

基于GEO数据库对高风险胃肠道间质瘤的 差异基因的生物信息学分析

邱小凤¹, 黄义莲¹, 赵文睫², 李娜², 孙焯煜¹, 贺延新^{2*}

¹青岛大学医学部, 山东 青岛

²青岛大学附属青岛市中心医院, 青岛市肿瘤医院消化内科, 山东 青岛

收稿日期: 2023年3月19日; 录用日期: 2023年4月15日; 发布日期: 2023年4月21日

摘要

目的: 采用生物信息学的方法探讨影响胃肠道间质瘤风险分级的差异基因表达及其生物学功能富集。方法: 检索并获取GEO数据库中关于胃肠道间质瘤恶性程度差异基因表达的数据, 利用GER20工具对差异基因进行筛选和分析, 采用DAVID数据库进行GO、KEGG富集分析, 使用STRING, Cytoscape软件行蛋白-蛋白互作(PPI)网络构建。结果: 筛选出两个数据集: GSE31802和GSE47911, 查找出其中共有的差异表达基因有42个, 它们的生物学过程可能主要富集在cAMP分解代谢过程、信号转导、基因表达的正向调控和酪氨酸激酶信号通路这几个方面。细胞学组分主要定位于质膜以及内质网腔。分子功能富集于信号素受体活性、序列特异性双链DNA结合。KEGG信号通路的富集分析显示差异基因主要富集于轴突的指导、癌症的转录失调等。对蛋白质相互作用分析中有关联的基因进行排序, 显示IL-6, PGR的节点分数最高。结论: 通过对影响胃肠道间质瘤风险分级的差异基因的分析, 本研究发现的关键差异基因有助于了解影响胃肠道间质瘤危险分级的潜在因子。

关键词

胃肠道间质瘤, 危险度分级, 生物信息学分析

Bioinformatics Analysis of Differential Genes in High-Risk Gastrointestinal Stromal Tumors Based on GEO Database

Xiaofeng Qiu¹, Yilian Huang¹, Wenjie Zhao², Na Li², Yeyu Sun¹, Yanxin He^{2*}

¹Medical College of Qingdao University, Qingdao Shandong

²Affiliated Qingdao Central Hospital of Qingdao University, Department of Gastroenterology, Qingdao Cancer Hospital, Qingdao Shandong

*通讯作者。

文章引用: 邱小凤, 黄义莲, 赵文睫, 李娜, 孙焯煜, 贺延新. 基于 GEO 数据库对高风险胃肠道间质瘤的差异基因的生物信息学分析[J]. 临床医学进展, 2023, 13(4): 6073-6079. DOI: 10.12677/acm.2023.134856

Abstract

Objective: To explore the differential gene expression and biological function enrichment affecting the risk classification of gastrointestinal stromal tumors by bioinformatics. **Methods:** The data of differential gene expression of malignant degree of gastrointestinal stromal tumors (GSE31802, GSE47911) in GEO database were retrieved and obtained. The differential genes were screened and analyzed by GER20 tool, GO and KEGG enrichment analysis were performed by DAVID database, and protein-protein interaction (PPI) network was constructed by STRING, Cytoscape software. **Results:** A total of 42 differentially expressed genes from the two data sets were screened, and their biological processes may be mainly concentrated in cAMP catabolism, signal transduction, positive regulation of gene expression and tyrosine kinase signal pathway; cellular components are mainly located in plasma membrane and endoplasmic reticulum lumen; molecular function is enriched in signal receptor activity and sequence-specific double-stranded DNA binding. The enrichment analysis of KEGG signal pathway showed that the differential genes were mainly enriched in the guidance of axons and transcriptional disorders in cancer. The sequencing of related genes in protein-protein interaction analysis showed that IL-6, PGR had the highest node score. **Conclusion:** Through the analysis of differential genes related to high-risk gastrointestinal stromal tumors, the key differential genes found in this study are helpful to understand the potential factors affecting the grading of gastrointestinal stromal tumors.

