Nova-seq 6000

OneNote > 研究雑感 > 5. Revision in Nature Biotechnology >
7. Prepare demultiplexed and trimmed reads

Step1:
Removing Illumina adapter sequences

File type:
.fastq

Software:
Trimgalore

File type:
.fq

Step2:
Demultiplexing reads using antibody barcodes

Software:
FlexBar

File type:
.fastq

Step3:
Convert reverse complement sequences to forward sequences

Software:
FastX tool kit

File type:
.fastq

Step4:
Combine fastq files of the converted reverse complement sequences and forward sequences

Operation by "cat" command in shell script

File type:
.fastq

Step5:
Remove left side of reads from sequences of antibody barcode-ligated sequence.

Software:
FlexBar

File type:
.fastq

Step6:
Remove right side of reads from a sequence of a cell-barcode.

Software:
FlexBar

File type:
.fastq

Step7:
Use the 8 nt sequence of the random primer as unique molecular identifier, and removed the 8 nt.

Software:
UMI tools

Step8:
Map the reads to the genome.

File type:
.fastq

Software:
Bowtie2

File type:
.sam

Step9:
Remove mapped reads, which have Q score < 1.

Software:
SAM tools

File type:
.sam

Step10:
Convert SAM files to BAM files.

Software:
SAM tools

File type:
.bam

Step11:
Sort order of reads.

Software:
SAM tools

File type:
.bam

Step12:
Generate index files of BAM files.

Software:
SAM tools

File type:
.bai

Step13:
Remove amplification duplicates using unique molecular identifier and mapping coordinates.

Software:
UMI tools

File type:
.bam

OneNote > 研究雑感 > 5. Revision in Nature Biotechnology > 8. Mapping

Bash script:
002_Figure1a-1b.sh

Step1:
Combine BAM files of 8 single cells in each experiment (1st, 2nd and 3rd experiment).

File type:
.bam

Software:
SAM tools

File type:
.bam

Step2:
Convert BAM files to Tag directories.

Software:
Homer

File type:
Tag directory

Step3:
Convert BAM files of 1st experiments to BED files.

Software:
BED tools

Step4:
Copy the BED file of 1st experiments to each Tag directory.

File type:
.bed

Step5:
Calculate distribution patterns of reads of 1st, 2nd and 3rd experiments in signal regions of 1st experiments.

Software:
Homer

Text file

Figure 1a & 1b

Bash script:
003_Figure1c-1f)Bootstrap.sh

Step1:
Combine BAM files of 1st, 2nd and 3rd experiments derived from the same single cell.

Software:
SAM tools

File type:
.bam

Step2:
Convert the combined BAM files to BED files.

Software:
BED tools

File type:
.bed

Step3:
Generate randomized controls from H3K27ac, H3K27me3 and control IgG of each cell.

Software:
BED tools

File type:
.bed

Step4:
Generate genome bins, 500bp bin sliding 250 bp.

Input file:
GRCh38.genome

Software:
BED tools

Output file:
Genome_bin_500bp_250bp
_sliding_GRCh38.bed

Step5:
Count reads in the bins.

Software:
BED tools

File type:
.bed

Step6:
Remove genomic bins, which does not contains single-cell reads.

Linux command
"cat" & "awk"

File type:
.bed

Step7:
Calculate Ab-IgG, Ab-Random IgG, Random Ab-IgG and Random Ab-Random IgG in each bin.

Linux command
"cat" & "awk"

File type:
.bed

Step8:
Add header to the BED files for the subsequent bootstrap test.

Linux command
"cat" & "sed"

File type:
.bed

Step9:
Make directories for bootstrap test.

Linux command
"mkdir"

Directories

Step10:
Copy the BED files into the directories with unified file name.

File type:
.bed

Step11:
Copy a R script file of bootstrap test into the directories.

c200_Rscript_boot.r

```
1  library(boot)
2  x=read.delim("Input_20000L.bed", header=TRUE)
3  SC <- x$RandAbDRandIgG
4  as.vector(SC)
5  sample <- function(SC, d) {
6  return(print(SC[d]))
7  }
8  Bout=boot(SC, sample, R=5000)
9  ci=boot.ci(Bout, conf=c(0.90, 0.95, 0.99, 0.999), type="basic")
10 sink("Output_CI_95.txt")
11 print (ci)
12 sink()
13 q()
```

Step12:
Adjust bin number to 200,000 bins for bootstrap test.

