**Advances in Clinical Medicine** 临床医学进展, 2020, 10(12), 2883-2889
Published Online December 2020 in Hans. http://www.hanspub.org/journal/acm
https://doi.org/10.12677/acm.2020.1012436

Hans汉斯

# 基于生物信息学方法对胰腺癌差异基因表达的分析

**王　朦[1]，朱梦迪[2]，许帮亮[1]，蒋桂星[3*]**

[1]浙江大学医学院，浙江　杭州
[2]大连医科大学医学院，辽宁　大连
[3]浙江大学医学院附属邵逸夫医院肝胆胰外科，浙江　杭州
Email: *2310036@zju.edu.cn, 21818380@zju.edu.cn

收稿日期：2020年11月15日；录用日期：2020年12月3日；发布日期：2020年12月10日

## 摘　要

**目的：该文旨在利用生物信息分析方法研究胰腺癌组织的差异表达基因(DEGs)，为进一步实验以及胰腺癌的诊疗和提供新途径。方法：分析来自美国国立生物技术信息中心(NCBI)的公共基因芯片数据库(GEO)中基因芯片数据(GSE15471、GSE16515)，其包括了胰腺癌及对应正常组织的基因表达数据，利用GEO在线分析工具GEO2R分别对两个基因芯片进行分析，初步筛选胰腺癌与正常组织的相关DEGs。对初步筛选出的两组相关差异表达基因进行Veen分析，得到显著DEGs。相关基因功能富集分析由DAVID在线数据库分析得到。同时，通过String在线数据库构建蛋白互作网络(PPI)并使用Cytoscape软件分析PPI网络。结果：设定GEO2R分析参数(|logFC| ≥ 2.0, P < 0.05)，并进行Veen分析，共筛选出111个显著DEGs，其中显著上调DEGs 84个，显著下调DEGs 27个。富集分析显示显著DEGs参与的主要生物过程有：胶原蛋白分解代谢、细胞外基质组织、胶原原纤维组织、细胞外基质分解；细胞成分有：细胞外隙、胞外区、胞外区、细胞外基质外来体；分子功能有：细胞外基质结构成分、钙离子结合、肝素结合。KEGG (Kyoto Encyclopedia of Genes and Genomes)通路分析显示，显著DEGs主要富集于癌症、小细胞肺癌、蛋白质消化吸收、胰腺分泌和PI3K-Akt信号通路。对显著DEGs编码的蛋白质所构建的PPI分析，发现COL1A2、ALB、COL12A1、COL5A1、COL5A2、COL11A1、MMP1、ITGA2、COL8A1SKA1等9个关键蛋白，由此确定关键DEGs。结论：我们的研究提示了胰腺癌与正常胰腺组织间关键DEGs，并发掘关键DEGs的相互作用关系，为胰腺癌的后续研究和诊疗提供新的方向。**

## 关键词

**胰腺癌组织，人类，表达**

# Expression Analysis of Different Genes in Pancreas Cancer Using Bioinformatic

**Meng Wang[1], Mengdi Zhu[2], Bangliang Xu[1], Guixing Jiang[3*]**

*通讯作者。

文章引用：王朦，朱梦迪，许帮亮，蒋桂星．基于生物信息学方法对胰腺癌差异基因表达的分析[J]．临床医学进展，2020, 10(12): 2883-2889. DOI: 10.12677/acm.2020.1012436

[1]School of Medicine, Zhejiang University, Hangzhou Zhejiang
[2]School of Medicine, Dalian Medical University, Dalian Liaoning
[3]Department of Hepatopancreatobiliary Surgery, Sir Run Run Shaw Hospital, School of Medicine, Zhejiang University, Hangzhou Zhejiang
Email: *2310036@zju.edu.cn, 21818380@zju.edu.cn

