



新生物学丛书

新生物学年鉴

2015

《新生物学年鉴2015》编委会 主编

- ▶ microRNA分子标志物研究
- ▶ 细胞内胆固醇运输机制
- ▶ 慢性代谢性疾病人群研究
- ▶ 植物发育单位与发育程序
- ▶ 植食性昆虫对植物的适应
- ▶ 转基因作物的生态风险评价



科学出版社

新生物学丛书

新生物学年鉴 2015

《新生物学年鉴 2015》编委会 主编

科学出版社

北 京

内 容 简 介

《新生物学年鉴 2015》为一本由《新生物学丛书》4 位编委组稿并把关、各地大学及研究所骨干研究人员撰写, 本书涉及动物学、植物学、医学、生态学等 6 大领域, 包括 microRNA 生物标志物研究、细胞内胆固醇运输、代谢性疾病研究、植物发育研究、昆虫生物防御研究, 以及转基因作物生态风险研究等内容, 共 6 篇文章。这些文章的内容均为撰写者的最新研究成果, 因此年鉴可以在一定程度上体现出我国生物学领域的发展现状。

本书适合各相关领域的高年级本科生、研究生和专业研究人员学习参考。对于希望了解生物学发展现状的科研爱好者, 本书也可作为很好的阅读材料。

图书在版编目(CIP)数据

新生物学年鉴. 2015/《新生物学年鉴 2015》编委会主编. —北京: 科学出版社, 2016.2

(新生物学丛书)

ISBN 978-7-03-047224-3

I. ①新… II. ①新… III. ①生物学-文集 IV. ①Q-53

中国版本图书馆 CIP 数据核字(2016)第 013597 号

责任编辑: 王 静 岳漫宇 / 责任校对: 陈玉凤

责任印制: 肖 兴 / 封面设计: 刘新新

科 学 出 版 社 出 版

北京东黄城根北街 16 号

邮政编码: 100717

<http://www.sciencep.com>

北京通州皇家印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2016 年 2 月第 一 版 开本: 787×1092 1/16

2016 年 2 月第一次印刷 印张: 11

字数: 243 000

定价: 98.00 元

(如有印装质量问题, 我社负责调换)

《新生物学丛书》专家委员会

主 任：蒲慕明

副 主 任：吴家睿

专家委员会成员 (按姓氏汉语拼音排序)

昌增益	陈洛南	陈晔光	邓兴旺	高 福
韩忠朝	贺福初	黄大昉	蒋华良	金 力
康 乐	李家洋	林其谁	马克平	孟安明
裴 钢	饶 毅	饶子和	施一公	舒红兵
王 琛	王梅祥	王小宁	吴仲义	徐安龙
许智宏	薛红卫	詹启敏	张先恩	赵国屏
赵立平	钟 扬	周 琪	周忠和	朱 祯

《新生物学年鉴 2015》编委会

主 编：蒲慕明

副 主 编：吴家睿

编 委 (按姓氏汉语拼音排序)

龙漫远 美国芝加哥大学

马克平 中国科学院植物研究所

吴家睿 中国科学院上海生命科学研究院

曾任森 福建农林大学生命科学学院

丛 书 序

当前，一场新的生物学革命正在展开。为此，美国国家科学院研究理事会于 2009 年发布了一份战略研究报告，提出一个“新生物学”(New Biology)时代即将来临。这个“新生物学”，一方面是生物学内部各种分支学科的重组与融合，另一方面是化学、物理、信息科学、材料科学等众多非生命学科与生物学的紧密交叉与整合。

在这样一个全球生命科学发展变革的时代，我国的生命科学研究也正在高速发展，并进入了一个充满机遇和挑战的黄金期。在这个时期，将会产生许多具有影响力、推动力的科研成果。因此，有必要通过系统性集成和出版相关主题的国内外优秀图书，为后人留下一笔宝贵的“新生物学”时代精神财富。

科学出版社联合国内一批有志于推进生命科学发展的专家与学者，联合打造了一个 21 世纪中国生命科学的传播平台——《新生物学丛书》。希望通过这套丛书的出版，记录生命科学的进步，传递对生物技术发展的梦想。

《新生物学丛书》下设三个子系列：科学风向标，着重收集科学发展战略和态势分析报告，为科学管理者和科研人员展示科学的最新动向；科学百家园，重点收录国内外专家与学者的科研专著，为专业工作者提供新思想和新方法；科学新视窗，主要发表高级科普著作，为不同领域的研究人员和科学爱好者普及生命科学的前沿知识。