Keywords

Gastrointestinal Stromal Tumor, Risk Classification, Bioinformatics Analysis

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

胃肠道间质瘤(Gastrointestinal stromal tumor, GIST)是胃肠道最常见的恶性上皮病变,它们起源于位于肌层内的 Cajal 间质细胞[1] [2]。胃肠道间质瘤主要由酪氨酸激酶受体 KIT 和/或血小板衍生生长因子受体- α (PDGFR- α)的致癌突变引起,大约 10%到 30%的胃肠道间质瘤有恶性的临床病程[1],大约占有胃肠道恶性肿瘤的 1%~3%,最常见的是发生在胃(50%~60%)和小肠(30%~35%),也可发生于食管、结直肠,有文献报道肠系膜、网膜等也可发生[3] [4],但不能完全排除是转移灶[5]。GIST 具有从良性小病变到致命肉瘤的恶性潜能[6],因此对该疾病进行科学的风险分级具有十分重要的意义。修改后的美国国立卫生研究院(National Institutes of Health, NIH)标准是评估胃肠道间质瘤潜在恶性风险的最常用工具,根据肿瘤的位置、大小、有丝分裂和肿瘤是否破裂,修改后的 NIH 标准将 GIST 分为极低风险、低风险、中风险和高风险[7]。往往中高危 GIST 患者手术切除后肿瘤转移和复发的风险显著增加,一项基于人群队列的综合分析构建了修改后的 NIH 标准的受试者工作曲线,曲线下面积为 0.78, 95%CI 为 0.75~0.80 [8]。但即使同是高危侵袭性患者,其预后情况仍不尽相同,所以单纯采用 NIH 分级标准评估 GIST 的恶性潜能并不完全准确[9]。

随着生物信息技术的发展，微阵列基因表达分析的方法以及生物信息学方法已被应用于获得基因表达数据，并可在 GEO (Gene Expression Omnibus) 数据库中下载，本研究中，我们在 GEO 数据库中下载了两个微阵列数据集，探究影响胃肠道间质瘤风险分级的差异表达基因，并对其进一步分析。

2. 资料与方法

2.1. 数据获取

在 GEO 数据库(<http://www.ncbi.nlm.nih.gov/geo>) 检索公开发表的基因芯片数据集，检索条件：1) 检索词为“gastrointestinal stromal tumor”；2) 物种为“Homo sapiens”。最终选取了 GSE31802, GSE47911 两个数据集，分为高风险组和非高风险组。

2.2. 差异基因的认识

利用 GEO 数据库提供的在线分析工具 GEO2R 下载差异表达的基因数据，以差异倍数 $|\log_{2}FC| > 2$, $p < 0.05$ 作为差异基因的筛选条件，并利用在线工具 Draw Venn Diagram (<https://bioinformatics.psb.ugent.be/webtools/Venn>) 绘制维恩图，识别两个数据集共有的差异性表达基因。

2.3. 功能富集分析

利用注释可视化和集成发现数据库(Database for Annotation Visualization and Integrated Discovery, DAVID) (<https://david.ncifcrf.gov/>) 对两个数据集共有的差异性表达基因进行基因本体论(Gene Ontology, GO)和京都基因与基因组百科全书(Kyoto Encyclopedia of Genes and Genomes, KEGG)的功能富集分析。GO 分析包括细胞成分(cellular component, CC)、生物过程(biological process, BP)和分子功能(molecular function, MF)。

2.4. 蛋白质相互作用分析

将筛选的两个数据集共有的差异性表达基因信息输入 STRING 数据库(<http://string-db.org/>)，评估它们之间的功能关联，再用 Cytoscape 软件(version 3.9.1)分析这些基因的关联程度。

3. 结果

3.1. 胃肠道间质瘤高风险组和非高风险组中差异表达基因的筛选

我们检索到了 GEO 数据库中关于胃肠道间质瘤恶性程度的差异基因表达数据集：GSE31802 和 GSE47911，利用 GEO2R 分别筛选出这两个数据集中差异性排名前 100 的基因及其相关数据，绘制热图如图 1 所示。进一步分析到两个数据集之间重叠的差异性表达基因有 42 个，如图 2 所示。

