# 肝细胞癌关键生物标志物的筛选和鉴定
## ——来自生物信息分析的证据

王乾坤
安徽医科大学第一临床医学院，中国·安徽 合肥 230032

【摘 要】为探究HCC发生的分子机制，从GEO芯片数据集GSE84402获取肝癌组织和肝癌旁正常组织的数据信息，通过GEO2R鉴定出差异表达基因(DEGs)，对DEGs进行功能GO富集分析和KEGG通路分析比较，并构建蛋白-蛋白相互作用网络(PPI)图导入Cytoscape进行模块分析得到Hub基因，分别对Hub基因在cBioPortal用Kaplan-Meier曲线进行总生存期和无病生存期分析，确定关键基因，最后用UCLAN进行验证。结果鉴定出93个基因，其中下调基因68个，上调基因25个。TTK、CYP3A4和MT1G是Hub基因，生存分析显示，TTK是肝癌的关键基因，可能是HCC的候选生物标志物，或成为肝细胞癌基因治疗的靶点，验证后支持上述结果。

【关键词】肝细胞癌；差异表达基因；蛋白-蛋白相互作用；富集分析；GSE84402

### 1 引言

原发性肝细胞癌(hepatocellular carcinoma,HCC)(以下简称肝癌)是我国发病率较高的的癌症之一,近几年发病率呈上升趋势，具有恶性程度高，进展快，转移早，发现晚，预后差的特点。因此，了解HCC早期诊断、治疗和生存预后相关基因，为制定有效诊断和治疗HCC的策略提供依据具有重大意义[1]。

从基因层面看,肝细胞癌的发生发展是一个多基因参与的多因素多阶段的复杂过程,涉及多个信号转导通路中大量基因表达的异常[2]。随着大数据时代的到来，人们广泛使用基因芯片技术和生物信息学的方法筛选参与HCC发生发展的关键基因，为肝癌的早期诊断和治疗带来福音。

目前虽然对HCC形成和发展的分子机制已有广泛的研究,但是肝细胞癌的关键基因尚未完全阐明，本实验选取了GSE84402数据集，对16例肝癌组织及16例相应的癌旁正常组织进行对比分析，经验证后选取具显著临床意义的基因，这可能是HCC的候选生物标志物，或成为肝细胞癌基因治疗的靶点。

### 2 数据的收集和预处理

2.1HCC组织基因表达谱数据获取及差异基因筛选

本实验从GEO数据库(http://www.ncbi.nlm.nih.gov/geo)选取了GSE84402数据集，其中包含了16例肝癌组织及16例相应的癌旁正常组织的基因表达谱，将其分别分组为cancer组和normal组使用GEO2R在线分析工具(http://www.ncbi.nlm.nih.gov/geo/geo2r),以矫正后P<0.01和作为截取标准,进行cancer组和normal组的DEGs分析。

2.2DEGs的GO和KEGG富集分析

在线数据库DAVID(http://david.ncifcrf.gov)(6.8版)是一个集生物数据和分析工具于一体的在线综合生物信息数据库,可为用户提供一套完整的基因与蛋白质功能注释信息[3]。本研究将DEGs导入DAVID数据库(http://david.ncifcrf.gov)(6.8版)，以人源基因为背景，进行GO功能富集分析KEGG通路分析，分析内容包括生物过程(biological process，BP)、分子功能(molecular function,MF)、细胞成分(cellular component,CC)及细胞信号转导通路。

2.3PPI网络建设与模块分析

本研究首先将cancer组特有的DEGs上传至STRING数据库(https://string-db.org/cgi/input.pl)主界面获得DEGs编码的蛋白质相互作用(protein-protein interaction，PPI)网络图。然后将得到的PPI以TSV文件下载后导入Cytoscape(版本3.6.1),用MCODE插件对PPI网络进行模块分析，基于重要性程度筛选关键Hub基因。选择标准: MCODE scores>5,degree cut-off=2,node score cut-off=0.2,k-score=2,Max depth=100。

2.4中枢基因的选择和分析与关键基因的验证

使用Cytoscape(版本3.6.1)的生物网络基因肿瘤学工具(BiNGO)(version 3.0.4)插件对Hub基因进行生物学过程分析并进行可视化。在cBioPortal(https://www.cbioportal.org/)中使用Kaplan-Meier曲线进行Hub基因的总生存期和无病生存期分析。最后利用UCLAN数据库(http://ualcan.path.uab.edu/)对最终确定的关键基因进行验证分析。

