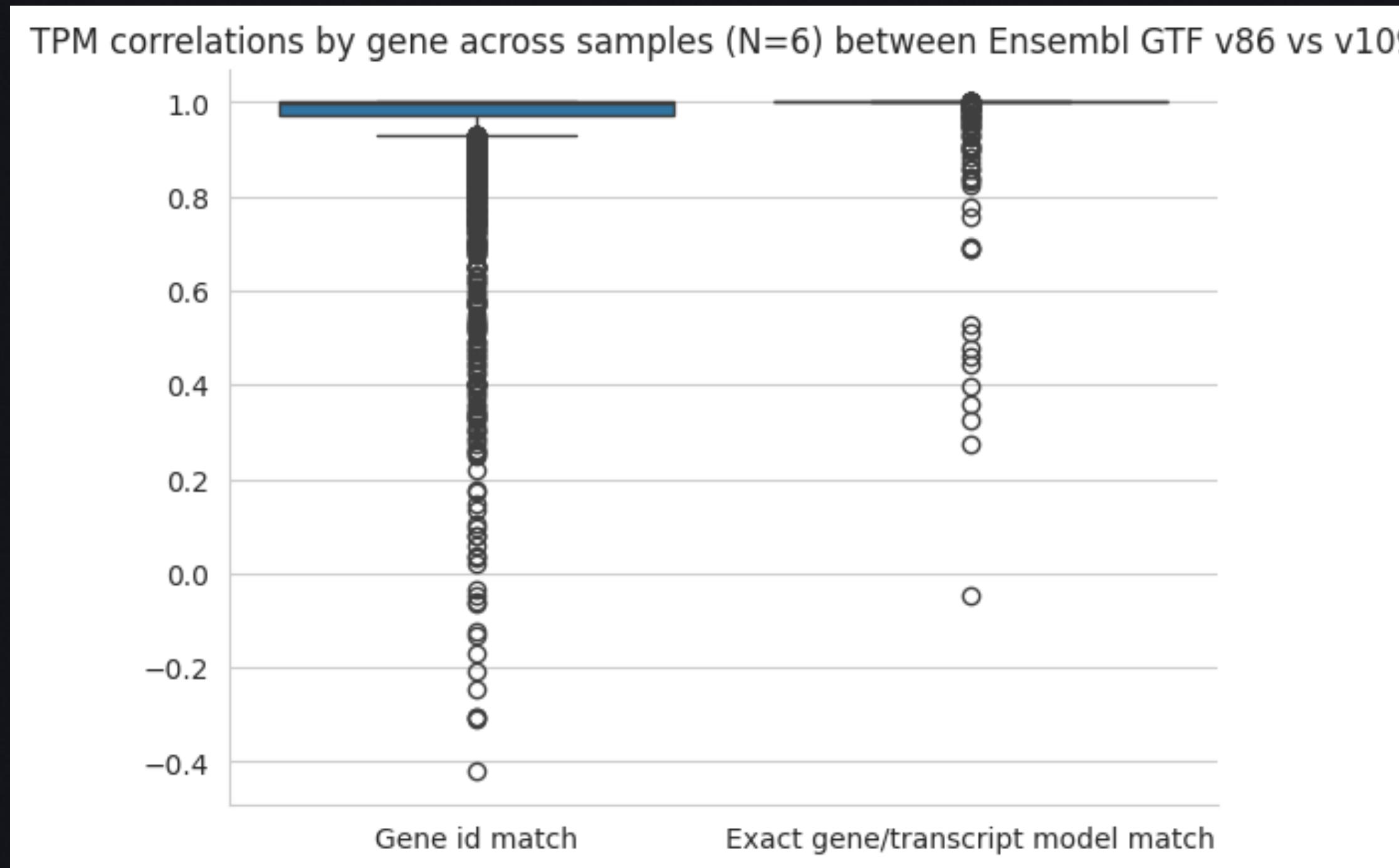
Bioinformagicians



The Challenge: Annotation-free analysis



Prototype 2: quantify the impact of mixing gene annotations and approach to minimise its impact



