Immune cell infiltration characteristics and related core genes of bioinformatic analysis in multiple sclerosis

jingyu zhao (496583663@qq.com)
Institute of Aviation Medicine https://orcid.org/0000-0002-7522-3992
Jianyong Zheng
Xijing Hospital
Qun Wang
Xijing Hospital
Qian Li
Xijing Hospital
Nan zhang
Xijing Hospital

Research article

Keywords: Multiple sclerosis, CIBERSORT, GSEA, Immune infiltration

Posted Date: May 21st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-540699/v1

Abstract

Background Introduction Multiple sclerosis(MS) is a common complication of uncontrolled or excessive neuroinflammation and autoimmunity disease. Advances in high-throughput technologies and available bioinformatics tools make it possible to evaluate different expressions in the whole genome instead of focusing on a limited number of genes.

Methods

Materials and methods Two public available databases GSE81279 and GSE21942 of multiple sclerosis samples were downloaded analyzed by CIBERSORT. Gene Ontology (GO) and KEGG pathway analysis based on GSEA was performed by cluster profile software to reveal the regulatory relations among genes and provided a systematic understanding of the functional differentially expressed genes at the transcriptional level.GSE81279 was used to validate the association between core genes and clinical information.

Results

For immune cells, T-cell gamma delta and monocyte showed a trend toward reduction. The connection between the most prominent GO terms showed HBB, GATA2, NAA35, TCL1A, SECISBP2L, CLC, AGPAT5, CCR3, LTF, MALAT1, MS4A3 were significantly differentially expressed in MS. Gene set enrichment result was presented CDKN1A, DDB2, MME HMGN1, XPC, RELA for subsequent analysis.GSE81279 showed five types of immune cells revealed important links with MS. GSEA and layered KEGG analyses revealed that enrichment of immune response-related in primary immunodeficiency, it also consistent with previous studies. We got 10 genes, including HLA-DR, IL7R, HBB, TNFRSF1A, CYP27B1, NR1H3, IL2RA, TNFR1, BAFF, and CYP2R1 had close connections to clinical features.

Conclusions

Our study identifies immune cell infiltration with microarray data of the plasma in MS by using CIBERSORT analysis, we also provide novel information for further study of genes of multiple sclerosis.

Background

Multiple sclerosis (MS) is the most common chronic inflammatory demyelinating central nervous system (CNS) disease, which is characterized by uncontrolled or excessive immune cell infiltration, demyelination, and neurodegeneration[1, 2]. Compelling data indicate that the gene human leukocyte antigen (HLA) -DRB1 is one of the most notable features of the disease-associated gene of MS[3]. However, with the gene-sequencing method was used to detect disease, there are many associations with markers that were nominally significant. Fischer found inflammatory cells composed of T and B lymphocytes that activated demyelination and neurodegeneration in the multiple sclerosis brain and spinal cord[4].

Previous research directions have suggested that CD8+ memory T cells are dominant and B cells are limited in lesions of multiple sclerosis[5, 6]. However, it is still unknown genes interaction with immune cells and the gene interactions and interaction-interaction relationships among the large poor of immune environment,

The advances in high-throughput sequencing technologies and the available bioinformatics tools are progressing at a very fast pace[7], making it possible to evaluate different expressions in the whole genome instead of focusing on a limited number of genes. Furthermore, Riedmaier I[8]. provided a better view of the immune cells and biological processes involved biomarkers through a way of cibersort. Accordingly, the gene expression analyses could not only

be early diagnostic tools, but also new early treatments, and people have a deeper understanding of the molecular mechanism of MS occurrence, which could provide a possible new direction in the treatment.

So, we performed a genome-wide gene-expression analysis by microarrays and new bioinformatics tools to identify the altered molecular mechanism and potential biomarkers to the identification of characteristic gene-expression profiles to make further diagnosis and treatments.

Methods

Materials

The multiple sclerosis samples were from the GEO database(www. NCBI.nlm. NIH. gov/geo), and the normalized expression matrix was also downloaded from the GEO database. We searched the keyword"multiple sclerosis", and made the criteria were Homo sapiens, the expression profiling of study type must be an array and the sources of the samples of each dataset must be a serum. According to the above items, the GSE21942 microarray dataset based on the Affymetrix Human GeneChip platform was downloaded and adopted for CIBERSORT. We applied a gene expression deconvolution algorithm, CIBERSORT applied a deconvolution algorithm to estimate the transcriptome profile of reference leucocyte cell types and defined a reference gene signature matrix which was used to deduce the leucocyte composition from the transcriptome of MS patients. The original CIBERSORT gene signature file 22 leucocyte cell was downloaded from the web(http://cibersort.stanford.edu/)[9].

We retained the results that were statistically significant. The GSE81279 microarray dataset demonstrated the association between selected genes and clinical features.

Evaluation of immune cell infiltration

The GSE21942 expression database was processed to remove the null values. We use the KNN method of "impute" package to supplement the missed values[10], the format was prepared under the accepted format of CIBERSORT, and then data were uploaded to the CIBERSORT web portal. The GSEA revealed the regulatory relations among genes and provided a systematic understanding of the underlying molecular

mechanisms[11].

Immune Infiltration Matrix

Statistical analysis was carried out using R-language (R-project.org) and packages available through the Bioconductor project (www.bioconductor.org). Limma package and Bayesian method were used to construct a linear model and to analyze the differential expression of various types of 22 immune cells[12]. We used the difference analysis between the multiple sclerosis group and the control group. The differences between different 22 immune cells were analyzed, using the Pearson correlation coefficient to further detect the relationship. P-values were corrected for multiple hypothesis testing and presented with FDR values to get the filtrated samples (significance level <0.05). We drew every histogram of the proportion of each immune cell by filtrating the sample. The immune cell subtype was plotted using the heatmap function of the R programming environment to show the relationship[13]. We drew a violin plot to visualize the expression differences in 22 immune cells by using the vioplot package of R.

GSEA preparation

GSEA is an analytical method for genome-wide expression profile microarray data, which could get enrichment by comparing genes with predefined gene sets[14]. A gene set is a group of genes that shares localization, pathways, functions, or other features, GSEA was conducted using a cluster profile package (version 3.5)[15] Following the change of |log2FC|, we calculated the fold change of gene expression between multiple sclerosis groups and the control group through the cluster profile package to acquire the gene list.

Enriched GO analysis and KEGG pathway

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses were the next analysis steps.The GO roots are BP (biological_process), MF (molecular_function), and CC (cellular_component)[16].GO analysis was performed through the GO function package, the adjusted p-value > 0.05 was excluded standard. We used the GO enrich package to detect a connection between the most significant GO terms and participating genes.KEGG pathway enrichment analyses were also conducted by the KEGG function in cluster profile package (significance level <0.05).

