Enhancing miRNA annotation confidence in miRBase by continuous cross dataset analysis

Thomas B. Hansen,* Jørgen Kjems and Jesper B. Bramsen

Department of Molecular Biology; Interdisciplinary Nanoscience Center (iNANO); Aarhus University; Aarhus, Denmark

*Correspondence to: Thomas B. Hansen;
Email: tbh@mb.au.dk

The immaculate annotation of all microRNAs (miRNAs) is a prerequisite to study their biological function on a genome-wide scale. However, the original criteria for proper miRNA annotation seem unsuited for the automated analysis of the immense number of small RNA reads available in next generation sequencing (NGS) datasets. Here we analyze the confidence of past miRNA annotation in miRBase by cross-analyzing publicly available NGS datasets using strengthened annotation requirements. Our analysis highlights that a large number of annotated human miRNAs in miRBase seems to require more experimental validation to be confidently annotated. Notably, our dataset analysis also identified almost 300 currently non-annotated miRNA*s and 28 novel miRNAs. These observations hereby greatly increase the confidence of past miRNA annotation in miRBase but also illustrate the usefulness of continuous re-evaluating NGS datasets in the identification of novel miRNAs.

The discovery of microRNAs (miRNAs) has truly revealed a whole new layer of gene regulation through RNA interference (RNAi)[1,2] and sparked the development of new classes of RNA-based therapeutics. Being broadly evolutionary conserved in eukaryotes, miRNAs play key roles in most biological processes including developmental timing, stem cell differentiation and disease development.[3-6]

The exploration of miRNA biology has proven to be a most difficult task; miRNAs are believed to act in concert as fine-tuners of gene expression[7,8] and most individual miRNAs typically have modest impact on their target mRNA and protein levels.[9-11] Furthermore, due to very limited base pairing required for miRNA-target interactions each miRNA has hundreds of potential targets[12] which in effect leaves miRNA target prediction very inefficient still. To understand these complex miRNA-mRNA relationships, target-site prediction along with genome wide-expression profiling and correlation seems an attractive approach.[13-15] However, the prerequisite for such analysis is a reliable miRNA registry that allows for valid target-site prediction as well as meaningful expression profiling attempts. Hence, it is imperative that the miRBase[16-18] is a resource devoid of erroneous annotations and false positives.

Too Many miRNAs Annotated

The recently released miRBase (version 16) holds 1,040 human pre-miRNAs. This number seems ever increasing and novel miRNAs will expectably be annotated proportionally to NGS sequencing depth. Initial predictions in silico estimated the human genome to hold at least a thousand miRNA genes.[19-21] These high expectations may have negatively impacted our standards for miRNA annotation and as a result miRBase now holds annotated miRNAs with questionable validity. This concern has been put forward in several cases. Schopman et al. reported that miR-1274b and miR-1274a are presumably small RNA fragments derived from tRNA processing.[22] Also Berezikov and colleagues[23] find in their analysis of a published small RNA sequencing set in Drosophila that a

high proportion of putative miRNA candidates are likely reflecting instead the degradation of diverse mRNAs, ribosomal RNA, other noncoding RNAs (ncRNAs) or endogenous small interfering RNAs (endo-siRNAs).

miRNA Characteristics, What to Look for?

Setting up criteria for miRNA annotation is likely an ongoing process in accordance with both technical and intellectual advancements. Quite naturally, the miRNA annotation criteria employed have focused on the categorical steps in the miRNA biogenesis pathway, albeit this is not immediately straightforward. miRNAs are 20–25 nucleotides long single-stranded RNA species typically derived from long primary polymerase II (polII) transcripts (pri-miRNAs), that are processed into mature miRNAs via short hairpin precursor miRNA intermediates (pre-miRNAs).[24] miRNA biogenesis involves processing of pri-miRNAs into pre-miRNAs by the nuclear Microprocessor complex[25,26] followed by cytoplasmic Dicer cleavage of pre-miRNAs into double-standed miRNA/miRNA* duplexes.[27,28] Hereby, a hallmark in miRNA biogenesis is the exact processing of endogenous stem-loop RNA structures into a miRNA/miRNA* duplex through a two-step RNase III cleavage reaction. However, with the recent discovery of both Microprocessor-independent (miRtrons[29,30]) or Dicer-independent miRNA processing (e.g., miR-451,[31-33]), the task of establishing one set of annotation criteria encompassing all subtypes of miRNAs is thus very complicated.

