

基于转录组数据筛选胸腺瘤致病基因及调控通路

姜佳辰¹, 张芝², 李金田^{1,3*}

¹甘肃中医药大学中医临床学院, 甘肃 兰州

²甘肃中医药大学第一临床医学院, 甘肃 兰州

³敦煌医学与转化教育部重点实验室, 甘肃 兰州

收稿日期: 2023年7月18日; 录用日期: 2023年8月9日; 发布日期: 2023年8月16日

摘要

目的: 综合生物信息学分析, 探究胸腺瘤的致病基因及调控通路。方法: GEO基因表达总数据库检索并下载3组胸腺瘤基因芯片数据, 并利用GEO2R筛选出差异基因(DEGs), 然后通过Metascape和STRING在线软件对差异基因进行GO、KEGG富集分析和蛋白互作网络构建; 然后应用Cytoscape软件筛选致病基因; 最后应用UALCAN和GEPIA数据库对关键基因进行验证、生存分析及调控机制通路探究。结果: 分析三组数据集获得1259个共同DEGs, GO、KEGG富集分析和SinoBiological网站检索显示DEGs参与肿瘤坏死因子信号通路、跨膜受体蛋白酪氨酸激酶信号通路、TNF信号通路、PI3K-Akt信号通路、鞘脂类信号通路、Hippo信号通路、钙信号通路、癌症通路、癌症中的转录失调和造血细胞系。30个关键致病基因, 包括ITGA4、ITGB3、ITGB5、ITGB6、ITGA8、ITGA11、COL6A1、THBS2、COL6A2等。通过对致病基因进行验证、通路探究和生存分析, 发现ITGA4、CD4、THY1、CD1D和IL2RB均可影响患者总体生存率, 提示可作为THY1的潜在研究方向。结论: 通过公共数据库的生物信息检索, 筛选出致病基因及其参与的机制通路, 其中ITGA4、CD4、THY1、CD1D和IL2RB基因将可能成为新的分子生物标志物和靶向基因治疗目标。

关键词

胸腺瘤, 生物信息学分析, 致病基因

To Screen Thymoma Pathogenic Genes and Regulatory Pathways Based on Transcriptome Data

Jiachen Jiang¹, Zhi Zhang², Jintian Li^{1,3*}

*通讯作者。

¹Clinical College of Chinese Medicine, Gansu University of Chinese Medicine, Lanzhou Gansu

²First School of Clinical Medicine, Gansu University of Chinese Medicine, Lanzhou Gansu

³Key Laboratory of Dunhuang Ministry of Medicine and Transformation, Lanzhou Gansu

Received: Jul. 18th, 2023; accepted: Aug. 9th, 2023; published: Aug. 16th, 2023

Abstract

Objective: To investigate the pathogenic genes and regulatory pathways of thymoma by comprehensive bioinformatics analysis. **Methods:** GEO Gene Expression Omnibus database was retrieved and downloaded for 3 sets of thymoma gene microarray data, and the differential genes (DEGs) were screened by GEO2R, then GO, KEGG enrichment analysis and protein interaction network construction were performed for the differential genes by Metascape and STRING online software; then the pathogenic genes were screened by Cytoscape software; finally, the UALCAN and GEPIA databases were applied to validate the key genes, analyze their survival and explore the regulatory pathways. **Results:** Analysis of the three datasets yielded 1259 common DEGs, GO, KEGG enrichment analysis and SinoBiological website search showed that DEGs were involved in tumor necrosis factor signaling pathway, transmembrane receptor protein tyrosine kinase signaling pathway, TNF signaling pathway, PI3K-Akt signaling pathway, sphingolipid signaling pathway, Hippo signaling pathway, calcium signaling pathways, cancer pathways, transcriptional dysregulation in cancer and hematopoietic cell lines. 30 key pathogenic genes include ITGA4, ITGB3, ITGB5, ITGB6, ITGA8, ITGA11, COL6A1, THBS2, COL6A2, etc. Through validation, pathway exploration and survival analysis of the pathogenic genes: ITGA4, CD4, THY1, CD1D and IL2RB could all affect the overall survival of patients, suggesting a potential research direction for THY. **Conclusion:** The causative genes and their involved mechanistic pathways were screened by bioinformatics search of public databases, in which ITGA4, CD4, THY1, CD1D and IL2RB genes will likely become new molecular biomarkers and targets for targeted gene therapy.

Keywords

Thymoma, Bioinformatics Analysis, Disease Genes

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

胸腺瘤(Thymoma, THY)是一种罕见、恶性程度较低的胸腺上皮细胞肿瘤,其特点是生长缓慢且极少转移,在成人患者中很常见。胸腺瘤与体液和细胞免疫缺陷有关,致病基因可能在胸腺瘤发展与免疫失调相关的病因机制中发挥相关作用。研究统计,近50%胸腺瘤患者会发展为重症肌无力(MG),而近80%的重症肌无力患者会发现胸腺瘤[1]。治疗胸腺瘤的主要方法是外科手术切除。WHO分型以病理组织学特征为依据,Masaoka分期以肿瘤侵袭范围为依据,胸腺瘤的预后与WHO分型和Masaoka分期显著相关[2][3][4]。近些年来,利用NCBI的生物信息学基因数据库和TCGG临床数据库,探索胸腺瘤患者发病的可能致病基因和机制通路,为肿瘤的临床早期诊断、个体化肿瘤药物的靶向治疗和预后提供指导。

2. 数据材料

检索 NCBI 数据库并下载 3 组胸腺瘤数据集, 分别为 GSE94769、GSE55852 和 GSE42977, 样本属于人类胸腺上皮肿瘤组织, 数据集来源于 GPL13534、GPL10123 和 GPL6790 芯片平台。

3. 处理方法

3.1. 基因数据处理及 DECs 筛选

GEO 中检索到胸腺瘤基因芯片数据, 同时利用 GEO2R 在线工具分析将 3 组数据集整合处理, 并进行差异性基因 DECs 筛选, 然后将 3 组数据集的差异基因筛选结果绘制火山图和 Venn 图。

