Feature selection based on differentially correlated gene pairs reveals the mechanism of IFN-β therapy for multiple sclerosis

Tao Jin ¹, Chi Wang ², Suyan Tian Corresp. ³

¹ Department of Neurology and Neuroscience Center, The First Hospital of Jilin University, Changchun, China
² Department of Biostatistics and Markey Cancer Center, University of Kentucky, Lexington, United States
³ Division of Clinical Research, The First Hospital of Jilin University, Changchuan, Jilin, China

Corresponding Author: Suyan Tian
Email address: windytian@hotmail.com

Multiple sclerosis (MS) is one of the most common neurological disabilities of the central nervous system. Immune-modulatory therapy with Interferon-β (IFN-β) is a commonly used first-line treatment to prevent MS patients from relapses. Nevertheless, a large proportion of MS patients on IFN-β therapy experience their first relapse within two years of treatment initiation. Feature selection, a machine learning strategy, is routinely used in the fields of bioinformatics and computational biology to determine which subset of genes is most relevant to an outcome of interest. The majority of feature selection methods focus on alterations in gene expression levels. In this study, we sought to determine which genes are most relevant to relapse of MS patients on IFN-β therapy. Rather than the usual focus on alterations in gene expression levels, we devised a feature selection method based on alterations in gene-to-gene interactions. In this study, we applied the proposed method to a longitudinal microarray dataset and evaluated the IFN-β effect on MS patients to identify gene pairs with differentially correlated edges (DCEs) that are consistent over time in the responder group compared to the non-responder group. The resulting gene list had a good predictive ability on an independent validation set and explicit biological implications related to MS. To conclude, it is anticipated that the proposed method will gain widespread interest and application in personalized treatment research to facilitate prediction of which patients may respond to a specific regimen.
Feature selection based on differentially correlated gene pairs reveals the mechanism of IFN-β therapy for multiple sclerosis

Tao Jin¹, Chi Wang², and Suyan Tian³*

¹Department of Neurology and Neuroscience Center, The First Hospital of Jilin University, 1 Xinmin Street, Changchun, Jilin, China, 130021

²Department of Biostatistics and Markey Cancer Center, University of Kentucky, 800 Rose St. Lexington, KY, USA, 40536

³Division of Clinical Research, The First Hospital of Jilin University, 1 Xinmin Street, Changchun, Jilin, China, 130021

Corresponding author:
Suyan Tian
Division of Clinical Research
The First Hospital of Jilin University
1 Xinmin Street
Changchun, Jilin, China, 130021
email: windytian@hotmail.com
Multiple sclerosis (MS) is one of the most common neurological disabilities of the central nervous system. Immune-modulatory therapy with Interferon-β (IFN-β) is a commonly used first-line treatment to prevent MS patients from relapses. Nevertheless, a large proportion of MS patients on IFN-β therapy experience their first relapse within two years of treatment initiation. Feature selection, a machine learning strategy, is routinely used in the fields of bioinformatics and computational biology to determine which subset of genes is most relevant to an outcome of interest. The majority of feature selection methods focus on alterations in gene expression levels. In this study, we sought to determine which genes are most relevant to relapse of MS patients on IFN-β therapy. Rather than the usual focus on alterations in gene expression levels, we devised a feature selection method based on alterations in gene-to-gene interactions. In this study, we applied the proposed method to a longitudinal microarray dataset and evaluated the IFN-β effect on MS patients to identify gene pairs with differentially correlated edges (DCEs) that are consistent over time in the responder group compared to the non-responder group. The resulting gene list had a good predictive ability on an independent validation set and explicit biological implications related to MS. To conclude, it is anticipated that the proposed method will gain widespread interest and application in personalized treatment research to facilitate prediction of which patients may respond to a specific regimen.

**Keywords:** Multiple sclerosis; differentially correlated edge; longitudinal; IFN-β therapy; treatment response; feature selection.
Introduction

Multiple sclerosis (MS) is an immune-mediated, inflammatory demyelinating disease of the central nervous system (CNS) that affects about 2.3 million people worldwide [1, 2]. The majority of patients are women between 20 and 40 years old, and the ratio of female to male is about 3:1 [1, 2]. Depending on the initial course of disease, MS may be categorized as relapsing-remitting or primary progressive disease [1]. Relapsing-remitting MS (RRMS) is the most prevalent form of MS, accounting for 85% of MS cases. As indicated by its name, it involves relapses followed by series of remission. In the early stage, patients with this type of MS may experience complete recovery or partial sequel; however, approximately 50% of patients progress over time from to secondary progressive disease, with or without acute relapses. In contrast, primary progressive MS begins with gradual neurologic deterioration from onset [1, 3].

