Investigation of multiple sclerosis-related pathways through the integration of genomic and proteomic data

Elif Everest¹, Ege Ülgen², Ugur Uygunoglu³, Melih Tutuncu³, Sabahattin Saip³, Osman Uğur Sezerman², Aksel Siva³ and Eda Tahir Turanlı⁴

¹ Department of Molecular Biology and Genetics, Faculty of Science and Letters, Istanbul Technical University, Istanbul, Turkey
² Department of Biostatistics and Medical Informatics, Faculty of Medicine, Acıbadem University, Istanbul, Turkey
³ Department of Neurology, Cerrahpaşa School of Medicine, Istanbul University-Cerrahpaşa, Istanbul, Turkey
⁴ Department of Molecular Biology and Genetics, Faculty of Science and Letters, Acıbadem University, Istanbul, Turkey

ABSTRACT

Background. Multiple sclerosis (MS) has a complex pathophysiology, variable clinical presentation, and unpredictable prognosis; understanding the underlying mechanisms requires combinatorial approaches that warrant the integration of diverse molecular omics data.

Methods. Here, we combined genomic and proteomic data of the same individuals among a Turkish MS patient group to search for biologically important networks. We previously identified differentially-expressed proteins by cerebrospinal fluid proteome analysis of 179 MS patients and 42 non-MS controls. Among this study group, 11 unrelated MS patients and 60 independent, healthy controls were subjected to whole-genome SNP genotyping, and genome-wide associations were assessed. Pathway enrichment analyses of MS-associated SNPs and differentially-expressed proteins were conducted using the functional enrichment tool, PANOGA.

Results. Nine shared pathways were detected between the genomic and proteomic datasets after merging and clustering the enriched pathways. Complement and coagulation cascade was the most significantly associated pathway (hsa04610, \( P = 6.96 \times 10^{-30} \)). Other pathways involved in neurological or immunological mechanisms included adherens junctions (hsa04520, \( P = 6.64 \times 10^{-25} \)), pathogenic Escherichia coli infection (hsa05130, \( P = 9.03 \times 10^{-14} \)), prion diseases (hsa05020, \( P = 5.13 \times 10^{-13} \)).

Conclusion. We conclude that integrating multiple datasets of the same patients helps reducing false negative and positive results of genome-wide SNP associations and highlights the most prominent cellular players among the complex pathophysiological mechanisms.

Subjects Bioinformatics, Genomics, Molecular Biology, Neurology

Keywords Bioinformatics, Genomics, Proteomics, Multiple sclerosis

INTRODUCTION

Multiple sclerosis (MS) is known to be an immune-mediated, neurodegenerative central nervous system (NS) disorder with complex inheritance and pathophysiological
mechanisms. Although approximately common 250 genetic variants with low to modest risk effects have been associated with MS mainly by genome-wide association studies (GWAS) using rather large sample groups (International Multiple Sclerosis Genetics Consortium, 2007; Sawcer et al., 2011; Patsopoulos et al., 2017; International Multiple Sclerosis Genetics Consortium et al., 2019), known variants not only fail to explain predicted MS heritability but also cannot be efficiently translated into disease mechanisms. To date, numerous studies have also revealed potential diagnostic and prognostic biomarkers and disease-related cellular pathways that emphasize the different pathological components of the disease; however, the exact underlying mechanisms in disease development and progression are largely unknown. In order to better translate the growing number of findings into disease pathophysiology, algorithms for pathway analyses of multiple high-throughput omics data seem essential. In this context, the integration of multiple omics data is essential to better describe the complex nature of MS.

In our previous study (Avsar et al., 2015), we have conducted a cerebrospinal fluid (CSF) proteome analysis using 2D-gel electrophoresis and mass spectrophotometry and identified 151 differentially expressed proteins between an MS cohort of 179 patients with different clinical MS phenotypes and 42 non-MS controls. Later, affected Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were identified using the functional enrichment tool Pathway and Network- Oriented GWAS Analysis (PANOGA) (Bakir-Gungor, Egemen & Sezerman, 2014), revealing MS-related pathways including aldosterone-regulated sodium reabsorption pathway, renin- angiotensin system, notch signaling pathway, and vitamin digestion and absorption pathway. Here, we further explored disease-related pathways, applying single nucleotide polymorphism (SNP) genotyping on 11 MS patients, who were included in the previous proteomic study, and 60 independent, healthy individuals. Pathway enrichment analyses of the genomic and proteomic data were conducted. The two datasets were merged, highlighting the most prominent pathways that may be affected in the studied patient group.

