网络出版时间:2020-8-21 14:48 网络出版地址:https://kns.cnki.net/kcms/detail/34.1065.r.20200820.1449.005.html

基因集富集分析探讨 HER2 基因对胃癌代谢的影响

刘 虎[1],吴思洽[1],包楚阳[2],李文娟[1],周守兵[1],金 伟[1]

摘要 **目的** 运用基因集富集分析(GSEA)探索人表皮生长因子受体2(HER2)基因表达状态对胃癌代谢通路影响情况。**方法** 下载癌症基因组图谱(TCGA)、基因表达综合数据库(GEO)和欧洲生物信息研究所(ArrayExpress)数据库的胃癌转录表达数据库,根据Her2拷贝数或者表达水平选取高低表达组,应用GSEA软件进行富集分析,取错误发现率q值(FDR q val)<25%和/或名义*P*值(nom *P* val)<0.01的代谢通路作为有意义的代谢基因集,结果绘制热图寻找共性。**结果** 在癌症基因组图谱胃腺癌集(TCGA STAD)、GEO数据集(GSE66229)等10个数据库中,Her2的高低表达对对过氧物酶体、N-聚糖生物合成和嘧啶代谢通路基因集有影响(FDR q val<25%且nom *P* val<0.01),糖基磷脂酰肌醇、甘油磷脂、鞘脂类代谢差异有统计学意义(nom *P* val<0.01)。**结论** 多个数据库GSEA分析结果提示癌基因Her2的状态与多个代谢通路基因集有相关性。

关键词 胃癌;癌症基因组图谱;基因集富集分析;人类表皮生长因子受体2;京都基因与基因组百科全书;代谢
中图分类号 R 735.2
文献标志码 A **文章编号** 1000-1492(2020)09-1339-05
doi:10.19405/j.cnki.issn1000-1492.2020.09.005

胃癌恶性度高,预后差,尤其晚期病情进展迅速,通常伴随恶病质状态。胃癌患者代谢异常导致生存期缩短是目前临床上的棘手问题。癌基因人类表皮生长因子受体2(human epidermal growth factor receptor 2 gene,Her2)过度表达是胃癌预后差的指标,也是靶向治疗的靶点,约有20%胃癌病人伴有Her2过表达,是胃癌明确的癌基因[1]。Her2是否会对胃癌代谢产生影响尚未有报道和关注。该文通过基因集富集分析(gene set enrichment analysis,GSEA)对癌基因Her2进行转录组测序技术及生物信息学方法研究。通过对癌症基因组图谱(the cancer genome atlas,TCGA)、基因表达综合数据库(the gene expression omnibus databases,GEO)和欧洲生物信息研究所(EMBL-EBI)ArrayExpress中的胃癌转录测序数据库进行分析,试图探讨Her2基因的高低表达水平对胃癌代谢通路的影响。

2020-04-20 接收
基金项目:国家自然科学基金(编号:81472750)
作者单位:[1]中国科学技术大学附属第一医院西区,安徽省肿瘤医院肿瘤化疗科,合肥 230031
[2]安徽医科大学第一附属医院肿瘤内科,合肥 230022
作者简介:刘 虎,男,博士,副主任医师,责任作者,E-mail:drliuhu@yahoo.com

1 材料与方法

1.1 研究对象 以TCGA的胃癌数据库(stomach adenocarcinoma,STAD)[2]、GEO数据库、ArrayExpress中搜索的胃癌RNAseq相关数据库(样本量在100以上)作为研究对象,筛选出10个数据库作为研究对象,见表1。通过R语言软件包TCGA2STAT[3]从Broad GDAC Firehose下载TCGA STAD RNA转录表达(RNAseq)和CNA数据,GEO在NCBI GEO直接下载,ArrayExpress在EMBL-EBI ArrayExpress直接下载。

