# A miRNA-tRNA mix-up: tRNA origin of proposed miRNA

Nick C.T. Schopman, Stephan Heynen, Joost Haasnoot & Ben Berkhout

# A miRNA-tRNA mix-up

## tRNA origin of proposed miRNA

Nick C.T. Schopman, Stephan Heynen, Joost Haasnoot and Ben Berkhout*

Laboratory of Experimental Virology; Department of Medical Microbiology; Center for Infection and Immunity Amsterdam (CINIMA); Academic Medical Center; University of Amsterdam; Amsterdam, The Netherlands

The rapid release of new data from DNA genome sequencing projects has led to a variety of misannotations in public databases. Our results suggest that next generation sequencing approaches are particularly prone to such misannotations. Two related miRNA candidates did recently enter the miRBase database, miR-1274b and miR-1274a, but they share identical 18-nucleotide stretches with tRNA$^{Lys3}$ and tRNA$^{Lys5}$, respectively. The possibility that the small RNA fragments that led to the description of these two miRNAs originated from the two tRNAs was examined. The ratio of the miR-1274b:miR-1274a fragments does closely resemble the known tRNA lys3:lys5 ratio in the cell. Furthermore, the proposed miRNA hairpins have a very low prediction score and the proposed miRNA genes are in fact endogenous retroviral elements. We searched for other miRNA-mimics in the human genome and found more examples of tRNA-miRNA mimicry. We propose that the corresponding miRNAs should be validated in more detail, as the small RNA fragments that led to their description are likely derived from tRNA processing.

The list of microRNA (miRNA) candidates encoded by the human genome is still growing at a steady pace. Some 533 human miRNAs were included in the miRBase database in 2007 (miRBase August 2007, release 10.0) and this list grew to 940 miRNAs more recently (miRBase April 2010, release 15). Besides finding of the characteristic small 18- to 21-nucleotide RNA fragment as putative miRNA, some of these miRNA entries have not been validated experimentally. For most of these, a computational prediction of the pre-miRNA hairpin structure is provided, which is very rarely experimentally verified. Mature miRNA products from one organism that are obvious homologues of well-described miRNAs from a related species are validated as such in the miRBase database. Up to August 2007, about 62% of the mature miRNA products in the database were validated experimentally in the original organism.[1] Ideally, the expression of miRNA candidates should be verified, followed by a search for their target mRNA(s) and cellular function. The rapid release of new data from DNA genome sequencing projects has led to a variety of misannotations in public databases.[2-5] Our results suggest that the output of next generation sequencing approaches, which includes the RNA transcriptome, is particularly prone to such misannotations.