The etiology of MS has not been fully elucidated, however, it is commonly believed to be triggered as an autoimmune response to an interaction between genetic and environmental factors [4]. In the pathogenesis of MS, T helper (Th)1 cells release pro-inflammatory cytokines such as interleukin-2 (IL-2), interferon-gamma (IFN-γ) and tumor necrosis factor-α (TNF-α) and play a crucial role in inflammation of MS. Th17 cells release IL-17 and have some impacts on MS as well. On the other hand, Th2 cells secrete IL-4, -5, and -10 to counter-regulate Th1 and Th17 responses [5]. Interferon-beta (IFN-β) is the first class of treatment. It was approved for relapsing-remitting MS in 1993, and was subsequently approved in Europe for secondary progressive MS with continued relapse. It is thought that IFN-β works to combat MS by stabilizing the blood brain barrier and altering relevant cytokines by redirecting pro-inflammatory Th1 and Th17 immune responses into anti-inflammatory Th2 responses.

There are an increasing amount of studies suggesting that genomic data such as gene expression profiles and genetic variants data may provide insightful clues on the development and molecular subtypes of MS. For example, Parnell et al. [6] tested the expression values of 110 known MS-related genes, and found that expression of the transcription factors (TF) controlling T and NK cell differentiation, EOMES, TBX21 and other TFs was significantly lower in MS compared to healthy controls. Besides the long-time well-known strong genetic effect of the HLA-DR2 locus, non-HLA variants (such as IL2R, IL7R and IL2R variants) have been implicated. Among them, IL2R and IL7R are focused the most since they are expressed on regulatory T cells which play crucial roles in regulating the immune response in MS patients[4].
Feature selection is a machine learning strategy of selecting, from among thousands of genes, a
gene signature (subset) that may be relevant for diagnosis of a disease, segmentation of disease
subtypes, patient drug response or survival rate prediction, and is becoming routine practice in the
fields of bioinformatics and computational biology [7, 8]. Until now, the majority of available
feature selection methods has focused on investigating alterations of gene expression levels. These
are referred to as node-centric methods. An alternative strategy is to focus on alterations of
relationships (correlation) among genes, which should also bear meaningful information about
which genes are likely to be associated with the outcomes of interest. Here, a “node” corresponds
to a gene in a gene-to-gene interaction network/graph and an “edge” corresponds to the correlation
between a gene pair. As pointed out by Kostka et al. [9], a differential expression analysis focuses
on the first moments (i.e., means) and a differential correlation analysis aims at the second
moments instead from the prospect of statistics. So far, development of such edge-centric methods
is far from being comparable to development of node-centric counterpart (see [9–11] for examples
of edge-centric methods).

Compared with classic feature selection methods which mainly deal with cross-sectional data, the
feature selection process for longitudinal gene expression data is more complicated. This is
unsurprising given the fact that longitudinal gene expression data involve more than a single time
point and consider both the expression value trajectory over time and their differences between
different phenotypes. So far, feature selection methods specifically to handle longitudinal data are
far from sufficient, to the best of our knowledge, the method proposed by Sun et al [12] is one of
the few edge-centric methods capable of dealing with longitudinal data. Specifically, in this so
called differentially expressed network method, an edge is selected if the directly linked gene pair
is strongly correlated (the absolute Spearman correlation coefficient is larger than a threshold) in
one group while not in the other group. Combining the differentially expressed genes of the two
groups builds a differentially expressed network. However, using a threshold to decide the
connection status between gene pairs may result in many falsely positive edges in which the
difference of correlation coefficients under two different conditions may be marginal.

Motivated by the differentially expressed network method [12], we proposed a bioinformatics
procedure to identify gene pairs with differentially correlated edges over time by replacing the
difference in connection statuses with that in actual correlation coefficients. Then we applied the
proposed method to a longitudinal microarray dataset which evaluates the effect of IFN-β on MS
patients. Our objective is to further elucidate the underlying therapeutic mechanisms of IFN-β and
to predict which subset of MS patients will respond to IFN-β therapy based on the identified
molecular signature.
Materials and Methods

Experimental data

A microarray experiment dataset was used as a training set in this study (GEO repository accession number GSE24427) [13]. We collected longitudinal gene expression profiles from 25 German relapsing-remitting MS patients treated with recombinant IFN-β-1b (250 µg every other day) for two years. The chips were hybridized on the Affymetrix hgu133A and B platforms. Only chips hybridized on the Affymetrix hgu133A platform were considered. This platform included 22,283 probe sets, corresponding to 12,437 unique genes.