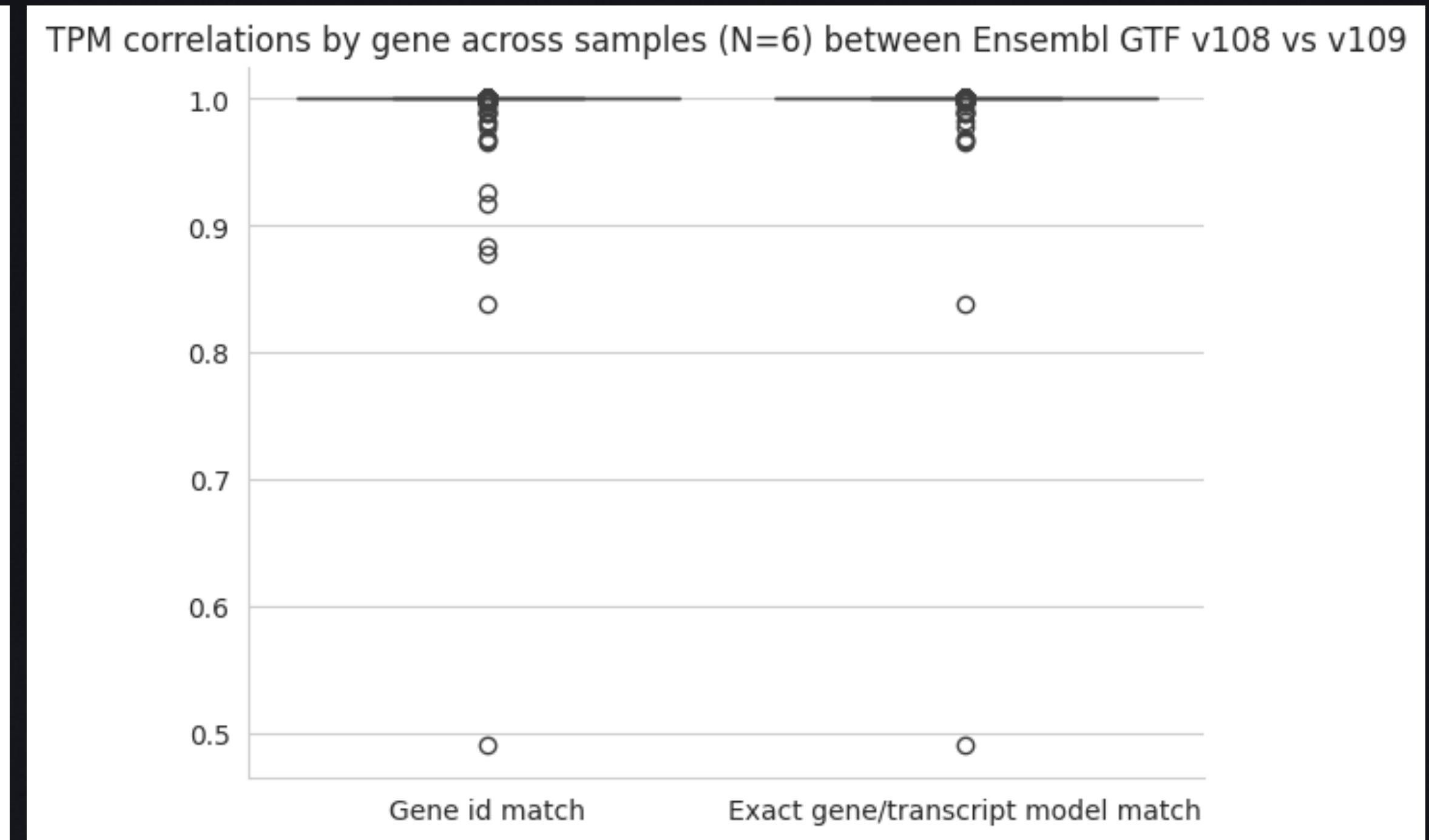
Matching gene ids: 3219

Jaccard by matching gene ids: 0.94

Identical genes: 1235

Identical genes with different ids : 13

Jaccard by exact model: 0.23



Matching gene ids: 3361

