Journal of Trends in Life Sciences
ISSN: 2960-0200
DOI: 10.61784/jtlsv2n276

SCREENING OF DIAGNOSTIC MARKERS FOR DEPRESSION BASED ON BIOINFORMATICS ANALYSIS

Wildman Santos
University of Michigan School of Public Health, Ann Arbor, MI, USA.

Abstract: Purpose: This study aimed to screen potential diagnostic markers in peripheral blood for major depressive disorder (MDD). Method: First, download the gene expression profile data set GSE32280 from the GEO database. The differentially expressed genes (DEGs) between MDD and normal control peripheral blood samples were screened using R software. The screened DEGs were subjected to GO functional annotation and KEGG pathway enrichment analysis; then, Cytoscape software was used to construct a protein-protein interaction (PPI) network, and key (hub) genes were screened out. ROC analysis was performed on hub genes using R software to identify hub genes with diagnostic value. Results: A total of 104 DEGs were screened out from the GSE32280 data set, including 47 up-regulated genes and 57 down-regulated genes. GO functional annotation showed that 104 DEGs were mainly involved in cell proliferation, inflammatory response, transport regulation and other functions. KEGG pathway enrichment analysis results showed that DEGs were mainly enriched in NK cell-mediated cytotoxicity, interaction between cytokines and their receptors, and chemokine signaling pathways. Obtain 16 hub genes from the PPI network. ROC analysis results of hub genes showed that CXCL1, EGF, IFNG and CXCL8 have high diagnostic value in MDD. Conclusion: CXCL1, EGF, IFNG and CXCL8 are important diagnostic markers for MDD.
Keywords: Major depressive disorder; Bioinformatics; Diagnostic markers

1 INTRODUCTION

Depression is a common neuropsychiatric disorder, ranking third in the global burden of disease [1]. Statistics from the World Health Organization show that there are approximately 350 million patients with depression worldwide[2]. Due to its high prevalence and disabling nature, depression brings serious health risks and heavy economic burden to individuals and society. Currently, the diagnosis and treatment of major depressive disorder (MDD) are based on the patient's symptoms and signs, and the objective criteria for early diagnosis of MDD patients still need to be elucidated [3, 4]. The development of genomics technology enables researchers to study gene expression and epigenetic changes in various diseases at the genome level, and the rapid development of bioinformatics analysis methods has brought new ideas to the interpretation of genomics results. It is widely used to analyze differentially expressed genes and screen biomarkers for disease diagnosis and treatment [5-8].

In this study, we downloaded the MDD expression profile data set GSE32280 from the GEO database (https://www. ncbi.nlm.nih. gov/geo/). R software was used to screen out differentially expressed genes (DEGs), and functional annotation and pathway enrichment analysis were performed on the DEGs. Then, a protein-protein interaction (PPI) network was constructed and disease-related key (hub) genes were identified. The diagnostic value of hub genes was analyzed through ROC curves in order to screen out molecular diagnostic markers for MDD.

1.1 Download the Data Set

From the National Center for Biotechnology Information, NCBI) GEO Database[9] Download the expression profile data set GSE32280. The GSE32280 data set includes a total of 16 peripheral blood samples, including 8 normal control samples. Example (GSM799727, GSM799728, GSM799729, GSM799730, GSM799731, GSM799732, GSM799733, GSM799734) and MDD Sample 8 For example (GSM799722, GSM799723, GSM799724, GSM799725, GSM799726, GSM799738, GSM799740, GSM799743). All samples in the data set adopt the AgentGPL570 platform ([HG-U133 Plus 2] Affymetrix Human Genome U133 Plus2.0Array) for analysis. The samples of the GSE32280 data set are downloaded from the same platform, and the uploaded data has been normalized and does not require further correction.

1.2 Identification of DEGs

Use the Limma package of R software to screen DEGs on the GSE32280 expression profile data set. Set the thresholds as adjusted $P < 0.05$, log FC <-0.5 (down-regulated genes), log FC > 0.5 (up-regulated genes) to define DEGs, and use the ggplot2 package [10] to draw a volcano plot of DEGs to visually display the differential expression of DEGs.

1.3 Functional Annotation and Pathway Enrichment Analysis of DEGs

Use David (https://david. ncifcrf. gov/home. jsp) to perform GO (geneontology) analysis. GO analysis includes cell components (cell component (CC), molecular function (MF) and biological process (BP). KEGG(kyoto enc y clopedia

of genes and genomes) is a database of pathway collections. We used the R software ClusterProfile package [11] to perform GO and KEGG pathway enrichment analysis on DEGs.