Linux command
"cat" & "head"

File type:
.bed

Step13:
Bootstrap test using R package "boot".

R package
"Boot"

Output file:
Text file

Confidence
interval 0.999

Step14:
Extract putative signal regions using upper side of 99.9% confidence interval determined by bootstrap test.

Linux command
"cat" & "awk"

Putative signal regions
File type: .bed

Step15:
Count region number of putative signals extracted in previous step.

Linux command
"find" & "wc -l"

**Number of putative
signal regions/cell**
File type: .CSV

Step16:
Count read number in the putative signal regions.

Linux command
"cat" & "awk"

File type:
.bed

Step17:
Count number of putative signals per cell.

Linux command
"find" & "wc -l"
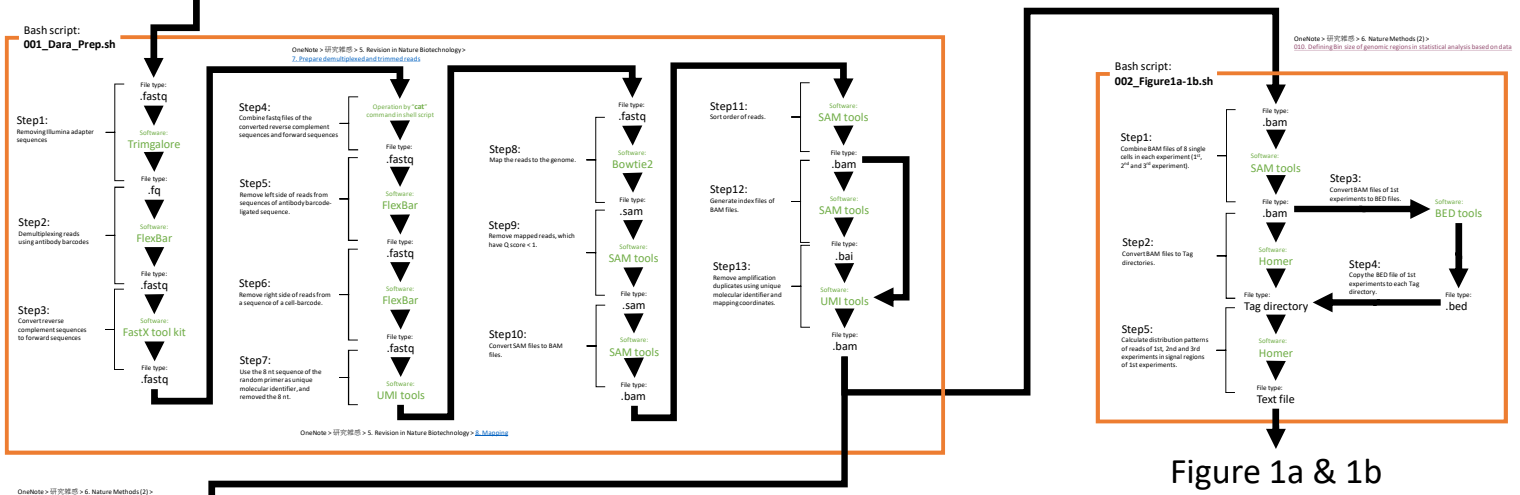
**Number of putative
signals/cell**
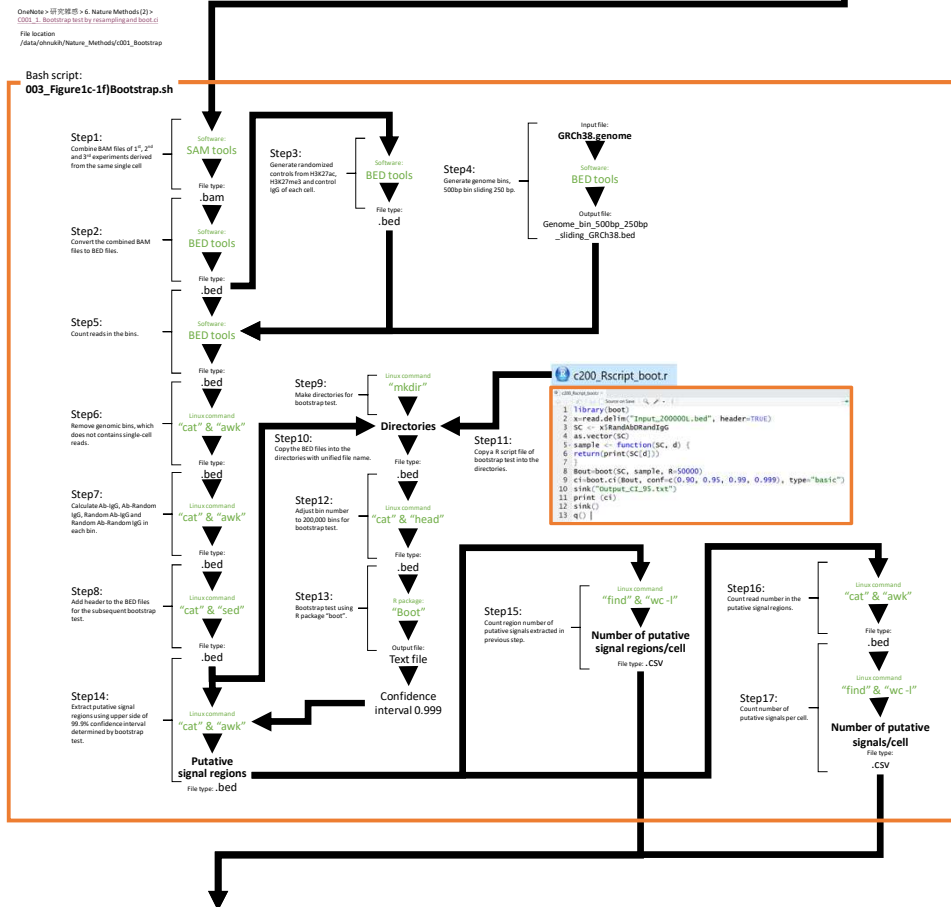File type:
.csv

