

乳腺癌和肺腺癌表达数据筛选骨癌痛潜在致痛基因

刘小旦

(北京大学基础医学院病理学系 北京大学第三医院病理科, 北京 100191)

摘要 目的: 通过乳腺癌和肺腺癌表达数据分析, 筛选出骨癌痛潜在的致痛基因。**方法:** 采用生物信息学技术分析 TCGA 数据库中乳腺癌和肺腺癌的基因表达数据, 并通过与人类分泌组蛋白基因比对, 获得乳腺癌和肺腺癌中表达趋势一致的分泌组蛋白差异基因。随后通过 KEGG (Kyoto Encyclopedia of Genes and Genomes) 通路富集分析这些差异基因的所在通路。**结果:** 和正常样本相比, 乳腺癌样本中有 1422 个基因表达上调, 2124 个基因表达下调; 肺腺癌中 1106 个基因表达上调, 3133 个基因表达下调。其中, 86 个分泌组蛋白对应基因在乳腺癌和肺腺癌中一致上调, 而 244 个分泌组蛋白对应基因在乳腺癌和肺腺癌中一致下调。对这 330 个差异基因进行 KEGG 富集分析, 发现了细胞因子-细胞因子受体相互作用等 20 条通路具有显著差异。随后, 通过文献回顾发现细胞因子 CXCL、IL-6 及 TNF- α 等与骨癌痛的发生有直接的关系。本研究中发现的基质金属蛋白酶家族多个成员表达明显上调, 这些蛋白对于肿瘤的骨转移以及骨质破坏和重塑发挥着重要作用, 间接参与了骨癌痛的发生。**结论:** 通过肿瘤表达数据挖掘, 本研究筛选出了数百个差异表达基因, 其中多个基因已有报道参与了骨癌痛的发生和维持, 其余可能是潜在的致痛基因。

关键词 骨癌痛; 乳腺癌; 肺腺癌; 生物信息学

Screening for potential pain genes in breast cancer and lung adenocarcinoma using bioinformatics LIU Xiaodan

(Department of Pathology, School of Basic Medical Sciences, Peking University Third Hospital, Peking University Health Science Center, Beijing 100191, China)

Abstract Objective: To screen for potential pain causing genes by analyzing the gene expression data of breast cancer and lung adenocarcinoma using bioinformatics. **Methods:** By analyzing the gene expression data of breast cancer and lung adenocarcinoma in the TCGA database using bioinformatics, and then comparing with human secretome genes, we obtained the secretome genes that were consistent between breast cancer and lung adenocarcinoma. Then, the pathways of these differentially expressed genes were analyzed by Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment. **Results:** Compared with normal samples, 1422 genes in breast cancer samples were up-regulated and 2124 genes were down-regulated. In lung adenocarcinoma, 1106 genes were up-regulated and 3133 genes were down-regulated. Among them, 86 secretome genes were both up-regulated in breast cancer and lung adenocarcinoma, while 244 secretome genes were both down-regulated. By KEGG pathway enrichment analysis of these 330 differentially expressed genes, there were 20 pathways such as cytokine-cytokine receptor interaction pathway that had changed significantly. Among them, CXCL, IL-6 and TNF- α et al. had been reported to contribute to the development of cancer-induced bone pain directly. What's more, increased expression of matrix metal proteinase family members found in our study had important roles in the metastasis of cancer and the remodeling of bone, which could contribute to the development of cancer-induced bone pain indirectly. **Conclusion:** By analyzing tumor gene expression data, we have screened out dozens of differentially expressed genes, in which many of them have been reported to be involved in the occurrence and maintenance of cancer induced bone pain, while the rest could be potential pain causing genes.

