Analysis of RNA-seq and ChIP-seq data from murine $CD4^+$ T cells during primary infection with Influenza A shows transcriptomic and H3K27ac differences compared to naive $CD4^+$ T cells

Sahi Hajirawala, Richard Kunze, Diana Lin, Emily Wong, Jessica Zhang

Abstract

Interleukin (IL)-10 production in $CD4^+$ T cells is a tightly-regulated process that plays a critical role in maintaining a balance between pathogen clearance and immunopathology in virally-infected hosts. IL-10 is a cytokine responsible restraining the immune response, which is necessary to prevent chronic infection and thus damage to host tissues and organs. It is known that interleukin (IL)-27 receptor signaling is needed for optimal expression of IL-10 during primary infections, but IL-10 is still expressed by reactivated T cells that have lost IL-27R after primary activation. To determine why IL-27R is no longer needed for IL-10 expression, we analyzed the transcriptome and epigenetic changes of murine $CD4^+$ T cell genomes with and without influenza infection. This report presents RNA-seq and ChIP-seq data that shows differentially expressed genes and pathways that may be involved in the expression of IL-10 in the absence of IL-27R. These two different datasets were integrated and analyzed through functional enrichment analysis using CistromeGO. We discovered that several genes and associated pathways are indeed differentially expressed, including many genes involved in the JAK-STAT pathway. Our results indicate possible epigenetic changes and an alternative pathways are responsible for the expression of IL-10 in reinfected IL-27-absent $CD4^+$ T cells.

Introduction

While the innate and adaptive immune systems play essential roles in host response to infection and pathogen clearance, their activity must be limited to prevent excessive inflammation and injury. Inflammation is an immune response induced by a variety of cytokines and chemokines. Anti-inflammatory cytokines are simultaneously induced by the proinflammatory response to prevent severe or chronic inflammation. When such counter-regulations fail, life threatening injury may result, such as seen in the 1918 influenza pandemic (1).

Interleukin (IL)-10 is among the major anti-inflammatory cytokines that negatively regulate the innate and adaptive immune response (1). IL-10 has been found to play a role in the clearance of viral infections in the lungs. IL-10 is expressed in infected tissue at the height of the T cell response by adaptive immune $CD8^+$ and $CD4^+$ effector T cells. CD8+ cytotoxic T lymphocytes (CTLs) require IL-27 and IL-2 derived from Helper T cells in order to produce IL-10 (2). During

infection, naive $CD4^+$ T cells are activated and produce IL-10 with assistance from IL-27 via the IL-27 receptor. When the $CD4^+$ T cells transition into resting memory cells, they lose their IL-27 receptor. However, when reactivated they still possess the ability to produce IL-10 without the need of IL-27 receptors. There is likely an epigenetic state change that occurs after primary infection which allow $CD4^+$ cells to continue producing IL-10 upon reinfection despite the lack of induction by IL-27.

To determine if any epigenetic state changes occur in $CD4^+$ T cells that would explain the production of IL-10 upon re-infection, ChIP-seq and RNA-seq data were taken from mice infected with Influenza A virus. C57Bl/6 mice were infected with Influenza A. Naive CD44lo $CD4^+$ T cells were taken from uninfected mice. 10 days after infection, primary T cells (Flu NP+ CD44hi $CD4^+$) were isolated. Memory T cells, defined as NP+ CD44hi $CD4^+$, were taken 30-35 days after infection and second degree effectors were taken 6 days after re-infection for transcriptomic analysis by RNA-seq. The low input Native ChIP-seq analysis produced data pertaining to H3K27ac, which is an epigenetic modification that indicates the acetylation of the 27th lysine residue of histone H3. The modification is associated with higher activation of transcription and is often found at distal and proximal regions of the transcription start site.

Using both the ChIP Seq and RNA Seq datasets, we determined which genes are upregulated or downregulated between the naive and primary effector cells. By comparing the transcriptional and epigenetic differences between the two data sets, we anticipate to determine which genes and pathways play a role in IL-10 production upon re-infection without the need for IL-27.

The initial quality of the RNA Seq and ChIP Seq reads was analyzed using FastQC v0.11.7 (https://github.com/s-andrews/FastQC). For analyzing the RNA-seq data set, we used STAR v2.5.3 (3) to map the reads to a genome template. Post-alignment quality control was performed using QoRTs v1.3.6 (4), count reads using HTSeq v0.11.1 (5), and differential expression analysis using DESeq2 v1.24 (6). ChIP-seq datasets were aligned using BWA v0.7.17 (7), and MACS2 v2.1.4 (8) was used to call enriched regions. RNA-seq and ChIP-seq data analysis was integrated using Cistrome-GO v1.0.0 (9). Additionally, samtools v1.9 (10) and sambamba v0.6.6 (11) were used to manipulate SAM/BAM files and IGV v2.7.0 (12) was used for visualization. This workflow is summarized in *Figure 1*, and the tools (including GitHub repository links) used in this workflow are presented in *Table 1*.

Materials and Methods

All scripts used can be found in the *scripts/* directory. An overview of our directory structure is provided in *Figure 2.*

The H3K27ac and RNA-Seq datasets used in subsequent analyses were generated by Dr. Georgia Perona-Wright (University of Glasgow) and Nicolette Fonseca (University of British Columbia).

Scripts were run on the provided ORCA (13) server. *Optional* parameter selection is explained, but *required* parameters (such as input/output filename specification) are *excluded*).

FastQC

To assess the quality of our sequencing runs, we ran FastQC on both the ChIP-seq and RNA-seq reads. FastQC was run using the *run_fastqc.sh* script.

RNA-seq

The RNA-seq data are grouped by three different conditions based on the T cell type: Naive, Primary, and Memory. In each condition, there are three replicates, each of which are Illumina paired-end, first (or reverse) strand. RNA-seq raw reads were aligned to the Mouse mm10 reference genome using STAR (*--runThreadN 8 --outSAMtype BAM SortedByCoordinate --outSAMunmapped Within --outSAMattributes Standard*). Eight threads were chosen to increase the efficiency of alignments without overloading the server. The output was sorted by coordinate for visualization and qualitative analysis in IGV. Following STAR, alignments were visually inspected by IGV to determine strandedness for downstream analysis. QoRTs was used to assess the quality of the alignments and the raw reads against the reference and annotation (*--generatePlots --stranded --genomeFA "${star_index_file}mm10_genome.fa" --rawfastq $rawread1,$rawread2*). STAR alignments were first sorted by name using samtools, and piped into HTSeq with the Mouse mm10 annotation GTF to generate count files (*-s reverse -r name -f bam -o ${cellType}_${i}.HT_out.sam*). These outputs were then analyzed with DESeq2 in R, using the Naive cell type as the control, to determine significantly expressed genes, or with an adjusted p-value of <0.05. The adjusted p-value threshold was chosen from industry standard. Since our focus is to determine why IL-10 expression is continued despite the absence of IL-27 following T cell activation, we selected Naive as control in order to determine the significant changes in gene expression following differentiation of T cells from their Naive state. Our adjusted p-value was selected to allow for enough genes for further integration analysis with ChIP-seq, as well as to determine statistically significant changes in gene expression. From these, we extracted the top 10 upregulated genes as well as the top 10 downregulated genes.

All RNA-seq alignments can be found in the *STAR_output/* directory, and are organized in subfolders by the cell type and the library/replicate. Within each subfolder, a STAR alignment BAM, sorted by coordinate, is present in addition to a HTseq count (*htseq.count*) file containing the gene counts. Each subfolder also includes a QC directory containing the results from QoRTs. STAR was run using the *run_STAR.sh* script and QoRTs was run using the *run_QoRTs.sh* script. To associate the gene names with gene IDs, we used the script *scripts/get-gene-names.sh* and the provided mm10 GTF file. Volcano plots were generated using *volcano.R* in the *scripts/* directory.

ChIP-seq

To align the ChIP-seq reads to the mouse mm10 reference genome, we ran BWA (*bwa mem -t 8*) to produce a SAM file. Eight threads were chosen (whenever threads could be specified) as to speed up the process but not to overload the ORCA server. The SAM file was then converted (*sambamba view -t 8 -S -f bam*) into a BAM file using sambamba, using *-t* to specify 8 threads, *-S* to specify a SAM file input, and *-f bam* to specify a BAM file output. Sambamba was also used to sort (*sambamba sort -t 8)* the BAM file and mark the duplicates (*sambamba markdup -t 8)*.Then we used MACS2 (*macs2 callpeak -f BAMPE -g mm -B -q 0.01*) to call the peaks, using *-f BAMPE* to specify paired-end BAM file, *-g mm* to specify the mm10 reference genome, *-B* to save a bedgraph of extended fragment pileups and local lambda tracks. Then MACS2 is used (*macs2 bdgdiff*) to find enriched/differential binding sites between primary and naive T cells (stored in *macs2/bdgdiff/*). The bedgraph (.bdg) files were converted to tiled data (.tdf) files, using igvtools v2.7.0 (12) *(igvtools totdf ${file}.bdg ${file}.tdf mm10)* before visualizing in IGV, stored in *macs2/tdfs/*.

ChIP-seq alignments were done by running the *run-bwa.sh* script. All ChIP-seq read alignments can be found in the *bwa_alignments/* directory. The *bwa_alignments/intermediates/* directory holds intermediate files for *bwa mem* and *sambamba* (as specified in the script to do so), but will be removed to reduce our digital footprint.

MACS2 was run using the *run-macs.sh* and *post-macs2.sh* scripts. All MACS2 output files can be found in the *macs/* directory. *The macs2/* directory is further broken down by condition: *Memory/ Naive/ Primary/*. Log files are stored in *macs2/logs/.* Bed files produced by *macs2 bdgdiff* can be found in the *macs2/bdgdiff/* directory.

Finally, both the MACS2 and DESeq2 output were fed into Cistrome-GO for functional enrichment analysis and ChIP-seq/RNA-seq data analysis integration, where we investigated the JAK/STAT pathway and others. Using the KEGG (14) defined JAK/STAT pathway (found in *CistromeGO/combined/JAK_STAT.full.txt*), we were able to extract the proteins that are also found in the mm10 annotation into *CistromeGO/combined/JAK_STAT.processed.mm10.txt*), using *scripts/get-mm10-JAK_STAT.sh.* To get the top 10 genes (by rank product) in our Cistrome-GO results, we used the script *scripts/get-top10-CisGO.sh* for our analysis.

Results and Discussion

Prior to analysis, quality control was done on raw ChIP-Seq and RNA-Seq reads using FastQC. Alignment quality was checked with QoRTs.

Determining Sequencing Quality with FASTQC

To assess the quality of our sequencing runs, we ran FASTQC on both the ChIP-seq and RNA-seq reads.

In the RNA-seq dataset (summarized in *Tables 2-4*), all primary T cell replicates (*Table 2*) had poor per tile sequence quality, per base sequence content, per sequence GC content, and excess overrepresented sequences. In naive T cell replicates (*Table 3*), the reads have poor per tile sequence quality, per base sequence content, per sequence GC content, and sequence length distribution, along with excess overrepresented sequences. In the memory T cell replicates (*Table 4*), reads have poor per base sequence content and per sequence GC content, and an excess of overrepresented sequences and adapter content.

In RNA sequencing of $CD4^+$ T cells from all three stages of infection, per base sequence content, per sequence GC content, and overrepresented sequence parameters had majority "Fail" or "Warn" flags from FastQC output. However, considering that the thresholds for these flags are based on whole genome shotgun DNA sequencing, these flags are not necessarily a marker of poor quality sequencing. Preparing a RNA seq library can skew base distribution, leading to poor per base sequence scores. RNA transcripts (as opposed to whole genomes) can have a greater- or lesser-than-normal distribution of mean GC content, leading to poor per base sequence content. Transcripts could register as overrepresented sequences if they are expressed abundantly (https://rtsf.natsci.msu.edu/genomics/tech-notes/fastqc-tutorial-and-faq/).

In the ChIP-seq dataset (summarized in *Tables 5-7*), the memory T cell reads (*Table 5*), both treatment and control, showed poor per base sequence content, per sequence GC content, sequence length distribution and overrepresented sequences. The primary T cell reads (*Table 6*) (control and treatment), showed poor per base sequence content, per sequence GC content and adapter content. The naive T cell reads (*Table 7*) (treatment and control), showed poor per base sequence content, per sequence GC content, sequence duplication levels, overrepresented sequences and adapter content.

For ChIP-seq, data from all three stages of infection showed poor per base sequence content and per sequence GC content. Poor per base sequence content can be explained by the enzymes used to fragment the DNA; they will always cut at the same sequences and can thus introduce bias to the start of the fragments. Per sequence GC content can be poor for the same reason as in RNA-seq; the whole genome was not sequenced. ChIP-seq enriches for sequences that bind specific proteins, in this case H3K27ac, so the overrepresented sequences seen for Memory and Naive are not unexpected either. (https://hbctraining.github.io/Intro-to-ChIPseq/lessons/02_QC_FASTQC.html)

From overall FastQC results, the quality of the data appears to be sufficient once considering that the data is RNA-seq and ChIP-seq, not whole genome. Parameters that have mostly "Fail"

flags can be mostly explained by the sequencing method used. The data can be used for further analysis, though quality should be kept in mind through analysis and result interpretation.