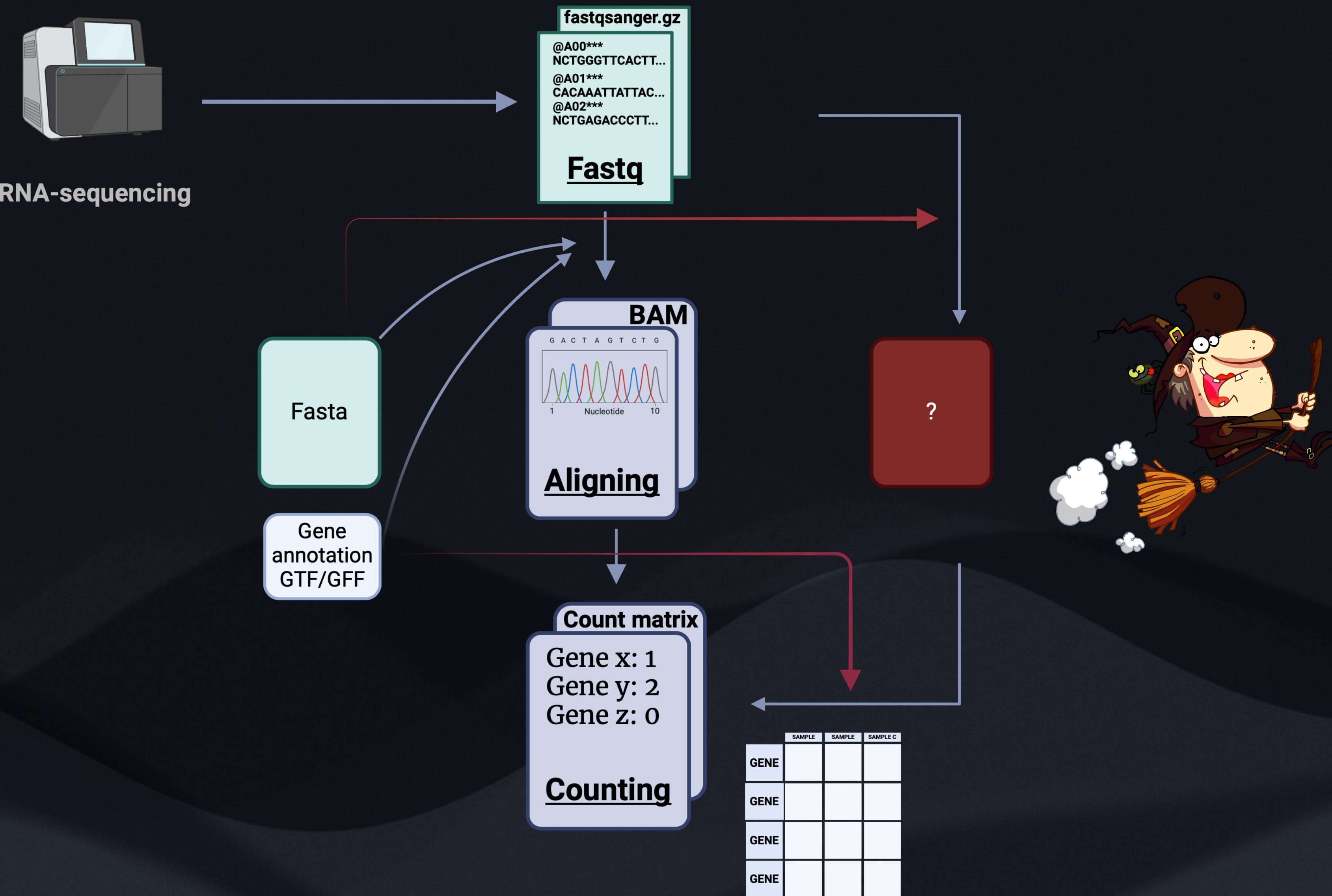
Jaccard by matching gene ids: 1.0

Identical genes: 3347

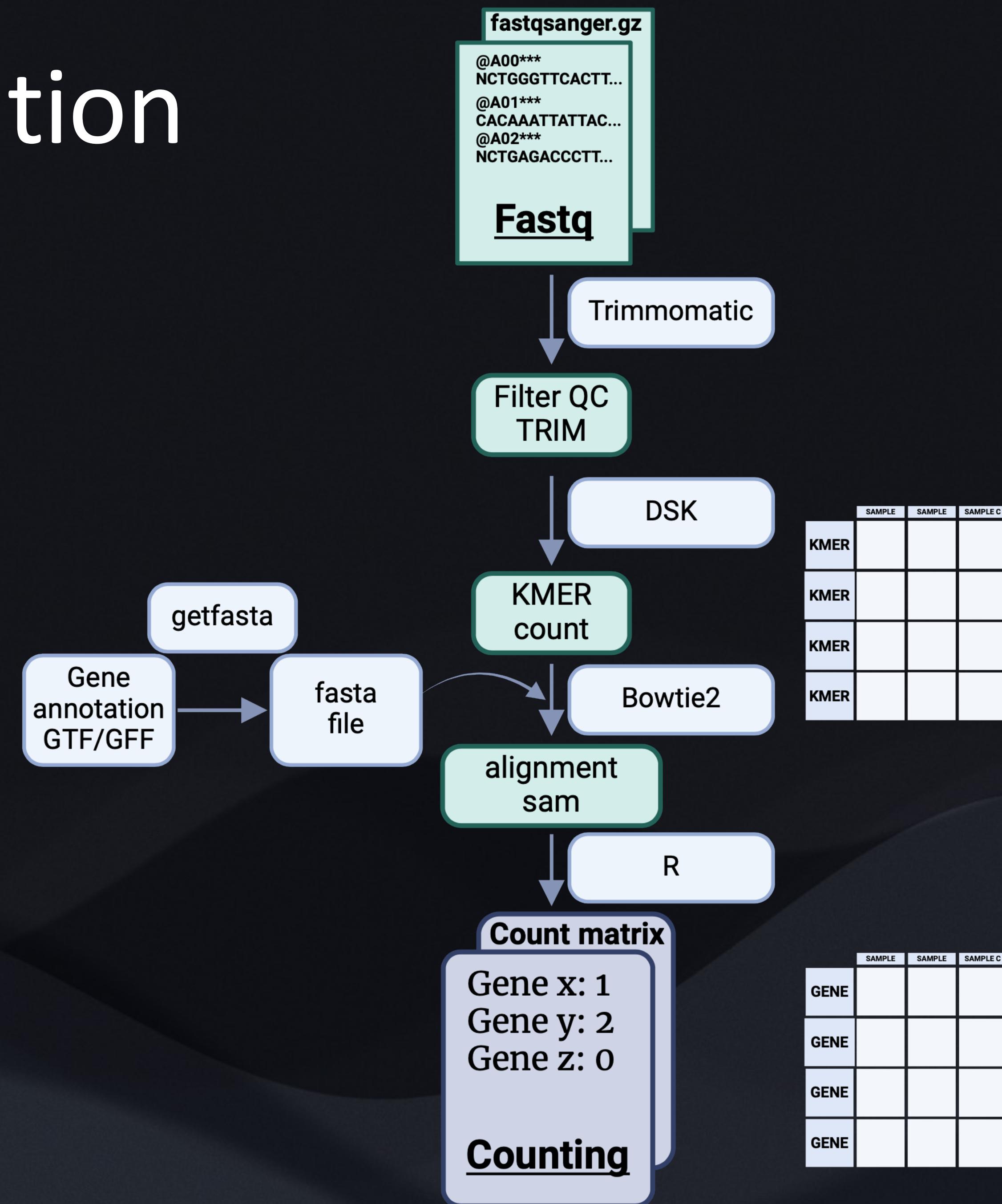
Identical genes with different ids : 0

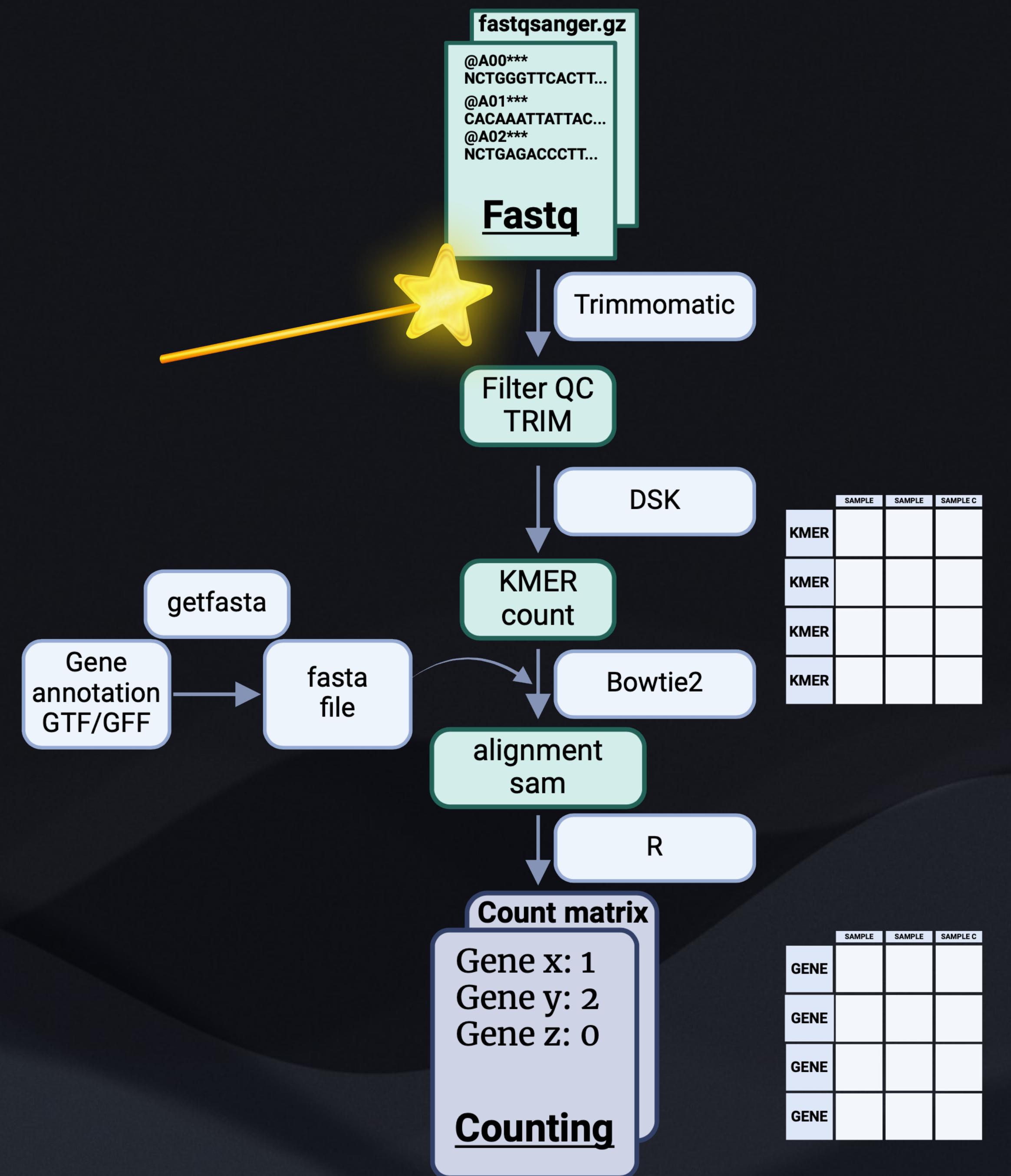