Received: Nov. 15th, 2020; accepted: Dec. 3rd, 2020; published: Dec. 10th, 2020

## Abstract

**Objective: The differentially expressed genes (DEGs) in pancreatic cancer tissues were studied by bioinformatics analysis, which provided a new way for further experiments and diagnosis and treatment of pancreatic cancer. Methods: Analysis from the national center for biotechnology information (NCBI) public gene chip databases (GEO) gene chip data (GSE15471, GSE16515), which include the pancreatic cancer and gene expression data of corresponding normal tissues using GEO online analytical tools GEO2R respectively to analyze two gene chip, preliminary screening of pancreatic cancer and normal tissue DEGs. Veen analysis was performed on the two groups of DEGs, and significant DEGs were obtained. Functional enrichment analysis of prominent DEGs was obtained by DAVID online database analysis. Meanwhile, protein interaction network (PPI) was constructed through String online database and ANALYZED by Cytoscape software. Results: A total of 111 prominent DEGs were screened out, including 84 upregulated and 27 downregulated genes (|logFC| ≥ 2.0, P < 0.05). Enrichment analysis showed that prominent DEGs were involved in the following biological processes: collagen catabolism, extracellular matrix tissue, collagen fibril tissue, extracellular matrix decomposition. The cell components include: extracellular space, extracellular region, extracellular region and extracellular matrix. Molecular functions include: extracellular matrix structural components, calcium ion binding, heparin binding. Results of KEGG pathway analysis showed that prominent DEGs were mainly concentrated in cancer, small cell lung cancer, protein digestion and absorption, pancreatic secretion and PI3K-Akt signaling pathways. PPI analysis of proteins encoded with significant DEGs revealed nine key proteins, including COL1A2, ALB, COL12A1, COL5A1, COL5A2, COL11A1, MMP1, ITGA2, and COL8A1SKA1, thus determining the key differential genes. Conclusion: Our study indicated that the key differentially expressed genes between pancreatic cancer and normal pancreatic tissue, and explored the interaction relationship between the key DEGs, providing a new direction for the follow-up research and diagnosis of pancreatic cancer.**

## Keywords

**Pancreas Cancer Tissue, Human, Express**



## 1. 引言

胰腺癌作为一种发病率逐年升高的恶性消化腺肿瘤，其恶性程度高，预后差，5 年生存率仅为 3%~6%。在美国，癌症相关死亡的最常见原因中胰腺癌排名第四，由于胰腺癌发病隐匿，有效的早期诊断方法及治疗手段缺乏，导致胰腺癌患者中 85.3%的死亡发生在新诊断病例中，在所有常见恶性肿瘤中居首位[1]

[2]。探索新的早期诊断和治疗靶点对胰腺癌的诊疗是十分迫切和必要的。已有研究表明胰腺癌的发生与基因突变，以及突变基因间相互密切作用有关，抑癌基因的沉默，原癌基因的显著表达，促进正常胰腺细胞向癌细胞转化过程。基因芯片是高通量信息数据集合体，记录了生物遗传信息。利用基因芯片，通过在线软件对已公布的胰腺癌基因数据进行分析探索，能得到胰腺癌与其对应正常胰腺组织的差异表达基因(DEGs)，为寻求疾病早期诊断及治疗靶点提供依据。本研究仔细筛选了公共基因芯片数据库(GEO)中的基因芯片，得到胰腺癌及其对应正常胰腺组织的基因芯片，进一步分析挖掘出胰腺癌关键差异表达基因。

## 2. 资料与方法

### 2.1. 资料

以“胰腺癌组织、人类、表达”为关键词，在美国国立生物技术信息中心(NCBI)的开放信息库 GEO 中共检索到了 560 个关于人类胰腺癌的基因芯片数据(https://www.ncbi.nlm.nih.gov/geo/) [3]。根据同公司实验平台分析所得基因芯片且包含相应正常胰腺样本基因芯片为标准，经过进一步仔细研究排除，我们下载得到，并分析了进行分析的编号 GSE15471、GSE16515 的基因芯片数据。编号 GSE15471 基因芯片数据包含了 72 个样本数据，包括 36 个胰腺癌样本，以及 36 个对照样本。编号 GSE16515 基因芯片数据包含了 52 个样本，包括 36 个癌样本和 16 个对照样本。

### 2.2. 方法

#### 2.2.1. 显著 DEGs 分析

GEO2R 为 GEO 数据库提供的基于 R 语言设计的在线分析软件。在进行初步分析时，我们设定了参数($|\log FC| \geq 2.0$, $P < 0.05$)，分别对编号 GSE15471、GSE16515 基因芯片原始数据进行分析理。之后，通过 Ecel 软件协助，初步分析筛选出胰腺癌组织 DEGs [3]。接着，使用 Venn diagram webtool 进行 Venn 分析，Venn 分析可识别 GSE15471、GSE16515 两个基因芯片数据分析初步所得 DEGs 的重叠部分，最终获得的此重叠部分定义为显著 DEGs。

