

文章编号: 1671-4229(2024)01-0012-08

基于自编码器的疾病相关 miRNAs 的预测方法

许鹏^{1,2}, 谢斌¹, 鲍振申¹, 李先彬¹, 刘文斌^{1*}

(1. 广州大学 计算科技研究院, 广东 广州 510006; 2. 黔南民族师范学院 计算机与信息学院, 贵州 都匀 558000)

摘要: MicroRNAs (miRNAs) 是一类由内源基因编码的长度约为 22 个核苷酸的非编码单链 RNA 分子, 它们在动植物中参与转录后基因表达调控。大量研究表明, miRNAs 在包括肿瘤在内的多种复杂疾病发生、发展过程中扮演着重要的角色。因此, 识别疾病相关的 miRNAs 对研究疾病的机理及治疗具有重要意义。鉴于湿实验验证方法存在耗时长、成本高的缺点, 当前许多研究工作聚焦于开发高效计算模型, 识别新的 miRNA-disease 关联关系。该研究提出一种基于自编码器数据驱动模型, 预测 miRNA-disease 关联关系。结果表明, 作者预测的疾病相关 miRNAs 在 HMDD 数据库中对应的疾病相关 miRNAs 列表上显著富集。此外, 通过对排名靠前的 miRNAs 分析, 发现这些 miRNAs 具有重要的生物学功能, 同时对于疾病的分类表现出较高的精度。总之, 文章提出的模型, 对于疾病相关 miRNAs 的发现具有重要的辅助作用。

关键词: 自编码器; miRNA-disease 关联; 数据驱动模型

中图分类号: TP333; Q811.4 **文献标志码:** A

Predictive method for disease-associated miRNAs based on autoencoders

XU Peng^{1,2}, XIE Bin¹, BAO Zhen-shen¹, LI Xian-bin¹, LIU Wen-bin^{1*}

(1. Institute of Computational Science and Technology, Guangzhou University, Guangzhou 510006, China;

2. School of Computer Science of Information Technology, Qiannan Normal University for Nationalities, Duyun 558000, China)

Abstract: MicroRNAs (miRNAs) are a class of non-coding single-stranded RNA molecules encoded by endogenous genes, with a length of about 22 nucleotides. They participate in post-transcriptional gene expression regulation in animals and plants. Numerous studies have shown that miRNAs play crucial roles in the occurrence and development of various complex diseases, including tumors. Therefore, identifying disease-related miRNAs is of significant importance for studying disease mechanisms and developing appropriate treatment strategies. Given the time-consuming and costly nature of wet-lab validation methods, many researchers focus on developing efficient computational models to identify novel miRNA-disease associations. This study proposes a data-driven model based on autoencoders to predict miRNA-disease associations. The results indicate that the disease-related miRNAs predicted are significantly enriched in the list of disease-related miRNAs from the HMDD database. Furthermore, by analyzing the top-ranked miRNAs, it was found that these miRNAs perform crucial biological functions and accurately exhibit disease classification. In conclusion, the model proposed in this paper serves as a valuable auxiliary tool for the discovery of disease-related miRNAs.

Key words: autoencoders; miRNA-disease associations; data-driven model

收稿日期: 2023-09-08; 修回日期: 2023-09-26

基金项目: 国家自然科学基金资助项目(62002079, 62072128, 62102104 和 62202115); 广东省自然科学基金资助项目(2023A1515011401); 广东省医学影像智能分析与应用重点实验室开放课题资助项目(2022B1212010011); 广州市科技局市校联合基金资助项目(SL2022A03J00935)

作者简介: 许鹏(1986—), 男, 副教授. E-mail: gzxupeng@gzhu.edu.cn

* 通信作者. E-mail: wbliu6910@gzhu.edu.cn

引文格式: 许鹏, 谢斌, 鲍振申, 等. 基于自编码器的疾病相关 miRNAs 的预测方法[J]. 广州大学学报(自然科学版), 2024, 23(1): 12-19.

MiRNAs 是一类长度约为 22 个核苷酸的非编码单链 RNA 分子,它们通过碱基互补配对原则与靶向 mRNA 的 3'端非编码区域特异性结合,从而调节转录后的基因表达^[1-3]。一般认为,miRNA 通过与靶 mRNA 的特异性结合对靶基因的表达起到抑制的效果^[4]。近年来,陆续有研究表明,miRNA 也能够促进靶基因的表达^[5-6]。通过文献检索,课题组将基于实验验证得到的正调控 miRNA-gene 关系对整理汇总并开发了一个数据库 PmiR-tarbase^[7]。自从 1993 年,miRNA 被首次报道以来,大量的证据表明,miRNA 参与包括细胞衰老、凋亡、分化及增殖等多种生物过程^[8-11]。此外,越来越多的研究发现,miRNA 在包括癌症、心血管疾病、神经系统疾病等多种疾病的发生和发展中发挥着重要的调控作用^[12-14]。因此,miRNA-disease 关联的预测可以从遗传角度为疾病发病机制提供新的视角。对于 miRNA-disease 关联的预测,生物学湿实验是最直接、有效的方法,但是该方法同时存在费时、费力等缺点。因此,越来越多的计算方法被开发出来,以预测潜在的 miRNA-disease 关联关系^[15-17]。现有的计算模型大体上可分为两类:基于相似性和基于机器学习的模型。