如果说科学出版社是一个“支点”，这套丛书就像一根“杠杆”，那么读者就能够借助这根“杠杆”成为撬动“地球”的人。编委会相信，不同类型的读者都能够从这套丛书中得到新的知识信息，获得思考与启迪。

《新生物学丛书》专家委员会

主 任：蒲慕明

副主任：吴家睿

2012 年 3 月

前 言

当前,生命科学正处于革命性变化的前夜,“新生物学时代”已经到来。创立《新生物学丛书》的一个主要目的,就是要及时总结当下最新的科研成果,展望未来的发展方向。为此,《新生物学丛书》编委会决定打造中国的生命科学年度报告——《新生物学年鉴》(以下简称《年鉴》)。《年鉴》以综述性文集的形式出版,系统总结国内外生命科学的研究成果,重点追踪和评述近年来生命科学各个领域的研究热点和前沿进展。《年鉴》的编委每年从《新生物学丛书》专家委员会中产生,负责撰写或组织相关领域的专家学者进行评述。

《年鉴》每年出版一期,涉及的研究领域都是当今生命科学最热点的前沿领域,参与文章撰写的所有人员均是活跃在科研第一线的权威专家学者,每篇文章都可以很好地体现出相关领域的前沿进展。我们希望能够将《年鉴》打造成与美国 *Annual Review* 系列一样高水平的中文版综述平台,及时地报道生命科学的成果和进步。

第一期《年鉴 2012》于 2013 年 1 月顺利出版,涉及了生物学的 7 个领域,共 12 篇文章。《年鉴 2013》于 2013 年 12 月出版,扩充了涉及的生物学子学科,涉及 9 个领域,共 10 篇文章。《年鉴 2014》于 2015 年 1 月出版,涉及生物学与生态学 6 大领域,共 6 篇文章。

在以上三本年鉴的基础上,今年的《年鉴 2015》也邀请了《新生物学丛书》专家组的 4 位专家组织并撰写。本书涉及动物学、植物学、医学、生态学等 6 大领域,包括 microRNA 生物标志物研究、细胞内胆固醇运输、代谢性疾病研究、植物发育研究、昆虫生物防御研究,以及转基因作物生态风险研究等内容,共 6 篇文章。通过这些精彩的综述性文章,可以了解到该学科的前沿发展、国际动向,开阔了视野,更为中国科学研究人员在该领域中的发展树立了信心。

本期年鉴从动物到植物、从微观到宏观,全面地介绍了新生物学的多个热点学科的最新动向。适合各相关领域的高年级本科生、研究生,以及专业研究人员学习参考。对于希望了解生物学发展现状和动态的科研爱好者,本书也是一本很好的阅读材料。

《新生物学年鉴 2015》编委会

2016 年 1 月

目 录

microRNA 生物标志物的生物信息学研究进展	1
林宇鑫 孙占东 张文字 严文颖 陈佳佳 沈百荣	
细胞内胆固醇运输机制概述	34
肖 健 宋保亮	
慢性代谢性疾病人群研究的进展与发展趋势	49
林 旭 黎怀星 孙 亮 刘 刚 胡 瑶 王飞杰 马一玮	
量体裁新衣：从植物发育单位到植物发育程序	73
白书农	
植食性昆虫适应植物蛋白酶抑制剂的机制	117
曾任森 孙仲享 宋圆圆 卢 凯 朱克岩	
转基因作物的生态风险评价研究进展	137
魏 伟 关正君 马克平	

* 策划编辑：龙漫远 美国芝加哥大学

microRNA 生物标志物的生物信息学 研究进展

作者：林宇鑫 孙占东 张文宇 严文颖 陈佳佳 沈百荣

苏州大学系统生物学研究中心

- ▶ 1. microRNA的功能与标志物的研究现状 /2
- ▶ 2. microRNA-mRNA调控网络的生物信息学研究 /7
- ▶ 3. microRNA标志物的网络特征与功能分析 /12
- ▶ 4. microRNA标志物的生物信息学模型及其应用 /18
- ▶ 5. microRNA与ceRNA /23
- ▶ 6. microRNA生物标志物研究展望 /25

摘要

microRNA 是一类非常重要的非编码 RNA, 主要存在于真核生物体中。由于高度的组织表达特异性及对基因表达的调控功能, microRNA 是疾病诊断和治疗的一类重要的生物标志物。本综述主要回顾了近年来 microRNA 生物标志物的生物信息学研究进展。重点从 microRNA-mRNA 调控网络的角度讨论作为标志物的 microRNA 具有的网络特征, 探索潜在 microRNA 标志物识别的有效方法, 并提出理论成果向临床实践转化可行性的建议。