Checking alignment quality with QoRTs

QoRTs was run on each cell type (Naive, Primary, Memory) and each of the three replicates for RNA-seq data. We focused on the GC content, gene-body coverage, strandedness, and insert size as assessments of our data and alignments. While all samples had acceptable mean GC content (between 40-60%), replicates in the Naive and Primary cell types all show a small sharp second peak higher than the first (*Figure 3*). This suggests either contamination or an overrepresentation of certain sequences. Though all samples exhibit the curve that is expected in gene-body coverage assessment, we see that samples in the Naive cell type showed higher coverage of the 3' end, while Primary showed higher coverage in the middle, and Memory showed higher coverage on both the 5' and 3' ends (*Figure 4*). This suggests sequencing or alignment bias in certain regions of the genes, but also may suggest where expression might be more greatly found. Both Primary and Memory samples demonstrate the expected strandedness from Illumina Stranded Paired-End RNA-Seq protocol, but it is clear that the strandedness was lost in all Naive samples (*Figure 5*). Lastly, in all samples, we see that the insert sizes are between 150-200bp which is standard for RNA-seq data (*Figure 6*). However, because mean insert sizes seem to lean more towards 200bp, there is potential for loss of coverage in smaller exons.

DESeq2 Analysis

The top 10 upregulated genes in Primary vs Naive $CD4^+$ T cells are shown in Table 8 (from most upregulated to least). The top 10 upregulated genes are *Cxcr6, Lilr4b, Fgl2, Gm37352, Ccr5, Myo1f, Farp1, Hist1h2bb, Pdnl1,* and *Rgs16*. These top 10 upregulated genes can be seen in the volcano plot in *Figure 7.*

The top 10 downregulated genes in Primary vs Naive are shown in Table 9 (from most downregulated to least). The top 10 downregulated genes are *CT010467.1, Mfsd4a, Gm34006, Gm11175, Gm37401, Gm20721, Gm15441, Gm26648, Ccdc152,* and *Bach2os*. Note that genes starting with "Gm" are predicted genes. These top 10 downregulated genes can be seen in the volcano plot in *Figure 7.*

The top 10 differentially expressed (regardless of up or down-regulated) genes in Primary vs Naive are shown in Table 10 (from most differentially expressed to least). The top 10 most differentially expressed genes are *CT010467.1, Mfsd4a, Gm34006, Gm11175, Gm37401, Gm20721, Gm15441, Gm26648, Ccdc152,* and *Cxcr6*. In this set of differentially expressed genes, only *Cxcr6* is an upregulated gene, whereas the rest are all downregulated. Among these top 10 most differentially expressed genes, only 1 is upregulated. Upon further analysis, it is seen that of the 8866 differentially expressed genes *of significance*, 55.4% (4914/8866) of the genes are upregulated, a slight majority over downregulated genes.

Cistrome-GO

Our combined RNA-seq/ChIP-seq Cistrome-GO results can be viewed online at http://go.cistrome.org/result?uid=1572837071*, o*r in our *CistromeGO/combined/* directory.

Our ChIP-seq only Cistrome-GO results can be viewed online at http://go.cistrome.org/result?uid=1573588661, or in our *CistromeGO/ChIP-seq-only/* directory.

Genes with highest regulatory potential based on H3K27ac

Processed results from ChIP-seq were first loaded into Cistrome-GO in solo mode without RNA-seq data to view genes that were differentially marked. The 'regulatory potential' parameter, calculated using the number of binding sites and their distance from a gene, in CistromeGO was used to generate the top 10 genes with the most potential to be affected by H3K27 acetylation (*Table 15*).

Some of these top 10 genes (*Table 15*) have been previously associated with influenza models. *Angptl2* is proinflammatory and associated with higher lung damage during influenza in a murine model (15). It is also upregulated by a STAT-3 mediated mechanism, part of the JAK-STAT pathway (15). *Ubash3A* and *Trib1* both have been found to decrease T cell activity and signalling, the latter in an antiviral context (16, 17).

Enriched Kegg Pathways

In ensemble mode using both RNA-seq and ChIP-seq results, the top 10 most enriched KEGG pathways (ordered by increasing FDR) is shown in *Table 11.* The top 5 most enriched KEGG pathways are: the T cell receptor signaling pathway, Th17 differentiation, viral carcinogenesis, endocytosis, and Th1 and Th2 cell differentiation. The fact that these T cell pathways are enriched are expected, considering the Primary vs Naive T cell condition. The viral carcinogenesis and endocytosis pathway enrichments are also expected, as the T cell is exposed to influenza in our treatments.

Enriched GO Molecular Functions

Based on our combined RNA-seq/ChIP-seq results, the top 10 most enriched GO Molecular Functions (ordered by increasing FDR) is shown in *Table 12.* The top 5 most enriched GO Molecular Functions are: kinase binding, transcription regulator activity, protein domain specific binding, RNA binding, and double-stranded DNA binding. These pathways are also expected, as one specific pathway of interest is the JAK/STAT signalling pathway, which involves many interactions between kinases. Additionally, the transcriptional regulation activity supports our hypothesis of epigenetic modifications.

GO Biological Processes

Based on our combined RNA-seq/ChIP-seq results, the top 10 most enriched GO Biological Processes (ordered by increasing FDR) is shown in *Table 13.* The top 5 most enriched GO Biological Processes are: regulation of immune system process, immune system process, positive regulation of gene expression, negative regulation of gene expression and regulation of transcription by RNA polymerase II. All of these pathways are expected for our primary vs naive T cell conditions with exposure to influenza.

Gene Products Ranked by Rank Product

Based on our combined RNA-seq/ChIP-seq results, the top 10 differentially expressed genes ranked by Rank Product are shown in *Table 14*. Of these top 10 genes, two of them are part of the JAK/STAT pathway: *Jak1* and *Socs3*. Activation of IL-27 receptor results in the subsequent activation of the JAK-STAT pathway, so it follows that genes involved in this pathway would be differentially expressed between naive and infected T cells. Several of these genes, as well as the top 10 differentially expressed, also show significant qualitative change in expression as seen in *Figures 10-20*.

These 10 genes have been labelled on a volcano plot in *Figure 8*. Note that only 3 genes are labelled on this volcano plot, as 5 other genes (including *Jak1* and *Socs3*) do not show up on the plot as their p-values and FDR values were 0, and the y-axis of the volcano plot is -log(p-value), where its y-value would be undefined. Additionally, all proteins of the JAK/STAT pathway (as defined by its KEGG pathway), have been labelled green in *Figure 9.* In most cases, these JAK/STAT proteins are upregulated or downregulated, and not insignificant.

Furthermore, a connection between the upregulated chemokine receptors *Cxcr6* and *Ccr5* and IL-10 production through the chemokine signaling pathway can be seen by the chemokine mediated activation of the JAK/STAT pathway. This involvement is supported by the absence of acetylation of *Cxcr6* and *Ccr5* in naive T cells, the presence of acetylation in primary T cells, and the continued acetylation in memory T cells, as shown in *Figure 21* and *Figure 22*. This suggests that the pathway is activated following T cell differentiation and remains active as a potential contributor to IL-10 production. Additionally, it has been shown that MAPK/Erk pathways in B cells result in IL-10 production (18). We note that in the chemokine receptors *Cxcr6* and *Ccr5* are also involved in activating the MAPK/Erk pathway, as seen in *Figure 23*, which may also result in IL-10 production in infected T cells post differentiation.

Conclusion

Altogether, after integrating our RNA-seq and CHiP-seq datasets, we posit that there is in fact some epigenetic change that allows for IL-10 expression in the absence of IL-27R. We hypothesize that the genes and proteins involved in the JAK-STAT pathway may be key players in this epigenetic change and provide an alternative pathway for IL-10 production. Specifically,

epigenetic changes to *CXCR6* and *CCR5* may play a role in this pathway, as we see increased acetylation and gene expression in primary $CD4^+$ T cells in comparison to naive $CD4^+$ T cells. Nonetheless, we note that our ability to identify pathways that cause IL-10 production is confounded by existing pathways that are caused by IL-10 signalling being expressed simultaneously. To investigate this further, future experiments and research of these JAK-STAT-related proteins and IL-10 in memory T cells could give insight on this phenomenon. Epigenetic changes in primary T cells (compared to naive) that persist in memory T cells would be of interest. Likewise, knockout experiments of specific JAK-STAT-related proteins and their impact on IL-10 production in naive/primary/re-infected T cells could help elucidate a specific mechanism.

References

1. Sun J, Madan R, Karp CL, Braciale TJ. 2009. Effector T cells control lung inflammation during acute influenza virus infection by producing IL-10. Nat Med 15:277–284.

2. Sun J, Dodd H, Moser EK, Sharma R, Braciale TJ. 2011. CD4+ T cell help and innate-derived IL-27 induce Blimp-1-dependent IL-10 production by antiviral CTLs. Nat Immunol 12:327–334.

3. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21.

4. Hartley SW, Mullikin JC. 2015. QoRTs: a comprehensive toolset for quality control and data processing of RNA-Seq experiments. BMC Bioinformatics 16:224.

5. Anders S, Pyl PT, Huber W. 2015. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 31:166–169.

6. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550.

7. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler

transform. Bioinformatics 25:1754–1760.

8. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. 2008. Model-based analysis of ChIP-Seq (MACS). Genome Biol 9:R137.

9. Li S, Wan C, Zheng R, Fan J, Dong X, Meyer CA, Liu XS. 2019. Cistrome-GO: a web server for functional enrichment analysis of transcription factor ChIP-seq peaks. Nucleic Acids Res 47:W206–W211.

10. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

11. Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. 2015. Sambamba: fast processing of NGS alignment formats. Bioinformatics 31:2032–2034.

12. Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 14:178–192.

13. Jackman SD, Mozgacheva T, Chen S, O'Huiginn B, Bailey L, Birol I, Jones SJM. 2019. ORCA: a comprehensive bioinformatics container environment for education and research. Bioinformatics 35:4448–4450.

14. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. 2017. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res 45:D353–D361.

15. Li L, Chong HC, Ng SY, Kwok KW, Teo Z, Tan EHP, Choo CC, Seet JE, Choi HW, Buist ML, Chow VTK, Tan NS. 2015. Angiopoietin-like 4 Increases Pulmonary Tissue Leakiness

and Damage during Influenza Pneumonia. Cell Rep 10:654–663.

16. Ge Y, Paisie TK, Newman JRB, McIntyre LM, Concannon P. 2017. UBASH3A Mediates Risk for Type 1 Diabetes Through Inhibition of T-Cell Receptor–Induced NF-κB Signaling. Diabetes.

17. Rome KS, Pear WS. 2019. Trib1 controls antiviral immunity by restraining CD4 and CD8 T cell effector responses during chronic infection. J Immunol 202.

18. Garaud S, Taher TE, Debant M, Burgos M, Melayah S, Berthou C, Parikh K, Pers J-O, Luque-Paz D, Chiocchia G, Peppelenbosch M, Isenberg DA, Youinou P, Mignen O, Renaudineau Y, Mageed RA. 2018. CD5 expression promotes IL-10 production through activation of the MAPK/Erk pathway and upregulation of TRPC1 channels in B lymphocytes. Cellular & Molecular Immunology.

Appendix (Figures and Tables)

***Figure 1**: Bioinformatics Workflow*

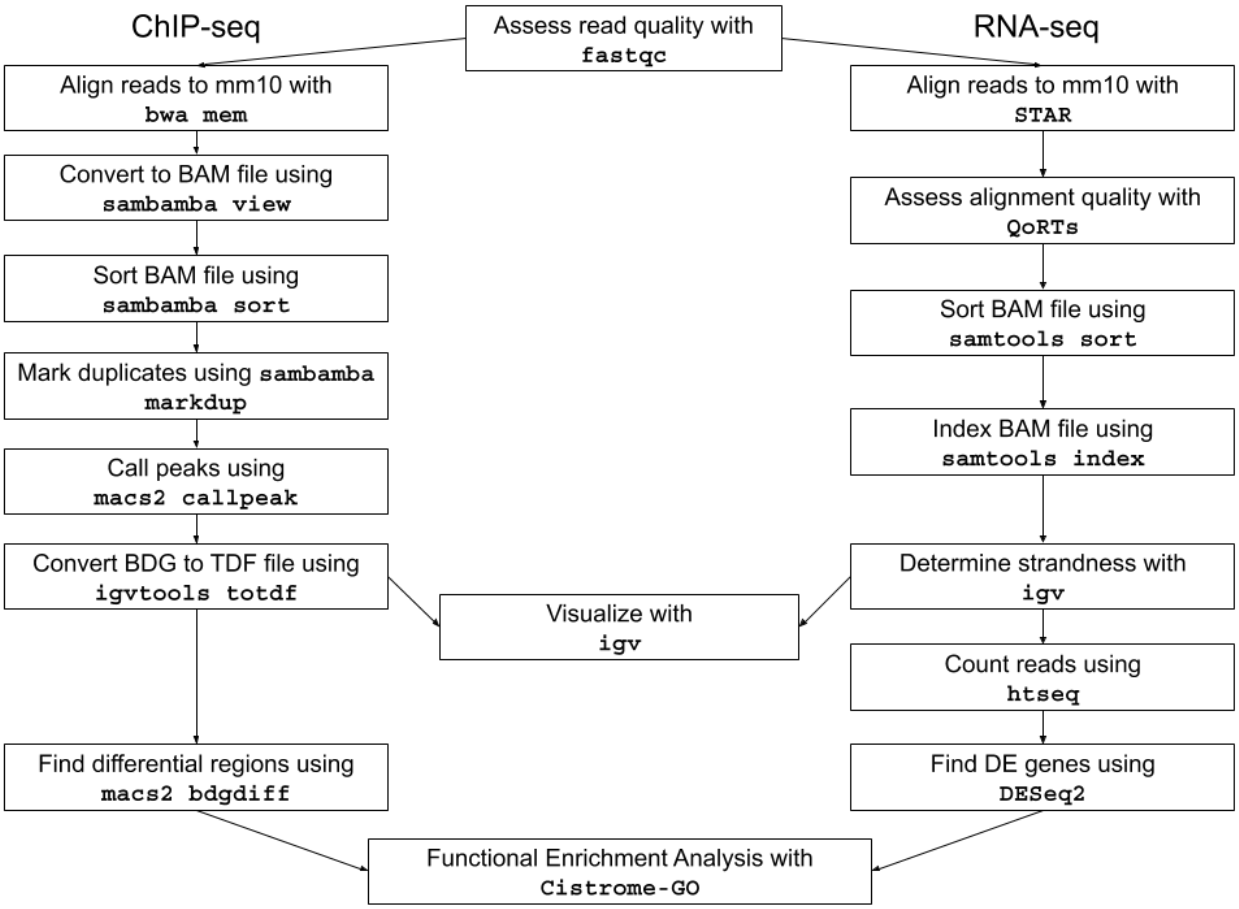


***Table 1:** Bioinformatics Tools*

Tool	GitHub Repository	Publication
FastQC	https://github.com/s-andrews/FastQC	—
STAR	https://github.com/alexdobin/STAR	(3)
QoRTs	https://github.com/hartleys/QoRTs	(4)
HTSeq	https://github.com/simon-anders/htseq	(5)
DESeq2	https://github.com/mikelove/DESeq2	(6)

