Clinical Medicine and Health Research Journal (CMHRJ)
Volume 03, Issue 06, November - December, 2023
Page No. 703-708,

Research Article

Using Gene Expression Profiling, İdentify Genetic Changes İn Multiple Sclerosis (MS) And Determine Molecular Pathways.

Zeynep Kucukakcali, Ipek Balikci Cicek

Department of Biostatistics and Medical Informatics, Inonu University Faculty of Medicine, 44280, Malatya, Turkey

Received: 20 October, 2023 Accepted: 25 November, 2023 Published: 02 December 2023

Abstract:

Aim: In this study, an open-access dataset of RNAseq data from proggressive multiple sclerosis (MS) samples and matched normal-appearing white matter (NAWM) samples was used to identify genetic alterations that play a role in the development of MS, the prevalence is on the rise in both industrialized and emerging nations.

Material and methods: The dataset utilized in this investigation comprises nine samples of lesions and corresponding samples of NAWM tissue. The gene expression analysis was conducted on the dataset using the limma package in the R programming language. The clusterProfiler software was utilized to do enrichment analysis.

Results: 493 genes showed different up- or down-regulation in lesion tissues compared to NAWM tissues. According to the results of the enrichment analysis with these genes, the processes that play a role in progressive MS cancer are intracelluler anatomical structure, membrane bounded organelle, non-membrane bounded organelle, intracelluler non-membrane bounded organelle.

Conclusion: As a result, it can be said that the genetic structure of progressive MS tissues (lesion) changes compared to NAWM tissues, with the information obtained from gene expression analysis and enrichment analysis. It might be argued that the therapeutic efficacy of the disease can be enhanced by the comprehensive and diverse analysis of biomarkers.

Keywords: Multiple sclerosis (MS), genetic alteration, gene expression analysis, bioinformatics

Introduction

Multiple sclerosis (MS) is a prevalent debilitating disorder that commonly affects young adults who do not have a prior history of trauma. The prevalence of MS is on the rise in both industrialized and emerging nations, however the precise underlying factors contributing to this phenomenon remain uncertain (1, 2). MS is a multifaceted medical disorder characterized by a combination of genetic variables that marginally elevate the likelihood of disease development, alongside various established environmental factors like vitamin D or ultraviolet B light (UVB) exposure, infection with the Epstein-Barr virus (EBV), obesity, and smoking. MS has historically been seen as an autoimmune disorder characterized by the attack of certain organs by T-cells. However, the efficacy of treatments targeting B-cells has presented a challenge to this conventional idea, as it contradicts the prevailing notion of T-cell-mediated autoimmunity in MS (3). MS is commonly conceptualized as a biphasic disorder, characterized by an initial period of inflammation that gives rise to relapsing-remitting disease. This is subsequently followed by a later phase of neurodegeneration, which leads to a progressive form of the disease that can be categorized as either secondary progressive MS or primary progressive MS (4, 5).

MS is a multifaceted disease that emerges from the intricate interplay between diverse genetic variations and extrinsic factors, which collectively contribute to its susceptibility. The interplay of hereditary and environmental variables synergistically triggers the onset of this illness. It is noteworthy that extensive analyses derived from observational research underscore the significance of specific environmental risk factors associated with MS, including obesity, vitamin D deficiency, Epstein-Barr virus infection, and smoking behavior. Modifying environmental and lifestyle factors has the potential to prevent or delay the onset of multiple sclerosis (MS), therefore highlighting the importance of establishing definitive causal links between these factors and the occurrence of MS (6, 7). In order to mitigate the occurrence of MS and its detrimental effects on individuals and healthcare systems, it is imperative to attain a comprehensive comprehension of the intricate interrelationships among many elements that exert influence on the development and progression of this condition. The successful completion of this task necessitates a collective endeavor including various scientific disciplines, including genetics, immunology, epidemiology, and molecular biology. By adopting this approach, it is possible to acquire a comprehensive comprehension of the etiology of MS and formulate targeted interventions aimed at reducing the likelihood of developing the condition. The pursuit of this objective is of utmost importance in light of the scientific progress and advancements in public health. It is imperative to decipher the intricate interplay between genetic and environmental elements that contribute to an increased susceptibility to MS (8, 9). The etiology of multiple sclerosis (MS) and, more precisely, the processes by which the disease progresses are intricate and not yet fully comprehended. However, a growing body of research emphasizes the

significant contribution of gene dysregulation within the central nervous system to this process (10).