3.2. GO 富集分析与 KEGG 信号通路富集分析

通过 GO 分析，两个数据集共有的差异表达基因的生物学过程可能主要富集在以下几个方面：cAMP 分解代谢过程，信号转导，基因表达的正向调控，酪氨酸激酶信号通路。细胞学组分主要定位于质膜以及内质网腔。分子功能富集于信号素受体活性，序列特异性双链 DNA 结合，如图 3 所示。KEGG 信号通路的富集分析显示：差异基因主要富集于轴突的指导，癌症中的转录失调等，如图 4 所示。

3.3. 蛋白质相互作用分析

将重叠的差异表达基因信息输入 STRING 评估他们之间的功能关联，再用 Cytoscape 软件分析这些基因的关联情况，并将关联程度进行等级排序，结果显示 IL-6, PGR 的节点分数最高，如图 5 所示。

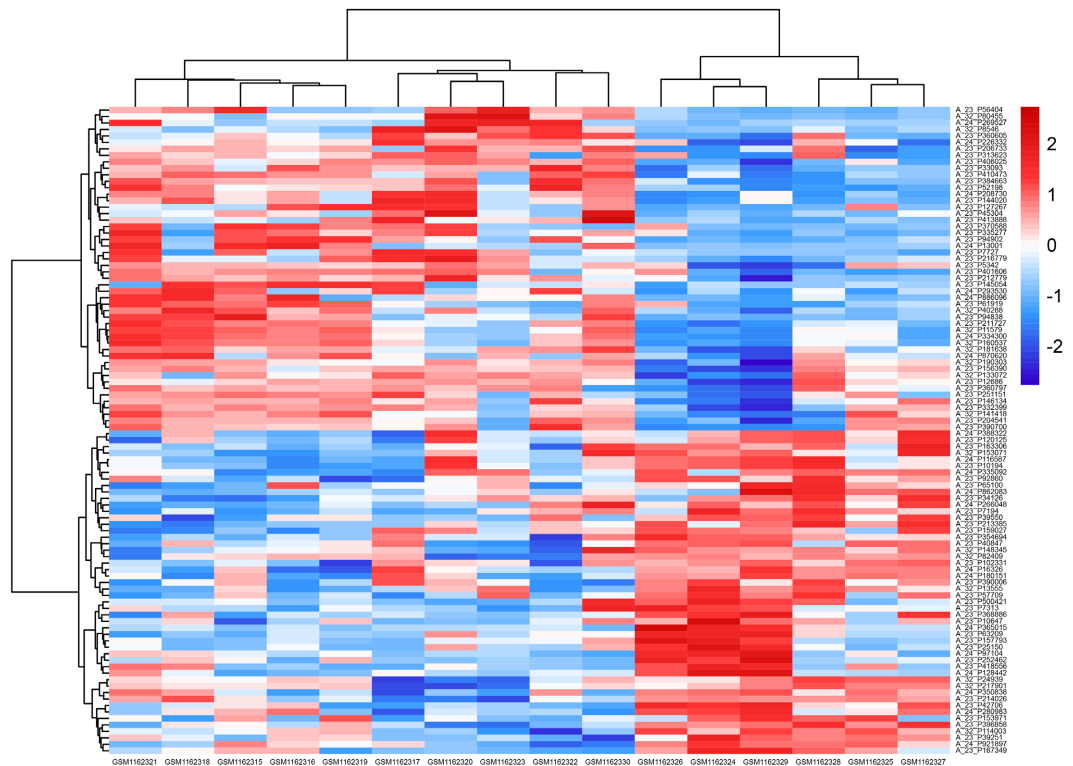
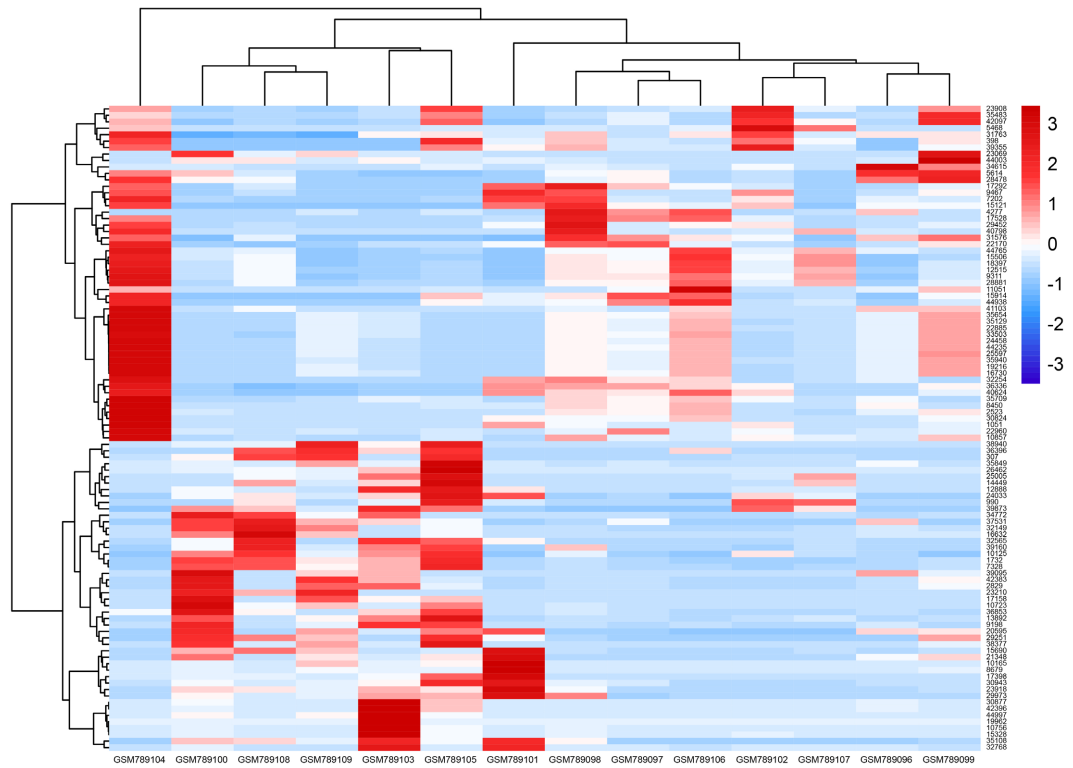