关键词

microRNA、生物标志物、microRNA-mRNA 调控网络、网络特征、生物信息学、转化医学

microRNA (简称 miRNA) 是一类长约 22 个碱基的内源性非编码 RNA, 广泛存在于多种真核生物中。它们在转录后水平上调控基因表达, 在发育、分化, 细胞的增殖、凋亡和代谢等生物过程中起到关键作用。生物标志物(biomarker)是一种重要的生化指标, 用于标记生命体各种分子、细胞、组织、系统等发生或可能发生的改变, 在疾病的分期判别、临床诊断、药物的疗效评估等方面具有广泛用途。一般而言, 生物标志物具备一定的特异性, 同时需要保证较高的灵敏度。近年来, 许多研究表明 microRNA 在疾病和正常样本中具有显著的表达差异, 并且在不同疾病样本中的表达情况也具有特异性。这在一定程度上揭示了 microRNA 可以作为一种理想的生物标志物。在大数据时代, 越来越多来源于疾病本身及患者临床样本的信息和知识亟待挖掘。microRNA 生物标志物的研究将为疾病的早期诊断及预后监测提供有力支持。随着系统医学特别是转化医学概念的提出, 将基础研究和临床治疗相结合的创新思维模式促进了生物标志物研究的不断发展。本文主要介绍近年来 microRNA 生物标志物的生物信息学研究进展, 结合作者自身的科研实践, 重点论述 microRNA 标志物在 microRNA-mRNA 调控网络中具有的一般规律和显著特征, 为潜在 microRNA 标志物的识别建立合理的模型体系, 推动研究成果向临床应用的转化。

1. microRNA 的功能与标志物的研究现状

1.1 microRNA 的来源与功能

据估计, microRNA 调控了约 30% 人类基因组编码基因 (Lewis et al., 2005)。它们多数位于基因间隔区, 可作为独立的转录单元进行表达。同时, 也有一些位于基因的内含子或外显子上, 能够与基因同步表达。microRNA 的生物合成由 RNA 聚合酶 II 启动。RNA

聚合酶 II 首先合成长约数千碱基的初始转录本 (pri-microRNA)，经细胞核内的 DGCR8/Drosha 蛋白复合体加工，产生长约 70 个核苷酸、具有特殊“发夹”结构的双链前体分子 (pre-microRNA)。pre-microRNA 由转运蛋白 Exportin-5 运输到细胞质，与 RNAase III 酶 Dicer 结合后被切割成长约 22 个碱基对的双链 RNA 分子。在解旋酶作用下双链解旋，其活性引导链加载到 RNA 诱导的沉默复合体 (RNA-induced silencing complex, RISC)，而非活性链则被移除并降解。microRNA-RISC 复合体可以识别并结合 mRNA 的互补序列，通过诱导 mRNA 降解或抑制 mRNA 的翻译来抑制基因的表达。

最近也有研究表明，microRNA 不仅能抑制翻译，也能增强翻译，其抑制还是增强作用取决于细胞周期所处的状态 (Vasudevan et al., 2007)。

1.2 microRNA 作为疾病标志物

生物标志物是一类能够反映疾病的发展阶段或状态的特征性生物指标 (Ghosh and Poisson, 2009)。高度灵敏和特异的生物标志物可以用于疾病的早期诊断和预后判断，还能作为潜在的治疗靶标，从而有效减少疾病的死亡率，为临床提供个性化靶向治疗的新策略。目前临床应用的生物标志物大多是血液中的特定蛋白质，如肌钙蛋白 (心血管疾病) (Ferreira et al., 2014)、癌胚抗原 (多种癌症) (Tiernan et al., 2013)、前列腺特异性抗原 (前列腺癌) (Nakayama et al., 2014) 和氨基转移酶 (肝功能) (Liu et al., 2014) 等。近来研究表明，microRNA 在病毒感染、心血管疾病、神经和肌肉系统疾病等多种疾病的发生发展中发挥作用 (Chen et al., 2014; Wang et al., 2014)。迄今发现的近半数 microRNA 位于染色体的脆性位点、杂合型丢失区、扩增区或断裂位点，提示 microRNA 在肿瘤中也扮演了重要角色 (Huang et al., 2013)。

