

doi: 10.3978/j.issn.2095-6959.2020.02.001
View this article at: <http://dx.doi.org/10.3978/j.issn.2095-6959.2020.02.001>

· 论著 ·

甲状腺乳头状癌颈部淋巴结转移相关基因的生物信息学分析

侯建忠，樊友本，伍波

(上海交通大学附属第六人民医院普外科，上海 200233)

[摘要] 目的：利用生物信息学方法分析甲状腺乳头状癌(papillary thyroid carcinoma, PTC)颈部淋巴结转移的关键基因。方法：从美国国立生物技术信息中心(National Center for Biotechnology Information, NCBI)公共数据平台高通量基因表达(Gene Expression Omnibus, GEO)数据库检索得到PTC的基因芯片GSE60542，使用GEO2R在线分析工具分别筛选PTC和瘤旁组织的差异表达基因(差异表达基因集A)以及PTC与颈部淋巴结转移癌组织之间的差异表达基因(差异表达基因集B)。进一步筛选出差异表达基因集A和差异表达基因集B交集的差异表达基因集C。通过生物信息学工具David, String, cBioPortal对差异表达基因集C进行生物学功能及其编码蛋白的互作分析，并对差异表达基因集C进行生存分析。结果：通过分析GSE60542芯片数据，一共获得1 086个差异表达基因A、194个差异表达基因B、39个差异表达基因C。基因本体(Gene Ontology, GO)分析显示：差异表达基因集C主要的分子功能与氧化还原酶活性、蛋白质同源二聚化活性、生长因子结合有关；这些基因主要参与细胞增殖、甲状腺激素的产生、神经细胞凋亡过程的负调节、 γ -氨基丁酸信号转导途径等生物学过程。通过cBioPortal在线分析网站，初步筛选出对患者生存时间影响较大的基因CCL21。结论：基因芯片结合生物信息学方法能够有效分析PTC组织、瘤旁组织以及颈部淋巴结转移癌组织间差异表达基因，并筛选出PTC颈部淋巴结转移相关的关键基因，为进一步研究PTC颈部淋巴结转移的分子机制提供一定的指导。

[关键词] 甲状腺乳头状癌；颈部淋巴结转移；生物信息学

Bioinformatics analysis for genes related to cervical lymph node metastasis of papillary thyroid carcinoma

HOU Jianzhong, FAN Youben, WU Bo

(Department of General Surgery, Sixth People's Hospital Affiliated to Shanghai Jiao Tong University, Shanghai 200233, China)

Abstract **Objective:** Bioinformatics methods were used to analyze key genes in cervical lymph node metastasis of papillary thyroid carcinoma (PTC). **Methods:** Microarray data of PTC GSE60542 was searched from Gene Expression Omnibus (GEO) database of National Center for Biotechnology Information (NCBI). The GEO2R online analysis tool was used to screen differentially expressed genes between PTC and paracarcinoma tissues (differentially expressed gene set A) and differentially expressed genes between PTC and cervical metastatic

收稿日期 (Date of reception): 2019-05-13

通信作者 (Corresponding author): 伍波, Email: wubo7421@sohu.com

基金项目 (Foundation item): 国家自然科学基金(81670718)。This work was supported by the National Natural Science Foundation of China (81670718).

lymph nodes (differentially expressed gene set B). The differentially expressed gene set C of the intersection of differentially expressed gene set A and differentially expressed gene set B was further screened. And bioinformatics tools including David, String and cBioPortal were applied to analyze the biological function of differentially expressed gene set C and their encoded proteins interaction. Then, we performed survival analysis on key genes. **Results:** By analyzing the GSE60542 microarray data, a total of 1 086 differentially expressed gene A, 194 differentially expressed gene B, and 39 differentially expressed gene C were obtained. GO analysis showed that these genes are mainly involved in the molecular function of oxidoreductase activity, protein homodimerization activity, growth factor binding and involved in the biological processes of cell proliferation, thyroid hormone generation, negative regulation of neuron apoptotic process, gamma-aminobutyric acid signaling pathway. Through the online cBioPortal analysis website, the gene CCL21, which had a great influence on the survival time of patients, was initially screened. **Conclusion:** Microarray combined with bioinformatics methods can effectively analyze the differentially expressed genes between PTC tissues, paracarcinoma tissues, and cervical metastatic lymph node tissues, which may provide guidance for molecular mechanisms research of PTC pathogenesis.

