

# 基于非参数正态图模型的差异网络分析 ——肝细胞癌生存时间枢纽基因的筛选

李晶, 张奇, 任雨冬, 刘艳

哈尔滨医科大学卫生统计学教研室, 黑龙江 哈尔滨 150081

**摘要:** **目的** 评价基于非参数正态图模型的差异网络分析方法筛选差异基因的效果, 并筛选影响肝细胞癌患者生存时间的枢纽基因, 用于预测患者的预后或作为新的药物靶点。 **方法** 模拟实验中用 AUC 值评价基于非参数正态图模型的差异网络分析筛选差异基因的效果, 再从 TCGA 数据库下载 3 个平台 5 个数据集的肝细胞癌患者的基因表达数据, 纳入通路等先验信息利用非参数正态图模型构建差异表达网络, 按照节点连接边的个数即度的大小选择枢纽基因。 **结果** 在 4 种条件的模拟实验中, 非参数正态图模型筛选差异基因的 AUC 值范围为 0.71~0.97; 在肝细胞癌数据的分析中, 筛选出 8 个与肝细胞癌生存时间相关的枢纽基因, 分别是 *IGF1*、*ErbB2*、*FGF9*、*GH2*、*cSH2*、*HSP90AA1*、*PPP2R5B* 和 *EPO*。 **结论**

基于非参数正态图模型的差异网络分析方法在模拟实验中筛选差异基因的效果较好, 在实例数据分析中筛选的肝细胞癌生存时间相关的枢纽基因都有较为明确的生物学功能。

**关键词:** 差异网络; 肝细胞癌; 生存时间; 枢纽基因; 非参数正态图模型

**中图分类号:** R735.7 **文献标识码:** A **文章编号:** 1006-3110(2019)04-0404-05 **DOI:** 10.3969/j.issn.1006-3110.2019.04.006

## Differential network analysis based on non-paranormal graphical models—screening of hub genes related to the survival time of hepatocellular carcinoma

LI Jing, ZHANG Qi, REN Yu-dong, LIU Yan

Department of Health Statistics, Harbin Medical University, Harbin, Heilongjiang 150081, China

Corresponding author: LIU Yan, E-mail: liuyan@ems.hrbmu.edu.cn

**Abstract:** **Objective** To evaluate the screening effect of differential network analysis based on non-paranormal graphical models, and to screen the hub genes affecting the survival time of patients with hepatocellular carcinoma using the differential network analysis model so as to predict the patients' prognoses or serve as a new drug target. **Methods** The screening effect of differential network analysis using non-paranormal graphical models was evaluated by the area under the ROC curve (AUC) values in the simulation experiments. The gene expression data of hepatocellular carcinoma from 3 platforms and 5 datasets were downloaded from The Cancer Genome Atlas (TCGA) database, and then incorporated prior information such as pathways into differential network using non-paranormal graphical models to construct the differential network. The hub genes were chosen by the degree, the number of edges connected to the node. **Results** In the simulation experiments performed under four conditions, the AUC values of the non-paranormal graphical models ranged from 0.71 to 0.97. In the application of hepatocellular carcinoma data, 8 hub genes related to the survival time of hepatocellular carcinoma were screened respectively, which were *IGF1*, *ErbB2*, *FGF9*, *GH2*, *cSH2*, *HSP90AA1*, *PPP2R5B* and *EPO*. **Conclusions** The simulation experiments of differential network analysis based on non-paranormal graphical models show good results. The biological function of hub genes affecting the survival time of patients with hepatocellular carcinoma is clearly defined.

