

Reuben's personal guide to Hi-C analysis

December 14, 2016

Introduction

Data Processing

Download Hi-C files

Files analysed throughout this guide were all obtained from the short read archive (SRA) under the accession .

Sequence Quality

Read mapping and filtering

Generate contact matrix

Use homer package to generate tag dir.

Use homer to convert tag dir to generate raw contact matrix.

Filter rows and columns of contact matrix. Remove rows and columns with poor read mapping.

Use HiCorrector to normalise contact matrices.

Normalising contact matrices

Downstream analysis

Topological domains

Genomic compartments

```
#!/usr/bin/perl

use strict;
use warnings;

5  for (my $i = 1; $i < 100; $i++) {
    if ($i == 99) {
        print $i." Luftballons reached!\n\n";
    } else {
10     print $i." Luftballons...\n";
    }
}

exit;
```

Listing 2: Nena would be proud.

```

#!/bin/bash

# Script to perform Hi-C read mapping for paired end reads
# using hicup software
# script is written for use on adelaide university phoenix
# cluster

# Invoked by:
#
# READPATH=<path to fastq file dir> GENOME=<name of genome
# assembly> sbatch alignHiC.sh

#SBATCH -p batch
#SBATCH -N 1
#SBATCH -n 16
#SBATCH --time=2-00:00
#SBATCH --mem=32GB

# Notification configuration
#SBATCH --mail-type=END
#SBATCH --mail-type=FAIL
#SBATCH --mail-user=reuben.buckley@adelaide.edu.au

#SBATCH --array=0-1 #result of 'ls *.fastq.gz | wc -l' less
# one.

module load R

FILES=$(ls $READPATH | rev | cut -c 9- | rev | uniq)

mkdir ./${FILES[$SLURM_ARRAY_TASK_ID]}

date > $GENOME.${FILES[$SLURM_ARRAY_TASK_ID]}.log.txt
hicup --bowtie2 /apps/software/Bowtie2/2.2.9-GCC-5.3.0-
      binutils-2.25/bin/bowtie2 --keep --longest 800 --shortest
      100 --threads 16 --index ../../genomes/genomeIndex/$GENOME
      /${GENOME}knownChr --digest ../../genomes/genomeDigest/
      $GENOME/DigestKnownChr* --outdir ${FILES[
      $SLURM_ARRAY_TASK_ID]} $READPATH${FILES[
      $SLURM_ARRAY_TASK_ID]}_1.fastq $READPATH${FILES[
      $SLURM_ARRAY_TASK_ID]}_2.fastq &>> $GENOME.${FILES[
      $SLURM_ARRAY_TASK_ID]}.log.txt
date &>> $GENOME.${FILES[$SLURM_ARRAY_TASK_ID]}.log.txt

```

Listing 1: Bash script used to map reads on phoenix