The objective of this investigation was to ascertain the specific genes responsible for multiple sclerosis (MS) and the corresponding metabolic pathways connected with these genes. This was achieved by utilizing publicly available RNA-sequencing data acquired from samples of lesions and normal-appearing white matter (NAWM) tissues. The primary aim was to elucidate the genetic mechanisms underlying the progression of MS.

Material and Methods

Dataset features

The open-access data set used in the study was obtained from The National Center for Biotechnology Information (NCBI) with the "GSE224377" GEO data code. The dataset contains transcriptomic profiles from normal-appearing white matter (NAWM) paired with lesions from 9 progressive MS patients.

Bioinformatics, gene expression, and enrichment analysis

Bioinformatics encompasses the comprehensive management of data, including its collection, storage, organization, archiving, analysis, and presentation. This field is grounded on the theoretical and practical foundations of various disciplines, such as biology, medicine, behavioral sciences, and health sciences. Moreover, the field of study is centered on the examination and advancement of computational tools and methods with the aim of expanding the utilization and manipulation of data generated from research or the implementation of established processes. Acquired through the process of scholarly investigation or the use of established procedures. Bioinformatic analyses are conducted by choosing a suitable database and application that facilitate the execution of bioinformatic analysis in accordance with the specific biological query, molecule, or structure under investigation. The data and findings obtained from the studies are synthesized and examined systematically in the context of existing literature on the subject (11).

Alterations in the physiological state of an organism or its constituent cells are invariably accompanied by corresponding modifications in the gene expression profile. Consequently, the measurement of gene expression assumes paramount significance in numerous domains of biological investigation. The DNA microarray technique, which is currently under development, is employed for the investigation of gene expression. This is achieved by the process of hybridization, where mRNA molecules are bound to a densely populated array of immobilized target sequences. Each of these target sequences corresponds to a unique gene. The impact of chemical substances on the regulation of gene expression can provide insights into both functional and toxicological attributes. Investigations including the analysis of clinical samples, encompassing both individuals with normal health conditions and those afflicted with illnesses, have the potential to unveil previously undiscovered biomarkers (12).

While RNA expression analysis have been a commonly employed technique in several studies, the task of extracting meaningful biological insights from the collected information and establishing correlations between the identified genes and disease states continues to pose a significant difficulty. The abundance of genes identified by expression analysis and their differential expression (either upregulated or downregulated) in the diseased state leads to a limited molecular association with the disease. Hence, it is imperative to group genes exhibiting shared properties in order to establish connections. Enrichment analysis is a method that has been derived directly from this particular methodology. Enrichment analysis pertains to the examination of genes and specifically emphasizes gene clusters, which refer to groupings of genes that exhibit shared biological roles, chromosomal position, or patterns. Therefore, a more comprehensive comprehension of the molecular mechanisms underlying the disease can be achieved by elucidating the structural arrangements resulting from the collective actions of the genes (13).

Bioinformatics analysis

In this study, an inquiry was conducted wherein gene expression analysis were performed on mRNA data obtained from primary prostate cancer samples and corresponding normal tissues. The inquiry utilized the limma package, which is available in the R programming language and enables expression analysis (14). The Limma software package, also known as Linear Models for Microarray Analysis, is a computational tool designed to analyze gene expression microarray data. Its primary objective is to employ linear models to evaluate specific experimental conditions and identify differential expression patterns. The capabilities of the packet can be utilized across several gene expression techniques, including microarrays, RNA-seq, and quantitative PCR. The findings are depicted through a tabular representation of genes arranged in a manner that reflects their significance, alongside a graphical visualization that highlights genes exhibiting differential expression. The table of findings presents the corrected P-values and log2-fold change (log2FC) values, where genes with the lowest p-values are considered to be the most reliable. Genes exhibiting up-regulation were identified using a criterion of Log2FC>1 and $p<0.05$, while genes exhibiting down-regulation were identified using a criterion of Log2FC-1 and $p<0.05$. The volcano plot was employed in the research to depict genes that exhibited differential expression. The volcano plot displays genes that exhibit overexpression in malignant tissue compared to normal tissue in the color red, genes that display downregulation in the color blue, and genes that do not demonstrate differential expression in either tissue are shown in the color black. The enrichment analysis was conducted using the clusterProfiler package in the R programming language, utilizing the results derived from the gene expression study (15). The findings are shown using dotplot, enrichment map, and category netplot, which allow for the visualization of shared functional processes among genes.