Core gene and clinical association

We got the most significant GO term through the CIBERSORT results (significance level <0.05).25 groups of correlation analysis data were obtained from the database GSE21942, only 1 gene sets was significantly enriched at FDR < 25%,8 gene sets were significantly enriched at normal value <1%,25 gene sets were significantly enriched at nominal value < 5%.In dataset GSE81279 with clinical information, we excluded parts of patients because of its source, and only sources of monocytes before and after treatment was included, which had approximately 608 genes obtained in the GO term associated with immune response, 19 were chosen for analysis.

Statistical analysis

Correlation coefficients between immune cells among all credible samples used Pearson and analysis towards the interrelationship between immune cells. Meanwhile, Wilcoxon signed-rank test was used to analyze the difference of immune cell infiltration between the normal group and multiple sclerosis group, P<0.05 was considered statistically significant.

Results

Bioinformation analysis

Based on cibersort way of 22 immune cells type, there were available to get 10 MS samples (p<0.05), and 15 normal samples.GSE81279 contained 6 MS untreated monocyte and 11 monocyte samples after treated as showed. The percentage of the infiltrated immune cell showed Monocytes, NK cells resting, T-cell gamma delta, B cell naive were the top 4 abundant immune infiltrates in normal people(Fig1 a). Compared with normal group, B cells naive(p= 0.001), T cells CD4 naive(p= 0.019), T cells gamma delta(p= 0.016), Eosinophils(p= 0.001), Neutrophils (p= 0.017)differed in MS group. Respectively, T-cell gamma delta and monocyte showed a trend toward reduction. In the MS group.It indicates the correlation between these differentially expressed types of immune cells. The five types of

immune cells were weakly to moderately correlated. Monocytes were negatively correlated with T cells CD4 naive and Dendritic cells activated, Neutrophils, Eosinophils, B cell naive (r = − 0.26 and r = − 0.49, r = − 0.41, r = − 0.42,r = − 0.09). T cells gamma delta were also negatively correlated with T cells CD4 naive and Dendritic cells activated, Neutrophils, Eosinophils, B cell naive (r =−0.56 and r =− 0.27,r =− 0.16,r = − 0.3 8,r = − 0.50), which indicated that the function of monocytes and T cells gamma delta may be antagonistic in MS .T cells CD4 naive and Dendritic cells activated Neutrophils, Eosinophils cells was synergistic.T cells CD4 naive were negatively correlated with T cells CD8, T-cells gamma delta(r = − 0.71 and r = − 0.54), were positively correlated with Neutrophils(r = 0.43), Eosinophils(r =0.5), T cells CD4 memory resting(r =0.41), B cells naive(r = 0.53)(Fig1 b,c,d).

GSEA analysis

Through the GO biological process, the top 10 most significantly enriched GO terms were presented in Table 1,2..GO terms of treatment were primarily associated with "Regulation Of Peptidyl Serine Phosphorylation Of Stat Protein, Response To Uv B, Negative Regulation Of Translational Initiation, Regulation Of Protein Kinase C Signaling, Positive Regulation Of Creb Transcription Factor Activity, Antibacterial Humoral Response, Extracellular Matrix Assembly, Digestive System Development, Positive Regulation Of Transcription By Rna Polymerase I, Modulation By Host Of Viral Process, Positive Regulation Of Carbohydrate Metabolic Process"($p<0.0001$, $p<0.0001$, $p<0.0001$,$p<0.0001$,$p<0.0001$,$p<0.0001$, $p<0.0001$, $p<0.0001$, $P=0.018$, $P=0.019$)were gene sets of MS.14 gene sets were significantly enriched in the untreated group($p<0.05$), and 6 gene sets were significantly enriched ($p<0.05$) of the treatment group. The details were shown in Fig 2.

The connection between the most prominent GO terms is shown. The network-presented numerous genes, such as HBB,GATA2,NAA35,TCL1A, SECISBP2L,CLC,AGPAT5,CCR3,LTF,MALAT1,MS4A3 and some other genes were significantly differentially expressed in MS.The two genes (PWP1, GGNBP2) expressed differently in the "Regulation Of Peptidyl Serine Phosphorylation Of Stat Protein" gene sets for GSEA.RELA,XPC,HMGN1,CDKN1A,MME,DDB2 expressed significantly in "Response To Uv B"gene sets for GSEA. Gene set enrichment result was presented in the figure. The enrichment showed that the gene set was enriched at the front of the sequence (ES = -0.65). We obtained the list of all core genes, such as CDKN1A, DDB2, MME HMGN1, XPC, RELA for subsequent analysis.

Table1 GSEA-based GO analysis of top 10 biological process terms

NAME	SIZE	ES	NES	NOM p-val	FDR q-val	FWER p-val
REGULATION_OF_PEPTIDYL_SERINE_PHOSPHORYLATION_OF_STAT_PROTEIN	20	-0.55770963	-1.8629167	0	0.2439	0.29
RESPONSE_TO_UV_B	17	-0.64523363	-1.6230046	0	1	0.98
MYOBLAST_PROLIFERATION	19	-0.50449675	-1.5804042	0.02040	1	1
NEGATIVE_REGULATION_OF_TRANSLATIONAL_INITIATION	17	-0.53856456	-1.57758	0	1	1
REGULATION_OF_PROTEIN_KINASE_C_SIGNALING	15	-0.61808467	-1.571803	0	1	1
POSITIVE_REGULATION_OF_CREB_TRANSCRIPTION_FACTOR_ACTIVITY	18	-0.6728891	-1.5715604	0.01818	1	1
ANTIBACTERIAL_HUMORAL_RESPONSE	40	-0.62112427	-1.5605229	0	1	1
SERINE_PHOSPHORYLATION_OF_STAT_PROTEIN	24	-0.4843511	-1.5514377	0.07317	1	1
EMBRYONIC_HINDLIMB_MORPHOGENESIS	28	-0.44431925	-1.5426135	0.04255	1	1
POSITIVE_REGULATION_OF_MESONEPHROS_DEVELOPMENT	22	-0.53573346	-1.5345056	0.02083	1	1

Table2. The top 10 GSEA-based kegg analysis terms.