### 3 结果

3.1肝癌中DEGs的鉴定

应用GEO2R对基因芯片GSE84402结果的标准化,鉴定出93个基因，其中下调基因68个，上调基因25个。

3.2DEGs的KEGG和GO富集分析

对93个差异基因使用DAVID工具进行GO功能富集分析和KEGG通路富集分析。GO功能分析结果显示，DEGs的生物过程(BP)变化显著富集于有丝分裂核分裂、有丝分裂细胞周期和细胞周期过程；分子功能(MF)变化主要富集于蛋白质结合和ATP结合。KEGG路径分析显示，DEGs主要富集于药物代谢和细胞周期(如表1所示)。

表1 KEGG和GO富集分析

| 类别 | 项目 | 描述 | 数量 | P值 |
|---|---|---|---|---|
| BP | GO:0051301 | cell division | 9 | $3.7\times10^{-10}$ |
| | GO:0007067 | Mitotic nuclear division | 7 | $6.9\times10^{-8}$ |
| | GO:0007062 | sister chromatid cohesio | 6 | $3.3\times10^{-8}$ |
| MF | GO:0042803 | protein homodimerization activity | 11 | $1.1\times10^{-3}$ |
| | GO:0019899 | enzyme binding | 6 | $1.5\times10^{-2}$ |
| | GO:0020037 | heme binding | 5 | $2.9\times10^{-3}$ |
| | GO:0005506 | iron ion binding | 5 | $4.3\times10^{-3}$ |
| KEGG | hsa05204 | Chemical carcinogenesis | 5 | $1.1\times10^{-3}$ |

3.3PPI网络建设与模块分析和Hub基因的筛选

利用String在线网络分析工具构建DEGs的PPI网络图并用Cytoscape标注上下调基因，其中下调基因用红色标注，上调基因用蓝色标注(如图1所示)，使用Cytoscape的MCODE插件得到连接度最高的三个Hub基因：TTK、CYP3A4和MT1G。

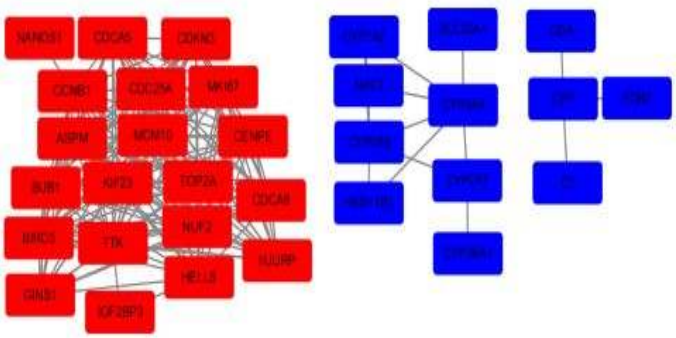

图1　HCC关键基因的蛋白互作网络图

3.4Hub 基因选择与分析

利用 cBioPortal 中的 Kaplan−Meier 曲线工具对三个 Hub 基因进行生存分析，我们发现一个关键基因(TTK)高表达时，样本的总生存率显著降低(如图2所示)，而其他基因的总存活率没有显著差异。因此，TTK 是本次实验初步确定的关键基因。

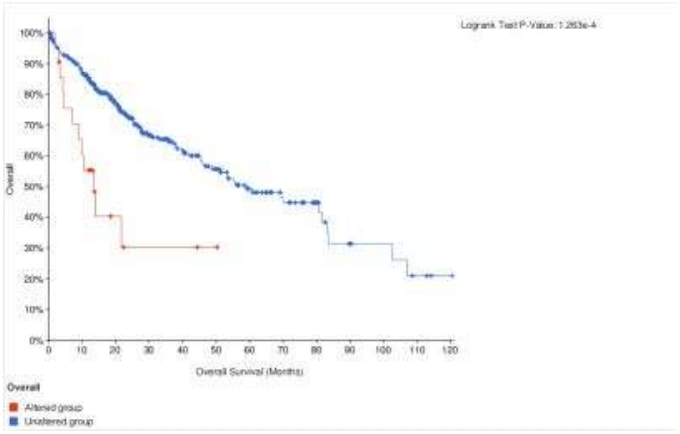

图2　TTK总生存率

3.5 关键基因的验证

在UCLAN数据库中对TTK在肝细胞癌中的表达情况进行验证，对 TTK 进行表达和生存分析，结果表明，TTK 在肝癌组织中高表达，且 P 值具有统计学意义（P<0.01）（如图 4 所示），对肝细胞癌患者具有较差的影响（如图 5 所示）。

