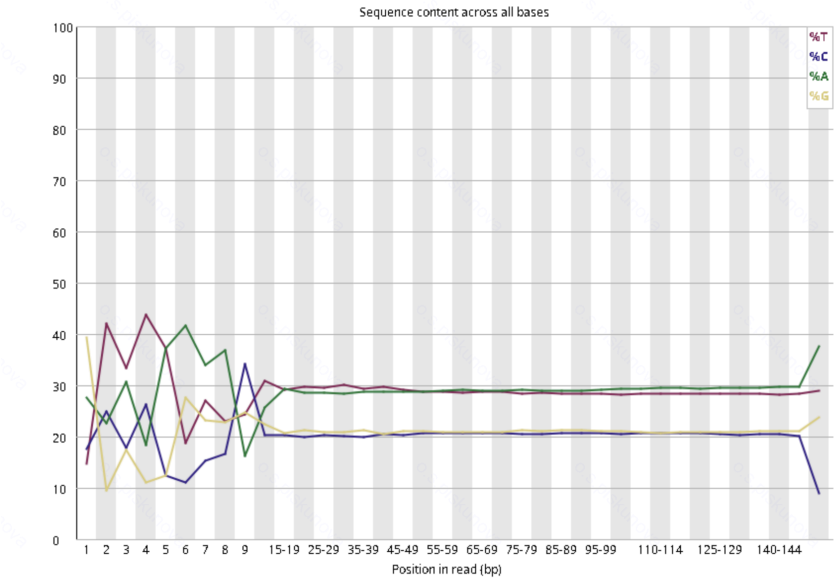
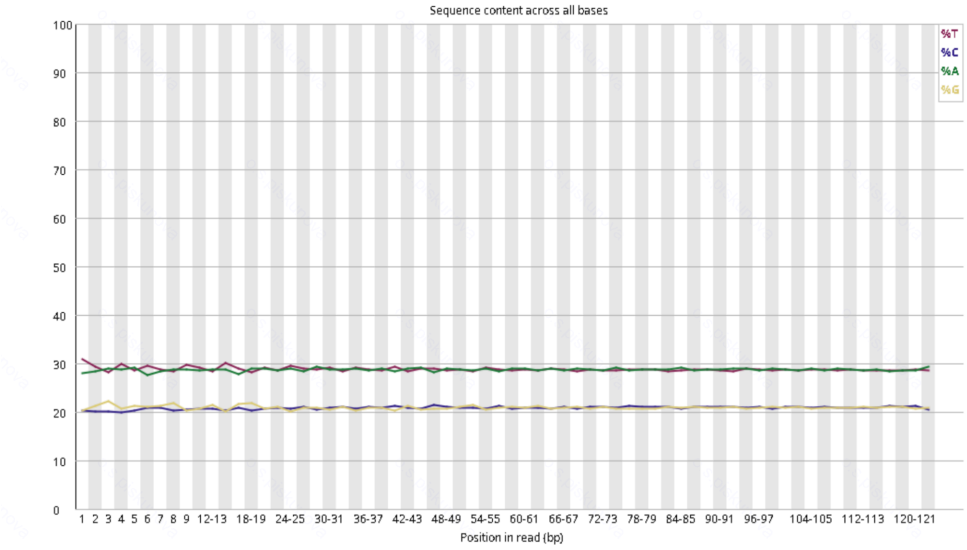
При первичном контроле качества получаем одну ошибку и одно предупреждение:

Per base sequence content

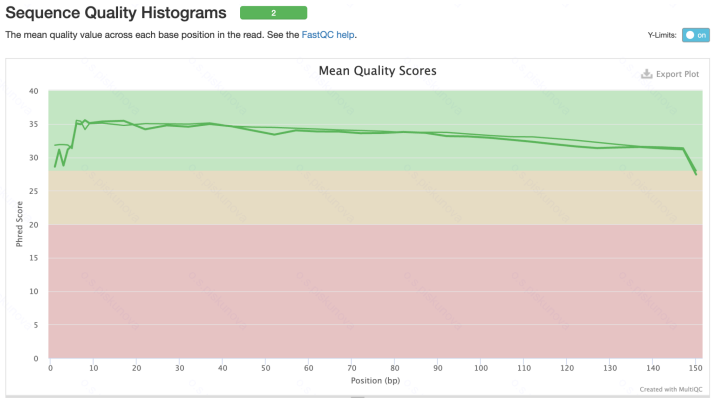


С помощью тримминга будем избавляться от ошибки Per base sequence content — колебаний содержания нуклеотидов в начальных позициях рида.
Анализируя график, видим, что можно отрезать адаптеры спереди и потримммить концы.
— зададим fastp параметры -f 15 -t 10 — отрезать 15 оснований спереди и 10 с хвоста рида (длина до 151). После тримминга ошибки нет и нет артефактов адаптеров на графике (мы не хотим выравнивать адаптеры на геном далее):

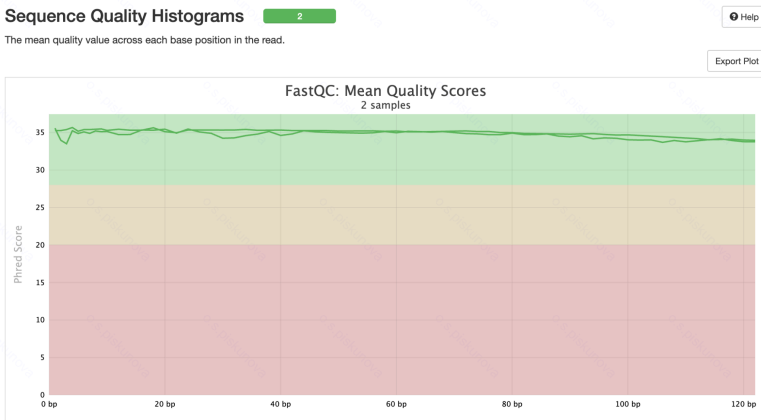
Per base sequence content



После тримминга у нас закономерно увеличилась доля оснований с более высоким качеством (мы отрезали то, что давало более низкое качество).
До



После



По совокупности этих причин будем использовать триммированные риды при дальнейшем анализе. С помощью 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. Тк 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	8866279	8866501	1000	.	.	gene_id "STRG.4"; transcript_id "STRG.4.1"; cov "5.739910"; FPKM "18.106720"; TPM "580.809082";
1	StringTie	exon	8866279	8866501	1000	.	.	gene_id "STRG.4"; transcript_id "STRG.4.1"; exon_number "1"; cov "5.739910";
1	StringTie	transcript	26281058	26281509	1000	.	.	gene_id "STRG.5"; transcript_id "STRG.5.1"; cov "15.070796"; FPKM "47.541279"; TPM "1524.98
```