中国科学: 生命科学 2011 年 第 41 卷 第 10 期: 958 ~ 964
SCIENTIA SINICA Vitae www.scichina.com life.scichina.com
《中国科学》杂志社
SCIENCE CHINA PRESS

论 文

人源 microRNA 前体的全基因组预测

应晓敏[①†], 朱娟娟[②†], 王小磊[③], 赵东升[③], 付汉江[②], 郑晓飞[②*], 李伍举[①*]

① 军事医学科学院基础医学研究所计算生物学中心, 北京 100850;
② 军事医学科学院放射与辐射医学研究所, 北京 100850;
③ 军事医学科学院卫生勤务与医学情报研究所, 北京 100850
† 同等贡献
* 联系人, E-mail: zhengxf@nic.bmi.ac.cn; liwj@nic.bmi.ac.cn

收稿日期: 2011-07-20; 接受日期: 2011-08-09
国家自然科学基金(批准号: 30500105 和 31071157)、国家重大科学研究计划(批准号: 2010CB912801)和国家科技重大专项"艾滋病和病毒性肝炎等重大传染病防治"(批准号: 2008ZX10002-016)资助项目
doi: 10.1360/052011-614

摘要 microRNA(miRNA)是一类不编码蛋白的调控小分子 RNA, 在真核生物中发挥着广泛而重要的调控功能. 由于miRNA的表达具有时空特异性, 因而通过计算方法预测miRNA而后有针对性的实验验证是 miRNA 发现的一条重要途径. 降低假阳性率是 miRNA 预测方法面临的重要挑战. 本研究采用集成学习方法构建预测 miRNA 前体的分类器 SVMbagging, 对训练集、测试集和独立测试集的结果表明, 本研究的方法性能稳健、假阳性率低, 具有很好的泛化能力, 尤其是当阈值取 0.9 时, 特异性高达 99.90%, 敏感性在 26%以上, 适合于全基因组预测. 采用 SVMbagging 在人全基因组中预测 miRNA 前体, 当取阈值 0.9 时, 得到 14933 个可能的 miRNA 前体. 通过与高通量小 RNA 测序数据的比较, 发现其中 4481 个 miRNA 前体具有完全匹配的小 RNA 序列, 与理论估计的真阳性数值非常接近. 最后, 对 32 个可能的 miRNA 进行实验验证, 确定其中 2 条为真实的 miRNA.

关键词
miRNA
预测
机器学习
集成学习

microRNA(miRNA)是一类长度为~22 nt 的调控小分子 RNA. 研究发现, miRNA 在生物体内发挥着重要的调控功能[1,2]. 人源 miRNA 还参与调控重大疾病的发生发展过程, 如肿瘤的发生、发展和转移[3~6]. 截止到 2011 年 4 月, 已经在动物、植物和病毒基因组中发现了共计 19724 个 miRNA, 其中人源 miRNA 为 1733 个[7]. 近年来, 高通量测序技术由于其高通量、高敏感、快速等特点, 已经成为 miRNA 发现的主要方法. 然而, 由于 miRNA 的表达具有时间、空间特异性, 部分 miRNA 只在特定组织或在特定条件下表达, 使得以高通量测序为代表的实验发现方法难以捕捉到这些 miRNA. 通过计算方法预测 miRNA 仍是一条行之有效的途径.

1 材料与方法

1.1 训练集和测试集

采用 miRBase release 9.0[8]中 391 例真实的人源 miRNA 前体(pre-miRNA, miRNA precursor)为阳性数据集. 随机抽取 300 例为阳性训练集, 余下的 91 例为

英文引用格式: Ying X M, Zhu J J, Wang X L, et al. Genome-wide prediction of human microRNA precursors. SCIENTIA SINICA Vitae, 2011, 41: 958–964, doi: 10.1360/052011-614

阳性测试集.

由于 3′ UTR 序列与 miRNA 以及绝大部分基因间区一样, 为不编码蛋白, 而且已发现的 miRNA 中只有少数几例位于 3′ UTR 区[9], 因而选择人 3′ UTR 序列作为阴性数据的来源, 3′ UTR 序列下载自 UTRdb 版本 22[10]. 通过滑窗取片段, 窗口长度为 1000 nt, 相邻窗口间重叠 150 nt. 滑窗取到的片段采用 RNAfold[11]折叠二级结构, 满足以下 4 个条件的茎环结构片段作为阴性数据集: (1) 总长度超过 55 个核苷酸; (2) 至少 18 个配对碱基对; (3) 末端环长度大于 3 个核苷酸; (4) 最低自由能(minimum free energy, MFE)≤−15 kcal/mol. 共得到 57994 例阴性样本, 随机抽取 44500 例为阴性训练集, 余下的 13494 例为阴性测试集.

miRBase release 10.0 人新增的 134 例真实 pre-miRNA 为独立阳性测试集, 随机抽取 1000 个人 19 号染色体中满足上述 4 个条件的茎环结构片段为独立阴性测试集.