Keywords cancer induced bone pain; breast cancer; lung adenocarcinoma; bioinformatics

[△] 通信作者 刘小旦 liuxiaodan@bjmu.edu.cn

临床上很多常见的晚期肿瘤（乳腺癌、前列腺癌以及肺癌等）都会发生骨转移，并且 75% 骨转移病人经历着严重的疼痛（cancer-induced bone pain, 骨癌痛）^[1]。肿瘤在骨中生长的过程中能够导致疼痛、高钙血症、贫血、感染以及骨折等，严重影响了病人的生活质量^[2]。

对骨癌痛机制的研究表明，它是一种包含神经病理性疼痛和炎症性疼痛成分的混合疼痛状态，但对周围组织和神经的改变又明显不同于神经病理性疼痛和炎症性疼痛，在脊髓水平也有其独特的神经化学变化^[2]。骨髓、矿化骨和骨膜受神经纤维的高度支配，当癌细胞侵入并开始增殖时，初级传入纤维的机械损伤、扩张和压迫会导致神经病理性疼痛^[3]。肿瘤是由肿瘤细胞以及间充质细胞组成的，后者包括巨噬细胞、中性粒细胞、T 细胞、成纤维细胞和内皮细胞等。肿瘤以及间充质细胞能够分泌各种生长因子、细胞因子、白细胞介素、趋化因子、前列腺素、内皮素和其他有助于疼痛发生或维持的炎症介质。这些炎症介质可直接或间接激活初级传入纤维或者脊髓背角上的伤害性感受神经元，从而导致骨癌痛的发生^[4-17]。临床上超过一半的转移性癌痛病人报告，目前的药物疗法不能充分缓解疼痛^[18]，说明癌痛有其独特的一面。因此，目前还迫切需要开发新的镇痛疗法，为癌症病人提供缓解疼痛和改善生活质量的姑息治疗。这就需要研究人员聚焦到肿瘤本身，特别是肿瘤分泌释放的一系列外分泌蛋白上，因为它们是肿瘤和伤害性感受器相互作用的桥梁。

肿瘤基因组学、蛋白组学、表观遗传学等的研究产生了大量的突变和基因表达等数据，这些数据大多储存在美国国家癌症研究所 TCGA (The Cancer Genome Atlas) 数据库中。本研究将运用生物信息学技术，挖掘 TCGA 数据库中乳腺癌和肺腺癌分泌组蛋白的基因表达情况，以期能够筛选出肿瘤新的致痛分子，同时探讨采用生物信息学技术对组学数据挖掘从而筛选新的分子靶点的可行性。

方 法

1. 研究方案及数据获取

人乳腺癌和肺腺癌表达基因数据分别来源于 TCGA 数据库 (<https://portal.gdc.cancer.gov>) 中 BRCA 和 LUAD 队列，其差异表达基因数据通过 GEPIA2 (<http://gepia2.cancer-pku.cn>) 数据库进行分析、下载获取。Log₂ (fold change) 绝对值和 q-value 值分别设

置为 1 和 0.05。人分泌组 (Secretome) 蛋白基因列表数据从人类蛋白数据集 (<https://www.proteinatlas.org>) 网站获取，总计 2793 个基因。研究方案见图 1。

2. 数据分析

采用 R studio 软件 (v1.3.1093)，通过 R 语言进行数据处理。主要使用 Tidyverse 和 limma 包进行数据分析，随后使用 EnhancedVolcano 包进行火山图绘制。采用 clusterProfiler 包进行 KEGG 通路富集分析，不同数据的交集韦恩图使用 VennDiagram 包进行绘制。

结 果

1. 人乳腺癌和肺腺癌差异表达基因结果

首先，通过 GEPIA2 网站对来源于 TCGA 数据库的人类乳腺癌 (BRCA) 和肺腺癌 (LUAD) 基因表达数据进行分析。和正常样本相比，乳腺癌样本中有 1422 个基因表达上调，2124 个基因表达下调；肺腺癌中 1106 个基因表达上调，3133 个基因表达下调。在表达上调的基因中，乳腺癌和肺腺癌表达变化一致的基因有 578 个；而在表达下调的基因中，乳腺癌和肺腺癌表达变化一致的基因有 1349 个（见图 2）。