NAME	SIZE	ES NES NOM p-val FDR q-val
PRIMARY_IMMUNODEFICIENCY	35	-0.5699937 -1.3643178 0.076923081
LINOLEIC_ACID_METABOLISM	24	-0.42973483 -1.2812052 0.08 1
GLYCINE_SERINE_AND_THREONINE_METABOLISM	30	-0.4299071 -1.2570243 0.232558151
LYSINE_DEGRADATION	43	-0.3849916 -1.2218008 0.1754386 1
TASTE_TRANSDUCTION	44	-0.25822443 -1.1707219 0.2244898 1
OLFACTORY_TRANSDUCTION	111	-0.23009725 -1.1693118 0.209302320.8867015
ABC_TRANSPORTERS	43	-0.36049527 -1.1056648 0.3043478 0.9979592
RENIN_ANGIOTENSIN_SYSTEM	16	-0.38465297 -1.0639392 0.326086971
RIG_I_LIKE_RECEPTOR_SIGNALING_PATHWAY	68	-0.33164003 -1.0420119 0.461538460.9984609
NOD_LIKE_RECEPTOR_SIGNALING_PATHWAY	62	-0.3369027 -0.9907027 0.530612231

Phosphorylation Of Stat Protein.C. representated kegg pathway:Primary Immunodeficiency.D.It representates Linoleic Acid Metabolism pathway.The depth of the inner arc area decreasing or increasing of the biological process

of the six pictures and shows the running ES score and positions of geneset members on the rank ordered List.E.heat map Heat Map of the top 50 features for each phenotype in 1 collapsed tosymbols.

KEGG analysis

There were 40 main KEGG pathways(Fig 3). The MS patients showed pathways of "Primary Immunodeficiency, Linoleic Acid Metabolism, Glycine Serine And Threonine Metabolism, Lysine Degradation, Taste Transduction, Olfactory Transduction, Abc Transporters, Renin-Angiotensin System, Rig I Like Receptor Signaling Pathway, Nod Like Receptor Signaling Pathway", the results indicated that the activation of signaling pathways in MS was related to immune, linoleic and amino acid metabolism, TNFRSF13B, PTPRC, CIITA, IL7R, CD40LG, DCLRE1C, BTK, JAK3, IKBKG, RFXAP, TNFRSF13C, ICOS, TAP2, RFX5, CD19, CD40, CD79A, BLNK were in Primary Immunodeficiency pathways. Meanwhile,25 gene sets are upregulated in phenotype MS.

Discovery of core genes

Six types of immune infiltrating cells are shown in Fig 4. A total of 41 genes showed a tight link with immune infiltrating cells. Genes such as CD79A, IGHD, IGHM, IGKC, IL4R, MS4A1, SELL, TCL1A were negatively correlated with B cells naive. CCR7, IL7R, ITK, SELL, TRAC, TRBC1 were negatively related to T cells CD4 naive. BCL2A1,CCR3,CLC,FOSB,NCF2,P2RY14 and C5AR1,CXCR2 ,

FCGR3B, FPR1, NCF2,SELL were negatively related to Eosinophils and Neutrophils respectively.CCL5,CD3D,CST7,GZMA,GZMK, IL2RB,KLRB1,PRF1 TRDC were positively correlated with T cells gamma delta ,CLEC7A, FCN1,HCK,MNDA, MS4A6A,S100A12 were positively correlated with monocytes. As we knew above, Monocyte and T cell gamma-delta were negatively correlated with the other 4 immune-cells.

Validation of core genes

GSE81279 shows the clinical information and the core gene list.We analysis between core gene list and clinical information CD79A,IL4R,MS4A1,SELL,CCR7, CD3D,FAIM3,IL7R,LCK,CCL5,GZMA,GZMK,IL2RB,KLRB1,PRF1,CLEC7A,FPR 1,MNDA,NCF2 worked in MS. Among them, CLEC7A, SELL, MS4A1 were positively correlated with the treated MS patients, it also indicated that these core genes had a close connection with immune cell infiltration and clinical manifestation. And, GZMA, IL7R, IL4R, GZMK, CD79A, IL2RB, PRF1, KLRB1,CD3D were positively correlated with untreated patients.

Discussion

Multiple sclerosis is an immune-mediated disease influenced by the interaction of genetics and environmental factors[17].it is still unknown genes interaction with immune cells and the gene interactions and interaction-interaction relationships among the large poor of immune environment, To calculate and identify the complex tissue cells, the deconvolution strategies have prevalent over the past few years, some researchers used a deconvolution algorithm to solve immune cell gene expression reversely. Previous studies have put forward an algorithm called "CSSAM" (cell type-specific significant analysis of microarrays), the development of the algorithm is based on the traditional microarray analysis method, ignoring the composition of the sample cell types that could not accurately distinguish between gene expression and the relationship between the different cell types. Newman[9] in 2015 published a paper that presents the CIBERSORT new development tools, its performance was superior to all other

methods. Therefore, we could uncover different expressional cell patterns of immune infiltration in MS and their association with clinical features.

In our study, we found T cell gamma delta and monocytes were two prominent differentially decreased cells. Monocytes are white cells, which account for the number of white blood cells in peripheral blood 4–10%, according to its function and cell transport characteristics, it could be divided into intrinsic and inflammatory monocytes[18].In the present study, inflammatory monocytes were important mediators in inflammatory diseases such as rheumatoid arthritis, and autoimmune encephalomyelitis, and multiple sclerosis of the central nervous system in which the monocyte migration works[19]. It also showed that monocytes constituted 24%-42% of immune cells in MS patients by our research. Although many types of leukocytes were involved in disease progression, activated monocytes were believed to be one of the first to arrive at the brain and initiate inflammation[20]. Most importantly, perivascular monocytes are not only involved in the process of atypical or severe presentation early in the disease, but also in patients with chronic MS. Other immune cells such as macrophages, B cells, and gamma delta T-cells, as well as natural killer (NK) cells also played a critical role in the process of pathology in MS patients[21–23]. We found T cells gamma delta decreased in MS patients. It comprised a small fraction (1–5%) of the lymphocytes, localized in mucosal tissue and skin as a major population(up to 50%) of lymphocytes[24]. As with the earlier study, Gamma-delta T cells were in the process of pathogenesis in MS, γδ T cells can produce IL-17 through exposure to some cytokines, such as IL-23 and IL-1 β to promote neutrophil infiltration[25, 26]. It is well recognized that IL-17 was found in the blood and CSF from MS patients during clinical relapses, furthermore, the expansion of IL-17 was associated with MS severity[27, 28].

