

慢性儿童免疫性血小板减少症差异表达基因的生物信息学分析

夏悦昕¹ 王霓¹ 张力¹ 宋文倩¹ 邵林楠¹ 戚凝¹ 于卫建¹

[摘要] 目的:探究慢性儿童免疫性血小板减少症(ITP)的关键基因和信号通路,为研究慢性 ITP 的潜在分子机制提供新的思路。方法:从 GEO 数据库获得 mRNA 表达芯片数据集 GSE46922,利用 GEO2R 筛选出 DEGs,并利用 DAVID 对 DEGs 进行 GO 功能富集分析和 KEGG 信号转导通路富集分析。随后,构建 PPI 蛋白互作网络,筛选核心靶标。结果:筛选出 274 个表达上调基因,主要参与细胞骨架组织,基于肌动蛋白丝的过程,肌动蛋白纤维组织等生物过程。筛选出 496 个表达下调基因,主要参与动作电位和肌肉收缩等过程。KEGG 分析发现,上调基因主要参与 HIF-1 信号通路。下调基因主要涉及焦点黏附、Rap1 信号通路、PI3K-Akt 信号通路、神经信号传递通路、肌动蛋白细胞骨架调控等信号通路。获得 11 个核心靶标:VEGFA,CCND1,ESR1,MYH14,ERBB2,CDKN3,PVALB,CDC20,CHEK1,MTOR 和 CFTR。结论:本研究通过生物信息学分析获得慢性 ITP 的关键基因和信号通路,为研究慢性 ITP 的发病机制提供了新的思路。核心靶标基因可作为诊断和预后评估的生物标志物。

[关键词] 生物信息学分析;慢性儿童免疫性血小板减少症;差异表达基因;血管内皮生长因子 A

doi:10.13201/j.issn.1004-2806-b.2019.10.006

[中图分类号] R457.1 **[文献标志码]** A

Identification of key genes and pathways associated with chronic pediatric ITP by bioinformatics analysis

XIA Yuexin WANG Ni ZHANG Li SONG Wenqian

SHAO Linnan QI Ning YU Weijian

(Dalian Blood Center, Dalian, 116000, China)

Corresponding author: YU Weijian, E-mail: yuweijiandl@163.com

Abstract Objective: To identify critical genes and pathways in order to unravel the molecular mechanisms associated with chronic ITP. **Method:** The gene expression profile dataset (GSE46922) was downloaded from GEO database. The DEGs were identified by GEO2R. GO and KEGG pathway enrichment analyses were carried out using DAVID database. Furthermore, PPI network was constructed by STRING database. **Result:** Totally 274 up-regulated genes were identified, which were enriched in biological processes of cytoskeleton organization, actin filament-based process and actin filament organization. Additionally, 496 down-regulated genes were identified, which demonstrated enrichment in biological processes including action potential, muscle contraction and behavior. Pathway analysis revealed that the up-regulated DEGs were particularly involved in HIF-1 signaling pathway, while the down-regulated DEGs were mainly involved in focal adhesion, Rap1 signaling pathway and PI3K-Akt signaling pathway. The top 11 hub genes of PPI network were VEGFA, CCND1, ESR1, MYH14, ERBB2, CDKN3, PVALB, CDC20, CHEK1, MTOR and CFTR. **Conclusion:** The present study identified the DEGs and pathways in chronic ITP by bioinformatics analysis, which might provide novel insights for unraveling pathogenesis of chronic ITP. The hub genes might serve as biomarkers for diagnosis and prognosis evaluation.