Results

The expression data used in the present study were obtained

from lesion and NAWM tissues from 9 progressive MS patients. Scatter plots of lesion and NAWM tissue from 9 progressive MS patients used in the study are given in Figure 1 and Figure 2.

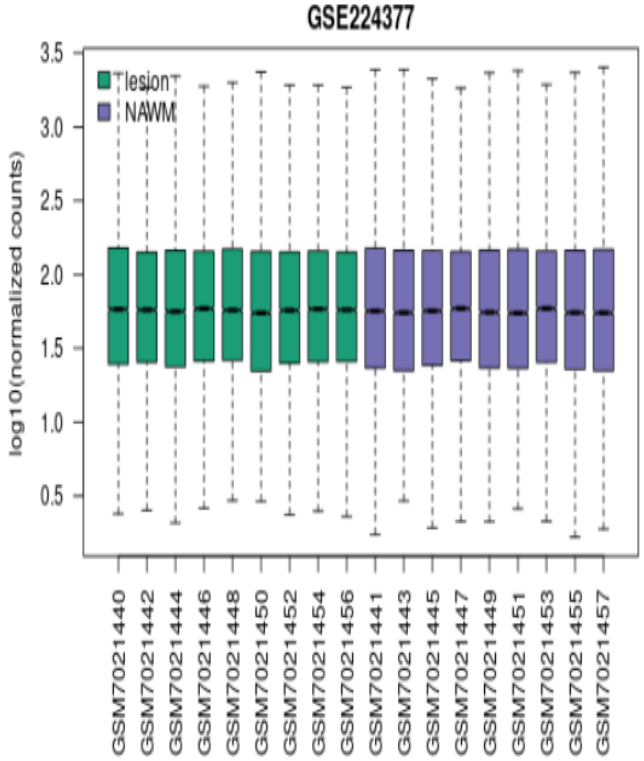

Figure 1: Distribution plot of the samples

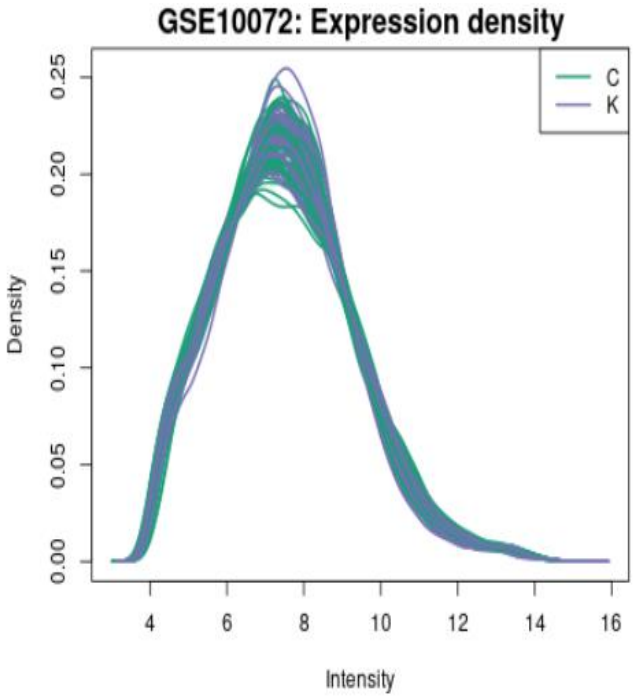

Figure 2: The expression density graph of the samples.

The UMAP graph is shown here in Figure 3, and it provides a graphical representation of the interrelationships between the samples. The graph depicts how samples that have been observed to have similar characteristics have been grouped together. On the graph, lesion samples are represented by green dots, while NAWM samples are represented by purple dots.