2. 人乳腺癌和肺腺癌中分泌组蛋白的基因表达结果

随后，将乳腺癌和肺腺癌中表达变化一致的基因（上调或下调）分别与人分泌蛋白组对应基因进行比对，发现 86 个分泌蛋白对应基因在乳腺癌和肺腺癌中一致上调，而 244 个分泌蛋白对应基因在乳腺癌和肺腺癌中一致下调（见图 2）。将这 330 个基因及其在乳腺癌中的表达数据作为分泌组蛋白差异基因表达数据进行后续分析。火山图展示了这 330 个基因在乳腺癌中的表达差异及其显著性（见图 3）。另外，表 1 展示了乳腺癌中前 20 个上调最

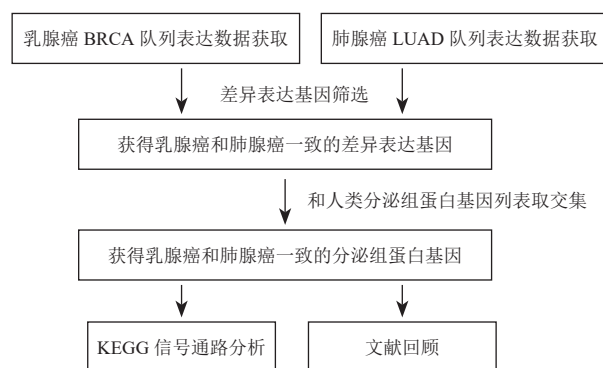


图 1 研究方案

Fig. 1 Research approach

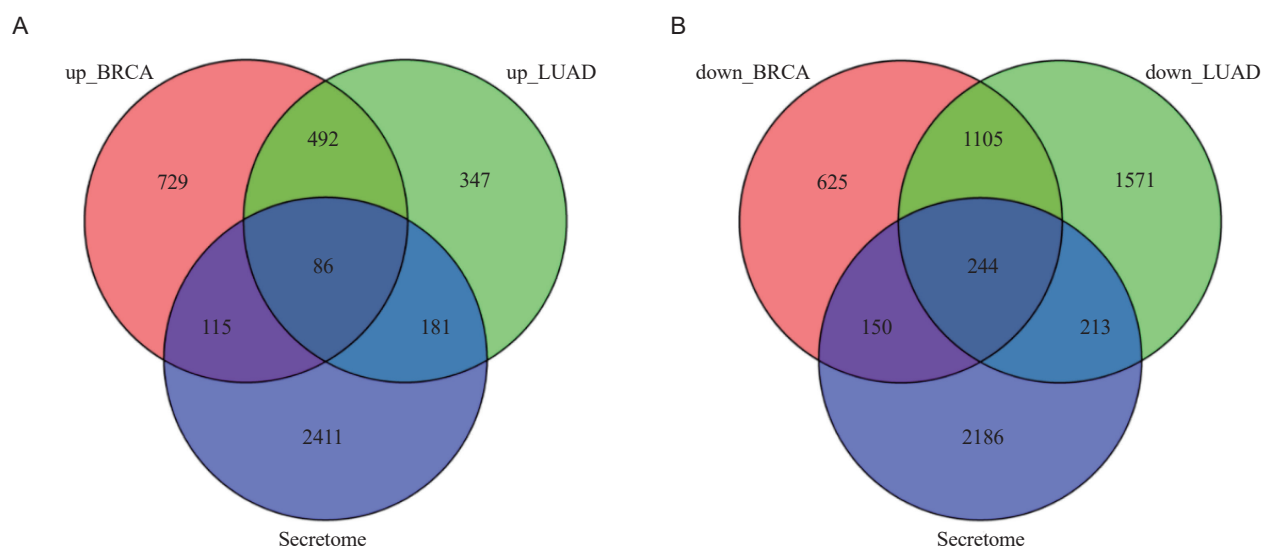


图2 乳腺癌和肺腺癌差异表达基因数量韦恩图

(A) 表达上调基因; (B) 表达下调基因。红色代表乳腺癌, 绿色代表肺腺癌, 紫色代表分泌组

Fig. 2 Venn diagrams of differentially expressed genes in breast and lung adenocarcinoma

(A) Up-regulated genes; (B) Down-regulated genes. Red represents breast cancer, green represents lung cancer and purple represents the secretome.

