Weller Day2 Session 2 - Sequencing QC and Assembly Script

Sequencing Technologies – part 2

# Quality Scores

>Labels Illumina etc. AGACCCAATTTTGGTTTGATAGTACATCCCAACAAAGGACGGCCATATTTGTTNNATTTATCTCTTTCAACTTGGATACCATGAGGCGGACCTTGAAATG

+

;??ABC;DDDDDDDDDDDDDDDDDDDDIDDDDDDDDDDDDDDDDDDDDJDDDDDDDDDDDDDDDDDDDDBDDDDDDDDDDDDDDDDDDDDCBDDD??;;8

>Labels Illumina etc.

CCATGAGGCGGACCTTGAAATGTTTTGGAATAAGCAGGGGGAATCCGCAAATCTTCCAAACGTAGAGCCCGTAGGGCCTTGAATCCAAATACATTACCTA

+

;??ABC;44567;DDDDDDDDDDDDDDIDDDDDDDDDDDDDDDDDDDDJDDDDDDDDDDDDDDDDDDDDBDDDDDDDDDDDDDDD>>=<<:::99??;;8

>Labels Illumina etc.

GAATCCAAATACATTACCTACAATGGAAGTGAACAAGTTAGTAACAGAACCTTCTTCGAAAAGGTCTAAGGGGTAAGCTACATAGGCAATAAATTGACTC

+

;??ABC;::99<<DDDDDDDDDDDDDDIDDDDDDDDDDDDDDDDDDDDJ89DDDDDDDDDDDDDDDDDDBDDDDDDDDDDDDDDDDDD99<<..???;;8

# Part 2

Assembly Fragments at 2X (known: length = 50, Fragments < 8nt are not sequenced)

Sequence 1

GATGCTACC TCGAGCCCGTT GCTGGAGAA GAAAGTCAATTTATTGCT GCTACCACA TCGAGCC CGTTGCT GAAGAAAGT CAATTTATTGCT

Sequence 2

CTGCAATAAT TAAATTTT GGATCTATTCT TTATGGATT GTGCTAAAG TAATTAAAT TTTGGATCTA TTCTGAGTTAT GGATTGTGCTAAAG

Sequence 3

ATTATAAAT ATTATACTCCT ATGAAACCAAA GATACTGAT ATTATAAATT ATTATACTCC TGACTATGA AACCAAAGATA

Sequence 4

GAAACCAAAGATA CTGATATCTTG GCAGCATTT CGAGTAACTCCTC GAAACCAAA GATACTGATAT CAGCATTTCGA CTCCTCAACC

Sequence 5

TCAACCTGGAGTT CCACCTGAAG CAGGGGCAGC GGTAGCTG AACTCCTCA ACCTGGAGTTCC AAGCAGGGG CAGCGGTAGCTG

Sequence 6

GGTAGCTGCTGAA TCTACTGGTA GACAACTGTATG GACTGACG GCTGCTGAATC ACTGGTACAT GTATGGACTGA

Sequence 7

ACTGACGGG CTTACTAGTCTT TTACAAAGGTCG GCTACCACAT CTGACGGGC TACTAGTCTTGAT TTACAAAGGTCGA TGCTACCACAT

To do 4X Coverage, add the following sets for each Sequence:

Sequence 1:

GATGCTACC GAGCCCGTTGC AAAGTCAATTTA GATGCTAC CACATCGAGC TGCTGGAGAAGA AAGTCAATTTA

Sequence 2:

GCAATAATTA AATTTTGGA TCTATTCTGA GTTATGGATTG CTGCAATAAT TAAATTTTGGAT CTATTCTGAGTT ATGGATTGTGCTAAAG

Sequence3:

ATTATAAATTG ACTTATTATACT ACTATGAAACC AAAGATACTGAT AAATTGACTTATTAT CCTGACTATGAA CAAAGATACTGAT

Sequence 4:

AAACCAAAG ATACTGATATCT GCATTTCGAGTAAC TCCTCAACC ACCAAAGATAC TGATATCTTGGCAGCA TTTCGAGTAAC TCCTCAACC

Sequence 5:

CAACCTGGAGTTCCACC AAGAAGCAGG GGCAGCGGTAG CCTCAACCTGGAGT CTGAAGAAGC AGGGGCA GCGGTAGCTG

Sequence 6:

GGTAGCTGCT AATCTTCTACTGG CATGGACAACTGT ATGGACTGAC TGCTGAATCTTCTA CTGGTACATGGACAACT GGACTGACG

Sequence 7:

CGGGCTTACTAG TCTTGATCGTTAC GGTCGATGCTACCACAT TAGTCTTGATCG TTACAAAGGTC GATGCTACCAC