



OPEN ACCESS

EDITED BY

Stephen M. Mount,
University of Maryland, College Park,
United States

REVIEWED BY

Jianlei Gu,
Yale University, United States
Yanchao Pan,
University of Michigan, United States

*CORRESPONDENCE

Sebastian Canzler,
✉ sebastian.canzler@ufz.de

RECEIVED 15 December 2024

ACCEPTED 12 February 2025

PUBLISHED 10 March 2025

CITATION

Lehmann J, Yazbeck A, Hackermüller J and
Canzler S (2025) An extended miRNA
repertoire in *Rattus norvegicus*.
Front. Bioinform. 5:1545680.
doi: 10.3389/fbinf.2025.1545680

COPYRIGHT

© 2025 Lehmann, Yazbeck, Hackermüller and
Canzler. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

An extended miRNA repertoire in *Rattus norvegicus*

Julienne Lehmann^{1,2}, Ali Yazbeck³, Jörg Hackermüller^{1,4} and
Sebastian Canzler^{1*}

¹Department Computational Biology and Chemistry, Helmholtz Centre for Environmental Research - UFZ, Leipzig, Germany, ²Department of Internal Medicine and Pediatrics, HIV Cure Research Center, Ghent University Hospital, Ghent University, Ghent, Belgium, ³Medical Faculty, Institute of Pathology, University of Cologne, Cologne, Germany, ⁴Department of Computer Science, Leipzig University, Leipzig, Germany

KEYWORDS

miRNA, micro RNA, ncRNA, non-coding RNA, homology search, *Rattus norvegicus*

1 Introduction

MicroRNAs (miRNAs) are small, non-coding RNA molecules, approximately 22 nucleotides in length, that play crucial roles in the regulation of gene expression. They function primarily by binding to complementary sequences in the 3' untranslated regions (UTRs) of target messenger RNAs (mRNAs), leading to mRNA degradation or translational repression (Bartel, 2004; Bushati and Cohen, 2007). Through this mechanism, miRNAs are involved in various biological processes, including development, differentiation, proliferation, and apoptosis (Ambros, 2004). The importance of miRNAs as regulatory elements is furthermore emphasized by their involvement in various diseases, particularly cancer, where they can act as either oncogenes or tumor suppressors (Budakoti et al., 2021; Hussien et al., 2021).

MicroRNAs are transcribed as primary miRNAs (pri-miRNAs) and processed into precursor miRNAs (pre-miRNAs), which are typically around 70 nucleotides long and form hairpin structures (Bartel, 2004). The miRNA duplex is generated from this precursor, consisting of a guide strand (mature miRNA) and a passenger strand (mature*). The mature miRNA is incorporated into the RNA-induced silencing complex (RISC) to guide gene silencing, while the mature* strand is usually degraded, although in some cases, it may also be functional (Bartel, 2004; Okamura et al., 2007).

Despite their critical functions, there is a significant discrepancy in the annotation of miRNAs between different model species, notably between rat (*Rattus norvegicus*) and mouse (*Mus musculus*). This discrepancy arises due to differences in sequencing efforts and annotation strategies but also through lineage-specific retrotransposons playing an essential role in the birth of new miRNA genes (Lehnert et al., 2011). Addressing this gap is essential for leveraging the rat as a model organism in biomedical research, particularly given its widespread use in pharmacology and toxicology studies (Jacob and Kwitek, 2002).

In this study, we corrected several incorrect homology assignments and identified and annotated novel rat miRNAs. Expanding the miRNA repertoire of this crucial model organism will enhance its utility, particularly for toxicological applications, where precise regulatory networks are critical for understanding the molecular basis of toxicity and drug responses.

2 Methods

To identify novel rat miRNAs, we utilized MIRf1x curated whole precursor miRNA family covariance models as described previously (Yazbeck et al., 2019). We focused on miRNA families that contained at least one mammalian miRNA sequence. The model building was based on miRBase version 21 (Kozomara and Griffiths-Jones, 2014).

We employed *infernal* v1.1.3 (Nawrocki and Eddy, 2013) to scan the rat genome (Rnor_6.0; Ensembl Release 102, accession number GCA_000001895.4) for potential miRNA candidates using default parameters. We chose Rnor_6.0 as reference since miRBase relies on this assembly for miRNA annotations, ensuring comparability with existing datasets. The candidate miRNAs were then subjected to a series of stringent filtering steps to ensure the accuracy and relevance of the identified sequences: (1) Candidates were filtered based on an e-value cutoff of 0.01 and a bit score threshold of 33, following the recommendations in the *infernal* tutorial [$\log_2(2 \times \text{genome size})$]. (2) Duplicated candidates located on unfinished chromosomes were eliminated. (3) Candidates overlapping with repeats annotated by RepeatMasker were excluded (Smit et al., 2015). (4) Candidates that were reverse complements of candidates with smaller e-value were also excluded.