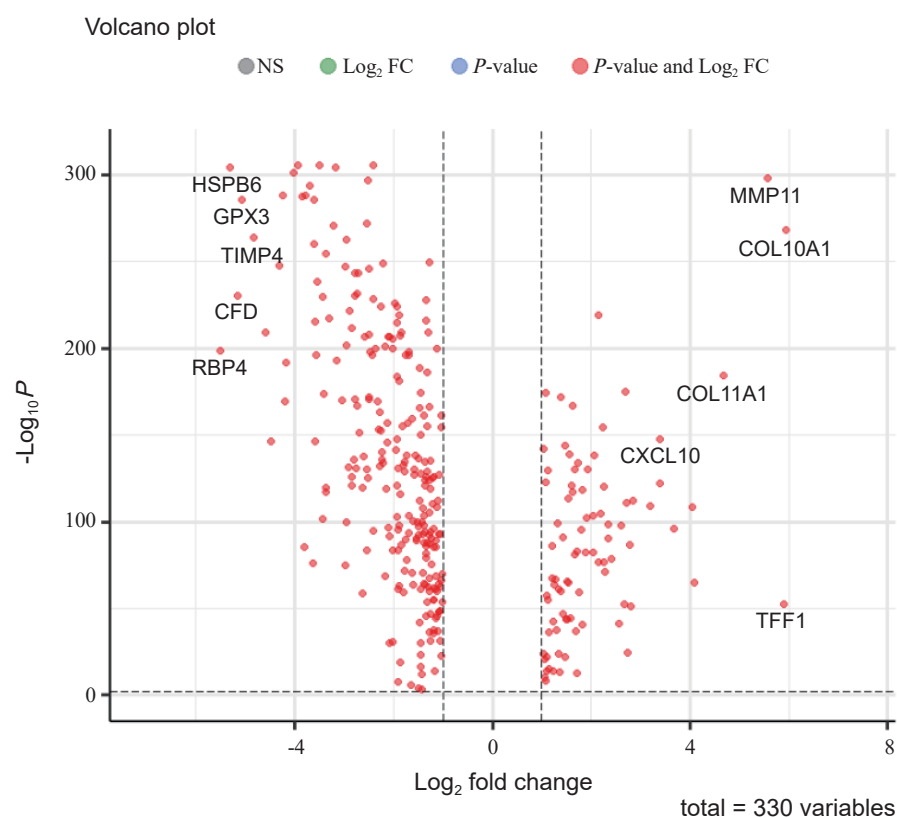


图3 乳腺癌中 330 个分泌蛋白组基因表达情况

横坐标为差异倍数的对数值 (以 2 为底), 纵坐标为校正后 P 值的对数值 (以 10 为底取负值)。图上选择性标记了前 10 个表达差异最大的基因。 $\log_2$ fold change 绝对值 < 1 或校正后 $P > 0.05$ 的基因未展示。

Fig. 3 330 differentially expressed genes in breast cancer

X axis shows $\log_2$ (fold change), while Y axis shows $(-\log_{10} P)$. Top 10 genes were selectively labeled, and the genes with $|\log_2(\text{fold change})| < 1$ or adjusted $P > 0.05$ were not shown.

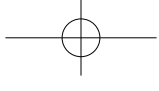


表 1 乳腺癌中前 20 个上调最明显外分泌蛋白组基因
Table 1 Top 20 most significantly upregulated genes of secretome in breast cancer

