

Alternative miRNAs: Human sequences misidentified as plant miRNAs in plant studies and in human plasma

Kenneth W. Witwer

The Johns Hopkins University School of Medicine
Department of Molecular and Comparative Pathobiology
Department of Neurology

733 N. Broadway
Miller Research Building Rm 829
Baltimore MD 21205
USA

p 1-410-955-9770
f 1-410-955-9823
email: kwitwer1@jhmi.edu

Abstract

A recent study reported that “Plant miRNAs found in human circulating system provide evidences of cross kingdom RNAi”¹. Analysis of two human blood plasma sequencing datasets was said to provide evidence for uptake of plant miRNAs into human plasma. The results were also purportedly inconsistent with contamination¹. However, a review of these data suggests that they do not support dietary xenomiR uptake, but instead confirm previous findings that detection of rare plant miRNAs in mammalian sequencing datasets is artifactual. Only one putative plant miRNA (“peu-MIR2910) in this study mapped consistently above background, and this sequence is found in a human rRNA. Several other rarer but consistently mapped plant miRNAs also have 100% or near 100% matches to human transcripts or genomic sequences, and some do not map to plant genomes at all. These misidentified “alternative miRNAs”—including MIR2910 and MIR2911—emphasize the need for rigorous filtering strategies when assessing possible xenomiRNAs.

INTRODUCTION

Reports of plant or other dietary miRNAs, or xenomiRs, entering mammalian circulation through the diet²⁻⁵ generated excitement for the xenomiR transfer hypothesis, yet negative results of replication and reproduction studies have cast doubt on xenomiR transfer as a general mechanism⁶⁻¹². A prominent claim of xenomiR function² has also failed rigorous reproduction⁸, unmasked as the result of an uncontrolled variable in the original experiment. Analyses of public datasets have revealed that studies of xenomiRs and other foreign-origin nucleic acids are fraught with artifacts: combinations of contamination, amplification or sequencing errors, permissive analysis pathways, and batch effects^{9,11,13-17}. A particularly comprehensive study recently found that foreign miRNAs in human biofluids and tissues do not match human food consumption, are marked by batch effects, and are thus most parsimoniously explained as artifacts¹⁴. Studies of organisms with no exposure to plants have also found evidence of the same types of apparent plant contamination that plague some measurements of human samples^{9,18}. Liu et al¹ mapped sequencing data from two studies of human plasma and other samples^{19,20} to various plant genomes using a 2010 plant miRNA database, PMRD²¹, concluding that previous reports of dietary xenomiR transfer are supported. In this brief report, these results are examined critically.

RESULTS

Data evaluation

A cross-check of the source files and articles shows that the plasma data evaluated by Liu et al were from 198 plasma samples, not 410 as reported. Ninomiya et al sequenced six human plasma samples, six PBMC samples, and 11 cultured cell lines¹⁹. Yuan et al sequenced 192 human plasma libraries (prepared from polymer-precipitated plasma particles)²⁰. Each library was sequenced once, and then a second time to increase total reads. Counts were presented as reads per million mapped reads (rpm)²⁰. In contrast, Liu et al appear to have reported total mapped reads in their data table¹. Yuan et al also set an expression cutoff of 32 rpm (log₂ rpm of 5 or above). With an average 12.5 million reads per sample (the sum of the two runs per library), and, on average, about half of the sequences mapped, the 32 rpm cutoff would translate to around 200 total reads in the average sample as mapped by Liu et al¹.