1.4 Construct PPI Network and Identify Hub Genes

STRING (https://string-db.org/) is an online tool designed to evaluate protein-protein interactions[12]. The MCODE module in Cy toscape software was used to analyze and visualize the STRING results. Use the Cy toHubba module to select the top 16 largest relevant criteria gene as hub Gene. Network Anal y st (https://www.networkanalyst.ca/faces/home.xht-ml) is a PPI network visual analysis platform. We entered 16 hub genes into NetworkAnalyst to visualize the PPI network.

1.5 Diagnostic Biomarker Screening

The pROC package of R software [13] was used to perform ROC curve analysis to evaluate the diagnostic value of 16 hub genes and screen out diagnostic biomarkers for MDD.

1.6 Statistical Analysis

Data were measured using R software (version 3.5.2), and $P < 0.05$ was considered statistically significant.

2 RESULT

2.1 Explicit DEGs

The GSE32280 data set includes a total of 16 Among the exceptional peripheral blood samples, 8 were normal samples and 8 were MDD samples. R software was used to screen DEGs between normal and MDD samples, and the thresholds were set as adjusted $P < 0.05$, $logFC < -0.5$ (down-regulated genes), and $logFC > 0.5$ (up-regulated genes).

2.2 GO Functional Annotation and KEGG Pathway Enrichment Analysis

GO functional annotation analysis was performed on the 104 DEGs screened out from the GSE32280 data set. The results showed that changes in BP that down-regulated DEGs were mainly concentrated in cell proliferation, inflammatory response, and transport regulation; changes in CC included platelet granules and secretory granules; changes in MF were mainly concentrated in cytokine activity, chemokine activity, and chemokines Degree of receptor binding (Figure 1 A). The BPs mainly involved in the up-regulation of DEGs include immune response, tumor necrosis factor production and defense mechanisms; changes in CC include the cytoplasmic membrane and cell surface; changes in MF are mainly enriched in MHC proteins and polysaccharide anchors (Figure 1B). KEGG pathway enrichment analysis results showed that DEGs were mainly enriched in NK cell-mediated cytotoxicity, interaction between cytokines and their receptors, and chemokine signaling pathways (Figures 1C and 1D).

2.3 GSEA Enrichment Analysis of MDD-Related Pathways and Genes

GSEA analysis results show that the enriched pathways include mature diabetes pathway, glycosaminoglycan biosynthesis, glycolysis synthesis pathway, calcium signaling pathway, dorsal-ventral axis formation, etc. (Figure 2). Ryanodine receptor 2 (RYR2), calcium voltage-gated channel subunit alpha1c (CACNA1C) and calcium voltage-gated channel subunit alpha1s (CACNA1S) are important genes in these pathways (Figure 3)

2.4 Construct PPI Network and Select Hub Genes

Analyze the DEGs selected from the GSE32280 data set through STRING. The analysis results were imported into Cytoscape, and the MCODE plug-in was used to construct the PPI network. The PPI network is shown in Figure 4A, and 10 nodes and 41 pairs of PPI relationships were obtained. Then, using the Cy to-Hubba plug-in, 16 hub genes closely related to MDD were obtained, namely CXCL8, IFNG, EGF, CXCL1, TLR3, PTGS2, CXCL5, CCL20, IL1RN, CXCL3, FASLG, CCR10, THBS1, CLEC7A, TNFSF4, OSM (Figure 4B).

2.5 Identify Diagnostic Markers for MDD

ROC analysis was used to evaluate the diagnostic value of 16 hub genes in MDD. CXCL1, EGF, CXCL8 and IFNG The areas under the ROC curve (area under the curve, AUC) are 0.865, 0.799, 0.737, and 0.705 respectively, and the AUC is between 0.700 and 0.900.
are considered to have high diagnostic value. Therefore, CXCL1, EGF, CXCL8 and IFNG are genes with high diagnostic value in MDD.

