GBP4 is an Accurate Diagnostic Biomarker and a Potential Treatment Target for Crohn's Disease

Heng Shi
The First Affiliated Hospital of Jinan University, Jinan University
Qin Peng
The Central Hospital of Shaoyang, University of South China
Xian-Ling Zhou
The First Affiliated Hospital of Jinan University, Jinan University
Shi-Ping Zhu
First Affiliated Hospital of Jinan University
Sheng-Yun Sun (shengyunsun2020@163.com)
First Affiliated Hospital of Jinan University

Research Article

Keywords: GBP4, Crohn's disease, immune cells, biomarker

Posted Date: December 10th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1144474/v1

Abstract

Background: Extensive evidence has shown that immune cell infiltration is associated with the pathogenesis of Crohn's disease (CD). In the present study, we explored the potential mechanism underlying the pathogenesis biomarkers for CD.

Methods: The GSE179285 dataset containing sequence data for intestinal mucosal was downloaded from the Gene Expression Omnibus (GEO) database. Differentially expressed genes (DEGs) in the intestinal mucosa of CD patients and healthy individuals were then identified. The infiltration pattern of 22 immune cell types was assessed using the CIBERSORT algorithm. The DEGs and 22 immune cell types were combined to find the key gene network using weighted gene co-expression network analysis (WGCNA), and pathway enrichment analyzes were performed on the hub module in the WGCNA. A linear regression model for the relationship between the expression of the hub genes in CD patients and infiltration of immune cells were also developed. The utility and accuracy of the hub genes for CD diagnosis were assessed using receiver operating characteristic (ROC) analysis. The accuracy of the model was validated using GSE20881 dataset.

Results: There were 1135 DEGs between the intestinal mucosal tissue of CD patients and healthy individuals. Of these DEGs, 711 genes were upregulated, whereas 424 of them were downregulated. There was also a significant difference in the infiltration of immune cells to the intestinal mucosal between the CD patients and healthy individuals. WGCNA revealed that the turquoise module genes were strongly correlated with the infiltration of M1 macrophages (cor=0.68, $p=10^{-16}$). Pathway enrichment analysis further showed the genes in the turquoise module mainly regulated the secretion of interferon-gamma and other immune effector molecules. Finally, the expression of GBP4, the identified hub gene, strongly correlated with the infiltration of M1 macrophages (adjusted r-squared=0.661, $p<2x10^{-16}$), and is a relatively good marker for CD diagnostic prediction (AUC=0.736). The relationship between GBP4 expression and infiltration of M1 macrophages (adjusted r-squared=0.435, $p<2x10^{-16}$) and prognostic value of the gene (AUC=0.702) were verified using the GSE20881 validation dataset.

Conclusion: GBP4 is a potential biomarker for accurate CD diagnosis. The expression of GBP4 promotes the infiltration of M1 macrophages to the intestinal mucosa of CD patients.

Introduction

Crohn's disease (CD) is a chronic inflammatory bowel disease caused by genetic and environmental factors, and alteration in the composition and abundance of gut microbiota. Research also shows the disease can lead to severely debilitating and dysregulated immune response [1, 2]. The incidence of CD is increasing worldwide, but it is highest in North America and Northern Europe [3-5]. In China, the economic growth in the country has paralleled an increase in the incidence of CD [6, 7]. Although the precise etiology of CD remains unclear, dysregulated and excessive immune responses against pathogenic gut microbiota have been implicated in the development of CD [8]. Obviously, immune responses, especially

the immune cells, play an important role in CD. Traditional techniques such as immunohistochemistry and flow cytometry, do not explicitly reveal the immune landscape in the intestinal mucosa of CD patients. Among the more well-studied genes, such as NOD2[9, 10], CARD15 [11, 12] and PRKCQ [13], have been implicated in CD occurrence and development. However, these genes are not entirely related to immune response and therefore are not ideal targets for immunotherapy. As the immunotherapy has been recommended by clinical guideline of CD treatment [14, 15], it is imperative to identify reliable targets in CD patients for immunotherapy. CIBERSORT is a gene expression-based algorithm that accurately reveals the infiltration pattern of immune cells based on gene expression profiles [16]. We investigated infiltration of 22 immune cell types to the intestinal mucosa of CD patients and healthy individuals. Weighted gene co-expression network analysis (WGCNA) is a bioinformatics analytical method for accurate exploration of the relationships between genes and phenotypes [17]. The distinct advantage of WGCNA is that genes can be clustered into co-expression modules, which connect the phenotypic characteristics and the changes in gene expression. The diagnostic value of hub genes can be assessed using receiver operating characteristic (ROC) curve analysis [18]. In the present study, the gene-sequence data for the infiltration of immune cells to the intestinal mucosa of CD patients and healthy individuals were downloaded from the Gene Expression Omnibus (GEO) database. The finding of this study will unpack the complex activities in the immune microenvironment of intestinal mucosa of CD patients, which may reveal new therapeutic targets for the treatment of the disease.

Results

CD microarray datasets

The diagrammatic flow of this study was shown in Figure 1. GSE179285[22] and GSE20881[23] datasets were used in this study. GSE179285 was the training set, whereas GSE20881 was the validation set. Data on GSE number, numbers of samples, gender, sites of mucosal collection, platform, and inflammation are shown in Table 1. There was no statistically significant difference ($p > 0.05$) between the training dataset and the validation dataset.

DEGs between CD patients and healthy individuals

Based on the GSE179285, there were 1135 DEGs between CD patients and healthy individuals, in which 711 genes were upregulated whereas 424 genes were downregulated (Figure 2A). The expression profile of the top 50 most upregulated genes and the top 50 most downregulated genes (Additional file 1) were displayed using a heatmap (Figure 2B). The upregulated genes occurred in the ileum, whereas the downregulated genes occurred in colon.

Immune cell infiltration

The proportion of immune cells varied between the intestinal mucosa tissues of CD patients and normal individuals (Figure 3A-3B, Table 2). Compared with normal tissue, the proportion of CD8 T cells, activated CD4 T cells memory, M1 Macrophages, and neutrophils were significantly higher in the

intestinal mucosa of CD patients. Contrarily, a reverse trend was observed for T regulatory cells (Tregs), gamma delta T cells, activated NK cells, M2 Macrophages, and resting Mast cells (Figure 4A). The proportions of plasma cells, CD4 naïve T cells, activated dendritic cells were almost insignificant. There was a strong positive correlation between infiltration of M1 Macrophages and neutrophils (Pearson correlation = 0.519, $p < 0.0001$), but a strong negative correlation between infiltration of resting Mast cells and activated Mast cells (Pearson correlation = -0.523, $p < 0.001$) (Figure 4B) Overall, these findings demonstrated the complex, intricate network of immune response in the intestinal mucosa of CD patients.

WGCNA and identification of hub genes

The soft thresholding power β was set at 18 in the subsequent analysis, because the scale independence reached 0.85 and had a relatively high-average connectivity (Figure 5A). A total of 23 outliner samples were detected, and the height cut-off value was set at 680 (Figure 5B). Four coexpression modules of DEGs were constructed by WGCNA (Figure 6A), and the relationship between modules and infiltration of the immune cells was performed. We found the most significant correlation between the turquoise module and infiltration of Macrophages M1 (cor=0.68, $p=1x10^{-25}$) (Figure 6B). The immune-related gene in the turquoise module (*GBP4*) was then identified based on MM > 0.9 and GS > 0.7 (Figure 7A). The expression level of the hub gene is shown in Figure 7B. Compared with healthy individuals, *GBP4* was significantly upregulated in the colon and ileum of CD patients.