Keywords papillary thyroid carcinoma; cervical lymph node metastasis; bioinformatics

2008—2012年我国甲状腺癌发病率为7.56/10万，病死率为0.52/10万，甲状腺癌的发病率和病死率分别居癌症发病率和病死率的第7位和第22位^[1]。最近10年中，甲状腺乳头状癌(papillary thyroid carcinoma, PTC)的增长率最快，其增长是甲状腺癌增长的主要因素，其他类型甲状腺癌构成比则逐年下降^[2]。甲状腺癌患者如若发生颈部淋巴结转移，可导致患者复发率增高和存活率降低^[3]。甲状腺癌相关分子标志物越来越引起人们的关注，灵敏度高、特异性强的肿瘤标志物有助于甲状腺癌的早期诊断，并为制订合理的治疗方案提供帮助^[4]。生物信息学是一门结合分子生物学与计算机信息技术的新型交叉学科，通过生物信息学分析能够高效地对研究所涉及的差异表达基因进行功能分析，并筛选出可能发挥作用的关键基因^[5]。

本研究采用生物信息学的方法筛选高通量基因表达(Gene Expression Omnibus, GEO)数据库基因芯片中PTC组织、癌旁组织以及颈部淋巴结转移癌组织差异表达基因，分析PTC颈部淋巴结转移相关差异表达基因的功能并进行关键基因的生存分析，以期为进一步研究PTC颈部淋巴结转移发生、发展的分子机制提供一定基础。

1 资料与方法

1.1 基因芯片的选取以及差异表达基因的筛选

本研究从美国国立生物技术信息中心(National Center for Biotechnology Information, NCBI)的GEO数据库(<https://www.ncbi.nlm.nih.gov/gds>)中

以“papillary thyroid carcinoma”为关键词进行检索，并筛选样本中包含PTC组织、癌旁组织以及颈部淋巴结转移癌组织对照的系列，检索时间截至2019年4月。从GEO检索到1个基因表达数据集GSE60542。

1.2 差异表达基因的筛选

GEO2R(<http://www.ncbi.nlm.nih.gov/geo/geo2r>)是一种交互式网络工具，允许用户比较GEO系列中的2个或多个基因芯片数据。通过GEO2R分析，设定筛选条件 $P<0.05$, logFC绝对值 >1 ，分别得到PTC和癌旁组织之间的差异表达基因(差异表达基因集A)以及PTC与颈部淋巴结转移癌组织之间的差异表达基因(差异表达基因集B)，并通过在线分析网站Bioinformatics & Evolutionary Genomics(<http://bioinformatics.psb.ugent.be/webtools/Venn/>)筛选出差异表达基因集A和差异表达基因集B之间的交集差异表达基因集C。

1.3 基因本体分析

基因本体(Gene Ontology, GO)分析可分为分子功能(molecular function), 生物过程(biological process)和细胞组成(cellular component)分析3个部分。GO分析广泛用于基因芯片数据的研究，David是常用的基因功能分析的在线生物信息学软件，能够对大规模的基因组数据进行功能分析。我们将差异表达基因集C导入David 6.8在线分析网站(<https://david.ncifcrf.gov/>)，以错误发现率(false discovery rate, FDR) <0.05 作为入选标准，对差异

表达基因集C进行功能注释, 分析这些基因主要参与的分子功能、生物过程和细胞组成。

1.4 蛋白互作网络构建

String(<https://string-db.org/>)是研究蛋白之间相互作用的在线生物信息学网站。我们将差异表达基因集C导入String 11.0在线分析网站(<https://string-db.org/>), 将最低互作分值(minimum required interaction score)设置成中度可信(medium confidence: 0.4)进行分析, 获得了蛋白与蛋白之间相互作用的网络。

1.5 差异表达基因集 C 对 PTC 预后的影响

通过可视化网站cBioPortal(<http://www.cbioportal.org/>)分析含有507例样本^[6]的PTC与正常标本的TCGA数据。通过对差异表达基因集C的单个基因进行生存分析, 分析差异表达基因集C的单个基因mRNA表达与PTC预后的关系, 得到单个基因对患者的总体生存Kaplan-Meier估计以及疾病/无进展的Kaplan-Meier估计(以log-rank test $P<0.05$ 为差异有统计学意义)。