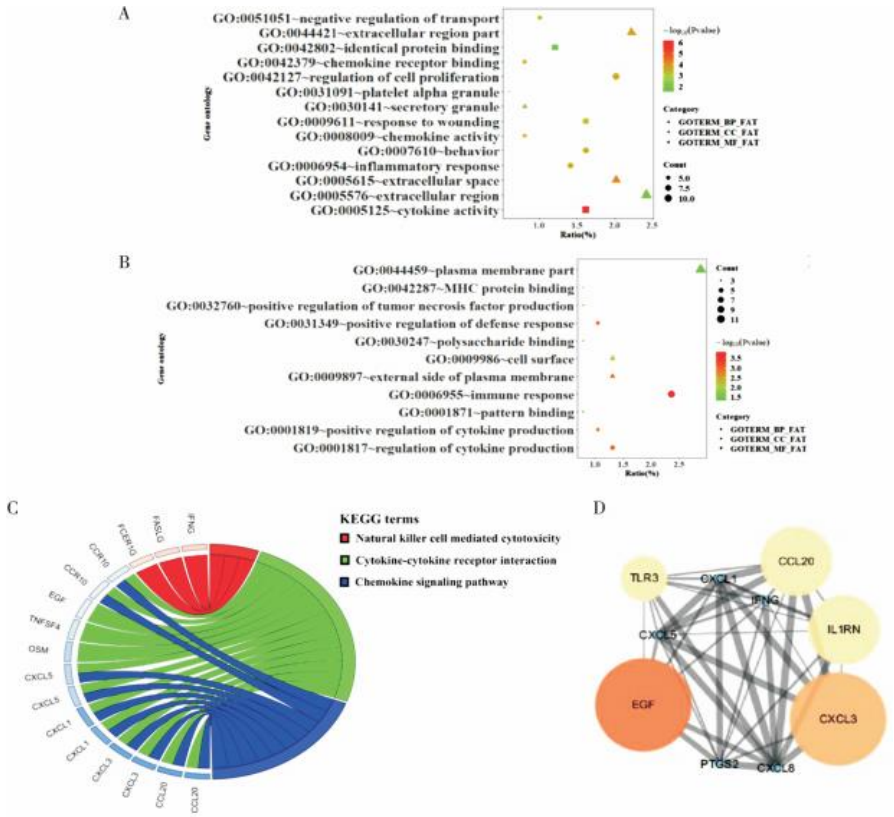

Figure 1 Perform GO functional annotation and KEGG pathway enrichment analysis on DEGs

Note: A and B: GO analysis of down-regulated and up-regulated DEGs respectively. The redder the color, the smaller the P, the bluer the color, the larger the P. The abscissa represents the number of genes, and the ordinate represents BP, CC and MF; C: DEGs KEGG analysis; D. Protein-protein interaction modules in DEGs. The color of the node represents the degree of connectivity. The greater the degree of connectivity, the more connections the node has; the degree of connectivity from large to small is orange, yellow and blue; the thickness of the line represents the comprehensive score, the higher the score, the thicker the line.