Functional enrichment analysis

The GO analysis showed that the brown module mainly regulated vesicle coating, vesicle targeting, Golgi vesicle budding, positive regulation of lipid biosynthetic process, and lipoprotein particle assembly, the grey module mainly regulated antigen processing and presentation, adaptive immune response, reactive oxygen species responses, interferon-gamma responses, immune effector process regulation, and the turquoise module mainly regulated interferon-gamma responses, immune effector process regulation, regulation of response to biotic stimulus, positive regulation of cytokine production, and leukocyte cell-cell adhesion (Figure 7C, Additional file 2). The KEGG analysis further revealed that the grey module mainly regulated antigen processing and presentation, allograft rejection, viral myocarditis, graft-versus-host disease, and Type I diabetes mellitus, and the turquoise module mainly regulated antigen processing and presentation, allograft rejection, viral myocarditis, staphylococcus aureus infection, pertussis, cytokine-cytokine receptor interaction, leishmaniasis, and viral protein interaction with cytokine pathways (Figure 7D, Additional file 3).

Linear model and ROC curve analysis

There was a positive linear correlation between the expression of *GBP4* and infiltration of M1 Macrophages to the intestinal mucosa of CD patients (Macrophage M1=0.0359382+0.0061959**GBP4*, adjust r-squared=0.661, $p < 2x10^{-16}$) (Figure 8A). The AUC for the diagnostic value of *GBP4* for CD was 0.736 (Figure 8B). The strong correlation between the expression of *GBP4* and infiltration of

Macrophages M1(Macrophage M1=0.0009155+0.1334921**GBP4*, adjust r-squared=0.435, $p < 2x10^{-16}$), as well as the good diagnostic value of the gene for CD (AUC=0.702) (Figure 8D) was confirmed using the validation set.

Discussion

CD is a relapsing inflammatory disease, mainly affecting the gastrointestinal tract, and frequently presents with abdominal pain, fever, bowel obstruction or as well as bloody or mucoid diarrhea [24]. The precise pathogenesis of CD remains unclear, but it has been linked to excessive immune response [25-27]. Unraveling the complex immune network underlying CD pathogenesis can uncover new targets for the treatment of the disease.

In the present study, we identified 1135 DEGs between CD patients and healthy individuals, some of which have been previously reported. *OLFM4*, which was the most upregulated gene, negatively regulates *H. pylori*-specific immune responses [28] and mucosal defense responses during inflammatory bowel disease [29]. The downregulated gene, *FABP1*, is a validated biomarker of CD diagnosis [30]. The function of other notable in CD such as *CHP2* is not well understood. Furthermore, the upregulated gene expression was observed in the ileum, which is the most common site for the disease [31].

CIBERSORT revealed a significant difference in proportion of immune cells in the intestinal mucosa of CD and healthy individuals. Macrophage and CD4+ T cells accounted for the largest proportion of the infiltrating immune cells. So far, it had already been reported that macrophage and CD4+ T cells played an important role in CD [32, 33]. Intestinal macrophages are a heterogeneous population of cells thought to be derived from classical blood monocytes, mediated by CCR2[34]. During inflammation, the recruited monocytes differentiate into inflammatory macrophages sensitive to stimulation by Toll-like receptors. The macrophages also secret proinflammatory cytokines, further promoting inflammation [35-37]. In CD patients, the CD14+ macrophages, which secret abundant TNF-α, are the largest proportion of immune cells on the inflamed mucosa [38, 39]. The proportion of infiltrating macrophages in the intestinal mucosa of CD patients is in line with our analysis by CIBERSORT. CD4+ T cells can also release a large amount of proinflammatory cytokines such as IFN-γ and IL-17/IL-22, and these cytokines contribute to the progression of CD [40]. We observed a significant difference in the proportion of resting NK cells, activated NK cells, monocytes, resting mast cells, and neutrophils in the intestinal mucosa of CD patients and normal individuals. Monocytes regulate the phagocytosis of pathogens, digesting processing and presentation of antigens, and releases of effector molecules such as chemokines and cytokines. Moreover, monocytes are thought to be the only source of intestinal macrophages, and changes in the composition of peripheral blood monocytes in CD patients have been reported [41, 42]. NK cells provide a rapid innate immune response, killing target cells without priming. Mast cells, which predominate at mucosal surfaces, are also crucial for early host defense. Mast cells selectively recruit and positively modulate the function of NK cells through soluble mediators such as interferons [43].

WGCNA of the GSE179285 dataset identified a strong link between the turquoise module and infiltration of macrophages M1. GO analysis revealed the genes in the turquoise module mainly regulate interferon-gamma response, regulation of immune effector process, regulation of response to biotic stimulus positive, regulation of cytokine production, and leukocyte cell-cell adhesion. Interferon-gamma can induce transcription of metal transporter, which contributes to CD pathogenesis [44]. Interferon-gamma-target therapy can be used in treating active CD [45]. Inflammation is closely related to regulating the immune effector process, response to biotic factors, production of cytokine, and adhesion of leukocytes to endothelial cells [46]. KEGG analyses demonstrated that *staphylococcus aureus* infection, pertussis, cytokine-cytokine receptor interaction, leishmaniasis, and interaction of viral protein with cytokine and cytokine receptor were important pathways in our study. *Staphylococcus aureus* [47], pertussis [48], and leishmaniasis [49] are some of the opportunistic infections in CD patients due to the immunomodulation and immunosuppressive therapies.

Herein, we found a strong linear relationship between the expression of *GBP4* and the infiltration of M1 macrophages in CD patients. Guanylate Binding Protein 4 (*GBP4*) regulates innate immune response via interferon gamma. GO annotations revealed *GBP4* regulates several biological processes, including *GTP* binding and *GTP*ase activity. Little is known about the *GBP* families. In mice, *GBP*s protect against lethal bacterial infections [50] through the *GBP*4 inflammasome-dependent production of prostaglandins [51]. Moreover, *GBP4* is an immune-related signature biomarker for predicting prognoses and immunotherapeutic responses in patients with muscle-invasive bladder cancer [52], and an immune microenvironment biomarker for the prognosis of ovarian cancer [53]. Also, *GBP4* takes part in the type-I interferon response and displays a positive correlation with macrophages [54]. However, there is no report about CD with *GBP4*.

Regarding limitations, first, the results are based on the computational algorithm. Although the accuracy of this technique has been validated, the finding of this study should be verified using in *vivo* experiments in the future. Second, given the small sample size, the finding of this study may have been exaggerated.

Conclusion

In conclusion, there is a significant difference in the infiltration of immune cells to intestinal mucosa tissues of CD patients and healthy individuals. Given that *GBP4* is a differently expressed gene between healthy individuals and CD patients and is a driver gene of macrophages, the gene is a potential biomarker for the CD diagnosis and prognosis as well as an immunotherapeutic target for CD treatment.

