

基于生物信息学分析并构建肝细胞癌血浆中潜在 miRNA-mRNA 调控网络

温明圆 白皓天 杨婧 王锐*
(黑龙江中医药大学 哈尔滨 150040)

摘要 目的 肝细胞癌(HCC)现已成为世界范围内发病和死亡的主要疾病。越来越多的实验数据表明,表达异常的 microRNA (miRNA) 参与了 HCC 的发病机制。然而,仍然缺乏对 HCC 血浆中 miRNA-mRNA 调控网络的综合研究。方法 在 GEO 数据库中下载微阵列数据集 GSE71540 和 GSE71545,使用 GEO2R 工具进行分析,识别 HCC 和正常血浆之间差异表达的 miRNAs (DEMs)。筛选出两个数据集中变化一致的 miRNA 作为候选 DEMs。通过使用 FunRich 和 miRNet 工具分别预测候选 DEMs 的上调转录因子与下调转录因子。接下来,利用 DAVID 对目的基因进行 GO 功能分析与 KEGG 通路富集分析。并且,利用 STRING 数据库和 Cytoscape 软件对 PPI 和 DEM-hub 基因网络进行了构建。最后,用 GSE102759 评估核心基因的表达。结果 在上述 2 个数据集中,共筛选出 3 个 DEMs。其中 1 个上调,2 个下调。RFX1 和 ARNT 可以调节大部分下调的 DEMs,而 SP4 和 ZIC1 可以调节大多数上调的 DEMs。预测共有 446 个靶基因,其中 161 个上调基因,285 个下调基因。DEMs 的靶基因主要富集于 Wnt 信号通路、细胞周期过程以及癌细胞中的粘蛋白中。通过 DEM-hub 基因网络构建,发现大部核心基因都可能受到 miR-5100、miR-1275、miR-530b-5p 和 miR-530-3p 的调控。在排名前 6 位的中心基因中,AGO2 和 IGF1R 的表达与 GSE102759 数据集一致。结论 本研究首次在 HCC 血浆中构建了潜在的 miRNA-mRNA 调控网络,可以为 HCC 的发病和治疗提供新的思路。

关键词 microRNA; 肝细胞癌; 生物信息学; miRNA-mRNA 调控网络

肝细胞癌 (Hepatocellular carcinoma,HCC) 是原发性肝癌最主要的类型,具有侵袭和转移率高、复发风险高、中晚期并发症严重等特点,是发病率和致死率较高的一种肿瘤疾病,多由慢性肝硬化、病毒感染 (HBV、HCV) 等慢性肝脏疾病导致。多数 HCC 患者确诊时已处于晚期,且多伴有转移,失去了最佳治疗时机,预后差,5 年存活率 <5%。HCC 的发生及发展参与的基因众多,由 DNA 突变引起的原癌基因激活或抑癌基因异常失活是其发生、发展的主要机制。作为一类内源性非编码 RNA,微小 RNA 可通过影响功能基因表达调控肿瘤细胞增殖、分化及凋亡等过程。近年来,随着人们对肿瘤的不断认识,发现 miRNA 在肿瘤的发生及发展过程中具有重要作用。

尽管对 miRNA 在 HCC 中的表达及功能的研究较多,但迄今对 miRNA-mRNA 调控网络在 HCC 血浆中的作用的研究较少。在本研究中,首次通过分析 GEO 数据库中的两个数据集(GSE71540、GSE71545),将 HCC 血浆中的 DEMs 与正常血浆进行对比筛选。然后分别进行转录因子-DEM、DEM-靶基因、靶基因功能分析和 DEM-hub 基因网络分析。此外,在 GSE102759 数据集中进一步验证了血浆中核心基因的水平。在此,旨在构建潜在的在 HCC 血浆中发挥作用 miRNA-mRNA 调控网络,为 HCC 的诊断和治疗提供有前景的靶点。

1 材料与方法

1.1 miRNA 微阵列排列

根据基因表达微阵列数据库(<https://www.ncbi.nlm.nih.gov/geo/>)中的选项进行如下选择:“HCC”(Keywords),“noncoding RNA profiles by array”(Study type),“Homo sapiens”(ism),“Plasma”(Attribute name)筛选与 HCC 血浆相关的 miRNA 数据集。最后选取 2 个 miRNA 数据集(GSE71540 和 GSE71545)进行后续分析。两个数据集均来源于 GPL18402 平台(febeit Homo Sapiens miRBase 13.0)。

1.2 DEMs 的筛选

GEO2R (<http://www.ncbi.nlm.nih.gov/geo/geo2r/>)是用于分析 GEO

数据的网络应用程序,分析 GEO 数据,用于 HCC 和正常血浆之间的 DEMs 比较和筛查。采用 Benjamini 和 Hochberg “假阳性”(FDR)方法对 P 值进行调整,进而纠正假阳性结果的发生。在此处,设置阈值,控制 P 值 < 0.05、|Log fold change (FC)| > 1.5,以此识别 DEMs。接着,利用维恩图分析 GSE71540 和 GSE71545 两个数据集中 DEMs 的重叠情况,使用仙桃学术 (<https://www.xiantaolove.com/products>) 生成 DEMs 的热图和候选 DEMs,仙桃学术是一个用于进行生信数据分析的功能型网站。

