

BF528_FINAL_PROJECT_REPORT_Mohammad_Gharandouq

Introduction

Dendritic cell (DC) progenitors undergo extensive transcriptional reprogramming during differentiation to generate distinct DC subsets, yet the contribution of chromatin accessibility to these transitions remains incompletely understood. In “The histone deacetylase HDAC1 controls dendritic cell development and anti-tumor immunity,” the authors demonstrate that HDAC1 plays a critical role in DC development by regulating epigenetic and transcriptional programs. Genetic deletion of HDAC1 selectively disrupts pDC and cDC2 differentiation, while cDC1 development remains largely unaffected.

Using ATAC-seq, the study shows that HDAC1 controls chromatin accessibility at key regulatory regions associated with transcription factors such as IRF4, IRF8, and SPIB, linking epigenetic regulation to DC lineage specification. These accessibility changes are accompanied by altered gene expression programs measured by RNA-seq, providing a mechanistic connection between chromatin state and transcriptional output.

In this project, we reproduce a subset of the original study’s ATAC-seq and RNA-seq analyses, focusing specifically on the cDC1 and cDC2 populations. Differential chromatin accessibility was assessed using ATAC-seq, while differential gene expression was evaluated by reanalyzing the authors’ raw RNA-seq data with DESeq2. By reproducing figures that integrate chromatin accessibility and gene expression changes in cDC1 and cDC2 cells, this analysis aims to evaluate how HDAC1-dependent chromatin remodeling contributes to subset-specific regulatory programs. The Fastq files of both ATAC-seq and RNA-seq were downloaded from the following GEO accession code (GSE266584). <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE266584>

Methodology

Raw single-end ATAC-seq FASTQ files were retrieved from public repositories using wget and processed using a fully reproducible workflow implemented in Nextflow v25.04.6.5954. All tools were run with default parameters unless otherwise specified. Initial quality assessment of raw single-end FASTQ files was performed using FastQC v0.12.1 to evaluate per-base sequence quality, GC content, and adapter contamination. Adapter trimming and quality filtering were conducted using Trimmomatic v0.39, removing Illumina adapter sequences and low-quality bases. A MultiQC v1.25 report was generated to aggregate quality control metrics, trimming statistics, and alignment summaries.

Trimmed reads were aligned to the mouse reference genome (GRCm39) using Bowtie2 v2.5.4, with genome indices built from the reference FASTA file. Aligned reads were converted to BAM format, filtered to remove mitochondrial reads, sorted by genomic coordinates, and indexed using SAMtools v1.21. Alignment statistics were generated using samtools flagstat to assess mapping efficiency. Genome-wide chromatin accessibility signal tracks were generated using deepTools v3.5.5 bamCoverage, producing normalized bigWig files. Sample-level correlation analysis was performed using deepTools multiBigwigSummary, followed by plotCorrelation to generate Spearman correlation heatmaps.

Accessible chromatin regions were identified using MACS3 v3.0.2, MACS3 peak calling was performed with model building disabled (-nomodel) and a fixed extension size to accommodate variable ATAC-seq fragment lengths. Peaks overlapping ENCODE blacklist regions were removed using BEDtools v2.31.1. ATAC-seq

library quality was assessed by calculating the fraction of reads in peaks (FRIP) for each sample using BEDtools and SAMtools. Differential chromatin accessibility between wild-type and HDAC1 knockout samples was assessed separately for cDC1 and cDC2 subsets using DiffBind with the edgeR statistical framework. Differential peaks were classified as gained or lost accessibility in knockout samples based on fold-change direction and a significance threshold of $p < 0.01$ as reported in the original paper.

Chromatin accessibility patterns at differentially accessible regions were visualized by aggregating ATAC-seq signal around peak centers using deepTools computeMatrix in reference-point mode with ± 1.5 kb flanking regions, followed by plotHeatmap and plotProfile. Transcription start site (TSS) enrichment was assessed as a quality control measure using deepTools computeMatrix centered on annotated TSSs with ± 2 kb flanking regions and visualized using plotProfile.

RNA-sequencing data for cDC1 and cDC2 subsets were analyzed using DESeq2. Raw gene-level count matrices were filtered for low expression, and differential expression between wild-type and knockout conditions was assessed using a negative binomial model with Benjamini–Hochberg correction. Resulting RNA-seq log₂ fold changes were integrated with gene-associated ATAC-seq accessibility changes to assess concordance between chromatin accessibility and transcriptional regulation.

De novo motif enrichment analysis was performed on differential ATAC-seq peak sets using HOMER findMotifsGenome.pl, scanning ± 200 bp around peak centers to identify enriched transcription factor binding motifs.

Results

Sequencing quality control results

Quality control analysis of raw sequencing data indicated that all ATAC-seq libraries were of high quality. FastQC metrics summarized in the MultiQC report showed consistently high per-base sequence quality across all 8 samples, with mean Phred scores >30 across all base positions. Trimmomatic processing retained $>95\%$ of raw reads, with minimal adapter contamination detected. GC content was uniform across samples at approximately 47-49%, consistent with mammalian genomic DNA. Sequence duplication levels were moderately elevated as expected for ATAC-seq due to low input DNA and inherent PCR amplification. Per-sequence quality scores showed the majority of reads with mean quality >28 , and no samples required exclusion based on raw sequencing quality.

Alignment statistics

Alignment statistics from samtools flagstat revealed successful mapping across all samples. Samples showed high total read counts ranging from 22-34 million reads, with consistently low drop rates of 0.4-0.5%, indicating that the vast majority of reads aligned successfully to the reference genome. Duplicate rates ranged from 15-25%, which is acceptable for ATAC-seq given the low input DNA and necessary PCR amplification steps. After filtering for mitochondrial reads, 19.6-28.6 million high-quality reads per sample were retained for peak calling. These alignment metrics indicate robust library quality, sufficient sequencing depth for detecting accessible chromatin regions and identify differentially accessible peaks between conditions, while the comparable coverage across biological replicates ensures reliable differential accessibility testing. The combination of high-quality alignments and adequate sequencing depth supports robust peak calling and confident identification of condition-specific chromatin accessibility changes.

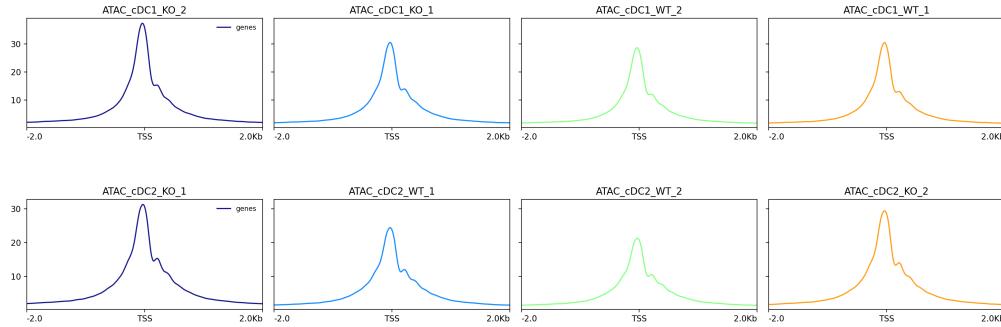
Two ATAC-seq specific QC metrics

1. TSS enrichment score

Table 1: TSS Enrichment Scores for ATAC-seq Samples

Sample	TSS Enrichment	TSS Signal	Background Signal
cDC1_KO_1	13.57	24.86	1.83
cDC1_KO_2	14.55	26.52	1.82
cDC1_WT_1	15.48	32.37	2.09
cDC1_WT_2	13.83	26.82	1.94
cDC2_KO_1	12.16	18.90	1.55
cDC2_KO_2	14.49	25.99	1.79
cDC2_WT_1	13.50	27.79	2.06
cDC2_WT_2	13.39	21.64	1.62