The remaining candidate miRNAs were curated using *MIRfix* on whole family alignments, which included the newly identified rat candidates. This was followed by an additional manual curation of the alignments involving a check for sequence conservation of mature and/or mature* regions and the assessment of the ability of novel sequences to fold into a hairpin secondary structure.

Potential miRNA candidates were manually assigned names in accordance with their homologous mouse miRNAs. Finally, the novel miRNA sequences were again blasted against the rat genome (Rnor_6.0) to extract the precise genomic coordinates using *blastn*¹. To ensure compatibility with the newer mRatBN7.2 assembly (de Jong et al., 2024) we mapped the coordinates using *CrossMap* (Zhao et al., 2014).

3 Data analysis

For our *infernal* scan, we utilized 781 mammalian miRNA families (excluding singletons), which included 435 already annotated rat miRNA sequences distributed across 247 miRNA families. This scan resulted in a total of 449 417 significant candidates scattered over 459 miRNA families.

Following a stringent filtering procedure to eliminate duplicates on unfinished chromosomes, overlaps with annotated miRNAs, repeats, and reverse complements, we identified 3521 potential novel miRNAs within 186 families. The three families with the most candidates accounted for nearly 2500 of those potential sequences. These families contain large numbers of annotated mouse sequences (up to 59 in MIPF0000316), hence introducing substantial variability. This circumstance leads to the detection of a high number of candidate sequences. For each of the 186 miRNA families with at least one candidate sequence, we conducted a *MIRfix* analysis and correction. Additionally, we manually curated the whole family alignments to further refine this set. The final set of new miRNAs in *R. norvegicus* contained 55 novel sequences, that have been uploaded to the European Nucleotide Archive (ENA) at EMBL-EBI with the accession numbers OZ078105–OZ078160². Notably, this included 39 families where no miRNA had previously been annotated in rat.

1 <https://blast.ncbi.nlm.nih.gov/Blast.cgi>

2 <http://www.ebi.ac.uk/ena/data/view/OZ078105-OZ078160>

TABLE 1 Corrected miRNA names and their respective families. Previously annotated miRNAs in miRBase that need to be renamed due to wrong homology assignments.

Previous miRNA name	Updated miRNA name	miRNA family ID
rno-mir-16	rno-mir-16-2	MIPF0000006
rno-mir-365	rno-mir-365-2	MIPF0000061
rno-mir-883	rno-mir-883a	MIPF0000389
rno-mir-17-1	rno-mir-17	MIPF0000001
rno-mir-17-2	rno-mir-106a	MIPF0000001
rno-mir-135a	rno-mir-135a-2	MIPF0000028
rno-mir-199a	rno-mir-199a-2	MIPF0000040
rno-mir-26a	rno-mir-26a-1	MIPF0000043
rno-mir-486	rno-mir-486b	MIPF0000220
rno-mir-3074	rno-mir-3074-1	MIPF0001103

With these discoveries, the updated miRNA repertoire in rats now contains 548 sequences distributed across 341 miRNA families. The complete dataset generated for this study has been deposited at Zenodo³ and GitLab⁴, including sequence files and curated alignments of families with novel miRNA candidates. Additionally, we identified 10 previously annotated rat miRNAs that require renaming due to incorrect homology assignments, as detailed in Table 1. An example of an interesting family requiring the renaming of an existing miRNA and featuring an additional new candidate is illustrated in Figures 1A, B.

3.1 Omics support of new miRNA candidates

Initially, the extended rat miRNA repertoire was generated to provide a more comprehensive miRNA layer for a case study aiming to demonstrate the benefits of multi-omics data integration as part of the CEFIC LRI C5 - XomeTox project⁵. As part of this larger project, we generated short RNA-Seq libraries from 75 rats, examining both thyroid and liver tissues⁶. A detailed description of the methods used is published elsewhere (Canzler et al., 2024).

