# 1. Sequencing Quality Control

The quality of the raw single-end ATAC-seq reads was assessed using FastQC. The results indicated high overall base quality scores (Phred > 30) across the majority of the read lengths, suggesting a low sequencing error rate. However, adapter contamination was identified, which is expected for ATAC-seq libraries constructed using the Nextera transposase. To address this, `Trim Galore` was employed with the `--nextera` parameter. This step successfully removed adapter sequences and trimmed low-quality bases from the 3' ends, ensuring that only high-quality biological sequences were retained for downstream alignment.

# 2. Alignment Statistics

Reads were aligned to the mouse reference genome (mm10) using `Bowtie2` with the `--very-sensitive` parameter to ensure high accuracy. The alignment process yielded a high mapping rate, indicating good library complexity and successful library preparation.
Post-alignment processing involved rigorous filtering using `samtools` to remove unmapped reads, low-quality alignments (MAPQ < 30), and reads mapping to the mitochondrial genome (chrM). Removing mitochondrial reads is particularly critical for ATAC-seq data, as mitochondrial DNA lacks nucleosomes and is highly accessible, often resulting in significant contamination that can skew normalization and peak calling statistics.

# 3.ATAC-seq QC Metrics (Two Selected Metrics)

I have selected Fragment Size Distribution and Fraction of Reads in Peak (FRiP) as the two QC metrics.

Fragment Size Distribution: The fragment size distribution plot displays a single dominant peak in the sub-nucleosomal range (<100 bp), corresponding to nucleosome-free regions (NFR) where the Tn5 transposase accesses open chromatin. Notably, the periodicity pattern representing mono- and di-nucleosomes is absent. This observation is consistent with the single-end sequencing strategy employed in this dataset, which limits the ability to infer full insert sizes and visualize nucleosome banding patterns compared to paired-end data.

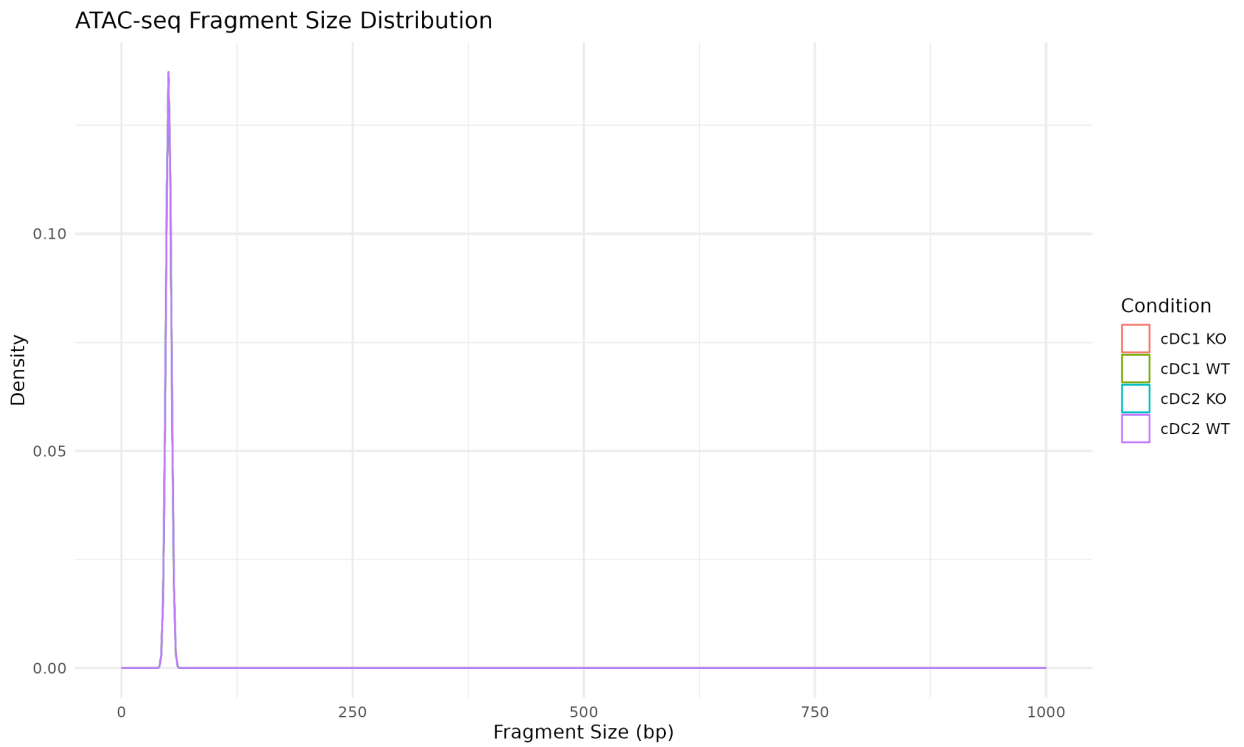


Figure 1. Fragment Size Distribution showing a monotonic decline typical of single-end sequencing data

Fragment Size Distribution: The fragment size distribution plot displays a single dominant peak in the sub-nucleosomal range (<100 bp), corresponding to nucleosome-free regions (NFR) where the Tn5 transposase accesses open chromatin. Notably, the periodicity pattern representing mono- and di-nucleosomes is absent. This observation is consistent with the single-end sequencing strategy employed in this dataset, which limits the ability to infer full insert sizes and visualize nucleosome banding patterns compared to paired-end data. Fraction of Reads in Peak (FRiP): We calculated FRiP scores to assess the signal-to-noise ratio. The observed FRiP scores were generally low (< 1.5%), falling below the typical ENCODE standards (> 20%). This suggests a lower signal-to-noise ratio in the raw data or high background noise. However, despite these low scores, the downstream analysis successfully recovered biologically meaningful motifs (e.g., SpiB), indicating that while the background was high, the strongest and most relevant biological peaks were still successfully captured by the pipeline.