MATERIALS & METHODS

Study groups

All patients were diagnosed at Istanbul University-Cerrahpaşa, Cerrahpaşa School of Medicine, Department of Neurology, according to the McDonald criteria (Polman et al., 2011). Eleven unrelated MS patients were randomly selected among the study group of our previous work comprising of 179 MS patients. The patient group had a heterogeneous disease presentation at the time of their CSF sample collection. Sixty independent age- and gender-matched intrinsic healthy controls were included in the analyses. All individuals in the study group are of Turkish origin. All procedures of the study were in accordance with the Helsinki Declaration of 1964 and its later amendments. The Ethics Committee of Istanbul Technical University approved the study (İTÜ-SM.INAREK-MBG-1), and each individual of the study group gave a written informed consent form prior to the sample collection.
**Genome-wide associations**

DNA isolation from blood samples was performed (Roche DNA Isolation Kit for Mammalian Blood), and genotyping for 300,000 SNP markers was applied for each individual on the Illumina HumanCytoSNP-12 array. Illumina GenomeStudio software was used for quality control. All individuals showed sample call rates of more than 98% (98.65–99.63%) and were therefore included in the study. Golden Helix SNP & Variation Suite software was used for identity-by-descent detection, suggesting no relatedness between the individuals. SNP filtering was performed: SNPs on the Y chromosome, with call rates lower than 95%, minor allele frequency lower than 0.01, and in strong linkage disequilibrium ($r^2 > 0.5$) were excluded from the study. A total of 129,547 SNPs was included in the analysis. Frequency differences of SNPs between cases and controls were assessed, and genotypic association $P$-values were obtained.

**Pathway enrichment analysis**

Genotypic associations and proteomic data were subjected to pathway enrichment analyses using the functional enrichment tool PANOGA, which reveals functionally relevant pathways by identifying genes within the pathways, incorporating protein-protein interaction (PPI) information, and extracting significant pathways. SNPs with genotypic association $P$-values lower than 0.05 and differentially-expressed proteins with $P$-values lower than 0.05 from our previous study were used for PANOGA procedure as follows: For SNP associations, because a SNP might affect more than one gene, each SNP was initially assigned to the gene on which the SNP has the most important functional effect, using the SPOT webserver (Saccone et al., 2010). Functional information of SNPs was obtained utilizing SPOT, F-SNP (Lee & Shatkay, 2007), SNPnexus (Dayem Ullah, Lemoine & Chelala, 2012), and SNPinfo (Xu & Taylor, 2009). Genes and proteins were then mapped onto a PPI network, for which Goh et al.’s human PPI network was used (Goh et al., 2007). Next, the jActive Modules algorithm (Ideker et al., 2002) was applied to identify active subnetworks containing a large number of the disease-affected genes and proteins in the PPI network. Each genotypic association and differential expression $P$-value was taken into account, and active subnetworks that overlap at most 50% with each other were extracted. In order to evaluate the biological importance of the subnetworks, the number of genes and proteins found in a specific KEGG pathway was compared to the total gene and protein number involved in the corresponding pathway. For this functional enrichment procedure, a two-sided hypergeometric test was used, and the Bonferroni method was applied for multiple testing corrections of the $P$-values. Significantly associated KEGG pathways to our patient group consisted of significantly enriched ($P < 0.05$) pathways for at least one of the active subnetworks. For the pathways that were enriched in multiple subnetworks, the one with the lowest $P$-value was reported. PANOGA was run 10 times for both genotypic associations and differentially-expressed proteins, and the lowest $P$-values over the 10 iterations were reported. The resulting enriched pathways for both datasets were then merged. If a given pathway was identified in both analyses, the corresponding $P$-values were merged using Fisher’s combined probability test. If the pathway was identified in only one of the analyses, the corresponding $P$-value was used as the final $P$-value.
Clustering of enriched pathways

Using a method previously described by Chen et al. (2014) we clustered the enriched pathways to identify similar groups and establish representative pathways. The clustering approach can be summarized as follows: initially overlap index matrix (OI) that consists of the overlap indices between all of the pairs of enriched pathways was calculated. The overlap index \( OI_{i,j} \) of a pair of pathways \( P_i \) and \( P_j \) (denoting the \( i \)th and \( j \)th pathways, respectively) was defined as:

\[
OI_{i,j} = \frac{|G_i \cap G_j|}{\min(|G_i|, |G_j|)}
\]

where \( G_i \) is the set of all genes in the \( i \)th pathway. In the overlap matrix, each row \( o_i \) represents the gene overlap profile of a pathway (i.e., \( o_i \) is the gene overlap profile of the \( i \)th pathway). To identify similarity between each pair of pathways Pearson correlation coefficients (R) between each pair of overlap profiles (e.g., \( o_i \) and \( o_j \)) were calculated. These correlation coefficients were then converted to pairwise distances (PD) as:

\[
PD = 1 - R.
\]

Using PD as the distance metric, hierarchical clustering (average-linkage) of the enriched pathways was performed. Examining the hierarchical clustering dendrogram, the dendrogram was partitioned into clusters at the manually selected PD cut-off value of 0.55. The pathway with the lowest \( P \)-value in each cluster was selected as the representative pathway for that cluster.

RESULTS

SNP associations

All 71 individuals and a total of 129,547 SNPs passed the quality check and therefore were included in the final analyses. The Manhattan plot summarizing the genome-wide associations is given in Fig. 1. The most significantly associated SNP was rs7873, located on 3’ UTR of the \( IGF2 \) gene, whose significance level was the only value close to a classical stringent \( P \)-value cut-off for a GWAS (\( P = 4.39 \times 10^{-07} \)). When considering SNPs with a significance level of lower than a looser \( P \)-value cut off (\( P < 10^{-05} \)) given the small sample size, 4 more SNPs have high associations with MS in the studied patient group, one of which is located on a gene: rs17187282 (\( P = 1.17 \times 10^{-06} \)), rs11688088 (\( P = 2.8 \times 10^{-06} \)), rs654188 (\( P = 5.24 \times 10^{-06} \)), and rs7092208 (\( P = 9.66 \times 10^{-06} \), intronic variant on the \( MGMT \) gene).