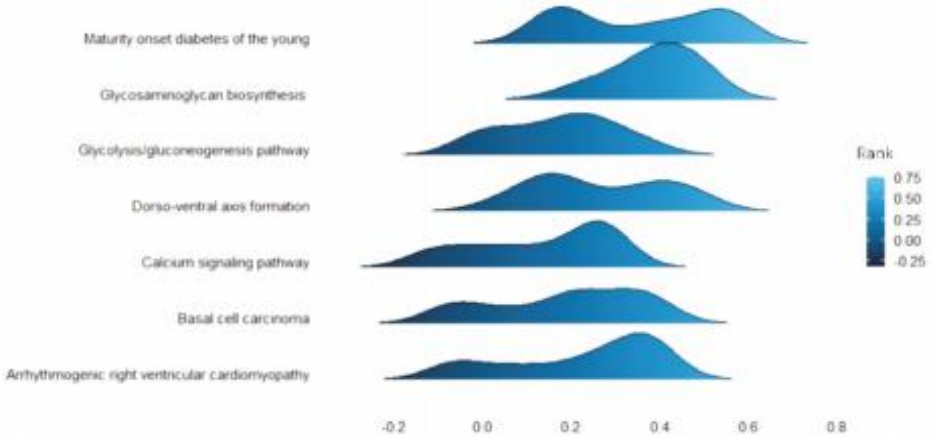

Figure 2 GSEA enrichment analysis of MDD-related pathways

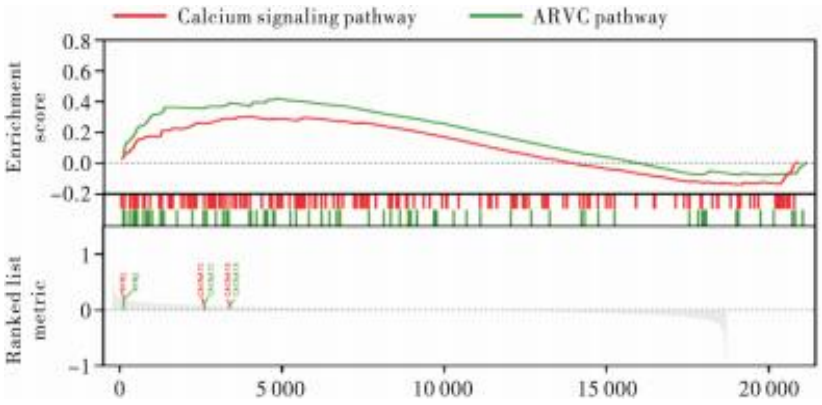

Figure 3 GSEA enrichment analysis of MDD-related genes

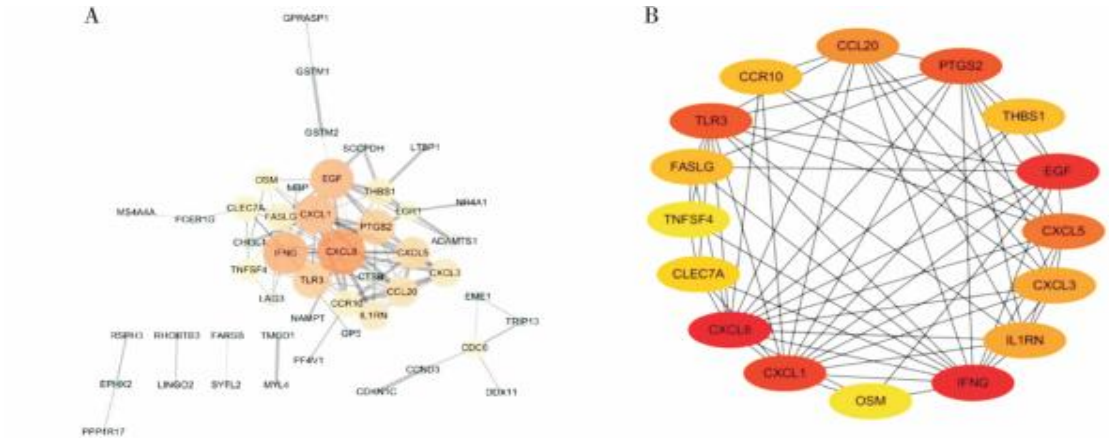

Figure 4 PPI network analysis of DEGs

Note: A: PPI network analysis diagram. The size of the node represents the clustering coefficient. The larger the node, the greater the clustering coefficient, indicating that the gene plays a greater role in the network; the color of the node represents the degree of connectivity. The greater the degree of connectivity, the more connections the node has; from Connectivity from large to small is orange, yellow and blue; the thickness of the line represents the comprehensive score, the higher the score, the thicker the line; the color of the line represents co-expression, and the same color represents the interaction between two proteins. B: Interaction network diagram of Hub genes. The darker the dot, the more important it is

3 DISCUSS

Depression has become a global mental health problem[14]. The prevalence of MDD has been increasing in recent years, but there are still no laboratory blood tests to support early diagnosis of MDD [15]. The rapid development and widespread application of microarray gene chip technology has revealed thousands of gene expression changes in the pathophysiological processes of diseases. Bioinformatics combined with microarray gene chips can systematically conduct a comprehensive study of genes with expression changes in diseases.

Analysis can screen out important biomarkers for early diagnosis of diseases [16]. Therefore, we hope to analyze the mRNA expression profile of MDD with the help of bioinformatics analysis and microarray gene chip results to screen out MDD-related diagnostic biomarkers.

In this study, we first downloaded the expression profile data set GSE32280 from the GEO database and used R software to screen out 104 DEGs in this data set. GO functional annotation analysis of 104 DEGs found that DEGs are mainly enriched in biological processes such as cell proliferation, inflammatory response, and transport regulation; they are involved in molecular functions such as cytokine activity, chemokine activity, and chemokine receptor binding; DEGs The cellular components are mainly concentrated in platelet granules and secretory granules. DEGs mainly mediate NK cell-related cytotoxicity, interactions between cytokines and their receptors, and chemokine signaling pathways. Secondly, we used GSEA to identify genes and pathways related to DEGs and MDD. The enriched pathways were mainly related to the calcium signaling pathway and arrhythmic right ventricular cardiomyopathy (arrthy thmia). right ventricular cardiomyopathy, ARVC) pathway related. The intracellular calcium channel Ryanodine receptor (RYR) and calcium voltage-gated channel subunit alpha1c (CACNA1C) are important genes in these pathways.