Key words bioinformatics analysis; chronic pediatric immune thrombocytopenia; differentially expressed genes; VEGFA

儿童免疫性血小板减少症(pediatric immune thrombocytopenia, ITP)是一种自身免疫系统介导的血液病,其特点是血小板计数低、自发性瘀伤、紫癜性皮疹和黏膜出血^[1],发病率为 4~5/10 万。根据血小板减少症的持续时间,将 ITP 分为 3 个阶段,即新诊断 ITP(病程<3 个月)、持续性 ITP(病

程 3~12 个月)和慢性 ITP(病程>12 个月)^[2]。儿童新诊断 ITP 和持续性 ITP 通常具有自限性^[3],约 20% 的患儿会发展成为慢性 ITP。由于长期的血小板减少,慢性 ITP 通常需要长期利用免疫抑制剂进行治疗,甚至在难治性病例中进行脾切除术。因此,在新诊断的 ITP 病例中,对其远期病程进行准确的预测有着极为重要的临床意义。

之前的研究发现,相对于自限性 ITP,慢性 ITP 的患儿年龄更大,血小板计数更高(>20 ×

¹大连市血液中心(辽宁大连,116000)

通信作者:于卫建, E-mail: yuweijiandl@163.com

$10^9/L$),黏膜出血较少,并且在诊断前发生病毒感染的可能性也较小^[1]。最近的研究发现,与自限性 ITP 相比,慢性 ITP 患儿血浆样本中白细胞介素 2(interleukin 2, IL-2)、白细胞介素 10(interleukin 10, IL-10)和干扰素 γ (interferon- γ)的表达水平较高^[4-5]。Zhang 等^[6]研究发现,与自限性 ITP 相比,慢性 ITP 患儿全血中 vanin-1 mRNA 表达水平较高。目前为止参与慢性 ITP 的关键调控基因和潜在分子机制尚未清晰,并且没有生物标志物能够在诊断早期准确地预测自限性 ITP 发展为慢性 ITP 的可能性。

基因芯片是一种高通量的基因表达分析平台,目前已得到广泛应用^[7]。本研究从 Gene Expression Omnibus(GEO)数据库获得 mRNA 表达芯片数据集 GSE46922,将慢性 ITP 患儿 T 细胞样本的表达谱与新诊断 ITP 患儿的表达谱进行比较,筛选出差异表达基因(differentially expressed genes, DEGs),并对 DEGs 进行 Gene ontology(GO)功能富集分析和 Kyoto Encyclopedia of Genes and Genomes(KEGG)信号转导通路富集分析。随后,通过蛋白-蛋白相互作用网络(protein-protein interaction, PPI)的构建,获得慢性 ITP 的核心靶标基因。本研究旨在利用生物信息学方法初步探究慢性 ITP 的关键基因和途径,为研究慢性 ITP 的发病机制提供了新的思路,有助于探索新的生物标志物以进行慢性 ITP 的诊断和预后评估。

1 资料与方法

1.1 研究对象

从 GEO 数据库(<http://www.ncbi.nlm.nih.gov/geo/>)获得 mRNA 表达芯片数据集 GSE46922。该数据集由 Jernås 等^[8]提交,包括 7 例新诊断 ITP 患儿 T 细胞样本和 6 例慢性 ITP 患儿 T 细胞样本。

1.2 数据来源

利用 GEO2R(<http://www.ncbi.nlm.nih.gov/geo/geo2r/>)筛选 DEGs;利用 DAVID 数据库(<https://david.ncifcrf.gov/>)对 DEGs 进行 Gene ontology-Biological process(GO-BP)功能富集分析和 KEGG 信号转导通路富集分析;利用 STRING(<http://www.string-db.org/>)数据库和可视化分析软件 Cytoscape 构建差异表达基因的 PPI 网络。

1.3 方法

1.3.1 DEGs 的筛选 利用在线工具 GEO2R(<http://www.ncbi.nlm.nih.gov/geo/geo2r/>)筛选 DEGs,筛选条件: $P < 0.05$, $|\log FC| \geq 2$ ^[9-11]。

1.3.2 GO 及 KEGG 分析 利用 DAVID 数据库(<https://david.ncifcrf.gov/>)对靶基因进行 GO 功能富集分析和 KEGG 信号转导通路富集分

析^[12-13]。筛选条件: $P < 0.05$ 。

1.3.3 PPI 网络构建 首先从 STRING(<http://www.string-db.org/>)数据库获得差异基因相互作用的数据^[14],筛选条件:confident interaction score > 0.4 。利用可视化分析软件 Cytoscape 获得差异基因 PPI 互作网络^[15]。节点连接边数 degrees > 20 的基因称为核心靶标。利用插件 Molecular Complex Detection(MCODE)进行模块分析,筛选条件:degree cut-off = 2, haircut on, node score cut-off = 0.2, k-core = 2, and max. depth = 100, MCODE scores ≥ 3.96 , hub nodes ≥ 6.83 and edges ≥ 14.83 , Pvalue < 0.05 ^[16]。