For enrichment analysis of the SNP associations, the \( P \)-value threshold was set lower than 0.05 (Fig. 1, blue line) in order to prevent the elimination of potentially MS-associated SNPs with falsely high \( P \)-values due to the low sample size and limitations of the genome-wide association methodology. Later, elimination of the false-positive SNP associations resulted from preferring not to set a conventional stringent \( P \)-value was aimed to be achieved during the identification of MS-related pathways using SNP subnetworks, therefore ruling functionally irrelevant SNPs out. A total of 6,594 SNPs had \( P \)-values under the threshold of 0.05 (Table S1), which were included in the subsequent analyses.
MS-related pathways

The combination of enriched pathways for the resulting SNPs and differentially expressed proteins resulted in 151 enriched pathways in total (Table S2). In order to identify pathways with similar content and function, this combined list of enriched pathways was clustered (Fig. 2). The dendrogram was then manually partitioned into biologically relevant clusters, and representative pathways were established. In total, 33 clusters, therefore 33 representative pathways were obtained (Table S3). Nine of the representative pathways emerged from both genomic and proteomic analyses, among which the complement and coagulation cascade (hsa04610) was the most significantly associated pathway in the studied group ($P = 6.96 \times 10^{-30}$). Figure 3 shows the complement and coagulation cascades with genes identified through the genomic dataset and differentially expressed proteins identified through the proteomic dataset as a representative pathway for the relationships between the genomic and proteomic findings.

Regulation of actin cytoskeleton (hsa04810) and focal adhesions (hsa04510) were the two systems with the second and third lowest $P$-values even though there were no significant protein level changes related to these pathways ($P = 9.64 \times 10^{-27}$ and $P = 3.29 \times 10^{-26}$, respectively). Other 8 shared pathways with high significance levels are as follows: Adherens junctions (hsa04520, $P = 5.38 \times 10^{-17}$), colorectal cancer (hsa05210, $P = 1.70 \times 10^{-15}$), pathogenic Escherichia coli infection (hsa05130, $P = 7.02 \times 10^{-9}$), prion diseases (hsa05020, $P = 2.48 \times 10^{-5}$), endometrial cancer (hsa05213, $P = 6.00 \times 10^{-9}$), non-small cell lung cancer (hsa05223, $P = 1.40 \times 10^{-8}$), bladder cancer (hsa05219, $P = 4.83 \times 10^{-6}$), and non-homologous end-joining (hsa03450, $P = 4.27 \times 10^{-5}$). Table 1 shows the detected SNP associations and differentially expressed proteins in the above-mentioned pathways.
Figure 2  Clustering dendrogram of the enriched pathways. The dendrogram shows all 151 enriched pathways, which were grouped in clusters based on their relevance to each other. The red line shows the cut-off for biologically relevant clusters, which was cut at the pairwise distance of 0.55, resulting in 33 representative clusters. Thirty-three pathways with the lowest $P$-values in each cluster constituted the representative pathways.

Figure 3 Complement and coagulation cascades. The illustration of complement and coagulation cascades shows the identified genes by SNP associations (blue) and differentially expressed proteins (yellow) as a representative pathway for the merged data analysis. The pathway was adapted from KEGG: Kyoto Encyclopedia of Genes and Genomes (Kanehisa & Goto, 2000; hsa04610, 2019) and created with BioRender.com.

The most significantly MS-associated pathways obtained from the SNP associations and differentially expressed proteins did not overlap when the analyses were performed.
### Table 1: The most prominent MS-related pathways emerged from the analyses.