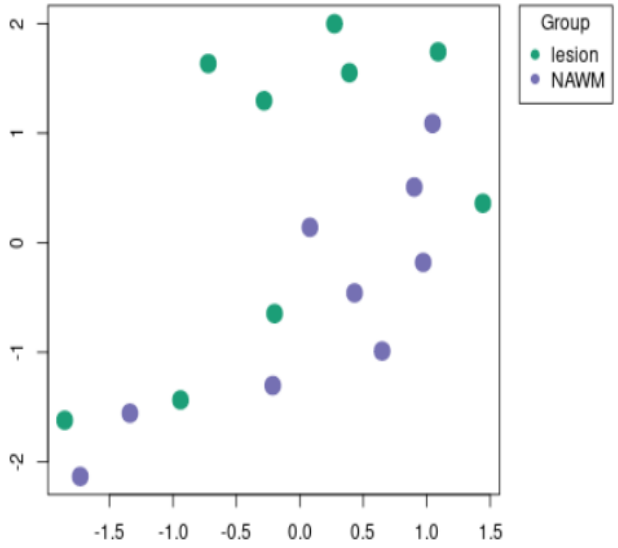

Figure 3: UMAP plot of the samples (Green Dots: Tumor tissues, Purple dots: non-tumor tissues).

Gene expression analysis revealed a total of 493 genes whose expression levels differed significantly (|log2FC| > 1.0, p 0.05) between the two groups. The top 10 genes whose expression was either up- or down-regulated between the two groups are listed in Tables 1 and 2, respectively.

Table 1: Transcripts found to be up-regulated in lesion compared to NAWM

ID	Adj.P Val	P Value	Log2FC	Gene name
200162	0,00377	5,22E-40	2,965016	SPAG17
1770	0,01503	3,04E-23	2,714513	DNAH9
126820	0,02455	2,47E-16	2,453199	DNAI3
4057	0,19226	9,83E-13	2,409883	LTF
100652824	0,01503	8,05E-25	2,356471	KIAA2012
342979	0,04013	1,52E-21	2,260114	PALM3
83657	0,10746	2,41E-26	2,24444	DYNLRB2
51364	0,04032	3,51E-23	2,242466	ZMYND10
3502	0,28119	1,15E-18	2,239731	IGHG3
146802	0,07622	1,57E-23	2,236028	SLC47A2

Table 2: Transcripts found to be down-regulated in lesion compared to NAWM

ID	Adj.P Val	P Value	Log2FC	Gene name
105375483	0,02778	5,37E-05	-1,72621	LOC105375483
100507336	0,17629	1,04E-02	-1,60446	LOC100507336
105375544	-	5,04E-04	-1,57644	LOC105375544
9200	-	1,95E-03	-1,54751	HACD1
101929637	0,12972	4,45E-03	-1,47907	NTM-AS1
101929467	0,18494	1,19E-02	-1,38144	LURAP1L-AS1
107984040	0,22059	1,82E-02	-1,37031	LOC107984040
286343	0,07272	1,00E-03	-1,34815	LURAP1L
284835	-	4,17E-02	-1,33359	LINC00323
100287225	0,12121	3,82E-03	-1,33345	LINC01630

Differentially expressed genes between groups can be shown using a volcano plot, which is depicted in Figure 4.

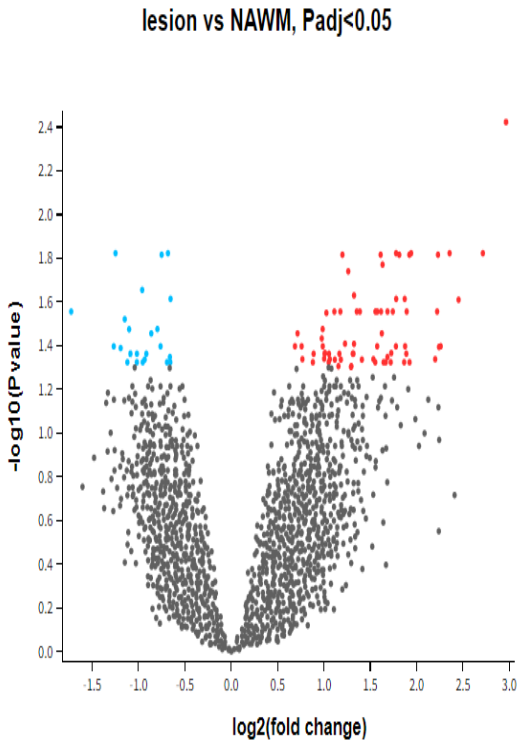

Figure 4: Volcano plot for transcripts in lesion and NAWM (Transcripts with an up in expression were down by red dots, those with a reduction in expression by blue dots, and those with stable expression by black dots)

Figures 5 and 6 depict the metabolic pathways linked to the illness based on the enrichment analysis results.