A Uniform System for miRNA Annotation

This first attempt to set up a uniform system for miRNA annotation was put forward by a group of leading laboratories in 2003,[34] proposing that candidate miRNAs should fulfill both expression and biogenesis characteristics to be annotated as proper miRNAs. In essence, miRNA candidates were required to be identified as ~22-nt RNA transcripts in a size-fractionated RNA sample as distinguished by
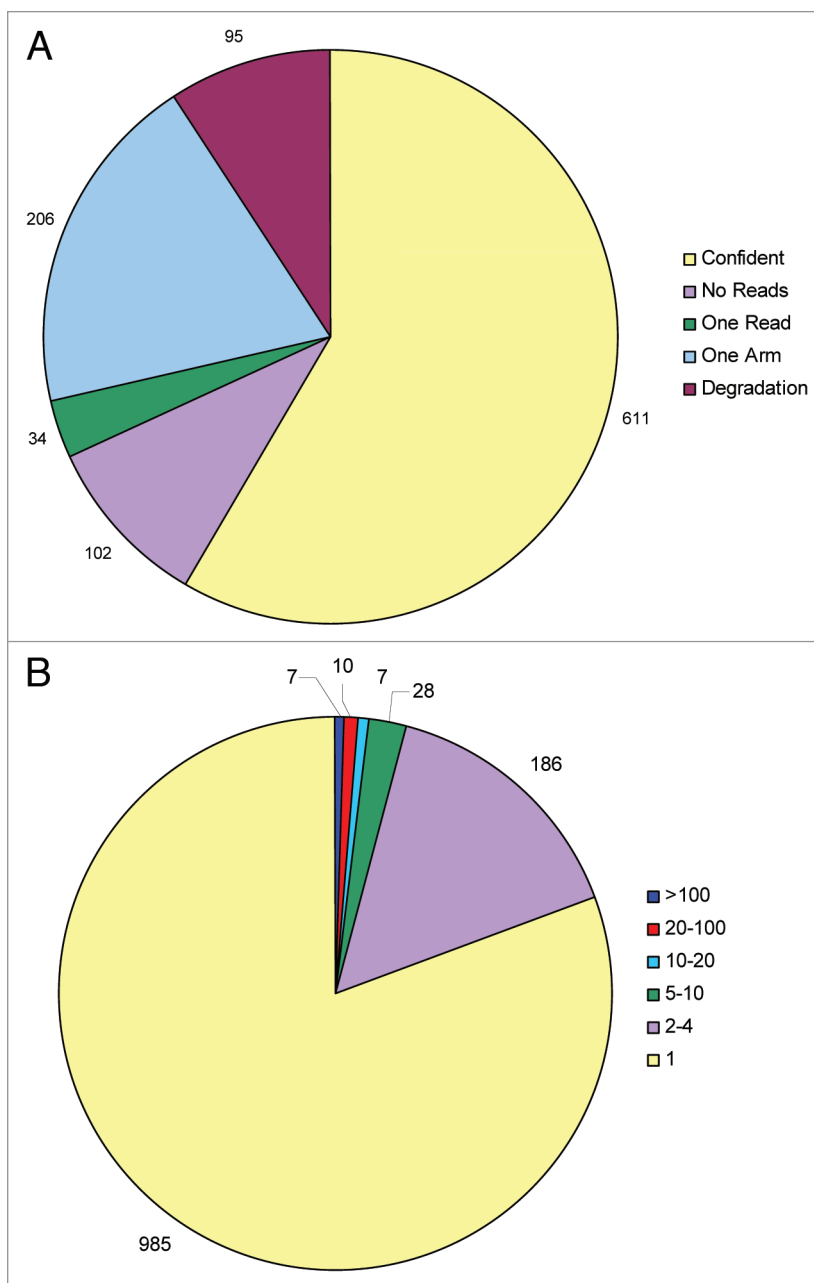

Figure 1. Grouping annotated miRNAs. (A) Based on datasets analyzed, annotated miRNAs distribute into following categories: confident, no reads, one-readers, one-armed or degradation signatures. (B) Subdividing mature miRNA sequences depending on the number of perfect matches in the human genome (hg18) using bowtie algorithm.

gel blot hybridization (northern blotting) or cDNA sequencing. Furthermore, candidate miRNAs should be contained within one arm of a predicted short-hairpin precursor satisfying rather loosely defined structural requirements. These early criteria still provide a broadly accepted standard for miRNA annotation, yet seem unsuited for the automated analysis of NGS data sets as miRNA candidates could qualify for annotation simply by identification of a single 21-mer read along with a plausible RNA structure. Considering the overwhelming depth of NGS and the extensive occurrence of pseudohairpins and miRNA-like structures in animal transcriptomes,[19] this will inevitably lead to false positives as noted above.[22,23,35]