Only one putative plant miRNA above background levels

Consulting the Liu et al mapping table¹ and the Sequence Read Archive (SRA), results from duplicate sequencing runs from the Yuan et al dataset were combined, and two samples without reliable replicates were eliminated. A total of 1294 putative plant miRNAs had at least one mapped read in at least one of the remaining 190 samples. However, many of these miRNAs were identical orthologs or paralogs, and most were mapped at one or fewer rpm on average, and in only a small minority of samples. Across all samples, only one putative plant miRNA mapped above a median 200 read cutoff, roughly corresponding to the 32 rpm cutoff of Yuan et al (Table 1). All other RNAs, including previously reported

xenomiRs such as MIR159a, MIR168a, and the plant ribosomal degradation fragment MIR2911²²⁻²⁴, were thus below the level of background noise established by the original investigators. Indeed, previously reported xenomiRs were mapped in few samples and below 1 rpm. The absence of these RNAs is confirmed by Liu et al's analysis of the Ninomiya, et al study¹⁹, where MIR159a, MIR168a, and MIR2911 mapped in none of the plasma samples. The single putative plant miRNA that mapped above background levels in this study was, again, peu-MIR2910 (Table 2).

Lowering the threshold: still only a handful of possible xenomiRs

Since only one plant miRNA appeared to map consistently above background, the inclusion threshold of Yuan et al was relaxed to include all miRNAs with three or more mapped reads (Liu et al data) in 10% or more of the samples from either study. These are rather permissive criteria but may at least screen out some false positives due to amplification and sequencing errors. All samples from the Ninomiya study were included, despite the fact that most were not plasma. 11 miRNAs satisfied these criteria for the Yuan et al data (Table 1). (One low-mapping miRNA was excluded because its sequence could not be found in miRBase^{25,26}, miRmaid²⁷, miRNEST 2.0²⁸ or indeed through any searches attempted.) 10 satisfied the criteria from the Ninomiya study (Table 2), including one sequence that was part of another (compare ath-MIRf10046-akr and ath-MIRf10045-ak, Table 3). However, if only the plasma samples from the latter study are considered, three miRNAs remain (Table 2). In total, 15 putative miRNAs satisfied the permissive inclusion criteria, including five (Yuan only), four (Ninomiya only), and six (both) (Table 3).

To miR or not to miR

As miRNA discovery, validation, and annotation has advanced, numerous reported miRNAs have been reclassified as degradation fragments of other noncoding RNAs (ncRNAs). A classic example is MIR2911, a plant rRNA degradation fragment that has been misidentified as a microRNA. Interestingly, only 2 of the 15 miRNAs identified as plant miRNAs in this study are annotated in miRBase. Although some of these sequences may represent rare or unusually structured miRNAs, several are part of non-miRNA ncRNAs or other sequences that seem unlikely, at least at first glance, to give rise to microRNAs. Among the apparently misidentified miRNAs is MIR2910, the most abundant plant miRNA identified by Liu et al. The MIR2910 sequence, UAGUUGGUGGAGCGAUUUGUC, is found in the highly conserved and expressed large subunit (LSU) rRNA of plants, and has been specifically removed from miRBase as a non-miRNA. Even the two identified miRNAs that remain in miRBase, MIR2916 and MIR894, are not above question. A 20 nucleotide stretch of MIR2916 map to rRNA, while the full MIR894 sequence appears to be found in a variety of plant transcripts.

Human sequences in the plant database and vice-versa

Curiously, several sequences did not map to the species to which they were ascribed by the PMRD²¹. Unfortunately, the PMRD could not be accessed directly during this study; however, other databases appear to provide access to its contents. Specifically, ptc-MIRf12412-akr and ptc-MIRf12524-akr did not map to *Populus* or to other plants. The poplar tree is also not a common dietary staple of human populations. In contrast, both sequences mapped with 100% identity and coverage to numerous human sequences (Table 3). ptc-MIRf10804-akr had numerous 100% identity human matches, plus a 1-mismatch

alignment to the human miR-3929 precursor. Other miRNAs, including MIR2911, also displayed some lesser degree of matching to human transcripts or the genome. Strikingly, the putative MIR2910 sequence is not only a fragment of plant rRNA; it has a 100% coverage, 100% identity match in the human 18S rRNA (see NR_003286.2 in GenBank; Table 3). These matches of putative plant RNAs with human sequences are difficult to reconcile with the statement of Liu et al that BLAST of putative plant miRNAs "resulted in zero alignment hit" ¹, suggesting that perhaps a mistake was made, and that the BLAST procedure was performed incorrectly.