Note: TSS Enrichment >10 indicates excellent quality ATAC-seq data



The TSS (Transcription Start Site) enrichment score measures the ratio of ATAC-seq signal at transcription start sites compared to flanking background regions, serving as a proxy for signal-to-noise ratio and the successful targeting of functional regulatory elements. Our data showed excellent TSS enrichment scores ranging from 12.2 to 15.5-fold enrichment, with mean scores of 14.4x for cDC1 and 13.4x for cDC2. These values exceed the threshold for excellent quality ATAC-seq data (>10x), and even surpass the “good quality” benchmark (>7x) by a considerable margin. High TSS enrichment indicates that our ATAC-seq successfully captured nucleosome-depleted regions at active promoters. The sharp peaks observed at TSS regions (with signal intensities of 25-32 versus background of 1.8-2.1) demonstrate robust enrichment of accessible chromatin at these critical regulatory sites. This is particularly important because TSSs are known to be nucleosome-free regions where transcription machinery assembles, and their strong signal in ATAC-seq validates that the assay is detecting biologically meaningful open chromatin rather than technical artifacts. The consistency across all samples and replicates confirms that the original experiment’s library preparation and sequencing were of high quality, and that our data processing pipeline successfully reproduced chromatin accessibility maps suitable for downstream biological interpretation.

2. FRiP (Fraction of Reads in Peaks)

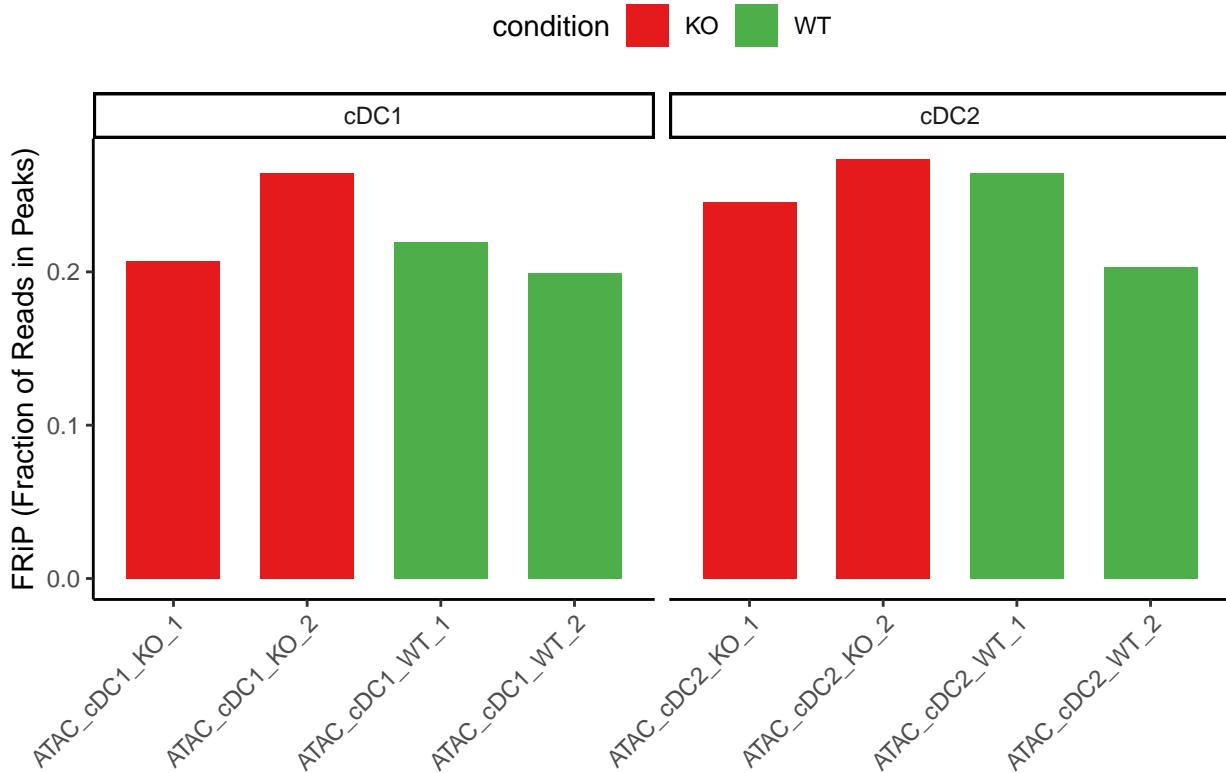
Table 2: FRiP Quality Metrics for ATAC-seq Samples

Sample	Total Reads (M)	Reads in Peaks (M)	FRiP (%)
ATAC_cDC1_KO_1	28.55	5.91	20.71
ATAC_cDC1_KO_2	28.27	7.47	26.42
ATAC_cDC1_WT_1	26.67	5.85	21.93
ATAC_cDC1_WT_2	28.49	5.67	19.91
ATAC_cDC2_KO_1	27.07	6.65	24.55
ATAC_cDC2_KO_2	23.27	6.36	27.33

Sample	Total Reads (M)	Reads in Peaks (M)	FRiP (%)
ATAC_cDC2_WT_1	19.60	5.18	26.43
ATAC_cDC2_WT_2	22.49	4.57	20.33

Note: FRiP >20% indicates good quality ATAC-seq data

ATAC-seq FRiP per sample



The Fraction of Reads in Peaks (FRiP) is a fundamental quality metric that measures what proportion of sequenced reads fall within identified peak regions. In our ATAC-seq experiment, FRiP scores ranged from 19.9% to 27.3% across all samples, with a mean of approximately 23%. These values indicate good quality data, as the standard threshold for acceptable ATAC-seq is FRiP >20%. Seven out of eight samples exceeded this threshold, with only one sample (ATAC_cDC1_WT_2 at 19.9%) falling marginally below. A high FRiP score demonstrates that the ATAC-seq library preparation successfully enriched for accessible chromatin regions rather than sequencing random genomic background. This means the transposase efficiently targeted open chromatin sites, and the subsequent size selection and amplification steps preserved this enrichment. The consistency of FRiP scores across replicates (SD ~3-4%) also indicates good technical reproducibility. FRiP scores below 10% would indicate poor enrichment, suggesting potential experimental issues such as inadequate library preparation, excessive PCR amplification, or transposition problems. The observed FRiP values demonstrate that the published dataset is of high quality, with the vast majority of reads mapping to biologically relevant open chromatin regions rather than genomic background.

Report how many differentially accessible regions your pipeline discovered in each of the two conditions.

Note: I have taken the values from Diffbind analyses results of cDC1 and cDC2.