Methods

Source of data

Gene expression data of CD patients and healthy individuals was downloaded from GEO database (http://www.ncbi.nlm.nih.gov/geo/). The screening criteria for the gene expression datasets were as

follows: (1) the study type was limited to expression profiling by array; (2) gene expression data in the intestinal mucosa of CD patients and normal individuals; (3) Each dataset contained for at least 100 samples; (4) analyzable processed data or raw data.

Data preprocessing and differential gene analysis

Data were preprocessed and analyzed using the R software (https://www.r-project.org/) through the following steps: (1) The probe names of each gene were converted to gene symbols, moreover, when a target gene corresponded to multiple probes, the average expression values of the probes was used to represent the expression level of the gene; (2) genes were excluded if the gene expression level was zero in more than half of the samples; (3) genes lacking expression level data for over 30% of the samples were also removed. Differential expression analysis was performed using the "*limma*" R package [19]. Adjusted p value < 0.05 and fold change >1.2 or fold change <-1.2 were set as the threshold for significant differential expression.

Immune infiltration analysis

The composition and proportion of 22 immune cells in the intestinal mucosa of CD patients and healthy individuals were estimated using the Cell-type Identification By Estimating Relative Subsets Of RNA Transcripts (CIBERSORT) tool in combination with leukocyte signature matrix (LM22) based on gene expression profiles of the cells [16]. The permutations (perm) of the deconvolution algorithm were set at 1000.

Construction of network and identification of hub genes

The coexpression network of DEGs and the infiltration of immune cells was performed as previously described [17]. First, the soft thresholding power β, to which coexpression similarity was raised to calculate adjacency, was calculated using the pickSoftThreshold function in the "*WGCNA*" R package. Second, the samples were clustered to identify any obvious outliers. Third, the coexpression network was then constructed. Fourth, key gene modules were identified using hierarchical clustering and the dynamic tree cut function. Gene significance (GS) and module membership (MM) were then calculated to match modules to specific immune cells. According to the correlation between the immune cells and ME and p value, and the module with the highest correlation coefficient and the smallest p value was selected as the most relevant module for the immune cells. Finally, the hub genes in the relevant module for the immune cells were identified based on MM > 0.9 and GS > 0.7.

Functional enrichment analysis

Biological process and pathway regulated by the genes in the modules were identified using Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway and Gene Ontology (GO) enrichment analysis via the "*clusterProfiler*" R package [20]. The cutoff of the q-value was set at 0.05.

Linear model and ROC curve analysis

The Best linear model for immune cell and hub genes was derived using a stepwise forward linear regression analysis. The following statistic model was developed: $y=\beta 0+\beta 1x1+\beta 2x2+\ldots+\beta ixi$, where y is the proportion of immune cell, xi is the expression value of hub genes, $\beta 0$ was the intercept of the regression equation, and βi is the regression coefficients.

The utility and accuracy of the hub genes for CD diagnosis were assessed by receiver operating characteristic (ROC) analysis using the "*ROCR*" R package [21]. The area under curve (AUC) was then calculated and screened for genes with AUC greater than 0.7.

Statistical analysis

Data were analyzed using R software (Rx64 4.0.3). Differences between two groups were analyzed using the Wilcoxon test, whereas the Kruskal-Wallis test used for multiple groups. The correlation between different immune cell subtypes to the intestinal mucosa of CD patients was performed using the Pearson correlation coefficient. Statistical significance was set at $p < 0.05$.

Abbreviations

CD, Crohn's disease; GEO, Gene Expression Omnibus; WGCNA, weighted gene co-expression network analysis; DEGs, differentially expressed genes; ROC, receiver operating characteristic; AUC, area under curve; CIBERSORT, Cell-type Identification By Estimating Relative Subsets Of RNA Transcripts; GS, Gene significance; MM, module membership; KEGG, Kyoto Encyclopedia of Genes and Genomes; GO, Gene Ontology.

Declarations

Author contributions

Conception and design: HS & QP; Administrative support: S-Y S; Provision of study materials or patients: HS & QP; Collection and assembly of data: HS & QP; Data analysis and interpretation: X-L Z, S-P Z; Manuscript writing: HS; Final approval of manuscript: All authors

Acknowledgements

Not applicable

Availability of data and materials

The data that support the findings of this study are openly available in GEO database (http://www.ncbi.nlm.nih.gov/geo/)

Conflicts of interests

The authors declare that they have no conflicts of interest.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Funding

None.

References

1. Ahmad T, Tamboli CP, Jewell D, Colombel JF: **Clinical relevance of advances in genetics and pharmacogenetics of IBD.***Gastroenterology* 2004, **126**(6):1533-1549.
2. Torres J, Mehandru S, Colombel JF, Peyrin-Biroulet L: **Crohn's disease.***Lancet* 2017, **389**(10080):1741-1755.
3. Benchimol EI, Mack DR, Nguyen GC, Snapper SB, Li W, Mojaverian N, Quach P, Muise AM: **Incidence, outcomes, and health services burden of very early onset inflammatory bowel disease.***Gastroenterology* 2014, **147**(4):803-813.e807; quiz e814-805.
4. Kaplan GG: **The global burden of IBD: from 2015 to 2025.***Nat Rev Gastroenterol Hepatol* 2015, **12**(12):720-727.
5. Rocchi A, Benchimol EI, Bernstein CN, Bitton A, Feagan B, Panaccione R, Glasgow KW, Fernandes A, Ghosh S: **Inflammatory bowel disease: a Canadian burden of illness review.***Can J Gastroenterol* 2012, **26**(11):811-817.
6. Leong RW, Lau JY, Sung JJ: **The epidemiology and phenotype of Crohn's disease in the Chinese population.***Inflamm Bowel Dis* 2004, **10**(5):646-651.
7. Mak WY, Mak OS, Lee CK, Tang W, Leung WK, Wong MTL, Sze ASF, Li M, Leung CM, Lo FH, et al: **Significant Medical and Surgical Morbidity in Perianal Crohn's Disease: Results from a Territory-Wide Study.***J Crohns Colitis* 2018, **12**(12):1392-1398.
8. Takagawa T, Kitani A, Fuss I, Levine B, Brant SR, Peter I, Tajima M, Nakamura S, Strober W: **An increase in LRRK2 suppresses autophagy and enhances Dectin-1-induced immunity in a mouse model of colitis.***Science translational medicine* 2018, **10**(444).
9. Fritz T, Niederreiter L, Adolph T, Blumberg RS, Kaser A: **Crohn's disease: NOD2, autophagy and ER stress converge.***Gut* 2011, **60**(11):1580-1588.
10. Hugot JP, Chamaillard M, Zouali H, Lesage S, Cézard JP, Belaiche J, Almer S, Tysk C, O'Morain CA, Gassull M, et al: **Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease.***Nature* 2001, **411**(6837):599-603.

