

• 临床研究 •

胃癌关键基因和通路的生物信息学和功能分析*

吴茜, 宋兴勃, 钟慧钰, 温阳, 应斌武[△]

610041 成都, 四川大学华西医院 实验医学科

[摘要] 目的: 本研究拟通过生物信息学分析的方法分析胃癌及癌旁正常组织差异表达的基因, 探索胃癌的发病机制, 为胃癌的早期诊断和治疗评估提供新思路。方法: 从基因表达数据库中获取 GSE79973 和 GSE19826 基因表达谱, 使用 GCB 在线实验室筛选差异表达基因, 并使用基因本体论分析、代谢通路分析、基因信号通路网络分析、共表达分析对所获得基因进行分析。结果: 与对照组相比, 在胃癌组织中筛选出共 1 206 个基因呈差异表达。其中, 542 个基因表达下调, 664 个基因表达上调; 差异基因主要涉及细胞粘附、细胞外基质组织、细胞外基质分解、胶原蛋白分解代谢过程等方面, pathway 分析发现核心信号通路主要涉及粘附力、糖酵解/糖异生、Wnt 信号通路、癌症通路等方面; 基因信号通路网络分析发现的关键节点基因为 *UGT2B15*、*ITGA2*、*ITGB1*、*CYP3A4*; 共表达网络分析推测的关键节点基因为 *SH3GL2*、*CKMT2*、*CHIA*、*ATP4A*。结论: *INHBA*、*UGT2B15*、*ITGA2*、*ITGB1*、*SH3GL2* 等基因及其相关的生物过程可能是胃癌的潜在生物标志物和治疗靶标, 生物信息学有助于全面深入研究疾病发生机制, 筛选可能的核心靶点, 为临床诊断及疾病治疗提供参考。

[关键词] 生物信息学; 胃癌; GEO 数据库; GCB

[中图分类号] R735.2; R319 **[文献标志码]** A doi:10.3969/j.issn.1674-0904.2020.02.009

引文格式: Wu Q, Song XB, Zhong HY, et al. Bioinformatics and functional analysis of key genes and pathways in gastric cancer [J]. J Cancer Control Treat, 2020, 33(2): 131-139. [吴茜, 宋兴勃, 钟慧钰, 等. 胃癌关键基因和通路的生物信息学和功能分析[J]. 肿瘤预防与治疗, 2020, 33(2): 131-139.]

Bioinformatics and Functional Analysis of Key Genes and Pathways in Gastric Cancer

Wu Qian, Song Xingbo, Zhong Huiyu, Wen Yang, Ying Binwu

Department of Laboratory Medicine, West China Hospital, Sichuan University, Chengdu 610041, Sichuan, China

Corresponding author: Ying Binwu, E-mail: docbwy@126.com

This study was supported by National Natural and Science Foundation of China (NO. 81672096).

[Abstract] **Objective:** The study aims to explore the pathogenesis of gastric cancer (GC) and provide new biomarkers for early diagnosis and treatment evaluation of GC. **Methods:** GSE79973 and GSE19826 gene expression profiles were obtained from the gene expression database. Differentially expressed genes (DEGs) were screened by using Gene-Cloud of Biotechnology Information platform. Gene ontology enrichment analysis, pathway analysis, pathway and network analysis and co-expression analysis were performed to analyze DEGs. **Results:** Compared with the control group, 1206 genes were differentially expressed in GC tissue. Among them, 542 genes were down-regulated, and 664 genes were up-regulated; DEGs mainly involved cell adhesion, extracellular matrix tissue, extracellular matrix disassembly, collagen catabolism. Pathway analysis revealed that the core signaling pathway mainly involved adhesion, gluconeogenesis, Wnt signaling pathway, cancer pathway and so on. Pathway and network analysis found that key hub genes were *UGT2B15*, *ITGA2*, *ITGB1* and *CYP3A4*. Co-expression network analysis speculated that key hub genes were *SH3GL2*, *CKMT2*, *CHIA* and *ATP4A*. **Conclusion:** Genes such as *GKN2*, *GKN1*, *ATP4B*, *UGT2B15*, *SH3GL2* and related biological processes may be potential biomarkers and therapeutic

targets for GC. Bioinformatics can help us to comprehensively study the mechanism of disease occurrence and identify possible core targets, which can provide potential targets for the diagnosis and treatment of the disease.

[收稿日期] 2019-07-23 **[修回日期]** 2019-11-01

[基金项目] * 国家自然科学基金(编号:81672096)

[通讯作者] [△] 应斌武, E-mail: docbwy@126.com

[Key words] Bioinformatics; Gastric cancer; GEO database; GCB

胃癌的发病率较高,在癌症中排名前五,也是癌症相关死亡的三大主要原因之一,其中约有 50% 的死亡报告来自于东亚国家,如中国和日本^[1]。虽然 80% ~ 90% 的肿瘤发生是随机的,但仍然有研究证明遗传因素与胃癌的发生相关^[2]。近年来,肿瘤生物学研究不断发展,胃癌的临床诊疗也随之进步,但由于胃癌发现时间一般较晚,造成临床诊疗延误,因而其预后效果仍不令人满意,胃癌患者的 5 年生存率较低,仅为 25.1%^[3-4]。相关学者针对胃癌的发病机制做了大量研究,许多基因被认为和胃癌的发生发展有关^[5-6],但胃癌的发病机制仍旧没有达成统一的观点。确定胃癌早期诊断的相关指标和治疗胃癌的关键作用靶点十分重要,但相关方面的研究并不成熟。近年来,基因芯片和 RNA 测序等高通量技术的不断成熟使许多基因表达的相关数据得以呈现,计算机技术结合生物信息学的方法开始广泛应用于肿瘤研究,其在寻找肿瘤的诊断和靶向治疗的新型标志物方面的有效性和可靠性已经得到了证明。