1.3 DEMs 上调与下调基因的预测

FunRich (<http://www.funrich.org/>)是一个独立的软件工具,主要用于基因和蛋白质的功能富集分析和相互作用网络分析,用于预测候选 DEMs 可能存在的上调基因。当 P 值 < 0.05 具有统计学意义。

miRNet (<https://www.mirnet.ca/>)是一个集成的工具包,包含关于 miRNA-靶基因相互作用的信息,并通过可视化分析显示其中的关联,用来预测候选 DEMs 潜在的下调基因。

1.4 目的基因的 GO 功能分析与 KEGG 通路分析

为研究 DEMs 目的基因的生物学功能,利用能够进行在线注释、可视化和集成的数据库工具 DAVID 进行基因本体分析(Gene Ontology, GO)和京都基因与基因组百科全书(Kyoto Encyclopedia of genes and Genomes, KEGG)路径分析。调整当 P 值 < 0.05 时具有统计学意义。

2 结果

2.1 HCC 血浆中 DEMs 的筛选与确定

筛选后调整阈值 Pvalue < 0.05 和 logFC | > 1.5, GSE71545 数据集中确定了 6 个 DEMs(3 个上调和 3 个下调),GSE71540 数据集中确定了 17 个数据集(8 个上调和 9 个下调)。以上两个数据集的 DEMs 的热图,在这些数据集中,一个上调的 DEM (miR-1275)和两个下调的 DEMs(miR-513b, miR-5100)在两个数据集中变化一致。

(下转第 48 页)

(上接第 39 页)

2.2 DEMs 上调基因与下调基因的预测

本研究使用 FunRich 软件预测候选 DEMs 的上调转录基因。上调和下调 DEMs 的转录因子如图 4 所示。上调 DEMs 的转录因子为 SP4、ZIC1、ITGAL、SPI、INSM1 和 CACD。下调的 DEMs 转录因子有 RFX1、ARNT、POU5F1、ELF3、IRF1 和 YY1。通过 miRNet 数据库共预测出候选 DEM 的靶向基因 446 个,其中上调 DEMs 的靶向基因 161 个,下调 DEM 的靶向基因 285 个(表 2)。为了更好地显示下调基因的关系,如图 5 所示,通过可视化分析展现了两个下调 DEMs 靶基因之间的关系,由于所筛选出上调 DEM 只有一个,这里对上调靶基因之间关系的可视化分析不做赘述。

2.3 目的基因 GO 功能和 KEGG 通路分析

接着,对 DEMs 的 446 个靶基因进行 GO 和 KEGG 分析。如图 6 所示,生物过程(BP)分析显示,DEMs 靶基因在细胞基质粘附调节、细胞热反应、细胞粘附的正向调节和细胞对生长因子刺激反应的调节等方面富集较多;细胞组分(CC)分析显示,DEMs 靶基因主要富集在细胞质处理小体、细胞质核糖核蛋白体、核糖核蛋白体和核斑点等;分子功能(MF)分析结果显示,DEMs 靶基因的组蛋白甲基转移酶活性、组蛋白赖氨酸 N-甲基转移酶、赖氨酸 N-甲基转移酶和蛋白赖氨酸 N-甲基转移酶显著富集。此外,如图 7 中 KEGG 通路分析显示,DEMs 靶基因在 Wnt 信号通路、人乳头瘤病毒感染、癌变组织中的蛋白聚糖、细胞周期、肝细胞癌中显著富集。

3 讨论

最近的研究表明 miRNA 的表达可以被转录因子所调控^[24-25],因此预测可以潜在调控候选 DEMs 的转录因子。其中,YY1 被预测为是能够调控下调 DEMs 表达的基因,在对 YY1 的研究中,Natanek 等^[26]发现 YY1 的表达与小四头肌纤维横截面积(CSA)有关,其核心定位于

HCC 患者的肌肉中,而非对照组。提示 YY1 在 HCC 相关肌肉萎缩中起重要作用。

尽管首次通过整合多个微阵列数据集来研究 HCC 血浆中潜在的 miRNA-mRNA 调控网络,但目前的研究存在一些局限性。首先,只关注了 HCC 与正常血浆之间的 miRNAs 和 mRNA,其中,一些 miRNAs 和 mRNA 在 HCC 的不同阶段可能存在差异,需要进行进一步的鉴定。其次,与生物标志物分析通常需要的样本量相比,所使用的每个数据集中的样本量还远远不够。最后,仅基于公共数据库的预测 miRNA-mRNA 的相互作用是远远不够的,还需要进行进一步的体内、体外实验来对分析进行验证。

综上所述,基于 GEO 数据库与生物信息学分析,构建了 HCC 血浆中四个潜在的 miRNA-mRNA 通路(miR-1275-ACTB、miR-1275-SUMO1、miR-1275-IGF1R 和 miR-513b-5p-HNRNPA1)作为 HCC 患者诊断和治疗的潜在生物标志物。希望这些发现能够为今后的深入研究做出贡献,达到改善 HCC 患者预后的目的。

参考文献:

[1]Pasquinelli AE. MicroRNAs and their targets: Recognition, regulation and an emerging reciprocal relationship. *Nat Rev Genet.* 13:271 - 282. 2012.

[2]Chen X, Liang H, Zhang J, et al. Secreted microRNAs: A new form of intercellular communication. *Trends Cell Biol.* 22:125 - 132. 2012.

基金项目:中央支持地方高校改革发展基金(2020YQ05)