Figure 1. Differential gene expression Heat map of GSE31802, GSE47911 dataset
图 1. GSE31802, GSE47911 数据集差异基因表达热图

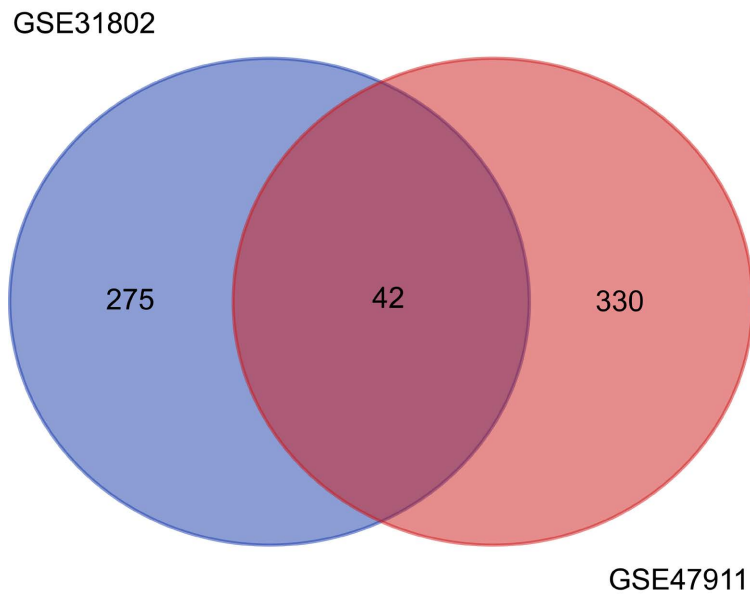


Figure 2. Venn diagram revealing the differentially expressed genes
图 2. 两组芯片数据差异表达基因韦恩图