The role of B cells in multiple sclerosis lesions is currently unclear and may include direct pro-inflammatory effects, antigen presentation to T cells in the peripheral immune system or the brain[29]. It is noteworthy that, the contribution of B cell to the inflammatory infiltrates in multiple sclerosis was lower compared to that of T cells, B cells appear to be a prominent source of IL-10 within the CNS of MS patients[30]. Naive CD4 + T cells could encounter specific antigens, they can differentiate into different effector subsets. MS patients endothelial cells express high levels of IL-17R[31], which favors the differentiation of naive CD4 + T cells into Th17 cells, transmigrating efficiently across BBB endothelial cells and disrupt BBB tight junctions, made destruction of the immune environment and brought inflammation factors of

CNS[32]"Regulation Of Peptidyl Serine Phosphorylation Of Stat Protein" was the top associated pathway under GSEA-based GO analysis. Signal transducers and activators of transcription (STAT) family in the peripheral blood mononuclear cells (PBMCs) of MS patients have been associated with the development of clinically definite MS in individuals, presenting a clinically isolated syndrome (CIS)[33]. The inhibitors of the JAK/STAT pathway have ameliorated the clinical signs of EAE using inhibiting CD4 T cell differentiation toward the proinflammatory Th1/Th17 subtypes[34]. It suggested that SATA family genes have not been formally associated with MS, however, their involvement in disease pathology seems plausible. GSEA and layered KEGG analyses revealed that enrichment of immune response-related in primary immunodeficiency, it also consistent with previous studies. By using Cytoscape, the significant genes of primary immunodeficiency were in visualization. What is strikingly noticeable is that taste transduction and olfactory transduction was on the list of top 5 items of KEGG analyses. As we all know, taste and olfactory disorders are typical clinical symptoms of diagnosing criteria.

Through CIBERSORT we found many novels commonly expressed genes, some of which were identified as a susceptibility factor in auto-immune disease. TNFRSF1 is member of the tumor necrosis factor receptor superfamily (TNFRSF), which plays a dominant role in the pathogenesis of many autoimmune diseases[35]. However, TNFRSF co-stimulatory pathways predominantly promote T cell self-reactivity, promotion or inhibition for Tcell depend on the

inflammatory environment[35]. S100A12, as an extracellular newly identified RAGE-binding protein, was highly expressed in monocytes, and S100A12 is a member of the S100 family[36].Hara Ma ect[37] found a strong correlation exists between S100A12 protein and mRNA in peripheral blood mononuclear cells, which corresponded with our researches. It associated with auto-immune disease of rheumatoid arthritis(RA) and psoriatic arthritis and with cardiovascular complications in lupus[38–40]

We found some documents about MS and related genes, HLA-DR, IL7R, HBB(, TNFRSF1A, CYP27B1, NR1H3, IL2RA, TNFR1, BAFF, CYP2R1 were part of the 10 core genes, which were related to clinical manifestation and associated with MS (Table 3.).People with HLA-DRB1 has three times higher risk than the general population to be an MS patient[41].IL7R plays a role in the pathophysiology and contributes to the non-HLA genetic risk in multiple sclerosis[42].IL2RA, also known as CD25, is central to the balance between immune tolerance and immunity, increased concentration of soluble CD25 in sera has been detected in patients[43, 44]. The significance of the HBB distributions expressing in internal cortical layers is not clear, but it may be a result of a contribution to MS pathology by exacerbating the effects of demyelination on neuronal energetics[45]. TNFRSF1A mutation is associated with multiple sclerosis, previous studies showed TNFRSF1A and IL2RA, IL7R are genes of coding for cytokine for auto-immune diseases, besides, CYP28B1 worked as environmental risk factors such as Vitamin D for an item of Uv BC related diseases[46, 47].CYP28B1 with variants is a known risk for MS, while, the role in the pathology of MS is still unknown. The MS family appears to have a highly susceptible NR1H3 genetic background, NR1H3, known as nuclear receptors, control transcriptional regulation of genes involved in lipid homeostasis, inflammation, and innate immunity, the Zhenya still cunzai about missense vibration of NRIH3 causes a novel familial form of multiple sclerosis[48]. BAFF was associated with multiple sclerosis as well as SLE, BAFF is quantitative parameter of intrathecal IgG synthesis in MS, and its level presents progress movement of MS[49, 50].CYP2R1 variants have lower levels of 25OHD by a clinically relevant degree, which have higher risk levels of MS[51]. we could make further studies to elucidate the interaction with multiple sclerosis features.

Table 3
The previous studies about core genes in MS

NAME	SOURCE	Introduction
HLA-DR[41]	Blood	HLA-DRB1 * 1501 is the most common MS susceptibility gene in the western population, and the risk of carriers is three times higher than that
IL7R[42]	Blood	IL7R PLAYS a role for the pathophysiology and contributeS to the non-HLA genetic risk in multiple sclerosis.
IL2RA[43]	Blood	The most associated IL2RA SNP for MS susceptibility is rs2104286 located in intron 1 of IL2RA,and it has immunopathological roles.
TNFRSF1A[35]	Blood	TNFRSF1A is slightly associated with multiple sclerosis,TNFRSF1A mutation is associated with multiple sclerosis.
CYP2R1[51]	Blood	Low-frequency CYP2R1 variant have lower levels of 25OHD by a clinically relevant degree,which have higher risk levels of MS
NR1H3[48]	Blood	The MS family appears to have a highly susceptible NR1H3 genetic background.
CYP28B1[46] BAFF[49]	Blood Blood	CYP27B1 with variants is a known biological outcome located within a locus associated with increased risk of MS. B-cell activating factor (BAFF), was associated with multiple sclerosis as well as SLE.
HBB[45]	nucleus	A part of a mechanism linking neuronal energetics with epigenetic

Conclusion

We analyzed immune cell infiltration of blood in MS. The relative data were derived from GEO which is reliable. Gene expression, immune and stromal scores, prognostic immune-related genes, and the abundance of infiltrating immune cells. The analytical ways were novel as well as scientific, we could make further research about the obtained core genes. However, some limitations existed in our research, the small samples and several restrictive databases are inevitable, well, we still found some significant differences among groups, and the clinical pieces of information support our results.

Declarations

Ethics approval and consent to participate

The observational clinical study was approved by the Ethical Committee and all participants provided written informed consent prior to entering the study.

Consent for publication

Not applicable.

Availability of data and materials

The data that support the findings of this study are available from the corresponding author upon reasonable request. Majority of the data analyses were performed using R × 64 3.5 All R codes written for this manuscript are available from the corresponding author upon request.

Competing interests

The authors declare that they have no competing interests.

Funding

No funding.

Author contributions

All authors contributed toward data analysis, drafting and critically revising the paper, approved the final version to be published, and agree to be accountable for all aspects of the work.