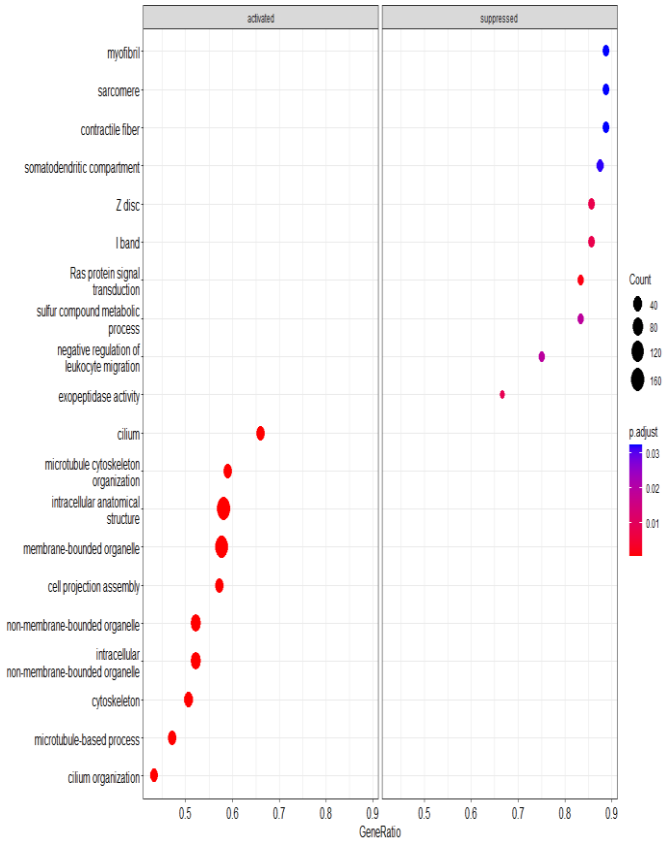

Figure5: the results of the enrichment analysis with dotplot

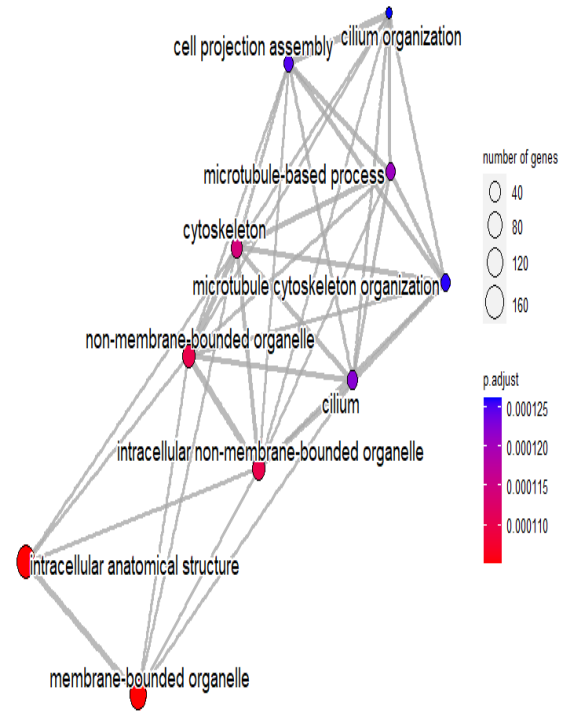

Figure 2: Enrichment map

Figure 7 is a diagram that shows the connections between genes and metabolic pathways in great detail. The processes and genes that may be most influenced by the illness state are depicted in this diagram.