2 结果

2.1 DEGs 的筛选

从数据集 GSE46922 中筛选出 770 个在儿童慢性与新诊断 ITP T 细胞样本中差异表达的基因,其中 274 个基因表达上调,496 个基因表达下调。

2.2 GO 及 KEGG 分析

GO 生物学过程分析发现上调的差异基因参与多种生物过程,排名前 5 的为:细胞骨架结构、基于肌动蛋白丝的过程、肌动蛋白纤维组织、细胞黏连负调控、上皮细胞迁移负调控(表 1)。下调基因主要参与动作电位和肌肉收缩等过程。KEGG 分析发现,上调基因主要参与 HIF-1 信号通路(表 2)。下调基因主要涉及焦点黏附、Rap1 信号通路、PI3K-Akt 信号通路、神经信号传递通路、肌动蛋白细胞骨架调控信号通路。

表 1 差异表达基因 GO-BP 分析

GO 编号	GO-BP 注释	基因/个数(%)	P
上调基因			
GO:0007010	细胞骨架结构	31(11.40)	9.80E-05
GO:0030029	基于肌动蛋白丝的过程	22(8.09)	1.15E-04
GO:0007015	肌动蛋白纤维组织	14(5.15)	3.26E-04
GO:0007162	细胞黏连负调控	11(4.04)	4.55E-04
GO:0010633	上皮细胞迁移负调控	6(2.21)	5.70E-04
下调基因			
GO:0001508	动作电位	15(3.02)	3.87E-06
GO:0006936	肌肉收缩	23(4.64)	1.98E-05
GO:0007610	behavior	31(6.25)	3.90E-05
GO:0045595	调节细胞分化	63(12.70)	4.86E-05
GO:0019228	神经元动作电位	7(1.41)	5.74E-05

2.3 PPI 网络构建

PPI 网络分析发现 11 个核心靶标:血管内皮生长因子 A(vascular endothelial growth factor A, VEGFA),细胞周期蛋白 D1(cyclin D1, CCND1),雌激素受体 1(estrogen receptor 1, ESR1),肌凝蛋

白重链 14(myosin, heavy chain 14, MYH14), erb-b2 受体酪氨酸激酶 2(erb-b2 receptor tyrosine kinase 2, ERBB2), 细胞周期蛋白依赖性激酶抑制剂 3(cyclin dependent kinase inhibitor 3, CDKN3), 小白蛋白(parvalbumin, PVALB), 细胞分裂周期 20(cell division cycle 20, CDC20), 细胞周期检测点激酶 1(checkpoint kinase 1, CHEK1), 雷帕霉素机制性靶标(mechanistic target of rapamycin, MTOR), 囊性纤维化跨膜传导调节蛋白(cystic fibrosis

transmembrane conductance regulator, CFTR)(图 1)。其中 ESR1、ERBB2 和 MTOR 表达上调, 其余基因表达下调。模块分析获得模块 1 和 2(图 2), KEGG 分析发现, 模块 1 中的基因主要参与刺激神经组织的配体受体相互作用通路, 细胞周期和钙离子信号通道(表 3)。模块 2 中的基因主要参与焦点黏附, PI3K-Akt 信号通路, 血小板激活和细胞外基质受体相互作用等信号通路。

