EMBL predoc course 2019 Day4 Mosaicatcher

Hyobin Jeong

October 19, 2019

Contents

1 Introduction

In this practical session, we will analyze 100 cells of RPE1. For this, we have mixed 80 cells of RPE-1 WT, and 20 cells of RPE-1 BM510 to simulate subclone. Our mission is to find out 20 cells in this synthetic subclone using Mosaicatcher pipeline.

```
## Copy the Mosaicatcher pipeline (for the practical session)
cd /scratch/name_abcd/
mkdir Practical_singlecell
cd Practical_singlecell
cp /g/korbel2/StrandSeq/Test_HJ/pipeline_20190625.tar.gz ./
tar -xvzf pipeline_20190625.tar.gz
cd pipeline_20190625/
```

2 Preparation of input files for Mosaicatcher

After copying the pipeline, we need to put input files to the 'bam' folder Inside of 'bam', you need to make two subfolders 'all' and 'selected'. You can copy the bam files to those folder, but to save the space we will bind the files rather than copying the files to this folder.

```
## Let's go inside of the bam folder, and then remove the bad quality librariedcd bam
mkdir RPE

## Inside of the RPE folder, we need to subfolders all and selected
cd RPE
mkdir all
mkdir selected

## Let's link the bam files and bai files for the 'all' first
cd all

cp /scratch/jeong/pipeline_20190625_RPE_mixture/bam/RPE/all/make_link_bam.sh ./
./make_link_bam.sh

cp /scratch/jeong/pipeline_20190625_RPE_mixture/bam/RPE/all/make_link_bai.sh ./
./make_link_bai.sh

## Let's link the bam files and bai files for the 'selected' also
cd ../selected

cp /scratch/jeong/pipeline_20190625_RPE_mixture/bam/RPE/selected/make_link_bam.sh ./
./make_link_bam.sh

cp /scratch/jeong/pipeline_20190625_RPE_mixture/bam/RPE/selected/make_link_bai.sh ./
./make_link_bai.sh
```

3 snv sites to genotype

As single-cell sequencing can be sparse to call single-nucleotide variants (SNV), we need the help of known heterozygous SNV sites to get more accurate haplotype information. Those snv sites information is in the 'snv sites to genotype' folder. If the index file is older than main file, it can make error during the run, so that we need to make sure the time points of copied files.

```
## Let's go inside the snv_sites_to_genotype folder
cd snv_sites_to_genotype

## Check the time points the files were generated
ls -lh

## If the index file is same or older than main file, we can try following code
cp *.tbi ../
cp ../*.tbi ./
```

4 Running Mosaicatcher

Now we are ready to execute mosicatcher pipeline

```
## Let's go inside of the folder with output files from plotting pipeline
cd /scratch/name_abcd/Practical_singlecell/pipeline_20190625
chmod +x run_pipeline_singularity.sh
sbatch -t 90:00:00 -N 1 -n 1 --mem=50000 --mail-type=FAIL,BEGIN,END \
--mail-user=hyobin.jeong@embl.de -o output.txt ./run_pipeline_singularity.sh
```

5 Session Info

```
toLatex(sessionInfo())
```

- R version 3.5.1 (2018-07-02), `x86_64-apple-darwin15.6.0`
- Locale: `C`
- Running under: `macOS High Sierra 10.13.3`
- Matrix products: default
- BLAS: `/System/Library/Frameworks/Accelerate.framework/Versions/A/Frameworks/vecLib.framework/Versions/A/libBL`
- LAPACK: `/Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRlapack.dylib`
- Base packages: base, datasets, grDevices, graphics, methods, parallel, stats, stats4, utils
- Other packages: BiocGenerics 0.26.0, GenomeInfoDb 1.16.0, GenomicRanges 1.32.7, IRanges 2.14.12, S4Vectors 0.18.3, breakpointR 1.0.0, breakpointRdata 1.0.0, cowplot 0.9.4, ggplot2 3.1.1, knitr 1.20
- Loaded via a namespace (and not attached): Biobase 2.40.0, BiocParallel 1.14.2, BiocStyle 2.8.2, Biostrings 2.50.2, DT 0.4, DelayedArray 0.6.6, DelayedMatrixStats 1.2.0, FNN 1.1.2.1, GenomeInfoDbData 1.1.0, GenomicAlignments 1.16.0, MOFAtools 0.99.0, Matrix 1.2-14, MultiAssayExperiment 1.6.0, R6 2.4.0, RColorBrewer 1.1-2, RCurl 1.95-4.12, Rcpp 1.0.1, Rhdf5lib 1.2.1, Rsamtools 1.34.0, SingleCellExperiment 1.2.0, SummarizedExperiment 1.10.1, XVector 0.20.0, assertthat 0.2.1, backports 1.1.4, beeswarm 0.2.3, bindr 0.1.1, bindrcpp 0.2.2, bitops 1.0-6, codetools 0.2-15, colorspace 1.4-1, compiler 3.5.1, corrplot 0.84, crayon 1.3.4, data.table 1.12.2, digest 0.6.19, doParallel 1.0.14, dplyr 0.7.6, dynamicTreeCut 1.63-1, edgeR 3.22.3, evaluate 0.11, foreach 1.4.4, ggbeeswarm 0.6.0, ggrepel 0.8.0, glue 1.3.1, grid 3.5.1, gridExtra 2.3, gtable 0.3.0, gtools 3.8.1, highr 0.7, htmltools 0.3.6, htmlwidgets 1.2, httpuv 1.5.1, igraph 1.2.2, iterators 1.0.10, jsonlite 1.6, later 0.8.0, lattice 0.20-35, lazyeval 0.2.2, limma 3.36.3, locfit 1.5-9.1, magrittr 1.5, matrixStats 0.54.0, mime 0.6, munsell 0.5.0, pheatmap 1.0.10, pillar 1.4.0, pkgconfig 2.0.2, plyr 1.8.4, promises 1.0.1, purrr 0.2.5, reshape2 1.4.3, reticulate 1.10, rhdf5 2.24.0, rjson 0.2.20, rlang 0.3.4, rmarkdown 1.10, rprojroot 1.3-2, rstudioapi 0.7, scales 1.0.0, scater 1.8.4, scran 1.8.4, shiny 1.3.2, shinydashboard 0.7.0, statmod 1.4.30, stringi 1.4.3, stringr 1.4.0, tibble 2.1.1, tidyr 0.8.1, tidyselect 0.2.4, tinytex 0.7, tools 3.5.1, tximport 1.8.0, vipor 0.4.5, viridis 0.5.1, viridisLite 0.3.0, withr 2.1.2, xfun 0.3, xtable 1.8-4, yaml 2.2.0, zlibbioc 1.26.0