

MicroRNA 靶基因的寻找及鉴定方法研究进展

夏伟, 曹国军, 邵宁生*

军事医学科学院基础医学研究所, 生物化学与分子生物学研究室, 北京 100850

* 联系人, E-mail: shaonsh@hotmail.com

收稿日期: 2008-09-17; 接受日期: 2008-11-28

国家重点基础研究发展计划(批准号: 2005CB724600)和国家自然科学基金(批准号: 30600110)资助项目

摘要 microRNA(miRNA)的生物学功能是人们非常关注的问题。而 miRNA 靶基因的确定是研究 miRNA 生物学功能的关键。目前有关 miRNA 靶基因的确定主要靠计算机生物信息学软件预测和生物学实验方法。其中生物信息学方法主要根据已证实的 miRNA 及其靶基因序列之间相互作用的规律性, 遵循几个常用原则设计的软件完成, 如 miRanda, TargetScan 和 TargetScanS, RNAhybrid, DIANA-microT, PicTar, RNA22 及 FindTar 等。生物学实验方法主要是利用免疫共沉淀寻找与 AGO 蛋白相互作用的 mRNA, 或研究受 miRNA 调控的 mRNA 水平和蛋白质水平变化来寻找 miRNA 靶基因。计算机预测方法和生物学实验方法相互补充和完善, 使人们能够更加方便地确定 miRNA 的靶基因, 从而进一步研究其生物学功能。本文主要介绍了 miRNA 靶基因的寻找及鉴定方法的研究进展。

关键词
miRNA
靶基因
预测
鉴定

microRNA(miRNA)是一类长度在 22 nt 左右的内源非编码小 RNA, 广泛存在于动物、植物、病毒等多种有机体中^[1,2]。1993 年, Lee 等人^[3]首先在秀丽隐杆线虫(*Caenorhabditis elegans*)体内发现了长度为 22 nt 的非编码小 RNA——*lin-4*, 同时发现 *lin-4* RNA 在 *lin-14* 基因的 3'UTR 内存在着互补位点。进一步研究表明, *lin-4* RNA 通过与 *lin-14* 基因的 3'UTR 特异性结合降低了 LIN-14 蛋白的表达水平, 但对 *lin-14* 基因的 mRNA 水平影响并不显著。而 *lin-4* RNA 便是现在所熟知的 miRNA^[4]。miRNA 基因通常位于基因间或内含子区域, 在核内由 RNA 聚合酶转录产生具有帽子结构多聚腺苷酸尾巴的 pri-miRNA。pri-miRNA 在核酸酶 Drosha 和其辅助因子 Pasha 的作用下被处理成由 70 个核苷酸组成的 pre-miRNA, 经 exportin 5 等蛋白转运到细胞质中。另一个核酸酶 Dicer 将其剪切成约 22 个核苷酸长度的 miRNA 双链, 其中一条为

成熟的 miRNA 分子。成熟的 miRNA 分子在细胞内与 Argonaute 蛋白等形成 RNA 诱导的沉默复合体(RISC), 并作用于特异 mRNA 的 3'UTR, 从而抑制翻译过程或直接降解 mRNA^[5]。

已有研究^[6~10]表明, miRNA 通过作用于相应靶 mRNA, 参与细胞增殖、凋亡、分化、代谢、发育、肿瘤转移等多种生物学过程。预测超过 1/3 的人类基因都是保守的 miRNA 靶基因。但目前为止, 仍有很多 miRNA 的功能并不清楚, 原因在于目前已经确定的 miRNA 靶基因数量很少, 而且 miRNA 靶基因的预测与鉴定难度较大。因此, 准确快速地预测并鉴定 miRNA 的靶基因, 对研究 miRNA 功能以及 miRNA 所参与的生物学过程具有十分重要的意义。

1 miRNA 靶基因的寻找

目前主要利用生物信息学方法和生物学实验方

法寻找 miRNA 的靶基因.

1.1 生物信息学方法

生物信息学方法主要是利用某种算法对靶基因样本进行评分及筛选. 与预测动物miRNA靶基因相比, 植物miRNA靶基因的预测要简单得多, 因为在植物中miRNA与靶基因几乎是以完全互补配对的方式结合, 预测时无需复杂的算法. 目前几个针对植物miRNA靶基因的预测方法基本是基于此原理设计的, 且准确率较高^[11,12]. 本文将重点讨论动物中miRNA靶基因的预测方法.

(1) 利用常用规则进行预测. 使用计算机预测动物中miRNA靶基因是一件十分困难的事情, 主要是因为目前已知的miRNA靶基因及其确切的靶点并不多, 在算法编写过程中没有足够的已知样本可参考, 不断增加和改进限制条件, 也难以获得同时具有高特异性和高灵敏度的算法. 并且, 目前对计算机预测出来的靶基因也没有一个快速的高通量鉴定方法, 这也在很大程度上影响了对算法准确性的评估, 无法对其进行优化^[13].

