

# 染色体上聚集的 microRNAs 具有更多共同的靶基因

李永庆<sup>1</sup>, 徐建震<sup>1</sup>, 孟志刚<sup>2</sup>, 李永进<sup>3</sup>

(1. 广东省工程技术研究所, 广州 510000; 2. 湖州师范学院信息工程学院, 浙江湖州 313000;

3. 哈尔滨医科大学生物信息学系, 哈尔滨 150086)

**摘要:** miRNAs 是一类 ~22nt, 在转录后调节 mRNAs 表达的内源性非编码 RNA。人类 miRNA 具有成簇聚集于染色体上的特点。文章系统分析了人基因组簇内 miRNA 各成员的预测靶基因之间的关系, 发现簇内 miRNA 各成员具有更多共同的靶基因, 表明簇内 miRNA 各成员功能相似。仔细分析簇内 miRNA 和 mRNA 的结合位点, 发现有两种类型: 一种是簇内 miRNAs 可能竞争性结合同一个靶基因的同个结合位点; 另一种是簇内 miRNAs 可能协同结合于同一个靶基因的不同结合位点。

**关键词:** microRNA; 前体 miRNA; 成簇 miRNA; miRNA 基因组学

**学科分类号:** Q6

## 0 引言

microRNA (miRNA) 是一类 ~22 nt 的内源性非编码 RNA。在细胞核内, miRNA 基因的初级转录产物 (pri-miRNA) 被 RNase III Droscha 切割成前体 miRNA (pre-miRNA)。然后, 由转运蛋白 Exportin-5 转运到细胞质中, 在这里被 RNaseIII Dicer切割加工成有功能的成熟 miRNA<sup>[1,2]</sup>。在动物体中, miRNA 通过非完全碱基配对的方式结合到被其调控的 mRNA 基因的 3' 端非翻译区 (3'UTR) 互补配对的位点上, 然后通过未知的机制抑制其翻译或者通过加速的脱腺苷化, 导致靶 mRNA 的降解<sup>[3]</sup>。目前已证明 miRNAs 在哺乳动物造血干细胞分化、脂肪代谢、癌症发生等方面发挥重要的调控作用。

计算生物学方法在 miRNA 研究中发挥重要作用。但目前研究集中于预测新的 miRNA 基因以及预测 miRNA 的靶基因方面, 而关注 miRNA 的基因组特征与其功能发挥之间关系的研究还刚刚开始。事实上, Altuvia 等<sup>[4]</sup>发现人类 42% 的 miRNA 都可以成簇 (cluster) 聚集于基因组上, 相互之间距离 <3 kbp。最近 Yu 等<sup>[5]</sup>发现在 51 个成簇的 miRNAs 中, 39 个 miRNA 簇的各个成员 miRNAs 在几种白血病细胞株中表达谱相似, 提示这些 miRNA 簇中的成员可能发挥相似的功能。本文研究成簇出现的 miRNA 各成员预测得到的靶基因之间的重合性, 从而探讨簇内 miRNA 各成员之间功能的关系。

## 1 方法

### 1.1 基因组注释信息

UCSC genome browser<sup>[6]</sup>提供了人类基因组的注释信息, 使用的版本是 hg18 (March 2006 assembly)。其中包括了 533 个人类前体 miRNAs 的基因组位置。

### 1.2 成簇出现的 miRNAs

按照前体 miRNA 在染色体上的位置, 利用簇内成员 miRNAs 相互之间在染色体上的距离 ( $D$ ) 定义 miRNA 簇。即定义  $D < X$  bp 的 miRNAs 属于同一个 miRNA 簇。为衡量结论对于不同簇定义的稳定性, 我们分别考察了  $D < 0.5$  kbp、 $D < 5$  kbp 和  $D < 50$  kbp 的各种情况。按照上述定义, 从 miR-Gen 数据库提取成簇出现的 miRNAs<sup>[7]</sup>。