CONCLUSION

In mammalian studies, mapping of MIR2910 and other dubious plant miRNAs is best explained as mapping of human degradome fragments to plant RNAs that are in some cases genuine sequences but not miRNAs, and in other cases, human sequences that have contaminated plant RNA samples and databases. Re-analysis of the results of Liu et al ¹ thus echoes the recent findings of Kang, Bang-Berthelsen, and colleagues ¹⁴, as well as previous negative findings surrounding dietary xenomiRs, summarized above. A stringent data analysis procedure, such as filtering all reads against the ingesting organism genome/transcriptome with one or two mismatches, then requiring perfect matches of remaining reads against plant or other foreign organisms, would engender higher confidence that "foreign" RNAs are not simply amplification or sequencing artifacts. Indeed, pre-mapping to the ingesting organism's genome may not be sufficient; as shown ¹⁴, the largest number of xenomiRs in some human studies are from rodents, likely because of proximity in research laboratories. Therefore, it may be best to screen against mammalian sequences in general, and perhaps also against widespread microbe contaminants. Of course, even the most stringent analysis procedures cannot distinguish a physical contaminant from a "real" read; therefore strict process controls are also needed to assess possible contamination. In general, such controls have not been done in existing studies.

This report underlines the danger in assuming that xenomiRs in mammalian material originate from the diet. When the species and roles are reversed—for example, with the finding of human sequences in a list of poplar tree miRNAs—few analysts would conclude that poplar trees consume humans. The simplest explanation is that the sequenced plant material was contaminated with human nucleic acid. In the same way, the extremely low-level, variable, and batch-effect prone concentrations of several plant sequences in human plasma and tissue could be due to uptake from the diet, albeit at levels far too low to affect physiologic processes. However, artifact remains the simplest explanation.

METHODS

Plant mapping results from Liu et al ¹ (total mapped counts) were downloaded from the BMC Genomics website. Accession numbers of sequencing datasets were checked against

the publications of Ninomiya et al ¹⁹ and Yuan et al ²⁰, as well as the Sequence Read Archive (SRA, <https://www.ncbi.nlm.nih.gov/sra>). Data were sorted and analyzed in Microsoft Excel. Plant miRNA sequences were obtained from miRBase (<http://mirbase.org/>) ²⁹. Because certain plant sequences have been removed from miRBase because they have been identified as ncRNA degradation artifacts, the plant microRNA database (PMRD) ²¹ was consulted; however, repeated attempts to access the site (<http://bioinformatics.cau.edu.cn/PMRD>) were unsuccessful, so information was retrieved instead from miRMaid (<http://140.mirmaid.org/home>) ²⁷ or miRNEST 2.0 (<http://rhesus.amu.edu.pl/mirnest/copy/home.php>) ²⁸. All analysis files are available on request (kwitwer1@jhmi.edu).

