

大理大学毕业设计

基于 **Bimax** 双聚类算法的功能性

miRNA-mRNA 调控模块挖掘

Identifying Functional miRNA-mRNA Regulatory Modules

Based on Bimax Biclustering Algorithm

学 院： 工程学院

项目组成员： 赵艳珊、张雪枝、朱秋平

专 业： 生物医学工程

年级（班级）： 2013 级（1）班

指导教师： 张俊鹏 职称： 副教授

起止日期： 2016 年 11 月至 2017 年 6 月

制表日期： 2016 年 11 月 15 日

[摘要]

微小 RNA (microRNA, 简称 miRNA) 在基因调控网络中发挥重要作用, 它所组成的调控网络已经成为生命科学研究的热点和前沿。基于此, 本设计基于双聚类方法挖掘功能性 miRNA-mRNA 调控模块 (FMRMs)。所采用的 miRNA-mRNA 调控关系数据源来自两种类型: 实验验证型数据库 miRTarBase 和计算机预测型数据库 TargetScan。网络拓扑分析表明: 两种数据库所形成的 miRNA-mRNA 调控网络是无标度网络, 节点分布服从幂律分布 ($R^2 = 0.9797$ 和 $R^2 = 0.9430$)。为了进一步研究网络的模块特性, Bimax 双聚类方法用来识别 miRNA-mRNA 子集或簇, 这些子集或簇也被认为 miRNA-mRNA 调控模块。通过 GO 生物过程和 KEGG 信号通道富集发现, 许多 FMRMs 与生物过程和信号通道显著性相关。总之, miRNA-mRNA 调控模块有助于了解 miRNA 调节机制。

[关键词] miRNA; mRNA; 功能性 miRNA-mRNA 调控模块; 富集分析

[Abstract]

MicroRNA (miRNA) plays an important role in gene regulatory networks, and the study of miRNA regulatory network has become a hotspot and frontier of life science research. In this thesis, a biclustering method is utilized to infer functional miRNA-mRNA regulatory modules (FMRMs). We collect two types of miRNA-mRNA regulation data sources: predicted database called TargetScan and experimentally validated database called miRTarBase. The network topology analysis indicates that the two miRNA-mRNA regulatory network are all scale-free networks, and the node distribution obeys a power-law distribution ($R^2 = 0.9797$ and $R^2 = 0.9430$). To further investigate the modularity of the two networks, a biclustering method called Bimax is used to identify miRNA-mRNA subsets or clusters. These subsets or clusters are also considered as miRNA-mRNA regulatory modules. Functional GO and KEGG enrichment analysis show that several FMRMs are closely associated with many biological processes and pathways. In summary, the miRNA-mRNA regulatory modules could help to understand regulatory mechanisms of miRNA.

[Key Words] miRNA; mRNA; functional miRNA-mRNA regulatory module; biclustering method; enrichment analysis

目 录

项目组成员分工.....	1
引言.....	2
第一章 miRNA-mRNA 调控数据源.....	4
1.1 miRNA-mRNA 调控数据库.....	4
1.2 miRNA-mRNA 调控数据预处理.....	5
1.3 本章小结.....	5
第二章 功能性 miRNA-mRNA 调控模块识别.....	6
2.1 概述.....	6
2.2 基于 Bimax 双聚类方法的功能性 miRNA-mRNA 调控模块挖掘	7
2.3 本章小结.....	9
第三章 miRNA-mRNA 调控模块功能验证.....	10
3.1 概述.....	10
3.2 富集分析结果与分析.....	10
3.3 本章小结.....	13
第四章 总结与展望.....	14
4.1 总结.....	14
4.2 展望.....	14
参 考 文 献.....	15
致 谢.....	19

项目组成员分工

	学生姓名	学号	具体分工
负责人	赵艳珊	2013166147	调控模块的功能验证、结果分析、撰写论文
第 2 负责人	张雪枝	2013166114	查找国内外文献、调控模块的有效识别、撰写论文
第 3 负责人	朱秋平	2013166155	获取 miRNA-mRNA 调控网络数据，并对数据进行预处理；查找国内外文献、撰写论文

引言

miRNA 是一类短小的非编码 RNA 分子，与靶基因 mRNA 经过互补配对原则来调节后转录的基因表达水平，最终导致靶基因 mRNA 的降解或翻译抑制^[1]。miRNA 与多种生物过程相关，其中包括发育时间，细胞增殖，凋亡，应激反应和癌症^[2-4]。因此，miRNA 被认为是生物网络中的关键调控子之一。

目前研究表明，miRNA 与 mRNA 之间的调控形式呈现模块化。模块是细胞功能常展现的一种形式，作为发挥生物作用的主要载体。李霞^[5]等人将模块描述为“相互协调工作进而执行一致功能且联系紧密的一组生物分子（节点）”，这些实现各不相同生物作用的模块就称之为功能模块^[6]。Joung^[7]等人将 miRNA-mRNA 调控模块定义为“参与类似生物过程的 miRNA 和其靶基因 mRNA 的组”，识别 miRNA-mRNA 调控模块能够更深入地研究两者相互作用之间的密切联系。