人全基因组序列采用 NCBI Build 36 版本, 下载自 GenBank[12], 正负链分别滑窗取茎环结构片段, 满足上述 4 个条件的片段用于预测 miRNA.

在完成人全基因组预测 pre-miRNA 的工作时, miRBase 数据库更新到 release 17, 包含 1733 条人的 miRNA. 因此, 预测得到的可能的 pre-miRNA 与这 1733 条人源 miRNA 进行序列比较.

1.2 特征提取

采用 128 个序列和二级结构特征描述样本, 如表 1 所示.

其中, 85 个序列特征是通过编程提取得到, 42 个结构特征是采用 RNAfold[11]折叠序列后在最低自由能结构中提取得到, 随机检验 p 值是采用 randfold 程序[13]计算待考查序列与 1000 条保持二联碱基成分的随机序列的最低自由能的随机检验 p 值得到.

1.3 集成分类器的构建

采用 bagging 方法构建集成分类器 SVMbagging, 流程如图 1 所示. 从阴性训练集随机放回抽样 300 例阴性样本, 与 300 例阳性样本构建基分类器. 该过程重复 11 次, 构建 11 个基分类器. 11 个基分类器的平均值作为集成分类器的最终输出结果.

基分类器采用支持向量机(support vector machines, SVM)方法进行训练, 核函数采用径向基函数(radial basis function, RBF), 参数 C 和γ采用网格搜索方法寻优. 支持向量机方法采用 libSVM 2.83[14]实现.

1.4 与其他方法的性能比较

将集成分类器 SVMbagging 与已经发表的 6 个基于机器学习方法预测 pre-miRNA 的方法进行比较, 分别是 TripleSVM[15], miPred[16], ProMiR[17], miRabela[18], BayesMiRNAfind[9]和 MiPred[19](为与 miPred 区分, 后文称之为 MiPred_RF).

1.5 高通量测序数据与分析流程

将人全基因组预测 pre-miRNA 的结果与已发表的 3 个高通量小 RNA 测序数据进行比较. 这 3 组数据分别是 Morin 等人[20]对人胚胎干细胞长度介于 15~30 nt 的小 RNA 的高通量测序数据、Bar 等人[21]对人胚胎干细胞长度介于 18~24 nt 的小 RNA 的高通量测序数据和 Friedländer 等人[22]对人 HeLa 细胞中长度介于 20~30 nt 的小 RNA 的高通量测序数据.

对上述 3 组高通量测序数据, 先根据原文献报道的 5′和 3′ linker 序列, 去除测序数据中包含的 linker,

表 1 描述样本的 128 个特征

特征	个数	特征	个数	特征	个数
一联碱基组成	4	最小膨胀圈的大小	1	膨胀圈碱基总数	1
二联碱基组成	16	大小分别为 1~10 nt 的内部环个数	10	单链区碱基总数	1
三联碱基组成	64	≤5 nt 的内部环个数	1	最大末端环的大小	1
GC 含量	1	大小为 6~10 nt 的内部环个数	1	最小末端环的大小	1
内部环个数	1	≥11 nt 的内部环个数	1	末端环的个数	1
膨胀圈个数	1	大小分别为 1~10 nt 的膨胀圈个数	10	碱基配对数	1
单链区个数	1	≤5 nt 的膨胀圈个数	1	最低自由能	1
最大内部环的大小	1	大小为 6~10 nt 的膨胀圈个数	1	序列长度	1
最小内部环的大小	1	≥11 nt 的膨胀圈个数	1	随机检验 p 值	1
最大膨胀圈的大小	1	内部环碱基总数	1		