Gene expression values were measured at five separate time points: before the first, second, 1-month, 12-month, and 24-month injections. Keeping in mind that several patients may have relapsed shortly after the two-year treatment (in which case their gene expression profiles may have undergone some changes similar to changes of non-responders), we restricted the responder category to patients whose first relapse time was more than 5 years (60 months), resulting in 9 responders and 9 non-responders fed into the downstream analysis.

To validate the predictive performance of the final model, data from the GSE19285 experiment [14] were downloaded from the GEO repository to serve as a validation set. In that study, 24 relapsing-remitting MS patients were injected with intramuscular IFN-β-1a once a week and the gene expression data were collected at the first, second and fifth injections. Likewise, the chips of that study were hybridized on both the Affymetrix hgu133A and B platform, but only data on the Affymetrix hgu133A platform were considered in this study. By adding the restriction of more than 5-years duration for the first relapse, 21 patients were included in this testing set; 12 were responders and 9 were non-responders.
Pre-processing procedures

Raw data (CEL files) of these two experiments were downloaded from the GEO repository, respectively. The expression values were obtained using the fRMA algorithm [15]. The expression summary values were normalized using quartile normalization and then log 2 transformed.

Statistical Methods

Identification of consistent DCEs

The procedure to identify consistently differentially correlated edges (DCEs) comprised three major steps: 1) interactions in String software [16] were used as a reference network (which slightly downsized the number of genes under consideration from 12,437 to 11,502), then Spearman correlation coefficients (SCCs) were calculated separately for the responder group and the non-responder group at each specific time point; 2) when the absolute difference of SCCs between two groups at a specific time point was larger than a pre-determined threshold (i.e., 0.6, this value may corresponds to two possible situations: either the correlation directions in two groups are the same with in one group being extremely large whereas in the other group being small; or the correlations in two groups are at least moderate but in different directions. In addition, with this step the number of gene pairs under consideration for p-value calculation was downsized to a manageable scale), the corresponding gene pair was deemed to be a DCE at that time point; and, 3) the intersection of DCEs across 5 time points was taken and the resulting subset was considered to represent consistently DCEs over time. A flowchart of the procedure is elucidated using a toy example with 5 genes (Figure 1).

Results
DCEs

In this study, the connection information in the String software was used as a reference. Only when the confidence score of a gene pair is larger than 0.6, these two genes are considered to be connected by an edge. The cutoff value for the absolute difference of SCCs between two groups is set at 0.6. Using the proposed procedure, 384 consistently differentially correlated edges (DCEs) over five time points were identified. These edges included 510 unique genes. By saying consistently, it is referred to the scenario that these edges were identified to be differentially correlated across all time points (i.e., the overlapped edges among these time points).

Using permutation tests and following the strategy proposed by[17] the significance levels of these gene pairs were determined and presented in Table 1 as well. Twenty-two DCEs are deemed to be statistically significant (with a p-value <0.01), which involve 41 unique genes (Table 1). Using these 41 genes as predictors, a linear support vector machine model was fit to the training set. The resulting model was applied to the validation set (an independent dataset from the training set) to calculate posterior probability of a MS patient being a responder at each time point. With 4 responders and 8 non-responders misclassified, a predictive accuracy of 80.95 % (51/63) was achieved by this 41-gene signature. Given that demographic heterogeneities may exist between patients in both experiments, and that the collection time points in the two studies differed hugely, with the early period of treatment being concentrated in the second study, this predictive accuracy is fairly good.

Discussion

The GeneCards database [18] indicates that of the 41-gene signature, 23 genes are directly associated to MS. Among these 23 genes, 7 genes (i.e., CXCL9, IL2RA, CXC3R, AKT1, CSF2, IL2RB and GCA) have a confidence score > 2 (Figure 2). Likewise, a SVM model was fit with these 7 genes as predictors and then applied to the validation set. This results showed that this 7-gene signature can achieve a predictive accuracy of 65.08 % (41/63) with 6 responders and 16 non-responders misclassified. Furthermore, these seven genes whose mean expression values for the responders and non-responders across time points are diagramed in Figure 3. According to the EDGE method [19], only AKT1 (adjusted p=0.003) and CXCL9 (adjusted p=0.012) are significantly differentially expressed over time between these two groups. Focusing specifically on the two extreme points, no genes are differentially expressed at the first time point while only CXCL9 is under-expressed in the responder group at the last time point according to Wilcoxon’s tests.
Next, a gene-to-gene interaction network of these 7 genes was constructed using the String software and presented in Figure 4. In this figure, we found that these genes were highly connected with all genes but GCA (which is an isolated gene) circling around one another to frame a loop. Interestingly, it was observed that the edges between IL2RA and IL2RB, and the edge between CXCR3 and CSF2 were negatively correlated to each other in the responder group while positively correlated in the non-responder group. Further study is highly desirable to explore if such correlation patterns have any impact on how a MS patient responds to the IFN-β treatment.