### 1.3 靶基因预测

为衡量结论对于不同预测算法的稳定性, 我们分别利用 PicTar<sup>[8]</sup>和 miRanda 算法<sup>[9]</sup>预测簇内各成员 miRNAs 的靶基因。从 <http://pictar.bio.nyu.edu/> 和 <http://www.microrna.org/mammalian/index.html> 下载了人类所有 miRNAs 预测的靶基因, 分别包含了对人类 178 和 157 个 miRNAs 的预测靶基因。

收稿日期: 2007-03-11

通讯作者: 徐建震, 电话: (020)86092216,

E-mail: [jianzhu@hotmail.com](mailto:jianzhu@hotmail.com)

### 1.4 评价靶基因重合的方法

首先按簇定义得到 miRNA 成员关系, 其次我们把各个簇中的 miRNA 成员两两组合成 miRNA 对, 然后对每对 miRNA 中的每个 miRNA 分别预测其靶基因, 最后统计靶基因的重合数目。由于不同的 miRNA 预测的靶基因数目有多有少, 我们设计如下指标考察重合率:

$$Ratio\_overlap = \frac{Num\_overlap}{(Num\_union\_miRNA\_A\_and\_B)}$$

$Num\_overlap$  表示组合成一对的 miRNA\_A 和 miRNA\_B 预测靶基因的交集中成员的数目, 即预测的靶基因重合数目。 $Num\_union\_miRNA\_A\_and\_B$  表示 miRNA\_A 和 miRNA\_B 预测靶基因的并集中成员的数目, 其中 miRNA\_A 和 miRNA\_B 位于同一个簇中。该指标越高, 重合的靶基因占这对 miRNA 靶基因并集中的比例越大, 说明这一对 miRNAs 具有更多共同的靶基因。同时为考察显著性, 构造随机抽样数据集如下: 不考虑 miRNAs 在染色体上的距离关系, 从所有被 PicTar 或者 miRanda 预测的 miRNAs 中随机抽取 1000 对 miRNAs, 然后对每对 miRNA 类似地计算该指标。用秩和检验计算实际分组和随机组之间的

统计显著性。所有计算过程在 Matlab release 13 (<http://www.mathworks.com/>) 平台下自编脚本实现。

## 2 结果与分析

### 2.1 不同距离定义下染色体上成簇出现 miRNA

簇内成员 miRNAs 相互之间在染色体上距离  $D < 0.5$  kb、 $D < 5$  kb 和  $D < 50$  kb 的情形下, 我们分别得到了 48、57 和 59 个在染色体上聚集的 miRNA 簇。在染色体距离定义  $D < 5$  kb 时, 除了第 2 号、10 号、15 号、16 号以及 Y 染色体外, 其他各条染色体上都发现有 miRNA 簇。其中 X 染色体有相对较多的 miRNA 簇。其他各条染色体上 miRNA 簇分布数目没有明显的偏歧。最小的 miRNA 簇中包含 2 个 miRNA 成员, 最大的簇包括 36 个 miRNA 成员, 位于 19 号染色体。当染色体距离定义延伸至  $D < 50$  kb 时, 在第 2 号、16 号染色体上都发现有 miRNA 簇。最大的簇包括 46 个 miRNA 成员, 位于 19 号染色体。各条染色体上 miRNA 簇统计数目见表 1。

**Table 1** Number of microRNA clusters in each human chromosome

Chromosome	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	Y
Number of miRNA cluster ( $D < 0.5$ kb)	3	0	2	1	2	0	4	0	2	0	2	2	2	8	0	0	5	0	4	1	0	0	10	0
Number of miRNA cluster ( $D < 5$ kb)	5	0	2	1	3	1	5	1	3	0	2	2	2	4	0	0	5	1	6	1	1	2	10	0
Number of miRNA cluster ( $D < 50$ kb)	6	1	3	1	3	2	5	1	3	0	3	2	2	2	0	1	6	1	4	1	1	2	9	0

在每个分组中, 各个簇内 miRNA 成员组对后利用 PicTar 或 miRanda 算法预测 miRNAs 的靶基因。由于目前这两种算法仅分别对 178 和 157 个人类 miRNAs 预测了靶基因, 一对 miRNA 中可能出现有的 miRNA 没有预测的靶基因的情况。此时, 我们不考虑这样的 miRNA 对。利用 PicTar 算法, 染色体距离为  $D < 0.5$  kb、 $D < 5$  kb 和  $D < 50$  kb 时, 分别得到了有预测靶基因的 27、36 和 71 个 miRNA 对, 位于 12 个、19 个和 35 个 miRNA 簇