#### 2.2.2. 显著 DEGs 的 GO 功能及 KEGG 通路富集分析

大规模功能富集研究已被应用于显著 DEGs 的研究，以了解显著 DEGs 的出现是由哪些功能过程所引起。其中，GO 功能分析由生物过程(BP)、细胞成分(CC)和分子功能(MF)三个部分组成。KEGG 作为公开生物数据储存整合了基因组、生物途径、疾病等信息数据，通过与世界上其他数据库连接，旨在揭示及绘制生命现象蓝图。在 DAVID 对上述筛选出的显著 DEGs 行 GO、KEGG 通路富集分析[4]。

#### 2.2.3. 差异表达基因的蛋白－蛋白相互作用网络(PPI)的构建

STRING 在线数据库对显著 DEGs 进行 PPI 绘制，并用 Cytoscape 软件对绘制 PPI 分析，得到关键蛋相互作用关系，以筛选出关键 DEGs [5]。

## 3. 结果

### 3.1. 显著差异基因筛选

本研究选择了两个基因芯片(GSE15471、GSE16515)，其中，GSE15471 胰腺癌组织样本 36 例，正常胰腺组织样本 36 例；GSE16515 为胰腺癌组织样本 36 例，正常胰腺组织样本 16 例(表 1)。通过比较胰腺癌组织样本和正常乳腺样本，利用 GEO2R 在线分析工具对两组样本进行分析处理，根据设定的参数($|\log FC| \geq 2.0$, $P < 0.05$)，GSE15471 中筛选出 239 个 DEGs，上调差异基因 205 个，下调差异基因 34 个；

GSE16515 中筛选出 396 个 DEGs，上调差异基因 301 个，下调差异基因 95 个。对两组基因芯片筛选出的差异基因进行 Venn 分析，得到两组 DEGs 的交集(图 1)。最后，确定交集中 111 个基因作为显著 DEGs，其中 84 个显著 DEGs 上调，27 个显著 DEGs 下调。

**Table 1.** Two pancreatic cancer microarray databases derived from the GEO
**表 1.** GEO 数据库胰腺癌基因芯片样本

| 基因芯片编号 | 胰腺癌样本量 | 正常组织样本量 | 总量 |
|---|---|---|---|
| GSE15471 | 36 | 36 | 72 |
| GSE16515 | 36 | 16 | 52 |

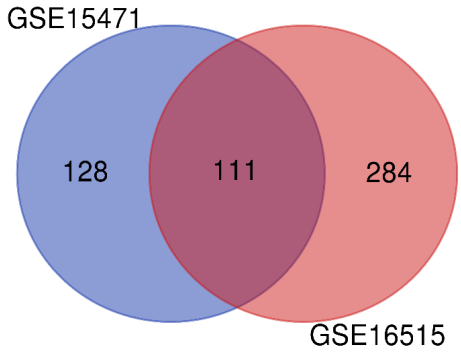



**Figure 1.** Venn analysis of DEGs
**图 1.** DEGs 维恩分析

## 3.2. 显著 DEGs 的 GO 功能及 KEGG 通路富集分析

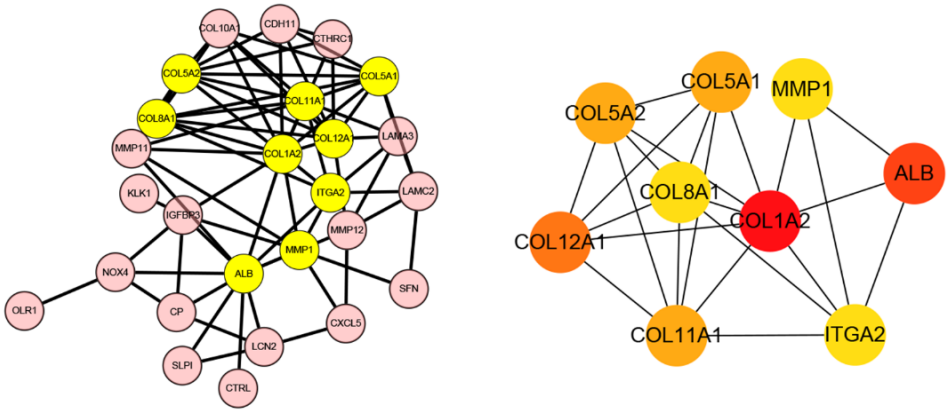
DAVID 在线数据库分析后，显著 DEGs 的 GO 功能富集主要在生物过程：胶原蛋白分解代谢、细胞外基质组织、胶原原纤维组织、细胞外基质合成与分解；细胞成分：细胞外间隙、胞外区、外泌体；分子功能：细胞外基质结构、钙离子与肝素结合(表 2)。KEGG 通路富集分析，胰腺癌显著 DEGs 富集于癌症、蛋白质消化吸收、胰腺分泌和 PI3K-Akt 信号通路、小细胞肺癌。