目前已有相关计算方法来挖掘 miRNA-mRNA 调节模块来阐明 miRNA-mRNA 调节关系^[8]。Lionetti 等人^[9]整合计算机预测数据和 miRNA-mRNA 双表达谱数据来推定功能性 miRNA-mRNA 调节关系。Liu 等人^[10]利用 miRNA 和 mRNA 双表达谱，提出概率图形模型挖掘功能性 miRNA 调节模块。随着基因表达数据的增长，许多基于序列和基于进化的信息的方法来鉴定 miRNA-mRNA 靶向关系就显得不够精确。基于此，Lu 等人^[11]建立 Lasso 回归模型来构建 miRNA 调节网络，用于复杂疾病的诊断和治疗。

近年来，miRNA 研究主要集中于两种方法：实验方法和计算机方法。实验方法受到低效率，时间消耗和高成本的限制^[12]。因此，计算机方法作为候选策略来阐明 miRNA 功能的调节机制。如图 1 所示，本设计基于双聚类计算机方法来挖掘功能性 miRNA-mRNA 调控模块，旨在挖掘。

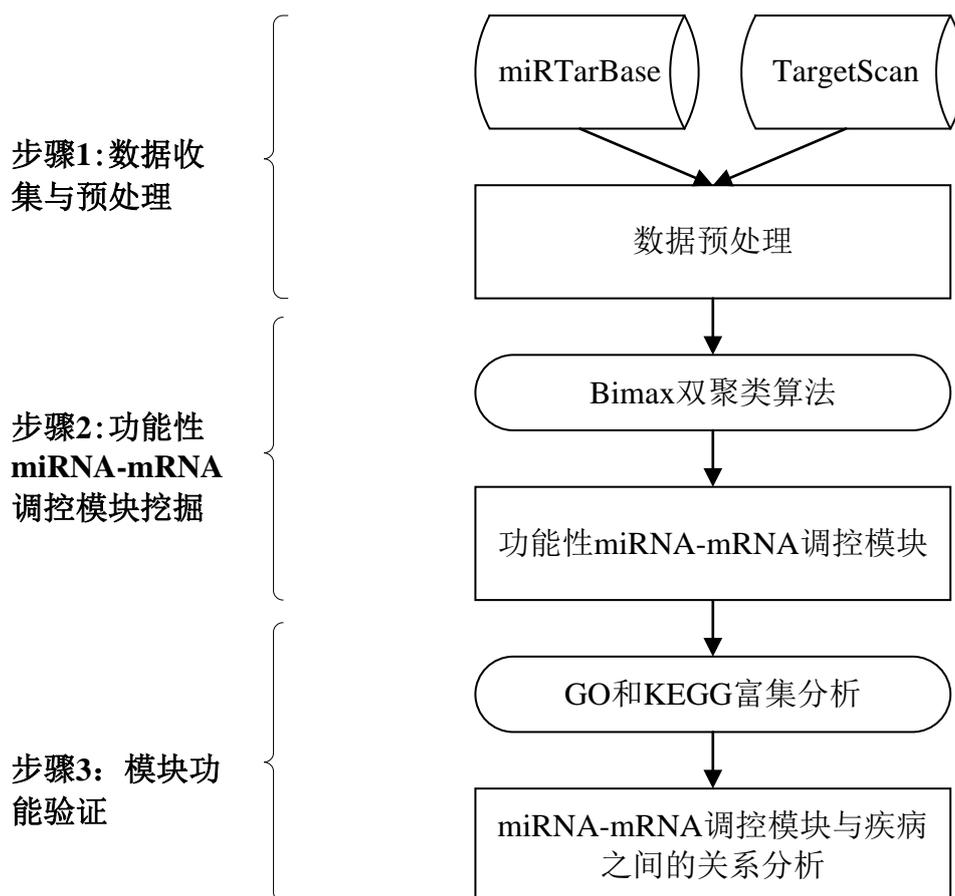


图 1 基于双聚类算法 Bimax 的功能性调控模块挖掘流程图

本设计主要由四个章节内容构成，具体如下：

第一章主要讲述了获取 miRNA-mRNA 调控网络数据源，选用了 miRTarBase 数据库和 TargetScan 数据库来构建 miRNA-mRNA 调控网络。

第二章针对功能性 miRNA-mRNA 调控模块概念以及模块识别算法进行阐述。在 R 软件 biclust 工具包中调用 Bimax 双聚类方法对 miRNA-mRNA 调控网络进行功能性模块挖掘。

第三章进行模块的功能验证。使用 R 工具包 clusterProfiler 对 miRNA-mRNA 调控模块进行富集分析，进而分析功能性调控模块与疾病之间的关系。

第四章是对本文进行总结，并对未来开展的研究内容进行展望。

第一章 miRNA-mRNA 调控数据源

1.1 miRNA-mRNA 调控数据库

目前，miRNA 数据资源众多，下表 1 汇总了一些常见调控数据库。本设计选用计算机预测型 TargetScan 数据库和实验验证型 miRTarBase 数据库分别来构建 miRNA-mRNA 调控关系网络。

表 1 常见 miRNA-mRNA 调控数据库汇总

序号	数据库	链接	参考文献
1	miRBase	http://www.mirbase.org/	[13]
2	miRanda	http://www.miRanda.org/	[14]
3	starBase	http://starbase.sysu.edu.cn/	[15]
4	miRTarBase	http://mirtarbase.mbc.nctu.edu.tw/	[16]
5	TargetScan	http://www.targetscan.org/	[17]
6	Tarbase	http://microrna.gr/tarbase/	[18]
7	PicTar	http://www.pictar.org/	[19]
8	microRA.org	http://www.microrna.org/	[20]
9	miRecords	http://mirecords.biolead.org/	[21]
10	PITA	http://genie.weizmann.ac.il/pubs/mir07/mir07_dyn_data.html	[22]
11	miRNAMap	http://mirnamap.mbc.nctu.edu.tw/	[23]
12	miRGen 2.0	http://www.diana.pcbi.upenn.edu/miRGen.html	[24]
13	StarScan	http://mirlab.sysu.edu.cn/starscan/	[25]
14	PMRD	http://bioinformatics.cau.edu.cn/PMRD/	[26]
15	RNA22	http://cbcsrv.watson.ibm.com/rna22.html	[27]

