



CLEVELAND  
INSTITUTE FOR  
COMPUTATIONAL  
BIOLOGY

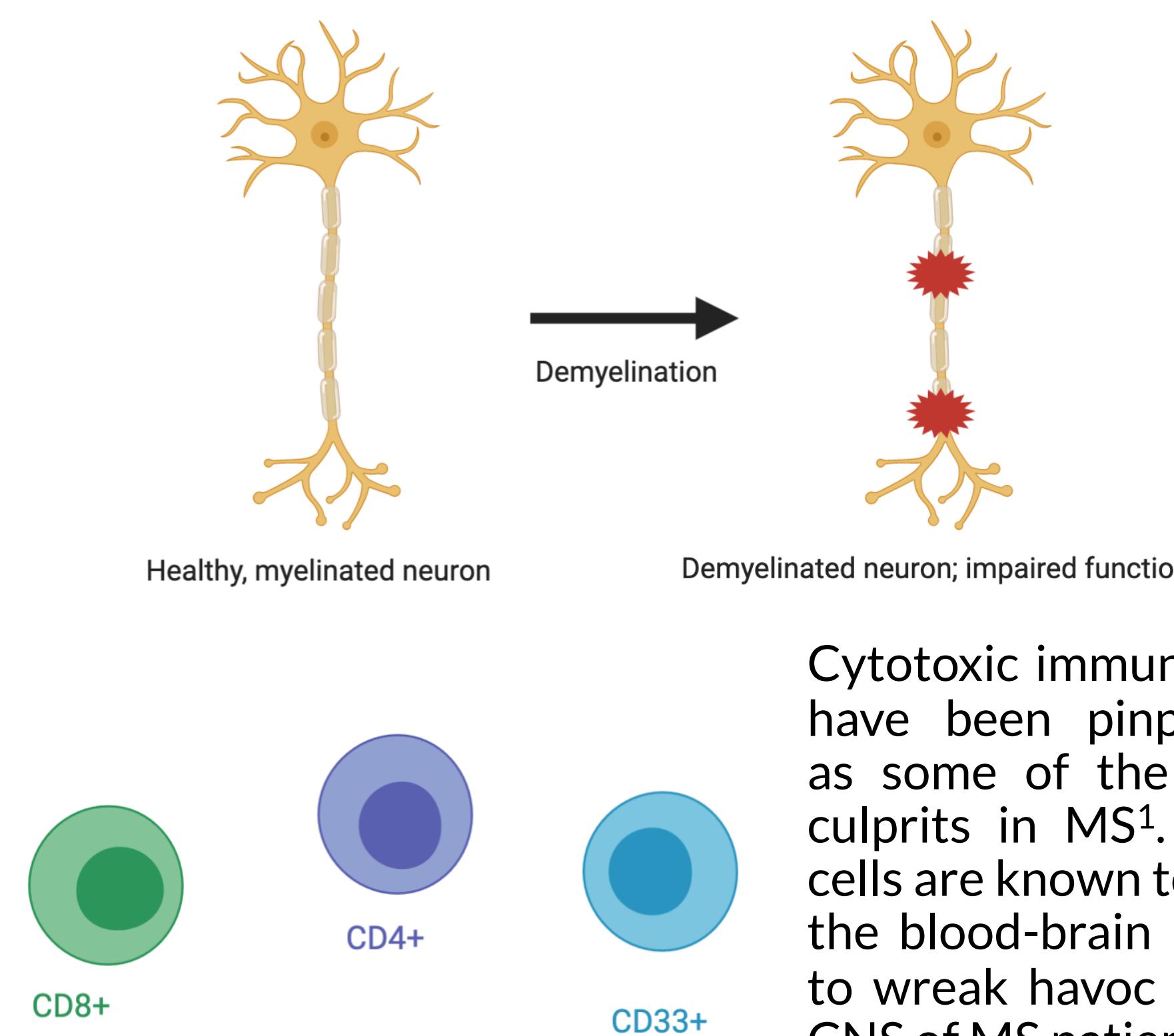
# EPIGENETIC AND GENE REGULATORY ARCHITECTURE OF MULTIPLE SCLEROSIS IN HISPANIC POPULATIONS