针对 microRNA 和疾病的相关性，已有大量研究采用 microRNA 克隆、定量 PCR，以及近来兴起的微球流式细胞、基因芯片和深度测序等高通量组学方法研究了 microRNA 在疾病中的整体表达，即 microRNA 组 (miRNome)，并已鉴定了众多在疾病中表达异常的 microRNA (Ozen et al., 2008; Wang et al., 2008; Lodes et al., 2009; Sun et al., 2009; Schaefer et al., 2010; Sorensen and Orntoft, 2010)。

microRNA 具有理想生物标志物的许多必要特征。

1) microRNA 分子比 mRNA 更小也更稳定，不易被核酸内切酶降解。从内脏组织、血液和其他生物体液 (如尿、唾液和痰) 中都能获得稳定的表达谱 (Dang et al., 2015)。

2) microRNA 能够完好保存于福尔马林固定石蜡包埋 (FFPE) 样品中并完整提取，从而有利于存档临床样本的准确分析 (Ferracin et al., 2011)。

3) microRNA 具有高度的组织表达特异性和时序表达特异性，因此能够可靠地反映疾病的进化和分化谱系 (Abrahamsson and Dabrosin, 2015)。

基于 microRNA 的生物标志物已成为当前的研究热点之一 (Garzon et al., 2009; Lee and Dutta, 2009)。我们在 PubMed 数据库中用关键词 “microRNA* [ti] AND *marker* [tiab]” 进行检索，共找到 1507 篇相关论文。其中，近五年来发表的共计 1318 篇。microRNA 在疾病的诊断、分类、预后、治疗效果评估等方面的潜力已经被越来越多的研究中所证

实,它们在改善临床诊断和后续治疗方面有着广阔的应用前景。

1.2.1 microRNA 作为诊断标志物

寻找疾病特别是像癌症这样复杂疾病的早期诊断标志物在临床上具有重要意义。由于患者的预后情况和存活时间在很大程度上取决于检出时所处的阶段,所以早期的发现可显著改善预后效果。近年来,多项研究报道了血清、血浆、尿液等体液循环 microRNA 能够准确区分正常样本和疾病样本,可以作为敏感特异的疾病诊断标志物。例如,在胰腺导管腺癌中 miR-205 和 miR-21 均过量表达 (du Rieu et al., 2010),且表达水平的变化先于导管中的表型变化,有望用于该疾病的早期检测。基于血液 microRNA 的诊断标志物 miR-642b、miR-885-5p 和 miR-22 可用于胰腺癌的早期检测 (Ganepola et al., 2014)。一组包含 4 个 microRNA 分子 (miR-19a-3p、miR-223-3p、miR-92a-3p 和 miR-422a) 的标记能准确诊断结直肠腺癌 (Zheng et al., 2014)。miR-141 血清水平可用于鉴别晚期前列腺癌患者和健康人 (Mitchell et al., 2008)。miR-126 和 miR-182 在尿液样本中的比例可用于检测膀胱癌 (Hanke et al., 2010)。唾液中 miR-125a 和 miR-200a 水平的降低与口腔鳞状细胞癌有关 (Park et al., 2009)。除此之外,组织特异性的 microRNA 表达水平还可以监测特定器官的病理生理条件,如急性组织损伤。例如,miR-122 的血浆水平可以精确地定量检测对乙酰氨基酚引起的肝损伤 (Wang et al., 2009)。由于 miR-122 是一种仅在肝脏中表达的肝脏特异性 microRNA,因此基于 microRNA 的方法检测和监测药物性肝损伤比血清 ALT 检测法更为敏感和可靠。除了肝损伤外, microRNA 还可用于检测心肌损伤。在心脏特异表达的 miR-499 的血浆水平和心肌梗死患者的血液肌钙蛋白水平相关 (Adachi et al., 2010)。同时, miR-1 (Ai et al., 2010) 和 miR-208 (Ji et al., 2009) 也已被证明是监测心肌损伤的敏感特异标志物。

1.2.2 microRNA 作为分类标志物

microRNA 的高度组织特异性使其在癌症分型和组织来源鉴定这两方面独具优势。利用它能准确区分不同类型的癌症亚型、识别低分化肿瘤的起源组织。

(1) 亚型分类

确定疾病的不同亚型是临床面临的重要问题。由于疾病不同亚型的发病分子机制不同,因此它们的临床表现、预后及对药物的反应也不尽相同。区别对待不同的亚型患者并给予个体化的靶向性治疗能够有效改善预后并提高存活率。microRNA 在各种癌症亚型中具有差异表达现象。例如, microRNA 分子在基底型乳腺癌和管腔型乳腺癌这两个亚型之间存在差异表达 (Blenkiron et al., 2007; Sempere et al., 2007)。miR-200 家族与管腔亚型相关,直接靶向 EMT 调节因子 ZEB1 和 ZEB2 (Gregory et al., 2008)。miR-342 表达水平在管腔 B 型乳腺癌中上升,且可以特异区分雌激素受体 (ER)、孕酮受体 (PR) 和原癌基因 HER2/neu 受体的表达状态 (Lowery et al., 2009)。而 miR-145、miR-205 和 miR-342 在基底细胞样三阴性乳腺癌 (ER⁻/PR⁻/HER2⁻) 中显著降低, miR-520g 在 ER⁻/PR⁻型乳腺癌中表达水平下调 (Gregory et al., 2008; Lowery et al., 2009)。Toffanin 等 (2011) 利用