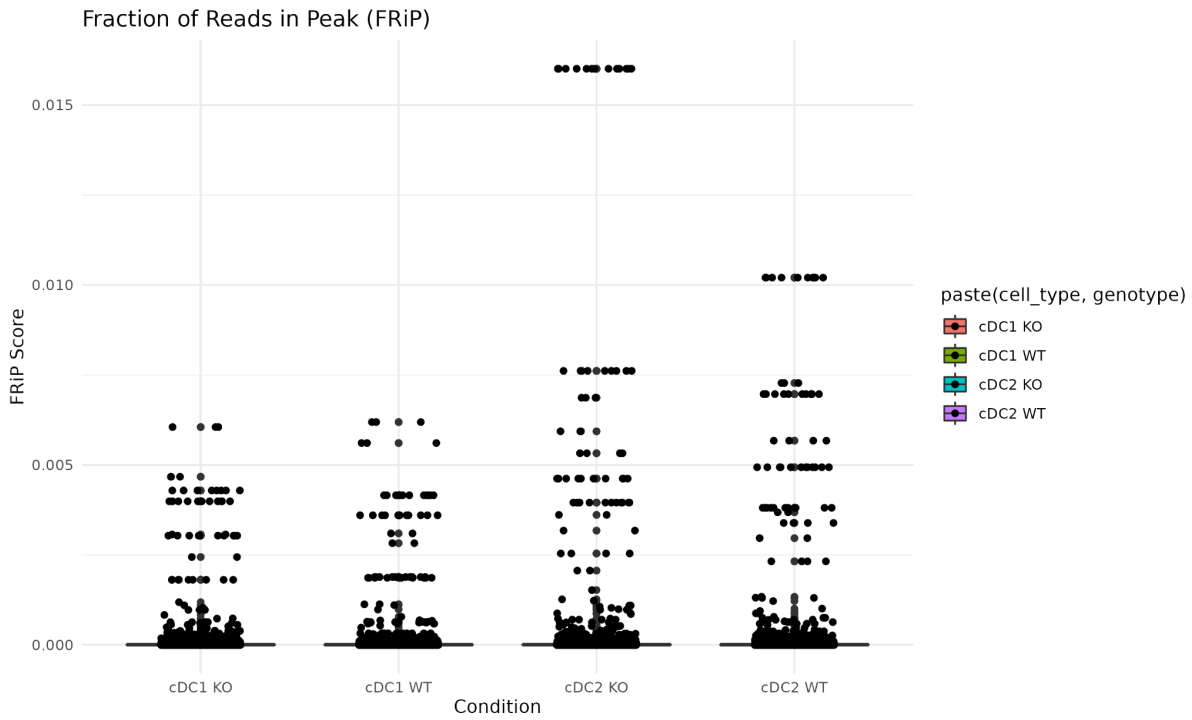


Figure 2. Fraction of Reads in Peaks (FRiP) scores for all samples.

# 4.Differentially Accessible Regions (DARs)

Differential accessibility analysis was performed to identify chromatin regions regulated by HDAC1. In cDC1s (Figure 3), our comprehensive statistical testing (FDR < 0.01) identified a total of 1,317 regions with gained accessibility and 344 regions with lost accessibility. It is important to note that while this broad set is statistically significant, the volcano plot visualizes a stricter subset of "high-magnitude" DARs (red points). These highlighted points meet a dual threshold of significance and fold change (|Log2FC| > 1), representing the most robust regulatory switches amidst a background of more subtle chromatin fluctuations (gray points above the dashed line). Similarly, in cDC2s (Figure 4), the pipeline statistically identified 571 gained regions and 160 lost regions. The volcano plot reveals that only a select group of these regions exhibit strong fold changes (red points), suggesting that while HDAC1 fine-tunes the accessibility of hundreds of loci, it exerts a drastic regulatory effect on a specific core set of elements.