本文在研究中拟选取基因表达综合数据库 GEO (Gene Expression Omnibus) 数据库中的 GSE79973 以及 GSE19826 数据集,将生物信息学方法应用到差异基因筛选和功能分析中,从而构建基

因网络图,完成关键基因靶点的筛选,以期为探索胃癌的发病机制提供新的见解,从而为胃癌的早期诊疗提供帮助。

1 材料和方法

1.1 数据来源

在隶属于美国国立生物技术信息中心的 GEO 数据库中使用“gastric cancer”关键词,获得 GSE79973 和 GSE19826 基因表达谱,分别由 He 等^[7]和 Wang 等^[8]上传,均基于 GPL570 平台[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array。共包含 47 例样本基因芯片数据,其中病例组(胃癌组织)22 例,对照组(正常胃组织)25 例。

1.2 数据分析流程

将基因芯片数据导入 GCB (Gene-Cloud of Biotechnology Information) 分析平台进行数据分析。首先,根据数据的样本特征将其分为数据组和对照组,进行差异基因的筛选,随后对其进行基因本体(gene ontology, GO)分析、Pathway 分析及在 Pathway 分析的基础上进行 Pathway 网络分析。最后,取 GO 分析和 Pathway 分析的交集结果进行基因信号通路网络分析及共表达网络分析。研究设计流程图如图 1 所示。

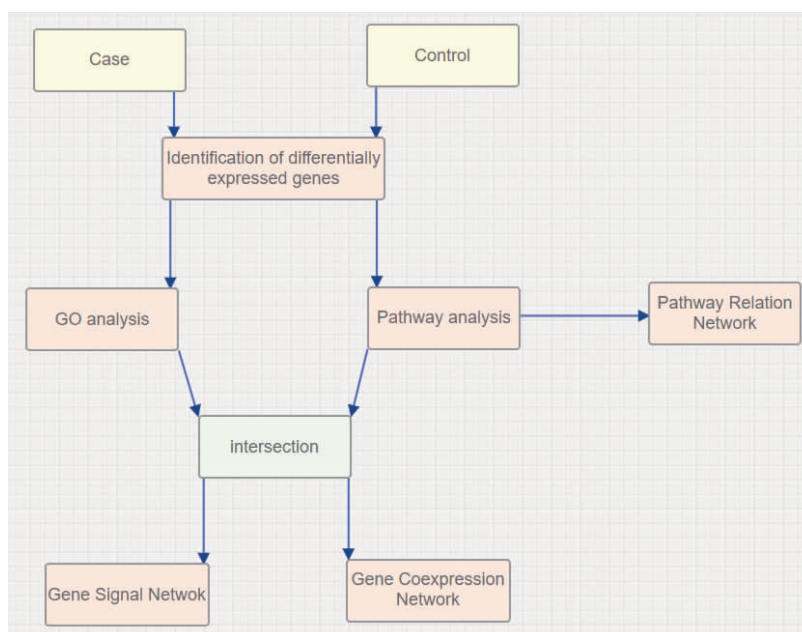


图 1 研究设计流程图

Figure 1. Flow Diagram of Study Design

1.3 差异基因筛选

基于 GCB 平台并对基因芯片数据进行标准处

理以利于分析并查看质控是否合格,在质控合格后设置参数(Q 值 < 0.05 , 差异倍数 > 2)进行差异基

因的筛选得到差异表达的基因 (differentially expressed genes, DEGs)。

1.4 GO 分析

GO 广泛应用于生物信息学领域中,可分为“生物过程”、“分子功能”以及“细胞组分”三个部分,GO 数据库可对基因产物进行简单注释,通常,GO 注释之后的步骤是 GO 富集分析,富集的显著性用 P 值表示,针对富集的显著性设定一个阈值,用于区分该富集是否显著。利用 GO 富集分析可以寻找差异基因富集在在哪些途径、细胞定位、生物学功能中^[9]。本研究设定错误发现率(false discovery rate, FDR) < 0.05,利用 GCBI 平台对差异表达基因进行 GO 分析,并筛选出差异最显著的 10 个基因功能进行排序及分析。

1.5 Pathway 分析

Pathway 分析是指代谢通路分析,通过对差异表达基因的 Pathway 显著性富集可以对其发挥最主要作用的代谢及信号传输途径进行了解。KEGG (Kyoto Encyclopedia of Genes and Genomes) 数据库是进行基因产物在细胞中的功能及其代谢途径分析的数据库。本研究使用 Fisher 精确检验,设定 $P < 0.05$,基于 GCBI 平台对差异表达基因进行 pathway 分析,从而筛选出差异最显著的 10 个功能并对其按照显著性进行排序及分析。

1.6 Pathway 关系网络分析

在 KEGG 数据库中,信号通路之间存在着上下游关系,根据此关系的不同,GCBI 平台构建了利用交互网络图进行 pathway 研究的分析方法。基于 pathway 分析结果,显著性 pathway 之间的信号传导关系以直观网络图的方式进行呈现,通过网络图可以确定最上游具有调控作用的信号通路和网络中处于最下游的效应信号通路。通过对于 pathway 之间的关系分析,可以对信号传递过程有更深刻的认识。

1.7 共表达网络分析

大量研究表明,基因在表达过程中相互之间存在影响,这种相互作用关系使得基因表达和调控过程形成了一个巨大而复杂的网络。GCBI 平台通过差异基因组成的表达谱数据进行基因共表达网络的构建。基因间的相互关系是具有层次性的,共表达网络可以将其清晰展现出来从而深入发现关键调控基因及其并呈现其相互作用的脉络。在共表达网络图中,根据基因之间调控关系的复杂程度,将每个基因用大小不同的点表示,点越大则说明调控关系更