microRNA 表达谱对丙型肝炎病毒相关的肝细胞癌进行分子分类,发现了肝癌 3 个主要的分子亚型: Wnt 信号相关、干扰素相关和增殖相关亚型。其中增殖相关亚型过度表达位于 19 号染色体 q13.41 的 miR-517a 和 miR-520c, 这些 microRNA 的表达促进肝癌细胞在体外增殖、迁移和侵袭。此外, miR-517a 还能促进体内肿瘤的生长和转移传播。

(2) 肿瘤组织来源识别

原发灶不明癌 (carcinoma of unknown primary, CUP) 指的是经组织学证实、但无法确定其原发部位的一类转移性恶性肿瘤。原发灶不明癌在临床诊断中较为常见, 约占新诊断癌症的 3%~5%。原发灶不明癌细胞早期即可发生转移且预后通常较差 (Oien and Evans, 2008), 是一个棘手的临床问题。目前, 临床上主要依赖成像技术结合免疫组化标志物对原发灶不明癌的肿瘤来源进行鉴定, 但仍有很大比例的原发灶部位无法明确, 难以系统性治疗, 通常只能选择广谱的化疗药物。然而广谱化疗的疗效并不理想, 无法明显改善患者的长期预后。只有当原发灶明确后才能根据原发部位的具体情况和特点进行位点特异性治疗。因此, 准确判断并掌握原发灶信息具有十分重要的临床价值。

近年来, 众多研究工作致力于寻找新型特异的原发灶诊断标志物。microRNA 的高度组织特异性使其成为鉴定肿瘤组织来源的新型分子诊断工具。这是因为当肿瘤扩散到多个转移灶的时候, 肿瘤自身会维持一个独特的、具有组织特异性的 microRNA 表达谱。例如, Rosenfeld 等 (2008) 分析了 336 个原发性和转移性肿瘤样本后获得一个包含 48 个 microRNA 的分类标志物, 能准确预测肿瘤的组织来源。Ferracin 等 (2011) 利用 microRNA 芯片分析了 101 个福尔马林固定石蜡包埋 (FFPE) 的原发癌和转移癌样品的 microRNA 表达, 鉴定了一个基于 47 个 microRNA 的分类器, 能够高度灵敏和特异地预测原发灶部位。上述研究表明, microRNA 的表达谱检测能够辅助现有的成像和免疫组化检测技术, 大大提高转移肿瘤组织来源的诊断精度。

1.2.3 microRNA 作为预后标志物

除了用于疾病的诊断和分类, microRNA 同样也可用于判断疾病的预后情况和预测疾病对治疗的反应。

microRNA 的预后功能首次在慢性淋巴细胞白血病中得到证明。2005 年, Calin 等报道了 microRNA 表达谱能够区分慢性淋巴细胞白血病患者的正常和恶性 B 淋巴细胞, 首次提出 microRNA 的表达标签和慢性淋巴细胞白血病的发展和预后有关。此后, 陆续有研究使用 microRNA 作为预后标志物来预测癌症的结果。例如, miR-26 的低水平表达可以作为肝细胞肝癌患者不良预后的独立预测因子 (Ji et al., 2009); 肺癌中 miR-155 的过表达和 let-7a 的下调与疾病预后不良相关 (Yanaihara et al., 2006)。Li 等 (2010) 报道了一个包含 7 个 microRNA 的表达标签, 可以预测胃癌患者总体生存率和无复发生存期。在黑色素瘤中, 较低的 miR-191 表达水平和较高的 miR-193a 表达水平往往预示着较短的 Kaplan-Meier 存活期 (Caramuta et al., 2010)。此外, 在各种癌症中 miR-21 的过度表达是预后不良的指标 (Rossi et al., 2010)。