但是, 人们在研究过程中仍然发现, miRNA 和靶基因间的作用具有一定规律性. 目前的预测方法虽然各不相同, 但主要遵循以下几个常用原则: (i) miRNA 与其靶位点的互补性; (ii) miRNA 靶位点在不同物种之间的保守性; (iii) miRNA-mRNA 双链之间的热稳定性; (iv) miRNA 靶位点处不应有复杂二级结构; (v) miRNA 5'端与靶基因的结合能力强于 3'端. 除以上几个基本原则外, 不同的预测方法还会根据各自总结的规律对算法进行不同的限制与优化, 下面概括介绍几种常见的预测方法.

(i) miRanda. miRanda是最早的一个利用生物信息学对miRNA靶基因进行预测的软件, 由Enright等人^[13]于2003年设计开发. 作为最早的miRNA靶基因预测软件, miRanda对 3'UTR的筛选依据主要是从序列匹配、miRNA与mRNA双链的热稳定性以及靶位点的保守性三个方面进行分析.

miRanda 软件首先选取了黑腹果蝇(*Drosophila melanogaster*)中已知的 73 条 miRNA 作为探针, 采用类似于 Smith-Waterman 的算法对 9805 个 3'UTR 进行碱基互补分析. 首先, 互补参数被设定为: G : C, A :

U 为+5, G : U 为+2, 其他配对方式为-3, 同时起始空位(gap-opening)罚分为-8, 延伸空位(gap-extension)罚分为-2; 其次, 为体现与靶基因作用过程中 miRNA 5'端和 3'端的不对称性, 5'端的前 11 个碱基的互补分值(complementarity scores)将乘以一个 scale 参数; 最后, miRNA 与靶 mRNA 的互补还要遵循以下 4 个规则: miRNA 第 2~4 位碱基和靶基因精确匹配; 第 3~12 位碱基和靶基因错配不得多于 5 个; 9~L-5(L 为 miRNA 总长)位碱基最少一个错配; 最后 5 个碱基错配不得多于 2 个. 在热稳定性方面, miRanda 采用 Vienna 软件包计算 miRNA 与 3'UTR 作用的自由能 (ΔG). 此外, 在物种间保守性方面, 要求靶位点在多物种 3'UTR 比对中相同位置处碱基相同.

综合以上 3 条原则, miRanda 选取每条 miRNA 相对的 3'UTR 中排名前 10 位的基因, 作为 miRNA 的候选靶基因, 对于多个 miRNA 对应于同一靶位点的情况, miRanda 则使用贪心算法(Greedy Algorithm)选取其中得分最高且自由能最低的那一对^[14].

(ii) TargetScan 和 TargetScanS. TargetScan 是 Lewis 等人^[15]在 2003 年开发的一款用于预测哺乳动物 miRNA 靶基因的软件, 该软件将 RNA 间相互作用的热力学模型与序列比对分析相结合, 预测不同物种间保守的 miRNA 结合位点.

与 miRanda 不同, 在 TargetScan 的算法中, 要求 miRNA 5' 端第 2~8 位碱基与 mRNA 的 3'UTR 完全互补, 这 7 个核苷酸被称作“miRNA 种子区”. 种子区向两侧延伸直到出现碱基错配为止, 这期间允许 G : U 配对, 同时利用 RNAFold 计算结合位点的自由能. 此外, TargetScan 中引入了信号噪声比来评估预测结果的准确度, 所谓信号噪声比即用已知(信号组)和随机生成的 miRNA(噪声组)分别对 mRNA 的 3'UTR 进行预测, 所得靶基因数目的比值. 当仅针对人(*Homo sapiens*)和小鼠(*Mus musculus*)的 3'UTR 对 miRNA 靶位点进行预测时, 信号噪声比为 2 : 1, 而针对人、小鼠、大鼠(*Rattus norvegicus*)的 3'UTR 进行预测时信号噪声比为 3.2 : 1, 当研究范围扩大到人、小鼠、大鼠和河豚(*Takifugu rubripes*)时, 信号噪声比为 4.6 : 1. 随着物种数目的增多, 预测得到的靶基因减少, 但准确性得到了相应的提高.