1.2 miRNA-mRNA 调控数据预处理

数据预处理是数据挖掘过程中的重要组成部分，主要满足数据处理的基本要求：噪声小、完整性好，属性全等。通过数据预处理，能够提高数据的质量，从而更加真实地反映实验结果。

本文从 miRTarBase 数据库和 TargetScan 数据库中获得 miRNA-mRNA 调控网络数据作为原始数据，从中筛选掉其他物种，只保留 Human 的，随后去除重复项，最终保存为.csv 格式。

1.3 本章小结

本文选用了验证型数据库 miRTarBase 和预测型数据库 TargetScan 分别构建 miRNA-mRNA 调控关系网络。并且对该调控网络数据进行清洗，去除重复项，确保每组调控关系是唯一的，使研究结果更加精确。

第二章 功能性 miRNA-mRNA 调控模块识别

2.1 概述

从许多研究中可以发现：miRNA 调控异常可能导致人类癌症的发生，例如乳腺癌、肺癌、前列腺癌等癌症。因此，通过靶基因 mRNA 间接了解 miRNA 参与的生物学过程和癌症的发生和发展具有重要作用。

miRNA 往往以模块的形式调控靶基因，本设计中，miRNA-mRNA 调控模块指在具有相似表达的一组 miRNAs 和 mRNAs^[28]，通过识别这种调控模块可以为进一步阐明复杂疾病的发病机制提供参考。

表 2 九种 miRNA-mRNA 调控模块识别算法的特点及其举例

识别方法	特点	算法举例	参考文献
基于双聚类方法	能够发现 miRNA 与其靶基因之间的潜在生物关系	Bimax、Cheng and Church 算法(CC)、FLOC 算法	[29]
基于关联规则	发现的 miRNA-mRNA 对具有较高的置信度并且 miRNA 与 mRNA 的表达模式高度相关	Apriori 算法、.基于划分的算法、FP-树频集算法	[30]
基于贝叶斯网络	可以用于表示人类基因和复杂疾病之间的关系	因子分析模型、独立分量分析模型	[31]
基于作者话题模型	能够发现 miRNA 与不同生物过程之间的关系	LDA 模型、TOT(Topic Over Time)模型、DTM (Dynamic Topic Model)模型	[32]
基于 Corr-LDA 模型	通过 miRNA 和 mRNA 的表达谱数据建立它们之间的相关系数	Corr-LDA 模型	[33]
基于 miRNA 调控网络	所识别模块内 miRNA 和 mRNA 之间具有较强的负相关性，而且具有显著的	Mirsynergy 算法	[34]

功能富集性			
7、基于二分团合并的算法	识别出的模块效果较好，且克服了在相对稠密 miRNA 调控网络中识别效果不明显的缺点	BiCliques Merging (BCM) 算法	[35]
8、基于多类型数据	能够预测未知的 miRNA 及其靶基因之间有意义的生物关系	MiRMD(miRNA-mRNA regulatory modules detection)算法	[36]
9、基于集体关系	识别出的模块内分子之间的紧密度较高，更能符合 miRNA 调控模块的紧密性特征	CGR(collective group relationships)算法	[37]

随着时间的推移，双向聚类方法应用的范围越来越广，在本文中，我们采用了 R 语言提供的 biclust 工具包，其中运用的是 Bimax 双聚类算法。该算法是 Prelic^[38]等人于 2006 年提出的一个简单的聚类模型，为双聚类验证而设计，在研究中它被认为是一种参考方法，它是反映双聚类基本思想的数据模型，且假设每个基因只有两个表达水平：变化（1）和不变（0），允许在合理的时间内找到最大包含的聚类，确定所有最优的双聚类，找到二元矩阵中 1 的集合。

Bimax 双聚类算法是迭代的，其思想是：

- (1) 重新排列行和列，将二元矩阵中的所有 1 集中在矩阵的右上角；
- (2) 将矩阵分为两个子矩阵。若一个子矩阵中总只有 1，则返回该子矩阵；
- (3) 继续在这两个子矩阵中迭代，直到得到最后的结果^[39]。

2.2 基于 Bimax 双聚类方法的功能性 miRNA-mRNA 调控模块挖掘

为了验证 Bimax 双聚类方法的有效性和可靠性，本设计分别基于 miRTarBase 数据库和 TargetScan 数据库获得验证型和预测型的两组数据集进行实验。

在获得验证型和预测型的 miRNA-mRNA 相互作用之后，我们评估由这些相互作用组成的网络是否是无标度的生物网络。我们使用幂率模型 ($y = ax^b$) 来拟合网络的节点度分布。如图 2、3 所示，该网络的节点度分布很好地服从幂率