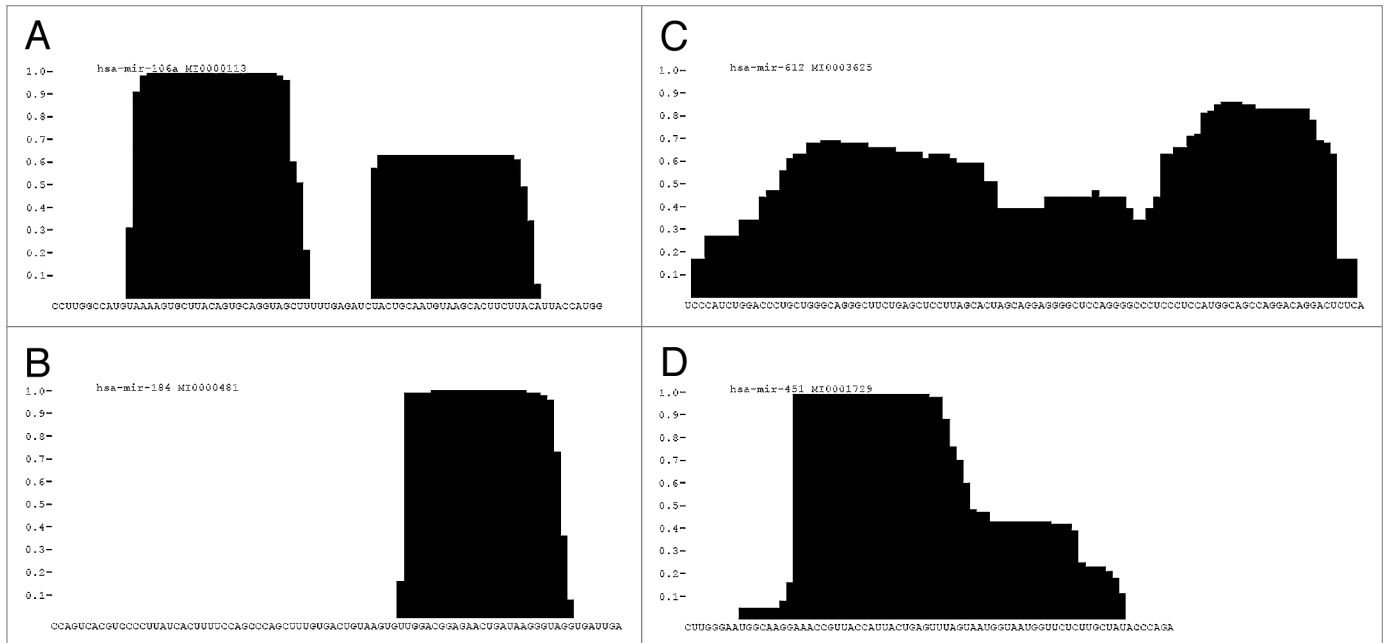To base miRNA annotation on NGS data calls for more stringent criteria. In

Figure 2. Read-plots. Representation of confident (A), one-armed (B), degradation (C) and miR-451 (D) pre-miRNA read-plot pattern compiled by plotting the log-scaled number of reads covering each nucleotide on the pre-miRNA sequence normalized to the log-scaled number of total reads matching the pre-miRNA sequence.

a recent discussion of plant miRNAs, Meyers and colleagues propose that "conclusive evidence of precise biogenesis from a qualifying stem-loop is now considered the sole criterion that is both necessary and sufficient for miRNA annotation."[36] That is, miRNA and miRNA* reads must be derived from opposing stem-arms of a single-stranded miRNA precursor hairpin forming a duplex with 2 nt 3' overhangs indicative of RNase III-enzyme processing. Even though the miRNA* strand might not be functionally incorporated into RISC, it is still an inescapable intermediate in canonical miRNA biogenesis. This precise biogenesis cleavage pattern must be convincingly reflected in the NGS reads so that sequences with highly heterogeneous 5' and 3' ends or with a non-RNase III derived signature can be disregarded.

Often, only one strand of the putative miRNA precursor is identified in datasets which may reflect both low-sequencing depth, rapid degradation of the miRNA* or Dicer/Drosha-independent miRNA biogenesis. In fact, numerous annotated miRNAs in miRBase are annotated without miRNA* strand evidence (see below) and single armed reads have been considered sufficient for miRNA annotation if

the miRNA is identified in multiple, independent libraries[36] or produces homogenous sequence reads indicative of precise nucleolytic processing.[23] We here advocate raising the stringency of miRNA annotation requirements; either the miRNA/miRNA* duplex is confidently reflected in the deep sequencing dataset or experimentally verified by miRNA candidate overexpression followed by northern blotting or luciferase reporter assays.