How these 7 genes are related to MS was mined by a PubMed literature searching. Since the autoimmune inflammatory process is believed to be essential for the development of MS, it is natural to observe that chemokines and chemokine receptors are involved in the pathogenesis of the disease [20]. For instance, a study [21] compared the chemokine receptor – CXCR3 – expression by gamma delta T cells derived from the blood and CSF of MS patients with health controls and observed its expression increased. Another study [22] also suggested significantly different expression of CXCR3 on peripheral blood lymphocytes in MS compared with controls.

Furthermore, CXCL9 is a ligand of CXCR3. A study [23] described two alternatively spliced variants of the human CXCR3-A receptor, termed CXCR3-B and CXCR3-alt. While human CXCR3-B binds CXCL9, CXCL10, CXCL11 as well as an additional ligand CXCL4, CXCR3-alt only binds CXCL11. A study by Mellergard et al. [24] suggested there was a marked decline in CSF levels of cytokines and chemokines in MS patients treated with natalizumab, thus including chemokines associated with both Th1 (CXCL9, CXCL10, CXCL11) and Th2 (CCL22). However, the chemokine system is complex and consists of numerous ligands and receptors. It is highly unlikely that targeting a particular simple chemokine or chemokine receptor will give beneficial effects.

A recent review [25] stated that in addition to IL7RA, IL2RA is a well-known immunogenetic marker of MS, explaining why IL2RA was ranked on the top of the directly related to MS list given by the GeneCards database [18]. Although the association between IL2RA and MS has been well established, the functional variation is still unknown. It is speculated that the effect of IL2RA on MS should be better described by several SNPs than by a single one [26], and they showed that the set of SNPs rs2256774 and rs3118470 provided a perfect discrimination between the MS patients and controls. Furthermore, there is a potential link between IL2RA and CXCR5 regarding MS. A study in China [27] showed that the genotypes and allele frequency distributions at the loci of IL2RA rs2104286 and rs12722489 were significantly different between the MS and control groups. Meanwhile, the gene polymorphisms at the loci of IL2RA rs2104286 and rs12722489 may
increase the onset risk of MS. IL2RA-rs2104286 showed a positive correlation with CXCR5-rs3922.

Using a meta-analysis, the association of polymorphisms in IL2RB, IL2RA and IL2 with MS were systemically reviewed, in which significant association for IL2RA and IL2 had been justified [28] and the association for IL2RB was deemed to be insignificant. No significant changes in serum IL-2 and soluble IL-2R levels were found between MS patients and controls, either [29].

The expression of CSF2, also known as GMCSF, by human TH cells has been reported to associated with MS disease severity. GMCSF is strongly induced by interleukin 2 (IL2) [30]; an MS-associated polymorphism in IL2RA specifically increased the frequency of GMCSF producing TH cells. The IL2RA polymorphism regulates IL2 responsiveness of naive TH cells and their propensity to develop into GMCSF producing memory TH cells. AKT is a critical mediator of growth factor-induced neuronal survival in the developing nervous system. Survival factors can suppress apoptosis in a transcription-independent manner by activating the serine/threonine kinase AKT1, which then phosphorylates and inactivates components of the apoptotic machinery. AKT activation relies on the PI3K pathway, and AKT1 is recognized as a critical node in the pathway. Furthermore, a recent study [31] found that compared with wild-type mice, AKT1-/- mice exhibited improved experimental autoimmune encephalomyelitis (EAE), a mouse model of MS. At the cellular level, AKT1 seems to inhibit the proliferation of thymogenic regulatory T cells, thereby promoting Ag-specific Th1/Th17 response. Additionally, Barca’s study [32] found that IFN-β promotes astrocyte survival by stimulating the PI3K/AKT pathway, and thus concluded that the beneficial effect of IFN-β in MS may depend partially on its ability to protect astrocytes from apoptotic cell death at the early stage of MS.

As far as GCA is concerned, Martinez et al. [33] examined four SNPs along the IFIH1-GCA-KCNH locus, rs13422767, rs2111485, rs1990760 and rs2068330 and reported an association of this locus with MS for the first time. Nevertheless, such an association had not been validated in another study [34]. The authors explained this discrepancy with three possible reasons: different ethnic background between the study population; different study designs used and the sample size issue.