分布, $R^2 = 0.9797$ 、 $R^2 = 0.9430$ 。因此, 两种数据库所形成的 miRNA-mRNA 调控网络是无标度网络。

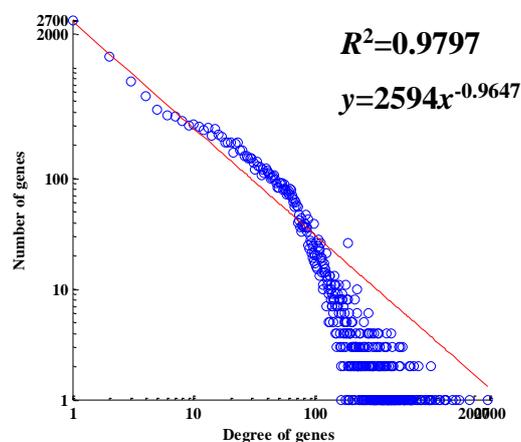


图 2 验证型调控网络的节点度分布图

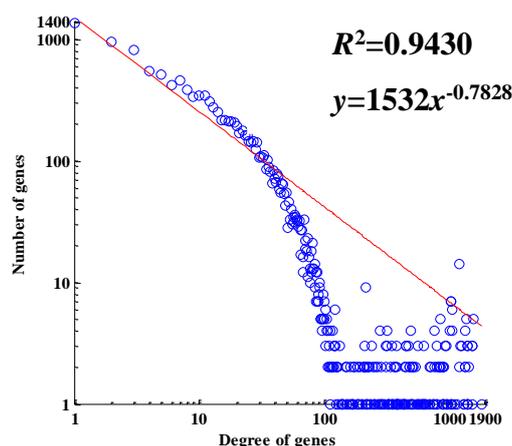


图 3 预测型调控网络的节点度分布图

我们分别将来自 miRTarBase 验证型的及 TargetScan 预测型的 miRNA-mRNA 调控网络转换为二进制矩阵, 用 R 工具包 biclust 中的 Bimax 方法来发现功能性调控模块。得到的模块表如下表 3、表 4 所示, 表中列出了 20 个聚类结果中有意义的模块, 及每个模块中 miRNA、mRNA、重要 GO 项 (Significant GO terms) 的数目, 重要 GO 项是指这个模块与多少生物过程相关。我们设置每个双聚类的 miRNA 和 mRNA 的最小数量分别为 5 和 10, 双聚类的数量设置为 20。

由表 3 可以看出: 大约一半的模块与几种生物过程功能相关, 这些模块被认为是功能性 miRNA-mRNA 调节模块。而表 4 的 20 个聚类结果中则只有 6 个模块具有显著性意义, 对比表 3、4, 说明对于不同的数据库, 使用 Bimax 算法挖

掘得到的调控模块的数目也不同。

表 3 验证型数据集上有意义的模块

Module ID	#miRNAs	#mRNAs	#Significant GO terms
2	5	41	1
5	5	15	29
6	5	19	113
8	5	16	2
12	5	12	6
13	6	11	6
18	5	22	28
19	5	20	2
20	5	19	2

表 4 预测型数据集上有意义的模块

Module ID	#miRNAs	#mRNAs	#Significant GO terms
1	5	841	2
3	6	112	3
5	6	22	4
6	7	122	8
8	7	138	2
12	8	68	2

2.3 本章小结

为了深入了解 miRNA 生物作用过程，探究其与靶基因之间的相互作用关系意义重大。本章中，我们在获得验证型和预测型的 miRNA-mRNA 相互作用之后，通过使用幂率模型（ $y = ax^b$ ）拟合网络的节点度分布，以此来评估由这些相互作用组成的网络是否是无标度的生物网络。结果表明，两种数据库所形成的 miRNA-mRNA 调控网络是无标度网络，节点度分布很好地服从幂率分布。目前已有许多算法用于识别调控模块，我们采用了 Bimax 双聚类算法从行和列两个方向，识别 mRNA 的子集在 miRNA 的子集上表现出相似的行为的双簇，并通过重要 GO 项指标来判断模块与多少生物过程相关，从而识别出功能性调控模块。

第三章 miRNA-mRNA 调控模块功能验证

3.1 概述

功能富集分析是一种识别在大量基因集或蛋白质集中过度表达的基因或蛋白质类别的方法,并且可能与疾病表型有关联。该分析侧重于基因组表达的变化,通过这种方法解决了单基因表达不可检测、变化小的问题^[40]。

本文将从 GO^[41]生物过程和 KEGG^[42]信号通道两个方面对挖掘得到的功能性调控模块进行功能富集分析。

3.2 富集分析结果与分析

为了揭示潜在的生物过程或途径,我们使用 R 工具包 clusterProfiler 对每个模块进行富集分析,校正 p 值截止值设置为 0.05。我们选择重要 GO 项数目最大的模块在 Cytoscape 软件中进行可视化,结果如下图 4、图 5 所示。图 4 是从表 3 中选取重要 GO 项数目最大的模块(即模块 6)的网络表示,红色节点表示 miRNA,白色节点是 mRNA。由图 4 可以看出:网络包含 5 个 miRNA, 19 个 mRNA 和 95 个边缘,是一个完整的二分网络。由于基于 TargetScan 数据库的模块 6 在 Cytoscape 软件中进行可视化的结果过大,不宜展示,因此选择稍小的模块(模块 5)来构建子网络,结果如下图 5 所示。由图 5 可以看出:网络包含 6 个 miRNA, 22 个 mRNA 和 132 个边缘,也是一个完整的二分网络。验证二分网络的依据:边缘数=miRNA 个数*mRNA 个数。