3.2. 差异表达基因的功能富集分析

应用 Metascape 在线软件, 检索差异表达基因(DECs)的生物学功能和调控机制通路, 并进行 GO、KEGG 功能富集分析、SinoBiological 网站的调控通路探究。富集分析就是分析一组基因在某个功能节点上的作用交集。京都基因和基因组百科全书(KEGG)是构成基因组、化学和功能分析的综合数据资源, 是了解生物体高级功能和生物系统的基因组重要数据库。

3.3. 差异表达基因的蛋白质相互作用分析

STRING 数据库是检索目标基因相互作用关系的数据信息合集, 其中包含丰富广泛的物种信息、多样蛋白质种类及蛋白相互作用联系, 可由于分析蛋白质与蛋白质间的相互作用关系(PPI), 有助探究主要基因的调控网络关系。

3.4. 关键基因筛选

将 STRING 构建致病基因的蛋白质互作网络 PPI 的结果导入 Cytoscape3.8.2 软件, 应用 cytoHubba 插件将 MCC 算法中交集前 30 的关键基因筛选, 选定在 12 种算法中出现频率大于 10 次的基因为致病基因。绘制关键基因的网络相关作用图并导出数据分析结果。

3.5. 验证关键基因表达水平及生存分析

应用 UALCAN 和 GEPIA 在线基因分析软件, 验证致病基因在胸腺瘤中的表达值及临床生存分析。通过不同数据库的双重验证分析, 对胸腺瘤的关键致病基因进行探究分析。

4. 结果

4.1. THY 差异表达基因筛选

检索 NCBI 数据库并应用 GEO2R 软件, 对 GSE94769、GSE55852 和 GSE42977 集合数据集, 进行差异表达基因进行分析, 共筛选出 1259 个差异表达基因(DEGs), 并绘制 Venn 图(图 1)和火山图(图 2)。

4.2. 差异表达基因的 GO 富集分析

通过 Metascape 软件对 1259 个差异基因进行 GO 富集分析, 结果显示差异基因表达产物主要参与细胞粘附的调控调节、细胞形态发生的调控、细胞的激活、细胞间质发展等; 管道形态发生、信息节点连接、调控复杂受体、跨膜受体蛋白酪氨酸激酶信号通路、肌动蛋白 filament-based 过程、大脑发育、脊索动物的胚胎发育及组织形态发生形成等(图 3)。