2 结果

2.1 差异表达基因筛选

通过检索GEO数据库, 得到PTC颈部淋巴结转移癌组织的基因芯片GSE60542。GSE60542含有33个PTC样品、29个癌旁样品、23个颈部淋巴结转移癌组织样品。通过GEO2R分析GSE60542芯片数据, 一共获得1 086个差异表达基因A、194个差

异表达基因B。通过在线分析网站Bioinformatics & Evolutionary Genomics分析, 得到39个差异表达基因C(图1)。

2.2 差异表达基因集 C 的 GO 分析

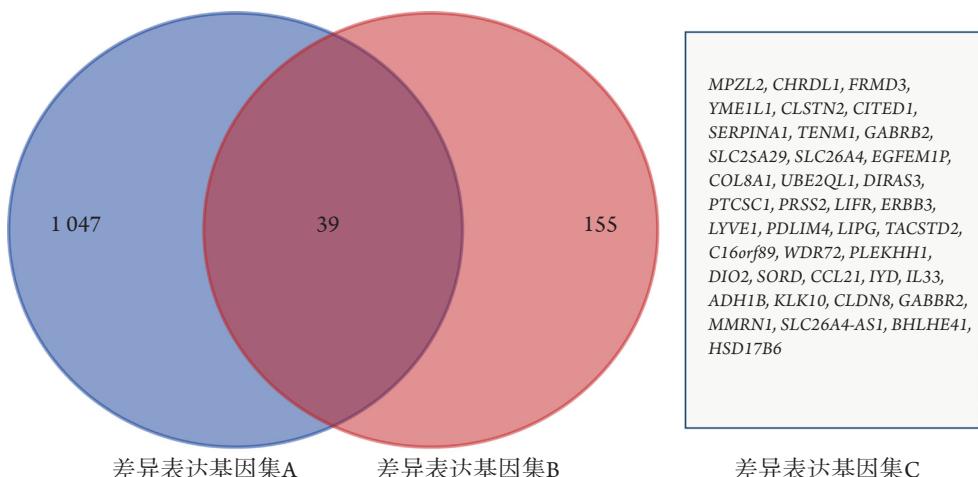
对差异表达基因集C进行GO分析。首先, 通过分析差异表达基因集C的主要分子功能、主要参与的生物学过程以及生物学组成, 发现这些基因主要的分子功能与氧化还原酶活性、蛋白质同源二聚化活性、生长因子结合有关; 这些基因主要参与细胞增殖、甲状腺激素的产生、神经细胞凋亡过程的负调节、 γ -氨基丁酸信号转导途径等生物学过程; 而细胞组成分析显示: 这些基因大多参与细胞外区域、细胞外空间、质膜的组成部分、膜的组成部分、细胞骨架等的组成(表1)。

2.3 蛋白互作网络分析

从蛋白水平对PTC组织、癌旁组织以及颈部淋巴结转移癌组织差异表达基因的生物学功能和调控作用进行研究, 我们通过在线分析网站String构建了差异表达基因集C编码蛋白的互作网络(图2)。

2.4 差异表达基因集 C 对 PTC 患者预后的影响

通过cBioPortal网站绘制单个基因生存曲线, 分析差异表达基因集C单个基因的表达与PTC患者预后的关系, 得到单个基因的总体生存Kaplan-Meier估计以及疾病/无进展的Kaplan-Meier估计。对患者生存时间影响较大的基因仅有CCL21, 差异有统计学意义(log-rank test $P=0.0362$, $P=0.0499$; 图3)。