表 2 差异表达基因 KEGG 通路路径列表

KEGG-ID	通路	%	P	基因
下调基因				
hsa04510	焦点黏连	2.62	1.76E-03	MYL7, CAV2, CCND1, COL3A1, PAK5, VEGFA, ITGB4, ITGB5, ITGA2, PDGFC, ITGB3, COL11A1, PARVB, RAP1GAP, GRIN2B, FGF14, CNR1, VEGFA, P2RY1, RAPGEF5, PDGFC, FGF12, PARD6G, ITGB3
hsa04015	Rap1 信号通路	2.22	1.64E-02	FGF14, COL3A1, ITGB4, ITGA2, ITGB5, FGF12, ITGB3, RPS6, IFNA1, CCND1, IFNA5, VEGFA, PDGFC, COL11A1
hsa04151	PI3K-Akt 信号通路	2.82	3.82E-02	GABRA2, GRIN2B, CCKBR, CHRM3, THRB, GRIK2, NMUR2, CNR1, P2RY1, ADRA1A, NPY1R, CHRNA1
hsa04080	刺激神经组织的配体受体相互作用通路	2.42	3.97E-02	MYL7, CHRM3, FGF14, PAK5, ITGB4, ITGB5, ITGA2, PDGFC, FGF12, ITGB3
hsa04810	肌动蛋白细胞骨架调节	2.02	4.12E-02	
上调基因				
hsa04066	HIF-1 信号通路	1.21	3.10E-02	ERBB2, IL6R, MTOR, ANGPT2, AKT2

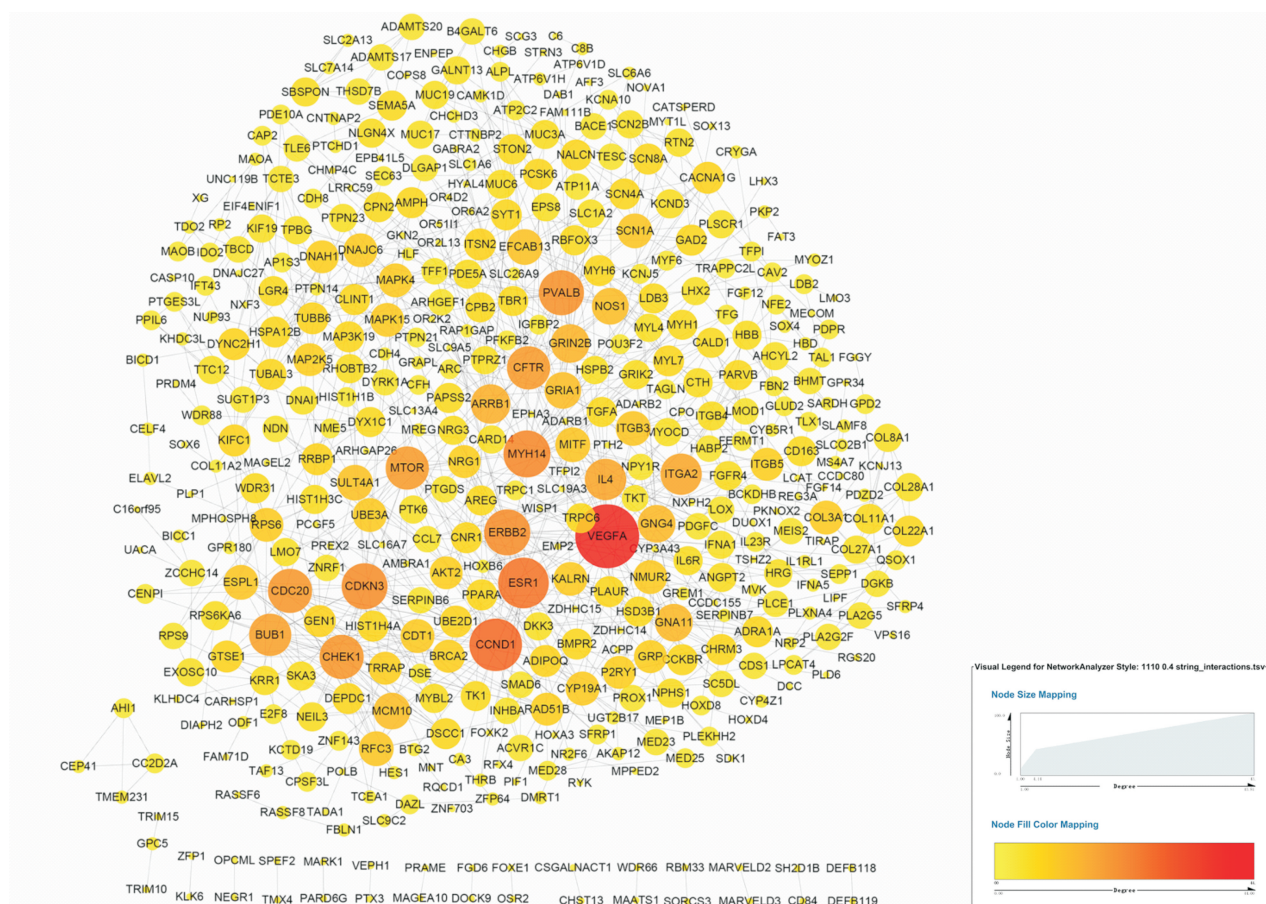
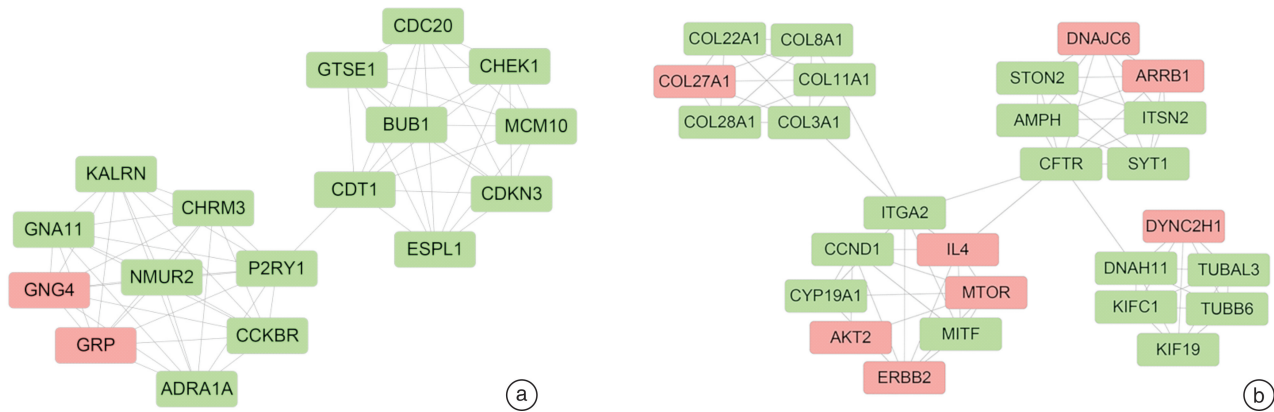