## Similarity of Two miRNA and tRNA Sequences

Two related miRNA candidates that are intriguing for multiple reasons did recently enter the miRBase database. The set of related human miR-1274a and 1274b were proposed based on the identification of small 18-nucleotide RNA fragments.[6] These transcripts adopt a characteristic pre-miRNA hairpin conformation (**Fig. 1A**). We indicated the putative Dicer cleavage sites that determine the
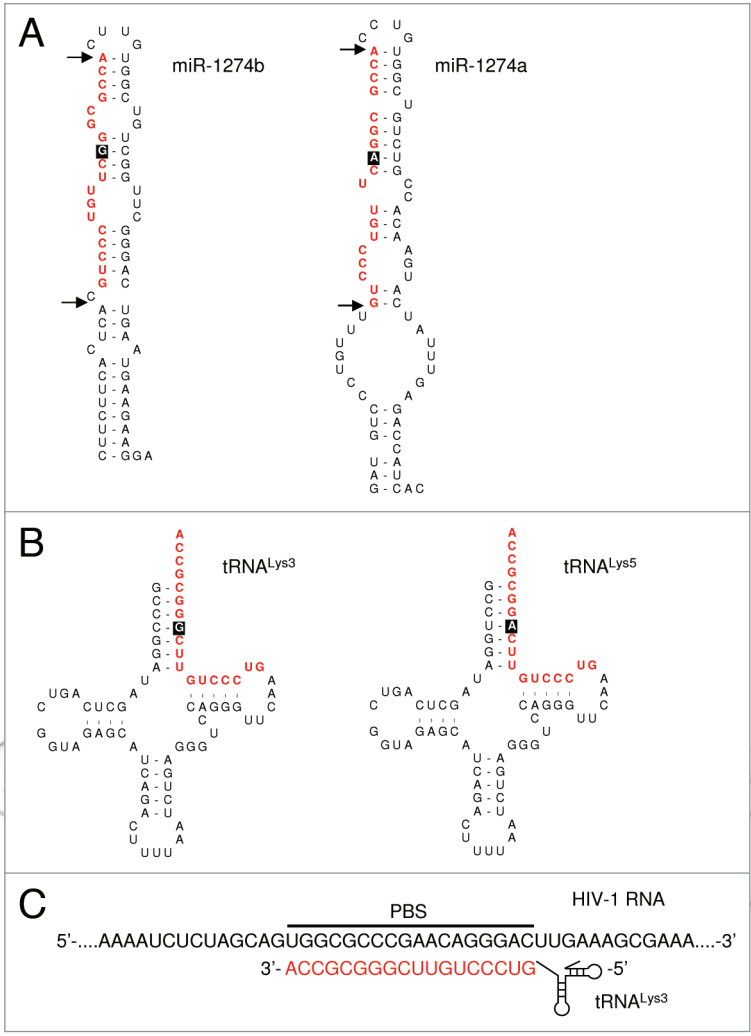
**Figure 1.** Mature human miR-1274a and miR-1274b are tRNA<sup>Lys</sup> mimics. (A) The predicted secondary structure of the miR-1274b and 1274a stem loop structure. Dicer cleavage sites are indicated by arrows and the mature miRNA sequence is marked in red. The single nucleotide difference between the mature miR-1274b and 1274a is shown as black boxed nucleotide. (B) The cloverleaf secondary structure of tRNA<sup>lys3</sup> and tRNA<sup>lys5</sup>, indicated in red is the sequence that mimics the proposed miRNAs and that was found by deep sequencing. The nucleotides in the black box indicate a single nucleotide difference between the two tRNAs, similar to the mature miRNAs. (C) Both tRNA<sup>Lys</sup> molecules are used as primer in HIV-1 RNA reverse transcription. The priming of reverse transcription by tRNA<sup>Lys3</sup> is depicted. The 3' end of the tRNA<sup>Lys3</sup> (red sequence) binds to the primer binding site (PBS) of HIV-1 RNA and the 3'-OH is subsequently extended.

mature miRNA strand. The two mature miRNAs differ in sequence only at a single nucleotide position that is marked in black. This single nucleotide polymorphism was suggested to be caused by RNA-editing,[7] although our analysis will clearly indicate that these RNAs are likely made as independent transcripts from separate genes. The putative miRNA genes encoding these transcripts are present on human chromosomes 5 and 19, respectively (**Table 1**).

Inspection of these candidate miRNA sequences struck us because of the similarity to the 3'-terminal domain of two related tRNA molecules, the major tRNA<sup>Lys3</sup> species and the minor tRNA<sup>Lys5</sup> variant (**Fig. 1B**). We marked the identical 18-nucleotide stretches in tRNA<sup>Lys3</sup> and miR-1274b and in tRNA<sup>Lys5</sup> and miR-1274a. Thus, the single nucleotide polymorphism that distinguishes miR-1274a from miR-1274b is also present in tRNA<sup>Lys3</sup> versus tRNA<sup>Lys5</sup>. This surprising finding raises several intriguing possibilities.