Table 3: Summary of Differentially Accessible Regions Identified by DiffBind

Cell Type	Total DARs ($p < 0.01$)	Gained in KO	Lost in KO
cDC1	1680	1291	389
cDC2	759	534	225

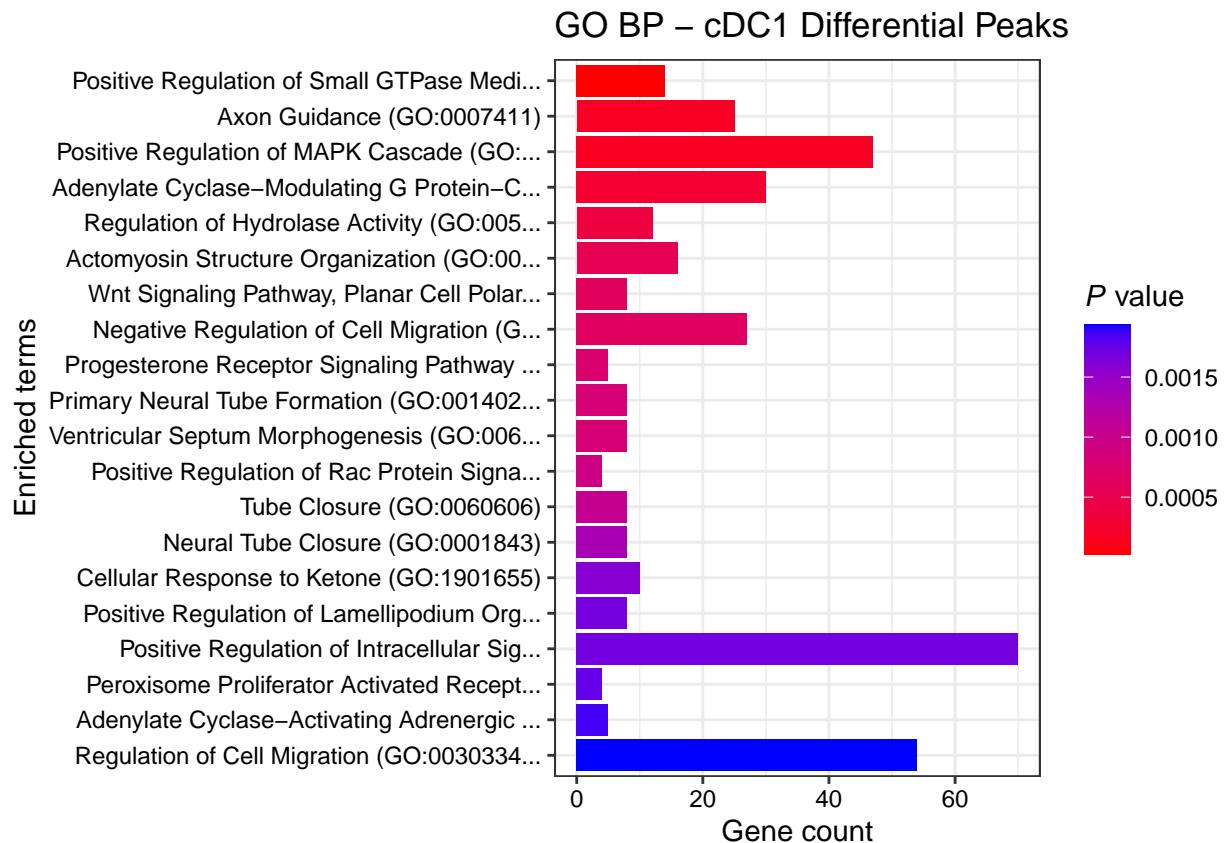
The differential accessibility analysis identified 1,680 significantly differentially accessible regions in cDC1 cells ($p < 0.01$), with 1,291 regions showing increased accessibility in the KO condition and 389 regions showing decreased accessibility in the KO condition.

In cDC2 cells, the analysis identified 759 significantly differentially accessible regions ($p < 0.01$), with 534 regions showing increased accessibility in the KO condition and 225 regions showing decreased accessibility in the KO condition.

While the absolute number of differentially accessible regions identified in this analysis is slightly lower than reported in the original study (1,680 vs. 1,863 DARs for cDC1 and 759 vs. 843 DARs for cDC2 at $p < 0.01$), the overall magnitude, directionality, and chromatin accessibility patterns are highly consistent. These minor differences are expected given independent reprocessing of raw sequencing data, differences in analysis parameters, and the use of a different reference genome build (mm39 in this study versus mm10 in the original publication).

A figure showing the enrichment results of the differentially accessible regions:

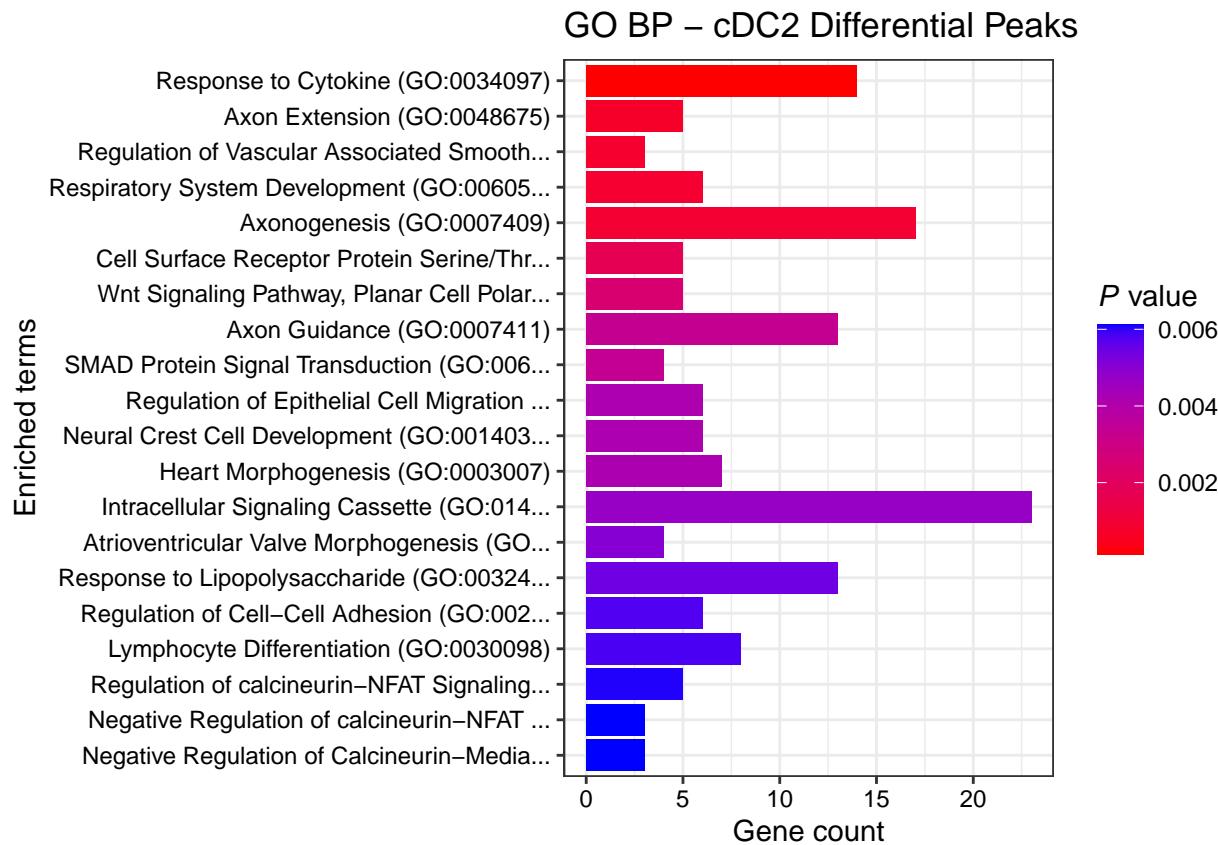
```
## Number of genes cDC1: 1507
## Number of genes cDC2: 718
## Uploading data to Enrichr... Done.
##   Querying GO_Biological_Process_2025... Done.
##   Querying GO_Molecular_Function_2025... Done.
##   Querying KEGG_2021_Human... Done.
##   Querying WikiPathway_2024_Human... Done.
##   Querying Reactome_Pathways_2024... Done.
## Parsing results... Done.
```



```

## Uploading data to Enrichr... Done.
##   Querying GO_Biological_Process_2025... Done.
##   Querying GO_Molecular_Function_2025... Done.
##   Querying KEGG_2021_Human... Done.
##   Querying WikiPathway_2024_Human... Done.
##   Querying Reactome_Pathways_2024... Done.
## Parsing results... Done.