11. Chamaillard M, Philpott D, Girardin SE, Zouali H, Lesage S, Chareyre F, Bui TH, Giovannini M, Zaehringer U, Penard-Lacronique V, et al: **Gene-environment interaction modulated by allelic heterogeneity in inflammatory diseases.***Proc Natl Acad Sci U S A* 2003, **100**(6):3455-3460.
12. Berrebi D, Maudinas R, Hugot JP, Chamaillard M, Chareyre F, De Lagausie P, Yang C, Desreumaux P, Giovannini M, Cézard JP, et al: **Card15 gene overexpression in mononuclear and epithelial cells of the inflamed Crohn's disease colon.***Gut* 2003, **52**(6):840-846.
13. Zhang YW, Xu XY, Zhang J, Yao X, Lu C, Chen CX, Yu CH, Sun J: **Missense mutation in PRKCQ is associated with Crohn's disease.***J Dig Dis* 2019, **20**(5):243-247.
14. Lichtenstein GR, Loftus EV, Isaacs KL, Regueiro MD, Gerson LB, Sands BE: **ACG Clinical Guideline: Management of Crohn's Disease in Adults.***Am J Gastroenterol* 2018, **113**(4):481-517.
15. Torres J, Bonovas S, Doherty G, Kucharzik T, Gisbert JP, Raine T, Adamina M, Armuzzi A, Bachmann O, Bager P, et al: **ECCO Guidelines on Therapeutics in Crohn's Disease: Medical Treatment.***J Crohns Colitis* 2020, **14**(1):4-22.
16. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA: **Robust enumeration of cell subsets from tissue expression profiles.***Nature methods* 2015, **12**(5):453-457.
17. Langfelder P, Horvath S: **WGCNA: an R package for weighted correlation network analysis.***BMC bioinformatics* 2008, **9**:559.
18. Metz CE: **Basic principles of ROC analysis.***Seminars in nuclear medicine* 1978, **8**(4):283-298.
19. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK: **limma powers differential expression analyses for RNA-sequencing and microarray studies.***Nucleic acids research* 2015, **43**(7):e47.
20. Yu G, Wang LG, Han Y, He QY: **clusterProfiler: an R package for comparing biological themes among gene clusters.***Omics : a journal of integrative biology* 2012, **16**(5):284-287.
21. Sing T, Sander O, Beerenwinkel N, Lengauer T: **ROCR: visualizing classifier performance in R.***Bioinformatics (Oxford, England)* 2005, **21**(20):3940-3941.
22. Keir ME, Fuh F, Ichikawa R, Acres M, Hackney JA, Hulme G, Carey CD, Palmer J, Jones CJ, Long AK, et al: **Regulation and Role of αE Integrin and Gut Homing Integrins in Migration and Retention of Intestinal Lymphocytes during Inflammatory Bowel Disease.***Journal of immunology (Baltimore, Md : 1950)* 2021, **207**(9):2245-2254.
23. Noble CL, Abbas AR, Lees CW, Cornelius J, Toy K, Modrusan Z, Clark HF, Arnott ID, Penman ID, Satsangi J, Diehl L: **Characterization of intestinal gene expression profiles in Crohn's disease by genome-wide microarray analysis.***Inflamm Bowel Dis* 2010, **16**(10):1717-1728.
24. Baumgart DC, Sandborn WJ: **Crohn's disease.***Lancet* 2012, **380**(9853):1590-1605.
25. Kemler BJ, Alpert E: **Inflammatory bowel disease associated circulating immune complexes.***Gut* 1980, **21**(3):195-201.
26. Cassinotti A, Sarzi-Puttini P, Fichera M, Shoenfeld Y, de Franchis R, Ardizzone S: **Immunity, autoimmunity and inflammatory bowel disease.***Autoimmunity reviews* 2014, **13**(1):1-2.

27. Geremia A, Biancheri P, Allan P, Corazza GR, Di Sabatino A: **Innate and adaptive immunity in inflammatory bowel disease.***Autoimmunity reviews* 2014, **13**(1):3-10.
28. Liu W, Yan M, Liu Y, Wang R, Li C, Deng C, Singh A, Coleman WG, Jr., Rodgers GP: **Olfactomedin 4 down-regulates innate immunity against Helicobacter pylori infection.***Proc Natl Acad Sci U S A* 2010, **107**(24):11056-11061.
29. Gersemann M, Becker S, Nuding S, Antoni L, Ott G, Fritz P, Oue N, Yasui W, Wehkamp J, Stange EF: **Olfactomedin-4 is a glycoprotein secreted into mucus in active IBD.***J Crohns Colitis* 2012, **6**(4):425-434.
30. Dooley TP, Curto EV, Reddy SP, Davis RL, Lambert GW, Wilborn TW, Elson CO: **Regulation of gene expression in inflammatory bowel disease and correlation with IBD drugs: screening by DNA microarrays.***Inflamm Bowel Dis* 2004, **10**(1):1-14.
31. Peyrin-Biroulet L, Loftus EV, Jr., Colombel JF, Sandborn WJ: **The natural history of adult Crohn's disease in population-based cohorts.***Am J Gastroenterol* 2010, **105**(2):289-297.
32. Kobayashi T, Okamoto S, Iwakami Y, Nakazawa A, Hisamatsu T, Chinen H, Kamada N, Imai T, Goto H, Hibi T: **Exclusive increase of CX3CR1+CD28-CD4+ T cells in inflammatory bowel disease and their recruitment as intraepithelial lymphocytes.***Inflamm Bowel Dis* 2007, **13**(7):837-846.
33. Gren ST, Grip O: **Role of Monocytes and Intestinal Macrophages in Crohn's Disease and Ulcerative Colitis.***Inflamm Bowel Dis* 2016, **22**(8):1992-1998.
34. Bain CC, Bravo-Blas A, Scott CL, Perdiguero EG, Geissmann F, Henri S, Malissen B, Osborne LC, Artis D, Mowat AM: **Constant replenishment from circulating monocytes maintains the macrophage pool in the intestine of adult mice.***Nature immunology* 2014, **15**(10):929-937.
35. Kamada N, Hisamatsu T, Okamoto S, Chinen H, Kobayashi T, Sato T, Sakuraba A, Kitazume MT, Sugita A, Koganei K, et al: **Unique CD14 intestinal macrophages contribute to the pathogenesis of Crohn disease via IL-23/IFN-gamma axis.***J Clin Invest* 2008, **118**(6):2269-2280.
36. Grimm MC, Pullman WE, Bennett GM, Sullivan PJ, Pavli P, Doe WF: **Direct evidence of monocyte recruitment to inflammatory bowel disease mucosa.***J Gastroenterol Hepatol* 1995, **10**(4):387-395.
37. Platt AM, Bain CC, Bordon Y, Sester DP, Mowat AM: **An independent subset of TLR expressing CCR2-dependent macrophages promotes colonic inflammation.***Journal of immunology (Baltimore, Md : 1950)* 2010, **184**(12):6843-6854.
38. Tamoutounour S, Henri S, Lelouard H, de Bovis B, de Haar C, van der Woude CJ, Woltman AM, Reyal Y, Bonnet D, Sichien D, et al: **CD64 distinguishes macrophages from dendritic cells in the gut and reveals the Th1-inducing role of mesenteric lymph node macrophages during colitis.***European journal of immunology* 2012, **42**(12):3150-3166.
39. Zigmond E, Varol C, Farache J, Elmaliah E, Satpathy AT, Friedlander G, Mack M, Shpigel N, Boneca IG, Murphy KM, et al: **Ly6C hi monocytes in the inflamed colon give rise to proinflammatory effector cells and migratory antigen-presenting cells.***Immunity* 2012, **37**(6):1076-1090.
40. Strober W, Fuss IJ: **Proinflammatory cytokines in the pathogenesis of inflammatory bowel diseases.***Gastroenterology* 2011, **140**(6):1756-1767.

