Extensive undocumented changes, disputable new entries, and records that are in conflict with 'NCBI Gene' and 'HGNC' permeate release 2.0 of gtRNAdb

Phillipe Loher, Venetia Pliatsika, Aristeidis G. Telonis, Yohei Kirino and Isidore Rigoutsos
Computational Medicine Center, Sidney Kimmel Medical College, Thomas Jefferson University, 1020 Locust Street, Philadelphia, PA 19107.

* Correspondence should be addressed to: Isidore Rigoutsos (isidore.rigoutsos@jefferson.edu)

NOTE: The described data are based on gtRNAdb's contents as it existed on January 17, 2016

Transfer RNA (tRNA) molecules have long been thought of as housekeeping molecules whose activities revolve solely around the translation process of messenger RNA into an amino acid sequence (1). This view is slowly being revised in light of accumulating evidence that tRNAs regulate cell phenotype (2,3) and are also sources of short non-coding RNAs (ncRNAs) (4,5) with largely unknown functions (6-8) that have variable lengths and are produced constitutively in different tissues in a manner that depends on a person's gender, population origin, race as well as tissue, tissue state and disease type/subtype (9).

For many years, the gtRNAdb repository (10) has been a key resource for researchers studying tRNAs. In light of the renewed research interest in this class of molecules, we have been eagerly anticipating version 2.0 (v2.0) of the gtRNAdb that was just released on line at http://gtrnadb.ucsc.edu/ and accompanied by a matching article (11). We were surprised to discover that v2.0 contains mistakes that are bound to adversely affect experimental studies of tRNAs and analytical pipelines, if not corrected promptly. We highlight these mistakes below by referencing the *Homo sapiens* portion of the database, which, according to the authors (11), is one of the most referenced sections of gtRNAdb. In our discussion, we will be using human genome assembly hg19 of v2.0 (accessible through http://lowelab.ucsc.edu/GtRNAdb/Hsapi19/) to facilitate comparisons with v1.0 of gtRNAdb that was based on hg19.

Sweeping, undocumented and disputable changes compared to version 1.0

In (11), Chan and Lowe state that they populated gtRNAdb v2.0 using a method that underwent major development but has not yet been peer-reviewed: the method is cited in (11) as "*Chan et al., in preparation*." The differences in the *Homo sapiens* tRNAs between the two versions of gtRNAdb are pervasive. In the absence of a description of the used methodology, we are limited to enumerating the changes between the two versions to enable a more detailed analysis when the description of the used method becomes available. Specifically for *Homo sapiens*:

- 74 of the 625 human tRNAs originally in v1.0 have been deleted and are now absent from v2.0 (Supp. Table 1);
- 41 of the 102 original human pseudo-tRNAs of v1.0 have been *elevated* to tRNAs in v2.0 (Supp. Table 2);
- 20 human tRNAs had their end-points modified in v2.0 compared to v1.0 (Supp. Table 3);
- 6 human tRNAs had their anticodon changed from v1.0 to v2.0 (Supp. Table 4); and,
- 60 *new* human tRNAs have been added in v2.0 (Supp. Table 5).

Of the 60 additions to v2.0, 39 were published in 2014 as part of a collection of 497 such sequences we reported in the human genome (see (12,13) and Supp. Table 5).

Many of the new or modified human tRNAs have secondary structures that deviate substantially from the tRNA cloverleaf

In v2.0 of gtRNAdb, Chan and Lowe use a covariance model to propose a more relevant secondary structure for the listed tRNAs: the covariance model accounts for nucleotide modifications that could be affecting folding. Because it relies on additional information, the resulting structure will generally differ from the one corresponding to the minimum folding energy. We inspected manually the secondary structures of those human tRNA entries that are new to v2.0 and evaluated them from the standpoint of whether they form properly folded secondary structures ("cloverleaf"). We found that:

- at least 9 of the 41 elevated pseudo-tRNAs,
- at least 15 of the 20 entries whose endpoints changed in v2.0, and
- at least 18 of the 60 newly added tRNAs

have secondary structures that deviate greatly from the tRNA cloverleaf (Supp. Table 6). In particular, in these cases the reported covariance model structures exhibit one or more instances of the following: mismatch at 1-72 base pairing; one or more mismatches at the stem; a bulge at the stem; unusual anticodon-loops; missing D-loop; missing T-loop; etc. Figure 1 shows six such examples.

