Gene Networks Reveal *LASP1*, *TUBA1C*, and *S100A6* as Likely Drivers of Multiple Sclerosis (MS)

Nafiseh Karimi

University of Isfahan

Majid Motovalli Bashi (✉ mbashi@sci.ui.ac.ir)

University of Isfahan

Mostafa Ghaderi-Zefrehei

Yasouj University

Research

Keywords: Multiple Sclerosis (MS), Bayesian Network, Transcriptome, qPCR

Posted Date: October 28th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1011436/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: Multiple sclerosis (MS), a non-contagious and chronic disease of the central nervous system, is an unpredictable and indirectly inherited disease affecting different people in different ways. Using Omics platforms data i.e. genomics, transcriptomics, proteomics, epigenomics, interactomics, and metabolomics it is now possible to construct sound systems biology models to extract full knowledge of the MS and path the way to likely construct personalized and therapeutic tools.

Methods: In this study, we learned many Bayesian Networks in order to find the transcriptional gene regulation networks that drive MS disease. We used a set of BN algorithms using the R add-on package bnlearn. The BN results underwent further downstream analysis and were validated using a wide range of Cytoscape algorithms, web based computational tools and qPCR amplification of blood samples from 56 MS patients and 44 healthy controls. The results were semantically integrated to come up with improved understanding of the complex molecular architecture underlying MS, distinguishing distinct metabolic pathways and providing a valuable foundation for the discovery of driven genes and possibly new treatments.

Results: Results show that the LASP1, TUBA1C, and S100A6 genes were most significant and likely playing a biological role in MS development. Results from qPCR showed a significant increase ($P < 0.05$) in LASP1 and S100A6 gene expression levels in MS patients compared to controls. However, a significant down regulation of TUBA1C was observed in the same comparison.

Conclusion: This study provides potential diagnostic and therapeutic biomarkers for increasing understanding of gene regulation underlying MS.

Introduction

Miller DH, 2008 #106; Miller DH, 2008 #106; Multiple sclerosis (MS) is a multifocal inflammatory autoimmune disease that severely affects the white matter of the CNS and leads to progressive neuronal damage in genetically sensitive hosts [1]. The integration of information gleaned from a variety of resources e.g. transcriptomics, genomics, proteomics and patient clinical data could boost our understanding of the mechanism(s) underpinning this disease [2]. In this regard, we can explore the signaling pathways involved in MS [3], apply logical networks to model signaling pathways in MS [4] and use networks to combine information on transcriptome-interactome data from MS studies [5]. We can also apply theory of biochemical systems for improving therapeutic drugs in re-myelination [6], create molecular networks based on transcription factors and genes expressed in mononuclear cells in MS patients [7], and design reactive networks between distinct miRNA and target genes in T cells [7] - all of which will help explain the molecular mechanisms involved in MS disease [8]. **Additional Table 1** shows some examples of network based studies used with different MS biological data. As one can see, Bayesian Network (BN) modeling paradigms have rarely been applied in this setting. BN uses probability theory to reason under uncertainty. BN as a graphical scheme (directed acyclic graph) consists of a qualitative part (structural model) and a quantitative part (local probability distributions), which allows for a different kind of probabilistic inference, and quantitatively measures even the smallest impact of a variable or set of variables on others. This sort of modeling is of great importance in transcriptomic studies, since it can reveal both qualitative and quantitative elements of learned gene networks. BN has previously been used in several transcriptomic studies [9-11].

Many existing categories of gene networks identify groups of related genes as gene sets, making experimental follow-up a formidable task. With BN, it is possible to determine whether a gene is a driving source of changes in its gene network or not, since both in-degree and out-degrees of connectivity of each gene can be readily verified. The more out-degree each gene has, the higher likelihood it has of being a key regulator gene. This would be a crucial characteristic when

looking for potential drug targets. Today, MS research is increasingly data-driven - a trend that arguably shall continue at a much higher rate in times to come. To tackle these large amounts of heterogeneous data, and to derive insight into MS disease, many interdisciplinary scientists have turned to a variety of computational tools. In this study we aim to gain much better insight into the regulatory transcriptional gene network underlying MS using systems biology approaches in the context of BN, that may yield mechanistically interpretable results.

Materials And Methods

Network analysis

In this study the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>) was scanned using a combination of several simple key words, and resulting DNA microarray experiments related to MS, that fulfilled our criteria, were explored. In the end, based on our criteria for choosing a suitable GEO data set, the microarray series with accession number GSE17048 was downloaded from GEO using the GEOquery package [12]. This accession was seen to have the highest number of arrays per probe - a fact that would help minimize the rate of false positives while training the regulatory gene BN. The GSE17048 study represents mRNA expression for all known genes in whole blood from 144 individuals (99 MS patients and 45 age-matched controls). In order to remove noise from the data, probes with the highest variance were obtained and used as an input to train the gene regulatory BN using bnlearn, an R add-on package [13, 14]. The following codes were used to filter the probes with highest variances: `qt <- quantile( t(data1) ); probs = c(0.0002,0.99) ); rows1 <- apply( t(data1) , function(x) any( x < qt | x > qt ) ); data2 <- t(data1)[ rows1]`

We obtained the best fitted BN model on our data using Bayesian information criterion (BIC) and its adjacency matrix, with the help of the *Cytoscape*-based *aMatReader* plugin, with the *Cytoscape* [15] environment used for further downstream scrutiny.