表1 入选的胃癌基因表达数据库

数据库编号	样本量
TCGA STAD	295
STAD CNA	295
GSE54129[4]	132
GSE27342[5]	160
GSE15460[6]	360
GSE14210[7]	167
GSE26253[8]	432
GSE37023[9]	250
GSE15459[10]	200
GSE66229[8]	400
E-MTAB-1338[11]	108

1.2 GSEA软件分析 在下载的数据库中,根据Her2基因表达水平排列次序,选取高低表达各20~30样本制作表型标签,以TCGA STAD的为例见表2。将表达数据库(下载的TCGA和GEO表达数据库)、表型标签加载到GSEA java软件中,设置排列数目为1 000个,chip platform选用对应数据所用的测序或者芯片平台,比如GPL5175平台。MSigDB选择C2 curated gene sets CP KEGG gene sets(包含186个基因集)。选取不同样本,制作不同的标签,运行3~4次GSEA java,取错误发现率(false discovery rate,FDR) q val<25%和nom *P* val<0.01基因

集最多的结果。

1.3 应用的软件格式 GSEA 富集分析 JAVA 软件下载于 Broad Institute，版本号 GSEA 4.0.3，R 语言 TCGA2STAT 软件包从 The Comprehensive R Archive Network(CRAN) 下载。

表 2 TCGA STAD 表型标签制作

Her2_低表达	Her2_高表达
TCGA-BR-A4J9-01	TCGA-CG-5726-01
TCGA-FP-7735-11	TCGA-D7-A4YY-01
TCGA-BR-7715-11	TCGA-D7-6520-01
TCGA-BR-A44T-01	TCGA-HU-A4H0-01
TCGA-BR-A4IV-01	TCGA-F1-6177-01
TCGA-BR-6802-11	TCGA-BR-8590-01
TCGA-HU-A4GJ-01	TCGA-BR-8373-01
TCGA-D7-6818-01	TCGA-D7-6817-01
TCGA-HU-A4GN-11	TCGA-BR-A4QI-01
TCGA-D7-6518-01	TCGA-D7-A4YT-01
TCGA-CG-5723-01	TCGA-BR-8682-01
TCGA-FP-8209-01	TCGA-BR-8289-01
TCGA-BR-8371-01	TCGA-D7-6820-01
TCGA-BR-8677-01	TCGA-D7-8579-01
TCGA-BR-A452-01	TCGA-BR-6565-01
TCGA-BR-6709-01	TCGA-D7-6526-01
TCGA-BR-4256-01	TCGA-BR-7715-01
TCGA-CD-8530-01	TCGA-IP-7968-01
TCGA-BR-8060-11	TCGA-D7-6815-01
TCGA-BR-A4PF-01	TCGA-HU-A4GF-01
TCGA-HU-A4GY-01	TCGA-FP-8631-01
TCGA-CG-5728-11	TCGA-FP-7735-01
TCGA-FP-7829-11	TCGA-CD-5800-01
TCGA-BR-6455-01	TCGA-D7-A4YX-01
TCGA-CD-8528-01	TCGA-HU-A4G6-01
TCGA-HU-8604-01	TCGA-D7-8573-01
TCGA-CG-5721-11	TCGA-CD-5804-01
TCGA-CD-5813-01	TCGA-HU-A4GH-01
TCGA-BR-8384-01	TCGA-HU-A4HD-01
TCGA-BR-4257-01	TCGA-BR-A4J4-01
TCGA-D7-6522-01	TCGA-CD-5799-01
TCGA-BR-A4J5-01	

