

## Отчёт. 10 задание.

1) Были получены файлы с сырыми последовательностями секвенирования - Прочтения - *illumina\_reads\_R1\_001.fastq*, *illumina\_reads\_R2\_001.fastq*.

2) Был проведён **fastqc** анализ.

а) Для *illumina\_reads\_R1\_001.fastq*:

Во-первых, в модуле "**Per base sequence quality**" видно, что качество нуклеотидов начинает заметно снижаться ближе к концу прочтений — примерно после 60-й позиции -> **лучше обрезать концы**.

Во-вторых, в модуле "**Overrepresented sequences**" обнаружены повторяющиеся последовательности, что указывает на присутствие остатков адаптеров -> **удаляем**. (см. Fig.1)

б) Для *illumina\_reads\_R2\_001.fastq*:

Аналогично. Низкое качество сиквенса и наличие адаптеров.(см. Fig.2)

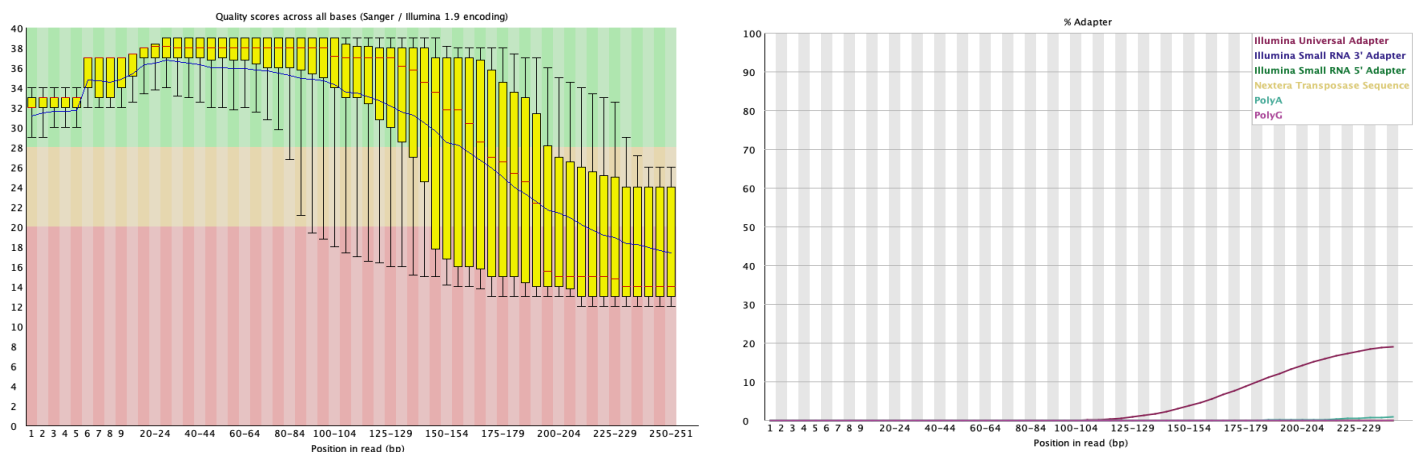


Figure 1. Per base sequence quality и Overrepresented sequences для *illumina\_reads\_R1\_001.fastq*