## ABSTRACT

Multiple Sclerosis (MS) is a chronic, inflammatory disease that affects the central nervous system (CNS), leaving individuals from a wide gamut of ages suffering from periodic "attacks" that severely damage the CNS and impair its functions. While the underlying cause of this disease remains unknown, cytotoxic T-cells have been identified as one of the main culprits behind the autoimmune response that characterizes MS, and thus are prime candidates for the study of MS and its genetic bases. While previous studies have assessed the molecular landscape of these cells and other immune cells, none have performed such experiments on cells from multi-ethnic MS patients. More specifically, chromatin accessibility and the transcriptome of these cells in MS patients with more diverse genome composition remain undescribed in the literature, as a majority of studies have focused on individuals of European descent. Studies in genetically admixed populations provide an important opportunity to observe new haplotype combinations and to elucidate how the interplay between chromatin accessibility and gene expression impacts the MS phenotype. To fill this gap in the literature, we have generated data from 16 Puerto Rican Hispanic MS patients by first performing cell sorting of whole blood to isolate CD4+ and CD8+ T-cells and CD33+ monocytes. We then performed RNAseq and ATACseq on these cell isolates. With these sequencing data, we hypothesized that the accessibility of certain regions of the chromatin would influence the expression levels of the genes that lie in those regions. Our preliminary data suggests that the gene expression levels for *MALT1*, *PGE2*, and *TNFRSF1A* are significantly influenced by chromatin conformation. Strikingly, some of these patterns are preserved across the various cell types surveyed from MS patients. Future work will be directed at further refinement of such conserved patterns and identification of additional relationships. In addition, we expect our data will provide new insights on the role that certain genomic elements, such as enhancers, may play in MS. The integration of gene expression and chromatin conformation data will allow us to reconstruct cell-specific gene regulatory networks, which will be invaluable tools for discovering underlying mechanisms of gene regulation in MS in multi-ethnic patients. This study represents the first ever cell-level examination of chromatin accessibility and its effect on gene expression levels in the context of an admixed population, which in turn highlights key differences in the immune system function of Hispanic MS.

## BACKGROUND

Multiple sclerosis (MS) is an autoimmune disease of the central nervous system (CNS). In it, the immune cells of a patient attack the myelin sheaths of neurons. Consequently, the axons of these neurons are left exposed, which disrupts electrical messaging of the CNS and leads to a wide array of symptoms in patients. The most severe are related to motor and sensory function.



Cytotoxic immune cells have been pinpointed as some of the major culprits in MS<sup>1</sup>. These cells are known to cross the blood-brain barrier to wreak havoc on the CNS of MS patients.

While MS is typically associated with European populations<sup>2</sup>, recent studies have shown that MS has a significant prevalence among Hispanic/Latinx populations. These individuals are multi-ethnic, which provides ample opportunities for genetic studies of MS. Therefore, we set out to characterize the molecular landscape of MS in a cohort of Puerto Rican MS patients ( $n = 16$ ). We focused specifically on the transcriptome, chromatin accessibility, and genetic variants of certain cytotoxic immune cells in these individuals.

Table 1. Characteristics of Patient Cohort

Gender	Age at Ascertainment	Age at 1 <sup>st</sup> Symptom	Age at 1 <sup>st</sup> Diagnosis	Family History of MS	Relapsing MS (Y/N)
Male 4	<= 30 2	<= 30 3	<= 30 2	Yes 2	Yes 4
Female 5	30-55 6	> 40 1	> 39 2	No 7	
	> 60 1				

\*Some clinical data are still being extracted from clinical charts and are currently unavailable.

S. Cruz-González<sup>1</sup>, H. Bai<sup>1,2</sup>, A.H. Beecham<sup>3,4</sup>, K.L. Miskimen<sup>1</sup>, C.P. Manrique<sup>3</sup>, W. Hulme<sup>3</sup>, AJ Griswold<sup>3,4</sup>, J.L. McCauley<sup>3,4</sup>, W.S. Bush<sup>1</sup>

<sup>1</sup>Cleveland Institute for Computational Biology, Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH, USA; <sup>2</sup>Department of Nutrition, Case Western Reserve University, Cleveland, OH, USA; <sup>3</sup>John P. Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, FL, USA; <sup>4</sup>Dr. John T. Macdonald Department of Human Genetics, Miller School of Medicine, University of Miami, Miami, FL, USA.