后来, Lewis 等人^[16]又对 TargetScan 进行了优化, 即 TargetScanS. TargetScanS 在人、小鼠、大鼠的基础上增加了狗(*Canis familiaris*)和鸡(*Gallus gallus*)的基因组数据, 同时在算法上做了改动, 将种子区由先前的 7 个核苷酸调整为 5' 端第 2~7 位碱基共 6 个核苷酸, 要求在种子区完全互补的情况下, miRNA 第 8 位碱基与靶基因互补或者 miRNA 第 1 位碱基是腺嘌呤。2007 年, Andrew 等人^[17]又将 TargetScanS 的算法中加入了新的条件, 即有效的 miRNA 结合位点多分布于 3'UTR 中 AU 富集的区域; 功能上具有协同作用的 miRNA 靶位点相近; miRNA 的第 12~17 个核苷酸与 3'UTR 互补促进两者的结合; 有效的 miRNA 结合位点优先分布于 3'UTR 的两侧, 但距离终止密码子至少 15 个核苷酸。

(iii) RNAhybrid. RNAhybrid 是 Rehmsmeier 等人^[18]在 2004 年开发的一种基于分析 miRNA 和靶基因间形成双链的二级结构, 从而预测 miRNA 靶基因的软件。RNAhybrid 在实质上是一种经典 RNA 二级结构预测软件的扩展, 它克服了当时如 Mfold 和 RNAFold 等 RNA 二级机构分析软件的不足, 能够更快速、更准确地计算一条短链 RNA 和一条长链 RNA 杂交(如 miRNA 和 mRNA 3'UTR)时的自由能, 并基于此来预测果蝇中 miRNA 的靶基因。RNAhybrid 的算法禁止分子内、miRNA 分子间及靶基因间形成二聚体, 根据 miRNA 和靶基因间结合自由能探测最佳的靶位点。

2006 年, Kruger 和 Rehmsmeier^[19]对 RNAhybrid 软件进行了升级, 在网络版的 RNAhybrid 上, 用户可以自由选择种子区的位置、长度以及是否允许 G : C 错配, 同时还可利用种子区对待分析的 3'UTR 数据库进行初步过滤, 加快了软件的分析速度。另一个新功能就是增加了对非配对区及环状区(loop region)长度的限制, 提高了软件预测的精确度。

(iv) DIANA-microT. DIANA-microT 是 Kiriakidou 等人^[20]于 2004 年开发的一款结合了生物信息学和实验学方法的靶基因预测软件。DIANA-microT 主要考虑单一结合位点的 miRNA 靶基因。此外在寻找结合位点时, 除了必需的 5' 端种子区外, 典型的中央突起以及 miRNA 在 3' 端与 mRNA 的结合也得到了考虑。在算法方面, DIANA-microT 主要基于以下两点原则

对 miRNA 靶基因进行判断: 首先, 通过动态规划算法计算经典的 Watson-Crick 碱基对及 G : U 错配的自由能, 进而衡量 miRNA 与靶基因间的结合能力; 其次, iRNA 相关蛋白影响 miRNA 与靶基因结合时形成的中央突起的大小与位置, 进而影响 miRNA 与靶基因的结合。

(v) PicTar. Krek 等人^[21]在 2005 年采用一种更先进的算法来预测脊椎动物、线虫和果蝇中 miRNA 的靶基因, 并通过实验方法进行了验证, 这种算法就是 PicTar, 即组合靶位点的概率识别(probabilistic identification of combinations of target sites)。

PicTar 的算法包括两个方面: 首先, 识别单个 miRNA 的靶位点。与大部分算法一样, PicTar 同样强调“种子区”在靶位点识别及在转录后调控中的关键作用, 强调 miRNA 和靶基因结合的自由能在靶基因翻译抑制中的关键作用。不同的是 PicTar 把种子序列分为“完全匹配种子序列”和“不完全匹配种子序列”, 前者要求种子序列和靶基因完全互补配对; 后者在满足 miRNA 与靶基因结合自由能不增加的前提下允许种子序列出现错配, 但不允许 G : U 配对。同时, PicTar 结合以前实验数据对两类种子序列对应的 miRNA 和靶基因结合自由能进行了限制, 要求“完全匹配种子序列”的 miRNA 与靶基因结合自由能小于 miRNA 和靶基因最优结合能的 33%; 而“不完全匹配种子序列”的 miRNA 与靶基因的结合自由能要求小于 miRNA 和靶基因最优结合能的 66%, 有效地降低了假阳性率; 其次, 为组合的靶位点评分。在 PicTar 的概率模型中, miRNA 在结合位点上互相竞争, 通过对重叠靶位点给予适当的分值来识别多个 miRNA 作用于同一靶位点, 或一个 miRNA 作用于多个靶位点的协同效应。

通过生物信息学和实验方法的分析, PicTar 算法估计会有 30% 左右的假阳性率。还预测在脊椎动物中, 平均每条 miRNA 大约会与 200 个转录本相互作用, 而 miRNA 簇在基因调控过程中起协同作用^[22]。

(vi) RNA22. RNA22 是由 Miranda 等人^[23]于 2006 年开发的一种识别 miRNA 靶位点以及相应 miRNA-mRNA 异源双链的软件。RNA22 与其他 miRNA 靶基因预测软件不同。首先, RNA22 的预测不依赖于物种间的保守性, 而是认为即使近亲物种