加复杂,即其所在位置在网络分析中的价值更高。

1.8 基因信号通路网络分析

基因信号网络对 KEGG 数据库进行了解构,突破了只能在单一 pathway 中获取基因表达过程中相互作用关系的局限。因而其可以在整个 KEGG 途径数据库中,寻找出某个蛋白质的上游或下游蛋白质。我们使用 GCBI 平台进行了基因信号通路网络分析,以确定关键通路和基因。

2 结果

2.1 DEGs 筛选结果

DEGs 的识别是筛选样本间的高通量遗传数据具有显著差异的基因的统计方法。在数据预处理之后,与对照组织相比,在胃癌组织中筛选出共 1 206 个基因差异表达。在这些 DEGs 中,542 个基因表达下调,664 个基因表达上调。结果展示在火山图和聚类图中(图 2 和图 3),可见 DEGs 能较好的区分出 2 个组。差异表达最显著的 20 个基因为 *GKN2*、*GKN1*、*ATP4B*、*GIF*、*LIPF*、*PGA3*、*ATP4A*、*PGC*、*AQP4*、*INHBA*、*ESRRG*、*DPCRI*、*KCNE2*、*KCNJ16*、*VSIG1*、*FUT9*、*CHIA*、*DPCRI*、*PGC* 和 *MFSD4*(表 1)。

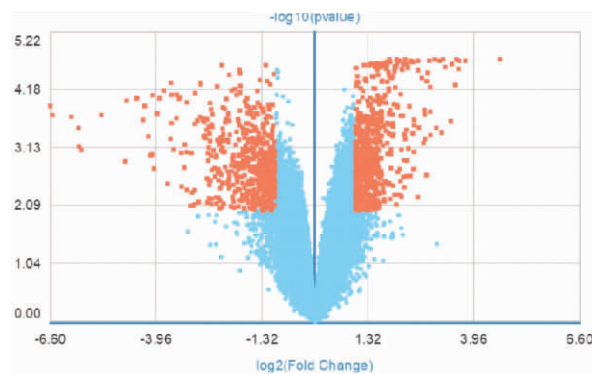


图 2 DEGs 差异火山图

Figure 2. Volcano Plot of DEGs

Orange represents DEGs; the down-regulated DEGs are on the left side of the midline, and the up-regulated DEGs are on the right. DEGs: Differentially expressed genes.

2.2 GO 分析结果

在本文中,使用 GCBI 平台获得的 DEGs 为 GO 富集分析形成了基础,计算 GO 的富集程度,展示了最可能与 DEG 相关的 10 个生物过程,包括:细胞粘附、细胞外基质组织、细胞外基质分解、胶原蛋白分解代谢过程、胶原纤维组织、小分子代谢过程、消化、血液凝固、血管生成、调节细胞增殖(表 2)。

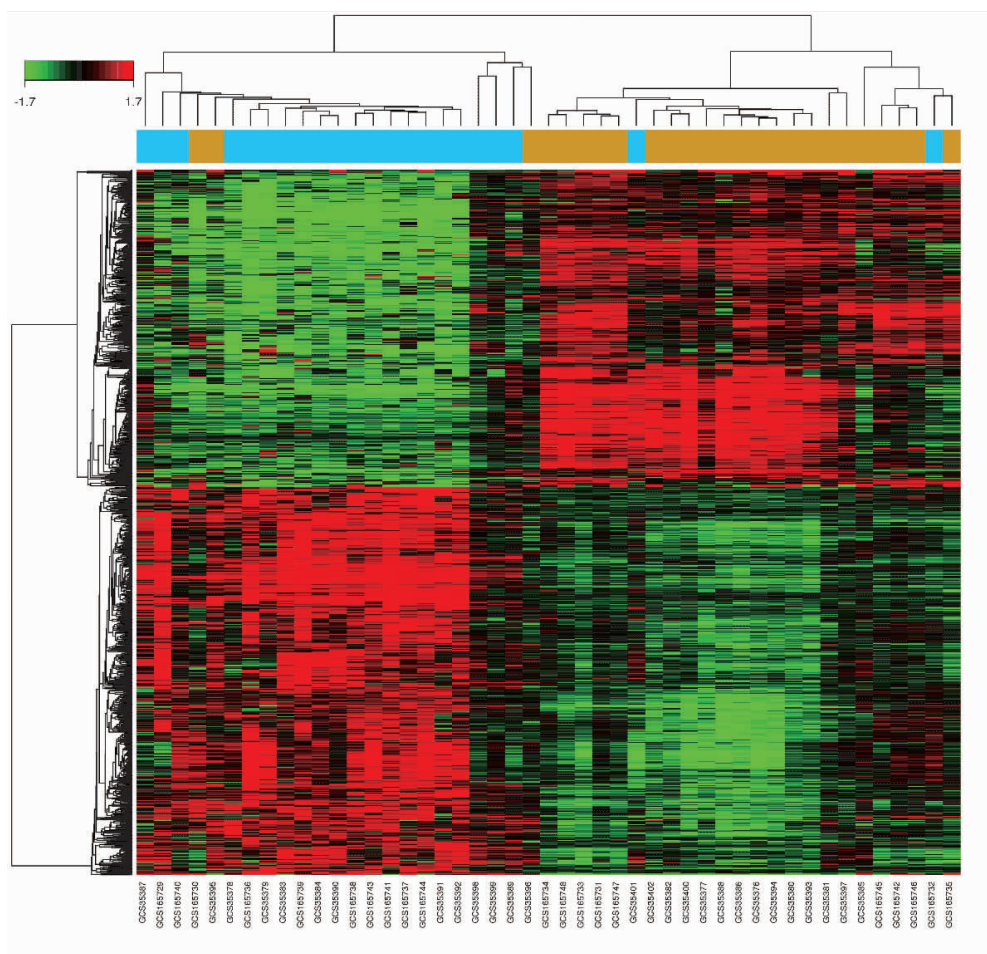


图 3 DEGs 聚类图

Figure 3. Dendrogram of DEGs

Blue indicates the gastric cancer group; Yellow indicates the control group; Red and green show differential gene expression in grouped samples; Red indicates that the expression value is high; Green indicates that the expression value is low.