中; 利用 miRanda 算法, 染色体距离为  $D < 0.5$  kb、 $D < 5$  kb 和  $D < 50$  kb 时, 分别得到了 13、23 和 51 个 miRNA 对, 位于 9 个、17 个和 31 个 miRNA 簇中。表 2 列出了这些 miRNA 对所在的 miRNA 簇的染色体位置, 可见这些 miRNA 簇在各条染色体上的分布没有明显的偏歧, 表明所使用的这些有预测靶基因的 miRNA 对数据具有基因组代表性。

**Table 2** Evaluation of miRNA clusters location

Prediction algorithms	Groups	Number of cluster	Chromosomelocation
PicTar	Group1 ( $D < 0.5$ kbp)	12	4, 7, 9, 11, 12, 13, 17, 19, X
	Group2 ( $D < 5$ kbp)	19	4, 5, 6, 7, 8, 9, 11, 12, 13, 14, 17, 19, 21, X
	Group3 ( $D < 50$ kbp)	35	1, 2, 3, 4, 5, 6, 7, 9, 11, 12, 13, 14, 17, 19, 20, 21, 22, X
miRanda	Group1 ( $D < 0.5$ kbp)	9	7, 11, 12, 17, 19, X
	Group2 ( $D < 5$ kbp)	17	1, 5, 7, 8, 9, 11, 12, 14, 17, 19, 21, X
	Group3 ( $D < 50$ kbp)	31	1, 2, 3, 6, 7, 8, 9, 11, 13, 14, 17, 19, 20, 21, X

## 2.2 成簇出现的 miRNAs 具有更多的共同靶基因

对上述 miRNA 对计算靶基因的重合率指标。表 3 列出了在不同距离组, miRNA 组对数、重合率指标的中位数以及统计显著性。结果可见, 在不同的染色体距离组, 簇内 miRNA 成员所调控的靶

基因重合率都显著高于随机抽样组。采用两种不同的预测算法得到一致的结果, 说明结论对于不同的预测算法具有稳定性。该计算结果表明染色体上成簇出现的 miRNAs 具有更多的共同靶基因。

**Table 3** Significance of overlapped target genes within clusters

Prediction algorithms	Groups	miRNAs pairs	Median	P value
PicTar	Group1 ( $D < 0.5$ kbp)	27	0.0589	2.42E-06
	Group2 ( $D < 5$ kbp)	36	0.0532	8.62E-06
	Group3 ( $D < 50$ kbp)	71	0.0480	3.14E-08
	Random sample	1000	0.0327	-----
miRanda	Group1 ( $D < 0.5$ kbp)	13	0.0785	0.0050
	Group2 ( $D < 5$ kbp)	23	0.0624	9.21E-04
	Group3 ( $D < 50$ kbp)	51	0.0543	1.43E-04
	Random sample	1000	0.0384	-----

## 2.3 两类不同的 miRNA 和 mRNA 的结合位点

一些研究表明 miRNA 利用不同的方式调控 mRNA 的表达。首先序列相似的 miRNA 可以竞争作用于同一个靶基因的同个结合位点上, 这样当不需要降低靶基因表达时, 可以通过竞争抑制来改变 miRNA 的作用<sup>[10]</sup>; 其次, 不同的 miRNAs 可

以作用于同一个靶基因的不同结合位点, 从而协同调控靶基因<sup>[8,9]</sup>。那么在染色体上邻近的 miRNA 簇内的 miRNA 成员之间是否依然可能存在上述的调控方式呢? 仔细分析簇内 miRNA 和 mRNA 的结合位点, 我们发现有些簇内 miRNAs 序列之间仅有个别核苷酸差异, 因此作用于同一个靶基因的同

