

基于生物信息学分析的骨肉瘤 hub 基因的识别与免疫机制的研究

冯磊 孙飞飞 钟伟^{通讯作者}

山东第二医科大学附属医院，山东省潍坊市，261042；

摘要：骨肉瘤是一种具有高度侵袭性和复发率的恶性骨肿瘤，其转移性和耐药性显著降低了现有治疗手段的有效性。尽管手术、化疗和放疗一定程度上改善了患者的生存率，但阐明骨肉瘤的发病机制、探索精准诊断和治疗策略仍具有重要临床意义。本研究构建了受试者工作特征曲线（ROC），评估了这些基因的诊断效能。此外，通过 CIBERSORT 算法分析骨肉瘤免疫图谱，探讨了核心基因与免疫细胞浸润之间的关联，深入解析了骨肉瘤免疫微环境的调控机制。研究表明，这些核心基因在骨肉瘤的发生、侵袭和转移过程中发挥重要作用，并与特定免疫细胞类型显著相关。本研究不仅为骨肉瘤的精准诊断和治疗提供了潜在的分子标志物，还为靶向治疗策略的开发和免疫治疗方案的优化提供了理论依据。这些发现有助于推动骨肉瘤研究从分子机制到临床应用的转化进程，为改善患者预后和生存质量提供了新的研究方向。

关键词：生物信息学分析；骨肉瘤；hub 基因；免疫机制；差异表达基因

DOI:10.69979/3029-2808.24.11.023

前言

近年来，生物信息学技术的快速发展为骨肉瘤研究提供了新的视角。通过高通量测序技术和公共数据库资源，研究人员能够系统性地分析骨肉瘤的基因表达特征，并从中筛选出具有诊断和预后价值的关键基因。本研究旨在通过系统性分析，明确骨肉瘤的关键分子标志物及其免疫调节机制，为骨肉瘤的精准诊断、预后评估和治疗策略优化提供理论依据。这些发现将有助于推动骨肉瘤研究从分子机制到临床应用的转化进程，为改善患者预后和生存质量提供新的研究方向。

1 骨肉瘤研究背景与研究目标

骨肉瘤（Osteosarcoma, OS）是最常见的原发性恶性骨肿瘤，主要起源于骨髓中的间充质细胞，具有高度侵袭性和转移倾向。该肿瘤多发于四肢长骨干骺端，其特征是肿瘤细胞能够直接形成骨或类骨组织，快速破坏周围组织，导致局部侵袭和远处转移。尽管近年来手术切除结合全身化疗和放疗的治疗方案显著提高了患者的远期生存率（65%~70%），但骨肉瘤的复发率仍然较高，成为患者临床死亡的主要原因之一。骨肉瘤的复杂性和不稳定性，以及对其发生机制的不完全了解，已经成为治疗进展的主要障碍。因此，深入研究骨肉瘤的分子机制，寻找新的治疗靶点显得尤为重要^[1]。

本研究的目标是通过生物信息学方法，从 GEO 数据

库中获取骨肉瘤转录组数据（GSE16088 和 GSE36001），结合差异基因分析和加权基因共表达网络分析（WGCNA），筛选出与骨肉瘤发生发展密切相关的 hub 基因。随后，通过基因本体论（GO）和京都基因组百科全书（KEGG）数据库进行功能富集分析，揭示这些基因在骨肉瘤中的生物学作用。同时，利用受试者工作特征曲线（ROC）评估这些基因的诊断效能，并结合 CIBERSORT 算法分析骨肉瘤免疫图谱，探讨核心基因与免疫细胞浸润之间的关联，深入解析骨肉瘤免疫微环境的调控机制。

2 研究方法 with 数据处理

2.1 数据来源与预处理

本研究的数据来源于 GEO（Gene Expression Omnibus）数据库，该数据库是 NCBI（美国国家生物技术信息中心）维护的公共基因表达数据库，整合了来自多种实验平台的基因表达数据，为研究者提供了丰富的资源。为了确保数据的可靠性和代表性，我们从 GEO 数据库中筛选并下载了两个骨肉瘤相关的转录组数据集：GSE16088 和 GSE36001。