Jaccard by exact model: 0.99

The Problem

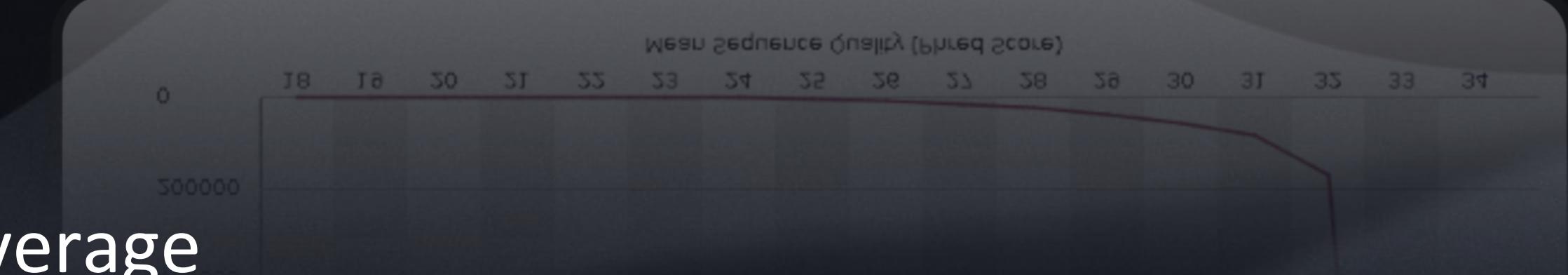
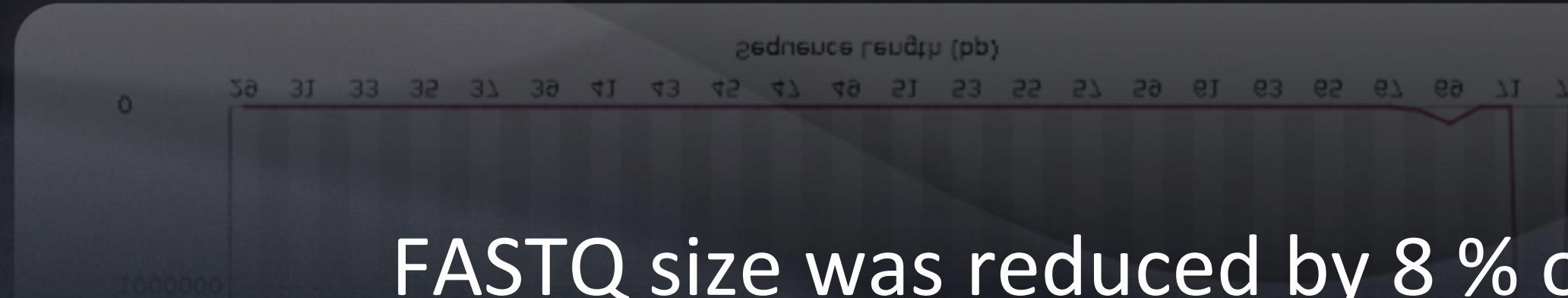
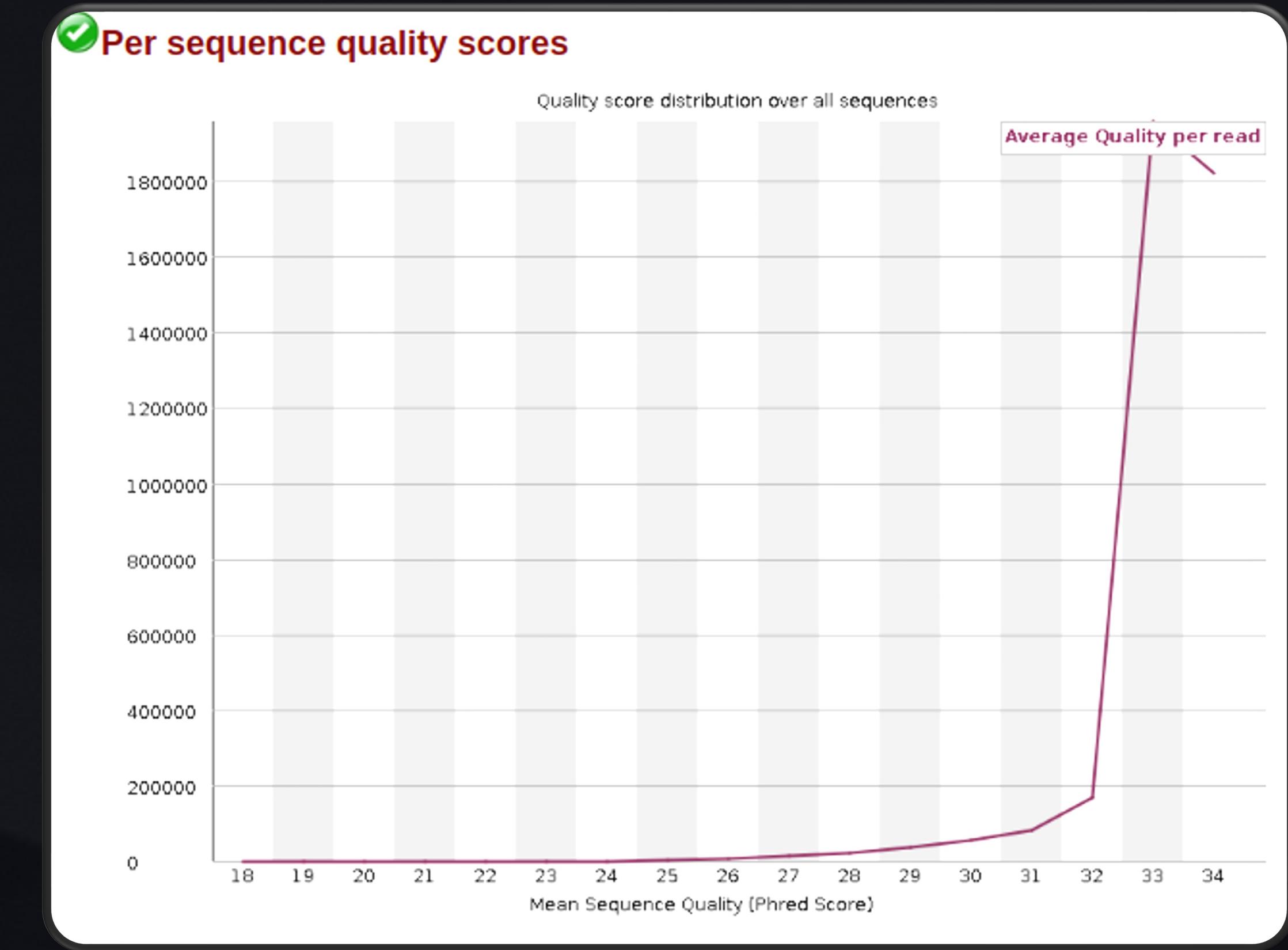
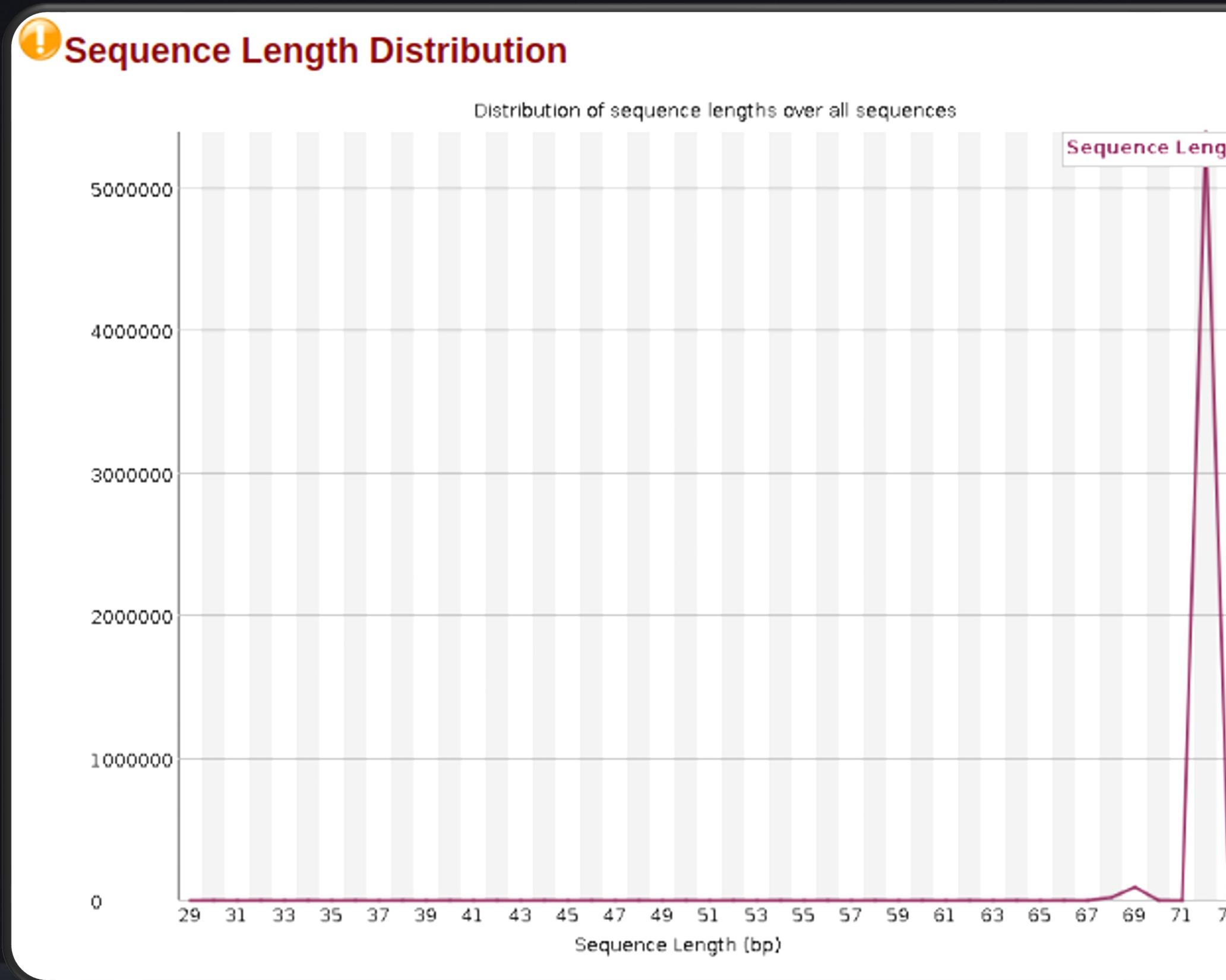