Acknowledgements

We appreciate the sharing of Perl code for by Kang luo at Xi'an psychiatric hospital

Footnotes

NAME	SOURCE	Introduction
HLA-DR[41]	Blood	HLA-DRB1 * 1501 is the most common MS susceptibility gene in the western population, and the risk of carriers is three times higher than that
IL7R[42]	Blood	IL7R PLAYS a role for the pathophysiology and contributeS to the non-HLA genetic risk in multiple sclerosis.
IL2RA[43]	Blood	The most associated IL2RA SNP for MS susceptibility is rs2104286 located in intron 1 of IL2RA,and it has immunopathological roles.
TNFRSF1A[35]	Blood	TNFRSF1A is slightly associated with multiple sclerosis,TNFRSF1A mutation is associated with multiple sclerosis.
CYP2R1[51]	Blood	Low-frequency CYP2R1 variant have lower levels of 25OHD by a clinically relevant degree,which have higher risk levels of MS
NR1H3[48]	Blood	The MS family appears to have a highly susceptible NR1H3 genetic background.
CYP28B1[46] BAFF[49]	Blood Blood	CYP27B1 with variants is a known biological outcome located within a locus associated with increased risk of MS. B-cell activating factor (BAFF), was associated with multiple sclerosis as well as SLE.
HBB[45]	nucleus	A part of a mechanism linking neuronal energetics with epigenetic
Table 3 The previous studies about core genes in MS		
S100A1237		changes to histones in the nucleus and provide neuroprotection in MS by Blood It associated with disease in rheumatoid arthritis(RA) and psoriatic arthritis and with cardiovascular complications in lupus.

References

1. Dendrou CA, Fugger L, Friese MA. Immunopathology of multiple sclerosis. Nat Rev Immunol. 2015. 15(9): 545-58.
2. Reich DS, Lucchinetti CF, Calabresi PA. Multiple Sclerosis. N Engl J Med. 2018. 378(2): 169-180.
3. International Multiple Sclerosis Genetics Consortium, Hafler DA, Compston A, et al. Risk alleles for multiple sclerosis identified by a genomewide study. N Engl J Med. 2007. 357(9): 851-62.
4. Patejdl R, Penner IK, Noack TK, Zettl UK. Multiple sclerosis and fatigue: A review on the contribution of inflammation and immune-mediated neurodegeneration. Autoimmun Rev. 2016. 15(3): 210-20.
5. van Nierop GP, van Luijn MM, Michels SS, et al. Phenotypic and functional characterization of T cells in white matter lesions of multiple sclerosis patients. Acta Neuropathol. 2017. 134(3): 383-401.
6. Frischer JM, Bramow S, Dal-Bianco A, et al. The relation between inflammation and neurodegeneration in multiple sclerosis brains. Brain. 2009. 132(Pt 5): 1175-89.
7. Gibrat J. A short note on dynamic programming in a band. 2018. 19(1).
8. Riedmaier I, Pfaffl MW. Transcriptional biomarkers – High throughput screening, quantitative verification, and bioinformatical validation methods. Methods. 2013. 59(1).

9. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015. 12(5): 453-7.
10. Kohbalan M, Mohd M, Safaai D. A Review on Missing Value Imputation Algorithms for Microarray Gene Expression Data. Curr Bioinform. 2019.9(1): 18-22.
11. Dall ' Era M. Treatment of lupus nephritis: current paradigms and emerging strategies. Curr Opin Rheumatol. 2017. 29(3): 1.
12. Diboun I, Wernisch L, Orengo CA, Koltzenburg M. Microarray analysis after RNA amplification can detect pronounced differences in gene expression using limma.2015. 7(1): 252-0.
13. Villman J, Happonen E. R-tools: mediawiki extension for full-scale statistical computing. 2013 .3(2):12-32.
14. Sun Y, Gao Y, Sun J, et al. Expression profile analysis based on DNA microarray for patients undergoing off? pump coronary artery bypass surgery. Experimental & Therapeutic Medicine.2017.6(3):112-131 .
15. Yu G, Wang L, Han Y, He Q. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. OMICS. 2018.16(5): 284-287.
16. Klopfenstein DV, Liangsheng Z, Pedersen BS, et al. GOATOOLS: A Python library for Gene Ontology analyses. Sci Rep. 8(1): 10872-.
17. Bisht B, Darling WG, Grossmann RE, et al. A multimodal intervention for patients with secondary progressive multiple sclerosis: feasibility and effect on fatigue. J Altern Complement Med. 2014. 20(5): 347-55.
18. Kamei M, Carman CV. New observations on the trafficking and diapedesis of monocytes. Curr Opin Hematol. 2010. 17(1): 43-52.
19. Cayrol R, Wosik K, Berard JL, et al. Activated leukocyte cell adhesion molecule promotes leukocyte trafficking into the central nervous system. Nat Immunol. 2008. 9(2): 137-45.
20. Reder Anthony T, Kursad G, Byskosh Paul V, Porrini. Monocyte activation in multiple sclerosis. Multiple Sclerosis. 1998. 4(3): 162-168.
21. Joerg S, Grohme DA, Erzler M, et al. Environmental factors in autoimmune diseases and their role in multiple sclerosis. Cellular and molecular life sciences: CMLS. 2016. 73(24): 4611-4622.
22. Shen D, Zheng X, Shi M, et al. The roles of macrophages and microglia in multiple sclerosis and experimental autoimmune encephalomyelitis. Journal of Neuroimmunology: Official Bulletin of the Research Committee on Neuroimmunology of the World Federation of Neurology. 2018. 318: 1-7.
23. Babic M, Romagnani C. The Role of Natural Killer Group 2, Member D in Chronic Inflammation and Autoimmunity. Front Immunol. 2018. 9: 1219.
24. Carding SR, Egan PJ. Gammadelta T cells: functional plasticity and heterogeneity. Nat Rev Immunol. 2002. 2(5): 336-45.
25. Wucherpfennig KW, Newcombe J, Li H, Keddy C, Cuzner ML, Hafler DA. Gamma delta T-cell receptor repertoire in acute multiple sclerosis lesions. Proc Natl Acad Sci U S A. 1992. 89(10): 4588-92.
26. Hvas J, Oksenberg JR, Fernando R, Steinman L, Bernard CC. Gamma delta T cell receptor repertoire in brain lesions of patients with multiple sclerosis. J Neuroimmunol. 1993. 46(1-2): 225-34.
27. Brucklacher-Waldert V, Stuerner K, Kolster M, Wolthausen J, Tolosa E. Phenotypical and functional characterization of T helper 17 cells in multiple sclerosis. Brain. 2009. 132(Pt 12): 3329-41.
28. Ferreira TB, Hygino J, Wing AC, et al. Different interleukin-17-secreting Toll-like receptor+ T-cell subsets are associated with disease activity in multiple sclerosis. Immunology. 2018. 154(2): 239-252.