**Key words:** differential network; hepatocellular carcinoma; survival time; hub gene; non-paranormal graphical model

无向图模型或马尔可夫网络是一种流行的统计模型, 广泛用于各种领域, 包括统计物理学、自然语言

**基金项目:** 国家自然科学基金(81172741)

**作者简介:** 李晶(1994-), 女, 山西省运城市人, 硕士在读, 主要从事医学领域统计学方法的应用与研究工作。

**通信作者:** 刘艳, E-mail: liuyan@ems.hrbmu.edu.cn。

处理、图像分析和医学<sup>[1]</sup>。高斯图模型是一类流行的无向概率图模型, 能很好地刻画服从正态分布的数据, 而高维数据通常不能严格服从正态分布, 并且往往含有噪声<sup>[2]</sup>, Zhang 等<sup>[3]</sup>采用了基于非参排序的统计量(肯德尔秩相关系数)估计相关系数矩阵, 提出了稳定的估计方法用于学习非参数概率图模型, 放宽了高斯图模型对数据正态分布的限制。肝细胞癌(hepatocel-

lular carcinoma cell, HCC) 是全世界最常见和最具侵袭性的癌症之一, HCC 是世界上第五大常见恶性肿瘤, 也是亚洲第二大癌症死亡原因<sup>[4]</sup>。恶性肿瘤不仅严重威胁人类的身体健康, 而且给国家、社会、家庭造成沉重的经济负担<sup>[5]</sup>, 肝切除仍是肝细胞癌的一线治疗手段, 肝功能储备良好的患者 5 年生存率约为 40%~50%<sup>[6]</sup>。本研究通过模拟实验评价基于非参数正态图模型的差异网络分析方法筛选差异基因的效果, 并通过对 TCGA 数据库中肝细胞癌患者多个数据平台的基因表达数据进行挖掘, 运用非参数正态图模型构建差异网络, 获取影响肝细胞癌患者生存时间的枢纽基因, 并进一步进行生物学功能验证, 明确枢纽基因在肝细胞癌预后中的作用。

1 材料与方法

1.1 pDNA 的方法原理 纳入先验信息基于非参数正态图模型的差异网络分析模型(the prior information-dependent differential network analysis model, pDNA) 是通过计算肯德尔相关系数矩阵得到协方差矩阵和精度矩阵, 结合 D-trace 损失函数和分组桥惩罚函数, 纳入通路等先验信息用于构建差异网络, 纳入的先验信息主要有以下三点: 第一点, 不属于同一通路的基因在差异网络中不存在差异边; 第二点, 网络的变化可能是由于特定的调控基因在状态间扰动形成的; 第三点, 由不同平台数据构建的差异网络可能享有共同结构<sup>[3]</sup>。

若随机向量  $X=(X_1, X_2, \cdots, X_p)^T$  服从非参数正态分布  $NPN(0, \Sigma, f)$ , 则稀疏精度矩阵  $\Sigma^{-1}$  表达了随机

变量  $(X_1, X_2, \cdots, X_p)$  的相关性, 当且仅当  $(\Sigma^{-1})_{ij}=0$  说明在给定其他变量时  $X_i$  与  $X_j$  条件独立。  $n_c(c=1, 2)$  表示第  $c$  组的样本例数,  $X^{kc}$  表示第  $k$  个平台第  $c$  组的  $n_c \times p$  基因表达数据矩阵, 矩阵中的第  $(l, i)$  个元素  $X^{kc}_{li}$  表示第  $l$  个样本第  $i$  个基因的表达量。为解决数据不严格服从正态分布的问题, 相关矩阵计算的是基因  $i$  和基因  $j$  的肯德尔相关系数进而得到两组样本的协方差矩阵  $\{\Sigma^{kc}\}$  和精度矩阵  $\{(\Sigma^{kc})^{-1}\}$ , 其中  $\Sigma^{kc}_{ij} = \begin{cases} \sin(\frac{\pi}{2}\hat{\tau}^{kc}_{ij}), i \neq j \\ 1, i=j \end{cases}$ 。纳入先验信息构建的差异网络模型