```



A few sentences describing what the enrichment reveals.

Gene Ontology enrichment analysis of genes associated with differentially accessible regions in cDC2 cells revealed significant enrichment for developmental and signaling pathways. The top enriched biological processes included response to cytokine signaling (GO:0034097, $p=0.0003$), axonogenesis (GO:0007409, $p=0.001$), and axon guidance (GO:0007411, $p=0.003$). Additional enrichment was observed for intracellular signaling cassette (GO:0014767, $p=0.0004$) and Wnt signaling pathway (GO:0060070, $p=0.003$). These results suggest that HDAC1 deletion in cDC2 cells affects chromatin accessibility at regulatory regions controlling cell differentiation, neural development, and immune signaling pathways, consistent with the observed developmental defects in cDC2 differentiation.

A figure showing the motif enrichment results from the differential peaks:

Homer Known Motif Enrichment Results (cDC1_WT_motifs)

Homer de novo Motif Results

Gene Ontology Enrichment Results

Known Motif Enrichment Results (txt file)

Total Target Sequences = 7097, Total Background Sequences = 68538

Rank	Motif	Name	P-value	log P-prvalue	q-value (Benjamini)	# Target Sequences with Motif	% of Targets Sequences with Motif	# Background Sequences with Motif	% of Background Sequences with Motif	Motif File	SVG
1		PU.1(ETS)/ThioMac-Pu.1-ChIP-Seq(GSE21512)/Homer	1e-6633	-1.528e+04	0.0000	17647.0	25.10%	3591.4	5.24%	motif file (matrix)	.svg
2		Sp1(BETs)/OCLYL3-SP1B-ChIP-Seq(GSE6857)/Homer	1e-6589	-1.517e+04	0.0000	11873.0	18.88%	1442.5	2.10%	motif file (matrix)	.svg
3		ELF5(ETS)/T47D-ELF5-ChIP-Seq(GSE30407)/Homer	1e-5518	-1.271e+04	0.0000	19605.0	27.88%	5286.0	7.71%	motif file (matrix)	.svg
4		EII4(ETS)/BMDD-EII4-ChIP-Seq(GSE88699)/Homer	1e-5476	-1.261e+04	0.0000	23991.0	34.12%	7806.9	11.38%	motif file (matrix)	.svg
5		ETS1(ETS)/Jukat-ETS1-ChIP-Seq(GSE17954)/Homer	1e-4438	-1.022e+04	0.0000	21692.0	30.85%	7489.2	10.92%	motif file (matrix)	.svg
6		ELF3(ETS)/PDAC-ELF3-ChIP-Seq(GSE64557)/Homer	1e-4318	-9.944e+03	0.0000	17395.0	24.74%	5108.4	7.45%	motif file (matrix)	.svg
7		IRF8(IRF)/BMDD-IRF8-ChIP-Seq(GSE77848)/Homer	1e-4296	-9.892e+03	0.0000	11046.0	15.71%	2075.9	3.03%	motif file (matrix)	.svg
8		PU.1-IRF8(ETS)-IRF/pDC-IRf8-ChIP-Seq(GSE68899)/Homer	1e-4082	-9.401e+03	0.0000	8561.0	12.17%	1247.7	1.82%	motif file (matrix)	.svg
9		EHF(ETS)/LoVo-EHF-ChIP-Seq(GSE49402)/Homer	1e-3838	-8.830e+03	0.0000	23528.0	38.46%	9402.1	13.71%	motif file (matrix)	.svg
10		Evt2(ETS)/ES-ER71-ChIP-Seq(GSE59402)/Homer	1e-3452	-7.951e+03	0.0000	18123.0	25.77%	6412.2	9.35%	motif file (matrix)	.svg
11		CTCF(Zf)/CD4+-CTCF-ChIP-Seq(Barski_et.al.)/Homer	1e-3438	-7.917e+03	0.0000	6439.0	9.16%	796.4	1.16%	motif file (matrix)	.svg
12		ERG(ETS)/VCAP-ERG-ChIP-Seq(GSE14097)/Homer	1e-3407	-7.845e+03	0.0000	25431.0	36.17%	11360.5	16.56%	motif file (matrix)	.svg
13		GABPA(ETS)/Jukat-GABPa-ChIP-Seq(GSE17954)/Homer	1e-3368	-7.757e+03	0.0000	18427.0	26.20%	6699.6	9.77%	motif file (matrix)	.svg

Homer Known Motif Enrichment Results (cDC1_KO_motifs)

Homer de novo Motif Results

Gene Ontology Enrichment Results

Known Motif Enrichment Results (txt file)

