## **Methods**

### **Bioinformatics Analysis Pipeline**

All sequencing data was processed using a custom, reproducible pipeline developed in Nextflow (v22.10.0 or later, DSL2) [1]. The pipeline automates quality control, read alignment, peak calling, differential analysis, and data visualization. All analyses were performed against the *Mus musculus* mm9 reference genome assembly.

### **Data Pre-processing and Quality Control**

Raw sequencing reads in FASTQ format were first assessed for quality using FastQC (v0.11.9). Adapters were trimmed using Trim Galore! (v0.6.7), which employs Cutadapt (v3.5) for adapter detection and removal. The trimming process used a Phred quality score threshold of 0, a stringency of 13 for adapter matching, and discarded reads that were shorter than 30 bp post-trimming.

### **Alignment and Filtering**

Trimmed reads were aligned to the mm9 mouse reference genome using Bowtie2 (v2.4.5) [4] with default parameters, generating alignments in SAM format. The resulting SAM files were converted to BAM format, sorted by coordinate, and indexed using Samtools (v1.15) [5]. To improve signal quality, reads mapping to ENCODE-blacklisted regions (mm9-blacklist) were removed. Subsequently, SAMbamba (v0.8.2) [8] was used to filter for uniquely mapped reads. For paired-end datasets, only properly paired reads were retained. This final set of filtered, sorted, and indexed BAM files was used for all downstream analyses.

### **Peak Calling**

Peak calling was performed on a per-sample basis using two alternative algorithms, depending on the data type.

1. **MACS2:** For transcription factor ChIP-seq, Model-based Analysis of ChIP-Seq 2 (MACS2, v2.2.7.1) [10] was the primary peak caller. Peaks were called without building a shifting model (--nomodel) and by treating paired-end data as fragments (BAMPE format). Both narrow (\*.narrowPeak) and broad (\*.broadPeak) peak calls were generated for each sample.
2. **EPIC2:** For histone modifications that generate broad peak regions (e.g., H3-K27me3), EPIC2 (v0.0.52) (Stovner and Sætrom, 2019) was used as an alternative. EPIC2 was run with an effective genome fraction of 0.80, a bin size of 400 bp, a gap size of 2400 bp, and a significance E-value of 100.

A consensus peak set was generated by merging the primary peak calls (.narrowPeak from MACS2 or .bed from EPIC2) from all samples using bedtools merge (v2.30.0) [6].

### **Signal Track Generation and Normalization**

For data visualization, signal tracks were generated in BigWig format. First, fragment coverage was calculated across the genome using bedtools genomecov, producing a bedGraph file. To normalize for sequencing depth and signal-to-noise ratio, a "Reads in Peaks per Million" (RiPPM) scaling factor was calculated for each sample. This factor was derived from the total number of fragments and the fraction of fragments in peaks (FRiP score), which was determined using bedtools coverage against the consensus peak set. The bedGraph signal was multiplied by this RiPPM factor, and the resulting normalized signal was converted into a BigWig file using the bedGraphToBigWig utility from UCSC tools (v424) [3]. For visualization of sample groups, BigWig files from constituent samples were merged and averaged using WiggleTools (v1.2.11).

### **Differential Peak Analysis**

To identify statistically significant differential binding sites, or differential chromatic accessibility, between experimental conditions, two complementary methods were employed using custom subworkflows:

1. **DIFFREPS:** A count-based differential analysis was performed using a custom workflow. This method utilized read counts within consensus peak regions and the pre-calculated normalization factors to identify significant changes.
2. **MAnorm2:** The MAnorm2 (v1.3.0) package in R [9] was used to compare and normalize ChIP-seq signals based on the assumption that most binding sites are common between conditions. MAnorm2 was run on fragment files and the per-sample peak sets.

Results from both methods were systematically compared to identify a high-confidence set of differential sites.

### **Comprehensive Quality Assessment and Reporting**

Post-alignment quality metrics were collected for each sample using Picard Tools (v2.27.1), including CollectMultipleMetrics and CollectInsertSizeMetrics for paired-end read libraries. Finally, a comprehensive quality control report was generated by aggregating results from FastQC, Bowtie2, Samtools, Picard, and MACS2 using MultiQC (v1.13) [2]. This provided a holistic overview of data quality, alignment rates, library complexity, and peak calling statistics across all samples.

### **References**

1. Di Tommaso, P., Chatzou, M., Floden, E. W., Barja, P. P., Palumbo, E., & Notredame, C. (2017). Nextflow enables reproducible computational workflows. *Nature Biotechnology*, 35(4), 316–319. PMID: 28398311.
2. Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32(19), 3047–3048. PMID: 27312411.
3. Kent, W. J., et al. (2002). The human genome browser at UCSC. *Genome Research*, 12(6), 996-1006. PMID: 12045153.
4. Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357–359. PMID: 22388286.
5. Li, H., et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. PMID: 19505943.
6. Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841–842. PMID: 20110278.
7. Stovner, E. B., & Sætrom, P. (2019). EPIC2: a software for enhancer-promoter interaction calling and epigenome-based gene regulatory network reconstruction. *BMC Bioinformatics*, 20(1), 405. PMID: 30923821.
8. Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J., & Prins, P. (2015). SAMbamba: a package for fast processing of high-throughput sequencing data. *Bioinformatics*, 31(12), 2032–2034. PMID: 25697820.
9. Tu, S., et al. (2021). MAnorm2 for quantitatively comparing groups of ChIP-seq samples. *Genome Research*, 31(1), 131-145. PMID: 33208455.
10. Zhang, Y., et al. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biology*, 9(9), R137. PMID: 18798982.