## On the Relationship between miRNA and tRNA Sequences

First, the production of the tRNA-like sequences from candidate miRNA molecules raises the possibility of an intricate interplay of the miRNA and tRNA pathways. We therefore tested whether miRNA-mimics can be found in the human genome for other tRNA species (tRNA<sup>Ala</sup>, tRNA<sup>Pro</sup>, tRNA<sup>Gly</sup>, tRNA<sup>Trp</sup>, tRNA<sup>Leu</sup> and tRNA<sup>Thr</sup>). In silico analysis revealed that other tRNA sequences could also be matched to human miRNAs in the miRBase database. Two tRNA species, tRNA<sup>Leu</sup> and tRNA<sup>Thr</sup> have an 18-nucleotide sequence at the 3' end that matches

**Table 1.** Candidate miRNA-tRNA misannotations

| miRNA/ tRNA | Length read (nt) | tRNA 5' or 3' end | tRNA gene position[a] | miRNA gene position[a] | miPred score (%) |
|---|---|---|---|---|---|
| miR-1274b/tRNA<sup>Lys3</sup> | 18 | 3' | 1, 6, 11, 17 | 19 (HERV-K) | 71.0 |
| miR-1274a/tRNA<sup>Lys5</sup> | 18 | 3' | 6, 11 | 5 (HERV-K) | 62.1 |
| miR-1280/tRNA<sup>Leu</sup> | 18 | 3' | 5, 6, 14, 16 | 3[b] | 67.2 |
| miR-720/tRNA<sup>Thr</sup> | 18 | 3' | 5, 14 | 3[b] | 64.4 |
| miR-1308/tRNA<sup>Gly</sup> | 18 | 5' | 1, 21 | X[b] | too low[c] |
| miR-886-5p/tRNA<sup>Ala</sup> | 18 | 5' | 6 | 5[b] | 73.3 |

[a]chromosome number; [b]no HERV identified; [c]free energy of the secondary structure below threshold.
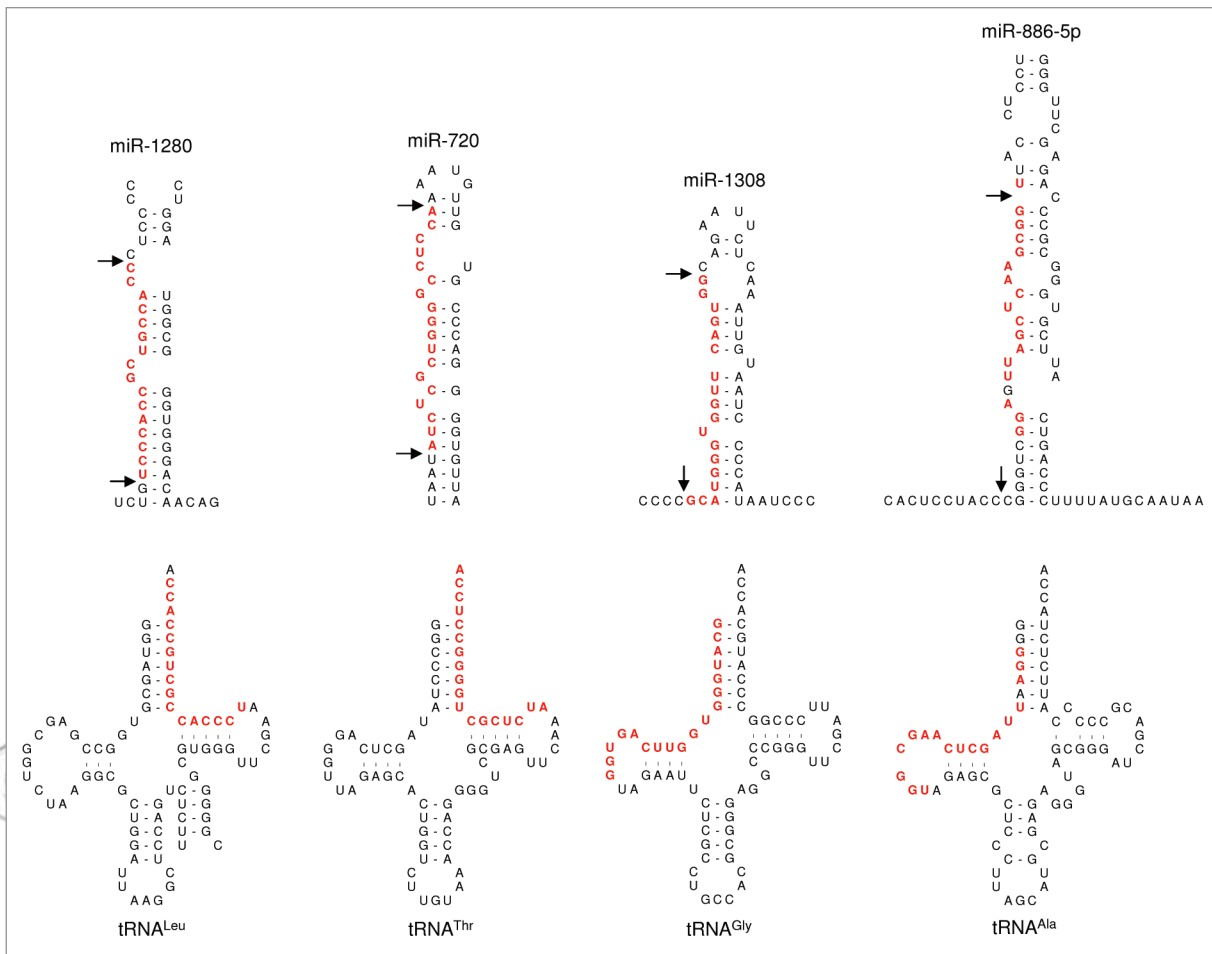
**Figure 2.** The proposed secondary structures of the human miR-1290, miR-720, miR-1308, miR-886-5p and their human tRNA-mimics. The sequence shared between miRNA and tRNA is indicated in red. The black arrows next to the miRNA hairpins indicate the Dicer processing sites.