In (11), Chan and Lowe do acknowledge that some of the covariance model structures that they report may be problematic and argue that their inclusion in v2.0 should spur research into tRNAs rather than hinder it. Considering that gtRNAdb is the go-to repository for those engaging in tRNA research (e.g. (7-9,12-25)), there is clear risk that users may not lend the proper weight to tRNAs whose secondary structures do not resemble the typical cloverleaf. Consequently, instead of being questioned, the corresponding "tRNA" sequences would become members of the collection of tRNAs that are encoded by the human genome. In our opinion, it would have been more prudent to delegate these sequences into a group of "candidate tRNAs" while awaiting the generation of evidence by the community either in favor of or against their tRNA nature. We note that in v2.0 no entries are labeled as "pseudo-tRNAs."

V2.0 contains numerous incorrect legacy identifiers

To help ensure continuity between versions, developers typically make available tables that link

new and old identifiers. Such tables are indispensable since they enable researchers to identify the correct tRNA sequence using an older label for it that they may encounter in a previous publication. GtRNAdb v2.0 contains 606 entries: for 405 of them the legacy identifiers have been reported incorrectly (Supp. Table 7). Specifically:

- legacy identifiers were associated with the 60 new entries of v2.0 – since these entries did not exist in v1.0 they should not have been assigned legacy identifiers;
- for 331 entries among those whose coordinates did not change between the v1.0 and v2.0 the legacy identifier claimed in v2.0 does not match the identifier that the entry had in v1.0; and,
- 14 entries that had their endpoint coordinates shifted between v1.0 and v2.0 were assigned legacy identifiers that do not match those of the original instance of the sequence in v1.0.

Many v2.0 tRNA records contain incongruent data that contradict their counterpart NCBI Gene and HGNC records

Each human tRNA record in v2.0 comprises a *unique identifier*, e.g. Lys-TTT-3-3, an associated set of global genomic coordinates (chromosome, strand, from/to endpoints), and the corresponding tRNA sequence. This identifier allows one to retrieve the unique NCBI Gene record for the respective human tRNA. For 116 of the 606 human tRNAs in gtRNAdb v2.0, their gtRNAdb records list different chromosome, strand, and endpoint information than the respective NCBI Gene record (Supp. Table 8). It is unclear which of the records (gtRNAdb or NCBI Gene) should be treated as the “correct” ones. In some instances, the records form long “chains” of incongruent labels and coordinates. This is highlighted by the example in Figure 2: gtRNAdb v2.0 associates tRNA-iMet-CAT-1-2 with coordinates 6_+_26286754_26286825; looking up tRNA-iMet-CAT-1-2 at NCBI Gene shows that this tRNA is associated with coordinates 17_-_80452597_80452668; looking up 17_-_80452597_80452668 in gtRNAdb v2.0 shows that it is associated with tRNA-iMet-CAT-1-8, which in NCBI Gene is associated with coordinates 6_-_27870271_27870342, which in gtRNAdb v2.0 is associated with tRNA-iMet-CAT-1-7, which in NCBI Gene is associated with coordinates 6_-_27560600_27560671, etc. Eventually, we reach NCBI Gene’s record for tRNA-iMet-CAT-1-3 that claims the latter’s coordinates as being 6_+_26286754_26286825, which gtRNAdb associates with tRNA-iMet-CAT-1-2 in gtRNAdb. Several more examples of incongruent such chains can be seen among the data of Supp. Table 8.

The incongruent contents of gtRNAdb v2.0 records are not confined to NCBI Gene. The database that is maintained by the "HUGO Gene Nomenclature Committee" (HGNC, accessible through http://www.genenames.org/) reveals similar inconsistencies. This is particularly problematic because in gtRNAdb v2.0 each of the *Homo sapiens* tRNA records links directly to the corresponding HGNC record. Examples of such incongruent records include: tRNA-Asp-GTC-2-10 (listed as appearing on chr. 12 by gtRNAdb v2.0 and on chr. 6 by HGNC), tRNA-Asp-GTC-2-6 (listed as appearing on chr. 6 by gtRNAdb v2.0 and on chr. 12 by HGNC), tRNA-Pro-TGG-3-2 (listed as appearing on chr. 14 by gtRNAdb v2.0 and on chr. 16 by HGNC), tRNA-Pro-TGG-3-5 (listed as appearing on chr. 16 by gtRNAdb v2.0 and on chr. 5 by HGNC), tRNA-Pro-TGG-3-1 (listed as appearing on chr. 5 by gtRNAdb v2.0 and on chr. 14 by HGNC), etc.