不存在 miRNA 结合位点, 仍有可能是 miRNA 的靶基因; 其次, RNA22 与其他软件的预测方向不同, RNA22 不是从 miRNA 入手寻找它的靶基因, 而是首先从感兴趣的序列入手, 寻找假定的 miRNA 结合位点, 再进一步确定其被哪条 miRNA 调控。RNA22 可以预测大部分已知的 miRNA-mRNA 异源双链, 它的出现对以前人们关于 miRNA 靶基因预测的相关认识提出了挑战。基因组中 miRNA 前体、结合位点以及相关基因的转录本数目可能多于之前的预测, 而 miRNA 的靶基因也可能存在于 3'UTR 以外的区域。其中 miRNA 靶位点可能存在与 5'UTR 的事实已经被证明。

上述不同算法各有其特点, 主要还是依据 miRNA 与其靶位点的互补性、miRNA 靶位点的保守性、miRNA-mRNA 双链之间的热稳定性及附近序列的二级结构等原则设计的。其中 miRanda 是将 miRNA-mRNA 之间的序列匹配, 保守性及热稳定性作为计算参数, 有比较好的检出率, 但是假阳性率也较高; TargetScan 提出了“种子区”的概念, 增加了预测的精确度; RNAhybrid 的研究重点在 RNA 二级结构预测方面, 能够更快速准确地计算 miRNA-mRNA 双链的自由能, 降低了假阳性率; DIANA-microT 则主要考虑 miRNA 调控单个靶基因的情况, 同时考虑中央突起以及 miRNA 在 3 端与 mRNA 的结合; RNA22 与其他算法不同, 不考虑物种间的保守性, 检出率较高。

利用常用规则进行 miRNA 靶基因预测的方法可以概括如表 1。可以看出, 目前常见的生物信息学预测方法各有其自身的特点和适用范围。相信通过生物学实验方法进一步验证, 各种生物信息学预测方

法也会互相补充, 不断改进和完善。

(2) 其他规则的引入对预测的影响。随着 miRNA 靶基因预测方法的不断增加, 拟定算法的标准及参考依据也在不断更新, 新的研究进展及实验结果都会对算法的更新起到推动和促进作用, 如机器学习法的引入 [24,25], 结合现有实验结果或基因芯片结果预测 miRNA 靶基因等 [26~28]。下面主要从新规则引入对预测方法的影响方面进行介绍。

(i) miRNA 和 mRNA 3'UTR 结合特性对靶基因预测的影响。预测 miRNA 靶基因的关键, 在于了解 miRNA 和 mRNA 3'UTR 上其靶位点之间的相互作用机制, 使得在拟定算法时有更多的限制条件, 从而增加预测精确度。总体来说, 了解 miRNA 与靶位点间相互作用主要从以下 3 个方面, 即 5'端种子区、中间的环状突起区以及 3'端的结合区。

Ye 等人 [29] 即是从中间环状区对 miRNA 与靶基因间的作用机制进行探讨, 结合其他参数开发了一个新的预测软件——FindTar。他们以血管内皮生长因子(VEGF)为研究对象, 将环状突起分成 3 种类型: 将起始于 miRNA 5'端第 9~11 个核苷酸的环称为标准环(standard loop), 将起始于第 9 个核苷酸之前的环称为 I 型偏心环, 将起始于第 11 个核苷酸之后的环称为 II 型偏心环。分析了 miRNA 和其已知靶位点的结合方式。同时验证了 VEGF 与不同 miRNA 的相互作用, 结果发现大部分 miRNA 与靶位点都是以标准环的形式结合的。通过对其他数据库的分析, 确定了环状区分值的标准: 就环状区的位置而言, 标准环为 1 分, I 型偏心环为 0.75 分, 开始于第 7 个核苷酸

表 1 动物中 miRNA 靶基因预测方法

预测方法	网址	检索范围	算法特点
miRanda	http://www.microrna.org/	人, 果蝇, 斑马鱼	序列匹配, 双链结合自由能, 物种间保守性
TargetScan/TargetScanS	http://www.targetscan.org/	人, 小鼠, 大鼠, 狗	提出“miRNA 种子区”的概念
RNAhybrid	http://bibiserv.techfak.uni-bielefeld.de/rnahybrid/	哺乳动物	快速准确计算 miRNA-mRNA 二聚体自由能
DIANA-microT	http://www.diana.pcbi.upenn.edu/	人, 小鼠, 大鼠, 果蝇	考虑 miRNA 调控单个靶位点的情况
PicTar	http://pictar.bio.nyu.edu/	脊椎动物	区分“完全匹配种子区”与“不完全匹配种子区”
RNA22	http://cbcdrv.watson.ibm.com/rna22.html	哺乳动物	不考虑保守性, 由 mRNA 入手预测相关 miRNA

或之前的型环为 0 分, 开始于第 8 个核苷酸的型环为 0.5 分; 就环状区的大小而言, 1 bp 为 10 分, 2 bp 或大于 3 bp 为 20 分, 3 bp 为 25 分; 最后将位置得分和大小得分相乘得到最终的环状系数.