BWA	https://github.com/lh3/bwa	(7)
MACS	https://github.com/taoliu/MACS	(8)
CistromeGO	https://github.com/changxinw/Cistrome-GO	(9)
samtools	https://github.com/samtools/samtools	(10)
sambamba	https://github.com/biod/sambamba	(11)
IGV	https://github.com/igvteam/igv	(12)
ORCA	https://github.com/bcgsc/orca	(13)

Figure 2: *Directory Structure*

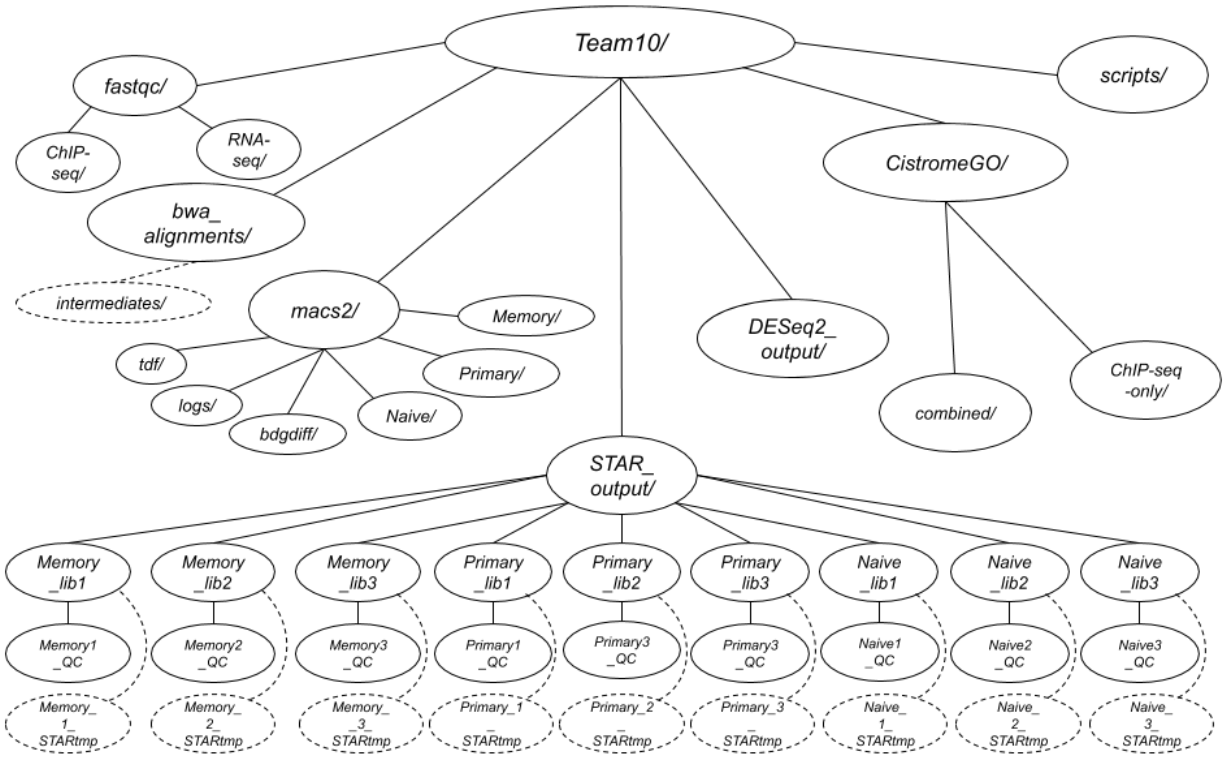


Table 2*: FastQC of Primary CD4⁺ T cell RNA-seq reads. ✔ = Passed; ⚠ = Warn; ✘ = Fail.*

	Primary CD4⁺ T cells	
	First Reads	**Second Reads**

Quality Control Metric	Rep 1	Rep 2	Rep 3	Rep 1	Rep 2	Rep 3
Basic Statistics	✔	✔	✔	✔	✔	✔
Per base sequence quality	✔	✔	✔	✔	✔	✔
Per tile sequence quality	⚠	⚠	⚠	⚠	✔	✔
Per sequence quality scores	✔	✔	✔	✔	✔	✔
Per base sequence content	✘	✘	✘	⚠	⚠	⚠
Per sequence GC content	⚠	⚠	⚠	✔	✔	✔
Per base N content	✔	✔	✔	✔	✔	✔
Sequence Length Distribution	✔	✔	✔	✔	✔	✔
Sequence Duplication Levels	✔	✔	✔	✔	✔	✔
Overrepresented sequences	✘	✘	✘	⚠	⚠	⚠
Adapter Content	✔	✔	✔	✔	✔	✔

Table 3: *FastQC of Naive CD4$^+$ T cell RNA-seq reads. ✔ = Passed; ⚠ = Warn; ✘ = Fail.*

	Naive CD4$^+$ T cells					
	First Reads			Second Reads		
Quality Control Metric	**Rep 1**	**Rep 2**	**Rep 3**	**Rep 1**	**Rep 2**	**Rep 3**
Basic Statistics	✔	✔	✔	✔	✔	✔
Per base sequence quality	✔	✔	✔	✔	✔	✔
Per tile sequence quality	⚠	⚠	⚠	⚠	⚠	⚠
Per sequence quality scores	✔	✔	✔	✔	✔	✔
Per base sequence content	⚠	⚠	⚠	⚠	⚠	⚠
Per sequence GC content	⚠	⚠	⚠	⚠	⚠	⚠

Per base N content	✔	✔	✔	✔	✔	✔
Sequence Length Distribution	⚠	⚠	⚠	⚠	⚠	⚠
Sequence Duplication Levels	✔	✔	✔	✔	✔	✔
Overrepresented sequences	✘	✘	✘	⚠	⚠	⚠
Adapter Content	✔	✔	✔	✔	✔	✔

Table 4: *FastQC of Memory CD4⁺ T cell RNA-seq reads. ✔ = Passed; ⚠ = Warn; ✘ = Fail.*

	Memory CD4⁺ T cells					
	First Reads			**Second Reads**		
Quality Control Metric	**Rep 1**	**Rep 2**	**Rep 3**	**Rep 1**	**Rep 2**	**Rep 3**
Basic Statistics	✔	✔	✔	✔	✔	✔
Per base sequence quality	✔	✔	✔	✔	✔	✔
Per tile sequence quality	—	—	—	—	—	—
Per sequence quality scores	✔	✔	✔	✔	✔	✔
Per base sequence content	✘	✘	✘	⚠	⚠	⚠
Per sequence GC content	⚠	⚠	⚠	✘	✘	✘
Per base N content	✔	✔	✔	✔	✔	✔
Sequence Length Distribution	✔	✔	✔	✔	✔	✔
Sequence Duplication Levels	✔	✔	✔	✔	✔	✔
Overrepresented sequences	✘	✘	✘	✘	✘	✘
Adapter Content	✘	✘	✘	⚠	⚠	⚠

Table 5: *FastQC of CD4⁺ Primary T cell ChIP-seq reads. ✔ = Passed; ⚠ = Warn; ✘ = Fail.*

	Primary CD4⁺ T cells

	Immunoprecipitation (H3K27ac)		Input (Background)	
Quality Control Metric	**First Read**	**Second Read**	**First Read**	**Second Read**
Basic Statistics	✔	✔	✔	✔
Per base sequence quality	✔	✔	✔	✔
Per tile sequence quality	✔	✔	✔	✔
Per sequence quality scores	✔	✔	✔	✔
Per base sequence content	✘	✘	✘	✘
Per sequence GC content	✘	✘	⚠	✘
Per base N content	✔	✔	✔	✔
Sequence Length Distribution	✔	✔	✔	✔
Sequence Duplication Levels	✔	✔	✔	✔
Overrepresented sequences	✔	✔	✔	✔
Adapter Content	✘	✘	✘	✘

Table 6: *FastQC of $CD4^+$ Naive T cell ChIP-seq reads. ✔ = Passed; ⚠ = Warn; ✘ = Fail.*

	Naive $CD4^+$ T cells			
	Immunoprecipitation (H3K27ac)		Input (Background)	
Quality Control Metric	**First Read**	**Second Read**	**First Read**	**Second Read**
Basic Statistics	✔	✔	✔	✔
Per base sequence quality	✔	✔	✔	✔
Per tile sequence quality	✔	✔	✔	✔
Per sequence quality scores	✔	✔	✔	✔

Per base sequence content	⚠	⚠	✘	✘
Per sequence GC content	✘	⚠	⚠	✘
Per base N content	✔	✔	✔	✔
Sequence Length Distribution	✔	✔	✔	✔
Sequence Duplication Levels	⚠	✔	✔	✔
Overrepresented sequences	✘	⚠	⚠	✔
Adapter Content	⚠	✔	✔	✔

Table 7: *FastQC of CD4$^+$ Memory T cell ChIP-seq reads. ✔ = Passed; ⚠ = Warn; ✘ = Fail.*

	Memory CD4$^+$ T cells			
	Immunoprecipitation (H3K27ac)		**Input (Background)**	
Quality Control Metric	**First Read**	**Second Read**	**First Read**	**Second Read**
Basic Statistics	✔	✔	✔	✔
Per base sequence quality	✔	✔	✔	✔
Per tile sequence quality	✔	✔	✔	✔
Per sequence quality scores	✔	✔	✔	✔
Per base sequence content	✘	✘	✘	✘
Per sequence GC content	⚠	⚠	⚠	⚠
Per base N content	✔	✔	✔	✔
Sequence Length Distribution	⚠	⚠	⚠	⚠
Sequence Duplication Levels	✔	✔	✔	✔
Overrepresented sequences	✘	⚠	✔	✔

Adapter Content	✔	✔	✔	✔

Figure 3: *QoRTs GC Content Plots for Naive, Memory and Primary CD4⁺ T cells.*

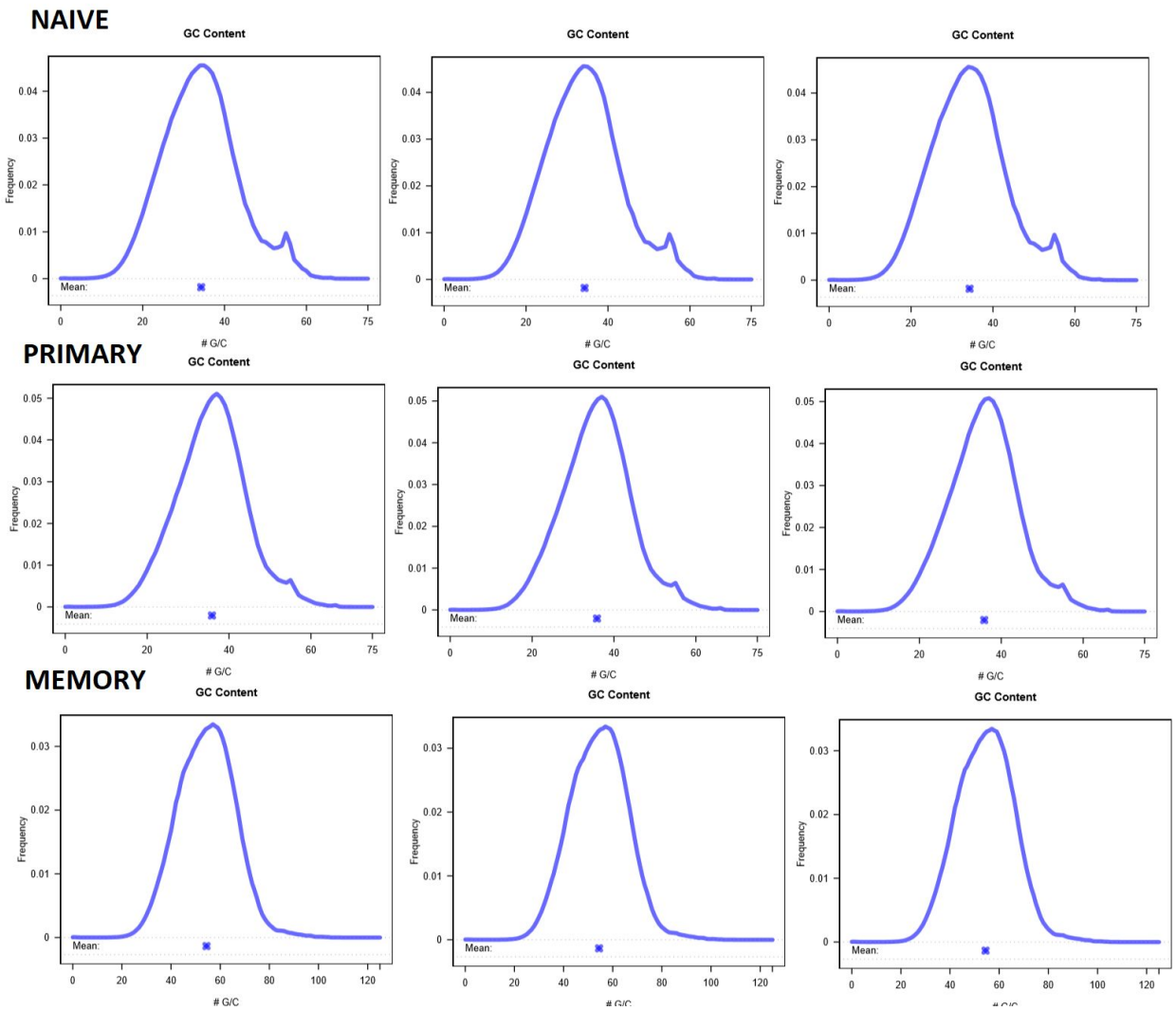


Figure 4: *QoRTs Gene-Body Coverage Plots for Naive, Memory and Primary CD4⁺ T cells.*