## APPROACH

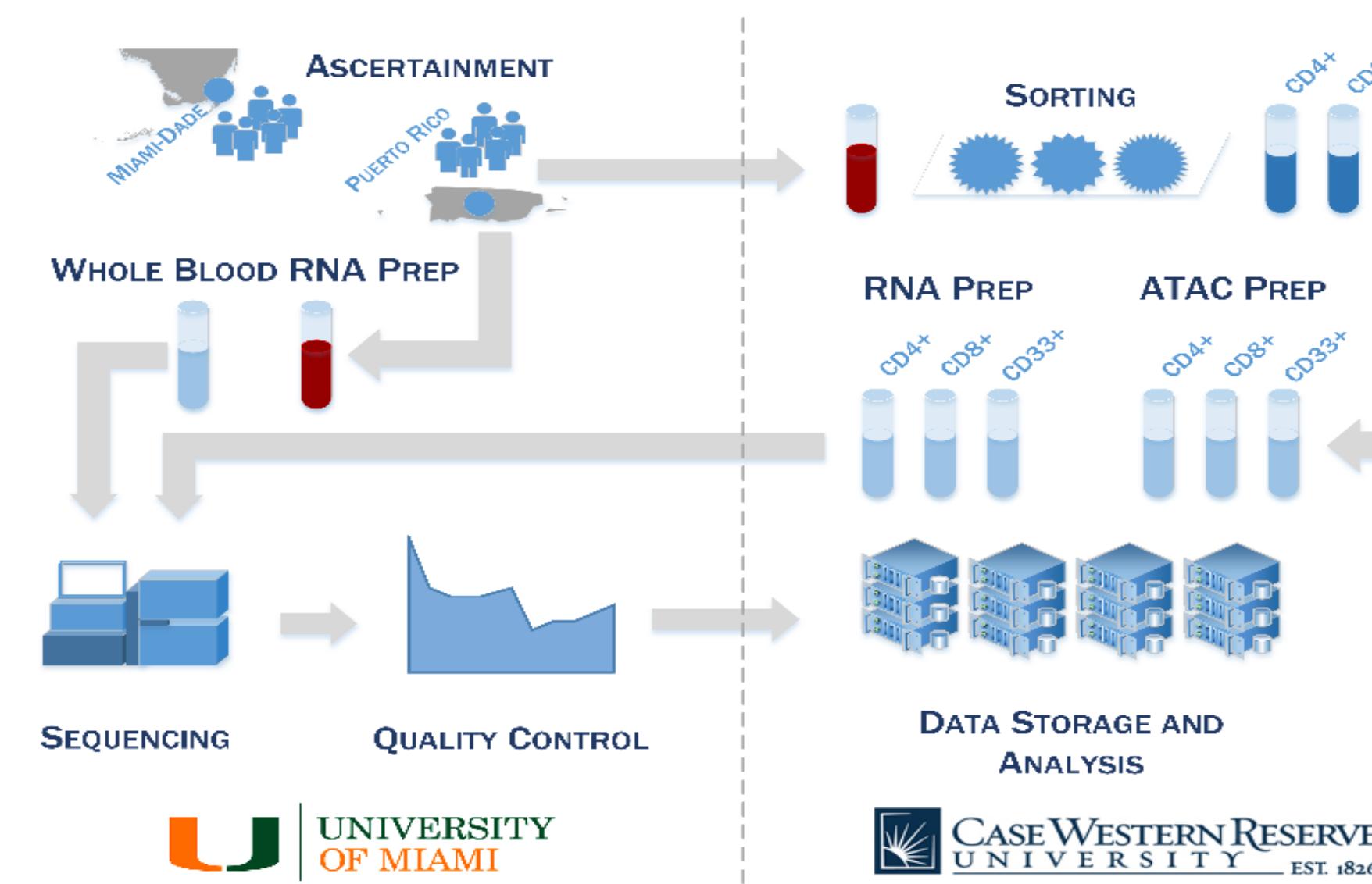


Figure 1: Our multi-omics approach provides a broad overview of the molecular landscape of cytotoxic cells of interest in multi-ethnic MS patients, with four datasets for each patient. These include including genotyping, cell-specific ATAC-seq, cell-specific RNA-seq, and whole-blood RNA-seq.