Furthermore, all these 7 genes except AKT1 and IL2RB are IFN-responsive genes according to the Interferome database (http://www.interferome.org/). In summary, further investigation is warranted to examine how IFN-β treatment establishes its effects by targeting these 7 genes.
Conclusions

Our procedure identified consistent DCEs over time for a responder group in comparison with a non-responder group, and the resulting gene signature has explicit biological relevance to MS. Based on simple calculations of SCCs and their corresponding difference between two different phenotypes, our method is easy to comprehend and can be implemented by an entry-level statistician or a clinician. Therefore, we anticipate its widespread application in relevant research areas, to help researchers identify the underlying therapeutic mechanism of a specific regimen (as shown in the Supplementary Materials, there the proposed procedure was applied to analyze the gene expression data of GSE 41846 [35] in which treated MS patients and non-treated MS patients were compared) and predict which patients are more likely to respond.

The statistical procedure used in this study has several limitations. First, it excludes genes that are not in overlapping DCEs, therefore, valuable information may be overlooked, especially for the first time point. In the future, a way to use the first time point as the reference may be considered such that the overlap of other time points with this point will become the focus. Second, the training dataset had only 25 samples in total, and to eliminate possible ambiguous results caused by patients having relapses in the first 2-5 years, 7 MS patients were excluded from the downstream analysis. This results in a smaller sample size. A large gene expression experiment with a better study design is highly desirable.

Availability of data and materials

Microarray data (accession numbers: GSE24427 and GSE19285) were downloaded from the Gene Expression omnibus (GEO) repository (https://www.ncbi.nlm.nih.gov/geo/).

Competing interests

No competing interests have been declared.
Funding
This study was supported by the Education Department of Jilin Province (grant No. JJKH20190032KJ), the Natural Science Foundation of China (grants No. 81671177 and 81401123), and the Natural Science Foundation of Jilin Province Science and Technology Development Plan Project (No. 20190201043JC).

Author’s Contributions
Conceived and designed the study: ST. Analyzed the data: ST TJ. Interpreted data analysis and results: CW ST TJ. Wrote the paper: ST CW TJ. All authors reviewed and approved the final manuscript.

Acknowledgements
We thank Donna Gilbreath for scientific editing and assistance in manuscript preparation.
References

1. Browne P, Chandraratna D, Angood C, Tremlett H, Baker C, Taylor BV, Thompson AJ. Atlas of multiple sclerosis 2013: a growing global problem with widespread inequity. Neurology. 2014;83:1022–4.

2. Brownlee WJ, Hardy TA, Fazekas F, Miller DH. Diagnosis of multiple sclerosis: progress and challenges. Lancet. 2017;389:1336–46.

3. Vargas DL, Tyor WR. Update on disease-modifying therapies for multiple sclerosis. Journal of Investigative Medicine. 2017;65:883–91.

4. Nafee T, Watanabe R, Fregni F. Multiple sclerosis. In: Neuromethods. 2018. p. 263–295.

5. Garg N, Smith TW. An update on immunopathogenesis, diagnosis, and treatment of multiple sclerosis. Brain Behav. 2015;5(9):e00362.

6. Parnell GP, Gatt PN, Krupa M, Nickles D, McKay FC, Schibeci SD, Batten M, Baranzine S, Henderson A, Barnett M, Slee M. The autoimmune disease-associated transcription factors EOMES and TBX21 are dysregulated in multiple sclerosis and define a molecular subtype of disease. Clin Immunol. 2014;151 (1):16–24. doi:10.1016/j.clim.2014.01.003.

7. Hira ZM, Gillies DF. A review of feature selection and feature extraction methods applied on microarray data. Adv Bioinformatics. 2015;2015:198363.

8. Ang JC, Mirzal A, Haron H, Hamed HN. Supervised, unsupervised, and semi-supervised feature selection: A review on gene selection. IEEE/ACM Trans Comput Biol Bioinforma. 2016;13:971–989.

9. Kostka D, Spang R. Finding disease specific alterations in the co-expression of genes. Bioinformatics. 2004;20 SUPPL 1:i194–9.

10. Carter SL, Brechbühl CM, Griffin M, Bond AT. Gene co-expression network topology provides a framework for molecular characterization of cellular state. Bioinformatics. 2004;20:2242–50.

11. Fuller TF, Ghazalpour A, Aten JE, Drake TA, Lusis AJ, Horvath S. Weighted gene coexpression network analysis strategies applied to mouse weight. Mamm Genome. 2007;18:463–72.