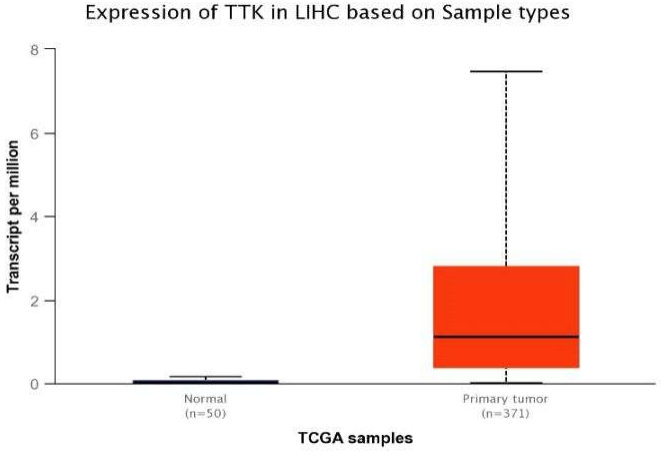


图3　TTK在肝癌组织表达情况（UCLAN数据库）

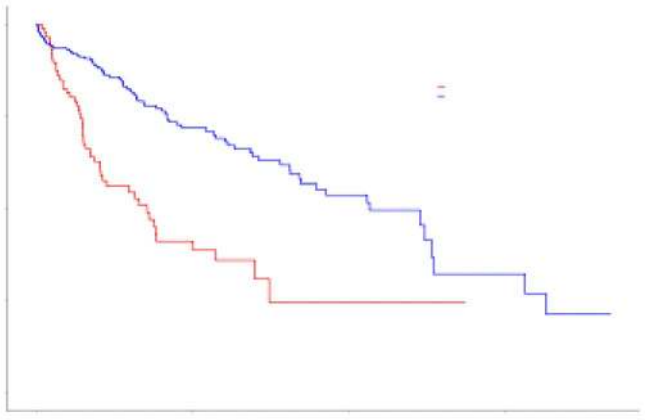

图4　TTK对肝癌病人预后生存分析（UCLAN数据库）

## 4　结论

本研究在 GSE84402 数据集的基础上，运用生物信息学方法筛选出 cancer 组和 normal 组内的 DGEs，随后将 93 个 DGEs 进行 GO 富集分析和KEGG通路分析比较，DGEs主要富集于有丝分裂核分裂、有丝分裂细胞周期和细胞周期过程，并构建PPI网络，选取连接度最高的3个Hub基因，生存分析显示只有TTK与总生存率相关，对TTK用UCLAN进行表达和生存验证，结果支持上述结论。总之，本研究通过生物信息学的方法发现，TTK 是 HCC 的关键基因，与肝癌的早期诊断、治疗与预后判断有重大关系，具有重要临床意义。但是TTK对于HCC的临床应用尚未有循证医学的证据，也没有实验可以直接证明，因此，下一步应该以预测结果为目的进行临床数据的收集和循证医学的探索。

**参考文献：**

[1]陆进,杨月,赵学影,胡超力.肝细胞癌关键基因的筛选及其临床意义[J].山西医科大学学报,2019,50(07):879−888.

[2]Kirstein M.M.,Vogel A..The Pathogenesis of Hepatocellular Carcinoma[J].Digestive Diseases,2014,32(5).

[3]连旭,孙唯秀,韩崇旭.基于生物信息学筛选肝细胞癌年轻患者特有关键枢纽基因及其临床意义[J].中国肿瘤生物治疗杂志,2020,27(02):161−169.

**作者简介：**

王乾坤（2000.5.7—），男，汉族，籍贯安徽霍邱，安徽医科大学第一临床医学院医学检验技术专业18级在校生。