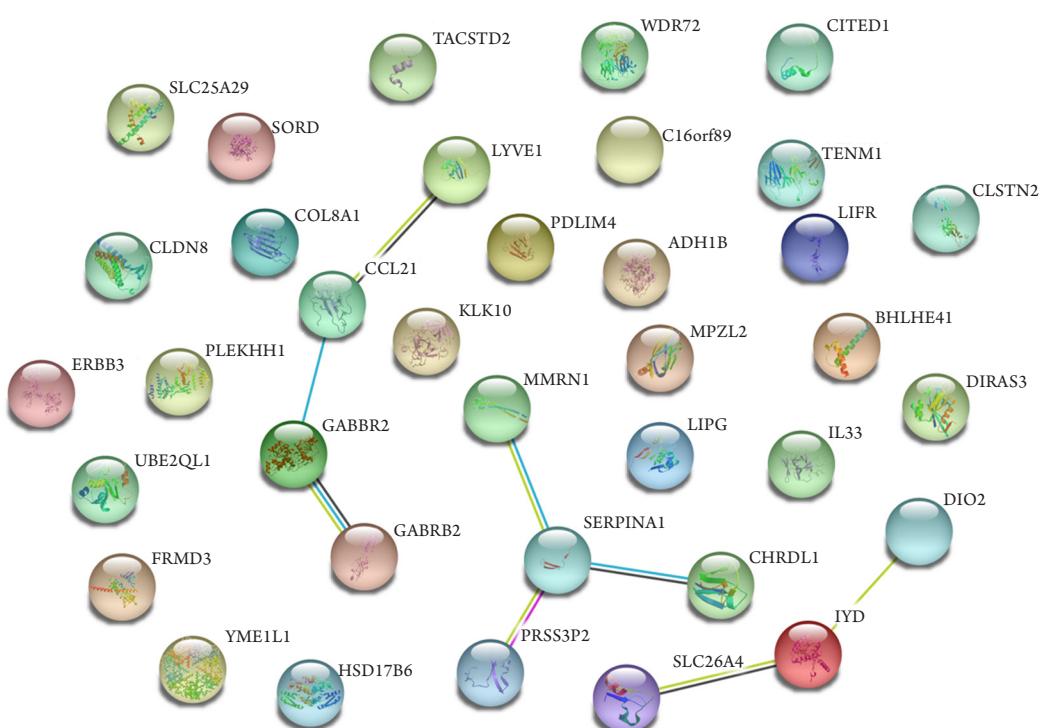
MPZL2, CHRDL1, FRMD3,
YME1L1, CLSTN2, CITED1,
SERPINA1, TENM1, GABRB2,
SLC25A29, SLC26A4, EGFBEM1P,
COL8A1, UBE2QL1, DIRAS3,
PTCSC1, PRSS2, LIFR, ERBB3,
LYVE1, PDLM4, LIPG, TACSTD2,
C16orf89, WDR72, PLEKHH1,
DIO2, SORD, CCL21, IYD, IL33,
ADH1B, KLK10, CLDN8, GABBR2,
MMRN1, SLC26A4-AS1, BHLHE41,
HSD17B6

图1 差异表达基因集A, B, C

Figure 1 Differentially expressed gene set A, B, C

表1 差异表达基因集C的GO分析**Table 1 GO analysis for differentially expressed gene set C**

GO分析	Count	P
分子功能		
氧化还原酶活性	4	0.006310094
蛋白质同源二聚化活性	5	0.047780146
生长因子结合	2	0.049977780
生物学过程		
细胞增殖	5	0.005514412
甲状腺激素的产生	2	0.023336985
神经细胞凋亡过程的负调节	3	0.027622762
γ-氨基丁酸信号转导途径	2	0.042380272
生物学组成		
细胞外区域	10	0.002197687
细胞外空间	8	0.010904214
质膜的组成部分	8	0.014063757
膜的组成部分	17	0.015482502
细胞骨架	4	0.031434217