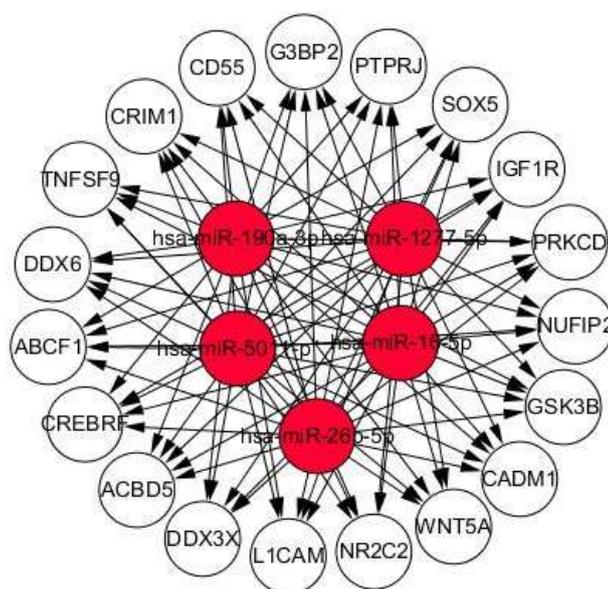


图 4 验证型调控模块网络图示例

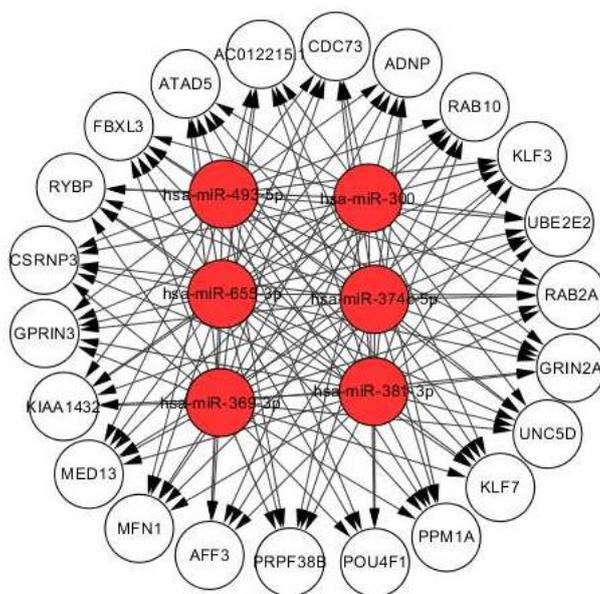


图 5 预测型调控模块网络图示例

两组数据集 GO 生物过程和 KEGG 信号通道功能富集分析的可视化结果如下图 6、7、8 所示：

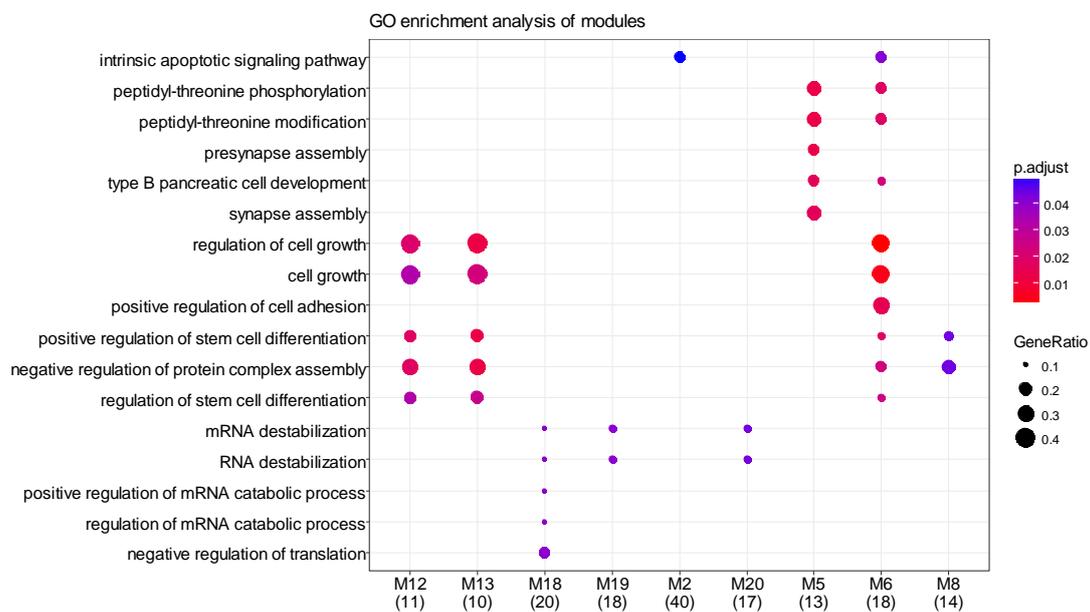


图 6 验证型调控模块的 GO 生物过程富集分析

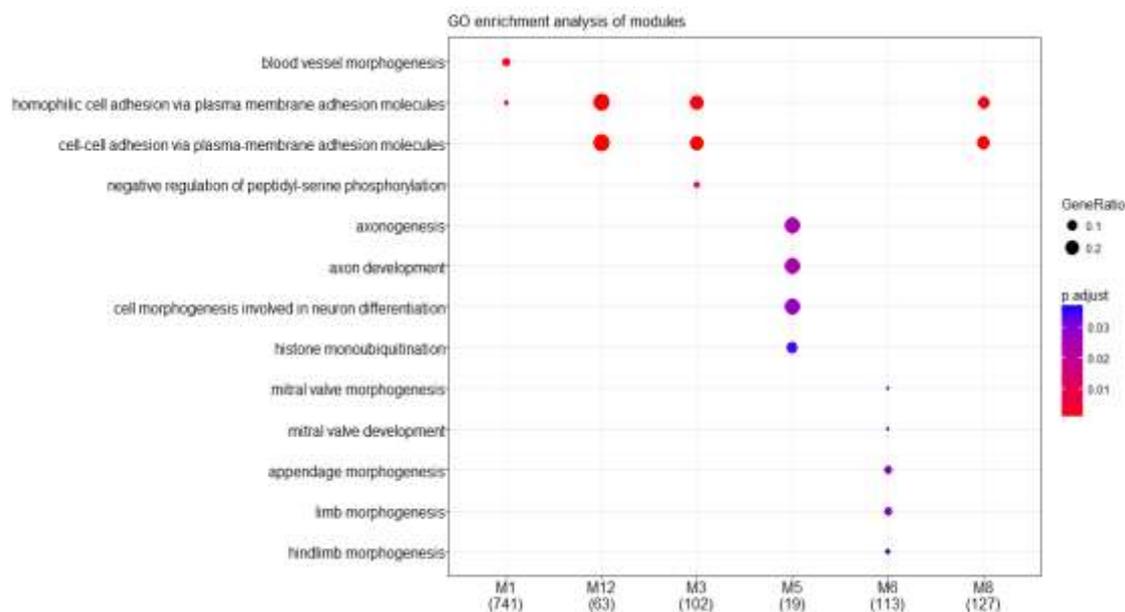


图 7 预测型调控模块的 GO 生物过程富集分析

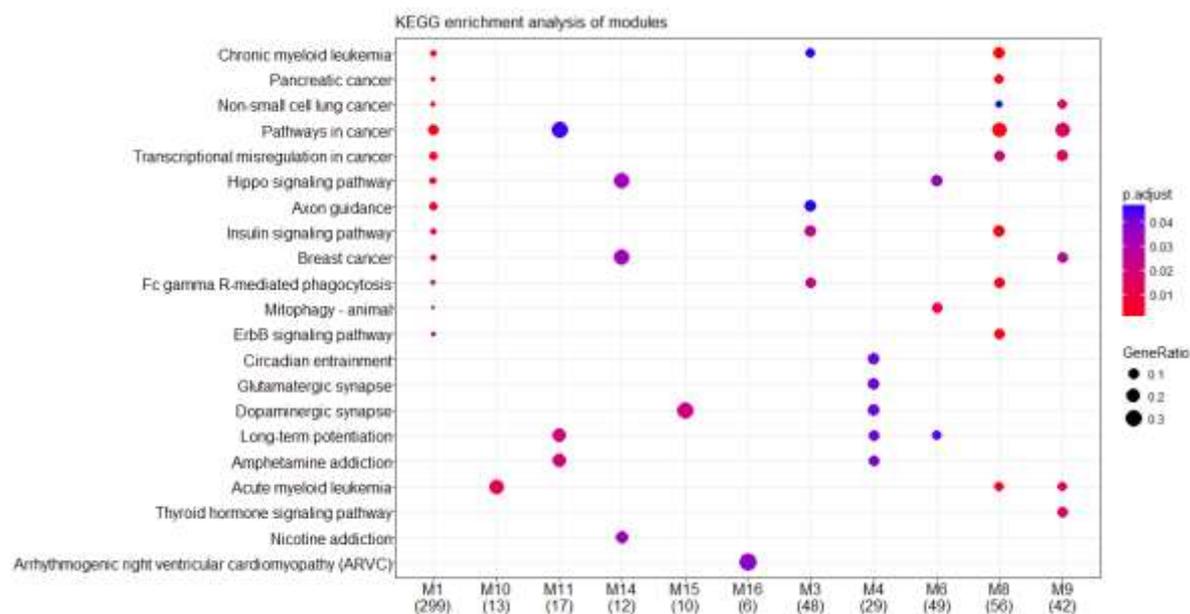


图 8 预测型调控模块的 KEGG 信号通道富集分析

图中横坐标是表示显著富集的模块，纵坐标则列出了显著富集的生物学过程。clusterProfiler 确定了与不同生物过程相关的不同模块，如图 6、7、8 中所示，有显著富集结果的模块分别有 9 个、6 个、11 个。以图 8 为例，与多巴胺能神经突触过程相关的有模块 4 和模块 15。

图中 p 值表明哪些类别更有可能具有生物学功能意义。图中的点是基于它们相应的 p 值进行颜色编码的，从红色到蓝色的颜色渐变对应于 p 值增加的顺序。

也就是说，红色表示低 p 值（高富集度），蓝色表示高 p 值（低富集度）；点的大小代表模块与生物过程相关的基因占比。由于基于 miRTarBase 数据库的 KEGG 信号通道分析没有富集结果，因此没有相应的图形展示。

3.3 本章小结

在上一章中我们通过 Bimax 双聚类算法发现了功能性调控模块，为了验证这些调控模块的生物功能，在本章中采用功能富集分析方法来挖掘该模块背后的生物学意义。为了揭示潜在的生物过程或途径，我们使用 R 工具包 clusterProfiler 对每个模块进行富集分析，核心思想都是将基因与生物学注释联系起来，站在统计学角度，从关联的生物学注释中标识出比较显著的生物通路或信息。其次，利用 Cytoscape 软件将重要 GO 项数目最大的模块进行可视化。基于上述两种数据库富集分析的结果，从参数中可以看出功能富集从更高的层次比如细胞水平来解释生物实验中的现象，进而分析功能性调控模块与疾病之间的关系，以此达到本设计的核心宗旨和目的。