miR-1280 and miR-720, respectively (**Fig. 2**). Interestingly, two cases represent 18-nucleotide RNA fragments from the 5'-terminal domain of a tRNA. The miR-1308 is identical to the first 18 nucleotides of tRNA$^{Gly}$ and miR-886-5p corresponds to nucleotides 3–20 of tRNA$^{Ala}$ with a central mismatch. These results are also fully compatible with a Dicer-mediated tRNA processing event, which results in 18–19 nucleotide fragments.[8] Due to this tRNA-miRNA mimicry, we propose that the corresponding miRNAs should be validated in more detail, as the small RNA fragments that led to their description can in fact be derived from tRNA processing.

Second, realizing that tRNA$^{Lys3}$ and tRNA$^{Lys5}$ have a rather unusual function in priming of the intricated reverse transcription process of the HIV-1 RNA genome in virus-infected cells,[9] an exciting scenario arises. It seems possible that

the 18-nucleotide mature miRNAs actually serve as primer for HIV-1 reverse transcription, which is an attractive idea because the primer-binding site (PBS) on HIV-1 RNA is the exact 18-nucleotide complement (**Fig. 1C**). In other words, how sure are we that HIV-1 replication is supported by a tRNA primer and not by the 18-nucleotide tRNA fragment or a miRNA-derived primer? Biochemical analysis of the priming species present in virion particles is consistent with a tRNA primer, although it does not rule out another source for priming.[10,11] Studies with mutant HIV-1 variants in which the PBS in the RNA genome was changed to accommodate another tRNA species do clearly disfavor the miRNA-route.[11]

A third less exciting possibility comes to mind. Perhaps the identified mature miRNA fragments represent cleavage products of the related tRNA molecules.