为  $\min_{\{\Delta^k\}, \{V^k\}} L(\{\Delta^k\}) + \lambda \sum_{j=1}^p \sqrt{\sum_{i=1}^p \sqrt{\sum_{k=1}^K |V^k_{ij}|}}$ , 该函数服从  $\Delta^k = V^k + (V^k)^T$ , 其中  $\lambda$  是非负的可调整的参数用于控制差异网络的稀疏度,  $\hat{\Delta}^k$  表示第  $k$  个平台两精度矩阵的差异, 为保证  $\hat{\Delta}^k$  是对称的需控制  $\Delta^k = (\Delta^k)^T$ ; 当  $F_{ij}=0$  时  $V^k_{ij}=0$ , 其中  $F$  是共通路指示矩阵, 若基因  $i$  和基因  $j$  至少同时出现在一条通路中则  $F_{ij}=1$ , 否则  $F_{ij}=0$ 。

1.2 参数选择的方法 参数的选择采用近几年提出的一种稳定性选择的方法 StARS<sup>[7]</sup> (stability approach to regularization selection), 该方法依据数据特征选取参数使得估计的差异网络稀疏且稳定, 在网络估计方面比包括  $K-CV$ 、 $AIC$ 、 $BIC$  等模型选择方法效果都要好<sup>[4]</sup>。

1.3 模拟数据的生成 考虑以下 6 个参数生成模拟数据, 进行 4 种条件下的模拟实验。见表 1。

表 1 模拟实验的参数设置

| 参数         | 意义              | 实验 1               | 实验 2               | 实验 3               | 实验 4               |
|------------|-----------------|--------------------|--------------------|--------------------|--------------------|
| $K$        | 平台数据库个数         | 5                  | 5                  | 5                  | 5                  |
| $p$        | 基因数             | 100                | 100                | 200                | 200                |
| $m_{pert}$ | 扰动基因个数          | 4                  | 4                  | 8                  | 8                  |
| $\rho_2$   | 特定数据集中差异边的比例    | 0.1                | 0.2                | 0.1                | 0.2                |
| $\rho_1$   | 每个扰动基因驱动的差异边的比例 | 0.3, 0.5, 0.7, 0.9 | 0.3, 0.5, 0.7, 0.9 | 0.3, 0.5, 0.7, 0.9 | 0.3, 0.5, 0.7, 0.9 |
| $n$        | 样本量             | 50, 100, 200       | 50, 100, 200       | 100, 200, 400      | 100, 200, 400      |

1.4 肝细胞癌数据的来源 为了解不同生存时间肝细胞癌患者的基因调控网络的差异, 选择 TCGA 数据库的肝细胞癌队列研究(version: 2017. 10. 13) 5 个基因表达数据集(3 个来自转录组, 1 个来自基因组, 1 个来自表观遗传组)。转录组的 3 个 RNA 测序数据集来自 Illumina HiSeq 2000 RNA Sequencing 平台, 基因组的拷贝数变异数据来自 Affymetrix Genome-Wide Human SNP Array 6.0 平台, 表观遗传组的甲基化数据来自 Illumina Infinium Human Methylation 450 平台, 如果多个

位点对应于一个基因, CpG 位点通过获取第一个主成分映射到基因。

1.5 统计学分析 使用 R 3.5.0 软件和 MATLAB2014a 软件完成统计学分析。在 R 软件中调用 MATLAB 程序进行模拟实验, 每种条件下的模拟实验各进行 10 次, 并用 ROC 曲线下的面积 AUC 值来评价模型对差异基因的筛选效果; 在实例数据分析中, 采用百分位数法确定阈值从而将样本分成两组; 缺失值填补采用中心趋势值填补法, 各数据集中存在超过 30% 缺失值的基因

直接删除;肝细胞癌生存时间差异基因的初步筛选采用  $t$  检验,通过查阅文献及 KEGG 富集分析结果选择富集在 7 条通路中的基因作为最终的候选基因;并将各数据集筛选过后的基因表达数据应用到 pDNA 模型当中构建差异网络,筛选影响肝细胞癌生存时间的枢纽基因。 $P<0.05$  为差异有统计学意义。