Our Solution



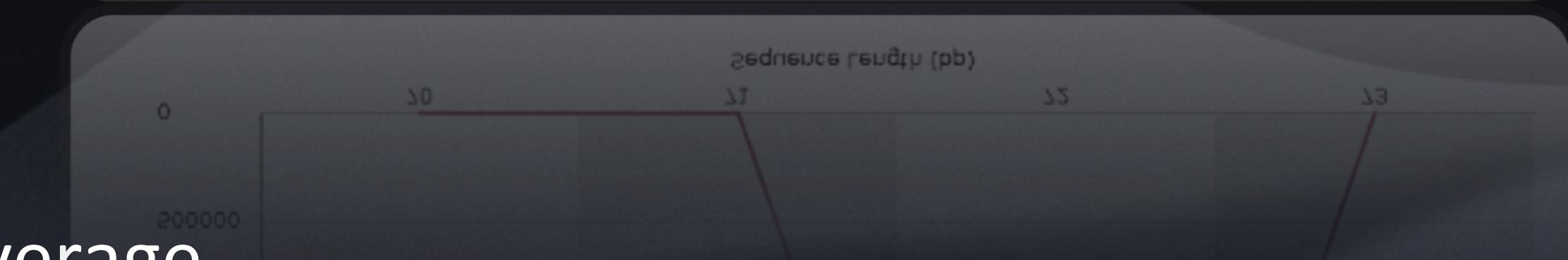
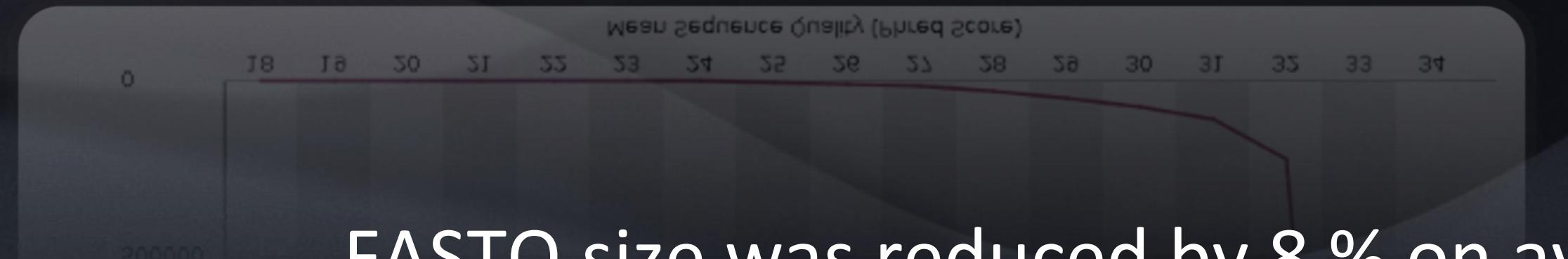
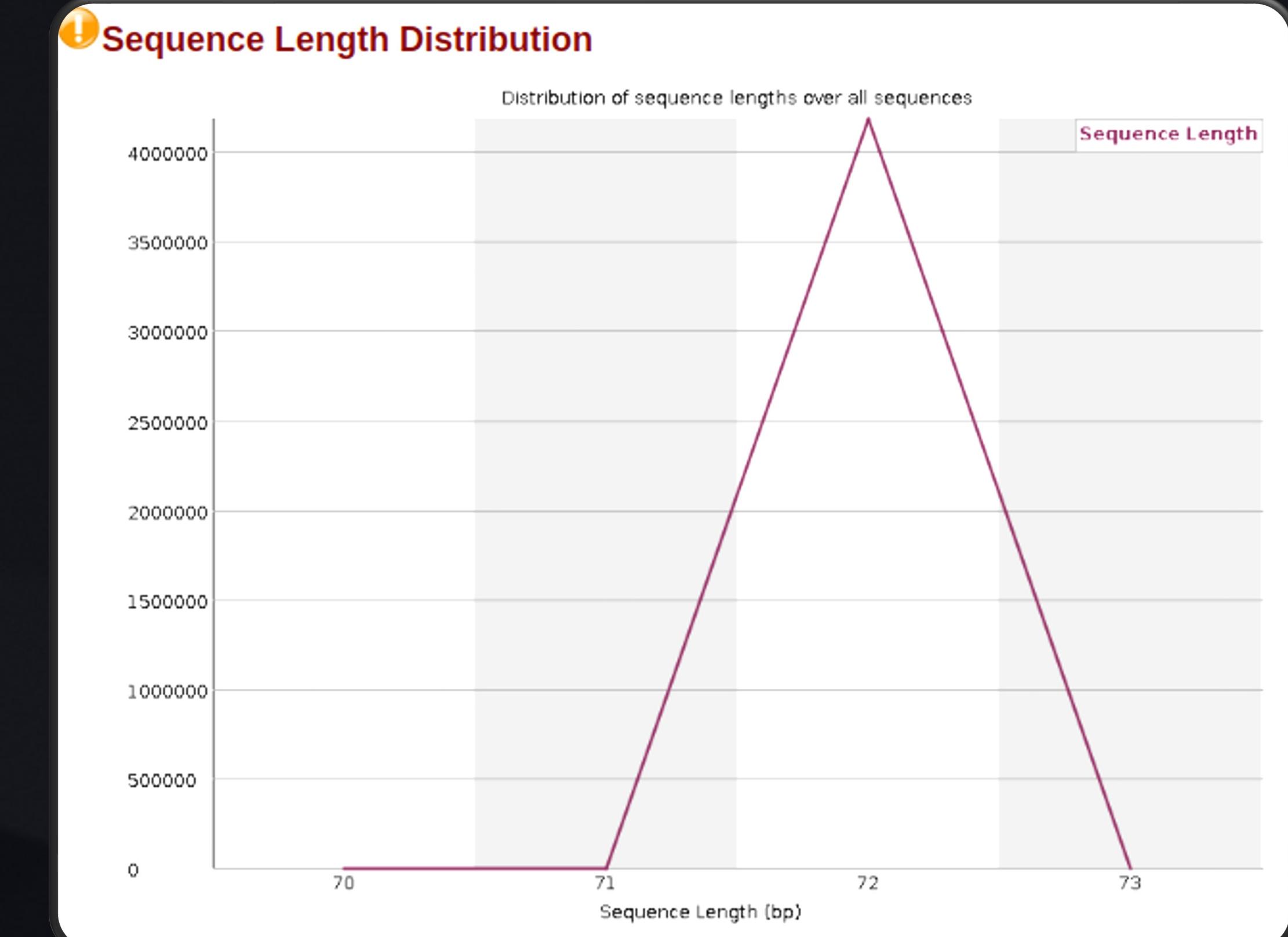
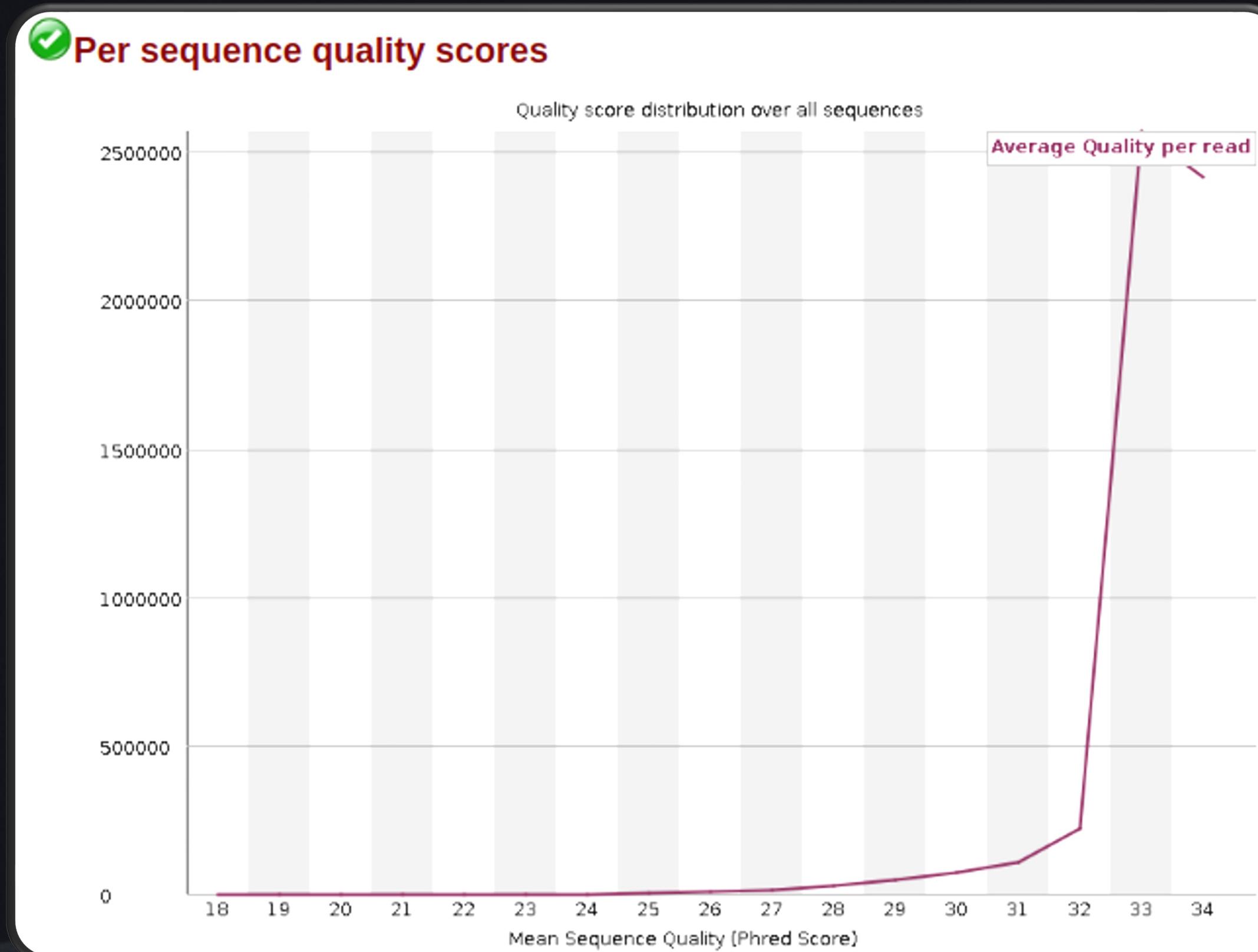