Downstream analysis

This was accomplished with the following *Cytoscape* add-on packages: the *jActiveModules* to explore the concept of gene modules and find sub-networks [16]; *MCODE* to identify putative complexes by finding regions of significant local density [17]; *CytoHubba*, to explore the PPI network of hub genes using eleven different methods [18]; the *NetworkAnalyzer* to determine the hub genes, taking into account the degree of topological criteria (e.g. the number of nodes, edges, and connected components, along with the network diameter, radius, density, centralization, heterogeneity, and clustering coefficient., The characteristic path length, and the distributions of node degrees, neighborhood connectivities, average clustering coefficients, and shortest path lengths were also considered) [19]; *iRegulon* to detect targets / motifs / paths from a set of genes; and the *CyTargetLinker* to integrate regulatory reactions in network analysis. In addition, we used *Metascape* [20] to annotate the multiple gene lists from our study. Even though transcriptomic statistical analysis is generally based on probe level data, the probe names were converted to their corresponding gene names using g:Profiler (<https://biit.cs.ut.ee/gprofiler/gost>) to get better insight into the data. Results from the aforementioned software were combined. **Figure 1** shows the analysis pipeline used in this study.

Validation of *LASP1*, *TUBA1C*, and *S100A6* genes using Quantitative Real-Time PCR (QRT-PCR)

Following up on outputs from the bioinformatics analysis, validation of significant DEGs (*LASP1*, *TUBA1C* and *S100A6*) was carried out using QRT-PCR. A total of 100 whole blood samples (56 MS cases, mean age: 39.5 years and 44 controls, mean age: 39.5) were provided by Al-Zahra Hospital and a Multiple Sclerosis medical center in Isfahan province. Within the 56 MS cases, 44 patients were in RRMS (Relapsing-remitting MS) phase and 12 patients were in SPMS (Secondary progressive MS) phase.

RNA extraction

Total RNA was extracted from each sample according to the standard TRIzol protocol (Bio BASIC, Canada) according to manufacturer's instructions. RNA concentration and quality were determined using both Nanodrop (Thermo Scientific Tm Nano Drope One C model) and gel electrophoresis -existence of two sharp bands representing 18S and 28S ribosomal RNA on a 1% (w/v) ethidium bromide stained agarose gel during electrophoresis through TAE buffer (40 mM Tris-acetate, 1 mM EDTA, pH 8.0) at 100 V for 30 min. Those RNA samples with a [RNA integrity Number](#) (RIN) < 1.8 were excluded from further analysis. High quality extracted total RNA was stored at -70°C until cDNA synthesis.

cDNA synthesis

Initially, DNase I (Fermentase Cat # ENO 521) treatment was used to remove genomic DNA before cDNA synthesis. Next, cDNA synthesis was carried out using a commercial kit provided by Yektatajhiz Company (Cat No:YT4500) according to manufacturer's instructions. e.g. keeping samples on ice under sterile conditions at 70°C for 5 mins, 37°C for 60 mins, 70°C for 5 mins, and finally storing all synthesized cDNAs at -20°C.

Quantitative Real Time PCR Analysis

To enable the validation of our candidate genes (*LASP1*, *TUBA1C* and *S100A6*), SYBR Green -based QRT-PCR was performed using a LightCycler® 96 (BioRad, Germany). The sequence of all primers used are listed in **Table 1**. These were designed using the PRIMER3 program (<http://frodo.wi.mit.edu>). QRT-PCR reactions were performed in duplicate and the values of average cycle threshold (Ct) were determined for each sample. The conditions of QRT-PCR amplification were: 1 cycle at 95 °C for 2 min, 40 cycles at 95 °C for 50 s, 60 °C for 30 s. The human beta-actin gene (ACTB_HUMAN) was used as the internal control. Hence, all calculated concentrations are relative to the concentration of the standard, expressed in arbitrary units and the quantification cycle (Cq) values were automatically calculated with Rotor-Gene software version 6.1. The results were analyzed using the $2^{-\Delta\Delta C_t}$ method [21].

Results

The fundamental idea behind this analysis was to shed some light into gene-gene interactions underpinning MS disease with regard to cause and effect. The *bnlearn* package was used to find the gene regulatory network behind this disease [22]. The results of comparison of the network structures determined from various algorithms, including Hill Climbing, Tabu Search, Max-Min Hill Climbing, and Restricted Maximize algorithms with different scoring functions, are shown in **Table 2**. Some key properties of BN are fundamental in judging estimated results.

Figure 2 shows some properties of trained BN gene networks.

Topological parameters can characterize the location of genes in a gene network [23]. Using *NetworkAnalyzer* the following network topological parameters were calculated in our data: clustering coefficient (0.003), number of nodes (1707), connected components (857), network diameter (27), network radius (1), shortest paths (192463), characteristics path (8.313), the average number of neighbors (1.992), network density (0.0), isolated nodes (854), number of self-loops (0), multi-edge node pairs (0) and analysis times (1.467). The nature of probability distribution induced by a gene regulatory BN will allow diverse probabilistic gene queries to be answered in linear time. However, many structural BN parameters may be important. One of the key parameters in **Table 2** is the branching factor. This parameter plays a significant role in development of the gene network. Each Node (gene) will have its own branching measure, which will determine the outdegree of that gene. If the branching factor value is not uniform in the network, an average branching factor can generally be calculated. This value turned out to be different depending on which algorithm was used. Max-

Min Hill Climbing returned a higher average than Restricted Maximize. In terms of system level understanding of research, the higher the branching factor, the more frequently gene regulators can be identified in the network.

Biological networks have a modular architecture [24]. MCODE can find connected and dense areas of the gene network based on network topology measures. In our analysis, 12 different modules were detected using MCODE, among which, 7 modules had 3 nodes; 3 edges with different interaction modes; 3 modules had 6 nodes and 7 edges; 1 module 15 nodes and 17 edges and finally 1 module had 6 nodes and 6 edges (**Additional Fig 1**). The active subnetworks were obtained using jActiveModules. The jActiveModules created 5 modules, in which ILLMN_1742167 (*TUBA1C*), ILLMN_1665909 (*LASP1*) and ILLMN_1713636 (*S100A6*) were seen to be enriched in those modules. (**Figure 3** and **Additional Fig 2**). The number of modules detected by this method was different than those identified with the MCODE based method.