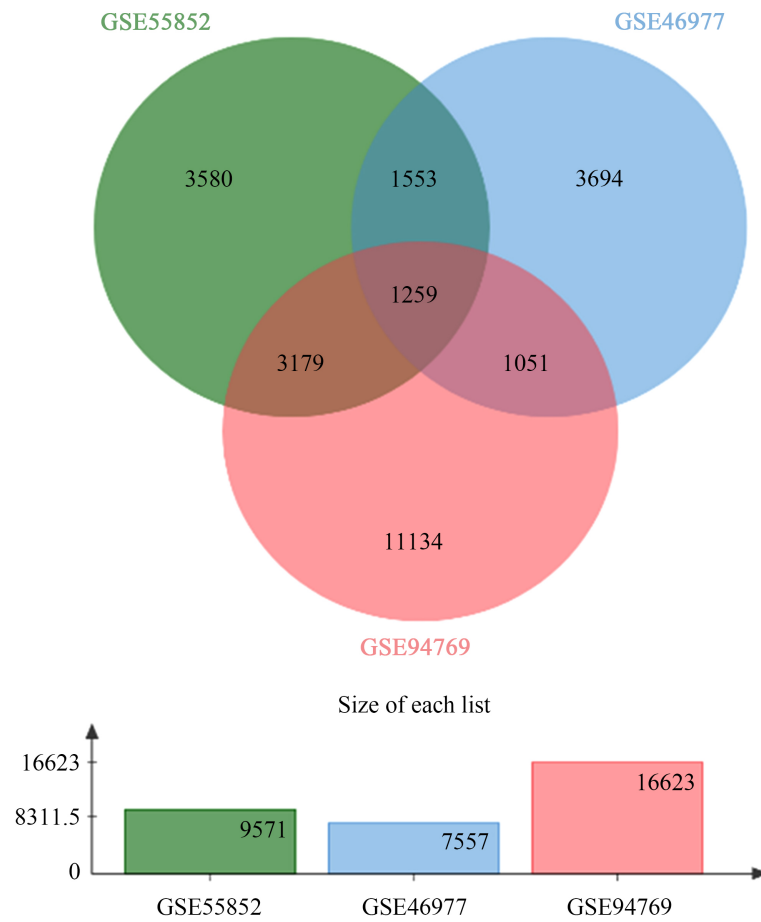


Figure 1. Venn diagram of differentially expressed genes
图 1. 差异表达基因 Venn 图

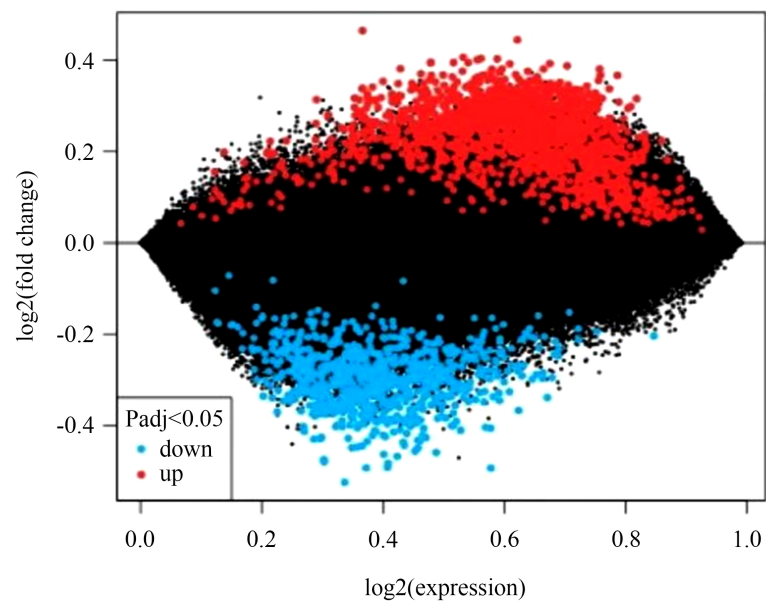


Figure 2. Volcano diagram of differentially expressed genes
图 2. 差异表达基因火山图

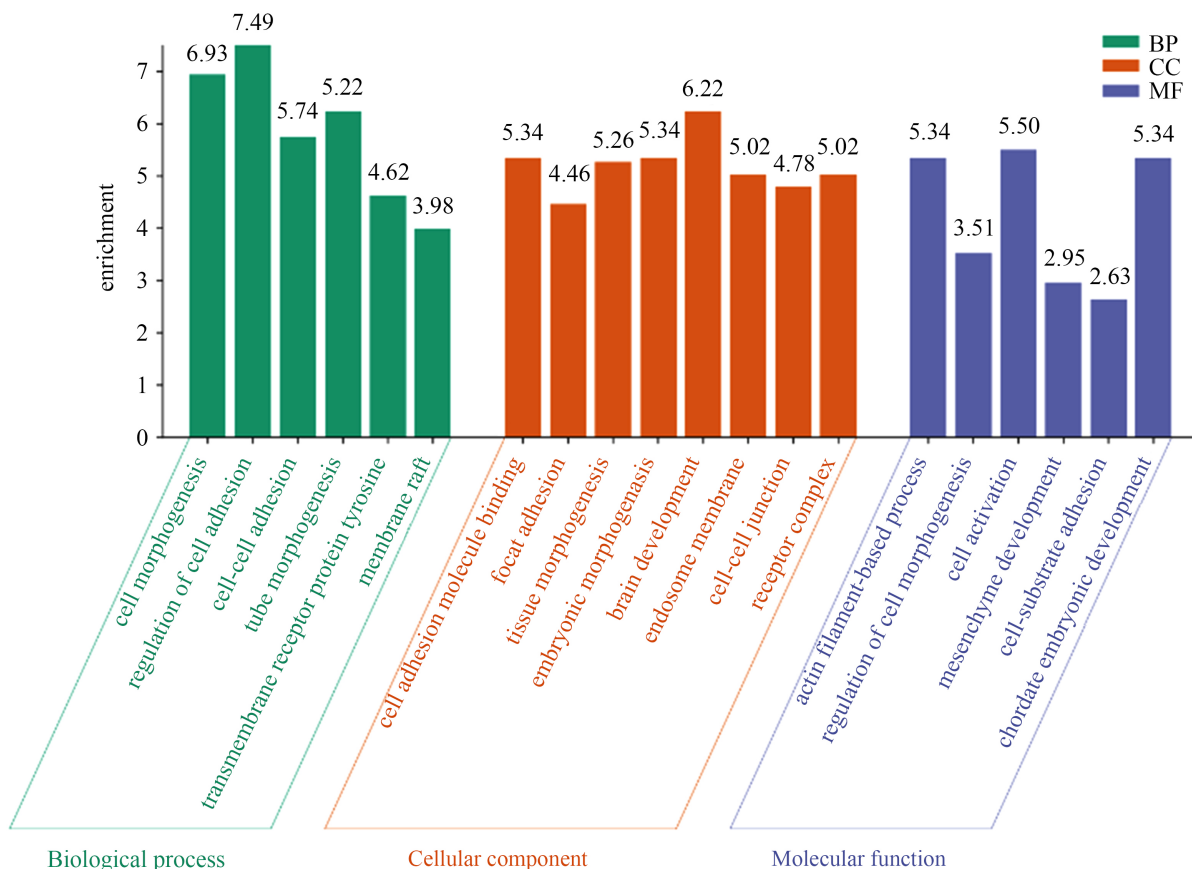


Figure 3. GO enrichment analysis of differential genes
图 3. 差异基因 GO 富集分析

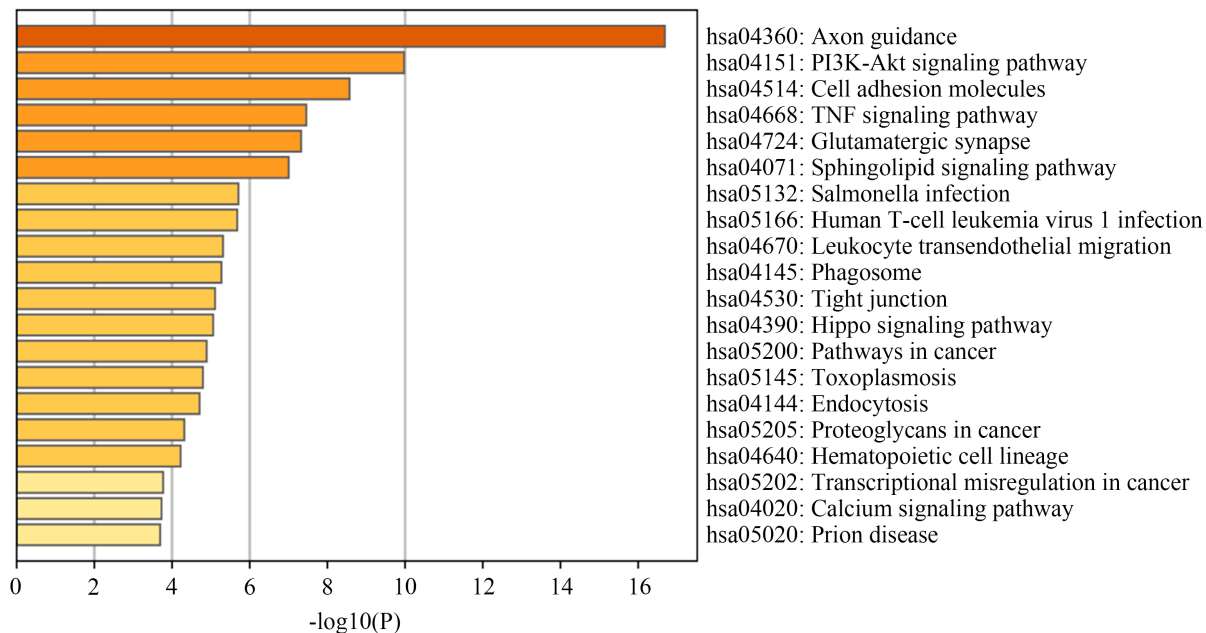


Figure 4. KEGG enrichment analysis of differential genes
图 4. 差异基因 KEGG 富集分析

4.3. 差异表达基因的 KEGG 富集分析

应用 Metascape 软件对 1259 个差异基因进行富集通路探究, 其主要参与的信号通路有肿瘤坏死因子信号通路、TNF 信号通路、PI3K-Akt 信号通路、鞘脂类信号通路、Hippo 信号通路、钙信号通路、癌症通路、癌症中的转录失调和造血细胞系等(图 4)。

4.4. 差异表达基因 PPI 作用分析结果