一个结合位点上。例如 hsa-miR-34b 和 hsa-miR-34c 位于 11 号染色体、距离 < 0.5 kbp 的同一个簇内，序列分别是：5'\_UAGGCAGUGU-CAUUAGCUGAUUG\_3' 和 5'\_AGGCAGUGUAGU-UA-GCUGAUUGC\_3'。它们被 miRanda 预测可能调控靶基因 ENST00000218172\_P，结合位点都是 5'\_TGAGAAGGT-GTGACACTGCCT\_3'。这提示这两个 miRNAs 可能竞争性调控同一个靶基因。但也存在序列完全不同的 miRNA 簇成员，作用于同一个靶基因不同结合位点。例如 hsa-miR-99b 和 hsa-miR-125a 位于同一个簇内，序列分别是：5'\_CACCCGUAGAACCGACCUUGCG\_3' 和 5'\_UCCCUGAGACCCUUUAACCUUGUG\_3'。虽然它们都被 miRanda 预测调控靶基因 ENST00000263733，但预测结合在不同的位点上。

hsa-miR-125a 的结合位点是 5'\_CCCTGTTTGC-CA---CTCAGGGT\_3'，而 hsa-miR-99b 的结合位点是 5'\_TCTAATGT--GTTCTATGGGTT\_3'。

这表明这两个 miRNAs 可能协同调控同一个靶基因。这些观察提示位于染色体上同一个簇的 miRNAs 成员之间具有竞争性或者协同调控共同靶的序列基础。

### 3 讨 论

本研究探讨了人类染色体上聚集的 microRNAs 的靶基因特点。首先，发现染色体上近邻的成簇 miRNAs 成员具有更多共同的靶基因。Baskerville 等<sup>[1]</sup>研究表明可能成簇的 miRNAs 通过共同的 pri-miRNAs 转录单位进行转录，其他研究结果也表明许多成簇出现的 miRNAs 在各个白血病细胞株中表达谱相似<sup>[2]</sup>。以上证据都提示，簇内的 miRNAs 基因可能具有共同或相似的调控功能。其次，我们发现这些同一个簇内的 miRNA 序列可以十分相似也可以十分不同，可以作用于共同靶基因的同个结合位点也可以分别有各自的结合位点，提示具有共同靶基因的 miRNAs 可能至少存在两种不同的调控机制，即通过竞争同一个结合位点从而竞争性调控同一个靶基因，或者通过结合不同的结合位点协同调控同一个靶基因。

由于目前已经实验证明的人类 miRNA 靶基因还很少，我们采用计算生物学的方法预测 miRNA 的靶基因。Stark 等<sup>[3]</sup>利用约 130 个实验证实的果蝇 miRNA-mRNA 调控实例，评估了几个预测算法

性能。结果表明 PicTar 算法的准确率 (accuracy) 约 90%，灵敏性 (sensitivity) 在 60%~80% 之间，是目前最好的预测算法之一，因此，我们首选该算法作为本研究的工具。同时，为衡量结论对于不同预测算法的稳定性，我们还选用了另一种常用的 miRanda 方法预测 miRNA 的靶基因。在本实验中采用两种不同的预测算法得到了一致的结论。

本研究工作还有进一步拓展的空间，比如我们在计算靶基因的重合指标时，要求两个 miRNA 的靶基因要完全一致。但是从生物学角度出发，不同的基因可以发挥相似的功能，例如位于同一个代谢通路上的不同基因。那么也许 miRNA 簇内不同的 miRNA 成员拥有较少直接重合的靶基因，但却分别调控了代谢通路上的不同的基因。另外，受数据所限，我们现在仅使用了大约 1/3~1/2 染色体上的 miRNA 簇。尽管这些 miRNA 簇数据具有较好的代表性，但随着数据完善，尤其是如果能积累足够的生物学实验已证实的 miRNA 和靶基因关系，我们可以更加完整地考察这个现象。

自从 miRNA 被发现以来，引起研究人员巨大的兴趣。miRNAs 如何调控靶基因以及自身的被调控机理等研究领域都方兴未艾。分析 miRNAs 基因的基因组特征与调控功能之间的关系有助于更好地理解其作用机制，同时可以更加精确地预测靶基因。揭示出 miRNA 与经典基因调控网络之间相互作用图并应用于疾病基因的预测将是今后 miRNA 研究的更大挑战<sup>[3]</sup>。