Figure 3 shows predicted modules in different modes of interaction. Module-level analysis explores the organization of biological systems and reconstructs module networks. **Figure 3** shows a module-level view of our gene regulatory BN network that denotes a high level representation of the regulatory machinery of the MS gene network topology. Dense module searching of two MS GWAS datasets identified several genes (*GRB2*, *HDAC1*, *IL2RA*, *JAK2*, *KEAP1*, *MAPK1*, *RELA* and *STAT3*). These genes are enriched for glial cell differentiation [25]. CytoHubba provides a user-friendly interface for discovering important nodes in biological networks [26]. Cytohubba considers the shortest path between groups of nodes. Among the 11 proposed algorithms, MCC fitted better than the others. In **Figure 4** and **Table 3** we present the top 10 identified probes.

Many genes e.g. *TUBA1C*, *LASP1* and *S100A6* indicated in **Figure 4** are close to being hub genes, and were indeed identified as hub genes by other algorithms. CyTargetLinker sheds some light into the network in terms of regulatory interactions i.e. miRNA-target, TF-target or drug-target interactions. The ErbB Signaling pathway for MS disease is available in the WikiPathways database and the network created for ErbB is shown in **Additional Fig 3**. The iRegulon software then allowed us to identify regulons using motif discovery in a set of regulated genes. Identified transcription factors affecting the hub genes are listed in **Additional Table 2** and **Additional Fig 4** and their common factors are identified in **Table 4**. The most significant, the *STAT5A* protein, mediates the responses of many cell ligands, such as IL2, IL3 and different growth hormones.

In this study, the gene identifiers were uploaded to Metascape and used in conjunction with KEGG pathways, GO biological processes, Reactome gene complexes, canonical and CORUM pathways [27]. The results of the enrichment analysis, including descriptions, function, ontology, expression, etc. are shown in **Table 5**, **Additional Table 3** and **Additional Fig 5**.

Genes were ranked from top to bottom based on degree, closeness and betweenness [higher degree (hub), higher betweenness (throat) and higher closeness centrality (shortest distance with other genes in the network)]. In terms of these parameters, three genes (*LASP1*, *TUBA1C*, *S100A6*) showed a significant correlation with MS disease. These were thus identified as hub genes (**Additional Table 4**). In this study, probes ILLMN_1665909, ILLMN_1742167, ILLMN_1713636 had high degrees of 15, 13 and 11, respectively and were identified as hub probes. 850 probes had zero input edges, 200 probes had 1 in-degree. ILLMN_1665909, with the highest out-degree (13 out-degree) and 2 in-degree (mapped to human *LASP1*) plays an important role in regulating activity. Its encoded protein is a cytoplasmic protein that binds focal adhesion proteins and plays a role in cell signaling, migration, and proliferation. ILLMN_1742167, with 12 out-degree and 1 in-degree mapped to the human tubulin gene (*TUBA1C*), and ILLMN_1713636 with 9 out-degree and 2 in-degree mapped to the *S100A6* gene. (**Figure 2** and **Additional Fig 6**).

Gene expression analysis

Patterns of relative gene expression for these genes (*LASP1*, *TUBA1C* and *S100A6*) showed significance between MS cases and controls ($P < 0.05$). Our results demonstrate that *LASP1* and *S100A6* genes show a significant up-regulated gene expression pattern while *TUBA1C* is significantly down-regulated when MS cases are compared to healthy controls (**Table 6** and **Figure 5**).

The results after studying the normality of the distribution of variables using the one sample Kolmogorov smirnov test and Unpaired t-test in GraphPad Prism 8 software show a significant difference in expression levels of *LASP1*, *TUBA1C* and *S100A6* genes between patients and healthy controls. P-values were: *TUBA1C* < 0.0001 , *S100A6* < 0.0001 , *LASP1* < 0.003 . Mean expression of *TUBA1C*, *LASP1* and *S100A6* genes in patient samples was 7.4, 5.6 and 2.9 respectively and 4.9, 8.1 and 8.0, respectively in healthy individuals. Results from statistical analysis also show a decrease in *TUBA1C* gene expression and an increase in *LASP1* and *S100A6* gene expression in MS patients compared to the control group (**Figure 6**).

Discussion

At present, the cause of MS is not fully understood, but knowledge of the genetic factors involved is essential for effective diagnosis and identification of the most appropriate MS therapeutic interventions. In this study, three genes (*LASP1*, *TUBA1c* and *S100A6*) with high degree, high closeness centrality and high betweenness measures, were highlighted as potential MS candidate regulator markers. These three genes (*LASP1*, *TUBA1c* and *S100A6*) seem to be the most significant genes in the MS disease process. Patients with MS are known to suffer from a number of digestive problems [28] and studies have shown that *LASP1* [29] and *S100A6* genes have high expression in the digestive system. A link can therefore be established between the expression of these genes, MS and gastrointestinal problems in MS patients.

S100A6 functioning in a wide range of cell types, is a member of the S100 family and may be involved in and expressed in several types of human cancers [30]. S100 family expression in MS patients could be considered as a diagnostic biomarker for MS, in which its inhibition of demyelinating nerve cells shows that S100 proteins could act as a candidate therapeutic target in MS [31]. Keiko Komatsu et al. (2000), reported increased expression of *S100A6* (Calcyclin), a calcium-bound protein of the S100 family, in human colorectal adenocarcinoma [32]; in this way, Eva Peterova (2021) reported an overexpression of S100 protein-encoding mRNA in both colorectal cancer cell lines [33]. A study by Bartkowska et al. [34] showed that in response to different stress conditions, the level of *S100A6* decreased in several brain structures, indicating that *S100A6* may modulate stress responses. In an autoimmune disease study, a genome-wide methylation array has identified a few hypomethylated immune-related genes, amongst them *S100A6* which shows up-regulation in autoimmune encephalitis patients [35].