STRING 软件导入乳腺癌的差异表达基因, 进行蛋白质 - 蛋白质相互作用网络 PPI 构建, 最终筛选出前 30 个致病基因制作网络关系图, 节点数量共 1252 个, 节点变数共 5482 条, 相互作用关系共获得 4676 条。

4.5. 关键基因筛选结果

通过 STRING 数据库构建差异表达基因蛋白质互作 PPI 网络关系图, 并将 PPI 分析结果导入 Cytoscape 中, 利用 cytoHubba 插件的 MCC 算法从 PPI 网络中筛选前 30 个关键致病基因, 包括 ITGA4、ITGB3、ITGB5、ITGB6、ITGA8、ITGA11、COL6A1、THBS2、COL6A2、COL9A1、THBS3、PSMA7 等(图 5)。

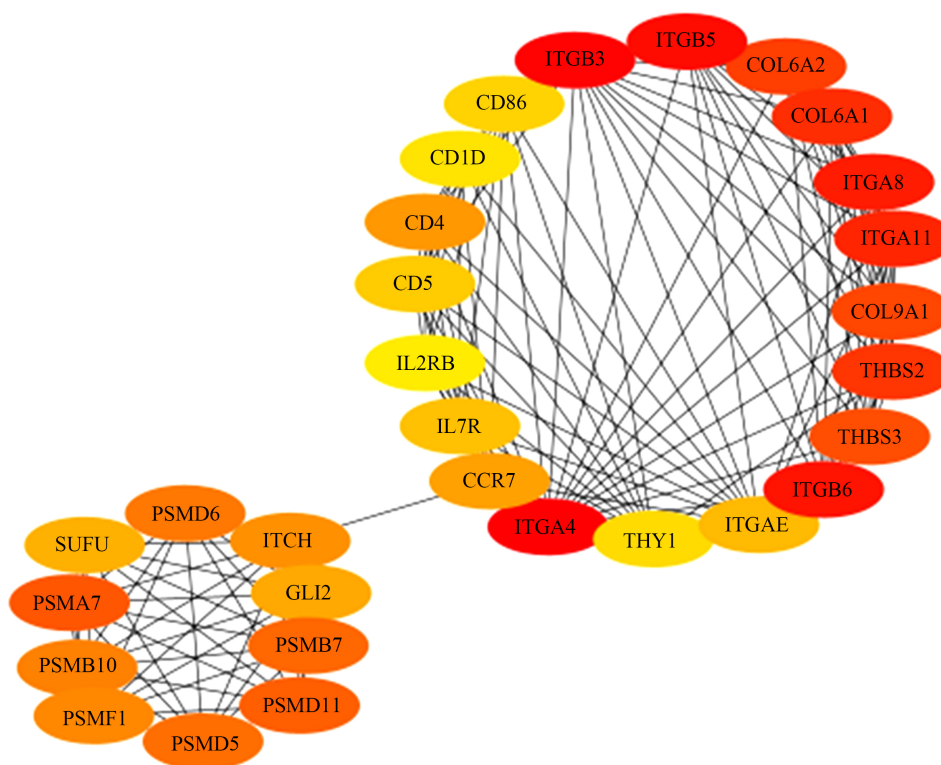


Figure 5. PPI network diagram of the top 30 key pathogenic genes
图 5. 前 30 个关键致病基因 PPI 网络关系图

4.6. 关键基因验证及生存分析

通过检索 UALCAN 和 GEPIA 两大数据库中乳腺癌的数据信息, 对选定的关键基因进行验证、通路探究和生存分析。ITGA4 ($P = 0.039$)、CD4 ($P = 0.03$)、THY1 ($P = 0.04$)、CD1D ($P = 0.034$)和 IL2RB ($P = 0.0017$)对患者的总生存率具有影响意义(图 6~10)。