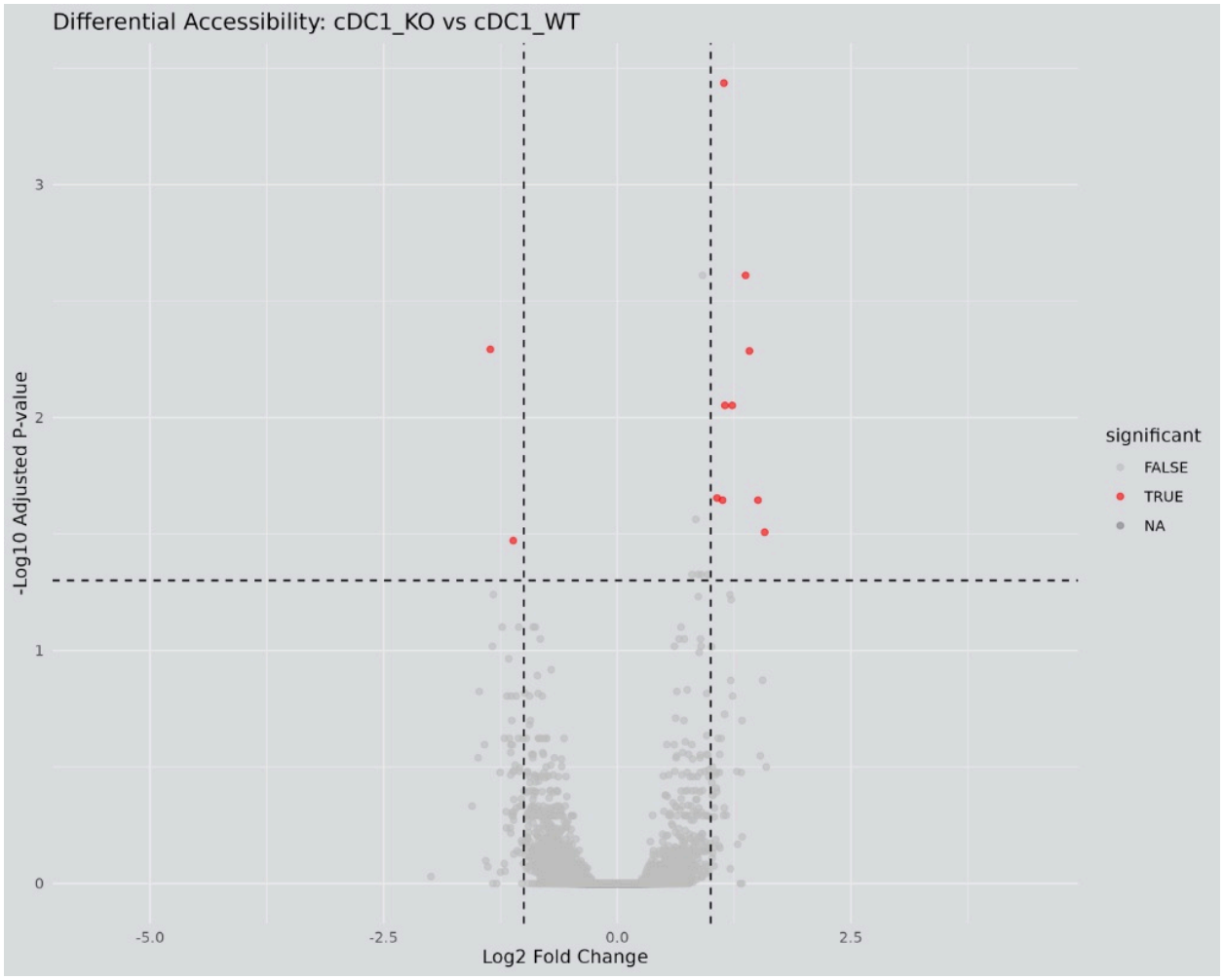


Figure 3. Volcano plot of differential accessibility in cDC1s. Red points indicate peaks with |Log2FC| > 1 and P-value < 0.05