数据集 GSE16088 包含 3 个正常对照样本和 14 个骨肉瘤患者样本，数据来源于 Affymetrix Human Genome U133 Plus 2.0 microarray 平台。数据集 GSE36001 则包括 6 个正常对照样本和 19 个骨肉瘤患者样本，数据来源于 Agilent-019888 Human 1A array platform。这

两个数据集均为公开发表的骨肉瘤研究数据，且样本量适中，具备较高的数据质量，能够满足后续分析的需求^[2]。

在数据预处理阶段，首先对原始数据进行质量控制和标准化处理。通过使用 R 语言中的 GEOquery 包从 GEO 数据库下载数据，并利用 limma 包进行标准化处理，以消除不同芯片间的系统性偏差^[3]。随后，我们对数据进行了批次效应校正。鉴于两个数据集分别来源于不同的实验平台和批次，批次效应可能对后续分析结果产生显著影响。因此，我们采用 CombatSeq 算法对数据进行批次效应校正，以确保不同批次数据的可比性。

为了进一步提高数据的可靠性和准确性，我们对预处理后的数据进行了缺失值处理。对于缺失的表达值，采用 k-近邻 (k-nearest neighbors, KNN) 方法进行插值填充。通过这些预处理步骤，确保了数据的完整性和一致性，为后续的差异基因分析和加权基因共表达网络分析 (WGCNA) 奠定了坚实的基础。

通过对两个数据集的整合分析，我们能够更全面地揭示骨肉瘤的基因表达特征，从而为筛选关键基因和分析其生物学功能提供可靠的数据支持。这些预处理步骤不仅保证了数据的高质量，也为后续研究结果的准确性和可信性提供了重要保障。

2.2 生物信息学分析方法

为了系统性地解析骨肉瘤的分子机制及其免疫调节机制，本研究采用了多种生物信息学分析方法，包括差异基因分析 (DEGs)、加权基因共表达网络分析 (WGCNA)、基因本体论 (GO) 富集分析、京都基因组百科全书 (KEGG) 通路分析、受试者工作特征曲线 (ROC) 分析以及 CIBERSORT 算法分析免疫图谱。这些方法的综合应用为筛选关键基因、揭示其功能及其在骨肉瘤免疫微环境中的作用提供了全面的分析框架。

差异基因分析是筛选骨肉瘤特异性基因表达模式的重要方法。我们使用 R 语言中的 limma 包对两个 GEO 数据集进行标准化处理和差异基因筛选。通过设定严格的筛选标准 (p 值 < 0.05 ，倍数变化 ≥ 2)，我们从两个数据集中分别识别出显著差异表达的基因，并通过 Venn 图分析筛选出两个数据集中共有的差异基因，从而确保了结果的可靠性和一致性。差异基因分析不仅能够揭示骨肉瘤与正常对照样本之间的基因表达差异，还为后续的网络分析提供了候选基因列表^[4]。

加权基因共表达网络分析 (WGCNA) 被用于构建基因共表达网络，并识别与骨肉瘤密切相关的基因模块。通过计算基因表达的相关性矩阵，我们构建了基因共表达网络，并通过动态剪切树算法将基因划分为不同的模

块。随后，通过模块-样本相关性分析，筛选出与骨肉瘤显著相关的基因模块。进一步结合差异基因分析结果，我们从这些模块中筛选出关键的 hub 基因，这些基因在骨肉瘤的发生发展中可能发挥重要作用。

基因本体论 (GO) 分析和京都基因组百科全书 (KEGG) 通路分析用于揭示这些关键基因的生物学功能及其参与的信号通路。通过 DAVID 工具对筛选出的核心基因进行 GO 功能富集分析，我们能够明确这些基因在细胞增殖、侵袭、迁移以及骨代谢等生物学过程中的潜在作用。同时，KEGG 通路分析则揭示了这些基因在骨肉瘤相关信号通路中的潜在调控机制，如 PI3K-AKT、MAPK 和 Wnt 信号通路。这些分析结果为理解核心基因的生物学功能及其在骨肉瘤中的作用机制提供了重要的理论依据。