基于相似性的模型在预测 miRNA-疾病关联关系方面取得了显著进展,它主要基于一个众所周知的生物学假设:功能相似的 miRNAs 往往与表型相似的疾病相关,相反亦然^[15, 18-19]。Jiang 等^[20]通过集成 miRNA 功能相似性网络、疾病相似性网络以及 miRNA-disease 网络,基于邻居连接信息提出一种计算模型预测潜在的 miRNA-disease 关联关系。然而,该方法主要是利用网络上邻居连接关系,这也为后续基于全局网络方法的提出留有余地。Mork 等^[21]利用验证的 protein-miRNA 关系和基于文本挖掘得到的 protein-disease 关联关系,提出了 miRNA-disease 关系预测方法 miR-PD。在他们的工作中,通过将 miRNA-protein 相互作用的得分与 protein-disease 关联的得分相乘,从而得到 miRNA-disease 关联评分。需要指出的是,对于没有共同的蛋白质连接的 miRNA-disease 关联不适用于该方法。Chen 等^[22],基于重启随机游走,提出了一种名为 RWRMDA 的预测模型。尽管该方法在预测方面表现出非常高的准确性,但无法应用于没有已验证疾病关联的 miRNA。

另外,许多基于机器学习的模型通过调整模

型参数或融合多源生物数据,成功提高了预测 miRNA-disease 关联关系的准确性。Chen 等^[23]提出了一种基于受限玻尔兹曼机的 miRNA-disease 关联预测模型,该模型不但在总体准确度方面有所提高,同时还能够区分多种类型的 miRNA-disease 关联关系。Ha 等^[24]提出了基于矩阵补全和网络正则化的方法,该模型使用了矩阵分解技术,将 miRNA 相似性网络数据作为正则化项,并通过考虑直接邻居关系实现 miRNA-disease 关系预测。Wang 等^[25]提出了一种名为 SAEMDA 的模型,在该方法中作者使用堆积自编码器以无监督的方式预训练了模型,然后采用同等数量的正负样本对模型进一步训练,最终在 miRNA-disease 关联关系预测方面得到了非常好的效果。借助堆积自编码器,Liu 等^[26]也提出了一种新的计算模型,该模型基于 miRNA 功能相似性网络、疾病语义相似性、miRNA 潜层特征以及疾病潜层特征,获得 miRNA-disease 特征向量,最后使用集成学习 XGBoost 模型预测新的 miRNA-disease 关系。

综上所述,不管是基于相似性还是机器学习的方法,在构建预测模型的过程中都需要应用到先验生物网络知识,以及关于 miRNA-disease 关联关系的验证集。虽然这些方法在预测性能上均得到非常可观的指标,但是在预测 miRNA-disease 关联时更多倾向于关注已知的关联,对于新的、未知的关联的预测能力受到一定限制。本研究中,为了避免预测结果的有偏性,不依靠生物网络知识、以及验证的 miRNA-disease 关联关系,应用深度学习学习方法提出一种新的预测模型。

1 数据和方法

1.1 数据

采用 TCGA (TCGA 是美国国家癌症研究所 (National Cancer Institute, NCI) 和美国人类基因组研究所 (National Human Genome Research Institute, NHGRI) 于 2006 年共同启动的一个项目) 中 miRNAs 表达数据,包括 33 种癌症及其对应正常组织的共 11 020 个样本,每个样本由 1 881 个 miRNAs 构成。为了避免批次效应和加速模型的训练,采用公式 (1) 将 miRNA 表达数据标准化到 0 ~ 1 之间:

$$e_{ij} = \frac{E_{ij} - \min(E_j)}{\max(E_j) - \min(E_j)},$$

$$i \in (1, 1881), j \in (1, 11020), \quad (1)$$

其中, E_{ij} 表示第 j 个样本、第 i 个 miRNA 的表达值, E_j 表示第 j 个样本中所有 miRNAs 的表达值, e_{ij} 表示标准化后的 miRNA 表达量。