References

1. Liu, Y.-C., Chen, W. L., Kung, W.-H. & Huang, H.-D. Plant miRNAs found in human circulating system provide evidences of cross kingdom RNAi. *BMC Genomics* **18**, 112 (2017).
2. Zhang, L. *et al.* Exogenous plant MIR168a specifically targets mammalian LDLRAP1: evidence of cross-kingdom regulation by microRNA. *Cell Res* **22**, 107–126 (2012).
3. Chin, A. R. *et al.* Cross-kingdom inhibition of breast cancer growth by plant miR159. *Cell Res* **26**, 217–28 (2016).
4. Liang, H. *et al.* Effective detection and quantification of dietetically absorbed plant microRNAs in human plasma. *J Nutr Biochem* **26**, 505–512 (2015).
5. Baier, S. R., Nguyen, C., Xie, F., Wood, J. R. & Zemleni, J. MicroRNAs Are Absorbed in Biologically Meaningful Amounts from Nutritionally Relevant Doses of Cow Milk and Affect Gene Expression in Peripheral Blood Mononuclear Cells, HEK-293 Kidney Cell Cultures, and Mouse Livers. *J Nutr* (2014). doi:jn.114.196436 [pii]10.3945/jn.114.196436
6. Snow, J. W., Hale, A. E., Isaacs, S. K., Baggish, A. L. & Chan, S. Y. Ineffective delivery of diet-derived microRNAs to recipient animal organisms. *RNA Biol* **10**, 1107–1116 (2013).
7. Witwer, K. W., McAlexander, M. A., Queen, S. E. & Adams, R. J. Real-time quantitative PCR and droplet digital PCR for plant miRNAs in mammalian blood provide little evidence for general uptake of dietary miRNAs: Limited evidence for general uptake of dietary plant xenomiRs. *RNA Biol* **10**, 1080–1086 (2013).
8. Dickinson, B. *et al.* Lack of detectable oral bioavailability of plant microRNAs after feeding in mice. *Nat Biotechnol* **31**, 965–967 (2013).
9. Tosar, J. P., Rovira, C., Naya, H. & Cayota, A. Mining of public sequencing databases supports a non-dietary origin for putative foreign miRNAs: underestimated effects of contamination in NGS. *RNA* **20**, 754–757 (2014).
10. Micó, V., Martín, R., Lasunción, M. A., Ordovás, J. M. & Daimiel, L. Unsuccessful Detection of Plant MicroRNAs in Beer, Extra Virgin Olive Oil and Human Plasma After an Acute Ingestion of Extra Virgin Olive Oil. *Plant Foods Hum. Nutr.* **71**, 102–8 (2016).
11. Auerbach, A., Vyas, G., Li, A., Halushka, M. & Witwer, K. Uptake of dietary milk miRNAs by adult humans: a validation study. *F1000Research* **5**, 721 (2016).
12. Witwer, K. W. Contamination or artifacts may explain reports of plant miRNAs in humans. *J. Nutr. Biochem.* **26**, 1685 (2015).
13. Lusk, R. W. Diverse and widespread contamination evident in the unmapped depths of high throughput sequencing data. *PLoS One* **9**, e110808 (2014).
14. Kang, W. *et al.* Survey of 800+ datasets from human tissue and body fluid reveals XenomiRs are likely artifacts. *RNA* rna.059725.116 (2017). doi:10.1261/rna.059725.116
15. Zhang, Y. *et al.* Analysis of plant-derived miRNAs in animal small RNA datasets. *BMC Genomics* **13**, 381 (2012).
16. Bağcı, C. & Allmer, J. One Step Forward, Two Steps Back; Xeno-MicroRNAs Reported in Breast Milk Are Artifacts. *PLoS One* **11**, e0145065 (2016).
17. Witwer, K. W. & Hirschi, K. D. Transfer and functional consequences of dietary microRNAs in vertebrates: Concepts in search of corroboration. *BioEssays* **36**, 394–