29. Lehmann-Horn K, Kinzel S, Weber MS. Deciphering the Role of B Cells in Multiple Sclerosis-Towards Specific Targeting of Pathogenic Function. Int J Mol Sci. 2017. 18(10).
30. Fillatreau S. Regulatory plasma cells. Curr Opin Pharmacol. 2015. 23(1): 1-5.
31. Larochelle C, Alvarez JI, Prat A. How do immune cells overcome the blood-brain barrier in multiple sclerosis. FEBS Lett. 2011. 585(23): 3770-80.
32. Astrocytes stimulate interleukin-17 and interferon-γ production in vitro.2019. 85(16): 3598-3606.
33. Liu Y, Holdbrooks AT, De Sarno P, et al. Therapeutic efficacy of suppressing the Jak/STAT pathway in multiple models of experimental autoimmune encephalomyelitis. J Immunol. 2014. 192(1): 59-72.
34. Qin H, Yeh WI, De Sarno P, et al. Signal transducer and activator of transcription-3/suppressor of cytokine signaling-3 (STAT3/SOCS3) axis in myeloid cells regulates neuroinflammation. Proc Natl Acad Sci U S A. 109(13): 5004-9.
35. Zhang Q, Vignali DA. Co-stimulatory and Co-inhibitory Pathways in Autoimmunity. Immunity.2018. 44(5): 1034-1051.
36. Bagheri V. S100A12: Friend or foe in pulmonary tuberculosis. Cytokine. 2017. 92(6): 80-82.
37. Hara M, Ando M, Morito T, et al. S100A12 gene expression is increased in peripheral leukocytes in chronic kidney disease stage 4-5 patients with cardiovascular disease. Nephron Clin Pract. 2013. 123(3-4): 202-8.
38. Nakashima A, Carrero JJ, Qureshi AR, et al. Effect of Circulating Soluble Receptor for Advanced Glycation End Products (sRAGE) and the Proinflammatory RAGE Ligand (EN-RAGE, S100A12) on Mortality in Hemodialysis Patients. Clin J Am Soc Nephrol. 2019.5(12): 2213-2219.
39. Shiotsu Y, Mori Y, Nishimura M, et al. Plasma S100A12 Level Is Associated with Cardiovascular Disease in Hemodialysis Patients. Clinical Journal of the American Society of Nephrology Cjasn. 2017.6(4): 718-723.
40. Zhao P, Wu M, Yu H, et al. Serum S100A12 Levels Are Correlated With the Presence and Severity of Coronary Artery Disease in Patients With Type 2 Diabetes Mellitus. Journal of Investigative Medicine the Official Publication of the American Federation for Clinical Research. 2018.61(5): 861-866.
41. Hollenbach JA, Oksenberg JR. The immunogenetics of multiple sclerosis: A comprehensive review. J Autoimmun. 2015. 64: 13-25.
42. Lundmark F, Duvefelt K, Iacobaeus E, et al. Variation in interleukin 7 receptor alpha chain (IL7R) influences risk of multiple sclerosis. Nat Genet. 2007. 39(9): 1108-13.
43. Anderson DE, Maier LM, Healy B, et al. Soluble IL-2RA levels in multiple sclerosis subjects and the effect of soluble IL-2RA on immune responses. The Journal of Immunology: Official Journal of the American Association of Immunologists. 2009. 182(3): 1541-1547.
44. Relationship between Multiple Sclerosis-Associated IL2RA Risk Allele Variants and Circulating T Cell Phenotypes in Healthy Genotype-Selected Controls .2018.15(3):1223-1245.
45. Brown N, Alkhayer K, Clements R, et al. Neuronal Hemoglobin Expression and Its Relevance to Multiple Sclerosis Neuropathology. J Mol Neurosci. 2016. 59(1): 1-17.
46. Pierrot-Deseilligny C, Souberbielle JC. Is hypovitaminosis D one of the environmental risk factors for multiple sclerosis. Brain. 2010. 133(Pt 7): 1869-88.
47. Australia and New Zealand Multiple Sclerosis Genetics Consortium (ANZgene). Genome-wide association study identifies new multiple sclerosis susceptibility loci on chromosomes 12 and 20. Nat Genet. 2009. 41(7): 824-8.
48. Minikel EV, MacArthur DG. Publicly Available Data Provide Evidence against NR1H3 R415Q Causing Multiple Sclerosis. Neuron. 2016. 92(2).

49. Ragheb S, Li Y, Simon K, et al. Multiple sclerosis: BAFF and CXCL13 in cerebrospinal fluid. Mult Scler. 2011. 17(7): 819-29.
50. Puthenparampil M, Miante S, Federle L, et al. BAFF is decreased in the cerebrospinal fluid of multiple sclerosis at clinical onset. J Neuroimmunol. 2016. 297: 63-7.
51. Despoina M, Tom D, Simon H, et al. Low-Frequency Synonymous Coding Variation in CYP2R1 Has Large Effects on Vitamin D Levels and Risk of Multiple Sclerosis. Am J Hum Genet. 2018. 103(6):266-291..