Re-Examination of miRBase Using Cross-Dataset Validation

The 1,040 registered human pre-miRNAs have been validated using techniques such as cDNA cloning, northern blotting and more recently NGS. Considering this lack in uniformity of miRNA annotation, we suggest that the more stringent miRNA annotation criteria of today (requirement for exact biogenesis, i.e., identification of miRNA and miRNA*) and the availability of massive NGS datasets should now be used to continuously re-evaluate past miRNA annotations. We therefore examined all pre-miRNAs using publicly available datasets,[37-52] as well as unpublished GEO accessions; GSE21279, GSE20892, discarding reads below 18 nucleotides in

length. Based on these datasets we subdivided miRNA genes into five categories inspired by Berezikov and coworkers[23] (**Fig. 1A**). We found that 654 out of 1,040 annotated human pre-miRNAs received matching 5p and 3p hits suggestive of RNase III processing and are thus considered bona fide miRNAs (**Figs. 1A and 2A**). This coheres with the fact that 309 out of 312 miRNAs with mature and star sequence annotation in miRBase are scored as confident in our analysis. Failing the test are miR-593 and miR-1207; in either case annotation seems inconsistent with RNase III based maturation (**Fig. 3A and B**), and the miRtron, miR-1225, which only receive very limited read coverage in the datasets investigated. Notably, our dataset analysis picked up 296 miRNA* sequences (**Sup. Table 1**) that are currently not annotated in miRBase as well as re-annotated eight erroneously annotated miRNAs (**Sup. Table 2**). Our analysis hereby greatly increases the confidence of these past miRNA annotations in miRBase but also illustrates the usefulness of continuous re-evaluation of NGS datasets.

Reproducible mature sequence hitters with no star sequence reads (one-armed pre-miRNAs, **Fig. 2B**) characterized 206
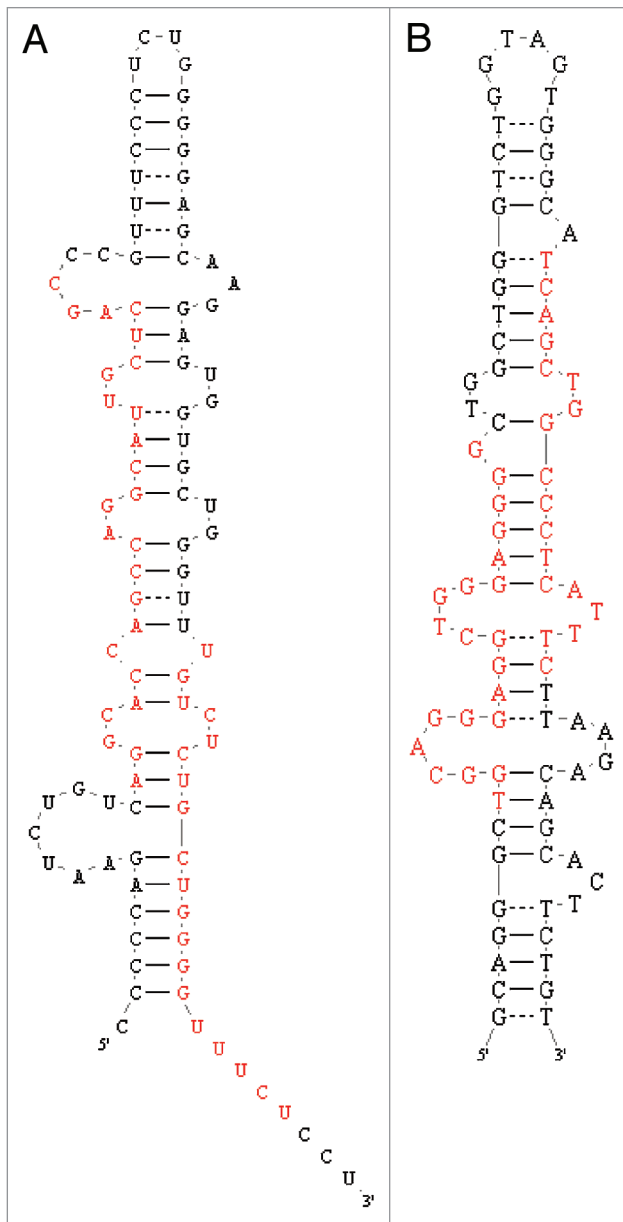
Figure 3. Annotated two-armed miRNAs not derived from two-step RNase III processing. Structural visualization of miR-593 (A) and miR-1207 (B) with annotated sequences depicted in red. Structural predictions were done by MFold and visualization by RNA Folder.

pre-miRNAs which is consistent with highly biased strand selection but evidently not sufficient to fulfil our criteria for unambiguous RNase III-derived processing. Therefore, one-armed pre-miRNAs should ideally undergo experimental validation prior to annotation. Finally, 95, 34 and 102 were classified as degradation products (**Fig. 2C**), single-hitters and no-hitters, respectively, all clearly falling short of confident miRNA annotation.