表 1 差异表达基因(前 20 个)

Table 1. Differentially Expressed Genes (Top 20)

No.	Gene symbol	Accession number	Gene description	Fold change	P	Gene feature
1	<i>GKN2</i>	NM_182536	Gastrokine 2	-97.35	<0.001	down
2	<i>GKN1</i>	NM_019617	Gastrokine 1	-94.05	<0.001	down
3	<i>ATP4B</i>	NM_000705	ATPase, H + /K + exchanging, beta polypeptide	-67.70	<0.001	down
4	<i>GIF</i>	NM_005142	Gastric intrinsic factor (vitamin B synthesis)	-59.69	<0.001	down
5	<i>LIPF</i>	NM_001198828	Lipase, gastric	-59.33	<0.001	down
6	<i>PGA3</i>	NM_001079807	Pepsinogen 3, group I (pepsinogen A)	-57.18	<0.001	down
7	<i>ATP4A</i>	NM_000704	ATPase, H + /K + exchanging, alpha polypeptide	-40.43	<0.001	down
8	<i>PGC</i>	NM_001166424	Progastricsin (pepsinogen C)	-26.69	<0.001	down
9	<i>AQP4</i>	NM_001650	Aquaporin 4	-26.07	<0.001	down
10	<i>INHBA</i>	NM_002192	Inhibin, beta A	24.39	<0.001	up
11	<i>ESRRG</i>	NM_001134285	Estrogen-related receptor gamma	-22.13	<0.001	down
12	<i>DPCR1</i>	NM_080870	Diffuse panbronchiolitis critical region 1	-21.79	<0.001	down
13	<i>KCNE2</i>	NM_172201	Potassium voltage-gated channel, Isk-related family, member 2	-19.64	<0.001	down

(Table 1 continues on next page)

(Continued from previous page)

No.	Gene symbol	Accession number	Gene description	Fold change	P	Gene feature
14	KCNJ16	NM_001270422	Potassium inwardly-rectifying channel, subfamily J, member 16	-19.08	<0.001	down
15	VSIG1	NM_001170553	V-set and immunoglobulin domain containing 1	-17.89	<0.001	down
16	FUT9	NM_006581	Fucosyltransferase 9 (alpha (1,3) fucosyltransferase)	-17.16	<0.001	down
17	CHIA	NM_001040623	Chitinase, acidic	-16.64	<0.001	down
18	DPCR1	NM_080870	Diffuse panbronchiolitis critical region 1	-16.53	<0.001	down
19	PGC	NM_001166424	Progastricsin (pepsinogen C)	-16.22	<0.001	down
20	MFS4	NM_181644	Major facilitator superfamily domain containing 4	-15.34	<0.001	down

表 2 差异表达基因 GO 分析

Table 2. GO Analysis of Differentially Expressed Genes

GO ID	GO name	Gene counts	Enrichment score	P	FDR
GO:0007155	Cell adhesion	64	7.31	<0.001	<0.001
GO:0030198	Extracellular matrix organization	46	11.6	<0.001	<0.001
GO:0022617	Extracellular matrix disassembly	26	17.07	<0.001	<0.001
GO:0030574	Collagen catabolic process	24	17.28	<0.001	<0.001
GO:0030199	Collagen fibril organization	17	22.60	<0.001	<0.001
GO:0044281	Small molecule metabolic process	77	2.93	<0.001	<0.001
GO:0007586	Digestion	17	16.03	<0.001	<0.001
GO:0007596	Blood coagulation	41	4.57	<0.001	<0.001
GO:0001525	Angiogenesis	25	6.45	<0.001	<0.001

GO: Gene ontology; FDR: False discovery rate.

2.3 Pathway 分析及网络分析结果

通过对 DEGs 的 pathway 分析,我们可以确定富集 DEGs 的 pathway 项目,并确定可能与不同样本的 DEGs 相关的细胞通路变化,选择显著性最高的 10 条通路列表如表 3。在网络分析中,信号通路具有更多的相互作用,因此其重要性更高。我们确定了共 83 条差异调节的 pathway。这些途径包括 ECM-

受体相互作用、蛋白质消化吸收、胃酸分泌、PI3K-Akt 信号通路等。在通路网络分析中,共筛选出来 38 条通路,其中包括 3 条上调通路,2 条下调通路和 33 条上调/下调通路,寻找到 38 条通路中的 71 种关系(图 4),其中粘附力、糖酵解/糖异生、Wnt 信号通路、癌症的通路与上游和下游的相互作用较强。

表 3 差异表达基因 Pathway 分析

Table 3. Pathway Analysis of Differentially Expressed Genes

Pathway ID	Pathway name	Gene count	Enrichment score	P	FDR
4512	ECM-receptor interaction	25	14.90	<0.001	<0.001
4974	Protein digestion and absorption	21	12.37	<0.001	<0.001
4151	PI3K-Akt signaling pathway	35	5.23	<0.001	<0.001
4510	Focal adhesion	27	6.80	<0.001	<0.001
980	Metabolism of xenobiotics by cytochrome P450	18	11.67	<0.001	<0.001
982	Drug metabolism-cytochrome P450	16	11.21	<0.001	<0.001
1100	Metabolic pathways	62	2.70	<0.001	<0.001
5204	Chemical carcinogenesis	16	10.00	<0.001	<0.001
5146	Amoebiasis	16	7.61	<0.001	<0.001

FDR: False discovery rate; ECM: Extracellular matrix.