| KEGG terms and IDs | SNP associations | Differentially expressed proteins | Final P-value |
|--------------------|------------------|------------------------------------|---------------|
| **Complement and coagulation cascades, hsa04610** | **FGA [rs2070018]; CR1 [rs1571344]; F13A1 [rs3901123]; rs34736558; rs1267843; rs1267914; rs749005]; PLG [rs9295131]; F5 [rs12131397]; C6 [rs2921184; rs1801033]; PROC [rs2069933; rs1799808]** | | **8.43 × 10^{-16}** | **6.96 × 10^{-30}** |
| **Adherens junction, hsa04520** | **INSR [rs10409516]; LMO7 [rs9573625]; WASL [rs10270793]; EGFR [rs6970262]; IGF1R [rs2684788]; FER [rs7715933; rs7700636; rs7713591]; CTNNAA3 [rs10997250; rs7076094; rs10997316]; RAC1 [rs2347338]; PVRL2 [rs519825]** | | **2.03 × 10^{-10}** | **6.64 × 10^{-25}** |
| **Colorectal cancer, hsa05210** | **TCF7L1 [rs6547608; rs10195517]; TGB2 [rs1473527; rs1417488]; AXIN1 [rs3916990]; MAPK9 [rs6867398; rs17080136; rs11956696]; MAPK8 [rs17780725]; RAC1 [rs2347338]** | | **2.88 × 10^{-9}** | **2.68 × 10^{-22}** |
| **Pathogenic Escherichia coli infection, hsa05130** | **ROCK2 [rs6432187]; WASL [rs10270793]** | | **3.72 × 10^{-7}** | **9.03 × 10^{-14}** |
| **Prion diseases, hsa05020** | **NOTCH1 [rs3812605; rs3812609; rs2229974]; NCAM2 [rs232398]** | | **6.31 × 10^{-10}** | **5.13 × 10^{-13}** |
| **Endometrial cancer, hsa05213** | **TCF7L1 [rs6547608; rs10195517]; AXIN1 [rs3916990]; EGFR [rs6970262]; CTNNAA3 [rs10997250; rs7076094; rs10997316]; SOS2 [rs2227276]** | | **5.77 × 10^{-6}** | **1.11 × 10^{-12}** |
| **Non-small cell lung cancer, hsa05223** | **ALK [rs7591913; rs7589120]; FHIT [rs643629; rs10510827; rs10510852; rs772479; rs6778312; rs12633994; rs13317933; rs13078190]; SOS2 [rs2227276]; EGFR [rs6970262]** | | **2.82 × 10^{-6}** | **1.26 × 10^{-12}** |
| **Bladder cancer, hsa05219** | **E2F3 [rs12527393]** | | **8.47 × 10^{-7}** | **1.11 × 10^{-10}** |
| **Non-homologous end-joining, hsa03450** | **XRCC4 [rs2662242]** | | **9.88 × 10^{-6}** | **9.53 × 10^{-9}** |

**Top 2 pathways emerged only from SNP associations**

| Regulation of actin cytoskeleton, hsa04810 | CYFIP1 [rs7182576; rs2120968; rs17137192]; ROCK2 [rs6432187]; WASL [rs10270793]; EGFR [rs6970262]; MYLK [rs678060; rs3911406; rs2124508; rs702032]; GNA13 [rs2994877]; PAK7 [rs618867; rs618709]; RAC1 [rs2347338]; PAK2 [rs6583176]; PAK4 [rs11083505; rs692364]; VAV3 [rs11576720]; ITGA4 [rs3770111]; ACTN1 [rs181484]; ITGAA2 [rs27504]; FNI [rs1404772]; VAV2 [rs7026263]; TIAM1 [rs2833359; rs845972; rs998449; rs11700792]; ARHGEF4 [rs2403238]; MYH9 [rs11703176]; CRK [rs6502707]; DOCK1 [rs10741150]; SOCS2 [rs482853; rs10458718]; ARHGEF6 [rs4132580]; ITGA9 [rs149815; rs1984311] | | – | **9.64 × 10^{-27}** |

(continued on next page)
### Table 1 (continued)

| KEGG terms and IDs | SNP associations | Differentially expressed proteins | Final P-value |
|--------------------|------------------|----------------------------------|---------------|
| Focal adhesion, hsa04510 | $3.29 \times 10^{-26}$ | $FIGF$ [rs7877192; rs6632528]; $SHC3$ [rs1331189]; $LAMA1$ [rs658121]; $IGF1R$ [rs2684788]; $MAPK9$ [rs6867398; rs17080136; rs11956696]; $MAPK8$ [rs17780725]; $KDR$ [rs7692791]; $ITGBe$ [rs6730023]; $RAC1$ [rs2347338]; $PAK4$ [rs11083505; rs692364]; $VASP$ [rs4803831]; $VAV3$ [rs1157620]; $CAV3$ [rs111713611]; $ITGGA4$ [rs3770111]; $ACTN1$ [rs181484]; $ITGAA2$ [rs27504]; $FN1$ [rs13044772]; $MAPK10$ [rs1469869]; $ITGAA10$ [rs12401622]; $ITGAA8$ [rs2039910; rs1451668]; $RAPGEF1$ [rs2296950]; $TLN1$ [rs2295794]; $CRK$ [rs6502707]; $DOCK1$ [rs10741150; rs6482855; rs10458718]; $SOS2$ [rs2227276]; $ITGAA9$ [rs149815; rs1984311] | – | $3.29 \times 10^{-26}$ |