In many cases faithful miRNA annotation is complicated by the mature miRNA sequence having multiple hits in the genome (referred to as multimappers). We find that only 1/3 of the miRNAs mapping perfectly to ten or more places in the genome are scored as confident in our analysis (and none considering miRNAs lacking star-strand annotation only), suggesting that multimappers are generally less likely to abide to the confident criteria employed here. As example, the putative

mature sequence of miR-4286 has been found in several datasets, however, it gives rise to more than 100 perfect hits in the human genome, unlike the annotated pre-miR-4286 which only has a single perfect match. The miRBase has no record of any star-sequence, there is no additional evidence in the original publication[53] and no circumstantial evidence, e.g., the saddle-back pattern of conservation, exists. Thus, it seems that one structure has been picked mimicking the loosely defined miRNA-like signature out of more than 100 possible genomic regions and that simply does not suffice for miRNA annotation. Other multimappers (miR-1912, miR-3674, miR-1303, miR-1268, miR-649, miR-1302, miR-1233, miR-1285, miR-3669, etc.,) similarly crave for additional supporting evidence in order to be properly annotated as miRNAs (**Fig. 1B**).

In conclusion our analysis highlights that a number of annotated miRNAs in miRBase seem to require more experimental validation to be confidently annotated. It should be clearly emphasized however, that the miRNAs not successfully passing our analysis may not be falsely annotated and we merely underscore that these miRNAs represent candidates with insufficient evidence for miRBase submission in our eyes.

Are We Unintentionally Excluding Valid miRNAs?

The requirement for miRNA/miRNA* identification may potentially disregard functional miRNAs produced by a non-canonical biogenesis pathways independently of either Drosha and/or Dicer. miR-451 is clearly a valid miRNA, is involved in erythropoiesis,[54,55] has been described as a tumor suppressor in human glioma cells[56] but has a dicer-independent biogenesis. Instead, the pre-miRNA is taken up and cleaved by Ago2-RISC and 3' trimmed,[31-33] producing a highly heterogeneous 3' end (**Fig. 2D**). This unusual biogenesis pathway quite naturally precludes the experimental identification of miR-451* and the predicted stem-loop structure would not immediately qualify miR-451 as a classical miRNA candidate. Still, the mature miR-451 is picked up
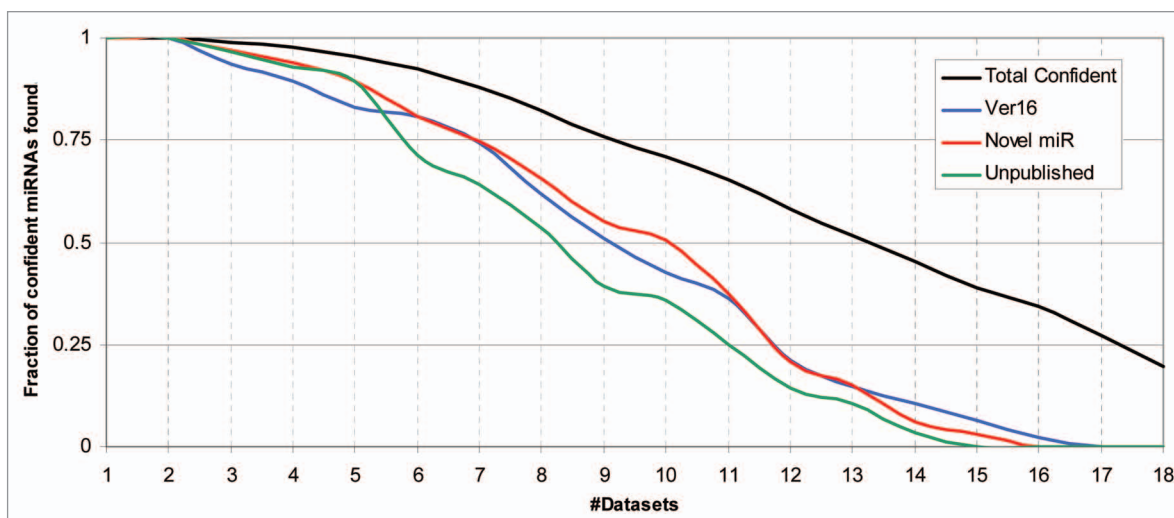
Figure 4. Cross-dataset analysis. Cumulative fraction of confident miRNAs as a function of the number of datasets wherein identified. The calculation was conducted on the confident subset miRBase accessions (n = 611), the confident subset of newly annotated miRNAs in miRBase version 16 (n = 47), the confident subset of miRNA candidates previously published (n = 67),[57] and, lastly, confident miRNAs from above list still unpublished (n = 28).