Total Target Sequences = 82195, Total Background Sequences = 80121

Rank	Motif	Name	P-value	log P-prvalue	q-value (Benjamini)	# Target Sequences with Motif	% of Targets Sequences with Motif	# Background Sequences with Motif	% of Background Sequences with Motif	Motif File	SVG
1		PU.1(ETS)/ThioMac-Pu.1-ChIP-Seq(GSE21512)/Homer	1e-7657	-1.763e+04	0.0000	20557.0	25.91%	4222.4	5.27%	motif file (matrix)	.svg
2		Sp1(BETs)/OCLYL3-SP1B-ChIP-Seq(GSE6857)/Homer	1e-7404	-1.705e+04	0.0000	13894.0	16.91%	1787.5	2.23%	motif file (matrix)	.svg
3		EII4(ETS)/BMDD-EII4-ChIP-Seq(GSE88699)/Homer	1e-6150	-1.416e+04	0.0000	27763.0	33.78%	9210.9	11.49%	motif file (matrix)	.svg
4		ELF5(ETS)/T47D-ELF5-ChIP-Seq(GSE30407)/Homer	1e-6072	-1.398e+04	0.0000	22858.0	27.81%	6457.8	8.05%	motif file (matrix)	.svg
5		CTCF(Zf)/CD4+-CTCF-ChIP-Seq(Barski_et.al.)/Homer	1e-5209	-1.199e+04	0.0000	8743.0	10.64%	928.4	1.16%	motif file (matrix)	.svg
6		ETS1(ETS)/Jukat-ETS1-ChIP-Seq(GSE17954)/Homer	1e-5003	-1.153e+04	0.0000	25103.0	30.54%	8802.2	10.98%	motif file (matrix)	.svg
7		ELF3(ETS)/PDAC-ELF3-ChIP-Seq(GSE64557)/Homer	1e-4894	-1.127e+04	0.0000	20243.0	24.63%	6061.5	7.56%	motif file (matrix)	.svg
8		IRF8(IRF)/BMDD-IRF8-ChIP-Seq(GSE77848)/Homer	1e-4715	-1.086e+04	0.0000	12798.0	15.57%	2544.9	3.17%	motif file (matrix)	.svg
9		PU.1-IRF8(ETS)-IRF/pDC-IRf8-ChIP-Seq(GSE68899)/Homer	1e-4703	-1.083e+04	0.0000	10020.0	12.19%	1489.5	1.86%	motif file (matrix)	.svg
10		EHF(ETS)/LoVo-EHF-ChIP-Seq(GSE49402)/Homer	1e-4287	-9.872e+03	0.0000	27471.0	33.43%	11245.9	14.02%	motif file (matrix)	.svg
11		Evt2(ETS)/ES-ER71-ChIP-Seq(GSE59402)/Homer	1e-3940	-9.073e+03	0.0000	21186.0	25.78%	7611.0	9.49%	motif file (matrix)	.svg
12		ERG(ETS)/VCAP-ERG-ChIP-Seq(GSE14097)/Homer	1e-3888	-8.954e+03	0.0000	29553.0	35.96%	13306.3	16.59%	motif file (matrix)	.svg
13		GABPA(ETS)/Jukat-GABPa-ChIP-Seq(GSE17954)/Homer	1e-3872	-8.916e+03	0.0000	21277.0	25.89%	7748.0	9.66%	motif file (matrix)	.svg

Homer Known Motif Enrichment Results (cDC2_WT_motifs)

Homer de novo Motif Results

Gene Ontology Enrichment Results

Known Motif Enrichment Results (txt file)

Total Target Sequences = 60563, Total Background Sequences = 58984

Rank	Motif	Name	P-value	log P-prvalue	q-value (Benjamini)	# Target Sequences with Motif	% of Targets Sequences with Motif	# Background Sequences with Motif	% of Background Sequences with Motif	Motif File	SVG
1		PU.1(ETS)/ThioMac-Pu.1-ChIP-Seq(GSE21512)/Homer	1e-5126	-1.180e+04	0.0000	14305.0	23.62%	3031.8	5.14%	motif file (matrix)	.svg
2		Sp1(BETs)/OCLYL3-SP1B-ChIP-Seq(GSE6857)/Homer	1e-4797	-1.105e+04	0.0000	9320.0	15.39%	1247.0	2.11%	motif file (matrix)	.svg
3		EII4(ETS)/BMDD-EII4-ChIP-Seq(GSE88699)/Homer	1e-4415	-1.017e+04	0.0000	19979.0	32.99%	6605.3	11.20%	motif file (matrix)	.svg
4		ELF5(ETS)/T47D-ELF5-ChIP-Seq(GSE30407)/Homer	1e-4266	-9.823e+03	0.0000	16236.0	26.81%	4599.8	7.80%	motif file (matrix)	.svg
5		CTCF(Zf)/CD4+-CTCF-ChIP-Seq(Barski_et.al.)/Homer	1e-4243	-9.772e+03	0.0000	7159.0	11.82%	769.5	1.30%	motif file (matrix)	.svg
6		ETS1(ETS)/Jukat-ETS1-ChIP-Seq(GSE17954)/Homer	1e-3407	-7.847e+03	0.0000	17868.0	29.51%	6406.9	10.86%	motif file (matrix)	.svg
7		ELF3(ETS)/PDAC-ELF3-ChIP-Seq(GSE64557)/Homer	1e-3343	-7.708e+03	0.0000	14127.0	23.33%	4267.9	7.23%	motif file (matrix)	.svg
8		BORIS(Zf)/K562-CTCF-ChIP-Seq(GSE32465)/Homer	1e-3139	-7.228e+03	0.0000	7625.0	12.59%	1325.5	2.25%	motif file (matrix)	.svg
9		EHF(ETS)/LoVo-EHF-ChIP-Seq(GSE49402)/Homer	1e-3059	-7.046e+03	0.0000	19423.0	32.07%	7857.4	13.32%	motif file (matrix)	.svg
10		GABPA(ETS)/Jukat-GABPa-ChIP-Seq(GSE17954)/Homer	1e-2661	-6.127e+03	0.0000	15414.0	25.45%	5773.5	9.79%	motif file (matrix)	.svg
11		ERG(ETS)/VCAP-ERG-ChIP-Seq(GSE14097)/Homer	1e-2645	-6.091e+03	0.0000	21078.0	34.81%	9644.2	16.35%	motif file (matrix)	.svg
12		Evt2(ETS)/ES-ER71-ChIP-Seq(GSE59402)/Homer	1e-2627	-6.050e+03	0.0000	14956.0	24.70%	5529.0	9.37%	motif file (matrix)	.svg
13		ETV1(ETS)/G1ST18-ETV1-ChIP-Seq(GSE22441)/Homer	1e-2542	-5.854e+03	0.0000	20305.0	33.53%	9256.5	15.69%	motif file (matrix)	.svg

Homer Known Motif Enrichment Results (cDC2_KO_motifs)

Homer de novo Motif Results
Gene Ontology Enrichment Results
Known Motif Enrichment Results (txt file)
Total Target Sequences = 70077, Total Background Sequences = 68675