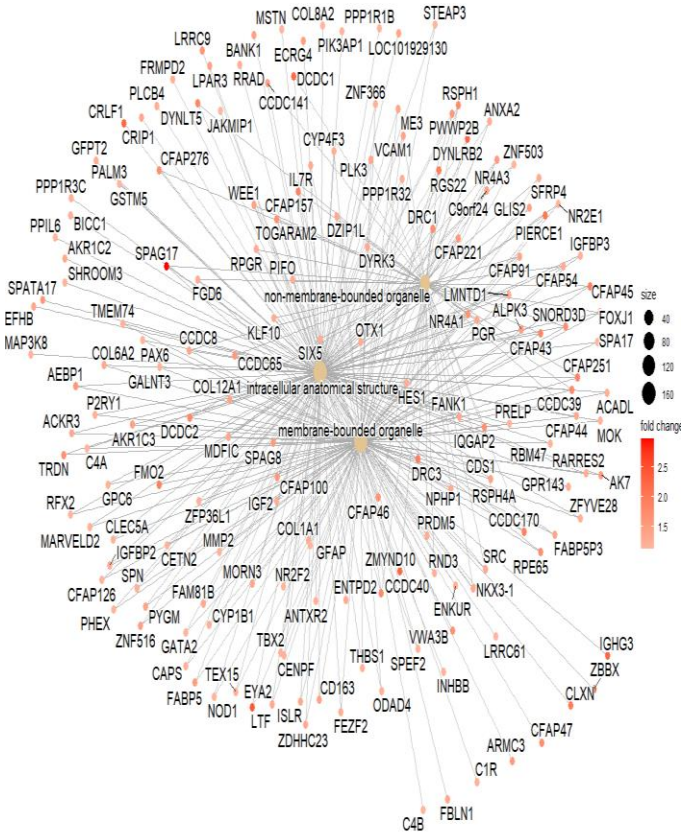

Figure 4: Category netplot

Discussion

MS is a pathological condition characterized by inflammation and neurodegeneration, resulting in the demyelination of the

central nervous system (CNS). It is widely believed to have an autoimmune etiology and typically manifests during early adulthood. The genesis of the disease remains poorly comprehended, with current understanding suggesting it to be a multifaceted condition arising from intricate interplay between genetic predisposition and environmental influences, some of which may be subject to modification. Enhancing our comprehension of these variables has the potential to yield novel and more efficacious strategies for patient counseling, and conceivably, the prevention and treatment of the ailment (16).

Over the course of the previous five years, there has been a notable convergence of information pertaining to environmental risk factors, as well as the discovery of novel genetic components. Additionally, there has been an enhanced recognition of the significance of gene-environment interactions, resulting in a heightened level of agreement among the academic community. The acquisition of this novel information has the potential to offer insights into the underlying biological mechanisms and strategies for preventing diseases in both general populations and individuals who are particularly susceptible. The examination of elements linked to the advancement and severity of diseases is equally crucial in order to direct the discovery of pathological mechanisms and, potentially, suggest lifestyle interventions to decelerate illness progression (16-18).

The utilization of genetic data has significance in the characterization of pathogenetic mechanisms and the elucidation of the intricate nature of MS onset, particularly within the framework of lifestyle and environmental factors. The integration of genetic and environmental factors has yielded comprehensive analyses, revealing that a significant amount of MS risk can be accounted for by presently identified risk factors. This information will facilitate the exploration of novel therapeutic methods in light of recent research findings. The prognosis of the condition and the techniques for therapy will be accurately determined. To attain these objectives, scholars have placed significant attention on genetic investigations pertaining to the impact of genetic alterations on the progression of MS.

Therefore, this study aims to identify the processes and genes involved in the development of MS using an open-access dataset of RNAseq data. For this purpose, the dataset used in the study consists of transcriptomic profiles from NAWM paired with lesions from 9 progressive MS patients. According to the results of the bioinformatics analysis, 493 genes showed different regulation. The SPAG17 gene showed 7.78 fold up-regulation in lesion samples compared to NAWM samples. Likewise, the DNAH9, DNAI3, LTF, KIAA2012, PALM3, DYNLRB2, ZMYND10, IGHG3, and SLC47A2 genes had up-regulated gene expression of 6.54, 5.46, 5.27, 5.09, 4.78, 4.72, 4.72, 4.69, and 4.69 fold, respectively. Moreover, the LOC105375483 gene showed 3.29-fold down-regulation in lesion samples compared to NAWM samples. Likewise, the SLOC100507336, LOC105375544, HACD1, NTM-AS1, LURAP1L-AS1, LOC107984040, LURAP1L, LINC00323, and LINC01630 genes had down-regulated gene expression of 3.03, 2.96, 2.90, 2.77, 2.60, 2.58, 2.53, 2.51, and 2.51 fold, respectively.

In a study, it was determined that SPAG7 plays a role in MS (19). In a different study, the relationship between SPAG17 and MS was revealed (20).