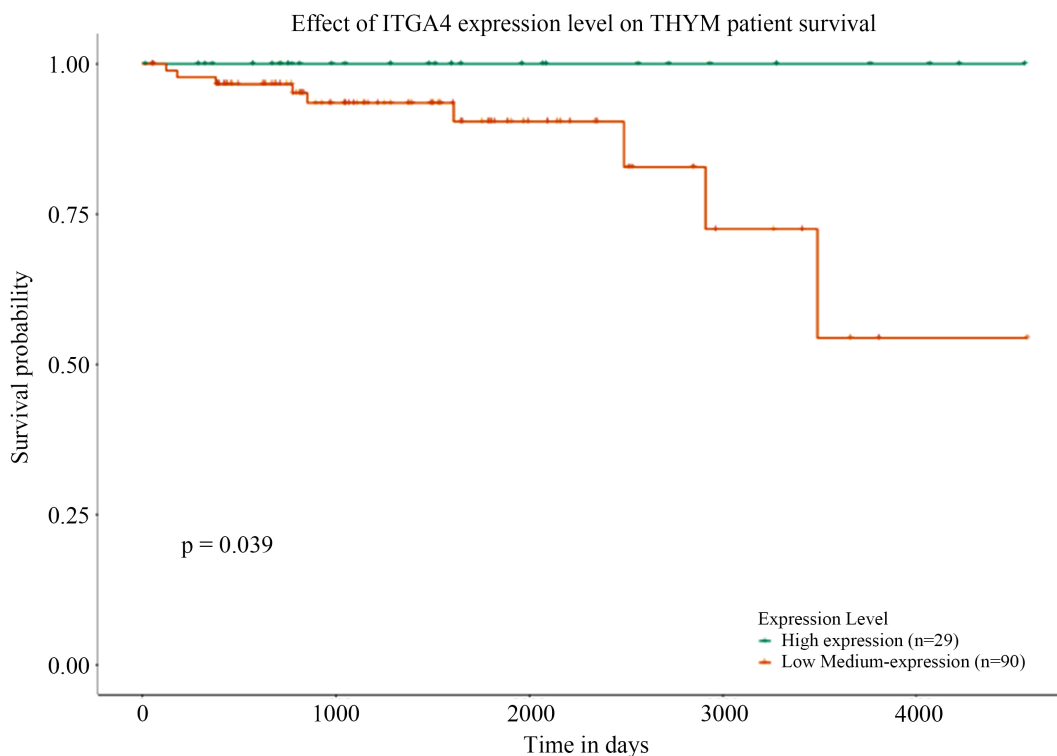


Figure 6. Survival analysis of ITGA4 gene
图 6. ITGA4 基因生存分析图

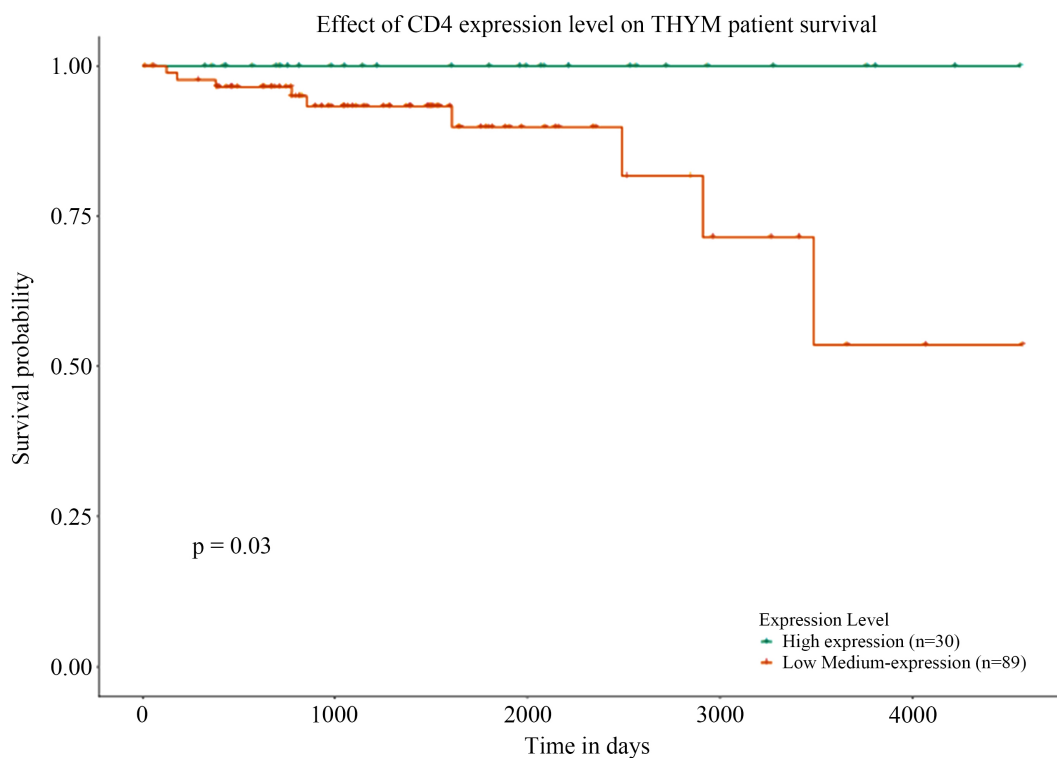


Figure 7. Survival analysis of CD4 gene
图 7. CD4 基因生存分析图

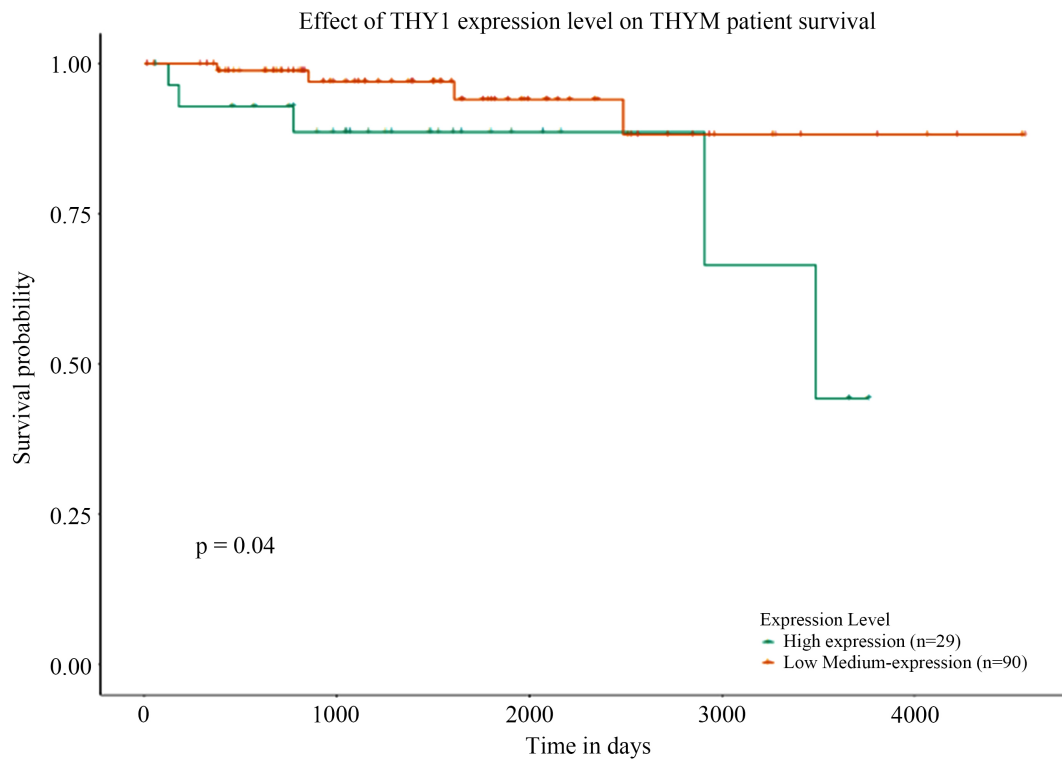


Figure 8. Survival analysis of THY1 gene
图 8. THY1 基因生存分析图

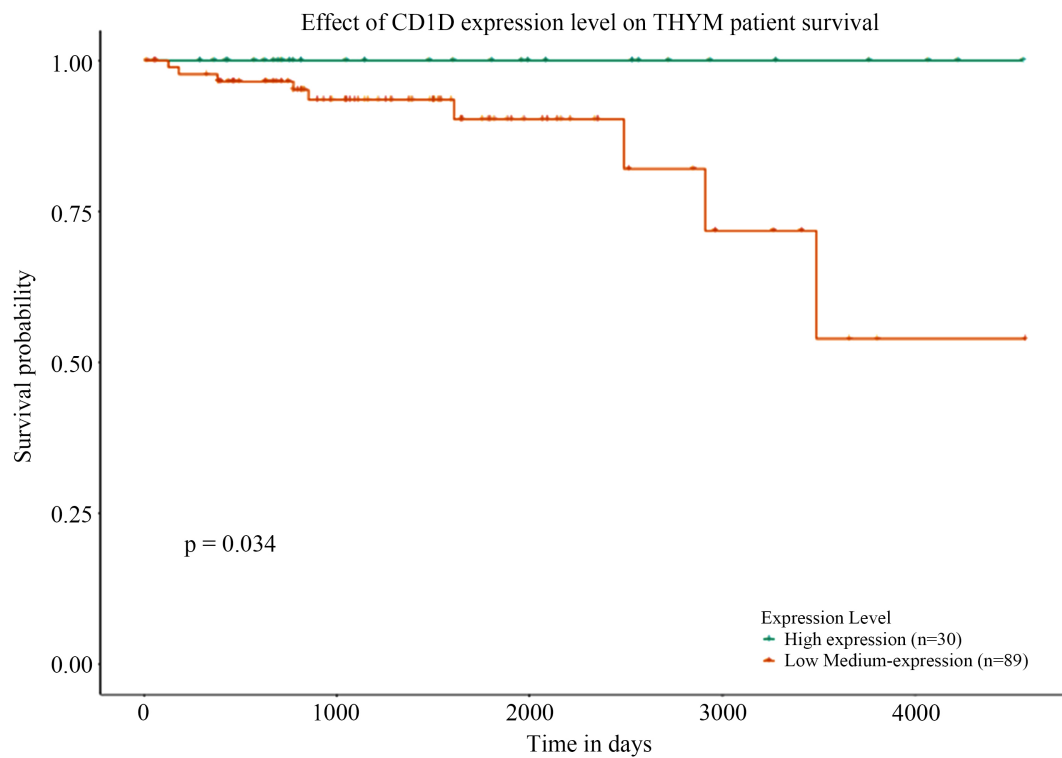


Figure 9. Survival analysis of CD1D gene
图 9. CD1D 基因生存分析图

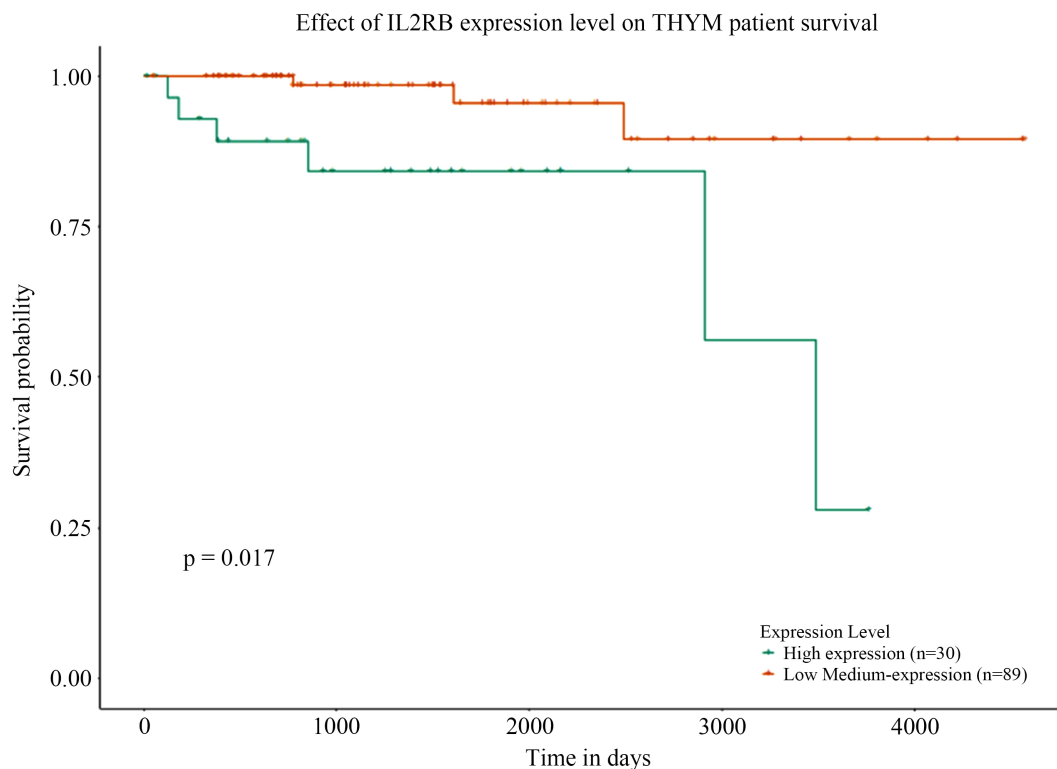


Figure 10. Survival analysis of IL2RB gene
图 10. IL2RB 基因生存分析图

5. 讨论

胸腺瘤(Thymoma, THY)是一种恶性程度较低的胸腺上皮细胞肿瘤,其特点是生长缓慢且极少转移。临床表现多为气短、咳嗽、胸痛等,1/5 病患可能合并重症肌无力(MG);而 MG 患者约半数以上有胸腺瘤或胸腺增生异常[5]。恶性肿瘤胸腺瘤是罕见的,胸腺瘤通常发生在中年患者[6]。Masaoka 临床分期系统将胸腺瘤分为 A、AB、B1、B2 和 B3 或混合型,手术切除胸腺是目前治疗该病最有效的手段[7],虽然胸腺瘤不常见且恶性肿瘤发生率低,但胸腺瘤与免疫系统疾病之间有密切关联。基于生物数据库及高通量技术的探究,通过对胸腺瘤发病的相关致病基因及机制通路的检索,揭示出相关基因及其发病机制可能成为胸腺瘤的分子生物标志物和靶向基因治疗目标。