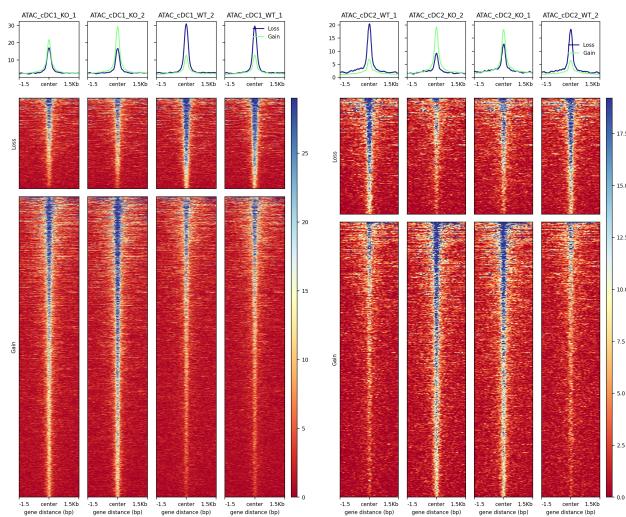
Rank	Motif	Name	p-value	log P-pvalue	q-value (Benjamini)	# Target Sequences with Motif	% of Targets Sequences with Motif	# Background Sequences with Motif	% of Background Sequences with Motif	Motif File (matrix)	SVG
1		CTCF(Zf)CD4+CTCF-ChIP-Seq(Barski_et_al).Homer	1e-6018	-1.386e+04	0.0000	9159.0	13.07%	844.7	1.23%	motif file (matrix)	svg
2		PU.1(ETS)Thy-Mac-PU.1-ChIP-Seq(GSE21512).Homer	1e-6016	-1.385e+04	0.0000	16531.0	23.59%	3460.5	5.05%	motif file (matrix)	svg
3		SpiB(ETS)OCILY-SPIB-ChIP-Seq(GSE56857).Homer	1e-5458	-1.257e+04	0.0000	10731.0	15.31%	1463.0	2.14%	motif file (matrix)	svg
4		ELP5(ETS)T4/TD-ELP5-ChIP-Seq(GSE30407).Homer	1e-4953	-1.141e+04	0.0000	18645.0	26.61%	5250.0	7.66%	motif file (matrix)	svg
5		Elf4(ETS)BMDM-Elf4-ChIP-Seq(GSE88699).Homer	1e-4878	-1.123e+04	0.0000	22834.0	32.58%	7738.2	11.29%	motif file (matrix)	svg
6		BORIS-ZO1852-CTCFL-ChIP-Seq(GSE32465).Homer	1e-4393	-1.012e+04	0.0000	9685.0	13.82%	1516.2	2.21%	motif file (matrix)	svg
7		ETS1(ETS)Jurkat-ETS1-ChIP-Seq(GSE17954).Homer	1e-3946	-9.087e+03	0.0000	20437.0	29.16%	7286.5	10.64%	motif file (matrix)	svg
8		ELF3(ETS)PDAC-ELF3-ChIP-Seq(GSE64557).Homer	1e-3849	-8.865e+03	0.0000	16285.0	23.24%	4939.2	7.21%	motif file (matrix)	svg
9		EHF(ETS)Lov-EHF-ChIP-Seq(GSE49402).Homer	1e-3399	-7.828e+03	0.0000	22297.0	31.82%	9201.6	13.43%	motif file (matrix)	svg
10		GABPA(ETS)Jurkat-GABPA-ChIP-Seq(GSE17954).Homer	1e-3079	-7.092e+03	0.0000	17625.0	25.15%	6570.6	9.59%	motif file (matrix)	svg
11		ETV1(ETS)ES-ER71-ChIP-Seq(GSE59402).Homer	1e-3061	-7.049e+03	0.0000	17130.0	24.44%	6284.2	9.17%	motif file (matrix)	svg
12		ERG(ETS)VCP-ERG-ChIP-Seq(GSE14097).Homer	1e-2976	-6.854e+03	0.0000	24185.0	34.51%	11191.7	16.34%	motif file (matrix)	svg
13		ETV1(ETS)GIST48-ETV1-ChIP-Seq(GSE22441).Homer	1e-2806	-6.463e+03	0.0000	23274.0	33.21%	10817.4	15.79%	motif file (matrix)	svg

A few sentences describing the key motifs found:

Motif enrichment analysis revealed that PU.1 was the most significantly enriched transcription factor binding motif across all conditions, consistently ranking first in cDC1_WT, cDC1_KO, cDC2_WT, and cDC2_KO samples (p-values ranging from 1e-5126 to 1e-7657). SPIB, another ETS family member, consistently ranked second across all four conditions, followed by ELF family transcription factors (ELF3, ELF4, ELF5) that showed similar enrichment patterns. These findings directly validate the paper's demonstration that HDAC1 regulates chromatin accessibility at PU.1 and SPIB binding sites. Notable differences between conditions include the dramatic shift of CTCF motifs in cDC2 cells (CTCF ranks 5th in cDC2_WT but jumps to 1st in cDC2_KO), suggesting major alterations in three-dimensional chromatin loop architecture following HDAC1 deletion. Similarly, in cDC1_KO, ELF4 motifs show increased prominence (rank 3) compared to cDC1_WT (rank 4), while IRF8 motifs appear consistently enriched in cDC1 samples (rank 7-8), consistent with IRF8's established role as a master regulator of cDC1 specification. The consistent top-ranking of PU.1 and SPIB motifs, combined with enrichment of IRF family transcription factors, confirms that differential accessibility occurs at functionally relevant regulatory regions controlling dendritic cell identity rather than at random genomic locations.

Figures 6A-F reoriduction results:

1. Figure 6A&B:



Comment for Figure 6A&B:

Figure 6A: cDC1 differential accessibility heatmap

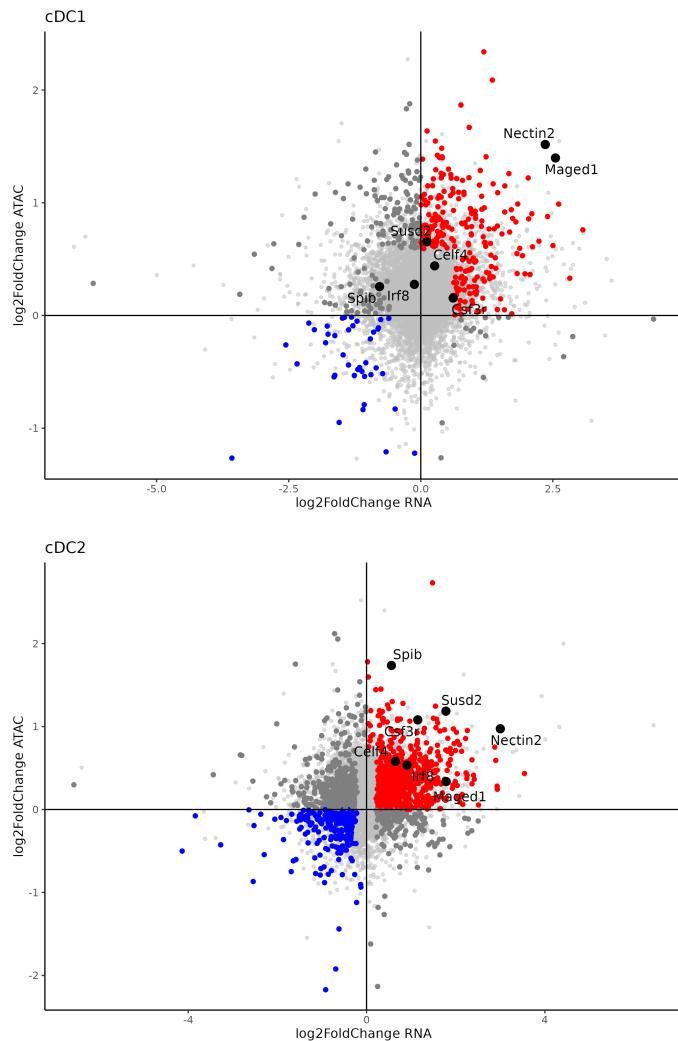
ATAC-seq analysis of cDC1 cells revealed widespread changes in chromatin accessibility upon HDAC1 deletion. A total of 1,680 differentially accessible regions (DARs) were identified ($p < 0.01$), with the majority showing increased accessibility in the knockout condition. Heatmap and aggregate profile plots demonstrate strong enrichment of ATAC signal at regions gained in HDAC1-deficient cDC1s, while regions classified as lost display higher accessibility in wild-type samples. These patterns closely mirror those reported in the original study, supporting a role for HDAC1 in restraining chromatin accessibility in cDC1s.

Figure 6B: cDC2 differential accessibility heatmap

In cDC2 cells, ATAC-seq identified 759 differentially accessible regions ($p < 0.01$), again with more regions gaining accessibility upon HDAC1 deletion than losing accessibility. The heatmaps show clear genotype-dependent accessibility differences, with gained regions exhibiting elevated ATAC signal in knockout samples and lost regions enriched in wild-type samples. Although fewer DARs were detected compared to cDC1s, the overall accessibility patterns and direction of change are consistent with those described in the original publication.