图 1 差异表达基因的 PPI 网络



a: 模块 1; b: 模块 2。红色框代表上调基因; 绿色框代表下调基因。

图 2 差异表达基因的模块分析

表 3 模块 1 和模块 2 差异表达基因 KEGG 通路列表

KEGG-ID	通路	%	P	基因
A, module 1				
hsa04080	刺激神经组织的配体受体相互作用通路	29.4	6.67E-04	CCKBR, CHRM3, NMUR2, P2RY1, ADRA1A
hsa04110	细胞周期	23.5	8.38E-04	BUB1, CHEK1, CDC20, ESPL1
hsa04020	钙信号通路	23.5	2.42E-03	CCKBR, CHRM3, GNA11, ADRA1A
hsa04114	减数分裂	17.6	1.24E-02	BUB1, CDC20, ESPL1
hsa04725	胆碱能突触	17.6	1.28E-02	CHRM3, GNA11, GNG4
B, module 2				
hsa04510	焦点黏连	25.9	1.78E-05	CCND1, COL27A1, ERBB2, COL3A1, ITGA2, COL11A1, AKT2
hsa04151	PI3K-Akt 信号通路	29.6	3.21E-05	IL4, CCND1, COL27A1, COL3A1, ITGA2, MTOR, COL11A1, AKT2
hsa04611	血小板激活	18.5	4.59E-04	COL27A1, COL3A1, ITGA2, COL11A1, AKT2
hsa04512	细胞外基质受体相互作用	14.8	1.88E-03	COL27A1, COL3A1, ITGA2, COL11A1
hsa04974	蛋白质的消化和吸收	14.8	1.95E-03	COL27A1, COL3A1, COL22A1, COL11A1