1. Filter sequences with a 2 bp window with quality <= 32



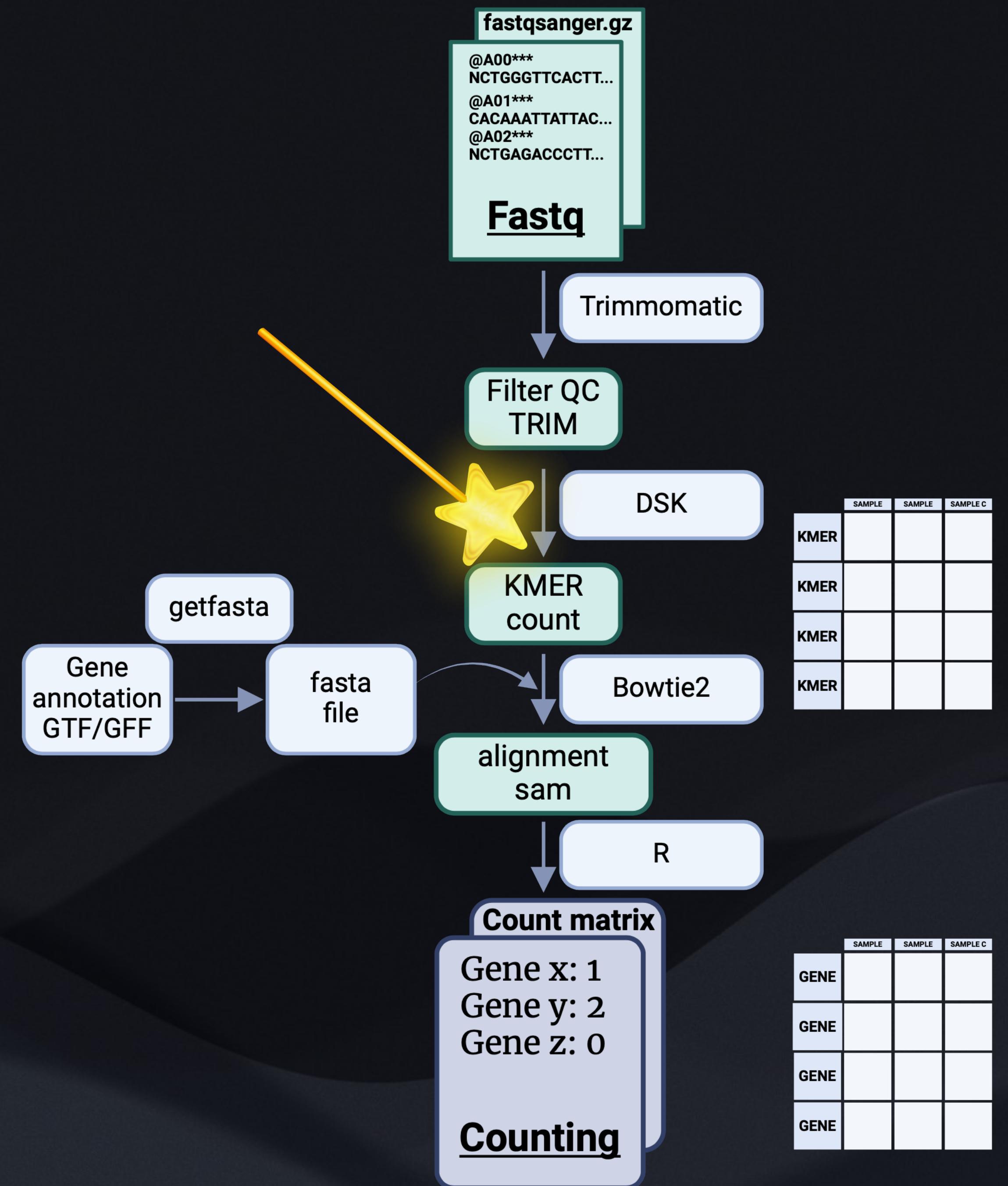
FASTQ size was reduced by 8 % on average
trimmomatic PE ... CROP:72 MINLEN:72 SLIDINGWINDOW:32:2