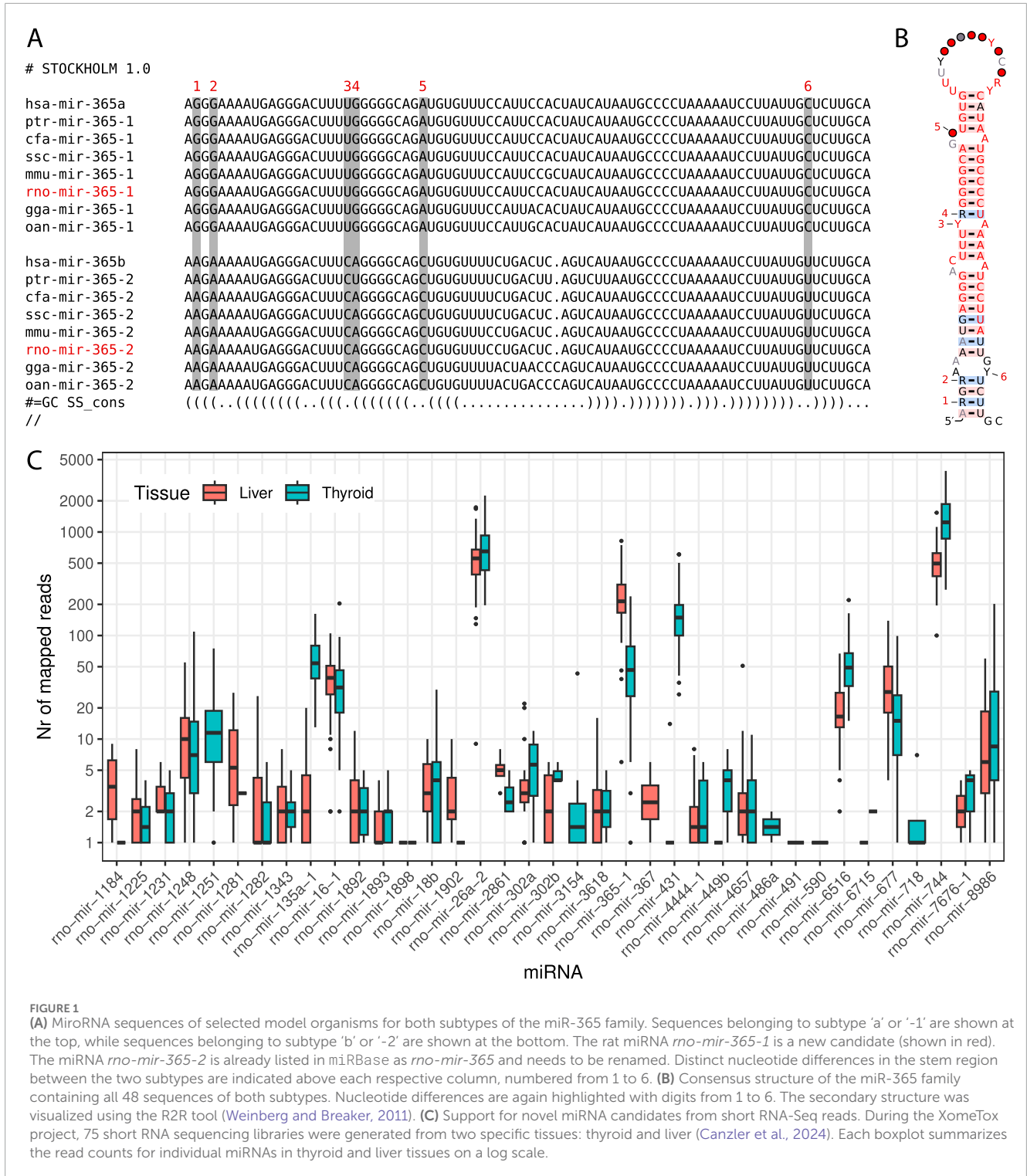
Using the extended miRNA repertoire, we analyzed the short RNA-Seq data to identify support for these sequences across all distinct samples. We discovered 37 miRNAs with overlapping reads in either or both tissues. Specifically, 35 miRNAs had read support

3 <https://doi.org/10.5281/zenodo.12626180>

4 https://codebase.helmholtz.cloud/departement-computational-biology/xometox/extended_rat_mirna_repertoire

5 <https://cefic-lri.org/projects/c5-xometox-evaluating-multi-omics-integration-for-assessing-rodent-thyroid-toxicity/>

6 <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA695243/>



in the thyroid and 32 miRNAs in the liver samples. The read counts for individual miRNA varied significantly, ranging from a few to several thousand per sample, as illustrated in Figure 1C. When miRNAs were detected in both tissues, the read counts were generally comparable.