3 讨论

阐明慢性 ITP 的关键基因和潜在分子机制对慢性 ITP 的诊断和预后评估具有重要意义。VEGFA 是连接度最高的核心靶标,相对于新诊断 ITP,其在慢性 ITP 患儿 T 细胞样本中的表达水平下调。其作为血管内皮生长因子(VEGF)超级家族的成员,编码血管内皮生长因子 A^[17-18]。VEGFA 在血管内皮细胞的增殖、分化和迁移过程中起着至关重要的作用,并参与血管通透性的维持^[19-20]。VEGFA 的异常表达可以破坏血管内皮屏障^[21]。VEGFA 的表达受 PI3K-Akt^[22] 和 HIF-1^[23] 等调控,此外,活化的血小板可以启动 VEGFA 的释放^[24]。VEGFA 能够诱导焦点黏附激酶(focal adhesion kinase, FAK)酪氨酸磷酸化,进而在内皮细胞中召集焦点黏附^[25]。焦点黏附是指内皮细胞与细胞外基质的特定定位点结合^[26]。大量证据表明,焦点黏附在血管内皮细胞的增殖、分化和迁移

中起着重要的作用。我们由此推测,相对于新诊断 ITP,慢性 ITP 患儿 T 细胞样本中 VEGFA 表达水平的下调影响了血管内皮细胞的增殖、分化和迁移,以及血管通透性。最近的研究表明,VEGFA 在多种肿瘤组织中表达上调,其表达与肿瘤分期和进展密切相关^[27]。然而迄今为止,关于 VEGFA 在慢性 ITP 中功能的研究尚未见报道。

本研究是第一篇关于儿童新诊断 ITP 和慢性 ITP 差异基因和关键信号通路的生物信息学分析,为研究慢性 ITP 的发病机制提供了新的思路,有助于探索新的生物标志物以进行慢性 ITP 的诊断和预后评估。包括 VEGFA、CCND1、ESR1、MYH14 和 ERBB2 等在内的 DEGs 可能成为慢性 ITP 诊断和预后评估的生物标志物。而焦点黏附,PI3K-Akt 信号通路,血小板激活和细胞外基质受体相互作用等通路可能在慢性 ITP 中起重要作用。然而,目前缺乏实验验证,因此需要进一步的分子生物学