### Table 2  Top five pathways emerged from each dataset analysis.

| KEGG Terms and IDs | P-value |
|--------------------|---------|
| **Top 5 pathways emerged from the proteomic-only data analysis** | |
| Complement and coagulation cascades, hsa04610 | $8.43 \times 10^{-16}$ |
| Adherens junction, hsa04520 | $2.03 \times 10^{-10}$ |
| Prion diseases, hsa05020 | $6.31 \times 10^{-10}$ |
| Colorectal cancer, hsa05210 | $2.88 \times 10^{-09}$ |
| Pathogenic *Escherichia coli* infection, hsa05130 | $3.72 \times 10^{-07}$ |
| **Top 5 pathways emerged from the genomic-only data analysis** | |
| Regulation of actin cytoskeleton, hsa04810 | $9.64 \times 10^{-27}$ |
| Focal adhesion, hsa04510 | $3.29 \times 10^{-26}$ |
| ErbB signaling pathway, hsa04012 | $1.28 \times 10^{-23}$ |
| Rap1 signaling pathway, hsa04015 | $2.35 \times 10^{-20}$ |
| Axon guidance, hsa04360 | $6.36 \times 10^{-20}$ |
| **Top 5 pathways emerged from the merged data analysis** | |
| Complement and coagulation cascades, hsa04610 | $6.96 \times 10^{-30}$ |
| Regulation of actin cytoskeleton, hsa04810 | $9.64 \times 10^{-27}$ |
| Focal adhesion, hsa04510 | $3.29 \times 10^{-26}$ |
| Adherens junction, hsa04520 | $6.64 \times 10^{-25}$ |
| ErbB signaling pathway, hsa04012 | $1.28 \times 10^{-23}$ |

separately (Table S2). The merged data analysis resulted in a combined list of pathways from both datasets, revealing a different set of pathways with the lowest P-values and emphasizing the most relevant pathways that could not have been prioritized by a straightforward, non-combinatorial omics approach. The top five pathways that emerged from the proteomic-only, genomic-only, and merged dataset analyses are given in Table 2.
DISCUSSION

The complement and coagulation cascade (CCC) pathway emerged as the most significantly associated pathway to MS in our patient group, with many SNP associations on different genes and differentially expressed CSF proteins as shown in Fig. 3. The complement and the coagulation systems are two closely linked cascades, both having roles in immunity. In a study by Magliozzi et al. (2019), CSF proteomic profiles of MS patients with low or high cortical lesion load were compared, revealing that the identified differentially expressed proteins were mainly involved in the complement and coagulation cascade. Examination of white matter lesions of MS patients has revealed dysregulation of coagulation-associated proteins in chronic active plaques involving Serpin A5 and tissue factor (Han et al., 2008). Also, significantly upregulated Serpin E1, another CCC component, was reported in the post-mortem cortex of progressive MS patients (Yates et al., 2017). In our study, even though the genes and differentially expressed proteins involved in CCC do not overlap, the pathway was found 10 times for each dataset in PANOGA, resulting in a high association with MS. Our results confirm the previous findings suggesting the involvement of CCC alterations in MS pathophysiology and highlight other pathway components that may also be responsible for these alterations.

Disruption of blood–brain barrier (BBB) integrity is one of the hallmarks in MS pathophysiology, during which massive leukocyte infiltration occurs across the damaged BBB into the CNS (Alvarez, Cayrol & Prat, 2011). Tight and adherens junctions between the endothelial cells of BBB have significant importance for normal immune surveillance in the CNS. Tight junctions consist of claudins, occludin, junctional adhesion molecules, and Zonula Occludens (Hawkins & Davis, 2005), whose dysfunctions leading to BBB abnormalities in MS have been previously reported (Kirk et al., 2003; Padden et al., 2007; Plumb et al., 2002). Cadherins form adherens junctions via homophilic interactions between endothelial cells and bind cytoskeletal components through cytoplasmic catenin proteins (Dejana, Orsenigo & Lampugnani, 2008). Although adherens junctions are required for the overall junctional organization to maintain the BBB integrity, their roles in normal physiology and diseased states have not been well-established. In a previous study, the level of an adherens junction protein, β-catenin, was similar in progressive MS and non-MS brain sections (Padden et al., 2007). In our patient group, adherens junctions, which were not detected in our previous proteomic study, emerged as the second most significantly altered cellular components with SNP associations on different genes and differentially expressed proteins indirectly related to the pathway. Our findings indicate that adherens junctions may have more influence on BBB function than it has been thought and are needed to be explored in both normal physiology and MS pathophysiology in more detail.

Focal adhesions and actin cytoskeleton regulation are two interrelated pathways that emerged with low final P-values from the analyses, even though no significant protein expression changes directly or indirectly related to these systems were detected. Organized clusters of focal adhesion complexes connect extracellular matrix through transmembrane integrin proteins to the intracellular actin cytoskeleton through other focal adhesion...
proteins in the complexes and mainly have roles in cell motility (Wozniak et al., 2004). During leukocyte infiltration across the BBB, α4-integrin proteins on leukocytes form firm connections with the endothelial surface, initiating the process (Berlin et al., 1995; Vajkoczy, Laschinger & Engelhardt, 2001). In this context, Natalizumab (Antegren, Elan Pharmaceuticals and Biogen), a monoclonal antibody against α4-integrin, has been used in MS treatment, which suppresses inflammatory activity by inhibiting leukocyte migration to the inflammation areas (Polman et al., 2006). We detected many SNP associations on genes encoding for many focal adhesion proteins, including the α4-integrin-coding ITGA4 gene. Focal adhesion molecules and their combinations to create different complexes are diverse, and regulation differences of these interactions for specific cell behaviors are yet to be elucidated.