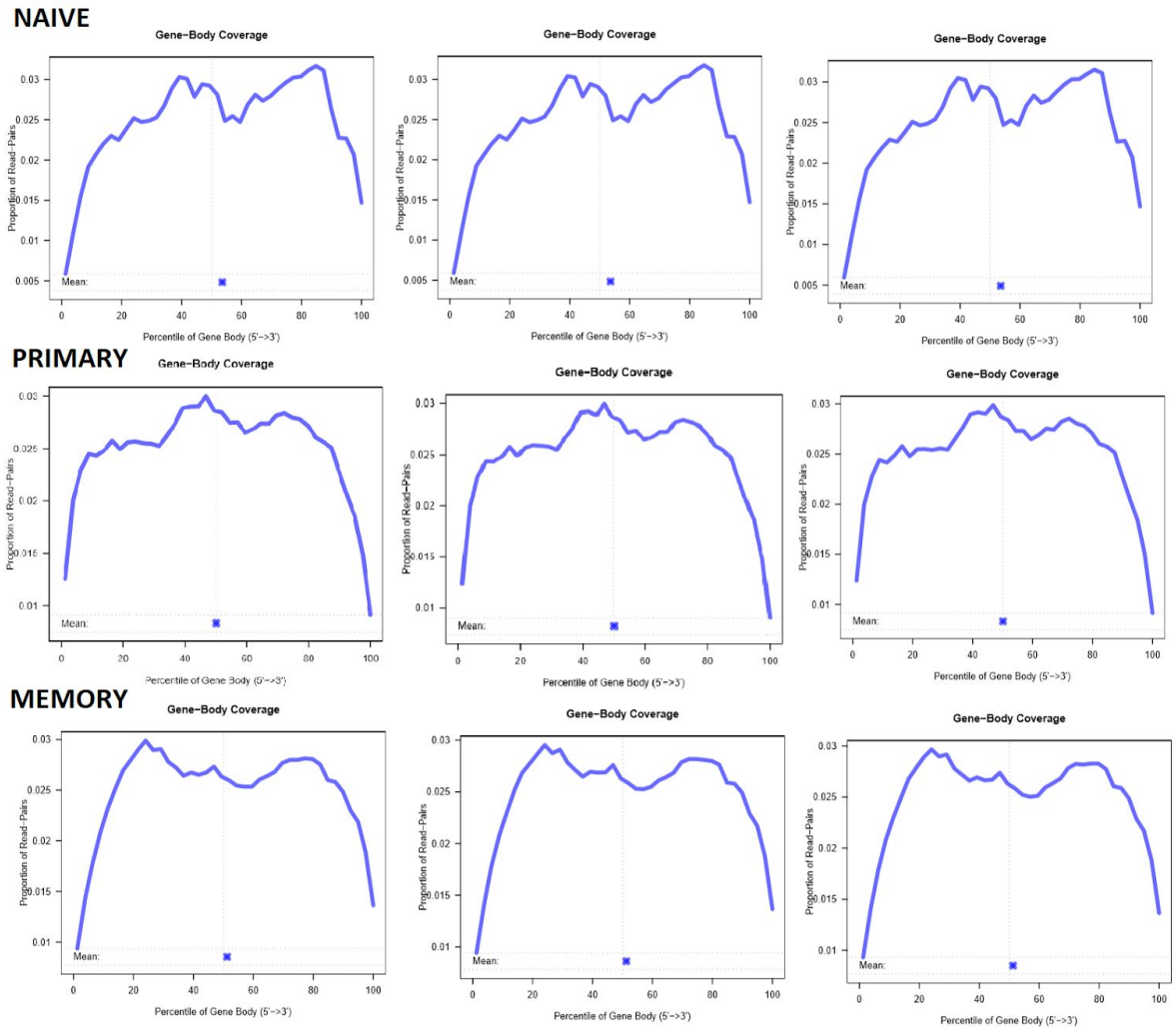


Figure 5: *QoRTs Strandedness Test Plots for Naive, Memory and Primary CD4⁺ T cells.*

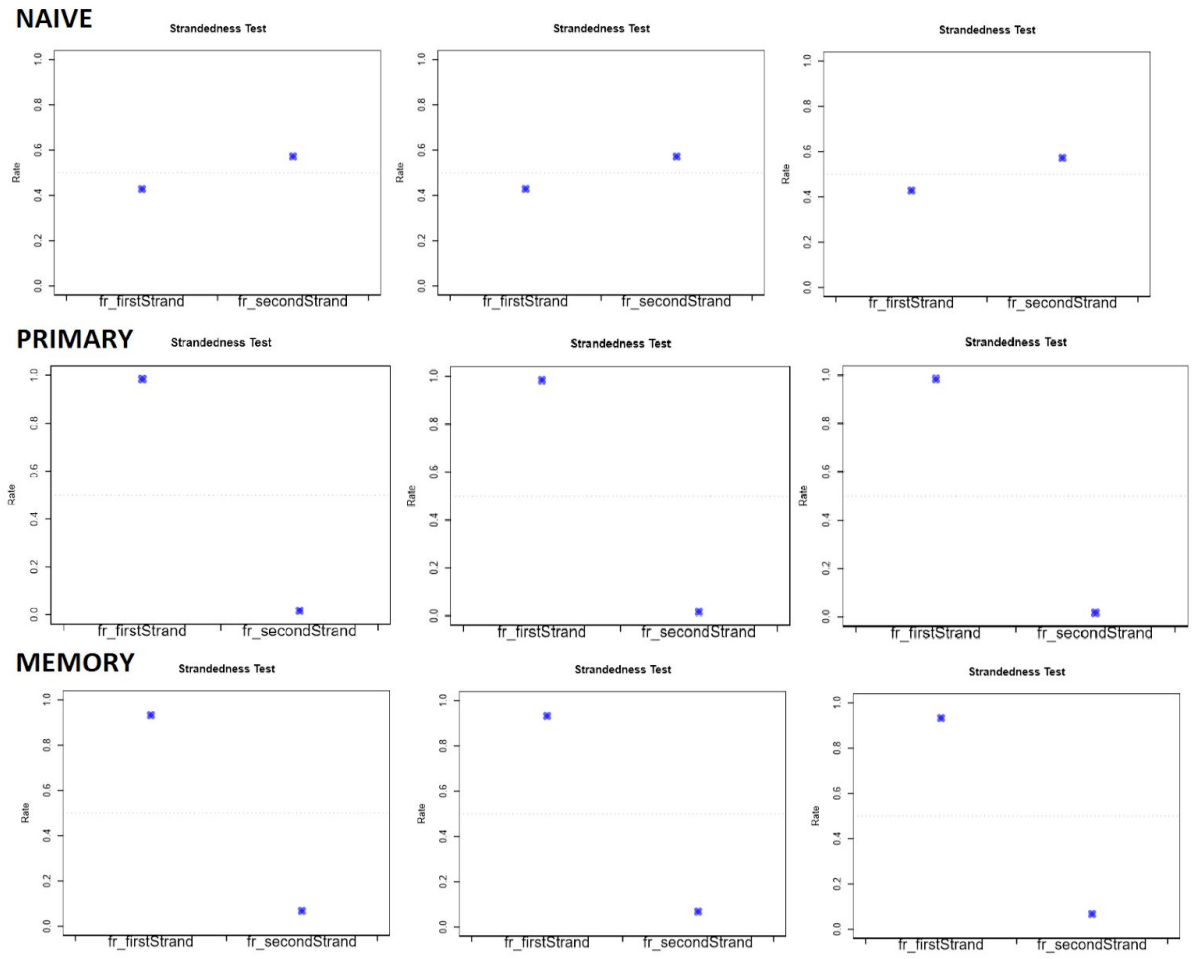


Figure 6: *QoRTs Insert Size Plots for Naive, Memory and Primary CD4⁺ T cells.*

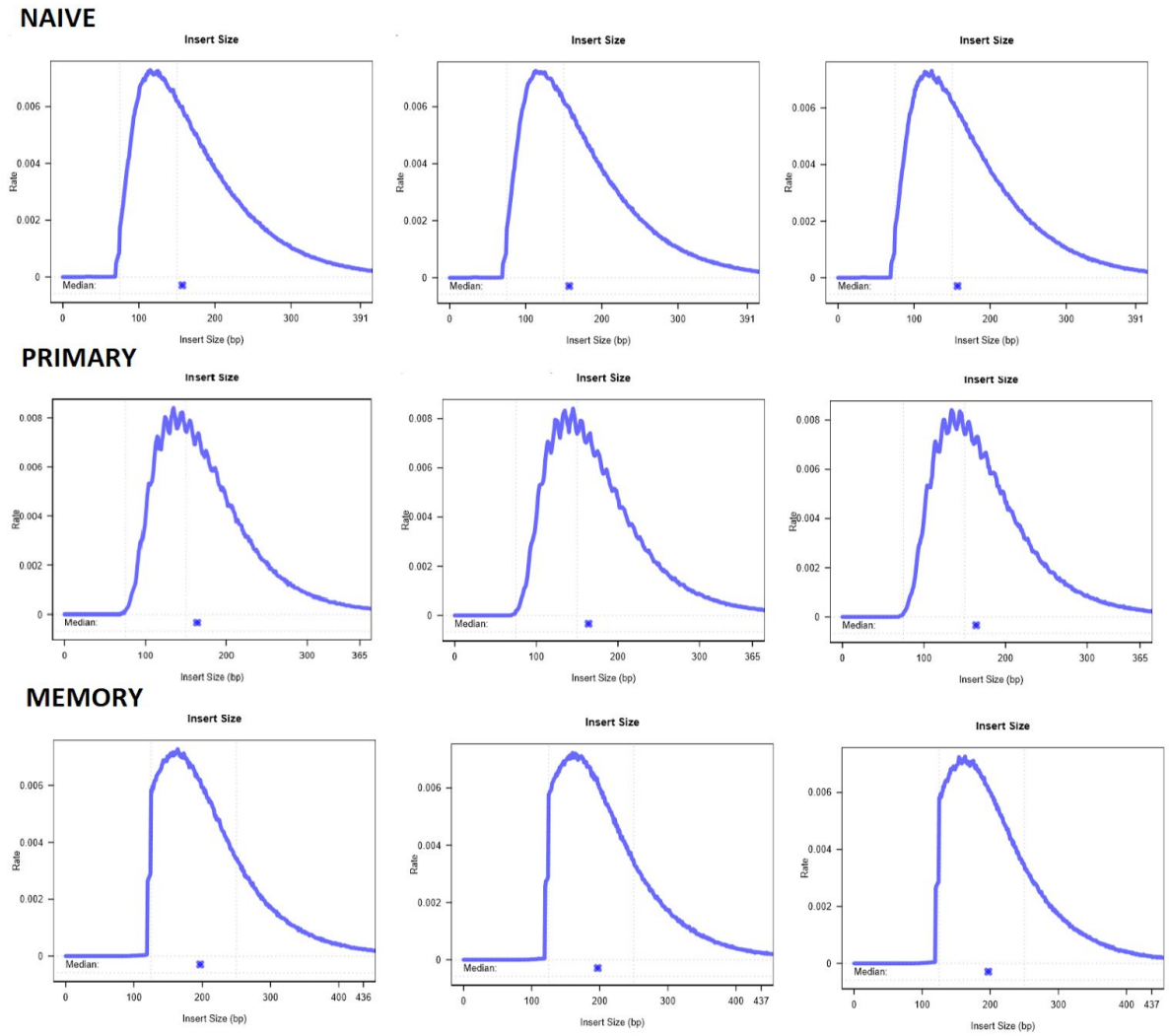


Table 8: *Most Upregulated Genes in Primary vs Naive T cells, from DESeq2 analysis.*

Gene	Fold Change	FDR	P-value	Gene Description
Cxcr6	1501x	3.10e-28	4.01e-19	chemokine (c-x-c motif) receptor 6
Lilr4b	1097x	8.77e-17	1.22e-17	leukocyte immunoglobulin-like receptor subfamily B
Fgl2	710x	7.06e-15	1.11e-15	fibrinogen-like protein 2
Gm37352	598x	3.97e-14	6.55e-15	

Ccr5	578x	4.89e-14	8.90e-15	chemokine (c-c motif) receptor 5
Myo1f	517x	1.62e-13	2.78e-14	
Farp1	487x	2.84e-13	4.95e-14	
Hist1h2bb	487x	3.37e-13	5.90e-14	histone cluster 1
Podnl1	477x	3.53e-13	6.19e-14	podocan-like 1
Rgs16	419x	9.80e-13	1.76e-13	regulator of G-protein signaling 16

Table 9: *Most Downregulated Genes in Primary vs Naive T cells, from DESeq2 analysis.*