1.4 统计学处理 GSEA 富集分析取 FDR q val < 25% 和/或 nom *P* val < 0.01 的富集结果认为差异有统计学意义。

2 结果

2.1 TCGA Her2 表达高低组的 GSEA 结果 在 Her2 高表达表型中 120/278 基因集上调，5 个基因集富集 FDR < 25%，5 个基因集富集 nom *P* < 1%，16 个基因集富集 nom *P* < 5%。FDR q val < 25% 且 nom *P* < 0.01 的基因集有 3 个：过氧化物酶体(KEGG PEROXISOME)、含硒氨基酸(KEGG SELENOAMINO ACID METABOLIS)、甘油酯类代谢(KEGG GLYCEROLIPID METABOLISM)(详见表 3)受 Her2 高表达影响最明显的前 50 基因热图见图 1。图 2 列出 3 个有意义基因集的富集图。

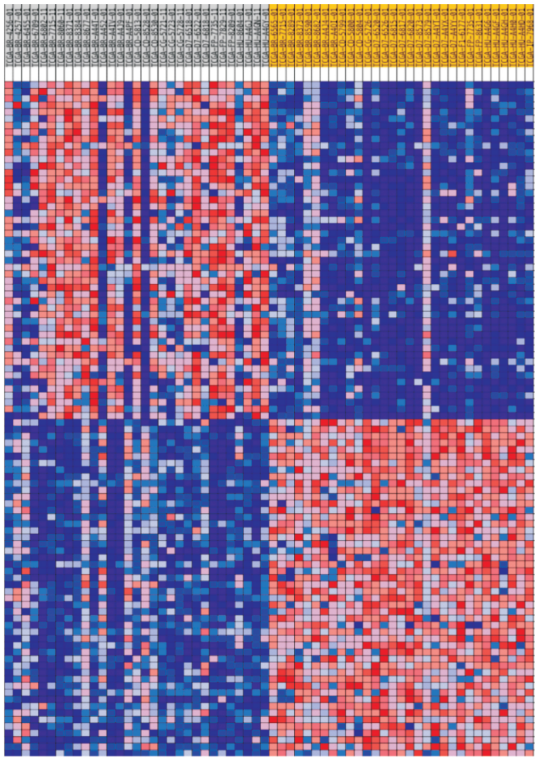

图 1 TCGA Her2 高低表达相关基因列表热图

2.2 TCGA 和 GEO 数据库的 GSEA 分析结果 用上述 TCGA 的 GSEA 分析方法对入选的 TCGA 的 CNA 和 GEO、ArrayExpress 所有数据进行分析，结果综合见表 4。以 FDR q val < 25% 和/或 nom *P* val < 0.01 判断为有意义。结果显示 Her2 的高表达在 TCGA、GSE15459 和 GSE66229 数据中 GSEA 分析有阳性结果。

2.3 不同数据库代谢通路影响的热图 通过绘制热图比较显示 Her2 对过氧物酶体、N-聚糖生物合成和嘧啶代谢通路基因集有非常影响(FDR q < 25% 且 nom *P* < 0.01)，糖基磷脂酰肌醇、甘油磷脂、鞘脂类代谢差异有统计学意义(nom *P* < 0.01)(见图 3)。

3 讨论

胃癌起病隐匿，疾病进展迅速，晚期伴有代谢异常，营养吸收障碍，加速病情恶化。本研究组前期转录组研究也发现胃癌中脂质代谢相关基因的异常表达[12]。Her2 是胃癌最明确的癌基因且有判断预后