Figure 1c-1f

Figure 2a-2d

Figure 2e-2h

/data/ohnukih/Nature_Methods/c001_Bootstrap/000_c001_bootstrap.sh

Supplementary
Figure 4a-4d

/data/ohnukih/Nature_Methods/c001_Bootstrap/000_c001_bootstrap.sh

Figure 4a-4b

Location of Bash script
"/data/ohnukih/Nature_Methods/CD06_Interacting_genes_with_enhancers/000_CD06_interacting_genes.sh"

Data location of figure shown in the right panel
"/data/ohnukih/Nature_Methods/CD06_Interacting_genes_with_enhancers/dd10_Summary_of_interacting_genes.xlsx"

Figure 4c-4g

OneNote>研究雑感>6. Nature Methods/C005. Pathway enrichment analysis with SC1-8, random controls of SC1-8, bulk ChIP-seq and RNA-seq

Bash script file
/data/ohnukih/Nature_Methods/c001_Bootstrap/000_c001_bootstrap.sh
/000_C005_BashScript_PathwayEnrichmentAnalysis.sh

Data location
"H:\Raw data\(D) 2019. 11\Comparing different cells of RNA-seq data\Comparizon of SC1-8 putative signals, KS62 bulk ChIP-seq, KS62 RNA-seq and H1 RNA-seq.xls"

Figure 5a

Data location
/data/ohnukih/Nature_Methods/C007_tSNE/

Bash script
/data/ohnukih/Nature_Methods/C007_tSNE/000_C007_data_prep_for_tSNE.sh

OneNote>研究雑感>6. Nature Methods/C007_tSNE plot

Figure 5b-5e +
Supplementary Fig.5a-5c

Data location
/data/ohnukih/Nature_Methods/C012_SE_heterogeneity/

Bash script
/data/ohnukih/Nature_Methods/C012_SE_heterogeneity/000_C012_SE_heterogeneity.sh

OneNote>研究雑感>6. Nature Methods/C012. Super-enhancers heterogeneity