41. Chapuy L, Sarfati M: **Single-Cell Protein and RNA Expression Analysis of Mononuclear Phagocytes in Intestinal Mucosa and Mesenteric Lymph Nodes of Ulcerative Colitis and Crohn's Disease Patients.***Cells* 2020, **9**(4).
42. Lissner D, Schumann M, Batra A, Kredel LI, Kühl AA, Erben U, May C, Schulzke JD, Siegmund B: **Monocyte and M1 Macrophage-induced Barrier Defect Contributes to Chronic Intestinal Inflammation in IBD.***Inflamm Bowel Dis* 2015, **21**(6):1297-1305.
43. Portales-Cervantes L, Dawod B, Marshall JS: **Mast Cells and Natural Killer Cells-A Potentially Critical Interaction.***Viruses* 2019, **11**(6).
44. Melia JMP, Lin R, Xavier RJ, Thompson RB, Fu D, Wan F, Sears CL, Donowitz M: **Induction of the metal transporter ZIP8 by interferon gamma in intestinal epithelial cells: Potential role of metal dyshomeostasis in Crohn's disease.***Biochem Biophys Res Commun* 2019, **515**(2):325-331.
45. Cui D, Huang G, Yang D, Huang B, An B: **Efficacy and safety of interferon-gamma-targeted therapy in Crohn's disease: a systematic review and meta-analysis of randomized controlled trials.***Clin Res Hepatol Gastroenterol* 2013, **37**(5):507-513.
46. Malizia G, Calabrese A, Cottone M, Raimondo M, Trejdosiewicz LK, Smart CJ, Oliva L, Pagliaro L: **Expression of leukocyte adhesion molecules by mucosal mononuclear phagocytes in inflammatory bowel disease.***Gastroenterology* 1991, **100**(1):150-159.
47. Bettenworth D, Nowacki TM, Friedrich A, Becker K, Wessling J, Heidemann J: **Crohn's disease complicated by intestinal infection with methicillin-resistant Staphylococcus aureus.***World J Gastroenterol* 2013, **19**(27):4418-4421.
48. Cleveland NK, Rodriquez D, Wichman A, Pan I, Melmed GY, Rubin DT: **Many Inflammatory Bowel Disease Patients Are Not Immune to Measles or Pertussis.***Dig Dis Sci* 2016, **61**(10):2972-2976.
49. Valdés Delgado T, Cordero Ruiz P, Bellido Muñoz F: **Visceral Leishmaniasis Infection in a Patient with Crohn's Disease Treated with Azathioprine.***J Crohns Colitis* 2017, **11**(10):1282-1283.
50. Wandel MP, Kim BH, Park ES, Boyle KB, Nayak K, Lagrange B, Herod A, Henry T, Zilbauer M, Rohde J, et al: **Guanylate-binding proteins convert cytosolic bacteria into caspase-4 signaling platforms.***Nature immunology* 2020, **21**(8):880-891.
51. Tyrkalska SD, Candel S, Angosto D, Gómez-Abellán V, Martín-Sánchez F, García-Moreno D, Zapata-Pérez R, Sánchez-Ferrer Á, Sepulcre MP, Pelegrín P, Mulero V: **Neutrophils mediate Salmonella Typhimurium clearance through the GBP4 inflammasome-dependent production of prostaglandins.***Nat Commun* 2016, **7**:12077.
52. Jiang W, Zhu D, Wang C, Zhu Y: **An immune relevant signature for predicting prognoses and immunotherapeutic responses in patients with muscle-invasive bladder cancer (MIBC).***Cancer Med* 2020, **9**(8):2774-2790.
53. Huo X, Sun H, Liu S, Liang B, Bai H, Wang S, Li S: **Identification of a Prognostic Signature for Ovarian Cancer Based on the Microenvironment Genes.***Front Genet* 2021, **12**:680413.
54. Ottenhoff TH, Dass RH, Yang N, Zhang MM, Wong HE, Sahiratmadja E, Khor CC, Alisjahbana B, van Crevel R, Marzuki S, et al: **Genome-wide expression profiling identifies type 1 interferon response**

pathways in active tuberculosis.*PLoS One* 2012, **7**(9):e45839.

Tables

Table 1 Characteristics of training and validation datasets

Characteristics		GSE179285 (Training set)	GSE20881 (Validation set)	P-value
Samples	CD	168	99	0.667
	Healthy controls	31	73	
Site of mucosal collection	Colon	109	150	0.667
	Ileum	90	22	
Gender	Female	13	27	0.793
	Male	27	26	
Inflammation	Inflamed	47	70	0.650
	Uninflamed	152	102	
Platform		GPL6480	GPL1708	

Notes: One patient or healthy individual may have one or more samples.

Table 2 Comparison of 22 proportion between CD and normal tissue

Immune cell	CIBERSORT fraction in % of all infiltrating immune cells (mean± SD)		
	CD tissue	Normal tissue	P-value
B cells naive	0.0233±0.0331	0.033±0.038	0.2538
B cells memory	0.0489±0.0542	0.0599±0.0713	0.7750
Plasma cells	0.0005±0.0024	0±0.0002	0.3943
T cells CD8	0.0441±0.0492	0.025±0.0369	0.0205
T cells CD4 naive	0.0002±0.0015	0.0002±0.0013	0.4076
T cells CD4 memory resting	0.1427±0.071	0.1452±0.0735	0.7587
T cells CD4 memory activated	0.0654±0.0517	0.0312±0.0328	0.0003
T cells follicular helper	0.0001±0.0008	0	0.5490
T cells regulatory Tregs.	0.0301±0.0228	0.0385±0.0226	0.0337
T cells gamma delta	0.0322±0.038	0.0444±0.0358	0.0201
NK cells resting	0.008±0.0175	0.0022±0.0096	0.0064
NK cells activated	0.0713±0.0472	0.0994±0.0472	0.0018
Monocytes	0.0087±0.0168	0.0015±0.0042	0.0188
Macrophages M0	0.1068±0.0643	0.0797±0.0613	0.0524
Macrophages M1	0.0814±0.0465	0.0595±0.0346	0.0162
Macrophages M2	0.156±0.0637	0.1939±0.0672	0.0024
Dendritic cells resting	0.0076±0.0142	0.0149±0.0273	0.0519
Dendritic cells activated	0	0	-
Mast cells resting	0.1088±0.0749	0.1437±0.0688	0.0114
Mast cells activated	0.032±0.0524	0.0105±0.0223	0.0625
Eosinophils	0.0094±0.0148	0.0096±0.0183	0.5390
Neutrophils	0.0226±0.0332	0.0077±0.008	0.0171

Notes: P values in red indicate statistical significance ($p < 0.05$).

