

Задание 13

Трещева Мария

1. Получение отчетов fastqc и multiqc и выбор файлов для выравнивания.

Код в slurm-скриптах. Посмотрим на отчеты multiqc до и после тримминга.

До тримминга:

multiqc Eg_Treg_S71_R1_001_fastqc.zip Eg_Treg_S71_R2_001_fastqc.zip -n multiqc_before.html

Sample Name	Dups	GC	Seqs
Eg_Treg_S71_R1_001	27.7 %	45.0 %	19.8 M
Eg_Treg_S71_R2_001	21.2 %	46.0 %	19.8 M

multiqc Eg_Treg_S71_R1_trimmed_fastqc.html Eg_Treg_S71_R2_trimmed_fastqc.html -n multiqc_trimmed.html

После тримминга:

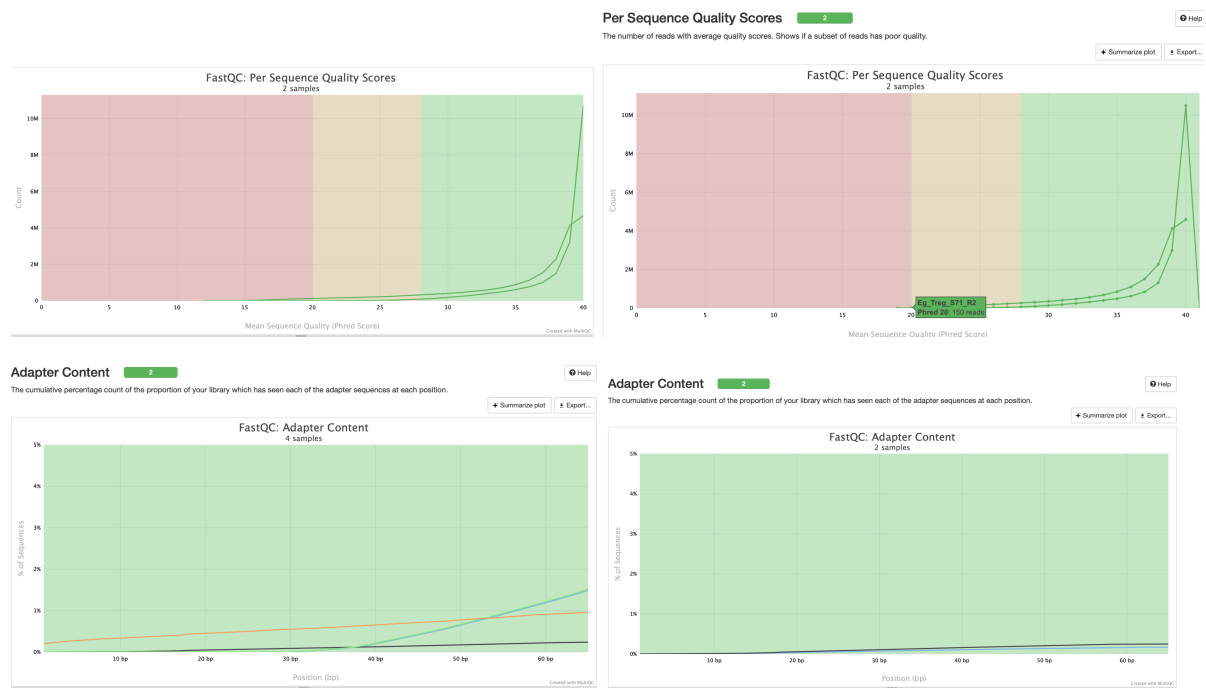
General Statistics

Copy tableConfigure columnsScatter plotViolin plotExport as CSV...Showing 2/2 rows and 3/6 columns.Summarize table

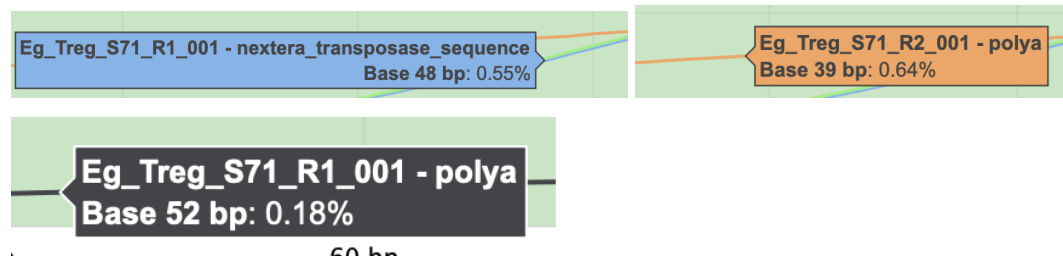
Sample Name	Dups	GC	Seqs
Eg_Treg_S71_R1	24.5 %	46.0 %	18.2 M
Eg_Treg_S71_R2	22.0 %	46.0 %	18.2 M

До/после:





До:

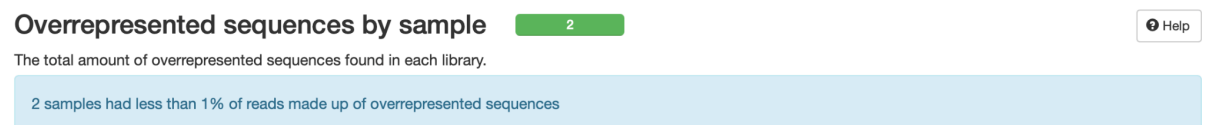


Есть незначительная перепредставленность адаптера:

Последовательность транспозазы Nextera (до 1,5%)

Последовательность PolyA (до 0,96%)

После:



Top overrepresented sequences

Top overrepresented sequences across all samples. The table shows 20 most overrepresented sequences across all samples, ranked by the number of samples they occur in.

Что изменилось после обработки fastp:

1. Удалены остаточные адаптеры
2. Удалены поли-хвосты

Видим, что вообще нет перепредставленных последовательностей, значит тримминг улучшил наши секи.

ИТОГ: Поскольку некоторые адаптеры оставались в изначальном датасете, и раз уж мы сделали тримминг, **все-таки воспользуемся результатами после тримминга**, хотя и исходные данные - хорошего качества.

2. Анализ выравнивания

Состав файла GTF

Каждая строка содержит 9 столбцов:

1. номер хромосомы
2. программа-генератор, у нас везде StringTie
3. тип элемента: transcript — описание транскрипта, exon — описание экзона
4. начальная позиция
5. конечная позиция
6. условный показатель качества
9. метаданные;

В метаданных содержится:

ID гена, ID транскрипта, ID из референсной аннотации (если есть), название гена, покрытие ридами, уровни экспрессии.

Для экзона еще порядковый номер экзона. Транскрипты с префиксом Eg_Treg_S71 — новые, предсказанные StringTie.