**Put reads into /pub/ehanse/Reads**

**Download and unzip complete mouse genome from Gencode into /pub/ehanse/Genome**

**“GRCm39.genome.fa”**

**Trim install**

conda install bioconda::trim-galore

**Trim**

#!/bin/bash

#SBATCH --job-name=trim\_galore

#SBATCH --partition=free

#SBATCH --cpus-per-task=16

#SBATCH --mem=100G

#SBATCH --time=4:00:00

#SBATCH --output=slurm-%j.out

#SBATCH --error=slurm-%j.err

# Run Trim Galore! with 16 CPUs

trim\_galore --paired Control1\_Read1.txt.gz Control1\_Read2.txt.gz \

--cores 16 \

--gzip \

--fastqc \

--path\_to\_cutadapt $(which cutadapt) \

--output\_dir /pub/ehanse/Trimmed/

**Build Reference genome with Bismark**

#!/bin/bash

#SBATCH --job-name=bismark\_index

#SBATCH --output=bismark\_index.log

#SBATCH --error=bismark\_index.err

#SBATCH --ntasks=1

#SBATCH --cpus-per-task=32

#SBATCH --mem=6400MB # 200MB per core \* 32 cores

#SBATCH --time=2:00:00 # Adjust based on expected runtime

#SBATCH --partition=free # Change if needed

module load bismark

bismark\_genome\_preparation --parallel 32 --verbose /pub/ehanse/Genome

**This will generate a series of Genomic files you will reference in the next step in your Genome directory**

**Align with Bismark**

#!/bin/bash

#SBATCH --job-name=bismark\_align

#SBATCH --output=bismark\_align.log

#SBATCH --error=bismark\_align.err

#SBATCH --ntasks=1

#SBATCH --cpus-per-task=32 # Adjust as needed

#SBATCH --mem=200GB # Increase if needed

#SBATCH --time=24:00:00 # Adjust based on dataset size

#SBATCH --partition=free

#SBATCH --error=%x.%A.err

module load bismark # Adjust if necessary

module load samtools # Ensure Samtools is available for sorting

# Define paths

GENOME\_DIR="/pub/ehanse/Genome"

READ1="Pio1\_Read1.txt.gz\_val\_1.fq.gz"

READ2="Pio1\_Read2.txt.gz\_val\_2.fq.gz"

OUTDIR="./bismark\_aligned" # Change to your preferred output directory

mkdir -p $OUTDIR

# Run Bismark alignment

bismark --genome $GENOME\_DIR \

-1 $READ1 -2 $READ2 \

--output\_dir $OUTDIR \cd

--parallel 32 \

--bowtie2 \

--non\_directional \

--gzip \

--bam

**“non\_directional” is for enzymatic-Seq produces BAM file that incorporates both reads. This takes 2-3 hours per sample. But can be batch issued to save time.**

**Extract Methylation Data**

**Use Bismark Methylation Extractor to generate output data files for downstream analysis**

#!/bin/bash

#SBATCH --job-name=bwameth\_align

#SBATCH -p free

#SBATCH --nodes=1

#SBATCH --ntasks=16

#SBATCH --mem=64G

#SBATCH --time=4:00:00

#SBATCH --error=%x.%A.err

#SBATCH --output=%x.%A.out

conda activate $ENVIRONMENT

bismark\_methylation\_extractor --paired-end --gzip --bedGraph --cytosine\_report \

--multicore 16 --genome\_folder /pub/ehanse/Genome \

--output Bismark\_Output $Sample1\_Aligned.sortedByName.bam

This will return a series of files that can be used for downstream analysis.

First to use Integrated Genome Viewer download it to your Desktop. Then open the .bedgraph files produced by Bismark\_Methylation\_Extractor. You can visualize data here. For more specific Analysis, we move to R.

Link to R code which is too long to post here and the above code is here:

<https://github.com/erichanse/MethylSeq.git>