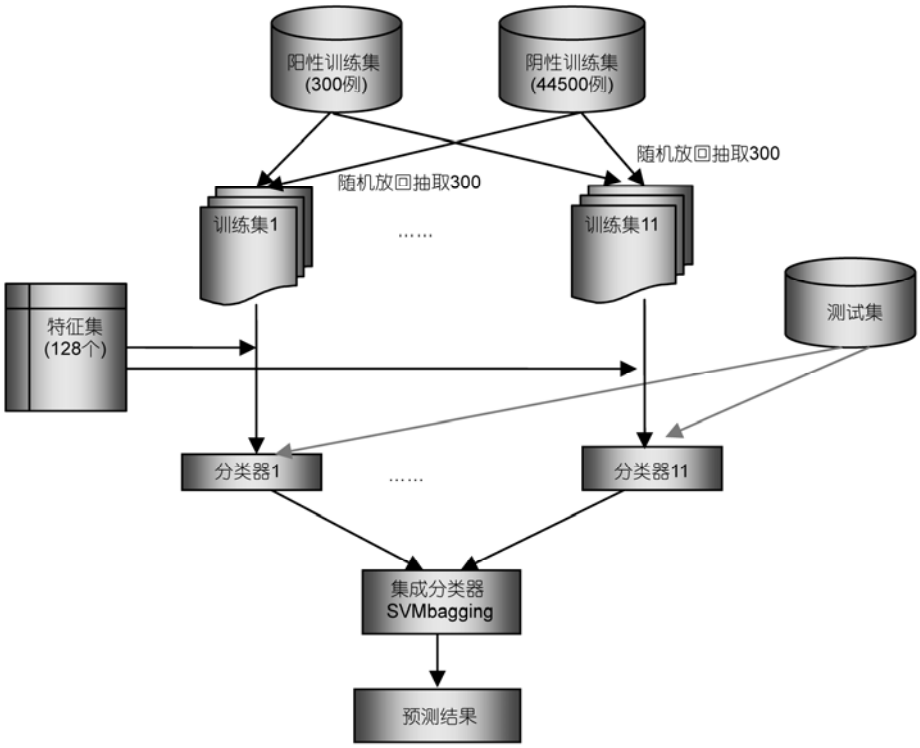

图 1　集成分类器 SVMbagging 的构建流程

而后采用 EMBOSS 软件包的 seqmatchall 函数对候选 pre-miRNA 与测序数据进行比较, wordsize 设为 15.

1.6　细胞系和质粒

HEK293 细胞由军事医学科学院二所五室保存, pMDTM18-T Vector 购自 TaKaRa 公司. HEK293 细胞用含 10%血清的 DMEM 培养基于 5% CO_2 的 37℃培养箱中培养.

1.7　试剂与材料

ImProm-Ⅱ™ 反转录酶为 Promega 公司产品, Trizol 为 Invitrogen 公司产品, polyA 加尾酶试剂盒为 Ambion 产品, PAGE 胶 DNA 回收试剂盒购自 OMEGA 公司, 20 bpDNA marker 购自 TakaRa 公司.

1.8　实验方法

提取 HEK293 细胞的总 RNA, 利用 miRNA 加尾技术和 RT-PCR 将其反转录成 cDNA; 利用所给的预测序列设计引物, Q-PCR 扩增产物, 将扩增好的扩增产物进行 PAGE 电泳, 用 PAGE 胶 DNA 回收试剂盒回收产物, 产物大小约 75 bp; 利用 TA 克隆技术, 将所回收的 DNA 插入到 pMDTM18-T vector. 菌落鉴定, 送阳性结果测序; 利用 boiedit 软件分析测序结果. 反转录的引物为:

QmiR-RT: GCGAGCACAGAATTAATACGAC-TCACTATAGG(T)18VN

PCR 扩增 3′通用引物为:

QmiR-3′: GCGAGCACAGAATTAATACGAC

详细实验操作参见附录: 实验材料和方法.

2　结果

2.1　集成分类器 SVMbagging 的性能

通过随机有放回抽取阴性样本、采用 SVM 方法构建基分类器, 得到由 11 个基分类器组成的集成分类器 SVMbagging. 该方法对训练集、测试集和独立测试集的性能如表 2 所示.

当阈值取 0.5 时, SVMbagging 对训练集的敏感性和特异性分别为 98.00%和 96.20%, 对测试集的敏感性和特异性分别为 79.12%和 96.12%, 对独立测试集的敏感性和特异性分别为 73.88%和 96.10%; 当阈值取 0.9 时, SVMbagging 对训练集的敏感性和特异性分

别为 48.67%和 99.90%, 对测试集的敏感性和特异性分别为 36.26%和 99.95%, 对独立测试集的敏感性和特异性分别为 26.12%和 99.90%. 通过这个结果可以看出, SVMbagging 对训练集、测试集和独立测试集的特异性非常接近; 而且当阈值取 0.9 时, 特异性均在 99.90%以上, 说明该方法的特异性很高, 而且泛化能力很好.

2.2　与其他方法的比较

将集成分类器 SVMbagging 与其他 6 种基于机器学习方法识别 pre-miRNA 的方法进行了比较. 图 2 是 SVMbagging 和其他 6 个分类器在独立阳性和独立阴性测试集中的 ROC 曲线. 从图 2 可以看出, SVMbagging 在独立测试集上的性能要优于其他 6 种方法, 尤其是当假阳性率低于 0.006 时, SVMbagging 的敏感性要远高于其他 6 种方法.

2.3　全基因组预测

采用 SVMbagging 在人全基因组序列中预测 pre-miRNA. 对人全基因组序列的正负链分别滑窗取片段折叠二级结构(详见“材料与方法 1.1”), 共得到 10928243 个满足条件的茎环结构片段. 采用 SVMbagging 对 10928243 个片段进行预测, 当取阈值 0.9 时, 14933 个片段被预测为可能的 pre-miRNA.