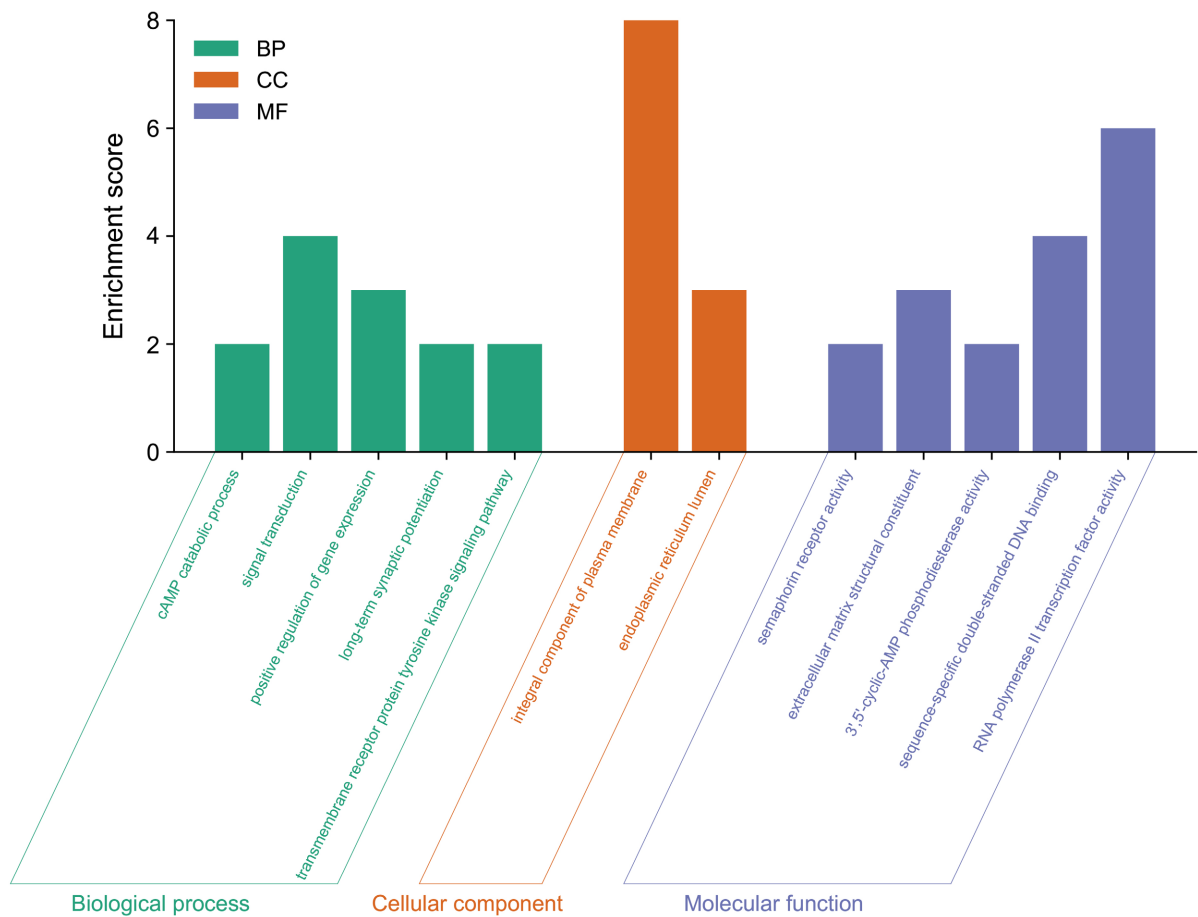


Figure 3. GO analysis of the differentially expressed genes
图 3. 差异性表达基因的 GO 分析

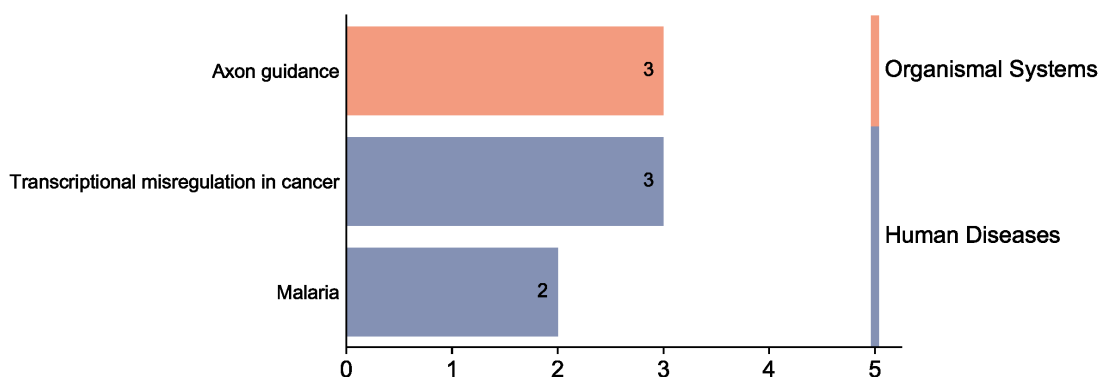


Figure 4. Enrichment analysis of KEGG signaling pathway of the differentially expressed genes
图 4. 差异性表达基因的 KEGG 信号通路富集分析

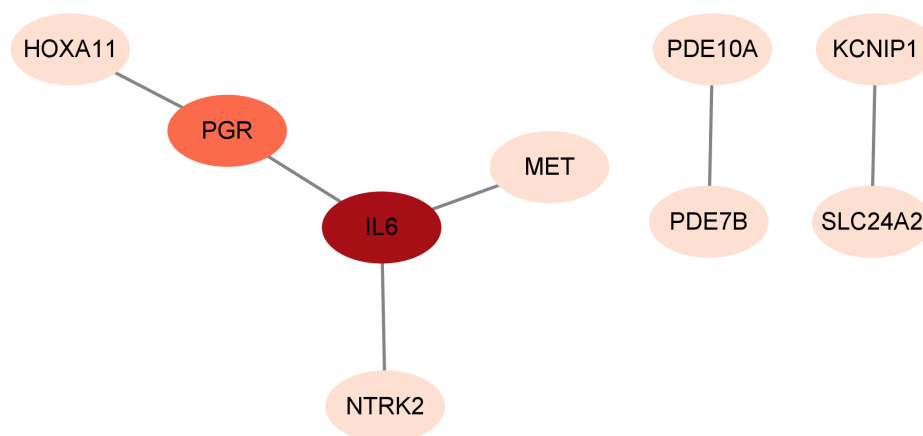


Figure 5. PPI network diagram of shared differentially expressed genes
图 5. 差异性表达基因的蛋白质相互作用

4. 讨论

在本研究中，我们以 GEO 数据库为基础，检索到了 2 个关于胃肠道间质瘤恶性程度的差异基因表达数据(GSE31802 和 GSE47911)，分析出这两个数据集之间重叠的差异性表达基因有 42 个。这 42 个共有的差异基因的生物学过程可能主要富集在 cAMP 分解代谢过程，信号转导，基因表达的正向调控和酪氨酸激酶信号通路这几个方面。