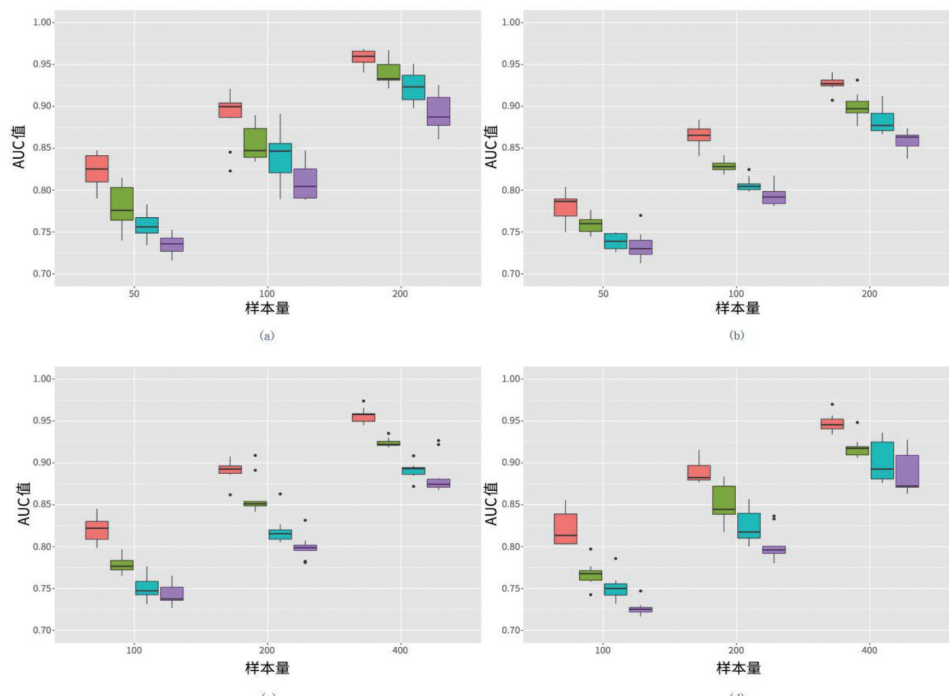
2 结果

2.1 模拟实验结果 4 种条件下的模拟实验,均显示出  $n$  最大且  $\rho_1$  最小时,基于非参数正态图模型的差异网络分析模型对差异基因的筛选效果最好。由表 2 和图 1 可以看出,非参数正态图模型筛选差异基因的 AUC 值总范围为 0.71~0.97;在基因数/扰动基因数分别是 100/4 和 200/8 的情况下;在样本量  $n$  相同时,每个扰动基因驱动的差异边的比例  $\rho_1$  越小筛选效果越好;在每个扰动基因驱动的差异边的比例  $\rho_1$  相同时,随样本量  $n$  的增加 AUC 值也逐渐增大,表明样本量越大

筛选效果越好。

表 2 4 种模拟实验条件下模型筛选差异基因的 AUC 值范围

| 每个扰动基因驱动的差异边的比例 | 样本量 | 实验 1      | 实验 2      | 实验 3      | 实验 4      |
|-----------------|-----|-----------|-----------|-----------|-----------|
| 0.3             | 50  | 0.79~0.85 | 0.75~0.80 | —         | —         |
|                 | 100 | 0.82~0.92 | 0.84~0.88 | 0.80~0.85 | 0.80~0.86 |
|                 | 200 | 0.94~0.97 | 0.91~0.94 | 0.86~0.91 | 0.88~0.92 |
|                 | 400 | —         | —         | 0.94~0.97 | 0.93~0.97 |
| 0.5             | 50  | 0.74~0.81 | 0.74~0.78 | —         | —         |
|                 | 100 | 0.83~0.89 | 0.82~0.84 | 0.76~0.80 | 0.74~0.80 |
|                 | 200 | 0.92~0.97 | 0.88~0.93 | 0.84~0.91 | 0.82~0.88 |
|                 | 400 | —         | —         | 0.92~0.94 | 0.91~0.95 |
| 0.7             | 50  | 0.73~0.78 | 0.73~0.75 | —         | —         |
|                 | 100 | 0.79~0.89 | 0.80~0.82 | 0.73~0.78 | 0.73~0.79 |
|                 | 200 | 0.90~0.95 | 0.87~0.91 | 0.81~0.86 | 0.80~0.86 |
|                 | 400 | —         | —         | 0.87~0.91 | 0.88~0.94 |
| 0.9             | 50  | 0.72~0.75 | 0.71~0.77 | —         | —         |
|                 | 100 | 0.79~0.85 | 0.78~0.82 | 0.73~0.77 | 0.72~0.75 |
|                 | 200 | 0.86~0.93 | 0.84~0.87 | 0.78~0.83 | 0.78~0.84 |
|                 | 400 | —         | —         | 0.87~0.93 | 0.86~0.93 |