Even though *S100A6* is involved in many biological phenomena e.g. cell proliferation, cytoskeletal dynamics and tumorigenesis, and some of its biological activity is still unknown [36]. At the transcriptional level, upstream stimulatory factor and Nuclear factor-kappa B (NF- κ B) activates the *S100A6* gene promoter, although p53 might act indirectly to suppress transcription of the *S100A6* gene [37].

Microtubules, that are essential for multiple cellular processes, are constructed from homologous tubulin proteins. Microtubules shape crucial cytoskeletal structures that play a pivotal role in creating and maintaining neuronal mechanistic, biochemical functions such as polarity, regulating neuronal morphology, dendrite growth, neuron migration throughout brain development, transporting cargo, and scaffolding signaling molecules to form signaling hubs, and are vulnerable to degradation and disorganization in a variety of neurodegenerative diseases [38-40]. Malfunction of microtubules (*Tuba1c*) is also considered as the central physiopathological mechanism of neurodegenerative diseases and abnormalities in the regulatory pathways of microtubules disrupt the properties and functions of microtubules,

leading to nerve damage [41]. In a 2006 study, a decreased expression of the *TUBA1C* gene in Parkinson's disease was demonstrated by quantitative analysis of gene expression [42].

LASP1 (The *LIM* and *SH3* protein 1) a focal adhesion adaptor protein, is an actin-binding, signaling pathway-regulated phosphoprotein and localizes within multiple sites of dynamic actin assembly. It has the potential to interact with various molecules, and is highly expressed in the adult central nervous system. It is ubiquitously expressed in normal tissues, and it transmits signals from the cytoplasm into the nucleus. It is a versatile structural, signaling, and biomarker protein and plays a crucial role in the growth and metastasis of gastric cancer, as well as being overexpressed in several other cancers [43-48]. Also, overexpression of *LASP1* is often seen in colon cancer tissues, especially in metastatic colon cancer tissues. *LASP1* can cause the progression and metastasis of colorectal cancer (CRC), but its mechanism is still unclear [49]. Another study showed that *LASP1* binds to the calcium-binding protein family (*S100A*) and increases its expression in colon cancer [50]. this way, statistical analysis revealed that *LASP1* and *S100P* are correlated ($\text{Kappa} = 0.347$, $P < 0.01$). Microarray data has revealed that alterations in *LASP1* proteins affect cell migration, adhesion, and cytoskeletal organization [47]. *LASP1*, significantly expressed by CNS neurons, is localized at synaptic sites [51].

A couple of significant transcription factors that interact with these hub genes were identified in this study. The YY1 transcription factor (Yi and Yang 1) is a multifactorial protein that, depending on the cell tissue, can activate or suppress gene expression [52]. It is expressed in the nervous system. The YY1 promoter lacks the usual TATA box but has a rich GC sequence and therefore resembles a large subset of housekeeping and growth regulator genes. These features suggest that it may play an important role in development. In the central nervous system, myelination is performed by oligodendrocytes. YY1 function in oligodendrocytes was first reported by Berndt et al (2001). YY1 activates the promoter of myelin lipids and has been identified as an important player in myelination of the central nervous system during growth. In multiple neurodegenerative diseases, YY1 function is degraded through distinct mechanisms, including protein utilization, protein degradation, and ectopic nuclear / cytoplasmic shuttle (N/C). These disorders inhibit YY1 transcriptional activity and lead to gene transcriptional abnormalities that contribute to disease pathogenesis. A future goal in YY1 research is to discover other potential mechanisms that lead to YY1 dysfunction in neurodegenerative diseases, such as ectopic changes after translation [53].

NFAT is a family of transcription factors originally described as key factors in the immune response. There are five major members of the NFAT family: NFATc3 is one of five members of the NFAT transcription factor family that act as signal integrators because their function is to bind STAT3, c-Jun, CREB, and ATF3 factors at specific DNA binding sites. NFATs cannot be regulated alone and act as calcium-dependent transcription factors. The antigen-mediated T cell receptor (TCR) mediates multiple signaling cascades, including phospholipase C (PLC) -dependent pathways that are secondary messengers of inositol-1,4,5-triphosphate (IP3) and diacylglycerol (DAG). IP3 binds to the IP3 receptor in the endoplasmic reticulum (ER) and releases Ca^{2+} ions into the cytoplasm. [54].

NFATc1-4 activates intracellular calcium via dephosphorylation. Calcineurin activates a calcium / calmodulin-dependent phosphatase, which detects the transmitter signal to the nucleus (NLS) and carries it to the nucleus. This acts as a link between calcium signaling and NFAT-dependent gene transcription. Conversely, NFAT inactivation is mediated by phosphorylation by kinases that either retain NFAT in the cytoplasm or enhance nuclear export. According to a recent study, native human astrocytes also express NFAT and transmit to the nucleus [21]. The findings suggest that NFATs control pathways involved in astrocyte activation and, therefore, may affect neuronal cell survival. Such a compelling idea suggests that NFAT is employed in neurological pathologies associated with neuro-inflammation in astrocytes. The findings show that NFATc3 is defined as a marker of a specific subset of astrocytes that are activated in response to lesions, as well as some degree of heterogeneity among astrocytes that may have consequences for cells in the nervous

system [55]. Preliminary findings in neuroblast cells have shown that various treatments that alter tubulin polymerization, such as reducing the mineral zinc, prevent the transfer of NFAT to the nucleus.

In addition, NFAT clusters gradually move away from the microtubules and communicate by transporting NFAT to the nucleus. Therefore, it has been shown that the association of NFAT with the microtubule network can increase the concentration of this transcription factor around the nucleus, and/or facilitate its interaction with nuclear pores. In agreement with a functional relationship between NFAT and microtubules, it has been observed that the degradation of several proteins that control the proper organization of the microtubule network, and the actin-cytoskeletal linker, disrupts the nucleus and transcriptional activity of NFAT. Overall, it indicates the involvement of microtubules in NFAT nuclear stimulation [56]. The *LASP1* gene enhances NFAT2 nuclear translocation by activating the nuclear factor Akt [57]. NFAT can affect processes such as axon growth, synaptogenesis, Schwann cell differentiation, and myelination [55]. In general, it can be concluded that increase of the expression of *LASP1* and *S100A6* genes and decrease the expression of the *TUBA1C* gene in multiple sclerosis disrupts NFAT transcriptional activity. Although the role of NFAT in regulating the immune system is well established, our knowledge of NFAT in human disease is limited. The function of NFAT in other aspects of human immune or inflammatory diseases is also largely unknown [58].