## RESULTS

We first performed a global ancestry analysis of the individuals in our cohort using the ADMIXTURE software<sup>3</sup>. The individuals in our cohort are mostly of European descent, but also show African and Native American ancestry (Figure 2).

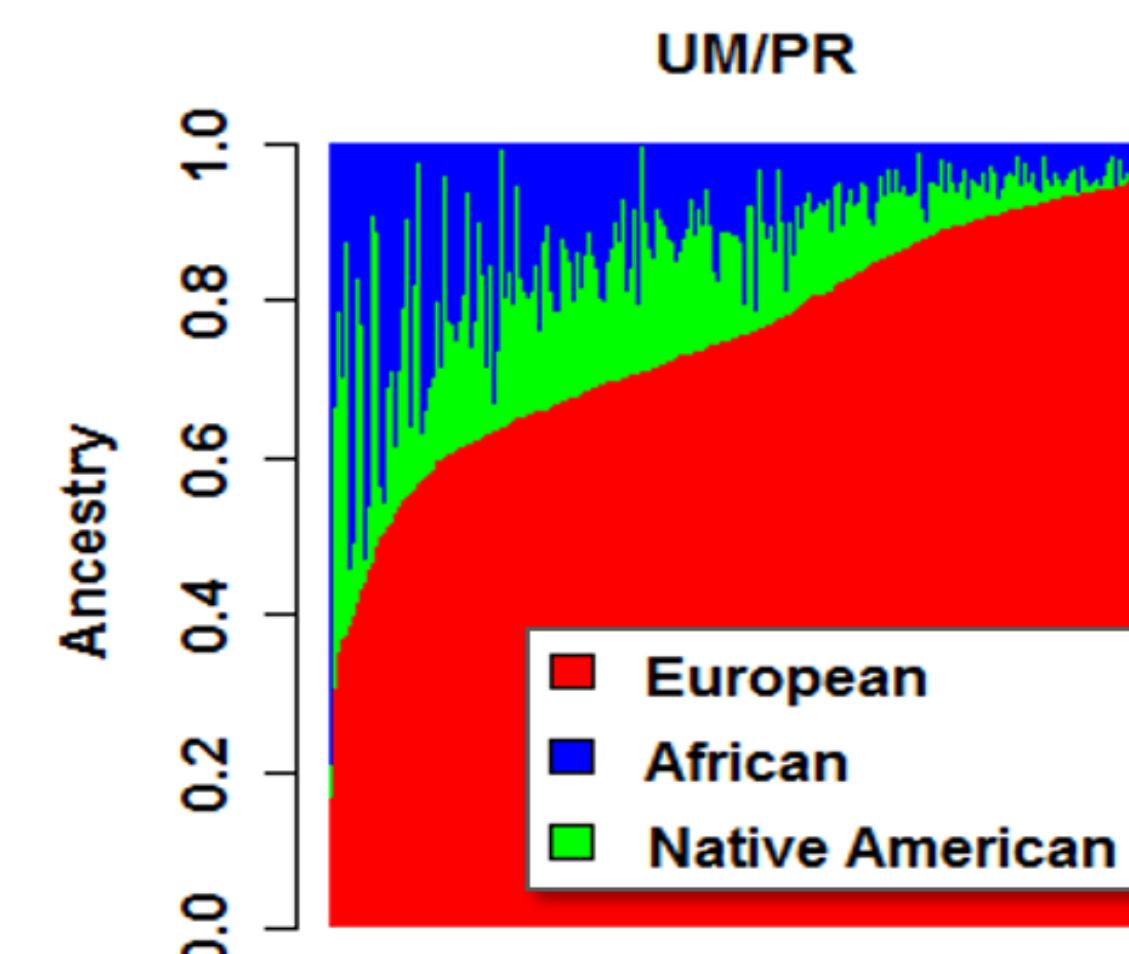


Figure 2: ADMIXTURE plot of the Puerto Rican MS individuals from which our samples were drawn.