1.2 方法

本文提出了一种基于纯数据驱动预测疾病相关 miRNAs 的方法。图 1 展示了该方法的主要流程:①基于所有的 miRNA 表达谱样本数据, 训练一个 miRNA 自编码器模型, 见图 1(a);②将疾病表型向量与图 1(a) 训练的编码器中的某个隐藏层进行反向回归训练, 进而与图 1(a) 后半部分模型进行拼接, 得到输入层为疾病表型与输出层为 miRNA 表达的多层神经网络, 见图 1(b);③采用类似于一种点亮机制, 点亮输入层中某一个节点的值为 1, 控制其余输入节点为 0, 观测输出层 miRNAs 的打分排序, 从而得到与点亮节点即疾病表型相关的 miRNAs, 见图 1(c)。

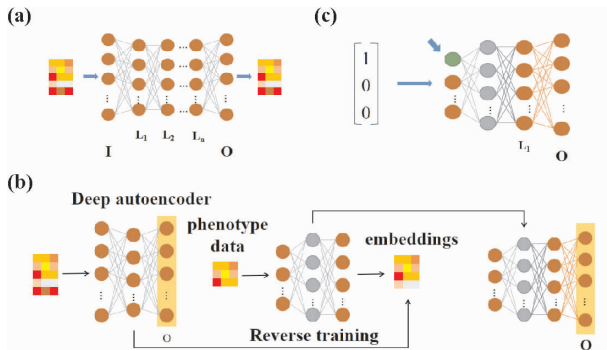


图 1 miRNA 自编码器和反向训练示意图

Fig. 1 Schematic diagram of miRNA autoencoders structure and reverse training

1.2.1 miRNA 自编码器模型

图 2 为 miRNA 自编码器模型示意图。其中, 归一化表达式矩阵 $[e_{ij}]$ 分别作为输入和输出, 自编码器采用全连接, 每层的激活函数采用 sigmoid 函数训练自编码器。使用优化器 ADAM, 初始化采用参数: 学习率 $l = 1.0 \times 10^{-3}$, $\beta_1 = 9.0 \times 10^{-1}$, $\beta_2 = 9.99 \times 10^{-1}$, $\varepsilon = 1.0 \times 10^{-8}$ 和衰减 $d = 1.0 \times 10^{-6}$ 。模型训练的批量规模为 100。为了挑选出适合本研究的最优编码器模型, 分别设置网络的层数为 1、2 和 3, 节点数为 100、200、300 和 400, 总共训练了 12 个模型。

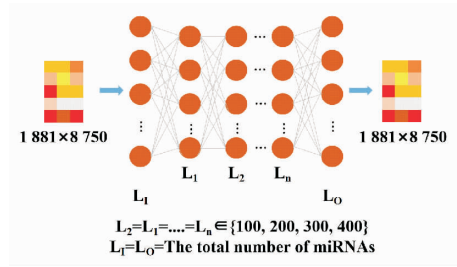


图 2 miRNA 自编码器结构示意图

Fig. 2 Schematic diagram of miRNA autoencoders structure

本文所采用的损失函数为均方误差 MSE :

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (2)$$

其中, n 表示 miRNAs 的数量, 在本文中为 1 881。

1.2.2 反向监督模型预测疾病相关 miRNAs

为了将疾病表型与 miRNA 的低维特征构建起联系, 采用反向监督模型, 该方法具体包含以下 4 个步骤:

(1) 利用训练好的自编码器将原始数据进行压缩;

(2) 在“反方向”上训练一个监督模型: 模型是一个单隐层监督神经网络, 在第二层和第三层有相同数量的节点, 分别加上 sigmoid 激活函数以及线性激活函数。模型的输出为 $H \times S$ 的表型矩阵 $[s_{kj}]$, H 是隐层中节点的数量。模型的输入矩阵为 $P \times S$ 的压缩矩阵 $[C_{ij}]$, $i \in [1, P]$, $j \in [1, S]$, 其中, P 和 S 分别为表型和样本的数量;

(3) 将(2)训练好的监督模型与 miRNA 自编码器模型左侧进行拼接;

(4) 根据输出端 miRNA 的表达得到疾病相关性评分: s^p 的绝对值, 用于确定与 p^{th} 表型相关的 miRNAs 的优先级, 通过堆叠神经网络的参数计算, 使用:

$$x^k = \begin{cases} f^k(W^k x^{k-1} + b^k) & \text{if } 1 < k \leq L-1, \\ r^p & \text{if } k = 1, \end{cases} \quad (3)$$

$$s^p = W^L x^{L-1}, \quad (4)$$

其中, $r^p = [\delta_{ip}]$ 是第 p 个表型向量, $i \in \{1, 2, 3, \dots, P\}$ 。在该模型中, 根据 miRNA 编码器隐层表示, 反向推测 miRNA 的表达, 因此, 也叫反向监督模型。