本研究通过 GEO 基因表达总数据库检索并下载 3 组胸腺瘤基因芯片数据集,筛选出 1259 个差异基因(DEGs),然后通过 Metascape 和 STRING 在线软件对差异基因进行生物学功能、机制通路分析和蛋白质-蛋白质作用网络构建。其中差异基因所涉及的信号通路有肿瘤坏死因子信号通路、跨膜受体蛋白酪氨酸激酶信号通路、TNF 信号通路、PI3K-Akt 信号通路、鞘脂类信号通路、Hippo 信号通路、钙信号通路、癌症通路、癌症中的转录失调和造血细胞系等。应用在线软件检索和关键基因的 PPI 网络构建,筛选出 30 个关键基因。通过 UALCAN 和 GEPIA 数据库对致病基因进行验证、生存分析和调控通路研究。通过生存分析验证证实,ITGA4 ($P = 0.039$)、CD4 ($P = 0.03$)、THY1 ($P = 0.04$)、CD1D ($P = 0.034$)和 IL2RB ($P = 0.0017$)基因均可影响患者总体生存率,提示生物标志物可能成为 THY 的潜在研究方向。

肿瘤坏死因子(TNF)信号通路:肿瘤坏死因子超家族的成员通过配体介导的三聚体发挥作用,引起多个细胞内适配器的募集,从而激活多种信号转导途径。其中 TNF 家族蛋白的募集可导致转录因子如 NF-kappaB 和 JNK 等激活,从而促进细胞分化、细胞免疫、炎症反应和肿瘤细胞死亡等[8] [9] [10]。跨

膜受体蛋白酪氨酸激酶信号通路(TPK-Ras-MAPK),其中RAS作为原癌基因,是TPK传递信号的重要部分,非受体酪氨酸蛋白激酶,在胸腺发育中的T细胞的生长增殖功能中起重要作用。PI3K-Akt信号通路属于G蛋白偶联受体(GPCRs)受体调节通路,可调节多种正常的生物学过程。