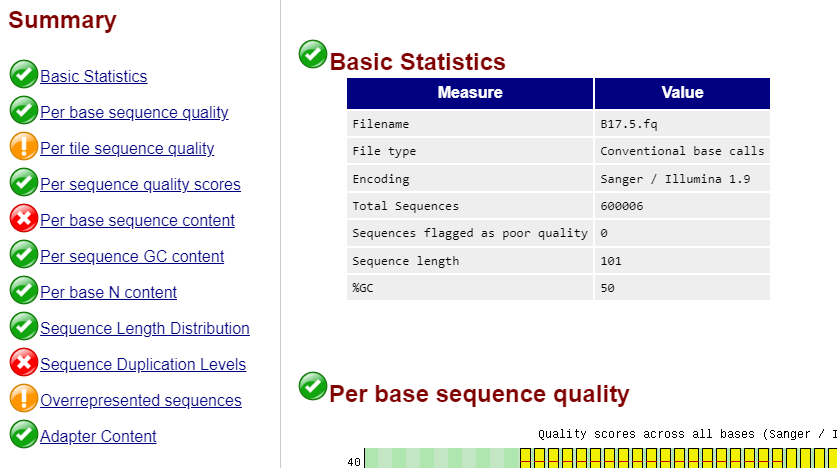
Ссылка на гугл коллаб

<https://colab.research.google.com/drive/1KJC11eDaCoTC1EQp4I5wj3rMK3tJVv9Q?usp=sharing>

Задания

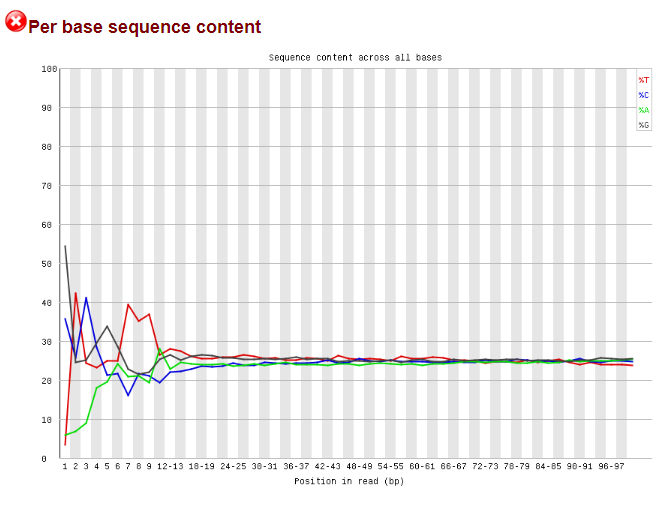
**1) Проверьте все входные данные при помощи fastqc.**

Пример анализа fastQC отчета для образца B17.5:

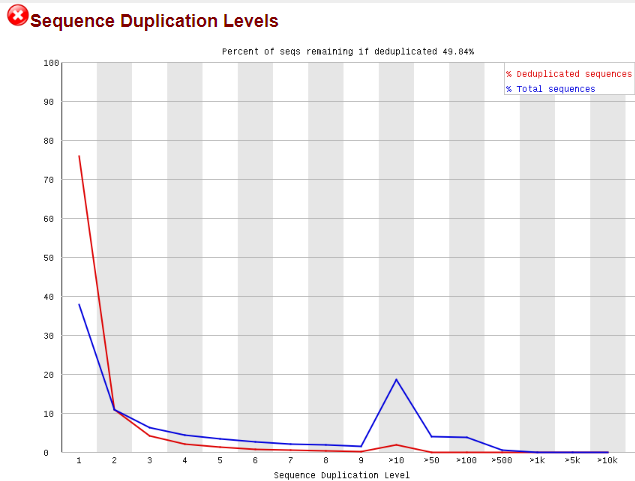


Из важного:

Наблюдается сдвиг в нуклеотидном распределении в первых 10 нуклеотидах ридов (в норме каждый нуклеотид встречается равновероятно, т.е. по 25%). Эта ошибка является нормой для РНК-секвенирования и возникает при подготовке библиотеки (<https://sequencing.qcfail.com/articles/positional-sequence-bias-in-random-primed-libraries/> ).



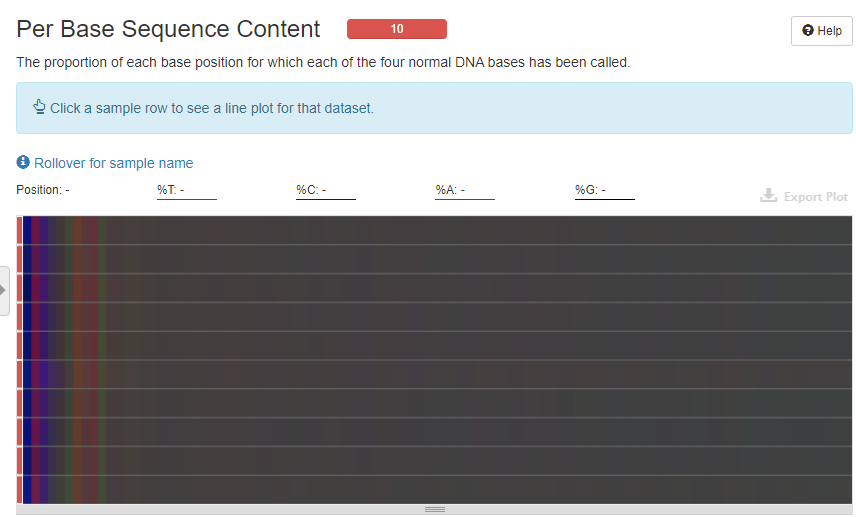
Ожидаемо высокое количество повторяющихся прочтений. Для RNA-Seq это может быть связано с активно транскрибирующимися РНК.