Peak-calling using the HMMRATAC software<sup>4</sup> on the ATAC-seq data provided us an overview of locations of the chromatin that were accessible. We plotted these peaks relative to transcription start sites (TSS) (Figure 3). Next, we asked what types of locations on the genome these peaks corresponded to. For this, we used the ChIPseeker R library<sup>5</sup> and annotated our peaks (Figures 4 and 5) relative to their genomic location.

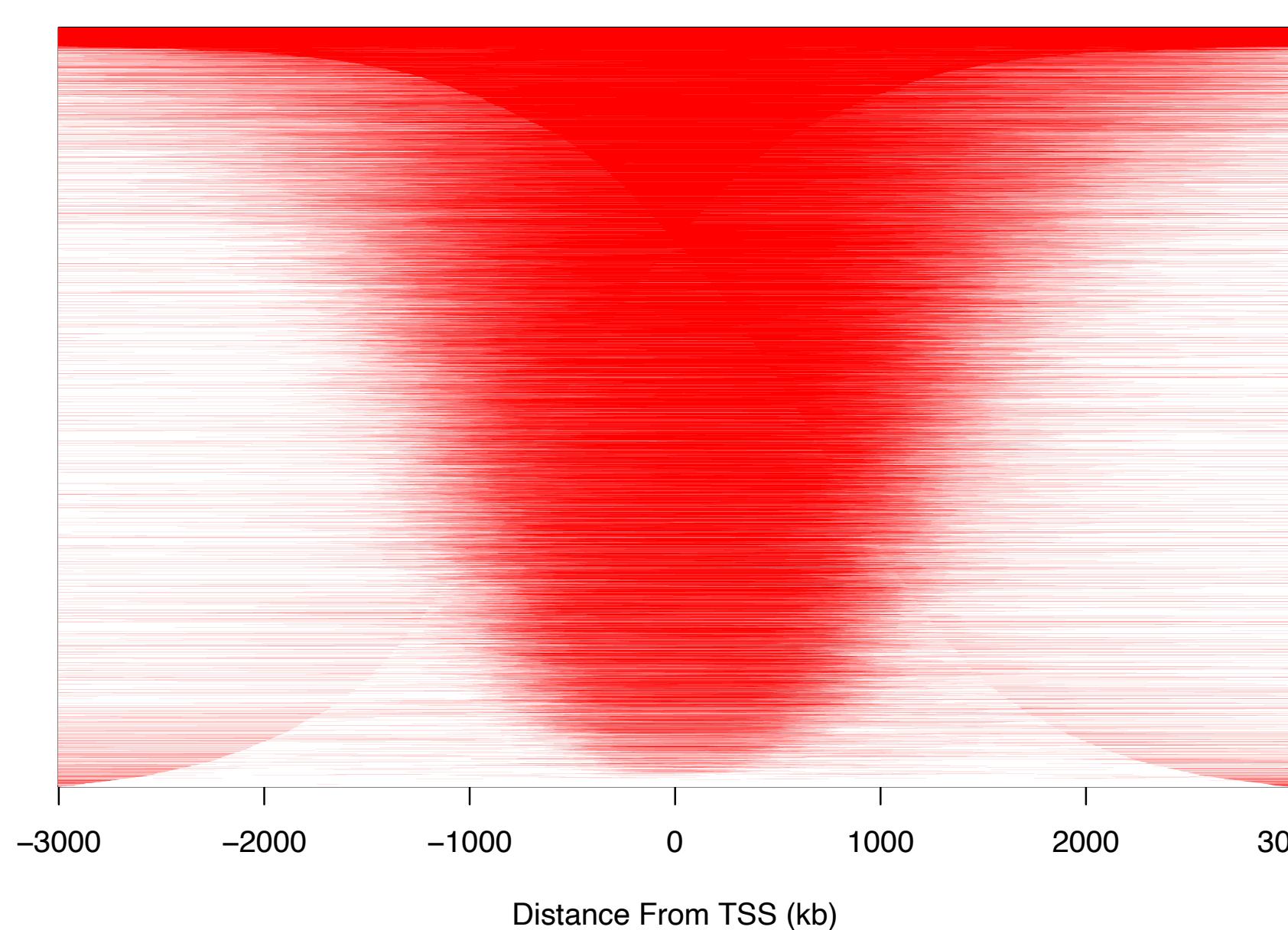


Figure 3: Heatmap of ATAC-seq peaks for all MS patients in our cohort.