注:a 图表示实验 1,b 图表示实验 2,c 图表示实验 3,d 图表示实验 4。  
图 1 4 种模拟实验条件下模型筛选差异基因的 AUC 值结果

2.2 肝细胞癌实例数据分析结果

2.2.1 肝细胞癌实例数据中样本的分组结果 根据生存时间的百分位数确定阈值  $K_1$  和  $K_2$ ,将样本分为两个不同的生存组:生存较短组(short-term survivors, STSs)和生存较长组(long-term survivors, LTSs)。生存时间小于  $K_1$  的患者被分到 STSs 组,生存时间大于  $K_2$  的被分到 LTSs 组,生存时间介于阈值中间( $K_1 \leq$

$x_i \leq K_2$ ) 的患者被剔除,根据不同的百分位数分组的情况见表 3。

表 3 不同百分位数的样本分组结果

| $K_1$              | $K_2$                | STSs 组例数(%) | LTSs 组例数(%) | 剔除例数(%)    |
|--------------------|----------------------|-------------|-------------|------------|
| 349.00( $P_{25}$ ) | 1 204.00( $P_{75}$ ) | 107(24.77)  | 108(25.00)  | 223(51.62) |
| 425.00( $P_{35}$ ) | 838.65( $P_{65}$ )   | 150(34.72)  | 151(34.95)  | 137(31.71) |
| 537.80( $P_{45}$ ) | 673.35( $P_{55}$ )   | 194(44.91)  | 194(44.91)  | 50(11.57)  |

根据表 3,选择选择  $K_1=537.80$  和  $K_2=673.35$ , 分别是第 45 位百分位数和第 55 位百分位数, 结果为 194 位患者 (占 44.91%) 被判定为生存较短组, 194 名患者 (占 44.91%) 被判定为生存较长组, 50 名患者 (占 11.57%) 被剔除。最终得到共同存在于 5 个数据集集中的 323 个样本, 其中 171 名患者属于生存较短组, 152 名患者属于生存较长组。

2.2.2 肝细胞癌实例数据中基因的初步筛选结果  
对初步筛选得到的差异基因做 KEGG 富集分析, 通过查阅文献<sup>[8-11]</sup> 以及表 4 所示的富集分析结果, 选择富集在以下 7 条通路中的基因作为本研究的候选基因集: 细胞周期调控、卵母细胞减数分裂、DNA 复制、碱基切除修复、PI3K-Akt 信号通路、MAPK 信号通路、ErbB 信号通路。在各数据集中删除存在超过 30% 缺失值的基因, 并对缺失值进行填补后, 有 177 个基因共同存在于 5 个数据集中并作为本研究的最终候选基因。

| 表 4 初步筛选差异基因的 KEGG 富集分析结果 |          |                      |                       |
|---------------------------|----------|----------------------|-----------------------|
| 通路                        | 通路编号     | 每个数据集中<br>该通路富集的基因个数 | 基因在该通路中<br>显著富集的数据集个数 |
| 细胞周期调控                    | hsa04110 | 50/50/50/50/50       | 5                     |
| 卵母细胞减数分裂                  | hsa04114 | 38/38/39/38/38       | 5                     |
| DNA 复制                    | hsa03030 | 14/14/18/14/14       | 5                     |
| 碱基切除修复                    | hsa03410 | 12/12/14/12/12       | 4                     |
| PI3K-Akt 信号通路             | hsa04151 | 69/69/67/69/69       | 1                     |
| MAPK 信号通路                 | hsa04010 | 49/49/48/49/49       | 1                     |
| ErbB 信号通路                 | hsa04012 | 14/14/13/14/14       | 1                     |