**图2 差异表达基因集C编码蛋白的互作网络****Figure 2 Interaction network of differentially expressed gene sets C-encoded proteins**

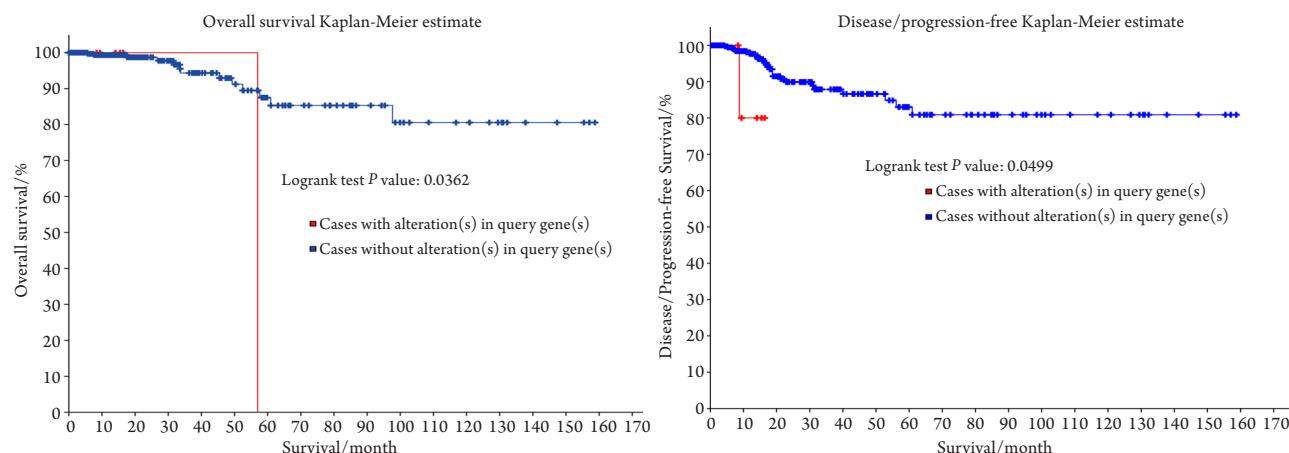


图3 CCL21基因表达与患者总体生存Kaplan-Meier估计以及疾病/无进展的Kaplan-Meier估计

Figure 3 Kaplan-Meier estimate and Kaplan-Meier estimate of disease/progression-free survival of CCL21 gene expression in patients

3 讨论

随着分子生物学技术的发展及其在临床领域中的应用，甲状腺癌发生、发展、转移及预后相关基因的研究及其相关分子标志物越来越引起人们的关注。微阵列型研究以及新一代测序技术等现代生物技术的发展以及生物信息学的应用为我们从分子水平揭示甲状腺癌的发生、发展机制提供了很好的手段^[7]。本研究采用生物信息学方法比较分析了GEO数据库中GSE60542数据集PTC组织、癌旁组织以及颈部淋巴结转移癌组织基因的表达差异，筛选出39个表达差异明显的基因，并采用David在线生物信息学分析工具对筛选出的差异表达基因进行GO功能注释分析，通过String网站分析PTC颈部淋巴结转移相关的蛋白-蛋白互作网络分析，最后通过cBioPortal网站对关键基因绘制生存曲线，分析差异表达基因与PTC预后的关系。

本研究GO功能注释结果显示：差异表达基因集C主要的分子功能与氧化还原酶活性、蛋白质同源二聚化活性、生长因子结合有关；这些基因主要参与细胞增殖、甲状腺激素的产生、神经细胞凋亡过程的负调节、γ-氨基丁酸信号转导途径等生物学过程；而细胞组成分析显示这些基因大多参与细胞外区域、细胞外空间、质膜的组成部分、膜的组成部分、细胞骨架等的组成。蛋白-蛋白互作网络分析结果显示，差异表达基因集C中存在3条蛋白质相互作用网络，分别为LYVE1-CCL21-GABBR2-GABRB2，DIO2-IYD-SLC26A4，CHRDL1-SERPINA1-MMRN1-PRSS3P2，但这些蛋白质之间

相互作用的关系，尚需进一步的实验证实。

为进一步筛选出PTC颈部淋巴结转移相关的关键基因，我们通过cBioPortal网站绘制生存曲线，筛选出对患者生存时间影响较大的基因CCL21。CCL21又称6 Ckine或次级淋巴组织趋化因子，是一种主要表达于外周免疫器官或组织的趋化因子，CCL21与其特异性受体CCR7在多种肿瘤的淋巴结转移中起重要作用^[8]。据文献[9]报道：CCL21/CCR7轴在促进乳腺癌诱导的淋巴管生成中具有重要作用。还有研究^[10]指出：CCL21/CCR7轴可以调节MEK/ERK1/2信号通路促进膀胱癌细胞侵袭和迁移，并与膀胱癌淋巴结转移相关。Sun等^[11]研究发现：CCL21/CCR7在肺癌中通过ERK/Akt信号通路诱导血管内皮生长因子VEGF-D上调从而促进淋巴管生成，并进一步研究发现CCL21和CCR7的表达水平与人非小细胞肺癌的临床分期、淋巴结转移以及患者生存时间相关。有学者在甲状腺癌中也进行了相关研究，Zhang等^[12]研究表明：CCL21/CCR7相互作用显著上调细胞周期蛋白A、细胞周期蛋白B1和细胞周期蛋白依赖性激酶1(CDK1)的表达，最终导致PTC细胞中G2/M期细胞的比例增加，从而影响PTC细胞的增殖。但CCL21基因是否与PTC淋巴结转移相关，尚需实验验证。