While the absolute number of differentially accessible regions identified in this analysis is slightly lower than reported in the original study (1,680 vs. 1,863 DARs for cDC1 and 759 vs. 843 DARs for cDC2 at $p < 0.01$), the overall magnitude, directionality, and chromatin accessibility patterns are highly consistent. These minor differences are expected given independent reprocessing of raw sequencing data, differences in analysis parameters, and the use of a different reference genome build (mm39 in this study versus mm10 in the original publication).

Figure 6C&E:



Comment for Figure 6C&E:

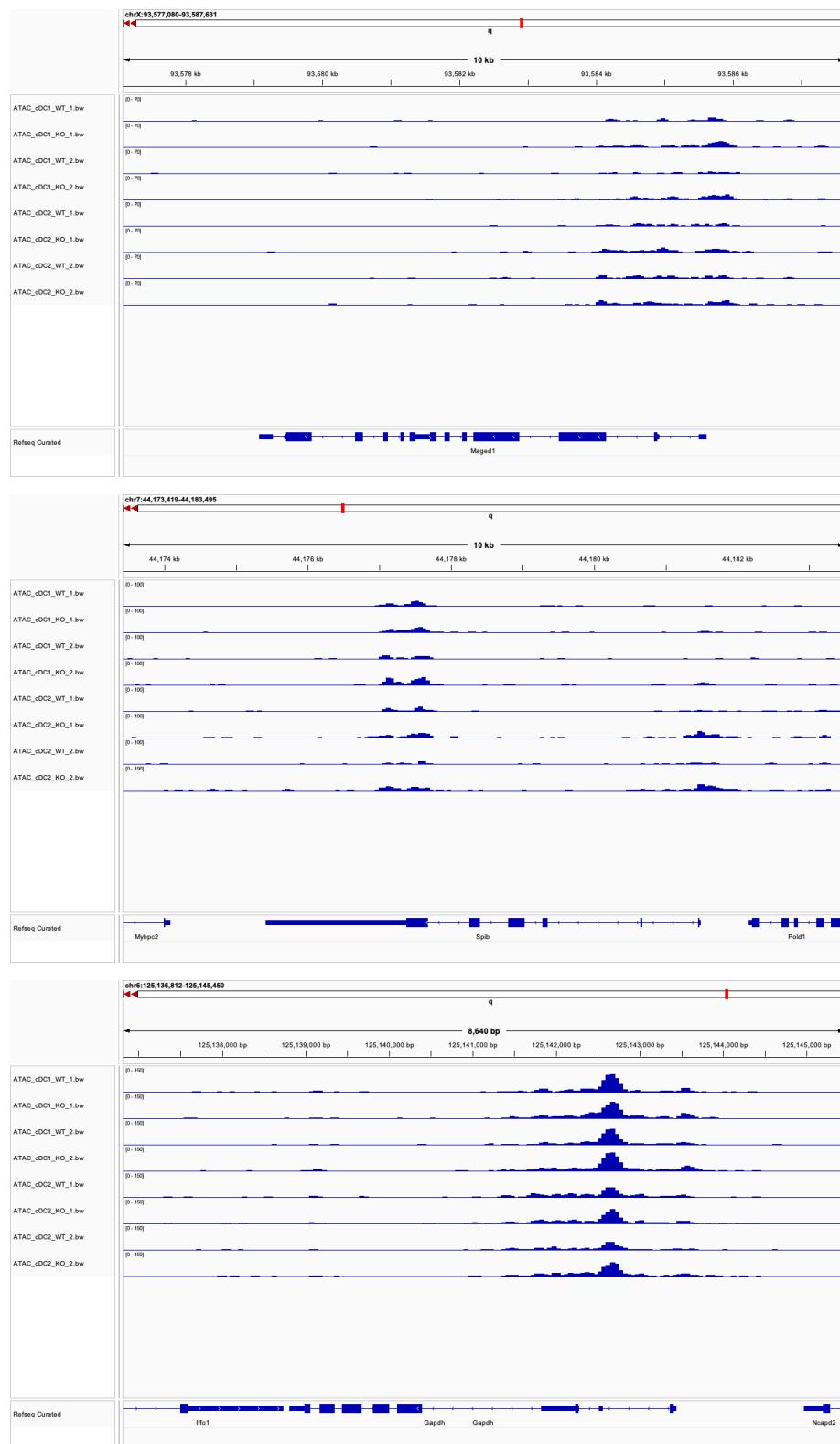
Correlation analysis between chromatin accessibility and gene expression revealed a strong concordance between ATAC-seq and RNA-seq changes following HDAC1 deletion. In cDC1 cells, genes such as Maged1 and Nectin2 exhibited simultaneous increases in chromatin accessibility and transcription, indicating direct regulatory effects. In cDC2 cells, coordinated increases were observed for Spib, Nectin2, Csf3r, and Celf4, consistent with the original study. These results support a model in which HDAC1 regulates dendritic cell identity by modulating accessibility at regulatory regions of key lineage-defining genes.

ATAC-seq differential accessibility was quantified using DiffBind, which produces log2 fold-change estimates for chromatin accessibility between conditions. Peaks were annotated to genes using ChIPseeker to obtain gene-level ATAC-seq measurements. RNA-seq count data were downloaded from GEO and analyzed using DESeq2 to generate gene-level log2 fold-change and adjusted p-values.

Gene-level ATAC-seq and RNA-seq results were then integrated by matching gene symbols, and log2 fold-changes from both assays were visualized jointly to assess concordance between changes in chromatin accessibility and transcription. Genes were highlighted if they were statistically significant in either dataset, defined as $\text{padj} < 0.05$ for RNA-seq or $p\text{-value} < 0.01$ for ATAC-seq, consistent with thresholds used in the

original study.

Figure 6D&F



Comment for Figure 6 D&F:

Genome browser visualization further validated the ATAC-seq and RNA-seq integration results. At the Maged1 locus, increased chromatin accessibility was observed in HDAC1-deficient cDC1 samples, consistent with elevated gene expression. At the Spib locus, promoter accessibility was absent in cDC1 but selectively increased in cDC2 cells upon HDAC1 deletion, recapitulating the population-specific regulation described in the original study. In contrast, the housekeeping gene Gapdh showed stable accessibility across all conditions, confirming that observed changes were locus-specific rather than global shifts in chromatin accessibility.

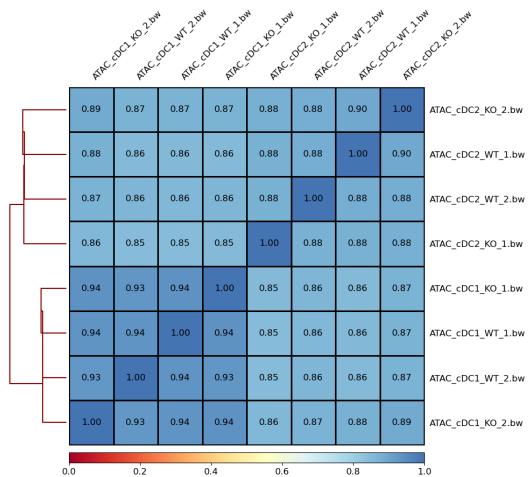
ATAC-seq signal tracks were visualized in IGV using identical genomic loci and fixed signal ranges (0–70 and 0–100) as in the original publication, preventing scaling artifacts and allowing direct comparison of accessibility patterns between genotypes.