In the studied MS patients, increased CSF nucleolin expression and SNP associations on WASL and ROCK2 genes were detected, all of which are components of the Pathogenic Escherichia coli infection pathway. Neural Wiskott-Aldrich syndrome protein encoded by WASL and Rho-associated protein kinase 2 encoded by ROCK2 are both regulators of the actin cytoskeleton (Miki, Miura & Takenawa, 1996; Gallo et al., 2012), and the involvement of the pathway may indicate induction of inflammatory cell mobility. Prion diseases are fatal conditions presenting neuronal degeneration, and the emergence of the pathway from our analyses emphasizes some shared molecular alterations with MS pathophysiology, as both neurodegeneration and neuroinflammation also occur in prion diseases (Perry, Cunningham & Boche, 2002; Eikelenboom et al., 2002). We have also detected a number of cancer pathways involving both SNP associations and differential CSF expressions. Cancer risk among MS patients has been investigated in a number of studies with diverse results, revealing unchanged (Nielsen et al., 2006), reduced (Bahmanyar et al., 2009), and increased (Grytten et al., 2020) cancer rates. Altered cancer pathways in our study group may be attributed to changes in immunological mechanisms involving the same components as in mechanisms against anti-tumor surveillance. Other pathways that emerged only from the SNP associations included a number of immunological and neurological mechanisms. T cell receptor signaling pathway (hsa04660, $P = 1.86 \times 10^{-19}$), chemokine signaling pathway (hsa04062, $P = 4.38 \times 10^{-19}$), and Fc-gamma R-mediated phagocytosis (hsa04666, $P = 8.78 \times 10^{-16}$) were the immune system-related pathways. Nervous system-related pathways included axon guidance (hsa04360, $P = 6.36 \times 10^{-20}$), retrograde endocannabinoid signaling (hsa04723, $P = 6.74 \times 10^{-18}$), neurotrophin signaling pathway (hsa04722, $P = 8.03 \times 10^{-18}$), glutamatergic synapse (hsa04724, $P = 1.74 \times 10^{-15}$), cholinergic synapse (hsa04725, $P = 7.79 \times 10^{-15}$), and dopaminergic synapse (hsa04728, $P = 5.83 \times 10^{-12}$). Also, the Rap1 signaling pathway (hsa04015, $P = 2.35 \times 10^{-20}$) was detected, further supporting the involvement of adherens junction, focal adhesion, and regulation of actin cytoskeleton pathways.

Working on small sample sizes for genome-wide associations leads to enormous numbers of false-positive SNP associations. Using large sample sizes, on the other hand, may result in false-negative results due to the low effect sizes of disease-associated SNPs. One limitation of our study is the considerably low sample size in the assessment of SNP associations. In this regard, we aimed to take advantage of having both DNA and CSF samples of the same
affected individuals. Therefore, we aimed to include both false-positive SNP associations and prevent missing out SNPs with falsely high \( P \)-values in the first place by setting a non-stringent \( P \)-value cut-off. We then aimed to exclude false-positive associations during the pathway enrichment analyses in which interacting SNP subnetworks were used to identify the pathways, eliminating functionally irrelevant SNPs. Integrative analyses of the genomic and proteomic datasets determined the shared pathways between the two datasets, which seem to represent the pathophysiological mechanisms. Another limitation of the study is that we have used a genotyping chip with a relatively low SNP number (300K) since it was one of the most powerful chips when the genotyping had been carried out. Lastly, we had not obtained mRNA samples of the study participants at the time of their sample collection, which would lead to a more comprehensive analysis of the altered pathways in the studied patient group.

**CONCLUSIONS**

Here, we provide an integrated pathway analysis of whole-genome SNP and CSF proteome datasets obtained from the same MS patients, highlighting the most prominent molecular pathways that may be altered in the studied patient group. We confirmed the involvement of some of the previously known players in MS pathophysiology and suggested possible roles of some others. We conclude that this approach provides a better placing of separate genomic and proteomic datasets into a biological context and is useful for elucidating disease mechanisms.

**ACKNOWLEDGEMENTS**

We thank the patients and volunteers who agreed to contribute to the study.

**ADDITIONAL INFORMATION AND DECLARATIONS**

**Funding**

This work was supported by the Scientific and Technological Research Council of Turkey (TÜBİTAK, 109S070) for proteomic analyses and the Istanbul Technical University, Scientific Research Projects (BAP, 37426) for genomic analyses. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Grant Disclosures**

The following grant information was disclosed by the authors:

Scientific and Technological Research Council of Turkey: TÜBİTAK, 109S070.
Istanbul Technical University, Scientific Research Projects: BAP, 37426.

**Competing Interests**

The authors declare there are no competing interests.
**Author Contributions**

- Elif Everest performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Ege Ülgen analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Uğur Uygunoğlu, Melih Tutuncu, Sabahattin Saip and Aksel Siva conceived and designed the experiments, authored or reviewed drafts of the paper, diagnosed the patients, and approved the final draft.
- Osman Uğur Sezerman and Eda Tahir Turanli conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

**Human Ethics**

The following information was supplied relating to ethical approvals (i.e., approving body and any reference numbers):

The Ethics Committee of Istanbul Technical University approved the study (İTÜ-SM.INAREK-MBG-1).