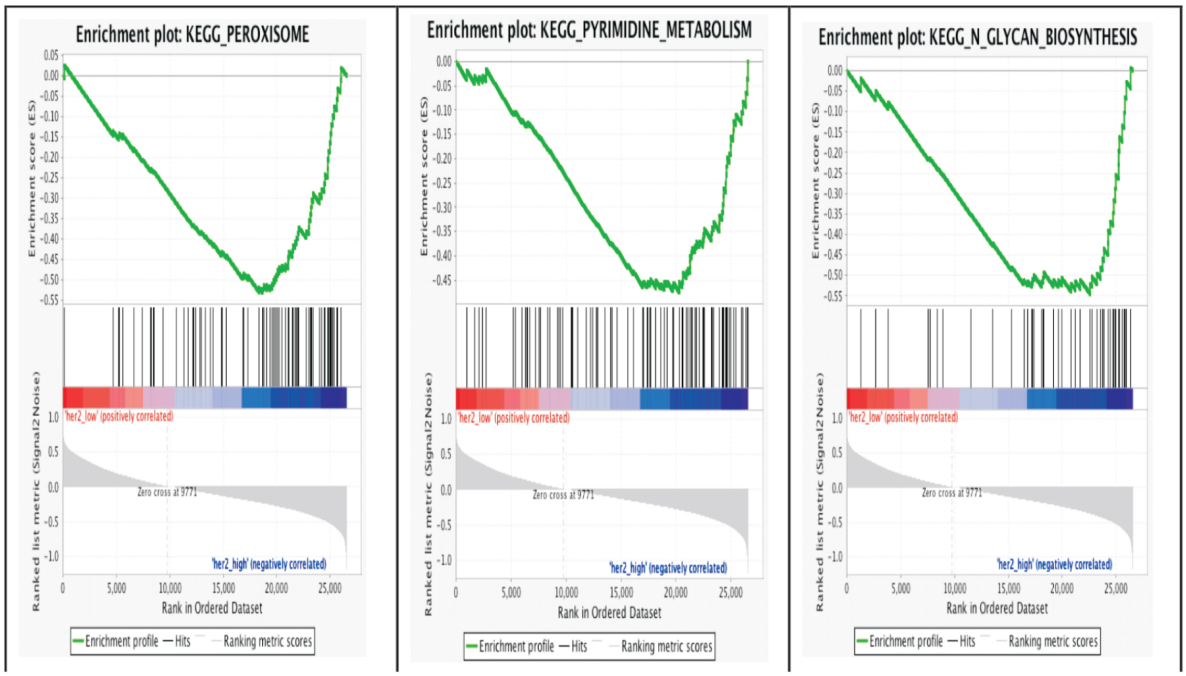

图 2　有意义的基因集的富集图

表 3　TCGA STAD 中 Her2 高表达富集结果列表

基因集名称	数量	标准	NOM *P*-val	FDR q-val	FWER p-val
KEGG_GLYCOSYLPHOSPHATIDYLINOSITOL_GPI_ANCHOR_BIOSYNTHESIS	25	-1.8308948	0	0.4274352	0.18
KEGG_SPHINGOLIPID_METABOLISM	39	-1.8086642	0.02222222	0.27447587	0.21
KEGG_BASAL_TRANSCRIPTION_FACTORS	33	-1.8022268	0.02	0.19395328	0.22
KEGG_PEROXISOME	76	-1.7493895	0	0.22516415	0.3
KEGG_GLYCEROPHOSPHOLIPID_METABOLISM	76	-1.6889246	0	0.28718305	0.44
KEGG_HOMOLOGOUS_RECOMBINATION	28	-1.6717087	0.04545455	0.26695168	0.48
KEGG_SELENOAMINO_ACID_METABOLISM	26	-1.6659013	0	0.24073945	0.49
KEGG_GLYCEROLIPID_METABOLISM	49	-1.6428105	0	0.24867189	0.51
KEGG_PYRIMIDINE_METABOLISM	94	-1.6353688	0.04651163	0.2459737	0.56
KEGG_N_GLYCAN_BIOSYNTHESIS	45	-1.5951	0.04347826	0.2951885	0.6
KEGG_ONE_CARBON_POOL_BY_FOLATE	17	-1.5632386	0.02325581	0.35539606	0.68
KEGG_TERPENOID_BACKBONE_BIOSYNTHESIS	15	-1.5499818	0.12195122	0.36099437	0.72
KEGG_BASE_EXCISION_REPAIR	33	-1.542029	0.07692308	0.34981707	0.74
KEGG_CELL_CYCLE	124	-1.5407037	0.11904762	0.32550505	0.74
KEGG_NOTCH_SIGNALING_PATHWAY	47	-1.5081813	0.04081633	0.38614118	0.84
KEGG_PENTOSE_PHOSPHATE_PATHWAY	27	-1.5017622	0.09090909	0.3713867	0.85
KEGG_SPLICEOSOME	124	-1.490956	0.02439024	0.37076354	0.86
KEGG_NUCLEOTIDE_EXCISION_REPAIR	43	-1.4884012	0.0952381	0.35905463	0.86
KEGG_AMINOACYL_TRNA_BIOSYNTHESIS	41	-1.4817811	0.14583333	0.35182452	0.87
KEGG_MISMATCH_REPAIR	23	-1.4817796	0.15	0.33423328	0.87
KEGG_ENDOCYTOSIS	178	-1.4739231	0.08510638	0.3379887	0.88
KEGG_TIGHT_JUNCTION	129	-1.4574409	0.03703704	0.35765696	0.91
KEGG_OTHER_GLYCAN_DEGRADATION	16	-1.4573224	0.04651163	0.34252074	0.91
KEGG_PROTEASOME	44	-1.4540299	0.12195122	0.3336786	0.92
KEGG_UBIQUITIN_MEDIATED_PROTEOLYSIS	134	-1.4481871	0.04255319	0.33357808	0.92