	基因名 Gene symbol	基因全称 Gene title	中位数 (肿瘤) Median (Tumor)	中位数 (正常组织) Median (Normal)	log ₂ FC	adj.P.value
1	COL10A1	collagen, type X, alpha 1	67.461	0.12	5.934	7.31E-269
2	TFF1	trefoil factor 1	133.796	1.27	5.892	1.98E-53
3	MMP11	matrix metalloproteinase 11	167.729	2.56	5.567	7.49E-299
4	COL11A1	collagen, type XI, alpha 1	37.131	0.5	4.668	9.11E-185
5	AGR2	anterior gradient 2	376.524	21.23	4.086	1.32E-65
6	CST1	cystatin SN	15.49	0	4.044	3.34E-109
7	IGHG4	Immunoglobulin Heavy Constant Gamma 4	68.071	4.39	3.68	1.46E-96
8	COMP	cartilage oligomeric matrix protein	44.47	3.32	3.396	9.04E-123
9	CXCL10	chemokine (C-X-C motif) ligand 10	25.419	1.53	3.384	1.24E-148
10	MMP9	matrix metalloproteinase 9	29.67	2.34	3.199	6.75E-110
11	COL1A1	collagen, type I, alpha 1	1055.93	146.017	2.846	4.89E-113
12	IGHG1	immunoglobulin heavy constant gamma 1	626.381	89.077	2.8	6.68E-52
13	CXCL9	chemokine (C-X-C motif) ligand 9	16.63	1.58	2.773	1.45E-87
14	TFF3	trefoil factor 3	297.737	43.949	2.733	4.63E-25
15	MMP13	matrix metalloproteinase 13 (collagenase 3)	5.71	0.02	2.718	6.85E-112
16	SDC1	syndecan 1	128.141	19.061	2.687	5.63E-176
17	IGHG3	Immunoglobulin Heavy Constant Gamma 3	79.781	11.69	2.67	5.11E-53
18	POSTN	periostin, osteoblast specific factor	404.921	65.48	2.61	8.58E-99
19	CXCL13	chemokine (C-X-C motif) ligand 13	6.93	0.35	2.554	5.01E-42
20	SPP1	secreted phosphoprotein 1	135.918	24.781	2.409	2.03E-79

明显的外分泌蛋白基因。其中包含胶原蛋白家族 3 个成员 (COL10A1、COL11A1 和 COL1A1)、基质金属蛋白酶家族 3 个成员 (MMP11/9/13)、趋化因子 CXCL 家族 3 个成员 (CXCL10/9/13) 以及免疫球蛋白 γ 重链 3 个成员 (IGHG4/1/3)。

3. KEGG 通路富集分析结果

随后对这 330 个分泌组蛋白差异表达基因进行 KEGG 通路富集分析, 发现 20 条通路其基因表达改变有显著性差异 (校正后 $P < 0.05$), 包括细胞因子-细胞因子受体相互作用通路、PI3K-Akt 通路、白介素-17 (interleukin, IL-17) 信号通路、TNF 信号通路等 (见图 4A)。其中, 前 2 条信号通路差异基因最多 (各 26 个), 图 4B 展示了细胞因子及其受体相互作用通路 26 个基因的表达情况, 其中 CXCL 家族多个成员 (CXCL9/10/13) 表达呈明显上调。这其中, CXCL10、IL-6 及 TNF- α 等多个蛋白已被报道参与了骨癌痛的发生和维持。

讨 论

本研究采用生物信息学技术, 通过挖掘乳腺癌和肺腺癌分泌蛋白组基因表达的数据, 发现细胞因

子及其受体相互作用等 20 条通路其基因表达具有显著差异。其中 CXCL 家族多个成员 (CXCL9/10/13) 表达呈明显上调。这与文献报道的骨癌痛模型中 CXCL10 表达明显上调一致^[11]。CXCL9/10/13 的共同受体是 CXCR3, 研究发现 CXCR3 的激活参与了骨癌痛的发生^[19,20]。另外, 也有文献报道了 CXCL1、CXCL12 等家族其他成员参与了骨癌痛的发生和维持^[10,21]。白介素也是一类重要的致痛物质, 多项研究表明 IL-1 β ^[15]、IL-6^[9,22]、IL-10^[23] 以及 IL-18^[7] 等参与了骨癌痛的发生和维持。本研究筛选出白介素家族的 2 个成员 (IL-6 和 IL-33) 在肿瘤中的表达下调。需要注意的是, 本研究中组学数据筛选基于转录组测序, 也就是 mRNA 的表达情况, 与真实的蛋白表达可能不太一致, 也有可能是因为 IL-6 和 IL-33 的大量翻译导致 mRNA 的耗竭造成的。