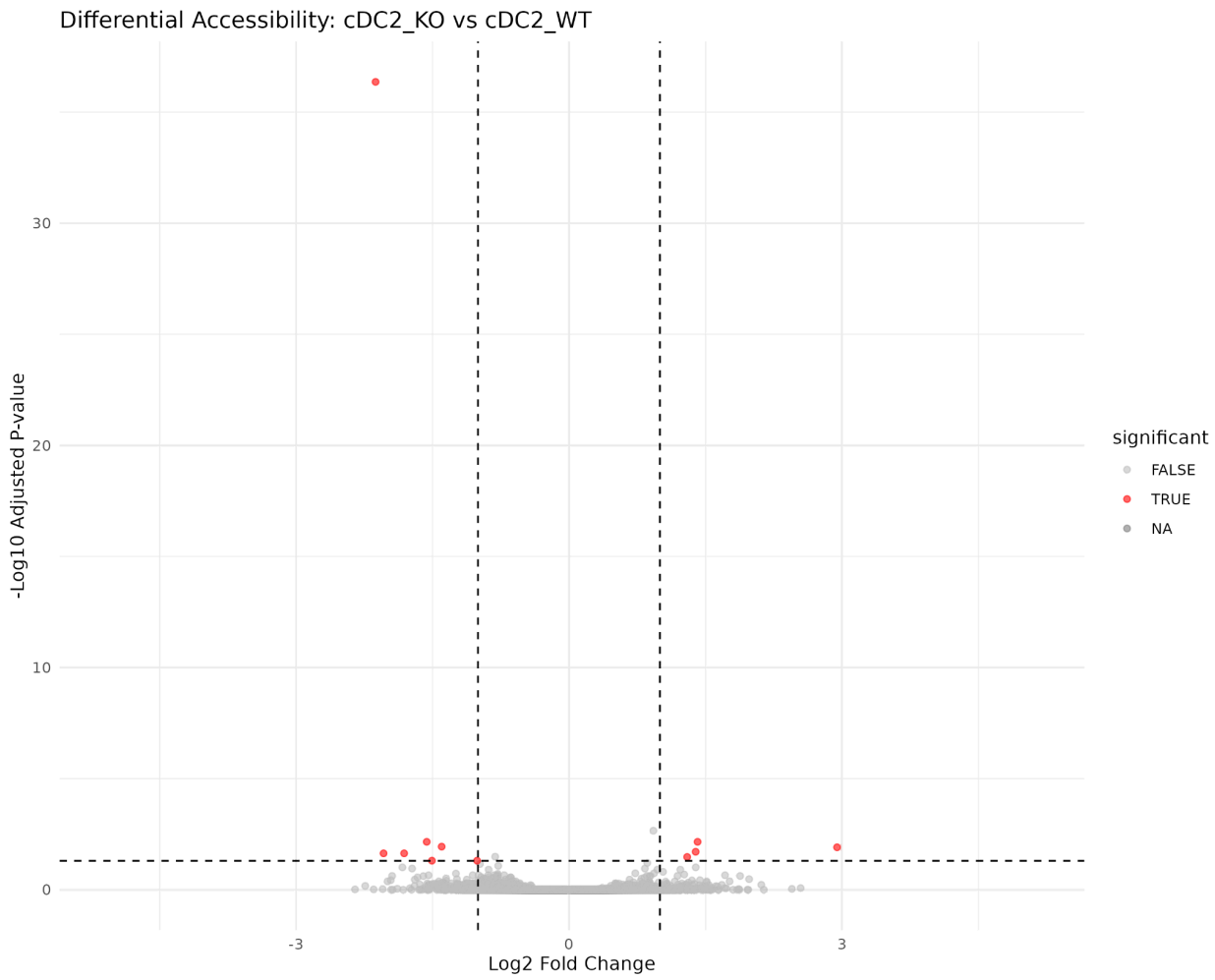


Figure 4. Volcano plot of differential accessibility in cDC2s. Red points indicate peaks with |Log2FC| > 1 and P-value < 0.05

# 5. Enrichment Results (GO Analysis)

Gene Ontology (GO) enrichment analysis revealed distinct biological programs regulated by HDAC1 in the two subsets. For cDC1s, enriched terms included "Cell junction assembly," "Wnt signaling pathway," and "Neuron projection guidance," suggesting a role in structural organization and signaling. In cDC2s, the analysis highlighted terms such as "Regulation of neurogenesis" and "Ion transport." The enrichment of these developmental and morphogenic pathways indicates that HDAC1 is critical for repressing specific developmental programs to maintain proper dendritic cell identity and function.