将 14933 个可能的 pre-miRNA 与 miRBase release 17 中的 1733 条人源 miRNA 进行了序列比对, 发现 687 个可能的 pre-miRNA 与已知 miRNA 完全匹配, 涉及 423 条已知的 miRNA, 占已知 miRNA 总数的 24.41%. 这个敏感性数值与 SVMbagging 对独立测试集的敏感性(26.12%)非常接近, 表明 SVMbagging 的敏感性稳健, 对 miRNA 预测性能的波动很小.

2.4　与高通量测序数据的比较

将预测的 pre-miRNA 与 3 组已经发表的高通量小 RNA 测序数据进行了序列比较. 发现 4316 个可能的 pre-miRNA 在 Morin 等人[20]的数据中有完全匹配的小 RNA 序列, 1553 个可能的 pre-miRNA 在 Bar 等人[21]的数据中有完全匹配的小 RNA 序列, 496 个可能

表 2　集成分类器 SVMbagging 的性能

阈值	精度(%)					
	训练集		测试集		独立测试集	
	阳性(300)	阴性(44500)	阳性(91)	阴性(13494)	阳性(134)	阴性(1000)
0.5	98.00	96.20	79.12	96.12	73.88	96.10
0.9	48.67	99.90	36.26	99.95	26.12	99.90

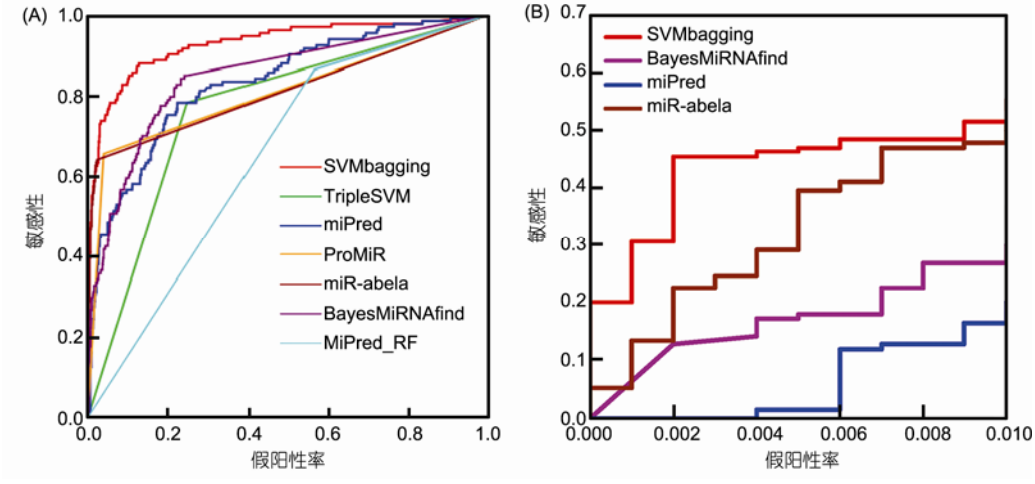

图 2　SVMbagging 与其他 6 种方法在独立测试集上的 ROC 曲线

(A) 假阳性率在 0~1 范围内的 ROC 曲线，ProMiR, TripleSVM 和 MiPred_RF 3 种方法由于仅给出是否为 pre-miRNA 的二值判断，因而不是连续变化的曲线，而是折线. (B) 假阳性率在 0~0.01 范围内的局部 ROC 曲线, 由于 ProMiR, TripleSVM 和 MiPred_RF 在该假阳性率范围内性能低于其他方法, 且没有连续变化的阈值, 因而该 3 种方法没有绘出

的 pre-miRNA 在 Friedländer 等人[22]的数据中有完全匹配的小 RNA 序列. 共计 4481 个可能的 pre-miRNA 在这 3 组高通量小 RNA 测序数据中有完全匹配的小 RNA 序列.

2.5 实验验证

挑选了 32 条在 3 组高通量测序数据中均存在完全匹配的可能的 pre-miRNA 进行了实验验证. 其中 15, 16, 35, 40, 42 和 45 号候选 miRNA 扩增效果较好, 经测序分析, 发现其中 4 条测序得到的序列比提供的候选 miRNA 长, 但全长序列跟候选 miRNA 对应的前体不一致; 只有 15 号和 42 号测序得到的序列与候选 miRNA 一致, 且在基因组中只定位在预测的 pre-miRNA 的位置上, 因而判断其为真实的 miRNA. 15 号和 42 号菌落鉴定结果如图 3 所示, 序列如下:

>候选_miRNA_15
AUCCCCAGAUACAAUGGACAAU
>候选_miRNA_42
UUUGGGACUGAUCUUGAUGUCUGC

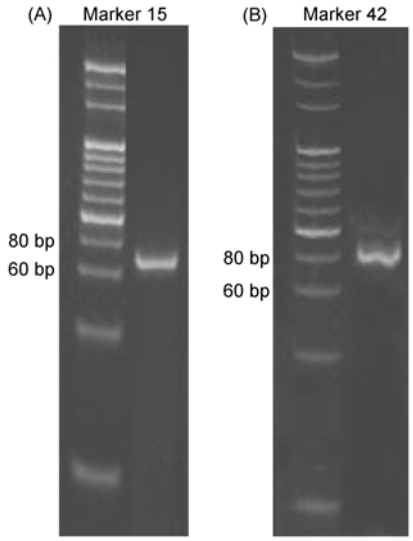

图 3 菌落鉴定结果

(A) 为 15 号候选 miRNA 的鉴定结果; (B) 为 42 号候选 miRNA 的鉴定结果

3 讨论

由于 miRNA 表达的时空特异性, 通过计算预测 miRNA 而后实验验证仍然是重要的途径. 计算预测 miRNA 面临的巨大问题是假阳性率高, 使得预测结果中存在大量的假阳性, 给实验验证带来困难. 降低假阳性率、提高特异性是 miRNA 预测方法面临的重要挑战. 为解决这一问题, 采用集成学习来构建预测 miRNA 前体的分类器, 对测试集和独立测试集的结果表明, 本研究的方法性能稳健, 具有很好的泛化能力. 尤其是当阈值取 0.9 时, 特异性高达 99.90%, 敏感性在 26%以上, 适合于全基因组预测.

到目前为止, 尽管很多 miRNA 预测方法被提出(参见文献[23]), 但系统地对人源 miRNA 进行全基因组预测的工作还只有 Bentwich 等人[24]于 2005 年发表的工作. 该研究基于有向图寻找最优分割路径, 在人全基因组中预测出了~5300 个可能的 miRNA, 并通过实验验证发现了 89 个新的 miRNA. Bentwich 等人的工作重点在发现新的人源 miRNA, 其方法不能公开使用, 也没有提供预测出的可能的 miRNA. 我们的工作重点在于提出适于基因组水平预测 miRNA 的方法, 预测方法 SVMbagging 和预测结果可以公开使用(http://ccb.bmi.ac.cn/miRNAprediction/).

通过实验验证, 确定预测的 2 条 miRNA 为真实的 miRNA. 经在最新版本的数据库 miRBase release 17 中比对发现, 这两条 miRNA 在实验验证和撰写论文的过程中被报道为真实的 miRNA, 分别是 hsa-miR-2355-5p 和 hsa-miR-3913-5p. 测得的序列比报道的序列在 3′端分别长 1 和 2 个碱基.

本研究共对 32 条预测的 miRNA 进行了验证, 确定其中 2 条为真实的 miRNA. 从验证的比例上看, 这一数值(2/32)较低. 分析主要有 3 个因素影响了验证的比例: (1) 选取的 32 个候选 miRNA 是在人胚胎干细胞和 HeLa 细胞系中表达的, 这些候选 miRNA 在 HEK293 细胞系中不一定也表达; (2) 即使候选 miRNA 在 HEK293 细胞系中表达, 如果表达量很低, 本研究的实验验证方法也可能检测不到; (3) 由于高通量测序技术在 miRNA 发现中的广泛应用, 目前尚未被发现的 miRNA 更多的应该是以时空特异表达的方式存在, 在常规细胞系中发现新 miRNA 的可能性应该会很低.

根据 SVMbagging 对独立测试集的性能, 估计在预测的 14933 个可能的 pre-miRNA 中, 10928 个为假阳性(10928243×0.001), 余下的 4005 个为真阳性, 这一数值与小 RNA 转录谱的比对结果(4481 例)非常接近, 说明 SVMbagging 的性能稳健. 根据 SVMbagging 对独立测试集和对 miRBase release 17 的性能, 估计人源

miRNA 的总数在 15000~16400 之间(4005/0.2612~4005/0.2441). 截止到2011年4月, 在人基因组中仅发现了 1733 个 miRNA, 推测还有大量的 miRNA 尚未被发现. 这些 miRNA 很可能是时空特异表达的, 因而采用常规实验方法或者高通量测序技术难以发现.