细胞学组分主要定位于质膜以及内质网腔。分子功能富集于信号素受体活性、序列特异性双链 DNA 结合。KEGG 信号通路的富集分析显示差异基因主要富集于轴突的指导、癌症中的转录失调等。我们采用 Cytoscape 软件进一步分析差异基因的功能关联，并进行了排序，其中 IL-6, PGR 的节点分数最高，它们可能在胃肠道间质瘤的风险分级中起到了一定的调控作用。

近年来，细胞因子相互作用的生化和免疫网络受到了极大的关注，这些网络有助于促进这些肿瘤的进展和转移扩散[10]。巨噬细胞来源的细胞因子 IL-6、TNF- α 、IL-1b 和干扰素-g 在炎症的诱导和放大中起着关键作用，Silke Cameron 等人研究发现，在 GIST 中，急性期细胞因子 RNA 表达很低，尤其是 TNF- α 和 IL-6 [11]。孕激素受体(progesterone receptor, PGR)通过配体依赖和非配体依赖的激活来调节细胞的生长和分化，在乳腺癌和妇科肿瘤中，它们的表达是内分泌治疗益处和良好的非治疗预后的已知预测因子 [12]。但在软组织肉瘤中，它们的表达谱不明确，其意义也不确定。本研究提供了一些可能性，但这些基因是否与胃肠道间质瘤的风险程度相关，在其中的作用机制如何，则有赖于更进一步的研究。