**Table 2.** Gene ontology (GO) function analysis of differentially expressed genes
**表 2.** 差异表达基因的 GO 功能分析

| 分类 | GO 术语 | 名称 | 富集基因数 | P |
|---|---|---|---|---|
| 生物过程 | GO: 0030574 | 胶原蛋白分解代谢 | 12 | 7.7E−14 |
| | GO: 0030198 | 细胞外基质组织 | 15 | 6.5E−12 |
| | GO: 0030199 | 胶原原纤维组织 | 8 | 2.3E−9 |
| | GO: 0022617 | 细胞外基质分解 | 8 | 2.7E−7 |
| 细胞成分 | GO: 0005615 | 细胞外隙 | 37 | 5.2E−16 |
| | GO: 0005576 | 胞外区 | 35 | 3.6E−12 |
| | GO: 0005578 | 细胞外基质 | 15 | 1.1E−9 |
| | GO: 0070062 | 外泌体 | 41 | 4.7E−9 |
| 分子功能 | GO: 0005201 | 细胞外基质结构成分 | 7 | 1.9E−6 |
| | GO: 0005509 | 钙离子结合 | 17 | 2.0E−6 |
| | GO: 0008201 | 肝素结合 | 8 | 3.1E−5 |

### 3.3. 蛋白互作网络(PPI)与关键差异表达基因筛选

经过 STRING 在线数据库的挖掘绘制，从而得到胰腺癌显著 DEGs 所表达的蛋白质之间的相互作用关系。图 2 展示我们研究发现的 54 个节点蛋白以及 158 组蛋白作用关系。紧接着，将 STRING 分析结果下载导入 Cytoscape 软件，得出 COL1A2、ALB、COL12A1、COL5A1、COL5A2、COL11A1、MMP1、ITGA2、COL8A1 等 9 个胰腺癌组织与正常胰腺组织相关 DEGs 编码的蛋白质为主要相互作用的蛋白。编码以上蛋白的基因被确定为关键 DEGs，胰腺癌关键 DEGs 的相互联系方式见图 2(B)所示。




(A) (B)

**Figure 2.** (A) The protein-protein interaction network of the differentially expressed genes, yellow is the 9 key proteins mentioned in the paper, and pink is the remaining protein; (B) The protein-protein interaction network of the hub genes
**图 2.** (A) 显著 DEGs 的 PPI，图中黄色为文中所说的 9 个关键蛋白，粉色为剩余的蛋白；(B) 关键蛋白互作网络

## 4. 讨论

近年来，胰腺癌已成为世界上发病率和死亡率增长最快的恶性肿瘤，5 年生存率仅为 3%~6% [6]。胰腺癌早期无特异性症状，并可发生早期得胰外扩散，同时由于针对胰腺癌早期诊断方法的缺乏。这意味着超过 3/4 的胰腺癌患者因未得到有效的早期诊断，导致癌细胞转移，从而失去了有效治愈并可能延长生存期的早期根治性手术治疗机会。由此可见，找到胰腺癌早期诊断以及新的治疗靶点是十分必要的。因此，本研究通过在 GEO 数据库里检索相关信息，从而获取胰腺癌以及正常胰腺组织的基因表芯片数据，通过 GEO2R 在线工具、Veen 分析处理数据，结合 DAVID、STRING 等生物信息数据库以及生物信息学方法对二者的 DEGs 进行分析，最后得到 COL1A2、ALB、COL12A1、COL5A1、COL5A2、COL11A1、MMP1、ITGA2、COL8A1 等显著 DEGs，相较于正常胰腺组织，这些关键 DEGs 表达显著上调(原癌基因)或下调(抑癌基因)。根据分析所得胰腺癌关键 DEGs 功能富集分析了解到，其主要参与胶原蛋白分解代谢、细胞外基质组织、细胞外基质分解、钙离子结合、肝素结合等功能。既往有研究和证据表明，本文筛选出这些显著 DEGs 的 GO、KEGG 富集通路在胰腺癌的发生发展中可能发挥重要作用。