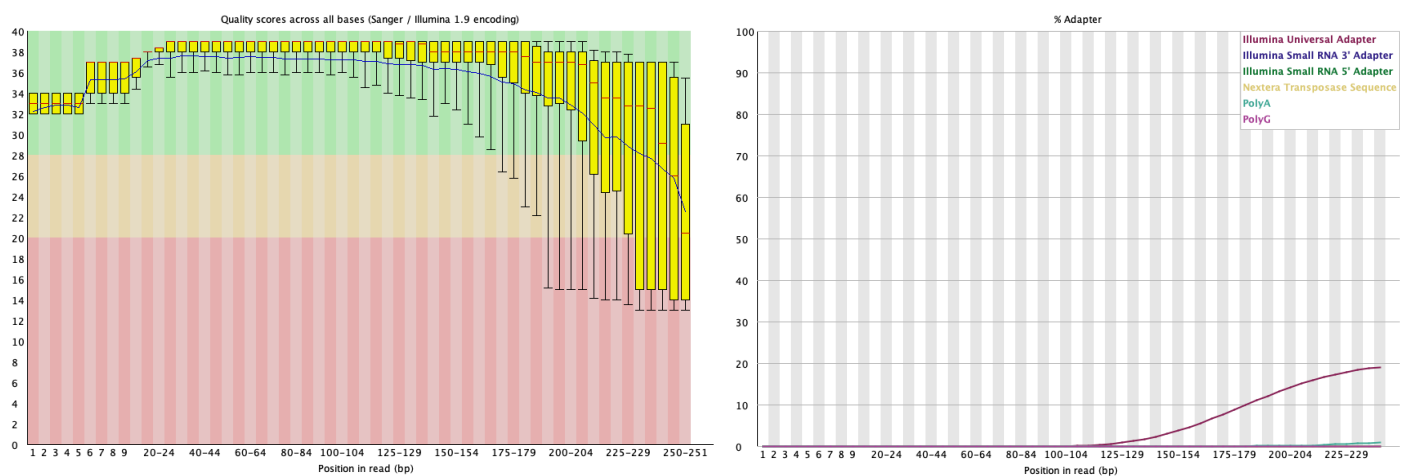


Figure 2. Per base sequence quality и Overrepresented sequences для *illumina\_reads\_R2\_001.fastq*

3) Проводили **тримминг** адаптеров и избавлялись от низкокачественных оснований и коротких ридов. Использовали **trimmomatic**.

**command:**

```
trimmomatic PE \
-threads 4 \ # используется 4 потока процессора;
illumina_reads_R1_001.fastq illumina_reads_R2_001.fastq \
new_R1.fastq.gz unpaired_R1.fastq.gz \
new_R2.fastq.gz unpaired_R2.fastq.gz \
ILLUMINACLIP:adapters.fa:2:30:10 \ # 2 - количество допустимых
mis-match мутаций; 30 - пороговый скор для обрезки; 10 - минимальное
совпадение;
LEADING:3
TRAILING:3
SLIDINGWINDOW:4:20
MINLEN:36 # отбросить риды, если после обрезки осталось 36
нуклеотидов.
```

4) Проводился повторный fastqc анализ. После тримминга, все модули показывали удовлетворительное качество. Per Base Sequence quality и Overrepresented sequences также стали приемлемы (см. Fig. 3)

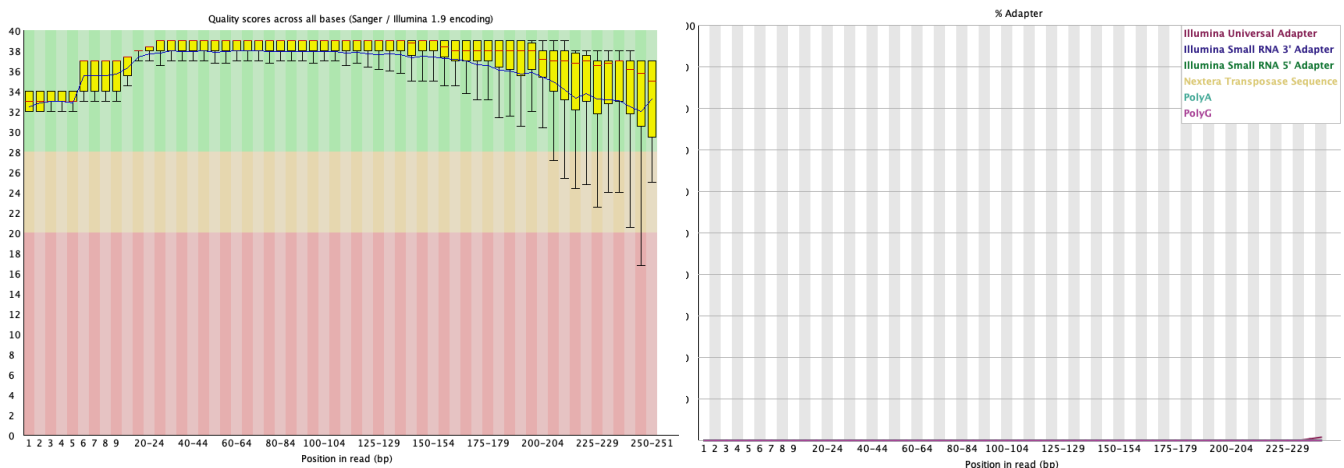


Figure 3. Per base sequence quality и Overrepresented sequences для illumina\_reads\_R1\_001.fastq после тримминга