在本研究的通路富集结果中, PI3K-Akt 通路同样有 26 个基因表达差异显著。既往的研究表明, 骨癌痛中初级伤害性感受神经元^[9]、脊髓背角^[19,24,25]、中脑导水管周围灰质^[26] 等多个部位均有 PI3K-Akt 通路的激活, 说明该通路的激活对于骨癌痛的发生和维持是至关重要的。PI3K-Akt 通路是多种细胞因子 (如 CXCL10、IL-6 和 MCP-1) 与其受体相互作

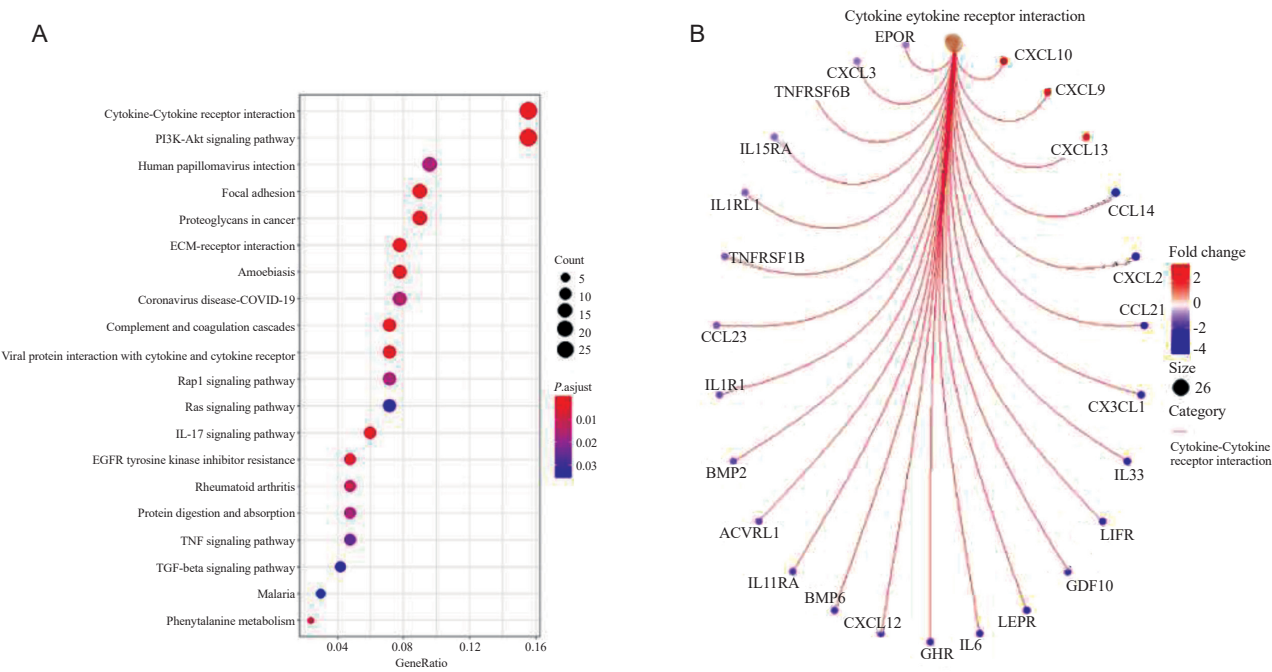


图 4 KEGG 通路富集分析结果
(A) 表达差异有显著改变的 20 条通路。圆点的大小反映差异显著基因的数量，而圆点颜色反映校正后 P 值的大小；
(B) 细胞因子及其受体相互作用通路 26 个基因的表达情况。红色越深表示基因表达上调越大，蓝色越深代表基因表达下调越大。

Fig. 4 Results of KEGG pathway enrichment analysis
(A) 20 significantly changed pathways were shown. The size of dot represented the number of differentially expressed genes, while the color of dot represented adjusted P value; (B) The results of 26 genes in the cytokine-cytokine receptor interaction pathway. Red: up-regulation; Blue: down-regulation.

