

End Joining Signatures - dev

Charlie Soeder

2/15/2019

25 February 2019

Rebuilding, starting with summary stats for the materials/methods section.

Reference genomes

Table 1: Size and Consolidation of Reference Genomes

Reference Genome:	dm6
number_bases	144 M
number_contigs	1.87 k

Sequenced reads

Table 2: Number of Sequenced Samples by Treatment

experimental	sample_count
control	15

Table 3: Sequenced Experimental Samples

name	paired	experimental	source
Mcm5-A7	TRUE	control	dannyMiller
mcm5-13	TRUE	control	dannyMiller
mcm5-12	TRUE	control	dannyMiller
mcm5-11	TRUE	control	dannyMiller
mcm5-10	TRUE	control	dannyMiller
mcm5-09	TRUE	control	dannyMiller
mcm5-08	TRUE	control	dannyMiller
mcm5-07	TRUE	control	dannyMiller
mcm5-06	TRUE	control	dannyMiller
mcm5-05	TRUE	control	dannyMiller
mcm5-04	TRUE	control	dannyMiller
mcm5-03	TRUE	control	dannyMiller
mcm5-02	TRUE	control	dannyMiller
mcm5-01	TRUE	control	dannyMiller
DfMcm5	TRUE	control	dannyMiller

Total Starting Reads: 1.97G Post-QC Reads: 1.85G.

Table 4: Read Count and Percent Retention

type	minimum	average	maximum
prefiltered	45.8 M	65.8 M	75.7 M
postfiltered	42.7 M	61.6 M	71.1 M
percent retention	92.4	93.6	95.5

This framework is general-purpose enough that it might be a good template.....

Dupes:

Table 5: Percentage Duplication

minimum	average	median	maximum
4.2	5.3	5.3	6.2