Figures

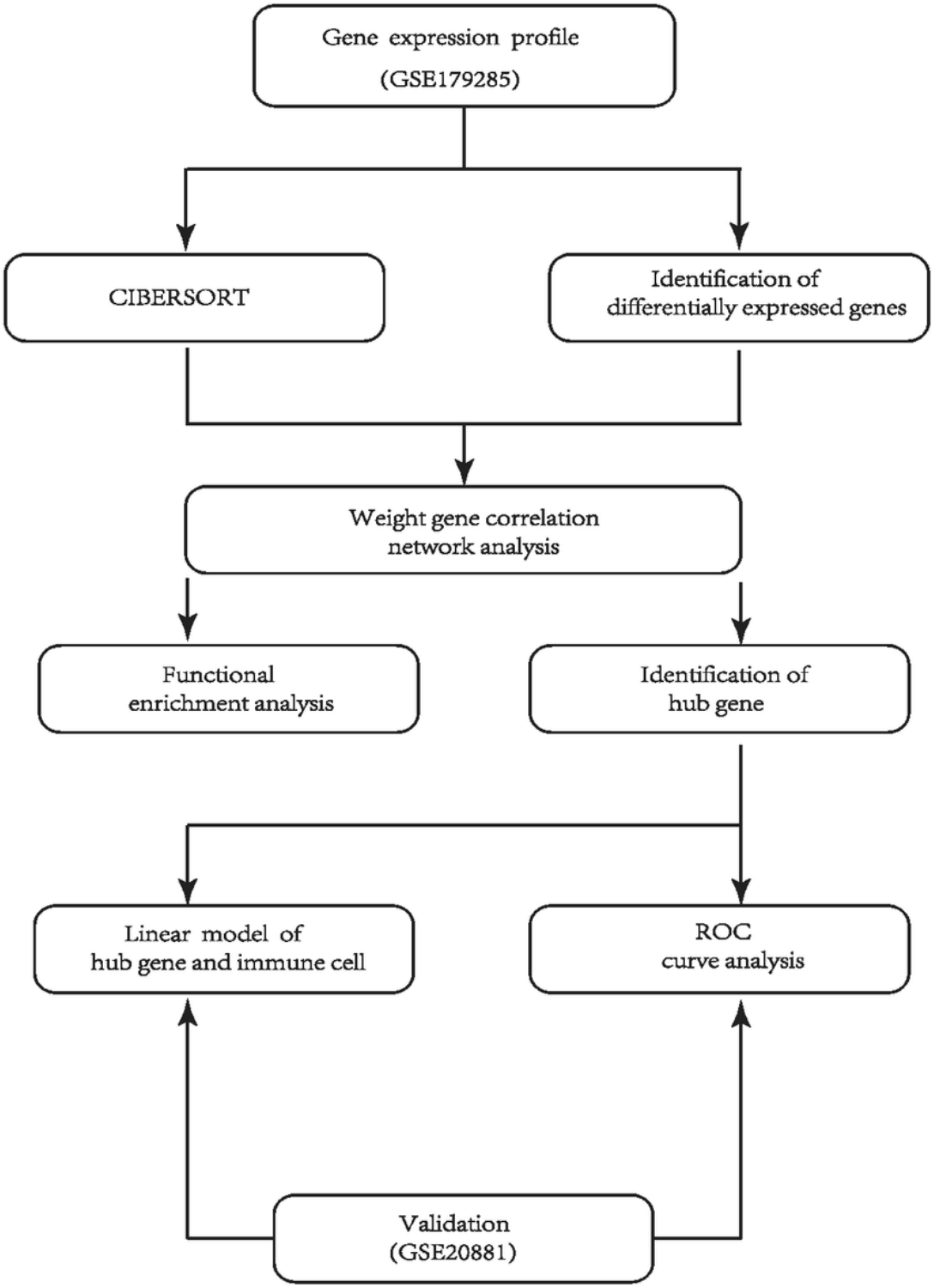

Figure 1

The diagrammatic workflow of the preset study.

Figure 2

(A) Volcano plots for the differentially expressed genes in the intestinal mucosa of CD patients and healthy individuals: red dots represent upregulated expressed genes, whereas blue dots represent downregulated expressed genes, and gray dots represent non-differentially expressed genes. (B). Heatmap for the top 50 most upregulated genes and the top 50 most downregulated genes.

Figure 3

The proportion of infiltrating immune cells in the intestinal mucosa. (A) Health individuals. (B) CD patients.