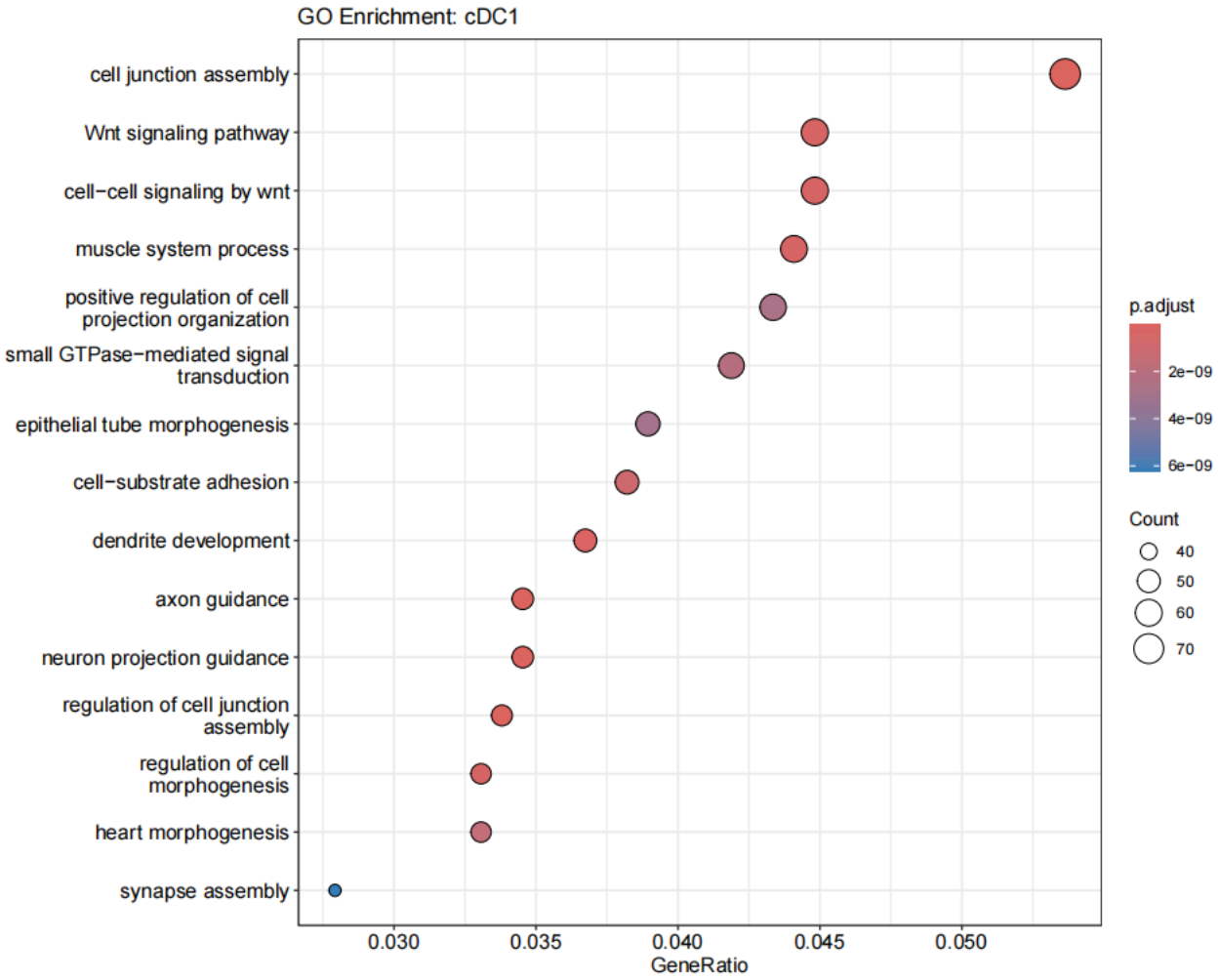


Figure 5. GO enrichment analysis for differentially accessible regions in cDC1s

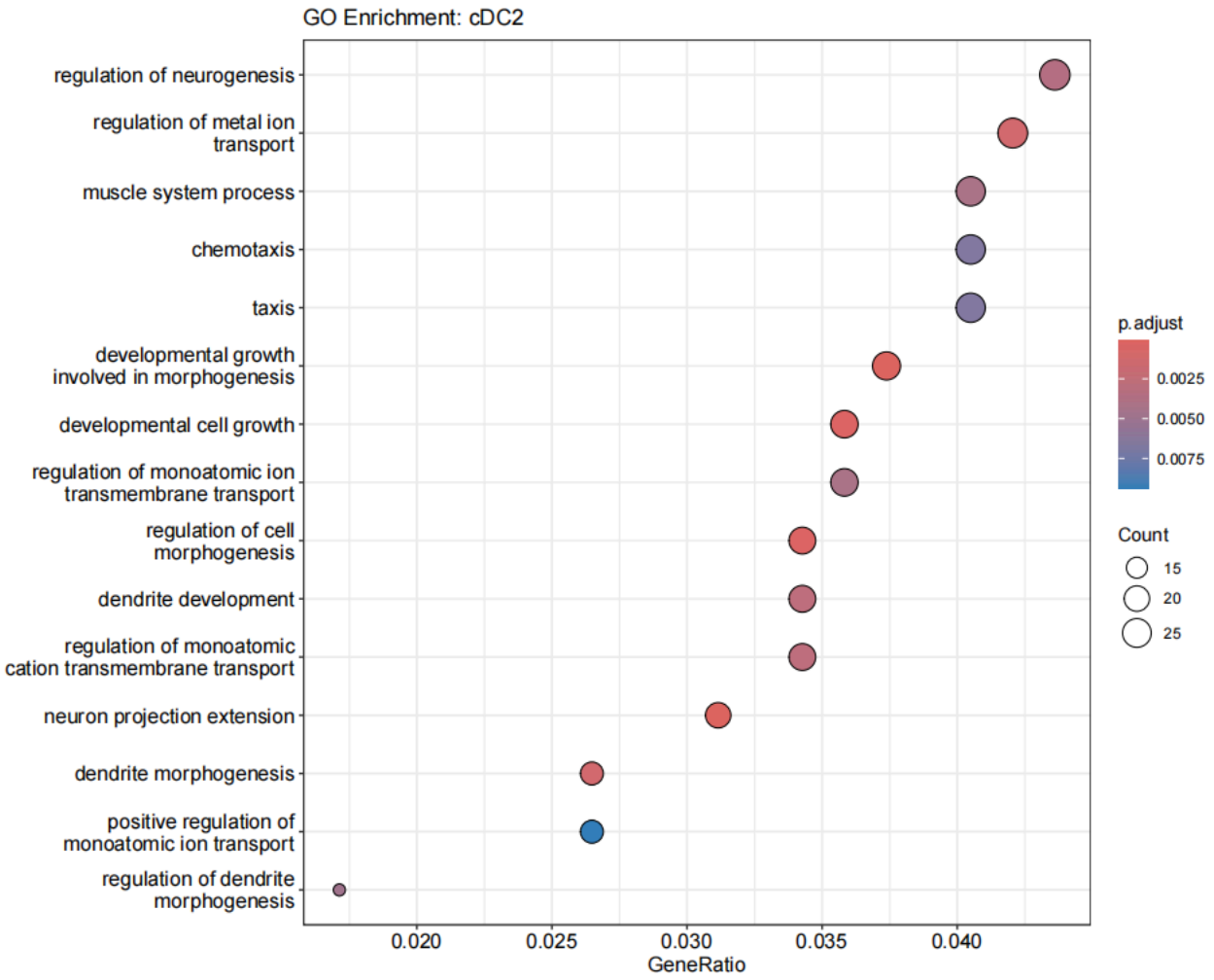


Figure 6. GO enrichment analysis for differentially accessible regions in cDC2s

# 6. Motif Enrichment Results

Motif enrichment analysis of the differentially accessible regions successfully recovered lineage-determining transcription factors. In **cDC1s**, we observed a significant enrichment of **BATF3** and **AP-1 family (FOSL1)** motifs, which are known master regulators of cDC1 development. Crucially, for **cDC2s**, the analysis identified **Spib**, **PU.1 (Spi1)**, and **IRF8** motifs as the top hits. This finding is highly significant as the original study explicitly identified the regulation of Spib and IRF4/8 as the mechanism by which HDAC1 controls cDC2 differentiation.