In summary, this study expands the known miRNA repertoire in *R. norvegicus* by identifying 55 novel miRNAs and correcting

misannotated sequences. By bridging the gap between rat and mouse miRNA annotations, this enhanced dataset, which now includes 341 miRNA families, improves the utility of the rat model. These advancements facilitate more comprehensive transcriptomic analyses, particularly in studies where understanding miRNA-regulated pathways is crucial for assessing molecular responses, such as after exposure to toxins and drugs.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

Author contributions

JL: Data curation, Formal Analysis, Investigation, Writing—original draft, Writing—review and editing. AY: Conceptualization, Data curation, Writing—review and editing. JH: Supervision, Writing—review and editing. SC: Conceptualization, Data curation, Investigation, Writing—original draft, Writing—review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was supported by CEFIC LRI through funding the project C5 - XomeTox.

References

- Ambros, V. (2004). The functions of animal microRNAs. *Nature* 431, 350–355. doi:10.1038/nature02871
- Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116, 281–297. doi:10.1016/s0092-8674(04)00045-5s0092-8674(04)00045-5
- Budakoti, M., Panwar, A. S., Molpa, D., Singh, R. K., Büsselberg, D., Mishra, A. P., et al. (2021). Micro-RNA: the darkhorse of cancer. *Cell Signal* 83, 109995. doi:10.1016/j.cellsig.2021.109995
- Bushati, N., and Cohen, S. M. (2007). microRNA functions. *Annu. Rev. Cell Dev. Biol.* 23, 175–205. doi:10.1146/annurev.cellbio.23.090506.123406cellbio.23.090506.123406
- Canzler, S., Schubert, K., Rolle-Kampczyk, U. E., Wang, Z., Schreiber, S., Seitz, H., et al. (2024). Evaluating the performance of multi-omics integration: a thyroid toxicity case study. *Arch. Toxicol.* 99, 309–332. doi:10.1007/s00204-024-03876-2
- de Jong, T. V., Pan, Y., Rastas, P., Munro, D., Tutaj, M., Akil, H., et al. (2024). A revamped rat reference genome improves the discovery of genetic diversity in laboratory rats. *Cell Genom* 4, 100527. doi:10.1016/j.xgen.2024.100527
- Hussen, B. M., Hidayat, H. J., Salihi, A., Sabir, D. K., Taheri, M., and Ghafouri-Fard, S. (2021). MicroRNA: a signature for cancer progression. *Biomed. Pharmacother.* 138, 111528. doi:10.1016/j.biopha.2021.111528
- Jacob, H. J., and Kwitek, A. E. (2002). Rat genetics: attaching physiology and pharmacology to the genome. *Nat. Rev. Genet.* 3, 33–42. doi:10.1038/nrg702
- Kozomara, A., and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* 42, D68–D73. doi:10.1093/nar/gkt1181
- Lehnert, S., Kapitonov, V., Thilakarathne, P. J., and Schuit, F. C. (2011). Modeling the asymmetric evolution of a mouse and rat-specific microRNA gene cluster intron 10 of the Sfmt2 gene. *BMC Genomics* 12, 257. doi:10.1186/1471-2164-12-257
- Nawrocki, E. P., and Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 29, 2933–2935. doi:10.1093/bioinformatics/btt509
- Okamura, K., Hagen, J. W., Duan, H., Tyler, D. M., and Lai, E. C. (2007). The mirtron pathway generates microRNA-class regulatory RNAs in *Drosophila*. *Cell* 130, 89–100. doi:10.1016/j.cell.2007.06.028
- Smit, A., Hubley, R., and Green, P. (2015). RepeatMasker open-4.0. Available at: <http://www.repeatmasker.org>.
- Weinberg, Z., and Breaker, R. R. (2011). R2R—software to speed the depiction of aesthetic consensus RNA secondary structures. *BMC Bioinforma.* 12, 3. doi:10.1186/1471-2105-12-3
- Yazbeck, A. M., Stadler, P. F., Tout, K., and Fallmann, J. (2019). Automatic curation of large comparative animal MicroRNA datasets. *Bioinformatics* 35, 4553–4559. doi:10.1093/bioinformatics/btz271
- Zhao, H., Sun, Z., Wang, J., Huang, H., Kocher, J. P., and Wang, L. (2014). CrossMap: a versatile tool for coordinate conversion between genome assemblies. *Bioinformatics* 30, 1006–1007. doi:10.1093/bioinformatics/btt730

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.