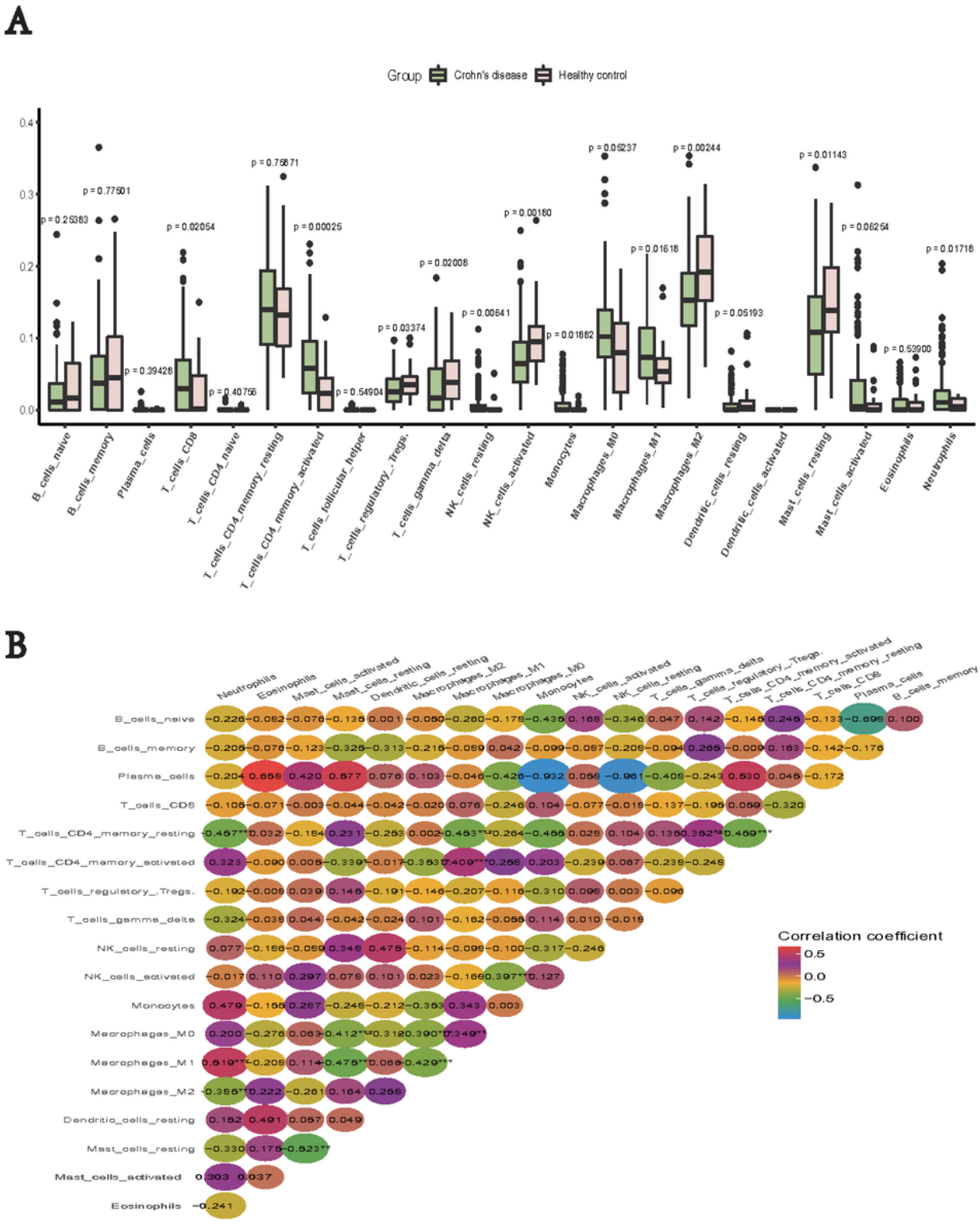

Figure 4

(A) Box plot for the differentially infiltrated immune cells in the intestinal mucosa between CD patients and normal individuals. (B) Correlation heatmap for the correlations between infiltrated immune cells in the intestinal mucosa of CD patients. The Pearson correlation coefficient was applied for the test, *, $p < 0.05$, **, $p < 0.01$, ***, $p < 0.001$.

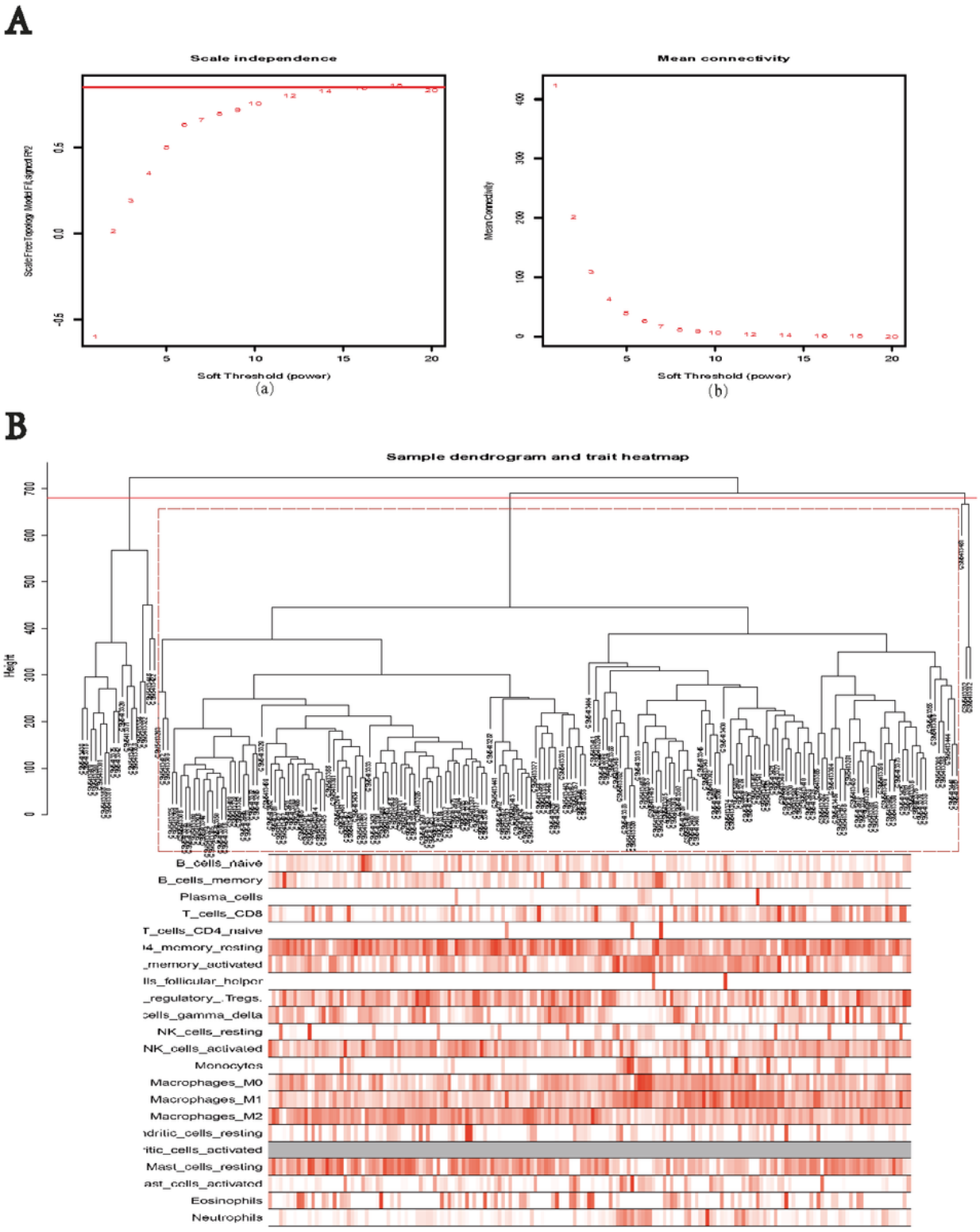

Figure 5

(A) The network topology of various soft-thresholding powers. (a) The x-axis represents the soft-thresholding power, whereas the y-axis represents the scale-free topology model fit index. (b) The x-axis reflects the soft-thresholding power. The y-axis reflects the mean connectivity (degree). (B) The sample dendrogram and the infiltration of immune cells heatmap. 23 outliner samples were identified. Only samples in a red dotted square box were included.

