Illumina data processing

Tong Liu

module load bioinfo-tools

module load Qiime/1.8.0

module load SeqPrep

module load cutadapt

#copy folder from service to self-created folder

cp -r "/gulo/glob/liutong/illumina\_project\_1/160105\_M00485\_0244\_000000000-AKNA5/Sample\_T1/" .

1) Trim off any remaining adapter, primer etc.

GTGBCAGCMGCCGCGGTAA for 16s forward primer

GACTACHVGGGTATCTAATCC for 16s reverse primer

cutadapt -g GTGBCAGCMGCCGCGGTAA -G GACTACHVGGGTATCTAATCC --max-n 0 --discard-untrimmed -q 10 --maximum-length 300 -o SameT.N0.dis.q10.M300.trimmed.1.fastq -p SameT.N0.dis.q10.M300.trimmed.2.fastq \*R1\_\*.fastq.gz \*R2\_\*.fastq.gz

2) Join the paired end reads.

join\_paired\_ends.py -f SameT.N0.dis.q10.M300.trimmed.1.fastq -r SameT.N0.dis.q10.M300.trimmed.2.fastq -m SeqPrep -o SameT.N0.dis.q10.M300.j150 -j 150

3) Cat all fastq into one fastq

#more info about SeqPrep: <https://github.com/jstjohn/SeqPrep>

cat \*/SameT.N0.dis.q10.M300.j150/seqprep\_assembled.fastq.gz >> all\_assembled.fastq.gz

4) Extra barcode

extract\_barcodes.py -f all\_assembled.fastq.gz -o barcode.fastq.gz -c barcode\_in\_label -l 16

5) Validate map

validate\_mapping\_file.py -m map\_tong\_test.txt -o map\_check

6) split\_libraries

split\_libraries\_fastq.py -o splited -i all\_assembled.fastq.gz -b barcode.fastq.gz/barcodes.fastq -m map\_tong\_test.txt --barcode\_type 16 --store\_qual\_scores --max\_bad\_run\_length 1 -q 19 -n 0 --max\_barcode\_errors 0

7)pick\_open\_reference\_otus.py

# own PC

pick\_open\_reference\_otus.py -i seqs.fna -o open\_otu\_pick/uclust\_97 -r /usr/local/lib/python2.7/dist-packages/qiime\_default\_reference/gg\_13\_8\_otus/rep\_set/97\_otus.fasta -s 0.1 -aO 2 --prefilter\_percent\_id 0.0 --min\_otu\_size 3

#Uppmax

pick\_open\_reference\_otus.py -i seqs.fna -o open\_otu\_pick/uclust\_97 -r /home/liutong/glob/illumina\_project\_1/97\_otus.fasta -s 0.1 -aO 16 --prefilter\_percent\_id 0.0 --min\_otu\_size 3