1. # Aligning the raw fast files:
The full pipeline till alignment is is process-chip-seq.sh

2. # Downloading the bed files and super enhancer coordinates:
downloaded from the reference papers

3. #This is just an intermediate step I had to do:
The original H1.bed file contained coordinates in hg19 format (from Hnisz et al. 2013)
I aligned our FASTQ data to hg38 reference genome using the bowtie2 index
This created a coordinate system mismatch - I was looking for hg19 coordinates in hg38-aligned data
The liftOver conversion fixed this mismatch by converting coordinates from hg19 to hg38
FASTQ files → Bowtie2 (hg38 reference) → BAM files (hg38 coordinates)
↓
H1.bed (hg19) → LiftOver → H1_hg38.bed → Signal quantification → CORRECTED results

```
# Re-quantify signal with CORRECT hg38 coordinates
echo "=== RE-RUNNING ANALYSIS WITH CORRECT COORDINATES ==="

multiBigwigSummary BED-file --BED H1_hg38_no_chr.bed \
    -b /gpfs/data/khodadadilab/home/temp/Di-Stefano-Lab-Assignment/Task-2/align/CTRL_rep1.bw \
       /gpfs/data/khodadadilab/home/temp/Di-Stefano-Lab-Assignment/Task-2/align/CTRL_rep2.bw \
       /gpfs/data/khodadadilab/home/temp/Di-Stefano-Lab-Assignment/Task-2/align/DDX6_rep1.bw \
       /gpfs/data/khodadadilab/home/temp/Di-Stefano-Lab-Assignment/Task-2/align/DDX6_rep2.bw \
    -o h3k27ac_CORRECTED_signals.npz \
    --outRawCounts h3k27ac_CORRECTED_data.tab

echo "Corrected analysis completed!"
```

4. #plotting the results:
Results were plotted using the file plot.r

What could be some possible reasons for different p-value

Potential Source of Divergence	Likely Choice in Paper	Choice in My run	Effect on *p*-value
Unit of replication	1 replicate per group (sgCTRL vs **one** DDX6 rep) → *n* = 1/group	Each of 684 SEs treated as an observation (after averaging 2 reps) → *n* = 684/group	Huge increase in sample size ⇒ far smaller standard error ⇒ *p* drops ≫
Variance estimate	With *n* = 1, must borrow/ pool variance across bins (limma/trend) → moderate *t*	Welch *t* on 684 values (high df)	Pooled region variance + high df ⇒ larger
Replicates included	Only “best” DDX6 replicate (“#5”) shown	Both CTRL reps and both DDX6 reps averaged	Averaging lowers within-group σ ⇒ *p* smaller
Signal transform	$\log_2$(RPKM) (zeros possibly dropped)	$\log_2$(RPKM + 1) (zeros kept)	Adding 1 compresses low CTRL values, increases mean gap ⇒ *p* smaller
Coordinate list	hg19 Hnisz SEs	Liftover hg38 SEs	Minor boundary shifts slightly change RPKM
Test direction	Likely one-tailed (expect ↑ in DDX6)	Two-tailed (default)	One-tailed halves *p* (cannot explain 10^2 gap alone)
Outlier handling	Boxplot caps whiskers at 1.5 × IQR (trims extreme highs)	Violin shows full distribution (keeps all)	Keeping high DDX6 points lowers σ & raises mean ⇒ *p* smaller