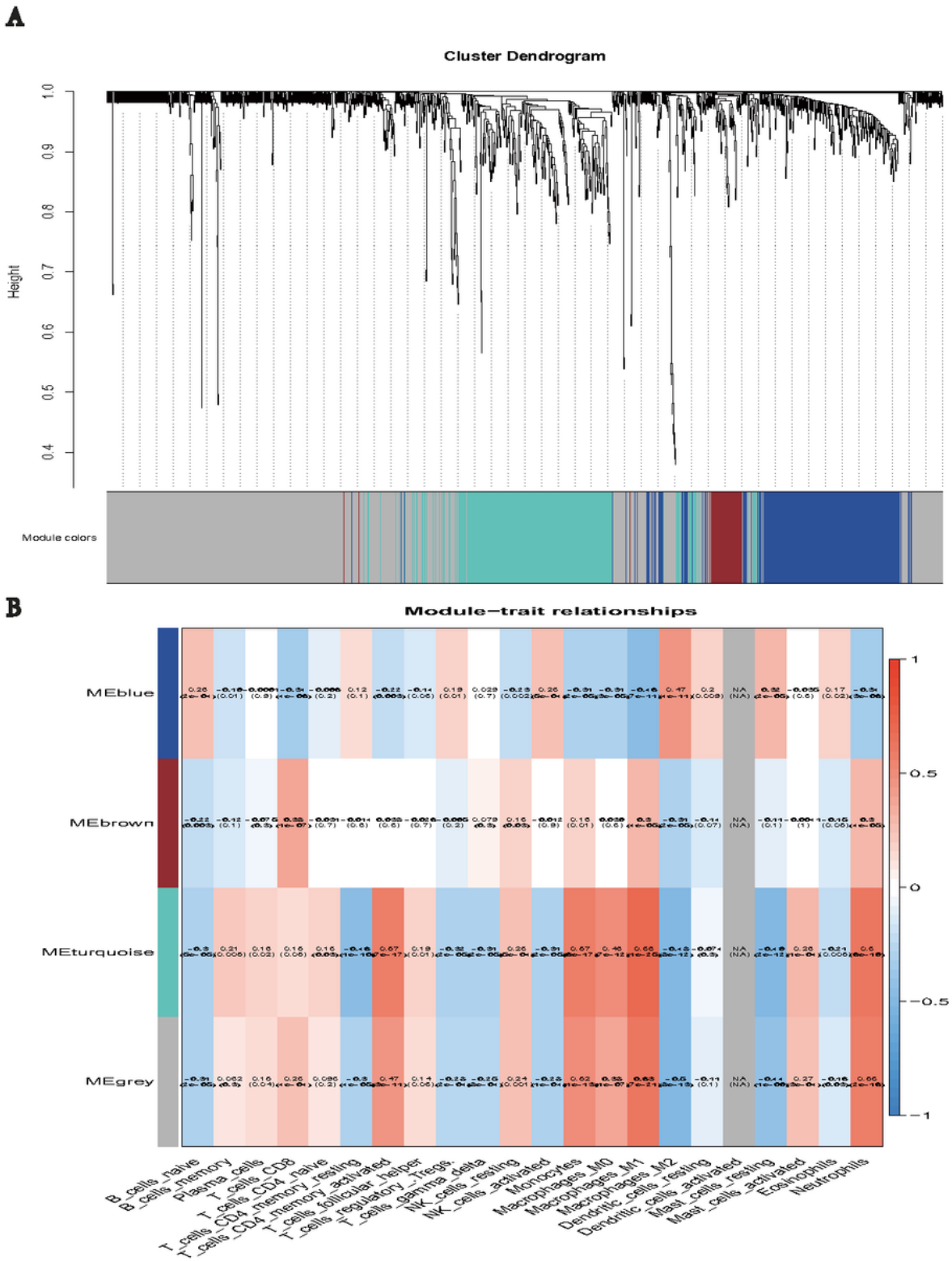

Figure 6

(A) Identification of modules by gene co-expression network. The branches of the cluster dendrogram represent the modules, whereas the leaves on the cluster dendrogram represent the genes. (B) Module-trait associations. Each row corresponds to a module, and each column corresponds to a proportion of infiltrating immune cell. Each cell contains the corresponding correlation and p-value.

Figure 7

(A) Scatter diagrams for each gene in turquoise module and M1 Macrophages. (B) Box plot for the expression level of the hub gene. (C) GO analysis of the genes in modules. The node size reflects the gene count, and the node color reflects the P -value [−log10 (P value)]. (D) KEGG analysis of genes in modules. The node size reflects the gene count, and the node color reflects the P-value [−log10 (P value)].

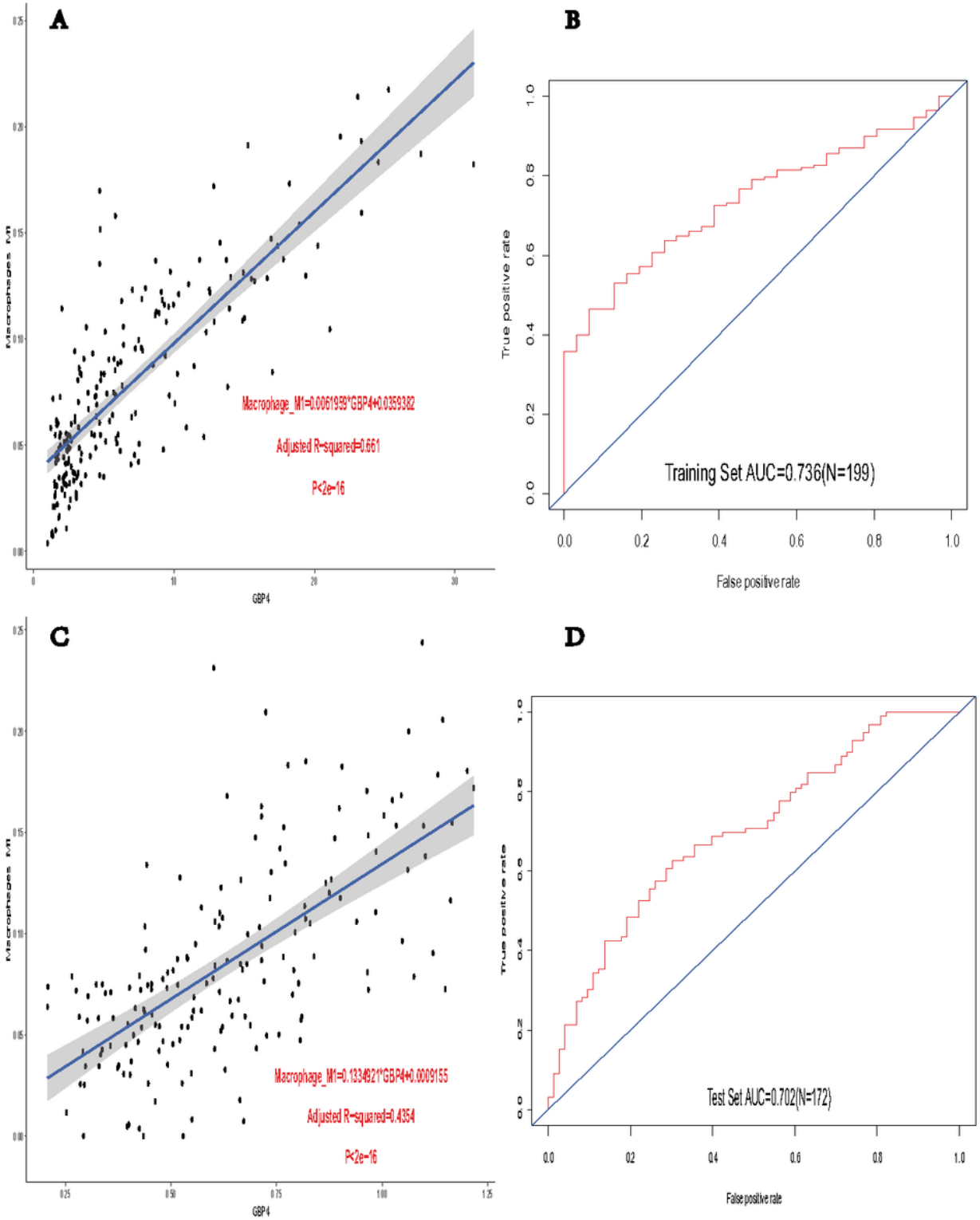

Figure 8

(A)Scatter plot and linear model of the hub gene and infiltration of M1 Macrophages based on GSE179285. (B) The ROC curve of the hub gene based on GSE179285. (C) Scatter plot and linear model of the hub gene and infiltration of M1 Macrophages based on GSE20881. (D) The ROC curve of the hub gene based on GSE20881.

Supplementary Files

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

- Additionalfile2.docx
- Additionalfile1.docx
- Additionalfile3.docx