和靶向药物，而癌基因对胃癌代谢的影响并未受关注。生物信息学的发展提供了数据挖掘分析的可能。GSEA 富集分析 java 软件可以初步探讨 Her2 对胃癌代谢的影响。

本研究搜索了 TCGA，GEO 和 ArrayExpress 数据库中胃癌的主要上规模的测序数据，共 10 个，并用

Her2 表达高低或者拷贝数高低(copy number analyses,CNA)作为表型标签,加载到 GSEA 中,和 MSigDB 中注释(curated)的基因集进行富集分析。当 FDR q val <25% 且 nom *P* val <0.01 时判断为有意义的基因集。在本文中详细列举了 TCGA STAD 中 Her2 高低表达的 GSEA 分析结果,GSEA 能获得与表型相关的基因热图,通过标化的富集评分(normalized enrichment score, NES)评判基因集与表型的关联。通过比较不同数据库中的 GSEA 分析结果获得一个共性结果,从而得到推断 Her2 高表达可能对过氧物酶体、N-聚糖生物合成和嘧啶代谢通路有影响。

表 4 Her2 高表达在不同基因数据库中对基因集的富集分析结果

基因集	样本	基因集上调比例	FDR <25%	nom *P* <1%	nom *P* <5%
TCGA STAD	295	120/178	5	5	16
STAD CNA	295	106/178	15	9	20
GSE54129	132	81/176	0	2	9
GSE27342	160	67/156	1	2	14
GSE 15460	360	80/176	0	2	6
GSE 14210	167	158/168	0	0	3
GSE 26253	432	91/166	0	0	4
GSE 37023	250	115/168	1	1	2
GSE 15459	200	89/176	6	7	13
GSE 66229	400	71/176	10	3	14
E-MTAB-1338	108	1/178	0	0	0

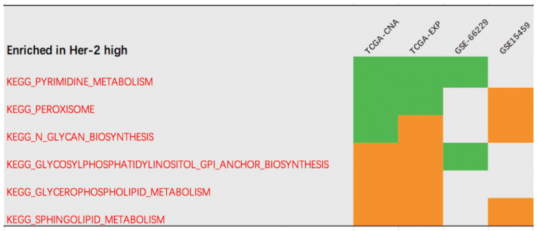

图 3 Her2 在不同数据中对基因集影响的热图

绿色: FDR q val <25% 且 nom *P* val <0.001,**黄色:** nom *P* val <0.001

Her2 高表达所影响的代谢通路本身与肿瘤就有密切联系并起核心作用,比如:过氧化物酶体分子和过氧化物酶体特异性蛋白在过氧化应激与肿瘤发生中有作用。过氧化物酶体本身可能充当信号枢纽,促进其他支持肿瘤的发生过程,例如自噬[13]。再比如:参与 N-聚糖生物合成的酶及其产物表达的变化可以调节结直肠癌细胞的细胞黏附、细胞信号传导和侵袭性[14]。而嘧啶作为 DNA 合成原料其代谢与肿瘤的失控生长关系更是密切[15]。Her2 与这些代谢之间的相互影响尚未被关注并需进一步实验证实。