Gene	**Fold Change**	**FDR**	**P-value**	**Gene Description**
CT010467.1	9103x	2.18e-216	1.19e-218	18sRNA, related sequence 5
Mfsd4a	1806x	4.01e-19	5.05e-20	major facilitator superfamily domain containing 4a
Gm34006	1613x	1.76e-18	2.24e-19	Predicted genes
Gm11175	1491x	2.20e-35	1.47e-36	
Gm37401	1449x	5.82e-18	7.62e-19	
Gm20721	1397x	7.30e-18	9.63e-19	
Gm15441	1317x	1.69e-17	2.26e-18	
Gm26648	1280x	2.31e-17	3.11e-18	
Ccdc152	1200x	4.76e-17	6.51e-18	coiled coil domain containing 152
Bach2os	1160x	6.89e-17	9.54e-18	

Table 10: *Most Differentially Expressed Genes in Primary vs Naive T cells, from DESeq2 analysis.*

Gene	**Fold Change**	**FDR**	**P-value**	**Differential Expression**
CT010467.1	9103x	2.18e-216	1.19e-218	Downregulated

Mfsd4a	1806x	4.01e-19	5.05e-20	Downregulated
Gm34006	1613x	1.76e-18	2.24e-19	Downregulated
Cxcr6	1501x	3.10e-28	4.01e-19	Upregulated
Gm11175	1491x	2.20e-35	1.47e-36	Downregulated
Gm37401	1449x	5.82e-18	7.62e-19	Downregulated
Gm20721	1397x	7.30e-18	9.63e-19	Downregulated
Gm15441	1317x	1.69e-17	2.26e-18	Downregulated
Gm26648	1280x	2.31e-17	3.11e-18	Downregulated
Ccdc152	1200x	4.76e-17	6.51e-18	Downregulated

Figure 7: *Volcano Plot - Top 10 Upregulated and Downregulated Genes of Significance. Generated using volcano.R on RNA-seq data.*

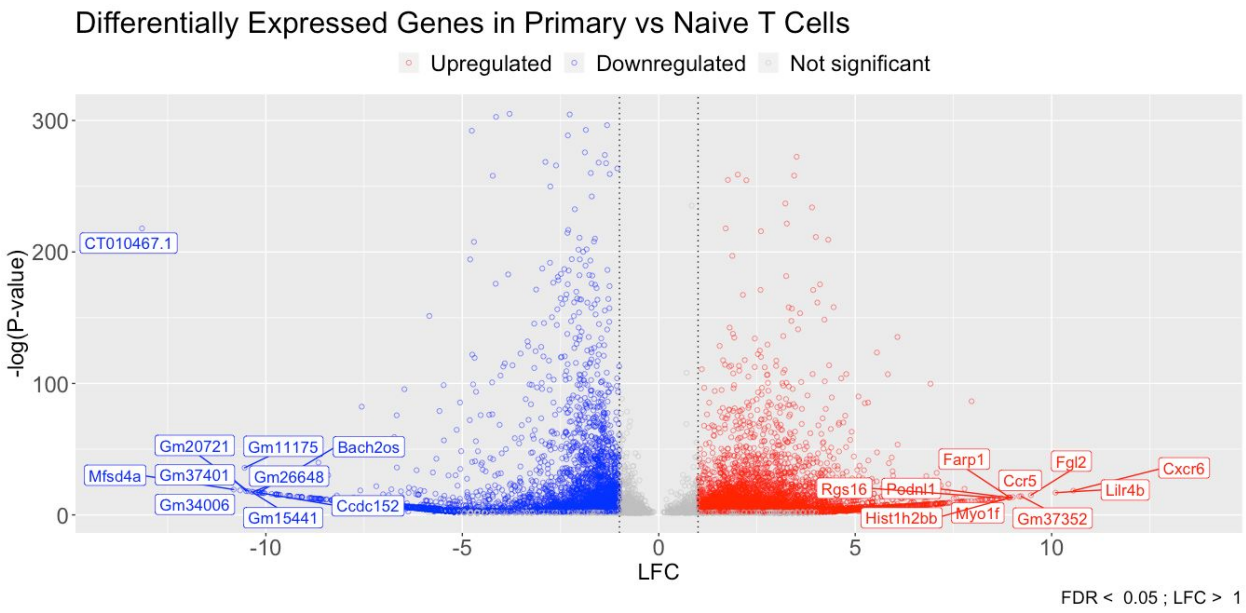


Table 11: *Top 10 Enriched KEGG Pathways of Significance. Generated and sorted using Cistrome-GO in ensemble mode using both RNA-seq and ChIP-seq data.*

KEGG Pathway	Enrichment	P-value	FDR	# of Genes
T cell receptor signaling pathway	3.128	2.545E-18	7.942E-16	101

Th17 cell differentiation	7.529	5.949E-18	9.281E-16	102
Viral carcinogenesis	2.084	1.000E-15	1.040E-13	227
Endocytosis	1.998	5.615E-15	4.380E-13	267
Th1 and Th2 cell differentiation	3.081	7.400E-15	4.617E-13	87
Human T cell leukemia virus 1 infection	2.016	9.444E-15	4.911E-13	243
Apoptosis	2.312	1.749E-13	7.797E-12	136
Hepatitis B	2.214	4.421E-13	1.379E-11	163
Epstein-Barr virus infection	2.059	4.268E-13	1.379E-11	225
Cellular senescence	2.283	4.231E-13	1.379E-11	181

Table 12: *Top 10 Enriched GO Molecular Functions of Significance. Generated and sorted using Cistrome-GO in ensemble mode using both RNA-seq and ChIP-seq data.*

GO Molecular Function	Enrichment	P-value	FDR	# of Genes
Kinase binding	2.439	2.888E-30	2.958E-27	801
Transcription regulator activity	2.098	1.740E-25	8.908E-23	1332
Protein domain specific binding	2.381	5.904E-25	1.816E-22	796
RNA binding	2.106	7.093E-25	1.816E-22	1024
Double-stranded DNA binding	2.281	5.164E-23	1.058E-20	984
Regulatory region nucleic acid binding	2.21	1.093E-22	1.865E-20	970
Transcription factor binding	2.31	2.331E-22	3.410E-20	691
Sequence-specific DNA binding	2.145	1.302E-21	1.667E-19	1121
Proximal promoter sequence-specific DNA binding	2.72	3.935E-21	4.477E-19	566
RNA polymerase II regulatory region DNA binding	2.35	5.331E-21	5.459E-19	819

Table 13: *Top 10 Enriched GO Biological Processes of Significance. Generated and sorted using Cistrome-GO in ensemble mode using both RNA-seq and ChIP-seq data.*

GO Molecular Function	Enrichment	P-value	FDR	# of Genes
Regulation of immune system process	2.407	3.481E-41	1.657E-37	1297
Immune system process	2.361	7.952E-39	1.892E-35	1445
Positive regulation of gene expression	1.923	2.582E-33	4.096E-30	1941
Negative regulation of gene expression	1.95	8.256E-31	9.824E-28	1647
Regulation of transcription by RNA polymerase II	1.903	2.977E-30	2.834E-27	1955
Positive regulation of nucleobase-containing compound metabolic process	1.886	9.406E-30	7.462E-27	1805
Positive regulation of immune system process	2.447	2.582E-29	1.756E-26	864
Positive regulation of biosynthetic process	1.848	3.217E-29	1.914E-26	1910
Regulation of protein modification process	1.915	5.387E-29	2.849E-26	1738
Regulation of lymphocyte activation	5.017	7.767E-29	3.697E-26	419

Table 14: *Top 10 Genes by Rank Product. Sorted using Cistrome-GO in ensemble mode using both RNA-seq and ChIP-seq data.*

Gene	RP Score	Adjusted RP score	logFoldChange	DE Significance	Ranked by Rank Product
Ahnak	2.603	0.835	6.713	0.000E+00	1
Ubash3a	3.116	1	3.107	1.005E-43	2
Gpr132	2.636	0.846	-1.062	7.285E-103	3

Socs3	0.786	0.252	-2.472	0.000E+00	4
Rps4x	0.773	0.248	-1.941	0.000E+00	5
Dennd4a	0.728	0.234	2.985	0.000E+00	6
Zc3hav1	0.695	0.223	-1.608	0.000E+00	7
Stk17b	0.685	0.215	-1.948	0.000E+00	8
H2-D1	1.878	0.6	1.759	3.016E-253	9
Jak1	0.652	0.209	-1.877	0.000E+00	10

Table 15: *Top 10 genes by H3K27ac regulatory potential. Sorted using Cistrome-GO in solo mode using only ChIP-seq data.*

Gene	RP Score	Adjusted RP score	Ranked by adjusted RP
Ubash3a	3.116	1	1
Rhof	3.016	0.968	2
Trib1	2.704	0.868	3
Gpr132	2.636	0.846	4
Inf2	2.616	0.84	5
Ahnak	2.603	0.835	6
Hist1h2bb	2.591	0.829	7
Angptl2	2.519	0.809	8
Gramd1a	2.514	0.807	9
Myd88	2.511	0.806	10

Figure 8: *Volcano Plot - Differential Expression of Top 10 Genes from CistromeGO. Generated using volcano.R on RNA-seq data.*

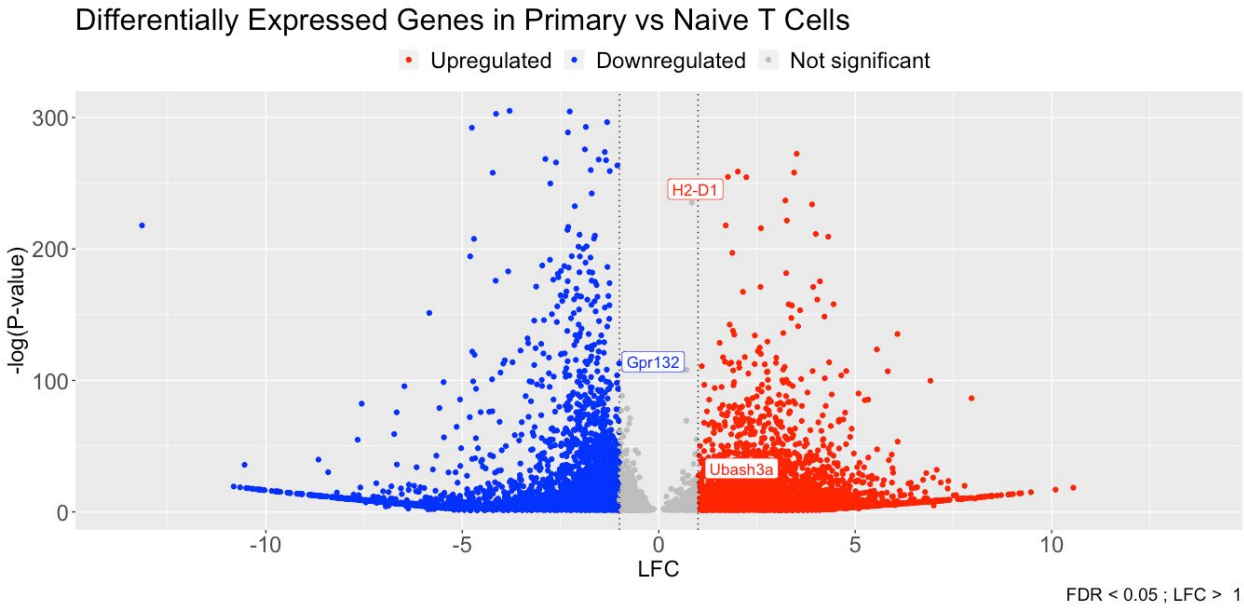
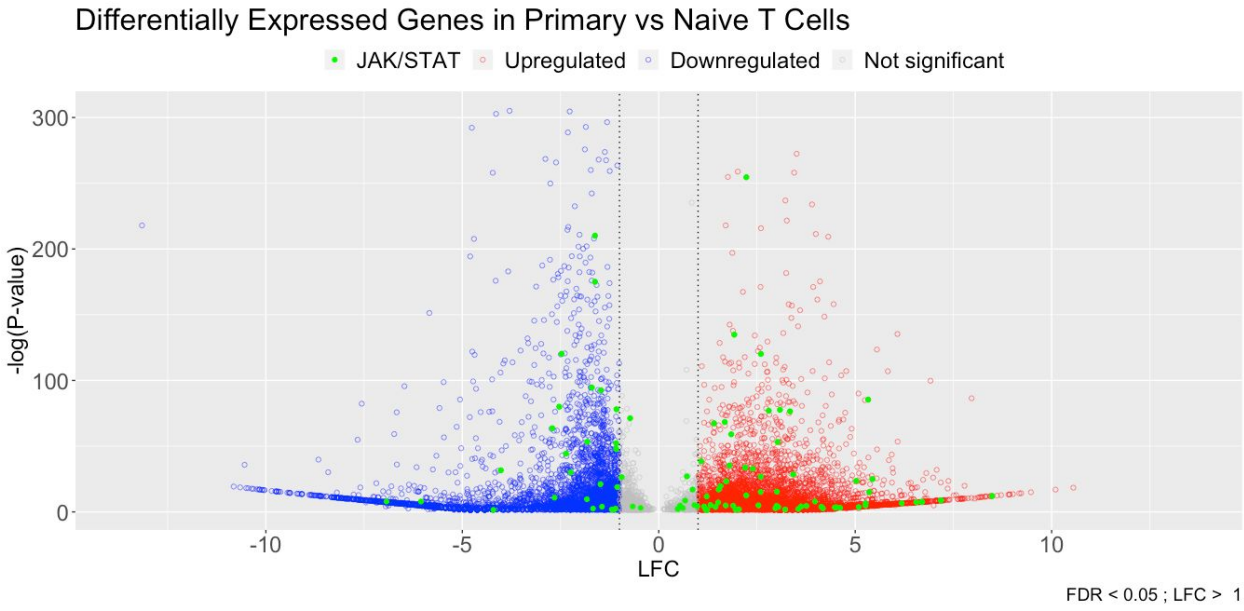