实验来验证 VEGFA、CCND1、ESR1、MYH14 和 ERBB2 等差异基因在慢性 ITP 中的功能。

参考文献

- [1] Glanz J, France E, Xu S, et al. A population-based, multisite cohort study of the predictors of chronic idiopathic thrombocytopenic purpura in children[J]. Pediatrics, 2008, 121: 506—512.
- [2] Rodeghiero F, Stasi R, Gernsheimer T, et al. Standardization of terminology, definitions and outcome criteria in immune thrombocytopenic purpura of adults and children: report from an international working group[J]. Blood, 2009, 113: 2386—2393.
- [3] Nugent DJ. Immune thrombocytopenic purpura of childhood[J]. Hematology Am Soc Hematol Educ Program, 2006: 97—103.
- [4] Semple JW, Milev Y, Cosgrave D, et al. Differences in serum cytokine levels in acute and chronic autoimmune thrombocytopenic purpura: relationship to platelet phenotype and antiplatelet T-cell reactivity[J]. Blood, 1996, 87: 4245—4254.
- [5] Del Vecchio GC, Giordano P, Tesse R, et al. Clinical significance of serum cytokine levels and thrombopoietic markers in childhood idiopathic thrombocytopenic purpura[J]. Blood Transfus, 2012, 10: 194—199.
- [6] Zhang B, Lo C, Shen L, et al. The role of vanin-1 and oxidative stress-related pathways in distinguishing acute and chronic pediatric ITP[J]. Blood, 2011, 117: 4569—4579.
- [7] Kuang Z, Guo L, Li X. Identification of key genes and pathways associated with classical Hodgkin lymphoma by bioinformatics analysis[J]. Mol Med Rep, 2017, 16: 4685—4693.
- [8] Jernäs M, Hou Y, Stromberg Celind F, et al. Differences in gene expression and cytokine levels between newly diagnosed and chronic pediatric ITP[J]. Blood, 2013, 122: 1789—1792.
- [9] Liu Z, Liang G, Tan L, et al. High-efficient Screening Method for Identification of Key Genes in Breast Cancer Through Microarray and Bioinformatics[J]. Anticancer Res, 2017, 37: 4329—4335.
- [10] Xu Z, Zhou Y, Cao Y, et al. Identification of candidate biomarkers and analysis of prognostic values in ovarian cancer by integrated bioinformatics analysis[J]. Med Oncol, 2016, 33: 130.
- [11] Liu J, Li H, Sun L, et al. Aberrantly methylated-differentially expressed genes and pathways in colorectal cancer[J]. Cancer Cell Int, 2017, 17: 75.
- [12] Liang B, Li C, Zhao J. Identification of key pathways and genes in colorectal cancer using bioinformatics analysis[J]. Med Oncol, 2016, 33: 111.
- [13] Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes[J]. Nucleic Acids Res, 2000, 28: 27—30.
- [14] Szklarczyk D, Franceschini A, Wyder S, et al. STRING v10: Protein-protein interaction networks, integrated over the tree of life[J]. Nucleic Acids Res, 2015, D447—D452.
- [15] Shannon P, Markiel A, Ozier O, et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks[J]. Genome Res, 2003, 2498—2504.
- [16] Liang B, Li C, Zhao J. Identification of key pathways and genes in colorectal cancer using bioinformatics analysis[J]. Med Oncol, 2016, 111.
- [17] Clauss M, Gerlach M, Gerlach H, et al. Vascular permeability factor: a tumor-derived polypeptide that induces endothelial cell and monocyte procoagulant activity, and promotes monocyte migration[J]. J Exp Med, 1990, 172: 1535—1545.
- [18] Wartiovaara U, Salven P, Mikkola H, et al. Peripheral blood platelets express VEGF-C and VEGF which are released during platelet activation[J]. Thromb Haemost, 1998, 80: 171—175.
- [19] Baeriswyl V, Christofori G. The angiogenic switch in carcinogenesis[J]. Semin Cancer Biol, 2009, 19: 329—337.
- [20] Chen HX, Xu XX, Tan BZ, et al. MicroRNA-29b Inhibits Angiogenesis by Targeting VEGFA through the MAPK/ERK and PI3K/Akt Signaling Pathways in Endometrial Carcinoma[J]. Cell Physiol Biochem, 2017, 41: 933—946.
- [21] Zhang RY, Liu YY, Qu HP, et al. The angiogenic factors and their soluble receptors in sepsis: friend, foe, or both? [J]. Crit Care, 2013, 17: 446.
- [22] Gerber HP, McMurtrey A, Kowalski J, et al. Vascular endothelial growth factor regulates endothelial cell survival through the phosphatidylinositol 3'-kinase/Akt signal transduction pathway. Requirement for Flk-1/KDR activation[J]. J Biol Chem, 1998, 273: 30336—30343.
- [23] Gerber HP, Condorelli F, Park J, et al. Differential transcriptional regulation of the two VEGF receptor genes. Flt-1, but not Flk-1/KDR, is up-regulated by hypoxia[J]. J Biol Chem, 1997, 272: 23659—23667.
- [24] Belair DG, Le NN, Murphy WL. Regulating VEGF signaling in platelet concentrates via specific VEGF sequestering[J]. Biomater Sci, 2016, 4: 819—825.
- [25] Mehta D, Malik AB. Regulating endothelial permeability[J]. Physiol Rev, 2006, 86: 279—367.
- [26] Mehta D. Focal adhesion kinase regulation of endothelial barrier function, apoptosis, and neovascularization[J]. Microvasc Res, 2012, 83: 1—2.
- [27] Naikoo NA, Dil-Afroze, Rasool R, et al. Upregulation of vascular endothelial growth factor(VEGF), its role in progression and prognosis of non-small cell lung carcinoma[J]. Cancer Genet, 2017, 216: 67—73.

(收稿日期: 2019-03-15)