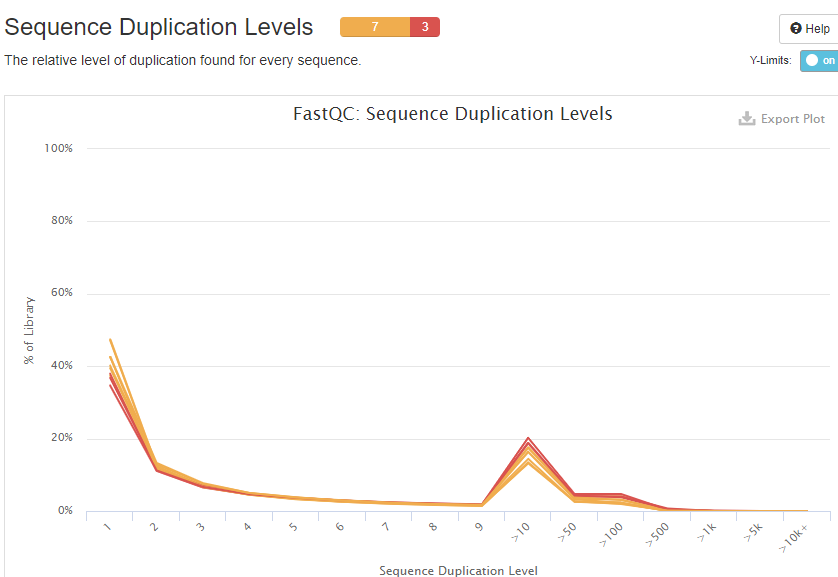
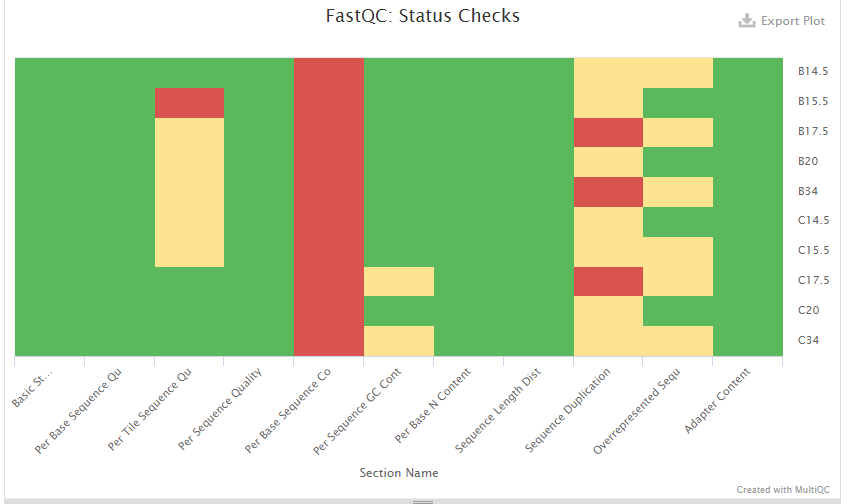


Отчет multiQC для всех образцов:

Изображение выглядит как стол

Автоматически созданное описание



**6)Прокартируйте все fq файлы (начните с одного) на 19ую хромосому при помощи hisat2 не допуская обрезания ридов и сообщив hisat2 координаты сайтов сплайсинга**

Пример для одного образца B17.5:

600006 reads; of these:

600006 (100.00%) were unpaired; of these:

75326 (12.55%) aligned 0 times

515175 (85.86%) aligned exactly 1 time

9505 (1.58%) aligned >1 times

87.45% overall alignment rate

**7)Выберите случайно один образец (был выбран B17.5)**

**8)Сколько ридов картируется в регион 19:12485000-12490000 в этом образце?**

В данный регион не откартировалось ни одного рида, ну и соответственно дальнейшее выполнение задания для данного региона невозможно. Взяла границы 19:12485000-12510000, в которых было откартировано 12 ридов

**9)Сколько из них картируются только в одно место генома?**

Все

**10)Сколько ридов картровалось без замен? Сколько с 1, 2 и т. д. заменами?**

Без замен 11 ридов

С 1 заменой 1 рид

**11)Сколько ридов картировалось на экзон-экзонные границы? Перечислите координаты всех интронов в данном интервале подтверждённых хотя бы одним ридом в формате: chr:from-to coverage**

4 рида картировались в один экзон (chr19:12506118-12507903)

остальные 8 ридов, видимо, ложатся в интроны

На экзон-экзонные границы не лег ни один рид

границы интронов:

chr19:12496187-12497112

chr19:12497289-12500605

chr19:12500939-12506118 (в этот интрон ложатся все оставшиеся 8 ридов)