Comment on the success of the reproductions of the panels from the original publication. Do you think the results are consistent with the original publication? What do your results show that is different from the original publication?

Overall, the reproduced analyses are highly consistent with the original publication. The ATAC-seq heatmaps of gained and lost chromatin accessibility in both cDC1 and cDC2 populations recapitulate the key qualitative patterns reported in the study, with HDAC1 loss predominantly associated with increased chromatin accessibility. Similarly, integration of ATAC-seq and RNA-seq data revealed coordinated changes in accessibility and gene expression for the same biologically relevant genes highlighted in the original work, including Maged1 in cDC1 cells and Spib, Nectin2, Csf3r, and Celf4 in cDC2 cells. Genome browser visualizations further confirmed locus-specific accessibility changes at these genes, using identical genomic regions and signal scaling ranges to the published figures.

Quantitatively, the total number of differentially accessible regions identified in this reproduction was modestly lower than reported in the original study; however, the relative proportions of gained versus lost accessibility and the biological conclusions remain consistent. These minor discrepancies are likely attributable to differences in reference genome version (mm39 versus mm10), peak calling and filtering parameters, and independent pipeline implementation. Importantly, none of these differences alter the central conclusions of the study. Together, these results demonstrate a successful and biologically faithful reproduction of the original findings while highlighting the expected variability introduced by updated genome annotations and computational workflows.

Extra: Correlation plot



Comment:

Spearman correlation analysis of ATAC-seq signal across all samples showed high concordance between biological replicates and clear clustering by cell type. Replicates within each population (cDC1 and cDC2) exhibited strong correlations ($\rho = 0.93\text{--}1.00$ for cDC1 and $\rho = 0.86\text{--}0.90$ for cDC2), indicating high technical and biological reproducibility. In contrast, correlations between cDC1 and cDC2 samples were consistently lower, reflecting distinct chromatin accessibility landscapes between the two dendritic cell subsets, suggesting that HDAC1 loss does not globally alter chromatin accessibility.

```
sessioninfo::session_info()
```

```
## - Session info -----
##   setting  value
##   version R version 4.4.3 (2025-02-28)
##   os        AlmaLinux 8.10 (Cerulean Leopard)
##   system   x86_64, linux-gnu
##   ui        X11
##   language (EN)
##   collate  en_US.UTF-8
##   ctype    en_US.UTF-8
##   tz       America/New_York
##   date     2025-12-15
##   pandoc   3.2 @ /usr/local/ood/rstudio-server-2024.12.0+467/bin/quarto/bin/tools/x86_64/ (via rmarkd)
##   quarto    1.5.57 @ /usr/local/ood/rstudio-server-2024.12.0+467/bin/quarto/bin/quarto
##
## - Packages -----
##   package * version date (UTC) lib source
##   cli      3.6.5   2025-04-23 [1] CRAN (R 4.4.0)
##   colorspace 2.1-2   2025-09-22 [1] CRAN (R 4.4.0)
##   curl      7.0.0   2025-08-19 [1] CRAN (R 4.4.0)
##   digest    0.6.37  2024-08-19 [1] CRAN (R 4.4.0)
##   dplyr     * 1.1.4   2023-11-17 [2] CRAN (R 4.4.3)
##   enrichR    * 3.4    2025-02-02 [2] CRAN (R 4.4.3)
##   evaluate   1.0.5   2025-08-27 [1] CRAN (R 4.4.0)
##   farver     2.1.2   2024-05-13 [1] CRAN (R 4.4.0)
```

```

## fastmap      1.2.0   2024-05-15 [1] CRAN (R 4.4.0)
## forcats      * 1.0.0   2023-01-29 [2] CRAN (R 4.4.3)
## generics     0.1.4   2025-05-09 [1] CRAN (R 4.4.0)
## ggplot2      * 3.5.1   2024-04-23 [2] CRAN (R 4.4.3)
## glue          1.8.0   2024-09-30 [1] CRAN (R 4.4.0)
## gtable        0.3.6   2024-10-25 [2] CRAN (R 4.4.3)
## hms           1.1.3   2023-03-21 [2] CRAN (R 4.4.3)
## htmltools     0.5.8.1  2024-04-04 [2] CRAN (R 4.4.3)
## httr          1.4.7   2023-08-15 [2] CRAN (R 4.4.3)
## knitr          * 1.49    2024-11-08 [2] CRAN (R 4.4.3)
## labeling       0.4.3   2023-08-29 [2] CRAN (R 4.4.3)
## lifecycle     1.0.4   2023-11-07 [2] CRAN (R 4.4.3)
## lubridate     * 1.9.4   2024-12-08 [2] CRAN (R 4.4.3)
## magick          * 2.8.5   2024-09-20 [2] CRAN (R 4.4.3)
## magrittr       2.0.4   2025-09-12 [1] CRAN (R 4.4.0)
## munsell        0.5.1   2024-04-01 [2] CRAN (R 4.4.3)
## pillar          1.10.1  2025-01-07 [2] CRAN (R 4.4.3)
## pkgconfig      2.0.3   2019-09-22 [2] CRAN (R 4.4.3)
## purrr          * 1.0.4   2025-02-05 [2] CRAN (R 4.4.3)
## R6              2.6.1   2025-02-15 [1] CRAN (R 4.4.0)
## Rcpp            1.1.0   2025-07-02 [1] CRAN (R 4.4.0)
## readr          * 2.1.5   2024-01-10 [2] CRAN (R 4.4.3)
## rjson           0.2.23  2024-09-16 [1] CRAN (R 4.4.0)
## rlang            1.1.6   2025-04-11 [1] CRAN (R 4.4.0)
## rmarkdown        2.29    2024-11-04 [2] CRAN (R 4.4.3)
## rstudioapi      0.17.1  2024-10-22 [1] CRAN (R 4.4.0)
## scales           1.3.0   2023-11-28 [2] CRAN (R 4.4.3)
## sessioninfo     1.2.3   2025-02-05 [2] CRAN (R 4.4.3)
## stringi          1.8.7   2025-03-27 [1] CRAN (R 4.4.0)
## stringr          * 1.5.1   2023-11-14 [2] CRAN (R 4.4.3)
## tibble          * 3.2.1   2023-03-20 [2] CRAN (R 4.4.3)
## tidyverse        * 2.0.0   2023-02-22 [1] CRAN (R 4.4.0)
## timechange       0.3.0   2024-01-18 [2] CRAN (R 4.4.3)
## tinytex          0.56    2025-02-26 [2] CRAN (R 4.4.3)
## tzdb             0.4.0   2023-05-12 [2] CRAN (R 4.4.3)
## vctrs            0.6.5   2023-12-01 [2] CRAN (R 4.4.3)
## withr            3.0.2   2024-10-28 [1] CRAN (R 4.4.0)
## WriteXLS         6.8.0   2025-05-22 [1] CRAN (R 4.4.0)
## xfun             0.53    2025-08-19 [1] CRAN (R 4.4.0)
## yaml            2.3.10  2024-07-26 [1] CRAN (R 4.4.0)
##
## [1] /usr4/bf527/mgdouq/R/x86_64-pc-linux-gnu-library/4.4
## [2] /share/pkg.8/r/4.4.3/install/lib64/R/library
## * -- Packages attached to the search path.
##
## -----

```