注: 基因数和显著富集的数据集个数分别表示 5 个数据集 (3 个 RNA 测序数据集、1 个拷贝数变异数据集和 1 个甲基化数据集) 的 KEGG 富集分析结果, 在某特定数据集中富集分析结果显示  $P<0.05$  说明初步筛选的差异基因显著富集在该通路中。

2.2.3 肝细胞癌实例数据中差异网络的构建及枢纽基因筛选结果  
根据稳定性选择方法 StARS, 可选参数  $\lambda$  从 0.05~1 之间等间距的 20 个值当中选择, 将  $\lambda$  设置为 0.35 构建差异网络。肝细胞癌患者生存较短组 (STSs) 和生存较长组 (LTSs) 的差异网络图见图 2。初步筛选得到 177 个差异基因, 共发现 62 条差异边至少出现在一个数据集中。从表 5 可以看出, 将差异基因的度从小到大排序, 选出前 8 个基因, 分别为 *IGF1*、*ErbB2*、*FGF9*、*GH2*、*cSH2*、*HSP90AA1*、*PPP2R5B*、*EPO* 作为枢纽基因。

为了进一步研究得到的差异网络, 将 5 个差异网络融合到一个权重网络  $W$  当中, 其中  $W_{ij}$  就是基因  $i$  和基因  $j$  之间的边被鉴定为差异边的数据库个数, 然后

计算每个基因的度。如图 3 所示, 共有 60 条的差异边 (占 96.77%) 是被超过 1 个数据集共同鉴定为差异边的, 表明基于非参数正态图模型的差异网络分析模型能够鉴别不同数据集共同的网络结构。由图 4 可以看出, 大部分基因的度都较低, 少数基因的度较高, 这是因为 pDNA 模型主要获取差异网络中驱动网络变化的枢纽基因。

表 5 枢纽基因对应度的情况

| 基因              | 度  |
|-----------------|----|
| <i>IGF1</i>     | 30 |
| <i>ErbB2</i>    | 27 |
| <i>FGF9</i>     | 27 |
| <i>GH2</i>      | 18 |
| <i>cSH2</i>     | 16 |
| <i>HSP90AA1</i> | 13 |
| <i>PPP2R5B</i>  | 13 |
| <i>EPO</i>      | 9  |