12. Sun SY, Liu ZP, Zeng T, Wang Y, Chen L. Spatio-temporal analysis of type 2 diabetes mellitus based on differential expression networks. Sci Rep. 2013;3:2268.
13. Hundeshagen A, Hecker M, Paap BK, Angerstein C, Kandulski O, Fatum C, Hartmann C, Koczan D, Thiesen HJ, Zettl UK. Elevated type I interferon-like activity in a subset of multiple sclerosis patients: Molecular basis and clinical relevance. J Neuroinflammation. 2012;9:140.

14. Hecker M, Goertsches RH, Fatum C, Koczan D, Thiesen HJ, Guthke R, Zettl UK. Network analysis of transcriptional regulation in response to intramuscular interferon-β-1a multiple sclerosis treatment. Pharmacogenomics J. 2012;12 (2):134–46.

15. McCall MN, Bolstad BM, Irizarry RA. Frozen robust multiarray analysis (fRMA). Biostatistics. 2010;11:242–53.

16. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, Von Mering C, Jensen LJ. STRING v9.1: Protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res. 2013;41:D808–15.

17. Anglani R, Creanza TM, Liuzzi VC, Piepoli A, Panza A, Andriulli A, Ancona N. Loss of connectivity in cancer co-expression networks. PLoS One. 2014;9(1): e87075.

18. Safran M, Dalah I, Alexander J, Rosen N, Iny Stein T, Shmoish M, Nativ N, Bahir I, Doniger T, Krug H, S1rota-Madi A, Olinger T, Golan Y, Stelzer G, Harel A, Lancet D. GeneCards Version 3: the human gene integrator. Database (Oxford). 2010;2010:baq020.

19. Storey JD, Xiao W, Leek JT, Tompkins RG, Davis RW. Significance analysis of time course microarray experiments. Proc Natl Acad Sci. 2005;102:12837–42.

20. Szczuciński A, Losy J. Chemokines and chemokine receptors in multiple sclerosis. Potential targets for new therapies. Acta Neurol Scand. 2007;115:137–46.

21. Murzenok PP, Matsuevicius D, Freedman MS. γ/δ T cells in multiple sclerosis: Chemokine and chemokine receptor expression. Clin Immunol. 2002;103 3 1:309–16.

22. Mahad DJ, Lawry J, Howell SJL, Woodroofe MN. Longitudinal study of chemokine receptor expression on peripheral lymphocytes in multiple sclerosis: CXCR3 upregulation is associated with relapse. Mult Scler. 2003;9:189–98.

23. Korniejewska A, Mcknight AJ, Johnson Z, Watson ML, Ward SG. Expression and agonist responsiveness of CXCR3 variants in human T lymphocytes. Immunology. 2011;132:503–15.

24. Mellerård J, Edström M, Vrethem M, Ernerudh J, Dahle C. Natalizumab treatment in multiple sclerosis: Marked decline of chemokines and cytokines in cerebrospinal fluid. Mult Scler. 2010;16:208–17.

25. Leray E, Moreau T, Fromont A, Edan G. Epidemiology of multiple sclerosis. Rev Neurol (Paris). 2016;172:3–13. doi:10.1016/j.neuro.2015.10.006.
26. Babron MC, Perdry H, Handel AE, Ramagopalan S V., Damotte V, Fontaine B, Muller-Myhsok B, Ebers GC, Clerget-Darpoux F. Determination of the real effect of genes identified in GWAS: The example of IL2RA in multiple sclerosis. Eur J Hum Genet. 2012;20:321–5.

27. Xia ZL, Qin QM, Zhao QY. A genetic link between CXCR5 and IL2RA gene polymorphisms and susceptibility to multiple sclerosis. Neurol Res. 2018;40:1040–7. doi:10.1080/01616412.2018.1517110.

28. Cavanillas ML, Alcina A, Ñez C, De Las Heras V, Fernández-Arquero M, Bartolomé M, de la Concha EG, Fernandez O, Arroyo R, Matesanz F, Urcelay E. Polymorphisms in the IL2, IL2RA and IL2RB genes in multiple sclerosis risk. Eur J Hum Genet. 2010;18:794–9.

29. Ott M, Demisch L, Engelhardt W, Fischer PA. Interleukin-2, soluble interleukin-2-receptor, neopterin, l-tryptophan and β2-microglobulin levels in CSF and serum of patients with relapsing-remitting or chronic-progressive multiple sclerosis. J Neurol. 1993;241:108–14.

30. Hartmann FJ, Khademi M, Aram J, Ammann S, Kockum I, Constantinescu C, Gran B, Piehl F, Olsson T, Codarri L, Becher B. Multiple sclerosis-associated IL2RA polymorphism controls GM-CSF production in human TH cells. Nat Commun. 2014;5:5056.