Figures

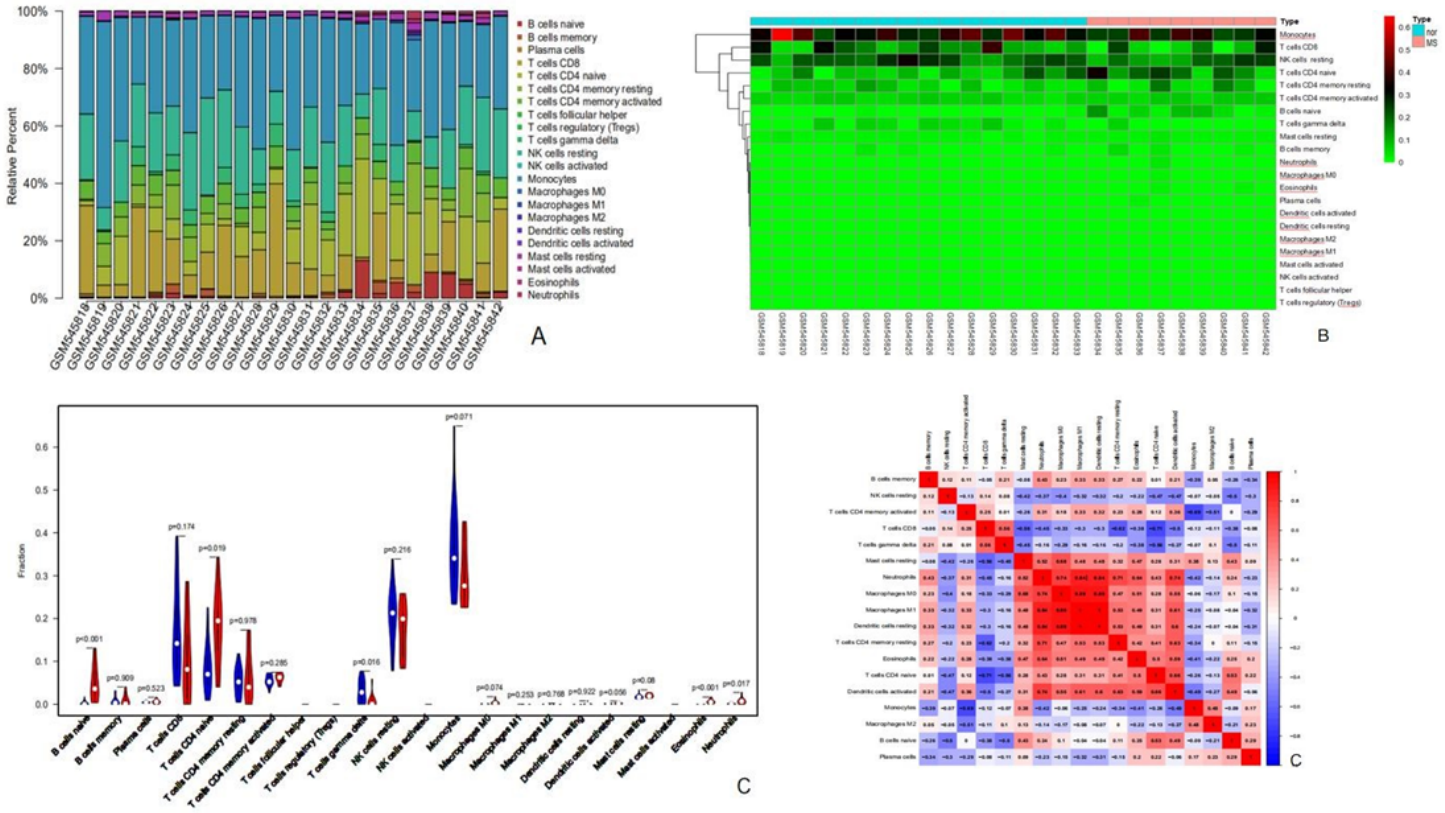

Figure 1

Immune infiltration of MS. a. Bar charts of 22 immune cell proportions in MS and normal groups. b. Correlation matrix of immune cell proportions. The differences are obvious by the matrix heat map. C. The median value range was in the box plots of the violin. The white points represented the mean fraction. The blue plots represented the normal group and the red plot for the MS group. d.The presentation of immune cells for correlation. R package ggplot (version 0.1.8) was used for the T test. Results visualization was performed by using R package (version 3.5.3). Correlation analysis and visualization were performed by using R package corrplot and corheatmap.

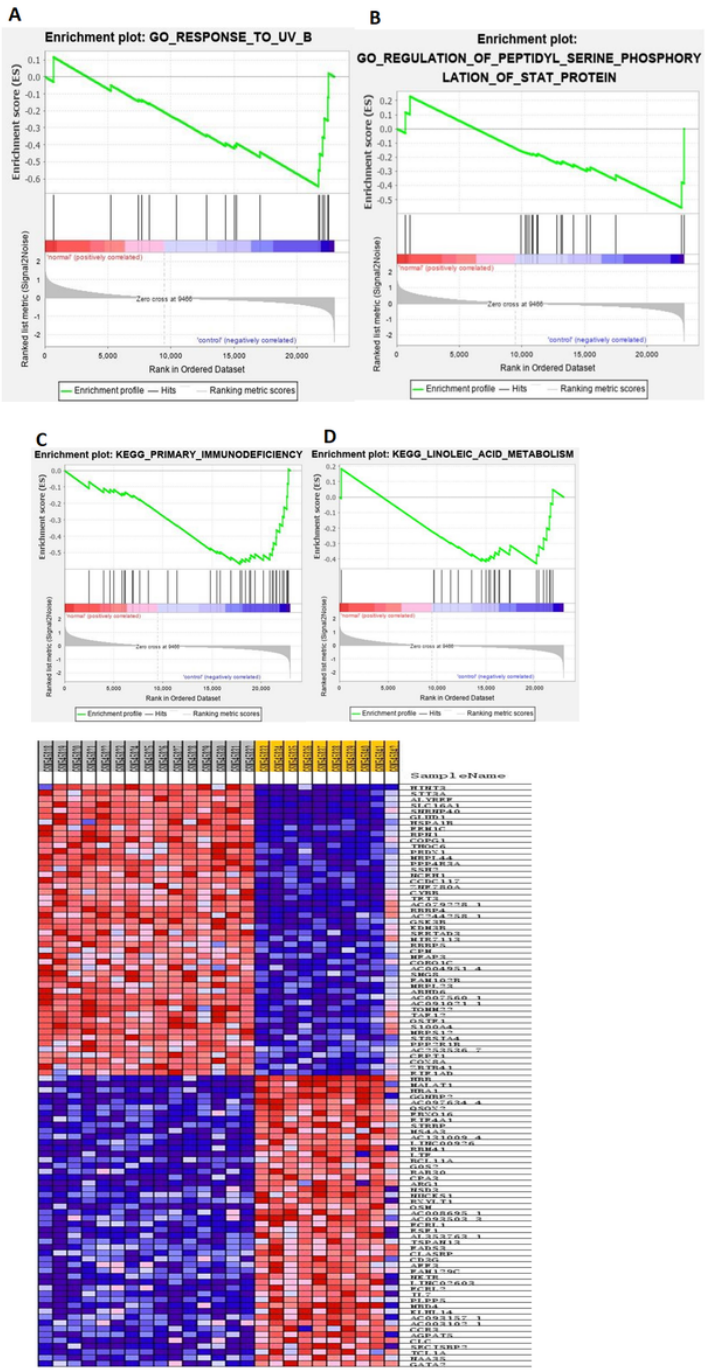

Figure 2

GO and KEGG analysis.A.The area of circle means gene sets quantities,it representates gene sets: response To Uv B,B.It representated gene sets:Regulation Of Peptidyl Serine Phosphorylation Of Stat Protein.C. representated kegg pathway:Primary Immunodeficiency.D.It representates Linoleic Acid Metabolism pathway.The depth of the inner arc area decreasing or increasing of the biological process of the six pictures and shows the running ES score and positions of geneset members on the rank ordered List.E.heat map Heat Map of the top 50 features for each phenotype in 1 collapsed tosymbols.