Conclusion

Even confining ourselves to the 606 records contained in the *Homo sapiens* portion of gtRNAdb v2.0 we encountered a significant number of incorrect legacy identifiers, inaccurate genomic information, conflicting references, etc. Because gtRNAdb is such a valuable resource for many scientists, it will be imperative that these problems be corrected promptly before they taint the pipelines of researchers worldwide.

Competing interests

All authors are actively involved in tRNA research or have published previously in this area. The authors declare no competing financial interests.

Authors' Contributions

All authors analyzed the data and wrote the commentary.

FIGURE CAPTIONS

Figure 1. Six examples of entries that are new to v2.0 of gtRNAdb. Even though the structures were established with the help of the covariance model, they deviate considerably from the typical tRNA cloverleaf structure. A: tRNA-Lys-CTT-chr15-5 – missing acceptor stem. B: tRNA-Leu-CTA-chr5-1 – missing anticodon stem. C: tRNA-Cys-ACA-1-1 – no cloverleaf structure. D: tRNA-Leu-CAA-7-1 – missing D arm/loop. E: tRNA-Gln-CTG-13-1 – missing anticodon arm/loop. F: tRNA-Cys-GCA-chr11-21 – missing T arm/loop.

Figure 2. An example of incongruent data contained in multiple gtRNAdb v2.0 records and NCBI Gene records. The first row's (tRNA-iMet-CAT-1-2) gtRNAdb v2.0 and NCBI Gene records can be seen at http://gtrnadb.ucsc.edu/genomes/eukaryota/Hsapi19/genes/tRNA-iMet-CAT-1-2.html and http://www.ncbi.nlm.nih.gov/gene/?term=trna-iMet-CAT-1-2 respectively; similarly for the other entries. See also text for an explanation.