近来研究发现 microRNA 参与调控肿瘤细胞对抗肿瘤药物的敏感性和耐药性, 能够

预测接受治疗的个体对于特定疗法的反应,从而区分特定疗法的受益人群,具有重大的临床价值。例如,低表达 miR-26 的肝细胞癌患者对于干扰素治疗的反应良好,存活率较高。因此,借助 miR-26 的表达水平能够筛选干扰素疗法的适用对象(Ji et al., 2009)。反之,一些 microRNA 的异常表达与肿瘤化疗药物的耐受性相关,而化疗耐药是现在临床肿瘤患者预后差的主要原因之一。Meng 等(2006)首次报道了 microRNA 在化疗药物抗性中的作用,并通过抑制 miR-21 和 miR-200b 的表达增加了胆管癌细胞系对吉西他滨的敏感性。Zhou 等(2010)则报道乳腺癌中 miR-125b 的高水平表达预示紫杉醇治疗反应不佳。Giovannetti 等(2010)发现 miR-21 高表达的胰腺癌患者在接受吉西他滨治疗时反应不佳。除了化疗以外, microRNA 还能影响患者对靶向疗法的反应。例如, miR-221 和 miR-222 的过表达导致多个和药物反应相关通路失调, miR-221/222 过表达可以作为雌激素受体靶向疗法的抗性标记。miR-221/222 通过负调控雌激素受体 α , 导致乳腺癌患者对它莫西芬的抗性(Zhao et al., 2008), 以及对 Fulvestran 的抗性(Rao et al., 2011)。而抑癌基因 miR-205 的异常表达能够直接靶向 HER3(Iorio et al., 2009), 从而提高对酪氨酸激酶抑制剂的响应速度。

microRNA 的异常表达和药物反应之间的相关性对揭示疾病耐药机制有重要意义。此外, 靶向耐药性相关 microRNA 有望成为克服疾病抗药性的有效手段。

1.2.4 microRNA 作为治疗靶点

microRNA 的异常表达与癌症间的相关性及 microRNA 功能分析表明, microRNA 在肿瘤发生中扮演了“抑癌基因”或“致癌基因”的双重角色, 调控 microRNA 的表达具有治疗癌症的可行性。鉴于在肿瘤发生中的双重角色, microRNA 在癌症治疗中既可以作为药物也可以作为药物靶标。

针对致癌 microRNA 的异常高表达, 一种方法是使用小分子药物调节 microRNA 的表达水平。例如, miR-21 是一个致癌 microRNA, 在多个实体瘤和血液系统恶性肿瘤中失调。Gumireddy 等(2008)用荧光素酶报告基因检测法筛选小分子文库, 寻找到可以抑制 miR-21 表达的化合物。经鉴定和体外测试证实, 该化合物可以抑制 miR-21 的表达, 诱导抗肿瘤反应。此后, 其他研究也利用荧光素酶法鉴定了可调节 microRNA 的表达的小分子药物(Young et al., 2010; Tripp and Young, 2014)。Young 等(2010)筛选小分子文库发现了调节 miR-122 表达的化合物。miR-122 表达富集于肝脏, 能够正调节丙型肝炎病毒的复制, 而小分子的 miR-122 抑制剂能够降低肝细胞中丙肝病毒复制。针对 microRNA 异常上调的另一种方法则是开发针对 microRNA 的反义寡核苷酸, 从而阻止致癌 microRNA 的表达(Garzon et al., 2010)。对于低表达的具有肿瘤抑制功能的 microRNA, 则可利用外源的 microRNA 模拟物或病毒载体编码的 microRNA 弥补低水平表达的 microRNA。靶向 microRNA 的治疗方法具有同时调控多个基因和通路的能力, 与现有的靶向疗法相结合, 可以大大改善疾病的反应并提高治愈率。

值得一提的是, 目前 microRNA 标志物的研究大多着重于单个 microRNA 分子层面, 并且相关标志物的应用价值尚未得到充分的临床验证。今后仍需基于多中心、大样本的数据做进一步挖掘和研究。

2. microRNA-mRNA 调控网络的生物信息学研究

2.1 microRNA-mRNA 靶标对的获取

2.1.1 microRNA 数据

近年来, microRNA 一直是生物领域的研究热点, 新的 microRNA 分子鉴定步伐不断加快。miRBase (Kozomara and Griffiths-Jones, 2014) 是常用的收录 microRNA 序列及其注释信息的数据库。在 2014 年 06 月发布的 21.0 版中共收录 28 645 条 microRNA 发夹前体序列, 35 828 条成熟 microRNA 序列, 涵盖动物、植物和病毒等 223 个物种。其中, 人类 microRNA 前体序列 1881 条, 成熟体序列 2588 条。相关数据可以在其官方网站查询下载: <http://www.mirbase.org/>。