RYR is located in the axons, dendritic spines and presynaptic terminals of neurons, and is highly expressed in the cerebellum, hippocampus, olfactory area, basal ganglia and cerebral cortex. RYR is responsible for mediating the release of Ca^{2+} from the intracellular calcium pool [17] and constitutes the intracellular calcium release channel on the endoplasmic reticulum membrane and sarcoplasmic reticulum membrane. Ca^{2+} entering the cell through the cell membrane Able to activate Ryanodine The receptor directly triggers the release of Ca^{2+} from intracellular calcium stores. During the physiological processes of development and aging, intracellular calcium homeostasis is mainly regulated by RYR. Abnormal expression of RYR can cause intracellular Ca^{2+} level imbalance, cell fragility, and damage to synaptic neuron function, leading to neuronal death, which may ultimately lead to depression [18].

CACNA1C encodes the α1C subunit of the L-type voltage-dependent Ca^{2+} channel, and the α1C subunit is the main subunit of Cav1.2. Cav1.2 is an important pathway that mediates Ca^{2+} entry into cells and plays an important role in dendritic development, neuron survival, synaptic plasticity, memory formation, learning and behavior [19]. Studies have shown that CACNA1C has become a candidate risk gene for neuropsychiatric diseases such as bipolar disorder, major depression, and schizophrenia [20]. CACNA1C heterozygous deletion mice have reduced Cav1.2 protein levels, LTCC currents, reduced exploratory behavior, reduced responses to amphetamines, and antidepressant-like behaviors in forced swimming and tail suspension tests [21]. After inhibiting the expression of CACNA1C in the prefrontal cortex of mice, the mice developed significant antidepressant-like behaviors [22]. The data in this study show that CACNA1C and RYR are likely to have significantly abnormal expression in the peripheral blood of patients with depression. Therefore, we speculate that CACNA1C and RYR may be involved in the pathophysiological mechanism of MDD.

This study evaluated the diagnostic value of hub genes and found that four genes (CXCL1, EGF, CXCL8 and IFNG) are closely related to the pathophysiological mechanism of MDD and may become diagnostic biomarkers for MDD. Among them, CXCL1 and CXCL8 are both chemokines. Chemokines are cytokines secreted by a variety of cells that

can induce the chemotactic properties of leukocytes. In addition to chemotactic and activating leukocytes, chemokines also have various biological activities such as stimulating cell proliferation and promoting the formation of new blood vessels, and play an important role in the pathogenesis of inflammation [23]. Studies have shown that CXCL1 is significantly elevated in the cerebrospinal fluid of chronic unpredictable depression model [24]. The expression of CXCL8 is significantly increased in the peripheral blood of patients with depression [25]. Epidermal growth factor (EGF) is an important cell growth factor. EGF levels are increased in the peripheral blood of patients with depression, and antidepressant treatment significantly reduces EGF levels [26]. From this, we speculate that CXCL1 and CXCL8 and EGF may be involved in the pathophysiological process of MDD and may become potential diagnostic biomarkers for MDD.