According to the results of the enrichment analysis performed by considering genes showing different regulation, intracelluler anatomical structure, membrane bounded organelle, non-membrane bounded organelle, intracelluler non-membrane bounded organelle processes were obtained as important processes.

Consequently, it can be asserted that the genetic composition of progressive multiple sclerosis (MS) undergoes alterations when considering the insights derived from gene expression analysis and enrichment analysis. Consequently, conducting comprehensive and diverse genetic investigations that take into account these changes may potentially enhance the therapeutic effectiveness of the disease. The utilization of biomarkers in the diagnostic process of multiple sclerosis (MS) enables the development of suitable treatment strategies and facilitates early intervention to mitigate disease progression. The novel approach to oncological treatment selection has demonstrated the potential to effectively target the disease and subsequently lower fatality rates.

References

1. Dimitrov LG, Turner B. What's new in multiple sclerosis? The British journal of general practice : the journal of the Royal College of General Practitioners. 2014;64(629):612-3.
2. Nicholas R, Rashid W. Multiple sclerosis. American family physician. 2013;87(10):712.
3. Waubant E, Lucas R, Mowry E, Graves J, Olsson T, Alfredsson L, et al. Environmental and genetic risk factors for MS: an integrated review. Annals of clinical and translational neurology. 2019;6(9):1905-22.
4. Dobson R, Giovannoni G. Multiple sclerosis–a review. European journal of neurology. 2019;26(1):27-40.
5. Cree BAC, Arnold DL, Chataway J, Chitnis T, Fox RJ, Pozo Ramajo A, et al. Secondary Progressive Multiple Sclerosis: New Insights. Neurology. 2021;97(8):378-88.
6. Hone L, Jacobs BM, Marshall C, Giovannoni G, Noyce A, Dobson R. Age-specific effects of childhood body mass index on multiple sclerosis risk. Journal of Neurology. 2022;269(9):5052-60.
7. Dyment DA, Ebers GC, Sadovnick AD. Genetics of multiple sclerosis. The Lancet Neurology. 2004;3(2):104-10.
8. Vandebergh M, Degryse N, Dubois B, Goris A. Environmental risk factors in multiple sclerosis: Bridging Mendelian randomization and observational studies. Journal of Neurology. 2022;269(8):4565-74.
9. Price E, Lucas R, Lane J. Experiences of healthcare for people living with multiple sclerosis and their healthcare professionals. Health expectations : an international journal of public participation in health care and health policy. 2021;24(6):2047-56.
10. Yang X, Wu Y, Zhang B, Ni B. Noncoding RNAs in multiple sclerosis. Clinical epigenetics. 2018;10(1):1-12.
11. Akalın PK. Introduction to bioinformatics. Molecular nutrition & food research. 2006;50(7):610-9.

12. van Hal NL, Vorst O, van Houwelingen AM, Kok EJ, Peijnenburg A, Aharoni A, et al. The application of DNA microarrays in gene expression analysis. Journal of biotechnology. 2000;78(3):271-80.
13. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(43):15545-50.
14. Smyth GK. Limma: linear models for microarray data. Bioinformatics and computational biology solutions using R and Bioconductor: Springer; 2005. p. 397-420.
15. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. The innovation. 2021;2(3).
16. Amato MP, Derfuss T, Hemmer B, Liblau R, Montalban X, Soelberg Sørensen P, et al. Environmental modifiable risk factors for multiple sclerosis: Report from the 2016 ECTRIMS focused workshop. Multiple sclerosis (Houndmills, Basingstoke, England). 2018;24(5):590-603.
17. Low-Frequency and Rare-Coding Variation Contributes to Multiple Sclerosis Risk. Cell. 2018;175(6):1679-87.e7.
18. Olsson T, Barcellos LF, Alfredsson L. Interactions between genetic, lifestyle and environmental risk factors for multiple sclerosis. Nature reviews Neurology. 2017;13(1):25-36.
19. Elkjaer ML, Frisch T, Reynolds R, Kacprowski T, Burton M, Kruse TA, et al. Molecular signature of different lesion types in the brain white matter of patients with progressive multiple sclerosis. Acta neuropathologica communications. 2019;7(1):1-17.
20. Lerma-Martin C, Badia-i-Mompel P, Ramirez Flores RO, Sekol P, Hofmann A, Thäwel T, et al. Spatial cell type mapping of multiple sclerosis lesions. bioRxiv. 2022:2022.11. 03.514906.]