2 结果与讨论

2.1 miRNA 自编码器模型训练

为了探究 miRNA 表达谱的低维特征表示, 基

于 miRNA 表达数据训练 miRNA 自编码器模型。训练神经网络需要大量的、质量良好的数据,因此,在本研究中,使用 TCGA 中 11 020 个样本的 miRNA 表达数据进行分析,其中,每个样本的特征维度为 1 881,即由 1 881 个 miRNAs 构成。该数据集总共包含了 33 种癌症及其癌旁正常组织样本。从该数据集中随机选取 8 750 个样本作为训练集,剩余的 2 270 个样本作为测试集。

训练从 100 到 400 个隐藏节点的不同宽度自动编码器,以 100 为步长逐步递增。另外,分别考虑了单隐层、双隐层、三隐层自编码器对数据的拟合能力,其中,双隐层和三隐层自编码器,称之为深度自编码器。本研究中,采用拟合优度 R^2 来评估模型对 miRNA 表达数据的拟合程度好坏。分别计算每个 miRNA 的 R^2 ,并统计所有 miRNAs 的 R^2 的平均值和中位数,具体结果如图 3 所示。

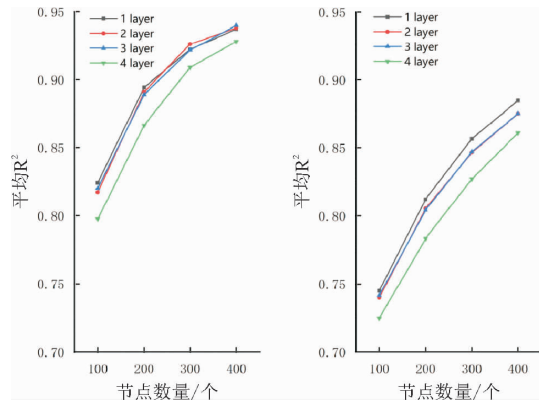


图 3 12 种模型训练结果 R^2 展示

Fig. 3 Display of training results for 12 models

通过比较不同结构自编码器的重构拟合度,发现相同结构的自编码器 R^2 的均值要大于中位

数,这一现象主要得益于部分 miRNA 的重构拟合度非常高。随着隐层数量的增加到 4 层的时候,不管是 R^2 平均值还是中位数均有明显的下降,这说明隐层数量 3 层及以内就足以表达编码 miRNA。此外,由图 3 可见,具有 400 个隐藏节点的单隐层自编码器与相同节点数的双层和三层深度自编码器表现相当。因此,最终选择具有 400 个隐藏节点的单隐层自编码器作为后续研究。

2.2 获取与疾病相关 miRNAs

为了后续“反向训练”能够提取疾病足够的特征,故对于待研究的疾病需要一定数量的样本。因此,本研究中选择数量最多的 3 种癌症:乳腺癌 (Breast Cancer, BRCA)、低质胶质瘤 (Lower Grade Glioma, LGG) 和肺腺癌 (Lung Adenocarcinoma, LUAD),作为分析研究对象。

将疾病表型向量(以 one-hot 编码表示 miRNA 样本对应的疾病类别)作为输入,miRNA 自编码器模型中的隐层节点表示作为输出,训练一个神经网络模型见图 1(b)。进而,与原始 miRNA 自编码器模型的左侧拼接,得到一个反向编码器模型。最后,根据输入的疾病表型向量,获取每种疾病的 miRNAs 排序。通过超几何检验方法,见公式(5),计算所得排序的 Top 30、Top 60 以及 Top 90 miRNAs 在 HMDD 数据上的富集程度^[27]。

$$p(x = k) = \frac{C_M^k \times C_{N-M}^{n-k}}{C_N^n}, \quad (5)$$

其中, N 表示数据集中 miRNAs 的总数量, M 表示 N 中与疾病相关 miRNA 数量, n 的值根据选取的排序分别 Top 30、Top 60 和 Top 90, k 表示 n 中与疾病相关 miRNAs 数量,如表 1 所示。

表 1 疾病相关 miRNAs 预测结果

Table 1 Prediction results of disease-related miRNAs

Top miRNAs	BRCA		LGG		LUAD	
	Count	p_value	Count	p_value	Count	p_value
Top 30	17	1.3×10^{-4}	6	1.1×10^{-2}	8	3.2×10^{-2}
Top 60	30	1.1×10^{-5}	12	4.2×10^{-4}	18	6.6×10^{-4}
Top 90	42	2.0×10^{-6}	18	1.6×10^{-5}	29	3.9×10^{-6}