利用荧光素酶报告基因的方法, 从灵敏度、特异性、约登指数(youden index)等方面对考虑和不考虑环状系数的两种方法进行比较. 发现增加了环状系数的选项后, 假阳性率降低, 约登指数更加接近 1(约登指数用于衡量预测准确性, 理想值为 1), 但假阴性率略有上升. 总体来说, 引入了环状系数后 miRNA 靶基因预测的灵敏度和精确度有所提高.

(ii) 非结合位点区域对靶基因预测的影响. Didiano 等人^[30]在 2008 年利用转基因线虫对 *lsy-6* miRNA 与它的靶基因 *cog-1* 3'UTR 间的相互作用机制进行了研究, 发现在 *cog-1* 的 3'UTR 中一个或多个种子区的匹配并不能作为预测 miRNA 与靶基因相互作用的标准. 在之前研究 *lsy-6* 调控 *cog-1* 及异源的 3'UTR 的过程中, 他们曾发现将 *cog-1* 3'UTR 靶位点突变后, *lsy-6* 确实失去了调控靶基因的能力, 但其种子区却依然与靶基因结合, 而且将 *lsy-6* 的靶位点构建到 13 条异源基因上之后, 并没有发现 *lsy-6* 对这些基因的表达起到调控作用.

cog-1 基因的 3'UTR 上存在着两个已知的 *lsy-6* 结合位点, 改变任何一个位点在 3'UTR 中的位置都会使得 *lsy-6* 失去对 *cog-1* 的调节作用. 通过实验证明, 每个 *lsy-6* 结合位点后的一段序列对其与靶基因的结合都起到至关重要的作用. 分别对这两段序列突变后显示, *lsy-6* 对靶基因的调节作用减弱, 而将这两段序列删除后发现, *lsy-6* 对靶基因的调控作用会完全消失. 研究几十个不同 3'UTR 之后, 认为在 miRNA 靶基因的预测过程中, 除了在种子区的互补序列以及针对靶位点的相关预测外, 3'UTR 的其他序列甚至整个 3'UTR 在 miRNA 的调控过程中也起到了关键作用. 最近研究表明, 线虫中有一组数量众多的 RNA 结合蛋白可以影响 miRNA 与靶 mRNA 结合效率. 他们分析, 这些位于 miRNA 靶位点 3' 端的序列有可能就是 RNA 结合蛋白的作用位点.

在今后预测 miRNA 靶基因的过程中, 除了传统的算法外, 还应考虑复杂的细胞内环境, 如 miRNA 和 mRNA 与 RNA 结合蛋白间的相互作用, 而这些是

无法通过目前的计算机模拟进行计算的, 需要新的方法来提高预测精确度, 基因组中受 miRNA 调控的基因数量可能并没有现在预测的那么多.

(iii) 靶基因二级结构的影响. 在 miRNA 与靶基因相互作用的过程中, mRNA 3'UTR 的二级结构也起着重要作用. Zhao 等人^[31]发现, 几乎所有的 miRNA 结合位点都位于 3'UTR 的不稳定区域内, 将靶位点上下游 70 nt 序列的自由能列为参考因素, 将为预测提供一定的帮助.

为研究靶基因二级结构对 miRNA 功能的影响, Long 等人^[32]在 2007 年利用 SFold 软件分析了 mRNA 的二级结构, 并提出了 miRNA 与 mRNA 间作用的“两步杂交模型”: 首先, miRNA 与靶基因上的 4 个连续未配对核苷酸结合形成核心碱基对; 之后, miRNA-mRNA 双链逐渐延伸, 破坏靶基因原有的二级结构, 并最终形成 miRNA-mRNA 的不完全互补双链区. 该模型的核心在于计算靶基因原有二级结构的破坏、核心碱基对的形成和最终的 miRNA-mRNA 双链形成过程中获得或损失的自由能. 他们分析了大量已公布的数据, 包括线虫中含有 *lin-41* 3'UTR 序列报告基因的体内活性, 以及其他在线虫和果蝇中得到实验证的 miRNA 与 mRNA 的相互作用等, 发现利用“两步杂交模型”将靶基因的二级结构作为参考指标之一, 和现有的数据吻合良好. 直接的实验证据也表明, 增加靶位点附近二级结构的稳定性可大大降低 miRNA 的作用. 因此新的预测方法还需利用这一点来提高预测精确度.

通过上述方法可见, 新的技术手段能够为靶基因的预测带来新的计算参数, 新的研究结果可以为预测提供更多的学习样本, 便于更好的总结规律、优化方法. 未来的靶基因预测方法将会更多地依赖生物学实验的研究进展, 尤其是高通量的靶基因鉴定方法将会对计算机预测产生重要的影响.