2. Crop reads to 72 bp and filter out reads < 72 bp

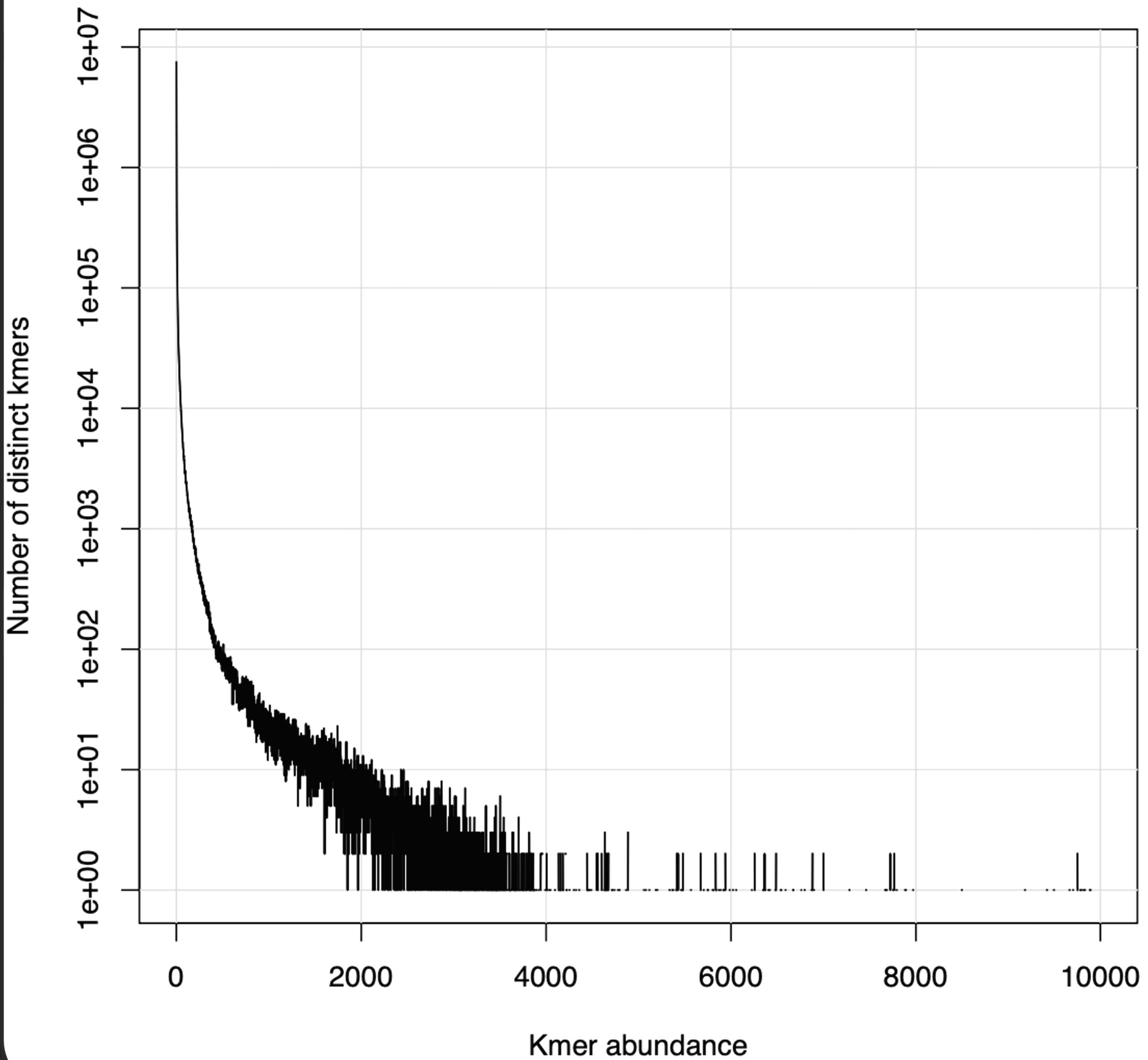


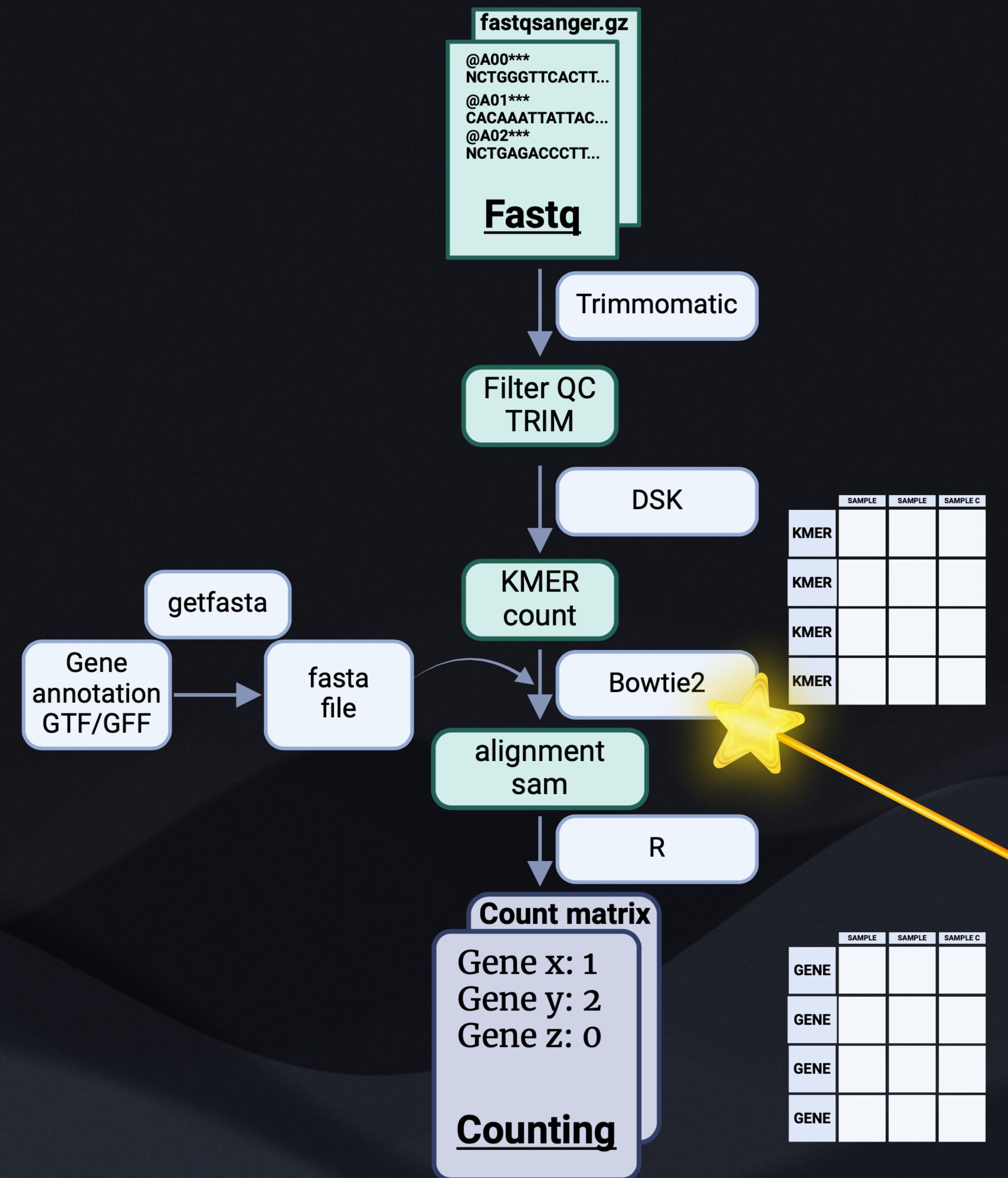
FASTQ size was reduced by 8 % on average

trimmomatic PE ... CROP:72 MINLEN:72 SLIDINGWINDOW:32:2



Kmer profile





V86

540228 reads; of these:

540228 (100.00%) were unpaired; of these:

135945 (25.16%) aligned 0 times

282874 (52.36%) aligned exactly 1 time

121409 (22.47%) aligned >1 times

74.84% overall alignment rate

V109

540228 reads; of these:

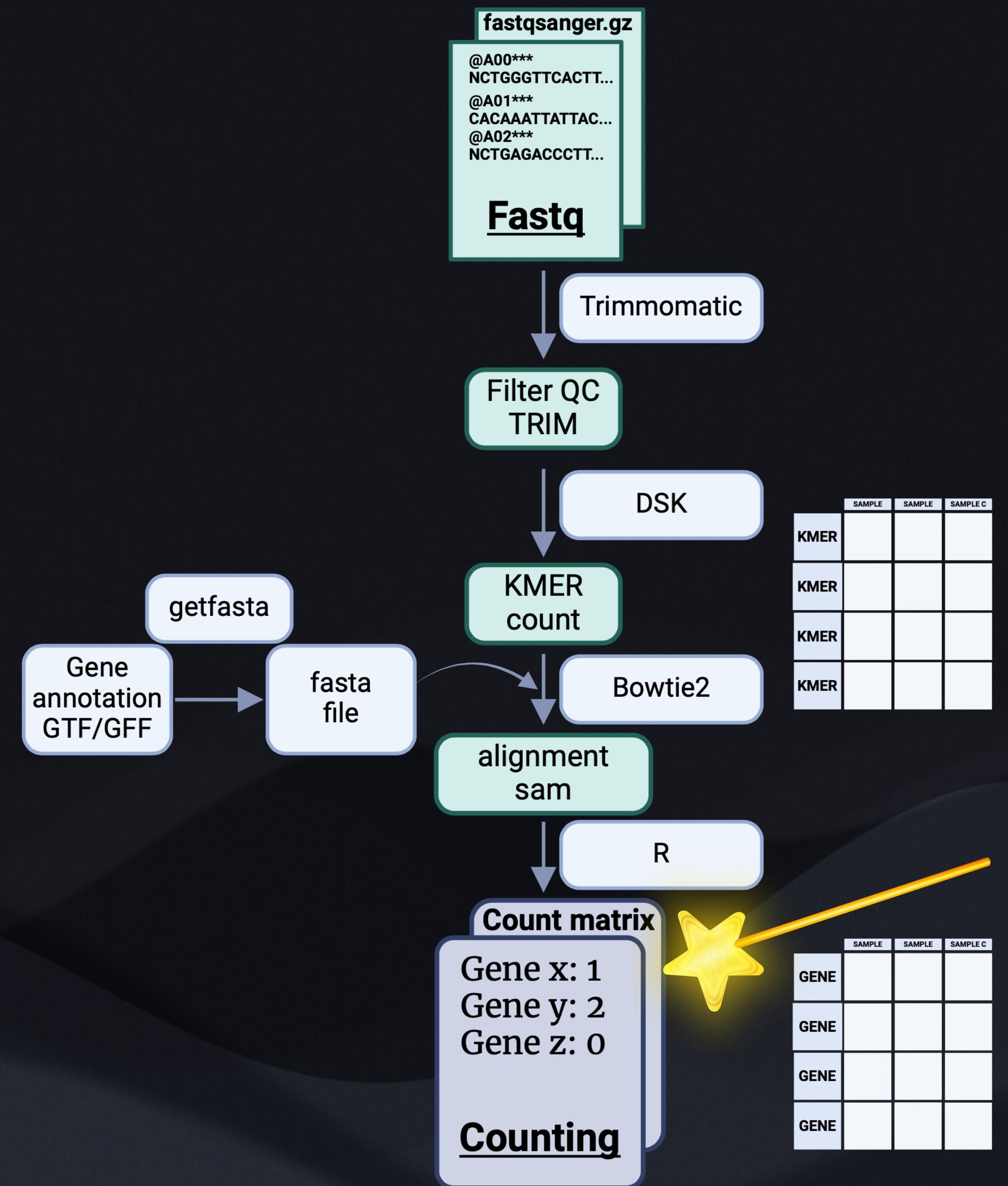
540228 (100.00%) were unpaired; of these:

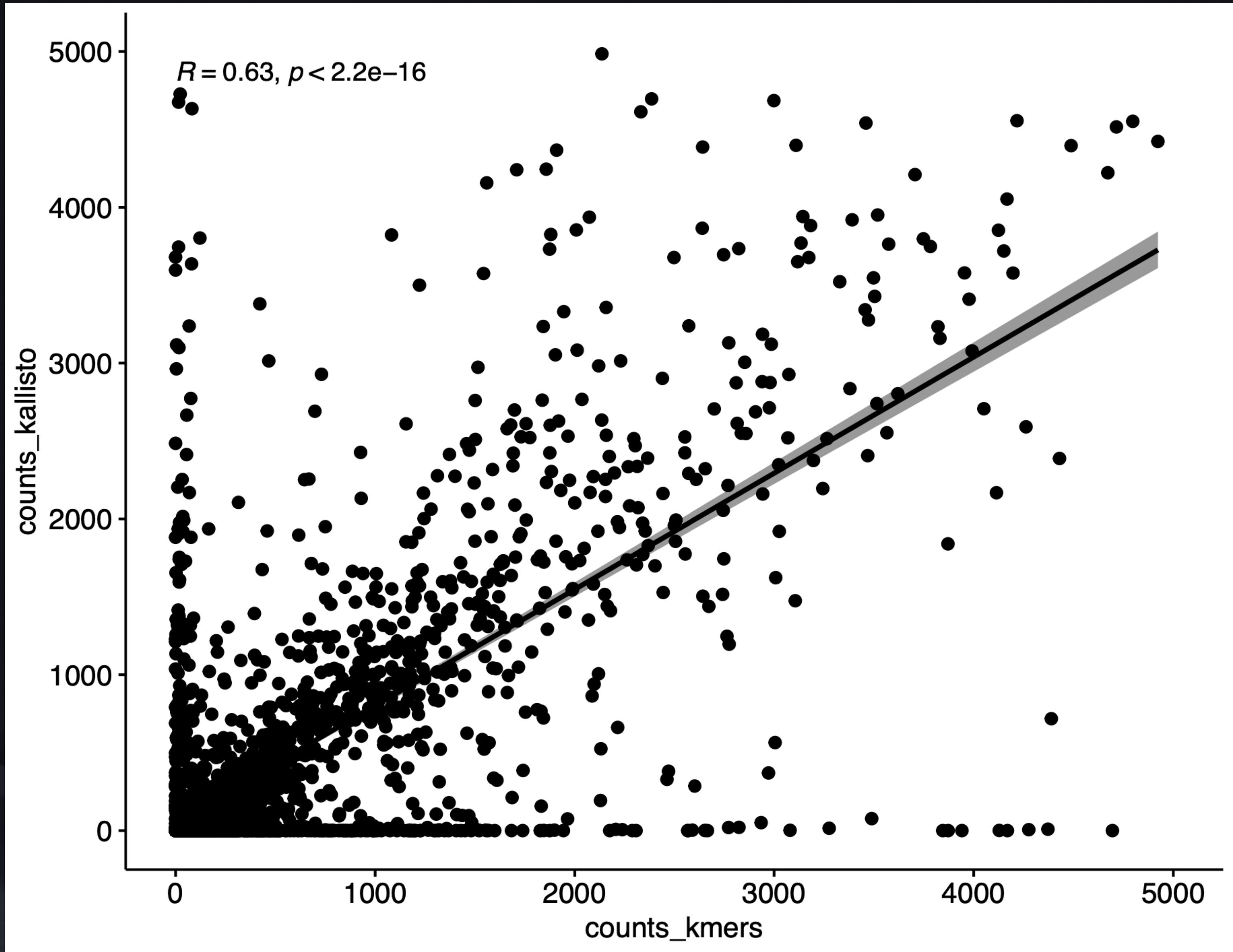
134058 (24.82%) aligned 0 times

270681 (50.10%) aligned exactly 1 time

135489 (25.08%) aligned >1 times

75.18% overall alignment rate





Thank you for your attention!

Supplementary