in 15 of the NGS datasets analyzed here and produces reads with consistent 5' end (**Fig. 2D**). Hereby we embrace the fact that additional Drosha/Dicer-independent miRNAs are likely to be identified, however, their annotation will require more experimental validation of their biogenesis pathways (overexpression analysis etc.) to support miRNA annotation in the absence of the miRNA* sequence. In this regard, we prefer high stringency in miRNA annotation to avoid false pre-miRNA candidates, even at the expense of sacrificing non-canonical miRNAs.

Using Cross-Dataset Analysis to Identify Novel miRNAs

Combining NGS datasets to enhance sequencing depth seems an attractive approach to help future miRNA annotation. So far miRNAs have been overlooked in individual NGS datasets, most likely due to experimental cut-off of low frequency sequence reads or lack of identifying the miRNA* strands. It has been speculated that the miRNAs to be discovered are more exclusively expressed species only found in specific tissues at distinct timepoints[53] which naturally would thwart cross-analysis with independent and unrelated datasets. To our knowledge such miRNAs have yet to be documented. In fact, newly annotated confident two-armed miRNAs in miRBase (version 16)

were all identified in two or more independent sequencing experiments (**Fig. 4**), and more than 50% were found independently in nine or more datasets, emphasizing the relevance of cross-examination. Likewise, we recently published 11 novel miRNAs out of 112 candidates based on both publicly available deep sequencing datasets and overexpression analysis.[57] Sixty-seven of the 112 miRNA candidates obtain a confident score when subjected to cross-examination, 34 of which are picked up in 10 or more datasets (**Fig. 4**) and 39 were published in miRBase by us and others. This leaves 28 confident, however unpublished, miRNAs in our list of candidates, and surprisingly, almost half of those are seen in nine or more sets (**Sup. Table 3 and Fig. 4**), which once again illustrates the advantages of cross-dataset analysis.

Concluding Thoughts

It is strictly required that we as researchers exercise greater care during submission and extraction of miRNA annotations to and from miRBase. In essence this requires establishing miRNA annotation criteria that can be implemented by most researchers, yet conclusively distinguishes precise miRNA biogenesis from experimental noise in NGS data sets. We support the notion that both strands of the miRNA duplex should preferentially be identified in NGS datasets and additional

experimental verification is needed upon lack of miRNA* reads. We acknowledge that a subset of bona fide miRNAs rely on non-canonical biogenesis and thus do not produce a miRNA* sequence, however in this case prompting for additional experimental evidence seems a reasonable request at this point. Furthermore, we propose to set up a "Candidate miRBase" in addition to a miRBase, where putative miRNA candidates be uploaded along with the supporting NGS data and subsequently, when sufficient evidence exists, be transferred to the "confident" miRBase.

Note

Supplemental materials can be found at: www.landesbioscience.com/journals/rnabiology/article/14333

References

1. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature 1998; 391:806-11.

2. Lee RC, Feinbaum RL, Ambros V. The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 1993; 75:843-54.

3. Ambros V. The functions of animal microRNAs. Nature 2004; 431:350-5.

4. Hannon GJ. RNA interference. Nature 2002; 418:244-51.

5. Chang TC, Mendell JT. microRNAs in vertebrate physiology and human disease. Annu Rev Genomics Hum Genet 2007; 8:215-39.

6. Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. Identification of novel genes coding for small expressed RNAs. Science 2001; 294:853-8.

7. Krek A, Grun D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, et al. Combinatorial microRNA target predictions. Nat Genet 2005; 37:495-500.

8. Wu S, Huang S, Ding J, Zhao Y, Liang L, Liu T, et al. Multiple microRNAs modulate p21$^{Cip1/Waf1}$ expression by directly targeting its 3' untranslated region. Oncogene 2010; 29:2302-8.

9. Ivanovska I, Cleary MA. Combinatorial microRNAs: Working together to make a difference. Cell Cycle 2008; 7:3137-42.

10. Baek D, Villen J, Shin C, Camargo FD, Gygi SP, Bartel DP. The impact of microRNAs on protein output. Nature 2008; 455:64-71.

11. Selbach M, Schwanhausser B, Thierfelder N, Fang Z, Khanin R, Rajewsky N. Widespread changes in protein synthesis induced by microRNAs. Nature 2008; 455:58-63.

12. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 2005; 120:15-20.

13. Gennarino VA, Sardiello M, Avellino R, Meola N, Maselli V, Anand S, et al. MicroRNA target prediction by expression analysis of host genes. Genome Res 2009; 19:481-90.

14. Wang X, Wang X. Systematic identification of microRNA functions by combining target prediction and expression profiling. Nucleic Acids Res 2006; 34:1646-52.

15. Huang JC, Babak T, Corson TW, Chua G, Khan S, Gallie BL, et al. Using expression profiling data to identify human microRNA targets. Nat Methods 2007; 4:1045-9.

16. Griffiths-Jones S. The microRNA registry. Nucleic Acids Res 2004; 32:109-11.

17. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: MicroRNA sequences, targets and gene nomenclature. Nucleic Acids Res 2006; 34:140-4.

18. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. miRBase: Tools for microRNA genomics. Nucleic Acids Res 2008; 36:154-8.

19. Bentwich I, Avniel A, Karov Y, Aharonov R, Gilad S, Barad O, et al. Identification of hundreds of conserved and nonconserved human microRNAs. Nat Genet 2005; 37:766-70.

20. Sheng Y, Engstrom PG, Lenhard B. Mammalian microRNA prediction through a support vector machine model of sequence and structure. PLoS One 2007; 2:946.

21. Yousef M, Showe L, Showe M. A study of microRNAs in silico and in vivo: Bioinformatics approaches to microRNA discovery and target identification. FEBS J 2009; 276:2150-6.

22. Schopman NC, Heynen S, Haasnoot J, Berkhout B. A miRNA-tRNA mix-up: TRNA origin of proposed miRNA. RNA Biol 2010; 7.

23. Berezikov E, Liu N, Flynt AS, Hodges E, Rooks M, Hannon GJ, Lai EC. Evolutionary flux of canonical microRNAs and mirtrons in drosophila. Nat Genet 2010; 42:6-9.

24. Kim VN, Han J, Siomi MC. Biogenesis of small RNAs in animals. Nat Rev Mol Cell Biol 2009; 10:126-39.

25. Denli AM, Tops BB, Plasterk RH, Ketting RF, Hannon GJ. Processing of primary microRNAs by the microprocessor complex. Nature 2004; 432:231-5.

26. Gregory RI, Yan KP, Amuthan G, Chendrimada T, Doratotaj B, Cooch N, Shiekhattar R. The microprocessor complex mediates the genesis of microRNAs. Nature 2004; 432:235-40.

27. Bernstein E, Caudy AA, Hammond SM, Hannon GJ. Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature 2001; 409:363-6.

28. Hutvagner G, McLachlan J, Pasquinelli AE, Balint E, Tuschl T, Zamore PD. A cellular function for the RNA-interference enzyme dicer in the maturation of the let-7 small temporal RNA. Science 2001; 293:834-8.

29. Ruby JG, Jan CH, Bartel DP. Intronic microRNA precursors that bypass drosha processing. Nature 2007; 448:83-6.

30. Okamura K, Hagen JW, Duan H, Tyler DM, Lai EC. The mirtron pathway generates microRNA-class regulatory RNAs in drosophila. Cell 2007; 130:89-100.

31. Cifuentes D, Xue H, Taylor DW, Patnode H, Mishima Y, Cheloufi S, et al. A novel miRNA processing pathway independent of dicer requires Argonaute2 catalytic activity. Science 2010; 328:1694-8.

32. Cheloufi S, Dos Santos CO, Chong MM, Hannon GJ. A dicer-independent miRNA biogenesis pathway that requires ago catalysis. Nature 2010; 465:584-9.

33. Yang JS, Maurin T, Robine N, Rasmussen KD, Jeffrey KL, Chandwani R, et al. Conserved vertebrate mir-451 provides a platform for dicer-independent, Ago2-mediated microRNA biogenesis. Proc Natl Acad Sci USA 2010; 107:15163-8.

34. Ambros V, Bartel B, Bartel DP, Burge CB, Carrington JC, Chen X, et al. A uniform system for microRNA annotation. RNA 2003; 9:277-9.

35. Chiang HR, Schoenfeld LW, Ruby JG, Auyeung VC, Spies N, Baek D, et al. Mammalian microRNAs: Experimental evaluation of novel and previously annotated genes. Genes Dev 2010; 24:992-1009.

36. Meyers BC, Axtell MJ, Bartel B, Bartel DP, Baulcombe D, Bowman JL, et al. Criteria for annotation of plant MicroRNAs. Plant Cell 2008; 20:3186-90.

37. Morin RD, O'Connor MD, Griffith M, Kuchenbauer F, Delaney A, Prabhu AL, et al. Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome Res 2008; 18:610-21.