用的下游通路，该通路的改变与细胞因子及其受体相互作用通路的改变是一致的。

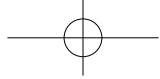
基质金属蛋白酶 (matrix metalloproteinases, MMPs) 是一类负责切割细胞外基质 (extracellular matrix, ECM) 蛋白质的主要蛋白酶家族。MMPs 在骨中的作用包括成骨细胞/骨细胞分化、骨形成、骨吸收过程中类骨的溶解、破骨细胞的募集和迁移，以及生理条件下骨的重塑^[27]。在本研究中，基质金属蛋白酶家族 3 个成员 (MMP11/9/13) 均在乳腺癌和肺腺癌中表达明显上调。它们的上调有利于骨微环境的重塑、肿瘤对骨的侵袭和转移^[28,29]。

本研究主要基于组学数据的筛选，缺乏后期基于动物实验数据的验证，这是本研究存在的主要不足。通过文献回顾，发现本研究所取得的主要发现可以在部分文献中得到验证，说明采用生物信息学技术挖掘组学数据可作为一种行之有效的目标基因筛选方法。本研究还筛选出其他几十个分泌蛋白组基因的表达明显上调，这些蛋白是否会参与癌痛的发生和维持需要更进一步的实验验证。

综上所述，通过乳腺癌和肺腺癌表达数据分析筛选出了一些潜在的致痛基因和信号通路，采用生物信息学筛选靶基因是一种简便、易行的方法。

参 考 文 献

- [1] Kane CM, Hoskin P, Bennett MI. Cancer induced bone pain[J]. BMJ, 2015, 350:h315.
- [2] Falk S, Dickenson AH. Pain and nociception: mechanisms of cancer-induced bone pain[J]. J Clin Oncol, 2014, 32 (16):1647-1654.
- [3] Mach DB, Rogers SD, Sabino MC, *et al.* Origins of skeletal pain: sensory and sympathetic innervation of the mouse femur[J]. Neuroscience, 2002, 113 (1):155-166.
- [4] Pezet S, McMahon SB. Neurotrophins: mediators and modulators of pain[J]. Ann Rev Neurosci, 2006, 29:507-538.
- [5] Sevcik MA, Ghilardi JR, Peters CM, *et al.* Anti-NGF therapy profoundly reduces bone cancer pain and the accompanying increase in markers of peripheral and central sensitization[J]. Pain, 2005, 115 (1-2):128-141.