Figure 9: *Volcano Plot - Differential Expression of Genes in the JAK/STAT Pathway. Generated using volcano.R on RNA-seq data.*



(Figures 10-20) show comparisons of significant genes in our analysis. The top 4 tracks are ChIP-Seq acetylations, and the remaining are RNA-seq coverage tracks. Reddish-pink tracks represent Primary T cells while Blueish-purple tracks represent Naive T cells

Figure 10: *Ahnak gene acetylation in Naive and Primary CD4⁺ T cells. Screenshot from Integrative Genomics Viewer.*

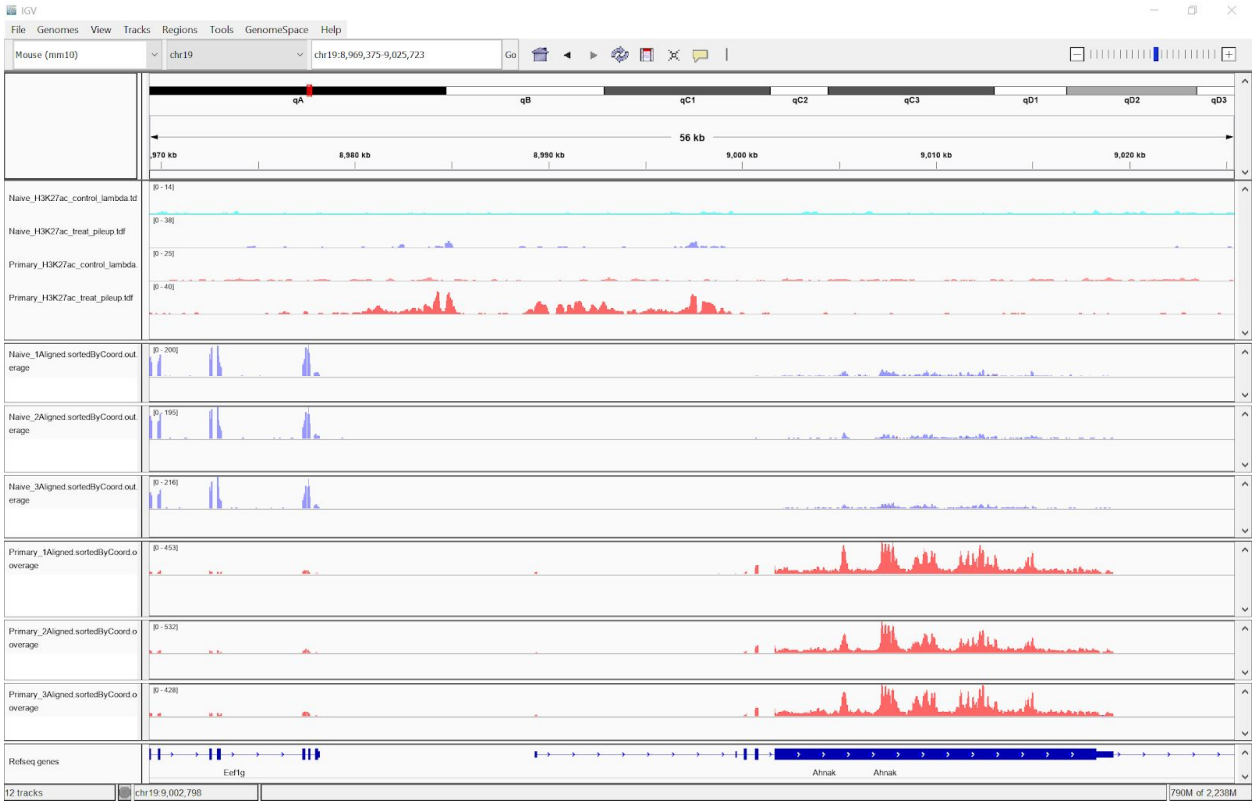


Figure 11: *Ccr5 gene acetylation in Naive and Primary CD4+ T cells. Screenshot from Integrative Genomics Viewer (IGV)* (12).

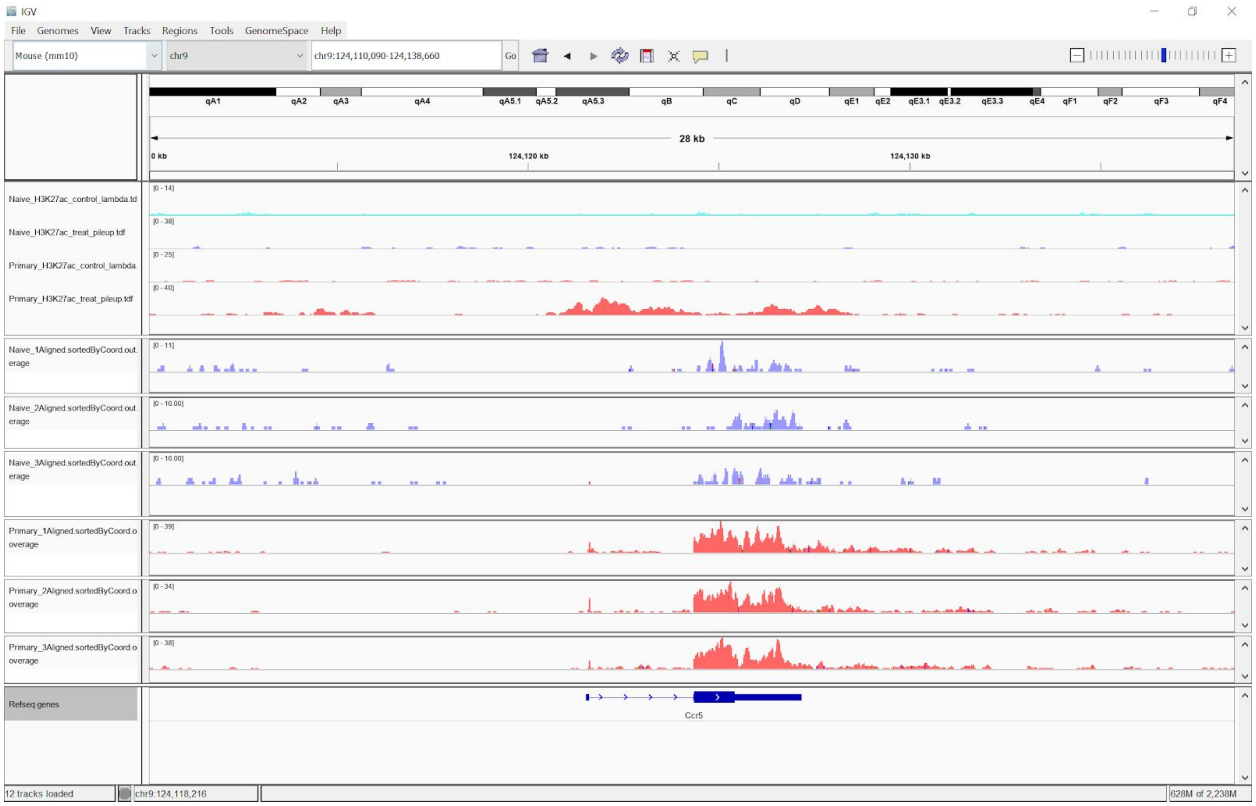


Figure 12: *Jak1 gene acetylation in Naive and Primary CD4⁺ T cells. Screenshot from Integrative Genomics Viewer (IGV)* ***(12)****.*

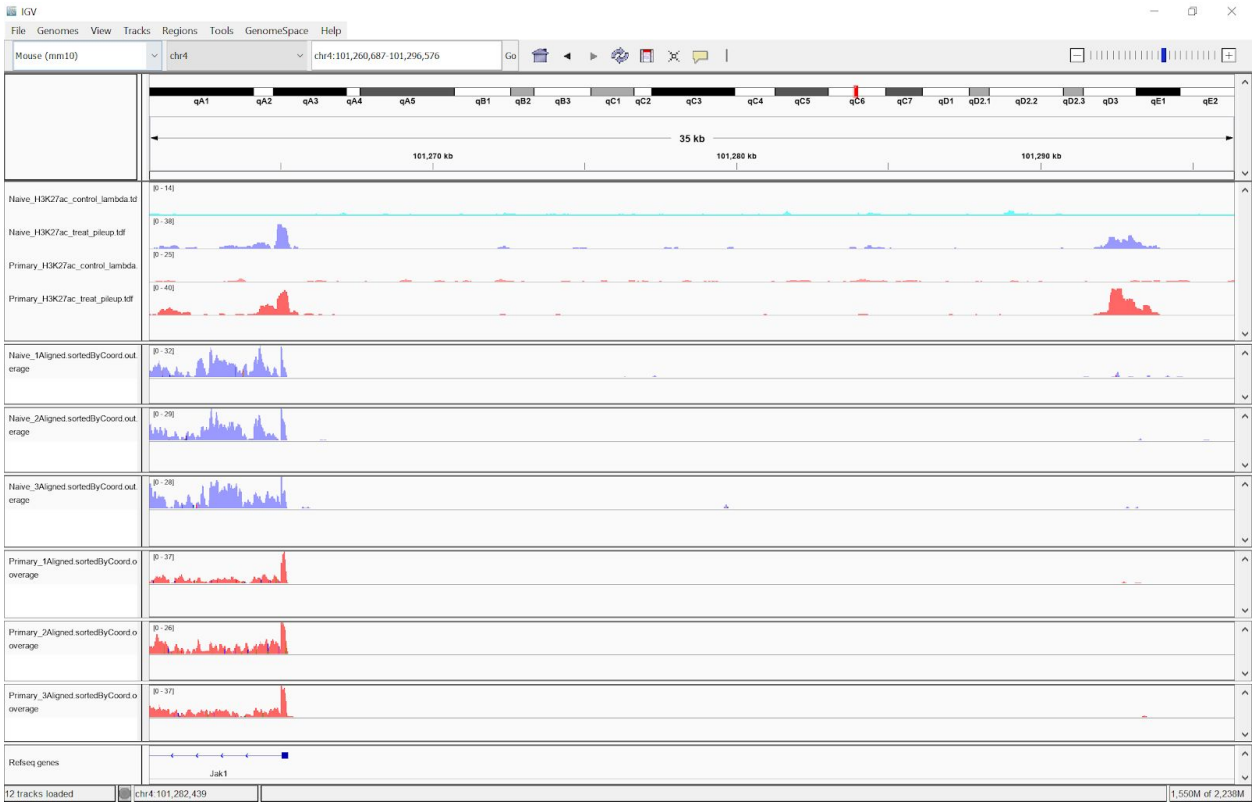


Figure 13: *Dennd4a gene acetylation in Naive and Primary CD4⁺ T cells. Screenshot from Integrative Genomics Viewer (IGV)* ***(12)****.*

Figure 14: *Gpr132 gene acetylation in Naive and Primary CD4⁺ T cells. Screenshot from Integrative Genomics Viewer (IGV)* (12).

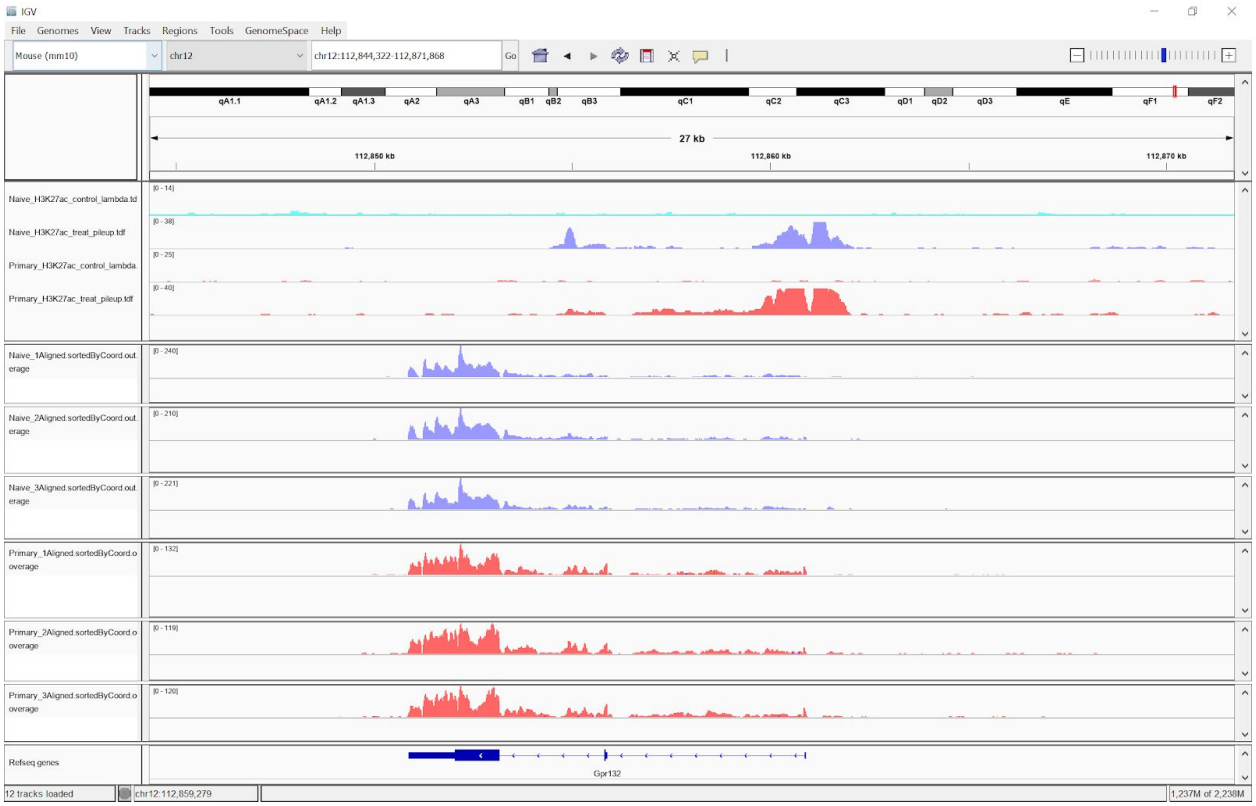


Figure 15: *H2D1 gene acetylation in Naive and Primary CD4⁺ T cells. Screenshot from Integrative Genomics Viewer (IGV)* ***(12)****.*