On this basis, the present study confirmed the importance of three gene expression patterns (*LASP1*, *S100A6*, *TUBA1C*) for understanding the transcriptome complexity of MS. This leads us to conclude that upregulation of *LASP1* and *S100A6* genes along with down-regulation of *TUBA1C* is central to MS pathology. To our knowledge, this is the first report to evaluate the level of expression of the above genes for discovery of a transcriptomic signature for MS disease. These findings provide a potential mechanism for some significant biomarkers responsible for the pathogenesis of MS. However, we still have a long way to go to understanding the larger transcriptomic profile for this disease. This study provides initial data to further investigate the possible role of these genes in the pathogenesis of MS.

Conclusions

NetworkAnalyzer is a versatile tool for the analysis of biological and other networks. In this study, we implement our network scoring methods into a Cytoscape plugin, *cytoHubba*. Results of the present study indicate that the analysis of gene expression data based on gene-gene interaction networks can provide more opportunities to determine the genes involved in MS. The importance of three candidate marker genes in this disease were highlighted -*LASP1*, *TUBA1C* and *S100A6*. These candidate genes identified by the biological systems approach, have been further confirmed in the laboratory. The significant difference in the expression of these three genes in patients with MS can be concluded to have a significant effect during this disease and will help further research on MS and its treatment. This useful tool can serve as a good starting point for identifying new therapies and understanding the basic mechanisms controlling normal cellular processes and disease pathologies.

Abbreviations

MS: Multiple Sclerosis

TFs: Transcription factors

GEO: Gene Expression Omnibus

BIC: Bayesian information criterion

BN: Bayesian Network

RRMS: Relapsing-remitting MS

SPMS: Secondary progressive MS

NF- κ B: Nuclear factor-kappa B

***LASP1*:** The *LIM* and *SH3* protein 1

YY1: Yi and Yang 1

TCR: T cell receptor

PLC: Phospholipase C

IP3: inositol-1,4,5-triphosphate

DAG: Diacylglycerol

ER: Endoplasmic reticulum

NLS: Nuclear localization signal

Declarations

Ethical Approval and Consent to participate

This project was in accordance with the ethical principles and national norms and standards for medical research in Iran and it is confirmed that all methods have been performed in accordance with the relevant instructions and regulations with Approval ID: IR.UI.REC.1399.076.

All experimental procedures were approved by the Ethics Committees of the University of Isfahan, and informed consent was obtained from all participant prior to in the study.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and/or analyzed in the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was funded by the University of Isfahan and the Biotechnology Development Headquarters.

Author's contributions

These authors contributed equally: Nafiseh Karimi, Majid Motovalli Bashi and Mostafa Ghaderi-Zefrehei.

Acknowledgments

The authors thank all the patients, their families, and caregivers. The authors thank all the clinicians for their involvement and contribution to the study. This work was supported by Jacqueline Smith from The University of Edinburgh, Easter Bush, Midlothian, EH25 9RG, UK for *Writing-review & editing*.

References

1. Hernandez AL, O'Connor KC, Hafler DA. Chapter 52 - Multiple Sclerosis. In: Rose NR, Mackay IR, editors. *The Autoimmune Diseases* (Fifth Edition). Boston: Academic Press; 2014. p. 735-56.
2. Hanafy KA, Sloane JA. Regulation of remyelination in multiple sclerosis. *FEBS Lett.* 2011;585(23):3821-8.
3. Han MH, Steinman L. Systems biology for identification of molecular networks in multiple sclerosis. *Multiple Sclerosis Journal.* 2009;15(5):529-30.
4. Kotelnikova E, Bernardo-Faura M, Silberberg G, Kiani NA, Messinis D, Melas IN, et al. Signaling networks in MS: A systems-based approach to developing new pharmacological therapies. *Multiple Sclerosis Journal.* 2014;21(2):138-46.
5. Safari-Alighiarloo N, Rezaei-Tavirani M, Taghizadeh M, Tabatabaei SM, Namaki S. Network-based analysis of differentially expressed genes in cerebrospinal fluid (CSF) and blood reveals new candidate genes for multiple sclerosis. *PeerJ.* 2016;4:e2775.
6. McGuinness MJ. *A Mathematical Model of Remyelination in Multiple Sclerosis: Undergraduate Honors Theses*; 2017.
7. Muñoz-San Martín M, Reverter G, Robles-Cedeño R, Buxò M, Ortega FJ, Gómez I, et al. Analysis of miRNA signatures in CSF identifies upregulation of miR-21 and miR-146a/b in patients with multiple sclerosis and active lesions. *Journal of Neuroinflammation.* 2019;16(1):220.
8. Liu M, Hou X, Zhang P, Hao Y, Yang Y, Wu X, et al. Microarray gene expression profiling analysis combined with bioinformatics in multiple sclerosis. *Mol Biol Rep.* 2013;40(5):3731-7.
9. Cervantes-Gracia K, Husi H. Integrative analysis of Multiple Sclerosis using a systems biology approach. *Sci Rep.* 2018;8(1):5633.
10. Diaz-Beltran L, Cano C, Wall DP, Esteban FJ. Systems Biology as a Comparative Approach to Understand Complex Gene Expression in Neurological Diseases. *Behavioral Sciences.* 2013;3(2):253-72.
11. Friedman N, Linial M, Nachman I, Pe'er D. Using Bayesian networks to analyze expression data. *Proceedings of the fourth annual international conference on Computational molecular biology*; Tokyo, Japan: Association for Computing Machinery; 2000. p. 127–35.
12. Davis S MP. a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinf.* 2007;14:1846–7.
13. Hao C. *Learning Bayesian Network Structure from Data* 2018.
14. Scutari M. Bayesian network constraint-based structure learning algorithms: Parallel and optimised implementations in the bnlearn r package. *arXiv preprint arXiv:14067648.* 2014.
15. Assenov Y. *Topological Analysis of Biological Networks*: Max Planck Institute for Informatics; 2006.
16. Su G, Morris JH, Demchak B, Bader GD. Biological network exploration with Cytoscape 3. *Current protocols in bioinformatics.* 2014;47:8.13.1-8..24.
17. Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, et al. Integration of biological networks and gene expression data using Cytoscape. *Nature protocols.* 2007;2(10):2366-82.
18. Chin C-H, Chen S-H, Wu H-H, Ho C-W, Ko M-T, Lin C-Y. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC systems biology.* 2014;8 Suppl 4(Suppl 4):S11-S.