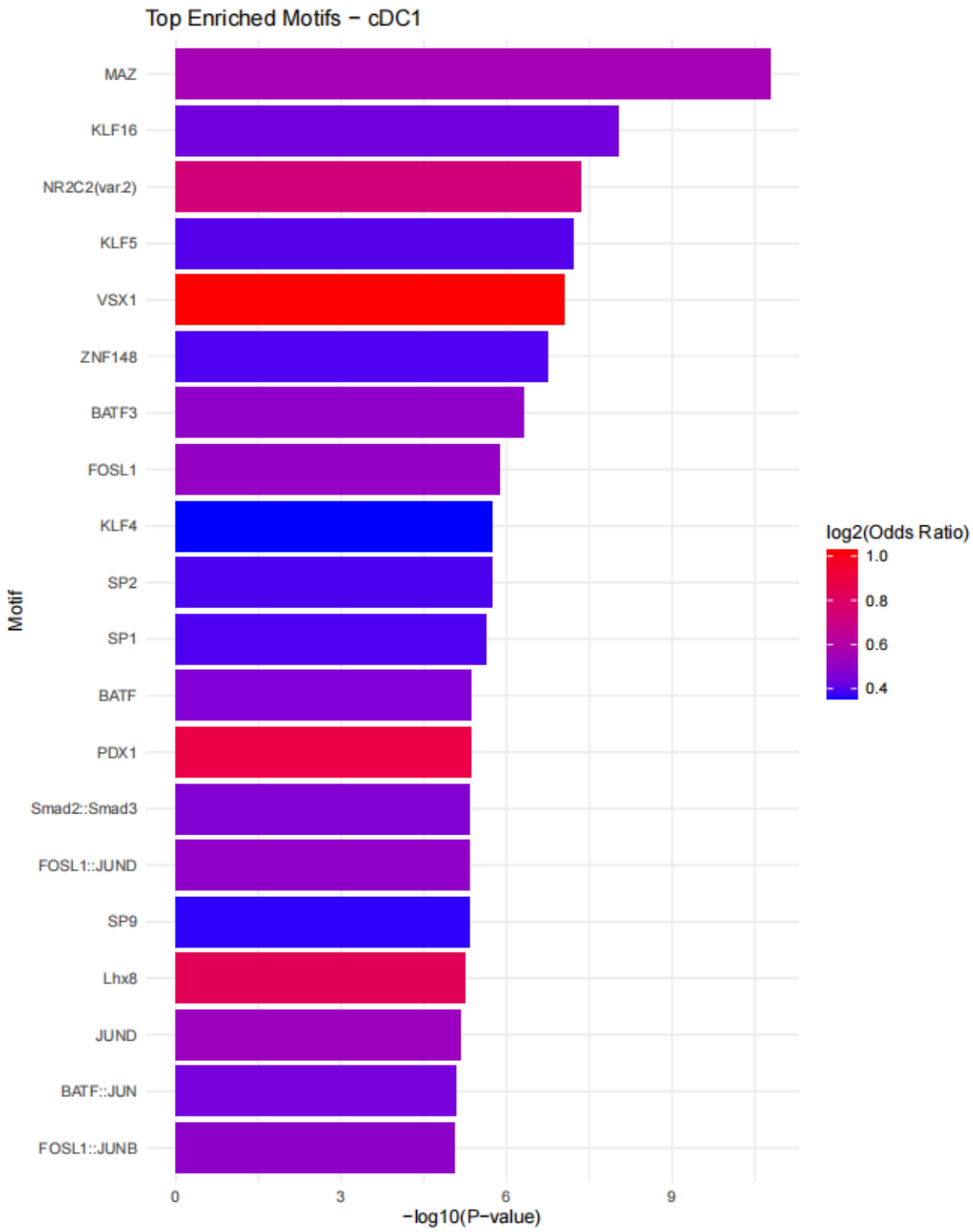


Figure 7. Top enriched motifs in cDC1 differential peaks showing BATF3/AP-1 enrichment