2.1.2 靶基因数据

目前, microRNA 靶基因数据的获取主要有两种途径: 计算机算法预测与生物学实验验证。

(1) 计算机算法预测

常用的 microRNA 靶基因预测软件它们的实现原理可归纳为 4 个方面。

1) “种子” (Seed) 匹配: microRNA 靶基因计算方法通常从最简单的 Watson-Crick (WC) 碱基配对开始, 即腺嘌呤 (A) 与尿嘧啶 (U) 匹配, 鸟嘌呤 (G) 与胞嘧啶 (C) 匹配。microRNA 5'端的第 2~8 位碱基作为种子序列与 mRNA 的 3'非编码区进行匹配。完美的“种子”匹配是在配对过程中 microRNA 与 mRNA 之间没有“间隔” (gap) 的存在。然而, 以序列互补性开发的算法存在着准确性低及大量的假阳性结果 (Bagby et al., 2009)。

2) 保守性: 一条序列在多个物种间维持稳定的性质。一般来说, microRNA 的种子区域比非种子区保守性更高。许多算法要求 microRNA 预测的靶点位于同源的 3'非翻译区 (3' Untranslated Region, 3' UTR), 且种子绑定区域是高度保守的。通过物种间预测靶标的序列保守性分析, 剔除了非保守性的绑定位点, 有效得减少了假阳性结果, 提高预测的准确性。

3) 自由能: 这里主要指最小自由能, 用于衡量 microRNA 与其靶标连接的稳定程度。如果 microRNA 与 mRNA 的绑定位点被预测具有很高的稳定性 (即自由能低), 那么这个 mRNA 将更有可能成为 microRNA 靶目标。自由能通常是一个负值, 单位为 kcal/mol (Yue et al., 2009)。

4) 位点的可结合性: 它是 microRNA 靶基因识别重要的生物因素之一, 用来衡量 microRNA 与 mRNA 定位及杂交的难易程度。转录之后, mRNA 的二级结构具有与

microRNA 形成绑定位点的能力。这个过程主要包含两个步骤：①microRNA 与 mRNA 上的短片段、可接和区域绑定；②mRNA 二级结构解旋，microRNA 完成位点绑定。因此，衡量一个位点与 microRNA 结合的总能量，可以作为评定一个 mRNA 是否是某个 microRNA 靶标的标志 (Peterson et al., 2014)。

尽管生物信息学 microRNA 靶基因预测软件被大量报道，但是它们的可靠性仍然具有争议。通常情况下，这些软件针对某一 microRNA 会得到大量靶基因，从几百到几千不等，其间存在很高的假阳性。不同算法的预测结果重合度很低 (Sethupathy et al., 2006)。因此，针对得到的基因数据进行深层次的“清洗” (data cleaning) 和分析显得尤为重要。表 1 列举了常用的 microRNA 靶基因预测软件。

表 1 常用的 microRNA 靶基因预测软件

软件名称	预测原理				网址	参考文献
	“种子” 匹配	保守性	自由能	位点可 结合性		
miRanda	✓	✓	✓	✓	www.microrna.org	Enright et al., 2003; John et al., 2004
TargetScan	✓	✓	×	×	www.targetscan.org	Garcia et al., 2011
DIANA-micro T-CDS	✓	✓	✓	✓	www.microrna.gr/microT-CDS	Paraskevopoulou et al., 2013
MirTarget2	✓	✓	✓	✓	http://mirdb.org	Wang, 2008
PITA	✓	✓	✓	✓	http://genie.weizmann.ac.il/pubs/mir07/mir07_data.html	Kertesz et al., 2007
RNAhybrid	✓	×	✓	×	http://bibiserv.techfak.uni-bielefeld.de/rnahybrid/	Rehmsmeier et al., 2004

研究表明，应用 CLIP-Seq 和 Degradome-Seq 方法能够显著减少 microRNA 绑定位点预测的假阳性，也能够减少 microRNA 靶位点搜索空间的范围 (Addo-Quaye et al., 2008; Chi et al., 2009; Pantaleo et al., 2010)。starBase (<http://starbase.sysu.edu.cn>) (Li et al., 2014) 通过把从 CLIP-Seq 和 Degradome-Seq 获取到的读段 (read) 与用 6 种不同算法预测得到的 microRNA 绑定位点进行匹配，从而发现大量精准的 microRNA 绑定位点。该数据库的最新版本 (V2.0) 包含人类在内的多个物种信息，用户可以通过输入 microRNA 或者基因官方名称 (gene symbol) 获取结果，结果可以以逗号分隔格式 (.CSV) 输出。另外，starBase 提供 microRNA 靶基因功能富集分析的工具，包括基因本体论 (gene ontology, GO) 和信号通路 (pathway) 分析 (Yang et al., 2011)。