2.4 共表达网络分析及信号通路网络分析结果

我们挑选出来自 GO 和 Pathway 分析的 263 个重叠基因,将它们应用于基因共表达网络构建和信号通路网络分析。共表达网络分析图如图 5 所示,共 46

个节点,代表 136 个基因,分布以聚集性分布呈现,符合共表达理论。其中前 10 个基因的信息如表 4 所示。信号通路网络分析共有 123 个中枢基因,212 个关系,前 10 个中枢基因的信息如表 5 所示。

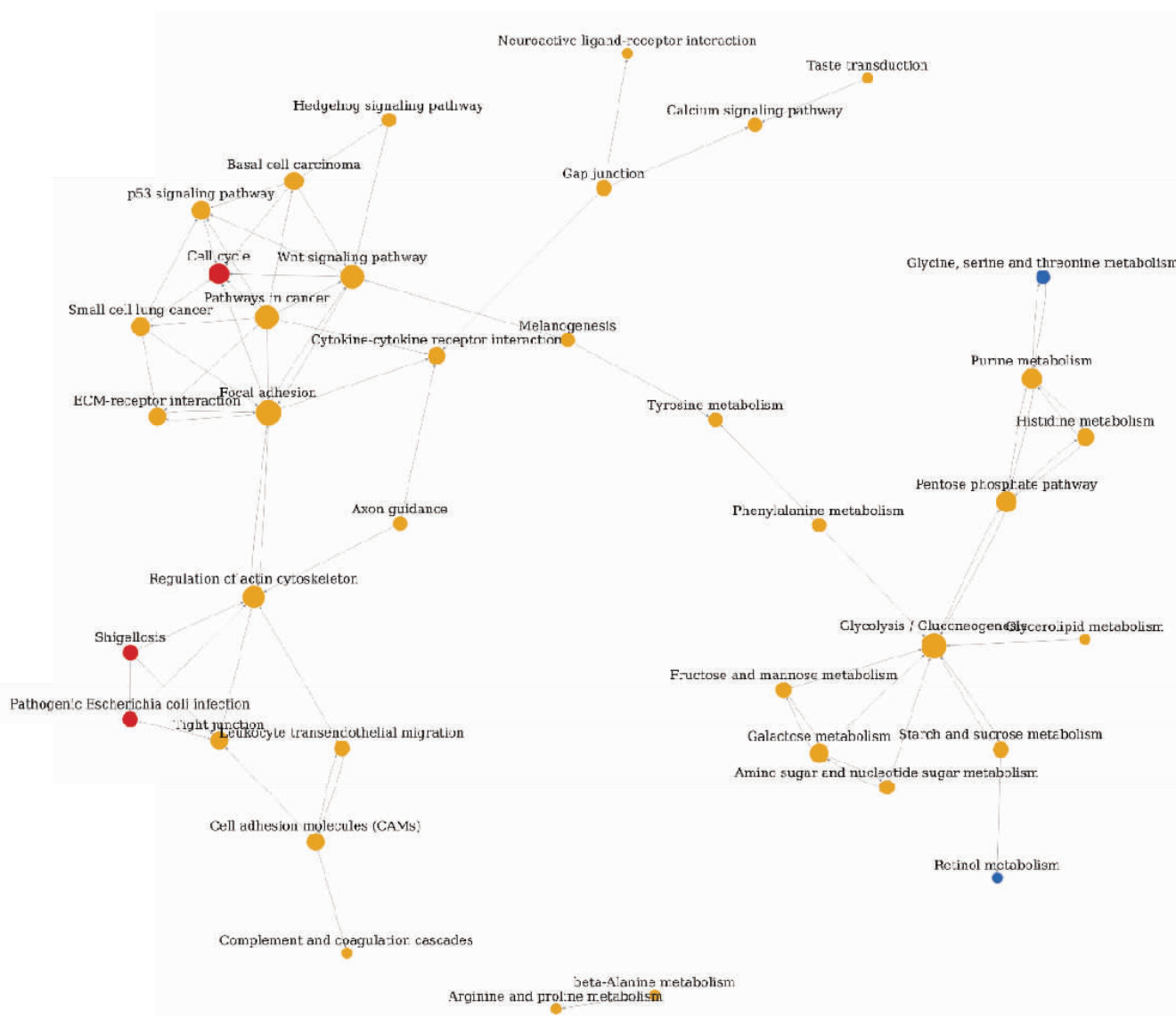


图 4 Pathway 网络分析图

Figure 4. Pathway and Network Analysis

Red represents up-regulation; Blue represents down-regulation; Yellow represents up-and-down-regulation. The larger the points are, the more signal paths with upstream and downstream interactions are.

表 4 差异表达基因共表达分析

Table 4. Co-Expression Analysis of Differentially Expressed Genes

No.	Gene symbol	Degree	Function	Description
1	SH3GL2	23	Coding	SH3-domain GRB2-like 2
2	CKMT2	18	Coding	Creatine kinase, mitochondrial 2 (sarcomeric)
3	CHIA	14	Coding	Chitinase, acidic
4	ATP4A	14	Coding	ATPase, H ⁺ /K ⁺ exchanging, alpha polypeptide
5	CCKBR	14	Coding	Cholecystokinin B receptor
6	TPCN2	14	Coding	Two pore segment channel 2
7	ATP4B	13	Coding	ATPase, H ⁺ /K ⁺ exchanging, beta polypeptide
8	KCNE2	13	Coding	Potassium voltage-gated channel, Isk-related family, member 2
9	CLCNKB	11	Coding	Chloride channel, voltage-sensitive Kb
10	KCNJ16	10	Coding	Potassium inwardly-rectifying channel, subfamily J, member 16