38. Affymetrix ENCODE Transcriptome Project, Cold Spring Harbor Laboratory ENCODE Transcriptome Project. Post-transcriptional processing generates a diversity of 5'-modified long and short RNAs. Nature 2009; 457:1028-32.

39. Ender C, Krek A, Friedlander MR, Beitzinger M, Weinmann L, Chen W, et al. A human snoRNA with microRNA-like functions. Mol Cell 2008; 32:519-28.

40. Zhu JY, Pfuhl T, Motsch N, Barth S, Nicholls J, Grasser F, Meister G. Identification of novel epstein-barr virus microRNA genes from nasopharyngeal carcinomas. J Virol 2009; 83:3333-41.

41. Wyman SK, Parkin RK, Mitchell PS, Fritz BR, O'Briant K, Godwin AK, et al. Repertoire of microRNAs in epithelial ovarian cancer as determined by next generation sequencing of small RNA cDNA libraries. PLoS One 2009; 4:5311.

42. Taft RJ, Glazov EA, Cloonan N, Simons C, Stephen S, Faulkner GJ, et al. Tiny RNAs associated with transcription start sites in animals. Nat Genet 2009; 41:572-8.

43. Bar M, Wyman SK, Fritz BR, Qi J, Garg KS, Parkin RK, et al. MicroRNA discovery and profiling in human embryonic stem cells by deep sequencing of small RNA libraries. Stem Cells 2008; 26:2496-505.

44. Vaz C, Ahmad HM, Sharma P, Gupta R, Kumar L, Kulshreshtha R, Bhattacharya A. Analysis of microRNA transcriptome by deep sequencing of small RNA libraries of peripheral blood. BMC Genomics 2010; 11:288.

45. Cummins JM, He Y, Leary RJ, Pagliarini R, Diaz LA Jr, Sjoblom T, et al. The colorectal microRNAome. Proc Natl Acad Sci USA 2006; 103:3687-92.

46. Somel M, Guo S, Fu N, Yan Z, Hu HY, Xu Y, et al. MicroRNA, mRNA and protein expression link development and aging in human and macaque brain. Genome Res 2010; 20:1207-18.

47. Jima DD, Zhang J, Jacobs C, Richards KL, Dunphy CH, Choi WW, et al. Deep sequencing of the small RNA transcriptome of normal and malignant human B cells identifies hundreds of novel microRNAs. Blood 2010; 116:e118-27.

48. Liao JY, Ma LM, Guo YH, Zhang YC, Zhou H, Shao P, et al. Deep sequencing of human nuclear and cytoplasmic small RNAs reveals an unexpectedly complex subcellular distribution of miRNAs and tRNA 3' trailers. PLoS One 2010; 5:10563.

49. Kuchen S, Resch W, Yamane A, Kuo N, Li Z, Chakraborty T, et al. Regulation of microRNA expression and abundance during lymphopoiesis. Immunity 2010; 32:828-39.

50. Mayr C, Bartel DP. Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. Cell 2009; 138:673-84.

51. Stark MS, Tyagi S, Nancarrow DJ, Boyle GM, Cook AL, Whiteman DC, et al. Characterization of the melanoma miRNAome by deep sequencing. PLoS One 2010; 5:9685.

52. Witten D, Tibshirani R, Gu SG, Fire A, Lui WO. Ultra-high throughput sequencing-based small RNA discovery and discrete statistical biomarker analysis in a collection of cervical tumours and matched controls. BMC Biol 2010; 8:58.

53. Goff LA, Davila J, Swerdel MR, Moore JC, Cohen RI, Wu H, et al. Ago2 immunoprecipitation identifies predicted microRNAs in human embryonic stem cells and neural precursors. PLoS One 2009; 4:7192.

54. Patrick DM, Zhang CC, Tao Y, Yao H, Qi X, Schwartz RJ, et al. Defective erythroid differentiation in miR-451 mutant mice mediated by 14-3-3zeta. Genes Dev 2010; 24:1614-9.

55. Rasmussen KD, Simmini S, Abreu-Goodger C, Bartonicek N, Di Giacomo M, Bilbao-Cortes D, et al. The miR-144/451 locus is required for erythroid homeostasis. J Exp Med 2010; 207:1351-8.

56. Nan Y, Han L, Zhang A, Wang G, Jia Z, Yang Y, et al. MiRNA-451 plays a role in tumor suppressor in human glioma cells. Brain Res 2010.

57. Hansen TB, Bramsen JB, Kjems J. Re-inspection of small RNA sequence datasets reveals several novel human miRNA genes. PLoS One 2010; 5:10961.