

Mapping HiC reads to reference genome

```
#!/bin/bash -l
#SBATCH --job-name=hic_map
#SBATCH --output=/scratch/grp/msc_appbio/Group2_ABCC/HiC/logs/%x_%j.out
#SBATCH --error=/scratch/grp/msc_appbio/Group2_ABCC/HiC/logs/%x_%j.err
#SBATCH --time=24:00:00
#SBATCH --mem=32G
#SBATCH --cpus-per-task=16

# Hi-C analysis requires genomic coordinates for each read. Alignment assigns each read to a location in
# the reference genome. Bowtie2 v2.5.1 and Samtools v1.17 were used for this.

module load bowtie2
module load samtools
#directory
BASE=/scratch/grp/msc_appbio/Group2_ABCC/HiC

# Folder containing raw Hi-C FASTQ files
RAW=$BASE/raw_fastq_HiC

# Folder containing Bowtie2 genome indices
IDX=$BASE/results/indices

# Where output of the mapping
ALIGN=$BASE/results/alignments

mkdir -p $ALIGN

# Each entry: SAMPLE_PREFIX  GENOME_INDEX_PREFIX

declare -a MAPS=(
    "BY4742_R1 BY4742_SRR6823436"
    "BY4742_R2 BY4742_SRR6823436"
    "SY14_R1 SY14"
    "SY14_R2 SY14"
)

# Loop over each (SAMPLE, GENOME) pair and perform:
#   1) Paired-end alignment with Bowtie2
#   2) SAM sorted BAM conversion
#   3) BAM indexing
#   4) Removal of intermediate SAM file to save space

for ENTRY in "${MAPS[@]}"
do
    set -- $ENTRY
    SAMPLE=$1
    GENOME=$2

    R1=${RAW}/${SAMPLE}_1.fastq.gz
    R2=${RAW}/${SAMPLE}_2.fastq.gz

    OUT_SAM=${ALIGN}/${SAMPLE}_${GENOME}.sam
    OUT_BAM=${ALIGN}/${SAMPLE}_${GENOME}.sorted.bam

    echo "==== Mapping $SAMPLE to $GENOME ==="
    echo "R1: $R1"
    echo "R2: $R2"
    echo "Index: ${IDX}/${GENOME}"

    bowtie2 -x ${IDX}/${GENOME} \
        -1 $R1 -2 $R2 \
        --very-sensitive-local \
        -p 8 \
        -S $OUT_SAM

    echo "Converting SAM to sorted BAM for $SAMPLE"

    samtools view -bS $OUT_SAM \
        | samtools sort -@ 4 -o $OUT_BAM

    samtools index $OUT_BAM
    rm $OUT_SAM

    echo "Done: $OUT_BAM"
done
```