1.2 生物学实验方法

由于计算机模拟在预测 miRNA 靶基因时存在一定的局限性, 因此很多学者也希望利用生物学实验方法直观地寻找 miRNA 靶基因.

(1) 从 mRNA 水平寻找 miRNA 靶基因

(i) miRNA 靶基因的大规模寻找. 一种方法是

将miRNA成熟体双链或腺病毒载体转染至细胞中,使得miRNA在细胞中过表达,之后利用基因芯片分析mRNA的变化以找出相应miRNA的靶基因^[33].由于miRNA在体内通过翻译抑制或影响mRNA的稳定性起作用,因此这种方法在寻找起翻译抑制作用的miRNA的靶基因时存在一定困难.

此外,Beitzinger等人^[34]利用AGO蛋白家族既能结合miRNA又能结合mRNA的特性,分别使用AGO-1和AGO-2蛋白的单克隆抗体在人类细胞中进行免疫共沉淀,得到与AGO-1和AGO-2蛋白结合的mRNA各600条,并通过克隆测序对这些mRNA进行鉴定.同时从与AGO-1蛋白结合的mRNA中随机挑选6条进行荧光素酶报告基因检测,发现其中有5条都是miRNA的靶基因.此外还发现,在与AGO蛋白结合的mRNA中,只有大约60%被3个最常用的预测软件(miRanda, TargetScan和PicTar)预测为miRNA的靶基因,说明现有的预测方法也许错过了相当数量的miRNA靶基因.不过他们也指出,只有高通量的miRNA靶基因验证方法出现之后,才能对现有的各种预测方法做一个相对准确的评估.

与上述工作类似,Easow等人^[35]也是利用纯化miRNP来分析miRNA的靶基因.他们利用基因芯片分析了miRNP结合的mRNA后发现,大量计算机预测的miRNA靶基因得到了富集,同时为了分析单个miRNA的靶基因,对野生型果蝇和miR-1缺失果蝇的miRNP结合mRNA进行了分析,得到并利用实验方法鉴定miR-1调节的mRNA,为miRNA靶基因的寻找提供了新的思路.

除上述工作外,我国武汉大学的张翼教授课题组也在通过利用miRNA与AGO抗体免疫共沉淀得到与其相关的mRNA,从而建立寻找特异miRNA靶基因的方法(未发表资料).本研究组也在尝试利用一种RNA伴侣分子,在体外钓取与miRNA特异结合的3'UTR序列,目前已取得了比较好的进展(未发表资料).

(ii) miRNA靶基因的精确寻找.最近,Liu等人^[36]在研究Mir-16家族的功能时建立了一种针对miRNA靶基因的反向筛选法.首先,确定了感兴趣的基因ccnd1,将它的3'UTR构建到荧光素酶报告载体

中,之后与构建的miRNA表达文库共转染HepG2细胞,得到荧光值明显下降的miRNA,再将得到的miRNA在体内验证其功能,最终得到Mir-16家族能够调节CCND1基因.该方法的关键在于,利用了自行构建的miRNA表达文库,能够方便同时研究多条miRNA,通过体外实验对候选miRNA进行初步筛选,从而快速地鉴定一条mRNA所对应的多条miRNA.

(2) 从蛋白质水平寻找miRNA靶基因

由于miRNA所介导的转录后翻译抑制过程不引起mRNA水平的改变,因此依据mRNA水平的变化寻找miRNA靶基因的方法在检出率上存在一定问题.不久前,Selbach^[37]和Baek^[38]两个研究组分别利用蛋白质组学的研究方法,建立了从蛋白质水平寻找miRNA靶基因的新方法.

两个课题组分别将成熟miRNA双链转染HeLa细胞,利用细胞培养稳定同位素标记技术(stable isotope labeling with amino acids in cell culture),将转染了成熟miRNA双链的细胞和正常细胞分别培养于含轻/重型稳定同位素标记必需氨基酸的培养基中,经过若干次细胞倍增后,稳定同位素按照序列特异的方式完全掺入到新合成的蛋白质中,通过比较标记前后同一肽段的质谱峰值变化实现对蛋白质的精确定量.在检测了2000~5000个蛋白后,两个研究组分别发现,转染一条miRNA后有几百个蛋白的表达水平发生了改变,许多改变在mRNA水平上不能得到体现.

蛋白质组学相关研究方法的引入,使得人们可以从蛋白质水平上寻找miRNA的靶基因,从而提高了靶基因的检出率.

2 miRNA靶基因的鉴定

与miRNA靶基因的预测方法相比,对miRNA靶基因进行实验验证的方法并不多,目前还没有一个快速、简便、高通量的鉴定方法.

最直接的鉴定方法是,利用荧光定量PCR及Western blot方法分别检测转染或敲低miRNA后细胞中mRNA水平及蛋白水平的变化,从而确定miRNA与靶基因的对应关系.这种方法能够直接鉴定出miRNA的靶基因,准确度高但不能鉴定miRNA的靶位点.