19. Assenov Y, Ramírez F, Schelhorn S-E, Lengauer T, Albrecht M. Computing topological parameters of biological networks. *Bioinformatics*. 2007;24(2):282-4.
20. Yingyao Zhou , Bin Zhou , Lars Pache , Max Chang , Alireza Hadj Khodabakhshi , Olga Tanaseichuk , et al. <https://metascape.org/>. 2019;10(1):1523.
21. Giaimo BD, Oswald F, Borggreffe T. Dynamic chromatin regulation at Notch target genes. *Transcription*. 2017;8(1):61-6.
22. Detanico T, Virgen-Slane R, Steen-Fuentes S, Lin WW, Rhode-Kurnow A, Chappell E, et al. Co-expression Networks Identify DHX15 RNA Helicase as a B Cell Regulatory Factor. *Frontiers in immunology*. 2019;10:2903.
23. Newman MEJ. The Structure and Function of Complex Networks. *SIAM Review*. 2003;45(2):167-256.
24. Barabási A-L, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nature Reviews Genetics*. 2010;12:56.
25. Manuel AM, Dai Y, Freeman LA, Jia P, Zhao Z. Dense module searching for gene networks associated with multiple sclerosis. *BMC Medical Genomics*. 2020;13(5):48.
26. Chin C-H, Chen S-H, Wu H-H, Ho C-W, Ko M-T, Lin C-Y. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC systems biology*. 2014;8 Suppl 4:S11-S.
27. <https://mips.helmholtz-muenchen.de/corum>. The comprehensive resource of mammalian protein complexes 2018 [
28. Gulick E, Namey M. Bowel Dysfunction in Persons with Multiple Sclerosis. 2012.
29. Orth MF, Cazes A, Butt E, Grunewald TGP. An update on the LIM and SH3 domain protein 1 (LASP1): a versatile structural, signaling, and biomarker protein. *Oncotarget*. 2015;6(1).
30. Wang XH, Du H, Li L, Shao DF, Zhong XY, Hu Y, et al. Increased expression of S100A6 promotes cell proliferation in gastric cancer cells. *Oncol Lett*. 2017;13(1):222-30.
31. Barateiro A, Afonso V, Santos G, Cerqueira JJ, Brites D, van Horssen J, et al. S100B as a Potential Biomarker and Therapeutic Target in Multiple Sclerosis. *Mol Neurobiol*. 2016;53(6):3976-91.
32. Komatsu K, Andoh A, Ishiguro S, Suzuki N, Hunai H, Kobune-Fujiwara Y, et al. Increased expression of S100A6 (Calcyclin), a calcium-binding protein of the S100 family, in human colorectal adenocarcinomas. *Clin Cancer Res*. 2000;6(1):172-7.
33. Peterova E, Bures J, Moravkova P, Kohoutova D. Tissue mRNA for S100A4, S100A6, S100A8, S100A9, S100A11 and S100P Proteins in Colorectal Neoplasia: A Pilot Study. *Molecules*. 2021;26(2).
34. Bartkowska K, Swiatek I, Aniszewska A, Jurewicz E, Turlejski K, Filipek A, et al. Stress-Dependent Changes in the CacyBP/SIP Interacting Protein S100A6 in the Mouse Brain. *PloS one*. 2017;12(1):e0169760-e.
35. Tsai M-H, Lin C-H, Tsai K-W, Lin M-H, Ho C-J, Lu Y-T, et al. S100A6 Promotes B Lymphocyte Penetration Through the Blood–Brain Barrier in Autoimmune Encephalitis. *Frontiers in Genetics*. 2019;10(1188).
36. Donato R, Cannon BR, Sorci G, Riuzzi F, Hsu K, Weber DJ, et al. Functions of S100 proteins. *Current molecular medicine*. 2013;13(1):24-57.
37. Leśniak. W, Słomnicki Ł P, Filipek A. S100A6 - new facts and features. *Biochem Biophys Res Commun*. 2009;390(4):1087-92.
38. Aiken J, Buscaglia G, Bates EA, Moore JK. The α -Tubulin gene TUBA1A in Brain Development: A Key Ingredient in the Neuronal Isotype Blend. *J Dev Biol*. 2017;5(3).
39. Dubey J, Ratnakaran N, Koushika SP. Neurodegeneration and microtubule dynamics: death by a thousand cuts. *Frontiers in cellular neuroscience*. 2015;9:343-.
40. Matamoros AJ, Baas PW. Microtubules in health and degenerative disease of the nervous system. *Brain research bulletin*. 2016;126(Pt 3):217-25.