在筛选到的 9 个关键 DEGs 中，COL1A2、ALB、MMP1、ITGA2 是我们非常感兴趣的基因。既往研究表明 MMP1、ITGA2 的高表达与胰腺导管腺癌不良预后有关[7] [8]。MMP1 与胆碱降解的功能有关，并有可能参与胰岛素诱导的胰腺癌的发生[9] [10]。并且有研究发现，ITGA2 基因高表达提高了前列腺、甲状腺癌细胞的侵袭性、增值与迁移能力[11] [12]。COL1A2 中的单核苷酸突变会引起基因翻译缺陷疾病，其主要表现包括骨缺乏症[13] [14]。有研究表明，COL5A1 基因缺乏症小鼠肝脏转氨酶增加，提示肝脏对

炎症反应的敏感性和肝脏免疫反应的改变。关于 ALB，研究提示 ALB 基因表达上调使免疫功能减弱[15] [16]。我们所得其余关键基因 COL12A1 异常表达与结直肠癌的癌变相关，其也是胃癌细胞体外培养高表达基因，COL12A1 基因的显著高表达，引起细胞外基质突变，在一定程度上促进了组织癌变的过程，对癌症的发生发展可能起到作用[17] [18]。COL5A2 基因在胃癌与宫颈癌中的高表达，引起III型和Ⅴ型胶原蛋白的表达水平异常，常示患者预后不佳[19]。COL8A1 在乳腺癌患者中表达高于正常对照组，尤其在三阴性乳腺癌中表达升高，其富集于蛋白多糖途径，一项研究表明，COL8A1 的高达与乳腺癌患者较差的预后有关，其高表达患者的总体生存率较 COL8A1 低表达患者生存预后差[20]。COL11A1 的肿瘤研究少，但 COL11A1 与 2 型 Stickler 综合征相关。COL11A1 是位于染色体 12p13.31 的非缩合蛋白 I 复合物亚基 2，在骨髓、脂肪组织中表达，参与在染色体浓缩的大蛋白质复合物。COL11A1 与在胰腺癌的发生发展以及诊断与治疗中功能现无明确研究[21]。

综上所述，在文研究筛选出的显著 DEGs 中 MMP1、ITGA2 的高表达与胰腺导管腺癌不良预后有关，这与我们的研究结果一致。我们得到的显著 DEGs 如 COL12A1、COL5A1、COL5A2、COL11A1、COL8A1 虽然其在各种疾病中都发挥着重要的作用，但是需要后续的临床病例验证和体内外实验研究去证实其在胰腺癌发生发展以及治疗中价值。但其可能作为潜在的关键生物标志物和治疗靶点，为进一步的实验研究提供参考信息和重要的研究线索。

## 参考文献


[1] Siegel, R.L., Miller, K.D. and Jemal, A. (2015) Cancer Statistics, 2015. *CA*: *A Cancer Journal for Clinicians*, **65**, 5-29. https://doi.org/10.3322/caac.21254

[2] Wu, C., Li, M., Meng, H., Liu, Y., Niu, W., Zhou, Y., *et al*. (2019) Analysis of Status and Countermeasures of Cancer Incidence and Mortality in China. *Science China Life Sciences*, **62**, 640-647. https://doi.org/10.1007/s11427-018-9461-5

[3] Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., *et al*. (2013) NCBI GEO: Archive for Functional Genomics Data Sets—Update. *Nucleic Acids Research*, **41**, D991-D995. https://doi.org/10.1093/nar/gks1193

[4] Huang da, W., Sherman, B.T. and Lempicki, R.A. (2009) Bioinformatics Enrichment Tools: Paths toward the Comprehensive Functional Analysis of Large Gene Lists. *Nucleic Acids Research*, **37**, 1-13. https://doi.org/10.1093/nar/gkn923

[5] Szklarczyk, D., Morris, J.H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., *et al*. (2017) The STRING Database in 2017: Quality-Controlled Protein-Protein Association Networks, Made Broadly Accessible. *Nucleic Acids Research*, **45**, D362-d8. https://doi.org/10.1093/nar/gkw937

[6] Saika, K. and Sobue, T. (2013) Cancer Statistics in the World. *Gan to Kagaku Ryoho*. *Cancer & Chemotherapy*, **40**, 2475-2480.

[7] 李孟畔, 沙磊. 胰腺导管腺癌相关基因的生物信息学分析及功能预测[J]. 解剖科学进展, 2020, 26(5): 540-543+9.

[8] Wang, S.S., *et al*. (2020) miR-216a-Mediated Upregulation of TSPAN1 Contributes to Pancreatic Cancer Progression via Transcriptional Regulation of ITGA2. *American Journal of Cancer Research*, **10**, 1115-1129.