5. 结论

胃肠道间质瘤是消化系统最常见的间叶肿瘤,可涉及整个胃肠道,多见于胃和小肠。在治疗手段上,主要是开腹、腹腔镜、内镜切除等手术方式,再结合分子靶向药物进行辅助治疗。尽管当前治疗方法取得了较好效果,但部分患者依然会出现复发或转移。因此明确影响患者危险分级的因素有助于更好地评估胃肠道间质瘤患者,为诊断和治疗提供参考。

参考文献

- [1] Akahoshi, K., Oya, M., Koga, T. and Shiratsuchi, Y. (2018) Current Clinical Management of Gastrointestinal Stromal Tumor. *World Journal of Gastroenterology*, **24**, 2806-2817. <https://doi.org/10.3748/wjg.v24.i26.2806>
- [2] Blay, J.Y., Kang, Y.K., Nishida, T. and von Mehren, M. (2021) Gastrointestinal Stromal Tumours. *Nature Reviews Disease Primers*, **7**, 22. <https://doi.org/10.1038/s41572-021-00254-5>
- [3] 李龙泉,王永庆,马明,黄田,梁元,张惠芸,黄缘壁,孙蔚明,王玉平. 胃肠道间质瘤治疗进展[J]. 医学综述, 2021, 27(8): 1539-1544.
- [4] 王芳芳,韩宇,张彩凤. 胃肠道间质瘤的研究进展[J]. 现代消化及介入诊疗, 2021, 26(10): 1318-1322.
- [5] 杜静,卢桂芳,马文慧,任牡丹,殷燕,卢新兰,和水祥. 胃间质瘤临床特征及免疫组化结果与危险度分级的相关性[J]. 现代肿瘤医学, 2022, 30(7): 1255-1259.
- [6] Joensuu, H., Hohenberger, P. and Corless, C.L. (2013) Gastrointestinal Stromal Tumour. *Lancet*, **382**, 973-983. [https://doi.org/10.1016/S0140-6736\(13\)60106-3](https://doi.org/10.1016/S0140-6736(13)60106-3)
- [7] Liu, L., Chen, J., Shan, J. and Sun, X. (2023) Development and Validation of an EUS-Based Nomogram for Prediction of the Malignant Potential in Gastrointestinal Stromal Tumors. *Scandinavian Journal of Gastroenterology*, 1-8. <https://doi.org/10.1080/00365521.2023.2175179>
- [8] Joensuu, H., Vehtari, A., Riihimäki, J., et al. (2012) Risk of Recurrence of Gastrointestinal Stromal Tumour after Surgery: An Analysis of Pooled Population-Based Cohorts. *The Lancet Oncology*, **13**, 265-274. [https://doi.org/10.1016/S1470-2045\(11\)70299-6](https://doi.org/10.1016/S1470-2045(11)70299-6)
- [9] 朱从波,廖国庆,赵丁民. Ki-67 对胃肠道间质瘤预后的评估价值[J]. 临床与病理杂志, 2018, 38(8): 1632-1639.
- [10] Madej-Michniewicz, A., Budkowska, M., Sałata, D., Dołęgowska, B., Starzyńska, T. and Błogowski, W. (2015) Evaluation of Selected Interleukins in Patients with Different Gastric Neoplasms: A Preliminary Report. *Scientific Reports*, **5**, 14382. <https://doi.org/10.1038/srep14382>
- [11] Cameron, S., Haller, F., Dudas, J., et al. (2008) Immune Cells in Primary Gastrointestinal Stromal Tumors. *European Journal of Gastroenterology & Hepatology*, **20**, 327-334. <https://doi.org/10.1097/MEG.0b013e3282f3a403>
- [12] Valkov, A., Sorbye, S., Kilvaer, T.K., et al. (2011) Estrogen Receptor and Progesterone Receptor Are Prognostic Factors in Soft Tissue Sarcomas. *International Journal of Oncology*, **38**, 1031-1040. <https://doi.org/10.3892/ijo.2011.920>