表 1 给出了疾病相关 miRNAs 预测结果,结果显示本文方法预测的 3 种癌症相关 Top 30、Top 60 和 Top 90 的 miRNAs 均显著富集($2.0 \times 10^{-6} \leq p_value \leq 3.2 \times 10^{-2}$) 在 HMDD 收集的疾病相关的 miRNAs 数据集上。此外,通过文献检索,虽然有

的 miRNA 没有出现在 HMDD 数据库中,但是有文献表明它与对应癌症有着重要的关系。如 hsa-mir-374c,其在 BRCA 疾病相关的 miRNAs 预测列表中均排名第一。Hao 等^[28]研究发现,通过上调 hsa-mir-374c 能够抑制乳腺癌细胞增殖、迁移以及

上皮-间质转化(EMT)。Lu 等^[29]实验证实 hsa-mir-627,能够通过靶向抑制 TRIM56 的表达从而影响 LUAD 的预后。综上所述,无偏非线性方法确实可以有效识别疾病相关的 miRNAs。但是,相较于其他基于相似性或者机器学习预测方法,本方法无法与其媲美,主要是由于本方法没有使用任何生物网络及关联关系方面的先验知识,而是依靠纯数据驱动的方式进行预测。

2.3 Top 30 miRNAs 作为疾病分类标志物验证

为了进一步验证模型筛选的与疾病相关 miRNA 的合理性,利用已预测的与疾病相关的 miRNAs 作为分类特征,对 3 种疾病进行分类。该分类任务主要研究两个问题:①以每种疾病相关 miRNAs 排序的前 30 个 miRNAs 作为特征,对癌症

及其癌旁正常组的进行二分类;②将每种疾病 miRNAs 排序的前 30 个 miRNAs 合并之后作为特征,对 3 种癌症进行分类,即三分类任务。两个分类任务中,分类器均选择支持向量机(SVM),采用 50 次 5 折交叉验证,结果取均值。

图 4 展示出基于 Top 30 miRNAs 采用 SVM 模型进行癌症分类的结果,其中前三幅图分别为疾病 BRCA、LGG 和 LUAD 与相对应正常组的二分类 ROC 曲线图,最后一幅图是 3 种疾病(BRCA、LGG、LUAD)的三分类效果图。无论是疾病与正常组,还是 3 种疾病的分类效果都表现良好(ROC > 0.94),这也间接说明了预测的与疾病相关的 miRNAs 的准确性。