目前最为常用的miRNA靶位点鉴定方法是荧光素酶报告基因法。其基本原理是首先构建荧光素酶表达载体, 将希望鉴定的miRNA靶基因的3'UTR构建到荧光素酶基因的3'UTR中, 之后将构建好的载体转染细胞并改变细胞中相应miRNA的表达水平, 最后检测荧光素酶的表达情况以分析转染3'UTR中是否含有miRNA的靶位点^[13,39]。

3 问题与展望

近几年来, 对miRNA的研究已经成为生命科学领域中的一个重要方向。miRNA分布范围广泛, 参与的生物学过程复杂, 调控的靶基因众多, 因此研究miRNA的功能及作用机制有着重要的意义。寻找并鉴定miRNA的靶基因是研究miRNA功能的基础, 然而由于miRNA的长度很短而且和靶基因间并非完全互补配对, 因此若想预测miRNA的靶基因是十分困难的。即使通过各种手段预测了miRNA的靶基因, 由于没有快速简便的验证方法, 也无法对预测的准确性进行评判。另外, 通过对现有miRNA与靶基因之间关系的研究, 还要考虑到miRNA的多靶基因特

性可能与不同的细胞状态有密切关系, 即miRNA与靶基因的关系可能是动态可变的, 因此有关miRNA靶基因的研究可能会比想象的还要复杂。但是经过不断探索, 找到了miRNA与靶基因作用的很多规律, 并基于此编写了各种预测miRNA靶基因的算法。

从目前文献中可以看出, 每一种新的预测方法的产生都有着新的实验结果、新的技术手段的支持。通过对miRNA与靶基因相互作用机制的研究, 使得人们在预测时能够加入更多的参数, 提高预测的灵敏度和精确度。随着实验验证的miRNA靶基因不断增多, 在设计算法时能够有更多的统计样本, 能够更准确地找到miRNA与靶基因间相互作用的特征, 优化计算方法。此外, 随着验证方法的不断更新, 对预测结果的实验鉴定也会更快速、更便捷, 预测方法将得到更有效的评估。计算机预测方法和生物学实验方法相互补充和完善, 不但能够减少miRNA靶基因寻找的盲目性, 节约实验成本, 而且可以使人们能够更有针对性地研究感兴趣的miRNA, 更加准确方便地阐明其在生命活动中的功能与意义。

致谢 感谢本研究室同事对本文提供的帮助。

参考文献

- 1 Bartel B, Bartel D P. MicroRNAs: at the root of plant development? *Plant Physiol.*, 2003, 132(2): 709—717 [[DOI](#)]
- 2 Lim L P, Glasner M E, Yekta S, et al. Vertebrate microRNA genes. *Science*, 2003, 299(5612): 1540 [[DOI](#)]
- 3 Lee R C, Feinbaum R L, Ambros V. The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, 1993, 75(5): 843—854 [[DOI](#)]
- 4 Lagos-Quintana M, Rauhut R, Lendeckel W, et al. Identification of novel genes coding for small expressed RNAs. *Science*, 2001, 294(5354): 853—858
- 5 Bartel D P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, 2004, 116(2): 287—297
- 6 Hwang H W, Mendel J T. MicroRNAs in cell proliferation, cell death, and tumorigenesis. *Brit J Cancer*, 2006, 94(6): 776—780 [[DOI](#)]
- 7 Cheng A M, Byrom M W, Shelton J. Antisense inhibition of human miRNAs and indications for an involvement of miRNA in cell growth and apoptosis. *Nucleic Acids Res*, 2005, 33(4): 1290—1297 [[DOI](#)]
- 8 Croce C M, Calin G A. miRNAs, cancer, and stem cell division. *Cell*, 2005, 122(1): 6—7 [[DOI](#)]
- 9 Ma L, Teruya-Feldstein J, Weinberg R A. Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. *Nature*, 2007, 449(7163): 682—688 [[DOI](#)]
- 10 Tavazoie S F, Alarco C. Endogenous human microRNAs that suppress breast cancer metastasis. *Nature*, 2008, 451(7175): 147—152 [[DOI](#)]
- 11 Rhoades M W, Reinhart B J, Lim L P, et al. Prediction of plant microRNA targets. *Cell*, 2002, 110(4): 513—520 [[DOI](#)]
- 12 Zhang Y. miRU: an automated plant miRNA target prediction server. *Nucleic Acids Res*, 2005, 33: W701—W704 [[DOI](#)]