综上所述,通过 GSEA 探讨了癌基因 Her2 对代谢通路的可能影响,并比较不同数据库中的结果初步形成了推断:Her2 高表达可能对多个代谢通路产生影响。

(致谢:感谢英国帝国理工大学 Stefan Antonowitz 博士提供的技术指导和帮助。)

参考文献

[1] Cutsem E V,Bang Y J,Feng F Y,et al. HER2 screening data from ToGA: targeting HER2 in gastric and gastroesophageal junction cancer[J]. Gastric Cancer,2015,18(3):476-84.

[2] Bass A,Thorsson V,Shmulevich I,et al. Comprehensive molecular characterization of gastric adenocarcinoma[J]. Nature,2014,513(7517):202-9

[3] Wan Y W,Allen G I,Liu Z. TCGA2STAT: simple TCGA data access for integrated statistical analysis in R[J]. Bioinformatics,2016,32(6):952-4.

[4] Sun C,Yuan Q,Wu D,et al. Identification of core genes and outcome in gastric cancer using bioinformatics analysis[J]. Oncotarget,2017,8(41):70271-80.

[5] Cui J, Chen Y, Chou W C,et al. An integrated transcriptomic and computational analysis for biomarker identification in gastric cancer[J]. Nucleic Acids Res,2011,39(4):1197-207.

[6] Subhash V V, Yeo M S, Wang L,et al. Anti-tumor efficacy of Selinexor(KPT-330) in gastric cancer is dependent on nuclear accumulation of p53 tumor suppressor[J]. Sci Rep,2018,8(1):12248.

[7] Kim H K, Choi I J, Kim C G,et al. Three-gene predictor of clinical outcome for gastric cancer patients treated with chemotherapy[J]. Pharmacogenomics J,2012,12(2):119-27.

[8] Oh S C, Sohn B H, Cheong J H,et al. Clinical and genomic landscape of gastric cancer with a mesenchymal phenotype[J]. Nat Commun,2018,9(1):1777.

[9] Wu Y, Grabsch H, Ivanova T,et al. Comprehensive genomic meta-analysis identifies intra-tumoural stroma as a predictor of survival in patients with gastric cancer[J]. Gut,2013,62(8):1100-11.

[10] Muratani M, Deng N, Ooi W F,et al. Nanoscale chromatin profiling of gastric adenocarcinoma reveals cancer-associated cryptic promoters and somatically acquired regulatory elements[J]. Nat Commun,2014,5:4361.

[11] Zhao C M,Hayakawa Y,Kodama Y,et al. Denervation suppresses gastric tumorigenesis[J]. Sci Transl Med,2014,6(250):250ra115.

[12] 向丽娟,汪圣毅,包楚阳,等. 胃癌脂代谢通路基因表达的转录组学高通量分析[J]. 安徽医科大学学报,2019,54(1):5-9.

[13] Sugiura A, Mattie S, Prudent J, et al. Newly born peroxisomes are a hybrid of mitochondrial and ER-derived pre-peroxisomes[J]. Nature,2017, 542(7640):251-4.