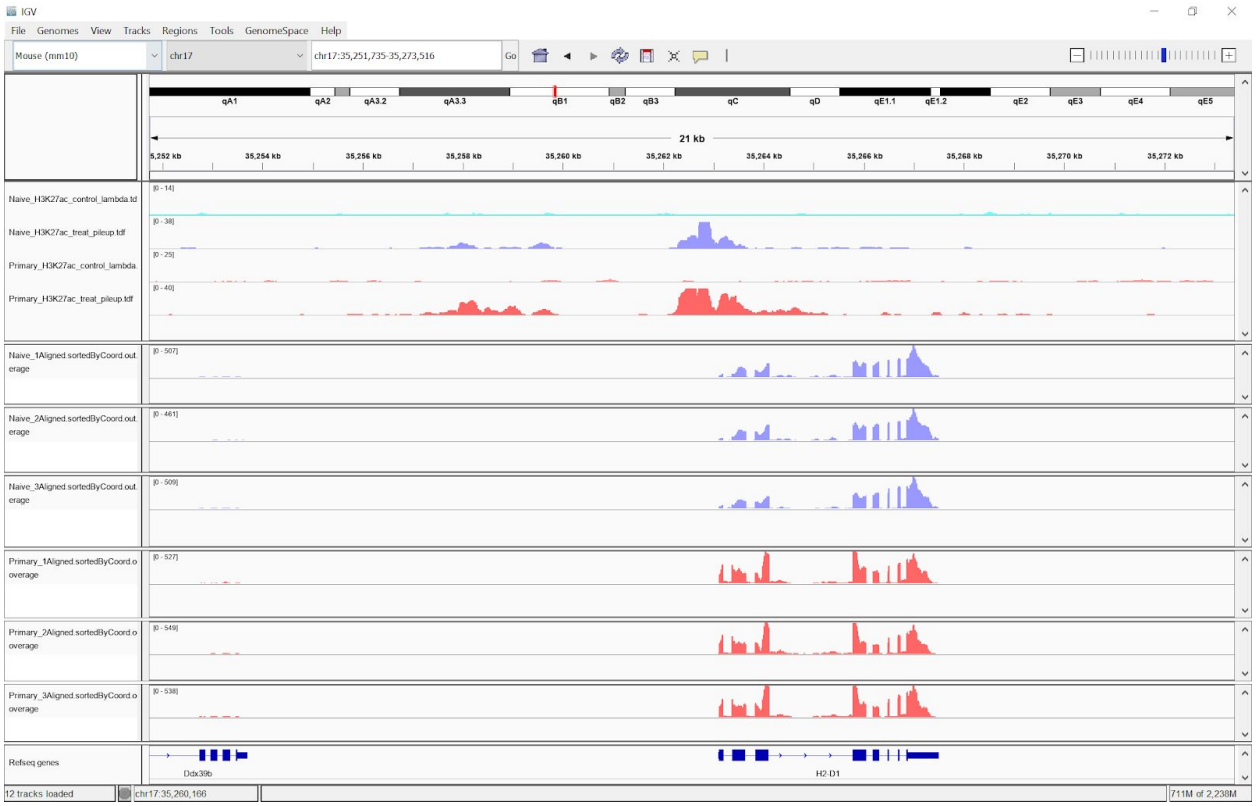


Figure 16: *Rps4x gene acetylation in Naive and Primary CD4⁺ T cells. Screenshot from Integrative Genomics Viewer (IGV)* (12).

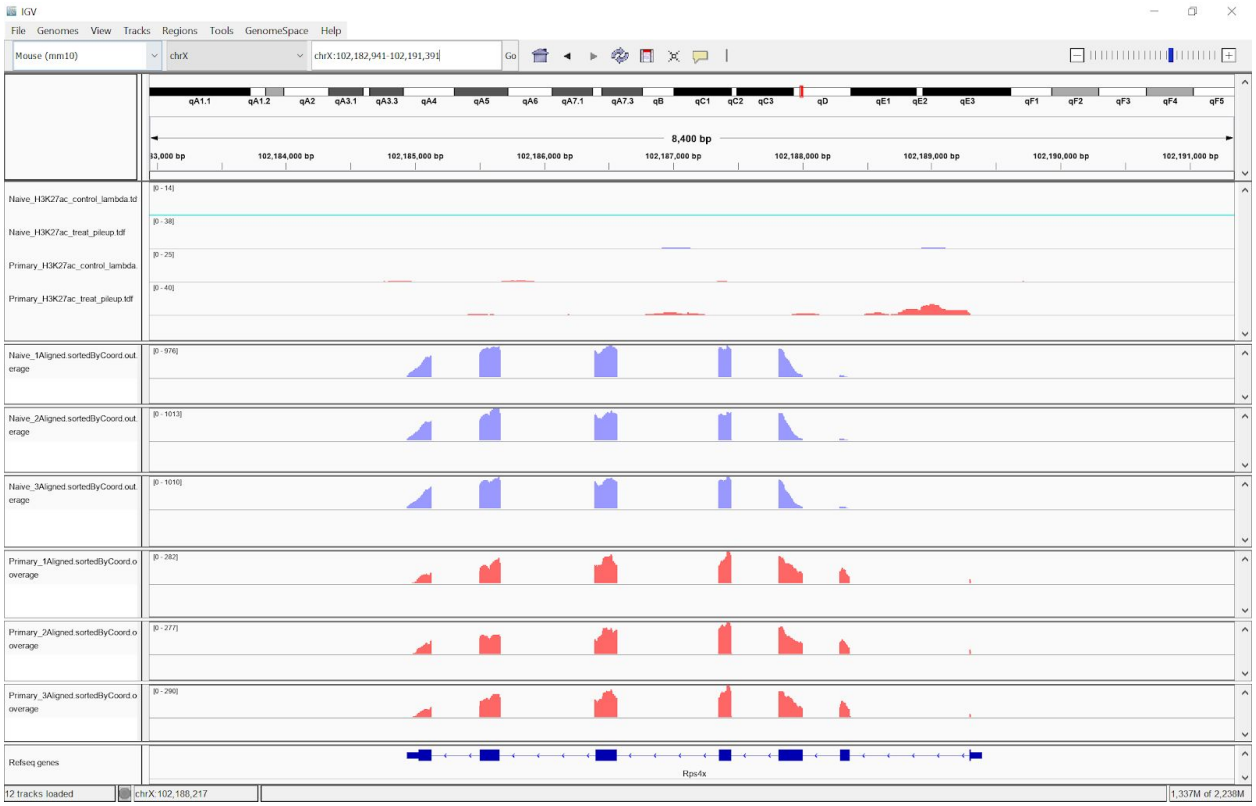


Figure 17: *Socs3 gene acetylation in Naive and Primary CD4⁺ T cells. Screenshot from Integrative Genomics Viewer (IGV)* (12).

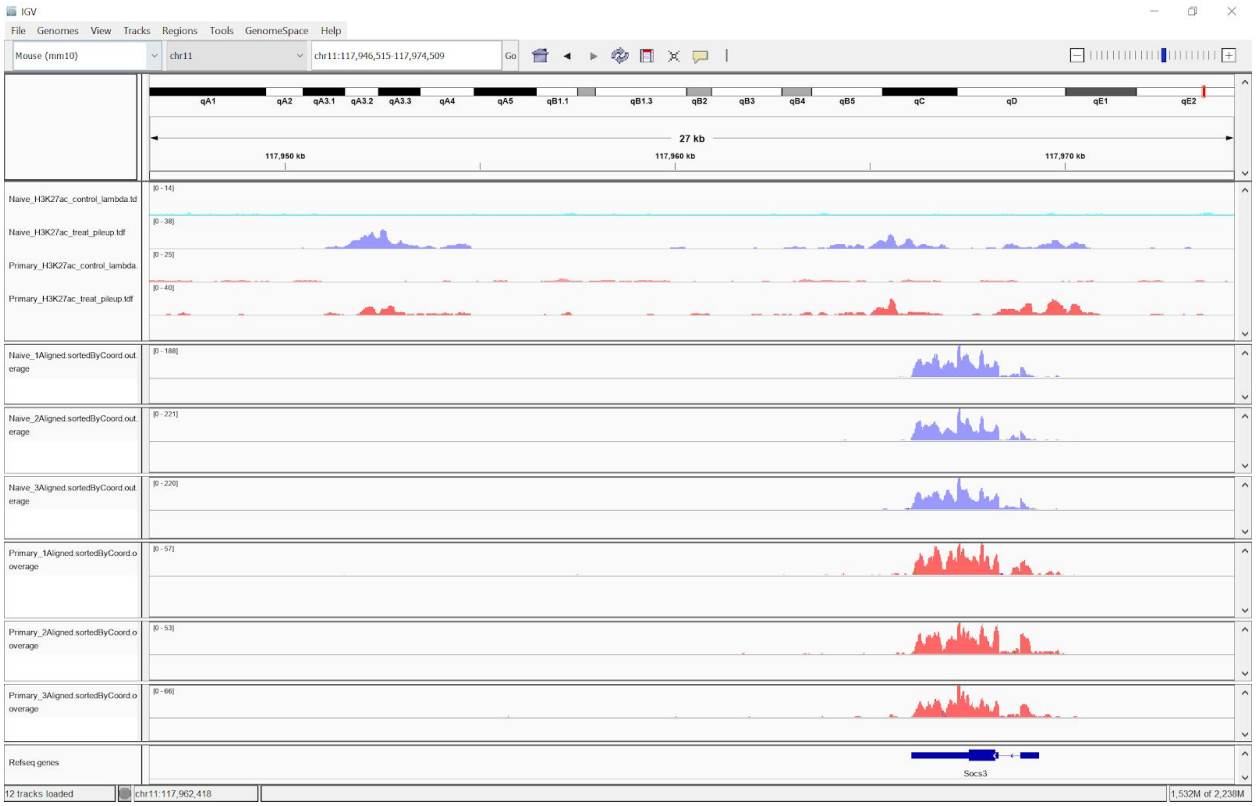


Figure 18: *Ubash3a gene acetylation in Naive and Primary CD4⁺ T cells. Screenshot from Integrative Genomics Viewer (IGV)* ***(12)****.*

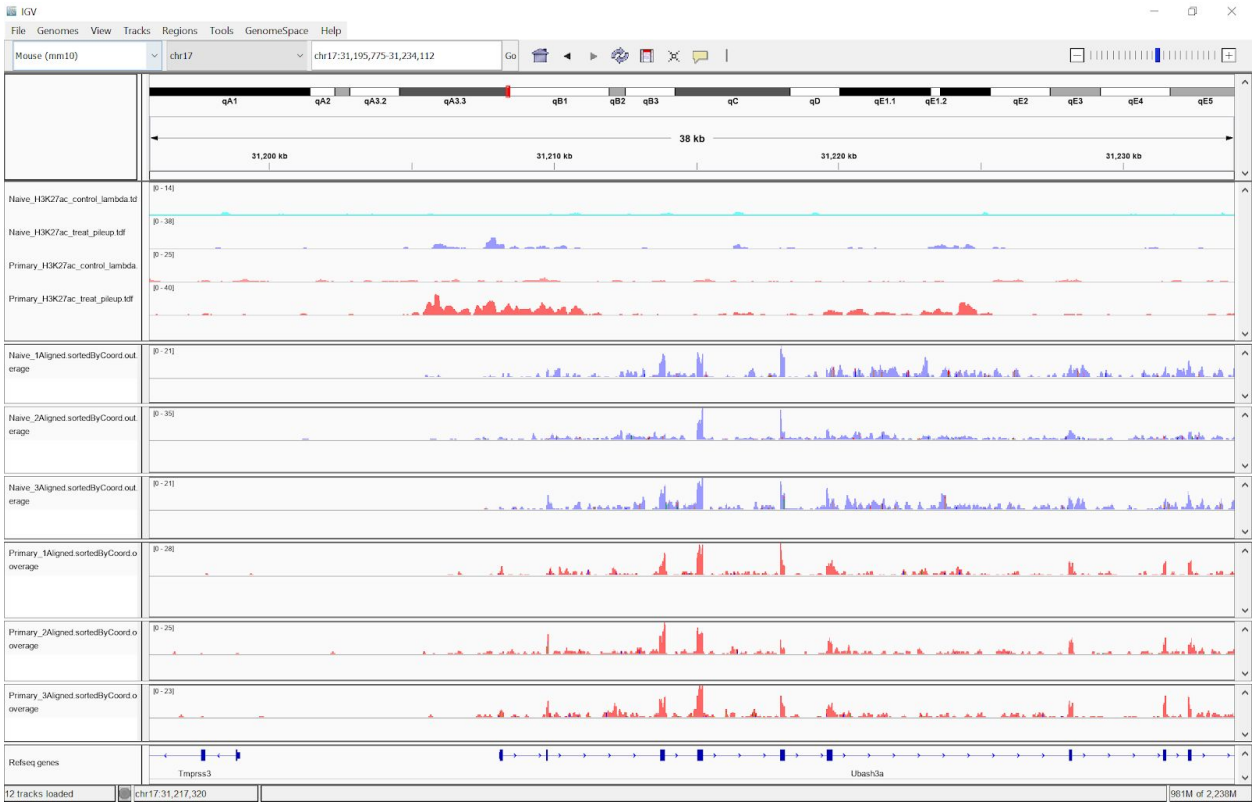


Figure 19: *Zc3hav1 gene acetylation in Naive and Primary CD4⁺ T cells. Screenshot from Integrative Genomics Viewer (IGV)* ***(12)****.*