- 13 Enright A J, John B, Gaul U, et al. MicroRNA targets in *Drosophila*. *Genome Biol*, 2003, 5(1): R1 [[DOI](#)]
- 14 John B, Enright A J, Aravin A, et al. Human microRNA targets. *PLoS Biol*, 2004, 2(11): e363 [[DOI](#)]
- 15 Lewis B P, Shih I H, Jones-Rhoades M W, et al. Prediction of mammalian microRNA targets. *Cell*, 2003, 115(7): 787—798 [[DOI](#)]
- 16 Lewis B P, Burge C B, Bartel D P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are MicroRNA targets. *Cell*, 2005, 120(1): 15—20 [[DOI](#)]
- 17 Andrew G, Kyle K F, Wendy K J. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Molecular Cell*, 2007, 27(1): 91—105 [[DOI](#)]
- 18 Rehmsmeier M, Steffen P, Hochsmann M, et al. Fast and effective prediction of microRNA/target duplexes. *RNA*, 2004, 10(10): 1507—1517 [[DOI](#)]
- 19 Kruger J, Rehmsmeier M. RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Res*, 2006, 34: W451—454 [[DOI](#)]
- 20 Kiriakidou M, Nelson P T, Kouranov A, et al. A combined computational/experimental approach predicts human microRNA targets. *Genes Dev*, 2004, 18(10): 1165—1178 [[DOI](#)]
- 21 Krek A, Grun D, Poy M N, et al. Combinatorial microRNA target predictions. *Nature Genet*, 2005, 37(5): 495—500 [[DOI](#)]
- 22 Grun D, Wang Y, Langenberger D, et al. MicroRNA target predictions across seven *Drosophila* species and comparison to mammalian targets. *PLoS Comput Biol*, 2005, 1(1): e13 [[DOI](#)]
- 23 Miranda K C, Huynh T, Tay Y, et al. A pattern-based method for the identification of microRNA binding sites and their corresponding heteroduplexes. *Cell*, 2006, 126(6): 1203—1217 [[DOI](#)]
- 24 Kim S K, Nam J W, Rhee J K, et al. MiTarget: microRNA target gene prediction using a support vector machine. *BMC Bioinform*, 2006, 7: 411 [[DOI](#)]
- 25 Thadani R, Tammi M T. MicroTar: predicting microRNA targets from RNA duplexes. *BMC Bioinform*, 2006, 7(5): S5—20
- 26 Moxon S, Moulton V, Kim J T. A scoring matrix approach to detecting miRNA target sites. *Algorithms Mol Biol*, 2008, 3: 3 [[DOI](#)]
- 27 Cheng C, Li L M. Inferring microRNA activities by combining gene expression with microRNA target prediction. *PLoS ONE*, 2008, 3(4): e1989
- 28 Wang X, El Naqa I M. Prediction of both conserved and nonconserved microRNA targets in animals. *Bioinformatics*, 2008, 24(3): 325—332 [[DOI](#)]
- 29 Ye W, Lv Q, Wong C K, et al. The effect of central loops in miRNA: MRE duplexes on the efficiency of miRNA-mediated gene regulation. *PLoS ONE*, 2008, 3(3): e1719 [[DOI](#)]
- 30 Didiano D, Hobert O. Molecular architecture of a miRNA-regulated 3' UTR. *RNA*, 2008, 14(7): 1297—1317 [[DOI](#)]
- 31 Zhao Y, Samal E, Srivastava D. Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis. *Nature*, 2005, 436(7048): 214—220 [[DOI](#)]
- 32 Long D, Lee R, Williams P, et al. Potent effect of target structure on microRNA function. *Nat Struct Mol Biol*, 2007, 14(4): 287—294 [[DOI](#)]
- 33 Lim L P, Lau N C, Garrett-Engle P, et al. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature*, 2005, 433(7027): 769—773 [[DOI](#)]
- 34 Beitzinger M, Peters L, Zhu J Y, et al. Identification of human microRNA targets from isolated argonaute protein complexes. *RNA Biol*, 2007, 4(2): 76—84
- 35 Easow G, Teleman A A, Cohen S M. Isolation of microRNA targets by miRNP immunopurification. *RNA*, 2007, 13(8): 1198—1204 [[DOI](#)]
- 36 Liu Q, Fu H, Sun F, et al. miR-16 family induces cell cycle arrest by regulating multiple cell cycle genes. *Nucleic Acids Res*, 2008, 36(16): 5391—5404 [[DOI](#)]
- 37 Selbach M, Schwanhäusser B, Thierfelder N, et al. Widespread changes in protein synthesis induced by microRNAs. *Nature*, 2008, 455(7209): 58—63 [[DOI](#)]
- 38 Baek D, Villén J, Shin C, et al. The impact of microRNAs on protein output. *Nature*, 2008, 455(7209): 64—71 [[DOI](#)]
- 39 Welch C, Chen Y, Stallings R L. MicroRNA-34a functions as a potential tumor suppressor by inducing apoptosis in neuroblastoma cells. *Oncogene*, 2007, 26(34): 5017—5022 [[DOI](#)]