REFERENCES

1. Barciszewska, M.Z., Perrigue, P.M. and Barciszewski, J. (2015) tRNA - the golden standard in molecular biology. Molecular bioSystems, **12**, 12-17.
2. Phizicky, E.M. and Hopper, A.K. (2010) tRNA biology charges to the front. Genes Dev, **24**, 1832-1860.
3. Raina, M. and Ibba, M. (2014) tRNAs as regulators of biological processes. Frontiers in Genetics, **5**, 171.
4. Keam, S. and Hutvágner, G. (2015) tRNA-Derived Fragments (tRFs): Emerging New Roles for an Ancient RNA in the Regulation of Gene Expression. Life, **5**, 1638-1651.
5. Sobala, A. and Hutvágner, G. (2011) Transfer RNA-derived fragments: origins, processing, and functions. WIREs RNA, **2**, 853-862.
6. Gebetsberger, J. and Polacek, N. (2013) Slicing tRNAs to boost functional ncRNA diversity. RNA Biol, **10**, 1798-1806.
7. Kumar, P., Anaya, J., Mudunuri, S.B. and Dutta, A. (2014) Meta-analysis of tRNA derived RNA fragments reveals that they are evolutionarily conserved and associate with AGO proteins to recognize specific RNA targets. BMC Biol., **12**, 78.
8. Malhotra, A. and Dutta, A. (2009) A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). Genes Dev, **23**, 2639-2649.
9. Telonis, A.G., Loher, P., Honda, S., Jing, Y., Palazzo, J., Kirino, Y. and Rigoutsos, I. (2015) Dissecting tRNA-derived fragment complexities using personalized transcriptomes reveals novel fragment classes and unexpected dependencies. Oncotarget, **6**, 24797-24822.
10. Chan, P.P. and Lowe, T.M. (2009) GtRNAdb: a database of transfer RNA genes detected in genomic sequence. Nucleic Acids Res, **37**, D93-D97.
11. Chan, P.P. and Lowe, T.M. (2015) GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. Nucleic Acids Res, gkv1309-1306.
12. Telonis, A.G., Kirino, Y. and Rigoutsos, I. (2015) Mitochondrial tRNA-lookalikes in nuclear chromosomes: Could they be functional? RNA Biol, **12**, 375-380.
13. Telonis, A.G., Loher, P., Kirino, Y. and Rigoutsos, I. (2014) Nuclear and mitochondrial tRNA-lookalikes in the human genome. Frontiers in Genetics.
14. Baras, A.S., Mitchell, C.J., Myers, J.R., Gupta, S., Weng, L.-C., Ashton, J.M., Cornish, T.C., Pandey, A. and Halushka, M.K. (2015) miRge - A Multiplexed Method of Processing Small RNA-Seq Data to Determine MicroRNA Entropy. PLoS ONE, **10**, e0143066-0143016.
15. Casas, E., Cai, G. and Neill, J.D. (2015) Characterization of circulating transfer RNA-derived RNA fragments in cattle. Frontiers in Genetics, **6**, 1-20.
16. Dhahbi, Spinder, S., Atamna, H., Boffelli, D. and Martin, D. (2014) Deep Sequencing of Serum Small RNAs Identifies Patterns of 5´ tRNA Half and YRNA Fragment Expression Associated with Breast Cancer. BIC, 37.
17. Gingold, H., Tehler, D., Christoffersen, N.R., Nielsen, M.M., Asmar, F., Kooistra, S.M., Christophersen, N.S., Christensen, L.L., Borre, M., Sørensen, K.D. et al. (2014) A Dual Program for Translation Regulation in Cellular Proliferation and Differentiation. Cell, **158**, 1281-1292.
18. Guy, M.P. and Phizicky, E.M. (2015) Two-subunit enzymes involved in eukaryotic post-transcriptional tRNA modification. RNA Biol, 0.
19. Honda, S., Loher, P., Shigematsu, M., Palazzo, J.P., Suzuki, R., Imoto, I., Rigoutsos, I. and Kirino, Y. (2015) Sex hormone-dependent tRNA halves enhance cell proliferation in breast and prostate cancers. Proc Natl Acad Sci USA, **112**, E3816-3825.
20. Lalaouna, D., Carrier, M.-C., Semsey, S., Brouard, J.-S., Wang, J., Wade, J.T. and Massé, E. (2015) A 3' External Transcribed Spacer in a tRNA Transcript Acts as a Sponge for Small RNAs to Prevent Transcriptional Noise. Mol Cell, 1-14.
21. Ortogero, N., Schuster, A.S., Oliver, D.K., Riordan, C.R., Hong, A.S., Hennig, G.W., Luong, D., Bao, J., Bhetwal, B.P., Ro, S. et al. (2014) A Novel Class of Somatic Small RNAs Similar to Germ Cell Pachytene PIWI-interacting Small RNAs. J Biol Chem.

22. Rao, B.S. and Jackman, J.E. (2015) Life without post-transcriptional addition of G-1: two alternatives for tRNAHis identity in Eukarya. RNA, **21**, 243-253.
23. Victoria, B., Dhahbi, J.M., Nunez Lopez, Y.O., Spinel, L., Atamna, H., Spindler, S.R. and Masternak, M.M. (2015) Circulating microRNA signature of genotype-by-age interactions in the long-lived Ames dwarf mouse. Aging Cell, n/a-n/a.
24. Yamtich, J., Heo, S.-J., Dhahbi, J., Martin, D.I.K. and Boffelli, D. (2015) piRNA-like small RNAs mark extended 3'UTRs present in germ and somatic cells. BMC Genomics, 1-12.
25. Zheng, G., Qin, Y., Clark, W.C., Dai, Q., Yi, C., He, C., Lambowitz, A.M. and Pan, T. (2015) Efficient and quantitative high-throughput tRNA sequencing. Nat Methods, 1-5.