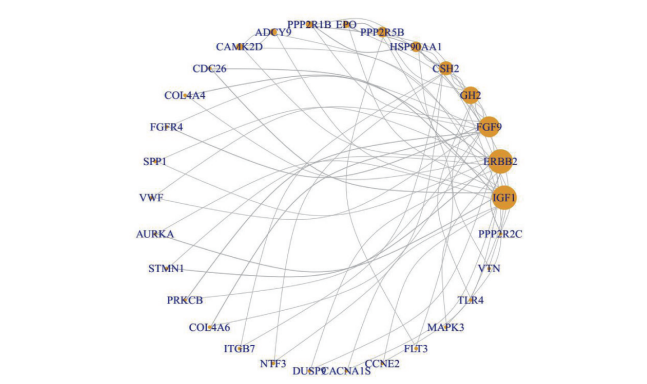


图 2 肝细胞癌患者 STSs 组和 LTSs 组基因表达的差异网络

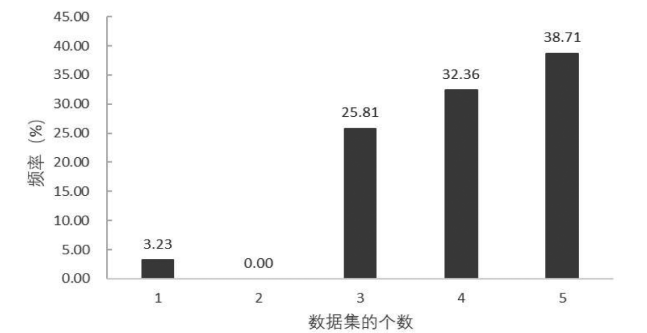


图 3 判定差异边的数据库个数频率分布

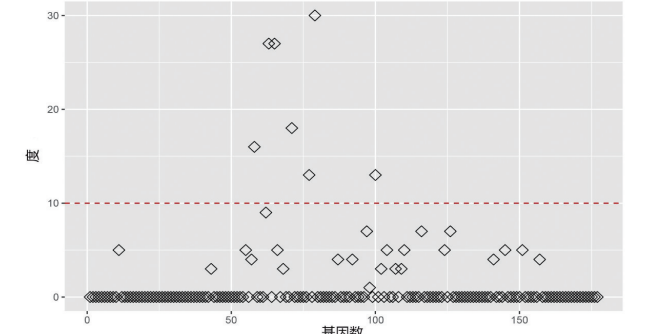


图 4 差异网络中节点度的分布



### 3 讨论

目前用来构建差异网络的方法一般都基于高斯图模型,要求数据服从正态分布,然而面对的组学数据通常是非正态的。在数据不服从正态分布时,精度矩阵不能够被翻译为可解释的图模型。而且,原先基于图模型的方法没有考虑来自生物医学研究的先验信息<sup>[3]</sup>。基于非参数正态图模型的差异网络分析方法在模拟实验中筛选差异基因的效果较好,能够综合考虑由不同的组学技术和实验平台测量得到的基因表达数据,包括 RNA 测序数据、拷贝数和甲基化数据,充分利用通路等先验信息,分析肝细胞癌患者的 5 个基因表达数据集,发现生存时间相关的枢纽基因,探讨筛选的枢纽基因和肝细胞癌预后的关系。

筛选得到的差异基因有较为明确的生物学功能。胰岛素生长因子 1(*IGF1*)是  $\beta$  细胞分裂素和促成长因子,具有调节免疫功能<sup>[12]</sup>,编码的蛋白质在功能和结构上与胰岛素相似,是参与调节生长和发育的蛋白质家族中的一员。表皮生长因子受体(ErbB2)属于酪氨酸激酶受体,过量表达 ErbB2 与人体肿瘤的高侵袭性和预后不良有密切关系,通过 FAK-Src-MAPK 信号通路诱导细胞转化和移动侵袭<sup>[13]</sup>。成纤维细胞生长因子(*FGF9*)编码的蛋白质是成纤维细胞生长因子家族(FGFs)的成员之一,参与了胚胎发育、细胞生长、形态发生、组织修复、肿瘤生长和侵袭等多种生理和病理的过程<sup>[14]</sup>,有效促进有丝分裂和细胞生长,这种蛋白质作为分泌因子被分离出来,对培养的神经胶质细胞有促进生长的作用。生长激素 2(*GH2*)、*CSH2* 编码的蛋白质是生长激素/催乳激素家族的成员,在生长控制中起着重要作用。HSP90AA1 与病毒蛋白 VP2 的结合导致自噬和 AKT-MTOR 途径失活,在早期感染中通过 HSP90AA1 - AKT - MTOR 途径诱导自噬<sup>[15]</sup>。*PPP2R5B* 产物属于 PP2A 调节亚基 B 家族,PP2A 是潜在的肿瘤抑制因子,其主要功能是磷酸化的蛋白发生去磷酸化,进而影响其目标蛋白活性及相应的信号通路活性,调节细胞增殖,分裂,分化,细胞凋亡,代谢,粘附和迁移等几乎所有的生理和病理过程,并被认为是可调节超过 30 种不同激酶的活性<sup>[16]</sup>。*EPO* 是 *EPO/TPO* 家族的成员,编码由四个螺旋束组成分泌的糖基细胞因子,该蛋白存在于血浆中,通过促进红细胞分化和启动血红蛋白合成来调节红细胞的生成,还具有神经保护作用,可抵抗多种潜在脑损伤和多种组织类型的抗凋亡功能。