张俊英等认为,PI3K-Akt信号通路通过激活、活化NF- κ B,从而介导机体细胞免疫及体液免疫应答,引起相关肺炎疾病[11]。Hippo信号通路参与机体细胞的生长、分化、增殖,在肿瘤的病程进展中发挥作用;赵昊等揭示了Hippo信号通路在免疫稳态调控中发挥重要作用[12]。

ITGA4(整合素A4)基因编码一条 α 4链。研究发现,ITGA4可能参与细胞溶解性T细胞相互作用,参与G蛋白偶联受体信号转导通路、MAPK-Erk信号通路和自噬信号通路等[13]。CD4(T细胞表面糖蛋白)通常位于T细胞内的部分,与T细胞表面的受体互相作用,激活相应免疫细胞并调节免疫作用,放大TCR产生的信号[14]。THY-1在细胞迁移、粘附、分化、再生、凋亡、机械转导和细胞分裂中发挥重要作用,因而参与T细胞活化、神经突生长、细胞凋亡、肿瘤抑制、伤口愈合和纤维化等过程[15][16]。

抗原递呈糖蛋白CD1D,CD1家族包括CD1A-CD1E,CD1受限的NKT细胞被抗原刺激后可以通过立即产生大量细胞因子来激活或抑制其他免疫活性细胞,因此,这些细胞被认为是免疫调节剂[17],可结合自身和非自身糖脂并将其呈现给自然杀伤性T细胞上的T细胞受体的抗原递呈蛋白。IL2RB(白细胞介素2受体),隶属I型细胞因子受体家族,通过调控T细胞的活化活性,从而促进T细胞的生长、分化和调节,并积极参与调节机体免疫功能。在大多数肿瘤类型中,研究发现,IL2RB高表达与预后不良相关,提示IL2RB在肿瘤发生、增殖、转移过程中起着重要作用。Li G等认为,IL2RB含量与肿瘤浸润淋巴细胞(如泛癌中的CD8+T细胞、树突状细胞和巨噬细胞)的丰度呈正相关[18]。ITGA4、CD4、THY-1、CD1D和IL2RB基因的发病机制与免疫反应相关联,同时其机制通路应需研究验证[19]。