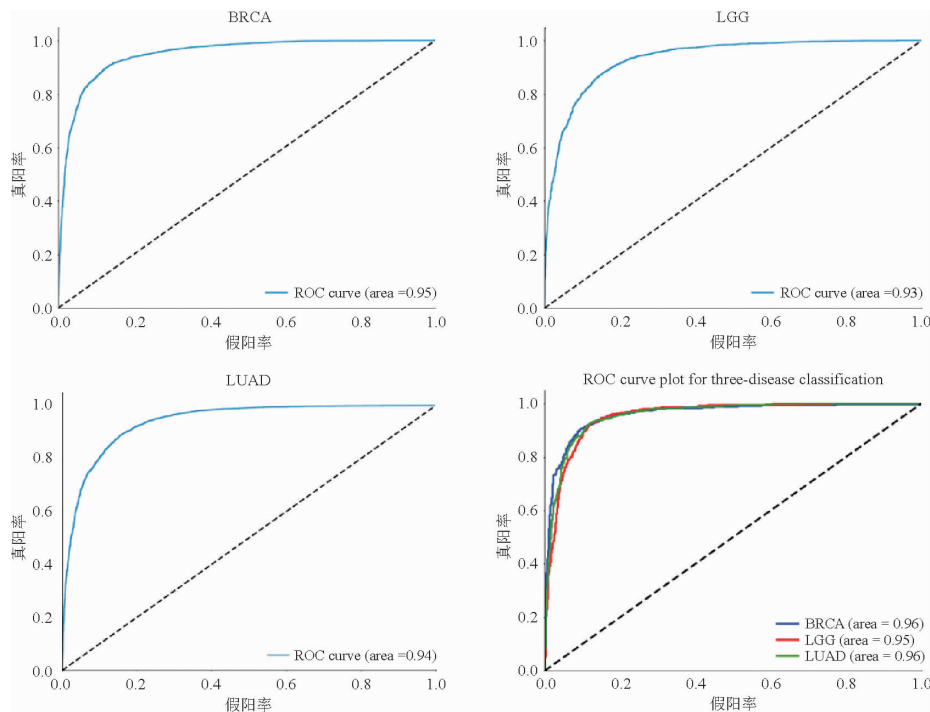


图 4 基于 Top 30 miRNAs 采用 SVM 模型进行癌症分类

Fig. 4 Using SVM model for cancer classification based on Top 30 miRNAs

2.4 miRNA 生物功能分析

癌症是一种复杂的系统性失调疾病,其发生、发展一定是伴随着多个 miRNAs 的表达紊乱,这些 miRNAs 应该是相互协作共同完成特定的功能。因此,通过绘制预测的疾病相关 miRNAs 功能相似性网络,来探究这些 miRNAs 是否会形成紧密的网络。MISIM^[30]提供绘制 miRNAs 功能相似性网络的工具,该工具整合了共表达相似性、共 GO 模块相似性、共文献相似性和共相似性疾病相似性等。

利用 MISIM,绘制每种疾病相关的前 50 个 miRNAs 功能相似性网络。图 5 展示了 3 种疾病相关 miRNAs 功能相似性网络,其中,蓝色椭圆表示 miRNA,红色的边表示 2 个 miRNA 功能相似性较大 $\in (0.5, 1]$,橙色的边表示 2 个 miRNA 功能相似性较小 $\in (0, 0.5]$ 。在前 50 个 miRNAs 中,有相当一部分 miRNAs 能够形成紧密的网络,这说明这些 miRNAs 在癌症发展过程中可能相互协作影响着癌症的进程。

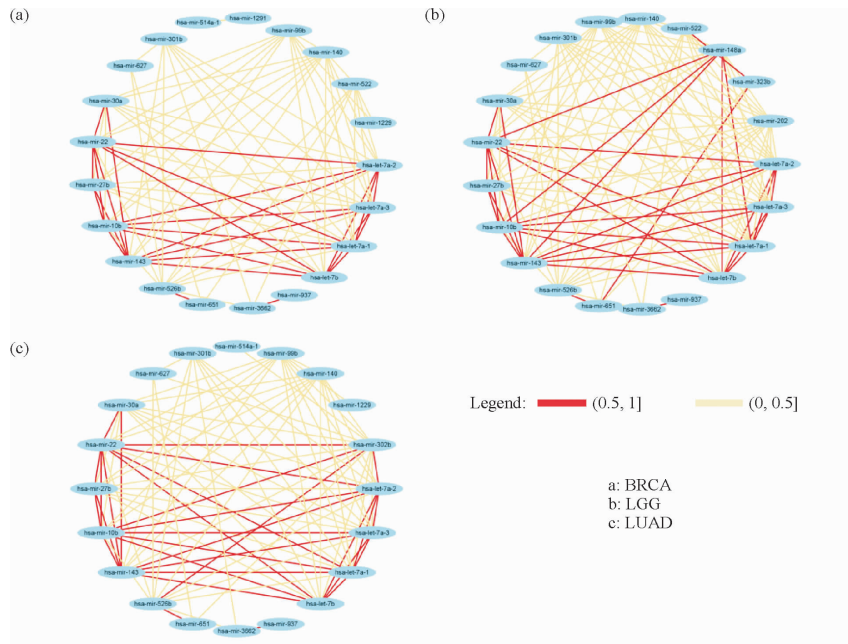


图 5 3 种疾病相关 miRNAs 功能相似性网络

Fig. 5 Functional similarity networks of miRNAs related to three disease

图 6 展示了本文方法预测 3 种疾病相关的前 50 个 miRNAs 显著富集的 ($P_{value} < 0.05$) 生物信号通路。由图 6 可见, ①这些 miRNAs 显著富集在一些癌症通路上, 如 Colorectal cancer、Gastric cancer、Non-small cell lung cancer、Breast cancer 以及 Bladder cancer 等通路, 这说明本文预测的 miRNAs

在癌症中确实发挥着重要角色。②这些 miRNAs 也富集在癌症发生、发展相关的相关通路上, 如 Cell cycle、Apoptosis、PI3K-Akt 和 P53 等信号通路, 已有大量实验研究结果表明, 这些信号通路与癌症发展密切相关。