基于非参数正态图模型的差异网络分析模型适用

于经过单变量单调函数变换后服从正态分布的连续型数据。然而在基因研究中存在许多离散型数据,例如下一代测序的基因表达数据,计算转录物映射到特定基因组位置的次数遵循泊松分布,并且可以使用多项分类或分类分布对突变类型建模,该模型对于离散型数据的差异网络筛选仍然具有局限性。

### 参考文献

- [1] Yang E, Ravikumar P, Allen GI, et al. Graphical models via univariate exponential family distributions[J]. J Mach Learn Res, 2015, 16(1):3813-3847.
- [2] 黄飞虎, 陈松灿. 联合学习动态半参数概率图模型[J]. 计算机科学与探索, 2018, 12(6):940-949.
- [3] Zhang XF, Ou-Yang L, Yan H. Incorporating prior information into differential network analysis using nonparanormal graphical models[J]. Bioinformatics, 2017, 33(16):2436-2445.
- [4] El Khodiry A, Afify M, El Tayebi HM. Behind the curtain of non-coding RNAs; long non-coding RNAs regulating hepatocarcinogenesis[J]. World J Gastroenterol, 2018, 24(5):549-572.
- [5] 孙高峰, 钦勤, 谢惠芳, 等. 乌鲁木齐市 2014 年恶性肿瘤发病情况分析[J]. 实用预防医学, 2017, 24(10):1240-1244.
- [6] Tang H, Lu W, Yang Z, et al. Risk factors and long-term outcome for postoperative intra-abdominal infection after hepatectomy for hepatocellular carcinoma[J]. Medicine, 2017, 96(17):e6795.
- [7] Liu H, Roeder K, Wasserman L. Stability approach to regularization selection (StARS) for high dimensional graphical models[J]. Adv Neural Inf Process Syst, 2010, 24(2):1432-1440.
- [8] Wong YH, Wu CC, Lin CL, et al. Applying ngs data to find evolutionary network biomarkers from the early and late stages of hepatocellular carcinoma[J]. Biomed Res Int, 2015(1):391475.
- [9] Cornella H, Alsinet C, Sayols S, et al. Unique genomic profile of fibrolamellar hepatocellular carcinoma[J]. Gastroenterology, 2015, 148(4):806-818.
- [10] Lin Z, Xu SH, Wang HQ, et al. Prognostic value of DNA repair based stratification of hepatocellular carcinoma[J]. Sci Rep, 2016, 6:25999.
- [11] Huang JL, Cao SW, Ou QS, et al. The long non-coding RNA PT-TG3P promotes cell growth and metastasis via up-regulating PTTG1 and activating PI3K/AKT signaling in hepatocellular carcinoma[J]. Mol Cancer, 2018, 17(1):93.
- [12] Mallol C, Casana E, Jimenez V, et al. AAV-mediated pancreatic overexpression of *Igf1* counteracts progression to autoimmune diabetes in mice[J]. Mol Metab, 2017, 6(7):664-680.
- [13] 何强, 梁力建, 彭宝岗, 等. ErbB2 通过 FAK-Src-MAPK 信号通路诱导细胞转化和移动侵袭[J]. 中国病理生理杂志, 2008, 24(12):2363-2365.
- [14] 易善勇, 杨晶, 官丽莉, 等. 成纤维细胞生长因子 9(FGF9)的研究进展[J]. 中国生物工程杂志, 2015, 35(7):94-101.
- [15] Hu B, Zhang Y, Lu J, et al. Binding of the pathogen receptor HSP90AA1 to avibirnavirus VP2 induces autophagy by inactivating the AKT-MTOR pathway[J]. Autophagy, 2015, 11(3):503-515.
- [16] Seshacharyulu P, Pandey P, Datta K, et al. Phosphatase: PP2A structural importance, regulation and its aberrant expression in cancer[J]. Cancer Lett, 2013, 335(1):9-18.

收稿日期:2018-07-26