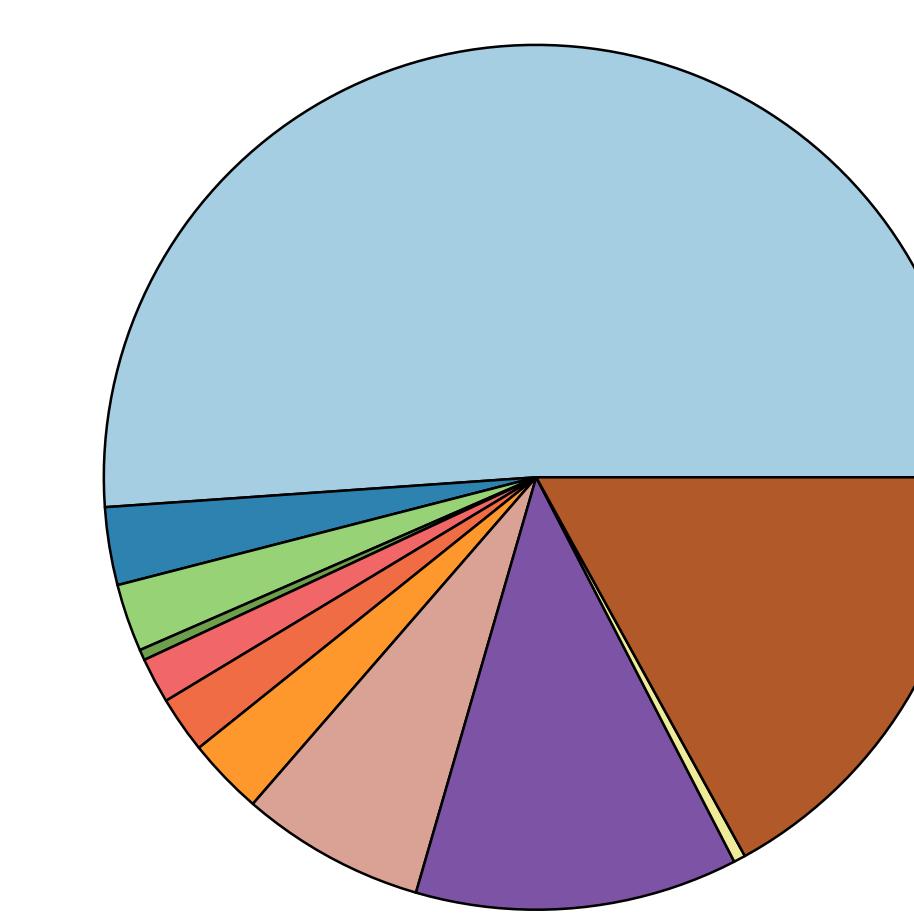


Figure 4: Pie chart showing distribution of peak location categories.

