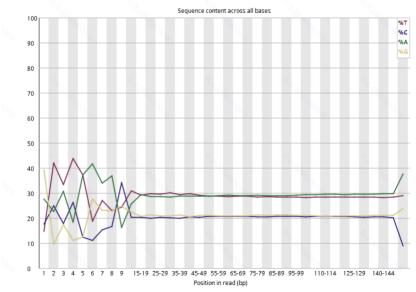
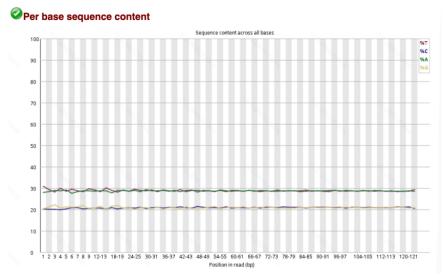
до тримминга получаем такое:

Per base sequence content - колебания содержания нуклеотид в начале и немного в конце цепочки.



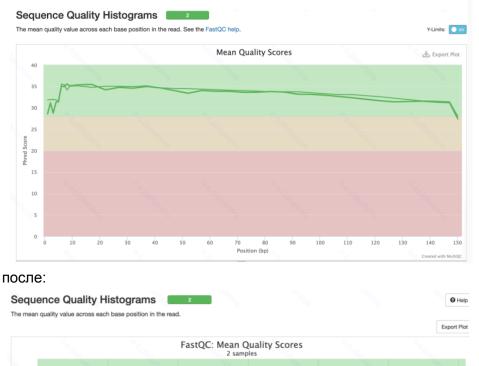


анализируя график, можно предложить подрезать 15 оснований в начале и 10 с конца, адаптеры нам не нужны при выравнивании на геном далее. после тримминга:



после тримминга увеличилась доля с высоким качеством (отрезали то, что давало плохое качество):

до:





именно поэтому и будем в дальнейшем использовать обрезанные данные. при помощи STAR, используя параметр для получения каунтов, выровняем риды на геном человека. далее применяем stringtie к полученному bam-файлу, чтобы получить модель транскриптов. файл gtf — это файл вложенных структур, в нем находится аннотация геномных особенностей (ген->транскрипт->экзоны/CDS) и их начальные и конечные позиции, а также айдишники, аттрибуты и другие характеристики, потому что один ген может кодировать несколько различных транскриптов. транскрипты могут отличаться количеством экзонов, типами экзонов и тд.

```
1 StringTie transcript 633888 634233 1000 ... gene_id "STRG.1"; transcript_id "STRG.1.1"; cov "13.426302"; FPKM "42.353672"; TPM "1358.578491";
1 StringTie exon 633888 634233 1000 ... gene_id "STRG.1"; transcript_id "STRG.1.1"; exon_number "1"; cov "13.426302";
1 StringTie transcript 7984894 7985251 1000 ... gene_id "STRG.2"; transcript_id "STRG.2.1"; cov "4.846369"; FPKM "15.288016"; TPM "490.393616";
1 StringTie exon 7984894 7985251 1000 ... gene_id "STRG.2"; transcript_id "STRG.2.1"; exon_number "1"; cov "4.846369";
1 StringTie transcript 8863890 8864092 1000 ... gene_id "STRG.3"; transcript_id "STRG.3.1"; cov "8.019705"; FPKM "25.298401"; TPM "811.496643";
1 StringTie exon 8863890 8864092 1000 ... gene_id "STRG.3"; transcript_id "STRG.3.1"; exon_number "1"; cov "8.019705";
1 StringTie transcript 866279 8866501 1000 ... gene_id "STRG.4"; transcript_id "STRG.4.1"; cov "5.739910"; FPKM "18.106720"; TPM "580.809082";
1 StringTie transcript 866279 8866501 1000 ... gene_id "STRG.4"; transcript_id "STRG.4.1"; cov "5.739910";
1 StringTie transcript 866279 8866501 1000 ... gene_id "STRG.4"; transcript_id "STRG.5.1"; cov "5.739910";
1 StringTie transcript 866279 8866501 1000 ... gene_id "STRG.4"; transcript_id "STRG.5.1"; cov "5.739910";
1 StringTie transcript 866279 8866501 1000 ... gene_id "STRG.4"; transcript_id "STRG.5.1"; cov "5.739910";
1 StringTie transcript 866270 866501 1000 ... gene_id "STRG.5"; transcript_id "STRG.5.1"; cov "5.739910";
```