6. 结论

综上所述,胸腺瘤(THY)是低度恶性罕见的免疫系统肿瘤,其发病基因机制与免疫系统疾病相互关联。然而,由于基因数据库样本的局限及胸腺瘤实验的欠缺,后期研究仍需要大量样本参数进行验证。基于生物数据库及高通量技术的探究,通过对胸腺瘤发病的相关致病基因及机制通路的探究,揭示ITGA4、CD4、THY-1、CD1D和IL2RB基因可能成为THY的潜在研究方向及其发病机制可能成为新的治疗靶点,亦为THY提供治疗新方向。

参考文献

- [1] Chang, R., Duan, S., Li, S., *et al.* (2021) Viral Infection in Thymoma and Thymic Tumors with Autoimmune Diseases. *Thoracic Cancer*, 12, 2971-2980. <https://doi.org/10.1111/1759-7714.14157>
- [2] 张博, 韩正祥, 秦晓冰, 桂鑫. 119例胸腺瘤临床特征分析[J]. 徐州医科大学学报, 2021, 41(6): 403-407.
- [3] 王立坤, 杜天宇, 肖旭阳. 胸腺瘤 WHO 分型和 Masaoka 分期的影响因素分析[J]. 江苏医药, 2022, 48(10): 1023-1026.
- [4] 李慧暘. 胸腺瘤和胸腺癌中胸腺上皮标记表达与分型、分期和预后相关性分析[D]: [硕士学位论文]. 长春: 吉林大学, 2022.
- [5] 张俊英, 赵晓莉, 王鑫, 单馨慧, 朱丕雷. 不同气道损害肺炎支原体肺炎患儿 PI3K-Akt 信号通路变化及治疗转归[J]. 中华医院感染学杂志, 2022, 32(7): 1106-1110.
- [6] 赵昊, 史一然, 章津嘉, 陈兰芬. 非经典 Hippo 信号通路调控免疫稳态[J]. 厦门大学学报(自然科学版), 2022, 61(3): 377-386.
- [7] 彭一鹏, 周晓飞, 王剑, 等. 肿瘤坏死因子 α 诱导蛋白3相互作用蛋白1调控核因子 κ B信号通路在IDH野生型胶质瘤中的作用机制[J]. 实用医学杂志, 2022, 38(16): 2002-2008.
- [8] 路蔓, 付强, 张静. 肿瘤坏死因子- α 诱导对膀胱癌细胞株NF- κ B信号通路分子表达及细胞增殖水平的影响[J].

- 检验医学与临床, 2021, 18(15): 2145-2147+2151.
- [9] 王蕊. 止麻消痰活血汤对老年心肌梗死大鼠神经调节蛋白-1/受体酪氨酸蛋白激酶 erbB3 通路及心室重构的影响[J]. 安徽医药, 2022, 26(7): 1315-1319+1486.
- [10] Mo, J., Zhang, J., Huang, H., Liu, C., Cheng, Y., Mo, Y., Wu, S., Zhong, Y., Zhong, C. and Zhang, B. (2022) The Early Predictive Effect of Low Expression of the *ITGA4* in Colorectal Cancer. *Journal of Gastrointestinal Oncology*, **13**, 265-278. <https://doi.org/10.21037/jgo-22-92>
- [11] 梁可克, 王志旺, 黄柯婷, 等. PI3K-Akt 信号通路调控慢性阻塞性肺疾病气道炎症反应的研究新进展[J]. 中国呼吸与危重监护杂志, 2022, 21(10): 751-755.
- [12] 廖君左, 蒋琳, 王城, 等. 孕鼠产前川芎嗪灌胃对先天性膈疝胎鼠肺组织发育及 Hippo 信号通路蛋白表达的影响[J]. 山东医药, 2023, 63(17): 38-41.
- [13] 鲁飞, 柳保丽, 张万宏, 等. ITGA4 基因在脑胶质瘤中的表达及临床意义[J]. 中国临床神经外科杂志, 2023, 28(3): 184-187+215.
- [14] Takeuchi, A. and Saito, T. (2017) CD4 CTL, a Cytotoxic Subset of CD4⁺ T Cells, Their Differentiation and Function. *Frontiers in Immunology*, **8**, Article 194. <https://doi.org/10.3389/fimmu.2017.00194>
- [15] 王宇欣, 周慈, 王和静, 等. 尘肺患者外周血胸腺细胞分化抗原-1 DNA 甲基化的改变及意义[J]. 环境与职业医学, 2019, 36(3): 242-247.
- [16] Yang, J., Zhan, X.Z., Malola, J., et al. (2020) The Multiple Roles of Thy-1 in Cell Differentiation and Regeneration. *Differentiation*, **113**, 38-48. <https://doi.org/10.1016/j.diff.2020.03.003>
- [17] Nishioka, Y., Masuda, S., Tomaru, U., et al. (2018) CD1d-Restricted Type II NKT Cells Reactive with Endogenous Hydrophobic Peptides. *Frontiers in Immunology*, **9**, Article 548. <https://doi.org/10.3389/fimmu.2018.00548>
- [18] Li, G., Wang, Y. and Cheng, Y. (2022) IL2RB Is a Prognostic Biomarker Associated with Immune Infiltrates in Pan-Cancer. *Journal of Oncology*, **2022**, Article ID: 2043880. <https://doi.org/10.1155/2022/2043880>
- [19] 朱帅, 祖玲玲, 徐嵩. RTN1 在肺腺癌中表达及对免疫微环境的影响[J]. 中国肺癌杂志, 2022, 25(6): 385-395.