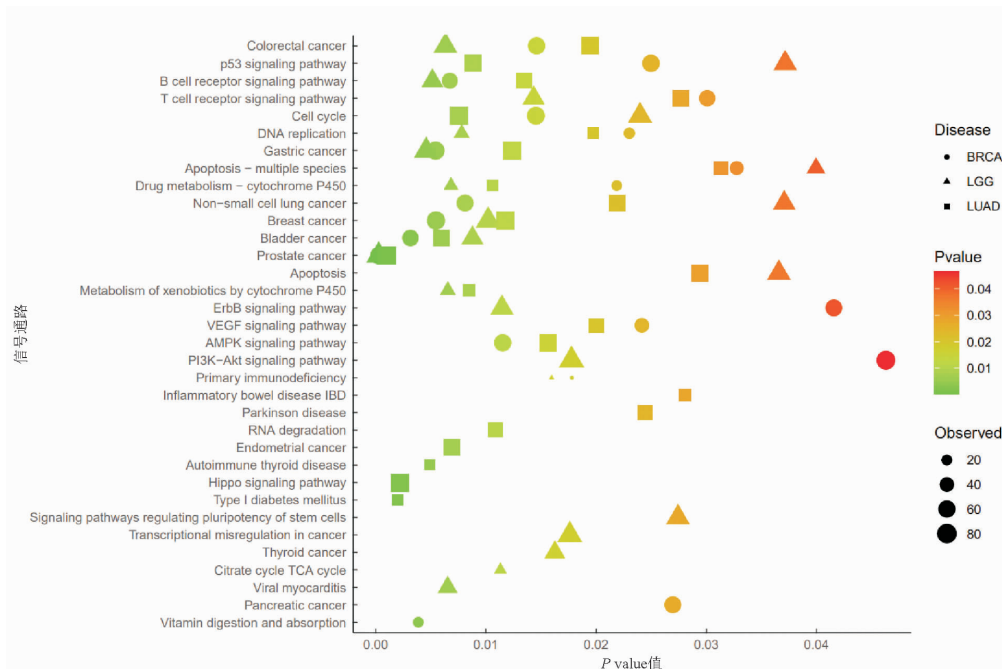


图 6 miRNA 生物信号通路富集分析

Fig. 6 Enrichment analysis of miRNA biological signaling pathways

3 总 结

miRNAs 是一类非编码短序列 RNA 分子,它能够通过对靶向调控编码基因的表达调节特定生物功能,进而影响疾病的进程。大量的研究表明,miRNAs 与疾病发生、发展,以及肿瘤的预后有重要的关系。本文提出一种基于自编码器数据驱动的模式,预测 miRNA-disease 关联关系。该方法主要基于 miRNA 表达谱数据,利用深度神经网络提取各种疾病在 miRNA 表达水平上的潜层特征,从而推测出与疾病相关的 miRNAs。基于深度神经网络自动提取特征,相较于人工提取特征来说,更具客观性、无偏性,这为本文成功预测疾病相关的 miRNAs 奠定了理论基础。在 BRCA、LGG 和 LU-

AD 3 种癌症数据集上的结果表明,预测的疾病相关 miRNAs 在 HMDD 数据库中对应的疾病相关 miRNAs 列表上显著富集。通过对排名靠前的 miRNAs 分析发现,这些 miRNAs 具备癌症分类标识物的潜质,并且参与癌症相关重要的生物信号通路。

与已有的预测疾病相关 miRNAs 方法相比,本文的方法不需要使用 miRNA 相似性网络、疾病相似性网络以及 miRNA-disease 关联关系等先验知识,因此,该方法适用性更强。此外,本文方法不会受到已有先验知识的限制,有利于发现新的 miRNA-disease 关联关系。未来,仍需要生物学实验对本方法预测的新的 miRNA-disease 关系作进一步的验证。

参考文献:

- [1] Ambros V. The functions of animal microRNAs[J]. *Nature*, 2004, 431(7006): 350-355.
- [2] Bartel D P. MicroRNAs: Genomics, biogenesis, mechanism, and function[J]. *Cell*, 2004, 116(2): 281-297.
- [3] 许鹏, 咎乡镇, 路德杨, 等. 疾病相关 miRNAs 识别方法研究进展[J]. *广州大学学报(自然科学版)*, 2019, 18(1): 54-64.
- [4] Wu Z W, Lu H F, Sheng J F, et al. Inductive microRNA-21 impairs anti-mycobacterial responses by targeting IL-12 and Bcl-2[J]. *Febs Letters*, 2012, 586(16): 2459-2467.
- [5] Tsai N P, Lin Y L, Wei L N. MicroRNA mir-346 targets the 5'-untranslated region of receptor-interacting protein 140 (RIP140) mRNA and up-regulates its protein expression[J]. *Biochemical Journal*, 2009, 424: 411-418.
- [6] Vasudevan S, Steitz J A. AU-rich-element-mediated upregulation of translation by FXR1 and Argonaute 2[J]. *Cell*, 2007, 128(6): 1105-1118.
- [7] Xu P, Li X, Liang Y, et al. PmiRtarbase: A positive miRNA-target regulations database[J]. *Computational Biology and Chemistry*, 2022, 98: 107690-107695.
- [8] Bartel D P. MicroRNAs: Target recognition and regulatory functions[J]. *Cell*, 2009, 136(2): 215-233.
- [9] Xu P, Wu Q, Lu D, et al. A systematic study of critical miRNAs on cells proliferation and apoptosis by the shortest path[J]. *BMC Bioinformatics*, 2020, 21(1): 396-410.
- [10] Xu P, Wu Q, Rao Y, et al. Predicting the influence of microRNAs on drug therapeutic effects by random walking[J]. *IEEE Access*, 2020, 8: 117347-117353.
- [11] Xu P, Zhu Y, Sun B, et al. Colorectal cancer characterization and therapeutic target prediction based on microRNA expression profile[J]. *Scientific Reports*, 2016, 6: 20616-20627.
- [12] Yu Z, Li Z, Jolicoeur N, et al. Aberrant allele frequencies of the SNPs located in microRNA target sites are potentially associated with human cancers[J]. *Nucleic Acids Research*, 2007, 35(13): 4535-4541.
- [13] Liu W, Cui Z, Zan X. Identifying cancer-related microRNAs based on subpathways[J]. *IET Systems Biology*, 2018, 12(6): 273-278.
- [14] Xu P, Wu Q, Yu J, et al. A systematic way to infer the regulation relations of miRNAs on target genes and critical miRNAs in cancers[J]. *Frontiers in Genetics*, 2020, 11: 278-291.
- [15] Chen X, Xie D, Zhao Q, et al. MicroRNAs and complex diseases: From experimental results to computational models[J]. *Briefings in Bioinformatics*, 2019, 20(2): 515-539.
- [16] Ha J, Park S. NCMD: Node2vec-based neural collaborative filtering for predicting MiRNA-Disease association[J]. *IEEE/*

- ACM Transactions on Computational Biology and Bioinformatics, 2023, 20(2):1257-1268.
- [17] Ha J. SMAP: Similarity-based matrix factorization framework for inferring miRNA-disease association[J]. Knowledge-Based Systems, 2023, 263:110295-110306.
- [18] Ha J, Park C, Park C, et al. IMIPMF: Inferring miRNA-disease interactions using probabilistic matrix factorization[J]. Journal of Biomedical Informatics, 2020, 102: 103358-103369.
- [19] Zou Q, Li J, Song L, et al. Similarity computation strategies in the microRNA-disease network: A survey[J]. Briefings in Functional Genomics, 2016, 15(1):55-64.
- [20] Jiang Q, Hao Y, Wang G, et al. Prioritization of disease microRNAs through a human phenome-microRNAome network [J]. BMC Systems Biology, 2010, 4(S1): 2-11.
- [21] Mork S, Pletscher-Frankild S, Paljeja C A, et al. Protein-driven inference of miRNA-disease associations[J]. Bioinformatics, 2014, 30(3): 392-397.
- [22] Chen X, Liu M X, Yan G Y. RWRMDA: Predicting novel human microRNA-disease associations[J]. Molecular Biosystems, 2012, 8(10): 2792-2800.
- [23] Chen X, Yan C, Zhang X, et al. RBMMMDA: Predicting multiple types of disease-microRNA associations[J]. Scientific Reports, 2015, 5: 13877-13890.
- [24] Ha J, Park C, Park S. PMAMCA: Prediction of microRNA-disease association utilizing a matrix completion approach[J]. BMC Systems Biology, 2019, 13(1):33-46.
- [25] Wang C C, Li T H, Huang L, et al. Prediction of potential miRNA-disease associations based on stacked autoencoder[J]. Briefings Bioinformatics, 2022, 23(2):1-11.
- [26] Liu D, Huang Y, Nie W, et al. SMALF: miRNA-disease associations prediction based on stacked autoencoder and XGboost [J]. BMC Bioinformatics, 2021, 22(1): 219:1-18.
- [27] Huang Z, Shi J, Gao Y, et al. HMDD v3.0: A database for experimentally supported human microRNA-disease associations[J]. Nucleic Acids Research, 2018, 47(D1): D1013-D1017.
- [28] Hao S, Tian W, Chen Y, et al. MicroRNA-374c-5p inhibits the development of breast cancer through TATA-box binding protein associated factor 7-mediated transcriptional regulation of DEP domain containing 1[J]. Journal of Cellular Biochemistry, 2019, 120(9): 15360-15368.
- [29] Lu K, Sui Y, Fu L. Identification of TRIM56 as a potential biomarker for lung adenocarcinoma[J]. Cancer Management and Research, 2021, 13: 2201-2213.
- [30] Li J, Zhang S, Wan Y, et al. MISIM v2.0: A web server for inferring microRNA functional similarity based on microRNA-disease associations[J]. Nucleic Acids Research, 2019, 47(W1): W536-W541.

【责任编辑: 陈 钢】