表 5 差异表达基因信号通路分析

Table 5. Analysis of Differentially Expressed Genes and Signaling Pathways

No.	Gene symbol	Gene description	Betweenness	Indegree	Outdegree
1	UGT2B15	UDP glucuronosyltransferase 2 family, polypeptide B15	793.9	12	12
2	ITGA2	integrin, alpha 2 (CD49B, alpha 2 subunit of VLA-2 receptor)	351.5	22	2
3	ITGB1	integrin, beta 1 (fibronectin receptor, beta polypeptide, antigen CD29 includes MDF2, MSK12)	351.5	22	2
4	CYP3A4	cytochrome P450, family 3, subfamily A, polypeptide 4	96.7	11	11
5	CYP3A5	cytochrome P450, family 3, subfamily A, polypeptide 5	96.7	11	11
6	GSTA1	glutathione S-transferase alpha 1	13.2	11	11
7	GSTA3	glutathione S-transferase alpha 3	13.2	11	11
8	GSTA4	glutathione S-transferase alpha 4	13.2	11	11
9	ALDH1A1	aldehyde dehydrogenase 1 family, member A1	77.8	9	9
10	PRKX	protein kinase, X-linked	894.0	4	4

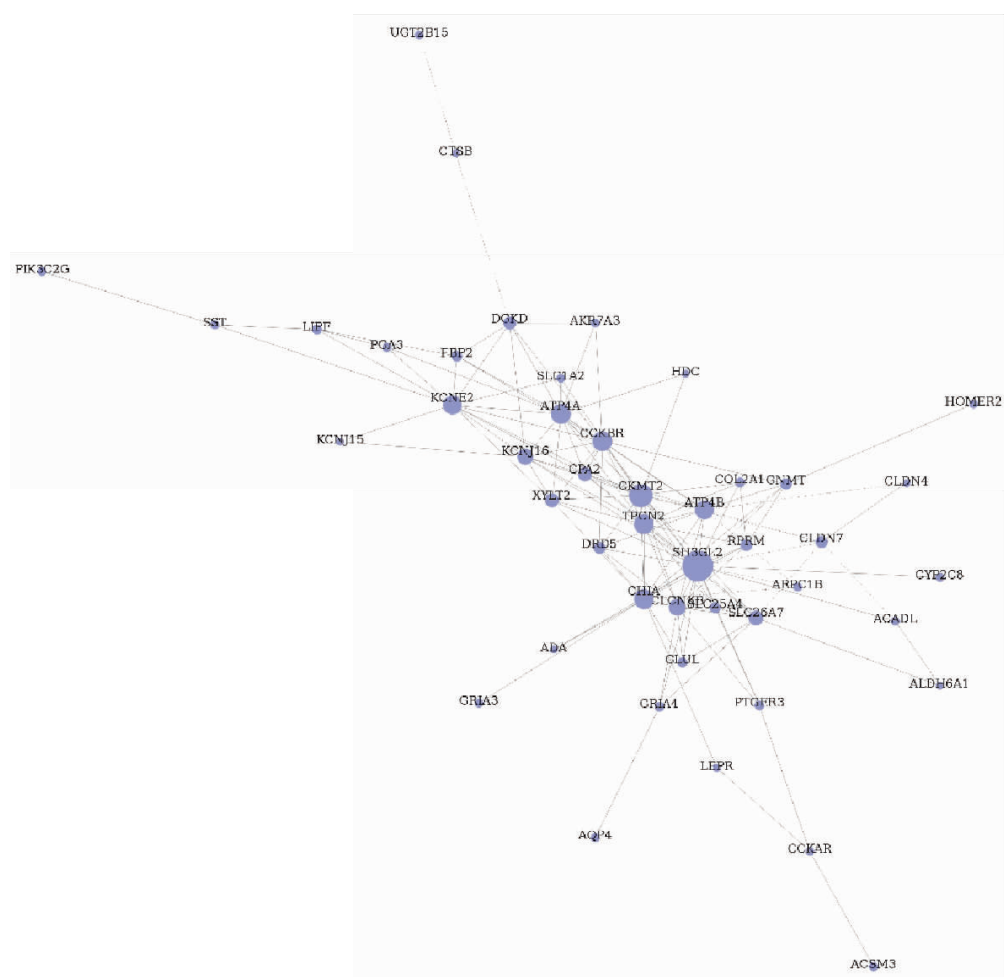


图 5 差异表达基因共表达网络图

Figure 5. Co-Expressing Network of Differentially Expressed Genes

Solid lines indicate positive correlations, and dashed lines indicate negative correlations. The larger the point is, the more genes it interacts with.

3 讨论

近年来,大量研究表明分子生物标志物在疾病

诊断、预后预测和靶向治疗中发挥重要作用。随着诊疗水平的不断进步,虽然过去的一个世纪里胃癌的发病率有所下降,但在世界范围内胃癌仍然是危

害极大的一种重大疾病。幽门螺旋杆菌感染、饮食不当、卫生条件落后和吸烟是常见的危险因素^[10]。另外,胃癌的诊断延误和转移扩散是胃癌死亡的主要原因。因此,寻找新的分子标志物对胃癌的早期诊断,针对性治疗和预后评估至关重要。