**Data Availability**

The following information was supplied regarding data availability:

The raw data are available in the Supplemental Files.

**Supplemental Information**

Supplemental information for this article can be found online at [http://dx.doi.org/10.7717/peerj.11922#supplemental-information](http://dx.doi.org/10.7717/peerj.11922#supplemental-information).

**REFERENCES**

**Alvarez JI, Cayrol R, Prat A. 2011.** Disruption of central nervous system barriers in multiple sclerosis. *Biochimica Et Biophysica Acta (BBA)-Molecular Basis of Disease* **1812**(2):252–264 DOI 10.1016/j.bbadis.2010.06.017.

**Avsar T, Durası İM, Uygunoğlu U, Tütüncü M, Demirci NO, Saip S, Sezerman OU, Siva A, Turanlı ET. 2015.** CSF proteomics identifies specific and shared pathways for multiple sclerosis clinical subtypes. *PLOS ONE* **10**(5):e0122045 DOI 10.1371/journal.pone.0122045.

**Bahmanyar S, Montgomery SM, Hillert J, Ekbom A, Olsson T. 2009.** Cancer risk among patients with multiple sclerosis and their parents. *Neurology* **72**(13):1170–1177 DOI 10.1212/01.wnl.0000345366.10455.62.

**Bakir-Gungor B, Egemen E, Sezerman OU. 2014.** PANOGA: a web server for identification of SNP-targeted pathways from genome-wide association study data. *Bioinformatics* **30**(9):1287–1289 DOI 10.1093/bioinformatics/btt743.

**Berlin CB, Bargatze RF, Campbell JJ, Von Andrian UH, Szabo MC, Hasslen SR, Nelson RD, Berg EL, Erlandsen SL, Butcher EC. 1995.** α4 integrins mediate lymphocyte attachment and rolling under physiologic flow. *Cell* **80**(3):413–422 DOI 10.1016/0092-8674(95)90491-3.
Chen YA, Tripathi LP, Dessailly BH, Nyström-Persson J, Ahmad S, Mizuguchi K. 2014. Integrated pathway clusters with coherent biological themes for target prioritisation. *PLOS ONE* 9(6):e99030 DOI 10.1371/journal.pone.0099030.

Dayem Ullah AZ, Lemoine NR, Chelala C. 2012. SNPnexus: a web server for functional annotation of novel and publicly known genetic variants (2012 update). *Nucleic Acids Research* 40(W1):W65–W70 DOI 10.1093/nar/gks364.

Dejana E, Orsenigo F, Lampugnani MG. 2008. The role of adherens junctions and VE-cadherin in the control of vascular permeability. *Journal of Cell Science* 121(13):2115–2122 DOI 10.1242/jcs.017897.

Eikelenboom P, Bate C, Van Gool WA, Hoozemans JJ, Rozemuller JM, Veerhuis R, Williams A. 2002. Neuroinflammation in Alzheimer’s disease and prion disease. *Glia* 40(2):232–239 DOI 10.1002/glia.10146.

Gallo RM, Khan MA, Shi J, Kapur R, Wei L, Bailey JC, Liu J, Brutkiewicz RR. 2012. Regulation of the actin cytoskeleton by Rho kinase controls antigen presentation by CD1d. *The Journal of Immunology* 189(4):1689–1698 DOI 10.4049/jimmunol.1101484.

Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabási AL. 2007. The human disease network. *Proceedings of the National Academy of Sciences of the United States of America* 104(21):8685–8690 DOI 10.1073/pnas.0701361104.

Grytten N, Myhr KM, Celius EG, Benjaminsen E, Kampman M, Midgard R, Vatne A, Aarseth JH, Riise T, Torkildsen Ø. 2020. Risk of cancer among multiple sclerosis patients, siblings, and population controls: a prospective cohort study. *Multiple Sclerosis Journal* 26(12):1569–1580 DOI 10.1177/1352458519877244.

Han MH, Hwang SI, Roy DB, Lundgren DH, Price JV, Ousman SS, Fernald GH, Gerlitz B, Robinson WH, Baranzini SE, Grinnell BW. 2008. Proteomic analysis of active multiple sclerosis lesions reveals therapeutic targets. *Nature* 451(7182):1076–1081 DOI 10.1038/nature06559.

Hawkins BT, Davis TP. 2005. The blood–brain barrier/neurovascular unit in health and disease. *Pharmacological Reviews* 57(2):173–185 DOI 10.1124/pr.57.2.4.

Ideker T, Ozier O, Schwikowski B, Siegel AF. 2002. Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics* 18(suppl_1):S233–S240 DOI 10.1093/bioinformatics/18.suppl_1.s233.

International Multiple Sclerosis Genetics Consortium. 2007. Risk alleles for multiple sclerosis identified by a genomewide study. *New England Journal of Medicine* 357(9):851–862 DOI 10.1056/NEJMoa073493.

International Multiple Sclerosis Genetics Consortium*†, ANZgene, IIBDGC, WTCCC2. 2019. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science* 365(6460):eaav7188 DOI 10.1126/science.aav7188.

Kancheisa M, Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* 28(1):27–30 DOI 10.1093/nar/28.1.27.