41. Sferra A, Nicita F, Bertini E. Microtubule Dysfunction: A Common Feature of Neurodegenerative Diseases. *Int J Mol Sci.* 2020;21(19).
42. Kim JM, Lee KH, Jeon YJ, Oh JH, Jeong SY, Song IS, et al. Identification of genes related to Parkinson's disease using expressed sequence tags. *DNA Res.* 2006;13(6):275-86.
43. Butt E, Raman D. New Frontiers for the Cytoskeletal Protein LASP1. *Frontiers in Oncology.* 2018;8(391).
44. Grunewald TGP, Butt E. The LIM and SH3 domain protein family: structural proteins or signal transducers or both? *Molecular Cancer.* 2008;7(1):31.
45. Li Z, Chen Y, Wang X, Zhang H, Zhang Y, Gao Y, et al. LASP-1 induces proliferation, metastasis and cell cycle arrest at the G2/M phase in gallbladder cancer by down-regulating S100P via the PI3K/AKT pathway. *Cancer Lett.* 2016;372(2):239-50.
46. Orth MF, Cazes A, Butt E, Grunewald TGP. An update on the LIM and SH3 domain protein 1 (LASP1): a versatile structural, signaling, and biomarker protein. *Oncotarget.* 2015;6(1):26-42.
47. Zhang H, Chen X, Bollag WB, Bollag RJ, Sheehan DJ, Chew CS. Lasp1 gene disruption is linked to enhanced cell migration and tumor formation. *Physiol Genomics.* 2009;38(3):372-85.
48. Zheng J, Yu S, Qiao Y, Zhang H, Liang S, Wang H, et al. LASP-1 promotes tumor proliferation and metastasis and is an independent unfavorable prognostic factor in gastric cancer. *J Cancer Res Clin Oncol.* 2014;140(11):1891-9.
49. Zhao L, Wang H, Liu C, Liu Y, Wang X, Wang S, et al. Promotion of colorectal cancer growth and metastasis by the LIM and SH3 domain protein 1. *Gut.* 2010;59:1226 - 35.
50. Niu Y, Shao Z, Wang H, Yang J, Zhang F, Luo Y, et al. LASP1-S100A11 axis promotes colorectal cancer aggressiveness by modulating TGFβ/Smad signaling. *Scientific Reports.* 2016;6(1):26112.
51. Phillips GR, Anderson TR, Florens L, Gudas C, Magda G, Yates JR, 3rd, et al. Actin-binding proteins in a postsynaptic preparation: Lasp-1 is a component of central nervous system synapses and dendritic spines. *J Neurosci Res.* 2004;78(1):38-48.
52. He Y, Casaccia-Bonnel P. The Yin and Yang of YY1 in the nervous system. *J Neurochem.* 2008;106(4):1493-502.
53. Chen ZS, Chan HYE. Transcriptional dysregulation in neurodegenerative diseases: Who tipped the balance of Yin Yang 1 in the brain? *Neural regeneration research.* 2019;14(7):1148-51.
54. Lee J-U, Kim L-K, Choi J-M. Revisiting the Concept of Targeting NFAT to Control T Cell Immunity and Autoimmune Diseases. *Frontiers in immunology.* 2018;9:2747-.
55. Serrano-Pérez MC, Martín ED, Vaquero CF, Azcoitia I, Calvo S, Cano E, et al. Response of transcription factor NFATc3 to excitotoxic and traumatic brain insults: identification of a subpopulation of reactive astrocytes. *Glia.* 2011;59(1):94-107.
56. Mastrogiiovanni M, Juzans M, Alcover A, Di Bartolo V. Coordinating Cytoskeleton and Molecular Traffic in T Cell Migration, Activation, and Effector Functions. *Frontiers in Cell and Developmental Biology.* 2020;8(1138).
57. Wu L, Lin W, Liao Q, Wang H, Lin C, Tang L, et al. Calcium Channel Blocker Nifedipine Suppresses Colorectal Cancer Progression and Immune Escape by Preventing NFAT2 Nuclear Translocation. *Cell Reports.* 2020;33(4):108327.
58. Pan MG, Xiong Y, Chen F. NFAT gene family in inflammation and cancer. *Current molecular medicine.* 2013;13(4):543-54.

Tables

Table 1: Primers designed for QRT-PCR.

NO.	Gene	Name	Seq.(5-3)	TM
1	H-TUBA1C	F	TTCCACCCTGAGCAACTC	60
		R	AACCAAGAAGCCCTGAAG	
2	H-S100A6	F	AGCACACCCTGAGCAAGA TCACCTCCTGGTCCTTGT	60
		R		
3	H-LASP1	F	GAGCAGCAGCCTCACCAC TACCGCTTCCCGCCAC	64
		R		
4	β -actin	F	TGGAGGTACCACCATGTACC CACATCTGCTGGAAGGTGGA	60
		R		

Table 2: Estimation of structural Bayesian network parameters with different algorithms.

Parameters	Score Based Algorithm		Hybrid Algorithm	
	Hill Climbing	Tabu Search	Max-Min Hill Climbing	Restricted Maximize
No. of Nodes	1707	1707	1707	1707
No. of Arcs	1700	1500	2485	2188
Undirected Arcs	0	0	0	0
Directed Arcs	1700	1500	2485	2188
Markov Blanket	3.32	2.69	4.64	3.93
Neighborhood Size	1.99	1.76	2.91	2.56
Branching Factor	1	0.88	1.46	1.28
No. of Tests	4354565	413365	9721786	7228057
loglik-g	-1308146	-1308146	-1318282	-1313436
AIC-g	-1312560	-1312560	-1323884	-1319335
BIC-g	-1319115	-1319115	-1332202	-1328094

loglik-g: The multivariate Gaussian log-likelihood (loglik-g) score, AIC-g: Akaike Information Criterion score, BIC-g: Bayesian Information Criterion score.

Table 3: The 10 top genes/probes identified by the MCC method.