我们使用生物信息学分析筛选了涉及胃癌发生发展的 DEGs。结果表明,与正常组织相比,胃癌组织中数千个基因发生了变化,本研究共确定了 1 206 个 DEGs,包括 664 个上调基因和 542 个下调基因。我们对其进行了进一步的筛选,找出了具有显著表达变化的前 20 个基因,其中 *GKN2* 和 *GKN1* 的表达明显下调。Dokhaee 等^[11] 纳入 27 例胃癌组织样本,通过 Real-time PCR 对 *GKN2* 和 *GKN1* 进行定量分析,结果表明与正常组织相比,癌症标本中 *GKN1* 和 *GKN2* 基因表达显著降低 (P 值分别为 0.008, 0.004)。另外, Song 等^[12] 通过高通量计算分析发现 *GKN2* 在胃癌组织中明显下调,后在细胞实验中验证了 *GKN2* 在胃癌体外增殖和侵袭中具有抗癌作用的具体机制,即 *GKN2* 的体外过度表达使 PTEN/PI3K/AKT/mTOR 和 JAK/STAT 通路沉默,进而抑制胃癌增殖和侵袭,这与本文章的预测结果一致。具有显著表达变化的前 20 个基因中, *INHBA* 表达明显上调,据报道 *INHBA* 表达水平在癌组织中显著高于在邻近正常粘膜中的表达水平,并且与低表达病例相比, *INHBA* 高表达与较差的 5 年存活率相关, *INHBA* 被证实是一个独立的预后因素^[13]。

GO 富集分析显示, DEGs 主要参与细胞粘附、细胞外基质、细胞外基质分解、胶原分解代谢、胶原纤维、血管生成等方面。Pathway 分析发现,差异最显著的富集通路主要涉及细胞外基质-受体相互作用、蛋白质消化吸收、PI3K-Akt 信号通路、细胞粘附、代谢途径、癌症通路等。将基因间相互作用形成网络后可见,胃癌组织和正常组织之间的差异也主要集中在细胞粘附、Wnt 信号通路、癌症通路等方面,以此可推测胃癌的发生发展与细胞外基质、细胞粘附、癌症通路的相关性较大。

基于 GCBI 平台的信号通路网络分析解构了 KEGG 数据库,突破了 KEGG-Pathway 数据库中基因间相互作用关系局限于某 Pathway 的限制,是在全 KEGG-Pathway 数据库的范围内筛选某个蛋白的上游或下游蛋白。差异表达基因的信号通路相互作用分析显示, *UGT2B15* 作为中枢基因,上下游各与 12 个基因有联系,该基因涉及的功能主要有外源性化合物代谢过程、细胞葡萄糖醛酸化、类固醇代谢过

程、黄酮类生物合成过程。葡萄糖醛酸化是消除外源性和内源性化合物(如环境致癌物质和体内雄激素)的主要途径,有研究表明, *UGT2B15* 的基因多态性与前列腺癌^[14] 和胃癌^[15] 的发生发展有关,可能是致癌基因,并且是癌症的治疗靶标。 *ITGA2* 和 *ITGB1* 均有 22 个上游基因和 2 个下游基因,属于整合素家族,整合素属于细胞粘附受体家族成员,是一类广泛分布于细胞表面的跨膜糖蛋白,多项研究证实整合素涉及多个肿瘤的侵袭转移,如 Lahlou 等^[16] 通过转基因小鼠肿瘤模型提供了较为明确的 *ITGB1* 在乳腺癌细胞侵袭与转移中起重要作用的体内实验证据。Chuang 等^[17] 发现 *ITGA2* 不仅抑制胃癌细胞迁移,还诱导胃癌细胞的凋亡,并且 *ITGA2* 的较高表达水平在抗 *ITGA2* 抗体治疗期间对凋亡进展的有着显著影响。 *GSTA1*、*GSTA3* 和 *GSTA4* 均有 11 个上游基因和 11 个下游基因,属于 *GST* 基因家族, *GST* 基因家族通过还原型谷胱甘肽与许多底物(如药物和环境污染物的结合,编码对某些生命过程以及解毒和毒性机制至关重要的基因。 *GST* 基因在氧化应激反应中被上调,并且在许多肿瘤中会过度表达,增强或减弱机体对化疗药物的反应^[18]。

共表达网络分析是通过差异基因组成的表达谱数据来构建基因共表达网络。共表达网络能够层次清晰地展现基因间的相互关系,更深入地发现关键调控基因以及相互作用脉络,可以揭示在健康状态和疾病状态之间存在差异的基因,并且有助于揭示潜在的疾病和其他表型调节因子,或揭示一些新的基因相互作用关系^[19]。如 Kong 等^[20] 通过共表达网络分析确定了与肝细胞癌的进展和预后相关的五个核心基因。差异表达基因的共表达网络分析显示 *SH3GL2*、*CKMT2*、*CHIA*、*ATP4A*、*CCKBR*、*TPCN2* 分别有 23、18、14、14、14、14 个基因呈现共表达趋势,这些基因的作用主要体现在信号转导、物质代谢过程、炎症反应的介质分子、免疫应答、跨膜运输等方面。有文献报道, *SH3GL2* 的敲除促进了胶质瘤细胞的迁移和侵袭,而 *SH3GL2* 的过表达则抑制了它们的迁移和侵袭。进一步的分子机制研究发现, *SH3GL2* 的沉默明显激活了 *STAT3* 信号,从而促进了 *MMP2* 的表达和分泌。 *SH3GL2* 通过负向调控 *STAT3*/*MMP2* 信号通路抑制胶质瘤细胞的迁移和侵袭行为, *SH3GL2* 的缺失可能会增强 *STAT3*/*MMP2* 信号通路,从而促进胶质瘤细胞的迁移和侵袭^[21]。 *CKMT2* 是维持正常能量代谢时必不可少的,有研究表明 *CKMT2* 可能是参与骨肉瘤形成的关键调节因

子^[22]。因此,共表达网络分析得到的 *SH3GL2*、*CKMT2*、*CHIA*、*ATP4A*、*CCKBR*、*TPCN2* 等节点基因,或许是胃癌进展过程中的关键基因,但这主要是基于数据推测,需要进一步实验验证。