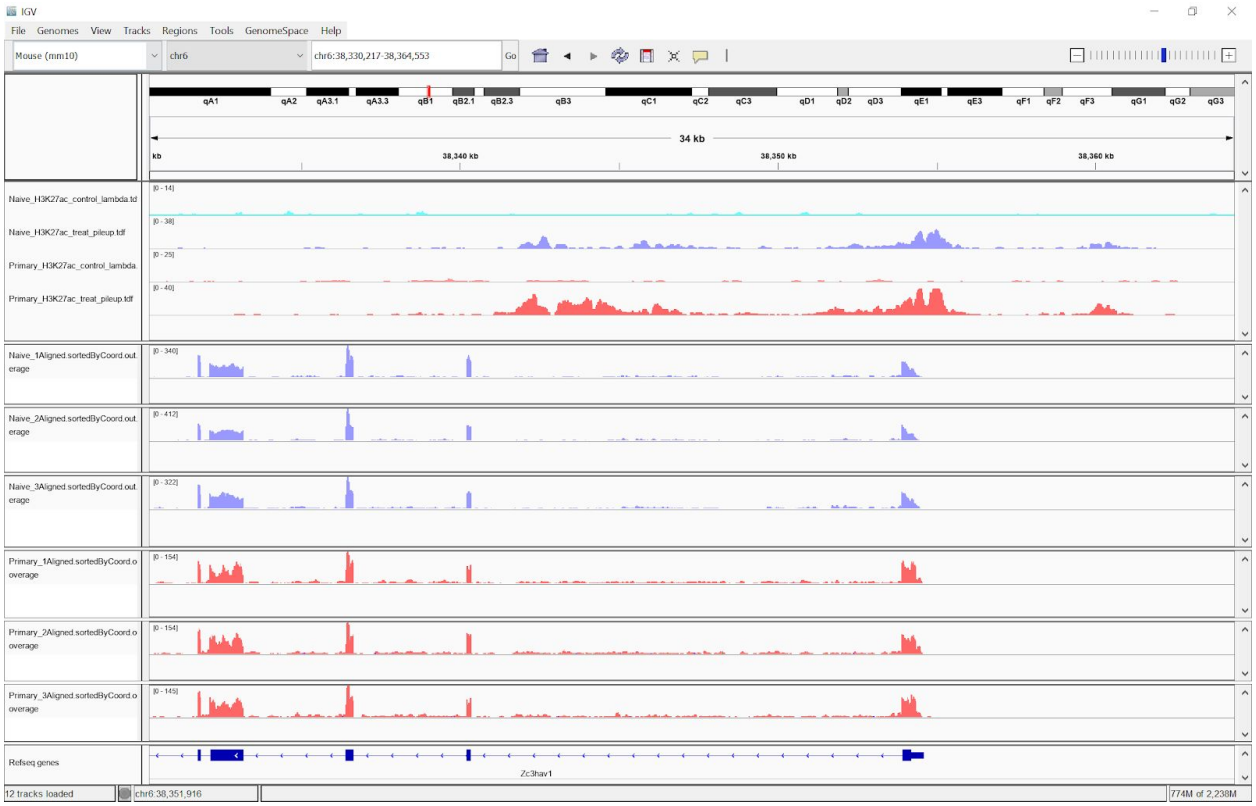


Figure 20: *Stk17b gene acetylation in Naive and Primary CD4⁺ T cells. Screenshot from Integrative Genomics Viewer (IGV)* ***(12)****.*

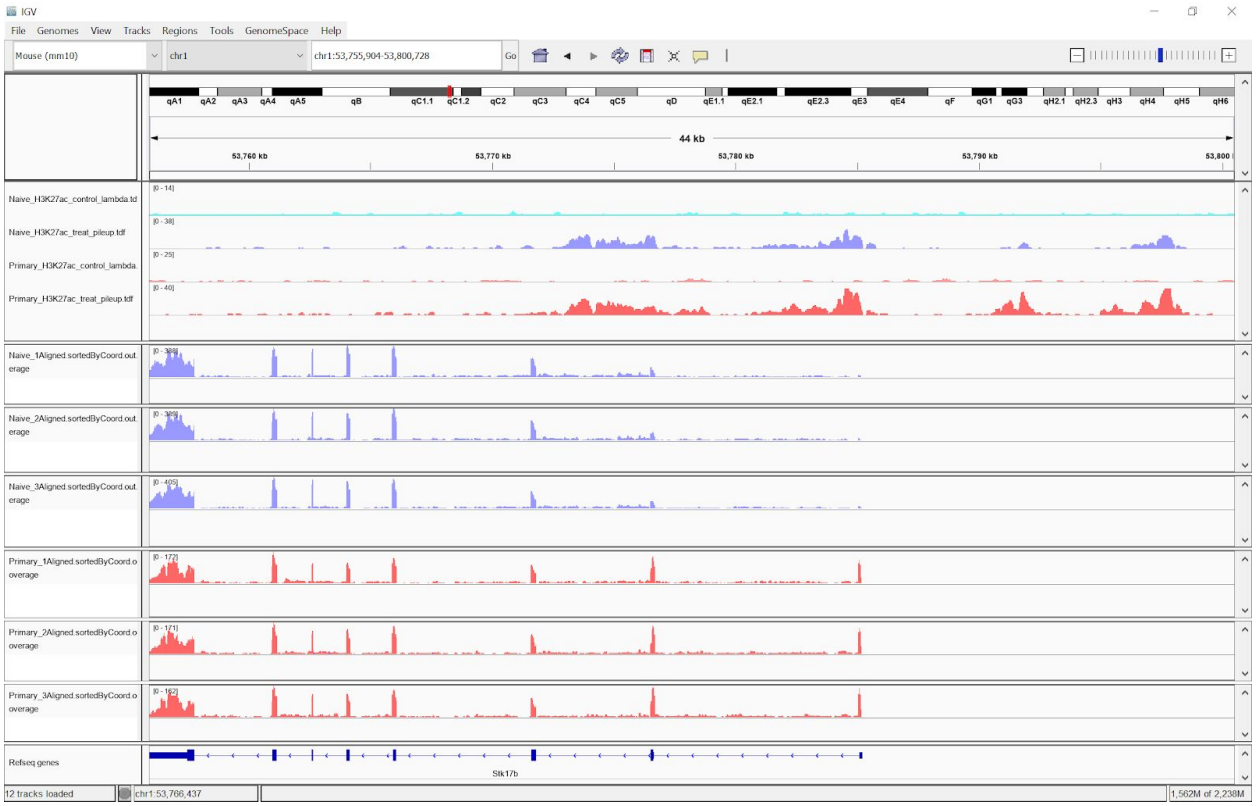


Figure 21: *Acetylation near CCR5 in Naive, Primary, and Memory T Cells. Screenshot from Integrative Genomics Viewer (IGV)* ***(12)****.*

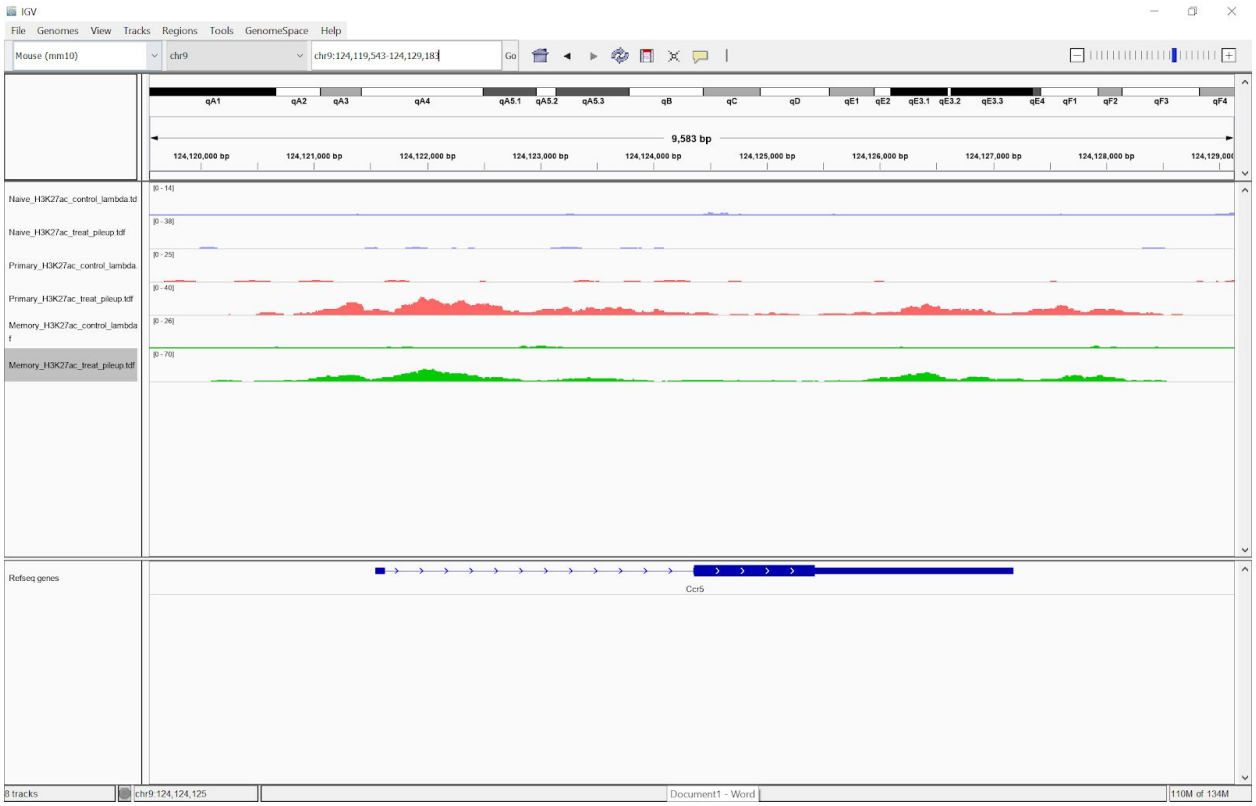


Figure 22: *Acetylation near Cxcr6 in Naive, Primary, and Memory T Cells. Screenshot from Integrative Genomics Viewer (IGV)* ***(12)****.*

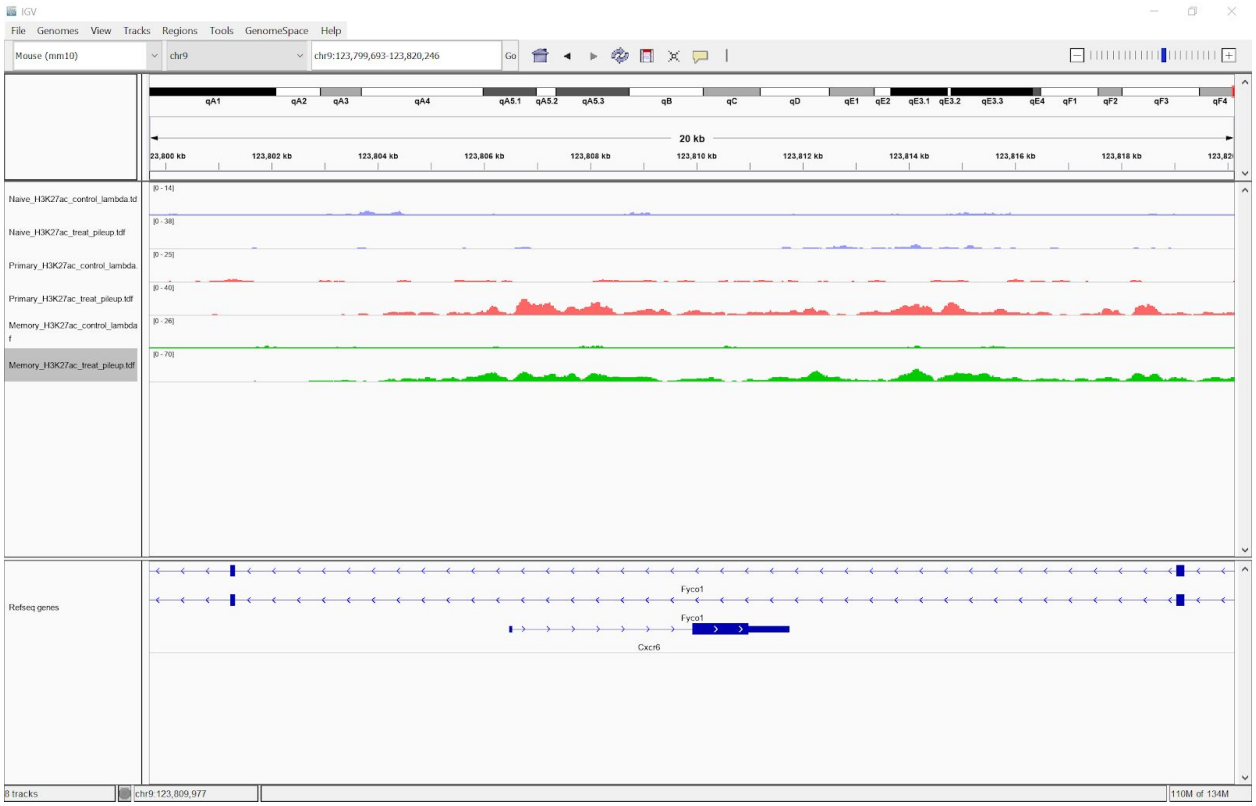


Figure 23: *Chemokine Signaling Pathway from KEGG* ***(14)****.*
(https://www.kegg.jp/kegg-bin/highlight_pathway?scale=1.0&map=map04062&keyword=cxcr6)