第四章 总结与展望

4.1 总结

研究内容分为三部分：（1）获取 miRNA-mRNA 调控网络数据源，并对该数据进行预处理：进行清洗，去除重复项；（2）功能性调控模块挖掘；（3）功能性调控模块 GO 生物过程和 KEGG 信号通道分析。

本文解决了 miRNA-mRNA 调控网络数据清洗问题、功能性调控模块的有效识别及功能验证问题。在实验过程中，基于验证型及预测型的 miRNA-mRNA 调控网络数据，采用当前应用较广的 Bimax 双聚类算法，借助 Cytoscape 软件将重要 GO 项数目最大的模块进行可视化，运用 R 的 clusterProfiler 包对挖掘得到功能性调控模块进行 GO 生物过程和 KEGG 信号通道功能富集分析。

在本文中，通过使用幂律率模型（ $y = ax^b$ ）来拟合网络的节点度分布，结果表明该网络的节点度分布很好地服从幂率分布。由表 3 发现大约一半的模块与几种生物过程相关，这些模块被认为是功能性调控模块。因此，可以认为 Bimax 双聚类算法是一个有前途的方法。

4.2 展望

本文使用 Bimax 双聚类算法来挖掘功能性 miRNA-mRNA 调控模块挖掘，通过实验我们发现 Bimax 双聚类算法的优点在于将原始数据矩阵化，减少了算法的空间，从而节省时间提高效率。但其缺点在于它仅适用于二进制矩阵，因此，在使用 Bimax 之前，应首先将其他类型的矩阵转换为二进制矩阵。将数据转换成二进制矩阵前，要先将数据离散成 0 和 1，离散过程中会造成数据失真影响聚类的质量。当前双聚类算法还存在局限性，如何克服这种局限且使该算法聚类效果达到最优是当前研究课题中面临的挑战之一。在将来，我们会进一步扩大 miRNA-mRNA 相互作用，这将有助于挖掘更有意义的模块。

参 考 文 献

- [1] Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function[J]. *Cell*, 2004, 116 (2): 281-297.
- [2] Ambros V. MicroRNA pathways in flies and worms: growth, death, fat, stress, and timing[J]. *Cell*, 2003, 113 (6): 673-676.
- [3] Ambros V. microRNAs: tiny regulators with great potential[J]. *Cell*, 2001, 107 (7) : 823-826.
- [4] Zhang B, Wang Q, Pan X. MicroRNAs and their regulatory roles in animals and plants [J]. *Cell Physiol.*, 2007, 210 (2) : 279-289.
- [5] 李霞, 李亦学, 廖飞. 生物信息学[M]. 人民卫生出版社 (LI Xia, LI Yi xue, LIAO Fei. *Bioinformatics*, 2010, 48.
- [6] 王正华, 刘齐军, 朱云平. 基因调控网络的模块化组织研究[J]. *遗传*, 2008, 30(1): 20-27.
- [7] Joung JG, Hwang KB, Nam JW, et al. Discovery of microRNA-mRNA modules via population-based probabilistic learning[J]. *Bioinformatics*, 2007, 23(9): 1141-1147.
- [8] Karim SMM, Liu L, Le TD, et al. Identification of miRNA-mRNA regulatory modules by exploring collective group relationships[J]. *BMC Genomics*, 2016, 17(1): 71.
- [9] Lionetti M, Biasiolo M, Agnelli L, et al. Identification of microRNA expression patterns and definition of a microRNA/mRNA regulatory network in distinct molecular groups of multiple myeloma[J]. *Blood*, 2009, 114(25): e20-e26.
- [10] Liu B, Liu L, Tsykin A, et al. Identifying functional miRNA-mRNA regulatory modules with correspondence latent dirichlet allocation[J]. *Bioinformatics*, 2010, 26(24): 3105-3111.
- [11] Lu Y, Zhou Y, Qu W, et al. A Lasso regression model for the construction of microRNA-target regulatory networks[J]. *Bioinformatics*, 2011, 27(17): 2406-2413.
- [12] Zhang B, Pan X, Wang Q. Computational identification of microRNAs and their targets[J]. *Birth Defects Research Part C Embryo Today Reviews*, 2006, 30(6):395-407.
- [13] Griffiths-Jones S, Saini HK, van Dongen S, et al. miRBase: tools for microRNA genomics[J]. *Nucleic Acids Research*, 2008, 36(suppl 1): D154-D158.
- [14] Enright AJ, John B, Gaul U, et al. MicroRNA targets in *Drosophila*[J]. *Genome*

Biology, 2003, 5(1): R1.

[15] Yang JH, Li JH, Shao P, et al. starBase: a database for exploring microRNA–mRNA interaction maps from Argonaute CLIP-Seq and Degradome-Seq data[J]. *Nucleic Acids Research*, 2011, 39(suppl 1): D202-D209.

[16] Hsu SD, Tseng YT, Shrestha S, et al. miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. *Nucleic Acids Res.*, 2014, 42(D1): D78-D85.

[17] Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets[J]. *Cell*, 2005, 120(1): 15-20.

[18] Sethupathy P, Corda B, Hatzigeorgiou AG. TarBase: A comprehensive database of experimentally supported animal microRNA targets[J]. *Rna*, 2006, 12(2): 192-197.