总之,我们目前的研究提供了一个全面的生物信息学分析,包括 DEGs 的筛选,以及对关注基因进行了 GO 分析、Pathway 分析、基因信号通路网络分析、共表达分析,得到可能与胃癌发生、演进相关的基因。目前的结果可以促进我们对胃癌相关潜在分子机制的理解。例如 *INHBA*、*UGT2B15*、*ITGA2*、*ITGB1*、*SH3GL2* 等基因,及其相关的生物过程如细胞粘附、胶原纤维、细胞外基质-受体相互作用等,或可提供胃癌诊断和治疗的潜在靶标。

作者声明:本文全部作者对于研究和撰写的论文出现的不端行为承担相应责任;并承诺论文中涉及的原始图片、数据资料等已按照有关规定保存,可接受核查。

学术不端:本文在初审、返修及出版前均通过中国知网(CNKI)科技期刊学术不端文献检测系统的学术不端检测。

同行评议:经同行专家双盲外审,达到刊发要求。

利益冲突:所有作者均声明不存在利益冲突。

文章版权:本文出版前已与全体作者签署了论文授权书等协议。

[参考文献]

- [1] Torrel A, Bray F, Siegel RL, et al. Global cancer statistics, 2012[J]. CA Cancer J Clin, 2015, 65(2):87-108.
- [2] Sitarz R, Skierucha M, Mielko J, et al. Gastric cancer: Epidemiology, prevention, classification, and treatment[J]. Cancer Manag Res, 2018, 10:239-248.
- [3] Roberta DA, Milena S, Coleman P, et al. Cancer survival in Europe 1999-2007 by country and age: Results of EURO-CARE-5-a population-based study[J]. Lancet Oncol, 2014, 15(1):23-34.
- [4] 罗会芹,何义富. 晚期胃癌转化治疗策略的研究进展[J]. 肿瘤预防与治疗, 2019, 32(8):718-724.
- [5] Li ML, Zhang JC, Li SG, et al. Characteristic gene expression profiles in the progression from normal gastric epithelial cells to moderate gastric epithelial dysplasia and to gastric cancer[J]. Chin Med J (Engl), 2012, 125(10):1777-1783.
- [6] 肖毓,张洪攀,谭榜宪. 基于 Oncomine 数据库荟萃分析 COL1A2 在胃癌患者中的表达及其临床意义[J]. 肿瘤预防与治疗, 2019, 32(6):503-510.
- [7] He J, Jin Y, Chen Y, et al. Downregulation of ALDOB is associated with poor prognosis of patients with gastric cancer[J]. Onco Targets Ther, 2016, 9:6099-6109.
- [8] Wang Q, Wen YG, Li DP, et al. Upregulated INHBA expression is associated with poor survival in gastric cancer[J]. Med Oncol, 2012, 29(1):77-83.
- [9] Roos DS. Computational biology. Bioinformatics trying to swim in a sea of data[J]. Science, 2001, 291(5507):1260-1261.
- [10] Yang D, Hendifar A, Lenz C, et al. Survival of metastatic gastric cancer: Significance of age, sex and race/ethnicity[J]. J Gastrointest Oncol, 2011, 2(2):77-84.
- [11] Dokhaee F, Mazhari S, Galehdari M, et al. Evaluation of GKN1 and GKN2 gene expression as a biomarker of gastric cancer[J]. Gastroenterol Hepatol Bed Bench, 2018, 11(S1):S140-S145.
- [12] Song WM, Lin X, Liao X, et al. Multiscale network analysis reveals molecular mechanisms and key regulators of the tumor micro-environment in gastric cancer. [J]. Int J Cancer, 2020, 146(5):1268-1280.
- [13] Katayama Y, Oshima T, Sakamaki K, et al. Clinical significance of INHBA gene expression in patients with gastric cancer who receive curative resection followed by adjuvant S-1 chemotherapy [J]. In Vivo, 2017, 31(4):565-571.
- [14] Chen X, Li D, Wang N, et al. Bioinformatic analysis suggests that UGT2B15 activates the HippoYAP signaling pathway leading to the pathogenesis of gastric cancer[J]. Oncol Rep, 2018, 40(4):1855-1862.
- [15] Habibi M, Mirfakhraie R, Khani M, et al. Genetic variations in UGT2B28, UGT2B17, UGT2B15 genes and the risk of prostate cancer: A case-control study[J]. Gene, 2017, 634:47-52.
- [16] Lahlou H, Muller WJ. β 1-integrins signaling and mammary tumor progression in transgenic mouse models: Implications for human breast cancer[J]. Breast Cancer Res, 2011, 13(6):229.
- [17] Chuang YC, Wu HY, Lin YL, et al. Blockade of ITGA2 induces apoptosis and inhibits cell migration in gastric cancer[J]. Biol Proced Online, 2018, 20:10.
- [18] Nebert DW, Vasiliou V. Analysis of the glutathione S-transferase (GST) gene family[J]. Hum Genomics, 2004, 1(6):460-464.
- [19] van Dam S, Vösa U, van der Graaf A, et al. Gene co-expression analysis for functional classification and gene-disease predictions [J]. Brief Bioinform, 2018, 19(4):575-592.
- [20] Kong J, Wang T, Zhang Z, et al. Five core genes related to the progression and prognosis of hepatocellular carcinoma identified by analysis of a coexpression network[J]. DNA Cell Biol, 2019, 38(12):1564-1576.
- [21] Zhu Y, Zhang X, Wang L, et al. Loss of SH3GL2 promotes the migration and invasion behaviours of glioblastoma cells through activating the STAT3/MMP2 signalling[J]. J Cell Mol Med, 2017, 21(11):2685-2694.
- [22] Wang H, Tang M, Ou L, et al. Biological analysis of cancer specific microRNAs on function modeling in osteosarcoma[J]. Sci Rep, 2017, 7(1):5382.