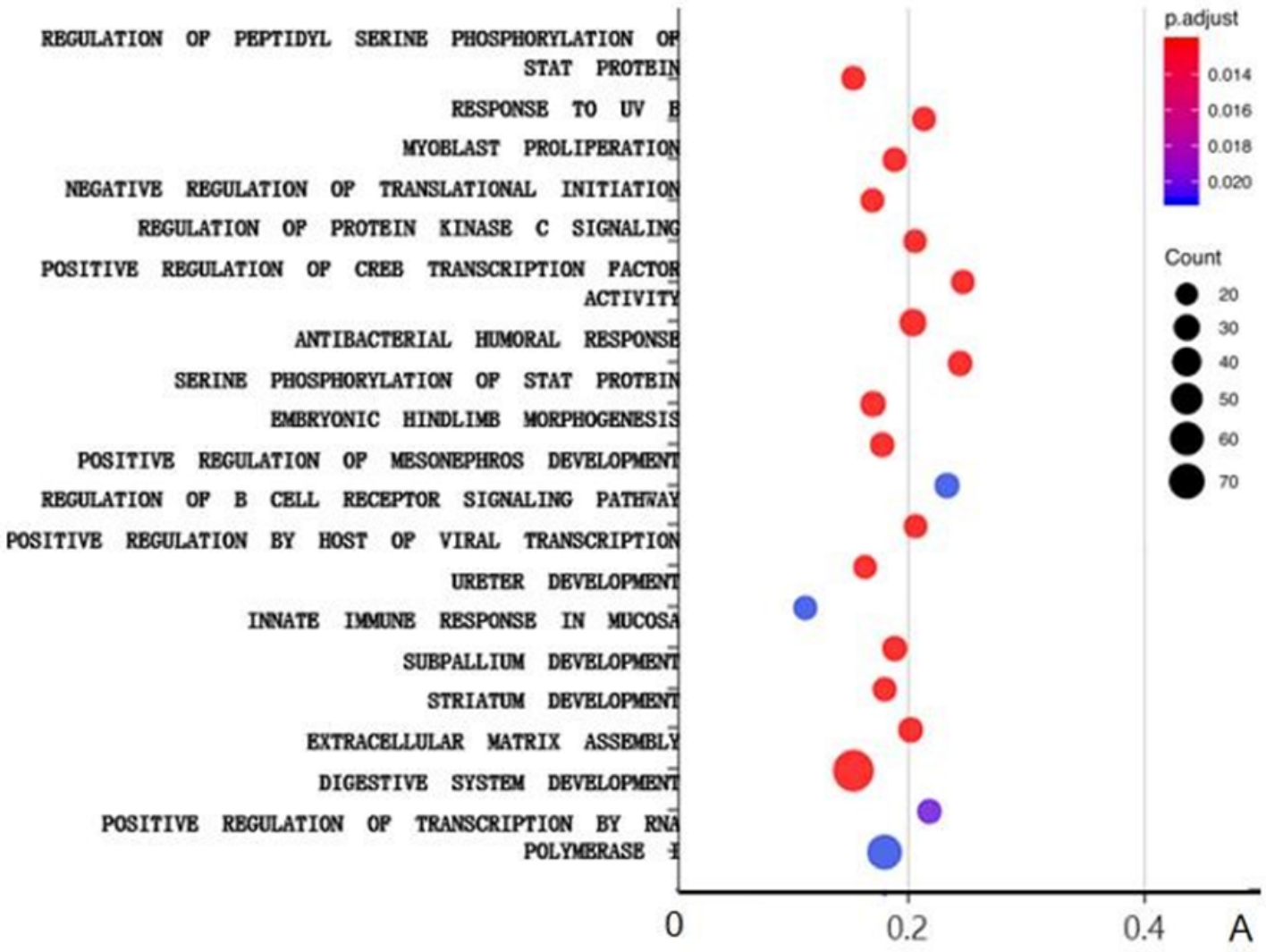

Figure 3

A. The vertical items are KEGG terms, and the horizontal graph are the gene ratio. The depth of the color represents the adjusted p-value. The area of the circle represents gene counts.

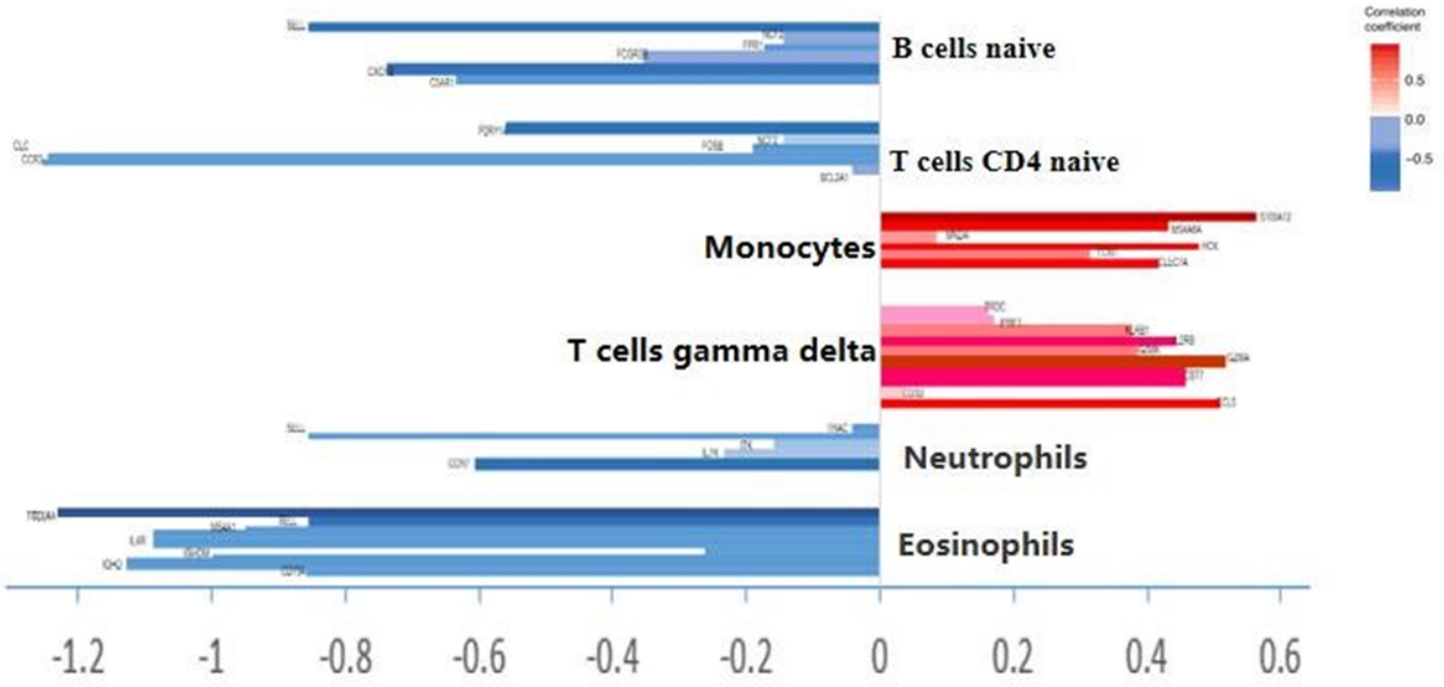

Figure 4

Correlation analysis between core genes in immune response and six types of immune infiltrating cells. The black bold items are the names of the six correlation immune cells. Red items represent positive correlation, the green items represent negative correlation.