参考文献

1 Bartel D P. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell, 2004, 116: 281–297
2 Seila A C, Sharp P A. Small RNAs tell big stories in Whistler. Nat Cell Biol, 2008, 10: 630–633
3 Voorhoeve P M, le Sage C, Schrier M, et al. A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors. Cell, 2006, 124: 1169–1181
4 Tavazoie S F, Alarcon C, Oskarsson T, et al. Endogenous human microRNAs that suppress breast cancer metastasis. Nature, 2008, 451: 147–152
5 Yu F, Yao H, Zhu P, et al. let-7 regulates self renewal and tumorigenicity of breast cancer cells. Cell, 2007, 131: 1109–1123
6 Huang Q, Gumireddy K, Schrier M, et al. The microRNAs miR-373 and miR-520c promote tumour invasion and metastasis. Nat Cell Biol, 2008, 10: 202–210
7 Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res, 2011, 39: D152–D157
8 Griffiths-Jones S, Saini H K, van Dongen S, et al. miRBase: tools for microRNA genomics. Nucl Acids Res, 2008, 36: D154–D158
9 Yousef M, Nebozhyn M, Shatkay H, et al. Combining multi-species genomic data for microRNA identification using a Naive Bayes classifier. Bioinformatics, 2006, 22: 1325–1334
10 Mignone F, Grillo G, Licciulli F, et al. UTRdb and UTRsite: a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. Nucleic Acids Res, 2005, 33: D141–D146
11 Hofacker I L, Fontana W, Stadler P F, et al. Fast folding and comparison of RNA secondary structures. Monatsh Chem, 1994, 125: 167–188
12 Benson D A, Karsch-Mizrachi I, Lipman D J, et al. GenBank. Nucleic Acids Res, 2011, 39: D32–D37
13 Bonnet E, Wuyts J, Rouze P, et al. Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences. Bioinformatics, 2004, 20: 2911–2917
14 Chang C C, Lin C J. LIBSVM: a Library for Support Vector Machine. 2001. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm
15 Xue C, Li F, He T, et al. Classification of real and pseudo microRNA precursors using local structure-sequence features and support vector machine. BMC Bioinformatics, 2005, 6: 310
16 Ng K L S, Mishra S K. *De novo* SVM classification of precursor microRNAs from genomic pseudo hairpins using global and intrinsic folding measures. Bioinformatics, 2007, 23: 1321–1330
17 Nam J W, Shin K R, Han J, et al. Human microRNA prediction through a probabilistic co-learning model of sequence and structure. Nucleic Acids Res, 2005, 33: 3570–3581
18 Sewer A, Paul N, Landgraf P, et al. Identification of clustered microRNAs using an ab initio prediction method. BMC Bioinformatics, 2005, 6: 267
19 Jiang P, Wu H, Wang W, et al. MiPred: classification of real and pseudo microRNA precursors using random forest prediction model with combined features. Nucl Acids Res, 2007, 35: W339–W344
20 Morin R D, O'Connor M D, Griffith M, et al. Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome Res, 2008, 18: 610–621
21 Bar M, Wyman S K, Fritz B R, et al. MicroRNA discovery and profiling in human embryonic stem cells by deep sequencing of small RNA libraries. Stem Cells, 2008, 26: 2496–2505
22 Friedländer M R, Chen W, Adamidi C, et al. Discovering microRNAs from deep sequencing data using miRDeep. Nat Biotechnol, 2008, 26: 407–415
23 侯妍妍, 应晓敏, 李伍举. microRNA 计算发现方法的研究进展. 遗传, 2008, 30: 687–696
24 Bentwich I, Avniel A, Karov Y, et al. Identification of hundreds of conserved and nonconserved human microRNAs. Nat Genet, 2005, 37: 766–770

Genome-wide Prediction of Human microRNA Precursors

YING XiaoMin[1], ZHU JuanJuan[2], WANG XiaoLei[3], ZHAO DongSheng[3], FU HanJiang[2], ZHENG XiaoFei[2] & LI WuJu[1]

1 Center of Computational Biology, Institute of Basic Medical Sciences, Academy of Military Medical Sciences, Beijing 100850, China;
2 Institute of Radiation Medicine, Academy of Military Medical Sciences, Beijing 100850, China;
3 Institute of Health Administration and Medicine Information, Academy of Military Medical Sciences, Beijing 100850, China

microRNAs (miRNAs) are a class of small regulatory non-coding RNAs. They are involved in diverse pathways and play important roles in gene regulation in eukaryotes. Since the expression of miRNAs is spatial and temporal-specific, computational prediction followed by experimental validation is still an important approach in miRNA discovery. Decreasing false positive ratio is very challenging in miRNA prediction. Here we employed ensemble learning to construct the classifier SVMbagging to predict miRNA precursors. The results of the training, test and independent test datasets demonstrate that SVMbagging is robust, has low false positive ratio and good generalization ability. Especially when the threshold is 0.9, the specificity of SVMbagging is as high as 99.90% and the sensitivity is higher than 26.00%. Therefore, SVMbagging is suitable for genome-wide prediction. We applied SVMbagging in genome-wide prediction of miRNA precursors in human genome. We obtained 14933 candidate miRNA precursors at the threshold of 0.9. Among them, 4481 miRNA precursors have perfect matches with small RNA reads when aligning with three groups of small RNA datasets from high-throughput sequencing technologies. The number of miRNAs is very close to the true positives estimated theoretically according to the performance of SVMbagging. Finally, we applied experimental methods to validate 32 candidate miRNAs. Two of them were confirmed to be true miRNAs.

miRNA, prediction, machine learning, ensemble learning

doi: 10.1360/052011-614

论 文

附录

1 实验材料

1.1 细胞系和质粒

293 细胞由军科院二所五室保存, pMD^{TM}18-T Vector 购自 TaKaRa 公司.