ProbelD	Transcript ID	Gene Name
ILMN_2180682	ENSG00000105887	"MTPN", myotrophin
ILMN_1797342	ENSG00000187239	"FNBP1",formin binding protein 1
ILMN_1713636	ENSG00000197956	"S100A6",S100 calcium binding protein A6
ILMN_1663512	ENSG00000135940	"COX5B",cytochrome c oxidase subunit 5B
ILMN_2333319	ENSG00000011304	"PTBP1",polypyrimidine tract binding protein 1
ILMN_1792150	ENSG00000012779"	"ALOX5",arachidonate 5-lipoxygenase
ILMN_1682993	ENSG00000105374	"NKG7",natural killer cell granule protein 7
ILMN_1665909	ENSG000000002834	"LASP1",LIM and SH3 protein 1
ILMN_1742167	ENSG00000167553	"TUBA1C",tubulin alpha 1c
ILMN_1691611	ENSG00000227453	"HNRNPA1P63",heterogeneous nuclear ribonucleoprotein A1 pseudogene 63

Table 4: The top 10 transcription factors (TFs) estimated to affect hub genes.

Transcription factors(TF)	NES*	AUC**	Target genes
<i>STAT5A</i>	3.01743	0.0435029	<i>LASP1, TUBA1C</i>
<i>NFATC1</i>	3.12016	0.0443181	
<i>MTA3</i>	3.19349	0.0448999	
<i>NFKB1</i>	4.00501	0.037604	
<i>ZNF362</i>	3.3813	0.0265967	
<i>SPI1</i>	6.96136	0.037604	<i>LASP1,S100A6</i>
<i>GABPB1</i>	4.86575	0.0307835	
<i>DLX1</i>	3.35429	0.0461759	<i>TUBAC,S100A6</i>
<i>YY1</i>	6.59215	0.0364023	<i>LASP1,TUBA1C,S100A6</i>
<i>NFATC3</i>	5.38372	0.0324693	

*Normalized enrichment score (NES). **Area Under the cumulative Recovery

Table5: Metascape results *LASP1*, *TUBA1C* and *S100A6* genes.

Input ID	Gene ID	Tax ID	Gene Symbol	Description	Biological Process (GO)	Subcellular Location (Protein Atlas)
<i>LASP1</i>	3927	H. sapiens	<i>LASP1</i>	LIM and SH3 protein 1	GO:0034220 ion transmembrane transport;GO:0009967 positive regulation of signal transduction;GO:0023056 positive regulation of signaling	Cytosol;Plasma membrane(Supported)Focal adhesion sites(Approved)
<i>TUBA1C</i>	84790	H. sapiens	<i>TUBA1C</i>	tubulin alpha 1c	GO:0030705 cytoskeleton-dependent intracellular transport;GO:0000226 microtubule cytoskeleton organization;GO:0051301 cell division	Microtubules(Supported)
<i>S100A6</i>	6277	H. sapiens	<i>S100A6</i>	S100 calcium binding protein A6	GO:0048146 positive regulation of fibroblast proliferation;GO:0048145 regulation of fibroblast proliferation;GO:0007409 axonogenesis	Cytosol;Plasma membrane(Enhanced)

Table 6: REST software data compared *LASP1*, *TUBA1c* and *S100A6* genes in MS patient and control groups.

Gene	Type	Reaction Efficiency	Expression	Std. Error	95% C.I.	P(H1)	Result
<i>ACTB</i>	REF	1.0	1.000				
<i>LASP1</i>	TRG	1.0	5.491	1.183 - 32.843	0.335 - 266.871	0.000	UP
<i>TUBA1C</i>	TRG	1.0	0.166	0.058 - 0.523	0.027 - 1.417	0.000	DOWN
<i>S100A6</i>	TRG	1.0	36.556	9.630 - 140.562	2.367 - 416.452	0.000	UP

Legend: P(H1) Probability of alternate hypothesis that difference between sample and control groups is due only to chance; TRG – Target; REF – Reference; Interpretation: *LASP1* is UP-regulated in sample group (in comparison to control group) by a mean factor of 5.491 (S.E. range is 1.183 - 32.843); *LASP1* sample group is different to control group. P(H1)=0.000. *TUBA1c* is DOWN-regulated in sample group (in comparison to control group) by a mean factor of 0.166 (S.E. range is 0.058 - 0.523). *TUBA1c* sample group is different to control group. P(H1)=0.000. *S100a6* is UP-regulated in sample group (in comparison to control group) by a mean factor of 36.556 (S.E. range is 9.630 - 140.562). *S100A6* sample group is different to control group. P(H1)=0.000

Figures

Figure 1.

Figure 1

The analysis pipeline used in this study.

A: Shortest path length distribution. The path length is the number of edges along the path. The distance d_{ij} between a pair of different nodes i and j is the length of the shortest path connection.

B: Betweenness centrality. The Betweenness centrality of a node reflects the amount of control that this node exerts over the interactions of other nodes in the network.

C: Distribution of in-degree gene connectivity measure.

D: Distribution of out-degree gene connectivity measure.

Figure 2.

Figure 2

Some topological measures of trained gene regulatory BN visualized by NetworkAnalyzer.

Fig 3.

Figure 3

Extracted molecular modules using jActiveModules.

Boxes represent the interquartile range, or the middle 50% of observations. The dotted line represents the median gene expression. Whiskers represent the minimum and maximum observations.

Fig 6.

Figure 6

Validation of the expression of TUBA1C LASP1, and S100A6 genes by real-time PCR. Unpaired t test was used to detect differences in gene expression between 2 groups patients & controls using the Graph Pad Prism 8 software. Significance: P-value TUBA1C<0.0001, P-value S100A6 <0.0001, P-value LASP1 <0.003.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [AdditionalFig1..tif](#)
- [AdditionalFig2..tif](#)
- [AdditionalFig3..tif](#)
- [AdditionalFig4..tif](#)
- [AdditionalFigure5..tif](#)
- [AdditionalFig6..tif](#)
- [AdditionalTable1..tif](#)
- [AdditionalTable2..tif](#)
- [AdditionalTable3..tif](#)
- [AdditionalTable4..tif](#)