- [6] Zhao D, Han DF, Wang SS, *et al.* Roles of tumor necrosis factor- α and interleukin-6 in regulating bone cancer pain via TRPA1 signal pathway and beneficial effects of inhibition of neuro-inflammation and TRPA1[J]. *Mol Pain*, 2019, 15:1744806919857981.
- [7] Liu S, Liu YP, Lv Y, *et al.* IL-18 contributes to bone cancer pain by regulating glia cells and neuron interaction[J]. *J Pain*, 2018, 19 (2):186-195.
- [8] Zhou YQ, Gao HY, Guan XH, *et al.* Chemokines and their receptors: potential therapeutic targets for bone cancer pain[J]. *Curr Pharm Des*, 2015, 21 (34):5029-5033.
- [9] Fang D, Kong LY, Cai J, *et al.* Interleukin-6-mediated functional upregulation of TRPV1 receptors in dorsal root ganglion neurons through the activation of JAK/PI3K signaling pathway: roles in the development of bone cancer pain in a rat model[J]. *Pain*, 2015, 156 (6):1124-1144.
- [10] Shen W, Hu XM, Liu YN, *et al.* CXCL12 in astrocytes contributes to bone cancer pain through CXCR4-mediated neuronal sensitization and glial activation in rat spinal cord[J]. *J Neuroinflamm*, 2014, 11:75.
- [11] Bu H, Shu B, Gao F, *et al.* Spinal IFN- γ -induced protein-10 (CXCL10) mediates metastatic breast cancer-induced bone pain by activation of microglia in rat models[J]. *Breast Cancer Res Tr*, 2014, 143 (2):255-263.
- [12] Bao Y, Hou W, Liu R, *et al.* PAR2-mediated upregulation of BDNF contributes to central sensitization in bone cancer pain[J]. *Mol Pain*, 2014, 10:28.
- [13] Han Y, Li Y, Xiao X, *et al.* Formaldehyde up-regulates TRPV1 through MAPK and PI3K signaling pathways in a rat model of bone cancer pain[J]. *Neurosci Bull*, 2012, 28 (2):165-172.
- [14] Schweizerhof M, Stöessner S, Kurejova M, *et al.* Hematopoietic colony-stimulating factors mediate tumor-nerve interactions and bone cancer pain[J]. *Nature Med*, 2009, 15 (7):802-807.
- [15] Zhang RX, Liu B, Li A, *et al.* Interleukin 1 beta facilitates bone cancer pain in rats by enhancing NMDA receptor NR-1 subunit phosphorylation[J]. *Neuroscience*, 2008, 154 (4):1533-1538.
- [16] 李琪琪, 林健清. 小胶质细胞活化参与骨癌痛机制的研究进展 [J]. 中国疼痛医学杂志, 2020, 26 (9): 685-688.
- [17] 杭黎华, 束薇薇, 罗红, 等. 骨癌痛大鼠脊髓背角趋化因子受体 5 的表达变化 [J]. 中国疼痛医学杂志, 2015, 21 (7): 495-498.
- [18] Ellingson HM, Vanderah TW. Potential therapeutic treatments of cancer-induced bone pain[J]. *Curr Opin Support Pa*, 2020, 14 (2): 107-111.
- [19] Guan XH, Fu QC, Shi D, *et al.* Activation of spinal chemokine receptor CXCR3 mediates bone cancer pain through an Akt-ERK crosstalk pathway in rats[J]. *Exp Neurol*, 2015, 263:39-49.
- [20] Guo G, Gao F. CXCR3: Latest evidence for the involvement of chemokine signaling in bone cancer pain[J]. *Exp Neurol*, 2015, 265:176-179.
- [21] Xu J, Zhu MD, Zhang X, *et al.* NF κ B-mediated CXCL1 production in spinal cord astrocytes contributes to the maintenance of bone cancer pain in mice[J]. *J Neuroinflamm*, 2014, 11:38.
- [22] Remeniuk B, King T, Sukhtankar D, *et al.* Disease modifying actions of interleukin-6 blockade in a rat model of bone cancer pain[J]. *Pain*, 2018, 159 (4):684-698.
- [23] Apriyani E, Ali U, Wang ZY, *et al.* The spinal microglial IL-10/ β -endorphin pathway accounts for cinobufagin-induced mechanical allodynia in bone cancer pain following activation of α 7-nicotinic acetylcholine receptors[J]. *J Neuroinflamm*, 2020, 17(1):75.
- [24] Guan X, Fu Q, Xiong B, *et al.* Activation of PI3K/ γ -Akt pathway mediates bone cancer pain in rats[J]. *J Neurochem*, 2015, 134 (3):590-600.
- [25] Jin D, Yang JP, Hu JH, *et al.* MCP-1 stimulates spinal microglia via PI3K/Akt pathway in bone cancer pain[J]. *Brain Res*, 2015, 1599:158-167.
- [26] Zhang J, Wang L, Wang H, *et al.* Neuroinflammation and central PI3K/Akt/mTOR signal pathway contribute to bone cancer pain[J]. *Mol Pain*, 2019, 15: 1744806919830240.
- [27] Paiva KBS, Granjeiro JM. Matrix Metalloproteinases in Bone Resorption, Remodeling, and Repair[J]. *Prog Mol Biol Transl Sci*, 2017, 148:203-303.
- [28] Zhu L, Tang Y, Li XY, *et al.* Osteoclast-mediated bone resorption is controlled by a compensatory network of secreted and membrane-tethered metalloproteinases[J]. *Sci Transl Med*, 2020, 12(529): eaaw6143.
- [29] Yoneda T, Sasaki A, Dunstan C, *et al.* Inhibition of osteolytic bone metastasis of breast cancer by combined treatment with the bisphosphonate ibandronate and tissue inhibitor of the matrix metalloproteinase-2[J]. *J Clin Investigation*, 1997, 99 (10):2509-2517.