This study uses bioinformatics methods to analyze MDD of The mRNA expression profile data set was analyzed and a total of 104 DEGs were obtained. and 16 hub genes. Further ROC analysis was performed on the hub gene and 4 The diagnostic markers are CXCL1, EGF, CXCL8 and IFNG. The results of this study provide a theoretical basis for the development of new MDD diagnostic biomarkers. Based on the results of this study, we will further study the pathophysiological mechanisms of MDD at the cellular and molecular levels.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Charlson FJ, Ferrari AJ, Flaxman AD. The epidemiological modelling of dysthymia: application for the Global Burden of Disease Study. 2010. J AffectDis- ord, 2013, 151(1): 111-120.
[2] Hawton K, Casanas I, Comabella C. Risk factors for suicide in individuals with depression: a systematic review. J Affect Disord, 2013, 147(1-3): 17-28.
[3] Fried EI, Epskamp S, Nesse R M. What are “good” depression symptoms? Comparing the centrality of DSM and non-DSM symptoms of depression in a network analysis. J Affect Disord, 2016, 189: 314-320.
[4] Pettersson A, Bostrom K B, Gustavsson P. Which instruments to support diagnosis of depression have sufficient accuracy? A systematic review. Nord J Psychiatry, 2015, 69(7): 497-508.
[5] Takahashi M, Lim PJ, Tsubosaka M. Effects of increased daily physical activity on mental healthand depression biomarkers in postmenopausal women. J Phys Ther Sci, 2019, 31(4): 408-413.
[6] Serati M, Redaelli M, Buoli M. Perinatal major depression biomarkers: a systematic review. J Affect Disord, 2016, 193: 391-404.
[7] Gururajan A, Clarke G, Dinan T G. Molecular biomarkers of depression. Neurosci Biobehavior Rev., 2016, 64: 101-133.
[8] Ferrúa CP, Giorgi R, daRosaL C. MicroRNAs expressed in depression and their associated pathways: a systematic review and a bioinformatics analysis. J Chem Neuroanat, 2019, 100: 101650.
[9] Barrett T, Wilhite SE, Ledoux P. NCBIGEO: archive for functional genomics data sets--update. NucleicAcids Res, 2013, 41(Database issue): D991- D995.
[10] Ginestet C. ggplot2: elegant graphics for data analysis. JR Stat Soca Stat, 2011, 174(1): 245.
[11] YuG, Wang LG, Han Y. Cluster Profiler: anR package for comparing biological themes among gene clusters. OMICS, 2012, 16(5): 284-287.
[12] Szklarczyk D, Gable AL, Lyon D. STRINGv11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res, 2019, 47 (D1): D607-D613.
[13] Robin X, Turck N, Hainard A. pROC: an open-source package for Rand S+ to analyze and compare ROC curves. BMC Bioinformatics, 2011, 12: 77.
[14] Rotenstein L S, Ramos M A, Torre M. Prevalence of depression, depressive symptoms, and suicidal ideation among medical students: a systematic review andmeta-analysis. JAMA, 2016, 316(21): 2214- 2236.
[15] Smith K. Mental health: a world of depression. Nature, 2014, 515(7526): 181.
[16] Hyman S. Mental health: depression needs large human-genetics studies. Nature, 2014, 515(7526): 189-191.
[17] Gyorke S, Terentyev D. Modulation of ryanodine receptor by luminalcal cium and accessory proteins in health and cardiac disease. Cardiovasc Res, 2008, 77(2): 245-255.
[18] Abu-Omar N, Das J, Szeto V. Neuronal ryanodine receptors in development and aging. Mol Neurobiol, 2018, 55(2): 1183-1192.
[19] Hofmann F, Flockerzi V, Kahl S. L-typeCaV1.2 calcium channels: frominvitro findings to in vivo function. Physiol Rev, 2014, 94(1): 303-326.
[20] Moon AL, Haan N, Wilkinson LS. CACNA1C: association with psychiatric disorders, behavior, and neurogenesis. Schizophr Bull, 2018, 44(5):958-965.
[21] Dao DT, Mahon PB, Cai X. Mooddis order susceptibility gene CACNA1C modifies mood-related behaviors in mice and interacts with sex to influence behavior in mice and diagnosis in humans. BiolPsychia- try, 2010, 68(9): 801-810.

[22] Kabir ZD, Lee AS, Burgdorf CE. Cacna1cinthe prefrontalcortexregulatesdepression-relatedbehaviors viaREDD1. Neuropsycho pharmacology, 2017, 42 (10): 2032-2042.
[23] BorrelliG M, Abrao MS, Mechsner S. Can chemokines be used as biomarkers for endometriosis? Asystematic review. Hum Reprod, 2014, 29(2): 253-266.
[24] Chai H H, Fu XC, Ma L. ThechemokineCXCL1 anditsreceptor CXCR2 contribute to chronic stress-induced depressionin mice. FASEBJ, 2019, 33(8): 8853-8864.
[25] Eyre H A, Air T, Pradhan A. A meta-analysis of chemokines in major depression. Prog Neuropsycho- pharmacol Biol Psychiatry, 2016, 68: 1-8.
[26] Memon A A, Sundquist K, Ahmad A. RoleofIL-8, CRP and epidermal growth factor indepression and anxiety patients treated with mindfulness-based therapy or cognitive behavioral therapy in primary health care. Psychiatry Res, 2017, 254: 311-316.