为了评估筛选出的核心基因的临床诊断价值，我们构建了受试者工作特征曲线 (ROC) 分析模型。通过计算每个基因的曲线下面积 (AUC)，评估其对骨肉瘤患者的诊断效能。ROC 分析不仅能够验证基因的诊断价值，还为后续的临床应用提供了重要的参考依据。

为了深入解析骨肉瘤的免疫微环境特征及其与核心基因的关系，我们采用了 CIBERSORT 算法对骨肉瘤样本中的免疫细胞浸润情况进行分析。通过计算 22 种免疫细胞类型的丰度，并结合核心基因的表达水平，我们探讨了这些基因与特定免疫细胞类型之间的关联。这不仅能够揭示核心基因在骨肉瘤免疫调节中的潜在作用，还为理解骨肉瘤免疫逃逸机制提供了新的视角^[5]。

3 研究结果与分析

3.1 差异表达基因筛选与功能富集分析

通过对 GSE16088 和 GSE36001 两个骨肉瘤转录组数据集的整合分析，我们首先进行了差异表达基因 (DEGs) 的筛选。基于 limma 算法，分别从 GSE16088 和 GSE36001 数据集中筛选出显著差异表达的基因，并通过 Venn 图分析确定了两个数据集中共有的 11 个核心基因。这些基因在骨肉瘤组织中表现出显著的上调或下调表达，提示其可能在骨肉瘤的发生和发展中发挥重要作用。

随后，我们对这些核心基因进行了基因本体论 (GO) 功能富集分析和京都基因组百科全书 (KEGG) 通路分析。GO 分析结果显示，这些基因主要富集在细胞增殖、迁移、侵袭和骨代谢相关的生物学过程中。例如，部分基因在细胞外基质重塑和胶原蛋白合成中具有重要作用，提示其可能通过调控细胞外基质的组成影响骨肉瘤细胞的侵袭和转移能力。此外，KEGG 通路分析表明，这些基因显著富集于 PI3K-AKT、MAPK 和 Wnt 信号通路，这些通路在骨肉瘤的细胞增殖、存活和侵袭中具有关键作用。例如，PI3K-AKT 通路的异常激活已被证实与骨肉瘤的转

移和耐药性密切相关。

为了进一步验证这些核心基因的临床诊断价值,我们构建了受试者工作特征曲线(ROC),评估了每个基因在区分骨肉瘤患者与正常对照样本中的诊断效能。结果显示,这些基因的曲线下面积(AUC)值均较高,表明其具有较高的诊断准确性。此外,结合 CIBERSORT 算法分析骨肉瘤免疫图谱,我们发现这些核心基因与特定免疫细胞类型(如 M2 型巨噬细胞和 CD8⁺ T 细胞)的浸润水平显著相关,提示其可能通过调控免疫微环境参与骨肉瘤的免疫逃逸机制。

本研究筛选出的 11 个核心基因不仅在骨肉瘤的分子机制中具有重要作用,还与其免疫微环境的调控密切相关。这些发现为骨肉瘤的精准诊断和治疗提供了潜在的分子标志物和靶点,同时也为后续实验验证和临床应用奠定了基础。

3.2 免疫图谱分析与 hub 基因验证

通过对骨肉瘤样本的免疫图谱分析,我们进一步探讨了核心基因在骨肉瘤免疫微环境中的潜在作用及其与免疫细胞浸润的关系。利用 CIBERSORT 算法,我们对 GSE16088 和 GSE36001 数据集中的骨肉瘤样本进行了免疫细胞类型丰度的计算,识别出 22 种免疫细胞在骨肉瘤组织中的浸润情况。分析结果显示,骨肉瘤组织中 M0 巨噬细胞和 M2 巨噬细胞的浸润水平显著较高,而 CD8⁺ T 细胞和休眠期 CD4⁺ 记忆性 T 细胞的浸润水平相对较低。这一结果提示,骨肉瘤的免疫微环境中可能存在免疫抑制性环境,从而促进肿瘤的逃逸和进展。