Kirk J, Plumb J, Mirakhur M, McQuaid S. 2003. Tight junctional abnormality in multiple sclerosis white matter affects all calibres of vessel and is associated with
blood–brain barrier leakage and active demyelination. *The Journal of Pathology* 201(2):319–327 DOI 10.1002/path.1434.

Lee PH, Shatkay H. 2007. F-SNP: computationally predicted functional SNPs for disease association studies. *Nucleic Acids Research* 36(suppl_1):D820–D824 DOI 10.1093/nar/gkm904.

Magliozzi R, Hametner S, Facchiano F, Marastoni D, Rossi S, Castellaro M, Poli A, Lattanzi F, Visconti A, Nicholas R, Reynolds R. 2019. Iron homeostasis, complement, and coagulation cascade as CSF signature of cortical lesions in early multiple sclerosis. *Annals of Clinical and Translational Neurology* 6(11):2150–2163 DOI 10.1002/acn3.50893.

Miki H, Miura K, Takenawa T. 1996. N-WASP, a novel actin-depolymerizing protein, regulates the cortical cytoskeletal rearrangement in a PIP2-dependent manner downstream of tyrosine kinases. *The EMBO Journal* 15(19):5326–5335 DOI 10.1002/j.1460-2075.1996.tb00917.x.

Nielsen NM, Rostgaard K, Rasmussen S, Koch-Henriksen N, Storm HH, Melbye M, Hjalgrim H. 2006. Cancer risk among patients with multiple sclerosis: a population-based register study. *International Journal of Cancer* 118(4):979–984 DOI 10.1002/ijc.21437.

Padden M, Leech S, Craig B, Kirk J, Brankin B, McQuaid S. 2007. Differences in expression of junctional adhesion molecule-A and β-catenin in multiple sclerosis brain tissue: increasing evidence for the role of tight junction pathology. *Acta Neuropathologica* 113(2):177–186 DOI 10.1007/s00401-006-0145-x.

Patsopoulos NA, Baranzini SE, Santaniello A, Shoostari P, Cotsapas C, Wong G, Beecham AH, James T, Replotle J, Vlachos I, McCabe C. 2017. The multiple sclerosis genomic map: role of peripheral immune cells and resident microglia in susceptibility. *BioRxiv*. 143933 DOI 10.1101/143933;%20t.

Perry HV, Cunningham C, Boche D. 2002. Atypical inflammation in the central nervous system in prion disease. *Current Opinion in Neurology* 15(3):349–354 DOI 10.1097/00019052-200206000-00020.

Plumb J, McQuaid S, Mirakhur M, Kirk J. 2002. Abnormal endothelial tight junctions in active lesions and normal—appearing white matter in multiple sclerosis. *Brain Pathology* 12(2):154–169 DOI 10.1111/j.1750-3639.2002.tb00430.x.

Polman CH, O’Connor PW, Havrdova E, Hutchinson M, Kappos L, Miller DH, Phillips JT, Lublin FD, Giovannoni G, Wajat A, Toal M. 2006. A randomized, placebo-controlled trial of natalizumab for relapsing multiple sclerosis. *New England Journal of Medicine* 354(9):899–910 DOI 10.1056/NEJMo044397.

Polman CH, Reingold SC, Banwell B, Clanet M, Cohen JA, Filippi M, Fujihara K, Havrdova E, Hutchinson M, Kappos L, Lublin FD. 2011. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Annals of Neurology* 69(2):292–302 DOI 10.1002/ana.22366.

Saccone SF, Bolze R, Thomas P, Quan J, Mehta G, Deelman E, Tischfield JA, Rice JP. 2010. SPOT: a web-based tool for using biological databases to prioritize SNPs after...
a genome-wide association study. *Nucleic Acids Research* **38**(suppl_2):W201–W209 DOI 10.1093/nar/gkq513.

Sawcer S, Hellenthal G, Pirinen M, Spencer CC, Patsopoulos NA, Moutsianas L, Dilthey A, Su Z, Freeman C, Hunt SE, Edkins S. 2011. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* **476**(7359):214–219 DOI 10.1038/nature10251.

Vajkoczy P, Laschinger M, Engelhardt B. 2001. α4-integrin-VCAM-1 binding mediates G protein–independent capture of encephalitogenic T cell blasts to CNS white matter microvessels. *The Journal of Clinical Investigation* **108**(4):557–565 DOI 10.1172/JCI12440.

Wozniak MA, Modzelewski K, Kwong L, Keely PJ. 2004. Focal adhesion regulation of cell behavior. *Biochimica Et Biophysica Acta (BBA)-Molecular Cell Research* **1692**(2–3):103–119 DOI 10.1016/j.bbamcr.2004.04.007.

Xu Z, Taylor JA. 2009. SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. *Nucleic Acids Research* **37**(suppl_2):W600–W605 DOI 10.1093/nar/gkp290.

Yates RL, Esiri MM, Palace J, Jacobs B, Perera R, De Luca GC. 2017. Fibrin (ogen) and neurodegeneration in the progressive multiple sclerosis cortex. *Annals of Neurology* **82**(2):259–270 DOI 10.1002/ana.24997.