1.2 主要试剂与材料

ImProm-Ⅱ^{TM} 反转录酶为 Promega 公司产品, Trizol 为 Invitrogen 公司产品, polyA 加尾酶试剂盒为 Ambion 产品, PAGE 胶 DNA 回收试剂盒购自 OMEGA 公司, 20 bpDNA marker 购自 TakaRa 公司.

1.3 引物序列

>candidate_miRNA_1
TTTAGTAGAGACGGGGTTT
>candidate_miRNA_7
GGCATGATCTCGGCTCAC
>candidate_miRNA_9
TGGGATTACAGGCGTGA
>candidate_miRNA_10
TCTCAGGAGTAAAGACAGAGTT
>candidate_miRNA_14
GAGGCAGGAGAATCGCT
>candidate_miRNA_15
ATCCCCAGATACAATGGACAAT
>candidate_miRNA_16
ATTGTCCTTGCTGTTTGG
>candidate_miRNA_17
TGGGACTACAGGCATGC
>candidate_miRNA_18
TAGTGGATGATGGAGACTCGG
>candidate_miRNA_20
AGGCAGGAGAATCACTTGAACC
>candidate_miRNA_22
GCTGGGATTATAGGCATGA
>candidate_miRNA_24
GCCCAGGCTGGAGTGCAATGG
>candidate_miRNA_27
TCCTGTGCTCTCCTGTCCT
>candidate_miRNA_28
CTGGAGTGTAGTGGCA
>candidate_miRNA_30
CAAAAACCGTGATTACTTTTGC
>candidate_miRNA_31
CAAAAGTAATTGCGGTC
>candidate_miRNA_32
GGAGGCAGAGGTTGCAG
>candidate_miRNA_33
CAGGCTGGTCTCGAACTC
>candidate_miRNA_34
GGGATTACAGGTGTGAGCCACCA
>candidate_miRNA_35
AAAAACCACAATTACTTTTGC
>candidate_miRNA_36
CACAGCAAGTGTAGACAGGCA
>candidate_miRNA_37
AAGGGCTTCCTCTCTGCAGGAC
>candidate_miRNA_38
GCTGGGACTACAGGCGTG
>candidate_miRNA_39
GTTGGTCAGGCTGGTCT
>candidate_miRNA_40
AAAAGTAATTGTGGTTTTTGC
>candidate_miRNA_41
AAAAACCACAATTACTTTTGCACCA
>candidate_miRNA_42
TTTGGGACTGATCTTGATGTCT
>candidate_miRNA_43
AGGCAGTGTATTGCTAGCGGCTG
>candidate_miRNA_44
AGAGTAGCCACTAGCCACATGT
>candidate_miRNA_45
TATGTGCCTAGTGGCTG
>candidate_miRNA_46
AGCTTTTGGGAATTCAGGTAG
>candidate_miRNA_47
AAGCAATACTGTTACCTGAAAT
GAPDH, PCR Primer(245bp)鉴定引物:
5′-TCAGTGGTGGACCTGACCTG-3′
5′-TGCTGTAGCCAAATTCGTTG-3′
反转录引物为 QmiR-RT:

GCGAGCACAGAATTAATACGACTCACTATAGG(T)18VN

PCR 扩增 3′通用引物为 QmiR-3′:

GCGAGCACAGAATTAATACGAC

2 实验方法和步骤

2.1 细胞培养

293 细胞用含 10%血清的 DMEM 培养基于 5%CO_2 的 37℃培养箱中培养.

2.2 总 RNA 的提取

Trizol 提 RNA:

用吸液器提取培养液, 加 1 mL Trizol/孔, 用枪吹打后收集到无 RNase 酶的 EP 管中, 震荡混匀室温放置 5 min;

加 0.2 mL 氯仿, 剧烈震荡 15 s, 室温放置 5 min, 12000 r/min, 4℃, 离心 15 min;

小心吸取上清(不要吸到中间层)于新的无 RNase 酶的 EP 管中, 加等体积异丙醇, 涡旋混匀, 室温放置 10 min, 12000 r/min, 4℃离心 10 min;

留沉淀, 加 1 mL 75%乙醇, 颠倒数次, 7500 r/min, 4℃离心 5 min;

去上清, 高速离心后彻底吸干乙醇, 敞盖放置 3 min, 加 20 μL 的 DEPC 水

测 OD 值

2.3 总 RNA 的 Poly(A)加尾

取 10 μg 总 RNA, 按如下体系对总 RNA 进行加尾:

5×Poly(A) 聚合酶缓冲液	10 μL
$MnCl_2$	5 μL
ATP(100 mmol/L)	0.5 μL
Poly(A)聚合酶	1 μL

加 DEPC 水补充至 50 μL. 水浴 37℃反应 60 min, 补充 150 μL DEPC 水, 加入等体积水饱和酚/氯仿/异戊醇(25∶24∶1), 震荡混匀, 4℃离心 12000 r/min×10 min. 小心吸上清(不要吸到中间层)于新的无 RNA 酶离心管中, 加入等体积氯仿/异戊醇(24∶1), 震荡混匀, 4℃离心 12000 r/min×10 min. 将上清小心吸到新的离心管中, 加入 1/10 体积的 3 mol/L NaAc(pH 5.2)混匀, 加入 2.5 倍体积无水乙醇, –20℃放置 60 min. 4℃离心 12000 r/min×15 min 后弃上清, 加入 1 mL 75%乙醇洗涤沉淀, 4℃离心 7500 r/min×5min.去上清, 敞盖室温放置晾干, 加入 20 μL DEPC 水溶解沉淀.

2.4 cDNA 合成

取 2 μg RNA 样品, 加入反转录引物(对于 mRNA 的反转录加入 1 μg oligo dT, 对于 miRNA 的反转录加入 2 μg miRNA 反转录引物), 用 DEPC 水补齐至 10 μL, 70℃变性 5 min, 置于冰上冷却 5 min.与此同时在一个新管中配制如下体系:

DEPC 处理水	12.2 μL
5×ImProm-Ⅱ™ Reverse Buffer	8 μL
$MgCl_2$ (25 mmol/L)	4.8 μL
dNTP(10 mmol/L)	2 μL
RNA 酶抑制剂(40 U/μL)	1 μL
ImProm-Ⅱ™ Reverse Transcriptase	2 μL

将此 30 μL 体系加入到上面的 10 μL 体系中, 混匀, 42℃反应 60 min, 70℃加热 15 min(灭活酶活性). cDNA 于–20℃保存, 作为实时定量 PCR 的模板.

2.5 实时定量 PCR 扩增 miRNA

配制 40 μL 反应体系如下:

2×uantiTect SYBR Green PCR Master Mix	20 μL
miRNA 特异性引物(10 μmol/L)	1.28 μL
3′公用引物(10 μmol/L)	1.28 μL
cDNA 模板	1.6 μL
无菌水	15.84 μL

混匀后分至两管, 每管 20 μL, 一个样品平行检测两管 PCR 反应条件如下: 95℃变性(酶去化学修饰)15 min, 然后(95℃ 15 s, 56℃ 30 s, 68℃ 30 s)40 个循环, 最后进行溶解曲线分析.同时扩增 U6 作为内参, 用$\Delta\Delta C_t$方法处理实时定量数据, 计算 miRNA 的表达变化.

2.6 PAGE 胶电泳和回收

利用所给的预测序列设计引物, Q-PCR 扩增产物, 挑选溶解曲线, 扩增曲线数值较好的扩增产物(15 号, 16 号, 35 号, 40 号, 42 号, 45 号), PAGE 电泳, 用 PAGE 胶 DNA 回收试剂盒回收产物, 产物大小约 75 bp.

PAGE 胶 DNA 回收步骤:

(1) 取一个新的 HiBind DNA 结合柱装在收集管中, 吸取适量体积的 Buffer GPS 平衡缓冲液至柱子中

(2) 室温放置 3~5 min

(3) 12000×*g* 2 min

(4) 弃滤液, 加 700 μL 无菌水至柱子中, 12000×*g*, 2 min.

(5) 向切下胶的 EP 管中加入 250 μL 的 Poly-Gel DNA Eution Buffer, 65 度 4H.

(6) 把溶胶液和胶块一起转入 Poly-Gel filter unit, 放置在 2 mL 的 collection tube, 10000×*g*, 10 min

(7) 向溶胶液中缴入 6 倍体积的 Buffer HB

(8) 将 750 μL 的上述溶液中加入 clean HiBind DNA column, 10000×*g*, 1 min, 室温, 弃去滤液

(9) 将剩余的液体按上述步骤操作, 弃滤液

(10) 加 700 μL 的 SPW wash buffer, 10000×*g*, 1 min, 室温

(11) 空转 2 min, 10000×*g*

(12) 转移至新的 EP 管, 加 15~30 μL 无菌水, 10000×*g*, 1 min.

2.7 AT 克隆

参见 TaKaRa pMD^{TM}18-T Vector 说明书.

2.8 测序鉴定

菌落鉴定, 送阳性结果测序, 利用 boiedit 软件分析测序结果.