- 406 (2014).
18. Zheng, L.-L. *et al.* Exo-miRExplorer: A Comprehensive Resource for Exploring and Comparatively Analyzing Exogenous MicroRNAs. *Front. Microbiol.* **8**, 126 (2017).
 19. Ninomiya, S. *et al.* Potential small guide RNAs for tRNase ZL from human plasma, peripheral blood mononuclear cells, and cultured cell lines. *PLoS One* **10**, e0118631 (2015).
 20. Yuan, T. *et al.* Plasma extracellular RNA profiles in healthy and cancer patients. *Sci. Rep.* **6**, 19413 (2016).
 21. Zhang, Z. *et al.* PMRD: plant microRNA database. *Nucleic Acids Res.* **38**, D806–D813 (2010).
 22. Zhou, Z. *et al.* Honeysuckle-encoded atypical microRNA2911 directly targets influenza A viruses. *Cell Res.* **25**, 39–49 (2015).
 23. Yang, J. *et al.* Anomalous uptake and circulatory characteristics of the plant-based small RNA MIR2911. *Sci. Rep.* **6**, 26834 (2016).
 24. Yang, J., Kongchan, N., Primo Planta, C., Neilson, J. R. & Hirschi, K. D. The atypical genesis and bioavailability of the plant-based small RNA MIR2911: bulking up while breaking down. *Mol. Nutr. Food Res.* 1600974 (2017). doi:10.1002/mnfr.201600974
 25. Kozomara, A. & Griffiths-Jones, S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* **42**, D68-73 (2014).
 26. Griffiths-Jones, S., Grocock, R. J., van Dongen, S., Bateman, A. & Enright, A. J. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* **34**, D140-4 (2006).
 27. Jacobsen, A., Krogh, A., Kauppinen, S. & Lindow, M. miRMaid: a unified programming interface for microRNA data resources. *BMC Bioinformatics* **11**, 29 (2010).
 28. Szcześniak, M. W. & Makałowska, I. miRNEST 2.0: a database of plant and animal microRNAs. *Nucleic Acids Res.* **42**, D74–D77 (2014).
 29. Kozomara, A. & Griffiths-Jones, S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* **42**, D68-73 (2014).

Putative miRNA	Samples w/ reads ≥ 3	Average total counts	Median total counts	Max	Est. median rpm in avg sample
peu-MIR2910	190	1143.4	1072.5	2020	180.6
peu-MIR2916	190	119.7	115	315	19.4
tae-MIR2005	178	8.7	9	23	1.5
peu-MIR2914	169	14.3	9	348	1.5
tae-MIR2018	167	5.9	6	14	1.0
ath-MIRf10482-akr	161	7.3	6	29	1.0
ppt-MIR896	147	4.6	4	15	0.7
ptc-MIRf12412-akr	47	2.7	2	18	0.3
peu-MIR2911	42	2.6	2	10	0.3
ppt-MIR894	39	2.3	2	18	0.3
ptc-MIRf12524-akr	28	1.8	1	5	0.2

Table 1. Summary of the most frequently mapping putative plant miRNAs in the Liu et al analysis of datasets from Yuan et al. Here, data from only 190 of 192 plasma samples were included, since all but the excluded 2 samples were successfully sequenced twice. miRNA inclusion criteria were: 1) Three total mapped reads according to Liu et al's data in at least 10% of the samples and 2) discoverable putative mature sequence through miRBase, miRmaid, or miRNEST 2.0. An "estimated median rpm" value was calculated based on median total counts, average reads, and the midpoint of the reported mapping percentage range. miRNAs with perfect human matches are in red, although most (see Table 1). Note that only MIR2910 consistently exceeds the rpm threshold set by Yuan et al.

Putative miRNA	Cells (n=17) and plasma (n=6)					Plasma only (n=6)				
	Samples w/ reads ≥ 3	Average	Median	Max	avg rpm	Samples w/ reads ≥ 3	Average	Median	Max	avg rpm
<i>peu-MIR2910</i>	21	370	41.5	5369	21.3	6	1210	480.5	5369	83.2
<i>ptc-MIRf10804-akr</i>	17	26	26	60	1.2	0			0	0
<i>ptc-MIRf12412-akr</i>	17	63	61	156	2.8	0			0	0
<i>tae-MIR2018</i>	17	39	29	163	1.7	0			0	0
<i>ptc-MIRf12524-akr</i>	10	13	10.5	31	0.4	2	16	16	31	0.4
<i>tae-MIR2005</i>	7	17	11	69	0.4	0			0	0
<i>ath-MIRf10045-akr</i>	4	5	4	10	0.1	0			0	0
<i>ath-MIRf10046-akr</i>	4	5	4	10	0.1	0			0	0
<i>peu-MIR2914</i>	3	26	11	81	0.3	3	33.7	19	81	1.2
<i>peu-MIR2915</i>	3	5	4.5	8	0.0	0			0	0