同时，基于整合 microRNA 和 mRNA 表达谱的研究方法也为 microRNA 靶基因的识别提供了新的策略。这是因为目前的研究已经表明 microRNA 可以直接影响 mRNA 的表达。因此，通过实验方法 (microarray、qPCR 等) 可以快速有效地检测出对应靶标在转录水平上的影响。HOCTAR (host gene oppositely correlated target) (Gennarino et al., 2011)——基于 microRNA 宿主基因与它们的靶标在表达上具有负相关性，通过整合表达图谱与基于序列开发的预测软件，设计出用来处理一系列 microRNA 转录本数据的新算法。它

利用转录本数据对假定的 microRNA 靶标进行打分,通过对 miRanda、TargetScan 和 PicTar 得到的靶基因进行再分析,有效减少了靶基因的数量,提高了准确度。目前,HOCTAR 提供 290 个人类 microRNA 的预测结果,同时提供基因本体论(GO)的分析工具。网站地址是: <http://hoctar.tigem.it>。

(2) 生物学实验验证

基于各种原理开发的计算预测工具,在极大程度上减少了 microRNA 靶标预测的假阳性,提高了结果准确度,使得 microRNA 与 mRNA 之间的调控关系更加清晰,推动了 microRNA 研究领域的蓬勃发展。但是,仅依赖于计算机算法预测靶基因在揭示 microRNA 调控模式上显然是不全面的,已有相关实验技术被用于验证计算结果。目前,主流的实验技术大致可以分为两类:直接验证和间接验证。

直接验证:报告基因法(reporter gene assay)是最常见的实验方法用于研究 microRNA-mRNA 的相互作用,它能够提供直接的证据表明 microRNA 与 mRNA 之间的调控关系。其基本原理是根据 microRNA 可以抑制报告蛋白(例如,荧光素酶和绿色荧光蛋白)的产生,进而导致蛋白的总量减少或影响蛋白活性(Zeng et al., 2003)。尽管这种方法可以证实 microRNA 与 mRNA 之间的相互作用,但是它不能够鉴别 microRNA 识别元件(对理解相互作用的结构特性很有帮助)。为了能够获取特异的 microRNA 识别元件,有效的办法是在基因的原始序列及突变序列上附着报告基因,进而测量 microRNA 进入细胞前后两个样本基因表达水平的变化。通过这种方法,可以有效识别相互作用的特异位点。

间接验证:基于表达量的验证方法是间接的。因为与 microRNA 相关联的 mRNA/protein 诱导的表达量的变化,包括直接靶点(结构相互作用)和间接靶点(靶点的表达是由直接靶点而不是 microRNA 引起的)。实验技术如基因芯片(microarray)和 PCR 常被用于衡量基因的表达情况。但是,由于 microRNA 的调控不仅是在 mRNA 水平上同时也影响蛋白质的表达水平,已有一些高通量的蛋白质组学技术被相继开发。细胞培养条件下稳定同位素标记技术(stable isotope labeling with amino acids in cell culture, SILAC) (Baek et al., 2008)和脉冲式细胞培养条件下稳定同位素标记技术(pulsed stable isotope labeling with amino acids in cellculture) (Selbach et al., 2008),通过质谱分析能够直接衡量受 microRNA 的改变造成的蛋白质产量的变化。Vinther 等(2006)使用 SILAC 方法研究了 miR-1 在 hela 细胞蛋白质组中的情况,发现 504 个蛋白质中的 12 个蛋白质都受其影响,当 miR-1 过表达时,这 12 个蛋白质的表达量减少。进一步的分析表明,其中 8 个蛋白质相应基因的 3' UTR 的种子区域互补性位点,可以显著的被 miR-1 种子区域富集。表 2 列举了常用的收录实验验证 microRNA-mRNA 调控关系数据库。

表 2 常用收录实验验证 microRNA-mRNA 调控关系的数据库

数据库名称	版本	网址	参考文献
miRecords	V4.0	http://mirecords.umn.edu/miRecords/download.php	Xiao et al., 2009
TarBase	V7.0	http://www.microrna.gr/tarbase	Vergoulis et al., 2012
miR2Disease	—	http://www.mir2disease.org/	Jiang et al., 2009
miRTarBase	V4.5	http://mirtarbase.mbc.nctu.edu.tw/	Hsu et al., 2014