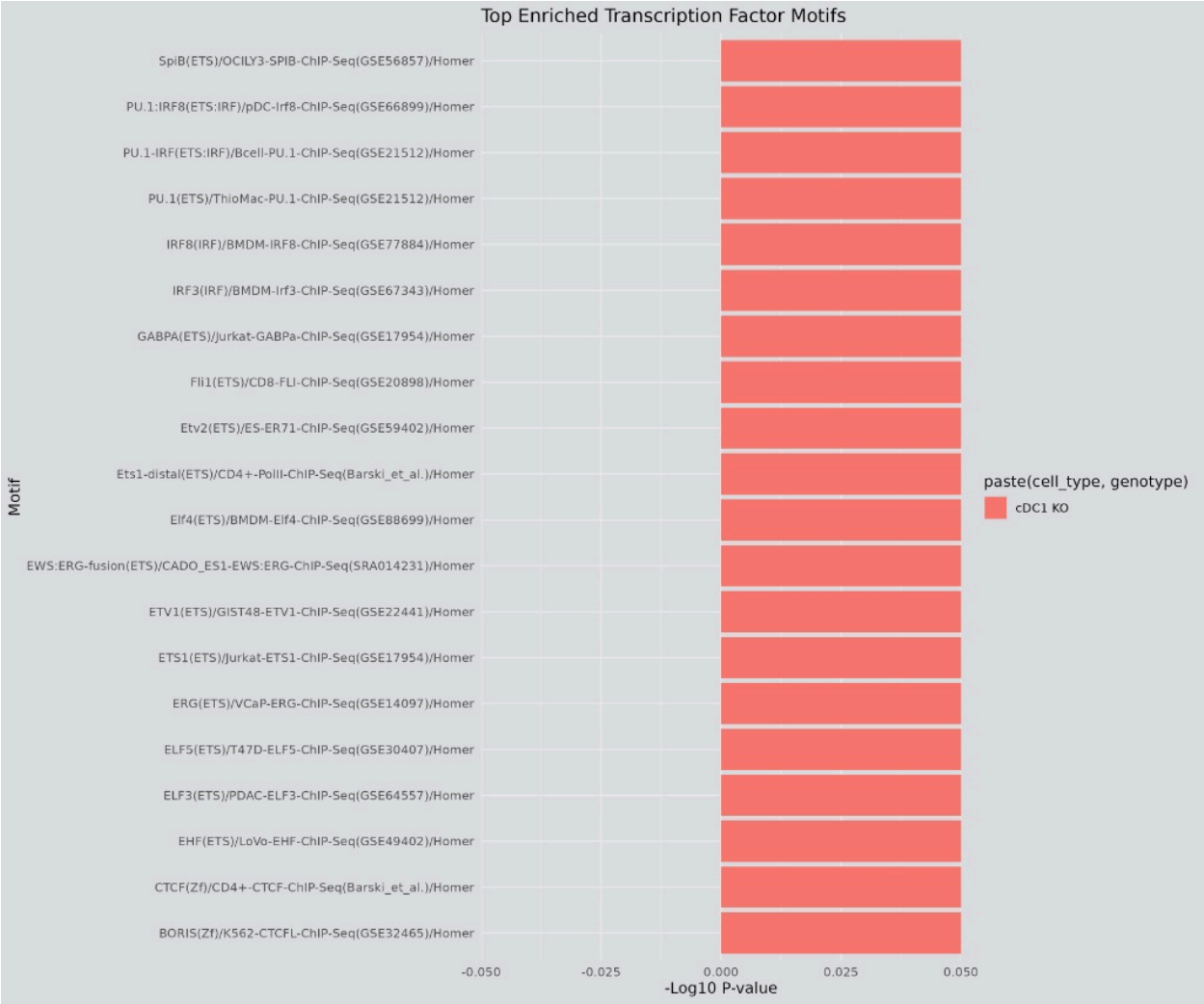


Figure 8. Top enriched motifs in cDC2 differential peaks showing SpiB/IRF8 enrichment

# 7. Reproduction Success and Conclusion

The reproduction of the original publication's findings was largely successful, particularly regarding the biological mechanisms.

- Consistency: Our multi-omics integration (Figure 9. & Figure 10.) mirrors panels 6C and 6E of the original paper. We successfully identified that HDAC1 loss leads to increased chromatin accessibility and upregulated expression of Maged1 and Nectin2 in cDC1s, and Spib and Csf3r in cDC2s. Furthermore, the motif analysis accurately re-identified the Spib/PU.1/IRF8 regulatory axis in cDC2s, confirming the core conclusion that HDAC1 epigenetically restricts these factors to define cDC2 identity.

- Differences: A key difference lies in the data quality metrics. Our analysis shows notably lower FRiP scores and fewer total differential peaks compared to the numbers reported in the publication (e.g., ~1600 DARs vs ~1800 for cDC1). Additionally, the single-end nature of the archived data prevented the visualization of nucleosome banding patterns seen in standard ATAC-seq libraries. Despite these technical limitations in the public dataset, the robustness of the biological signal allowed us to successfully replicate the key molecular discoveries of the study.