31. Ouyang S, Zeng Q, Tang N, Guo H, Tang R, Yin W, Wang A, Tang H, Zhou J, Xie H, Langdon WY, Yang H, Zhang J. Akt-1 and Akt-2 Differentially Regulate the Development of Experimental Autoimmune Encephalomyelitis by Controlling Proliferation of Thymus-Derived Regulatory T Cells. J Immunol. 2019;202(5):1441–52.

32. Barca O, Ferré S, Seoane M, Prieto JM, Lema M, Señarís R, Arce VM. Interferon beta promotes survival in primary astrocytes through phosphatidylinositol 3-kinase. J Neuroimmunol. 2003;139:155–9.

33. Martínez A, Santiago JL, Cénit MC, de las Heras V, de la Calle H, Fernández-Arquero M, Arroyo R, de la Concha EG, Urcelay E. IFIH1-GCA-KCNH7 locus: Influence on multiple sclerosis risk. Eur J Hum Genet. 2008;16(7):861–4.

34. Couturier N, Gourraud PA, Cournu-Rebeix I, Gout C, Bucciarelli F, Edan G, Babron MC, Clerget-Darpoux F, Clanet M, Fontaine B, Brassat D. IFIH1-GCA-KCNH7 locus is not associated with genetic susceptibility to multiple sclerosis in French patients. Eur J Hum Genet. 2009;17(6):844–7.

35. Nickles D, Chen HP, Li MM, Khankhanian P, Madireddy L, Caillier SJ, Santaniello A, Cree BA, Pelletier D, Hauser SL, Oksenberg JR, Baranzine SE. Blood RNA profiling in a large cohort of multiple sclerosis patients and healthy controls. Hum Mol Genet. 2013;22(20):4194–205.


**Figure 1. Flowchart of the proposed procedure.** (A) Identification of differentially correlated edges. (B) Generation of consistently differentially correlated edges. Here, a toy-example with 5 genes was used to graphically elucidate the proposed procedure. First, a reference network was created according to the String software. Only the interactions recorded in this software were considered. Then the Spearman correlation coefficients for these gene pairs were calculated for the responder group and the non-responder group, respectively. If the absolute value of the correlation coefficient difference between these two groups is larger than a predetermined cutoff, say, 0.6, the corresponding edge is deemed as differentially correlated edge (DCE). Those common DCEs across time points are consistently DCEs. Then the corresponding p-values for these DCEs were calculated using permutation tests. Upon the genes involved in the statistically significant DCEs, the biological relevance was investigated.

**Figure 2. Venn-diagram showing 41 genes involved in the statistically significant DCEs.** Among them, 23 genes are indicated to be directly related to multiple sclerosis, whose symbols are given in the Venn-diagram. 8 of 23 genes have a confidence score > 2, which are highlighted in red. DCEs: consistently differentially correlated edges.

**Figure 3. Mean expression value trajectories of the identified 7-gene list for the responder and the non-responder.** (A) For CXCL9. (B) For IL2RA. (C) For CXCR3. (D) For AKT1. (E) For CSF2. (F) For IL2RB. (G) For GCA. Here, R: the responder group; NR: the non-responder group.

**Figure 4. Gene-to-gene interaction network of the identified gene list.** (A) The responder group. (B) The non-responder group. Here, + stands for the correlation coefficient between the corresponding gene pair is positive; - stands for the correlation coefficient between the corresponding gene pair is negative.
Figure 1

Flowchart of the proposed procedure

(A) Identification of differentially correlated edges. (B) Generation of consistently differentially correlated edges.
A

Reference network based on the String software

Correlation matrix

B

Consistently differentially correlated edges (CDREs)

Unique genes involved in these CDREs
Figure 2

Venn-diagram showing 41 genes involved in the statistically significant DCEs.
Figure 3

Mean expression value trajectories of the identified 7-gene list for the responder and the non-responder.

(A) For CXCL9. (B) For IL2RA. (C) For CXCR3. (D) For AKT1. (E) For CSF2. (F) For IL2RB. (G) For GCA.
Figure 4

Gene-to-gene interaction network of the identified gene list.