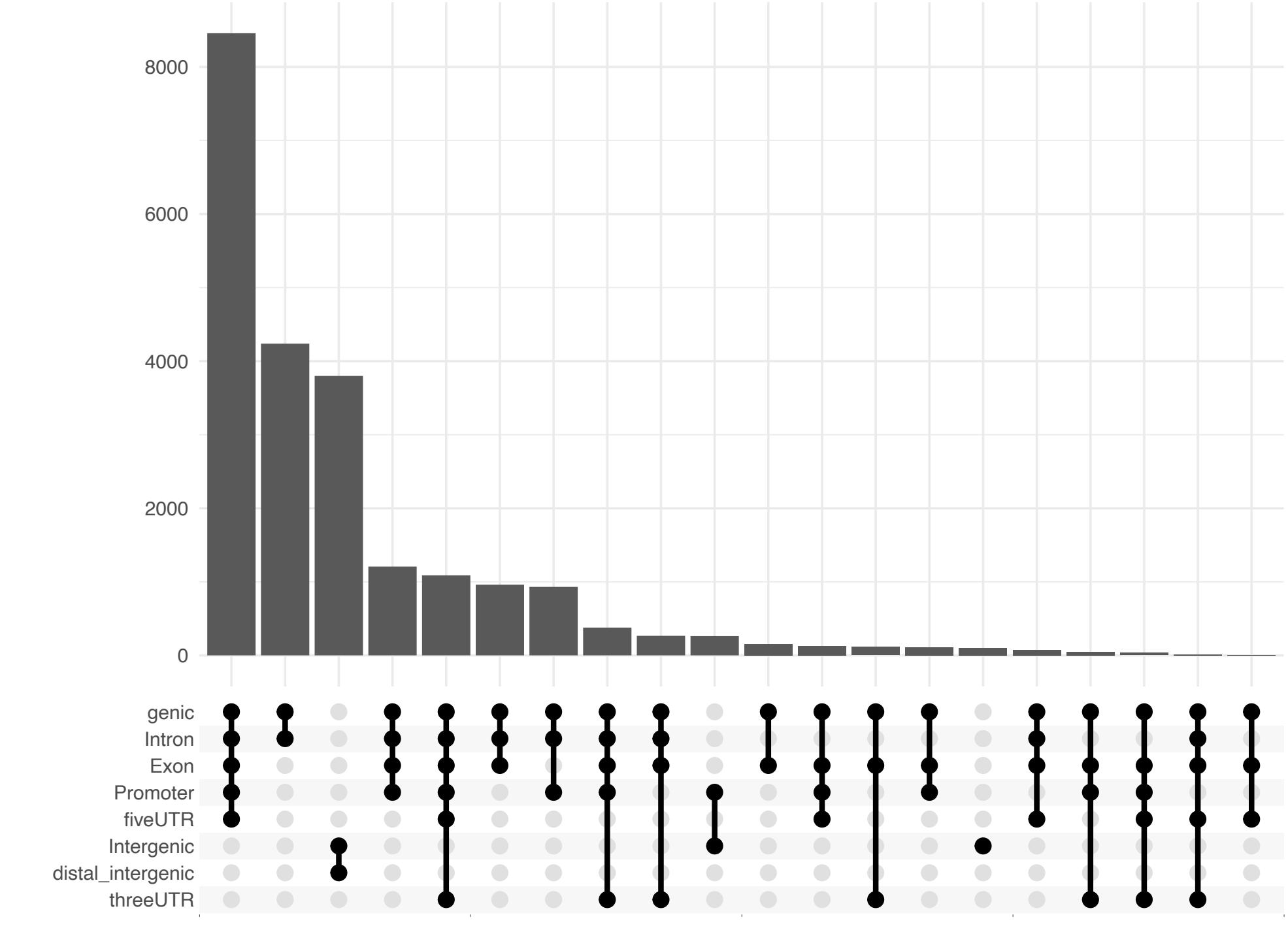
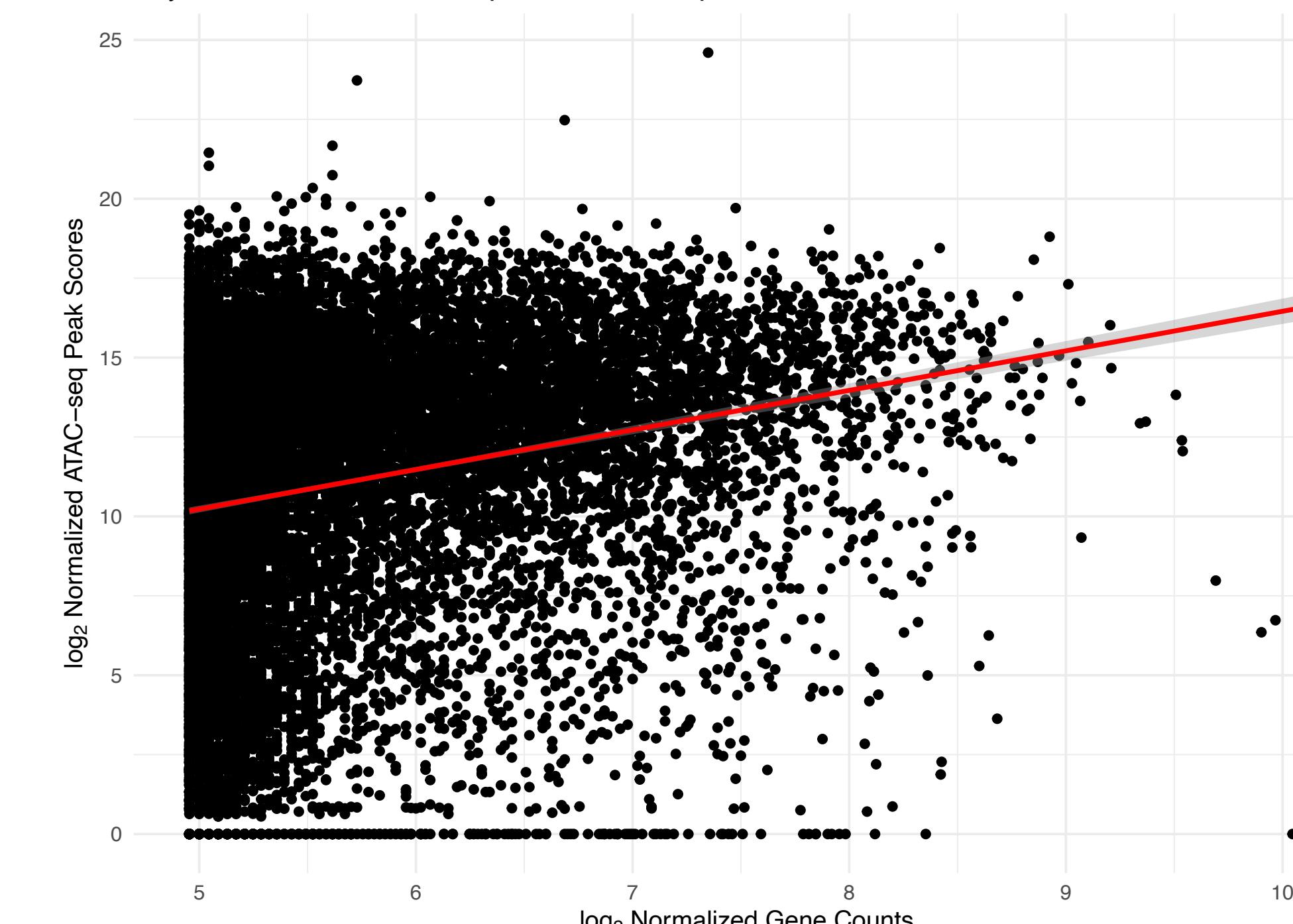