本研究从生物信息学角度对PTC颈部淋巴结转移相关基因进行探索，但缺少相应的实验学验证，此为本研究的不足。由于本研究仅纳入1个基因芯片数据(GSE60542)，一定程度上限制了差异表达基因的筛选，最终只发现1个与患者预后生存

相关的基因。综上，本研究利用生物信息学方法通过对GEO芯片数据进行差异表达基因的功能分析以及蛋白-蛋白互作网络分析、并通过生存分析进一步探究了关键基因与PTC颈部淋巴结转移患者预后的关系，为进一步研究PTC颈部淋巴结转移分子机制具有一定的指导意义。

参考文献

1. Loret-Tieulent J, Franceschi S, Dal Maso L, et al. Thyroid cancer “epidemic” also occurs in low- and middle-income countries[J]. Int J Cancer, 2019, 144(9): 2082-2087.
2. 高明, 郑向前. 甲状腺癌过去与未来十年[J]. 中国肿瘤临床, 2018, 45(1): 2-6.
GAO Ming, ZHENG Xiangqian. Thyroid cancer over the past decade and the next decade[J]. Chinese Journal of Clinical Oncology, 2018, 45(1): 2-6.
3. Lee YC, Na SY, Park GC, et al. Occult lymph node metastasis and risk of regional recurrence in papillary thyroid cancer after bilateral prophylactic central neck dissection: a multi-institutional study[J]. Surgery, 2017, 161(2): 465-471.
4. 杨旭辉. 甲状腺癌分子标记物的研究进展[J]. 临床与病理杂志, 2015, 35(6): 1148-1156.
YANG Xuhui. Research progress of thyroid cancer molecular markers[J]. Journal of Clinical and Pathological Research, 2015, 35(6): 1148-1156.
5. Zhang L, Huang Y, Ling J, et al. Screening and function analysis of hub genes and pathways in hepatocellular carcinoma via bioinformatics approaches[J]. Cancer Biomarkers, 2018, 22(3): 511-521.
6. Cancer Genome Atlas Research Network. Integrated genomic characterization of papillary thyroid carcinoma[J]. Cell, 2014, 159(3): 676-690.
7. Bai Y, Sartor M, Cavalcoli J. Current status and future perspectives for sequencing livestock genomes[J]. J Anim Sci Biotechnol, 2012, 3(1): 8.
8. Raman D, Baugher PJ, Thu YM, et al. Role of chemokine in tumor growth[J]. Cancer Lett, 2007, 256(2): 137-165.
9. Tutunea-Fatan E, Majumder M, Xin X, et al. The role of CCL21/CCR7 chemokine axis in breast cancer-induced lymphangiogenesis[J]. Mol Cancer, 2015, 14: 35.
10. Xiong Y, Huang F, Li X et al. CCL21/CCR7 interaction promotes cellular migration and invasion via modulation of the MEK/ERK1/2 signaling pathway and correlates with lymphatic metastatic spread and poor prognosis in urinary bladder cancer[J]. Int J Oncol, 2017, 51(1): 75-90.
11. Sun L, Zhang Q, Li Y, et al. CCL21/CCR7 up-regulate vascular endothelial growth factor-D expression via ERK pathway in human non-small cell lung cancer cells[J]. Int J Clin Exp Pathol, 2015, 8(12): 15729-15738.
12. Zhang YY, Liu ZB, Ye XG, et al. Iodine regulates G2/M progression induced by CCL21/CCR7 interaction in primary cultures of papillary thyroid cancer cells with RET/PTC expression[J]. Mol Med Rep, 2016, 14(4): 3941-3946.

本文引用: 侯建忠, 樊友本, 伍波. 甲状腺乳头状癌颈部淋巴结转移相关基因的生物信息学分析[J]. 临床与病理杂志, 2020, 40(2): 259-264. doi: 10.3978/j.issn.2095-6959.2020.02.001

Cite this article as: HOU Jianzhong, FAN Youben, WU Bo. Bioinformatics analysis for genes related to cervical lymph node metastasis of papillary thyroid carcinoma[J]. Journal of Clinical and Pathological Research, 2020, 40(2): 259-264. doi: 10.3978/j.issn.2095-6959.2020.02.001