Identifying Differentially Expressed Genes Shared Between Progressive Types of Multiple Sclerosis to Uncover Drivers of Disease Progression

Saad Khalid A Aldawsari

Supervised by: Hugo Tavares, Andries J. van Tonder

Project objectives and goals:

The aim of this project is to investigate the molecular mechanisms driving disease progression in multiple sclerosis (MS) by analyzing gene expression profiles across different MS subtypes. Specifically, the project focuses on identifying differentially expressed genes (DEGs) shared between Primary Progressive MS (PPMS) and Secondary Progressive MS (SPMS), which are associated with irreversible disability and limited therapeutic options.

The key goals include:

- Identifying shared upregulated and downregulated genes unique to progressive MS.
- Highlighting potential regulatory pathways, including coding and non-coding transcripts, that may contribute to disease progression.
- Providing a foundation for future studies on potential drivers for disease progression.

Methodology and approach:

Raw RNA-seq data comprising post-mortem brain tissue from non-MS controls, relapsing–remitting MS (RRMS), primary progressive MS (PPMS) and secondary progressive MS (SPMS) were obtained from the NCBI Gene Expression Omnibus (GEO accession: GSE290552) using **nf-core/fetchngs**

pipeline and processed through **nf-core/rnaseq**. Preprocessing involved read quality assessment, adapter trimming, and alignment to the reference human genome, followed by transcript quantification. Gene-level counts were generated using **tximport**, and subsequent normalization was carried out in **DESeq2** to correct for differences in sequencing depth and library size. Lowly expressed genes (less than 5 reads) were filtered to improve the statistical power. Differential expression across MS subtypes was assessed using **DESeq2** with a likelihood ratio test, and log2 fold changes were stabilized via ashr shrinkage. Gene annotation was retrieved from **AnnotationHub** and **biomaRt**, and functional enrichment analysis was performed with **clusterProfiler** to identify overrepresented Gene Ontology categories. Data visualization, including principal component analysis, heatmaps, and Venn diagrams, was conducted in R using **ggplot2** and **pheatmap** to assess variance, expression patterns, and genes uniquely shared between progressive MS subtypes.

Results

Table 1. Differentially Expressed Genes Shared only between the Progressive Types of Multiple Sclerosis

upregulated genes		downregulated genes	
Ensembl gene id	**gene name**	**Ensembl gene id**	**gene name**
ENSG00000157005	SST	ENSG00000138722	MMRN1
ENSG00000163359	COL6A3	ENSG00000143951	WDPCP
ENSG00000169914	OTUD3	ENSG00000173714	WFIKKN2
ENSG00000174576	NPAS4	ENSG00000205038	PKHD1L1
ENSG00000181045	SLC26A11	**ENSG00000240497**	
ENSG00000181773	GPR3	**ENSG00000270164**	**LINC01480**
ENSG00000256569		**ENSG00000288696**	
ENSG00000258791	**LINC00520**	**ENSG00000308230**	

Upregulated genes in progressive MS:

- **SST, NPAS4, GPR3:** Form a coordinated axis in inhibitory interneurons. SST reduces cortical hyperexcitability, NPAS4 drives GPR3 expression, and GPR3 reinforces inhibitory feedback, which might indicate an effort in stabilizing circuits undergoing demyelination (1-4).

- **COL6A3**: Encodes collagen VI α3 chain; have been found in experimental studies to protect against age-related neurodegeneration (5).

Downregulated genes in progressive MS:

- **WDPCP:** it regulates ciliogenesis and cell polarity (6), Although not directly studied in oligodendrocyte progenitor cells (OPCs), proper primary cilia are essential for OPC proliferation and differentiation. Therefore, reduced WDPCP expression may indirectly impair OPC function, decrease myelination, contributing to the progression of the disease **(7).**

Differentially expressed long non-coding RNAs (lncRNAs):

- Previous studies have found that lncRNAs are differentially expressed in MS. They play critical roles in normal neural development and immune regulation, so their dysregulation may contribute to neurodegeneration and immune dysfunction (8). However, their functional role in MS remains largely understudied.

Challenges and obstacles encountered:

- the **small sample size** (n = 16), which limits statistical power and generalizability of findings.
- **The use of post-mortem tissue introduces variability** due to differences in post-mortem interval, and potential confounding factors from patient comorbidities.
- **Annotation of long non-coding RNAs (lncRNAs)** and interpretation of their functional relevance in MS also remain limited, constraining biological inference for these transcripts.
-

Next steps and future work:

Future work will focus on validating the differentially expressed genes and the SST–NPAS4–GPR3 axis in independent cohorts and experimental models. Integration with single-cell transcriptomic data could provide cell-

type-specific insights, particularly in oligodendrocyte progenitor cells and inhibitory interneurons. Moreover, given their substantial representation among the identified DEGs, further studies are needed to explore the functional roles of long non-coding RNAs in the context of multiple sclerosis progression.

Resources and tools used:

Tools used:

1. **nf-core pipelines: nc-core/fetchngs** for downloading raw RNA-seq data from GEO **& nf-core/rnaseq** for preprocessing, including quality control, adapter trimming, and alignment.
2. **R Packages:**
   - **Tximport**: for importing transcript-level quantifications and summarizing to gene-level counts
   - **DESeq2**: for normalization, differential expression analysis, and log2 fold change shrinkage.
   - **AnnotationHub** and **biomaRt** for gene annotation.
   - **ggplot2** and **pheatmap** for visualization, including PCA, heatmaps, and venn diagrams.

References:

1. GeneCards – The Human Gene Database. (2025). SST Gene. Retrieved August 21, 2025, from https://www.genecards.org/cgi-bin/carddisp.pl?gene=SST
2. GeneCards – The Human Gene Database. (2025). NPAS4 Gene. Retrieved August 21, 2025, from https://www.genecards.org/cgi-bin/carddisp.pl?gene=NPAS4
3. Fu, J., Guo, O., Zhen, Z., & Zhen, J. (2020). Essential functions of the transcription factor NPAS4 in neural circuit development, plasticity, and

diseases. Frontiers in Neuroscience, 14. https://doi.org/10.3389/fnins.2020.603373

4. GeneCards – The Human Gene Database. (2025). GPR3 Gene. Retrieved August 21, 2025, from https://www.genecards.org/cgi-bin/carddisp.pl?gene=GPR3
5. Gregorio, I., Braghetta, P., Bonaldo, P., & Cescon, M. (2018). Collagen VI in healthy and diseased nervous system. Disease Models & Mechanisms. https://doi.org/10.1242/dmm.032946
6. GeneCards – The Human Gene Database. (2025). WDPCP Gene. Retrieved August 21, 2025, from https://www.genecards.org/cgi-bin/carddisp.pl?gene=WDPCP
7. Ki, S. M., Jeong, H. S., & Lee, J. E. (2021). Primary Cilia in Glial Cells: An Oasis in the Journey to Overcoming Neurodegenerative Diseases. Frontiers in Neuroscience, 15. https://doi.org/10.3389/fnins.2021.736888
8. Jalaiei, A., Asadi, M. R., Sabaie, H., Dehghani, H., Gharesouran, J., Hussen, B. M., Taheri, M., Ghafouri-Fard, S., & Rezazadeh, M. (2021). Long Non-Coding RNAs, Novel offenders or guardians in Multiple sclerosis: A scoping review. Frontiers in Immunology, 12. https://doi.org/10.3389/fimmu.2021.774002

Source code:

https://github.com/saad-aldawsari/KAUST-Cambridge-BI-Final-Project