[19] Krek A, Grün D, Poy MN, et al. Combinatorial microRNA target predictions[J]. *Nature Genetics*, 2005, 37(5): 495-500.

[20] Betel D, Wilson M, Gabow A, et al. The microRNA. org resource: targets and expression[J]. *Nucleic Acids Research*, 2008, 36(suppl 1): D149-D153.

[21] Xiao F, Zuo Z, Cai G, et al. miRecords: an integrated resource for microRNA–target interactions[J]. *Nucleic Acids Research*, 2009, 37(suppl 1): D105-D110.

[22] Kertesz M, Iovino N, Unnerstall U, et al. The role of site accessibility in microRNA target recognition[J]. *Nature Genetics*, 2007, 39(10): 1278-1284.

[23] Hsu PW, Huang HD, Hsu SD, et al. miRNAMap: genomic maps of microRNA genes and their target genes in mammalian genomes[J]. *Nucleic Acids Research*, 2006, 34(suppl 1): D135-D139.

[24] Alexiou P, Vergoulis T, Gleditsch M, et al. miRGen 2.0: a database of microRNA genomic information and regulation[J]. *Nucleic Acids Research*, 2009: gkp888.

[25] Liu S, Li JH, Wu J, et al. StarScan: a web server for scanning small RNA targets from degradome sequencing data[J]. *Nucleic Acids Research*, 2015, 43(1):480-6.

[26] Zhang Z, Yu J, Li D, et al. PMRD: plant microRNA database[J]. *Nucleic Acids Research*, 2010, 38(suppl 1): D806-D813.

[27] Miranda KC, Huynh T, Tay Y, et al. A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes[J]. *Cell*, 2006, 126(6): 1203-1217.

- [28] 张俊鹏. 基于概率话题模型的功能性 miRNA-mRNA 调控模块识别[D]. 昆明理工大学, 2012.
- [29] Pio G, Ceci M, D'Elia D, et al. A novel biclustering algorithm for the discovery of meaningful biological correlations between microRNAs and their target genes[J]. BMC Bioinformatics, 2013, 14(7): S8.
- [30] Tran DH, Satou K, Ho TB. Finding microRNA regulatory modules in human genome using rule induction[J]. BMC Bioinformatics, 2008, 9(12): S5.
- [31] Friedman N, Linial M, Nachman I, et al. Using Bayesian networks to analyze expression data[J]. Journal of computational biology, 2000, 7(3-4): 601-620.
- [32] Peng X, Li Y, Walters KA, et al. Computational identification of hepatitis C virus associated microRNA-mRNA regulatory modules in human livers[J]. BMC Genomics, 2009, 10(1): 373
- [33] Steyvers M, Smyth P, Rosen-Zvi M, et al. Probabilistic author-topic models for information discovery[C]//Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2004: 306-315.
- [34] Liu B, Liu L, Tsykin A, et al. Identifying functional miRNA - mRNA regulatory modules with correspondence latent dirichlet allocation[J]. Bioinformatics, 2010, 26(24): 3105-3111.
- [35] Li Y, Liang C, Wong KC, et al. Mirsynergy: detecting synergistic miRNA regulatory modules by overlapping neighbourhood expansion[J]. Bioinformatics, 2014: btu373.
- [36] Liang C, Li Y, Luo J. A novel method to detect functional microRNA regulatory modules by bicliques merging[J]. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2016, 13(3): 549-556.
- [37] 刘斌. 基于多源数据的 miRNA 调控模块识别算法研究[D]. 湖南大学, 2016.
- [38] Prelić A, Bleuler S, Zimmermann P, et al. A systematic comparison and evaluation of biclustering methods for gene expression data. Bioinformatics, 2006, 22(9): 1122-1129.
- [39] Eren K, Deveci M, Küçüktunç O, et al. A comparative analysis of biclustering algorithms for gene expression data[J]. Briefings in bioinformatics, 2013, 14(3): 279-292.
- [40] Mootha VK, Lindgren CM, Eriksson KF, et al. PGC-1 α -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes[J]. Nature Genetics, 2003, 34(3): 267-273.
- [41] Ashburner M, Ball CA, Blake JA, et al. Gene Ontology: tool for the unification of biology[J]. Nature Genetics, 2000, 25(1): 25-29.

[42] Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes[J]. Nucleic Acids Research, 2000, 28(1): 27-30.

致 谢

在大理大学度过了四年的紧张学习时光，这段时光使我们受益匪浅。在论文即将完成之际，我们要向所有给予我们支持、帮助和鼓励的人表示我们最诚挚的谢意。

首先，我们感谢我们的导师张俊鹏老师。在论文撰写的整个过程中，从论文选题、到撰写开题报告、最后到正文撰写，张老师都提出了很多宝贵意见。无论从格式规范、论文要点、还是文章结构，张老师都不厌其烦，给予我们悉心指导，使我们能够最后顺利完成论文写作工作。张老师对工作的认真负责、对学术的钻研精神和严谨的学风，都是值得我们终生学习的。

其次，我们也要感谢工程学院的各专业老师们，在他们严格无私的指导下，我们系统地学习了生物医学工程专业的各方面知识，还学习到很多做人、做事的道理。于此同时，我们还要感谢一直以来为我们无私奉献的父母以及四年来给我们热心提供帮助的各位同学。

最后，我们会在今后的工作、学习中加倍努力，心怀感恩，用自己的成绩来回报社会，为自己的母校增添光彩。