A first indication that this scenario may be true comes from inspection of the ratio of the molecules: tRNA$^{Lys3}$ is more abundant than tRNA$^{Lys5}$ and the former is also used more frequently as primer for HIV-1 reverse transcription. We based this on the presence of the identifying point mutation in the PBS of HIV-1 and HIV-2 isolates because this sequence of the viral progeny is in fact copied from the priming tRNA species during reverse transcription.[12,13] We inspected a total of 265 HIV isolates, of which 9 showed the Lys5-specific point mutation (ratio Lys3:Lys5 = 29:1). More recently, deep sequencing of small tRNA fragments in extracts from the SupT1 T-cell line revealed 1966 Lys3 and 53 Lys5 reads, yielding a similar 37:1 ratio (unpublished results). Strikingly, the reported number of miR-1274b and miR-1274a reads (670 versus 24) also yields a similar ratio of 28:1.[6]

Another indication that tRNA degradation can lead to 18-nucleotide fragments that mimic miRNA molecules comes from recent deep sequencing studies, which revealed many of such tRNA fragments.[8,14,15] If true, this scenario puts into doubt the miRNA nature of miR1274a and miR1274b (**Fig. 1**) and the other miRNAs (**Fig. 2**). Consistent with this idea is the relatively low score of the pre-miRNAs of 1274b and 1274a in the miPred algorithm that gives a 71.0% and 62.1% prediction confidence, respectively. For comparison, two randomly chosen miRNAs, miR-16-1 and miR-122, have a predicted score of 83.7% and 82.9%, respectively. The other four miRNA candidates also obtained a low score ranging from below the threshold to 73.3% (**Table 1**).

The set of miRNAs for which we suspect that they may be of tRNA origin do not represent minor small RNA species. For instance, the early Morin study on small RNAs in human embryonic stem cells revealed three of the miRNA/tRNA sets that we discussed, including miR-1274b or tRNA$^{Lys3}$ that ranked as the second most abundant small RNA, miR-1308 or tRNA$^{Gly}$ as third most abundant and miR-1274a or tRNA$^{Lys5}$ that ranked 19th.[6]

## Endogenous Retroviral Origin of the Proposed miRNA Genes

But what then is the origin of the candidate miRNA genes on chromosome 19 and 5 that have an identical 18-nucleotide match with the mature tRNA$^{Lys3}$ and tRNA$^{Lys5}$, respectively. Please note that the predicted mature miRNA/tRNA fragments share the 3'-terminal CCA sequence that is only added post-transcriptionally to all tRNA molecules.[16] In other words, the miRNA-like genes cannot be tRNA genes. The miRNAs could represent a pseudogene that is made by reverse transcription of a mature tRNA, but the 5'-flanks of the miRNAs do not resemble the flanking sequences of the tRNA molecules, making this explanation less likely. Separate gene locations have also been proposed for the four other miRNAs that mimic cellular tRNAs (**Fig. 2**, **Table 1**).

The 3'-part of a mature tRNA molecule is copied into the retroviral genome during the process of reverse transcription, including the 3' terminal CCA-end. Thus, is it possible that the proposed miR-1274b and 1274a genes represent integrated proviral genomes of an endogenous retrovirus. Inspection of the flanking sequences did indeed reveal that these two miRNA-like genes on chromosome 19 and 5 represent members of the extended HERV-K family of endogenous retroviruses (**Table 1**), more specifically the HML-6 group.[17,18] In fact, these results also suggest that this retrovirus could use both tRNA$^{Lys3}$ and tRNA$^{Lys5}$ as primer for reverse transcription, very similar to the exogenous HIV-1 retrovirus. Previous reports already hinted at the possibility that small RNA fragments with a tRNA-like sequence may be derived from the antisense strand of endogenous retroviral elements, which have a sense PBS sequence complementary to the 3' end of a cellular tRNA.[7,14]

In conclusion, we propose that two recent additions to the miRNA database may not represent bonafide miRNA molecules. We demonstrated that the small RNA fragments that led to their identification can be derived from two related tRNA molecules, and the observed ration of the two fragments resembles the known tRNA ratio. In addition, the proposed miRNA structures are not supported by a high prediction value and their proposed genes are instead the PBS elements of endogenous retroviruses. The confusion is due to the generation of small 3'-terminal tRNA fragments that are fully complementary to retroviral PBS elements, giving rise to the misinterpretation of retroviral sequences as miRNA-encoding genes. Several deep sequencing studies have reported "cross-mapping" of miRNA sequences with cellular tRNAs.[7] In general, tRNAs are highly expressed and can be processed into discrete small RNA species. Our results indicate that tRNA-derived miRNA-like products form a substantial fraction of small RNA deep sequencing libraries. It is important to realize the danger of this miRNA-tRNA mix-up as there are many more tRNAs and endogenous retroelements, such that misinterpretation may occur more frequently.