(A) The responder group. (B) The non-responder group.
Table 1 (on next page)

Consistent edge differences (network differences) across five time points
|       |       | T1  | T2  | T3  | T4  | T5  | p-value |
|-------|-------|-----|-----|-----|-----|-----|---------|
| ADCY1 | AKT1  | -0.567 | 0.65 | -0.717 | 0.7 | -0.383 | 0.667 | -0.883 | 0.283 | -0.683 | 0.167 | 0.0001 |
| FKB10 | HTR5A | 0.633 | -0.4 | 0.717 | -0.433 | 0.833 | -0.033 | 0.617 | -0.217 | 0.967 | -0.083 | 0.0087 |
| DCAF13 | KRR1  | -0.533 | 0.467 | -0.133 | 0.6 | -0.6 | 0.35 | 0.533 | -0.5 | -0.7 | 0.217 | 0 |
| ATRX  | XRCC6 | 0.233 | -0.533 | 0.283 | -0.55 | -0.75 | -0.117 | 0.517 | -0.467 | 0.567 | -0.5 | 0.0095 |
| CSF2  | CXCR3 | -0.1 | -0.883 | 0.3 | -0.7 | -0.033 | -0.783 | 0.283 | -0.6 | 0.333 | -0.767 | 0.0083 |
| FKBP10| PNLIP | 0.033 | 0.783 | 0.533 | -0.267 | 0.317 | -0.567 | 0.8 | 0.1 | 0.433 | -0.5 | 0.0059 |
| EP300 | GTF2H5 | 0.933 | -0.3 | 0.633 | -0.267 | 0.817 | -0.733 | 0.617 | -0.5 | 0.817 | -0.017 | 0.0024 |
| EP300 | POLR2K | 0.817 | -0.35 | 0.783 | -0.117 | 0.8 | -0.733 | 0.65 | -0.533 | 0.833 | -0.167 | 0.0094 |
| UBE2D2 | UBXN7 | 0.467 | -0.367 | 0.467 | -0.267 | 0.633 | -0.533 | 0.617 | -0.033 | 0.717 | -0.15 | 0.0068 |
| ALMS1 | CEP290 | -0.683 | 0.25 | -0.617 | 0.567 | -0.717 | -0.05 | 0.733 | -0.617 | 0.783 | -0.05 | 0.0047 |
| RSRC1 | SNU13 | -0.683 | 0.433 | -0.683 | 0.433 | -0.45 | 0.333 | -0.633 | 0.667 | -0.217 | 0.717 | 0.0023 |
| BCR   | CUX1  | 0.217 | -0.5 | 0.633 | -0.767 | 0.533 | -0.533 | 0.35 | -0.65 | 0.333 | -0.367 | 0.007 |
| FGF21 | FLT1  | 0.6 | -0.433 | 0.717 | -0.517 | 0.183 | -0.667 | 0.983 | -0.417 | 0.833 | -0.117 | 0.0014 |
| DYNC1H1 | TUBGCP2 | -0.7 | 0.3 | -0.767 | 0.817 | -0.767 | 0.317 | -0.55 | 0.3 | -0.467 | 0.517 | 0.0006 |
| CXCL9 | GRM4  | 0.1 | -0.85 | 0.05 | -0.583 | 0.25 | -0.5 | 0.767 | -0.083 | 0.8 | -0.167 | 0.0007 |
| DYNC1H1 | PCNT  | 0.483 | -0.75 | 0.4 | -0.933 | 0.267 | -0.05 | 0.567 | -0.06 | 0.4 | -0.367 | 0.004 |
| CROCC | PCNT  | -0.467 | 0.617 | -0.283 | 0.817 | -0.45 | 0.683 | -0.517 | 0.733 | -0.483 | 0.233 | 0.0079 |
| IL2RA | IL2RB | -0.5 | 0.317 | -0.583 | 0.017 | -0.75 | 0.15 | -0.783 | 0.517 | -0.467 | 0.433 | 0.0072 |
| MELK  | RFC4  | -0.367 | 0.683 | -0.117 | 0.533 | -0.633 | 0.617 | -0.033 | 0.583 | -0.867 | 0.133 | 0 |
| HNRNPA2B | 1  | -0.733 | 0.233 | -0.767 | 0.4 | -0.617 | 0.2 | -0.683 | 0.017 | -0.9 | -0.183 | 0.0026 |
| ADCY3 | NUDT2 | 0.6 | -0.267 | 0.483 | -0.65 | 0.417 | -0.667 | 0.65 | -0.433 | 0.25 | -0.517 | 0.0095 |
| AGA   | GCA   | 0.817 | -0.317 | 0.817 | -0.133 | 0.8 | -0.65 | 0.617 | -0.617 | 0.633 | -0.45 | 0.0099 |

Note: R: responders; NR: non-responders. The gene pairs fit two conditions: 1. The difference for Spearman correlation coefficient between responders and non-responders >0.6 and 2. the corresponding p-value <0.01.