Figure 5: UpSet plot of annotation overlaps of the ATAC-seq peaks.

Finally, we determined if peak height (given by the peak scores obtained from HMMRATAC) significantly informed gene expression levels (given by normalized gene counts obtained from the DESeq2 R library) in our cohort of Puerto Rican MS patients. (Figure 6)

Adj R<sup>2</sup> = 0.049295 Intercept = 4.0023 Slope = 1.2456 P = 1.206e-147



## CONCLUSIONS AND FUTURE DIRECTIONS

- Our integrated Multi-Omics approach provides an overview of the molecular landscape of MS in Puerto Rican patients
- We found a significant positive relationship between ATAC-seq peak height and the expression levels of the corresponding gene
- Peak annotation allowed us to discover which genomic elements overlap our ATAC-seq peaks, a factor which will inform future analyses
- Perform De novo motif enrichment for ATAC-seq peaks located in promoter regions
- Further analysis of cis-regulatory elements (such as enhancers)
- Construct gene regulatory networks for genes of interest in MS

## Contact Us

Sebastián Cruz-González: sxc1506@case.edu

William S. Bush: wsb36@case.edu

## Acknowledgements

This work was funded by NIH Grants from the National Institute of Neurological Disorders and Stroke R01 NS096212 (McCauley) and the National Institute of General Medical Sciences R25 GM075207 (MacDonald).

## References

- Höglund, R. A., & Maghazachi, A. A. (2014). Multiple sclerosis and the role of immune cells. *World Journal of Experimental Medicine*, 4(3), 27.
- Negrotto, L., & Correale, J. (2018). Evolution of multiple sclerosis prevalence and phenotype in Latin America. *Multiple Sclerosis and Related Disorders*, 22, 97-102.
- Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(9), 1655-1664.
- Tarbell, E. D., & Liu, T. (2018). HMMRATAC: A Hidden Markov Model for ATAC-seq.
- Yu, G., Wang, L., & He, Q. (2015). ChIPseeker: An R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*, 31(14), 2382-2383.
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12).