Figure 1

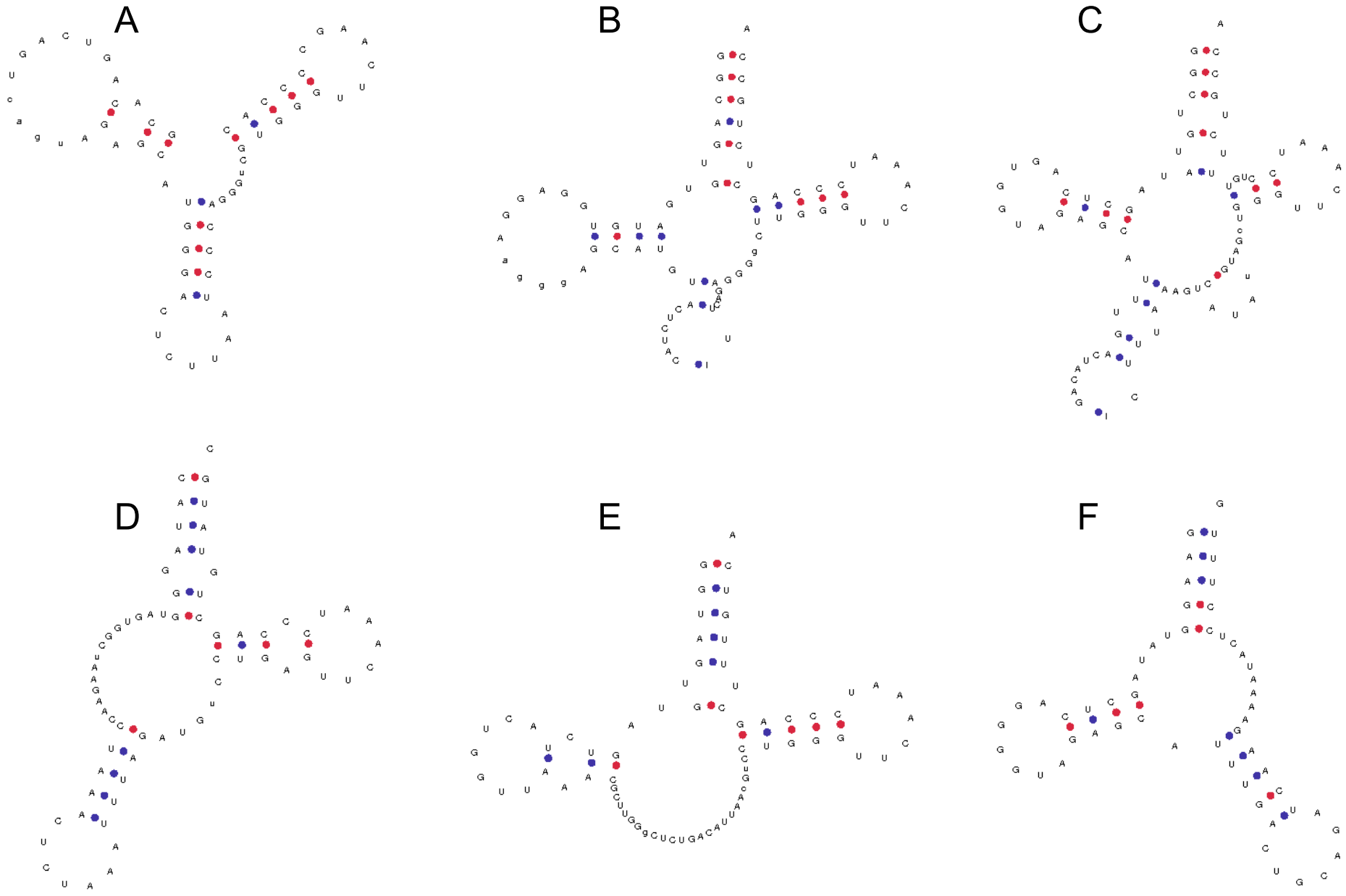
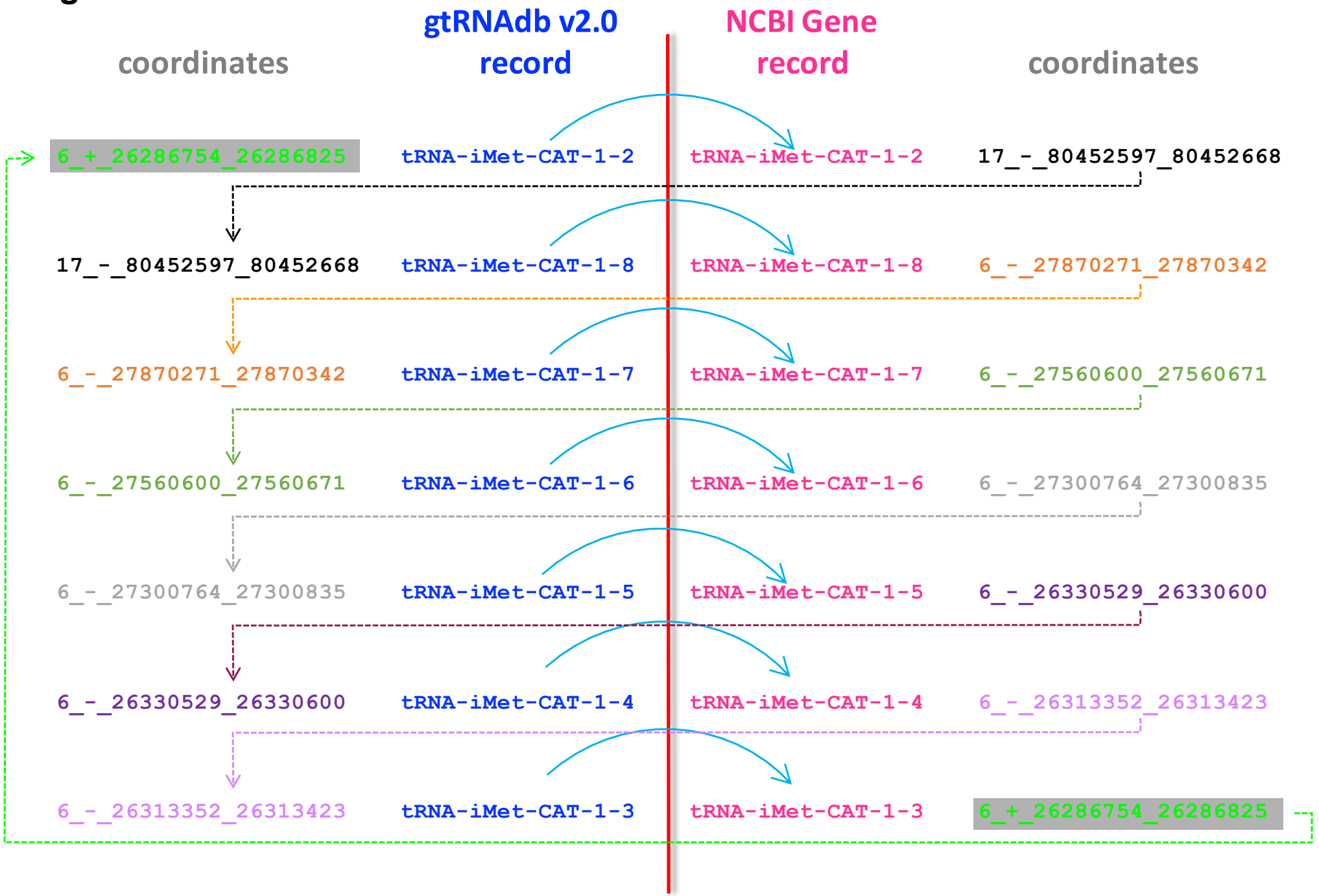


Figure 2


gtRNAdb v2.0 record
NCBI Gene record
coordinates
coordinates
6_+_26286754_26286825
tRNA-iMet-CAT-1-2
tRNA-iMet-CAT-1-2
17_-_80452597_80452668
17_-_80452597_80452668
tRNA-iMet-CAT-1-8
tRNA-iMet-CAT-1-8
6_-_27870271_27870342
6_-_27870271_27870342
tRNA-iMet-CAT-1-7
tRNA-iMet-CAT-1-7
6_-_27560600_27560671
6_-_27560600_27560671
tRNA-iMet-CAT-1-6
tRNA-iMet-CAT-1-6
6_-_27300764_27300835
6_-_27300764_27300835
tRNA-iMet-CAT-1-5
tRNA-iMet-CAT-1-5
6_-_26330529_26330600
6_-_26330529_26330600
tRNA-iMet-CAT-1-4
tRNA-iMet-CAT-1-4
6_-_26313352_26313423
6_-_26313352_26313423
tRNA-iMet-CAT-1-3
tRNA-iMet-CAT-1-3
6_+_26286754_26286825