### 参考文献:

- [1] 李伟, 金由辛. miRNAs 的生物合成过程. *生物化学与生物物理进展*, 2005,32(8):707~711
- [2] Bartel DP. MicroRNAs genomics, biogenesis, mechanism and function. *Cell*, 2004,116(2):281~291
- [3] Wu LG, Fan JH, Belasco JG. MicroRNAs direct rapid deadenylation of mRNA. *Proc Natl Acad Sci USA*, 2006,103(11):4034~4039
- [4] Altuvia Y, Landgraf P, Lithwick G, Elefant N, Pfeffer S, Aravin A, Brownstein MJ, Tuschl T, Margalit H. Clustering and conservation patterns of human microRNAs. *Nucl Acids Res*, 2005,33(8):2697~2706
- [5] Yu J, Wang F, Yang GH, Wang FL, Ma YN, Du ZW, Zhang JW. Human microRNA clusters: genomic organization and expression profile in leukemia cell lines. *Biochem Biophys Res Commun*, 2006,349:59~68
- [6] Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, Kent WJ. The UCSC Table Browser data retrieval tool. *Nucl Acids Res*, 2004,32(Suppl 1):D493~D496

- [7] Megraw M, Sethupathy P, Corda B, Hatzigeorgiou AG. miRGen: a database for the study of animal microRNA genomic organization and function. *Nucl Acids Res*, 2007, 35:D149~D155
- [8] Krek A, Grün D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, MacMenamin P, da Piedade I, Gunsalus KC, Stoffel M, Rajewsky N. Combinatorial microRNA target predictions. *Nature genetics*, 2005,37(5):495~500
- [9] John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS. Human microRNA targets. *PLoS Biol*, 2004,2:e363
- [10] Hua Z, Lü Q, Ye W, Wong CK, Cai G, Gu D, Ji Y, Zhao C, Wang J, Yang BB, Zhang Y. MiRNA-Directed Regulation of VEGF and other angiogenic factors under hypoxia. *PLoS ONE*, 2006,1:e116
- [11] Baskerville S, Bartel DP. Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA*, 2005,11:241~247
- [12] Stark A, Brennecke J, Bushati N, Russell RB, Cohen SM. Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution. *Cell*, 2005, 123:1133~1146
- [13] Xu JZ, Li YJ. Discovering disease-genes by topological features in human protein-protein interaction network. *Bioinformatics*, 2006,22:2800~2805

## CLUSTERED microRNAs TEND TO REGULATE MORE COMMON TARGET GENES

LI Yong-qing<sup>1</sup>, XU Jian-zhen<sup>1</sup>, MENG Zhi-gang<sup>2</sup>, LI Yong-jin<sup>3</sup>

(1. Guangdong Engineering Technology Research Institute, Guangzhou 510000, China;

2. School of information engineering, Huzhou Normal college, Huzhou 313000, Zhejiang, China;

3. Department of bioinformatics, Harbin Medical University, Harbin 150000, China)

**Abstract:** MicroRNAs (miRNAs) are one class of ~22 nt, endogenous single-stranded RNAs which can regulate the mRNA expression at the post-transcriptional level. Previous analysis revealed that human miRNAs tended to cluster in regions on chromosomes. In this paper, the authors analyzed genome-wide predicted target genes of the miRNAs members within each cluster. They found that the ratio of overlapped target genes between miRNAs member in the clusters is significantly higher than that of random sampled miRNAs pairs, which indicates miRNA members within the same cluster tend to have more common target genes. Detailed analysis of the binding site between miRNAs and the target genes indicate that there are two types of binding sites: one is that two miRNAs within the same cluster competitively bind to the same binding site in the same target mRNAs; the other is that two miRNAs within the same cluster synergically bind to different binding sites in the same target mRNAs.

**Key Words:** microRNA; pre-miRNA; Clustered miRNAs; microRNA genomics

---

**Received:** Mar 11, 2007

**Corresponding author:** XU Jian-zhen, Tel: +86(20)86092216, E-mail: [jianzxu@hotmail.com](mailto:jianzxu@hotmail.com)