进一步分析发现,核心基因(如 P4HA1 和 ITGAV)与特定免疫细胞类型的浸润水平显著相关。P4HA1 作为胶原蛋白合成的关键酶,其上调表达不仅与骨肉瘤的侵袭和转移密切相关,还与 M2 型巨噬细胞的浸润水平呈正相关。M2 型巨噬细胞在肿瘤微环境中通常表现出促肿瘤特性,能够分泌多种促血管生成和免疫抑制因子,从而促进肿瘤细胞的增殖和转移。P4HA1 的高表达可能通过调控细胞外基质的重塑,间接促进 M2 型巨噬细胞的募集和功能活化,从而形成一个促肿瘤的免疫微环境。

ITGAV 基因的表达水平与 CD8⁺ T 细胞的浸润水平呈负相关。CD8⁺ T 细胞是抗肿瘤免疫反应的重要效应细胞,其浸润不足可能与肿瘤免疫逃逸机制有关。ITGAV 基因编码的整合素 αV 蛋白在肿瘤血管生成和细胞迁移中发挥重要作用,其高表达可能通过调节肿瘤内血管微环境,抑制 CD8⁺ T 细胞的浸润和活性,从而削弱宿主的抗肿瘤免疫应答。

为了进一步验证这些核心基因的临床价值,我们结

合 ROC 分析评估了它们的诊断效能。结果显示, P4HA1 和 ITGAV 基因在区分骨肉瘤患者与正常对照样本中的曲线下面积(AUC)均较高,分别为 0.85 和 0.82,表明这些基因具有较高的诊断准确性。同时,这些基因的表达水平与患者的临床特征(如肿瘤分期和预后)也表现出显著相关性,提示其可能作为潜在的诊断和预后标志物。

综合以上分析,本研究筛选出的核心基因不仅在骨肉瘤的发生和发展中发挥关键作用,还通过调控免疫微环境中的免疫细胞浸润,参与肿瘤免疫逃逸机制。这些发现为骨肉瘤的免疫治疗提供了新的研究方向,例如针对 P4HA1 和 ITGAV 的靶向治疗可能能够调节肿瘤免疫微环境,增强抗肿瘤免疫应答,从而改善患者的治疗效果和生存质量。

4 结束语

本研究通过生物信息学方法系统性地筛选和验证了骨肉瘤的关键基因及其免疫调节机制,为骨肉瘤的分子机制研究、精准诊断和治疗策略优化提供了重要的理论依据。未来的研究需要结合实验验证和多组学技术,进一步深化对骨肉瘤分子机制和免疫微环境的理解,为改善患者的预后和生存质量提供新的研究方向。

参考文献

- [1] 沈家亮,王琳,尚文强. 骨肉瘤患者癌组织中 AHA1 mRNA 和 LOXL2 mRNA 表达与侵袭转移基因 mRNA 表达的相关性及临床意义[J]. 现代检验医学杂志, 2024, 39(02): 39-45.
- [2] Chen C, Xie L, Ren T, Huang Y, Xu J, Guo W. Immunotherapy for osteosarcoma: Fundamental mechanism, rationale, and recent breakthroughs. *Cancer letters* 2021 Mar 1;500:1-10.
- [3] Xiao X, Wang W, Li Y, Yang D, Li X, Shen C, Liu Y, Ke X, Guo S, Guo Z. HSP90AA1-mediated autophagy promotes drug resistance in osteosarcoma. *Journal of experimental & clinical cancer research* : CR 2018 Aug 28;37(1):201.
- [4] Sheng G, Gao Y, Yang Y, Wu H. Osteosarcoma and Metastasis. *Frontiers in oncology* 2021;11: 780264.
- [5] 魏宇. 骨肉瘤患者 MTHFR C677T 基因多态性与 HD-MTX 不良反应的相关性[J]. 沈阳药科大学学报, 2022, 39(11): 1368-1373. DOI: 10.14066/j.cnki.cn21-1349/r.2021.1435.