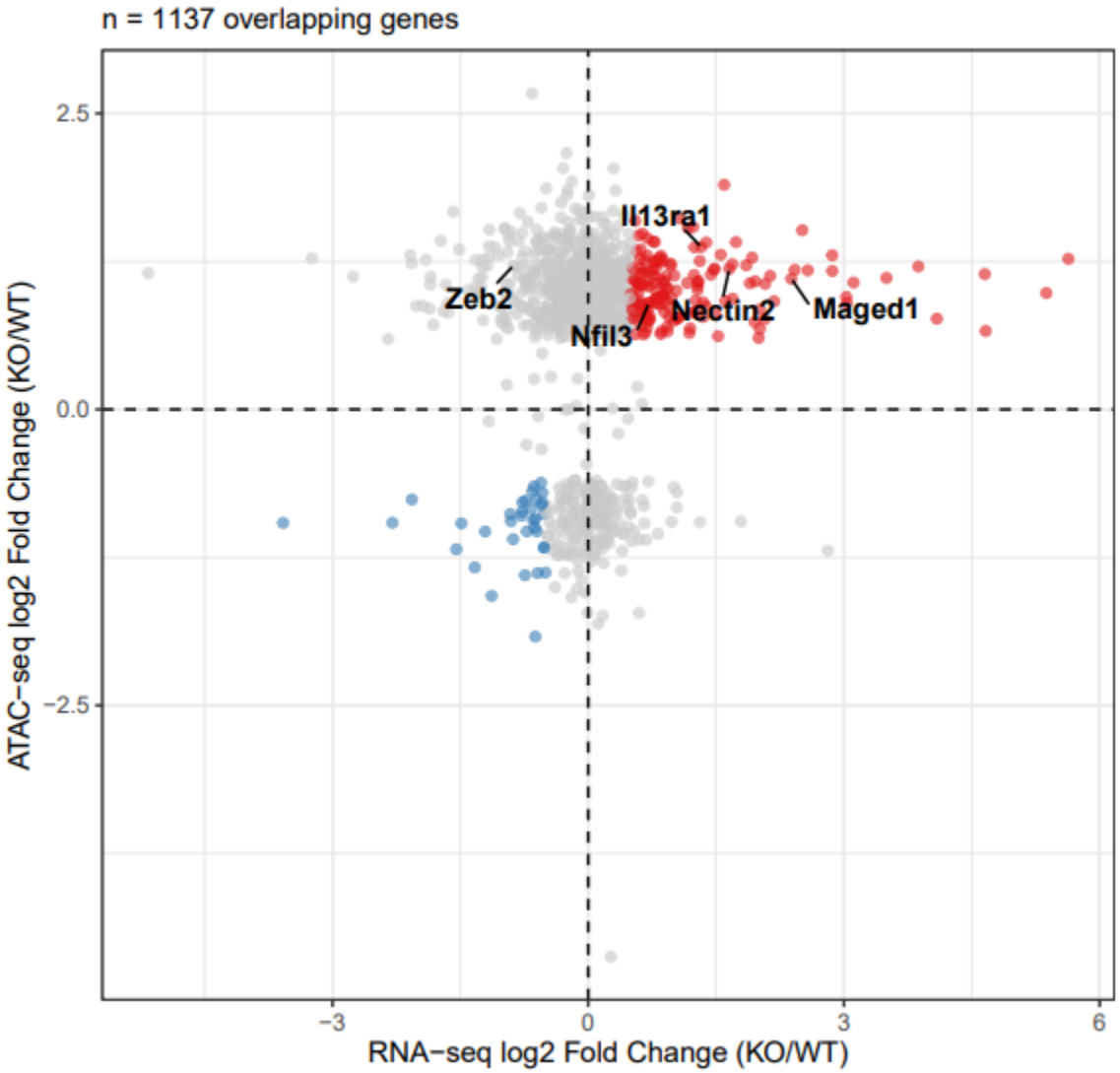


Figure 9. Integration of ATAC-seq and RNA-seq for cDC1

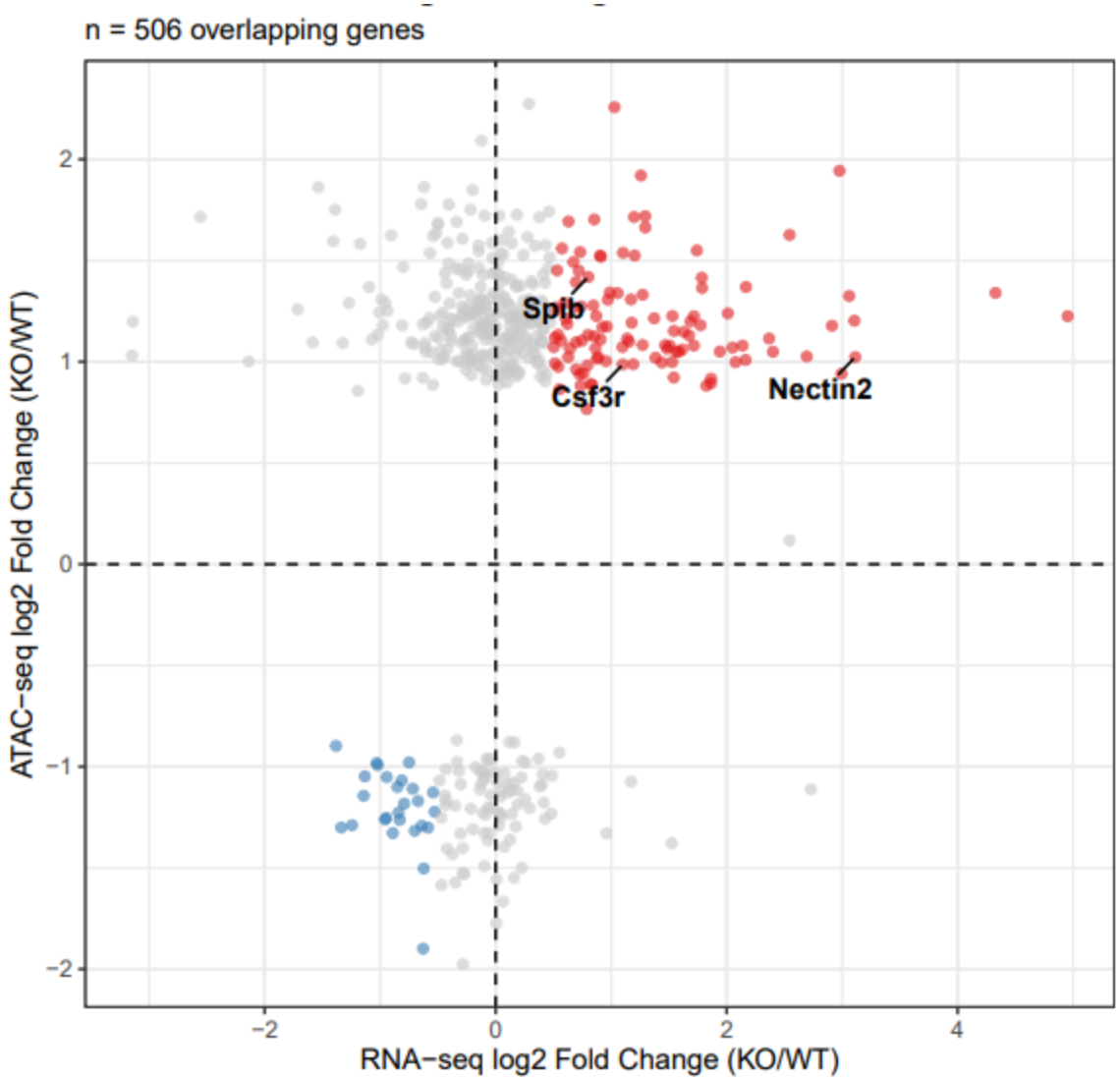


Figure 10. Integration of ATAC-seq and RNA-seq for cDC2