### References

1. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. miRBase: tools for microRNA genomics. Nucleic Acids Res 2008; 36:154-8.
2. Bidartondo MI. Preserving accuracy in GenBank. Science 2008; 319:1616.
3. Korning PG, Hebsgaard SM, Rouze P, Brunak S. Cleaning the GenBank *Arabidopsis thaliana* data set. Nucleic Acids Res 1996; 24:316-20.
4. Nilsson RH, Ryberg M, Kristiansson E, Abarenkov K, Larsson KH, Kõljalg U. Taxonomic reliability of DNA sequences in public sequence databases: a fungal perspective. PLoS ONE 2006; 1:59.
5. Zaranek AW, Levanon EY, Zecharia T, Clegg T, Church GM. A survey of genomic traces reveals a common sequencing error, RNA editing and DNA editing. PLoS Genet 2010; 6:1000954.
6. 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.
7. de Hoon MJ, Taft RJ, Hashimoto T, Kanamori-Katayama M, Kawaji H, Kawano M, et al. Cross-mapping and the identification of editing sites in mature microRNAs in high-throughput sequencing libraries. Genome Res 2010; 20:257-64.
8. Cole C, Sobala A, Lu C, Thatcher SR, Bowman A, Brown JW, et al. Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs. RNA 2009; 15:2147-60.
9. Abbink TE, Berkhout B. HIV-1 reverse transcription initiation: a potential target for novel antivirals? Virus Res 2008; 134:4-18.
10. Jiang M, Mak J, Ladha A, Cohen E, Klein M, Rovinski B, et al. Identification of tRNAs incorporated into wild-type and mutant human immunodeficiency virus type 1. J Virol 1993; 67:3246-53.
11. Das AT, Klaver B, Berkhout B. Reduced replication of human immunodeficiency virus type 1 mutants that use reverse transcription primers other than the natural tRNA(3Lys). J Virol 1995; 69:3090-7.
12. Das AT, Klaver B, Berkhout B. Sequence variation of the HIV primer-binding site suggests the use of an alternative tRNA$^{Lys}$ molecule in reverse transcription. J Gen Virol 1997; 78:837-40.
13. Das AT, Vink M, Berkhout B. Alternative tRNA priming of human immunodeficiency virus type 1 reverse transcription explains sequence variation in the primer-binding site that has been attributed to APOBEC3G activity. J Virol 2005; 79:3179-81.
14. Kawaji H, Nakamura M, Takahashi Y, Sandelin A, Katayama S, Fukuda S, et al. Hidden layers of human small RNAs. BMC Genomics 2008; 9:157.
15. Haussecker D, Huang Y, Lau A, Parameswaran P, Fire AZ, Kay MA. Human tRNA-derived small RNAs in the global regulation of RNA silencing. RNA 2010; 16:673-95.
16. Shi PY, Maizels N, Weiner AM. CCA addition by tRNA nucleotidyltransferase: polymerization without translocation? EMBO J 1998; 17:3197-206.
17. Medstrand P, Mager DL, Yin H, Dietrich U, Blomberg J. Structure and genomic organization of a novel human endogenous retrovirus family: HERV-K (HML-6). J Gen Virol 1997; 78:1731-44.
18. Zsiros J, Jebbink MF, Lukashov VV, Voute PA, Berkhout B. Evolutionary relationships within a subgroup of HERV-K-related human endogenous retroviruses. J Gen Virol 1998; 79:61-70.