(下转第 1349 页)

methacrylate and ethylene glycol dimethacrylate copolymer segments as the core, and oligoethylene glycol methacrylate together with segment of 1,4,7,10-tetraazacyclododecane-1,4,7-triacetic acid as the shell respectively, followed by the modification of Oct as corona. These polymeric nanoparticles were chelated with gadolinium ions to obtain a novel tumor-targeted branched copolymer nanosized MR contrast agent [P(DO3A-Gd)-Oct]. The morphology of the products was analyzed by nuclear magnetic resonance (NMR), dynamic light scattering(DLS) and transmission electron microscope(TEM). The stability, hemolysis rate, cytotoxicity and *in vivo* safety of P(DO3A-Gd)-OCT were preliminarily tested. Magnetic resonance relaxation rates were measured using a clinical 3.0T magnetic resonance imager and *in vivo* magnetic resonance imaging was performed in H22 hepatoma bearing mice. ***Results*** According to the DLS and TEM results, the hydrodynamic diameter of P(DO3A-Gd)-Oct was about 20 nm. After the exposure to P(DO3A-Gd)-Oct, the hemolysis rate was only 0.34 %, and no obvious cyto-toxicity was observed. Compared with normal mice without any treatment, there was no obvious change of the pathological structures of major organs or the major blood biochemical indexes after the intravenous (i.v.) injection of P(DO3A-Gd)-Oct. The relaxivity of P(DO3A-Gd)-Oct (r1) was 8.33 $mM^{-1}s^{-1}$. After 60 min and 120 min of the i.v. injection, there were obvious increase of the signal in the tumor region in T1 weighted images of H22 tumor bearing mice, which was attributed to the increased concentration of gadolinium ions in tumor tissues compared with the non-targeted control. ***Conclusion*** P(DO3A-Gd)-Oct exhibits bio-compatibility and excellent MR contrast ability, which can contribute to the enhancement of the sensitivity and accuracy in the diagnosis of malignant tumors such as liver tumor.

Key words magnetic resonance contrast agents; branched copolymer; octreotide; targeting; tumor

(上接第 1342 页)

[14] de Freitas Junior J C, Morgado-Diaz J A. The role of N-glycans in colorectal cancer progression: potential biomarkers and therapeutic applications [J]. Oncotarget, 2016, 7(15): 19395 - 413.

[15] Robinson A D, Eich M L, Varambally S. Dysregulation of de novo nucleotide biosynthetic pathway enzymes in cancer and targeting opportunities [J]. Cancer Lett, 2020, 470: 134 - 40.

Study on the impact of HER2 expression on the metabolism of gastric cancer with gene set enrichment analysis

Liu Hu[1], Wu Sihan[1], Bao ChuYang[2], et al

([1]*Dept of Oncology, West District of Affiliated Hospital of University of Science and Technology of China, Anhui Provincial Cancer Hospital, Hefei* 230031;
[2]*Dept of Oncology, The First Affiliated Hospital of Anhui Medical University, Hefei* 230022)

Abstract ***Objective*** To explore the impact of the status of human epidermal growth factor receptor 2 gene (HER2) gene on the metabolism of gastric cancer with computational gene sets enrichment analysis (GSEA). ***Methods*** The expression datasets of gastric cancer were downloaded from the cancer genome atlas(TCGA), gene expression omnibus(GEO) and ArrayExpress. Phenotype labels of Her2 high and low expression were made, which were applied in the GSEA. The meaningful metabolic genesets were found at false discovery rate(FDR) q value < 25% and/or nominal *P* value < 0.01. Heatmap was made to compare the meaningful gene sets in different datasets. ***Results*** Among the 10 datasets including TCGA STAD and GSE66229, peroxisome, N-glycan and pyrimidine metabolisms were significantly affected at FDR q val < 25% and nom *P* val < 0.01, glycosylphospatidylinositol GPI anchor biosynthesis, glycerolphosphalipid metabolism and sphingolipid metabolism at *P* value < 0.01. ***Conclusion*** Multiple metabolic gene sets are affected by Her2 status with evidences in multiple datasets.

Key words gastric cancer; TCGA; GSEA; HER2; KEGG; metabolism