Table 2. Putative plant miRNA mapping from the Ninomiya et al dataset. This dataset consisted of both cellular and plasma miRNA. Here, all results are shown in the left half of the table, and plasma results on the right. miRNA inclusion criteria were: 1) Three total mapped reads according to Liu et al's data in at least 10% of the samples (cells and plasma together) and 2) discoverable putative mature sequence through miRBase, miRMaid, or miRNEST 2.0. "Avg rpm" is calculated from the total mapped reads and total reads per sample (not mapped reads). Putative miRNAs that met inclusion criteria in the Yuan et al study are italicized, and sequences with perfect human matches are in red.

Putative miRNA	Yuan et al or Ninomiya et al	Sequence	miRBase	miRNA status in plant	Human matches?	Example match
peu-MIR2910	Both	UAGUUGGUGGAGCGAUUUGUC	No	rRNA fragment	18s rRNA	NR_003286.2
peu-MIR2916	Yuan	UGGGGACUCGAAGACGAUCAUUAU	Yes	Possible, but 20 nt maps to rRNA	Partial	
tae-MIR2005	Both	GGGUGUAUAGCUCAGUUGG	No	Unlikely: plant mitochondrial genome	Partial	
peu-MIR2914	Both	CAUGGUGGUGACGGGUGACGGAG	No	rRNA fragment	Partial	
tae-MIR2018	Both	GCCCCUCUAGCUCAGUUGGU	No	Unlikely; maps to many plant transcripts	Partial	
ath-MIRf10482-akr	Yuan	UCUACUCGACUAGGUGGUCGAGUGG	No	Maps to <i>Arabidopsis</i>	Partial	
ppt-MIR896	Yuan	GUCAAUUUGGCCGAGUGGUUAAGGC	No	tRNA fragment	Partial	
ptc-MIRf12412-akr	Both	GCUGGGAUUACAGGCGUGAGCCACC	No	Does not map to <i>Populus</i>	Many exact	XR_001736898.1
peu-MIR2911	Yuan	GGCCGGGGGACGGCUGGGGA	No	rRNA fragment	1 mismatch in a 20-nt stretch	NC_018921.2
ppt-MIR894	Yuan	CGUUUCACGUCGGGUUACCC	Yes	Yes, but other transcriptome matches	No	
ptc-MIRf12524-akr	Both	CCUGUAAUCCAGCUACUCGGG	No	Does not map to <i>Populus</i>	Many exact	NG_053018.1
ath-MIRf10045-akr	Ninomiya	UCUACUCGACCUAGGUGGUCGAGUGGU	No	Unlikely; chromosomal region	Partial	
ath-MIRf10046-akr	Ninomiya	CUCGACCUAGGUGGUCGAGUGGU	No	Part of above sequence	Partial	
peu-MIR2915	Ninomiya	CCCGUCUAGCUCAGUUGGUA	No	tRNA fragment; sequence found in many transcripts	Partial	
ptc-MIRf10804-akr	Ninomiya	CCUGUAAUCCAGCACUUUGG	No	Unlikely; microsatellite sequence	Many exact, also hsa-miR-3929 precursor (1 mismatch)	XR_001736898.1 (antisense)

Table 3. Putative plant miRNAs mapped by Liu et al from the Yuan et al or Ninomiya et al studies. Inclusion criteria were: 1) Three total mapped reads according to Liu et al's data in at least 10% of the samples in the respective studies and 2) discoverable putative mature sequence through miRBase, miRMaid, or miRNEST 2.0. miRNA status was considered unlikely if miRBase listed the miRNA as a non-miRNA or if the sequence mapped to non-miRNA regions. Human matches were exact (with an example given), "partial" (at least 15 nt stretches with 100% identity), or as otherwise described. Note that the ath-MIRf10046-akr is found within the ath-MIRf10045-akr sequence.