[9] 岳天翔, 易燕, 黄冬琴. 抗阻运动对胰岛素抵抗大鼠胰岛素敏感性和肝脏细胞外基质的调节作用[J]. 基因组学与应用生物学, 2020, 39(7): 3279-3285.

[10] Zhou, H., *et al*. (2020) Identification of MMP1 as a Potential Gene Conferring Erlotinib Resistance in Non-Small Cell Lung Cancer Based on Bioinformatics Analyses. *Hereditas*, **157**, 32. https://doi.org/10.1186/s41065-020-00145-x

[11] Qin, A., Liu, Q. and Wang, J. (2020) Ropivacaine Inhibits Proliferation, Invasion, Migration and Promotes Apoptosis of Papillary Thyroid Cancer Cells via Regulating ITGA2 Expression. *Drug Development Research*, **81**, 700-707. https://doi.org/10.1002/ddr.21671

[12] Gaballa, R., Ali, H.E.A., Mahmoud, M.O., Rhim, J.S., Ali, H.I., Salem, H.F., *et al*. (2020) Exosomes-Mediated Transfer of Itga2 Promotes Migration and Invasion of Prostate Cancer Cells by Inducing Epithelial-Mesenchymal Transition. *Journal of Cancer*, **12**, 2300. https://doi.org/10.3390/cancers12082300

[13] Sini, S., Han, X.J., Qin, Z., Marika, L., Alice, C., W, R.L., *et al*. (2020) Exome Sequencing Reveals a Phenotype Mod-

ifying Variant in ZNF528 in Primary Osteoporosis with a COL1A2 Deletion. *Journal of Bone and Mineral Research*: *The Official Journal of the American Society for Bone and Mineral Research*.

[14] Bruce, B., Karen, W., Ella, O., *et al*. (2020) High Bone Mineral Density Osteogenesis Imperfecta in a Family with a Novel Pathogenic Variant in COL1A2. *Hormone Research in Paediatrics*.

[15] Youngji, K., Seul, Y.Y., Sup, P.S., Jin, K.M., Min, S.C. and Hee, C.S. (2019) Congenital Analbuminemia in a Korean Male Diagnosed with Single Nucleotide Polymorphism in the ALB Gene: The First Case Reported in Korea. *Yonsei Medical Journal*, **60**, 700-703. https://doi.org/10.3349/ymj.2019.60.7.700

[16] Lorenzo, M., Gianluca, C., Monica, C., Francesca, L., Monica, G. and Ulrich, K.-H. (2019) Diagnosis, Phenotype, and Molecular Genetics of Congenital Analbuminemia. *Frontiers in Genetics*, **10**, 336. https://doi.org/10.3389/fgene.2019.00336

[17] Wang, F., *et al*. (2020) Identifying the Hub Gene in Gastric Cancer by Bioinformatics Analysis and *in Vitro* Experiments. *Cell Cycle*, **19**, 1326-1337. https://doi.org/10.1080/15384101.2020.1749789

[18] Wu, Y. and Xu, Y. (2020) Integrated Bioinformatics Analysis of Expression and Gene Regulation Network of COL12A1 in Colorectal Cancer. *Cancer Medicine*, **9**, 4743-4755. https://doi.org/10.1002/cam4.2899

[19] Shen, H.Y., *et al*. (2020) The Prognostic Value of COL3A1/FBN1/COL5A2/SPARC-mir-29a-3p-H19 Associated ceRNA Network in Gastric Cancer through Bioinformatic Exploration. *Journal of Cancer*, **11**, 4933-4946. https://doi.org/10.7150/jca.45378

[20] Peng, W., Li, J.-D., Zeng, J.-J., Zou, X.-P., Tang, D., Tang, W., *et al*. (2020) Clinical Value and Potential Mechanisms of COL8A1 Upregulation in Breast Cancer: A Comprehensive Analysis. *Cancer Cell International*, **20**, Article No. 392. https://doi.org/10.1186/s12935-020-01465-8

[21] Nixon, T., Richards, A.J., Lomas, A., Abbs, S., Vasudevan, P., McNinch, A., *et al*. (2020) Inherited and de Novo Biallelic Pathogenic Variants in COL11A1 Result in Type 2 Stickler Syndrome with Severe Hearing Loss. *Molecular Genetics & Genomic Medicine*, **8**, e1354. https://doi.org/10.1002/mgg3.1354